Adversarial Multi-Task Learning of Decorrelated Features for Age-Invariant Face Recognition

Raluca-Maria Rusu
Student

Lecturer, PhD. Diana-Laura Borza
Supervisor

Abstract—**The difficulty of matching faces with significant age differences in traditional Face Recognition (FR) systems has contributed to a rising interest in Age-Invariant Face Recognition (AIFR) research. Decoupling identity, age, and gender features ensures that the recognition system focuses purely on identity markers when verifying or identifying an individual, without interference from age or gender variations. This paper proposes a method in which face features are factorized by a Feature Residual Factorization Module (FRFM) into three components: identity, age, and gender. To ensure the identity features are effectively isolated from age and gender influences we use the Decorrelated Adversarial Learning (DAL) algorithm and Batch Canonical Correlation Mapping Module (BCCM) to maximize the correlation between features, and the Feature Residual Factorization Module (FRFM) and Backbone Network to minimize the correlation. This process enhances the model's ability to accurately recognize an individual's specific identity characteristics regardless of face feature differences, as demonstrated by achieving a 94.61% recognition accuracy on the FG-NET dataset. Identifying long-term missing people, confirming travelers' identities despite age changes, enabling historical and genealogical research, enhancing facial authentication, and assisting in medical diagnostics by monitoring age-related facial changes are just a few real-world applications in which our AIFR model can be used.**

1. Introduction

A key characteristic people use to identify one another is their faces. Face Recognition (FR) has been extensively researched, and with advances in computer hardware and imaging technologies, it has become increasingly common in daily life. Applications include passport verification, video surveillance [30], biometric authentication [31], and medical diagnostics. Despite FR's success, algorithms struggle to minimize age disparities as significant age differences can overpower identity features in cross-age face identification [1].

Age-Invariant Face Recognition (AIFR) is a sub-task of FR aiming to reduce the impact of age variance on recognition. Age-Invariant Face Verification (AIFV) and Age-Invariant Face Recognition (AIFR) are the two main categories of face recognition across aging. The inputs for the AIFR task include a set of facial images spanning different ages of various individuals. Each input image is accompanied by metadata indicating the individual's identity and possibly their age at the time the photo was taken and gender. The output depends on the specific category: for AIFV, the output is a binary decision indicating whether two images are of the same person given as a similarity percentage between the features; for AIFR, the output is a multi-class decision involving identifying or grouping all different-aged images of the same individual from a pool of faces. The core objective of both categories is to isolate discriminative features that are effective for both recognition and verification tasks.

AIFR research focuses on how aging affects facial features, which vary due to numerous factors. A person's work habits, diet, stress, health, smoking, substance abuse, and cosmetic surgery are all important environmental elements that contribute to a person's aging process. Furthermore, biological aspects should be taken into consideration as aging might manifest differently in different genders and ethnicities [33]. Constructing a system akin to the human vision system is very challenging. Humans can recognize hundreds of faces via lifelong learning and even recognize faces they haven't seen in a few years. We possess this aptitude to such a reasonable degree that it is rarely impaired by the passage of time or different visual changes brought on by aging, facial expressions, or factors like hair color or wearing glasses.

Figure 1. Images for illustrating face-by-age variations from the AgeDB dataset.

A further significant concern in this subtask is the alterations in the same person's photos as they age, or intra-subject variations, and the resemblance between the photos of different people, or inter-subject similarities. Because of this, many FR systems in use today struggle to recognize faces over large age differences. Age-Invariant Face Recognition has to consider the intra-subject variance, but also the inter-subject variance caused by age data.

Datasets are crucial for AIFR research as the face images that are utilized directly affect how accurate face recognition algorithms are [31]. The wide variances in the facial images of an individual and the similarities between those of different individuals are all critical challenges [32]. The ideal dataset has multiple face identities, many photos at various ages with small age gaps, and annotations for identity, age, and optionally gender or ethnicity. Many currently available datasets have issues such as imbalanced age-progressive images per person, large age gaps, and variations in illumination, pose, clarity, or facial expressions, complicating training and hindering optimal results.

Paper [2] details the methods usually used to tackle AIFR: Generative, Discriminative, and Deep Neural Network based methods. All techniques are used to train the models to learn specific traits that are constant with age. These techniques have evolved throughout the years in response to continuous efforts to develop AIFR systems that are more resilient, flexible, and effective. Generative methods focus on simulating the aging process of the face by generating images of a person at different ages to compare it to test images. The selection of parameters in a Generative model is frequently not successful in modeling the entire range of variations. Discriminative methods focus on identifying characteristics that are both age-invariant and sufficiently unique to distinguish between people. These are especially valuable since they prioritize stability over variability and directly address the problem of telling apart individuals within the same age group as well as identifying individuals at various ages. AIFR has been transformed by Deep Neural Network based methods, which make use of large-scale data processing power and the capacity to learn unique, hierarchical characteristics from data. These demonstrated a remarkable capacity to learn characteristics from the raw pixels that are both age and gender-invariant and age and gender-specific, while also deducing identity-specific features.

We propose a cohesive Deep Neural Network Multi-Task Learning Framework to learn feature invariant and identity-preserving characteristics. Specifically, we decompose high-level mixed features into three uncorrelated components—identity, age, and gender—through a Feature Residual Factorization Module (FRFM) and generate age group, gender, and identity classification tasks. We then calculate the correlation between these three components with a Batch Canonical Correlation Mapping Module (BCCM). The model employs a Decorrelated Adversarial Learning (DAL) method to maximize and minimize their correlation. Our model presents competitive performance for AIFR and is suitable for both face identification and verification purposes.

Our main contributions presented in this paper can be summarized as follows. Firstly, the Decorrelated Adversarial Learning (DAL) algorithm is based on the Feature Residual Factorization Module to regularize the learning of three decomposed elements from facial features: age, gender, and identity. Thus we obtain feature invariant and identity-preserving characteristics to improve recognition. To our knowledge, this is the first attempt to include decorrelated adversarial feature learning for three facial characteristics. Secondly, the Batch Canonical Correlation Mapping Module (BCCM) used to calculate the correlation between the three face features: age, gender, and identity. Lastly, the proposed method has significantly improved accuracy performance on the AIFR dataset FG-NET [17] which strongly demonstrates its effectiveness.

2. Related Work

Shallow Age-Invariant Face Learning, as described by [20], involves grouping image pairs with broad age gaps and applying State-Of-The-Art (SOTA) methods for face verification. This study assessed six deep face models across different metric learning techniques: Open-Face, VGG-Face, Face-Net, Deep-ID, DeepFace, and ArcFace. Three cross-age datasets were rearranged into positive and negative pairings with various age ranges. The research concluded that ArcFace and Face-Net perform better for shallow AIFR issues than other SOTA methods. Most Face Recognition algorithms degrade for shallow AIFR and overfit in terms of feature dimensions, making matching image pairs across wide age ranges a challenging problem.

Dataset	No. Images	No. Subjects	Age Labels	Age Range	Gender Labels
FG-NET [17]	$1,002$	82	Yes	0 to 69	No
AgeDB [18]	$16,488$	568	Yes	1 to 101	Yes
MORPH [34]	$55,134$	$13,617$	Yes	16 to 77	Yes
CACD [19]	$163,446$	$2,000$	Yes	16 to 62	Yes
VGGFace [7]	$3,310,000$	$9,131$	Yes	-	Yes

TABLE 1. COMPACT OVERVIEW OF THE CURRENTLY AVAILABLE AGING DATABASES.

Canonical Correlation Analysis (CCA) measures the linear relationship between two multidimensional variables. Paper [26] developed a multi-feature CCA method for face-sketch recognition. In [3] they proposed factorizing facial features into two uncorrelated parts, age-related and identity-related components, through a deep feature factorization framework. Then [1] enhanced the linear factorization module of [3] through an attendance-based feature decomposition in high-dimensional space, using GANs for training. Adversarial approaches, such as Generative Adversarial Networks (GANs), have shown effectiveness in tasks like face aging, super-resolution, and improving discriminative models. Paper [27] used GAN to generate high-resolution small faces for improved detection, [28] developed an adversarial UV completion framework (UV-GAN) for pose-invariant face recognition. Our work extends [3] deep feature factorization module to factorize the three uncorrelated parts: age-related, gender-related, and identity-related.

Prior AIFR studies have typically segregated facial attributes without considering their potential interdependencies [3]. For instance, the High-level Feature Attribution (HFA) [26] and the Latent Feature Convolutional Neural Network (LF-CNN) created to improve HFA, the Age Estimation guided CNN (AE-CNN) introduced in [24] and the Orthogonal Embedding CNN (OE-CNN), an orthogonal embedding decomposition encoding identity information in the angular space and age information in the radial direction proposed in [4]. These have shown limitations in addressing the underlying connections between facial features. Paper [3] proposed a Decorrelated Adversial Learning (DAL) method on the correlation between decoupled features of age and identity of a similar person. In response, our model adopts the DAL method to adversely maximize and minimize the mutual correlations between identity (x_{id}), age (x_{age}), and gender (x_{gender}) features to account for both the intra-subject variance and the inter-subject variance.

Age-Invariant Face Recognition has seen significant advancements, but there are still a number of major problems that make it difficult to create and implement efficient systems. These difficulties result from the intricacy of face aging on its own, as well as the technological constraints of recognition algorithms. Based on the premise that the decomposed components accurately reflect face information, our survey points to the importance of feature decomposition in invariant feature learning. Nonetheless, the identity characteristics could contain details like gender or age, and the dissected components could have latent relationships with one another, but correctly classifying the decomposed components and the identity-dependent component being rich in information is a step towards AIFR.

3. Method

This section offers a detailed overview of our methodology, outlining the specific approaches and procedures employed, as well as the reasons for choosing different techniques and their contribution to the reliability and validity of our results.

3.1. Backbone Network

In the initial stages of developing the Age-Invariant Face Recognition model, we used a Custom CNN Backbone inspired by ResNet, following [3] and [4] Orthogonal Embedding CNN (OE-CNN) with added Batch Normalization. This 64-layer CNN has 4 stages with 3, 4, 10, and 3 stacked Residual Blocks, and a final Fully Connected (FC) layer outputting 512-dimensional face features. A Residual Block consists of 3 units of 3x3 Batch Normalization, Convolution, and ReLU layers. The results we achieved using this as a Backbone were insignificant, as the dataset's size and diversity were limited, and the task it had to learn was immensely dependent on these resources.

Given these limitations, we transitioned to Transfer Learning. We conducted research for design and performance on different architectures and concluded that the best suited for our task is the **Inception-Resnet-v1** from [5], pre-trained by [6] on the VGGFace2 [7] dataset. This serves as our feature extractor, converting input images into embeddings or initial face features.

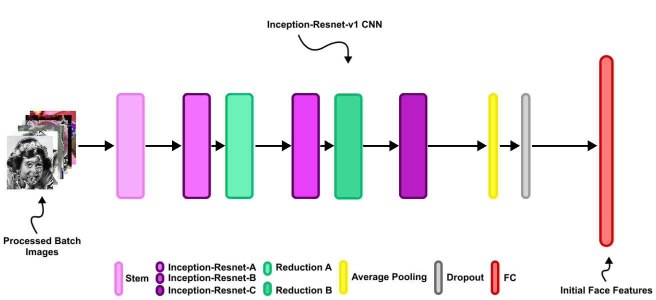

Figure 2. The Backbone Inception-Resnet-V1's architecture.

As described in [5], **Inception-Resnet-v1**, shown in Figure 2, combines Inception modules and residual connections. Inception modules capture information at various scales, while residual connections facilitate gradient flow in deep networks. The architecture introduced by [5] begins with a **Stem** block that performs initial convolutions to reduce dimensionality before further processing. This block combines convolutions of varying sizes to capture a broad spectrum of features from the input images, setting the stage for deeper analysis in subsequent layers. Following the **Stem**, [5] used several Inception blocks: **Inception-Resnet-A**, **Inception-Resnet-B**, and **Inception-Resnet-C**. Each of these blocks is designed to handle different aspects of the image. **Inception-Resnet-A** blocks focus on capturing smaller details with a combination of smaller convolutions. **Inception-Resnet-B** blocks apply larger convolutions to abstract higher-level features that are more spatially spread out. **Inception-Resnet-C** blocks, positioned deeper within the network, optimize the feature channels, enhancing the high-level features captured in previous blocks.

To reduce dimensionality and prepare the data for the next stages, [5] used two reduction blocks, **Reduction A** and **Reduction B**, strategically placed in the network. These blocks perform down-sampling, which helps in reducing computational complexity and focuses the network on the most salient features. Towards the end of the network, [5] append an **Average Pooling** layer, a **Dropout** layer, and a **Fully Connected (FC)** layer to the network.

These features enable **Inception-Resnet-v1** to effectively learn and generalize from diverse facial images, making it ideal for our Age-Invariant Face Recognition models.

3.2. Single-Task Model

To evaluate the improvement of our Multi-Task Model for AIFR over a simple FR model, we first developed a Baseline Single-Task Model that only classified identity. Feature extraction is the first stage in the Face Recognition process, where faces are compared against a collection. Similar to [3], we supervise identity learning using a margin loss. The **Large Margin Cosine Loss** or **CosFace**, introduced in [8], modifies the traditional softmax loss into a cosine loss by normalizing both features and weight vectors. This introduces a cosine margin to enhance the decision boundary between classes, achieving the lowest intra-class margin and maximum inter-class margin. The loss function is represented as:

$$\mathcal{L}_{ID} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s(\cos(\theta_{y_i,i})-m)}}{e^{s(\cos(\theta_{y_i,i})-m)}+\sum_{j\neq y_i}e^{s\cos(\theta_{j,i})}} \tag{1}$$

where N is the number of identities, y_i is the true class label for the i-th sample, $\theta_{y_i,i}$ is the angle between the feature vector of the i-th sample and its true class weight vector, $\theta_{j,i}$ is the angle between the feature vector of the i-th sample and any other class j weight vector, s is a scaling factor, and m denotes the margin separating identities by increasing inter-class variance and decreasing intra-class variance.

3.2.1. Identity Classification. For classification, we use a Softmax layer with Cross Entropy Loss. This setup maps the raw embeddings from the Backbone into a probability distribution over the target classes. Transforming embeddings to logits and then to normalized probabilities enables straightforward comparison against true class labels.

3.3. Multi-Task Model

In the pursuit of a robust AIFR system, we continued our study by developing a Multi-Task Model that simultaneously classifies identity, age, and gender from facial images. This model integrates a Feature Residual Factorization Module (FRFM) and three dedicated discriminators or classifiers, one for each attribute, facilitating a joint supervised learning approach to optimize feature extraction and enhance discriminative performance.

3.3.1. Feature Residual Factorization Module. Drawing inspiration from advanced factorization techniques in Deep Neural Networks [9], the Feature Residual Factorization Module (FRFM), inspired from [3], separates facial embeddings into three parts: age features (x_{age}), gender features (x_{gender}), and identity features (x_{id}). This disentanglement is shown in Figure 3 and is achieved through linear decomposition.

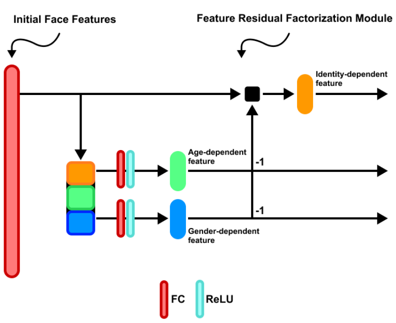

Figure 3. The Feature Residual Factorization Module (FRFM) architecture.

Given embeddings x extracted by the backbone CNN B from an input picture p, the linear decomposition is:

$$x = x_{\text{age}} + x_{\text{gender}} + x_{\text{id}}, \tag{2}$$

where x_{id} denotes identity features, x_{age} denotes age features, and x_{gender} denotes gender features. Age and gender features are encoded through the residual mapping function $x_{\text{age}} + x_{\text{gender}} = R(x)$ and $x = x_{\text{id}} + R(x)$. Using the FRFM, we obtain age and gender features, with the residual part being identity features:

$$x_{\text{age}} + x_{\text{gender}} = R(x), \qquad x_{\text{id}} = x - R(x) = x - x_{\text{age}} - x_{\text{gender}} \tag{3}$$

3.3.2. Identity Discriminator. The Identity Discriminator is a crucial part of our Multi-Task Model, designed to recognize and verify individual identities based on extracted facial features. This module uses the same architecture as our Single-Task model classification but uses as input only the residual identity features.

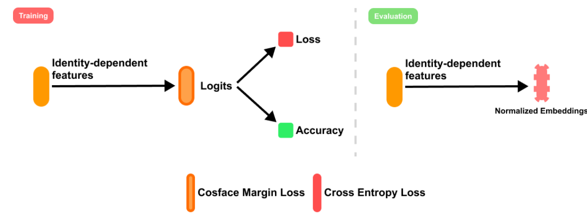

Figure 4. Identity Discriminator architecture.

3.3.3. Age Discriminator. To learn age-related information, x_{age} is passed into the Age Discriminator to learn age-specific information. Our data contains age labels, and we classify three age groups: $0 - 25$ (young), $25 - 55$ (adults), and $56+$ (elderly). We stack a Fully Connected Layer, Batch Normalization, ReLU activation, and another Fully Connected Layer over the x_{age} features to compute the logits. A Softmax layer with Cross-Entropy Loss is used for age group classification to calculate the loss and accuracy.

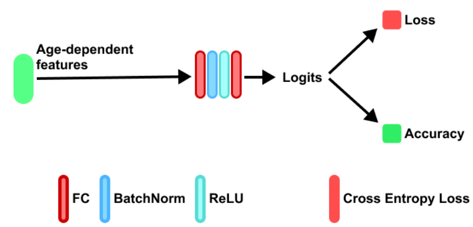

Figure 5. Age Discriminator architecture.

3.3.4. Gender Discriminator. To learn gender-related information, we feed x_{gender} into the Gender Discriminator. Our data has gender labels for two classes, $male$ and $female$. We stack a Fully Connected Layer, Batch Normalization, ReLU activation, and another Fully Connected Layer over the x_{gender} features to compute the logits. A Softmax layer with Cross-Entropy Loss is used for gender classification to calculate the loss and accuracy.

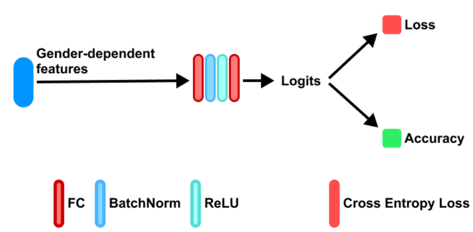

Figure 6. Gender Discriminator architecture.

3.4. Multi-Task + DAL Model

Facial attributes like identity, age, and gender are complex and intertwined. Traditional systems address these attributes independently, potentially missing intrinsic correlations. To manage these correlations, we use a Multi-Task

Model combined with Decorrelated Adversarial Learning (DAL), first introduced in [3], minimizing correlations between identity (x_{id}), age (x_{age}), and gender (x_{gender}) features.

The model's architecture, pictured in 7, facilitates simultaneous learning and refinement of each attribute's discriminative power while also enhancing the separation of these intertwined features. This dual-objective process is realized through an adversarial framework that decorrelates the features, ensuring that improvements in one attribute's accuracy do not detrimentally affect the others. This strategy not only improves the robustness of the model but also enhances its general applicability across diverse demographic groups.

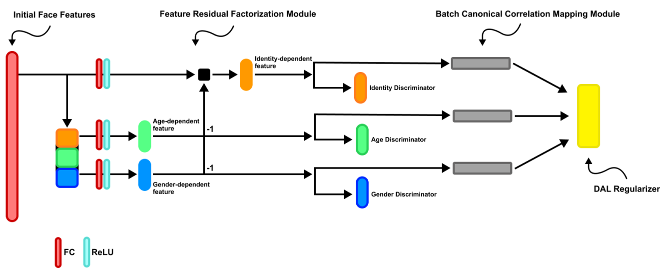

Figure 7. Overview on how the three factorized components x_{age}, $\mathrm{x}_{\text{gender}}$, and x_{id} are used for classification and DAL regularization.

3.4.1. Batch Canonical Correlation Mapping Module (BCCM). The BCCM, similar to [3], is integral to our Multi-Task Model. It quantifies correlations among identity, age, and gender features, enhancing the model's ability to discriminate and correlate these attributes effectively. Each feature set (identity, age, gender) is reduced to one dimension using a linear predictor, essential for calculating scalar correlations. During the forward pass, the method processes three inputs: identity features, age features, and gender features. These are transformed using their corresponding predictors, and predictions are used to calculate means and variances. Pairwise correlations are then computed using covariance divided by the product of standard deviations.

$$
\begin{aligned}
id\_age\_corr &= \frac{(\hat{y}_{age} - y_{age}) \cdot (\hat{y}_{id} - y_{id})}{\sqrt{var(y_{age}) \cdot var(y_{id})}} \\
id\_gender\_corr, &= \frac{(\hat{y}_{gender} - y_{gender}) \cdot (\hat{y}_{id} - y_{id})}{\sqrt{var(y_{gender}) \cdot var(y_{id})}} \\
age\_gender\_corr &= \frac{(\hat{y}_{age} - y_{age}) \cdot (\hat{y}_{gender} - y_{gender})}{\sqrt{var(y_{age}) \cdot var(y_{gender})}}
\end{aligned}
$$

where $\hat{y}_{age}$, $\hat{y}_{gender}$, and $\hat{y}_{id}$ are the predicted values for the age, gender, and id features, y_{age}, y_{gender} and y_{id} are the mean values of the age, gender, and id features, and $var(y_{age})$, $var(y_{gender})$, and $var(y_{id})$ are the variances of the age, gender, and id features.

Each correlation coefficient id_age_corr, id_gender_corr, and age_gender_corr is calculated as the covariance between two features divided by the product of their standard deviations, resulting in a measure of the linear relationship between the features. The overall correlation coefficient is obtained by averaging these three pairwise correlations, providing a scalar value that quantifies interdependence among identity, age, and gender features.

3.4.2. Decorrelated Adversarial Learning (DAL). The DAL approach fine-tunes the process of managing inter-subject and intra-subject correlations. This module dynamically adjusts to maximize or minimize the correlation between feature sets based on training objectives. Our model uses four key supervision modules: Identity Discriminator, Age Discriminator, Gender Discriminator, and the DAL Regularizer (Figure 7). The DAL Regularizer guides feature learning to maximize correlations calculated by the BCCM among x_{id}, x_{age}, and $\mathrm{x}_{\text{gender}}$. This adversarial learning technique ensures distinct and invariant feature extraction, enhancing generalization across tasks. We strategically manipulate gradients during training. In the minimization process, we train only the Backbone and FRFM parameters, freezing the BCCM parameters. In the maximization process, we freeze the Backbone and FRFM, train the BCCM parameters, and inverse its gradients for adversarial training, ensuring maximization of correlations among features.

Following [3], the model training is governed by a multi-task loss function designed to minimize classification errors and reduce correlation effects among features:

$$
\begin{aligned}
TL = L_{CE}(x_{\text{id}}) + \lambda_1 L_{CE}(x_{\text{age}}) + \lambda_2 L_{CE}(x_{\text{gender}}) \\
+ \lambda_3 L_{DALR}(x_{\text{id}}, x_{\text{age}}, x_{\text{gender}})
\end{aligned} \tag{4}
$$

where TL denotes the total loss, incorporating the Cross-Entropy Loss L_{CE} for identity, age, and gender features, alongside a specialized DAL Regularization Loss L_{DALR}, and λ_1, λ_2, and λ_3 are hyperparameters that help balance the impact of each loss component on the training process.

4. Experiments

4.1. Implementation Details

4.1.1. Datasets and Data Preprocessing. To achieve the best results, we required a large amount of high-quality data. We used the AgeDB dataset [18] and a small part of the CACD dataset [19] to create our datasets. We manually verified, added, and updated age and gender labels, ensured identity uniqueness, and checked image quality. This process resulted in two datasets for objective comparative tests.

Small Dataset: Contains 500 unique identities, each with over 20 images, spanning ages 0 to 101. It includes 6626 images of female subjects and 9126 images of male subjects, totaling $15,752$ images.

Large Dataset: Contains 1035 unique identities, each with over 5 images, spanning ages 0 to 101. It includes 8244 images of female subjects and 12143 images of male subjects, totaling $20,387$ images.

Proper data preprocessing is crucial for developing reliable face recognition models. These adjustments enhance the model's accuracy and adaptability by improving its ability to generalize from training data to new evaluation data. Figure 8 illustrates the transformations used to improve the data and simulate real-world variability.

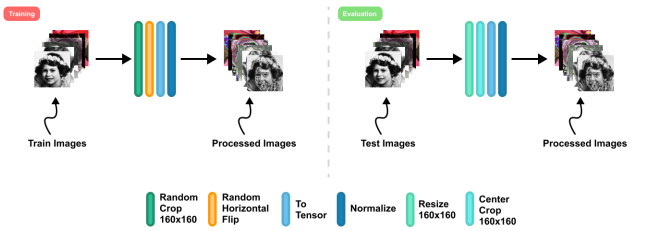

Figure 8. Data processing steps in both the training and evaluation phases.

4.1.2. Training. We trained the models on both small and large datasets. Training is initialized using a configuration singleton class that sets key parameters and paths for the model's training, evaluation, and saving procedures. **Model Name and Weights:** The models are named 'singletask', 'multitask', or 'multitask_dal', with pre-trained weights loaded if available to enhance training efficiency for complex datasets. **Embedding Size:** Set at 512, defining the dimensionality of facial embeddings for capturing distinct features. **Loss Function:** The Cosface Margin Loss is used to enhance feature discrimination. **Hyperparameters:** Learning rate, batch size, and number of epochs are set to 0.01, 64, and 40, respectively, to balance speed and accuracy. The model is instantiated through a handler that selects the appropriate model type based on the configuration. This model initializes a trainer class to manage the training process. **Optimizer:** An SGD optimizer with momentum 0.9 is used to adjust model weights, balancing quick convergence and stability. The Single-Task model optimizes the Backbone parameters, the Multi-Task model optimizes the Backbone and FRFM module, and the Multi-Task + DAL model alternates between focusing on direct task learning and adversarial decorrelation to develop a robust understanding of each attribute while maintaining their independence. For 40 iterations, the model minimizes the loss by optimizing the Backbone and FRFM parameters. For 30 iterations, it maximizes the loss by flipping the BCCM gradients and optimizing only its parameters. During each epoch, input data is processed, losses are computed, and weights are updated. **Loss Computation and Backpropagation:** for Single-Task model, identity loss only; for Multi-Task model, combined identity, age, and gender losses weighted by predefined lambdas (1, 0.9, 0.9); for Multi-Task + DAL model, combined identity, age, gender, and DALR losses weighted by predefined lambdas (1, 0.9, 0.9, 0.9). Training progress and key metrics are logged using **Wandb** for real-time insights and quick adjustments. At the end of each epoch, the model's state is saved, facilitating periodic evaluations and retaining the best-performing models.

4.1.3. Testing. We evaluated our models using the public AIFR face dataset FG-NET [17] and 20% of our own datasets (excluded from training). For face verification, we used cosine similarity of the concatenated identity features from the original and flipped images. For face identification, we measured the accuracy of matching each image to its correct label.

4.2. Models Comparative Analysis

4.2.1. 80/20 Split. The models were evaluated using an 80/20 training/testing data split. This standard method ensures a balanced and fair comparison of the models' performance under identical conditions.

Model	Trained on	80/20 Split
Single-Task	large	85.65%
Multi-Task	large	85.92%
Multi-Task + DAL	large	**86.24%**
Single-Task	small	93.02%
Multi-Task	small	93.12%
Multi-Task + DAL	small	**93.89%**

TABLE 2. ACCURACY RESULTS FOR SINGLE-TASK, MULTI-TASK, AND MULTI-TASK + DAL MODELS ON LARGE AND SMALL DATASETS UNDER THE 80/20 SPLIT PROTOCOL.

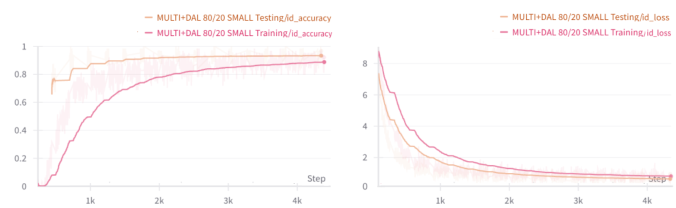

Figure 9. Wandb graphs of Multi-Task+DAL model training on the small dataset.

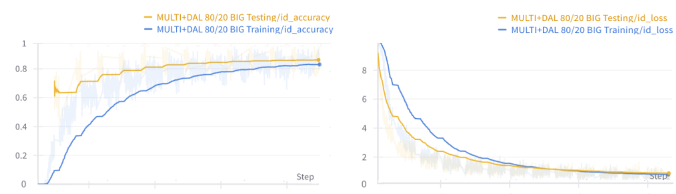

Figure 10. Wandb graphs of Multi-Task+DAL model training on the large dataset.

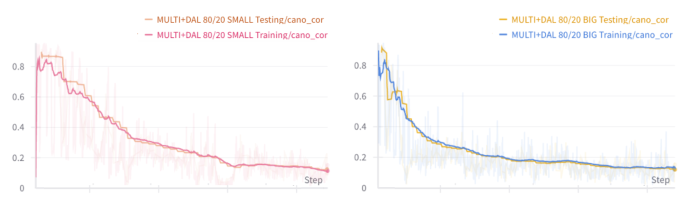

Figure 11. Wandb graphs of the canonical correlation minimizing and maximizing procedure while training on both datasets.

From the results presented in Table 2, several observations can be made. The Multi-Task and Multi-Task + DAL models consistently outperform the Single-Task model on both small and large datasets. This suggests that integrating multiple learning tasks enhances the model's ability to recognize facial features effectively. The Multi-Task +

DAL model shows a slight improvement over the standard Multi-Task model, particularly in the large dataset scenario. This improvement underscores the benefit of using DAL to minimize feature correlation, which potentially enhances the model's ability to generalize across more complex or diverse datasets. The highest accuracy is observed in the Multi-Task + DAL model trained on the small dataset, which may indicate that DAL's decorrelation capabilities are particularly effective when handling more nuanced data.

These findings suggest that the integration of multitasking learning frameworks, especially when combined with advanced techniques like DAL, can significantly improve the performance of facial recognition systems. The enhanced ability to handle complex interactions between different facial attributes in the Multi-Task + DAL model also suggests promising applications in fields requiring robust, scalable, and highly accurate facial recognition capabilities.

4.3. Experiments on the FG-NET Dataset

The FG-NET Aging Dataset from [17] consists of 1002 facial images representing 82 distinct individuals across various age groups. The age distribution, particularly prominent in the younger age provides a robust foundation for evaluating AIFR models.

4.3.1. Two-Pairs Evaluation. For this evaluation, we created 1000 positive and 1000 negative pairs of images, with equal gender distribution and a minimum age difference of 15 years between paired images. The model analyzes both the original and flipped versions of each image, combining the extracted features. Cosine similarity is then computed to measure the similarity between the feature sets. A threshold of 0.5 is used to determine if two images likely depict the same person. If the similarity score exceeds 0.5 for a positive pair or falls below it for a negative pair, it counts as a correct prediction. These results are aggregated to calculate the model's overall accuracy in distinguishing between positive and negative pairs.

Model	**Trained on**	**Evaluation two-pairs**
Single-Task	large	79.31%
Multi-Task	large	79.83%
Multi-Task + DAL	large	**80.00%**
Single-Task	small	78.85%
Multi-Task	small	79.42%
Multi-Task + DAL	small	**82.21%**

TABLE 3. Performance comparison of models on two-pair evaluation using the FG-NET dataset.

We tested the three models trained on both small and large datasets. As shown in Table 3, the Multi-Task + DAL model consistently outperformed the others, highlighting the benefits of combining multi-task and decorrelated adversarial learning for improving recognition accuracy across age-diverse facial data.

4.3.2. Leave-one-out Evaluation. In this rigorous testing protocol, each image in the FG-NET dataset was used once as a test image while the rest formed the training set. This process was repeated until all images had been used as the test image. Importantly, no images from FG-NET were employed for training or fine-tuning the models prior to this evaluation, ensuring that the test conditions were strictly unbiased and indicative of true model performance in unseen scenarios.

Method	**Rank-1**
Park et al. (2010) [10]	37.40%
Li et al. (2011) [11]	47.50%
HFA (2013) [12]	69.00%
MEFA (2015) [13]	76.20%
CAN (2017) [14]	86.50%
LFCNNs (2017) [15]	88.10%
AIM (2018) [16]	93.20%
Age + DAL (2019) [3]	94.50%
Multi-Task + DAL	**94.61%**

TABLE 4. Comparative performance of our Multi-Task + DAL model against other methods under the leave-one-out evaluation protocol on the FG-NET dataset.

Table 4 illustrates the results along with a comparative analysis with related works in AIFR. The Multi-Task + DAL model achieved notable improvements over previous methods, highlighting the advances enabled by our model's architecture and training approach.

5. Conclusion

This research proposed a cohesive Deep Neural Network Multi-Task Learning Framework for Age-Invariant Face Recognition (AIFR), demonstrating significant improvements by leveraging techniques such as Decorrelated Adversarial Learning and Multi-Task Learning based on multiple facial characteristics: age, gender, and identity. Our model decomposes high-level mixed features into three uncorrelated components through a Feature Residual Factorization Module and calculates the correlation between these components within a Batch Canonical Correlation Mapping Module. The results showed competitive performance for AIFR, making it suitable for both face identification and verification purposes. The proposed methodology has shown significant improvements in recognizing faces across varying age groups, thereby addressing a critical challenge in the domain of Face Recognition.

6. Future improvements

We identified several potential directions for further development to enhance the solution's resilience, robustness, and practical use in real-world situations. Future research on the Age-Invariant Face Recognition model may focus on optimizing hyperparameters and incorporating more diverse datasets to improve the accuracy of similarity scores. Integrating more biometric characteristics, such as ethnic features, could further enhance the system's accuracy and reliability. Additionally, experimenting with different loss functions may potentially enhance its discriminative capabilities. By focusing on these future improvements, this research can pave the way for the development of a more resilient and comprehensive Age-Invariant Face Recognition system.

References

[1] Zhang Junping Shan Hongming Huang, Zhizhong. When age-invariant face recognition meets face age synthesis: A multi-task learning framework and a new benchmark. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45:7917–7932, 2023.

[2] S. Patel S. Mittal. Age invariant face recognition techniques: A survey on the recent developments, challenges, and potential future directions. International Journal of Engineering Trends and Technology, 71:1–26, 2023.

[3] Zhifeng Li Wei Liu Hao Wang, Dihong Gong. Decorrelated adversarial learning for age-invariant face recognition, 2019.

[4] Y. Wang, D. Gong, Z. Zhou, X. Ji, H. Wang, Z. Li, W. Liu, and T. Zhang. Orthogonal Deep Features Decomposition for Age-Invariant Face Recognition. In European Conference on Computer Vision (ECCV), 2018. 2, 5, 7, 8

[5] V. Vanhoucke A. Alemi C. Szegedy, S. Ioffe. Inception-v4, inception-resnet and the impact of residual connections on learning, 2016.

[6] Tim Esler. Face Recognition Using Pytorch, 2017. v2.5.3 https://github.com/timesler/facenet-pytorch

[7] W. Xie O. M. Parkhi A. Zisserman Q. Cao, L. Shen. Vggface2: A dataset for recognising faces across pose and age. In International Conference on Automatic Face and Gesture Recognition, 2018.

[8] Z. Zhou X. Ji D. Gong J. Zhou Z. Li W. Liu H. Wang, Y. Wang. Cosface: Large margin cosine loss for deep face recognition. Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[9] C. Li X. Tang C. C. Loy K. Cao, Y. Rong. Pose-robust face recognition via deep residual equivariant mapping. IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 5187–5196, 2018.

[10] A. K. Jain U. Park, Y. Tong. Age-invariant face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2010.

[11] A. K. Jain Z. Li, U. Park. A discriminative model for age invariant face recognition. IEEE transactions on Information Forensics and Security (TIFS), 2011.

[12] D. Lin J. Liu X. Tang D. Gong, Z. Li. Hidden factor analysis for age invariant face recognition. International Conference on Computer Vision (ICCV), 2013.

[13] D. Tao J. Liu X. Li. D. Gong, Z. Li. A maximum entropy feature descriptor for age invariant face recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), page 5289–5297, 2015.

[14] M. Ye C. Xu, Q. Liu. Age invariant face recognition and retrieval by coupled auto-encoder networks. Neurocomputing, 2017.

[15] K. Luu M. Savvides C. N. Duong, K. G. Quach. Temporal non-volume preserving approach to facial age progression and age-invariant face recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[16] J. Zhao, Yu Cheng, Yi Cheng, Y. Yang, H. Lan, F. Zhao, L. Xiong, Y. Xu, J. Li, S. Pranata, S. Shen, J. Xing, H. Liu, S. Yan, J. Feng. Look Across Elapse: Disentangled Representation Learning and Photorealistic Cross-Age Face Synthesis for Age-Invariant Face Recognition, 2018.

[17] Lanitis, A., Taylor, C.J., Cootes, T.F. Toward automatic simulation of aging effects on face images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002.

[18] Papaioannou Athanasios Sagonas Christos Deng Jiankang Kotsia Irene Zafeiriou Stefanos Moschoglou, Stylianos. Agedb: The first manually collected, in-the-wild age database. pages 1997–2005, 2017.

[19] Chen Chu-Song Hsu Winston Chen, Bor-Chun. Face recognition and retrieval using cross-age reference coding with crossage celebrity dataset. IEEE Transactions on Multimedia, page 804–815, 2015.

[20] Islam, Khawar, Lee, Sujin, Han, Dongil, Moon, Hyeonjoon. 2021 36th International Conference on Image and Vision Computing New Zealand (IVCNZ), Face Recognition Using Shallow Age-Invariant Data, 2021, 1-6.

[21] Z. Li, U. Park, and A. K. Jain. A discriminative model for age invariant face recognition. IEEE transactions on Information Forensics and Security (TIFS), 2011. 2, 8

[22] D. Gong, Z. Li, D. Lin, J. Liu, and X. Tang. Hidden factor analysis for age invariant face recognition. In International Conference on Computer Vision (ICCV), 2013. 2, 4, 5, 7, 8

[23] Y. Wen, Z. Li, and Y. Qiao. Latent factor guided convolutional neural networks for age-invariant face recognition. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. 2, 4, 5, 7

[24] T. Zheng, W. Deng, and J. Hu. Age Estimation Guided Convolutional Neural Network for Age-Invariant Face Recognition. In IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017. 2, 7

[25] W. Yang, D. Yi, Z. Lei, J. Sang, and S. Z. Li. 2D-3D Face Matching using CCA.

[26] D. Gong, Z. Li, J. Liu, and Y. Qiao. Multi-feature Canonical Correlation Analysis for Face Photo-Sketch Image Retrieval. In Proceedings of ACM international conference on Multimedia, pages 617–620, 2013. 2

[27] Y. Bai, Y. Zhang, M. Ding, and B. Ghanem. Finding tiny faces in the wild with generative adversarial network. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR'18), 2018.

[28] J. Deng, S. Cheng, N. Xue, Y. Zhou, and S. Zafeiriou. Uvgan: Adversarial facial uv map completion for pose-invariant face recognition. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR'18), 2018.

[29] Y. Zhao, Z. Jin, G. jun Qi, H. Lu, and X. sheng Hua. An adversarial approach to hard triplet generation. In European Conference on Computer Vision (ECCV), 2018. 3

[30] G. Guo and N. Zhang. A survey on deep learning based face recognition. Computer Vision and Image Understanding, 189:102805, 2019.

[31] M. M. Sawant, K. M. Bhurchandi.Age Invariant Face Recognition: a Survey on Facial Aging Databases, Techniques and Effect of Aging. Artificial Intelligence Review. 10.1007/s10462-018-9661-z, 2019, 52, 1-28.

[32] Tong Yiying Jain Anil Park, Unsang. Facial aging databases, techniques and effects of aging: A survey. International Journal of Engineering Trends and Technology (IJETT), 67:8–13, 2019.

[33] Indiramma M. Nayak, J. S. An approach to enhance age invariant face recognition performance based on gender classification. Journal of King Saud University - Computer and Information Sciences, 34:5183–5191, 2022.

[34] Tesafaye T. Ricanek, Karl. Morph: A longitudinal image database of normal adult age-progression. pages 341 – 345, 2006.

[35] Krizhevsky, A., Sutskever, I., & Hinton, G.E. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60, pages 84 - 90, 2012.