EXTRACTION OF ROCKS AND BOULDERS ON NATURAL TERRAIN USING SEMANTIC SEGMENTATION

Shenlu Jiang (1), Coco Yin Tung Kwok (1), Man Sing Wong (1,*)

[1] Department of Land Surveying and Geo-informatics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China

Email: sljiang@polyu.edu.hk; yt-coco.kwok@connect.polyu.hk; lswong@polyu.edu.hk;

KEY WORDS: Rock outcrops; Boulders; Deep Learning; Semantic Segmentation; Remote Sensing

ABSTRACT: The geological mapping for the rock outcrop and boulders distribution on countryside is necessary for support hazard and risk management. Different approaches were proposed to detect and identify the rock outcrop and boulders based on remote sensing technology in literature. Given the latest development in technology, using traditional remote sensing methods to detect the rocks and boulders in VHR aerial imagery is still challenging. The complicated background noise prohibits the precision and accuracy of identifying the targets as well as their boundaries detection. Currently, deep neural networks (DNNs) have been demonstrating its outstanding performance on feature extraction which has been widely used in the computer vision. The semantic image segmentation is a pixel-level classification model, which is appropriate to the rock and boulder detection due to the model itself can segment very dense and detailed features over VHR imagery. In this paper, an optimized FCN-DenseNet was proposed to detect the rock outcrops and boulders in Hong Kong. Since the sizes of the rocks and boulders are variated, the neural network should be employed with multi-scale pooling to extract low/high-level features. As the contexts of the rocks and background terrain are similar, the pyramid dense blocks were employed to enhance the capability of self-feature-extraction. Our proposed boulder-net was evaluated over the VHR imagery in Hong Kong. The results show that the rocks and boulders are identified and classified and the boundaries of the rocks are accurately segmented.

1. INTRODUCTION

Machine Learning is an effective empirical approach for classification applications, which is widely used in the community of remote sensing, i.e., ocean (Yi and Prybutok, 1996), climate change (Zahabiyoun et al., 2013) and crop/agriculture practices (Carpenter et al., 1997). Current methods mostly employed in human-crafted features, e.g., scale-invariant feature transform (David, 1999) and histogram of oriented gradients (Sebastian et al., 2013) and traditional classifier, e.g., support vector machines (Cortes and Vapnik, 1995) and self-organizing map (Chon and Park, 2008), to categorize the objects/features. Currently, detecting the rock outcrops and boulders over VHR imagery is still under research due to the extremely complicated background in rural environment. Some researchers evaluated the distribution of rocks by taking photos (Bonilla-Sierra et al., 2015), laser (Afana et al., 2013) or combining photo and laser (Beraldin, 2004) on the ground. However, detecting the boulders and rock outcrops by fieldwork is challenging as it is difficult to access to the ridged terrain and it costs exhaustive human resources and times. Comparing with the in-situ method, detecting from the aerial has a merit on rapidly acquiring ground information. Some approaches employed the UAVs imagery (Salvini et al., 2015), and satellite imagery (Coggan et al., 2007) for boulder detection. Three major challenges remain in the existing methods, namely, 1.) Rapidly and accurately identifying the targets against complicated background; 2.) Precisely classifying the boulders and rock outcrops; 3.) Extracting accurate boundaries of the boulders and rock outcrops. To classify the targets/objects via traditional classifier, selection of the scheme for identifying the objects is essential, in which selective search and brute forcing scanning window are mostly utilized. However, these methods can only provide a general position of the targets. On the other hand, it consumes exhaustive computing cost and always fails during multi-scale detection. On the contrary, some methods used non-linear architecture, but the architecture is not complicated enough to cover all situations, which makes the classifier be sensitive and causing errors. Moreover, the exact boundaries of the boulders and rock outcrops cannot be extracted using this approach.

To overcome the aforementioned challenges, we employed a semantic segmentation deep neural network to detect the rock outcrops and boulders in pixel level in Hong Kong. The existing CNN (Krizhevsky et al., 2012) is used to extract the features by neural network itself, which has shown promising performance in feature extraction than the traditional methods. The semantic segmentation DNN employed the convolution layer to extract multi-scale features through down samplings and recover the size of feature map through transposed convolutional layer to derive a pixel-level mask image (Long et al., 2015). However, loss of feature information of low-level features (e.g. edge of boulders) is inevitable during the down-sampling; vice versa, high-level features can be extracted by a series of down-samplings. This confliction leads the challenge in designing the neural network, i.e., a question of how many times of down-sampling should be deployed in the DNN to allow multi-scale detection. Several DNNs were proposed to overcome this issue, for example, Unet (Ronneberger et al., 2015) employed the skipping connection architecture to compensate low-level features from early/middle stage feature maps to the feature maps during upscale. However, noises caused by extracted features are contained in the early-stage feature maps, and it leads the ambiguous classification for the final outputs. The atrous convolution (Chen et al., 2018) was proposed to utilize dilated convolution (i.e., convolute the feature maps with holes) to extract features without down-sampling, but the receptive field is fixed, and it restricted the detection of objects in different scales. The above DNNs mostly employed the basenet to utilize the pre-trained features to fine-turn the features, which minimizes the training time, however, the dataset of basenet is mostly trained from the ground perspective view, which represented different features compared with our proposed task. Therefore, this study employs DenseNet datasets for modeling, the DenseNet (Landola et al., 2015) builds up a dense inline connection among different scales' feature maps which costs less computing cost than the ResNet, VGG etc., but with simpler architecture. We proposed DNN applied FCN-DenseNet (Jegou et al., 2017) architecture, and it was optimized with the multi-scale detection. To detect the rock outcrops and boulders in different sizes, five down-sampling layers were designed to extract the features by the DenseNet. The deconvolution was utilized to gradually increase the feature map to the original size in order to accurately extract the details of the boulders. The multi-scale features are utilized and being self-extracted through the dense connection. At the end of the neural network, the softmax classifies boulders, rock outcrops and the background. Differing from the semantic segmentation from in-situ datasets, the scales of boulders and rock outcrops are greatly different in aerial view (i.e., some boulders have a large spatial extent, but others cover only very slight pixels). In our proposed neural network, the number of down-sampling layers was increased to segment the boulders and rock outcrops at higher level. The parameters of our neural network (e.g., optimizer, loss function) are optimized in this study. Increasing the depth of each layer would lead over-fit problem and incredible cost on GPU memory. Therefore, our optimized neural network utilized an optimized depth for each convolutional layer which maintains effective performance without significant loss of accuracy.

To demonstrate the efficiency of our proposed method, the neural network was evaluated in a real-world case. The results show that boulders and rock outcrops in different scales can be accurately identified and the boundaries are able to be segmented precisely. Our proposed neural network allows the DNN self-extraction of the features. Our optimized architecture also enables the DNN to classify the rock outcrops and boulders in different scales. To the best of our knowledge, this is the first study on applying the semantic image segmentation DNN in the geological mapping using VHR aerial imagery.

2. PROPOSED METHOD

The architecture and methodology of our proposed neural network are explained in this section. This part can be divided as follow: 1.) Residual Convolution, 2.) Dense Connection, 3.) Deconvolution, 4.) Skipping Connection, 5.) Summary of our optimization.

2.1 Modified Residual Convolution Unit

The DenseNet is an image recognition neural and achieves the remarkable score in the ImageNet classification test, which illustrates its outstanding feature extraction ability. The DenseNet takes both merits from residual convolution unit (RCU) (He et al., 2016) and inception (Szegedy et al., 2017) to avoid gradient vanishing and maximize the utilization rate of multi-scale features, the formula of RCU in the DenseNet is defined as follow:

$$X_l = X_l(X_{l-1}) + X_{l-1} \tag{1}$$

Assuming each neural layer as X_l, each X_l is residually connected by former layer X_{l-1}. At the same time, the bypass connection is employed to connect X_l and X_{l-1}. The inline connection is built up through this architecture among the layers.

2.2 Dense Connectivity

The residual convolution layer allows the architecture of DNN to be very deep, but the Fractal Net (Larsson et al., 2017) illustrates the information redundancy still exists during the depth increases. Otherwise, the low-level features in the early stage layers are inevitably lost during the down-sampling. To solve this problem, the DenseNet is proposed to maximize the utilization rate of the features. The architecture of Dense Block is shown in Figure 1.

$$X_l = H_l\{[X_0, X_1, \dots, X_{l-1}]\} \tag{2}$$

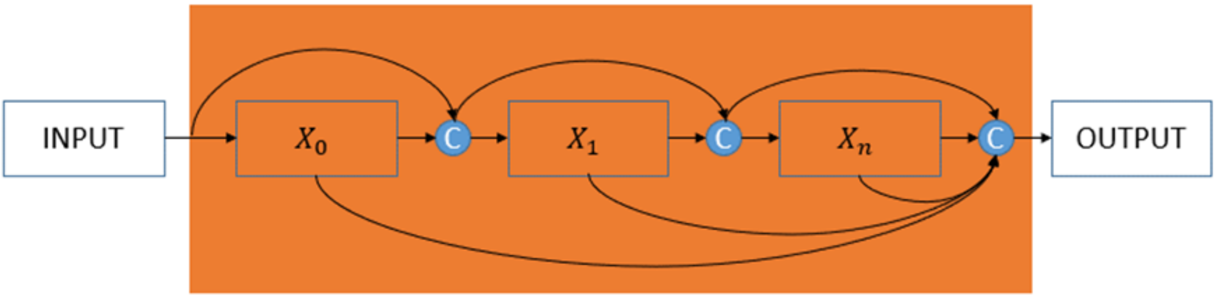

Figure 1. The Framework of Dense Connection

The formula of Dense Block is defined in formula 2, the symbol [] means concatenation. Each layer X_l combines the feature maps X_0 - X_{l-1} and be fused through modified RCU.

The entire architecture of the DenseNet is shown in Figure 2. The features in each scale are extracted through dense block, and features in each scale are combined by dense connection. At the end of DenseNet, the DNN is output by the softmax classifier to determine the classification of the target. However, our ultimate task is to obtain the pixel-level mask imagery with the same size as the original input, therefore further processing is required to recover the size of the feature map.

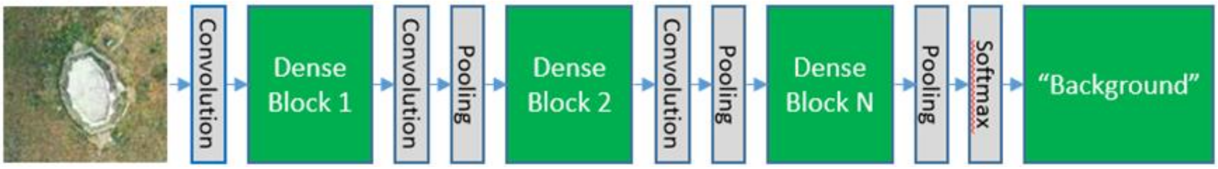

Figure 2. The Entire Framework of DenseNet

2.3 Deconvolution

To recover the size of feature map, an up-scale convolution architecture, i.e., deconvolution/transposed convolution was employed. The formula of deconvolution is defined as:

$$o = s(i - 1) + k - 2p \quad (3)$$

where s means the stride size, i donates the size of input, p means the padding size. Comparing with bilinear recovery, the deconvolution gradually adjusts the size of feature map to avoid ambiguous classification.

2.4 Skip Connection

Although the size of feature map was recovered to the original input through the deconvolution layers, accurately segmentation and detection of the boundary of rock outcrops and boulders during the up-scale are still uncertain due to the inevitable feature loss happens during the deconvolution. To avoid the ambiguous classification for the background and our targeting features, the skipping connection was employed to compensate the early/middle stages features to the deconvolution layers and compensate low/middle-level features to the output. The bypass-connection was employed to connect their corresponding up-scale feature maps.

2.5 Our Optimization for the Boulder Detection

As stated in the previous sections, we further discussed our optimization for the task of the rock outcrop and boulder detection. The architecture diagram of our proposed neural network is shown in Figure 3 and the detailed deployment is shown in Table 1. To enable detection invariance on scale, the number of down-sampling was increased to allow larger receptive fields in multi-scale feature maps. The loss function and batch normalization of each layer were employed to improve the grouping capability, i.e., the batch normalization was deployed after Relu. In Table 1, the detailed architecture of our designed neural network is shown.

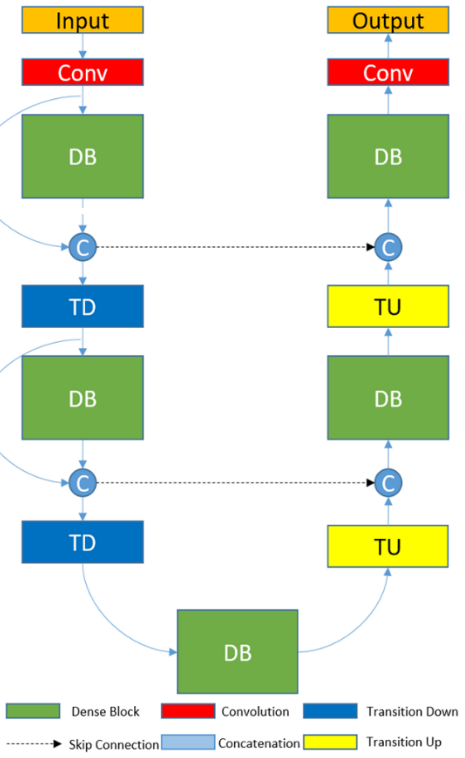

Figure 3. The diagram of our proposed DNN

Table 1. The Architecture of Our Proposed Neural Network

Architecture of our proposed DNN
Input, n = 3
3x3 Convolution, n = 36
DB (5 layers) + TD, n = 96
DB (6 layers) + TD, n = 168
DB (8 Layers) + TD, n = 264
DB (10 Layers) + TD, n = 384
DB (12 Layers) + TD, n = 528
DB (15 Layers) + TD, n = 708
DB (17 Layers), n = 912, Bottom
TU + DB (15 layers), m = 1092
TU + DB (12 layers), m = 852
TU + DB (10 layers), m = 648
TU + DB (8 layers), m = 480
TU + DB (6 layers), m = 336
TU + DB (5 layers), m = 248
1x1 Convolution, m = c
SoftMax

The DB donates the Dense Block, TD means Transition Down, TU stands for transition up, m is the depth of feature map and c is the number of classes. In the down-sampling procedure, the depth of each dense block is calculated as listed in the following formula:

$$d_n = L * G + d_{n-1} \tag{4}$$

The G donates the growth rate and the L stands for the number of layers. The depth of each dense block equals to the sum number layer multi growth rate and the depth of last dense block. During the deconvolution, the skipping connection compensates early-stage feature map to corresponding up-scale feature map.

Comparing with other semantic segmentation neural networks, our proposed method has three merits in rock outcrop and boulder detection. 1.) The DNN extracts the features totally by the neural network itself instead of basenet, which allows the DNN to better adapt the aerial view and identify the difference between boulders and background; 2.) The neural network maximizes the utilization rate of features among scales, which enables the DNN to detect the rock outcrops and boulders in very detailed scale; 3.) The increased number of pooling steps allows the neural network to be able to segment the rock outcrops and boulders in a smaller scale.

3. EXPERIMENT

In this section, our proposed neural network was discussed. The image dataset was acquired on the VHR images of Hong Kong. A pixel resolution is 10 cm on ground.

3.1 Training data and test data

Sufficient training data is essential to train the deep neural network. However, a very limited dataset for rock outcrop and boulder detection in VHR imagery can be found online, thus we built the dataset by our own. The boulders were labelled in the ArcGIS as polygon shapefile, then being modified to the corresponding label image as shown in Figure 4. A random region in the dataset was selected as training dataset. Around 8,900 boulders were labelled on a 15,082 x 11,537 pixel VHR image. Due to the limitation of hardware, the neural network cannot input such large resolution imagery at the same time. Therefore, the image is split into 512 x 512 images with overlapping of 256 pixels. To evaluate our proposed neural network, another image with 1,000 labels was used.

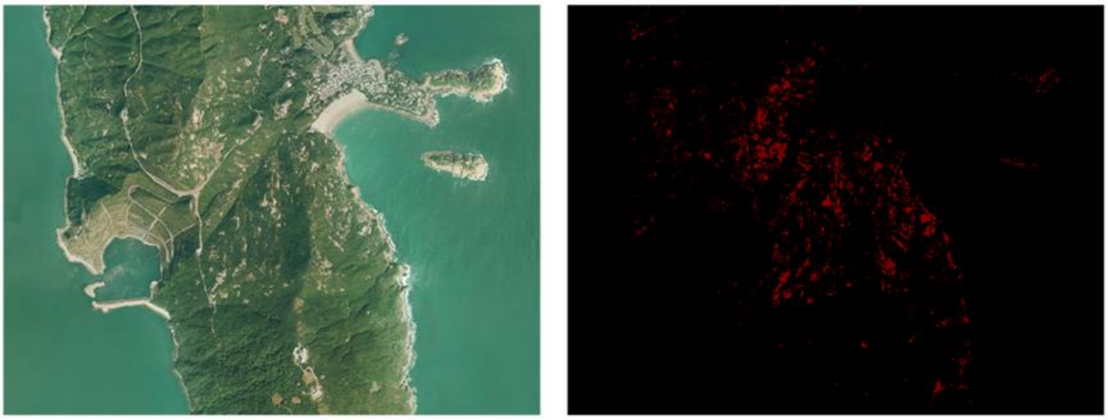

Figure 4. Training Dataset, the left is the original imagery, the right is the ground truth where black represents the background and the red represents the boulders

3.2 Deployment and Parameters of Neural Network

Our neural network was deployed in Tensorflow on an Ubuntu operating system with the graphics processing unit (GPU) platform. Our desktop employs an I7 9700 central processing unit (CPU), 16 GB of memory and a GTX 1080Ti GPU. The Adam (Kingma and Ba, 2015) optimizer was utilized to allow the DNN to be rapid coverage. The neural network processes 100 epochs to ensure sufficient training for the scene.

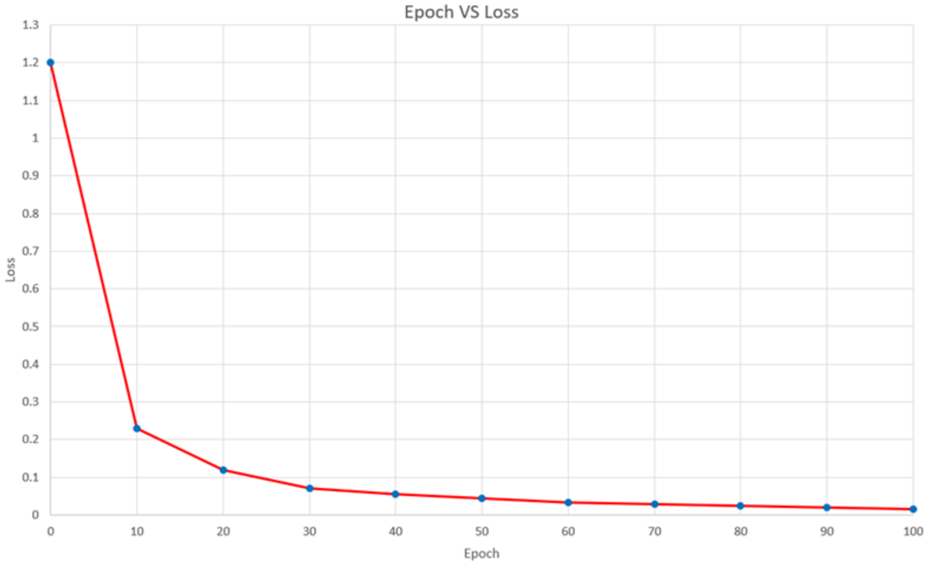

Figure 5. Training Procedure of Our Proposed Neural Network

3.3 Results of Semantic Segmentation

The training procedures are shown in Figure 5. The average loss of each epoch sharply decreases in the first 10 epochs and remains at around 0.2, then it gradually declines until 50 epochs which is less than 0.03. In the final 100 epochs, the loss of the neural network keeps slightly decreasing which reaches 0.003 at the end of training. Due to the limited number of training data, we employed the data augmentation for increasing the number of samples.

Table 2. Quantitative Evaluation of the Boulder Detection

Items	Accuracy
Precision Rate	94.6%
Recall Rate	93.5%
MIoU	82.1%

There are three indices, precision rate, recall rate, MIoU which were employed in this evaluation. The precision rate is defined as follow:

$$P = \frac{TP}{TP+FP} \tag{6}$$

The precision rate is calculated by the true positive (TP) result divided by the amount of TP and false positive results (FP). The precision rate shows the accuracy of detected targets. Another index is the recall rate which is defined as follow:

$$R = \frac{TP}{TP+FN} \tag{7}$$

The recall rate is counted by the TP divided by the amount of TP and false negative (FN) results. The recall rate illustrated the rate of detected targets. The final index Mean Intersection over Union (MIoU), is defined as follow:

$$\mathrm{MIoU} = \frac{1}{k+1}\sum_{i-0}^{k}\frac{p_{ij}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}p_{ji}-p_{ii}} \tag{8}$$

The p_{ii} donates the overlap value and the p_{ji} donates the estimation value. The MIoU represents the pixel-level accuracy of boulder detection.

According to the results of both visualization (in Figure 6) and quantization evaluation, our proposed DNN enables to detect the rock outcrops and boulders even in complicated environments. The false positive detection rarely happens, which illustrates the promising classification accuracy of our proposed neural network. The scores of recall rate reveal that our proposed DNN identifies most boulders on the image. The MIoU score illustrates that our proposed neural network extracts the ideal boundary of the boulders. The optimized architecture enables the neural network to group the boulders in different scales. The DenseNet blocks allow our DNN to extract the features adapting to the aerial view. To demonstrate the segmentation quality more clearly, the enlarged images were revealed with more details.

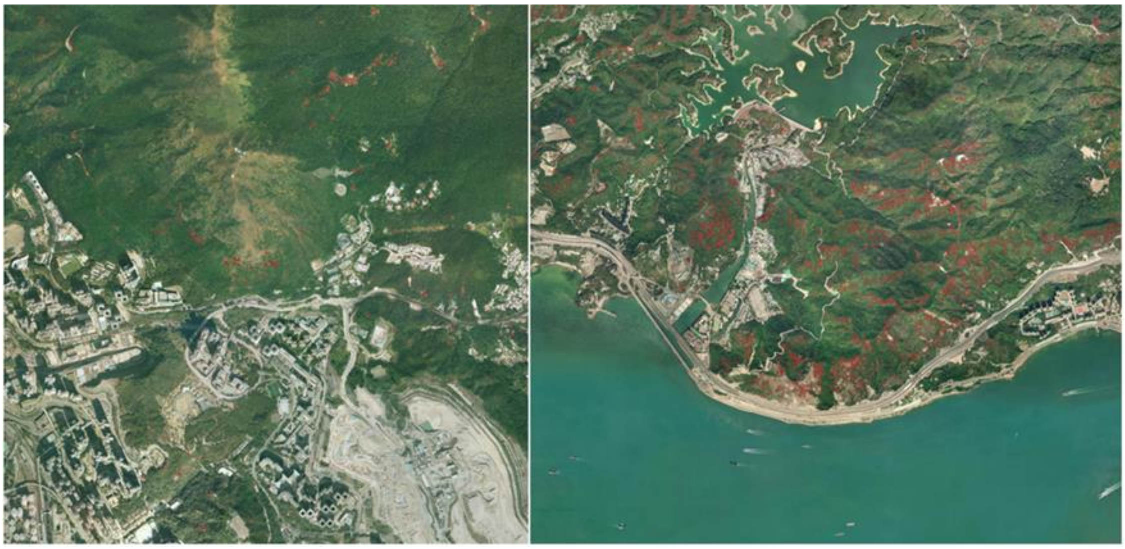

Figure 6. Visualization of Our Results

The detailed results are shown in Figure 7, our proposed neural network achieved promising segmentation results. The DNN miss-recognizes other man-made features, infrastructures, and road rarely. The rock outcrops and boulders were accurately mapped. The boundaries of the boulders were also precisely segmented. The precise segmentation and classification results illustrate that the semantic segmentation could obtain high accuracy for rock outcrop and boulder detection. One factor should be noted that only one imagery was used in this paper which is less than the norm for training the deep neural network, thus an expected improvement could be obtained if the dataset could be increased.

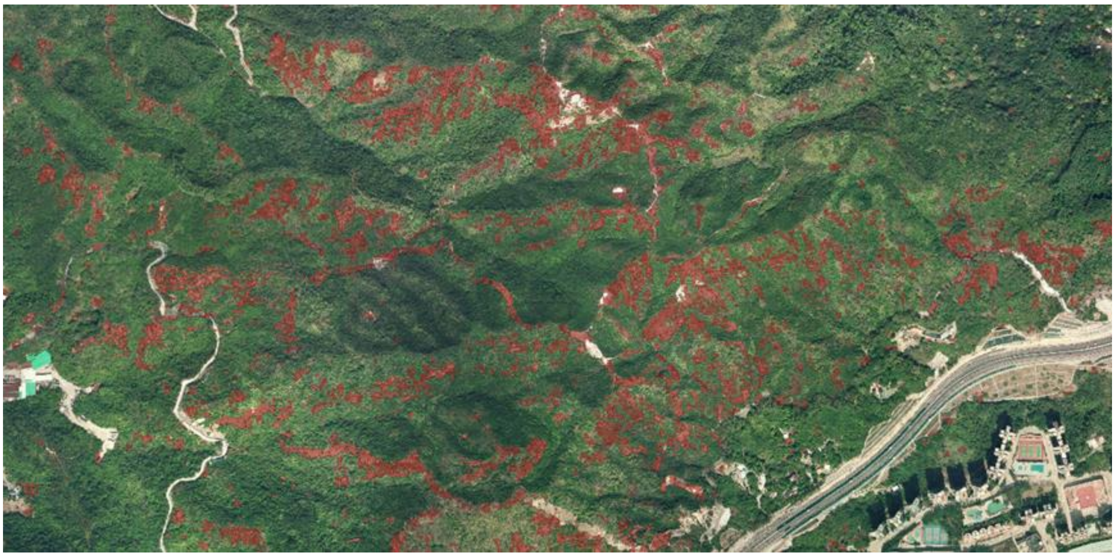

Figure 7. Detailed Visualization of Our Result

4. DISCUSSION

The developed DNN identified and segmented the rock outcrops and boulders in complicated rural areas, in Hong Kong. In the experiment, it obtained 94.6% precision rate, 93.5% recall rate and 82.1% MIoU, which illustrate the promising accuracy of the developed DNN. However, there still remains two technical limitations in this work: 1)

Labelling the training dataset is time-consuming and it requires extensive human resource; 2) Only mask image is output but corresponding information, i.e. area, size, x, y coordinates, slope of each rock outcrop and boulder requires further processing. In contrast to the existing methods, the developed method shows the self-extraction features have better quality than human-craft features.

5. REFERENCES

Afana, A., Graham, H., Davis, J., Williams, J., Hardy, R., Rosser, N., 2013. Integrating full-waveform terrestrial laser scanning into automated slope monitoring. Poster presented at XV International ISM Congress, pp.16-20.

Beraldin, J. A., 2004. Integration of laser scanning and close-range photogrammetry – The last decade and beyond. Proceedings of the XXth ISPRS Congress, 35 (B), pp. 12-23.

Bonilla-Sierra, V., Scholtes, L., Donzé, F. V., Elmouttie, M.K., 2015. Rock slope stability analysis using photogrammetric data and DFN – DEM modelling. Acta Geotechnica, 10 (4), pp. 497-511.

Carpenter, G.A., Gjaja, M.N., Gopal, S., Woodcock, C.E., 1997. Art neural networks for remote sensing: vegetation classification from Landsat TM and terrain data. IEEE Transactions on Geoscience and Remote Sensing, 35 (2), pp. 308-325.

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K. and Yuilee, A.L., 2018. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40 (4), pp. 834-848.

Chon, T.S., Park, Y.S. 2008, Self-Organizing Map, Encyclopedia of Ecology, pp. 3203-3210.

Coggan, J.S., Wetherelt, A., Gwynn, X. P., Flynn, Z.N., 2007. Comparison of hand-mapping with remote data capture systems for effective rock mass characterization. Congress of the International Society for Rock Mechanics, pp. 201-206.

Cortes, C., Vapnik, V.N. 1995, Support-vector-networks. Machine learning, 20 (3), pp. 273-297.

David., G.L., 1999. Object recognition from local scale-invariant features. Proceedings of the International Conference on Computer Vision.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Jegou, S., Drozdzal, M., Vazquez, D., Romeroa, A., Bengio, Y., 2017. The one hundred layer Tiramisu: Fully convolutional DenseNets for Semantic Segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Kingma, D.P., Ba. J., 2015. Adam: A Method for Stochastic Optimization. International Conference on Learning Representations.

Krizhevsky, A., Llya, S., Sutskever, L., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, pp. 1097-1105.

Landola, F., Moskewicz, M., Karayev, S., Girshick, R., Darrell, T., Keutzer, K., 2015. DenseNet: Implementing Efficient ConvNet Descriptor Pyramids. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Larsson, G., 2017. FractalNet: Ultra-Deep Neural Networks without Residuals. International Conference on Learning Representations.

Long, J., Shelhamer, E., Darrell, E., 2015. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical Image Computing and Computer-assisted Intervention.

Salvini, R., Riccucci, S., Gullì, D., Giovannini, R., Vanneschi, C. and Francioni, M., 2015. Geological application of UAV photogrammetry and terrestrial laser scanning in marble quarrying (Apuan Alps, Italy). Engineering Geology for Society and Territory, 5, pp. 979-983.

Sebastian, T., Franz, K., Peter, R, Uwe. S, Airborne., 2013. Vehicle Detection in Dense Urban Areas Using HoG Features and Disparity Maps. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 6 (6), pp. 2327-2337.

Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A., 2017. Inception-v4, inception-resnet and the impact of residual connections on learning. In Thirty-First AAAI Conference on Artificial Intelligence.

Yi, J.S., Prybutok, V.R., 1996. A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area. Environmental Pollution, 92 (3), pp. 349-357.

Zahabiyoun, B., Goodarzi., M.R., Bavani., A.R.M., Azamathulla., H.M., 2013. Assessment of climate change impact on the Gharesou river Basin using SWAT Hydrological model Clean. Soil, Air, Water, 41 (6), pp. 601-609.