COMP3419 Facial Attribute Analysis

SID: 520141575

***Abstract*—The objective of this assignment was to utilize deep learning architectures to create a model that can accurately describe facial attributes given a batch of images using the CelebA dataset provided. Additionally, optimizing through data augmentation and learning rates were introduced to produce a more robust and flexible model.**

I. Introduction

THE aim of this facial attribute analysis assignment is to develop a machine learning framework to classify various facial attributes for each face in the CelebA dataset. The model would need to be trained in order to accurately predict the binary labels for each image and indicate whether or not each attribute is present in the image provided.

The given CelebA dataset consists of 202,599 images labeled with 40 different facial attributes and labeled 1 or -1 depending on whether it is present or not.

For this project, there were various methods used in order to achieve the desired outcome. The use of Convolutional Neural Network (CNN) was used for extracting the features and the classifications of the 40 different facial attributes. Additionally frameworks such as PyTorch, PIL and pandas was used for model implementation, image processing and handling data respectively. Finally, random horizontal flipping was used as a form of data augmentation. This was applied to ensure the model was able to predict the necessary facial attributes despite the images being altered. In other words, improving the generalization of the model for a much more robust outcome.

II. Methods

This section will outline the methods used for this project, and explaining its use and roles within. The major methods and tools used were Convolutional Neural Networks (CNN), Data Augmentation, Loss Function and Optimization, Data Handling and Processing and the training process.

A. Convolutional Neural Network (CNN)

The Convolutional Neural Network is a type of deep learning model that can recognize various patterns and objects within an image, which is done so by applying grids (referred to as Kernels) over the image to analyze small patches at a time to recognize and identify patterns [1]. CNN was implemented to perform the classification of facial attributes for each image provided. The model consists of three convolutional layers with the Rectifier Linear Unit (ReLU) to extract features from the input images. After each layer, a max-pooling layer is used to reduce the size of data for optimization. The final layer then outputs the probabilities for each attributes post-extraction.

B. Data Augmentation

In order for the model to be more flexible, a random horizontal flipping was implemented. This was implemented for a more robust deep learning model that allows the model to recognize facial features despite some images being flipped, and thus handling more variations, and effectively generalizing the model for better precision.

C. Loss Function and Optimization

The model used the Binary Cross Entropy Loss (BCELoss) to measure its prediction of the facial attributes. This works by comparing the probabilities to the labels 0 and 1 for each attribute [2]. Additionally, the use of an Adam optimizer helped the model learn efficiently by adjusting how much it changes the model's weights during each step of the training.

D. Data Handling and Preprocessing

The CelebA dataset was split into training, validation and testing sets based on the partitions. Each image was resized to 128x128 pixels and then normalized to be between -1 and 1, allowing the model to train faster. The PIL library was used to load the images and pandas was used to manage the attribute labels.

E. Training

The model was trained for 5 epochs, referring to the number of times it passes through a specified batch of images. The model then made predictions, calculated the error using BCELoss and updates its weights. After each epoch, the performance of the model was checked on the validation set to ensure the model was learning correctly.

III. Experimental Setup

A. Dataset

The dataset used for this assignment was the CelebA dataset containing 202,599 images of celebrity faces, and each are annotated with 40 binary facial attributes. The dataset is also divided into training, validation and test sets based on the official partition (162,000+ images for training and 19,000+ for validation and testing). The images were resized to 128 pixels and normalized between the range of -1 to 1 to facilitate faster training.

B. Evaluation Metrics

The main evaluation metric for this was the accuracy in which the model predicts the attributes for each image. This was done by predicting the presence/absence of each attribute, and was considered correct if the model's probability was greater than 0.5. The final run on the test set was measured by computing the overall accuracy across all images.

IV. Results & Discussions

Method	Accuracy	Validation Loss	Training Loss
LR 0.001	83.50%	0.3756	0.3572
RHF	83.95%	0.3673	0.3490
LR 0.0005	84.03%	0.3642	0.3355
RHF	84.27%	0.3547	0.3469

The table above shows the four different runs that were tested with the model implemented. The model went through 1000 images over 5 runs for each method.

The first run was the base model with a learning rate of 0.001. It achieved a total test accuracy of 83.50%, which gives a solid starting point for the model to work with. In addition, there was a decrease in both training and validation loss throughout the five runs it made, suggesting that the model was learning over time.

The second run included the data augmentation, which in this case was the random horizontal flipping. This slightly improved the test accuracy by 0.45% and also lowered both the validation and training loss. This indicates that by generalizing the tests by including data augmentation, it helped the model to make better, more flexible predictions during its training process.

The third run was done by lowering the learning rate to 0.0005. This improved the test accuracy making it 84.03%, thus suggesting that with a slower learning rate, it allowed the model to undergo its training process in a more stable manner.

Finally, the fourth and last run was done by both lowering the learning rate to 0.0005 and implementing random horizontal flipping. The outcome was the most accurate out of the four with a higher test accuracy at 84.27% and a lower validation and training loss. This indicates that combining the data augmentation in addition to a slower learning rate allowed the model to both generalize and stabilize during its training process, thus yielding an accurate and robust result.

V. Conclusion

For this project, the Convolutional Neural Network was implemented to perform multi-label classification of facial attributes using the given dataset. Throughout the process, methods such as data augmentation and learning rates were adjusted in order to capture a variety of performances to observe the impact in which each method had.

The results showcased that, whilst the base method with a learning rate of 0.001 produced reasonable results, by implementing random horizontal flipping and allowing the model to run at a slower learning rate, enhanced the stability and reliability of the model's performance, thus increasing the overall test accuracy with a lower validation and training loss.

In general, by tuning the hyper-parameters and augmentation techniques, the model can be optimized for better accuracy in recognizing facial attributes in a given image. To optimize the model even further, future work could explore a variety of data augmentation techniques as well as more extensive models of architecture.

VI. References

References

[1] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: An overview and application in radiology," *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, 2018.

[2] R. Wali, "Xtreme margin: A tunable loss function for binary classification problems," *ArXiv*, 2022. Cornell University.