An Illumination Augmentation Approach for Robust Face Recognition

Zhanxiang Feng[1], Xiaohua Xie[2,3], Jianhuang Lai[2,3(✉)], and Rui Huang[2]

[1] School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou, China
fengzhx@mail2.sysu.edu.cn

[2] School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China
{xiexiaoh6,stsljh}@mail.sysu.edu.cn

[3] Guangdong Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, Guangzhou, China

Abstract. Deep learning has achieved great success in face recognition and significantly improved the performance of the existing face recognition systems. However, the performance of deep network-based methods degrades dramatically when the training data is insufficient to cover the intra-class variations, e.g., illumination. To solve this problem, we propose an illumination augmentation approach to augment the training set by constructing new training images with additional illumination components. The proposed approach first utilizes an external benchmark to generate several illumination templates. Then we combine the generated templates with the training images to simulate different illumination conditions. Finally, we conduct color correction by using the singular value decomposition (SVD) algorithm to confirm that the color of the augmented image is consistent with the input image. Experimental results demonstrate that the proposed illumination augmentation approach is effective for improving the performance of the existing deep networks.

Keywords: Face recognition · Deep learning
Illumination augmentation

1 Introduction

Face recognition has been a hot research topic in the past decades and attracted considerable research attention. In recent years, with the development of deep learning techniques and the emergences of large-scale face datasets, deep network-based methods have significantly advanced face recognition techniques [1–3]. Applications of face recognition are emerging in video surveillance, social security, company attendance, and identity authentication.

Although deep models have proved their effectiveness in improving the performance of face recognition techniques, most of the existing face recognition systems are trained with large-scale datasets. In some applications, each person

J. Zhou et al. (Eds.): CCBR 2018, LNCS 10996, pp. 409–417, 2018.
https://doi.org/10.1007/978-3-319-97909-0_44

contains only 1–2 samples, and the training data may be inadequate to cover the changing illumination, pose, and image quality. Particularly, the performance of the existing deep networks will decrease dramatically when dealing with extreme lighting condition. Therefore, the topic of learning robust deep representations with insufficient training samples will be worthwhile for face recognition.

A natural idea is to augment the training data and generate additional training samples. Some recent works have focused on synthesizing novel face images with changing poses, attributes, and identities by GAN (generative adversarial network) [4–6]. Learning deep networks with the synthesized face images improves the performance of deep networks to some specific problem. However, the generalizability of GAN-based approaches to other datasets is under study. Besides, controlling the facial details and ID of the generated image by GAN is extremely difficult. Furthermore, the training process of the GAN models is time-consuming, and the labeling of face image attributes is costly.

In this paper, we focus on improving the performance of face recognition systems using the data augmentation technique. We propose to perform illumination augmentation to increase the diversity of the training data. Firstly, we generate different reference illumination templates from other datasets. For each training sample, our approach simulates different illumination conditions using the pre-defined illumination templates. Eventually, we utilize the singular value decomposition (SVD) algorithm to transform the output image to the color subspace which is consistent with the input image. Furthermore, we construct a new dataset by collecting images stored in the second-generation ID cards and images captured in the realistic surveillance environment. We also build a testing set which comprises of images captured in the railway station. Experiments demonstrate that the proposed illumination augmentation approach is effective for improving the performance of deep network-based face recognition models.

2 Related Works

Deep networks have achieved remarkable success in face recognition and dramatically improved the performance of the state-of-the-art methods [1–3]. Taigman et al. [1] proposed a pioneer CNN model named DeepFace which outperformed traditional face recognition methods and closely approached human-level performance. Sun et al. [2] proposed a DeepID network which employed identification and verification supervisory signals to improve the recognition performance. Schroff et al. [3] proposed a network named FaceNet which adopted triplet loss to enforce a margin between distances of intra-class samples and those of inter-class samples.

3 Proposed Method

3.1 Overall Framework

Figure 1 demonstrates the overall framework of the proposed illumination augmentation approach. We first perform Gaussian filtering on the reference images

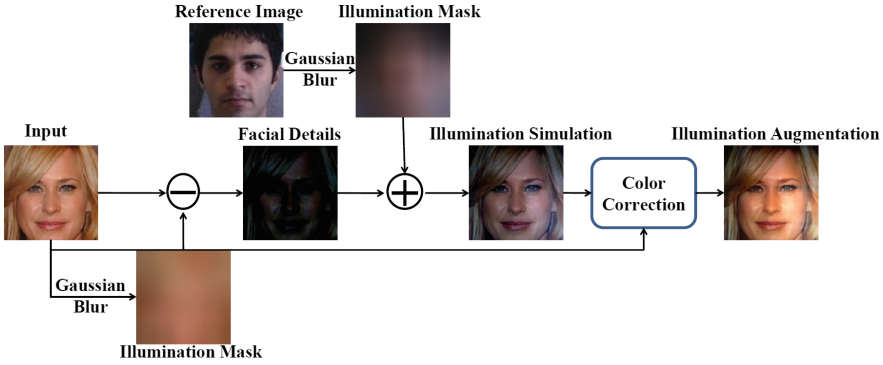

Fig. 1. Proposed framework. We first simulate different illumination situations using the reference images from an external dataset and then perform color correction to obtain the illumination augmentation output.

from an external benchmark to extract the reference illumination masks. Then, we extract the facial details of the input image by subtracting the illumination mask of the input. After that, we combine the facial details of the input image with the reference illumination masks to generate face images with different illumination situations. Eventually, we perform color correction to make sure that the color components of the augmented image is consistent with the input image.

3.2 Illumination Variation Simulation

We perform Gaussian filtering with a large blur kernel on the reference image and input image to extract the corresponding illumination mask. Denote the input image and the reference image as $\boldsymbol{X}_i$ and $\boldsymbol{X}_r$, we can compute the illumination masks $\boldsymbol{X}_i^m$ and $\boldsymbol{X}_r^m$ as follows:

$$\begin{aligned} \boldsymbol{X}_i^m &= \boldsymbol{X}_i * \mathcal{G}, \\ \boldsymbol{X}_r^m &= \boldsymbol{X}_r * \mathcal{G}, \end{aligned} \tag{1}$$

where $\mathcal{G}$ denotes the function of Gaussian filtering and can be defined as follows:

$$\mathcal{G} = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}. \tag{2}$$

We can obtain the facial details of the input image by subtracting the illumination mask. Denote $\boldsymbol{X}_i^d$ as the facial details, the computation is as follows:

$$\boldsymbol{X}_i^d = \boldsymbol{X}_i - \boldsymbol{X}_i^m. \tag{3}$$

Then we combine the facial details with the reference illumination mask to simulate different illumination conditions $\boldsymbol{X}_i^v$, which is formulated as follows:

$$\boldsymbol{X}_i^v = \boldsymbol{X}_i^d + \boldsymbol{X}_r^m. \tag{4}$$

3.3 Color Correction

The color components of the simulated image may be different to the input image. We propose to conduct color correction to ensure that the color components of the final output is consistent with that of the input image. We first perform SVD [11] algorithm on each channel of both the input image and the simulated output to extract their color components. Denote $\boldsymbol{X}_{iA}, A = \{R, G, B\}$ and $\boldsymbol{X}^v_{iA}, A = \{R, G, B\}$ as input and simulated image, we have:

$$\begin{aligned} \boldsymbol{X}_{iA} &= U_{iA}\Sigma_{iA}V_{iA}, A = \{R, G, B\}, \\ \boldsymbol{X}^v_{iA} &= U^v_{iA}\Sigma^v_{iA}V^v_{iA}, A = \{R, G, B\}. \end{aligned} \tag{5}$$

Note that Σ_{iA} and Σ^v_{iA} contains the color components of the input and simulated image, we can correct the color condition of the simulated image according to the input image by replacing Σ^v_{iA} with Σ_{iA}. Denote $\boldsymbol{X}_{oA}$ as the augmented output, then we have:

$$\boldsymbol{X}_{oA} = U^s_{iA}\Sigma_{iA}V^s_{iA}, A = \{R, G, B\} \tag{6}$$

4 Experiment

4.1 Experimental Settings

Training Set. We utilize CASIA-WebFace [7], a popular public face dataset, to train the baseline model. CASIA-WebFace contains 494,414 samples of 10,575 subjects detected from the Internet.

We also construct a domestic dataset for training a stronger model for the domestic face recognition. The domestic training dataset contains 864,652 samples of 386,847 subjects. Most of the subjects contain only 2–3 images, of which one image is from the second-generation ID cards and other images are from the surveillance videos. Training a robust model for the domestic dataset is challenging because of the lack of training sample for each person.

Testing Set. We evaluate the performance of the proposed illumination augmentation approach on the LFW dataset [8]. The LFW dataset contains 13,233 images of 5,749 subjects captured in the unconstrained environment. The LFW dataset is now the most popular benchmark for face recognition. We adopt the standard verification protocol to conduct fair comparison with other methods.

We also construct a domestic testing set to evaluate the performance of the face recognition models under realistic surveillance environment. The domestic testing dataset contains 3,722 prob images of 39 subjects captured in a railway station. The challenges include illumination, pose, and occlusion. For testing, we conduct matching between the domestic testing dataset and a gallery set comprised of 10,039 images captured in the second-generation ID cards.

Implementation Details. We select 20 images from the CMU-PIE [9] dataset to generate reference illumination templates. For each training sample, we randomly select 2 reference templates and obtain 2 illumination augmentation

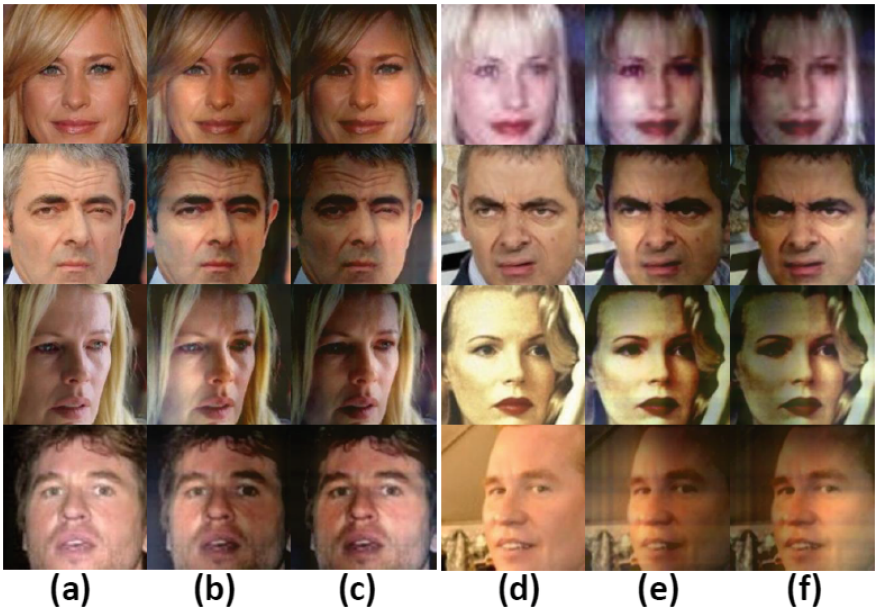

Fig. 2. Illumination augmentation results on CASIA-WebFace. (b) and (c) are the augmented images of (a), while (e) and (f) are the augmented images of (d).

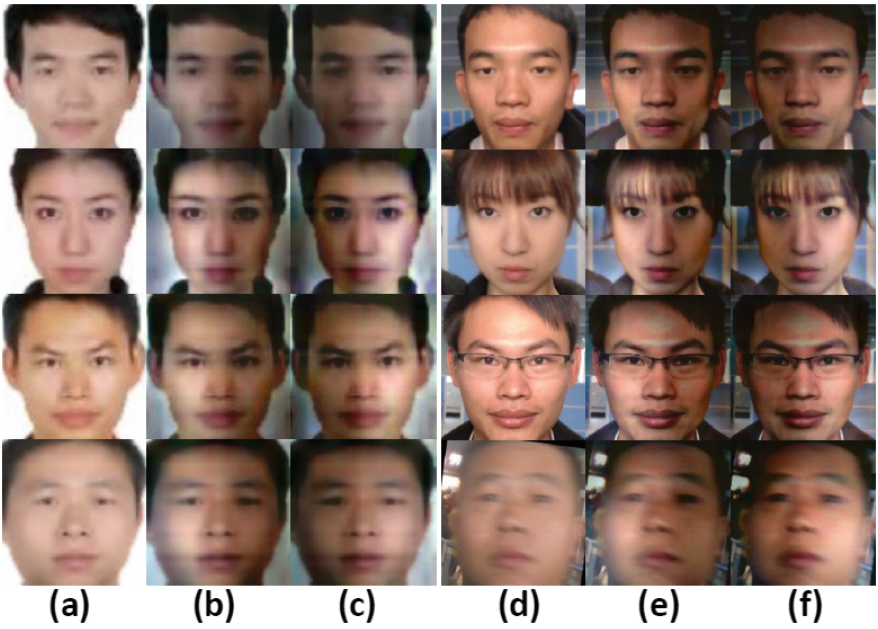

Fig. 3. Illumination augmentation results on the domestic training set. (b) and (c) are the augmented images of (a), while (e) and (f) are the augmented images of (d).

Table 1. The DenseNet structure

Layer name	Input size	Output size	Parameters
Conv1a	112 × 96 × 3	112 × 96 × 64	3 × 3 × 64 conv
Conv1b	112 × 96 × 64	112 × 96 × 128	3 × 3 × 128 conv
Pool1	112 × 96 × 128	56 × 48 × 128	2 × 2 max pooling, stride 2
Dense Block 1	56 × 48 × 128	56 × 48 × 320	$\begin{bmatrix} 1 \times 1\, conv \\ 3 \times 3\, conv \end{bmatrix} \times 3$
Transition 1	56 × 48 × 320	56 × 48 × 128	1 ×1 × 128 conv
Pool2	56 × 48 × 128	28 × 24 × 128	2 × 2 max pooling, stride 2
Dense Block 2	28 × 24 × 128	28 × 24 × 384	$\begin{bmatrix} 1 \times 1\, conv \\ 3 \times 3\, conv \end{bmatrix} \times 4$
Transition 2	28 × 24 × 384	28 × 24 × 256	1 × 1 × 256 conv
Pool3	28 × 24 × 256	14 × 12 × 256	2 × 2 max pooling, stride 2
Dense Block 3	14 × 12 × 256	14 × 12 × 576	$\begin{bmatrix} 1 \times 1\, conv \\ 3 \times 3\, conv \end{bmatrix} \times 5$
Transition 3	14 × 12 × 576	14 × 12 × 512	1 × 1 × 512 conv
Pool4	14 × 12 × 512	7 × 6 × 512	2 × 2 max pooling, stride 2
Dense Block 4	7 × 6 × 512	7 × 6 × 896	$\begin{bmatrix} 1 \times 1\, conv \\ 3 \times 3\, conv \end{bmatrix} \times 6$
Transition 4	7 × 6 × 896	7 × 6 × 512	1 × 1 × 512 conv
Pool5	7 × 6 × 512	4 × 3 × 512	2 × 2 max pooling, stride 2
Dense Block 5	4 × 3 × 512	4 × 3 × 896	$\begin{bmatrix} 1 \times 1\, conv \\ 3 \times 3\, conv \end{bmatrix} \times 6$
Transition 5	4 × 3 × 896	4 × 3 × 1024	1 × 1 × 1024 conv
Pool6	4 × 3 × 1024	1 × 1 × 1024	2 × 2 Global pooling
fc7	1 × 1 × 1024	10,575	Full connection

results. Our training process is two step. First we train a baseline model with CASIA-WebFace using the DenseNet [10] structure. Table 1 demonstrates the details of the network. Then we utilize the triplet loss [3] to fine-tune the baseline model with the domestic training set. For the first step, we set the batch size as 128, the learning rate as 0.1 and will be decreased by half every 40,000 iterations, and the weight decay as 5×10^{-4}. For the second step, we set the batch size as 120, the learning rate as 0.01 and will be decreased by half every 40,000 iterations, and the weight decay as 5×10^{-4}.

4.2 Qualitative Evaluation of Illumination Augmentation

Figures 2 and 3 demonstrate the illumination augmentation results on CASIA-WebFace and the domestic training set. We can see that the illumination augmentation approach manage to add additional illumination variations to the input image without changing the facial details for both CASIA-WebFace and the domestic training set. Our approach is proved to be adaptive to any training sample with changing illumination, pose, and image quality.

4.3 Quantitative Evaluation of Illumination Augmentation

Evaluation on LFW. Table 2 demonstrates the quantitative evaluation of the proposed illumination augmentation (IA) approach on the LFW dataset. We compare our method with DeepFace [1], DeepID2+ [2], and FaceNet [3]. The experimental results verify that the proposed IA approach is effective for improving the performance of the existing deep models. Implementing the proposed network with the augmented training samples results in an improvement of 0.27% verification accuracy. We also notice that the verification accuracy of the proposed approach outperforms DeepFace and DeepID2+. Note that with the same training samples, the accuracy of our method is higher than that of FaceNet. Consequently, our method is competitive against the state-of-the-art methods.

Evaluation on the Domestic Testing Set. Table 3 demonstrates the evaluation results on the domestic testing set. We can see that training deep models

Table 2. Evaluation on LFW

Method	DeepFace [1]	DeepID2+ [2]	FaceNet [3]	FaceNet [3]	Proposed	Proposed (IA)
Number of samples	4M	-	0.49M	200M	0.49M	0.49M
Verification accuracy	97.35%	98.70%	98.3%	99.63%	98.45%	98.72%

Table 3. Evaluation on the domestic testing set

Method	Training set	Number of samples	Accuracy
FaceNet (Softmax)	CASIA-WebFace	0.49M	57.1%
FaceNet (Triplet)	Domestic set	0.12M	74.7%
FaceNet (Triplet)	Domestic set	0.86M	81.7%
Proposed (Softmax)	CASIA-WebFace	0.49M	65.66%
Proposed (Triplet)	Domestic set	0.12M	76.7%
Proposed (Triplet+IA)	Domestic set	0.12M	80.91%
Proposed (Triplet)	Domestic set	0.86M	85.43%
Proposed (Triplet+IA)	Domestic set	0.86M	89.45%

with the domestic training set is beneficial to improving the recognition accuracy on the test set captured in the realistic surveillance environment. Compared with the deep models trained with CASIA-WebFace, an improvement of 24.6% is obtained for FaceNet trained with the domestic training set. Similarly, an improvement of 19.77% is also observed for the proposed network. With more training data, the performance of deep networks continue to improve. As the number of training data increases from 0.12M to 0.86M, we can see a performance gain of 7% for FaceNet and 8.73% for our network. Furthermore, we notice that the proposed IA approach is effective for improving the performance of deep networks with the domestic dataset. With IA approach, an improvement of 4.02% is observed for the proposed network. Note that the performance of the proposed network trained with CASIA-WebFace outperforms that of FaceNet with a margin of 8.56%. Consequently, our method achieves better generalizablity than FaceNet.

5 Conclusion

In this paper, we study the topic of data augmentation for face recognition and propose an illumination augmentation (IA) method. We first simulate different illumination conditions from the external benchmark and then perform color correction to obtain the augmented training samples with additional illumination variations while preserving the facial details. The IA approach is suitable for any face image with changing illumination, pose, and image quality. To further improve the performance of the deep networks towards robust face recognition under realistic environment, we construct a domestic training set together with a domestic testing set. Experimental results on the LFW and the domestic testing set verify the effectiveness of the proposed approach.

Acknowledgments. This project was supported by the NSFC (U1611461, 61573387, 61672544) and Tip-top Scientific and Technical Innovative Youth Talents of Guangdong special support program (NO. 2016TQ03X263).

References

1. Taigman, Y., Yang, M., Ranzato, M.A., Wolf, L.: DeepFace: closing the gap to human-level performance in face verification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1701–1708 (2014)
2. Sun, Y., Wang, X., Tang, X.: Deeply learned face representations are sparse, selective, and robust. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2892–2900 (2015)
3. Schroff, F., Kalenichenko, D., Philbin, J.: FaceNet: a unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 815–823 (2015)
4. Yin, W., Fu, Y., Sigal, L., Xue, X.: Semi-latent GAN: learning to generate and modify facial images from attributes. arXiv preprint arXiv:1704.02166

5. Bao, J., Chen, D., Wen, F., Li, H., Hua, G.: CVAE-GAN: fine-grained image generation through asymmetric training. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2745–2754 (2017)
6. Tran, L., Yin, X., Liu, X.: Disentangled representation learning GAN for pose-invariant face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1415–1424 (2017)
7. Yi, D., Lei, Z., Liao, S., Li, S.Z.: Learning face representation from scratch. arXiv preprint arXiv:1411.7923
8. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Technical report 07–49, University of Massachusetts, Amherst (2007)
9. Sim, T., Baker, S., Bsat, M.: The CMU pose, illumination, and expression (PIE) database. In: Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 53–58 (2002)
10. Huang, G., Liu, Z., Weinberger, K.Q., van der Maaten, L.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708 (2017)
11. Demirel, H., Anbarjafari, G.: Pose invariant face recognition using probability distribution functions in different color channels. IEEE Sig. Process. Lett. 537–540 (2008)