AN EFFICIENT DEEP NEURAL NETWORKS TRAINING FRAMEWORK FOR ROBUST FACE RECOGNITION

Canping Su[1], Yan Yan[1,*], Si Chen[2], Hanzi Wang[1]

[1]Fujian Key Laboratory of Sensing and Computing for Smart City, School of Information Science and Engineering, Xiamen University, Xiamen 361005, China

[2]School of Computer and Information Engineering, Xiamen University of Technology, Xiamen 361024, China

Introduction

We propose an efficient deep neural networks training framework for face recognition. The framework contains two stages:

- ✓ The DNN initialization. A deep architecture based on the softmax loss function is designed to initialize the DNN.
- ✓ The adaptive fine-tuning. The formulation is described as follow.

$$Loss = \sum_{i=1}^{N} \max\left(err(S_i), 0\right) \quad (1)$$

$$err(S_i) = D_{ap_i} + \frac{D_{max}}{\sqrt[n]{D_{ap_i}}}\tau - D_{an_i} \quad (2)$$

The completed method

Algorithm 1: An Efficient DNN Training Framework

```
Input: Training dataset S, sampling interval K, and
       maximal epoch T
Output: The trained network parameters W
Initialization: Randomize W, 𝒯 = ∅, t = 1;
while not converge do   // The DNN initialization
    for each training sample x_i ∈ S do
        Forward pass to obtain the face representation;
        Backpropagate to update the network parameters W
          via the original softmax.
    end
end
while t < T do          // The adaptive fine-tuning
    if t mod K then // Generate triplet samples
        Generate positive pair (x_i^a, x_i^p) according the current
          model parameters W;
        Select the negative sample x_i^n via Eq. (2);
        𝒯 = 𝒯 ∪ (x_i^a, x_i^p, x_i^n);
    end
    for each triplet sample (x_i^a, x_i^p, x_i^n) ∈ 𝒯 do
        Forward pass to obtain the face representation;
        Backpropagate to update the network parameters W
          via Eq. (1);
    end
    t ← t + 1;
end
```

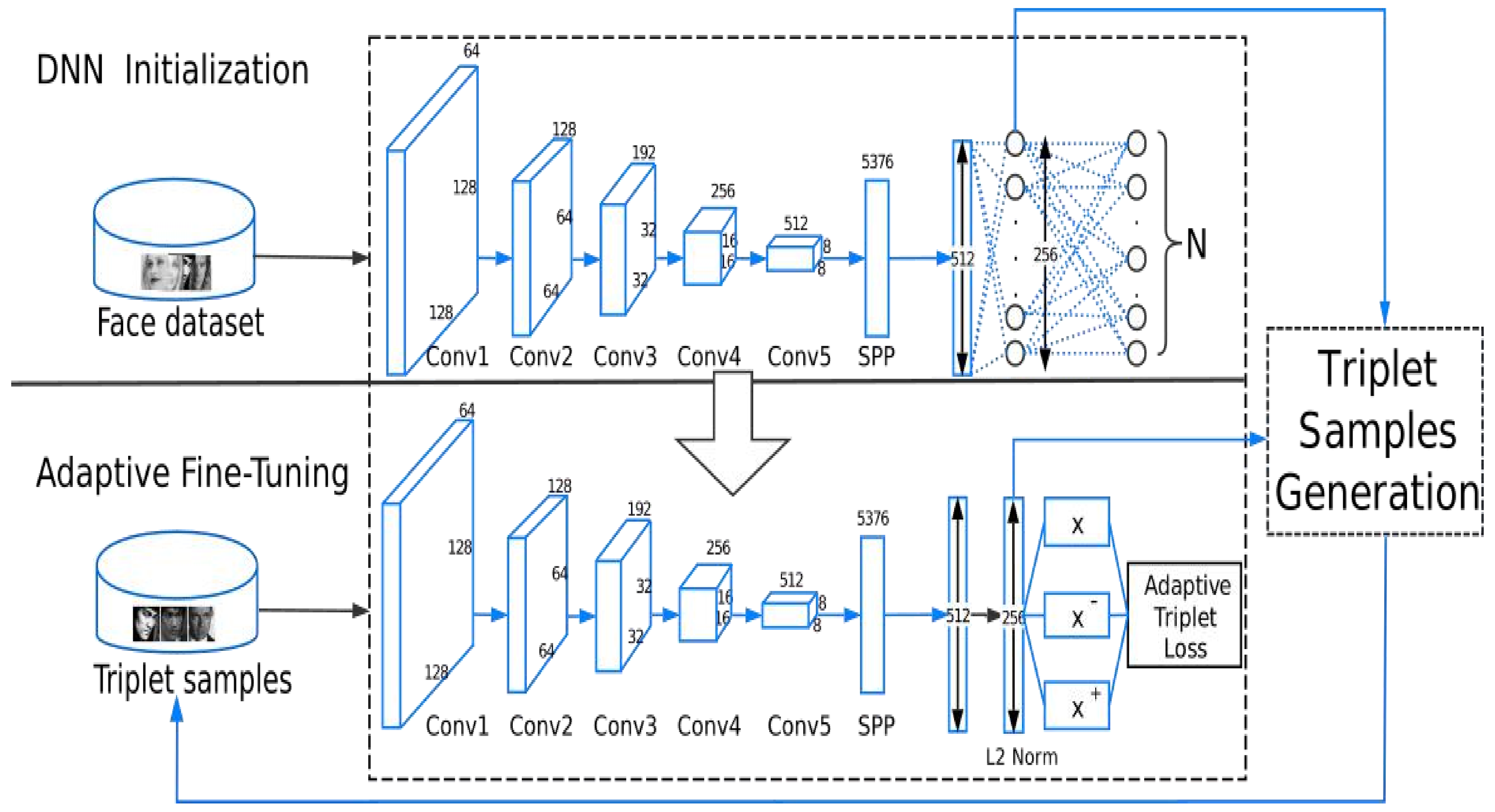

Fig. 1. The schematic diagram of the proposed training framework.

Results

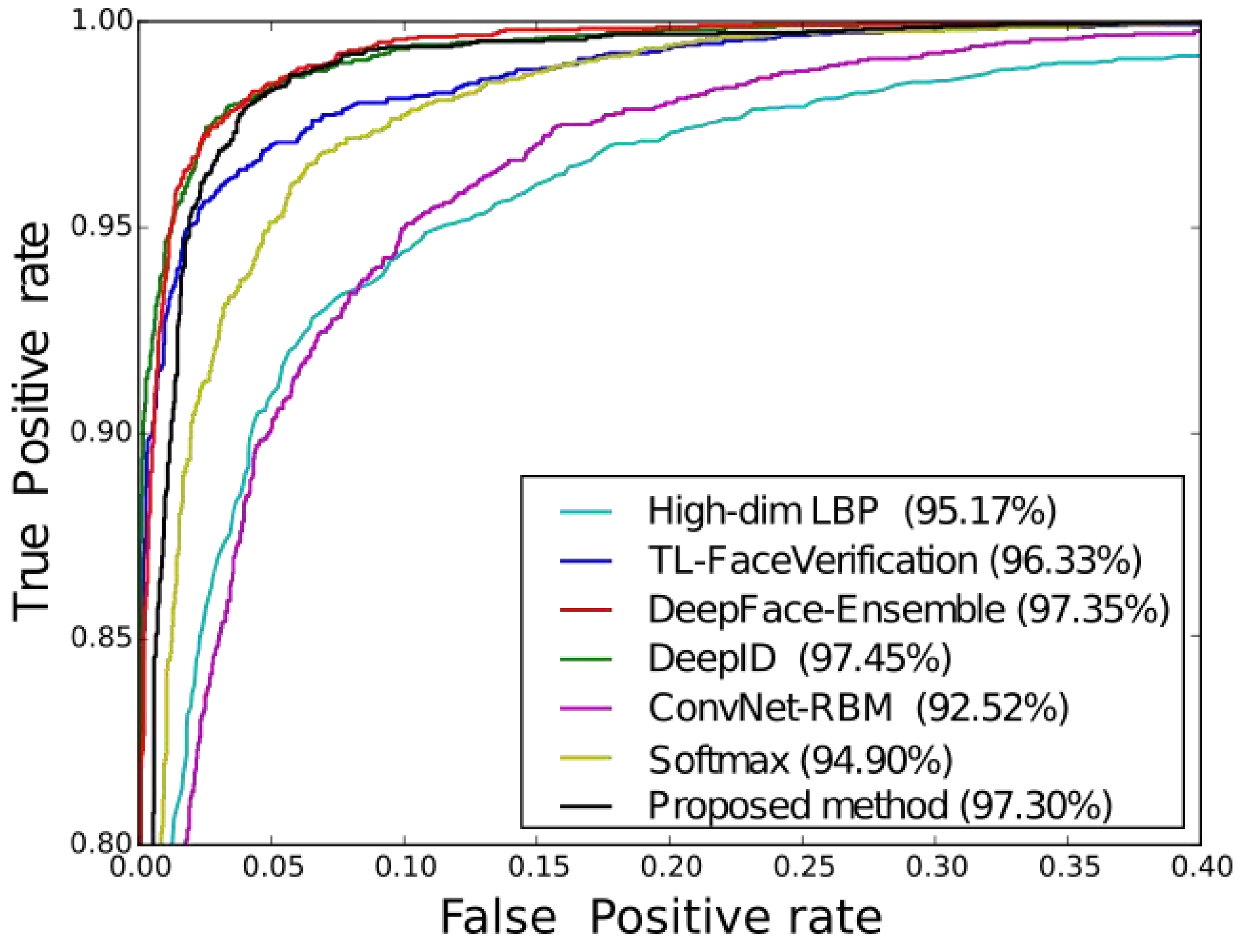

Fig. 2. Performance on the LFW dataset.

Table 1. The training time obtained by the different methods

Method	Time (hours)
Softmax	24
Triplet	Failed (more than 168)
FaceNet[1]	more than 1000
Proposed method	96

Table 2. The recognition accuracy on the different datasets.

Method	FERET	MultiPIE	FEI	Carmera 12
DAE [2]	84.80%	82.50%	N/A	N/A
SPAE [3]	92.50%	91.40%	N/A	N/A
Softmax	99.96%	97.26%	98.72%	98.44%
Proposed method	**99.99%**	**99.31%**	**99.96%**	**99.52%**

Conclusions

- In this paper, a novel DNN training framework, which takes advantage of both the softmax loss and triplet loss functions, has been proposed for efficient face recognition.
- A specific softmax loss-based DNN architecture is designed to initialize the DNN. Based on it, we improve the discrimination capability of the DNN with a triplet loss function, where an adaptive margin is adopted.
- We have verified the effectiveness of the proposed DNN training framework on the LFW dataset and four different face datasets.

References

[1] Florian Schroff, Dmitry Kalenichenko, and James Philbin, “Facenet: A unified embedding for face recognition and clustering,” in CVPR, 2015, pp. 815–823.

[2] Yoshua Bengio, “Learning deep architectures for ai,” FTML, vol. 2, no. 1, pp. 1–127, 2009.

[3] Meina Kan, Shiguang Shan, Hong Chang, and Xilin Chen, “Stacked progressive auto-encoders (spae) for face recognition across poses,” in CVPR, 2014, pp. 1883–1890.

Contact us

yanyan@xmu.edu.cn