Age Detection with Convolutional Neural Network

Nerya Reznickovich Haim Goldfisher Eliya Shlomo

9 March 2024

Abstract

This project presents a comprehensive study on age detection classifiers employing three distinct architectures: Softmax (50% accuracy), Neural Network with Softmax (61% accuracy), and CNN - Convolutional Neural Network (67.5% accuracy). The research utilizes the unit of UTKFace dataset with Age Facial dataset, a diverse collection of facial images annotated with age labels, categorized into seven age groups. The primary objective is to evaluate and compare the performance of these classifiers in accurately predicting age groups based on facial features. The experimental results provide insights into the strengths and weaknesses of each architecture, shedding light on their suitability for age detection tasks.

1 Introduction

Age estimation from facial images is a complex task in computer vision, prompting the application of deep learning techniques to enhance accuracy. This study, anchored by the UTKFace dataset[1] and the Age Facial dataset[2], delves into the challenges of predicting age accurately from facial features. The continuous nature of age poses inherent challenges in accurate prediction from facial images. To address this, we adopt a pragmatic approach, discretizing age into seven groups: '0-3', '4-12', '13-21', '22-35', '36-50', '51-69', and '70+'. This transformation turns age estimation into a classification problem, where the model predicts the probability distribution across predefined age categories[3]. Treating age prediction as classification "relaxes" the problem a bit, making it easier to solve — typically, we don't need the exact age of a person; a rough estimate is sufficient[4].

The significance of age estimation within deep learning extends to diverse applications, from personalized user experiences to bolstering security and surveillance systems. This project underscores the pivotal role of deep learning in overcoming the nuanced challenges of age prediction from facial images. This project aims to develop and evaluate age estimation models with varying architectures. Beginning with simpler models like softmax classification, our study progresses to explore more sophisticated approaches, particularly emphasizing convolutional neural networks (CNNs). The objective is to achieve accurate age predictions and provide a comparative analysis of each architecture's strengths and weaknesses. In subsequent sections, the paper delves into related work and required background, followed by a detailed project description, experiments/simulation results, and a comprehensive analysis of the findings.

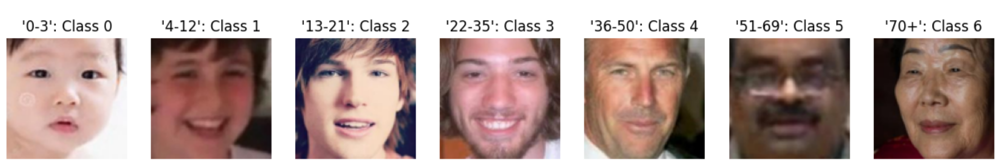

Figure 1. Age group classes example.

2 Related Work and Background

2.1 Related Work

2.1.1 Incorporating "Facial Age" Dataset

Recognizing the importance of achieving balanced model performance across all age groups, the Facial Age Dataset was selectively incorporated into the pre process for "weaker" age categories. This targeted integration aimed to mitigate potential biases and inaccuracies that may arise due to imbalanced data distribution.

2.1.2 Train-Test and Train-Validation Split

Before applying oversampling and augmentation techniques, a thoughtful data partitioning was implemented to maintain a balanced and unbiased distribution. Initially, the dataset underwent an 80-20 train-test split (3031 test samples), determined by the size of the smallest age group to ensure proportional representation across categories. Subsequently, the remaining training set was subjected to an 80-20 train-validation split (1211 validation samples), once again based on the size of the smallest age group. This sequential approach was crucial in preventing data overlap between the training, validation, and testing sets.

2.1.3 Data Augmentation and Balanced Sampling Strategy

In addressing the imbalanced distribution of age groups within the dataset, a thoughtful decision was made to prioritize oversampling over undersampling. This choice was influenced by a substantial disparity in image counts among age categories, notably with the '22-35' age group containing almost 10,000 images, while '0-3', '4-12', '13-21', and '70+' age groups each had less than 4,000 images. To mitigate the potential drawbacks of oversampling, a nuanced strategy was employed. The '22-35' age group underwent strategic undersampling, reducing its count to 4,000 instances, thereby preventing undue dominance during model training. Subsequently, smaller age groups were carefully augmented through oversampling, and further diversity was introduced by applying augmentation techniques to every age group until reaching a uniform count of 5,000 images. This comprehensive approach aims to strike a balance, avoiding excessive undersampling while still addressing the challenges of overfitting[5], resulting in a more balanced and diverse training dataset.

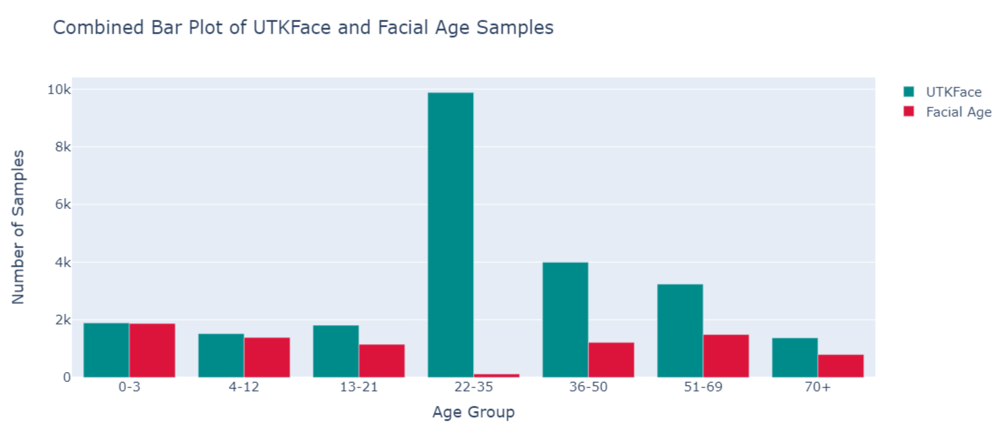

Figure 2. Distribution of Samples by Age Group.

2.1.4 Normalization

In the chosen approach, pixel values are normalized by dividing them by 255.0. This rescaling operation adjusts intensity values to a consistent range between 0 and 1, promoting uniform data representation. The normalization process contributes to improved model convergence during training, and helps alleviate the influence of varying illumination conditions.

2.2 Background

2.2.1 The UTKFace and Facial Age Datasets

The UTKFace Dataset is a comprehensive resource for facial analysis, offering over 20,000 annotated face images covering a wide age spectrum from 0 to 116 years. With rich diversity in expressions, poses, and environmental conditions, it serves as a valuable asset in various computer vision tasks.
Complementing this, the Facial Age Dataset significantly augments the dataset by adding annotated facial images focused on age. This inclusion expands the dataset size, contributing to a more extensive and diverse training set.

3 Project Description

3.1 Convolutional Neural Network Model Architecture

The model comprises two convolutional layers, each followed by max-pooling layers. The first convolutional layer utilizes 64 filters of size 3×3 with ReLU activation, adept at capturing basic facial features such as edges and textures. Subsequent to the first convolutional operation, a max-pooling layer with a 2×2 window reduces spatial dimensions, focusing on the most salient features. The second convolutional layer increases the filter count to 128, allowing the model to learn more complex features such as parts of the face relevant to age estimation. Another max-pooling step follows, further condensing the feature representation.
After the convolutional and pooling layers, the model proceeds to flatten the feature maps, transforming them into a one-dimensional vector. This serves as the input for the subsequent dense layers. The first dense layer consists of 256 neurons with ReLU activation, forming a fully connected layer that enables the network to learn non-linear combinations of high-level features extracted by the preceding convolutional layers. To prevent overfitting, a Dropout layer with a rate of 0.5 is applied after this dense layer. Following this, a second dense layer is introduced, containing 128 neurons with ReLU activation, contributing to a deeper representation of the learned features. Another Dropout layer with a rate of 0.5 is applied after the second dense layer for additional regularization. The final dense layer, featuring 7 neurons and softmax activation, produces the probability distribution across the seven age groups.
The model is compiled with the Adam optimizer, with a learning rate of 0.0001, and uses sparse categorical crossentropy as the loss function, suitable for multi-class classification. Accuracy serves as the metric for performance evaluation. Training is conducted over 100 epochs with a batch size of 100, using both training and validation datasets to fine-tune and validate the model's performance iteratively.

4 Experiments/Simulation Results

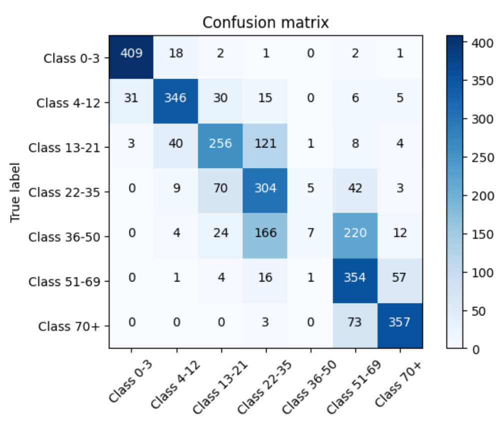

Figure 3. Confusion Matrix for Optimal CNN Model.

4.1 Softmax Model and Neural Network with Softmax

The Softmax Model, featuring a single Dense layer with softmax activation, classifies flattened one-dimensional input into predefined age categories. Compiled with the Adam optimizer (learning rate 0.0001), sparse categorical crossentropy serves as the loss function, it achieves an accuracy 50% on test.
The Neural Network with Softmax enhances complexity with additional Dense layers, utilizing ReLU activation functions. Comprising five layers with increasing neurons, it concludes with an output layer for age classification. Compiled similarly to the Softmax Model, it achieves an accuracy 61% on test.
Both architectures establish a baseline for age detection, paving the way for more sophisticated model exploration.

4.2 Regularization

In the pursuit of enhancing the robustness and generalization capabilities of the Convolutional Neural Network (CNN) model, various regularization techniques were explored[6]. The primary regularization methods investigated included L2 regularization, dropout, and batch normalization.

4.2.1 L2 Regularization

L2 regularization, also known as Ridge, was employed to penalize large weights in the network, aiming to prevent overfitting[7]. Despite its effectiveness in some scenarios, the application of L2 regularization did not yield substantial improvements in the age detection model's performance (57% on test).

4.2.2 Dropout

Dropout, a technique involving randomly deactivating a fraction of neurons during training, emerged as the most successful regularization method (67.5%). By preventing reliance on specific neurons and promoting a more distributed learning process, dropout effectively mitigated overfitting. The inclusion of dropout layers after the first and second dense layers significantly contributed to the model's robustness and improved performance on the test set.

4.2.3 Batch Normalization with Early Stopping

Batch normalization, which normalizes the inputs of a layer, was integrated into the model to further enhance training stability and performance. The combination of batch normalization with an early stopping mechanism proved to be particularly effective. Early stopping was implemented to halt training when no improvement in validation accuracy was observed, preventing overfitting. Remarkably, this combination yielded promising results after only 16 epochs, with a test accuracy of 67%.

5 Previous Attempts

5.1 First Attempt: Basic Modeling

In the preliminary phase, only UTKFace dataset was constructed by randomly sampling 1000 images from each of the seven designated age groups, resulting in a total of 7000 images. Subsequently, all images underwent uniform resizing to 100x100 pixels and conversion to grayscale. After this preprocessing stage, the dataset was partitioned into training (80%) and testing (20%) subsets. Within the training set, an additional segmentation into training (80%) and validation (20%) subsets was carried out. Initial modeling efforts included the application of the default softmax algorithm without additional layers, serving as a standalone investigation. The accuracy results for this model on the validation set were around 40% accuray , and on the test set were the same. Following this, a specific two-layer softmax classification model was introduced, establishing a distinct phase in our experimentation. The accuracy results for this model on the validation set were 45%, and on the test set were 48% This step paved the way for the subsequent incorporation of the CNN, marking the early stages of refining the age estimation model. The accuracy results for the CNN model on the validation set were 53%, and on the test set were 54%.

Figure 4. Visualization of Dimensionality Reduction.

5.2 Second Attempt: Augmentation Strategies

In the second phase of experimentation, an augmentation sequence was applied to the training images to enhance the robustness of the age estimation model. The sequence, defined using the imgaug library, included various transformations such as horizontal flipping (50% probability), random rotations (-15 to 15 degrees), Gaussian blur (sigma 0 to 3.0), pixel value multiplication (0.8 to 1.2), contrast normalization (0.8 to 1.2), and gamma contrast adjustment (50% probability). The augmented training dataset was created by applying these transformations, and the augmented images were then combined with the original training set. This augmentation strategy aimed to diversify the dataset, contributing to improved model generalization. The resulting dimensions of the augmented training dataset were verified through self-check outputs, while the dimensions of the testing dataset remained unchanged. Although the augmentation strategy aimed to diversify the dataset and contribute to improved model generalization, the observed improvements in results were moderate.

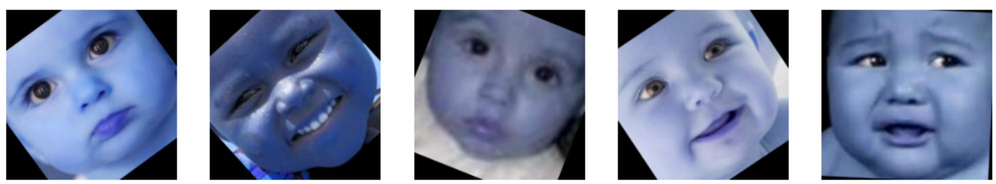

Figure 5. Augmented Image Examples.

6 Conclusions

6.1 Discussion

The findings underscore the significant potential of CNNs in handling the intricacies involved in age detection from facial images. Notably, the CNN model outperformed the other architectures.
The impact of age group partitioning on the predictive accuracy of the models further emphasizes the importance of considering how age categories are defined. The study adopted age partitions based on demographic and physiological considerations; however, findings suggest that alternative partitioning could potentially enhance the model's performance. The challenge in distinguishing between ages, especially near the predefined age group thresholds, underscores the complexity of the age estimation task.

6.2 Future Work

6.2.1 Improved Results and Fine-Grained Age Estimation

The accuracy of age detection model has shown promise, but there is room for improvement, especially with the incorporation of more data and higher resolution. In advancing age detection models, a crucial direction involves transitioning from age group estimation to precise age prediction. This entails exploring high-resolution imagery to capture nuanced facial details for improved accuracy, albeit with potential increased computational demands. Moreover, expanding the dataset size to encompass a broader range of ages is pivotal, enhancing the model's generalization capabilities. The integration of these approaches aims to refine age detection models, enabling more accurate and detailed age predictions as technology progresses.

6.2.2 Multi-Face Age Estimation through Improved Preprocessing

The UTKFace and Facial Age datasets primarily provide cropped images of individual faces, limiting its applicability to scenarios with multiple people in a single image. An intriguing avenue for future exploration involves the development of preprocessing techniques capable of detecting and extracting faces from crowded scenes. By incorporating state-of-the-art face detection algorithms, such as those based on deep learning architectures, we can extend the applicability of age estimation models to diverse scenarios with numerous individuals. This approach would involve not only refining existing face detection models but also integrating them seamlessly into the age estimation pipeline. The resulting system could provide age estimates for each person in a given image, contributing to a more comprehensive understanding of age distribution within complex social settings.

7 Code Snippets

```
# Define the augmentation techniques
augmentation_techniques = [
    iaa.Affine(rotate=(-45, 45)),  # random rotation between -45 and 45 degrees
    iaa.Fliplr(0.5),  # horizontal flip with 50% probability
]
```

Figure 6. Augmentation Techniques.

```
# Create a Sequential model
model = Sequential()

# Convolutional layers
model.add(Conv2D(64, kernel_size=(3, 3), activation='relu', input_shape=(200, 200, 1)))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Conv2D(128, kernel_size=(3, 3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))

# Flatten layer
model.add(Flatten())

# Dense layers with dropout
model.add(Dense(256, activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(128, activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(7, activation='softmax'))

# Compile the model
model.compile(optimizer=tf.keras.optimizers.Adam(learning_rate=1e-4),
              loss='sparse_categorical_crossentropy',
              metrics=['accuracy'])

history = model.fit(x_train, y_train, epochs=100,batch_size=100, validation_data=(x_valid, y_valid))
```

Figure 7. CNN Code Overview.

References

[1] *UTKFace Dataset*: https://www.kaggle.com/datasets/jangedoo/utkface-new/code

[2] *Facial Age Dataset*: https://www.kaggle.com/datasets/frabbisw/facial-age

[3] Phyo-Kyaw Sai, Jian-Gang Wang, and Eam-Khwang Teoh. *Facial age range estimation with extreme learning machines.*

[4] Adrian Rosebrock. *OpenCV Age Detection with Deep Learning*: https://pyimagesearch.com/2020/04/13/opencv-age-detection-with-deep-learning/

[5] Christina Ellis. *Oversampling vs undersampling for machine learning*: https://crunchingthedata.com/oversampling-vs-undersampling/

[6] *How to Treat Overfitting in Convolutional Neural Networks*: https://www.analyticsvidhya.com/blog/2020/09/overfitting-in-cnn-show-to-treat-overfitting-in-convolutional-neural-networks/

[7] Ujwal Tewari. *Regularization — Understanding L1 and L2 regularization for Deep Learning*: https://medium.com/analytics-vidhya/regularization-understanding-l1-and-l2-regularization-for-deep-learning-a7b9e4a409bf