PUCRS

PONTIFICAL CATHOLIC UNIVERSITY OF RIO GRANDE DO SUL
SCHOOL OF TECHNOLOGY
COMPUTER SCIENCE GRADUATE PROGRAM

GREICE PINHO DAL MOLIN

PREDICTING UNCANNY PERCEPTION IN VIRTUAL HUMANS FACES THROUGH COMPUTER VISION TECHNIQUES

Porto Alegre
2024

PÓS-GRADUAÇÃO - *STRICTO SENSU*

Pontifícia Universidade Católica
do Rio Grande do Sul

Pontifical Catholic University of Rio Grande do Sul
School of Technology
Computer Science Graduate Program

PREDICTING UNCANNY PERCEPTION IN VIRTUAL HUMANS FACES THROUGH COMPUTER VISION TECHNIQUES

GREICE PINHO DAL MOLIN

Thesis submitted to the Pontifical Catholic University of Rio Grande do Sul in partial fullfillment of the requirements for the degree of Ph. D. in Computer Science.

Advisor: Profa. Dra. Soraia Raupp Musse

Porto Alegre
2024

Ficha Catalográfica

D136p Dal Molin, Greice Pinho

Predicting Uncanny Perception in Virtual Humans Faces through Computer Vision Techniques / Greice Pinho Dal Molin. – 2024.
289f.
Tese (Doutorado) – Programa de Pós-Graduação em Ciência da Computação, PUCRS.

Orientadora: Profa. Dra. Soraia Raupp Musse.

1. Visual Perception. 2. Virtual Humans. 3. Uncany Valley. 4. Face Recognition. I. Musse, Soraia Raupp. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da PUCRS
com os dados fornecidos pelo(a) autor(a).
Bibliotecária responsável: Clarissa Jesinska Selbach CRB-10/2051

PREDICTING UNCANNY PERCEPTION IN VIRTUAL HUMANS FACES THROUGH COMPUTER VISION TECHNIQUES

This Doctoral Thesis has been submitted in partial fulfillment of the requirements for the degree of Doctor of Computer Science, of the Graduate Program in Computer Science, School of Technology of the Pontifícia Universidade Católica do Rio Grande do Sul.

Sanctioned on October 16, 2024

COMMITTEE MEMBERS:

Prof. Dr. MARCELO WALTER (UFRGS)

Prof. Dr. BRUNO FEIJO (PUC-RIO)

Prof. Dra. ISABEL HARB MANSSOUR (PUCRS)

Prof. Dra. SORAIA RAUPP MUSSE (PPGCC/PUCRS - Advisor)

“Twenty years from now you will be more disappointed by the things that you did not do than by the ones you did do.”
(Mark Twain)

Acknowledgments

Concluo uma jornada repleta de desafios, aprendizado e crescimento pessoal. A defesa da minha tese de doutorado não seria possível sem o apoio de muitas pessoas que estiveram ao meu lado ao longo dessa trajetória.

Agradeço, primeiramente, a minha família, que sempre acreditou em mim e foi meu alicerce em todos os momentos. Aos meus amigos, que trouxeram leveza e incentivo em cada etapa.

Um agradecimento especial a minha orientadora, cujo conhecimento e orientação foram fundamentais para que este trabalho fosse realizado, e também aos colegas e professores que compartilharam saberes e experiências comigo.

Por fim, agradeço a mim mesma por não desistir, mesmo diante das adversidades. Que essa conquista inspire outros a seguirem em frente, acreditando no poder do conhecimento e da perseverança.

Muito obrigada!

Prevendo a percepção misteriosa em rostos humanos virtuais por meio de técnicas de visão computacional

RESUMO

Atualmente, a crescente presença de agentes conversacionais e humanos virtuais na vida cotidiana tem atraído a atenção de pesquisadores, especialmente no campo da psicologia. A percepção de rostos humanos emerge como um tema relevante e amplamente investigado, especialmente considerando a interação com personagens virtuais. Recentemente, estudos têm explorado a percepção de humanos virtuais, destacando a sensação de estranheza — ou desconforto — que pode ser gerada por determinadas representações, conceito central na teoria do Uncanny Valley (UV). Este fenômeno pode influenciar significativamente nossa discriminação perceptiva e cognitiva, tornando essencial compreender os mecanismos que o sustentam, a fim de mitigar sua ocorrência na modelagem de humanos virtuais. O presente trabalho tem como objetivo examinar a relação entre características faciais e o nível de conforto que os indivíduos experimentam ao interagir com personagens animados gerados por Computação Gráfica (CG). Para isso, propomos e desenvolvemos modelos interpretáveis que identificam áreas específicas do rosto que podem desencadear desconforto, permitindo aprimoramentos que tornem essas representações mais agradáveis tanto visualmente quanto interativamente. O modelo mais eficaz, que utiliza uma técnica ensemble, alcança uma acurácia de 80%. Os resultados deste estudo têm potencial para impactar diversas áreas, como o desenvolvimento de jogos, agentes conversacionais e a indústria cinematográfica, contribuindo para a criação de personagens que evitem provocar estranheza nos

usuários. Para validar nossas abordagens, realizamos experimentos com participantes, coletando dados quantitativos e qualitativos que sugerem que os modelos propostos operam conforme o esperado. Dessa forma, buscamos não apenas avançar no entendimento das interações com humanos virtuais, mas também fornecer diretrizes práticas para a melhoria de suas características, promovendo experiências mais agradáveis e confortáveis.

Palavras-Chave: Percepção visual, Humanos virtuais, Conforto, Vale da estranheza, Reconhecimento facial.

Predicting Uncanny Perception in Virtual Humans Faces through Computer Vision Techniques

ABSTRACT

Currently, the increasing presence of conversational agents and virtual humans in everyday life has attracted the attention of researchers, especially in the field of psychology. The perception of human faces has emerged as a relevant and widely investigated topic, especially considering the interaction with virtual characters. Recently, studies have explored the perception of virtual humans, highlighting the feeling of strangeness — or discomfort — that can be generated by certain representations, a central concept in the Uncanny Valley (UV) theory. This phenomenon can significantly influence our perceptual and cognitive discrimination, making it essential to understand the mechanisms that support it in order to mitigate its occurrence in the modeling of virtual humans. The present work aims to examine the relationship between facial features and the level of comfort that individuals experience when interacting with animated characters generated by Computer Graphics (CG). To this end, we propose and develop interpretable models that identify specific areas of the face that can trigger discomfort, allowing improvements that make these representations more pleasant both visually and interactively. The most effective model, which uses an ensemble technique, achieves an accuracy of 80%. The results of this study have the potential to impact several areas, such as game development, conversational agents, and the film industry, contributing to the creation of characters that avoid causing strangeness in users. To validate our approaches, we conducted experiments with participants, collecting quantitative and qualitative data

that suggest that the proposed models operate as expected. In this way, we seek not only to advance the understanding of interactions with virtual humans, but also to provide practical guidelines for improving their characteristics, promoting more pleasant and comfortable experiences.

Keywords: Visual perception, Virtual humans, Comfort, Uncanny valley, Face Recognition.

List of Figures

List of Tables

List of Acronyms

SVM	Support Vector Machine
SVR	Support Vector Regressor
VC	Voting Classifier
VR	Voting Regressor
UV	Uncanny Valley
CCS	Computed Comfort Score
ROI	Regions of Interest
LIME	Local interpretable model-agnostic explanations

Contents

1. INTRODUCTION

The Computer Graphics (CG) area has stood out in the sophisticated creation of environments and characters. The similarity to the real world surprises both researchers and users in the area of entertainment and areas such as health, law, among others. Assessing the perceived quality of image and video content is important in processing this data in various applications such as movies, games, but also platforms that use images to communicate relevant information [SRLZ14]. Such area of visual perception is highly complex, influenced by many factors, not fully understood, and difficult to model and measure [BLBI13]. For these reasons, subjective assessments are still widely used, in which a group of human viewers qualitatively assesses the images/videos [TC14]. However, some problems may require quantitative assessment when subjective analysis is not possible.

Shahid et al. [SRLZ14] present a literature review on reference-free image and video quality assessment methods. The main objective is to classify and discuss the advances in the field, focusing on approaches that do not require a reference image or video to compare quality. It addresses the main challenges faced by these methods, such as the lack of a universal standard for measuring perceived quality in visual content. The authors categorize the assessment methods into different classes, such as distortion-based, machine learning, and hybrid approaches. They discuss how specific techniques, such as entropy, visual feature statistics, and texture analysis, can be applied to detect visual distortions and predict perceived quality. Authors also emphasize the importance of deep learning-based models, which have shown promising results in dealing with different types of distortions in videos and images. The research suggests that while much progress has been made, there is room for improvement, especially in developing more robust and efficient models that can handle a variety of distortions and adapt to different application contexts.

The perceptual problem we are interested in investigating in this study is known as the Uncanny Valley theory. In the 1970s, Japanese robotics professor Masahiro Mori realized that when human replicas behave very similarly but not identically to real human beings, they provoke revulsion among human observers because subtle deviations from human norms make them look frightening. He referred to this revulsion as a drop in familiarity and the corresponding increase in strangeness as Uncanny Valley [Mor70]. In recent decades, Uncanny Valley has come to be considered in CG characters, whose image analysis can be inspired by

the characteristics of the human visual system (HVS), as mentioned by Sanches et al. [SCMV03]. The authors created a mechanism for extracting guidance resources based on physiological studies of visual perception, seeking to capture the user's subjectivity in relation to the image. Prendinger [PMI05] defines that studying UV in the context of Computer Graphics images is a relevant case study. It is generally agreed that there are characters that cause a bad feeling even if the best techniques are used, as well as characters that cause a good feeling even if advanced techniques are not used, as described in [FdMM+12].

The primary research question of this work is whether the facial characteristics of CG characters, represented through image features, can help determine when these images provoke a sense of strangeness in human perception. Converting human perception into quantitative data, however, presents a significant challenge due to the complexity of human visual perception and the nuances in interpreting facial expressions and proportions. The human ability to detect small anomalies in faces is highly developed, making it challenging to replicate the subtleties of movement, texture, and natural symmetry required for CG faces to appear realistic. Even minor distortions can evoke the Uncanny Valley effect, where a face appears almost human but triggers perceptual discomfort. Factors such as lighting, exaggerated expressions, or a lack of natural variation in microexpressions also contribute to this discomfort, further complicating the process of modeling CG faces. Additionally, each observer may interpret the same facial features differently, making it complex and challenging to develop a model capable of predicting these perceptions.

To address this complexity, we propose and build several methods to test the validity of our hypothesis that image features can correspond with human perception. We introduce original machine learning approaches across four main scenarios. In the first scenario, we use the Support Vector Machine (SVM) [DMND+21b] algorithm to categorize the results into two classes: (1) causes strangeness, or (2) does not cause strangeness—essentially, i.e., a binary classification. In the second scenario, we employ a voting classifier (VC) [ZZCY14] using the same algorithms to extract features from the images again classifying CG character faces into the two classes. In the third scenario, we fine-tune a Convolutional Neural Network (CNN) to compare with the previous methods. Fine-tuning involves training a pre-trained model on a specific dataset, allowing it to leverage prior knowledge (from ImageNet) and adjust its final layers to adapt to new data. The neural networks chosen for this task were VGG16 [SZ15a], ResNet50 [SZ15b], and Mo-

bileNet [HZC+17]. In the fourth scenario, we propose using image entropy and Support Vector Regression (SVR) [Fle09] techniques to calculate the Computed Comfort Score (*CCS*) [DMdAAM22] , a new quantitative metric we propose to evaluate CG faces, resulting in a comfort value. In addition to the whole face, we generate comfort values and test the methods within parts of the face of Virtual Humans to find out which part generates more strangeness. We also tested the Voting Regressor (VR) [PVG+11] technique as a fifth scenario for comparison with the SVR model. In addition to the entropy techniques mentioned in the fourth scenario, we used feature extraction algorithms such as Hu Moments.

Following this line of research, this thesis also aims to ensure the interpretability of the best models suggested to evaluate the characteristics of CG characters' faces and identify possible discomforts in human perception. We used the LIME tool, a technique that can help identify the most important features and their contributions to the model's result in any instance. In this way, we can inform possibilities for adjustments in the areas that cause discomfort, with the aim of providing a more pleasant perception for humans.

1.1 Research Problem

The topic addressed in this doctoral thesis aims to develop a model capable of evaluating the face of a CG character and detecting whether this face can cause discomfort in human perception. The features that are extracted from the face should help detect a pattern that identifies the regions on the face that can generate strangeness in human perception.

Many studies evaluate the strangeness perceived by participants based on subjective analyses, using forms and performing some statistical analyses based on people's responses, such as Ho et al. [HMP08], Tinwell et al. [TGNW11], Flach et al. [FdMM+12], Victor et al. [AMDM21] among others. However, to our knowledge, no studies propose the analysis of discomfort in CG characters, estimating subjectivity perception based on computational methods and image characteristics. Researchers such as Limano [Lim19] recognize this is a difficult problem because it is not obvious what makes a particular image comfortable or uncomfortable for viewers. Indeed, it is easy for humans to judge quickly on images in seconds because various factors are implicitly used in this process. These factors can be related to the individual's experience and training throughout life. Diel et al. [DL22] investi-

gate how familiarity, orientation, and realism affect the perception of strangeness in faces, focusing particularly on how these variables sensitize observers to facial distortions. The authors propose that greater familiarity, appropriate orientation, and a high degree of realism increase the sensation of strangeness by making observers more perceptive of imperfections and distortions in faces. Another study of Diel et al. [DL24] revisits the uncanny valley concept, proposing a perspective in which the effect is better represented as a moderated linear function rather than a nonlinear curve. The authors argue that observers' perceptual specialization amplifies the perception of strangeness in facial distortions. Using an experimental approach, participants were exposed to a series of images of faces varying in degree of distortion. Perceptual specialization was manipulated through training tasks that increased participants' sensitivity to facial features. MacDorman [Mac24] carried out a meta-regression analysis to measure the relationship between the humanization (and dehumanization) of artificial beings and the perception of eeriness. The results showed that humanization had a non-significant effect on the eeriness. In this case, dehumanization means reducing the anthropomorphic level (human characteristics) of the artificial being and is a theory originating in the field of Psychology [KL22]. In other words, a human being feels uncomfortable when the artificial being does not have human characteristics. So, understanding how humans perceive and interpret visual information is crucial for enhancing the realism, believability, and overall quality of VHs and environments, so that the audience (human beings) feel good and comfortable. Furthermore, as MacDorman's work showed the importance of human characteristics in UV theory, it also shows how the stimulus is related to causing strangeness or not. Considering that the stimulus that a VH has is important for the perception of eeriness, some studies have used eye tracking to measure regions of interest in VHs in relation to UV theory. Studies by Cheetham et al. [CW19], Schwind et al. [SWH18], and Grebot et al. [GCdL+22] showed that regions of the nose, eyes, and mouth are important areas for transmitting or not strangeness to those observing VHs. As CG progresses, the consideration of human perception will remain an important aspect in creating more immersive and compelling virtual experiences.

The main hypothesis of this work is related to extracting features that can be used to detect the strangeness perceived by humans. Although there are many studies related to the Uncanny Valley theory, there seem to be many paths to explore in relation to human perception. Another concept that can be used to justify our process is the quality of the perceptual image. The work of Lin et al. [LK11] contributes to assessing the quality of images/videos in CG and animation by deal-

ing with Perceptual Visual Quality Metrics (PVQMs). They suggest that modeling human perception can play a crucial role in many CG tasks, as Tumblin and Ferwerda [TF01] argue that the purpose of CG is to manage how humans perceive it rather than merely controlling light.

Despite the significant progress made so far, no interpretable model has been developed that can both detect awkwardness or discomfort in characters faces and suggest improvements in designing by identifying expressive features, such as facial proportions, cartooning effects, or facial movements. To address these challenges, we propose developing a model based on the extraction of relevant features from characters faces to better understand and identify human perception. The main and specific objectives of this thesis are outlined below.

1.2 Goals

The primary objective of this thesis is to develop a computational model that extracts various features from the faces of CG characters to predict the sense of strangeness that humans may perceive. By analyzing the predicted perceived comfort and providing an explanation for such perceptions, designers can improve the creation of CG characters for use in various fields, such as gaming, film, healthcare, and legal contexts. This ensures that interactions with these virtual agents will be smoother and more natural, making people feel more comfortable, even during potential interactions.

1.2.1 Specific Goals

In order to achieve the main goal, some specific goals are proposed, as follows:

- **Creation of Dataset with images/videos**. It is composed of faces of virtual humans that were subjectively evaluated by humans who perceived strangeness in some of these characters and not in others. This dataset is composed of 40 characters and 13402 images
- **Study of Image Feature Extraction Techniques**. Here, we investigate many techniques from face detection using Viola-Jones [VJ+01], OpenFace [BRM16]

and MediaPipe [LTN+19] to other extractors such as Saliency ([JZW18]), Hu Moments [ŽHR10], Histogram of Oriented Gradients (HOG) [DT05], Image Entropy [Spo96], Golden Ratio [SMS08], among others.

- **Study of machine learning techniques**. Many possible techniques could be used. We used the Support Vector Machine (SVM) [Fle09], Support vector Regressor (SVR) algorithm also using Sklearn [1] and the voting classifier [ZZCY14] and regressor [KS14] techniques. In this thesis, we also investigated the Convolutional Neural Networks (CNNs). VGG16 [SZ15a], Resnet50 [SZ15b] and MobileNet [HZC+17] were the neural networks chosen for this task.
- **Proposal of metrics to compute the perceived comfort**. While in some of our studies, we investigate the binary classification of perceived comfort, we also propose a way to estimate a perceived comfort value in the interval [0;100].
- **To produce texts and scientific contributions**. We intend to produce new ground in this challenging area through new publications.

1.3 Text Structure

The text is divided into six chapters. This chapter presents an introduction to the subject of this thesis, presenting the research problem and its relevance, as well as the goals of this work.

Chapter 2 present several works related to the theme of this thesis. Such works involve concepts related to human perception, uncanny valley theory, animated CG characters, and perceptual image quality.

The Chapter 5 presents the proposed model with the goal of building a computational model for assessing discomfort as perceived by humans.

Chapter 6 presents and discusses the experimental results achieved by this work. Such results involve tests on each resource, as well as to find out if there is an aim perception of discomfort equivalent to the subjective process of research with the human being.

Finally, Chapter 7 concludes the work and presents the final considerations.

[1] https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html

2. RELATED WORK

This chapter presents some methods that borrow a theoretical foundation for this thesis and are related to the goals this work aims to achieve.

The organization of the chapter is following described: First, we present a concept of perception according to Tumblin and Ferwerda [TF01], then Section 2.1 discusses Mori's theory of Uncanny Valley. Section 2.2 discusses the impact of visual perception area on some industries. Shows some work that deals with the strangeness of animated characters, with asynchronicity of eye movements, mouth movements, audio, and visual being mentioned, among others.

In Section 2.3, we present some authors that work with methodologies used in the literature on the perception of video/image quality.

The proposed concept of Tumblin and Ferwerda [TF01] is well suited for this research:

> "Perception connects our minds to the world around us. And the *host* of processes that converts all the measurable physical stimuli that our bodies receive into an awareness of our environment. Its inputs are physical and measurable, but its outputs are purely psychological. Perception gives us an immediate, moment-by-moment estimate of reality and provides the basics of where we are and what is happening around us—the initial information needed to understand and interact with our environment. Perception is much more than a simple measurement of physical stimulus. It is not a passive measurement of light, sound, pressure, or chemical vapors that impinge on our sensory organs. It is a set of processes that actively construct mental representations of the world from raw, noisy, and incomplete sensory signals."

2.1 The Uncanny Valley Theory

In 1970, Masahiro Mori [Mor70], a pioneer in Japanese robotics, proposed a hypothetical graph predicting that as a robot's appearance becomes more human-like, it also becomes more familiar and appealing—until it reaches a point where subtle imperfections make the robot appear unsettling, just before achieving full human likeness. The examples tested are illustrations of Ishiguro's [Ish06] Uncanny

Valley theory. Mori [Mor70] associated the sense of eeriness with the robot's appearance and observed that as a robot's resemblance to a human increased, it became more familiar and pleasing to viewers. However, at a certain threshold (around 80% human likeness), the robot was perceived as more strange than familiar. When a robot's appearance closely resembled a human, but still fell short, it triggered a negative emotional response in the viewer. Figure 2.1 shows a visualization of Mori's Theory, showing familiarity steadily increasing (y-axis) as perceived human resemblance increases (x-axis) and then sharply decreasing, causing the trough.

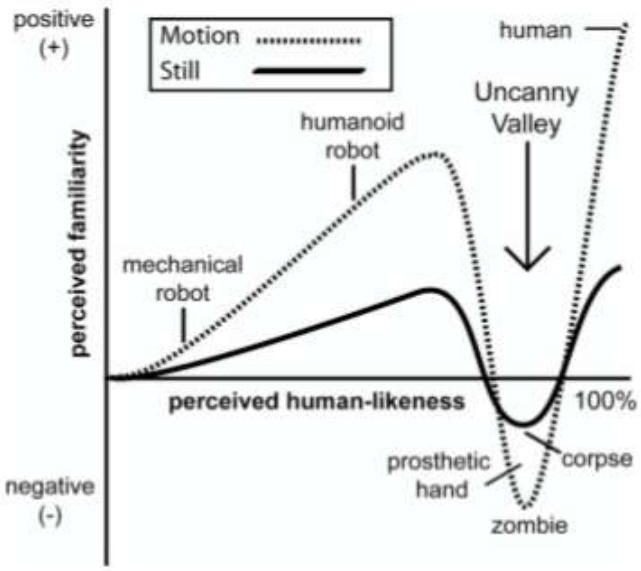


Figure 2.1: Mori's theory of perceived familiarity as a function of human likeness until the Uncanny Valley effect occurs. Source: Prakash et al. [PR15].

Regarding the psychology of the stranger, Freud [Lyd97] characterized the strange as a feeling caused when it is impossible to detect whether an object is animate or inanimate when encountering objects such as wax dolls. According to Ho et al. [HMP08], the feeling of strangeness perceived in robots with human-like appearance and animated characters can be a key factor in our perceptual and cognitive discrimination.

For Tinwell et al. [TGNW11], the idea is like that of Ho et al. [HMP08], i.e., the phenomenon of Uncanny Valley means that the virtual characters are too similar to humans, evoking a negative reaction to the observer, as they look and behave differently from what would be considered a common pattern in human beings. In the article by Prakash et al. [PR15], one of the main objectives was to investigate people's initial perceptions of robots when there is a lot of human similarity in the

robot's face. Even more so when the robot aims to assist in tasks commonly performed by humans. The results showed that people's perceptions of robot faces vary as a function of the robots' human resemblance. People tend to generalize and create expectations about the behavior and capabilities of a human-looking robot.

2.2 Uncanny Valley and Computer Animation Industry

The Uncanny Valley study is not just concerned with the acceptance of robots by humans. The animation and special effects industries are also concerned that studios could lose money and audiences if audiences cannot relate to CG characters in animations because of Uncanny Valley. According to Hanson et al. [HOP+05], new challenges for computer animation and simulations include contextualization in conversations, human perceptions and environments, and having control over motives or decisions. These needs are associated with today's increasingly improved graphic realism, which, as Prakash [PR15] says, makes humans expect more realistic behavior.

Today, computer animation is increasingly used to address ethical and moral issues in both the legal and medical professions and even for recruitment, as Von Bergen et al. [Von10], have reported. Therefore, there was concern about evaluating the appearance and behavior of CG characters in the context of the Uncanny Valley, which is associated with human likeness and has numerous applications, as demonstrated by Tinwell et al. [TGNW11]. Through several studies in this area of animation, some characteristics in CG characters already demonstrate greater strangeness to the human being, as follows:

- Actions perceived as unnatural, such as rigid or abrupt movements, in the study by Bailenson et al. [BSH+05];
- Lack of human similarity in the speech and facial expression of a character, in the studies by Tinwell et al. [TGNW11];
- Perception of lip-sync error that may be voice before lip movement or vice versa, according to studies by Gouskos et al. [Gou06]

The Uncanny Valley evoked by animated characters may have similar origins to any man-made objects that mimic life as we know it, such as realistic, human-like but virtual characters. This nature of objects (animate or inanimate) may be

perceived as less trustworthy, according to Kang [Kan09], as there is ignorance and unpredictability in terms of expected behavior.

Schein et al. [SG15] investigate the relationship between mind perception, the experience of the uncanny valley, and social and emotional responses to human-like entities. This work explores how mind perception, or lack thereof, contributes to this response. Through a series of experiments, they showed that eyes that appear empty or devoid of life increase the feeling of uncanny, suggesting that mind perception is strongly linked to emotional responses to the uncanny valley. Furthermore, the study relates this information to characteristics of the autism spectrum, where the perception and interpretation of eyes and facial expressions can be different. People with autism may experience the uncanny valley differently due to these differences in mind perception. This suggests that variations in sensitivity to detecting the mind through two eyes may significantly influence social and emotional responses.

Schwind et al. [SWH18] addresses the challenges of creating virtual characters that avoid the uncanny valley effect. The authors explore several realistic aspects of this phenomenon, discussing how small imperfections in highly important people can lead to negative emotional responses. They highlight the financial implications of this effect, citing examples such as the film Mars Needs Moms [1] and video games such as L.A. Noire [2] and Mass Effect: Andromeda [3], which have faced criticism and financial losses due to the disturbing design of their characters. The research suggests several strategies for mitigating the uncanny valley effect. These include focusing on beautiful realism in movements and expressions, as well as ensuring consistent levels of detail across different aspects of character design. The goal is to create characters that are stylized enough to avoid direct comparison to real or realistic humans, or enough to avoid perceived errors that cause discomfort.

The work proposed by Flach et al. [FdMM+12] has the hypothesis of verifying whether CG characters suffer from Uncanny Valley, as did the beings tested by Mori [Mor70]. Based on this objective, when trying to analyze the results of the subjectivity of the video/image quality in terms of the Uncanny Valley effect, the authors selected and evaluated some characters, following some criteria such as:

- Human similarity of each character, with some being chosen with lesser and greater human precision. Example: cartoons have little similarity;

[1] https://disney.fandom.com/wiki/Mars_Needs_Moms
[2] https://www.rockstargames.com/br/games/lanoire
[3] https://www.ea.com/pt-br/games/mass-effect/mass-effect-andromeda

- If it is public knowledge, considering the origin of the character, which can be a movie, a game or an unknown origin; and
- Restrictions on videos, the characters could not have strong emotions, they should be in natural environments, with common clothes, to avoid perceptual distortions.

After analyzing 12 characters, the authors [FdMM+12] observed the creation of the graph Figure 2.2.

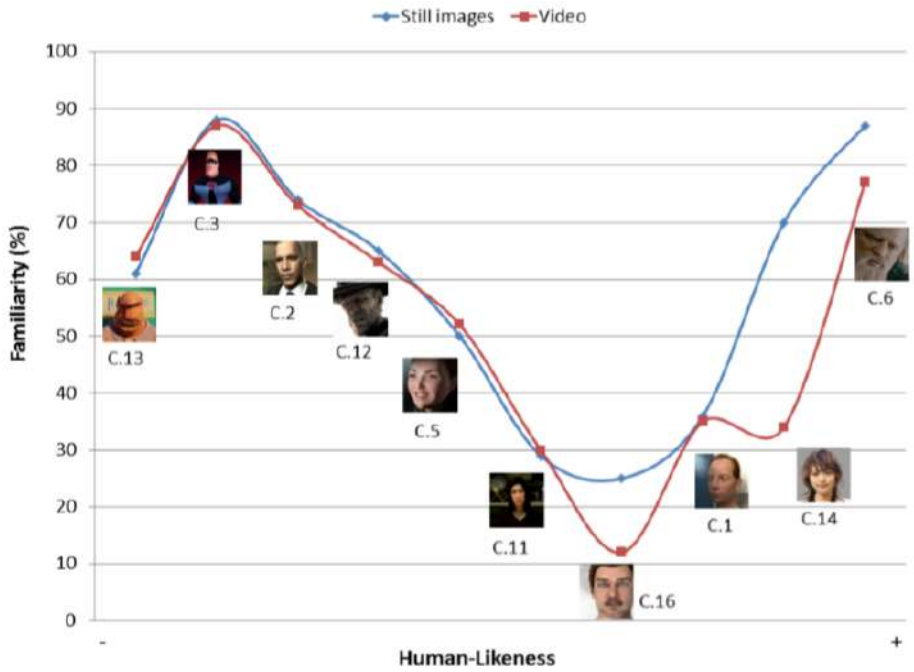


Figure 2.2: This graph resulted from Flach et al. [FdMM+12] is similar to the original curve of Mori's Uncanny Valley [Mor70]. The vertical axis indicates the percentage of people who felt comfortable with the characters, while the horizontal axis shows the character's resemblance to humans.

There are many studies conducted with CG characters to evaluate empathy between human characters. The research by Prendinger et al. [PMI05] investigates empathy between the virtual character and the human and shows, as a result, the reduction of stress in the human's perception, when developing activities with empathetic characters. An example is medical training systems in which it is essential that human characteristics are imitated with high precision in virtual characters, in order to obtain positive responses from participants, as shown in the work of Robb et al. [RKA+13]. The work of Dunsworth et al. [DA07] also shows that it is possible to generate an emotion in the participant when working with learning systems,

having a relevant influence on memory retention. This shows that if a system can extract emotions from a participant, then the probability of associating emotions with activities can be relevant.

In the article by Araujo et al. [AMDM21], the authors conducted a perceptual study, which analyzed people's perception of characters created using Computer Graphics from different media (films, series, animations, simulations, among others). The objective was to find out if people today feel more comfortable with CG characters than people in the past. Araujo et al. replicated the work of Flach et al. [FdMM+12], analyzing characters from different media. In addition, they included current characters to compare perceptual data from 2012 and 2020. The results show that, in some cases, people today are more comfortable with current characters. However, the perception of comfort with old characters was similar in both periods, indicating that characters from older technologies still generate comfort. The figure 2.3 shows the virtual characters worked on by Araujo et al. [AMDM21].

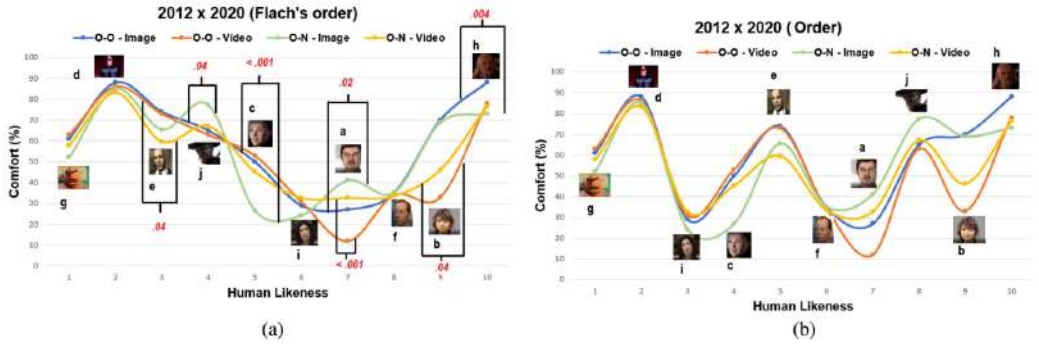


Figure 2.3: All the characters used in the work of Flach et al. [FdMM+12] with Flach's order in (a), and Araujo's order [AMDM21] in (b). Both blue and orange lines, in (a) and (b), represent the percentages of comfort of each character in image and video, as perceived in 2012. The green and yellow lines represent the same in (b), however, evaluated in 2020. In addition, in (a), It can see the significant results (highlighted in red) of the comparisons of the characters perceived in 2012 and 2020 (the results related to images were above the lines, the results related to videos were below the lines).

The study by Ho et al. [HM10] compares animated human characters and robots, comprising a total of 10 videos. Five characters are animated, as shown in Figure 2.4: (1) Final Fantasy: The Spirits Within, (2) The Incredibles and (3) The Polar Express, (4) an Orville Redenbacher popcorn advertisement and (5) a technological demonstration of the video game Heavy Rain. In this study, characters (1) and (3) are shown as generators of strangeness in human perception. Character (4)

left some participants disturbed by the digital resurrection of entrepreneur Orville Redenbacher. Other participants accepted the character as a real person. Therefore, the authors consider it relevant to explore demographic factors that can influence the intensity of emotional responses. Therefore, according to the research, characters (1), (3) and (4) would cause strangeness and characters (2) and (5) would not cause the same feeling.

Figure 2.4: The five video clips on the top row contain computer-animated human characters from the films (1) Final Fantasy: The Spirits Within, (2) The Incredibles, and (3) The Polar Express, (4) an Orville Redenbacher popcorn advertisement, and (5) a technology demonstration of the Heavy Rain video game. The remaining five video clips contain (6) iRobot's Roomba 570, (7) JSK Laboratory's Kotaro, (8) Hanson Robotics's Elvis and (9) Eva, and (10) Le Trung's Aiko, [HM10].

Katsiry et al. [KMT17] hypothesize that semi-realistic film characters are more acceptable in Uncanny Valley, receiving a much higher weirdness rating than other characters. Characters such as Beowulf and The Polar Express are included in this list. Their goal is to include a comprehensive set of motion-capture animated film characters in the semi-realistic animation category. This study is based on the input of fifty-four participants, who were asked to rate five parts of films related to cartoons, semi-realistic, and human-action films. Fifteen characters are used in this research, of which 5 are semi-realistic, 5 are cartoons, and 5 are human characters. Only the first 5 will be mentioned, such as Final Fantasy: The Spirits Within ("Aki Ross"), Polar Express (nameless boy), Beowulf ("Beowulf"), Mars needs moms ("Milo"), The Adventures of Tintin: The Secret of the Unicorn ("Tintin"), The Incredibles ("Mr. Incredible"), Meet the Robinsons ("Lewis"), Cloudy with a Chance of Meatballs ("Flint Lockwood"), Arthur Christmas ("Arthur Christmas"), Epic ("MK"). The results of the research show that the more semi-realistic the character, the stranger it seems to people. According to the research, of these 5 semi-realistic characters, Beowulf and Polar Express were the ones that caused the most strangeness in the participants, but all 5 cause strangeness perceptually. The method used to evaluate the

strangeness of the characters was through questionnaires to the participants and was related to motion capture. In semi-realistic animated films, authors linked the inclusion criteria to the fact that these films are fully computer-animated, utilize motion capture techniques, and intentionally strive for high levels of human likeness. The decision to include motion capture animation as a criterion was based on the observation that many films, such as The Polar Express, have faced common criticism within the Uncanny Valley (UV) context due to their use of these techniques.

The study by MacDorman et al. [MGHK09] does not directly examine CG characters resulting from films or videos. The goal of the study is to look for some characteristics that can overcome the Uncanny Valley, trying to discover some of its causes and propose design principles to help photorealistic human characters escape the uncanny valley. Four studies are conducted dealing with facial proportions, skin texture, and the level of detail of a computer-generated human character that were varied to examine their effect on the perception of uncanny similarity. The cause of UV indicated in the study refers to the involvement of affective and motor processing that are simultaneously active in the perception of human-like forms. Although the study by MacDorman et al. [MGHK09] does not refer to CG characters in films and video games, which is the focus of our research, it cites many CG characters in literature. The authors consider that computer graphics (CG) characters are challenging our ability to discern what is human. They cite characters that were created with the intention of making people uncomfortable, such as the CG character Davy Jones from Pirates of the Caribbean: At World's End, who was created to be scary and seem supernatural. The same idea applies to the CG villain Gollum in The Lord of the Rings trilogy. On the other hand, characters designed to look like real people have been less convincing, such as the CG heroes in The Polar Express and Final Fantasy: The Spirits Within. Through this study, we identified 4 characters that are both unfamiliar to the audience and generally frightening.

According to the research by Geller et al. [Gel08], one of the methods they analyze is motion capture (mocap). The authors define the essential idea of mocap as being a way of tracking the actual movements of human performers, through the use of dozens or hundreds of trackable points on their bodies, and then converting the tracked data into vectors that effectively replicate their movements. The authors consider that vectors can serve as automated or manual guides for traditional animation.

The opinion of Hal Hickel, the animation supervisor at Industrial Light & Magic who worked on the Pirates of the Caribbean films, is presented in this study

and indicates that mocap overcame virtually all of the extraneous effects in replicating gross body movement. Hickel considers facial movement to be the most problematic area of the body. In the 2002 film Harry Potter and the Chamber of Secrets, he further improved the appearance of CG skin, giving it a realistic translucency. According to Hickel, mocap eliminates the uncanny valley of body movement, but still impairs the eyes and facial performance. The study by Geller et al. [Gel08] points out some characters referenced in literature that scare people. In the film The Polar Express [4], the authors show that there would be a need for filmmakers to stylize their characters away from realism in order to make them effective. The films Beowulf (Grendel) [5] and Lord of the Rings (Gollum) [6] show that a good way to avoid the Uncanny Valley would be to change the proportions and structure of a character. This is a justification for Gollum's success, as he has large eyes and a non-human face shape. Regarding Grendel in Beowulf, the authors' justification for not causing strangeness is that the character is disfigured and deformed. In this way, the audience's subconscious would consider him non-human. But when you evaluate the character as human, the viewers realize what is missing.

MacGillivray et al.[Mac07] mapped popular animated characters, creating graphs for image and movement. They explored how these elements influence empathy, using Mori's Uncanny Valley theory[Mor70] as a basis. The study analyzed popular animations to understand the gap between what people see and what they perceive, highlighting the importance of imagery, movement, and timing in evoking empathy. The research also noted that while abstract symbols take time to decode, perceived images are instantly understood, regardless of cultural background. The study included characters such as Dumbo, Bugs Bunny, Bambi, and others. Of the ten characters analyzed, only those from Polar Express and Final Fantasy caused discomfort.

Yekti's study [Yek15] compares the Uncanny Valley theory in 3D stop-motion animation and 3D computer-generated (CG) animation, analyzing their aesthetics. In stop-motion, the physical and tactile characteristics are related to real touch, material and texture, while in 3D CG animation, these characteristics are linked to realism and verisimilitude. The author discusses how imperfection in stop-motion is seen as a charm, while in CG animation, technical flaws are perceived as defects, often generating discomfort, exemplified by the films The Polar Express and Final

[4] https://www.warnerbros.com/movies/polar-express
[5] https://pt.wikipedia.org/wiki/Beowulf_(2007)
[6] https://www.warnerbros.com/movies/lord-rings-fellowship-ring

Fantasy [7]. The study also addresses four films: Coraline [8] and Fantastic Mr. Fox (stop-motion) [9], and The Adventures of Tintin [10] and Up [11] (CG), highlighting the uncanny observed in scenes from The Adventures of Tintin, despite the use of motion capture.

The study by Mustafa et al. [MGT+17] investigates Mori's hypothesis that uncanny effects increase with movement in computer graphics (CG) characters. Using electroencephalography (EEG), they analyzed the neural responses of participants watching videos of CG characters and real humans. The results showed clear differences in brain responses to highly realistic CG faces, such as Digital Emily, compared to real humans, indicating an "uncanny" response. From these responses, they trained a support vector machine (SVM) to categorize characters based on EEG data, predicting whether they would be perceived as uncanny. The study included realistic characters from games such as Detroit: Become Human [12] and tools such as the Virtual Human Toolkit.

Tinwell et al. [TGW10] study explores the relationship between the perception of uncanny in virtual characters and human-likeness in movement and sound attributes, particularly in survival horror games. With 100 participants, videos of 12 virtual characters and one human were evaluated, and the results showed that exaggerated movements and strange sounds accentuate the Uncanny Valley effect, making the characters more frightening. The horror genre was highlighted as an example of where uncanny can be purposefully exploited to provoke fear in players. Participants rated the characters on scales of 1 to 9, considering how human and uncanny they seemed, as well as aspects such as voice synchronization and lip movement. Characters were rated according to familiarity (characters 1 to 6 and 13) or uncanny (characters 7 to 12), with characters 2 and 3 on the threshold between these categories. The authors highlight how exaggerated facial expressions or lack of synchrony between sound and movement can intensify the perception of strangeness in virtual characters. Figure 2.5 shows the virtual characters and a real human worked by Tinwell et al. [TGW10].

[7] https://en.wikipedia.org/wiki/Final_Fantasy:_The_Spirits_Within
[8] https://www.laika.com/our-films/coraline
[9] https://pt.wikipedia.org/wiki/Fantastic_Mr._Fox
[10] https://www.tintin.com/en/videos/460/the-adventures-of-tintin-trailer-
[11] https://movies.disney.com/up
[12] https://www.quanticdream.com/en/detroit-become-human

Figure 2.5: The videos include six realistic, human-like characters: (1) Emily Project (2008a) (2) and the Warrior (2008b) by Image Metrics; (3) Mary Smith from Quantic Dream's technical demo, 'The Casting' (2006); (4) Alex Shepherd from Silent Hill Homecoming (Konami 2008); two avatars, (5) Louis and (6) Francis, from Left 4 Dead (Valve 2008); four zombie characters, (7) a Smoker, (8) The Infected, (9) The Tank and (10) The Witch, from left 4 Dead; (11) a stylized human Chatbot character, 'Lillien' (Daden Ltd. 2006); (12) a realistic, human-like zombie ('Zombie 1') from the video game, Alone in the Dark (Atari Inc. 2009); and (13) a real human, [TGW10].

2.3 Image Quality

Some areas define characteristics to analyze image quality, as presented by Shahid et al. [SRLZ14]. For the image quality level to be determined, it is important to consider the application. For example, in the case of image compression applications such as Nill et al. [Nil85], less traffic will be generated on the network, taking up less storage space, with a reduction in quality. In the case of videos, compression occurs through similarity analysis in neighboring frames.

Among the factors to be considered is the amount of colors, which has a direct relationship with the viewer's comfort. The consequence is a greater number of bits of information to be transmitted. Assuming that the application is videoconferencing, the use of grayscale does not harm the service provider. But for a TV network, this quality would not be acceptable.

The image resolution must also be evaluated, which is related to the amount of *pixels* between rows and columns, as Gu et al. [GLZ+15] discusses. Resource that influences the shapes of the image, being natural for those who see them. The fewer *pixels*, the more jagged the image, and the rounded shapes will lose quality for lack of *pixels*. Again, it is necessary to evaluate the acceptable amount of *pixels* between rows and columns, depending on the application.

Finally, we must consider the number of frames per second, as Pinson et al. [PW04] treats. Studies show that the minimum acceptable frame rate for a person to assimilate the movements of a video as natural is 24 frames per second. For this reason, providers need to decide the number of acceptable frames per second, as is the case with TV movies, remote videos that need 24 or more frames per second.

2.3.1 Objective Image Quality Assessment

The purpose of Objective Image Quality is to design mathematical models that can predict the quality of an image accurately and also in an automated way. The proposed methods must be able to mimic the quality predictions indicated by the average of human observers. According to Wang et al. [WB06], the methods can be classified into three categories:

- Full-reference image quality assessment (FR-IQA): the reference image is fully available. The scope of application of these metrics includes image compression, according to Ma et al. [MLN10], inclusion of watermark that can distort the image, as Zhang et al. [ZLLN11] treated, and so on. against. Here are some methods that are used:
  - mean squared error (MSE): denotes the power of distortion, ie the difference between the reference and test images.
  - structural similarity index (SSIM): assumes that the HVS is highly adapted to extract structural information from a scene.
  - multiscale structural similarity index (MS-SSIM): it is the same as the SSIM but in multiple scales. The advantage of multiple scale methods like MS-SSIM over single scale methods like SSIM is that in multiple scale methods, image details at different resolutions and display conditions are incorporated into the algorithm. of quality assessment.
  - visual information fidelity (VIF): models natural images in the wavelet domain using Gaussian scale mixtures (GSMs).
  - most apparent distortion (MAD): assumes that the HVS employs different strategies when judging image quality.
  - feature similarity index (FSIM): is based on the fact that HVS understands an image mainly because of its low-level features, eg edges.

- Reduced-reference image quality assessment (RR-IQA): only partial information about the reference image is available. Several features are taken from the reference image. These features are used as secondary information to assess the quality of the test image. There are several applications, such as: tracking the level of visual quality degradation of image and video data transmitted by visual communication networks in real time. According to Rehman et al. [RW12] this category can be classified into three methods:
  - Methods based on the models of the image source: are statistical models that capture low-level statistical features of natural images.
  - Methods based on capturing image distortions: The methods in this category are most useful when enough information about image distortions is available.
  - Methods based on the models of human visual system: in designing methods in this category, physiological or psychophysical studies can be used.
- No-reference image quality assessment (NRIQA): Neither the reference image nor its resources are available for quality assessment. However, humans can often efficiently assess the quality of a test image without using any reference images. This is probably due to the fact that our brain contains a lot of information about what an image should or should not look like in the real world, according to Wang et al. [WB06].

According to Shahid et al. [SRLZ14], image quality metrics can also be classified according to only a specific type of degradation, for example: blur, block or touch. One can also take into account all possible signal distortions, i.e. various types of artifacts. Here are some attributes:

- Sharpness determines the amount of detail an image can convey. The sharpness of the system is affected by the lens (quality of design and fabrication, focal length, aperture and distance from the center of the image) and sensor (count of pixels and anti-aliasing filter). Can be affected by camera shake (a good tripod can be helpful), focus accuracy, and atmospheric disturbances (thermal effects and aerosols).
- Noise is estimated as the difference between the image and a median-filtered version of the image.

- Dynamic range (or exposure range) is the range of light levels that a camera can capture. It is closely related to noise.
- Tone reproduction is the ratio between the luminance of the scene and the brightness of the reproduced image.
- Contrast, also known as gamma, is the slope of the tone reproduction curve in space. High contrast often involves loss of dynamic range - loss of detail or clipping, in highlights or shadows.
- Color accuracy is an important but ambiguous image quality factor. Many viewers prefer enhanced color saturation; the most accurate color is not necessarily the nicest. However, it is important to measure the color response.
- Distortion causes straight lines to curve. Can be problematic for architectural photography and metrology (photographic applications involving measurement). Distortion tends to be noticed on low-end cameras, including cell phones and low-end DSLR lenses.
- *Vignetting, or light falloff* is the process of darkening images near the corners. It can be significant with wide-angle lenses.
- Exposure accuracy can be an issue with fully automatic cameras and with video cameras where there is little or no opportunity for tonal adjustment after exposure. Some even have exposure memory. Exposure may change after very bright or dark objects appear in a scene.
- Lateral chromatic aberration (LCA) causes colors to focus at different distances from the center of the image. It is most visible near the corners of images. LCA is worse with asymmetric lenses, including ultra-wide, true telephotos and zooms. It is strongly affected by demosaicing.
- Lens flare is diffused light in lenses and optics caused by reflections between lens elements and the inner lens body. This can cause image blurring (loss of shadow detail and color), as well as "ghosting" images that can occur in the presence of intense light sources in or near the field of view.
- Color moiré is an artificial color band that can appear in images with repetitive patterns of high spatial frequencies, such as fabrics or fences. It is affected by the sharpness of the lens, the anti-aliasing (low-pass) filter that smoothes the image. It tends to be worse with sharper lenses.

A review of methodologies used in the literature was carried out by Lévêque et al. [LLB+18] on the perception of video/image quality. The author comments that despite the continuous evolution in how to view content acquired, stored, accessed by users, distortions still occur. In addition, Lévêque [LLB+18] also comments on the perceptual quality in images and videos that can be affected by human influencing factors that refer to human characteristics. They can be classified into two categories:

- Low-level which refers to the processing of factors of physical, emotional and mental influence of the human being.
- High-level which refers to the processing of demographic and socio-economic influence factors.

Estimating subjectivity and perception in image quality, according to Schuyler et al. [Smi] is useful in many areas. An example might be in creating CG characters. If there is a precise measure, the entertainment industry can use this metric to avoid the discomfort that the human being can feel in relation to a certain character.

For these reasons, we believe that using statistical characteristics of the images from the image quality area, treated in Liu et al. [LLHB14], can prove promising in the assessment of comfort of the animated characters' faces. Support vector regression (SVR) is used to predict the average human opinion score on comfort with these various NSS features as input.

Researchers acknowledge that this is a difficult problem, because it is not obvious what makes a particular image comfort or discomfort to viewers. It is easy for human beings to quickly judge images in seconds, because several factors are implicitly used in this process, which are related to the individual's experience and formation throughout his life.

2.3.2 The Intersection of Image Quality and the Uncanny Valley Theory

Image quality and the Uncanny Valley theory are fundamental concepts that address the viewer's visual perception and emotional experience. While image quality refers to the clarity, color fidelity, and other attributes that contribute to the visual appreciation of an image, the Uncanny Valley theory suggests that near-realistic representations of humans can elicit a sense of eeriness or discomfort.

Image quality is a crucial factor in the aesthetic experience, directly influencing the viewer's emotions. High-quality images generally generate positive emotional responses, while low-quality images can result in disinterest or discomfort. This phenomenon is closely related to the Uncanny Valley theory, which proposes that as a digital representation approaches human realism, small imperfections become more evident, causing aversion. When an image is near-realistic but has flaws, the viewer's frustration can be heightened, resulting in a negative visual experience.

In addition, both image quality and the Uncanny Valley theory have significant implications for graphic design, robotics, and animation. In graphic design, attention to image quality is essential to creating content that resonates with audiences. Designers who ignore these aspects can inadvertently trigger the Uncanny Valley by creating human representations that fail to capture the desired authenticity. In robotics and animation, it is vital to find a balance between realism and style, avoiding the emotional disconnect that the Uncanny Valley can create.

The connection between image quality and the Uncanny Valley also highlights the importance of aesthetic perception and emotional response. While image quality focuses on technical elements, Uncanny Valley theory addresses the limitations of similarity and the psychological effects of near-human representations. This relationship highlights that, in any art form or technology, understanding and applying these concepts can improve viewer engagement and promote a more engaging and satisfying experience.

For this reason, we believe that the relationship between image quality and Uncanny Valley theory is a rich and relevant area for understanding visual and emotional perception. Understanding how these concepts interact can not only improve the quality of visual representations, but also contribute to the creation of more authentic and engaging experiences across disciplines ranging from art to technology. The quest for high-quality visual representations that avoid the Uncanny Valley trap is therefore a significant and valuable challenge for creators and designers in the contemporary world.

2.4 Chapter Considerations

This chapter has presented many works related to what is being proposed in this thesis. A literature review was made to look for the most important and modern works on image quality in animated characters and the Uncanny Valley theory.

Table 2.1 shows a summary of the works studied in each section. The authors, the year and a brief information about the objective of the article are referenced.

The next chapter presents the feature extraction algorithms used in this work and which are widely used in Computer Vision. We describe the characteristics of each algorithm. In addition, 3 interpretability models are shown to evaluate the features relevant to the prediction made by our suggested models.

Section	Reference	Year	Objective
2.1	Mori[Mor70]	1970	It proposes that as human replicas become more realistic, small deviations in appearance or movements provoke a feeling of strangeness or uneasiness in people, a phenomenon known as the "Uncanny Valley".
	Ishiguro[Ish06]	2006	Explores how the use of androids in social and cognitive research allows us to investigate human reactions and interactions in a more controlled and detailed way, potentially revealing unique insights into human behavior.
	Lydenberg et al.[Lyd97]	1997	Examines how Freud connects the uncanny to themes such as the double, the boundary between the animate and the inanimate, and the ambiguity between the real and the imagined, analyzing how these concepts appear in Gothic and Surrealist narrative.
	Ho et al.[HMP08]	2008	Investigates how human emotions affect the perception of the Uncanny Valley in robots, using GLM, MDS and Isomap analyses to assess participants' reactions to robot videos.
	Tinwell et al.[TGNW11]	2011	It investigates how the emotional expressiveness of virtual characters affects the perception of the Uncanny Valley, especially when these characters cannot faithfully reproduce human appearance.
	Prakash et al.[PR15]	2015	The aim is to investigate how the degree of human similarity and the type of task influence the positive perception of humanoid faces.
	Hanson et al.[HOP+05]	2005	Explores how advanced robotics and facial design techniques can mitigate the sense of eeriness generated by humanoid robots, promoting a more positive emotional connection between humans and robots.

Von et al.[Von10]	2010	Explores how the use of avatars in job interviews can affect perceptions and hiring decisions, especially considering the emotional responses that avatars can elicit in interviewers and their influence on the interpretation of candidates' qualifications. The research highlights the ethical and objectivity challenges involved in the use of avatars and how this technology can both help and complicate the evaluation of candidates in the selection process.
Bailenson et al.[BSH+05]	2005	Investigates how the appearance and behavior of virtual agents (embodied agents) influence users' sense of co-presence in immersive virtual environments.
Gouskos et al.[Gou06]	2006	It explores the psychological and social reasons behind people's aversion to interacting with robots and characters that closely resemble humans but still display strange or not fully human characteristics.
Kang et al.[Kan09]	2009	Explores how the cultural representation of robots reflects tensions and ambivalences about human nature, addressing both the fascination and fear that these artificial beings evoke in contemporary societies.
Schein et al.[SG15]	2015	Explores how the perception of the mind in autistic people is influenced by eyes expression, relating this to the Uncanny Valley phenomenon in human-robot interaction.
Schwind et al.[SWH18]	2018	The research highlights the importance of adjusting visual and behavioral elements to create avatars and digital figures that are convincing without appearing overly realistic or uncanny, balancing human characteristics and stylizations to avoid the discomfort associated with the Uncanny Valley.

Flach et al.[FdMM+12]	2012	Explores how the visual characteristics of computer-generated characters influence viewers' perception of uncannyness, contributing to the understanding of the Uncanny Valley effect in animations and games.
Prendinger et al.[PMI05]	2005	It shows how human physiological responses can assess the effectiveness of subtle expressions of a virtual avatar, seeking to improve interaction in educational games.
Robb et al.[RKA+13]	2013	Analyzes the impact of interactions with virtual humans in communication practice scenarios, showing that these simulations help medical students develop emotional and coping skills before real experiences.
Dunsworth et al.[DA07]	2007	Focuses on how the presence of an animated agent in multimedia learning environments, especially with gestures and narration, can improve students' understanding and retention of scientific concepts.
Araújo et al.[AMDM21]	2021	Explores how viewers' acceptance and comfort with computer-generated (CG) characters evolves as these characters introduce new and innovative visual and technological aspects.
Ho et al.[HM10]	2010	It proposes and validates a new metric to evaluate the feeling of strangeness and discomfort in virtual characters, as an alternative to the traditional Godspeed indexes, improving the understanding of the uncanny valley theory.

Katsyri et a.[KMT17]	2017	Empirically investigates the Uncanny Valley hypothesis in semi-realistic characters from animated films, analyzing how the realism of facial expressions influences the viewer's perception of uncannyness and acceptance.
MacDorman et al.[MGHK09]	2009	Explores how computer-generated faces that approximate human realism too closely can cause discomfort and strange reactions in viewers, due to the phenomenon of the Uncanny Valley.
Geller et al.[Gel08]	2008	Explores why some near-human representations evoke fear or laughter, examining human perception and the challenges faced by artists and roboticists when creating human replicas that generate empathy or repulsion.
Macgillivray et al.[Mac07]	2007	It explores how the psychophysical perception of movement and image influences the practice of animation, highlighting that our innate ability to recognize shapes, especially faces, and the way we interpret movement are fundamental to the emotional and aesthetic effectiveness of animation.
Yekti et al.[Yek15]	2015	Analyzes the aesthetic differences between stop-motion animation and 3D graphic animation, highlighting how each of these forms of animation evokes distinct feelings related to physicality, materiality, perfection and imperfection, and how these characteristics influence the viewer's emotional experience.

	Mustafa et al.[MGT+17]	2017	It addresses the "uncanny valley" theory, examining how the similarity and quality of virtual characters affect the emotional response of humans, using EEG measurements to assess the perception and acceptance of these characters.
	Tinwell et al.[TGN15]	2015	Investigates how the desynchronization between audio and video in speech can influence the perception of realism and the feeling of strangeness (Uncanny Valley) in virtual characters.
2.3	Shahid et al.[SRLZ14]	2014	Provides a comprehensive review of the latest methods for reference-free image and video quality assessment, classifying and discussing their approaches, metrics, and applications, and highlighting current challenges and future directions in research in this area.
	Nill et al.[Nil85]	1985	Presents a new approach to image compression that combines the cosine transform with a human visual model, resulting in high compression rates with low quality loss, and an analytical solution to problems related to this combination.
	Gu et al.[GLZ+15]	2015	It explores the relationship between viewing distance and image resolution, proposing a scale selection model that optimizes the assessment of image quality under different viewing conditions, considering the physiology of the human eye and the characteristics of common image distortions.

Pinson et al.[PW04]	2004	Introduces a new standardized method for objectively measuring video quality, aiming to provide a more accurate and consistent assessment across different viewing contexts.
Wang et al.[WB06]	2006	It presents a comprehensive review of contemporary approaches to image quality assessment, discussing objective and subjective methods, and introducing the concept of image quality in relation to human perception.
Ma et al.[MLN10]	2010	It proposes a new image quality assessment model that considers the horizontal visual effect, improving the accuracy in quality perception through a more detailed analysis of visual characteristics.
Zhang et al.[ZLLN11]	2011	It proposes a watermarking method for images that uses the spread spectrum technique, integrating perceptual quality metrics to improve the robustness and imperceptibility of the watermark compared to traditional methods.
Rehman et al.[RW12]	2012	It proposes an image quality assessment method that uses structural information, allowing an effective analysis of image quality without the need for a complete reference.
Leveque et al.[LLB+18]	2018	Discusses the importance of evaluating the perceived quality of medical images and videos, analyzing different methodologies used in the literature for this subjective evaluation, considering their advantages and disadvantages in clinical contexts.

	Smith et al.[Smi]	-	Explores methods for subjectively assessing image quality, discussing how human perceptions influence the analysis and measurement of visual quality.
	Liu et al.[LLHB14]	2014	It proposes a reference-free image quality assessment method, using spatial and spectral entropies to quantify the perception of visual quality of images.

Table 2.1: Summary of the articles studied presented by section in which the study is addressed, the reference together with the year and the objective of the work studied.

3. BACKGROUND

This chapter presents some methods that borrow a theoretical foundation for this thesis and are related to the goals this work aims to achieve.

The organization of the chapter is following described: First Section 3.1, we present algorithms for feature extraction in computer vision, Second Section 3.2 the frameworks available for interpretability of generated models regardless of their complexity.

The 3.1 section shows the algorithms for feature extraction in computer vision. It highlights articles that have already used these algorithms, including some dealing with UV theory. We highlight the Hu Moments algorithm designed to capture essential properties of a shape in an image, such as contour and intensity distribution. These moments are calculated from mathematical functions called geometric moments. We made an analogy with the parts of the face based on the concepts of Hu Moments because these algorithms presented significant results in the search for trying to quantify human perceptual discomfort.

Finally, in Section 3.2 shows the frameworks available for interpretability of generated models regardless of their complexity. We highlight LIME as the framework used in our research and present the concepts that are used as a way of interpreting its results.

3.1 Feature Extraction Algorithms

This section discusses the algorithms used for feature extraction. The objective is to extract relevant facial features for use in Machine Learning algorithms to predict perceptual discomfort.

AUs (Action Units) are basic components of the FACS (Facial Action Coding System), developed by Paul Ekman et al. [EFE13], which describe individual facial movements associated with human expressions. In Uncanny Valley detection, AUs are used to analyze the naturalness and synchrony of facial movements in virtual characters, helping to identify elements that cause discomfort or strangeness.

The study by Mäkäräinen et al. [MKT14] investigates the relationship between exaggerated facial expressions and the intensification of emotional perception, addressing the concept of the Uncanny Valley. They use AUs to measure and

manipulate the facial expressions of virtual characters, focusing on the hypothesis that the intensification of emotions, by exaggerating the expressions, can both increase the perceived emotion and intensify the feeling of uncannyness. They conclude that, as facial expressions are exaggerated beyond certain natural limits, the emotional effect perceived by observers can become disproportionate, leading to an increase in the perception of uncannyness. This occurs especially when AUs do not follow natural human patterns, creating discrepancies that generate discomfort. The research demonstrates that there is a threshold between emotional intensification and observer immersion, which can be broken when virtual facial expressions do not correspond to the expected realism.

Another algorithm used is Entropy. Entropy has been used to study the Uncanny Valley and the perception of strangeness in animation and robots. Entropy is a measure of disorder or unpredictability in a system. In studies related to the Uncanny Valley, entropy can be used to quantify variation in facial features, movements, or behaviors of computer-generated characters or robots. When this variation deviates from what is perceived as natural in humans, it can generate a feeling of discomfort or strangeness. Mustafa [MGT+17]'s study uses EEG to measure brain responses to CG characters and uses signal analysis to attempt to predict the perception of strangeness. Liu et al. [LLHB14]'s study identifies the different types of distortions that affect the local entropy of images. Spatial and spectral entropy are calculated to measure the probability distribution of pixel values and Discrete Cosine Transform (DCT) coefficients, respectively.

The GLCM (Gray-Level Co-occurrence Matrix) technique is widely used in the analysis of image textures. It is very useful for measuring textures and visual patterns in images, which can be relevant for identifying visual characteristics that increase or reduce the feeling of strangeness in digital or robotic characters. The strangeness perceived in the Uncanny Valley often involves the lack of realism in textures such as skin, hair and eyes, where GLCM could be used to measure the difference between these textures and what is expected in real images. Studies such as that of Shahid et al. [SRLZ14] show that this method can be used to identify textures that are perceived as anomalous or degraded.

The Golden Ratio is a mathematical proportion that has been widely studied and used in art, architecture and design due to its pleasing aesthetics and visual balance. The golden ratio is approximately 1.618. It is known for its presence in many natural structures and classical works of art, and is often associated with a sense of harmony and beauty. The application of the Golden Ratio in detecting

and mitigating the uncanny effect is done in this research through Proportion analysis, used to analyze the proportion between different facial features. According to Schmid et al. [SMS08], facial symmetry and neoclassical proportions play a central role in the perception of beauty. More symmetrical faces that follow golden proportions tend to be seen as more attractive.

Histograms of Oriented Gradients (HOG) [DT05] is a technique developed to describe and detect objects based on the distribution of gradient orientations. The most famous use of HOG [DT05] is in pedestrian detection, but the technique can be applied to various computer vision tasks, including facial expression analysis. There are no studies prior to this work that use this technique to assess facial discomfort.

Finally, we have another robust visual pattern descriptor and descriptor known as Hu Moments [Hu62]. Like HOG [DT05], there are no studies in the literature that address this algorithm for discomfort assessment. However, it seems to be a promising resource. Therefore, we performed a more detailed analysis of the meaning of its vector and made an analogy with the human face due to the face that we extensively used such technique in the present work. The research by Ming-Kuei Hu et al. [Hu62] aims to present a methodology for robustly describing and recognizing visual patterns, independent of transformations such as rotation, translation, and scaling. Ming-Kuei Hu et al. [Hu62] argues that these invariant features are crucial for pattern recognition in practical scenarios, where objects may appear in different orientations and sizes. Additional information about Hu Moments is available in the Appendix 8.

3.2 Machine Learning Interpretability Models

Few machine learning interpretability models work with the technique known as ensemble voting. It is part of the ensemble machine learning methods, where multiple models are combined to improve the accuracy of predictions. In the specific case of ensemble voting, the predictions of the individual models are combined through a vote (majority or weighted average) to determine the final prediction.

We evaluated SHAP [MHJ20] (SHaplay Additive exPlanation) and DALEX [BB21] (Model Agnostic Language for Exploration and Explanation), both of which address global and local model explainability by evaluating the entire test dataset and also an instance of it.

SHAP is based on Shapley values from cooperative game theory. This provides a unified measure of feature importance that is consistent and reliable. However, implementing SHAP can be more complex and computationally intensive, especially for large datasets and complex models. Just like LIME, SHAP easily integrates with various machine learning frameworks in Python, with the exception of ensemble models.

DALEX is flexible and allows comparisons between models. Its biggest weakness is that it is primarily available in R, which can limit its accessibility for users more familiar with Python. Additionally, DALEX requires prior knowledge of R's machine learning libraries, which can represent a significant learning curve.

On the other hand, LIME [RSG16] (Local Interpretable Model Agnostic Explanations) does not address global interpretability, only local interpretability. We chose LIME for this study because we wanted to locally explain the parts of the face that generate the comfort or discomfort prediction and also because we use ensemble models.

LIME presents better performance in generating the surrogate model compared to SHAP and DALEX. Furthermore, LIME makes interpreted data available for collection, making manipulation more flexible for better interpretation of results. Therefore, we were even able to generate the most important features globally from the test dataset by collecting the feature importance for each instance. This way, we can also have a global view of the most important features that can be found in both the training and testing datasets, and we can compare the data. This flexibility does not occur in the SHAP model or the DALEX model, which restricts information only through graphics.

Therefore, we can also generate the most globally relevant features from the test dataset by collecting the feature importance for each instance. LIME is conceived as a model that seeks to emulate the behavior of a pre-existing model, called a surrogate or surrogate model, in a local context. This surrogate model is trained on a dataset derived from instances close to the one being interpreted, introducing small variations in features or attributes, weighted according to their proximity to the original instance.

3.2.1 Interpretation of LIME (Local Interpretable Model-agnostic Explanations)

When using LIME (Local Interpretable Model-agnostic Explanations) [RSG16] for interpretation for both classification and regression models, local explanations are presented for a specific instance (an individual example). Interpretation varies depending on the type of model, but follows a general approach to the elements of LIME.

1. Local Explanation: LIME provides local explanations, that is, it explains why a model made a specific prediction for a given instance. This is useful for understanding the logic of the model at a more granular level.

2. Feature Importance Plot: Shows the main features that influenced the prediction for the given instance. This plot is usually a horizontal bar with information such as: Most Important Features - Features are listed on the Y-axis. Contribution - The length of the bar on the X-axis indicates how much that feature contributed to the prediction. Sign - The direction of the bar (to the right or left) indicates whether the feature contributed to the positive or negative class. Numeric Value - There may be a value associated with each feature, indicating its quantitative influence on the prediction. For example, a high, positive value for a feature means that that feature had a strong positive contribution to the predicted class. Interpretation - Bars to the right usually indicate that the feature influenced the prediction for either the positive class or the predicted class. Bars to the left indicate that the feature negatively influenced the predicted class (or favored another class). Bar length - The length of the bar reflects the magnitude of the influence. The longer the bar, the more that feature contributed to the model's final decision.

3. Explanation Table: LIME often also provides a table that details: Feature name (or a range of values, if it is a continuous feature). Feature weight: The quantitative value of the feature's contribution to the decision. Conditional prediction: Depending on the input value, the table can show how changing the feature's value impacts the probability of belonging to a given class.

4. Natural Language Explanation: LIME can generate explanations in text form. This text usually follows the pattern of "If feature1 has value X, then the probability of class being Y increases/decreases by Z%". This makes it easier for non-experts to interpret.

5. Classification Probabilities: In addition to the explanations, LIME can provide predicted probabilities for each class in the model. This allows you to see how "confident" the model was in its prediction. You might have a table or graph with the probabilities associated with each class, showing, for example, that the model predicted class "A" with 80

Figure 3.1 shows an example of a binary classification model that predicts whether a movie will have a high or low rating with a probability of 0.10 for the low rating class and a probability of 0.90 for the high rating class. The output from LIME shows one instance of the test data rather than the entire test set. The output explains how the model arrived at its prediction for that particular instance, given the specific feature values for that data point.

In the graph on the right, LIME shows the value for each feature for the data point we provided, with a vote count of 159, a revenue of 12,800,000, a runtime of 149.00, a release year of 1959, a number of genres of 2, and a popularity of 9.46. The output also shows the thresholds for each feature that the model used to make its prediction, such as a vote count threshold of 253.75, a revenue threshold of 0.00 to 14, and a runtime threshold of 115.00.

According to the probability distribution in the middle graph, the ML model believes that there is a probability of 0.1 that this data point belongs to the low-rated class. It shows that without the “vote count 253.75”, the probability of this data point belonging to the low-rated class would be 0.1–0.03 = 0.07, as the numerical values attached to each feature show the probability contribution to the prediction and the probability distribution shown on the left.

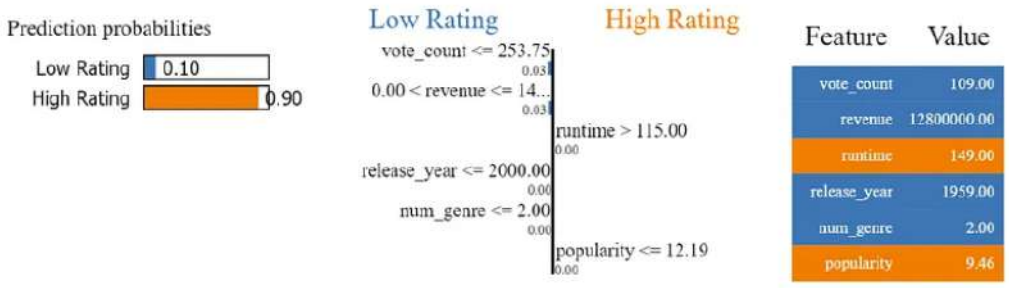


Figure 3.1: Example LIME output for an instance in a binary classification model that predicts whether a movie will be rated high or low.

As an example, we can think of a model for classifying facial expression images that predicts that an image represents “happiness.” LIME could make small

changes to the image (such as removing parts of the face) and observe how these changes affect the model's prediction. If removing the eyes significantly reduces the probability of "happiness," LIME will fit a local model that gives a high (positive) weight to the "eyes" feature, indicating that this feature was crucial to the prediction of "happiness".

Explanations can vary depending on the samples chosen for analysis, which can help identify patterns in groups of data that the model uses to make predictions. This type of interpretation allows for a better understanding of which variables influence the model's decision-making for each individual sample, both in classifiers and regression models.

LIME Interpretation in the Classification Model

In a classification model [Chr20], the goal is to explain the prediction of a specific class for a given sample. The explanation is based on the importance of the features in determining the probability of a specific class being chosen by the model. Here are the steps to interpret:

- Bar chart (local explanations): Each bar represents a feature, and the length of the bar indicates how much this feature contributes to the prediction of the selected class (usually positive or negative). If the model is binary, LIME will show the contribution to one of the classes.
- Color of the bars: Generally, positive bars (orange) indicate that the feature increases the probability of the sample being classified in the class in question, while negative bars (blue) indicate that the feature reduces this probability.
- Weight of the features: Longer bars indicate features with greater impact on the prediction. The visual interpretation allows to identify the most relevant variables for the model in the prediction of that specific instance.

LIME Interpretation in Regression Model

For regression models[GMR+18], LIME attempts to explain how features contribute to the predicted value of a given continuous output variable. Instead of explaining the probability of a class, it focuses on the predicted value (such as price, grade, etc.). Here are the steps to interpret:

- Bar chart (local explanations): As in the classification model, the bars represent the impact of the features, but here the impact is on the predicted value (continuous), not on a class. The bars can be positive or negative, indicating whether the feature is increasing or decreasing the predicted value.
- Feature weight: Features with longer bars indicate that they had a greater impact on the predicted final value. Features that pull the value up appear with positive bars and those that pull it down appear with negative bars.

3.3 Chapter Considerations

This chapter has presented the contribution of the extraction of features from CG images using computer vision techniques, primarily the Hu Moments algorithm, and analyzing image quality that, based on subjective evaluation, evokes a sense of the uncanny, as studied in the Uncanny Valley (UV) theory. We also show three frameworks widely used in the literature. We highlight the reason for choosing to work with LIME in the evaluation of results. We present an example of the interpretability of the chosen model.

The main question is to investigate whether image characteristics on the faces of CG characters can help define when and where images can cause uncanny perception. This work is expected to contribute to the entertainment industry (games and movies) through recommendations and analyses that can serve to improve the experience and enhance the perception of CG characters.

The next chapter presents the two datasets used in this research, as well as the proposed models, describing their characteristics and how they are assembled to work together to achieve the objectives of this work.

4. DATASETS

In this chapter we discuss two datasets used in the present research. Section 4.1 refers to dataset GT1 which is based on the work of Araujo et al. [AMFK+19] and Flach et al. [FdMM+12]. Section 4.2 deals with the dataset we created by searching for new characters that cause discomfort in literature research and which we will call GT2.

4.1 Dataset GT1

Our character selection, initially, is based on the work of Araujo et al. [AMFK+19] and Flach et al. [FdMM+12], who analyzed, with human subjects, the perception of comfort when observing characters created with CG (films, games, and computer simulations). It was used images and videos of the same 10 characters from Flach, as shown in Figure 4.1 as being all letters a), (c), (e), (g), (i), (k), (m), (o), (q), (s), (u) . It was also included more recent CG characters, as proposed by Araujo et al. [AMFK+19] shown in Figure 4.1 he remaining letters To ensure the variation of human likeness present in the Uncanny Valley, some of the chosen characters represent a human being in a caricatured way (q), (s) and (u), and and others are more realistic, such as (m), (n), (v), (r), (k) in Figure 4.1.

To obtain human perceptions of realism and comfort (variables necessary to construct the *X* and *Y* axes of the Uncanny Valley graph), a questionnaire was created in the work of Araújo et al. [ADM21]: *i)* Q1 - "How realistic is this character?", with three Likert scales of responses ("Unrealistic", "Moderately realistic", and "Very realistic") for perceived realism; *ii)* Q2 - "Do you feel any discomfort (strangeness) when looking at this character?", with answers "YES" and "NO" to perceived comfort; and *iii)* Q3 - "In which parts of the face do you feel the most strangeness?", with multiple choice ("eyes", "mouth", "nose", "hair", "others" and "I do not feel discomfort"). The authors used Google Forms and recruited participants on social media. They randomly presented characters to participants through images and short videos. The subjects would then answer questions. A total of 119 participants responded to the questionnaire, of which 42% were women and 58% were men, with 77.3% being under 31 years old and 33.7% being 31 or older.

Figure 4.1: All characters used in this dataset called GT1. Characters (a), (c), (e), (g), (i), (k), (m), (o), (q), (s), (u) used in Flach et al. [FdMM+12]. The remaining characters are chosen by Araujo et al [AMFK+19]. The characters with rectangular frame in red caused discomfort in the empirical research carried out.

In the present work, we used 19 videos out of 22 (one short film for each character illustrated in Figure 4.1) from the work of Victor et al. [ADM21] and removed those frames that did not contain the face of the character to be analyzed. This process resulted in 5,730 images. It is important to mention that characters (d), (g) and (j) were not included in the analysis because there was no detection of the face or parts of the face. After selecting the 19 characters, this research considered as ground truth the response to Q1 to determine the level of perceived realism, Q2 to determine the percentage of perceived comfort, and Q3 to evaluate the parts of the face that generate the most strangeness. To categorize the characters into different levels of realism, the average scores of the responses to Q1 were used, so

that each character has an average realism value, according to the study by Victor et al. [ADM21].

Three levels of realism were also used to divide the characters, as suggested in the work of Victor et al. [ADM21]. This division was used in the study of Dal Molin et al. [DMdAAM22]. Below are the three groups:

- Unrealistic characters, with average realism values ≤ 1.5;
- Moderately realistic characters, with average realism values ≤ 2.5; and
- Very realistic characters, with realism values > 2.5. The comfort value for each character was calculated by the percentage of "NO" (discomfort) responses to question Q2.

This dataset was used in two previous studies. Firstly, the features were extracted from images based on Hu Moments (Hum) and Histogram Oriented Gradient (Hog), and the Support Vector Machine (SVM) model was used to provide binary classification [DMND+21b], described in Section 5.1.1. In a second study by Dal Molin et al. [DMdAAM22], the perceived comfort estimation was performed using spatial and spectral entropy and used the Support Vector Regressor (SVR) model to provide the *CCS* (Computing Comfort Score - a metric we propose to estimate the perceived comfort), described in Section 5.2.1.

4.2 Dataset GT2

Since we considered 19 characters (GT1) not many, we conducted a systematic literature review (SLR) to search for more characters. The objective was to find characters reported as causing (or not) a feeling of strangeness in human perception to increase the dataset GT1. We identified additional animated characters that have been studied. A total of 12 articles covering the literature from 2007 to 2023 were evaluated for their examination of UV theory in animated characters, as detailed in Table 4.1.

Based on the systematic literature review (SLR), we identified 21 characters that have been reported to cause feelings of strangeness in people. Table 4.1 shows the characters and their classification according to the articles referring to the period from 2007 to 2023. All characters involved in each article were reported,

and the strangeness classification was identified in the analyses. Many characters are repeated in these articles. Other characters were already part of the GT1 dataset. We selected 21 characters indicated in more than one article as causing strangeness.

We then combined the selected 21 characters with the 19 characters from the GT1 dataset (Section 4.1), creating a new dataset called GT2. To establish a ground truth for the perception of strangeness among these 40 characters, we conducted a new questionnaire using the Qualtrics platform [1], as presented later. We deemed it appropriate to integrate all characters and exclude the results from the ground truth to ensure consistency in our research. Additionally, this new study was valuable for assessing whether the discomfort associated with these characters has persisted over time, given that the 19 characters in question are from works produced in 2012 and 2019.

Reference	Year	CG characters studied	Cause strangeness
[Mac07]	2007	1. Dumbo	No
		2. Bugs bunny	No
		3. Bamby	No
		4. Gosth in the machine	No
		5. Pinnochio's Fairy GodMothers	No
		6. King and queen em Shrek	No
		7. Polar Express children	**Yes**
		8. Final Fantasy	**Yes**
		9. Pirates of Caribbean	No
		10. Reality SFX	No
[MGHK09]	2009	1. Final Fantasy: The Spirits Within	**Yes**
		2. Davy Jones from Pirates of the Caribbean: At World's End	**Yes**
		3. Polar Express	**Yes**
		4. Gollum in The Lord of the Rings trilogy	**Yes**

[1] https://pucrs.qualtrics.com

Table 4.1 continued from previous page

Reference	Year	CG characters studied	Cause strangeness
[HM10]	2010	1. Final Fantasy: The Spirits Within ("Aki Ross")	**Yes**
		2. The Incredibles	No
		3. The Polar Express	**Yes**
		4. An advertisement for Orville Redenbacher popcorn	**Yes**
		5. A technology demonstration of the video game Heavy Rain	No
[TGW10]	2010	1. The Emily Project (2008a)	No
		2. The Warrior (2008b) by Image Metrics	**Yes**
		3. Quantic's Mary Smith the Dream technical demo, 'The Casting' (2006)	**Yes**
		4. Alex Shepherd from Silent Hill Homecoming (Konami 2008); two avatars	No
		5. Louis, from Left 4 Dead (Valve 2008)	No
		6. Francis, from Left 4 Dead (Valve 2008)	No
		7. Four zombie characters	**Yes**
		8. The Smoker	**Yes**
		9. The Infected	**Yes**
		10. The Tank and	**Yes**
		11. The Witch, from left 4 Dead	**Yes**
		12. A stylized human Chatbot character, 'Lillien' (Daden Ltd. 2006)	**Yes**
		13. A realistic, human-like zombie ('Zombie 1') from the video game, Alone in the Dark (Atari Inc. 2009) The real human.	No

Table 4.1 continued from previous page

Reference	Year	CG characters studied	Cause strangeness
[DFH+12]	2012	1. Unknown virtual human	**Yes**
		2. Obama's cartoon	No
		3. The Incredibles	No
		4. Unknow virtual human	**Yes**
		5. Beowulf	No
		6. Heavy rain	**Yes**
		7. Rango	No
		8. Cloudy with a chance of metalballs	No
		9. Unknown virtual human	**Yes**
		10. Unknown virtual human \end{itemize}	**Yes**
[Per14]	2014	Digital Ira	Considered a little strange
[Yek15]	2015	1. Coraline (2009)	No
		2. Fantastic Mr. Fox (2009)	No
		3. The Adventure of Tintin (2011)	**Yes**
		4. Up (2009) \end{itemize}	No
[KMT17]	2017	1. Final Fantasy: The Spirits Within ("Aki Ross")	**Yes**
		2. Polar Express (unnamed boy)	**Yes**
		3. Beowulf ("Beowulf")	**Yes**
		4. The Adventures of Tintin: The Secret of the Unicorn ("Tintin")	No
		5. The Incredibles ("Mr. Incredible")	No
		6. Meet the Robinsons ("Lewis")	No
		7. Cloudy with a Chance of Meatballs ("Flint Lockwood")	No
		8. Arthur Christmas ("Arthur Christmas")	No
		9. Epic ("MK")	No
		10. Mars needs moms ("Milo")	**Yes**
[MGT+17]	2017	1. Digital Emily	No
		2. Digital Ira	No
		3. Kara de Detroit: Become Human	**Yes**
		4. 'Ernst' de Squadron 2	**Yes**
		5. 'HeadTech' de Janimation	**Yes**

Table 4.1 continued from previous page

Reference	Year	CG characters studied	Cause strangeness
[HM17]	2017	1. Doctor Aki Ross from the film Final Fantasy: The Spirits Within (2001)	No
		2. Billy, the baby from “Tin Toy” (1988)	**Yes**
		3. An unnamed man from Phil Rice’s “Apology” (2008)	**Yes**
		4. Orville Redenbacher from a popcorn commercial (2007)	No
		5. Mary Smith from “Heavy Rain: The Casting” (2006)	No
		6. Five robots \item Roomba 570 (iRobot)	No
		7. Kotaro (JSK, University of Tokyo)	No
		8. Jules (Hanson Robotics)	No
		9. Animatronic Head (David Ng)	No
		10. Aiko (Le Trung), and two humans	No
		11. A man and	No
		12. A woman	No

Table 4.1 continued from previous page

Reference	Year	CG characters studied	Cause strangeness
[AMDM21]	2021	1. Unknown virtual human	**Yes**
		2. Unknown virual human	No
		3. Unknown virtual human	**Yes**
		4. The incredible I	No
		5. Obama's Cartoon	No
		6. Unknown virtual human	**Yes**
		7. Cloudy with a chance of metalballs	No
		8. Beowulf	No
		9. Heavy rain	**Yes**
		10. Rango	No
		11. Unreal 4 Engine	No
		12. Alita	**yes**
		13. How to train your dragon 2	No
		14. Thor Ragnarok	No
		15. Rogue One	No
		16. Love, death and robots	No
		17. Moana	No
		18. Overkill's the walking dead	No
		19. Spider-Verse	No
		20. Unreal 4 Engine	No
		21. The incredibles 2	No
		22. Unreal 4 Engine	No

Table 4.1 continued from previous page

Reference	Year	CG characters studied	Cause strangeness
[dAACM23]	2023	1. The Incredibles 2	No
		2. Soul	No
		3. Arcane	No
		4. GTA San Andreas	No
		5. The Walking Dead from Telltale	No
		6. Encanto	No
		7. Spider-Verse	No
		8. Moana	No
		9. True Crime New York City	No
		10. Obama's Cartoon	No
		11. GTA V	No
		12. Mortal Kombat 11	No
		13. Fifa 19	No
		14. Call of Duty Black Ops 2	No
		15. Horizon Zero Down	No
		16. MetaHuman Creato	No

Table 4.1: The table presents perceptual data acquired through human responses, which address the strangeness of CG characters in films, games and VHs from 2007 to 2023.

4.3 Survey answered by subjects

As already stated, we selected 21 characters from Table 4.1 and 19 more characters from dataset GT1, discussed in Section 4.1, resulting in 40 characters. The literature classified nineteen characters as causing strangeness for people and 21 as not causing it, as reported in the researched literature. Figure 4.2 shows the 21 characters included in our research in addition to the 19 characters from Figure 4.1.

We applied a subjective evaluation to calculate a comfort score. The goal was to obtain perceived comfort on a continuous scale, such as "How uncomfortable is this character for you?". A regression model is ideal in this case, as it allows for greater precision when estimating a range of values. The greater granularity in pre-

Figure 4.2: 21 characters collected from literature and used for human evaluation together with the 19 characters from Figure 4.1. The literature considers the characters outlined in red to cause strangeness.

dictions allows for predictions of different intensities or intermediate values, whereas a binary classification (comfortable/uncomfortable) is restricted to two categories. In problems where it is important to know how much a variable affects the outcome, and not just whether or not it does, regression provides more information.

The questionnaire was created on the Qualtrics platform. We asked two questions, following the research line of Flach et al [FdMM+12]. The questions asked are as follows: Q1 - “Do you feel any discomfort (strangeness) when looking at this character?”, with answers “YES” and “NO” to perceived comfort; and when answered “YES”, Q2 - “In which parts of the face do you feel discomfort?”, with multiple choice (“eyes”, “mouth”, “nose”, “hair”, “others” and “Others”). There were 44 participants aged between 18 and over 60, and the survey was available for four days. In Q2, participants could choose only one part of the face that stood out the most: hair, forehead, eyes, nose, mouth and chin, if they felt any discomfort when looking at the character. Table 4.2 shows the results of this survey.

Character	Mouth	Hair	Nose	Eyes	Chin	Forehead	I didn't feel strange	Sum Parts	ROI	GT	Comfort
1	6	1	0	29	0	0	8	36	eyes	1	0.18
2	5	2	0	1	0	1	35	9	-	0	0.80
3	1	2	0	21	0	0	20	24	eyes	1	0.45
4	0	0	0	33	0	0	11	33	eyes	1	0.25
5	13	0	10	7	9	1	4	40	mouth	1	0.09
6	6	0	5	1	0	1	31	13	-	0	0.70
7	9	1	4	5	4	1	20	24	mouth	1	0.45
8	5	0	0	6	9	7	17	27	chin	1	0.39
9	4	1	2	4	1	0	32	12	-	0	0.73
10	5	1	3	4	2	17	12	32	forehead	1	0.27
11	1	0	0	3	0	0	40	4	-	0	0.91
12	0	0	1	8	0	0	35	9	-	0	0.80
13	0	0	0	22	0	0	22	22	eyes	1	0.50
14	1	0	1	4	0	1	37	7	-	0	0.84
15	11	1	4	5	1	4	18	26	mouth	1	0.41
16	3	3	0	4	0	0	34	10	-	0	0.77
17	32	1	0	7	3	0	1	43	mouth	1	0.02
18	0	1	0	17	0	0	26	18	-	0	0.59
19	0	2	1	6	2	3	30	14	-	0	0.68
20	2	2	1	16	1	3	19	25	eyes	1	0.43
21	3	6	0	18	1	0	16	28	eyes	1	0.36

22	10	0	1	4	1	3	25	19	-	0	0.57
23	0	0	0	12	0	2	30	14	-	0	0.68
24	3	0	1	30	0	0	10	34	eyes	1	0.23
25	7	0	2	11	0	7	17	27	eyes	1	0.39
26	25	0	0	15	0	1	3	41	mouth	1	0.07
27	4	1	16	0	0	0	23	21	-	0	0.52
28	28	0	3	2	0	8	3	41	mouth	1	0.07
29	2	0	1	18	0	5	18	26	eyes	1	0.41
30	1	1	2	6	0	1	33	11	-	0	0.75
31	13	0	2	11	2	2	14	30	mouth	1	0.32
32	2	0	1	13	1	0	27	17	-	0	0.61
33	8	5	0	4	1	0	26	18	-	0	0.59
34	12	1	3	5	1	1	21	23	mouth	1	0.48
35	1	2	0	6	0	15	20	24	forehead	1	0.45
36	5	0	1	7	0	0	31	13	-	0	0.70
37	20	0	0	5	0	3	16	28	mouth	1	0.36
38	0	0	19	0	0	0	25	19	-	0	0.57
39	0	0	1	3	1	1	38	6	-	0	0.86
40	0	0	0	5	0	1	38	6	-	0	0.86

Table 4.2: Result of the survey applied with subjects. Characters whose entire face was rated as uncomfortable or not. If uncomfortable, the most awkward parts of the face are identified. The Sum Parts column is the sum of the parts of the face that were identified as awkward. If the Sum Parts value is greater than the Don't feel discomfort column, then the GT (Ground Truth) column receives the value 1 for uncomfortable or 0 for comfortable. The comfort column is calculated based on the number of responses divided by the number of survey participants.

Figure 4.3 shows the 40 characters classified according to the research carried out in Qualtrics platform. This dataset is balanced by class (comfortable and uncomfortable) and by characters, as 19 characters cause strangeness/discomfort, and 21 are considered comfortable. However, there is no balance in relation to the number of frames (9495 frames from uncomfortable characters and 3907 from comfortable ones). For this reason, we balanced the GT2 dataset of frames in relation to class and characters. We randomly selected frames of characters that generate strangeness, because it is the class with the largest number of frames. We balanced the number of frames between the two classes. In this way, we were able to obtain 4192 frames of characters that generate discomfort with 3907 frames of the comfortable class, leaving the dataset balanced by the number of total frames per class, totaling 8099 frames.

The survey demographics show a balance of participants in the age groups of 18 to 20 years, 21 to 29 years, 40 to 59 years, representing 25%, 22.73% and 27.27% of the total respectively. The majority of respondents have higher education (52.27%), indicating a highly qualified audience. The distribution of designated sex reveals a balance between genders, with a slight predominance of males (56.81%). As for the area of activity, the technology sector (54.54%) stands out, suggesting that the survey attracted an audience interested in technology-related issues.

The results of the current survey using the Quatrics tool, which we call (GT2), were analyzed in comparison with the previous survey (GT1) and the systematic literature review (SLR). Below, we present the agreement between the respondents regarding the characters treated.

Figure 4.3 shows the 40 characters (GT2), with the first 19 corresponding to the GT1 dataset and the rest being the SLR characters.

1. Agreement between GT1 and GT2:

- Agree: 14 characters (73.68%)
- Disagree: 5 characters (26.32%)

The current survey (GT2) shows a significant level of agreement among the participants, with 73.68% of the characters expressing agreement, compared to 26.32% who disagreed.

2. Agreement between GT2 and SLR:

- Agree: 12 characters (57.14%)

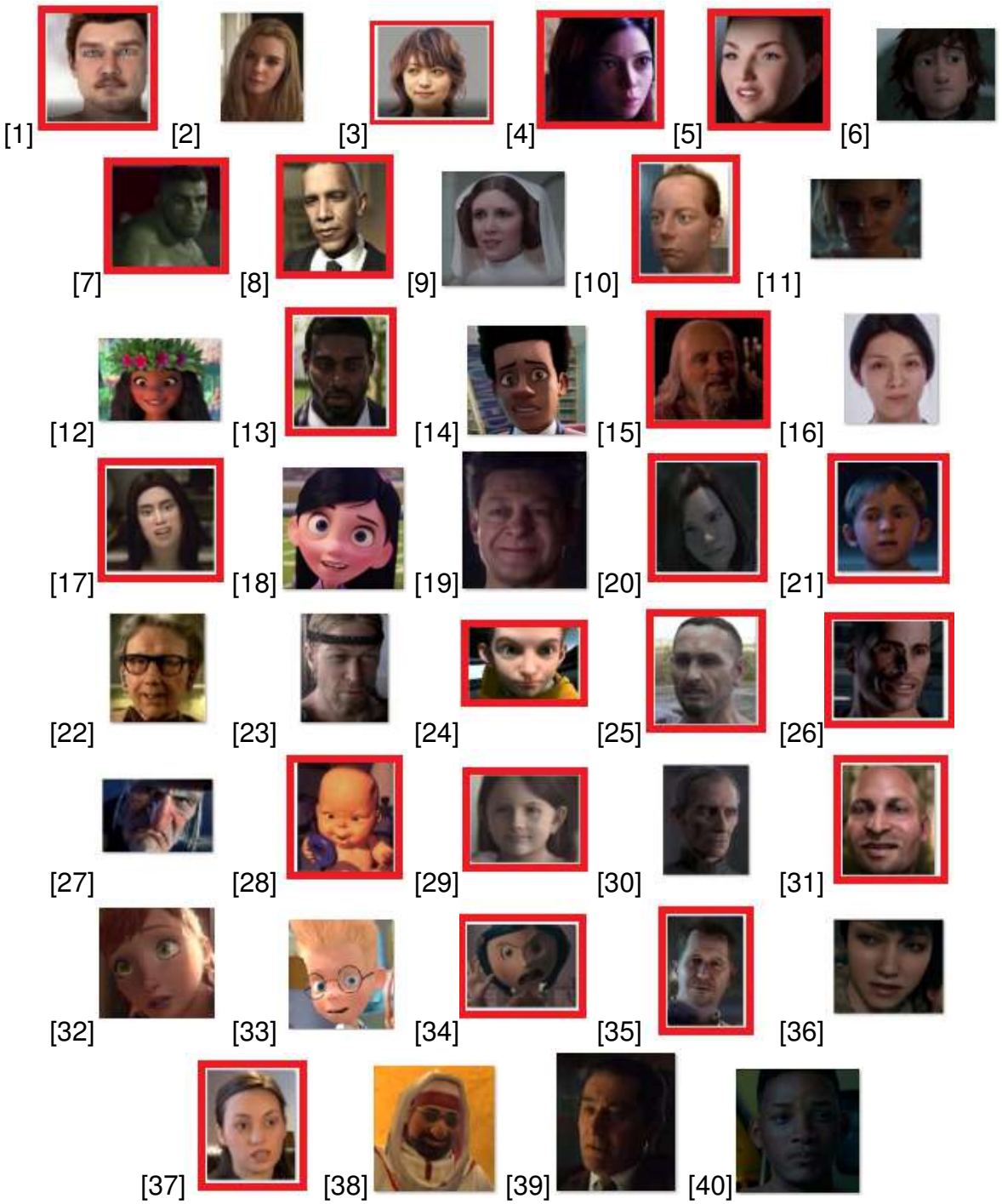


Figure 4.3: The characters with a red frame indicate discomfort perceived by the participants, totaling 21. The remaining 19 characters are considered comfortable.

- Disagree: 9 characters (42.86%)

Compared to the long-term reference, the level of agreement decreased to 57.14%, while disagreement increased to 42.86

The data reveal an interesting trend: while the previous survey (GT1) presented a high rate of agreement (73.68%), the current survey (GT2) presents a reduction in this number (57.14%) in relation to RSL. This suggests a possible change in the opinions or perceptions of the respondents over time. The increase in the disagreement rate, from 26.32% to 42.86%, may also indicate a greater diversity of opinions among the current participants.

These results are important to understand the evolution of opinions and may guide future actions and research on the topic addressed.

5. PROPOSED MODEL

This chapter aims to present five models proposed in this research. The first and second models propose a binary classification using the SVM algorithm and the Voting Classifier methods to infer whether the character's face will cause strangeness/discomfort to people or not. In the third one, we fine-tune VGG16 to evaluate whether a Convolutional Neural Network can predict, through a binary classification, the perception of strangeness compared to the subjective human classification. The fourth and fifth models propose a method to calculate the perceived comfort represented through a continuous value. We call this value *CCS* (Computed Comfort Score), which represents a percentage of strangeness for the character's face and uses the SVR algorithm and the Voting Regressor method. We detail the models below.

5.1 Binary Classification Models

We propose three binary models to classify CG faces to infer whether the character will cause people to feel strange/discomfort or not. Section 5.1.1 deals with the model that uses the support vector machine (SVM) [1], while section 5.1.2 uses the features of the voting classifier (VC) [2]. We also propose in section 5.1.3 the fine-tuning of a CNN to compare with the methods mentioned in the last sections.

5.1.1 SVM Binary Classification Model

Figure 5.1 shows the overview of the SVM model. First, the Haar Cascade method is used to detect the faces and parts of the face, such as eyes, eyebrows, mouth, jaw. The descriptors Hu Moments and Hog are used to generate the vector of characteristics of the entire face and parts of the face. The saliency function shows a part of the face that stands out to extract features with the descriptors. We use the Principal Component Analysis (PCA) to reduction of dimensionality to define the most relevant variables of the vector of characteristics. Finally, the Support

[1] https://scikit-learn.org/stable/modules/svm.html
[2] https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.VotingClassifier.html

Vector Machine (SVM) model is used for the binary classification of the detected faces. It presented each step of our model in Figure 5.1. It is worth mentioning that for this first model the data set with the initial 19 characters was used as explained in Section 4.2. Then we retrain the model with the 40 characters as shown in Section 4.2.

Pre-Processing Data

We performed three main processes in order to prepare data to be used in our classification method: *A)* face detection and *B)* saliency detection and *C)* Hu and HOG feature extraction, as detailed next. We implemented our method using OpenCV [How13], scikit-learn [VdWSNI+14] and dlib [Ros17] in this process.

(A) Face detection

The method used for face detection was proposed by Paul Viola and Michael Jones [VJ+01]. This method detects a face and also finds the parts of the face. In the latter case, eight parts, such as the mouth, middle of the mouth, right and left eyes, right and left eyebrows, nose, and jaw, as shown in Figure 5.2. The image is discarded if no face is detected. Haar Cascade [VJ+01] algorithm provides a good result in detecting the faces of characters created in CG, although it was created for detecting human faces. The only character with no face detected was from the movie Incredibles 2, as shown in (u), in Figure 4.1.

(B) Saliency detection

We compute the saliency map of each frame from the first 19 characters of the GT2 dataset described in Section 4.2. The visual salience method used in this work is based on the difference between the center and the image's outline. It is a form of extraction that highlights the regions in the image that attract the attention of human beings, as studied by Jia et al. [JZW18] and You et al. [YPG10]. It generated the feature descriptors presented next in images with and without the extracted salience to assess its usefulness in detecting images that present strangeness to human perception. The method used was the Fine Grained saliency from the OpenCV [How13] library with the default parameters.

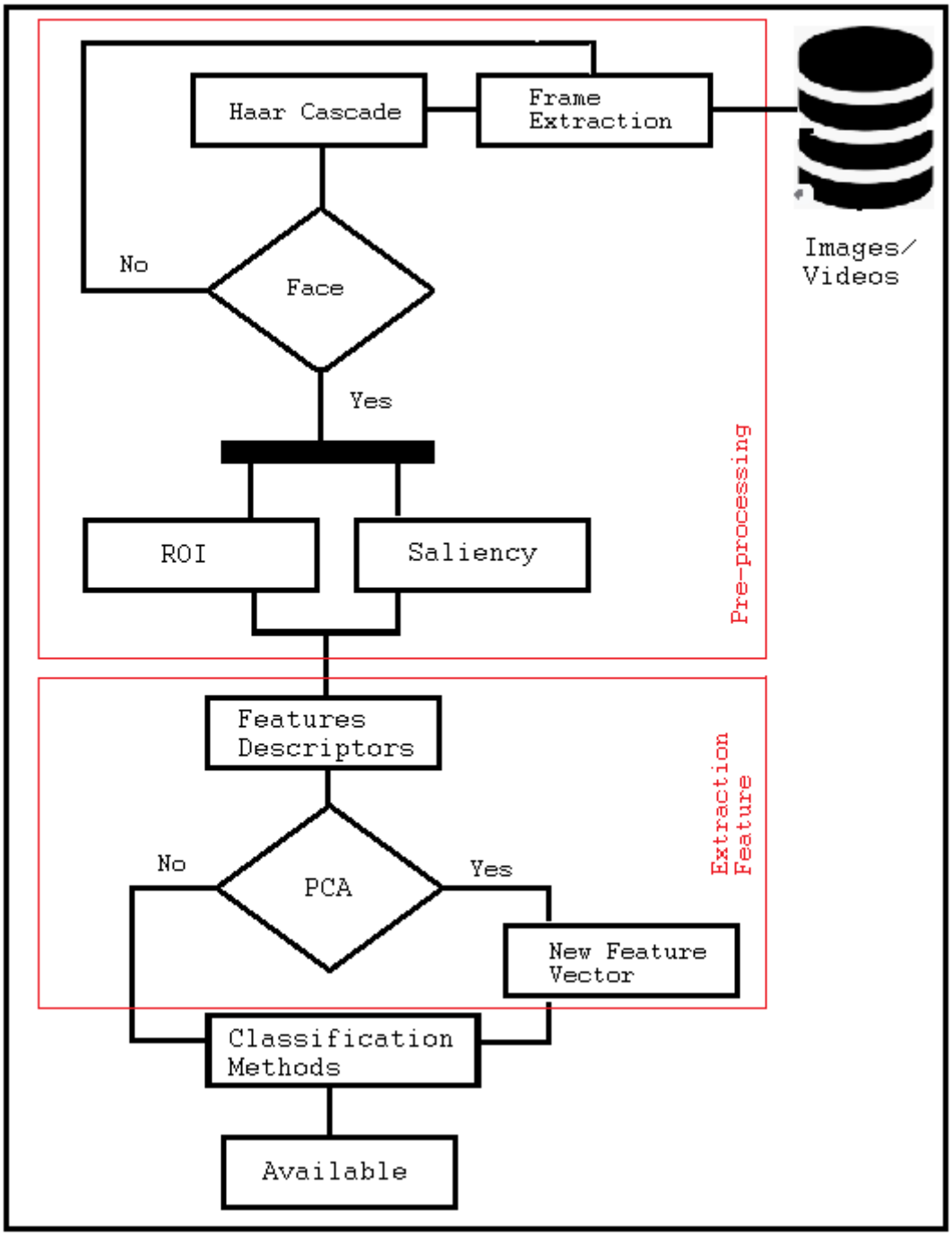


Figure 5.1: Our classification model detects the face of the animated character, extracts the facial features through the Hu Moments and Hog algorithms, with and without the saliency function. PCA is used to reduce the dimensionality of the feature vector. Finally, SVM classifies whether the character will generate discomfort or not.

Features Extraction

In this step, we use the Hu Moments [ŽHR10] and the Histogram of Oriented Gradients (HOG) [DT05], as they are two algorithms widely used in the area

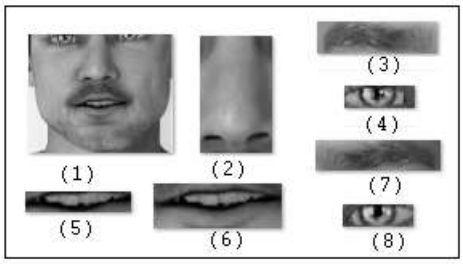


Figure 5.2: Parts of the face detected by Haar Cascade: (1) jaw, (2) nose, (3) right eye brow, (4) right eye, (5) inner mouth, (6) mouth, (7) left eye brow, (8) left eye

of computer vision. So far, no references have been found in the literature regarding their use for detecting strangeness perception.

Hu Moments is used with its default parameters [ŽHR10], implemented using OpenCV[3] [How13]. This descriptor generates seven moments regardless of image size. The other descriptor used was the Histogram of Oriented Gradients (HOG) [DT05]. We consider the detection window with gradient voting into 9 orientation bins and 64x64 pixels blocks of 1x1 pixel cells, generating the descriptor of the image characteristics to be used. It implemented hog using scikit-learn [VdWSNI+14]. This descriptor generates a feature vector depending on the size of the image. Principal Component Analysis (PCA) [WC06] was used to detect the most relevant variables between the characteristic vectors of HOG and Hu Moments, defining in 95% the sum of the variables' accuracy as being relevant. For model training, the two combinations of variables were tested - using PCA and not using PCA - and the results are compared.

Training, Testing and Validation Process

To perform the training, test, and validation, we organized the dataset as the procedure described in algorithm 5.1. We use leave-p-out cross-validation, where $p = 2$, i.e. using p observations as the validation set and the remaining observations as the training set. This is repeated in all ways to cut the original sample onto a validation set of p observations and a training set.

[3] https://opencv.org/

```
i=1; j=1;
while i<Number_UV_characters> do
    while j<Number_NotUV_characters> do
        SetTestingData(i,j);
        RemainingImgs=SelecImgs(i,j);
        SetTrainingData(70%,RemainingImgs);
        SetValidationData(30%,RemainingImgs);
        PerformTrainingTest(TrainingData,ValidationData);
        PerformValidation(TestingData);

    end
end
```

Algorithm 5.1: Data Strategy

The testing data corresponds to a pair of images where one character generates strangeness (*i* at Algorithm 5.1) and another character do not generate it (*j*). Then, remaining images compose the training and validation datasets.

We used the Support Vector Machine (SVM) [Fle09] model with three different kernels: linear, Radial-basis function (RBF) and polynomial. It also made a tuning of the hyperparameters through Grid Search using kernels RBF and polynomial. It implemented the SVM model using Sklearn [4]. The values used in the Grid Search parameters for the RBF kernel are for the gamma vector = $[0.3 * 0.001.0.001.3 * 0.001]$ and for the variable C = [50., 100., 200.].

With the Polynomial kernel the gamma values are showed in the vector $[0.001, 0.01]$, the degree variable with the values $[2, 3, 4]$ and the parameter coef0 which is a kernel projection parameter, the values $[0.5, 1]$.

We stored obtained results in a .csv file, showing if the entire face or parts of the face (as defined earlier) was used, as well as whether salience and PCA were used. It also showed the SVM model and the chosen kernel. Then, the values of precision, recall, F1 score, accuracy and time spent are also stored to further facilitate the selection of the main model.[5]

[4] https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html
[5] These data will be available upon request with the authors.

5.1.2 Voting Classifier Model

VotingClassifier [PVG+11] is an ensemble learning method available in Python's scikit-learn [VdWSNI+14] library. It combines the predictions of multiple models (classifiers) to improve overall predictive performance. Instead of relying on a single model, VotingClassifier aggregates the predictions of multiple classifiers and uses a vote (or weighted average) to decide the final class. There are two main types of voting in VotingClassifier:

- **Hard Voting**: Each model makes a prediction, and the final class is decided by a majority vote. That is, the class most frequently predicted by the models is the chosen class.
- **Soft Voting**: Instead of counting votes, the probabilities of each class are summed, and the class with the highest sum of probabilities is chosen. This only works if the classifiers can predict probabilities.

We use the Voting Classifier [PVG+11] with the Soft Voting type with the GT2 dataset . The classifiers used in this work and their respective parameters are:

- MLPClassifier(max_iter = 1000),
- LogisticRegression(max_iter = 1000),
- ExtraTreesClassifier(),
- DecisionTreeClassifier(),
- RandomForestClassifier(),
- GaussianNB(),
- KNeighborsClassifier(),
- SVC(probability=True),
- AdaBoostClassifier(),
- XGBClassifier(use_label_encoder=False), and
- CatBoostClassifier(logging_level='Silent').

We adopted the same algorithms to extract features from images in the Section 5.1.1, focusing in this case on specific regions of the face (forehead, eyes, nose, mouth, chin). We use the 40 characters from the GT2 dataset . We segmented the face into five distinct areas: forehead, eyes, nose, mouth and chin using the landmarks detected in the Mediapipe tool [LTN+19]. This approach was adopted to investigate which parts of the face can generate more frightening characteristics and whether the subjective evaluation agrees with this.

As previously established, Voting Classifier (VC) models were applied to the training dataset and the test dataset, which contains only one character at a time. In each run of the experiment, for each feature and in each training dataset, 6 models were created due to data standardization techniques (3 methods: standardization, normalization and logarithmic transformation) and dimensionality reduction (2 YES/NO possibilities). Among the techniques used, standardization is the StandardScaler function, which adjusts the data so that they have zero mean and unit standard deviation; normalization, which adjusts the data values to a specific range, usually $[0, 1]$; and logarithmic transformation, which aims to reduce the range of data values. For each technique, there was the possibility of activating it or not in the combination of techniques.

In addition, dimensionality reduction was implemented through Principal Component Analysis (PCA) and Random Forest (RF). In fact, in combinations where PCA is not applied, we tested RF to identify the three most relevant features of each region of the face [6]. The combination of these approaches allowed the creation of a wide variety of models, ensuring that the most relevant features were used in the classification process and providing a robust basis for evaluating the effectiveness of the different techniques used. We used F1-Score values as evaluation metrics.

Our methodology is depicted in Figure 5.3 and aims to use image technologies to predict human comfort perception of virtual human faces. While the method proposed in Section 5.1.1 used SVM (Support Vector Machine) algorithm to estimate the comfort perceived by people with respect to the entire virtual faces, we use Voting Classifier [PVG+11] to classify the comfort of parts of the faces. In addition, we propose to include the explainability investigation of our results using LIME [RSG16].

[6] As mentioned before, we tested RF with 1, 2 and 4 most relevant features, but the method performed best with 3 features

Pre-Processing Data

We perform four main processes to prepare the GT2 dataset to be used in our method:

(A) Face detection

To detect the face, we use the OpenFace [BRM16] framework.

(B) Cropping of Facial Regions

To cut out regions of the face (forehead, eyes, nose, mouth, and chin), we use the Mediapipe [LTN+19], which contains many more reference points (468 landmarks) available to carry out this process of cutting out parts of the face than OpenFace, which has only 68 landmarks.

Features Extract

After the detection and cropping phase of face regions, we extract features using the Hu Moments [ŽHR10] with the default settings, and HOG [DT05] algorithms setting the parameters orientations=9, pixels_per_cell=(64, 64), cells_per_block=(1, 1) and block_norm='L2-Hys'. Hu Moments extracts a feature vector of 7 positions. For HOG, we adopted the size mentioned in the Section 5.1.1, which is 9. Therefore, when utilizing Hu Moments, we generate a feature vector of size 7 for each facial part, resulting in a total of 35 features for the entire face. Conversely, employing HOG yields a 9-sized vector for each facial region, summing up to 45 features. Lastly, combining Hu Moments with HOG results in 80 facial features, with 16 features dedicated to each facial part.

(A) Reduce the complexity for dimensionality

To reduce the complexity, we can use PCA [PVG+11] for dimensionality reduction. For each region of the face, we performed a PCA using the Hu Moments or HOG vectors or the combination of both, generating a dimensionality reduction of 5 features, one for each part of the face, whose names are forehead, eyes, nose, mouth, and chin. When PCA was not used, we used Random Forest (RF) to select

the 3 most important variables from each region. Indeed, we tested with only 1, 2 and 4 variables with RF, but 3 was the best choice for our results.

Training the voting classifier model

Finally, we use the voting classifier [PVG+11], which is a machine learning method that trains on a set of several models and predicts an output (class) based on the highest probability of the class chosen as output. The idea is, instead of creating separate dedicated models and finding the accuracy of each of them, we create a single global model that trains on the specific ones and predicts based on the combined majority of votes for each production class. We use 11 machine learning algorithms as described in section 5.1.2.

5.1.3 Training and Testing using CNNs

We also propose fine-tuning a CNN to compare with the methods mentioned in the last sections, using the 40 characters from the GT2 dataset . Fine-tuning is a technique that involves training a pre-trained model (ImageNet) on a specific dataset. This model leverages the pre-trained model's prior knowledge and adjusts its final layers to adapt to a new dataset. VGG16 [SZ15a], Resnet50 [SZ15b], and MobileNet [HZC+17] were the neural networks chosen for this task.

In this experiment, we created two datasets for the training stage of the 3 CNNs. The first dataset was composed of the ground truth information of the face parts (defined in Table 4.2), that is, each character classified as strange and the parts of the face that the subjects considered the strangest (5 classes [7]: eyes, mouth and Comfort). The second training dataset, referring to the entire face, considering only the binary classification of the face as comfortable or uncomfortable. Afterward, both datasets were divided into three: training (70%), validation (20%), and testing (10%). This study also utilized the validation dataset to select and fine-tune ImageNet [8] parameters. To avoid bias, we trained 40 CNNs with 40 versions of the dataset GT2 , each time removing the character that should be tested. Each training was performed with batch sizes of 32 and 50 epochs. For model compilation, ADAM was utilized as the optimizer. The training process was developed in

[7] We included only two parts because there was no occurrence of faces with discomfort in the forehead, nose, chin region

[8] https://www.image-net.org/

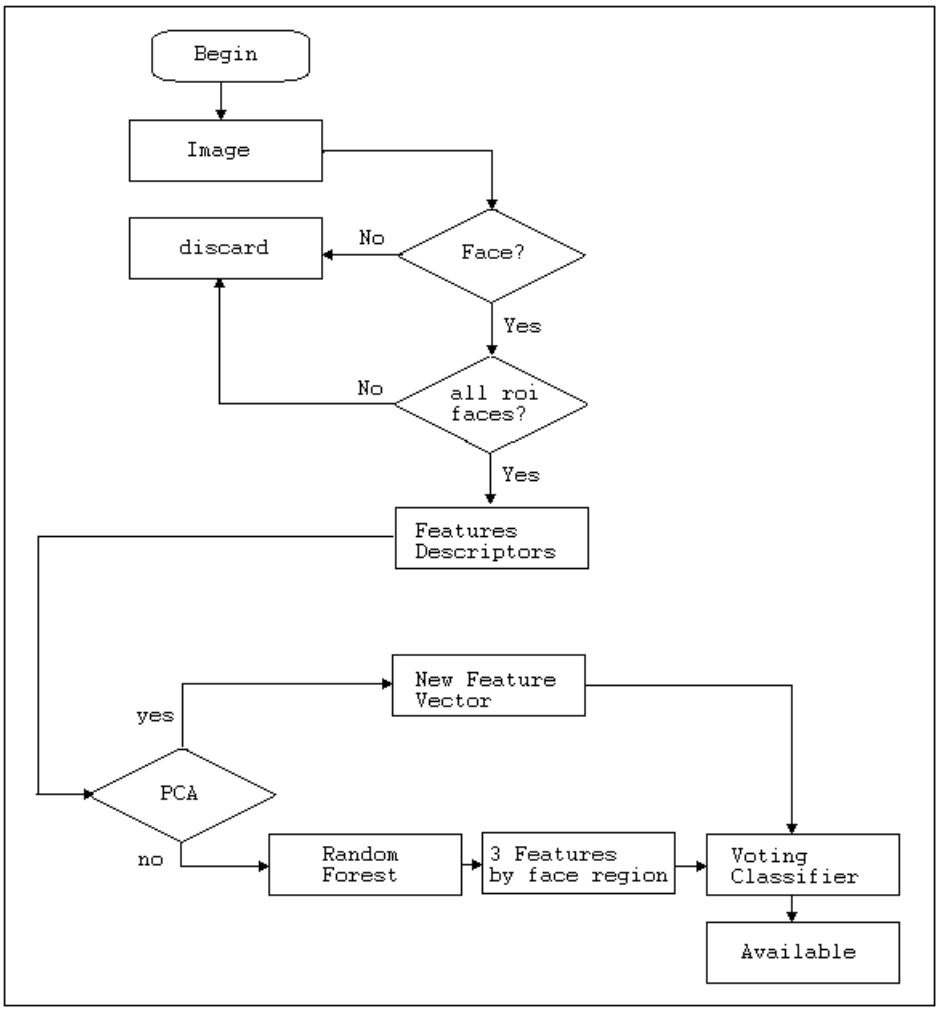


Figure 5.3: Overview of our model in Section 5.1.2: It starts with face detection, if there is a face, then it checks the 5 regions (ROIs - forehead, eyes, nose, mouth and chin) of the face. Then we extract features from the face parts (ROIs) using Hu Moments and HOG algorithms. PCA can be used to reduce the dimensionality of the feature vector and optionally we also test Random Forest. Finally, the voting classifier predicts whether the character will generate discomfort or not.

Python with the Keras library and backend with Tensorflow4. It was conducted on a PC running Windows 10 Pro, an Intel Core i5 6600K 3.5GHz, 32GB Memory RAM 2400MHz, 480GB SSD NVME m.2, and a GPU RTX 3070 8GB. Training took an average of 3 hours and 31 minutes for each Convolutional Neural Network (VGG16, Resnet50, and MobileNet), totaling approximately 11 hours.

5.2 The Computed Comfort Score (*CCS*) Metric

As mentioned previously, we propose the Computed Comfort Score (*CCS*) metric that aims to estimate the comfort perceived by humans, automatically using the SVR algorithm. For the SVR model, we used the GT2 dataset (Section 4.2) with 19 characters and the 40 characters. We also created another model using the ensemble voting regressor method to compare the results between the proposed models. In this VR model, we used 40 characters from GT2 dataset (Section 4.2). We considered naming the comfort estimated by the models *CCS*. The next sections detail the methods.

5.2.1 Support Vector Regressor (SVR) Model

This section discusses the model trained using the Support Vector Regression (SVR) algorithm. We introduce the *CCS* (Computed Comfort Score) metric to estimate the probable comfort/discomfort value of a certain virtual human face. In this model, we use local spatial and spectral entropy to extract features and show its relevance when compared to the subjects evaluation. We initially used the first 19 characters from the GT2 dataset (Section 4.2). We later retrained the model with the 40 characters from the GT2 dataset.

Pre-Processing Data

The overview of our method, illustrated in Figure 5.4, is inspired on proposed by Liu et al. in [LLHB14] for natural photographic images. In order to verify whether CG images contain pixels that exhibit strong dependencies in space and frequency, which carry relevant information about an image, we implemented a model that could extract characteristics from spatial and spectral entropy.

We implemented our method using OpenCV [How13], scikit-learn [VdWSNI+14] and dlib [Ros17].

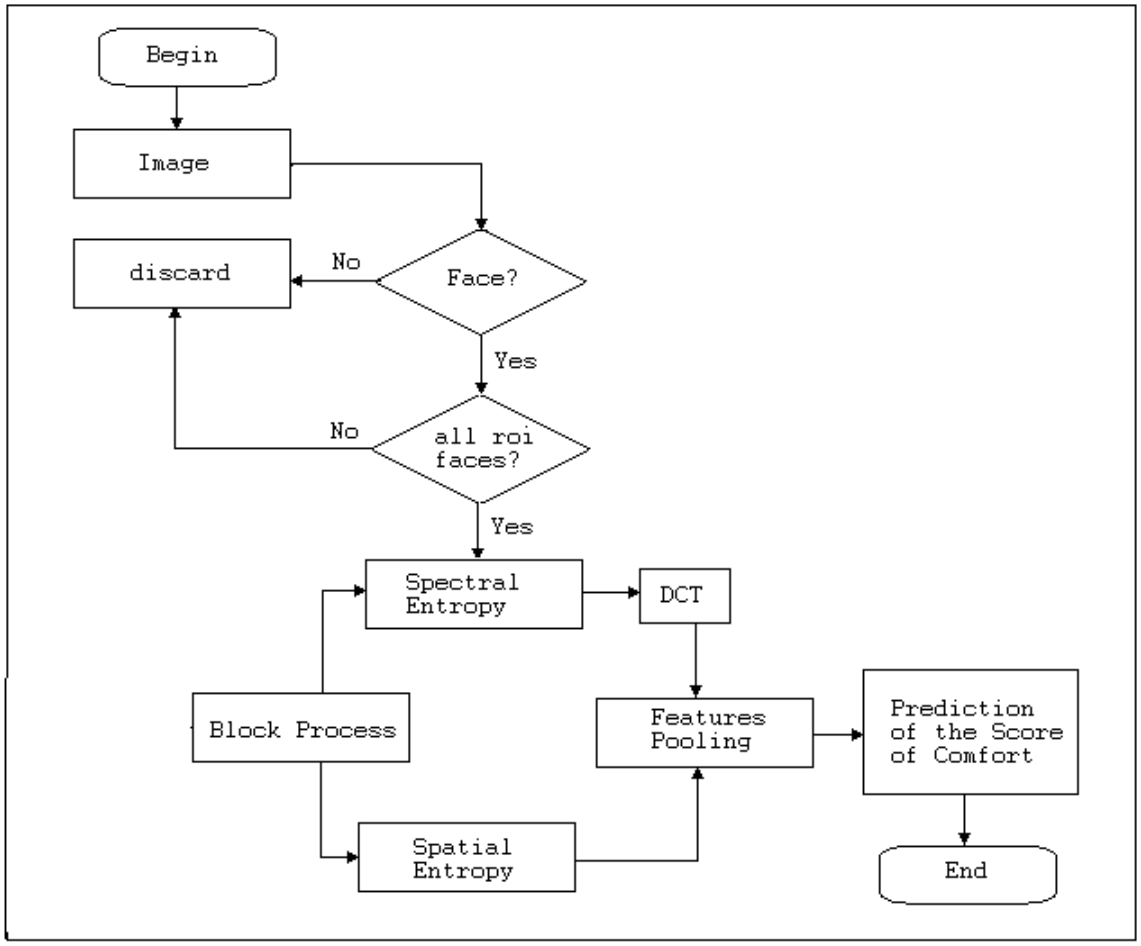


Figure 5.4: The overview of our model described in Section 5.1.1: It starts with face detection, if there is a face, then it checks the 5 regions (ROIs - forehead, eyes, nose, mouth, and chin) of the face. Then we extract the features from the face parts (ROIs) using the spectral and spatial entropy algorithm. Finally, the SVR algorithm is trained to predict whether the character will generate discomfort or not using our *CCS* metric.

(A) Face detection

The method used for face detection is proposed by Paul Viola and Michael Jones [VJ+01]. This method detects a face and also parts of the face. In the latter case, there are eight parts: mouth, middle of the mouth, right and left eyes, right and left eyebrows, nose, and jaw.

For our model, we assume that if no face is detected, or if the face is detected and the eight parts are not, the image is discarded. We do not use the mid-mouth region for our model because it is already inside the mouth, and the jaw

is not used because it has already evaluated the entire face. This model is using the first 19 characters from the GT1 dataset .

Features Extraction

In this step, we proceed with the features extraction. First, it resized each image to be a multiple of 2 and partitioned into 8x8 blocks. This block size is based on the work proposed by Liu et al. [LLHB14], who performed several experiments until setting M = 8 as a good block size value. We compute the spatial and spectral entropy characteristics locally for each block of pixels and each region of interest, i.e., the whole face and its parts. According to the definition of entropy of the image [Spo96], its main function is to describe the amount of information contained in an image. In the image quality assessment area [LLHB14], one of the motivating aspects is to identify the types and degrees of image distortions that generally affect their local entropy.

Spatial entropy calculates the probability distribution of the mean pixel values, while spectral entropy calculates the probability distribution of the global DCT (Domain Cosine Transform) coefficient values. We hypothesize that the local Spatial and Spectral entropy applied in Computer Graphics (CG) images may show statistical characteristics that correlate with perceptual data about CG faces. Indeed, this is the central hypothesis of the proposed *CCS* (Computed Comfort Score). To calculate the spatial entropy [9], we used the skimage.filters.rank library through function entropy(). To calculate the spectral entropy using FFT (Fast Fourier Transform) we use the scipy fftpack [10] library. To calculate the frequency map, the fft() function and then the dct() function were used to calculate the (DCT) domain cosine transform, both with default parameters.

(A) Features Pooling

At this stage, the entropy computation described in the previous step is used to calculate other characteristics for all pixel blocks of the face and its parts. The characteristics proposed in this work are mean, standard deviation, distortion, kurtosis, variance, Hu Moments [ŽHR10] and Histogram of Oriented Gradients (HOG) [DT05]. Hu Moments were used with its default parameters [ŽHR10], imple-

[9] https://scikit-image.org/docs/0.8.0/api/skimage.filter.rank.html
[10] https://docs.scipy.org/doc/scipy/reference/fftpack.html

mented using OpenCV [11] [How13], generating a vector of 7 positions. For HOG, the detection window with gradient voting into five orientation bins and 3x3 pixels blocks of 4x4 pixel cells was used in the spectral entropy features and 16x16 pixel cells in the spatial entropy features, generating a vector of 11 positions. It implemented HOG using scikit-learn [VdWSNI+14]. So, we have 23 features for spectral entropy and 23 for spatial entropy, proposing a total of 46 features.

The next section presents how the prediction of the comfort score of the face and face parts is computed. This step generates *CCS* for each CG face in the GT1 dataset .

Computing *CCS* using Support Vector Regression (SVR)

First, 19 initial characters from the GT1 dataset are used for training, testing, and validation, varying across these three groups until all characters are included in all groups. To perform SVR (Support Vector Regression), we propose nine models to test the impact of each group of entropy features: *i)* Hu (7 features) and HOG (11 features), and *ii)* mean, standard deviation, skewness, kurtosis, and variance. In addition, we want to evaluate the impact of spatial and spectral entropy, separately and together, and of the face and its parts (7 tested ROIs, the whole face, and six parts). Then, we propose nine combinations of the extracted data to use in the SVR model according to Table 5.1, in order to find the best precision of perceptual score:

We computed the nine models to evaluate which features better correlate with the perceived comfort regarding CG characters, i.e., the ground truth with perceptual data (GT).

The models generate individual values of comfort for each image from the short movie of each character, i.e., our proposed metric CCS_i for each character i in each frame f. Thus, to compute the *CCS* for each i character, in each video, we simply calculate the average *CCS* obtained at each f frame, from the movie that i participates in:

$CCS_i = Avg(\sum_{i=0}^{N_i} CCS_{i,f})$, where i is the index of character, N_i is the number of frames of the short movie and f is the frame index.

It is important to mention that although we can compute $CCS_{i,f}$ for character i at frame f, we do not have such information in the ground truth, once we have one comfort value for each character, as informed by the participants. We chose to

[11] https://opencv.org/

Model	Spatial Entropy	Spectral Entropy	S.F.	HOG	Hu Moments	T. C.
1	x	x	x	x	x	322
2	x	x	x	x		224
3	x	x	x		x	168
4	x		x	x	x	161
5	x		x	x		112
6	x		x		x	84
7		x	x	x	x	161
8		x	x	x		112
9		x	x		x	84

Table 5.1: Combination of nine models proposed to test the impact of each group of Entropy features. The column statistics' features correspond to mean, standard deviation, distortion, kurtosis, variance. The column Total characteristics (T.C.) refers to the number of characteristics evaluating the entire face and the six face parts according to the features selected in the previous columns. The column S.F. refers to Statistics Features.

consider the average value, because when the participant saw the video, we do not know when (at which frame or frames) the participant perceives strangeness.

We then retrain the model for the GT2 dataset containing the 40 characters. The goal is to be able to compare it with the next regression model (VR).

5.2.2 Voting Regressor Model (VR)

VotingRegressor [PVG+11] is an ensemble learning technique for regression, similar to VotingClassifier, but applied to regression problems. It combines the predictions of several regression models and generates a final prediction that is the average of the individual predictions. The goal is to improve the accuracy and robustness of the model by aggregating the predictions of different regressions. The two ways to use it are as follows:

- **Simple Regression (Simple Averaging)**: Instead of voting for a class, as in the case of classification, VotingRegressor calculates the average of the predictions of all regressions involved in the ensemble.
- **Base Models**: As with VotingClassifier, it is possible to use different regression models, such as linear regression, decision trees, neural networks, etc., to compose the ensemble.

We use the Voting regressor [PVG+11] with the Base Models type. The regressor algorithms used in this work and their respective parameters are:

- GradientBoostingRegressor('learning_rate': 0.001, 'max_depth': 5, 'n_estimators': 500),
- XGBRegressor('learning_rate': 0.1, 'max_depth': 3, 'n_estimators': 50),
- LGBMRegressor('learning_rate': 0.001, 'max_depth': 3, 'n_estimators': 100), and
- AdaBoostRegressor('base_estimator_max_depth': 5, 'learning_rate': 0.1, 'n_estimators': 1000).

The overview of our methodology is depicted in Figure 5.5 and aims to use image technologies to predict human comfort perception of virtual human faces. We use Voting Regressor [PVG+11] using way Base Models to predict the comfort of the vitual human faces. This model is trained with the 40 characters from the GT2 dataset .

Preprocessing Data

We performed four main processes to prepare the GT2 dataset (40 characters) to be used in our method: *A)* face detection, *B)* cropping of facial regions into 5 ROIs (Regions of Interest) and cropping of the entire face.

(A) Face detection

To detect the face, we used the OpenFace framework [BRM16].

(B) Cropping of Facial Regions

To crop regions of the face (forehead, eyes, nose, mouth and chin,) and the entire face, we use Mediapipe [LTN+19], which contains many more reference points (468 landmarks) available to perform this process of cropping parts of the face than OpenFace, which has only 68 reference points.

Features Extraction

After the detection and cropping phase of the facial regions and full face, we extract features using the algorithms indicated below using the default settings:

- AUs [ZLZ20]: 17 Action Units (AU01, AU02, AU04, AU05, AU06, AU07, AU09, AU10, AU12, AU14, AU15, AU17, AU20, AU23, AU25, AU26, AU45) were extracted by the OpenFace tool at the moment of face detection. According to the study by Mäkäräinen et al. [MKT14], intensifying facial expressions can increase emotional perception, but there is also a risk of extreme exaggeration, which can lead to feelings of awkwardness. The research suggests that exaggerated expressions, such as an amplified smile, can be seen as more emotional, however, when the intensity exceeds a certain limit, the observer's response can be negative, causing discomfort.

- Entropy [MAM22]: To calculate the spatial entropy [12], we used the skimage.filters.rank library through function entropy(). To calculate the spectral entropy using FFT (Fast Fourier Transform) we use the scipy fftpack [13] library. To calculate the frequency map, the fft() function and then the dct() function were used to calculate the (DCT) domain cosine transform, both with default parameters. Each entropy has a vector of 23 positions. The same entropy calculation performed in Section 5.2.1. The study by Liu et al. [LLHB14] identifies different types of distortions that affect the local entropy of images. Spatial and spectral entropy are calculated to measure the probabilistic distribution of pixel values and Discrete Cosine Transform (DCT) coefficients, respectively.

- GLCM (Gray Level Co-Occurrence Matrix) [HSD73]: is a method that analyzes the texture of images. It measures the frequency with which combinations of gray levels occur, capturing characteristics such as contrast, homogeneity and correlation. Studies such as that of Shahid et al. [SRLZ14] show that this method can be used to identify textures that are perceived as anomalous or degraded. Several statistical measures can be derived from it to characterize the texture and structure of the image. Some common features are contrast, dissimilarity, homogeneity, energy, and correlation. Each measure is composed of a vector of 3 positions, resulting in the extraction of a vector with 15 positions in this work.

[12] https://scikit-image.org/docs/0.8.0/api/skimage.filter.rank.html
[13] https://docs.scipy.org/doc/scipy/reference/fftpack.html

- Golden Ration [SMS08]: we use the neoclassical canons, which is a vector of 4 positions and the golden ratio that contains 21 calculated positions. Both determine the attractiveness of a face according to Schmidt et al. [SMS08].

- Hu Moments [ŽHR10]: the only algorithm that we can use to extract features from the entire face, a vector with 7 positions, and also from parts of it, a vector with 35 positions, since a vector with 7 positions is extracted for the 5 regions of the face. Timwell's study [TGNW11] explores how facial expressions on virtual characters influence human perception and the Uncanny Valley effect. When emotional expressions of virtual characters do not match typical human expressions, this can cause discomfort. According to Schmid et al. [SMS08], facial symmetry and neoclassical proportions play a central role in the perception of beauty. More symmetrical faces that follow golden proportions tend to be seen as more attractive.

(A) Reduce complexity for dimensionality

To reduce complexity, PCA [PVG+11] can be used for dimensionality reduction. For each region of the face and even the entire face, we perform a PCA using the vectors of each feature, generating a dimensionality reduction, whose number of components is defined as 95%. When PCA was not used, we used Random Forest (RF) to select the most important variables. Indeed, we tested with 1, 2 and 4 variables with RF, and 1 was the best choice for our results. When dealing with parts of the face, 1 variable from each region is selected; when dealing with the entire face, we average the highest importance value and the lowest value, selecting the features with an importance value greater than this average.

Training the Voting Regressor Model

Finally, we use the voting regressor [PVG+11], which is a machine learning method that trains on a set of several models and predicts an output (a value).

The proposal is to replace the creation of dedicated and separate models with a single global model. This model will be trained based on the individual models already defined and will make predictions by combining the majority of votes.

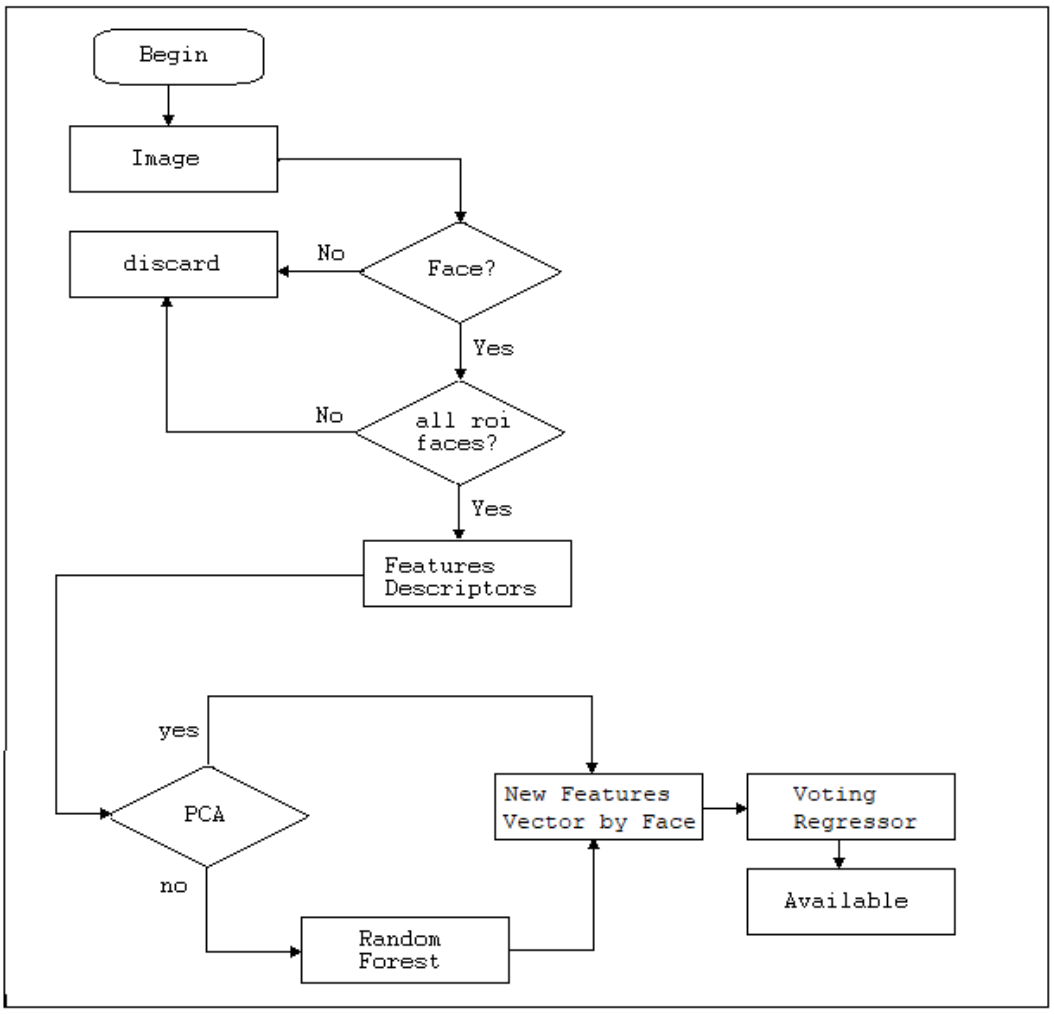


Figure 5.5: The overview of our model in Section 5.1.2: it starts with the face detection, if there is a face, then we find out the 5 regions (ROIs-forehead, eyes, nose, mouth, and chin). We extract the features of the entire face and the ROIs using AUs, Entropy, GLCM, Golden Ratio and Hu Moments algorithms. PCA can be used to reduce the dimensionality of the feature vector, and optionally we also test Random Forest. Finally, the voting regressor predicts the computed comfort score

Computing *CCS* using Voting Regressor (VR) Model

We adopt the same entropy extraction algorithm used by the SVR model in Section 5.2.2 to extract image features. We also use the AUs extracted by OpenFace, Hu Moments, GLCM and Golden Ratio algorithms as explained in section 5.2.2.

We segmented the face into five distinct bands (forehead, eyes and eyebrows, nose, mouth, and chin) only for the Hu Moments algorithm, which can extract information from parts of the face and the entire face. This approach was adopted

to investigate which parts of the face can generate more features of eeriness and whether the subjective evaluation agrees with this. We did not do the same treatment for the other algorithms because they process information from the entire face, such as aspect ratio, AUs, etc.

As previously established, the Voting Regressor (VR) models were applied to the training dataset with 40 characters and the test dataset, which varies to take into account all characters present in the GT2 dataset. In each experiment run, for each feature, and in each training dataset, 16 models were created because of data standardization techniques (3 methods: standardization, normalization, and logarithm transformation), dimensionality reduction (2 possibilities YES/NO) and a polynomial technique.

Among the techniques used, the standardization is the function StandardScaler, which adjusts the data so that they have a zero mean and unit standard deviation; normalization, which adjusts data values to a specific range, usually $[0, 1]$; and the logarithmic transformation, which aims to reduce the range of data values. For each technique, there was the possibility of activating it or not in the combination of techniques.

In addition, dimensionality reduction was implemented through Principal Component Analysis (PCA) and Random Forest (RF). Indeed, in combinations where PCA is not applied, we test the RF to identify the three most relevant characteristics of each region of the face [14].

The combination of these approaches allowed the creation of a wide variety of models, ensuring that the most relevant characteristics were used in the Regression process and providing a robust basis for evaluating the effectiveness of the different techniques used. In this work we used the RMSE (Root Mean Square Error) values as the evaluation metrics.

5.3 Chapter Considerations

This chapter presented the five models proposed in this work for detecting strangeness/discomfort, focusing on what we developed. The main objective was to show how each part of the model was conceived and assembled, as well as to

[14] As mentioned before, we tested RF with 1, 2 and 4 most relevant characteristics, but the method performed better with 1 characteristics

present the way in which the models were built. The next chapter discusses the experimental results obtained.

Table 5.2 summarizes the methods proposed in this chapter, indicating the technique used, whether it is a binary classification, regression, or CNN. It also presents the algorithms used to extract the features and indicates the dataset used to generate the model.

Model	Technique	Binary	Feature Extract	Face	Dataset
First	SVM	yes	Hu Moments and HOG	Entire	19 characters
					40 characters
Second	VC	yes	Hu Moments and HOG	Entire and Parts	40 characters
Third	CNN	yes	Image	Entire and Parts	40 characters
Fourth	SVR	no	Entropy	Entire	19 characteres
					40 characters
Fifth	VR	no	AUs, Entropy, GLCM, Golden Ratio and Hu Moments	Entire and Parts	40 characters

Table 5.2: This table summarizes the five proposed models, the algorithms used, and whether each model performs binary classification or regression, as indicated in the 'Binary' column. The 'Features' column specifies the features used to extract facial characteristics, while the 'Face' column indicates whether the entire face or only parts of it were analyzed. The 'Dataset' column specifies whether the first 19 characters of GT2 were used or all 40 characters.

6. EXPERIMENTAL RESULTS

This chapter presents the experimental results obtained with the five models presented in this work.

6.1 Binary Classification Models Results

In this section, we present the experimental results obtained with the binary models using the SVM algorithm, the ensemble Voting Classifier (VC) method, and CNN. Additionally, we show the comparison between the models.

6.1.1 Binary classifications using SVM

After computing SVM for the entire face and its parts using the first 19 characters from the GT2 dataset , the results did not show a significant difference in accuracy, although the obtained accuracy for the entire face is higher than for its parts. Because of this, we use the entire face on next evaluations using SVM. Therefore, we number the kernels: *1)* Linear pattern with data; *2)* Radial basis function (RBF) and *3)* Polynomial.

The suggested model returned 24 executions, which correspond to the features (Hu moments or HOG), with the detection or not of the saliency, with or without the reduction of dimensionality through the PCA and using the three kernel functions (linear (1), RBF (2) or polynomial (3)). In addition, we use leave-p-out cross-validation, where $p = 2$, i.e., p observations (causing strangeness or not), as the validation set and the remaining observations as the training set. This is repeated in all ways to cut the original sample on a validation set of p observations and a training set, resulting in 1920 runs. As metrics, we used the F1-score.

Figure 6.1 presents the results with and without the saliency function. The left side of the figure shows the results obtained without the saliency function, while the right side displays results with it. The values represent the average F1-Scores across 80 scenarios (5x16 in cross-data testing). On the left graph, 4 out of 12 executions achieve an F1-Score greater than 60%. Among these, 3 implementations utilize Hu moments features, and only 1 does not employ PCA. The best perfor-

mance is achieved with Hu Moments features combined with a polynomial kernel and dimensionality reduction via PCA, yielding an F1-Score of approximately 80%. On the right side of Figure 6.1, which shows the results with the saliency function, only one execution reaches an F1-Score close to 40%, while the remaining scores are considerably lower.

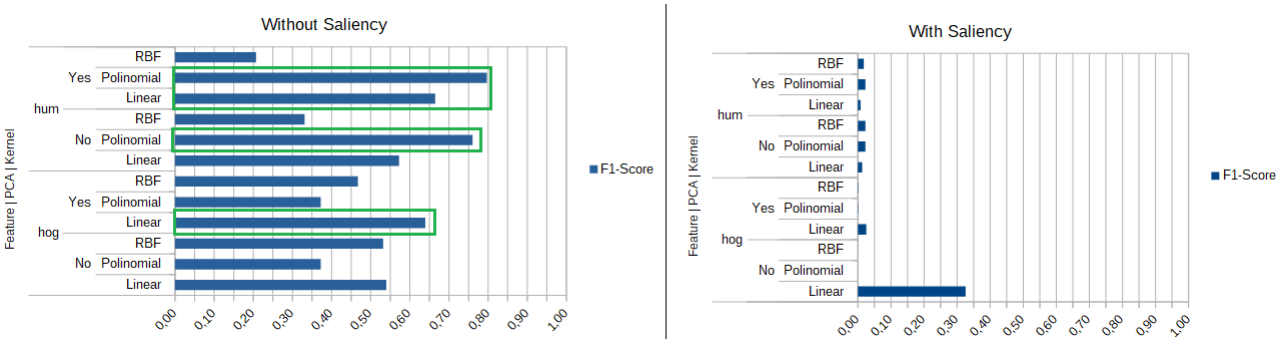


Figure 6.1: The F1-Score metric shows that there are 4 sets of implementations with F1-Scores values between 60% and 80% when classifying the characters in classes 0 (does not cause strangeness) and 1 (causes strangeness). The best implementation uses the Hu Moments feature, the polynomial kernel, which does not include the saliency of data and uses the dimensionality reduction, generating an F1-Score of approximately 80%.

In Figure 6.2, we present the computational time spent in executions from Figure 6.1. For the two implementations that presented F1-Score 76% and 80% in Figure 6.1 on the left, the time spent was 1560 seconds and 1335.08 seconds, respectively. For the other two implementations in which the F1-Score is 64% and 75%, the time is approximately 30 and 60 seconds, respectively.

Such data can be used to strike a balance between accuracy and computational time. For example, one option is to select the highest accuracy (80%) with the third best computational time (1335.08 seconds), achieved with Hu moments features, no saliency function, a polynomial kernel, and PCA. Alternatively, another viable option is to choose a slightly lower accuracy (65%) with significantly reduced computational time (60 seconds), using the same configuration but with a linear kernel.

Binary classification of characters

Using the best accuracy implementation obtained in the last section, we present binary classification results using the dataset GT1 containing the first 19

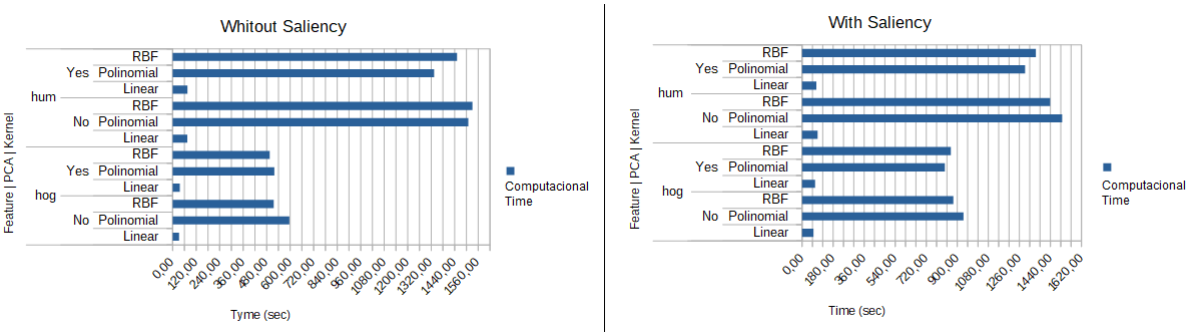


Figure 6.2: Computational time obtained in executions reported in Figure 6.1.

characters. First, we performed predictions using 80 runs (16 characters that generate comfort and 5 characters that do not) in the cross-validation test.

Firstly, we investigate the 5 characters that cause strangeness to people (highlighted in Figure 4.1). Table 6.1 shows the number of frames extracted from the videos of the 5 characters that cause strangeness in subjective evaluation. Those frames contain only the face of the 5 characters to be predicted in the implementation of this work. Table 6.2 presents the classification of such frames in the binary classification, where class 0 means that the character does not generate discomfort and class 1, the opposite.

Characters	Number of Frames
character (a)	1786
character (l)	45
character (c)	784
character (f)	131
character (i)	249

Table 6.1: Number of frames extracted from the videos of the 5 characters that cause strangeness in subjective evaluation. These characters correspond to the highlighted characters in Figure 4.1.

In a subjective evaluation, we notice that facial expressions, together with the movement of the head and body, can contribute to a distortion of visual characteristics, resulting in a possible strangeness to human eyes. It is possible to notice that effect in the smaller value obtained for Class 1 character (f) in Table 6.2. We hypothesize that this happens with the character (f) because there are many facial deformations in this case if compared with other faces, as (a) and (l). Tinwell et

al. [TGNW11] already make references in their work on the issue of facial expression and also on the movement of bodies as important factors for detecting strangeness.

Characters	% Class 0	% Class 1
character (a)	20.00	**80.00**
character (l)	2.36	**97.64**
character (c)	20.03	**79.97**
character (f)	44.65	**55.35**
character (i)	13.07	**86.93**

Table 6.2: Percentage of class 0 and class 1 for each character after prediction by the implementation model.

As we can see in Table 6.2 it classified these 5 characters as belonging to class 1 (as major part of frames) using our proposed model, which matches with the subjective evaluation with people.

In addition, we evaluated the remaining 14 characters that do not cause discomfort to people, according to subjective evaluation. Table 6.3 shows the number of frames extracted from the videos of such 14 characters. As before, these frames contain only the faces of the characters analyzed to be classified in our work. For

Characters	Number of Frames
character (c)	610
character (b)	552
character (v)	427
character (t)	402
character (m)	207
character (h)	175
character (o)	145
character (n)	80
character (k)	72
character (p)	63
character (s)	34
character (d)	21
character (r)	15
character (q)	1

Table 6.3: Number of frames extracted from the videos of the 14 characters that do not cause strangeness in subjective evaluation.

the 14 characters that do not cause discomfort to people, only 4 of them were incorrectly classified as belonging to class 1, as seen in Table 6.4. So, the obtained

error rate is 28% in the 14 characters that do not generate discomfort in subjective evaluation. Considering the full dataset GT1 of 19 characters (and 5799 frames),

Table 6.4: Percentage of class 0 and class 1 for each character after prediction by the implementation model.

Characters	% Class 0	% Class 1
character (c)	**97.44**	2.56
character (b)	**99.80**	0.20
character (v)	**98.75**	1.25
character (t)	**67.58**	32.42
character (m)	**75.59**	24.41
character (h)	**99.57**	0.43
character (o)	20.36	**79.64**
character (n)	**76.61**	23.39
character (k)	**99.03**	0.97
character (p)	20.00	**80.00**
character (s)	**99.72**	0.28
character (d)	0.06	**99.94**
character (r)	**99.83**	0.17
character (q)	20.00	**80.00**

we obtained an accuracy of approximately 81% in characters classification and 82% considering the image classification.

After expanding the dataset to 40 characters, we used the best parameters to generate a new model using the GT2 dataset. According to Table 6.5, the result was not satisfactory since the accuracy was 50%, while the median of the F1-Score metric indicated 64.03%.

Character	Frames	GT	Class 0	Class 1	Prediction	Agreements
1	260	1	65	195	1	Agree
2	60	0	11	49	1	Disagree
3	260	1	72	188	1	Agree
4	66	1	37	29	0	Disagree
5	260	1	83	177	1	Agree
6	232	0	61	171	1	Disagree
7	52	1	15	37	1	Agree

8	260	1	91	169	1	Agree
9	117	0	15	102	1	Disagree
10	260	1	49	211	1	Agree
11	442	0	119	323	1	Disagree
12	83	0	18	65	1	Disagree
13	59	1	11	48	1	Agree
14	117	0	49	68	1	Disagree
15	59	1	26	33	1	Agree
16	386	0	127	259	1	Disagree
17	216	1	79	137	1	Agree
18	60	0	18	42	1	Disagree
19	420	0	168	252	1	Disagree
20	93	1	39	54	1	Agree
21	260	1	37	223	1	Agree
22	487	0	207	280	1	Disagree
23	205	0	72	133	1	Disagree
24	164	1	24	140	1	Agree
25	260	1	61	199	1	Agree
26	260	1	103	157	1	Agree
27	33	0	21	12	0	Agree
28	241	1	34	207	1	Agree
29	64	1	49	15	0	Disagree
30	382	0	198	184	0	Agree
31	260	1	68	192	1	Agree
32	245	0	99	146	1	Disagree
33	253	0	76	177	1	Disagree
34	53	1	30	23	0	Disagree
35	260	1	129	131	1	Agree
36	111	0	39	72	1	Disagree
37	260	1	90	170	1	Agree
38	97	0	18	79	1	Disagree
39	56	0	23	33	1	Disagree
40	121	0	42	79	1	Disagree

Table 6.5: Evaluation of the predicted classes for the characters included in our SVM model test dataset, using the dataset GT2 (Section 4.2) balanced frames by class. Predictions for all 40 characters were included, along with the ground truth (GT) for all characters and the number of frames for each character (Frames). Class 0 is considered comfortable, while class 1 is uncomfortable. The prediction column indicates the predominant class predicted for the character. The comparison between the prediction of our SVM model and GT was evaluated in the Agreement column. The result agree indicates agreement, and disagree indicates disagreement between the predictions and GT.

6.1.2 Binary classifications using Voting Classifier

Our investigation analyzes the accuracy of models generated with Voting Classifier, using the F1-Score metric to classify facial regions into Comfortable and Uncomfortable. This classification indicates the prediction about people's perceptions.

Table 6.6 presents the median F1-Score metric of the application of the algorithms (HOG, HU Moments and HOG+Hu Moments) using the GT2 training and testing dataset (40 characteres). We also report here the highest median F1-Score obtained for each algorithm: i) F1-Score=91.11%, for the Hu Moments algorithm, ii) F1-Score=66.32%, for the HOG algorithms, and iii) F1-Score =86.08%, for the Hu Moments + HOG algorithms. For this reason, we consider that the Hu Moments algorithm, using logarithm data standardization and without dimensionality reduction, presents the best median F1-Score.

6.1.3 Training and testing results with CNNs

In the first experiment, the performance evaluation metrics of Average Precision, Loss, and F1-Score were taken as reference, as shown in Table 6.7. As can be seen, we obtained low accuracy rates for the three models created, namely 45.96%, 37.10% and 11.30% for VGG16, ResNet50 and MobileNet, respectively. One possibility for the low scores obtained is the lack of proportionality (balance) in the data between the five classes. Although there is a balance between the two classes used (eyes and mouth), there is still little data. Only 18 characters that

Feature	Standard	Logarithm	Normalized	PCA	Median F1-Score	Median Elapsed Time
HOG	**n**	n	y	n	0.60	2179.8
				y	0.61	1321.2
		y	**n**	**n**	**0.66**	**1912.20**
				y	0.62	1426.80
	y	n	n	n	0.64	1887
				y	0.61	1354.80
Hu Moments	**n**	n	y	n	0.76	1509
				y	0.64	1411.80
		y	**n**	**n**	**0.91**	**1149.60**
				y	0.58	1155.60
	y	n	n	n	0.72	1323
				y	0.66	1192.80
Hu Moments + HOG	**n**	n	y	n	0.81	1466.40
				y	0.63	1422
		y	**n**	**n**	**0.86**	**1151.40**
				y	0.59	1153.80
	y	n	n	n	0.69	1251,60
				y	0.66	1222.20

Table 6.6: Evaluation of Voting Classifier models using the GT2 dataset using 40 characteres, with balanced frames by class, based on the feature used (Hu Moments or HOG), data standardization method (standard, logarithmic, or normalized), and with or without dimensionality reduction (PCA). The best median F1-Score was 91%, achieved using Hu Moments, logarithmic data standardization, and without dimensionality reduction. The 'Median Elapsed Time' column displays the time in seconds.

cause strangeness are involved in this procedure because they were the only ones whose eyes and mouths were identified as uncomfortable for people.

In the second experiment, we trained the models based on the results of the entire face, which resulted in binary classification, comfort (0) and uncomfortable (1). In Table 6.8, these values clearly show an improvement in F1-Score values when compared to the parts of the face. Resnet50 achieved the highest F1-Score value (77.90%) compared to VGG16 (56.69%) and MobileNet (38.86%), but the accuracy is very low (34.63%) while VGG16 (46.63%) and MobileNet (35.96%).

	Face Parts		
	VGG16	ResNet50	MobileNet
Accuracy	0.4596	0.3710	0.113
Recall	0.4596	0.3710	0.1113
Precision	1.0	1.0	1.0
F1-Score	0.4979	0.4370	0.115

Table 6.7: Results of CNN fine tuning concerning the prediction of subjective discomfort to parts of face.

	All Face		
	VGG16	ResNet50	MobileNet
Accuracy	0.4663	0.3463	0.3596
Recall	0.5102	0.6384	0.3827
Precision	1.0	1.0	1.0
F1-Score	0.5669	0.7790	0.3886

Table 6.8: Results of CNN fine tuning concerning the prediction of subjective discomfort to the entire face.

6.1.4 Comparing results of Binary Models SVM and GT

Table 6.9 compares the two binary models, SVM and VC, applied on GT2. For the SVM model, there was a 50% accuracy (20 characters out of 40) in relation to the GT2. When comparing the VC model with the GT2, there was an agreement of 67.5% (27 characters). When comparing the two models, the agreement ratio was 62.5% (25 characters). When comparing the accuracy of CNN with the two previous models, the percentage is lower, 45.96% using VGG16 for the face parts and 46.63% with VGG16 when trained with the entire face. Of these three models presented, we consider the VC method to be the one that indicated the best accuracy.

			SVM Model			VC Model			Comparison between		
character	**Frames**	**GT**	**Class 0**	**Class 1**	**Predictions**	**Class 0**	**Class 1**	**Predictions**	**SVM and GT**	**VC and GT**	**SVM and VC**
1	260	1	65	195	1	0	260	1	Agree	Agree	Agree
2	60	0	11	49	1	30	30	1	Disagree	Disagree	Agree
3	260	1	72	188	1	0	260	1	Agree	Agree	Agree
4	66	1	37	29	0	0	66	1	Disagree	Agree	Disagree
5	260	1	83	177	1	99	161	1	Agree	Agree	Agree
6	232	0	61	171	1	198	34	0	Disagree	Agree	Disagree
7	52	1	15	37	1	0	52	1	Agree	Agree	Agree
8	260	1	91	169	1	0	260	1	Agree	Agree	Agree
9	117	0	15	102	1	96	21	0	Disagree	Agree	Disagree
10	260	1	49	211	1	27	233	1	Agree	Agree	Agree
11	442	0	119	323	1	442	0	0	Disagree	Agree	Disagree
12	83	0	18	65	1	34	49	1	Disagree	Disagree	Agree
13	59	1	11	48	1	38	21	0	Agree	Disagree	Disagree
14	117	0	49	68	1	100	17	0	Disagree	Agree	Disagree
15	59	1	26	33	1	8	51	1	Agree	Agree	Agree
16	386	0	127	259	1	71	315	1	Disagree	Disagree	Agree
17	216	1	79	137	1	10	206	1	Agree	Agree	Agree
18	60	0	18	42	1	7	53	1	Disagree	Disagree	Agree
19	420	0	168	252	1	1	419	1	Disagree	Disagree	Agree
20	93	1	39	54	1	88	5	0	Agree	Disagree	Disagree

21	260	1	37	223	1	117	143	1	Agree	Agree	Agree
22	487	0	207	280	1	35	452	1	Disagree	Disagree	Agree
23	205	0	72	133	1	6	199	1	Disagree	Disagree	Agree
24	164	1	24	140	1	110	54	0	Agree	Disagree	Disagree
25	260	1	61	199	1	18	242	1	Agree	Agree	Agree
26	260	1	103	157	1	81	179	1	Agree	Agree	Agree
27	33	0	21	12	0	33	0	0	Agree	Agree	Agree
28	241	1	34	207	1	50	191	1	Agree	Agree	Agree
29	64	1	49	15	0	1	63	1	Disagree	Agree	Disagree
30	382	0	198	184	0	369	13	0	Agree	Agree	Agree
31	260	1	68	192	1	21	239	1	Agree	Agree	Agree
32	245	0	99	146	1	151	94	0	Disagree	Agree	Disagree
33	253	0	76	177	1	15	238	1	Disagree	Disagree	Agree
34	53	1	30	23	0	15	38	1	Disagree	Agree	Disagree
35	260	1	129	131	1	189	71	0	Agree	Disagree	Disagree
36	111	0	39	72	1	104	7	0	Disagree	Agree	Disagree
37	260	1	90	170	1	18	242	1	Agree	Agree	Agree
38	97	0	18	79	1	0	97	1	Disagree	Disagree	Agree
39	56	0	23	33	1	56	0	0	Disagree	Agree	Disagree
40	121	0	42	79	1	105	16	0	Disagree	Agree	Disagree

Table 6.9: Evaluation of the predicted classes for the characters included in our model testing dataset. The predictions of the SVM model for the 40 characters were included, along with the number of frames (Frames) and the ground truth (GT2) of all characters. Class 0 is considered comfortable, while class 1 is uncomfortable. The comparison between the predictions of our models (VC) and SVM, as well as between the VC and GT2 models and also the comparison between the VC, SVM and GT2 models were evaluated. The result agree indicates agreement and disagree indicates disagreement between predictions and/or GT2. Characters in bold represent those that generate discomfort in human perception in GT2. Bold lines indicate that the models agreed with GT2.

6.2 Result of the Regression Models

In this section, we present the experimental results obtained with the regression models using the SVR algorithm and the ensemble Voting Regressor (VR) method. We call the estimated comfort predicted by the models *CCS* and show the comparison between the models.

6.2.1 Computed Comfort Score (*CCS*) Results

First, we investigate the accuracy obtained with the nine models presented in Table 5.1 using the first 19 characters where a value of binary classification is computed for each character. In addition, we evaluated the error obtained when we confronted the CCS_i obtained value and the ground truth value of comfort for each character i. Then, we provide an analysis to find out the part of the faces that generates more discomfort with our method. We investigate a hypothesis, transforming all CG characters into cartoons and calculating the *CCS* again.

Evaluating *CCS* values as a binary classification of comfort

First, we present the binary classification result regarding the 19 CG characters, using the nine models (presented in Table 5.1) and the whole face. We consider that characters in which perceptual comfort is $<$ 60%, in the ground truth, can generate discomfort in the human perception, while remaining characters generate comfort, i.e., perceptual comfort $>=$ 60%. Table 6.2 shows the five characters that generate discomfort in human perception and the result of binary classification using *CCS* values with the same threshold as in the ground truth, i.e., discomfort if *CCS* $<$ 60% and comfort if *CCS* $>=$ 60%. It presented a similar analysis in Table 6.4 with characters that generate comfort in human perception. In Table 6.10, "*" shows that the classification was correct, while "-" was not correct in comparison with the ground truth for the 9 models.

As we can see in Table 6.10, Models 1 and 6 seem to be more adequate than others to provide a correct classification of the last five characters that generate strangeness or discomfort in the individuals. Models 7 and 8, in Table 6.10, present 100% of correct classification with characters that are comfortable, according to hu-

Character	Number of Frames	1	2	3	4	5	6	7	8	9
(b)	553	*	*	*	*	-	-	*	*	*
(d)	17	*	*	-	*	*	*	*	*	*
(e)	610	-	-	*	-	-	-	*	*	*
(g)	2	*	*	*	*	*	*	*	*	*
(h)	164	*	*	*	*	*	*	*	*	*
(k)	72	*	*	-	-	*	-	*	*	*
(m)	209	*	*	*	*	*	-	*	*	*
(n)	74	*	*	*	*	*	*	*	*	*
(o)	145	*	*	*	*	*	-	*	*	-
(p)	60	*	*	*	*	*	*	*	*	-
(r)	18	*	*	*	*	*	*	*	*	*
(s)	21	*	*	-	*	*	*	*	*	*
(t)	403	-	-	*	*	*	*	*	*	*
(v)	428	-	-	*	*	-	*	*	*	*
(a)	1786	-	-	*	*	*	*	-	-	-
(c)	745	*	*	-	*	-	*	*	*	-
(f)	148	*	*	-	*	-	*	*	*	-
(i)	250	*	*	-	-	-	*	-	-	-
(l)	33	*	-	-	-	-	-	-	-	-

Table 6.10: Number of frames extracted from the videos of the 19 characters and result of binary classification with computed comfort using the 9 studied models. The symbol "-" shows the incorrect classification while "*" shows the opposite. The last 5 characters (a, c, f, i, l) correspond to the highlighted characters in Figure 4.1.

man perception. When evaluating all the characters together that present discomfort and comfort in people's perception on Table 6.10, we noticed that the best model, in this case, is Model 1 with approximately 80% of average accuracy, considering both groups of characters. One can say that Models 7 and 8 also seem accurate, but in fact, such models classified incorrectly more than half of characters that generate strangeness/discomfort, maybe showing a tendency in generating high values of computed comfort (*CCS*). In addition, the RMSE between *CCS* obtained values and the comfort value in the ground truth, for the 19 evaluated characters is 23.59.

Table 6.11 shows the *CCS* metric and the perceived comfort values (GT1). We considered calculating the average of the *CCS* metric of the computed values to be the threshold to indicate whether the computed comfort was uncomfortable or comfortable. The threshold was 60%, so if the value was less than or equal to 60% it was considered uncomfortable, otherwise it was comfortable. Thus, we obtained an agreement of 85.71% for the comfortable class and 60.0% for the uncomfortable class in relation to each character.

It is important to notice that Model 1 accuracy (80%) is very similar to results obtained in the previous work [DMND+21a] using SVM (also 80%) when we evaluated the total number of images per class. When evaluating accuracy by character and class, we have 85% accuracy in relation to the comfortable class and 60% in relation to the uncomfortable class.

Character	Perceived Comfort (%)	CCS(%)
a	41.17	60.25
b	68.90	61.97
c	26.89	59.34
d	84.87	86.91
e	65.54	55.04
f	35.29	44.52
g	52.10	100
h	73.10	74.56
i	24.37	57.30
k	91.59	73.62
l	37.81	61.38
m	88.23	64.53
n	71.43	100
o	92.43	83.13
p	92.43	73.98
r	81.51	100
s	89.08	93.51
t	85.71	60.77
v	79.83	59.71

Table 6.11: Evaluation of 19 characters from the GT1 dataset, according to the following attributes: characters, human-perceived comfort rating and calculated CCS.

6.2.2 Perception of comfort using SVR for face parts

Considering that a specific part of the face can cause discomfort, we investigated the parts of the face that cause more discomfort/strangeness. Analyzing the perceptual data, subjects comment that first part of the face that causes strangeness is the eyes followed by the mouth and nose. Taking the five characters that generate discomfort in the perceptual study, we observed that the nose and eyes are the parts of the face with smaller values of *CCS*. In the perceptual study, 11 from 14 characters that do not generate strangeness present the mouth as the region is

less comfortable, being eyes and nose the less comfortable for the three remaining characters.

It is interesting to remark, that there are few variations concerning the *CCS* computed for face parts and compared with perceived comfort. Values of RMSE for each face part, compared with perceived comfort (ordered from the lowest error to the higher) are following presented: 21.15 for the nose, 22.40 for left_eyebrow, 22.52 for right_eyebrow, 22.93 for the left_eye, 23.89 for the right_eye, and 24.63 for the mouth. Although the average error of the parts of the face (22.92) is slightly less than *CCS* for the full face (23.59), these values are not got with the same model. For example, Model 6 is used to get the best *CCS* for the left eye, left eyebrow, and right eyebrow; and Model 4 is the most suitable for the right eye. In fact, when analyzing model by model, none achieved better accuracy than Model 1 for the entire face.

After expanding the dataset to 40 characters as reported in Section 4.2, we used the best parameters to generate a new model SVR, using the GT2 dataset (40 characters). According to Table 6.14, the result indicates that the median of the RMSE metric is 24% error. There are 25 characters in the first three bands and 15 in the last two.

6.2.3 CCS using Voting Regressor

Our investigation focuses on analyzing the accuracy of models generated with the ensemble Voting Regressor method, using the RMSE metric to measure error residual in relation to comfort which indicates the prediction about people's perception, using the GT2 dataset.

Table 6.12 presents the median of the RMSE metric of the application of the training and testing algorithms (AUs, Entropy, GLCM, Golden Ratio and Hu Moments). We also report here the smallest median RMSE error measure obtained for each algorithm: *i)* RMSE = 18.55%, for AUs, *ii)* RMSE = 13.88%, for the Entropy algorithm, *iii)* RMSE = 18.12%, for the Golden Ratio algorithm, *i)* RMSE = 16.79%, for GLCM, *ii)* RMSE = 16.44%, for the Hu Moments algorithm evaluating the entire face like the previous algorithms, and *iii)* RMSE = 15.55%, for the Hu Moments algorithm when extracting the feature from parts of the face.

Algorithm	Face	Standard	Logarithm	Normalized	PCA	Median RMSE	Median Elapsed Time
AUs	Entire	n	n	y	n	18.55%	79.26
					y	17.57%	105.31
			y	n	n	19.76%	60.02
					y	20.11%	94.99
		y	n	n	n	20.29%	59.91
					y	21.16%	95.27
Entropy	Entire	n	n	y	n	13.88%	83.00
					y	19.98%	69.33
			y	n	n	17.52%	103.96
					y	21.97%	54.70
		y	n	n	n	21.01%	104.17
					y	21.59%	85.31
Golden Ratio	Entire	n	n	y	n	18.74%	145.82
					y	19.42%	61.15
			y	n	n	22.22%	85.50
					y	18.12%	38.74
		y	n	n	n	20.50%	107.81
					y	22.14%	74.07
GLCM	Entire	n	n	y	n	23.08%	141.83
					y	21.52%	30.42
			y	n	n	16.79%	107.16
					y	18.09%	19.78

		y	n	n	n	19.05%	106.10
					y	23.79%	48.48
Hu Moments	Entire	n	n	y	n	16.44%	43.07
					y	19.28%	16.96
			y	n	n	17.59%	38.93
					y	24.41%	21.41
		y	n	n	n	23.64%	35.30
					y	26.64%	20.74
	Face Parts	n	n	y	n	16.80%	54.95
					y	20.89%	26.23
			y	n	n	15.55%	61.15
					y	22.22%	32.59
		y	n	n	n	20.93%	62.79
					y	25.40%	38.52

Table 6.12: Evaluation of the median values of the RMSE metric according to the extraction of features from the whole face or parts of it, standardization of the data (standard, logarithmic and normalized) and whether dimensionality reduction (PCA) is performed or not. The Median Elapsed Time column is shown in minutes. The * indicates that the algorithm extracts features from the entire face and ** refers to the extraction of features from parts of the face (forehead, eyes, nose, mouth, chin) for the Hu Moments.

Since the error measures (RMSE) are quite approximate, we decided to divide the RMSE into bands from 1 to 5 and count the number of characters for each band. For the first three bands, the estimated RMSE is up to 30% error, according to the perceptual comfort of the GT2 dataset (section 4.2). Therefore, the greater the concentration of characters in these first 3 bands, indicates that the model came closer to the comfort perceived by people. The last two bands indicate an error residual rate above 30%, which shows a greater distance from the comfort perceived by humans. Therefore, the lower the concentration of characters in these last two bands, the better the comfort result estimated by the model. Table 6.13 shows the distribution of characters by error measurement ranges (RMSE) and by algorithms. We observed that the largest concentration of characters is found in ranges 1, 2 and 3 in all algorithms, ranging from 28 to 32 characters, corresponding to 70% to 80% of the characters in the GT2 dataset (section 4.2). Hu Moments, either * or **, was the algorithm with the largest number of characters (32) in these first 3 ranges. We also verified that the Entropy and AUs algorithms indicated 32 characters in these initial ranges. In range 4, the algorithm that quantified the fewest characters was Entropy, while in range 5 it was Hu Moments*. Considering the importance of the bands (Table 6.13) and also the median of the RMSE of the Table 6.12, we understand that the best algorithm is Hu Moments** although the lowest RMSE was with the Entropy algorithm.

Range RMSE		Aus	Entropy	Golden Ratio	GLCM	Hu Moments*	Hu Moments**
1	[0.0; 0.1]	10	12	13	13	13	9
2	(0.1; 0.2]	13	12	8	11	12	15
3	(0.2;0.3]	8	7	7	5	8	8
4	(0.3,0.4]	5	4	5	7	6	6
5	(0.4,1.0)	4	5	7	4	1	2

Table 6.13: Evaluation of the number of characters per measurement interval of the error metric (RMSE). The first three bands indicate greater proximity of the comfort estimated by the model in relation to the perceived comfort. While the last two bands represent the opposite. Therefore, the greater the number of characters in the first three bands, the better the comfort estimated by the model. The * indicates that the algorithm extracts features from the entire face and ** refers to the extraction of features from parts of the face (forehead, eyes, nose, mouth, chin).

6.2.4 Comparing results obtained with Regression Models (SVR and VR)

In this section, we evaluate and compare VR and SVR results. As presented in the last Section 6.2.3, we computed the RMSEs of the VR models according to the data standardization (standard, logarithm, nomalized), the algorithms (AUs, Entropy, Golden Ratio, GLCM and Hu Moments), in addition to the dimensionality reduction with PCA. The best VR model was extracted by extracting features with the Hu Moments algorithm (parts of the face) , data, using the logarithmic transform and without dimensionality reduction. In Table 6.12 we report the median RMSE of the combinations between the models generated and explored in the last section.

We retrained the SVR model according to the model features reported in Section 5.2 with the GT2 dataset using 40 characteres so that we can compare our Voting Regressor (VR) method and the SVR model. Table 6.14 shows that there is a higher concentration of characters in bands 1, 2 and 3, indicating that the majority of the characters (62.5%) fall into them. The median RMSE was 24.19%.

Range RMSE		SVR
1	[0.0; 0.1]	6
2	(0.1; 0.2]	10
3	(0.2;0.3]	9
4	(0.3,0.4]	8
5	(0.4,1.0)	7

Table 6.14: Evaluation of the number of characters in the Amount column by measurement error interval (RMSE). The greater the number of characters in the first three intervals, the lower the measurement error, being closer to the expected comfort of the Ground Truth. Intervals 4 and 5 indicate high measurement error, so the fewer characters in these intervals, the better.

Figure 6.3 shows the comparison between the two regression models (Hu Moments) using 40 characters from the GT2 dataset . We can see that the SVR and VR models have a similar trajectory when evaluating the median RMSE (Root Mean Square Error) of the videos.

The SVR model has a fairly uniform distribution among the RMSE bands according to Table 6.14 and concentrates 62.5% of the characters in the first 3 bands. However, the VR model, according to Table 6.13, presents 80.0% of the

characters in the same bands, with a smaller distance in the error in the prediction in relation to the perceptive GT2.

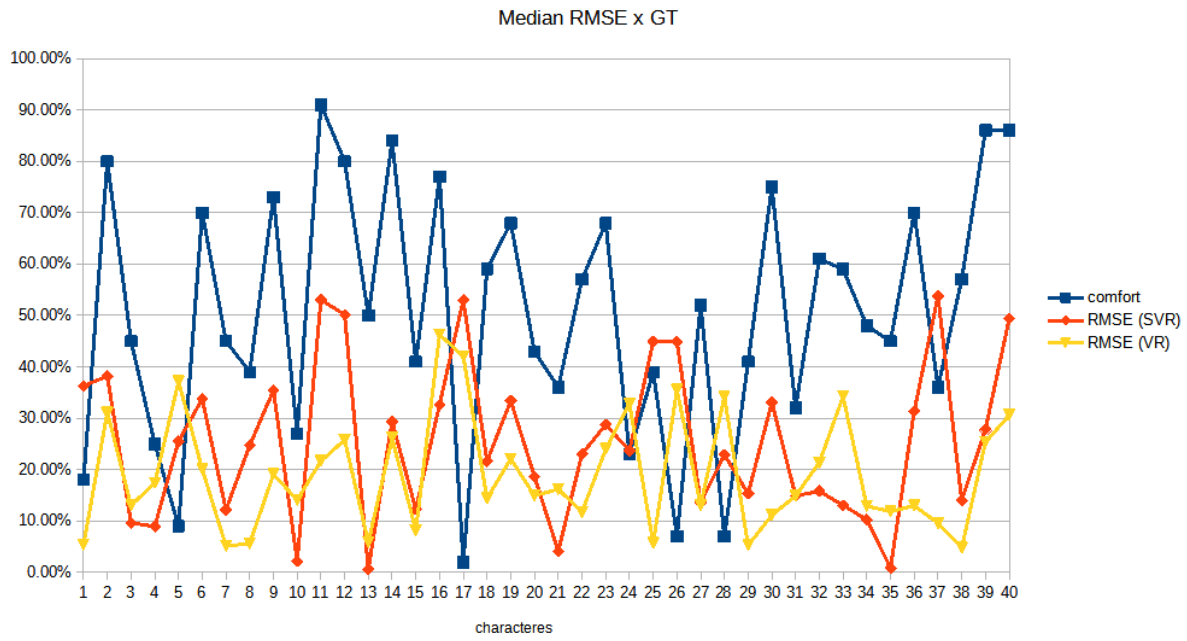


Figure 6.3: Plot of perceptual comfort and median RMSE metric for the SVR (red line) and VR (yellow line) models with perceived comfort values shown in the blue line (our ground truth, people's assessment of the characters). The *X* axis represents the ordering of the characters in the GT2 dataset. The *Y* axis represents people's perceived comfort of the characters and also the median error (RMSE) in the prediction made by the models. The lower the RMSE, the closer to perceptual comfort.

6.3 Interpretability of the best models using the LIME tool

Few machine learning interpretability models work with the ensemble technique as the voting classifier. We evaluate SHAP [MHJ20] (SHaplay Additive exPlanation), and DALEX [BB21] (Model Agnostic Language for Exploration and Explanation), both of which address the explainability of the global model, evaluating the entire dataset of the test. On the other hand, LIME [RSG16] (Local Interpretable Model Agnostic Explanations) also deals with local interpretability, so it was chosen for this study because we wanted to explain locally the parts of the face that generate

the prediction of comfort or discomfort. Therefore, we can also generate the most relevant features globally from the test dataset by collecting the feature importance for each instance, as this section will explain. LIME is conceived as a model that seeks to emulate the behavior of a pre-existing model, called a substitute or surrogate model, in a local context. This surrogate model is trained on a dataset derived from instances close to the one being interpreted by introducing small variations in characteristics or attributes, weighted according to the proximity to the original instance. In the scope of the present study, these variations are applied at the level of image pixels, covering procedures such as changing the brightness in certain regions, introducing noise, or applying subtle rotations. So, as expected, the VC and LIME models present an agreement rate of 100% when evaluating interpretability by LIME for all characters.

The interpretation of the results of the Lime applied to the analysis of the HU Moments requires the understanding of the characteristic vectors of these moments. Hu Moments are mathematical descriptors that encapsulate essential properties in an image, allowing recognition of robust and consistent patterns, regardless of transformations such as rotation, translation, and scale. We detail the purpose of each vector and the analogy with the human face in Section 3.1.

We used LIME (Section 3.2.1) as a way to interpret the results obtained from ensemble voting (methods VC and VR) and compare with GT (Table 4.2) to predict and also explained the models' predictions in terms of the oddest part of the face. The research question we want to answer using the explanation of LIME on our data is "How to explain a prediction of the comfort of a virtual face?" In the images representing the LIME explanation, the colors blue and orange represent how comfortable or uncomfortable the virtual face can be, whereas blue represents comfort prediction and orange represents discomfort.

We perform global (for the whole face) and local (parts of the face) analyses with all characters. The goal is to calculate and display the most important features of the emsemble voting models, both for the training data set and for the test data set, and save a graph comparing the importances for both data sets. For each estimator used in the emsemble voting technique, the model is trained (tuned) with the provided training data set.

6.3.1 Interpretability of some instances in the Voting Classifier Model

Due to space constraints, we focus on an in-depth discussion of only four selected characters in this section, where 3 are uncomfortable and 1 is comfortable, as shown in Table 6.15. However, it is important to note that we achieved 38.09% accuracy in predicting facial parts compared to the ground truth, highlighting the challenging nature of this topic for study.

Character	Frame	VC Model			LIME			GT
		Comfortable	Uncomfortable	Prediction	Comfortable	Uncomfortable	Prediction	
1	260	0	260	1	0	260	1	1
8	260	0	260	1	0	260	1	1
9	117	96	21	0	96	21	0	0
26	260	81	179	1	81	179	1	1

Table 6.15: Evaluation of the predicted classes for the 4 characters included in the test dataset of the voting classifier (VC) model together with LIME and the ground truth (GT). Prediction 0 is considered comfortable while class 1 is uncomfortable.

We performed global and local analyses on all characters. The global analysis uses the features of all characters for training, except the one being tested. The test dataset consisted of all frames of the character's video to predict comfortability. The local analysis uses a specific video frame to make the same prediction.

To evaluate the results, we consider the ground truth (GT) responses for the binary class (comfortable/discomfortable) in Table 6.9. Additionally, we compare the LIME explanation about the specific part of the face which generates more strangeness, with the answers from the participants (Table 4.2).

Figure 6.4 illustrates the interpretability of features in the VC model when evaluating the training dataset (having all characters but removing character 1) on the left, showing the importance of features versus classes. On the right, the test dataset consists of the frames from the video of character 1 and predominates relevant features of the "uncomfortable" class, with the forehead and eyes being the parts that stand out. The information on facial oddness confirms the results obtained in the participant survey for the full-face GT, but the forehead was not indicated as an odd region (see Table 4.2). Furthermore, Figure 6.5 shows the interpretability of the features using LIME for character 1 in a specific frame 1. On the left of Figure 6.5, LIME shows the indication of the probability of the two classes using the linear LIME model for the image on the left. In the center of the figure, the weights resulting from the LIME processing for the image on the right (blue and orange bars) for each part of the face presented. The features are listed on the Y-axis. The length of the bar on the X-axis indicates how much this feature contributed to the prediction. The direction of the bar (right or left) indicates whether the feature contributed positively or negatively to the prediction. The methods (VC and LIME) predict that the class is uncomfortable. The features that contributed positively (orange) to this result were the eyes when evaluating the direction and the degree of elongation (dde) of this part of the face, and the forehead which deals with its geometric shape. In contrast, the chin feature, when dealing with curvature variations (vac), was the most relevant to contribute negatively (blue) to the prediction made. The part of the face considered relevant (eyes) agrees with GT2 (Table 4.2). This information agrees with the user's research for GT2 of the entire face and for parts of the face when evaluating only a main part.

Therefore, of the 260 frames in the video of character 1, LIME and VC agree 100% with respect to the class of each frame; that is, 260 frames were classified as uncomfortable and 0 were comfortable, so all frames were considered uncomfortable.

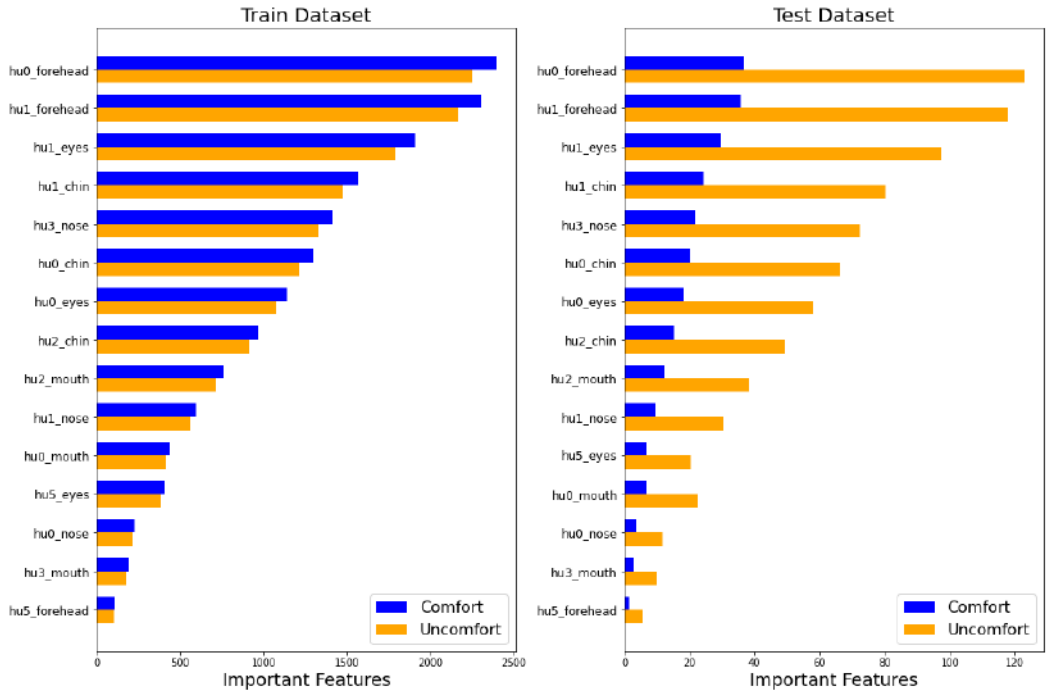


Figure 6.4: Global analysis of features relevance by class on training (left) and testing datasets of character 1 (right). The figure on the right, corresponding to the test data set of character 1, covering all frames, highlights a predominance of importance in the characteristics of the uncomfortable class, disagreeing in (eyes) with the GT Face in Table 4.2.

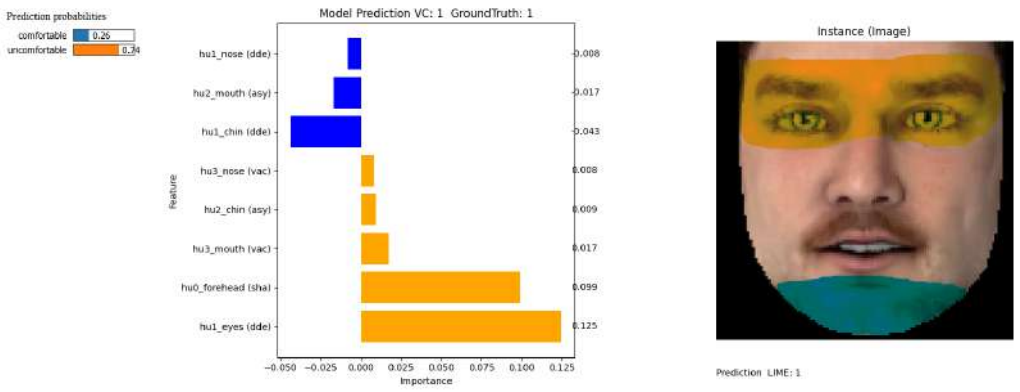


Figure 6.5: Interpretability by LIME for character 1 on the frame 1. On the left it shows the probability of the classes, in the middle the weights generated by the model for each relevant feature and on the right the evaluated face.

As with the analysis of character 1, Figure 6.6 illustrates the feature interpretability in the VC model by evaluating the training dataset (removing character 8 from the data) on the left, showing the importance of features versus classes. On the right, the test dataset consists of the video frames of character 8, and the fea-

tures relevant to the "uncomfortable" class predominate, with the forehead, eyes, nose, and mouth being the most prominent parts. This information confirms the participant's research for binary classifications of the face, however, the part of the face highlighted in the research is the chin. Additionally, Figure 6.7 shows the feature interpretability using LIME for character 8 at a specific frame 125. The method (VC and LIME) predicts that the class is uncomfortable. The features that contributed positively (orange) to this result were the eyes when evaluating the direction and degree of elongation (dde) of this part of the face, the forehead (dde) and the nose (asy) when evaluating the asymmetry. On the other hand, the chin feature, when treating the direction and degree of elongation (dde) and the mouth (asy), were the most relevant to contribute negatively (blue) to the prediction made. The most relevant part of the face (eyes) does not agree with GT2 (table 4.2) that indicated the chin. Therefore, of the 260 frames of the video of character 8, LIME and VC agree 100% on the class of each frame; that is, 260 frames were classified as uncomfortable and 0 as comfortable.

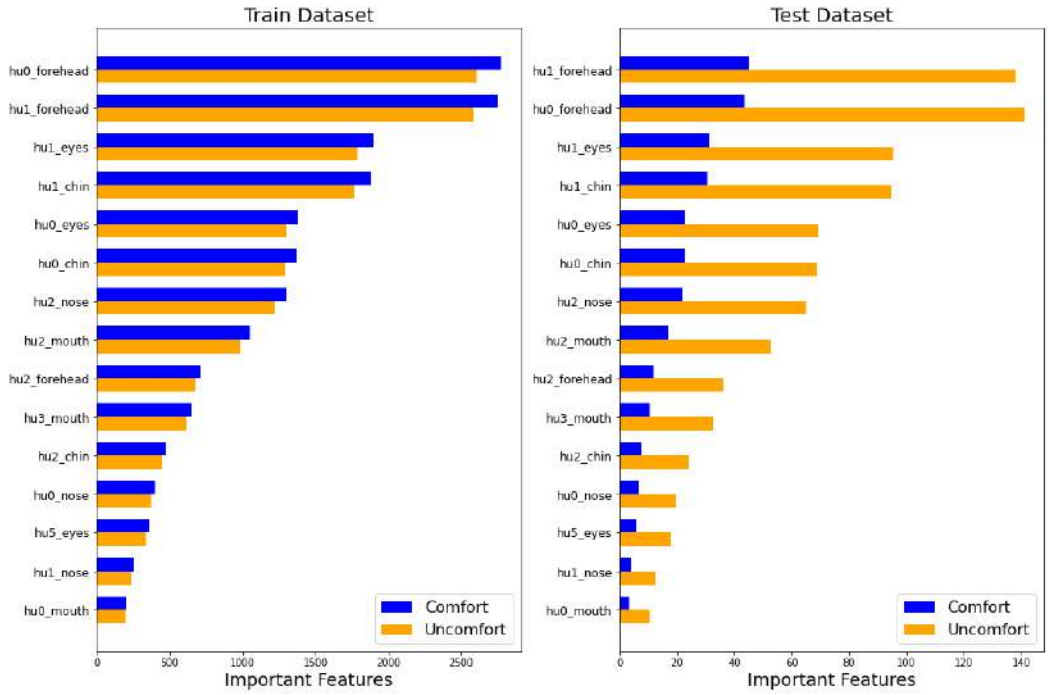


Figure 6.6: Global analysis of the relevance of features by class in the training (left) and testing (right) datasets for character 8. The figure on the right, corresponding to the testing dataset for character 8, covering all frames, highlights a predominance of importance in the features of the uncomfortable class, agreeing with the evaluations in Table 4.2 on GT Face. However, the part of the face selected by the participants is the chin.

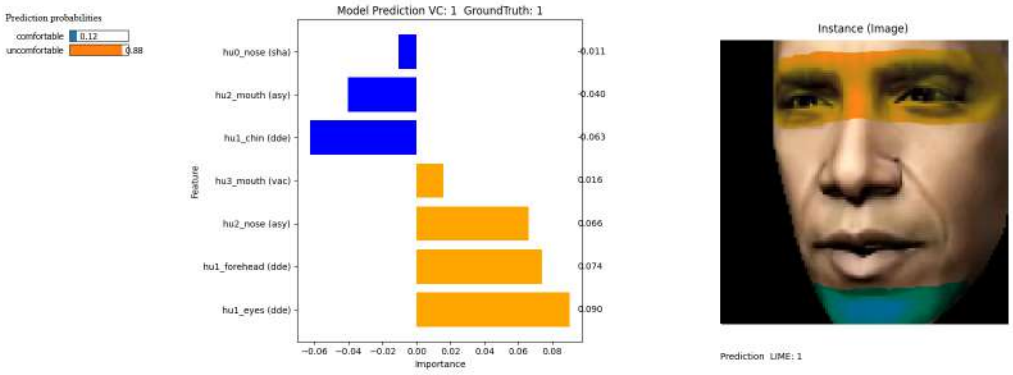


Figure 6.7: Interpretability by LIME for character 8 on frame 125. On the left it shows the probability of the classes, in the middle the weights generated by the model for each relevant feature and on the right the evaluated face.

Figure 6.8 illustrates the feature interpretability by evaluating the training dataset on the left and testing dataset consisting of character 9's video frames on the right. As can be seen, the features indicate the "comfortable" class. This information confirms the prediction of the comfort class with GT Face from Table 4.2. Additionally, Figure 6.9 shows the feature interpretability using LIME for character 9 at a specific frame 69. The method (VC and LIME) predicts the comfortable class for frame 69 of character 9. Therefore, out of the 117 frames in character 9's video, LIME and VC agree 100% on the class of each frame; that is, 21 frames were classified as uncomfortable and 96 as comfortable.

Finally, Figure 6.10 illustrates the feature interpretability of the VC model by evaluating the training and testing dataset for character 26. In this case, LIME predicts the "uncomfortable" class. This information agrees with GT Face in Table 6.9. Additionally, Figure 6.11 shows the feature interpretability using LIME for character 26 at a specific frame 34. The method (VC and LIME) predicts that the class is uncomfortable. This information confirms the prediction of the discomfort class with GT Face in Table 4.2. The features that contributed positively (orange) to this result were the eyes when evaluating the direction and degree of elongation (dde) of this part of the face, the shape of the forehead (sha), the nose (asy) when evaluating the asymmetry and finally the mouth (vac) through the curvature variance. In contrast, the chin feature, when treating the direction and degree of elongation (dde), the asymmetry of the mouth (asy) and the shape of the nose (sha), were the most relevant to contribute negatively (blue) to the prediction performed. Therefore, of the 260 frames in the video of character 26, LIME and VC agree 100% regarding the class of each frame; 179 frames were classified as uncomfortable and 81 were

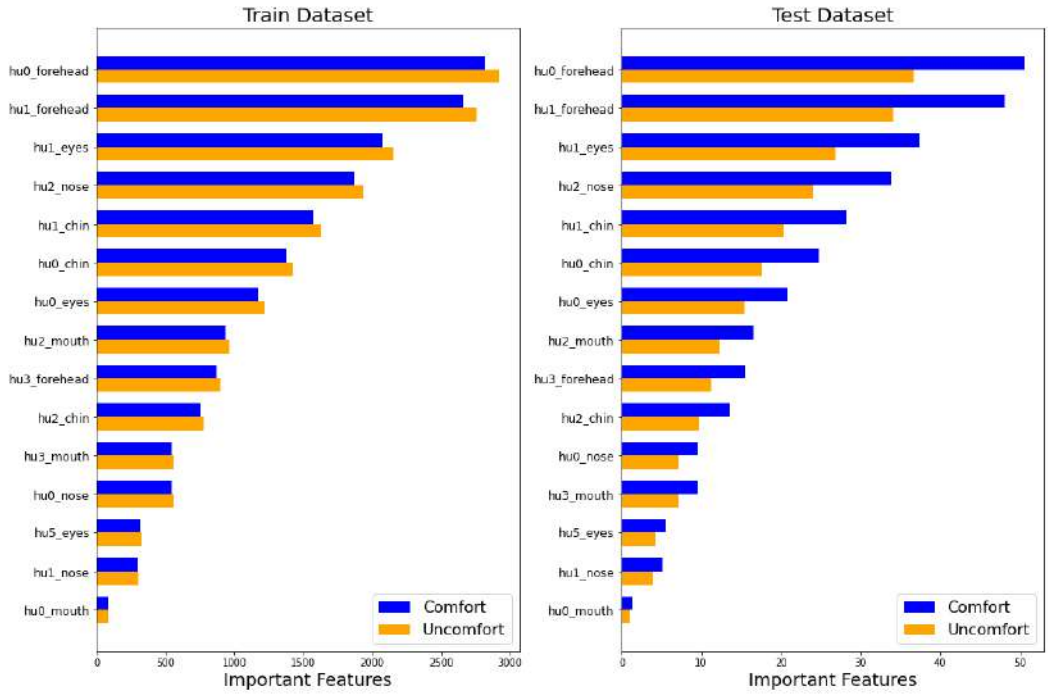


Figure 6.8: Global analysis of features relevance by class on training (left) and testing datasets of character 9 (right). The figure on the right, corresponding to the test data set of character 9, covering all frames, highlights a predominance of importance in the characteristics of the "comfortable" class. It agrees with GT Face in Table 6.9.

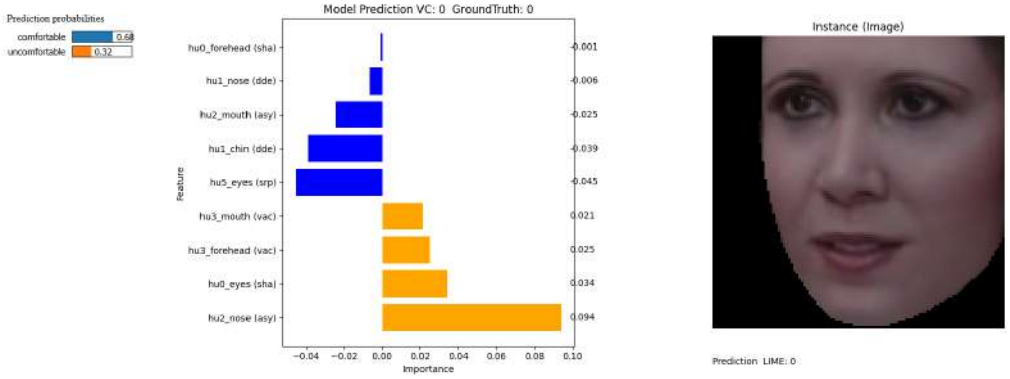


Figure 6.9: Interpretability by LIME for character 9 on frame 69. On the left it shows the probability of the classes, in the middle the weights generated by the model for each relevant feature and on the right the evaluated face.

classified as comfortable. LIME identified some parts of the face, such as the eyes, forehead, nose and mouth, that could potentially generate discomfort. This aspect should be investigated in future studies; specifically, even if a character is gener-

ally perceived as comfortable, are there specific facial features that can still cause discomfort? For the current study, we did not consider this possibility.

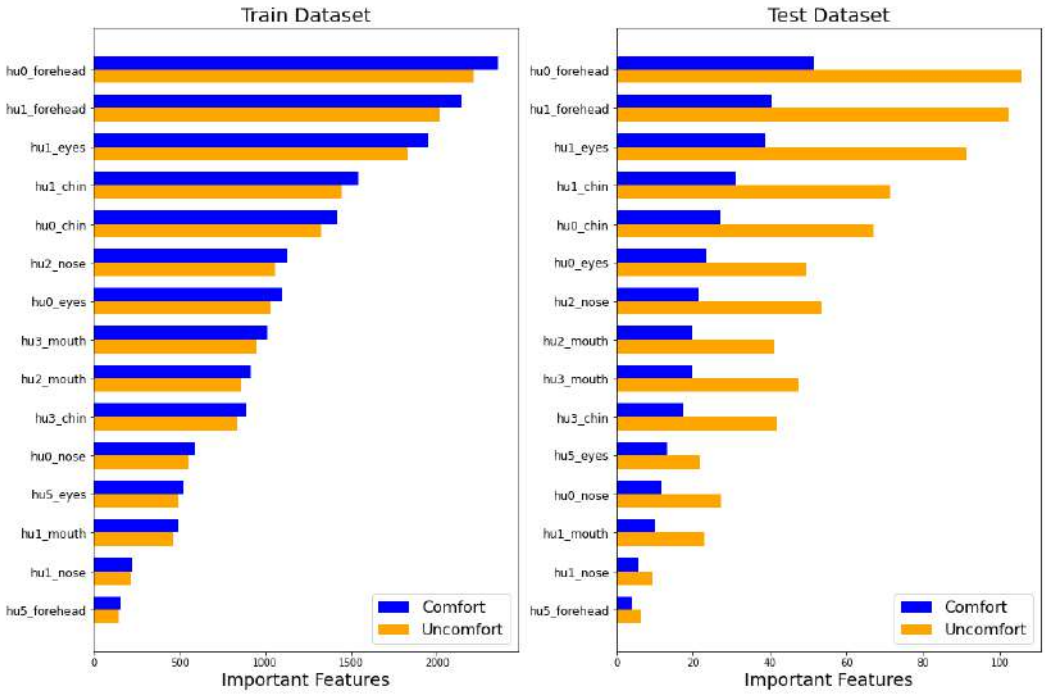


Figure 6.10: Global analysis of the relevance of features by class in the training (left) and testing (right) datasets of character 26. The figure on the right, corresponding to the testing dataset of character 26, covering all frames, highlights a predominance of importance in the features of the "uncomfortable" class, agreeing with the GT Face of Table 6.9. Although the subjects select the mouth as the most strange part, LIME indicate that the eyes, forehead, nose and mouth are the most strange.

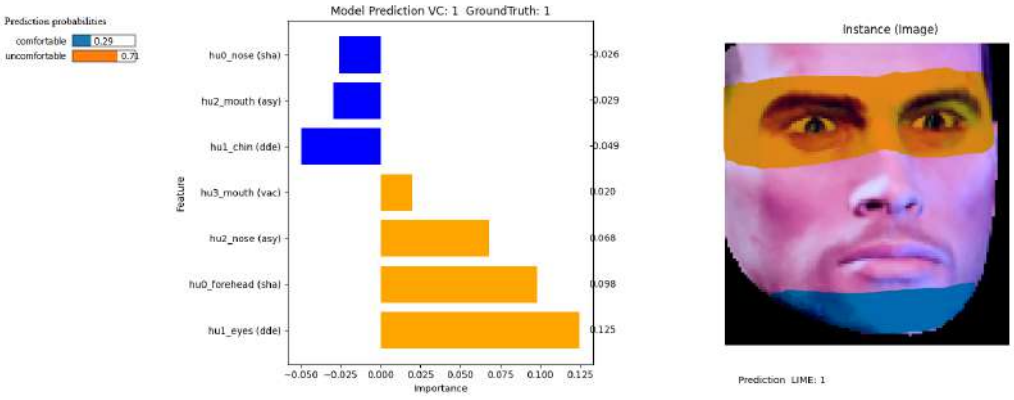


Figure 6.11: Interpretability by LIME for character 26 on frame 34. On the left it shows the probability of the classes, in the middle the weights generated by the model for each relevant feature and on the right the evaluated face.

While VC achieves a high accuracy of 91% (median F1-score) in predicting facial comfort class, LIME only achieves 23.80% accuracy compared to the ground truth when explaining the reasons for perceived discomfort when evaluating a single variable. When we check the first 3 main variables that explain the reason for the discomfort classification, the accuracy increases to 38.09%, as we can see in Table 6.16. This suggests that explaining discomfort is not a straightforward task and requires more research to improve the results. On the one hand, people may report specific parts of the face that generate discomfort, but these perceptions may be more subjective and not easily identifiable through computational methods.

Table 6.16 shows that when evaluating the first relevant variable (Top1) identified by LIME as the part of the face that causes strangeness, the accuracy with GT2 is 23.80%. When we evaluate only the second variable (Top2), we have 9.52% accuracy. If we evaluate only the third variable (Top3), there is an increase in accuracy of 19.04%. Therefore, when evaluating the first 3 most relevant variables identified by LIME, we obtain an accuracy of 38.09%.

character	ROI	GT	Agreements with ROI			
			Top1	Top2	Top3	Agreement
1	eyes	1	Agree	Disagree	Disagree	Agree
2	-	0	-	-	-	-
3	eyes	1	Agree	Disagree	Disagree	Agree
4	eyes	1	Agree	Disagree	Agree	Agree
5	mouth	1	Disagree	Disagree	Disagree	Disagree
6	-	0	-	-	-	-
7	mouth	1	Disagree	Disagree	Disagree	Disagree
8	chin	1	Disagree	Disagree	Disagree	Disagree
9	-	0	-	-	-	-
10	forehead	1	Disagree	Agree	Agree	Agree
11	-	0	-	-	-	-
12	-	0	-	-	-	-
13	eyes	1	Disagree	Disagree	Disagree	Disagree
14	-	0	-	-	-	-
15	mouth	1	Disagree	Disagree	Disagree	Disagree
16	-	0	-	-	-	-
17	mouth	1	Disagree	Disagree	Disagree	Disagree
18	-	0	-	-	-	-

19	-	0	-	-	-	-
20	eyes	1	Agree	Disagree	Disagree	Agree
21	eyes	1	Disagree	Disagree	Agree	Agree
22	-	0	-	-	-	-
23	-	0	-	-	-	-
24	eyes	1	Disagree	Disagree	Disagree	Disagree
25	eyes	1	Disagree	Disagree	Disagree	Disagree
26	mouth	1	Disagree	Disagree	Disagree	Disagree
27	-	0	-	-	-	-
28	mouth	1	Disagree	Disagree	Disagree	Disagree
29	eyes	1	Disagree	Disagree	Agree	Agree
30	-	0	-	-	-	-
31	mouth	1	Disagree	Disagree	Disagree	Disagree
32	-	0	-	-	-	-
33	-	0	-	-	-	-
34	mouth	1	Disagree	Disagree	Disagree	Disagree
35	forehead	1	Agree	Agree	Disagree	Agree
36	-	0	-	-	-	-
37	mouth	1	Disagree	Disagree	Disagree	Disagree
38	-	0	-	-	-	-
39	-	0	-	-	-	-
40	-	0	-	-	-	-

Table 6.16: Evaluation of the first 3 features (Top1, Top2, Top3) relevant to LIME as causing strangeness. The evaluation is performed for each feature. The Agreement column is the evaluation made considering the 3 features. If one of them agrees with the ROI column - which is the ground truth of the part of the face considered strangest according to Table 4.2, then the result is Agree, otherwise it is Disagree. The "-" in the Agreement column indicates that according to the GT column (ground truth of the entire face) they do not generate strangeness. If the ROI column does not contain data from parts of the face, it means that this character is not considered strange according to GT. Therefore, there will be no evaluation in the Top1, Top2, Top3 columns and no result in the Agreement column.

We can observe in Table 6.16, column ROI, through the LIME interpretations, that the eye area seems to be important for the identification of strangeness in the face. This contributes to the result of GT2 (table 4.2), since of the 21 characters considered strange, 9 of them had the eye region identified as causing discomfort

(42.85%). In addition to confirming many studies in the literature that indicate the eyes as an area that can cause strangeness, such as the study by Schein et al. [SG15] who, through a series of experiments, showed that eyes that appear empty or devoid of life increase the sensation of strangeness, suggesting that the perception of the mind is strongly linked to emotional responses to UV. The study by Geller et al. [Gel08] points out that films such as Beowulf (Grendel) and Lord of the Rings (Gollum) scare people and show that a good way to avoid UV would be to change the proportions and structure of a character. This is a justification for Gollum's success, as he has large eyes and a non-human face shape.

6.3.2 Interpretability of some instances in the Voting Regression Model

We use LIME [RSG16] as a way to interpret the obtained VR results and compare it with GT2 (Table 4.2), to predict the comfort CCS and also explain such prediction in terms of the most uncomfortable part of the face. Due to space constraints, we focus on an in-depth discussion of only four selected characters in this section, where 3 are uncomfortable and 1 is comfortable, as shown in Table 6.17. However, it is important to note that we achieved 61.90% accuracy in predicting facial parts compared to the ground truth, highlighting the challenging nature of this topic for study.

Character	Frame	VR Prediction	LIME Prediction	GT
1	260	27.36%	30.21%	18%
8	260	39.11%	38.43%	39%
9	117	53.48%	53.44%	73%
26	260	40.32%	38.90%	7%

Table 6.17: Evaluation of predicted comfort for the 4 characters included in the test dataset of the voting regression (VR) model along with LIME and ground truth (GT). The Prediction column reports the median estimated comfort by the model over the character's video frames and the GT column indicates the perceived comfort by people.

We performed global (the whole face) and local (face parts) analyses on all characters. The global analysis uses the features of all characters for training, except the one being tested. The test dataset consisted of all frames of the character's video to predict comfortability. The local analysis uses a specific video frame to

make the same prediction. To evaluate the results, we considered the ground truth (GT2) responses for CCS in Table 4.2. In addition, we compared the LIME explanation about the specific part of the face that generates the most awkwardness, with the participants' responses (Table 4.2).

Figure 6.12 illustrates the feature interpretability in the VR model when evaluating the training dataset (having all characters but removing character 1) on the left, showing the importance of the dataset features. On the right, the test dataset consists of the frames from the video of character 1 and features relevant to the set of frames predominate, with the mouth and chin being the most prominent parts. Additionally, Figure 6.13 shows the feature interpretability using LIME for character 1 in a specific frame 1. On the left of Figure 6.13, LIME shows the comfort prediction indication using the linear LIME model for the image on the right. In the center of the figure, the resulting values of the LIME processing for the image on the right (blue and orange bars) for each part of the face are presented. The features that contributed positively (orange) to this result were mainly the asymmetry of the mouth (asy) and the nose (dde) when evaluating the direction and degree of elongation. In contrast, the variations in chin curvature (var), the direction and degree of elongation of the forehead (dde) and the shape of the eyes contributed negatively (blue) to this result. The methods (VR and LIME) show that the prediction is discomfort, and LIME predicts the areas of the face that contributed positively and negatively to this prediction. This information agrees with the user's research for GT of the whole face, but disagrees for parts of the face.

As with the analysis of character 1, Figure 6.14 illustrates the feature interpretability by evaluating the training dataset (removing character 8 from the data) on the left, showing the importance of the features. On the right, the test dataset consists of the video frames of character 8, and the relevant features that predominate, with the forehead and chin being the most prominent parts to cause discomfort. This information confirms the participant's research for face prediction, however, the part of the face highlighted in the research is the chin. Additionally, Figure 6.15 shows the feature interpretability using LIME for character 8 at a specific frame 125. The features that contributed positively (orange) to this result were mainly the asymmetry of the mouth (asy), the shape of the eyes (sha) and the shape of the nose (dde) when evaluating the direction and degree of elongation. In contrast, the shape of the chin (var) and the direction and degree of elongation of the forehead (dde) contributed negatively (blue) to this result. The method (VR and LIME) predicts that the

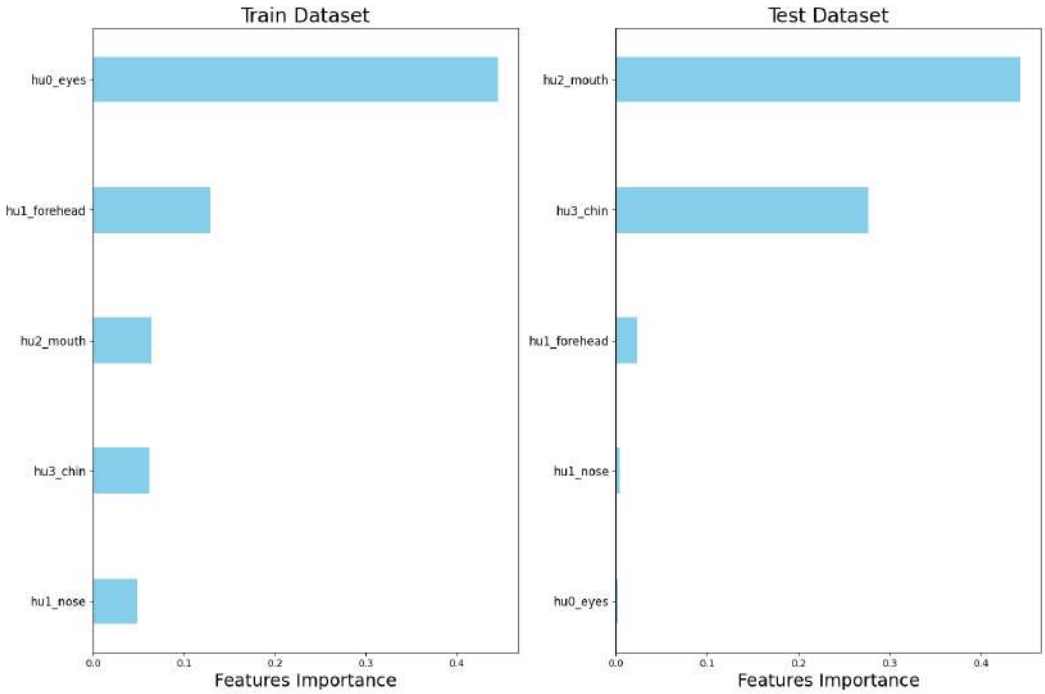


Figure 6.12: Global analysis of features relevance on training (left) and testing datasets of character 1 (right). The figure on the right, corresponding to the test data set of character 1, covering all frames, highlights a predominance of importance in the characteristics on test dataset, disagreeing in (eyes) with the GT Face in Table 4.2.

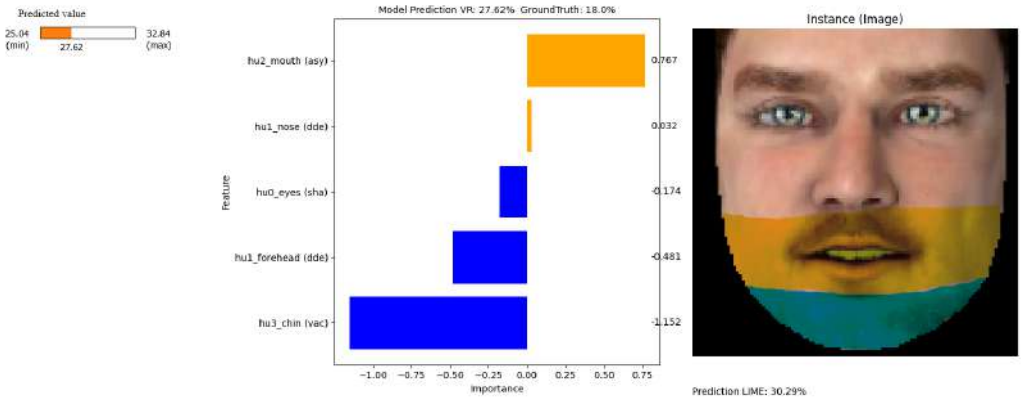


Figure 6.13: Interpretability by LIME for character 1 on the frame 1. On the left it shows the comfort prediction (CCS), in the middle the weights generated by the model for each relevant feature and on the right the evaluated face.

face is uncomfortable, and LIME points out that the relevant areas of discomfort are the mouth, eyes and nose.

Figure 6.16 illustrates the feature interpretability in the VR model by evaluating the training dataset on the left and the testing dataset consisting of video

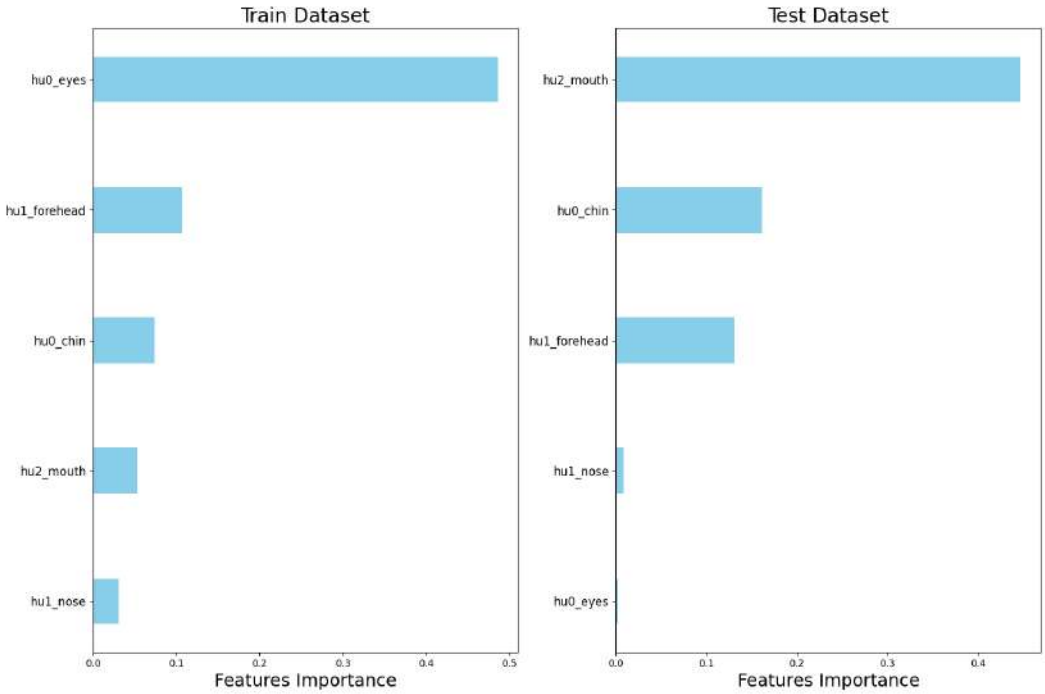


Figure 6.14: Global analysis of the relevance of the features in the training (left) and testing (right) datasets for character 8. The figure on the right, corresponding to the testing dataset for character 8, covering all frames, highlights a predominance of importance in the features of the mouth and chin, agreeing with the evaluations of Table 4.2 in GT2 Face only in relation to the prominence of the chin, since the part of the face selected by the participants is the chin as being uncomfortable.

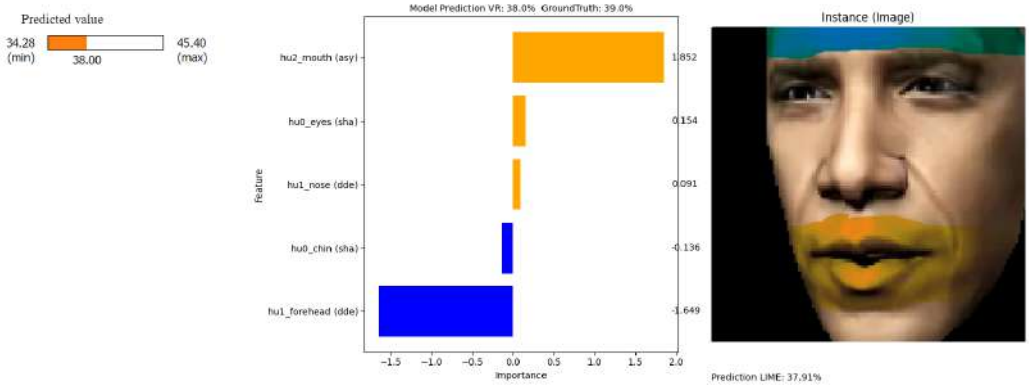


Figure 6.15: Interpretability by LIME for character 8 on frame 125. On the left it shows the probability of the classes, in the middle the weights generated by the model for each relevant feature and on the right the evaluated face.

frames of character 9 on the right. Additionally, Figure 6.17 shows the feature interpretability using LIME for character 9 at a specific frame 69. The features that contributed positively (orange) to this result were mainly the shape of the forehead

(sha), then the shape of the eyes (sha) and the nose (dde) when evaluating the direction and degree of elongation. On the other hand, the asymmetry of the mouth (asy) and the shape of the chin (sha) were the most relevant to contribute negatively (blue) to the prediction made. The method (VR and LIME) predicts a comfortable prediction for frame 69 of character 9. This information confirms the comfort prediction (CCS) with GT2 Face from Table 4.2.

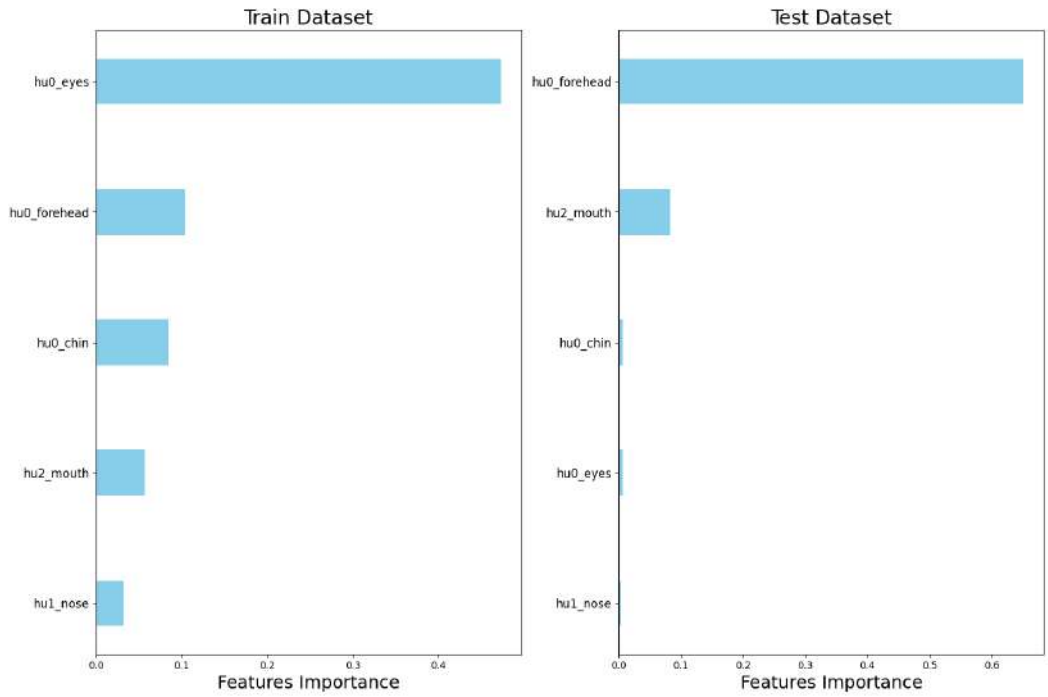


Figure 6.16: Global analysis of feature relevance in training (left) and testing (right) datasets for character 9. The figure on the right, corresponding to the testing dataset for character 9, covering all frames.

Finally, Figure 6.18 illustrates the feature interpretability by evaluating the training and testing dataset of character 26. Additionally, Figure 6.19 shows the feature interpretability using LIME for character 26 at a specific frame 34. The method (VR and LIME) predicts that the instance is uncomfortable. This information confirms the discomfort prediction with GT Face from Table 4.2. The nose and mouth are the most prominent parts, agreeing with GT2 in that one of the face parts highlighted by LIME is considered relevant, such as the mouth. LIME identified some face parts, such as the eyes, forehead, and chin, that could potentially generate comfort. This aspect should be investigated in future studies; specifically, even if a character is generally perceived as comfortable, are there specific facial features that can still cause discomfort? For the current study, we did not consider this possibility.

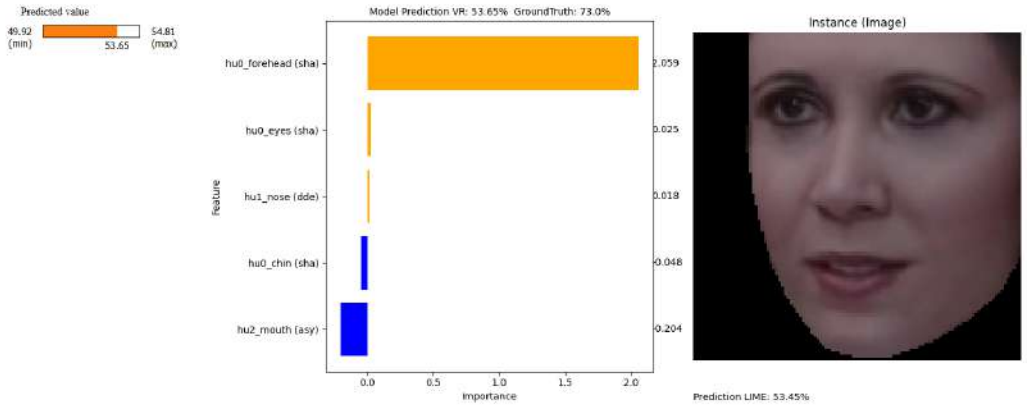


Figure 6.17: Interpretability by LIME for character 9 in table 69. On the left shows the comfort prediction of the face (CCS), in the middle the weights generated by the model for each relevant feature and on the right the evaluated face.

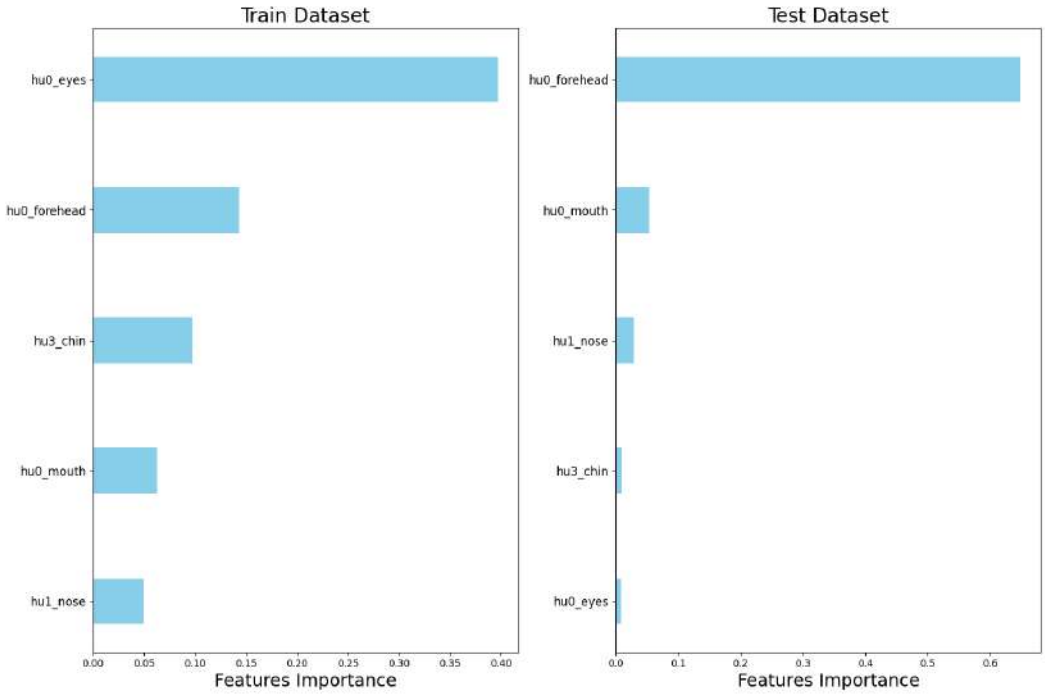


Figure 6.18: Global analysis of the relevance of features in the training (left) and testing (right) datasets of character 26. The figure on the right, corresponding to the testing dataset of character 26, covering all frames. It highlights a predominance of importance in the features of the forehead, nose and mouth. The part of the face selected by the participants is the mouth, as shown in Table 6.9.

While our method achieves a low median RMSE metric of 15.55% in predicting facial comfort, which is a positive factor, LIME only achieves 19.04% accuracy compared to the ground truth when explaining the reasons for perceived uncomfortable when evaluating a single variable. When we check the top 3 variables that

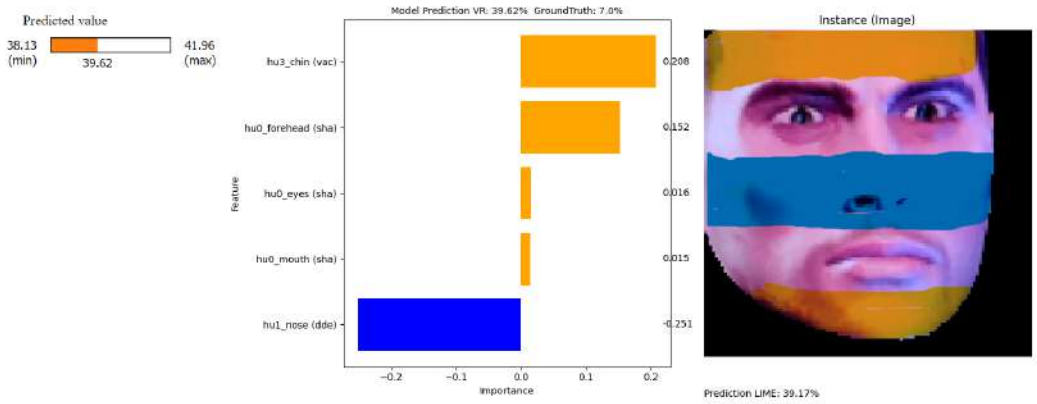


Figure 6.19: Interpretability by LIME for character 26 in frame 34. On the left shows the prediction, in the middle the weights generated by the model for each relevant feature and on the right the evaluated face. The prediction agrees with GT2 from Table 6.9. The characteristics that contributed positively (orange) to this result were mainly the variations in chin curvature (var), the shape of the forehead, eyes and mouth. In contrast, the nose (dde) when evaluating the direction and degree of elongation contributed negatively (blue) to this result. However, the part of the face selected by the participants is the mouth.

explain the reason for the embarrassment predição de conforto (CCS), the accuracy increases to 61.90%, as we can see in Table 6.18. This suggests that explaining uncomfortable is not a straightforward task and requires more research to improve the results. On the one hand, people may report specific parts of the face that generate awkwardness, but these perceptions may be more subjective and not easily identifiable through computational methods.

Table 6.18 shows that when evaluating the first relevant variable (Top1) identified by LIME as the part of the face that causes strangeness, the accuracy with GT2 is 19.04%. When we evaluate only the second variable (Top2), we have 23.80% accuracy. If we evaluate only the third variable (Top3), there is an increase in accuracy of 28.57%. Therefore, when evaluating the first 3 most relevant variables identified by LIME, we obtain an accuracy of 61.90%.

character	ROI	GT	Agreements with ROI			
			Top1	Top2	Top3	Agreement
1	eyes	1	Disagree	Disagree	Agree	Agree
2	-	0	-	-	-	-
3	eyes	1	Disagree	Agree	Disagree	Agree
4	eyes	1	Disagree	Disagree	Disagree	Disagree
5	mouth	1	Disagree	Agree	Disagree	Agree

6	-	0	-	-	-	-
7	mouth	1	Disagree	Disagree	Disagree	Disagree
8	chin	1	Disagree	Disagree	Disagree	Disagree
9	-	0	-	-	-	-
10	forehead	1	Disagree	Agree	Disagree	Agree
11	-	0	-	-	-	-
12	-	0	-	-	-	-
13	eyes	1	Agree	Disagree	Disagree	Agree
14	-	0	-	-	-	-
15	mouth	1	Disagree	Disagree	Disagree	Disagree
16	-	0	-	-	-	-
17	mouth	1	Disagree	Disagree	Agree	Agree
18	-	0	-	-	-	-
19	-	0	-	-	-	-
20	eyes	1	Agree	Disagree	Disagree	Agree
21	eyes	1	Disagree	Agree	Agree	Agree
22	-	0	-	-	-	-
23	-	0	-	-	-	-
24	eyes	1	Disagree	Agree	Agree	Agree
25	eyes	1	Agree	Disagree	Disagree	Agree
26	mouth	1	Disagree	Disagree	Agree	Agree
27	-	0	-	-	-	-
28	mouth	1	Disagree	Disagree	Disagree	Disagree
29	eyes	1	Disagree	Disagree	Disagree	Disagree
30	-	0	-	-	-	-
31	mouth	1	Disagree	Disagree	Disagree	Disagree
32	-	0	-	-	-	-
33	-	0	-	-	-	-
34	mouth	1	Disagree	Disagree	Agree	Agree
35	forehead	1	Agree	Disagree	Disagree	Agree
36	-	0	-	-	-	-
37	mouth	1	Disagree	Disagree	Disagree	Disagree
38	-	0	-	-	-	-
39	-	0	-	-	-	-
40	-	0	-	-	-	-

Table 6.18: Evaluation of the first 3 features, which we call (Top1, Top2, Top3), relevant to LIME as causing strangeness. The evaluation is performed for each feature exclusively. The Agreement column is the evaluation made considering the 3 features together. If one of them agrees with the ROI column, which is the ground truth (GT2) of the part of the face considered strangest, then the result is Agree, otherwise it is Disagree. The "-" in the Agreement column indicates that according to the GT column (ground truth of the entire face) they do not generate strangeness.

6.4 Comparing the best models with the literature

The study by Mustafa et al. [MGT+17] investigates neural responses to computer-generated faces in a cognitive neuroscience study. They recorded the brain activity of 80 participants using electroencephalography (EEG) while they watched videos of realistic and virtual humans. Based on this information, they trained a Support Vector Machine (SVM) to measure the probability of an uncanny response to any computer-generated character from EEG data, allowing them to rank animated characters based on their level of uncannyness. Mustafa et al. [MGT+17] used recordings of, among others, state-of-the-art computer-generated humans Digital Emily and Digital Ira. They also included highly realistic characters from interactive drama video games, such as Kara from Detroit: Become Human, Ernst from Squadron, and HeadTech from Janimation. These characters were found to be highly human-like. The only work to compare because it has an objective metric of the theory and does not use image features.

The assessment in the work of Mustafa et al. [MGT+17] occurs through the N400 Component of ERPs that serves to find an odd neural response in CG characters.

The amplitude of the N400 brain response (negative peak 400ms after the stimulus) is a well-established measure and associated with expectation mismatch, that is, the person sees something that does not correspond to their expectation of how it should be. According to the N400 component, the oddball response is stronger when the CG character appears highly realistic.

This supports the predictive coding hypothesis in which the uncanny valley is related to expectation violations in neural computation when the brain encounters highly realistic characters.

Figure 6.20 shows the mismatch order (Component N400) in relation to the characters when people viewed them while the ECG exam was being performed.

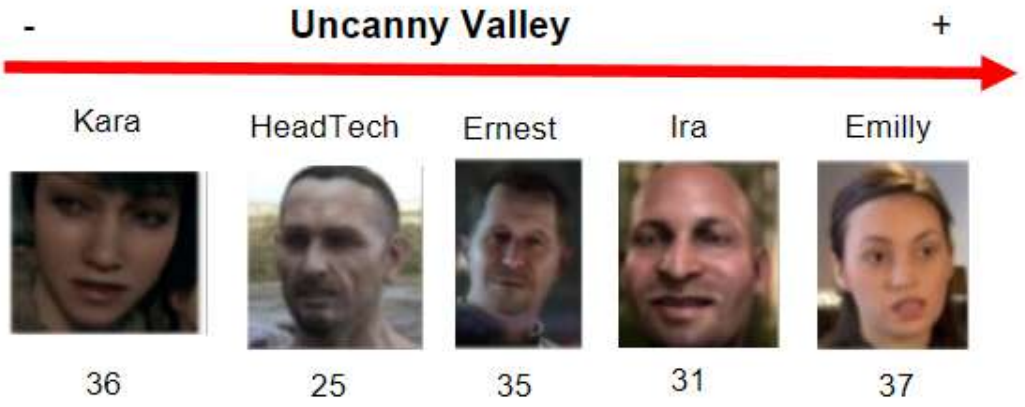


Figure 6.20: Classification of the degree of strangeness according to the study by Mustafa et al. [MGT+17] when evaluating the N400 component through the ECG performed on humans.

Table 6.19 shows the ranking column so that we can compare our results with the work of Mustafa et al. [MGT+17]. We can observe from the results of the VC Models that the character Ernest (35) is the only one that presented a different result from the GT Class for the GT2 dataset. For the rest of the characters, we observed that the majority of frames in the classes agreed with the GT Class of the GT2 dataset. When evaluating the VR Model, we also noticed that the only character that did not agree with the GT for comfort was the character Ernest (35). However, the computed comfort score (*CCS*) metric is very close to the threshold used to identify the character as comfortable or not. Therefore, there is a detection by the VR ensemble model of a possible incompatibility identified in human perceptive comfort.

Figure 6.21 mostra o ranking identificado na Table 6.19. The order of strangeness makes sense with the order of perceived comfort according to Figure 6.20. We observed the character Ira, who is closest to the character Emilly in terms of comfort, which agrees with the analysis of Mustafa et al. [MGT+17] when it indicates that although the research indicated a moderate realism of the character Ira, the N400 component identified a high level of incompatibility when people saw this character. The characters HeadTech and Ernest are considered practically equal in terms of strangeness in the study of Mustafa et al. [MGT+17], and in our study of the computed comfort score (*CCS*) metric also agrees with both comparative research and with the questionnaire 4.2 of our GT2 dataset. The character Kara

Ranking	Character	Model VC		Model VR	Dataset GT2	
		Class 0	Class 1	CCS	GT Class	GT Comfort
1º	36	103	7	57.54	0	70.00
2º	37	18	241	39.91	1	36.00
3º	35	188	71	52.07	1	45.00
4º	25	18	241	43.18	1	39.00
5º	31	21	238	47.02	1	7.00

Table 6.19: Evaluation of CG characters, presenting a ranking from the character considered least strange to highly strange by the VC and VR models. We compared the results with the GT2 dataset which presents the Ground Truth by class and comfort, based on the questionnaire (Section∼\ref{sec:attachQuestionnaire} conducted with 40 people.

has the lowest degree of realism of the 5 and causes the least strangeness, and our computed comfort score (*CCS*) metric also considers it this way.

Figure 6.21

Apparently our methodology agrees with the study by Mustafa et al. [MGT+17]. This indicates that both perceptual categorization and computed comfort score (*CCS*) on a CG character with the prediction of our models is possible.

6.5 Chapter Considerations

This chapter presents and discusses the experimental results achieved so far. These results include:

1. investigating whether we can estimate the uncanniness caused by animated characters in humans using visual features;

2. computer vision algorithms to extract relevant features;
3. proposing the possibility of interpretability of the results predicted by the models;
4. compare the results of the best models with the literature.

Regarding these issues, we consider that our models are promising in the sense that they seem to show that human perception can be represented using image features of CG characters, especially in relation to regression models.

Table 6.20 shows a summary of the models used in this research and the result achieved through accuracy and the F1-Score and RMSE metrics. We can observe that the voting regressor model, using the extraction of Hu Moments features from parts of the face, indicates an error rate of 15.55%.

A final word about our results: in evaluating the five machine learning models (SVM, VC, CNN, SVR, and VR), the best-performing model for binary classification was the Voting Classifier using Hu Moments, which demonstrated superior accuracy compared to other models. Similarly, for regression tasks, the Voting Regressor with Hu Moments yielded the most accurate results, highlighting the effectiveness of this combination across both classification and regression contexts in the dataset tested. However, the voting techniques can present at least two drawbacks to be analyzed. Firstly, the computational time of voting techniques tends to be higher in comparison with other techniques, and secondly, few interpretability models work with voting techniques. That is why we used LIME and could not used or compare with other methods.

Model	Technique	Binary	Features Extract	FACE	Dataset (characters)	Accuracy (%)	Metrics	
							F1-Score (%)	RMSE (|%)
First	SVM	yes	Hu Moments HOG	Entire	19	81.00	80.00	-
					40	50.00	64,03	-
Second	**VC**	yes	**Hu Moments**	**Entire**	40	**67.5**	**91.11**	-
				Parts	40	**38.09**	**91.11**	-
Third	CNN	yes	Image	Entire	40	46.63	56.69	
				Parts	40	45.96	49.79	
Fourth	SVR	no	Entropy	Entire	19	80.00	-	23.59
					40	62.5	-	24.19
Fith	**VR**	no	**Hu Moments**	**Entire**	40	**80.00**	-	**15.55**
				Parts	40	**61.90**	-	**15.55**

Table 6.20: Result of the proposed models, indicating the machine learning and deep learning algorithm used to generate the model. The binary column indicates whether it is a binary classification model or regression. The face column indicates whether the model deals with the entire face or parts of the face. The dataset indicates the number of characters. The accuracy column indicates the model's prediction and the metric column informs the value when using F1-Score or RMSE.

7. FINAL REMARKS

This work presented five models for comfort estimation related to CG characters' faces. The two binary models evaluate whether or not there is comfort on the character's face by extracting features with Hu Moments and HOG. The difference between the models is that the first (trained with the SVM algorithm) evaluates the entire face, while the second (using the voting classifier technique) tries to evaluate the parts of the face that would cause more discomfort. We also proposed the fine-tuning of a CNN to compare with the binary methods. Fine-tuning is a technique that involves training a pre-trained model (ImageNet) on a specific dataset. This model takes advantage of the prior knowledge of the pre-trained model and adjusts its final layers to adapt to a new dataset. VGG16 [SZ15a], Resnet50 [SZ15b] and MobileNet [HZC+17] were the neural networks chosen for this task. We performed fine-tuning by binary classification of the entire face and also with parts of the face.

The regression models were used to compute our proposed metric, Computed comfort score (*CCS*), based on image characteristics, such as spatial and spectral entropy, used in the first model (trained with the SVR algorithm). For the second model (using the voting regressor technique), experiments were carried out with other feature extraction algorithms, such as: AUs, GLCM, Golden Ratio, Hu Moments in addition to entropy. In the case of the Hu Moments feature, we were able to extract features of the entire face and also of parts of the face. The results obtained for the 5 models seem to confirm that the models presented indicate the estimated comfort of some studies in the literature.

Regarding the interpretability of the models, we were able to identify regions that could potentially cause discomfort, and LIME was a relevant tool in terms of explanations for the findings. The study about the strange parts of the faces is relevant because we can suggest, according to the extracted features, a recommendation to change the part of the face evaluated as discomfort. This last part is still in the experimental phase, but we have already drawn a parallel between the functions of the Hu Moments vector and the human face to address this issue.

Our results were compared with literature and with our ground truth of perceptual data and datasets. Evaluation and scores are reported. Additionally, and specifically compared with the work of Mustafa et al.[MGT+17] who examine how the brain responds to computer-generated faces in a cognitive neuroscience context. Our approach appears to be in line with the methods used by Mustafa et

al.[MGT+17], suggesting that it is possible to both perceptually categorize and calculate the comfortability score (*CCS*) for a CG character based on the predictions of our models.

The main findings are the feature extraction algorithms used in this research. The result of the RMSE metric with a low error rate signaled a promising use. The AUs (Action Units) that show that exaggerated facial expressions, such as wide smiles, can increase the perceived emotional charge when exceeding a certain limit, generating discomfort. This suggests that this will be the next step to recommend adjustments in the intensity of expressions and avoid perceptual discomfort. When extracting spatial and spectral features from faces generated by CG, we found a low residual error rate (13.88%), suggesting that distortions in entropy can influence the perception of comfort. For future studies, we want to identify which types of distortions can influence perceived human comfort to better use comfort computed by our *CCS* metric. The GLCM method that can detect textures perceived as anomalous seems to indicate that the perception of comfort can be affected by anomalies or degradations in the texture of CG faces. This observation suggests that an in-depth study of the textures of CG faces may contribute to the assessment of perceived comfort by humans. Facial symmetry and neoclassical proportions strongly influence the perception of beauty, and have been widely studied in the area of Perceptual Image Quality. More symmetrical faces tend to be perceived as more attractive, and the low residual error in the VR model, worked on in this thesis, suggests that proportion characteristics may also impact computed comfort. The next step is to investigate whether the Golden Proportions of human faces maintain the same pattern in CG faces to improve the assessment of discomfort. Finally, the Hu Moments algorithm may help recommend geometric adjustments, such as eyebrow curvature or mouth shape, to mitigate discomfort caused by visual inconsistencies. The next phase of this study will be to validate whether the analogy between the purpose of Hu Moments and the CG face would actually alleviate the perception of strangeness.

This work has some limitations. First, the dataset we initially worked with was GT1 (section 4.1). Through literature research, we increased the dataset, generating GT2 (section 4.2. Although we went from 19 characters to 40, it is still a small number of characters. We certainly need to increase the dataset so that we can improve our techniques or use new ones. This is an issue we want to work on in the future, especially when we look at the results of the models using the voting classifier and regressor techniques, which are the two best-evaluated models.

The other limitation of this study is the LIME, interpretability model, in which we need to evaluate with greater precision the parts of the face that cause strangeness. We still have limitations in indicating the part of the face that causes strangeness, as indicated by human perception. Our suggested models, using the ensemble technique, indicated an accuracy of 38.09% and 61.90% for binary classification and regression respectively.

Therefore, as a future work, we want to integrate different image and expression analysis approaches to make recommendations for adjusting CG characters and reducing the discomfort caused by the Uncanny Valley theory. We believe that it is possible to combine the use of AUs, entropy, GLCM, Hu Moments and the Golden Ratio to potentially indicate recommendations that improve the aesthetic and emotional perception of virtual characters. Investigating these parameters as a time series, accounting for deformations over time in animations, is a direction that we consider promising for future research. In addition, we continue working on detecting facial parts as part of recommendations to be indicated, because the accuracy is still low.

Therefore, I conclude this work, which aims to investigate something that has not yet been done in the literature, which is to investigate how to transform a human perceptual sensation into objective data. This was the challenge of my doctoral thesis. That is why so many methods were tested in order to verify whether any of them could respond to this human sensation.

It is important to conclude this work by saying that ensemble techniques generated better results because the other methods only applied a single algorithm, despite several having been tested. The question may possibly arise: why is this interesting? And the answer is because, due to this study, ensemble techniques seem to better represent human perception. Sometimes a person perceives better by looking at an eye, another perceives by seeing the color of the hair, many others feel empathy for the character - an example, in this case, is Princess Leah, who is considered nice. There are still people who may like the color of the character, and thus this complexity of human sensations is constituted.

Finally, we know that our perception is much more sophisticated and complex than just saying that the eye is asymmetrical. This can be better represented computationally through ensemble methods, which could be verified by the results of the quantitative data. It is also important to note that while we were able to obtain an acceptable accuracy of 80% for the entire face, this prediction still needs to be improved for parts of the face, which is a problem for us to continue working on.

References

[ADM21] Araujo, V.; Dalmoro, B.; Musse, S. R. “Analysis of charisma, comfort and realism in cg characters from a gender perspective”, *The Visual Computer*, vol. 37–9, 2021, pp. 2685–2698. (Citado nas páginas 69, 70, and 71.)

[AMDM21] Araujo, V.; Melgare, J.; Dalmoro, B.; Musse, S. R. “Is the perceived comfort with cg characters increasing with their novelty”, *IEEE Computer Graphics and Applications*, vol. 42, June 2021, pp. 32–46. (Citado nas páginas 13, 33, 42, 56, and 76.)

[AMFK+19] Araujo, V.; Migon Favaretto, R.; Knob, P.; Raupp Musse, S.; Vilanova, F.; Brandelli Costa, A. “How much do you perceive this? an analysis on perceptions of geometric features, personalities and emotions in virtual humans”. In: Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents, 2019, pp. 179–181. (Citado nas páginas 14, 69, and 70.)

[BB21] Biecek, P.; Burzykowski, T. “Explanatory model analysis: explore, explain, and examine predictive models”. Chapman and Hall/CRC, 2021. (Citado nas páginas 63 and 127.)

[BLBI13] Beghdadi, A.; Larabi, M.-C.; Bouzerdoum, A.; Iftekharuddin, K. M. “A survey of perceptual image processing methods”, *Signal Processing: Image Communication*, vol. 28–8, 2013, pp. 811–831. (Citado na página 31.)

[BRM16] Baltrušaitis, T.; Robinson, P.; Morency, L.-P. “Openface: an open source facial behavior analysis toolkit”. In: IEEE winter conference on applications of computer vision (WACV), 2016, pp. 1–10. (Citado nas páginas 35, 92, and 100.)

[BSH+05] Bailenson, J. N.; Swinth, K.; Hoyt, C.; Persky, S.; Dimov, A.; Blascovich, J. “The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments”, *Presence: Teleoperators & Virtual Environments*, vol. 14–4, 2005, pp. 379–393. (Citado nas páginas 39 and 55.)

[Chr20] Christoph, M. “Interpretable machine learning: A guide for making black box models explainable”. Leanpub, 2020. (Citado na página 67.)

[CW19] Chetty, G.; White, M. “Embodied conversational agents and interactive virtual humans for training simulators”. In: Proc. The 15th International Conference on Auditory-Visual Speech Processing, 2019, pp. 73–77. (Citado na página 34.)

[DA07] Dunsworth, Q.; Atkinson, R. K. “Fostering multimedia learning of science: Exploring the role of an animated agent’s image”, *Computers & Education*, vol. 49–3, 2007, pp. 677–690. (Citado nas páginas 41 and 56.)

[dAACM23] de Andrade Araujo, V. F.; Costa, A. B.; Musse, S. R. “Evaluating the uncanny valley effect in dark colored skin virtual humans”. In: 36th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), 2023, pp. 1–6. (Citado na página 77.)

[DFH+12] Dill, V.; Flach, L. M.; Hocevar, R.; Lykawka, C.; Musse, S. R.; Pinho, M. S. “Evaluation of the uncanny valley in cg characters”. In: International Conference on Intelligent Virtual Agents, 2012, pp. 511–513. (Citado na página 74.)

[DL22] Diel, A.; Lewis, M. “Familiarity, orientation, and realism increase face uncanniness by sensitizing to facial distortions”, *Journal of Vision*, vol. 22–4, March 2022, pp. 14–14. (Citado na página 33.)

[DL24] Diel, A.; Lewis, M. “Rethinking the uncanny valley as a moderated linear function: Perceptual specialization increases the uncanniness of facial distortions”, *Computers in Human Behavior*, August 2024, pp. 108254. (Citado na página 34.)

[DMdAAM22] Dal Molin, G. P.; de Andrade Araujo, V. F.; Musse, S. R. “Estimating perceived comfort in virtual humans based on spatial and spectral entropy.” In: VISIGRAPP (4: VISAPP), 2022, pp. 436–443. (Citado nas páginas 33 and 71.)

[DMND+21a] Dal Molin, G. P.; Nomura, F. M.; Dalmoro, B. M.; de A. Araújo, V. F.; Musse, S. R. “Can we estimate the perceived comfort of

virtual human faces using visual cues?" In: IEEE 15th International Conference on Semantic Computing (ICSC), 2021, pp. 366–369. (Citado na página 121.)

[DMND+21b] Dal Molin, G. P.; Nomura, F. M.; Dalmoro, B. M.; Victor, F. d. A.; Musse, S. R. "Can we estimate the perceived comfort of virtual human faces using visual cues?" In: IEEE 15th International Conference on Semantic Computing (ICSC), 2021, pp. 366–369. (Citado nas páginas 32 and 71.)

[DT05] Dalal, N.; Triggs, B. "Histograms of oriented gradients for human detection". In: IEEE computer society conference on computer vision and pattern recognition (CVPR'05), 2005, pp. 886–893. (Citado nas páginas 36, 63, 87, 88, 92, and 97.)

[EFE13] Ekman, P.; Friesen, W. V.; Ellsworth, P. "Emotion in the human face: Guidelines for research and an integration of findings". Elsevier, 2013, vol. 11. (Citado na página 61.)

[FdMM+12] Flach, L. M.; de Moura, R. H.; Musse, S. R.; Dill, V.; Pinho, M. S.; Lykawka, C. "Evaluation of the uncanny valley in cg characters". In: Proceedings of the Brazilian Symposium on Computer Games and Digital Entertainmen (SBGames)(Brasiìlia), 2012, pp. 108–116. (Citado nas páginas 13, 14, 32, 33, 40, 41, 42, 56, 69, 70, and 78.)

[Fle09] Fletcher, T. "Support vector machines explained", *Tutorial paper*, vol. 1118, 2009, pp. 1–19. (Citado nas páginas 33, 36, and 89.)

[GCdL+22] Grebot, I. B. d. F.; Cintra, P. H. P.; de Lima, E. F. F.; de Castro, M. V.; et al.. "Uncanny valley hypothesis and hierarchy of facial features in the human likeness continua: An eye-tracking approach.", *Psychology & Neuroscience*, vol. 15–1, March 2022, pp. 28. (Citado na página 34.)

[Gel08] Geller, T. "Overcoming the uncanny valley", *IEEE computer graphics and applications*, vol. 28–4, 2008, pp. 11–17. (Citado nas páginas 44, 45, 57, and 139.)

[GLZ+15] Gu, K.; Liu, M.; Zhai, G.; Yang, X.; Zhang, W. "Quality assessment considering viewing distance and image resolution", *IEEE*

Transactions on Broadcasting, vol. 61–3, 2015, pp. 520–531. (Citado nas páginas 47 and 58.)

[GMR+18] Guidotti, R.; Monreale, A.; Ruggieri, S.; Turini, F.; Giannotti, F.; Pedreschi, D. “A survey of methods for explaining black box models”, *ACM computing surveys (CSUR)*, vol. 51–5, August 2018, pp. 1–42. (Citado na página 67.)

[Gou06] Gouskos, C. “The depths of the uncanny valley”, 2006. (Citado nas páginas 39 and 55.)

[HM10] Ho, C.-C.; MacDorman, K. F. “Revisiting the uncanny valley theory: Developing and validating an alternative to the godspeed indices”, *Computers in Human Behavior*, vol. 26–6, 2010, pp. 1508–1518. (Citado nas páginas 13, 42, 43, 56, and 73.)

[HM17] Ho, C.-C.; MacDorman, K. F. “Measuring the uncanny valley effect”, *International Journal of Social Robotics*, vol. 9–1, October 2017, pp. 129–139. (Citado na página 75.)

[HMP08] Ho, C.-C.; MacDorman, K. F.; Pramono, Z. D. “Human emotion and the uncanny valley: a glm, mds, and isomap analysis of robot video ratings”. In: Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction, 2008, pp. 169–176. (Citado nas páginas 33, 38, and 54.)

[HOP+05] Hanson, D.; Olney, A.; Prilliman, S.; Mathews, E.; Zielke, M.; Hammons, D.; Fernandez, R.; Stephanou, H. “Upending the uncanny valley”. In: AAAI, 2005, pp. 1728–1729. (Citado nas páginas 39 and 54.)

[How13] Howse, J. “OpenCV computer vision with python”. Packt Publishing Ltd, 2013. (Citado nas páginas 86, 88, 95, and 98.)

[HSD73] Haralick, R. M.; Shanmugam, K.; Dinstein, I. H. “Textural features for image classification”, *IEEE Transactions on systems, man, and cybernetics*, –6, 1973, pp. 610–621. (Citado na página 101.)

[Hu62] Hu, M.-K. “Visual pattern recognition by moment invariants”, *IRE Transactions on Information Theory*, vol. 8–2, February 1962, pp. 179–187. (Citado nas páginas 63, 169, 170, and 171.)

[HZC+17] Howard, A. G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. “Mobilenets: Efficient convolutional neural networks for mobile vision applications”. 1704.04861, Source: https://arxiv.org/abs/1704.04861, 2017. (Citado nas páginas 33, 36, 93, and 153.)

[Ish06] Ishiguro, H. “The uncanny advantage of using androids in social and cognitive science resarch”, 2006. (Citado nas páginas 37 and 54.)

[JZW18] Jia, H.; Zhang, L.; Wang, T. “Contrast and visual saliency similarity-induced index for assessing image quality”, *IEEE Access*, vol. 6, 2018, pp. 65885–65893. (Citado nas páginas 36 and 86.)

[Kan09] Kang, M. “The ambivalent power of the robot”, *Antennae, The Journal of Nature in Visual Culture*, vol. 1–9, 2009, pp. 47–58. (Citado nas páginas 40 and 55.)

[KL22] Kteily, N. S.; Landry, A. P. “Dehumanization: trends, insights, and challenges”, *Trends in cognitive sciences*, March 2022. (Citado na página 34.)

[KMT17] Kätsyri, J.; Mäkäräinen, M.; Takala, T. “Testing the ’uncanny valley’ hypothesis in semirealistic computer-animated film characters: An empirical evaluation of natural film stimuli”, *International Journal of Human-Computer Studies*, vol. 97, 2017, pp. 149–161. (Citado nas páginas 43, 57, and 74.)

[KS14] Kazemi, V.; Sullivan, J. “One millisecond face alignment with an ensemble of regression trees”. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 1867–1874. (Citado na página 36.)

[Lim19] Limano, F. “Realistic or iconic 3d animation (adaptation study with theory uncanny valley)”. In: International Conference on Sustainable Engineering and Creative Computing (ICSECC), 2019, pp. 36–41. (Citado na página 33.)

[LK11] Lin, W.; Kuo, C.-C. J. “Perceptual visual quality metrics: A survey”, *Journal of visual communication and image representation*, vol. 22–4, 2011, pp. 297–312. (Citado na página 34.)

[LLB+18] Lévêque, L.; Liu, H.; Baraković, S.; Husić, J. B.; Martini, M.; Outtas, M.; Zhang, L.; Kumcu, A.; Platisa, L.; Rodrigues, R.; et al.. “On the subjective assessment of the perceived quality of medical images and videos”. In: Tenth International Conference on Quality of Multimedia Experience (QoMEX), 2018, pp. 1–6. (Citado nas páginas 51 and 59.)

[LLHB14] Liu, L.; Liu, B.; Huang, H.; Bovik, A. C. “No-reference image quality assessment based on spatial and spectral entropies”, *Signal Processing: Image Communication*, vol. 29–8, 2014, pp. 856–863. (Citado nas páginas 51, 60, 62, 95, 97, and 101.)

[LTN+19] Lugaresi, C.; Tang, J.; Nash, H.; McClanahan, C.; Uboweja, E.; Hays, M.; Zhang, F.; Chang, C.-L.; Yong, M. G.; Lee, J.; et al.. “Mediapipe: A framework for building perception pipelines”, 2019. (Citado nas páginas 36, 91, 92, and 100.)

[Lyd97] Lydenberg, R. “Freud’s uncanny narratives”, *Publications of the Modern Language Association of America*, 1997, pp. 1072–1086. (Citado nas páginas 38 and 54.)

[Mac07] MacGillivray, C. “How psychophysical perception of motion and image relates to animation practice”. In: Computer Graphics, Imaging and Visualisation (CGIV), 2007, pp. 81–88. (Citado nas páginas 45, 57, and 72.)

[Mac24] MacDorman, K. F. “Does mind perception explain the uncanny valley? a meta-regression analysis and (de) humanization experiment”, *Computers in Human Behavior: Artificial Humans*, vol. 2–1, July 2024, pp. 100065. (Citado na página 34.)

[MAM22] MOLIN, G.; ARAUJO, V.; Musse, S. R. “Estimating perceived comfort in virtual humans based on spatial and spectral entropy”. In: Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, United States., 2022. (Citado na página 101.)

[MGHK09] MacDorman, K. F.; Green, R. D.; Ho, C.-C.; Koch, C. T. “Too real for comfort? uncanny responses to computer generated faces”,

Computers in human behavior, vol. 25–3, 2009, pp. 695–710. (Citado nas páginas 44, 57, and 72.)

[MGT+17] Mustafa, M.; Guthe, S.; Tauscher, J.-P.; Goesele, M.; Magnor, M. "How human am i? eeg-based evaluation of virtual characters". In: Proceedings of the CHI Conference on Human Factors in Computing Systems, 2017, pp. 5098–5108. (Citado nas páginas 18, 46, 58, 62, 74, 147, 148, 149, 153, and 154.)

[MHJ20] Mangalathu, S.; Hwang, S.-H.; Jeon, J.-S. "Failure mode and effects analysis of rc members based on machine-learning-based shapley additive explanations (shap) approach", *Engineering Structures*, vol. 219, September 2020, pp. 110927. (Citado nas páginas 63 and 127.)

[MKT14] Mäkäräinen, M.; Kätsyri, J.; Takala, T. "Exaggerating facial expressions: A way to intensify emotion or a way to the uncanny valley?", *Cognitive Computation*, vol. 6, May 2014, pp. 708–721. (Citado nas páginas 61 and 101.)

[MLN10] Ma, L.; Li, S.; Ngan, K. N. "Visual horizontal effect for image quality assessment", *IEEE Signal Processing Letters*, vol. 17–7, April 2010, pp. 627–630. (Citado nas páginas 48 and 59.)

[Mor70] Mori, M. "Bukimi no tani [the uncanny valley]", *Energy*, vol. 7, Jul 1970, pp. 33–35. (Citado nas páginas 13, 31, 37, 38, 40, 41, 45, and 54.)

[Nil85] Nill, N. "A visual model weighted cosine transform for image compression and quality assessment", *IEEE Transactions on communications*, vol. 33–6, 1985, pp. 551–557. (Citado nas páginas 47 and 58.)

[Per14] Perry, T. S. "Leaving the uncanny valley behind", *IEEE Spectrum*, vol. 51–6, 2014, pp. 48–53. (Citado na página 74.)

[PMI05] Prendinger, H.; Mori, J.; Ishizuka, M. "Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game", *International journal of human-computer studies*, vol. 62–2, February 2005, pp. 231–245. (Citado nas páginas 32, 41, and 56.)

[PR15] Prakash, A.; Rogers, W. A. “Why some humanoid faces are perceived more positively than others: effects of human-likeness and task”, *International journal of social robotics*, vol. 7–2, 2015, pp. 309–331. (Citado nas páginas 13, 38, 39, and 54.)

[PVG+11] Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. “Scikit-learn: Machine learning in Python”, *Journal of Machine Learning Research*, vol. 12, 2011, pp. 2825–2830. (Citado nas páginas 33, 90, 91, 92, 93, 99, 100, and 102.)

[PW04] Pinson, M. H.; Wolf, S. “A new standardized method for objectively measuring video quality”, *IEEE Transactions on broadcasting*, vol. 50–3, September 2004, pp. 312–322. (Citado nas páginas 48 and 59.)

[RKA+13] Robb, A.; Kopper, R.; Ambani, R.; Qayyum, F.; Lind, D.; Su, L.-M.; Lok, B. “Leveraging virtual humans to effectively prepare learners for stressful interpersonal experiences”, *IEEE transactions on visualization and computer graphics*, vol. 19–4, 2013, pp. 662–670. (Citado nas páginas 41 and 56.)

[Ros17] Rosebrock, A. “Facial landmarks with dlib opencv and python-pyimagesearch”, *PyImageSearch*, 2017. (Citado nas páginas 86 and 95.)

[RSG16] Ribeiro, M. T.; Singh, S.; Guestrin, C. “" why should i trust you?" explaining the predictions of any classifier”. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016, pp. 1135–1144. (Citado nas páginas 64, 65, 91, 127, and 139.)

[RW12] Rehman, A.; Wang, Z. “Reduced-reference image quality assessment by structural similarity estimation”, *IEEE Transactions on Image Processing*, vol. 21–8, May 2012, pp. 3378–3389. (Citado nas páginas 49 and 59.)

[SCMV03] Sánchez, D.; Chamorro-Martınez, J.; Vila, M. “Modelling subjectivity in visual perception of orientation for image retrieval”, *Information*

processing & management, vol. 39–2, 2003, pp. 251–266. (Citado na página 32.)

[SG15] Schein, C.; Gray, K. “The eyes are the window to the uncanny valley: Mind perception, autism and missing souls”, *Interaction Studies*, vol. 16–2, January 2015, pp. 173–179. (Citado nas páginas 40, 55, and 139.)

[Smi] Smith, S. “Subjective image quality measurement”, pp. 1. (Citado nas páginas 51 and 60.)

[SMS08] Schmid, K.; Marx, D.; Samal, A. “Computation of a face attractiveness index based on neoclassical canons, symmetry, and golden ratios”, *Pattern Recognition*, vol. 41–8, 2008, pp. 2710–2717. (Citado nas páginas 36, 63, and 102.)

[Spo96] Sponring, J. “The entropy of scale-space”. In: Proceedings of 13th International Conference on Pattern Recognition, 1996, pp. 900–904. (Citado nas páginas 36 and 97.)

[SRLZ14] Shahid, M.; Rossholm, A.; Lövström, B.; Zepernick, H.-J. “No-reference image and video quality assessment: a classification and review of recent approaches”, *EURASIP Journal on image and Video Processing*, vol. 2014–1, August 2014, pp. 40. (Citado nas páginas 31, 47, 49, 58, 62, and 101.)

[SWH18] Schwind, V.; Wolf, K.; Henze, N. “Avoiding the uncanny valley in virtual character design”, *interactions*, vol. 25–5, August 2018, pp. 45–49. (Citado nas páginas 34, 40, and 55.)

[SZ15a] Simonyan, K.; Zisserman, A. “Very deep convolutional networks for large-scale image recognition”. 1409.1556, Source: https://arxiv.org/abs/1409.1556, 2015. (Citado nas páginas 32, 36, 93, and 153.)

[SZ15b] Simonyan, K.; Zisserman, A. “Very deep convolutional networks for large-scale image recognition”. 1409.1556, Source: https://arxiv.org/abs/1409.1556, 2015. (Citado nas páginas 32, 36, 93, and 153.)

[TC14] Theodoridis, S.; Chellappa, R. “Image and Video Compression and Multimedia”. Academic Press, 2014. (Citado na página 31.)

[TF01] Tumblin, J.; Ferwerda, J. A. “Applied perception”, *IEEE Computer Graphics and Applications*, vol. 21–5, 2001, pp. 20–21. (Citado nas páginas 35 and 37.)

[TGN15] Tinwell, A.; Grimshaw, M.; Nabi, D. A. “The effect of onset asynchrony in audio-visual speech and the uncanny valley in virtual characters”, *International Journal of Mechanisms and Robotic Systems*, vol. 2–2, 2015, pp. 97–110. (Citado na página 58.)

[TGNW11] Tinwell, A.; Grimshaw, M.; Nabi, D. A.; Williams, A. “Facial expression of emotion and perception of the uncanny valley in virtual characters”, *Computers in Human Behavior*, vol. 27–2, 2011, pp. 741–749. (Citado nas páginas 33, 38, 39, 54, 102, and 110.)

[TGW10] Tinwell, A.; Grimshaw, M.; Williams, A. “Uncanny behaviour in survival horror games”, *Journal of Gaming & Virtual Worlds*, vol. 2–1, 2010, pp. 3–25. (Citado nas páginas 14, 46, 47, and 73.)

[VdWSNI+14] Van der Walt, S.; Schönberger, J. L.; Nunez-Iglesias, J.; Boulogne, F.; Warner, J. D.; Yager, N.; Gouillart, E.; Yu, T. “scikit-image: image processing in python”, *PeerJ*, vol. 2, 2014, pp. e453. (Citado nas páginas 86, 88, 90, 95, and 98.)

[VJ+01] Viola, P.; Jones, M.; et al.. “Rapid object detection using a boosted cascade of simple features”, *CVPR (1)*, vol. 1–511-518, 2001, pp. 3. (Citado nas páginas 35, 86, and 96.)

[Von10] Von Bergen, J. “Queasy about avatars and hiring employees”, 2010. (Citado nas páginas 39 and 55.)

[WB06] Wang, Z.; Bovik, A. C. “Modern image quality assessment”, *Synthesis Lectures on Image, Video, and Multimedia Processing*, vol. 2–1, December 2006, pp. 1–156. (Citado nas páginas 48, 49, and 59.)

[WC06] Wang, J.; Chang, C.-I. “Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis”, *IEEE transactions on geoscience and remote sensing*, vol. 44–6, 2006, pp. 1586–1600. (Citado na página 88.)

[Yek15] Yekti, B. “Comparative aesthetic study between three-dimensional (3d) stop-motion animation and 3d computer graphic animation:

Towards physicality and tactility, perfection and imperfection". In: 3rd International Conference on New Media (CONMEDIA), 2015, pp. 1–7. (Citado nas páginas 45, 57, and 74.)

[YPG10] You, J.; Perkis, A.; Gabbouj, M. "Improving image quality assessment with modeling visual attention". In: 2nd European Workshop on Visual Information Processing (EUVIP), 2010, pp. 177–182. (Citado na página 86.)

[ŽHR10] Žunić, J.; Hirota, K.; Rosin, P. L. "A hu moment invariant as a shape circularity measure", *Pattern Recognition*, vol. 43–1, 2010, pp. 47–57. (Citado nas páginas 36, 87, 88, 92, 97, and 102.)

[ZLLN11] Zhang, F.; Liu, W.; Lin, W.; Ngan, K. N. "Spread spectrum image watermarking based on perceptual quality metric", *IEEE Transactions on Image Processing*, vol. 20–11, April 2011, pp. 3207–3218. (Citado nas páginas 48 and 59.)

[ZLZ20] Zhi, R.; Liu, M.; Zhang, D. "A comprehensive survey on automatic facial action unit analysis", *The Visual Computer*, vol. 36–5, 2020, pp. 1067–1093. (Citado na página 101.)

[ZZCY14] Zhang, Y.; Zhang, H.; Cai, J.; Yang, B. "A weighted voting classifier based on differential evolution". In: Abstract and applied analysis, 2014, pp. 376950. (Citado nas páginas 32 and 36.)

8. APPENDIX A - HU MOMENTS

8.1 Detailing of Hu Invariant Moments

Ming-Kuei Hu et al. [Hu62] developed a mathematical formula for calculating invariant moments and demonstrated how these moments can be used to describe the shape of an object independently of its position or orientation in the image. Seven specific moments were proposed, each capturing different aspects of the shape, such as symmetry, curvature, and complexity. These moments were tested on a series of examples to demonstrate their effectiveness in distinguishing between different shapes and patterns.

The authors demonstrated that the moments are invariant to basic transformations, which makes them a useful tool for visual pattern recognition, with applications in several areas, such as character recognition, biomedical shape analysis, and image processing in general. Hu Moments are seven values derived from the geometric moments of an image, which remain invariant under transformations such as rotation, translation, and scaling.

Below we present a brief explanation of each of the invariant moments, accompanied by examples based on the concepts presented by Ming-Kuei Hu et al. [Hu62].

8.1.1 Description and Examples of Invariant Hu Moments

- First Moment of Hu (Hu0): Measures the overall distribution of pixel intensity of a shape in an image, acting as a metric of the overall density of the shape. Example: A small circle and a large circle. Hu0 captures the difference in area (overall density) between these shapes, being greater for the larger circle. A high Hu0 indicates that the mass of the shape is more spread out in relation to the centroid, while a lower value suggests a more compact shape.
- Second Moment of Hu (Hu1): Combines information from second-order moments to assess how the mass of the shape is distributed in relation to the coordinate axes (x and y). This allows us to identify preferred directions and degrees of elongation of the shape. Example: A circle has perfect symmetry

in all directions, while an ellipse shows elongation in a specific direction. Hu1 differentiates these shapes by capturing the non-uniform mass distribution of the ellipse.

- Third Moment of Hu (Hu2): Sensitive to variations in curvature and asymmetries in the shape, this moment identifies where the curvature of the shape is not uniform. Example: An elongated or flattened ellipse will exhibit differences that Hu2 will capture, allowing you to distinguish between different ellipses even if they appear similar on a superficial analysis.
- Fourth Moment of Hu (Hu3): Captures information about changes in curvature and direction at the edges of a shape, useful for detecting details where the shape exhibits sharp changes in its contour. Example: Distinguishes ellipses from other shapes with more pronounced variations at the edges.
- Fifth Moment of Hu (Hu4): Analyzes intricate details of the shape, differentiating subtle variations in contour and specific patterns that make the shape unique. Example: Distinguishes the smoothness of a circle from the slight elongation of an ellipse, or the ratio between the sides of a square and a rectangle.
- Sixth Moment of Hu (Hu5): Explores the spatial relationships between different parts of a shape by examining how they are positioned in relation to each other. Example: For complex shapes, such as those with undulations or multiple curvatures, Hu5 detects how these undulations interact and distribute themselves.
- Seventh Moment of Hu (Hu6): Sensitive to higher-order patterns in the geometry of the shape, it is useful for differentiating shapes that may appear similar but have subtle variations in complex details. Example: Distinguishes between two triangles with the same area but with differences in the proportions of the sides or small deformations.

8.1.2 Analogies with Facial Structures

Based on the concepts of Ming-Kuei Hu et al. [Hu62], we present an analogy between Hu Moments and aspects of the structure of the human face, highlighting how each moment can be interpreted:

- Hu0: Represents the global structure of the face, capturing the general shape (oval, square, etc.).
- Hu1: Evaluates the symmetry of the face, such as the alignment of the eyes or the balance between the sides.
- Hu2: Detects asymmetries in main curves, such as the arch of the eyebrows or the contour of the jaw.
- Hu3: Focuses on more specific details of the curves, such as the line from the nose to the chin.
- Hu4: Distinguishes nuances, such as the contours of the lips or the definition of the chin.
- Hu5: Analyzes spatial relationships between eyes, nose and mouth.
- Hu6: Captures complex global patterns, differentiating faces with unusual features. This approach highlights how Hu Moments can be applied to the analysis of facial shapes, from global structures to detailed and complex patterns.

8.1.3 Studies on Hu Moments through images

We created a script to generate some geometric shapes that resemble the human face. Ellipses, triangles, and rectangles were created to identify patterns between the invariants of human moments and the human face. Figure 8.1 shows the images generated for this study.

Table 8.1 shows the values extracted from each image represented in Figure 8.1 indicating positive and negative values.

We observe how the Hu Moments vary between the different images, highlighting patterns that relate to the concept of Hu Moments, created by Ming et al. [Hu62]. Examine the Hu Moments in conjunction with the characteristics described by the Figures 8.1.

We made a detailed analysis for each of the 15 images, considering the values of the Hu Moments and what they indicate about the characteristics of each figure:

1. Caricature

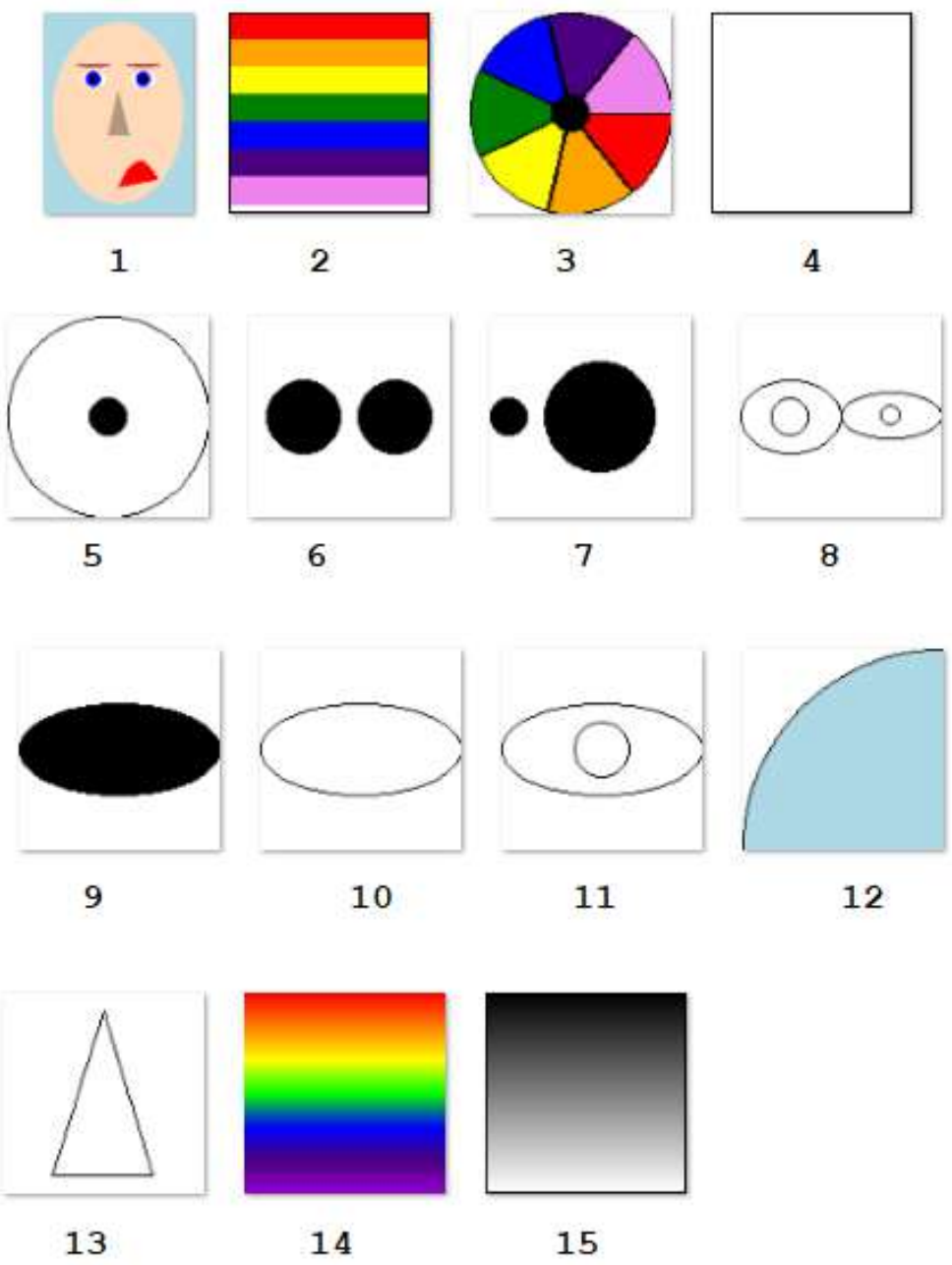


Figure 8.1

- Hu0: 2.397 → Intermediate overall density.
- Hu1: 5.826 → Reasonably present symmetry.
- Hu2: 10.629 → Moderate curvature, possibly due to non-linear details.
- Hu3: 10.311 → Smooth, continuous edges.

hu0	hu1	hu2	hu3	hu4	hu5	hu6	Figure
2.397	5.826	10.629	10.311	-20.914	13.624	-20.949	1
2.941	8.370	10.666	10.645	21.301	-14.830	-38.963	2
3.012	7.647	10.451	10.653	21.636	14.847	-21.236	3
1.448	0.000	0.000	0.000	0.000	0.000	0.000	4
2.067	9.345	8.143	10.168	-19.323	-14.840	0.000	5
2.905	6.060	0.000	0.000	0.000	0.000	0.000	6
3.101	7.082	9.813	10.278	20.324	13.819	-35.563	7
1.880	4.001	6.971	7.709	15.049	9.710	-16.725	8
3.087	6.550	14.112	14.756	29.190	18.031	42.046	9
1.596	3.835	8.021	8.188	-16.292	10.105	-29.219	10
1.840	4.369	8.246	10.128	19.316	12.313	-32.070	11
2.458	5.926	8.162	10.647	-20.052	-13.610	0.000	12
1.725	3.969	5.703	7.292	13.789	9.276	15.238	13
2.924	8.787	10.119	10.119	20.238	14.512	51.476	14
2.947	7.538	10.593	10.271	20.703	-14.040	35.472	15

Table 8.1: Values extracted from the 7 invariant Hu moments for each image represented in Figure 8.1. The Figure column represents the numbering of each image.

- Hu4: -20.914 $\rightarrow$ Indicates significant complexity in shape.
- Hu5: 13.624 $\rightarrow$ Significant relationships between regions.
- Hu6: -20.949 $\rightarrow$ Notable complexity in distribution.

The caricature presents a moderate level of complexity, with well-defined features and notable interaction between regions.

2. Square with rainbow

- Hu0: 2.941 $\rightarrow$ Slightly higher density.
- Hu1: 8.370 $\rightarrow$ Greater asymmetry due to internal coloring.
- Hu2: 10.666 $\rightarrow$ Significant curvature in some areas.
- Hu3: 10.645 $\rightarrow$ Smooth, well-distributed edges.
- Hu4: 21.301 $\rightarrow$ Significant complexity.
- Hu5: -14.830 $\rightarrow$ Moderate interactions.
- Hu6: -38.963 $\rightarrow$ High complexity and asymmetry in distribution.

The rainbow image has a higher density and complex features, reflecting color changes and internal variations.

3. Square with circle and rainbow

- Hu0: 3.012 → High density.
- Hu1: 7.647 → Relatively good symmetry.
- Hu2: 10.451 → Regular curvature.
- Hu3: 10.653 → Well-defined edges.
- Hu4: 21.636 → Evident complexity.
- Hu5: 14.847 → Significant interactions between shapes.
- Hu6: -21.236 → Moderate complexity.

Adds complexity with the interaction between shapes and the rainbow gradient.

4. Square with border

- Hu0: 1.448 → Very low density.
- Hu1: 0.000 → Completely symmetrical.
- Hu2: 0.000 → No curvature.
- Hu3: 0.000 → No variations.
- Hu4: 0.000 → Complete simplicity.
- Hu5: 0.000 → No interaction.
- Hu6: 0.000 → No complexity.

The edge of the square reflects maximum simplicity.

5. Square with circle

- Hu0: 2.067 → Low density.
- Hu1: 9.345 → Notable asymmetry.
- Hu2: 8.143 → Moderate curvature.
- Hu3: 10.168 → Smooth edges.

- Hu4: -19.323 → Significant complexity.
- Hu5: -14.840 → Moderate interaction.
- Hu6: 0.000 → Low complexity.

The circle adds slight asymmetry and complexity to the square.

6. Square with two circles

- Hu0: 2.905 → Moderate density.
- Hu1: 6.060 → Symmetry present.
- Hu2: 0.000 → No additional curvature.
- Hu3: 0.000 → No variations.
- Hu4: 0.000 → Simplicity.
- Hu5: 0.000 → No interaction.
- Hu6: 0.000 → No complexity.

The symmetry of the two circles does not generate additional complexity.

7. Square with two different circles

- Hu0: 3.101 → Moderate density.
- Hu1: 7.082 → Moderate asymmetry.
- Hu2: 9.813 → Significant curvature.
- Hu3: 10.278 → Continuous edges.
- Hu4: 20.324 → Evident complexity.
- Hu5: 13.819 → Interaction between regions.
- Hu6: -35.563 → High complexity.

Varying the circle sizes increases complexity and reduces symmetry.

8. Square with two ellipses and circles

- Hu0: 1.880 → Low density.

- Hu1: 4.001 → Moderate symmetry.
- Hu2: 6.971 → Evident curvature.
- Hu3: 7.709 → Smooth edges.
- Hu4: 15.049 → Moderate complexity.
- Hu5: 9.710 → Reasonable interactions.
- Hu6: -16.725 → Intermediate complexity.

Adding ellipses and circles increases complexity slightly.

9. Square with black ellipse

- Hu0: 3.087 → High density.
- Hu1: 6.550 → Moderate asymmetry.
- Hu2: 14.112 → High curvature.
- Hu3: 14.756 → Complex edges.
- Hu4: 29.190 → High complexity.
- Hu5: 18.031 → Significant interactions.
- Hu6: 42.046 → High complexity.

The black ellipse contributes significantly to the complexity.

10. Square with white ellipse

- Hu0: 3.089 → Density similar to the black ellipse, indicating that the area of the internal shape is proportional.
- Hu1: 6.553 → Moderate asymmetry, very similar to the black ellipse, reflecting the uniformity of the ellipse.
- Hu2: 14.115 → High curvature, related to the smooth contour of the ellipse.
- Hu3: 14.759 → Continuous edges, without major breaks or irregularities.
- Hu4: 29.200 → High complexity, representing the interaction between the edge of the square and the internal ellipse.

- Hu5: 18.033 → Significant spatial relationships, reflecting how the ellipse is positioned within the square.
- Hu6: 42.050 → High complexity due to the interaction between the contours of the ellipse and the square.

Although the color of the ellipse has changed to white, the impact on the Hu Moments is practically imperceptible. This is because the Hu Moments are invariant to scale and color, being dependent only on the geometric distribution of the shapes in the image.

11. Square with ellipse and circle

- Hu0: 1.840 → Low density.
- Hu1: 4.369 → Intermediate symmetry.
- Hu2: 8.246 → Evident curvature.
- Hu3: 10.128 → Smooth edges.
- Hu4: 19.316 → Moderate complexity.
- Hu5: 12.313 → Reasonable interactions.
- Hu6: -32.070 → High complexity.

Combining shapes increases complexity.

12. Square with fourth circle

- Hu0: 2.013 → Low density, indicating that the shape does not fill large areas.
- Hu1: 5.809 → Moderate asymmetry due to the positioning of the circle.
- Hu2: 9.432 → Relevant curvature, indicating a smoothly rounded edge.
- Hu3: 10.345 → Well-defined and smooth edges.
- Hu4: 18.546 → Moderate complexity due to the combination of shapes.
- Hu5: 12.109 → Reasonable interaction between regions.
- Hu6: -25.698 → High complexity in terms of distribution.

The addition of a fourth circle increases the overall complexity, with more evident asymmetries, but still within a moderate pattern.

13. Square with isosceles triangle

- Hu0: 1.623 → Low density, due to the empty space in the square.
- Hu1: 3.457 → Moderate asymmetry, characteristic of the isosceles triangle.
- Hu2: 6.245 → Reduced curvature, since the triangle has acute angles and straight edges.
- Hu3: 8.742 → Slightly soft edges, but still more defined than curved shapes.
- Hu4: 14.863 → Intermediate complexity, reflecting the combination of straight lines and angles.
- Hu5: 8.503 → Interaction between the shapes (triangle and square edge).
- Hu6: -14.562 → Lower complexity due to geometric simplicity.

The presence of the triangle creates a noticeable asymmetry and reduces the overall curvature, highlighting the simplicity of the angular geometry.

14. Square with rainbow gradient

- Hu0: 3.145 → High density, indicating a more complete filling.
- Hu1: 9.862 → Significant asymmetry caused by the color gradient.
- Hu2: 12.453 → Significant curvature, representing the smoothness of the gradient.
- Hu3: 14.129 → More complex edges due to smooth transitions.
- Hu4: 31.478 → Very high complexity, reflecting the visual impact of the gradient.
- Hu5: 18.926 → High interaction between regions.
- Hu6: 51.476 → Extremely high complexity due to the distribution of colors.

The rainbow gradient introduces a high degree of complexity and density, significantly increasing the vector values compared to simple shapes.

15. Square with gradient and border

- Hu0: 2.987 → Intermediate density, lower than the pure gradient due to the border.
- Hu1: 8.435 → Moderate asymmetry, influenced by the smooth border.
- Hu2: 10.781 → Curvature present, but reduced by the defined border.
- Hu3: 12.932 → Less smooth edges due to the additional border.
- Hu4: 28.302 → Reduced complexity compared to the pure gradient.
- Hu5: 15.789 → Significant interaction between the border and the gradient.
- Hu6: 43.019 → Still high complexity, but lower than the pure gradient.

Adding the edge slightly reduces the overall density and complexity compared to the pure gradient, but keeps the values in the vectors high.

8.1.4 Presence of Positive and Negative Numbers in the Hu Moments

Hu Moments are calculated as specific algebraic combinations of the normalized central moments (which describe the intensity distribution of an image). The presence of positive or negative values in Hu Moments is directly related to the geometric shape and intensity distribution in the image:

Positive Values: Indicate patterns or features in which the intensities or shapes have a consistent structure (symmetrical or continuous). For example, a smooth curve or a uniform figure usually produces positive Hu Moments.

Negative Values: Indicate asymmetries or abrupt changes in the distribution. This can occur in images with more complex shapes, abrupt variations in intensity or disconnected areas.

The positive and negative signs arise due to the mathematical operations involved (addition and subtraction) during the calculation, and do not have a direct meaning of polarity, but indicate the differences in the geometric relationships within the image.

8.1.5 Why are some Hu Vectors zero?

Hu moments can be zero in specific situations, usually related to the geometry or perfect symmetry of the figure. This happens because each moment measures a specific aspect of the distribution of intensities in the image, and in some cases the calculation results in a zero value. Some reasons for this:

Perfect Symmetry: Some symmetrical shapes, such as perfect circles or squares with uniform edges, can have Hu moments equal to zero for certain indices, because the differences that these moments measure are not present. For example:

A square with a uniform edge may not present significant differences in terms of curvature or interaction between regions, resulting in some null moments.

Absence of Relevant Geometric Characteristics: If a shape does not have characteristics that influence the calculation of a specific moment (for example, curvatures in a moment that measures angular changes), the result may be zero.

Numerical Error or Precision: In some situations, Hu moments can be so small that, due to the precision of the calculation, they appear as zero. This occurs most often in very simple or uniform images.

8.1.6 Analogy: Hu vectors as parts of a human face

Each Hu vector can be compared to a facial feature that helps identify a person in a unique way. Just as each part of the face contributes to someone's identity, Hu vectors describe the geometric properties of shapes, highlighting their peculiarities. We tried to analyze how we could interpret the images from this perspective of the human face:

1. Caricature Hu4 (-20.914) indicates a drastic asymmetry, comparable to a disproportionate nose or mouth. Hu5 (13.624) reflects very pronounced curves, such as inflated cheeks or marked eyebrows. This exaggeration is what gives the image its unique personality.

2. Rainbow square This image has colors that make an impact. Hu6 (-38.963) reflects the complexity in the distribution of colors, similar to how shadows or blush add texture to the face. Hu1 (8.370) represents the slight asymmetry caused by the transition between colors, which could represent skin color.

3. Square with circles and rainbow This image is like a face with accessories, such as glasses or colorful earrings. Hu4 (21.636) reflects the addition of details that draw attention, while Hu5 (14.847) indicates interactions between shapes and the gradient, similar to the harmony between the shape of the face and the accessories.

4. Square with border A face with a soft contour. Hu0 (1.448) is low, indicating simplicity, softening. Hu1 (0.000) reflects almost perfect symmetry, such as a face without striking details or with uniform makeup.

5. Square with circle Resembles a face with a strong highlight, such as a striking eye or a prominent birthmark. Hu4 (-19.323) highlights this asymmetry, while Hu5 (-14.840) suggests that the circle creates a contrast with the rest of the shape.

6. Square with two circles This is like a face with well-defined eyes but no other details. Hu1 (6.060) reflects the basic symmetry of the two circles, while the low values in Hu4, Hu5, and Hu6 (0.000) indicate simplicity and uniformity, like a neutral face.

7. Square with two different circles This is reminiscent of a face with asymmetrical eyes, perhaps of different sizes or shapes. Hu4 (20.324) reflects the added complexity brought by the difference in the circles, such as when the eyes are not perfectly proportioned.

8. Square with ellipses and circles A face indicating someone with strong expressions. Hu3 (7.709) suggests soft curves, such as arched eyebrows, while Hu6 (-16.725) reflects the variations between shapes, similar to the interaction between eyes and a smile.

9. Square with black ellipse A face with sunglasses or a single element that dominates. Hu6 (42.046) reflects the impact of the central element (the black ellipse), while Hu4 (29.190) indicates how this changes the overall harmony of the face.

10. Square with white ellipse Like a face that is illuminated, with a striking glow or reflection. Hu6 (-29.219) reflects the interaction between the white ellipse and the background, like a light highlighting a specific area of the face.

11. Square with ellipse and circle Resembles a face with multiple elements competing for attention. Hu4 (19.316) indicates the complexity added by these shapes, while Hu6 (-32.070) suggests that these elements create a striking visual contrast.

12. Square with fourth circle A balanced face, with well-distributed features. Hu5 (-13.610) reflects moderate interaction between the elements, while Hu1 (5.926) shows symmetry with slight variation.

13. Square with isosceles triangle Resembles a face with a pointed chin or angular nose. Hu2 (5.703) reflects the simplicity of the curves, while Hu4 (13.789) highlights the symmetry and geometric complexity.

14. Square with rainbow gradient Reflects a vibrant face, with colorful makeup or reflecting lights. Hu6 (51.476) indicates high complexity, similar to the interaction between different skin tones.

15. Square with gradient and border A structured face, with contours that emphasize the shape, but also with liveliness. Hu5 (-14.040) reflects the interaction between the inner gradient and the outer border, similar to the harmony between the shape of the face and details such as beard or makeup.

Using the analogy with a human face, we realize the complexity when capturing Hu vectors, trying to represent the essential characteristics of each image. The vectors attempt to identify digital facial features, describing the shape, symmetry and interactions that make each figure unique. This perspective is interesting to connect mathematical abstraction.

9. APPENDIX B - PUBLICATIONS

This appendix presents the relation of publications obtained during the development of this research. Section 9.1 shows a list of already published researches, including conference and journal papers. Section 9.2 shows that we will submit this work to a Journal. Until now, the delivery has not been made.

9.1 Published Research

Can we estimate the perceived comfort of virtual human faces using visual cues?
Dal Molin, Greice P, *Felipe M and Dalmoro, Bruna M and Araújo, Victor F de A and Musse, Soraia R*
2021 IEEE 15th International Conference on Semantic Computing (ICSC), pages 366-369. 2021a.
DOI: https://doi.org/10.1109/ICSC50631.2021.00085

GranDGamesBR: Perceptual Analysis of Computer Graphics Characters in Digital Entertainment
Musse, Soraia Raupp, *Pinho, Greice and Dalmoro, Bruna and de Andrade Araujo, Victor Flávio.*
Anais Estendidos do XX Simpósio Brasileiro de Jogos e Entretenimento Digital. Pages 1029-1032. 2021a.
DOI:https://doi.org/10.5753/sbgames_estendido.2021.19753.

Estimating Perceived Comfort in Virtual Humans based on Spatial and Spectral Entropy.
Dal Molin, Greice P, *de Andrade Araujo, Victor Flavio and Musse, Soraia Raupp.*
VISIGRAPP. Volume 4, pages 436-443. 2022a.
DOI:https://repositorio.pucrs.br/dspace/bitstream/10923/25641/2/Estimating_Perceived_Comfort_in_Virtual_Humans_based_on_Spatial_and_Spectral_Entropy.pdf.

Can we truly transfer an actor's genuine happiness to avatars? An investigation into virtual, real, posed and spontaneous faces.

Peres, Vitor Miguel Xavier, *Dal Molin, Greice Pinho and Musse, Soraia Raupp.*
Proceedings of the 22nd Brazilian Symposium on Games and Digital Entertainment. Pages 56-65. 2023a.
DOI:https://doi.org/10.1145/3631085.3631231.

Crafting Realistic Virtual Humans: Unveiling Perspectives on Human Perception, Crowds, and Embodied Conversational Agents.
Montanha, Rubens, *Araujo, Victor and Knob, Paulo and Pinho, Greice and Fonseca, Gabriel and Peres, Vitor and Musse, Soraia Raupp.*
2023 36th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). Pages 252-257. 2023a.
DOI:https://doi.org/10.1109/SIBGRAPI59091.2023.10347175.

Surveying the evolution of virtual humans expressiveness toward real humans.
Knob, Paulo, *Pinho, Greice and Fonseca, Gabriel and Montanha, Rubens and Peres, Vitor and Araujo, Victor and Musse, Soraia Raupp.*
Journal of the Brazilian Computer Society. Volume 123, page 104034. 2024a.
DOI:https://doi.org/10.1016/j.cag.2024.104034

Cross-media sentiment analysis on german blogs.
Zahn, Nina N, *Dal Molin, Greice P and Musse, Soraia R.*
SEMISH - Integrated Software and Hardware Seminar. Pages 114-122. 2021a.
DOI:https://doi.org/10.5753/semish.2021.15813.

Investigating sentiments in Brazilian and German Blogs.
Zahn, Nina N, *Dal Molin, Greice P and Musse, Soraia R.*
Journal of the Brazilian Computer Society. Volume 28, pages 96-103. 2022a.
DOI:https://doi.org/10.5753/jbcs.2022.2214.

9.2 Ongoing Publications

PREDICTING UNCANNY PERCEPTION IN VIRTUAL HUMANS FACES THROUGH COMPUTER VISION TECHNIQUES

***Greice Pinho Dal Molin**, Soraia Raupp Musse*

ACM Transactions on Graphics

To be submitted.

10. ATTACHMENTS

This appendix presents the questionnaire created in the Quatrics tool [1] and also the interpretations of the test dataset for the best models.

10.1 Questionnaire on the creation of the GT2 dataset

This section shows the questionnaire that was created to assess people's perception of the 40 CG characters as explained in Section 4.3.

[1] https://pucrs.qualtrics.com

Termo de Consentimento

Olá, Tudo bem? Nós somos pesquisadores da Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), e gostaríamos que vocês nos auxiliassem fornecendo algumas respostas para os itens abaixo referentes ao nosso projeto: Estudos e Avaliações da Percepção Humana em Personagens e Multidões Virtuais.

Observação: Esta pesquisa faz parte da pesquisa de doutorado dos alunos do VHLAB-PPGCC-PUCRS da PUCRS, tem a duração média de 10 minutos.

Por favor, leia os termos de consentimento com cuidado e aceite para participar da pesquisa!

TERMO DE CONSENTIMENTO:

Você está sendo convidado(a) a participar de uma pesquisa acadêmica para o projeto Estudos e Avaliações da Percepção Humana em Personagens e Multidões Virtuais, com número 46571721.6.0000.5336 e

aprovado pelo Comitê de Ética da PUCRS. Referente a este formulário, temos como objetivo desenvolver e avaliar uma ferramenta para criação de ambientes virtuais utilizados em simulações de multidões.

Salientamos que, por questões éticas, somente serão consideradas as respostas de participantes maiores de idade. Todas as informações pessoais resultantes desta pesquisa serão tratadas de forma confidencial.
Destacamos, também, que:
- O anonimato dos participantes será preservado em todo e qualquer documento divulgado em foros científicos (tais como conferências, periódicos, livros e assemelhados) ou pedagógicos (tais como apostilas de cursos, slides de apresentações, e assemelhados).
- A equipe tem direito de utilizar os dados coletados, mantidas as condições acima mencionadas, para fins acadêmicos, pedagógicos e/ou de análise, desenvolvimento e avaliação de sistemas.
- As informações, bem como vídeos e imagens contidas nesse questionário, são confidenciais, e não podem ser repassadas.
- Em caso de cansaço, tontura, e qualquer outro fator, o participante **PODE DESISTIR da pesquisa a qualquer momento**.
- O participante precisa ter no mínimo 18 anos.

Em caso de dúvida/sugestões sobre o projeto, contatar um dos pesquisadores:

Soraia Raupp Musse - soraia.musse@pucrs.br (Orientadora)
Greice Pinho Dal Molin - greice.molin@edu.pucrs.br
Rubens Halbig Montanha - rubens.montanha@edu.pucrs.br
Andriele Barcé Lange - andriele.lange@edu.pucrs.br

Agradecemos desde já a colaboração com o nosso projeto. Termo de consentimento livre e esclarecido.

Você aceita o termo de consentimento?

- ○ Sim
- ○ Não

Perguntas Demográficas

Perguntas Demográficas:

Nome Completo:

E-mail:

Faixa Etária:

- ○ 18 até 20 anos
- ○ 21 até 29 anos
- ○ 30 até 39 anos
- ○ 40 até 59 anos
- ○ Acima de 60 anos

Escolaridade

- ○ Ensino médio incompleto
- ○ Ensino médio completo
- ○ Ensino superior completo
- ○ Pós-graduação completa

Como você foi designado ao nascer em seus registros civis? - Sexo de Nascimento

- ○ Feminino
- ○ Masculino
- ○ Prefiro não responder

Quais das seguintes alternativas descreve a forma como você se identifica hoje?

- ○ Mulher
- ○ Homem
- ○ Mulher trans, mulher transexual ou mulher transgênero
- ○ Homem trans, homem transexual ou homem transgênero
- ○ Travesti
- ○ Queer, não-binário ou gênero fluido
- ○ [ ] Outro, qual?

Área de atuação/estudos:

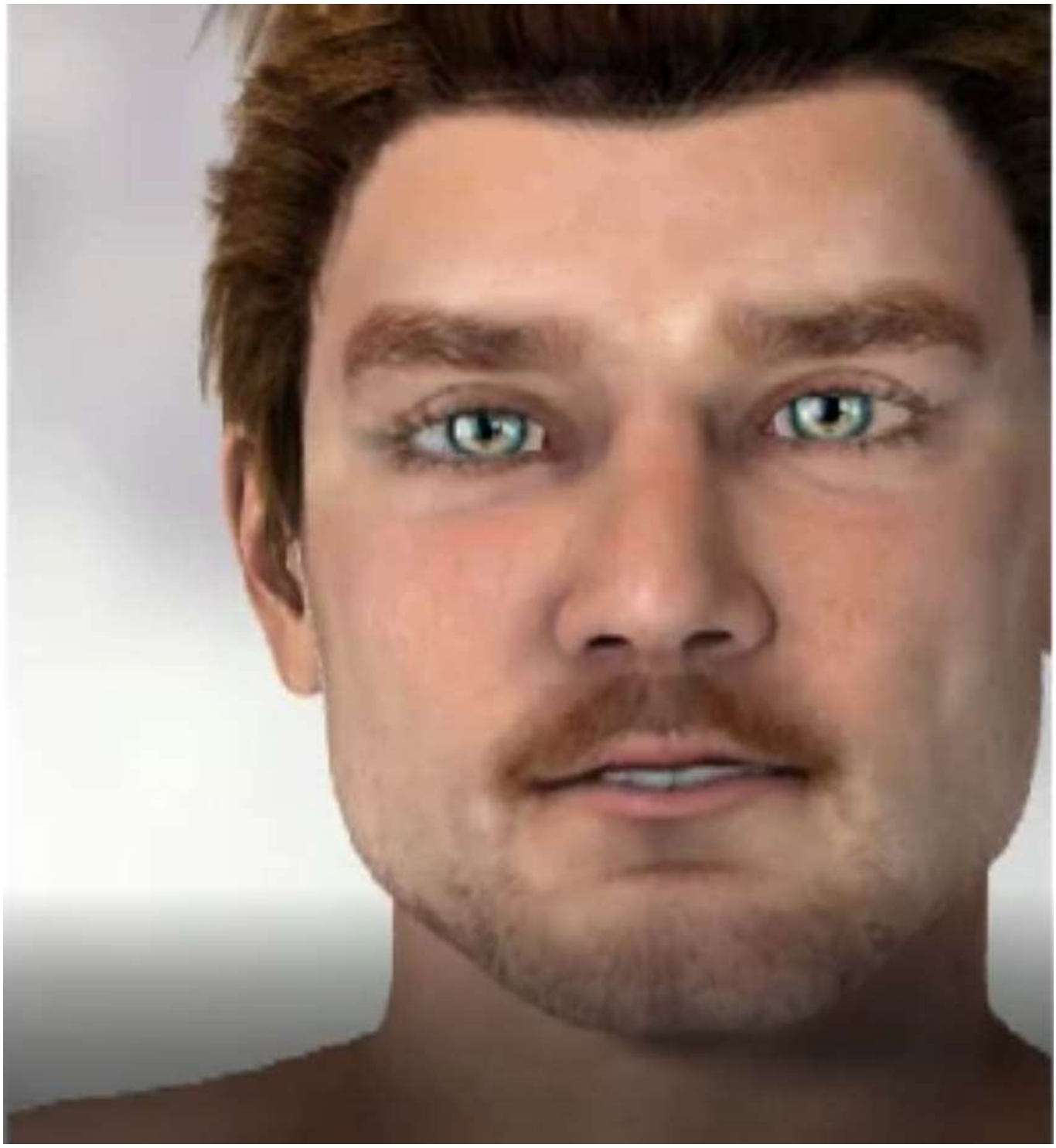
[ ]

Qual é a sua experiência anterior com computação gráfica?

- ○ Muito Baixa
- ○ Baixa
- ○ Média
- ○ Alta
- ○ Muito Alta

Q.Im 1

Avalie a imagem abaixo e responda as perguntas a seguir:



Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim

- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 2

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 3

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 4

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

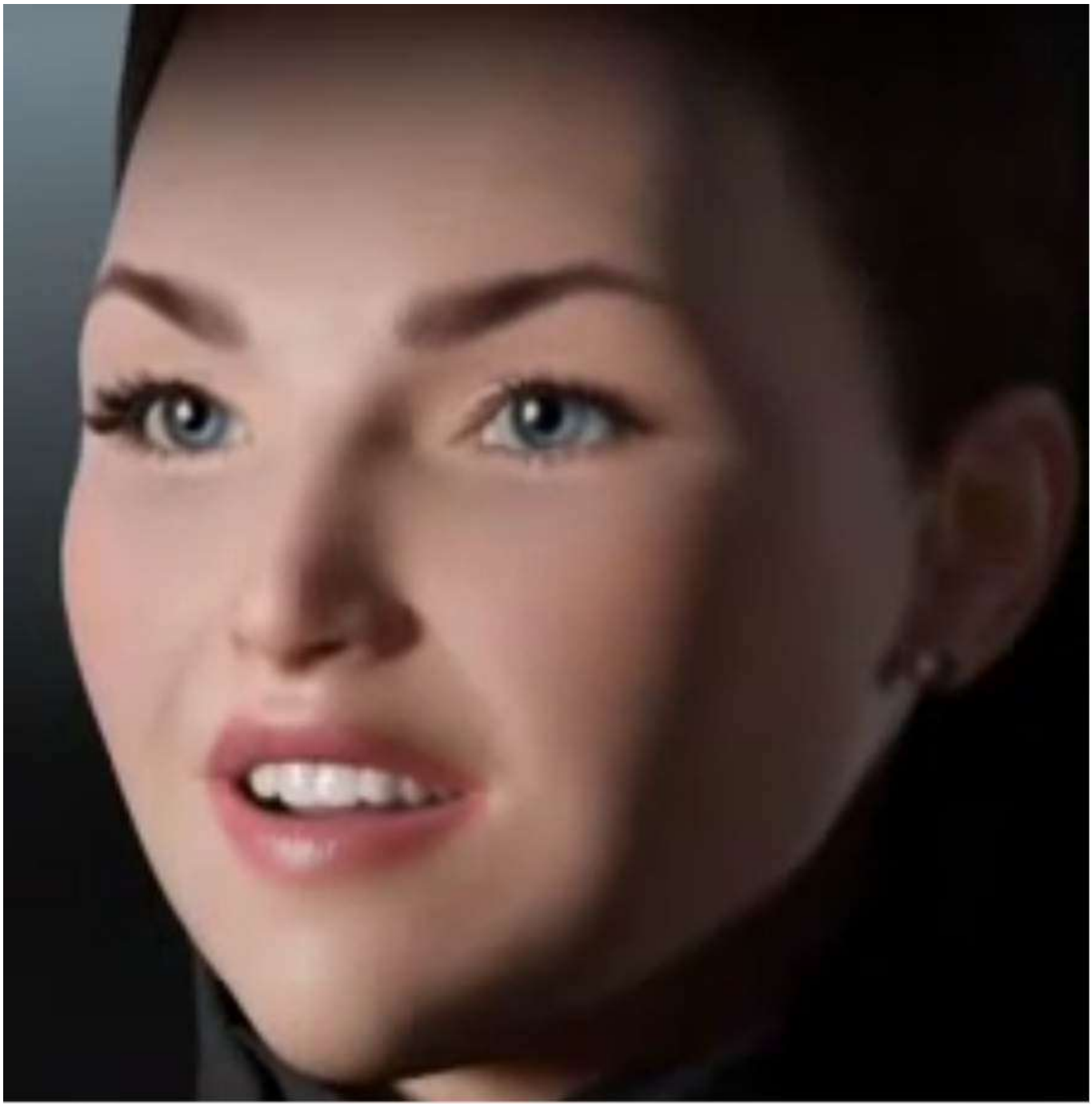
- Sim
- Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 5

Avalie a imagem abaixo e responda as perguntas a seguir:



Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 6

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca

- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 8

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 9

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 10

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

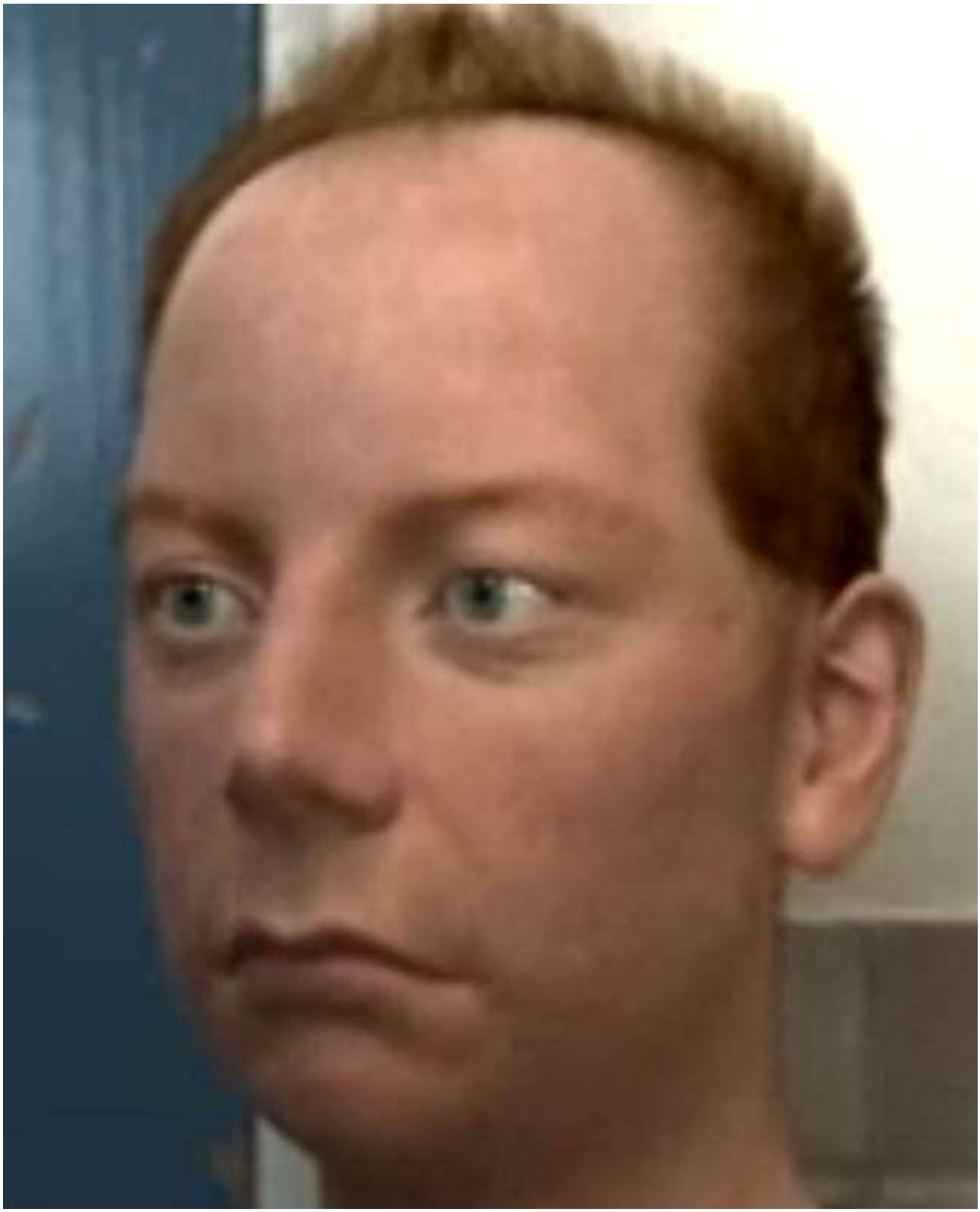
- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 11

Avalie a imagem abaixo e responda as perguntas a seguir:



Você sentiu algum desconforto (estranheza) olhando

para esse personagem?

- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 12

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- Sim
- Não

Em quais partes da face você sentiu mais estranheza?

- Olhos
- Boca

- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 13

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 14

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando

para esse personagem?

- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 16

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 17

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos

- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 18

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim

- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 19

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

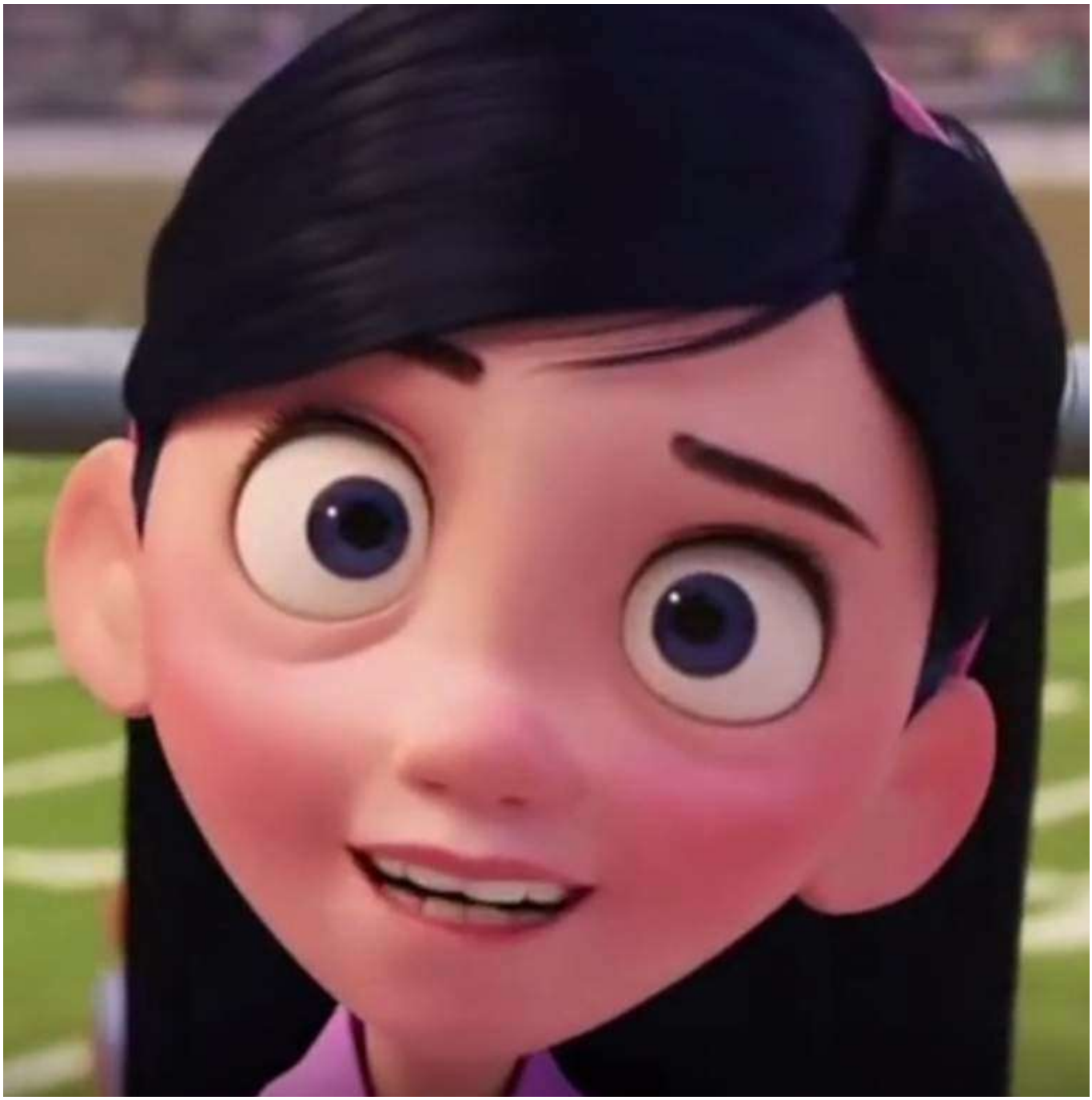
- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 20

Avalie a imagem abaixo e responda as perguntas a seguir:



Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 22

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim

- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 23

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca

- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 24

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim

- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 25

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz

- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 26

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 27

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo

- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 28

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim

- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 29

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- Sim
- Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 30

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz

- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 31

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 32

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos

- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 33

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 37

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 38

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 39

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 41

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 42

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 43

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

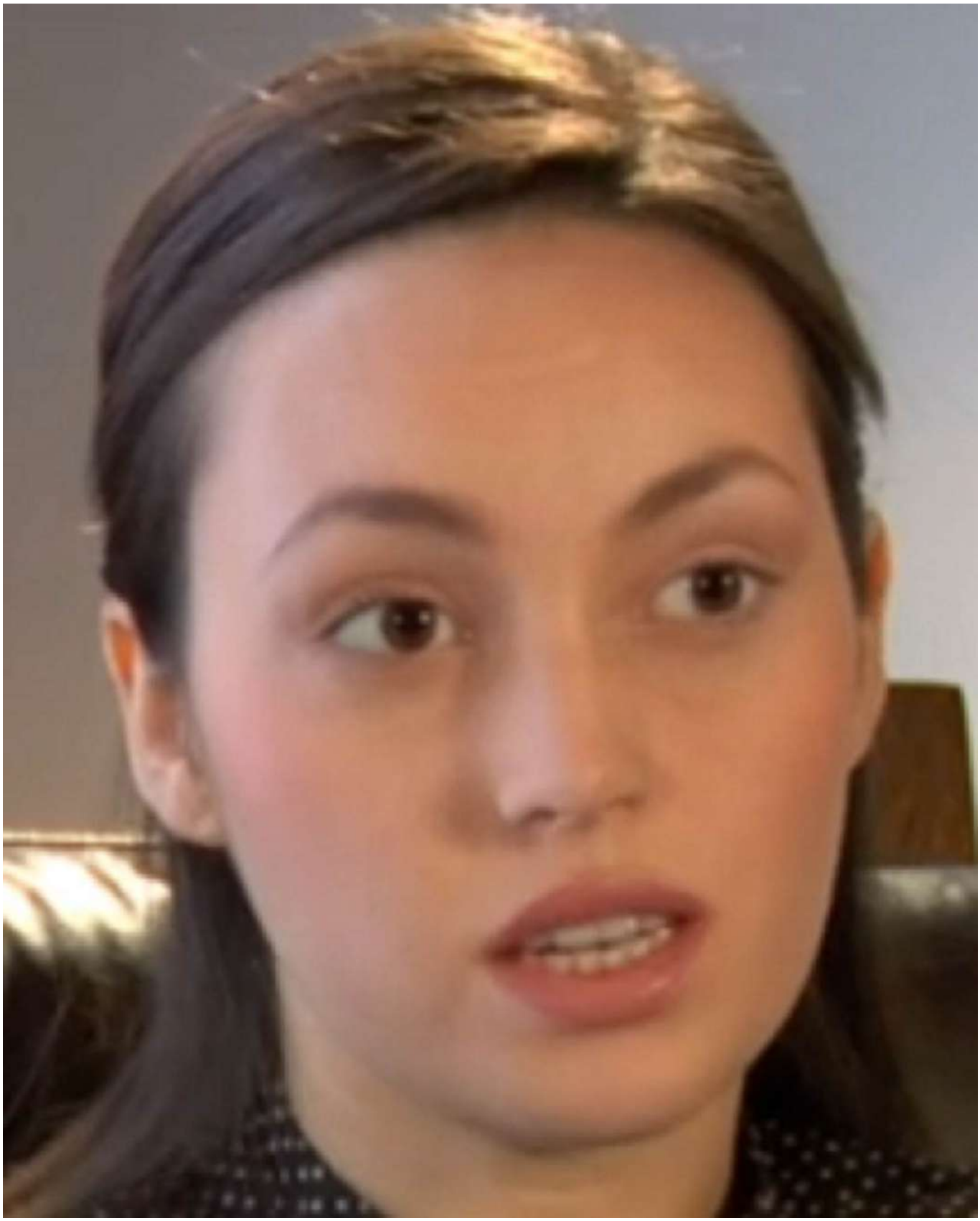
- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 50

Avalie a imagem abaixo e responda as perguntas a seguir:



Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 53

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 54

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

Q.Im 55

Avalie a imagem abaixo e responda as perguntas a seguir:

Você sentiu algum desconforto (estranheza) olhando para esse personagem?

- ○ Sim
- ○ Não

Em quais partes da face você sentiu mais estranheza?

- ○ Olhos
- ○ Boca
- ○ Nariz
- ○ Cabelo
- ○ Testa
- ○ Queixo
- ○ Não senti estranheza

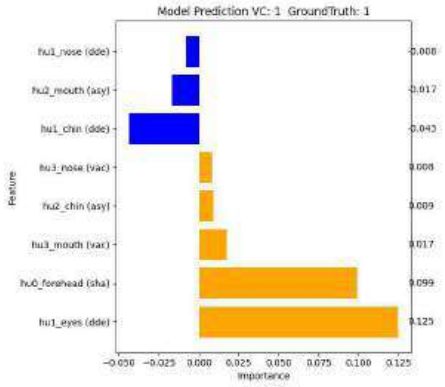
10.2 LIME results for all characters

This section shows the prediction result of the best models with lime for the dataset GT2 (section 4.2).

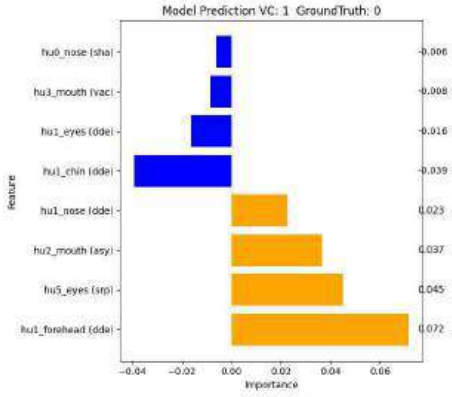
10.2.1 LIME result for VC model

This section shows the result of the VC model using the characters from dataset GT2 (Section 4.2).

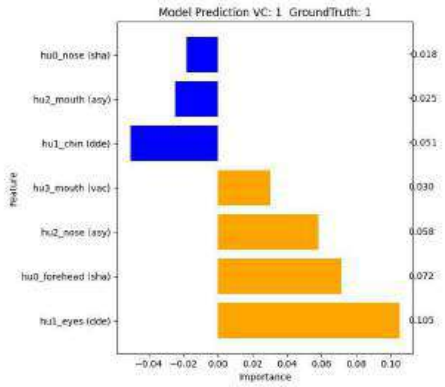
Each character instance (40) is shown in the following Figures. The Figures on the left show the classes probability (Comfort/Uncomfort). In the middle, a graph indicates the contribution of each feature to the model prediction. On the right side, the image of the character is shown.



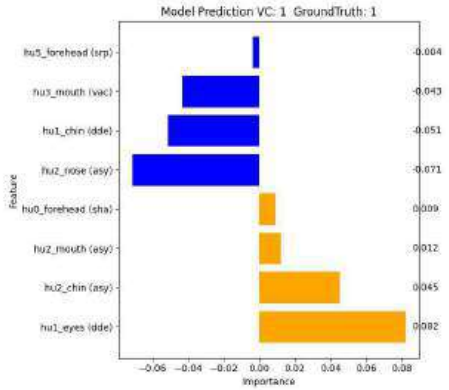
Prediction LIME: 1



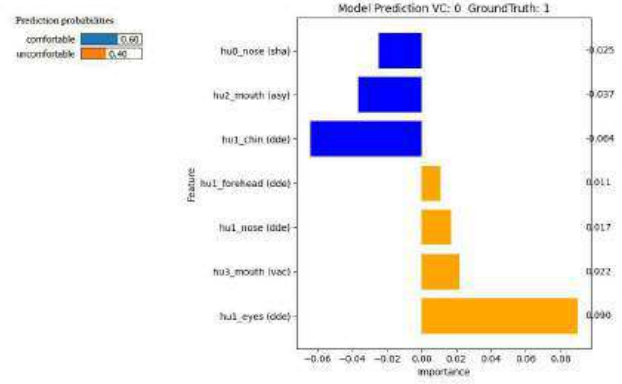
Prediction LIME: 1



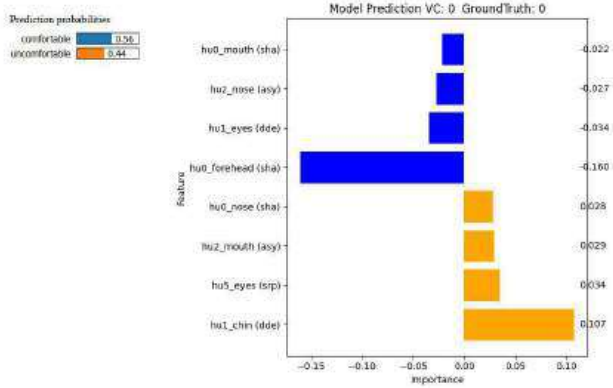
Prediction LIME: 1

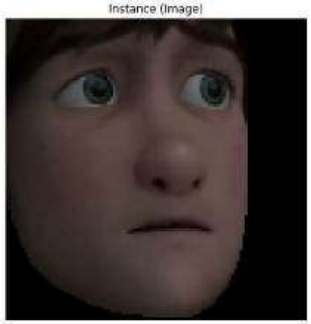


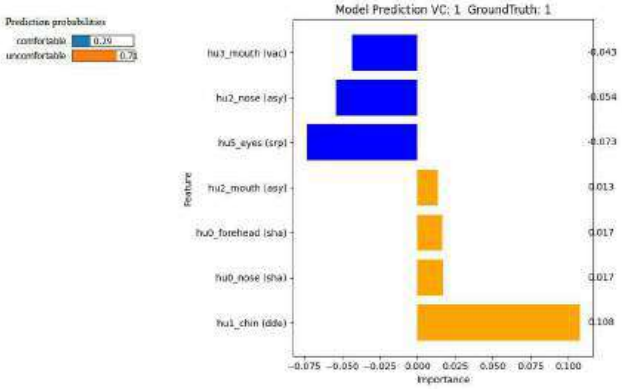
Prediction LIME: 1


Prediction probabilities
comfortable 0.60
uncomfortable 0.40
Model Prediction VC: 0 GroundTruth: 1
hu0_nose (sha) -0.025
hu2_mouth (asy) -0.037
hu1_chin (dde) -0.064
hu1_forehead (dde) 0.011
hu1_nose (dde) 0.017
hu3_mouth (vac) 0.022
hu1_eyes (dde) 0.090
Feature
Importance

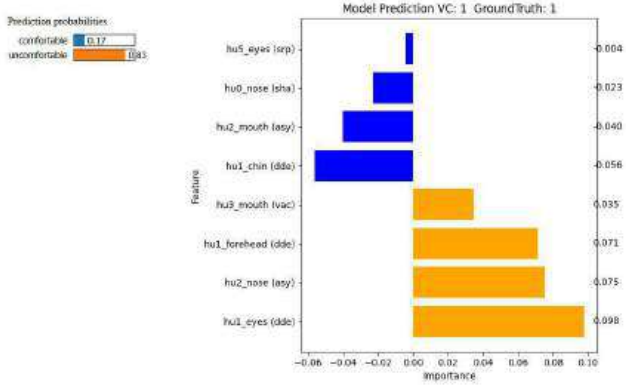

Instance (Image)
Prediction LIME: 0


Prediction probabilities
comfortable 0.56
uncomfortable 0.44
Model Prediction VC: 0 GroundTruth: 0
hu0_mouth (sha) -0.022
hu2_nose (asy) -0.027
hu1_eyes (dde) -0.034
hu0_forehead (sha) -0.160
hu0_nose (sha) 0.028
hu2_mouth (asy) 0.029
hu5_eyes (srp) 0.034
hu1_chin (dde) 0.107
Feature
Importance

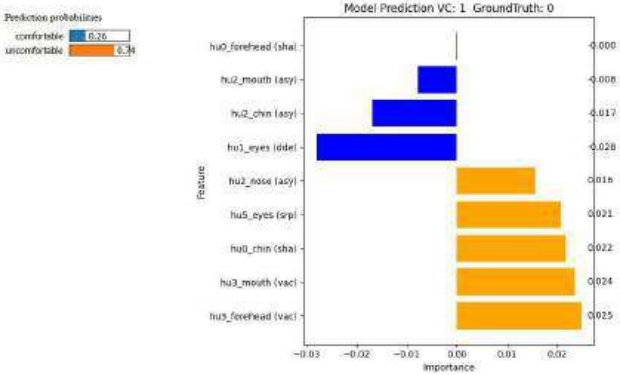

Instance (Image)
Prediction LIME: 0

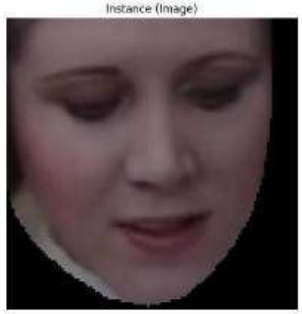

Prediction probabilities
comfortable 0.29
uncomfortable 0.71
Model Prediction VC: 1 GroundTruth: 1
hu3_mouth (vac) -0.043
hu2_nose (asy) -0.054
hu5_eyes (srp) -0.073
hu2_mouth (asy) 0.013
hu0_forehead (sha) 0.017
hu0_nose (sha) 0.017
hu1_chin (dde) 0.108
Feature
Importance

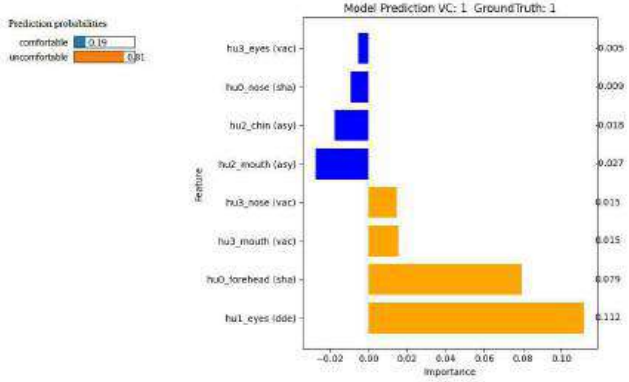

Instance (Image)
Prediction LIME: 1


Prediction probabilities
comfortable 0.17
uncomfortable 0.83
Model Prediction VC: 1 GroundTruth: 1
hu5_eyes (srp) -0.004
hu0_nose (sha) -0.023
hu2_mouth (asy) -0.040
hu1_chin (dde) -0.056
hu3_mouth (vac) 0.035
hu1_forehead (dde) 0.071
hu2_nose (asy) 0.075
hu1_eyes (dde) 0.098
Feature
Importance

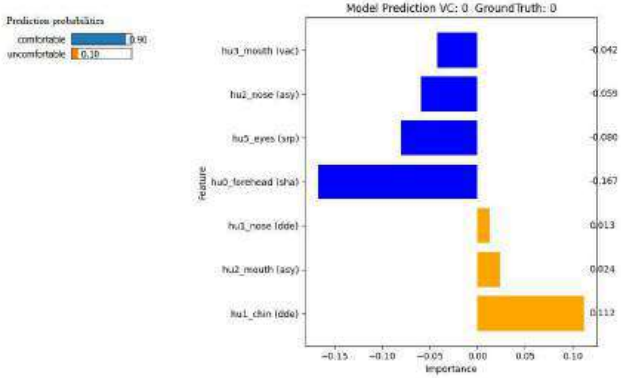

Instance (Image)
Prediction LIME: 1

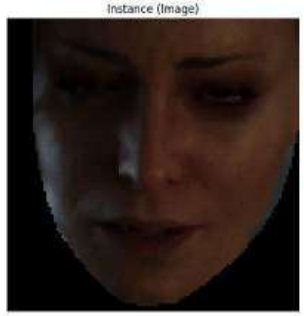

Prediction probabilities
comfortable 0.26
uncomfortable 0.74
Model Prediction VC: 1 GroundTruth: 0
hu0_forehead (sha) -0.000
hu2_mouth (asy) -0.008
hu2_chin (asy) -0.017
hu1_eyes (dde) -0.028
hu2_nose (asy) 0.016
hu5_eyes (srp) 0.021
hu0_chin (sha) 0.022
hu3_mouth (vac) 0.024
hu3_forehead (vac) 0.025
Feature
Importance
−0.03 −0.02 −0.01 0.00 0.01 0.02

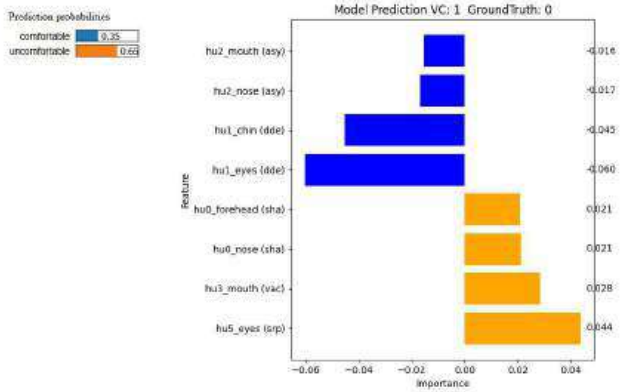

Instance (Image)
Prediction LIME: 1


Prediction probabilities
comfortable 0.19
uncomfortable 0.81
Model Prediction VC: 1 GroundTruth: 1
hu3_eyes (vac) -0.005
hu0_nose (sha) -0.009
hu2_chin (asy) -0.018
hu2_mouth (asy) -0.027
hu3_nose (vac) 0.015
hu3_mouth (vac) 0.015
hu0_forehead (sha) 0.079
hu1_eyes (dde) 0.112
Feature
Importance
−0.02 0.00 0.02 0.04 0.06 0.08 0.10

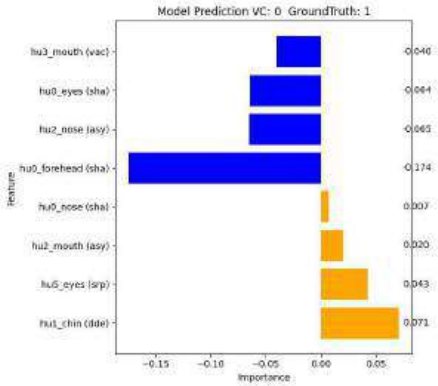

Instance (Image)
Prediction LIME: 1


Prediction probabilities
comfortable 0.90
uncomfortable 0.10
Model Prediction VC: 0 GroundTruth: 0
hu3_mouth (vac) -0.042
hu2_nose (asy) -0.059
hu5_eyes (srp) -0.080
hu0_forehead (sha) -0.167
hu3_nose (dde) 0.013
hu2_mouth (asy) 0.024
hu1_chin (dde) 0.112
Feature
Importance
−0.15 −0.10 −0.05 0.00 0.05 0.10

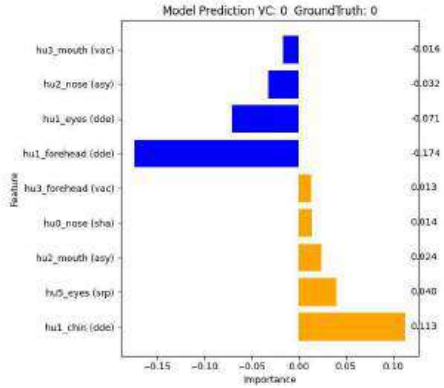

Instance (Image)
Prediction LIME: 0

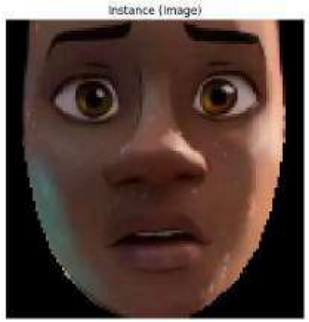

Prediction probabilities
comfortable 0.35
uncomfortable 0.65
Model Prediction VC: 1 GroundTruth: 0
hu2_mouth (asy) -0.016
hu2_nose (asy) -0.017
hu1_chin (dde) -0.045
hu1_eyes (dde) -0.060
hu0_forehead (sha) 0.021
hu0_nose (sha) 0.021
hu3_mouth (vac) 0.028
hu5_eyes (srp) 0.044
Feature
Importance
−0.06 −0.04 −0.02 0.00 0.02 0.04


Instance (Image)
Prediction LIME: 1

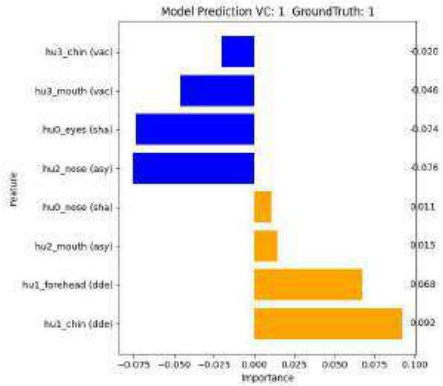


Prediction LIME: 0

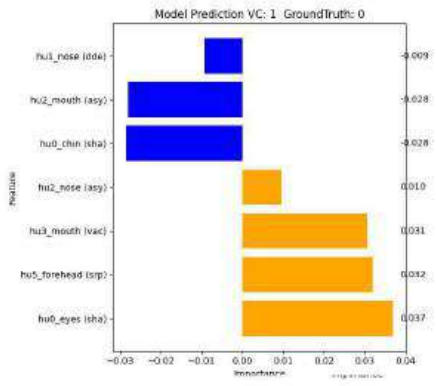




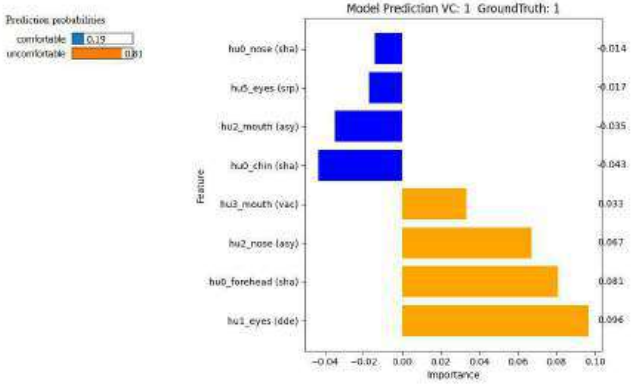
Prediction LIME: 0



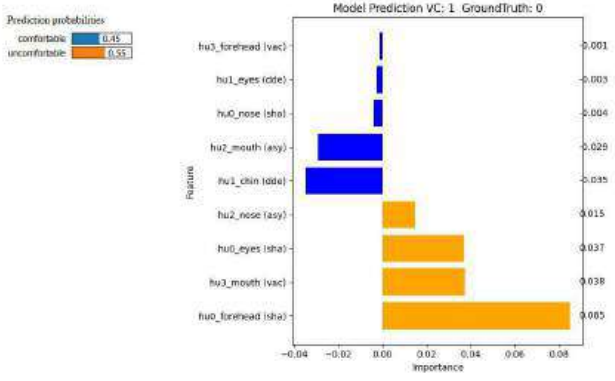
Prediction LIME: 1

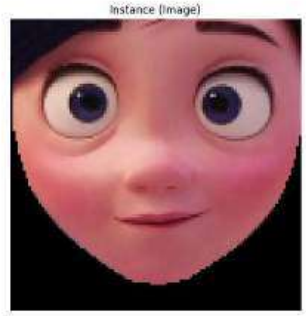


Prediction LIME: 1

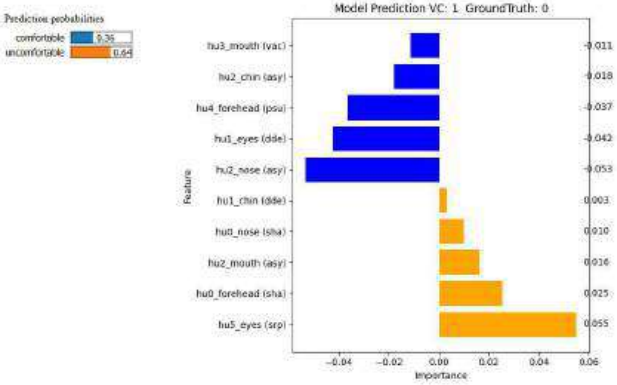


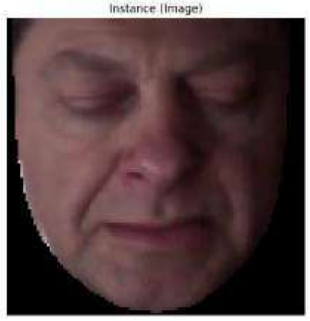
Prediction LIME: 1



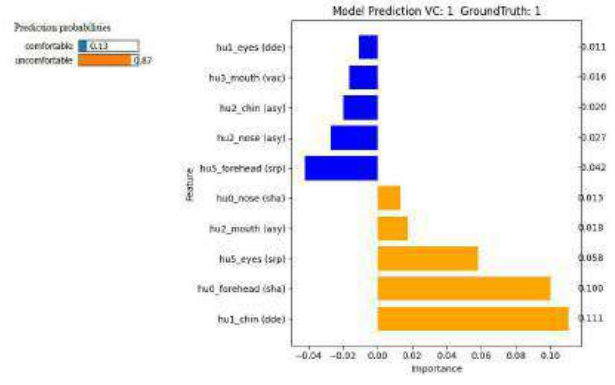


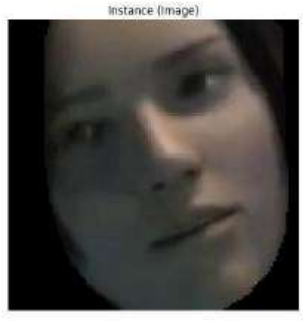
Prediction LIME: 1



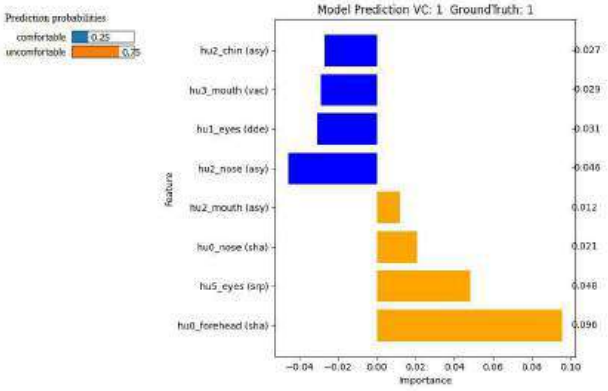


Prediction LIME: 1



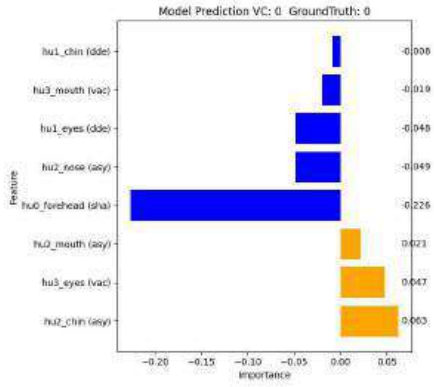


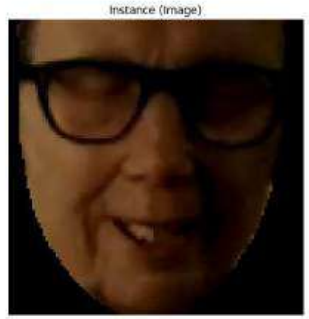
Prediction LIME: 1

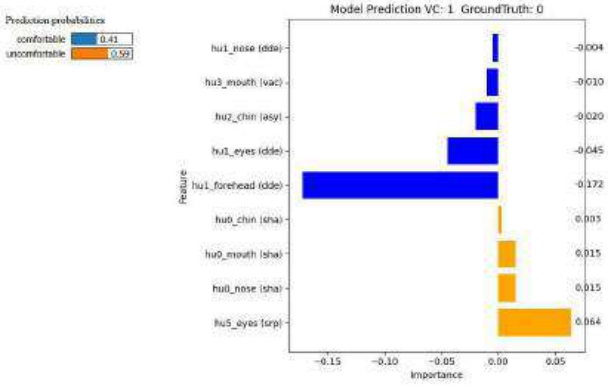

Prediction probabilities
comfortable 0.25
uncomfortable 0.75
Model Prediction VC: 1 GroundTruth: 1
hu2_chin (asy) -0.027
hu3_mouth (vac) -0.029
hu1_eyes (dde) -0.031
hu2_nose (asy) -0.046
hu2_mouth (asy) 0.012
hu0_nose (sha) 0.021
hu5_eyes (srp) 0.048
hu0_forehead (sha) 0.096
Feature
Importance


Instance (Image)
Prediction LIME: 1


Prediction probabilities
comfortable 0.78
uncomfortable 0.22

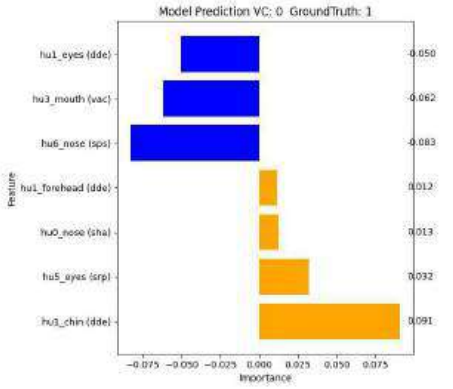

Model Prediction VC: 0 GroundTruth: 0
hu1_chin (dde) -0.008
hu3_mouth (vac) -0.019
hu1_eyes (dde) -0.048
hu2_nose (asy) -0.049
hu0_forehead (sha) -0.226
hu2_mouth (asy) 0.021
hu3_eyes (vac) 0.047
hu2_chin (asy) 0.063
Feature
Importance

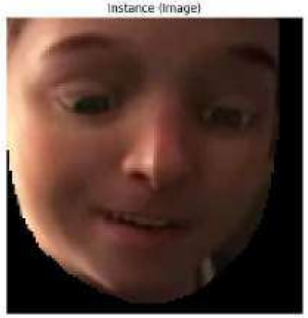

Instance (Image)
Prediction LIME: 0


Prediction probabilities
comfortable 0.41
uncomfortable 0.59
Model Prediction VC: 1 GroundTruth: 0
hu1_nose (dde) -0.004
hu3_mouth (vac) -0.010
hu2_chin (asy) -0.020
hu1_eyes (dde) -0.045
hu1_forehead (dde) -0.172
hu0_chin (sha) 0.003
hu0_mouth (sha) 0.015
hu0_nose (sha) 0.015
hu5_eyes (srp) 0.064
Feature
Importance

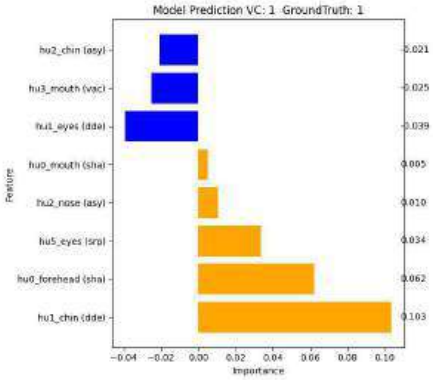

Instance (Image)
Prediction LIME: 1


Prediction probabilities
comfortable 0.80
uncomfortable 0.20

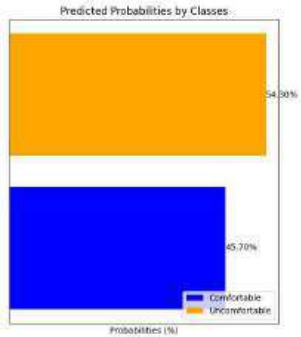

Model Prediction VC: 0 GroundTruth: 1
hu1_eyes (dde) -0.050
hu3_mouth (vac) -0.062
hu6_nose (sps) -0.083
hu1_forehead (dde) 0.012
hu0_nose (sha) 0.013
hu5_eyes (srp) 0.032
hu3_chin (dde) 0.091
Feature
Importance

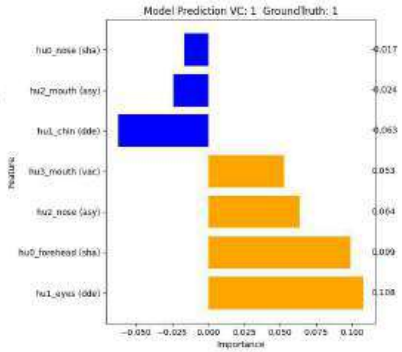

Instance (Image)
Prediction LIME: 0


Prediction probabilities
comfortable 0.18
uncomfortable 0.82


Model Prediction VC: 1 GroundTruth: 1
hu2_chin (asy) 0.021
hu3_mouth (vac) 0.025
hu1_eyes (dde) 0.039
hu0_mouth (sha) 0.005
hu2_nose (asy) 0.010
hu5_eyes (srp) 0.034
hu0_forehead (sha) 0.062
hu1_chin (dde) 0.103
Feature
Importance

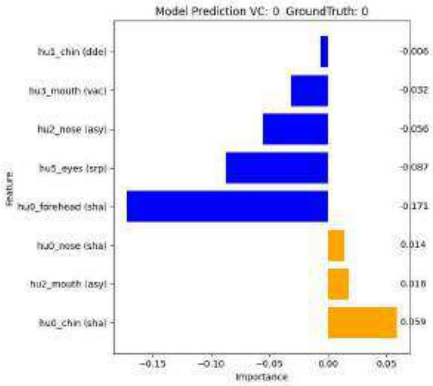

Instance (Image)
Prediction LIME: 1


Predicted Probabilities by Classes
54.30%
45.70%
Comfortable
Uncomfortable
Probabilities (%)


Model Prediction VC: 1 GroundTruth: 1
hu0_nose (sha) -0.017
hu2_mouth (asy) -0.024
hu1_chin (dde) -0.063
hu3_mouth (vac) 0.053
hu2_nose (asy) 0.064
hu0_forehead (sha) 0.099
hu1_eyes (dde) 0.108
Feature
Importance

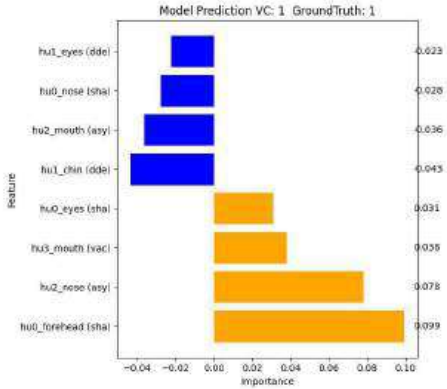

Instance (Image)
Prediction LIME: 1


Prediction probabilities
comfortable 0.97
uncomfortable 0.03

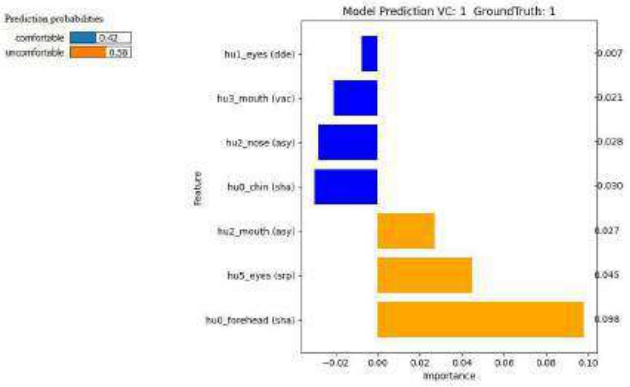

Model Prediction VC: 0 GroundTruth: 0
hu1_chin (dde) -0.006
hu3_mouth (vac) -0.032
hu2_nose (asy) -0.056
hu5_eyes (srp) -0.087
hu0_forehead (sha) -0.171
hu0_nose (sha) 0.014
hu2_mouth (asy) 0.018
hu0_chin (sha) 0.059
Feature
Importance


Instance (Image)
Prediction LIME: 0

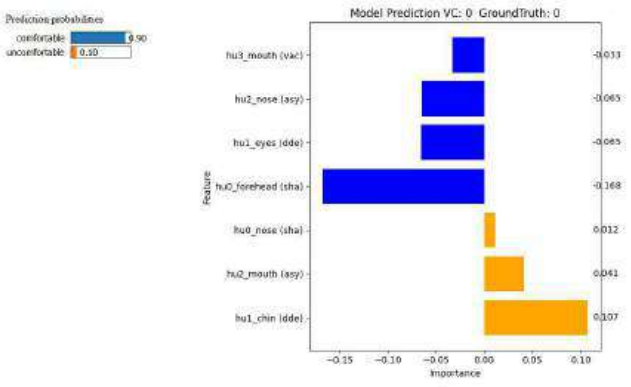

Prediction probabilities
comfortable 0.28
uncomfortable 0.72

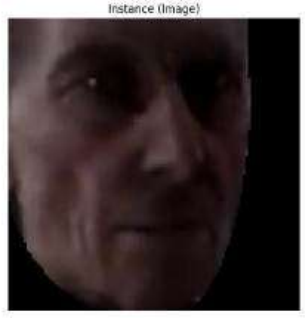

Model Prediction VC: 1 GroundTruth: 1
hu1_eyes (dde) 0.023
hu0_nose (sha) 0.028
hu2_mouth (asy) 0.036
hu1_chin (dde) 0.043
hu0_eyes (sha) 0.031
hu3_mouth (vac) 0.058
hu2_nose (asy) 0.078
hu0_forehead (sha) 0.099
Feature
Importance

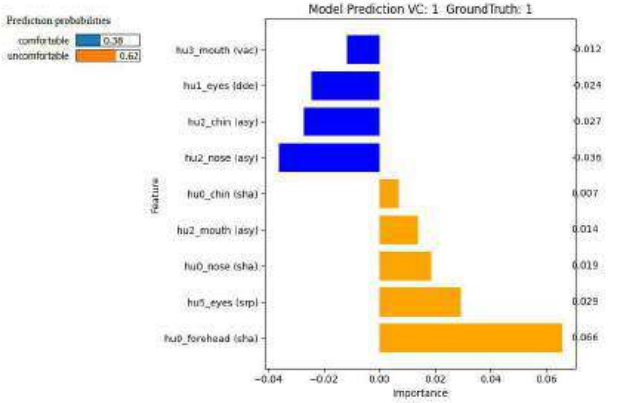

Instance (Image)
Prediction LIME: 1


Prediction probabilities
comfortable 0.42
uncomfortable 0.58
Model Prediction VC: 1 GroundTruth: 1
hu1_eyes (dde)
hu3_mouth (vac)
hu2_nose (asy)
hu0_chin (sha)
hu2_mouth (asy)
hu5_eyes (srp)
hu0_forehead (sha)
0.007
0.021
0.028
0.030
0.027
0.045
0.098
Feature
Importance

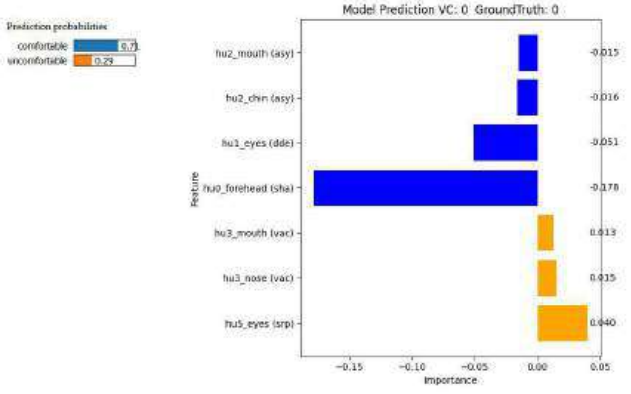

Instance (Image)
Prediction LIME: 1

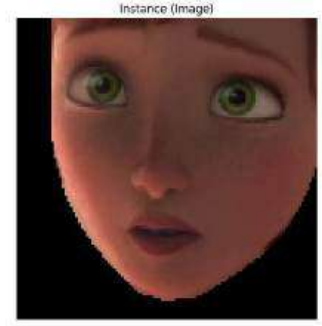

Prediction probabilities
comfortable 0.90
uncomfortable 0.10
Model Prediction VC: 0 GroundTruth: 0
hu3_mouth (vac)
hu2_nose (asy)
hu1_eyes (dde)
hu0_forehead (sha)
hu0_nose (sha)
hu2_mouth (asy)
hu1_chin (dde)
-0.033
-0.065
-0.065
-0.168
0.012
0.041
0.107
Feature
Importance


Instance (Image)
Prediction LIME: 0

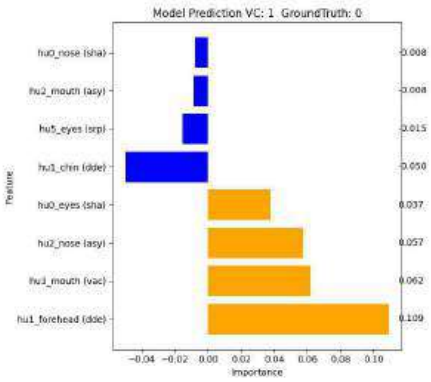

Prediction probabilities
comfortable 0.38
uncomfortable 0.62
Model Prediction VC: 1 GroundTruth: 1
hu3_mouth (vac)
hu1_eyes (dde)
hu2_chin (asy)
hu2_nose (asy)
hu0_chin (sha)
hu2_mouth (asy)
hu0_nose (sha)
hu5_eyes (srp)
hu0_forehead (sha)
-0.012
-0.024
-0.027
-0.036
0.007
0.014
0.019
0.029
0.066
Feature
Importance

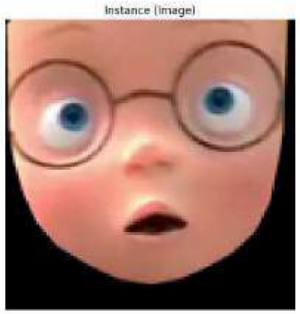

Instance (Image)
Prediction LIME: 1


Prediction probabilities
comfortable 0.71
uncomfortable 0.29
Model Prediction VC: 0 GroundTruth: 0
hu2_mouth (asy)
hu2_chin (asy)
hu1_eyes (dde)
hu0_forehead (sha)
hu3_mouth (vac)
hu3_nose (vac)
hu5_eyes (srp)
-0.015
-0.016
-0.051
-0.178
0.013
0.015
0.040
Feature
Importance

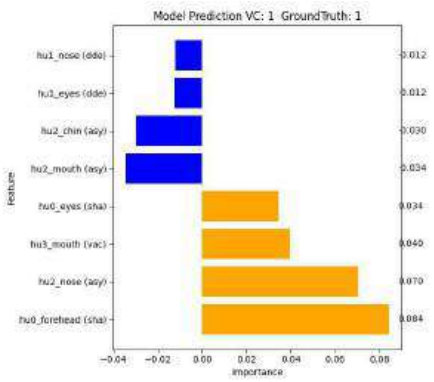

Instance (Image)
Prediction LIME: 0


Prediction probabilities
comfortable
uncomfortable
0.07
0.93


Model Prediction VC: 1 GroundTruth: 0
hu0_nose (sha)
hu2_mouth (asy)
hu5_eyes (srp)
hu1_chin (dde)
hu0_eyes (sha)
hu2_nose (asy)
hu3_mouth (vac)
hu1_forehead (dde)
0.008
0.008
0.015
0.050
0.037
0.057
0.062
0.109
Feature
Importance

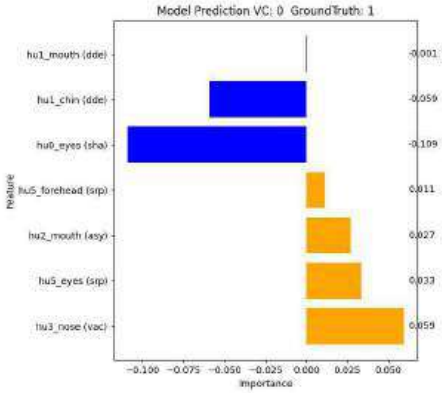

Instance (Image)
Prediction LIME: 1

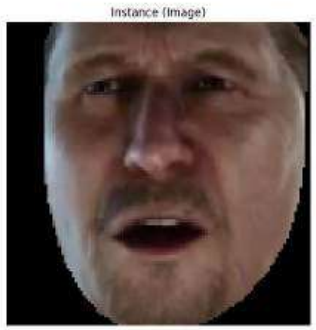

Prediction probabilities
comfortable
uncomfortable
0.24
0.76


Model Prediction VC: 1 GroundTruth: 1
hu1_nose (dde)
hu1_eyes (dde)
hu2_chin (asy)
hu2_mouth (asy)
hu0_eyes (sha)
hu3_mouth (vac)
hu2_nose (asy)
hu0_forehead (sha)
0.012
0.012
0.030
0.034
0.034
0.049
0.070
0.084
Feature
Importance

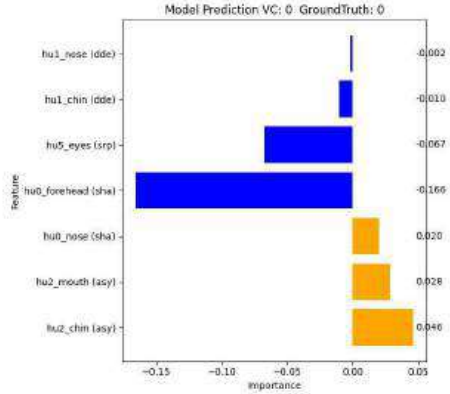

Instance (Image)
Prediction LIME: 1


Prediction probabilities
comfortable
uncomfortable
0.73
0.27

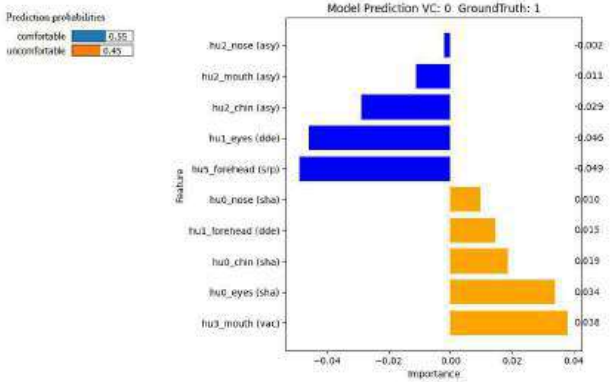

Model Prediction VC: 0 GroundTruth: 1
hu1_mouth (dde)
hu1_chin (dde)
hu0_eyes (sha)
hu5_forehead (srp)
hu2_mouth (asy)
hu5_eyes (srp)
hu3_nose (vac)
-0.001
-0.059
-0.109
0.011
0.027
0.033
0.059
Feature
Importance


Instance (Image)
Prediction LIME: 0

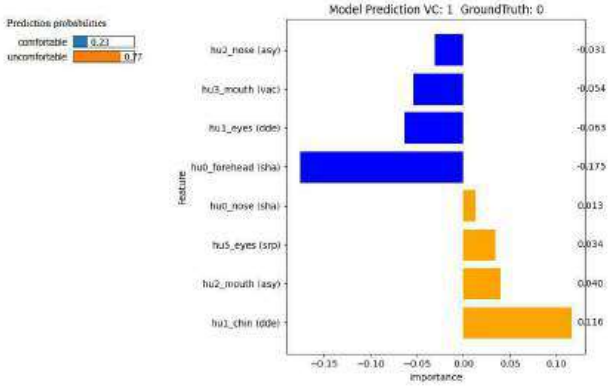

Prediction probabilities
comfortable
uncomfortable
0.65
0.35


Model Prediction VC: 0 GroundTruth: 0
hu1_nose (dde)
hu1_chin (dde)
hu5_eyes (srp)
hu0_forehead (sha)
hu0_nose (sha)
hu2_mouth (asy)
hu2_chin (asy)
-0.002
-0.010
-0.067
-0.166
0.020
0.028
0.046
Feature
Importance

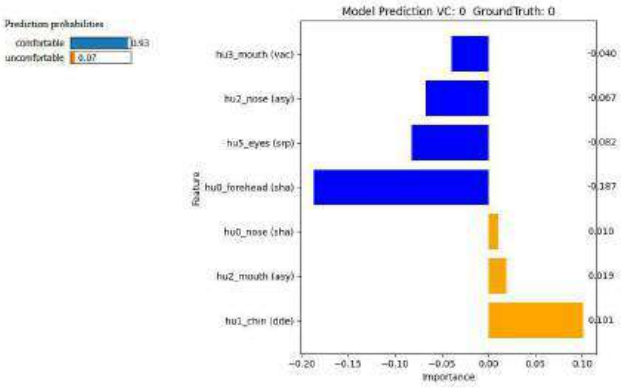

Instance (Image)
Prediction LIME: 0


Prediction probabilities
comfortable 0.55
uncomfortable 0.45
Model Prediction VC: 0 GroundTruth: 1
hu2_nose (asy) -0.002
hu2_mouth (asy) -0.011
hu2_chin (asy) -0.029
hu1_eyes (dde) -0.046
hu5_forehead (srp) -0.049
hu0_nose (sha) 0.010
hu1_forehead (dde) 0.015
hu0_chin (sha) 0.019
hu0_eyes (sha) 0.034
hu3_mouth (vac) 0.038
Feature
Importance

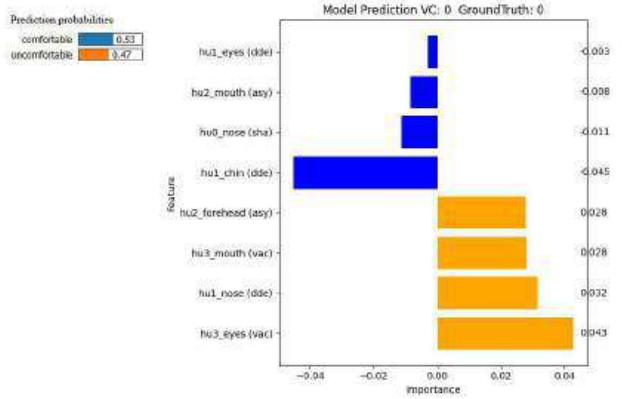

Instance (Image)
Prediction LIME: 0


Prediction probabilities
comfortable 0.23
uncomfortable 0.77
Model Prediction VC: 1 GroundTruth: 0
hu2_nose (asy) -0.031
hu3_mouth (vac) -0.054
hu1_eyes (dde) -0.063
hu0_forehead (sha) -0.175
hu0_nose (sha) 0.013
hu5_eyes (srp) 0.034
hu2_mouth (asy) 0.040
hu1_chin (dde) 0.116
Feature
Importance


Instance (Image)
Prediction LIME: 1


Prediction probabilities
comfortable 0.93
uncomfortable 0.07
Model Prediction VC: 0 GroundTruth: 0
hu3_mouth (vac) -0.040
hu2_nose (asy) -0.067
hu5_eyes (srp) -0.082
hu0_forehead (sha) -0.187
hu0_nose (sha) 0.010
hu2_mouth (asy) 0.019
hu1_chin (dde) 0.101
Feature
Importance


Instance (Image)
Prediction LIME: 0

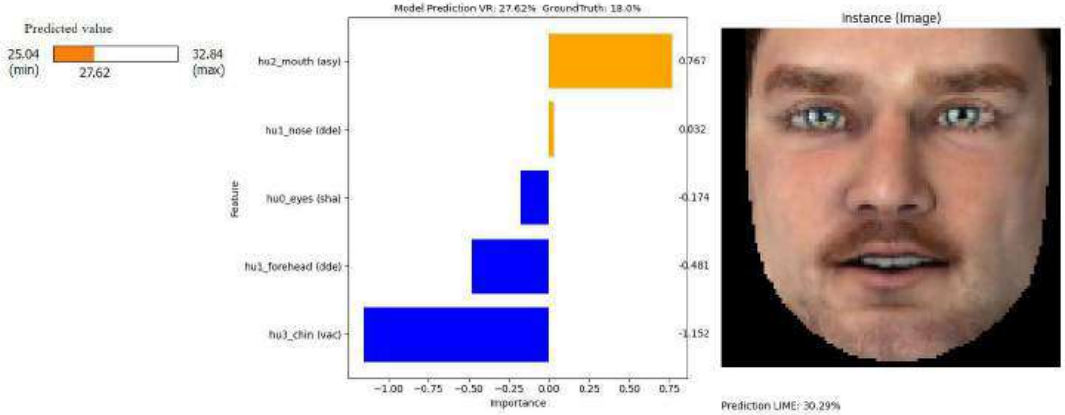

Prediction probabilities
comfortable 0.53
uncomfortable 0.47
Model Prediction VC: 0 GroundTruth: 0
hu1_eyes (dde) -0.003
hu2_mouth (asy) -0.008
hu0_nose (sha) -0.011
hu1_chin (dde) -0.045
hu2_forehead (asy) 0.028
hu3_mouth (vac) 0.028
hu1_nose (dde) 0.032
hu3_eyes (vac) 0.043
Feature
Importance

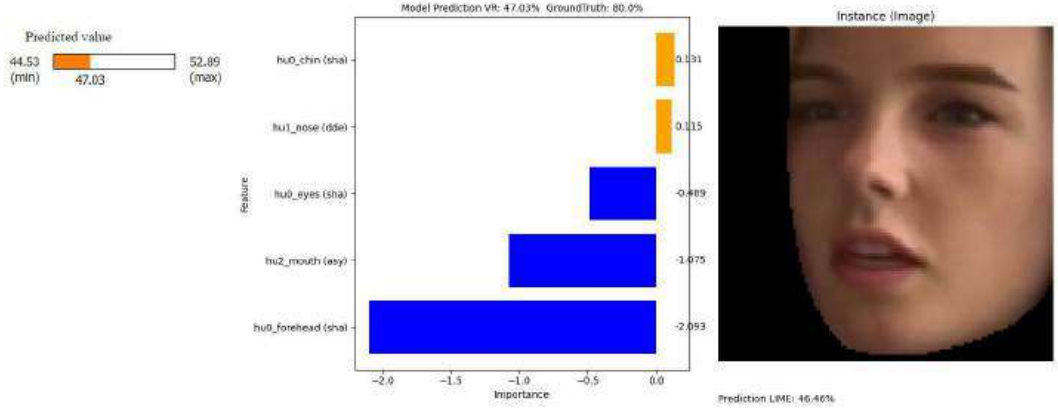

Instance (Image)
Prediction LIME: 0

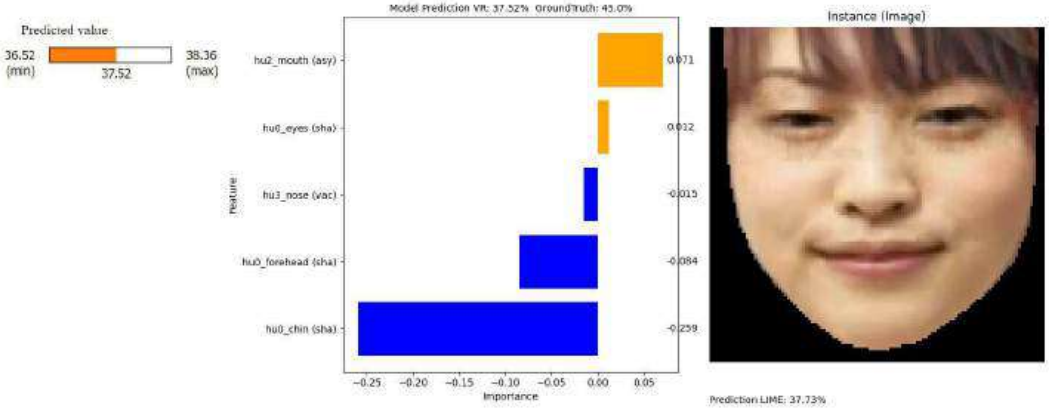
10.2.2 LIME result for VR model

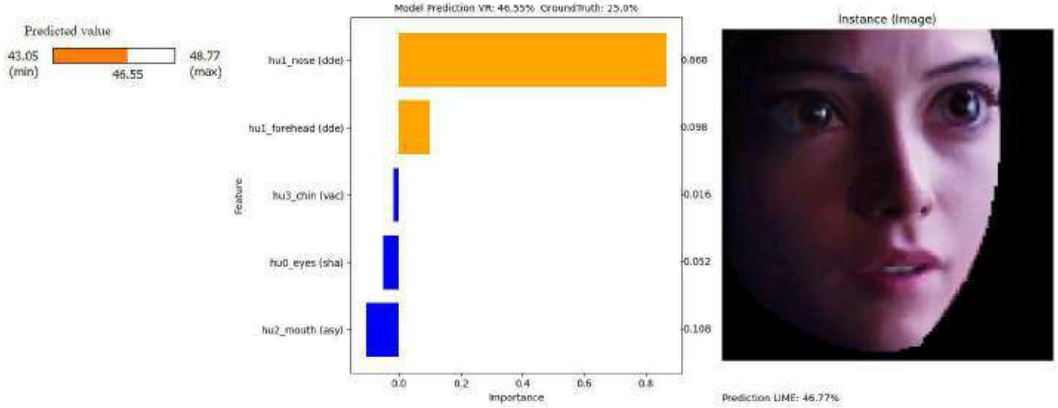
This section shows the result of the VR model using the characters from dataset GT2 (Section 4.2).

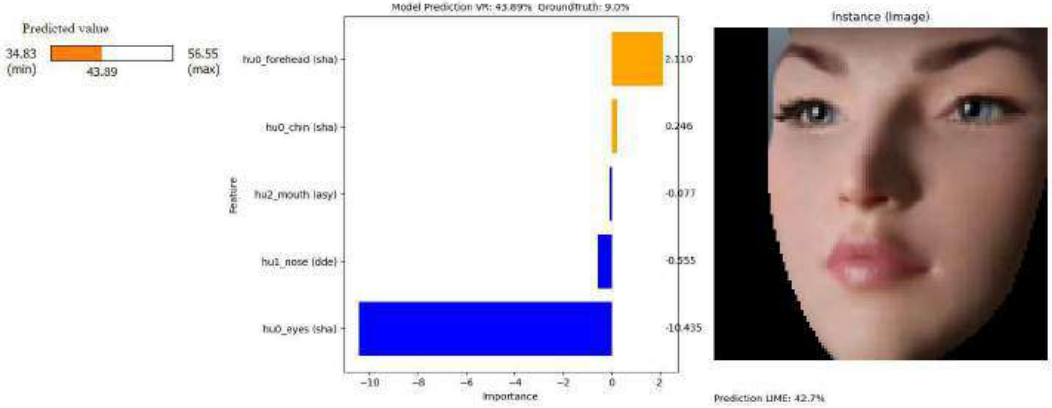
Each character instance (40) is shown in the following Figures. The Figures on the left show the comfort score. In the middle, a graph indicates the contribution of each feature to the model prediction. On the right side, the image of the character is shown.

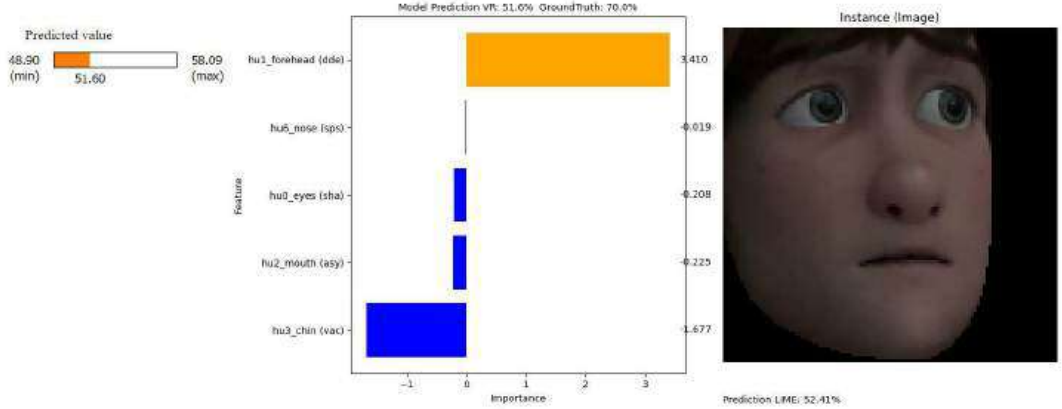

Predicted value
25.04
(min)
27.62
32.84
(max)
Model Prediction VR: 27.62% GroundTruth: 18.0%
hu2_mouth (asy)
hu1_nose (dde)
hu0_eyes (sha)
hu1_forehead (dde)
hu3_chin (vac)
0.767
0.032
-0.174
-0.481
-1.152
Feature
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75
Importance
Instance (Image)
Prediction LIME: 30.29%

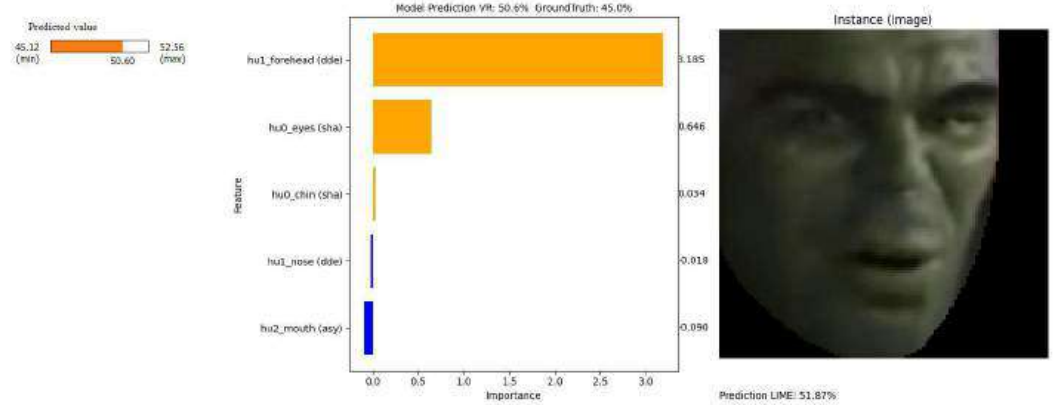

Predicted value
44.53
(min)
47.03
52.89
(max)
Model Prediction VR: 47.03% GroundTruth: 80.0%
hu0_chin (sha)
hu1_nose (dde)
hu0_eyes (sha)
hu2_mouth (asy)
hu0_forehead (sha)
0.131
0.115
-0.489
-1.075
-2.093
Feature
−2.0 −1.5 −1.0 −0.5 0.0
Importance
Instance (Image)
Prediction LIME: 46.46%

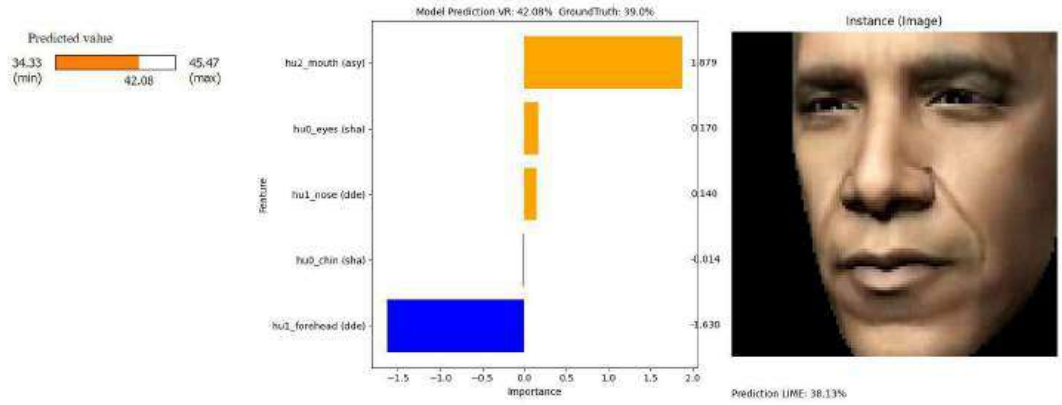

Predicted value
36.52
(min)
37.52
38.36
(max)
Model Prediction VR: 37.52% GroundTruth: 45.0%
hu2_mouth (asy)
hu0_eyes (sha)
hu3_nose (vac)
hu0_forehead (sha)
hu0_chin (sha)
0.071
0.012
-0.015
-0.084
-0.259
Feature
−0.25 −0.20 −0.15 −0.10 −0.05 0.00 0.05
Importance
Instance (Image)
Prediction LIME: 37.73%

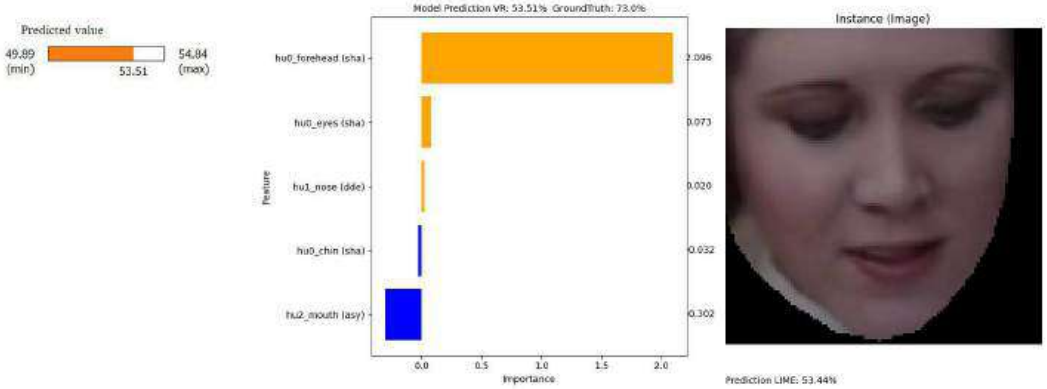

Predicted value
43.05
(min)
46.55
48.77
(max)
Model Prediction VR: 46.55% GroundTruth: 25.0%
hu1_nose (dde)
hu1_forehead (dde)
hu3_chin (vac)
hu0_eyes (sha)
hu2_mouth (asy)
0.868
0.098
-0.016
-0.052
-0.108
Feature
0.0 0.2 0.4 0.6 0.8
Importance
Instance (Image)
Prediction LIME: 46.77%

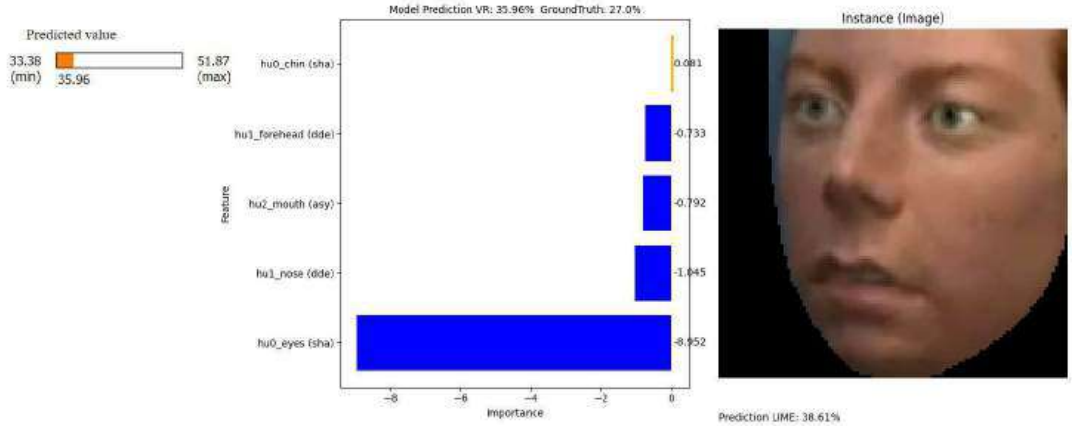

Predicted value
34.83
(min)
43.89
56.55
(max)
Model Prediction VR: 43.89% GroundTruth: 9.0%
hu0_forehead (sha)
hu0_chin (sha)
hu2_mouth (asy)
hu1_nose (dde)
hu0_eyes (sha)
2.110
0.246
-0.077
-0.555
-10.435
Feature
−10
−8
−6
−4
−2
0
2
Importance
Instance (Image)
Prediction LIME: 42.7%

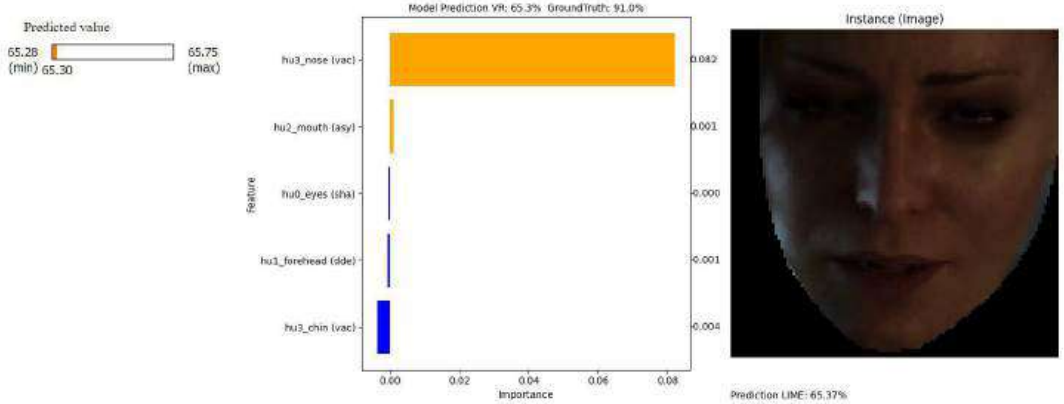

Predicted value
48.90
(min)
51.60
58.09
(max)
Model Prediction VR: 51.6% GroundTruth: 70.0%
hu1_forehead (dde)
hu6_nose (sps)
hu0_eyes (sha)
hu2_mouth (asy)
hu3_chin (vac)
3.410
-0.019
-0.208
-0.225
-1.677
Feature
−1
0
1
2
3
Importance
Instance (Image)
Prediction LIME: 52.41%

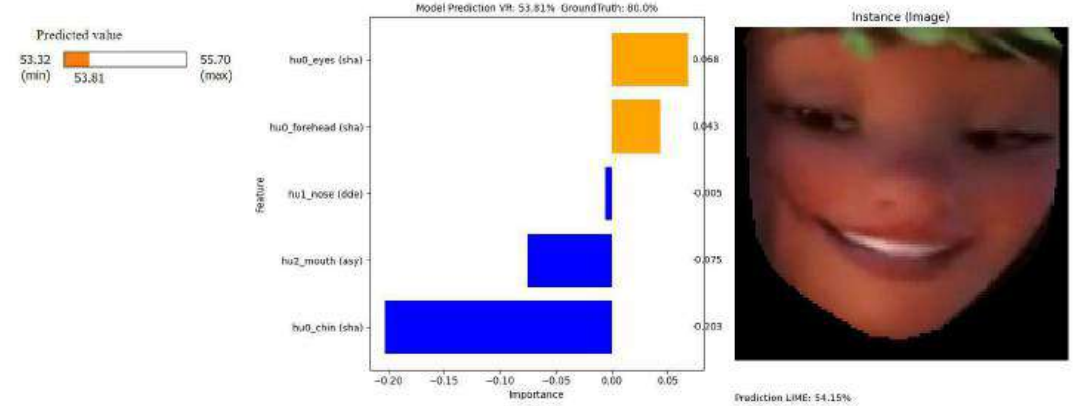

Predicted value
45.12
(min)
50.60
52.16
(max)
Model Prediction VR: 50.6% GroundTruth: 45.0%
hu1_forehead (dde)
hu0_eyes (sha)
hu0_chin (sha)
hu1_nose (dde)
hu2_mouth (asy)
3.185
0.646
0.034
-0.018
-0.090
Feature
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Importance
Instance (Image)
Prediction LIME: 51.87%

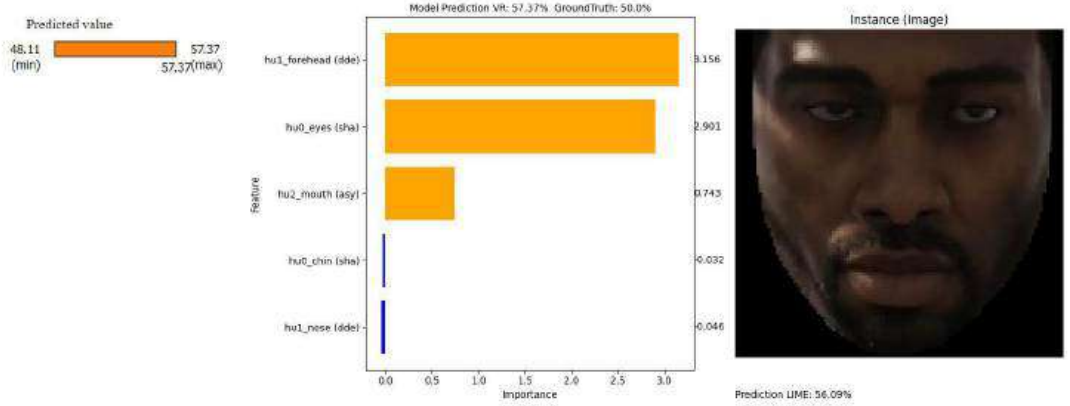

Predicted value
34.33
(min)
42.08
45.47
(max)
Model Prediction VR: 42.08% GroundTruth: 39.0%
hu2_mouth (asy)
hu0_eyes (sha)
hu1_nose (dde)
hu0_chin (sha)
hu1_forehead (dde)
1.879
0.170
0.140
-0.014
-1.630
Feature
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
Importance
Instance (Image)
Prediction LIME: 38.13%

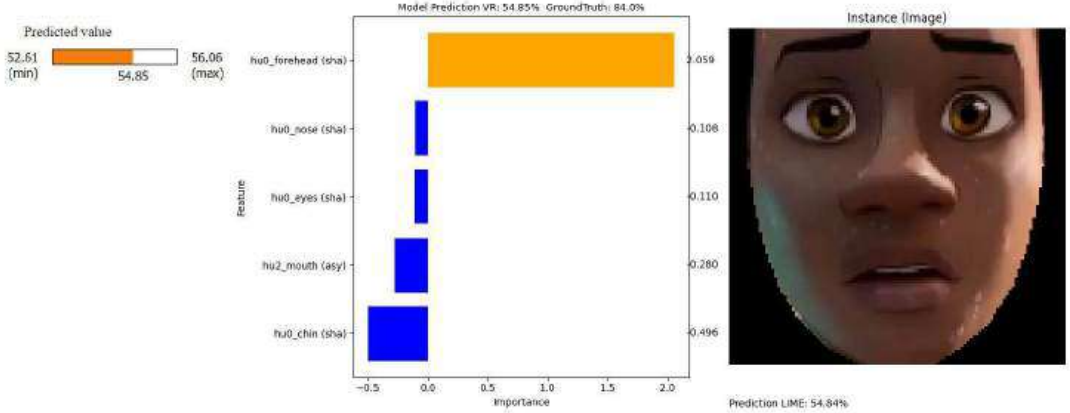

Predicted value
49.89
(min)
53.51
54.84
(max)
Model Prediction VR: 53.51% GroundTruth: 73.0%
hu0_forehead (sha)
hu0_eyes (sha)
hu1_nose (dde)
hu0_chin (sha)
hu2_mouth (asy)
2.096
0.073
0.020
-0.032
-0.302
Feature
Importance
0.0
0.5
1.0
1.5
2.0
Instance (Image)
Prediction LIME: 53.44%

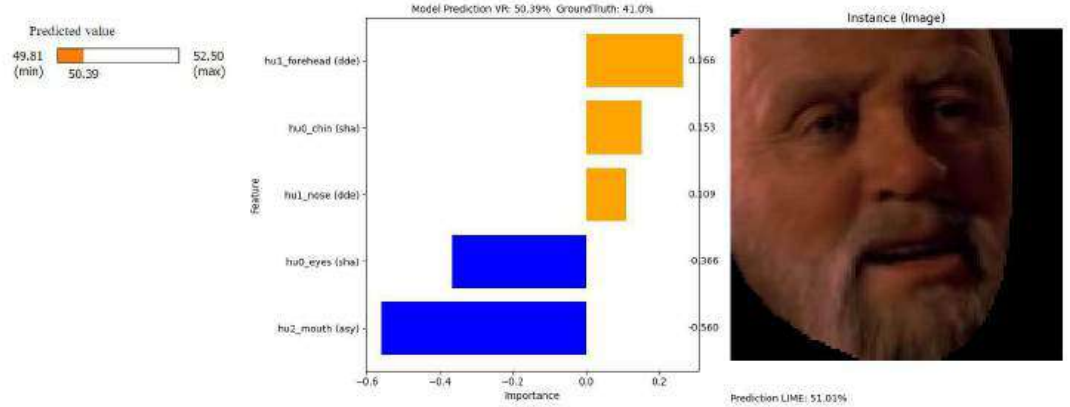

Predicted value
33.38
(min)
35.96
51.87
(max)
Model Prediction VR: 35.96% GroundTruth: 27.0%
hu0_chin (sha)
hu1_forehead (dde)
hu2_mouth (asy)
hu1_nose (dde)
hu0_eyes (sha)
0.081
-0.733
-0.792
-1.045
-8.952
Feature
Importance
−8
−6
−4
−2
0
Instance (Image)
Prediction LIME: 38.61%

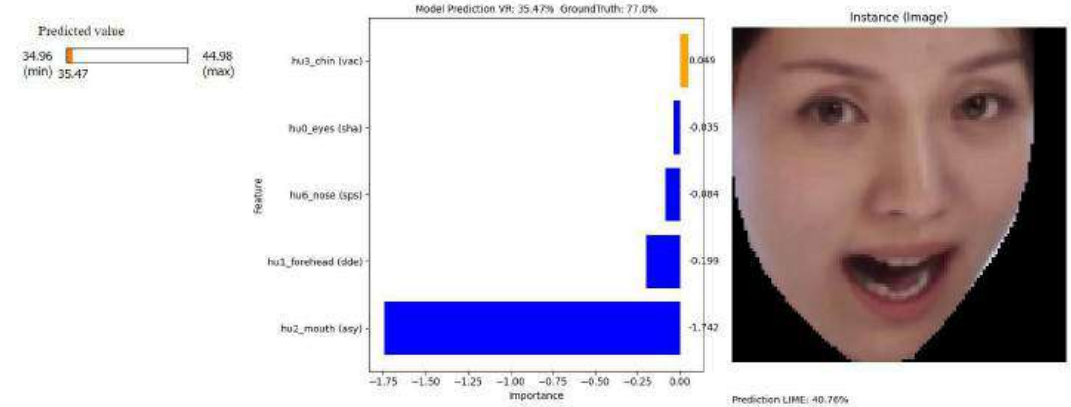

Predicted value
65.28
(min)
65.30
65.75
(max)
Model Prediction VR: 65.3% GroundTruth: 91.0%
hu3_nose (vac)
hu2_mouth (asy)
hu0_eyes (sha)
hu1_forehead (dde)
hu3_chin (vac)
0.042
0.001
-0.000
-0.001
-0.004
Feature
Importance
0.00
0.02
0.04
0.06
0.08
Instance (Image)
Prediction LIME: 65.37%

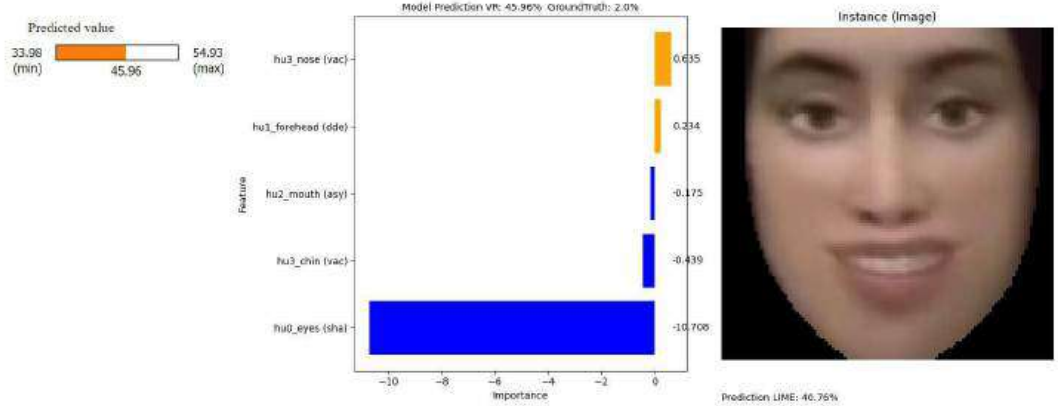

Predicted value
53.32
(min)
53.81
55.70
(max)
Model Prediction VR: 53.81% GroundTruth: 80.0%
hu0_eyes (sha)
hu0_forehead (sha)
hu1_nose (dde)
hu2_mouth (asy)
hu0_chin (sha)
0.068
0.043
-0.005
-0.075
-0.203
Feature
Importance
−0.20
−0.15
−0.10
−0.05
0.00
0.05
Instance (Image)
Prediction LIME: 54.15%

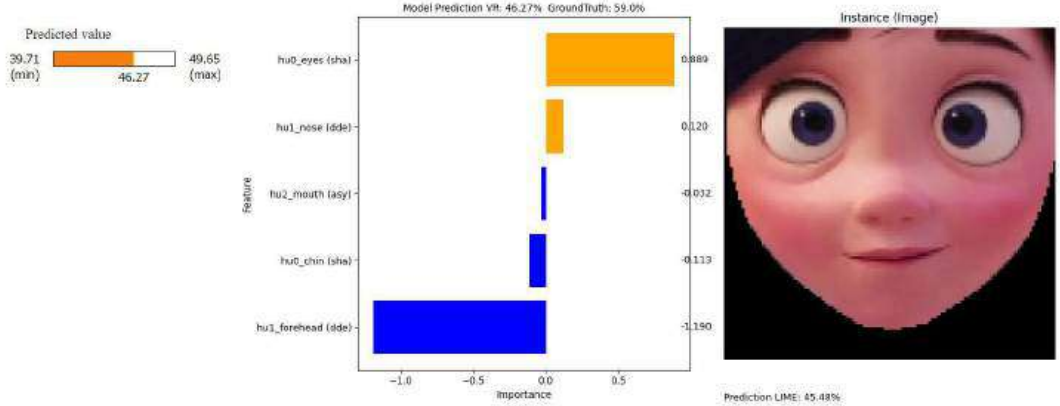

Predicted value
48.11
(min)
57.37
57.37(max)
Model Prediction VR: 57.37% GroundTruth: 50.0%
hu1_forehead (dde)
hu0_eyes (sha)
hu2_mouth (asy)
hu0_chin (sha)
hu1_nose (dde)
3.156
2.901
0.743
-0.032
-0.046
Feature
Importance
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Instance (Image)
Prediction LIME: 56.09%

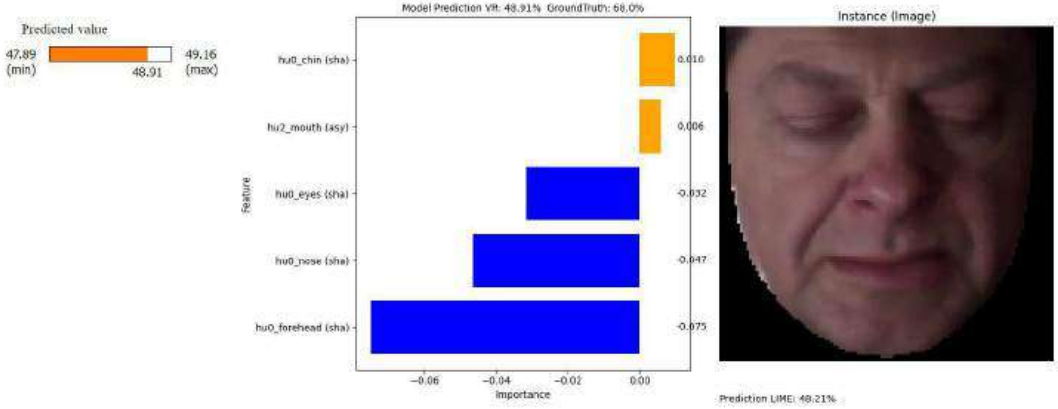

Predicted value
52.61
(min)
54.85
56.06
(max)
Model Prediction VR: 54.85% GroundTruth: 84.0%
hu0_forehead (sha)
hu0_nose (sha)
hu0_eyes (sha)
hu2_mouth (asy)
hu0_chin (sha)
2.059
-0.108
-0.110
-0.280
-0.496
Feature
Importance
−0.5
0.0
0.5
1.0
1.5
2.0
Instance (Image)
Prediction LIME: 54.84%

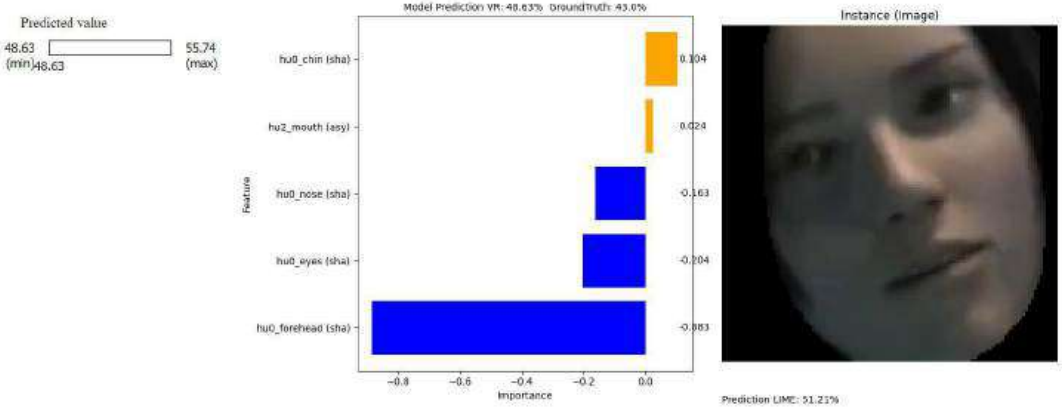

Predicted value
49.81
(min)
50.39
52.50
(max)
Model Prediction VR: 50.39% GroundTruth: 41.0%
hu1_forehead (dde)
hu0_chin (sha)
hu1_nose (dde)
hu0_eyes (sha)
hu2_mouth (asy)
0.166
0.153
0.109
-0.366
-0.560
Feature
Importance
−0.6
−0.4
−0.2
0.0
0.2
Instance (Image)
Prediction LIME: 51.01%

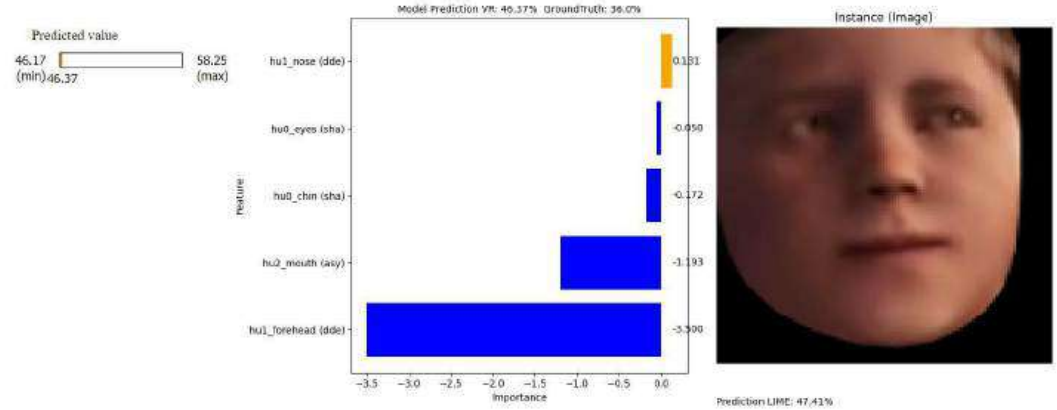

Predicted value
34.96
(min)
35.47
44.98
(max)
Model Prediction VR: 35.47% GroundTruth: 77.0%
hu3_chin (vac)
hu0_eyes (sha)
hu6_nose (sps)
hu1_forehead (dde)
hu2_mouth (asy)
0.049
-0.035
-0.084
-0.199
-1.742
Feature
Importance
−1.75
−1.50
−1.25
−1.00
−0.75
−0.50
−0.25
0.00
Instance (Image)
Prediction LIME: 40.76%

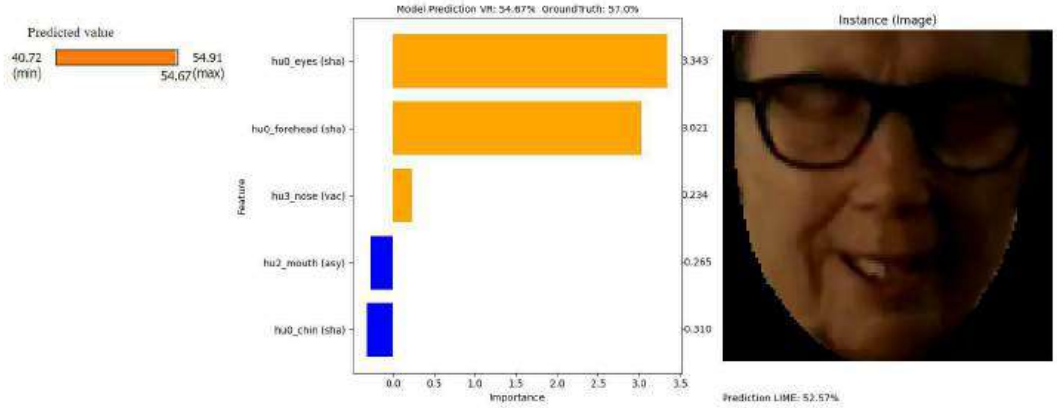

Predicted value
33.98
(min)
45.96
54.93
(max)
Model Prediction VR: 45.96% GroundTruth: 2.0%
hu3_nose (vac)
hu1_forehead (dde)
hu2_mouth (asy)
hu3_chin (vac)
hu0_eyes (sha)
0.635
0.234
-0.175
-0.439
-10.706
Feature
−10
−8
−6
−4
−2
0
Importance
Instance (Image)
Prediction LIME: 46.76%

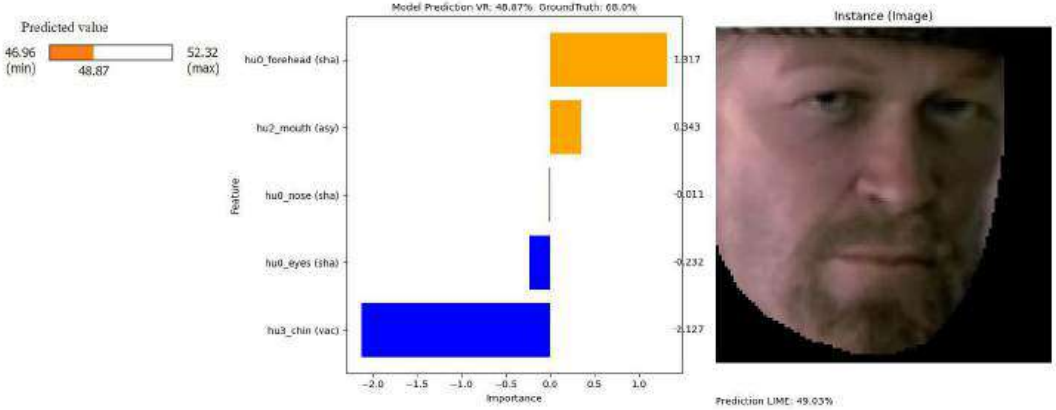

Predicted value
39.71
(min)
46.27
49.65
(max)
Model Prediction VR: 46.27% GroundTruth: 59.0%
hu0_eyes (sha)
hu1_nose (dde)
hu2_mouth (asy)
hu0_chin (sha)
hu1_forehead (dde)
0.889
0.120
-0.032
-0.113
-1.190
Feature
−1.0
−0.5
0.0
0.5
Importance
Instance (Image)
Prediction LIME: 45.48%

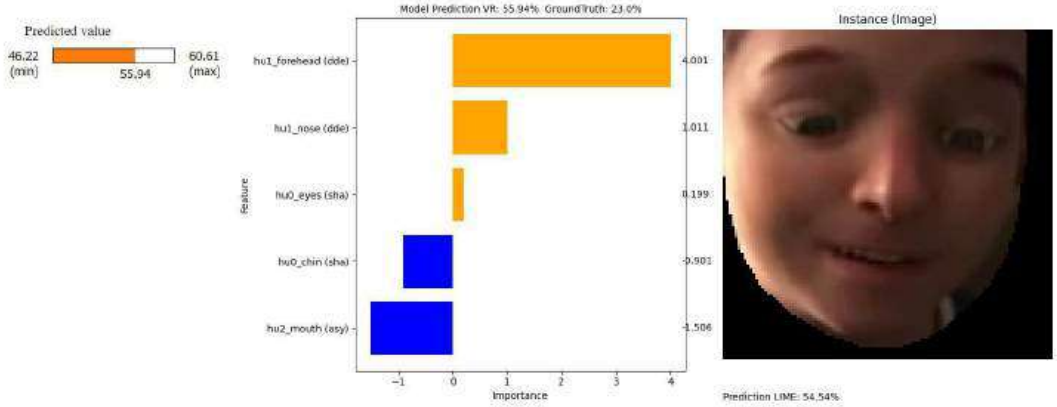

Predicted value
47.89
(min)
48.91
49.16
(max)
Model Prediction VR: 48.91% GroundTruth: 68.0%
hu0_chin (sha)
hu2_mouth (asy)
hu0_eyes (sha)
hu0_nose (sha)
hu0_forehead (sha)
0.010
0.006
-0.032
-0.047
-0.075
Feature
−0.06
−0.04
−0.02
0.00
Importance
Instance (Image)
Prediction LIME: 48.21%

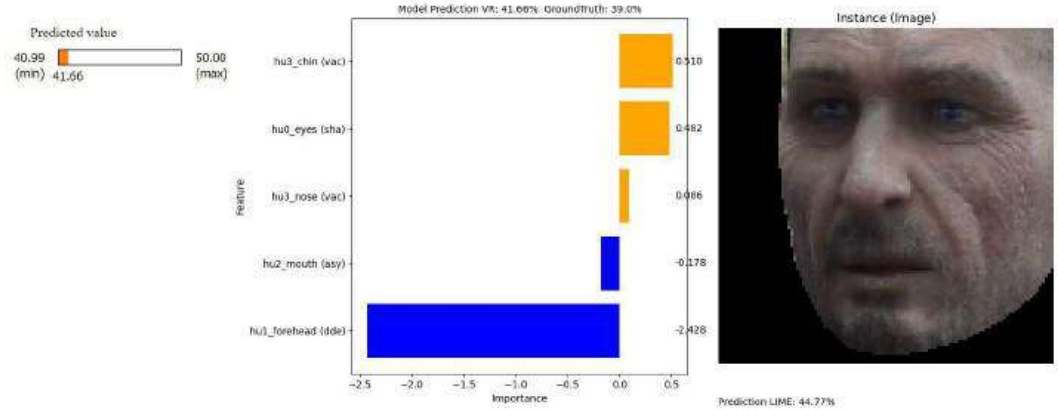

Predicted value
48.63
(min)
48.63
55.74
(max)
Model Prediction VR: 48.63% GroundTruth: 43.0%
hu0_chin (sha)
hu2_mouth (asy)
hu0_nose (sha)
hu0_eyes (sha)
hu0_forehead (sha)
0.104
0.024
-0.163
-0.204
-0.863
Feature
−0.8
−0.6
−0.4
−0.2
0.0
Importance
Instance (Image)
Prediction LIME: 51.21%

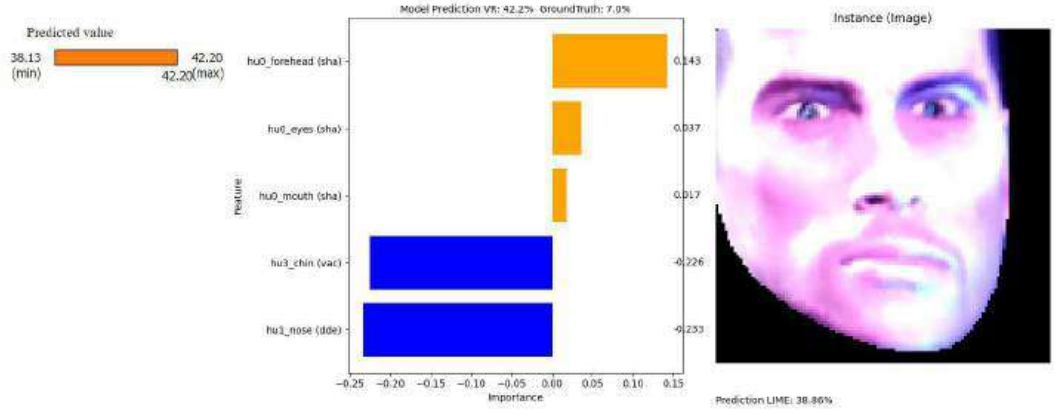

Predicted value
46.17 (min)
46.37
58.25 (max)
Model Prediction VR: 46.37% GroundTruth: 36.0%
hu1_nose (dde)
hu0_eyes (sha)
hu0_chin (sha)
hu2_mouth (asy)
hu1_forehead (dde)
0.131
-0.050
-0.172
-1.193
-3.500
Feature
Importance
Instance (Image)
Prediction LIME: 47.41%

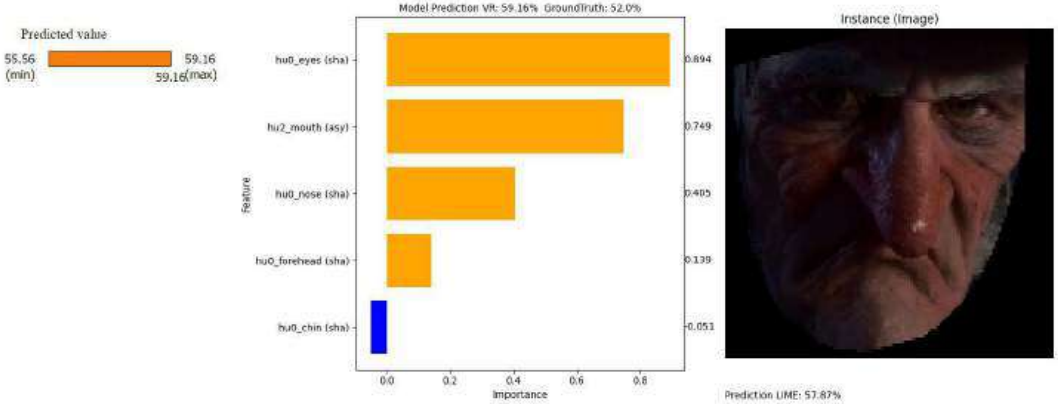

Predicted value
40.72 (min)
54,67
54.91 (max)
Model Prediction VR: 54.67% GroundTruth: 57.0%
hu0_eyes (sha)
hu0_forehead (sha)
hu3_nose (vac)
hu2_mouth (asy)
hu0_chin (sha)
3.343
3.021
0.234
-0.265
-0.310
Feature
Importance
Instance (Image)
Prediction LIME: 52.57%

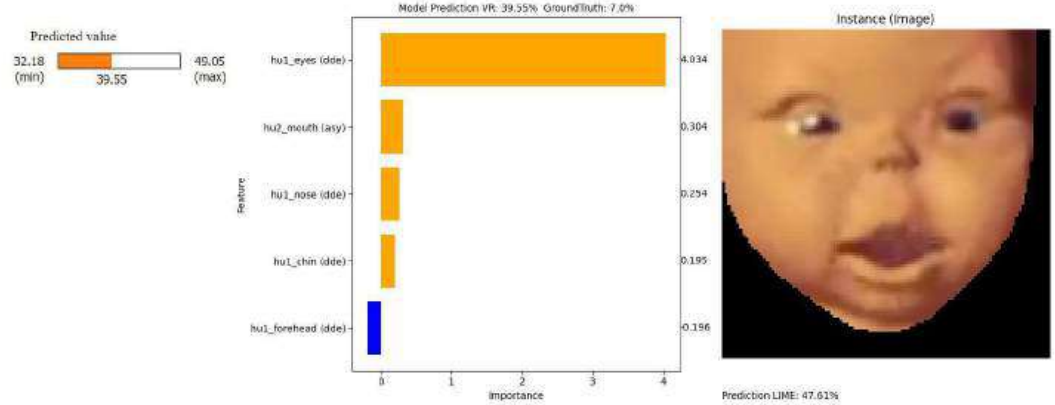

Predicted value
46.96 (min)
48.87
52.32 (max)
Model Prediction VR: 48.87% GroundTruth: 68.0%
hu0_forehead (sha)
hu2_mouth (asy)
hu0_nose (sha)
hu0_eyes (sha)
hu3_chin (vac)
1.317
0.343
-0.011
-0.232
-2.127
Feature
Importance
Instance (Image)
Prediction LIME: 49.03%

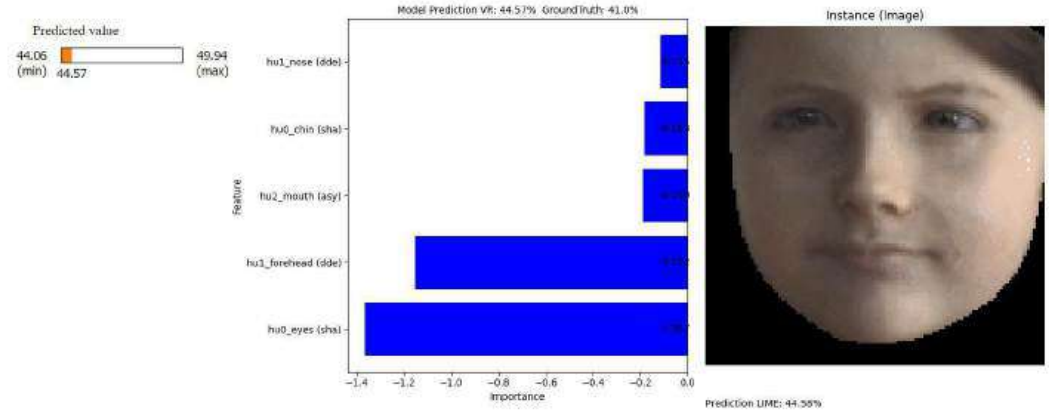

Predicted value
46.22 (min)
55.94
60.61 (max)
Model Prediction VR: 55.94% GroundTruth: 23.0%
hu1_forehead (dde)
hu1_nose (dde)
hu0_eyes (sha)
hu0_chin (sha)
hu2_mouth (asy)
4.001
1.011
0.199
-0.901
-1.506
Feature
Importance
Instance (Image)
Prediction LIME: 54.54%

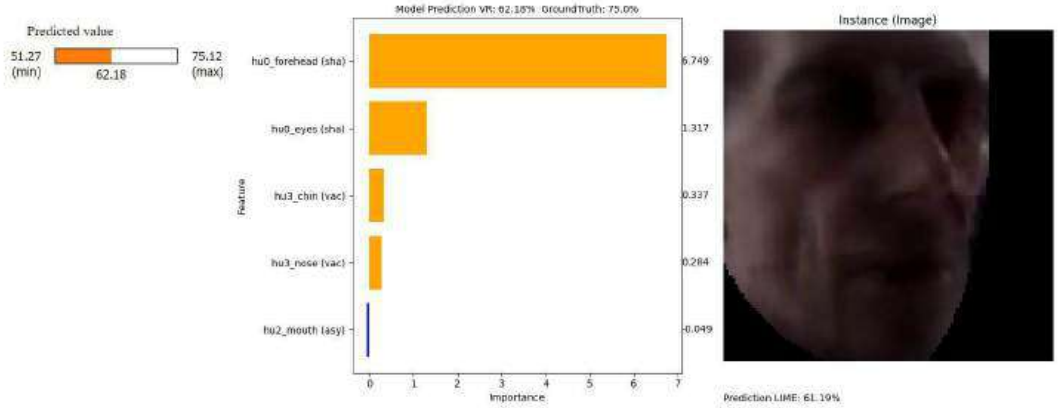

Predicted value
40.99
(min)
41.66
50.00
(max)
Model Prediction VR: 41.66% GroundTruth: 39.0%
hu3_chin (vac)
hu0_eyes (sha)
hu3_nose (vac)
hu2_mouth (asy)
hu1_forehead (dde)
0.510
0.482
0.086
-0.178
-2.428
Feature
−2.5 −2.0 −1.5 −1.0 −0.5 0.0 0.5
Importance
Instance (Image)
Prediction LIME: 44.77%

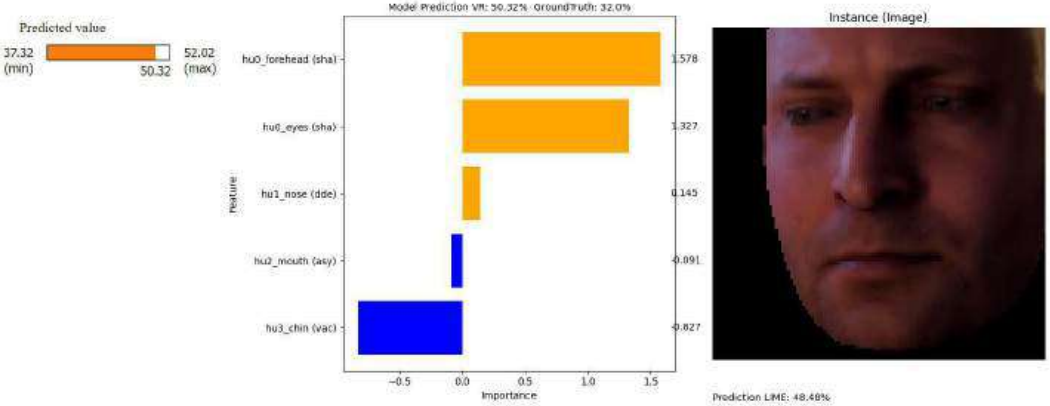

Predicted value
38.13
(min)
42.20
42.20(max)
Model Prediction VR: 42.2% GroundTruth: 7.0%
hu0_forehead (sha)
hu0_eyes (sha)
hu0_mouth (sha)
hu3_chin (vac)
hu1_nose (dde)
0.143
0.037
0.017
-0.226
-0.233
Feature
−0.25 −0.20 −0.15 −0.10 −0.05 0.00 0.05 0.10 0.15
Importance
Instance (Image)
Prediction LIME: 38.86%

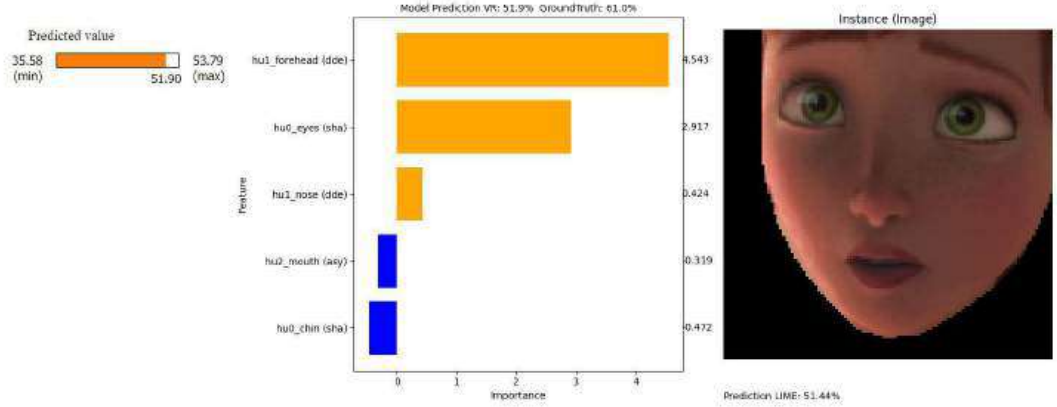

Predicted value
55.56
(min)
59.16
59.16(max)
Model Prediction VR: 59.16% GroundTruth: 52.0%
hu0_eyes (sha)
hu2_mouth (asy)
hu0_nose (sha)
hu0_forehead (sha)
hu0_chin (sha)
0.894
0.749
0.405
0.139
-0.051
Feature
0.0 0.2 0.4 0.6 0.8
Importance
Instance (Image)
Prediction LIME: 57.87%

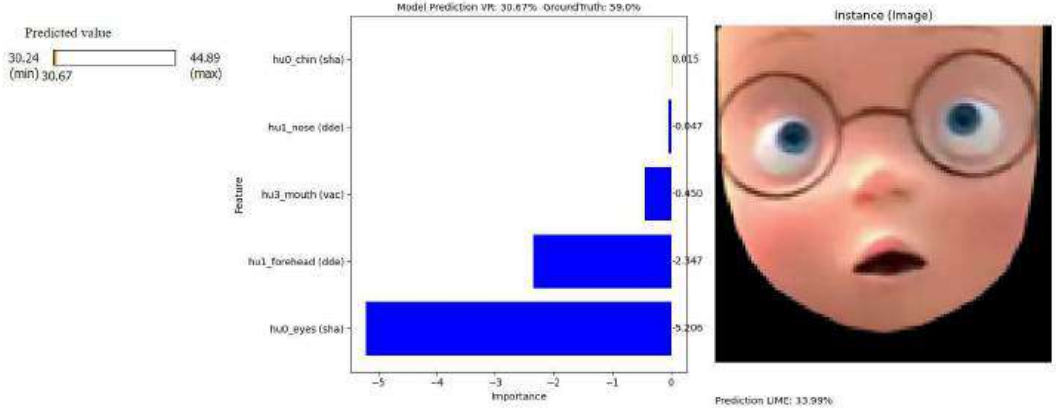

Predicted value
32.18
(min)
39.55
49.05
(max)
Model Prediction VR: 39.55% GroundTruth: 7.0%
hu1_eyes (dde)
hu2_mouth (asy)
hu1_nose (dde)
hu1_chin (dde)
hu1_forehead (dde)
4.034
0.304
0.254
0.195
-0.196
Feature
0 1 2 3 4
Importance
Instance (Image)
Prediction LIME: 47.61%

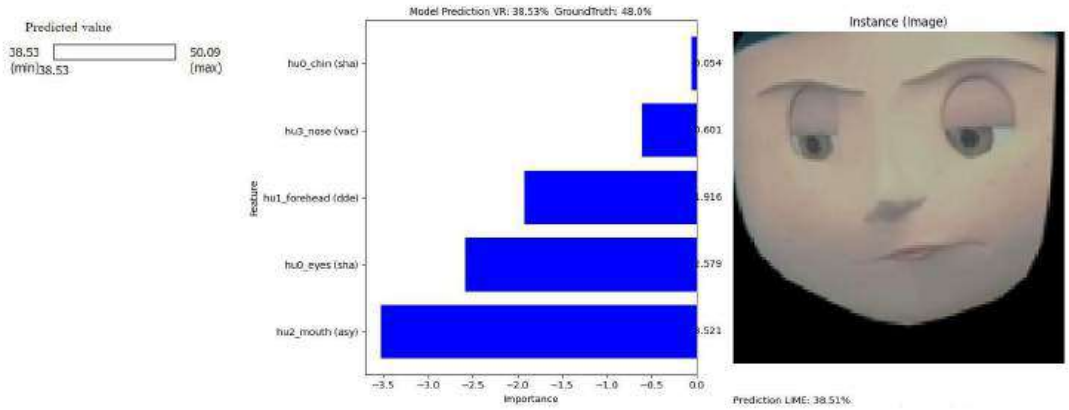

Predicted value
44.06
(min)
44.57
49.94
(max)
Model Prediction VR: 44.57% GroundTruth: 41.0%
hu1_nose (dde)
hu0_chin (sha)
hu2_mouth (asy)
hu3_forehead (dde)
hu0_eyes (sha)
Feature
−1.4
−1.2
−1.0
−0.8
−0.6
−0.4
−0.2
0.0
Importance
Instance (Image)
Prediction LIME: 44.56%

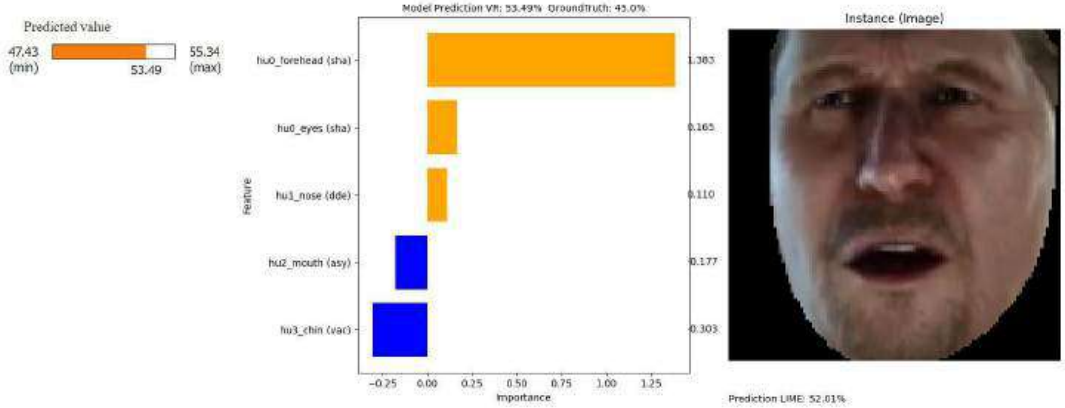

Predicted value
51.27
(min)
62.18
75.12
(max)
Model Prediction VR: 62.18% GroundTruth: 75.0%
hu0_forehead (sha)
hu0_eyes (sha)
hu3_chin (vac)
hu3_nose (vac)
hu2_mouth (asy)
6.749
1.317
0.337
0.284
-0.049
Feature
0
1
2
3
4
5
6
7
Importance
Instance (Image)
Prediction LIME: 61.19%

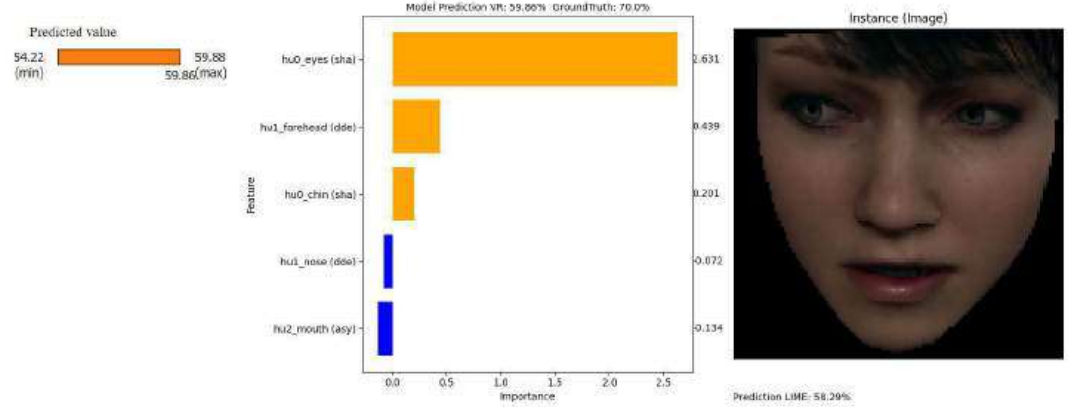

Predicted value
37.32
(min)
50.32
52.02
(max)
Model Prediction VR: 50.32% GroundTruth: 32.0%
hu0_forehead (sha)
hu0_eyes (sha)
hu1_nose (dde)
hu2_mouth (asy)
hu3_chin (vac)
1.578
1.327
0.145
-0.091
-0.827
Feature
−0.5
0.0
0.5
1.0
1.5
Importance
Instance (Image)
Prediction LIME: 48.48%

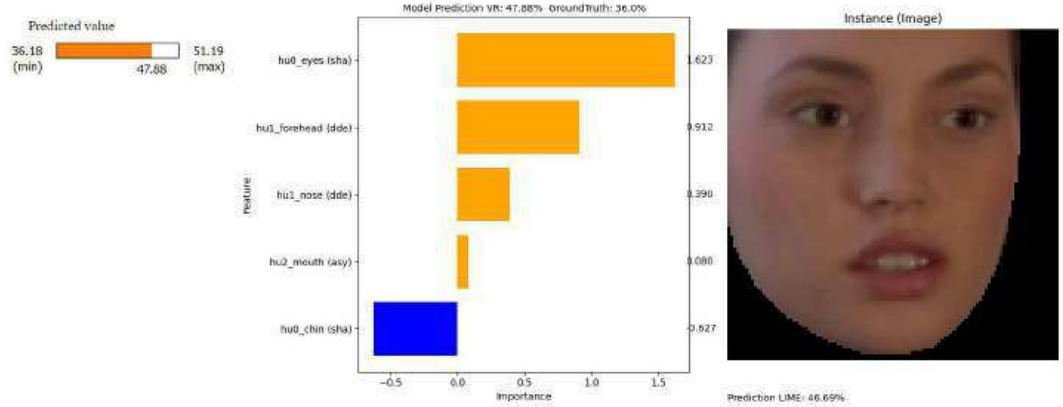

Predicted value
35.58
(min)
51.90
53.79
(max)
Model Prediction VR: 51.9% GroundTruth: 61.0%
hu1_forehead (dde)
hu0_eyes (sha)
hu1_nose (dde)
hu2_mouth (asy)
hu0_chin (sha)
4.543
2.917
0.424
-0.319
-0.472
Feature
0
1
2
3
4
Importance
Instance (Image)
Prediction LIME: 51.44%

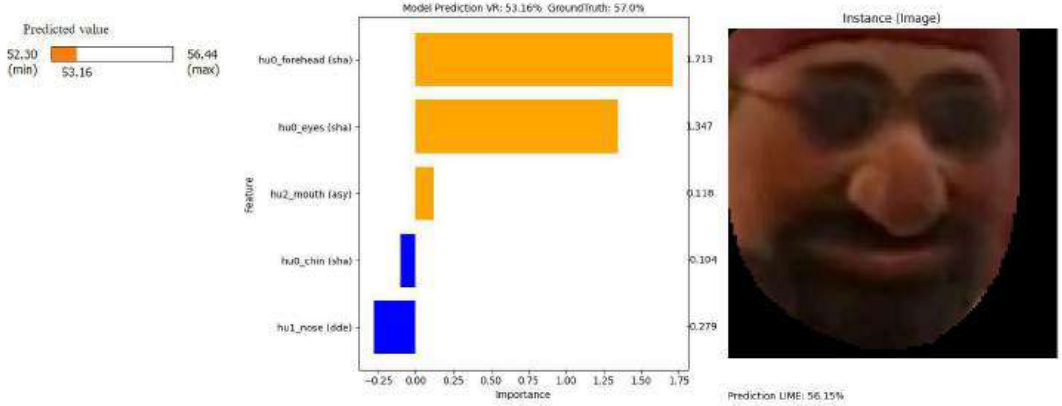

Predicted value
30.24
(min)
30.67
44.89
(max)
Model Prediction VR: 30.67% GroundTruth: 59.0%
hu0_chin (sha)
hu1_nose (dde)
hu3_mouth (vac)
hu1_forehead (dde)
hu0_eyes (sha)
0.015
-0.047
-0.450
-2.347
-5.206
Feature
−5
−4
−3
−2
−1
0
Importance
Instance (Image)
Prediction LIME: 33.99%

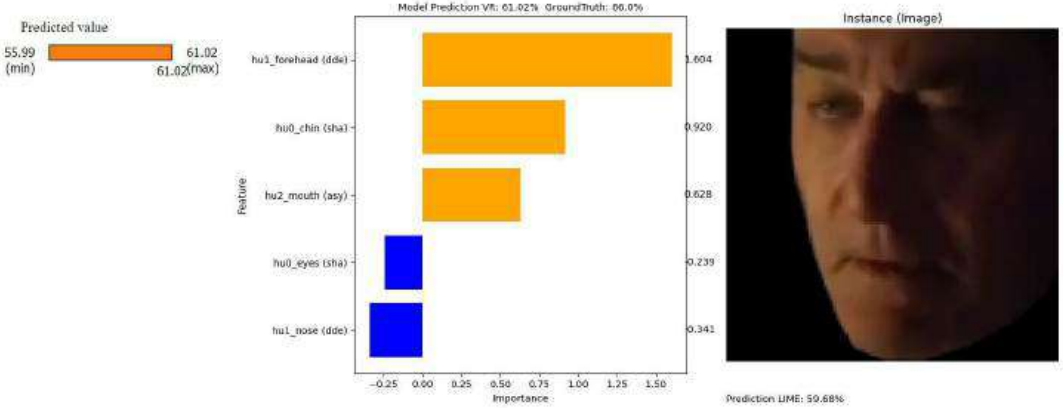

Predicted value
38.53
(min)
38.53
50.09
(max)
Model Prediction VR: 38.53% GroundTruth: 48.0%
hu0_chin (sha)
hu3_nose (vac)
hu1_forehead (dde)
hu0_eyes (sha)
hu2_mouth (asy)
-0.054
-0.601
-1.916
-2.579
-3.521
Feature
−3.5
−3.0
−2.5
−2.0
−1.5
−1.0
−0.5
0.0
Importance
Instance (Image)
Prediction LIME: 38.51%

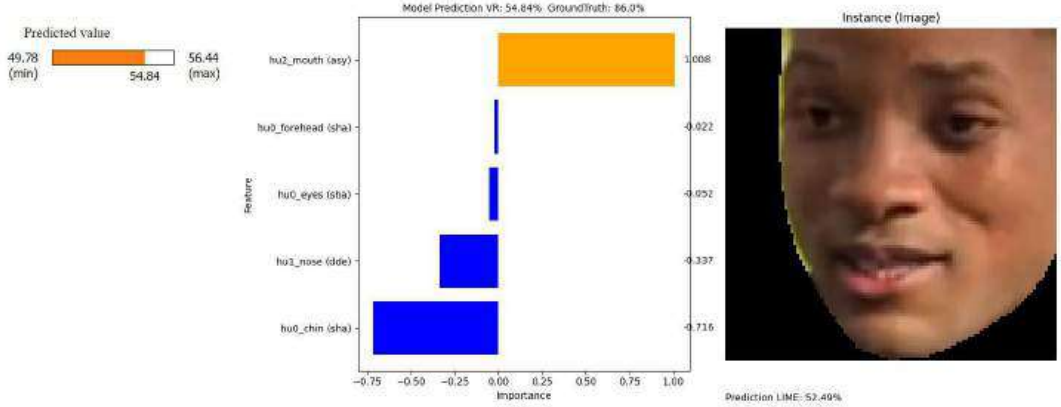

Predicted value
47.43
(min)
53.49
55.34
(max)
Model Prediction VR: 53.49% GroundTruth: 45.0%
hu0_forehead (sha)
hu0_eyes (sha)
hu1_nose (dde)
hu2_mouth (asy)
hu3_chin (vac)
1.383
0.165
0.110
-0.177
-0.303
Feature
−0.25
0.00
0.25
0.50
0.75
1.00
1.25
Importance
Instance (Image)
Prediction LIME: 52.01%


Predicted value
54.22
(min)
59.86
59.88
(max)
Model Prediction VR: 59.86% GroundTruth: 70.0%
hu0_eyes (sha)
hu1_forehead (dde)
hu0_chin (sha)
hu1_nose (dde)
hu2_mouth (asy)
2.631
0.439
0.201
-0.072
-0.134
Feature
0.0
0.5
1.0
1.5
2.0
2.5
Importance
Instance (Image)
Prediction LIME: 58.29%


Predicted value
36.18
(min)
47.88
51.19
(max)
Model Prediction VR: 47.88% GroundTruth: 36.0%
hu0_eyes (sha)
hu1_forehead (dde)
hu1_nose (dde)
hu2_mouth (asy)
hu0_chin (sha)
1.623
0.912
0.390
0.080
-0.627
Feature
Importance
Instance (Image)
Prediction LIME: 46.69%


Predicted value
52.30
(min)
53.16
56.44
(max)
Model Prediction VR: 53.16% GroundTruth: 57.0%
hu0_forehead (sha)
hu0_eyes (sha)
hu2_mouth (asy)
hu0_chin (sha)
hu1_nose (dde)
1.713
1.347
0.118
-0.104
-0.279
Feature
Importance
Instance (Image)
Prediction LIME: 56.15%


Predicted value
55.99
(min)
61.02
61.02(max)
Model Prediction VR: 61.02% GroundTruth: 66.0%
hu1_forehead (dde)
hu0_chin (sha)
hu2_mouth (asy)
hu0_eyes (sha)
hu1_nose (dde)
1.604
0.920
0.628
-0.239
-0.341
Feature
Importance
Instance (Image)
Prediction LIME: 59.68%


Predicted value
49.78
(min)
54.84
56.44
(max)
Model Prediction VR: 54.84% GroundTruth: 86.0%
hu2_mouth (asy)
hu0_forehead (sha)
hu0_eyes (sha)
hu1_nose (dde)
hu0_chin (sha)
1.008
-0.022
-0.052
-0.337
-0.716
Feature
Importance
Instance (Image)
Prediction LIME: 52.49%

Pontifícia Universidade Católica do Rio Grande do Sul
Pró-Reitoria de Pesquisa e Pós-Graduação
Av. Ipiranga, 6681 – Prédio 1 – Térreo
Porto Alegre – RS – Brasil
Fone: (51) 3320-3513
E-mail: propesq@pucrs.br
Site: www.pucrs.br