CrossMark

# Transfer Learning for Classification of Optical Satellite Image

**MaoYang Zou**[1,2] · **Yong Zhong**[1]

**Abstract** Deep convolutional neural network (DCNN) has achieved great success in the classification of natural images, but it requires numerous labelled data for training. In the absence of a large number of optical satellite images and labelled data, how to guarantee the effect of classification of the optical satellite images with DCNN? In this case, this paper has discussed how to fine-tune a pre-trained DCNN in a layer-wise manner by transfer learning. In our experiment, DCNN is pre-trained with ImageNet which is a large labelled dataset of natural images, and then optical remote sensing images are used to fine-tune the learnable parameters of pre-trained DCNN. The experimental results show that transfer learning is feasible to deal with the above problem. In the process of transfer training, if the second half of the layers are fine-tuned, compared with the fine-tuning of the entire network, the almost same accuracy can be achieved, but the convergence is more rapid. The experimental results provide a solution for how to achieve the incremental classification performance in practical applications.

**Keywords** Optical remote sensing image · Deep convolutional neural network · Classification · Transfer learning · Fine-tuning

---

---

✉ MaoYang Zou
zoumy@cuit.edu.cn

1 Chengdu Institute of Computer Application, University of Chinese Academy of Sciences, No.9 The 4th Section of South Renmin Road, Chengdu 610041, China

2 Chengdu University of Information Technology, No.24 Block 1, Xuefu Road, Chengdu 610225, China

 Springer

# 1 Introduction

In recent years, the neural networks which simulate the human brain to analyze and learn has become one of the research focuses and has achieved great success in the fields including speech recognition, natural language and image processing.

In the field of image processing, Deep Convolutional Neural Network (DCNN), which has forward and backward passes, has already outperformed other approaches in natural image classification [1, 2], natural image segmentation [3], and object detection [4], etc. If DCNN is applied to the classification of optical remote sensing images, what will happen? And what can we do to achieve good results?

Training models from scratch in DCNN requires a great deal of labelled training data, because the limited availability of labelled data may lead to an undesirable local minimum for the cost function. Some remote sensing images are scarce to obtain and expensive to be labelled. If only a few thousand images are input to train DCNN from scratch, it would easily yield a overfitting. Transfer learning is proposed as the solution. More specifically, a DCNN should be pre-trained in classification on a very large dataset, such as ImageNet which is a natural image dataset. Then the weights of the pre-trained DCNN is used as an initial value of a new network, in other word, the pre-trained DCNN is regarded as a new network. This means through transfer learning, the pre-trained models are applied to new task as a feature generator. The new network is trained by remote sensing images unceasingly and Back Propagation algorithm (BP) is used to adjust its weights, until the new network can classify remote sensing images accurately. The primary reason for success of this approach is that the low-level features can be preserved from one dataset to another and reused without training from scratch, even when the final classification object is different [5].

There are additionally two advantages in this case:

- It saves some cost for manual label. The approach of DCNN is supervised learning, which requires manual label of data, but transfer learning with fine-tuning uses the pre-trained classifier to train a new classifier and the useful information of the source data is effectively utilized, so the demand for new label is reduced.
- It greatly accelerates the convergence speed and saves the training time.

How can you achieve better performance through fine-tuning in the transfer learning process? The papers [5, 6] propose that the features transferred from different tasks are better than random weights for initializing network weight, even when the features are transferred from distant objects. The above factor is considered in the design of our experiment. Unlike the experiment in the paper [5] in which part of the layers are initialized randomly, in our experiment, the weights of all layers are initialized by the features transferred from other tasks instead of random weights. And then, the pre-trained DCNN is fine-tuned in a layer-wise manner for optical remote sensing image classification until the best performance is found.

Contributions of this paper are shown as follows:

- The paper discusses how to classify remote sensing images in the absence of a great deal of labelled data.
- The experiment finds a way to fine-tune a pre-trained DCNN in a layer-wise manner to obtain incremental performance.

## 2 Related Work

### 2.1 Remote Sensing Image Classification

The resolution of the optical remote sensing image has become higher and higher, carrying rich information. Many researchers extract the features by the traditional machine learning method, and identify them by the classifier [7], such as Linear regression, neural network [8], Bayesian network, fuzzy clustering [9], SVM based on statistical learning [10],etc. For example, Zhu et al. [10]extracted the Local Binary Pattern (LBP), shape and the gray-scale distribution feature, and used the support vector machine (SVM) to classify the ships.

DCNNs have been applied to remote sensing image classification [11–13] and object detection [14, 15] , which has achieved a certain success. Chen et al. [11] "present a new all-convolutional networks (A-ConvNets), which only consists of sparsely connected layers, without fully connected layers being used", "and achieve an average accuracy of 99% on classification of ten-class targets". Luus et al. [12] propose that " The end-to-end learning system learns a hierarchical feature representation with the aid of convolutional layers to shift the burden of feature determination from hand-engineering to a deep convolutional neural network (DCNN)." " It is shown that a single DCNN can be trained simultaneously with multiscale views to improve prediction accuracy over multiple single-scale views."

DCNN has been applied to remote sensing image classification as mentioned above. We focus on how to classify the optical remote sensing images through transfer learning and  with the lack of a great deal of labelled data. Currently, few research results were published in this area, concerning remote sensing images.

### 2.2 Deep Convolution Neural Network

The excellent performance of DCNN in image processing can be attributed to its capability of extracting a set of discriminating features on multiple levels. "The kth output feature map $Y_k$ can be computed as: $Y_k = f(W_k^* x)$, where the input image is denoted by x; the convolutional filter related to the k$th$ feature map is denoted by $W_k$; the multiplication sign in this context refers to the 2D convolutional operator, which is used to calculate the inner product of the filter model at each location of the input image; and f() represents the nonlinear activation function" [2, 16].

DCNN has four obvious advantages for image processing as follows:

Firstly, The neurons in the receptive field are not fully connected. In addition, weight with a convolution filter in the same layer is shared.Receptive field and parameter sharing can reduce the number of training parameters. Consequently, the dimension disaster is minimized [17].

Secondly, The alternation of the convolution layer and the pooling layer makes the DCNN sensitive to the local small features [17].

Thirdly, ReLu is an activation function defined as the positive part of its argument: $f(x) = \max(0,x)$ ,where x is the input to a neuron. ReLu can alleviate gradient diffusion in the DCNN, and it can make the output of the network have high sparsity [18].

Finally, the optimization methods about network weights can greatly improve the neural network performance.For example, the model of MBGD(Mini-batch gradient descent) update frequency is higher than BGD (Batch gradient descent). The batched updates of MBGD provide a computationally more efficient process than SGD(Stochastic gradient descent). Furthermore, Konecny et al. [19] proposed mini-batch semi-stochastic gradient descent and they prove that they can "reach any predefined accuracy with less overall work than without mini-batching."

## 2.3  Transfer Learning and Fine-Tuning

The traditional machine learning assumes that the training dataset and the testing dataset obey the same data distribution. But in many cases, the assumption of same distribution is not valid. The goal of transfer learning is to learn the knowledge from a dataset to facilitate the learning tasks in the new dataset. Therefore, it is not necessary to make assumption of same distribution for transfer learning.

At present, transfer learning can be divided into three categories according to whether there are labelled data in the source dataset and the target dataset [20]: inductive transfer learning [21], transductive transfer learning [22], and unsupervised transfer learning [23, 24].

Transfer learning strategies depend on various factors, but the two most important ones are the size of the new dataset, and its similarity to the original dataset. For the optical remote sensing images that we want to classify and the optical images of ImageNet, the low-level features of the two kinds of images share very strong similarity, which helps to transfer. We mainly consider the size of the dataset.

For different database sizes, the pre-trained DCNN has two approaches to apply the pre-trained model to the new tasks of image classification.

- The first approach: If the new dataset is small, the pre-trained DCNN weights is used as a fixed feature extractor wherein the network remains the same [25, 26]. It extracts the feature vector before the last layer of fully-connected layers and then trains a linear classifier for classification.
- The second approach: If the new dataset is relatively large, we fine-tune pre-trained DCNN with the new dataset [27], and the Back Propagation algorithm is used to fine-tune the weights again. Finally the DCNN is updated to solve a new problem [28, 29], and the accuracy can be higher than that of the first approach.

This paper aims to seek for the best fine-tuning method, so the second approach is adopted.

Fine-tuning refers to the process in which parameters of a model must be adjusted very precisely. Fine-tuning is regarded as one of the tricks in machine learning. In the experiment of the paper [5] about DCNN fine-tuning, there are two datasets named dataset A and dataset B. The first n (n ranges from 1 to 7) layers are copied from dataset A and frozen. The left higher layers are initialized randomly and trained by dataset B. The paper points that "the extent to which transfer is successful has been carefully quantified layer by layer."

## 3 Experiments

### 3.1 Samples

There are a total of 2100 remote sensing satellite images in UC Merced Land Use Dataset, Specifically, which include 21 categories of remote sensing satellite images, and each category has 100 pictures. This dataset has some characteristics that are not good for image classification, because some images from different categories are very similar, as shown in Figs. 1 and  2, but some images in the same category are quite different, as shown in Figs. 3 and 4.

With the number of images in UC Merced Land Use Dataset, if the network is trained from scratch, it will certainly yield a serious overfitting. Even if we adopt the approach of pre-training and transfer learning, the data still need to be augmented, so we quadrupled UC Merced Land Use Dataset by means of horizontally flipping, color jittering(adjusting the image brightness, saturation or contrast), random crop, shift and so on.



**Fig. 1**  Mobile Home Park 10

**Fig. 2** Buildings 25
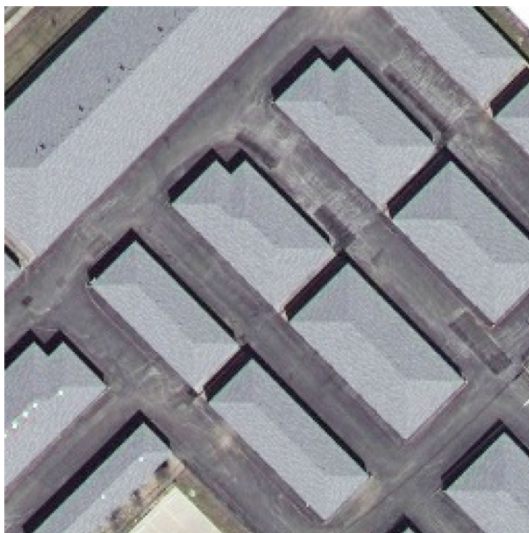


**Fig. 3** Storage Tanks 73



### 3.2 Network Architecture and Parameters

Deep convolutional neural network has several typical network structures. AlexNet is one of the most famous convolutional neural networks. It is designed by Geoffrey Hinton and Alex Krizhevsky and wins the championship in 2012. It is used for ImageNet classification and almost halved the error rate of best algorithm, so it has attracted the attention of computer vision community. AlexNet consists of 5 convolutional layers and 3 fully connected layers. In order to adapt to the classification

**Fig. 4** Storage Tanks 97
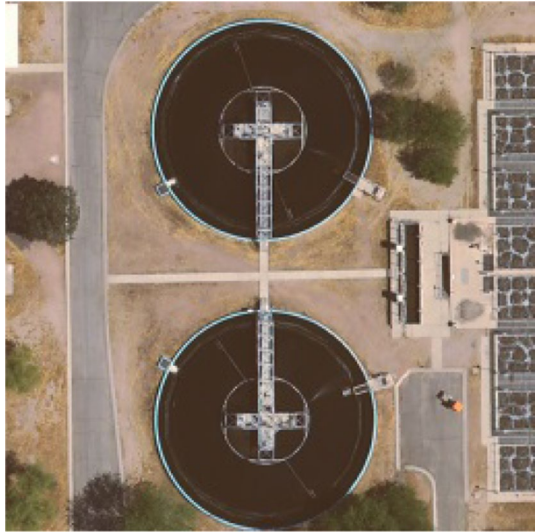


task, we modify the last full connected layer to 21 nodes, each of which represents a category in the image dataset.

The experimental network and parameters are shown in the Fig. 5. For each layer, the upper part of the Fig. 5 represents the input, the middle part represents the parameters, and the lower part represents the output.

It must be noted that the learning rate for the last layer is 0.01, the learning rate for the rest fine-tuned layers is 0.001, the learning rate for the frozen layer is 0, and learning_rate_decay is 0.95.

### 3.3 Experimental Approach

The 80% of the dataset is used as the training set, and 20% is the testing set. All of the image chips taken from the same wide area image are included in the same training or test set. Five-fold cross-validation is used ten times, and the average value of accuracy of the results is calculated as the evaluation criterion.

The experiment differs from [25, 26] wherein the network remains the same and serves as a feature generator. It differs from [28] wherein the entire network was fine-tuned at once. It also differs from [5] wherein the first n layers are transferred from other network and the left higher layers are initialized randomly. In our experiment, the weights of all layers are transferred from AlexNet pre-trained by ImageNet, because the features transferred from different task are better than random weights for initializing network weight. We have conducted eight rounds of experiments after AlexNet is pre-trained. In the first round, the parameters of the last layer of pre-trained AlexNet is trained with UC Merced Land Use Dataset until convergence while freezing all the parameters in the previous layers, and the accuracy is calculated in this case. Similarly, in the second round, we train the last two layers of pre-trained AlexNet and freeze all the parameters of other layers in the update
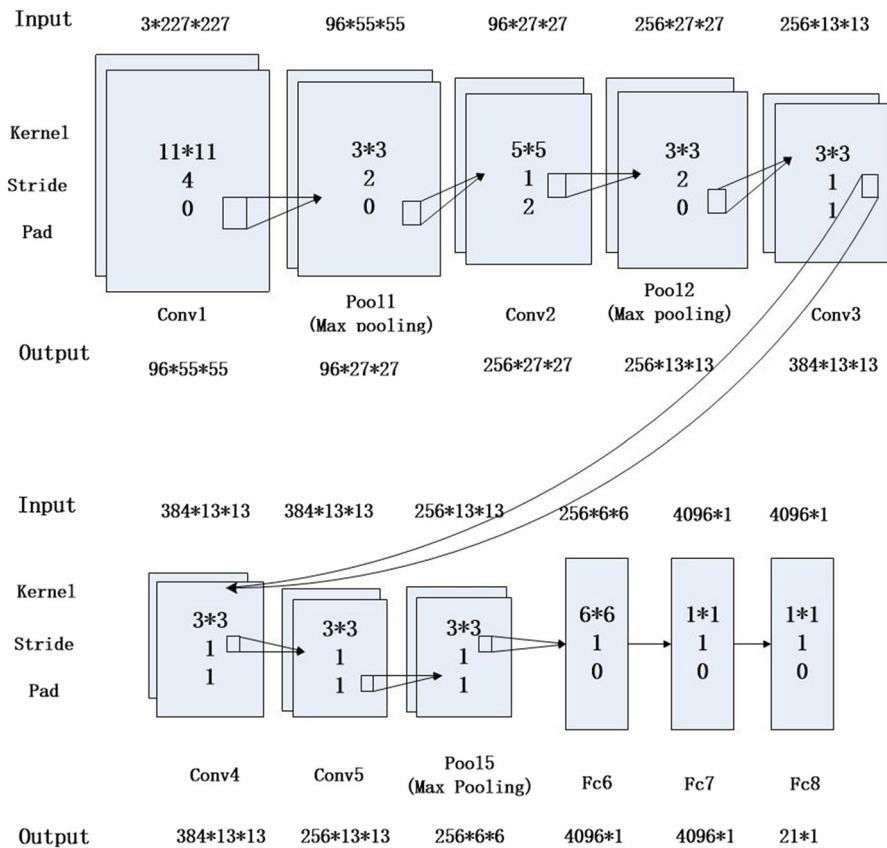
**Fig. 5** Network architecture

process. Next, the training incrementally includes one more layers in the update process. In this way, finally the entire network undergoes fine-tuning at once. Overall, the network is fine-tuned in a layer-wise manner after AlexNet is pre-trained by the natural images.

## 3.4 Experimental Results

For an image, AlexNet shows its possibilities of belonging to each of the 21 categories. With our statistics, the Fig. 6 shows the classification accuracy of images in each category in five typical cases. The five typical cases are training from scratch without transfer learning (called "Training from Scratch"), transfer learning without fine-tuning(called "Transfer Learning without Fine-tuning"), the fine-tuning of the last two layers in AlexNet (called "Fine-tuned AlexNet: fc7 - fc8"), the fine-tuning of layers from fc8 to conv5 (called "Fine-tuned AlexNet:conv5-fc8"), and the fine-tuning of the entire network (called "Fine-tuned AlexNet:conv1-fc8").
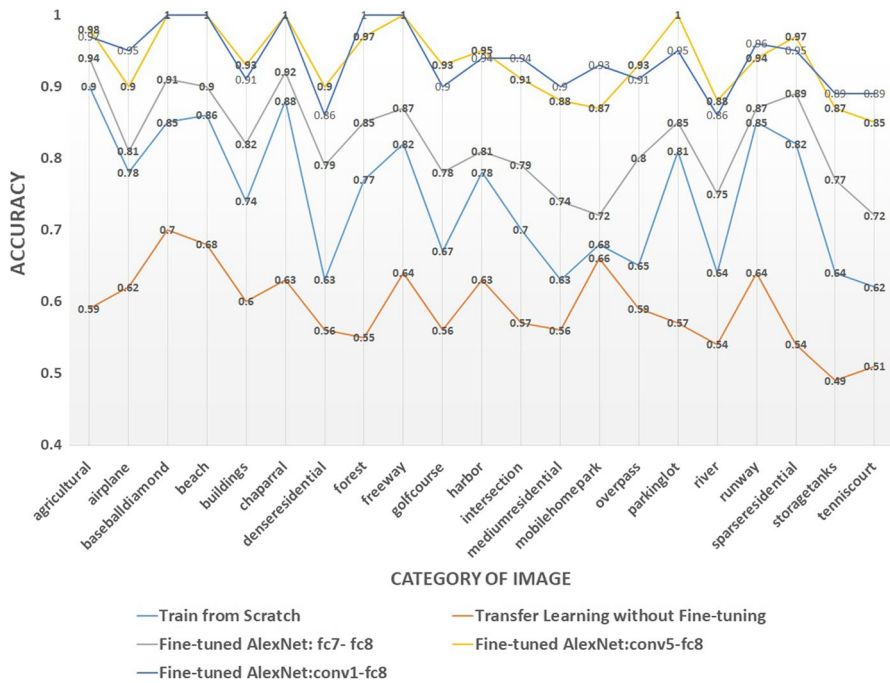
**Fig. 6** Classification accuracy

The horizontal axis of Fig. 6 is the image category of UC Merced Land Use Dataset, and the vertical axis is the accuracy.

In Fig. 6, tenniscount, denseresidential, golfcourse, mediumresidential, and storagetanks are the categories with low classification accuracy in our experiment. Overall, for all categories, the accuracy of training from scratch is 74.86%, the accuracy of transfer learning without fine-tuning is only 59.19%, the accuracy of "Fine-tuned AlexNet: fc7 - fc8" is 82.38%, the accuracy of "Fine-tuned AlexNet:conv5 - fc8" is 93.62%, and the accuracy of "Fine-tuned AlexNet:conv1 - fc8" is 93.86%. The data show that transfer learning with fine-tuning is feasible, because the accuracies of "Fine-tuned AlexNet:conv5 - fc8" and "Fine-tuned AlexNet:conv1 - fc8" are both high. The accuracies of "Fine-tuned AlexNet:conv5 - fc8" and "Fine-tuned AlexNet:conv1 - fc8" are very similar, and the accuracy of "Fine-tuned AlexNet:conv1 - fc8" is 0.24% higher.

# 4 Discussion

## 4.1 Why Is AlexNet Chosen as Experimental Network?

There are three reasons why we choose AlexNet as the experimental network:

- A large number of experiments show that it has excellent classification effect.
- A pre-trained AlexNet model was available in the Caffe library.
- AlexNet has more layers than networks such as LeNet and CompactNet, which are too shallow to obtain sufficient image features. Of course, there are still GoogleNet and VGGNet with deeper network structure, but they are slow to converge. The purpose of our experiment is to find the appoach of fine-tuning to obtain incremental performance on the pre-trained network, so AlexNet is a reasonable choice.

### 4.2  Setting Up the Learning Rate

Setting up the learning rate is a key point in the experiment. In order to have better performance for gradient descent, we need to set the learning rate in an appropriate range. If learning rate is too small, the algorithm takes a long time to converge. On the contrary, the excessive learning rate will lead the object function to oscillate near the lowest point.

In the process of training in AlexNet, we find that when the loss oscillates in a certain area but does not converge, if we lower the learning rate by one order of magnitude, the loss will drop in the one or two epoch, the accuracy will change dramatically at this moment. To perform statistical comparisons,in our experiment, the learning rate of the last layer is 0.01, the learning rate of the other fine-tuned layers is 0.001, the learning rate of frozen layer is 0, and learning-rate-decay is 0.95.

### 4.3  Analysis for Experimental Results

We train the AlexNet on the large dataset to get the pre-trained model and then conduct feature-based transfer learning. According to the experimental results, transfer learning with fine-tuning is feasible, because no matter which method of fine-tuning is adopted, pre-trained AlexNet with fine-tuning performs much better than the AlexNet trained from scratch and the transfer learning without fine-tuning. By comparison, we find that the pre-trained model has some advantages over the randomly initialized model. For example, the pre-trained model clearly extracts the high-level features of images, such as the edge features and shape features, and removes the background of objects. The randomly initialized model only simply smooth the image to a certain extent, and the background is still prominent.

We analyzed the experimental results of fine-tuning, and the 4096 dimension feature extracted from fc7 of AlexNet is analyzed by means of dimension reduction. We find that the categories of lower accuracy, such as tenniscount, denseresidential, mediumresidential and storagetanks, are close to each other, in other words, the distance is small between categories, so it is difficult to distinguish.

We then analyzed the classification accuracy of various fine-tuning approaches. The accuracies are almost the same for "Fine-tuned AlexNet:conv5-fc8" and "Fine-tuned AlexNet:conv1-fc8". The reason is that the initial layers describe the general features of the image, such as color and edge. The last few layers, such as the last

fully connected layers, describe the high order features which are related to image classification task. Thus the fine-tuning for the first few layers of the AlexNet is not absolutely necessary, but the fine-tuning for the last few layers of the AlexNet is very important.

### 4.4 Limitations of The Experiment

Finally, we discuss the limitations of the experiment. Because we use pre-training network, the model architecture is limited slightly. When we fine-tune the DCNN in Caffe, the network structure should be consistent with pre-training model to ensure the parameters are correctly loaded, so we can not remove the layer of pre-training network arbitrarily.

## 5 Conclusion

The experiment shows that transfer learning with fine-tuning is feasible to classify optical remote sensing images in the absence of a large number of labelled images. In the experiments of fine-tuning a pre-trained DCNN in a layer-wise manner to get good performance, we find that the optimal solution is to freeze the first half of layers and fine-tune the second half of layers. The performance of "Fine-tuned AlexNet:conv5-fc8" is almost the same with that of "Fine-tuned AlexNet:conv1-fc8", but the former takes shorter training time. Our experiment provides the solution for how to achieve the good classification performance in practical applications.

## References

1. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *NIPS'12 Proceedings of the 25th international conference on neural information processing systems*, Lake Tahoe, Nevada December 03–06, 2012, (Vol. 1, pp. 1097–1105).
2. Rawat, W., & Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. *Neural Computation*, *29*(9), 2352–2449. https://doi.org/10.1162/NECO_a_00990 .
3. Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, Boston, Massachusetts June 08–10, 2015 (pp. 3431–3440). https://doi.org/10.1109/CVPR.2015.7298965
4. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, Montreal, Canada December 07–12, 2015 (pp. 91–99).
5. Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks?. Montreal, Canada December 08 - 13, 2014, In *Advances in neural information processing systems* (Vol. 27, pp. 3320–3328).
6. Long, M., Cao, Y., Wang, J., & Jordan, M. (2015). Learning transferable features with deep adaptation networks. In *Proceedings of the 32nd international conference on machine learning*, Lille, France July 06–11, 2015 (Vol. 37, pp. 97–105).
7. Sun, X., Fu, K., Long, H., Hu, Y., Cai, L., & Wang, H. (2008, July). Contextual models for automatic building extraction in high resolution remote sensing image using object-based boosting

method. In *Geoscience and remote sensing symposium*, 2008. IGARSS 2008. IEEE International (Vol. 2, pp. II-437). IEEE. https://doi.org/10.1109/IGARSS.2008.4779022

8. Qu, J. S., Qu, S. B., & Wang, Z. J. (2009). Feature-based fuzzy-neural network approach for target classification and recognition in remote sensing images. *Journal of Remote Sensing*, *13*(1), 68–74.

9. Zhao, C., & Qi, B. (2012). Hyperspectral image classification based on fuzzy kernel weighted c-means clustering. *Chinese Journal of Scientific Instrument*, *33*(9), 2016–2021.

10. Zhu, C., Zhou, H., Wang, R., & Guo, J. (2010). A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features. *IEEE Transactions on geoscience and remote sensing*, *48*(9), 3446–3456. https://doi.org/10.1109/TGRS.2010.2046330.

11. Chen, S., Wang, H., Xu, F., & Jin, Y. Q. (2016). Target classification using the deep convolutional networks for SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(8), 4806–4817. https://doi.org/10.1109/TGRS.2016.2551720.

12. Luus, F. P., Salmon, B. P., van den Bergh, F., & Maharaj, B. T. J. (2015). Multiview deep learning for land-use classification. *IEEE Geoscience and Remote Sensing Letters*, *12*(12), 2448–2452. https://doi.org/10.1109/LGRS.2015.2483680.

13. Zhou, Y., Wang, H., Xu, F., & Jin, Y. Q. (2016). Polarimetric SAR image classification using deep convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters*, *13*(12), 1935–1939. https://doi.org/10.1109/LGRS.2016.2618840.

14. Chen, X., Xiang, S., Liu, C. L., & Pan, C. H. (2014). Vehicle detection in satellite images by hybrid deep convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters*, *11*(10), 1797–1801. https://doi.org/10.1109/LGRS.2014.2309695.

15. Tang, J., Deng, C., Huang, G. B., & Zhao, B. (2015). Compressed-domain ship detection on spaceborne optical image using deep neural network and extreme learning machine. *IEEE Transactions on Geoscience and Remote Sensing*, *53*(3), 1174–1185. https://doi.org/10.1109/TGRS.2014.2335751.

16. Yu, D., Wang, H., Chen, P., & Wei, Z. (2014). Mixed pooling for convolutional neural networks. In *the 9th international conference on rough sets and knowledge technology (RSKT'14),* Shanghai, China October 24–26, 2014 (pp. 364–375). Springer, Cham. https://doi.org/10.1007/978-3-319-11740-9_34

17. He, J., Zou, M., & Liu, P. (2017). Convolutional neural networks for chinese sentiment classification of social network. In *2017 IEEE international conference on mechatronics and automation (ICMA)*, Takamatsu, Japan August 06-09, (pp. 1877–1881). IEEE. https://doi.org/10.1109/ICMA.2017.8016104

18. Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, Haifa, Israel June 21–24, 2010 (pp. 807–814).

19. Konecny, J., Liu, J., Richtarik, P., & Takac, M. (2016). Mini-batch semi-stochastic gradient descent in the proximal setting. *IEEE Journal of Selected Topics in Signal Processing*, *10*(2), 242–255.

20. Zhuang, F. Z., LUO, P., HE, Q., & Shi, Z. (2015). Survey on transfer learning research. *Journal of Software*, *26*(1), 26–39.

21. Fengmei, W., Jianpei, Z., Yan, C., & Jing, Y. (2014). FSFP: Transfer learning from long texts to the short. *Applied Mathematics & Information Sciences*, *8*(4), 2033. https://doi.org/10.12785/amis/080462.

22. Dai, W., Xue, G. R., Yang, Q., & Yu, Y. (2007). Co-clustering based classification for out-of-domain documents. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, California, USA August 12–15, 2007 (pp. 210-219). ACM. https://doi.org/10.1145/1281192.1281218

23. Samanta, S., Selvan, A. T., & Das, S. (2013). Cross-Domain clustering performed by transfer of knowledge across domains. In 2013 Fourth *International conference on computer vision, pattern recognition, image processing and graphics (NCVPRIPG)*, Jodhpur, India December 18–21, 2013 (pp. 1–4). IEEE. https://doi.org/10.1109/NCVPRIPG.2013.6776213

24. Dai, W., Yang, Q., Xue, G. R., & Yu, Y. (2008, July). Self-taught clustering. In *Proceedings of the 25th international conference on machine learning*, Helsinki, Finland July 05–09, 2008 (pp. 200–207). ACM. https://doi.org/10.1145/1390156.1390182

25. Sharif Razavian, A., Azizpour, H., Sullivan, J., & Carlsson, S. (2014). CNN features off-the-shelf: an astounding baseline for recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, Columbus, Ohio June 23–28, 2014 (pp. 806–813). https://doi.org/10.1109/CVPRW.2014.131

26. Penatti, O. A., Nogueira, K., & dos Santos, J. A. (2015). 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*, Boston, Massachusetts June 11–12,2015 (pp. 44–51).
27. Maggiori, E., Tarabalka, Y., Charpiat, G., & Alliez, P. (2017). Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(2), 645–657. https://doi.org/10.1109/TGRS.2016.2612821.
28. Azizpour, H., Sharif Razavian, A., Sullivan, J., Maki, A., & Carlsson, S. (2015). From generic to specific deep representations for visual recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, Boston, Massachusetts June 11–12,2015 (pp. 36–45). https://doi.org/10.1109/CVPRW.2015.7301270
29. Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., & Darrell, T. (2014). Decaf: A deep convolutional activation feature for generic visual recognition. In *Proceedings of the 31st international conference on machine learning*, Beijing, China June 21C26, 2014 (Vol. 32, pp. 647–655).