ROBUST BUILDING FOOTPRINT EXTRACTION FROM BIG MULTI-SENSOR DATA USING DEEP COMPETITION NETWORK

M. Khoshboresh Masouleh *, M.R. Saradjian

School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran - (m.khoshboresh, sarajian)@ut.ac.ir

Commission VI, WG VI/4

KEY WORDS: Building Extraction, Big Multi-Sensor Data, Deep Learning, Convolutional Network, Cloud Computing, Colab

ABSTRACT:

Building footprint extraction (BFE) from multi-sensor data such as optical images and light detection and ranging (LiDAR) point clouds is widely used in various fields of remote sensing applications. However, it is still challenging research topic due to relatively inefficient building extraction techniques from variety of complex scenes in multi-sensor data. In this study, we develop and evaluate a deep competition network (DCN) that fuses very high spatial resolution optical remote sensing images with LiDAR data for robust BFE. DCN is a deep superpixelwise convolutional encoder-decoder architecture using the encoder vector quantization with classified structure. DCN consists of five encoding-decoding blocks with convolutional weights for robust binary representation (superpixel) learning. DCN is trained and tested in a big multi-sensor dataset obtained from the state of Indiana in the United States with multiple building scenes. Comparison results of the accuracy assessment showed that DCN has competitive BFE performance in comparison with other deep semantic binary segmentation architectures. Therefore, we conclude that the proposed model is a suitable solution to the robust BFE from big multi-sensor data.

1. INTRODUCTION

In recent years, robust automated algorithm development for the extraction of building footprints from remotely sensed data is a hot topic for research and commercial projects (Bi et al., 2019). In practice, there are two issues that are essential in building footprint extraction (hereafter called BFE for short). First, data source selection that plays an important role in information extraction. Second, the appropriate knowledge such as deep learning (DL) for accurate and efficient data processing

1.1 Data Source

The main factors of data source selection for BFE are related to separation between buildings from non-buildings (spatial resolution considerations), the confusing effect of vegetation-cover on building detection (spectral resolution considerations), and the shade of building/non-building objects and lighting conditions (types of sensors considerations).
The majority of related works that uses the multi-sensor data consist of very high spatial resolution multispectral images and light detection and ranging (LiDAR) data (Huang et al., 2019; Li et al., 2013; Rottensteiner et al., 2003; Volpi and Tuia, 2018; Yang et al., 2018). LiDAR data (also known as point clouds) and digital surface models (DSMs) generated by aerial platform equipped with airborne laser scanning, such as unmanned aerial vehicle or aircraft are applicable for the automatic BFE, because these data provide the geometrical features of buildings shapes (Cai et al., 2019; Jung and Sohn, 2019; Rottensteiner et al., 2007; Sohn and Dowman, 2007). Moreover, fusion of LiDAR point clouds and very high spatial resolution multispectral images offers an efficient data source for BFE (Huang et al., 2017). Hence, on the basis of the fresh information mentioned above, still, the use of these combined data can be a very convenient source of data for BFE and also has many key issues in processing unresolved, particularly suited to big multi-sensor data. Since the big multi-sensor data has a massive volume of geospatial aerial or satellite data, it is extremely difficult or impossible to process using traditional algorithms (Philip Chen and Zhang, 2014; Yang et al., 2017).

1.2 Deep Learning

In the past few years, DL approaches play a crucial role in analysing the big image data (Khoshboresh Masouleh and Shah-Hosseini, 2019a; Maggiori et al., 2017; Masouleh and Shah-Hosseini, 2018; Samuel R. et al., 2019). The DL approaches incorporate two influential concepts in an optimal big data analysis workflow for image data (Wu et al., 2018). Sparse topological connectivity in convolutional neural networks which is the most common type of DL and weight sharing across deep network can be used to improve the generalization in DL. DL allows computational algorithms of complex structures that are fused of various processing layers to learn features of input image with various levels of abstraction (LeCun et al., 2015).
In remote sensing image processing, DL approaches have been widely used in classification of the land-use and land-cover using low-resolution images (Fu et al., 2017), segmentation and object detection from high-resolution image (Mou and Zhu, 2018a), and extracting and interpreting ambiguous information from single image, such as DSM (Mou and Zhu, 2018b). Moreover, due to difficult challenges of extracting building from remotely sensed data, BFE is one of the most important objectives in geomatics science. Table 1 presents the overview of the methods recently published in applied DL for BFE, based

* Corresponding author

https://doi.org/10.5194/isprs-archives-XLII-4-W18-615-2019 |

on highlighting characteristics with focus on research innovations. Although the related methods are powerful, but it's not still an outstanding performance for BFE, particularly, in robust BFE from big image data.

Algorithm	Highlight of characteristics
ConvNet+ SignedDist (Yuan, 2018)	- Integrating multi-layer information and a unique output representation - Combine signed-distance labels with ConvNet
ABF+SegNet (Masouleh and Shah-Hosseini, 2018)	- Fusion of convolutional layer with adaptive bilateral filter - The minimum bounding rectangle were used for outline regularization
SegNet-Dist-Fused (Yang et al., 2018)	- Combine signed-distance labels with SegNet - No requirement of post-processing
Res-U-Net (Xu et al., 2018)	- Feature extraction based on several residual blocks - A concatenation with the corresponding block from the encoding part is designed
MC–FCN (Wu et al., 2018)	- A bottom-up / top-down multi-constraint fully convolutional network - Basic structure based on fusion of U-Net and three extra multi-scale constraints
GRRNet (Huang et al., 2019)	- Encoding stage based on residual learning network - Improving feature learning with a gated feature labelling unit

Table 1. Overview of recent BFE researches using DL models

In this study, we focus on the key challenges for creating robust BFE model. In this regard, the major contributions of this study to the robust BFE from big multi-sensor data using DL algorithms are as follows:

- For the architectural structure, an efficient deep competition network (DCN) is proposed based on the encoder vector quantization with classified structure and superpixel for BFE from big multi-sensor data using very high spatial resolution multispectral images and LiDAR data.
- In the feature learning step, the very high spatial resolution multispectral image superpixel-based features of big multi-sensor data are combined with LiDAR data (*i.e.* DSM) superpixel-based features in variety of complex roof shapes and textures in large urban areas.

This paper is organized as follows: section (2), provides essential context around the proposed method. The experiment results on big multi-sensor data are demonstrated in section (3). Insight discussion on gains from the study is presented in final section.

2. METHODOLOGY

In this study, two semantic segmentation methods based on DL architectures, including Res-U-Net (Xu et al., 2018), and ABF+SegNet (Masouleh and Shah-Hosseini, 2018), have been used for BFE results comparisons. The models have been selected because of their good performance in BFE from multi-sensor data.

2.1 Res-U-Net

Res-U-Net is a new fully convolutional network for semantic binary segmentation. The architecture uses the modified versions of U-Net and ResNet (Xu et al., 2018). This model features a robust encoder-decoder structure for BFE, because upsampling features in the decoder block and the corresponding max-pooling features in encoder block are made separately and concatenated for other upsampling layers (Huang et al., 2019).

2.2 ABF+SegNet

Khoshboresh Masouleh and Shah-Hosseini (2018) proposed a fusion-based architecture, called ABF+SegNet for building outline enhancement using remote sensing big image data, where the SegNet model (Badrinarayanan et al., 2017) acts as the high-level features generator based on adaptive bilateral filter. ABF+SegNet achieves excellent performance on RGB images for building extraction.

2.3 Deep Competition Network (DCN)

In this paper, we proposed an efficient DCN architecture for BFE from big multi-sensor data. DCN is a deep superpixelwise convolutional encoder-decoder architecture using the encoder vector quantization (Kohonen, 1995) with classified structure for semantic binary segmentation. Our proposed DCN consists of five encoding-decoding blocks with convolutional weights for robust binary representation (feature) learning.
Figure 1 shows the processing chain of the proposed algorithm. In this algorithm, a superpixel segmentation method called Simple Linear Iterative Clustering - SLIC (Achanta et al., 2012), which is one of the mostly used image segmentation algorithms is used to generate basic processing unit.
The competition function to BFE in DCN can be described as follows:

$$n = \arg\min_i \{ \| \frac{1}{2} \sum f(I - O)^2 \| \} \quad (1)$$

where n = winner index
i = binary value
f = sigmoid function
I = input
O = prediction (output)

In Equation 1, a loss function (*e.g.,* sigmoid) is defined monotonically increasing as follows:

$$f(I) = \frac{1}{1 + e^I} \quad (2)$$

where f = sigmoid function
e = Napier's constant (= 2.7182)

In each block, batch normalization function (Laurent et al., 2016), and dropout-based regularization technique have been used to improve performance in training stage with focus on reducing overfitting (Kingma et al., 2015). The activation function on this model is the rectified linear unit (ReLU) function (LeCun et al., 2015) and the proposed competition function appears in the final output layers. Batch normalization, and ReLU functions are computed as follows, respectively:

$$I_{BN} = \frac{I_i - B_m}{B_v} \quad (3)$$

$$ReLU(I_i) = \max(0, I_i) = \begin{cases} I_i, if & I_i \geq 0 \\ 0, if & I_i < 0 \end{cases} \quad (4)$$

where I_i = input
B_m = batch mean
B_v= batch variance

https://doi.org/10.5194/isprs-archives-XLII-4-W18-615-2019 | © Authors 2019. CC BY 4.0 License.

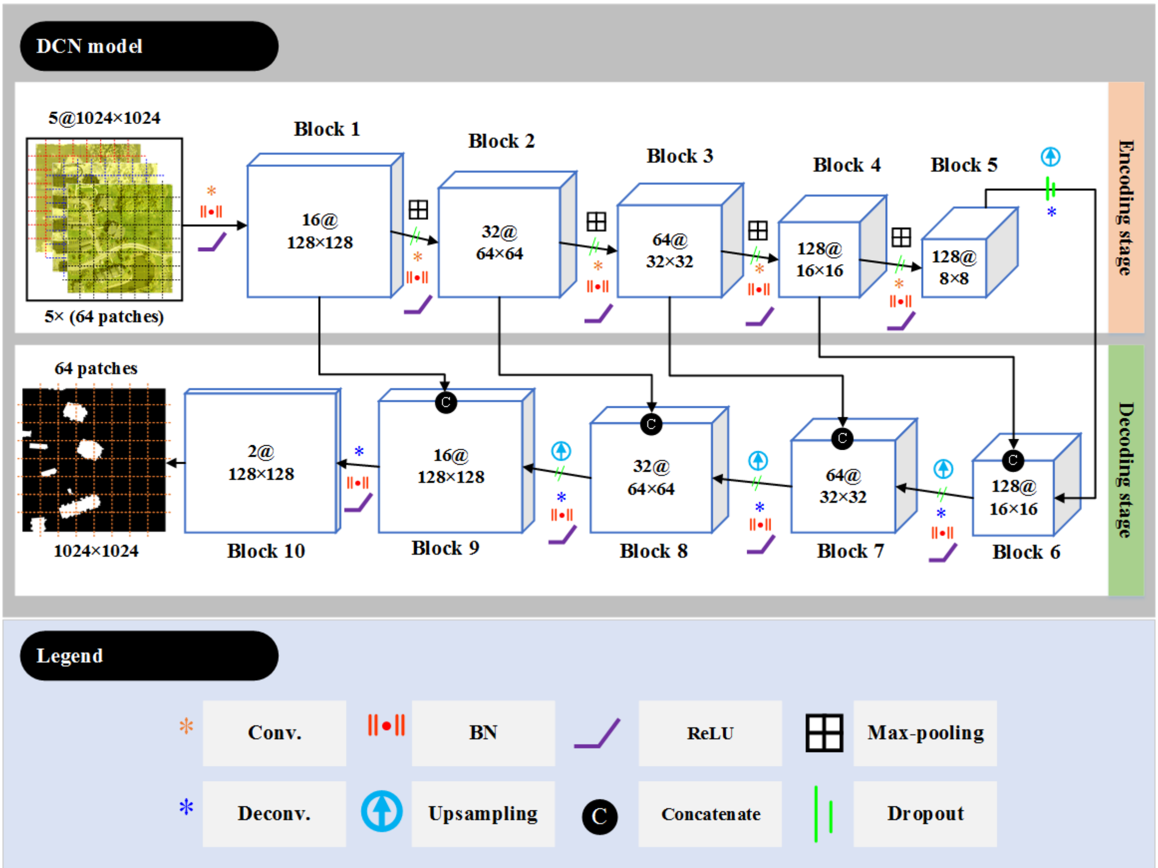

Figure 1. Architecture of the proposed DCN

3. EXPERIMENTS AND RESULTS

3.1 Big Multi-Sensor Data

The experiments on big multi-sensor data consisting of:
(1) Very high spatial resolution multispectral images with the four spectral bands (*i.e.* red, green, blue, and NIR) and a ground sampling distance (GSD) of 0.5 foot obtained from the State of Indiana in the US with multiple building scenes,
(2) DSM generated from LiDAR point clouds with a GSD of 0.5 foot by the Indiana Office of Information Technology,
(3) The normalized difference vegetation index (NDVI) generated from red and NIR bands, and
(4) OpenStreetMap shapefiles used as ground truth map to validate results.
Figure 2 displays the research site of the big multi-sensor data. The research site covers about 950 km^2 from the Indianapolis city in the US. Moreover, big multi-sensor data consist of RGB, DSM, NDVI and building footprint map without major misalignment in the projection of the North American Datum (NAD). Table 2 shows the splitting statistics of the big multi-sensor data.

	Tiles
Training	256
Validation	40
Testing	3
Total	299

Table 2. Dataset splitting statistics

For more information about the coordinate system of data, please see (http://gis.iu.edu/datasetInfo/statewide/in_2011.php). In Figure 2, the orange boundary indicates the study area.

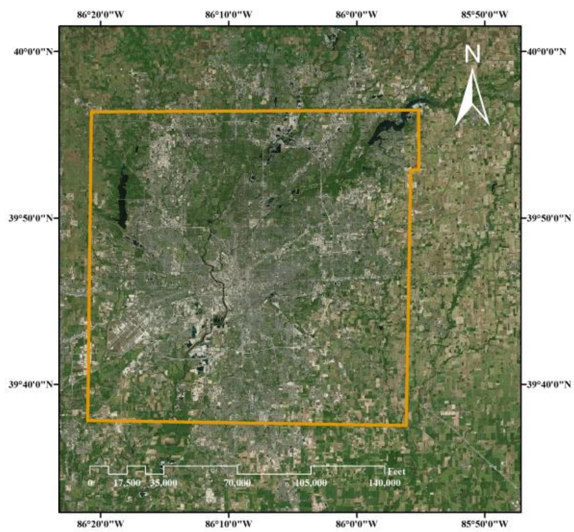

Figure 2. Location of study area

3.2 DCN Implementation

DCN was implemented using Keras (Chollet, 2018) on the free cloud Tesla K80 GPU and 12GB of RAM in google Colarboratory (Colab), but there is not enough RAM for big data storage. For this reason, we integrated google drive (free cloud storage) with Colab for memory enhancing. Keras is a powerful and open source DL framework written in Python. DCN was trained and validated with adaptive moment estimation (ADAM) optimizer using the default parameters (Kingma and Ba, 2014) and with a batch size of 64 for 250 epochs for BFE. Moreover, 299 tiles from Indiana each of the size 1024×1024 pixels are processed using a 128×128 pixels sliding window in order to reduce memory consumption.

3.3 Building Footprint Extraction

As shown in Figures 3-5, three representative samples are selected from the test area with densely distributed buildings for assessing the performance of the baseline models in comparison with our proposed model. The evaluation metrics of overall accuracy (OA) and intersection over union (IoU) are used to evaluate the performance of the models based on the values of true positive (white), false positive (red), false negative (blue), and true negative (black). The OA and IoU are the most used evaluation measures for BFE in many previous studies (Khoshboresh Masouleh and Shah-Hosseini, 2019b; Maggiori et al., 2017; Masouleh and Sadeghian, 2019; Volpi and Tuia, 2018; Xu et al., 2018; Yang et al., 2018). The OA and IoU are defined as:

$$OA = \frac{TP + TN}{TP + FP + FN + TN} \tag{2}$$

$$IoU = \frac{TP}{TP + FN + FP + e} \tag{3}$$

where TP = true positive
FP = false positive
FN = false negative
TN = true negative
$e = 10^{-15}$ (to avoid division by zero)

In Table 3 we compared DCN's performance with Res-U-Net and ABF+SegNet models for three test samples. Bold fonts denote the best results and the underlined fonts denote the second best results. The DCN model's OA reaches 94%, 99% and 99% while the IoU reaches the 93%, 99% and 98% for each test samples, respectively. The results demonstrated that DCN model is a suitable solution to robust BFE from big multi-sensor data, especially on the different types of roofs.

	Model	Res-U-Net	ABF+SegNet	DCN
Case-1	OA	91%	91%	**94**%
	IoU	89%	90%	**93**%
Case-2	OA	**99**%	**99**%	**99**%
	IoU	98%	98%	**99**%
Case-3	OA	96%	97%	**99**%
	IoU	91%	96%	**98**%

Table 3. Quantitative evaluation on the test area

3.4 Computational Cost Analysis

Computational cost is an important factor in big data processing, particularly for real world applications such as BFE, because hardware limitations (*e.g.* memory consumption, processing system, etc.) in the real world. Therefore, selection of appropriate computational space for big data processing with optimal computational cost is necessity. Cloud computing is an internet-based space for reducing the computational cost in DL experiments. In this paper, we used Colab (cloud computing platform) for training efficiency and computational cost analysis. For this purpose, we trained all models (Res-U-Net, ABF+SegNet, and DCN) based on big multi-sensor data in Colab. The evaluation metric for computational cost analysis is defined as (Justus et al., 2018):

$$CC = \frac{NE \times TT}{60} \tag{4}$$

where CC = computational cost (in min)
NE = number of epochs
TT = training time per epoch (in sec)

The computational cost results are presented in Table 4. Bold font denotes the best result and the underlined font denotes the second best result.

Model	Res-U-Net	ABF+SegNet	DCN
CC (min)	535	590	**461**

Table 4. Computational cost results on the test area

4. CONCLUSION

In this paper, we focus on tackling the regularization of building outlines problem in very high spatial resolution remote sensing images by proposing a model based on DL and superpixel segmentation called DCN. Most important feature of this model is exploitation of vector quantization theory and convolutional layers in creating a DL network for BFE. In order to train the proposed model, INDIANA dataset was used. Results of applying the proposed method on three test samples indicate improvement in training speed and increase in accuracy and validity of BFE from big multi-sensor data that contains very high spatial resolution multispectral images and LiDAR data. DCN model, automatically extracts the buildings from input data based on the encoder vector quantization framework (supervised learning). In order to evaluate the results, we compared our model with two powerful DL models. Based on the statistical results shown in Table 3, the accuracy is somewhat better, but the IoU (a scale invariant metric) is obviously improved. Future studies can be conducted to increase performance of DCN through optimizing network depth and improving superpixel segmentation methods to reinforce BFE.

ACKNOWLEDGEMENTS

The INDIANA dataset were obtained freely from the Indiana Office of Information Technology. We thank the Indiana Office of Information Technology. Finally, we would like to thank the anonymous referee for his/her helpful comments.

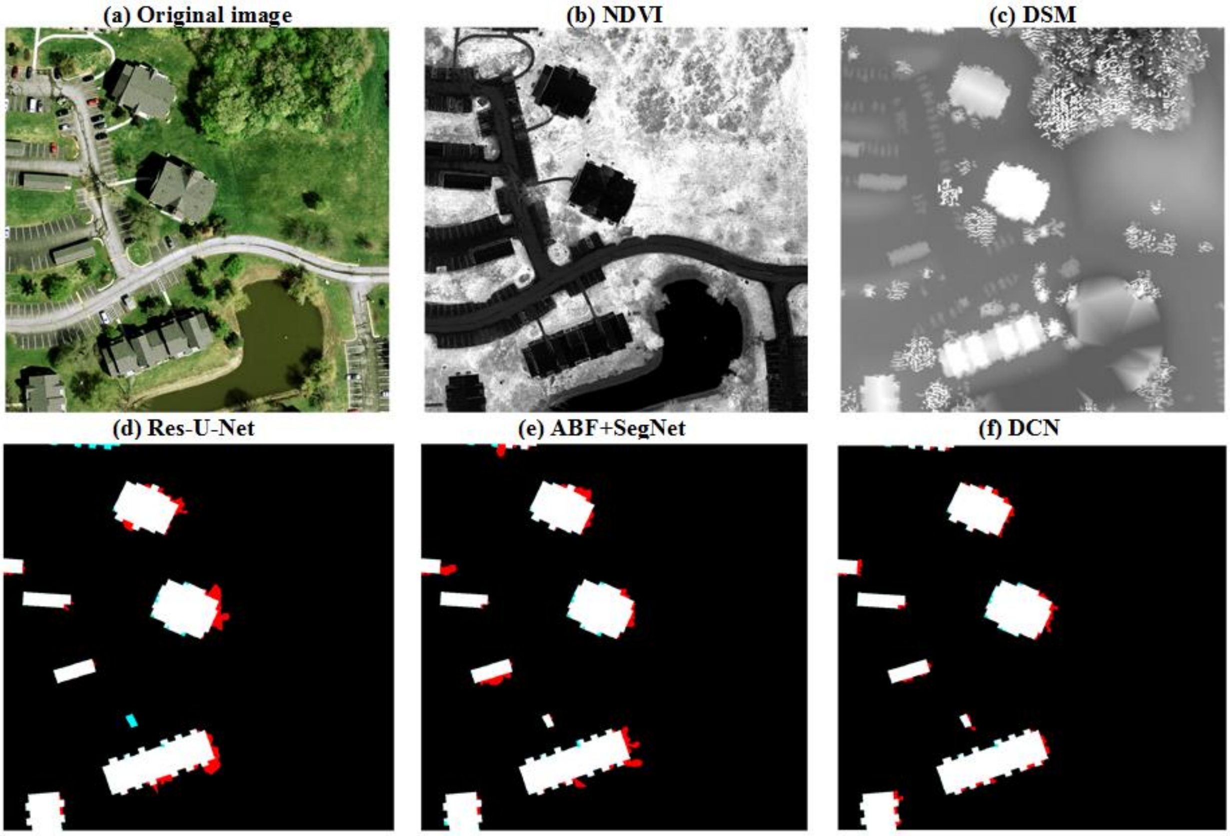

Figure 3. Comparison of BFE results using Res-U-Net, ABF+SegNet and DCN with ground truth in Case-1

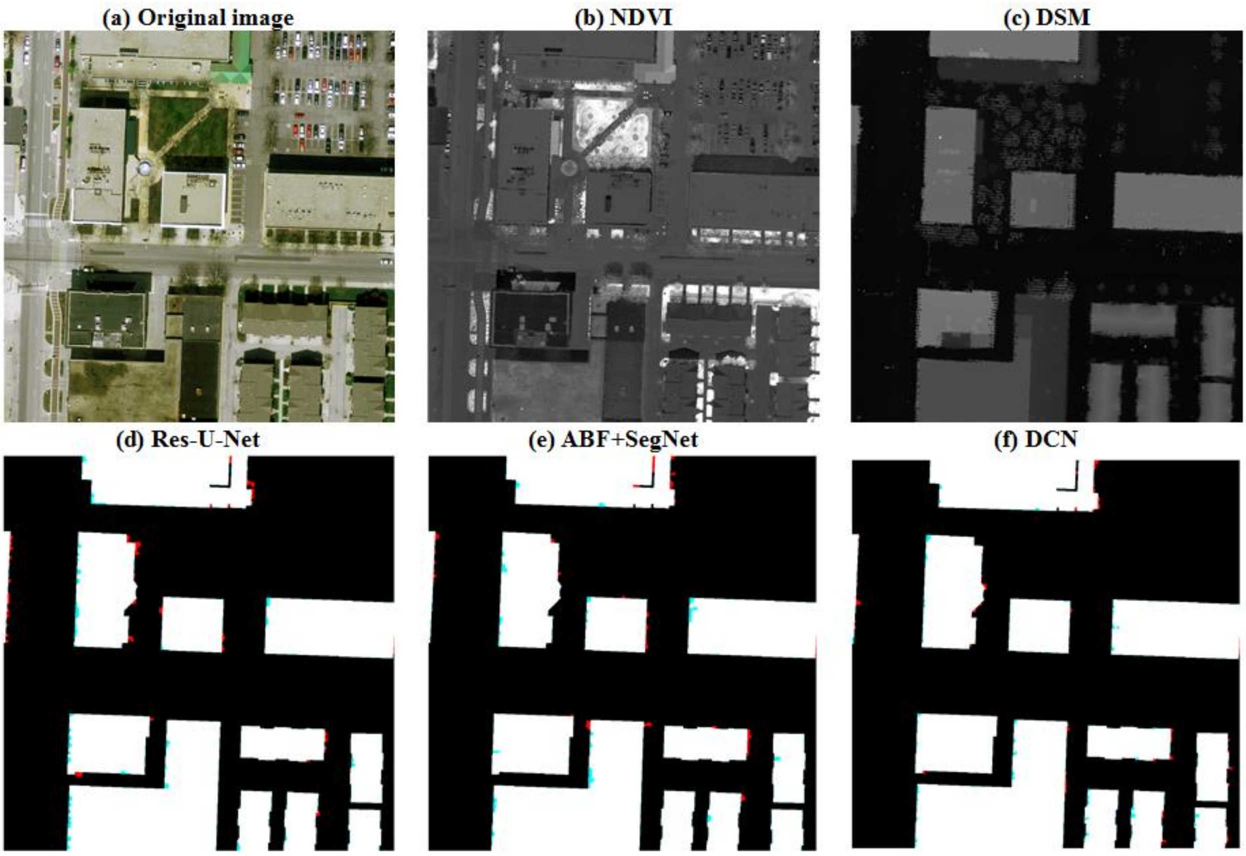

Figure 4. Comparison of BFE results using Res-U-Net, ABF+SegNet and DCN with ground truth in Case-2

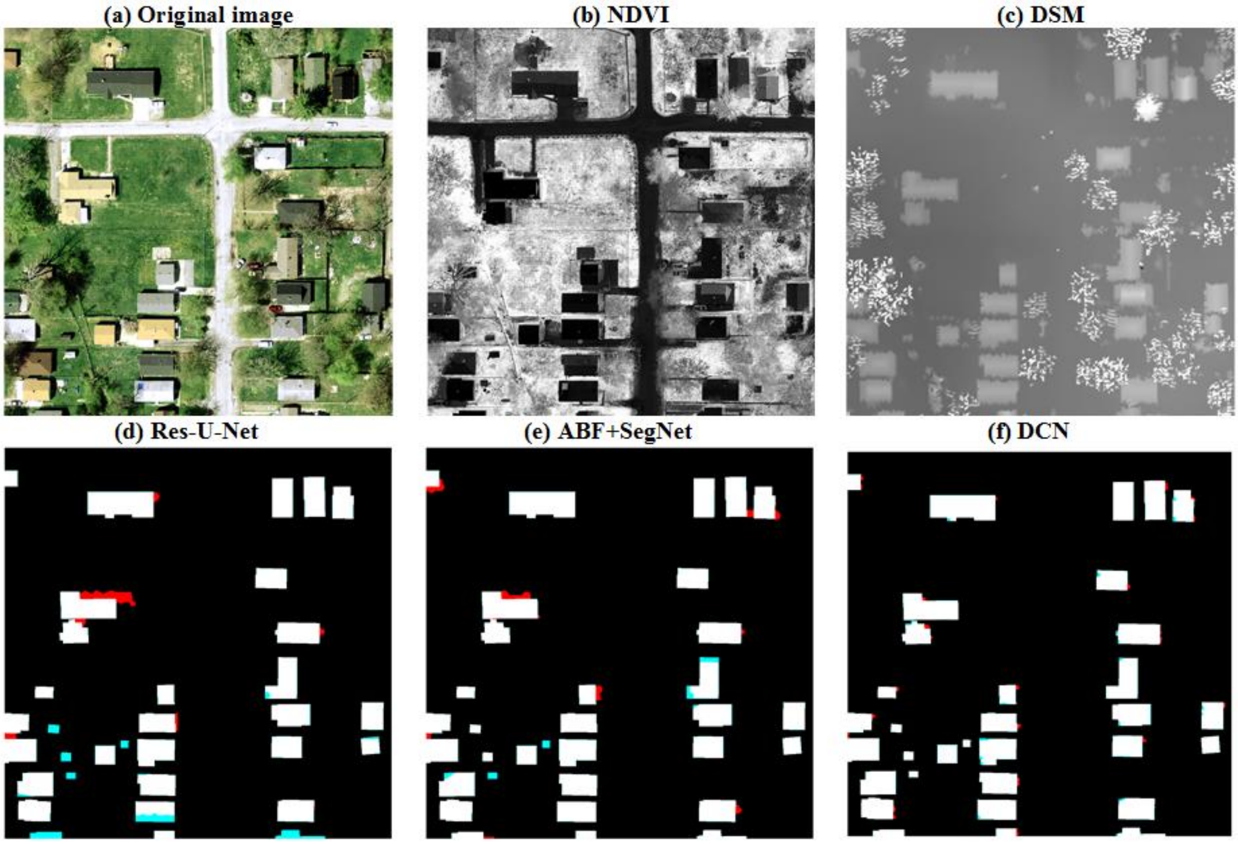

Figure 5. Comparison of BFE results using Res-U-Net, ABF+SegNet and DCN with ground truth in Case-3

REFERENCES

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S., 2012. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* 34, 2274–2281. https://doi.org/10.1109/TPAMI.2012.120

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 2481–2495. http://dx.doi.org/10.1109/tpami.2016.2644615

Bi, Q., Qin, K., Zhang, H., Zhang, Y., Li, Z., Xu, K., 2019. A Multi-Scale Filtering Building Index for Building Extraction in Very High-Resolution Satellite Imagery. *Remote Sens.* 11, 482. https://doi.org/10.3390/rs11050482

Cai, Z., Ma, H., Zhang, L., 2019. A Building Detection Method Based on Semi-Suppressed Fuzzy C-Means and Restricted Region Growing Using Airborne LiDAR. *Remote Sens.* 11, 848. https://doi.org/10.3390/rs11070848

Chollet, F., 2018. Deep learning with Python. Manning Publications Co, Shelter Island, New York.

Fu, G., Liu, C., Zhou, R., Sun, T., Zhang, Q., 2017. Classification for High Resolution Remote Sensing Imagery Using a Fully Convolutional Network. *Remote Sens.* 9, 498. https://doi.org/10.3390/rs9050498

Huang, J., Zhang, X., Xin, Q., Sun, Y., Zhang, P., 2019. Automatic building extraction from high-resolution aerial images and LiDAR data using gated residual refinement network. ISPRS J. Photogramm. *Remote Sens.* 151, 91–105. https://doi.org/10.1016/j.isprsjprs.2019.02.019

Huang, Y., Zhuo, L., Tao, H., Shi, Q., Liu, K., 2017. A Novel Building Type Classification Scheme Based on Integrated LiDAR and High-Resolution Images. *Remote Sens.* 9, 679. https://doi.org/10.3390/rs9070679

Jung, J., Sohn, G., 2019. A line-based progressive refinement of 3D rooftop models using airborne LiDAR data with single view imagery. *ISPRS J. Photogramm. Remote Sens.* 149, 157–175. https://doi.org/10.1016/j.isprsjprs.2019.01.003

Justus, D., Brennan, J., Bonner, S., McGough, A.S., 2018. Predicting the Computational Cost of Deep Learning Models, in: 2018 IEEE International Conference on Big Data (Big Data). Presented at the 2018 IEEE International Conference on Big Data (Big Data), pp. 3873–3882. https://doi.org/10.1109/BigData.2018.8622396

Khoshboresh Masouleh, M., Shah-Hosseini, R., 2019a. Development and evaluation of a deep learning model for real-time ground vehicle semantic segmentation from UAV-based thermal infrared imagery. *ISPRS J. Photogramm. Remote Sens.* 155, 172–186. https://doi.org/10.1016/j.isprsjprs.2019.07.009

Khoshboresh Masouleh, M., Shah-Hosseini, R., 2019b. A hybrid deep learning–based model for automatic car extraction from high-resolution airborne imagery. *Appl. Geomat.* https://doi.org/10.1007/s12518-019-00285-4

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. *3rd Int. Conf. Learn. Represent.* ICLR.

Kingma, D.P., Salimans, T., Welling, M., 2015. Variational Dropout and the Local Reparameterization Trick, in: Cortes, C., Lawrence, N.D., Lee, D.D., Sugiyama, M., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 28. Curran Associates, Inc., pp. 2575–2583.

Kohonen, T., 1995. Learning Vector Quantization, in: Kohonen, T. (Ed.), Self-Organizing Maps, Springer Series in Information Sciences. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 175–189. https://doi.org/10.1007/978-3-642-97610-0_6

Laurent, C., Pereyra, G., Brakel, P., Zhang, Y., Bengio, Y., 2016. Batch normalized recurrent neural networks, in: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Presented at the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, Shanghai, pp. 2657–2661. https://doi.org/10.1109/ICASSP.2016.7472159

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521, 436–444. https://doi.org/10.1038/nature14539

Li, Y., Wu, H., An, R., Xu, H., He, Q., Xu, J., 2013. An improved building boundary extraction algorithm based on fusion of optical imagery and LIDAR data. *Optik* 124, 5357–5362. https://doi.org/10.1016/j.ijleo.2013.03.045

Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2017. Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification. *IEEE Trans. Geosci. Remote Sens.* 55, 645–657. https://doi.org/10.1109/TGRS.2016.2612821

Masouleh, M.K., Sadeghian, S., 2019. Deep learning-based method for reconstructing three-dimensional building cadastre models from aerial images. *J. Appl. Remote Sens.* 13, 024508. https://doi.org/10.1117/1.JRS.13.024508

Masouleh, M.K., Shah-Hosseini, R., 2018. Fusion of deep learning with adaptive bilateral filter for building outline extraction from remote sensing imagery. *J. Appl. Remote Sens.* 12, 1. https://doi.org/10.1117/1.JRS.12.046018

Mou, L., Zhu, X.X., 2018a. Vehicle Instance Segmentation from Aerial Image and Video Using a Multitask Learning Residual Fully Convolutional Network. *IEEE Trans. Geosci. Remote Sens.* 56, 6699–6711. http://dx.doi.org/10.1109/tgrs.2018.2841808

Mou, L., Zhu, X.X., 2018b. IM2HEIGHT: Height Estimation from Single Monocular Imagery via Fully Residual Convolutional-Deconvolutional Network. *ArXiv180210249 Cs.*

Philip Chen, C.L., Zhang, C.-Y., 2014. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Inf. Sci.* 275, 314–347. https://doi.org/10.1016/j.ins.2014.01.015

Rottensteiner, F., Trinder, J., Clode, S., Kubik, K., 2007. Building detection by fusion of airborne laser scanner data and multi-spectral images: Performance evaluation and sensitivity analysis. *Remote Sens.* 15.

Rottensteiner, F., Trinder, J., Clode, S., Kubik, K., 2003. Building detection using LIDAR data and multispectral images, in: In: *Proceedings of DICTA.* pp. 673–682.

Samuel R., D.J., E, F., Manogaran, G., G.n, V., T, T., S, J., A, A., 2019. Real time violence detection framework for football stadium comprising of big data analysis and deep learning through bidirectional LSTM. *Comput. Netw.* 151, 191–200. https://doi.org/10.1016/j.comnet.2019.01.028

Sohn, G., Dowman, I., 2007. Data fusion of high-resolution satellite imagery and LiDAR data for automatic building extraction. *ISPRS J. Photogramm. Remote Sens.* 62, 43–63. https://doi.org/10.1016/j.isprsjprs.2007.01.001

Volpi, M., Tuia, D., 2018. Deep multi-task learning for a geographically-regularized semantic segmentation of aerial images. *ISPRS J. Photogramm. Remote Sens.* 144, 48–60. https://doi.org/10.1016/j.isprsjprs.2018.06.007

Wu, G., Shao, X., Guo, Z., Chen, Q., Yuan, W., Shi, X., Xu, Y., Shibasaki, R., 2018. Automatic Building Segmentation of Aerial Imagery Using Multi-Constraint Fully Convolutional Networks. *Remote Sens.* 10, 407. https://doi.org/10.3390/rs10030407

Xu, Y., Wu, L., Xie, Z., Chen, Z., 2018. Building Extraction in Very High Resolution Remote Sensing Imagery Using Deep Learning and Guided Filters. *Remote Sens.* 10, 144. https://doi.org/10.3390/rs10010144

Yang, C., Yu, M., Hu, F., Jiang, Y., Li, Y., 2017. Utilizing Cloud Computing to address big geospatial data challenges. *Comput. Environ. Urban Syst.*, Geospatial Cloud Computing and Big Data 61, 120–128. https://doi.org/10.1016/j.compenvurbsys.2016.10.010

Yang, H.L., Yuan, J., Lunga, D., Laverdiere, M., Rose, A., Bhaduri, B., 2018. Building Extraction at Scale Using Convolutional Neural Network: Mapping of the United States. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 11, 2600–2614. https://doi.org/10.1109/JSTARS.2018.2835377

Yuan, J., 2018. Learning Building Extraction in Aerial Scenes with Convolutional Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 40, 2793–2798. https://doi.org/10.1109/TPAMI.2017.2750680