SEGMENTATION OF ELECTRICAL SUBSTATIONS USING DEEP CONVOLUTIONAL NEURAL NETWORK

M.Mesvari[1], R.Shah-Hosseini [1*]

[1] School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran,
(m.mesvari, rshahosseini)@ut.ac.ir

Commission IV, WG IV/3

KEY WORDS: Segmentation, Encoder-Decoder Convolutional Neural Network, Electrical substations, High-resolution satellite images.

ABSTRACT:

The location of electrical substations is one of the factors affecting the improvement of electrical energy distribution, as well as the management and control of this energy source. Less cost and manpower will be spent through automating the process of detection and segmentation of these features with the help of deep neural networks and the potential of existing high spatial resolution satellite images. In this study, a deep encoder-decoder neural network was used. This network is one of the most updated deep learning methods in image processing and segmentation. This network has been trained in three RGB bands with the help of high-resolution satellite images (~1m) and eventually segmented the areas related to electrical substations with relatively high accuracy. As the results of this convolutional neural network, the IOU and Precision parameters were obtained, and their values were 88.2 and 93.7%, respectively, indicating the efficiency of the proposed deep learning method in the segmentation of existing satellite images.

1. INTRODUCTION

Today, the increase in energy consumption makes it necessary to study, optimize, and develop energy sources. Electrical substations as an energy source can play a key role in the supply of electrical energy. Thus, the location of such substations can be considered a fundamental step in the future planning of electrical energy distribution and the management and control of these energy sources. High-resolution satellite images and deep learning (LeCun, Bengio, and Hinton 2015) techniques present the context for detecting and segmenting (Minaee et al. 2021; Ghosh et al. 2019) the areas related to electrical substations.

As Remote Sensing acquisition techniques develop rapidly, data volume is increasing, requiring scaling and improving processing tools. Nowadays, Deep learning has become a popular technique in remote sensing, but it depends on the quality of the training data (Kampffmeyer, Salberg, and Jenssen 2016).

Due to the extensive uses of deep neural networks, detecting and segmenting electrical substations can be conducted automatically, and the results with high accuracy can be obtained. This can be regarded in the process of reducing detection errors. Nowadays, the cost and time can be reduced due to the potential of high-resolution satellite images compared to photogrammetric imaging.

Segmenting objects needs both object-level data and low-level pixel data. Feedforward networks face the challenge of capturing spatial information in the lower layers of convolutional networks; object-level knowledge is encoded in upper layers, subject to certain factors such as pose and appearance, but not affected by these. A new top-down refinement approach is proposed to augment feedforward nets for object segmentation (Pinheiro et al. 2016). Using this bottom-up/top-down architecture, high-fidelity object masks can be generated efficiently. Sharp Mask is 50% faster than the original Deep Mask network due to optimizing its overall architecture. Therefore, their model achieved new levels of performance and speed for generating object proposals.

Regarding pixel labeling, the proposed general refinement approach could be applied to many more applications.

In 2018, (Kemker, Salvaggio, and Kanan 2018) showed the efficiency of FCN architectures for semantic segmentation of remote sensing Multispectral imagery. They demonstrated that learning global relationships between objects is more efficient by combining convolution and pooling operations in an end-to-end segmentation model. The researchers in this paper found that supervised DCNN frameworks provide better classification performance than unsupervised learning methods in fourteen out of eighteen classes in RIT-18. There is a high cost and manpower requirement to obtain multispectral imagery (Wang et al. 2017) (MSI), so it is unavailable. So the paper adapted state-of-the-art DCNN frameworks using MSI imagery for semantic segmentation, and in order to initialize a DCNN framework, they replaced the generated synthetic MSI with true MSI data. The resolution of these images is about 0.05 to 100m; according to this resolution, the number of classes is different.

(Yuan et al. 2021) presented a new DCNN model multichannel water body detection network (MC-WBDN). A multichannel fusion module, a spatial pyramid pooling module, and space-to-depth and depth-to-space operations were integrated in this model. In comparison to other DCNN models, the MC-WBDN model performed well in detecting water bodies regardless of light and weather conditions, and even detected tiny water bodies more accurately.

(Noh, Hong, and Han 2015) trained a deep deconvolution network to achieve semantic segmentation. They learned the network based on the convolutional layers adopted from the VGG16-layer net. In this paper, pixel-level(Li et al. 2018) class labels and predict segmentation masks were identified by deconvolution and un-pooling layers in the deconvolution network. Then, trained network is performed on each proposal in an input image and the results are combined simply to form the final semantic segmentation map. Unlike fully convolutional algorithms, their proposed algorithm integrated deep deconvolution and proposal-wise prediction. In this way, they

---

* Corresponding author

could also identify detailed structures and objects at multiple scales. They achieved the highest accuracy (72.5%) between these methods in the absence of Microsoft COCO dataset by the ensemble with the fully convolutional network.

(Long, Shelhamer, and Darrell 2015) has shown that when convolutional networks are trained end-to-end and pixels-to-pixels can achieve superior results in semantic segmentation. By building fully convolutional networks, they could produce correspondingly-sized output by taking inputs of arbitrary sizes. This paper presents a detailed description of the space of fully convolutional networks, their application to spatially dense prediction tasks, and draws connections with previous work. By fine-tuning their learned representations to segmentation tasks, they adapted contemporary classification networks (AlexNet(Yu et al. 2016), VGG(Simonyan and Zisserman 2014), and Google Net) into fully convolutional networks. To produce accurate and detailed segmentations, they developed a skip architecture. This study achieved state-of-the-art segmentation of PASCAL VOC by applying a fully convolutional neural network.

(Hamida et al. 2017) new Deep Learning methods for estimating tiny or coarse semantic segmentation of land cover. The resolution of these images varies from 20m to 300m. Different 2D architectures were examined to process the spatial and spectral dimensions of the data together.

When dealing with smaller datasets, optimizing this kind of neural network is challenging due to the noise in the reference land cover features. The best solution is to estimate coarse image scales up to the reference image scale. Higher resolution estimation is challenging during the training and validation steps to enable reliable comparisons. Developing multiscale estimation capabilities for each proposed model would be a first step forward.

(Prathap and Afanasyev 2018) propose a deep learning approach using numerous enhancements throughout the process of building detection (Pan et al. 2020), which is a challenging problem. A 2-sigma percentile is applied to the initial dataset to normalize it. Data preparation includes ensemble modeling, incorporating Open Street Map data into three models. By adding batch normalization wrappers to the U-Net(Shi et al. 2021; Pan et al. 2020), the Binary Distance Transformation (BDT) improves the data labeling process. The analysis of a multispectral image of Las Vegas, Paris, Shanghai, and Khartoum shows that this solution can enhance building detection accuracy compared to the winning solutions of SpaceNet2. Selecting the training dataset with the highest variance and using unsupervised deep learning approaches can overcome these issues.

In 2020, (Tong et al. 2020) proposed a deep model derived from labeled land cover datasets that can be used to classify unlabeled high spatial resolution images. A pseudo-labeling and sample selection scheme was proposed to improve the transferability of deep models by using deep neural networks for presenting contextual information about different types of land covers. An annotated land cover dataset was also used as the source data to pre-train deep Convolutional Neural Networks (CNNs). Then, a CNN is trained to classify images patch-wise using a target image without labels. In order to retrieve related samples from the source data, they assigned pseudo-labels to patches with high reliance. Pre-trained deep models were fine-tuned based on pseudo-labels that were confirmed with the retrieved results. A hybrid classification using patch-wise classification combined with hierarchical segmentation was developed by combining fine-tuned CNN and target image to classify pixel-level land cover.

U-Net was proposed in 2020 by (Zheng and Chen 2021) for the multi-classification and binary classification of high spatial resolution images from Gaofen-2. There are six different types of features labeled in this image: farmland, water, meadows, forests, and more. U-Net neural network multi-classification obtained an overall accuracy of 93.83% for the training data, 82.27% for the test data, and 797.5% for the binary classification algorithm. It was shown that both models were highly accurate at segmenting remote sensing images.

According to the weak feature representation ability and noise information that can affect the segmentation performance, (Liu et al. 2021) proposed an Adaptive Multi-Scale Module (AMSM) and Adaptive Fuse Module (AFM) to solve these problems, respectively. They improved target identification while maintaining overall accuracy with the network of two relationship modules, which was tested on the Vaihingen dataset.

Encoder-decoder models have two problems. The first one is structural stereotyping, which results from imbalances in the receptive fields within the frameworks. Another problem is insufficient learning, which is prevalent in deeper neural networks. These problems are independent of each other. random sampling training (RST) and ensemble inference (EI) methods were proposed by (Sun, Tian, and Xu 2019) in order to mitigate the adverse outcomes of the first problem. These methods improved the overall accuracy. Finally, they used a residual topology to solve insufficient training.

This paper represents an encoder-decoder decoder (Ji et al. 2021) convolutional neural network method to detect and segment the electrical substations in high-resolution satellite images. The dataset, proposed method and evaluation metrics are described in section 2. In section 3, the results obtained are detailed. Finally, in section 4, the conclusion of this research and the important factors to improve the results are described.

2. MATERIAL AND METHOD

2.1 Dataset

The dataset used in this study involved high-resolution satellite images (~1m) in three spectral bands; each image has 750 rows and columns. Each image includes an electrical substations feature. The feature is a target that the used algorithm should detect. The labelling process in this study was conducted in arc map software using shapefiles related to the features in the images. This dataset was already discussed in the ICETCI 2021 competition that the areas related to electrical substations should be detected using machine learning algorithms(Mahesh 2020).

This dataset can be downloaded from the coda lab competition website. Because of the limited number of these images (in the first phase, the images related to the training process were uploaded on the related website), data augmentation (Van Dyk and Meng 2001; Shorten and Khoshgoftaar 2019) methods were used to increase the number of available data, including the vertical, horizontal and horizontal-vertical flip. Finally, these images were resized and entered into the convolutional neural network as input. An example of such images is shown in Figure 1.
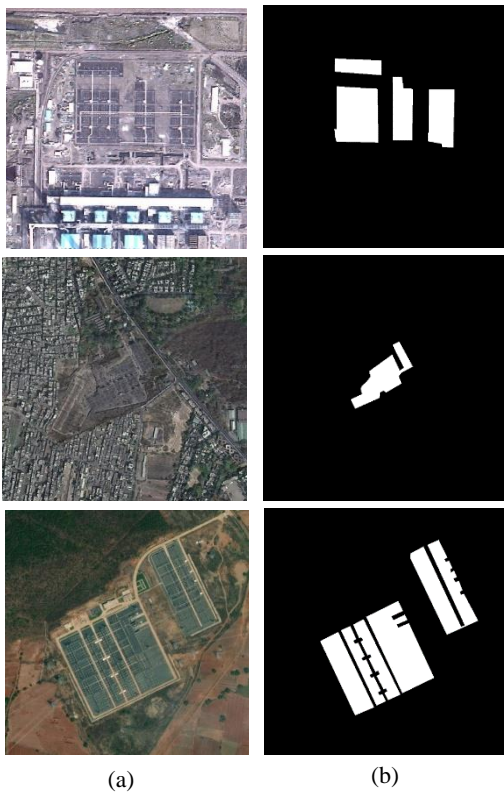
(a)   (b)

Figure 1. Illustrate a) image and b) ground truth

2.2 Methodology

In this study, the encoder-decoder neural network was used due to its basic application in image processing. Such networks have found numerous applications in photogrammetry and remote sensing in extracting information from images.

The structure of this deep learning network was created by using two-dimensional convolutional blocks. The general design of this network was conducted according to the ideas of U-Net and SegNet networks (Badrinarayanan, Kendall, and Cipolla 2017).

Every two-dimensional convolutional block was composed of two convolutional layers, each of which was a convolutional neuron enabling this block to extract more features. The existing kernel size was considered to be 3x3 in each of these convolutional blocks.

In addition, batch normalization was used to prevent the overfitting process in this convolutional neural network. The dimensions of input and output images in this convolutional neural network were the same.

Each convolutional neuron in this block required an activation function.

Like many other studies, ReLU activation function was used in this study due to its functional features. Here are two significant features of this activation function:

1. This function considers the negative input values as zero.

Since the probability of negative values is very low in image processing, this feature significantly helps to avoid the error caused by negative values.

2. The derivative of this activation function for positive values is equal to 1.

If an activation function is applied that the derivative in positive values is opposite to one, the output values may be regarded as out of the intended range and cause an error in the convolutional network.

In this network, an encoder path extracts the information from existing images. In the decoder path, we should retrieve such extracted features.

Ultimately, such features were reconstructed, and the final map (involving the target and non-target classes) was associated with a loss function. In the encoder path, five convolutional layers were considered, and convolutional blocks were used in each block.

Regarding the available images and features (electrical substations), the number of target pixels was less than the background classes, and a kind of imbalanced problem was observed.

A weight loss function is used to solve this problem. The production of an activation function is one of the subjects in charge of activating and deactivating the neurons. Indeed, it plays a key role in updating the weights and the process of derivation and gradient of weights.

Hence, selecting an appropriate activation function can prevent network overfitting and act well in solving nonlinear problems and base pixel patterns. The loss function used in this network was binary-cross entropy (Ruby and Yendapalli 2020), giving a weight according to the target and non-target pixels in the model.

The activation function used in this layer was soft-max with good capability in segmentation.

In the decoder path, the extracted features were reconstructed in four decoding layers to present a high-level feature representation of the original image.

In the end, the last function used could pass the network outputs to place the extracted features as binary into two classes.

The decoder path is followed by a final pixel-wise segmentation layer. The high-level feature representation at the output of the final decoder is fed to a trainable soft-max classifier (Qi, Wang, and Liu 2017). For each pixel, the predicted segmentation corresponds to the class with the highest probability.

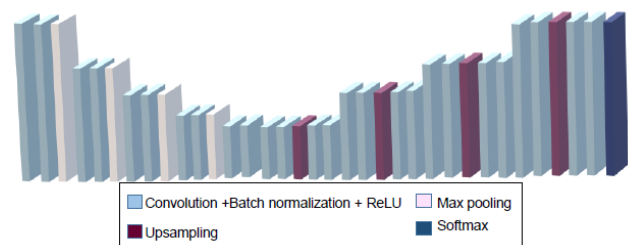The architecture of this encode decoder convolutional neural network is shown in Figure2.

Figure 2. Architecture of encoder-decoder convolutional neural network

2.3 Experiment

Google Colaboratory platform was used to implement the model. Additionally, libraries such as Tensorflow, Keras and OPENCV were used in this method. The training dataset uploaded on the coda lab competition website was used in the training process. These images were resized to 512×512 resolution. 10% of these images (not used in the training process) were used to test the model. The batch size set is 10, and the number of epochs is 350. The time of the training process is about 6.5 hours. After training the model, the model was tested on the test data. The outline of this study is shown in Figure 3.
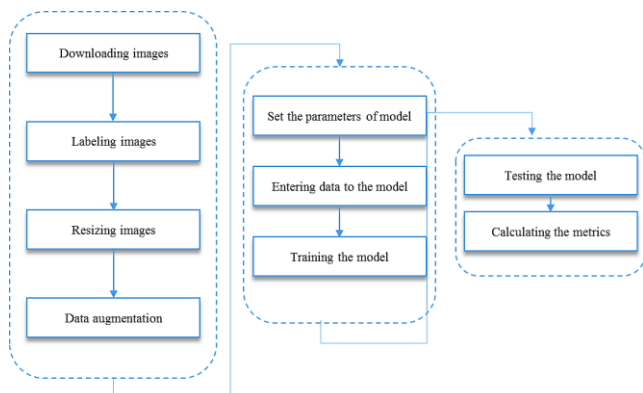

Figure 3. Flowchart of the study

2.4 Evaluation metrics

Participant	IoU
1	0.858
2	0.851
3	0.842
4	0.837
5	0.836
Our proposed method	0.882

According to the results obtained from the method used in this paper, the evaluation metrics are: Precision, Recall, IoU, F1Score and Kappa

Precision is calculated as the proportion of relevant instances that appear among the retrieved instances. Recall is the proportion of relevant instances that appear among the retrieved instances. Therefore, these two parameters are based on relevance. These metrics are calculated as follows:

Table 2. Comparison between participant's methods and the proposed method

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

When the ground truth and predicted points have the same label, it is known as a true positive (TP). False positives (FP) are points that are predicted to be positive but have a negative label.

False negatives (FN) refer to points that are predicted to be negative but have a positive label.

F1-Score incorporates precision and recall measurement in

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (4)$$

order to calculate its harmonic mean. F1-score can be defined as:

$$F1\ score = \frac{2 \times (Precision \times Recall)}{(Precision + Recall)} \quad (3)$$

IOU (Intersection-Over-Union) is a statistic used to compare sample sets' similarities by calculating the ratio of intersection and union. These two sets represent the prediction and the reference in image semantic segmentation. Based on the equation below, we can calculate the value of IoU.

That A is the prediction, while B is the ground truth.

As a statistical criterion, Kappa is commonly used in machine learning for classes with multiple or imbalanced data sets (De Raadt et al. 2019). Its maximum and minimum values are 1 and -1, respectively. It is indicated that a kappa value over 0.75 indicates an excellent agreement, whereas a value below 0.4 indicates a poor classifier.

$$K = \frac{P_0 - P_e}{1 - P_e} \quad (5)$$

That P0 is the relatively observed agreement between raters, and P0 is the hypothetical probability of agreement.

3. RESULT AND DISCUSSION

Based on the general performance of evaluation criteria of the proposed method, the values of Precision, Recall, F1-Score, IoU and Kappa were considered. These results indicate that the convolutional neural network performed well in detecting the electrical substations and can be used to detect and segment these features from satellite images. These results are shown in Table1.

Table 1. Result of the quantitative evaluation

METRICS	Deep convolutional neural network
Precision	93.7
Recall	93.73
IoU	88.2
F1 score	93.71
Kappa	87.4

The IoU obtained in this method was more than the highest value of the IoU determined in the competition mentioned above. According to the IoU values reported on the competition website, a comparison between the participant's methods is presented in Table2. In general, such an algorithm could visually detect the range related to these electrical substations. This can be effective in controlling and managing the existing energy resources, as well as future planning. Figure4 displays the results obtained from the deep learning algorithm.

| (a) | (b) | (c) |
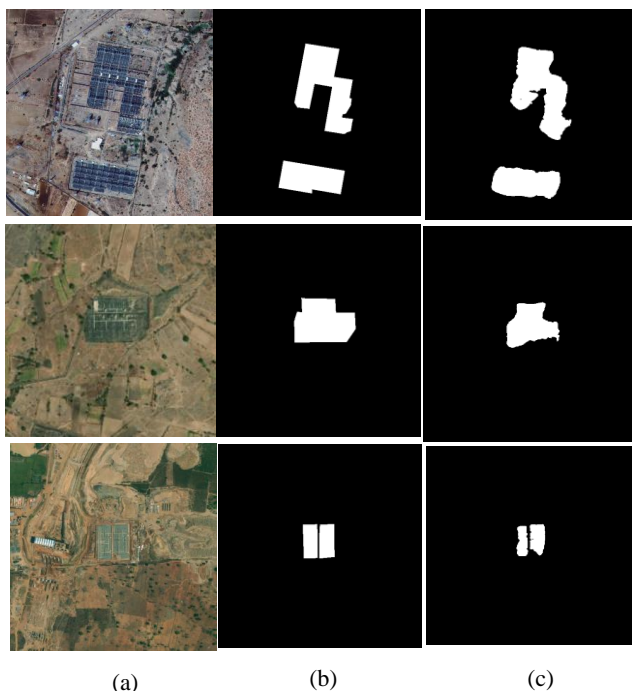
Figure 4. illustrate the results, a) RGB image, b) Ground truth, c) Predicted image

4. CONCLUSION

This study used the applications of an encoder-decoder neural network to detect and segment electrical substations. These areas were well detected visually and regarding the values related to evaluation criteria. The factors which can increase the accuracy of results involve the increased number of high-resolution satellite images to improve the training process of the convolutional neural network since the lack of satellite images including the desired feature was one of the challenges of this study. Furthermore, more reliable results can be achieved if the data related to other sources, such as three-dimensional layer data, are combined with these images. Although this convolutional neural network is one of the most updated methods in image segmentation, developing deep learning methods in the field of segmentation and detection can be a key step to improving these results.

REFERENCES

Badrinarayanan, Vijay, Alex Kendall, and Roberto Cipolla. 2017. 'Segnet: A deep convolutional encoder-decoder architecture for image segmentation', *IEEE transactions on pattern analysis and machine intelligence*, 39: 2481-95.

De Raadt, Alexandra, Matthijs J Warrens, Roel J Bosker, and Henk AL Kiers. 2019. 'Kappa coefficients for missing data', *Educational and Psychological Measurement*, 79: 558-76.

Ghosh, Swarnendu, Nibaran Das, Ishita Das, and Ujjwal Maulik. 2019. 'Understanding deep learning techniques for image segmentation', *ACM Computing Surveys (CSUR)*, 52: 1-35.

Hamida, A Ben, Alexandre Benoit, Patrick Lambert, Louis Klein, C Ben Amar, Nicolas Audebert, and Sébastien Lefèvre. 2017. "Deep learning for semantic segmentation of remote sensing images with rich spectral content." In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2569-72. IEEE.

Ji, Yuzhu, Haijun Zhang, Zhao Zhang, and Ming Liu. 2021. 'CNN-based encoder-decoder networks for salient object detection: A comprehensive review and recent advances', *Information Sciences*, 546: 835-57.

Kampffmeyer, Michael, Arnt-Borre Salberg, and Robert Jenssen. 2016. "Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks." In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 1-9.

Kemker, Ronald, Carl Salvaggio, and Christopher Kanan. 2018. 'Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning', *ISPRS Journal of Photogrammetry and Remote Sensing*, 145: 60-77.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. 'Deep learning', *nature*, 521: 436-44.

Li, Ruirui, Wenjie Liu, Lei Yang, Shihao Sun, Wei Hu, Fan Zhang, and Wei Li. 2018. 'DeepUNet: A deep fully convolutional network for pixel-level sea-land segmentation', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11: 3954-62.

Liu, Yifan, Qigang Zhu, Feng Cao, Junke Chen, and Gang Lu. 2021. 'High-Resolution Remote Sensing Image Segmentation Framework Based on Attention Mechanism and Adaptive Weighting', *ISPRS International Journal of Geo-Information*, 10: 241.

Long, Jonathan, Evan Shelhamer, and Trevor Darrell. 2015. "Fully convolutional networks for semantic segmentation." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3431-40.

Mahesh, Batta. 2020. 'Machine learning algorithms-a review', *International Journal of Science and Research (IJSR).[Internet]*, 9: 381-86.

Minaee, Shervin, Yuri Y Boykov, Fatih Porikli, Antonio J Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. 2021. 'Image segmentation using deep learning: A survey', *IEEE transactions on pattern analysis and machine intelligence*.

Noh, Hyeonwoo, Seunghoon Hong, and Bohyung Han. 2015. "Learning deconvolution network for semantic segmentation." In *Proceedings of the IEEE international conference on computer vision*, 1520-28.

Pan, Zhuokun, Jiashu Xu, Yubin Guo, Yueming Hu, and Guangxing Wang. 2020. 'Deep learning segmentation and classification for urban village using a worldview satellite image based on U-Net', *Remote Sensing*, 12: 1574.

Pinheiro, Pedro O, Tsung-Yi Lin, Ronan Collobert, and Piotr Dollár. 2016. "Learning to refine object segments." In *European conference on computer vision*, 75-91. Springer.

Prathap, Geesara, and Ilya Afanasyev. 2018. "Deep learning approach for building detection in satellite multispectral imagery." In *2018 International Conference on Intelligent Systems (IS)*, 461-65. IEEE.

Qi, Xingqun, Tianhui Wang, and Jiaming Liu. 2017. "Comparison of support vector machine and softmax classifiers in computer vision." In *2017 Second International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*, 151-55. IEEE.

Shi, Jiyuan, Ji Dang, Mida Cui, Rongzhi Zuo, Kazuhiro Shimizu, Akira Tsunoda, and Yasuhiro Suzuki. 2021. 'Improvement of damage segmentation based on pixel-level data balance using vgg-unet', *Applied Sciences*, 11: 518.

Shorten, Connor, and Taghi M Khoshgoftaar. 2019. 'A survey on image data augmentation for deep learning', *Journal of big data*, 6: 1-48.

Simonyan, Karen, and Andrew Zisserman. 2014. 'Very deep convolutional networks for large-scale image recognition', *arXiv preprint arXiv:1409.1556*.

Sun, Yi, Yan Tian, and Yiping Xu. 2019. 'Problems of encoder-decoder frameworks for high-resolution remote sensing image segmentation: Structural stereotype and insufficient learning', *Neurocomputing*, 330: 297-304.

Tong, Xin-Yi, Gui-Song Xia, Qikai Lu, Huanfeng Shen, Shengyang Li, Shucheng You, and Liangpei Zhang. 2020. 'Land-cover classification with high-resolution remote sensing images using transferable deep models', *Remote Sensing of Environment*, 237: 111322.

Van Dyk, David A, and Xiao-Li Meng. 2001. 'The art of data augmentation', *Journal of Computational and Graphical Statistics*, 10: 1-50.

Wang, Hongzhen, Ying Wang, Qian Zhang, Shiming Xiang, and Chunhong Pan. 2017. 'Gated convolutional neural network for semantic segmentation in high-resolution images', *Remote Sensing*, 9: 446.

Yu, Wei, Kuiyuan Yang, Yalong Bai, Tianjun Xiao, Hongxun Yao, and Yong Rui. 2016. "Visualizing and comparing AlexNet and VGG using deconvolutional layers." In *Proceedings of the 33 rd International Conference on Machine Learning*.

Yuan, Kunhao, Xu Zhuang, Gerald Schaefer, Jianxin Feng, Lin Guan, and Hui Fang. 2021. 'Deep-learning-based multispectral satellite image segmentation for water body detection', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14: 7422-34.

Zheng, Xiaoxiong, and Tao Chen. 2021. 'High spatial resolution remote sensing image segmentation based on the multiclassification model and the binary classification model', *Neural Computing and Applications*: 1-8.