Peer-reviewed version available at Sensors 2019, 19, 1164; doi:10.3390/s19051164

Semantic Segmentation of Building Roof in Dense and Diverse Urban Environment with Deep Convolutional Neural Network

Yuchu Qin,[1] Yunchao Wu,[2] Bin Li,[1] Shuai Gao,[1] Miao Liu,[3] Yulin Zhan,[1]

1.State Key Lab of Remote Sensing Sciences, Institute of Remote Sensing and Digital Earth (RADI), Chinese Academy of Sciences (CAS), Beijing, China

2.Beijing Municipal Institute of City Planning & Design, Beijing, China

3 South Dakota State University,Brookings,SD,USA

Abstract: This paper presents a novel approach for semantic segmentation of building roof in dense urban environment with Deep Convolution Neural Network (DCNN) using imagery acquired by a Chinese Very High Resolution (VHR) satellite mission, i.e. GaoFen-2 (GF-2). To provide an operational end-to-end work flow for accurate build roof mapping with feature extraction as well as image segmentation, a fully convolutional DCNN with both convolutional and deconvolutional layers is designed to perform the VHR image analysis for labeling pixels. Since the diverse urban patterns and building styles in large areas, sample image data sets of building roof and non-building roof are collected over different metropolitan regions in China. We selected typical cities with dense urban environment in each metropolitan region as study areas for collecting training and test samples. High performance cluster with GPU-mounted workstations is employed to perform the model training and optimization. With the building roof samples collected over different cities, the predictive model with multiple NN layers is developed for building roof labeling. The validation of the building roof map shows that the overall accuracy(OA) and the mean Intersection Over Union(mIOU) of DCNN based segmentation are 94.67%, 0.85 respectively, while CRF-refined segmentation achieved OA of 94.69% and mIOU of 0.83. The results suggest that the proposed approach is a promising solution for building roof mapping with VHR images over large areas across different urban and building patterns. With the operational acquisition of GF2 VHR imagery, it is expected to develop an automated pipeline for operational built-up area monitoring and timely update of building roof map over large areas.

Keywords: VHR image, Building roof, Segmentation, GF2, Deep Convolution Neural Network

1. Introduction

Urbanization is a process that human being significantly affect the natural environment of land surfaces, it not only induces changes of land cover and land use, it also has profound influences in daily life of our society[1]. Building is the essential part of urban environment, and it plays a vital role in human life as basic infrastructure of human settlement[2]. Accurate and timely map of building roof is crucial for urban planning, infrastructure construction and environmental management, especially in rapid urbanization areas[3],[4]. However, the conventional solutions of building roof mapping, e.g. field survey, manual delineation upon imagery, are time-consuming and labor-intensive endeavors. The widely available Very High Resolution (VHR) satellite imagery is an unique data source for building roof mapping[5],[6]. Many efforts have been carried out to develop approaches for automatic building roof mapping, However, most of these methods are based on specific features and rules, and low-level features were extracted from VHR images for the building mapping, e.g. histogram of oriented gradients (HOG)[7],[8]. In addition, some other studies for building detection also tried use shadow information[9],[10]，the graph theory[11],[12], MRF-based approach[13]. The classification methods employing the advantages of multispectral information[14],[15]. [16] proposed a post-processing framework for building extraction with VHR imagery, the framework relies on a morphological building index (MBI) which integrating spectral, geometrical, and contextual information for building mapping, the experiments suggested that the proposed framework would achieve a promising results. However, this type of methods rely on the selection of features and image type, it is hard to develop a model for operational building mapping over large areas, especially for dense urban environment with different building types.

As an import machine learning approach, Neural Networks(NN) is inspired by the process of simulating the recognition process of brains to perform recognition tasks. Since 2006, a series techniques and strategies, e.g. layer-wise training and pre-training, Restricted Boltzmann Machine(RBM), Recurrent Neural Networks(RNN) were proposed for NN with multiple hidden layers for large scale learning problems[17]. Deep NN with multiple hidden layers between the input and output layers are applied for learning data representations in recognition tasks, e.g. image classification, scene understanding, speech recognition. Convolution Neural Network (CNN) is designed for feature learning as well as inference in image classification, segmentation and scene understanding [18]. CNN takes a moving window as a filter for capturing features in image space, and the features then can be applied for classification or segmentation [19]. CNN not only could reduce the number of weights and bias in NN by sharing parameters of the deep neural networks, it also provides a promising solution to combine features both spectral and spatial textural domains for scene understanding and segmentation. Currently, with high performance computational computers with GPUs, CNN has been widely applied in vision recognition tasks and has achieved the state of the art results [18], [20]-[23]. There are many efforts for building mapping with DCNN, [24] proposed the automatic building extraction methods in aerial scenes using convolutional neural networks. A network was designed with novel components which are easy to implement, it enables the network to learn hierarchical features for segmenting buildings. [25] proposed a method which automatically generates a full resolution binary building mask using a Fully Convolution Network (FCN) architecture and achieved a promising result for building roof mapping. [26] proposed a

two-stage CNN model to detect rural buildings in high-resolution imagery. The experiments showed that the two-stage CNN model can effectively reduce the complexity of the background and improve the efficiency, achieving a building detection accuracy of 88%. Besides these, the building detection task can also be solved as part of the land use classification problem with the deep learning methods. For example, [27] investigated the convolutional neural networks for large-scale remotely sensed image classification, they proposed an end-to-end framework for the dense, pixel-wise classification of satellite imagery with CNNs.

Currently, high resolution satellite could capture VHR imagery that covers large areas timely, with pre-trained CNN model, it is feasible to conduct large area built-up mapping at individual building level. However, operational mapping with VHR imagery over large areas is still a costly task with current commercial VHR images. GF-2 (Gaofen-2) is one of the series satellite mission in China High-resolution Earth Observation System (CHEOS) [28]. It was launched is on August 19, 2014 with PMS which acquires panromatic and multispectral image simultaneously. It is expected to provide a long-term VHR satellite imagery with an affordable cost for China and other areas, especially developing countries. The image data acquired by GF-2 has been widely used in land resource monitoring, environmental protection, urban land cover mapping as well as forest applications. This study presents our effort of building roof segmentation with DCNN over mega-cities in China, using the GF-2 PMS data and DCNN model, it is expected to develop an automatic pipeline for operational building mapping with Chinese GF-2 VHR imagery over large areas at low cost, regardless of the location of the areas and the acquisition data of the image.

2. Data sets and study areas

The GF-2 PMS is capable of collecting satellite imagery with GSD (Ground Sampling Distance) of 0.8m panchromatic and 3.2m multispectral bands on a swath of 45 km. Table 1 illustrate basic configurations of the sensor.

Table 1. Configuration of PMS

Sensor	Spatial Resolution(m)	Spectral Bands(μm)
PMS-panromatic	0.8	0.45-0.90
PMS-multispectral	3.2	0.45-0.52;0.52-0.59; 0.63-0.69;0.77-0.89

To cover the diverse urban patterns and building styles in China, we selected typical cities in different metropolitan regions as study areas for collecting training and test GF-2 PMS images, Table 2 shows the cities and acquisition date of the images. With the assumption that imagery acquired in growing season would provide better vision effect in both spectral and spatial domains, we only choose images acquired from June to September in 2016 with cloud cover less than 5% for further processing, and the total 7 images are selected for this study.

Table 2. Area and acquisition date of images

City	Region of China	Acquisition Date
Beijing	North	20160827
Shenyang	Northeast	20160612
Chengdu	Southwest	20160711
Guangzhou	South	20160723
Wuhan	Central China	20160901
Shanghai	Southeast	20160602
Urumqi	Northwest	20160625

2.1. Image preprocessing and training data collection

Though the orientation of PMS images are well calibrated, high precision co-registration between the panchromatic and multispectral images acquired by PMS is performed to keep rigorous geometrical alignment in image space. With the reference of corresponding panchromatic imagery, visual selection of ground control points is carried out for georectification of multispectral images. The overall mean registration error for all images is less than 1.0m. With the panchromatic and the corresponding rectified multispectral image, the image pan-sharping is conducted to obtain fused images with spatial resolution of 1.0m with Gram-Schmidt algorithm. Figure 1 illustrates the PMS panchromatic image, the true color composites of PMS multispectral image, and the true color composites of pan-sharped images, these images suggests that the pan-sharped images of PMS could characterize the building boundaries with pixel resolution of 1.0m.

2.2. Collection of sample images

With the pan-sharped high resolution imagery, building roof samples for training and test are collected by manual delineation, though the basic unit of delineation is individual building, for areas with close connected buildings, neighbour buildings are delineated together since it is hard to separate individual buildings. The manual delineated polygons are converted to binary images as building roof mask, where the binary image values of 1 and 0 means building roof, rest of land cover types respectively.

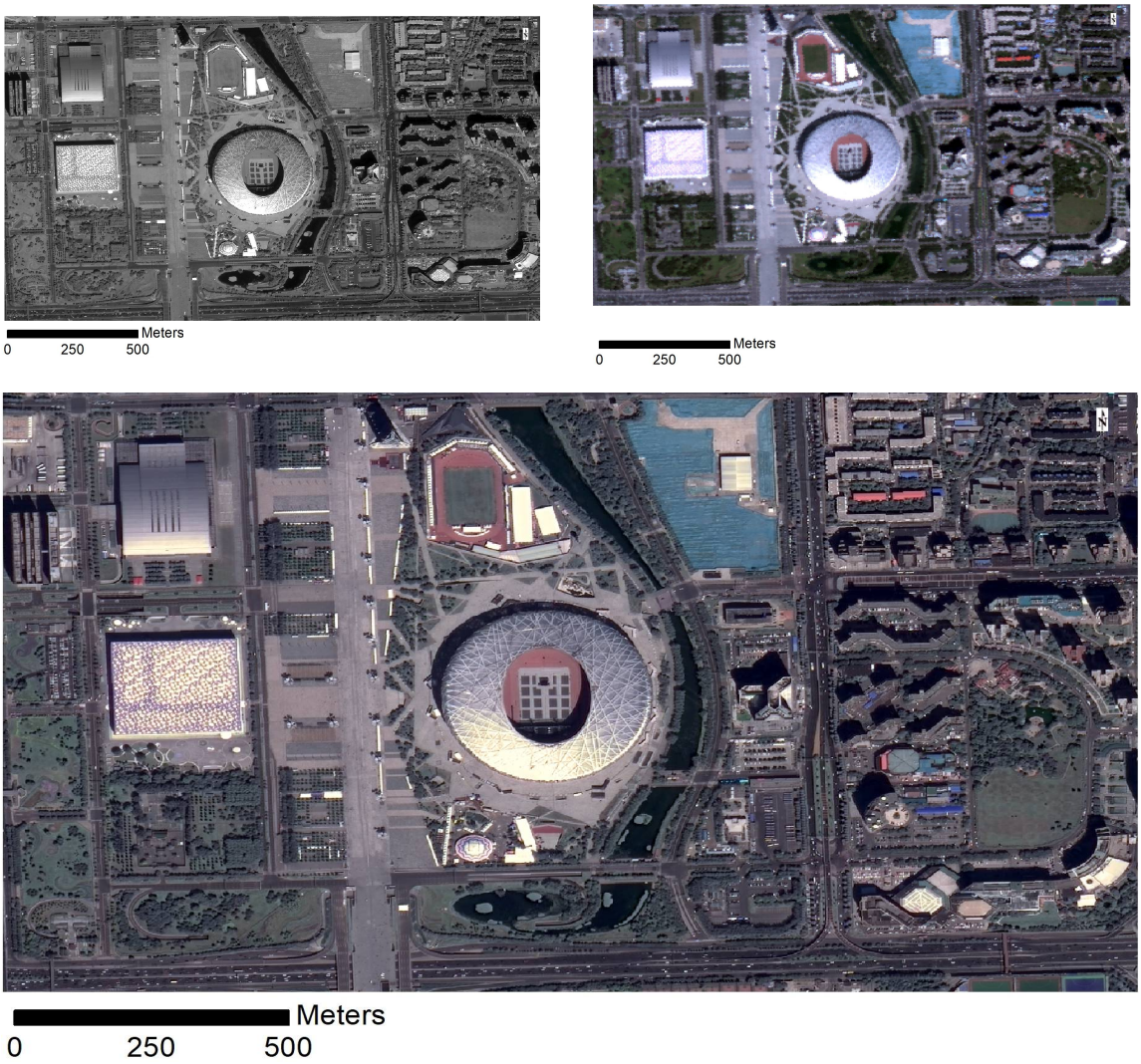

Figure 1. Sample images captured by GF-2 PMS sensors: the 3 images cover the same area around National Stadium at Beijing, upleft is panchromatic image with resolution of 1.0m, upright is the true color composite of multi-spectral image with resolution of 4.0m, and the bottom one is the pan-sharped image of the panchromatic and multi-spectral images, and the acquisition date of the images is 2016/08/27.

3.Methodology

In this paper, a DCNN is developed to perform the image segmentation and per-pixel labeling for building roof with VHR satellite imagery. Generally, a DCNN consists of convolutional and pooling layers, the convolutional layers perform feature extraction and the pooling layers summarize the features by aggregating neighbour pixels in imagery, while both convolutional and pooling layers reduce image size, the deconvolutional layer provide a way to upsample image to original resolution, the detailed description of the three type operations is well introduced in [29].

3.1. Design of DCNN

There are many well known CNN models for vision recognition tasks, e.g. AlexNet, VGG,

GoogLeNet, RestNet. To develop an accurate model with high computational efficiency, the VGG-16 is selected as basic frame of DCNN for building roof segmentation with GF2 VHR satellite imagery. The VGG-16 model is defined as a combination of series convolutional and pooling layers, it is a promising model for image analysis. However, the original VGG-16 model was designed to provide summarized semantic information at image level, [30] proposed a fully connected network (FCN) for dense prediction of images, with the use of deconvolutional operation to upsample CNN layers, FCN is an end-to-end framework for image segmentation at pixel level. Since the building roof mapping requires dense prediction at single pixel level, the deconvolutional layer is also adopted to recover the image size for dense per-pixel prediction. Figure 2 illustrates the architecture of the DCNN for building roof segmentation with GF-2 imagery, with both convolutional and deconvolutional layers. It is well addressed that the distribution drift of data, i.e. image Digital Number(DN) in this study, between layers in DCNN may significantly reduce the computational efficiency and segmentation accuracy. Batch normalization layers, as additional layers in fully convolutional DCNN, are therefore placed to perform feature normalization between layers. The fully convolutional network could perform feature extraction with spatial-spectral information of imagery as well as predict per-pixel class label.

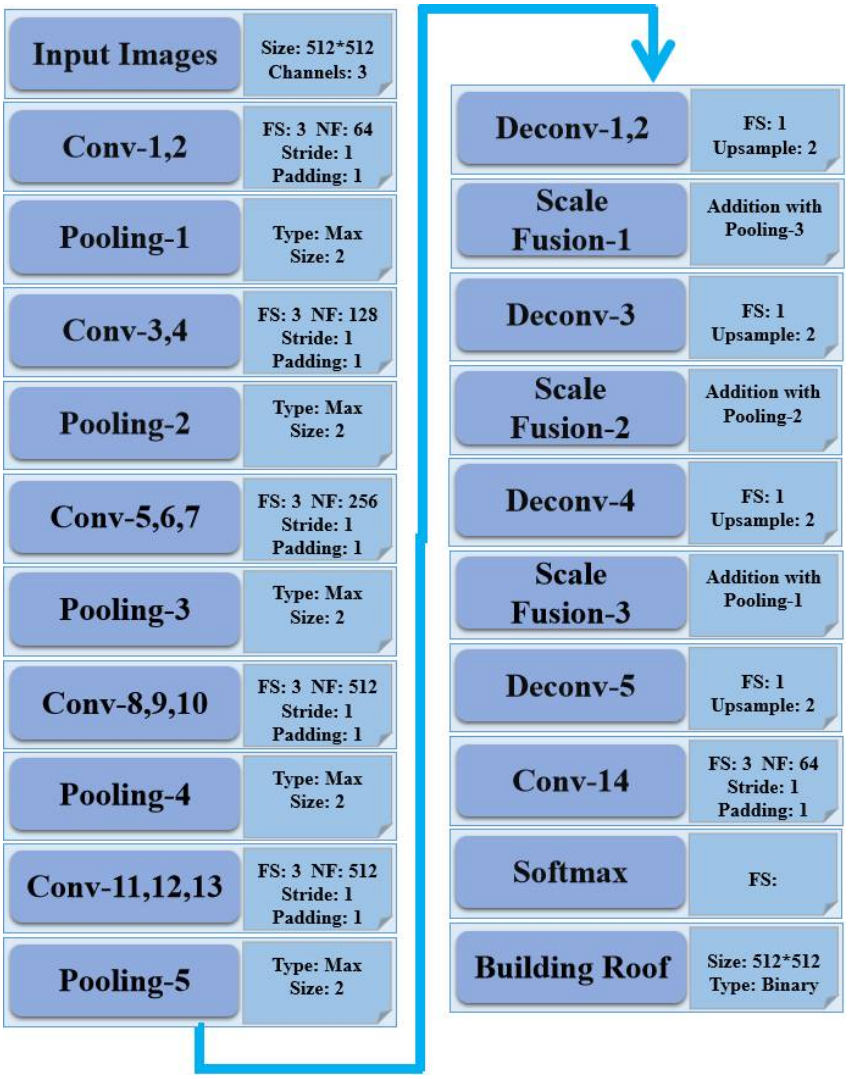

Figure 2. Architecture of the designed DCNN for building roof segmentation

Peer-reviewed version available at Sensors 2019, 19, 1164; doi:10.3390/s19051164

3.2. DCNN model training and inference

To reduce the effect of varied illumination and atmosphere in images acquired at different time over different areas., Eq(1) is employed to normalize the original image DN values at scene level. It is expected to provide comparable images for the training and test. The normalized images are then cliped to sample images with the size of 512x512 pixels. Totally 1460 sample images are selected, 80% of the sample images are randomly selected for training, while the 20% are selected for validation.

$$DN_T = DN_{Ori} - \frac{\sum_{i=0}^{M \bullet N} DN_i}{M \bullet N} \quad (1)$$

DN_T DN_{Ori}

Where DN_{ori} and DN_T are the DN values of original and transformed images, respectively, and M,N are the image height and width.

The DCNN model is implemented upon the open source deep learning package developed by Google, i.e. Tensorflow. High performance computational workstation with GPU, i.e. NVIDIA TitanX, is employed to perform the DCNN model training and inference, and the batch size in the training is 8. To prevent the DCNN from over-fitting, the dropout rate in training stage is 0.50.

2.2. Post-processing with Conditional Random Field(CRF)

It is well established that image segmentation with DCNN could smooth the boundaries of objects [Jonathan Long et al;., 2015], to refine the segmentation results, CRF is employed to regularize the shape of building segmentation. As a probabilistic graphical model, CRF has been widely applied in image analysis, it provides a solution for image segmentation by connecting image pixels with neighbors using a graph, the graph consists of nodes, which are single image pixels. The nodes, together with edges of the graph, are utilized to characterize spatial-spectral relationship. The posterior probability of the model can be characterized by the Gibbs distribution.

$$P(Y \mid X) = \frac{1}{Z(X)} \bar{P}(Y, X) \quad (2)$$

$$\bar{P}(Y, X) = \exp(\sum_i w_i * f_i(Y, X))$$

$$Z(X) = \sum_Y \exp(\sum_i w_i * f_i(Y, X))$$

Where P(Y|X) is the normalized conditional probability of the event X and Y, which is characterized by a Gibbs distribution, P(Y,X) is the joint distribution probability, which is defined as an exponential function of weighted factors, and the factors are customized model for specific problems. [31] proposed a fully connected CRF model for image segmentation upon pixel level, the joint distribution probability is parameterized with the summary of unary and pairwise potentials(Eq 3).

$$E(x) = \sum_i \psi_u(x_i) + \sum_{i<j} \psi_p(x_i, x_j) \quad (3)$$

E(x) is the Gibbs energy, $\sum_i \psi_u(x_i)$ is the unary term, which is computed independently for each pixel, denotes pairwise potentials, which $\sum_{i<j} \psi_p(x_i, x_j)$ is designed for characterizing the relationship between pixel with its neighbors. The detailed definitions of the unary and pairwise potentials can be found in [31]. With mean field approximation algorithm, the maximization of the posterior probability of Gibbs distribution is calculated for inference of the model. In this study, the fully connected CRF model was employed for building roof map segmentation.

3. Results and Discussion

3.1. Training of DCNN model

The batch size for the interactive training is 8, while the total number of sample images is 1200, with 3000 epochs for the optimization of the DCNN model, theoretically, the GPU workstation would run the optimization procedure (1200 images * 3000 epochs)/8 images = 450,000 times. Figure 3 shows the change of loss in the training step, it suggest that the loss gradually decreased during the training stage. The moving average loss is calculated with moving window size of 100, while the maximum value of moving average error approximately is 1.2, the minimum value of moving average error is 0.05, and it does not have significant change in the end of training, that means the DCNN model converges to moving average error of 0.05.

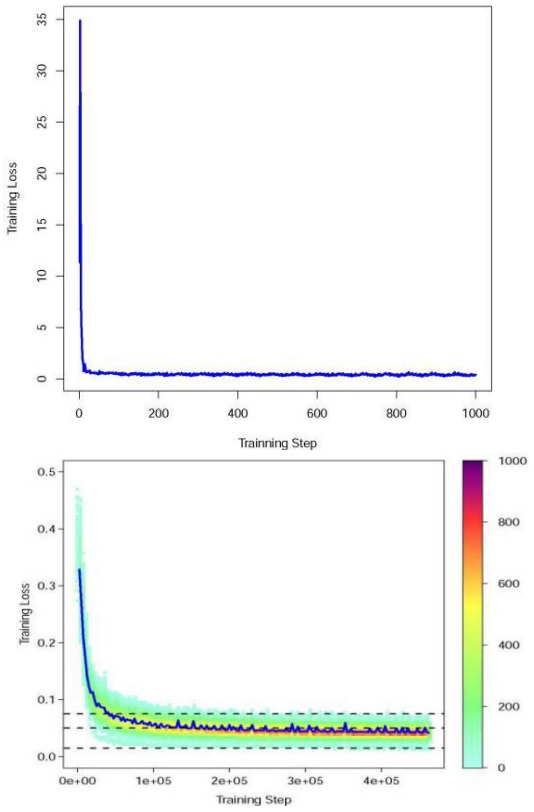

Figure 3. The change of loss value in model training stage: X axis and Y axis denote the training steps and the training loss, respectively; the upper figure shows the process with training step of 1-1000,the bottom figure shows the entire training process; For the bottom figure, the color denotes the density of steps, while the dash lines are training bounds with loss values of 0.075,0.005,0.015, respectively; the blue line is moving average of training losses with window size of 1000.

3.2. Qualitative assessment

Basically, the semantic segmentation of building roof is a binary classification, i.e. image pixels are labeled either building roof or non-building roof. With the DCNN model trained by the interactive optimization, the segmentation of building roof, i.e. the prediction of per-pixel label, is performed by the trained DCNN model, and CRF is also employed to refine the segmentation results. Figure 4 illustrates the building roof segmentation results with fused GF-2 PMS imagery over urban areas, i.e. dense areas with business buildings, dense urban area with apartments, low density area with apartments, urban area with single family houses and sparse urban areas, are selected for qualitative assessment, the sample images cover different urban types. The indicators for accuracy assessment, i.e. True Positive (TP, building roof pixel was correctly classified), True Negative (TN, non-building roof pixel was correctly classified), False Positive (FP, non-building roof pixel was classified as building roof), False Negative (FN, building roof pixel was classified as non-building roof), are identified for visualizing the segmentation results in details. Visual inspection upon the segmentation results suggests that the DCNN could generate a promising building roof map over different urban types. From the colorized accuracy indicator map, we can observe that many individual buildings are segmented as connected image objects. While CRF was introduced as an efficient solution for optimizing the segmentation results, it doesn't significant change the building roof map. It is also observed that the segmented building roof have smooth boundaries, even after the regularization operations with CRF.

3.3.Quantitative assessment

The results of building roof segmentation are binary images, to obtain a quantitative assessment of the segmentation results, the accuracy at both pixel level and image segment level are calculated, the per-pixel level accuracy is characterized by overall accuracy(OA), it is defined by Eq(4).

$$OA = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{TN} N_{FP} + N_{FN}} \tag{4}$$

Where NT_P, NF_N are the numbers of pixels which are correctly labeled, while N_{TN}, N_{FP} are the numbers of pixels which are incorrectly labeled. The mean overall accuracy is averaged with OA values of all image samples for test. The mean Intersection-over-union (mIOU) is applied to characterize accuracy at image segment level, and mIOU is calculated using Eq(5).

$$IOU = \frac{N_{GT} \bigcap N_{\mathrm{DR}}}{N_{GT} \bigcup N_{\mathrm{DR}}} \tag{5}$$

$$mIOU = \frac{\sum_{i=0}^{k} IOU_i}{k}$$

Where N_{GT} is the total number of pixels of the ground truth patch, i.e.manual delineation mask of building roof, NDR is the total number of pixels of the corresponding building roof detected by the DCNN,k is the total number of segmentation patches.

The OA and mIOU for quantitative assessment of segmentation result are given in Table 3, the OA values of both the two results, i.e. DCNN segmented and CRF-refined building roof are 94.67% and 94.69%, respectively, it suggests the DCNN worked very well for building roof segmentation upon the Chinese GF-2 imagery, and the CRF could improve the segmentation results. However, the CRF based refinement didn't make significant change upon the original segmentation.

TABLEIII. OVERALL ACCURACY AND MIOUS OF THE SEGMENTATION RESULTS.

Methods	OA	mIOU
DCNN	94.67%	0.83
DCNN-CRF	94.69%	0.83

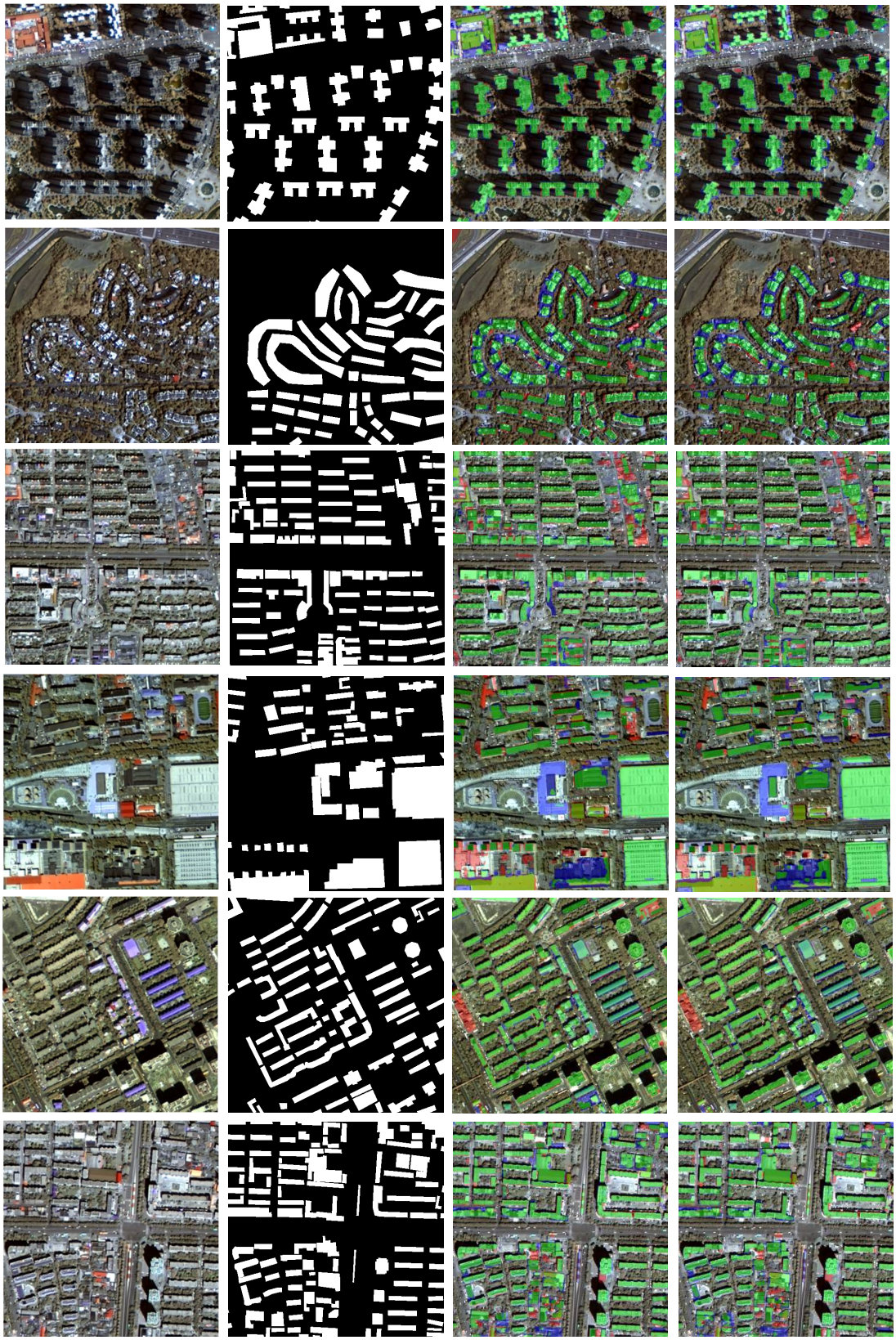

Figure 4. Visual comparison of building roof segmentation in different urban environment: for each row, images from left to right are true color composite GF-2 PMS imagery(512Pixel*512Pixel), manual delineation of building roof, DCNN segmentation results and CRF-refined building roof segmentation, respectively; for the images with segmentation results, green mask is TP, blue mask is FN and red mask is FP.

4.Conclusion

This paper presents approach for semantic segmentation of building roof in dense and diverse urban environment with Chinese GF2 PMS images, a fully connected DCNN was introduced to carry out the feature extraction as well as pixel labeling, the experiment is conducted for VHR images acquired by PMS on-board of Chinese GF2 satellite, the DCNN could extract features at both spectral and spatial domains, and the features are also fused together for dense prediction at pixel level for the pixel level labeling. The quantitative assessment of the semantic segmentation results suggest that the fully connected DCNN achieved high accuracy building roof map. However, the convolutional operation performs feature extraction with filters of moving windows, and the pooling operation aggregates neighbour pixels with summation, both of the two operations would smooth images and eventually change boundaries of building. The future work on the algorithm improvement would focus on the fusion of low level geometrical features, i.e. boundaries of objects with high level semantic features for discriminating edges of building roof.

The generation of DCNN model relies on large volume of positive and negative samples, the collection of the massive building roof samples is not only costly, it is also difficult to cover all scenarios in the practice of inference, especially in dense and diverse urban environment over large areas. Currently, there are some efforts are conducting to collect building roof globally over different areas, e.g. SpaceNet, hosted by DigitalGlobe on Amazon Web Service(AWS), and Volunteering Geographical Inventory (VGI) based approach is also investigated for sample collection, it would be a alternative for large scale sample collection.

In summary, we can conclude that DCNN performs well on semantic segmentation of building roof in dense urban environment using VHR images, with the advances in high performance computers and the collection of massive sample images, it is also a promising solution for operational building roof mapping over large areas with different building types and distribution patterns.

Acknowledgments:

This work was supported by the 100 Talents Program of the Chinese Academy of Sciences.

REFERENCES

[1]United Nations, Department of Economic and Social Affairs, Population Division. *World Urbanization Prospects: The 2014 Revision*.2015, (ST/ESA/SER.A/366).

[2] Alshehhi, R.; Marpu, P. R.; Wei, L. W.; Mura, M. D. Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks. *ISPRS J PHOTOGRAMM*. **2017**,*130*, pp. 139-149.

[3] Xie, Y.;Weng, Q. Updating urban extents with nighttime light imagery by using an object-based thresholding method. *Remote Sens. Environ*. **2016**,187, pp. 1-13.

[4] Seto, K. C.; Golden, J. S.; Alberti, M.; Nd, T.B. Sustainability in an urbanizing planet. *Proc Natl Acad Sci USA*. **2017**,114, 34, pp. 8935-8938.

[5] Sohn, G.; Dowman, I. Data fusion of high-resolution satellite imagery and lidar data for automatic building extraction. *ISPRS J PHOTOGRAMM*. **2007**, 62, pp. 43-63.

[6]Awrangjeb, M.; Zhang, C.; Fraser, C. S. Automatic extraction of building roofs using lidar data and multispectral imagery. *ISPRS J PHOTOGRAMM*. **2013**, 83, pp. 1-18.

[7] Zhang, Y. Optimisation of building detection in satellite images by combining multispectral classification and texture filtering. *ISPRS J PHOTOGRAMM*. **1999**, 54, pp. 50-60.

[8] Cheng, G.; Han, J. A survey on object detection in optical remote sensing images. *ISPRS J PHOTOGRAMM*. **2016**,117, pp. 11-28.

[9]Ok, A.O. Automated detection of buildings from single VHR multispectral images using shadow information and graph cuts. *ISPRS J PHOTOGRAMM*. **2013**, 86, pp. 21-40.

[10]Ngo, T. T.; Mazet, V.; Collet, C.; Fraipont, P.D. Shape-based building detection in visible band images using shadow information. *IEEE J SEL TOP APPL*. **2016**, 99, pp. 1-13.

[11] Kim, T.; Muller, J. P. Development of a graph-based approach for building detection. *IMAGE VISION COMPUT*. **1999**,17, pp. 3-14.

[12] Sirmacek, B.; Unsalan, C. Urban-area and building detection using sift keypoints and graph theory. *IEEE T GEOSCI REMOTE*. **2009**,47, pp. 1156-1167.

[13] Grinias, I.; Panagiotakis, C.; Tziritas, G. MRF-based segmentation and unsupervised classification for building and road detection in peri-urban areas of high-resolution satellite images. *ISPRS J PHOTOGRAMM*. **2016**,122, pp. 145-166.

[14]San, A. D. K.; Turker, M. Support vector machines classification for finding building patches from ikonos imagery: the effect of additional bands. *J APPL REMOTE SENS*. **2014**, 8, 683-694.

[15]Sumer, E.; Turker, M. An adaptive fuzzy-genetic algorithm approach for building detection using high-resolution satellite images. *COMPUT ENVIRON URBAN*. **2013**, 39, pp. 48-62.

[16]Huang, X.; Yuan, W.; Li, J.; Zhang, L. A new building extraction postprocessing framework for high-spatial-resolution remote-sensing imagery. *IEEE J SEL TOP APPL*. **2017**,10, pp. 654-668.

[17] Lecun, Y.; Bengio, Y.; Hinton, G. *Deep learning. Nature.* **2015**, 521, pp. 436-444.

[18] Krizhevsky, A.;Sutskever, I.; Hinton, G. E. ImageNet classification with deep convolutional neural networks. International Conference on Neural Information Processing Systems. **2012**, 60, pp.1097-1105.

[19]LeCun, Y.; Kavukvuoglu, K.; Farabet, C. Convolutional Networks and Applications in Vision. In Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS'10). **2010**, pp. 253-256.

[20]Ross, G.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. **2014**, pp.580-587.

[21]Ren,S.; Girshick,R.; Sun, J. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE T PATTERN ANAL.***2017**, 39, pp. 1137-1149,.

[22]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. **2016**, pp.770-778.

[23] He,K.; Gkioxari,G.; Dollár,P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision. **2017**, pp. 2980-2988.

[24]Yuan,J. Automatic building extraction in aerial scenes using convolutional networks. arXiv preprint, **2016**, arXiv:1602.06564.

[25]Bittner, K.; Cui, S.; Reinartz, P. Building Extraction from Remote Sensing Data using fully convolutional Networks.ISPRS Hannover Workshop: Hrigi . **2017**, XLII-1/W1, pp.481-486.

[26]Sun, L.; Tang, Y.; Zhang,L. Rural building detection in high-resolution imagery based on a two-stage cnn model. *IEEE GEOSCI REMOTE S.* **2017**, 14, pp. 1998-2002.

[27]Maggiori, E.;Tarabalka, Y.; Charpiat, G.; Alliez, P. Convolutional neural networks for large-scale remote-sensing image classification. *IEEE T GEOSCI REMOTE*. **2016**, 55, pp. 645-657.

[28]Tong, X.; Zhao, W.; Xing, J.; Fu, W. Status and development of China High-Resolution Earth Observation System and application. In Proceedings of IEEE Symposium on Geoscience and Remote Sensing. **2016**, pp. 3738-3741.

[29] Ian, G.; Yoshua, B.; Aaron, C. Deep Learning. MIT Press. **2016**.

[30] Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. **2015**, 79, pp. 3431-3440.

[31] Krähenbühl, P.; Koltun, V. Efficient inference in fully connected crfs with gaussian edge potentials. *In Advances in neural information processing systems*. **2011**, pp. 109-117.