ALIGNFACE: ENHANCING FACE VERIFICATION MODELS THROUGH ADAPTIVE ALIGNMENT OF POSE, EXPRESSION, AND ILLUMINATION

Sahar Husseini[1,2] *Jean-Luc Dugelay*[2]

[1] Docaposte Biometrics Lab, Sophia Antipolis, France
[2] Department of Digital Security, Eurecom Research Center, Sophia Antipolis, France

ABSTRACT

In the field of face recognition and verification, the practice of face frontalization is conventionally regarded as a standard technique. However, traditional frontalization methods often manipulate original facial images, relying on symmetric cues or data distributions from machine learning model training, which may lead to the distortion of genuine facial features. To tackle these challenges, this paper presents AlignFace, a novel face normalization algorithm specifically designed for preprocessing in the context of face verification. Distinct from existing methods, AlignFace uniquely aligns head pose, expression, and illumination conditions between image pairs. This is achieved by estimating these parameters in one image and reconstructing the other to correspond, all while meticulously preserving each image's distinct identity features. Such an approach not only ensures a more authentic representation of facial characteristics but also maintains the integrity of real features in one of the images. Our extensive experimental evaluations, conducted on benchmark datasets such as LFW, CFP, AgeDB, and IJB-B, underscore the effectiveness of AlignFace. The comparative analysis with existing methods demonstrates its state-of-the-art performance, highlighting substantial advancements in face verification accuracy. For further research and replication, the code for our method is accessible at: https://github.com/SaharHusseini/ALIGNFACE.

***Index Terms*—** Face verification, Normalization, 3DMM

1. INTRODUCTION

Face Verification (FV) has gained significant attention due to its great potential value in practical applications such as access control and video surveillance. Recent progress in face verification heavily depends on the utilization of deep Convolutional Neural Networks (CNNs), consistently showcasing notable accuracy that frequently exceeds human-level performance. In face verification models, where two images are used as input, effectiveness is indeed influenced by several factors, including scene illumination during image capture, camera parameters, image quality, alterations in facial expressions, and changes in the head pose of the subjects. Hence, it is crucial to direct the model's focus exclusively towards distinctive features crucial for individual recognition while neglecting extraneous elements. For this purpose, diverse strategies have been investigated, falling into two primary categories: incorporating image quality-related factors, such as head pose and illumination, into the loss function [1]. The other approach focuses on the implementation of preprocessing techniques to normalize elements such as head pose and expression [2], and to address variations in illumination [3].

Recent advancements, focusing on enhancing performance through improved loss functions, often involve the incorporation of margin-based loss functions [4], with the primary objective of minimizing intra-class variation and maximizing inter-class distinction. A widely adopted margin-based loss function is ArcFace [5], which introduces an angular margin term into the standard softmax classification loss, significantly enhancing class separability. Nevertheless, recent investigations have pointed out that ArcFace exhibits a degree of quality-agnostic behavior, leading to instability in within-class distributions [1]. To address these challenges and improve performance, AdaFace [4] integrates image quality information into the loss function.

In parallel, another crucial technique for improving face verification is face normalization. This involves synthesizing and transforming a face with arbitrary pose, illumination, and expression into a desired pose, balanced illumination, and neutral expression to enhance recognition. Through the normalization of images to a shared representation, the model is enabled to concentrate its discriminative capacity on the intrinsic characteristics of individuals, thereby fostering more reliable and accurate face verification outcomes.

Normalization of face pose is widely adopted in the field, typically with the desired pose specified as frontal [6]. In [7], a combination of a 3D Morphable Model (3DMM) and a Generative Adversarial Network (GAN) is employed to generate frontal face images from input profile images. Likewise, in [2], face frontalization is accomplished entirely through a Generative Adversarial Network (GAN). The DVN [8] utilizes two layers of dual-view generators to normalize a face in dual views - one in frontal view and the other in a yaw 45° side view. MVN [6] is designed to learn the transformation from an input set to seven output sets, encompassing seven face poses from 0° to 90° in yaw with a 15° interval, utilizing

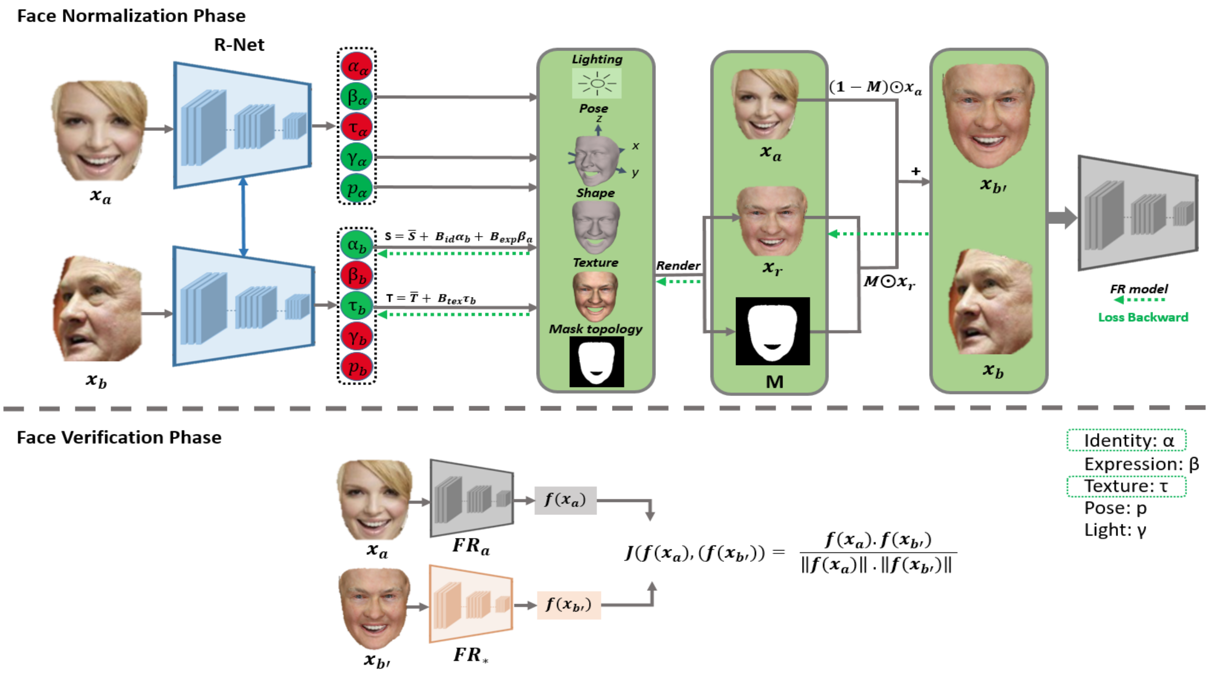

Fig. 1: **Overview of Face Normalization by AlignFace:** For each image, 3DMM coefficients, including identity (α), expression (β), texture (τ), illumination (γ), and head pose (p) are extracted using the R-Net model. The image x_b is then normalized to produce $x_{b'}$, aligning its expression, head pose, and lighting conditions with those of x_a. During normalization process $x_{b'}$'s identity and texture coefficients are iteratively updated (n iterations) while keeping the parameters of the FR and R-Net models frozen. Although images generated as $x_{b'}$ closely follow the distribution of real images, discrepancies might exist between the distributions of generated $x_{b'}$ and real faces x_a. To ensure accuracy at the face verification phase, the FR model used for extracting face embeddings for $x_{b'}$ is fine-tuned, denoted as FR_*. x_a and $x_{b'}$ represent different identities.

seven generators. However, transferring faces to specific head poses is not always advisable due to several reasons:

- **Training Data Distribution:** The majority of the training data may not be centered within the frontal pose range and could be distributed across various angles. As illustrated in the DVN [8] framework, the face encoder exhibits greater expertise with faces within a 45° range, reflecting the predominant distribution of training data in their database. Therefore, to ensure effective of face verification in diverse scenarios, it is essential not to exclusively rely on normalization at specific poses, as optimal results may vary.

- **Photo-Realism and Texture Loss:** Generated frontalized (or at any other specific degree) face images from GANs may lack photo-realism and exhibit artifacts and texture loss, especially in occluded regions. Counterfeiting features in synthetic generated images may degrade recognition performance. For example, if a particular facial feature, such as a birthmark or mole, is obscured in the original image and remains ungenerated by the GAN model during frontalization, while being visible in the second image, the face verification model may incorrectly categorize these two images as representing different identities.

The reasons mentioned above could also be applicable to the normalization of illumination, expression, and other extraneous elements.

In this paper, we introduce an innovative normalization algorithm designed for preprocessing input images in the context of face verification. Diverging from conventional methods, our approach places a distinctive emphasis on achieving consistency in head pose, expression, and illumination conditions between two images, avoiding an exclusive focus on the normalization of extraneous elements at specific values. Specifically, our methodology involves estimating the head pose, expression, and illumination conditions in one image, followed by the reconstruction of the second image to align with the same head pose, expression, and illumination conditions while preserving its own unique identity features. This ensures the constancy of real features in one of the images, providing a more authentic representation of the facial distribution. By adopting this approach, our algorithm allows the verification process to concentrate solely on identity evaluation, unaffected by variations in non-essential extraneous and synthesized features. This refined focus contributes to a more accurate and reliable assessment of facial identity in face verification scenarios.

This paper is organized as follows: Section 2 offers an overview of related work. In Section 3, we present our proposed face normalization algorithm. Section 4 covers the experimental setup and the results. Finally, Section 5 summarizes the conclusion of our study.

2. RELATED WORK

In recent years, deep neural networks have shown notable success in face verification. The State-of-The-Art (SoTA) methods aim to map each face image to a latent space representation that is closely associated with the individual's identity, clustering representations of the same person together. Challenges arise when face images contain uncertainty, making learned representations unreliable and error-prone. Moreover, variations in extraneous elements between image pairs can lead to the loss of crucial identity information, resulting in images that cannot be identified.

Early research investigated into various loss functions, including contrastive loss [9], triplet loss [10]. However, a notable transformation has taken place more recently, as researchers have shifted their focus towards optimizing loss functions to reduce the demand for extensive training data. Central to these innovative methods is the adoption of margin-based softmax loss functions for training Face Recognition (FR) models. The incorporation of a margin is crucial in these loss functions, as it empowers the learned features to become more discerning and discriminative. Pioneering contributions to this field include SphereFace [11], CosFace [12], and ArcFace [5], each introducing distinct variations of margin functions. However, these loss functions share a common limitation: they rely on fixed margin values that do not account for inherent variations, such as differences in image quality, within the same class. This limitation has prompted the development of solutions based on adaptive margin loss. MagFace [1] incorporates the quality of a face image sample, into the margin calculation which aims to concentrate high-quality samples in a compact region around the class centers and the low quality samples further from the class center. This approach helps prevent the algorithm from overly emphasizing noisy or difficult samples, which could otherwise compromise its effectiveness and lead to overfitting.

In addition to refining loss functions, numerous studies have focused on optimizing input images before they are fed into the FV model or subjected to feature extraction. A crucial preprocessing step in this regard is **face normalization**, which addresses various aspects, including illumination, expression, and head pose normalization. Illumination normalization seeks to reduce the impact of lighting conditions on facial appearance, ensuring that the texture and color of the face remain consistent [3, 13, 14]. On the other hand face frontalization is aimed at transforming facial images into a frontal view, even in the presence of potential occlusions. In recent years, deep learning-based solutions [2, 7, 15] have addressed both face frontalization and neutralization of facial expressions, leveraging the capabilities of neural networks. Despite showcasing promising synthesis quality, these methods encounter challenges in preserving face identity details, especially in scenarios with substantial pose variations.

R-Net: In this paper, we employ R-Net [16], a CNN-based model, to perform 3D face reconstruction from a single image. This model is trained using a hybrid-level loss function that seamlessly integrates both low-level and perception-level information, enhancing its reconstruction capabilities. The model's strength lies in its robustness in handling challenges such as occlusion and extreme poses. It achieves this robustness by incorporating a skin color-based photometric error attention strategy, making it adaptable to scenarios with occlusions and other intricate appearance variations, such as beards and heavy makeup. The backbone of this model is the ResNet-50 network, which plays a crucial role in regressing the 3D Morphable Model (3DMM) coefficients required for accurate 3D reconstruction.

3. METHODOLOGY

In face verification, a pair of images $\{x_a, x_b\} \subset X$ is examined using a face recognition model denoted as $f(x) : X \to \mathbb{R}^d$. This model extracts feature embeddings from the faces in the images, placing them in the $\mathbb{R}^d$ space. The similarity between a pair of images can be commonly calculated using the cosine similarity formula:

$$J(f(x_a), f(x_b)) = \frac{f(x_a) \cdot f(x_b)}{\|f(x_a)\| \cdot \|f(x_b)\|} \tag{1}$$

where $\langle \cdot, \cdot \rangle$ represents the inner product of the vectors. The function J denotes the cosine similarity between the feature embeddings of x_a and x_b, with values ranging from 0 to 1. The prediction for face verification is formulated as:

$$C(x_a, x_b) = \begin{cases} 1 & \text{if } J(f(x_a), f(x_b)) \geq \delta \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Here, δ represents the threshold. When $C(x_a, x_b)$ equals 1, the two images are considered to depict the same identity; otherwise, they represent different identities.

3.1. 3D Face Model reconstruction

Given a facial image, denoted as x, R-Net model [16] is employed to regress the 3D Morphable Model (3DMM) coefficients denoted as $\alpha \in \mathbb{R}^{80}$, $\beta \in \mathbb{R}^{64}$, and $\tau \in \mathbb{R}^{80}$ corresponding to the image x. Once these coefficients are obtained, the 3D face shape (S) and texture (T) can be represented by an affine model:

$$\begin{aligned} S &= S(\alpha, \beta) = \bar{S} + B_{id}\alpha + B_{exp}\beta \\ T &= T(\delta) = \bar{T} + B_{tex}\tau \end{aligned} \tag{3}$$

where $\bar{S}$ and $\bar{T}$ denote the averages of face shape and texture, while B_{id}, B_{tex}, and B_{exp} represent the Principal Component Analysis (PCA) bases of identity, texture, and expression, respectively. The values of $\bar{S}$, $\bar{T}$, B_{id}, and B_{tex} are derived from the well-established 2009 Basel Face Model [17] and the expression bases B_{exp} are sourced from [18], which constructed using data from Face-Warehouse [19]. Furthermore, the R-Net model regresses the illumination coefficients $\gamma \in \mathbb{R}^9$, and the head pose $p \in \mathbb{R}^6$.

With access to both the facial texture and shape, we are able to represent the complete 3D mesh model of the face as $M_{sh} = (S, T)$, where $S \in \mathbb{R}^{n\times 3}$ represents the XYZ coordinates of n vertices, and $T \in \mathbb{R}^{n\times 3}$ corresponds to the RGB values of these vertices [20].

3.2. Proposed Method

The face verification system operates on a pair of facial images as its input. In environments without constraints, these images may exhibit variations in head pose, facial expressions, and lighting conditions, thereby significantly impacting the system's performance. To effectively tackle this challenge, we propose a dedicated pipeline designed specifically for face verification, as illustrated in Figure 1. Our primary objective is to normalize one of the faces within the pair, ensuring alignment in terms of head pose, expression, and illumination. This normalization process optimizes the system's workload, enabling it to concentrate exclusively on identity verification. Specifically, given an image pair, our methodology entails selecting one image, denoted as x_a, which possesses a head pose closest to the frontal pose, to serve as the reference. Subsequently, the second image, x_b, undergoes normalization to become $x_{b'}$, aligning its expression, head pose, and lighting condition with those of the original image x_a. To achieve this, we follow these steps:

- Utilize the R-Net to extract 3DMM coefficients for both provided images. As face verification models exhibit more sensitivity to pose variations than to scene illumination and facial expression [7], we specifically focus on the head pose coefficient p. An image with the closest deviation from the frontal pose is denoted as x_a, while the second image x_b undergoes normalization. The coefficients for these image pairs are as follows:

$$x_a : \{\alpha_a, \beta_a, \tau_a, \gamma_a, p_a\}, x_b : \{\alpha_b, \beta_b, \tau_b, \gamma_b, p_b\}$$

- To reconstruct the 3D face model x_b with the same illumination, head pose, and expression as x_a, we initialize the 3D mesh model using the following coefficients:

$$\{\alpha^{[0]}_{b'}, \beta^{[0]}_{b'}, \tau^{[0]}_{b'}, \gamma^{[0]}_{b'}, p^{[0]}_{b'}\} \leftarrow \{\alpha_b, \beta_a, \tau_b, \gamma_a, p_a\}$$

- Since the regressed 3DMM coefficients are all differentiable, we employ $L_f = -J(f(x_{b'}), f(x_b))$ as loss function and update coefficients $\alpha_{b'}$, $\tau_{b'}$ for n iterations. The objective function can be expressed as:

$$\min_{\alpha_{b'}, \tau_{b'}} L_f(x_{b'}, x_b) \quad (4)$$

where,

$$x_{b'} = M \odot x_r + (1 - M) \odot x_a \quad (5)$$

and x_r, M are computed using the rendering function $R(\bar{S} + B_{id}\alpha_{b'} + B_{exp}\beta_a, \bar{T} + B_{tex}\tau_{b'}, \gamma_a, p_a)$. The symbol $\odot$ denotes element-wise multiplication, and R represents the rendering function, which takes into consideration factors such as camera position and illumination. The variable M signifies the binary mask used in this process.

- After successfully reconstructing $x_{b'}$, the next step involves conducting face verification between the image pair x_a and $x_{b'}$ using a trained FR model as a feature extractor and computing the cosine distance between their feature vectors. Models such as ArcFace [5], MagFace [1], or AdaFace [4] can be employed for this purpose, considering that our algorithm serves as a face normalization tool.

- The images generated as $x_{b'}$ follow the distribution of real images; however, FR models are typically trained on real datasets, and a potential discrepancy may exist between the distributions of generated $x_{b'}$ and real faces x_a. To ensure result precision, we created a training dataset normalized by our proposed normalization tool. Subsequently, we fine-tuned the selected FR model with the generated $x_{b'}$ data, denoted as FR_*.

- After completing the fine-tuning process, the face recognition models FR_a and FR_* are employed to extract feature embeddings from x_a and $x_{b'}$, followed by computing the cosine similarity distance between these embeddings.

4. EXPERIMENTAL EVALUATION

4.1. Datasets

In our experiment, the MS1M-V2 dataset [5], containing 5.8 million images and 85,000 identities, was utilized to fine-tune face recognition model. For the evaluation purposes, we selected four widely recognized unconstrained face verification benchmarks, namely Labeled Faces in the Wild (LFW) [21], Celebrities in Frontal-Profile (CFP) [22], AgeDB [23], and IARPA Janus Benchmark-B (IJB-B) [24] dataset. The LFW dataset comprises 13,233 facial images from 5,749 individuals, showcasing various poses, facial expressions, and lighting conditions. The CFP dataset, with 7,000 facial images, emphasizes extreme head poses, such as profiles, leading to significant occlusion. AgeDB, consisting of 16,516 images, focuses on age-related variations. The IJB-B dataset features 21.8K still images and 55K frames from 7,011 videos, representing 1,845 subjects with diverse qualities. All images are resized to 112×112 dimensions before the verification step.

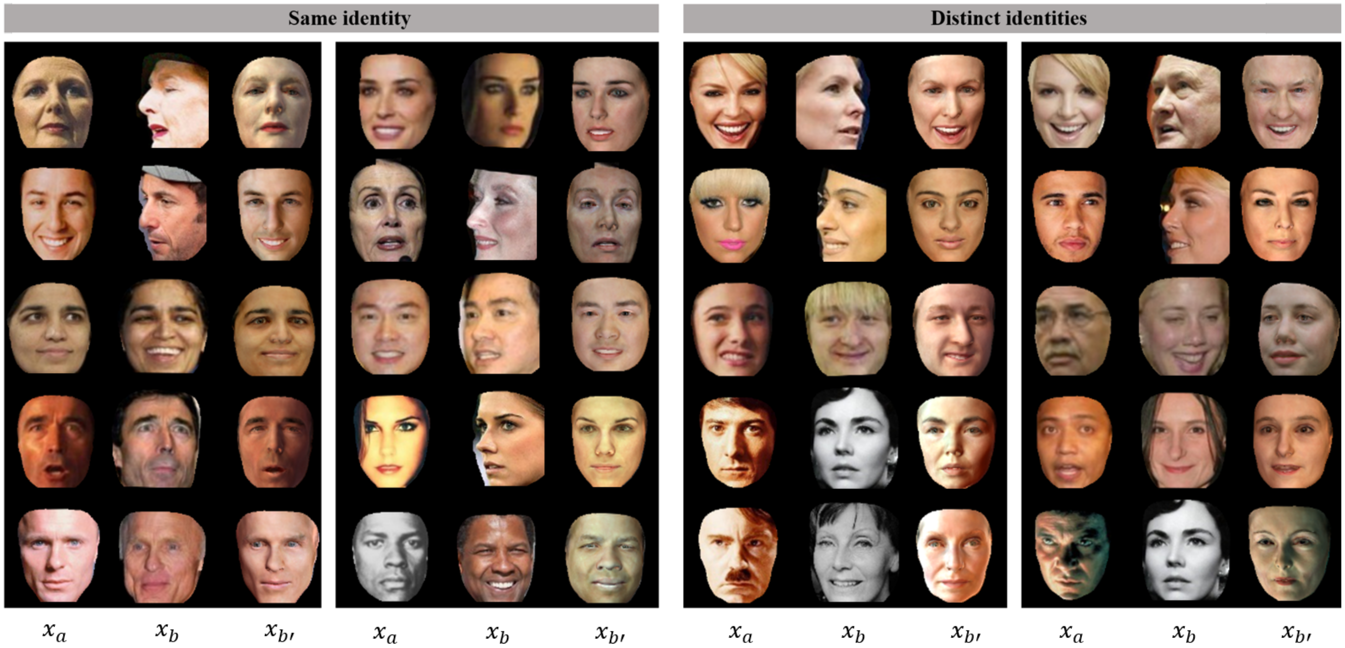

Fig. 2: **AlignFace Efficacy in Normalizing Pose, Expression, and Illumination**. Displayed are the original image x_a, the comparative image x_b, and AlignFace's reconstructed image $x_{b'}$. This demonstrates AlignFace's ability to effectively transfer the extraneous conditions of x_a to x_b while preserving the unique identity features in x_b. Note that in the examples on the left, the identities are the same, whereas in the examples on the right, the identities are different.

4.2. Face normalization

We employ a PyTorch implementation of R-Net [16] to acquire the 3DMM coefficients for image pairs. Within our pipeline, the FR encoder model is utilized to extract feature embeddings from both x_b and $x_{b'}$. It's important to note that the FR model is pretrained and fixed under the normalization framework. Since the entire pipeline, including the rendering procedure, is differentiable, $x_{b'}$ can be iteratively updated through backpropagation on the low-dimensional identity (α) and texture (β) coefficients of the 3DMM. We set the number of iterations to $N = 300$, the learning rate to $\alpha = 1.5$, and the decay factor to $\mu = 1$. This iterative process results in the reconstruction of x_b, aligning its expression, head pose, and lighting conditions with those of the image x_a while preserving its unique identity features.

4.3. Face verification models

In our experiments, we benchmark and utilize the encoders of two SoTA face recognition models: MagFace [1] and AdaFace [4], to serve as facial feature extractors. We employed the official implementations of MagFace and AdaFace, both utilizing ResNet100 backbones trained on the MS1M-V2 dataset. The encoder used for x_a feature extraction does not require fine-tuning. However, since a potential discrepancy may exist between the distributions of generated $x_{b'}$ and real faces, on which the original face recognition models are trained, we ensure result precision by creating a training dataset. This dataset, normalized using our proposed normalization tool and derived from MS1M-V2, serves as the basis for fine-tuning the selected FR model. The fine-tuned model, denoted as FR_* in Figure 1, is trained with the generated $x_{b'}$ datatype. The fine-tuning process follow the same parameters and instructions specified in the official implementation.

4.4. Comparisons with state-of-the-art methods

To assess the efficacy of our proposed method, we conducted a comprehensive comparative analysis with SoTA methods. The results, encompassing 1:1 verification accuracy for LFW, CFP, and AgeDB, as well as TAR@FAR=0.01% for the IJB-B dataset, are showcased in Table 1. Notably, all models featured in this table were trained utilizing the MS1M-V2 dataset and the ResNet100 backbone. In our evaluation, we incorporated our novel normalization method as a preprocessing step for two specific models: AdaFace and MagFace. These modified models are referred to as "AlignFace+MagFace" and "AlignFace+AdaFace," respectively. The results presented in Table 1 for the LFW, CFP, and AgeDB datasets demonstrate that, although face verification performance is approaching saturation on these benchmarks, our proposed enhancements have yielded significant improvements. However, this increased accuracy results in higher processing times and resource consumption.

Table 1: **Comparative analysis on benchmark datasets:** Accuracy metrics for 1:1 verification are presented for LFW, CFP, and AgeDB datasets. For the IJB-B dataset, we report the TAR@FAR=0.01%. Red: best, blue: second-best.

Method	Dataset				
	LFW [21]	CFP [22]	AgeDB [23]	AVG	IJB-B [24]
LFW CosFace [12]	99.81	98.12	98.11	98.68	94.80
ArcFace [5]	99.83	98.27	98.28	98.79	94.25
MV-Softmax [25]	99.80	98.28	97.95	98.68	93.60
MagFace [1]	99.83	98.46	98.17	98.82	94.51
AdaFace [4]	99.82	98.49	98.05	98.79	95.67
R-Net α coefficient	92.76	84.65	87.25	86.22	87.13
R-Net α coefficient after normalization	97.46	95.32	94.11	95.63	93.46
AlignFace+MagFace	99.82	98.73	98.33	99.29	94.46
AlignFace+AdaFace	99.82	98.85	97.95	98.87	96.02

In particular, on the CFP benchmark, the incorporation of our normalization technique with MagFace (denoted as AlignFace+MagFace) led to an improvement in performance by 0.24% in accuracyn compared to the previous best method. Additionally, when combined with AdaFace (AlignFace+AdaFace), there was a further increase of 0.36% in accuracy, thereby exceeding the capabilities of the previously established best-performing method. This improvement can be attributed to the distinct advantages of our normalization method in minimizing head pose differences between image pairs. This is particularly significant in the CFP dataset, which comprises images with both frontal and profile head poses.

The results from the IJB-B dataset indicate that the integration of AlignFace with AdaFace (AlignFace+AdaFace) yields a 0.36% increase in performance compared to using AdaFace alone. The IJB-B dataset is specifically designed to incorporate low-quality images within its validation protocol. The improvement underscores our algorithm's robustness with varying image qualities. Additionally, the average values (AVG) presented in Table 1 indicate that the accuracy for the LFW, CFP, and AgeDB datasets generally improves when our normalization method is incorporated, further validating the efficacy of our approach. Figure 2 highlights the efficacy of the proposed method in normalizing faces in scenarios with variations in pose, expression, and illumination between input pairs.

4.5. Ablation Study

Assessment of Verification Accuracy via 3DMM Identity Coefficients: To assess verification accuracy using 3DMM identity coefficients, we conducted an ablation analysis in this study. We compared the identity coefficients directly extracted from the input pairs using the R-Net, labeled as 'R-Net α coefficient' in Table 1, with those coefficients post-normalization, termed 'R-Net α coefficients after normalization'. Initial results indicated that the verification accuracy with 'R-Net α coefficients' was lower than that of SoTA methods. Nevertheless, upon updating these coefficients to derive $x_{b'}$ ('R-Net α coefficients after normalization'), a significant improvement was observed. It is crucial to note that even with the enhanced coefficients from the R-Net post-normalization, the verification accuracies did not exceed those of SoTA methods. The further improvement was observed only after processing the normalized faces, utilizing optimized identity coefficients, through the FR model for feature embedding. This advancement can be attributed to the training of FR models (such as AdaFace) and the evolution of margin-based loss functions, which have markedly increased the discriminative power of face embeddings.

5. CONCLUSION

Our proposed solution, AlignFace, introduces a novel approach to face normalization, specifically designed for preprocessing input images within the context of face verification. AlignFace stands out by focusing on aligning head pose, expression, and illumination conditions between two images. It proficiently estimates these conditions in one image and reconstructs the other to match, while carefully preserving the unique identity features of each image. This method ensures the preservation of genuine facial features in one image, providing a more accurate representation of facial characteristics. Our experimental results underscore AlignFace's superiority over existing state-of-the-art methods in face verification across multiple benchmark datasets.

Acknowledgment

This research was supported by Docaposte for Biometrics Research. Thanks to Fabien Aili and Emmanuel Nars for their valuable contributions.

6. REFERENCES

[1] Qiang Meng, Shichao Zhao, Zhida Huang, and Feng Zhou, "Magface: A universal representation for face recognition and

quality assessment," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 14225–14234.

[2] Yichen Qian, Weihong Deng, and Jiani Hu, "Unsupervised face normalization with extreme pose and expression in the wild," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9851–9858.

[3] Mannan Hassan, Muhammad Suhail Shaikh, and Munsif Ali Jatoi, "Image quality measurement-based comparative analysis of illumination compensation methods for face image normalization," *Multimedia Systems*, pp. 1–10, 2022.

[4] Minchul Kim, Anil K Jain, and Xiaoming Liu, "Adaface: Quality adaptive margin for face recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 18750–18759.

[5] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4690–4699.

[6] Chia-Hao Tang, Yi-Mei Chou, and Gee-Sern Jison Hsu, "Multi-view normalization for face recognition," in *2021 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2021, pp. 2343–2347.

[7] Hao Wu, Jianyang Gu, Xiaojin Fan, He Li, Lidong Xie, and Jian Zhao, "3d-guided frontal face generation for pose-invariant recognition," *ACM Transactions on Intelligent Systems and Technology*, vol. 14, no. 2, pp. 1–21, 2023.

[8] Gee-Sern Hsu and Chia-Hao Tang, "Dual-view normalization for face recognition," *IEEE Access*, vol. 8, pp. 147765–147775, 2020.

[9] Sumit Chopra, Raia Hadsell, and Yann LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*. IEEE, 2005, vol. 1, pp. 539–546.

[10] Florian Schroff, Dmitry Kalenichenko, and James Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823.

[11] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song, "Sphereface: Deep hypersphere embedding for face recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 212–220.

[12] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu, "Cosface: Large margin cosine loss for deep face recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 5265–5274.

[13] Sahil Dalal and Virendra P Vishwakarma, "A novel approach of face recognition using optimized adaptive illumination–normalization and kelm," *Arabian Journal for Science and Engineering*, vol. 45, pp. 9977–9996, 2020.

[14] Sahar Husseini, Pouria Babahajiani, and Moncef Gabbouj, "Color constancy model optimization with small dataset via pruning of cnn filters," in *2021 9th European Workshop on Visual Information Processing (EUVIP)*, 2021, pp. 1–6.

[15] Sahar Husseini and Jean-Luc Dugelay, "A comprehensive framework for evaluating deepfake generators: Dataset, metrics performance, and comparative analysis," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 372–381.

[16] Yu Deng, Jiaolong Yang, Sicheng Xu, Dong Chen, Yunde Jia, and Xin Tong, "Accurate 3d face reconstruction with weakly-supervised learning: From single image to image set," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2019, pp. 0–0.

[17] Pascal Paysan, Reinhard Knothe, Brian Amberg, Sami Romdhani, and Thomas Vetter, "A 3d face model for pose and illumination invariant face recognition," in *2009 sixth IEEE international conference on advanced video and signal based surveillance*. Ieee, 2009, pp. 296–301.

[18] Yudong Guo, Jianfei Cai, Boyi Jiang, Jianmin Zheng, et al., "Cnn-based real-time dense face reconstruction with inverse-rendered photo-realistic face images," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 6, pp. 1294–1307, 2018.

[19] Chen Cao, Yanlin Weng, Shun Zhou, Yiying Tong, and Kun Zhou, "Facewarehouse: A 3d facial expression database for visual computing," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 3, pp. 413–425, 2013.

[20] Xiao Yang, Chang Liu, Longlong Xu, Yikai Wang, Yinpeng Dong, Ning Chen, Hang Su, and Jun Zhu, "Towards effective adversarial textured 3d meshes on physical face recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 4119–4128.

[21] Gary B Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller, "Labeled faces in the wild: A database forstudying face recognition in unconstrained environments," in *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*, 2008.

[22] Soumyadip Sengupta, Jun-Cheng Chen, Carlos Castillo, Vishal M Patel, Rama Chellappa, and David W Jacobs, "Frontal to profile face verification in the wild," in *2016 IEEE winter conference on applications of computer vision (WACV)*. IEEE, 2016, pp. 1–9.

[23] Stylianos Moschoglou, Athanasios Papaioannou, Christos Sagonas, Jiankang Deng, Irene Kotsia, and Stefanos Zafeiriou, "Agedb: the first manually collected, in-the-wild age database," in *proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 51–59.

[24] Cameron Whitelam, Emma Taborsky, Austin Blanton, Brianna Maze, Jocelyn Adams, Tim Miller, Nathan Kalka, Anil K Jain, James A Duncan, Kristen Allen, et al., "Iarpa janus benchmark-b face dataset," in *proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 90–98.

[25] Xiaobo Wang, Shifeng Zhang, Shuo Wang, Tianyu Fu, Hailin Shi, and Tao Mei, "Mis-classified vector guided softmax loss for face recognition," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, vol. 34, pp. 12241–12248.