UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Posgrados

Emotion Recognition through Fine-tuned CNNs: Analyzing Facial Expressions in Videos

Proyecto de Titulación

Edison Daniel Marin Alquinga

Felipe Grijalva, Ph.D.

Director de Trabajo de Titulación

Trabajo de titulación de posgrado presentado como requisito para la obtención del título de Magíster en Inteligencia Artificial

Quito, 02 de diciembre de 2024

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ
COLEGIO DE POSGRADOS

HOJA DE APROBACIÓN DE TRABAJO DE TITULACIÓN

Emotion Recognition through Fine-tuned CNNs: Analyzing Facial Expressions in Videos

Edison Daniel Marin Alquinga

Nombre del Director del Programa:	Felipe Grijalva
Título académico:	Ph.D. en Ingeniería Eléctrica
Director del programa de:	Inteligencia Artificial
Nombre del Decano del colegio Académico:	Eduardo Alba
Título académico:	Doctor en Ciencias Matemáticas
Decano del Colegio:	Ciencias e Ingenierías
Nombre del Decano del Colegio de Posgrados:	Dario Niebieskikwiat
Título académico:	Doctor en Física

Quito, diciembre 2024

© DERECHOS DE AUTOR

Nombre del estudiante:	Edison Daniel Marin Alquinga
Código de estudiante:	00339450
C.I.:	0503910903
Lugar y fecha:	Quito, 02 de noviembre de 2024

ACLARACIÓN PARA PUBLICACIÓN

Nota: El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en http://bit.ly/COPETheses.

UNPUBLISHED DOCUMENT

Note: The following graduation project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on http://bit.ly/COPETheses.

RESUMEN

Gracias a los avances de inteligencia artificial, la capacidad en descifrar las emociones humanas a partir del rostro se ha convertido en un modo crucial de comunicación. Por lo que se propone un modelo de reconocimiento de emociones basado en aprendizaje profundo. A partir de 283,9k imágenes, se realiza la preparación de los datos. Este incluye el preprocesamiento, división y aumento, seguido de un ajuste fino de una red neuronal convolucional preentrenada. Para el entrenamiento escogen las imágenes provenientes de AffectNet, donde el modelo clasifica en 7 categorías originales. Luego, se realiza una evaluación, tomando como base los fotogramas de los rostos en videos detectados mediante MTCNN, teniendo como resultado una linea de tiempo de las emociones detectadas en el video de manera agrupada en positivo, negativo y neutro.

Palabras clave: AffectNet, expresiones faciales, mobileNet, MTCNN.

ABSTRACT

Thanks to advances in artificial intelligence, the ability to decipher human emotions from the face has become a crucial mode of communication. Therefore, an emotion recognition model based on deep learning is proposed. From 283.9k images, data preparation is performed. This includes preprocessing, splitting, and augmentation, followed by fine-tuning a pre-trained convolutional neural network. For training, images from the AffectNet dataset are used, where the model classifies them into the original 7 categories. Then an evaluation is performed by analyzing frames of faces detected in videos using MTCNN. The result is an emotion timeline for the video, with the emotions grouped into positive, negative, and neutral categories.

Key words: AffectNet, facial expressions, mobileNet, MTCNN.

TABLA DE CONTENIDO

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

Emotion Recognition through Fine-tuned CNNs: Analyzing Facial Expressions in Videos

Felipe Grijalva, *Senior Member, IEEE,* Daniel Marin, *Member, IEEE*

***Abstract*—Thanks to advances in artificial intelligence, the ability to decipher human emotions from the face has become a crucial mode of communication. Therefore, an emotion recognition model based on deep learning is proposed. From 283.9k images, data preparation is performed. This includes preprocessing, splitting, and augmentation, followed by fine-tuning a pre-trained convolutional neural network. For training, images from the AffectNet dataset are used, where the model classifies them into the original 7 categories. Then an evaluation is performed by analyzing frames of faces detected in videos using MTCNN. The result is a timeline of emotions for the video, grouped into positive, negative, and neutral categories.**

***Index Terms*—AffectNet, facial expressions, mobileNet, MTCNN.**

I. Introduction

The recognition of human emotions based on facial expressions (FE) is an important task that has applications in fields such as human-computer interaction, behavioral analysis, and affective computing [1]. According to Zhu et al. [2], more than 55% of facial expressions are reflected as emotional information, which is transmitted by people. Based on this, the development of models capable of interpreting signals automatically is very valuable to improve interaction with digital systems [3]. However, FE are complex and vary depending on cultural context, physical characteristics, and individual subtleties, making classification difficult [4].

Advances in convolutional neural networks (CNNs) have greatly improved accuracy in computer vision tasks, including emotion detection [5]. Normally, when deep network training is done from scratch, it involves the use of large amounts of data and computational resources, resulting in a limitation [6]. For this reason, the fine-tuning of a model called MobileNetV2 has become popular.

The MobileNetV2 architecture is designed to perform efficient image classification tasks such as facial expression recognition (FER) [7]. In addition, it is chosen to use the FE dataset from AffectNet; which contains a base of more than 1 million images [8].

This research aims to design and build a facial expression recognition system based on the fine-tuning of CNNs to classify emotions into three broad categories: positive, neutral, and negative grouped from the original set of emotions. The project aims to implement a test system capable of detecting faces in the video, recognizing the dominant emotions at each moment, and generating a timeline that summarizes the emotional behavior throughout the recording.

II. State of the art

This section covers some previous studies on the RES, databases, networks to be trained and pre-trained in order to know the accuracy during training and tests performed, among other noteworthy parameters.

Kuruvayil and Palaniswamy [9] use the images from the CMU Multi PIE dataset with 0,12% and 0,02% of 750k images [10] to carry out network training and validation; respectively. By applying ResNet with 4 residual blocks containing 2 convolutions for each block, a meta-learning, a learning rate of 1 hundredth and 1250 epochs, a 90% accuracy in training is obtained. In addition, to make it more efficient, the stochastic gradient descent (SGD) is carried out and through a classification 5 classes of emotions are obtained, such as disgust, joy, neutral, anger and surprise. Finally, they use 475 images from the AffectNet and CMU Multi PIE databases, in order to determine an average result of 68% of the performance of these tests.

Wang et al. [11] used 4 databases, one of them being AffectNet. When applying 28,45% of the total images, 280k and 4,5k are taken as samples for training and respective tests. From a learning rate of 0.0001, batch size equal to 128, 300 epochs and the GCANet architecture, they obtain an accuracy that does not exceed 61% for 7 classes of emotions, joy, anger, fear, disgust, sadness, surprise, neutral.

Likewise, other authors such as Ullah et al. [12] use 2 databases called AffectNet and RAF-DB to compare their trainings with different architectures. Of which the AffectNet study base is taken as a reference. Of the total number of images, 29,16% (291651) and 0,4% (4k) are used for training and testing the model; respectively. Applying a combination of VGG-19, ResNet-50, and Inception-V3 extracts the features of the AffectNet images and forges a more robust assembly. Consequently, it goes through a CNN and you get an accuracy of 89% for either training or testing, as they handle the same dataset. Finally, they apply a Softmax classifier to define the 7 emotions mentioned in the previous case.

Although there have been several studies that seek to improve network performance, there are difficulties such as recognizing between anger and disgust or between the expression of fear and surprise. Magherini et al. [13] is a case study, where they used 70%, 10% and 20% of

a total of 320k images, during training, validation and testing (TVT); respectively. They applied a learning rate of 0,001, inceptionResNetv2 architectures for classification and ResNet50v2 in validation, reaching accuracies of more than 96% in any of its TVT stages. Emotion detection was performed on subjects viewed through a webcam.

Based on 9k images made up of videos of 30 participants and complemented with the AffectNet base, 70% of the training is carried out and both validation and testing 15%. This was established by Shomoye and Zhao. [14] They also applied a MobileNet-based CNN whose batch size is 16, used a dropout rate of 0,5, regularization equal to 0,3 and about 100 epochs to obtain an accuracy of 77% in training. However, it should be noted that of the 11 types of emotions, four were carried out: joy, neutral, boredom and confusion.

III. Methodology

The proposed model involves developing a FER system to obtain information about a person's emotional state. Figure 1 shows a general procedure that starts from the preprocessing of the data to obtaining a detailed view of the person's emotional change through a video.

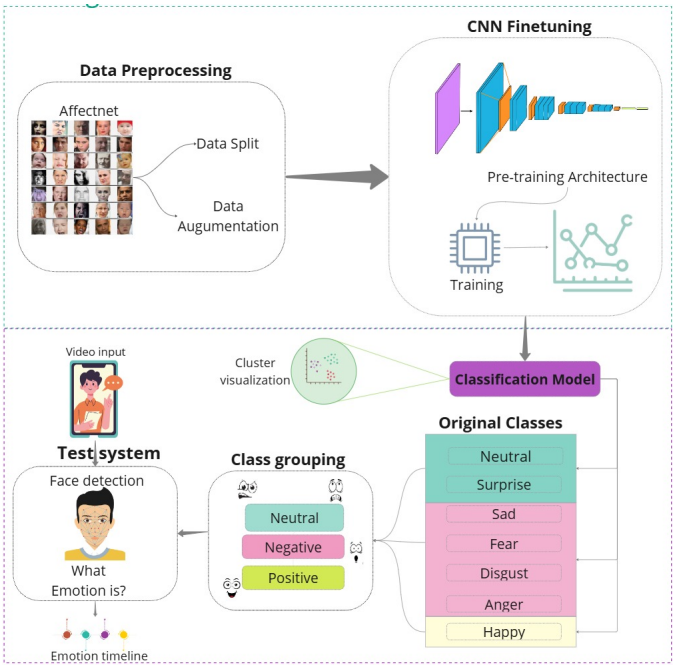

Figure 1. Block diagram of the proposal

A. Data preprocessing

By using 28.39% (283,9k photos) of the AffectNet database, the data is pre-processed. This consists of classifying the annotated images according to the type of emotion, as illustrated in Figure 2; which has 7 classes.

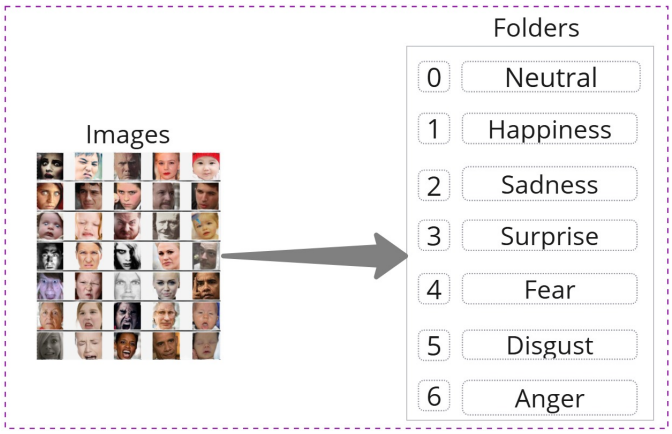

Figure 2. Directories for FE

Then, data management is carried out. By applying PyTorch Lighting, it is ensured that the information is processed efficiently [15]. To do this, initialization, configuration and loading of the data is performed.

At the beginning, the main arguments are placed, such as the address, batch size for training and validation, number of parallel processes and a proportion in the validation of the data, which in this case is 20% of the selected images. In addition, data augmentation is performed for training, seen in Figure 3.

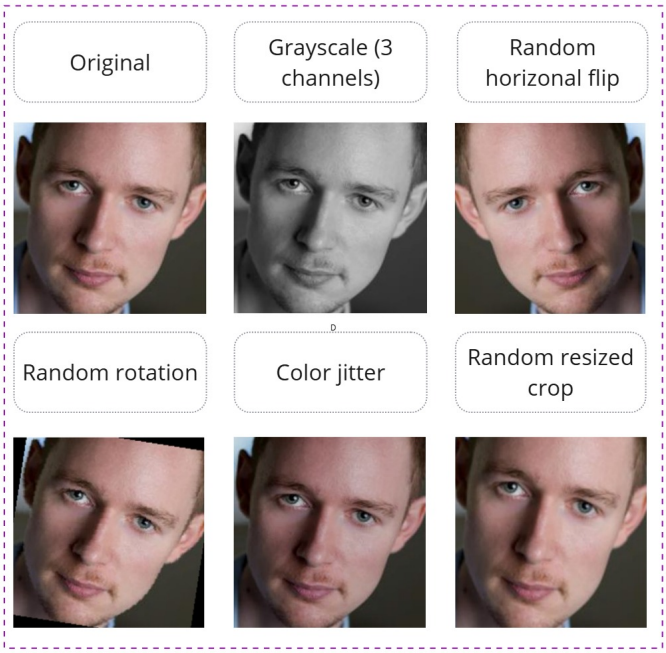

Figure 3. Data Augmentation

This involves a series of random conversions to increase the diversity of the images, without the need to collect new samples. This transformation includes grayscale conversion, defining a size of 224x224 pixels, rotations, flipping, cropping, color fluctuation, resizing, normalization, and conversion to tensors. As far as validation and testing are

concerned, a transformation similar to training is applied, except that the data augmentation is not placed.

The configuration and loading of the data consists of a staged organization of training/validation, testing and prediction, as shown in Figure 4.

Within the first stage, the images are uploaded by class and a percentage is divided for training and another for validation. The latter does not require an increase in data since it focuses on evaluating the performance of the model [16].

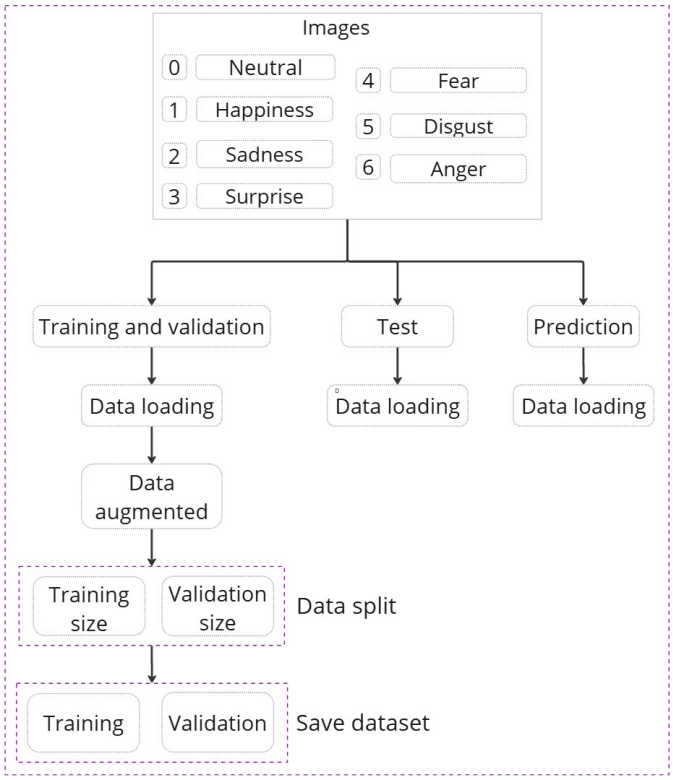

Figure 4. Data setup and upload

In the same way, data augmentation is not applied during testing, since it consists of evaluating the real performance of the model. As for the predictions, it ensures that the results of the images are consistent.

B. CNN finetuning

The finetuning of the CNN is based on a hidden MobileNetV2 structure and capability placement, defined through Figure 5.

The MobileNetV2 architecture is composed of several layers either convolutional complete with 32 filters, maximum groupings, and rectified linear unit (ReLU) activation functions [17].

Therefore, hidden layers are used to reduce dimensionalities, stabilize and improve speed during training, and introduce a non-linearity function with a low negative gradient, in order to reduce the problem of possible dead neurons [18],[19].

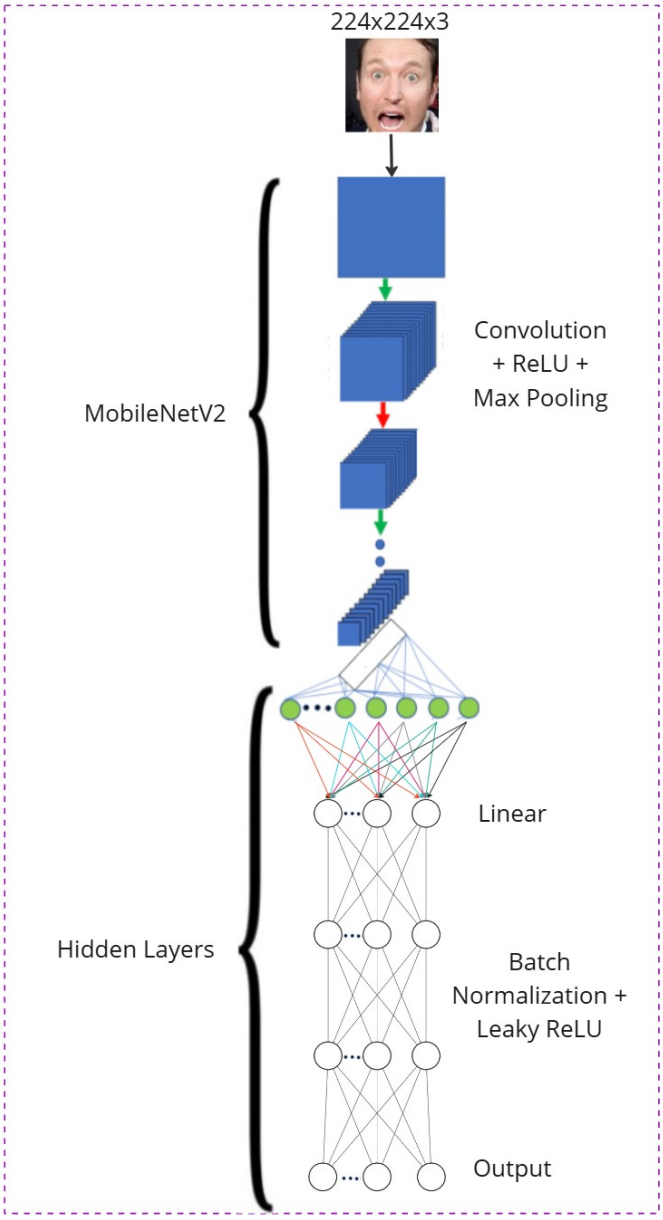

Figure 5. CNN

C. Model classification

At the exit of the hidden layers, there are 128 neurons that pass through classifiers based on emotions, as shown in Figure 6. The FE classifier contemplates 7 emotions such as:

- Joy.
- SadneFs.
- Anger.
- Disgust
- Fear.
- Neutral.
- Surprise.

Then, they are grouped into 3 classes, obtaining positive, negative and neutral emotions.

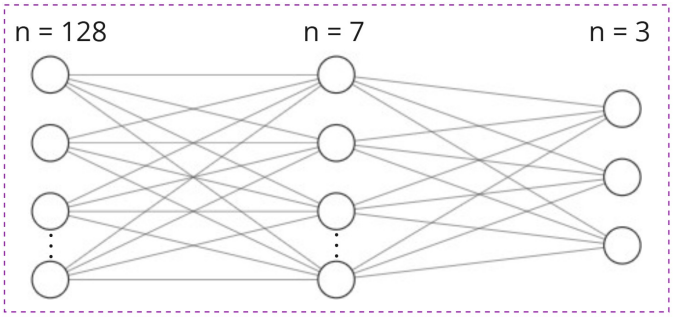

Figure 6. Model classification

For the tests, the detection and classification of EF is carried out from a video with MTCNN. This comprises 3 stages CNN:

- P-net: used to define bounding box regression windows and vectors to perform a preliminary detection where faces might be present.
- R-net: refines the P-net output proposal and performs calibration in the regression to eliminate false positives.
- O-net: similar to the previous network, except that it detects the key facial points for the eyes, nose and mouth.

Therefore, MTCNN can be described as a CNN framework that aims to detect and align faces (see Figure 7).

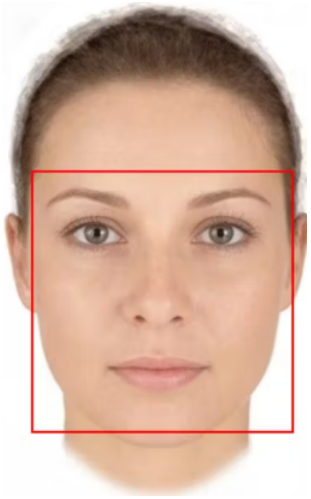

Figure 7. Face detection with MTCNN

Then, the video is uploaded and a resize is performed on the image. Thanks to this model, the type of FE is predicted, whether neutral, positive or negative.

For the implementation of face expression recognition model, the Python code available in this `repository`, which provides a practical and detailed approach to leveraging AffectNet for facial expression analysis.

IV. Results and discussion

According to the selected database, each type of emotion includes between 3803 and 134415 images, seen in Figure 8.

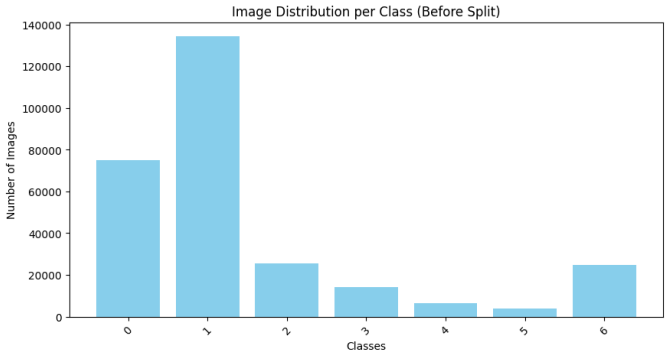

Figure 8. Initial database per class

Expressed as a percentage, 1,3%, 2,2%, 5%, 8,8%, 9%, 26,4% and 47,3% of FE are obtained such as disgust, fear, surprise, anger, sadness, neutral and happiness; respectively, as shown in Figure 9. This implies that there is a majority in the number of images that express joy, with the least amount being the emotion called disgust.

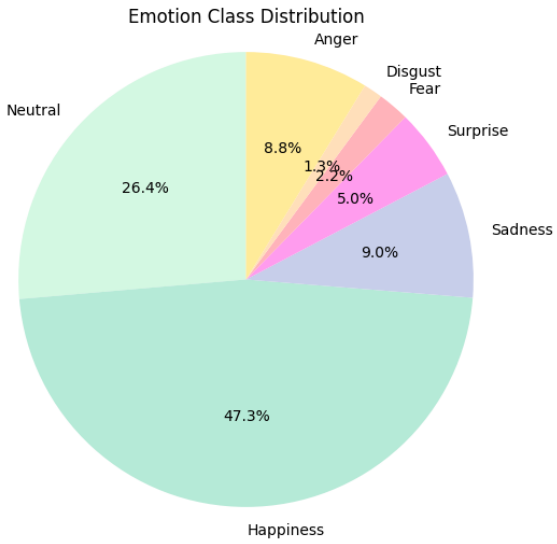

Figure 9. Emotion class distribution

Of the images, 80% and 20% are chosen for training and validation; respectively, for the proposed CNN. Taking as a starting point, hyperparameters are configured to balance effectiveness and efficiency during training. Among them, the following hyperparameters are defined:

- A learning rate equal to 0,001.
- A batch of 32 implies a balance between training speed and GPU memory usage.
- Approximately 50 epochs.
- 4 threads are set for parallel processing, so that data processing can be accelerated.

- 7 classes are defined, mentioned above.
- Differentiated learning rates are applied, for regularization and to avoid overadjustment; where there is a rate of 1e-5 for the base layer and the head of the proposed model with a value of 0,001. In addition, a weight decay of 0,01 is used in the trainable layers.
- Finally, the adjustment of the model parameters is done through the Adam optimizer [20]. For each epoch, the loss function is sought to be reduced. Figure 10 shows values equal to 0.68 in training and 0.63 for validation.

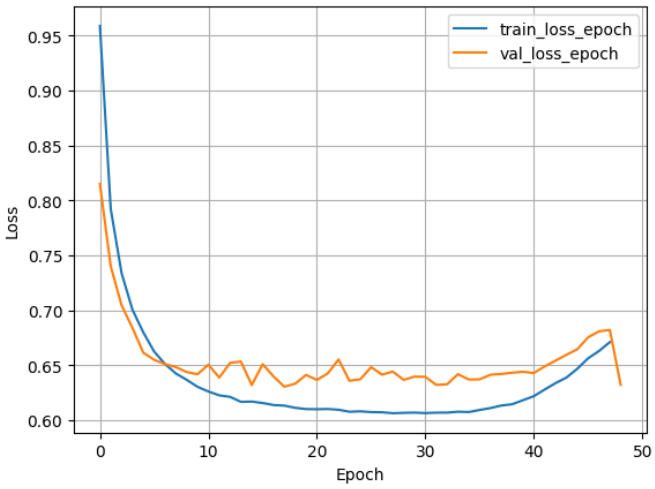

Figure 10. Loss training and validation – 7 class

Figure 11 shows how the accuracy of the training and validation of the proposed CNN reaches values equal to 77,9% and 77,08%; respectively.

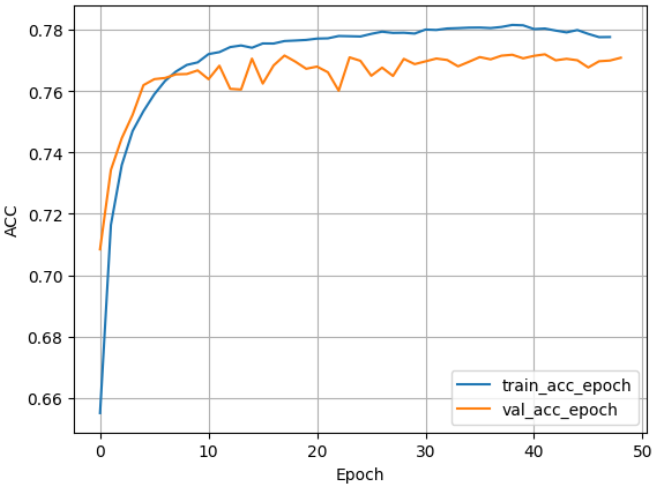

Figure 11. Accuracy training and validation – 7 class

Figure 12 shows the performance of the model according to the 7 classes mentioned above. The confusion matrix shows a representation of the true and predicted classes in the model. In addition, it indicates the predictions for each class. Neutral emotions represent the highest value in the prediction, that is, they have a value of 81,8%. However, the lowest prediction is 27,6% corresponding to disgust.

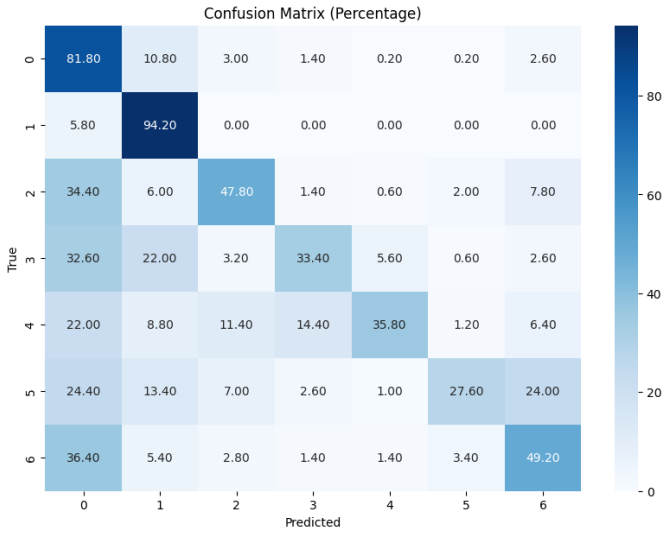

Figure 12. Confussion matrix CNN modified

There are certain emotions that have similar characteristics, so the model can confuse and represent an image that is not in accordance with the type of FAITH, as happens with disgust, fear, and sadness. On the other hand, expressions such as joy and neutrality are far from the others, as can be seen in Figure 13.

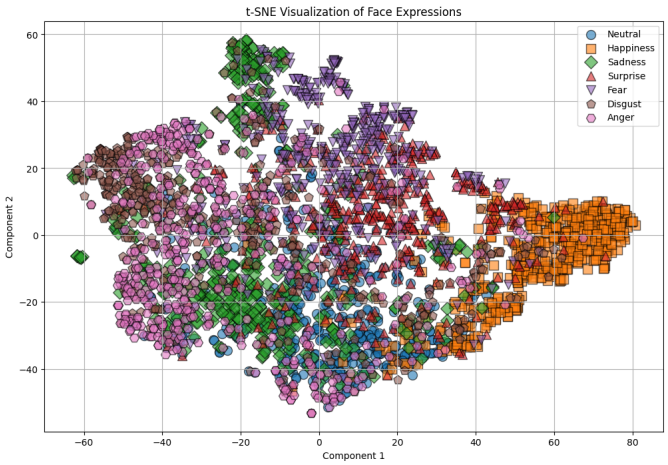

Figure 13. Visualization of FE by 7 classes

From this, a grouping of three classes is made as follows:

- Neutral:
  - Neutral.
  - Surprise.
- Negative:
  - Sadness.
  - Fear.
  - Disgust.
  - Angry.
- Positive: happiness.

In this way, predictions of 57,35%, 74,6% and 94,2% of the neutral, positive and negative emotions shown in Figure 14 are obtained.

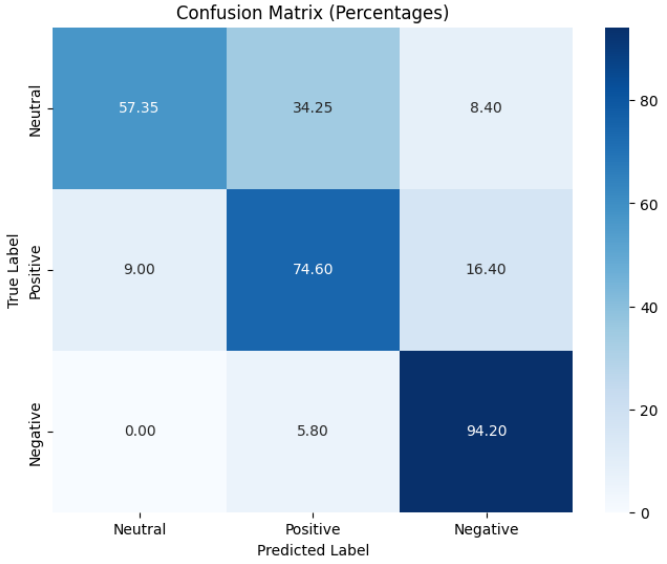

Figure 14. Percentage confussion matrix – 3 classes

Figure 15 shows how the group of expressions are well differentiated, especially with negative and positive emotions.

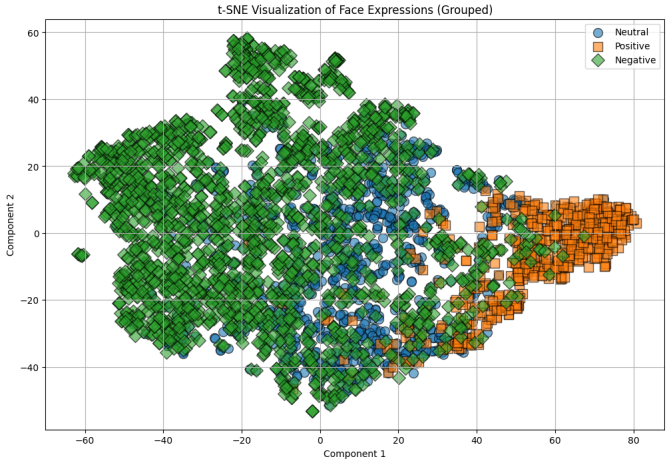

Figure 15. Visualization of FE by 3 classes

Figure 16 shows the evolution of each detected image in a period of less than 20 seconds. That implies that 2 to 4 seconds shows a different emotion.

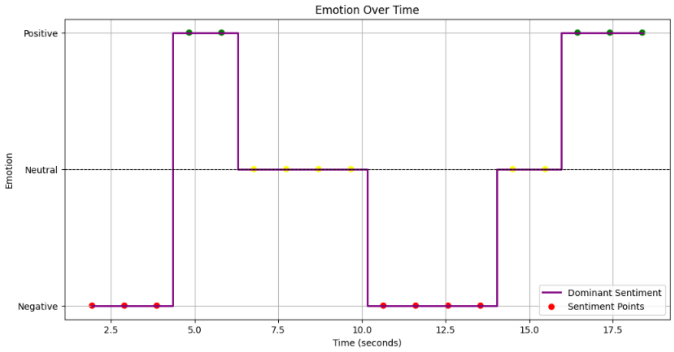

Figure 16. Emotions probability

Some studies revealed that the CNN models comprise a dataset not only from AffectNet, but from others observed in Table I. As is the case of Kuruvayil and Palaniswamy [9] who use the CMU Multi PIE for training and network validation; meanwhile in the tests they are handled with part of the CMU Multi PIE and AffectNet. On the other hand, Wang et al. [11] and Ullah [12] use several datasets but independently during each TVT process. However, there are authors such as Magherini [13], Shomoye and Zhao [14], and the proposed CNN that through AffectNet carry out the network model. However, the latter 2 determine the RES through real-time videos.

Table I
DATABASE TVT

Database	**[9]**	**[11]**	**[12]**	**[13]**	**[14]**	**CNN**
AffectNet	X	X	X	X	X	X
CMU Multi PIE	X					
RAF-DB		X	X			
FERPlus		X				
SFEW2.0		X				
Videos					X	X

With respect to CNN architectures, it can be seen from Table II that most authors employ a specific network, either ResNet or MobileNet V2. However, there are other authors such as Ullah et al. combine several architectures with the purpose of increasing precision in training.

Table II
CNN ARCHITECTURES

CNN	**[9]**	**[11]**	**[12]**	**[13]**	**[14]**	**CNN**
ResNet	X		X	X		
GCANet		X				
VGG-19			X			
Inception-V3			X	X		
MobileNetV2					X	X

Based on the hyperparameters and CNNs established, the accuracies in the training and validation are obtained, as indicated in Table III. However, the accuracy of the tests is also shown. It is observed that some precisions are greater than those mentioned by Mobilenet, this is because it uses different architectures combined, so it requires more processing.

Table III
TVT Percentage

Sources	Training	Validation	Test
[9]	90%	N.M.	68%
[11]	61%	N.M.	N.M.
[12]	89%	N.M.	89%
[13]	96,9%	96,7%	96,9%
[14]	77%	N.M.	N.M.
CNN base	77,9%	77,08%	75,38%

Where N.M. means "does not mention".

Table IV shows various FE with which each model of the network was determined. Some models were grouping the FE to recognize each emotion.

Table IV
Class by FER - AffectNet

CNN	**[9]**	**[11]**	**[12]**	**[13]**	**[14]**	**CNN**
Neutral	X	X	X	X	X	
Happiness	X	X	X	X	X	X
Sadness			X			
Surprise	X		X	X		
Fear					X	X
Disgust	X				X	X
Anger	X				X	X
Neutral 1						X
Negative						X
Surprise-fear				X		
Disgust-anger				X		

Where neutral 1 (neutral and surprise), negative (sad, fear, disgust and anger).

V. Conclusions

This study confirmed the efficiency of fine-tuning MobileNetV2 model for facial expression recognition, benefiting from pre-trained networks for faster learning and reducing the need for extensive computational resources. Adapting these models to the AffectNet dataset helped address the specific emotion classification task.

The model's accuracy improved by grouping emotions into broader categories like Positive, Negative, and Neutral. This simplification reduced misclassifications, enhancing the model's ability to capture core emotional states.

Hyperparameter tuning further optimized training, stabilizing learning while preventing overfitting. This allowed the pre-trained layers to retain their generalization capabilities.

Finally, MTCNN was used for video face detection, enabling the creation of emotional timelines, which is valuable for applications in human-computer interaction and affective computing, offering insights into emotional dynamics across a video.

Acknowledgment

A heartfelt thank you to A. Mollahosseini, B. Hasani, and M. H. Mahoor for their incredible work in creating and providing the AffectNet dataset. Your efforts have made a significant contribution to advancing research in emotion recognition and affective computing. This invaluable resource enables researchers and practitioners to explore and develop innovative solutions in the field. Thank you for your dedication and generosity in sharing your work with the community!

References

[1] M. Kaur and M. Kumar, "Facial emotion recognition: A comprehensive review," *Expert Systems*, vol. 41, no. 10, p. e13670, 2024, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/exsy.13670. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/exsy.13670

[2] Q. Zhu, H. Zhuang, M. Zhao, S. Xu, and R. Meng, "A study on expression recognition based on improved mobilenetV2 network," *Scientific Reports*, vol. 14, no. 1, p. 8121, Apr. 2024, publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/s41598-024-58736-x

[3] G. Pei, H. Li, Y. Lu, Y. Wang, S. Hua, and T. Li, "Affective Computing: Recent Advances, Challenges, and Future Trends," *Intelligent Computing*, vol. 3, p. 0076, Jan. 2024, publisher: American Association for the Advancement of Science. [Online]. Available: https://spj.science.org/doi/10.34133/icomputing.0076

[4] G. K. Kaur, A. Seram, K. Ansari, D. Patel, P. Singh, and S. Singla, "Facial Emotion Recognition Through Quantum Machine learning," in *2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS)*, vol. 1, Apr. 2024, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/document/10616819

[5] H. Kumar, "Facial Emotion Recognition and Detection Using Convolutional Neural Networks," *International Journal for Research in Applied Science and Engineering Technology*, vol. 12, no. 5, pp. 4690–4693, May 2024. [Online]. Available: https://www.ijraset.com/best-journal/facial-emotion-recognition-and-detection-using-convolutional-neural-networks-279

[6] M. Huh, B. Cheung, J. Bernstein, P. Isola, and P. Agrawal, "Training Neural Networks from Scratch with Parallel Low-Rank Adapters," Jul. 2024, arXiv:2402.16828. [Online]. Available: http://arxiv.org/abs/2402.16828

[7] P. T. Huong, L. T. Hien, N. M. Son, and T. Q. Nguyen, "Enhancing deep convolutional neural network models for orange quality classification using MobileNetV2 and data augmentation techniques," Jul. 2024, iSSN: 2693-5015. [Online]. Available: https://www.researchsquare.com/article/rs-4641084/v1

[8] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild," *IEEE Transactions on Affective Computing*, vol. 10, no. 1, pp. 18–31, Jan. 2019, conference Name: IEEE Transactions on Affective Computing. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8013713

[9] S. Kuruvayil and S. Palaniswamy, "Emotion recognition from facial images with simultaneous occlusion, pose and illumination variations using meta-learning," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 9, pp. 7271–7282, Oct. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1319157821001452

[10] W. Mellouk and W. Handouzi, "Facial emotion recognition using deep learning: review and insights," *Procedia Computer Science*, vol. 175, pp. 689–694, Jan. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050920318019

[11] S. Wang, A. Zhao, C. Lai, Q. Zhang, D. Li, Y. Gao, L. Dong, and X. Wang, "GCANet: Geometry cues-aware facial expression recognition based on graph convolutional networks," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 7, p. 101605, Jul. 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1319157823001593

[12] Z. Ullah, M. Mohmand, S. Rehman, M. Zubair, M. Driss, W. Boulila, R. Sheikh, and I. Alwawi, "Emotion Recognition from Occluded Facial Images Using Deep Ensemble Model," *Computers, Materials & Continua*, vol. 73, no. 3, pp. 4465–4487, 2022, publisher: Tech Science Press. [Online]. Available: https://www.techscience.com/cmc/v73n3/49024

[13] R. Magherini, E. Mussi, M. Servi, and Y. Volpe, "Emotion recognition in the times of COVID19: Coping with face masks," *Intelligent Systems with Applications*, vol. 15, p. 200094, Sep. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2667305322000345

[14] M. Shomoye and R. Zhao, "Automated emotion recognition of students in virtual reality classrooms," *Computers & Education: X Reality*, vol. 5, p. 100082, Dec. 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2949678024000321

[15] K. Sawarkar, *Deep Learning with PyTorch Lightning: Swiftly build high-performance Artificial Intelligence (AI) models using Python*. Packt Publishing Ltd, Apr. 2022, google-Books-ID: SfJWEAAAQBAJ.

[16] G. Iyengar, H. Lam, and T. Wang, "Is Cross-Validation the Gold Standard to Evaluate Model Performance?" Aug. 2024, arXiv:2407.02754. [Online]. Available: http://arxiv.org/abs/2407.02754

[17] M. Aparna and B. S. Rao, "A novel automated deep learning approach for Alzheimer's disease classification," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 12, no. 1, pp. 451–458, Mar. 2023, number: 1. [Online]. Available: https://ijai.iaescore.com/index.php/IJAI/article/view/21922

[18] M. Davel, M. Theunissen, A. Pretorius, and E. Barnard, "DNNs as Layers of Cooperating Classifiers," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, pp. 3725–3732, Apr. 2020, number: 04. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/5782

[19] J. Xu, Z. Li, B. Du, M. Zhang, and J. Liu, "Reluplex made more practical: Leaky ReLU," in *2020 IEEE Symposium on Computers and Communications (ISCC)*, Jul. 2020, pp. 1–7, iSSN: 2642-7389. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9219587

[20] J. J. Cuevas, E. P. Martínez, J. d. J. G. Cortés, S. S. Pérez, and J. A. C. Campos, "Comparativa de desempeño de los optimizadores Adam vs SGD en el entrenamiento de redes neuronales convolucionales para la clasificación de imágenes ECG (comparative performance of Adam vs. SGD optimizers in convolutional neural network training for the classification of ECG images)," *Pistas Educativas*, vol. 42, no. 137, Nov. 2020, number: 137. [Online]. Available: https://pistaseducativas.celaya.tecnm.mx/index.php/pistas/article/view/2300