Research on Image Super-resolution Technology Based on Sparse Constraint SegNet Network

Ruikai Lin, Liyuan Wang*, Tao Xia
Northeast Forestry University, Harbin city, Heilongjiang Province, 150000 China

*Corresponding author e-mail: rickie_lim@nefu.edu.cn

Abstract. Aiming at the problems of long time-consuming and low accuracy in extracting buildings with traditional machine learning methods. In this paper, the SegNet semantic segmentation models on deep learning is used to improve the algorithm, and a high-resolution remote sensing image building extraction algorithm based on sparse constrained SegNet are proposed. First, regular terms and Dropout are added to the SegNet model, which greatly reduces the occurrence of model over-fitting; secondly, in order to extract richer semantic features for the model, the algorithm introduces a pyramid pooling module; finally, the Lorentz function sparse constraint factor is introduced to the SPNet model, Construct a new semantic segmentation model LSPNet. In order to verify the reliability and applicability of the proposed algorithm, the optimized LSPNet model is used to identify and extract the buildings in the high-resolution data set. Experimental results show that compared with traditional machine learning methods, this method has the advantages of fast convergence and high accuracy, and has a good application prospect.

Keywords: Deep learning; feature extraction; semantic segmentation; sparse constraint.

1. Introduction
The massive amount of ground feature information contained in remote sensing images plays a vital role in the research of many fields. How to mine these remote sensing images of deeper information is particularly important. Among them, the extraction of remote sensing image buildings is the most critical item in the basic surveying and mapping content. The accuracy of recognition and extraction is directly related to the level of surveying and mapping of features, and buildings are also a category that is prone to increase or decrease in the category of features. Therefore, it needs to be updated continuously. In recent years, due to the large imaging area of remote sensing images, automatic extraction of buildings from remote sensing images has faced the following difficulties: First, the extraction accuracy is usually related to the size of the building, and the extraction accuracy of larger buildings is often higher than that of small buildings; The second is that the imaging color of remote sensing images is low, and the color changes of the ground objects are not obvious, So, it is easy to cause the mixed extraction of the building and the background, and the separability is deteriorated. Accurately extract all types of buildings [1].

With the introduction to the concept of deep learning and the publication of a series of cases and research reports, the field of image recognition has also ushered in its spring. Zuo Tongchun proposed
an end-to-end improved full convolutional neural network structure HF-FCN to realize the area recognition of buildings, but the accuracy of building extraction still needs to be improved; Mnih established a remote sensing image feature based on DCNN Feature automatic extraction system, and adding random condition fields to improve the model; Chen Leishi and others used convolutional neural networks and BP neural networks to extract construction land in urban areas, and concluded that the extraction accuracy of convolutional neural networks is better[2]. However, the impact on remote sensing image bands of the accuracy of classification results is not considered; He Hao et al. designed an Encoder-Decoder model to extract roads to remote sensing images. This method retains more local information on features, but does not consider the topological structure information about roads, the extraction accuracy still needs to be improved.

Aiming at the time-consuming and low-precision problems of traditional machine learning methods for extracting buildings, a sparse constrained semantic segmentation model LSPNet is constructed in this paper, which achieves the purpose of rapid model convergence and effectively improves the accuracy of semantic segmentation.

2. Algorithm principle

2.1. SegNet model
In the network structure of Figure 1, the encoder is a VGG16 model composed of 13 convolutional layers. Each convolutional layer contains batch normalization (Batch Normalization) and the operation of activation function ReLU (as shown in Figure 2). The pooling layer is a $2 \times 2$ window, and the maximum pooling method is used for down-sampling.
The feature map obtained by the encoder is up-sampled by using the index value of the largest pooled pixel; at the same time, the generated feature map is convolved again to obtain a dense feature map; finally, the generated multi-channel feature map is classified by Sigmoid The device operates and outputs the end-to-end semantic cut image [3].

2.2. Pyramid pooling module

In the original SegNet model, the pooling layer operation shown in Figure 3 will cause the image to lose a small amount of high-frequency components, produce passivated fuzzy blocks and lose pixel position and spatial information. To solve this question, consider introducing the pyramid module.

In the original SegNet model, the pooling layer operation will cause the image to lose a small amount of high-frequency components, produce passivated fuzzy blocks and lose pixel position and spatial information. To solve this problem, consider introducing a pyramid module [4].

The pyramid pooling module performs feature fusion according to different coarse and fine scales. The output of different scale levels contains feature maps of different sizes, but all use 1*1 convolutional layers to reduce the dimension of the context representation to the original 1/N, where N means The size of the refinement level is increased, and then the low-dimensional feature map is up-sampled through bilinear interpolation to obtain features of the same size. The coarsest scale uses global average pooling to output the feature map in a single grid, and the feature map is divided into different sub-regions at the finest scale to form different location features. The problem of fuzzy blocks is as little as possible, and the original feature information extracted by the convolutional layer is retained to the maximum extent. Including the space, color, and location information of the pixel [5]. Finally, the different levels of features are assembled into the final pyramid pooling global feature. Figure 3 and Fig. 4 is the pyramid pooling module, which uses different scales to perform pooling operations on the same feature image, and then uses the subsampling layer of the deep convolutional neural network to connect it to obtain the output response to / from each location.

![Pyramid pooling module](image1)

![LSPNet model architecture](image2)
2.3. Sparse constraint based on Lorentz function.
The least square error function is a loss function of the neural network. The adjustment range and training duration of the model's hyper parameters are related to the initial error of the model. Suppose the sample set composed of n training samples is

\[ \{(x^1, y^1), (x^2, y^2), \ldots, (x^n, y^n)\} \]

In the formula, \(y_i (i = 1, 2, \ldots, n)\) is the real output, and the least square error function is used to calculate the error sum of the predicted output \(r\) and the real output \(y\)

\[ J(w, b) \]

However, using the least square error function as the loss function has the disadvantages of more iterations and slower convergence speed. Adding the Lorentz function of the loss function can improve the sparsity of the model and speed up the convergence speed of the model. Lorentz function is often used in image processing and analysis, such as image edge detection and image modeling, etc. The basic model is shown in the formula, where \(a\) and \(B\) are constants.

\[ F = \frac{A}{B^2 + x^2} \]  

The Lorentz function is a standard one-parameter Cauchy distribution, and the probability density function is shown in formula (3).

\[ (x; s) = \ln (1 + (x/s)^2) \]

In the formula, \(s\) is the scale variable, which controls the sparseness of the function.

2.4. LSPNet model structure
The algorithm models / modelled to refer to the SPSNet model, and the pooling layer is changed from / into the pyramid pooling module; at the same time, the difference between the output value and the label valued is calculated by using the loss function of the increased sparse constraint factor, which speeds up [6]. The speed at which the error back propagation algorithm updates the weights and bias terms to reduce the training time.

3. Results evaluation index
Generally, in the field of deep learning, there are three kinds of accuracy judgment indexes for feature extraction.

**Fig.5** Features tag example
As shown in Fig. 5, tn is the non-building feature that is correctly detected; fp is the non-building feature that is incorrectly detected as a building feature; fn is the building feature that is incorrectly detected as a non-building feature; tp is the correct detection Building characteristics.

4. Experiment and result analysis

4.1. Experimental data set

Usually when training a neural network, enough training samples will make the network training effect better. When the training samples are insufficient, data enhancement is needed to the existing data. The data format of the dataset used in this paper is too large. Import the dataset into ArcGIS software, crop it into 13 sub-images according to the density of the building, and perform data enhancement such as rotating the data, adding random noise and gamma transformation Operation, the original data set capacity is increased to 18,532, showing part of the image and its corresponding label map after data enhancement. The image map is in RGB format. The black part in the label map represents the background, and the white part represents the building. The data resolution of both is cropped to 256 × 256 pixels [7].

In the aviation data set, 15,000 pieces of data were selected for training, 2,560 pieces of data for testing, and 2 pieces of data from different regions for verification. The experiment uses the deep learning framework TensorFlow to train the neural network [8]. Complete multiple training sessions on Nvidia GeForce GTX 1070Ti (8GB). GPU The initial learning rate is set to 0.0001, and the attenuation value is 0.000001 and the momentum is 0.9, the batch is set to 5 batches, and the number of iterations is 16000 times, As shown in Fig 6.

![Data enhancement samples and labels](image)

Fig.6 Data enhancement samples and labels

4.2. Experimental results and analysis

First, train the SegNet model and the SegNet model with Dropout and regular terms separately, and introduce the pyramid pooling model. Analysis of the loss value and accuracy changes / changed curve of the training set and the validation set for 3 trainings, it can be seen that the SegNet model with the dropout and regular terms / termed added has a 0.5% increase in IOU and a precision increase in 1.2
compared to the existing SegNet models / modelled. %, Recall has increased by 6.8%, and the model convergence speed has been improved. Effectively avoid the appearance of over-fitting phenomena [9]. Introduce the pyramid pooling module to get the SPNet model. Compare with the SegNet model, IoU increased by 1.1%, Precision increased by 1.8%, and Recall increased by 7.5%.

The Lorentz function is introduced to construct a sparse constrained semantic segmentation model LSNet for training. The hyper parameters set are consistent with the previous experiment. The loss values / valued and accuracy of the training set and the validation set can be compared. Compare with the SPNet model training situation, after introducing the sparse constraint factor, the model converges faster, the loss value decreases, and the accuracy rate increases [10]. This shows that the semantic segmentation capability of the model is enhanced.

5. Conclusion

(1). The Lorentz function sparse constraint is introduced into the SegNet model, and the sparse constraint semantic segmentation model LSPNet is constructed, and the global sparse network structure is constructed. Extract buildings from the data set. The results show that the semantic segmentation models / modelled that introduces the pyramid module and the sparse constraint factor improves the accuracy and speed of building extraction. (2). In order to test the applicability of the algorithm, the algorithm is applied to the extraction of high-resolution buildings from other data sets; at the same time, a comparison experiment with the traditional algorithm has been done, and the results show that the proposed algorithm has various accuracy. In index evaluation and visual analysis, it is superior to other algorithms and has a good application prospect.

(3). Both the SPNet model and the LSPNet model failed to recognize small buildings, undetected ground objects, and failed in segment the edges of some buildings well.

References

[1] Machine learning for aerial image labeling. MNIH V. University of Toronto. 2013
[2] Automatic Color Correction for Multisource Remote Sensing Images with Wasserstein CNN. Guo J, Pan Z, Lei B, et al. Remote Sensing of Environment. 2017
[3] Building change detection for high-resolution remotely sensed images based on a semantic dependency. Zhong C, Xu Q, Yang F, Hu L. 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). 2015
[4] Dense semantic labeling of subdecimeter resolution images with convolutional neural networks. Volpi M, Tuia D. IEEE Transactions on Geoscience and Remote Sensing. 2017
[5] SegNet: A deep convolutional encoder-decoder architecture for image segmentation. Alex Kendall, Vijay Badrinarayanan, Roberto Cipolla. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017
[6] Building detection from aerial images using invariant color features and shadow information. Sirmacek B, Unsalan C. 23rd International Symposium on Computer and Information Sciences. 2008
[7] A new method of building detection from a single aerial photograph. Zhong S, Huang J, Xie W. 2008 9th International Conference on Signal Processing. 2008
[8] Scale space and edge detection using anisotropic diffusion. P. Perona, J. Malik. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1990
[9] Texture and Scale in Object-Based Analysis of Subdecimeter Resolution Unmanned Aerial Vehicle (UAV) Imagery. Laliberte, A.S, Rango, A. Geoscience and Remote Sensing, IEEE Transactions on. 2009
[10] Gradient-based learning applied to document recognition. LeCun L, Bottou L, Bengio Y, Haffner P. Proceedings of Tricom. 1998