Age Prediction with Deep Convolutional Neural Networks

Rashmi Hiregoudar
Hochschule Furtwangen
rhi55178@stud.hs-furtwangen.de

Stephen Thomas
Hochschule Furtwangen
sth58964@stud.hs-furtwangen.de

***Abstract*—Accurately estimating a person's age from facial images in unconstrained environments remains a challenging task due to factors such as lighting variations, head pose, and facial expressions. In this study, we present a deep transfer learning approach using the ResNet50V2 architecture, initially trained on ImageNet and adapted for age regression [1]. The model is trained on the UTKFace dataset using a two-phase strategy: an initial feature extraction stage with frozen layers, followed by fine-tuning of the entire network. To enhance robustness and generalization, the training pipeline incorporates data augmentation techniques. Experimental evaluation results in a Mean Absolute Error (MAE) of 10.01 years, demonstrating that the proposed method performs reliably under real-world, unconstrained conditions. *Index Terms*—Age Estimation, Deep Learning, Transfer Learning, Convolutional Neural Networks, Age Analysis**

I. Introduction

Predicting a person's age from a facial image is a fascinating yet challenging task within the field of computer vision. It has become increasingly relevant due to its broad range of applications, including biometric verification, demographic profiling, personalized content delivery, healthcare monitoring, and enforcing age-based access controls [5]. As technology continues to integrate more deeply into our daily lives, the ability to estimate age accurately and efficiently has grown in importance.

Unlike classification tasks such as detecting gender or facial expressions, age estimation is inherently more complex. Aging is a continuous, non-linear process influenced by genetic, environmental, and lifestyle factors. Furthermore, external variations such as lighting conditions, facial expressions, head pose, and image quality make it even harder to generalize across diverse images. These factors introduce a high degree of variability, making age prediction a difficult problem to solve—especially in real-world, unconstrained scenarios.

Earlier methods typically relied on handcrafted features and traditional machine learning algorithms like support vector machines or random forests. While these approaches had some success, they often lacked the robustness needed to perform well across diverse datasets, particularly when faced with non-ideal conditions.

The rise of deep learning has brought significant improvements in this area. Convolutional Neural Networks (CNNs), in particular, have proven highly effective in extracting meaningful and hierarchical features directly from images, eliminating the need for manual feature engineering [1]. However, training such deep networks from scratch requires extensive labeled data and computational resources, which are not always feasible.

To overcome these limitations, transfer learning has emerged as a practical solution. In this project, we explore a transfer learning-based approach to age estimation using deep CNNs. We build upon the ResNet50V2 architecture [4], which has been pre-trained on the large-scale ImageNet dataset. By modifying its final layers for regression, we adapt the network specifically for predicting age.

We conduct our experiments using the UTKFace dataset [2], a widely used collection of facial images labeled with age, gender, and ethnicity. The dataset offers a rich variety of facial appearances, making it well-suited for training models in real-world conditions. Our training strategy involves two phases: initially freezing the base layers to preserve learned features, followed by fine-tuning the entire network to adapt more closely to the age prediction task.

Through extensive testing, we demonstrate that our model achieves a low Mean Absolute Error (MAE) on the test set, indicating strong performance even under challenging conditions. This work highlights the effectiveness of combining transfer learning with deep neural networks for age estimation and contributes to ongoing research in facial analysis.

II. Dataset

This project utilizes the UTKFace dataset, a well-known and widely adopted benchmark for facial analysis tasks [9]. It contains over 20,000 images of human faces, each labeled with three essential attributes: age, gender, and ethnicity. The age annotations span from

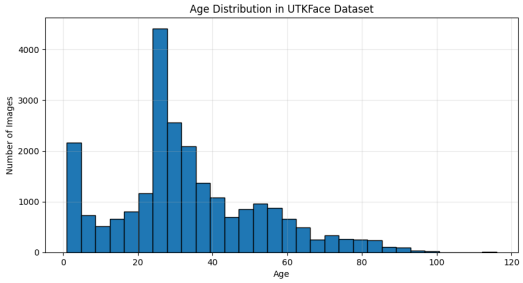

Fig. 1: Age distribution in the UTKFace dataset, showing a concentration in the 20-40 age range and fewer samples at the extremes.

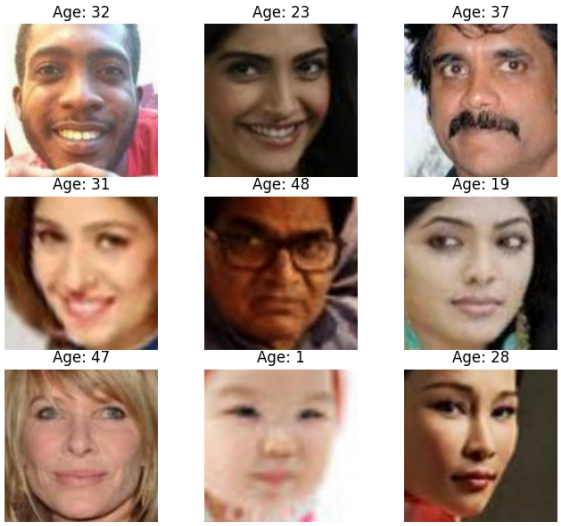

Fig. 2: Sample Pictures with Age

0 to 116 years, making the dataset particularly valuable for modeling age estimation across all stages of life. The dataset, containing over 20,000 images, was divided into three subsets: a training set with 16,577 images (70%), a validation set with 3,552 images (15%), and a test set with 3,552 images (15%).

A unique aspect of this dataset is its file-naming convention, where each image is named following the format: age_gender_ethnicity_date_and_time.jpg This format simplifies the process of extracting labels directly from filenames, removing the need for additional annotation files or metadata processing.

What sets the UTKFace dataset apart is its rich diversity. It includes individuals from various ethnic backgrounds, across different age groups and genders. The images reflect real-world, unconstrained conditions—with variations in lighting, facial expressions, image quality, and even partial occlusions such as eyeglasses, hands, or hats. This variety adds to the complexity of the task but also improves the model's ability to generalize to new, unseen data.

A histogram showing the distribution of ages in the UTKFace dataset would visually confirm the point made in the Results section about class imbalance (Fig. 1).

III. Architecture

The age estimation system developed in this project is based on a deep convolutional neural network (CNN) framework that leverages transfer learning. At its core, the model uses ResNet50V2 [4] — a 50-layer residual network known for its efficient feature extraction and stable training behavior. Originally designed for image classification tasks on the large-scale ImageNet dataset, ResNet50V2 is repurposed here for a regression task, where the goal is to predict age as a continuous variable.

To tailor the model for age estimation, we begin by removing the original classification head (`include_top=False`) and retaining the convolutional base. A custom regression head is added on top. A Global Average Pooling 2D layer was added after the convolutional base. This layer significantly reduces the number of model parameters compared to a traditional Flatten layer, which helps in mitigating overfitting and reducing computational cost. A Dropout layer with a rate of 0.5 was included as a regularization technique. This randomly sets half of the inputs to 0 during training, preventing co-adaptation of neurons and improving the model's ability to generalize to unseen data. Finally, a Dense output layer with a single neuron and a linear activation function allows the model to output continuous age predictions.

The structure of the custom head is summarized below:

TABLE I: Summary of Custom Head Structure

Layer Type	Output Shape	Param #
resnet50v2 (base)	(None, 7, 7, 2048)	23,564,800
global_average_pooling2d	(None, 2048)	0
dropout	(None, 2048)	0
dense (prediction)	(None, 1)	2,049

Training is performed in two stages to balance learning efficiency and generalization. Initially, only the newly added regression layers are trained while the pre-trained base remains frozen. This helps the model learn task-specific patterns related to age without disrupting the general visual features learned from ImageNet. In the second phase, we unfreeze the entire network and fine-tune it using a lower learning rate. This step allows the model to refine the pre-trained weights and better adapt to the age estimation task.

The model is compiled using the Adam optimizer [8], with Mean Absolute Error (MAE) serving as both

the loss function and the performance metric. MAE is especially suitable for this regression problem, as it directly measures the average magnitude of errors in predicted ages.

Formally, the goal is to learn a mapping function $f(x;\theta)$, where x is the input facial image, $\hat{y} = f(x;\theta)$ is the predicted age, and θ represents the learnable model parameters. The loss function is defined as:

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

where y_i is the true age, $\hat{y}_i$ is the predicted age, and n is the number of samples in the batch.

IV. Implementation and Training Strategy

IV.1 Data Pipeline

A robust data pipeline was constructed using TensorFlow's `tf.data` API to efficiently handle data loading, preprocessing, and augmentation [6].

a) Label Extraction

Age labels were extracted by parsing the image filenames. String splitting functions were used to isolate the age component, which was then converted to an integer data type.

b) Image Loading and Resizing

Images were loaded using `tf.io.read_file` and decoded into tensors. As the ResNet50V2 architecture expects a fixed input size, all images were resized to 224×224 pixels using bilinear interpolation.

c) Pixel Value Normalization

Pixel values were normalized from the standard $[0, 255]$ integer range to the $[-1, 1]$ floating-point range, which is the expected input format for the ResNetV2 model.

d) Data Augmentation

To improve model generalization and mitigate overfitting, online data augmentation was applied to the training set. This included random horizontal flips and random brightness adjustments, where the brightness factor was varied by a maximum delta of 0.2.

The pipeline was further optimized with batching and prefetching to ensure maximum GPU utilization during training [6].

IV.2 Training Protocol

The training process was programmed using the Keras API [7] and was divided into two distinct phases, guided by the hyperparameters listed in Table II.

Training was monitored using TensorFlow Keras callbacks [7]. A `ModelCheckpoint` callback was configured to save the weights of the model that achieved the lowest validation MAE, ensuring that the best-performing model was retained regardless of potential overfitting in later epochs.

TABLE II: Hyperparameter Settings for Training Phases

Hyperparameter	**Value**	**Phase**
Image Dimensions	$224 \times 224 \times 3$	Both
Batch Size	32	Both
Optimizer	Adam	Both
Learning Rate	0.001	1 (Feature Extraction)
Learning Rate	1×10^{-5}	2 (Fine-Tuning)
Epochs	10	1 (Feature Extraction)
Epochs	15	2 (Fine-Tuning)
Loss Function	Mean Absolute Error	Both

a) Feature Extraction Phase

The regression head was trained while keeping the ResNet50V2 base frozen for 10 epochs.

b) Fine-Tuning Phase

The entire model was unfrozen and trained for 15 additional epochs with a reduced learning rate (1×10^{-5}).

This two-stage strategy allows the model to first learn high-level, task-specific features in the new layers and then gently adapt the pre-learned, low-level features of the entire network to the specific domain of age estimation.

V. Results and Evaluation

The final model was tested on the designated test dataset, where it achieved a Mean Absolute Error (MAE) of approximately 10.01 years. While this result is higher than an ideal target, it is still considered reasonable given the unconstrained and diverse nature of the UTKFace dataset, which includes a wide range of age groups, facial expressions, lighting conditions, and occlusions [9].

A scatter plot comparing the predicted ages with the actual ages (Fig. 3) shows a clear positive correlation, indicating that the model generally predicts in the correct direction. However, some variance is observed, particularly at the youngest and oldest ends of the age spectrum. This is likely due to class imbalance in the dataset, with fewer samples representing infants and elderly individuals, which can reduce the model's accuracy in these ranges [2].

As shown in Figure 5, a set of test images is displayed alongside their true and predicted ages. Visual inspection confirms that the model is often successful at estimating the correct age group, though some errors are noticeable. For example, in one extreme case, the model predicted an age of -1.05 years for a sample with a true age of 1 year, highlighting potential weaknesses in handling edge cases like very young children.

Despite such outliers, the model shows promising generalization capabilities. The overall performance suggests that the model has learned meaningful age-related features from the data. However, there is still room for improvement. Potential enhancements could include balancing the dataset across age groups,

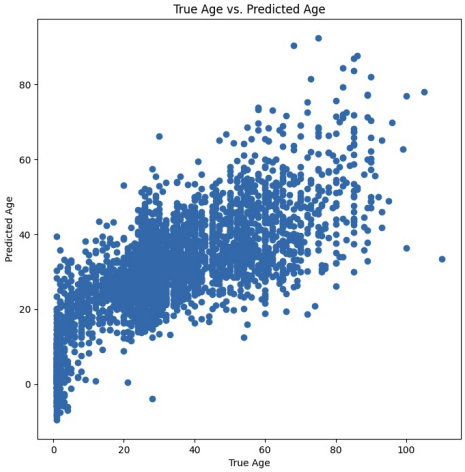

Fig. 3: Scatter plot of predicted vs. actual ages.

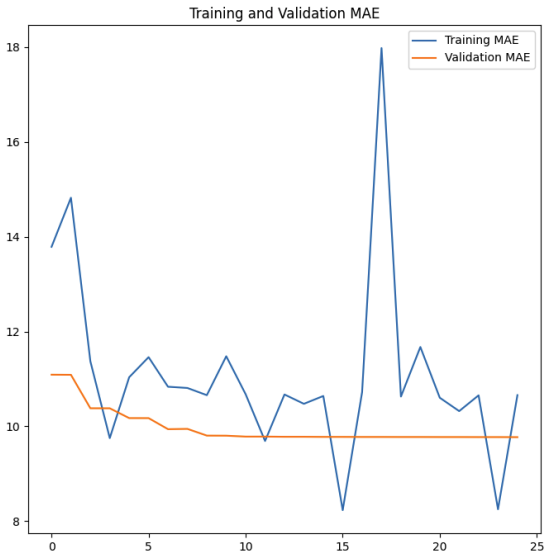

Fig. 4: MAE trends during training

incorporating attention mechanisms to focus on age-relevant facial regions, or using multi-task learning to jointly learn age and other facial attributes like gender or ethnicity [5].

The prediction process was completed without critical runtime errors. The TensorFlow runtime did generate standard messages indicating that oneDNN optimizations were active, which may cause slight numerical variations due to floating-point operations being executed in different orders [6]. These warnings are expected and do not indicate any fault in the model's logic or accuracy.

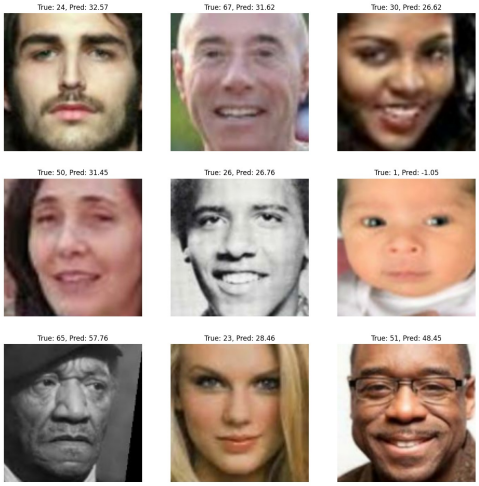

Fig. 5: Example test images with true and predicted ages.

VI. Conclusion

This study confirms that deep transfer learning, particularly using a pre-trained ResNet50V2 model [4], is a viable and effective approach for age estimation from unconstrained facial images. The model, adapted with a custom regression head and trained in two phases, achieved reasonable accuracy on the diverse UTKFace dataset [9]. While performance was strong overall, especially in middle age ranges, challenges remain in predicting outlier age groups such as infants and elderly individuals. Future improvements could focus on addressing data imbalance, incorporating attention mechanisms to emphasize age-relevant features, and exploring multi-task learning frameworks that jointly predict age alongside other facial attributes like gender and ethnicity [5].

References

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2016, pp. 770–778.

[2] Z. Zhang, Y. Song, and H. Qi, "Age Progression/Regression by Conditional Adversarial Autoencoder," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2017.

[3] A. G. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv preprint arXiv:1704.04861*, 2017.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Identity Mappings in Deep Residual Networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 630–645.

[5] R. Ranjan, V. M. Patel, and R. Chellappa, "A Deep Learning Approach for Face Detection, Pose Estimation, and Landmark Localization," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, 2016.

[6] M. Abadi et al., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems," *arXiv preprint arXiv:1603.04467*, 2016.
[7] F. Chollet et al., "Keras," *https://keras.io*, 2015.
[8] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv preprint arXiv:1412.6980*, 2014.
[9] The UTKFace Dataset, *[Online]. Available: https://susanqq.github.io/UTKFace/*.