LARNet: Lie Algebra Residual Network for Face Recognition

Xiaolong Yang [1 2 3] **Xiaohong Jia** [1 2] **Dihong Gong** [3] **Dong-Ming Yan** [4 2] **Zhifeng Li** [3] **Wei Liu** [3]

Abstract

Face recognition is an important yet challenging problem in computer vision. A major challenge in practical face recognition applications lies in significant variations between profile and frontal faces. Traditional techniques address this challenge either by synthesizing frontal faces or by pose invariant learning. In this paper, we propose a novel method with Lie algebra theory to explore how face rotation in the 3D space affects the deep feature generation process of convolutional neural networks (CNNs). We prove that face rotation in the image space is equivalent to an additive residual component in the feature space of CNNs, which is determined solely by the rotation. Based on this theoretical finding, we further design a Lie Algebraic Residual Network (LARNet) for tackling pose robust face recognition. Our LARNet consists of a residual subnet for decoding rotation information from input face images, and a gating subnet to learn rotation magnitude for controlling the strength of the residual component contributing to the feature learning process. Comprehensive experimental evaluations on both frontal-profile face datasets and general face recognition datasets convincingly demonstrate that our method consistently outperforms the state-of-the-art ones.

1. Introduction

The recent development of deep learning models and an increasing variety of datasets have greatly advanced face recognition technologies (Liu et al., 2017; Wang et al., 2018a; Deng et al., 2019). Although many deep learning models are strong and robust to face recognition conducted in unconstrained environments, there remain quite a lot of challenges for recognizing faces across different age levels (Gong et al., 2013a; Wang et al., 2019; 2018b; Gong et al., 2015; Li et al., 2016b), different modalities (Li et al., 2014; Gong et al., 2017; Li et al., 2016a; Gong et al., 2013b; Luo et al., 2021), different poses (Huang et al., 2000; Cao et al., 2018a; Masi et al., 2016a; AbdAlmageed et al., 2016), and occlusions (Song et al., 2019; Zhang et al., 2007). In this paper, we develop a robust recognition algorithm to address the challenges for general face recognition, with a particular effect on matching faces across different poses (e.g., frontal vs. profile). Since the generalization ability of the deep model is closely related to the size of the training data, given an uneven and insufficient distribution of frontal and profile face images, the deep features tend to focus on frontal faces, and the learning results are only biased incomplete statistics. In order to tackle this problem, some work has reconstructed more datasets by different data augmentation methods. A typical way is to enrich input sources either by the synthesis of profile faces with appearance variations (Masi et al., 2016b) or by a set of images as one image input (Xie & Zisserma, 2018), so that the need for profile data is alleviated. Another way is to combine more data information, including multi-task learning (AbdAlmageed et al., 2016; Masi et al., 2016a; Yin & Liu, 2017) and template adaptation (Hassner et al., 2016; Crosswhite et al., 2018). Here, multi-task learning focuses on pose-aware targets, combined with richer information such as illumination, expression, gender, and age, to comprehensively boost the recognition performance; while the method based on template adaptation learning always creates a mean 3D model face, and by means of migration and mapping it avoids processing the 3D transformation at the image level. Nevertheless, these strategies tend to increase the unnecessary computational burden. Some other approaches use profile faces to synthesize frontal faces so that they can avoid large pose variations (Tran et al., 2017; Wang et al., 2017; Yin et al., 2017; Zhou et al., 2020). However, these methods would suffer from artifacts caused by occlusions and non-rigid expressions.

The above mentioned work mostly relies on additional data sources or additional labels. A recent new work called Deep Residual Equivalent Mapping (DREAM) (Cao et al., 2018a)
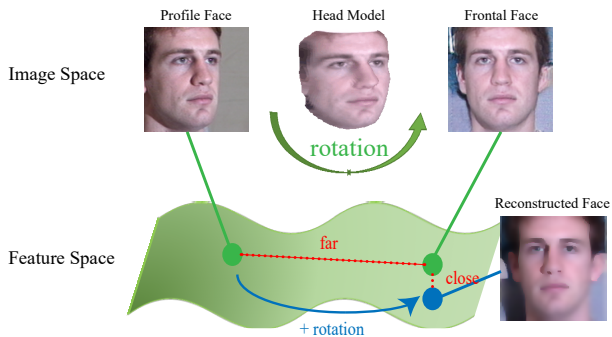

Figure 1. *Frontalization* or *rotation* in the feature space. Naturally in the real world, a frontal-profile pair is generated by head rotation, and we prove that a face rotation in the image space is equivalent to an additive residual component in the deep feature space. To show the equivalence, we reconstruct the image corresponding to the modified feature (blue dot) and provide the visual result for the expected frontal face.

has further discussed the gap between those features of frontal-profile pairs simply by approximating the difference using a learning model. Their approach of exploring this gap is similar to generative adversarial network (GAN), which makes the target sample (frontal face feature) and the generated sample (profile face feature) as close as possible through encoding and decoding.

We observe a natural fact that frontal-profile pairs are generated by head rotations, which should not be ignored in profile face recognition. However, rotation matrices are not easy to be embedded in CNNs. This is because the group of rotation matrices is closed under multiplication but not closed under addition, while the addition operation appears frequently in all gradient descent calculations. Benefiting from the pose estimation work (Tuzel & abd Peter Meer, 2008) in the field of *simultaneous localization and mapping* (SLAM), we develop a novel approach of using Lie algebra to achieve the updates of rotation matrices in CNNs.

We prove that for each frontal-profile pair linked by a rotation, their corresponding deep features also preserve a corresponding rotation relationship. To the best of our knowledge, this is the first attempt to theoretically explore and explain the physical relationship between the features of a frontal face and its profile counterpart. Based on this theoretical result, we propose the *Lie Algebra Residual Network* (LAR-Net). LARNet achieves face frontalization or rotation-and-render in the feature space, as shown in Fig. 1. Meanwhile, we conduct comparative experiments with more than 30 solutions under various evaluation criteria and metrics, while our method outperforms representative state-of-the-art competitors. In summary, our contributions are three-fold:

1. We theoretically prove that the features of frontal-profile pairs have a physical relationship based on ro-

tations in a residual network using Lie algebra, which is equivalent to an additive residual component in the deep feature space of CNNs.

2. We design a novel gating subnet based on the proof, which neither needs to modify the original backbone network structure nor relies on a large number of modules, but brings a great performance improvement.

3. LARNet enhances the deep model's ability in feature representation and feature classification, and accomplishes superior performance on various datasets and in various evaluation criteria, including frontal-profile face verification-identification and general face recognition tasks.

2. Related Work

We briefly discuss the most related work of profile face recognition and large-pose face recognition.

Insufficient Dataset. Many methods try to solve the profile face recognition problem by avoiding the unevenness of datasets. Masi *et al.* (2016b) proposed domain-specific data augmentation with increasing training data sizes for face recognition systems, and focused on important facial appearance variations. Multicolumn Network (Xie & Zisserma, 2018) and Neural Aggregation Network (NAN) (Yang et al., 2017) propose to use more information, such as a set of images or videos as input, to tackle the potential shortcomings of merely using a single image. These methods are not easy to avoid falsely matching profile faces of different identities or missing frontal and profile faces of the same identity. **Pose Variation.** Many existing methods have conducted in-depth researches on large poses. Template-adaptation-based work (Hassner et al., 2016; Crosswhite et al., 2018) is mainly engaged with transfer learning by means of a constructed classifier and synthesizer, and pooling based on image quality and head poses. As opposed to those techniques which expect a template model to learn pose invariance, pose aware deep learning methods (AbdAlmageed et al., 2016; Masi et al., 2016a) use multiple pose-specific models and rendered face images, which reduce the sensitivity to pose variations. More work favors using more labels instead of pose itself. Multi-task learning (MTL) has been widely used, which consists of pose, illumination, and expression estimations. Yin *et al.* (2017) proposed a pose-directed multi-task CNN and united the balance between different tasks. DebFace (Zhou et al., 2020) (de-biasing adversarial network) additionally takes gender, age, and race into consideration, and minimizes the correlation among feature factors so as to abate the bias influence from the other factors. These methods are effective; however, the use of multiple models and tasks tends to increase the computational cost, and the accuracy of their results is confined.

Frontalization. Since profile and large-pose faces bring more challenges, some methods directly use an available dataset to synthesize frontal faces to perform face recognition. Due to the widespread use of GAN, FF-GAN (Yin & Liu, 2017) and DR-GAN (Tran et al., 2017) surpass the performances of many competitors, with disentangled encoder-decoder structure towards learning a generative and discriminative representation. With the rapid progress of 3D face reconstruction technology, projecting rendering of frontal faces after reconstruction has also risen. Rotate-and-Render (Zhou et al., 2020) is a representative work from single-view images, and can leverage the recent advances in 3D face modeling and high-resolution GAN to constitute building blocks, since the 3D rotation-and-render of faces can be applied to arbitrary angles without losing facial details. Note that the reconstruction and synthesis approaches have their advantages in visualization performance, but their feature representation capability is inadequate for applications in unconstrained environments.

Feature Space. Some work considered features rather than image itself. Shi *et al.* (2019) proposed *Probabilistic Face Embeddings* (PFEs), which represent each face image as a Gaussian distribution in the latent space. Feature Transfer Learning (Yin et al., 2019) encourages the under-represented distribution to be closer to the regular distribution. Both of these two approaches target at making the sample distribution close to a Gaussian prior. Although in the above work features have been paid attention, specific datasets or face recognition in the real world cannot yet guarantee that samples are in a Gaussian distribution. Another representative work is DREAM (Cao et al., 2018a), which uses a residual network to directly modify the features of a profile face to the frontal one. DREAM bridges frontal-profile feature pairs by a mapping learned by deep learning; however, their ad-hoc designed results reach a bottleneck of feature-representation-based methods. Although both DREAM and our work explore the gap between frontal-profile feature pairs, DREAM mostly follows empirical observations and makes approximate estimation for the differences between the feature pairs, while our approach theoretically captures and embeds the rotation relationship between the feature pairs. To the best of our knowledge, our work is the first attempt to explore how face rotation in the 3D space impacts on the deep feature generation process of CNNs, and it is critical to mathematically reveal the relationship between the rotation and resulting features. Benefiting from the theoretical finding, LARNet has achieved superior performance, demonstrated by extensive experiments.

3. Methodology

In this section, we assume that a frontal face and its profile face have a corresponding rotation relationship in the origi-

nal 3D space. For ease of understanding, only the rotation with the orthogonal transformation relationship is discussed here. The derivation of more complex Euclidean transformation relationships, including translation and zooming, is referred to our supplementary material.

3.1. Problem Formulation

Our goal is to find a transformation between the features of an input profile face image and the expected frontal face image, to realize *frontalizition* in the deep feature space and to achieve a powerful feature representation robust to pose variations, as shown in Fig. 1.

We denote $\mathcal{F}(\mathbf{x})$ as a feature extraction function in CNNs for an input image $\mathbf{x}$. Here, for each pixel (u, v) in the image $\mathbf{x}$, we adopt its homogeneous coordinate representation $(u, v, 1)^{\top}$, and for convenience still denote the collection of these 3D homogeneous coordinates as $\mathbf{x}$.

Let d be the dimension of layers to be considered. Then the extracted feature $\mathcal{F}(\mathbf{x}) \in \mathbb{R}^d$. We shall prove that there exists a map $\mathcal{R}_{map}(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ that acts as a rotation in the deep feature space corresponding to the rotation $\mathbf{R} \in SO(3)$ of the (homogenized) image $\mathbf{x}$:

$$\mathcal{F}(\mathbf{R} \cdot \mathbf{x}) = \mathcal{R}_{map}(\mathcal{F}(\mathbf{x})). \tag{1}$$

For the frontal face image $\mathbf{x}_f$ and its profile face image $\mathbf{x}_p$, the homography transformation matrix of these two images degenerates into a rotation matrix: $\mathbf{x}_f = \mathbf{R} \cdot \mathbf{x}_p$, and we have:

$$\mathcal{F}(\mathbf{x}_f) = \mathcal{F}(\mathbf{R} \cdot \mathbf{x}_p) = \mathcal{R}_{map}(\mathcal{F}(\mathbf{x}_p)). \tag{2}$$

Furthermore, we try to use Lie group theory (Rossmann, 2002) and prove that the mapping $\mathcal{R}_{map}(\cdot)$ can be decomposed into an additive residual component, which is solely determined by the rotation:

$$\mathcal{F}(\mathbf{x}_f) = \mathcal{F}(\mathbf{x}_p) + \omega(\mathbf{R}) \cdot \mathbf{C}_{res}(\mathbf{R}, \mathbf{x}_p). \tag{3}$$

Thus, we only need a residual subnet $\mathbf{C}_{res}$ for decoding pose variant information from the input face image, and a gating subnet ω to learn rotation magnitude for controlling the strength of the residual component contributing to the feature learning process. Eq. (3) is the core principle of LARNet we propose, and the detailed derivations and experimental design will be presented in the following sections.

3.2. Rotation in Networks and Lie Algebra

To find what $\mathcal{R}_{map}$ exactly is, we have tried to directly explore and analyze the role of rotation $\mathbf{R}$ in networks from Eq. (2). From the original paper of ResNet (He et al., 2016), it proposes a novel *shortcut*, which not only retains the

depth of deep networks, but also has the advantages of shallow networks in avoiding the overfitting issue. The feature learning from shallow layer l to deep layer L is described as:

$$\mathbf{x}_L = \mathbf{x}_l + \sum_{i=l}^{L-1} H\left(\mathbf{x}_i, w_i\right), \tag{4}$$

$$
\begin{aligned}
\frac{\partial \mathbf{Loss}}{\partial \mathbf{x}_l} &= \frac{\partial \mathbf{Loss}}{\partial \mathbf{x}_L} \cdot \frac{\partial \mathbf{x}_L}{\partial \mathbf{x}_l} \\
&= \frac{\partial \mathbf{Loss}}{\partial \mathbf{x}_L} \left(1 + \frac{\partial}{\partial \mathbf{x}_l} \sum_{i=l}^{L-1} H\left(\mathbf{x}_i, w_i\right) \right),
\end{aligned} \tag{5}
$$

where $\mathbf{x}_l$ represents the input of the l-th residual block, and $H(\cdot)$ is the residual function with weights w. Since the second term at big brackets of Eq. (5) will quickly drop to 1, we focus on the first principal term $\partial \mathbf{Loss}/\partial \mathbf{x}_L$.

Note that the rotation matrix $\mathbf{R} \in SO(3)$ is not closed under matrix additions. Hence in nonlinear optimization in CNNs, an update of $\mathbf{R}$ using derivations does not yield a new rotation matrix (McKenzie, 2015). Therefore, a direct use of $\mathbf{R}$ is not appropriate, while we need to explore a new approach of embedding $\mathbf{R}$ in the network.

Inspired by the prior work (Tuzel & abd Peter Meer, 2008), we adopt Lie algebra with its own addition, multiplication, and derivative to replace the rotation matrix $\mathbf{R}$ in CNNs. First, for each rotation matrix $\mathbf{R} \in \mathbb{R}^{3 \times 3}$, it corresponds to a vector ϕ through the exponential mapping (Carmo, 1992):

$$\mathbf{R} = \exp(\phi^\wedge), \tag{6}$$

where $^\wedge$ is the skew-symmetric operator. The detailed definition of operator $^\wedge$ and a proof for Eq. (6) are placed into the supplementary material.

On the other hand, the vector ϕ can be obtained from $\mathbf{R}$ by the following Rodriguez' rotation formula (Rodriguez, 1840) and Taylor expansion:

$$
\begin{aligned}
\mathbf{R} = \exp(\theta \psi^\wedge) &= \sum_{n=0}^{\infty} \frac{1}{n!} (\theta \psi^\wedge)^n \\
&= \cos\theta \mathbf{I} + (1 - \cos\theta)\psi\psi^T + \sin\theta\psi^\wedge.
\end{aligned} \tag{7}
$$

Here $\phi = \theta\psi$ is in the *Axis-Angle representation* form, with a unit vector $\psi \in \mathbb{R}^3$ being the direction of the rotation axis and θ being the rotation angle according to the right hand rule, respectively. Since $\mathbf{R}\psi = \psi$, ψ is the eigenvector of the matrix $\mathbf{R}$ for eigenvalue $\lambda_{\mathbf{R}} = 1$. Eq. (7) leads:

$$tr(\mathbf{R}) = 2\cos\theta + 1. \tag{8}$$

Hence, we can solve ϕ as:

$$\phi = \theta\psi = \arccos(\frac{tr(\mathbf{R}) - 1}{2})\psi. \tag{9}$$

Next, we show the addition and multiplication in Lie algebra by *Baker-Campbell-Hausdorff* (BCH) formula (Wulf, 2002; Brian, 2015) and *Friedrichs'* theorem (Wilhelm, 1954; Jacobson, 1966):

$$
\begin{aligned}
\exp\left(\Delta\phi^\wedge\right)\exp\left(\phi^\wedge\right) &= \exp\left(\left(\phi + \mathbf{J}_l(\phi)^{-1}\Delta\phi\right)^\wedge\right), \\
\exp\left((\phi + \Delta\phi)^\wedge\right) &= \exp\left((\mathbf{J}_l\Delta\phi)^\wedge\right)\exp\left(\phi^\wedge\right).
\end{aligned} \tag{10}
$$

$\mathbf{J}_\ell$ is the *left Jacobian* of $SO(3)$. For a point $\mathbf{p} \in \mathbb{R}^3$, the derivative of $\mathbf{R}\mathbf{p}$ with respect to a perturbed rotation is:

$$
\begin{aligned}
\frac{\partial(\mathbf{R}\mathbf{p})}{\partial(\Delta\phi)} &= \lim_{\Delta\phi \to 0} \frac{\exp\left(\Delta\phi^\wedge\right)\exp\left(\phi^\wedge\right)\mathbf{p} - \exp\left(\phi^\wedge\right)\mathbf{p}}{\Delta\phi} \\
&= -(\mathbf{R}\mathbf{p})^\wedge.
\end{aligned} \tag{11}
$$

For a current $\mathbf{R}_i$, we choose a perturbation $\Delta\phi^\wedge$, such that $\mathbf{R}_{i+1} = \exp(\Delta\phi^\wedge)\mathbf{R}_i$. Then for a point $\mathbf{p}$, Eq. (11) leads:

$$\mathbf{R}_{i+1}\mathbf{p} = \exp(\Delta\phi^\wedge)\mathbf{R}_i\mathbf{p} \approx \mathbf{R}_i\mathbf{p} - (\mathbf{R}_i\mathbf{p})^\wedge\Delta\phi. \tag{12}$$

Then for the target function that is to be optimized, denoted by u, we use Taylor expansion to write:

$$
\begin{aligned}
u(\mathbf{R}_{i+1}\mathbf{p}) = u\left(\exp\left(\Delta\phi^\wedge\right)\mathbf{R}_i\mathbf{p}\right) &\approx u\left(\left(1 + \Delta\phi^\wedge\right)\mathbf{R}_i\mathbf{p}\right) \\
&\approx u\left(\mathbf{R}_i\mathbf{p}\right) - \underbrace{\frac{\partial u}{\partial \mathbf{d}}\bigg|_{\mathbf{d} = \mathbf{R}_i\mathbf{p}} (\mathbf{R}_i\mathbf{p})^\wedge}_{\boldsymbol{\delta}^T} \Delta\phi \\
&= u\left(\mathbf{R}_i\mathbf{p}\right) + \boldsymbol{\delta}^T\Delta\phi.
\end{aligned} \tag{13}
$$

We need to determine $\Delta\phi$ such that the value of u decreases. A possible choice is to pick $\Delta\phi = -\alpha D\delta$, where $\alpha > 0$ is a small step size and D is an arbitrary positive-definite matrix. Applying this perturbation within the scheme, we can update the rotation matrix by $\mathbf{R}_{i+1} \leftarrow \exp\left(-\alpha D\delta^\wedge\right)\mathbf{R}_i$.

Back to the original problem, given Eqs. (10-13), we can rewrite the first principal term of Eq. (5) as follows:

$$
\begin{aligned}
\frac{\partial \mathbf{Loss}}{\partial \mathbf{x}_L^f} &\approx \lim_{\Delta\phi \to 0} \frac{\partial \mathbf{Loss}}{\exp\left((\phi + \Delta\phi)^\wedge\right) \cdot \mathbf{x}_L^p - \exp\left(\phi^\wedge\right)\mathbf{x}_L^p} \\
&= \frac{\partial \mathbf{Loss}}{-(\mathbf{R} \cdot \mathbf{x}_L^p)^\wedge \cdot \partial\Delta\phi} \\
&= \frac{\partial \mathbf{Loss}}{\partial(\mathbf{R} \cdot \mathbf{x}_L^p)}.
\end{aligned} \tag{14}
$$

Note that in Eq. (2), we mentioned that the homography relationship between the two original images $\mathbf{x}_p$ and $\mathbf{x}_f$ is connected by a rotation, but this relationship generally cannot be guaranteed in the CNNs. However, Eq. (14) suggests

that this relationship is inherited in another way during gradient decent at each layer. In fact, since $\mathbf{R} \in SO(3)$, $\mathbf{R} \cdot \mathbf{x}_L^p$ and $\mathbf{x}_L^f$ are asymptotically stable according to Lyapunov's second method (Lyapunov, 1992; Bhatia & Szegö, 2002). With the gradual training progress of the ResNet, the feature vectors of $\mathbf{R} \cdot \mathbf{x}_L^p$ and $\mathbf{x}_L^f$ have the same convergent representation: $\mathcal{F}(\mathbf{x}_f) = \mathcal{F}(\mathbf{R} \cdot \mathbf{x}_p)$. Furthermore, we decouple the rotation relation from face features into Eq. (9) and Eq. (12). Let $V_{res} = \mathcal{F}(\mathbf{R} \cdot \mathbf{x}_p) - \mathcal{R}_{map}\mathcal{F}(\mathbf{x}_p) \in \mathbb{R}^d$ be the residual vector, and we have:

$$\mathcal{R}_{map}^{-1}\mathcal{F}(\mathbf{x}_f) = \mathcal{F}(\mathbf{x}_p) + \mathcal{R}_{map}^{-1} \cdot V_{res},$$
$$\mathcal{F}(\mathbf{x}_f) = \mathcal{F}(\mathbf{x}_p) + \mathcal{R}_{map}^{-1}(V_{res} + \mathcal{R}_{map}\mathcal{F}(\mathbf{x}_f) - \mathcal{F}(\mathbf{x}_f)). \tag{15}$$

Since during our training stage, the feature $\mathcal{F}(\mathbf{x}_p)$ is approaching to $\mathcal{R}_{map}\mathcal{F}(\mathbf{x}_f)$ (we shall show a corresponding analysis in Eq. (17) in Sec. 3.3), Eq. (15) leads to Eq. (16):

$$\mathcal{F}(\mathbf{x}_f) \approx \mathcal{F}(\mathbf{x}_p) + \mathcal{R}_{map}^{-1}(\mathcal{F}(\mathbf{x}_p) - \mathcal{R}_{map}\mathcal{F}(\mathbf{x}_p)). \tag{16}$$

This agrees exactly with Eq. (3). Hence, we can design the gating control function $\omega(\mathbf{R})$ as $\mathcal{R}_{map}^{-1}$ to filter the feature flow, and the component $\mathbf{C}_{res}(\mathbf{R}, \mathbf{x}_p) = \mathcal{F}(\mathbf{x}_p) - \mathcal{R}_{map}\mathcal{F}(\mathbf{x}_p)$ is obtained through residual network training.
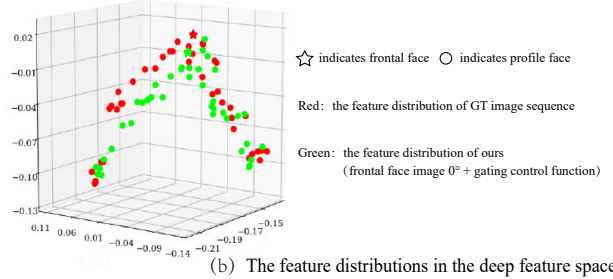
3.3. The Architecture of Subnet

As previously stated, we expect to design a residual subnet $\mathbf{C}_{res}$ for decoding pose variant information from input face images. The residual formulation in Eq. (3) allows us to use a succinct enough network structure for learning the residual compensation from the clean deep features, which is a relatively easy task. The most convenient way is to add the gating control residuals directly to the existing backbone (Arcface (Deng et al., 2019) with ResNet-50 in our paper). Residual learning can be arranged before the final fully-connected (FC) layer of the ResNet-50 without revising any learned parameters of the original backbone model. Our residual learning has two fully-connected layers with Parametric Rectified Linear Unit (PReLU) (He et al., 2015) as the activation function. We train it by minimizing ℓ_2 norm of the difference between the profile features $\mathcal{F}(\mathbf{x}_p)$ and frontal features under the rotation $\mathcal{R}_{map}\mathcal{F}(\mathbf{x}_f)$ using stochastic gradient descent.

$$\min_{\Omega_p} \Sigma||\mathcal{F}(\mathbf{x}_p) - \mathcal{R}_{map}(\Omega_p)\mathcal{F}(\mathbf{x}_f)||_2^2, \tag{17}$$

where Ω_p denotes the learnable parameters. We train this subnet on frontal-profile pairs sampled from the MS-Celeb-1M dataset (mentioned in Sec.4.2), and fix these parameters for the testing. Applying a subnet with complicated structure may increase the risk of overfitting, and the design with two FC layers is on the consideration of both the task difficulty and the risk of model robustness.

(a) A ground-truth image sequnce：pose varirant is from -90° to 90°.

☆ indicates frontal face ◯ indicates profile face

Red: the feature distribution of GT image sequence

Green: the feature distribution of ours
(frontal face image 0° + gating control function)

(b) The feature distributions in the deep feature space

Figure 2. The effect of our gating control function for the same identity. (a) The top is a sequence of images taken in real life, with pose variant being from $-90°$ to $90°$ for the same individual. (b) The bottom is the feature distribution in the deep feature space. The dots represent profile faces while stars denote frontal faces. The red dots are the feature vectors generated by the image sequence, and the green dots are the feature vectors of the frontal face image ($0°$) with different yaw angle variants simulated by our gating control function. Their similar distributions indicate that our gating control function maps the features of the frontal and profile faces closer, thereby enhancing the feature representation ability to accommodate pose variations.

Furthermore, we design a gating control function ω to analyze the rotation magnitude for controlling the strength of the residual component contributing to the deep feature learning process. In our problem context, ω needs to satisfy the following properties:

● $\omega \in [0, 1]$. Intuitively, when the input is frontal face $\mathbf{x}_0$, there is almost no difference in the feature representation in the same network, and $\mathbf{C}_{res}$ of residual learning will bring errors and weaken the classification ability. Therefore, it is expected that the gating control function is 0 in this case; ideally, the magnitude of the residual is thus the largest at the complete profile pose: $\mathcal{F}(\mathbf{x}_0) - \mathcal{F}(\mathbf{x}_{\pi/2})$, so the maximum value of the gating control function is 1:

$$\mathcal{F}(\mathbf{x}_0) = \mathcal{F}(\mathbf{x}_{\pi/2}) + 1 * (\mathcal{F}(\mathbf{x}_0) - \mathcal{F}(\mathbf{x}_{\pi/2})) = \mathcal{F}(\mathbf{x}_0).$$

● ω has symmetric weights. A gating control function learns rotation magnitude for controlling the strength of the residual component contributing to the feature learning process, and the same deflection angles should bring the same influence on frontal faces (for example, yaw angle with turning left or right). We will also use data flipping augmentation to strengthen the symmetry of the model during training.

Besides, it is worth noting that roll, yaw, and pitch angles have different contributions to the final face recognition performance. The effect of roll will be eliminated by face alignment (mentioned in Sec. 4.2), while face images with

false positive pair

frontal face data

profile face data
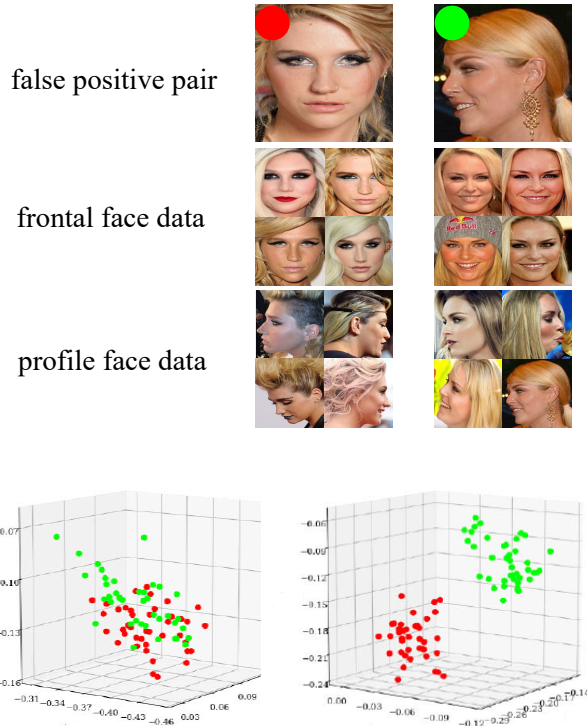
The feature distributions of the other method (left) and ours (right)

Figure 3. The effect of our gating control function for different identities. This is a challenging false positive example for a general face recognition model (Wang et al., 2018a). We collect more frontal and profile face data of those two individuals from the Celebrities Frontal-Profile dataset (Sengupta et al., 2016), and visualize feature distributions of all the images. It is obvious that our model with the gating control function has a better classification and clustering ability.

large pitch angles are relatively rare. Combining all of the above constraints and solving Eq. (9) by Chebyshev polynomial approximation, we get $\omega = |\sin\theta|$ with $\sin\theta = ||(\sin_{pitch}, \sin_{yaw}, \sin_{roll})||_{\infty}$ for all angles $\in [-\pi/2, \pi/2]$, which ensures that there is a one-to-one correspondence between the elements in Lie algebra ϕ and the rotation $\mathbf{R}$ and also guarantees the completeness of the proposed theory.

As a way to demonstrate the effectiveness of our proposed subnet, Fig. 2 illustrates for the same identity. When inputs are image sequences of the same individual with different yaw angles, ResNet can extract features and display the distribution of those vectors as red dots. Meanwhile, we use our gating control function with the only frontal face image to simulate pose variations. Our results' distribution is marked by green dots. Through the visualization, it can be clearly seen that our model can accurately simulate the feature vector distribution of different faces varying from yaw angles, which proves that our gating control function improves the feature representation capability, especially amenable to pose variations.

In addition, Fig. 3 illustrates the effect of our subnet for different identities. This is a challenging example even for the almost blameless face recognition model (Wu et al., 2016). We collect more frontal and profile face data of those two individuals, and visualize feature vectors of all the images corresponding to this sample. It is obvious that our subnet with the gating control function is instrumental in obtaining better classification and clustering performance.

4. Experimental Results

In this section, we first provide a description of implementation details (Sec. 4.1). Besides, we list all the datasets used in the experiments and briefly explain their own characteristics (Sec. 4.2). Furthermore, we present two ablation studies on the architecture and gating control function, respectively, which explain the effectiveness of our experimental design on recognition performance (Sec. 4.3). We also compare with existing methods and some findings about profile face representation, and conduct extensive experiments on frontal-profile face verification-identification and general face recognition tasks (Sec. 4.4).

4.1. Implementation Details

Network Architecture. (1) Backbone: our backbone network is ResNet with a combined margin loss: CM(1, 0.3, 0.2) proposed by ArcFace. According to recent researches, *i.e.*, Saxe *et al.* (2014), Highway Networks (Srivastava et al., 2015), and Balduzzi *et al.* (2017), ResNet-50 has the best layers balancing efficiency and accuracy. Therefore, we choose the ResNet with 50 layers as it is very popular in a large amount of existing work and also convenient to compare. (2) LARNet: after the backbone, the clean deep features are mapped to *rotation* through the subset and gating control function. (3) LARNet+: similar to DREAM (Cao et al., 2018a), we also use the end-to-end mechanism to further improve the performance of our results. Based on LARNet, we make residual learning together with the backbone network in an end-to-end manner. We train the ResNet and residual learning together and then train the residual learning separately with pose-variant frontal-profile face pairs.

Data Preprocessing. As shown in Fig. 4, we use MTCNN (Zhang et al., 2016) to detect face areas and facial landmarks on both training and testing sets. We use a flipping strategy to achieve data augmentation and enhance our model's ability to learn symmetry. In addition, face alignment and scaling (224×224) are taken into account to reduce the impact caused by translation and zooming, when we only consider $SO(3)$ instead of $SE(3)$.

Training Details. The model is trained with 180K iterations. The initial learning rate is set as 0.1, and is divided by

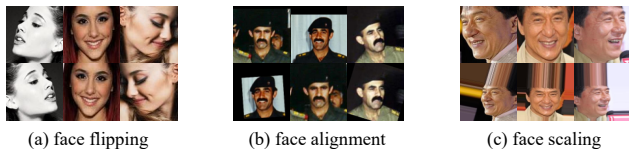(a) face flipping   (b) face alignment   (c) face scaling

Figure 4. Data prepossessing on frontal and profile faces. (a) Face flipping: data enhancement for strengthening the model's ability to learn symmetry; (b) face alignment: reduce the impact caused by translation and rotation in a plane such that the eyes lie along a horizontal line; (c) face scaling: reduce the impact caused by zooming because focal length differs in every image such that faces are approximately identical in size.

10 after 100K, 160K iterations. The SGD optimizer has momentum 0.9, and weight decay $5e-4$. For rotation angles, we take rotation estimation via state-of-the-art work (Yang et al., 2019; 2020) and obtain pose labels. The protocols of training and testing for different datasets will be explained in the next section.

4.2. Datasets Exhibition

Training Data. We separately employ the two most widely used face datasets as training data in order to conduct fair comparison with the other methods, *i.e.*, cleaned MS-Celeb-1M database (**MS1MV2**) (Guo et al., 2016) and **CASIA-WebFace** (Yi et al., 2014). MS1MV2 is a clean version of the original MS-Celeb-1M face dataset that has too many mislabeled images, containing 5.8M images of 85,742 celebrities. CASIA-WebFace uses tag-similarity clustering to remove noise of the data source, containing 500K images of 100K celebrities from IMDb.

Testing Data. We explore several efficient face verification datasets for testing. Celebrities in Frontal-Profile (**CFP**) (Sengupta et al., 2016) is a challenging frontal to profile face verification dataset, containing 500 celebrities, each of which has 10 frontal and 4 profile face images. We further test another challenging dataset IARPA Janus Benchmark A (**IJB-A**) (Klare et al., 2015) that covers extreme poses and illuminations, containing 500 identities with 5,712 images and 20,414 frames extracted from videos. Besides focusing on frontal-profile face verification, we also conduct experiments on general face recognition datasets to verify that our method can reach the state of the art on general face recognition tasks. Including the most widely used **LFW** (Huang et al., 2008) dataset (13,233 face images from 5,749 identities) and **YTF** (Wolf et al., 2011) dataset (3,425 videos of 1,595 different people), we also report the performance of Cross-Pose LFW (**CPLFW**) (Zheng & Deng, 2018), which deliberately searches and selects 3,000 positive face pairs with pose differences to add pose variations to intra-class variance, so the effectiveness of several face verification methods can be fully justified. Furthermore, we also extensively conduct a more in-depth ablation experiment on the Large-scale CelebFaces Attributes (**CelebA**) dataset (Liu

Table 1. Ablation study on architecture. Evaluation is conducted on the CFP-FP dataset.

Architecture	Verification (%)
Backbone	92.96
LARNet	98.84
LARNet+	**99.21**

Table 2. Ablation study on gating control function. Evaluation is conducted on the CelebA dataset with metric Equal Error Rate.

Gating Control Function	EER (%)
Identity mapping: $\omega \equiv 1$	15.35
Linear mapping: $\omega = 2\theta/\pi$	9.68
Nolinear mapping: $\omega = sigmoid(4\theta/\pi - 1)$	8.45
PReLU	9.72
cReLU with OW	7.92
LARNet: $\omega = |\sin\theta|$	**6.26**

et al., 2015b), containing 10,177 celebrities and 202,599 face images, which covers large pose variations.

4.3. Ablation Studies

To justify that our LARNet does improve the performance of profile face recognition, we conduct two ablation experiments: 1) the architectures with/without the gating control function, and 2) the forms of the gating control function.

4.3.1. THE ARCHITECTURE OF LARNET

We study the effectiveness of architectures with and without the gating control function as well as an end-to-end optimization manner. All the results are using the same backbone ResNet-50, combined margin loss in Arcface (Deng et al., 2019), and MS1MV2 training dataset, and evaluation is conducted on the CFP-FP dataset. The baseline method is Arcface without refinement. As shown in Table 1, we observe that compared with the strong baseline without the gating control function, our LARNet brings a significant advancement by 5.88% in verification accuracy. LARNet+ with the end-to-end optimization manner also plays an important role in further performance improvement, and achieves a more superior result 99.21% than LARNet 98.84%.

4.3.2. THE GATING CONTROL FUNCTION

Next, we further analyze that which type of gating control function has a greater contribution to the performance. For facilitating a fair comparison, all methods take the same CASIA-WebFace dataset and ResNet-50 backbone for training, and evaluation is conducted on CeleA with metric Equal Error Rate (EER). In Table 2, identity mapping means $\omega \equiv 1$, which represents some GAN-based work, ignoring the internal connection of frontal-profile face pairs, and only relying on the generator and the discriminator to produce results. Linear mapping: $\omega = 2\theta/\pi$ is a natural attempt and meets the design constraints. Besides,

Table 3. Quantitative evaluation on the IJB-A dataset, where o.s. denotes the optimal setting, and f. is fine-tuning/refinement. Symbol '-' indicates that the metric is not available for that protocol.

Method	TAR@FAR=0.01	TAR@FAR=0.001	Rank-1 Acc.	Rank-5 Acc.
Wang *et al.* (2016)	0.729	0.510	0.822	0.931
Pooling Faces (Hassner et al., 2016)	0.819	0.631	0.846	0.933
Multi Pose-Aware (AbdAlmageed et al., 2016)	0.787	—	0.846	0.927
DCNN Fusion (f.) (Chen et al., 2016)	0.838	—	0.903	0.965
PAMs (Masi et al., 2016a)	0.826	0.652	0.840	0.925
Augmentation+Rendered (Masi et al., 2016b)	0.886	0.725	0.906	0.962
Multi-task learning (Yin & Liu, 2017)	0.787	—	0.858	0.938
TPE(f.) (Sankaranarayanan et al., 2017)	0.900	0.813	0.932	—
DR-GAN (Tran et al., 2017)	0.831	0.699	0.901	0.953
FF-GAN (Yin et al., 2017)	0.852	0.663	0.902	0.954
NAN (Yang et al., 2017)	0.921	0.861	0.938	0.960
Multicolumn (Xie & Zisserma, 2018)	0.920	—	—	—
VGGFace2 (Cao et al., 2018b)	0.904	—	—	—
Template Adaptation(f.) (Crosswhite et al., 2018)	0.939	—	0.928	—
DREAM (Cao et al., 2018a)	0.872	0.712	0.915	0.962
DREAM(E2E+retrain,f.) (Cao et al., 2018a)	0.934	0.836	0.939	0.960
FTL with 60K parameters (o.s.) (Yin et al., 2019)	0.864	0.744	0.893	0.947
PFEs (Shi & Jain, 2019)	0.944	—	—	—
DebFace (Gong et al., 2020)	0.902	—	—	—
Rotate-and-Render (Zhou et al., 2020)	0.920	0.825	—	—
HPDA (Wang et al., 2020)	0.876	0.803	0.84	0.88
CDA (Wang & Deng, 2020)	0.911	0.823	0.936	0.957
LARNet	0.941	0.842	0.936	0.968
LARNet+	**0.951**	**0.874**	**0.949**	**0.971**

our gating control function essentially acts as a filtering activation function, and we compare it with two widely used activation functions PReLU (He et al., 2015) and cReLU with OW (Balduzzi et al., 2017). Nolinear mapping: $\omega = sigmoid(4\theta/\pi - 1)$ is reported by DREAM (Cao et al., 2018a). Our method $\omega = |\sin\theta|$ still achieves the best result: EER 6.26%. This observation ascertains our design of exerting a higher degree of correction to a profile face, and has a better understanding about the effectiveness of our proposed LARNet for profile face recognition.

4.4. Quantitative Evaluation Results

We compare our method with more than 30 competitors published in the recent five years with respect to different metrics, which include template based, GAN, residual learning, and three-dimensional reconstruction, aiming at handling various tasks such as face search, face recognition, face verification, large pose recognition, etc. All numerical statistics are the best results obtained from original quotation, cross-reference, and experimental reproduction.

4.4.1. IJBA DATASET: VERIFICATION AND IDENTIFICATION WITH STATE OF THE ARTS

In this experiment, we evaluate our method on the challenging benchmark IJBA that covers full pose variations and complies with the original standard protocol. The evaluation metrics include popular True Acceptance Rate (TAR) at False Acceptance Rate (FAR) of 0.01 and 0.001 on the

verification task, and Rank-1/Rank-5 recognition accuracy on the identification task. All methods employ the same MS1MV2 dataset and ResNet-50 backbone for training.

In Table 3, we compare with various state-of-the-art techniques. Our LARNet reaches 0.941 (TAR@FAR=0.01), and after refinement with end-to-end retraining, LARNet+ achieves a better performance with 0.951. They have surpassed the other existing methods by a large margin. Our method also brings significant improvement in a more challenging metric TAR@FAR=0.001, with the results of 0.842 and 0.874, respectively. Furthermore, for face identification, LARNet has an advantage in both Rank-1 Acc. (0.936) and Rank-5 Acc. (0.968), and LARNet+ pushes the result to a higher level: Rank-1 Acc. (0.949) and Rank-5 Acc. (0.971). Our method has achieved superior performance on both recognition and verification tasks.

4.4.2. CFP-FP DATASET: PROFILE FACE VERIFICATION CHALLENGE

We employ CFP-FP as the frontal profile face verification dataset with the protocol that the whole dataset is divided into 10 folds each containing 350 same and 350 not-same pairs of 50 individuals. All methods employ the same MS1MV2 dataset and ResNet-50 backbone for training. From Table 4, we observe that the face verification results of state-of-the-art face recognition models are basically at 94%+. It is worth noting that in 2020, a latest work named universal representation learning face (URFace) (Shi et al.,

Table 4. Quantitative evaluation on the CFP-FP dataset, where o.s. denotes the optimal setting, and f. is fine-tuning/refinement.

Method	Verification (%)
SphereFace (o.s.+f.)	94.17
CosFace (o.s.)	94.40
ArcFace (o.s.+f.)	94.04
URFace (all modules, MS1MV2, o.s.)	98.64
Human-level	98.92
LARNet	98.84
LARNet+	**99.21**

Table 5. Quantitative evaluation on the general face recognition datasets: LFW, YTF, and CPLFW, where o.s. denotes the optimal setting, and f. is fine-tuning/refinement. Symbol '-' indicates that the metric is not available for that protocol. For fairness, the training network and dataset: ResNet-50 + CASIA-WebFace. Symbol '*' indicates the methods under different training designs. The evaluation metric is verification rate (%).

Method	LFW	YTF	CPLFW
HUMAN-Individual	97.27	-	81.21
HUMAN-Fusion	**99.85**	-	85.24
DeepID (Sun et al., 2014)	99.47	93.20	-
Deep Face (Taigman et al., 2014)	97.35	91.4	-
VGG Face (Parkhi et al., 2015)	98.95	97.30	90.57
FaceNet (Schroff et al., 2015)	99.63	95.10	-
Baidu (Liu et al., 2015a)	99.13	–	-
Center Loss (Wen et al., 2016)	99.28	94.9	85.48
Range Loss (Zhang et al., 2017)	99.52	93.70	-
Marginal Loss (Deng et al., 2017)	99.48	95.98	-
SphereFace(o.s.) (Liu et al., 2017)	99.42	95.0	81.4
SphereFace+(o.s.)	99.47	–	90.30
CosFace(o.s.) (Wang et al., 2018a)	99.51	96.1	-
CosFace*(MS1MV2,R64, o.s.)	99.73	97.6	-
Arcface(o.s.) (Deng et al., 2019)	99.53	–	92.08
ArcFace*(MS1MV2,R100,f.)	99.83	**98.02**	95.45
Ours: LARNet	99.36	96.55	95.51
Ours: LARNet+	99.71	97.63	**96.23**

2020) has reached an astonishing 98.64% under the training of the MS1MV2 dataset and the auxiliary learning of a large number of modules, such as variation augmentation, confidence-aware identification loss, and multiple embeddings. However, the result of our LARNet has reached 98.84%, which outperforms almost all competitors. Regarding further advanced LARNet+, to our best knowledge, the 99.21% performance is the first to surpass the reported human-level performance (98.92%) on the CFP-FP dataset.

4.4.3. LFW, YTF, AND CPLFW DATASETS: GENERAL FACE RECOGNITION

To get a better understanding about our LARNet, we conduct a more in-depth comparison on general face recognition. LFW and YTF datasets are the most widely used benchmark for unconstrained face verification on images and videos. In this experiment, we follow the unrestricted with labelled outside data protocol to report the performance. CPLFW

emphasizes pose difference to further enlarge intra-class variance. All methods employ the same CASIA-WebFace dataset and ResNet-50 backbone, while the results of some different experimental designs (marked with '*') are also reported.

From Table 5, we find that because the LFW dataset is too small, almost all methods can achieve 99%+. Although the meaning is very weak, the result 99.74% of our method is also at the forefront, only lower than human-level 99.85% and Arcface* 99.83%. It is worth noting that Arcface* is trained with the MS1MV2 dataset (5.8M) and ResNet-100, while Arcface with CASIA-WebFace (100K) and shallower ResNet-50 (the same as ours) only reaches 99.53% and is inferior to ours. We provide this more comparative result for an extensive study. For the video sampled dataset YTF, there exists the same situation, and our LARNet and LARNet+ are still superior to all ResNet-50 based face recognition methods and slightly inferior to deeper ResNet-100 based Arcface*. We also introduce a more challenging CPLFW dataset with large poses, which has more realistic consideration on pose intra-class variations. Our method achieves the best results 95.51% and 96.23%, respectively.

5. Conclusions

We proposed the Lie Algebra Residual Network (LARNet) to boost face recognition performance. First, we presented a novel method with Lie algebra theory to explore how face rotation in the 3D space affects the deep feature generation process, and proved that a face rotation in the image space is equivalent to an additive residual component in the deep feature space. Furthermore, we designed a gating control function, which is derived on the foundation of Lie algebra to learn rotation magnitude and control the impact of the residual component on the feature learning process. Moreover, we provided the results of ablation studies to validate the effectiveness of our Lie algebraic deep feature learning. The comprehensive experimental evaluations demonstrate the superior performance of the proposed LARNet over the state-of-the-art methods on frontal-profile face verification, face identification, and general face recognition tasks. In future work, we will continue to pay attention to the interpretability of Euclidean transformation in other CNNs, and intend to explore more mathematical tools to decouple potential geometric properties hidden in the image space.

Acknowledgements

References

AbdAlmageed, W., Wu, Y., Rawls, S., Harel, S., Hassner, T., Masi, I., Choi, J., Lekust, J., Kim, J., Natarajan, P., Nevatia, R., and Medioni, G. Face recognition using deep multi-pose representations. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1–9, 2016.

Balduzzi, D., Frean, M., Leary, L., Lewis, J., Ma, K. W.-D., and McWilliams, B. The shattered gradients problem: If ResNets are the answer, then what is the question? In *The International Conference on Machine Learning (ICML)*, pp. 342–350, 2017.

Bhatia, N. P. and Szegö, G. P. *Stability Theory of Dynamical Systems*. Springer Science & Business Media, 2002.

Brian, H. *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*. Springer, 2015.

Cao, K., Rong, Y., Li, C., Tang, X., and Loy, C. C. Pose-robust face recognition via deep residual equivariant mapping. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5187–5196, 2018a.

Cao, Q., Shen, L., Xie, W., Parkhi, O. M., and Zisserman, A. Vggface2: A dataset for recognising faces across pose and age. In *IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, pp. 67–74, 2018b.

Carmo, M. P. D. *Riemannian Geometry*. Birkhäuser, 1992.

Chen, J.-C., Patel, V. M., and Chellappa, R. Unconstrained face verification using deep CNN features. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1–9, 2016.

Crosswhite, N., Byrne, J., Stauffer, C., Parkhi, O., Cao, Q., and Zisserman, A. Template adaptation for face verification and identification. *Image and Vision Computing*, 79: 35–48, 2018.

Deng, J., Zhou, Y., and Zafeiriou, S. Marginal loss for deep face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 60–68, 2017.

Deng, J., Guo, J., Xue, N., and Zafeiriou, S. ArcFace: Additive angular margin loss for deep face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4690–4699, 2019.

Gong, D., Li, Z., Lin, D., Liu, J., and Tang, X. Hidden factor analysis for age invariant face recognition. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 2872–2879, 2013a.

Gong, D., Li, Z., Liu, J., and Qiao, Y. Multi-feature canonical correlation analysis for face photo-sketch image retrieval. In *Proceedings of the 21th ACM International Conference on Multimedia*, pp. 617–620, 2013b.

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

Gong, D., Li, Z., Huang, W., Li, X., and Tao, D. Heterogeneous face recognition: A common encoding feature discriminant approach. *IEEE Transactions on Image Processing(TIP)*, 26(5):2079–2089, 2017.

Gong, S., Liu, X., and Jain, A. K. Jointly De-biasing face recognition and demographic attribute estimation. In *European Conference on Computer Vision (ECCV)*, pp. 330–347, 2020.

Guo, Y., Zhang, L., Hu, Y., He, X., and Gao, J. MS-Celeb-1M: A dataset and benchmark for large-scale face recognition. In *European Conference on Computer Vision (ECCV)*, pp. 87–102, 2016.

Hassner, T., Masi, I., Kim, J., Choi, J., Harel, S., Natarajan, P., and Medioni, G. Pooling faces: Template based face recognition with pooled face images. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 59–67, 2016.

He, K., Zhang, X., Ren, S., and Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 1026–1034, 2015.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

Huang, F. J., Zhou, Z., Zhang, H.-J., and Chen, T. Pose invariant face recognition. In *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 245–250, 2000.

Huang, G. B., Mattar, M., Berg, T., and Learned-Miller, E. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In *Workshop on faces in Real-Life Images: detection, alignment, and recognition*, 2008.

Jacobson, N. *Lie Algebras*. John Wiley & Sons, 1966.

Klare, B. F., Klein, B., Taborsky, E., Blanton, A., Cheney, J., Allen, K., Grother, P., Mah, A., and Jain, A. K. Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus Benchmark A. In *IEEE Conference*

on Computer Vision and Pattern Recognition (CVPR), pp. 1931–1939, 2015.

Li, Z., Gong, D., Qiao, Y., and Tao, D. Common feature discriminant analysis for matching infrared face images to optical face images. *IEEE Transactions on Image Processing(TIP)*, 23(6):2436–2445, 2014.

Li, Z., Gong, D., Li, Q., Tao, D., and Li, X. Mutual component analysis for heterogeneous face recognition. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 7(3):1–23, 2016a.

Li, Z., Gong, D., Li, X., and Tao, D. Aging face recognition: A hierarchical learning model based on local patterns selection. *IEEE Transactions on Image Processing(TIP)*, 25(5):2146–2154, 2016b.

Liu, J., Deng, Y., Bai, T., Wei, Z., and Huang, C. Targeting ultimate accuracy: Face recognition via deep embedding. In *arXiv preprint:1506.07310*, 2015a.

Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., and Song, L. SphereFace: Deep hypersphere embedding for face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 212–220, 2017.

Liu, Z., Luo, P., Wang, X., and Tang, X. Deep learning face attributes in the wild. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 3730–3738, 2015b.

Luo, Y., Zhang, Y., Yan, J., and Liu, W. Generalizing face forgery detection with high-frequency features. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

Lyapunov, A. M. The general problem of the stability of motion. *International Journal of Control*, 55(3):531–534, 1992.

Masi, I., Rawls, S., Medioni, G., and Natarajan, P. Pose-aware face recognition in the wild. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4838–4846, 2016a.

Masi, I., Tran, A. T., Hassner, T., Leksut, J. T., and Medioni, G. Do we really need to collect millions of faces for effective face recognition? In *European Conference on Computer Vision (ECCV)*, pp. 579–596, 2016b.

McKenzie, D. W. *An Elementary Introduction to Lie Algebras for Physicists*. Cornell University, 2015.

Parkhi, O. M., Vedaldi, A., and Zisserman, A. Deep face recognition. In *British Machine Vision Association*, pp. 1–12, 2015.

Rodriguez, O. Des lois geometriques qui regissent les deplacements d'un systeme solide dans l'espace et de la variation des coordonnees provenant de deplacements consideres independamment des causes qui peuvent les produire. *J Mathematiques Pures Appliquees*, 5:380–440, 1840.

Rossmann, W. *Lie Groups: An Introduction Through Linear Groups*. Oxford Press, 2002.

Sankaranarayanan, S., Alavi, A., Castillo, C. D., and Chellappa, R. Triplet probabilistic embedding for face verification and clustering. In *IEEE International Conference on Biometrics theory, Applications and Systems (BTAS)*, 2017.

Saxe, A. M., McClelland, J. L., and Ganguli, S. Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. In *The International Conference on Learning Representations (ICLR)*, 2014.

Schroff, F., Kalenichenko, D., and Philbin, J. FaceNet: A unified embedding for face recognition and clustering. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 815–823, 2015.

Sengupta, S., Chen, J.-C., Castillo, C., Patel, V. M., Chellappa, R., and Jacobs, D. W. Frontal to profile face verification in the wild. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1–9, 2016.

Shi, Y. and Jain, A. K. Probabilistic face embeddings. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 6902–6911, 2019.

Shi, Y., Yu, X., Sohn, K., Chandraker, M., and Jain, A. K. Towards universal representation learning for deep face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6817–6826, 2020.

Song, L., Gong, D., Li, Z., Liu, C., and Liu, W. Occlusion robust face recognition based on mask learning with pairwise differential siamese network. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 773–782, 2019.

Srivastava, R. K., Greff, K., and Schmidhuber, J. Highway networks. In *The International Conference on Machine Learning (ICML) Deep Learning workshop*, 2015.

Sun, Y., Chen, Y., Wang, X., and Tang, X. Deep learning face representation by joint identification-verification. In *NeurIPS*, pp. 1988–1996, 2014.

Taigman, Y., Yang, M., Ranzato, M., and Wolf, L. DeepFace: Closing the gap to human-level performance in face verification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1701–1708, 2014.

Tran, L., Yin, X., and Liu, X. Disentangled representation learning GAN for pose-invariant face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1415–1424, 2017.

Tuzel, O. and abd Peter Meer, F. P. Learning on Lie groups for invariant detection and tracking. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8, 2008.

Wang, D., Charles, O., and Anil, K. J. Face search at scale. *IEEE Transactions on Pattern Analysis and Machine Intelligence(T-PAMI)*, 39(6):1122–1136, 2016.

Wang, H., Wang, Y., Zhou, Z., Ji, X., Gong, D., Zhou, J., Li, Z., and Liu, W. CosFace: Large margin cosine loss for deep face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5265–5274, 2018a.

Wang, H., Gong, D., Li, Z., and Liu, W. Decorrelated adversarial learning for age-invariant face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3527–3536, 2019.

Wang, J., Zhang, J., Luo, C., and Chen, F. Joint head pose and facial landmark regression from depth images. *Computational Visual Media*, 3(3):229–241, 2017.

Wang, M. and Deng, W. Deep face recognition with clustering based domain adaptation. *Neurocomputing*, 393(1): 1–14, 2020.

Wang, Q., Wu, T., Zheng, H., and Guo, G. Hierarchical pyramid diverse attention networks for face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8326–8335, 2020.

Wang, Y., Gong, D., Zhou, Z., Ji, X., Wang, H., Li, Z., Liu, W., and Zhang, T. Orthogonal deep features decomposition for age-invariant face recognition. In *European Conference on Computer Vision (ECCV)*, pp. 738–753, 2018b.

Wen, Y., Zhang, K., Li, Z., and Qiao, Y. A discriminative feature learning approach for deep face recognition. In *European Conference on Computer Vision (ECCV)*, pp. 499–515, 2016.

Wilhelm, M. On the exponential solution of differential equations for a linear operator. *Communications on Pure and Applied Mathematics*, 7(4):649–673, 1954.

Wolf, L., Hassner, T., and Maoz, I. Face recognition in unconstrained videos with matched background similarity. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 529–534, 2011.

Wu, Y., Li, J., Kong, Y., and Fu, Y. Deep convolutional neural network with independent softmax for large scale face recognition. In *Proceedings of the 24th ACM International Conference on Multimedia*, pp. 1063–1067, 2016.

Wulf, R. *Lie Groups – An Introduction Through Linear Groups*. Oxford Graduate Texts in Mathematics, Oxford Science Publications, 2002.

Xie, W. and Zisserma, A. Multicolumn networks for face recognition. In *arXiv preprint:1807.09192*, 2018.

Yang, J., Ren, P., Zhang, D., Chen, D., Wen, F., Li, H., and Hua, G. Neural aggregation network for video face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4362–4371, 2017.

Yang, T.-Y., Chen, Y.-T., Lin, Y.-Y., and Chuang, Y.-Y. FSA-Net: Learning fine-grained structure aggregation for head pose estimation from a single image. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1087–1096, 2019.

Yang, X., Jia, X., Yuan, M., and Yan, D.-M. Real-time facial pose estimation and tracking by coarse-to-fine iterative optimization. *Tsinghua Science and Technology*, 25(5): 690–700, 2020.

Yi, D., Lei, Z., Liao, S., and Li, S. Z. Learning face representation from scratch. In *arXiv preprint:1411.7923*, 2014.

Yin, X. and Liu, X. Multi-task convolutional neural network for pose-invariant face recognition. *IEEE Transactions on Image Processing(TIP)*, 27(2):964–975, 2017.

Yin, X., Yu, X., Sohn, K., Liu, X., and Chandraker, M. Towards large-pose face frontalization in the wild. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 3990–3999, 2017.

Yin, X., Yu, X., Sohn, K., Liu, X., and Chandraker, M. Feature transfer learning for face recognition with under-represented data. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5704–5713, 2019.

Zhang, K., Zhang, Z., Li, Z., and Qiao, Y. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10): 1499 – 1503, 2016.

Zhang, W., Shan, S., Chen, X., and Gao, W. Local Gabor binary patterns based on Kullback - Leibler divergence for partially occluded face recognition. *IEEE Signal Processing Letters*, 14(11):875–878, 2007.

Zhang, X., Fang, Z., Wen, Y., Li, Z., and Qiao, Y. Range loss for deep face recognition with long-tailed training data. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 5409–5418, 2017.

Zheng, T. and Deng, W. Cross-pose LFW: A database for studying cross-pose face recognition in unconstrained environments. In *Technical Report of Beijing University of Posts and Telecommunications*, 2018.

Zhou, H., Liu, J., Liu, Z., Liu, Y., and Wang, X. Rotate-and-Render: Unsupervised photorealistic face rotation from single-view images. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5911–5920, 2020.