

Research on Remote Sensing Image Classification Based on Transfer Learning and Data Augmentation

Liyuan Wang^[0009-0002-5271-4905], Yulong Chen^{(✉)[0000-0003-1821-3330]}, Xiaoye Wang, Ruixing Wang, Hao Chen, and Yinhai Zhu

Hubei Normal University, Hubei Huangshi 43500, China
ylchen0424@stu.hbnu.edu.cn

Abstract. Traditional algorithms are no longer effective in the context of the current proliferation of remote sensing image data and resolution, and the remote sensing image classification algorithm based on convolutional neural network architecture needs a significant amount of annotated datasets, and the creation of these training data is labor-intensive and time-consuming. Therefore, using a small sample dataset and a mix of transfer learning and data augmentation, this paper suggests a method for classifying remote sensing images. In this paper, the parameters from the Resnet50 model's pre-training on the Imagnet dataset are migrated to the Resnet50-TL model and ultimately classified using Log softmax. The NWPU-RESISC45 dataset is used in this study to train the model and for data Augmentation procedures. The experimental findings demonstrate that the ResNet50-TL model performs better than other popular network architectures currently in use. The model can classify objects with an accuracy of 96.11% using only 700 data points per class, resulting in a high accuracy rate in a limited amount of data. In the future, the dataset will be increased and the network architecture will be updated frequently to make remote sensing picture interpretation more intelligent and portable.

Keywords: ResNet50 · Image classification · Remote sensing imagery · Transfer learning · Data Augmentation.

1 Introduction

Remote sensing image technology has advanced to a new level with the quickening pace of worldwide science and technology development, and China's remote sensing technology is also advancing. The accuracy of remote sensing pictures currently obtained in China has reached the sub-meter level. In addition to improving accuracy, remote sensing technology advancements have also resulted in a massive rise in data volume. A large number of details progressively emerge in the high-resolution remote sensing images, increasing their complexity, and statistics show that the number of these images grows at a terabyte level every

day, and it is already challenging to manage such enormous and complex data using manual and machine learning interpretation techniques, leading to the development of efficient classifiers. High-performance methods for classifying and interpreting remote sensing images are therefore crucial for study[1].

The maximum likelihood method[2], the minimum distance method[3], K-means[4], and other traditional classification methods for remote sensing images have gradually lost accuracy as a result of the growth of data volume and image resolution, making it challenging to process these images effectively. Miller et al. used neural networks[5] to classify remote sensing images in 1995, expanding the area of machine learning. In 2011, Mountrakis used support vector machines[6] to classify remote sensing images, demonstrating that machine learning algorithms based on image categorization can produce superior results to conventional statistical techniques. However, for sub-meter high-resolution images, the features that must be extracted and the expressions made up of the corresponding functions are more complex, and the shallow learning network only has a small number of computational units to effectively represent the complex functions. As a result, the shallow model gradually becomes unable to adapt to the complex samples as the number of samples and sample diversity increase. Deep learning networks, on the other hand, shine at complex classification thanks to their excellent features, including strong function representation and the capacity to extract high-level semantic features layer by layer, reflecting the intrinsic character of the data.

Since deep learning's rapid growth in 2010, the technology it represents has been gradually revealing its benefits. Deep learning is frequently used in remote sensing image analysis tasks[7], including scene classification[8], target detection[9], image fusion[10], and other uses. The image technology represented by convolutional neural network (CNN)[11] shows great advantages in scene classification and target detection. Compared with traditional machine learning algorithms and manual interpretation, CNN does not require manual processing of features, has high accuracy and strong generalization, and can mine deeper features to build an effective and accurate classification model, which solves the current problems in remote sensing image classification.

In recent years, a large number of CNN designs have emerged from convolutional neural networks. In 2012, Alex Krizhevsky et al. introduced Alexnet[12], which is the first deep learning computation on GPU and addresses the computational bottleneck of deep learning. Alexnet11 has a deeper network than Lenet and adds a dropout layer and activation function Relu. To create a very deep network, the Visual Geometry Group at Oxford University suggested VGG[13] in 2014. This innovation had a significant influence on the design of later CNN architectures. In 2014, Christian Szegedy et al. proposed the Googlenet[14] network, which is based on Inception modules. It addresses the issue of overfitting in deeper networks by superimposing multiple Inception modules and creates a sparse, high-performance network structure by integrating the processing of various filters. By introducing the design of residual blocks, the Resnet[15] network, proposed by Kaiming He et al. in 2015, greatly eliminated the issue of network

degradation brought on by too many layers of the network. As a result, the "depth" of the neural network was able to surpass 100 layers for the first time, and the largest neural network even exceeded 1000 layers. However, since neural networks frequently have tens or even hundreds of layers, a lot of datasets are required for training, and most of the time, these datasets must be manually la-beled. As a result, the focus of current research is on finding small sample datasets to train high-precision models[16]..Additionally, the data augmentation operation can create data actively to increase the dataset and aid in training the model using small batch samples, allowing the model to learn more details and features about the picture. In 2020, Shawky O. A. et al. suggested combining data augmentation and CNN, and they had success with remote sensing picture classification tasks[17].

Nowadays, using Transfer learning[18] to train on datasets has become the choice of many. Many pre-trained models can be used for prediction using Transfer learning, which can maximize training time, reduce model overfitting, and still have good accuracy in the case of small samples. The majority of commonly used CNN models have already trained weights on the dataset ImageNet.

To improve the accuracy of remote sensing image classification and reduce the workload of manual annotation of datasets, this paper proposes a Transfer learning model with ResNet50 as the architecture for the development characteristics of remote sensing images[19], combining deep learning techniques to obtain pre-trained models through Transfer learning, and adding a fully connected layer behind the ResNet50 network to fine-tune the model for the dataset, so as to achieve deep feature extraction of remote sensing images and obtain high prediction accuracy under the background of a small batch of data. This study also compares and contrasts experimental data from the VGG and Denesnet models, demonstrating that Resnet has the benefit of accuracy in this classification of species.

2 The Resnet50 model based on Transfer learning

2.1 Analysis of the ResNet50 model

As the neural network architecture becomes deeper and deeper, from a few layers at the beginning to more than 100 layers at present, we begin to find that as the network gets deeper and deeper, it will lead to a decline in accuracy, that is, the network degradation problem and the ResNet (residual neural network) proposed by Kai-Ming. His team at Microsoft Research solves this problem, the network uses the residual function to achieve a constant mapping, avoiding accuracy degradation. avoiding the degradation of accuracy.

The ResNet network structure is mainly composed of numerous residual blocks. As shown in Fig. 1, each residual block consists of two weight layers and a ReLu activation function.

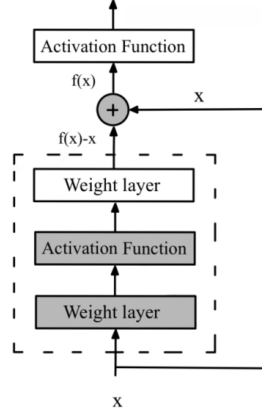


Fig. 1. ResNet residual blocks.

ResNet residual block uses to provide a short circuit connection x to make the output become $(f(x)-x)+x$, where x is the output of the previous layer, $f(x)$ is the total output of the residual block, $(f(x)-x)$ is the output of x through the network. In this case if the network layer is too deep, resulting in a very small output $f(x)-x$ when x enters the network, this will cause the gradient to disappear and thus lead to a loss of accuracy. The model uses a residual block in order to output $(f(x)-x) + x$ after the residual block. Even though $f(x)-x$ is very small, adding x is equivalent to the output of the previous layer, which avoids the disappearance of the gradient of the current layer and thus helps the neural network to become deeper without loss of accuracy.

The output of the fully-connected layer of the original ResNet50 model is a feature dimension vector of 1000, and since there are five classifiers in this case, a Log softmax classifier is added after the fully-connected layer to classify the output to complete our multi-classification task.

2.2 Log softmax classifier

To achieve the final output classification in this study, we add a fully connected layer and a Log softmax classifier at the end of the ResNet network. Log softmax will accept the feature matrix returned from the fully connected layer and obtain the probability value of each category by operation, and finally, the model outputs the classification results. In this paper, we classify five remote sensing images, where we are provided with M inputs (x_i, y_i) , where x_i ($i=1, 2, 3, 4, 5$) is the output of the fully connected layer and y_i ($i=1, 2, 3, 4, 5$) is the corresponding category. i is the category of the output, our paper K is 5, K is the target number of categories, let the function $f(x_i)$ output the probability value of the category corresponding to each input as $P(y_i = j|x_i)$, the function is as follows:

$$f(x_i) = \begin{pmatrix} p(y_i = 1 | x_i) \\ p(y_i = 2 | x_i) \\ M \\ p(y_i = k | x_i) \end{pmatrix} = \frac{1}{\sum_{i=1}^k e^{x_i}} \begin{pmatrix} e^{x_i} \\ e^{x_i} \\ M \\ e^{x_i} \end{pmatrix} \quad (1)$$

Compared with ordinary softmax, Log softmax will penalize larger errors in the likelihood space more highly, and the computation process is smoother without overflow problems due to too large or too small x_i . The function is as follows:

$$\begin{aligned}
\log_e [f(x_i)] &= \log_e \left(\frac{e^{x_i}}{e^{x_1} + e^{x_2} + \dots + e^{x_n}} \right) \\
&= \log_e \left(\frac{e^{(x_i - M)}}{\sum_j^n e^{(x_j - M)}} \right) \\
&= \log_e (e^{(x_i - M)}) - \log_e \left(\sum_j^n e^{(x_j - M)} \right) \\
&= (x_i - M) - \boxed{\log_e \left(\sum_j^n e^{(x_j - M)} \right)}
\end{aligned} \tag{2}$$

where M is the maximum of the input x_i . When x_i is a positive number, $x_i - M$ is always less than 0; when $x_i - M$ is a very large negative value, there must be an $x_i - M = 0$. This ensures that the result of the formula in the black box is equal to $\log_e(x)$, where $x_i > 1$ is not a very large number, so that the problem of underflow is solved.

While in the image classification task, the values keep changing in the deep network may lead to too large or too small feature matrices in the final output, thus leading to errors in the softmax classifier, and the Log softmax classifier solves this situation by logarithmic operations, which can speed up the operation and improve data stability.

2.3 Classification methods of transfer learning models

Transfer learning is the transfer of knowledge from a model that has been trained and learned on other similar data sets to an existing problem to help solve the problem, where the most critical aspect is the transfer of knowledge, where the knowledge gained from training on other tasks is used to develop models for new tasks.[20]

Knowledge transfer between the source domain X and the target domain Y is really the foundation of transfer learning. When $X \neq Y$ and $T1 \neq T2$, the source domain X 's knowledge of the task $T1$ solved is used to assist the target domain Y in solving the task $T2$.

This paper will examine how to train ResNet50 on small batch samples using Transfer[21] learning and apply it to remote sensing image classification in order to lessen the dependence of convolutional neural networks on datasets and to obtain models with higher accuracy.

2.4 Data augmentation

Data augmentation is a method to fictitiously increase the amount of data by generating new data from existing data, as convolutional neural network (CNN) require a large data set for training and CNN are not well trained with small

sample sizes. This involves conducting operations linked to image fusion as well as making minor changes to the data, such as flipping, cropping, changing the color, etc., or creating new data using deep learning models. The enhanced data sets have more features, allowing the model to extract features and magnify feature points more effectively. This enhances the deep learning model’s performance and output[22].

The findings of the remote sensing images vary depending on the spectral range and contain complex data information. The model can adapt to remote sensing images with various resolutions and spectra by adding data augmentation operations, which can increase the accuracy of image classification.

2.5 ResNet50-TL

The block diagram of this study model is shown in Fig. 2.

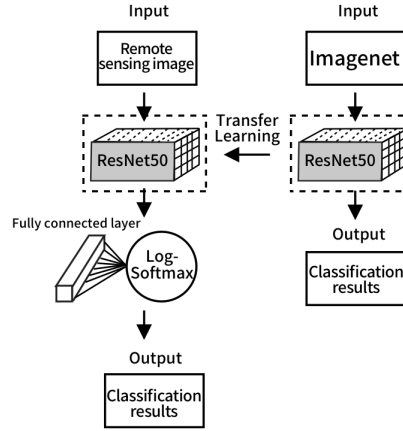


Fig. 2. Block diagram of ResNet50-TL model.

As the feature extraction layer of this model, we first take the pre-training model, which is the ResNet50 model trained on ImageNet. After that, we move the trained ResNet50 model’s parameters to our ResNet50-TL model so that our model can pick up information from the previously trained large model. Finally, we add the fully connected layer and log softmax for classification after the neural network layer, and the output result is the ResNet50-TL model used in this experiment.

In order to maximize the knowledge and performance gained from Transfer learning, we first freeze the pre-trained layers during the training process. This prevents them from back propagating during the following training period. In order to obtain its own fully connected layer parameters and subsequently classify the dataset, we redefine the final fully connected layer and train it using our image dataset.

2.6 Evaluation metrics

Accuracy was used to assess the outcomes of this experiment, where Accuracy is the proportion of correctly predicted outcomes across all samples.

$$Accuracy = \frac{\text{number of correctly classified samples}}{\text{total number of samples}} \quad (3)$$

We are able to determine the accurate rate of image classification using this evaluation metric.

3 Model training and validation

3.1 Experimental dataset

The remote sensing image scene classification (RESISC) dataset used in this study is the NWPU-RESISC45[23] dataset, which was produced by Northwestern Polytechnic University (NWPU). The dataset has a total of 403.74 MB in size and includes 31,500 images with 256*256 pixel resolution, 700 images per class. 45 scene classes are addressed, The 45 scene categories include aircraft, airports, baseball fields, basketball courts, beaches, bridges, jungles, etc. In terms of the quantity of images and the variety of classifications, the NWPU-RESISC45 dataset is one of the most intricate and significant remote sensing scene datasets. The data examples are shown in Fig. 3 below.

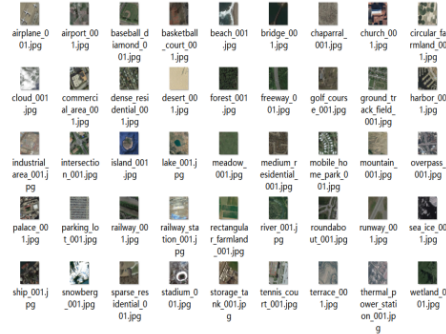


Fig. 3. Image of the dataset.

The aircraft, bridge, palace, building, and arena are chosen in this experiment. Additionally, the data set is split into three categories: the training set (70%), the confirmation set (10%), and the test set (20%).

Additionally, the RSSCN7 dataset, which consists of seven different categories of remote sensing photographs and 400 images in each category, was donated by Wuhan University in 2015 and is used in this study to assess how well the model performs when tested against other datasets.

3.2 Data pre-processing

The experiments in this article were carried out using the pytorch-1.11.0 environment and jupyternotebook. Once we had the dataset, we used PyTorch's ImageFolder to import the images and resize them all to 254*254 before performing image Augmentation procedures on the dataset. After performing RandomResizedCrop and RandomHorizontalFlip, the data is finally converted to Tensor format.

3.3 Image Data Augmentation and Feature Extraction Figure

The deeper network levels and increased feature extraction of the Resnet network are one of its benefits. Under the assumption of small batch samples, this research conducts data Augmentation on the original dataset by cutting, cropping, flipping, changing the color of the data, and other operations to improve the dataset and assist the model in extracting more features. Fig. 4 depicts the picture after a color change. Fig. 5 shows the image with random flipping and cropping.

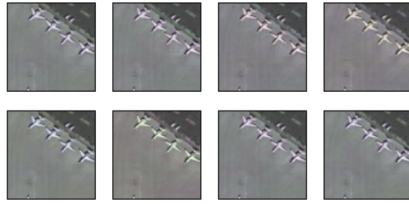


Fig. 4. Data augmentation image 1.

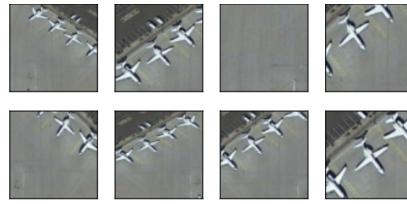


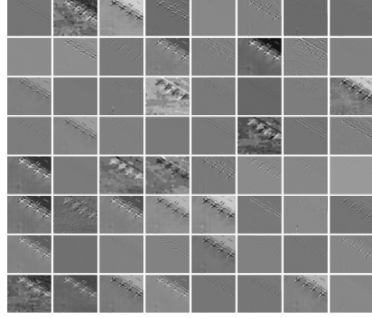
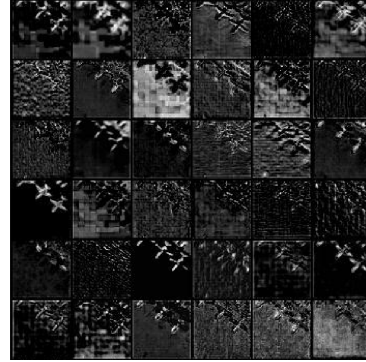
Fig. 5. Data augmentation image 2.

As can be seen, data augmentation procedures significantly increase the dataset, enabling the model to extract more features.

Based on the data Augmentation operation, this experiment employs ResNet50 to extract the image texture and shape features. After network computation, the feature map is visualized, and Figure 5 shows that the features extracted by the network ignore the background and other redundant factors and only extract the essential information.

Fig. 6 displays the outcomes of the second layer of convolutional rendering. Fig. 7 displays the outcomes of the second convolutional layer's rendering component.

As can be seen, as the network's layer count rises, the shallow convolutional layer's feature information remains comparatively rich and the feature data is fairly similar to the original picture data. But as the number of layers increases, the obtained characteristics become less and less useful and more and more abstract. This is so because, while the deeper convolutional kernel begins to extract features like lines and edges, the shallower convolutional kernel prefers to learn basic features like points and colors. The learned characteristics become more precise and abstract as the layers are added to.

**Fig. 6.** Second layer feature map.**Fig. 7.** Fifth layer feature map.

3.4 Training and testing of remote sensing picture classifier models

The pre-trained Resnet50 network architecture in Torchvision, which was trained on the imagenet competition dataset, is used in this exercise. We added a fully connected layer and a pooling layer to the end of the network design and set the number of outputs to 5 to create a classifier for remote sensing images. The original resnet50 parameters were then adjusted in accordance with the dataset's features.

The loss function used in this experiment in the model training is NLLLoss, the parameters are set as follows while the optimizer is used to Adam: Learning Rate = 0.03, Epochs = 10 and Batch size = 256.

The dataset is first trained without the Transfer learning model and Data Augmentation, and the findings are as Table 1 follows.

Table 1. Accuracy under no Transfer learning.

Model	Accuracy
SVM	72.23%
Alexnet	80.03%
VGG16	83.12%
GOOGLeNet	90.21%
Denesnet121	91.03%
ResNet50	91.54%

Then we add the data augmentation operation to get the Table 2 result.

Table 2. Precision after Data Augmentation.

Data Augmentation Operation	Accuracy
Alexnet	83.13%
VGG16	85.33%
GOOGLeNet	92.29%
Denesnet121	93.63%
ResNet50	93.74%

You can see that the data augmentation has improved the overall accuracy of the model. And the deep learning model has higher accuracy compared to the traditional machine learning model SVM.

For experimental comparison, we simultaneously trained VGG, SVM, and Denesnet using the Transfer learning model. The training set and test set errors are displayed in Table 3.

Table 3. Precision after Transfer learning.

Model	Accuracy
Alexnet	89.21%
VGG16	91.34%
GOOGLeNet	93.06%
Denesnet121	94.71%
ResNet50	96.11%

It is clear that transfer learning has a significant positive impact on these models' accuracy, and by gaining knowledge of the source domain, the models are able to perform better in classification tasks even with small batch sizes.

And from among them, we have selected the ResNet50 architecture. Compared to other existing models, ResNet50-T1 model accuracy is more accurate than the conventional algorithm. It also runs more quickly and has a superior ability to extract features.

We also execute the code both before and after fine-tuning, and both times the accuracy is approximately 94% before and 96% after. The best accuracy was 96% with a learning rate of 0.03 after we tried the accuracy at various learning rates. Then, we evaluated the accuracy using various batch sizes and epochs, and the best accuracy was obtained with batch sizes of 256 and 10 respectively. The overall accuracy varied by about 3%. The accuracy varies by about 3% overall, and the findings are as Table 4 follows.

Table 4. Accuracy at different hyperparameters.

Learning Rate	Epochs	Batch size	Accuracy
0.03	10	256	96.11%
0.03	12	128	96.01%
0.03	8	64	93.22%
0.10	10	256	95.45%
0.10	12	128	95.44%
0.10	8	64	93.01%
0.07	10	256	95.18%
0.07	12	128	94.64%
0.07	8	64	93.22%

And we can see that changing hyperparameters like learning rate has little effect on the model. Ultimately, we select the collection of hyperparameters that has the highest accuracy.

We validated the experiments using the RSSCN7 dataset and only 400 photos per class, achieving 93.14% classification accuracy using the ResNet50 model.

This was done to further confirm the validity of the experiments. This model's great accuracy in classifying images with few samples is further demonstrated.

In conclusion, Transfer learning based on the neural network achieved under the training of small batch samples gives the Resnet network a greater advantage in remote sensing image classification tasks. At the same time, Transfer learning based on the neural network is also able to have high accuracy under these conditions.

4 Conclusion

In light of the current rapid growth in the quantity and resolution of remote sensing data, this paper proposes a method of remote sensing image classification based on the combination of transfer learning and data augmentation. It does this by transferring the parameters obtained from the pre-training of the ResNet50 network on Imagenet to the ResNet50-Tl model, which significantly enhances the feature extraction ability and makes the model more general. The network's overall accuracy eventually reaches about 96%, which is a significant improvement compared to other models, after the fully connected layer and Log softmax classifier are added to the network's back. This is followed by a data augmentation operation on a real dataset and the training prediction. By using transfer learning and data augmentation, this experiment accomplishes the classification of high-precision remote sensing images with small batch samples, thereby reducing the workload associated with manually labeling the dataset and achieving a high degree of accuracy. The number of network layers will be increased, more remote sensing image dataset types added, the accuracy and robustness of the model improved, high-precision and low-cost remote sensing image classification sought after, and the speed of light weight and intelligence of remote sensing image interpretation work accelerated in the ensuing research.

References

1. Zhu, X.X., Tuia, D., Mou, L., Xia, G.S., Zhang, L., Xu, F., Fraundorfer, F.: Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE geoscience and remote sensing magazine* **5**(4), 8–36 (2017)
2. Strahler, A.H.: The use of prior probabilities in maximum likelihood classification of remotely sensed data. *Remote sensing of Environment* **10**(2), 135–163 (1980)
3. Wacker, A., Landgrebe, D.: Minimum distance classification in remote sensing. *LARS Technical Reports* p. 25 (1972)
4. Lv, Z., Hu, Y., Zhong, H., Wu, J., Li, B., Zhao, H.: Parallel k-means clustering of remote sensing images based on mapreduce. In: *Web Information Systems and Mining: International Conference, WISM 2010, Sanya, China, October 23–24, 2010. Proceedings.* pp. 162–170. Springer (2010)
5. Miller, D.M., Kaminsky, E.J., Rana, S.: Neural network classification of remote-sensing data. *Computers & Geosciences* **21**(3), 377–386 (1995)
6. Mountrakis, G., Im, J., Ogole, C.: Support vector machines in remote sensing: A review. *ISPRS journal of photogrammetry and remote sensing* **66**(3), 247–259 (2011)

7. Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A.: Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS journal of photogrammetry and remote sensing* **152**, 166–177 (2019)
8. Zou, Q., Ni, L., Zhang, T., Wang, Q.: Deep learning based feature selection for remote sensing scene classification. *IEEE Geoscience and Remote Sensing Letters* **12**(11), 2321–2325 (2015)
9. Chang, C.I., Heinz, D.C.: Constrained subpixel target detection for remotely sensed imagery. *IEEE transactions on geoscience and remote sensing* **38**(3), 1144–1159 (2000)
10. Liu, Y., Chen, X., Wang, Z., Wang, Z.J., Ward, R.K., Wang, X.: Deep learning for pixel-level image fusion: Recent advances and future prospects. *Information Fusion* **42**, 158–173 (2018)
11. Zhang, W., Tang, P., Zhao, L.: Remote sensing image scene classification using cnn-capsnet. *Remote Sensing* **11**(5), 494 (2019)
12. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. *Communications of the ACM* **60**(6), 84–90 (2017)
13. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014)
14. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 1–9 (2015)
15. 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)
16. Zhang, H., Liu, Y., Fang, B., Li, Y., Liu, L., Reid, I.: Hyperspectral classification based on 3d asymmetric inception network with data fusion transfer learning. *arXiv preprint arXiv:2002.04227* (2020)
17. Shawky, O.A., Hagag, A., El-Dahshan, E.S.A., Ismail, M.A.: Remote sensing image scene classification using cnn-mlp with data augmentation. *Optik* **221**, 165356 (2020)
18. Alem, A., Kumar, S.: Transfer learning models for land cover and land use classification in remote sensing image. *Applied Artificial Intelligence* **36**(1), 2014192 (2022)
19. Shabbir, A., Ali, N., Ahmed, J., Zafar, B., Rasheed, A., Sajid, M., Ahmed, A., Dar, S.H.: Satellite and scene image classification based on transfer learning and fine tuning of resnet50. *Mathematical Problems in Engineering* **2021**, 1–18 (2021)
20. Zhang, D., Liu, Z., Shi, X.: Transfer learning on efficientnet for remote sensing image classification. In: *2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*. pp. 2255–2258. IEEE (2020)
21. Xie, M., Jean, N., Burke, M., Lobell, D., Ermon, S.: Transfer learning from deep features for remote sensing and poverty mapping. In: *Proceedings of the AAAI conference on artificial intelligence*. vol. 30 (2016)
22. Lv, N., Ma, H., Chen, C., Pei, Q., Zhou, Y., Xiao, F., Li, J.: Remote sensing data augmentation through adversarial training. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **14**, 9318–9333 (2021)
23. Cheng, G., Han, J., Lu, X.: Remote sensing image scene classification: Benchmark and state of the art. *Proceedings of the IEEE* **105**(10), 1865–1883 (2017)