Likelihood Ratio based Loss to finetune CNNs for Very Low Resolution Face Verification

Dan Zeng[1], Raymond Veldhuis[1], Luuk Spreeuwers[1], Qijun Zhao[2]
[1] Faculty of EEMCS, University of Twente, Enschede, the Netherlands
[2] College of Computer Science, Sichuan University, China
{d.zeng,r.n.j.veldhuis,l.j.spreeuwers}@utwente.nl, qjzhao@scu.edu.cn

Abstract

In this paper, we propose a likelihood ratio based loss for very low-resolution face verification. Existing loss functions either improve the softmax loss to learn large-margin facial features or impose Euclidean margin constraints between image pairs. These methods are proved to be better than traditional softmax, but fail to guarantee the best discrimination features. Therefore, we propose a loss function based on likelihood ratio classifier, an optimal classifier in Neyman-Pearson sense, to give the highest verification rate at a given false accept rate, which is suitable for biometrics verification. To verify the efficacy of the proposed loss function, we apply it to address the very low-resolution face recognition problem. We conduct extensive experiments on the challenging SCface dataset with the resolution of the faces to be recognized below 16×16*. The results show that the proposed approach outperforms state-of-the-art methods.*

1. Introduction

Face verification is a common computer vision task, determining whether a pair of faces belongs to the same identity, that is widely used for identity authentication. Face verification performance has been boosted due to advanced deep CNN architectures [1, 2, 4, 3, 5] and the development of discriminative learning approaches. The main purpose of discriminative learning approaches is to ensure the features of the same person have a small distance while features of different individuals have a considerable distance. Loss functions are commonly used to shape the discriminative learning.

Loss functions are categorized into classification loss (identification loss) functions [6, 7, 8, 9, 10, 11] and metric learning loss (verification loss) functions [12, 13] according to whether an image or image pair is explicitly used. The most popular classification loss, i.e., the softmax loss is widely used for deep CNN learning. However, it forces the network to learn separating features (i.e., separating different classes) which is not necessarily discriminative enough and cannot ensure that the features of the same identity have a small distance. Many prevailing classification loss functions (see Fig. 1(a)) are proposed to address this issue. These methods improve the embedded features by incorporating various margin-based constraints to strengthen the classifier part and then use the softmax function and cross-entropy loss to supervise network training. Compared with the traditional softmax loss function, these embedded features are more discriminative mainly because of the use of margin constraints, which improves the performance of face recognition.

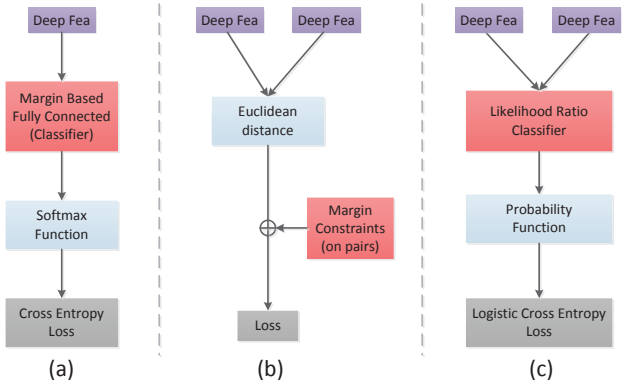

Figure 1. Comparison between existing loss functions including (a) margin based softmax loss, (b) image pair magrgin based loss and the proposed loss function (c) likelihood ratio based loss.

Metric learning loss functions (see Fig. 1(b)) such as contrastive loss [13] or triplet loss [12] explicitly constrain the relationship between image pairs and force positive pairs to have a small distance and negative pairs to have a large distance. More specifically, Euclidean distance based margin constraint is employed to get the negative pairs far

978-1-7281-3640-0/19/$31.00 ©2019 IEEE

away from positive pairs, which can improve the discriminative capacities of embedding features.

All of the above loss functions are margin-constraint based learning methods, either Euclidean distance margin based [13, 12, 6], cosine similarity metric based [9, 10, 11], or angular margin based [7, 8]. These methods can achieve better results compared with traditional softmax, but they cannot guarantee the best discriminative ability of embedded deep features.

In this paper, we propose a likelihood ratio loss function which directly optimizes the constraints between image pairs (see Fig. 1(c)). It is well-known that an optimal classifier in Neyman-Pearson sense is obtained by thresholding the likelihood ratio, cf. for example [14], which means at a given false-acceptance rate the classifier minimized the false rejection rate and vice versa. The likelihood ratio will give the highest verification rate at a given false accept rate. Therefore, the proposed loss function can be deployed to solve the face verification task and would achieve the optimal verification performance in theory. A comparison between the proposed method and the existing loss functions is shown in Fig. 1.

Ideally, we could train deep CNNs from scratch by likelihood ratio loss. But it requires a large amount of training data, and the training is time-consuming. In this paper, we combine the deep CNN with the proposed loss function sequentially and then fine-tune the parameters using likelihood ratio loss function. The deep CNN is assumed to be trained already by arbitrary loss functions, and any advanced deep models can be incorporated. More specifically, we apply the proposed loss function to the specific situation with only limited data available. Our application domain is very low resolution (VLR) face recognition, which is challenging and more difficult than low resolution (LR) face recognition. The size of LR face images is less than 25×25 pixels [16] while the resolution of the VLR face image is lower than 16×16 pixels [15].

The three main contributions of this paper are described below.

- We propose a likelihood ratio based loss function for very low resolution face verification and can achieve good performance.
- We propose a training procedure so that CNNs trained by arbitrary losses can be further fine tuned with only limited data at hand.
- We can use situation-relevant prior odds to benefit the VLR face recognition task.

The rest of this paper is organized as follows. Related work is shown in Sec. 2. Sec. 3 describes the proposed method. Experimental results are shown in Sec. 4. Concluding remarks are drawn in the end.

2. Related Work

In this section, we first elaborate the existing loss functions and highlight the difference in the proposed method. Then we focus on our application domain and illustrate the related work on very low resolution face recognition.

2.1. Loss Functions

Loss functions play a very important role in the deep learning techniques. The traditional softmax loss function involves the fully connected layer, softmax function and cross-entropy loss function. The fully connected layer realises the classifier and does not require explicit margin based constraints. Thus the feature embedding can be separable but it cannot ensure sufficient discrimination of the features to reduce intra-class distances. Center face [6] penalizes the Euclidean distance between the deep features and their corresponding class center such that the deep features of each class are pulled to its center. It only explicitly optimizes the intra-class compactness while ignoring the inter-class variances.

L-Softmax loss [7] enforces an angular margin constraint between the deep features of different classes. It implicity incorporates the angle which may render the features angularly distributed thus this does not guarantee optimal discrimination. As an improvement, A-Softmax loss [8] explicitly projects the original Euclidean space of features to an angular space and then imposes angular margins on the projected features. Angular margin based loss functions are more appropriate than Euclidean distance margin based loss functions because the traditional softmax function has an intrinsic consistency with cosine of the angle [8]. Recently, cosine margin based loss functions [9, 10, 11] are introduced to learn large margin features in a natural way. It seems reasonable to apply cosine margin to the features between different classes because cosine distance metric is frequently used for recognition. Compared with angular margin based constraints, cosine margin based loss functions are more robust to the noises around the boundary [11].

As Fig. 1 shows, most prevailing loss functions are proposed to engage in various margin constraints based on softmax loss. More specifically, these methods develop the fully connected layer (regarded as a classifier) in the softmax loss by enforcing margin constraints. In other words, the core task of softmax loss based methods is how to improve the classifier part, i.e., fully connected layer part. More elaborated classifier produces the better performance. As for face verification, the most intuitive way is to employ constraints on image pairs rather than the single image. Contrastive loss [13] minimizes the Euclidean distances between faces for the same identity and enforces a margin between different identities. Triplet loss [12] aims to pull the positive pair separately from the negative by a Euclidean

distance. In conclusion, these methods define the rule which acts as 'classifier' to improve network training. However, the mining for appropriate image pairs can be troublesome. Differently, we merge the likelihood ratio classifier as the part of the proposed loss function. It offers two benefits: first, the likelihood ratio classifier is an optimal classifier in Neyman-Pearson sense. Second, it works when the training data is limited.

2.2. Very Low Resolution Face Recognition

Faces captured at some distance by surveillance cameras usually have a very low resolution, i.e., the size is less than 16×16 pixels, whereas enrolled faces are collected in a controlled scenario with high resolution. The comparison of both constitute the very low resolution face recognition problem.

To our knowledge, there are few works proposed to address this challenging task. To solve the problem, Zou et al. [15] design two constraints including a data constraint and a discriminant constraint. The data constraint is to estimate the reconstruction error in the high resolution (HR) image space to make use of the information from HR training images. The discriminative constraint is to use class label to boost recognition performance. This method focuses more on face super resolution (SR) rather than face recognition, which leads to a suboptimal recognition performance. Ref. [36] utilizes singular value decomposition to represent face images and considers face hallucination and low resolution recognition simultaneously to improve the performance for each task. PCSRN (Partially Coupled SR Networks) [17] is introduced to solve the very low resolution problem not limited to face recognition by assuming that a part of HR feature and VLR feature are shared. It generalizes VLR face recognition to a common VLR recognition problem while ignoring the characteristic brought by face, which may lead to suboptimal solutions.

Recently, Peng et al. [18] propose Mixed Resolution Classifier (MRC) to map HR faces and VLR faces to a common feature space and obtain its likelihood ratio for face verification. Discriminative MDS method [19] learns mapping matrix to project the HR and LR images to a common space while considering both interclass distances and intraclass distances. However, this method concerns LR instead of VLR face recognition. Deep coupled resnet [37] is presented to combine one trunk network and two branch networks. The trunk network is used to extract discriminative features and the branch networks are trained on image pairs with HR and specific target LR faces to further minimize their feature differences. GenLR-Net [20] is developed to deal with low resolution face and object recognition. However, it mainly focuses on object recognition and only use simulated faces (not realistic low resolution faces) for testing. Resolution invariant deep network (RIDN) [21] is proposed to solve the LR face recognition and has achieved best performance, but the performance decreases dramatically when the images degrade from LR to VLR. The proposed RIDN is the basis of the method proposed here and is used for feature extraction.

3. The Proposed Approach

3.1. Problem Statement

Most margin based softmax losses incorporate a better classifier, i.e., margin based fully connected layer, to improve the face recognition performance. Similarly, we propose a loss function based on the traditional likelihood ratio classifier in which the margin constraints are explicitly optimised which results in a better feature embedding. The likelihood ratio classifier has been proven to be an optimal classifier in Neyman-Pearson sense which can achieve theoretically optimal performance for face verification.

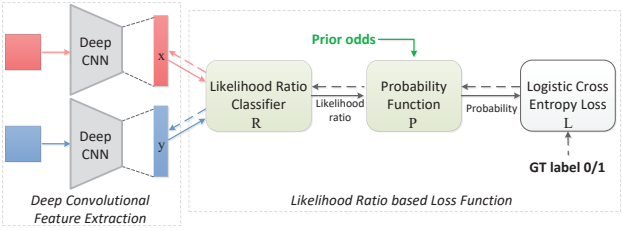

Figure 2. The framework of the proposed likelihood ratio loss function. The gray box indicates that parameters of the network are fixed during the fine-tuning stage.

The framework of the proposed approach is shown in Fig. 2. We use the deep CNN for convolutional feature extraction and the architecture of deep CNN can be arbitrary. In this paper, we use the RIDN [21] (see Fig. 3) as a example to illustrate how the likelihood ratio loss function improves the embedded features for the specific task, such as very low resolution face recognition. The likelihood ratio based loss function consists of three parts: The likelihood ratio classifier unit takes the deep feature pair as input and outputs its likelihood ratio; The probability function unit considers the situation-relevant prior odds to predict the probability that a pair of features belongs to the same person; The logistic cross entropy loss unit calculates the loss to supervise the network training and update the parameters.

The initial deep features of the HR image and the VLR image are presented by $\mathbf{x}$ and $\mathbf{y}$, respectively. Likelihood ratio score, obtained by likelihood ratio classifier $\mathrm{R}(\cdot)$, that is denoted as,

$$s = \mathrm{R}(\mathbf{x}, \mathbf{y}; \mathbf{W}) \tag{1}$$

where $\mathbf{W}$ represents the covariance and cross-covariance matrices that needs to be estimated during a training process. It will be fixed after being estimated. Then we use

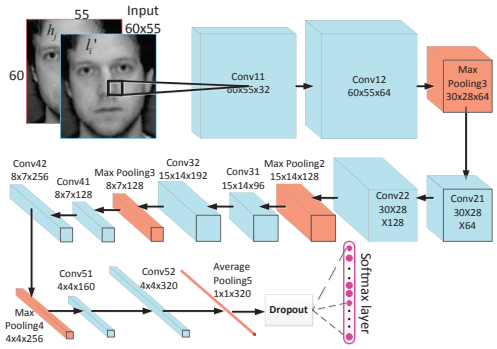

Figure 3. Architecture of Resolution Invariant Deep Network (RIDN) [21] for resolution-robust feature extraction.

the probability function $\mathrm{P}(\cdot)$ and logistic cross entropy loss function $\mathrm{L}(\cdot)$ sequentially to update the initial features.

During face recognition, we first obtain the updated features, and then use the likelihood ratio classifier as the metric for face verification.

3.2. Likelihood Ratio Classifier

Given two embedded deep features $\mathbf{x} \in \mathbb{R}^{\mathrm{n}}$ and $\mathbf{y} \in \mathbb{R}^{\mathrm{n}}$. We look for support for hypothesis $\mathrm{H_s}$: the features originate from the same person versus hypothesis $\mathrm{H_d}$: the features originate from different individuals. The decision that provides a maximum verification rate at a given false-acceptance rate follows by thresholding the likelihood ratio:

$$\mathrm{lr}\left(\mathbf{x},\mathbf{y}\right)=\frac{\mathrm{p}\left(\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}|\mathrm{H_s}\right)}{\mathrm{p}\left(\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}|\mathrm{H_d}\right)}. \tag{2}$$

The similarity score, which is used as an approximation of log likelihood ratio, is denoted as,

$$\mathrm{s}(\mathbf{x}_c,\mathbf{y}_c)=-\sum_{\mathrm{i}=1}^{\mathrm{d}}\frac{\nu_{\mathrm{i}}}{1-\nu_{\mathrm{i}}}(\mathbf{x}_{c,i}-\mathbf{y}_{c,i})^2+\sum_{\mathrm{i}=1}^{\mathrm{d}}\frac{\nu_{\mathrm{i}}}{1+\nu_{\mathrm{i}}}(\mathbf{x}_{c,i}+\mathbf{y}_{c,i})^2. \tag{3}$$

The derivation of formulas is ignored here for simplicity and can be be found in Ref. [18]. The block diagram of the likelihood ratio classifier according to Eq. (3) is shown in Fig. 4. It describes how to get a similarity score from a given image pair in detail.

The likelihood ratio classifier contains feature reduction and similarity score calculation. During feature reduction, whitening transforms is first applied to $\mathbf{x}$ and $\mathbf{y}$ and $\mathbf{U},\mathbf{D},\mathbf{V}$ are obtained by singular value decomposition. Here subscript $*,1d$ in $(\mathbf{U}_{*,1d})^{\mathrm{T}}$ and $(\mathbf{V}_{*,1d})^{\mathrm{T}}$ denotes the first d columns of matrix are used. The $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$ are the average HR and LR image features, respectively. In short, feature reduction phase reduces the dimension of features from dimension n to a compact feature dimension d. The

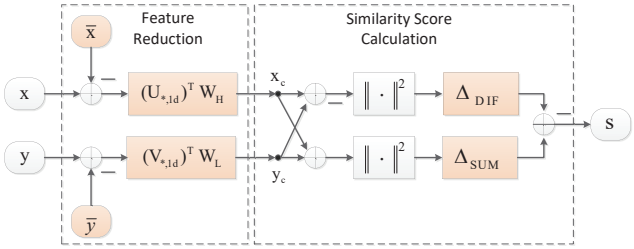

Figure 4. Block diagram of the likelihood ratio classifier. All the boxes in color denotes matrices that need learning during training process.

color blocks perform matrix multiplications while the white block computes a squared vector norm. In similarity score calculation, the matrices Δ_{DIF} and Δ_{SUM} are diagonal matrices learned from training process which are defined as follows,

$$\Delta_{\mathrm{DIF},i}=\frac{\nu_i}{\nu_i-1}\,,i=1,...,d \tag{4}$$

$$\Delta_{\mathrm{SUM},i}=\frac{\nu_i}{\nu_i+1}\,,i=1,...,d \tag{5}$$

where d is the number of singular value ν_i on the diagonal of dignonal matrix $\mathbf{D}$.

3.3. Probability Function

For any pair of features, its log likelihood ratio can be obtained as described in Sec. 3.2. $\mathrm{H_s}$ hypothesizes the features originate from the same identity and $\mathrm{H_d}$ hypothesizes the features originate from the different individuals. Based on the posterior probability, we obtain,

$$\mathrm{p}(\mathrm{H_s}|s)=\frac{\mathrm{p}(s|\mathrm{H_s})\mathrm{p}(\mathrm{H_s})}{\mathrm{p}(s)} \tag{6}$$

$$\mathrm{p}(\mathrm{H_d}|s)=\frac{\mathrm{p}(s|\mathrm{H_d})\mathrm{p}(\mathrm{H_d})}{\mathrm{p}(s)} \tag{7}$$

Divide Eq.(6) and Eq.(7) and bring $\mathrm{p}(\mathrm{H_s}|s)+\mathrm{p}(\mathrm{H_d}|s)=1$ into the formula. Then we can arrive at:

$$\frac{\mathrm{p}(\mathrm{H_s}|s)}{1-\mathrm{p}(\mathrm{H_s}|s)}=\frac{\mathrm{p}(s|\mathrm{H_s})}{\mathrm{p}(s|\mathrm{H_d})}\cdot\frac{\mathrm{p}(\mathrm{H_s})}{\mathrm{p}(\mathrm{H_d})} \tag{8}$$

where $\frac{\mathrm{p}(\mathrm{H_s})}{\mathrm{p}(\mathrm{H_d})}$ is the prior odds, which is a hyper-parameter and needs to be given according to the specific situation. For simplicity, we abbreviate prior odds to A.

Here likelihood ratio $\frac{\mathrm{p}(s|\mathrm{H_s})}{\mathrm{p}(s|\mathrm{H_d})}$ can be written as $\mathrm{lr}(s)$, denoting the likelihood ratio on s. After training likelihood ratio classifier, the similarity score s would be the $\mathrm{lr}(\mathbf{x},\mathbf{y})$ or the $\mathrm{llr}(\mathbf{x},\mathbf{y})$ which is short for $\log(\mathrm{lr}(\mathbf{x},\mathbf{y}))$. If we train the method sufficiently, the ratio of similarity score function $\mathrm{lr}(s)$ would become $\mathrm{lr}(\mathbf{x},\mathbf{y})$. We use llr to represent $\frac{\mathrm{p}(s|\mathrm{H_s})}{\mathrm{p}(s|\mathrm{H_d})}$. The Eq.(8) is denoted as,

$$\mathrm{P}(\mathrm{H_s}|s)=\frac{\mathrm{e}^{\mathrm{llr}+\log A}}{1+\mathrm{e}^{\mathrm{llr}+\log A}} \tag{9}$$

where llr is short for $\mathrm{llr}(s)$ and can become $\mathrm{llr}(\mathbf{x}, \mathbf{y})$ from sufficiently training. And llr can be approximated by likelihood ratio classifier.

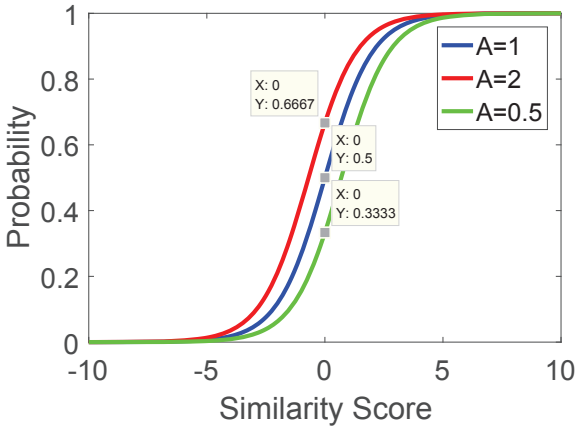

Figure 5. How prior odds A affects the predicted probability. When the similarity score is fixed, the prior odds produce the probability that an image pair belongs to the same person.

From Eq.(9), we could see if prior odds A equals to 1, the formula becomes the sigmoid function. As Fig. 5 shows, with the increase of A, higher confidence will be given on positive image pairs. With the decrease of A, bias will be given on negative image pairs.

3.4. Logistic Cross Entropy Loss Function

The probability function predicting the probability that an image pair belongs to the same identity is incorporated into the final loss calculation. We rewrite the Eq.(9) as follows for concise,

$$\hat{\mathrm{P}} = \frac{1}{1 + A^{-1}\mathrm{e}^{-s}} \tag{10}$$

with similarity score s equals to llr, where $\hat{\mathrm{P}}$ represents the predicted probability. The loss function is denoted as,

$$\begin{aligned} \mathrm{L} &= -\mathrm{P}\log\hat{\mathrm{P}} - (1-\mathrm{P})\log(1-\hat{\mathrm{P}}) \\ &= -\mathrm{P}(\log A + s) - \log\frac{\mathrm{e}^{-s}}{A + \mathrm{e}^{-s}} \end{aligned} \tag{11}$$

where P is the target probability, and $\mathrm{P} = 1$ means the image pair are from the same identity, and $\mathrm{P} = 0$ means the image pair are from different individuals. The loss is minimized over the parameters of the deep CNN by computing the gradient of L, and stochastic gradient descent (SGD) is used in back-propagation.

4. Experimental Results

In this section, we first introduce the training phases and training data, and then discuss the factors that affect the performance. Finally, we report the experimental results on SCface [22], which is the public dataset we can find that contain the realistic VLR faces, to verify the efficacy of the proposed likelihood ratio based loss function.

Table 1. Composition of RIDN dataset. The table presents the number of subjects (# Sub), the images per subject (# Ips), and the average inter-pupillary distance (IPD) as well as the standard deviation

Dataset	# Sub	# Ips	IPD [pixel]
WebFace [23]	10069	$1-534$	$58\,(5)$
FERET [24]	1195	$1-24$	$60\,(2)$
CAS-PEAL [25]	1040	$3-43$	$61\,(3)$
FRGC v2 [26]	466	$1-88$	$126\,(2)$
Multi-PIE [27]	337	$83-486$	$72\,(5)$
MUCT [28]	176	$7-12$	$89\,(6)$
Faces94 [29]	153	$7-20$	$48\,(4)$
AR [30]	100	$2-6$	$57\,(3)$
PIE [31]	68	$2-5$	$80\,(8)$
ORL [32]	40	$6-10$	$34\,(3)$
Pointing 04 [33]	15	$32-42$	$53\,(5)$
Grimace [34]	12	$2-20$	$51\,(5)$

4.1. Datasets

RIDN dataset is composed of 12 public face datasets. The training dataset has $13,671$ subjects, including $438,139$ images in total. Table 1 lists details of the data. The facial images display illumination, expression and pose variations. Only facial images with pose less than $30°$ in the yaw orientation and $15°$ in the pitch orientation are included. All images are of relatively high resolution according to inter-pupillary distance.

SCface dataset contains facial images of 130 subjects taken in an uncontrolled indoor environment. The facial images are captured by five surveillance cameras at three distances, distance1 $(4.20m)$, distance2 $(2.60m)$ and distance3 $(1.00m)$, and one frontal mugshot per subject was taken by a digital camera is included. The surveillance cameras are placed slightly above the subject's head. Some of the collected images are blurred. Moreover, pose and lighting as well as quality varies for different cameras at different distances. Facial images captured at distance1 $(4.20m)$ are of the poorest quality, where inter-pupillary distance is lower than 10 pixels. Example images captured at distance1 are shown in Fig. 6.

4.2. Experimental Settings

The proposed method consists of three training phases: (1) training the deep CNN; (2) training the likelihood ratio classifier; (3) employing the likelihood ratio loss to fintune the deep CNN. We will elaborate each phase below.

As for the deep CNN training, we use the RIDN as a representative example and take the RIDN dataset for training.

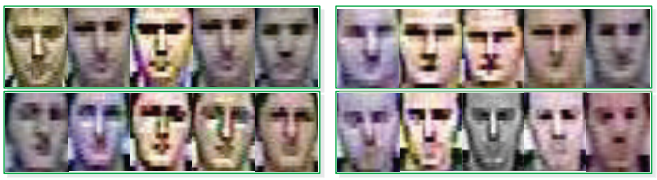

Figure 6. Examples of face images in SCface dataset which are captured at distance1 ($4.20m$). Faces images originates from the same identity are in a green box.

Any advanced network combined can be used because we take the deep model as a feature extractor.

The second phase is the likelihood ratio classifier training. To compare fairly with the latest MRC method [18], we stay with the same experimental settings as Ref. [18]. We consider the frontal mugshot as the enrolled image and distance1 ($4.20m$) as the test images. The first 100 subjects are selected as reference/probe combinations for likelihood ratio classifier off-line training.

The third phase is to use the proposed loss to supervise the deep CNN finetune. Here we randomly generate genuine and impostor pairs from the 100 subjects in the SCface dataset for training. Each person contains five VLR images captured at distance1 and one HR image. Thus we could get at most five genuine pairs per person, and we chose the same number of impostor pairs. Only the last layer of deep CNN is trained during this phase. We will discuss how the number of image pairs affects the performance in the below.

For face recognition, we use images captured at distance1 ($4.20m$) as test faces, while using the HR face images taken under a controlled environment for enrollment. The block diagram is shown in Fig. 7. More specifically, we first extract features of images via the updated deep CNN, and then we use the likelihood ratio classifier for face verification.

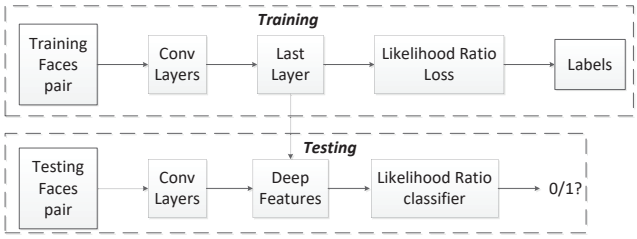

Figure 7. Block diagram about how to use the proposed loss for recognition.

In our experiments, all facial images from RIDN dataset are preprocessed through face detection and facial landmarking [35] and aligned by applying affine transformation using four landmarks, i.e., left eye center, right eye center, nose tip, and mouth center. We use the provided landmarks from SCface and repeat the stated preprocessing process. For all datasets mentioned above, HR and LR facial images are cropped to 60×55 and 30×24, respectively.

The CNN is trained following the method in [21], except that we mix five resolutions instead of four as used in Ref. [21]. The resolution of 55×50 is added, and this results in improved performance.

4.3. Exploratory Experiments

Effect of prior odds. Normally we assume the positive image pair and the negative pair would appear at the same probability. However, prior odds need to be changed under different situations requirements. For example, when passing by the checkpoint of customs, high security is needed, we could vary the prior odds to emphasize more to the negative pairs. Changing the prior probability can achieve different performance in the end (see Fig. 8). The horizontal axis indicates the ratio of negative sample pairs to positive sample pairs which equal to A^{-1}. The smaller the ratio is, the more consideration is given on the positive pairs, that is, positive pairs have higher prior probability. If the ratio equal is 1, it is just the same as the commonly used assumption. With the ratio increases, the higher prior probability is shifted on negative image pairs, and it gains more attention than positive pairs do.

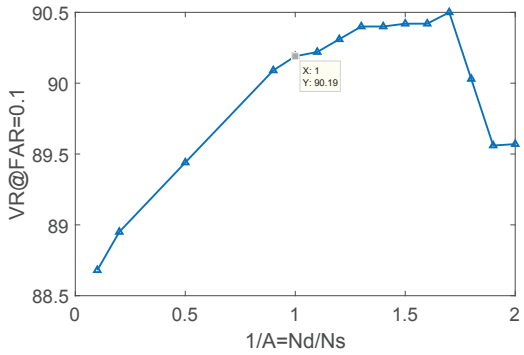

Figure 8. Effect of prior odds on performance. N_s and N_d denote the number of image pairs originates from the same person and different individuals, respectively.

As Fig. 8 shows, we can see that higher prior odds on a negative pair in a certain range can improve the performance. It suggests it is more important to enlarge inter-class variances rather than diminish intra-class variances to improve recognition result.

Effect of number of training image pairs. Here, the training image pairs are used to train the deep CNN supervised by the proposed likelihood ratio loss. We keep the number of genuine and impostor pairs equal. Then we vary the number of genuine and impostor pairs from 100 to 500 to compare its effect. Prior odds is set to 1, and the result is shown in Table. 2.

From the Table. 2, we find that the use of training image pairs can improve the performance compared with the

Table 2. A different number of training image pairs affect the recognition performance.

# genuine pairs	# impostor pairs	VR@FAR=0.1
100	100	88.47%
200	200	89.90%
300	300	**90.19%**
400	400	89.15%
500	500	89.73%

baseline performance 87.65%. The baseline performance is achieved by the same architecture but without the use of the proposed loss for fine-tuning. We choose the 300 genuine and impostor pair, respectively, in the below experiment.

4.4. Results on SCface

Comparison with State-of-the-art Methods. In this experiment, facial images of SCface captured at distance1 ($4.20m$) are compared to mugshots. The best performance reported in the literature is of RIDN [21] and MRC [18]. The comparison results of face verification protocols are listed in Table 3. The proposed baseline approach uses the deep CNN and likelihood ratio classifier but don't apply the likelihood ratio loss function to finetune the network. The best performance is achieved under the prior odds $N_d/N_s = 1.7$ and the detail about how the prior odds affect the recognition results can be seen in Fig. 8.

Table 3. Face recognition results on the SCface dataset.

Methods	VR@FAR=0.1
RIDN [21]	$70(3)\%$
MRC [18]	$73(6)\%$
Proposed Baseline	87.65%
Proposed Loss	**90.50%**

5. Conclusion

The results obtained by the proposed likelihood ratio based loss are promising. The proposed loss can be used in any specific applications with the existing deep model, and only small training image pairs are needed for the network training. In this paper, we use the VLR face recognition task as a case to show that the performance can be improved with the likelihood ratio loss. To our best knowledge, this work is the first (i) for exploring the likelihood ratio, into the loss design, and (ii) for introducing the proposed loss function to improve the performance of VLR face recognition.

References

[1] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. (2012) 1097–1105

[2] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)

[3] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2016) 770–778

[4] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2015) 1–9

[5] Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A. A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: AAAI. vol. 4, p. 12 (2017)

[6] Wen, Y., Zhang, K., Li, Z., Qiao, Y.: A discriminative feature learning approach for deep face recognition. In: European Conference on Computer Vision, Springer. (2016) 499–515

[7] Liu, W., Wen, Y., Yu, Z., Yang, M.: Large-margin softmax loss for convolutional neural networks. In: ICML. (2016) 507–516

[8] Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., Song, L.: Sphereface: Deep hypersphere embedding for face recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. vol. 1, p. 1 (2017)

[9] Wang, F., Liu, W., Liu, H., Cheng, J.: Additive margin softmax for face verification. arXiv preprint arXiv:1801.05599 (2018)

[10] Deng, J., Guo, J., Zafeiriou, S.: Arcface: Additive angular margin loss for deep face recognition. arXiv preprint arXiv:1801.07698 (2018)

[11] Wang, H., Wang, Y., Zhou, Z., Ji, X., Li, Z., Gong, D., Zhou, J., Liu, W.: Cosface: Large margin cosine loss for deep face recognition. arXiv preprint arXiv:1801.09414 (2018)

[12] 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. (2015) 815–823

[13] Sun, Y., Chen, Y., Wang, X., Tang, X.: Deep learning face representation by joint identification-verification. In: Advances in neural information processing systems. (2014) 1988–1996

[14] Bazen, A.M., Veldhuis, R.N.: Likelihood-ratio-based biometric verification. IEEE Transactions on circuits and systems for video technology **14**(1) (2004) 86–94

[15] Zou, W.W., Yuen, P.C.: Very low resolution face recognition problem. IEEE Transactions on Image Processing **21**(1) (2012) 327–340

[16] Choi, J. Y., Ro, Y. M., and Plataniotis, K. N.: Color face recognition for degraded face images IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) **39**(5) (2009) 1217–1230

[17] Wang, Z., Chang, S., Yang, Y., Liu, D., Huang, T.S.: Studying very low resolution recognition using deep networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2016) 4792–4800

[18] Peng, Y., Spreeuwers, L., Veldhuis, R.: Low-resolution face alignment and recognition using mixed-resolution classifiers. IET Biometrics **6**(6) (2017) 418–428

[19] Yang, F., Yang, W., Gao, R., Liao, Q.: Discriminative multidimensional scaling for low-resolution face recognition. IEEE Signal Processing Letters **25**(3) (2018) 388–392

[20] Prasad Mudunuri, S., Sanyal, S., Biswas, S.: GenLR-Net: Deep Framework for Very Low Resolution Face and Object Recognition With Generalization to Unseen Categories. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. (2018) 489–498

[21] Zeng, D., Chen, H., Zhao, Q.: Towards resolution invariant face recognition in uncontrolled scenarios. In: Biometrics (ICB), 2016 International Conference on IEEE. (2016) 1–8

[22] Grgic, M., Delac, K., Grgic, S.: SCface–surveillance cameras face database. Multimedia tools and applications **51**(3) 863–879.

[23] Yi, D., Lei, Z., Liao, S., Li, S.Z.: Learning face representation from scratch. arXiv preprint arXiv:1411.7923 (2014)

[24] Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The FERET evaluation methodology for face-recognition algorithms. IEEE Transactions on pattern analysis and machine intelligence **22**(10) (2000) 1090–1104

[25] Gao, W., Cao, B., Shan, S., Chen, X., Zhou, D., Zhang, X., Zhao, D.: The CAS-PEAL large-scale Chinese face database and baseline evaluations. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans **38**(1) (2008) 149–161

[26] Phillips, P.J., Flynn, P.J., Scruggs, T., Bowyer, K.W., Chang, J., Hoffman, K., Marques, J., Min, J., Worek, W.: Overview of the face recognition grand challenge. In: Computer vision and pattern recognition, IEEE computer society conference on. vol. 1, pp. 947–954 (2005)

[27] Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S.: Multi-pie. Image and Vision Computing **28**(5) (2010) 807–813

[28] Milborrow, S., Morkel, J., Nicolls, F.: The MUCT landmarked face database. Pattern Recognition Association of South Africa **201**(0) (2010)

[29] Faces94. http://cswww.essex.ac.uk/mv/allfaces/faces94.html.

[30] Martinez, A.M.: The AR face database. CVC Technical Report24 (1998)

[31] Sim, T., Baker, S., Bsat, M.: The CMU pose, illumination, and expression (PIE) database. In: Automatic Face and Gesture Recognition, fifth IEEE International Conference on. (2002) 53–58

[32] ORL. http://www.cam-orl.co.uk/.

[33] Gourier, N., Letessier, J.: The Pointing 04 data sets. In: Proceedings of Pointing 2004, ICPR International Workshop on Visual Observation of Deictic Gestures. (2004) 1–4

[34] Grimace. http://cswww.essex.ac.uk/mv/allfaces/grimace.zip.

[35] Sun, Y., Wang, X., Tang, X.: Deep convolutional network cascade for facial point detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2013) 3476–3483

[36] Jian, M., Lam, K. M.: Simultaneous hallucination and recognition of low-resolution faces based on singular value decomposition IEEE Transactions on Circuits and Systems for Video Technology **25**(11) (2015) 1761–1772

[37] Lu, Z., Jiang, X., Kort, A.: Deep coupled resnet for low-resolution face recognition IEEE Signal Processing Letters **25**(4) (2018) 526–530