

## Using Fully Convolutional Networks and Pléiades Satellite Imagery with Pre-trained VGG-19 Model to Detect Impervious Surface in Farmland

Wei Li (1), Chi-Kuei Wang (1)

<sup>1</sup> National Cheng Kung Univ., Department of Geomatics, 1 University Road, Tainan, Taiwan  
Email: i013yamtli@gmail.com; chikuei@mail.ncku.edu.tw

**KEY WORDS:** Impervious Surface, Multispectral Satellite Imagery, Fully Convolutional Networks, Transfer Learning

**ABSTRACT:** The rapid growth of artificial impervious surface areas such as buildings and pavements has caused the loss of arable land and pose threat to food production. Periodic monitoring the use of farm land is necessary. It is on an urgent need to develop laborsaving methods since commonly practices of tallying impervious surfaces are based on manual digitization. Emerging deep convolutional neural networks (CNNs)-based techniques have shown a great potential for remote sensing image classification tasks. Transfer learning refers to the technique that transfers pre-learned representation from one domain to another. And, it has been generally applied in CNNs on specific datasets to reduce the data labelling effort and shorten the training process by adopting pre-trained networks. This study aims to explore the potential of detecting impervious surfaces using fully convolutional networks (FCNs) with pre-trained network parameters and high resolution multispectral satellite imagery. Since off-the-shelf pre-trained networks are usually trained with RGB bands, the number of input bands is restricted to three. We propose adding a convolutional layer, which generates three-layer feature maps before the pre-trained network. Experiments on a pansharpened Pléiades satellite image dataset with a pre-trained VGG-19 network were conducted. The classification accuracy F1-score of 94.2% was achieved.

## 1. INTRODUCTION

### 1.1 Background

Impervious surfaces in farm land not only pose threat to food crops production, but also eliminate water infiltration and soil water evaporation. Due to the rapid growth of industrial and residential development, the investigation of impervious surfaces in farm land should be carried out annually. Manually digitizing the areas of impervious surfaces is a commonly adopted method to investigate the farm land use status. However, such method is tedious and time consuming, making it difficult to carry out the investigation annually. Emerging deep convolutional neural network (CNN)-based techniques have shown the great potential for remote sensing data classification tasks (Kumar, Pandey, Lohani, & Misra, 2019; Maggiori, Charpiat, Tarabalka, & Alliez, 2016; Marcos, Hamid, & Tuia, 2016; Wei, Wang, & Xu, 2017). This study aims to tally buildings using multispectral satellite imagery and CNN-based methods with greater efficiency and less manual work.

### 1.2 Related Work

The image recognition tasks in the field of computer vision can be divided into: (a) image classification; (b) image classification and object localization; (c) semantic segmentation; and (d) instance segmentation. Image classification is to assign one label or a rank of labels to an input image from a fixed set of classes from the whole scene. Image classification and object localization provides not only the class labels but also spatial information such as bounding boxes or centroids for each labelled object. Semantic segmentation labels each pixel in the image rather the whole scene. Instance segmentation detects the appearance of each object instance by giving different labels to different instances of objects with the same class.

Similar to pixel-based classification tasks in the field of remote sensing, semantic segmentation makes pixel-wise class prediction to input images. A number of CNNs have been developed to fulfill the task. Some of the networks are introduced as follows. Fully convolutional network (FCN) (Long, Shelhamer, & Darrell, 2015) first introduced the concept of making pixel-wise prediction by replacing fully connected layers with convolutional layers and upscaling the heatmap to the original resolution by transposed convolution operation. Segnet (Badrinarayanan, Kendall, & Cipolla, 2015) reduces the effort to learning to upscaling by non-linear methods using pooling indices computed in the pooling layers. U-Net (Ronneberger, Fischer, & Brox, 2015) was at first developed for biomedical image segmentation and then found to be useful in other segmentation tasks. Efficient neural network (ENet) (Paszke, Chaurasia, Kim, & Culurciello, 2016) was proposed to perform pixel-wise classification in real time applications. DeepLabv3+ (Chen, Zhu, Papandreou, Schroff, & Adam, 2018) combines the advantages of spatial pyramid pooling module and skip architecture and achieves state-of-the-art performance.

## 2. MATERIALS

### 2.1 Study Area

Kaohsiung City is a special municipality located in the southern Taiwan with an area of 2,952 km<sup>2</sup>. Being the third most populous city in Taiwan, Kaohsiung is an industrial city and a major international port. There are 47,355 hectares of the arable land in Kaohsiung. Various kinds of crops such as fruit, vegetable, rice, and grain are grown in this area. The complex agricultural use makes Kaohsiung a suitable site for developing a robust method to tally buildings in farm land.

### 2.2 Satellite Imagery

Pléiades satellites provide 2-meter resolution multispectral (blue, green, red, and near infrared bands) and 0.5-meter resolution panchromatic satellite imagery. The pansharpened product, taking advantage of both multispectral and panchromatic bands, comprises blue, green, red, and near infrared bands with the spatial resolution of 0.5m. Pléiades satellite imagery has shown its potential to investigate land use and land cover status using either traditional remote sensing approaches (Feng et al., 2017; Sublime, Troya-Galvis, &Puissant, 2017) or deep learning techniques (D.Wang et al., 2018) in previous studies.

## 3. METHODS

### 3.1 Image Masking

Image masking is considered to be a useful technique to improve modeling performance by restricting the analysis to specific groups of pixels rather than analyzing the whole scene (Kastens et al., 2005). Masks derived from cadastral records were applied to narrow down the regions of interest since the focus of this study lies on the detection of impervious surfaces in farm land. Reference data were produced by manually digitizing the impervious surfaces from the masked images.

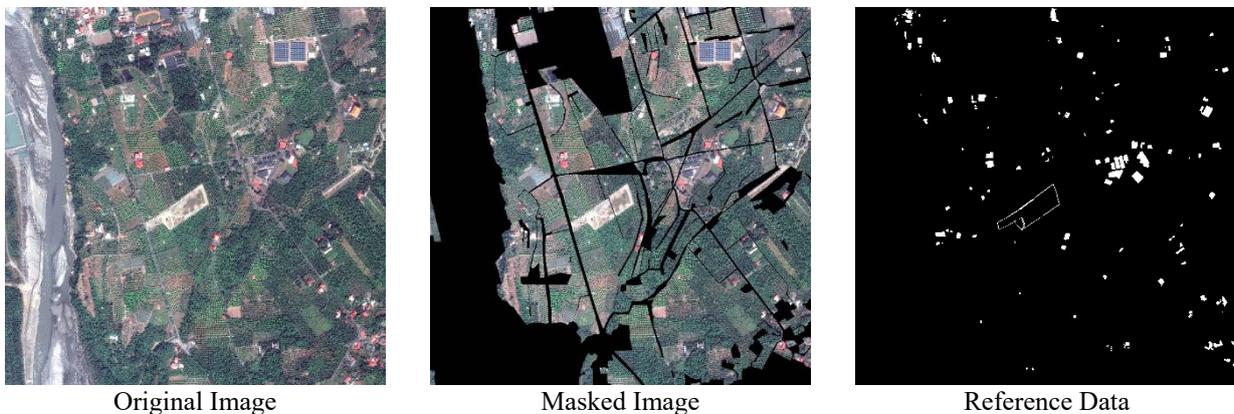


Figure 1 Image masking using cadastral records.

### 3.2 Data Augmentation

Data augmentation methods (Krizhevsky, Sutskever, &Hinton, 2012) enlarge datasets using label-preserving transformations to reduce the effort of labelling. To prevent the networks from overfitting on the image dataset, the dataset was augmented with random cropping and flipping methods. Both methods generated transformed images from the masked images with little computing power. To generate image translations, 1,000 patches with the size of 224×224 pixels were randomly extracted from each masked image. Compared to producing image dataset of non-overlapping patches, random cropping method increased the size of image dataset by a factor around 7. The extracted patches using random cropping method were horizontally flipped, vertically flipped, and 180° rotated. The flipping methods increased the size of image dataset by another factor of 4. Totally 20,000 patches were generated from 5 masked images.

### 3.3 Fully Convolutional Network

FCNs take advantage of pre-trained network parameters with a succinct encoding-decoding network structure. FCNs, which elegantly adapt the recent success of classification networks into semantic segmentation by transferring the pre-learned representations into themselves, have been widely used by the remote sensing community (Fu, Liu, Zhou, Sun, &Zhang, 2017; Piramanayagam, Saber, Schwartzkopf, &Koehler, 2018; J.Wang, Shen, Qiao, Dai, &Li, 2019; Wu et

al., 2019). The networks used in this study are based on FCN-8s with pre-trained VGG19 and implemented in Tensorflow using python.

### 3.3.1 Transfer Learning

Transfer learning is a technique that attempts to reduce the efforts in training process via transferring pre-learned knowledge from one data domain to another in the field of machine learning (S. J.Pan &Yang, 2010). Since deep convolutional neural networks tend to extract low level features such as edges and corners in lower hierarchy layers (Zeiler &Fergus, 2013), it is asserted that the lower hierarchy layers are similar in different image datasets. Only the higher hierarchy layers, which may be specific to different image dataset and classification tasks, need to be fully trained. Transfer learning technique has been widely implemented in various image classification tasks using CNNs, for instance, fine-tuning networks pre-trained on natural image dataset to medical image tasks (Shin et al., 2016).

Long, Shelhamer, &Darrell (2015) adapted several representative CNNs, namely AlexNet, VGG16, and GoogLeNet to FCNs. And, the FCN adopting VGG16 achieved the best classification accuracy. VGG19 (Simonyan &Zisserman, 2014), an off-the-shelf pre-trained model, which has a deeper structure and higher accuracy than VGG16, was adopted in this study. The number of channels of the input data would be restricted to 3 since VGG19 was trained on natural images with RGB color only, while multispectral satellite imagery contains more than 3 bands. Networks with different input layer designs were tested and compared to address this problem.

Copy Initialization Network (CoinNet) was developed to take advantage of all spectral bands and make full use of the parameters from the first layer of the pre-trained network for multispectral imagery semantic segmentation task (B.Pan et al., 2019). Another strategy that adds a convolutional layer before the transferred pre-trained network was proposed in this study. Experiments carried out to compare the performance are described in the following section.

### 3.3.2 Network Architecture

Experiments of network architectures were designed based on different strategies. Experiments conducted in this study, namely FCN-8s-RGB, FCN-8s-CIR, FCN-8s-RI, FCN-8s-TL, FCN-8s-CoinNet and FCN-8s-Conv-TL, are described as follows.

Dimension reduction techniques such as selection of most uncorrelated bands (Mahdianpari et al., 2018) and principal component analysis (PCA) (Jiao et al., 2017) are usually pursued to adjust the data to the pre-trained network by eliminating some data redundancy beforehand.

The input data were adapted to FCN-8s-RGB by selecting red, green, and blue bands. The spectral bands are the same as the dataset on which VGG19 was trained.

Color-Infrared imagery is commonly used in remote sensing applications. The input data were adapted to the standard FCN-8s-CIR by selecting near infrared, red, and green bands, which are the band compositions of Color-Infrared imagery.

To make good use of every spectral band, we designed networks which take all spectral bands as the input. The first layer of the pre-trained VGG19 is a convolutional layer with the filter kernel size of  $3 \times 3 \times 3 \times 64$ . In order to make use of all multispectral bands, the size of the filter kernel needs to be adjust to  $3 \times 3 \times 4 \times 64$  by expanding the input dimension from 3 to 4.

FCN-8s-RI randomly initialize the weight parameters of the first layer of the pre-trained network since the size has been changed. The weight parameters of the first layer are described as

$$w^{FCN-8s-RI} = [w_1 \ w_2 \ w_3 \ w_4] \dots\dots\dots(1)$$

where each element in  $w_i$  is randomly initialized.

The new weight parameters produced by the dimension expansion are randomly initialize in FCN-8s-TL. The other parameters, having the same size with the pre-trained VGG19, are initialized using the pre-trained VGG19 parameter values. The weight parameters of the first layer are described as

$$w^{FCN-8s-TL} = [w_1^{VGG19} \ w_2^{VGG19} \ w_3^{VGG19} \ w] \dots\dots\dots(2)$$

where each element in  $w$  is randomly initialized.

Copy Initialization Network (CoinNet) was developed to take advantage of all spectral bands and make full use of the parameters from the first layer of the pre-trained network for multispectral imagery semantic segmentation tasks (B.Pan et al., 2019). The new weight parameters produced by the dimension expansion using a duplication strategy based on the assumption that although the reflectance in different bands present different characteristics, the overall pattern of them should be similar. As a result, it is reasonable to adopt weight parameters from other channels to extract low level features in the first layer. The weight parameters of the expanded channel are copied from the pre-trained parameters of the first channel. The weight parameters of the first layer of FCN-8s-CoinNet are described as

$$w^{FCN-8s-coin} = [w_1^{VGG} w_2^{VGG} w_3^{VGG} w_1^{VGG}] \dots \dots \dots (3)$$

It is assumed that a convolutional layer can learn to produce output feature maps that suit the following transferred network. In order to fully utilize the unsupervised adaptivity of a CNN, a convolutional layer which produces 3-channel feature maps along with ReLU activation function was added before the standard FCN-8s network using VGG19. The weight parameters of the first layer of FCN-8s-Conv-TL are described as

$$w^{FCN-8s-Conv-TL} = [w_1 w_2 w_3 w_4], w_i \in \mathbb{R}^{3 \times 3 \times 3} \dots \dots \dots (4)$$

where each element in  $w_i$  is randomly initialized.

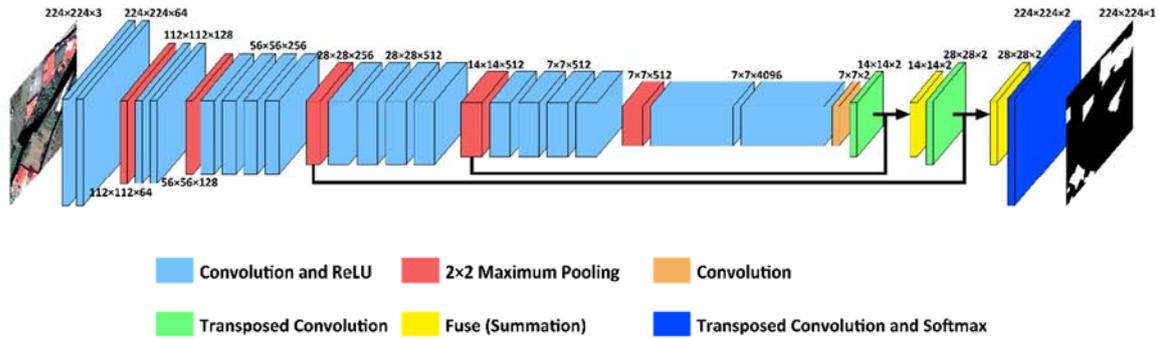


Figure 2 FCN-8s-RGB and FCN-8s CIR architecture.

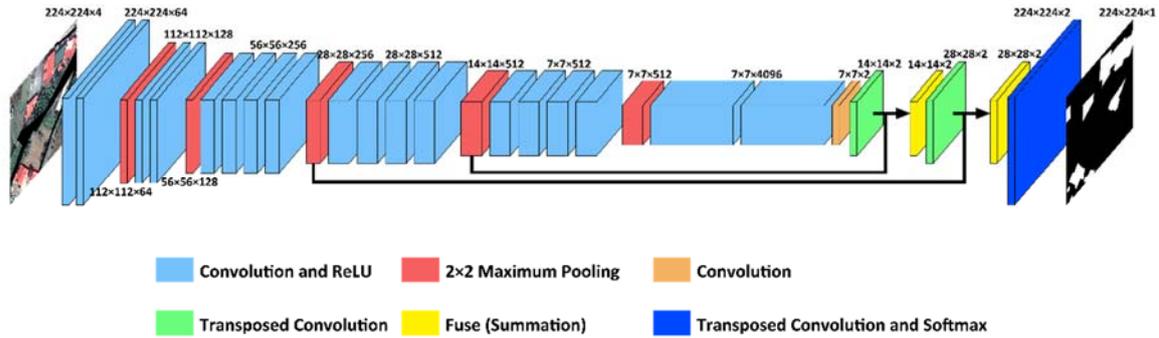


Figure 3 FCN-8s-RI, FCN-8s-TL and FCN-8s-CoinNet architecture.

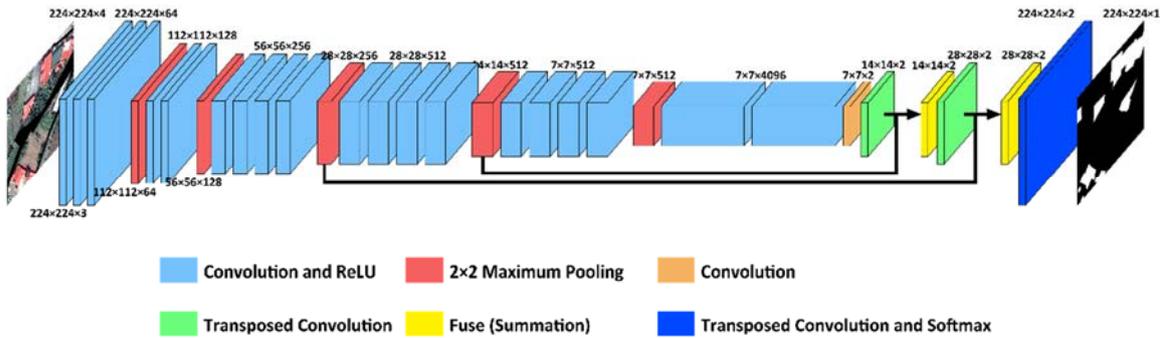


Figure 4 FCN-8s-Conv-TL architecture.

## 4. RESULTS AND DISCUSSION

Experiments were carried out by training different FCN-8s-based networks using fixed hyperparameters (the batch size was set as 16 and the learning rate was set as  $1 \times 10^{-6}$ ). 16,000 image patches were used to train the networks, and 4,000 image patches were used to test the networks. Networks were trained using a work station equipped with a NVIDIA GeForce RTX 2080 Ti GPU.

### 4.1 Accuracy Assessment

Performance of the networks were evaluated using 100 images from the testing samples. FCN-8s-CIR achieved higher recall value than FCN-8s-RGB with the same training iteration number. FCN-8s-RI and FCN-8s-TL used all 4 bands as input and spent twice the training iterations compared to FCN-8s-RGB and FCN-8s-CIR. However, the accuracy indices values are the worst among all networks. Expanding the first layer of the pre-trained network with randomly initialized parameters may affect the network performance. It is found that appropriate initial values of parameters play an import role in network training. FCN-8s-CoinNet achieved a higher F1 Score value with lower training iteration number compared to FCN-8s-RI and FCN-8s-TL. Results have shown that FCN-8s-Conv-TL, which is one layer deeper than other networks, achieved the highest accuracy in all measurement metrics with the largest training iteration number among all networks.

Smoothness and the scale of detail of the classification result are important issues for semantic classification. Comparisons between predictions and reference data are shown to compare the classification accuracy and segmentation detail. Individual rooftops and complex impervious surface such as rooftops and concrete surfaces are the two typical appearances of buildings in the study area. Every individual rooftop in Figure 5 was successfully detected by all the networks. Some areas were misidentified as buildings by FCN-8s-CIR, FCN-8s-RI, and FCN-8s-TL. FCN-8s-CoinNet and FCN-8s-Conv-TL produced neck and neck predictions with less salt-and-pepper noise and higher accuracies than other networks. The complex impervious surface as shown in Figure 6 was detected by FCN-8s-Conv-TL with complete shape with sharp edge while other networks failed to detect the whole impervious surface completely.

Generally speaking, all the networks provide passable results detecting buildings in farm land. FCN-8s-Conv-TL presented the best results with finest boundaries and highest classification accuracy values.

Table 1 FCN-8s-based network evaluation.

	FCN-8s- RGB	FCN-8s- CIR	FCN-8s- RI	FCN-8s- TL	FCN-8s- CoinNet	FCN-8s-Conv- TL
Input Bands	3	3	4	4	4	4
Training Iterations	120,000	120,000	230,000	230,000	180,000	250,000
Overall Accuracy	0.981	0.983	0.979	0.980	0.984	0.987
Precision	0.932	0.933	0.917	0.922	0.925	0.944
Recall	0.899	0.912	0.899	0.898	0.935	0.940
F1 Score	0.915	0.922	0.908	0.909	0.930	0.942



True Color



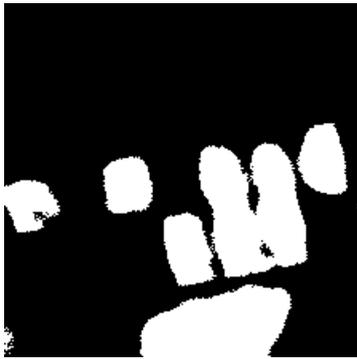
Color-Infrared



Reference Data



FCN-8s-RGB



FCN-8s-CIR



FCN-8s-RI



FCN-8s-TL



FCN-8s-CoinNet



FCN-8s-Conv-TL

Figure 4 Detection of individual rooftops.

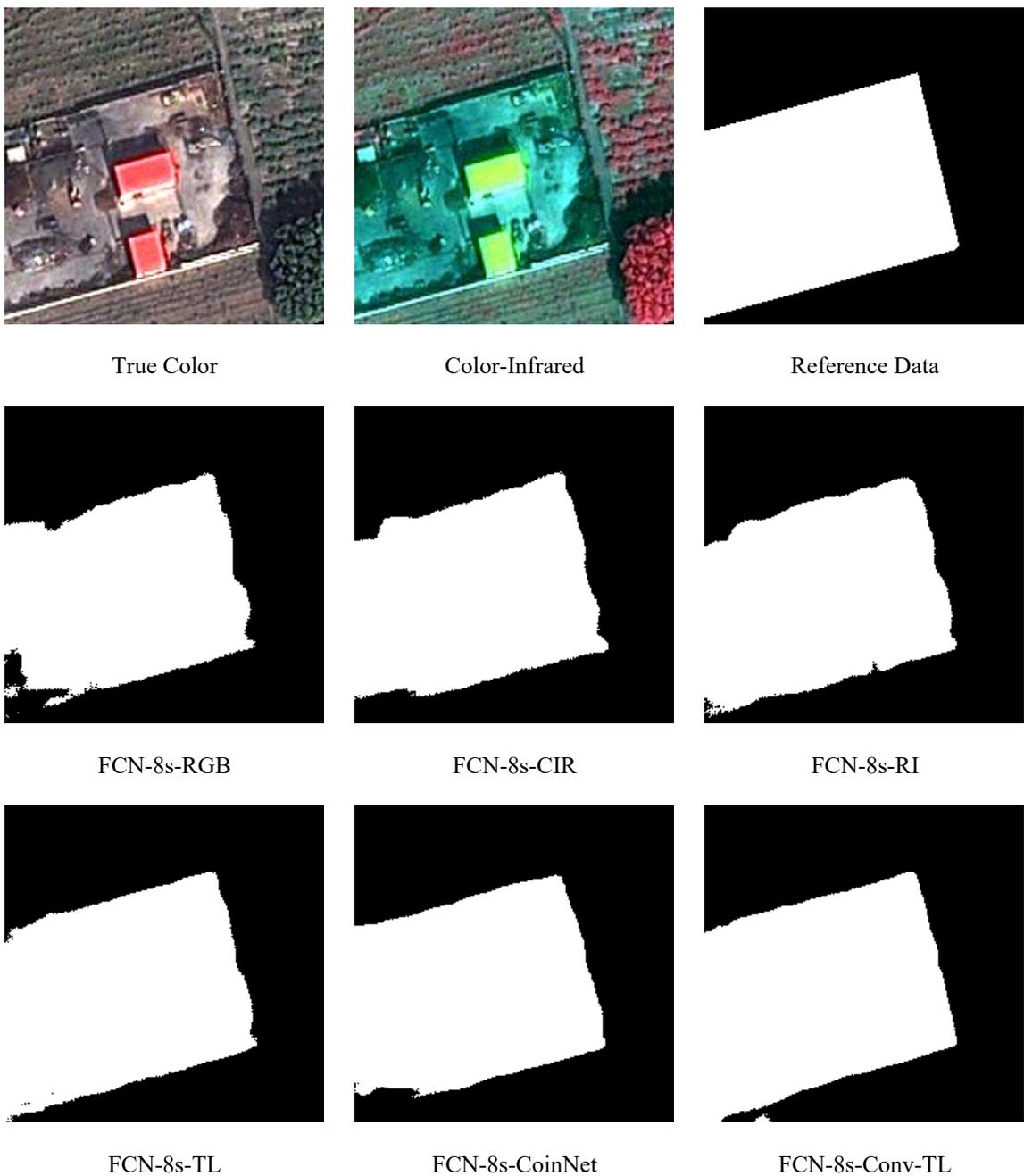


Figure 5 Detection of complex impervious surface.

## 4.2 Transferability

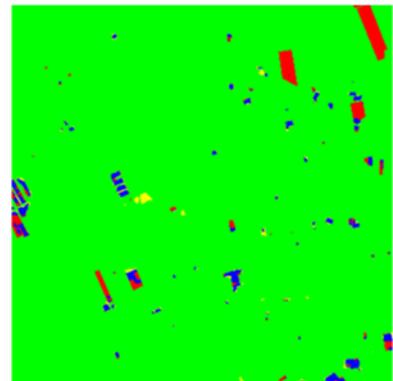
The transferability of FCN-8s-Conv-TL, which outperforms other networks from the previous tests, was validated using coastal area of Kaohsiung City. The overall accuracy, precision, recall, and F1 score of 98.4%, 76.6%, 44.1% and 56.0% were achieved, respectively. Buildings with small footprints, which may be omitted to be digitized during manual work, were detected. However, the sizes of the buildings are usually underestimated. Complex impervious surfaces contain rooftops, rooftop water towers, balconies, and concrete pavements. The complex impervious surfaces were detected with incomplete borders. The land use status of this area is commonly seen in residential area but rarely found in farm land. As a result, the network tends to extract rooftops, the major buildings in the testing dataset, rather than the whole complex impervious surfaces. Sheet metal houses which appear to be bright and dark stripes in satellite images were not detected by the network. Such buildings are usually chicken coops in farm land. Chicken coops were omitted because buildings with such kind of pattern were not included in the training image dataset. The network shows poor recall and F1 score accuracy because of the high omission rate of chicken coops.



Coastal area of Kaohsiung City



Detection of impervious surfaces



Comparisons between the prediction and reference data. Blue for true positive, yellow for false positive, red for false negative, and green for true negative.

Figure 6 Transferability test in coastal area of Kaohsiung City.



Figure 7 Complex impervious surface captured from Google Street View.



Figure 8 Chicken coops captured from Google Street View.

## 5. CONCLUSION AND OUTLOOK

This study aims to investigate the potential for detecting impervious surfaces in farm land using fully convolutional networks and multispectral satellite imagery. Since there were limited training samples, transfer learning technique was implemented to adapt existing classification network into pixel-wise classification. The pre-trained networks are usually trained with RGB datasets, which make the number of input bands restricted to 3. To address this problem, we proposed FCN-8s-Conv-TL which adds a convolutional layer before the pre-trained network to generate three-channel feature maps. Experiments were conducted to compare the performance of FCN-8s-Conv-TL and other networks with different strategies, namely FCN-8s-RGB, FCN-8s-CIR, FCN-8s-RI, FCN-8s-TL, and FCN-8s-CoinNet. Quantitative and qualitative results have shown that FCN-8s-Conv-TL have the best performance among all networks. The overall accuracy, precision, recall, and F1 score of 98.7%, 94.4%, 94.0%, and 94.2% was achieved, respectively.

Transferability of FCN-8s-Conv-TL was validated using the satellite image in coastal area. Buildings which appears to be small segments were successfully detected. Although the network underestimates the area of each small segment, the results are still valuable for the task since such buildings are likely to be omitted by human eyes. Detection of complex impervious surfaces was incomplete since the network tends to detect buildings rather than the whole impervious surfaces. Sheet metal houses, with the distinctive pattern of a series of bright and dark stripes, were failed to be detected due to the absence of training samples.

The potential of detecting impervious surfaces in farm land using deep convolutional networks and multispectral satellite imagery has been shown in this study. A complete image dataset which embraces a wide variety of buildings in farm land is required to further improve the performance. The next move will be the time series building change detection. Besides the area of build-up surfaces, the number of buildings is an useful information for building change detection. It is suggested that state-of-the-art networks for instance segmentation such as Mask R-CNN are worth trying since they provide individual image objects with finer boundaries.

## REFERENCES

- Badrinarayanan, V., Kendall, A., & Cipolla, R. (2015). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. Retrieved from <http://arxiv.org/abs/1511.00561>
- Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. Retrieved from <https://arxiv.org/abs/1802.02611>
- Feng, Y., Lu, D., Moran, E., Dutra, L., Calvi, M., deOliveira, M., ...DeOliveira, M. A. F. (2017). Examining Spatial Distribution and Dynamic Change of Urban Land Covers in the Brazilian Amazon Using Multitemporal Multisensor High Spatial Resolution Satellite Imagery. *Remote Sensing*, 9(4), 381. <https://doi.org/10.3390/rs9040381>
- Fu, G., Liu, C., Zhou, R., Sun, T., & Zhang, Q. (2017). Classification for High Resolution Remote Sensing Imagery Using a Fully Convolutional Network. *Remote Sensing*, 9(5). <https://doi.org/10.3390/rs9050498>
- Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., & Garcia-Rodriguez, J. (2017). A Review on Deep Learning Techniques Applied to Semantic Segmentation. Retrieved from <http://arxiv.org/abs/1704.06857>
- Jiao, L., Liang, M., Chen, H., Yang, S., Liu, H., & Cao, X. (2017). Deep Fully Convolutional Network-Based Spatial Distribution Prediction for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(10), 5585–5599. <https://doi.org/10.1109/TGRS.2017.2710079>
- Kastens, J. H., Kastens, T. L., Kastens, D. L. A., Price, K. P., Martinko, E. A., & Lee, R. Y. (2005). Image masking for crop yield forecasting using AVHRR NDVI time series imagery. *Remote Sensing of Environment*. <https://doi.org/10.1016/j.rse.2005.09.010>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, 1097–1105. Retrieved from <http://dl.acm.org/citation.cfm?id=2999134.2999257>
- Kumar, B., Pandey, G., Lohani, B., & Misra, S. C. (2019). A multi-faceted CNN architecture for automatic classification of mobile LiDAR data and an algorithm to reproduce point cloud samples for enhanced training. *ISPRS Journal of Photogrammetry and Remote Sensing*, 147, 80–89. <https://doi.org/10.1016/j.isprsjprs.2018.11.006>
- Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3431–3440. <https://doi.org/10.1109/CVPR.2015.7298965>
- Maggiori, E., Charpiat, G., Tarabalka, Y., & Alliez, P. (2016). Recurrent Neural Networks to Correct Satellite Image Classification Maps. <https://doi.org/10.1109/TGRS.2017.2697453>
- Mahdianpari, M., Salehi, B., Rezaee, M., Mohammadimanesh, F., Zhang, Y., Mahdianpari, M., ...Zhang, Y. (2018). Very Deep Convolutional Neural Networks for Complex Land Cover Mapping Using Multispectral Remote Sensing Imagery. *Remote Sensing*, 10(7), 1119. <https://doi.org/10.3390/rs10071119>
- Marcos, D., Hamid, R., & Tuia, D. (2016). Geospatial Correspondences for Multimodal Registration. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 5091–5100. <https://doi.org/10.1109/CVPR.2016.550>
- Pan, B., Shi, Z., Xu, X., Shi, T., Zhang, N., & Zhu, X. (2019). CoinNet: Copy Initialization Network for Multispectral Imagery Semantic Segmentation. *IEEE Geoscience and Remote Sensing Letters*, 16(5), 816–820.

<https://doi.org/10.1109/LGRS.2018.2880756>

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345–1359. <https://doi.org/10.1109/TKDE.2009.191>

Paszke, A., Chaurasia, A., Kim, S., & Culurciello, E. (2016). ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation. Retrieved from <https://arxiv.org/abs/1606.02147>

Piramanayagam, S., Saber, E., Schwartzkopf, W., & Koehler, F. W. (2018). Supervised classification of multisensor remotely sensed images using a deep learning framework. *Remote Sensing*, 10(9), 1–25.

<https://doi.org/10.3390/rs10091429>

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. Retrieved from <http://arxiv.org/abs/1505.04597>

Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., ... Summers, R. M. (2016). Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Transactions on Medical Imaging*, 35(5), 1285–1298. <https://doi.org/10.1109/TMI.2016.2528162>

Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. Retrieved from <http://arxiv.org/abs/1409.1556>

Sublime, J., Troya-Galvis, A., & Puissant, A. (2017). Multi-Scale Analysis of Very High Resolution Satellite Images Using Unsupervised Techniques. *Remote Sensing*, 9(5). <https://doi.org/10.3390/rs9050495>

Wang, D., Wan, B., Qiu, P., Su, Y., Guo, Q., Wang, R., ... Wu, X. (2018). Evaluating the Performance of Sentinel-2, Landsat 8 and Pléiades-1 in Mapping Mangrove Extent and Species. *Remote Sensing*, 10(9).

<https://doi.org/10.3390/rs10091468>

Wang, J., Shen, L., Qiao, W., Dai, Y., & Li, Z. (2019). Deep Feature Fusion with Integration of Residual Connection and Attention Model for Classification of VHR Remote Sensing Images. *Remote Sensing*, 11(13).

<https://doi.org/10.3390/rs11131617>

Wei, Y., Wang, Z., & Xu, M. (2017). Road Structure Refined CNN for Road Extraction in Aerial Image. *IEEE Geoscience and Remote Sensing Letters*, 14(5), 709–713. <https://doi.org/10.1109/LGRS.2017.2672734>

Wu, G., Guo, Y., Song, X., Guo, Z., Zhang, H., Shi, X., ... Shao, X. (2019). A Stacked Fully Convolutional Networks with Feature Alignment Framework for Multi-Label Land-cover Segmentation. *Remote Sensing*, 11(9).

<https://doi.org/10.3390/rs11091051>

Zeiler, M. D., & Fergus, R. (2013). Visualizing and Understanding Convolutional Networks. Retrieved from <https://arxiv.org/abs/1311.2901>