Simultaneous Age, Gender and Race Classification Using Facial Image Characteristics

Arora Manav: U1822077D   Fazili Numair: U1822056F

Abstract

Gender and age classification are well-defined problems in Machine Learning. These Computer Vision tasks have gained significant traction attributed to the veritable rise in circulation and availability of image data over the past decade. However, despite data availability, the problems associated with these tasks are uniquely difficult due to the inherent complexity of the tasks and exacerbated by the lack of quality images and labeling across the prevalent datasets. In this project, we leverage a CNN model to classify attributes of age, gender, and race of a person from an image. These classifications are generated simultaneously to leverage gender-specific age characteristics, race and age-specific gender characteristics, and gender and age-specific race characteristics inherent to images. Our findings report a marginal performance increase over the prevalent methods while the novelty introduced by race classification promotes further understanding of the datasets.

Keywords: Machine Learning, Classification Image Recognition, CNN.

1. INTRODUCTION

Age and gender play fundamental roles in social interactions. Languages reserve different salutations and grammar rules for men or women, and often different vocabularies are used when addressing disparate age groups. Despite the basic roles these attributes play in our day-to-day lives, the ability to automatically estimate them accurately and reliably from facial images is still distant from meeting the requirements of commercial applications. This is perplexing when considering recent claims to super-human capabilities in the related domain of face recognition. Facial age variations have a significant impact on gender classification since shape and texture vary across age groups. This requires reexamination of the gender classification system to consider this information. In this project, we propose a model to simultaneously consider age and gender recognition and take advantage of the distinctive characteristics inherent to these images. Subsequently, the race attribute is also incorporated in this implementation. Furthermore, the model is trained and cross-evaluated across two datasets of UTK and Adience to evaluate consistency and generalizability.

2. LITERATURE REVIEW

Gil Levi and Tal Hassner delineated in their paper [1] that by learning representations through the use of deep-convolutional neural networks (CNN), a significant increase in performance can be obtained on these tasks. They proposed a simple convolutional net architecture that can be used even when the amount of learning data is limited. This method was evaluated on the Adience benchmark for age and gender estimation which will also be used for this implementation and further comparison.

3. IMPLEMENTATION

3.1. Models Implemented

1. Age & Gender Classification

2. Age, Gender and Race Classification

3.2. Datasets

The age and gender classification models are common to UTKFace and Adience datasets while the race attribute was only trained on the UTK Dataset as Adience does not have a race/ethnicity-based training attribute.

3.2.1 UTK Dataset

UTKFace dataset is a large-scale image dataset consisting of 20,708 face images. The images have age annotations between (0-116), binary gender values of Male or Female, and five unique ethnicities comprising White, Black, Asian, Indian, and Others. The images cover variations in pose, facial expression, illumination, occlusion, resolution, etc.

3.2.2. ADIENCE Dataset

The Adience dataset [2], published in 2014, contains 26,580 photos across 2,284 subjects with a binary gender label and contains 25 distinct age values of either an integer type or a range class. The key principle of the data set is to capture the images as close to real-world conditions as possible, including all variations in appearance, pose, lighting condition, and image quality, to name a few. The dataset size was reduced to 17,900 after removing the null values.

3.3. Age & Gender Classification

3.3.1. Objective

To classify a given person's gender and categorise their age into the respective bins, simultaneously.

3.3.2. Data Processing

1. Data extraction and cleaning.
2. One hot encoding of age,gender and race attributes using Keras library.
3. Data Scaling and Normalization
4. Images converted to numpy arrays for faster computations.

3.3.3. Model Design & Architecture

The CNN model uses several Convolutional and Pooling layers to learn feature representations from the images. This is followed by standard Fully Connected (FC) layers, which are used towards final classification. The feature extraction (CNN) layers are shared, while separate FC layers are used for the three attributes with branching points immediately after the feature extraction. The feature extraction was kept common to consider age and gender recognition simultaneously.

The CNN network characteristics are outlined in Table 1.

Input Dimensions (RGB)	198 × 198 × 3
Filter Size	3 × 3
Pooling	MaxPool2D (2 × 2)
Dropout FC	0.2 - 0.3
Dropout CNN	Spatial Dropout of 0.1 - 0.2 where filters exceed 128
Loss Function	Categorical Cross Entropy
Batch Size	32
Learning rate	Decaying LR halved every 5 epochs - initial LR: 0.008
Normalization	Batch
Activation	Relu
Activation Outputs	Softmax
Optimizer	Adam

Table 1: CNN Network Characteristics

The Gender network characteristics are outlined in Table 2.

Input	Output of Feature Extraction
Flatten	Input
FC Layers	64-32-2
Dropout FC	0.25 - 0.3
Loss Function	Categorical Cross Entropy
Batch Size	32
Learning rate	Decaying LR halved every 5 epochs - initial LR: 0.008
Normalization	Batch
Activation	Relu
Activation Outputs	Softmax
Optimizer	Adam

Table 2: Gender Network Characteristics

The Age network characteristics are outlined in Table 3.

Input	Output of Feature Extraction
Flatten	Input
FC Layers	128-64-32-4/7
Dropout FC	0.2 - 0.3
Loss Function	Categorical Cross Entropy
Batch Size	32
Learning rate	Decaying LR halved every 5 epochs - initial LR: 0.008
Normalization	Batch
Activation	Relu
Activation Outputs	Softmax
Optimizer	Adam

Table 3: Age Network Characteristics

3.3.4. Parameter Tuning

3.3.4.1 . Bin Size

This was one of the most important parameters especially given the fact that both out datasets had a different data distribution. We experimented using different bin sizes like 25,10 and 5 years.

3.3.4.2. Regularization

A myriad of regularization techniques were tried from Dropout to L2 regularization. The model obtained best performance while using dropout with probability between 0.2 - 0.35.

3.3.4.3. Learning Rate

We implement an adaptive learning rate (LR) scheduler based on epoch decay. This ensures a time-dependent relationship with LR towards finding the optimal local minima as higher learning rates inhibit the discovery of local minimum as the number of epochs increases. Based on empirical analysis, the optimal setting was achieved with LR 0.008 and halved every 5 epochs.

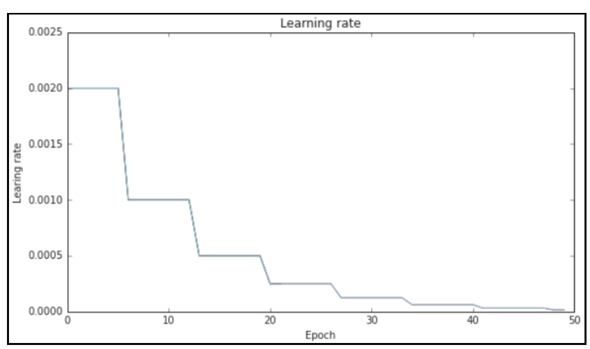

Figure 1: Learning rate decay across 50 epochs.

3.3.4.4. Batch Size

Experimented with 32, 64, 128, 256. The model produced best results with batch size 32.

3.4. Race, Age & Gender Classification

The key aspects in which this implementation was different from the one above are in terms of objective, model design and architecture.

This model aims to classify a given person's gender and categorise their race as well as age into the respective bins, simultaneously.

The key difference in model design is the number of branches created after feature extraction. This model creates three branches, one each for gender, age, and race attributes, as compared to the previous model's 2 branches.

The Race network characteristics are outlined in Table 4.

Input	Output of Feature Extraction
Flatten	Input
FC Layers	128-64-32-5
Dropout FC	0.2 - 0.3
Loss Function	Categorical Cross Entropy
Batch Size	32
Learning rate	Decaying LR halved every 5 epochs - initial LR: 0.008
Normalization	Batch
Activation	Relu
Activation Outputs	Softmax
Optimizer	Adam

Table 4: Race Network Characteristics

4. RESULTS

Table 5 outlines the model iterations that were implemented on both datasets.

Model	Dataset	# Age category	Epoch	Class weights
M1	Adience	25	22	False
M1	UTK	4	22	False
M2	Both	7	22	False
M3	Both	7	50	False
M4	Both	7	50	True

Table 5: Summary of All Model Configurations.

M1 for UTK and Adience datasets is different because Adience has 25 unique age values. Some values are specific (age = 10) while others are classified as a range (20,23). Therefore the datasets are standardized by converting the range values to an integer by computing the mean of the two endpoints. Subsequently, these values are binned into 7 distinct age groups of range 10 (M2 - M4) as Adience does not support individual age values from age 60 onwards, and thus all age's beyond 60 are assigned to a single class.

Class weights were used to try and mitigate the class imbalance in both the distributions in terms of ages.

4.1. UTK Database

4.1.1. Age, Gender and Race Classifier Validation Performance

Model	Overall Loss	Age	Gender	Race
M1	1.86 ± 0.39	71.5 ± 6.8	89.8 ± 2.0	76.1 ± 7.8

M2	2.20 ± 0.42	57.0 ± 7.8	90.0 ± 1.7	76.8 ± 8.1
M3	2.34 ± 0.26	58.8 ± 2.3	90.2 ± 1.6	79.2 ± 4.9
M4	2.42 ± 0.33	57.6 ± 6.4	90.3 ± 1.4	78.1 ± 4.4

Table 6: Validation Performance of all models on UTK Data.

4.1.2. Overall Model Loss

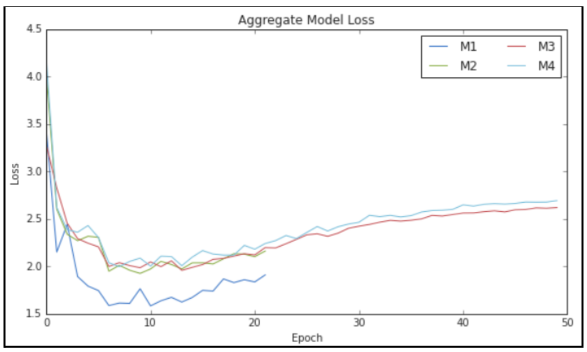

Figure 2: UTK Overall Model Loss.

4.1.3. Age Classifier

Age classifier validation accuracy for UTK shown in Figure 3.

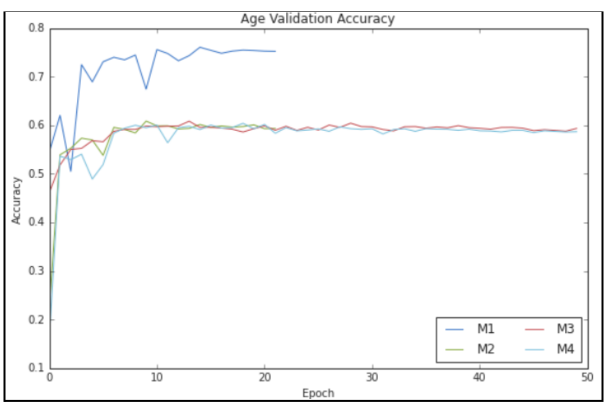

Figure 3: UTK Age classification performance (validation accuracy)

Age classifier validation loss for UTK shown in Figure 4.

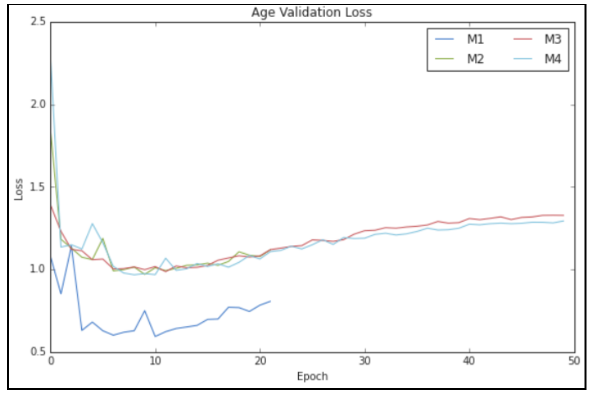

Figure 4: UTK Age classification performance (validation accuracy)

4.1.4. Gender Classifier

Gender classifier validation accuracy for UTK shown in Figure 5.

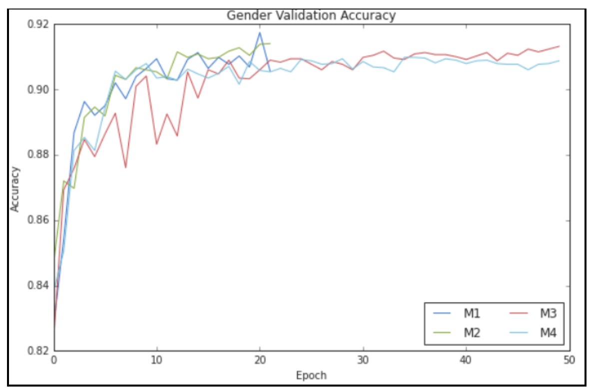

Figure 5: UTK Gender classification performance (validation accuracy)

Gender classifier validation loss for UTK shown in Figure 6.

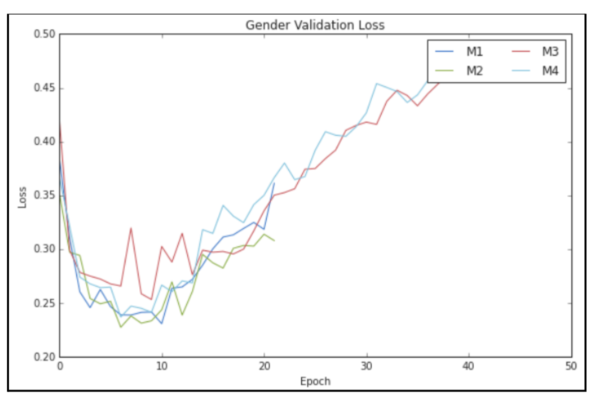

Figure 6: UTK Gender classification performance (validation loss)

4.1.5. Race Classifier

Race classifier validation accuracy for UTK shown in Figure 7.

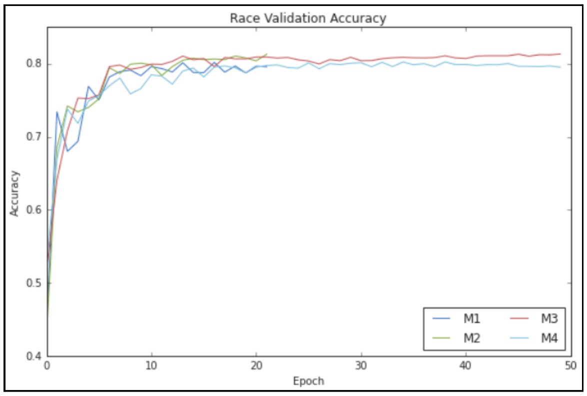

Figure 7: UTK Race classification performance (validation accuracy)

Race classifier validation loss for UTK shown in Figure 8.

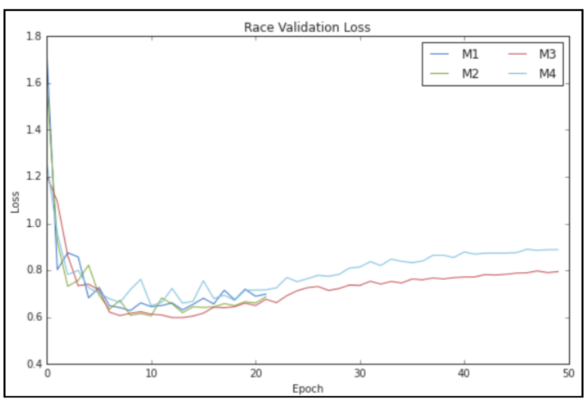

Figure 8: UTK Race classification performance (validation loss)

4.2. Adience Dataset

4.2.1. Age and Gender Classifier Validation Performance

Model	Overall	Age	Gender
M1	1.87 $\pm$ 0.36	50.9 $\pm$ 9.0	86.4 $\pm$ 7.3
M2	1.41 $\pm$ 0.29	62.9 $\pm$ 8.0	86.8 $\pm$ 5.8
M3	1.43 $\pm$ 0.21	67.0 $\pm$ 7.6	89.6 $\pm$ 4.7
M4	1.49 $\pm$ 0.18	66.2 $\pm$ 6.8	89.8 $\pm$ 4.6

Table 7: Validation Performance of all models on Adience Data.

4.2.2. Overall Model Loss

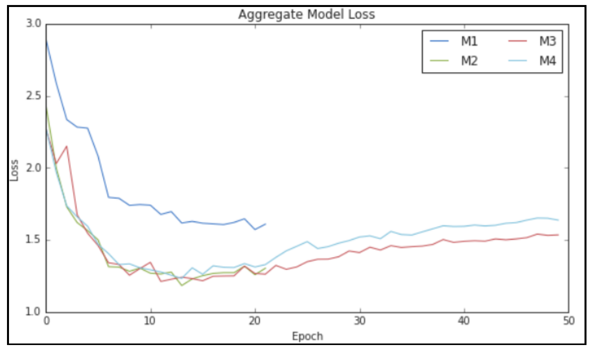

Figure 9: Adience Overall Model Loss.

4.2.3. Age Classifier

Age classifier validation accuracy for Adience shown in Figure 10.

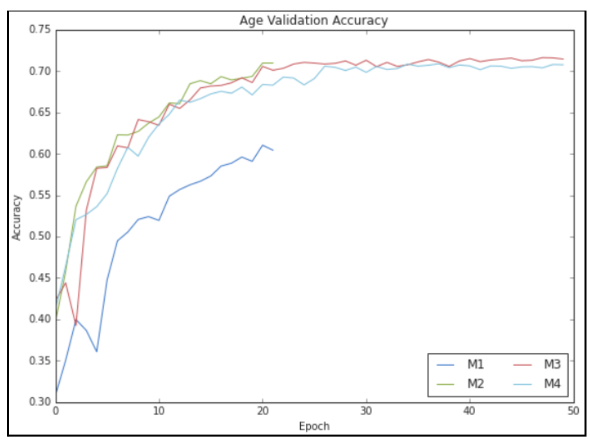

Figure 10: Adience Age classification performance (validation accuracy)

Age classifier validation loss for Adience shown in Figure 11.

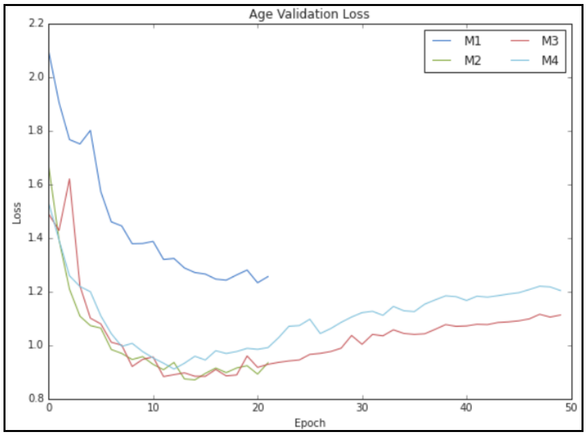

Figure 11.: Adience Age classification performance (validation loss)

4.2.4. Gender Classifier

Gender classifier validation accuracy for Adience shown in Figure 12.

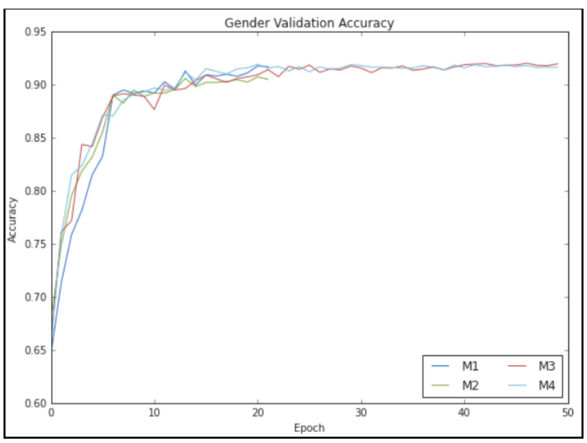

Figure 12: Adience Gender classification performance (validation accuracy)

Gender classifier validation loss for Adience shown in Figure 13.

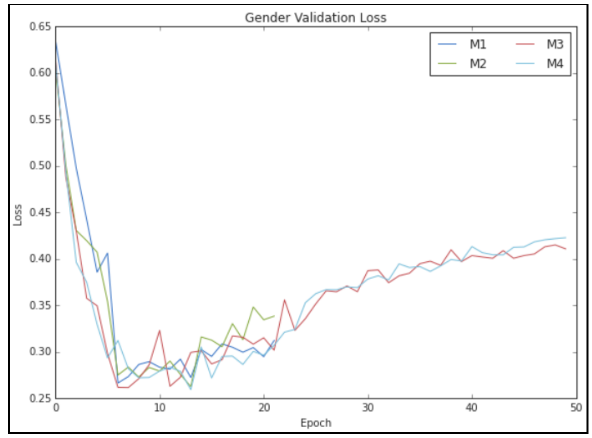

Figure 13: Adience Gender classification performance (validation loss)

4.3. Cross Evaluation

Cross evaluation was conducted to gauge a sense of generalizability of the models. This was done by evaluating the models trained on UTK on the Adience dataset and vice-versa. The experiment was set up and a 10 fold random sampling using a sample size of 100 random images for each model.

4.3.1. UTK Performance on Adience Data

Model	Age	Gender
M1	x	57.3 ± 6.8
M2	17.4 ± 3.3	56.8 ± 3.2
M3	23.0 ± 4.6	55.2 ± 4.6
M4	16.6 ± 4.3	56.2 ± 5.6

Table 8: Cross Evaluation performance of UTK on Adience Images. x means cross evaluation applicable.

4.3.2. Adience Performance on UTK Data

Model	Age	Gender
M1	x	70.0 ± 6.8
M2	30.8 ± 4.3	73.4 ± 5.8
M3	21.8 ± 3.8	74.7 ± 4.2

M4	28.6 ± 3.1	74.9 ± 3.4

Table 9: Cross Evaluation performance of Adience on UTK Images. x means cross evaluation applicable.

4.3.3. Sample Cross Evaluation Results

Sample cross evaluation of UTK models on Adience dataset for supervised race and gender and unsupervised race classification respectively, are given in Table 10. The labels are described as follows:

actual age range:predicted age range | actual gender : predicted gender | predicted race

M3 UTK	M4 UTK
 (30-40]:(30-40]|male:male|black	 (30-40]:(20-30]|male:male|asian
 (20-30]:(20-30]|female:male|black	 (20-30]:(20-30]|female:female|black
 (20-30]:(20-30]|male:female|black	 (20-30]:(20-30]|male:male|indian

Table 10: Sample Results of UTK on Adience Cross Evaluation.

Empirical analysis suggests that these models are not generalizable. This may be attributed to different distributions of the two datasets, especially for the age attribute as shown in Figure 14. The image types and quality also vary quite significantly across the two datasets. Moreover, both datasets are relatively small (~20,000 images each) and hence having a larger, uniform dataset will enable higher generalizability.

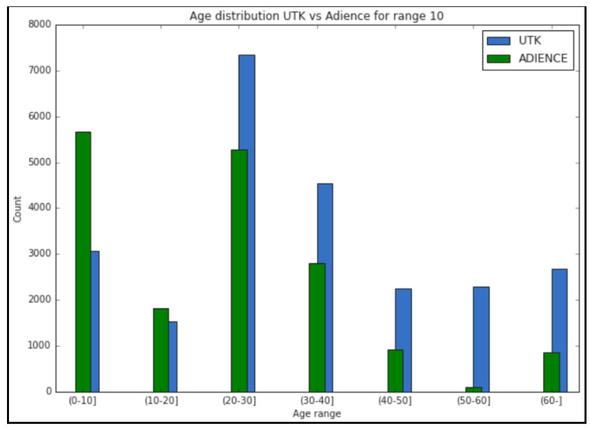

Figure 14: Difference in data distribution in terms of age for custom binning user in M2-M4.

4.4. Comparison with Existing Methods

Age and Gender Classification Comparison to existing approaches on the Adience Dataset.

4.4.1. Gender Classifier

Method	Accuracy
Best from [1]	86.8 ± 1.4
Best form [3]	77.8 ± 1.3
Best from [4]	79.3 ± 0.0
Proposed trained on UTK Dataset*	90.3 ± 1.4 (M4)
Proposed Trained on Adience	89.8 ± 4.6 (M4)

Table 11: Gender Classifier comparison with existing methods.

*Model is not exactly comparable to the others as all other model results are on the Adience Data

4.4.2 Age Classifier

Method	Accuracy
Best from [4]	45.1 ± 2.6
Best form [1]	50.7 ± 5.1
Proposed trained on UTK Dataset*	71.5 ± 6.8 (M1)
Proposed Trained on Adience	67.0 ± 7.6 (M4)

Table 12: Age Classifier comparison with existing methods.

*Model is not exactly comparable to the others as all other model results are on the Adience Data

5. CONCLUSION

We discussed a Convolution Neural Network-based approach to classify the age, gender, and race attributes from a facial image. This approach focused on simultaneously classifying these attributes to take advantage of the combined characteristics. The ensuing models achieved modest performance and in some instances surpassed the prevalent approaches. However, these models are not generalizable as analyzed during cross-evaluation. Therefore, future work includes implementing this approach over a larger, diverse, and uniformly distributed dataset in order to improve the model generalizability.

References

[1] G. Levi and T. Hassner, “Age and gender classification using convolutional neural networks.” in IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) workshops, 2015.

[2] Unfiltered faces for gender and age classification,” *Adience Dataset*. Available online https://talhassner.github.io/home/projects/Adience/Adience-data.html. [Accessed: 15-Nov-2021].

[3] E. Eidinger, R. Enbar, and T. Hassner. Age and gender estimation of unfiltered faces. Trans. on Inform. Forensics and Security, 9(12), 2014

[4] T. Hassner, S. Harel, E. Paz, and R. Enbar. Effective face frontalization in unconstrained images. Proc. Conf. Comput. Vision Pattern Recognition, 2015.

6. Appendix

Appendix A: Age & Gender Classifier

Appendix A.png

Can be found in the 4042_Appendix folder in the submission file.

Appendix B: Race, Age & Gender Classifier

Appendix B.png

Can be found in the 4042_Appendix folder in the submission file