Deep Residual Network with Transfer Learning for High Spatial Resolution Remote Sensing Scenes Classification

Ziteng Wang, Wenmei Li, and Guan Gui(✉)

College of Telecommunications and Information Engineering,
Nanjing University of Posts and Telecommunications, Nanjing 210003, China
{1019010409,liwm,guiguan}@njupt.edu.cn

Abstract. Deep residual network (DRN) is considered a promising image classification method for high spatial resolution remote sensing (HSRRS) images due to its great feature extraction capabilities. The classification performance of the DRN is greatly relies on the size of training samples. However, the sample size of HSRRS images is relatively small due to the high acquisition cost. Blindly increasing the sample size would requires huge computing resources and image annotation cost, but would not necessarily improve the classification performance of DRN. In this paper, a transfer learning-aided DRN method (TL-DRN) is proposed for a few shot learning to address the performance challenges associated with HSRRS with relatively small sample size and explore the impact of sample size on classification performance. In the experiment, the weights (shared knowledge) obtained by training the ImageNet datasets on the DRN model are transfered to the TL-DRN model. Experiments with ten different small-scale training sample sizes are conducted. Experimental results show that when the total training sample size is increased from 10 to 100, the classification performance of the TL-DRN model tends to be stable and the mean accuracy of its testing set has stabilized at around 94%. TL-DRN shows a superiority of up to 16% over DRN, in terms of classification accuracy.

Keywords: Convolutional neural network · Images classification · High spatial resolution remote sensing · Transfer learning

1 Introduction

With the enhancement of remote sensing satellite images resolution, we can obtain more information to conduct more impactful research related to land planning, disaster prevention and so on. Classification is one of the most important tasks in HSRRS image processing. HSRRS images have rich spatial, shapes, textures and colors features, which provide good basis for the classification [1].

Deep learning (DL)-based image classification methods usually adopt convolutional neural networks (CNNs) to automatically extract image feature, which

S. Liu and X. Ma (Eds.): ADHIP 2021, LNICST 416, pp. 315–325, 2022.
https://doi.org/10.1007/978-3-030-94551-0_26

gets rid of the complex artificial feature design [2–4]. Generally, the deeper the neural network, the better the classification performance [5]. On account of its excellent feature extraction capabilities, CNNs have been widely applied in the tasks of signal modulation recognition [6,7], voice signal processing and other fields. In recent years, CNNs have developed rapidly, and more and more image classification models based on CNNs achieved excellent results like VGG [8], ResNet [9], InceptionV3 [10] and so on. However, these classification models mainly rely on a great quantity of training samples to obtain high accuracy. Otherwise overfitting may occur [11], which means that trained models have the prefect performance on training samples, while perform badly on the independent test samples. Due to the difficulty of acquiring HSRRS images, the dataset of HSRRS is small. In recent years, a great deal of research have shown that transfer learning could greatly alleviate the phenomenon of over-fitting of the classification model with small sample size [12]. Y. Boualleg *et al.* proposed a CNN-DeepForest based on deep forest and CNNs transfer learning for HSRRS images classification [13]. Xue *et al.* proposed a MSDFF model based on multi-structure deep features fusion for HSRRS image classification [14]. However, there is still a gap in the drive to find optimum sample size for transfer learning to achieve the best classification performance.

In this paper, we propose a TL-DRN model for HSRRS image classification. The proposed model is used to train ten groups of datasets with different sample sizes to explore the influence of the sample size on the model training. The mean accuracy is used to evaluate the performance of the model. The main contributions of this paper can be summarized as below:

1. TL-DRN for HSRRS image classification with limited sample size is proposed. Experiments have proved that the TL-DRN model is more suitable for HSRRS image classification in the case of small sample size than the DRN model.
2. The impact of sample size on TL-DRN are studied. Experimental results confirm that the performance of TL-DRN tends to be stable when the sample size of training reaches six per category.

2 Theoretical Basis

In this section, CNNs, ResNet50 and transfer learning are introduced in detail.

2.1 CNNs

CNNs is a special artificial neural network. Its main feature is the ability to perform convolution operations. Therefore, CNNs is excellent in image classification, detection and segmentation [15].

The input of CNNs is often raw data such as images and audio. The structure of CNNs is a hierarchical model composed of convolutional layers, pooling layers, fully connected layers [16], and activation functions. The original input information undergoes layer-by-layer operation to extract feature information.

Then, those information is used for classification through the fully connected layer [17].

When designing a CNNs, the number of channels in the convolution layer should be equals to the input data. The number of convolution kernels should be the same as that of channels output from this layer. The convolution kernel generally has two attributes: stride (s) and padding (p). Output size of the feature layer after convolution can be obtained by the following calculation formula [18]:

$$n = \frac{N + 2p - f}{s} + 1 \tag{1}$$

where N, f and n represent the size of the input, the convolution kernel and the output respectively.

When the image passes through the convolutional part, low-level convolutional layers extract low level semantic features like texture and shape etc., while high level convolutional layers extract high level semantic features [19]. In general, high level semantic features are more convenient for image classification. Finally, the feature information output by the convolutional layer is mapped to the labeled sample space through the fully connected layer to complete the classification task.

2.2 ResNet50 and DRN

As the number of convolutional layers is increased, the high-level semantic features of the image can be better extracted. However, the deep network may have the problem of gradient disappearance or explosion, which hinders the convergence of the network, otherwise known as the degradation problem. To solve the issue of degradation, ResNet50 is proposed. The ResNet50 network is a stack of residual networks. The Fig. 1 shows the structure of the residual network. The principles of the residual network are as follows.

The residual network consists of one residual unit. First, the residual unit can be written as

$$y_l = h(x_l) + F(x_l, W_l) \tag{2}$$

$$x_{l+1} = f(y_l) \tag{3}$$

where x_l and x_{l+1} respectively represent the input and output of the l-th residual unit [20], and every residual unit includes a multi-layer structure generally. F is the residual function, indicating the learned residual. In addition, $h(x_l) = x_l$ indicates the identity mapping and f represents the rectified linear unit (relu) activation function which is expressed in Eq. (4). The way of adding a highway between the output and input of the network allows us to easily solve the problem of gradient dispersion and network performance degradation.

$$f(x) = \max(x, 0) \tag{4}$$

It should be note that the convolution part of the ResNet50 network is called as deep residual network (DRN) for the convenience of writing.

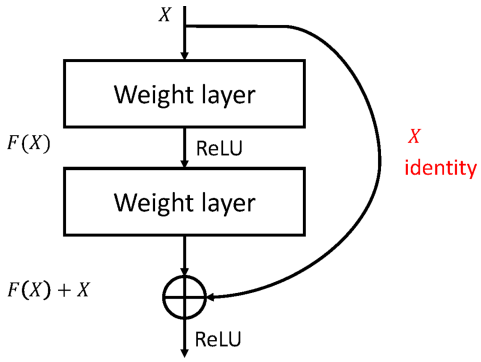

Fig. 1. The architecture of the residual block.

2.3 Transfer Learning

Transfer learning is a learning method for small sample training [21]. It applies the knowledge and experience learned in other tasks to the current task.

In transfer learning, domains (D) and tasks (T) are defined, domains are divided into source domains (D_s) and target domains (D_t), and tasks are divided into source tasks (T_s) and target tasks (T_t). The domain includes feature space and edge probability distribution. Given D_s, T_s, D_t, and T_t, transfer learning uses the knowledge learned from D_s and T_s to enhance the learning of the prediction function f for D_t, where $T = f(D)$, $T_s \neq T_t$ and $D_s \neq D_t$ [22].

In terms of image classification processing, some studies have found that no matter which image dataset is input into the CNNs, the features extracted from the low level convolutional layers are similar. A great deal of researches have proved that the features extracted by a dataset after CNNs are often applicable to another dataset [23]. Therefore, based on this feature, we conduct transfer learning on small samples.

3 The Proposed HSSRS Classification Based on DRN and Transfer Learning

In this section, the structure of TL-DRN network, the training method of TL-DRN, the objective function and evaluation indexs used in the experiments will be introduced in detail.

3.1 The Structure of TL-DRN

TL-DRN is a deep learning network model composed of DRN and transfer learning. First, we build the convolution part of the TL-DRN according to the structure of DRN to prepare for migration. After that, two sets of fully connected layers are added behind the DRN for classification. The framework of the model is presented in Fig. 2.

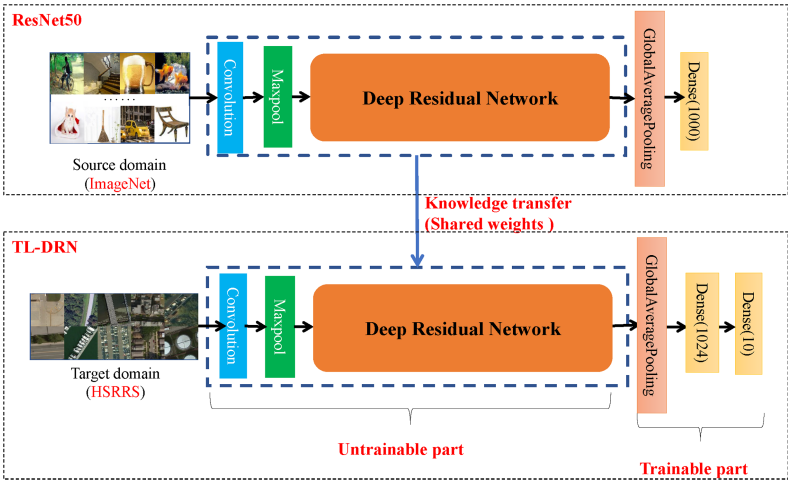

Fig. 2. The framework of ResNet50 and TL-DRN.

3.2 The Method of TL-DRN Training

The TL-DRN training method consists of three parts: ResNet50 training, network reconstruction and feature transfer, and TL-DRN training. The process is presented in Algorithm 1.

ResNet50 Training. ResNet50 is composed of CNNs and residual blocks, which can solve the gradient explosion problem caused by the increase of the network level. Figure 2 shows the model structure. ImageNet [24] dataset is used as the source domain to training ResNet50 and it can be expressed as:

$$D_S = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\} \tag{5}$$

where $\boldsymbol{x}_i = [x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(k)}]^T$ $(i = 1, 2, \ldots\ldots, n)$, $x_i^{(j)}$ represents the j-th feature of the i-th input data of source domain, y_i represents the true label category of source domain. The model can be expressed as:

$$F_S = f_{ResNet50}(\theta_{DRN}, \theta_{fc}; x_i) \tag{6}$$

where F_S represents the output of the model. θ_{DRN} indicates the weight parameter obtained by training the deep residual convolution part and θ_{fc} indicates the weight parameter obtained from the fully connected layer training.

Network Reconstruction and Feature Transfer. TL-DRN is constructed according to Fig. 2. Since ResNet50 and TL-DRN have the same DRN structure, the weight of the convolution part obtained by ResNet50 training can be extracted and loaded into the convolution part of TL-DRN. The dataset used

Algorithm 1. The proposed TL-DRN method.

Input: 10 categories of HSRRS images;
Output: The TL-DRN;
Training and testing the ResNet50 on the ImageNet dataset;
Constructing TL-DRN network based on DRN network of ResNet50;
for j-th layer in TL-DRN_layers **do**
  for i-th layer in DRN_layers **do**
    if j-th layer name == i-th layer name **then**
      Load the weights of layer i to layer j;
      Freeze the weights of layer j;
    else
      Pass;
    end if
  end for
end for
Training and testing the TL-DRN on the HSRRS images ;
return TL-DRN.

in TL-DRN is ten-category images of HSRRS. In transfer learning, this dataset represents the target domain, so it can be written as:

$$D_T = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\} \tag{7}$$

where $\boldsymbol{x}_i = [x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(k)}]^T$, $x_i^{(j)}$ represents the j-th feature of the i-th input data of target domain, y_i represents the true label category of target domain. The model can be expressed as:

$$F = f_{TL-DRN}(\theta_{DRN}, \theta_{newfc}; x_i) \tag{8}$$

where, the value of θ_{DRN} in Eq. (8) is the value of θ_{DRN} trained in Eq. (6). θ_{newfc} indicates the weight parameter obtained from the TL-DRN fully connected layer training.

TL-DRN Training. After network reconstruction and feature transfer, the TL-DRN model can learn the knowledge θ_{DRN} obtained by ResNet50 training. Therefore, we only need to train the parameters θ_{newfc} of the fully connection layer.

3.3 Loss Function

The multi-class categorical cross entropy loss function is used as the loss function and it can be written as:

$$L = -\sum_{i=1}^{N} y_i \log(F_i) \tag{9}$$

where y_i is the true label of i-th input and F_i is the result of the i-th output of the model. N is the number of categories. We aims to train the model to find a suitable set of θ to minimize the loss function.

3.4 Evaluation Index of Experimental Results

Firstly, line chart is used to show the training process, where the horizontal axis represents the epoch of training, the left vertical axis indicates the mean accuracy, and the right vertical axis indicates the loss value.

Secondly, mean accuracy (MA) judge the overall performance of the classification and its calculation formula can be written as [25]:

$$\mathrm{MA} = \frac{1}{N}\sum_{i=1}^{n} C_{ii} \tag{10}$$

where N is the total sample size of the testset, n is the total number of categories to be classified, C_{ii} is the number correctly classified for class i.

4 Experimental Results

In this section, we collect different datasets, conduct experiments according the Section III, and analyze the results of the experiments.

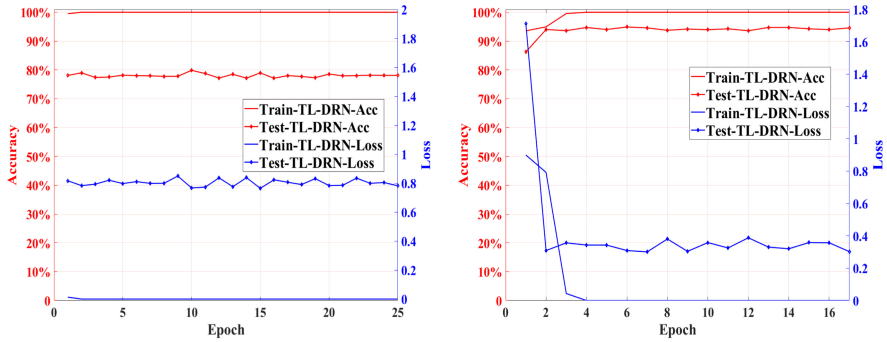

Fig. 3. Accuracy and loss line chart during the period of training and testing for training sample size is (a) one, (b) ten per category with TL-DRN.

4.1 Data Description

From the UCMerce LandUse [26] and RSI-CB128 datasets [27], ten-category HSRRS image samples are selected for classification. One image per category is randomly selected to form a training dataset. Afterwards, $\{2, 3, 4, \cdots, 10\}$ images are selected in the same way as in the above. These ten sample sets are used as small training samples to explore the influence of sample size on TL-DRN classification. Ten images per category are selected randomly to form a training dataset. After that, $\{20, 30, \cdots, 100\}$ images per category are selected in the

same way as in the above. As a control group, these 10 sample sets with larger training samples are directly trained by ResNet50. The specific training sample size is shown in Table 1 and Table 2. In the experiment, all test sets used are the same, with one hundred samples in each category. It should be noted that the following operations are performed in the image preprocessing stage: (1) The input picture size is uniformly changed to the format of $224 \times 224 \times 3$; (2) The data is effectively expanded by rotating and translating the image. Therefore, the data set has been expanded twice on the original basis.

4.2 Experiments Setting

There are three experiments in this part. The experiment one explores the influence of sample size on the transfer learning model. The experiment two and experiment three serve as control learning groups. First of all, for the first experiment, ten sets of training data with a small sample size were selected. The sample sizes of these ten training sets are $\{1, 2, 3, \cdots, 10\}$ per category. The TL-DRN model is built according to Fig. 2. Experiments with 10 sets of data are conducted according to the TL-DRN training method in Section III. Secondly, for the second experiment, the same data set as TL-DRN experiment is selected. The ResNet50 model is built according to Fig. 2 and used to train 10 sets of training samples separately. At last, for the third experiment, ten sets of training data with a larger sample size are selected. The sample sizes of these ten training sets are $\{10, 20, 30, \cdots, 100\}$ per category. After that, ResNet50 is used to train ten sets of training samples directly.

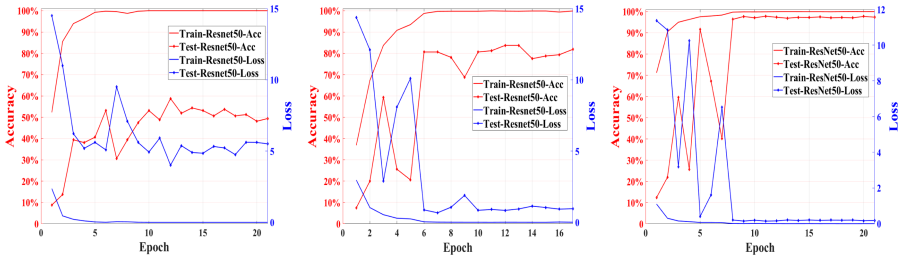

Fig. 4. Accuracy and loss line chart during the period of training and testing for training sample size is (a) one, (b) ten and (c) one hundred per category with ResNet50.

4.3 Experimental Results

First of all, the line graphs of the training process of the three experiments when the sample size is the smallest and the sample size is the largest are shown in Figs. 3 and 4. Each line graph shows the changes of the accuracy and loss of the training set and testing set as the number of training iterations increases. Comparing the training process diagram with the smallest sample size and the largest

Table 1. Testing set mean accuracy on TL-DRN and ResNet50.

TSSPC[a]	1	2	3	4	5	6	7	8	9	10
TMA/T[b] (%)	76.60	83.90	89.00	93.30	94.30	94.80	94.40	94.20	94.80	94.30
TMA/R[c] (%)	52.70	45.00	62.40	71.00	73.70	71.90	72.40	82.10	75.20	79.40

[a] Training sample size per category
[b] Test Mean Accuracy/TL-DRN
[c] Test Mean Accuracy/ResNet50

Table 2. Testing set mean accuracy on ResNet50.

TSSPC[a]	10	20	30	40	50	60	70	80	90	100
TMA[b] (%)	79.40	91.50	90.10	90.60	92.90	95.50	95.80	96.10	94.60	95.90

[a] Training sample size per category
[b] Test Mean Accuracy/ResNet50

sample size of each experiment, we can see that as the sample size increases, the overfitting problem of the model is better mitigated. For example, the MA of the training set and testing set of Fig. 3(a) differs by about 20% while differs by about 6% of Fig. 3(b). Moreover, for small samples, transfer learning can reduce the overfitting situation (This can be understood by comparing to the image of Fig. 3(b) and Fig. 4(b)).

For the first experiment, the test MA obtained by the test experiment in the case of each sample size are given in Table 1. It can be seen clearly that with the increase of the sample size, the classification performance of our proposed TL-DRN is also greatly improved. However, the classification performance of the test fluctuates little, by about 0.3% when the sample size reaches six per category. The MA of the testing set is around 94.4%.

For the second experiment, comparing the test results of TL-DRN and the test results of ResNet50 in the same sample size in Table 1, it can be found that transfer learning greatly improves the performance of classification with small samples size. At the highest level, TL-DRN improves accuracy by nearly 39% compared to ResNet50, and at the lowest level, it has a nearly 12% improvement.

For the third experiment, the test MA obtained by the test experiment in the case of each sample size are given in Table 2. Compared with the first experiment, in the case of ten times the training sample size of the first experiment, the classification performance of ResNet50 is slightly better than that of TL-DRN generally. At the highest level, ResNet50 improves accuracy by nearly 7.6% compared to TL-DRN, and at the lowest level, it has a nearly 2.7% drop.

5 Conclusion

In this paper, the influence of the sample size on the classification of TL-DRN model for ten-category of HSRRS images was investigated. When the sample size of ten types of HSRRS images reaches six per category, the classification performance of TL-DRN network tends to be stable. In addition, the classification

effect of the TL-DRN model is far better than that of ResNet50 on training samples of the same magnitude. And when the training sample size of ResNet50 is 10 times that of TL-DRN, the classification effect of TL-DRN is only slightly lower than ResNet50. It was also shown that TL-DRN is a good candidate for classification of HSRRS images. However, when the sample size increases to a certain level, the continued increase of sample size has little effect on performance. We will continue to be committed to improving the classification performance of the model through other methods in the case of small samples.

References

1. Zhong, Y., Han, X., Zhang, L.: Multi-class geospatial object detection based on a position-sensitive balancing framework for high spatial resolution remote sensing imagery. ISPRS J. Photogramm. Remote Sens. **138**, 281–294 (2018)
2. Arredondo-Velzquez, M., Diaz-Carmona, J., Torres-Huitzil, C., Barranco-Gutirrez, A.I., Padilla-Medina, A., Prado-Olivarez, J.: A streaming accelerator of convolutional neural networks for resource-limited applications. IEICE Electron. Express **16**(23), 20190633 (2019)
3. Gui, G., Huang, H., Song, Y., Sari, H.: Deep learning for an effective nonorthogonal multiple access scheme. IEEE Trans. Veh. Technol. **67**(9), 8440–8450 (2018)
4. Wang, Y., Yang, J., Liu, M., Gui, G.: LightAMC: lightweight automatic modulation classification via deep learning and compressive sensing. IEEE Trans. Veh. Technol. **69**(3), 3491–3495 (2020)
5. Chen, Y., Jiang, H., Li, C., Jia, X., Ghamisi, P.: Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. IEEE Trans. Geosci. Remote Sens. **54**(10), 6232–6251 (2016)
6. Wang, Y., Liu, M., Yang, J., Gui, G.: Data-driven deep learning for automatic modulation recognition in cognitive radios. IEEE Trans. Veh. Technol. **68**(4), 4074–4077 (2019)
7. Wang, Y., et al.: Automatic modulation classification for MIMO systems via deep learning and zero-forcing equalization. IEEE Trans. Veh. Technol. **69**(5), 5688–5692 (2020)
8. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
10. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826 (2016)
11. Liu, S., Deng, W.: Very deep convolutional neural network based image classification using small training sample size. In: 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), pp. 730–734. IEEE (2015)
12. Torrey, L., Shavlik, J.: Transfer learning. In: Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques, pp. 242–264. IGI global (2010)
13. Boualleg, Y., Farah, M., Farah, I.R.: Remote sensing scene classification using convolutional features and deep forest classifier. IEEE Geosci. Remote Sens. Lett. **16**(12), 1944–1948 (2019)

14. Xue, W., Dai, X., Liu, L.: Remote sensing scene classification based on multi-structure deep features fusion. IEEE Access **8**, 28746–28755 (2020)
15. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. **25**, 1097–1105 (2012)
16. Wu, H., Gu, X.: Max-pooling dropout for regularization of convolutional neural networks. In: Arik, S., Huang, T., Lai, W.K., Liu, Q. (eds.) ICONIP 2015, Part I. LNCS, vol. 9489, pp. 46–54. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-26532-2_6
17. Wiatowski, T., Bölcskei, H.: A mathematical theory of deep convolutional neural networks for feature extraction. IEEE Trans. Inf. Theory **64**(3), 1845–1866 (2017)
18. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
19. Long, M., Zhu, H., Wang, J., Jordan, M.I.: Deep transfer learning with joint adaptation networks. In: International Conference on Machine Learning, pp. 2208–2217. PMLR (2017)
20. Zagoruyko, S., Komodakis, N.: Wide residual networks. arXiv preprint arXiv:1605.07146 (2016)
21. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Trans. Knowl. Data Eng. **22**(10), 1345–1359 (2009)
22. Li, W., et al.: Classification of high-spatial-resolution remote sensing scenes method using transfer learning and deep convolutional neural network. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **13**, 1986–1995 (2020)
23. Yosinski, J., Clune, J., Bengio, Y., Lipson, H.: How transferable are features in deep neural networks? Adv. Neural Inf. Process. Syst. **27**, 3320–3328 (2014)
24. Marmanis, D., Datcu, M., Esch, T., Stilla, U.: Deep learning earth observation classification using ImageNet pretrained networks. IEEE Geosci. Remote Sens. Lett. **13**(1), 105–109 (2016)
25. Luo, W., Yang, W., Yu, X., Wang, Y., Tan, K.: Lightweight convolutional neural network for high-spatial-resolution remote sensing scenes classification. In: 2020 International Conference on Wireless Communications and Signal Processing (WCSP), pp. 104–108 (2020)
26. Yang, Y., Newsam, S.: Bag-of-visual-words and spatial extensions for land-use classification. In: Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 270–279 (2010)
27. Li, H., et al.: RSI-CB: A large scale remote sensing image classification benchmark via crowdsource data. arXiv preprint arXiv:1705.10450 (2017)