Facial Aging Project

Finding Missing Children

Computer Vision Team

Internship 2021

NajahNow

Table Of CONTENTS

INTRODUCTION

Finding Missing Children

Face recognition is important because parents are more likely to have the picture of the lost child rather than the fingerprint or the iris which is why many researches have been conducted. The problem is that the more time passes between the person in the photo and the "probe" gets larger, the harder it is to recognize them.

There have been several approaches that were made, as explained in the following:

1. Discriminative approaches.
2. Generative approaches.

1. **Discriminative approaches** depended on two main things, one, is that identity and age can be separated. And two, that identity is suitable for face recognition. It discarded age-related information.

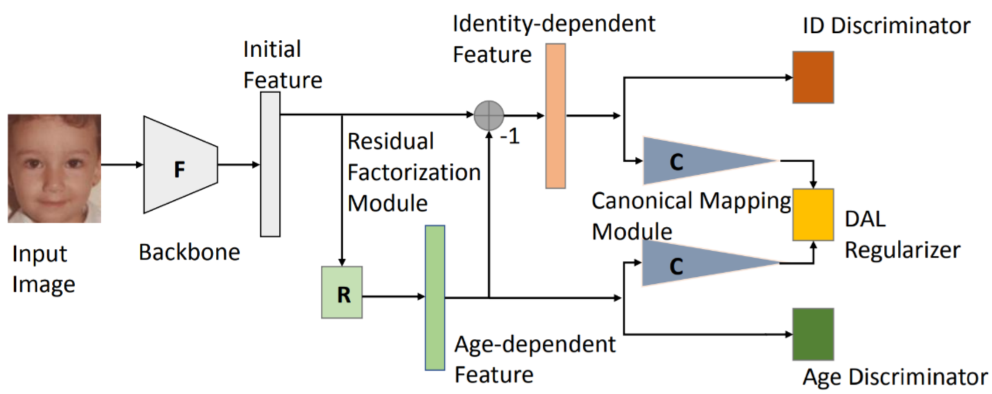
2. **Generative approaches** depended on Autoencoders and GANs. To synthesize the face to the desired age.

DISCRIMINATIVE APPROACHES

Decorrelated Adversarial Learning for Age-Invariant Face Recognition

A deep feature factorization learning framework is used to factorize the mixed face features into **two uncorrelated components**:

- Identity-dependent component (**xid**)
- Age-dependent component (**xage**)



Modules

- **F**: represents a backbone CNN that extracts an initial feature vector x from an input image. After extracting the initial feature vector, we define linear factorization as follows:

$$x = x_{id} + x_{age}$$

- **R**: represents a deep residual mapping module to extract the age-dependent component (x age) from the initial vector x, the we use x and x age to obtain x id as follows:

$$x_{id} = x - x_{age}$$

- **C**: maps xid, xage to the canonical variables v_{id}, v_{age}
- **DAL Regularizer**: calculates the canonical correlation between the paired features of the decomposed components.

Multi-Task Training Process

- **Age Discriminator:** Ensure age discriminating information
- **Identity Discriminator:** Ensure the identity-preserving information
- **DAL Regularizer:** The joint supervision to guide the feature learning
- **Canonical Correlation Maximizing Process:** freeze F, R and train C
- **Canonical Correlation Minimization Process:** freeze C and train F and R

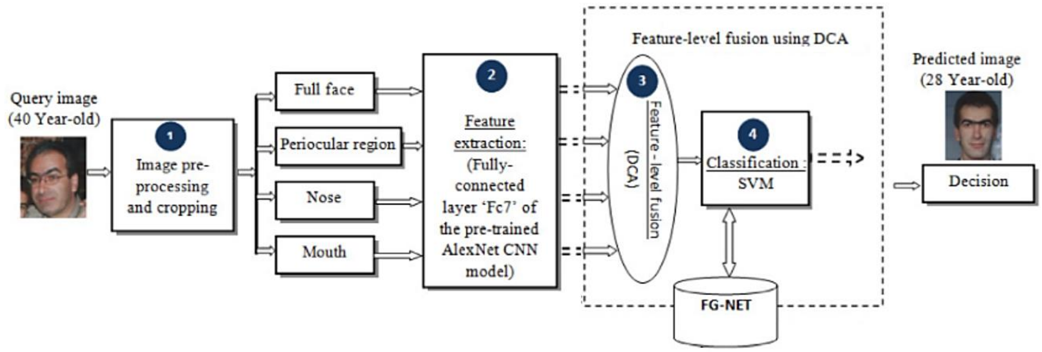
Loss Function

$$\mathcal{L} = \mathcal{L}_{ID}(\mathbf{x_{id}}) + \lambda_1 \mathcal{L}_{SM}(\mathbf{x_{age}}) + \lambda_2 \mathcal{L}_{DAL}(\mathbf{x_{id}}, \mathbf{x_{age}})$$

Component-Based Age-Invariant Face Recognition Using Discriminant Correlation Analysis

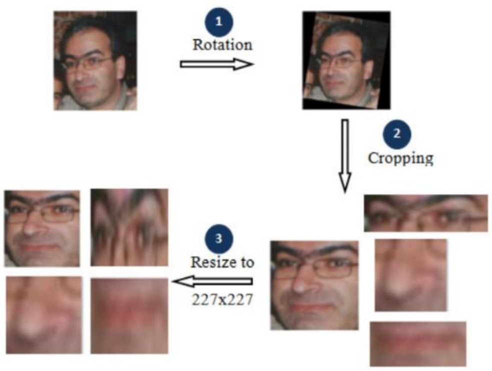
Face is an important characteristic for biometric identification and verification systems, as it holds the needed information about individual's identity. However, it is the characteristic most impacted by aging process, and facial aging is the main effect which decreases significantly the performance of face recognition algorithms. The main idea of this study relies on the fact that aging affects facial components such as mouth, eyes and nose differently.

Therefore, we suggest to consider face as an independent component set, and each facial component (eyes, mouth, and nose) will be processed separately, and we propose an effective component-based method for age-invariant face recognition using a **Discriminant Correlation Analysis** (DCA) as a feature-level fusion algorithm to combine Deep-based features computed from separated facial components, and a **Support Vector Machine** (SVM) as a classifier.

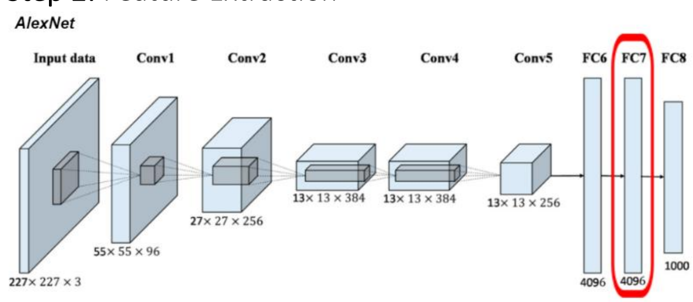


Proposed Approach Description

Step 1: Image Preprocessing

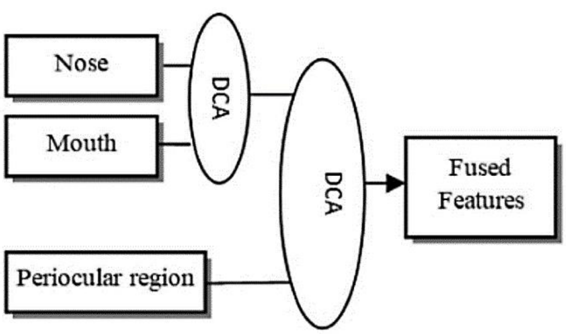


Step 2: Feature Extraction



Extract 4096-dimensional vector from FC7 for each facial component.

Step 3: Feature-Level Fusion



DCA performs an effective feature fusion by maximizing the pairwise correlations across the two feature sets.

Step 4: Classification

In the recognition stage, Support Vector Machine (SVM) is used as a classifier.

GENERATIVE APPROACHES

Generative Adversarial Networks (GANs)

The generator's primary goal is to generate data (such as images, video, audio, or text) from a randomly generated vector of numbers, called a **latent space**.

Face Aging can be very useful for both the entertainment and surveillance industries. It is particularly useful for face verification because it means that a company doesn't need to change their security systems as people get older. An age-cGAN network can generate images at different ages, which can then be used to train a robust model for face verification.

Important Concepts Related to GANs

We will first look at KL divergence. It is also very important to understand JS divergence, which is an important measure to assess the quality of the models. We'll then look at the Nash equilibrium, which is a state that we try to achieve during training. Finally, we will look closer at objective functions, which are very important to understand in order to implement GANs well.

Kullback-Leibler Divergence

Kullback-Leibler divergence (KL divergence), also known as relative entropy, is a method used to identify the similarity between two probability distributions.

$$D_{KL}(p||q) = \int_x p(x) log \frac{p(x)}{q(x)} dx$$

The KL divergence will be zero, or minimum, when *p(x)* is equal to *q(x)* at every other point.

Jensen-Shannon Divergence

The Jensen-Shannon divergence (also called the Information Radius (IRaD) or the total divergence to the average) is another measure of similarity between two probability distributions.
The following equation represents the Jensen-Shannon divergence between two probability distributions, p and q:

$$D_{JS}(p||q) = \frac{1}{2} D_{KL}(p||\frac{p+q}{2}) + \frac{1}{2} D_{KL}(q||\frac{p+q}{2})$$

Nash Equilibrium

The Nash Equilibrium describes a particular state in game theory, a non-cooperative game in which each player tries to pick the best possible strategy to gain the best possible outcome for themselves. Eventually, all the players reach a point at which they have all picked the best possible strategy for themselves.

Inception Score

The Inception score is the most widely used scoring algorithm for GANs. It uses a pre-trained inception V3 network (trained on Imagenet) to extract the features of both generated and real images. The quality of the model is good if it has a high inception score.

Fréchet Inception Distance

To overcome the various shortcomings of the inception score, the Fréchlet Inception Distance (FID) was proposed.

Disadvantages

GANs face some issues such as mode collapse, internal covariate shifts, and vanishing gradients.

Mode Collapse

Mode collapse is a problem that refers to a situation in which the generator network generates samples that have little variety or when a model starts generating the same images. Sometimes, a probability distribution is multimodal and very complex in nature. This means that it might contain data from different observations and that it might have multiple peaks for different sub-graphs of samples. Sometimes, GANs fail to model a multimodal probability distribution of data and suffer from mode collapse. Solutions include:

- By training multiple models (GANs) for different modes
- By training GANs with diverse samples of data

Vanishing Gradients

During backpropagation, gradient flows backward, from the final layer to the first layer. As it flows backward, it gets increasingly smaller. Sometimes, the gradient is so small that the initial layers learn very slowly or stop learning completely.
To overcome this problem, we can use activation functions such as ReLU, LeakyReLU, and PReLU. The gradients of these activation functions don't saturate during backpropagation.

Internal Covariate Shift

An internal covariate shift occurs when there is a change in the input distribution to our network. When the input distribution changes, hidden layers try to learn to adapt to the new distribution. This slows down the training process. If a process slows down, it takes a long time to converge to a global minimum.

Architecture of the Age-cGAN

The architecture of a cGAN for face aging is slightly more complicated. The Age-cGan consists of four networks: an **encoder**, the **FaceNet**, a **generator**, and a **discriminator**.

With the **encoder**, we learn the inverse mapping of input face images and the age condition with the latent vector. **FaceNet** is a face recognition network that learns the difference between an input image x and a reconstructed image. We have a **generator** network, which takes a hidden representation consisting of a face image and a condition vector and generates an image. The **discriminator** network is to discriminate between the real images and the fake images.

Encoder Network

The primary goal of the encoder network is to generate a latent vector of the provided images. Basically, it takes an image of a dimension of (64, 64, 3) and converts it into a 100-dimensional vector.

Generator Network

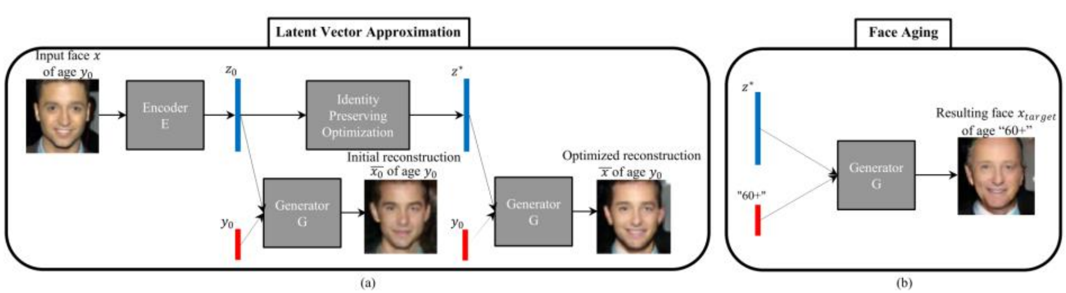
The primary goal of the generator is to generate an image of a dimension of (64, 64, 3). It takes a 100-dimensional latent vector and some extra information, y, and tries to generate realistic images. The generator network is a deep convolutional neural network too. It is made up of dense, upsampling, and convolutional layers. It takes two input values: a noise vector and a conditioning value. The conditioning value is the additional information provided to the network. For the Age-cGAN, this will be the age.

Face Recognition Network

The primary goal of the face recognition network is to recognize a person's identity in a given image.

Stages of the Age-cGAN

1. **Conditional GAN Training:** In this stage, we train the generator network and the discriminator network.
2. **Initial Latent Vector Approximation:** In this stage, we train the encoder network.
3. **Latent Vector Optimization:** In this stage, we optimize both the encoder and the generator network.



When Age-Invariant Face Recognition Meets Face Age Synthesis

To minimize the effects of age variation in face recognition, previous work either extracts identity-related discriminative features by minimizing the correlation between identity- and age-related features, called **age-invariant face recognition** (AIFR), or removes age variation by transforming the faces of different age groups into the same age group, called **face age synthesis** (FAS); however, the former lacks visual results for model interpretation while the latter suffers from artifacts compromising downstream recognition. How to minimize the effects of age variation is a lingering challenge for current face recognition systems to correctly identify faces in many practical applications such as finding lost children.

AIFR remains extremely challenging in the following three aspects. First, when the age gap becomes large in cross-age face recognition, age variation can largely affect the facial appearance, compromising the face recognition performance. Second, face age synthesis (FAS) is a complex process involving face aging since the facial appearance drastically changes over a long time and differs from person to person. Last, it is infeasible to obtain a large paired face dataset to train a model in rendering faces with natural effects while preserving identities. To overcome these issues, current methods for AIFR can be roughly divided into two categories, generative and discriminative models.

Recently, generative adversarial networks (GANs) have been successfully used to enhance the image quality of synthesized faces; they typically use the one-hot encoding to specify the target age group. However, the one-hot encoding represents the age group-level face transformation, ignoring the identity level personalized patterns and leading to unexpected artifacts. As a result, the performance of AIFR cannot be significantly improved due to the unpleasing synthesized faces and unexpected changes in identity.

On the other hand, the discriminative models focus on extracting age-invariant features by disentangling the identity-related information from the mixed information so that only the identity-related information is expected for the face recognition systems.

A Multi-Task Learning Framework

The study proposes a unified, multi-task learning framework to simultaneously achieve AIFR and FAS, termed **MTLFace**, which can enjoy the best of both worlds; i.e. learning age-invariant identity-related representation while achieving pleasing face synthesis.

First, decompose the mixed face feature into two uncorrelated components—identity- and age-related feature— through an attention mechanism, and then decorrelate these two components using multi-task training and continuous domain adaptation. Then, decorrelate these two components in a multi-task learning framework, in which an age estimation task is to extract age related features while a face recognition task is to extract identity-related features. Moreover, an identity conditional module is proposed to achieve identity-level transformation patterns for FAS, with a weight-sharing strategy to improve the age smoothness of synthesized faces.

Second, an attention-based feature decomposition is proposed to separate the age- and identity-related features on high-level feature maps, which can constrain the decomposition process in contrast to the previous unconstrained decomposition on feature vectors.

Third, compared to previous one-hot encoding achieving age group-level face transformation, a novel identity conditional module is proposed to achieve identity-level face transformation, with a weight sharing strategy to improve the age smoothness of synthesized faces.

Fourth, extensive experiments demonstrate the effectiveness of the proposed framework for AIFR and FAS on five benchmark datasets, and competitive performance on two popular GFR datasets.

Last, a large cross-age dataset of millions of faces with age and gender annotations was collected and released, which can advance the development of the AIFR and FAS.

An overview of the proposed MTLFace including two tasks.

- AIFR: The encoder E first extracts the mixed feature maps from input faces, which are then decomposed into two disjoint identity- and age-related feature maps by the multitask training and continuous domain adaptation.
- FAS: The decoder D produces synthesized faces through identity conditional module based on multi-level features; the PatchDiscriminator D_{img} penalizes the framework for better visual quality

This method differs from the previous work in the following aspects:

1- The proposed MTLFace achieves AIFR and FAS simultaneously to enhance the visual quality with identity-related information from AIFR.

2- The proposed identity conditional module (ICM) achieves an identity-level face age synthesis in contrast to the previous group-level face age synthesis.

3- A weight-sharing strategy in ICM can improve the age smoothness of synthesized faces.

Methodology

As the faces change a lot over time, the critical problem of AIFR is that the age variation usually introduces the increasing intra-class distances. As a result, it is challenging to correctly recognize two faces of the same person with a large gap, since the mixed facial representations are severely entangled with unrelated information such as facial shape and texture changes. Recently, a linear factorization module was designed to decompose the feature vectors into these two unrelated components. Formally, given the feature vector $x \in R^d$ extracted from an input image $I \in R^{3\times H\times W}$, their linear factorization module is defined as: $x = x_{age} + x_{id}$

However, it has the following drawbacks:

1- This decomposition performs on one dimensional feature vector, the resultant identity-related component lacks spatial information of face, not suitable for FAS.

2- This decomposition is unconstrained, which may lead to unstable training.

To address these drawbacks, we instead propose to decompose the mixed feature-maps at a high-level semantic space through an attention mechanism, termed attention based feature decomposition or AFD. The main reason is that manipulating on the feature vectors is more complicated than on the feature maps since the aging/rejuvenation effects, such as beards and wrinkles, appear in the semantic feature space but lose in the one-dimensional features.

Formally, a ResNet-like backbone is used as encoder E to extract mixed feature maps $X \in R^{C \times H' \times W'}$ from an input image I, i.e. X = E(I), the AFD can be defined as follows:

$$X = \underbrace{X \circ \sigma(X)}_{x_{age}} + \underbrace{X \circ 1 - \sigma(X)}_{x_{id}}$$

where $\circ$ denotes element-wise multiplication and σ represents an attention module.

REFERENCES

K. Zhang, Z. Li, and Y. Qiao (2016). Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks, IEEE signal processing letters, 10.

B.Kishore kumar, G. N. (2020). Missing children identification system using Deep Learning and. *JOICS* , 4.

Debayan Deb, D. A. (2019). Finding Missing Children: Aging Deep Face Features. *arXiv*, 10.

Grigory Antipov, M. B.-L. (2017). Face Aging With Conditional Generative Adversarial Networks. *arXiv*, 5.

Hao Wang, D. G. (2019). Decorrelated Adversarial Learning for Age-Invariant Face Recognition. *IEEE Xplore*, 10.

Leila Boussaad, A. B. (2020). An effective component-based age-invariant face recognition using Discriminant Correlation Analysis. *ScienceDirect*, 9.

Zhizhong Huang, J. Z. (2021). When Age-Invariant Face Recognition Meets Face Age Synthesis. *arXiv*, 12.