A MULTI-MODEL FUSION OF CONVOLUTION NEURAL NETWORK AND RANDOM FOREST FOR DETECTING SETTLEMENTS WITHOUT ELECTRICITY

Yu Xia, Qi Huang and Hongyan Zhang*

The State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, P.R. China
*Corresponding author: zhanghongyan@whu.edu.cn

IGARSS 2021 - 2021 IEEE International Geoscience and Remote Sensing Symposium | 978-1-6654-0369-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/IGARSS47720.2021.9553087

ABSTRACT

In this paper, a multi-model fusion framework is proposed for automatic detection of settlements without electricity (DSE) based on the multimodal and multitemporal remote sensing data. To settle the problems of data noise and data redundancy, the data preprocessing step, which consists of band selection, cloud removal, grayscale stretch and data augmentation, is firstly applied. Two models of single-task and dual-task are further constructed for DSE. The single-task model builds a global context convolutional neural network (GC-CNN) for the detection of settlements without electricity and the dual-task model employs the GC-CNN for settlement detection and the random forest classifier for electricity detection. Moreover, a model fusion principle and a post-processing method is designed to integrate and improve the results above, thus producing the final segmentation result. Verified through the competition website, the proposed method achieved a F1-score of 0.8806, ranking second in the first track of 2021 IEEE GRSS Data Fusion Contest.

***Index Terms*-**Data fusion, detection of settlements without electricity, convolution neural network, random forest

1. INTRODUCTION

Since the second industrial revolution, electricity has brought light and heat to mankind and has become an inseparable part of human life [1]. However, on a global scale, there are still many people living in areas without electricity, surrounded by darkness and eroded by coldness. Hence, accurately locating the areas without electricity is the key to ending the darkness and expelling the coldness.

In recent years, remote sensing has become an effective technique in earth observations, which plays a vital role in many geographic applications, such as land use classification [2], disaster monitoring [3] and urban analysis [4]. Compared with manual survey, remote sensing technology is a more economical means of manpower and material resources for the automatic detection of settlements without electricity. The multimodal and multitemporal remote sensing data consists of Sentinel-1, Sentinel-2, Landsat-8 and VNP46A1 product. Among the given data, optical and radar images are helpful to extract residential buildings [5]. Moreover, the night time data is closely related to light intensity, which is a valuable indicator for power supply [6]. However, some data mentioned above are severely polluted by the noise from cloud and moonlight. Our goal is to accurately extract settlements without electricity at 500m resolution on the basis of these data sampled to 10m resolution, which are provided by the Track 1 of IEEE-GRSS 2021 Data Fusion Contest [7]. In essence, the detection of settlements without electricity (DSE) contains two main difficulties: 1) data diversity: there exist multimodal and multitemporal data, from which the valuable and useful information needs to be screened out; 2) task complexity: the composite target task consists of two challenge elements: the detection of buildings and the detection of electricity.

In order to detect settlements without electricity, we propose a multi-model fusion framework to incorporate a convolution neural network and a random forest algorithm. An indispensable data preprocessing step is firstly adopted based on multimodal and multitemporal remote sensing data to remove noises and retain valid data, thus obtaining a considerable and feasible dataset. On the basis of the dataset preprocessed, the single-task model and the dual-task model are established to detect settlements without electricity. In the single-task model, all data are directly input into the global context convolution neural network to segment settlements without electricity. In the dual-task model, optical data is utilized to detect buildings with the neural network, while the night time data is used for electricity detection with random forest method. After that, a model fusion principle and a post-processing method is finally applied to acquire the output result.

2. METHODOLOGY

2.1. The Proposed Framework

The purpose of our work is to extract settlements without electricity by segmenting binary labels with a resolution of 500m from multimodal and multitemporal remote sensing data with a resolution of 10m. In order to effectively utilize the available data, a multi-model fusion framework is designed. As shown in Fig. 1, there exist three steps: data pre-processing, multi-model fusion and post-processing, which will be introduced in the following sections.

2.2. Data pre-processing

The step of data preprocessing is of vital importance in the whole framework, which aims at removing data noise and preserving valid data. It contains four parts: data and band selection, cloud removal, grayscale stretch and data augmentation. Due to the serious noises in Sentinel-1 and Landsat-8, only Sentinel-2 and VNP46A1 products are

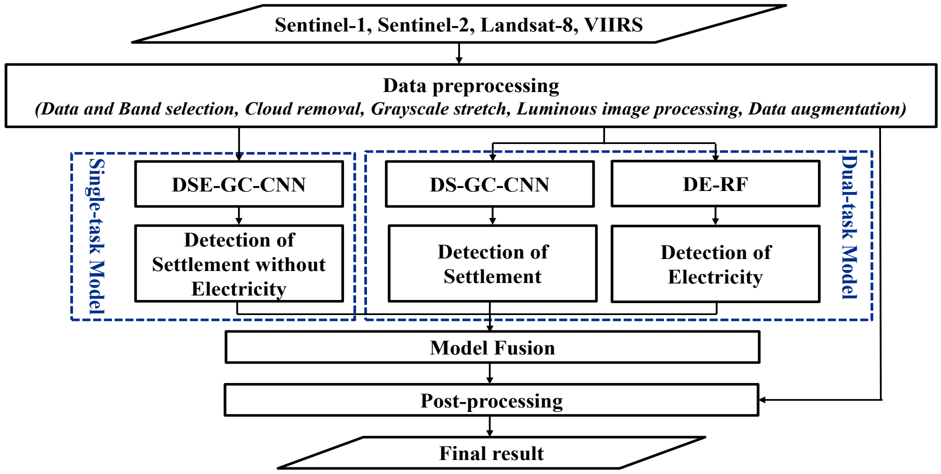


Fig. 1. Flow chat of the proposed framework.

preserved. According to some commonly used building indices (UI, NDBI) [8], six bands of Red, Green, Blue, NIR, SWIR-1 and SWIR-2 of Sentinel-2 images are selected for the following steps. For optical images contaminated by clouds, it is necessary to composite cloud-free images from multi-temporal data, where the median filter method is adopted [9]. For VNP46A1 product data, in order to avoid the influence of light variations in different times, multi-temporal night time data are all accumulated to generate new images. In addition, the median and maximum values of multitemporal night time data are composited to minimize the effects of the data noise. Furthermore, the 2% linear grayscale stretch algorithm is then implemented on all selected bands to eliminate extreme noisy points and enhance the image contrast. In the end, we apply data augmentation methods such as rotation, flipping to generate a larger dataset and provide the foundation of robust network training [10].

2.3. Multi-model fusion

The task of DSE can be divided into two parts: building detection and electricity detection. In view of this, two models named the single-task model and the dual-task model are then established and further integrated, which will be described in detail as follows.

2.3.1 Single-task model

To detect settlements without electricity, an end-to-end global context convolutional neural network (DSE-GC-CNN) is proposed. As is shown in Fig. 2, the network accepts 800*800 Sentinel-2 images with six bands and night time images with three bands, outputting the 16*16 segmentation maps. In order to extract the global context information in the imagery and capture long distance dependencies of deep neural networks, a global context module is adopted in our detection network [11], which is depicted in Fig.3. It includes three parts: global attention pooling for context modeling, bottleneck transformation for capturing channel dependencies and broadcast element-wise addition for feature fusion.

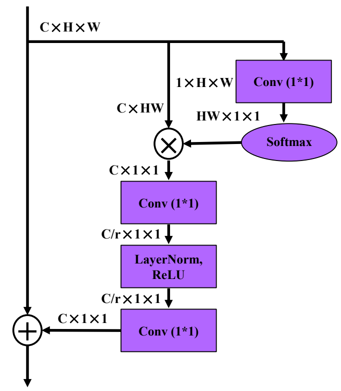


Fig. 3. An overview of global context block.

Furthermore, in view of the categorical imbalance in the dataset, we adopt the focal loss function in network training [12], which is described as follows:

$$FL(p,y) = -\alpha y(1-p)^{\gamma}\log(p) - (1-\alpha)(1-y)p^{\gamma}\log(1-p) \qquad (1)$$

where y is the true label, and p is the predicted result of the network, α denotes weight parameter for the sample categories, which is set as 0.6 in our experiment, and γ represents the weight parameter for the difficulty of sample classification, which is set as 2.

In the testing phase, a creative voting strategy is applied to the detection results. Among the eight results of images generated by different data augmentation methods, if more than two prediction results show that there exist settlements without electricity, then this pixel will be labeled as “DSE”.

2.3.2 Dual-task model

As depicted in Fig. 4, the same convolutional neural network with global context block is applied to detect settlements (DS-GC-CNN), while random forest classifier [13] is used for detecting areas with electricity (DE-RF). The inputs of DS-GC-CNN are optical images, whereas the inputs of DE-RF are VNP46A1 products with 9 phases and median, minimum, maximum, summation feature images. The outputs of the two methods are both binary graphs. By

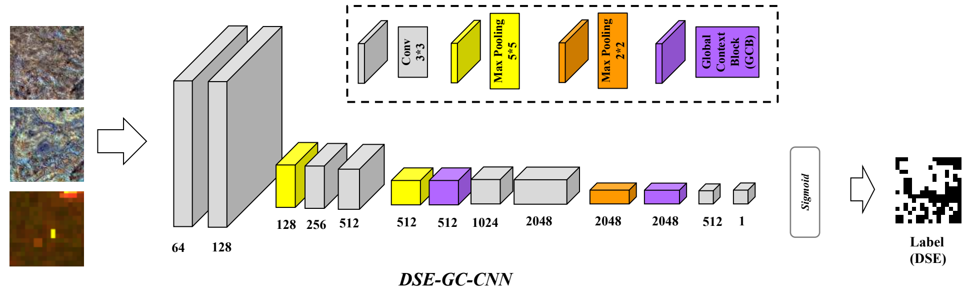


Fig. 2. An overview of the DSE-GC-CNN.

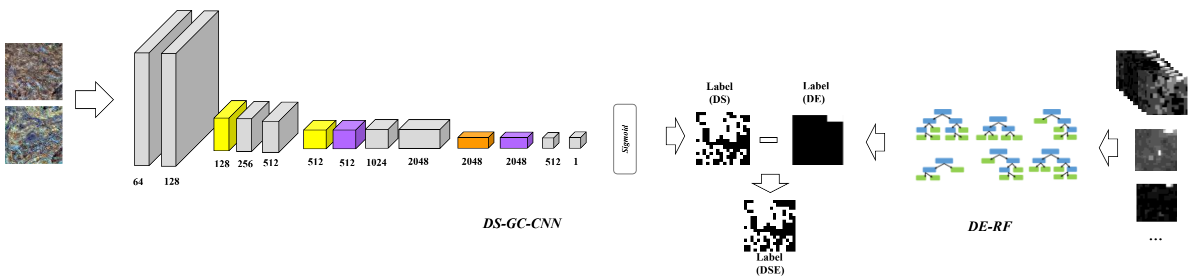


Fig. 4. A workflow of dual-task model with DS-GC-CNN and DE-RF.

subtracting these two results, the final residential area without electricity can be obtained. It is worth mentioning that the same voting strategy in DSE-GC-CNN is adopted in DS-GC-CNN.

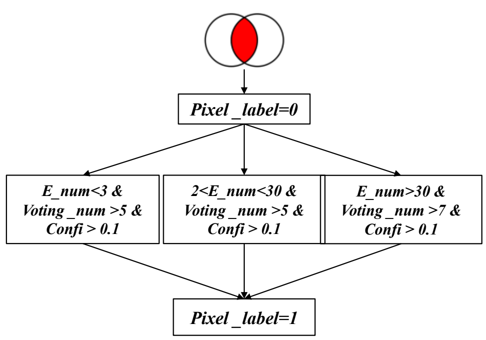


Fig. 5. The principle of model fusion.

2.3.3 Model fusion

Although the single-task model can produce the final result concisely and conveniently, the network unevenly uses night time data and optical data, leading to the integration insufficient. On the other hand, although the dual-task model makes full use of data characteristic, there is an accumulation of errors in the process of multiple tasks. All in all, compared with the single-task model, the dual-task model pays more attention to the detection of buildings, which can achieve higher accuracy of building detection, assisting the subsequent integration of results. As is depicted in Fig. 5, *E_num* is the number of powered pixels in each tile, while *Voting_num* represents the number of positive results in the voting strategy (DS-GC-CNN). *Confi* is the confidence value of each pixel in DSE-GC-CNN. The first step of model fusion is taking the intersection of two model results, which produces high-confidence labels, yet there exist many omission errors. On the basis of these results, the confidence maps of the networks and the results of DE-RF are further processed with the threshold segmentation method, so that some high-confidence residential areas with no electricity can be added to the intersection results.

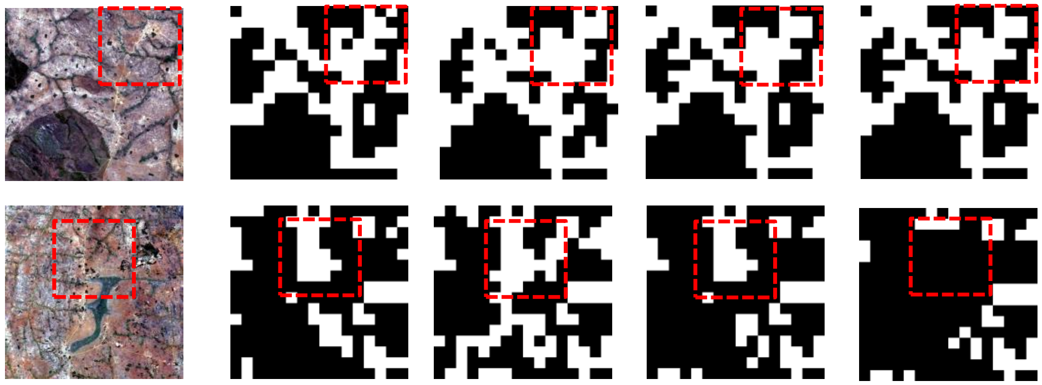
2.4. Post-processing

After the above steps, the multi-model fusion result of settlements without electricity still exist some noisy points, which can be further refined with threshold segmentation and morphological processing. For each tile with *E_num* greater than 30, pixels with max value between 18 and 30 and pixels with sum value greater than 30 are considered as powered pixels, which will be further excluded. In addition, we assume that when the number of residential areas without electricity in a tile is less than 3, the entire tile has no residential areas.

3. EXPERIMENTS

The performance of our proposed framework is evaluated on the test data of 19 tiles, each of which has the size of 800*800. The segmentation maps are 16*16, which are assessed via F1 score for comparison, since it can comprehensively evaluate and balance the precision and recall rate.

Visualized results selected are shown in Fig. 6, where different results of intermediate processes and methods can be compared. In addition, Table 1 lists the quantitative results for the different models in detection of settlements



RGB **Single-task Model** **Dual-task Model** **Model Fusion** **Post-processing**

Fig. 6. Detection results of different methods and periods.

without electricity. By comparing visual results and quantitative results, several conclusions can be drawn. Firstly, the precision is lower than the recall in the results of the single-task model and the dual-task model, indicating that the results of two models have the problem of over-detection. Besides, the quantitative results of the single-task model and the dual-task model are close, yet the visual detection results are inconsistent. Finally, multi-model fusion operation accounts for a 1% improvement in terms of F1 score, which demonstrates that the integration of different models works.

Table 1. Performance on the test data in the contest

Method	F1	Recall	Precision
Single-task Model	0.8623	0.8857	0.8401
Dual-task Model	0.8581	**0.9014**	0.8188
Model Fusion	0.8744	0.8717	0.8771
Post-processing	**0.8806**	0.8711	**0.8896**

4. CONCLUSIONS

In this paper, we propose a novel framework for detection of settlements without electricity based on multi-model fusion. An enhanced training dataset is generated by cloud removal, band selection, grayscale stretch and data augmentation. An end-to-end global context convolutional neural network and random forest classifier are further combined to form two models, integrated for predicting potential residential areas without electricity. During the post-processing step, the results from different models are merged to improve the prediction accuracy. The experiment results have illustrated that our multi-model fusion framework effectively improves the final prediction results.

5. ACKNOWLEDGMENT

The authors would like to thank the IEEE GRSS Image Analysis and Data Fusion Technical Committee, Hewlett Packard Enterprise, SolarAid, and Data Science Experts for organizing the Data Fusion Contest. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB0504500, in part by the National Natural Science Foundation of China under Grant 61871298, and in part by the Wuhan Application Foundation Frontier Project No. 2020010601012184.

6. REFERENCES

[1] C. D. Elvidge, M. L. Imhoff, K. E. Baugh, V. R. Hobson, I. Nelson, J. Safran, J. B. Dietz, and B. T. Tuttle, "Night-time lights of the world: 1994–1995," ISPRS J. Photogramm. Remote Sens., vol. 56, no. 2, pp. 81-99, 2001.

[2] M. Weiss, F. Jacob, and G. Duveiller, "Remote sensing for agricultural applications: A meta-review," Remote Sens. Environ., vol. 236, p. 111402, 2020.

[3] O. M. Bello and Y. A. Aina, "Satellite remote sensing as a tool in disaster management and sustainable development: towards a synergistic approach," Procedia-Social and Behavioral Sciences, vol. 120, pp. 365-373, 2014.

[4] P. A. Longley, M. J. Barnsley, and J.-P. Donnay, "Remote sensing and urban analysis: a research agenda," Remote Sensing and Urban Analysis: GISDATA 9, p. 245, 2000.

[5] M. Pesaresi, C. Corbane, A. Julea, A. J. Florczyk, V. Syrris, and P. Soille, "Assessment of the added-value of Sentinel-2 for detecting built-up areas," Remote Sens., vol. 8, no. 4, p. 299, 2016.

[6] C. N. Doll, J.-P. Muller, and J. G. Morley, "Mapping regional economic activity from night-time light satellite imagery," Ecol. Econ., vol. 57, no. 1, pp. 75-92, 2006.

[7] N. a. G. Yokoya, Pedram and Hansch, Ronny and Prieur, Colin and Malha, Hana and Chanussot, Jocelyn and Robinson, Caleb and Malkin, Kolya and Jojic, Nebojsa, "2021 Data Fusion Contest: Geospatial Artificial Intelligence for Social Good [Technical Committees]," IEEE Geosci. Remote Sens. Mag., vol. 9, no. 1, pp. 287-C3, 2021.

[8] H. Xu, "A new index for delineating built-up land features in satellite imagery," J. Remote Sens., vol. 29, no. 14, pp. 4269–4276, 2008.

[9] G. Mateo-García, L. Gómez-Chova, J. Amorós-López, J. Muñoz-Marí, and G. Camps-Valls, "Multitemporal Cloud Masking in the Google Earth Engine," Remote Sens., vol. 10, no. 7, p. 1079, 2018.

[10] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," Journal of Big Data, vol. 6, no. 1, pp. 1-48, 2019.

[11] Y. Cao, J. Xu, S. Lin, F. Wei, and H. Hu, "Gcnet: Non-local networks meet squeeze-excitation networks and beyond," in Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops 2019, pp. 0-0.

[12] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in Proceedings of the IEEE international conference on computer vision 2017, pp. 2980-2988.

[13] M. Pal, "Random forest classifier for remote sensing classification," Int. J. Remote Sens., vol. 26, no. 1, pp. 217-222, 2005.