Classification of Occluded Images for Large-Scale Datasets With Numerous Occlusion Patterns

FENG CEN[ID], XIAOYU ZHAO, WUZHUANG LI, AND FANGLAI ZHU
Department of Control Science and Engineering, Tongji University, Shanghai 201804, China

Corresponding author: Feng Cen (feng.cen@tongji.edu.cn)

ABSTRACT Large-scale image datasets with numerous occlusion patterns prevail in real applications. The classification scheme based on subspace decomposition-based estimation with squared l_2-norm regularization (SDBE_L2) has shown promising performance for the classification of partially occluded images. For the large-scale image datasets with numerous occlusion patterns, it however suffers from a high labor intensity in acquiring extra image pairs and a large consumption of computational resources in the training stage. To reduce the labor intensity, this paper enumerates several useful types of extra image pairs to guide the collection of extra images and introduces an intra-class random pairing method to semi-automatically form the extra image pairs. To alleviate the consumption of computational resources, this paper proposes two dictionary compression approaches: 1) uncentered PCA-based single partition compression (UPSPC), which compresses the dictionary to a size not larger than twice the column vector length without affecting the classification accuracy, and 2) uncentered PCA-based intra-class partition compression (UPIPC), which can further shrink the occlusion error dictionary (or class dictionary) when it has a small number of occlusion classes (or image classes). The proposed approaches are based on the property of SDBE_L2 being invariant to the uncentered PCA of sub-dictionaries. The extensive experiments on the Caltech-101 dataset and Oxford-102 flower dataset demonstrate the enumerated examples and the intra-class random pairing method facilitate acquiring the extra images and forming the extra image pairs only with a small loss in the classification accuracy. The experimental results on a large-scale occluded image dataset synthesized from the ILSVRC 2012 classification dataset with numerous occlusion patterns show that the proposed dictionary compression approaches reduce the dictionary size by over 11 times and shorten the training time by more than 39 times without loss in the classification accuracy.

INDEX TERMS Convolutional neural networks, SDBE, occlusion, image classification, principal component analysis.

I. INTRODUCTION

CClassification of partially occluded images is a long-standing challenge in computer vision [1]–[6]. Recently, many research attempts have been made to introduce the rapidly developing deep learning techniques [7]–[15] into this field. Under the deep learning framework, occlusion can be tackled in a low-level representation, *e.g.* the image itself [16]–[22] or low-level deep feature map [23], or a high-level representation, *e.g.* the deep feature vector (DFV) [24], of the image. None of these approaches, however, can handle the classification on large-scale generic

The associate editor coordinating the review of this manuscript and approving it for publication was Wenming Cao[ID].

image datasets with numerous occlusion patterns, which prevails in real applications.

To cope with the occlusion, the variations introduced by occlusion are usually needed to be modeled. For the low-level representation, the diversity of the occlusion-related variations are much richer than that for the high-level representation, since the low-level representation contains more details than the high-level representation. Consequently, for the low-level representation, much more training images, especially occluded training images, are usually required to model the occlusion-related variations than for the high-level representation.

For a large-scale image dataset with numerous occlusion patterns, the diversity of the occlusion-related variations in

the low-level representation is extremely rich and thus a huge number of occluded training images are usually required for the approaches based on the the low-level representation. Acquiring and labeling the occluded training images for the classification task, however, is a labor-intensive and time-consuming work. Lack of sufficient task-specific occluded training images discourages the development and application of the low-level representation based approaches. In contrast, much fewer occluded training images are needed for the high-level representation due to less diversity. Therefore, handling occlusion in the high-level representation seems more practicable for the large-scale datasets with numerous occlusion patterns.

In [24], a high-level representation based approach, namely subspace decomposition-based estimation (SDBE), was proposed to improve the classification performance against the occlusion on the occluded generic image datasets. The SDBE_L2 implementation of SDBE, which employs a squared l_2-norm to regularize the decomposition coefficients introducing very low extra computational cost in testing, is applicable to the large-scale datasets.

However, for the large-scale datasets with numerous occlusion patterns, the SDBE_L2-based classification scheme presented in [24] can exhaust the computational resources on many computers in the training stage. In [24], for each image class or occlusion pattern, at least tens of DFVs or occlusion error vectors (OEVs) were employed to construct the class dictionary (CD) or occlusion error dictionary (OED), respectively. For numerous image classes or occlusion patterns, this dictionary construction method gives rise to a dictionary of catastrophically large size, which requires a huge amount of memory to store and extremely long time to process. Besides, constructing an OED from exact-matching extra image pairs, which was adopted in [24], is a labor-intensive task for real applications. The exact-matching extra image pairs each is composed of a non-corrupted and a corrupted version of an image. In practice, this type of image pairs usually requires intensive manual labor to prepare due to lack of an automatic approach to precisely find out the corrupted versions of an occlusion-free image from plenty of occluded images.

In this paper, we focus on overcoming the above-mentioned deficiencies of the SDBE_L2-based classification scheme on the large-scale datasets with numerous occlusion patterns. We improve the SDBE_L2-based classification scheme from two aspects: 1) reducing the labor intensity in acquiring and forming the extra image pairs and 2) alleviating the computational resource consumption in the training stage.

To reduce the labor intensity, we first investigate what types of extra image pairs can contribute positively to the improvement on the classification of occluded images. Then, we enumerate several examples that can achieve a high overlap between the linear span of the generated OED and the *occlusion error subspace* as the empirical guidelines for collecting the extra images. Also, we introduce an intra-class random pairing method to effectively exploit the extra images

that lack exact-matching counterparts and reduce the labor intensity in forming the extra image pairs.

To alleviate the computational resource consumption, we propose two novel dictionary compression approaches: 1) uncentered PCA-based single partition compression (UPSPC) and 2) uncentered PCA-based intra-class partition compression (UPIPC). The proposed dictionary compression approaches are based on the uncentered principal component analysis (PCA). We prove that SDBE_L2 is invariant to the uncentered PCA of sub-dictionaries. Two methods are introduced to divide the CD and OED into sub-ditionaries: 1) *single parirition* and 2) *intra-class partition*. The *single partition* treats the CD or OED as a single sub-dictionary. While for the *intra-class partition*, the CD or OED is divided into sub-dictionaries according to the image classes or occlusion classes, respectively. In UPSPC, the compression is achieved by reserving the non-zero uncentered principal components (PCs) of each sub-dictionary of the *single partition*. The maximum size of the dictionary compressed by UPSPC is only twice the length of the column vectors. When the number of the image classes or occlusion classes is small, the size of the CD or OED can be further decreased by UPIPC. In UPIPC, the *intra-class partition* is employed and the compression is achieved by only reserving the first few PCs of each sub-dictionary.

In summary, our main contributions to improve the SDBE_L2-based classification scheme are as follows.

- We expand the admissible types of the extra image pairs and provide the empirical guidelines for the collection of the extra images to diminish the difficulty in acquiring the extra images.
- We introduce the intra-class random pairing method to reduce the workload in pairing the extra images.
- We propose two novel dictionary compression approaches: UPSPC and UPIPC, to alleviate the consumption of the computational resources in the training stage.

The paper is organized as follows. Section II summarizes the related works. Section III briefly reviews the SDBE_L2-based classification scheme. Section IV investigates the useful types of the extra image pairs and introduces the intra-class random pairing method. Section V describes the proposed dictionary compression approaches. Section VI presents experimental results and Section VII concludes the paper.

II. RELATED WORK

In the image space, deep generative models are commonly employed to restore the occluded image regions. A large number of exact-matching occluded and occlusion-free image pairs that can cover the variations caused by the occlusions are usually required to train the generative models. In [16], [18]–[20], the generative adversarial networks (GANs) were used to yield the missing portions of the image. Without particular considerations on the classification, however, the images recovered by purely inpainting are unsuitable for

image classification due to the inter-class information introduced by the generated patches, which degrades the classification accuracy. Furthermore, these approaches need to know the shapes and locations of the occlusions or missing portions in advance. Unfortunately, automatic occlusion detection in generic images is still a tough challenge in computer vision.

To avoid providing prior knowledge about the occlusion structure, learning the occlusion structure from data is developed for the generative model based approaches. In [25], an average relative difference between the activations of the occluded and the clean image was introduced to discriminate between the corrupted and the non-corrupted elements of the DFVs for the stacked sparse denoising auto-encoder. The corrupted elements are then replaced with the averaged non-corrupted elements to reconstruct the occlusion-free face image. Because of the holistic attribute of the DFVs and inter-class information introduced by the averaged non-corrupted elements, this approach only exhibits improvement for small-scale face datasets with a few types of synthetic occlusions.

To improve the classification performance for the generated model based approaches, in [26], an identity-based supervised CNN was proposed to provide an extra guidance for the training of robust LSTM-autoencoders to preserve the identity information. The trained robust LSTM-autoencoders, which have an LSTM branch to learn the structure of the occlusion were then employed to de-occlude face images. However, in order to cover the variations caused by the occlusions, massive exact-matching occluded and occlusion-free face image pairs are required to train the robust LSTM-autoencoders. The robust LSTM-autoencoders only shows improvement in face recognition.

In the deep feature space, the correspondence between the occluded image regions and the corrupted elements of the deep feature map, training with data augmentation, and recovering uncorrupted DFVs were investigated to improve the robustness of deep features to the occlusion. In [23], a pairwise differential Siamese network (PDSN) was proposed to learn the correspondence between the occluded facial regions and the corrupted activations of the top convolution layer. The occlusion-associated feature elements, which will be discarded in classification, are indicated by a discarding mask that is generated according to the learned correspondence. This approach however is highly rely on an accurate occlusion detector and well-aligned face images because the correspondence is built on a predefined image grid and the generation of the feature discarding mask requires an accurate detection of the occlusion. Consequently, this approach is not applicable to the classification of generic occluded images. In [27], a strategy for data augmentation with synthetic occluded face images was proposed to train the CNNs so as to extract the features more locally and equally. The occluded face images were synthesized by placing the occluder with a high probability at the face regions sensitive to the classification result. The sensitive face regions were identified by an occlusion map that was obtained via the visualization technique

proposed in [28]. The occlusion map, however, is only effective for well-aligned images, *e.g.*, aligned face images. In addition, to cover the variations caused by numerous occlusion patterns, an extremely large number of synthetic occluded images should be employed in training. This will prolong the training time dramatically. In [27], the improvement was only shown on small face datasets with a few occlusion patterns. In [24], the SDBE-based image classification scheme was proposed. The SDBE approach recovers the non-corrupted DFV by projecting the DFV of the occluded image onto the linear span of a CD along that of an OED. Two implementations of SDBE were introduced in [24], SDBE_L1 and SDBE_L2. For a large-scale dataset, SDBE_L2 has much less computational complexity in testing than but achieves classification accuracy similar to SDBE_L1. While, the implementation of SDBE_L2 in [24] is computational expensive in training for the large-scale datasets and numerous occlusion patterns. This is because the similarities between the OEVs associated with the same occlusion class and between the DFVs of the same image class are not effectively exploited. In addition, the redundancy in the dictionary is not removed.

From the above analysis, we can learn that for the large-scale generic image datasets with numerous occlusion patterns, the practicable classification scheme is still a challenge.

III. SDBE_L2-BASED CLASSIFICATION

In this section, we briefly describe SDBE_L2-based classification scheme for the sake of completeness. SDBE_L2 is designed to estimate the non-corrupted DFV by projecting the corrupted DFV onto the *class subspace* $\mathcal{A}$ along the *occlusion error subspace* $\mathcal{B}$. Because the projection is usually nonunique in practice, SDBE_L2 constrains the projection by minimizing the squared l_2-norm of the projection coefficients.

Suppose $N_{\mathbf{A}}$ training images of $K_{\mathbf{A}}$ image classes and $N_{\mathbf{B}}$ extra image pairs that are associated with $K_{\mathbf{B}}$ occlusion patterns are available. A base-CNN is employed to extract the m-dimensional DFV from each image. For the ith image class, there are $N_{\mathbf{A}_i}$ training images. Let $\mathbf{a}_{ij}$ denote the DFV of the jth training image in the ith image class. For the ith occlusion pattern, there are $N_{\mathbf{B}_i}$ extra image pairs. Let $\mathbf{b}_{ij}$ denote the OEV obtained from the jth extra image pair and associated with the ith occlusion pattern.

The OEV $\mathbf{b}_{ij}$ is calculated by $\mathbf{b}_{ij} = \mathbf{b}_{ij}^{o} - \mathbf{b}_{ij}^{f}$, where $\mathbf{b}_{ij}^{f}$ and $\mathbf{b}_{ij}^{o}$ are the DFVs of the occlusion-free and the occluded image of the jth extra image pair associated with the ith occlusion pattern, respectively. The CD $\mathbf{A}$ and OED $\mathbf{B}$ are constructed by concatenating the DFVs of the training images and the OEVs of the extra image pairs, respectively, *i.e.*, $\mathbf{A} = [\mathbf{A}_1, \ldots, \mathbf{A}_i, \ldots, \mathbf{A}_{K_{\mathbf{A}}}]$ with $\mathbf{A}_i = [\mathbf{a}_{i1}, \mathbf{a}_{i2}, \ldots, \mathbf{a}_{iN_{\mathbf{A}_i}}] \in \mathbb{R}^{m \times N_{\mathbf{A}_i}}$ and $\mathbf{B} = [\mathbf{B}_1, \ldots, \mathbf{B}_i, \ldots, \mathbf{B}_{K_{\mathbf{B}}}]$ with $\mathbf{B}_i = [\mathbf{b}_{i1}, \mathbf{b}_{i2}, \ldots, \mathbf{b}_{iN_{\mathbf{B}_i}}] \in \mathbb{R}^{m \times N_{\mathbf{B}_i}}$.

Then, the DFV of the ith test image, $\mathbf{v}_i$, is decomposed as follows,

$$\mathbf{v}_i = \mathbf{A}\boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\beta} + \mathbf{n}, \tag{1}$$

where $\mathbf{n}$ is a noise term and $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1^T, \ldots, \boldsymbol{\alpha}_i^T, \ldots, \boldsymbol{\alpha}_{K_A}^T]^T$ and $\boldsymbol{\beta} = [\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_i^T, \ldots, \boldsymbol{\beta}_{K_B}^T]^T$ are coefficients.

SDBE_L2 employs the squared l_2-norm regularized LS estimates to solve equation (1) as follows,

$$\hat{\boldsymbol{\omega}} = \underset{\boldsymbol{\omega}}{\operatorname{argmin}}\{||\mathbf{v}_i - \mathbf{D}\boldsymbol{\omega}||_2^2 + \lambda||\boldsymbol{\omega}||_2^2\}, \tag{2}$$

where λ is a positive hyperparameter, $\mathbf{D} = [\mathbf{A}\ \mathbf{B}]$, and $\boldsymbol{\omega}_i = [\boldsymbol{\alpha}^T\ \boldsymbol{\beta}^T]^T$.

The SDBE_L2-based classification scheme has two stages: training stage and testing stage. The primary purpose of the training stage is to construct the CD and OED and then figure out the projection matrix. The details of the training stage are listed below.

- **Input**: a set of training images and a set of extra image pairs.
1) For each image, use the base-CNN to extract the DFV.
2) Calculate the OEV $\mathbf{b}_{ij}$ for each extra image pair.
3) Concatenate all $\mathbf{a}_{ij}$'s to form the CD $\mathbf{A}$ and all $\mathbf{b}_{ij}$'s to create the OED $\mathbf{B}$.
4) Normalize each column of $\mathbf{A}$ and $\mathbf{B}$ to have unit l_2-norm (optional).
5) Calculate

$$\mathbf{P} = \left(\mathbf{D}^T\mathbf{D} + \lambda\mathbf{I}\right)^{-1}\mathbf{D}^T, \tag{3}$$

where $\mathbf{P} = [\mathbf{P}_\alpha^T\ \mathbf{P}_\beta^T]^T$, $\mathbf{P}_\alpha \in \mathbb{R}^{n_A \times m}$, $\mathbf{P}_\beta \in \mathbb{R}^{p_B \times m}$.
6) Calculate the class subspace projection matrix

$$\mathbf{O} = \mathbf{A}\mathbf{P}_\alpha. \tag{4}$$

7) Train the classifier $\mathcal{C}$ with the column vectors of $\mathbf{A}$ (optional).
- **Output**: $\mathbf{O}$ and $\mathcal{C}$.

The function of the testing stage is to predict the image class of the test image. The details of the testing stage are listed as follows.

- **Input**: $\mathbf{O}$, $\mathcal{C}$ and test image $\mathbf{y}_i$.
1) Use the base-CNN to extract the DFV of the test image $\mathbf{v}_i$.
2) Normalize $\mathbf{v}_i$ to have unit l_2-norm (optional).
3) Estimate the projection of the DFV $\mathbf{v}_i$ onto the *class subspace* by

$$\hat{\mathbf{v}}_{0i} = \mathbf{O}\mathbf{v}_i. \tag{5}$$

4) Predict the image class of $\hat{\mathbf{v}}_{0i}$ with the classifier $\mathcal{C}$.
- **Output**: The image class of $\hat{\mathbf{v}}_{0i}$.

While the testing stage is independent of the size of the dictionary $\mathbf{D}$ because $\mathbf{O} \in \mathbb{R}^{m \times m}$, the computational resources consumed to figure out $\mathbf{O}$ in the training stage, such as the memory and computing time, increase with the size of $\mathbf{D}$.

To calculate $\mathbf{O}$, the memory should be able to store the matrix $\mathbf{D}^T\mathbf{D}$, which is a matrix of size $(N_A + N_B) \times (N_A + N_B)$.

For a large-scale image dataset with numerous occlusion patterns, $(N_A + N_B)$ is very large such that the memory required to compute $\mathbf{O}$ can easily exceed the memory capacity of many computers. For instance, for a dataset with 1000 image classes and 2000 occlusion patterns, adopting the setting used in [24]: 20 DFVs for each image class and 100 OEVs for each occlusion pattern, the size of $\mathbf{D}^T\mathbf{D}$ is 220000×220000. Considering the commonly used double-precision representation of real value (8 bytes per value), the minimum memory required to store a 220000×220000 matrix is over 360GB, which exceeds the memory capacity of desktop computers.

The most time-consuming procedure in the training stage is inverting $\left(\mathbf{D}^T\mathbf{D} + \lambda\mathbf{I}\right)$, which has a computational complexity of over $\mathcal{O}((N_A + N_B)^{2.37})$ [31], [32]. Therefore, to process a huge size matrix, such as the above-mentioned example, the computing time will be extremely long.

Besides, in [24], employing the exact-matching extra image pairs to generate the OED restrains the practical application of the SDBE_L2-based classification scheme. In many applications, the exact-matching counterparts of the extra images are difficult to be acquired.

IV. ACQUIRING EXTRA IMAGE PAIRS

In this section, we investigate the types of the extra image pairs that can contribute positively to the classification of occluded images. Acquiring the extra image pairs includes two steps: 1) collecting the occluded images and occlusion-free images and 2) matching the occluded images to the occlusion-free images to yield the extra image pairs.

As mentioned in [24], the linear span of the OED $\mathbf{B}$ is regarded as an approximation of the *occlusion error subspace* $\mathcal{B}$. To achieve a small estimation error, the linear span of $\mathbf{B}$ should heavily overlap with $\mathcal{B}$, i.e., the overlapping ratio,

$$\delta = \frac{\mathbf{B} \cap \mathcal{B}}{\mathcal{B}}, \tag{6}$$

should be sufficiently large. In addition to the type of the extra image pairs used in [24], *i.e.*, covering all the occlusion patterns in the test images and having exact-matching between the occluded and the occlusion-free images, it is apparent that many alternatives can yield the OEDs with high δ.

Through a toy example, we first take a close look at the variation, caused by the occlusion, of the DFV. As shown in Fig.1, the DFVs of the occluded images, which reflect the features of both the occlusion patch and original images, move from the *class subspaces* $\mathcal{A}_i$'s to the linear span $\mathcal{A}_{oj}$ that is associated with the jth occlusion pattern. For instance, the black arrow line $\mathbf{v}_{0i}$ (the DFV of the original occlusion-free beaver image) moves from $\mathcal{A}_i$ to $\mathcal{A}_{oj}$ and becomes the black arrow line $\mathbf{v}_i$ (the DFV of a beaver image with 25% occlusion). The OEV between $\mathbf{v}_i$ and $\mathbf{v}_{0i}$, which falls into the *occlusion error subspace* $\mathcal{B}$, is drawn out as a solid red arrow line. In Fig. 1, the blue arrow lines are additional OEV examples associated with $\mathcal{B}$.

From Fig.1, we can learn that for a given occlusion pattern, the relative position between the *class subspaces* associated
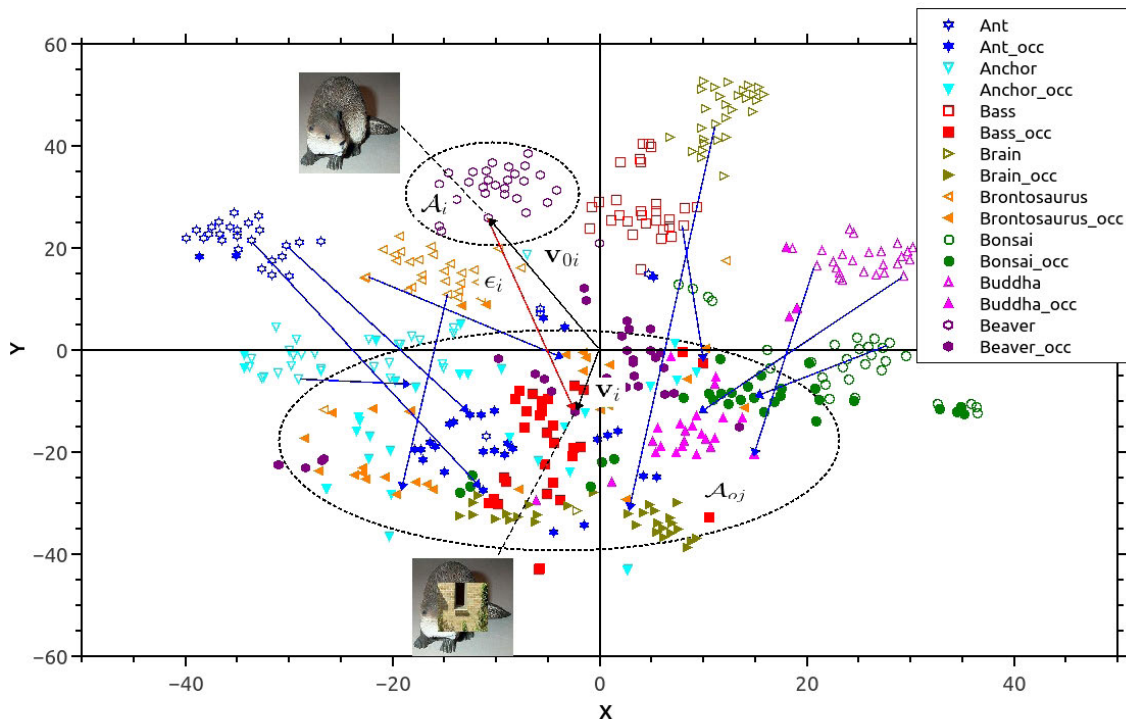
FIGURE 1. 2-D visualization of the DFVs of the occlusion-free and occluded images. The DFVs are extracted by using the pre-trained ResNet-152 network [8]. The t-SNE algorithm [29] is adopted to project the DFVs from the original 2048-D to the 2-D for visualization. The occlusion-free images are drawn from Caltech-101 dataset [30] and resized to 224 × 224. Only one occlusion pattern is illustrated. The occluded images are synthesized by contaminating the occlusion-free images at the center with an occlusion patch scaled to 25% of the occlusion-free images. The examples of occlusion-free images and occluded images are shown. To make the illustration clear, merely 8 image classes (30 occlusion-free images each) are plotted. The DFVs of occlusion-free images and occluded images are marked with hollow symbols and solid symbols, respectively. The red and blue arrow lines represent the OEVs and the black arrow lines the DFVs. (better viewed in color) .

with the occlusion-free extra images, *e.g.*, $\mathcal{A}_i$'s, and the linear span associated with the occlusion pattern, *e.g.*, $\mathcal{A}_{oj}$, determines the linear span of the OEVs. This indicates that any occluded and occlusion-free image pairs with appropriate relative positions in the deep feature space can be adopted as the extra image pairs because the linear span of their generated OEVs can have a high δ.

In practice, many types of occluded and occlusion-free image pairs can meet the requirement of high δ. To guide acquiring useful extra image pairs in practice, we enumerate some examples from three aspects: occlusion-free extra image, occluded extra image, and pairing method. The effectiveness of these examples is demonstrated in section VI-A.

A. OCCLUSION-FREE EXTRA IMAGE

The occlusion-free extra image determines the start point of the OEV. Suppose the test images are drawn from an image class set $\mathcal{T}$. The occlusion-free extra images can be drawn from either inside or outside $\mathcal{T}$.

For the case inside $\mathcal{T}$, the OEVs start from the space close to the start points of the OEVs of the occluded test images[1].

---

[1]The OEV of the occluded test image refers to the residual vector between the DFV of the occluded test image and its original occlusion-free image. The OEVs of the occluded test images definitely fall into $\mathcal{B}$.

While for the case outside $\mathcal{T}$, the start points of the OEVs locate in the space far from those of the OEVs of the occluded test images. The linear span of the OED, therefore, more likely deviates from $\mathcal{B}$ for the case outside $\mathcal{T}$ than for the case inside $\mathcal{T}$. Consequently, drawing from outside $\mathcal{T}$ usually shows smaller improvement than from inside $\mathcal{T}$.

Despite achieving smaller improvement, drawing from outside $\mathcal{T}$ facilitates the collection of extra images. For instance, in a flower classification task, the animal images can be adopted as the extra images. Here, we give two examples regarding the case outside $\mathcal{T}$.

1) Intra-dataset occlusion-free extra images: the occlusion-free extra images are drawn from the same dataset as the test images but distinct from the image classes of the test images.

2) Inter-dataset occlusion-free extra images: the occlusion-free extra images are drawn from the datasets distinct from the task-specific dataset, For instance, the OED used for the classification on the Oxford-102 dataset [33] is constructed by drawing the occlusion-free extra images from the Caltech-101 dataset [30](see the experiment in section VI-A2).

Since the images from both inside and outside $\mathcal{T}$ can be adopted as the occlusion-free extra images, the occluded extra

images and the pairing method become the deterministic factors for the usefulness of the extra image pairs.

B. OCCLUDED EXTRA IMAGE

The occluded extra image determines the endpoint of the OEV. Suppose the occluded test images and the occluded extra images are associated with the jth and j'th occlusion patterns, respectively. If $\mathbf{A}_{oj'}$, the linear span of the DFVs of the occluded extra images, is heavily overlapped with $\mathcal{A}_{oj}$, the OEVs of the extra image pairs would mainly end within the space that can produce $\mathbf{B}_{j'}$, the linear span of the OEVs of the extra image pairs, of high δ.

In addition to the jth occlusion pattern, it is easy to infer that the following occlusion patterns can give rise to an $\mathbf{A}_{oj'}$ heavily overlapped with $\mathcal{A}_{oj}$.

1) The intra-class occlusion pattern: the occlusion patch is distinct from but belongs to the same occlusion class as the occlusion patch of the jth occlusion pattern.
2) The linear transformation occlusion pattern: the occlusion pattern is a translation, rotation, or scaling of the jth occlusion pattern.
3) The noised occlusion pattern: the occlusion patch is a noised version, *e.g.*, corrupted by the Gaussian noise, of the occlusion patch of the jth occlusion pattern.
4) The cropped occlusion pattern: the occlusion patch is cropped from the occlusion patch of the jth occlusion pattern or vice versa.

For an extreme scaling in the second item, *e.g.*, enlarging the occlusion patch from 25% occlusion to 80% (see the experiment in section VI-A), $\mathcal{A}_{oj}$ and $\mathbf{A}_{oj'}$ may not be overlapped with each other. In such situation, however, both $\mathcal{A}_{oj}$ and $\mathbf{A}_{oj'}$ are in the same direction from $\mathcal{A}_i$ to the linear span associated with the image class of the occlusion patch. It means that the OEVs associated with the extremely scaled occlusion pattern point towards the direction similar to the OEVs associated with the jth occlusion pattern. Therefore, $\mathcal{B}_j$ and the linear span of $\mathbf{B}_{j'}$ can have large overlapping.

C. PAIRING METHOD

In practical applications, many extra images may not have their exactly matched counterparts. For instance, for face recognition in the wild, it is almost unable to find an occlusion-free face image and an occluded face image having the same head pose, face expression, and illumination. In such situation, the exact-matching method used in [24] is unable to efficiently exploit the extra images. In particular, for a small set of extra images, the extra image pairs generated by using the exact-matching method will be insufficient to produce an OED to approximate $\mathcal{B}$ with a small error. On the other hand, finding the exactly matched occlusion-free and occluded images from plenty of extra images usually consumes a vast amount of labor.

To alleviate these problems, we introduce an *intra-class random pairing* method, where each extra image pair is formed by randomly combining an occlusion-free extra

image with an occluded extra image of the same image class. The *intra-class random pairing* method is a semi-automatic method only requiring to determine the image classes of the occluded extra images manually. Therefore, much less labor is needed to pair the extra images. It can also effectively exploit the extra images lacking the exactly corresponding counterparts in the set of collected extra images.

For a well-trained CNN, the DFVs of the occlusion-free images of the same image class cluster together, and thus the image class determines the approximate location of the start point of an OEV. Meanwhile, the DFVs associated with the same occlusion pattern locate close to each other since these DFVs reflect the features of the same occlusion patch. For a given occlusion pattern, it is apparent that the start points and endpoints of the *intra-class random pairing* OEVs have small deviations from the respective points of the *exact-matching* OEVs since these two types of OEVs are associated with the same image class and the same occlusion pattern. Therefore, the intra-class random pairing OED, which is constructed by using the extra image pairs that are formed with the *intra-class random pairing* method, can have high δ.

V. DICTIONARY COMPRESSION

In this section, we introduce the uncentered PCA-based dictionary compression approaches. As the basis of the proposed approaches, the uncentered PCA of sub-dictionaries is first described, and then, the strategies of dictionary partition and compression are presented.

A. UNCENTERED PCA OF SUB-DICTIONARIES

The dictionary $\mathbf{D}$ can be split into multiple sub-dictionaries. A sub-dictionary is a subset of the column vectors of $\mathbf{D}$. Let $\mathbf{D}_i$ denote the ith sub-dictionary and $\boldsymbol{\omega}_i$ the vector of coefficients associated with $\mathbf{D}_i$. Then, $\mathbf{D}$ can be written as $\mathbf{D} = [\ldots, \mathbf{D}_i, \ldots]$ and the associated coefficients $\boldsymbol{\omega} = [\ldots, \boldsymbol{\omega}_i^T, \ldots]^T$.

The uncentered PCA is given via a singular value decomposition (SVD) of the matrix[2]. The SVD of $\mathbf{D}_i$ is given by

$$\mathbf{D}_i = \mathbf{L}_i \Sigma_i \mathbf{R}_i^T \,, \tag{7}$$

where $\mathbf{L}_i$ and $\mathbf{R}_i$ are the orthogonal matrices of the left and the right singular vectors, respectively, and Σ_i is a diagonal matrix of singular values in descending order. The column vectors of $\mathbf{R}_i$ are actually the eigenvectors of $\mathbf{D}_i^T \mathbf{D}_i$, thereby being viewed as the the principal axes of the uncentered PCA [36] of $\mathbf{D}_i$.

To avoid confusing with the standard PCA [37], [38], where the matrix is column-centered, in this paper, the notations for the uncentered PCA is modified by ''uncentered'', *e.g.*, the principal axes of the uncentered PCA is named uncentered principal axes.

Equation (7) can be rewritten as $\mathbf{D}_i \mathbf{R}_i = \mathbf{L}_i \Sigma_i$. The left term represents the projection of $\mathbf{D}_i$ onto the uncentered

---

[2]In the literature, many algorithms have been proposed to solve the SVD of an extremely large matrix [34], [35].

principal axes. The columns of $\mathbf{L}_i\Sigma_i$ therefore can be interpreted as the uncentered PCs[3]. Let

$$\mathbf{D}_i^{pca} = \mathbf{L}_i\Sigma_i \qquad (8)$$

and $\boldsymbol{\omega}_i^{pca} = \mathbf{R}_i^T\boldsymbol{\omega}_i$. $\boldsymbol{\omega}_i^{pca}$ is the projection of $\boldsymbol{\omega}_i$ onto the uncentered principal axes. Then, we have

Lemma 1: Let $F(\boldsymbol{\omega}) = ||\mathbf{v}_i - \mathbf{D}\boldsymbol{\omega}||_2^2 + \lambda||\boldsymbol{\omega}||_2^2$ and $\widetilde{F}(\widetilde{\boldsymbol{\omega}}) = ||\mathbf{v}_i - \widetilde{\mathbf{D}}\widetilde{\boldsymbol{\omega}}||_2^2 + \lambda||\widetilde{\boldsymbol{\omega}}||_2^2$, where $\widetilde{\mathbf{D}} = [\ldots, \mathbf{D}_{i-1}, \mathbf{D}_i^{pca}, \mathbf{D}_{i+1}, \ldots]$ and $\widetilde{\boldsymbol{\omega}} = [\ldots, \boldsymbol{\omega}_{i-1}^T, (\boldsymbol{\omega}_i^{pca})^T, \boldsymbol{\omega}_{i-1}^T, \ldots]^T$. Then, $F(\boldsymbol{\omega}) = \widetilde{F}(\widetilde{\boldsymbol{\omega}})$.

Proof: Because $\mathbf{D}_i^{pca}\boldsymbol{\omega}_i^{pca} = \mathbf{L}_i\Sigma_i\mathbf{R}_i^T\boldsymbol{\omega}_i = \mathbf{D}_i\boldsymbol{\omega}_i$, we have $\widetilde{\mathbf{D}}\widetilde{\boldsymbol{\omega}} = \mathbf{D}\boldsymbol{\omega}$. Then, we obtain

$$||\mathbf{v}_i - \widetilde{\mathbf{D}}\widetilde{\boldsymbol{\omega}}||_2^2 = ||\mathbf{v}_i - \mathbf{D}\boldsymbol{\omega}||_2^2. \qquad (9)$$

Recall that $\mathbf{R}_i$ is an orthogonal matrix, *i.e.*, $\mathbf{R}_i\mathbf{R}_i^T = \mathbf{I}$, where $\mathbf{I}$ denotes the identity matrix. We have

$$(\boldsymbol{\omega}_i^{pca})^T\boldsymbol{\omega}_i^{pca} = \boldsymbol{\omega}_i^T\mathbf{R}_i\mathbf{R}_i^T\boldsymbol{\omega}_i \qquad (10)$$
$$= \boldsymbol{\omega}_i^T\boldsymbol{\omega}_i. \qquad (11)$$

From equation (11), we obtain

$$||\boldsymbol{\omega}||_2^2 = \boldsymbol{\omega}^T\boldsymbol{\omega} \qquad (12)$$
$$= \sum_j \boldsymbol{\omega}_j^T\boldsymbol{\omega}_j \qquad (13)$$
$$= \sum_{j\neq i} \boldsymbol{\omega}_j^T\boldsymbol{\omega}_j + (\boldsymbol{\omega}_i^{pca})^T\boldsymbol{\omega}_i^{pca} \qquad (14)$$
$$= \widetilde{\boldsymbol{\omega}}^T\widetilde{\boldsymbol{\omega}} \qquad (15)$$
$$= ||\widetilde{\boldsymbol{\omega}}||_2^2. \qquad (16)$$

From equation (9) and (16), we have $F(\boldsymbol{\omega}) = \widetilde{F}(\widetilde{\boldsymbol{\omega}})$. ∎

It is easy to infer from Lemma 1 that in the uncentered principal space of $\mathbf{D}_i$, which is defined by $\mathbf{R}_i$, equation (2) keeps unchanging.

By repeatedly replacing each $\mathbf{D}_i$ with $\mathbf{D}_i^{pca}$, we have the following theorem.

Theorem 1: If $\mathbf{D}_i$'s in $F(\boldsymbol{\omega})$ are substituted with $\mathbf{D}_i^{pca}$'s, then the solution of equation (2) can be obtained by replacing $\hat{\boldsymbol{\omega}}_i$'s in $\hat{\boldsymbol{\omega}}$ with $\mathbf{R}_i^T\hat{\boldsymbol{\omega}}_i$'s.

Theorem 1 indicates that SDBE_L2 is invariant to the uncentered PCA of sub-dictionaries. Moreover, the projection of a sub-dictionary onto the uncentered principal axes gives rise to a significant-first form for the sub-dictionary since the uncentered PCs carrying more information come before those capturing less information.

These properties facilitates the dictionary compression. First, the size of $\mathbf{D}_i^{pca}$ is not larger than $m \times m$ because the number of the non-zero columns of $\mathbf{D}_i^{pca}$ is equal to the rank of $\mathbf{D}_i$, which is smaller than or equal to m. Second, if the original column vectors of a sub-dictionary are highly correlated, the information of the sub-dictionary concentrates on the first few uncentered PCs, which correspond to the large singular values. Consequently, the sub-dictionary can be compressed by only reserving the first few uncentered PCs.

---

^{3}In the literature, some authors confusingly call the column vectors of $\mathbf{R}_i$ "principal components" (*e.g.*, [38]), but we reserve this name for $\mathbf{D}_i\mathbf{R}_i$ in keeping with the terminology in [37].

B. DICTIONARY PARTITION AND COMPRESSION

The above analysis encourages us to compress the dictionary via the uncentered PCA of sub-dictionaries. However, how to partition the dictionary $\mathbf{D}$ into sub-dictionaries has yet to be investigated.

An intuitive partition is to consider the CD $\mathbf{A}$ or OED $\mathbf{B}$ as a single sub-dictionary. We name this type of partition *single partition*. The *single partition* can ensure that $\mathbf{A}$ or $\mathbf{B}$ is compressed to a size not larger than $m \times m$. Let $m_\mathbf{A}$ or $m_\mathbf{B}$ denote the numbers of non-zero PCs of $\mathbf{A}$ or $\mathbf{B}$, respectively. Because the rank of $\mathbf{A}$ or $\mathbf{B}$ is not larger than m, we have $m_\mathbf{A} \leq m$ or $m_\mathbf{B} \leq m$. By dropping off the zero PCs, $\mathbf{A}$ or $\mathbf{B}$ is shrunk to a size smaller than or equal to $m \times m$ in the unentered principal space. We call this type of compression uncentered PCA-based single partition compression (UPSPC). The UPSPC algorithm is presented in Algorithm 1. UPSPC is a type of lossless compression due to the invariability of SDBE_L2 to the uncentered PCA.

Algorithm 1 The Proposed UPSPC Algorithm

Input: $\mathbf{A}$ or $\mathbf{B}$.
1) Let $\mathbf{D}_1 = \mathbf{A}$ or $\mathbf{B}$.
2) Conduct SVD on $\mathbf{D}_1$ according to equation (7).
3) Compute $\mathbf{D}_1^{pca}$ by equation (8).
4) Drop off all of the zero vectors in $\mathbf{D}_1^{pca}$.
Output: $\mathbf{A}^{\mathrm{UPSPC}} = \mathbf{D}_1^{pca}$ or $\mathbf{B}^{\mathrm{UPSPC}} = \mathbf{D}_1^{pca}$.

UPSPC provides a lower bound for the compression. For $\mathbf{D}$ associated with a small number of image classes or occlusion classes, a higher compression ratio is able to be achieved. We divide the CD or OED into sub-dictionaries according to the image classes or occlusion classes, respectively, *i.e.*, put the CD column vectors of the same image class into the same sub-dictionary or the OED column vectors of the OED associated with the same occlusion class into the same sub-dictionary. We name this type of partition *intra-class partition*. The column vectors of the same image class or occlusion class are usually highly correlated. Therefore, for the *intra-class partition*, the first few uncentered PCs of each sub-dictionary capture most of the information of the sub-dictionary and can represent the sub-dictionary approximately.

Suppose there are $K_\mathbf{A}$ image classes and $\kappa_\mathbf{B}$ occlusion classes in $\mathbf{D}$. Let $m_{\mathbf{A}_i}$ and $m_{\mathbf{B}_j}$ denote the numbers of the reserved uncentered PCs for the ith sub-CD and the jth sub-OED, respectively. If $\sum_1^{K_\mathbf{A}} m_{\mathbf{A}_i} < m_\mathbf{A}$ or $\sum_1^{\kappa_\mathbf{B}} m_{\mathbf{B}_j} < m_\mathbf{B}$, then $\mathbf{A}$ or $\mathbf{B}$ can be compressed to a size smaller than that achieved by UPSPC by only reserving the first few uncentered PCs of each sub-dictionary. We call this type of compression uncentered PCA-based intra-class partition compression (UPIPC). The UPIPC algorithm is presented in Algorithm 2. Unlike UPSPC, the classification accuracy for UPIPC has a small deviation from that for the uncompress dictionary due to discarding some uncentered PCs in each sub-dictionary.
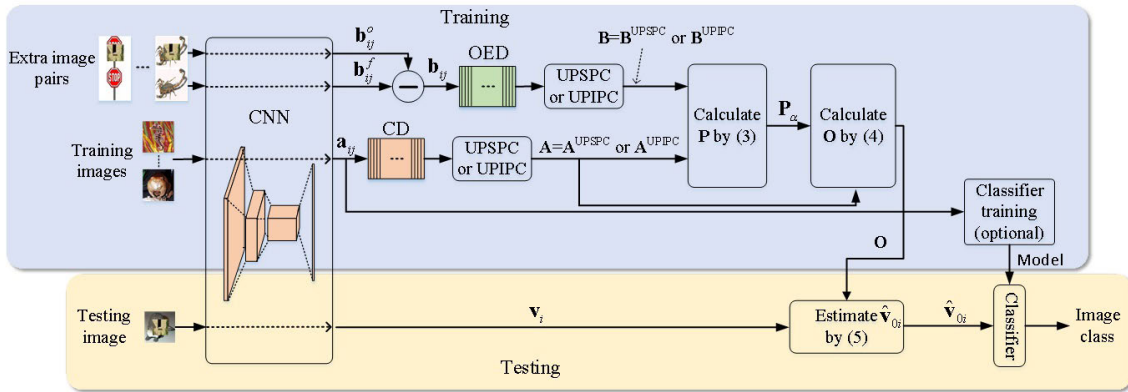
FIGURE 2. The SDBE_L2-based classification scheme with the proposed dictionary compression.

Algorithm 2 The Proposed UPIPC Algorithm

Input: $\mathbf{A}$ and $\{m_{\mathbf{A}_i}\}_{i=1}^{K_{\mathbf{A}}}$, or $\mathbf{B}$ and $\{m_{\mathbf{B}_i}\}_{i=1}^{\kappa_{\mathbf{B}}}$.

1) Divide $\mathbf{A}$ into sub-matrices according to the image classes: $\mathbf{A} = [\mathbf{A}_1, \ldots, \mathbf{A}_i, \ldots, \mathbf{A}_{K_{\mathbf{A}}}]$, or divide $\mathbf{B}$ into sub-matrices according to the occlusion classes: $\mathbf{B} = [\mathbf{B}_1, \ldots, \mathbf{B}_i, \ldots, \mathbf{B}_{\kappa_{\mathbf{B}}}]$.

2) Let $\mathbf{D}_i = \mathbf{A}_i, i \in \{1, \ldots, K_{\mathbf{A}}\}$ or $\mathbf{D}_i = \mathbf{B}_i, i \in \{1, \ldots, \kappa_{\mathbf{B}}\}$.

3) Conduct SVD on $\mathbf{D}_i$ according to equation (7).

4) Compute $\mathbf{D}_i^{pca} = [\mathbf{d}_{i1}, .., \mathbf{d}_{ij}, \ldots]$ by equation (8), where $\mathbf{d}_{ij}$ is the jth uncentered PC of $\mathbf{D}_i^{pca}$.

5) Let $\mathbf{D}_i^{\text{UPIPC}} = [\mathbf{d}_{i1}, .., \mathbf{d}_{im_{\mathbf{A}_i}}]$ or $\mathbf{D}_i^{\text{UPIPC}} = [\mathbf{d}_{i1}, .., \mathbf{d}_{im_{\mathbf{B}_i}}]$.

Output: $\mathbf{A}^{\text{UPIPC}} = [\mathbf{D}_1^{\text{UPIPC}}, \ldots, \mathbf{D}_{K_{\mathbf{A}}}^{\text{UPIPC}}]$ or $\mathbf{B}^{\text{UPIPC}} = [\mathbf{D}_1^{\text{UPIPC}}, \ldots, \mathbf{D}_{\kappa_{\mathbf{B}}}^{\text{UPIPC}}]$.

C. CLASSIFICATION SCHEME

The overall SDBE_L2-based classification scheme integrated with the proposed dictionary compression approaches is illustrated in Fig. 2. In the training stage, the extra image pairs can be generated with either the exact-matching or the proposed *intra-class random pairing* method. The constructed CD and OED are compressed with the proposed UPSPC or UPIPC algorithm independently. Then, the compressed CD and OED are concatenated to form the compressed dictionary $\mathbf{D} = [\mathbf{A} \ \mathbf{B}]$, where $\mathbf{A} = \mathbf{A}^{\text{UPSPC}}$ or $\mathbf{A}^{\text{UPIPC}}$ and $\mathbf{B} = \mathbf{B}^{\text{UPSPC}}$ or $\mathbf{B}^{\text{UPIPC}}$. The remaining training procedures and the testing stage are the same as the original SDBE_L2-based classification scheme.

D. ANALYSIS OF COMPUTATIONAL RESOURCE CONSUMPTION

For the large-scale datasets with numerous occlusion patterns, it is apparent that $N_{\mathbf{A}} \gg m$ and $N_{\mathbf{B}} \gg m$. Therefore, the maximum size of the matrices needed to be stored is $m \times (N_{\mathbf{A}} + N_{\mathbf{B}})$, which is the size of the dictionary before

compression. Such a maximum size is much smaller than that of the original SDBE_L2-based scheme, which is $(N_{\mathbf{A}} + N_{\mathbf{B}}) \times (N_{\mathbf{A}} + N_{\mathbf{B}})$.

Regarding the computational complexity, unlike the original SDBE_L2-based classification scheme, inverting $(\mathbf{D}^T\mathbf{D} + \lambda\mathbf{I})$ is not a computationally intensive procedure anymore, because the size of $\mathbf{D}$ becomes not larger than $m \times m$ after dictionary compression. The most time-consuming procedure is the SVD of the uncompressed CD or OED. The SVD has a computational complexity of $\mathcal{O}(mN_{\mathbf{A}})$ or $\mathcal{O}(mN_{\mathbf{B}})$ for the CD or OED, respectively [39]. This complexity is much lower than that of inverting $(\mathbf{D}^T\mathbf{D} + \lambda\mathbf{I})$ in the original SDBE_L2-based scheme, since $N_{\mathbf{A}} \gg m$ and $N_{\mathbf{B}} \gg m$ for the large-scale datasets with numerous occlusion patterns.

VI. EXPERIMENTS

In the experiments, the proposed algorithms were implemented with Matlab. Like [24], the MatConvNet [40] implementation of the ResNet-152 network [8], which was pre-trained on the ILSVRC2012 classification dataset [41], was adopted as the base-CNN to extract the 2048-D DFVs. The experiments are conducted on a PC with an i7 CPU and 64GB memory and without GPU acceleration.

A. EXPERIMENTS ON ACQUIRING EXTRA IMAGE PAIRS

In this section, we illustrate the effectiveness of the various examples enumerated in Section IV. The experiments were conducted on the Caltech-101 dataset [30] and Oxford-102 flower dataset [33].

The l_2-regularized l_2-loss linear SVM [42] was adopted as the classifier, which was trained on the normalized DFVs of the training images for each experiment. The regularization parameter of the linear SVM was drawn from the grid set $\Theta = \{2^{-15}, \ldots, 2^0, \ldots, 2^{15}\}$. To match the input size of the base-CNN, all of the images were directly resized to 224×224. The occluded images were synthesized by superimposing the occlusion patches scaled to specific occlusion ratios on the resized images.

The optional l_2 normalization procedures (step 4 in the training stage and step 2 in the testing stage) were adopted in the experiments due to a strong preference of the linear SVM for normalized vectors. The hyperparameter λ of SDBE_L2 is drawn from the grid set $\mathbf{\Lambda} = \{10^{-6}, \ldots, 0.5, 1, \ldots, 10\}$.

1) CALTECH-101 DATASET

The Caltech-101 dataset excluding the ''background'' image class is adopted for evaluation. The settings on the training images, occlusion-free test images, and occlusion-free extra images are the same as in [24]. In particular, the dataset is split into a *class set* (80 image classes with names from ''accordion'' to ''schooner'' in alphabet order) and an *extra set* (remaining 21 image classes). The training images and the occlusion-free test images are drawn from the *class set*. The occlusion-free extra images are drawn, unless otherwise specified, from the *extra set*. From each image class in the *class set*, 30 images and a maximum of 50 images are randomly drawn as the training images and occlusion-free test images, respectively. The occlusion-free extra images, 30 images per image class, are randomly drawn from each image class of the *extra set*.

Two experiments were conducted on the Caltech-101 dataset to assess the enumerated examples on the collection of occluded extra images and pairing methods. In each trial, only the best result with respect to the regularization parameter of the linear SVM and the hyperparameter λ of SDBE_L2 is reported.

In the first experiment, the cropping, noising, scaling, rotation, and translation of the occlusion patch and intra-class random pairing method with intra-dataset occlusion-free images (example 1 in Section IV-A, example 2, 3, and 4 in SectionIV-B, and example in Section IV-C) were evaluated. The test images were synthesized by using a single occlusion pattern (25% occlusion at the center of the image), named original occlusion pattern, to contaminate the occlusion-free test images. The example of the occluded image for the original occlusion pattern is indexed by 1 in Fig.3a.

For each trial, the OED was constructed by employing the occluded extra images synthesized only with the testing occlusion pattern. In Fig.3, each trial, except the 22nd trial, is numbered by the index of the occlusion pattern used to synthesize the occluded extra images. The occluded extra images of the first and 22nd trials were corrupted by using the original occlusion pattern. The first 21 trials adopted the exact-matching OEDs, while the last trial employed the intra-class random pairing OED. The OED used in each trial includes 630 OEVs associated with 21 image classes, each 30 OEVs.

Suppose the test images are contaminated with the kth occlusion pattern and the OED is constructed by using the extra image pairs associated with the lth occlusion pattern. The intersection between $\mathcal{B}_k$ and the linear span of $\mathbf{B}_l$ is unable to be measured directly since the exact span of $\mathcal{B}_k$ is unavailable. Instead, we appraise the intersection between the linear spans of $\mathbf{B}_l$ and $\mathbf{B}_k$ as an approximation since the linear
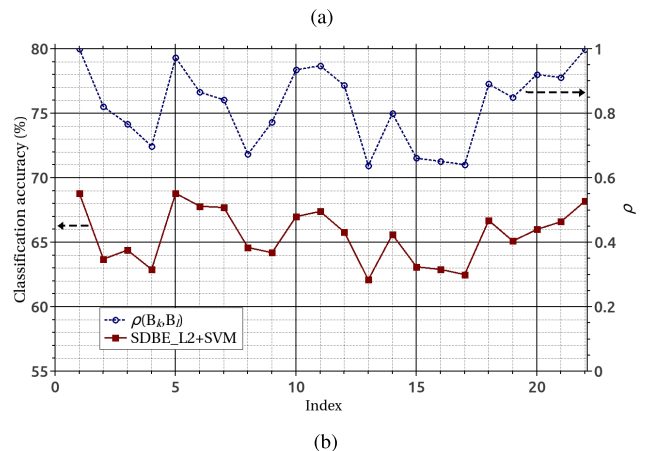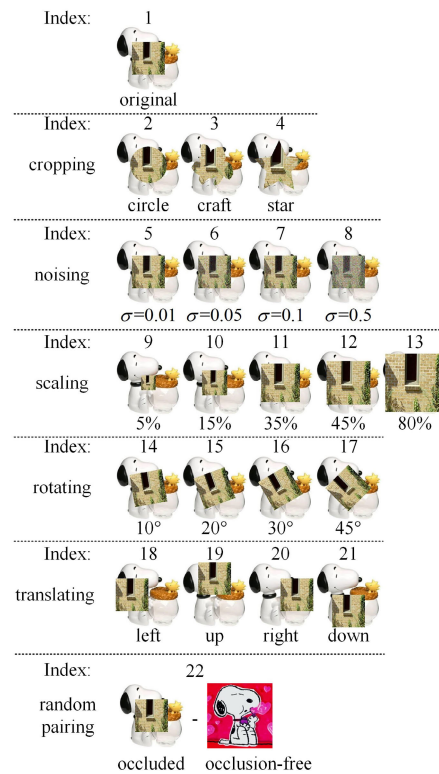

(a)

(b)

FIGURE 3. Comparison of the OEDs constructed with various occlusion patterns and distinct pairing methods. (a) shows the example extra images employed to construct the OED for each trial. (b) plots the classification accuracy and normalized mean correlation $\rho(\mathbf{B}_k, \mathbf{B}_l)$ with respect to the index of the evaluation. The occlusion ratio of the test image is 25%.

span of $\mathbf{B}_k$ apparently has the smallest deviation from the $\mathcal{B}_k$ among all of the occlusion patterns.

The normalized mean correlation between the column vectors of $\mathbf{B}_k$ and $\mathbf{B}_l$ is employed to assess the degree of the intersection between their linear spans. The normalized mean correlation is defined as

$$\rho(\mathbf{B}_k, \mathbf{B}_l) = \frac{\sum_i \sum_j \rho_{ij}(\mathbf{B}_k, \mathbf{B}_l)}{\sqrt{\sum_i \sum_j \rho_{ij}(\mathbf{B}_k, \mathbf{B}_k)}\sqrt{\sum_i \sum_j \rho_{ij}(\mathbf{B}_l, \mathbf{B}_l)}},$$

(17)

where $\rho_{ij}(\mathbf{B}_k, \mathbf{B}_l)$ is the Pearson correlation coefficient [43] between the ith OEV of $\mathbf{B}_k$ and the jth OEV of $\mathbf{B}_l$. A high magnitude of $\rho(\mathbf{B}_k, \mathbf{B}_l)$ indicates a strong correlation and hence large intersection.

The result for each trial is shown in Fig.3b. Comparing with the approach without employing SDBE_L2, which achieves 55.4% classification accuracy, all of the evaluated variants of the original OED (index 1) improve the performance significantly, though smaller than the original OED. Therefore, in practice, we can collect the occluded images associated with the occlusion patterns being of the cropping, nosing, scaling, rotation, and translation of the original occlusion patterns. Also, we note that the intra-class random pairing method achieves the improvement similar to the exact-matching method (index 22 vs. index 1). So, in practice, we can employ the intra-class random pairing method to save much labor.

By comparing the dash(blue) line and the solid(red) line in Fig.3b, we can learn that the improvement is highly related to the normalized mean correlation ρ. The results manifest that the overlapping between the linear span of the OED and the *occlusion error subspace* is a deterministic factor for the performance of SDBE_L2. This also indicates that in practice, ρ is able to be employed to estimate the effectiveness and contribution of the acquired extra image pairs to the performance improvement.

In the second experiment, we evaluated the performances for the intra-class occlusion pattern (example 1 in Section IV-B). The classification results are plotted in FIGURE4. For each testing occlusion ratio, the OED is constructed by using the extra image pairs associated merely with the testing occlusion ratio, *e.g.*, for the test of 25% occlusion, the occluded extra images only with 25% occlusion are employed to construct the OED. At each testing occlusion ratio, the intra-class occlusion pattern $\mathbf{z}_2$ achieves a significant improvement, close to that achieved by the original occlusion pattern $\mathbf{z}_1$, over that without SDBE_L2, whereas the inter-class occlusion pattern $\mathbf{z}_3$ achieves a very small improvement.

From the above two experiments, we know that if the occlusion patterns in the extra image pairs are associated with the same occlusion classes as those in the occluded test images, the OED can provide positive contribution to the classification. This indicates that the OEVs associated with the same occlusion class are highly correlated.

2) OXFORD-102 FLOWER DATASET
We conducted a performance comparison between the intra-dataset occlusion-free extra images and the inter-dataset occlusion-free extra images (example 1 and 2 in Section IV-A) on the Oxford-102 flower dataset.

The first 83 image classes of the Oxford-102 flower dataset were regarded as the *class set*. The images of the *class set* in the training set, validation set, and testing set, which are defined in [33], were adopted as the training images (10 images per image class), occlusion-free validation images
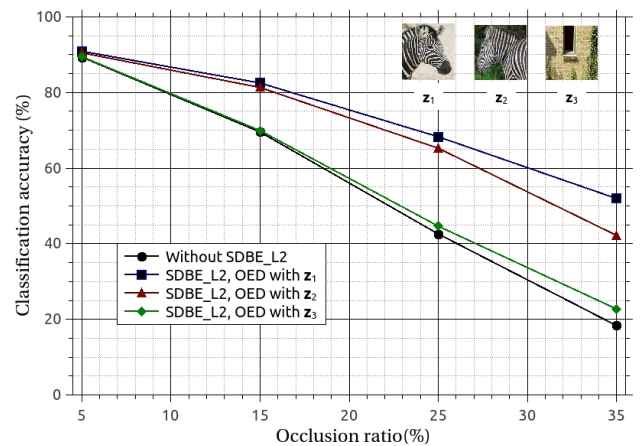

FIGURE 4. Comparison of the classification accuracies with respect to the occlusion ratio for the OEDs associated with the original occlusion patch $\mathbf{z}_1$, intra-class occlusion patch $\mathbf{z}_2$, and inter-class occlusion patch $\mathbf{z}_3$. $\mathbf{z}_1$ and $\mathbf{z}_2$ are drawn from the same occlusion class and $\mathbf{z}_1$ and $\mathbf{z}_3$ from distinct occlusion classes. The occluded test images are synthesized by contaminating the occlusion-free test images with the occlusion patch $\mathbf{z}_1$ scaled to the testing occlusion ratios.

(10 images per image class), and occlusion-free test images, respectively.

Two types of *extra sets* were evaluated for comparison. The first type, where the occlusion-free extra images were randomly drawn from the remaining 19 image classes each with 30 images, is a case for the intra-dataset occlusion-free extra images; the second one is a case for the inter-dataset occlusion-free extra images, which adopts occlusion-free images used in section VI-A1, except for those in "sunflower" and "water_lily", which are two image classes of flowers.

For each occlusion ratio, eight occlusion patterns - four occlusion patches (shown in Fig.5a) at two occlusion positions: center and off-center (illustrated in the examples of occluded images in Fig.5b) - were employed to synthesize the occluded images.

Similar to the experiment in FIGURE4, for each testing occlusion ratio, the occluded extra images were synthesized with the occlusion patterns of the testing occlusion ratio. The exact-matching method was employed to form the extra image pairs. The occluded and occlusion-free validation images were employed to determine the regularization parameter for the linear SVM and the hyperparameter λ for SDBE_L2. After the hyperparameters were obtained, the training images and the occlusion-free validation images were combined as the set of training images to train the linear SVM for testing.

The experimental result is reported in Fig.5c. From the results, we can observe that the extra sets of the second type achieves significant improvement, though smaller than the extra sets of the first type, for the occluded test images. This result demonstrates that both the intra-dataset and the inter-dataset occlusion-free extra images are applicable to SDBE_L2. It is worth noting that the intra-dataset
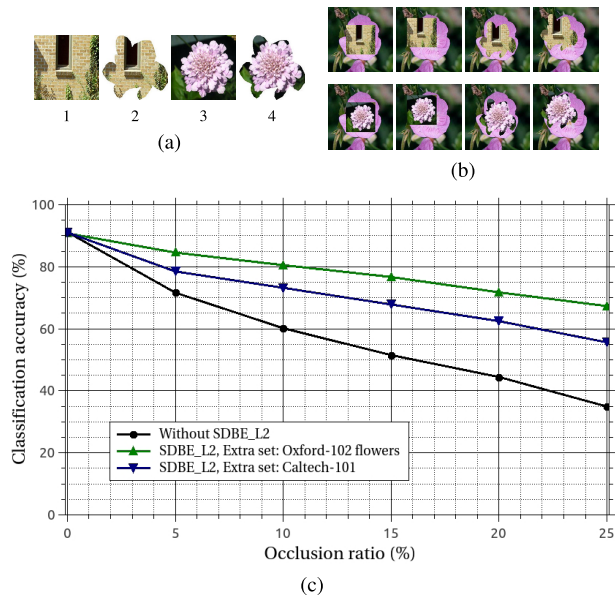
FIGURE 5. Comparison of classification accuracy for the intra-dataset occlusion-free extra images and inter-dataset occlusion-free extra images on the Oxford-102 flower dataset. (a) Four occlusion patches used in the experiment. (b) Examples of occluded images with 20% occlusion and two occlusion positions. (c) The classification accuracy with respect to the occlusion ratio.

occlusion-free extra images can be irrelevant to the task. In practice, exploiting the task-irrelevant images alleviates the difficulty in collecting the extra images dramatically.

B. EXPERIMENTS ON DICTIONARY COMPRESSION

In this section, we concern ourselves with the dictionary compression. We evaluated both UPSPC and UPIPC on a large-scale synthetic occluded image dataset with numerous occlusion patterns. The original "fc1000" and "prob" layers of the pre-trained ResNet-152 network, which actually constitute a softmax classifier, were adopted as the classifier in the experiments.

Due to lack of publicly available large-scale datasets with numerous occlusion patterns, we adopted the ILSVRC2012 dataset as the set of occlusion-free images and synthesize the occluded images from the occlusion-free images. To match the input size of the base-CNN, the occlusion-free images were first resized to the shorter side of 256 and then cropped the center with the size of 224×224 out as the occlusion-free images for further processing.

The ILSVRC2012 dataset was split into the *class set*, *extra set*, and *occlusion patch set*. The splitting configuration is the same as in [24], *i.e.*, 900 image classes in the *class set*, 20 image classes in the *extra set*, and 80 image classes in the *occlusion patch set*.

To make the uncompressed CD tractable for comparison in our computer, a subset of original training images, which was formed by randomly drawing 20 original training images from each image class of the *class set*, was adopted as the training set for SDBE_L2.

The original validation images in the *class set* were treated as the occlusion-free test images. To reduce the evaluation workload, the occluded test images were synthesized from a subset of the occlusion-free test images, which was obtained by randomly drawn from the occlusion-free test images each image class with five images. We compared the classification accuracies of the subset and the whole set of the occlusion-free test images with the original ResNet-152 network and observed merely a very small increase (0.34%) in classification accuracy for the subset. This indicates that the statistic of the subset has a negligible deviation from that of the whole set, and thus the occluded test images synthesized from the subset can statistically represent those generated from the whole set.

A total of 288 occlusion patterns were employed to synthesize the occluded test images. 36 occlusion patches, as shown in FIGURE 6a, were segmented from the images that are randomly drawn from 12 image classes, each 3 images, of the *occlusion patch set*. Each occlusion patch has 2 occlusion ratios (10% and 20%) and 8 corruption positions (4 positions for each occlusion ratio), which are randomly and independently generated for each occlusion patch.

A total of 100 occlusion-free extra images, which were randomly drawn from 20 image classes of the *extra set*, each image class 5 images, were used in the experiment. 28800 occluded extra images were synthesized with all 288 occlusion patterns from the occlusion-free extra images. Two types of original OEDs, exact-matching OED and intra-class random pairing OED, which are denoted by the superscripts *e* and *i*, respectively, were evaluated in the experiment.

By employing the OED construction method presented in [24], the generated OED is a 2048×28800 matrix. Taking the CD of 2048×18000 into account, we got the matrix **D** of 2048×46800. Accordingly, $\mathbf{D}^T\mathbf{D}$, which is used to compute **P**, is a matrix of 46800×46800. For the double-precision representation, $\mathbf{D}^T\mathbf{D}$ requires over 17GB memory to store, which is unable to be handled by many desktop computers.

We first evaluated UPSPC for the CD and OED. To show the advantage of the uncentered PCA over the centered PCA, we also evaluated the center PCA-based single partition compression (CPSPC), where the uncentered PCA of UPSPC is replaced with the centered PCA. We employ the superscriptions "UPSPC" and "CPSPC" to denote the CDs or OEDs compressed by using UPSPC and CPSPC, respectively.

The classification results on the synthetic occluded test images are tabulated in TABLE 1. The original CDs or OEDs are indicated by the superscription "org". UPSPC successfully reduces the size of the dictionary to $m \times 2m$ (over 11 times smaller than the original dictionary) without affecting the classification accuracy for both the cases of the exact-matching OED and intra-class random pairing OED ($\mathbf{D} = [\mathbf{A}^{\text{UPSPC}} \ \mathbf{B}^{\text{UPSPC},e}]$ vs. $\mathbf{D} = [\mathbf{A}^{\text{org}} \ \mathbf{B}^{\text{org},e}]$ and $\mathbf{D} = [\mathbf{A}^{\text{UPSPC}} \ \mathbf{B}^{\text{UPSPC},i}]$ vs. $\mathbf{D} = [\mathbf{A}^{\text{org}} \ \mathbf{B}^{\text{org},i}]$). While, CPSPC fails to compress the dictionary because the classification

(a)

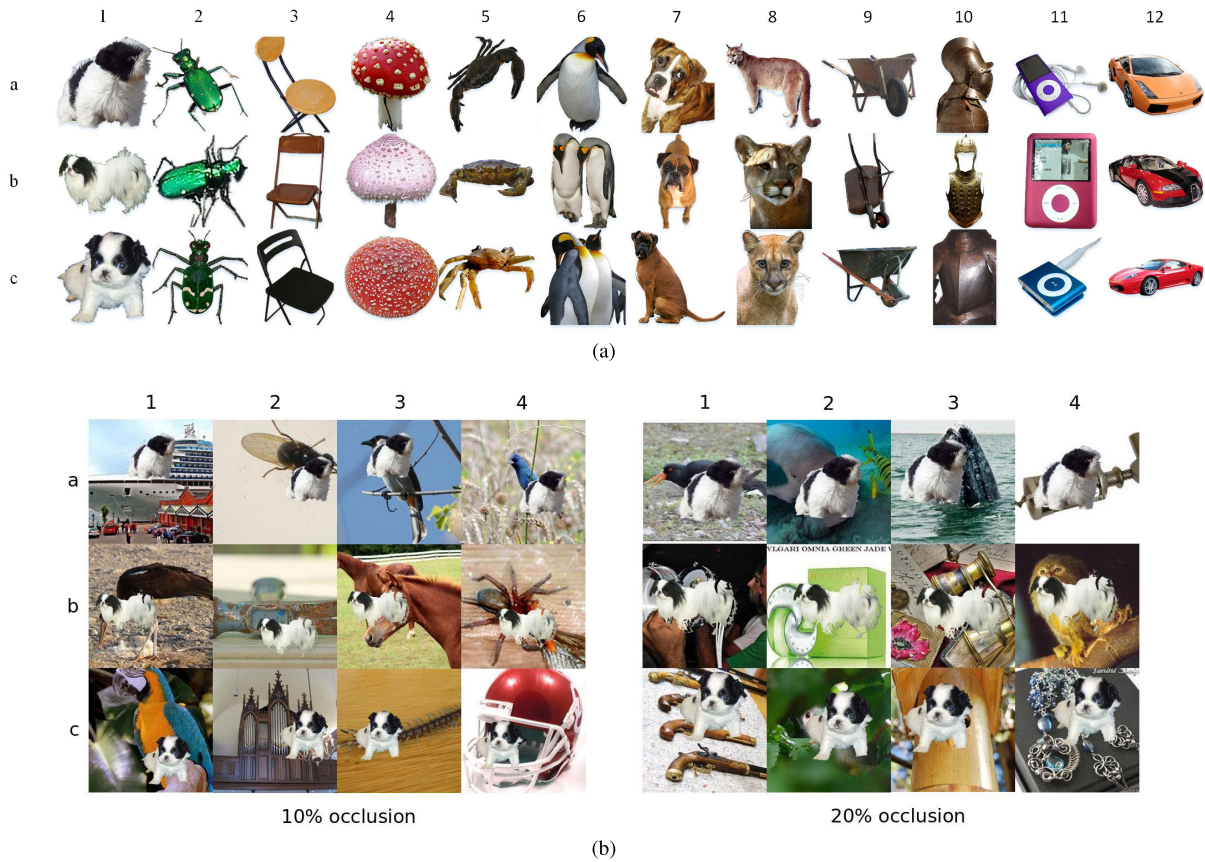10% occlusion   20% occlusion

(b)

FIGURE 6. (a) 36 occlusion patches that were used to synthesize the occluded images. Each column corresponds to an occlusion class. (b) Examples of the occluded test images for the first occlusion class. Two occlusion ratios: 10% and 20%, and four random occlusion positions for each occlusion patch at given occlusion ratio are shown. The occlusion positions for each occlusion patch at each occlusion ratio were independently and randomly sampled.

TABLE 1. Comparison of dictionary compression for the *single partition*. The results for SDBEŁ2 with uncompressed dictionary and the original ResNet-152 networks are also list for comparison. OR: occlusion ratio.

Method	$\mathbf{D}$	CD		OED		OR		Avg.
		$M_{\mathbf{A}}$	$m_{\mathbf{A}_i}$	$M_{\mathbf{B}}$	$m_{\mathbf{B}_j}$	10%	20%	
Original ResNet-152	-	-	-	-	-	51.03%	26.27%	38.65%
SDBE_L2	$[\mathbf{A}^{\mathrm{org}}\ \mathbf{B}^{\mathrm{org},e}]$	18000	20	28800	2400	62.19%	49.06%	55.62%
SDBE_L2	$[\mathbf{A}^{\mathrm{UPSPC}}\ \mathbf{B}^{\mathrm{UPSPC},e}]$	2048	-	2048	-	62.19%	49.06%	55.62%
SDBE_L2	$[\mathbf{A}^{\mathrm{CPSPC}}\ \mathbf{B}^{\mathrm{CPSPC},e}]$	2047	-	2047	-	2.15%	0.74%	1.45%
SDBE_L2	$[\mathbf{A}^{\mathrm{org}}\ \mathbf{B}^{\mathrm{org},i}]$	18000	20	28800	2400	61.79%	48.52%	55.15%
SDBE_L2	$[\mathbf{A}^{\mathrm{UPSPC}}\ \mathbf{B}^{\mathrm{UPSPC},i}]$	2048	-	2048	-	61.79%	48.52%	55.15%
SDBE_L2	$[\mathbf{A}^{\mathrm{CPSPC}}\ \mathbf{B}^{\mathrm{CPSPC},i}]$	2047	-	2047	-	3.74%	1.44%	2.59%

accuracy decreases almost to zero, *i.e.*, the cases of $\mathbf{D} = [\mathbf{A}^{\mathrm{CPSPC}}\ \mathbf{B}^{\mathrm{CPSPC},e}]$ and $\mathbf{D} = [\mathbf{A}^{\mathrm{CPSPC}}\ \mathbf{B}^{\mathrm{CPSPC},i}]$.

In our experiments, for the evaluations on the dictionaries compressed with UPSPC, *i.e.*, $\mathbf{D} = [\mathbf{A}^{\mathrm{UPSPC}}\ \mathbf{B}^{\mathrm{UPSPC},e}]$ and $\mathbf{D} = [\mathbf{A}^{\mathrm{UPSPC}}\ \mathbf{B}^{\mathrm{UPSPC},i}]$, the maximum memeory consumption is lower than 3.5GB and the training time shorter than 100 seconds. While for the evaluations of the original SDBE_L2-based classification scheme, *i.e.*, $\mathbf{D} = [\mathbf{A}^{\mathrm{org}}\ \mathbf{B}^{\mathrm{org},e}]$ and $\mathbf{D} = [\mathbf{A}^{\mathrm{org}}\ \mathbf{B}^{\mathrm{org},i}]$, the maximum memory

consumption is higher than 40GB (over 11 times than UPSPC) and the training time longer than 1.1 hours (over 39 times than UPSPC).

Then, we evaluated UPIPC. Similarly, the centered PCA-based intra-class partition compression (CPIPC), where the uncentered PCA of UPIPC is substituted with the centered PCA, was evaluated for comparison. By the superscriptions "UPIPC" and "CPIPC", we denote the CDs or OEDs compressed by using UPIPC and CPIPC, respectively.
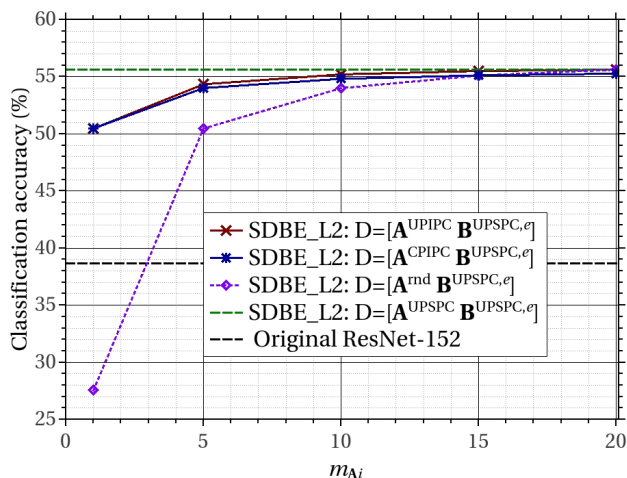
FIGURE 7. Classification accuracy with respect to m_{A_i} for SDBE_L2 with different types of compressed CDs. The result of the original ResNet-152 is also shown for comparison.

In addition, for comparison, the random selection-based compression approach, denoted by the superscription "rnd", was also evaluated for the *intra-class partition*. In the random selection-based compression approach, m_{A_i} (or m_{B_j}) column vectors are randomly drawn from the ith (or jth) instra-class sub-CD (or sub-OED) to construct the CD (or OED).

We first evaluated the compression for the CD. Without loss of generality, only the exact-matching OED was adopted for evaluation. For simplicity, m_{A_i}'s were set to the same value for all i's. The classification accuracy averaged over all of the occluded test images is shown in FIGURE7, where only the results for $\mathbf{B}^{\text{UPSPC},e}$ are plotted because $\mathbf{B}^{\text{org},e}$ and $\mathbf{B}^{\text{UPSPC},e}$ show the same results.

From FIGURE7, we can observe that the uncenter PCA-based approach is slightly better than the centered PCA-based approach and much better than the random selection-based approach. Although for *intra-class partition*, the compressed CD can achieve a classification accuracy close to the uncompressed CD by only reserving 5 PCs for each sub-dictionary, the size of the compressed CD is much lager than for the *single partition* (2048 × 4500 vs. 2048 × 2048) owing to the large number of the image classes. This result shows that for a large number of image classes, the *intra-class partition* is worse than the *single partition*.

Next, we evaluated the compression for the OED. Similar to the experiment for the CD, m_{B_j}'s were set to the same value for all j's. UPIPC, CPIPC, and the random selection based compression approach were evaluated. Both the exact-matching OED and the intra-class random pairing OED were adopted for compression. The classification accuracy averaged over all of the occluded test images are shown in FIGURE8.

From FIGURE8, we note that by only reserving 5 PCs for each sub-dictionary, *i.e.*, a total of 60 column vectors for the compressed OED (much less than those for the *single partition*, 60 vs. 2048), both the OEDs compressed by using UPIPC and CPIPC achieve a slightly higher classification accuracy than that compressed by using UPSPC, which has the same results as the original OED. This result demonstrates that for a small number of occlusion classes, the *intra-class partition* can achieve higher compression ratio than the *single partition*.

The slightly higher classification accuracy can be contributed to the mitigation of the influence of the outlier OEVs. The original OEDs usually contain some outlier OEVs that are harm to SDBE_L2 and mainly captured by the PCs associated with small singular values. By ignoring the less
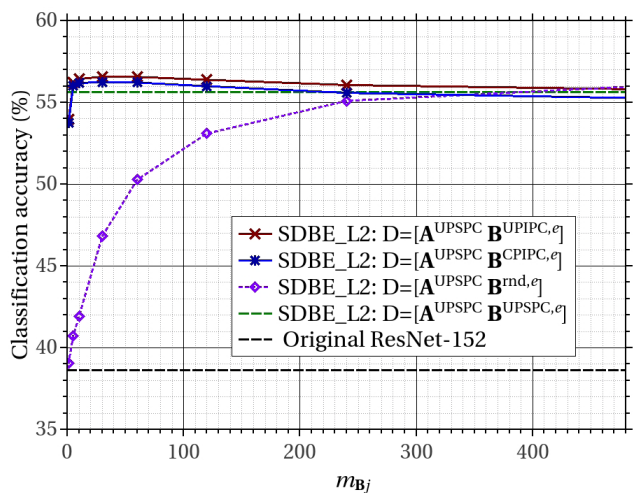
(a)
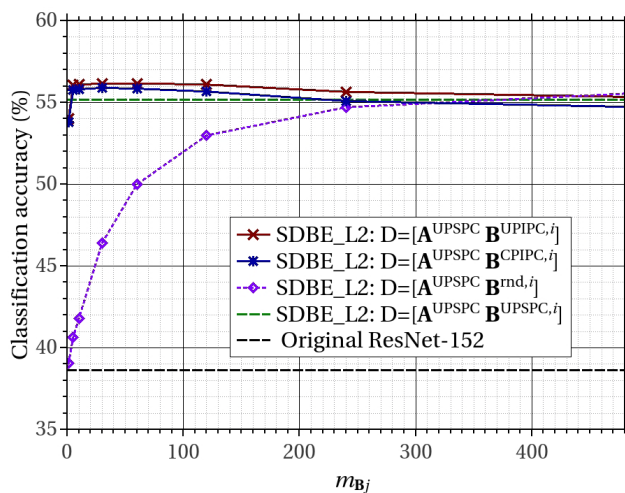
(b)

FIGURE 8. Classification accuracy with respect to m_{B_j} for SDBE_L2 with different compressed OEDs. The original OED is constructed from (a) exact-matching extra image pairs or (b) intra-class random pairing extra image pairs. The result of the original ResNet-152 is also shown for comparison.

important PCs, the compressed OED can alleviate the influence of the outlier OEVs, thus achieving better result.

We can also observe that the uncentered PCA-based approach, UPIPC, achieves the best results for both the exact-matching OED and intra-class random pairing OED. This indicates that the advantage of the proposed approach is irrelevant to the pairing method.

In the above experiment for compression, we should note that if a centered PCA-based compression approach is applied to both the CD and OED, the classification accuracy is degraded dramatically; while, if it is applied only to one of the CD and OED, the classification result is just slightly worse than that achieved by the uncentered PCA-based compression approach. The SDBE approach is based on the uncorrelation or independency between the CD and the OED [24]. In the centered principal subspaces, the CD and the OED are not nearly uncorrelated or indepenbent. Therefore, applying the centered PCA to both the CD and the OED gives rise to low classification accuracy.

We should also note that the intra-class random pairing OED has a similar performance (a small performance loss, less than 0.5% in classification accuracy) comparing to the exact-matching OED. This result demonstrates that the intra-class random pairing method is an effective alternative to reduce the construction workload of the OED.

VII. CONCLUSION

To reduce the difficulty and workload in acquiring the extra image pairs for the SDBE_L2-based classification scheme, we have given some examples on the useful types of occlusion-free extra images and occluded extra images and introduced the intra-class random pairing method to semi-automatically form the extra image pairs. The extensive experiments on various synthetic occluded image datasets show that comparing to the original OED in [24], the enumerated examples and the intra-class random pairing method only result in a small loss in classification accuracy. We have also observed that the classification performance for a variant of the OED is highly related to the normalized mean correlation to the original OED.

In addition, in order to decrease the dictionary size, we have proved that SDBE_L2 is invariant to the uncentered PCA and proposed two novel uncentered PCA-based dictionary compression approaches, UPSPC and UPIPC. For UPSPC, the size of the dictionary can be reduced to not larger than twice the column vector length. For the OED (or CD) associated with a small number of occlusion classes (or image classes), the dictionary can be shrunk further by UPIPC. The proposed dictionary compression approaches facilitate the application of the SDBE_L2-based classification scheme to the large-scale datasets with numerous occlusion patterns. The experiments conducted on the large-scale synthetic occluded image dataset have demonstrated the effectiveness of the proposed dictionary compression approaches.

In this paper, although UPIPC can achieved better compression under the specific condition, the adoption of UPIPC is manually determined. An automatic adoption approach need to be investigated in the future. In addition, we only propose the improvements from the perspective of the dictionary generation. Promoting the classification accuracy on the occluded images from the perspective of the base-CNN training is deserved investigation in the future. Integrating the SDBE_L2-based approach into other computer vision tasks, such as object detection and semantic segmentation, to improve the performance against the occlusion is also our future research direction.

REFERENCES

[1] W. A. Perkins, "A model-based vision system for industrial parts," *IEEE Trans. Comput.*, vol. C-27, no. 2, pp. 126–143, Feb. 1978.

[2] B. Bhanu and O. D. Faugeras, "Shape matching of two-dimensional objects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 2, pp. 137–156, Mar. 1984.

[3] J. L. Turney, T. N. Mudge, and R. A. Volz, "Recognizing partially occluded parts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-7, no. 4, pp. 410–421, Jul. 1985.

[4] P. W. M. Tsang, P. C. Yuen, and F. K. Lam, "Classification of partially occluded objects using 3-point matching and distance transformation," *Pattern Recognit.*, vol. 27, no. 1, pp. 27–40, Jan. 1994.

[5] J. Zhang, X. Zhang, H. Krim, and G. G. Walter, "Object representation and recognition in shape spaces," *Pattern Recognit.*, vol. 36, no. 5, pp. 1143–1154, May 2003.

[6] F. Cen and G. Wang, "Dictionary representation of deep features for occlusion-robust face recognition," *IEEE Access*, vol. 7, pp. 26595–26605, 2019.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[9] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[10] H.-Y. Zhou, A.-A. Liu, W.-Z. Nie, and J. Nie, "Multi-view saliency guided deep neural network for 3-D object retrieval and classification," *IEEE Trans. Multimedia*, vol. 22, no. 6, pp. 1496–1506, Jun. 2020.

[11] C. Shi and C.-M. Pun, "Multiscale superpixel-based hyperspectral image classification using recurrent neural networks with stacked autoencoders," *IEEE Trans. Multimedia*, vol. 22, no. 2, pp. 487–501, Feb. 2020.

[12] M. N. Bajwa, S. Khurram, M. Munir, S. A. Siddiqui, M. I. Malik, A. Dengel, and S. Ahmed, "Confident classification using a hybrid between deterministic and probabilistic convolutional neural networks," *IEEE Access*, vol. 8, pp. 115476–115485, 2020.

[13] F. Johnson, A. Valderrama, C. Valle, B. Crawford, R. Soto, and R. Nanculef, "Automating configuration of convolutional neural network hyperparameters using genetic algorithm," *IEEE Access*, vol. 8, pp. 156139–156152, 2020.

[14] X.-F. Xu, L. Zhang, C.-D. Duan, and Y. Lu, "Research on inception module incorporated siamese convolutional neural networks to realize face recognition," *IEEE Access*, vol. 8, pp. 12168–12178, 2020.

[15] W. Ma, K. Li, and G. Wang, "Location-aware box reasoning for anchor-based single-shot object detection," *IEEE Access*, vol. 8, pp. 129300–129309, 2020.

[16] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2536–2544.

[17] R. A. Yeh, C. Chen, T. Y. Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do, "Semantic image inpainting with deep generative models," 2016, *arXiv:1607.07539*. [Online]. Available: http://arxiv.org/abs/1607.07539

[18] A. Nguyen, J. Clune, Y. Bengio, A. Dosovitskiy, and J. Yosinski, "Plug & play generative networks: Conditional iterative generation of images in latent space," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3510–3520.

[19] C. Yang, X. Lu, Z. Lin, E. Shechtman, O. Wang, and H. Li, "High-resolution image inpainting using multi-scale neural patch synthesis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, vol. 1, no. 2, p. 3.

[20] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative image inpainting with contextual attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5505–5514.

[21] J. Bao, D. Chen, F. Wen, H. Li, and G. Hua, "CVAE-GAN: Fine-grained image generation through asymmetric training," 2017, *arXiv:1703.10155*. [Online]. Available: http://arxiv.org/abs/1703.10155

[22] Y. Jiang, J. Xu, B. Yang, J. Xu, and J. Zhu, "Image inpainting based on generative adversarial networks," *IEEE Access*, vol. 8, pp. 22884–22892, 2020.

[23] L. Song, D. Gong, Z. Li, C. Liu, and W. Liu, "Occlusion robust face recognition based on mask learning with pairwise differential siamese network," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 773–782.

[24] F. Cen and G. Wang, "Boosting occluded image classification via subspace decomposition-based estimation of deep features," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 3409–3422, Jul. 2020.

[25] L. Cheng, J. Wang, Y. Gong, and Q. Hou, "Robust deep auto-encoder for occluded face recognition," in *Proc. 23rd ACM Int. Conf. Multimedia (MM)*, 2015, pp. 1099–1102.

[26] F. Zhao, J. Feng, J. Zhao, W. Yang, and S. Yan, "Robust LSTM-autoencoders for face de-occlusion in the wild," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 778–790, Feb. 2018.

[27] D. S. Trigueros, L. Meng, and M. Hartnett, "Enhancing convolutional neural networks for face recognition with occlusion maps and batch triplet loss," *Image Vis. Comput.*, vol. 79, pp. 99–108, Nov. 2018.

[28] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. ECCV*, 2014, pp. 818–833.

[29] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

[30] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories," *Comput. Vis. Image Understand.*, vol. 106, no. 1, pp. 59–70, Apr. 2007.

[31] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 3rd ed. Cambridge, MA, USA: MIT Press, 2009.

[32] *Computational Complexity of Mathematical Operations*. Accessed: Aug. 25, 2020. [Online]. Available: https://en.wikipedia.org/wiki/Computational_complexity_of_mathematical_operations

[33] M.-E. Nilsback and A. Zisserman, "Automated flower classification over a large number of classes," in *Proc. 6th Indian Conf. Comput. Vis., Graph. Image Process. (ICVGIP)*, Dec. 2008, pp. 722–729.

[34] L. Wu, E. Romero, and A. Stathopoulos, "PRIMME_SVDS: A high-performance preconditioned SVD solver for accurate large-scale computations," 2017, *arXiv:1607.01404*. [Online]. Available: https://arxiv.org/abs/1607.01404

[35] R. Larsen, "Lanczos bidiagonalization with partial reorthogonalization," in *Proc. DAIMI Rep. Ser.*, vol. 27, no. 537. Aarhus, Denmark: Aarhus Univ., Department of Computer Science, Dec. 1998. [Online]. Available: https://tidsskrift.dk/daimipb/article/view/7070, doi: 10.7146/dpb.v27i537.7070.

[36] J. Cadima and I. Jolliffe, "On relationships between uncentred and column-centred principal component analysis," *Pakistan J. Statist.*, vol. 25, no. 4, pp. 473–503, 2009.

[37] I. T. Jolliffe, *Principal Component Analysis* (Springer Series in Statistics), 2nd ed. New York, NY, USA: Springer-Verlag, 2002.

[38] C. M. Bishop, *Pattern Recognition and Machine Learning* (Information Science and Statistics). Berlin, Germany: Springer-Verlag, 2006.

[39] A. K. Cline and I. S. Dhillon, "Computation of the singular value decomposition," in *Handbook of Linear Algebra*, L. Hogben, R. Brualdi, A. Greenbaum, and R. Mathias, Eds. Boca Raton, FL, USA: CRC Press, 2006.

[40] A. Vedaldi and K. Lenc, "MatConvNet: Convolutional neural networks for MATLAB," in *Proc. 23rd ACM Int. Conf. Multimedia (MM)*, 2015, pp. 689–692.

[41] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[42] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "LIB-LINEAR: A library for large linear classification," *J. Mach. Learn. Res.*, vol. 9, pp. 1871–1874, Jun. 2008.

[43] M. James, *Pattern Recognition*. New York, NY, USA: Wiley, 1988.

FENG CEN received the Ph.D. degree in computer application technology from the Shanghai Jiaotong University, Shanghai, China, in 2003.

He is currently an Associate Professor with the Department of Control Science and Engineering, Tongji University, Shanghai. His research interests include computer vision, image processing, and pattern recognition.

XIAOYU ZHAO received the B.S. degree in control science and engineering from Tongji University, China, in 2018, where he is currently pursuing the M.E. degree with the Department of Control Science and Engineering.

His current research interests include computer vision, image processing, and deep learning.

WUZHUANG LI received the B.S. degree in mechatronics engineering from the China University of Mining and Technology, China. He is currently pursuing the M.E. degree with the Department of Control Science and Engineering, Tongji University, Shanghai, China.

His current research interests include natural language processing, image processing, and deep learning.

FANGLAI ZHU received the Ph.D. degree in control theory and control engineering from Shanghai Jiao Tong University, Shanghai, China, in 2001.

He worked with Purdue University, West Lafayette, IN, USA, as a Visiting Scholar, for six months in 2015, supported by the Chinese government. Before 2007, he was with the Guilin University of Electronic Technology (GLIET), Guilin, China, as an Associate Professor for three years and as a Professor for two years, respectively, where he was honored by the Guangxi local government as one of the 100 Young and Middle-aged Disciplinary Leaders in Guangxi Higher Education Institutions. Because of the excellent research work he has done in GLIET, he was accepted to join Tongji University, Shanghai, China, in July 2007, as a Professor. His research interests include nonlinear observer design, sliding-model control, model-based fault diagnosis, fault-tolerate control, and T–S fuzzy models.

Dr. Zhu was a recipient of the third prize of the Natural Science Award of Shanghai Science and Technology in 2011.

• • •