2017 IEEE 12th International Conference on Automatic Face & Gesture Recognition

Noisy Face Image Sets Refining Collaborated with Discriminant Feature Space Learning

Xin Liu[1,2], Meina Kan[1,2], Shiguang Shan[1,2,3], Xilin Chen[1,2]
[1] Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS),
Institute of Computing Technology, CAS, Beijing, 100190, China
[2] University of Chinese Academy of Sciences, Beijing 100049, China
[3] CAS Center for Excellence in Brain Science and Intelligence Technology

***Abstract*— Large-scale face data together with deep learning technology have significantly improved the performance of face recognition in the wild. Hereinto, the large-scale face data plays a fundamental role, and it is nontrivial to collect a large-scale face dataset with accurate class labels. No wonder it is quite money and effort consuming by collecting manually, however it is easy to access large scale face images by using a search engine with names as keywords. Unfortunately, the retrieved face images from search engine are usually messed up with some noise images with wrong labels, which forms a great need of developing algorithms to refine the retrieved noisy face image set. In this work, we propose a joint framework in which multiple noisy face image sets refining collaborates with the discriminant feature space learning. Specifically, the two modules, refining each noisy face image set by conducting one-class classification based on learnt discriminant feature and learning discriminant feature space based on refined face image sets, are updated iteratively inducing an effective refinement model. To investigate the proposed method, we collect a real-world dataset for the evaluation including 15,515 images of 46 subjects with $40\% \sim 63.5\%$ noise images per subject. The experimental results demonstrate that state-of-the-art one-class classification methods can be significantly improved when being embedded in the proposed framework, and the proposed framework exhibits strong robustness even when the mean noise proportion is up to $50\% \sim 80\%$.**

I. Introduction

Nowadays, big face data together with deep learning techniques [2], [4] achieve state-of-the-art performance in real-world face recognition [12], [3], [13], [1]. A few large-scale face datasets such as SCF [3], CelebFaces+ [13] and CASIA-WEB [1] have been collected to facilitate the training of DCNN. As a result, large-scale real-world face dataset with label annotation becomes the basis for face recognition research.

One common way of building a large-scale face dataset is to collect face images manually, but it is quite money and effort consuming. Recently, another popular way is to collect face images semi-automatically, i.e., a large-scale candidate face images with noise images are firstly retrieved from the search engine using the person's names as keywords, and then the noisy face image sets are refined manually to remove those noise images. This semi-automatic way has saved a lot cost, but still needs lots of manual effort to eliminate those noise images. Therefore, it is in great need to develop techniques that can fully automatically refine the *Noisy Face Image Set* (NFIS) to achieve large-scale and clean face images set, and this is called as NFIS refining in this work. To make this problem more general, we assume that only the visual information of NFIS is available, and neither the initial rank-order list nor text around the images is provided.

The NFIS refining seems related with face clustering [15], [16] and face tagging [17], [14], [18], but they are quite different. Face clustering attempts to split a whole image set into clusters, with each cluster containing only those images of one subject. Face tagging intends to associate the faces in an image with names appeared in the surrounding text. In the NFIS refining problem this work focuses on, each of a few subjects is associated with an noisy image set, which contains images of this subject but corrupted with some irrelevant images, and the goal is to remove those irrelevant images in each noisy face images set. It should be noticed noisy images in NFIS means wrong label, which is different from the image-level noise studied in [26], [27].

A straightforward strategy is turning to a little human effort. [20] proposes an active learning approach that allows human to label a small number of representative images to benefit the refining. [1] employs a semi-automatic pipeline to refine noisy face image set with a few pre-labeled images as seeds. Some other pioneer works exploit auxiliary information to facilitate the refining, such as the surrounding text, initial rank-list [21], [22]. Among these works, [21] fuses both the text and visual feature to re-rank the noisy image set. [22] adopts an incremental model learning approach to re-rank the noisy image set, started from the initial rank-order list returned by the search engine.

In many more challenging scenarios, neither manual effort nor auxiliary information is available, leaving the visual information as the only information. [19] proposes an extended probabilistic latent semantic analysis method (pLSA) encoding the spatial information of visual word to learn object classifier from the noisy image sets. In [11], an unsupervised label refinement approach is proposed by forming the visual relevance of images in a graph-based and low-rank learning framework. These two methods model all the noisy image sets jointly. Another promising paradigm attempts to refine each noisy image set independently, called as one-class classification, and some representative works include Locally One-Class SVM [28], UOCL [6], RKDE [5] and SMRS

978-1-5090-4023-0/17 $31.00 © 2017 IEEE
DOI 10.1109/FG.2017.138

[7]. Among these works, Locally One-Class SVM learns multiple discriminative hyper-spheres locally for the normal data through a top-down procedure. UOCL jointly learns label assignment and max-margin one-class classifier for the noisy image set. RKDE is a nonparametric density estimation method, of which low density sample has lower confidence of being clean. SMRS is a sparse based representative selection method which assumes that the noise sample only takes part in the representation of itself and very few other samples.

The basic principal of most methods is that those clean images in each NFIS are strongly related to each other while those noise images are weakly or not related to them. This kind of relevance is usually characterized based on the raw feature, which may be inaccurate. Therefore, we proposed a joint framework in which the NFIS refining collaborates with the discriminant feature space learning. Specifically, one-class classification is conducted on each NFIS to determine each sample's confidence of being clean. Furthermore, based on those high confidence samples, the supervised feature space learning method FLDA [8] is adopted to learn discriminant feature, based on which the one-class classification can be conducted in a more discriminant and accurate feature space. The one-class classification and discriminant feature are updated iteratively, and they can benefit each other, leading to an self-taught framework for NFIS refining. We term this approach as *Noisy Face Image Sets Refining Collaborated with Discriminant Feature Space Learning* .

The contributions of our work can be summarized as:

1. We propose a joint framework of *Noisy Face Image Sets Refining Collaborated with Discriminant Feature Space Learning* (NFIS-DFSL) to refine those noisy face image sets. Our approach jointly conducts feature space learning and one-class classification via a self-taught mechanism. The experimental results demonstrate that state-of-the-art one-class classification methods can be significantly improved when being embedded in NFIS-DFSL, exhibiting strong robustness even when the mean noise proportion is up to $50\% \sim 80\%$.

2. A new dataset, *Bing Noisy Face Dataset* (BNF), is collected to study the NFIS refining problem. BNF dataset contains 15,515 images of 46 celebrities retrieved from Microsoft Bing search engine and the ground truth label of each sample is manually labeled for evaluation. To our best knowledge, this is the first real-world dataset to study the NFIS refining.

The rest of this paper is organized as follows: section 2 formally describes NFIS refining, discusses the robustness of FLDA to random noise, and introduces our proposed framework of NFIS-DFSL. Section 3 presents a new collected dataset, i.e., *Bing Noisy Face Dataset* (BNF) and the experimental evaluation, and finally section 4 concludes this work.

II. Noisy Face Image Sets Refining Collaborated with Discriminant Feature Space Learning

This section firstly gives a formal definition of *Noisy Face Image Set* (NFIS) refining and then discusses the robustness of Fisher Linear Discriminant Analysis (FLDA) to random noise which will be exploited for discriminant feature space learning in our framework, and finally presents our joint framework of NFIS-DFSL.

A. Problem Description

Formally, let us assume there are C NFISs, denoted as $\{X_i\}_{i=1}^C$. For the i-th NFIS with N_i samples, its data matrix is denoted as $X_i = [x_{i1}, x_{i2}, ..., x_{iN_i}] \in \mathbb{R}^{d\times N_i}$, which is obtained from the search engine with a subject name as the keyword. The rank-order from the search engine for X_i is assumed to be unavailable in this work. Inevitably, X_i contains a few images that are irrelevant to the given subject, called as noise images. The goal of this problem is to determine those noise images, so as to achieve a set of clean face images for this subject. It is generally formulated as a re-ranking problem in most existing works, i.e., re-rank X_i as an ordered image set $Z_i = [x_{s_{i1}}, x_{s_{i2}}, ..., x_{s_{iN_i}}]$, in which the face images are ranked in descending order according to their confidence score of being clean.

In this work, the Precision-Recall curve is exploited as the measurement to evaluate the performance of NFIS refining. The Precision-Recall curve of multiple NFISs is calculated by the average of all the Precision-Recall curves. Besides, the Area under Curve (AUC) of the Precision-Recall curve is exploited as another measurement.

B. Analysis of the Robustness of FLDA to Random Noise

FLDA [8] is one of the most typical supervised feature space learning approach, and in this section we show that FLDA is robust to random noise which makes it a favorable choice of discriminant feature space learning in our framework for NFIS refining problem.

For each NFIS X_i, those clean and noise images are denoted as $X_i^R \in \mathbb{R}^{d\times N_i^R}$ and $X_i^O \in \mathbb{R}^{d\times N_i^O}$ respectively. μ, μ^R, μ_i^R and μ_i^O represent the mean of all samples from C NFISs, the mean of all clean samples from C NFISs, the mean of all clean samples of X_i, and the mean of all noise samples of X_i respectively. Suppose each sample in $\{X_i\}_{i=1}^C$ has been subtracted by the mean of all samples and makes $\mu = 0$. As we assume the noise images are totally randomly sampled, it is safe to assume the mean of noise images, i.e., expectation of μ_i^O to be zero. Furthermore, the expectation of μ_i and μ^R can be reformulated as below:

$$E(\mu_i) = \frac{E(N_i^R\mu_i^R + N_i^O\mu_i^O)}{N_i} = \frac{N_i^R}{N_i}\mu_i^R,$$
$$E(\mu^R) = E\left(\frac{\sum_{i=1}^C (N_i\mu - N_i^O\mu_i^O)}{\sum_{i=1}^C N_i^R}\right) = 0 = u, \quad (1)$$

with $N_i = N_i^R + N_i^O$. Eq. (1) means that μ^R has the same expectation as μ and μ_i also shares same expectation as μ_i^R but with a different scale. With this, the expectation of between-class scatter can be reformulated as follows:

$$\begin{aligned} E(S_B) &= \sum_{i=1}^C N_i E((\mu_i - \mu)(\mu_i - \mu)^T) \\ &= \sum_{i=1}^C \frac{N_i^R}{N_i} N_i^R (\mu_i^R - \mu^R)(\mu_i^R - \mu^R)^T. \quad (2) \end{aligned}$$

Eq. (2) demonstrates that the between-class scatter calculated with noise samples approximately equals to that calculated with only clean samples but with a different scale, which means between-class scatter is not affected by the random noise images.

Similarly, the total scatter S_T can be rewritten as follows:

$$S_T = \sum_{i=1}^{C} \sum_{x_k \in X_i^R} (x_k - \mu)(x_k - \mu)^T + \sum_{i=1}^{C} \sum_{x_k \in X_i^O} (x_k - \mu)(x_k - \mu)^T. \quad (3)$$

In Eq. (3), the first term is S_T of those clean samples of the C subjects, which can be considered as a specific S_T from C subjects, and the second term is the S_T of a few random noise images, which can be considered as a random sampling of the S_T of all population. Based on the assumptions that these noise images are totally random, it can be seen as a random sampling of all facial images from multiple subjects. So the scatter of the random noise is just slightly discriminative, and will not be biased which can be seen as a regularization term. In summary, based on C NFISs, the between-class scatter S_B is not affected by the noise images and total scatter S_T is actually a regularization of S_T from only those clean images.

These characteristics demonstrate that FLDA is robust to random noise, and considering its favorable robustness, FLDA is adopted for discriminant feature space learning in our joint framework for NFIS problem. We will prove the robustness of FLDA to random noise in the experimental section.

C. Proposed Method

In this subsection, we tackle the NFIS refining by proposing the NFIS-DFSL. The feature space learning and NFIS refining are jointly conducted in NFIS-DFSL via a self-taught mechanism. By design, NFIS-DFSL adopts the FLDA to learn discriminant feature on the higher confidence subsets produced by the one-class classification method and conducts one-class classification in the FLDA feature space to update each sample's confidence score of being clean again. NFIS-DFSL method teach itself to choose the more confident samples and update the confidence model in an iterative method via FLDA learning and one-class classification. An overview of the joint framework of NFIS-DFSL is presented in Figure 1, consisting of three components, initialization of discriminant feature representation, refining each NFIS given the discriminant feature representation, and learning discriminant feature representation given refined NFISs.

Initialization of discriminant feature representation. To initialize a discriminant feature representation for the following refinement of NFIS, the first that springs to mind is directly applying FLDA on the images from NFISs including those noise images as it is robust to random noise, as follows:

$$W = arg\max_{W}(|W^T S_B W| / |W^T S_T W|). \quad (4)$$

With discriminant projection W, the discriminant feature of the i-th NFIS can be obtained as $W^T X_i$. An alternative is using the raw feature space directly, where W can be seen as an identity matrix. We will compare these two initialization strategies in Sec. III-C.

Given discriminant feature representation, refine each NFIS. As each NFIS distributes differently from the others, they are refined separately based on the discriminant feature representation aiming for better refinement. For each NFIS, the one-class classification method is exploited to refine it as follows:

$$f_{OC_i}(W^T X_i), \ i = 1, 2, ..., C, \quad (5)$$

where OC indicates the one-class classification method and f_{OC_i} denotes the one-class classier which outputs each sample's confidence score of being clean for the i-th NFIS. Then each NFIS is re-ranked according to its confidence score in descending order. As a joint framework, NFIS-DFSL can accompany any kind of one-class classification method, such as [6], [7], [5].

Given refined NFISs, learn discriminant feature. In the re-ranked NFIS, those samples in front have higher confidence of being clean, and the subset only with those samples with higher confidence from all subjects are much less noisy. Based on the subset which is much less noisy, a better discriminant feature can be achieved. Specifically, the first $K\%$ samples from each re-ranked NFIS are merged together to train the FLDA model denoted as W_K according to Eq. (4).

Repeating. NFIS-DFSL repeats NFIS refining and discriminant feature space learning iteratively, and each NFIS is expected to be refined more and more accurately. The parameter K will be small at the first iteration, and becomes larger with more iterations until it reaches a pre-defined threshold. In this work, K is increased with step of 5 until all the samples are selected, i.e, $K, 2K, 3K, ..., 19K, 20K$, but other types of increasing step is also applicable.

The FLDA training in anterior iteration uses fewer but cleaner samples, and the FLDA training in the posterior iteration uses more but maybe noisy samples. However, it is hard to determine which iteration to stop or output the best performance. Besides, to make use of the complementary of the models from each iteration, we propose to ensemble all one-class classifiers from all iterations to form a more robust once-class classifier as below:

$$F_{OC_i}(X_i) = \sum_{k \in \{K, 2K, 3K, ..., 20K\}} f_{OC_{ik}}(W_k^T X_i), \ i = 1, 2, ..., C. \quad (6)$$

The ensemble classifier in Eq. (6) actually increases the weight of those samples with higher confidence score of being clean, i.e., those samples have more chance to participate the training of FLDA model. This strategy is more flexible and robust leading to better result, as evaluated in sec. 3.3. The detailed procedures of NFIS-DFSL is presented in Algorithm 1.

III. Experimental Results

In this section, we firstly present a newly collected benchmark, *Bing Noisy Face Dataset* (BNF), for the evaluation of NFIS refining. Then, we evaluate the proposed joint

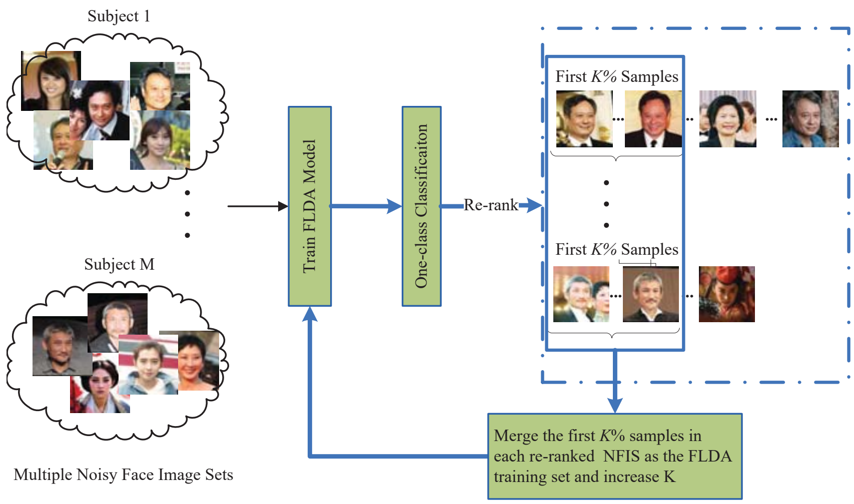

Fig. 1. Overview of the NFIS-DFSL.

Algorithm 1 NFIS-DFSL.

Input:

The C NFISs: $\{X_i\}_{i=1}^{C}$. One-class classification method: OC.

The set of K in ascending order: S_K:, e.g., $K = 5$, $S_K = \{K, 2K, 3K, ..., 20K\}$

Output:

The C re-ranked NFISs

Initialize the discriminant feature space according to Eq. (4) and conduct one-class classification to re-rank each NFIS.
for each K in S_K **do**
  For each re-ranked NFIS, select the first $K\%$ samples in front and use all these selected samples to re-train a FLDA model denoted as W_K.
  In the learned FLDA feature space, conduct one-class classification to re-rank each NFIS. Denote $f_{OC_{iK}}(W_K^T X_i)$ as one-class classifier for the i-th NFIS in this iteration.
end for
Calculate the ensemble one-class classifier $F_{OC_i}(X_i)$ for the i-th NFIS according to Eq. (6), and use $F_{OC_i}(X_i)$ to re-rank the i-th NFIS, where $i = 1, 2, ..., C$.
return The C re-ranked NFISs.

---

framework of NFIS-DFSL on this real-world dataset and a synthetic dataset.

A. Evaluation Datasets

To collect a real-world noisy face dataset for evaluating techniques aiming for NFIS refining problem, we use the Microsoft Bing search engine to crawl face images using the names of 46 celebrities. For each subject, the crawler automatically downloads $300 \sim 500$ images and removes broken and duplicate images. For the raw images, two preprocessing steps are applied, i.e., face detection and face alignment. In this study, the OpenCV face detector [23] and the CFAN face alignment toolbox [9] are employed to align the raw images into 300×300 pixel. Then manual identity annotation is done with the help of three volunteers. For each person, an image is labeled as clean if all volunteers agree in that, and similarly an image is labeled as noise if all volunteers agree in that. Those images with disagreements among all volunteers will be dropped directly. Totally, we get 15, 515 images of 46 NFISs and term this dataset as *Bing Noisy Face Dataset* (BNF).

Fig. 2. Examples of NFIS in BNF. (a) Jacee Chen; (b) Jackie Chen. Images with green boarder are clean ones and images with red boarder are noise ones.

Fig. 2 shows two example NFISs in BNF. The noise samples in BNF dataset can be either random outlier or mislabeled sample:

1) Random outlier, e.g, the second noise image in Fig. 2-(b).

2) Mislabeled sample, e.g, the first noise image in Fig. 2-(a), who is actually the subject in Fig. 2-(b).

The noise proportion of BNF dataset ranges in [0.4, 0.635], which forms a very challenging dataset for NFIS refining. However, to investigate the problem in more challenging case, a subset with noise proportion ranging in [0.5, 0.635] is selected, denoted as BNF-Hard, which includes 10,085 images from 23 subjects.

Furthermore, to investigate the NFIS refining in extremely challenging case, a synthetic dataset is used. Firstly, we collect images of another 1,001 celebrities with Bing search engine as 1,001 NFISs. It takes less than 5 minutes to label a noisy face set with about 300 face images with the first 200 images sorted by NFIS-DFSL being manually checked. While it takes about 15 minutes if labeled without the assistance of NFIS-DFSL method to get the same number of clean face images. Secondly, the NFIS of these 1,001 subjects are re-ranked using our proposed NFIS-DFSL method, and the first 200 images of each NFIS are manually checked to remove those noise images. Finally, those clean images of the first 200 images from all subjects form as a *Bing Celebrity Face Dataset* (BCF). In summary, BCF dataset contains 181,373 images of 1,001 persons with no noise images for each subject. To build a synthetic noisy dataset which is extremely challenging, 100 persons with 100 images per person are selected, and images randomly from the rest 901 persons are added into each subject as noise images. In this way, two synthetic datasets are established with percentage of noise images as 60% and 80% respectively, and the two datasets are termed as BCF-N60 and BCF-N80. Table. I presents the statistics of all four evaluation sets in BNF and BCF.

B. Experimental Settings

Face Representation:. In all experiments, the face images are normalized into 80×64 pixel with two eye center locations fixed at (17, 31) and (41, 31). Rather than the raw grey pixel, the Local Binary Patterns (LBP) feature [24] known as a popular local feature for face representation is employed for more expressive representation. Specifically, the face image is divided into 10*8 cells, and then, the uniform LBP pattern [10] is extracted leading to the LBP feature of 4720 dimensions. Finally, the square root of the LBP is used as the feature representation following the suggestions in [25].

One-class Classification Methods: The methods of UOCL[6], RKDE[5] and SMRS[7] are the three state-of-the-art one-class classification methods, and our method is compared with them. As our method is a joint framework, any one-class classification method can be embedded in it. Therefore, UOCL, RKDE and SMRS are respectively embedded in our method for a fair comparison with these methods. The parameters of each method are set as same as the embedded one in our method. For UOCL, the Gaussian Kernel $k(x,y) = exp(-\|x-y\|^2/2\sigma^2)$ is employed with $\sigma = \sum_{i,j=1}^{n} \|x_i - x_j\|^2/n^2$ and k of the KNN graph is set as 6 in terms of squared Euclidean distance. In RKDE, the bandwidth is chosen via least square cross validation. For SMRS, the *row sparsity index* (rsi) is adopted, of which lower rsi means higher confidence of being clean.

C. Evaluation of the proposed NFIS-DFSL

This subsection compares the proposed NFIS-DFSL method with three state-of-the-art one-class classification methods, i.e., UOCL, RKDE and SMRS. The evaluations of all methods are shown in Fig. 3 and Table II. Fig. 3-(a) compare our method with UOCL. Among the comparison methods, UOCL-RAW directly conduct the UOCL in the LBP feature space, and it performs the worst as the LBP is not discriminant and thus cannot promise a good result of UOCL. Furthermore, UOCL-FLDA0 is conducted in the FLDA0 feature space, which performs much better as the discriminant feature can significantly facilitate the refinement problem. Here, FLDA0 is the initial discriminant feature space achieved by using all NFISs. Moreover, our method NFIS-FDL outperforms both UOCL-raw and UOCL-FLDA0 benefitted from the iterative update of both feature space learning and NFIS refining. Two initialization strategies of the feature representation are employed for our methods, i.e., raw LBP feature and FLDA0 feature, and as seen UOCL-NFIS-DFSL(RAW) and UOCL-NFIS-DFSL(LDA0) are comparative to each other on average. Similarly, our NFIS-DFSL is also compared with RKDE and SMRS as shown in Fig. 3-(b) and Fig. 3-(c) respectively. From these comparisons, the same conclusion can be obtained as follows:

1) The state-of-the-art one-class classification methods, i.e., UOCL, RKDE and SMRS can be significantly improved when being embedded in NFIS-DFSL, as the discriminant feature space learning and NFIS refining can benefit each other in an iterative way.

2) The NFIS-DFSL with RKDE embedded achieves the best performance on average on the four evaluation sets, as shown in Table II. The experimental results also demonstrate that, for both real-world BNF and synthetic BCF dataset, the bigger the minimum noise proportion, the larger performance improvement NFIS-FDL achieves.

3) FLDA is robust to noise. Although the mean of noise images may be not exactly zero, it approximates zero with promising performance. This can be proved by that the UOCL-FLDA0, RKDE-FLDA0 and SMRS-FLDA0 significantly outperform UOCL-RAW, RKDE-RAW and SMRS-RAW. As shown in Table II, the performance of all one-class classification methods conducted in FLDA0 is double of that in raw LBP feature space on BCF-N80 with 80% noise.

4) The experimental results also demonstrate that the real-world dataset can be more challenging. As shown in Table. II, both the one-class classification methods conducted in FLDA0 and our NFIS-DFSL perform better on synthetic BCF-N80 with 80% noise than on real-world BNF-Hard with 56.46% noise.

To further investigate the self-taught mechanism of NFIS-DFSL, Fig. 4 demonstrates the intermediate results for a NFIS during different iterations. The UOCL-NFIS-DFSL(RAW) method and BNF-Hard dataset are employed. With the increase of iteration, the re-ranked NFIS become more and more accurate. NFIS-DFSL method teaches itself to choose the more confident samples and update the confidence model in Eq. (6) in an iterative method via FLDA learning and one-class classification.

TABLE I

STATISTICS OF DIFFERENT EVALUATION SETS IN BNF AND BCF

Evaluation Set	#Identity	#Samples	Range of Noise Proportion	Mean Proportion
BNF	46	15,515	[0.4, 0.635]	0.5112
BNF-Hard	23	10,085	[0.5, 0.635]	0.5646
BCF-N60	100	50,000	[0.6, 0.6]	0.6
BCF-N80	100	25,000	[0.8, 0.8]	0.8

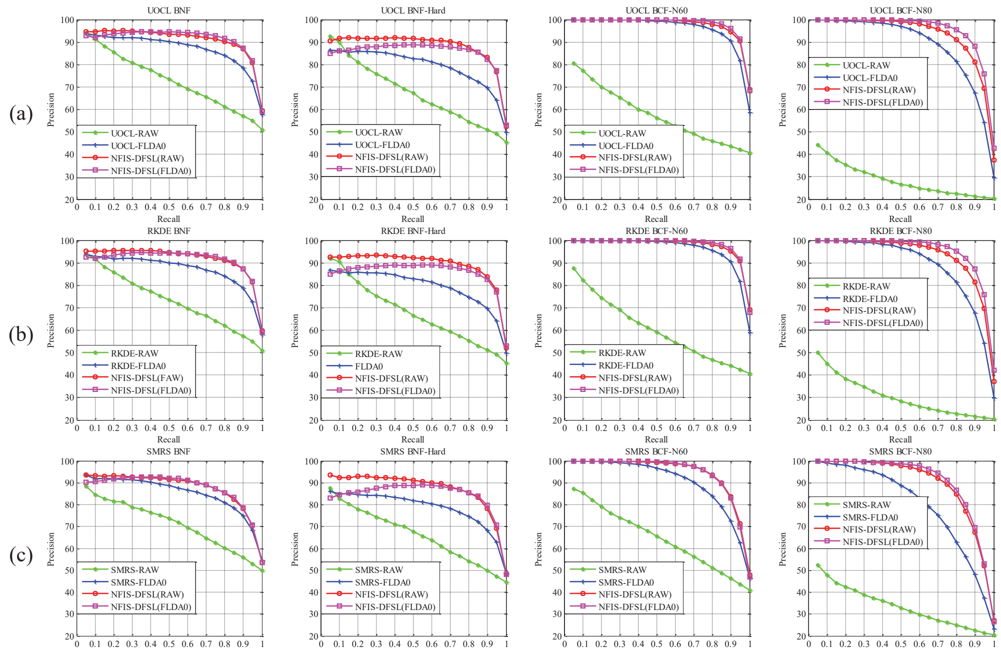

Fig. 3. Comparisons of our NFIS-DFSL with state-of-the-art one-class classification methods: a) UOCL; b) RDKE; c) SMRS.

TABLE II

COMPARISONS OF THE AUC OF ONE-CLASS CLASSIFICATION METHODS AND NFIS-DFSL ON THE FOUR EVALUATION SETS, AND THE BEST PERFORMANCE IN EACH COLUMN IS SHOWN IN BOLD.

Method	BNF	BNF-Hard	BCF-N60	BCF-N80	**Average**
UOCL-RAW	0.7360	0.6831	0.5861	0.3028	0.5770
UOCL-FLDA0	0.8766	0.8023	0.9650	0.8937	0.8844
UOCL-NFIS-DFSL(RAW)	0.9176	0.8885	0.9803	0.9390	0.9313
UOCL-NFIS-DFSL(FLDA0)	0.9184	0.8615	0.9835	**0.9581**	0.9304
RKDE-RAW	0.7389	0.6844	0.6118	0.3206	0.5889
RKDE-FLDA0	0.8765	0.8030	0.9654	0.8942	0.8848
RKDE-NFIS-DFSL(RAW)	**0.9239**	**0.8980**	0.9813	0.9388	**0.9353**
RKDE-NFIS-DFSL(FLDA0)	0.9174	0.8642	**0.9837**	0.9567	0.9305
SMRS-RAW	0.7195	0.6708	0.6559	0.3517	0.5995
SMRS-FLDA0	0.8588	0.7945	0.9102	0.8082	0.8429
SMRS-NFIS-DFSL(RAW)	0.8840	0.8797	0.9498	0.9032	0.9042
SMRS-NFIS-DFSL(FLDA0)	0.8811	0.8509	0.9490	0.9127	0.8984

IV. CONCLUSIONS AND FUTURE WORKS

In this work, we attempt to deal with the *Noisy Face Image Set* (NFIS) refining, which plays an important role in the building of large-scale face dataset. Firstly, we set up the *Bing Noisy Face Dataset* with 15,515 images of 46 NFISs collecting from Microsoft Bing Search Engine. Furthermore, we propose a joint framework of *Noisy Face Image Sets Refining Collaborated with Discriminant Feature Space Learning* (NFIS-DFSL) to deal with NFIS refining problem, which conducts feature space learning and refinement of NFISs in a self-taught mechanism. The experimental results demonstrate that state-of-the-art one-class classification methods can be significantly improved when being embedded in NFIS-DFSL, and our proposed framework exhibits strong robustness when the mean noise proportion

Fig. 4. The intermediate result for a NFIS during different iterations, and only the top-8 re-ranked images are shown for simplicity. Images with green boarder are clean ones and images with red boarder are noise ones.

is up to 50% $\sim$ 80%.

For future work, our approach can be naturally improved by incorporating the initial rank from search engine. Besides, we will replace FLDA with more advanced feature learning method such as convolution neural network and develop method to estimate the noise proportion of the NFIS.

V. Acknowledgement

This work was partially supported by 973 Program under contract No. 2015CB351802, Natural Science Foundation of China under contracts Nos. 61390511, 61650202, 61402443, 61272321, and the Strategic Priority Research Program of the CAS (Grant XDB02070004).

References

[1] Yi, D., Lei, Z., Liao, S., Li, S.Z, "Learning face representation from scratch", *arXiv preprint arXiv:1411.7923*, 2014
[2] Krizhevsky, A., Sutskever, I., Hinton, G.E, "Imagenet classification with deep convolutional neural networks", *Advances in Neural Information Processing Systems*, pp 1097-1105, 2012
[3] Taigman, Y., Yang, M., Ranzato, M., Wolf, L, "Deepface: Closing the gap to human-level performance in face verification", *IEEE Conference on Computer Vision and Pattern Recognition*, pp 1701-1708, 2014
[4] Jia, Yangqing and Shelhamer, Evan and Donahue, Jeff and Karayev, Sergey and Long, Jonathan and Girshick, Ross and Guadarrama, Sergio and Darrell, Trevor, "Caffe: Convolutional architecture for fast feature embedding", *Proceedings of the ACM International Conference on Multimedia*, pp 675-678, 2014
[5] Kim, JooSeuk and Scott, Clayton D, "Robust kernel density estimation", *The Journal of Machine Learning Research*, vol. 13, pp 2529-2565, 2012
[6] Liu, Wei and Hua, Gang and Smith, John R, "Unsupervised one-class learning for automatic outlier removal", *IEEE Conference on Computer Vision and Pattern Recognition*, pp 3826-3833, 2014
[7] Elhamifar, Ehsan and Sapiro, Guillermo and Vidal, Rene, "See all by looking at a few: Sparse modeling for finding representative objects", *IEEE Conference on Computer Vision and Pattern Recognition*, pp 1600-1607, 2012
[8] Belhumeur, P.N. and Hespanha, J.P. and Kriegman, D.J., "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection", *IEEE Transactions on Pattern Analysis and Machine Intelligence* , vol. 19, pp 711-720, 1997
[9] Zhang, Jie and Shan, Shiguang and Kan, Meina and Chen, Xilin, "Coarse-to-fine auto-encoder networks (cfan) for real-time face alignment", *European Conference on Computer Vision* , pp 1-16, 2014
[10] Ojala, Timo and Pietikainen, Matti and Maenpaa, Topi, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns", *IEEE Transactions on Pattern Analysis and Machine Intelligence* , vol. 24, pp 971-987, 2002
[11] Wang, Dayong and Hoi, Steven CH and He, Ying and Zhu, Jianke, "Mining weakly labeled web facial images for search-based face annotation", *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, pp 166-179, 2014
[12] Gary B. Huang, Manu Ramesh, Tamara Berg and Erik Learned-Miller, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments", *Techincal Report of University of Massachusetts, Amherst*, 2007
[13] Sun, Yi and Wang, Xiaogang and Tang, Xiaoou, "Deep learning face representation from predicting 10,000 classes", *IEEE Conference on Computer Vision and Pattern Recognition*, pp 1891-1898, 2014
[14] Zhang, Xiao and Zhang, Lei and Wang, Xin-Jing and Shum, Heung-Yeung, "Finding celebrities in billions of web images", *IEEE Transactions on Multimedia*, vol. 14, pp 995-1007, 2012
[15] Cui, Jingyu and Wen, Fang and Xiao, Rong and Tian, Yuandong and Tang, Xiaoou, "EasyAlbum: an interactive photo annotation system based on face clustering and re-ranking", *IEEE Conference on Computer Vision and Pattern Recognition*, pp 367-376, 2007
[16] Zhu, Chunhui and Wen, Fang and Sun, Jian, "A rank-order distance based clustering algorithm for face tagging", *IEEE Conference on Computer Vision and Pattern Recognition*, pp 481-488, 2011
[17] Ozcan, Mert and Luo, Jie and Ferrari, Vittorio and Caputo, Barbara, "A Large-Scale Database of Images and Captions for Automatic Face Naming", *British Machine Vision Conference*, pp 1-11, 2011
[18] Kapoor, Ashish and Hua, Gang and Akbarzadeh, Amir and Baker, Simon, "Which faces to tag: Adding prior constraints into active learning", *IEEE Conference on Computer Vision and Pattern Recognition*, pp 1058-1065, 2009
[19] Fergus, Robert and Fei-Fei, Li and Perona, Pietro and Zisserman, Andrew, "Learning object categories from Google's image search", *International Conference on Computer Vision*, pp 1816-1823, 2005
[20] Collins, Brendan and Deng, Jia and Li, Kai and Fei-Fei, Li, "Towards scalable dataset construction: An active learning approach", *European Conference on Computer Vision*, pp 86-98, 2008
[21] Schroff, Florian and Criminisi, Antonio and Zisserman, Andrew, "Harvesting image databases from the web", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, pp 754-766, 2011
[22] Li, Li-Jia and Fei-Fei, Li, "Optimol: automatic online picture collection via incremental model learning", *International Journal of Computer Vision*, vol. 88, pp 147-168, 2010
[23] Bradski, Gary and Kaehler, Adrian, "Learning OpenCV: Computer vision with the OpenCV library", *O'Reilly Media, Inc.*, 2008
[24] Ahonen, Timo and Hadid, Abdenour and Pietikäinen, Matti, "Face recognition with local binary patterns", *European Conference on Computer Vision*, pp 469-481, 2004
[25] Nguyen, Hieu V and Bai, Li, "Cosine similarity metric learning for face verification", *Asian Conference on Computer Vision*, pp 709-720, 2010
[26] Li, Liangyue and Li, Sheng and Fu, Yun, "Learning low-rank and discriminative dictionary for image classification", *Image and Vision Computing*, pp 814-823, 2014
[27] Li, Sheng and Fu, Yun, "Robust subspace discovery through supervised low-rank constraints", *Proceedings of the 2014 SIAM International Conference on Data Mining*, pp 163-172, 2014
[28] Li, Sheng and Shao, Ming and Fu, Yun, "Locality linear fitting one-class SVM with low-rank constraints for outlier detection", *International Joint Conference on Neural Networks*, pp 676-683, 2014