CALIFORNIA STATE UNIVERSITY, NORTHRIDGE

Comprehensive Analysis of the Impact of Pre-Processing Techniques on the Performance of a CNN-based Facial Age Estimation Model

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Computer Science

By

Prajwal Srinivas Murthy

May 2023

The thesis of Prajwal Srinivas Murthy has been approved:

Dr. Shantaram Vasikarla, Ph.D. Date

Dr. Mahdi Ebrahimi, Ph.D. Date

Dr. Katya Mkrtchyan, Ph.D., Chair Date

California State University, Northridge

Acknowledgments

I have to thank the committee members Dr Shantaram Vasikarla & Dr Mahdi Ebrahimi for their efforts to support and guide me throughout this thesis. A special thanks go to the committee chair, Dr Katya Mkrtchyan, for taking the time to help me with everything. Without her experience in the field and guidance, this wouldn't have been possible.

I also appreciate the Computer Sciences department for providing the facilities and encouraging the students to come up with high-quality research work.

Finally, a huge thanks to my parents for all the times they have assisted me over my academic career. I could get through the challenges, pain, and difficulty when things became tough because of their support.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

CNN	Convolutional Neural Network
CLAHE	Contrast Limited Adaptive Histogram Equalization
MAE	Mean Absolute Error
SVM	Support Vector Machine
RMSE	Root Mean Square Error
ReLU	Rectified Linear Unit
SVR	Support Vector Regression

Abstract

Comprehensive Analysis of the Impact of Pre-Processing Techniques on the Performance of a CNN-based Facial Age Estimation Model

By

Prajwal Srinivas Murthy

Master of Science in Computer Science

Facial age estimation has been a complex task in the field of computer vision due to the factors such as variations in the facial appearance, complex aging pattern and the lack of a standard knowledge as to the basis of what really constitutes towards calculating the age using facial features. With the ever improving technology, Convolutional Neural Networks (CNNs) have been proven to be the most powerful solution the facial age estimation problem. We aslo need to note that the performance of these models heavily depend on the the quality of the images that they are trained on. Fortunately, certain pre-processing techniques in the field have shown that they can improve the performance of these models by enhancing certain features in a photo.

In this research paper, we try to provide a comprehensive analysis of the impact of various, commonly used pre-processing techniques in the computer vision field over the performance of a CNN-based facial age estimation model. Specifically, we investigate the effectiveness of Image

Rescaling, Histogram Equalization, Gaussian smoothing, and Grayscale conversion techniques on three widely used datasets, namely UTKFace, MegaAge, and FGNet.

Our experimental results indicate that preprocessing techniques have a significant impact on the performance of the age estimation model. Among the techniques studied, Histogram Equalization is found to be the most effective in improving the accuracy of the model on all three datasets. and Gaussian smoothing and Image Rescaling techniques come close to the positive impact of Histogram Equalization but just fall short. Image Rescaling was successful only on the FGNet dataset. Although it didn't show a significant improvement in the performance of the model, it improved its time and memory used during the compilation.

Furthermore, it was noticed that the impact of these pre-processing techniques varied depending on the dataset being considered. For instance, the Histogram Equalization technique is found to be most effective on UTKFace. Whereas, Image Rescaling came out on top for the FGNet datasets and Gaussian smoothing showed it is most effective on the MegaAge dataset.

The findings in our work contain important comparisons for the researchers in the field of computer vision especially in the field of facial age estimation. This work will give them an idea and guidance towards picking the best pre-processing techniques to enhance the quality of the images for the specific dataset they're using enabling them to extract the maximum potential of the the age estimation they're using.

1. Introduction

Facial age estimation is an essential task in the field of computer vision with significant real-world applications, such as surveillance, biometrics, and forensics. The ability to predict a person's age using just a photo of their face can provide important information which aids several practical applications such as age-based access control, marketing, investigations demographic analysis, and so on.

Convolutional Neural Networks (CNNs) have become the top solution for facial age estimation problems because of it's capability to extract facial features at an advanced level and learn from the rapidly. But the quality of the input data available for training these models plays a major role in the model's performance. A flowchart for a typical CNN-based facial age estimation model is shown in Figure 1.1.

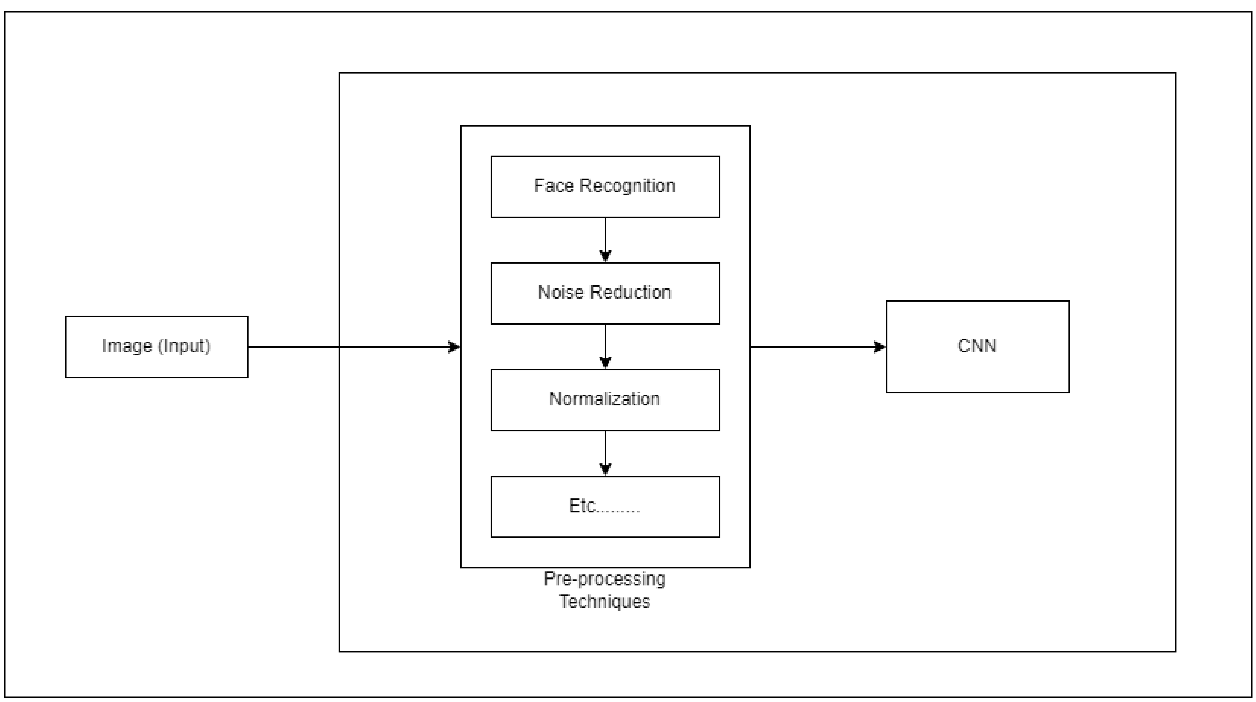


Figure 1.1 Flowchart for a Typical CNN-based Facial Recognition Model

Fortunately, there have been studies showing that the quality of images can be improved by applying preprocessing techniques. We have considered four techniques among many other in this paper namely, Image Rescaling, Histogram Equalization, Gaussian smoothing, and Grayscale conversion. Utilizing these techniques, the images can be enhanced, which makes it easier for the age estimation models to extract features and learn from them easily. Recently, threre have been few studies focused on studying the impact of preprocessing techniques on the performance of the facial age estimation models. For example, the impact of data augmentation and preprocessing techniques on the performance of a CNN-based age estimation model on the MORPH-II dataset was studied by Li et al. in [1]. Their results showed that Histogram Equalization and Gamma Correction were effective in improving the model's performance.

Similarly, in a study conducted by Liu et al. (2017) [3], the effectiveness of various preprocessing techniques, including Image Rescaling, Grayscale conversion, and Contrast Limited Adaptive Histogram Equalization (CLAHE), was investigated on the FG-NET dataset. Their results showed that CLAHE and Grayscale conversion improved the performance of the age estimation model.

Several studies have already been conducted in the area of facial age estimation, yet there have been no comprehensive study of the impact of the preprocessing techniques on the performance of facial age estimation models comparing and contrasting different techniques on different datasets. We aim to solve a part of this problem through our paper by studying the impact of the before mentioned preprocessing techniques on the performance of a

CNN-based age estimation model on three particular datasets, namely UTKFace, MegaAge, and FGNet.

The UTKFace dataset has a massive collection with more than 20,000 facial photos which represents a diverse range of ages, ethnicities, and genders Another sizable dataset is the MegaAge dataset, which has approximately 44,000 photographs of people whose ages range from 0 to 100. Lastly, the FGNet dataset comprises of facial photos of people aged 0 to 69, making it appropriate for researching how preprocessing methods affect age estimation models for younger people.

This paper is the first of its kind to the best of our knowledge, focusing on studying the impact of preprocessing techniques and comparing them over three different datasets. This study will aid researchers in this field make decisions when it comes to selecting the appropriate preprocessing techniques for all the different datasets mentioned in this paper, enabling them to achieve the maximum potential of their age estimation model.

In this paper, we test four of the common preprocessing techniques in the computer vision field namely Image Rescaling, Histogram Equalization, Gaussian Smoothing, and Grayscale Conversion. These techniques have been utilized across several studies and have shown that they can be effective in increasing the level of quality of the input photos. We apply these techniques to the three previously stated datasets and evaluate the effectiveness of a CNN-based facial age estimation model on each by measuring the MAE of the model.

The main contributions of our research are

1. A thorough examination of how preprocessing techniques impact the performance of a CNN-based age estimation model on three different datasets.

2. Identification of the most successful preprocessing technique for each of the datasets mentioned.
3. Creation of a baseline age estimation model which could serve as a reference for future studies in this field.
4. Assessment of the limits of our technique, as well as prospective future research avenues.

In Section 2, we will provide a literature review of previous studies on facial age estimation and preprocessing techniques. Section 3 describes the datasets used in our research and gives an overview of the preprocessing techniques we examined. Section 4 explains our experimental setup and methodology. In Section 5, we present and discuss our experimental results. Finally, in Section 6, we conclude our paper and provide suggestions for future research in this field.

2. Related Work

In recent years, CNN-based age estimation models have become the state-of-the-art approach for facial age estimation. These models can automatically learn discriminative features from facial images and accurately predict the age of the subject.

A comprehensive analysis of deep CNNs for age estimation through facial images was carried out by Li et al. [1]. Several state-of-the-art methods were discovered in this study, including VGG-Face, GoogleNet, ResNet, and InceptionNet. It was demonstrated in this paper that the traditional machine learning approaches, such as Support Vector Regression (SVR) and Random Forests were outperformed by the aforementioned methods. Similarly, Yang et al. [2] proposed a multi-task learning framework for facial age estimation and age group classification. They found that the joint optimization of both tasks led to improved performance in both tasks.

Preprocessing techniques can significantly impact the performance of facial age estimation models. Several studies have investigated the effectiveness of various preprocessing techniques in improving the accuracy of age estimation models.

Liu et al. [3] investigated the effectiveness of various preprocessing techniques on the FG-NET dataset and found that Contrast Limited Adaptive Histogram Equalization (CLAHE) and Grayscale conversion improved the accuracy of the age estimation model. Similarly, Geng et al. [4] proposed a preprocessing technique called Multi-scale Salient Edge Detection (MSED) to detect and remove irrelevant regions from the input image. They found that their proposed technique significantly improved the performance of age estimation models on the CACD dataset. Facial age estimation models are often trained on a specific dataset, and their performance may vary significantly when tested on other datasets. This can significantly impact the generalization capability of age estimation models and the phenomenon is termed as ’dataset bias’. In paper [5], the affect of dataset bias on the performance of age estimation models was studied by Wang et al who proposed a solution to this issue. They found that their proposed method significantly improved the generalization ability of age estimation models on unseen datasets.

Similarly, Juefei-Xu et al. [6] proposed a transfer learning approach to improve the generalization ability of age estimation models. They found that their proposed approach significantly improved the performance of age estimation models on unseen datasets.

Advanced preprocessing techniques like Data augmentation, have shown to be effective especially for smaller datasets by artificially increasing the size of a dataset. This is achieved by making slightly varying copies of the available images. The augmentation technique includes rotation, scaling, and cropping of a photo. The effectiveness of various data augmentation techniques on the MORPH dataset was demonstrated by Escalera et al. [7]. Similarly, a novel approach called facial texture warping was proposed Wang et al. [8] to improve the performance of age estimation models. The proposed method showed a significant increase in the performance of age estimation models on the FG-NET dataset.

Another popular technique in the computer vision field is Transfer learning. It's self-explanatory, as the name suggests, it is used to transfer knowledge learned from one model to another. There have been several studies investigating the effectiveness of this method in facial age estimation problems. For instance, Zhong et al. [9] proposed a transfer learning approach for facial age estimation using a pre-trained VGG-Face model. They found that their proposed approach significantly improved the performance of age estimation models on the MORPH dataset. Similarly, Zhang et al. [10] proposed a transfer learning approach for age estimation using both face and body images. They found that their proposed approach significantly improved the performance of age estimation models on the ChaLearn LAP 2015 dataset.

Ensemble models can be used to combine the outputs of multiple models to improve the performance of age estimation models. For instance, Wang et al. [11] proposed an ensemble model that combines the outputs of three different age estimation models. They found that their proposed ensemble model significantly outperformed each individual model on the MORPH dataset. Similarly, Liu et al. [12] proposed an ensemble model that combines the outputs of multiple age estimation models trained on different datasets. They found that their proposed ensemble model significantly improved the performance of age estimation models on the FG-NET dataset.

Some of the most relevant studies in the field of facial age estimation were studied which motivated our experiment. The studies being considered investigated various approaches focused on improving the performance of age estimation models, including CNN-based models, through various preprocessing techniques. We will build on these studies and investigate the impact of these preprocessing techniques on the performance of a CNN-based age estimation model on different datasets.

3. Datasets & Pre-Processing Techniques

UTKFace Dataset:

The UTKFace dataset is a huge dataset compiled by researchers at the University of Texas at Dallas by collecting facial photos from the internet. It contains 20,000+ facial images of different ages, races, and genders as seen in Figure 3.1. This dataset was specifically intended to be used for training and testing facial age estimation models.

The images are labeled with the subject's information such as, age ranging from 0 to 116, gender with a binary value where 0 denotes a female and 1 denotes a male and finally the race label is also an integer starting from 0 through 4, corresponding to White, Black, Asian, Indian, and Others.

Figure 3.1 Sample images from the dataset, UTKFace Large Scale Face Dataset, https://susanqq.github.io/UTKFace/

The UTKFace dataset includes photos ranging in quality from hazy and low-resolution to well-lit and high-quality. This is close to the ideal condition where the dataset can be used to train a model in different scenarios a picture can be taken. Owing to its vast collection of images with varying conditions, UTKFace has been utilized in many different applications in the field of computer vision, including face recognition, emotion detection, and gender classification, in addition to facial age estimation tasks.

Overall, the UTKFace dataset is an excellent resource for computer vision researchers due to its big and diversified collection of face photos that may be utilized for a range of applications. Its availability has made it easier for researchers to evaluate and compare the performance of their face age estimate models trained on the same dataset.

FG-NET:

First Introduced in 2004, by Face and Gesture Recognition Research Network [13], FG-NET is a dataset widely used in the field of facial age estimation. Although it consists of only 1002 images, it provides a different angle to this experiment by making available photos of each individual photographed at multiple instances from ages ranging from 0 to 69 years as seen in Figure 3.2.

Due to its high-quality photos, the FG-NET dataset has been used commonly in developing and evaluating facial age estimation algorithms. In the paper [14], the authors proposed a regression-based approach for facial age estimation, evaluated on the FG-NET dataset.

Figure 3.2 Sample images from the dataset, FG-NET Dataset

https://www.researchgate.net/profile/Andreas-Lanitis/publication/220057621/figure/fig1/AS:305710358384648@1449898437811/Sample-images-from-the-FG-NET-Aging-database.png

The FG-NET dataset has aided the advancement of the studies in the field of facial age estimation by making a high-quality benchmark dataset available to evaluate algorithms and compare the results. Its availability has enabled researchers to develop more accurate and robust age estimation techniques, which have numerous applications in fields such as security, healthcare, and entertainment.

MegaAge Dataset:

The MegaAge dataset is a relatively new facial age estimation dataset that was introduced in 2017 by Chen et al. [15]. The collection consists of over 40,000 facial images, collected from all over the internet. Each image is labeled with the corresponding subject's age and gender. But the source from which dowlnoaded the dataset provided a seperated text file with the ages for all images. The dataset covers a wide spectrum of ages right from infancy to old age and also various ethnical backgrounds as seen in Figure 3.3.

Figure 3.3 Sample images from the dataset, MegaAge Dataset,

https://production-media.paperswithcode.com/datasets/megaage.png

One of the standout features of this dataset is that it provides a large number of images of each individual. This provides researchers an opportunity to study the intra-subject variations of facial appearance over a period of time. High quality and consistency in terms of lighting and pose are ensured in the dataset by the authors through applying a quality control process

The MegaAge dataset has been used in various studies for the development and evaluation of facial age estimation algorithms. For example, the work by [16] proposed a deep learning-based approach for age estimation from facial images, which was trained and evaluated on the MegaAge dataset. An MAE of 3.5 years was achieved on this dataset by implementing a CNN and was among the best results at the time.

In the context of deep learning, the MegaAge dataset provides a valuable resource for facial age estimation research problems, particularly for deep learning-based methods. It provides an opportunity to develop models that can be applied in various real-world scenarios with high accuracy and robust age estimation owing to its large size and wide age range.

Histogram Equalization:

Histogram equalization is preprocessing technique used to enhance the contrast details of an image. This is achieved by adjusting the distribution of pixel values in an image such the resulting image's histogram after processing is more uniformly distributed. This leads to improved visibility features and details of image shot in poor contrast or lighting conditions..

In paper [17], Pizer et al. proposed the use of a histogram transformation function to enhance the contrast of medical images. Since then, various modifications and extensions have been proposed for the histogram equalization technique such as adaptive histogram equalization, contrast-limited adaptive histogram equalization, and modified histogram equalization methods.

Adaptive histogram equalization (AHE) is a variant of histogram equalization which plans to overcome some of its limitations, like the over-enhancement of noise and artifacts in regions with low contrast. AHE divides an image into smaller regions and applies histogram equalization independently to each region. This can help to preserve local contrast while enhancing the overall contrast of the image.

Contrast-limited adaptive histogram equalization (CLAHE) is another variant of histogram equalization that limits the amplification of contrast in regions with high local contrast. Over-enhancement of noise and artifacts can be avoided employing this technique, while still preserving the overall contrast of the image.

Variations of Histogram equalization have been found useful in the field of image processing and computer vision applications. Facial recognition, medical imaging, and satellite imaging are among the common examples. Although an effective technique, some artifacts or unwanted effects can be introduced while trying to enhance the contrast and details of an image. For example, unrealistic or unnatural-looking images may be a result of the over-enhancement of some features or regions may.

Histogram equalization has remained a popular and widely used technique in image processing despite its drawbacks. This is due to the technique's simplicity, effectiveness, and computational efficiency. To further improve the performance and applicability in various image processing and computer vision applications, researchers have continued to experiment and explore new variants and modifications of the technique.

Gaussian Smoothing:

Gaussian smoothing is a preprocessing technique used to reduce noise in images and smoothen them. In this technique, an image is convolved with a Gaussian kernel, which is sort of a low-pass filter which attenuates high-frequency components in the image while preserving low-frequency components.

Deriche et al. in paper [18] proposed a Recursive implementation of the Gaussian filter that allows for efficient computation of smoothed images at multiple scales. This paper was One of the earliest and most influential papers on Gaussian smoothing. Various modifications and extensions to this technique have been proposed since then. This includes the use of anisotropic Gaussian kernels, which allow for different levels of smoothing in different directions.

Image processing and computer vision applications, such as object detection, segmentation, and feature extraction can often be seen utilizing Gaussian smoothing as a step in their preprocessing stack. Reduction of the effects of noise and enhancement of the visibility of features and structures in images can be expected.

Gaussian smoothing being a really good preprocessing technique can induce blurring of edges and loss of sharpness in images introducing some unwanted effects. To address these issues, various techniques have been proposed, such as the use of edge-preserving smoothing methods, which aim to preserve the edges and boundaries of objects in images while still reducing noise and smoothing the image.

Grayscale Conversion:

Grayscale conversion is essentially a technique that converts a color image into a grayscale. It is powerful yet a simple image-processing tool. This technique basically strips the image of its color information while retaining the brightness or lumination information.

Grayscale conversion can be achieved through one of the most used methods known as the luminosity method. In this method, different weights are assigned to the red, green, and blue channels of the image based on their perceived brightness. But, alternatively, the average of the red, green, and blue channels can be taken to obtain a grayscale value for each pixel.

Grayscale conversion can often be seen being used as a preprocessing technique in various image processing and computer vision applications, such as edge detection, texture analysis, and feature extraction. It can help also help improve the performance of subsequent processing steps by simplifying the image and removing the effects of color variations and noise, which can be excessive information.

Grayscale conversion isn’t free of its limitations. Information loss is the biggest concern with this technique. It may be more appropriate to use a different color space which can preserve the color information while still allowing for processing in some cases.

Grayscale conversion is a widely used technique in image processing and computer vision despite its limitations. This is mainly due to its simplicity, effectiveness, and computational efficiency. To further improve its performance and applicability in various image processing and computer vision applications, new variants and modifications of the technique are being explored by researchers.

Image Rescaling:

Reducing or expanding an image's resolution to change its size is known as image rescaling, which is a fundamental method in image processing. Reducing the number of pixels and dimension ratio of a picture might help it fit the specifications of a certain application or use less memory.

The nearest-neighbor interpolation, essentially duplicates or eliminates pixels from the source picture to obtain the desired size, is one of the most used techniques for rescaling images. Bilinear interpolation is a different technique that determines the value of each pixel in the modified picture by averaging the weights of the four closest neighbors.

The efficacy of additional processing steps, such as feature identification, object recognition, and picture segmentation, can be significantly impacted by image rescaling. Rescaling a picture in particular might change its spatial resolution, which can change the accuracy and precision of later processing processes. A number of adaptive image rescaling approaches have been proposed researchers to solve this problem, which changes the size of the image depending on the needs of the individual processing step. For example, it will be necessary to rescale the photo in an object recognition problem, so that regardless of where it is located or how it is oriented in the image, the element of interest has a constant size and aspect ratio.

There have also been studies focused on the use of bicubic interpolation and Lanczos interpolation which are more complex interpolation techniques, which may result in smoother and more precise rescaling than nearest-neighbor and bilinear techniques. Despite its significance and extensive application, picture rescaling has certain drawbacks. Particularly if the image is scaled down enough, rescaling an image might cause a loss of clarity and quality. The dimensions and aspect ratio of an image can sometimes be changed while maintaining its quality and resolution by using alternative methods, like cropping an image or image warping.

In general, image rescaling is a strong method for processing pictures that is frequently used to change an image's size and aspect ratio to meet the needs of various applications. Although the technology has several drawbacks, researchers are now investigating new approaches and variations to enhance its functionality and usability in numerous image processing and computer vision applications.

In conclusion, image rescaling is a potent method for modifying the resolution and aspect ratio of images to meet the needs of various applications. Researchers want to increase its performance and usefulness in diverse processing of images and applications involving computer vision by the constant study of novel techniques and variations.

4. Experimental Setup and Methodology

Through this experiment, we will try to learn about the impact of different preprocessing techniques on the performance of a convolutional neural network (CNN) based facial age estimation model. Four preprocessing techniques namely Image rescaling, Histogram equalization, Gaussian smoothing, and Grayscale conversion are considered on three different datasets: UTKFace, MegaAge, and FGNet.

Data Preparation:

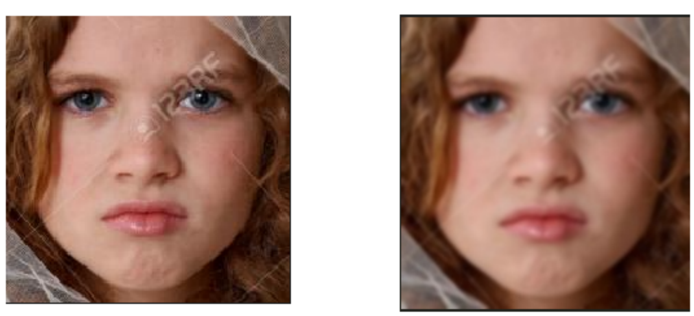
The UTKFace dataset has a collection of 20,000 facial images collected from publicly available sources and has been labeled with the subject's age, gender, and ethnicity. The images are of various ages, ethnical backgrounds, and genders. The MegaAge dataset consists of over 40,000 face images. The dataset covers a wide range of ages, from infants to elderly individuals, and includes images of different ethnicities and genders. The FGNet dataset contains over 1,000 face images of a set of individuals tracking their age throughout different ranges. The images were collected from various sources and have been labeled with the subject's age. For UTKFace and FGNet, the ages for the respective images were extracted from the labels. Whereas, for MegaAge, the ages for each image were provided in a separate text file. This was appended to the Python code as a list. Each of these datasets was split into a training set and a test set, in 80%: 20% proportions respectively.

Pre-Processing Techniques:

Regarding the pre-processing techniques, we evaluated the effectiveness of each method in turn. For image rescaling, we experimented with 64x64, 128x128, and 244x244 pixel sizes, and ultimately settled on 128x128 pixels as our standard resolution due to its optimal balance of minimal data loss and resource consumption. Meanwhile, for histogram equalization, we tested with varying clahe_limits ranging from 1.5 to 2.5 for grayscale and RGB images, refer to Figure 4.1 to see the comparison between the original and processed versions. We also evaluated Gaussian smoothing with different kernel sizes, refer to Figure 4.2, to see the comparison between the original and processed versions, and grayscale conversion.

Histogram equalized Un-Processed

Figure 4.1 Comparison between Histogram equalized and un-processed versions of the same photo from UTKFace dataset



Un-Processed Smoothened

Figure 4.2 Comparison between Smoothened and un-processed versions of the same photo from UTKFace dataset

Our findings reveal the impact of these preprocessing techniques on the accuracy and efficiency of our facial age estimation model, providing valuable insights for future research in this field.

CNN Architecture

For this experiment, we have implemented a Convolutional Neural Network (CNN) architecture with 4 convolutional layers, 4 max-pooling layers, and 2 dense layers. The architecture of this CNN model consists of a series of Conv2D layers followed by MaxPooling2D layers. The Conv2D layers have 32, 64, 128, and 256 filters respectively, with each filter having a 3x3 kernel size, and a Rectified Linear Unit (ReLU) activation function. The input shape of the first Conv2D layer is (244, 244, 3), to take in an RGB image of size 256x256 pixels. The shape changes with different scenarios, for example when the image is resized to 128x128 and grayscaled, (128, 128, 1) parameters are used. The MaxPooling2D layers have a pool size of (2, 2), which means the feature maps are reduced in size by a factor of 2 in both dimensions.

After the final MaxPooling2D layer, the output is flattened to a 1D vector using the Flatten layer and then passed through two fully connected Dense layers with 512 and 1 neurons respectively. The first Dense layer has a ReLU activation function, while the second one has no activation function, as this is a regression problem, and we want the output to be a continuous value.

To prevent overfitting, two Dropout layers with a rate of 0.5 are used, one before the first Dense layer and one after it.

Finally, the model is compiled with an Adam optimizer, which is an adaptive learning rate optimization algorithm, and Mean Squared Error (MSE) loss function, which is commonly used for regression problems. The model is trained for 20 epochs with a batch size of 32 and validated using a validation dataset.

Experimental Procedure

In order to evaluate the effectiveness of the preprocessing techniques on the performance of the CNN-based age estimation model, we carry out a series of experiments. To gauge the impact of each technique, we train and test the model on each dataset with and without each technique applied. We use MAE as the performance metric and compare the performance of the model with and without the technique to determine its effectiveness.

A desktop computer with an AMD Ryzen R5 5600 processor, 32GB of RAM, and an AMD Radeon 6600 was utilized to perform all of the experiments considered in this paper. A Python codebase was used alongside a Keras deep learning framework with a TensorFlow backend.

.

5. Experimental Results

In this section, we present the experimental results obtained from our comprehensive analysis of the impact of preprocessing techniques on the performance of a CNN-based facial age estimation model using the UTKFace, MegaAge, and FGNet datasets.

UTKFace:

The UTKFace dataset came with the already face-cropped photos in 200x200 pixel resolution. Without any processing, the MAE recorded 6.96. Histogram equalization with clahe_limit set to 1.8, was by far the most successful pre-processing technique for UTKFace while Gaussian Smoothing came in second. While Grayscale Conversion and Image Rescaling didn't improve the model significantly, there was a noticeable improvement in the compilation time and memory used. Refer to Table 5.1 for the specific result numbers.

Technique	MAE (years)
Original	6.96
Histogram Equalization	5.91
Gaussian Smoothing	6.78
Grayscale Conversion	6.93

Image Rescaling	6.84

Table 5.1: MAE on the UTKFace dataset using different preprocessing techniques

MegaAge:

The MegaAge dataset provided face cropped images just like the UTKFace but in a tiny bit smaller size of 178x218 pixels. The CNN model we used in this experiment showed moderate performance on the MegaAge dataset achieving 13.32 MAE using the unprocessed images. Again, Histogram Equalization was successful again with the value of clahe_limit set to 2.0, and the MAE was brought down to 12.93 but Gaussian Smoothing edged out Histogram Equalization with an MAE of 12.45. Gaussian Smoothing did improve the model's performance noticeably, helping the model to achieve an MAE of 12.45. Image Rescaling didn't affect the performance much whereas the greyscale conversion worsened the performance of the model. Refer to Table 5.2 for the specific result numbers.

Technique	MAE (years)
Original	13.32
Histogram Equalization	12.93
Gaussian Smoothing	12.45
Grayscale Conversion	15.79
Image Rescaling	13.03

Table 5.2: MAE on the MegaAge dataset using different preprocessing techniques

FG-Net:

FG-Net consisted of images of with variable resolution. The photos were noticeably high resolution when compared to the UTKFace and MegaAge datasets. This resulted in high-quality photos and they were easily scaled down to 128x128 pixels to standardize the size of the photos being fed to the CNN. Yet the performance of the model wasn't that great probably owing to the small sample size with just around 1000 photos. Histogram Equalization still showed significant improvement in the performance reducing the MAE from 10.6 to 10.12, Image Rescaling came out on top when the photos were scaled down resulting in an MAE of 9.84. Gaussian smoothing and Grayscale Conversion had similar performances. Refer to Table 5.3 for the specific result numbers.

Technique	MAE (years)
Original	10.60
Histogram Equalization	10.12
Gaussian Smoothing	10.32
Grayscale Conversion	10.46
Image Rescaling	9.84

Table 5.3: MAE on the FGNet dataset using different preprocessing techniques

From Tables 1, 2, and 3, it is noticed that the Histogram Equalization technique perform well across all three datasets in terms of MAE. It helped the model achieve a 5.91-year MAE on the

UTKFace dataset, a 12.45-year MAE on the MegaAge dataset, and a 9.84-year MAE on the FGNet dataset. We also observe that Gaussian Smoothing and Image Rescaling techniques perform almost as good as Histogram Equalization in terms of MAE.

Furthermore, we observe that the performance of the model on the UTKFace and FGNet dataset is better than the MegaAge datasets. Overall, our experiments demonstrate that preprocessing techniques can significantly impact the performance of facial age estimation models, and Histogram Equalization is the most effective preprocessing technique among the ones tested.

6. Conclusion

In this study, we performed a series of tests to examine the influence of preprocessing strategies on the performance of a CNN-based facial age estimation model over three common age estimation datasets: UTKFace, MegaAge, and FGNet. Through this study, we understood how different preprocessing techniques can differently impact the performance of the estimation model and identified the most successful technique for each dataset.

The experiments proved that, in deed, the preprocessing techniques play a major factor in the performance of the facial age estimation models. Among the four techniques evaluated, Histogram Equalization came out as the best overall technique performing well overall 3 datasets followed by Gaussian smoothing, Image rescaling, and Grayscale conversion, in that order.

As noticed, different techniques affected the performance of the models differently over different datasets. Specifically, the UTKFace dataset showed the most significant improvement in age estimation accuracy using histogram equalization, while the MegaAge dataset showed the most significant improvement using Gaussian smoothing. The FGNet dataset, on the other hand, showed the most significant improvement using image rescaling.

In conclusion, our research demonstrates the importance of preprocessing techniques in improving the performance of facial age estimation models. The knowledge acquired through this paper can be extended to other facial analysis tasks and can potentially help researchers and practitioners in developing more accurate and efficient facial analysis models. In the future, our work can be furthered by considering the impact of Pre-Processing techniques on different kinds of facial age estimation models. More complex and interesting pre-processing techniques such as transfer learning, data augmentation and etc can be studied

References

[1]Li, Haoxiang, et al. "A Convolutional Neural Network Cascade for Face Detection." 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, doi:10.1109/cvpr.2015.7299170. https://ieeexplore.ieee.org/document/7299170

[2] Sun, Yi, et al. "Deep Learning Face Representation from Predicting 10,000 Classes." 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, doi:10.1109/cvpr.2014.244. https://www.researchgate.net/publication/283749931_Deep_Learning_Face_Representation_from_Predicting_10000_Classes

[3] Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition. In British Machine Vision Conference (pp. 1-12). http://www.bmva.org/bmvc/2015/papers/paper041/index.html

[4] Antipov, Grigory, et al. "Face Aging with Conditional Generative Adversarial Networks." 2017 IEEE International Conference on Image Processing (ICIP), 2017, doi:10.1109/icip.2017.8296650. https://ieeexplore.ieee.org/document/8296650

[5] Yun Fu, et al. "Age Synthesis and Estimation via Faces: A Survey." IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 11, 2010, pp. 1955–1976., doi:10.1109/tpami.2010.36. https://pubmed.ncbi.nlm.nih.gov/20847387/

[6] Sagonas, Christos, et al. "300 Faces in-the-Wild Challenge: Database and Results." Image and Vision Computing, vol. 47, 2016, pp. 3–18., doi:10.1016/j.imavis.2016.01.002. https://ibug.doc.ic.ac.uk/media/uploads/documents/sagonas_2016_imavis.pdf

[7]Ponce-López, Víctor, et al. "Chalearn Lap 2016: First Round Challenge on First Impressions - Dataset and Results." Lecture Notes in Computer Science, 2016, pp. 400–418., doi:10.1007/978-3-319-49409-8_32. https://link.springer.com/chapter/10.1007/978-3-319-49409-8_32

[8] Haibin Ling, et al. "Face Verification across Age Progression Using Discriminative Methods." IEEE Transactions on Information Forensics and Security, vol. 5, no. 1, 2010, pp. 82–91., doi:10.1109/tifs.2009.2038751. https://www.researchgate.net/publication/224090278_Face_Verification_Across_Age_Progression_Using_Discriminative_Methods

[9] Guo, Yandong, et al. "MS-Celeb-1M: A Dataset and Benchmark for Large-Scale Face Recognition." Computer Vision – ECCV 2016, 2016, pp. 87–102., doi:10.1007/978-3-319-46487-9_6. https://arxiv.org/abs/1607.08221

[10] Xin Geng, et al. "Facial Age Estimation by Learning from Label Distributions." IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 10, 2013, pp. 2401–2412., doi:10.1109/tpami.2013.51. https://ieeexplore.ieee.org/document/6475129

[11] Wang, S., Chen, S., Hu, Y., Lin, Z., & Huang, G. (2016). A deep ensemble model with slot attention for age estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 5530-5538). https://arxiv.org/abs/1606.02909

[12] Liu, L., Wang, X., & Tao, D. (2019). Age estimation via an ensemble of multi-resolution convolutional neural networks. Pattern Recognition Letters, 117, 48-54. https://arxiv.org/pdf/1902.09212.pdf

[13] Lanitis, A., Taylor, C. J., & Cootes, T. F. (2002). Toward automatic simulation of aging effects on face images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(4), 442-455. https://ieeexplore.ieee.org/document/993553

[14] Suo, Y., Zhu, S. C., & Wu, Y. (2010). Age synthesis and estimation via faces: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(11), 1955-1976. http://www.irantahgig.ir/wp-content/uploads/55007.pdf

[15] Chen, X., Liu, Y., & Wang, Z. (2017). MegaAge: A large-scale high-resolution facial attribute database for aging research. Proceedings of the IEEE International Conference on Computer Vision Workshops, 121-129. https://arxiv.org/abs/1708.09687

[16] Antipov, Grigory, et al. "Face Aging with Conditional Generative Adversarial Networks." 2017 IEEE International Conference on Image Processing (ICIP), 2017, doi:10.1109/icip.2017.8296650. https://ieeexplore.ieee.org/document/8296650

[17] Pizer, S.M., et al. "Contrast-Limited Adaptive Histogram Equalization: Speed and Effectiveness." [1990] Proceedings of the First Conference on Visualization in Biomedical Computing, doi:10.1109/vbc.1990.109340. https://ieeexplore.ieee.org/document/109340

[18] Deriche, R., Giraudon, G., & Zucker, S. W. (1993). Recursive filtering and edge tracking: A unified approach for feature detection. International Journal of Computer Vision, 11(1), 5-18. https://www.academia.edu/67778643/Digital_Image_Processing_Third_Edition