CYSHIELD

Age Invariant Face Recognition Report

Introduction

This project aimed to solve a key challenge in face recognition: accurately identifying people as they age. This is important for things like long-term surveillance and finding missing people. The main issue is that a person's face changes a lot over their lifetime, which makes traditional recognition systems unreliable. The project explored two different approaches to create a system that could recognize faces regardless of age.

Approach 1: Multi-task Learning

The first approach was inspiration from paper: “WhenAge-Invariant Face Recognition Meets Face Age Synthesis: AMulti-Task Learning Framework” involved building a single, complex **multi-task deep learning model**. This model was designed to do two things at once:

- **Recognize a person's identity.**
- **Predict their age.**

The idea was that learning to predict age would help the system create more robust facial features for recognition. However, this method proved to be very demanding. The model took a long time to train, and debugging the combined process was difficult with limited computational resources.

Approach 2: Separate Models

Due to the challenges of the first approach, I pivoted to a second strategy: creating **two separate, specialized models**.

- **An age prediction model.**
- **A face recognition model.**

This method was much more practical. It simplified the architecture, made it easier to train and debug each model independently, and required less computational power. Both models were trained on a smaller portion of the dataset to ensure the project could be completed within the given timeframe.

Dataset

The selection of an appropriate dataset is critical for developing an effective age-invariant face recognition system. The dataset must provide sufficient age diversity, multiple images per individual across different time periods, and adequate sample size for robust model training. The following table presents a comprehensive comparison of available datasets commonly used in age-invariant face recognition research.

Dataset	Details	Limitations
FERET	14,126 images of 1,199 subjects, minimal age variation.	Limited age differences, not ideal for (AIFR).
FG-NET	1,002 images from 82 subjects, used for facial aging research.	Few subjects, limits its generalization for large scale age-based recognition.
MORPH	Large dataset with 79,897 images of 21,194 subjects.	Still lacks enough age diversity for certain age groups.

CACD	163,446 images from 2,000 celebrities.	Celebrity-focused, introduces bias for positions, illuminations and lightings.
PCSO-LS	1.5 million images of 18,007 criminals, with images taken over 5 years.	Excludes ages 0-15 and focuses on criminal profiles.
WIT	1,109 images from 110 subjects, sourced from the internet.	Small size, potential variability due to internet sourced images.
FACES	1,026 images of 171 subjects with 6 facial expressions.	Fixed lighting and frontal views limit natural pose and lighting variations.
ADIENCE	26,580 images from 2,284 subjects with variation in posture, lighting, and noise.	Wild images limit controlled condition testing, affecting recognition accuracy.
In-house Dataset	200 images from various subjects with limited environmental diversity.	Small sample size and limited age variation affect generalizability

The initial plan for this project was to use the **MORPH (Morph Album-2) dataset** for training the face recognition system. MORPH was chosen because of its large size, age diversity, and longitudinal data, which are all crucial for developing a robust age-invariant model. It is also a well-established benchmark in academic research, allowing for easy comparison with existing methods. However, a key challenge arose: the MORPH dataset was **not publicly accessible** and required a licensing agreement that couldn't be secured within the project's timeframe.

Due to this accessibility issue, the project pivoted to the **Cross-Age Celebrity Dataset (CACD)** as an alternative. Although this dataset is publicly available and large-scale, it was still **too big for the project's computational resources**. Consequently, the team decided to train their models using only

top 200 identifier in age variance of the full CACD dataset to manage resource constraints and complete the project on time.

Dataset Preprocessing

Face Detection and Alignment:

MTCNN-based Face Processing:
The preprocessing pipeline employed the Multi-task Convolutional Neural Network (MTCNN) for robust face detection and alignment

Key Parameters and Rationale:

- Image Size (224x224)
- Margin (20 pixels)

Applied Data Augmentation:

- Random horizontal flip
- Color Jitter
- Normalization

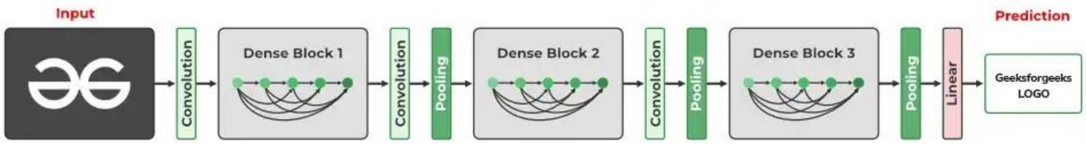
Data splitting:

For Age model: The splitting ratio is 80%:20%

For Face recognition: The splitting was by identifiers: 160 id for train, 20 id for validation, 20 for testing

Age Prediction model

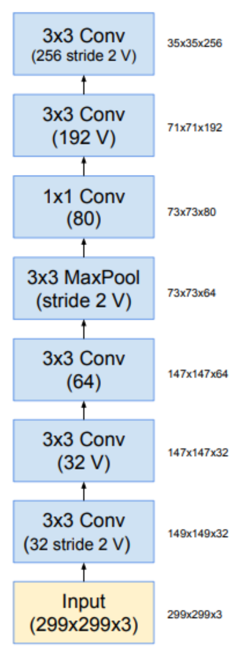
The age prediction model utilized **DenseNet121** as the backbone architecture, selected after initial experimentation with ResNet50. DenseNet121 was chosen due to its superior feature reuse capabilities and gradient flow properties.



This model uses a **dense connectivity pattern**, where each layer receives features from all previous layers. This method improves feature reuse and helps prevent the vanishing gradient problem. The model's architecture is defined by a **growth rate** of 32, which determines the number of new feature maps added at each layer. **Transition layers** are used between dense blocks to reduce feature map dimensions. For age prediction, the model first extracts **1024-dimensional features**. These features are then fed into a **single linear regression layer** that outputs a continuous age value.

Face Recognition model

The face recognition system employs a sophisticated architecture combining **InceptionResnetV1** as the feature extraction backbone with **ArcFace loss** for discriminative embedding learning. This combination leverages state-of-the-art deep metric learning techniques specifically designed for face verification and identification tasks. Outputs 512-dimensional feature embeddings.



Architecture Advantages:

1. **Multi-scale Feature Learning**: Inception modules capture both fine-grained details (facial texture) and global structure (face geometry)

2. **Domain-Specific Pre-training**: VGGFace2 initialization provides face-optimized feature representations
3. **Computational Efficiency**: Balanced architecture providing strong performance without excessive computational overhead

Loss Function

Age Prediction Loss

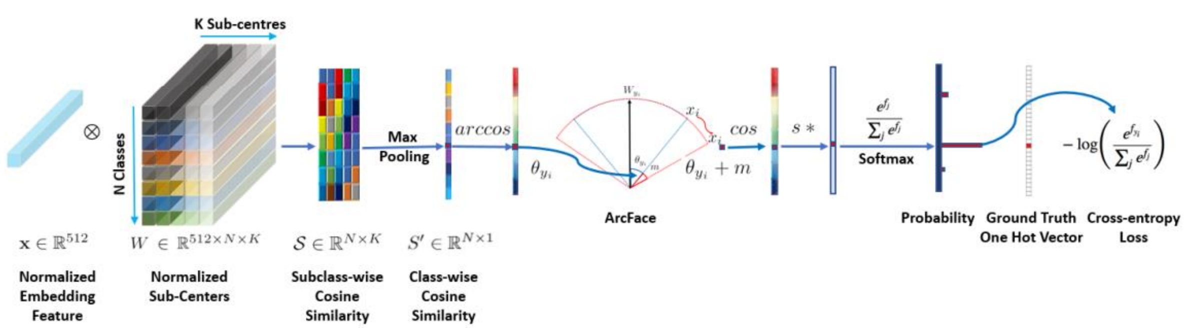
Age prediction loss is Mean Absolute Error (MAE)

Advantages:

1. **Robust to Outliers**: MAE is less sensitive to age labeling errors compared to Mean Squared Error (MSE), which is crucial given that celebrity age information may sometimes be inaccurate or estimated
2. **Interpretable Units**: Loss values directly correspond to years of age error, providing intuitive understanding of model performance
3. **Linear Penalty**: Provides consistent gradient magnitude across different error ranges, avoiding the quadratic penalty of MSE that can dominate training with large errors

Face Matching Loss

ArcFace Loss (Additive Angular Margin Loss)



Advantages:

1. **Angular Margin Enforcement**: Unlike traditional softmax loss, ArcFace adds an angular margin (m=0.50) that explicitly increases the decision boundary between different identities, leading to more discriminative embeddings
2. **Geometrically Interpretable**: Operates in angular space where face similarity is naturally measured, making the learned representations more meaningful for face matching tasks
3. **Scale Factor Control**: The scale parameter (s=30.0) controls the sharpness of the decision boundary, providing better gradient flow during training compared to standard softmax
4. **State-of-the-Art Performance**: ArcFace has demonstrated superior performance compared to alternative metric learning losses in face recognition benchmarks

Comparison with Alternative Losses:

Traditional Softmax Loss:

- **Limitation**: Only ensures separability, doesn't explicitly maximize inter-class margins
- **Problem**: Can lead to embeddings that are separable but not necessarily discriminative for similarity matching

Triplet Loss:

- **Limitation**: Requires careful triplet mining and can suffer from slow convergence
- **Comparison**: ArcFace provides more stable training without complex sampling strategies

Performance Analysis

Age Model Results:

Model	Training MAE	Validation MAE
Resnet50	3.82	5.09
DenseNet121	**3.44**	**4.87**
DenseNet169	3.86	4.98

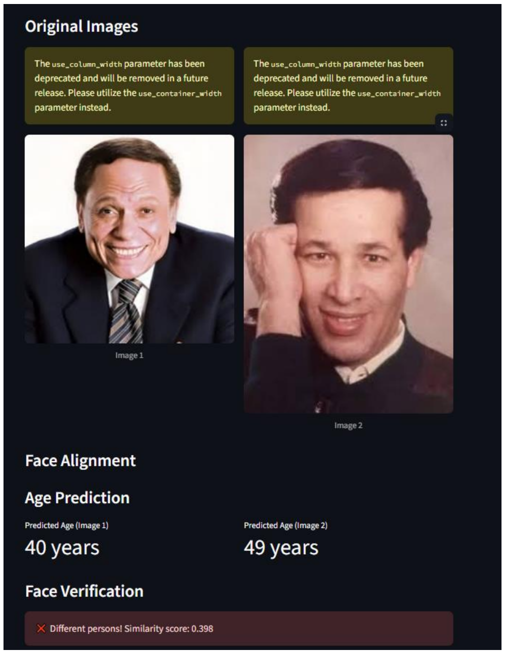
Face Matching Model Results:

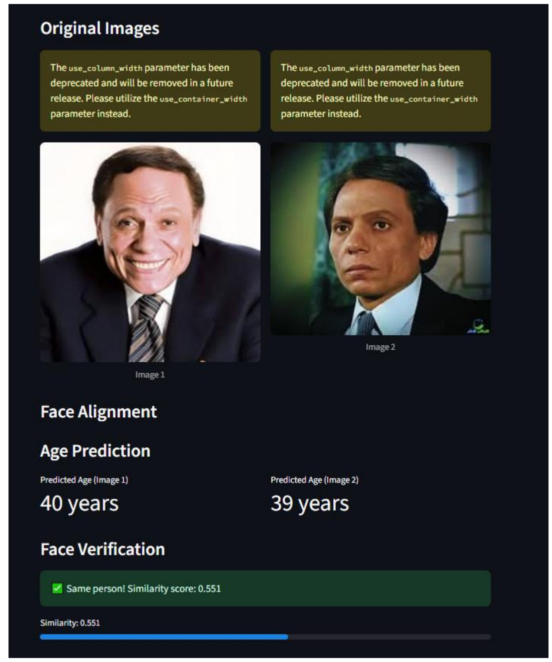
The face recognition model demonstrated strong discriminative performance on the validation set, achieving an **Area Under the Curve (AUC) of 90.54%** for the face verification task.

Performance Interpretation:

Excellent Discriminative Capability: An AUC of 90.54% indicates that the InceptionResnetV1 + ArcFace combination successfully learned discriminative facial embeddings. This score means that in 90.54% of cases, the model assigns higher similarity scores to genuine positive pairs (same identity) compared to negative pairs (different identities).

Deployment Using Streamlit results:





Project Summary

This project developed an age-invariant face recognition system with a two-stage approach, adapting to significant computational constraints. The system used separate models for age prediction and face recognition, achieving strong results despite limited resources.

Key Achievements:

- **Age Prediction:** A DenseNet121 model achieved a mean absolute error (MAE) of 4.87 years.
- **Face Recognition:** An InceptionResnetV1 model with ArcFace loss achieved an Area Under the Curve (AUC) of 90.54%.
- **Methodology:** The research successfully shifted from a complex multi-task model to separate, specialized models to stay within resource limits, proving that a pragmatic approach can still yield functional results.

- **Technical Contributions:** Developed a robust preprocessing pipeline and validated the effectiveness of specific architectures and loss functions for this problem.

Limitations and Future Work

The main limitation was training on only 10% of the dataset due to computational constraints, which are likely limited performance. The use of separate models also prevented the exploration of potential synergies between age and identity tasks.

Future work includes:

- **Immediate Enhancements:** Retraining the models on the full dataset with better computational resources to improve performance and conducting a demographic analysis to address potential biases.
- **Advanced Research:** Reviving the multi-task learning approach with more efficient architectures and exploring new techniques like age-conditioned embeddings and temporal face synthesis.
- **Long-Term Vision:** Applying technology to real-world applications such as security systems, missing person identification, and healthcare.