Computer Science and Information Systems 17(2):427–443 https://doi.org/10.2298/CSIS191222010Z

Land-use classification via ensemble dropout information discriminative extreme learning machine based on deep convolution feature

Tianle Zhang[1], Muzhou Hou[1], Tao Zhou[2*], Zhaode Liu[2], Weirong Cheng[3], and Yangjin Cheng[4]

[1] School of Mathematics and Statistics, Central South University
410083 Changsha, China
houmuzhou@sina.com, csuztl@csu.edu.cn
[2] School of Economics, Guangdong University of Finance and Economics
510320,Guangzhou, China
zhoutaoscut@hotmail.com*, lzhaode@163.com
*Corresponding author
[3] School of Mathematics and Science, Peking University
100871 Beijing, China
chenweirong@pku.edu.cn
[4] School of Mathematics and Computational ScienceXiangtan University
411105, Xiangtan, Hunan, China
yjcheng@xtu.edu.cn

Abstract. Classifying land-use scenes with high quality and accuracy is an important research direction in current hyperspectral remote sensing images, which is conducive to scientific management and utilization of land. An effective classifier and feature extractor can improve classification stability and accuracy. Therefore, based on deep learning technique, a dropout-based ensemble learning method is proposed in this paper, which combines convolutional neural network (CNN) and information discriminating extreme learning machine (IELM). Pre-trained CNN is used to learn effective and robust features, and deep convolution features are fed to the IELM classifier. Then the adoption of dropout technique and ensemble method can improve generalization capabilities and stability. The effectiveness of the proposed algorithm is tested by hyperspectral remote sensing image classification experiments. The experimental results show that the proposed E-CNN-dropIELM has achieved satisfactory results compared to state-of-the-art methods in terms of classification accuracy and stability.

Keywords: Extreme learning machine (ELM), Convolutional neural network (CNN), Dropout, Ensemble, Land-use classification.

1. Introduction

Nowadays remote sensing imaging technology and deep learning have developed rapidly, and now high-resolution remote sensing (HRRS) images have become a hot research field. Its most important research goal is to classify high-resolution remote sensing images, for instance, agricultural, airplane, beach, etc. It is of great importance in many remote sensing applications, such as land resource management, urban development and nature conservation. However, there are lots of challenges in the land-use scenes classification such as insufficient training sample data, extracting effective classification features, and improving classifier accuracy.

In recent decades, in order to overcome these problems, a large number of previous work use machine learning methods to improve the accuracy of land use classification, such as support vector machine (SVM) [28], manifold learning (ML) [2], random forests (RF) [24], artificial neural network (ANN) [11], unsupervised learning (UL), etc. Recently, many methods have made very good progress in the field of remote sensing image classification. Most of the previous works mainly focused on classifying by low-level features. Lowe extracted distinctive invariant features from images [16]. Oliva, et. used the gist descriptor to achieve probable semantic category of scenes [18]. Ren, et. proposed a novel scheme for local binary patterns (LBP) to tackle pixel correlation [23]. Although these low-level features have proven their validity for scene classification, these features are not ideal due to the inability to reflect semantic information [18,23,37]. Many researchers have developed statistical methods to process shallow features for mid-level features. Bag-of-visual-words (BOVWs) model is the most famous approach, which has been great successful in scene classification [33,38]. Of course, there are some other ways to get the semantic information of the images, such as locality-constrained linear coding (LLC) [15], vector of locally aggregated descriptors (VLAD) [10], improved fisher kernel (IFK) [22], et. However, the semantic information of hand-crafted features is insufficient. In the field of image recognition, convolutional neural network (CNN) as one of deep learning models has achieved successful achievements and rapid development [26,35,12,13].

Recently, CNN which can extract discriminative and robust features, has been applied to land-use image classification [8,19]. Features extracted from the pretrained CNN show outstanding performance in scene classification[21,1]. So instead of using these CNN architectures directly as the final classifiers, many researchers use pre-trained CNNs as feature extractors and combined it with effective and simple classifiers. The Extreme Learning Machine (ELM) is a powerful learning algorithm based on single hidden layer feedforward neural networks (SLFNs). It obtains the output weight by solving the least squares solution of the SLFNs while the input weight of the hidden node is randomly generated. At the same time the ELM can avoid the possibility of slow convergence and local optimal solution of back propagation algorithm (BP). Because of its very efficient and impressive approximation and generalization capabilities, ELM has been applied to many areas [6,34,36,17,27].ELM is applied to land cover classification as a classifier [20,14,29]. ELM achieves better classification results compared with BP and SVM, and its computational time is much smaller than BP and SVM. Yan, et. proposed information discriminative extreme learning machine (IELM) to overcome the shortcoming that ELM is in insufficient with limited hyperspectral remote sensing images [31]. Qian, et. proposed a model for land-use scene classification by integrating CNN and constrained extreme learning machine (CELM) [30]. Zhu, et. integrated CNN and KELM, and KELM is used to enhance the discriminative ability of classifier by kernel function [39]. While the effectiveness of such features and classifiers have been verified, some papers consider ensemble methods. Actually, it is able to provide a more effective way to solve the land-use classification problem. Although, the ensemble algorithm is very effective in the general pattern recognition tasks[5,3,4], it is still a challenging task to land use classification in practical applications. In order to make full use of effective deep convolution features and overcome the limitations of traditional classifiers and fully connected layer (FCL), a new framework for land-use scene classification is proposed in this paper. First, CNN is trained with a large land-use scene image dataset. The pre-trained CNN removing FCL is used as a feature extractor to learn deep and discriminative features. Then, the IELM-based basic classifiers with dropout technology are built, and excellent classification result is obtained via majority voting. To sum up, the following contributions have been made in this paper:

i A new framework for land-use scene classification is present combining deep convolutional features by CNN with ensemble classifier via majority voting.
ii By introducing the dropout method, IELM-based basic classifiers with different subsets of discriminative features are constructed. Such a classification framework has better generalization ability and can be effectively for extensive applications.
iii The experimental results on the remote scene image land-use classification show that E-CNN-dropIELM can improve the classification performance compared to the several the-state-of-art techniques.

The rest of this paper is organized as follows. Section 2-5 introduces the related background knowledge of IELM and CNN, et. Section 6 describes in detail the proposed E-CNN-dropIELM approach. Experiments and the results are presented in Section 7, while some concluding remarks are drawn in Section 8.

2. The pretrained CNN

CNN is considered to be one of the most successful structures for deep learning due to its good performance. A typical CNN consists of two parts. One is the feature extraction part, which is usually alternately connected to the pooling layers by alternating convolution layers. The output of the upper layer is the input of the next layer. The shallow layer extracts relatively specific features, and the deep layer often extracts relatively abstract features by features learned in previous layers. The other part is the classifier, fully-connected layer. And a typical feed-forward neural network uses a Softmax layer as a classifier to calculate the likelihood of each class. Parameters of CNN are usually optimized through a classical stochastic gradient descent algorithm based on the backpropagation. A typical CNN architecture for land-use scene classification in Fig. 1.

The deep convolution features learned by the pre-training CNN are effective for the classification of land use scenarios, referring to the previous papers [7]. Therefore, the pre-training architecture of CNN provided by MatConvNet toolbox is used, which is pre-trained on Imagenet [25] and has similar structure with AlexNet. The architecture used in this paper is shown in Fig. 2, and it consists of five convolution layers and three fully-connected layer. One pooling layer is connected behind each convolution layer. We remove full-connection layers and use the remaining CNNs as a feature extractor.

In order to achieve better performance in a small sample dataset, this paper uses a series of random transforms to augment images for obtaining better features to increasing accuracy. The formulas are as follows,

$$\mathbf{I}' = s(\lambda)\mathbf{R}(\theta)\mathbf{I} + \mathbf{T}(\delta) \tag{1}$$

where $\mathbf{I}$ is the input image matrix. $s(\lambda)$ is the scaling factor. $\mathbf{R}(\theta)$ is the rotation matrix. $\mathbf{T}(\delta)$ is the shift matrix. Then bilinear interpolation is used to resize the image.

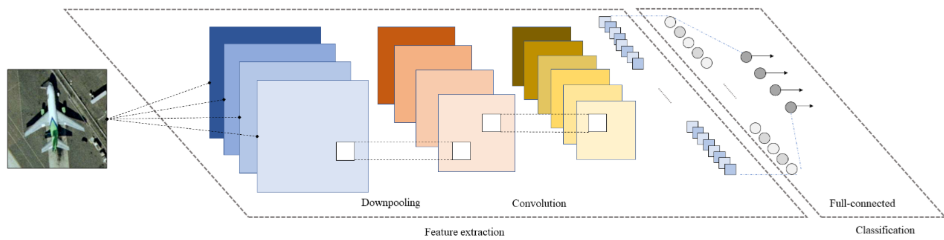

Fig. 1. A typical CNN architecture for land-use scene classification.

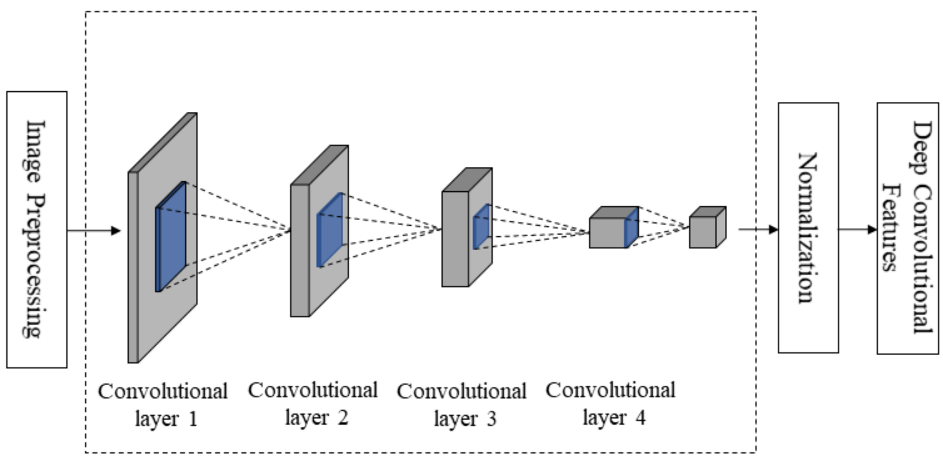

Fig. 2. The CNN architecture used in this paper, and the max-pool layers are not displayed.

3. Dropout method

Dropout means that neurons are randomly removed during network training, and it is an effective way to improve network's generalization capabilities. As shown in Fig. 3, removing a neuron equal to drop it temporarily from the network with all its input and output connections. Choosing which neuron to be removed is random, and usually each neuron is preserved with a fixed probability independent of other neurons. We introduce the basic idea of dropout by linear neurons. Given an input vector $\mathbf{I} = (I_1, I_2, I_3, \cdots, I_n)$

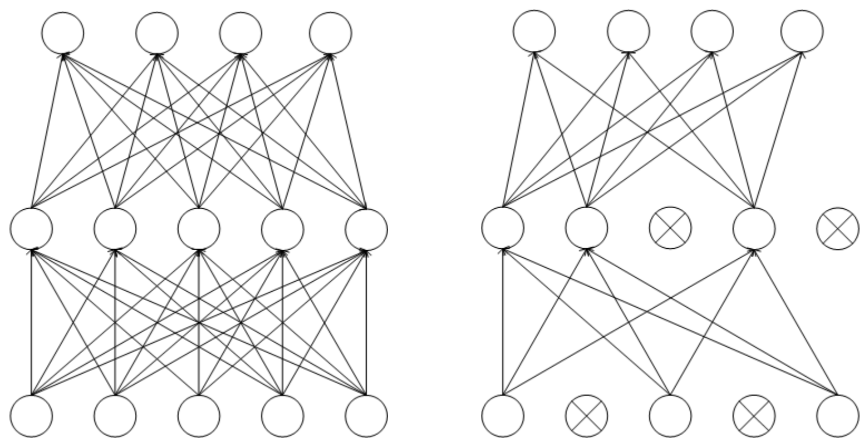

Fig. 3. The diagram of basic idea of the proposed algorithm.

, the output of the linear neuron is $S(\mathbf{I}) = \sum_{i=1}^{n} \omega_i I_i$. If I_i is deleted with a uniform distribution, or deleted with an equal probability of 0.5, $2n$ sub-networks will be generated including the empty network. The average of outputs of all sub-networks is:

$$E(S) = \frac{1}{2^n} \sum S(\omega_i, I) \tag{2}$$

where $S(\omega_i, I)$ means a set all sub-networks.

Suppose $\delta_i (1 \leq i \leq n)$ is a stochastic variable satisfying Bernoulli distribution and they are independent with each other. $p_i = P(\delta_i = 1), q_i = P\left(\delta_{(i)} = 0\right) = 1 - p_i$. If remove $I_i (1 \leq i \leq n)$ or $\omega_i (1 \leq i \leq n)$ with probability q_i , then the output of the linear element can be expressed as follow:

$$S(\mathbf{I}) = \sum_{i=1}^{n} \omega_i \delta_i I_i \tag{3}$$

According to the linear property by mathematical expectation,

$$E(S) = \sum_{i=1}^{n} \omega_i E(\delta_i) I_i = \sum_{i=1}^{n} \omega_i p_i I_i \tag{4}$$

If $p_1 = p_2 = \cdots = p_n = p$,Eq. (4) can be written as

$$E(S) = \sum_{i=1}^{n} \omega_i E(\delta) I_i = \sum_{i=1}^{n} \omega_i p I_i \tag{5}$$

If there is a bias in the linear neuron, output of the linear neuron can be written as:

$$S(\mathbf{I}) = \sum_{i=1}^{n} \omega_i \delta_i I_i + b\delta_b \tag{6}$$

Similarly,

$$E(S) = \sum_{i=1}^{n} \omega_i p_i I_i + b p_b \tag{7}$$

where $p_b = P\left(\delta_b = 1\right)$.

4. ELM algorithm

The extreme learning machine is a novel learning algorithm based on the structure of single-hidden layer feedforward networks (SLFNs) that was proposed by G B Huang et al. in 2006 and then extended to "generalized" SLFNs where the hidden layer neurons need not be neuron alike [9]. The hidden neuron's parameters are randomly generated without iterative tuning and are independent of the data. The network topology of ELM is shown in Fig. 4.

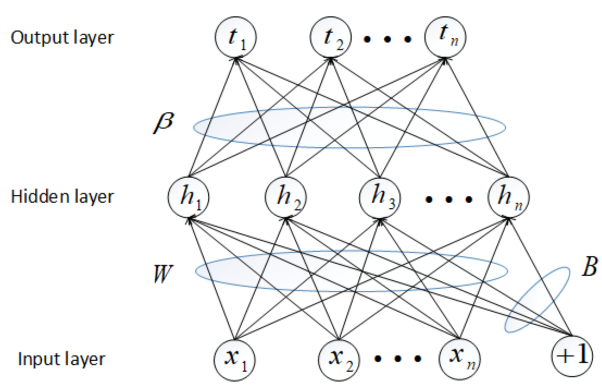

Fig. 4. The network topology of ELM.

Given a training set with N distinct samples $\Gamma = \{(\mathbf{x}_j, \mathbf{t}_j) \,|\mathbf{x}_j \in \mathbb{R}^n, \mathbf{t}_j \in \mathbb{R}^m,$ $\mathbf{j} = 1, \ldots, N\}$, where X is the predictive variable with dimension n and t is the objective variable with dimension m. The mathematical model of SLFNs with M hidden neurons and an activation function $G(\cdot)$ are described as

$$\sum_{i=1}^{n} \beta_i G\left(\mathbf{w_i}, b_i, \mathbf{x_j}\right), j = 1, 2, \ldots, N \tag{8}$$

where $b_i \in R$ is the randomly assigned bias of the ith hidden node and $\mathbf{w_i} \in R$ is the randomly assigned input weight vector connecting the ith hidden node to the output node. $G(\mathbf{w_i}, b_i, \mathbf{x_j})$ is the output of the ith hidden node with respect to the input sample $\mathbf{x_j}$.

When the SLFN is completely approximating the data, that is, when the error between the output $\hat{t}_i$ and the actual t_i is zero, the relationship is:

$$\sum_{i=1}^{n} \beta_i G(\mathbf{w}_i, b_i, \mathbf{x}_j) = \mathrm{t}_j, j = 1, 2, \ldots, N \tag{9}$$

Eq. (9) can be written as

$$\mathbf{H}\beta = \mathbf{T} \tag{10}$$

where

$$\mathbf{H} = \left[\mathbf{h}_1^T\ \mathbf{h}_2^T \ldots \mathbf{h}_N^T\right] = \begin{bmatrix} G(\mathbf{w}_1, b_1, \mathbf{x}_1) & \cdots & G(\mathbf{w}_n, b_n, \mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ G(\mathbf{w}_1, b_1, \mathbf{x}_N) & \cdots & G(\mathbf{w}_n, b_n, \mathbf{x}_N) \end{bmatrix}_{N \times n} \tag{11}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \beta_2^T \\ \vdots \\ \beta_N^T \end{bmatrix}_{n \times m}, T = \begin{bmatrix} t_1^T \\ t_2^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m} \tag{12}$$

The output weights can be computed by finding the least-square solutions of the SLFNs in Eq. (10), which is given as

$$\beta = \mathbf{H}^{\dagger}\mathbf{T} \tag{13}$$

where $\mathbf{H}^{\dagger}$ is the Moore-Penrose generalized inverse of the matrix $\mathbf{H}$. If $\mathbf{H}^{\dagger}\mathbf{H}$ is not a singular matrix, then, Eq. (13) can be written as

$$\beta = \mathbf{H}^{\dagger}\mathbf{T} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{T} \tag{14}$$

In addition, we add an l-2 regularization constraint to the loss function of ELM which can improve the generalization and robustness. Eq. (14) becomes the following formula:

$$\beta = \begin{cases} \mathbf{H}^T\left(\frac{I}{C} + \mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{T} & \text{if } N \leq L \\ \left(\frac{I}{C} + \mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{T} & \text{if } N > L \end{cases} \tag{15}$$

where C assigns the regularization, I assigns the unit matrix, N assigns the number of samples and L assigns the number of hidden layer neurons.

5. IELM algorithm

The paper [32] proposed a kind of regularized extreme learning machine based on discriminative information (IELM). For the classification problem, IELM considers the geometric features and the discriminant information of data samples, and optimizes the output weight of the extreme learning machine by maximizing the heterogeneous dispersion and minimizing the same kind of dispersion. It can improve accuracy and generalization of the traditional ELM.

Given N distinct samples $\Gamma = \{(\mathbf{x}_j, \mathbf{t}_j) | \mathbf{x}_j \in \mathbb{R}^n, \mathbf{t}_j \in \mathbb{R}^m, \mathbf{j} = 1, \ldots, N\}$, S_B is the inter-class scatter matrix, S_W is the within-class scatter matrix. They are described as:

$$S_B = \sum_{i=1}^{C} \sum_{j=1}^{N_c} \left(x_j - u^i\right) \left(x_j - u^i\right)^T \tag{16}$$

$$S_W = \sum_{i=1}^{C} \left(u^i - u\right) \left(u^i - u\right)^T \tag{17}$$

where C is the number of categories of data samples, u_i is the mean of ith class, and u is the mean of data samples. Formation difference matrix can be described as:

$$S = S_B - (1 - \eta) S_W \tag{18}$$

where η is a parameter from 0 to 1. The parameter η plays the role of adjusting the intra-class and the within-class discriminant information. So IELM makes good use of the discriminant information in the data and improves the classification ability of ELM.

The loss function of IELM is described as:

$$\begin{aligned} &\min_{\beta} \frac{1}{2} \beta^T S \beta + \frac{1}{2} C \sum_{i=1}^{N} \varepsilon^2 \\ &\text{s.t. } \sum_{i=1}^{n} \beta_i G\left(\mathbf{w}_i, b_i, \mathbf{x}_j\right) - \mathrm{t}_j = \varepsilon_i, j = 1, 2, \ldots, N \end{aligned} \tag{19}$$

where $\mathcal{E}$ assigns training error.

The Lagrange function of Eq. (19) is:

$$L = \frac{1}{2} \beta^T S \beta + \frac{1}{2} C \sum_{i=1}^{N} \varepsilon^2 + \sum_{i=1}^{N} \sum_{j=1}^{m} \left(\alpha_{ij} \beta_j G\left(\mathbf{x}_i\right) - \mathbf{t}_{ij} + \varepsilon_{ij}\right) \tag{20}$$

where β_j is the output weight of jth output node, and the KKT condition is:

$$\frac{\partial L}{\partial \beta_j} = 0 \rightarrow \beta_j = \sum \alpha_{ij} G\left(x_i\right)^T \rightarrow WS = H^T \alpha \tag{21}$$

$$\frac{\partial L}{\partial \varepsilon_j} = 0 \rightarrow \alpha_i = C \varepsilon_i, i = 1, \ldots, N \tag{22}$$

$$\frac{\partial L}{\partial \alpha_j} = 0 \rightarrow G\left(x_i\right) \beta - t_i^T + \varepsilon_i^T = 0 \tag{23}$$

where $\alpha_i = \left[\alpha_{i1}, \ldots, \alpha_{im}\right]^T, \alpha = \left[\alpha_1, \ldots, \alpha_n\right]$

According to the above, Eq. (15) becomes the following formula:

$$\beta = \begin{cases} \mathbf{h}^r \left(\frac{S}{C} + \mathbf{h}^r\mathbf{h}\right)^{-1} S\mathbf{T} & \text{if } N \le L \\ \left(\frac{S}{C} + \mathbf{H}^T\mathbf{h}\right)^{-1} \mathbf{h}^T\mathbf{T} & \text{if } N > L \end{cases}$$

And the output of IELM is:

$$\mathrm{g}(x) = \mathrm{G}(x)\beta = \begin{cases} \mathrm{G}(x) \left(\mathbf{H}^T \left(\frac{S}{C} + \mathbf{H}^T\mathbf{H}\right)^{-1} S\mathbf{T}\right) & \text{if } N \le L \\ \mathrm{G}(x) \left(\frac{S}{C} + \mathbf{H}^T\mathbf{H}\right)^{-1} \mathbf{H}^T\mathbf{T} & \text{if } N > L \end{cases}$$

6. The proposed E-CNN-dropIELM algorithm

In this section, we introduce the basic idea of the proposed E-CNN-dropIELM. The accuracy of image classification is mainly affected by two parts, feature extractor and classifier. Firstly, this method uses CNN as a feature extractor, which is necessary because effective features can improve the performance of the classification. And then add dropout method to IELM, which considers the inter-class distribution and discriminant information of the data. And neurons are randomly removed, so the IELM is changing for each training. These trained IELM networks are used as basic classifiers by repeatedly “dropout” the neurons. Therefore, all basic classifiers work on different feature subspaces. Finally, these basic classifiers used for data classification are ensembled by simple majority voting method, and it can effectively avoid over-fitting of a local feature by the network.

The diversity and complementarity of basic classifiers have a great influence on the performance of the ensembled classification. Diversity is the error diversity of basic classifiers, which means the samples misclassified by basic classifiers should be different. If the samples classified by each basic classifier are different, the diversity of basic classifiers is best. Obviously, such basic classifiers are also the best complementarity. In other words, classifiers can learn from each other when classifying the sample. The typical method of constructing basic classifiers is to divide the training set into several subsets of data, and then train basic classifiers with these subsets of data. Because these basic classifiers are trained with different subsets of data, basic classifiers trained in this way have better diversity. In this paper, the method of constructing basic classifiers is different from the typical method, and use the Dropout technique to construct the basic classifier, IELM. These basic classifiers are trained with the same dataset, but the network topology and feature subspace of basic classifiers are different. Naturally, when classifying the samples, they have good diversity and can better learn from each other.

The proposed E-CNN-dropIELM algorithm can be described as follow:

7. Remote sensing image land-use classification experiments

7.1. The UC-Merced land use Dataset

In this paper, we investigate the performance of the proposed E-CNN-dropIELM on the UC-Merced (UCM) data set [33]. The data set contains high-resolution remote sensing images of 21 different land-use classes, such as airplanes, beaches, etc. Each class has

Algorithm 1 Ensemble dropout information discriminative extreme learning machine based on deep convolutional features for image classification.

Input:
$\Gamma = \{(\mathbf{x}_j, \mathbf{t}_j) \,|\mathbf{x}_j \in \mathbb{R}^n, \mathbf{t}_j \in \mathbb{R}^m, \mathbf{j} = 1, \ldots, N\}$, training set; T, testing set; l, the number of basic classifiers; C, regularization parameter; η, discriminant information parameter

Output:
j^*, the class label of $\mathbf{x} \in T$

Extract feature via a fix CNN removed the last layer softmax;
Initialize a larger IELM with m hidden nodes;
// Train l basic classifiers: $L = L_1, L_1, \cdots, L_i$;
for all $(i = 1; i \le l; i = i + 1)$ **do**
  Generate a m-dimensional binary random vector whose elements are drown independently from a Bernoulli distribution;
  Dropout some hidden nodes with the generated m-dimensional binary random vector, obtain a IELM$_i$ denoted by L_i;
  Train L_i with IELM and soft-maximize its outputs, obtain a probability distribution $(p_{i1}(\mathbf{x}), p_{i2}(\mathbf{x}), \ldots, p_{ik}(\mathbf{x}))$;
end for
// Test $\mathbf{x} \in T$ via basic classifiers: $L = L_1, L_1, \cdots, L_i$;
for all $(\mathbf{x} \in T)$ **do**
  Extract feature via a fix CNN removed the last layer softmax;
  for all $(i = 1; j \le l; j = j + 1)$ **do**
    Calculate label vector $(p_{i1}(\mathbf{x}), p_{i2}(\mathbf{x}), \ldots, p_{ik}(\mathbf{x}))$ via L_i with IELM;
  end for
  Calculate $p_{j^*}(\mathbf{x}) = \max_{1 \le j \le k} \sum_{i=1}^{l} p_{ij}(\mathbf{x})$ via majority voting;
end for
return j^*;

100 images and their size is 256 x 256 pixels. The sample images of 12 classes in UCM is shown in Fig. 5.

Fig. 5. Image examples from UCM land use data set.

All samples were resized to 227 x 227 pixels because input image of the pre-trained CNN requires. 4096-dimensional vector is fed into basic classifiers IELMs as feature.

7.2. Experiment and results

In order to evaluate the performance of the proposed algorithm, we compare it with the classical algorithms such as SVM, ELM and some successful algorithms in mentioned papers in UCM dataset [37,35,21,29,7]. In this paper, all the experiments were performed 30 times using MATLAB R2017b on a computer with 2.60 GHz CPU and 16.0 GB RAM.

In this experiment, 70% of each class is used as training set and the rest is used as training set. We average accuracies of 30 times experiments on testing set and use grid search method to obtain the optimal values. The optimal regularization parameter C is 2^{11} by employing $C = 2^x, x = -5, \ldots, 30$, and the optimal discrimination information parameter η and the optimal dropout probability $p$are 0.5 and 0.9 from a set $\{0, 0.1, \ldots, 1\}$. Impacts of the proposed E-CNN-dropIELM by different factors are shown in Fig. 6.

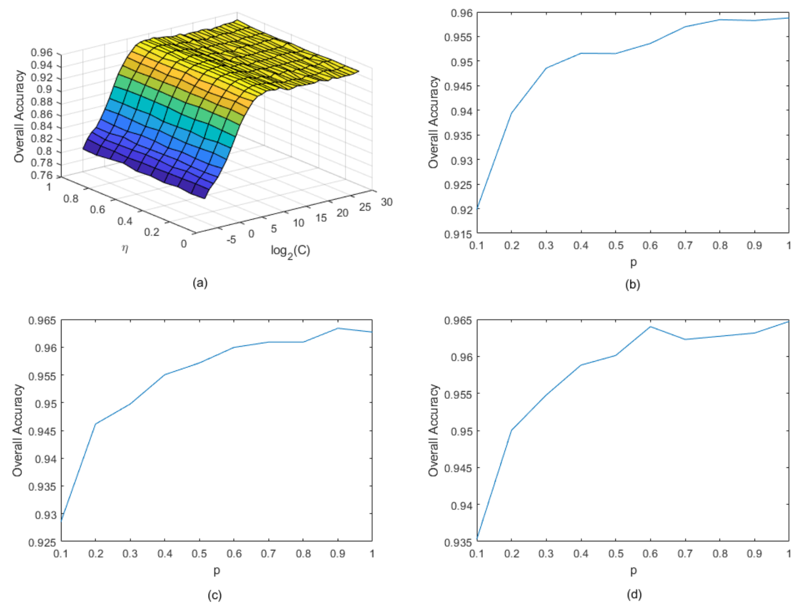

Fig. 6. Impacts on OA of E-CNN-dropIELM by different factors. (a) Overall Accuracy by discrimination information parameter η and regularization parameter C (b) Overall Accuracy by dropout probability p when the number of basic classifiers is 3. (c) Overall Accuracy by dropout probability p when the number of basic classifiers is 5. (d) Overall Accuracy by dropout probability p when the number of basic classifiers is 7.

In order to test the performance of the proposed classifier, compare the performance of the proposed algorithm, IELM, ELM, SVM and KNN with different numbers of training sets on CNN features. Fig. 7 shows the proposed E-CNN-dropIELM has better performance on the CNN features, in contrast to these traditional algorithms and the accuracy is higher with the increase of training data.

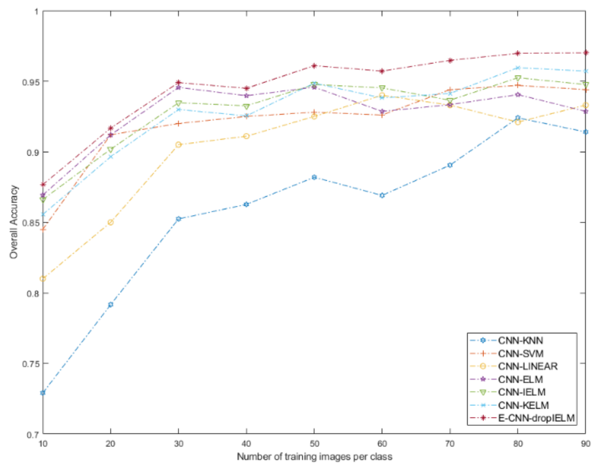

Fig. 7. Performance evaluations on the classification accuracies of KNN, SVM, LINEAR, ELM, IELM, KELM and E-CNN-dropIELM with varied proportions as training samples of the CNN feature.

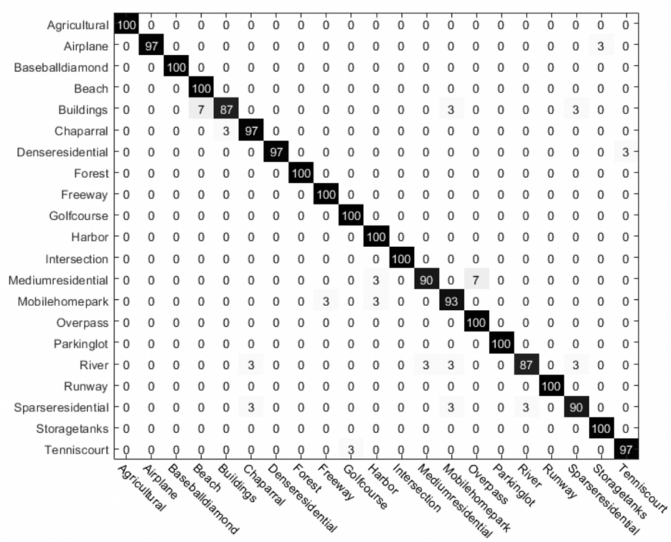

Fig. 8. Confusion matrix of the classification results achieved by the proposed E-CNN-dropIELM method using 70% of the CNN feature for training when C, η, p and number of basic classifiers are 2^{11}, 0.5,0.9 and 5.

The confusion matrix of the classification results shows that he most error-prone categories are buildings and river, and the most confused categories are buildings and beach, medium residential and overpass, as shown in Fig. 8. But actually, we can find the proposed method has resulted in more than 96% accuracy for most classes. We compare the best results with the reported classification accuracy of some existing methods on the UCM Land Use dataset. The results are shown in Table 1. As expected, the high classification accuracy of the proposed algorithm indicates that our method performs primely, and the reason is twofold. First, the features extracted by the pretrained CNN are sufficiently discriminative. On the other hand, IELMs with the random feature subspace constructed via the dropout method are used as basic classifiers, and then effectively ensembles these basic classifiers to achieve satisfactory result. In addition, to verify the classification accuracy of the proposed algorithm, experimental results show that the proposed method outperform some existing method proposed in [37,35,21,29,7].

Table 1. Classification accuracy (%) of existing methods and the proposed method on the UCM Land Use data set

Method	Accuracy(%)
MCMI-LBP [37]	88.20
CNN-SVM [21]	93.42
CaffeNet [7]	94.43
AlexNet [7]	94.37
VGG-S [7]	94.60
RCNet [35]	94.53
CNN-ELM [29]	95.62
proposed	96.90

8. Conclusion

In this paper, a novel land scene classification framework is proposed. This framework uses the pre-trained CNN with the fully connected layer removed as the feature extractor, and then it allows IELM with dropout method to be the basic classifier. Finally, the excellent classification result is obtained via majority voting, which effectively utilizes these basic classifiers based on different sub-feature spaces. The performance of the proposed E-CNN-dropIELM for land-use scene classification is proved on the UCM Land Use database. The experimental results show that the proposed algorithm can achieve more competitive results compared with the algorithm mentioned above paper. The properties of the proposed hybrid model is twofold: One is the proposed model uses pre-trained CNN to automatically extract high-level features, while most other traditional methods rely heavily on hand-designed low-level and mid-level features. The other is that the proposed model employs IELM with dropout as the basic classifiers, and then effectively ensembles these basic classifiers via majority voting. This is able to improve generalization and accuracy, and also can avoid sample shortages and overfitting problems.

Acknowledgments. This study was funded by the Major project of National Social Science Foundation (15zdc012),and by the National Natural Science Foundation of China (Grant number 61375063, 61271355, 11301549 and 11271378) and by the Fundamental Research Funds for the Central Universities of Central South University (2018zzts322).

References

1. Castelluccio, M., Poggi, G., Sansone, C., Verdoliva, L.: Land use classification in remote sensing images by convolutional neural networks. arXiv preprint arXiv:1508.00092 (2015)
2. Chen, Y., Crawford, M.M., Ghosh, J.: Applying nonlinear manifold learning to hyperspectral data for land cover classification. In: Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS'05. vol. 6, pp. 4311–4314. IEEE (2005)
3. Dietterich, T.G.: Ensemble methods in machine learning. In: International workshop on multiple classifier systems. pp. 1–15. Springer (2000)
4. Dietterich, T.G.: An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. Machine learning 40(2), 139–157 (2000)
5. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
6. Hou, M., Zhang, T., Weng, F., Ali, M., Al-Ansari, N., Yaseen, Z.M.: Global solar radiation prediction using hybrid online sequential extreme learning machine model. Energies 11(12), 3415 (2018)
7. Hu, F., Xia, G.S., Hu, J., Zhang, L.: Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. Remote Sensing 7(11), 14680–14707 (2015)
8. Hu, F., Xia, G.S., Zhang, L.: Deep sparse representations for land-use scene classification in remote sensing images. In: 2016 IEEE 13th International Conference on Signal Processing (ICSP). pp. 192–197. IEEE (2016)
9. Huang, G.B., Zhu, Q.Y., Siew, C.K.: Extreme learning machine: theory and applications. Neurocomputing 70(1-3), 489–501 (2006)
10. Jegou, H., Perronnin, F., Douze, M., Sánchez, J., Perez, P., Schmid, C.: Aggregating local image descriptors into compact codes. IEEE transactions on pattern analysis and machine intelligence 34(9), 1704–1716 (2011)
11. Kadavi, P.R., Lee, C.W.: Land cover classification analysis of volcanic island in aleutian arc using an artificial neural network (ann) and a support vector machine (svm) from landsat imagery. Geosciences Journal 22(4), 653–665 (2018)
12. Karpathy, A., Fei-Fei, L.: Deep visual-semantic alignments for generating image descriptions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3128–3137 (2015)
13. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. pp. 1097–1105 (2012)
14. Li, Y., Xie, W., Li, H.: Hyperspectral image reconstruction by deep convolutional neural network for classification. Pattern Recognition 63, 371–383 (2017)
15. Liu, G., Liu, Y., Guo, M., Liu, P., Wang, C.: Non-negative locality-constrained linear coding for image classification. In: International Conference on Intelligent Science and Big Data Engineering. pp. 462–471. Springer (2015)
16. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International journal of computer vision 60(2), 91–110 (2004)
17. Muzhou, H., Ming, C., Yangchun, Z.: A self-organizing mixture extreme leaning machine for time series forecasting. In: Proceedings of ELM-2014 Volume 1, pp. 225–236. Springer (2015)

18. Oliva, A., Torralba, A.: Modeling the shape of the scene: A holistic representation of the spatial envelope. International journal of computer vision 42(3), 145–175 (2001)
19. Othman, E., Bazi, Y., Alajlan, N., Alhichri, H., Melgani, F.: Using convolutional features and a sparse autoencoder for land-use scene classification. International Journal of Remote Sensing 37(10), 2149–2167 (2016)
20. Pal, M.: Extreme-learning-machine-based land cover classification. International Journal of Remote Sensing 30(14), 3835–3841 (2009)
21. Penatti, O.A., Nogueira, K., Dos Santos, J.A.: Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 44–51 (2015)
22. Perronnin, F., Sánchez, J., Mensink, T.: Improving the fisher kernel for large-scale image classification. In: European conference on computer vision. pp. 143–156. Springer (2010)
23. Ren, J., Jiang, X., Yuan, J.: Learning lbp structure by maximizing the conditional mutual information. Pattern Recognition 48(10), 3180–3190 (2015)
24. Ruiz Hernandez, I.E., Shi, W.: A random forests classification method for urban land-use mapping integrating spatial metrics and texture analysis. International journal of remote sensing 39(4), 1175–1198 (2018)
25. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: Imagenet large scale visual recognition challenge. International journal of computer vision 115(3), 211–252 (2015)
26. Sharif Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S.: Cnn features off-the-shelf: an astounding baseline for recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 806–813 (2014)
27. Sun, H., Hou, M., Yang, Y., Zhang, T., Weng, F., Han, F.: Solving partial differential equation based on bernstein neural network and extreme learning machine algorithm. Neural Processing Letters 50(2), 1153–1172 (2019)
28. Ustuner, M., Sanli, F.B., Dixon, B.: Application of support vector machines for landuse classification using high-resolution rapideye images: a sensitivity analysis. European Journal of Remote Sensing 48(1), 403–422 (2015)
29. Weng, Q., Mao, Z., Lin, J., Guo, W.: Land-use classification via extreme learning classifier based on deep convolutional features. IEEE Geoscience and Remote Sensing Letters 14(5), 704–708 (2017)
30. Weng, Q., Mao, Z., Lin, J., Liao, X.: Land-use scene classification based on a cnn using a constrained extreme learning machine. International journal of remote sensing 39(19), 6281–6299 (2018)
31. Yan, D., Chu, Y., Li, L., Liu, D.: Hyperspectral remote sensing image classification with information discriminative extreme learning machine. Multimedia Tools and Applications 77(5), 5803–5818 (2018)
32. Yan, D., Chu, Y., Zhang, H., Liu, D.: Information discriminative extreme learning machine. Soft Computing 22(2), 677–689 (2018)
33. Yang, Y., Newsam, S.: Bag-of-visual-words and spatial extensions for land-use classification. In: Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems. pp. 270–279 (2010)
34. Yang, Y., Hou, M., Luo, J., Liu, T.: Neural network method for lossless two-conductor transmission line equations based on the ielm algorithm. AIP Advances 8(6), 065010 (2018)
35. Zhang, L., Zhang, L., Du, B.: Deep learning for remote sensing data: A technical tutorial on the state of the art. IEEE Geoscience and Remote Sensing Magazine 4(2), 22–40 (2016)
36. Zhang, T., Hou, M., Weng, F., Yang, Y., Sun, H., Wang, Z., Gao, Z., Luo, J.: An online learning algorithm for voice activation detection based on a pretrained online extreme learning machine. In: Proceedings of the 2nd International Conference on Computer Science and Application Engineering. pp. 1–8 (2018)

37. Zhao, F., Sun, H., Liu, S., Zhou, S.: Combining low level features and visual attributes for vhr remote sensing image classification. In: MIPPR 2015: Remote Sensing Image Processing, Geographic Information Systems, and Other Applications. vol. 9815, p. 98150C. International Society for Optics and Photonics (2015)
38. Zhao, L.J., Tang, P., Huo, L.Z.: Land-use scene classification using a concentric circle-structured multiscale bag-of-visual-words model. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7(12), 4620–4631 (2014)
39. Zhu, X., Li, Z., Zhang, X.Y., Li, P., Xue, Z., Wang, L.: Deep convolutional representations and kernel extreme learning machines for image classification. Multimedia Tools and Applications 78(20), 29271–29290 (2019)

Tianle Zhang is a graduate student in the School of Mathematics and Statistics of Central South University. His research interests lie in applied statistics and deep learning.

Muzhou Hou received the B.S. degrees from Hunan Normal University, Changsha, China, M.S. and Ph.D. degree from Central South University, Changsha, China in 1985, 2002, and 2009, respectively. He is currently a Professor and doctorial advisor in School of Mathematics and Statistics, Central South University, PR China. His current research interests include machine learning, computational intelligence, neural networks, data mining and numerical approximation.

Tao Zhou is received the B.S. degrees from Hunan Normal University, Changsha, China, M.S. degree from Central South University, Changsha, China in 2002, 2010 respectively. and a Ph.D.degree from South China University of Technology, Guangzhou, Guangdong, China, in 2018. He is currently a teacher in School of Economics, Guangdong University of Finance & Economics,Guangzhou, Guangdong,China. His current research interests include machine learning, computational intelligence, neural networks, numerical approximation and Management science and Real options.

Zhaode Liu received a M.S. degree from Chongqing Normal University, Chongqing, China, in 2002 and a Ph.D.degree from South China University of Technology, Guangzhou, Guangdong, China, in 2013.candidate at Guangdong University of Finance & Economics, Guangzhou, Guangdong, China, in 2002. he is an associate professor in the Faculty of Economics, Guangdong University of Finance & Economics,Guangzhou,Guangdong,China. His research in interests lie in Comprehensive evaluation, Management science and Real options.

Weirong Chen is a graduate student in the School of Mathematics Sciences, Peking University. Her research interests include machine learning and natural language processing.

Yangjin Cheng is currently a Professor in School of Mathematics and Computational ScienceXiangtan University, PR China. His current research include interests computational intelligence, neural networks and Uncertainty optimization theory and its application in Finance.

Received: December 22, 2019; Accepted: March 20, 2020.