Project Name

Face-Matching System for Identifying Individuals at Different Ages

Project Overview

This project consists of two main components:

1. **Age Estimation**
2. **Face Matching**

---

1. Age Estimation

Dataset Choice and Reason

UTKFace Dataset

The UTKFace dataset is a large-scale face dataset containing images annotated with **age, gender, and ethnicity**. It is widely used for computer vision tasks such as **age estimation, face recognition, and demographic analysis**.

Key Features of UTKFace Dataset:

- **Diverse Age Range:** Covers ages from **0 to 116 years**, making it useful for age progression and regression tasks.
- **High Variability:** Includes variations in **pose, facial expression, and lighting conditions**.
- **Annotations:** Each image is labeled with:
  - **Age (numerical value)**
  - **Gender (0 for male, 1 for female)**
  - **Ethnicity (White, Black, Asian, Indian, and Others)**
- **Dataset Size:** Over **20,000 images** of different individuals.

Age Prediction Model Architecture

The **AgeEstimationModel** is a deep learning model designed for age prediction, supporting two backbone architectures: **ResNet-50** and **Vision Transformer (ViT)**.

(A) ResNet-50 Based Model

- Uses **ResNet-50**, a **convolutional neural network (CNN)** designed for feature extraction.
- The original **fully connected (fc) layer** is replaced with:
  - **Dropout layer (0.2 probability):** Helps prevent overfitting.

    - **Linear layer (2048 → 256 neurons):** Reduces feature dimensions.
    - **Final Linear layer (256 → output_nodes):** Maps extracted features to the predicted age.

Why Use ResNet-50?

- Strong feature extraction capabilities due to deep CNN layers.
- Pretrained models on large datasets (e.g., **ImageNet**) allow for effective transfer learning.
- Works well with **structured visual patterns** (e.g., facial features).

(B) Vision Transformer (ViT) Based Model

- Uses **ViT-Small Patch 14 (DINOv2 model)**, a self-attention-based transformer for image processing.
- The final classification head is modified with:
    - **Dropout layer (0.2 probability)**
    - **Linear layer (384 → 256 neurons)**
    - **ReLU activation**
    - **Final Linear layer (256 → output_nodes) for age prediction**

Why Use ViT?

- Captures **long-range dependencies** in images better than CNNs.
- Handles **occlusions and complex textures** more effectively.
- Self-attention mechanisms provide improved feature representation.

Loss Function Selection and Reason

The model uses **L1 Loss (Mean Absolute Error - MAE)**, which is well-suited for age estimation as a regression problem.

Why L1 Loss?

- **Minimizes Absolute Differences**: Computes the absolute difference between predicted and actual age values.
- **Robust to Outliers**: Unlike **Mean Squared Error (MSE)**, which squares errors, L1 Loss treats all errors equally.
- **Stable Training**: Provides smooth and stable gradients, helping prevent overfitting.

Performance Analysis and Evaluation Metrics

1. Mean Absolute Error (MAE)

- **Directly measures** the average absolute difference between predicted and actual ages.
- **More robust to outliers** than MSE.
- **Easy to interpret** (lower MAE = better accuracy).

2. Training and Validation Loss Tracking

- The model **saves and plots** loss values for both training and validation.
- **Best Model Selection:** The model saves the weights with the lowest validation loss.

2. Face Matching

Dataset Choice and Reason

Labeled Faces in the Wild (LFW) Dataset

The **LFW dataset** is one of the most widely used benchmark datasets for **face recognition and verification**. It contains **real-world images collected from the internet** with labeled pairs for evaluating face-matching algorithms.

Key Features of LFW Dataset:

- **Number of Images: 13,233 images** of **5,749 individuals**.
- **Source:** Images collected from online sources, ensuring **natural variations** in lighting, background, and pose.
- **Annotations:** Labeled pairs for **face verification tasks**.
- **Unconstrained Conditions:** Unlike controlled datasets, LFW images exhibit:
  - **Pose variations**
  - **Facial expressions**
  - **Illumination changes**
  - **Background clutter**
- **Pairwise Evaluation:** Designed for **face verification**, where models compare two images and determine if they belong to the same person.

Face Matching Model Architecture

The model is implemented in FaceMatch.py and consists of two main components:

(A) Backbone Network (Feature Extractor)

- A **CNN-based feature extractor** that processes face images into **128-dimensional embeddings**.
- Uses **six convolutional layers** followed by **fully connected layers**.
- Outputs a **128-dimensional feature vector**.
- Uses **ReLU activation, max pooling, and dropout** for regularization.

(B) Face Embedding Model (FaceEmb)

- Takes **three images** as input:
  - **Anchor (A)** → Face of the target person.
  - **Positive (P)** → Another image of the same person.
  - **Negative (N)** → An image of a different person.
- This structure is used for **triplet loss training** to improve face-matching accuracy.

Loss Function Selection and Reason

The model uses **Triplet Margin Loss**:

```
criterion = torch.nn.TripletMarginLoss(margin=1.0, p=2)
```

Why Triplet Margin Loss?

- Used in **face verification models** like **FaceNet**.
- **Minimizes intra-class distance** (A ↔ P) while **maximizing inter-class distance** (A ↔ N).
- Helps the model learn **discriminative features** for face-matching.

Performance Analysis and Evaluation Metrics

1. Training and Validation Monitoring

- The model logs **training loss** after every few batches.
- **Best Model Selection:** Saves the model with the lowest validation loss.

2. Possible Additional Metrics for Future Improvement

- **Rank-1 Accuracy**: Measures how often the correct face appears as the top match.
- **False Acceptance Rate (FAR) & False Rejection Rate (FRR)**: Measures the reliability of face matching.
- **t-SNE Visualization**: Helps in **embedding space analysis** to ensure distinct clustering of different identities.

This structured approach ensures a robust system for **age estimation and face matching across different ages**.