LAB 7: Face-based age estimation

Gr.5

1. Introduction

The main goal of this project is to create a neural network model that can estimate a person's age just by looking at their face in an image. To do this, we will use well-known datasets and modern deep learning techniques. This document explains all the key assumptions, methods, technical needs, and evaluation strategies for the project.

2. Data Quality

1. The datasets chosen for training and testing have accurate age labels for each face image.

2. The datasets will be checked to remove any corrupted or incomplete images.

3. Dataset Diversity

1. The datasets cover a wide range of ages, ideally from 0 to 100 years, to help the model learn to estimate ages for people of all age groups.

2. Images include people of different ethnic backgrounds, genders, and face structures to prevent bias.

3. The data also has different lighting conditions, backgrounds, and image quality to make the model more robust.

4. If we find the datasets are imbalanced (like too many young faces and not enough old ones), we will use techniques to fix this, such as:

    - Sampling techniques to balance the data

    - Data augmentation to create more varied images

4. Datasets

To train and evaluate the model effectively, we have selected the following datasets:

4.1 Kaggle Faces: Age Detection from Images

https://www.kaggle.com/datasets/arashnic/faces-age-detection-dataset/data

This dataset contains **19,906 images** labeled in three broad categories:

- **YOUNG**: 34% of the dataset
- **MIDDLE**: 54% of the dataset
- **OLD**: 12% of the dataset

While it's good for learning the difference between young, middle-aged, and old faces, it is **not suitable for exact age prediction** since there are no exact ages—just age groups. We can still use it for testing general predictions but not for detailed age estimation.

4.2 UTKFace

https://paperswithcode.com/dataset/utkface

UTKFace is a big dataset with over **20,000 face images** that includes:

- **Age**, **Gender**, and **Ethnicity** labels
- Ages from **0 to 116 years**
- Different poses, expressions, lighting, occlusions, and resolutions

This dataset is really good for training the model to estimate specific ages because it has clear, exact age labels.

4.3 IMDB-WIKI – 500k+ Face Images with Age and Gender Labels

https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/

This is one of the largest datasets available, containing:

- **460,723 face images** from IMDb celebrities
- **62,328 images** from Wikipedia profiles
- **523,051 images** in total

Ages are calculated based on birth dates and when the photos were taken. While there might be some noise in the data, it is still very useful for training a model that can handle a wide variety of face types and ages.

5. Preprocessing

1. Standard face-detection and alignment techniques (like **MTCNN** or **dlib**) will be used to crop and standardize the face images.
2. Images will be resized to **at least 64×64 pixels** for good detail.
3. Preprocessing steps include:
    - Cropping and aligning faces
    - Normalizing pixel values (0 to 1 scaling)
    - Converting to grayscale
    - Standardizing the image data to help the model learn faster

6. Modeling Framework

1. The project will be implemented using **Python**:
    - **PyTorch** for building the deep learning model
    - **TensorFlow/Keras** if additional experiments are needed
    - **Scikit-Learn** for data preprocessing and evaluation

7. Proposed Models for Age Estimation

To accurately estimate age from faces, we will try different types of models:

7.1 Convolutional Neural Networks (CNN)

CNNs are great for analyzing images. Models like **ResNet**, **VGGNet**, and **Inception** can be fine-tuned to learn age-related features in faces.

7.2 Residual Networks (ResNet)

ResNet is a type of CNN that helps avoid problems when training deep models by using "shortcut" connections. This helps the model learn more details from the images.

7.3 ResNet-50 (Proposed Model)

We propose to use **ResNet-50** as the main architecture for age estimation in this project. ResNet-50 is a deeper version of ResNet with **50 layers**, which allows it to learn complex features and fine-grained details that are essential for accurate age estimation.
Why ResNet-50?

- It has proven to work well on large datasets with high accuracy.
- Its skip connections help avoid problems with vanishing gradients, making it more stable during training.
- Pretrained weights are available, allowing us to fine-tune the model instead of training from scratch, speeding up the process.
- It generalizes well on different facial features due to its depth and ability to capture complex patterns.

7.4 Wide Residual Networks (WRN)

WRNs are a variation of ResNet but with wider layers. This helps the model understand even more details in the images, which is useful for age prediction.

7.5 EfficientNet

EfficientNet is designed to be very efficient—achieving good results with fewer parameters. This is helpful if we want to train faster without losing accuracy.

7.6 Transformers for Vision (ViT)

Vision Transformers (ViT) are a new way of analyzing images. Instead of traditional layers, they use "attention mechanisms" to focus on important parts of the image, which could help with detailed age prediction.

8. Conclusion

The proposed model for this project is **ResNet-50**, due to its strong performance, efficient use of pretrained weights, and proven success in similar image-based tasks. This model, combined with the selected datasets and preprocessing steps, is expected to achieve high accuracy in age estimation tasks.