Infrared Image Segmentation for Photovoltaic Panels Based on Res-UNet

Chapter *in* Lecture Notes in Computer Science · October 2019

DOI: 10.1007/978-3-030-31654-9_52

Hao Zhang
Nanjing University of Aeronautics and Astronautics

Infrared Image Segmentation for Photovoltaic Panels Based on Res-UNet

Hao Zhang[1[0000−0003−1923−589X]⋆], Xianggong Hong[1,*], Shifen Zhou[1], and Qingcai Wang[1]

School of Information Engineering, Nanchang University, Nanchang, China
*Corresponding author: hongxianggong@ncu.edu.cn

Abstract. Infrared image segmentation is the basis of error detection for photovoltaic panels. In this work, the infrared image data are collected by infrared thermal imager from the view of unmanned aerial vehicle (UAV). A semantic segmentation neural network named Deep Res-UNet, which combines the strengths of residual learning, transfer learning, and U-Net, is proposed for infrared image segmentation. Residual units are applied in both the encoding and decoding path, which makes the whole deep network ease to train. A modified ResNet-34 with pre-trained weights is utilized to get better feature representation. In the modified ResNet-34, maxpooling layer is removed for reducing the loss in resolution, an additional conv1 stage is added to copy features for the corresponding decoding path and the skip block with dilated residual block is added to generate features with larger resolution and larger receptive field. A new loss function combining Binary Cross-Entropy (BCE) and Dice is proposed to get better results. Additionally, Conditional Random Field (CRF) is integrated into the model as a post-process. The experimental results show that the prediction results of the proposed model are **97.11%** and **94.47%** respectively on the two evaluation indexes of F_1 and Jaccard index, which is better than FCN-8s, SegNet, U-Net and two descendants of U-Net and ResNet: ResUnet and ResNet34-Unet.

Keywords: Infrared image · Semantic segmentation · Photovoltaic panels · U-Net.

1 Introduction

Infrared Thermography (IRT) plays a vital role in monitoring and inspecting thermal defects of photovoltaic panels. It has many advantages such as non-contact detection, safety, reliability and providing broad inspection coverage. The unmanned aerial vehicle (UAV) equipped with infrared thermal imager inspects the solar panel group overhead, getting infrared images of the photovoltaic plate area. The limitation of the infrared thermal imager, the flight height of UAV and other factors will result in the low-resolution photos which are hard for the human view. How to analyze these infrared images automatically is the

[⋆] Student

key to detect the fault area of solar panels efficiently and accurately. Infrared image segmentation is a fundamental and challenging problem in IRT for its over-centralized distributions and low-intensity contrasts.

A variety of methods have been proposed to segment infrared images in recent years. These algorithms can be categorised into threshold based [13, 5], textural analysis [8, 14], fuzzy based [23], region based [12, 21]. The key of the threshold based methods is to choose a appropriate threshold. J.N. Kapur [13] adopted the entropy of the histogram to get grey-level picture thresholding. Chen J. [5] used the gray level and local entropy information to construct a new tow-dimension histogram to achieve the segmentation. Edge, color, texture, and motion are integrated to segment objects within an image [8, 14]. Wu, J. [23] utilized fast fuzzy c-means with spatial information to segment infrared images. Region-based segmentation algorithms can be classified into region growing and region division and merge. According to [12], region-based methods are better than other algorithms for image segmentation. For infrared image segmentation, a region of interest (RoI) extraction method based on gray level co-occurrence matrix (GLCM) characteristic imitate gradient proposed in [21], and it achieves better performance than grab cut. In [27], a FAsT-Match algorithm is proposed for infrared images segmentation of electrical equipment, getting comparable result with other traditional methods. However, these methods can only segment images based on low-level semantic information, which perform worse and often cause over-segmented.

Recent years have witnessed the broad application of deep learning in the field of computer vision and image processing, such as image classification [9], object detection [25], and image denoising [2]. The powerful feature extraction ability of convolutional neural network has also been applied in semantic segmentation [17, 20, 1]. Semantic segmentation can be treated as a pixel-level classification problem, labeling each pixel in the image with a specific category. The common pipeline of semantic segmentation can be divided into an encoder and a decoder, where the encoder is responsible for feature extraction for the classification of pixels while the decoder is utilized to restore information lost in the encoder process. The encoder could be a reduced classification model, like VGG-16 [22] or ResNet [11], which has excellent feature representation ability. The decoder often consists of several up-sampling layers or deconvolution layers, recovering the resolution of downsampled features. Long *et al.* utilizes FCN [17] (fully convolutional network) which replaced the fully connected layer with convolutional layer, achieving pixel-level segmentation. Chen *et al.* combined the FCN with a conditional random field (CRF) [15] to extract a DeepLab [7] model, using FCN to segment the image and optimizing the result through CRF. Yu *et al.* [24] introduced dilated convolutions into the segmentation pipeline instead of pooling layer to realize multi-scale contextual information for better accuracy.

Lots of works have suggested that deeper networks would bring better performance [22]. However, it is very diffcult to train a very deep network for the emergence of problems such as vanishing gradients with the increasing of layers. He *et al.* [11] introduced the deep residual learning framework that employs an

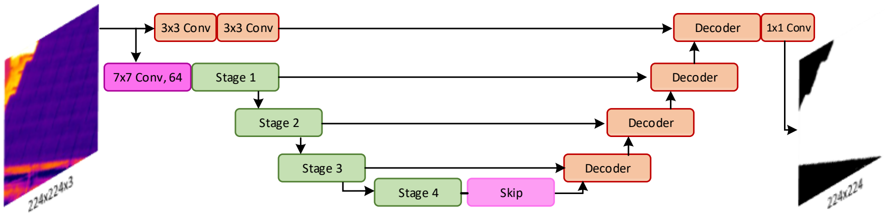


Fig. 1. The pipeline of our proposed Deep Res-UNet.

identity mapping. Instead of using skip connection in fully convolutional networks [17], Ronneberger *et al.* proposed that which combines different level of semantic information to get a finer result for medical image segmentation. To achieve this, U-Net copies low-level features to the corresponding high levels through concatenation operation.

In this work, we inherit the benefit of the U-Net [20] architecture with improvement by substituting the plain unit with multiple residual units [11]. We propose a deep residual U-Net, named Deep Res-UNet, an architecture that takes advantage of strengths from both deep residual learning and U-Net architecture. A modified ResNet-34 is utilized to extract features in the encoding path, and residual block is adopted in decoding path to ease the training of deep networks. We remove the maxpooling layer in the original ResNet-34 to keep a larger resolution, and a skip block with dialted residual unit is used to enlarge the receptive field while keep the resolution simultaneously. A new loss function which combines the Binary Cross-Entropy and Dice coefficient is proposed in this work. Additionally, Conditional Random Field (CRF) is integrated into the model as a post-processing to achieve a better boundary recovery. The experimental results show that the results of the proposed model are 97.11% and 94.47% respectively on the two evaluation indexes of F_1 and Jaccard index, which is better than FCN-8s [17], SegNet [1], U-Net [20], and two variants of U-Net: ResUnet [26] and ResNet34-Unet [3].

2 Method

2.1 The Structure of Proposed Deep Res-UNet

The proposed Deep Res-UNet (Figure 1 and Table 1) in this paper was designed based on ResNet [11], which has shown excellent performance in image classification task, and has been applied in many tasks. ResNet with a series of stacked residual blocks is powerful enough to extract features and strength the feature propagation during training and testing. Meanwhile, encouraged by the excellent performance the symmetrical structure of U-Net [20] for biomedical image segmentation, we utilize a multi-stage architecture of Deep Res-UNet for infrared image segmentation with two parts (Figure 1). The first part (encoding part) is designed to extract the features using the modified ResNet-34, also known as the ”contracting” path in U-Net. The second part (decoding part, or ”expansive” path in U-Net) is utilized to generate the classification map using the extracted features at different stages of the encoding part. The proposed Deep Res-UNet inherits both the benefits of ResNet and U-Net, and to get better feature representations from the encoding part, a modified ResNet-34 is utilized.

The conventional ResNet [11] starts with 7×7 convolutional layers with stride 2 and a max-pooling layer to down-sample the size of feature map for fast computation. Four stages with different numbers of residual blocks with two convolutional layers (shown in Figure 2(b)) are followed by, and the first convolutional layer of the first residual block in stage $2-4$ is with stride 2 while the other is 1. Specificly, ResNet-34 has $[3, 4, 6, 3]$ residual blocks and total 33 convolutional layers and 1 fully connected layer, so called ResNet-34. There are 5 downsampling operations in ResNet, and the size of the final feature maps is $32\times$ smaller than the original input size, *e.g.*, a 7×7 feature map is generated with the input of 224×224. We modified ResNet-34 as the Encoder part (as illustrated in Figure 1 and Table 1) from three aspects: 1) an additional `conv1` stage with two 3×3 convolution layers to keep the resolution of the input size instead of downsampling it directly with convolutional layers with a kernel of size 7×7 and stride 2; 2) removal of the maxpooling operation after the first convolution layer which will bring loss in information; 3) an additional skip block is attached after Stage 4 to expand the receptive field while keep the resolution.

Actually, the combination of ResNet [11] and U-Net [20] has two options: 1) replace the plain block of U-Net with residual block in ResNet ; 2) replace the Encoder part of U-Net with a reduced ResNet . Zhang *et al.* [26] proposed the combination of residual learning and U-Net with replacing the plain block of U-Net with residual units for road extraction. Buslaev *et al.* [3] adopted ResNet-34 pre-trained on ImageNet [9] and decoder adapted from vanilla U-Net. Besides these two options, we inherit both the benefits of the two choices: the residual blocks ease training of deep networks in both the Encoder and Decoder parts; the pre-trained ResNet-34 has more powerful feature extraction ability.

As illustrated in Figure 1, the Encoder consists of both plain block, residual block and dilated residual block, which are shown in Figure 2 (a)-(c). These three types of blocks allow us to effectively capture the multi-scale context, where the plain block in the first layer can extract the shallow information of the input image, the residual blocks capture the semantic information through deep stacked network, and the last skip layer with dilated convolution can effectively enlarge the field of view of filters to incorporate deep semantic information without downsampling the size of feature maps. The Encoder part encodes the input image into compact representations.

The right part of Figure 1 and Figure 2 (d) have shown the Decoder of the proposed Deep Res-UNet. Low-level features are upsampled first and then concated with feature maps from the corresponding encoding path. The plain block of the vanilla decoder of U-Net is replaced by the residual block. The last stage of the decoding path, a 1×1 convolution layer and a sigmoid activation

Table 1. Network Structure

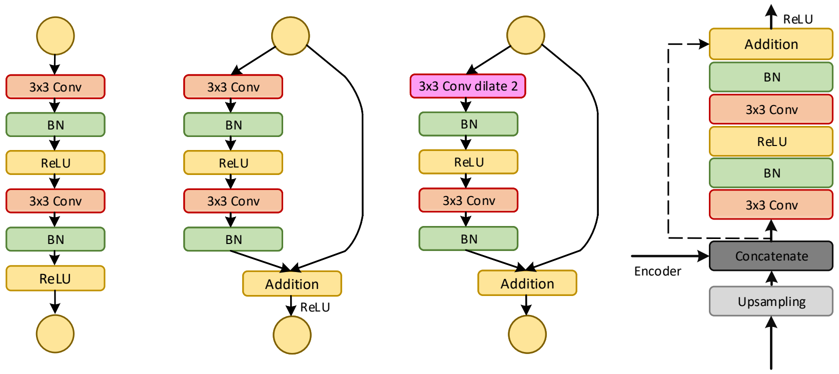
layer name	output size	Encoder	Decoder
conv1	224×224	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 1$	$1 \times 1, 64$
Stage1	112×112	$7 \times 7, 64$ stride 2 $\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$
Stage2	56×56	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$
Stage3	28×28	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$
Stage4	14×14	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$
Skip	14×14	$[3 \times 3, 512]$, dilation 2	

layer is utilized to project the multi-channel feature maps into the desired segmentation result. The Decoder part recovers the representations to a pixel-wise classification, i.e., semantic segmentation.

2.2 Residual Block

As illustrated in Figure 2 (a)(b), U-Net [20] adopted the plain blocks with two 3×3 convolutional layers with batch normalization (BN) and ReLU layers, and the residual blockd with skip connection can be stacked to a very deep network. We replace the original plain block in U-Net with residual block in both the Encoder and the Decoder parts. In the encoding path, the modified ResNet-34 pre-trained on ImageNet [9] has powerful feature extraction ability than randomly initialized.

More importantly, a dilated residual block is proposed to fix the removal the maxpooling in the conventional ResNet-34 [11]. Dilated convolution, also known as atrous convolution, has been demonstrated effective in semantic segmentation [6, 24, 7]. Dialted convolution allows us to repurpose ImageNet [9] pre-trained models to extract denser feature maps by removing teh downsampling operation, which can control the resolution wihtout requiring learning extra parameters. Actually, Li *et al.* [16] pointed out that there is gap between the image classification and object detection, so large downsampling factor in image classification networks brings large valid receptive field which is good for image classification but compromises the object location ability. As for semantic segmentation, pre-trained ResNet [11] with 32× downsampling factor performs



(a) Plain Block (b) Encoder Block (c) Skip Block (d) Decoder Block

Fig. 2. Basic blocks of the Encoder & Decoder.

well for extracting semantic features for classification, but it harms pixel-level classification or segmentation for the downsampled information can not be recoveried accurately through up-sampling operation. So we modified the original ResNet-34 with the removing of maxpooling layer and adding a dialted residual block, named Skip Block.

As illustrated in Figure 2 (c), the Skip Block is based on the residual block in ResNet-34 with simply replacing the first convolutional layer with dialted convolution with dilated rate 2. The dilated convolution will enlarge the receptive field of the filters without downsampling the resolution.

2.3 Loss Function

Semantic segmentation can be treated as pixel-level classification task, so it often adopts cross entropy loss or mean square error (MSE) as loss function. In this paper, we use pixel-wise cross entropy as loss function following U-Net [20], which can be formulated as:

$$L(g,p) = -\sum_i g_i \log p_i \tag{1}$$

where i is the categor y, specificly $(0,1)$ in our task, p_i is the predicted label, and g_i is the true label. The loss caculate the result of every pixel averagely without considering the imbalance problem between classes.

Dice coefficient is often used to evaluate the performance of sementation, representing for the overlapping between two samples.

$$D(G,P) = \frac{2|G \cap P|}{|G| + |P|} \tag{2}$$

where G is the true object region, and P is the predicted object region. The Dice coefficient can be generalized in 2D segmentation[4, 10], where predict label

$p_i \in 0, 1$, soft Dice can be formulated as:

$$D_{soft}(g, p) = \frac{2\sum_i p_i \times g_i}{\sum_i p_i + \sum_i g_i} \tag{3}$$

where $p_i \in [0, 1]$ is the value after sigmoid function, $g_i \in [0, 1]$ is the true label. We combine the Dice coefficient and Binary Cross-Entropy (BCE) Loss as blew:

$$L(g, p) = (1 - \beta)(\sum_i g_i \log(p_i)) - \beta \log(\frac{2\sum_i p_i \times g_i + smooth}{\sum_i p_i + \sum_i g_i + smooth}) \tag{4}$$

where *smooth* is often an extreme small number, e.g. $1e - 15$ to prevent from dividing by zero. β is set as 0.75 empirically.

2.4 Post Processing

For accurate boundary recovery, fully connected Conditional Random Field (CRF) is integrated into our system as a post-process from [15]. The model employs the energy function

$$E(x) = \sum_i \theta_i (x_i) + \sum_{ij} \theta_{ij} (x_i, x_j) \tag{5}$$

where x is the label assignment for pixels. We use as unary potential $\theta_i (x_i) = -\log P (x_i)$, where $P (x_i)$ is the label assignment probability at pixel i as computed by a DCNN. The pairwise potential has a form that allows for efficient inference while using a fully-connected graph, i.e., when connecting all pairs of image pixels, i; j. In particular, as in [15], we use the following expression

$$\theta_{ij} (x_i, x_j) = \mu (x_i, x_j) \left[w_1 \exp \left(-\frac{\|p_i - p_j\|^2}{2\sigma_\alpha^2} - \frac{\|I_i - I_j\|^2}{2\sigma_\beta^2} \right) + w_2 \exp \left(-\frac{\|p_i - p_j\|^2}{2\sigma_\gamma^2} \right) \right] \tag{6}$$

where $\mu(x_i, x_j) = 1$ if $x_i \neq x_j$, and zero otherwise, which, as in the Potts model, means that only nodes with distinct labels are penalized. The remaining expression uses two Gaussian kernels in different feature spaces; the first, *appearance kernel* depends on both pixel positions (denoted as p) and RGB color (denoted as I), and the second kernel, *smoothness kernel* only depends on pixel positions. The degrees of nearness and similarity are controlled by parameters θ_α and θ_β.

3 Results

3.1 Dataset and Augmentation

The image data are collected from Jiangsu LINYANG Power Station, located at the No.666, LINYANG Road, Qidong Economic Development Area, Jiangsu

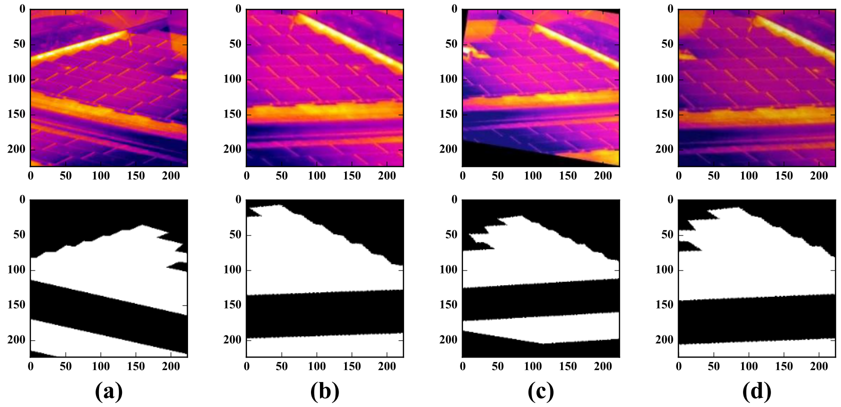


Fig. 3. Data augmentation with random crop (**a**), horizontal flip (**b**), shift scale rotate (**c**), random brightness contrast (**d**).

Province 226220, China, with the coordinate position on the map of (121.639278, 31.817825). The imaginary thermal instrument adopted here is DM63 series, manufactured by Zhejiang Dali Technology Co., Ltd... The images were shot in the morning, high noon and evening respectively from Jul. to Aug. 2016, when sunny and cloudy weather. The pixel of the adopted thermal imager instrument is 320 × 240. The training set involves 216 images and the testing set is 19. All images are annotated manually with two kinds of pixels: including ”panels” and ”others”.

To prevent overfitting and make the model more robust, we use some data augmentation methods when training. These methods include random crop, horizontal flip, shift scale rotate, random brightness contrast, as shown in Figure 3. All these methods are combined only when training for all experiments in this paper.

3.2 Experimental Setup and Results

The proposed model was implemented using PyTorch [19]. The encoding part (Stage 1-4) of our deep Res-UNet is initialized from the pre-trained model on ImageNet[9], and the other part is initialized randomly. The initial learning rate is 0.0001, and SGD is used to optimize the model. We start training the model with a minibatch size of 16 on an NVIDIA GTX 1080 GPU with 8 GB onboard memory, Intel(R) Xeon(R) CPU E5-2683 v3 @2.00GHz.

Evaluation Metrics To assess the quantitative performance of the proposed Deep Res-UNet for infrared image segmentation, the precision (P) and recall (R) are introduced, as well as the F_1 score [18] and Jaccard index. The F_1 score is calculated by P and R, which is an evaluation metric for the harmonic mean of P and R, and it can be calculated as follows:

$$F_1 = 2 \times \frac{P \times R}{P + R} \tag{7}$$

where

$$P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN} \tag{8}$$

Here, TP, FP and FN represent for the number of ture positives, false positives and false positives, respectively. P measures the proportion of matched pixels in the predicted results and R is the percentage of matched pixels in the ground truth. Jaccard index (Intersection Over Union) measures the similarity of the predicted result and the ground truth, and it can be calculated as:

$$J(Gr, Pr) = \frac{|Pr \cap Gr|}{|Pr| + |Gr| - |Pr \cap Gr|} \tag{9}$$

where Gr is the true object region, and Pr is the predicted object region.

Results As shown is Table 2, our proposed Deep Res-UNet achieves a performance of 96.30%, 98.03%, 97.11%, and 94.47% with the highest in R, F_1, and Jaccard metrics under $\beta = 0.75$. Compared to the original U-Net [20], the Deep Res-UNet gets 1.06 and 1.82 points in F_1 and Jaccard respectively. ResUnet [26] got the highest P of 97.25% also the lowest R of 94.29%, F_1 of 95.54% and Jaccard of 91.83% mainly for it only has 15 convolutional layers which has weak ability to extract features. As for ResNet34-Unet [3], it performs quite better compared to U-Net and ResUnet mainly due to the pre-trained ResNet34 [11]. To demonstrate the pre-trained model helps extract features, we trained our model without per-trained on ImageNet [9], we received a performance of 96.04%, 95.76%, 95.90%, and 91.84% in P, R, F_1, and Jaccard, even worse than U-Net.

We visualised some of the results of FCN-8s[17], SegNet[1], UNet [20], ResUnet [26], ResNet34-Unet [3], and our Deep Res-UNet in Figure 4. All the output of these models are post-processed with conditional random field (CRF) to get a better boundary. As shown in the second row, our proposed Deep Res-UNet performs better in the area with complex shapes, which is mainly because of the pre-trained model of the deep network. Similarly, ResNet34-UNet has shown the same signature mainly for the same reason. After the post-process of CRF, all models can detect the main area of photovoltaic panels correctly.

4 Conclusions

In this work, we propose Deep Res-UNet for segmentation of UAV-based infrared images for photovoltaic panels. Infrared images are collected by the UAV equipped with infrared thermal imager, inspecting the solar panel group overhead. The proposed network combines the strengths of residual learning and

Table 2. Experimental Comparison

Method	β	Precision	Recall	F1	Jaccard
FCN8s [17]	0	95.64	95.85	95.43	92.01
SegNet [1]	0	95.84	97.07	96.35	93.17
UNet [20]	0	96.11	96.30	96.05	92.65
ResUnet [26]	0	**97.25**	94.29	95.54	91.83
ResNet34-UNet [3]	0	95.52	97.21	96.24	92.94
Ours w/o	0	96.04	95.76	95.90	91.84
Ours	0	96.34	96.59	96.25	93.10
Ours	0.75	96.30	**98.03**	**97.11**	**94.47**

U-Net. Both the Encoder and Decoder are constructed by residual units, which make the whole deep network easy to train. ResNet-34 with modifications is utilized to learn the feature representations. The modifications including: the addition of conv1 stage for copying features for the original input size, the removal of maxpooling layer for remaining a larger resolution, and the adding of Skip block with dilated residual unit to enlarge the received field. To get a finer result of the network, a modified loss function, which inherits both Binary Cross-Entropy (BCE) and Dice is utilized, and we reached the best result of our model at $\beta = 0.75$. Besides, Conditional Random Field (CRF) is integrated into the segmentation pipeline as a post-process. Experiments in this work demonstrate the superiority of our proposed Deep Res-UNet, which exceeds FCN-8s [17], SegNet [1], U-Net [20] and its variants ResUnet [26] and ResNet34-Unet [3].

References

1. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence **39**(12), 2481–2495 (2017)
2. Burger, H.C., Schuler, C.J., Harmeling, S.: Image denoising: Can plain neural networks compete with bm3d? (2012)
3. Buslaev, A., Seferbekov, S.S., Iglovikov, V., Shvets, A.: Fully convolutional network for automatic road extraction from satellite imagery. In: CVPR Workshops. pp. 207–210 (2018)
4. Chang, H.H., Zhuang, A.H., Valentino, D.J., Chu, W.C.: Performance measure characterization for evaluating neuroimage segmentation algorithms. Neuroimage **47**(1), 122–135 (2009)
5. Chen, J., Guan, B., Wang, H., Zhang, X., Tang, Y., Hu, W.: Image thresholding segmentation based on two dimensional histogram using gray level and local entropy information. IEEE Access **6**, 5269–5275 (2018)

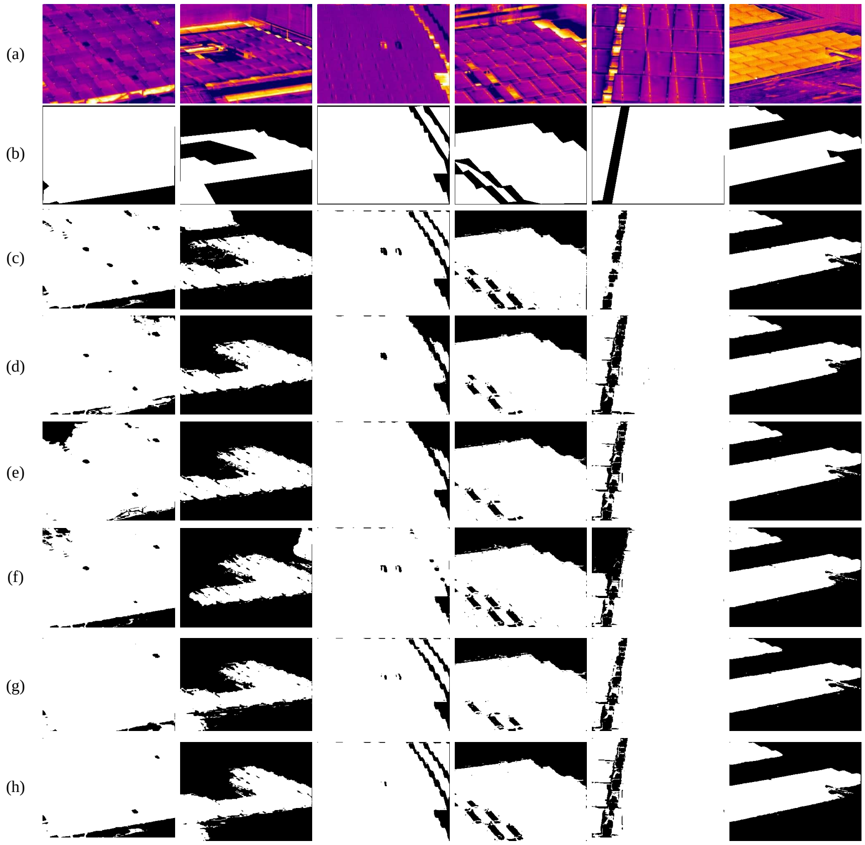


Fig. 4. Some visulized results of different models. (a) Image; (b) Groun-truth; (c) FCN-8s [17]; (d) SegNet [1]; (e) U-Net [20]; (f) ResUnet [26]; (g) ResNet34-UNet [3]; (h) Our Deep Res-UNet.

6. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Semantic image segmentation with deep convolutional nets and fully connected crfs. arXiv preprint arXiv:1412.7062 (2014)
7. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence **40**(4), 834–848 (2018)
8. Cremers, D., Rousson, M., Deriche, R.: A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. International journal of computer vision **72**(2), 195–215 (2007)
9. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. IEEE (2009)
10. Fidon, L., Li, W., Garcia-Peraza-Herrera, L.C., Ekanayake, J., Kitchen, N., Ourselin, S., Vercauteren, T.: Generalised wasserstein dice score for imbalanced

multi-class segmentation using holistic convolutional networks. In: International MICCAI Brainlesion Workshop. pp. 64–76. Springer (2017)

11. 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. pp. 770–778 (2016)
12. Jaffery, Z., , I.: Performance comparison of image segmentation techniques for infrared images (12 2015). https://doi.org/10.1109/INDICON.2015.7443391
13. Kapur, J.N., Sahoo, P.K., Wong, A.K.: A new method for gray-level picture thresholding using the entropy of the histogram. Computer vision, graphics, and image processing **29**(3), 273–285 (1985)
14. Kaur, S., Kaur, P.: An edge detection technique with image segmentation using ant colony optimization: A review. In: 2016 Online International Conference on Green Engineering and Technologies (IC-GET). pp. 1–5. IEEE (2016)
15. Krähenbühl, P., Koltun, V.: Efficient inference in fully connected crfs with gaussian edge potentials. In: Advances in neural information processing systems. pp. 109–117 (2011)
16. Li, Z., Peng, C., Yu, G., Zhang, X., Deng, Y., Sun, J.: Detnet: Design backbone for object detection. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 334–350 (2018)
17. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3431–3440 (2015)
18. Martin, D.R., Fowlkes, C.C., Malik, J.: Learning to detect natural image boundaries using local brightness, color, and texture cues. IEEE Transactions on Pattern Analysis & Machine Intelligence (5), 530–549 (2004)
19. Paszke, A., Gross, S., Chintala, S., Chanan, G.: Pytorch: Tensors and dynamic neural networks in python with strong gpu acceleration. PyTorch: Tensors and dynamic neural networks in Python with strong GPU acceleration **6** (2017), https://pytorch.org/
20. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
21. Shen, H., Zhu, L., Hong, X., Chang, W.: Roi extraction method of infrared thermal image based on glcm characteristic imitate gradient. In: CCF Chinese Conference on Computer Vision. pp. 192–205. Springer (2017)
22. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
23. Wu, J., Li, J., Liu, J., Tian, J.: Infrared image segmentation via fast fuzzy c-means with spatial information. In: 2004 IEEE International Conference on Robotics and Biomimetics. pp. 742–745. IEEE (2004)
24. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122 (2015)
25. Zhang, H., Hong, X.: Recent progresses on object detection: a brief review. Multimedia Tools and Applications (Jun 2019). https://doi.org/10.1007/s11042-019-07898-2
26. Zhang, Z., Liu, Q., Wang, Y.: Road extraction by deep residual u-net. IEEE Geoscience and Remote Sensing Letters **15**(5), 749–753 (2018)
27. Zou, H., Huang, F.: Infrared image segmentation for electrical equipment based on fast-match algorithm. Infrared Technology **38**(1), 21–27 (2016)