SolarNet: A Deep Learning Framework to Map Solar Plants In China From Satellite Imagery

Xin Hou, Biao Wang, Wanqi Hu, Lei Yin, Anbu Huang, Haishan Wu *
WeBank AI group

Abstract

Renewable energy such as solar power is critical to fight the ever more serious climate, how to effectively detect renewable energy has became an important issue for governments. In this paper, we proposed a deep learning framework named SolarNet which is designed to perform semantic segmentation on large scale satellite imagery data to detect solar farms. SolarNet has successfully mapped 439 solar farms in China, covering near 2000 square kilometers, equivalent to the size of whole Shenzhen city or two and a half of New York city. To the best of our knowledge, it is the first time that we used deep learning to reveal the locations and sizes of solar farms in China, which could provide insights for solar power companies, climate finance and markets.

1 Introduction

While climate change has become one of greatest threats to our world, renewable energy such as solar power is critical to fight climate change(Chu & Majumdar, 2012; Agnew & Dargusch, 2015). China, as the world's leading installer of solar photovoltaics (PV), is the world's largest producer of solar PV power and massive solar farms were built not only to produce clean energy but also to reduce poverty.

However, one question remains to be answered: where are those solar farms located? Mapping the location of solar farms and tracking its installation progress is particularly important for the following aspects: first, it allows the government to gauge the development of solar power industry and make strategies; second, it helps the solar power company to quantify and optimize the efficiency of solar panels; third, it is useful for investors to evaluate the operation of solar power companies. Obviously, it is impractical to locate solar farms with maps manually. Most recently, more and more companies have launched satellites into space, produced massive satellite imagery data and therefore accelerated its commercialization in various fields.

In this paper, we proposed a deep learning framework named **SolarNet**, which is used to analyze large-scale high-resolution satellite imagery data and is able to accurately identify hundreds visible large solar farms in China while many of those are built in deserts, mountains and even lakes. To the best of our knowledge, it is the first time that the locations and sizes of solar farms in China are tracked by mining satellite imagery data through deep learning algorithms.

2 Related Works

In this section, we give a brief review of related works, including semantic segmentation and solar panel detection.

Semantic Segmentation: In order to evaluate the use of solar energy, we need to use segmentation algorithms to evaluate the construction of solar power stations. If we can segment the area of the solar power board, we can estimate its power generation. Semantic segmentation(Long et al., 2015) is an important computer vision technique that has been widely applied to detect objects from

---

* haishanwu@webank.com

remote sensing imagery data, such as urban architectural segmentation(Wei et al., 2004; Bischke et al., 2019), road extraction(Mokhtarzade & Zoej, 2007), crop segmentation(Rydberg & Borgefors, 2001), etc. Deep learning has achieved great success in semantic segmentation task(Long et al., 2015).

However, compared with natural images, segmentation on satellite imagery data is much more challenging due to: 1) the resolution of different satellite may vary, 2) the size of satellite is huge which may lead to huge computational cost, 3) the background, cloud, reflection of sunshine etc. could also complicate the segmentation tasks, 4)the texture of solar panels may also vary due to various sensor specs.

Solar Panel Detection: Most recently, Yu et al.(Yu et al., 2018) proposed a framework called Deep-Solar which successfully located the civil solar panels in the United States and developed a public data set. Their data set mainly focused on household solar power planes in the US, by contrast, most of the large solar power plants in China were built in the fields with complex background such as deserts, mountains and even lakes as shown in Figure 1, which pose more challenges to the detection task. In order to fully evaluate the proposed segmentation method, we also particularly created a satellite imagery data set of the solar plants in China to train our model.

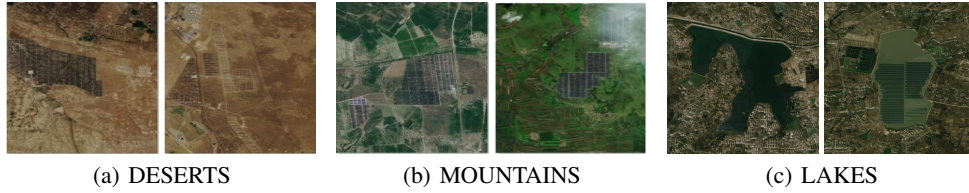(a) DESERTS   (b) MOUNTAINS   (c) LAKES

Figure 1: Part of solar farms in China. (a). solar power plants in the deserts, (b). solar power plants in the mountains, (c). solar power plants in the lakes. All these images contain complex backgrounds.

.

3 METHOD

SolarNet is a semantic segmentation method based on the combination of Expectation-Maximization Attention Unit(EMAU) and multitask optimization. In order to compare the performance, we used UNet as a baseline algorithm, which is one of most popular deep learning based semantic segmentation methods.

3.1 UNET

UNet is composed of multiple layers convolution and deconvolution. This architecture can produce a prediction for each pixel, while retaining the spatial information in the original input image. The UNet was first proposed by (Ronneberger et al., 2015), is used as a baseline model and the net architecture is illustrated in Figure 2.
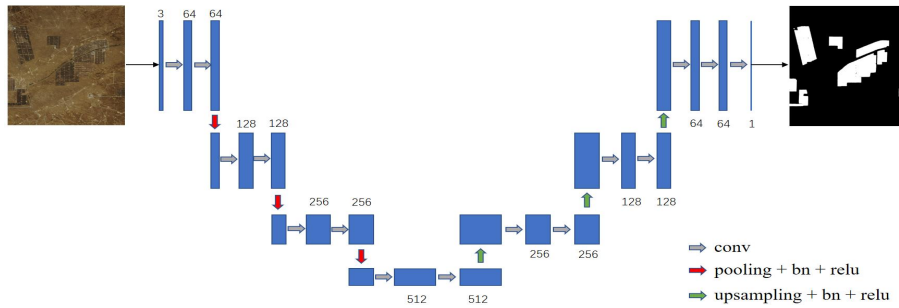

Figure 2: UNet Architecture.

3.2 SOLARNET: A MULTITASK EXPECTATION-MAXIMIZATION ATTENTION NETWORKS

EMAU: One shortcoming of UNet segmentation structure is that its multiple local convolution operations is not able to capture sufficient global information, and thus harms the performance in discontinuous object segmentation. Inspired by Xia's work(Li et al., 2019), they poposed the Expectation-Maximization Attention Unit(EMAU) module, which replaced the local convolution operation with the global unsupervised clustering EM algorithm in the feature extraction process, thus could effectively captures the global information. In our case, the solar power plants usually scatter in various discontinuous areas as shown in Figure 3, and EMAU moudle is able to deal with such case.
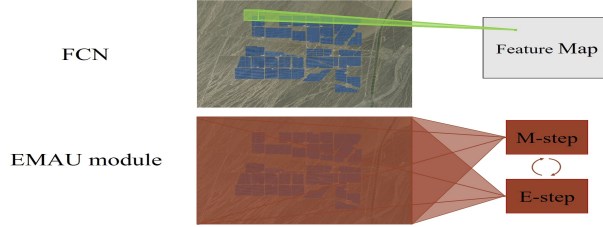

Figure 3: When performing convolution operation, each convolution operator only extracts the local spatial features. The EMAU module performs clustering operation of element wise, and could capture more the global information in space.

Multitask Optimization: Besides, we proposed an optimized multitask-EMANet, which combines local pixel-level segmentation and global image-level classification. Many existing studies(Zhou et al., 2016; Le et al., 2019) show that the feature map of classification network usually corresponds the area of the object to be segmented, which could improve the segmentation performance.
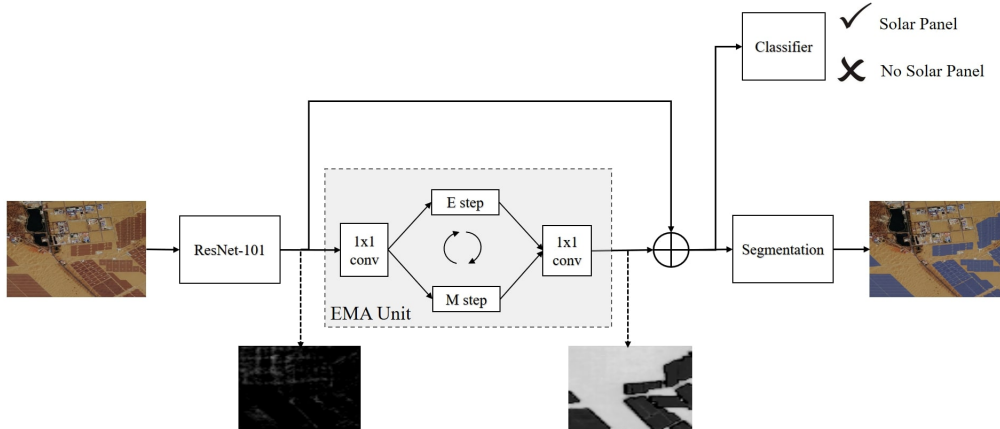

Figure 4: SolarNet Architecture: in addition to the EMA operator, two 1×1 convolutions at the beginning and the end of EMA, sum the output with original input, to form a residual-like block.

The proposed SolarNet complete architecture is shown in Figure 4, it used pretrained ResNet-101 as backbone(He et al., 2016) and the EMAU module to extract features. After re-configuring the features of EMAU module, the feature of ResNet-101 were then summed together and the last summed one was used to the last segmentation task. SolarNet adopted the classification network to further enhance the segmentation results. Meanwhile, the classification network shares the same weight with segmentation network, and the final layer is a fully connected layer which is used to classify whether contains the solar planes or not. More detailed structure and training process have been described in appenidx section.

4 RESULTS AND CONCLUSION

Results: We used the trained SolarNet framework to map all the solar farms in China by mining large scale satellite imagery data that covered the whole China. We successfully detected about 500 solar farms covering the area of 2000 square kilometers or 770 square miles in total, equivalent to the size of whole Shenzhen city or two and a half or New York city. Figure 5 visualized the locations of all detected solar farms in China marked by blue dots. One can see that most of the solar farms were built in the northwestern part of China where the sunlight is abundant and thus is ideal for solar power. Among all the provinces in China, Qinghai has installed the most solar farms with the area of near 400 square kilometers in total as shown in Figure 8. One can also visit our website ms.webank.com/moonshot-demo/solarnet-beta.html to interact with the results.
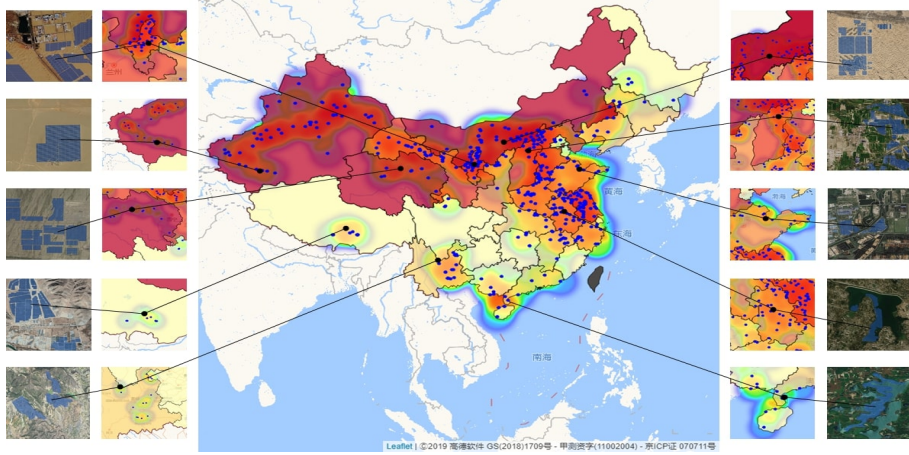

Figure 5: Solar farm map in China. Each blue dot indicates a detected solar farm from satellite imagery. We colored each province according to the area of solar farms (darker color indicates larger areas). A heat map of solar farm density was also overlaid. Eight representative solar farms built on deserts, mountains, lakes or the fields were also displayed.

We then used mean Intersection over Union (mIoU) as the criteria to evaluate segmentation performance and compared the SolarNet with two other methods in three kinds of dataset. Our dataset include 938 images, deepsolar dataset(Yu et al., 2018) include 1555 images. The results in Table 1 shows that the SolarNet outperformed two others.

Model	mIOU		
	our dataset	deepsolar dataset	our+deepsolar dataset
Resnet101-Unet	84.65%	84.22%	86.54%
Resnet101-EMANet-single	94.00%	**90.98%**	93.79%
SolarNet-Multitask-1.0	**94.21%**	90.39%	**93.94%**

Table 1: With the multi-task embedding, SolarNet could beat the orignal EMANet and UNET on our dataset evaluation.

Conclusion: Mapping and tracking the installment of solar panel from satellite imagery data is potentially helpful for the following fields: 1) it could help the solar PV power company to optimize the location and direction of solar panels so that to maximize the production of such renewable energy; 2) it could help the investors and market researchers to track the latest trends of solar PV power; 3) government could use our result to evaluate the efficiency of regarding policies, for example, how the subsidiary policy is impacting the development of solar power industry. Therefore, we plan to build a Solar Power Index in China by analyzing longer historical satellite imagery data with SolarNet so that we could track long term trends. And we also plan to apply the proposed framework to map the location and develop the index of other type of renewable energy such wind turbine.

REFERENCES

Scott Agnew and Paul Dargusch. Effect of residential solar and storage on centralized electricity supply systems. *Nature Climate Change*, 5(4):315–318, 2015.

Dimitrios Alexakis, Apostolos Sarris, Theodoros Astaras, and Konstantinos Albanakis. Detection of neolithic settlements in thessaly (greece) through multispectral and hyperspectral satellite imagery. *Sensors*, 9(2):1167–1187, 2009.

Paul Arellano, Kevin Tansey, Heiko Balzter, and Doreen S Boyd. Detecting the effects of hydrocarbon pollution in the amazon forest using hyperspectral satellite images. *Environmental Pollution*, 205:225–239, 2015.

Ursula C Benz, Peter Hofmann, Gregor Willhauck, Iris Lingenfelder, and Markus Heynen. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for gis-ready information. *ISPRS Journal of photogrammetry and remote sensing*, 58(3-4):239–258, 2004.

Benjamin Bischke, Patrick Helber, Joachim Folz, Damian Borth, and Andreas Dengel. Multi-task learning for segmentation of building footprints with deep neural networks. In *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 1480–1484. IEEE, 2019.

Christopher Burges, Tal Shaked, Erin Renshaw, Ari Lazier, Matt Deeds, Nicole Hamilton, and Gregory N Hullender. Learning to rank using gradient descent. In *Proceedings of the 22nd International Conference on Machine learning (ICML-05)*, pp. 89–96, 2005.

Steven Chu and Arun Majumdar. Opportunities and challenges for a sustainable energy future. *nature*, 488(7411):294, 2012.

Amod V Dandawate and Georgios B Giannakis. Asymptotic theory of mixed time averages and kth-order cyclic-moment and cumulant statistics. *IEEE Transactions on Information Theory*, 41 (1):216–232, 1995.

Timothy Dozat. Incorporating nesterov momentum into adam. 2016.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pp. 1026–1034, 2015.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *European conference on computer vision*, pp. 630–645. Springer, 2016.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pp. 1097–1105, 2012.

Thi-Lam-Thuy Le, Nicolas Thome, Sylvain Bernard, Vincent Bismuth, and Fanny Patoureaux. Multitask classification and segmentation for cancer diagnosis in mammography. *arXiv preprint arXiv:1909.05397*, 2019.

Xia Li, Zhisheng Zhong, Jianlong Wu, Yibo Yang, Zhouchen Lin, and Hong Liu. Expectation-maximization attention networks for semantic segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 9167–9176, 2019.

Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440, 2015.

Tomáš Mikolov, Martin Karafiát, Lukáš Burget, Jan Černockỳ, and Sanjeev Khudanpur. Recurrent neural network based language model. In *Eleventh annual conference of the international speech communication association*, 2010.

Mehdi Mokhtarzade and MJ Valadan Zoej. Road detection from high-resolution satellite images using artificial neural networks. *International journal of applied earth observation and geoinformation*, 9(1):32–40, 2007.

Todd K Moon. The expectation-maximization algorithm. *IEEE Signal processing magazine*, 13(6): 47–60, 1996.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015. URL `http://arxiv.org/abs/1505.04597`.

Anna Rydberg and Gunilla Borgefors. Integrated method for boundary delineation of agricultural fields in multispectral satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, 39(11):2514–2520, 2001.

Ke Sun, Bin Xiao, Dong Liu, and Jingdong Wang. Deep high-resolution representation learning for human pose estimation. *arXiv preprint arXiv:1902.09212*, 2019.

Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He. Non-local neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7794–7803, 2018.

Yanfeng Wei, Zhongming Zhao, and Jianghong Song. Urban building extraction from high-resolution satellite panchromatic image using clustering and edge detection. In *IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium*, volume 3, pp. 2008–2010. Ieee, 2004.

Jiafan Yu, Zhecheng Wang, Arun Majumdar, and Ram Rajagopal. Deepsolar: A machine learning framework to efficiently construct a solar deployment database in the united states. *Joule*, 2(12): 2605–2617, 2018.

Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2921–2929, 2016.

5 APPENDIX

UNET architecture: The network architecture is described in detail in Table 2. It has tow parts: a contracting path and an expansive path. The contracting path follows the typical architecture of a convolutional network. we uses two repeated convolutions with 3×3 kernerl size, while each is followed by a batch normalization layer and a rectified linear unit, then a 2x2 max pooling operation with stride 2 for downsampling. At each downsampling step we made the number of feature channels becomes to double times. In the expansive process every step consists of upsampling feature map followed by a 2x2 convolution that halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3x3 convolutions, each followed by a BN layer and a ReLU layer. In the final layer, a 1x1 convolution is used to map each 2-component feature vector to the desired number of classes whether this pixel is solar panel or not. The network has 17 convolutional layers in total.

INPUT			
3x3 conv 64 dim→	3x3 conv 64 dim→	pooling →	BN & RELU
3x3 conv 128 dim→	3x3 conv 128 dim→	pooling →	BN & RELU
3x3 conv 256 dim→	3x3 conv 256 dim→	pooling →	BN & RELU
3x3 conv 512 dim→	3x3 conv 512 dim→	pooling →	BN & RELU
3x3 conv 512 dim→	3x3 conv 512 dim→	upsampling →	BN & RELU
3x3 conv 256 dim→	3x3 conv 256 dim→	upsampling →	BN & RELU
3x3 conv 128 dim→	3x3 conv 128 dim→	upsampling →	BN & RELU
3x3 conv 64 dim→	3x3 conv 64 dim→	upsampling →	BN & RELU
1x1 conv 2 dim	→	SoftMax	

Table 2: UNet architecture detail

EMAU module: Attention mechanism have been widely used for various tasks. The proposed Expectation-Maximization Attention (EMA) module (Li et al., 2019) is robust with regard to the variance of input and is also efficient in terms of memory and computational power(Wang et al., 2018). For a simple introudction, we consider an input feature map X of size $C \times H \times W$ from a single image. X was the intermediate activated feature map of a CNN. We reshaped X into $N \times C$, where $N = H \times W$. Briefly, given the input $X \in \mathbb{R}^{N \times C}$, the initial bases $\mu \in \mathbb{R}^{K \times C}$ and $Z \in \mathbb{R}^{N \times N}$ are the latent variables. The E-step is used to estimates the latent variables Z, and then used the M-step updated bases μ. After T times iteration, we reconstruct the $\hat{X}$ since $K << N$, $\hat{X}$ lies in a subspace of X. This method removes much unnecessary noise and makes the final classification of each pixel more segmentable. Moreover, this operation reduces the complexity from $O(N^2)$ to $O(NK)$ in the pixel segmentation process.

E-step:

$$z_{nk} = \frac{\kappa(x_n, \mu_k)}{\sum_{j=1}^{K} \kappa(x_n, \mu_j)} \tag{1}$$

where κ represents the general kernel function, we simply take the exponential inner dot $exp(a^T, b)$ in our implementation.

M-step:

$$\mu_k^t = \frac{z_{nk}^t X_n}{\sum_{m=1}^{N} z_{mk}^t} \tag{2}$$

Multitask Optimization: we used this formulaton to get the overall loss:
$Loss_{total} = \lambda \cdot Loss_{cls} + (1 - \lambda) \cdot Loss_{seg}$

Implementation of training details: When training the model, we also adopted adam gradient descent method(Burges et al., 2005; Dozat, 2016). In order to fully incorporate the EMAU's into deep neural networks, we here describe how to train EMAU in each iteration. As each image X has different pixel feature distributions compared to others, using the μ to reconstruct feature maps of a new image is not suitable. So we need to run EMAU moudle on each image simultaneously. For the first mini-batch, the Kaiming's initialization(He et al., 2015) has been used to initialize μ_0 , where the matrix multiplication can be treadted as a 1×1 convolution. For the following batches, we can simple used back propagation to update μ_0 by standard. However, since iterations of E-step and M-step can be expanded as a recurrent neural network (RNN) (Mikolov et al., 2010), the gradients propagating though them will generate the vanishing or explosion problem. Therefore, the updating of μ_0 is unstable, moving averaging(Dandawate & Giannakis, 1995) has been used to update μ_0 in the training process. After several iterations over an image, the generated μ_T can be considered as a biased update of μ_0, where the bias comes from the image sampling process: $\mu_0 = \alpha\mu_0 + (1-\alpha)\mu_T$

2058 images were used to train the mode while 435 images were used to test the model. The size of all the images ranges from 512×512 to 10000×10000. In order the create more dataste to train the model, we adopted the following data augmentation methods:

- **Crop**: Choosed a random ROI area from a original image: $X_{arg} = ROI(X)$.

- **Scale**: Choosed a random scale size $s \in (0.8, 1.2)$, rescaled the original image: $X_{arg} = Rescale(X, s)$

- **Rotation**: Choosed a random angle $\theta \in (-180, 180)$, rotated the orignal image:$X_{arg} = Rotate(X, \theta)$

- **Reflection**: Flipped the original image horizontally: $X_{arg} = FlipH(X)$, or flipped the original image vertically: $X_{arg} = FlipV(X)$

Parameter	Learning Rate	Iteration	Training Set	Testing Set
Value	$1e^{-3}$	20000	819	119
Parameter	EM Iteration		EM Latent Variables Size	
Value	10		1024	

Table 3: Parameters of SolarNet to train the model.

The pseudo code of the training process of SolarNet is shown in Algorithm 1. It is important to note that in each iteration a semi-supervised clustering process of T-round EMAU module is required. And in the test process, each image was performed a clustering process with T-round iteration.
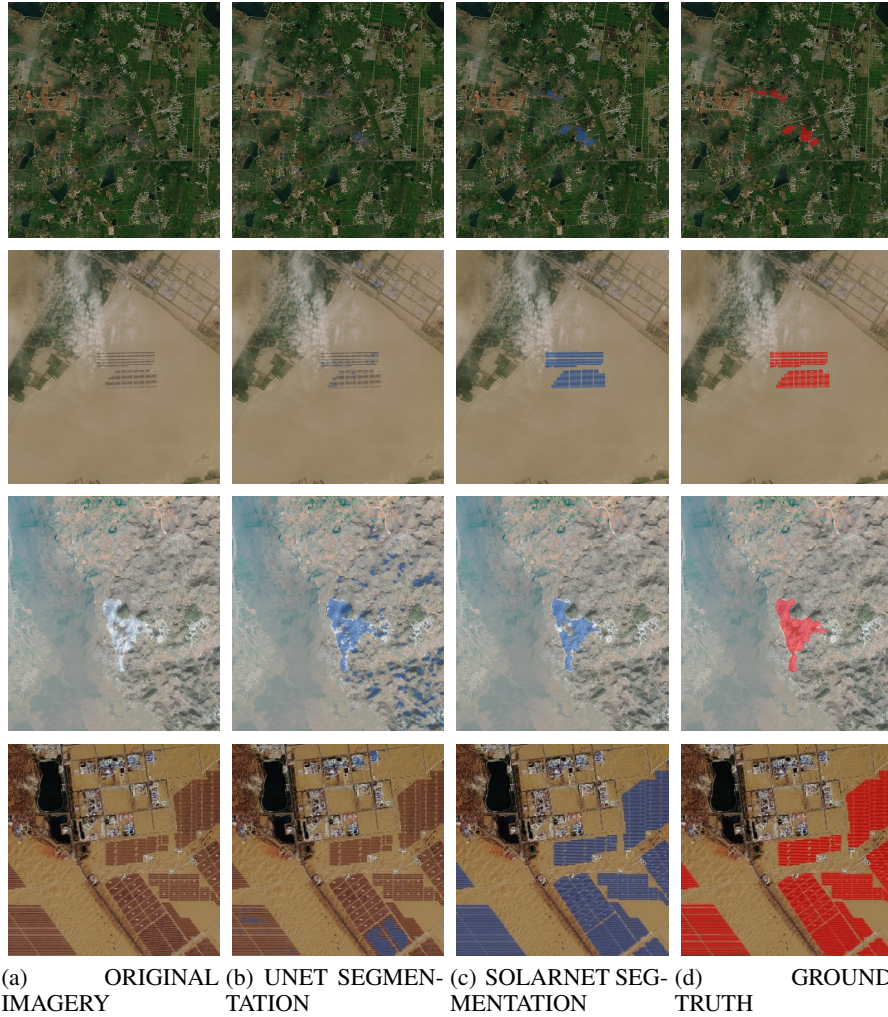
Experimental details:

(a) ORIGINAL IMAGERY	(b) UNET SEGMEN-TATION	(c) SOLARNET SEG-MENTATION	(d) GROUND TRUTH

Figure 6: Solar farms located by SolarNet. The first column is the orignal satellite imagery data. The blue area indicates the detected solar farms by UNet (second column) and EMANet (third column) and red area in the fourth column indicates the ground-truth labeled manually. One can see how SolarNet was able to accurately detect solar farms under very complicated backgrounds.
.

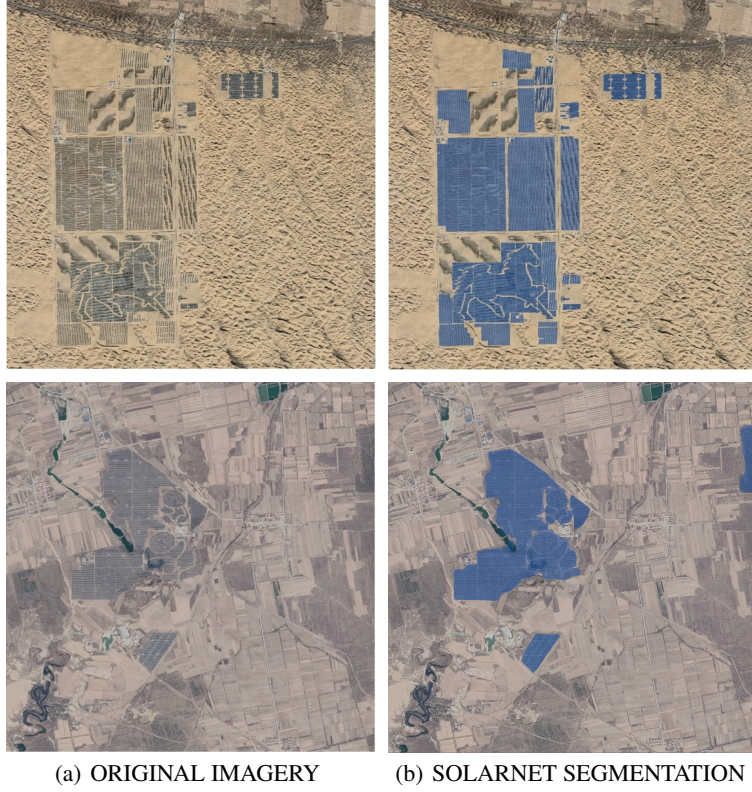(a) ORIGINAL IMAGERY   (b) SOLARNET SEGMENTATION

Figure 7: Two massive animal-shaped (horse and panda) solar farms detected by SolarNet.
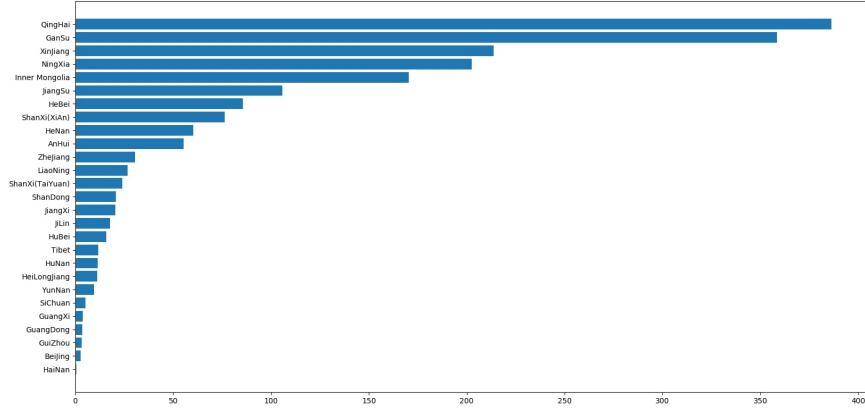

Figure 8: The area of detected solar farms in various provinces in China (unit: km^2)

SolarNet may fail to detect the solar farms when the it resembles its surrounding background as shown in Figure 9. In the future, we plan to improve our methods in the following way:

1) Labeling more solar panels from the satellite imagery data in various circumstances, such as the solar panels on the roof in residential areas.

2) Adapting SolarNet to handle the satellite imagery data with various resolutions Benz et al. (2004). For example, HRNet proposed by Sun et al. (2019) is an effective super-resolution method to deal with various resolution images.

3) Using hyperspectral imagery data to enhance the segmentation performance. As showed in Alexakis et al. (2009); Arellano et al. (2015) could provide more information when detecting objects from satellite.

(a) ORIGINAL IMAGERY   (b) SOLARNET SEGMENTA-   (c) GROUND TRUTH
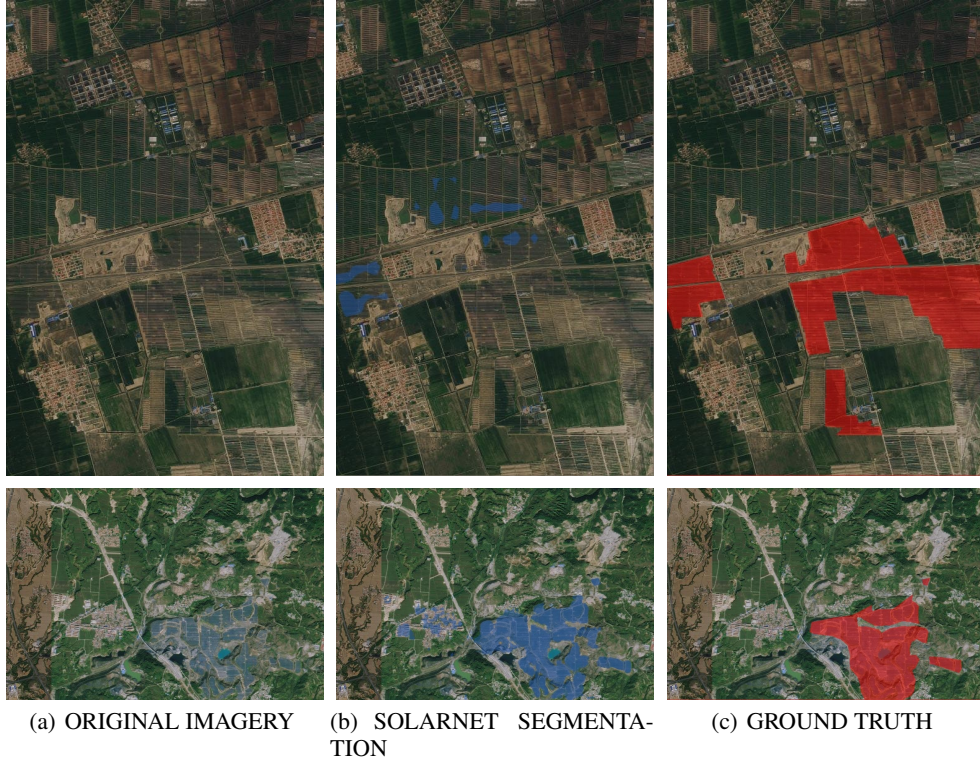   TION

Figure 9: SolarNet may fail to detect the solar farm when it resembles its surrounding environment.

Algorithm 1 SolarNet Training Procedure

Initial: Random Initial network's weights: W_0
Input: Original Satellite Imagery: X
              Semantic Segmentation Imagery: X_s
              Whether it contains solar panels: Y
Procedure:
1: **function** ESTEP(μ, X)
2:      **return** $z_{nk} = \frac{\kappa(x_n, \mu_k)}{\sum_{j=1}^{K} \kappa(x_n, \mu_j)}$
3: **end function**
4: **function** MSTEP(Z, X)
5:      **return** $\mu_k^t = \frac{z_{nk}^t X_n}{\sum_{m=1}^{N} z_{mk}^t}$
6: **end function**
7: **for** $i = 0 \rightarrow MaxIter$ **do**
8:      $X_{res} = ResNet(X)$
9:      $Logit = Cls(X_{res})$
10:      $L_{cls} = CrossEntropy(Logit, Y)$
11:      Random initial μ_0
12:      $Z_0 = $ ESTEP(μ_0, X_{res})
13:      **for** $t = 0 \rightarrow T$ **do**
14:          $u_t = $ MSTEP(Z_t, X_{res})
15:          $Z_{t+1} = $ ESTEP(μ_t, X_{res})
16:      **end for**
17:      $\widetilde{X} = Z_t \cdot \mu_t$
18:      $L_{seg} = CrossEntropy(\widetilde{X}, X_s)$
19:      $L_{total} = \lambda \cdot L_{cls} + (1 - \lambda) \cdot L_{seg}$
20:      $W_{i+1} = W_i + \frac{\partial L_{total}}{\partial W_i}$
21: **end for**