INDUCTIVE CONFORMAL PREDICTOR FOR SPARSE CODING CLASSIFIERS: APPLICATIONS TO IMAGE CLASSIFICATION

Sergio Matiz and Kenneth E. Barner

Department of Electrical and Computer Engineering, University of Delaware, Newark, DE, USA

ABSTRACT

Conformal prediction uses the degree of strangeness (nonconformity) of new data instances to determine the confidence values of new predictions. We propose an inductive conformal predictor for sparse coding classifiers, referred to as ICP-SCC. Our contribution is twofold: first, we present two nonconformity measures that produce reliable confidence values; second, we propose a batch mode active learning algorithm within the conformal prediction framework to improve classification performance by selecting training instances based on two criteria, informativeness and diversity. Experiments conducted on face and object recognition databases demonstrate that ICP-SCC improves the classification accuracy of state-of-the-art dictionary learning algorithms while producing reliable confidence values.

Index Terms— Conformal prediction, sparse coding, dictionary learning, image classification, active learning.

1. INTRODUCTION

Conformal prediction (CP) was proposed by Vovk, Shafer and Gammerman [1] based on the principles of algorithmic randomness and transductive inference. CP uses the degree of strangeness (nonconformity) of new data instances to determine the confidence values of new predictions.

The CP framework yields a set of predicted class labels with guaranteed error rate, a property referred to as *validity*. Moreover, unlike Bayesian methods [2], CP is only based on the assumption that the data are independent and identically distributed, *i.e.*, no knowledge on the prior is required.

Transductive conformal prediction for active learning has been reported in the literature [3, 4]. Active learning selects a set of instances from an unlabeled pool based on several criteria to improve classification performance. Ho *et al.* [3] proposed the query by transduction, which sequentially selects the most informative instances from an unlabeled pool. The disadvantage of transductive inference is computational inefficiency, which restricts its applicability.

Inductive conformal prediction emerged as an alternative to transductive inference [1, 5, 6, 7]. The application of inductive conformal predictors (ICP) to decision trees is studied in [6]. Balasubramanian *et al.* [7] use informativeness to perform active learning within the CP framework, improving the performance of support vector machines. Although the efforts mentioned above are shown to enhance performance, they only use informativeness as the selection criterion for active learning, which is only optimal for the selection of one instance at each iteration.

Sparse coding has recently gained interest in a variety of problems in image processing and computer vision, including face recognition, image classification, and image denoising [8, 9, 10, 11]. The goal of sparse coding is to approximate an input signal by a sparse linear combination of atoms (columns) of an overcomplete dictionary. Jiang *et al.* [12] incorporate the class label information and a label consistency term in the objective function to simultaneously learn a discriminative dictionary and a linear classifier. They refer to this method as label consistent K-SVD (LC-KSVD). Gu *et al.* [8] propose projective dictionary pair learning, which learns a synthesis dictionary, and an analysis dictionary.

Despite these advances, sparse coding algorithms require modifications [13, 14], or additional techniques to be implemented in conjunction with them [15, 16], to perform active learning, since confidence values and a measure of informativeness are required for those purposes. In addition, predictions accompanied by confidence values are desirable, since they provide information on the reliability of such predictions.

In light of the above, we propose an inductive conformal predictor for sparse coding classifiers, referred to as ICP-SCC. Our contribution is twofold: first, we present two nonconformity measures that produce reliable confidence values; second, we propose a batch mode active learning algorithm within the conformal prediction framework to improve classification performance by selecting training instances based on two criteria, informativeness and diversity.

This paper is organized as follows. First, an introduction to conformal prediction and dictionary learning (DL) is provided in Section 2. The ICP-SCC algorithm is described in Section 3. Furthermore, two nonconformity measures for sparse coding classifiers and the ICP-SCC query function for active learning are introduced. Experiments conducted on face and object recognition databases are presented in Section 4.

This material is based upon work supported by the National Science Foundation under Grant No. 1319598

2. BACKGROUND

2.1. Conformal prediction

CP uses the nonconformity of new data instances to determine the confidence values of new predictions. For an arbitrary significance level $\epsilon \in [0,1]$, CP yields a set Ψ^{ϵ} containing the correct class label of a given data instance with probability $(1-\epsilon)$, a property referred to as validity [17]. Define a bag of size $n \in \mathbb{R}$ as a collection of n elements, some of which may be identical with each other. Let that bag be denoted as $[\![z_1, \dots, z_n]\!]$. Define $z_i = (\mathbf{x}_i, h_i)$, where $\mathbf{x}_i$ represents a data instance and h_i its corresponding class label.

A nonconformity measure $A([\![z_1, \dots, z_n]\!], z)$ is a function producing a nonconformity score $\alpha \in \mathbb{R}$, representing how different z is from the elements in the bag $[\![z_1, \dots, z_n]\!]$. The nonconformity score of an element z_i in $[\![z_1, \dots, z_n]\!]$ is obtained as $\alpha_i = A([\![z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_n]\!], z_i)$.

In addition, we can measure the conformity of $\mathbf{x}_{n+j}$ to class q using *p-values*, which are defined as [1]:

$$p(\alpha_{n+j}^{(\mathcal{H}_q)}) = \frac{count\{i : \alpha_i > \alpha_{n+j}^{(\mathcal{H}_q)}\}}{n+1}, \tag{1}$$

where $\alpha_{n+j}^{(\mathcal{H}_q)}$ is the nonconformity score of $\mathbf{x}_{n+j}$, under the null hypothesis $\mathcal{H}_q$, and $p(\alpha_{n+j}^{(\mathcal{H}_q)})$ is its p-value. Notice that the p-value is highest when all previous nonconformity scores, $\alpha_1, \dots \alpha_n$, are higher than that of the new instance, $\alpha_{n+j}^{(\mathcal{H}_q)}$, meaning that $\mathbf{x}_{n+j}$ best conforms to class q. CP uses Equation (1) to predict the label for $\mathbf{x}_{n+j}$ using the highest p-value. In addition, for each new instance $\mathbf{x}_{n+j}$ and significance level $\epsilon \in [0,1]$, we form a set of labels $\Psi_{n+j}^{\epsilon} = \{i : p(\alpha_{n+1}^{(\mathcal{H}_i)}) > \epsilon\}$ containing the correct class label for $\mathbf{x}_{n+j}$ with probability $(1-\epsilon)$, according to the validity property.

The p-values are also used to quantify the informativeness [3, 4] . Ho and Wechsler [3] define the quality of information of instance $\mathbf{x}_{n+j}$ as

$$I(\mathbf{x}_{n+j}) = p_{n+j}^{(1)} - p_{n+j}^{(2)}, \tag{2}$$

where $p_{n+j}^{(1)}$ and $p_{n+j}^{(2)}$ are the largest and second largest p-values for instance $\mathbf{x}_{n+j}$, respectively.

2.2. Dictionary Learning

Two types of DL approaches are considered in this work: synthesis dictionary learning (SDL), and dictionary pair learning (DPL). The two aforementioned techniques are briefly introduced in this section. For the following definitions, let $\mathbf{Y} \in \mathbb{R}^{N \times n}$ be a matrix composed of n training vectors $\mathbf{y} \in \mathbb{R}^{N \times 1}$, $\mathbf{X} \in \mathbb{R}^{K \times n}$ be a matrix composed of vectors $\mathbf{x} \in \mathbb{R}^{K \times 1}$, which are the sparse representations of the training vectors in matrix $\mathbf{Y}$, and M be the number of classes. Let $\mathbf{D} \in \mathbb{R}^{N \times K}$ be the dictionary, constituted by K atoms $\mathbf{d} \in \mathbb{R}^{N \times 1}$ that are the columns of $\mathbf{D}$.

2.2.1. *Synthesis Dictionary Learning*

A reconstructive dictionary $\mathbf{D} \in \mathbb{R}^{N \times K}$ is learned by solving $< \mathbf{X}, \mathbf{D} >= \arg\min_{\mathbf{X},\mathbf{D}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2$. This optimization problem is solved by alternating between the updates of $\mathbf{D}$ and $\mathbf{X}$ [18]. LC-KSVD [12] and LC-RLSDLA [10] use an augmented version of matrix $\mathbf{Y}$, including the class label information, to simultaneously obtain a linear classifier $\mathbf{W} \in \mathbb{R}^{M \times K}$. Define $\mathbf{u} = [u_1, \dots, u_M]^T = \mathbf{W}\mathbf{x}$. The predicted label for $\mathbf{x}$ is obtained as $\hat{h} = \arg\max_j u_j$, for $j = 1, \dots, M$.

2.2.2. *Dictionary Pair Learning*

DPL [8, 11] learns M synthesis dictionaries $\mathbf{D}_j \in \mathbb{R}^{N \times K}$, and M analysis dictionaries $\mathbf{P}_j \in \mathbb{R}^{K \times N}$ $(j = 1, \dots, M)$. DPL solves $< \mathbf{P}, \mathbf{D} >= \arg\min_{\mathbf{P},\mathbf{D}} \sum_{j=1}^{M} \left\|\mathbf{Y}_j - \mathbf{D}_j\mathbf{P}_j\mathbf{Y}_j\right\|_F^2 + \left\|\mathbf{P}_j\bar{\mathbf{Y}}_j\right\|_2^2$, where $\mathbf{Y}_j$ is a matrix containing the training vectors of class j, and $\bar{\mathbf{Y}}_j$ is a the complementary data matrix of $\mathbf{Y}_j$. Define $v_j = \left\|\mathbf{x} - \mathbf{D}_j\mathbf{P}_j\mathbf{x}\right\|_2$. The predicted label for $\mathbf{x}$ is obtained as $\hat{h} = \arg\min_j v_j$, for $j = 1, \dots, M$.

3. INDUCTIVE CONFORMAL PREDICTOR FOR SPARSE CODING CLASSIFIERS

We propose an inductive conformal predictor for sparse coding classifiers, referred to as ICP-SCC. Furthermore, we present an active learning algorithm within the CP framework, in which instances are selected from an unlabeled pool based on two criteria, informativeness and diversity. In the remainder of this section the proposed nonconformity measures and query function are introduced, and the ICP-SCC algorithm is described.

3.1. ICP-SCC Nonconformity Measures

We propose two nonconformity measures, the first one is designed for SDL, and the second one for DPL. A description of the nonconformity measures is provided below.

3.1.1. *Nonconformity measure for SDL*

Let $\mathbf{W} \in \mathbb{R}^{M \times K}$ be a linear classifier, for M distinct class labels, constituted by row vectors $\mathbf{w}_q \in \mathbb{R}^K$, $q \in \{1, 2, \dots, M\}$. Define $\hat{\mathbf{w}}_q = \mathbf{w}_q / \left\|\mathbf{w}_q\right\|$. The proposed nonconformity measure for SDL under the null hypothesis $\mathcal{H}_q$ is given by

$$A_{SDL}^{(\mathcal{H}_q)} := -\hat{\mathbf{w}}_q\mathbf{x} + \frac{1}{M-1}\sum_{i \neq q}\hat{\mathbf{w}}_i\mathbf{x}, \tag{3}$$

3.1.2. *Nonconformity measure for DPL*

Let $\mathbf{D}_j \in \mathbb{R}^{N \times K}$, and $\mathbf{P}_j \in \mathbb{R}^{K \times N}$ be the synthesis and analysis dictionaries for class j $(j = 1, \dots, M)$, respectively.

Algorithm 1 ICP-SCC

1: **Input:** Proper training set $T_{prop} = \{z_1, ..., z_\ell\}$, calibration set $T_{cal} = \{z_{\ell+1}, ..., z_{\ell+r}\}$, unlabeled pool $U = \{\mathbf{x}_{n+1}, \ldots, \mathbf{x}_{n+v}\}$, classification rule C_{prop}, number of desired instances N_{AL}, and number of class labels M
2: Use Equation (3) or (4) and the classification rule C_{prop} to calculate:
- The nonconformity scores $\{\alpha_{\ell+1}, \ldots, \alpha_{\ell+r}\}$ corresponding to the instances in the calibration set.
- The nonconformity scores $\{\alpha^{\mathcal{H}_i}_{n+1}, \ldots, \alpha^{\mathcal{H}_i}_{n+v}\}$ corresponding to the instances in the unlabeled pool, where $i = \{1, \ldots, M\}$

3: Use Equation (1) to calculate the p-values associated with the instances in U, and obtain their informativeness $I(\mathbf{x}_{n+j})$ through equation (2), where $j \in \{1, \ldots, v\}$
4: Apply equation (5) to select the N_{AL} most informative and diverse instances. Then group such instances and their corresponding class labels as $T_d = \{z^d_1, \ldots, z^d_{N_{AL}}\}$
5: Construct $T_{AL} = T_{prop} \cup T_d$
6: **Output:** T_{AL}

The proposed nonconformity measure for DPL under the null hypothesis $\mathcal{H}_q$ is given by

$$A^{(\mathcal{H}_q)}_{DPL} := \|\mathbf{x} - \mathbf{D}_q\mathbf{P}_q\mathbf{x}\|_2 - \frac{1}{M-1}\sum_{i \neq q}\|\mathbf{x} - \mathbf{D}_i\mathbf{P}_i\mathbf{x}\|_2 \,, \quad (4)$$

Assuming that the classifiers are accurate and the null hypothesis $\mathcal{H}_q$ is true, the values of $A^{(\mathcal{H}_q)}_{SDL}$ and $A^{(\mathcal{H}_q)}_{DPL}$ will decrease, since the term $\hat{\mathbf{w}}_q\mathbf{x}$ in (3) increases, and the term $\left\|\mathbf{x} - \mathbf{D}_q\mathbf{P}_q\mathbf{x}\right\|_2$ in (4) decreases, indicating that $\mathbf{x}$ conforms to class q. Conversely, if the null hypothesis $\mathcal{H}_q$ is false, the value of $A^{(\mathcal{H}_q)}_{SDL}$, and $A^{(\mathcal{H}_q)}_{LM}$ will tend increase, indicating that $\mathbf{x}$ does not conform to that particular class.

3.2. ICP-SCC Query Function

Different from previous work on ICP [5, 7], the proposed approach considers both informativeness and diversity as the selection criteria for active learning. The proposed query function is given by

$$\mathbf{x}_t = \underset{\mathbf{x}_i \in U/T_d}{\arg\min} \left\{ \rho|I(\mathbf{x}_i)| + (1-\rho) \max_{\mathbf{x}_j \in T_d} \left[\frac{|\mathbf{x}_i \cdot \mathbf{x}_j|}{\|\mathbf{x}_i\|\|\mathbf{x}_j\|} \right] \right\}, \quad (5)$$

where U, T_d and, U/T_d are the unlabeled pool U, the set of instances selected for training, and the set of instances in U that are not contained in T_d, respectively. Diversity is measured by the second term using cosine angle distance [19]. The parameter ρ provides the trade-off between informativeness and diversity. The first instance of T_d is selected as the most informative instance in U. The algorithm stops when the number of selected instances in T_d is equal to the desired number N_{AL}.

3.3. ICP-SCC for Active Learning

ICP-SCC selects the most informative and diverse instances from an unlabeled pool using the proposed query function described in (5). Informativeness is computed through equation (2). The selected instances, along with their corresponding class labels, are used in a subsequent training stage to improve performance, instead of relying on instances that are selected at random.

Define $T_{train} = \{z_1, ..., z_n\}$ as the training set and $U = \{\mathbf{x}_{n+1}, \ldots, \mathbf{x}_{n+v}\}$ as the unlabeled pool. Following the steps described for ICP in [1, 7], we split T_{train} into $T_{prop} = \{z_1, ..., z_\ell\}$, the proper training set, and $T_{cal} = \{z_{\ell+1}, ..., z_{\ell+r}\}$, the calibration set, where the size of the training set satisfies $n = \ell + r$. C_{prop} is the classifier trained on the proper training set T_{prop}, which is used to compute informativeness. Let N_{AL} and M be the number of desired instances from U and the number of class labels, respectively. Let $T_{AL} = T_{prop} \cup T_d$, where $T_d = \{z^d_1, \ldots, z^d_{N_{AL}}\}$ is the set of pairs containing the N_{AL} most informative and diverse instances in U and their corresponding class labels. The proposed active learning approach is summarized in Algorithm 1.

4. EXPERIMENTAL RESULTS

The focus of ICP-SCC is twofold: 1) to improve the performance of sparse coding classifiers through active learning; and 2) to produce reliable confidence values. Therefore, ICP-SCC is to evaluated based on the improvement achieved in classification performance and the quality of the produced confidence values.

4.1. Experimental Setup

The performance of ICP-SCC is evaluated using two different sparse coding algorithms: LC-KSVD [12], and DPL [8]. The baseline for our experiments is random sampling. The databases used in our experiments are described below.

The *Extended YaleB database* [20] consists of 2,414 frontal-face images of 38 people (38 different classes), taken under varying lighting conditions. There are about 64 images for each person. The feature descriptors used are randomfaces [21] of size $N = 504$.

The *AR database* [22] contains over 4,000 frontal-face images of 100 people (100 different classes), each of size 165×120. Images include facial variations and also disguises, such as sunglasses and scarves. The feature descriptors used are randomfaces of size $N = 540$.

The *Caltech101 database* [23] contains 9,144 images from 101 object classes, and a background class, including animals, vehicles, etc. The number of images in each category varies from 31 to 800. For Caltech101, SIFT descriptors are first extracted. Next, spatial pyramid features are obtained

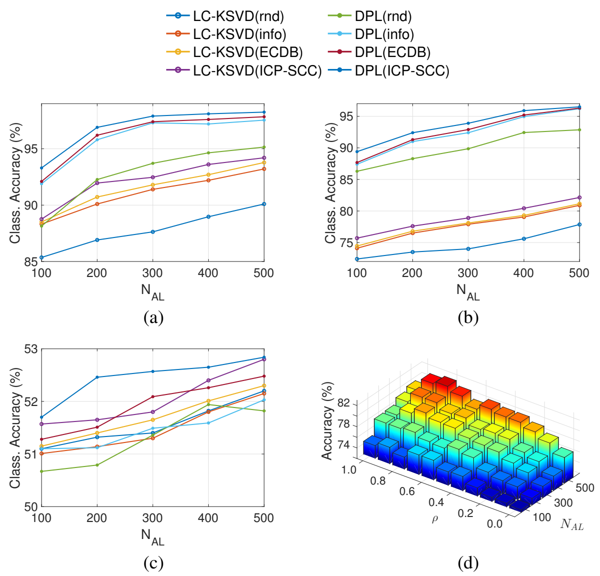

Fig. 1. Classification accuracy for DPL and LC-KSVD as a function of N_{AL}, (a) YaleB ($K = 380$), (b) AR ($K = 400$), (c) Caltech101 ($K = 510$), (d) effect of ρ on the performance of LC-KSVD (AR)

from the SIFT descriptors. Then, the dimensionality of the resulting features is reduced to 3000 through PCA [12].

For each of the experiments, 5 trials are conducted. In each trial, the order of the training instances is permuted. The average classification accuracy is presented. The number of images per class in the proper training set for the Extended YaleB, AR, and Caltech101 databases is 8, 5, and 5, respectively. The calibration set consists of 199 instances, which results in a resolution of 0.5% in the confidence values, according to (1). Optimization is performed over the parameter ρ through exhaustive search, and the best results are presented.

4.2. Results: ICP-SCC for Active Learning

The performance improvement obtained through ICP-SCC is compared with that of: random sampling, active learning based on informativeness [3, 5, 24], and MCLU-ECDB [25], which are denoted as (rnd), AL(info), and AL(ECDB) respectively. The performance of LC-KSVD and DPL as a function of the number of selected instances N_{AL}, for the different databases and query functions, is shown in Fig. 1. It is observed that the performance of both algorithms is improved when ICP-SCC is used, for all the considered databases.

Table 1 shows that for the AR database (LC-KSVD, $N_{AL} = 300$) the performance of (rnd), AL(info), and AL(ECDB) is 74.0%, 77.9%, and 78.1%, respectively, whereas that of ICP-SCC is 79.9%. Similar results are observed for the Extended YaleB and Caltech101 databases, with ICP-SCC achieving the best performance among the considered active learning approaches.

Table 1. Classification accuracy for different query functions as a function of the number of selected instances N_{AL}

Algorithm	Query func.	YaleB N_{AL} 200	YaleB N_{AL} 300	AR N_{AL} 200	AR N_{AL} 300	Cal101 N_{AL} 200	Cal101 N_{AL} 300
LC-KSVD	(rnd)	86.9	87.6	73.5	74.0	51.3	51.4
	AL(info)	90.1	91.4	76.5	77.9	51.1	51.3
	AL(ECBD)	90.7	91.8	76.8	78.1	51.4	51.6
	ICP-SCC	**91.9**	**92.5**	**77.6**	**79.9**	**51.7**	**51.8**
DPL	(rnd)	92.3	93.7	88.3	89.9	50.8	51.4
	AL(info)	95.8	97.3	91.0	92.4	51.1	51.5
	AL(ECBD)	96.2	97.4	91.3	92.9	51.5	52.1
	ICP-SCC	**96.9**	**97.9**	**92.4**	**93.9**	**52.5**	**52.6**

Table 2. Experimental results of the validity property

Algorithm	Confidence (%)	Error (%) YaleB	Error (%) AR	Error (%) Cal101
LC-KSVD	95	4.8	4.7	3.9
	90	10.0	12.2	9.9
	85	15.5	15.9	14.3
DPL	95	5.1	5.7	4.2
	90	10.4	11.7	9.2
	85	15.8	16.2	15.6

The effect of the parameter ρ on the performance of LC-KSVD (AR database) is shown in Fig. 1(d). Notice that ρ has to be optimized for the each value of N_{AL}. Similar results are obtained for the YaleB and Caltech101 databases.

4.3. Results: ICP-SCC Confidence Values

The quality of the ICP-SCC confidence values is demonstrated through the evaluation of validity property [5]. We define the percentage of errors as the number of times the correct label for instances $\mathbf{x}_{n+j}$ is not in Ψ^{ϵ}_{n+j}, for a given ϵ, divided by the total number of test instances. The experimental results in Table 2 show that for a given confidence level, $(1 - \epsilon)$, the percentage of errors closely approximates the significance level, ϵ, which agrees with the validity property. This demonstrates the quality of the p-values calculated through the proposed nonconformity measures and thus the usefulness of its confidence measures.

5. CONCLUSIONS

An inductive conformal predictor for sparse coding classifiers, referred to as ICP-SCC is proposed. Two nonconformity measures, one for synthesis dictionary learning, and the other one for dictionary pair learning are proposed. Furthermore, an active learning algorithm within the conformal prediction framework is presented to improve classification performance. Experiments conducted on face and object recognition databases demonstrate that ICP-SCC improves the classification accuracy of state-of-the-art dictionary learning algorithms, while producing reliable confidence values.

6. REFERENCES

[1] V. Vovk, A. Gammerman, and G. Shafer., *Algorithmic Learning in a Random World*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

[2] T. Melluish, S.Craig, I. Nouretdinov, and V. Vovk, "Comparing the bayes and typicalness frameworks," in *Proc. European Conference on Machine Learning (ECML 2001)*. 2009, vol. 2167, pp. 360–371, Springer.

[3] Shen-Shyang Ho and Harry Wechsler, "Query by transduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 9, pp. 1557–1571, Sep. 2008.

[4] V. Balasubramanian, S. Chakraborty, and S. Panchanathan, "Generalized query by transduction for online active learning," *Proc. Int. Conf. Computer Vision Workshops (ICCV Workshops)*, pp. 1378–1385, 2009.

[5] H. Papadopoulos, V. Vovk, and A.Gammerman, "Conformal prediction with neural networks," *Proc. IEEE Int. Conf. Tools with Artificial Intelligence (ICTAI)*, vol. 2, pp. 388–395, Oct. 2007.

[6] H. Papadopoulos, H. Boström, and T. Löfström, "Conformal prediction using decision trees," *Proc. IEEE Int. Conf. Data Mining (ICDM)*, pp. 330–339, Dec. 2013.

[7] Vineeth Balasubramanian, Shen-Shyang Ho, and Vladimir Vovk, *Conformal prediction for reliable machine learning: theory, adaptations and applications*, Morgan Kaufmann, 2014.

[8] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng, "Projective dictionary pair learning for pattern classification," in *Advances in neural information processing systems*, 2014, pp. 793–801.

[9] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Trans. Image Process.*, vol. 17, no. 1, pp. 53–69, Jan 2008.

[10] S. Matiz and K. E. Barner, "Label consistent recursive least squares dictionary learning for image classification," *Proc. IEEE Int. Conf. Image Processing (ICIP)*, pp. 1888–1892, 2016.

[11] Meng Yang, Heyou Chang, and Weixin Luo, "Discriminative analysis-synthesis dictionary learning for image classification," *Neurocomputing*, vol. 219, pp. 404–411, Jan 2017.

[12] Z. Jiang, Z. Lin, and L. S. Davis, "Label consistent K-SVD: Learning a discriminative dictionary for recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2651–2664, Nov 2013.

[13] J. J.Y. Wang and X. Gao, "Semi-supervised sparse coding," in *Proc. International Joint Conference on Neural Networks (IJCNN)*, 2014, pp. 1630–1637.

[14] X. Wang, X.Guo, and S. Z. Li, "Adaptively unified semi-supervised dictionary learning with active points," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2015, pp. 1787–1795.

[15] T. Wu, A.D. Sarwate, and W.U. Bajwa, "Active dictionary learning for image representation," in *Proc. SPIE Unmanned Systems Technology XVII*, 2015, vol. 9468, pp. 946809(1) –946809(10).

[16] M. Li and I. K. Sethi, "Confidence-based active learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 8, pp. 1251–1261, Aug. 2006.

[17] G. Shafer and V. Vock, "A tutorial on conformal prediction," *J. Mach. Learn. Res. 9*, pp. 371–421, 2008.

[18] H. Lee, A. Battle, R. Raina, and A.Y. Ng, "Efficient sparse coding algorithms," in *Advances in neural information processing systems*, 2007, pp. 801–808.

[19] K. Brinker, "Incorporating diversity in active learning with support vector machines," in *Proc. Int. Conf. Machine Learning (ICML)*, 2003, pp. 59–66.

[20] 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. Intelligence*, vol. 23, no. 6, pp. 643–660, Jun 2001.

[21] W. John, 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.

[22] A.M. Martinez and R. Benavente, "The AR face database," *CVC Tech. Report # 24*, 1998.

[23] Li Fei-Fei, Rob Fergus, and Pietro Perona, "Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories," *Computer vision and Image understanding*, vol. 106, no. 1, pp. 59–70, 2007.

[24] U. Johansson T. Löfström and H. Boström, "Effective utilization of data in inductive conformal prediction using ensembles of neural networks," *Proc. The International Joint Conference on Neural Networks (IJCNN)*, vol. 2, pp. 1–8, Aug. 2013.

[25] B. Demir, C. Persello, and L. Bruzzone, "Batch-mode active-learning methods for the interactive classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens*, vol. 49, no. 3, pp. 1014–1031, Oct. 2010.