System Architecture Details & Analysis

Age Prediction & Face Matching Project

1. System Architecture Details

Age Detection Model (EfficientNet-B4):

The core of the system is the **EfficientNet-B4** architecture, implemented via the timm library. It was chosen for its superior parameter efficiency compared to older ResNet models.

- **Input Size:** 224x224. This resolution is a standard trade-off, preserving enough facial detail (wrinkles, texture) without excessive computational cost.
- **Dropout:** A dropout rate of 0.3 is applied after the feature extraction layer to prevent overfitting.

Multi-Task Heads:

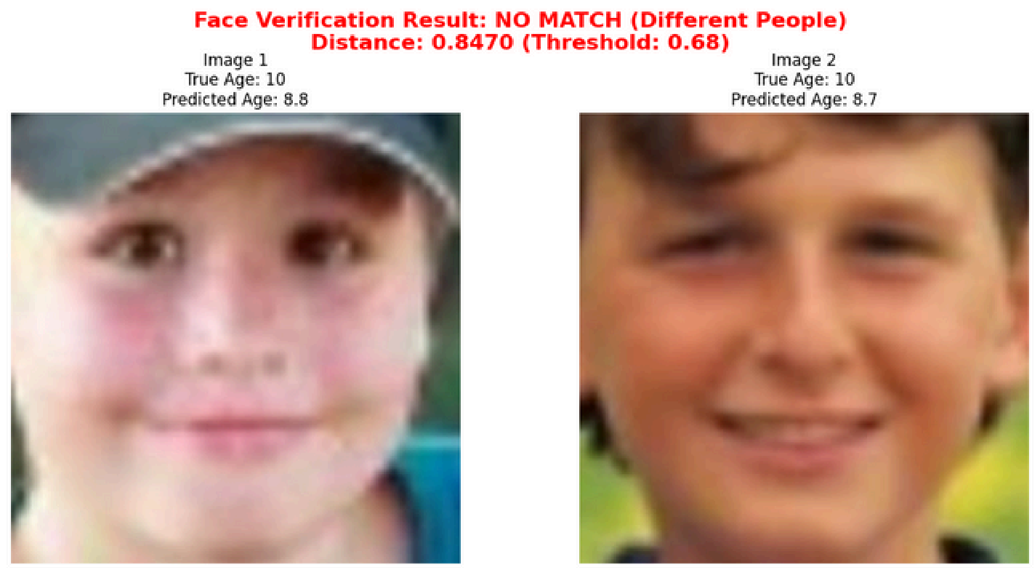
- **Regression Head:** A dense layer (256 units) followed by a single neuron outputs a continuous scalar age.
- **Distribution Head (Soft Labels):** Captures uncertainty by predicting a probability distribution over 101 age classes (0-100).

2. Face Matching Algorithm

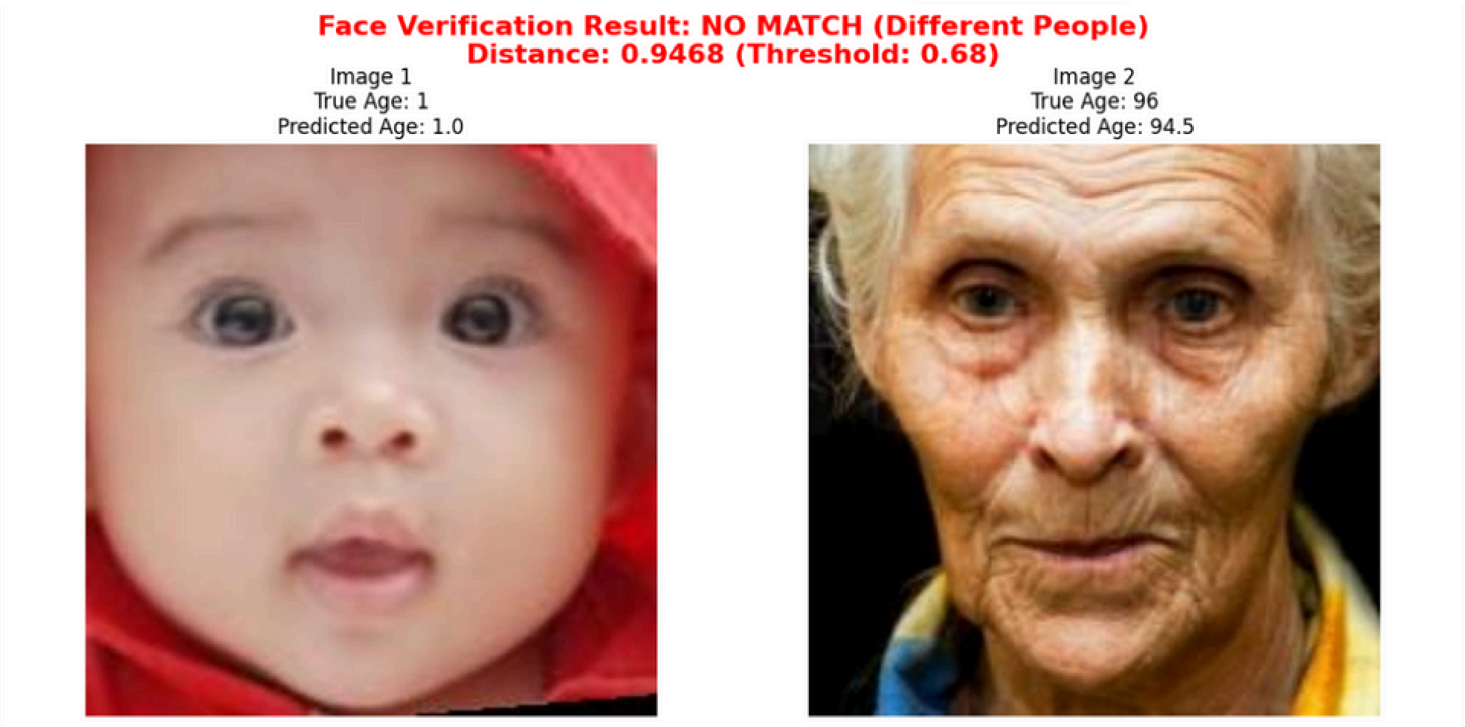
The system integrates the **DeepFace** library to handle identity verification:

- **Detection (RetinaFace):** Used to locate and align faces with high precision, even in difficult poses.
- **Verification (ArcFace):** A state-of-the-art model that generates embeddings for face comparison. It computes the cosine similarity between two face embeddings to determine if they belong to the same person.

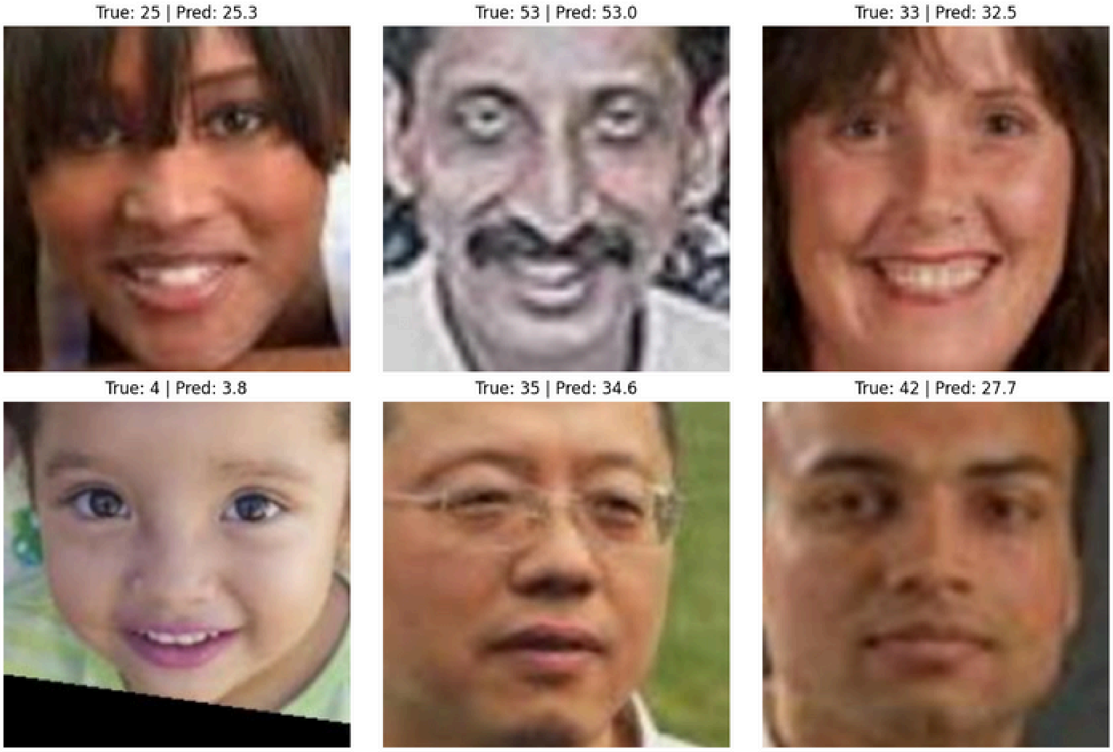
Face Matching Test Result:

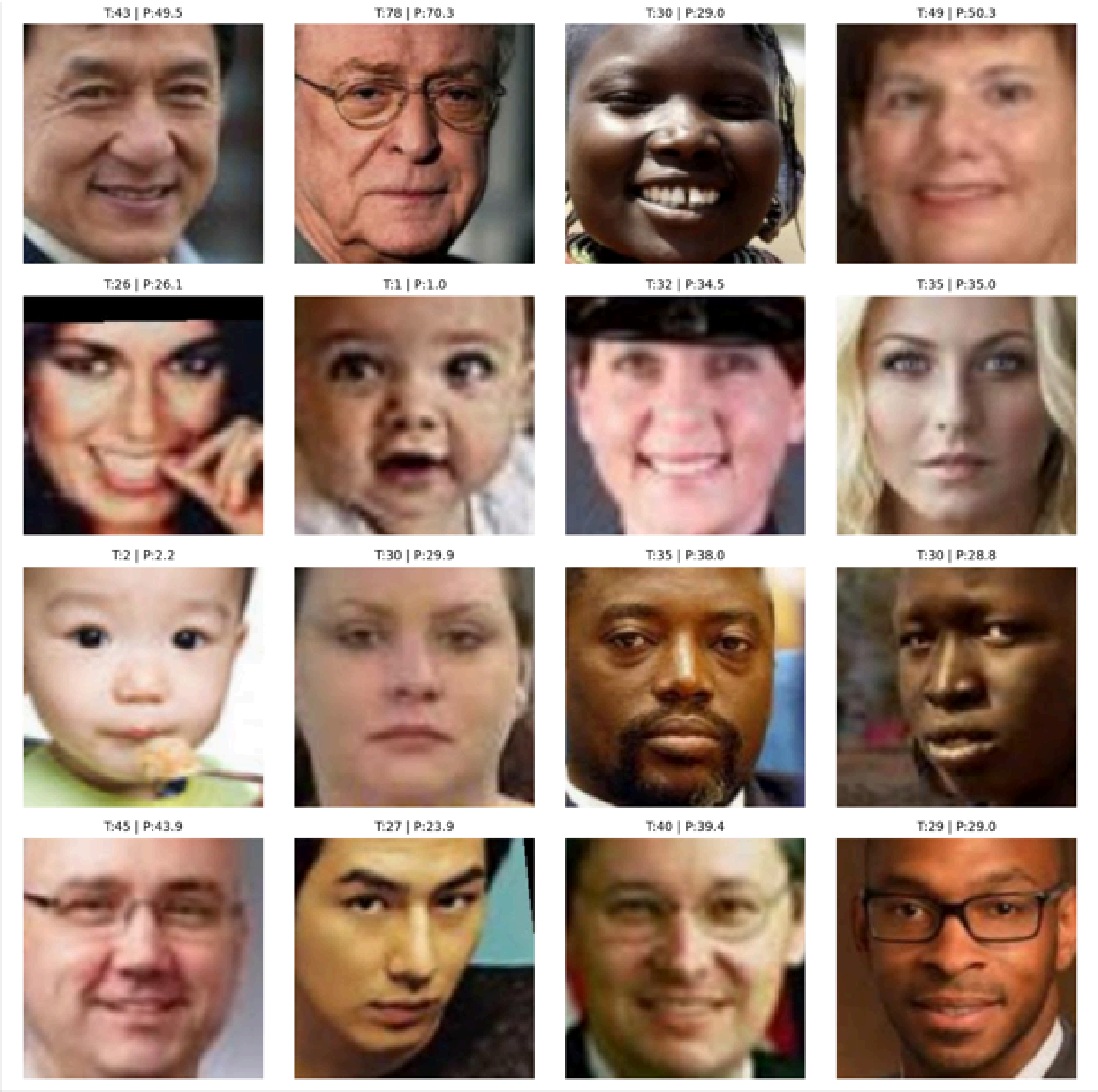



Face Verification Result: MATCH (Same Person)
Distance: 0.3682 (Threshold: 0.68)
Image 1
True Age: 18
Predicted Age: 19.1
Image 2
True Age: 16
Predicted Age: 16.8


Face Verification Result: NO MATCH (Different People)
Distance: 0.9468 (Threshold: 0.68)
Image 1
True Age: 1
Predicted Age: 1.0
Image 2
True Age: 96
Predicted Age: 94.5

Sample Predictions:





3. Previous Work & Implementation Decisions

Why Removed ImageNet Normalization?

I replaced standard ImageNet normalization (mean/std) with simple 0-1 scaling. Since age estimation relies heavily on texture and lighting nuances, standard normalization sometimes skewed the input distribution unnecessarily. Simple scaling is more interpretable and sufficient for this domain.

Why Removed Explicit Face Detection in Training?

The training pipeline initially included MTCNN for face detection. This was removed because the UTKFace dataset consists of already cropped/aligned faces. Re-running detection was redundant, computationally expensive, and occasionally failed on difficult samples, interrupting the training loop.

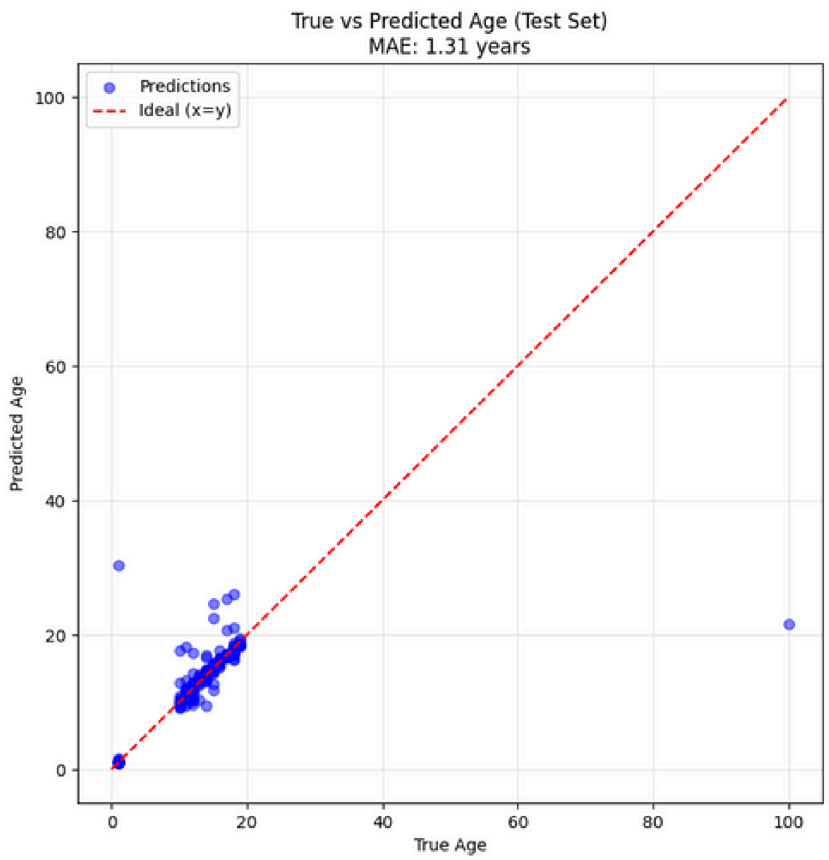
Augmentation Strategy:

employed specific augmentations via Albumentations to mimic real-world variations:

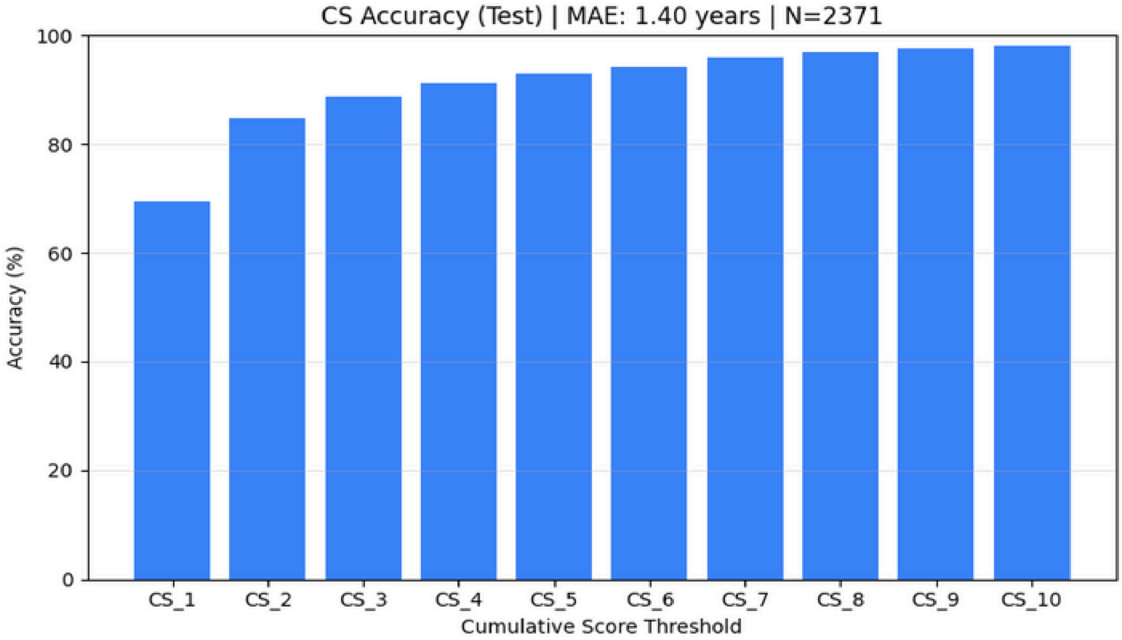
- **Horizontal Flip:** Standard data multiplication.
- **Shift/Scale/Rotate:** Handles misalignment in real-world inputs.
- **Random Brightness/Contrast:** Accounts for varying lighting conditions.
- **Gaussian Blur:** Simulates low-quality camera inputs.
- **Coarse Dropout:** Simulates occlusions (e.g., masks, glasses, hair) to force the model to look at the whole face.

4. Performance Analysis

True vs Predicted Age (Test Set):



Accuracy Measurements:



Test MAE: 1.3994 years

Threshold	Accuracy (%)
CS_1	69.34
CS_2	84.65
CS_3	88.70
CS_4	91.14
CS_5	92.87
CS_6	94.14
CS_7	95.78
CS_8	96.88
CS_9	97.60
CS_10	98.19

Validation Metric:

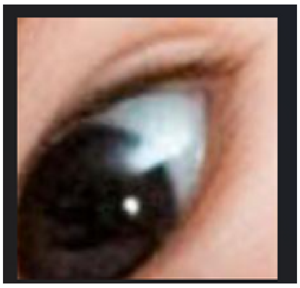
Best Validation MAE: 4.4395 years

5. Dataset Limitations & Issues

Several critical issues were identified in the UTKFace and FG-NET datasets that hinder optimal performance:

Noisy & Poor Quality Images:

- **Occlusions/Artifacts:** Some images contain random noise or confusing features, such as a baby with a mustache drawn on, or images where only a single eye is visible.



- **Blurry Images:** A significant number of images are blurred, losing high-frequency texture details (wrinkles) essential for age estimation.

Outliers & Labeling Errors:

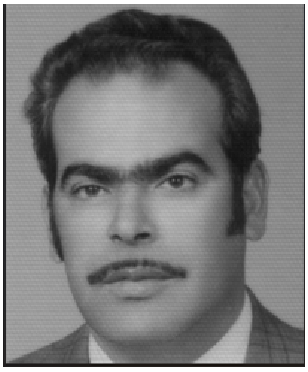
- **The '116-Year-Old' Outlier:** There is a sample labeled 116 years old where the subject looks very old, but due to image quality/smoothing, the wrinkles have disappeared. The model detects 'old' but cannot regress to such an extreme outlier without texture cues.

- **Wrong Labels:** We found instances of elderly subjects

mislabeled as
177 years old, which confuses the regression loss.

Color vs. Grayscale Bias (FG-NET):

- In the FG-NET dataset, we observed inconsistencies where the same individual is present in both RGB and Grayscale formats but labeled with different ages (e.g., a 10-year gap).
- The dataset often labels grayscale images with younger ages (or vice versa), creating a spurious correlation where 'color' becomes a predictor for age rather than facial features.



6. Future Work & Proposed Solutions

To address the limitations above, we propose the following roadmap:

1. Standardization to Grayscale:

Convert all training images to grayscale. Age estimation should rely on wrinkles, skin texture, and geometry, not color. This eliminates the bias where color quality (sepia/black-and-white) implies a certain era or age.

2. Image Sharpening:

Apply sharpening filters (e.g., Unsharp Masking) to blurry images during preprocessing. This enhances edge visibility and brings out texture details like wrinkles that are currently lost.

3. Dataset Cleaning via Face Matching

Since UTKFace is supposed to contain unique persons (or at least consistent identities), we will run a Face Matching pass (using our ArcFace module) on the entire dataset to cluster identities. This will help identify duplicates with conflicting age labels and clean the training set.

4. Manual Label Verification:

Implement a manual review loop (or semi-automated using high-confidence model disagreement) to filter out obviously wrong labels (e.g., 'Granny labeled 20') and remove noisy artifacts (mustache babies)