IGARSS 2021 - 2021 IEEE International Geoscience and Remote Sensing Symposium | 978-1-6654-0369-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/IGARSS47720.2021.9553860

MULTISOURCE DATA FUSION FOR THE DETECTION OF SETTLEMENTS WITHOUT ELECTRICITY

Yanbiao Ma,Yuxin Li,Kexin Feng,Xueli Geng,Licheng Jiao,Fang Liu,Yuting Yang

Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education,
School of Artificial Intelligence,Xidian University,Xi'an,Shaanxi Province 710071,China

ABSTRACT

The international charity SolarAid aims to provide access to lights in areas without electricity, and it is a challenge to accurately and efficiently transmit the lights to the areas in need. Multisource, multitemporal, and multimodal remote sensing images can provide rich information about the target area, so using multisource remote sensing images for accurate detection of human settlements without electricity is a feasible solution. In this paper two separate detection tasks are formulated: building two attention SENet for settlements detection and light detection using the Sentinel-2 dataset and the Suomi Visible Infrared Imaging Radiometer Suite (VIIRS) night time dataset, respectively. In addition, we study a new outlier removal method based on the pixel distribution characteristics of the VIIRS dataset for data pre-processing, and propose a post-processing method based on region continuity for further correction of the results. Experiments show that our method can maximize the use of multisource data information and rank first in the detection of settlements without electricity challenge track (Track DSE) of the 2021 IEEE GRSS Data Fusion Contest.

***Index Terms*—** multisource data fusion; attention mechanism; outlier removal; Sentinel-2; VIIRS;

1. INTRODUCTION

Electricity has become one of the essential sources of energy for human life, but in sub-Saharan Africa, 548 million people still have no access to electricity. The international charity SolarAid aims to provide access to solar lights in areas without electricity. In order to keep an up to date mapping of the regions to prioritize when distributing their solar lanterns to the population without electricity, the accurate detection of settlements is therefore very important. With the rapid development of deep neural networks, image processing techniques have also ushered in new changes. For remote sensing images, the generalization and classification accuracy of the classification model based on deep neural networks are significantly better than those of traditional image processing methods, providing a feasible technology for the detection work based on remote sensing images.

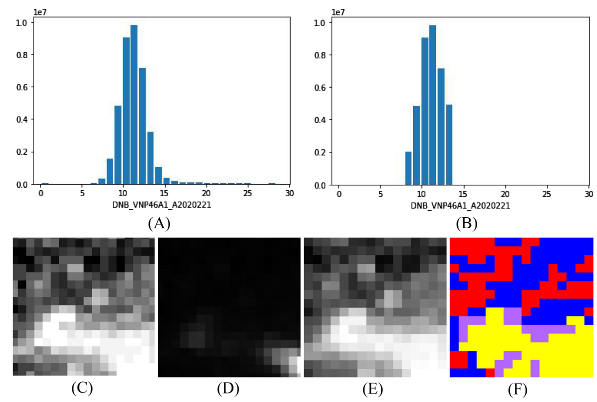

Fig. 1. Comparison of The Effects of Our Outlier Removal Method. (C) The single-channel image obtained from 9 night time images after using our outlier removal method. (D) The single-channel image obtained from 9 night time images without using the outlier removal method. (E) The image generated after using average pooling on C. (F) Ground truth.

In our work, the task can be divided into the detection of human settlements and the detection of the presence of electricity. It can be obtained from observing remote sensing images that, contrary to the distribution characteristics of settlements in urban areas, settlements in non-urban areas are more scattered and smaller in area, thus requiring higher feature extraction capabilities for spatial location information. Considering the need to determine whether there are settlements in an area of each $50*50$ pixels, the classification-based methods are preferred. The neural network with attention mechanism can extract richer and more accurate spatial information [1] and has a significant improvement in the detection of inconspicuous features. Therefore, we chose SENet, and the experimental results show that the SENet154 model outperforms other models.

The Suomi Visible Infrared Imaging Radiometer Suite (VIIRS) dataset can be used to detect the presence or absence of electricity in an area. However, the problem is that even if the ground truth shows that there is no electricity in an area, the images images obtained by pre-processing still show irregular light and dark variations. To address this problem, we study an outlier removal method based on the characteristics of the pixel distribution of the VIIRS dataset, and the processed images can more accurately reflect the location of areas with electricity. In the post-processing stage, a novel method is proposed for correcting the results of settlements detection, resulting in a significant performance improvement.

2. DATA PRE-PROCESSING

For each dataset, Sentinel-1 dataset did not find an effective data processing method and the experimental results were not satisfactory, and the ground sampling distance of Landsat 8 dataset was larger than that of Sentinel-2, which made the image resolution lower, so Sentinel-2 dataset [2] was finally selected. In this section, considering the existence of noise, we perform outlier removal process on the data and propose a new outlier removal method for the pixel distribution characteristics of the VIIRS dataset.

2.1. Pre-processing of Sentinel-2 Multispectral Dataset

The Sentinel-2 multispectral dataset [3] contains 12 channels, each with $800 * 800$ pixels, where the synthesis of bands 4, 3, and 2 can yield true color (RGB) images. By comparing the experimental results of the images synthesized by the other bands of Sentinel-2, RGB images are finally selected to detect human settlements, and the specific comparison results will be presented in the Experiments.

Firstly, the outliers of the three channels are processed. To remove extreme outliers, we experimented with multiple truncation parameters, and the best results were experimented with a parameter of 2000, so the pixel values of each channel are clipped to the interval [0,2000] [4], and then the synthesized images are normalized to [0,255]. Secondly, since the ground sampling distance of the RGB image is $10m * 10m$ and the ground truth sampling distance is $500m * 500m$, the submitted ground truth is $16 * 16$ pixels, which means that for the image with $800 * 800$ pixels, it is necessary to determine whether there are human settlements in each $50 * 50$ pixels area, so each RGB image is cropped to 256 images with $50 * 50$ pixels. Thirdly, the cropped images with class names "Human settlements without electricity" and "Human settlements with electricity" are merged into a new class named "Human settlements", and similarly, the images of the remaining two classes are merged into a new class named "No human settlements", thus obtaining a dataset with two new classes that can be used to determine whether there are settlements in a image. Finally, we expand the dataset using five data augmentation methods: rotate 90 degrees, rotate 180 degrees, rotate 270 degrees, flip horizontally and flip vertically, and the experiments show that the expanded dataset shows better results.

2.2. Pre-processing of The VIIRS Night Time Dataset

The VIIRS dataset provides 9 multitemporal single-channel night time images for each tile, which can be used to detect the presence or absence of electricity in an area. In order to eliminate the noise interference in the image acquisition process, the outliers in the data need to be processed. First, in the training set, we count the distribution of all pixels in a total of 60 images from the same channel. Next, all the pixel values of each channel are sorted from smallest to largest, and the values at 5% and 95% in the sort are recorded as L_{c_min} and L_{c_max}, respectively, where c denotes the channel name. To make the overall variance of the retained pixel values smaller, we also tried various truncation parameters, and the values at 5% and 95% were finally chosen because they performed best in the experiment. Then, the method of removing outliers can be represented by the formula

$$pixelc_{i,j} = \begin{cases} L_{c\_min}, & pixelc_{i,j} \leqslant L_{c\_min} \\ pixelc_{i,j}, & L_{c\_min} < pixelc_{i,j} < L_{c\_max} \\ L_{c\_max}, & pixelc_{i,j} \geqslant L_{c\_max} \end{cases} \quad (1)$$

We refer to this method as an outlier removal method based on the characteristics of the pixel distribution of the VIIRS dataset. Taking the channel DNB_VNP46A1_A2020221 as an example, with the pixel value as the horizontal axis and the number of pixels as the vertical axis, the pixel distribution of this channel is shown in Fig.1(A), and it can be seen that there are extremely small or extremely large pixel values. By calculation, the $L_{\text{DNB\_VNP46A1\_A2020221\_min}}$ is 8.0 and the $L_{\text{DNB\_VNP46A1\_A2020221\_max}}$ is 14.0. The pixel distribution after removing the outliers is shown in Fig.1(B).

For each tile, we sum the pixel values of the 9 night time images after removing outliers and normalize them to the interval [0,255] and thereby obtain a single-channel night time image. Since grouth truth is a coarse label of $16 * 16$ pixels, we use $50 * 50$ average pooling instead of max-pooling for the obtained single-channel image to generate a new image of $16 * 16$ pixels. The reason for not using max-pooling is that it would cause the original area of detected light to be excessively expanded and more prone to false detection. The generated $16 * 16$ pixels images are also expanded using the five data augmentation methods described above to obtain a new night time dataset for training. Taking tile14 in the training set as an example, and Fig.1 shows that the areas with electricity can be more accurately identified using our method.

3. THE DETECTION OF HUMAN SETTLEMENTS WITHOUT ELECTRICITY

In this section, we formulate two separate detection tasks, settlements detection and light detection, and propose a post-processing method to correct the results of settlements detection based on regional continuity. The sentinel-2 images and VIIRS night time images, which were processed using

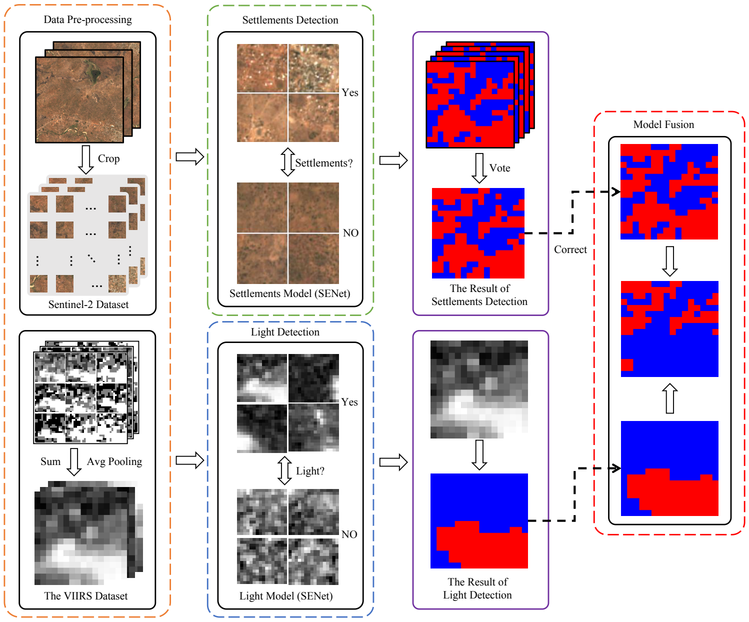

Fig.2

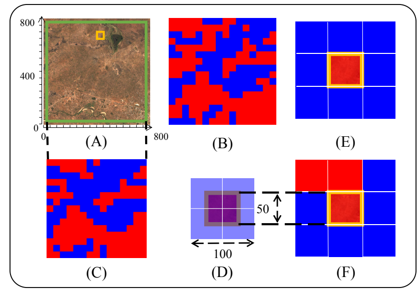

Fig.3

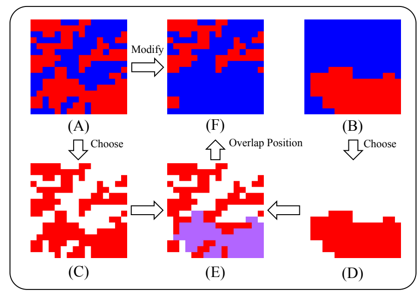

Fig.4

Fig. 2. The Detection Process for Human Settlements without Electricity. For a binary map with $16 * 16$ pixels, the red color represents 1 and the blue color represents 0. **Fig. 3**. Correction Process for The Results of Dettlements Detection. The area marked by the green line indicates the image with $750 * 750$ pixels. **Fig. 4**. The Process of Fusion of Two Results. The purple color represents the overlapping regions.

the pre-processing method introduced in Section 2, are used as the training set, and the detection process for settlements without electricity is shown in Fig.2.

3.1. Settlements Detection

After pre-processing, an expanded Sentinel-2 dataset with $50 * 50$ pixels is obtained. By setting different hyper-parameters, the pre-processed training set is used to train multiple convolutional neural networks SENet154 (hereafter referred to as Settlements Model). On the test set, the predictions of the top 5 best performing classification models on the validation set are used for fusion. The presence of settlements is recorded as 1 and the absence of settlements is recorded as 0. The fusion rule is that for the same area, if there are more votes for 1 than for 0, the prediction is set to 1, otherwise it is set to 0. For each tile, a binary map with $16 * 16$ pixels can be obtained as the final prediction to represent the result of settlements detection.

3.2. Light Detection

The new dataset obtained by preprocessing the VIIRS dataset is used to train a SENet154 model (hereinafter referred to as Light Model) to determine whether there is an area with electricity in a night time image. We define the $50 * 50$ pixels part of an $800 * 800$ pixels image as a region, which means that an image consists of 256 regions. If a night time image is classified by Light Model as having the presence of electricity, then the specific regions with electricity in the image need to be detected.

If there is light in a region, the region is recorded as 1 and no light is recorded as 0. By setting a threshold value, regions with values greater than the threshold value are set to 1 and regions with values less than the threshold value are set to 0, thus obtaining the specific regions with electricity in the image. Since the optimal threshold value is different for different night time images, and it is an empirical value, our threshold selection method is to ensure that all the bright enough regions in the image are 1, when the corresponding threshold value is the most appropriate. For each tile, a binary map with $16 * 16$ pixels can also be obtained to represent the results of light detection.

3.3. Transfer Learning

For both Settlements Model and Light Model, we use SENet154 pre-trained on ImageNet for transfer learning. The channel attention mechanism allows the network to learn the weights of channels adaptively and assign different weights to different channels. We set the batchsize to 32, epoch to 15, and learning rate to 0.001, and the learning rate decays to 90% of the original one for each epoch during training.

3.4. Correction of The Results of Settlements Detection

As shown in Fig.3, (A) is an RGB image with $800 * 800$ pixels from the Sentinel-2 dataset, and (B) is a binary map with $16 * 16$ pixels obtained from (A) by settlements detection. In this subsection, we propose a post-processing method based on regional continuity for the correction of (B).

An image with $750 * 750$ pixels is acquired at the center of (A), and similarly, (C) is a binary map with $15 * 15$ pixels obtained from this image by settlements detection. A region with $50 * 50$ pixels is randomly selected in (A), and taking the region marked by the orange line as an example, we can see that the position of this region is predicted as 1 in (B), but is predicted as 0 in (C), as shown in (D), which means that the two predicted results conflict. At this point, we choose to check the pixel values of the other 8 regions around the conflicting region in (B), and if the pixel values are all 0, as shown in (E), the pixel value 1 of the region will be corrected to 0. Otherwise, due to the continuity of the settlements distribution, which we refer to as regional continuity, as shown in (F), the region is not corrected.

3.5. Model Fusion

The obtained corrected result of settlements detection (Fig.4(A)) and the result of light detection (Fig.4(B)) are used for fusion to generate the image with $16 * 16$ pixels of the settlements without electricity. The fusion method is shown in Fig.4, where all the regions set to 1 in (A) are chosen as (C), and all the regions set to 1 in (B) are chosen as (D). If there is an overlap between (C) and (D) as shown in (E), the final submission result (F) is obtained by modifying 1 to 0 at the overlap position in (A).

4. EXPERIMENTS

We have tested the performance of multiple networks on the validation set [1, 5, 6] and compared the effectiveness of using multiple operations. As shown in Table 1, SENet154, which used all operations, performed the best, and therefore SENet154 was chosen to be used on the test set.

Table 1. Performance of Different Combinations of Methods

Settlements Model	ResNet34	ResNet50	SENet50	PNASNet	SENet154 (All Channels)	SENet154 (RGB)
Recall	0.8224	0.8052	**0.8927**	0.8575	0.8356	0.8549
Precision	0.8576	**0.8745**	0.8037	0.8275	0.7946	0.8600
F1 score	0.8396	0.8384	0.8459	0.8422	0.8151	**0.8575**
F1 score ([0,2000] Truncation)	0.8410	0.8396	0.8457	0.8466	0.8204	**0.8592**
F1 score (Data Augmentation)	0.8527	0.8446	0.8525	0.8572	0.8311	**0.8678**

Table 2 shows the improvement of F1 score on the test set by adding various operations. After fusing the results of settlements models with the results of light detection, the F1 is 88.64%, and after post-processing, it reaches 89.39%.

Table 2. F1 Score for Different Combinations of Operations

Operation	Settlements Model	Settlements Models Fusion	+Light Model	+Post-processing
F1 score	84.14%	85.30%	88.64%	89.39%

5. CONCLUSION

In this paper, we make full use of the multisource remote sensing data and formulate two detection tasks to detect human settlements without electricity. The studied outlier removal method based on the pixel distribution characteristics of the VIIRS dataset significantly improves the performance of Light Model, which can be widely used in pre-processing for the VIIRS dataset. The proposed post-processing method based on the regional continuity incorporates a priori knowledge and effectively corrects the prediction results.

6. ACKNOWLEDGEMENT

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.

7. REFERENCES

[1] H. Jie, S. Li, S. Gang, and S. Albanie, "Squeeze-and-excitation networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, no. 99, 2017.

[2] Naoto Yokoya, Pedram Ghamisi, Ronny Hansch, Colin Prieur, Hana Malha, Jocelyn Chanussot, Caleb Robinson, Kolya Malkin, and Nebojsa Jojic, "2021 data fusion contest: Geospatial artificial intelligence for social good [technical committees]," *IEEE Geoscience and Remote Sensing Magazine*, vol. 9, no. 1, pp. 287–C3, 2021.

[3] "2021 ieee grss data fusion contest," [REF. NO.], online:`www.grss-ieee.org/community/technical-committees/data-fusion`.

[4] M. Schmitt, J. Prexl, P. Ebel, L. Liebel, and X. X. Zhu, "Weakly supervised semantic segmentation of satellite images for land cover mapping – challenges and opportunities," 2020.

[5] "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision & Pattern Recognition*, 2016.

[6] C. Liu, B. Zoph, M. Neumann, J. Shlens, W. Hua, L. J. Li, L. Fei-Fei, A. Yuille, J. Huang, and K. Murphy, "Progressive neural architecture search," 2017.