ELEKTROTEHNIŠKI VESTNIK 87(5): 281-286, 2020
ORIGINAL SCIENTIFIC PAPER

# Building Extraction from High-Resolution Remote-Sensing Images Based on Deep Learning

**Haihui You[1 2], Linhui Li[1] , Weipeng Jing[1]**

[1] *College of Information and Computer Engineering. Northeast Forestry University, China*
[2] *College of Computer and Information Engineering. HeiLongJiang University of Science Techonlogy,China*
*E-mail: jwp@nefu.edu.cn*

**Abstract.** The efficient and accurate extraction of building feature information in remote-sensing images has become one of the most important elements of satellite remote-sensing image research. The paper proposes a convolutional neural network with a symmetric encoding-decoding structure. Alternating convolutional blocks and maximum pooled under-sampling at the encoder end are used to complete the relevant operations. The convolutional blocks are operated by linear residual blocks, and complementary zeros are added after $3 \times 3$ convolutional layers to ensure consistency in feature-map dimensions. A traditional ReLU activation function is replaced with a SELU activation function in order to retain more feature information during training and to solve the problem of dead neurons. A $1 \times 1$ convolutional layer and a Sigmoid function are finally introduced to complete the final building extraction. The experimental results show that the model is more effective in densely-populated urban areas than in Alpine towns, but the overcrowding of buildings also causes difficulties in accurate edge segmentation.

**Keywords:** encoder-decoder; building identification; deep learning; convolutional neural network; SELU

**Razpoznava zgradb iz visokoločljivih daljinsko zajetih slik na osnovi globokega učenja**

Eden izmed ključnih korakov pri daljinskem zaznavanju s sateliti je učinkovito in natančno pridobivanje informacij o zgradbah. V članku predlagamo konvolucijsko nevronsko mrežo s simetrično kodirno-dekodirno zgradbo. Pri delovanju enkoderja smo uporabili izmenjujoče konvolucijske gradnike in največje podvzorčenje. Konvolucijski bloki delujejo hkrati s preostalimi linearnimi bloki. Po konvolucijski plasti 3 x 3 smo dodali ničle za zagotovitev doslednosti pri ugotavljanju dimenzij. Da bi ohranili več informacij med postopkom učenja, smo uveljavljeno aktivacijsko funkcijo ReLU zamenjali s funkcijo SeLU. Za končno razpoznavo zgradb smo uvedli konvolucijsko plast 1 x 1 in funkcijo Sigmoid. Poskusni rezultati kažejo, da je predlagani pristop bolj učinkovit v gostonaseljenem mestnem okolju kot na podeželju, vendar so pri natančni robni razčlenitvi težava prebivalci v mestih.

## 1 INTRODUCTION

With the emergence of high-resolution remote-sensing satellites, such as QuickBird and ZYG3, high-resolution remote-sensing images have become one of the main sources of important geoscience information [1]. The significant improvement of the spatial resolution of a remote-sensing image makes the remote-sensing image show more detailed information, geometric structure and texture features. As one of the most important artificial targets on the ground, buildings play an important role in urban planning, military reconnaissance and map. Therefore, it is important to extract buildings quickly and accurately from remote-sensing images.

In recent years, with the development of deep learning algorithms, target detection and recognition and image semantic segmentation based on the convolutional neural network have greatly improved the accuracy of remote sensing image target segmentation. In papers [2][3], the convolutional neural network algorithm was applied to extract buildings, and the accuracy is significantly improved, but a large number of samples are needed and the boundary of the extraction results is not complete. Aiming at the problems of high-resolution remote-sensing image building information extraction, we propose a network model based on the Segnet network. It has a symmetric encoder-decoder structure for image feature extraction and fusion. The model takes a high-resolution remote sensing image as an input and performs a pixel-level building extraction to achieve the final segmentation result and adopts the structure of the residual. It combines the SELU activation function and zero padding operation to design a convolutional neural network with a symmetric encoding-decoding structure. Finally, we extract the building information based on the Inria aerial image data set. The good results are obtained.

In the construction of a convolutional neural network model, the traditional activation function is the ReLU function. However, in the calculation process of ReLU, the negative data are often reduced to zero, which will cause “dying Relu” in the model and most components of the network will never be updated. Therefore, valuable information will inevitably be lost in the

*Received 22 September 2020*
*Accepted 6 November 2020*

activation process. Therefore, replacing the ReLU function with the SELU function can save more picture information during training, effectively prevent generation of dead neurons and accelerate the convergence speed of the model. Better results can be achieved by combining SELU with batch normalization.

The next sections are organized as follows. Section 2 illustrates the related work in the literature of semantic image segmentation. Section3 explains our methodology to approach high-quality segmentation. Section 4 presents different experiments and results to demonstrate the power of our methodology. In section 5, conclusions are drawn and directions given for future work.

## 2 Related Works

Traditional approaches usually rely on the domain knowledge to extract features, such as Texton Forests [7], Random Forest[8], [9], SVM[10], [11] and CRFs[12]. As shown in [4,5], SVM is applied to a high-resolution remote-sensing image building information. In [6], a morphology building index (MBI) is proposed. The above methods of building extraction are all based on pixel features, which lead to problems of a fuzzy and incomplete boundary and low-extraction accuracy.

In recent years, deep-learning methods have shown an excellent performance in many fields including semantic segmentation. A series of deep Neural Network models based on Convolutional Neural Network (CNN) have emerged [13]. Characterization of the models have a strong ability. A series of such models has appeared, Fast R–CNN[14], YOLO[15], SSD[16], target detection model, the FCN[17], SegNet[18] and U-net[19] target-segmentation model . Therefore, many scholars have improved the model algorithm and applied it to building extraction. For example, Maggiori et al. enhance the model ability to extract architectural details by integrating low-resolution features with high-resolution local features [20]. Zhang et al. propose a multilevel classifier method, i.e. adaptive-image segmentation, and improve the extraction accuracy of buildings [21]. Yuan et al. use a method of multi-stage combined mapping of building features to classify image pixels [2].

Volodymyr Mnih uses convolutional neural networks in his PhD thesis [22] to train an aerial image labeling system for roads and buildings. He tries neural networks and conditional random Fields as post-processing to CNN. His model shows a good performance on Massachusetts roads and building dataset.

Saito and Aoki [23] use CNN for road and building detection. They use a normal down-sampling architecture of CNN. At the end, a fully connected layer with Dropout is added to infer prediction of the input image. Their model outperforms the Mnih models for both the roads and buildings using a single model for each class.

Andrew Khalel and Motaz el-Saban [24] introduce a new DCNN semantic image-segmentation architecture. It bases on stacked U-nets where each network enhances results of the previous one, and performs verification on the Inria Aerial image dataset.

## 3 Methodology

The essence of building extraction in a high-resolution remote sensing image is actually an image-segmentation problem. Image segmentation is a special problem of image recognition. We use a classical convolutional neural network with a symmetric encoder-decoder structure to perform a pixel-level building extraction. Our model adopts the structure of a symmetric encoder and decoder to extract and fuse image features (see Fig.1). The model takes a high-resolution remote-sensing image as an input and performs a pixel-level building extraction to achieve the final segmentation result.

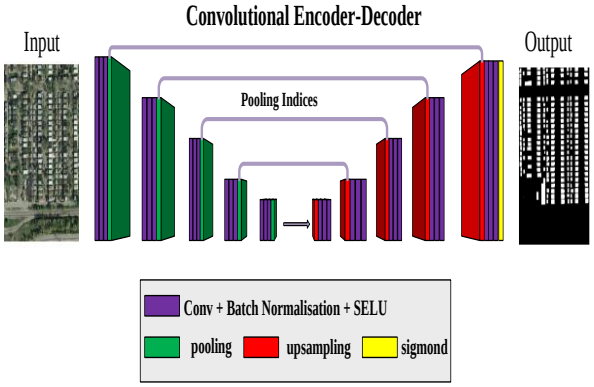


Figure 1. Encoder-decoder architecture.

### 3.1 Encoder

The input of the encoder is a remote-sensing image, and the image resolution features are extracted through a convolution operation and a down-sampling operation, in which the down-sampling operation is completed by a maximum pooling layer. In the convolutional layer, the residual mode is used to complete the feature extraction of a remote-sensing image. As shown in Figure 1, each convolution layer consists of 1×1, 3×3, and 1×1 filters and performs a unit plus to learn more image features. The 3×3 filter reduces the convolutional layer to ensure that the feature map of a remote-sensing image maintains a certain size after a convolution operation.

Each layer also has a batch normalization layer for a faster convergence. The operation of the down-sampling at the encoder end is completed by a maximum pooling operation with the step size of 2 and core size of 2×2. After the pooling operation, the size of the feature map changes and its length and width become half of the original size. In this way, after four times of sampling, the feature image will be 1/16 of the original input image.

### *3.2 Decoder*

In the construction of a convolutional neural network model, the traditional activation function is ReLU. In the ReLU activation function, if the input is less than zero, the output is equal to zero, otherwise the output is equal to the input. As a result, when a function is introduced into the neural network, the model also introduces a good nonlinearity and sparsity, thus reducing the time and space complexity. But in the calculation process of ReLU, the negative data is often reduced to zero, which will cause dying ReLU in the model and most components of the network will never be updated. Therefore, valuable information will inevitably be lost in the activation process. Many authors choose to map the no-zero input to the non-zero output to solve the problem of dead neurons, and propose a variety of alternative activation functions, such as LReLU[26]、ELU[27] and SELU[28]. Their function formula and image are as follows:

$$LReLu(x) = \begin{cases} x & if\ x > 0 \\ \alpha x & if\ x < 0 \end{cases} \tag{1}$$

LReLU activation function $\alpha$ usually values in（*0.1*, *0.3*）, but in the training process of a neural network model not learn the $\alpha$ values, so you need to debug the assignment. $\alpha$ select *0.2*. Since the output value is no longer limited to zero, the problem of dead neurons is solved to some extent. The negative output can also push the weight and bias the change of the network in the right direction. Since the function formula does not include an exponential operation, the calculation speed of the network is improved

$$ELU(x) = \begin{cases} x & if x > 0 \\ \alpha(e^x - \alpha) & if x < 0 \end{cases} \tag{2}$$

$$SELU(x) = \lambda \begin{cases} x & if\ x > 0 \\ \alpha e^x - \alpha & if\ x \le 0 \end{cases} \tag{3}$$

The ELU activation function $\alpha$ value is in (0.1, 0.3), $\alpha$ also select 0.2. If the activation function input is less than zero, it outputs a value slightly less than zero, thus solving the problem of dead neurons. But the introduction of exponential computing increases the cost of computing.

Gunter et al.[35] give the optimal value of $\alpha$ and $\beta$ in the SELU activation function, i.e. $\alpha \approx 1.67326, \beta \approx 1.05070$. With the positive and negative outputs, the SELU activation function enables the network model to be self-normalized, which results in all the outputs being 0 and the standard deviation is 1, which means that the network model can converge at a faster rate. The use of the SELU function also means that the gradient disappearance or gradient explosion will not occur [35].

Therefore, substituting the ReLU function with the SELU function not only save more picture information during training and effectively prevent the generation of dead neurons, but also accelerate the convergence speed of the model. Better results can be achieved by combining the SELU function with a batch normalization.

### *3.3 Decoder*

The decoder is mainly used to decode the features extracted from the encoder end. By up-sampling the input feature map of the encoder, the decoder end connects with the same level feature map of the encoder end through channels to splicing the feature vectors, and then USES multiple convolutional layers for feature fusion and amplification of the receptive field. Finally, a 1×1 convolution operation is used at the end of the decoder to convert the image channel number to one. As the sigmoid function has good mapping between 0 and 1, the probability that each pixel point is obtained by using the sigmoid function, thus achieving the effect of extracting buildings from high-resolution remote-sensing images.

$$sigmoid(x) = \frac{1}{1+exp^{-x}} \tag{4}$$

The up-sampling operation at the decoder end is performed by a *2×2* deconvolution. The convolution block is also composed of a *1×1*, a *3×3* and a *1×1* convolution, with batch regularization and SELU activation function operations between them. At the same time, after a 3×3 convolution operation, it also introduces a zero-complement operation. This is because without zero padding, reduces the size of the convolution operation and could make a characteristic figure resulting in an up-sampling and down-sampling. The figure size is not equal and performs a cross-layer the samples under the same level for connection to the operating characteristics of the figure for cutting operation. This not only promotes the complexity of the network, but also makes the network owners with a different size of the input and output. For the zero padding operation, therefore, the whole network can be made of only the largest pooling and the deconvolution operation causes changes in the characteristics of the image size, as a result of the model with a symmetry and operation, which makes the model in the channel connection does not need to carry on a cutting operation, but also ensures an equal the size of the input and output image, satisfying the requirements of the extraction of remote- sensing image building.

### *3.4 Loss Function*

Establishment of a loss function is extremely important in the process of the network-model construction. A good loss function can provide a better learning effect for the model. The loss function commonly used in image segmentation is the cross-entropy loss function.

(See formula 5). All pixels in the image learn equally by comparing the predicted value of each pixel with the true value and then averaging all pixels. N is the total number of pixels, y is the real value of pixels, f(x) predict values of pixels, $p$ collection is the pixel prediction, $t$ is a real tag set of pixels.

$$L_{cross\ entropy\ loss} = -\sum_{n=1}^{N} ylog\big(f(x)\big) + (1-y)log\big(1-f(x)\big) \quad (5)$$

However, the problem of the disequilibrium between the building pixels and the background pixels in a remote-sensing image makes it difficult for the model to learn the features of smaller objects. One way to solve this problem is to make the weight for each category in the calculation of the loss, such as adding a larger weight to the building and a smaller pixel to the background. However, the introduction of weights will introduce extra super parameters to the network model needing to be adjusted continuously in the training process. The other way is to choose a function with a small bias and combine it with the cross-entropy loss function. We use the Dice loss function. The formula is :

$$L_{Dice} = 1 - \frac{2\sum_{n=1}^{N} p_n \times t_n}{\sum_{n=1}^{N} p_n + \sum_{n=1}^{N} t_n} \quad (6)$$

The Dice loss function can be applied in the case of uneven samples. It essentially measures the overlap between the predicted value and the true value. The more matching parts between samples, the more the Dice coefficient approaches 1 and the loss function approaches 0. Therefore, the formula of the combined loss function is:

$$Loss\ function = L_{cross\ entropy\ loss} + L_{Dice} \quad (7)$$

## 4 EXPERIMENT

### 4.1 Datasets

To illustrate the power of the proposed model, we use the Inria Aerial Image Labeling dataset. It is a remote-sensing image data set specially built for the urban building detection. There are only two kinds of the marks of the images, i.e. building and non-building, so it can be used for the pixel-level semantic segmentation. The dataset provides remote-sensing images with a resolution of *0.3*m covering an area of *810* square kilometers

Figure 2. Chicago, Kitsap County buildings truth labels.

### 4.2 Evaluation Criteria

The semantic segmentation is essentially a dichotomy task with two classes, so the predicted results are usually as follows:

TP (true Positive): the pixel is judged as a positive sample, and the actual result is the same as the predicted value;

FP (false Positive): the pixel is judged as a positive sample, but the actual result is contrary to the predicted value;

TN (True Negative): the pixel is judged as negative sample, and the actual result is the same as the predicted value;

FN (false negative): the pixel is judged to be a negative sample, but the actual result is contrary to the predicted value.

We use precision, recall and IoU as evaluation indexes

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

$$IoU = \frac{TP}{FP+TP+FN} \quad (10)$$

The precision index means how many of the samples that are predicted to be positive are actually positive. The recall-rate index is for the original sample, and it shows how many of the positive examples in the sample are correct. IoU index mainly measures the degree of the overlap between the output area generated by the model and the real area of a building

### 4.3 Experimental results and analysis

Figures 3 and 4 show the original images, standard building label and predictive renderings of the trained models for five different urban areas in Austin, Chicago, Key SAP County, Washington, Tyrol, and Vienna .

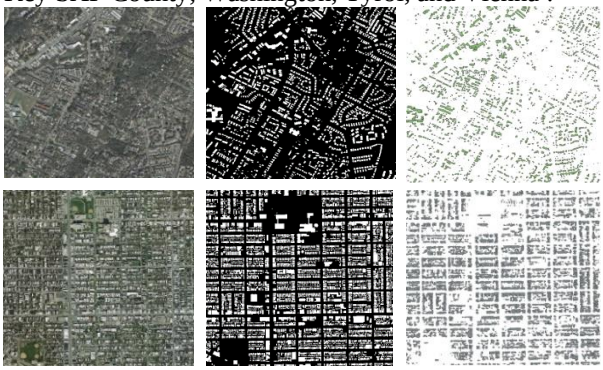

Figure 3.Austin and Chicago images and buildings label images and predicted label image.

The structure of the image of the above densely populated urban areas is simple due to the structure of the distributed equably neat, similar geometric features of the same direction and most of the buildings. In these areas, the test results of buildings are better. Wherever, the flat terrain and the color features of the easily distinguishable buildings and backgrounds also enable the model to learn better the architectural features. The

test results are consequently generally better than those for the Alpine towns and rural areas with farmlands, such as Kitsap and Tyrol. However, due to the small spacing and high density of buildings in these areas, it is difficult for the model to recognize the boundary features of buildings in the training process, which leads to segmentation difficulties and mistakes identification of the building pixels as the background pixels.

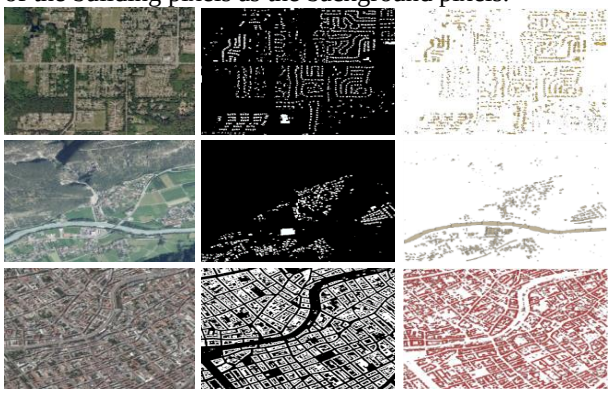

Figure 4. Waston, Tyrol and Vienna images, buildings label images and predicted label images.

For the Alpine towns and rural areas with farmlands, such as Kitsap and Tyrol, similar color features and complex buildings present great challenges for building extraction. Also, due to the topography, the background terrain of some areas is higher than the buildings and vice versa. Such staggered terrain structures also bring difficulties to the training of the model and reduce accuracy in the identification of buildings. In the Tyrol area, there are differences between the river and the surrounding green farmland background which may cause the model to misidentify it as a building.

After running all experiments and choosing the best model (section3), our results are compared with the results of other methods.

Table 1. IoU results obtained with different methods for the Aerial Image Labeling validation set

| Method | Austin | Chicago | Kitsap | Tyrol | Vienna |
|---|---|---|---|---|---|
| FCN+MLP[25] (Baseline) | 61.20 | 61.30 | 51.50 | 57.95 | 72.13 |
| SegNet [29] (Single-Loss) | 76.49 | 66.77 | 72.69 | 66.35 | 76.25 |
| SegNet+ MultiTask-Loss [29] | 76.76 | 67.06 | 73.30 | 66.91 | 76.68 |
| 2-levels[24] U-Netsaug. | 77.29 | 68.52 | 72.84 | 75.38 | 78.72 |
| Our model | 77.21 | 70.52 | 73.42 | 75.30 | 79.47 |

Table 2. Precision and Recall result of our method

| | Precision(%) | Recall(%) |
|---|---|---|
| Austin | 83.68 | 76.64 |
| Chicago | 87.20 | 79.45 |
| Kitsap.co | 78.82 | 75.24 |
| Tyrol | 76.73 | 74.38 |
| Vienna | 85.12 | 85.70 |

As above, our model performs best. Among all the comparisons, the FCN performance is the weakest. Since FCN is a general-purpose semantic segmentation model which does not consider the specific characteristics of aerial images. Compared with FCN, SegNet performs better because of its sufficient de-convolution layers and specific up-sampling scheme. Due to the multi-scale aggregation scheme, the behavior of U-Net is stronger than that of FCN and SegNet. For the overall scenario, our model achieves 75.2% in IoU. For the scenario of different cities, our method is still superior to other methods.

Based on the above results, the following conclusions are drawn. First, FCN and SegNet have smooth-boundary prediction and perform well on small buildings. Second, U-Net can extract more diverse buildings since it could learn multi-scale information from aerial images. Third, FCN achieves a good behavior on the building locations and boundaries as a result of the fused intension-frequency information.

## 5 CONCLUSION

In the paper, we propose an end-to-end convolutional neural network with a symmetric encoding-decoding to extract buildings from high-resolution remote-sensing images. In the encoding end, the model introduces a residual block for the convolution operation in the down-sampling and replaces the ReLU activation function with the SELU function to solve the problem of dead neurons. The up-sampling of the decoder is accomplished by a deconvolution and convolution block operation, and the characteristic channel connection corresponding to the same layer of the encoder is also introduced. After each convolution operation, a zero-complement any strategy is adopted to ensure the invariability of the size of the input and output data of the model. Finally, the Inria data set proves the feasibility of the model.

Our experiment shows that the test results in densely populated urban areas are better than in the Alpine town areas, but the over-dense cluster of buildings also brings difficulties to the edge segmentation. Therefore, our future focus will be on this problem, and we will try to increase the depth and the width of the model and to obtain more subtle image features with a deeper and wider model structure. Although the results obtained by using different comparable methods are acceptable, our model outperforms them.

## ACKNOWLEDGEMENT

The work described in this paper is supported by National Natural Science Foundation of China (31770768), Fundamental Research Funds for the Central Universities (2572017PZ04), Heilongjiang Province Applied Technology Research and Development Program Major Project (GA18B301,

GA20A301) and China State Forestry Administration Forestry Industry Public Welfare Project (201504307).

## REFERENCES


[1] CHERIYADAT AM. Unsupervised feature learning for aerial scene classification[J]. IEEE Transactionson Geoscience and Remote Sensing,2014,52(1):439G451.

[2] YUAN Jiangye, Automatic building extraction in aerial Scenes using convolutional networks[J/OL]. (2016-02-21).[2017-05-23]. https:‖arxiv.org/abs/1602. 06564.

[3] Vakalopoulou M.Karantzalosk,Komodakisn,et al . Building detectionin very highre solution multi-Spectral data with deeplearning features[C]‖Proceedings of 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Milan:IEEE,2015:1873G1876.

[4] HAN Junwei, ZHANG Dingwen, CHENG Gong,etal. Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning[J]. IEEE Transactions on Geoscience and RemoteSensing,2015,53(6):3325-3337.

[5] HUANG Xin,ZHANG Liangpei.An SVM ensemble approach Combining spectral,structural,and semantic features for the classification of high-resolution remotely sensed imagery[J]. IEEE Transactions on Geoscience and Remote Sensing,2013,51(1):25 7-272.

[6] LIN Xiangguo, ZHANG Jixia. Object-based morphological Building index for building extraction from high resolution Remote sensing imagery[J]. Acta Geodaeticaet Cartographica Sinica,2017,46(6):724-733.

[7] J. Shotton, M. Johnson, and R. Cipolla, "Semantic texton forests for image categorization and segmentation," in Computervision and pattern recognition, 2008. CVPR 2008. IEEEConference on. IEEE, 2008, 1–8.

[8] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, and R. Moore, "Real-time human pose recognition in parts from single depth images," Communications of the ACM, vol. 56, no. 1, 116–124, 2013.

[9] F. Schroff, A. Criminisi, and A. Zisserman, "Object class segmentation using random forests." in BMVC, 2008, pp. 1–10.

[10] Y. Yang, S. Hallman, D. Ramanan, and C. C. Fowlkes, "Layered object models for image segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34,no. 9, 1731–1743, 2012.

[11] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan,"Object detection with discriminatively trained partbasedmodels," IEEE transactions on pattern analysis andmachine intelligence, vol. 32, no. 9, 1627–1645, 2010.

[12] C. Russell, P. Kohli, P. H. Torr et al., "Associative hierarchical crfs for object class image segmentation," in Computer Vision, 2009 IEEE 12th International Conference on. IEEE, 2009, 739–746.

[13] Lecun Y L, Boottou L, BengioYet al.Gradient-Based Learning Applied to Document Recognition [J]. Proceedings of the IEEE,1998.86(11):2278-2324

[14] Girshick R.Fast R-CNN[J].Computer Science,2015

[15] Redom J,Divvala S,Girshick R,et al.You Only Look Once:Unified,Real-Time object Detectiion[C].Computer vision&Pattern Recognition.2016

[16] Liu W, AngueLov D, Erhan D .SSD: Single Shot MultiBox Detector[C].European Conference on Computer vision.2016:21-37.

[17] Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on pattern Analysis & Machine Intelligence, 2014,39(4):640-651.

[18] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," arXiv preprint arXiv:1511.00561.2015.

[19] Ronneberger O,Fischer P,Brox T.U-net:Convolutional networks for biomedical image segmentation[C].Lecture Notes in Computer Science .2015

[20] Maggiori E, Tarabalka Y, Charpiat G, et al. Convolutional Neural Networks forLarge-Scale Remote Sensing Image Classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016:1-13.

[21] Zhang X, Du S. Learning selfhood scales for urban land cover mapping with very high resolution satellite images[J]. Remote Sensing of Environment, 2016,178:172-190.

[22] V. Mnih, "Machine learning for aerial image labeling," Ph.D.dissertation, University of Toronto, 2013.

[23] S. Saito and Y. Aoki, "Building and road detection from large aerial imagery," in SPIE/IS&T Electronic Imaging. International Society for Optics and Photonics, 2015, DOI: 10.1117/12.2083273

[24] Khalel A, El-Saban M .Automatic Pixelwise Object Labeling for Aerial Imagery Using Stacked U-Nets[J]. 2018.arXiv:1803.04953v1 [cs.CV] 13 Mar 2018

[25] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Can semantic labeling methods generalize to any city? The Inria Aerial image labeling benchmark," in IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE.2017.

[26] Andrew L. Maas, Awni Y. Hannun, Andrew Y. Ng. Rectifier Nonlinearities Improve Neural Network Acoustic Models. Proceedings of the International Conference on Machine Learning, Atlanta, Georgia, USA, 2013.

[27] Djork-Arn´e Clevert, Thomas Unterthiner, Sepp Hochreiter. Fast and accurate deep networklearning by exponential linear units (ELUS)[J]. arXiv 2016, arXiv:1511.07289.

[28] Günter Klambauer, Thomas Unterthiner, et al., Self-Normalizing Neural Networks[J]. arXiv 2017, arXiv:1706.02515.

[29] B. Bischke, P. Helber, J. Folz, D. Borth, and A. Dengel,"Multi-task learning for segmentation of building footprintswith deep neural networks," arXiv preprint arXiv:1709.05932,2017


**Haihui You** received her MSE degree in Computer Science from the Harbin Normal University, China, in 2005. She is a master student majoring in computer technology at the Key Laboratory of Forestry Data Science and Cloud Computing of the State Forestry Adiminstration, Northeast Forestry University. She works in Hei Long Jiang University of Science Techonlogy. Her current research interests include remote image processing and deep learning.

**Linhui Li** was born in Hun Lunbeier City, Inner Mongolia Autonomous Region, China in1979. she received a Bachelor's degree
from Northeast Forestry University in Harbin, China, in 2002 and a Master's degree from Northeast Forestry University in Harbin, China, in 2005. Now she is the Ph. D student of School of Forestry Information Engineering in Northeast Forestry University, china. Her major is computer application. Her research interests are in the area high performance computing, data miming, remote sensing image processing and application. He has hosted and participated in national and provincial research project and published several papers

**Weipeng Jing** received his Ph.D. degree from the Harbin Institute of Technology of China. He is currently an Associate Professor with the Northeast Forestry University, China. His research interests include modeling and scheduling for distributed computing systems, fault- tolerant computing and system reliability, cloud computing, and spatial-data mining. He has published over 50 research papers in refereed journals and conference proceedings, such as CPC.