Emotionalyzer: Player's Facial Emotion Recognition ML Model for Video Game Testing Automation

Rebeca Bravo-Navarro, Luis Pineda-Knox and Willy Ugarte[a]

Universidad Peruana de Ciencias Aplicadas (UPC), Lima, Peru

Keywords: Sentiment Analysis, Human-Computer Interaction, Player Testing, Gameplay Experience Testing, Facial Emotion Recognition.

Abstract: In video game development, the play testing phase is crucial for evaluating and optimizing user perception before launch. These tests are often costly and require significant time investment, as they are conducted by experts observing gameplay sessions, which makes capturing real-time data, such as facial and bodily expressions, challenging. Additionally, many independent studies lack the necessary resources to conduct professional testing. Therefore, smaller developers need more cost-effective and time-efficient alternatives to improve their products and streamline the development process. This project aims to develop a real-time facial emotion recognition model using machine learning, which will be integrated into an application that records the player's emotions during the gameplay session. It seeks to benefit Peruvian indie companies by reducing costs and time associated with traditional testing and providing a more precise and detailed evaluation of the user experience. Additionally, the use of machine learning technology ensures continuous adaptation and progressive improvements in the model over time.

1 INTRODUCTION

In the realm of video game development, the process of user experience testing holds significant importance prior to a game's launch. This step is critical for evaluating and refining the game from the perspective of the player (Dumas and Redish, 1993).

Major players in the video game industry invest substantial resources into this phase, ensuring that games are released only when they meet a certain level of quality and completeness (El-Nasr et al., 2013).

However, conducting user testing can be both costly and time-consuming, often requiring the presence of an expert to observe gameplay sessions. One notable challenge of this approach is its limited ability to capture spontaneous reactions, such as facial expressions and body language during gameplay.

Typically, player feedback is gathered through post-game interviews and surveys[1,2], although these methods may introduce biases due to memory and self-reporting tendencies.

Turning our attention to Peru, the local chapter of the International Game Developers Association (IGDA) [3] reports over 50 registered companies engaged in video game development, alongside numerous independent developer teams, with a focus on indie games.

Despite the growing presence of the industry, the videogame landscape in Peru remains relatively young, with most companies emerging after the year 2000.

Nevertheless, due to resource constraints, few, if any, independent developers can conduct professional-level user testing. Notable examples of this situation include Colorful by Peruvian developer Hitoshi Kanno, which remains in development ten years after its initial conceptualization[4], and Peruvian studio Bamtang Games, which only recently implemented its own Quality Assurance department despite developing games for over 20 years[5].

[a] https://orcid.org/0000-0002-7510-618X

[1] Live-game satisfaction survey: https://www.hoyolab.com/article/3523425

[2] Closed alpha survey announcement: https://twitter.com/KAGESMG/status/1784255738174468196

[3] IGDA Peru website: https://igda.pe/

[4] Colorful website: https://www.facebook.com/ColorfulTheGame

[5] Jesus Blas's talk on 'Cambios en el desarrollo de videojuegos': https://www.facebook.com/igdaperu/videos/250518101166436

Bravo-Navarro, R., Pineda-Knox, L. and Ugarte, W.
Emotionalyzer: Player's Facial Emotion Recognition ML Model for Video Game Testing Automation.
DOI: 10.5220/0013439400003929
In *Proceedings of the 27th International Conference on Enterprise Information Systems (ICEIS 2025) - Volume 1*, pages 937-943
ISBN: 978-989-758-749-8; ISSN: 2184-4992

This underscores the need for more cost-effective and time-efficient alternatives that not only enhance the quality of their products but also streamline the development process, particularly during the design phase.

Our proposed solution entails the development of an autonomous learning model capable of recognizing user emotions through facial recognition technology.

This model would be implemented and tested using a desktop application that runs during gameplay sessions, capturing user inputs to generate an emotion report upon completion. Our work primarily targets independent companies within Peru's video game development sector.

Given the growing nature of this industry in the country, many such companies operate with modest budgets and development teams, typically classified as small or medium-sized enterprises.

Our scope encompasses the development of a facial emotion recognition (FER) model tailored for use in video game testing by independent entities, alongside data analysis to predict emotions in single-player experiences. Additionally, the model will be implemented through a desktop application.

Altough, it is important to note that the project excludes testing multiplayer experiences, emotion recognition through voice or bodily movements, and implementation for mobile devices or consoles.

Despite its promising potential, our proposal faces several constraints. Limited financial resources may hinder its execution, necessitating careful budgetary planning. Additionally, potential shortages in human resources, such as experience and skill sets, could pose implementation challenges.

Moreover, legal and ethical considerations related to the collection and processing of facial data may impact project development.

Lastly, a critical limitation is the availability of a diverse range of facial expressions in the training data, which is crucial for refining the facial emotion recognition model.

Ultimately, our project aims to serve as an innovative and accessible solution to enhance user testing within Peru's independent video game development sector.

By addressing prevailing limitations in cost and accuracy associated with emotional data collection, our initiative seeks to empower independent developers to compete on the global stage with products of superior quality.

This paper is distributed in the following sections: First, we review the related works of Facial Emotion Recognition (FER) in Section 2. Then, we talk about relevant concept and theory related to the background of our research and describe in more detail our main contribution in Section 3. Furthermore, we will explain procedures performed and the experiments that were carried out in this work in Section 4. In the end, we will show the main conclusions of the project and we will indicate some recommendations for future work in Section 5.

2 RELATED WORK

Regarding our issue, we have reviewed a wide range of documents and previous research on emotion recognition through machine learning, as well as a limited number of similar applications of this technology in the realm of video games.

However, unfortunately, we have not identified solutions comparable to our proposed project. Below, we will present the most relevant works related to our proposal.

2.1 Towards Personalised Gaming via Facial Expression Recognition

In (Blom et al., 2014), the authors address the issue of personalizing the gaming experience through real-time emotion recognition.

The growing significance of AI in gaming underscores the need for tailored experiences. Researchers proposed a technique based on modifying game levels guided by player expressions, leveraging computer vision.

Using IN-SIGHT SDK for facial expression recognition, they tracked emotions and mapped them to gameplay challenges in Infinite Mario Bros.

Results showed effective adaptation of difficulty levels based on player emotions, with user preference for this dynamic approach over static level systems.

Both (Blom et al., 2014) and our proposal center around the enhancement of gaming experiences through the implementation of emotion recognition techniques during gameplay.

They exhibit commonalities in their recognition of the pivotal role of personalization, utilization of facial recognition technologies, and acknowledgment of the crucial significance of advanced technologies such as computer vision and machine learning.

2.2 Facial Emotion Recognition Using Deep Learning Detector and Classifier

In (Kit et al., 2023), the authors focus the issue of facial expression recognition, highlighting the importance of non-verbal elements in human communication.

A deep learning-based system is proposed, utilizing the MobileNetv-1 model to predict emotions in video sequences, prioritizing speed and accuracy.

Training datasets are prepared in both color and grayscale, followed by model training and evaluation.

It is concluded that grayscale facial recognition achieves an accuracy of 86.42%, surpassing color recognition due to the influence of lighting on color variation.

Facial alignment and image color space significantly affect the accuracy and computational cost of facial emotion recognition.

2.3 A Systematic Review on Affective Computing: Emotion Models, Databases, and Recent Advances

In (Wang et al., 2022), the authors address the challenge of emotional recognition and sentiment analysis through physical and physiological data, aiming to enhance human-computer interaction.

Over 380 studies were reviewed, categorizing affective computing into unimodal recognition and multimodal analysis. Models based on physical and physiological information were examined, concluding with a comprehensive analysis of model efficacy.

The fusion of physical and physiological data enables the extraction of useful features to improve affective computing models.

A systematic review of emotion models, databases, and recent advances is presented, intending to guide academic and industrial researchers towards promising new directions in this rapidly advancing field.

Within (Wang et al., 2022) , the authors conducts a thorough review of existing research and analyzes the efficacy of emotional recognition models within the realm of human-computer interaction.

In contrast, our proposal outlines a plan to develop a specific emotional recognition model tailored for gaming experiences.

Furthermore, they address a broader context of human-computer interaction, while our proposal specifically targets the gaming domain.

Lastly, the authors provide a comprehensive overview of existing models and advancements, whereas our proposal concentrates on outlining objectives and steps for the development and implementation of a gaming-specific emotional recognition model.

In essence, both works aim to enhance human interaction through emotional recognition, but they diverge in focus, context, and level of detail.

2.4 A System of Emotion Recognition and Judgment and Its Application in Adaptive Interactive Game

In (Lin et al., 2023), the authors propose system for recognizing and assessing emotions based on optimal physiological signals for interactive gaming applications is established.

Ten participants played the Super Mario game while their physiological responses were recorded to assess the game's effect on their emotions.

The results of this system were compared with conventional machine learning methods, demonstrating the superiority of the former.

The system enabled the detection of emotional changes in players during gameplay, enhancing their experience.

It was observed that players' perceptions of emotional changes varied and that prior testing experience affected the results. This underscores the effectiveness of the proposed system and its potential to enhance interactivity in technology-based games.

Comparing (Lin et al., 2023) with our proposal, the former provides concrete evidence of the effectiveness of an emotion recognition system in gaming, whereas our proposal outlines a plan for future development in a related area.

Both contribute to the understanding and advancement of emotion recognition in gaming, but they differ in their stages of development and presentation of results.

2.5 Towards Automated Video Game Testing: Still a Long Way to Go

In (Politowski et al., 2022), the authors explore the escalating complexity of game development, driving up costs and necessitating larger, higher-quality games. It delves into the challenges of manual playtesting, particularly for smaller companies lacking in-house QA teams.

There's a growing interest in automated video game testing, although skepticism persists among de-

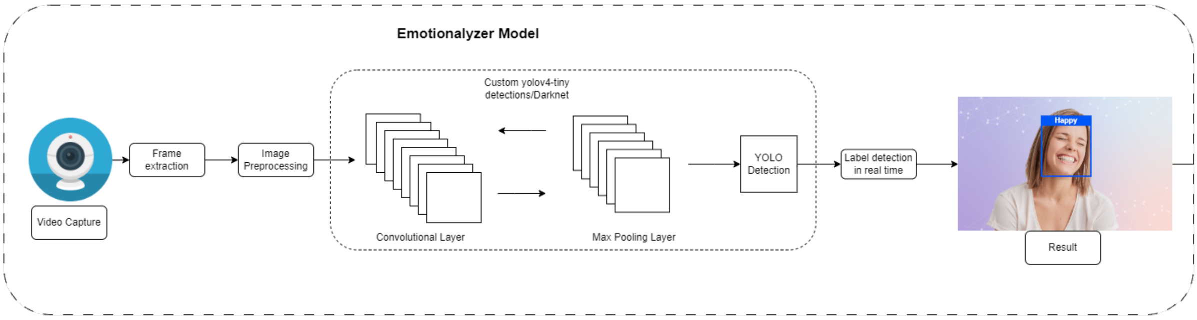

Figure 1: Model architecture.

velopers.

Academic research highlights machine learning and AI-based approaches, yet practical implementation remains a concern. Notable solutions like Wuji and ICARUS show promise in automating game testing processes.

Key challenges include maintaining game functionality and addressing the lack of automated test maintenance. Bridging the gap between theoretical advancements and practical implementation is crucial for the future of game testing.

Overall, while automated testing holds potential, collaboration between academia and industry is essential for its successful integration.

Comparing (Politowski et al., 2022) with our proposal, both address the challenges and potential solutions related to game development and testing.

In (Politowski et al., 2022), the authors discuss the escalating complexity of game development, particularly focusing on the challenges of manual playtesting and the growing interest in automated video game testing.

Similarly, our proposal acknowledges the need for more efficient and cost-effective testing methods, particularly in the context of smaller game development companies.

3 CONTRIBUTION

The primary contribution of this research lies in the application of facial emotion recognition techniques to streamline the testing phase of video games.

By automating player observation and accurately identifying the emotions experienced during testing sessions, this approach transforms raw data into refined information that developers can readily use to enhance their creations.

3.1 Preliminary Concepts

In this section, the primary concepts utilized in our research are introduced.

Definition 1 (Facial emotion recognition (Vedantham and Reddy, 2023))**.** *Pertains to the ability to discern emotional states based on gathered data, whether in the form of video clips or images. It has been an area of ongoing research development for several years.*

Definition 2 (Game experience testing (Dumas and Redish, 1993))**.** *This stage is conducted prior to the release of a video game and is essential for evaluating the game from the user's perspective and enhancing their experience.*

Definition 3 (Player experience (Nacke and Drachen, 2011))**.** *It focuses on the qualitative aspects of player interaction with games, taking into account factors such as enjoyment and difficulty, among others.*

Definition 4 (Sentiment analysis (Chaturvedi et al., 2018))**.** *Seeks to categorize text (and sometimes audio and video) as either positive or negative. It is closely linked to information retrieval and fusion since it involves collecting, integrating, and classifying data. It's a complex research problem that involves addressing various NLP tasks, including named entity recognition, concept extraction, sarcasm detection, aspect extraction, and subjectivity detection.*

In this preceding section, we delved into the concepts of facial emotion recognition, player experience, and game experience testing.

3.2 Method

In this section, we detail the method we've developed for our research, in which we utilize "Darknet," an open-source neural network framework, for detecting and classifying facial emotions.

The main contribution of this research consists of applying a machine learning model capable of providing accurate emotion recognition during the play test of a video game, with the purpose of streamlining the process.

In the Fig. 1, we can observe the beginning of the process, from image capture via the webcam to the output of results.

3.3 Video Capture

The process initiates with capturing video using a webcam, continuously obtaining video frames.

3.4 Frame Extraction

The video stream is divided into individual frames for further processing and analysis, as the model processes single images rather than a continuous stream.

3.5 Image Preprocessing

The extracted frames are preprocessed to prepare them for the convolutional neural network input.

3.6 Our Model

Convolutional Layer. : The preprocessed images are processed through a convolutional layer, applying filters to extract key features such as edges, textures, and patterns important for emotion detection.

Max Pooling Layer. The output from the convolutional layer undergoes max pooling, reducing the spatial dimensions and emphasizing the most prominent features to reduce computational complexity.

YOLO Detection. The processed features are input into a YOLO (You Only Look Once) detection model, customized using yolov4-tiny and the Darknet framework for real-time object detection. YOLO detects faces in the frame and identifies emotions by labeling them accordingly.

3.7 Result

The final output is an image with each detected face annotated with the corresponding emotion, displayed in real time to show the detected emotions for each face in the video frame.

4 EXPERIMENTS

In this section, we will cover the experiments conducted in our project, the requirements for replicating these experiments, and an analysis of the results obtained from this process.

4.1 Experimental Protocol

This subsection provides information about the environment setup for the experiments, including details on the local hardware configuration and the applications utilized.

This project was developed on a Google Colab notebook using the L4 GPU. The model was built with Darknet v4.5.4. The system specifications are as follows:

- NVIDIA-SMI 535.104.05 Driver Version: 535.104.05
- CUDA Version: 12.2

4.2 Face Emotion Dataset

For training the model, two datasets were selected because the available public datasets did not provide a sufficient number of images to ensure effective training.

The first dataset is FER-2013[6], consisting of grayscale facial images with a resolution of 48x48 pixels.

These images are automatically aligned to be centered and occupy a uniform space in each image.

The public test dataset contains 3,589 examples. The second dataset is a public sample from the AffectNet-HQ [7] dataset, which includes 29,042 examples with a resolution of 96x96 pixels.

To effectively integrate both datasets, a complete standardization of the images to the .png format was carried out, converting all images to grayscale and adjusting them to a uniform resolution of 48x48 pixels.

During the preparation of the training set, a thorough review was conducted to identify and remove images that did not show faces, thus ensuring the quality and consistency of the final dataset.

Additionally, emotional labels were assigned to organize the images into corresponding folders. Each folder was then divided into two subsets: one for training and one for validation, distributed in a 70% and 30% ratio, respectively.

This approach ensures a robust and well-structured preparation for the facial emotion classification model.

[6] Public FER-2013 dataset sample: https://www.kaggle.com/datasets/msambare/fer2013

[7] Public AffectNet-HQ dataset sample: https://www.kaggle.com/datasets/noamsegal/affectnet-training-data

Table 1: Our model results by emotions.

Name	Avg. Precision	TP	FN	FP	TN	Accuracy	Error Rate	Precision	Recall	Specificity	False Pos Rate
angry	74.1064	17	3	24	59	.7379	.2621	.4146	.8500	.7108	.2892
disgust	66.8204	19	1	29	70	.7479	.2521	.3958	.9500	.7071	.2929
fear	62.7131	12	8	15	42	.7013	.2987	.4444	.6000	.7368	.2632
happy	67.2481	13	7	14	63	.7835	.2165	.4815	.6500	.8182	.1818
sad	74.7543	17	3	10	52	.8415	.1585	.6296	.8500	.8387	.1613
surprise	58.9889	18	2	30	54	.6923	.3077	.3750	.9000	.6429	.3571
neutral	53.0684	16	4	34	68	.6885	.3115	.3200	.8000	.6667	.3333

Table 2: Our model general results.

Precision	.42
recall	.80
F1-score	.55
TP	112
FP	156
FN	28
Mean avg.precision	.653856

Table 3: Models Comparison.

Model	Avg. Precision
MobileNet_v1_1.00_224	69.62%
MobileNet_v1_0.75_160	66.72%
RANDA	88.71%
R-emo	65.4%
Our Model	**65.39%**

4.3 Results

In this subsection, the experiments carried out and the results obtained in each of these are detailed.

As depicted in Table 2, the model evaluation revealed a precision of 0.42, indicating that 42% of positively classified detections were accurate. Furthermore, the recall achieved was 0.8, signifying the model's correct identification of 80% of positive instances within the dataset.

The resulting F1-score stood at 0.55, harmonizing precision and recall into a consolidated metric. The assessment identified 112 true positives alongside 156 false positives and 28 false negatives.

The average Intersection over Union (IoU) was 33.63%, representing the mean overlap between predicted bounding boxes and ground truth. Applying an IoU threshold of 50%, the average precision reached 65.39%, evaluating the model's object detection capability across varying confidence thresholds.

Considering the precision metric and comparing our model with other approaches, it can be observed from Table 3 that our model shows the lowest precision percentage, with a difference of 23.32% compared to the reference model, which was RANDA.

5 CONCLUSIONS

Based on our findings, the current model demonstrates acceptable performance in recognizing facial emotions during video game testing. However, there is considerable potential for improvement.

One key area for enhancement is the incorporation of facial landmarks in future evaluations. These landmarks can provide more detailed information about facial expressions, which could significantly improve the model's accuracy.

Additionally, fine-tuning the model parameters through more extensive training and validation could further enhance its performance by reducing false positive rates. When compared to the RANDA model, our current model exhibits lower accuracy, underscoring the necessity for additional optimizations.

These optimizations might include refining the feature extraction process, experimenting with different machine learning algorithms, or employing more sophisticated data augmentation techniques to better handle the variability in facial expressions. By addressing these areas, we aim to achieve precision levels that are comparable to, or even surpass, those of the RANDA model.

Moreover, conducting more comprehensive testing with a larger and more diverse dataset could help identify specific weaknesses and areas for further refinement. (Burga-Gutierrez et al., 2020) Continuous iteration and feedback from real-world testing scenarios will be crucial in evolving our model to meet the high standards required for effective emotion recognition in video game development.

Looking forward, we aim to integrate our model into a computer application designed for real-time analysis of facial emotions during video game testing. (Guillermo et al., 2023) This application will leverage the improved accuracy and reduced false positive rates achieved through incorporating facial landmarks and fine-tuning model parameters.

By enabling real-time emotion detection, this tool could provide invaluable insights into player experiences, helping developers identify areas of frustration, excitement, or disengagement. (de Rivero et al., 2023) This immediate feedback can streamline the development process, allowing for timely adjustments to improve overall game design and user experience.

The development of this application will also involve optimizing the model's computational efficiency to ensure it operates effectively within the constraints of real-time processing during video game testing sessions.

REFERENCES

Blom, P. M., Bakkes, S., Tan, C. T., Whiteson, S., Roijers, D. M., Valenti, R., and Gevers, T. (2014). Towards personalised gaming via facial expression recognition. In Horswill, I. and Jhala, A., editors, *Proceedings of the Tenth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, AIIDE 2014, October 3-7, 2014, North Carolina State University, Raleigh, NC, USA*. AAAI.

Burga-Gutierrez, E., Vasquez-Chauca, B., and Ugarte, W. (2020). Comparative analysis of question answering models for HRI tasks with NAO in spanish. In *SIMBig*, volume 1410 of *Communications in Computer and Information Science*, pages 3–17. Springer.

Chaturvedi, I., Cambria, E., Welsch, R. E., and Herrera, F. (2018). Distinguishing between facts and opinions for sentiment analysis: Survey and challenges. *Inf. Fusion*, 44:65–77.

de Rivero, M., Tirado, C., and Ugarte, W. (2023). Formalstyler: Gpt-based model for formal style transfer with meaning preservation. *SN Comput. Sci.*, 4(6):739.

Dumas, J. S. and Redish, J. C. (1993). A practical guide to usability testing. *Intellect*.

El-Nasr, M. S., Drachen, A., and Canossa, A., editors (2013). *Game Analytics, Maximizing the Value of Player Data*. Springer.

Guillermo, L., Rojas, J., and Ugarte, W. (2023). Emotional 3d speech visualization from 2d audio visual data. *Int. J. Model. Simul. Sci. Comput.*, 14(5):2450002:1–2450002:17.

Kit, N. C., Ooi, C.-P., Tan, W. H., Tan, Y.-F., and Cheong, S.-N. (2023). Facial emotion recognition using deep learning detector and classifier. *International Journal of Electrical and Computer Engineering (IJECE)*, 13(3):3375–3383.

Lin, W., Li, C., and Zhang, Y. (2023). A system of emotion recognition and judgment and its application in adaptive interactive game. *Sensors*, 23(6):3250.

Nacke, L. and Drachen, A. (2011). Towards a framework of player experience research (pre-print). In *Foundations of Digital Games Conference*.

Politowski, C., Guéhéneuc, Y., and Petrillo, F. (2022). Towards automated video game testing: Still a long way to go. In *6th IEEE/ACM International Workshop on Games and Software Engineering, GAS@ICSE, Pittsburgh, PA, USA, May 20, 2022*, pages 37–43. ACM.

Vedantham, R. and Reddy, E. S. (2023). Facial emotion recognition on video using deep attention based bidirectional LSTM with equilibrium optimizer. *Multim. Tools Appl.*, 82(19):28681–28711.

Wang, Y., Song, W., Tao, W., Liotta, A., Yang, D., Li, X., Gao, S., Sun, Y., Ge, W., Zhang, W., and Zhang, W. (2022). A systematic review on affective computing: emotion models, databases, and recent advances. *Inf. Fusion*, 83-84:19–52.