Enhancing Facial Age Estimation: Integrating Image Augmentation in Delta Age AdaIN Operation for Improved Network Robustness and Reliability

Shayan Khosravi & Abdur Rahman Mohammed

***Abstract*—This research paper presents an enhanced approach to facial age estimation by building upon the Delta Age AdaIN (DAA) operation, as detailed in the referenced CVPR 2023 paper [1]. Our work adds to the existing method by assessing the network's performance on a different facial age-related dataset and assessing the impact of the implementation of various image augmentation techniques on the overall performance of the proposed method. The dataset, "UTKFace", was specifically chosen to test the adaptability and performance of the proposed method on a dataset characterized by a more diverse range of ethnicities and ages. As for image augmentation, a specific set of augmentation techniques were used in isolation and combination to simulate a wide array of real-world conditions often encountered when dealing with raw image data such as brightness and contrast variability. The focus behind the implementation of these image augmentation techniques was to showcase the benefits of data augmentation for model robustness and reliability thereby improving the proposed method's adaptability and performance.**

I. Introduction

The problem of facial age estimation has so far been approached from a classical computer vision perspective such as feature extraction and modeling to predict age [2] [3]. This approach differs from how the human eye recognizes age, which involves comparing current experience information with a broader human context. Humans tend to estimate an individual's age by relying on their intuition which is comprised of a series of encounters with individuals of varying ages. It is through this process that we can estimate age by subconsciously comparing the facial features of an individual with facial features we have seen in the past. However, It is difficult to get representative age images of different ethnicities and genders for all ages. This would require a significantly balanced dataset of many ethnicities, genders, and ages. Such a dataset would be very difficult to build and procure; hence the idea of comparative learning has been ignored in the area of facial age estimation.

The original paper proposes an innovative method to circumvent the difficulties of requiring a complex dataset for this issue. The paper proposes a method that utilizes the power of generative adversarial networks, more specifically, StyleGAN [4] to generate facial images to represent all ages. The method focuses on applying facial features at varying degrees with respect to a specific age using the input image. The network's binary code mapping layer is responsible for learning the difference in facial features to derive a standard deviation and mean vector representing the rate of change in terms of facial features over time [1]. The Delta Age AdaIN (DAA) operation of the network is responsible for generating the representative age images based on the standard deviation and mean vector produced by the binary code mapping layer. Once all representative age images are generated, the network is tasked with comparing the input images with the representative age images to determine the most similar facial features for final age estimation.

A. Objectives

The primary objective of this research is to increase the performance of the Delta Age AdaIN (DAA) operation by assessing its capabilities on a more diverse facial dataset comprised of a wide range of ethnicities and ages. This objective is possible through the implementation of specific image augmentation techniques designed to simulate real-world conditions encountered in raw data and improve the network's robustness in processing raw data. The goal is to achieve a noticeable improvement in both the performance and robustness of the network.

B. Hypothesis

Implementing an image augmentation layer to the FaceEncoder module, trained on the UTKFace dataset through transfer learning, will improve the overall performance of age estimation across different ethnicities and gender.

II. Methodology

To achieve our objective, we needed to comprehensively assess the impact of the image augmentation layer and the UTKFace dataset on the original architecture from [1]. With a multitude of augmentations available with numerous possible combinations, establishing a structure for our approach is crucial. We noticed that most of the augmentations fall under two types; one adjusts the color of the pixels, and another changes the image itself. Therefore, we divided our models into four categories:

- **Base DAA model (No augmentation)**: Serves as a baseline to check the performance with the original architecture
- **Color-based image augmented model**: Applies augmentations that update the pixel color
- **Distortion-based image augmented model**: Applies augmentations that distort the image
- **Color & Distortion based image augmented model** : A model that considers combinations of the two types.

The following sections provide a detailed insight into our methodology:

A. Dataset

The UTKFace dataset [5] was chosen for our hypothesis due to several significant reasons. Firstly, it is composed of approximately 25,000 images, offering a substantial pool of data for our training and testing data split. The age range of the individuals within the dataset spans from a range of 0 to 116 years, giving us a full spectrum from infancy to old age. Another significant aspect of the UTKFace dataset is that each image is labeled with not only age but also gender and ethnicity. This multi-ethnic (Figure 3) labeling provides a richer set of data points, allowing for more nuanced and sophisticated age estimation models if needed. Such models can understand and adjust for potential variations in aging appearances across different genders and ethnicities, leading to more accurate and universally applicable results.

The ethnic diversity of the UTKFace dataset is a crucial advantage, especially when compared to the MegaAge Asian dataset used in the original paper [1]. While the MegaAge Asian dataset is larger, with 45,000 images, and covers a similar age range of 0 to 100 years, it primarily consists of individuals of Asian descent. This lack of diversity can lead to biased models that perform well only for specific ethnicities. In contrast, the ethnic diversity in the UTKFace dataset ensures that the developed models are more generalizable and perform well across a variety of ethnic groups. This inclusivity is critical in deploying age estimation models in global, multicultural contexts where diversity is the norm.

B. Preprocessing

For dataset pre-processing, a significant right skew in the age distribution was identified, particularly within the 25-30 age range (Figure 2). This segment showed an over-representation compared to other age categories, creating a right skew in the dataset's age distribution. Such an imbalance could lead to a biased age estimation model, favoring accuracy in the 25-30 age bracket while underperforming in others. To reduce sample bias, the under-sampling technique was used to specifically target this age range with the highest number of images. Furthermore, the images were of varying sizes and were part of various environments. Fortunately, the authors had a face detection and cropping pipeline that was customized for the UTKFace dataset. There were other images that were pruned from the dataset since they didn't have all the required facial features for age estimation. The final age distribution for the dataset can be depicted in Figure 4.

C. Image Augmentation Layer

Image augmentations essentially change the original image to a new image. These augmentations help avoid over-fitting of models and make them robust. The final list of augments are mentioned below for each of the models:

- **Color augmented model**
    1) Brightness (probability: 20%, strength: 0.2)

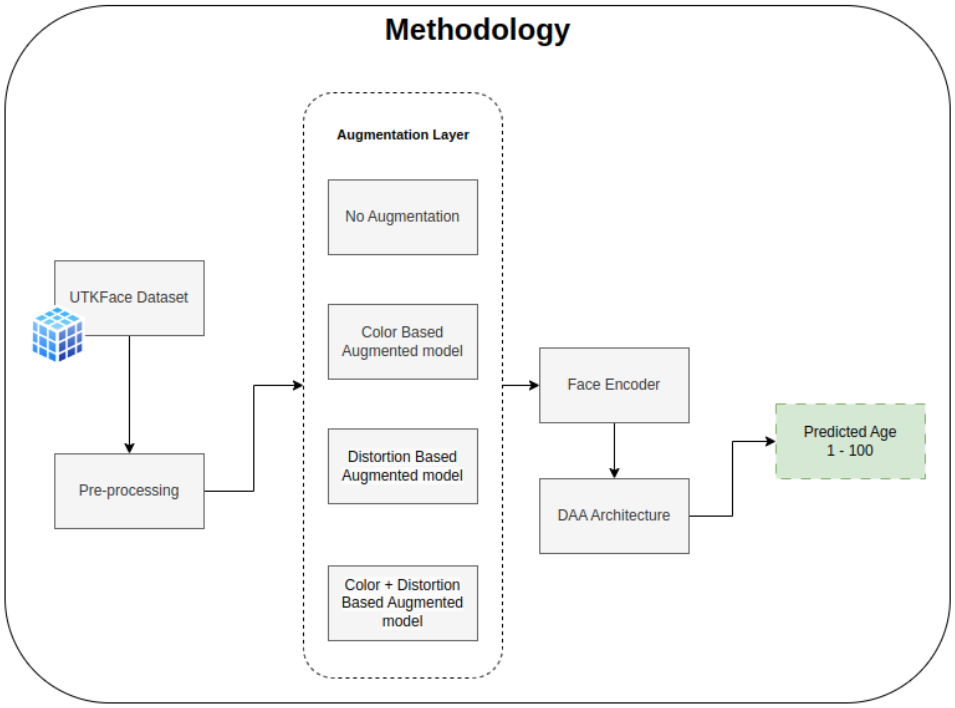


Figure 1: Methodology

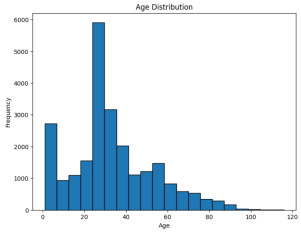


Figure 2: UTKFace Age Distribution Before Preprocessing

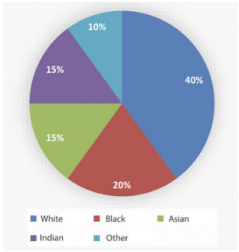


Figure 3: UTKFace Ethnic Diversity Distribution

2) Contrast (probability: 20%, strength: 0.2)
3) Saturation (probability: 20%, strength: 0.2)
4) Random grayscale (probability: 5%)

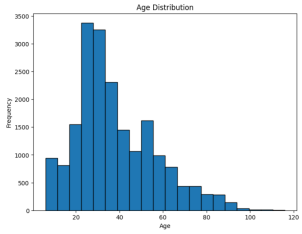


Figure 4: UTKFace Age Distribution After Preprocessing

- **Distortion augmented model**
  1) Horizontal flip (probability: 10%)
  2) Gaussian blur (probability: 20%, sigma=2, kernel_size=3)
- **Color & Distortion augmented model**
  1) Brightness (probability: 20%, strength: 0.2)
  2) Contrast (probability: 20%, strength: 0.2)
  3) Saturation (probability: 20%, strength: 0.2)
  4) Random grayscale (probability: 5%)
  5) Horizontal flip (probability: 10%)
  6) Gaussian blur (probability: 20%, sigma=2, kernel_size=3)

These augmentations were applied randomly to the training and validation set. Various other image augmentations were applied to the model, some of them are: random noise, elastic transform, etc.

D. Implementation

The models were trained on Google Colab Pro with the V100 Nvidia Tesla GPU, which has 16 GB of GPU RAM and 85 GB of system RAM. The processor used is the Intel(R) Xeon(R) CPU @ 2.00GHz, running Python v3.10.12. The GitHub repository for this research can be found here[1]. Around 100 models were trained and tested. The models were saved at every 5 epochs to Google Drive.

These are the parameters that were used for the experiment after trial and error:

- Epochs: 50
- backbone: ResNet18 [6]
- batch size: 64
- Learning rate: 1e-3
- Input image size: 128 x 128
- seed: 42
- CA threshold: 7
- Min Age: 1
- Max Age: 100

E. Metrics

For the experiment, we used two types of evaluation metrics to assess the performance of our model: Mean Absolute Error (MAE) and Cumulative Accuracy (CA).

1) Mean Absolute Error (MAE): Mean Absolute Error is a measure of the difference between two continuous variables. It calculates the average magnitude of errors in a set of predictions, without considering their direction. The MAE is given by the formula:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{1}$$

where n is the number of observations, y_i is the true value, and $\hat{y}_i$ is the predicted value.

[1]https://github.com/codezart/Delta_Age_AdaIN

2) Cumulative Accuracy (CA): Cumulative Accuracy, on the other hand, is a metric used to determine the accuracy of a classification model. K is the total number of testing images and Kn represents the number of testing images whose absolute errors are smaller than n (margin of error).

$$CA = \frac{K_n}{K} \times 100\% \tag{2}$$

For example, CA(7) would represent the percentage of predicted ages which fall between a margin of error of ±7.

III. Experiment Results and Analysis

The experiment involved testing various models on two different datasets: Mega Age Asian and UTK Face. The results are summarized in table I & II.

Dataset: Mega Age Asian		
Model Type	CA(7)	MAE
No Augmentation (Base DAA)	88.11	3.399

Table I: Comparison of CA(7) & MAE on Mega Age Asian Dataset

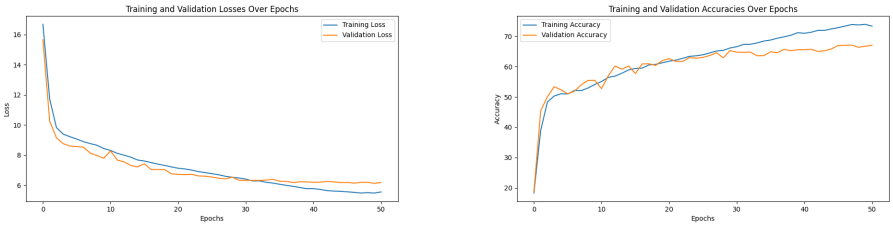
Dataset: UTKFace		
Model Type	CA(7)	MAE
No Augmentation (Base DAA)	64.23	6.900
Color Based Augmentation	65.43	6.767
Distortion Based Augmentation	**66.73**	**6.644**
Color + Distortion Based Augmentation	65.48	6.718

Table II: Comparison of CA(7) & MAE on UTKFace Dataset

A. Analysis

The performance of the models varied between the two datasets. The base model trained on the MegaAge Asian dataset achieved an impressive performance of 88.11% CA(7) and a MAE of 3.399. In contrast, a base model trained on the UTK Face dataset performed lower with 64.23% CA(7) and a MAE of 6.900. This does make sense because there were multiple ethnic groups (Figure 3) in the UTKFace dataset and this dataset is more than half the size of MegaAge Asian after preprocessing.

From table II, it is clear that introducing distortion-based image augmentations does improve the model performance with CA(7) improving by 2.5%. Color base augmentations, however lower than distortion-based, did also improve CA(7) by



(a) Distortion Augmentation loss (b) Distortion Augmentation accuracy

Figure 5: Loss and accuracy curves for different models

1%. A similar case is seen for the combined version. This suggests that adding distortion-based augmentations to the images does enhance the model's ability to learn more for the same architecture. While training our models, we saw that the models started to overfit after 25 to 35 epochs (Figure 5a). This is evident with the loss curves for training steadily decreasing.

Overall our hypothesis for the paper is not completely nullified as the model did worse on the UTKFace dataset, but this was due to the relatively smaller dataset size and more ethnic diversity. Adding distortion-based image augmentations did improve the model's capacity to learn better, be more generalized, and robust.

B. Other Experiments & Analysis

While we trained our main models, we also trained other models for better insight. First, we tried to train UTKFace with MegaAge Asian learned model (transfer learning), this had surprisingly worse results. Second, to test with new data from outside of UTKFace, we tested the UTKFace trained model with MegaAgeAsian test data and vice versa. UTKFace trained model had better generalization and was able to perform better than MegaAgeAsian trained model. Third, we trained 2 models (base - no augmentation, distortion based) with MegaAge Asian and UTKFace for 200 epochs and it does good for validation data but is less generalized. the increase was up to 69% for UTKFace distortion model with 200 epochs, while the base version only improved till 66.7%.

IV. Challenges Faced

While developing a pipeline and studying the results of augmentations we experienced several challenges. First, we did not have enough compute, therefore we set our epoch limit to max 50, in comparison to the 200 epochs that the original paper used. This tackled the time needed to train the models. Second, the models started to overfit prematurely, for this we increased the regularization strength. Third, the UTKFace dataset included images of varying sizes with some not having all the facial features, these were removed. Fourth, there was a lot of data imbalance, particularly for the ages between 20 and 40, we performed the undersampling technique to deal with this. Fifth, when training the model on Google Colab, our pipeline would sometimes get stuck and be out of connection with the runtime getting deleted. We got Colab Pro which mitigated this issue and we saved the model at every few steps.

V. Conclusion

In this research, we explored the performance and capability of the Delta Age AdaIN (DAA) operation on an ethnically diverse dataset such as UTKFace and various image augmentation techniques. Our aim was to assess how these factors affect the model's accuracy and overall performance in age estimation across various ethnicities, ages, and synthetically augmented images representing raw image data.

We discovered a noticeable interaction between dataset diversity and image augmentation. While the UTKFace dataset posed challenges, leading to a lower performance in contrast to the original MegaAge Asian dataset, implementing image augmentation techniques showed the promising results we were speculating originally in our hypothesis; specifically, distortion-based augmentations that improved the network performance, indicating their potential in enhancing the network learning capability and robustness.

Throughout our research, we faced several challenges, such as computational limitations, premature overfitting, and data imbalance within the UTKFace dataset. We solved these issues through various methods like limiting the number of epochs and applying under-sampling techniques and image cropping. Overall, our study illustrates the significance of image augmentation in not only the area of facial age estimation but learning-based computer vision as a whole, particularly in the context of building robust and reliable models.

REFERENCES

[1] P. Chen, X. Zhang, Y. Li, J. Tao, B. Xiao, B. Wang, and Z. Jiang, "Daa: A delta age adain operation for age estimation via binary code transformer," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 15 836–15 845.

[2] T.-Y. Yang, Y.-H. Huang, Y.-Y. Lin, P.-C. Hsiu, and Y.-Y. Chuang, "Ssr-net: A compact soft stagewise regression network for age estimation." in *IJCAI*, vol. 5, no. 6, 2018, p. 7.

[3] C. Zhang, S. Liu, X. Xu, and C. Zhu, "C3ae: Exploring the limits of compact model for age estimation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 12 587–12 596.

[4] F. Makhmudkhujaev, S. Hong, and I. K. Park, "Re-aging gan: Toward personalized face age transformation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 3908–3917.

[5] S. Y. Zhang, Zhifei and H. Qi, "Age progression/regression by conditional adversarial autoencoder," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.