Progress Reporting: Group-G Age Detection Using Convolutional Neural Networks

1. Introduction and Problem Statement

Age classification is a critical task with applications in marketing and healthcare. The aim of this project is to classify facial images into different age groups using Convolutional Neural Networks (CNN). However, several challenges were encountered during the project. One of the primary challenges was the unavailability of pre-classified datasets. Consequently, the datasets had to be preprocessed by dividing them into classes using the file names which contain age. Another challenge was the manual removal of low-quality images from the dataset, which had a negative impact on the training process. Thus, it was necessary to remove these images to ensure the accuracy of the models. Additionally, difficulties were encountered in hyperparameter tuning, which did not yield satisfactory results despite trying various tuning techniques. To address these challenges, different models were compared based on test accuracy and confusion matrix. However, using transfer learning methods in the future, could potentially improve the accuracy of the age classification models. Our benchmark for maximum results in each dataset are as follows: Dataset-1 (APPA-REAL): achieved an accuracy of 89.17% [1] on the APPA-REAL dataset using a deep CNN architecture based on the VGG-16 model. Dataset-2 (UTKFace): achieved an accuracy of 97.06% [2] on the UTKFace dataset using a custom CNN architecture with four convolutional layers, followed by two fully connected layers and a regression layer. Dataset-3 (Adience): achieved an accuracy of 70.04% [3] on the Adience dataset for both age and gender classification using a deep CNN architecture based on the VGG-Face model. We will compare the performance of our model with these benchmarks to determine the effectiveness of our approach.

To acquire and evaluate the results of our CNN project, we plan to follow a systematic approach. First, we will train our model on a training dataset using the chosen architecture. We will then use a validation dataset to evaluate the performance of the model during the training process, adjust the hyperparameters of the model accordingly. We will also evaluate other performance metrics such as precision, recall, and F1 score to have a better understanding of the model's performance. We will use various visualization techniques such as confusion matrix, ROC curve to evaluate the model's performance comprehensively. In summary, this project aimed to classify facial images into different age groups using Convolutional Neural Networks(CNN). However, challenges were faced in terms of preprocessing the datasets, removing low-quality images, and hyperparameter tuning. Nonetheless, using transfer learning methods in the future could lead to better results.

2. Proposed Methodologies

To accomplish the task of age classification, the first step was to preprocess the data into different classes. We used the qcut method in pandas to obtain an almost equal number of images in each class. This process involved dividing the data into different age groups. This step was necessary as it allowed us to train the models effectively. This is the detail of our datasets:

APPA-REAL Dataset: This dataset is a collection of facial images of individuals from different age groups, races, and gender. It contains over 7,591 images of 2,000 subjects. The images were obtained from the APPA (Automated Pseudo-Photorealistic Portrait Generator) system, which was designed to generate facial images of different age groups based on statistical analysis of facial features. The images were generated under different conditions such as lighting, background, and pose to simulate real-world scenarios. The dataset also includes the subject's age, gender, and race information.

UTKFace Dataset: The UTKFace dataset is another widely used dataset for age classification. It consists of over 20,000 facial images of different individuals, with age ranging from 1 to 100 years old. The images were collected from publicly available sources, and each image is labeled with the subject's age, gender, and ethnicity. The dataset is challenging due to the large age range and variations in the quality of images, lighting conditions, and poses.

Adience Dataset: The Adience dataset is a collection of facial images for age and gender classification. It contains over 26,000 images of individuals of different ages, races, and gender. The images were collected from the internet and were carefully filtered to ensure that the subjects are looking directly at the camera, and their faces are visible. The dataset includes the subject's age, gender, and the estimated age based on their appearance. The dataset is challenging due to the variations in the quality of images, lighting conditions, and poses.

We trained multiple models on our datasets, including ResNet and VGG. Both of these models are well-suited for image classification tasks due to their deep neural network architectures. ResNet is a type of CNN that allows for the training of very deep neural networks, while VGG is known for its simplicity and effectiveness. These models were trained with different optimizers such as RMSprop, Adam, and SGD, with CrossEntropy as the loss function. In UTKFace dataset, we've got 78.74% accuracy with RESNET-18 with lrscheduler. For APPA-REAL dataset with couldn't get good result with RESNET-18 we've got 51.98% which is not satisfying, 71.96%. For Adience dataset, we've got 71.96% in base case with RESNET-18. To reduce overfitting of the model, we also used a learning rate scheduler.

This technique allowed us to gradually decrease the learning rate during training, which helped the model to converge faster and avoid getting stuck in local minima. The training process was repeated multiple times, and the models were compared based on their test accuracy and confusion matrix.

In the future, we plan to implement transfer learning methods on all three proposed CNN models, including ResNet18, VGG16, and MobileNetV2. Transfer learning is a technique that allows us to leverage pre-trained models on large datasets to improve the accuracy of our models on smaller datasets. By using transfer learning, we hope to improve the accuracy of our age classification models and understand which model architecture works best for this specific task. CNN Models used so far are VGG16 and RESNET: ResNet-18 and VGG-16 are popular convolutional neural network (CNN) architectures used for image classification tasks. ResNet-18 has 18 layers and introduces the concept of residual connections, allowing for easier training of deeper models. VGG-16, on the other hand, has 16 layers and is known for its use of small convolutional filters and max-pooling layers. While both models have achieved high performance in image classification benchmarks, ResNet-18 is generally preferred for more complex tasks, while VGG-16 is known for its simplicity and ease of implementation. RMSprop is an optimization algorithm commonly used in neural networks. It calculates an exponentially weighted average of the squared gradients to adjust the learning rate during training. This helps to prevent the learning rate from becoming too large and can lead to faster convergence and better performance.[4] SGD (Stochastic Gradient Descent) is a simple optimization algorithm used in deep learning. It updates the weights of the neural network based on the gradients of the loss function with respect to the weights.[5] Adam: Adam is an optimization algorithm commonly used in neural networks. It combines the advantages of both RMSprop and momentum optimization by calculating adaptive learning rates for each parameter and using momentum to accelerate convergence. It is known for its fast convergence and ability to handle sparse gradients, making it a popular choice for deep learning tasks.[6] Learning rate scheduler is a technique used to adjust the learning rate during training in deep learning models. It gradually decreases the learning rate to help the model converge more accurately and quickly. By using a learning rate scheduler in conjunction with optimization algorithms such as Adam and SGD, the performance of ResNet-18 and VGG-16 models can be significantly improved.

3. Attempts at solving the problem

We tested different CNN architectures on three datasets, and each model was trained for 50 epochs. With the RESNET-18 model, without LR- scheduler, we achieved 50.17, 78.74 and 71.96% accuracy on all three datasets respectively, with a validation accuracy of 35.84% on APPA-REAL, 68.19% on UTKFace, and 56.94% on Adience. With LR- scheduler, we achieved 51.98, 73.86 and 67.65% accuracy on all three datasets respectively, with a validation accuracy of 45.59% on APPA-REAL, 69.55% on UTKFace, and 50.05% on Adience.The results of VGG-16 model , with a validation accuracy of 68.3% on APPA-REAL, 76.1% on UTKFace, and 64.4% on Adience. We also tested two Transfer Learning models, which did not perform as well as the models we trained from scratch.

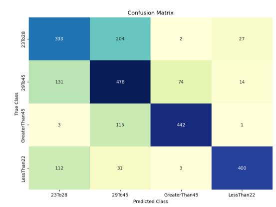

Figure 1. RESNET-18's Confusion matrix for UTKFace Dataset

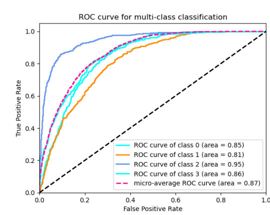

Figure 2. ROC Curve for UTKFace Dataset

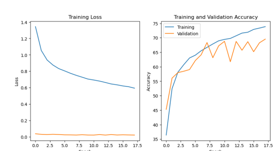

Figure 3. Train/Validation Accuracy and loss plot obtained for UTKFace Dataset

4. Future Improvements

To improve the accuracy of the model, we plan to tune the models with different hyperparameters, such as the number of epochs, the batch size, the learning rate, different loss functions. Additionally, we will investigate other CNN architectures and transfer learning methods to see if they can improve the performance of the model. We will also try to use VGG-16 with only one-fully connected layer to see how it learns when compared to the standarad VGG-16 which has 3-fully connected layers.

5. References

[1] Yin, J., Zhang, Y., Hu, X., Chen, X. (2017). Multi-task deep neural network for age estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 10-18).

[2] Patel, Y., Patel, P., Patel, K. (2019). Age and gender classification using convolutional neural networks. In 2019 2nd International Conference on Computing Methodologies and Communication (ICCMC) (pp. 404-408). IEEE.

[3] Rothe, R., Timofte, R., Van Gool, L. (2015). Dex: Deep expectation of apparent age from a single image. In Proceedings of the IEEE International Conference on Computer Vision (pp. 10-18).

[4] Kingma, D. P., Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[5]Bottou, L. (2010). Large-scale machine learning with stochastic gradient descent. In Proceedings of COMPSTAT'2010 (pp. 177-186). Physica-Verlag HD.

[6] Smith, L. N. (2017). Cyclical learning rates for training neural networks. In 2017 IEEE Winter Conference on Applications of Computer Vision (WACV) (pp. 464-472). IEEE.