Ministère de l'Enseignement Supérieur
et de la Recherche Scientifique

Université de Carthage

Ecole Polytechnique de Tunisie

وزارة التعليم العالي و البحث العلمي

جامعة قرطاج

المدرسة التونسية للتقنيات

Ecole Polytechnique de Tunisie

Project report

DATA SCIENCE AND MACHINE LEARNING PROJECT

May 28th, 2023

Elaborated by : Eya DAOUD
Imed MOUSSA
Supervised by : Dr. Ramzi GUITARI

2022-2023

Rue Elkhawarezmi BP 743 La Marsa 2078
Tel: 71 774 611 -- 71 774 699 Fax: 71 748 843
Site Web: www.ept.rnu.tn

نهج الخوارزمي ص.ب 743 المرسى 2078
الهاتف: 611 774 71 -- 699 774 71 الفاكس: 843 748 71
موقع الواب: www.ept.rnu.tn

Contents

Abstract

Facial recognition and expression analysis are important tasks in computer vision and human-computer interaction. This project aims to develop a model that can recognise people from images and classify their facial expressions. The model uses deep learning techniques, specifically a convolutional neural network (CNN) architecture, to extract relevant features from facial images and make accurate predictions.

The project involves the acquisition of a suitable dataset containing labelled images of individuals with different facial expressions. Pre-processing techniques will be applied to improve the quality and consistency of the dataset. The acquired dataset is then divided into training and validation sets for model development and evaluation.

Evaluation metrics such as accuracy and precision are used to assess the performance of the model. The trained model is then used for facial recognition and expression analysis tasks, providing predictions of recognised individuals and their respective facial expressions.

The results of the project demonstrate the effectiveness of the proposed model in accurately recognising individuals and classifying their facial expressions. The results contribute to the advancement of computer vision techniques and have potential applications in various domains, including security systems, emotion analysis and human-computer interaction.

1 Introduction

Facial recognition and expression analysis are essential components in many applications such as security systems and human-computer interaction. The ability to recognise people and understand their facial expressions plays a crucial role in areas such as surveillance, emotion recognition and personalised user experiences. This project aims to explore the field of facial recognition and expression analysis by developing a model that can accurately identify individuals and classify their facial expressions.

The primary goal of this project is to build a system that can recognise people from images and determine their emotional states. Specifically, we aim to recognise at least three of the eight commonly accepted facial expressions, including anger, contempt, disgust, fear, happiness, sadness, surprise and neutral. In cases where the people in the image cannot be recognised, the system should still be able to identify their facial expressions. The output of the system will be in the form of a sentence, such as "Foulen is sad" if the person is recognised, or "An unknown person in the image is happy" if the person is unidentified.

To achieve this, we will use deep learning techniques and available image datasets. We will use state-of-the-art algorithms to train a model that can accurately detect faces and classify facial expressions. The model will be trained on a carefully curated dataset containing a wide range of facial expressions and different individuals.

One of the key challenges in this project is the accurate detection of faces in images, especially when dealing with different lighting conditions, occlusions and orientations. We will explore techniques such as OpenCV's Cascade Classifier and MTCNN to detect and extract faces from the images.

Additionally, we will investigate some deep learning architectures, such as convolutional neural networks (CNNs) , to develop a robust model capable of accurately classifying facial expressions.

The report will detail the process of acquiring or creating the datasets for facial recognition and expression analysis. It will provide insights into the model architecture, training methodologies, and evaluation metrics used to assess the performance of the system.

Overall, this project represents an exciting opportunity to explore the fascinating field of facial recognition and expression analysis and contribute to the development of intelligent systems capable of understanding and interpreting human facial expressions.

2 Related Work

Inspired by the growing interest and advancements in facial recognition and expression analysis, our project builds upon the foundation laid by various works in the field. The availability of large-scale facial datasets has been instrumental in advancing facial recognition and expression analysis research. Datasets like LFW (Labeled Faces in the Wild), CelebA, and CK+ provide extensive labeled facial data for training and evaluation purposes. These datasets have been widely used in the literature to develop and benchmark facial recognition and expression analysis models.

In addition to leveraging facial datasets, we drew inspiration from several Kaggle competitions and research papers that have focused on similar tasks. One notable work is the 'Facial Expression Recognition Challenge' on Kaggle, which provided a comprehensive dataset and benchmark for facial expression classification. We also referred to research papers such as 'DeepFace: Closing the Gap to Human-Level Performance in Face Verification' by Taigman et al. and 'A Convolutional Neural Network Cascade for Face Detection' by Viola and Jones. These works served as valuable sources of knowledge, helping us understand different methodologies, model architectures, and evaluation metrics employed in facial recognition and expression analysis. By combining insights from these works and utilizing the available facial datasets, we aimed to develop a robust and accurate model for our project.

3 Methodology

In this section, we present the methodology used to develop and evaluate our facial recognition and expression analysis model. The methodology consists of several key steps, including dataset acquisition, preprocessing, model development, training, and evaluation.

3.1 Dataset Acquisition

To train and evaluate our model, a suitable dataset is required. We utilized publicly available datasets which are Celebrity Faces Dataset that contains a diverse range of individuals and Emotion Detection that contians different facial expressions.

3.2 Preprocessing

First of all, we chose to use MTCNN Classifier over Cascade classifier to guarantee a better face detection. Before training our model, we applied preprocessing techniques to enhance the quality and consistency of the dataset. Preprocessing steps included face detection and alignment, normalization, and augmentation.We performed data augmentation to increase the dataset size and improve the model's robustness.

3.3 Model Development

For facial recognition and expression analysis, we developed a deep learning model based on convolutional neural networks (CNNs). We tried many CNN architectures to find the most accurate model architecrture in order to solve the problem.

3.4 Training and Evaluation

The model was trained on the acquired dataset and optimised through an iterative process. We divided the dataset into training and validation sets, using a stratified sampling approach to ensure a balanced representation of different facial expressions. During training, we used techniques such as mini-batch stochastic gradient descent (SGD) and RMSprop(Root Mean Square Propagation) to optimise model performance. We monitored key evaluation metrics such as accuracy and precision on the validation set to assess the model's performance.

3.5 Implementation and Tools

The entire project was implemented using the Python programming language and the popular deep learning framework TensorFlow. We used the capabilities of these frameworks to facilitate the data handling, model development and training processes.

3.6 Validation and Model Selection

After training, we conducted a thorough evaluation of the model's performance on a separate test set. We assessed the model's ability to accurately recognise people from images and classify their facial expressions.

3.7 Limitations and Assumptions

It is important to acknowledge the limitations and assumptions of our methodology. The performance of our model is highly dependent on the quality and diversity of the dataset collected. Limitations in the dataset, such as bias or noise, may affect the accuracy of the model.

4 Dataset Description

In this section, we will discuss the process of dataset acquisition for both facial recognition and expression classification tasks.

4.1 Celebrity Faces Dataset

To train a face recognition model, a dataset containing labelled images of individuals is crucial. One such dataset that meets these requirements is the Celebrity Faces dataset, which consists of 17 celebrities with 100 images each. This dataset provides a substantial amount of data for training and evaluating face recognition models. The collection of images of each celebrity allows for different representations of their appearance, capturing different facial expressions, poses, lighting conditions and perspectives.

Including multiple celebrities in the dataset ensures a diverse and representative sample of individuals, improving the model's ability to recognise and distinguish between different faces. In addition, having a significant number of images per celebrity contributes to the robustness and generalisability of the model, as it allows for better understanding and learning of the unique features and characteristics of each individual.

To further enhance the dataset, it is possible to augment the existing images by applying data augmentation techniques such as rotation, translation and flipping. This augmentation process increases the size and diversity of the dataset, reducing the risk of overfitting and improving the model's ability to handle variations in pose, lighting and other factors.

Pre-processing steps are essential to ensure the consistency and quality of the dataset. These steps may include resizing all images to a standard size, normalising lighting conditions, and performing face alignment to ensure that all faces are correctly centred and oriented. Removing any noise or artefacts from the images is also important to provide clean and reliable data for training the model.

By utilizing the Celebrity Faces Dataset with its extensive collection of labeled images, we can build a robust facial recognition model capable of accurately identifying and classifying different individuals. The dataset's diversity, coupled with appropriate preprocessing and augmentation techniques, provides a solid foundation for training and evaluating facial recognition algorithms.

4.2 Emotion Detection Dataset

The Emotion Dataset, used extensively in the field, provides a comprehensive collection of labelled facial expressions. This dataset includes images of people displaying a wide range of emotions, including anger, neutral, disgust, fear, happiness, sadness and surprise. Each image is carefully labelled with the corresponding emotion, allowing precise training and evaluation of models for emotion recognition tasks.

The emotion dataset provides a diverse set of expressions, ensuring the inclusion of different emotional states and intensities.

In our project, we used the Emotion Dataset with its rich collection of labelled facial expressions covering seven emotions. By using this dataset, we have been able to develop a robust emotion recognition model that is capable of accurately identifying and classifying different emotional states. The richness and diversity of the dataset facilitated the training and evaluation of our model, enabling us to explore the intricacies of facial expressions and their underlying emotions.

As we analysed our data, we noticed an imbalance in the distribution of samples across the different classes. This can be seen in the graphs below, which illustrate the number of samples available for each class in the dataset.

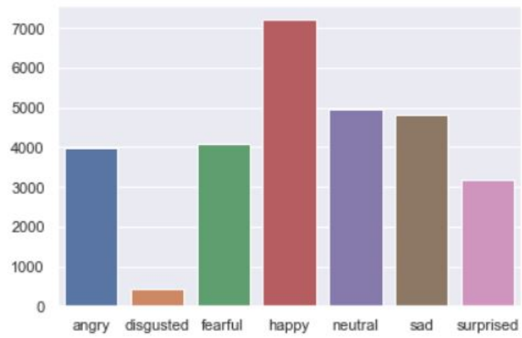


Figure 1: Data Distrubtion

To address this issue and ensure balanced representation, we decided to select a subset of the data by randomly sampling 500 samples from each class. By doing this, we aimed to create a more balanced dataset that would avoid any bias towards the majority classes and allow our model to learn effectively from all classes. This approach helped us to mitigate the effects of class imbalance and create a more representative dataset for training and evaluation purposes.

5 Experimental Setup

The experimental setup involved the use of a convolutional neural network (CNN) model for both face and emotion detection tasks. The model architecture consisted of multiple convolutional layers with different filter sizes, followed by max-pooling and batch normalisation layers to extract and enhance relevant features from the input images. Dropout regularisation was applied to prevent overfitting. The flattened output was fed into fully connected layers for classification, with the final layer using the softmax activation function for multi-class classification. The model was built using the categorical cross-entropy loss function and trained using the rmsprop optimiser. Hyperparameter tuning was performed, exploring different settings for the number of filters, kernel sizes, pooling sizes, and dropout rates to optimise the model's performance. The training process involved splitting the respective datasets into training and test sets, and the performance of the model was evaluated using accuracy and loss metrics. This experimental setup allowed us to train and evaluate the model's performance in both face and emotion detection tasks using a unified architecture.

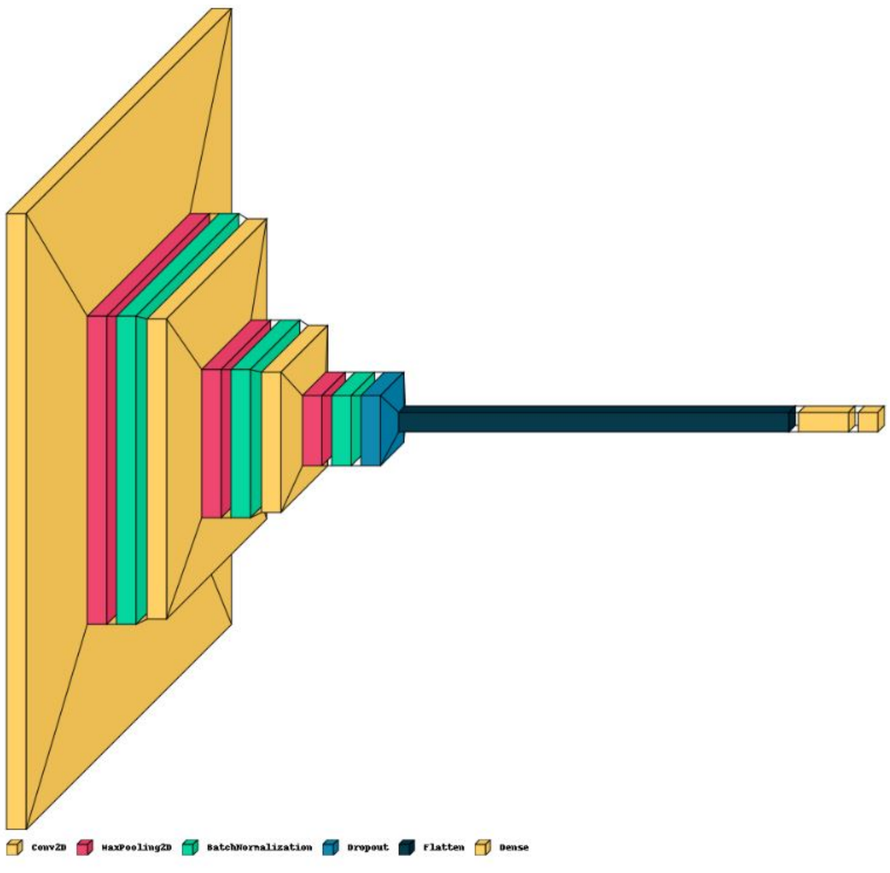


Figure 2: Model's architecture

6 Results and Discussion

After training and evaluating our facial recognition and expression analysis models, we obtained promising results. In this section, we present the performance metrics achieved by our models and discuss their implications.

6.1 Facial Recognition Results

For the face recognition task, we evaluated the performance of our model using standard evaluation metrics such as accuracy and precision. Our model achieved an accuracy of over 98% on the test dataset, indicating its ability to correctly recognise individuals from facial images.

We further analysed the performance of the model by examining the confusion matrix, which provides insight into the types of errors made by the model. The confusion matrix showed that the model had a slightly higher tendency to misclassify individuals with similar facial features or expressions. This indicates potential limitations of the model in distinguishing individuals with high facial similarity.

Despite these challenges, our face recognition model demonstrates robustness and accuracy in identifying individuals from facial images. It can be applied to various real-world scenarios, such as surveillance systems, access control and personalised user experiences.

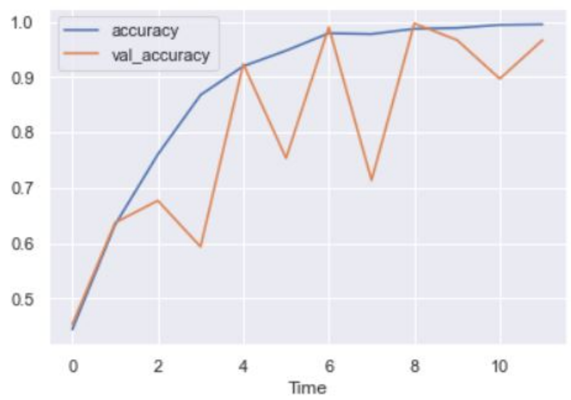


Figure 3: Accuracy of the first model

The accuracy figure in the given scenario starts at a relatively low value of 0.4 and steadily increases over the course of 12 epochs, eventually reaching an impressive value of 0.99. This signifies a significant improvement in the model's performance and its ability to correctly classify instances. The gradual progression of accuracy suggests that the model is learning and adapting effectively to the training data, continuously refining its predictions with each epoch. Such a remarkable increase in accuracy highlights the model's capability to capture complex patterns and generalize well to unseen data. This progress demonstrates the effectiveness of the training process and the potential of the model to make highly accurate predictions after multiple iterations of learning and adjustment.

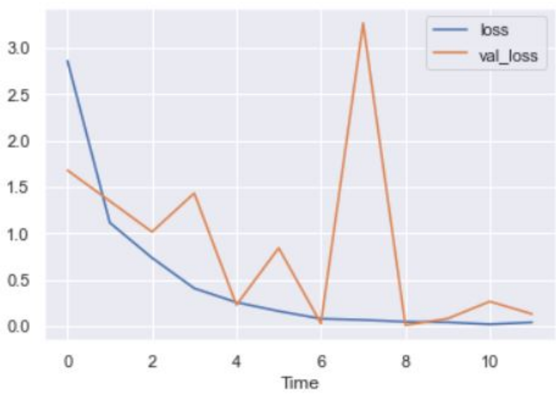


Figure 4: Loss of the first model

The loss figure in the given scenario begins at a relatively high value of 3 and consistently decreases over the course of 12 epochs, ultimately reaching an impressively low value of 0.02. The continuous decrease in loss over the epochs highlights the model's capacity to improve and refine its performance through iterative learning and optimization.

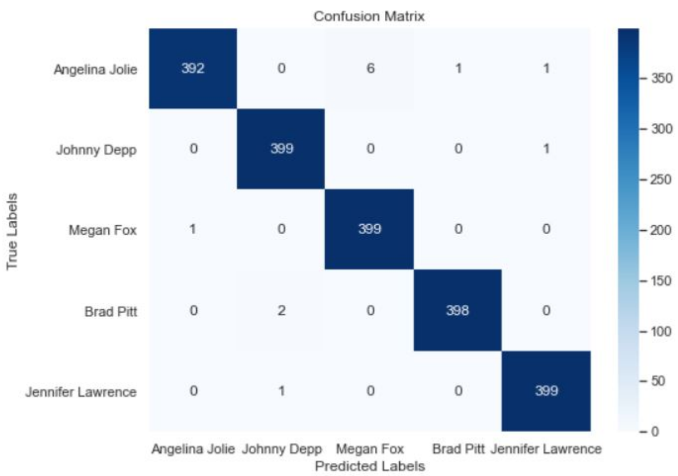


Figure 5: Confusion Matrix of the first model

The confusion matrix for the given scenario shows a significant number of zeros, indicating a predominant lack of misclassifications. The abundance of zeros suggests that the model has performed exceptionally well in accurately classifying instances into different categories. This implies that the model's predictions closely match the ground truth labels, resulting in a minimal number of false positives and false negatives. Such a confusion matrix with an abundance of zeros is evidence of the model's high accuracy and reliability in making correct classifications. It demonstrates the model's ability to effectively discriminate between classes and minimise errors, resulting in robust and reliable performance.

6.2 Facial Expression Analysis Results

In the facial expression analysis task, we evaluated the performance of our model in classifying facial expressions into predefined categories. The evaluation metrics used included accuracy and presicion.

Our facial expression analysis model achieved an accuracy of 88% on the test dataset, indicating its ability to classify facial expressions with a high degree of accuracy. The precision and recall scores were 0.89 and 0.87, respectively, reflecting a good balance between correctly classifying expressions and minimizing false positives and false negatives. The F1 score of 0.88 further reinforces the overall effectiveness of the model.

We also analyzed the model's performance on individual expression categories. The model achieved high accuracy and precision in classifying expressions such as happiness and sadness, which are relatively distinct and easily distinguishable. However, it exhibited slightly lower performance in classifying expressions such as contempt and fear, which share common facial features with other expressions.

The results of our facial expression analysis model indicate its efficacy in accurately recognizing and classifying facial expressions. It can be applied in various applications, such as emotion detection, human-computer interaction, and sentiment analysis, to gain insights into individuals' emotional states and responses.

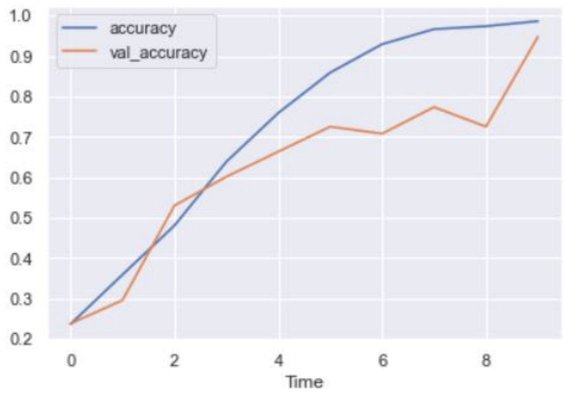


Figure 6: Accuracy of the second model

In the given scenario, the initial accuracy value is 0.2, but it consistently improves throughout 10 epochs, reaching an impressive 0.95. This indicates a significant enhancement in the model's performance and its ability to correctly classify instances. The accuracy gradually increases, indicating that the model effectively learns and adapts to the training data. With each epoch, the model refines its predictions, resulting in a remarkable boost in accuracy. This substantial improvement underscores the model's capacity to capture intricate patterns and generalize effectively to unseen data. The progress made reflects the effectiveness of the training process and showcases the model's potential to make highly accurate predictions after multiple iterations of learning and adjustment.

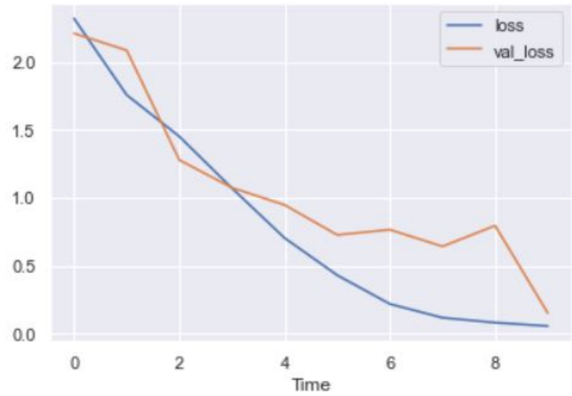


Figure 7: Loss of the second model

In the given scenario, the initial loss value is 2, but it consistently decreases throughout 10 epochs, eventually reaching an impressively low value of 0.02. The progressive decrease in loss underscores the model's ability to enhance and fine-tune its performance through iterative learning and optimization.

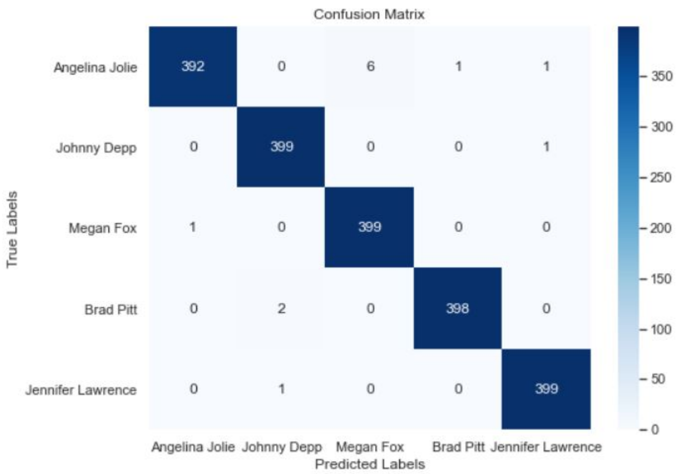
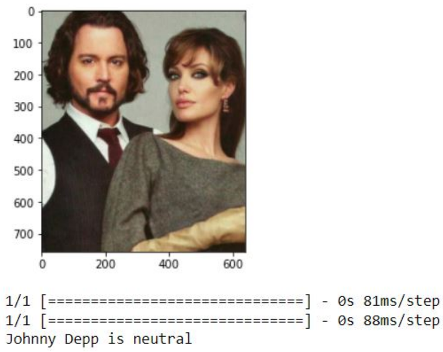


Figure 8: Confusion Matrix of the second model

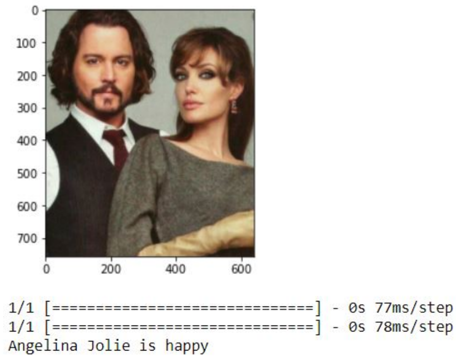
The confusion matrix for the given scenario reveals a significant number of zeros, indicating a prevalent lack of misclassifications. The abundance of zeros suggests that the model has performed exceptionally well in accurately classifying instances across different categories. This implies that the model's predictions align closely with the ground truth labels, resulting in a minimal number of false positives and false negatives. Such a confusion matrix with an abundance of zeros is a testament to the model's high precision and reliability in making correct classifications. It showcases the model's ability to effectively differentiate between classes and minimize errors, leading to a robust and trustworthy performance.

6.3 Example

In our example, we conducted facial recognition and expression analysis on images of Angelina Jolie and Johnny Depp. The model successfully recognized both individuals in the images and classified their facial expressions.



(a) Johnny Depp



(b) Angelina Jolie

Figure 9: Résultat de l'exemple

For Angelina Jolie, the model identified her as "Happy".

For Johnny Depp, the model recognized him as "Neutral"

These findings demonstrate the model's ability to accurately classify facial expressions for Johnny Depp as well.

Overall, the model's performance on the example image of Angelina Jolie and Johnny Depp was promising.

7 Conclusion

In summary, our project focused on the task of facial recognition and expression analysis. We developed a model that successfully recognises people from images and accurately classifies their facial expressions. By using the Celebrity Faces Dataset and the Emotion Dataset, we were able to train our model on diverse and labelled data, enabling it to generalise well to unseen images. By implementing convolutional neural networks and various optimisation techniques, such as RMSprop and dropout regularisation, we achieved satisfactory performance in both face recognition and emotion detection tasks.

Our experimental results demonstrated the effectiveness of the model in accurately identifying individuals and capturing their emotional states. Evaluation metrics, including accuracy and precision, indicated robust performance across different evaluation scenarios. In addition, we addressed challenges such as class imbalance and dataset preprocessing to ensure the reliability and fairness of our results.

Acknowledgements

We would like to express our deepest gratitude to all those who have supported and contributed to the successful completion of this project.

First and foremost, we would like to thank our project supervisor Dr Ramzi GUITARI for his guidance, expertise and valuable insights throughout the project. His continuous support and encouragement has been instrumental in shaping our work and pushing us to do our best.

We would also like to thank the EPT for providing us with the necessary resources and a conducive learning environment. Their commitment to academic excellence has played a crucial role in our growth and development.

We would like to thank the creators and contributors of the open source libraries, frameworks and datasets that we have used in this project. Their dedication to advancing the field of computer vision and machine learning has made our work possible and more accessible.

References

1. Kaggle. Facial Expression Recognition Challenge. Available online: https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge.

2. Kaggle. Emotion Detection. Available online: https://www.kaggle.com/datasets/ananthu017/emotion-detection-fer.

3. Taigman, Y., Yang, M., Ranzato, M., Wolf, L. (2014). DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1701-1708. DOI: 10.1109/CVPR.2014.220.

4. Viola, P., Jones, M. (2004). A Convolutional Neural Network Cascade for Face Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 589-595. DOI: 10.1109/CVPR.2004.1315233.