ITERATIVE DYNAMIC GENERIC LEARNING FOR SINGLE SAMPLE FACE RECOGNITION WITH A CONTAMINATED GALLERY

Meng Pang, Yiu-ming Cheung*, Qiquan Shi†, and Mengke Li**

*Department of Computer Science, Hong Kong Baptist University, Hong Kong SAR
†Huawei Noah's Ark Lab, Hong Kong SAR

ABSTRACT

This paper studies a new challenging problem in face recognition (FR) with single sample per person (SSPP), i.e., SSPP FR with a contaminated gallery (SSPP-CG FR), where the gallery is contaminated by variations. In SSPP-CG FR, the popular generic learning methods will suffer serious performance degradation because the applied prototype plus variation (P+V) model is not suitable in such scenarios. The reasons are twofold: 1) The contaminated gallery samples yield bad prototypes to represent the persons; 2) The generated variation dictionary is simply based on the subtraction of average face from generic samples of the same person and cannot well depict the intra-personal variations. To tackle SSPP-CG FR, we propose a novel Iterative Dynamic Generic Learning (IDGL) method, where the labeled gallery and unlabeled query sets are fed into a dynamic label feedback network for learning. Specifically, IDGL first recovers the *prototypes* via a semi-supervised low-rank representation (SSLRR) framework and learns a representative *variation dictionary* by extracting the "sample-specific" corruptions from an auxiliary generic set. Then, it puts them into the P+V model to estimate labels for query samples. Subsequently, the estimated labels are used as the feedbacks to modify the SSLRR, thus updating new prototypes for the next round of P+V based label estimation. With the dynamic learning network, the accuracy of the estimated labels is improved iteratively owing to the steadily enhanced prototypes. Experiments on various benchmark databases have verified the superiority of IDGL.

Index Terms— Face recognition, single sample per person, low-rank representation, contaminated gallery set

1. INTRODUCTION

Single sample per person face recognition (SSPP FR), i.e. recognizing a person with a single face image only for training, has several attractive multimedia applications such as surveillance security and criminal identification [1]. However, SSPP FR is still one of the most challenging problems in FR due to the unavailability of intra-class information [2].

To address the SSPP FR problem, some attempts have been made in the last decade, which can be roughly grouped into two types: *patch-based methods* and *generic learning methods*. Patch-based methods [2, 3, 4] partition each sample in the gallery set (i.e., training set) into multiple image patches, then perform feature extraction and recognition based on these local patches. For generic learning methods, they usually introduce an auxiliary generic set to provide new and useful information. Typically, Deng *et al.* [5] proposed a superposed sparse representation classification (SSRC)-based P+V model provided that a query face equals its *prototype* plus the *intra-personal variation*. In the P+V model, each prototype is directly approximated by the gallery sample since it is assumed to be standard and variation-free under the assumption. In other words, the P+V model is actually implemented as the gallery plus variation (G+V) model. Besides, the variation dictionary is generated by subtracting the average face from generic samples of each person. Based on the P+V/G+V model, a variety of generic learning methods [6, 7, 8, 9] have been proposed recently to address the SSPP FR problem.

These aforementioned methods have achieved promising performance for SSPP FR provided that each gallery sample is a standard face with neutral expression and under uniform lighting (like an ID photo). However, from the practical viewpoint, the gallery samples can be collected in a less constrained environment. For example, for criminal identification, the suspects can be illegal immigrants, smugglers, or the persons without residence registration. In such cases, the gallery samples (i.e., reference photos) of suspects are hardly acquired through standard photograph, but may be provided by witnesses with unaligned mobile photos or intercepted from the blurred surveillance videos. Therefore, various nuisance variations, e.g., expressions, lightings and disguises, could exist in gallery samples, thus increasing more difficulty for practical SSPP FR. Such a new and practical issue is called SSPP FR with a *contaminated* gallery (SSPP-CG FR).

In SSPP-CG FR, existing methods will suffer heavy performance degeneration. Particularly for patch-based method-

This work was supported in part by NSFC (Grant No. 61672444), in part by Hong Kong Baptist University (HKBU), Research Committee, IG-FNRA 2018/19 (Grant No. RC-FNRA-IG/18-19/SCI/03), in part by the Innovation and Technology Fund of Innovation and Technology Commission of the Government of the Hong Kong SAR (Project No. ITS/339/18), and in part by the SZSTI (Grant No. JCYJ20160531194006833). Yiu-ming Cheung is the corresponding author (email: ymc@comp.hkbu.edu.hk).

978-1-7281-1331-9/20/$31.00

s, discriminative learning and feature extraction from local patches can be sensitive to the variations in contaminated gallery samples [6]. Worse still, some patches may even be corrupted and capture meaningless information of persons. In contrast, generic learning methods usually perform better than patch-based methods because they will introduce useful supplementary information from the generic set. Nevertheless, the P+V model applied in the existing generic learning methods is still not suitable for SSPP-CG FR. The plausible reasons are twofold: *First*, the contaminated gallery samples cannot be treated as proper prototypes for the P+V model; *second*, the variation dictionary in the P+V model is simply based on the subtraction of average face from generic samples of the same person. Under the circumstances, the important variation details can usually be subtracted, as the average face is unable to characterize the neutral image of the person well. Therefore, a query sample will be easily misclassified as a gallery sample with the similar facial variation in these generic learning methods.

To address the above two issues, we propose a novel Iterative Dynamic Generic Learning (IDGL) method for SSPP-CG FR. IDGL is based on a new observation that a face sample is composed of 1) an invariant low rank part (LRP) characterizing the neutral prototype of the person, and 2) the corruptions representing the intra-personal variants. Motivated by this, IDGL learns *proper prototypes* for contaminated gallery samples by recovering their LRPs through a semi-supervised low-rank representation (SSLRR) framework, and learns *representative variation dictionary* by extracting the "sample-specific" corruptions from the auxiliary generic set. Moreover, to *enhance the prototypes*, IDGL constructs a dynamic label feedback network to update the prototypes iteratively.

As shown in Fig. 1, IDGL includes two learning stages, i.e., *prototype learning via SSLRR* and *P+V based label estimation*, and a dynamic label feedback step. In Stage I, with the labeled gallery and unlabeled query sets, we propose the SSLRR framework to learn proper prototypes to represent the persons by recovering the LRP of each gallery sample. In Stage II, rather than simply subtracting the average face, we introduce a "sample-specific" corruption strategy to learn a representative variation dictionary from an auxiliary generic set, which avoids the important variation details being subtracted. Then, with the learned prototypes and learned variation dictionary, we could estimate the labels for query samples to further update the prototype learning process. In a dynamic updating manner, the estimated query labels are used as the feedbacks to modify the label indicator in SSLRR of Stage I, so as to update new and better prototypes.

Benefiting from the positive dynamic learning network, 1) the qualities of the learned prototypes are enhanced because both linear and non-linear variations are gradually decreased and the useful information in the query set is effectively employed; and 2) the accuracy of the estimated labels is improved iteratively owing to the constantly enhanced prototypes. It is worth noting that, after recovering proper prototypes for the contaminated gallery, IDGL can also be applicable in an inductive scenario for online recognition. That is, when a new unlabeled query sample comes, it will be directly fed into the learned prototypes plus learned variation dictionary (i.e., learned P + learned V) model for recognition, which is prone to real-time face retrieval scenarios.

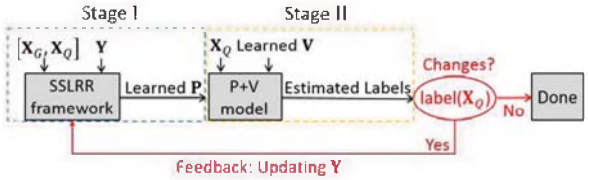


Fig. 1. The flowchart of IDGL, where $\mathbf{X}_G$ and $\mathbf{X}_Q$ denote the labeled gallery and unlabeled query sets, respectively, and $\mathbf{Y}$ is the label indicator matrix.

We highlight the contributions of our work as follows: 1) To the best of our knowledge, this work is among the first to study the new challenging SSPP-CG FR problem, where the gallery set is contaminated by nuisance variations. 2) We propose a novel IDGL method by developing a dynamic label feedback network, to tackle the SSPP-CG FR problem. 3) We present a new "sample-specific" corruption strategy to learn a more representative variation dictionary for the P+V model, compared with the existing generic learning methods.

2. THE PROPOSED IDGL METHOD

We first define some basic notations thereinafter. Let $\mathbf{X} = [\mathbf{X}_G, \mathbf{X}_Q] = [\mathbf{x}_1, \cdots, \mathbf{x}_m, \mathbf{x}_{m+1}, \cdots, \mathbf{x}_n] \in \Re^{d\times n}$ be the sample set matrix, where $\mathbf{X}_G = \{\mathbf{x}_i|_{i=1}^m\}$ and $\mathbf{X}_Q = \{\mathbf{x}_i|_{i=m+1}^n\}$ are the labeled gallery set and unlabeled query set, respectively. The labels of labeled samples are denoted as $y_i \in \{1, 2, \cdots, c\}$, where c is the total number of classes. In SSPP FR, each person has only one single sample, thus m is initialized as c. The label indicator binary matrix $\mathbf{Y} = [\mathbf{Y}_1; \mathbf{Y}_2; \cdots; \mathbf{Y}_n] \in \Re^{n\times c}$ is defined as follows: if $\mathbf{x}_i$ has label $y_i = j$, $\mathbf{Y}_{ij} = 1$; otherwise, $\mathbf{Y}_{ij} = 0$. The auxiliary generic data matrix is defined as $\mathbf{A} = [\mathbf{A}_1, \cdots, \mathbf{A}_s] = [\mathbf{a}_1, \cdots, \mathbf{a}_S] \in \Re^{d\times S}$ ($S = sT$), with s persons *not of interest* and each having T different variations, and $\mathbf{A}_i = [\mathbf{a}_{(i-1)\cdot T+1}, \cdots, \mathbf{a}_{i\cdot T}]$. The prototypes and the variation dictionary to be learned in our IDGL are denoted as $\mathbf{P}$ and $\mathbf{V}$, respectively.

2.1. Stage I: Prototype Learning via SSLRR

In this stage, we aim to learn proper prototypes for contaminated gallery samples by extracting their LRPs through low-rank representation (LRR). However, in SSPP FR, the gallery set only contains single training sample for each person, which makes the existing unsupervised LRR-based methods fail to work in this case due to the extreme lack of training

samples. Based on this consideration, we thus introduce the unlabeled query set into the SSPP-based gallery, and present an SSLRR framework for prototype learning as follows:

$$\begin{aligned}\min_{\mathbf{F},\mathbf{Z},\mathbf{E}} & \sum_{i,j=1}^{n}||\mathbf{F}_i-\mathbf{F}_j||_2^2\mathbf{Z}_{ij}+\lambda_1\sum_{i=1}^{m}||\mathbf{F}_i-\mathbf{Y}_i||_2^2 \\ & +\alpha||\mathbf{Z}||_*+\beta||\mathbf{E}||_{2,1} \\ s.t.\ & \mathbf{X}=\mathbf{X}\mathbf{Z}+\mathbf{E},\end{aligned} \tag{1}$$

where the first two terms encourage the predicted label matrix $\mathbf{F}\in\Re^{n\times c}$ to capture both the label fitness and the manifold smoothness on the semi-supervised graph [10]. $||\mathbf{Z}||_*$ is the nuclear norm of $\mathbf{Z}$ to capture the low rank structure of image data. $\mathbf{E}_{2,1}$ is the $l_{2,1}$ norm of $\mathbf{E}$ that encourages the columns of $\mathbf{E}$ to be zero. λ_1, α and β are the balanced parameters.

The solution of Eq. (1) can be computed based on the linearized alternating direction method with adaptive penalty (LADMAP) [11]. Specifically, we first introduce an auxiliary variable, i.e, $\mathbf{S}$, and convert Eq. (1) to the following problem:

$$\begin{aligned}\min_{\mathbf{F},\mathbf{Z},\mathbf{E},\mathbf{S}} & \sum_{i,j=1}^{n}||\mathbf{F}_i-\mathbf{F}_j||_2^2\mathbf{S}_{ij}+Tr((\mathbf{F}-\mathbf{Y})^T\mathbf{U}(\mathbf{F}-\mathbf{Y})) \\ & +\alpha||\mathbf{Z}||_*+\beta||\mathbf{E}||_{2,1} \\ s.t.\ & \mathbf{X}=\mathbf{X}\mathbf{Z}+\mathbf{E},\mathbf{Z}=\mathbf{S},\end{aligned} \tag{2}$$

where $Tr(.)$ denotes the trace of a matrix, and $\mathbf{U}\in\Re^{n\times n}$ is a diagonal matrix with the first m and the rest $n-m$ diagonal elements as λ_1 and 0, respectively. Then, we can obtain the augmented Lagrangian function of Eq. (2) as follows:

$$\begin{aligned}& L(\mathbf{Z},\mathbf{F},\mathbf{E},\mathbf{S},\mathbf{\Lambda}_1,\mathbf{\Lambda}_2,\mu) \\ =& \sum_{i,j=1}^{n}||\mathbf{F}_i-\mathbf{F}_j||_2^2\mathbf{S}_{ij}+Tr((\mathbf{F}-\mathbf{Y})^T\mathbf{U}(\mathbf{F}-\mathbf{Y})) \\ & +\alpha||\mathbf{Z}||_*+\beta||\mathbf{E}||_{2,1}+Q(\mathbf{Z},\mathbf{E},\mathbf{S},\mathbf{\Lambda}_1,\mathbf{\Lambda}_2,\mu) \\ & -\tfrac{1}{2\mu}(||\mathbf{\Lambda}_1||_F^2+||\mathbf{\Lambda}_2||_F^2)\end{aligned} \tag{3}$$

where $\mathbf{\Lambda}_1$, $\mathbf{\Lambda}_2$ are Lagrange multipliers, $\mu\geq 0$ is a penalty parameter, and $Q(\mathbf{Z},\mathbf{E},\mathbf{S},\mathbf{\Lambda}_1,\mathbf{\Lambda}_2,\mu)=\mu/2(||\mathbf{X}-\mathbf{X}\mathbf{Z}-\mathbf{E}+\mathbf{\Lambda}_1/\mu||_F^2+||\mathbf{Z}-\mathbf{S}+\mathbf{\Lambda}_2/\mu||_F^2)$. We then update the variables $\mathbf{Z}$, $\mathbf{F}$, $\mathbf{E}$ and $\mathbf{S}$ alternately, by minimizing L with other variables fixed. With some algebra, the updating rules for $\mathbf{Z}$, $\mathbf{F}$, $\mathbf{E}$ and $\mathbf{S}$ are as follows:

$$\mathbf{Z}^{k+1}=\mathcal{D}_{\frac{\alpha}{\eta\mu^k}}(\mathbf{Z}^k-\nabla_{\mathbf{Z}}Q(\mathbf{Z}^k,\mathbf{E}^k,\mathbf{S}^k,\mathbf{\Lambda}_1^k,\mathbf{\Lambda}_2^k,\mu^k)/\eta) \tag{4}$$

$$\mathbf{F}^{k+1}=(\mathbf{L}^k+\mathbf{U})^{-1}\mathbf{U}\mathbf{Y} \tag{5}$$

$$\mathbf{E}^{k+1}=\Omega_{\frac{\beta}{\mu^k}}(\mathbf{X}-\mathbf{X}\mathbf{Z}^{k+1}+\mathbf{\Lambda}_1^k/\mu^k) \tag{6}$$

$$\begin{aligned}\mathbf{S}^{k+1}=&\arg\min_{\mathbf{S}}\sum_{i,j=1}^{n}||\mathbf{F}_i^{k+1}-\mathbf{F}_j^{k+1}||_2^2\mathbf{S}_{ij} \\ &+\tfrac{\mu^k}{2}||\mathbf{S}-(\mathbf{Z}^{k+1}+\mathbf{\Lambda}_2^k/\mu^k)||_F^2\end{aligned} \tag{7}$$

where $\eta=||\mathbf{X}||_F^2$, $\nabla_{\mathbf{Z}}Q$ is the partial differential of Q w.r.t. $\mathbf{Z}$, i.e., $\nabla_{\mathbf{Z}}Q=-\mathbf{X}^T(\mathbf{X}-\mathbf{X}\mathbf{Z}^k-\mathbf{E}^k+\mathbf{\Lambda}_1^k/\mu^k)+(\mathbf{Z}^k-\mathbf{S}^k+\mathbf{\Lambda}_2^k/\mu^k)$. $\mathbf{L}\in\Re^{n\times n}$ in Eq. (5) is the graph Laplacian matrix and computed as $\mathbf{L}^k=\mathbf{W}^k-\mathbf{S}^k$, where $\mathbf{W}_{ii}^k=\Sigma_j\mathbf{S}_{ij}^k$. $\mathcal{D}$ and Ω are the singular value thresholding [12] and $l_{2,1}$ minimization operators [13], respectively. Eq. (7) is solved by decomposing it into n independent sub-problems with each having a closed-form solution

$$\mathbf{S}_i=\mathbf{Z}_i^{k+1}+(\mathbf{\Lambda}_2^k-\mathbf{H}_i)/\mu^k, \tag{8}$$

where $\mathbf{H}$ is a n by n matrix whose values are defined as $\mathbf{H}_{ij}=\frac{1}{2}||\mathbf{F}_i^{k+1}-\mathbf{F}_j^{k+1}||_2^2$, and $\mathbf{S}_i$ and $\mathbf{H}_i$ are the i-th ($i=1,\cdots,n$) columns of matrices $\mathbf{S}$ and $\mathbf{H}$, respectively. The algorithm for SSLRR is outlined in **Algorithm 1**. Note that, although it is nontrivial to theoretically prove the convergence for **Algorithm 1**, as the SSLRR involves four iterating variables, i.e., $\{\mathbf{F},\mathbf{Z},\mathbf{E},\mathbf{S}\}$, and the objective function in Eq. (2) is not smooth, **Algorithm 1** can still have good convergence property under mild conditions, according to [13, 14].

After obtaining the optimal solution $\{\mathbf{F}^*,\mathbf{Z}^*,\mathbf{E}^*\}$, the recovered prototype $\mathbf{P}_i$ for the i-th person in the contaminated gallery can be calculated from $\mathbf{X}\mathbf{Z}^*$ w.r.t. the samples predicted as the i-th person.

Algorithm 1 IDGL Stage I: SSLRR

Input: $\mathbf{X}\in\Re^{d\times n}$, $\mathbf{Y}\in\Re^{n\times c}$, $\mathbf{U}\in\Re^{n\times n}$; $\lambda_1,\alpha,\beta>0$

1: Initialization: $\mathbf{Z}^0=\mathbf{S}^0=\mathbf{E}^0=\mathbf{F}^0=\mathbf{\Lambda}_1^0=\mathbf{\Lambda}_2^0=\mathbf{O}$; $\mu^0=0.11$, $\mu^{max}=10^6$, $\rho=1.1$, $\epsilon_1=\epsilon_2=10^{-6}$, $\eta=||\mathbf{X}||_F^2$, $k=0$

2: **while** $||\mathbf{X}-\mathbf{X}\mathbf{Z}^k-\mathbf{E}^k||_F/||\mathbf{X}||_F>\epsilon_1$ or $\mu^k\max(\sqrt{\eta}||\mathbf{Z}^k-\mathbf{Z}^{k-1}||_F,||\mathbf{F}^k-\mathbf{F}^{k-1}||_F,||\mathbf{E}^k-\mathbf{E}^{k-1}||_F,||\mathbf{S}^k-\mathbf{S}^{k-1}||_F)>\epsilon_2$ **do**

3: Update $\mathbf{Z}^{k+1}$, $\mathbf{F}^{k+1}$, $\mathbf{E}^{k+1}$ and $\mathbf{S}^{k+1}$ as Eq. (4)-Eq. (7)

4: Update the multipliers $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ as follows:

$$\mathbf{\Lambda}_1^{k+1}\leftarrow\mathbf{\Lambda}_1^k+\mu^k(\mathbf{X}-\mathbf{X}\mathbf{Z}^{k+1}-\mathbf{E}^{k+1})$$

$$\mathbf{\Lambda}_2^{k+1}\leftarrow\mathbf{\Lambda}_2^k+\mu^k(\mathbf{Z}^{k+1}-\mathbf{S}^{k+1})$$

5: Update the parameter μ as follows:

$$\mu^{k+1}=\min(\mu^{max},\rho\mu^k)$$

6: Update k: $k\leftarrow k+1$.

7: **end while**

Output: An optimal solution $\{\mathbf{F}^*,\mathbf{Z}^*,\mathbf{E}^*,\mathbf{S}^*\}$

Fig. 2. Illustration of the learned variation dictionary of some generic samples from one person on AR database.

2.2. Stage II: P+V based Label Estimation

Variation Dictionary Learning: We present a new way to learn a representative variation dictionary $\mathbf{V}$ from the auxiliary generic set $\mathbf{A}$. Different from the existing methods that simply treat average face as the neutral image and subtract average face from generic samples to generate variations, our

Algorithm 2 IDGL Method

Input: $\mathbf{X} = [\mathbf{X}_G, \mathbf{X}_Q] \in \Re^{d\times n}$, $\mathbf{Y} \in \Re^{n\times c}$, $\mathbf{U} \in \Re^{n\times n}$; $\lambda_1, \lambda_2, \lambda_3, \alpha, \beta, t_{max} > 0$

1: **repeat**
2: Stage I: Learning prototypes $\mathbf{P}$ based on **Algorithm 1**
3: Stage II: Learning variation dictionary $\mathbf{V}$ in Eq. (9)-(10)
4: Stage II: Estimating $label(\mathbf{X}_Q)$ in Eq. (11)-(12)
5: Updating $\mathbf{Y}$ through $label(\mathbf{X}_Q)$
6: **until** the maximum number of iterations τ_{max} is reached or $label(\mathbf{X}_Q)$ is not changed between two successive iterations

Output: Estimated labels for the query set, i.e., $label(\mathbf{X}_Q)$

method models the neutral image by the class-specific low-rank part (LRP) and uses the rest part (i.e., sample-specific corruptions) as the variations. The LRP is more suitable to represent the neutral image than the average face, and enables the important variation details not to be subtracted. Specifically, for each generic subset of the i-th class, i.e., $\mathbf{A}_i \in \Re^{d\times T}$, we solve the following LRR-based optimization problem:

$$\min_{\mathbf{L}_i, \mathbf{V}_i} ||\mathbf{L}_i||_* + \lambda_2||\mathbf{V}_i||_{2,1}, \quad s.t.\ \mathbf{A}_i = \mathbf{A}_i\mathbf{L}_i + \mathbf{V}_i, \tag{9}$$

where $\mathbf{A}_i\mathbf{L}_i$ describe the LRPs of generic samples for the i-th class, while $\mathbf{V}_i$ model the "sample-specific" corruptions that can be treated as the intra-personal variations. Hence, the learned variation dictionary $\mathbf{V}$ is formed as

$$\mathbf{V} = [\mathbf{V}_1, \cdots, \mathbf{V}_s] \in \Re^{d\times S}. \tag{10}$$

Fig. 2 illustrates the learned variation dictionary of some generic samples from one person on AR database, where we observe that it has intuitive explanations and can well characterize the variations such as expressions, lightings and disguises (sunglasses and scarf).

P+V model: Based on the learned variation dictionary $\mathbf{V}$ and the learned prototypes $\mathbf{P}$ in Stage I, we then perform label estimation for the query set, i.e., $\mathbf{X}_Q = \{\mathbf{x}_i|_{i=c+1}^{n}\}$. Specifically, for each query sample $\mathbf{x}_i$, we solve the P+V model based minimization problem as follows:

$$\begin{bmatrix}\boldsymbol{\theta}^* \\ \boldsymbol{\varphi}^*\end{bmatrix} = \arg\min_{\boldsymbol{\theta},\boldsymbol{\varphi}} \left\|\mathbf{x}_i - [\mathbf{P}\ \mathbf{V}]\begin{bmatrix}\boldsymbol{\theta} \\ \boldsymbol{\varphi}\end{bmatrix}\right\|_2^2 + \lambda_3\left\|\begin{bmatrix}\boldsymbol{\theta} \\ \boldsymbol{\varphi}\end{bmatrix}\right\|_1, \tag{11}$$

where $\boldsymbol{\theta} \in \Re^{c\times 1}$ and $\boldsymbol{\varphi} \in \Re^{S\times 1}$ denote the coefficient vectors of $\mathbf{P}$ and $\mathbf{V}$, respectively, λ_3 is a regularization parameter. Eq. (11) is solved via the basis pursuit de-nosing (BPDN)-homotopy algorithm [15]. Next, we compute the residual for each class $k = 1, \cdots, c$ by

$$r_k(\mathbf{x}_i) = \left\|\mathbf{x}_i - [\mathbf{P}\ \mathbf{V}]\begin{bmatrix}\delta_k(\boldsymbol{\theta}^*) \\ \boldsymbol{\varphi}^*\end{bmatrix}\right\|_2^2, \tag{12}$$

where $\delta_k(\boldsymbol{\theta}^*)$ is a vector whose nonzero entries are the entries in $\boldsymbol{\theta}^*$ that are associated with class k. Therefore, the label of the query sample $\mathbf{x}_i$ will be classified into the class with the *smallest* $r_k(\mathbf{x}_i)$, i.e., $label(\mathbf{x}_i) = \arg\min_k r_k(\mathbf{x}_i)$.

2.3. Dynamic Label Feedback

After obtaining the estimated labels for the query set, i.e., $label(\mathbf{X}_Q)$, in Stage II, we then leverage them as the feedbacks to modify the label indicator matrix $\mathbf{Y}$ of the SSLRR in Eq. (1), thus updating new prototypes $\mathbf{P}$ to facilitate the next round of P+V based label estimation. Overall, the complete algorithm of IDGL is presented in **Algorithm 2**.

It is worth noting that, in real-world scenarios, the whole query set always cannot be obtained in advance. To mimic practical face retrieval applications, we thus first collect a few antecedent query samples for batch processing, and use them to recover proper prototypes. Subsequently, IDGL can be extended to an inductive scenario for online recognition. That is, when a new query sample comes, it does not need to join the dynamic learning network but can be directly fed into the learned P + learned V model in Eq. (11)-(12) for label estimation, which is effective and efficient.

3. EXPERIMENTAL RESULTS

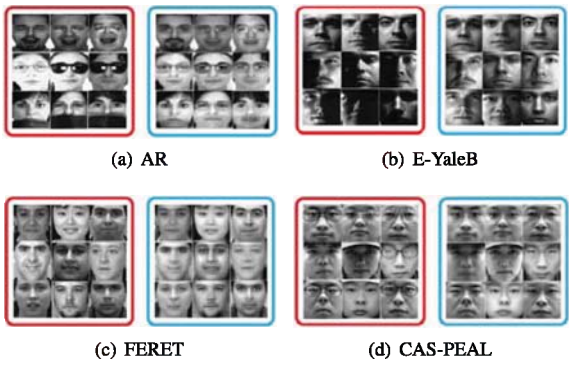
3.1. Evaluation of IDGL on SSPP-CG FR

This subsection evaluates the performance of our IDGL for SSPP-CG FR on AR [16], E-YaleB [17], FERET [18] and CAS-PEAL [19] databases in inductive setting. In this case, the unlabeled query set is divided into two equal parts, i.e., half of the query samples join the dynamic learning network to recover prototypes, while the rest half are used as new query samples for recognition. For each tested database, we randomly construct 5 gallery sets with the contamination ratios ranging from 10% to 90% with the interval of 40%. The partitions of the evaluated and generic subjects on the four databases follow the protocol in [4]. We repeat each experiment 5 times, and report the average results.

There exist few methods specifically designed for SSPP-CG FR. In the experiments, we choose 5 representative methods for comparison, including 1 popular patch-based method, i.e., DMMA [2], and 4 recent generic learning methods, i.e., SSRC [5, 9], SVDL [6], CPL [7], and the state-of-the-art S^3RC [8]. The parameters of these methods are tuned to achieve the best results. For our IDGL, λ_1, α, β in Eq. (1), λ_2 in Eq. (9), λ_3 in Eq. (11), and τ_{max} are empirically set as 15, 1, 2, 0.05, 0.001, and 10, respectively.

Fig. 3 illustrates the learned prototypes for some contaminated gallery samples on the four databases. It is clear that IDGL can successfully remove various linear variations, especially for the shadow and disguises (e.g., sunglasses and scarf), from these samples. Besides, even facing with the nonlinear variations of expressions such as laugh (see Fig. 3 (a)), IDGL has also shown good robustness and acquires appropriate prototypes that can well represent the persons.

Table 1 presents the performance of different methods. We can observe that, as the contamination ratio increases, all the methods suffer from performance decline to different extents.



(a) AR (b) E-YaleB

(c) FERET (d) CAS-PEAL

Fig. 3. Some contaminated gallery samples (left) and the learned prototypes by IDGL (right) on (a) AR, (b) E-YaleB, (c) FERET and (d) CAS-PEAL databases, respectively.

Table 1. Recognition accuracies of different methods for SSPP-CG FR on AR, E-YaleB, FERET and CAS-PEAL databases, where the contamination ratios are 10%, 50%, and 90%, respectively.

Gallery		DMMA	SSRC	SVDL	CPL	S^3RC	IDGL
AR	10%	54.9	84.9	83.1	82.9	90.9	**95.0**
	50%	37.5	74.7	70.7	70.7	82.8	**90.6**
	90%	25.3	68.7	62.3	64.4	75.1	**86.6**
E-YaleB	10%	43.2	68.5	70.0	63.0	66.7	**76.6**
	50%	33.2	62.2	61.0	56.9	53.0	**71.4**
	90%	23.3	55.7	51.3	49.1	42.3	**65.7**
FERET	10%	44.5	70.1	69.3	62.8	72.3	**79.9**
	50%	30.9	61.6	54.9	48.6	62.7	**69.5**
	90%	21.0	47.5	40.1	36.9	55.4	**63.1**
CAS-PEAL	10%	63.5	80.9	82.1	78.5	82.5	**88.5**
	50%	59.9	74.1	74.7	70.0	73.5	**79.7**
	90%	52.1	68.1	68.3	68.7	66.7	**75.7**

However, IDGL consistently outperforms the other generic learning methods including S^3RC, SVDL and SSRC, and the superiority of IDGL has shown to be more significant. For example, as the contamination ratio increases from 10% to 90%, IDGL has a gain over the second best S^3RC, from 4.1% to 11.5% on AR database. The superiority of IDGL is owing to its two advantages. The first advantage is the learned prototypes that can better represent the persons and narrow the gap between a query sample and the gallery sample of the same person but with different types of variations. The second one is the learned variation dictionary that can provide representative intra-personal variations to better reconstruct query samples. Besides, the patch-based DMMA is not competitive with SSRC, and perform much worse than our IDGL.

3.2. Computational Complexity Analysis

This subsection analyzes the computational complexity of IDGL in inductive setting. In this setting, a few query samples are collected first to train prototypes followed by the recognition of new query samples. Let $\widehat{\mathbf{X}} \in \Re^{d\times l}$ $(l = c + q)$ be the sample set matrix to be processed, q be the number of query samples for training prototypes, k be the rank of $\widehat{\mathbf{X}}$, and τ be the number of iterations in **Algorithm 1**, then the time complexity for Stage I is $O(\tau(lk^2+l^3+l^2k))$. In Stage II, the time complexity of variation dictionary learning in Eq. (9)-(10) is $O(sd^3)$ [11], and the label estimation in Eq. (11)-(12) requires $O(\tau_1 d^2 q + \tau_1 d(c + S)q)$, where τ_1 is the number of iterations for BPDN-homotopy. Let τ_{max} be the maximum number of iterations in **Algorithm 2**, then the time complexity for training prototypes is $O(\tau_{max}\tau l^3 + \tau_{max}\tau_1 dq(d + S) + sd^3)$ $(k < l, c \ll S)$. In recognition phase, the time complexity for recognizing a new query sample is $O(\tau_1 d(d + S))$.

Table 2. Recognition accuracies (%) of IDGL using the Light CNN and InsightFace features and the other deep learning-based methods on LFW database.

Methods	Accuracy (%)
DeepID	70.7
VGG-face	84.7
JCR-ACF	86.0
NN+InsightFace	94.5
NN+LightCNN-29	98.3
IDGL+InsightFace	98.1
IDGL+LightCNN-29	**99.7**

3.3. Evaluation on Deep Learning-based Features

This subsection evaluates the performance of IDGL with deep learning-based features under unconstrained environments. We first compare our IDGL using the state-of-the-art Light CNN (CNN-29 model) [20] and InsightFace [21] features, i.e., IDGL+LightCNN-29 and IDGL+InsightFace, with 3 recent deep learning-based methods including DeepID [22], VGG-face [23] and joint and collaborative representation with local adaptive convolution feature (JCR-ACF) [24], on the unconstrained LFW database [25]. For reference, we also present the results of the nearest neighbor classier using the two deep learning-based features, i.e., NN+LightCNN-29 and NN+InsightFace. We follow the experimental setting in [24] and report the recognition accuracies of all the methods. As shown in Table 2, NN+LightCNN-29 and NN+InsightFace have obtained quite high recognition accuracies of 98.3% and 94.5%, respectively, on LFW database. But even more surprising is that our IDGL still achieves the highest recognition accuracy of 99.7% using the Light CNN feature, which far outperforms the other deep learning-based methods.

Furthermore, we introduce two more challenging unconstrained databases, i.e., CelebA [26] and IJB-C [27], to evaluate the performance of IDGL+LightCNN-29 and IDGL+InsightFace. We also leverage the NN+LightCNN-29 and NN+InsightFace as two baseline methods. On CelebA, we randomly select 300 persons with 10 images per person for testing, where the first 200 persons are used for evaluation and the rest 100 ones for generic learning. On IJB-C, we select 200 videos from 200 persons for testing, where the first

Table 3. Recognition accuracies (%) of IDGL+LightCNN-29 and IDGL+InsightFace on the unconstrained CelebA and IJB-C databases. The improvements of the two methods w.r.t. their corresponding baselines are highlighted in the brackets.

Methods	CelebA	IJB-C
NN+LightCNN-29	87.9	70.9
NN+InsightFace	89.0	79.1
IDGL+LightCNN-29	93.7 (↑ **5.8**)	81.8 (↑ **10.9**)
IDGL+InsightFace	92.6 (↑ **3.6**)	86.2 (↑ **7.1**)

half are used for evaluation and the rest half for generic learning. For CelebA (or IJB-C), we randomly select a sample (or frame) of each person (or video) as the gallery sample, and select another 9 samples (or frames) for recognition. We repeat the experiment 5 times and report the average recognition results in Table 3. It is observed that IDGL+LightenCNN-29 can further enhance the recognition performance over the baseline NN+LightCNN-29 on two databases. The same situation applies to IDGL+InsightFace and NN+InsightFace. The promising results again verifies the feasibility and effectiveness of combining IDGL with deep learning-based features for practical SSPP-CG FR under unconstrained environments.

4. CONCLUSION

This paper has proposed the IDGL method to address the new challenging SSPP-CG FR problem. IDGL develops a dynamic label feedback network to update proper prototypes as well as to recognize query samples. Besides, IDGL presents a new "sample-specific" corruption strategy to learn the variation dictionary from an auxiliary generic set. Moreover, IDGL can be further enhanced by combining it with deep learning-based features under unconstrained environments. The experiments have demonstrated the superiority of IDGL, with significant improvements over the state-of-the-art counterparts.

5. REFERENCES

[1] Xiaoyang Tan, Songcan Chen, Zhi-Hua Zhou, and Fuyan Zhang, "Face recognition from a single image per person: A survey," *Pattern Recognit.*, vol. 39, no. 9, pp. 1725–1745, 2006.

[2] Jiwen Lu, Yap-Peng Tan, and Gang Wang, "Discriminative multimanifold analysis for face recognition from a single training sample per person," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 39–51, 2013.

[3] Pengyue Zhang, Xinge You, Weihua Ou, CL Philip Chen, and Yiu-ming Cheung, "Sparse discriminative multi-manifold embedding for one-sample face identification," *Pattern Recognit.*, vol. 52, pp. 249–259, 2016.

[4] Meng Pang, Yiu-ming Cheung, Binghui Wang, and Risheng Liu, "Robust heterogeneous discriminative analysis for face recognition with single sample per person," *Pattern Recognit.*, vol. 89, pp. 91–107, 2019.

[5] Weihong Deng, Jiani Hu, and Jun Guo, "In defense of sparsity based face recognition," in *CVPR*, 2013, pp. 399–406.

[6] Meng Yang, Luc Van Gool, and Lei Zhang, "Sparse variation dictionary learning for face recognition with a single training sample per person," in *ICCV*, 2013, pp. 689–696.

[7] Hong-Kun Ji, Quan-Sen Sun, Ze-Xuan Ji, Yun-Hao Yuan, and Guo-Qing Zhang, "Collaborative probabilistic labels for face recognition from single sample per person," *Pattern Recognit.*, vol. 62, pp. 125–134, 2017.

[8] Yuan Gao, Jiayi Ma, and Alan L Yuille, "Semi-supervised sparse representation based classification for face recognition with insufficient labeled samples," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2545–2560, 2017.

[9] Weihong Deng, Jiani Hu, and Jun Guo, "Face recognition via collaborative representation: its discriminant nature and superposed representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 10, pp. 2513–2521, 2018.

[10] Xiaojin Zhu, Zoubin Ghahramani, and John D Lafferty, "Semi-supervised learning using gaussian fields and harmonic functions," in *ICML*, 2003, pp. 912–919.

[11] Zhouchen Lin, Risheng Liu, and Zhixun Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *NIPs*, 2011, pp. 612–620.

[12] Jian-Feng Cai, Emmanuel J Candès, and Zuowei Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.

[13] Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, 2013.

[14] Xiaozhao Fang, Yong Xu, Xuelong Li, Zhihui Lai, and Wai Keung Wong, "Robust semi-supervised subspace clustering via non-negative low-rank representation," *IEEE Trans. Cybern.*, vol. 46, no. 8, pp. 1828–1838, 2015.

[15] David L Donoho and Yaakov Tsaig, "Fast solution of l_1-norm minimization problems when the solution may be sparse," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 4789–4812, 2008.

[16] Aleix M Martinez, "The AR face database," *CVC Tech. Rep.*, vol. 24, 1998.

[17] Athinodoros S. Georghiades, Peter N. Belhumeur, and David J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, 2001.

[18] P Jonathon Phillips, Hyeonjoon Moon, Syed A Rizvi, and Patrick J Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, 2000.

[19] Wen Gao, Bo Cao, Shiguang Shan, Xilin Chen, Delong Zhou, Xiaohua Zhang, and Debin Zhao, "The CAS-PEAL large-scale chinese face database and baseline evaluations," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 38, no. 1, pp. 149–161, 2008.

[20] Xiang Wu, Ran He, Zhenan Sun, and Tieniu Tan, "A light CNN for deep face representation with noisy labels," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 11, pp. 2884–2896, 2018.

[21] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *CVPR*, 2019, pp. 4690–4699.

[22] Yi Sun, Xiaogang Wang, and Xiaoou Tang, "Deep learning face representation from predicting 10,000 classes," in *CVPR*, 2014, pp. 1891–1898.

[23] Omkar M Parkhi, Andrea Vedaldi, Andrew Zisserman, et al., "Deep face recognition," in *BMVC*, 2015, vol. 1, p. 6.

[24] Meng Yang, Xing Wang, Guohang Zeng, and Linlin Shen, "Joint and collaborative representation with local adaptive convolution feature for face recognition with single sample per person," *Pattern Recognit.*, vol. 66, pp. 117–128, 2017.

[25] Gary B Huang and Erik Learned-Miller, "Labeled faces in the wild: Updates and new reporting procedures," *Dept. Comput. Sci., Univ. Massachusetts Amherst, Amherst, MA, USA, Tech. Rep*, pp. 14–003, 2014.

[26] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang, "Deep learning face attributes in the wild," in *ICCV*, 2015, pp. 3730–3738.

[27] Brianna Maze, Jocelyn Adams, James A Duncan, Nathan Kalka, Tim Miller, Charles Otto, Anil K Jain, W Tyler Niggel, Janet Anderson, Jordan Cheney, et al., "Iarpa janus benchmark-c: Face dataset and protocol," in *ICB*, 2018, pp. 158–165.