DEAD WOOD DETECTION BASED ON SEMANTIC SEGMENTATION OF VHR AERIAL CIR IMAGERY USING OPTIMIZED FCN-DENSENET

Shenlu Jiang[1], Wei. Yao[1,2,*], Marco Heurich[3]

[1] Department of Land Surveying and Geo-informatics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China
[2] Research Institute for Sustainable Urban Development, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China
- (sljiang, wei.hn.yao)@polyu.edu.hk
[3] Dept. for Conservation and Research, Bavarian Forest National Park, 94481 Grafenau, Germany
- marco.heurich@npv-bw.bayern.de

ABSTRACT:

The assessment of the forests' health conditions is an important task for biodiversity, forest management, global environment monitoring, and carbon dynamics. Several research works were proposed to evaluate the state condition of a forest based on remote sensing technology. Concerning existing technologies, employing traditional machine learning approaches to detect the dead wood in aerial colour-infrared (CIR) imagery is one of the major trends due to its spectral capability to explicitly capturing vegetation health conditions. However, the complicated scene with background noise restricted the accuracy of existing approaches as those detectors normally utilized hand-crafted features. Currently, deep neural networks are widely used in computer vision tasks and prove that features learnt by the model itself perform much better than the hand-crafted features. The semantic image segmentation is a pixel-level classification task, which is best suitable to dead wood detection in very high resolution (VHR) mode because it enables the model to identify and classify very dense and detailed components on the tree objects. In this paper, an optimized FCN-DenseNet is proposed to detect dead wood (i.e. standing dead tree and fallen tree) in a complicated temperate forest environment. Since the appearance of dead trees generally occupies greatly different scales and sizes; several pooling procedures are employed to extract multi-scale features and dense connection is employed to enhance the inline connection among the scales. Our proposed deep neural network is evaluated over VHR CIR imagery (GSD-10cm) captured in a natural temperate forest in Bavarian national forest park, Germany, which has undergone on-site bark beetle attack. The results show that the boundary of dead trees can be accurately segmented, and the classification are performed with a high accuracy, even though only one labelled image with moderate size is used for training the deep neural network.

1. INTRODUCTION

Rapid detecting the health conditions of trees in forests is important for various studies, i.e., biodiversity and ecosystem (Lindenmayer and Noss, 2006), forest management (A. Bobiec, 2002) and carbon dynamics (P. Asante et al., 2011). In the assessment of above studies, the distribution of dead trees is one of the key indicators because many evaluations are based on living and dead components in the forest. There has been a great number of research works using remote sensing techniques to evaluate the condition of a whole forest. Currently, the aerial laser scanner (ALS) is one of the most promising methods which allows accurate identification of biomass, live and dead trees (Sterenczak.et al., 2019) through canonical correlation analysis on vertical variables. However, the expense of generating dense point clouds is greatly high and the accuracy of detecting fallen dead trees is highly dependent on the quality of generated point clouds. Furthermore, method solely using point clouds can normally only detect fallen dead trees. Multi-sensor fusion solution is essential to detect other type of dead trees. Another trend is to utilize the methods based on multispectral camera. For the task of dead tree detection, the Colour-Infrared camera (CIR) is widely utilized as the image pixel value in the near-infrared spectral channel obtained by the camera retains magnificent strength on identifying the tree health in forests (Polewski et al., 2015). However, most of current methods based on optical camera failed to accurately detect multi-class of dead trees and pixel-level tree segmentation is vulnerable to different factors,

although the benefits of pixel level detection are significant. Various parameters (such as the diameter of trees' trunk (DBH), tree height, detail of branches and leaves etc.) can be obtained for further evaluation if the boundary of dead trees can be accurately segmented.

Currently, the approaches based on deep learning obtained marvellous accuracy in various object detection tasks. Thus, we propose to apply the DNN technique to the task of dead tree detection. In this paper, a deep neural network for semantic segmentation is proposed to classify different types of dead trees (e.g. standing/fallen dead trees) at pixel-level through aerial CIR camera. There are three challenges remaining on accurately classifying the dead trees. 1.) The dead trees differ greatly in scales in aerial images. 2.) The branches and leaves of the dead trees only contain very few pixels. 3.) Two target dead tree types do not have significant difference in context. Therefore, the proposed DNN is required to be scale-invariance, precise in pixel level and equipped with powerful classification ability. The semantic image segmentation is utilized to classify dead trees. There exists a recent work (Safonova et al., 2019) that utilized the deep object detection scheme to localize damaged trees, but no detailed object boundary information can be yielded which poses the difficulty for feature derivation. Unlike object detection, semantic image segmentation is a dense classification problem for the image understanding, which allows the deep neural network to categorize different classifications in pixel-level.
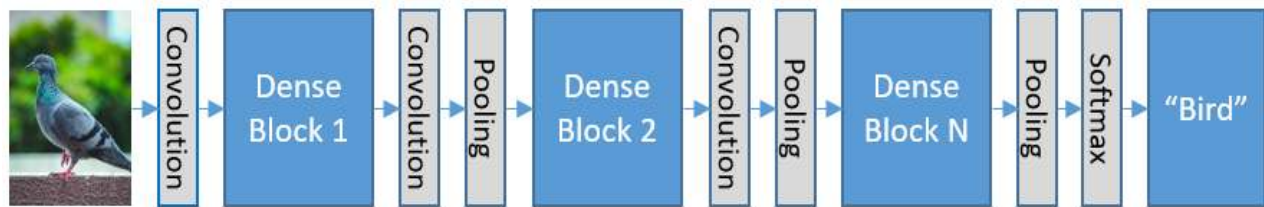
---

* Corresponding author

Figure. 1 The framework of DenseNet

Starting from FCN (Long et al., 2015), the semantic segmentation DNN employed the convolutional layer to extract from low level to high level features through down-sampling and recover the resolution by transposed convolutional layer. But reducing the map size causes the information loss of low-level features (e.g. branches and leaves). High-level features are required to be extracted by a series of down-sampling, but it leads to the difficulty on up-sampling and ambiguity of grouping the objects (e.g. accurately segmenting the large-size trees and precisely determine their semantic labels). To solve the above contradictory, different deep neural networks are proposed. Unet (O.Ronneberger et al., 2015) and SegNet (Badrinarayanan et al., 2017) utilized the skipping connection to compensate low-level features in early/middle stage feature maps to the up-sampling feature maps. However, the features in the early stage layers contain a great number of noises apart from low-level features which influences the accuracy of classification for the final output. The atrous convolution (Chen et al., 2018) proposed to utilize dilated convolution (i.e., a kind of convolution layer with varies rates of holes) to extract features without down-sampling, but the receptive field of the above methods are fixed which leads the challenges on grouping the objects. Another trend, i.e. which we also utilized, is to employ the DenseNet (Landola et al., 2015) to build up a dense inline connection among different scales' feature maps. Our proposed method applied the optimized FCN-DenseNet (Jegou et al., 2017) for this task. For the purpose of detecting dead trees of different scales, a series of down-sampling layers are designed to extract the features in multi scales. The transposed convolution layer is utilized to gradually increase the feature map to original resolution to enable the neural network to segment the branches and leaves of dead trees. The dense connection is employed to link feature maps of different scales. The feature maps for down-sampling is connected by the Dense Block. During the up-sampling, the feature map in each scale is compensated by its corresponding feature map during the down-sampling to maximize the utilization rate of low-level feature. There are two benefits that can be obtained by dense connection, i.e., the multi-scale features are sufficiently utilized, and the features can be extracted by the neural network itself. The softmax is employed to classify the types of dead trees. Different from the semantic segmentation in general view, the scales of objects can be greatly different (e.g., cover the whole image or cover very slight pixels) in aerial view. In our proposed neural network, we increase the number of down-sampling layers to allow the neural network to group the objects in higher level scales. Additionally, the channels of CIR imagery are different from the traditional RGB imagery, which requires the deep neural network to extract the features totally by itself but without using fine-turning technology. In our proposed neural network, the inner parameters (e.g., optimizer, loss function etc.) are optimized which allows the neural network to rapidly converge. Finally, the DenseNet demands exhaustive memory cost accompanying with the augment of neural network layers. Blindly increasing the depth of each layer would lead to over-fitting problem and incredible cost on GPU memory. Our optimized neural network utilizes a more reasonable depth of
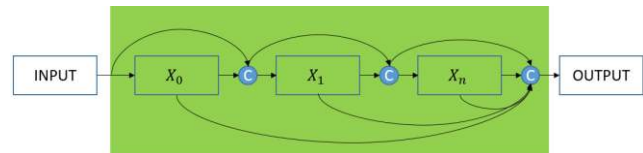

Figure. 2 The framework of Dense Block

convolutional layers which maintains effective performance without accuracy loss.

To prove the efficiency of our proposed method, the neural network is evaluated in a real-world forest dataset. There are two classes are labelled, i.e., standing dead tree and fallen dead tree in one image. The results illustrate that different types of the dead trees can be accurately classified and the boundary of them are precisely segmented in pixel level. Our proposed strategy allows the DNN self-extract the features of dead tree which are the most suitable to the environment and optimized architectures enables the DNN to precisely classify the dead trees of greatly different scales. To the best of our knowledge, this is the first work that applies the semantic image segmentation based on deep neural network to the task of dead tree detection in VHR aerial CIR imagery.

2. METHOD

In this section, the architecture and preliminary knowledge of our optimized FCN-DenseNet are explained. Our proposed DNN follows the classical architecture, i.e., FCN, in order to extract the dead trees in extremely different scales. As the pre-trained BaseNet cannot be used to provide well-extracted features in this task, we also employ the concept of DenseNet to extract the features of dead trees by the neural network itself.

2.1. Modified Residual Convolution Unit

According to the review (Voulodimos et al., 2018), DenseNet achieves the remarkable score in the ImageNet classification test (e.g. a contest which is required to classify thousands of classifications) which illustrates its amazing strength of feature extraction. Between two layers, the residual convolution unit (RCU) (He et al., 2016) is employed to avoid the gradient vanishing in the very deep neural network. Differing from original RCU, the formula of RCU in the DenseNet is defined as follow:

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

Assuming each neural layer as X_l, each X_l is residually added by its last layer X_{l-1}. Meanwhile, the concept of inception (Szegedy et al., 2017) is employed which adds a bypass connection to X_l. The architecture is employed in order to allow the dense connection from the neighbour layers. The modified RCU block enables a stronger inline relationship among the layers which is the key factor of Dense Block.
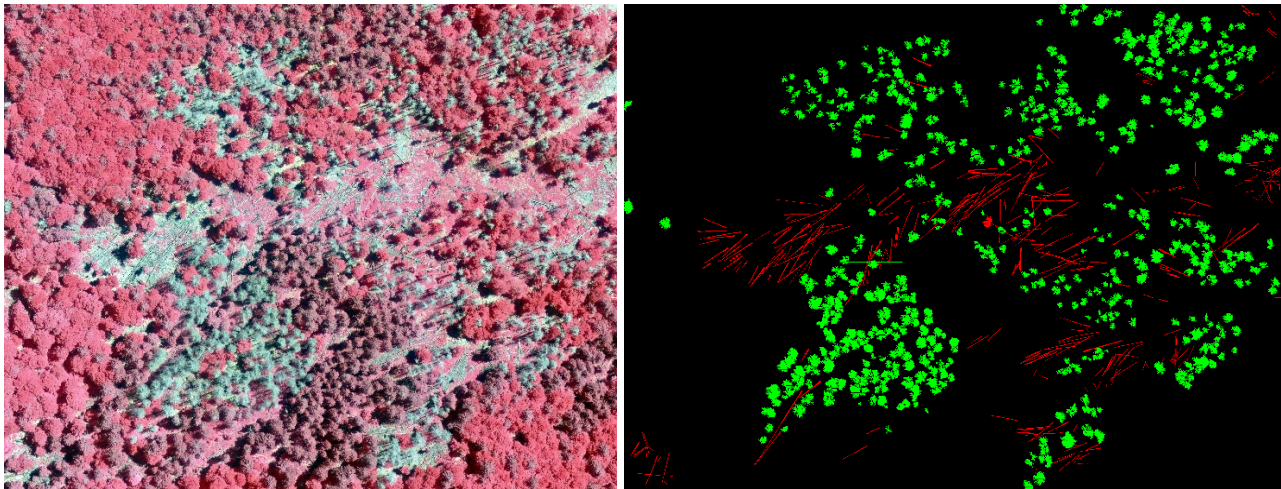
Figure. 3 The training data (left) and its corresponding labelled image (right), the red label means the fallen dead trees and the green means the standing dead trees
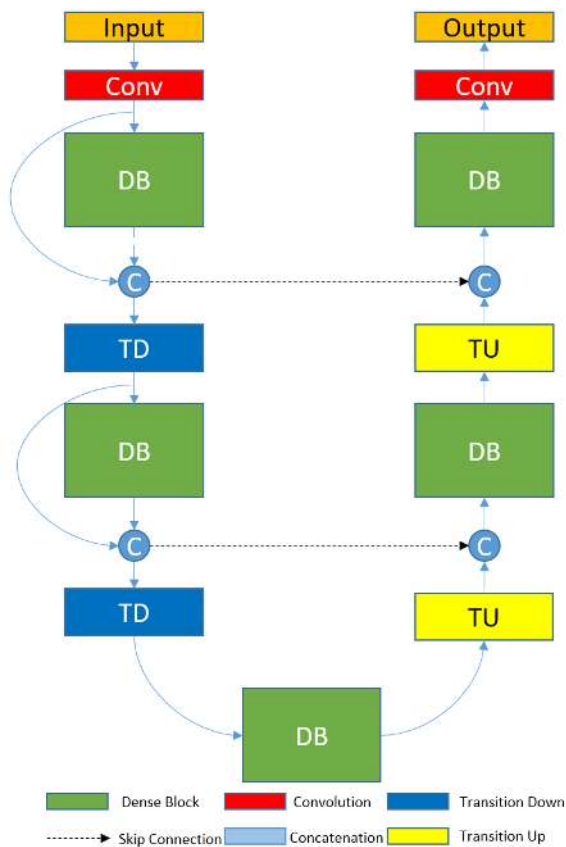

Figure. 4 The diagram of our proposed DNN

2.2. Dense Connectivity

RCU block enables the neural network to extract features in great depth, but the Fractal Net (Larsson et al., 2017) illustrated that the information redundancy also appeared during the depth increasing. Otherwise, due to that the RCU only forward transforms the feature information, the low-level features in the early stage layers are inevitably lost during the down-sampling. Therefore, the DenseNet is proposed to maximize the utilization rate of the features. The architecture of Dense Block is shown in Fig 2.

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

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 Fig 1. The original image is inputted into the deep neural network and modified by a convolutional layer. After each dense block, the pooling layers is employed to down-sample the deep neural network. Finally, the two fully connection layers and a softmax classifier are employed to classify the result. In such structure, each neural network layer is connected by all early-stage features which sufficiently utilize the different scales' features in order to increase inline relationship inside the neural network.

2.3. Transposed Convolution

While the DenseNet is usually employed for image recognition, our proposed method is utilized in the semantic image segmentation. The major difference between the two tasks is the output. The image recognition is only required to output a single value which represents the class label of image. However, the semantic segmentation requires the deep neural network to pixel-wise classify the objects and the resolution of output image remains the same as original input. Faced with such situation, the fully convolution neural network represents an up-scaling convolutional architecture, i.e. deconvolution/transposed convolution. The formula of output stride can be defined as:

$$o = s(i - 1) + k - 2p \tag{3}$$

s means the stride size, i donates the size of input, p means the padding size. In the example, the output size is calculated as 2*(3-1)+3-2=5. Comparing with other methods such as bilinear, the transposed convolution gradually adjusts the resolution of feature maps during the up-sampling which avoids the blur boundary for the output of dead trees, especially, the branches and leaves of the standing dead trees.

2.4. Skip Connection

Although the transposed convolution enables the DNN to recover the resolution of the feature map from shallow to original resolution, the up-sampling procedures is still at risk for accurately grouping the dead trees because transposed convolution leads to inevitable information loss during the up-scaling. To avoid the ambiguous classification for the both types
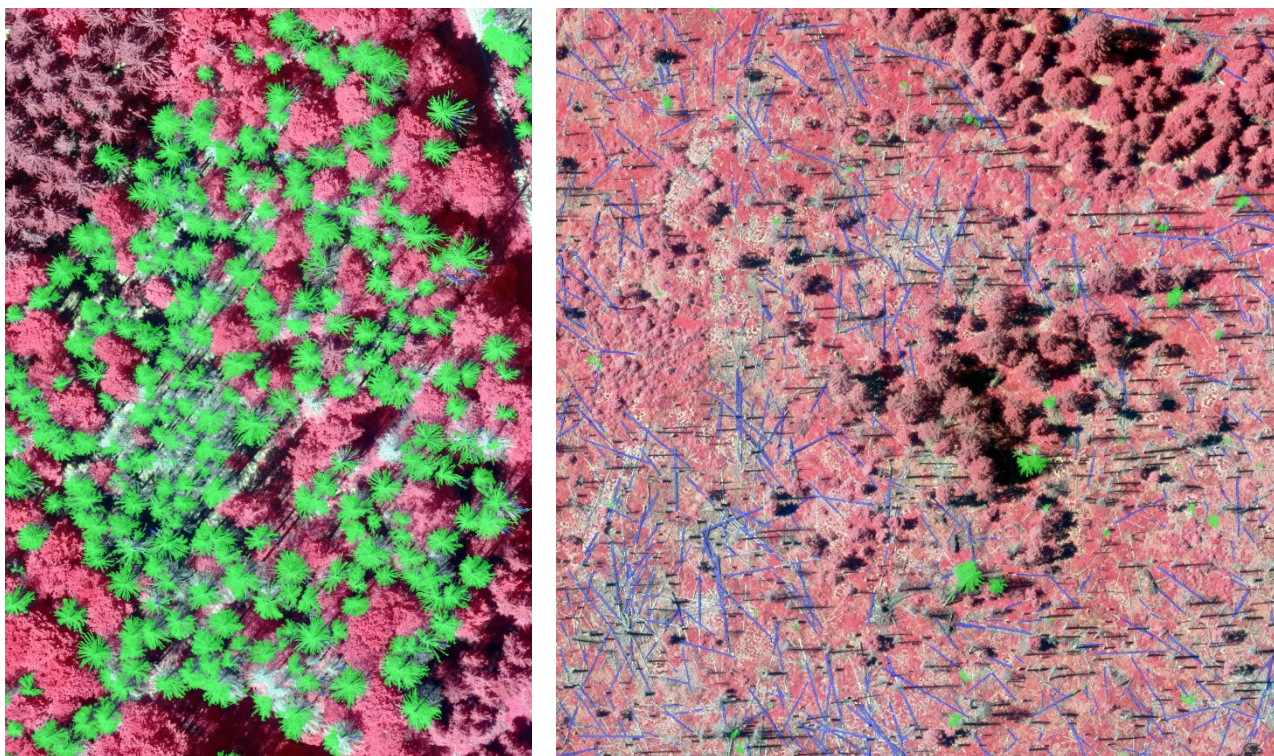
Figure. 5 Semantic segmentation of our proposed methods. Left: the green label means the standing dead trees; Right: the blue label means the fallen dead trees

of dead trees, the skipping connection is employed to compensate the low-level features to the up-scale feature map in order to increase the inline connection of the neural network and sufficiently utilize the multi-scale features. The skipping connection employs the early/middle stage feature map to bypass-connect the corresponding up-scale map and the structure already proved its advances in UNet, segNet etc.

2.5. Optimization for Dead Tree Detection

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

Table. 1 The Architecture of our proposed neural network

As the preliminary knowledge of the FCN-DenseNet is explained, we further introduce our optimization scheme for the task of dead tree detection. The architecture diagram of the proposed neural network is shown in Fig. 4 and the detailed deployment is shown

in Table 1. To enable the scale invariance, we enlarge the number of down-sampling procedures which allows the neural network to scan the objects in a larger receptive field. Moreover, since the features of two target classifications are similar, we decrease the number of each convolutional layer but enlarge the length of Dense Block in order to increase the ability of sub-classification for the neural network. Due to the increased depth and steps of down-sampling, we change the queue of the loss Function and Batch normalization of each layer, i.e., we employ the batch normalization after the Relu which is contrast to the original version. The modified queue of loss function and batch normalization proved to be more robust when the neural network is very complicated. In the Table 1, the detailed architecture of our designed neural network is shown.

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

$$d_n = L * G + d_{n-1} \qquad (4)$$

The G donates the growth rate and the L stands for the number of layers. The depth of each dense block is computed by the number of layer * the growth rate and the depth of last dense block. During the up-sampling, the skipping connection is employed to compensate each feature map with corresponding scale features. The formula can be summarized as follow:

$$d_{b+x} = L * G + d_{b-x} \qquad (5)$$

The b donates the bottom layer, i.e., the shallowest layer. The depth of each up-sampling feature map is the amount by its generated layers and its corresponding down-sampling feature map.
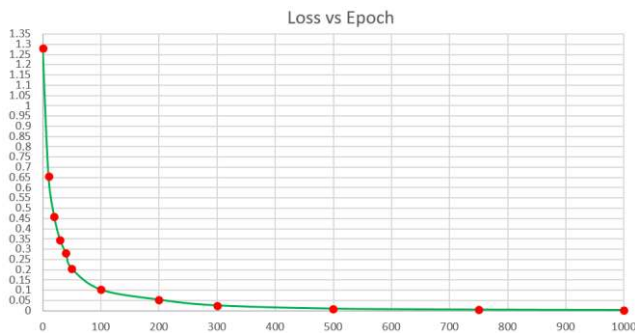
Figure. 6 Training Procedure of our proposed method

Figure. 7 Visualization of the results in Detail

Compared with other state-of-art DNNs, our proposed method holds three advantages for the task of dead tree detection. 1.) The DNN extracts the features through the dense extraction instead of utilizing the BaseNet pre-trained by common RGB channels datasets, which is more suitable to the task under specific camera, i.e., CIR camera. 2.) The neural network sufficiently utilizes the multi-scale features and compensate the corresponding features to the up-scale layers, which enables the DNN to detect extremely detailed parts (e.g., branches and leaves) and classify the objects at pixel-wise. 3.) The enlarged number of pooling steps allows the neural network to be able to group dead trees of greatly larger scale.

3. EXPERIMENTAL ANALYSIS

In this section, our proposed neural network is evaluated in our designed experiments. The image dataset was acquired in the leaf-on state during a flight campaign carried out in August 2012 using a DMC aerial camera. The mean above-ground flight height was 1900 m, corresponding to a pixel resolution of 20 cm on the ground. The evaluation dataset is collected in the Bavarian Forest National Park (49.00' 19'' N, 13. 12' 9''E), which is located in south-eastern part of Germany and close to the border between Czech Republic and Germany. The mean above-ground flight height was 1900 m, corresponding to a pixel resolution of 20 cm on the ground. The images contain 3 spectral bands: near infrared, red and green.

3.1. Training data and test data

To train the deep neural network, a sufficient number of training samples is essential. Since benchmark datasets for dead tree detection can be rarely found online, we build up and label the dataset manually by ourselves.

The dead trees were labelled in the ArcGIS as polygon shape file, then being modified to the corresponding label image as shown on Fig 3. A random region in the dataset is selected as training dataset. There are 261 fallen dead trees and 305 standing dead trees are labelled on a 4826 x 3726 image. However, the neural network cannot directly train the model with such high resolution. The image is split into a series of 512 x 512 images with stride equalling 256. To evaluate our proposed neural network, two regions (one of them is full of standing dead trees and another one is full of fallen dead trees) are selected. Due to the difficulty of labelling the dataset, we utilize the human eyes to do the quantitative evaluation of the test dataset.
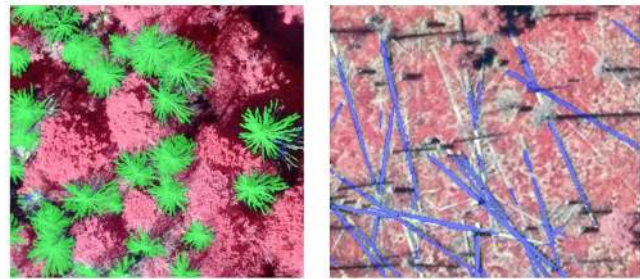
3.2. Deployment and Parameter Setting of Neural Network

Our neural network is deployed in Tensorflow on an Ubuntu operating system, which provides a good compatibility with the graphics processing unit (GPU) platform. Our desktop uses an I7 8700k central processing unit (CPU) with 16 GB of memory and a GTX 1080ti GPU, which allow a high-quality performance. The Adam (Kingma and Ba, 2015) optimizer is utilized. The neural network operates 1000 epoches in total in order to ensure sufficient and complete training of the neural network.

3.3. Results of Semantic Segmentation

The training procedures are shown in Fig. 6. The average loss of each epoch sharply decreases in the first 50 epochs and kept at around 0.2, then it gradually declines until 500 epoch which is less than 0.01. In the last 500 epochs, the loss of the neural network keeps slightly decreasing which reaches 0.003 in the end of training. Due to the limited number of training data, we employ the data augmentation, e.g. v_flip, h_flip, rotation, etc.

Data\Evaluation	Recall Rate	Precision
Dataset1	94.6%	100%
Dataset2	67.8%	99.0%

Table. 2 Object-level quantitative evaluation of fallen and standing dead trees, the dataset1 shows the result of standing dead trees and dataset2 shows the result of fallen dead trees.

The visual results obtained by applying our proposed method to two randomly selected regions are shown on Fig. 5. As the input of neural network is 512 x 512, we employed the multi-scale test technology to output the segmentation map. The results of object-level quantitative evaluation are also revealed in the Table 2. In the dataset 1, our proposed DNN successfully detects 263 standing dead trees in totally 278 samples and 0 false positive samples are detected. In the dataset 2, our proposed DNN detects 303 fallen dead trees in totally 447 fallen dead trees and 3 false positive samples are detected. The evaluation index is the precision rate whose function is defined as follow:

$$P = \frac{TP}{TP+FP} \qquad (6)$$

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 is the most direct index which shows the accuracy in detected results. Another index is the recall rate whose formula is defined as follow:

$$R = \frac{TP}{TP+FN} \qquad (7)$$

The recall rate is counted by the TP divided by the amount of TP and false negative (FN) results. According to the results of both visual and quantitative evaluations, the proposed deep neural

network achieves remarkable grouping and classification strength for the dead tree detection even in such greatly complicated environments. The false alarms almost have not occurred according to the reported scores of precision rate. The scores of recall rate reveal that our proposed DNN successfully detected most of standing dead trees in the scene while the fallen trees were outperformed. The optimized and enlarged pooling steps enabled the neural network to group the dead tree pixels in a large scale. The multi-scale feature extraction allows precise detection for multi-scale dead trees. The DenseNet blocks enable the neural network to exactly extract multi-scale features which seems to benefit the task of dead tree detection. To show the segmentation quality more clearly, the enlarged images are revealed showing the results in more details.

As shown on Fig.7, our proposed neural network achieves very promising segmentation results as well as the accurate classification of dead tree types. The branches and leaves of the standing dead trees are accurately delineated. The brunch of fallen dead trees is precisely detected. The precise segmentation and classification results illustrate that the semantic segmentation can obtain marvellous accuracy for the dead tree detection. One factor should be noted that only a very limited number of training data is used in this paper which is greatly less than that required for training the deep neural network, thus an expectable improvement can be obtained if the training dataset is enlarged.

4. DISCUSSION

Although the number of training dataset is very limited, the deep neural network still could deliver very promising results compared with the traditional methods based on handcraft features extracted from CIR and ALS. However, there still remains two challenges: 1. the dataset for semantic segmentation requires exhaustive time for labelling. 2. Although the objects are detected at pixel level, but the objects are not grouped into instance level, which two restrict its further application. Vice versa, the advantage of our proposed method is also significant. The existing methods employed the traditional hand-craft features, all of which are based on human conception. Specific features can only express single-mode characteristic, which leads to limited object expression power and restricts the robustness during detection. In opposite to traditional methods, the deep neural network excavates the features by the neural network itself which proves that it is more efficient and robust through machine learning. In addition, many of traditional methods can only perform a binary classification task, but the deep neural network easily upscales the model to categorize several classes at the same time, which increases the utilization rate of features. Therefore, we argue that the semantic segmentation will be the future trend and applied in the real world once the problem of instance detection is solved.

5. CONCLUSION

In this paper, we presented an optimized FCN-DenseNet to deal with the task of dead tree detection by aerial CIR camera. The DenseNet architecture allows the deep neural network self-extract the features which are the most suitable to the forest environment. The improved architecture enables the deep neural network to sufficiently utilize the extracted features and scan the objects in greatly different scales. Through the experiment, our proposed deep neural network obtained marvellous accuracy in both dead tree segmentation and classification even though the number of training data is rather limited. The results reveal that the semantic image segmentation DNN can obtain very

promising results for the dead tree detection and delineate the boundary of them at high-accuracy pixel-level.

However, there still remains two issues in this work. The training dataset for semantic segmentation is demanding, which leads to the exhaustive time cost for preparing the dataset. The dead tree objects are not output at instance level, which restricts its further utilization. In the future work, we plan to develop system which allows to rapidly label the dataset and detect the tree objects directly at instance level.

REFERENCES

Lindenmayer, D.B., Noss, R.F.2006. Salvage logging, ecosystem processes, and biodiversity conservation, *Conservation Biology*, 20(4).

Bobiec, A., 2002. Living stands and dead wood in the Białowieża forest: suggestions for restoration management, *Forest Ecology and Management*,165(1-3), 125-140.

Asante, P., Amstrong, G.W., Adamowicz, W.L., 2011. Carbon sequestration and the optimal forest harvest decision: A dynamic programming approach considering biomass and dead organic matter, *Journal of Forest Conomics*, 17(1), 3-17.

Sterenczak, K., Mielcarek, M., Modzelewska, A., Kraszewski, B., Fassnacht, F. E., Hislszcanski, J., 2019. Intra-annual Ips typographus outbreak monitoring using a multi-temporal GIS analysis based on hyperspectral and ALS data in the Białowieża Forests, *Forest Ecology and Management*, 422, 105-116.

Long, J., Shelhamer, E., Darrell, E., 2015. Fully convolutional networks for semantic segmentation, In *IEEE Conference on Computer Vision and Pattern Recognition,* Boston, USA. 2015. doi.org/10.1109/CVPR.2015.7298965.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In *the International Conference on Medical image computing and computer-assisted intervention*, Munich, Germany, 2015. doi.org/10.1007/978-3-319-24574-4_28.

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12), 2481-2495.

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., 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), 834-848.

Landola, F., Moskewicz, M., Karayev, S., Girshick, R., Darrell, T., Keutzer, K., 2014. DenseNet: Implementing Efficient ConvNet Descriptor Pyramids. *arXiv preprint*, arXiv:1404.1869.

Jegou, S. Drozdzal, M., Vazquez, D., Romeroa, A., Bengio, Y., 2017. The one hundred layer Tiramisu: Fully convolutional DenseNets for Semantic Segmentation. In *IEEE Conference on*

Computer Vision and Pattern Recognition, Honolulu, Hawaii, USA, 2017. doi,org/10.1109/CVPRW.2017.156.

Voulodimos, A., Doulamis, N., Doulamis, A., Protopapadakis, E., 2018. Deep learning for computer vision: a brief review. In *the Computational intelligence and neuroscience*, Montreal, Canada, 2018. doi.org/10.1155/2018/7068349.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition,* Las Vegas, Nevada, USA, 2016. doi.org/10.1109/CVPR.2016.90

Szegedy, C., Loffe, S., Vanhoucke, V. and Alemi, A.A., 2017. Inception-v4, inception-resnet and the impact of residual connections on learning. *Thirty-First AAAI Conference on Artificial Intelligence*, San Francisco, California, USA, 2017.

Larsson, G., 2017. FractalNet: Ultra-Deep Neural Networks without Residuals. In *the International Conference on Learning Representations*, Palais des Congrès Neptune, Toulon, France, 2017.

Kingma, D.P., Ba, J., 2015. Adam: A Method for Stochastic Optimization. In *the International Conference on Learning Representations*, San Diego, California, USA, 2015.

Safonova, A., Tabik, S., Alcaraz-Segura, D., Rubtsov, A., Maglinets, Y., Herrera, F., 2019. Detection of Fir Trees (Abies sibirica) Damaged by the Bark Beetle in Unmanned Aerial Vehicle Images with Deep Learning, *Remote Sensing*, 11(6), 643. doi.org/10.3390/rs11060643.

Polewski, P., Yao, W., Heurich, M., Krzystek, P., Stilla, U., 2015. Detection of single standing dead trees from aerial color infrared imagery by segmentation with shape and intensity priors, *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, II-3/W4, 181-188, doi.org/10.5194/isprsannals-II-3-W4-181-2015.