

# INTELLIGENT DETECTION OF STRUCTURE FROM REMOTE SENSING IMAGES BASED ON DEEP LEARNING METHOD

Liang Xin \*

<sup>1</sup> Shanghai Institute of Surveying and Mapping, Shanghai, China, 476422556@.com

**KEY WORDS:** Remote Sensing Image, Land Use Monitoring, Deep Learning, Neural Network, Building

## ABSTRACT:

Utilizing high-resolution remote sensing images for earth observation has become the common method of land use monitoring. It requires great human participation when dealing with traditional image interpretation, which is inefficient and difficult to guarantee the accuracy. At present, the artificial intelligent method such as deep learning has a large number of advantages in the aspect of image recognition. By means of a large amount of remote sensing image samples and deep neural network models, we can rapidly decipher the objects of interest such as buildings, etc. Whether in terms of efficiency or accuracy, deep learning method is more preponderant. This paper explains the research of deep learning method by a great amount of remote sensing image samples and verifies the feasibility of building extraction via experiments.

## 1. INTRODUCTION

### 1.1 The history of Artificial intelligence

Artificial intelligence was founded as an academic discipline in 1956. In the years since has experienced several waves of optimism, followed by disappointment and the loss of funding (known as an "AI winter") (Crevier et al., 1993), followed by new approaches, success and renewed funding. For most of its history, AI research has been divided into subfields that often fail to communicate with each other. These sub-fields are based on technical considerations, such as particular goals (e.g. "robotics" or "machine learning"), the use of particular tools ("logic" or "neural networks"), or deep philosophical differences. Subfields have also been based on social factors (particular institutions or the work of particular researchers). This paper focuses on the deep learning method which is a branch of machine learning.

### 1.2 The meaning of building extraction

As the main means of earth observation, high-resolution remote sensing is currently the major approaches to data acquisition. Its resolution can reach the sub-meter, which is able to identify the main features clearly and extract constructions from remote sensing images. Yet the method of recognizing buildings from the images by manual visual interpretation has a large workload, low efficiency and can be affected considerably by subjective consciousness. Through the traditional target extraction or shallow machine learning algorithm, it is difficult to extract the buildings because of the heavy mixing of buildings and surrounding environment. The shadow effect of tall buildings is obvious, which influences the extraction effect of other buildings in the shadow. Under the circumstances of vigorous development of artificial intelligence, using the depth neural network to achieve the intelligent recognition and the building extraction through the deep learning method, can greatly reduce the workload of manual visual interpretation and improve the work efficiency.

### 1.3 The traditional method

In the research of remote sensing image information extraction and target recognition, traditional algorithm makes target identification mainly through extracting the image feature

manually. These features include the texture feature (Peng et al., 2006), gradient direction histogram (HOG) (Tang, 2015), Gabor transformation (Sastry et al., 2007), etc. Then these features are input in the form of characteristic vector to a traditional classifier such as SVM (Seetha et al., 2012), AdaBoost (Ramzi et al., 2014), decision tree (Zhuo et al., 2013) and so on to conduct the classification. It has achieved effective results in several target recognition tasks, such as the identification of aircrafts, ships. However, this method often performs poorly in manual feature extraction, algorithm robustness, displacement and rotation invariance.

Compared to the traditional algorithm, the depth of the learning algorithm in information extraction of remote sensing image, represented by the convolution neural network (CNN) (Arun et al., 2013), has many natural advantages: firstly, it can get rid of the manual calculation steps and study directly from remote sensing data in accordance with the data distribution model; secondly, it can effectively deal with high-dimensional and massive data for pattern recognition and classification.

### 1.4 Building extraction method based on deep learning

Deep learning is a branch of machine learning, from Turing put forward the idea of AI for the first time in 1950, in the development of more than 60 years, artificial intelligence has experienced two ups and two downs, every rise is that algorithms had a new breakthrough while every low tide is the bottleneck of computing performance and the data size. In 2012, represented by deep learning algorithms, it ushered in the third great development, the identification rate exceeds 99% and 95% on the voice and visual recognition respectively.

The recognition error rate of deep neural network, represented by GoogleNet is only 3.6% on ImageNet. This is the first time that the machine defeats humans in the aspect of image recognition (human error recognition rate is between 5% and 10%). Deep learning can achieve such superior performance, which is inseparable with its ability of feature extraction. In terms of feature extraction, deep learning method can not only learn the underlying characteristics such as color, edge, but also can learn intermediate features such as texture, shape. More senior features like a dog's head, partial of objects are all part of the features it can extract.



Figure 1. Features extracted by deep learning method

Therefore, extracting buildings from remote sensing images based on the deep learning can extract the characteristic information of buildings from different scales, different levels and different grades. This is the main basis to accurately identify and locate the building, which highly exceeds the recognition accuracy of traditional method.

## 2. TECHNIQUE METHODS

### 2.1 Overall technical process

The technical scheme considered the characteristics of the buildings on the remote sensing image and the superiority and inferiority of different deep learning models in the field of target detection, target segmentation. Simultaneously it also took the detection efficiency and accuracy into account. The design technical scheme is as follows:

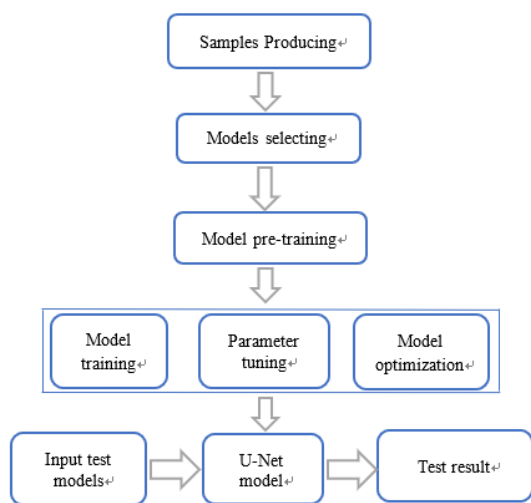


Figure 2. Design technical scheme flowchart

Firstly, capture sample data of buildings on Google Earth and produce buildings sample library of multi-source, multi-scale and multi-spatial resolution, which covers domestic and foreign building entities of different shapes, different attitudes. Then, use ArcGIS to make the label by vectorization. Through the preprocessing, the sample and label are transformed into the input data format of the network model.

The sample data can be input into network model to be trained after they were produced. What we used is the U-Net network model (Ronneberger, 2015). This model uses the transfer learning to pre-train the model. This technical route can guarantee high recall rate and efficient detection efficiency, at the same time it improved the initial performance of the model through transfer learning, which can effectively shorten the training cycle.

### 2.2 Development framework selection

The deep learning frameworks of the current mainstream are shown in Table 1. Tensor Flow is the official development framework of deep learning provided by Google. It has quantities of advantages, such as high-quality code, high-activity community and strong-maintainability team. Moreover, it supports multiple GPU parallel computing and GPU cluster computing and the mainstream deep neural network model such as CNN, FCN and RNN, etc. Hence Tensor Flow is preferred to be our development framework.

Frame	Caffe	MXNet	Tensor Flow	Theano	Torch
Source	UC Berkeley	Baidu distributed machine learning community	Google	University of Montreal	Facebook, Google, DeepMind
Main language	C++	C++	C++	Python	Lua
Secondary language	Python, Matlab	Python, R, JS, Julia, Scala, Matlab, Go	Python	--	--
Hardware	CPU/GPU	CPU/GPU/Mobile	CPU/GPU/mobile	CPU/GPU	CPU/GPU/FPGA
Distributed	--	support	support	--	--
Operation System	cross-platform	cross-platform	cross-platform	cross-platform	Do not support windows
Strengths	Fast train speed; Code easy, rapidly development, more official examples	good optimized memory External dependence	Many network types more than CNN	Python Deep learning library Better speed and stability	Simple Model Highly modular
Weaknesses	Less flexible Distributed not support	High technical threshold	High technical threshold	Difficult to debug	Lua language less common Less other machine library

Table 1. Current mainstream deep learning frameworks

### 2.3 Building sample production scheme

The production of building samples is briefly illustrated in the overall technical scheme. In general, we use screenshots tools to capture building sample library of multi-source and multi-scale spatial resolution on Google Earth, covering domestic and foreign the building entity of different shapes, different attitudes. Then, use ArcGIS to make labels by vectorizing. Through the preprocessing, the samples and labels are transformed into the input data format of the network model.

### 2.4 Neural network model

#### 2.4.1 U-Net

Network structure is as shown in Figure 3, blue represents the convolution and activation function, the grey indicates the

replication, red for the down sampling, green means up sampling and then convolution, *conv* 1 \* 1 represents the convolution operation whose nuclear is 1 \* 1. This network has only

convolution and down sampling instead of full connection. This is an end-to-end network model where the input and output are both images.

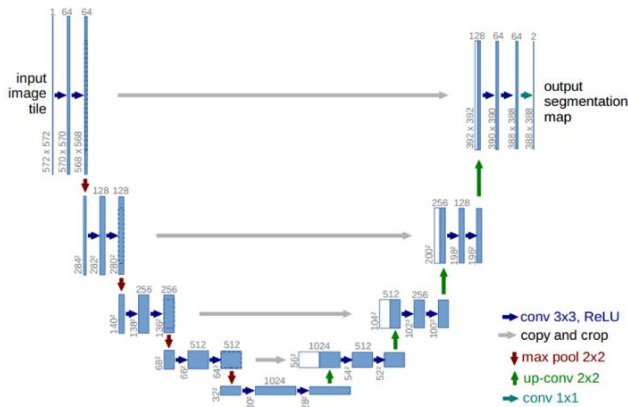
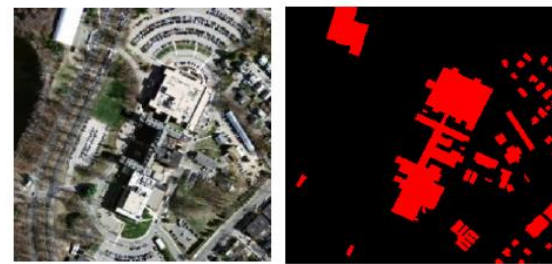


Figure 3. U-Net structure

### 2.5 Model training and prediction

The following Figure 4 shows the process of model training and prediction:



American building sample library, resolution: 1 m



U-Net training



Worldview3 or Gf-2 and other domestic data.

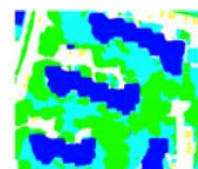
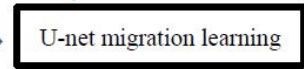


Figure 4. Process of model training and prediction

### 3. EXPERIMENTAL ANALYSIS

Worldview3 satellite, which has 0.31 meters resolution, is adopted in this experiment. The experimental data are Worldview3 satellite images of three areas where the buildings are relatively dense in some state in America. This experiment detected the buildings from those satellite images respectively. The training sample uses the construction sample library in U.S. and Worldview3 data. Deep learning model trains using the 20th level of U-net model.

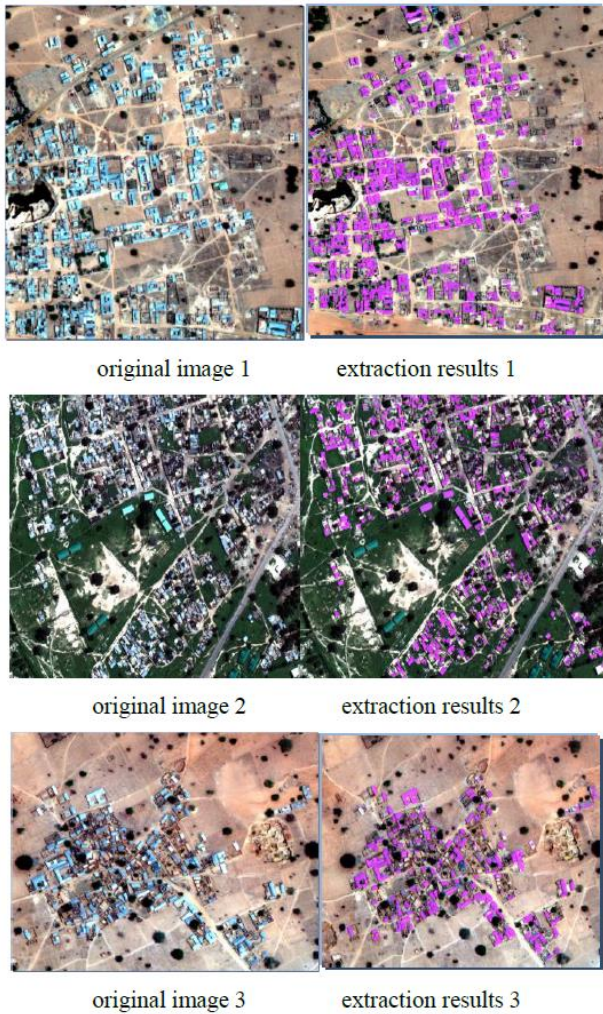


Figure 5. Experimental Worldview3 satellite images

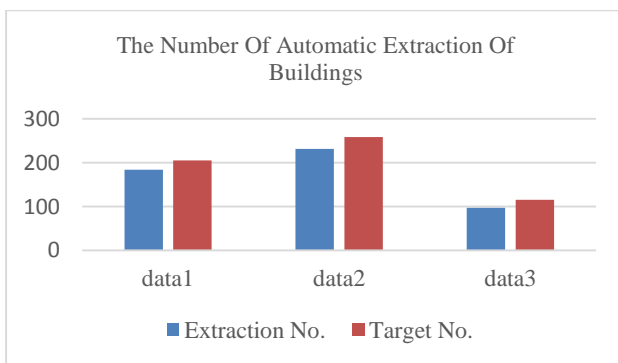


Figure 6. Chart of statistics of automatic extraction of buildings

From the experimental results indicated in Figure 6, the recognition rate of the three experimental data groups was 89.7%, 89.5% and 84.3% respectively. The overall recognition rate was over 87%, which basically reached the level of artificial interpretation recognition. The main reason for the missing recognition of buildings is that the area is too small or blocked by other buildings. Also, it reflects that the noise is quite small and the ability of resistance to noise of the training sample is insufficient.

#### 4. CONCLUSION

Compared with the traditional classification of artificial interpretation, it can identify a wide range of buildings rapidly and effectively using artificial intelligence deep learning through the study of vast amounts of samples. It can not only identify efficiently, but also ensure the accuracy. This method plays an important role in land use monitoring, illegal land discriminant and survey of land and resources. In the future the diversity and complexity of the training samples will be increased, which makes the deep learning model more robust and has higher accuracy rate of extraction.

#### REFERENCES

- Crevier, Daniel, 1993. *AI: The Tumultuous Search for Artificial Intelligence*, New York, NY: BasicBooks, pp. 115-117.
- Arun P V., 2013. A CNN based hybrid approach towards automatic image registration. *Geodezijos Darbai*, pp. 121-128.
- Peng L, Zhao ZM, Yang J, et al., 2006. A study on multi-spectral remote sensing image texture segmentation based on wavelet domain hidden Markov tree model. *Journal of wuhan university of technology (traffic science and engineering)*, pp. 561-564.
- Ramzi P, Samadzadegan F, Reinartz P., 2014. An AdaBoost Ensemble Classifier System for Classifying Hyperspectral Data. *Photogrammetrie - Fernerkundung - Geoinformation*, pp. 27-39.
- Ronneberger O., 2015. Invited Talk: U-Net Convolutional Networks for Biomedical Image Segmentation, pp. 234-241.
- Sastry C S, Ravindranath M, Pujari A K, et al., 2007. A modified Gabor function for content based image retrieval. *Pattern Recognition Letters*, pp. 293-300.
- Seetha H, Saravanan R, Murty M N., 2012. Pattern Synthesis Using Multiple Kernel Learning for Efficient SVM Classification. *Cybernetics & Information Technologies*, pp. 77-94.
- Tang CH, 2015. A detection method based on gradient directional histogram for overlooking pedestrians. *Pattern recognition and artificial intelligence*, pp. 19-26.
- Zhuo L, Zhang J, Zhao Y, et al., 2013. Compressed domain based pornographic image recognition using multi-cost sensitive decision trees. *Signal Processing*, pp. 2126-2139.