Age Detection Using Convolutional Neural Networks

A. Introduction and Problem Statement

Age detection from facial images is vital for security, healthcare, and marketing applications, yet remains challenging due to variations in aging influenced by genetics, lifestyle, and environmental factors. Traditional methods relying on handcrafted features often miss the complex patterns of human aging. Convolutional Neural Networks (CNNs), commonly used for image classification, require adaptation to address challenges like class imbalance, inter-age similarity, and image quality effects on performance.

While CNNs can automatically extract hierarchical features from facial images, existing strategies face limitations, including imbalanced datasets, noisy age labels, and inter-age similarities. A report by the National Institute of Standards and Technology (NIST) [1], a U.S. agency that sets measurement standards and technology, highlights critical limitations in age prediction algorithms, noting that accuracy varies across demographics due to factors such as image quality, gender, and age. Female faces typically have higher error rates than male faces, and different demographic groups exhibit varying prediction sensitivities. Although accuracy has improved over the past decade, current advancements may still fall short in applications requiring precise age verification. These constraints also extend to CNN based approaches where age classification primarily happens through features extracted by the CNN based on images provided.

This project addresses these constraints by leveraging a diverse dataset, incorporating advanced CNN architectures, and adopting techniques to enhance classification accuracy and generalization across age groups. To improve generalizability and accuracy, this project will assess pre-trained CNN-based models optimized for age classification, investigating architectures that can reliably classify age groups with high accuracy

B. Proposed Methodologies

This project utilizes three distinct image datasets to evaluate the performance of various pretrained models in age classification tasks.

The first dataset, FacialAge [2], includes 9,778 images of cropped faces in various age groups, with distribution shown in Fig. 1. Classes are arranged to maintain a 10% density difference, though age 49+ is underrepresented, and youth has high representation. This dataset consists primarily of sharp images. The second dataset, Adience [3], comprises 19,400 images from Flickr, taken with mobile phones (iPhone 5 or later), divided into 8 age categories, as shown in Fig. 2. Younger age groups are significantly represented, with substantial images of infants and toddlers. The third dataset, UTKFaces [4], has 20,000 images with metadata for age, gender, and ethnicity, categorized into 20 age groups, as shown in Fig. 3 The oldest age category (80+) has fewer than 100 images, while other classes have relatively consistent densities outside of "youth."

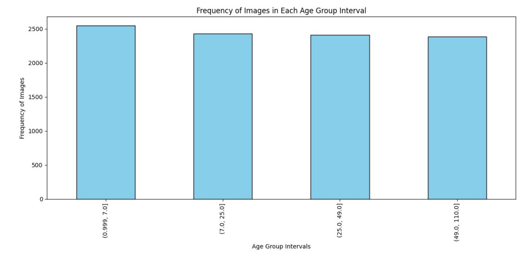

Figure 1. FacialAge classwise distribution

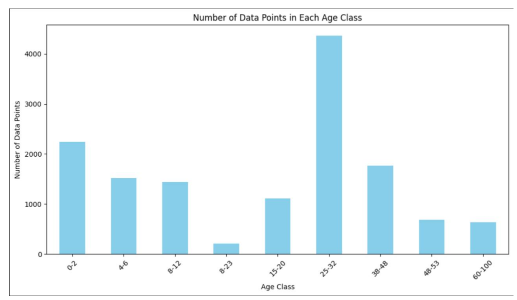

Figure 2. Adience classwise distripution

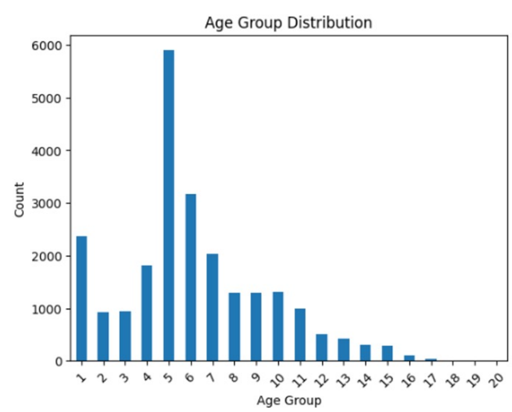

Figure 3. UTKFace classwise distripution

Images across all datasets were captured in natural settings, often outdoors, with varying levels of blur due to spontaneous, non-studio conditions. Some images contain multiple faces, adding complexity and realism. Each dataset exhibits diversity in age, gender, race, and ethnicity, capturing a wide range of human characteristics.

Image blurriness was quantified using Laplace variance [5], where low variance indicates less pixel color variation in blurry images. Images falling in the 1^{st} quartile for blur within each dataset were marked as "excessively hazy" and removed to improve model performance while maintaining dataset size. Noisy data, such as images with invalid or null values, were also excluded.

For model architecture, we employed VGGNet-16 [6], a deep CNN for image classification and feature extraction. VGGNet-16 uses 3x3 convolutional layers in a simple stacked configuration, which is effective for detecting key facial features. DenseNet121 [7], the second architecture, features dense blocks where each layer connects to all previous layers, making it particularly useful for smaller datasets by enhancing feature reuse. This dense connectivity retains fine-grained details such as textures, wrinkles, and facial structures, essential for precise age differentiation. Lastly, EfficientNet-B0 [8], optimized for both accuracy and efficiency, utilizes compound scaling to balance depth, width, and resolution, enabling robust classification with fewer parameters. We modified its classifier layer to output specific age group classes, supporting efficient and accurate age categorization. We have used the baseline models provided by PyTorch [9]. Furthermore, we have used PyTorch [9] transformers for our preprocessing.

C. Attempts at solving the problem

In this project, we aim to classify three large image datasets using three different pre-trained models, resulting in nine computations. Each dataset undergoes a preprocessing phase tailored to its specific content before being processed through the various models. Initially, data preprocessing is performed. The image preprocessing steps include resizing each image to 224×224 pixels for uniformity, followed by random adjustments in brightness, contrast, and saturation to introduce variation in lighting and color tones. Images are also randomly converted to grayscale, retaining three color channels, and undergo random perspective changes, simulating different viewing angles. A slight rotation of up to ± 20 degrees is applied to enhance orientation diversity, and images are randomly flipped along the y-axis. Finally, images are converted to tensors and normalized according to ImageNet [10] values, ensuring compatibility with pre-trained model standards.

The primary reason for VGGNet's [6] poor performance can be attributed to its deep, parameter-heavy architecture, which leads to slower convergence and difficulty in generalizing to new data, particularly in age classification tasks with complex and diverse features. From Table 1, it is clear that VGGNet [6] struggles to converge, as evidenced by its low accuracy of 31.40% on the test set for the Adience [3] dataset, alongside a high validation loss of 1.9006. These values indicate that VGGNet [6] is unable to effectively learn from the dataset. In contrast, DenseNet and EfficientNet [8], which utilize more efficient architectural features like dense connections and compound scaling, performed much better. DenseNet [7] achieved a training accuracy of 73.69% and a test accuracy of 72.82% on the Adience [3] dataset, with a validation loss of 0.73, demonstrating steady improvement and better generalization. EfficientNet [8] outperformed both models, achieving 79.03% training accuracy, 74.45% validation accuracy, and 76.05% test accuracy, with a validation loss of 0.71, suggesting superior performance on the same dataset. These results highlight the limitations of VGGNet's [6] architecture, especially its inability to effectively manage the dataset's complexities, compared to the more modern, efficient architectures of DenseNet [7] and EfficientNet [8]. The lack of advanced features like skip connections and attention mechanisms in VGGNet [6] limits its ability to focus on important features.

Dataset	Model	Accuracy (%)	Loss
UTKFaces	EfficientNet	72.82 (Training)	1.53
Facial Age	VGGNet	35 (Training)	1.8
Adience	DenseNet	73.69 (Training) 71.28 (Validation) 72.82 (Test)	0.6736 0.73
Adience	EfficientNet	79.03 (Training) 74.45 (Validation) 76.05 (Test)	0.57 0.71
Adience	VGGNet	31.40 (Training) 32.90 (Validation 31.40 (Test)	1.9006 1.9006

Table 1. Summary of Preliminary Results

D. Future Steps

Moving forward, several steps will be taken to further enhance model performance and optimize accuracy, especially for VGGNet [6]. First, we will test different combinations of datasets and models to complete the remaining four model trainings. Additional hyperparameter tuning will be conducted, focusing on key parameters such as learning rate, batch size, and the number of epochs to maximize model accuracy. Additionally, a comparative analysis will be conducted using tools like Grad-CAM or t-SNE on different model architectures to evaluate their performance across various datasets. This step will help us identify the strengths and weaknesses of each model and understand how well each one performs on datasets with differing characteristics, such as image quality, class distribution, and noise levels. Future work will also involve implementing transfer learning to leverage pre-trained models, such as those trained on ImageNet [10], as a starting point. We plan to choose ResNet [11] and ShuffleNet [12] for that purpose. This approach will enable the model to benefit from learned feature representations, potentially improving performance with reduced training time. By combining transfer learning with targeted fine-tuning on our specific dataset, we aim to achieve higher accuracy and efficiency.

References

[1] Kayee Hanaoka, Patrick J. Grother, Mei Lee Ngan, Joyce Yang, George W. Quinn, and Austin Hom. Face analysis technology evaluation: Age estimation and verification, 2024-05-29 04:05:00 2024. 1

[2] Fazle Rabbi. Facial age dataset. https://www.kaggle.com/datasets/frabbisw/facial-age. 1

[3] Eran Eidinger, Roee Enbar, and Tal Hassner. Age and gender estimation of unfiltered faces. *IEEE Transactions on Information Forensics and Security*, 9:2170–2179, 2014. 1, 2

[4] Zhifei Zhang, Yang Song, and Hairong Qi. Age progression/regression by conditional adversarial autoencoder, 2017. 1

[5] Raghav Bansal, Gaurav Raj, and Tanupriya Choudhury. Blur image detection using laplacian operator and open-cv, 2016. 1

[6] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2015. 2

[7] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks, 2018. 2

[8] Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks, 2020. 2

[9] Jason Ansel, Edward Yang, Horace He, Natalia Gimelshein, Animesh Jain, Michael Voznesensky, Bin Bao, Peter Bell, David Berard, Evgeni Burovski, Geeta Chauhan, Anjali Chourdia, Will Constable, Alban Desmaison, Zachary DeVito, Elias Ellison, Will Feng, Jiong Gong, Michael Gschwind, Brian Hirsh, Sherlock Huang, Kshiteej Kalambarkar, Laurent Kirsch, Michael Lazos, Mario Lezcano, Yanbo Liang, Jason Liang, Yinghai Lu, CK Luk, Bert Maher, Yunjie Pan, Christian Puhrsch, Matthias Reso, Mark Saroufim, Marcos Yukio Siraichi, Helen Suk, Michael Suo, Phil Tillet, Eikan Wang, Xiaodong Wang, William Wen, Shunting Zhang, Xu Zhao, Keren Zhou, Richard Zou, Ajit Mathews, Gregory Chanan, Peng Wu, and Soumith Chintala. PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation, April 2024. 2

[10] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database, 2009. 2

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. 2

[12] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun. Shufflenet: An extremely efficient convolutional neural network for mobile devices, 2017. 2