SEMI-SUPERVISED FEATURE LEARNING FOR REMOTE SENSING IMAGE CLASSIFICATION

Xiaoshuang Yin[1], *Wen Yang*[1], *Gui-Song Xia*[2], *Lixia Dong*[1]

[1] School of Electronic Information, Wuhan University, Wuhan 430072, China
[2] Key State Laboratory LIESMARS, Wuhan University, Wuhan 430079, China

ABSTRACT

This paper presents a semi-supervised method for learning informative image representations, which is a crucial but challenging step for remote sensing image classification. More precisely, we propose to represent an image by projecting it onto an ensemble of prototype sets sampled from a Gaussian approximation of multiple feature spaces. Given a set of images with a few labeled ones, we first extract preliminary features, e.g. color and textures, to form a low-level image description. We then build an ensemble of informative prototype sets by exploiting these feature spaces with a Gaussian normal affinity. Discriminative functions are subsequently learned from the resulting prototype sets, and each image is represented by concatenating their projected values onto such prototypes for final classification. Experiments on two high-resolution remote sensing image sets demonstrate the efficiency of the proposed method on remote sensing image classification with different classifiers.

Index Terms— Semi-supervised feature learning, ensemble projection, remote sensing image classification

1. INTRODUCTION

Image classification plays an important role in the interpretation of remote sensing imagery. Popular classification techniques rely on labeled samples for training the classifiers, where the classification accuracy heavily depends on the amount and the quality of the training samples. However, the annotations of such labeled samples is usually time consuming and sometimes impossible to acquire in many real world problems. In order to enrich the information given as input to classification algorithm, alternative methods, such as active learning and semi-supervised learning [1] have been proposed, which jointly exploit labeled and unlabeled samples for training classifiers to improve classification accuracy.

This paper investigates the problem of semi-supervised feature learning for remote sensing image classification. In remote sensing literature, many efforts have been made to develop semi-supervised methods to improve the classification performance, most of which focus on regularizing the classification boundaries with unlabeled data [1]. For instance, Bruzzone *et al.* [2] introduced a transductive support vector machines (SVMs) for semi-supervised classification of remote sensing images, by exploiting both labeled and unlabeled samples. Recently, Dópido *et al.* [3] developed a semi-supervised remote sensing image classification method by intelligently selecting unlabeled samples using self-learning. Instead of optimizing classification boundaries, this paper proposes to learn an image representation in a semi-supervised way.

Our goal is to establish a discriminative image representation by exploiting a few labeled data and tremendous unlabeled ones. The idea is inspired by the work in [4], which introduced a unsupervised feature learning approach relying on ensemble projection. In particular, this paper proposes to represent an image by projecting it onto an ensemble of prototype sets sampled from a Gaussian approximation of multiple feature spaces, called *semi-supervised ensemble projection* (SSEP). It can improve the diversity and accuracy of the ensemble, which effects the classification performance a lot. For an image set with a few labels, we first extract preliminary features as low-level image descriptions. An ensemble of informative prototype sets are then constructed by exploiting feature spaces with a Gaussian normal affinity [5]. Each image is represented by concatenating the projected values onto such prototypes for final classification. In contrast with previous semi-supervised classification methods, our contributions are two-fold: (1) unlike [4], we learn features with a semi-supervised algorithm via ensemble projection, which makes better use of given training samples; (2) we use Gaussian normal affinity to approximate the feature space and develop a new sampling method which enables one to fast sample the prototypes in different feature spaces with a high accuracy. To our knowledge, some recent works close to us include [6], which presented a semi-supervised algorithm for non-linear feature extraction, and [7], who introduced a semi-supervised local discriminant analysis method for feature extraction. However, these methods are under the framework of multivariate analysis, which is very different from ours.

The rest of this paper is organized as follows. Section 2 presents the flowchart of our semi-supervised feature learning by SSEP. Section 3 presents and analyses experimental results. Conclusions are discussed in the last section.

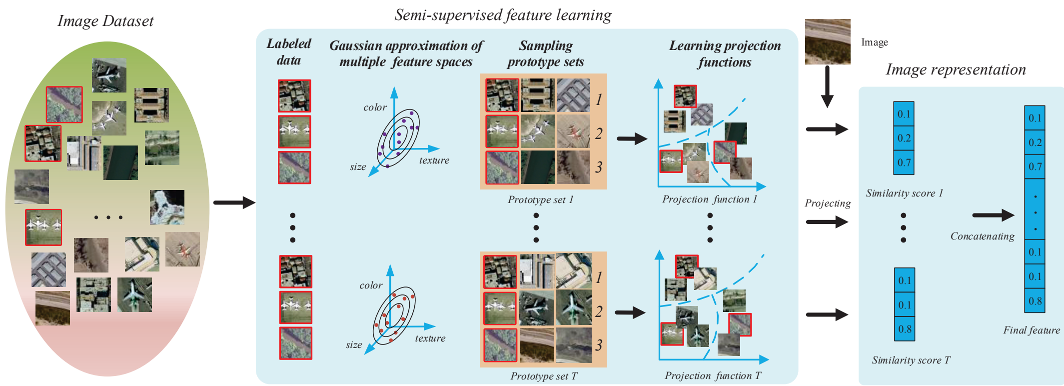


Fig. 1. The flowchart of SSEP. From a few labeled data (in red box) and unlabeled data, we build an ensemble of T prototype sets. Images are represented by multi-features and the sampling algorithm is performed in multi-feature spaces for great diversity. Each prototype set consists of all labeled data and their neighbors, which is then used to learn a discriminative classifier for projection function. Images are represented by concatenating their projection values on all the prototype sets.

2. SEMI-SUPERVISED FEATURE LEARNING

This section presents the flowchart of SSEP. Our method aims to learn a high-level image representation by exploiting a few labeled data, which is then fed into different classifiers to obtain final classification results. The process of semi-supervised feature learning by SSEP is shown in Fig. 1, and the details are as follows.

2.1. Sampling with Gaussian normal affinity

The training data is composed of both labeled data $D_l = \{x_i, y_i\}_{i=1}^{kl}$ and unlabeled data $D_u = \{x_j\}_{j=kl+1}^{kl+u}$, where x_i is the feature descriptor of image i, and $y_i \in \{1, ..., k\}$ is its label. k is the number of categories, l is the number of labeled data in each category and u is the number of unlabeled data.

With the labeled data, a new sampling algorithm based on Gaussian normal affinity is proposed to produce T prototype sets $P^{t, t\in\{1,...,T\}} = \{(s_i^t, c_i^t)\}_{i=1}^{pk}$ with great diversity from all data, where $s_i^t = \{1, ..., kl+u\}$ is the index of the ith chosen image, $c_i^t \in \{1, ..., k\}$ is its label and p is number of images selected for each class. As mentioned before, the sampling algorithm is performed in different feature spaces for great diversity. Therefore, there are multiple feature descriptors $x_i = \{x_i^1, ..., x_i^r\}$, where r is the number of feature spaces (e.g. $k = 3$, $l = 1$, $p = 3$ and $r = 3$ in Fig. 1). The algorithm for creating a single P^t in the t trial is given in Alg. 1.

Specifically, each prototype set consists of all labeled data and their neighbors. For each category, l labeled images are served as seeds and the sampling procedure is as follows :

(1) In each feature space, we first find n neighbors of each seed, resulting in ln neighbors in total. Then m neighbors are randomly selected.

(2) Second, the enriched training data for each category is obtained by concatenating the randomly selected neighbors in all feature spaces.

Therefore, the training data for each category has p ($p = mr+l$) images. The final prototype set is constructed by concatenating the training data in all categories. The sampling algorithm is repeated T times to construct the ensemble. Being different from [4], our sampling algorithm is performed in different feature spaces with exploiting the labeled data, which increase both the reliability and diversity of the ensemble. Moreover, we introduce Gaussian normal affinity instead of KNN, which can be used to find neighbors efficiently in a high-dimensional feature space.

As data distributed in a high-dimensional feature space, it is difficult and inefficient to find their nearest neighbors. For efficiency, we employ Gaussian normal affinity [5] to find neighbors of the given image. For a given image x, the normal to the Gaussian at x is computed as

$$w_x = \Sigma^{-1}(x - \mu) \tag{1}$$

where $\Sigma = \frac{1}{kl+u}\sum_i (x_i - \mu)(x_i - \mu)^T$, $\mu = \frac{1}{kl+u}\sum_i x_i$ and $kl + u$ is the number of images. The normal is then used to rank the similarities between x and other images. Therefore, neighbors of the given image are obtained by Gaussian normal affinity.

2.2. Ensemble projection (EP)

Ensemble learning is a key step of EP, which requires the ensemble should be both diverse and accurate [8]. As mentioned above, the sampling algorithm ensures the diversity of

Algorithm 1: Semi-supervised sampling with Gaussian normal affinity in tth trial

input : Dataset D;
  Labeled training data $D_l = \{x_i, y_i\}_{i=1}^{kl}$.
output: Prototype set P^t.

Find corresponding training image indexes in D, i.e. $V = \{kl\}$ image index;
for $i \leftarrow 1$ **to** k **do**
  $V_i = \{l$ image index of class $i\}$;
  for $f \leftarrow 1$ **to** r **do**
    $s =$ indexes of the ln nearest neighbors of data in V_i in feature space f;
    $s_{i,f}^t =$ randomly select m indexes from s;
  $s_i^t = (s_{i,1}^t, ..., s_{i,r}^t) \cup V_i$;
  $c_i^t = (i, i, ..., i) \in \mathcal{R}^{mr+l}$;
let $p = mr + l$ then
$s^t = \{s_1^t, ...s_k^t\} \in \mathcal{R}^{kp}$, $c^t = \{c_1^t, ...c_k^t\} \in \mathcal{R}^{kp}$;
$P^t = \{(s_i^t, c_i^t)\}_{i=1}^{kp}$.

the ensemble. For good precision, EP introduces discriminative learning method that is logistic regression as the base learner. For an input image x, each base learner projects x (i.e. classify it), resulting a similarity vector

$$S_i = (S_{i,1}, ..., S_{i,k}) \tag{2}$$

where $i \in \{1, ..., T\}$, element $S_{i,c}$ ($c \in \{1, ..., k\}$) measures the probability of x belonging to category c using base learner i. A new image representation $S = [S_1, S_2, ..., S_T]$ is obtained by concatenating all the similarity scores for the final classification. Therefore, the dimension of S is $k \times T$.

3. EXPERIMENTAL RESULTS

3.1. Datasets and Experimental Settings

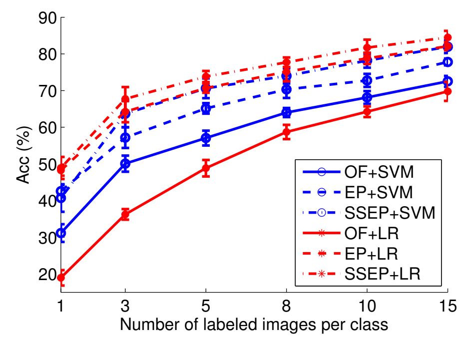
We evaluate our method on two high-resolution satellite datasets: (1) a 19-class satellite scene dataset[1]: 19 categories and each of them has 50 images, with size of 600×600 pixels; (2) a 21-class satellite scene dataset[2]: 21 categories and it has 100 images for each , with size of 256×256 pixels. Three feature descriptors were used in our implementation, scale invariant features transform (SIFT), combined scattering (CS) and bag of colors (BOC).

To investigate the classification performance of our feature learned by SSEP, the original feature (OF, obtained by the concatenation of SIFT, CS and BOC) and feature learned by EP were adopted for comparison. Two traditional supervised classifiers were adopted, logistic regression (LR) and linear support vector machines (SVMs). We used L2-regularized LR of LIBLINEAR with $C = 15$ and the linear SVMs of LIBSVM with $C = 15$. As to the parameters of our method, we used the following for experimental sets, $r = 3$, $m = 8$, $n = 10$ and $T = 300$. Overall accuracy (Acc) was used as the evaluation criteria.

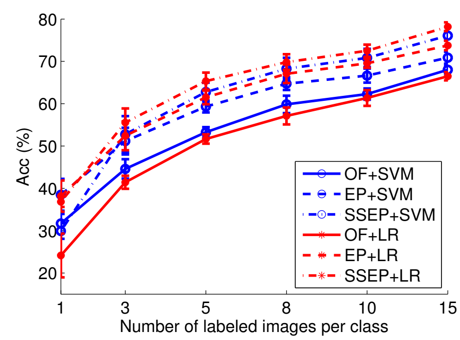
[1]data available at http://dsp.whu.edu.cn/cn/staff/yw/HRSscene.html.
[2]data available at http://vision.ucmerced.edu/datasets/landuse.html.

3.2. Classification results with supervised classifiers

The dimensions of the SIFT, CS and BOC are 512, 595 and 512 respectively for both datasets. The dimensions of feature learned by EP and SSEP are the same, 5700 for 19-class dataset and 6300 for 21-class dataset. For each dataset, we randomly selected $\{1, 3, 5, 8, 10, 15\}$ images per category as labeled training data. The rest images were used for testing. Five-fold experiments were performed and the mean and standard deviation for Acc were recorded.



(a) 19-class satellite scene dataset



(b) 21-class satellite scene dataset

Fig. 2. Classification results with different features.

Fig. 2 shows all the results and Tab. 1 and Tab. 2 shows the classification results with 5 training images per class, from which we can see our feature learned by SSEP has the best performance. The feature learned by EP and SSEP performs better than OF, which demonstrates the effectiveness of ensemble projection. LR working with our feature performs better than SVM, obtaining the best classification results. In general, the feature learned by SSEP combines well with SVM and LR, and obtains promising classification results on high-resolution satellite datasets.

3.3. Classification results with semi-supervised classifiers

To illustrate the applicability of our feature, we further combine our feature with semi-supervised classifiers. Except SVM and LR, two more semi-supervised classifiers LSVM [9] and meanS3VM [10] were adopted. For LSVM, we used the the scheme suggested by [9] with means of Newtons method, $\gamma_I = 1$, $\gamma_A = 1 \times 10^{-5}$. For meanS3VM, $C1 = 1$, $C2 = 0.1$, alternating optimization and Gaussian kernel was used. We used the 5 random selected images in Sec. 3.2 as training data for each class. The rest images were used for testing. Five-fold experiments were performed and the mean and standard deviation for Acc were recorded. Note that when other number of training data are used, we get the same trends for LSVM and meanS3VM as for SVM and LR (shown in Fig. 2).

Table 1. Acc (%) of classification on 19-class satellite scene dataset, with 5 training data per class.

	SVM	LR	LSVM	meanS3VM
OF	57.05(2.01)	48.84(2.25)	50.13(1.17)	61.15(2.09)
EP	65.14(1.42)	70.66 (1.26)	60.63(1.01)	66.92 (1.07)
SSEP	70.61(2.69)	**73.82 (1.52)**	72.26(1.05)	73.75(0.87)

Table 2. Acc (%) of classification on 21-class satellite scene dataset, with 5 training data per class.

	SVM	LR	LSVM	meanS3VM
OF	53.26(1.20)	51.66(1.09)	49.97(2.51)	55.31(1.25)
EP	59.31(1.39)	61.50(1.62)	58.92(3.24)	60.47(2.25)
SSEP	62.76(2.21)	65.34(2.01)	**66.49(2.18)**	66.29(1.58)

Tab. 1 and Tab. 2 show the classification results with 5 labeled images on both datasets, where the best results are shown in bold black. Among different features, OF has the worst result and feature learned by SSEP has the best performance as in 3.2. As we know, LSVM and meanS3VM exploit both labeled and unlabeled data when classify images and their performances are supposed to be better than plain classifiers. However, with our feature, their classification performance are worse than LR in 19-class satellite scene dataset. Both LR and SVM can obtain comparable classification results to semi-supervised classifiers with our feature. This suggests our semi-supervised feature learning and other semi-supervised classifiers are marginally complementary, and nearly all knowledge in unlabeled data are already exploited by our feature learning. Therefore, even plain classifiers (LR and SVM) can obtain promising results with our feature. The semi-supervised experimental results further confirm the superiority of our feature learned by SSEP.

4. CONCLUSION

This paper addressed the task of semi-supervised feature learning using ensemble projection. A new category-level image representation for high-resolution satellite images was proposed. To create an ensemble of diverse prototype sets, a novel sampling algorithm is designed and performed in different feature spaces. As data distributed in a high-dimensional feature space can be approximated by the covariance matrix of the data, we introduce Gaussian normal affinity to find neighbors in the sampling algorithm. Then discriminative classifiers are learned on these prototype sets, and images are represented by concatenating their projection values on the prototype sets. Results on two challenging datasets demonstrate the superiority of our feature learned by SSEP.

5. ACKNOWLEDGEMENT

The research was supported by the Chinese National Natural Sciences Foundation[Grant no. 61271401 and no.91338113].

6. REFERENCES

[1] L. Bruzzone and C. Persello, “Recent trends in classification of remote sensing data: active and semisupervised machine learning paradigms,” in *Proc. IGARSS*, 2010, pp. 3720–3723.

[2] L. Bruzzone, M. Chi, and M. Marconcini, “A novel transductive svm for semisupervised classification of remote-sensing images,” *IEEE T. Geoscience and Remote Sensing*, vol. 44, no. 11-2, pp. 3363–3373, 2006.

[3] I. Dópido, J. Li, P. R. Marpu, A. Plaza, J. Bioucas-Dias, and J. Benediktsson, “Semisupervised self-learning for hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7-1, pp. 4032–4044, 2013.

[4] D. Dai and L. V. Gool, “Ensemble projection for semi-supervised image classification,” in *Proc. ICCV*, 2013.

[5] M. Gharbi, T. Malisiewicz, S. Paris, and F. Durand, “A gaussian approximation of feature space for fast image similarity,” *Technical report, Massachusetts Institute of Technology*, 2012.

[6] E. Izquierdo-Verdiguier, L. Gómez-Chova, L. Bruzzone, and G. Camps-Valls, “Semisupervised nonlinear feature extraction for image classification,” in *Proc. IGARSS*, 2012.

[7] W. Liao, A. Pizurica, P. Scheunders, W. Philips, and Y. Pi, “Semi-supervised local discriminant analysis for feature extraction in hyperspectral images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 184–198, 2013.

[8] D. Dai, M. Prasad, C. Leistner, and L. Van Gool, “Ensemble partitioning for unsupervised image categorization,” in *Computer Vision–ECCV 2012*, 2012, pp. 483–496.

[9] S. Melacci and M. Belkin, “Laplacian support vector machines trained in the primal,” *J Mach Learn Res*, vol. 12, pp. 1149–1184, 2011.

[10] Y. F. Li, J. T. Kwok, and Z. H. Zhou, “Semi-supervised learning using label mean,” in *Proc. ICML*, 2009, pp. 633–640.