Convolutional Neural Network for Satellite Image Classification

Mohammed Abbas Kadhim and Mohammed Hamzah Abed

Abstract Multimedia applications and processing is an exciting topic, and it is a key of many applications of artificial intelligent like video summarization, image retrieval or image classification. A convolutional neural networks have been successfully applied on multimedia approaches and used to create a system able to handle the classification without any human's interactions. In this paper, we produce effective methods for satellite image classification that are based on deep learning and using the convolutional neural network for features extraction by using AlexNet, VGG19, GoogLeNet and Resnet50 pretraining models. The Resnet50 model achieves a promising result than other models on three different dataset SAT4, SAT6 and UC Merced Land. The accuracy of classification of this model for UC Merced Land dataset is 98%, for SAT4 is 95.8%, and the result for SAT6 is 94.1%.

Keywords Satellite image classification · Deep learning · Convolutional neural network · Features extraction

1 Introduction

In recent years, remote sensing technologies have been developed quickly. That means, acquiring an extensive collection of remote sensing images with high resolution have become much more accessible. Based on this notion, many researchers of remote sensing recognition and classifications have been moving from traditional methods to recent techniques. The traditional methods depend on the intensity of pixel level interpretation while the modern techniques are focused in the semantic understanding of the images. The semantic understanding aims to classify the data

M. A. Kadhim (✉) · M. H. Abed
College of Computer Science and IT, Computer Sciences Department, University of Al-Qadisiyah, Diwaniyah, Iraq
e-mail: mohammed.abbas@qu.edu.iq

M. H. Abed
e-mail: mohammed.abed@qu.edu.iq

M. Huk et al. (eds.), *Intelligent Information and Database Systems: Recent Developments*, Studies in Computational Intelligence 830,
https://doi.org/10.1007/978-3-030-14132-5_13

into a set of semantic categories and a set of classes depending on remote sensing image content [1–3]. The image classification can be divided into three main classes according to its features [1]. The ‘handcrafted feature-based method’ focuses on different properties such as colors and shape information, which are possible properties of sense images [2–4], while ‘unsupervised feature learning-based methods’ aim to learn a set of basic functions such as a bag of words model that is used for features encoding. The most common encoding method is called quantization, and more effective method is fisher encoding, where the input in the Fisher method is a set of handcrafted characteristics, and the output is a set of learned features [5–7]. Finally, the ‘deep feature learning-based methods’ which is called Deep Learning (DL) [8–10]. In recent years, deep learning of remote sensing image features has shown an impressive capability for classification by selection of appropriate features for the problem of remote sensing image classification [1]. Selection of appropriate the deep learning is a subfield of machine learning based on multiple layers of learning. The deep learning structure extends from the classic Neural Network (NN) by adding more layers to the hidden layer part. There are many architectures of deep learning, one of them is a Convolutional Neural Network (CNN). The CNN is widespread and has been used in recent years for handling a variety and complex problems such as image recognition and classification by using a sequence of feed-forward layers. The CNN is similar to the traditional neural network, and it is made by neurons that have learnable weights and biases. The neurons receive a set of inputs and performing some non-linear processing, and it can be considered as a feed-forward artificial neural network [11]. Convolutional network architectures use the images as inputs which allow the encoding of certain properties into the architecture. The typical structure of CNN is a series of layers including a convolutional layer, a pooling layer, and full connection layers [1]. It can be said that it is a special case of the neural network that consists of one or more convolutional layers that are responsible for extracting low-level features such as lines, edges and corners; pooling/subsampling layers that make the features robust against distortion and noise; non-linear layers that work as a trigger function to signal different identification of likely features on each hidden layer; and fully connected layers that mathematically sum up a weighting of the previous layer of features [12]. In this paper, four effective strategies and architecture of CNN have been proposed to improve the performance of satellite images classification, four approaches of CNN (AlexNet, VGG19, GoogLeNet and Resnet50) have been used as a pre-trained for features extraction, each of them trained on imageNet dataset. We evaluate our methods by combining the earlier features with more in-depth features in a fully connected layer and compare all the results of the models with several novel methodologies on three datasets SAT 4, SAT6 and UCMD.

The structure of this paper is organized as the following: in Sect. 2, we present the related works with CNN for image classification and recognition. Section 3 gives an overview of the datasets that used in our system. The proposed work and its components have been discussed in Sect. 4. Finally, Sects. 5 and 6 contain the experiment results and conclusions of this work respectively.

2 Related Works

Classification of the satellite image is a process of categorizing the images depend on the object or the semantic meaning of the images so that classification can be categorized into three major parts: methods that are based on low features, or the other methods that are based on high scene features [13]. The first method of classification that are depend on low features is used a simple type of texture features or shape features, the most common methods of low features is local binary pattern or features based on histogram same as with paper [14], the researcher in that paper used the texture with LBP as a classification tool. The methods based on mid features are suitable for a complex type of images and structure [5]. The methods that are based on high features compare with other can be considered the most effective methods for complex images. The CNN is one of the most and widely used in deep learning algorithm with image processing [9].

Saikat Basu, Sangram Ganguly, and others proposed method that is a learning framework for satellite imagery "DeepSat", they focus on classification based on deep unsupervised learning "Deep Belief Network for classification" with Convolutional Neural Networks and achieve accuracy result 97.946 for SAT4-dataset and 93.916 for SAT6-dataset [10]. Ju et al. [15] produce a research paper for investigated of a widely used ensemble approaches for image classification and recognition tasks using deep convolutional neural networks. These approaches include majority voting, the Bayes Optimal Classifier, and super learner. Albert et al. [16] analyze patterns in land use in urban neighborhoods by using large-scale satellite imagery data and state-of-the art computer vision techniques basing on deep CNN. They obtain ground truth land by using class labels carefully sampled from open-source surveys, in particular, the Urban Atlas land classification dataset of 20 land use classes across 300 European cities. They also show that the deep representations extracted from satellite imagery of urban environments can be used to compare neighborhoods across several cities. Robinson et al. [17] proposed a deep learning convolutional neural networks model for creating high-resolution population estimations from satellite imagery. The proposed CNN model has been trained to predict population in the USA at a 0.01×0.01 resolution grid from 1-year composite Landsat imagery. The CNN model evaluated and validated in two ways: quantitative and qualitative. In quantitative validation, the proposed model's grid cell estimates aggregated at a county-level comparing with several US Census county-level population projections, and qualitatively, by directly interpreting the model's predictions in terms of the satellite image inputs. In general, the proposed model is an example of how machine learning techniques can be a useful tool for extracting information from inherently unstructured, remotely sensed data to provide practical solutions to social problems. Pratt et al. [18] propose Convolutional Neural Networks approach for Diabetic Retinopathy (DR) diagnosis from digital fundus images and classify its severity. They develop CNN architecture and data augmentation which can identify the intricate features that involved in the classification task such as micro-aneurysms, exudate and hemorrhages on the retina and consequently provide a diagnosis automatically without user input. They trained the proposed CNN approach using a high-end graphics processor unit (GPU) on the Kaggle dataset and demonstrate exciting results. In this work, we will focus on CNN as a classification method. Shamsolmoali et al. [19] proposed have a new classification pipeline to facilitate a high dimensional multimedia data analysis basing on a unified deep CNN and the modified residual network which can be integrate with the other feed-forward network style in an endwise training fashion.

3 Datasets

In the proposed work, we will use three different dataset SAT4, SAT 6 and UCMD. The first two types "SAT4 and SAT6" images are extracted from the NAIP program, this data set consists of 330,000 scenes spanning of all United States images. The images consist of 4 layers red, green, blue and Near Infrared (NIR). The third dataset is UC Merced Land Use Dataset contain "tif" file image format.

3.1 SAT 4

This version of the dataset consists of 500,000 image patches that are covering four lands included barren land, trees, grassland and a class that are contain all land cover classes. 400,000 classes are chosen for the training set, and the 100,000 remain are used for a testing dataset. All images are normalized into 28 × 28 pixels [10].

3.2 SAT6

This version of the dataset contains 405,000 images each of size 28 × 28 pixels, and covering six land classes barren land, trees, grassland, roads, buildings and water bodies. 324,000 images are choosing as a training dataset, and the remain 81,000 are used for testing dataset [10]. Also, Fig. 1 shows samples image of SAT 4 and SAT 6 datasets.

3.3 UC Merced Land

This dataset consists of 21 classes land use image dataset each class contains 100 image each image measures 256 × 256 pixel. The images extracted manually from large dataset images from the USGS National Map Urban Area Imagery collection. Figure 2 shows selected samples of the images from 20 class [20].

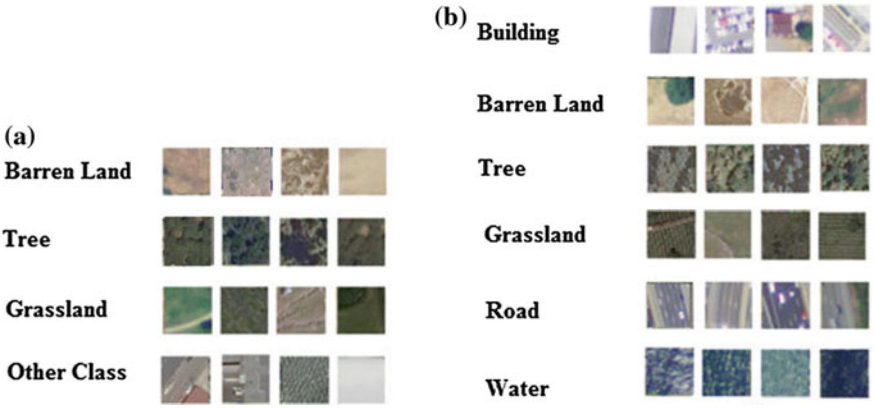

Fig. 1 Sample images "28 × 28 × 4" from **a** SAT4 and **b** SAT6 dataset

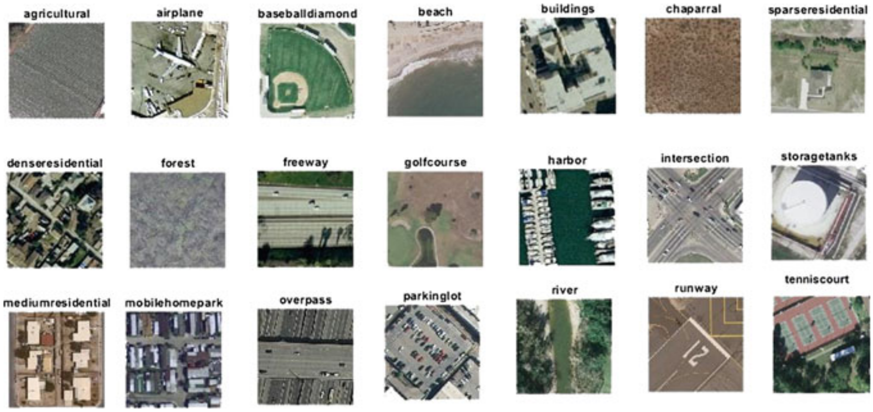

Fig. 2 Sample images from UC Merced Land dataset

4 Proposed Work

The structure of the proposed work was planned after studying the literature work on satellite image classification as in Fig. 3 that illustrates a general overview of the proposed model of satellite image classification that based on CNN. The proposed work is divided into two parts: the training phase and testing phase. The datasets are divided into two sets initially the first one is used as a training image and the second one used for testing of our models. SAT dataset consists of SAT4 and SAT6 each one contains 400,000, 324,000 images are selected as a training set consecutively and 100,000, 81,000 images are selected as a testing set.

The other datasets UC Merced Land Use that contain 21 class each one has 100 images, we have selected 70 images as training set and 30 images as testing set for all the classes. Moreover, because of the model implemented and tested on two different

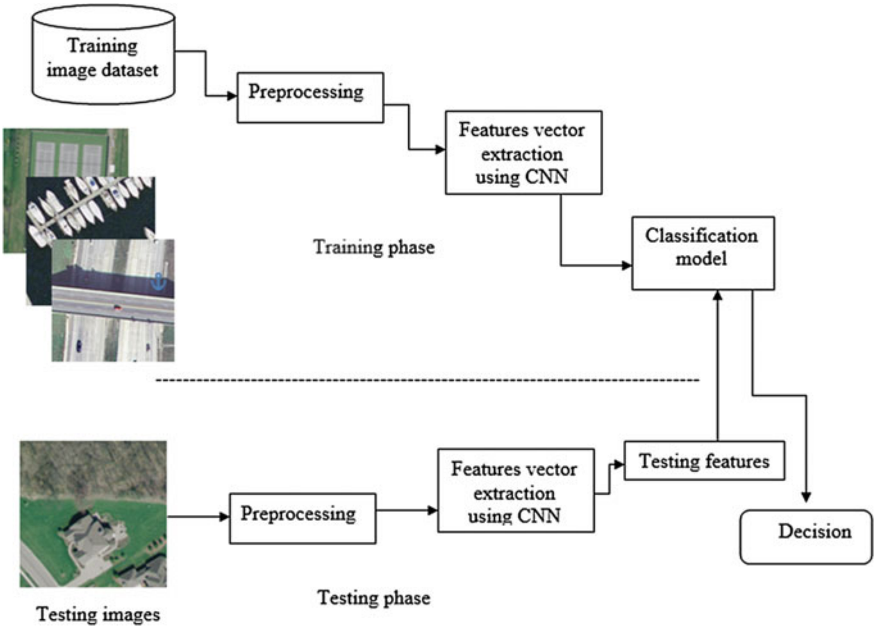

Fig. 3 The diagram of the proposed system

datasets, the preprocessing phase is such an important step to make the input images sharing the same characteristics.

4.1 Training Phase

The first stage in our model is the training phase. In this part, the selected images from both of datasets as training images are going through steps starting from preprocessing features vector extraction based on CNN.

Preprocessing. The datasets that used in our model are different, the color images of SAT airborne datasets consist of four bands 28 × 28 uint8, and the other dataset UCMD 256*256 uint8 three bands for red, green and blue. So to build a model that used for classification you must be starting with color normalization all image by reducing the invisible band NIR of the SAT datasets, convert the images into grayscale, and then the entire satellite images are ready to the next step for extracting features vector that belongs to each image in training set.

Features extraction based on CNN. The efficiency of satellite image classification is based on the power of the features that extracted from the training dataset. The power of that features will be reflected on testing phase. So by proposed off-the-

Table 1 Pretrained network, layers and features layers

	AlexNet	VGGNet-19	GoogleNet	Resnet50
Input data	Input image 227 × 227 × 3	Input image 224 × 224 × 3	Input image 224 × 224 × 3	Input image 224 × 224 × 3
Layers	25 × 1 nnet.cnn.layer.Layer	47 × 1 nnet.cnn.lay er.Layer	144 × 1 nnet.cnn.layer.Layer	177 × 1 nnet.cnn.layer.Layer
Features layer	fc8	fc8	loss3-classifier	fc1000

shelf features extraction from the images, we provide high-level features to be set of training data to train the CNN.

In this work, we have used several pretrained networks all of them have been trained on the ImageNet dataset as can visit the link http://www.image-net.org/ which have one thousand object categories. Table 1 shows the characteristic of each one that used and the fully connected layer that we have considered it as a features vector. Every CNN's layer produces an activation or response to an input image. In all these layers there are only a few layers within CNN architecture that can be suitable for features extraction of the input image. The feature that have been extracted from the deeper layer can be used as a training feature because it gives advance features contrariwise the beginning layer of the CNN capture only the primary image features like edge and blobs. The first layer of the CNN has learned for detecting the edge and blob features, and these original features are processed by deeper layer in this case the first features are combined with more in-depth high-level features in full connections layer, that can be used in recognition or classification tasks, so the fully connected layer is chosen to be features's layer.

4.2 Testing Phase

The second phase of the satellite image classification model is a testing phase. In this part, the 30% remaining of each dataset will be tested to check and measure the accuracy of the classifier method. Same as with a prepare the input data for training phase it will occur the testing images starting with preprocessing and extract set of features for all categories in the datasets and save it as two-dimensional matrices each row belongs to the one image. we will explain the experimental result of the satellite image classification based on CNN.

5 Experimental Results

We evaluate the performance of the satellite image classification that is used datasets which mention in the datasets section above. In this part, we will discuss the experimental results that are implemented based on a combination of deep features and earlier features of CNN by using four models AlexNet, VGGNet-19, GoogleNet and Resnet50 which are pretrained on imageNet dataset. The features are extracted from different layer based on the model type and full connection layers have shown in Table 1. Due to we have used different datasets and varying dimensions, we kept the size of an image and normalized the four bands into visible layers only red, green and blue. The features layer are selected in four models from last pooling full connection layer: AlexNet is layer number 23 "fc8", VGGNet-19 is 45 layer "fc8", googleNet is layer number 142 "loss3-classifier" and Resnet50 is layer number 175 "fc1000". Table 2 shows the configuration of the four models on UCMD dataset.

In this work, we have tested four pretrained CNN with their configuration that are listed in Table 2, on different datasets SAT 4, SAT 6 and UC Merced Land. Each dataset is divided randomly into two part: training and a testing subset of images, Table 3 shows the datasets setting in our experimental results.

The proposed method that is based on combination of deep features and earlier features with Resnet50 that extracted from "fc1000" layer achieve better result than features extracted from first convolutional or deep convolutional, also it shows better performance than other pretrained convolutional neural network like AlexNet, VGG-19 and GoogleNet because the feature that extracted from Resnet50 are deeper than the others under the selected percentage 70% of training with the configuration of Resnet50 that shows in Table 2.

Also, Fig. 4 shows the accuracy and achieve of the Resnet50 model has the better result than other models, and Fig. 5 shows the loss of training of the samemodel in 250 epochs both of them by using UC Merced Land Datasets. Figure 6 presents the comparison among the models that used for features extraction, its visible that the Resnet50 model used for features extraction has a better result of classification than other models and loss function is less than others.

So Table 4 show the accuracy of all datasets that used with different models and algorithms.

As shown above in Table 4, the accuracy values that produced by the research paper [10] is achieved a classification ratio on SAT4 and SAT6 reached to 97.946 and 93.916 respectively. They suggested a mechanism for extracting data and features of an input image and used the principle of normalization of that features as a vector in Deep Belief Network for classification. They presented two datasets SAT4 and SAT6 and that proposed work didn't test on UC Merced Land. The second research paper in Table 4 [21] that investigated in our experiments, the researchers proposed an agile CNN architecture named SatCNN for HSR-RS image scene classification. Based on recent improvements to modern CNN architectures and they are used a smaller kernel with effective convolutional layers to build an effective and fast CNN. The method tested on SAT4 and SAT6 with achievement ratio 99.65 and 99.54

Table 2 Configuration of the pretrained models

AlexNet		VGGNet-19		GoogleNet		Resnet50	
Layer	Configuration	Layer	Configuration	Layer	Configuration	Layer	Configuration
Conv1	Filter 96 11 × 11 × 3 Stride 1 [4 4] Pooling 3 × 3 Stride2 [2 2]	conv1_1	Filter 64 3 × 3 × 3 Stride1 [1 1] Pooling 2 × 2 Stride2 [2 2]	conv1-7 × 7_s2	Filter 64 7 × 7x3 Stride1 [2 2] Pooling 3 × 3 Stride2 [2 2]	conv1	Filter 64 7 × 7 × 3 Stride 1 [2 2] Pooling 3 × 3 Stride2 [2 2]
Conv2	Filter 256 5 × 5 × 48 Stride 1 [1 1] Pooling 3 × 3 Stride2 [2 2]	conv2_1	Filter 128 3 × 3 × 64 Stride1 [1 1] Pooling 2 × 2 Stride2 [2 2]	conv2-3 × 3_reduce	Filter 64 1 × 1 × 64 Stride 1 [1 1] Pooling 3 × 3 Stride2 [2 2]	res2a_branch2a	Filter 64 1 × 1 × 64 Stride 1 [2 2] Pooling 3 × 3 Stride2 [2 2]
Conv3	Filter 384 3 × 3 × 256 Stride 1 [1 1] Pooling 3 × 3 Stride2 [1 1]	conv3_1	Filter 256 3 × 3 × 128 Stride1 [1 1] Pooling 2 × 2 Stride2 [2 2]	inception_3a-1 × 1	Filter 64 1 × 1 × 192 Stride 1 [1 1] Pooling 3 × 3 Stride2 [1 1]	res3d_branch2c	Filter 512 1 × 1 × 128 stride [1 1]

(continued)

Table 2 (continued)

AlexNet		VGGNet-19		GoogleNet		Resnet50	
Layer	Configuration	Layer	Configuration	Layer	Configuration	Layer	Configuration
Conv4	Filter 384 3 × 3 × 192 Stride 1 [1 1] Pooling 3 × 3 Stride2 [1 1]	conv4_1	Filter 512 3 × 3 × 256 Stride1 [1 1] Pooling 2 × 2 Stride2 [2 2]	inception_3b-1 × 1	Filter 128 1 × 1 × 256 Stride1 [1 1] Pooling 3 × 3 Stride2 [1 1]	res4e_branch2c	Filter 1024 1 × 1 × 256 Stride1 [1 1]
Full connection 1	fc7 4096 fully connected layer	Full connection 1	fc7 4096 fully connected layer	–	–	–	–
Full connection 2	fc8 1000 fully connected layer	Full connection 2	fc8 1000 fully connected layer	Full connection 1	loss3-classifier 1000 fully connected layer	Full connection 1	fc1000 1000 fully connected layer

Table 3 Dataset setting for experimental results

Dataset	Subset of images		Images resolution	
	Training	Testing	Width	Height
SAT 4	400,000	100,000	28	28
SAT 6	324,000	81,000	28	28
UC Merced Land	21 class × 70	21 class × 30	256	256

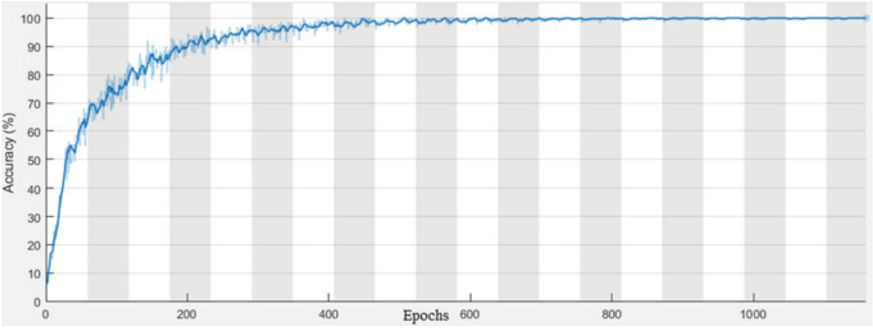

Fig. 4 Training accuracy of the Resnet50 model

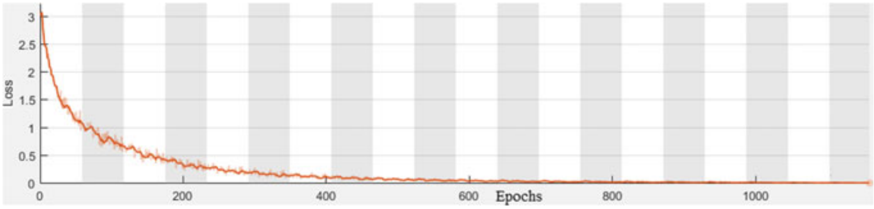

Fig. 5 The training loss of Resnet50 model

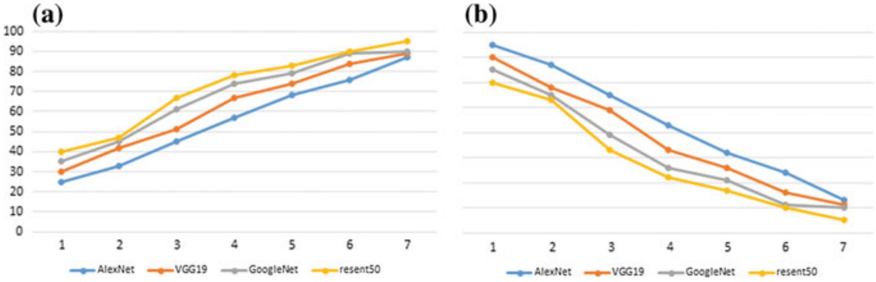

Fig. 6 Comparison between models based on UC Merced Land dataset. **a** Accuracy of training. **b** Loss of training

Table 4 Accuracy of the SAT4, SAT6 and UC Merced Land dataset based on different models and algorithms that are used

Algorithms	Classifier accuracy % SAT4	Classifier accuracy % SAT6	Classifier accuracy % UC Merced Land
DeepSat [10]	97.946	93.916	–
Agile CNN SatCNN [21]	99.65	99.54	–
Triplet networks [22]	99.72	99.65	97.99
Features extraction based on AlexNet	84	82	87
Features extraction based on VGG19	89	84	89
Features extraction based on GoogleNet	91	89	90
Features extraction based on Resnet50 (proposed method)	95.8	94.1	98

respectively and it is not tested on UC Merced Land. The performance accuracy of the third research paper [22] is 99.72, 99.65, and 97.99 on SAT4, SAT6, and UC Merced Land datasets respectively. The researchers produce a novel classification method via triple networks. In our experiment results on proposed methods based on features extraction depend on Resnet50 achievement produce the best model for classifying image set of UC Merced Land dataset. The performance of our proposed model (Resent50) is better than results yielded from research paper [10] for SAT6 dataset and it is the worst for SAT4 dataset.

6 Conclusions

In this paper, we present useful models for satellite image classification that are based on convolutional neural network, the features that are used to classify the image extracted by using four pretrained CNN models: AlexNet, VGG19, GoogleNet and Resnet50 and compare the result among them. The features are extracted from a combination layer or full connection layer of earlier layers and deep layers. After the experiment result of the datasets and the pretrained models, the Resnet50 model achieves a better result than other models for all the datasets that are used "SAT4, SAT6 and UC Merced Land". The result of the classification based on Resnet50 as features extraction has better accuracy and minimum loss value than other methods and able to work on different datasets. The achievement of our proposed method based on Resnet50 is better result than research paper [10] for image classifying of SAT6, it is also a better than research paper [22] for classify UC Merced Land dataset.

References

1. Cheng, G., Li, Z., Yao X., Guo, L., Wei, V.: Remote sensing image scene classification using bag of convolutional features. IEEE Geosci. Remote Sensing Lett. **14**(10), (2017)
2. Bian, X., Chen, C., Tian, L., Du, Q.: Fusing local and global features for high-resolution scene classification. IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens. **10**(6), 2889–2901 (2017)
3. Huang, L., Chen, C., Li, W., Du, Q.: Remote sensing image scene classification using multi-scale completed local binary patterns and Fisher vectors. Remote Sens. **8**(10), (2016)
4. Chen, C., Zhang, B., Su, H., Li, W., Wang, L.: Land-use scene classification using multi-scale completed local binary patterns. Signal Image Video Process. **10**(4), 745–752 (2016)
5. Zhang, F., Du, B., Zhang, L.: Saliency-guided unsupervised feature learning for scene classification. IEEE Trans. Geosci. Remote Sens. **53**(4), 2175–2184 (2015)
6. Li, Y., Tao, C., Tan, Y., Shang, K., Tian, J.: Unsupervised multilayer feature learning for satellite image scene classification. IEEE Geosci. Remote Sens. Lett. **13**(2), 157–161 (2016)
7. Yuan, Y., Wan, J., Wang, Q.: Congested scene classification via efficient unsupervised feature learning and density estimation. Pattern Recogn. **56**, 159–169 (2016)
8. Yao, X., Han, J., Cheng, G., Qian, X., Guo, L.: Semantic annotation of high-resolution satellite images via weakly supervised learning. IEEE Trans. Geosci. Remote Sens. **54**(6), 3660–3671 (2016)
9. Zou, Q., Ni, L., Zhang, T., Wang, Q.: Deep learning based feature selection for remote sensing scene classification. IEEE Geosci. Remote Sens. Lett. **12**(11), 2321–2325 (2015)
10. Basu, Saikat, Ganguly, Sangram, Mukhopadhyay, Supratik, DiBiano, Robert, Karki, Manohar, Nemani, Ramakrishna: DeepSat—A Learning Framework For Satellite Imagery, SIGSPATIAL'15, Nov 03–06, 2015. Bellevue, WA, USA (2015)
11. Yu, X., Wu, X., Luo, C., Ren, P.: Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework. GISci. Remote Sensing (2017)
12. Hijazi, S., Kumar, R., Rowen, C.: Using Convolutional Neural Networks for Image Recognition, IP Group, Cadence
13. Yu, Y., Liu, F.: Dense connectivity based two-stream deep feature fusion framework for aerial scene classification. www.mdpi.com/journal/remotesensing (2018)
14. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**, 971–987 (2002)
15. Ju, C., Bibaut, A., van der Laan, M.J.: The relative performance of ensemble methods with deep convolutional neural networks for image classification, ArXiv e-prints, Apr (2017)
16. Albert, A., Kaur, J., Gonzalez, M.: Using convolutional networks and satellite imagery to identify patterns in urban environments at a large scale. In: Proceeding of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining pp. 1357–1366 (2017)
17. Robinson, C., Hohman, F., Dilkina, B.: A deep learning approach for population estimation from satellite imagery. In: Proceedings of the 1st ACM SIGSPATIAL Workshop on Geospatial Humanities, pp. 47–54 (2017)
18. Pratt, H., Coenen, F., Broadbent, D.M., Harding, S.P., Zheng, Y.: Convolutional neural networks for diabetic retinopathy. In: International Conference On Medical Imaging Understanding and Analysis, MIUA 2016, Loughborough, UK, (2016)
19. Shamsolmoali, P., Jain, DK., Zareapoor, M., Yan, J., Alam, M.A.: High-dimensional multimedia classification using deep CNN and extended residual units. Multimedia Tools Appl. https://doi.org/10.1007/s11042-018-6146-7 (2018)
20. Yang, Y., Newsam, S.: Bag-of-visual-words and spatial extensions for land-use classification. In: ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM GIS), (2010)

21. Zhong, Yanfei, Fei, Feng, Liu, Yanfei, Zhao, Bei, Jiao, Hongzan, Zhang, Liangpei: SatCNN: satellite image dataset classification using agile convolutional neural networks. Remote Sensing Lett. **8**(2), 136–145 (2017)
22. Liu, Yishu, Huang, Chao: Scene Classification via Triplet Networks. IEEE J Selected Topics Appl Earth Observ. Remote Sensing **11**(1), 220–237 (2018)