

Received December 16, 2019, accepted December 29, 2019, date of publication January 20, 2020, date of current version April 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2968082

A Novel Active Learning Algorithm for Robust Image Classification

XINGLIANG XIONG¹, MINGYU FAN², CHUANG YU³, AND ZHENJIE HONG⁴

¹Key Laboratory of Child Development and Learning Science, Ministry of Education, School of Biological Science and Medical Engineering, Southeast University, Nanjing 210096, China

²College of Mathematics, Physics and Electronic Information Engineering, Wenzhou University, Wenzhou 325035, China

³School of Mechanical, China University of Mining and Technology, Xuzhou 221000, China

⁴College of Computer Science and Artificial Intelligence, Wenzhou University, Wenzhou 325035, China

Corresponding author: Zhenjie Hong (hong@wzu.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant U1831116.

ABSTRACT Training samples need to be labeled before being used to train classification model, which usually takes too much labor and material resources. Recently, this problem has attracted widespread attention. In order to reduce the workload of labeling samples, we propose a novel active learning methodology, which uses locally linear reconstruction coefficients to construct semi-supervised data manifold adaptive kernel space. Comparing the new method with other sampling approaches on several real-world image datasets, experimental results indicate that the novel algorithm has preferable classification ability. Especially, it can show higher classification accuracy under the condition that only a few samples are selected to train the classifier model.

INDEX TERMS Active learning, experimental design, image classification, local linear reconstruction, manifold learning.

I. INTRODUCTION

Improving the accuracy of image classification is one of the most important and interesting problems in machine learning field [1]–[3]. Many researchers carried out various image studies in recent years, such as human face image classification [4]–[6], handwritten digital image recognition [7]–[9], remote sensing image classification [10], [11] and so on. Traditional supervised data classification models almost all based on statistical methodology [12], [13], which first need a user to label numerous samples and then apply these labeled sample points to train the classification model. However, in a host of real-world applications of the image classification, labeling unclassified data samples are costly and time-consuming, especially for some special image recognition problems that need prior professional knowledge. For instance, people who label the samples that come from functional magnetic resonance imaging need to master some biomedicine knowledges. It is hard to obtain enough labeled samples that are used for training the classifier model, while in the case of training samples shortage possibly make the classification ability of the classifier decrease severely.

The associate editor coordinating the review of this manuscript and approving it for publication was Inês Domingues¹.

Consequently, how to collect the most representative sample subset from the whole dataset, cut down the cost of labeling samples, and train an excellent classifier has become one of the most critical problems in data classification [14], [15]. In order to solve these issues efficiently, a lot of people try to apply active learning in different kinds of classification tasks [16], [17]. Many studies show that active learning algorithm can greatly reduce the workload of labeling samples and successfully obtain higher classification accuracy. Thereby, it has attracted wide attention in pattern recognition, data mining, etc. [15]–[17]. With active learning, a learner can actively select the unlabeled samples that contain the most information and transmit them to an expert to label, and then add these labeled samples to the training set to train the classifier. This is very different from the previous passive learning that passively labels and processes for all data samples.

The problem in active learning is to determine which sample contains the most information. According to the difference of selecting strategies, active learning algorithms mainly include three kinds of methods that based on uncertainty [18], committee [19] and generalized error reduction [20], [21]. Many active learning techniques, such as Optimal Experimental Design (OED) [22] and Transductive Experimental Design (TED) [23], [24], have been proposed for image

and text classification. Recently, the algorithm which bases on the OED is very popular in machine learning due to its good efficiency and the absence of priori knowledge of the samples. OED is related to the experimental design of which goal is to minimize measurement error. Three most famous methodologies of OED include A-Option Design (AOD), D-Option Design (DOD) and E-Option Design (EOD), which optimize for trace, determinant, and eigenvalue of a matrix respectively [22]. In the OED algorithm, each sample point is viewed as an experiment, and its label is considered as a measure. Selecting the best samples is thought to design an experiment that allows one of the parameters in the measurement model to have the least variance. However, this method only let the optimization target focus on the variance of the correlation parameters. It does not optimize the classifier's prediction error with the selected data directly. Considering that there are some shortcomings in OED, related researchers proposed TED algorithm on the basis of the OED. Although TED basically comes from OED, it estimates the expected prediction error on all examples, not only labeled but also unlabeled. It is important to note that both OED and TED algorithms are strictly correlated to statistics. Choosing the samples that can minimize the expected prediction error is equivalent to select the data points which can linear reconstruct the rest samples optimally. However, TED only pays close attention to the global and discriminant structure of data space. The intrinsic geometric structure that is propitious to increase the classification accuracy of the classifier is usually discarded by TED [25], [26].

For studying the potential geometric structure properties of the samples space, researchers have designed many different manifold learning methodologies, e.g., ISOMAP [27], Locality Preserving Projections (LPP) [28] and so on. Although these methods have some incoherence differences between each other, locally invariant conception is the most significant common idea among them [29]. With regard to locally invariant conception, all the neighbors have common properties, i.e., the samples that are close together have the same class attributes. Based on the TED and given consideration to the locally invariant conception of manifold learning, D. Cai et al. present a new method that named Manifold Adaptive Experimental Design (MAED) by using graph Laplacian [30], [31]. It solves active learning problems by implementing the data classification in manifold kernel space [32]. Classification accuracy of the MAED algorithm closely relates to the selection strategy of the constructed graph. Hence, how to obtain an efficient and effective neighbor graph is the most important problem in the MAED approach. There are many options for constructing the manifold structure graph [31]. A new manifold adaptive active learning methodology which called Semi-Supervised Adaptive Experimental Design (SSAED) for image active learning under the circumstance of semi-supervised learning [33]–[36] is proposed in this study. By calculating the locally linear reconstruction coefficients [37] of sample points with the semi-supervised strategy and incorporating graph structure into the manifold

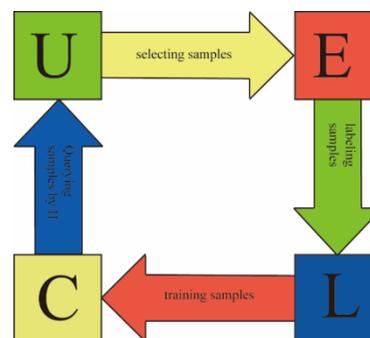


FIGURE 1. An example of the active learning process. U denotes the unlabeled sample database that has never been tagged and E the expert system that can correctly label the samples in U. L denotes the labeled sample dataset which is used to train a classifier, C is a certain kind of classifier, and H is a query function that is used to query information in the pool of U.

kernel, the new manifold structure can be used in the learning process. Then, the OED can be efficiently completed in the manifold adaptive kernel space. It is noteworthy that many active learning algorithms don't pay attention to the class information when building the intrinsic manifold structure. Interestingly, the novel algorithm that proposed by us can effectively solve this problem.

We make elaborative arrangements for the remainder of our article with this type: details of method and materials that are adopted in this study are described in section two. In the third section, we present experimental results and discussion. Finally, we summarize this study and give out the conclusion.

II. METHOD AND MATERIALS

In the process of human learning, people often first use the existing experience to learn new knowledge, then rely on the acquired knowledge to summarize and accumulate the experience. Hence, these experience and knowledge are constant interactive. Similarly, machine learning simulates the process of human learning by using the knowledge of existing to train model and obtain new knowledge firstly, then correct the model with the accumulation of information. By this way, a new model that is more accurate and useful can be obtained. Active learning selectively acquires a small number of samples that contain the largest amount of information, which is different from previous passive learning that passively labels and processes all data samples. In active learning, the learner actively selects the most representative unlabeled samples and transmits these samples to an expert to label, then adds these labeled samples to training dataset to train the classifier [15], [16].

Fig.1 shows the brief procedure of active learning. Learners use a small amount of initial labeled samples L to learn, the query function H first chooses one or a group of the most useful samples from the unlabeled samples pool U, then asks these samples' labels from the supervisor E, finally uses the new knowledge to train the classifier C and implements the next round of inquiry. Active learning is a cyclic process that does not stop until a certain criterion is reached.

The fundamental purpose of active learning is to search the most representative samples. In Euclidean space, given a sample dataset $S = \{s_1, s_2, \dots, s_i, \dots, s_n\}$, where s_i denotes a sample that is represented by a vector of some dimension, $i \in [1, n]$, and n denotes the total number of samples. Searching the most representative samples is to search a subset $Q = \{q_1, q_2, \dots, q_j, \dots, q_m\} \subset S$ that contains the maximally informative data samples, where $m < n, j \in [1, m]$. That is to say, it can obtain the highest classification accuracy when these subsamples are labeled to treat as training data. Seeking the maximally informative samples to label is routinely described as experimental design in statistics [22]. In the sample dataset S , each sample point is usually considered to be an experiment and its corresponding class label is considered to as a measurement.

A. THE METHOD OF OED

The research task of OED is strictly related to the experimental design that is expected to minimize variance of a parameterized model [22]. Take into account $f(s) = w^T s$ from $c = w^T s + \Delta$, where $f(s)$ denotes the linear function, c the measurement, and Δ the measurement noise that follows the zero mean normal distribution of which variance is equal to σ^2 . Assuming that there exist an array of labeled samples $\{(q_1, c_1), \dots, (q_j, c_j), \dots, (q_m, c_m)\}$, where c_j is the class label of the j -th sample q_j in the subset Q . Therefore, we have:

$$\hat{w} = \arg \min_w \{L(w) = \sum_{j=1}^m (w^T q_j - c_j)^2\} \quad (1)$$

where w denotes the weight vector and \hat{w} is its maximum likelihood estimate. Naturally, we can get the estimate error as follows:

$$w' = \hat{w} - w. \quad (2)$$

Reference [31] points out that the above error has zero mean and a covariance matrix produced from $\sigma^2 H_w$. Where H_w denotes the inverse of the Hessian operator of the objective function $L(w)$ in (1), and it can be written by:

$$H_w = \left(\frac{\partial^2 L}{\partial w^2}\right)^{-1} = \left(\sum_{j=1}^m q_j q_j^T\right)^{-1} = (QQ^T)^{-1} \quad (3)$$

where $Q = \{q_1, q_2, \dots, q_j, \dots, q_m\}$. Thereby, OED aims to select the subsamples by minimizing the error that is produced from the matrix H_w . The most famous optimal designs include the AOD, DOD and EOD, which optimize for trace, determinant and eigenvalue of the matrix H_w respectively [22].

B. THE METHODOLOGY OF TED

The TED method that is essentially based on the OED, evaluates the expected prediction error on both labeled and unlabeled examples. Giving out the simple dataset S , minimizing the average expected square predictive error of the S can be

equivalently viewed as a simplified optimization target as follows:

$$\min_{Q \subset S} Tr(S^T (QQ^T)^{-1} S). \quad (4)$$

The most important problem in (4) is how to search the subset Q , however, it is difficult to obtain a global solution. This is because that the issue is the NP-hard [23]. After a few transformations, it is able to convert the task (4) into the following form:

$$\min_{a_i \in IR^m, Q=(q_1, \dots, q_m)} \sum_{i=1}^n \|s_i - Q^T a_i\|^2 + \xi \|a_i\|^2 \quad (5)$$

where $a_i = (a_{i,1}, a_{i,2}, \dots, a_{i,m})$ is a vector of auxiliary reconstructing coefficient that uses the linear combination of the subset $Q = \{q_1, q_2, \dots, q_j, \dots, q_m\}$ to fit the data point s_i , i.e., the whole dataset S can be reconstructed by the subset Q . The ξ is a regularization parameter that is used to control the amount of shrinkage [38]–[40]. Whereas, the above formula is still a disadvantageous approach. This is due to the solution in (5) is suboptimal.

In order to thoroughly solve the above optimization problem, a series of auxiliary variables $b = (b_1, b_2, \dots, b_i, \dots, b_n)$ are introduced to control the noise of training samples, i.e., b_i and s_i are one-versus-one, which help us to determine the sample s_i of the whole dataset S is selected or not selected as a training sample by the b_i . Do some transformations, it can convert the target (5) into another form:

$$\min_{b, a_i \in IR^n} \sum_{i=1}^n (\|s_i - S^T a_i\|^2 + \sum_{j=1}^n \frac{a_{i,j}^2}{b_j}) + \varphi \|b\|_1 \quad (6)$$

s.t. $b_j \geq 0, \quad j = 1, \dots, n$

where φ is a regularization parameter. Minimizing $\|b\|_1$ can obtain a sparse vector of which some elements are zero. Supposing that $b_j = 0$, we would have $a_{i,j} = 0$, or else the (6) would become infinity, that is to say, the whole optimization would not get a satisfying solution. Thereby, the j -th sample s_j that corresponds to the b_j would not be selected, then we can select the most informative samples according to the value of b_j . Reference [24] reveals that the (6) is convex optimization, hence, it can make sure that the calculations of the global optimization solution come out.

C. THE ALGORITHM OF MAED

Considering that the TED approach has some shortcomings and the potential geometric attributes of sample data space are exceedingly useful for increasing classification accuracy, it is natural to think that put the manifold structure into the process of active learning. As for how to carry out this idea concretely, an effective way is to implement the active learning tasks in the manifold adaptive kernel space. We demonstrate the details as follows:

Giving out the sample dataset S and defining a mathematical formula $r_s = (R(s, s_1), \dots, R(s, s_n))^T$, then the reproducing kernel of Hilbert space can use the following equation to

represent itself [31], [32]:

$$\hat{R}(s, q) = R(s, q) - \rho r_s^T (E + FR)^{-1} FR_q \quad (7)$$

where E denotes an identity matrix, R is a kernel matrix that comes from Hilbert space, and $\rho \geq 0$ represents a constant that is used to regulate the smoothness of $\hat{R}(s, q)$. With regard to the matrix R , it can use the classic kernels, such as the linear kernel, Gaussian kernel, to obtain it. As for the construction of the reproducing kernel, the most important thing is the selection of the matrix F . Based on different kinds of graph construction methods, we can obtain various matrix F .

By choosing the matrix F that bases on graph Laplacian [30], [41], MAED method presents the data-dependent kernel and performs convex TED problem in the manifold adaptive kernel space. For the sample dataset S , let $R(s_i, s_j) = \langle \psi(s_i), \psi(s_j) \rangle$, ψ denote a feature map that from raw input data space to the reproducing kernel Hilbert space [32], [42]. Then the convex TED of (6) is rewritten by another form:

$$\min_{b, a_i \in \mathbb{R}^n} \sum_{i=1}^n (\|\Psi(s_i) - \Psi(S)a_i\|^2 + \sum_{j=1}^n \frac{a_{i,j}^2}{b_j}) + \varphi \|b\|_1 \quad (8)$$

$$s.t. \quad b_j \geq 0, \quad j = 1, \dots, n$$

where $\psi(S)$ denotes the data matrix of the reproducing kernel Hilbert space, i.e., $\psi(S) = (\psi(s_1), \dots, \psi(s_n))$. With (8), we consider a $n \times n$ kernel matrix \hat{R} like this $\hat{R}_{i,j} = \hat{R}(s_i, s_j)$, and let v_i be its i -th column (or row, because R is symmetric) vector. Then, after some mathematical deductions, we can get the following equation:

$$a_i = (\text{diag}(b)^{-1} + \hat{R})^{-1} v_i \quad (9)$$

where $\text{diag}(b)$ denotes a diagonal matrix whose elements on the main diagonal are b_1, b_2, \dots, b_n . Once a_i is gotten, fix a_i and search the minimal value for b_j . Do some similar mathematical derivations again, we can obtain the following equation:

$$b_j = \sqrt{\frac{\sum_{i=1}^n a_{i,j}^2}{\varphi}}. \quad (10)$$

Therefore, by iteratively compute until convergence, $a_{i,j}$ and b_j can be successfully calculated. Hence, it can choose the most informative samples according to $b_j (j = 1, 2, \dots, n)$ that are arranged in descending order.

D. THE NOVEL ALGORITHM

Obviously, the performance of the MAED approach mainly depends on the choice of the matrix F . Thereupon, it is possible to design a robust active learning method, which uses a new graph construction style to calculate the manifold adaptive matrix. By calculating the locally linear reconstruction coefficients of sample data points with the semi-supervised manner and incorporating the manifold structure into the active learning process, we propose a novel active learning

algorithm that is called semi-supervised adaptive experimental design (SSAED) for image classification. In this subsection, we first give out the graph construction method used in this paper and then present the new active learning algorithm that bases on this graph adaptive kernel space. Some details of graph embedding can be found in reference [43].

1) GRAPH CONSTRUCTION

- (1) **Construct a nearest neighbor graph with the semi-supervised manner.** Given a set of sample data points $S = \{s_1, s_2, \dots, s_n\}$ in \mathbb{R}^m , for any data point s_i with known class, firstly, find the samples which have the same class information, and then use Euclidean distance to compute its k nearest data samples from the same class dataset. Otherwise, for a sample without class information, we directly use Euclidean distance to compute the k nearest data points from the whole dataset.
- (2) **Compute reconstruction weight matrix and reconstruction graph.** The weight matrix W is computed by solving the following optimization problem:

$$\varepsilon(W) = \min \sum_i \left\| s_i - \sum_{j \in N_i} W_{ij} s_j \right\|^2 \quad (11)$$

where W_{ij} denotes the contribution of the j -th sample data point to the i -th sample data point, N_i denotes the nearest neighbors of the sample point s_i . In order to make the problem (11) well-posed, it should minimize the above cost function under the condition that (11) meets two constraints:

- (i) s_i is reconstructed only from its neighbors, in other words, $W_{ij} = 0, \forall j \notin N_i$;
- (ii) the elements of each row of the coefficient matrix add up to one, i.e., $\sum_{j=1}^N W_{ij} = 1, \forall i$.

After obtaining the weight matrix, then we have:

$$\Phi = (E - W)^T (E - W) \quad (12)$$

Give $S = \{s_1, s_2, \dots, s_n\}$, let $\Phi = G(S)$ as the reconstruction graph [43].

2) THE NEW ACTIVE LEARNING ALGORITHM

Applying the graph Φ constructed above, the procedure of our novel active learning algorithm can be summarized as follows:

Input: Giving the samples $S = \{s_1, s_2, \dots, s_n\}$ in Euclidean space, which is partially labeled with class information.

- (1) **Build local line reconstruct graph based on LLE [37].** Under the semi-supervised condition, construct a nearest neighbor graph with weight W as demonstrated in the (11), and then calculate the locally linear reconstruction graph Φ as presented in the (12).
- (2) **Compute the kernel matrix R with the input kernel type.** With regard to the kernel type, there are several

uncommon kernels can be selected. For instance, it can convert polynomial and Gaussian kernel into manifold adaptive kernels.

- (3) **Set up a new kernel in reproducing kernel Hilbert space.** Replace the matrix F of reproducing kernel (7) by Φ , let r_i be the i -th column of the above R , and let $R(i, j)$ be the element of the i -th row, j -th column of R . Finally, we obtain the manifold adaptive kernel:

$$R_F(i, j) = R(i, j) - \rho r_i^T (E + \Phi R)^{-1} \Phi r_j \quad (13)$$

- (4) **Calculate deformation of convex TED in reproducing kernel Hilbert space.** Let v_i be the i -th column (or row, because R_F is symmetric) of the new kernel R_F that is calculated in the previous step. Initialize $a_{i,j} = 1$, then we can ultimately obtain:

$$b_j = \sqrt{\frac{\sum_{i=1}^n a_{i,j}^2}{\varphi}}, \quad j = 1, 2, \dots, n \quad (14)$$

$$a_i = (\text{diag}(b)^{-1} + R_F)^{-1} v_i, \quad i = 1, 2, \dots, n \quad (15)$$

- (5) **Choose samples.** The samples of the original dataset $S = \{s_1, s_2, \dots, s_n\}$ are arranged from big to small on the basis of the numerical value of b_j . Then, the maximally informative samples can be selected by the top t samples, because b_j and s_j are one-versus-one. It is noteworthy that people can label for the subset that is combined by the top t samples, which is easier than label for the whole dataset.

Output: All the original data samples are rearranged according to their importance, and give out the optimal subset that is used to do the classification tasks.

After choosing the most useful samples, many kinds of image classification tasks can be successfully carried out. Our novel algorithm, SSAED, is substantially based on the TED, but the SSAED method attaches great importance to the geometric structure of data space. Comparing with the MAED algorithm, our novel active learning method also bases on manifold assumption, which defines the adjacency graph with a certain neighborhood size for all samples. The differences between these two manifold learning methods mainly have two aspects. In the first place, the styles of constructing graph are remarkably dissimilar. Our novel algorithm adopts LLE to construct the graph, while MAED uses LE to do this. In the second place, the ways that use the class information of samples are different. The novel algorithm constructs the manifold adaptive kernel with the semi-supervised manner, which not only uses the structure of data but also considers the category information of the data. It makes the proposed method more robust, efficient and effective. However, MAED only adopts the supervised manner to construct the manifold adaptive kernel.

E. EXPERIMENTAL MATERIALS

In this study, we evaluate the performance of our novel active learning algorithm, SSAED, on several benchmark datasets.

TABLE 1. Properties of the maximal image databases used in this study, where n , d and c denote the number of sizes, dimensionality and class respectively.

Dataset	N	D	C
MNIST	4000	784	10
USPS	900	256	10
PIE	1700	1024	10
ORL32	400	1024	40
ORL64	400	4096	40
Yale32	165	1024	15
Yale64	165	4096	15
Yale-B	2414	1024	38
CBCL	400	361	2

These datasets are all image data that are very popular in the machine learning field.

1) DESCRIPTION OF DATASETS

Programs are run on nine extremely famous real-world image datasets: MNIST, USPS, PIE, ORL32, ORL64, Yale32, Yale64, Yale-B, and CBCL. Except for the MNIST and USPS handwritten digital datasets, the other seven datasets are all human face images. ORL32 and ORL64 are datasets of two different dimensions in the ORL dataset. Similarly, Yale32 and Yale64 are datasets of two different dimensions in the Yale dataset. Table 1 gives out the attributes of the seven broad categories that are used in this study.

2) EXPERIMENTAL SETTING

In the data classification, a k -fold cross-validation strategy is adopted by us. As for the parameter k , we let it be 10. All experiments are performed in the software environment of MATLAB2017a and Windows 7 operation system. Concomitantly, the hardware specifications are Core i5 processor, 4G memory, independent graphics card, and 500G mechanical hard disk.

III. RESULTS AND DISCUSSION

We estimate the performance of the novel algorithm, SSAED, on several benchmark datasets. The comparison is made with two classification methods: the MAED algorithm and Random sampling approach (Random).

- Random sampling method conformably chooses examples as the training data, which does not consider the importance of each sample. That is to say, the strategy of the Random method is to select the samples in a disorderly manner. In this study, the Random method is used as the baseline of active learning.
- MAED algorithm, proposed by D. Cai et al. [31], selects the most informative samples with a restrictive parameter, which has been detailedly described in the method and materials section.

The fundamental aim of this study is to investigate whether the novel method can obtain a satisfactory classification result

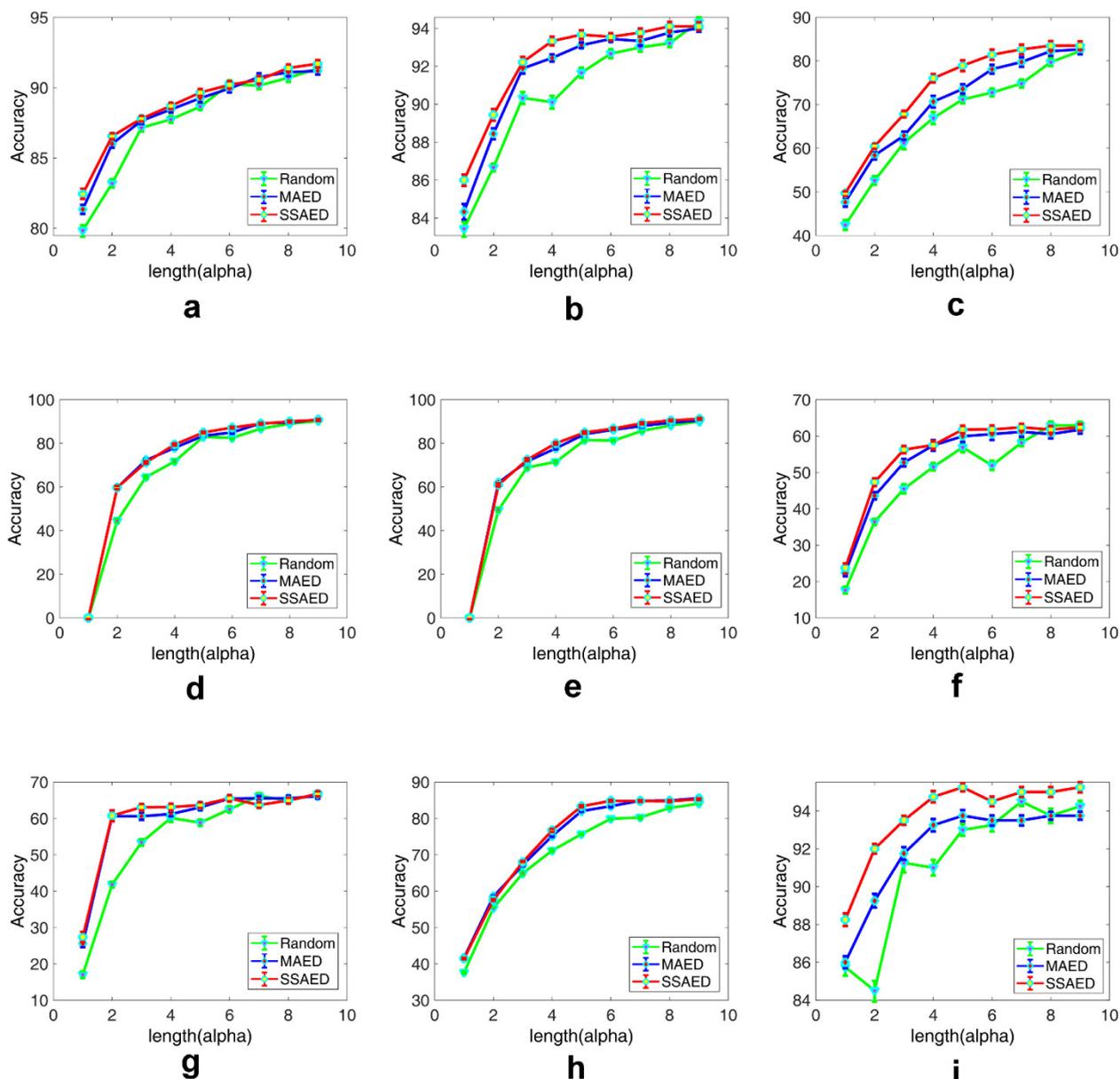


FIGURE 2. Average classification accuracy. From subfigure ‘a’ to ‘i’, they denote the classification results of the MNIST, USPS, PIE, ORL32, ORL64, Yale32, Yale64, Yale-B, and CBCL image dataset respectively.

under the condition that the training sample dataset’s size is small. We let proportion of the labeled training data (the samples that are selected by Random, MAED and SSAED methods) change from 0.1 to 0.9. In Fig.2, the ‘length(alpha)’ denotes the proportion of the selected training sample data. We use the 10-point system, for instance, in the figure, ‘2’, ‘4’, ‘6’, and ‘8’ denote ‘0.2’, ‘0.4’, ‘0.6’, and ‘0.8’ respectively, i.e., alpha=2 means that twenty percentage samples in the dataset $S = (s_1, s_2, \dots, s_n)$ are selected to be labeled and used as the training set.

A. AVERAGE CLASSIFICATION ACCURACIES ON WIDE RANGE TRAINING SETS

For each kind of image datasets (MNIST, USPS, PIE, ORL32, ORL64, Yale32, Yale64, Yale-B, CBCL), we give out the

classification accuracies based on the three algorithms. The range of selected training sample set varies from 10% to 90%. The classification results on all real-world datasets are shown in the Fig.2. The numerical values of the average classification accuracies and standard deviations on the nine datasets with the three sample selection approaches are shown in table 2.

As is shown in the Fig.2, the classification accuracy is increasing with more samples are selected as the training set. When compared with MAED and Random methods, there is no consistent winner. Overall, our novel method performs the best, while the Random method leads to the lowest classification accuracies in all datasets. In the following passages, we present out the detailed descriptions for each subfigure of the Fig.2.

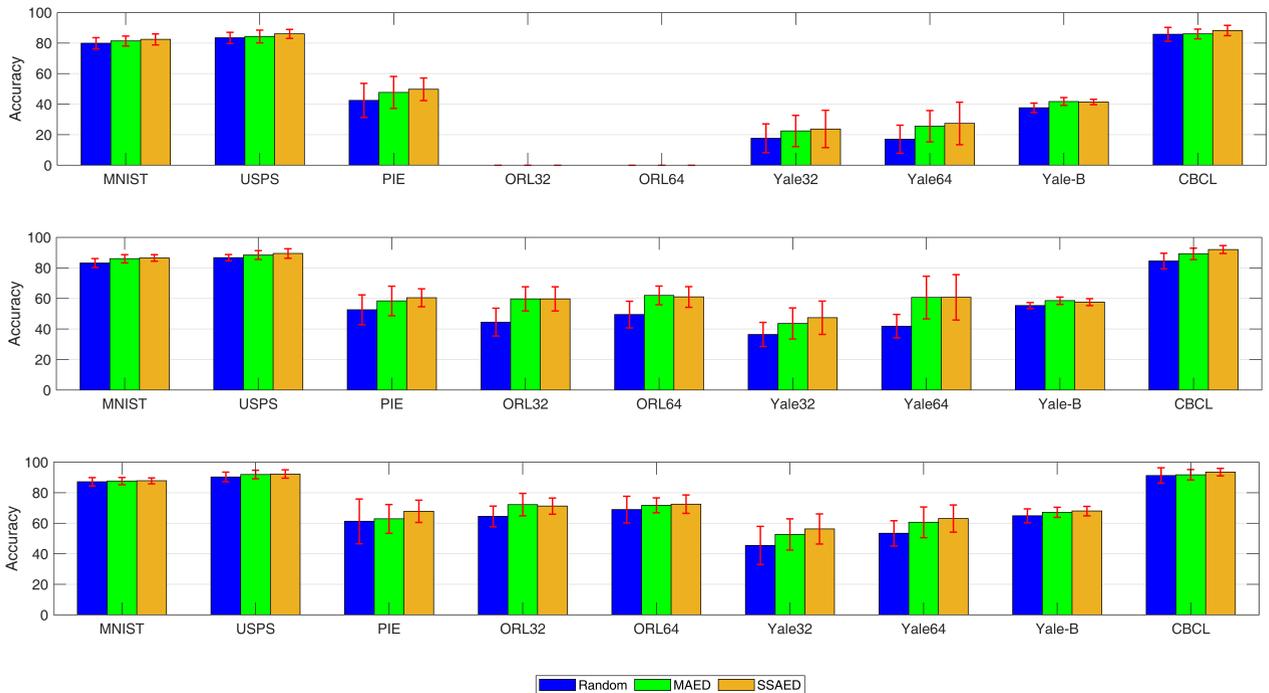


FIGURE 3. Comparisons among three algorithms with small size training sample datasets. The upper, middle and lower subfigures represent the classification accuracies under the condition that 10%, 20% and 30% of the data points are marked as training samples, respectively. The results of ORL32 and ORL64 in the top subfigure have not been shown out, which are owing to the classification accuracies are equal to zero in the case of only 10% of data points are selected as a training set.

As is shown in the Fig.2.a, our novel algorithm SSAED performs well when the size of the training sample set is small in the MNIST dataset. When 10% of samples are selected as the training set, the classification accuracy of SSAED is over 82%, while the Random method is lower than 80%. Except for the condition that 70% of samples used as the training set, SSAED performs better than the MAED algorithm. In the USPS handwritten digital image dataset, our new method is a consistent winner when it is compared with the MAED method (see the Fig.2.b). The discriminative ability of the selected informative samples by the SSAED algorithm is much better than that of Random methods on the USPS dataset when there are less than 50% of data points are selected. It is noteworthy that the SSAED method performs very well when the size of the training set is small. For instance, the classification accuracy of SSAED is over 86% and approximately overpass 3 percentage points compared with the result of Random method under the condition that only 10% of data points are selected to use as the training samples.

As can be seen from the Fig.2.c, the SSAED continuously shows better results than both MAED and Random methods in the PIE dataset. The accuracy of the new method overpasses 7 percentage points compared with the Random and 2 percentage points compared with the MAED method when only 10 percentage data samples are selected to use as the training samples. It can be seen from both Fig.2.d and Fig.2.e, SSAED shows a little bit better performance than the

MAED method in the ORL datasets (ORL32 and ORL64), but performs much better than the Random method. When the alpha changes from 0.2 to 0.7, the SSAED is less sensitive to outliers than the other methods. As for the classification accuracies in the Yale datasets (Yale32 and Yale64), we can know that the SSAED performs better than the MAED and Random in most of the cases (see the Fig.2.f and Fig.2.g). However, all the three methods show their pool classification abilities that are all lower than 70% on this kind of face image dataset. As can be seen from the Fig.2.h, comparison among the three methods shows that our novel method performs the best in the Yale-B dataset. Particularly, the SSAED continuously obtains higher classification accuracy than the Random method when alpha ranges from 0.1 to 0.9. As is shown in the Fig.2.i, the SSAED consistently outperforms both MAED and Random in the CBCL dataset. With alpha increasing, when it reaches at 0.5, the SSAED method gets the highest classification accuracy that is over 95%. It is noteworthy that the classification accuracy of the SSAED is over 88% under the condition that only 10% samples are labeled as the training data. Whereas, the accuracy of MAED is 86% on the same condition, and the Random method needs more than twice samples to be chosen as the training set to achieve 88% classification accuracy. The results in the CBCL clearly show the advantages of our novel algorithm in selecting the best samples for labeling, i.e., these labeled samples that are chosen by the SSAED can improve the classifier the most.

TABLE 2. The numerical value of average classification accuracy and standard deviation.

datasets	algorithms	10%	20%	30%	40%	50%	60%	70%	80%	90%
MNIST	Random	79.85±3.69	83.2±2.96	87.15±2.73	87.75±2.45	88.65±2.27	90.25±2.61	90.15±2.52	90.70±2.89	91.35±2.82
	MAED	81.35±3.26	86.00±2.71	87.65±2.38	88.45±2.19	89.25±2.88	89.95±2.96	90.75±2.76	91.10±2.81	91.20±2.70
	SSAED	82.45±3.59	86.55±2.10	87.80±1.96	88.70±2.24	89.65±2.48	90.2±2.63	90.55±2.70	91.40±2.56	91.70±2.64
USPS	Random	83.44±3.68	86.67±2.10	90.33±3.10	90.11±3.33	91.67±2.68	92.67±2.35	93.00±2.29	93.22±2.06	94.33±2.48
	MAED	84.33±4.14	88.44±2.88	91.89±2.72	92.44±1.95	93.11±1.46	93.44±1.52	93.33±2.34	93.78±1.90	94.00±1.83
	SSAED	86.00±2.93	89.44±3.11	92.22±2.72	93.33±2.28	93.67±2.35	93.56±1.95	93.78±2.17	94.11±2.29	94.11±2.16
PIE	Random	42.43±11.20	52.63±9.74	61.27±14.64	66.95±13.80	71.22±9.00	72.78±9.12	74.83±8.97	79.75±9.37	82.27±7.39
	MAED	47.68±10.53	58.38±9.65	62.90±9.42	70.68±13.53	73.60±10.73	78.12±10.04	79.77±11.03	82.25±10.78	82.67±9.79
	SSAED	49.75±7.42	60.45±5.86	67.83±7.28	76.05±9.92	78.93±12.27	81.43±11.83	82.68±10.71	83.52±10.36	83.52±9.79
ORL32	Random	0.00±0.00	44.50±9.11	64.50±6.75	71.75±7.91	83.00±5.37	82.50±6.12	86.75±5.66	89.00±5.30	90.25±3.43
	MAED	0.00±0.00	59.75±7.95	72.25±7.31	78.00±4.83	83.50±4.12	85.00±5.40	89.25±3.34	89.50±3.50	91.00±3.76
	SSAED	0.00±0.00	59.75±7.95	71.25±5.30	79.50±4.05	85.00±3.54	87.25±3.99	89.00±3.94	90.00±4.41	90.75±4.42
ORL64	Random	0.00±0.00	49.50±8.72	69.00±8.68	71.50±6.69	81.50±5.16	81.25±4.12	85.75±5.01	88.25±3.92	90.00±3.54
	MAED	0.00±0.00	62.00±6.10	71.75±4.87	77.75±4.48	84.25±6.98	86.25±6.15	88.00±5.99	89.50±6.10	90.50±4.22
	SSAED	0.00±0.00	61.00±6.79	72.50±6.01	80.00±5.27	85.00±5.14	86.75±5.01	89.25±3.74	90.50±3.87	91.25±3.77
Yale32	Random	17.65±9.40	36.43±7.98	45.51±12.48	51.54±10.35	56.91±13.12	51.95±12.61	58.2±9.96	62.90±11.41	62.98±9.41
	MAED	22.46±10.17	43.6±10.23	52.72±10.22	57.46±14.36	59.96±15.64	60.55±16.18	61.18±12.98	60.55±12.02	61.76±10.89
	SSAED	23.75±12.15	47.35±10.92	56.29±9.81	57.54±10.79	61.80±10.37	61.84±10.72	62.39±9.98	61.80±9.20	62.39±8.74
Yale64	Random	17.06±9.12	41.84±7.73	53.42±8.27	60.15±11.07	58.86±8.61	62.54±8.46	66.21±7.31	64.89±4.25	66.76±7.07
	MAED	25.59±10.19	60.63±13.96	60.63±10.02	61.21±7.04	63.05±7.73	65.51±7.75	65.55±6.33	65.55±6.91	66.14±6.51
	SSAED	27.43±13.93	60.77±14.85	63.09±8.91	63.13±9.25	63.68±7.87	65.55±8.59	63.71±7.64	64.93±8.03	66.76±7.07
Yale-B	Random	37.66±3.10	55.34±1.98	64.92±4.53	71.17±3.43	75.68±2.25	79.91±2.63	80.28±2.39	82.89±2.13	84.09±1.80
	MAED	41.72±2.61	58.53±2.43	67.19±3.30	75.19±2.51	82.02±2.28	83.39±2.97	84.88±2.18	84.92±2.28	85.59±2.15
	SSAED	41.47±1.76	57.58±2.27	68.02±3.08	76.72±4.17	83.35±2.65	84.84±2.36	84.80±2.15	84.76±2.34	85.29±2.18
CBCL	Random	85.75±4.57	84.50±5.11	91.25±5.03	91.00±4.12	93.00±3.07	93.25±3.34	94.50±2.30	93.75±3.77	94.25±2.90
	MAED	86.00±3.16	89.25±3.74	91.75±3.34	93.25±3.34	93.75±2.95	93.50±2.69	93.50±2.69	93.75±2.12	93.75±2.12
	SSAED	88.25±3.34	92.00±2.58	93.50±2.42	94.75±2.99	95.25±2.19	94.50±2.58	95.00±2.64	95.00±2.64	95.25±2.75

The first column of the table denotes the datasets that are used in this study. The second column is the three algorithms, Random, MAED and SSAED. The columns that are from the third to eleventh column denote the results which are on different sizes of training sample datasets.

From the subfigure ‘a’ to ‘i’ that is orderly presented out in the Fig.2, we can clearly see that the classification accuracies in the handwritten datasets (MNIST and USPS) are all extremely satisfactory. Whether in a small training sample set or a large training sample set, good classification accuracy can be obtained. However, the classification results are unsatisfactory in some of the human face image datasets, e.g., Yale32 and Yale64. Additionally, we can also know that the SSAED can continuously perform better than the other two methods on the three datasets: USPS, PIE, and CBCL. It is important to note that the SSAED can obtain satisfactory results of which classification accuracies are more than 82% in the case of only choose 10% samples as the training set on the MNIST, USPS and CBCL datasets.

B. AVERAGE CLASSIFICATION ACCURACIES ON SMALL RANGE TRAINING SETS

Due to the fundamental aim of active learning is to select the best samples to train the classifier model and expect to obtain

a satisfactory classification accuracy, experimental result comparison bars on three kinds of small size training sets are given in this part. In terms of the percentage, we choose the training set of which size accounts for 10%, 20%, and 30% of the whole dataset respectively. All the comparisons are implemented in the nine image datasets (MNIST, USPS, PIE, ORL32, ORL64, Yale32, Yale64, Yale-B and CBCL) with the three algorithms (Random, MAED and SSAED).

From Fig.3, we can see that SSAED performs better than both Random and MAED methods in almost all circumstances. Except for the ORL64 and Yale-B datasets with 20% and ORL32 dataset with 30% of data samples are selected as the training set, our novel algorithm is not lower than the MAED. The Random method performs the worst in all conditions (see three subfigures), this indirectly proves that active learning can reduce the cost of artificial marking. It is noteworthy that the SSAED algorithm performs extremely well in the MNIST, USPS and CBCL datasets. In these three datasets, all the classification accuracies with small size

TABLE 3. Comparison of processing time among the three algorithms, Random, MAED and SSAED, on all the image datasets applied in this paper (time/s).

database/algorithm	Random	MAED	SSAED
MNIST	22.54	223.51	226.86
USPS	0.97	15.39	15.89
PIE	0.27	0.65	0.63
ORL32	0.65	1.80	2.24
ORL64	2.75	4.11	5.24
Yale32	0.15	0.24	0.32
Yale64	0.52	0.66	0.86
Yale-B	67.59	1.43E+04	866.30
CBCL	0.34	1.49	1.84

training samples are higher than 80%, especially, in the CBCL dataset, the accuracy even approximates 90% with 10% of training samples and overpasses 90% with 20% and 30% of training samples.

C. EFFICIENCY ANALYSIS OF THE THREE ALGORITHMS

For testing the operation efficiency of the SSAED algorithm, comparative experiments were implemented in the MNIST, USPS, ORL32, ORL64, Yale32, Yale64, Yale-B and CBCL datasets. Some details of these datasets have been presented in table 1. The running time of each algorithm in the nine datasets is presented in table 3. We can see that the Random method has the shortest running time, which is attributed to its simple structure that doesn't consider the intrinsic geometry of the sample data space. The running time of MAED and SSAED is relatively close, which due to the processing steps are similar. Both MAED and SSAED methods attach great importance to the geometric structure of data space. It is important to note that the dimensionality of the data set affects the processing times of the algorithms, the higher the dimensionality of the dataset, the longer the algorithms run, e.g., ORL32, ORL64, Yale32 and Yale64 datasets confirm this viewpoint (see table 1 and table 3). Theoretically, due to the SSAED algorithm adopts the semi-supervised strategy to construct the special geometric structure graph, which has to consider both labeled samples and unlabeled samples information. Hence, it needs to take more time than the other two methods. Actually, the results in the table 3 confirm this viewpoint.

We have so far compared the three algorithms (Random, MAED and SSAED) on two handwritten digital image datasets (MNIST, USPS) and seven human face image datasets (PIE, ORL32, ORL64, Yale32, Yale64, Yale-B and CBCL). Many data suggest that the new algorithm can achieve higher classification accuracy. SSAED has some advantages over the MAED and Random. As is shown in Fig.2, Fig.3 and table 2, SSAED algorithm yields impressive results on the compared datasets. Two reasons could account for these phenomena. Firstly, the applied datasets

always have partially labeled data samples, which make the SSAED method adequately takes account of the class information of these partially labeled samples. For the unlabeled samples, SSAED still can use them effectively, because the graph is constructed with the semi-supervised style. Secondly, many samples that come from diverse categories overlap with each other seriously. All these indicate that construct graph with semi-supervised manner has obvious merits [33]–[36]. Among the three methods, the Random method performs the worst all the time on all the image datasets. Both MAED and SSAED algorithms are much better than the Random. The main reason is that the active learning method spontaneously selects the most useful samples to label as the training set, which is helpful for achieving maximum efficiency [31]. However, in the Random method, the training set is constructed by choosing the samples randomly, which easily leads to putting the useless samples into the training dataset. Therefore, it is difficult to obtain a satisfactory classification accuracy. Our experimental results further prove that the active learning algorithm has great application value.

IV. CONCLUSION

As is shown in section two, a novel active learning method has been presented for data classification by incorporating the locally linear reconstruction coefficients matrix into the learning course, where the coefficients are computed with a semi-supervised manner. The new algorithm attaches great importance to the potential manifold structure of the data space. The classification accuracies which are on nine real-world datasets indicate that SSAED surpasses both the MAED and Random methods, especially under the condition that only a few examples are chosen to use as the training dataset. However, it still has some shortcomings in this research. One weakness of the novel approach is that the experimental result is not satisfying in all the image datasets. The new method is only sensitive to some datasets, but not to all. In the future, we need to design a universal approach that can perform well in more datasets. Additionally, the key improvement of our algorithm is the locally linear reconstruction frame, which comes from the LLE. We adopt this frame to calculate the reconstruction coefficients that are used to construct the novel graph, but it is difficult to determine which constructive graph is the best. Thereby, we will try to use other methods, such as sparse reconstruction with the semi-supervised or supervised manner, to construct the new graph in the next classification study.

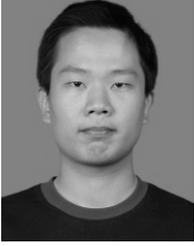
REFERENCES

- [1] M. Peña, R. Liao, and A. Brenning, "Using spectrotemporal indices to improve the fruit-tree crop classification accuracy," *ISPRS J. Photogram. Remote Sens.*, vol. 128, pp. 158–169, Jun. 2017, doi: [10.1016/j.isprsjprs.2017.03.019](https://doi.org/10.1016/j.isprsjprs.2017.03.019).
- [2] S. Rathore, M. Habes, M. A. Iftikhar, A. Shacklett, and C. Davatzikos, "A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages," *NeuroImage*, vol. 155, pp. 530–548, Jul. 2017, doi: [10.1016/j.neuroimage.2017.03.057](https://doi.org/10.1016/j.neuroimage.2017.03.057).

- [3] A. Boukharouba and A. Bennia, "Novel feature extraction technique for the recognition of handwritten digits," *Appl. Comput. Informat.*, vol. 13, no. 1, pp. 19–26, Jan. 2017, doi: [10.1016/j.aci.2015.05.001](https://doi.org/10.1016/j.aci.2015.05.001).
- [4] F. Shen, C. Shen, X. Zhou, Y. Yang, and H. T. Shen, "Face image classification by pooling raw features," *Pattern Recognit.*, vol. 54, pp. 94–103, Jun. 2016, doi: [10.1016/j.patcog.2016.01.010](https://doi.org/10.1016/j.patcog.2016.01.010).
- [5] L. Zhang, Q. Liang, Y. Shen, M. Yang, and F. Liu, "Image set classification based on synthetic examples and reverse training," *Neurocomputing*, vol. 228, pp. 3–10, Mar. 2017, doi: [10.1016/j.neucom.2016.04.067](https://doi.org/10.1016/j.neucom.2016.04.067).
- [6] Y. Xu, Z. Zhang, G. Lu, and J. Yang, "Approximately symmetrical face images for image preprocessing in face recognition and sparse representation based classification," *Pattern Recognit.*, vol. 54, pp. 68–82, Jun. 2016, doi: [10.1016/j.patcog.2015.12.017](https://doi.org/10.1016/j.patcog.2015.12.017).
- [7] C. Jou and H.-C. Lee, "Handwritten numeral recognition based on simplified structural classification and fuzzy memberships," *Expert Syst. Appl.*, vol. 36, no. 9, pp. 11858–11863, Nov. 2009, doi: [10.1016/j.eswa.2009.04.025](https://doi.org/10.1016/j.eswa.2009.04.025).
- [8] H. Karimi, A. Esfahanimehr, M. Mosleh, F. M. J. Ghadam, S. Salehpour, and O. Medhati, "Persian handwritten digit recognition using ensemble classifiers," *Procedia Comput. Sci.*, vol. 73, pp. 416–425, Dec. 2015, doi: [10.1016/j.procs.2015.12.018](https://doi.org/10.1016/j.procs.2015.12.018).
- [9] L. B. Saldanha and C. Bobda, "An embedded system for handwritten digit recognition," *J. Syst. Archit.*, vol. 61, no. 10, pp. 693–699, Nov. 2015, doi: [10.1016/j.sysarc.2015.07.015](https://doi.org/10.1016/j.sysarc.2015.07.015).
- [10] M. Wang, Y. Wan, Z. Ye, and X. Lai, "Remote sensing image classification based on the optimal support vector machine and modified binary coded ant colony optimization algorithm," *Inf. Sci.*, vol. 402, pp. 50–68, Sep. 2017, doi: [10.1016/j.ins.2017.03.027](https://doi.org/10.1016/j.ins.2017.03.027).
- [11] G. Revollo Sarmiento, M. Cipolletti, M. Perillo, C. Delrieux, and G. M. Perillo, "Methodology for classification of geographical features with remote sensing images: Application to tidal flats," *Geomorphology*, vol. 257, pp. 10–22, Mar. 2016, doi: [10.1016/j.geomorph.2015.12.020](https://doi.org/10.1016/j.geomorph.2015.12.020).
- [12] A. P. D. Silva, "Optimization approaches to supervised classification," *Eur. J. Oper. Res.*, vol. 261, no. 2, pp. 772–788, Sep. 2017, doi: [10.1016/j.ejor.2017.02.020](https://doi.org/10.1016/j.ejor.2017.02.020).
- [13] L. Ma, M. Li, X. Ma, L. Cheng, P. Du, and Y. Liu, "A review of supervised object-based land-cover image classification," *ISPRS J. Photogram. Remote Sens.*, vol. 130, pp. 277–293, Aug. 2017, doi: [10.1016/j.isprsjprs.2017.06.001](https://doi.org/10.1016/j.isprsjprs.2017.06.001).
- [14] D. A. Cohn, Z. Ghahramani, and M. I. Jordan, "Active learning with statistical models," *J. Artif. Intell. Res.*, vol. 4, pp. 129–145, Jul. 2018, doi: [10.1613/jair.295](https://doi.org/10.1613/jair.295).
- [15] S. Dasgupta and D. Hsu, "Hierarchical sampling for active learning," in *Proc. 25th Int. Conf. Mach. Learn.-ICML*, 2008, pp. 208–215.
- [16] L. Liu, Y. Xie, Z. Wang, J. Liu, and Z. Shi, "A new graph based active learning method," *Procedia Eng.*, vol. 29, pp. 2610–2620, 2012, doi: [10.1016/j.proeng.2012.01.360](https://doi.org/10.1016/j.proeng.2012.01.360).
- [17] D. J. C. Mackay, "Information-based objective functions for active data selection," *Neural Comput.*, vol. 4, no. 4, pp. 590–604, Jul. 1992, doi: [10.1162/neco.1992.4.4.590](https://doi.org/10.1162/neco.1992.4.4.590).
- [18] B. Settles and M. Craven, "An analysis of active learning strategies for sequence labeling tasks," in *Proc. Conf. Empirical Methods Natural Lang. Process.-EMNLP*, 2008, pp. 1070–1079.
- [19] Y. Freund, H. S. Seung, N. Tishby, and E. Shamir, "Selective sampling using the query by committee algorithm," *J. Mach. Learn.*, vol. 28, pp. 133–168, Aug. 2017, doi: [10.1023/A:1007330508534](https://doi.org/10.1023/A:1007330508534).
- [20] Y. Guo and R. Greiner, "Optimistic active learning using mutual information," in *Proc. Int. Joint Conf. Artif. Intell.*, Hyderabad, India, 2007, pp. 823–829.
- [21] M. Lindenbaum, S. Markovitch, and D. Rusakov, "Selective sampling for nearest neighbor classifiers," *Mach. Learn.*, vol. 54, no. 2, pp. 125–152, Feb. 2004, doi: [10.1023/b:mach.0000011805.60520.fe](https://doi.org/10.1023/b:mach.0000011805.60520.fe).
- [22] A. C. Atkinson, A. N. Donev, *Optimum Experimental Designs*. Oxford, UK: Oxford Univ. Press, 2007.
- [23] K. Yu, J. Bi, and V. Tresp, "Active learning via transductive experimental design," in *Proc. 23rd Int. Conf. Mach. Learn.*, Pittsburgh, PA, USA, 2006, pp. 1081–1088.
- [24] K. Yu, S. Zhu, W. Xu, and Y. Gong, "Non-greedy active learning for text categorization using convex transductive experimental design," in *Proc. 31st Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.-SIGIR*, Singapore, 2008, pp. 635–642.
- [25] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *J. Mach. Learn. Res.*, vol. 7, pp. 2399–2434, Nov. 2006.
- [26] N. Loeff, D. Forsyth, and D. Ramachandran, "ManifoldBoost: Stage-wise function approximation for fully-, semi- and un-supervised learning," in *Proc. 25th Intern. Conf. Mach. Learn.*, Helsinki, Finland, 2008, pp. 600–607.
- [27] J. B. Tenenbaum, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000, doi: [10.1126/science.290.5500.2319](https://doi.org/10.1126/science.290.5500.2319).
- [28] X. He and P. Niyogi, "Locality preserving projections," in *Proc. Conf. Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2003, pp. 153–160.
- [29] R. Hadsell, S. Chopra, and Y. Lecun, "Dimensionality reduction by learning an invariant mapping," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Washington, DC, USA, vol. 2, Jul. 2006, pp. 1735–1742.
- [30] M. Belkin and P. Niyogi, "Laplacian Eigenmaps and spectral techniques for embedding and clustering," in *Proc. Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2001, pp. 585–591.
- [31] D. Cai and X. He, "Manifold adaptive experimental design for text categorization," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 4, pp. 707–719, Apr. 2012, doi: [10.1109/tkde.2011.104](https://doi.org/10.1109/tkde.2011.104).
- [32] V. Sindhwani, P. Niyogi, and M. Belkin, "Beyond the point cloud: From transductive to semi-supervised learning," in *Proc. Int. Conf. Mach. Learn.*, Bonn, Germany, 2005, pp. 824–831.
- [33] X. J. Zhu, "Semi-supervised learning literature survey," *Comput. Sci.*, vol. 37, no. 1, pp. 63–77, Jan. 2005.
- [34] S. Zhu, X. Sun, and D. Jin, "Multi-view semi-supervised learning for image classification," *Neurocomputing*, vol. 208, pp. 136–142, Oct. 2016, doi: [10.1016/j.neucom.2016.02.072](https://doi.org/10.1016/j.neucom.2016.02.072).
- [35] H. Cecotti, "Active graph based semi-supervised learning using image matching: Application to handwritten digit recognition," *Pattern Recognit. Lett.*, vol. 73, pp. 76–82, Apr. 2016, doi: [10.1016/j.patrec.2016.01.016](https://doi.org/10.1016/j.patrec.2016.01.016).
- [36] D. Zhou, O. Bousquet, J. Weston, B. Schölkopf, and T. N. Lal, "Learning with local and global consistency," in *Proc. Conf. Adv. Neural Inf. Process. Syst.*, Vancouver, Canada, 2003, pp. 321–328.
- [37] S. T. Roweis, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000, doi: [10.1126/science.290.5500.2323](https://doi.org/10.1126/science.290.5500.2323).
- [38] C. Li, H. Liu, and D. Cai, "Active learning on manifolds," *Neurocomputing*, vol. 123, pp. 398–405, Jan. 2014, doi: [10.1016/j.neucom.2013.08.002](https://doi.org/10.1016/j.neucom.2013.08.002).
- [39] Q. Li, X. Shi, L. Zhou, Z. Bao, and Z. Guo, "Active learning via local structure reconstruction," *Pattern Recognit. Lett.*, vol. 92, pp. 81–88, Jun. 2017, doi: [10.1016/j.patrec.2017.04.022](https://doi.org/10.1016/j.patrec.2017.04.022).
- [40] L. Zhang, C. Chen, J. Bu, D. Cai, X. He, and T. S. Huang, "Active learning based on locally linear reconstruction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 10, pp. 2026–2038, Oct. 2011, doi: [10.1109/tpami.2011.20](https://doi.org/10.1109/tpami.2011.20).
- [41] X. He, W. Min, K. Zhou, and D. Cai, "Laplacian optimal design for image retrieval," in *Proc. 30th Ann. Int. ACM SIGIR Conf.*, Amsterdam, The Netherlands, 2007, pp. 119–126.
- [42] B. Schölkopf, A. J. Smola, *Learning With Kernels*. Cambridge, MA, USA: MIT Press, 2002.
- [43] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007, doi: [10.1109/tpami.2007.250598](https://doi.org/10.1109/tpami.2007.250598).



XINGLIANG XIONG received the master's degree from the College of Mathematics and Information Science, Wenzhou University, Wenzhou, China, in 2014. He is currently pursuing the Ph.D. degree with the School of Biological Science and Medical Engineering, Southeast University, Nanjing, China. His research interests include machine learning, statistical pattern recognition, and EEG signal processing.



MINGYU FAN received the B.Sc. degree in information and computing science from the Minzu University of China, Beijing, China, in 2006, and the Ph.D. degree in applied mathematics from the Academy of Mathematics and System Science, Chinese Academy of Science, Beijing, in 2011. He is currently a Professor with Wenzhou University, Wenzhou, China. His current research interests include data mining and pattern recognition algorithms and the applications of them on solving industrial problems.



ZHENJIE HONG received the B.Sc. and M.A. degrees from Zhejiang University, Hangzhou, China, in 1986 and 1989, respectively, and the Ph.D. degree from Shanghai University, Shanghai, China, in 2005. From 2005 to 2007, he was a Post-doctoral Fellow with the Shanghai Astronomical Observatory, Chinese Academy of Sciences. He is currently a Professor with Wenzhou University, Wenzhou, China. His current research interests include mathematical optimization, statistics, and machine learning algorithm.

• • •



CHUANG YU received the bachelor's degree from the Nanjing Institute of Technology, Nanjing, China, in 2017. He is currently pursuing the master's degree with the School of Mechanical, China University of Mining and Technology. His main direction is the processing of EEG signals.