International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:10, No:6, 2016

Satellite Imagery Classification Based on Deep Convolution Network

Zhong Ma, Zhuping Wang, Congxin Liu, Xiangzeng Liu

Abstract**—**Satellite imagery classification is a challenging problem with many practical applications. In this paper, we designed a deep convolution neural network (DCNN) to classify the satellite imagery. The contributions of this paper are twofold — First, to cope with the large-scale variance in the satellite image, we introduced the inception module, which has multiple filters with different size at the same level, as the building block to build our DCNN model. Second, we proposed a genetic algorithm based method to efficiently search the best hyper-parameters of the DCNN in a large search space. The proposed method is evaluated on the benchmark database. The results of the proposed hyper-parameters search method show it will guide the search towards better regions of the parameter space. Based on the found hyper-parameters, we built our DCNN models, and evaluated its performance on satellite imagery classification, the results show the classification accuracy of proposed models outperform the state of the art method.

Keywords**—**Satellite imagery classification, deep convolution network, genetic algorithm, hyper-parameter optimization.

I. Introduction

NOWADAYS, massive imagery generated by satellites every day, putting forward urgent requirements for using the AI research to analyze these images to inform decision-making. One fundamental problem among these analyses is satellite imagery classification, which has a wide range of applications.

While efficient classification of scenes from satellite image data is a challenging problem, due to the high variability inherent in satellite data, and the large-scale variance of objects, most of the current object classification approaches are not suitable for handling satellite datasets [1].

Deep Learning has gained popularity over the last decade due to its ability to learn data representations in an unsupervised manner and generalize to unseen data samples using hierarchical representations. And it has yielded superior performance in many image classification tasks. Applying the deep learning technology to satellite imagery analysis is becoming an active research topic.

Mnih and Hinton [2] proposed a method that uses Deep Neural Networks to detect roads in Aerial imagery. But their work focused on the initialization of the weights of neural network using Restricted Boltzmann Machines (RBMs), and did not investigate the classification of aerial imagery. Basu et al. [1] built a satellite imagery classification database based on this data and investigated the classification of satellite imagery using various deep learning algorithms, including Deep Belief Network (DBN), deep convolutional neural networks (DCNN), and Stacked Denoising Autoencoder (SDAE). Their results show that directly using the current deep learning algorithms did not go to satisfying performance, the accuracy of classification of these algorithms are all between 70%~90%. They reckon the key to improve the performance is to extract better features. Thus, they extracted 150 features using traditional methods, normalized them and fed the normalized feature vectors to a Deep Belief Network for classification. But the idea of deep learning is to learn features from data automatically, therefore their method did not take full advantage of deep learning. Xueyun Chen et al. [3] proposed a hybrid deep convolutional neural network method to detect vehicles in satellite images. They think that the most challenging part in the analysis of satellite images is the large-scale variance of object, which the traditional DCNN is difficult to tolerate. Hence, they presented a hybrid DNN (HDNN), by dividing the maps of the last convolutional layer and the max-pooling layer of DNN into multiple blocks of variable receptive field sizes or max-pooling field sizes, to enable the HDNN to extract variable-scale features.

With respect to general image classification, the result from the Image Net Large Scale Visual Recognition Challenge (ILSVRC) 2014 shows that the most advanced image classification methods worldwide are all based on deep convolutional neural networks (DCNN). Thus, the most promising way to improve the classification of the satellite imagery is the method building on the success of the DCNN. Particularly, GoogLeNet [4] won the first place on the ILSVRC2014 classification task. The basic component of their model is “Inception module”, which is designed based on the Hebbian principle and can capture multi-scale features within the same level. This property may help to deal with the large-scale variance in satellite imagery. But this Inception module and the whole DCNN model both have lots of architectural hyper-parameters, and for now, there is no scientific way to determine these hyper-parameters.

In this paper, we designed a DCNN to classify the satellite imagery. To tackle the large-scale variance in satellite imagery, we choose the Inception module as the basic component of our DCNN model. We proposed a genetic algorithm based method to optimize the hyper-parameters of the DCNN. The model is evaluated on the DeepSat dataset [1], the experimental results in satellite imagery classification justify the benefits of this Inception module based structure and the genetic algorithm based hyper-parameter optimization.

Zhong Ma is with the Xi'an Microelectronics Technology Institute, Xi’an, China (phone: 0086-029-8730; e-mail: mazhong@mail.com).
Zhuping Wang, Congxin Liu, and Xiangzeng Liu are with the Xi'an Microelectronics Technology Institute, Xi’an, China (e-mail: zxjwl@126.com, sxphillx@163.com, lxzccy20062008@126.com).

II. Deep Convolutional Neural Network Design

We first present the general architecture of our DCNN model, then describe how to optimize the hyper-parameters using genetic algorithm.

A. Deep Conventional Neural Network Architecture

The GoogleNet is built in the Inception Module, which is designed following the Hebbian principle [5], which can be interpreted intuitively as: Neurons that fire together, wire together. The architecture of Inception Module is shown in Fig. 1. It combines filters with different size of 1×1, 3×3, 5×5 into one layer, and we think that such design makes it more robust to scale variance.

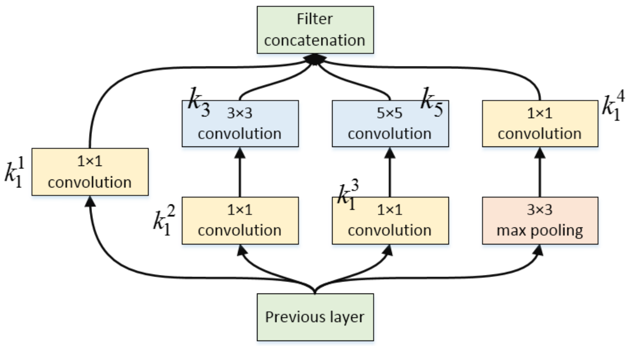


Fig. 1 The architecture of Inception Module. The notations of each filter are shown near corresponding filter.

Such design also brings a lot of hyper-parameters; it includes:

- The number of each 1×1 filter k_1^i, $i \in \{1,2,3,4\}$ represent different 1×1 filter, respectively.
- The number of 3×3 filter k_3 and 5×5 filter k_5.

All these notations are also shown on Fig. 1, close to each corresponding filter for clarity.

Our whole DCNN model is built by stacking the Inception Modules on top of each other. Therefore, with respect to the whole DCNN model, there are some more hyper-parameters, includes: the number of Inception Modules N_I, and the type of loss $T_l \in \{\text{hinge loss, softmax}\}$.

With computational efficiency and practicality in mind, we set these parameters $k_1^i, k_3, k_5 \in \{4,8,16,32,64,128,256\}$, $N_I \in \{1,2,3,4,5\}$. Thus, there are (4+1+1)×7×5×2=13,720 possible combinations.

B. Genetic Algorithm Based Architectural Parameter Optimization

The performance of any single model instantiation may range from chance to state-of-the-art performance depending on parameter configurations. In this paper, the parameter space is huge, and each configuration evaluation is very time consuming, so it is not practicable to evaluate every single configuration.

A practicable way is random search, and it was found to be an effective approach to search for particularly discriminative representations for recognition tasks [6]. However, for large enough problems, random search is still prohibitively computationally expensive. Therefore, the model search should not be random but guided towards better regions of the parameter space. Recently, automated hyper-parameter optimization [7] was proposed to use Bayesian optimization methods to guide search in a large parameter space. However, the Bayesian optimization methods make the next search choice based on the previous choice, which makes it hard to parallelize.

In this paper, we proposed a genetic algorithm (GA) based architectural parameter optimization method, using genetic algorithm to guide the parameter search. And the evaluation of each configuration within one generation is independent, makes this method easy to parallelize on a distributed computing system.

Our optimization method consists of the following steps: 1. The chromosome representation of the solution domain; 2. Evaluating the fitness the chromosomes; 3. Genetic operator construction, includes selection, crossover and mutation. The whole procedure of the method is shown in Fig. 2.

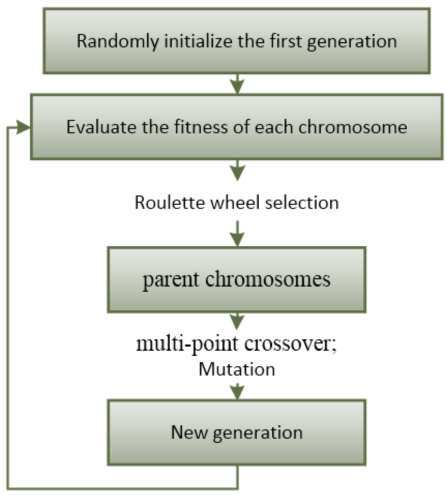


Fig. 2 The whole process of proposed genetic algorithm based architectural parameter optimization method.

The most popular chromosome representation method is binary coding, but it's not applicable for our case. In this paper, we use a different representation method, each chromosome c_i is a set of all hyper-parameters: $c_i \in \{k_1^1, k_1^2, k_1^3, k_1^4, k_3, k_5, N_I, T_l\}$. Each hyper-parameter in such chromosome is called a gene. The first generation is initialized randomly.

Then, we use each of these c_i as a configuration to build a DCNN model, after train these models, the accuracy of these models on test dataset is the fitness of corresponding c_i. Note that the evaluation of each chromosome within one generation is independent, so this step can be easily parallelized on a distributed computing system.

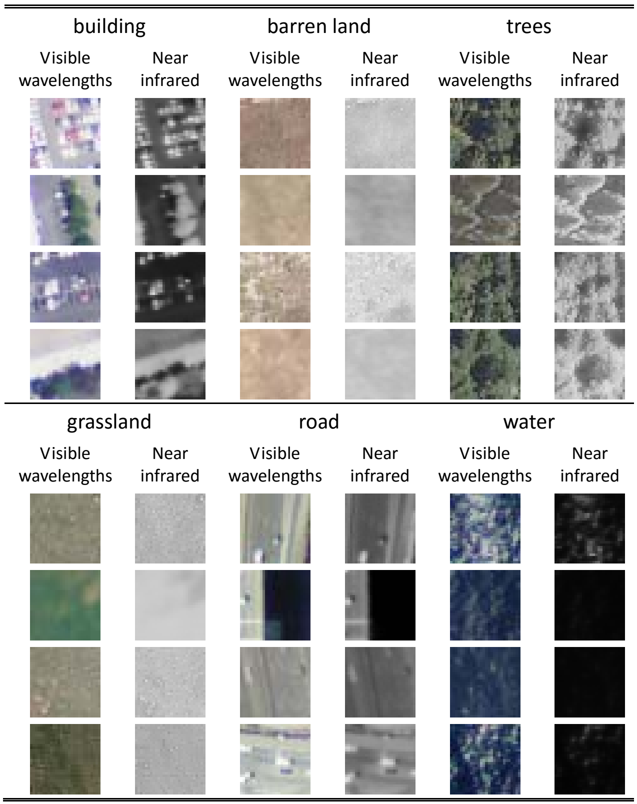


Fig. 3 Sample images of six different classes from the DeepSat dataset. Left columns of each class is images with visible wavelengths band, right column of each class is the same images with near infrared band

The roulette wheel selection is used to select the chromosomes according to their fitness. Concretely, the probability of a chromosome a selected is proportional to their fitness. To avoid 0 or 1 probabilities, the fitness (accuracy) a is first converted using:

$$f = \frac{1}{1+e^{(1/2-a)\bullet m}}. \tag{1}$$

where m is a parameter controls how close the converted fitness f can be to 0 or 1. In this paper, we set $m = 10$.

After selecting a pair of chromosomes, we use multi-point crossover to generate a child chromosome, the crossover rate is set to 0.5. Intuitively, each gene in a child chromosome comes from either one of parent chromosomes with equal probability.

Then, for each gene, there is a small probability of mutation, which means each gene has a slight chance to reselect randomly from the hyper-parameters space.

III. Implementation and Experimental Results

Hyper-parameters search and model performance evaluation were conducted on the DeepSat dataset [1], which is a benchmark database of satellite imagery classification. It includes two subsets: SAT-4 and SAT-6. SAT-4 consists of a total of 500,000 images covering four broad land cover classes. These include — barren land, trees, grassland and a class that consists of all land cover classes other than the above three. 400,000 images were chosen for training and the remaining 100,000 were chosen as the testing dataset. SAT-6 consists of a total of 405,000 images covering 6 land cover classes — barren land, trees, grassland, roads, buildings and water bodies. 324,000 images were chosen as the training dataset and 81,000 were chosen as the testing dataset. The images in this dataset are from the National Agriculture Imagery Program (NAIP), each consists of 4 channels — red, green, blue and Near Infrared (NIR), and the size of each image is all 28×28. Sample images from the dataset are shown in Fig. 3.

A. Architectural Parameter Optimization

The training of DCNN model requires a large amount of computation, while the search space is huge. To find the optimized hyper-parameter in a reasonable time, the search of hyper-parameter is conducted on two subsets of DeepSat dataset. Each subset consists of 4,000 training images, and 1,000 testing images, that randomly choose from SAT-4 and SAT-6, respectively. Each image in both subsets subtracted mean of the subset before evaluating the hyper-parameter.

The population size of each generation p is set to 100. The number of generations is set to 10. So the total configurations we have evaluated are 1,000. The mutation rate is set to $1/p$.

To verify the search efficiency of the proposed method, we compared it with random search. 1,000 configurations were chosen randomly from the search space, and evaluated on the same subsets.

Considering the memory efficiency and computational complexity, so that the model can be run on devices including even those with limited computational resources, we start using Inception modules only at higher layers while keeping the lower layers in traditional convolutional fashion. Besides, in traditional DCNN model, the late fully connected layers have the most parameters, and are prone to overfitting. To further reduce the computation complexity and improve the generalization ability, we use a global average pooling layer [8] replace the traditional fully connected layers. Concretely, instead of adding fully connected layers on top of the stacked Inception module, we take the average of each feature map, and the resulting vector is fed directly into the loss layer. The whole DCNN model architecture is shown in Fig. 4.

To verify the search efficiency of the proposed optimization method, we compared it with random search. 1000 configurations were chosen randomly, and evaluated on the test set. The results along with the result of the proposed method are shown in Fig. 5. The results show that rather than randomly test configurations from parameters space, the search of the proposed method is guided towards better regions of the parameter space, and eventually found a better configuration within 1,000 trails.

International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:10, No:6, 2016

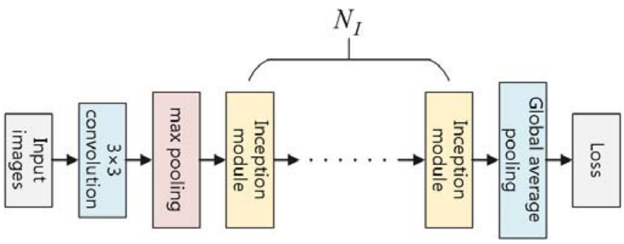
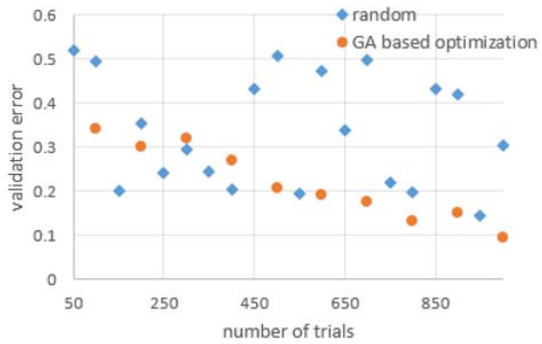
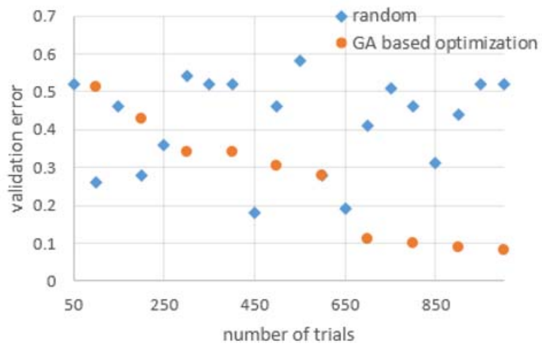


Fig. 4 The whole DCNN model architecture



(a)



(b)

Fig. 5 Optimization of validation set performance on data sets SAT-4 (a) and SAT-6 (b)

The hyper-parameters found by the proposed method after 10 generations are shown in TABLE I.

TABLE I
THE RESULTS OF PROPOSED ARCHITECTURAL PARAMETER OPTIMIZATION METHOD

hyper-parameters	k_1^1	k_1^2	k_1^3	k_1^4	k_3	k_5	N_I	T_l
On subset of SAT-4	256	64	32	64	256	64	2	Softmax
On subset of SAT-6	32	256	256	64	64	4	5	Softmax

B. Satellite Imagery Classification

After we found the optimized hyper-parameters, we built a DCNN model based on it, and evaluated its performance of satellite imagery classification on DeepSat dataset.

Since the data augmentation [9] significantly improved the performance of CNN, it almost became a standard procedure for training the CNN model now. In this paper, we also augmented the data before we trained our model on it.

For each image in the training set, we rotated it with $\theta \in \{90,180\}$ clockwise, then together with the original image, all these three images were flipped horizontally to produce three new images. Thus, the amount of images was augmented to 6 times as many as its original data. To get a fair comparison with the state-of-the-art method, only the two training sets were augmented, the tests of the proposed models are still conducted on the original test set.

The accuracy of the proposed DCNN models as well as the state of the art method is shown in

TABLE . The results show the proposed DCNN models outperform the state of the art method on satellite imagery classification. And note that the DeepSat method used 150 human crafted features, the proposed DCNN models are end to end models, it takes the original image data as its input.

TABLE II
CLASSIFICATION ACCURACY OF PROPOSED DCNN MODEL AS WELL AS THE STATE OF THE ART METHOD ON SAT-4 AND SAT-6

	Classifier Accuracy on SAT-4 (%)	Classifier Accuracy on SAT-6 (%)
Proposed DCNN model	98.408%	96.037%
DeepSat	97.946%	93.916%

IV. CONCLUSION AND FUTURE DIRECTIONS

In this paper, we have presented an approach for automatically classifying satellite imagery using DCNN. The satellite imagery classification is a challenging problem. We introduced the Inception module as the building block to build our DCNN model to cope with the large-scale variance in the satellite images. Another tricky part in building DCNN model is there are lots of hyper-parameters need to be determined. We proposed a genetic algorithm based hyper-parameters optimization method to address this issue and efficiently searched a large pool of proposed DCNN models. Comparing with random search, the search results of the proposed method are guided towards better regions of the parameter space, and can find a better configuration within limited trails. We built DCNN models based on the found hyper-parameters, and evaluated the models on the benchmark database, the results show the proposed models outperform the state of the art method.

For further work, we intend to analyze the computational complexity of the proposed model, trying to reduce the prediction time while keep the classification accuracy. Another challenging part we encountered with in satellite imagery classification research is the lack of labeled data, while the unlabeled data are relatively easy to access. Therefore, as another aspect of our further work, we plan to investigate unsupervised feature learning on unlabeled satellite imagery, and clustering the unlabeled satellite imagery based on this kind of features.

ACKNOWLEDGMENT

The authors thank Tang Lei from Xi'an Microelectronics Technology Institute for his valuable comments to improve the quality of this paper.

REFERENCES

[1] S. Basu, S. Ganguly, S. Mukhopadhyay, R. DiBiano, M. Karki, and R. Nemani, "DeepSat-A Learning framework for Satellite Imagery," *arXiv preprint arXiv:1509.03602*, 2015.

[2] V. Mnih and G. E. Hinton, "Learning to Detect Roads in High-Resolution Aerial Images," in *Computer Vision - ECCV 2010, PT VI*, 2010, vol. 6316, pp. 210–223.

[3] Xueyun Chen, Shiming Xiang, Cheng-Lin Liu, and Chun-Hong Pan, "Vehicle Detection in Satellite Images by Hybrid Deep Convolutional Neural Networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 10, pp. 1797–1801, Oct. 2014.

[4] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *arXiv preprint arXiv:1409.4842*, 2014.

[5] Sanjeev Arora, Rong Ge, and Tengyu Ma, "provable bounds for learning some deep representations," presented at the International conference on machine learning, 2014.

[6] Cox D. and Pinto N., "Beyond Simple Features: A Large-Scale Feature Search Approach to Unconstrained Face Recognition," presented at the IEEE International conference on Automated Face and Gesture Recognition, Santa Barbara, CA, 2011, pp. 8–15.

[7] J. Bergstra, D. Yamins, and D. Cox, "Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures," in *Proceedings of The 30th International Conference on Machine Learning*, 2013, pp. 115–123.

[8] Min Lin, Qiang Chen, and Shuicheng Yan, "Network In Network," presented at the International conference on learning representations, Banff, Canada, 2014.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems 25*, 2012, pp. 1097–1105.

Zhong Ma was born in China, in 1984. He received the PhD degree from Northwestern Polytechnical University, Xi'an, China, in 2015. In 2013, he was a Visiting Researcher with the Centre for Computational Intelligence, De Montfort University, UK. He is now a post doc researcher in the Xi'an Microelectronics Technology Institute, China. His research interests include computer vision, machine learning and eye tracking research. He is the author and coauthor of more than 10 articles.

Zhuping Wang was born in China, in 1963. He received his MS from the Xi'an Microelectronics Technology Institute, Xi'an, China, in 1989. Since 1986, he has been with the Xi'an Microelectronics Technology Institute. He is now the director of the research and development department of the Institute. His research interests include high performance computing and redundant systems.

Congxin Liu received a B.S. degree from Wuhan University of Hydraulic and Electric Engineering (Yi Chang), China, in 1997, and Ph.D. degree from Shanghai Jiao Tong University, China, in 2012. He is currently a senior software engineer in Xi'an Microelectronics Technology Institute. He has taken charge of many research projects (e.g. Core Electronic Devices, High-end Generic Chips and Basic Software, 863 National High Tech.Plan). His research interests include local invariant feature, machine learning and Parallel Computing.

Xiangzeng Liu received both his M.S. and his Ph.D. degrees in applied mathematics department from the Northwest Polytechnical University in 2008 and 2011, respectively. Since 2012, he has worked as an intelligent image processing algorithm designer in Xi'an Microelectronics Technology Research Institute. His research interests are computer vision and image processing, which includes infrared image processing, remote sensing processing, objects detection and recognition.