Semi-Supervised Learning in the Framework of Data Multiple 1-D Representation

Jianzhong Wang
College of Sciences
Sam Houston State University
Huntsville, Texas 77341–2206, USA
e-mail: jzwang@shsu.edu

***Abstract*—The paper develops 1D-based ensemble method for semi-supervised learning (SSL). The method integrates the classifier based on *data 1-D representations* and *label boosting* in a serial ensemble. In each stage, the data set is first represented by several 1-D stacks, which preserve the local similarity between data samples. Then, a *1-D ensemble labeler* (1DEL) is constructed and used to create a *newborn labeled subset* from the unlabeled set. United with the subset, the original labeled is boosted for the next learning stage. The boosting process is repeated till the updated labeled set reaches a certain size. Finally, a 1DEL is applied again to build the classifier. The validity and effectiveness of the method are confirmed by experiments. Comparing to several other popular SSL methods, the results of the proposed method are very promising.**

***Keywords*–Data 1-D representation; regularization; label boosting; ensemble; semi-supervised learning.**

I. INTRODUCTION

In this paper, we introduce a novel ensemble method for SSL based on *data 1-D representation*. In SSL, the essential problem is data binary classification, which can be briefly described as follows: Assume that the samples (or members, points) of a given data set $X = \{\vec{x}_i\}_{i=1}^n \subset \mathbf{R}^m$ belong to two classes *A* and *B*, labeled by 1 and -1, respectively. Denote by y_j the label of the sample $\vec{x}_j$, where $y_j \in \{1, -1\}, 1 \le j \le n$. In a SSL problem, X is divided into two disjoint subsets: $X = X_\ell \cup X_u, X_\ell \cap X_u = \emptyset$, where the members in $X_\ell = \{\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_{n_0}\}$ have known labels $Y_\ell = \{y_1, y_2, \cdots, y_{n_0}\}$, while the labels for the members in $X_u = \{\vec{x}_{n_0+1}, \vec{x}_{n_0+2}, \cdots, \vec{x}_n\}$ are unknown. We often call a function $f : X \to \{1, -1\}$ a *classifier* (or *labeler*) on X. The *classification error* is measured by the misclassified number:

$$E(f) = |\{\vec{x} \in X |\ f(\vec{x}_i) \neq y_i\ 1 \le i \le n\}|,$$

where $|S|$ denotes the cardinality of a set S. Then, the quality of a classifier is measured by the *error rate* $E(f)/|X|$. The task of SSL is to find a classifier f with the error rate as small as possible.

The monograph [1] and the survey paper [2] gave a comprehensive review of various SSL methods, among which the popular ones are based on kernel technique such as transductive support vector machines, manifold regularization, and other graph-based methods [3] [4]. In these methods, using *kernel trick*, people construct a kernel function to map original samples onto a reproducing kernel Hilbert space (RKHS) [5], where the non-linear decision boundary in the raw data space becomes nearly linear. Thus, people can construct classifiers in the RKHS using regularization methods. The success of a kernel-based method strongly depends on the exploration of data structure by kernels. However, it is often difficult to design suitable kernels, which precisely explore the feature spaces. Recently, researchers have developed new SSL models, which construct classifiers without adopting kernel technique, for instance, the data-tree based method [6] [7] constructs the classifier based on the data multi-layer structure.

In all of the models above, a single classifier is employed to label unlabeled points. However, when a data set has a complicate intrinsic structure and high-dimensionality, a single classifier usually cannot complete the task satisfactorily. The proposed method takes the idea of ensemble methodology in the *multiple classifier systems* (MCSs) [8]: It build a final classifier by integrating multiple pre-classifiers. Since MCSs perform information fusion at different levels, they overcome the limitations of the traditional approaches [9]–[11].

The novelty of the introduced ensemble SSL method is the following: It adopts the framework of data 1-D representation, in which the data set is represented by several different 1-D sequences, then a classifier is constructed as an ensemble of pre-classifiers built on these sequences. Here, we choose 1-D models because 1-D decision boundary is a set of points on a line, which has the simplest topological structure. As a result, the pre-classifiers can be easily constructed by classical 1-D regularization methods without using kernel trick or data trees. Furthermore, the simplicity of 1-D models makes the algorithm for building the final classifier relatively reliable and stable. We new describe the architecture and technological process of our method in the following.

1) The data set X is first mapped to several 1-D sets $\{T^i\}_{i=1}^k$, which preserve the local similarity of members in X. Correspondingly, the couple $\{X_\ell, X_u\}$ is mapped to $\{T_\ell^i, T_u^i\}$ for each 1-D set T^i.
2) A pre-classifier g^i on X is constructed based on T^i by a 1-D regularization method. Then an ensemble labeler g on X is assembled from $\{g^i\}_{i=1}^k$ to label all members of X.
3) A *feasibly confident subset* $L \subset X_u$ is produced by g. According to the *class weights* of the members of L, a half of members in L is chosen into the *newborn labeled subset* S. Then, the initial labeled set X_ℓ is boosted to $X_\ell^{new} = X_\ell \cup S$.
4) The procedure above is repeated till the updated labeled set X_ℓ^{new} reaches a certain size. Finally, the classifier f is obtained by applying the ensemble labeler g on the newest couple $\{X_\ell^{new}, X_u^{new}\}$.

Our strategy adopts *Model-guided Instance Selection* (MIS) approach [9], but is slightly different from AdaBoost algorithm [12] in the sense that AdaBoost updates the misclassified weights on X_u, while our method updates the set X_u itself.

The paper is organized as follows: In Section II, we develop the 1-D based ensemble SSL method. In Section III, we

 ISBN: 978-1-61208-477-0

demonstrate the validity of our method in two examples and give the comparison of our results with other methods. The conclusion is given in the last section.

II. THE 1-D BASED ENSEMBLE SSL METHOD

In this section, we introduce the novel SSL method based on data 1-D representation.

A. Data 1-D Representations

Assume that the data set X is initially arranged in a stack $\mathbf{x} = [\vec{x}_1, \cdots, \vec{x}_n]$, where the first n_0 members are in Class A and others are in Class B. Let $d(\vec{x}, \vec{y})$ be a metric on X that measures the dissimilarity between the points of X. Let π be an index permutation of the index sequence $[1, 2, \cdots, n]$, which induces a permutation P_π on the initial stack $\mathbf{x}$, yielding a stack of X headed by $\vec{x}_{\pi(1)}$: $\mathbf{x}_\pi = P_\pi \mathbf{x} = [\vec{x}_{\pi(1)}, \cdots, \vec{x}_{\pi(n)}]$. We define the set of all permutations of X headed by $\vec{x}_\ell$ by

$$\mathcal{P}_\ell = \{P_\pi; \quad \pi(1) = \ell\}.$$

According to [13], the *shortest-path sorting of* X headed by $\vec{x}_\ell$ is the stack $\mathbf{x}_\pi$ that minimizes the path starting from $\vec{x}_\ell$ and though all points in X, i.e., $\mathbf{x}_\pi = P_\pi \mathbf{x}$, where P_π is given by

$$P_\pi = \underset{P \in \mathcal{P}_\ell}{\arg\min} \sum_{j=1}^{n-1} d((P\mathbf{x})_j, (P\mathbf{x})_{j+1}). \tag{1}$$

Let the stack $\mathbf{x}_\pi$ be the shortest-path sorting of X headed by $\vec{x}_\ell$. Set

$$t_1 = 0, \ t_{j+1} - t_j = \frac{d(\vec{x}_{\pi(j)}, \vec{x}_{\pi(j+1)})}{\sum_{k=1}^{n-1} d(\vec{x}_{\pi(k)}, \vec{x}_{\pi(k+1)})}. \tag{2}$$

Then, the stack $\mathbf{t} = [t_1, \cdots, t_n]$ is called the 1-D (shortest-path) representation of X headed by $\vec{x}_\ell$.

The problem (1) has NP computational complexity. A greedy algorithm to find an approximation of P_π in (1) is referred to [13]. Once, P_π is found, the corresponding 1-D representation is obtained by (2).

Denote by T the set of the components of $\mathbf{t}$. The bijective mapping $h : T = h(X_\ell)$ is called a 1-D (shortest-path) embedding of X headed by $\vec{x}_\ell$, which also map the unlabeled set X_u onto $T_u = h(X_u) \subset T$. Then, a classifier on T induces a classifier on X. Since T is a 1-D set, its class decision boundary is reduced to a discrete set in $[0, 1]$.

B. The 1-D based ensemble labeler

Although the simplest topological structure of data 1-D representation reduces the decision boundary to a discrete set in $[0, 1]$ points, a single 1-D representation cannot truly preserve the data similarity because the sorting is a serial process that makes earlier selected adjacent pairs are more similar than the later selected ones. To overcome the drawback of a single 1-D embedding, we employ the *spinning technique* to build several 1-D representations. Based on each of them, we first construct a pre-classifier, then assemble an ensemble labeler from them. The following is the details.

Let $\vec{h} = [h_1, \cdots, h_k]$ be a k-ple 1D-embedding and P_i be the permutation operator on X corresponding to h_i such that the stack $\mathbf{x}_{\pi_i} = P_i \mathbf{x}$ is headed by a randomly selected point $\vec{x}_{\pi_i(1)}$. The embedding h_i produces a 1-D representation of X: $\mathbf{t}^i = h_i(\mathbf{x}_{\pi_i})$. For a function f on X, $s^i = f \circ h_i^{-1}$ is a function on $\mathbf{t}^i$. We now represent a function f on X by its vector form $\mathbf{f} = [f_1, \cdots, f_n]$, $f_j = f(\vec{x}_j)$, and a function s on $\mathbf{t}^i$ by the vector $\mathbf{s} = [s_1^i, \cdots, s_n^i]$, $s_j^i = s(t_j^i)$.

Let $T_\ell^i = h_i(X_\ell)$ and $T_u^i = h_i(X_u)$. Using a classical regularization method, we construct a pre-classifier g_i for X based on the couple $\{T_\ell^i, T_u^i\}$. For instance, denote by $C^1[0, 1]$ the space of smooth functions on $[0, 1]$ and by $Ds_j = (s(t_{j+1}^i) - s(t_j^i))/(t_{j+1}^i - t_j^i)$ the difference quotient of $s \in C^1[0, 1]$ on the stack $\mathbf{t}^i$ at t_j^i. Let q^i be the solution of the following constrained minimization problem:

$$q^i = \underset{s \in C^1[0,1]}{\arg\min} \frac{1}{n_0} \sum_{j=1}^{n_0} \left(s(h^i(\vec{x}_j)) - y_j\right)^2 + \frac{\lambda}{2} \sum_{j=1}^{n-1} (Ds_i)^2, \tag{3}$$

subject to the constraint

$$\frac{1}{n} \sum_{j=1}^{n} s(t_j^i) = M,$$

where M can be chosen to $M = \frac{1}{n_0} \sum_{j=1}^{n_0} y_j$. We denote by $\vec{1}$ the vector whose all entries are 1, denote by I_{n_0} the $n \times n$ diagonal matrix, in which only $(\pi^i(j), \pi^i(j))$-entries are 1, $1 \le j \le n_0$, but others are 0. Set $w_0 = w_n = 0, w_j = 1/(t_{j+1}^i - t_j^i)^2$, and denote by $D = [D_{i,j}]$ the $n \times n$ three-diagonal matrix, in which

$$\begin{cases} D_{j,j} = w_{j-1} + w_j & 1 \le j \le n, \\ D_{j,j+1} = D_{j+1,j} = -w_j & 1 \le j \le n-1, \end{cases}$$

Then, the vector representation of q^i on the stack $\mathbf{t}^i$ is the following solution

$$\mathbf{q}^i = (I_{n_0} + n_0 \lambda D)^{-1} (\vec{y} + \mu \vec{1}), \tag{4}$$

with

$$\mu = \frac{M - \mathcal{E}\left((I_{n_0} + n_0 \lambda D)^{-1} \vec{y}\right)}{\mathcal{E}\left((I_{n_0} + n_0 \lambda D)^{-1} \vec{1}\right)},$$

where $\mathcal{E}(\vec{v})$ denotes the mean value of the vector $\vec{v}$. We define the pre-classifier on X associated with the 1-D embedding h_i by $g^i = q^i \circ h_i^{-1}$. Finally, we define 1DEL on X by

$$g(\vec{x}) = \frac{1}{k} \sum_{i=1}^{k} \operatorname{sign}(g^i(\vec{x})), \quad x \in X. \tag{5}$$

C. The newborn labeled subset selector

Using the 1-D ensemble labeler g in (5), we construct

$$L^+ = \{\vec{x} \in X_u; \ g(\vec{x}) = 1\}, \quad L^- = \{\vec{x} \in X_u; \ g(\vec{x}) = -1\}.$$

In a great chance, L^+ contains the members in Class A, while L^- contains the members in Class B. We call $L = L^+ \cup L^-$ the *feasibly confident subset* created by g. For convenience, we denote the set operator that create the feasibly confident subset L from X_u by $\mathbf{G} : \mathbf{G}(X_u) = L$.

We now select a half of the members in L to form a *newborn labeled subset* $S = S^+ \cup S^-$, where S^+ contains all Class-A members in L^+ and S^- contains all Class-B members in L^-. They are constructed as follows. Let X_ℓ^+ contain all Class-A members of X_ℓ and X_ℓ^- contain all Class-B members of X_ℓ. For each $\vec{x} \in L$, define $d(\vec{x}, X_\ell^+) = \min_{\vec{y} \in X_\ell^+} d(\vec{x}, \vec{y})$

 ISBN: 978-1-61208-477-0

and $d(\vec{x}, X_\ell^-) = \min_{\vec{y} \in X_\ell^-} d(\vec{x}, \vec{y})$. We now associate $\vec{x}$ with the *class weight*

$$w(\vec{x}) = \frac{d(\vec{x}, X_\ell^-)}{d(\vec{x}, X_\ell^-) + d(\vec{x}, X_\ell^+)}.$$

Finally, let the set S^+ contain the half of members of L^+ with the greatest class weights and S^- contain the half of members in L^- with the smallest class weights. We call the operator $\mathbf{S} : \mathbf{S}(L) = S$ a *newborn labeled subset selector* and call the composition $\mathbf{M} = \mathbf{S} \circ \mathbf{G}$ a *newborn labeled subset creator* because the newborn labeled subset $S = M(X_u)$.

D. Construction of the final classifier

We now build the (final) classifier by a serial ensemble, in which the labeled set is cumulatively boosted. Let the initial labeled set be equipped with the index 0: $X_\ell^0 = X_\ell$. Starting from X_ℓ^0, we apply the newborn labeled subset creator $\mathbf{M}_1$ to create a newborn labeled set S^1, which is united with X_ℓ^0 to produce $X_\ell^1 = X_\ell^0 \cup S^1$. Repeating the procedure n times, the labeled set will be cumulatively boosted to a labeled set X_ℓ^n:

$$X_\ell^0 \Longrightarrow X_\ell^1 \Longrightarrow \cdots \Longrightarrow X_\ell^n.$$

We set a *boosting-stop parameter* $p, 0 < p < 1$. The process will not be terminated until the labeled set X_ℓ^n reaches the size $|X_\ell^n| \geq p|X|$. Finally, we apply 1DEL on the couple $\{X_\ell^n, X_u^n\}$ to construct the final classifier f on X, which labels each $\vec{x} \in X$ by $\operatorname{sign} f(\vec{x})$.

III. Experiments

We use two benchmark databases of handwritten digits, MNIST [21] and USPS [22] in the experiments to present the validity and effectiveness of the proposed method. In the literature of machine learning, MNIST is often used to test the error rate of classifiers obtained by supervised learning. The best result for the error rate up to 2012 was 0.23%, reported in [14] by using the convolutional neural network technique. In 2013, the authors of [15] claimed to achieve 0.21% error rate using DropConnect, which is based on regularization of neural networks. Because in SSL no large training set is available for producing classifiers, the error rates obtained by SSL methods usually are much higher than the claimed error rates obtained by supervised learning. Besides, the error rates of SSL are strongly dependent the size of the initial label set X_ℓ. In general, the smaller the size of X_ℓ, the higher the error rate. Hence, it is unfair to compare the error rates obtained by SSL methods to the above recorded ones.

In all of our experiments, the spin number 3 is used for constructing 1DEL while 20 for building the final classifier, and the boosting-stop parameter p is set to 0.7.

For comparison, we choose the same data setting as in [7]: In MINST, for each of the digits $\{3, 4, 5, 7, 8\}$, 200 samples were selected at random so that the cardinality of the data set is $|X| = 1000$, where the digit 8 is assigned to Class B, and others belong to Class A. In USPS, for each of the digits $0-9$, 150 samples are selected at random so that $|X| = 1500$, where the digits 2 and 5 are assigned to Class B, and others belong to Class A. In all experiments, the initial labeled set X_0 is preset to 10 various sizes of $10, 20, \cdots, 100$, respectively, and the labeled digits are distributed evenly on each chosen digit.

TABLE I. ERROR RATE OF THE PROPOSED 1-D BASED ENSEMBLE SSL METHOD FOR 50 RANDOMLY SELECTED SUBSETS FROM MNIST WITH $|X| = 1000$.

$\lvert X_0 \rvert$	10	20	30	40	50	60	70	80	90	100
Mean%	7.8	7.9	4.6	2.5	2.1	1.9	1.9	1.9	1.2	1.2
Min%	7.6	7.9	4.6	1.9	2.1	1.9	1.9	1.9	1.2	1.2
Max%	19.4	7.9	4.6	3.5	2.1	1.9	1.9	1.9	1.2	1.2
STD%	1.7	0	0	0.7	0	0	0	0	0	0

TABLE II. ERROR RATE OF THE PROPOSED 1-D BASED ENSEMBLE SSL METHOD FOR 50 RANDOMLY SELECTED SUBSETS FROM USPS WITH $|X| = 1500$.

$\lvert X_0 \rvert$	10	20	30	40	50	60	70	80	90	100
Mean%	3.3	2.1	1.5	1.5	1.3	1.4	1.4	1.4	1.4	1.2
Min%	2.0	1.3	1.5	1.5	1.3	1.4	1.4	1.4	1.4	1.2
Max%	16.8	2.9	1.7	1.5	1.3	1.4	1.4	1.4	1.4	1.2
STD%	1.99	0.8	0.02	0	0	0	0	0	0	0

Note that a vector $\vec{x} \in X$ is originally represented by a $c \times c$ matrix $[x_{i,j}]_{i,j=1}^c$, where $c = 20$ for MNIST and $c = 16$ for USPS. To reduce the shift-variance, we define the 1-shift distance between two digit images [16]:

$$d(\vec{x}, \vec{y}) = \min_{\substack{|i'-i| \leq 1 \\ |j'-j| \leq 1}} \sqrt{\sum_{i=2}^{c-1} \sum_{j=2}^{c-1} (x_{i,j} - y_{i',j'})^2}.$$

We first run our algorithm on 50 subsets (with 1000 members) randomly chosen from the MNIST database and show the test results in Table I, where the first row is the number of samples in X_ℓ, and the $2^{nd} - 5^{th}$ rows are the mean, minimum, maximum, and standard deviation of the error rates of the 50 tests, respectively. In the second experiment, we run our algorithm for USPS is a similar way: 50 subsets with 1500 members are randomly chosen from USPS database. The test results are shown in Table II, where the setting for the rows is the same as in Table I. The Tables I and II show that the standard deviations of the error rates are quite small, particular when the known labeled members are more than 1%. This indicates the high stability of the proposed SSL algorithm.

In Figure 1, we give the comparison of the average error rates (of 50 tests) of our 1-D based ensemble method to Laplacian Eigenmaps (Belkin & Niyogi, 2003 [3]), Laplacian Regularization (Zhu et al., 2003 [17]), Laplacian Regularization with Adaptive Threshold (Zhou and Belkin, 2011 [18]), and Haar-Like Multiscale Wavelets on Data Trees (Gavish et al., 2011 [7]) on the subsets randomly chosen from both MNIST and USPS databases. The results show that our method achieves competitive results comparing to other SSL methods.

We have also applied the proposed method on the real-world applications, such as the classification of hyperspectral images [19] and the face recognition [20]. In these experiments, we have even adopted a much simpler label boosting method: Choosing the newborn labeled subset at random. The obtained results are still very promising and superior over many other popular methods. It is also worth to point out that the method is not very sensitive to the parameters. For instance, in our experiments, if spinning numbers are set to 3–5, and the boosting-stop parameter is set in the range of

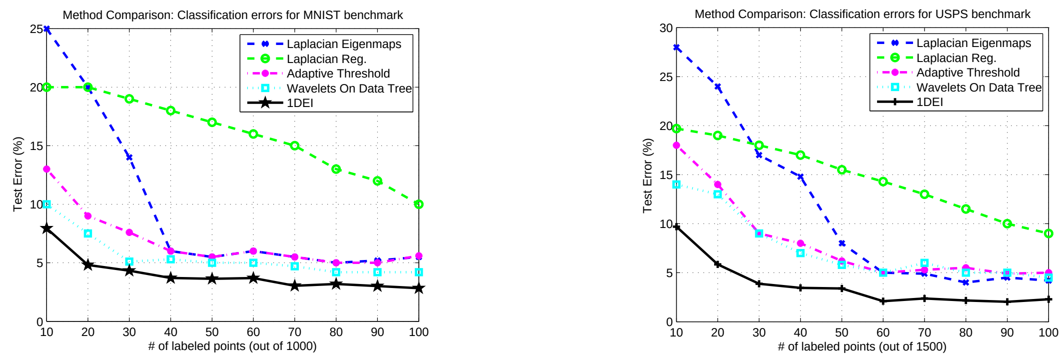

Figure 1. RESULT COMPARISON WITH DIFFERENT SSL MODELS.

0.6–0.8, the results are similar. The detailed discussion on the parameter tuning can be found [19] [20].

IV. Conclusion

We proposed a new ensemble method for SSL based on data 1-D representations, which enable us to construct ensemble classifiers assembled from several pre-classifiers for the same data set using classical 1-D regularization technique. Furthermore, a label boosting technique is applied for robustly enlarging the labeled set to a certain size so that the final classifier is built based on the boosted labeled set. The experiments show that the performance of the proposed method is superior to many popular methods in SSL. The new method also exhibits a clear advantage for learning the classifier when only a small labeled set is given. Because the method is independent of the data dimensionality, it can also be applied to various types of data. Since the algorithms to construct the classifiers in the proposed method only employ 1-D regularization technique, avoiding the complicate kernel trick, they are simple and stable. It can be expected that the created 1-D framework in this paper will be applied to the development of more machine learning methods for different purposes. In the future work, we will study how to accelerate the sorting algorithm in 1-D embedding and consider to integrate the data-driven wavelets with the proposed method.

Acknowledgment

This work is supported by SHSU-2015 ERG Grant no. 250711. The authors would like to thank the anonymous reviewers for their highly insights and helpful suggestions.

References

[1] O. Chapelle, A. Zien, and B. Schölkopf, Semi-supervised Learning. MIT Press, 2006.

[2] X. Zhu, "Semi-supervised learning literature survey," University of Wisconsin-Madison, Computer Sciences TR-1530, July 2008.

[3] M. Belkin and P. Niyogi, "Using manifold structure for partially labeled classification," Advances in Neural Information Processing Systems, vol. 15, 2003, pp. 929–936.

[4] T. Joachims, "Transductive learning via spectral graph partitioning," in Proceedings of the 20th International Conference on Machine Learning, 2003, pp. 290–297.

[5] V. Vapnik, Statistical Learning Theory. Wiley-Interscience, New York, 1998.

[6] R. Coifman and M. Gavish, "Harmonic analysis of digital data bases," Applied and Numerical Harmonic Analysis. Special issue: Wavelets and Multiscale Analysis, 2011, pp. 161–197.

[7] M. Gavish, B. Nadler, and R. Coifman, "Multiscale wavelets on trees, graphs and high dimensional data: Theory and applications to semi supervised learning," in Proceedings of the 27th International Conference on Machine Learning, 2010, pp. 367–374.

[8] T. K. Ho, J. J. Hull, and S. N. Srihari, "Decision Combination in Multiple Classifier Systems," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 16, no. 1, Jan. 1994, pp. 66–75.

[9] L. Rokach, "Ensemble-based classifiers," Artif. Intell. Rev., vol. 33, 2010, pp. 1–39.

[10] M. Wozniak, M. Grana, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," Information Fussion, vol. 16, 2014, pp. 3–17.

[11] Z-H. Zhou, J. Wu, and W. Tang, "Ensembling neural networks: Many could be better than all," Artif. Intell. vol. 137, 2002, pp. 239–263.

[12] D. Z. Li, W. Wang, and F. Ismail, "A selective boosting technique for pattern classification," Neurocomputing, vol. 156, 2015, pp. 186–192.

[13] I. Ram, M. Elad, and I. Cohen, "Image processing using smooth ordering of its patches," IEEE Trans. on Image Processing, vol. 22, no. 7, July 2013, pp. 2764–2774.

[14] C. Dan, U. Meier, and J. Schmidhuber, "Multi-column deep neural network for image classification," IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 3642–3649.

[15] W. Li, M. Zeiler, S. Zhang, Y. LeCun, and R. Fergus, "Regularization of neural network using dropconnect," Journal of Machine Learning Research, vol. 28, no. 3, 2013, pp. 1058–1066.

[16] J. Z. Wang, "Semi-supervised learning using multiple one-dimensional embedding based adaptive interpolation," International Journal of Wavelets, Multiresolution and Information Processing, vol. 14, no. 2, 2016, pp. 1640002: 1–11.

[17] X. Zhu, Z. Ghahramani, and J. Lafferty, "Semi-supervised learning using gaussian fields and harmonic functions," in Proceedings of the 20th International Conference on Machine Learning, vol. 3, 2003, pp. 912–919.

[18] X. Zhou and M. Belkin, "Semi-supervised learning by higher order regularization," in Proceedings of the 14th International Cofference on Artificial Intelligence and Statistics, 2011, pp. 892–900.

[19] H. Luo et al., "Hyperspectral image classification based on spectral-spatial 1-dimensional manifold," IEEE Trans. Geosci. d Remote Sens., in press.

[20] Y. Wang, Y. Y. Tang, L. Li, and J. Z. Wang, "Face recognition via collaborative representation based multiple one-dimensional embedding," International Journal of Wavelets, Multiresolution and Information Processing, vol. 14, no. 2, 2016, pp. 1640003:1–15.

[21] Y. LeCun, C. Cortes, and C. J. C. Burges, "THE MNIST DATABASE of handwritten digits," http://yann.lecun.com/exdb/mnist/, accepted March 17, 2016.

[22] "USPS handwritten digit data," http://www.gaussianprocess.org/gpml/data/, accepted March 17, 2016.