Iterative Dynamic Generic Learning for Face Recognition From a Contaminated Single-Sample Per Person

Meng Pang, Yiu-Ming Cheung, *Fellow, IEEE*, Qiquan Shi, *Student Member, IEEE*, and Mengke Li

***Abstract*— This article focuses on a new and practical problem in single-sample per person face recognition (SSPP FR), i.e., SSPP FR with a contaminated biometric enrolment database (SSPP-ce FR), where the SSPP-based enrolment database is contaminated by nuisance facial variations in the wild, such as poor lightings, expression change, and disguises (e.g., wearing sunglasses, hat, and scarf). In SSPP-ce FR, the most popular generic learning methods will suffer serious performance degradation because the prototype plus variation (P+V) model used in these methods is no longer suitable in such scenarios. The reasons are twofold. First, the contaminated enrolment samples could yield bad prototypes to represent the persons. Second, the generated variation dictionary is simply based on the subtraction of the average face from generic samples of the same person and cannot well depict the intrapersonal variations. To address the SSPP-ce FR problem, we propose a novel iterative dynamic generic learning (IDGL) method, where the labeled enrolment database and the unlabeled query set are fed into a dynamic label feedback network for learning. Specifically, IDGL first recovers the prototypes for the contaminated enrolment samples via a semisupervised 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 will be used as feedback 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 by virtue of the steadily enhanced prototypes. Experiments on various benchmark face data sets have demonstrated the superiority of IDGL over state-of-the-art counterparts.**

***Index Terms*— Contaminated biometric enrolment database, face recognition (FR), low-rank representation (LRR), single-sample per person (SSPP).**

Manuscript received May 27, 2019; revised November 18, 2019 and March 4, 2020; accepted March 28, 2020. Date of publication April 20, 2020; date of current version April 5, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61672444, in part by Hong Kong Baptist University (HKBU), Research Committee, Initiation Grant, Faculty Niche Research Areas (IG-FNRA) 2018/19, under Grant 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 under Project ITS/339/18, and in part by the Shenzhen Science and Technology Innovation Commission (SZSTI) under Grant JCYJ20160531194006833. *(Corresponding author: Yiu-Ming Cheung.)*

Meng Pang, Yiu-Ming Cheung, and Mengke Li are with the Department of Computer Science, Hong Kong Baptist University, Hong Kong (e-mail: mengpang@comp.hkbu.edu.hk; ymc@comp.hkbu.edu.hk; csmkli@comp.hkbu.edu.hk).

Qiquan Shi is with the Huawei Noah's Ark Lab, Hong Kong (e-mail: shi.qiquan@huawei.com).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNNLS.2020.2985099

I. Introduction

SINGLE-SAMPLE per person face recognition (SSPP FR), i.e., recognizing one person by his/her single face image in the biometric enrolment database[1] only for training (refer to Fig. 1), has a number of attractive real-world applications in criminal identification, law enforcement, access control, video surveillance, and person reidentification, to name a few [1]–[7]. However, SSPP FR is still one of the most challenging problems in FR due to the unavailability of intraclass information [8]. In such a case, a flurry of the conventional Fisher-based subspace learning methods [9]–[12], e.g., locality sensitive discriminant analysis (LSDA) [9], cannot be directly applied. Moreover, many existing sparse representation and dictionary learning methods [13]–[17], e.g., sparse representation-based classification (SRC) [13], will also suffer heavy performance decline because these methods still require multiple-sample per person (MSPP) to reasonably represent query samples.

To address the SSPP FR problem, many attempts have been made in the past decade, which can be roughly classified into two types [18]: patch-based methods and generic learning methods. Patch-based methods [19]–[22] partition each sample in the enrolment database (i.e., gallery set) into several image patches and then perform feature extraction and recognition based on these local patches. Typically, Zhu *et al*. [19] proposed a patch-based collaborative representation-based classification (PCRC) method by integrating the CRC outputs from all partitioned patches for recognition. Lu *et al*. [20] proposed a discriminative multimanifold analysis (DMMA) method to convert SSPP FR to a manifold-to-manifold matching problem provided that the patches of each person lie in an individual manifold. Moreover, Zhang *et al*. [21] extended DMMA and proposed a sparse discriminative multimanifold embedding (SDMME) method, by using sparse graph embedding to learn a discriminative subspace for partitioned patches.

For generic learning methods, they usually introduce an auxiliary generic set to provide new and useful information. Deng *et al*. [23] proposed an extended SRC (ESRC) framework by adding a generic variation set into the SSPP-based enrolment database for sparse coding. Moreover, Deng *et al*. [24] proposed a superposed SRC (SSRC)-based

[1]More standardized biometric vocabularies can refer to the website of https://www.christoph-busch.de/standards.html

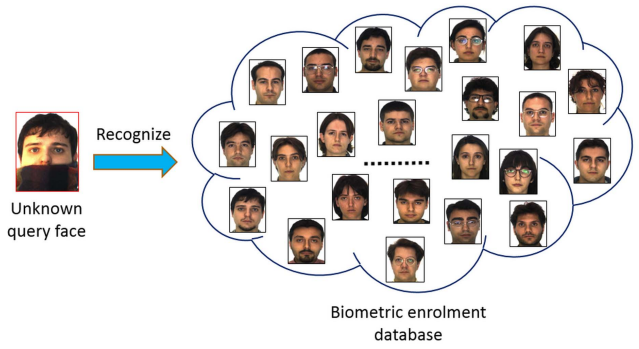


Fig. 1. Illustration of SSPP FR.

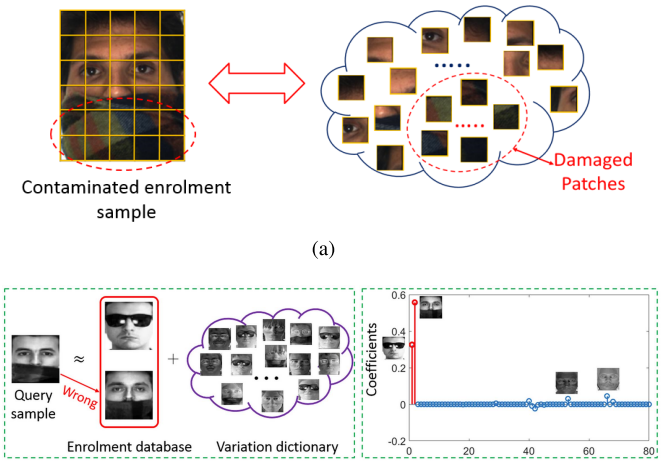


Fig. 2. (a) One contaminated enrolment sample wearing a scarf (left) and its partitioned patches (right). In these patches, some of them are damaged and contain useless information for discriminative learning and feature extraction. (b) Failed binary classification example of SSRC for SSPP-ce FR. A query sample wearing a scarf is misclassified as the wrong person wearing a similar type of scarf (left), according to the representation coefficients (right).

P+V model provided that a query face equals its prototype plus the intrapersonal variation. In the P+V model, each prototype is directly approximated by the enrolment sample provided that it is standard and variation-free. In other words, the P+V model is actually implemented as the enrolment database plus variation (E+V) model. Besides, the variation dictionary is generated by subtracting the average face from samples of each person in the auxiliary generic set. Based on the P+V/E+V model, a variety of generic learning methods [25]–[29] have been proposed recently to address the SSPP FR problem. For example, Yang *et al.* [25] proposed a sparse variation dictionary learning (SVDL) method to generate the intrapersonal variations by additionally using the relationship between the enrolment database and generic set. Ji *et al.* [26] extended SVDL by proposing a collaborative probabilistic labels (CPLs) method, to further consider the contributions of different persons in the generic set.

Although these methods mentioned earlier have achieved promising performances for SSPP FR, they still assume that each sample in the biometric enrolment database is a standard unoccluded face under uniform lighting and with a neutral expression (such as an ID photograph). Nevertheless, in some real-world scenarios, the enrolment samples are likely to be collected in less constrained environments. For example, for criminal identification, the suspects can be illegal immigrants, smugglers, or persons without residence registration. In such cases, the enrolment samples (i.e., reference photographs) of suspects are hardly acquired through standard shooting in the police but may be provided by witnesses with unaligned mobile photographs or intercepted from blurred surveillance videos. Therefore, various nuisance variations, such as expressions, lightings, poses, and disguises (e.g., wearing sunglasses or scarf), could exist in these enrolment samples, thus increasing much more difficult for practical SSPP FR. Such a new and practical issue in FR is called SSPP FR with a contaminated biometric enrolment database (SSPP-ce FR), while the previous problem of SSPP FR with a standard biometric enrolment database is called SSPP-se FR, as denoted in [30].

In general, the SSPP-ce FR problem is more challenging than SSPP-se FR, as samples in the biometric enrolment database will be contaminated and make the existing methods not amenable in this case. Particularly, for patch-based methods, discriminative learning and feature extraction from partitioned patches can be sensitive to the variations in contaminated enrolment samples [25]. Worse still, some patches may even be corrupted and capture meaningless information of persons [see Fig. 2(a)]. In contrast, generic learning methods usually perform better than patch-based methods because they introduce useful auxiliary information from the external generic set. Nevertheless, the P+V model applied in the existing generic learning methods is still not suitable to handle the new SSPP-ce FR problem. The plausible reasons are twofold.

1) The contaminated enrolment samples can no longer be treated as proper prototypes for the P+V model.
2) The generated variation dictionary cannot represent the intrapersonal variations well as it 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 because the average face is unable to characterize the neutral image of the person well, especially when the number of generic samples per person is insufficient.

Consequently, for SSPP-ce FR, a query sample will be easily misclassified as the wrong enrolment sample (i.e., person) with a similar variation in the existing generic learning methods. An illustrative classification example of SSRC for SSPP-ce FR is shown in Fig. 2(b), where a query sample wearing a scarf is misclassified as the enrolment sample wearing a similar type of scarf. As far as we know, the SSPP-ce FR problem has yet to be studied in the literature.

To address these two issues, we propose a novel iterative dynamic generic learning (IDGL) method for SSPP-ce 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 intrapersonal variants. Motivated by this, IDGL learns proper prototypes for contaminated enrolment samples by recovering their LRPs through a semisupervised

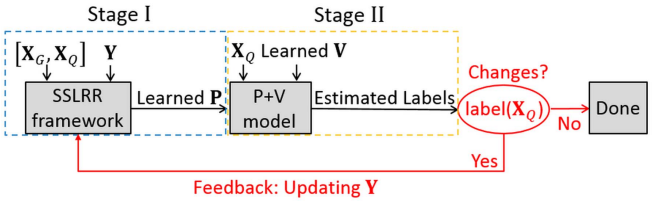


Fig. 3. Flowchart of the proposed IDGL. $\mathbf{X}_G$ and $\mathbf{X}_Q$ are the labeled gallery and unlabeled query sets, respectively. $\mathbf{P}$ and $\mathbf{V}$ denote the prototypes and variation dictionary, respectively. $\mathbf{Y}$ is the label indicator matrix.

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. 3, 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 enrolment and unlabeled query sets, we present an SSLRR framework to learn neutral prototypes to represent the persons by recovering the LRP of each contaminated enrolment sample. In Stage II, rather than simply subtracting the average face, we introduce a new "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 will be used as the feedback 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, the qualities of the learned prototypes are enhanced because the nuisance variations are gradually decreased and the useful information in the query set is effectively employed, and the accuracy of the estimated labels is improved iteratively due to the steadily enhanced prototypes. It is worth noting that after recovering proper prototypes for the contaminated enrolment database, 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.

Moreover, motivated by the success of deep learning [31]–[37] in FR and face verification, we further employ the pretrained deep neural networks to extract high-semantic features for enrolment and query samples. Subsequently, we incorporate IDGL with these deep learning-based features to verify the feasibility and generalization of IDGL for practical SSPP-ce FR under unconstrained environments.

The contributions of our work are highlighted as follows.

1) A novel IDGL by developing a dynamic learning network is proposed, to address the new and challenging SSPP-ce FR problem, where samples in the biometric enrolment database are contaminated by various nuisance facial variations.
2) A new "sample-specific" corruption strategy is presented to learn a more representative variation dictionary for the P+V model compared with the existing generic learning methods.
3) The performance of IDGL is enhanced by combining it with deep learning-based features for practical SSPP-ce FR under unconstrained environments.

The remainder of this article is organized as follows. In Section II, we review the preliminaries and related works. In Section III, we introduce the proposed IDGL method in detail. In Section IV, we conduct the experiments on nine benchmark data sets to evaluate the performance of IDGL. Finally, we draw a conclusion in Section V.

II. Preliminaries and Related Works

A. Basic Notations

In this article, matrices and vectors are represented with capital bold and lowercase bold symbols, respectively. For a matrix, e.g., $\mathbf{M}$, $\mathbf{M}(i, j)$, $\mathbf{M}(i, :)$, and $\mathbf{M}(:, j)$ denote its (i, j)th entry, ith row, and jth column, respectively. $||\mathbf{M}||_F$, $||\mathbf{M}||_2$, and $||\mathbf{M}||_*$ indicate the Frobenius norm, l_2-norm, and nuclear norm (the sum of singular values), respectively. The $l_{2,0}$- and $l_{2,1}$-norms of $\mathbf{M}$ are defined by $||\mathbf{M}||_{2,0} = \#\{j : ||\mathbf{M}(:, j)||_2 \neq 0\}$ and $||\mathbf{M}||_{2,1} = \Sigma_j ||\mathbf{M}(:, j)||_2$, respectively. $\mathbf{M}'$, $\mathbf{M}^{-1}$, and $\mathrm{Tr}(\mathbf{M})$ denote the transpose, the inverse, and the trace of $\mathbf{M}$, respectively. $\mathbf{I}$ represents the identity matrix, and $\mathbf{1}_d \in \Re^{d\times 1}$ denotes the unit column vector with d-dimension.

Let $\mathbf{X} = [\mathbf{X}_G, \mathbf{X}_Q] = [\mathbf{x}_1, \ldots, \mathbf{x}_m, \mathbf{x}_{m+1}, \ldots, \mathbf{x}_n] \in \Re^{d\times n}$ 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 enrolment database (i.e., gallery set) and unlabeled query set, respectively. The labels of labeled samples are denoted as $y_i \in \{1, 2, \ldots, c\}$, 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 introduced auxiliary generic data matrix is defined as $\mathbf{A} = [\mathbf{A}_1, \ldots, \mathbf{A}_s] = [\mathbf{a}_1, \ldots, \mathbf{a}_S] \in \Re^{d\times S}$ $(S = sT)$, with s persons not of interest and each having T different variations, and $\mathbf{A}_i = [\mathbf{a}_{(i-1)\cdot T+1}, \ldots, \mathbf{a}_{i\cdot T}]$ is the generic subsect of the ith person. The variation dictionary generated by SSRC is denoted as $\mathbf{V}_{\mathrm{ssrc}}$. The prototypes and the variation dictionary learned in our IDGL are denoted as $\mathbf{P}$ and $\mathbf{V}$, respectively.

B. Related Works

1) P+V Model: The popular P+V model [24] is developed to handle the SSPP-se FR problem, which is based on the assumption that a query sample of one person can be represented as a superposition of two different subsignals, i.e., the prototype of this person plus the intrapersonal variations. In the P+V model, the prototype is directly approximated by the enrolment sample; thus, the P+V model is actually implemented as the enrolment database plus variation (E+V) model. Formally, for a query sample $\mathbf{y}$, it can be represented as

$$\mathbf{y} = \mathbf{G}\boldsymbol{\theta} + \mathbf{V}_{\mathrm{ssrc}}\boldsymbol{\varphi} + \mathbf{e} \tag{1}$$

where $\mathbf{G}$, $\mathbf{V}_{\text{ssrc}}$, and $\mathbf{e}$ denote the enrolment samples dictionary, the variation dictionary, and a small noise, respectively, $\boldsymbol{\theta}$ is the sparse coefficient vector that selects a few of enrolment samples (i.e., persons) from $\mathbf{G}$, and $\boldsymbol{\varphi}$ is another sparse coefficient vector that selects a few types of variations from $\mathbf{V}_{\text{ssrc}}$. In this case, $\mathbf{V}_{\text{ssrc}}$ is generated by simply subtracting the average face from samples of the same person in the generic set $\mathbf{A}$ as follows:

$$\mathbf{V}_{\text{ssrc}} = \left[\mathbf{A}_1 - \mathbf{c}_1\mathbf{1}_T', \ldots, \mathbf{A}_m - \mathbf{c}_m\mathbf{1}_T'\right] \tag{2}$$

where $\mathbf{c}_i = (1/T)\mathbf{A}_i\mathbf{1}_T$ indicates the class centroid of the ith class.

Subsequently, the sparse coefficient vectors $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$ can be computed through solving the l_1-based minimization problem as follows:

$$\begin{bmatrix}\boldsymbol{\theta}^* \\ \boldsymbol{\varphi}^*\end{bmatrix} = \arg\min_{\boldsymbol{\theta},\boldsymbol{\varphi}} \left\|\mathbf{y} - [\mathbf{G} \quad \mathbf{V}_{\text{ssrc}}] \begin{bmatrix}\boldsymbol{\theta} \\ \boldsymbol{\varphi}\end{bmatrix}\right\|_2^2 + \lambda \left\|\begin{bmatrix}\boldsymbol{\theta} \\ \boldsymbol{\varphi}\end{bmatrix}\right\|_1 \tag{3}$$

where λ is a regularization parameter. Finally, similar to SRC, the query sample $\mathbf{y}$ will be classified into the enrolment sample (i.e., person) with the smallest reconstruction residual.

Based on the P+V/E+V model, various generic learning methods, such as SSRC and SVDL, have been proposed for SSPP-se FR and achieved good performances. However, when encountering the new SSPP-ce FR problem, these methods will suffer serious performance degradation due to the bad prototypes and the unrepresentative variation dictionary.

2) Prototype Learning-Based Methods: Recently, two prototype learning-based methods, namely, semisupervised sparse representation-based classification (S^3RC) [38] and synergistic generic learning (SGL) [30], have been proposed to tackle SSPP-ce FR. S^3RC estimates prototypes by the Gaussian mean for the enrolment and query sets via a Gaussian mixture model (GMM). SGL pursues prototypes by preserving the more discriminative portions of enrolment samples on account of a linear Fisher information-based feature regrouping (FIFR). However, S^3RC executes prototype learning only once based on the simple GMM, and the qualities of the learned prototypes depend heavily on its one-trail clustering accuracy. Consequently, S^3RC is unstable and sensitive to large variations (e.g., shadows) in enrolment samples. Furthermore, for SGL, the linear-based FIFR is unable to separate nonlinear variations (e.g., expressions and poses) from enrolment samples.

Other related prototype learning-based approaches include the conditional alternatives [39]–[42] based on the framework of generative adversarial nets (GANs) [43] that aim to synthesize identity-preserved neutral samples and treat them as new prototypes. However, these methods require large-scale training pairs to train the generator (or mapping) and need to know the input-out patterns beforehand, which makes them unsuitable under the SSPP-ce FR setting. Besides, although some of them perform very well for specific variations, such as poses or occlusions, they are incapable to handle multiple variations simultaneously. For example, the trained mapping in the two-pathway GAN (TP-GAN) [40] is to correct the ill-posed enrolment samples but cannot eliminate other types of variations (e.g., occlusions, shadows, and expressions) in contaminated enrolment samples. Generally speaking, the performance of SSPP-ce FR using the existing prototype learning-based methods is limited and needs to be improved.

3) Low-Rank Representation: Low-rank representation (LRR) jointly learns a lowest-rank representation that can uncover the intrinsic subspaces of data [44]. The objective function of LRR is formulated as follows:

$$\min_{\mathbf{Z},\mathbf{E}} \text{rank}(\mathbf{Z}) + \beta||\mathbf{E}||_{2,0}, \quad \text{s.t. } \mathbf{X} = \mathbf{A}\mathbf{Z} + \mathbf{E} \tag{4}$$

where $\mathbf{X}$, $\mathbf{A}$, $\mathbf{Z}$, and $\mathbf{E}$ denote the data matrix, representation dictionary, reconstruction coefficient matrix, and noise matrix, respectively, rank($\mathbf{Z}$) is the rank of $\mathbf{Z}$, and $||\mathbf{E}||_{2,0}$ models the sample-specific corruptions and outliers.

Usually, when solving the problem in (4), the representation dictionary $\mathbf{A}$ is chosen by the data matrix itself, i.e., $\mathbf{X}$, rank($\mathbf{Z}$) is approximated by $||\mathbf{Z}||_*$, and $||\mathbf{E}||_{2,0}$ is relaxed as $||\mathbf{E}||_{2,1}$. Thus, (4) can be rewritten as follows:

$$\min_{\mathbf{Z},\mathbf{E}} ||\mathbf{Z}||_* + \beta||\mathbf{E}||_{2,1}, \quad \text{s.t. } \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E} \tag{5}$$

where $\mathbf{X}\mathbf{Z}$ can be treated as the invariant LRP of the original data $\mathbf{X}$, while the rest part of $\mathbf{X}$, i.e., $\mathbf{E}$, represents the variant part depicting the sample-specific corruptions.

III. Proposed Method

IDGL is presented in two iterative learning stages, i.e., prototype learning via SSLRR and P+V-based label estimation, and a dynamic label feedback step. The flowchart of IDGL is illustrated in Fig. 3.

A. Stage I: Prototype Learning via SSLRR

In practice, it is observed that a face sample of one person is composed of: 1) an invariant LRP characterizing the neutral prototype of the person and 2) the "sample-specific" corruptions representing the intrapersonal variants. Hence, in this stage, we attempt to learn proper prototypes for contaminated samples in the biometric enrolment database, by extracting their LRPs through LRR. However, in SSPP FR, the enrolment database 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 enrolment database and present a compact SSLRR framework for prototype learning as follows:

$$\begin{aligned}\min_{\mathbf{F},\mathbf{Z},\mathbf{E}} & \sum_{i,j=1}^{n} ||\mathbf{F}_i - \mathbf{F}_j||_2^2\mathbf{Z}_{ij} + \lambda_1 \sum_{i=1}^{m} ||\mathbf{F}_i - \mathbf{Y}_i||_2^2 \\ & + \alpha||\mathbf{Z}||_* + \beta||\mathbf{E}||_{2,1} \\ \text{s.t. } & \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}\end{aligned} \tag{6}$$

where the first and second 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 semisupervised graph [45]. $||\mathbf{Z}||_*$ captures the low-rank structure of image data $\mathbf{X}$, and $\mathbf{E}_{2,1}$ encourages the columns of $\mathbf{E}$ to be zero provided that the

noise is "sample-specific". λ_1, α, and β are the balanced parameters.

Due to the high dimensionality of the original data, it is always time-consuming to compute the solution for (6). Thus, analogous to the work in [46], we present an equivalent problem with the low-rank factorization and offer a faster solution for (6) without performance loss.

Specifically, we first factorize the original data $\mathbf{X}$ via the skinny singular value decomposition (SVD) as

$$\mathbf{X} = \mathbf{W}_r \mathbf{\Sigma}_r \mathbf{H}_r' \tag{7}$$

where $\mathbf{\Sigma}_r = \text{diag}(\sigma_1, \ldots, \sigma_r)$ is a diagonal matrix with r ($r \ll d$) positive singular values in a descending order, and $\mathbf{W}_r \in \Re^{d\times r}$ and $\mathbf{H}_r \in \Re^{n\times r}$ are two columnwise orthogonal matrices with $\mathbf{W}_r'\mathbf{W}_r = \mathbf{H}_r'\mathbf{H}_r = \mathbf{I}$ satisfied. Let $\mathbf{X}_r = \mathbf{\Sigma}_r\mathbf{H}_r'$, then (6) can be converted to a factorized problem as

$$\begin{aligned}\min_{\mathbf{F},\mathbf{Z},\mathbf{E}_r} &\sum_{i,j=1}^{n} ||\mathbf{F}_i - \mathbf{F}_j||_2^2 \mathbf{Z}_{ij} + \lambda_1 \sum_{i=1}^{m} ||\mathbf{F}_i - \mathbf{Y}_i||_2^2 \\ &+ \alpha||\mathbf{Z}||_* + \beta||\mathbf{E}_r||_{2,1} \\ \text{s.t. } &\mathbf{X}_r = \mathbf{X}_r\mathbf{Z} + \mathbf{E}_r. \end{aligned} \tag{8}$$

It is easy to prove that $||\mathbf{E}||_{2,1} = ||\mathbf{X} - \mathbf{X}\mathbf{Z}||_{2,1} = ||\mathbf{W}_r(\mathbf{X}_r - \mathbf{X}_r\mathbf{Z})||_{2,1} = ||\mathbf{X}_r - \mathbf{X}_r\mathbf{Z}||_{2,1} = ||\mathbf{E}_r||_{2,1}$ according to [46], and thus, the objective functions of (6) and (8) can be equivalent. For (6), the solution $\{\mathbf{F}, \mathbf{Z}\}$ is unchanged in (8), and $\mathbf{E}$ can be computed by $\mathbf{E} = \mathbf{W}_r\mathbf{E}_r$.

The problem in (8) can be efficiently solved by a variant of alternating direction method (ADM) called linearized ADM with adaptive penalty (LADMAP) [47]. In the first step, we introduce an auxiliary variable, i.e., $\mathbf{S}$, and convert (8) to the following equivalent problem:

$$\begin{aligned}\min_{\mathbf{F},\mathbf{Z},\mathbf{E}_r,\mathbf{S}} &\sum_{i,j=1}^{n} ||\mathbf{F}_i - \mathbf{F}_j||_2^2 \mathbf{S}_{ij} + \text{Tr}\left((\mathbf{F} - \mathbf{Y})'\mathbf{U}(\mathbf{F} - \mathbf{Y})\right) \\ &+ \alpha||\mathbf{Z}||_* + \beta||\mathbf{E}_r||_{2,1} \\ \text{s.t. } &\mathbf{X}_r = \mathbf{X}_r\mathbf{Z} + \mathbf{E}_r, \mathbf{Z} = \mathbf{S} \end{aligned} \tag{9}$$

where $\mathbf{U} \in \Re^{n\times n}$ 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 (9) as follows:

$$\begin{aligned} &L(\mathbf{Z}, \mathbf{F}, \mathbf{E}_r, \mathbf{S}, \mathbf{\Lambda}_1, \mathbf{\Lambda}_2, \mu) \\ &= \sum_{i,j=1}^{n} ||\mathbf{F}_i - \mathbf{F}_j||_2^2 \mathbf{S}_{ij} + \text{Tr}\left((\mathbf{F} - \mathbf{Y})'\mathbf{U}(\mathbf{F} - \mathbf{Y})\right) \\ &\quad + \alpha||\mathbf{Z}||_* + \beta||\mathbf{E}_r||_{2,1} + Q(\mathbf{Z}, \mathbf{E}_r, \mathbf{S}, \mathbf{\Lambda}_1, \mathbf{\Lambda}_2, \mu) \\ &\quad - \frac{1}{2\mu}\left(||\mathbf{\Lambda}_1||_F^2 + ||\mathbf{\Lambda}_2||_F^2\right) \end{aligned} \tag{10}$$

where $\mu \geq 0$ is a penalty parameter, $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ are two Lagrange multipliers, and $Q(\mathbf{Z}, \mathbf{E}_r, \mathbf{S}, \mathbf{\Lambda}_1, \mathbf{\Lambda}_2, \mu) = \mu/2(||\mathbf{X}_r - \mathbf{X}_r\mathbf{Z} - \mathbf{E}_r + \mathbf{\Lambda}_1/\mu||_F^2 + ||\mathbf{Z} - \mathbf{S} + \mathbf{\Lambda}_2/\mu||_F^2)$.

The LADMAP is to update $\mathbf{Z}$, $\mathbf{F}$, $\mathbf{E}_r$, and $\mathbf{S}$ alternately by minimizing L with other variables fixed, where the quadratic term Q is linearized by its first-order approximation at the previous iteration and adding a proximal term [47]. With some algebra, the updating rules for $\mathbf{Z}$, $\mathbf{F}$, $\mathbf{E}_r$, and $\mathbf{S}$ are as follows:

$$\begin{aligned} \mathbf{Z}^{k+1} &= \arg\min_{\mathbf{Z}} \alpha||\mathbf{Z}||_* \\ &\quad + \left\langle \nabla_Z Q\left(\mathbf{Z}^k, \mathbf{E}_r^k, \mathbf{S}^k, \mathbf{\Lambda}_1^k, \mathbf{\Lambda}_2^k, \mu^k\right), \mathbf{X}_r - \mathbf{X}_r\mathbf{Z} - \mathbf{E}_r \right\rangle \\ &\quad + \frac{\eta\mu^k}{2}||\mathbf{Z} - \mathbf{Z}^k||_F^2 \\ &= \mathcal{D}_{\frac{\alpha}{\eta\mu^k}}\left(\mathbf{Z}^k - \nabla_{\mathbf{Z}} Q\left(\mathbf{Z}^k, \mathbf{E}_r^k, \mathbf{S}^k, \mathbf{\Lambda}_1^k, \mathbf{\Lambda}_2^k, \mu^k\right)/\eta\right) \end{aligned} \tag{11}$$

$$\begin{aligned} \mathbf{F}^{k+1} &= \arg\min_{\mathbf{F}} \text{Tr}\left(\mathbf{F}'\mathbf{L}^k\mathbf{F}\right) + \text{Tr}\left((\mathbf{F} - \mathbf{Y})'\mathbf{U}(\mathbf{F} - \mathbf{Y})\right) \\ &= (\mathbf{L}^k + \mathbf{U})^{-1}\mathbf{U}\mathbf{Y} \end{aligned} \tag{12}$$

$$\begin{aligned} \mathbf{E}_r^{k+1} &= \arg\min_{\mathbf{E}_r} \beta||\mathbf{E}_r||_{2,1} + \frac{\mu^k}{2}||\mathbf{X}_r - \mathbf{X}_r\mathbf{Z}^{k+1} \\ &\qquad + \mathbf{\Lambda}_1^k/\mu^k - \mathbf{E}_r||_F^2 \\ &= \Omega_{\frac{\beta}{\mu^k}}\left(\mathbf{X}_r - \mathbf{X}_r\mathbf{Z}^{k+1} + \mathbf{\Lambda}_1^k/\mu^k\right) \end{aligned} \tag{13}$$

$$\begin{aligned} \mathbf{S}^{k+1} &= \arg\min_{\mathbf{S}} \sum_{i,j=1}^{n} \left|\left|\mathbf{F}_i^{k+1} - \mathbf{F}_j^{k+1}\right|\right|_2^2 \mathbf{S}_{ij} \\ &\quad + \frac{\mu^k}{2}\left|\left|\mathbf{S} - \left(\mathbf{Z}^{k+1} + \mathbf{\Lambda}_2^k/\mu^k\right)\right|\right|_F^2 \end{aligned} \tag{14}$$

where η is a relaxation parameter that satisfies $\eta > ||\mathbf{X}_r||_F^2$, and $\nabla_{\mathbf{Z}} Q$ is the partial differential of Q with respect to $\mathbf{Z}$, i.e., $\nabla_{\mathbf{Z}} Q = -\mathbf{X}_r'(\mathbf{X}_r - \mathbf{X}_r\mathbf{Z}^k - \mathbf{E}_r^k + \mathbf{\Lambda}_1^k/\mu^k) + (\mathbf{Z}^k - \mathbf{S}^k + \mathbf{\Lambda}_2^k/\mu^k)$. $\mathbf{L} \in \Re^{n\times n}$ in (12) is the graph Laplacian matrix and computed as $\mathbf{L}^k = \mathbf{M}^k - \mathbf{S}^k$, where $\mathbf{M}_{ii}^k = \Sigma_j \mathbf{S}_{ij}^k$. $\mathcal{D}$ and Ω are the singular value thresholding [48] and $l_{2,1}$ minimization operators [44], respectively. Equation (14) is solved by decomposing it into n independent subproblems as

$$\arg\min_{\mathbf{S}_i} \mathbf{S}_i^T\mathbf{H}_i + \frac{\mu^k}{2}\left|\left|\mathbf{S}_i - \left(\mathbf{Z}_i^{k+1} + \mathbf{\Lambda}_2^k/\mu^k\right)\right|\right|_2^2 \tag{15}$$

with each having a closed-form solution as follows:

$$\mathbf{S}_i = \mathbf{Z}_i^{k+1} + \left(\mathbf{\Lambda}_2^k - \mathbf{H}_i\right)/\mu^k \tag{16}$$

where $\mathbf{H}$ is a n by n matrix whose values are defined as $\mathbf{H}_{ij} = 1/2||\mathbf{F}_i^{k+1} - \mathbf{F}_j^{k+1}||_2^2$, and $\mathbf{S}_i$ and $\mathbf{H}_i$ are the ith ($i = 1, \ldots, n$) columns of matrices $\mathbf{S}$ and $\mathbf{H}$, respectively. The algorithm for solving SSLRR by LADMAP is outlined in Algorithm 1.

Although the global convergence of LADMAP with two variables has been proven in [47], it is still difficult to prove the convergence of LADMAP with three or more variables. Since the proposed SSLRR involves four iterating variables, i.e., $\{\mathbf{F}, \mathbf{Z}, \mathbf{E}_r, \mathbf{S}\}$ and the objective in (9) is not smooth, it would be not easy to prove the convergence in theory. Fortunately, there actually exist some conditions for facilitating the convergence of SSLRR solved by LADMAP according to the theoretical results in [44] and [49]. Specifically, the three conditions (sufficient but may not necessary) for Algorithm 1 to converge are presented as follows.

1) *Condition 1:* The so-called representation dictionary, i.e., $\mathbf{X}_r$, is of full column rank.
2) *Condition 2:* The optimality gap produced in each iteration step is monotonically decreasing, namely, the residual $\epsilon_k = ||(\mathbf{F}^k, \mathbf{Z}^k, \mathbf{E}_r^k, \mathbf{S}^k) - (\mathbf{F}^*, \mathbf{Z}^*, \mathbf{E}_r^*, \mathbf{S}^*)||_F^2$ is monotonically decreasing, where $\mathbf{F}^k$, $\mathbf{Z}^k$, $\mathbf{E}_r^k$, and $\mathbf{S}^k$

Algorithm 1 IDGL Stage I: Solving SSLRR by LADMAP

Input: Factorized data matrix: $\mathbf{X}_r \in \Re^{r\times n}$, $\mathbf{Y} \in \Re^{n\times c}$, $\mathbf{U} \in \Re^{n\times n}$; Parameters: $\lambda_1, \alpha, \beta > 0$

1: Initialization: $\mathbf{Z}^0 = \mathbf{S}^0 = \mathbf{E}_r^0 = \mathbf{F}^0 = \mathbf{\Lambda}_1^0 = \mathbf{\Lambda}_2^0 = \mathbf{O}$; $\mu^0 = 0.11$, $\mu^{max} = 10^6$, $\rho^0 = 1.1$, $\epsilon_1 = \epsilon_2 = 10^{-6}$, $\eta = 1.02 * ||\mathbf{X}_r||_F^2$, $k = 0$

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

3: Fix $\mathbf{F}^k$, $\mathbf{E}_r^k$, $\mathbf{S}^k$ and update $\mathbf{Z}^{k+1}$ via Eq. (11)

4: Fix $\mathbf{Z}^{k+1}$, $\mathbf{E}_r^k$, $\mathbf{S}^k$ and update $\mathbf{F}^{k+1}$ via Eq. (12)

5: Fix $\mathbf{Z}^{k+1}$, $\mathbf{F}^{k+1}$, $\mathbf{S}^k$ and update $\mathbf{E}_r^{k+1}$ via Eq. (13)

6: Fix $\mathbf{Z}^{k+1}$, $\mathbf{E}_r^{k+1}$, $\mathbf{F}^{k+1}$ and update $\mathbf{S}^{k+1}$ via Eq. (14)

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

$$\mathbf{\Lambda}_1^{k+1} \leftarrow \mathbf{\Lambda}_1^k + \mu^k(\mathbf{X}_r - \mathbf{X}_r\mathbf{Z}^{k+1} - \mathbf{E}_r^{k+1})$$
$$\mathbf{\Lambda}_2^{k+1} \leftarrow \mathbf{\Lambda}_2^k + \mu^k(\mathbf{Z}^{k+1} - \mathbf{S}^{k+1})$$

8: Update the parameter μ as follows:

$$\mu^{k+1} = \min(\mu^{max}, \rho\mu^k), \text{ where}$$
$$\rho = \begin{cases} \rho^0 \text{ if } \mu^k \max(\sqrt{\eta}||\mathbf{Z}^{k+1}-\mathbf{Z}^k||_F, ||\mathbf{F}^{k+1}-\mathbf{F}^k||_F, \\ \quad ||\mathbf{E}_r^{k+1} - \mathbf{E}_r^k||_F, ||\mathbf{S}^{k+1} - \mathbf{S}^k||_F) \le \epsilon_2 \\ 1 \quad \text{Otherwise.} \end{cases}$$

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

10: **end while**

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

denote the solution produced at the kth iteration, respectively, and $(\mathbf{F}^*, \mathbf{Z}^*, \mathbf{E}_r^*, \mathbf{S}^*)$ indicates the "ideal" solution obtained by minimizing the Lagrangian function L with respect to $\mathbf{F}$, $\mathbf{Z}$, $\mathbf{E}_r$, and $\mathbf{S}$, simultaneously.

3) *Condition 3:* The parameter μ is nondecreasing and upper bounded.

In our algorithm, Condition 1 is easy to obey by using the orthogonal basis of $\mathbf{X}_r$ in practice. For Condition 2, although it is not easy to strictly prove the monotonically decreasing, the convexity of the Lagrangian function could guarantee its validity to some extent as discussed in [50] and [51]. Condition 3 of the upper boundedness for μ is usually used by the traditional theory of ADM to guarantee the convergence, which is also adopted in our algorithm and can be satisfied by step 8. Furthermore, we empirically study the convergence of Algorithm 1 and show its convergence process on AR data set in Fig. 4, where we can observe that Algorithm 1 converges gradually as the number of iterations increases.

After running Algorithm 1, the optimal solution of (6) can be obtained by letting $\mathbf{F} = \mathbf{F}^*$, $\mathbf{Z} = \mathbf{Z}^*$, and $\mathbf{E} = \mathbf{W}_r\mathbf{E}_r^*$. Subsequently, the recovered prototype $\mathbf{P}_i$ for the ith person in the contaminated gallery can be calculated from $\mathbf{XZ}^*$ with respect to the samples predicted as the ith person. Accordingly, the prototype learning is presented in Algorithm 2, which enables the learned prototypes to correctly represent the target persons.

B. Stage II: P+V-Based Label Estimation

1) 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, e.g., SSRC, that simply treat average face as the neutral image and subtract average face from generic samples to generate variations, our method models the neutral image by the class-specific 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 ith class, i.e., $\mathbf{A}_i \in \Re^{d\times T}$, we solve the following LRR-based optimization problem:

$$\min_{\mathbf{L}_i, \mathbf{V}_i} ||\mathbf{L}_i||_* + \lambda_2||\mathbf{V}_i||_{2,1}, \quad \text{s.t. } \mathbf{A}_i = \mathbf{A}_i\mathbf{L}_i + \mathbf{V}_i \tag{17}$$

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

$$\mathbf{V} = [\mathbf{V}_1, \ldots, \mathbf{V}_s] \in \Re^{d\times S}. \tag{18}$$

Fig. 5 illustrates the learned variation dictionary of generic samples from one person on AR data set, where we observe

Algorithm 2 IDGL Stage I: Prototype Learning

Input: $\mathbf{X} \in \Re^{d\times n}$, $\mathbf{F}^* \in \Re^{n\times c}$, and $\mathbf{Z}^* \in \Re^{n\times n}$

1: Initialization: $\mathbf{P} = \mathbf{O}$, $\mathbf{F}_{temp} \in \Re^{n\times c} = \mathbf{O}$

2: **for** $i = 1 : n$ **do**

3: $k = \arg\max_k \mathbf{F}^*(i, :)$

4: $\mathbf{F}_{temp}(i, k) = 1$

5: **end for**

6: **for** $j = 1 : c$ **do**

7: **for** $i = 1 : n$ **do**

8: **if** $\mathbf{F}_{temp}(i, j) == 1$ **then**

9: $\mathbf{P}_j = \mathbf{P}_j + \mathbf{X}(:, i) \times \mathbf{Z}^*(i, j)$

10: **end if**

11: **end for**

12: **end for**

Output: Learned prototype set: $\mathbf{P} = [\mathbf{P}_1, \ldots, \mathbf{P}_c] \in \Re^{d\times n}$

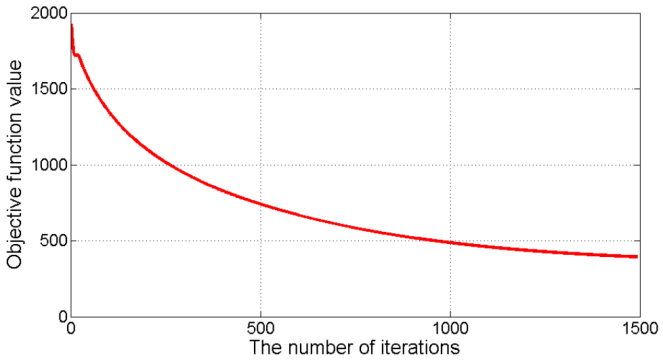


Fig. 4. Convergence process for Algorithm 1 on the AR data set.

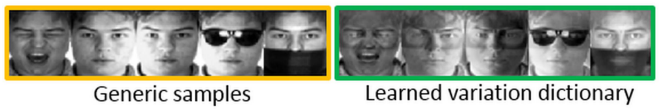


Fig. 5. Illustration of the learned variation dictionary of the generic samples from one person on the AR data set.

Algorithm 3 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 $\mathbf{P}$ via **Algorithm 1-2**
3: Stage II: Learning variation dictionary $\mathbf{V}$ in Eq. (17)-(18)
4: Stage II: Estimating $label(\mathbf{X}_Q)$ in Eq. (19)-(21)
5: Updating $\mathbf{Y}$ through $label(\mathbf{X}_Q)$
6: **until** 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., $label(\mathbf{X}_Q)$

that it has intuitive explanations and can well characterize the variations, such as expressions, lightings, and disguises (i.e., wearing sunglasses and scarf).

2) Label Estimation: 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:

$$\begin{bmatrix}\boldsymbol{\theta}^* \\ \boldsymbol{\varphi}^*\end{bmatrix} = \arg\min_{\boldsymbol{\theta},\boldsymbol{\varphi}} \left\|\mathbf{x}_i - [\mathbf{P} \quad \mathbf{V}] \begin{bmatrix}\boldsymbol{\theta} \\ \boldsymbol{\varphi}\end{bmatrix}\right\|_2^2 + \lambda_3 \left\|\begin{bmatrix}\boldsymbol{\theta} \\ \boldsymbol{\varphi}\end{bmatrix}\right\|_1 \quad (19)$$

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. In this article, (19) is solved via the basis pursuit denosing (BPDN)-homotopy algorithm [52]. Next, we compute the residual for each class $k = 1, \ldots, c$ by

$$r_k(\mathbf{x}_i) = \left\|\mathbf{x}_i - [\mathbf{P} \quad \mathbf{V}] \begin{bmatrix}\delta_k(\boldsymbol{\theta}^*) \\ \boldsymbol{\varphi}^*\end{bmatrix}\right\|_2^2 \quad (20)$$

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 $r_k(\mathbf{x}_i)$ as follows:

$$\text{label}(\mathbf{x}_i) = \arg\min_k r_k(\mathbf{x}_i). \quad (21)$$

C. 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 (6), thus updating new prototypes $\mathbf{P}$ to facilitate the next round of P+V-based label estimation. This dynamic learning process will be terminated when the maximum number of iterations τ_{max} is reached or $label(\mathbf{X}_Q)$ is not changed between two successive iterations. In summary, the complete algorithm of IDGL is presented in Algorithm 3. It is worth mentioning that, line 3 in Algorithm 3 only needs to be executed once in the first iteration because the learned variation dictionary can be shared in later iterations.

With the dynamic learning network in IDGL, the estimated labels for query samples will be more and more accurate due to the consistently improved prototypes. For better understanding, we select a contaminated enrolment sample wearing sunglasses on the AR data set and illustrate its learned prototype for each iteration in Fig. 6. It can be observed that IDGL converges fast after just five iterations. Besides, as the number of iterations increases, the disguise of sunglasses in the enrolment sample can be removed gradually, thus making the learned prototype better represent the target person.

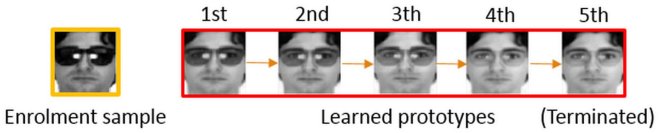


Fig. 6. Illustration of a contaminated enrolment sample wearing sunglasses on the AR data set and its learned prototype for each iteration in IDGL.

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 (19)–(21) for label estimation, which is effective and efficient.

IV. Experimental Results

This section includes four experimental parts to evaluate the performance of the proposed IDGL method as follows.

1) In Section IV-B, we verify the effectiveness of the learned variation dictionary in the IDGL model and compare its performance with that of the state-of-the-art variation dictionary learning methods on AR, E-YaleB, Multi-PIE, CAS-PEAL, and FERET data sets.
2) In Section IV-C, we evaluate the performance of IDGL for SSPP-ce FR on the abovementioned five benchmark data sets, in both transductive and inductive settings.
3) In Section IV-D, we investigate the contributions of the learned prototypes and learned variation dictionary in the IDGL model, respectively, when addressing the SSPP-ce FR problem.
4) In Section IV-E, we analyze the computational complexity of IDGL.
5) In Section IV-F, motivated by the success of deep learning in FR [53], we further evaluate the performance of IDGL by combining it with the deep learning-based features on the more challenging Face Recognition Grand Challenge version 2.0 (FRGC v2.0), the unconstrained Labeled Faces in the Wild (LFW), CelebFaces Attributes (CelebA), and IJB-C data sets.

The abovementioned experiments are conducted on a host (CPU: Dual 6-core Intel Xeon X5650 2.66-GHz 12-MB L3 Cache; Memory: 32 GB).

A. Data Set Description

The AR data set [54] consists of over 4000 face images from 126 persons across two sessions (i.e., S-I and S-II), and

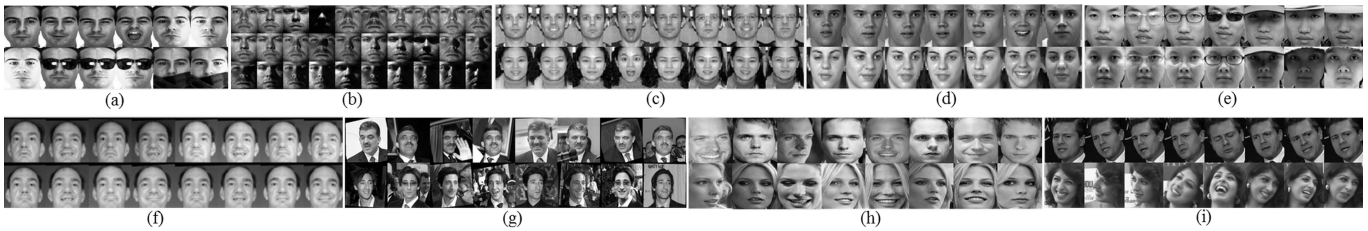


Fig. 7. Illustration of some face samples on nine face data sets. (a) AR. (b) E-YaleB. (c) Multi-PIE. (d) FERET. (e) CAS-PEAL. (f) FRGC v2.0. (g) LFW. (h) CelebA. (i) IJB-C.

each session has 13 images per person, including different facial variations of expressions, illuminations, and disguises (i.e., sunglasses and scarf).

The E-YaleB data set [55] consists of 2414 images from 38 persons under varieties of lighting conditions, which are divided into five subsets. Subset 1 is under normal lighting condition (lighting angles: 0°–12°), Subsets 2 and 3 depict slight-to-moderate luminance variations (angles: 13°–25° and 26°–50°), and Subsets 4 and 5 characterize severe lighting variations (angles: 51°–77° and $>77°$).

The Multi-PIE data set [56] consists of 337 persons with each containing face images with six different expressions across four sessions (Session 1–4), 15 poses, and 20 illuminations.

The CAS-PEAL data set [57] contains 99 594 images of 1040 persons (595 males and 445 females) with variations including expression, facing direction, accessory, lighting, and age. CAS-PEAL is believed to be the largest public data set with occluded face images available.

The FERET data set [58] is sponsored by the U.S. Department of Defense through the DARPA Program and consists of 14 126 images from 1199 persons who are diverse across ethnicity, gender, and age.

The FRGC v2.0 data set [59] consists of 50 000 images of 4003 persons with two different facial expressions, taken under different illumination conditions.

The LFW data set [60] contains over 13 000 images of 5749 persons collected in uncontrolled environments with large variations in expressions, poses, illuminations, and so on.

The CelebA data set [61] consists of more than 200 000 celebrity images, each annotated with 40 attributes. The images in this data set cover large pose variations and background clutter and are center cropped to 128×128 pixels.

The IJB-C data set [62] contains 3531 subjects with a total of 31 334 still images and 117 542 frames from 11 779 videos. All the images and videos are collected in unconstrained environments, which show large variations in expression, pose, image qualities, and so on.

Fig. 7 show some gray face samples on AR, E-YaleB, Multi-PIE, CAS-PEAL, FERET, FRGC v2.0, LFW, CelebA, and IJB-C face data sets.

B. Evaluation of the Learned Variation Dictionary

This section validates the effectiveness of the learned variation dictionary in the IDGL model. Specifically, we follow the setting of SSPP-se FR and choose unoccluded neutral faces to build the standard biometric enrolment database.

TABLE I
DATA SET CONFIGURATION

Dataset	Dimension	Evaluated persons	Generic persons
AR	2304 (48×48)	50	50
E-YaleB	2304 (48×48)	20	18
Multi-PIE	2304 (48×48)	80	40
FERET	2304 (48×48)	100	100
CAS-PEAL	2304 (48×48)	100	100

In the experiment, our learned variation dictionary in IDGL is denoted as Dict-IDGL, and the prototype learning via the SSLRR stage has not been used in the case for a fair comparison. Subsequently, we evaluate the performance of Dict-IDGL under various query variations on AR, E-YaleB, Multi-PIE, FERET, and CAS-PEAL data sets.

On the AR data set, we randomly select 50 persons from S-I for evaluation and choose another 50 persons as the generic set for generic learning methods. The neutral images of all evaluated persons are used to build the biometric enrolment database, and the rest 12 images are formed as five query sets (expression, illumination, illumination+sunglasses, illumination+scarf, and disguises). Moreover, we also leverage the face images from S-II for testing. On the E-YaleB data set, the first 20 persons are used for evaluation, and the rest 18 persons are used for generic learning. The first image of each person in Subset 1 (lighting angle: 0°–12°) is chosen as enrolment sample, and Subsets 2–5 form four query sets with different lighting angles. On the Multi-PIE data set, we select 120 persons in expression subset across four different sessions, where the first 80 persons are used for evaluation, while the rest 40 persons are used for generic learning. The neutral image of each person in Session 1 is used as an enrolment sample, while the rest nine images in Session 2–4 form three query sets. On the FERET data set, we select 200 persons from seven subsets (ba, bj, bk, bd, bf, and bg), including variations of poses, illuminations, and expressions. We use the first 100 persons for evaluation, while the rest 100 persons are chosen as the generic set. The neutral images of the evaluated persons are used as enrolment samples, and the rest six images are formed as three query sets. On the CAS-PEAL data set, we use 200 persons from the Normal and the Accessory categories; thus, each person has one neutral image and six images wearing different glasses and hats. The first 100 persons and the rest 100 ones are used for evaluation and generic learning, respectively. The neutral images of the evaluated persons are used as enrolment samples, and the rest six images are formed as two query sets (glasses and hats). The configurations of the five tested data sets are listed in Table I.

We select three representative methods for comparison, including the baseline SRC [13], and two state-of-the-art

TABLE II
RECOGNITION RATES (%) OF DIFFERENT METHODS ON THE AR, E-YALEB, MULTIPIE, FERET, AND CAS-PEAL BENCHMARK DATA SETS FOR SSPP-SE FR

Query set		SRC	SSRC	SVDL	Dict-IDGL
AR	S-I: Expression	84.0	87.7	89.3	**96.7**
	S-I: Illumination	74.0	94.7	95.0	**98.0**
	S-I: Ill.+Sunglasses	41.3	86.7	86.7	**92.7**
	S-I: Ill.+Scarf	26.7	77.0	78.7	**86.7**
	S-I: Disguise	36.0	81.0	87.0	**90.0**
	S-II: Expression	62.0	73.7	74.0	**80.0**
	S-II: Illumination	42.7	81.0	84.7	**87.3**
	S-II: Ill.+Sunglasses	21.3	64.3	68.7	**72.7**
	S-II: Ill.+Scarf	16.0	60.3	65.3	**70.7**
	S-II: Disguise	26.0	68.0	70.0	**75.0**
	Average	43.0	77.4	79.9	**85.0**
E-YaleB	Subset 2 (13°-25°)	96.7	**100.0**	**100.0**	**100.0**
	Subset 3 (26°-50°)	56.7	97.3	98.6	**99.2**
	Subset 4 (51°-77°)	15.0	69.3	71.1	**79.6**
	Subset 5 (>77°)	7.6	18.9	20.8	**31.0**
	Average	44.0	71.4	72.6	**77.5**
Multi-PIE	Session 2	55.4	75.0	73.8	**75.8**
	Session 3	53.8	72.1	65.8	**77.5**
	Session 4	56.7	68.3	66.7	**72.1**
	Average	55.3	71.8	68.8	**75.1**
FERET	Expression	75.0	82.0	85.0	**89.0**
	Illumination	72.0	79.0	82.0	**87.0**
	Poses	38.5	60.7	56.0	**61.3**
	Average	61.8	73.9	74.3	**79.1**
CAS-PEAL	Accessory: Glasses	90.3	94.3	95.7	**97.0**
	Accessory: Hats	54.0	64.3	68.9	**76.3**
	Average	72.2	79.3	82.3	**86.7**

variation dictionary learning methods, i.e., SSRC [24], [29] and SVDL [25]. For SRC and SSRC, the values of the regularization parameter λ are searched from $\{0.001, 0.005, 0.01, 0.05, 0.1\}$ to achieve the best results over five tested data sets. For SVDL, the parameters λ_1, λ_2, and λ_3 are set to be 0.001, 0.01, and 0.0001, respectively, according to the suggestion in [25]. As to our Dict-IDGL, the only related parameter is λ_2 in (17) and is empirically set as 0.05.

Table II lists the accuracies of all the methods over five tested data sets. It can be seen that Dict-IDGL consistently outperforms the other variation dictionary learning methods, including SVDL and SSRC, in almost all cases. For example, it delivers 7.6%, 6.1%, 3.3%, 5.2%, and 7.4% improvements, on average, over SSRC on AR, E-YaleB, Multi-PIE, FERET, and CAS-PEAL, respectively. The superior performances of Dict-IDGL on the abovementioned five tested data sets demonstrate the effectiveness and rationality of the proposed new way of learning variation dictionary in IDGL. Besides, SVDL is based on the generated variation dictionary in SSRC but additionally uses the relationships between the enrolment and generic samples; thus, SVDL achieves better performance than SSRC in most cases. In addition, the baseline SRC is not competitive with the variation dictionary learning methods, such as SSRC and SVDL, and performs much worse than our Dict-IDGL, which explains the importance of introducing auxiliary variation dictionary for SSPP FR.

C. Evaluation of IDGL on SSPP-Ce FR

This section evaluates the performance of our IDGL for SSPP-ce FR on AR, E-YaleB, Multi-PIE, FERET, and CAS-PEAL benchmark data sets. In this section, the prototype learning via the SSLRR stage is involved in the IDGL model since the enrolment samples can be contaminated by nuisance variations in this scenario. All the experiments are performed in both transductive and inductive settings as follows.

1) *Transductive:* In this setting, the data are partitioned into two parts, i.e., the labeled enrolment sample and the unlabeled query samples. For each tested data set, we randomly select a contaminated sample of each person as the enrolment sample and use all the rest samples for recognition. For example, on AR data set, we select the contaminated sample from the expression, illumination, illumination+sunglasses, illumination+scarf, and disguises subsets and use the rest 12 samples for recognition.
2) *Inductive:* In this setting, the unlabeled query samples are further divided into two equal parts, i.e., half of the query samples join the dynamic learning network in IDGL, while the rest half are used as new query samples for recognition. In this case, we only focus on the performance of the new query samples.

In both transductive and inductive settings, we randomly construct five biometric enrolment databases with the contamination ratios of 10%, 30%, 50%, 70%, and 90%, respectively, for each data set. The partitions of the evaluated and generic persons on these data sets follow the configurations in Table I. We repeat the experiment five times and report the average results.

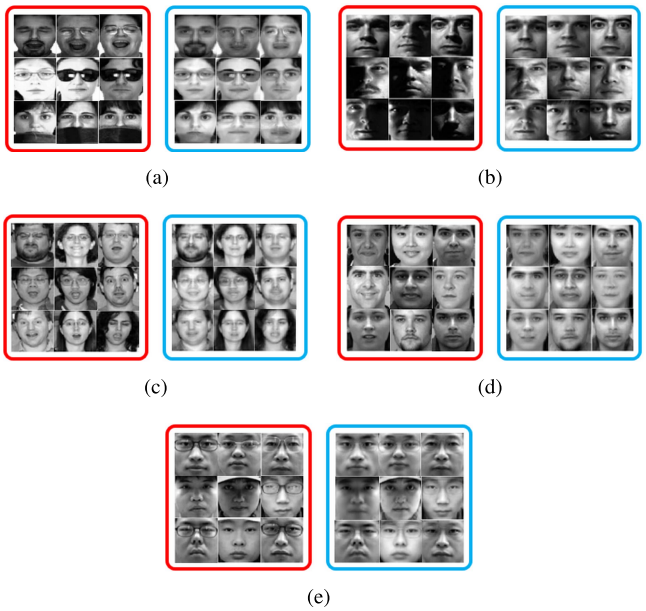
We choose nine representative methods for comparison, including the baseline SRC, three popular patch-based methods, i.e., PCRC [19], DMMA [20], and SDMME [21], four recent generic learning methods, i.e., ESRC [23], SSRC, SVDL, and CPL [26], and the state-of-the-art prototype learning-based S^3RC [38]. Among the nine comparing methods, we implement DMMA, SDMME, and CPL by ourselves and obtain the source codes of the other six methods from the original authors. For the parameter settings, the nonoverlapped patch sizes for DMMA, SDMME, and PCRC are set as 8×8 pixels. Besides, k_1, k_2, k, and σ in DMMA are empirically tuned to be 30, 2, 2, and 100, respectively, and the balance factor λ in SDMME is tuned to be 0.001. According to the suggestions in [26] and [38], λ, δ_1, and δ_2 in CPL are set as 0.01, 0.3, and 3, respectively, and λ in S^3RC is set as 0.001. The parameters of SRC, ESRC, SSRC, and SVDL are kept the same as that in Section IV-B. For our IDGL, λ_1, α, and β in (8), λ_2 in (17), λ_3 in (19), and t_{max} are set as 15, 1, 2, 0.05, 0.001, and 10, respectively, over five tested data sets. Moreover, for each iteration in IDGL, we empirically give a penalty of β by $\beta = (\beta/1.2)$ to weaken the effect of the residual term and, thus, strengthen the effect of the low-rank term $||\mathbf{Z}||_*$ for better prototype recovering. For a fair comparison, the values of the parameters in these methods are retained unchanged in both transductive and inductive settings.

In Fig. 8, we illustrate the learned prototypes for some contaminated enrolment samples in transductive setting on AR, E-YaleB, Multi-PIE, FERET, and CAS-PEAL benchmark data sets. It is clear that our IDGL can successfully remove various linear facial variations, especially for the shadow and disguises (e.g., sunglasses and scarf), from the contaminated enrolment samples. Besides, even facing the more challenging

TABLE III

AVERAGE RECOGNITION RATES (%) AND STANDARD ERRORS (%) OF DIFFERENT METHODS ON AR, E-YALEB, MULTI-PIE, FERET, AND CAS-PEAL BENCHMARK DATA SETS FOR SSPP-CE FR IN TRANSDUCTIVE SETTING. IN THE BRACKETS, WE SHOW THE IMPROVEMENT OF OUR IDGL WITH RESPECT TO THE SECOND BEST METHOD IN THE CASE

Enrolment database		Baseline	Patch-based methods			Generic learning methods					Our method
		SRC	DMMA	SDMME	PCRC	ESRC	SSRC	SVDL	CPL	S^3RC	IDGL
AR	10%	50.5±3.2	56.4±1.2	55.9±0.8	66.3±4.6	81.6±5.3	84.4±1.0	83.0±1.7	84.5±6.8	90.9±1.2	**96.3±1.0** (↑ 5.4)
	30%	46.6±1.0	46.7±1.3	47.0±1.9	59.6±4.3	75.9±1.1	80.0±1.7	73.8±0.9	73.9±0.9	89.0±0.6	**93.3±1.1** (↑ 4.3)
	50%	43.0±1.0	38.2±1.5	38.7±1.1	57.3±1.6	68.5±1.3	73.8±2.3	64.4±2.6	66.5±1.7	85.4±3.1	**92.9±1.8** (↑ 7.5)
	70%	41.1±2.7	30.2±1.7	31.0±1.5	57.4±3.0	63.1±1.9	67.1±2.4	56.3±2.5	59.7±3.1	79.4±1.9	**88.3±2.4** (↑ 8.9)
	90%	38.6±1.9	27.3±1.4	28.2±1.6	54.3±2.6	59.0±1.8	64.9±1.1	52.8±1.5	57.8±2.3	78.8±3.7	**87.8±1.7** (↑ 9.0)
	Average	44.0	39.8	40.2	59.0	69.6	74.0	66.1	68.5	84.7	**91.7** (↑ 7.0)
E-YaleB	10%	44.1±1.3	43.0±1.0	42.0±2.0	64.5±3.1	63.3±1.1	66.1±1.4	67.4±1.3	64.6±1.8	59.7±1.9	**77.2±1.5** (↑ 9.8)
	30%	38.9±1.9	36.0±1.3	37.4±1.4	53.9±3.2	58.3±1.8	61.7±2.2	57.5±2.3	56.9±2.1	52.1±2.9	**74.0±0.9** (↑ 12.3)
	50%	37.1±2.4	32.7±3.4	31.7±2.4	54.4±1.1	56.2±5.4	58.7±4.4	54.8±5.0	53.0±3.6	45.1±3.2	**72.3±5.1** (↑ 13.6)
	70%	33.5±2.7	26.9±1.4	26.2±1.1	50.8±1.7	50.9±1.9	54.5±1.0	47.7±2.4	47.2±2.6	42.2±3.5	**68.1±2.8** (↑ 13.6)
	90%	31.6±1.6	24.0±1.6	23.5±1.6	51.2±1.6	49.2±1.0	52.6±0.5	44.8±1.9	44.9±1.9	39.7±4.2	**67.2±1.1** (↑ 14.6)
	Average	37.0	32.5	32.2	55.0	55.6	58.7	54.4	53.3	47.8	**71.8** (↑ 13.1)
Multi-PIE	10%	56.0±0.9	62.8±1.2	59.5±0.9	60.4±1.5	59.6±1.9	66.1±1.5	67.1±1.2	65.2±1.5	71.5±1.5	**75.9±1.6** (↑ 4.4)
	30%	54.2±1.2	63.6±0.7	60.4±1.0	60.5±1.1	59.3±2.0	64.8±1.0	66.5±1.7	64.5±1.2	69.7±1.2	**76.3±1.1** (↑ 6.6)
	50%	52.8±2.6	60.8±1.7	57.5±2.6	58.0±2.5	58.3±1.2	63.8±1.2	64.8±2.2	61.8±2.0	67.8±2.2	**74.4±1.9** (↑ 6.6)
	70%	53.4±3.1	58.9±2.3	53.4±2.4	57.4±3.4	58.3±0.7	63.1±2.3	64.2±2.4	62.5±1.7	67.7±3.8	**74.5±2.8** (↑ 6.8)
	90%	52.6±1.8	58.2±2.6	51.6±2.1	54.3±1.8	55.8±2.3	61.6±1.1	63.8±1.8	59.9±0.5	66.4±1.6	**74.0±2.4** (↑ 7.6)
	Average	53.8	60.9	56.5	58.1	58.3	63.9	65.3	62.8	68.6	**75.0** (↑ 6.4)
FERET	10%	40.6±2.0	44.1±2.3	45.8±2.3	28.4±1.0	51.3±2.3	62.7±2.8	50.9±2.7	49.3±1.9	72.2±3.4	**75.8±1.0** (↑ 3.6)
	30%	36.1±1.1	38.3±1.5	39.5±1.9	24.3±0.6	45.0±1.1	60.0±1.9	50.3±2.2	44.6±2.3	68.6±2.6	**73.2±2.8** (↑ 4.6)
	50%	34.9±0.7	32.7±1.8	33.9±2.0	24.7±0.7	40.2±1.9	59.5±2.6	47.2±1.0	43.3±0.9	62.7±1.7	**68.7±0.8** (↑ 6.0)
	70%	34.2±0.8	29.2±0.4	30.8±0.3	22.5±1.1	37.8±1.6	56.3±1.4	45.6±2.3	41.7±1.6	59.2±1.5	**67.7±2.1** (↑ 8.5)
	90%	30.3±1.4	25.1±0.8	26.1±0.7	20.8±0.9	34.2±1.2	53.2±0.8	40.4±0.9	36.8±2.0	55.4±2.5	**64.6±1.2** (↑ 9.2)
	Average	35.2	33.9	35.2	24.1	41.7	58.3	46.9	43.1	63.6	**70.0** (↑ 6.4)
CAS-PEAL	10%	68.6±1.3	55.1±1.2	56.1±0.7	66.3±2.5	72.5±1.7	73.2±1.5	72.6±2.6	73.5±2.3	79.2±2.3	**83.8±1.4** (↑ 4.6)
	30%	65.8±1.1	52.3±1.0	52.0±1.1	63.0±1.3	69.1±1.1	70.3±1.5	68.4±1.7	69.4±2.2	74.6±0.4	**79.8±1.2** (↑ 5.2)
	50%	64.8±1.1	50.8±1.0	50.3±1.7	59.7±3.5	66.9±1.0	67.4±1.6	66.5±1.9	68.2±1.5	70.6±2.6	**75.6±2.4** (↑ 5.0)
	70%	62.6±1.3	46.8±2.3	46.2±1.9	61.0±0.9	62.6±3.3	64.2±2.1	62.6±2.4	63.9±0.8	69.2±0.9	**74.7±2.5** (↑ 5.5)
	90%	60.0±1.8	44.4±1.9	42.7±2.0	60.5±0.8	60.5±1.5	61.9±1.9	60.3±2.5	61.4±1.9	65.2±1.7	**71.2±2.9** (↑ 6.0)
	Average	64.4	49.9	49.5	62.1	66.3	67.4	66.1	67.3	71.8	**77.0** (↑ 5.2)



(a) (b) (c) (d) (e)

Fig. 8. Some contaminated samples in the biometric enrolment database (left) and the corresponding learned prototypes by IDGL (right) on (a) AR, (b) E-YaleB, (c) Multi-PIE, (d) FERET, and (e) CAS-PEAL data sets, respectively.

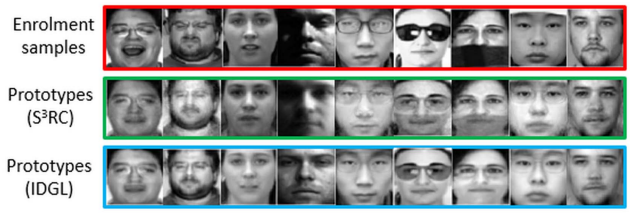


Fig. 9. Comparisons between the learned prototypes by S^3RC and the ones by our IDGL for the contaminated enrolment samples with nine different types of variations, i.e., expressions of laugh and disgust, slight illumination, shadow, disguises of ordinary glasses, sunglasses, scarf and hat, and pose.

nonlinear variations, such as exaggerated expressions and poses, IDGL has also shown good robustness and acquires appropriate prototypes that can well represent the persons. Furthermore, considering that S^3RC also contains prototype learning stage, we, thus, compare the learned prototypes by our IDGL and S^3RC, with nine different types of variations, i.e., expressions of laugh and disgust, slight illumination, shadow, disguises of ordinary glasses, sunglasses, scarf and hat, and pose. It can be observed from Fig. 9 that the learned prototypes by S^3RC always contain noises and even cannot represent the correct person in a few cases (e.g., under slight illumination and shadow). In contrast, the learned prototypes by our IDGL look very smooth and all of them can correctly characterize the neutral image of the target persons. This is because, S^3RC executes prototype learning only once based on the simple GMM, and the qualities of the learned prototypes depend heavily on its clustering accuracy, while our IDGL dynamically updates the prototypes with the label-feedback SSLRR framework, which will lead the learned prototypes to be more reasonable and reliable.

Tables III and IV present the performances of all the involved methods in the transductive setting and inductive setting, respectively. From Tables III and IV, we have the following key observations.

TABLE IV

AVERAGE RECOGNITION RATES (%) AND STANDARD ERRORS (%) OF DIFFERENT METHODS ON THE AR, E-YALEB, MULTI-PIE, FERET, AND CAS-PEAL BENCHMARK DATA SETS FOR SSPP-CE FR IN INDUCTIVE SETTING. IN THE BRACKETS, WE SHOW THE IMPROVEMENT OF OUR IDGL WITH RESPECT TO THE SECOND BEST METHOD IN THE CASE

Enrolment database		Baseline	Patch-based methods			Generic learning methods					Our method
		SRC	DMMA	SDMME	PCRC	ESRC	SSRC	SVDL	CPL	S^3RC	IDGL
AR	10%	43.7±3.1	54.9±2.1	52.5±1.4	57.9±7.2	80.5±2.2	84.9±1.7	83.1±1.9	82.9±1.7	90.9±1.4	**95.0±1.2** (↑ 4.1)
	30%	38.7±1.7	43.1±2.8	41.0±3.1	50.8±4.9	74.0±1.4	79.4±2.0	82.1±3.0	76.6±1.9	86.5±1.5	**91.6±0.6** (↑ 5.1)
	50%	39.4±0.8	37.5±1.7	34.9±0.6	54.7±1.9	66.9±3.9	74.7±3.7	70.7±3.0	70.7±2.1	82.8±2.5	**90.6±2.2** (↑ 7.8)
	70%	37.9±1.7	29.7±2.3	28.1±3.1	52.1±4.2	58.8±3.5	70.5±2.2	64.7±4.7	65.6±2.3	75.8±1.6	**86.7±2.0** (↑ 10.9)
	90%	37.7±2.4	25.3±1.4	24.6±1.4	51.3±1.1	58.3±3.3	68.7±1.4	62.3±1.0	64.4±2.6	75.1±2.0	**86.6±1.2** (↑ 11.5)
	Average	39.5	38.1	36.2	53.4	67.7	75.6	72.6	72.0	82.2	**90.1** (↑ 7.9)
E-YaleB	10%	45.8±1.4	43.2±0.6	41.5±0.6	67.3±3.1	65.4±0.8	68.5±1.2	70.0±0.6	63.0±1.0	66.7±0.4	**76.6±1.2** (↑ 6.6)
	30%	45.3±2.0	37.3±1.6	35.7±1.0	56.7±3.0	61.0±1.6	64.6±2.6	63.6±2.4	58.5±1.8	58.7±3.6	**72.1±3.0** (↑ 7.5)
	50%	41.9±2.6	33.2±2.4	31.3±1.9	58.4±2.8	58.7±4.6	62.2±4.9	61.0±5.4	56.9±3.3	53.0±3.7	**71.4±4.0** (↑ 9.2)
	70%	36.9±2.8	27.3±1.5	25.9±1.5	52.0±1.6	54.0±2.3	57.8±2.8	54.5±3.0	50.3±2.2	49.3±2.4	**67.0±2.9** (↑ 9.2)
	90%	33.7±2.4	23.3±1.4	22.4±1.6	51.4±1.5	52.5±1.0	55.7±1.2	51.3±1.3	49.1±0.5	42.3±6.8	**65.7±1.2** (↑ 10.0)
	Average	40.7	32.9	31.4	57.2	58.3	61.8	60.1	55.6	54.0	**70.6** (↑ 8.8)
Multi-PIE	10%	64.3±0.8	71.0±1.2	68.7±1.1	63.8±1.2	68.4±1.6	72.5±0.7	75.7±2.1	71.8±1.6	71.7±1.4	**79.8±1.8** (↑ 4.1)
	30%	62.5±1.7	69.7±2.2	67.4±1.0	63.0±1.3	65.9±1.5	70.9±2.4	72.0±1.3	70.8±1.4	70.3±4.1	**79.6±1.6** (↑ 7.6)
	50%	59.6±2.1	65.2±2.5	63.3±2.8	58.3±1.7	62.4±2.0	68.0±1.4	69.8±2.0	66.0±3.0	70.1±1.8	**78.9±2.1** (↑ 8.8)
	70%	58.8±2.9	64.1±2.9	57.2±4.3	57.4±4.1	62.5±3.6	65.9±2.6	68.3±3.7	66.8±2.8	65.5±1.7	**78.1±1.6** (↑ 9.8)
	90%	57.3±2.5	63.3±2.3	54.1±1.6	55.9±2.9	57.4±1.4	65.3±1.7	65.6±2.9	64.0±2.3	65.6±4.2	**76.4±1.7** (↑ 10.8)
	Average	60.5	66.7	62.1	59.7	63.3	68.5	70.3	67.9	68.6	**78.6** (↑ 8.3)
FERET	10%	50.6±3.1	44.5±3.7	44.3±2.5	43.7±4.6	62.9±2.9	70.1±3.8	69.3±1.6	62.8±2.0	72.3±2.6	**79.9±2.2** (↑ 7.6)
	30%	47.2±2.0	37.0±1.7	37.5±1.4	35.1±0.8	54.8±1.6	64.2±2.1	61.6±1.8	53.8±2.1	67.6±1.5	**73.9±2.1** (↑ 6.3)
	50%	40.3±1.5	30.9±1.5	31.9±1.6	32.6±2.2	47.7±3.4	61.6±2.2	54.9±1.8	48.6±1.5	62.7±3.5	**69.5±1.7** (↑ 6.8)
	70%	39.7±1.4	28.4±1.5	28.9±1.1	30.3±2.3	42.9±1.2	57.5±1.4	50.8±1.6	46.7±0.3	60.9±2.5	**68.2±3.5** (↑ 7.3)
	90%	30.4±1.6	21.0±1.1	21.8±0.9	23.1±1.1	32.9±1.2	47.5±1.4	40.1±1.8	36.9±1.3	55.4±2.0	**63.1±1.5** (↑ 7.7)
	Average	41.6	32.4	32.9	33.0	48.2	60.2	55.3	49.8	63.8	**70.9** (↑ 7.1)
CAS-PEAL	10%	75.8±0.8	63.5±1.5	62.3±1.5	67.0±2.6	78.5±2.7	80.9±1.8	82.1±1.2	78.5±0.9	82.5±1.7	**88.5±0.4** (↑ 6.0)
	30%	72.5±0.8	61.3±1.0	59.5±1.3	60.4±3.6	74.9±0.9	76.6±1.3	78.5±0.9	76.1±3.2	77.2±1.6	**82.9±1.4** (↑ 4.4)
	50%	70.7±1.7	59.9±2.1	58.4±1.9	63.6±3.9	72.8±1.9	74.1±2.8	74.7±3.4	70.0±6.3	73.5±2.5	**79.7±3.8** (↑ 5.0)
	70%	67.3±2.2	54.7±2.5	52.8±1.9	63.1±3.0	67.6±3.7	71.5±1.8	71.0±1.5	68.9±3.1	68.3±1.7	**78.0±2.0** (↑ 6.5)
	90%	66.3±1.8	52.1±2.4	49.5±2.8	62.7±3.5	65.5±1.5	68.1±2.5	68.3±1.8	68.7±2.7	66.7±1.8	**75.7±1.8** (↑ 7.0)
	Average	70.5	58.3	56.5	63.4	71.9	74.2	74.9	72.4	73.6	**81.0** (↑ 6.1)

1) As the contamination ratio increases, all the methods will suffer from performance degradation to a certain degree, no matter in the transductive or inductive setting.
2) IDGL still consistently outperforms the other comparing methods in all cases over five tested data sets, and the superiority of IDGL has shown to be more significant as the contamination ratio increases. Specifically, when the contamination ratio increases from 10% to 90%, IDGL has a gain over the second-best method of this case, from 5.4% to 9.0% (4.1% to 11.5%) on AR, from 9.8% to 14.6% (6.6% to 10.0%) on E-YaleB, from 4.4% to 7.6% (4.1% to 10.8%) on Multi-PIE, from 3.6% to 9.2% (7.6% to 7.7%) on FERET, and from 4.6% to 6.0% (6.0% to 7.0%) on CAS-PEAL, in the transductive (inductive) setting. The superior performances of IDGL can attribute to its two advantages. On the one hand, IDGL updates proper prototype to represent the neutral image of each person, which will narrow the gap between a query sample and the enrolment sample of the same person but with different variations and, meanwhile, enlarge the gap between a query sample and the enrolment samples of different persons but with the similar variation. On the other hand, IDGL presents a new way to learn a representative variation dictionary that can provide additive and sharable variations for reconstructing query samples.
3) S^3RC achieves better results than the generic learning methods, including SSRC, CPL, and SVDL in most cases because it contains a prototype learning stage to recover contaminated enrolment samples. However, S^3RC performs poorly on the E-YaleB data set. The reason is that the clustering of GMM in S^3RC is sensitive to illumination variations and shadows on the E-YaleB data set.
4) CPL and SVDL obtain poor performance over five tested data sets and even perform worse than SSRC in some cases, especially when the contamination ratio is high. This is because the two methods directly use contaminated samples as prototypes and then generate variation dictionaries under the premise. In this case, the generated variation dictionaries are probably unsuitable for the SSPP-ce FR problem.
5) The patch-based DMMA, SDMME, and PCRC are not competitive with the generic learning methods, such as SSRC and ESRC, in most cases and perform much worse than our IDGL.

In summary, this experiment verifies the superior performance of IDGL compared with the existing methods for SSPP-ce FR.

D. Investigation of the Learned P and Learned V

This section investigates the contributions of the learned prototypes and the learned variation dictionary in IDGL for SSPP-ce FR. To this end, we construct two typical methods denoted as Proto-IDGL and Dict-IDGL (refer to Section IV-B), by removing the variation dictionary learning step and the prototype learning via SSLRR stage in IDGL, respectively. In Proto-IDGL, the variation dictionary is borrowed from SSRC, while in Dict-IDGL, the original enrolment samples are directly used as the prototypes. Subsequently,

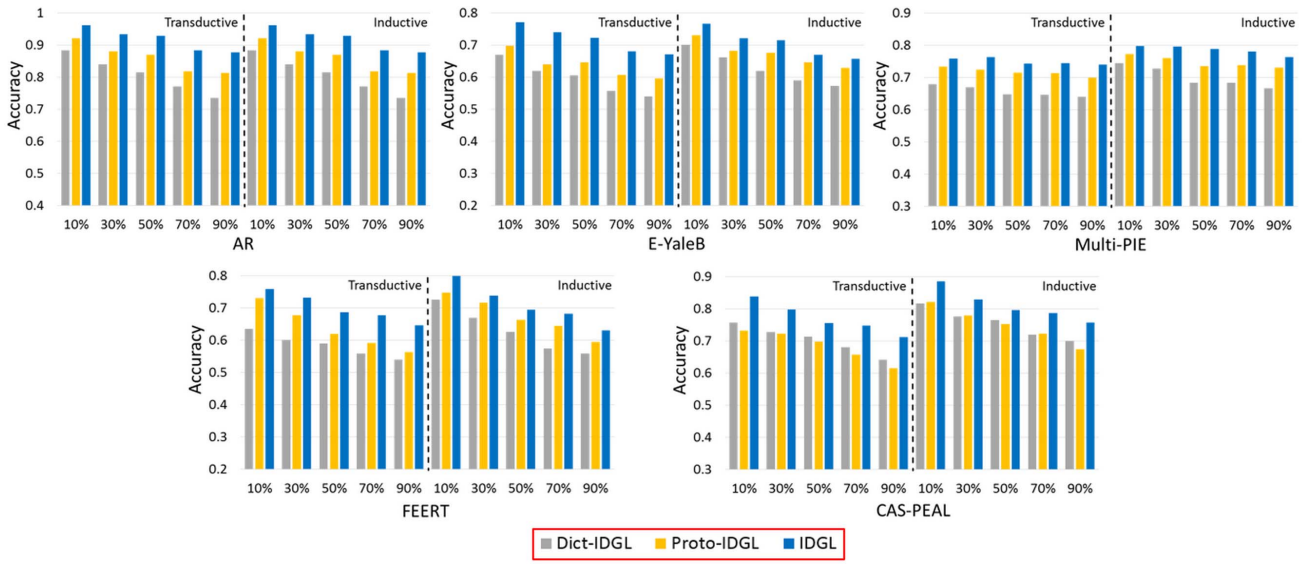


Fig. 10. Comparison results of Dict-IDGL, Proto-IDGL, and IDGL for SSPP-ce FR on AR, E-YaleB, Multi-PIE, FERET, and CAS-PEAL benchmark data sets.

TABLE V
TIME COMPLEXITY OF IDGL IN INDUCTIVE SETTING

Time complexity	Training			Recognition
	Stage I	Stage II: VDL	Stage II: Label estimation	
IDGL	$O(\tau(lk^2+l^3+l^2k))$	$O(sd^3)$	$O(\tau_1 d^2 q+\tau_1 d(c+S)q)$	$O(\tau_1 d(d+S))$
	$O(\tau_{max}\tau l^3+\tau_{max}\tau_1 dq(d+S)+sd^3)$			

l: the number of samples in the sample set matrix $\widehat{\mathbf{X}}$.
s: the number of persons in the generic set.
q: the number of query samples for training prototypes.
τ: the number of iterations in **Algorithm 1**.
τ_{max}: the maximum number of iterations in **Algorithm 3**.
c: the number of enrolment samples.
S: the total number of generic samples.
d: the dimension of the samples in $\widehat{\mathbf{X}}$.
τ_1: the number of iterations in BPDN-homotopy.
k: the rank of $\widehat{\mathbf{X}}$.

we compare the performances of Proto-IDGL and Dict-IDGL with IDGL for SSPP-ce FR in both transductive and inductive settings. The experimental settings on the five tested data sets, including AR, E-YaleB, Multi-PIE, FEERT, and CAS-PEAL follow the protocol in Section IV-C, and the values of the parameters in IDGL are retained unchanged. To keep a fair comparison, the parameters λ_1, α, β, and λ_3 in Proto-IDGL are also set as 15, 1, 2, and 0.001, respectively, and λ_2 and λ_3 in Dict-IDGL are set to be 0.05 and 0.001, respectively. We show the comparison results of Proto-IDGL, Dict-IDGL, and IDGL in Fig. 10.

From Fig. 10, it is observed that Proto-IDGL achieves better performances than Dict-IDGL on AR, E-YaleB, Multi-PIE, and FERET data sets and obtains comparable results with Dict-IDGL on CAS-PEAL data set. This observation indicates that the prototype learning can play a more critical role compared with the variation dictionary learning for SSPP-ce FR. Moreover, the learned variation dictionary also helps to address the SSPP-ce FR problem, as the integration of Proto-IDGL and Dict-IDGL, i.e., IDGL, has shown to consistently outperform either of the Proto-IDGL and Dict-IDGL in all the cases over five tested data sets. In a nutshell, the learned prototypes and the learned variation dictionary are complementary to each other in the learned P + learned V model, and both contribute to the performance of IDGL for SSPP-ce FR.

E. Computational Complexity Analysis

In this section, we analyze the computational complexity of IDGL in the 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, 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 (VDL) in (17) and (18) is $O(sd^3)$ [44], and the label estimation in (19)–(21) requires $O(\tau_1 d^2 q + \tau_1 d(c + S)q)$ [63], where τ_1 is the number of iterations for BPDN-homotopy. Let τ_{max} be the maximum number of iterations in Algorithm 3; then, the time complexity for training prototypes is $O(\tau_{max}\tau l^3+\tau_{max}\tau_1 dq(d+S)+sd^3)$ $(k < l$ and $c \ll S)$. In recognition phase, the time complexity for recognizing a new query sample is $O(\tau_1 d(d+S))$. For clarity, we summarize our complexity analysis of IDGL in Table V.

Furthermore, we evaluate the time cost of our IDGL method on the CAS-PEAL data set in the inductive scenario. The experiments are conducted on a host (CPU: Dual 6-core Intel Xeon X5650 2.66-GHz 12-MB L3 Cache; Memory: 32 GB). The training time that recovers proper prototypes requires 126.3861 s. Moreover, the recognition time on a new query sample is 0.0855 s, on average, which is fast and less than the acceptable 0.5 s.

TABLE VI

RECOGNITION ACCURACIES (%) OF IDGL AND THE OTHER GENERIC LEARNING AND PROTOTYPE LEARNING-BASED METHODS ON FRGC V2.0, WITH PIXELS AND THE OPENFACE FEATURE (DIMENSION = 128)

Methods	ESRC	SSRC	SVDL	CPL	S^3RC	IDGL
Pixels	67.4	69.8	67.8	70.9	70.4	**77.9**
Openface	89.6	91.7	92.1	91.6	93.2	**94.5**

F. Evaluation on Deep Learning-Based Features

This section includes two experimental parts. In Part I, we compare our IDGL with four recent generic learning methods, i.e., ESRC, SSRC, SVDL, and CPL, and the state-of-the-art prototype learning-based S^3RC on FRGC v2.0 data set, with both pixels and the deep learning-based Openface feature [64]. For the FRGC v2.0 data set, a subset of 5000 images of 250 persons is used. The first 200 persons are chosen for evaluation, while the rest 50 persons are chosen for generic learning. Then, we test the performances of all the methods by randomly selecting one image of each evaluated person as the enrolment sample and another image from the rest images for testing. The values of the parameters in IDGL and the other five comparing methods are kept the same as that in Section IV-C. We repeat the experiment five times and report the average accuracies of these methods on the FRGC v2.0 data set in Table VI. From Table VI, we observe that our IDGL consistently outperforms the other five comparing methods with either of the two types of features. Moreover, the Openface feature enhances the performance of IDGL remarkably compared with raw pixels, which verifies the power of deep learning-based features.

In Part II, we evaluate 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) [65] and InsightFace [37] features, i.e., IDGL+LightCNN-29 and IDGL+InsightFace, with four recent deep learning-based methods, including DeepID [32], VGG-face [31], center loss-based CNN [34], and joint and collaborative representation with local adaptive convolution feature (JCR-ACF) [35], on the unconstrained LFW data set. 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. Following the protocol in [35], we use a subset of 158 persons with no less than ten images per person from LFW-a for testing. The first 50 persons are selected for evaluation, while the rest 108 persons are used for generic learning. We randomly select a sample of each person as the enrolment sample and use the rest samples for recognition. We repeat the experiment five times and report the average recognition accuracies of all the methods. As shown in Table VII, NN+LightCNN-29 and NN+InsightFace have obtained quite high recognition accuracies of 98.3% and 94.5%, respectively, on the LFW data set. However, 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.

TABLE VII

RECOGNITION ACCURACIES (%) OF IDGL USING THE LIGHT CNN AND INSIGHTFACE FEATURES AND THE OTHER DEEP LEARNING-BASED METHODS ON LFW DATA SET

Methods	Accuracy (%)
DeepID	70.7
Center loss-based CNN	72.8
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**

TABLE VIII

RECOGNITION ACCURACIES (%) OF IDGL+LIGHTCNN-29 AND IDGL+INSIGHTFACE ON CELEBA AND IJB-C DATA SETS. WE HIGHLIGHT THE IMPROVEMENTS OF IDGL+LIGHTCNN-29 AND IDGL+INSIGHTFACE WITH RESPECT TO THE CORRESPONDING BASELINE METHODS IN **BOLD**

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**)

Furthermore, we introduce two more challenging unconstrained data sets, i.e., CelebA and IJB-C, 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 the CelebA data set, we randomly select 300 persons with ten images per person for testing, where the first 200 persons are used for evaluation and the rest 100 ones for generic learning. On the IJB-C data set, we select 200 videos from 200 persons for testing, where the first 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 enrolment sample and select another nine samples (or frames) for recognition. We repeat the experiment five times and report the average recognition results in Table VIII. It is observed that IDGL+LightenCNN-29 can further enhance the recognition performance over the baseline NN+LightCNN-29 on two tested data sets. The same situation applies to IDGL+InsightFace and NN+InsightFace. For example, IDGL+LightCNN-29 delivers 5.8% and 10.9% improvements over the NN+LightCNN-29 on the CelebA and IJB-C data sets, respectively. The promising results again verify the feasibility and effectiveness of combining our IDGL with deep learning-based features for practical SSPP-ce FR under unconstrained environments.

V. CONCLUSION

This article has proposed a novel IDGL method to address a new and more challenging problem in SSPP FR, i.e., SSPP-ce FR, where the biometric enrolment database is contaminated by nuisance facial variations in the wild. IDGL develops a dynamic label feedback network to update proper prototypes for contaminated enrolment samples. Moreover, IDGL introduces a new way to learn a representative variation dictionary

via extracting the "sample-specific" corruptions from an auxiliary generic set. The experiments on various constrained and unconstrained face data sets have demonstrated the superiority of IDGL, with the significant improvement of the performance over the state-of-the-art counterparts. In the future work, we will focus on the derivational problem of SSPP-ce FR, i.e., recognizing query samples from a heterogeneous domain, such as eyewitness sketches or infrared photographs, with a contaminated SSPP-based enrolment database.

REFERENCES

[1] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, "Face recognition from a single image per person: A survey," *Pattern Recognit.*, vol. 39, no. 9, pp. 1725–1745, Sep. 2006.

[2] S. Gao, K. Jia, L. Zhuang, and Y. Ma, "Neither global nor local: Regularized patch-based representation for single sample per person face recognition," *Int. J. Comput. Vis.*, vol. 111, no. 3, pp. 365–383, Feb. 2015.

[3] L. Best-Rowden, H. Han, C. Otto, B. F. Klare, and A. K. Jain, "Unconstrained face recognition: Identifying a person of interest from a media collection," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 12, pp. 2144–2157, Dec. 2014.

[4] F. Mokhayeri, E. Granger, and G.-A. Bilodeau, "Domain-specific face synthesis for video face recognition from a single sample per person," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 3, pp. 757–772, Mar. 2019.

[5] M. Ye, Y. Cheng, X. Lan, and H. Zhu, "Improving night-time pedestrian retrieval with distribution alignment and contextual distance," *IEEE Trans. Ind. Informat.*, vol. 16, no. 1, pp. 615–624, Jan. 2020.

[6] C. Yan *et al.*, "STAT: Spatial-temporal attention mechanism for video captioning," *IEEE Trans. Multimedia*, vol. 22, no. 1, pp. 229–241, Jan. 2020.

[7] C. Yan, L. Li, C. Zhang, B. Liu, Y. Zhang, and Q. Dai, "Cross-modality bridging and knowledge transferring for image understanding," *IEEE Trans. Multimedia*, vol. 21, no. 10, pp. 2675–2685, Oct. 2019.

[8] F. Liu, J. Tang, Y. Song, L. Zhang, and Z. Tang, "Local structure-based sparse representation for face recognition," *ACM Trans. Intell. Syst. Technol.*, vol. 7, no. 1, pp. 1–20, Oct. 2015.

[9] D. Cai, X. He, K. Zhou, J. Han, and H. Bao, "Locality sensitive discriminant analysis," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, Jan. 2007, pp. 1713–1726.

[10] Y. Zhou and S. Sun, "Manifold partition discriminant analysis," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 830–840, Apr. 2017.

[11] S. Li and Y. Fu, "Learning robust and discriminative subspace with low-rank constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 11, pp. 2160–2173, Nov. 2016.

[12] M. Pang, Y.-M. Cheung, R. Liu, J. Lou, and C. Lin, "Toward efficient image representation: Sparse concept discriminant matrix factorization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 11, pp. 3184–3198, Nov. 2019.

[13] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[14] M. Pang, B. Wang, Y.-M. Cheung, and C. Lin, "Discriminant manifold learning via sparse coding for robust feature extraction," *IEEE Access*, vol. 5, pp. 13978–13991, 2017.

[15] Z. Li, Z. Lai, Y. Xu, J. Yang, and D. Zhang, "A locality-constrained and label embedding dictionary learning algorithm for image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 2, pp. 278–293, Feb. 2017.

[16] T. Shu, B. Zhang, and Y. Y. Tang, "Sparse supervised representation-based classifier for uncontrolled and imbalanced classification," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Dec. 28, 2018, doi: 10.1109/TNNLS.2018.2884444.

[17] S. Yi, Z. He, Y.-M. Cheung, and W.-S. Chen, "Unified sparse subspace learning via self-contained regression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 10, pp. 2537–2550, Oct. 2018.

[18] T. Pei, L. Zhang, B. Wang, F. Li, and Z. Zhang, "Decision pyramid classifier for face recognition under complex variations using single sample per person," *Pattern Recognit.*, vol. 64, pp. 305–313, Apr. 2017.

[19] P. Zhu, L. Zhang, Q. Hu, and S. C. Shiu, "Multi-scale patch based collaborative representation for face recognition with margin distribution optimization," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Oct. 2012, pp. 822–835.

[20] J. Lu, Y.-P. Tan, and G. 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, Jan. 2013.

[21] P. Zhang, X. You, W. Ou, C. L. Philip Chen, and Y.-M. Cheung, "Sparse discriminative multi-manifold embedding for one-sample face identification," *Pattern Recognit.*, vol. 52, pp. 249–259, Apr. 2016.

[22] M. Pang, Y.-M. Cheung, B. Wang, and R. Liu, "Robust heterogeneous discriminative analysis for face recognition with single sample per person," *Pattern Recognit.*, vol. 89, pp. 91–107, May 2019.

[23] W. Deng, J. Hu, and J. Guo, "Extended SRC: Undersampled face recognition via intraclass variant dictionary," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 9, pp. 1864–1870, Sep. 2012.

[24] W. Deng, J. Hu, and J. Guo, "In defense of sparsity based face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 399–406.

[25] M. Yang, L. Van, and L. Zhang, "Sparse variation dictionary learning for face recognition with a single training sample per person," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 689–696.

[26] H.-K. Ji, Q.-S. Sun, Z.-X. Ji, Y.-H. Yuan, and G.-Q. Zhang, "Collaborative probabilistic labels for face recognition from single sample per person," *Pattern Recognit.*, vol. 62, pp. 125–134, Feb. 2017.

[27] C.-P. Wei and Y.-C.-F. Wang, "Undersampled face recognition via robust auxiliary dictionary learning," *IEEE Trans. Image Process.*, vol. 24, no. 6, pp. 1722–1734, Jun. 2015.

[28] Y.-F. Yu, D.-Q. Dai, C.-X. Ren, and K.-K. Huang, "Discriminative multi-scale sparse coding for single-sample face recognition with occlusion," *Pattern Recognit.*, vol. 66, pp. 302–312, Jun. 2017.

[29] W. Deng, J. Hu, and J. 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, Oct. 2018.

[30] M. Pang, Y.-M. Cheung, B. Wang, and J. Lou, "Synergistic generic learning for face recognition from a contaminated single sample per person," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 195–209, 2020.

[31] O. M. Parkhi *et al.*, "Deep face recognition," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, vol. 1, no. 3, 2015, p. 6.

[32] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1891–1898.

[33] Y. Sun, D. Liang, X. Wang, and X. Tang, "DeepID3: Face recognition with very deep neural networks," 2015, *arXiv:1502.00873*. [Online]. Available: http://arxiv.org/abs/1502.00873

[34] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 499–515.

[35] M. Yang, X. Wang, G. Zeng, and L. 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, Jun. 2017.

[36] W. Wen, X. Wang, L. Shen, and M. Yang, "Adaptive convolution local and global learning for class-level joint representation of face recognition with single sample per person," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 3537–3542.

[37] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4690–4699.

[38] Y. Gao, J. Ma, and A. 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, May 2017.

[39] J. Zhao *et al.*, "Dual-agent GANs for photorealistic and identity preserving profile face synthesis," in *Proc. Adv. Neural Inf. Process. Syst. (NIPs)*, 2017, pp. 66–76.

[40] R. Huang, S. Zhang, T. Li, and R. He, "Beyond face rotation: Global and local perception GAN for photorealistic and identity preserving frontal view synthesis," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2439–2448.

[41] Y.-A. Chen, W.-C. Chen, C.-P. Wei, and Y.-C.-F. Wang, "Occlusion-aware face inpainting via generative adversarial networks," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 1202–1206.

[42] F. Qiao, N. Yao, Z. Jiao, Z. Li, H. Chen, and H. Wang, "Geometry-contrastive GAN for facial expression transfer," 2018, *arXiv:1802.01822*. [Online]. Available: http://arxiv.org/abs/1802.01822
[43] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst. (NIPs)*, 2014, pp. 2672–2680.
[44] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.
[45] X. Zhu, Z. Ghahramani, and J. D. Lafferty, "Semi-supervised learning using Gaussian fields and harmonic functions," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2003, pp. 912–919.
[46] S. Xiao, W. Li, D. Xu, and D. Tao, "FaLRR: A fast low rank representation solver," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 4612–4620.
[47] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Proc. Neural Inf. Process. Syst. (NIPs)*, 2011, pp. 612–620.
[48] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, Jan. 2010.
[49] X. Fang, Y. Xu, X. Li, Z. Lai, and W. K. Wong, "Robust semi-supervised subspace clustering via non-negative low-rank representation," *IEEE Trans. Cybern.*, vol. 46, no. 8, pp. 1828–1838, Aug. 2016.
[50] J. Eckstein and D. P. Bertsekas, "On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, nos. 1–3, pp. 293–318, 1992.
[51] M. Hong, Z.-Q. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," *SIAM J. Optim.*, vol. 26, no. 1, pp. 337–364, Jan. 2016.
[52] D. L. Donoho and Y. 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, Nov. 2008.
[53] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[54] A. M. Martinez, "The AR face database," CVC, New Delhi, India, Tech. Rep. 24, 1998.
[55] A. S. Georghiades, P. N. Belhumeur, and D. 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, Jun. 2001.
[56] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," *Image Vis. Comput.*, vol. 28, no. 5, pp. 807–813, May 2010.
[57] W. Gao *et al.*, "The CAS-PEAL large-scale chinese face database and baseline evaluations," *IEEE Trans. Syst., Man, Cybern. A, Syst. Hum.*, vol. 38, no. 1, pp. 149–161, Jan. 2008.
[58] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.
[59] P. J. Phillips *et al.*, "Overview of the face recognition grand challenge," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2005, pp. 947–954.
[60] G. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Univ. Massachusetts, Amherst, MA, USA, Tech. Rep. 07-49, Oct. 2007.
[61] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3730–3738.
[62] B. Maze *et al.*, "IARPA janus Benchmark–C: Face dataset and protocol," in *Proc. Int. Conf. Biometrics (ICB)*, Feb. 2018, pp. 158–165.
[63] A. Y. Yang, Z. Zhou, A. Ganesh, S. Shankar Sastry, and Y. Ma, "Fast L1-minimization algorithms for robust face recognition," 2010, *arXiv:1007.3753*. [Online]. Available: http://arxiv.org/abs/1007.3753
[64] B. Amos *et al.*, "Openface: A general-purpose face recognition library with mobile applications," *CMU School Comput. Sci.*, vol. 6, p. 2, Jun. 2016.
[65] X. Wu, R. He, Z. Sun, and T. Tan, "A light CNN for deep face representation with noisy labels," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 11, pp. 2884–2896, Nov. 2018.

Meng Pang received the B.Sc. and M.Sc. degrees in software engineering from the Dalian University of Technology, Dalian, China, in 2013 and 2016, respectively, and the Ph.D. degree from the Department of Computer Science, Hong Kong Baptist University, Hong Kong, in 2019.

He is a Research Associate with the Department of Computer Science, Hong Kong Baptist University. His research interests include image processing and adversarial machine learning.

Yiu-Ming Cheung (Fellow, IEEE) received the Ph.D. degree from the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong.

He is currently a Full Professor with the Department of Computer Science, Hong Kong Baptist University, Hong Kong. His research interests include machine learning, pattern recognition, visual computing, and optimization.

Dr. Cheung is also a fellow of The Institution of Engineering and Technology (IET), British Computer Society (BCS), and Royal Society of Arts (RSA) and an IETI Distinguished Fellow. He also serves as an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON CYBERNETICS, and *Pattern Recognition*, to name a few. For more details, please refer to http://www.comp.hkbu.edu.hk/~ymc.

Qiquan Shi (Student Member, IEEE) received the B.E. degree in information security from the Computer School, Wuhan University, Wuhan, China, in 2013, and the Ph.D. degree in computer science from Hong Kong Baptist University, Hong Kong, in 2018.

He is currently a Researcher with the Huawei Noah's Ark Lab, Hong Kong. His current research interests include tensor decomposition, time-series analysis, large-scale optimization, and machine learning.

Mengke Li received the B.Sc. degree in communication engineering from Southwest University, Chongqing, China, in 2015, and the M.Sc. degree in electronic engineering from Xidian University, Xi'an, China, in 2018. She is currently pursuing the Ph.D. degree with the Department of Computer Science, Hong Kong Baptist University, Hong Kong, under the supervision of Prof. Y.-M. Cheung.

Her current research interests include image restoration and related fields.