

# Deep Learning-Based Classification of Remote Sensing Image

Jian-min Liu<sup>1,2\*</sup>, Min-hua Yang<sup>1</sup>

<sup>1</sup> School of Geosciences and Info-Physics, Central South University, Changsha, 41000, China

<sup>2</sup> School of Information, Hunan Institute of Humanities Science and Technology, Loudi, 417000, China

\* Corresponding author. Email: liujianmin@csu.edu.cn

Manuscript submitted May 8, 2016; accepted January 3, 2017.

doi: 10.17706/jcp.13.1.44-48

---

**Abstract:** Deep Learning networks have sharply increased over the past 10 years, and deep Learning-Based Classification of Remote Sensing Image has attracted extensive interest. We trained a multilayer deep learning network to classify the 8 thousand unlabeled remote sensing images from Internet into the 600 different classes. In order to improve the efficiency, and shorten the experiment time, we also used pre-trained model, NVIDIA GTX970 GPU, 32GB internal memory, 10T hard-disk. We achieved error rates of 9.7% which work went relatively well than the traditional machine learning techniques. Deep learning-based network can achieve the classification of unlabeled data without any manual intervention. Compared to those usual machine learning algorithm, accuracy and speed of deep learning-based classification network is more faster and accurately.

**Key words:** Deep Learning network (DLN), remote sensing image, classification.

---

## 1. Introduction

In geography observation, land planning, military techniques and many other fields, demand for object recognition sharply increases. In the recent 20 years, various types classification algorithm have played a role of object classifiers with acceptable efficiency and low cost, but high resolution remote sensing image also brought coexistence of the opportunities and challenges to the classification issues, the effect is also positive and negative. In general, conventional classification algorithms of remote sensing image mainly based on spectrum, and texture, and profile, included regression-based, k-means, support vector machines, decision tree model, expectation-maximization, k-nearest neighbor classification, AdaBoost, naive bayes and maximum likelihood.

The output of conventional machine learning algorithms rely heavily on art of representation or feature selecting method of input data, which can be called an acrobatic of scientific research rather than a skill which is used in production and manufacturing, application scenarios are extremely narrow and can not be separated from some specific experienced person.

Use visual attention mechanism like human, which uses features of input image from top to bottom to obtain saliency map and abstracted salient objects from remote sensing image, and gets satisfying result (Gang Xu, Hong Huo, Tao Fang, Deren Li, 2007) [1].

Assemble several co-existing independent objects called Compound object into new feature sets, and work better in image retrieval area (Dihua Guo, Hui Xiong, Vijayalakshmi Atluri, Nabil R. Adam, 2009) [2].

Object recognition of HRRS (high resolution remote sensing images) can be based on domain knowledge

provided by industry experts. Experiments show good results and high success rate when distinguish objects in urban and rural areas (Nicolas Durand, Sebastien Derivaux, Germain Forestier, Cedric Wemmert, Pierre Ganc, arski, 2007) [3].

Traditional image processing algorithm work well on LRRS (low resolution remote sensing images) but obtain low success rate HRRS (high resolution remote sensing images). Based on feature template, achieve image and template matching, and accurate object recognition (Jun Zhanga, Xiyuan Zhou, 2004) [4].

The NDI and Morphological Operators based method for shadow detection is quite effective and obtain high accuracy (Krishna Kant Singh Kirat Pal M. J. Nigam, 2012) [5].

Support Vector Machine(SVM) with basic nature as a binary recognizer work better than the other methods when abstract various object similar to vegetation, water, and oil spill from remote sensing image(Wardaya, 2014) [6].

Object recognition depended on what combines local and global features of objects, and use a ranking support vector machine (RSVM). The experiment proved that its obvious advantages on HRRS (high resolution remote sensing images) over some other object recognition methods (Xiao Bai, Huigang Zhang, and Jun Zhou, 2014) [7].

Object classification based on support degree of evidence theory and D-S decision tree method, and the ability of reclassification and high accuracy make the method work better than Bayesian on land use classification (LI Xuerong, XING Qianguo, KANG Lingyan, 2010) [8].

For a Reinforced concrete three-layers buildings, only if it did not change in essence, including but not limited to illumination change, change of vegetation in rainy and foggy days, facade renovation, and external decoration change, the corresponding characteristics output remain unchanged, which means core structure unchanged.

But if it did change in essence, including but not limited to, sealing the balcony for illegal structures, and heightening room, private residential green space, which will trigger enforcement warning soon.

Deep network imitation of human brain, which included multi-level abstraction and self-learning features, abstract non-monitoring characteristics from a large number of urban image data, and had high practical application value better than other machine learning methods.

## 2. Algorithm Analysis

For instance, an object recognition assignment, we need to differentiate among hundreds or more different objects based on remote sensing images. It is necessary we prepared a large number of training samples for traditional supervised machine learning algorithm, which means a variety of forms of remote sensing images belong to hundreds different kinds objects which had been manually labeled into accurate classification. It is accepted to be very expensive and extremely time-consuming we identify and label a large number Trained sample, with accurate tags.

Deep learning-based network consists of the following layers, auto-encoder layer, convolutional layer, pooling layer, and fully-connected layer. To perfectly implement auto-encoder's parameter optimization tuning, we use the back-propagation algorithm and derive the equations of overall cost of back-propagation as show below:

$$C(V, d) = \left( \frac{1}{n} \sum_{j=1}^n \frac{1}{2} \|k_{v,d}(x^{(i)} - y^{(i)})\|^2 \right) + \frac{\gamma}{4} \sum_{u=1}^{m_u-1} \sum_{j=1}^{s_j} \sum_{i=1}^{s_u+1} (V_{ij}^{(u)})^2 \quad (1)$$

We obtain:

$$C_{sp} = C(V, d) + \beta \sum_{i=1}^{s^2} \rho \log \frac{\rho}{\rho_i} + (1 - \rho) \log \frac{1 - \rho}{1 - \rho_i} \tag{2}$$

$$\delta_j^{(2)} = \left( \left( \sum_{i=1}^{s^2} V_{ij}^{(2)} \delta_0^{(3)} + \beta \left( \frac{-\rho}{\rho_i} + \frac{1 - \rho}{1 - \rho_i} \right) \right) g'(z^2) \right) \tag{3}$$

At last, an advanced logistic model, namely softmax model is applied in multi-class classification and recognition, and after that we obtain an available model which play a role in multi-class classification.

$$k_w(x^{(j)}) = \begin{bmatrix} p(y^{(j)} = 1 | x^{(j)}; w) \\ p(y^{(j)} = 2 | x^{(j)}; w) \\ p(y^{(j)} = 3 | x^{(j)}; w) \\ \cdot \\ \cdot \\ \cdot \\ p(y^{(j)} = h | x^{(j)}; w) \end{bmatrix} \tag{4}$$

Because the cumulative sum of probability is equal to 1, we obtain:

$$k_w(x^{(j)}) = \frac{1}{\sum_{i=1}^h e^{w_i^T x^{(j)}}} \begin{bmatrix} e^{w_1^T x^{(j)}} \\ e^{w_2^T x^{(j)}} \\ e^{w_3^T x^{(j)}} \\ \cdot \\ \cdot \\ \cdot \\ e^{w_h^T x^{(j)}} \end{bmatrix} \tag{5}$$

Currently there are hundreds kinds of innumerable free downloading images, and we designed a web crawler to gain a lot of free pictures. These free images are clearly distinct from those dataset of manual annotation, which may include certain objects or not containing any objects, in other words, these images do not label. By multilayer convolutional neural network, we obtain a pre-trained model to classify the 79 thousand high-resolution remote sensing images but without tag downloading from the Internet into the 319 different classes.

### 3. Experimental Results and Analysis

In series experiments, we focus on aircraft detection on high-resolution and optical remote sensing images. Based on caffe and cuda, we trained a multilayer deep learning networks to classify the 8 thousand unlabeled remote sensing images from Internet into the 600 different classes.

We achieved error rates of 9.7% which work went relatively well than the traditional machine learning techniques. In order to increase the efficiency, and shorten the experiment time, we also used pre-trained model, NVIDIA GTX970 GPU, 32GB internal memory, 10T Hard-Disk. All these models and devices are huge

time savers, and it makes it an extremely desirable classification platform.

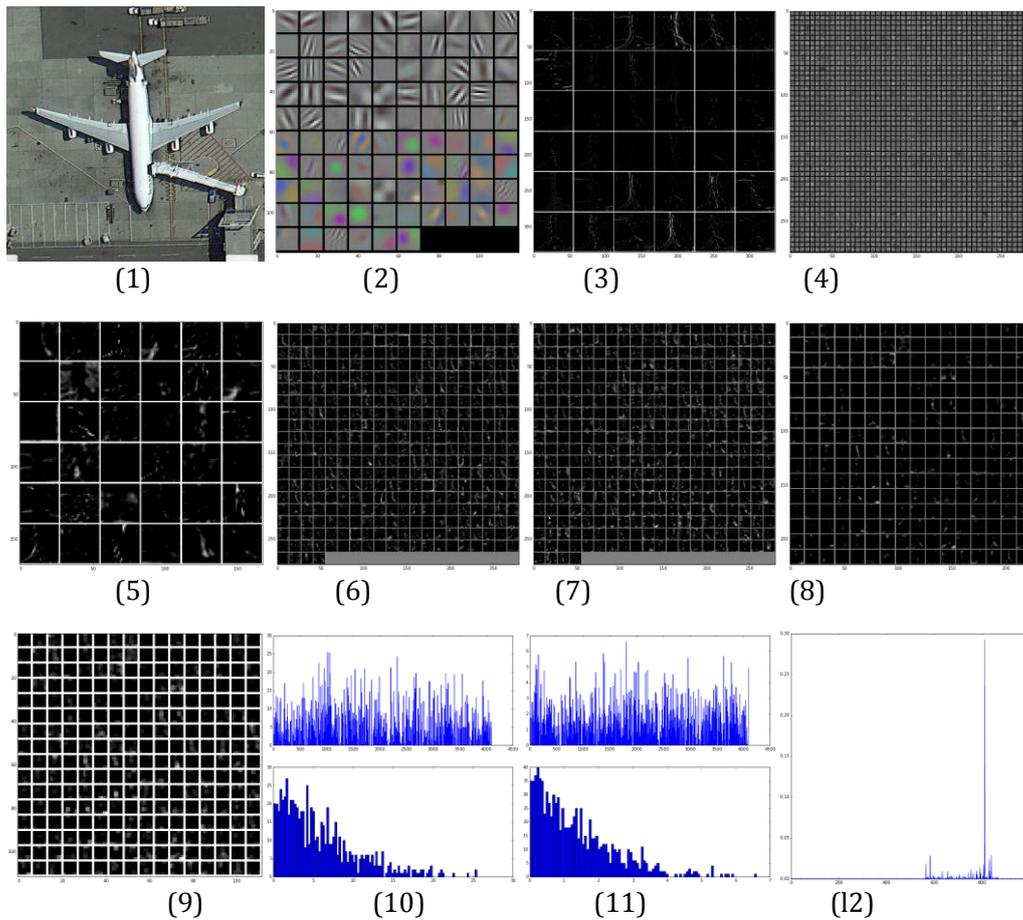


Fig. 1. Intermediate state parameter diagram and classification results: (1) The input image. (2)The first layer output, conv1. (3) The second layer filters, conv2. (4) The second layer output, conv2. (5) The third layer output, conv3. (6) The fourth layer output, conv4. (7) The fifth layer output, conv5. (8) The fifth layer after pooling, pool5. (9) The first fully connected layer, fc6. (10) The second fully connected layer, fc7. (11) The final probability output. (12) The top 5 predicted labels.

The process from the first layer to the seventh layer of the classification of an image is given in Fig. 1-1(1-12). Fig. 1(1) show the input raw image, Fig. 1(2) show the output of first layer named conv1 layer, Fig. 1-3) show the filters of second layer named conv2 layer, Fig. 1(4) show the output of second layer named conv2 layer, Fig. 1 (5) show the output of third layer output named conv3 layer, Fig. 1(6) show the output of fourth layer named conv4 layer, Fig. 1(7) show the output of fifth layer named conv5 layer, Fig. 1(8) show the fifth layer after pooling named pool5 layer, Fig. 1(9) show the first fully connected layer named fc6 layer, Fig. 1-10) show the second fully connected layer named fc7 layer, Fig. 1(11) show the output of final probability, and Fig. 1(12) show the top 5 detected labels.

Top 1 detected label is jet aircraft, the top 2 detected label is space shuttle, the top 3 detected label is liner, ocean liner, the top 4 detected label is aircraft carrier:, and the top 5 detected label is aircraft.

We found that all of selected features are basically on the edge of the line of different objects of the image, and the shape of these line segments is similar, but the direction is different.

#### 4. Conclusion

Deep learning-based network can achieve the classification of unlabeled data without any manual

intervention. Compared to those usual machine learning algorithm, accuracy and speed of deep learning-based classification network is more faster and accurately.

That is, complex graphics, often by some of the basic structure of the composition. A map can be made by using several kinds of orthogonal edges linear representation. For example, the sample  $x$  can use the orthogonal edges in three according to the weighted and blended.

And other basic edge on the elementary of the basic structure of the object does not contribute to the weight, can be set to 0. From the experimental results, we can see that deep learning-based classification network can extract features from the image, achieve hierarchical abstraction, founded on a certain number of labeled data-set and generate a pre-training model.

In general, providing an extension to the existing analytical methods of classification of high-resolution remote sensing image is just one part of this paper, the other part is making it easy-using and ease of learning access of deep learning-based classification method.

## Acknowledgment

Supported by Provincial Natural Science Foundation of Hunan (2012JJ5029)

Supported by Science and Technology Planning Project of Loudi (2014KJ04)

## References

- [1] Xu, G., Huo, H., Fang, T., *et al.* (2007). Extracting salient object from remote sensing image based on guidance of visual attention. *Proceedings of International Symposium on Multispectral Image Processing and Pattern Recognition. International Society for Optics and Photonics.*
- [2] Guo, D., Xiong, H., Atlur, V., *et al.* (2009). Object discovery in high-resolution remote sensing images: A semantic perspective. *Knowledge & Information Systems, 19(19)*, 211-233.
- [3] Durand, N., Derivaux, S., Forestier, G., *et al.* (2007). Ontology-based object recognition for remote sensing image interpretation. *Proceedings of IEEE International Conference on Tools with Artificial Intelligence* (pp. 472-479).
- [4] Jun, Z. G., & Zhou, X. Y. (2004). Object recognition based on template correlation in remote sensing image. *World Congress on Intelligent Control & Automation, 5(5)*, 4070-4074.
- [5] Kantsingh, K., Pal, K. J., & Nigam M. (2012). Shadow detection and removal from remote sensing images using NDI and morphological operators. *International Journal of Computer Applications, 42(10)*, 810.
- [6] Wardaya, P D. (2014). Support vector machine as a binary classifier for automated object detection in remotely sensed data. *Proceedings of IOP Conference Series: Earth and Environmental Science. IOP Conference Series: Earth and Environmental Science.*
- [7] Bai, X., Zhang, H., & Zhou, J. (2014). VHR object detection based on structural feature extraction and query expansion. *IEEE Transactions on Geoscience & Remote Sensing, 52(10)*, 6508-6520.
- [8] Li, X., & Xing, Q. (2010). Remote sensing image classification method based on evidence theory and decision tree. *Proceedings of SPIE - The International Society for Optical Engineering.*



**Jianmin Liu** was born in Fujian Province of China, in Dec. 1976. He received the B.S. degree from the University of Hunan (UHN), Changsha, in 1997 and the M.S. degree from the University of Xiamen (UX), Xiamen, in 2011. He is currently pursuing the Ph.D. degree with the Department of GIS of the University of Central South (UCS). His research interests include remote sensing, GIS, and machine learning.