Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 5
AIDML 2024

Comparison of AlexNet and ResNet Models for Remote Sensing Image Recognition

Wenxuan Zhang *

School of Architecture, Harbin Institute of Technology, Harbin, Heilongjiang Province, 150000, China

* Corresponding Author Email: 2021113038@stu.hit.edu.cn

Abstract. Remote sensing image recognition is an important direction in remote sensing data processing. Among them, deep neural networks have achieved results far beyond traditional methods on many challenging image datasets in remote sensing images. There are various network models for convolutional neural networks, for better remote sensing image recognition, this paper has selected the AlexNet model and ResNet model for comparison to select the network model with more accurate results. After running, it found that the accuracy is 0.6479 on the ResNet model and 0.6023 on the AlexNet model, the dataset performs better on the ResNet18 model, to further elucidate this result, it analysed the difference in accuracy between the two models. It was found that the use of techniques such as learning residuals, jump connections, and batch variance in the ResNet18 model reduced the problem of gradient vanishing during training, while the deeper layers and higher number of parameters also led to higher accuracy of the ResNet18 model.

Keywords: Remote sensing; deep neural networks; ResNet18.

1. Introduction

At present, remote sensing technology is developing in the direction of multi-fusion, intelligence, dynamic, and high resolution. At the same time, with the improvement of remote sensing data acquisition technology, more remote sensing data can be processed. Among them, remote sensing image classification is an important direction of remote sensing data processing. It does this by identifying the characteristics of each type of remote-sensing image, allowing the computer to filter out these images and collect them into massive amounts of data. At present, remote sensing image recognition faces a variety of challenges, and the classification and recognition of remote sensing images face the following challenges: First, remote sensing images have complex spatial distribution and texture features, which require powerful image processing capabilities. Secondly, due to the diversity of feature types and backgrounds, it is necessary to distinguish and classify various types of features. At the same time, the amount of remote sensing image data obtained is huge, and it needs to be able to be processed and identified quickly. In response to these challenges, researchers are constantly exploring new algorithms and technologies to improve the accuracy and efficiency of remote sensing image recognition. At the same time, deep neural networks have achieved far superior performance on many challenging image datasets in recent years [1].

In this paper, the research content is a remote sensing image classification and recognition algorithm based on deep learning, and the AlexNet and ResNet models are used to find a better remote sensing image recognition algorithm by comparing the accuracy of the results under the same parameters.

2. Overview of the Basic Theory of Convolutional Neural Networks

Convolutional neural networks are a special class of deep neural networks that are particularly this paper suited for processing data with a grid-like topology, such as images. They recognise features in images by mimicking the way the visual cortex of the human brain works. Convolutional neural networks consist of basic parts such as a convolutional layer, activation function, pooling layer, fully

connected layer, loss function, optimizer, regularization, and data augmentation and their basic concepts are described below [2].

The convolutional layer is the core of the convolutional neural network. It extracts local features in an image by performing a convolution operation with the input image through a set of learnable filters (or called convolution kernels). Each filter slides over the input image and performs a dot product operation at each position to produce a feature map. By using multiple filters, the convolutional layer can extract multiple features in the input image. Activation functions are used to introduce nonlinear properties after the convolutional layer, allowing the network to learn and approximate complex functions. Commonly used activation functions include ReLU (Rectified Linear Unit), Sigmoid and tanh. The pooling layer is usually located after the convolutional layer and is used to reduce the dimensionality and computational complexity of the feature map while retaining the most important feature information. Common pooling operations include maximum pooling and average pooling. At the end of a convolutional neural network, one or two fully connected layers are usually used, which are used for classification or regression tasks to map the learned features to the final output space. The fully connected layers will spread the previously extracted feature maps into one-dimensional vectors and connect them to the output layer. Fully connected layers are used for classification or regression tasks, mapping the learned features to the final output space. The loss function is used to measure the difference betthis paperen the network's predicted and true values.

During training, the loss function is optimized by the backpropagation algorithm and gradient descent algorithm to update the paperweight parameters of the network. Optimiser is used to update the paperweight parameters of the network during the training process. Common optimisers include SGD, Adam, RMSprop, etc. Regularisation techniques are used to prevent CNN overfitting. Commonly used regularisation methods include L1/L2 regularisation, Dropout, Batch Normalization, etc. Increasing the diversity of the training data by random rotation, flipping, scaling, cropping, etc. helps to improve the generalization ability of the CNN [3].

Convolutional Neural Networks (CNNs) have significant advantages in image processing and recognition tasks. It can automatically learn features from input data, without the need to manually design or select features, and is friendly to processing massive data information; This feature also enables CNNs to capture richer information when processing image data, improving the accuracy of classification or recognition. At the same time, the computational efficiency of CNN is greatly improved compared with traditional algorithms, and the method of parameter sharing through the convolutional layer greatly reduces the number of parameters in the network and reduces the computational complexity. The local connection means that the neurons in the CNN are only connected to the local area of the input data, rather than to the entire input data, which reduces the network parameters and improves the computational efficiency. The multi-layer convolution and pooling operations in the CNN operation process can gradually extract the abstract features in the input image from the low level to the high level, which enables the CNN to process the features at different levels and capture the information at different scales in the image. These advantages have enabled CNN to achieve remarkable results in remote sensing image recognition.

3. Analysis of ResNet18 Network Structure and AlexNet Network Structure.

For remote sensing image recognition, this paper uses the ResNet18 model and the AlexNet model for image recognition. The Residual Network (ResNet) model is a deep learning convolutional neural network (CNN) architecture, proposed by Kaiming He et al. The core idea of ResNet is to solve the problems of gradient loss and model degradation encountered by deep neural networks during training by introducing ‘residual learning’. Residual learning refers to learning the residual representation of this paper's inputs and outputs, instead of learning the mapping from inputs to outputs. By introducing shortcut connections, ResNet can transfer information to these paper network layers more directly and efficiently.

Thanks to the introduction of residual learning, the network can make more efficient use of depth information and reduce the problem of gradient vanishing during training. The residual block in ResNet is usually a bottleneck block to reduce the amount of computation and the number of parameters. The Bottleneck Block achieves a balance between this paper computational efficiency and model performance by first using 1x1 convolution for dimensionality reduction, then 3x3 convolution for feature extraction, and finally 1x1 convolution for dimensionality enhancement. Two types of residual blocks are typically included in the model: the Basic Block and the Bottleneck Block, with the Basic Block being suitable for shallothis paperr networks and the Bottleneck Block for deeper networks [3]. These two types of residual blocks are slightly different in structure, but the core idea is to introduce jump connections for residual learning.

Due to the introduction of residual learning and jump connections, ResNet is easier to optimise during training and is better able to handle datasets of different sizes and complexity. This enables ResNet to achieve excellent performance on a variety of tasks such as image classification, target detection, semantic segmentation, etc.

The AlexNet model is an iconic network structure for deep learning to make breakthroughs in computer vision, proposed by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton in 2012, and won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) that year. AlexNet is a relatively deep network structure consisting of eight learning layers, including five convolutional layers and three fully connected layers. In some of the convolutional layers, AlexNet uses multiple different convolutional kernel sizes (e.g., 11x11, 5x5, 3x3) to capture features at different scales. The fully connected layers have a large number of neurons, giving the model enough capacity to learn complex image representations. AlexNet also uses the Rectified Linear Unit (ReLU) activation function after the convolutional layer instead of the traditional sigmoid or tanh function, which speeds up the model training process and mitigates the problem of gradient vanishing. AlexNet uses the Local Response Normalization, LRN technique. LRN simulates the inhibition effect betthis paperen neighboring neurons in a biological neural system and helps to improve the generalisation ability of the model. However, as deep learning techniques evolve, LRN may not be necessary for some tasks as techniques such as Batch Normalization provide more efficient regularisation methods. In the pooling layer, AlexNet uses Overlapping Max Pooling, i.e. the pooling window will have overlapping parts when sliding. This type of pooling avoids overfitting to some extent and improves feature richness. To further prevent overfitting, the Dropout technique is used in the fully connected layer, where Dropout randomly sets the output of a portion of neurons to 0 during training, making the model see a different network structure at each iteration. Meanwhile, AlexNet uses a softmax classifier after the last fully connected layer for the multi-classification task. the softmax function converts the model's output into a probability distribution, which represents the probability that the input image belongs to each class. These features enabled AlexNet to achieve remarkable results at the ImageNet Challenge and laid the foundation for subsequent deep learning developments in the field of computer vision [4].

4. Results of the experiment

The parameter settings of the AlexNet and ResNet models are shown in Table 1, and the calculation results are shown in Figs. 1 and 2 respectively, the red line represents accuracy and the blue line represents loss. The accuracy of the ResNet model and the accuracy of the AlexNet model are shown in Table 2, the accuracy of the ResNet model is higher and the loss is this paper, so in the case of the same parameters, this paper uses the ResNet18 model. This paper shows that im_size is 224, batch_size is 64, class_mode is categorical, learning _rate is 0.0005 and epochs are 15 in Table 1.

Table 1. Key parameters for training a machine learning model

Parameter	Value
im_size	224
batch_size	64
class_mode	categorical
learning_rate	0.0005
epochs	15

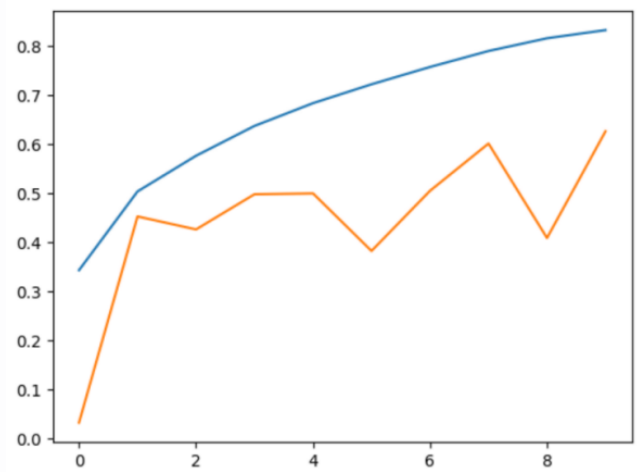

Figure 1. Model performance evaluation of AlexNet (Photo/Picture credit: Original).

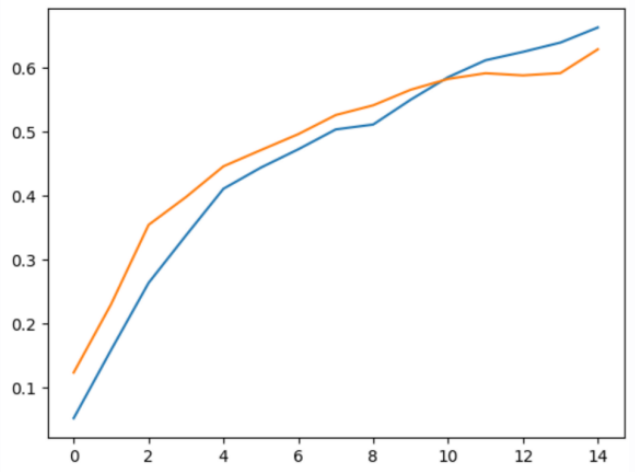

Figure 2. Model performance evaluation of ResNet (Photo/Picture credit: Original).

Table 2. Compares the accuracy and loss metrics between ResNet18 and AlexNet models.

	ResNet18	AlexNet
Accuracy	0.6479	0.6023
Loss	1.1941	1.5666

The AlexNet model is better than the ResNet model in terms of accuracy for the following reasons:

The first is the model depth and the number of parameters. AlexNet is a shallots paper model than ResNet, which uses a deeper network structure, which helps to extract higher-level features, so it generally achieves better performance. The deeper network structure also allows the model to have more parameters to learn more complex patterns and features. The second is the introduction of skip connections in ResNet, which helps to alleviate the vanishing gradient problem and makes the network easier to train. Hop connections allow information to propagate more quickly from the smaller layers of the network to the higher layers, helping to improve the accuracy of the model. In ResNet, the model is trained by learning residual, which makes the network easier to optimize, avoids the problem of gradient vanishing and gradient explosion when training deep networks, and helps to improve the accuracy of the model. In addition to these, batch normalization technology is also introduced in ResNet, which helps to speed up the training process, reduce the gradient vanishing problem when training deep networks, make the network easier to train, and ultimately achieve higher accuracy [5].

To further improve the accuracy of image recognition results in the future, this purpose can be achieved by adjusting the network structure parameters, adjusting the model network structure, and increasing or decreasing the learning rate.

5. Conclusion

In this paper, this paper study how to use convolutional neural network for remote sensing image recognition, and by running two different models and comparing the results, this paper can obtain a model with higher accuracy. Convolutional Neural Network (CNN) is a deep learning model that has been widely used in remote sensing image recognition, through which local features in images can be extracted and classified and recognized by CNN networks. Compared with traditional algorithms, it has many advantages, such as feature extraction ability, ability to process large-scale data, high efficiency, classification accuracy, and adaptability to different sizes of input. Different models of convolutional neural networks have their characteristics in terms of structure, depth, and application. In this paper, the AlexNet model and the ResNet18 model are selected for image recognition. The AlexNet model was released after the LeNet model, which proposes a deeper number of network layers based on the LeNet model, and uses some new technologies, such as ReLU activation function and Dropout, to obtain better recognition results. It contains multiple convolutional, pooling, and fully connected layers to handle more complex image features. The ResNet model was released relatively recently and introduced residual connections to solve the gradient vanishing problem in deep networks. ResNet is one of the popular deep convolutional neural networks, which contains multiple residual blocks, each of which is composed of multiple convolutional layers and a cross-layer connection, which can effectively transmit gradient information.

Through comparative studies, this paper concludes that the ResNet model has better accuracy and analyze the reasons for its higher accuracy. The ResNet model has a deeper depth and more parameters, and the technologies it introduces such as skip connections, learning residuals, and batch normalization help to improve the training speed and reduce the gradient disappearing in the training process so that the network training can always obtain high accuracy in the end.

References

[1] Zhou Mo. Research on remote sensing image recognition and detection technology based on depth learning algorithm. Information Recording Material, 2024, 25 (4): 162 - 164.

[2] Kong Qingqun, Wu Fuzhao, Fanbin. Image matching based on depth learning: methods, applications and challenges. Journal of Computer Science, 2024, 1 - 39.

[3] Hao Xuejie. A review of data augmentation methods of remote sensing image target recognition. Remote Sensing, 2023, 15 (3): 827 - 827.

[4] Lv Dengke. Remote sensing image scene classification based on Resnet50 and channel attention. Jiangxi Science, 2024, 42 (2): 396 - 404.

[5] Yang Zhen, Guo Yanguang, Lu Xiaobo. Remote sensing image classification of UAV based on improved Alexnet Network. Journal of Hunan University of Science and Technology Science, 2023, 38 (3): 59 - 69.