Developing a Segmenter for Hyperspectral Wood Images with Encoder-Decoder Architecture of Convolutional Neural Network

Phyu Phyu Htun
University of Computer Studies, Yangon
Yangon, Myanmar
phyuphyuhtun@ucsy.edu.mm

Ah Nge Htwe
University of Computer Studies, Yangon
Yangon, Myanmar
anhtwe@ucsy.edu.mm

Tammam Tillo
Department of Electronics and Communication Engineering
Indraprastha Institute of Information Technology,
Delhi, India
tammam@iiitd.ac.in

***Abstract*— The aim of this paper is to propose a segmenter which can predict pixel wise class label of heartwood and sapwood area in hyperspectral wood images. This process is developed by using skip connection in an encoder-decoder architecture. We compared the results with respect to the number of skip connection in the segmenter. According to the results, one skip connection is enough for our proposed segmenter.**

Keywords—hyperspectral images, convolutional neural network, image segmentation, skip connection

I. INTRODUCTION

Hyperspectral image (HSI) information is presented in the format of three-dimension hypercubes with different electromagnetic wavelengths. The three-dimension hypercube might be viewed as being made of two dimensions representing spatial information, and the third dimension representing spectral information. Hyperspectral imaging is achieved by capturing an image at different electromagnetic wavelengths. The wavelength range might cover the visible spectrum and it might include ultraviolet (UV) and it might include long wave infrared (LWIR) wavelengths. The most well-known are the visible, near-infrared, and mid-infrared wavelength bands. A hyperspectral imaging sensor might capture images with narrow and contiguous wavelengths within a specified spectral range [2].

Hyperspectral images might get used to detect objects in the scene and it might get used to classify different areas in a scene. For example, in the wood industry hyperspectral images might get used to identify areas in the wood boards which are regarded as suitable for certain usage. This kind of task, this is to say classification of different areas in an image, might get regarded as a segmentation task of the image.

There are two types of image segmentation. Semantic segmentation performs pixel-wise class labels with a set of object categories (for example, cats or dogs) for all pixels of the image. Whereas, instance segmentation aims to identify for each object in the set of objects categories the instances of the object [1].

Nowadays, Convolutional Neural Networks (CNN) are applied in hyperspectral images segmentation due to their high accuracy results. But these methods might require high computation resources.

As mentioned before, an application domain of hyper_rspectral imaging segmentation is the segmentation of HSI wood images. In this domain, the most common classes to be recognized are heartwood and sapwood. Sapwood transports water and dissolved minerals between the roots and the crown of the tree, it contains moisture and living cells, hence it is usually lighter in color. Heartwood is more resistant to attacks (by insects and fungi), and give the wood a distinctive darker color. Thus, the heartwood is preferred by stakeholders working in this domain [3][1].

For this paper, we proposed a segmenter to predict heartwood and sapwood area in HSI wood images. Firstly, we made a spatial classifier for hyperspectral wood images by adopting a CNN network, Cifar10Net, which can accept three bands images. The adopted classifier provides a good accuracy results and did not use high computing machine for training [3]. Secondly, the classifier was used to make as a segmenter by using encoder-decoder architecture of CNN.

II. DATA DESCRIPTION

In this work, we used hyperspectral images of eucalyptus wood, where the dimension of the wood board is 200cm × 25cm × 2cm. These images were acquired over the visible and near-infrared range (400-1000 nm), using a push-broom hyperspectral imaging system, the image spatial size is 2500×512 pixels and they contain 320 bands. There are 14 wood boards for training data and 6 wood boards for testing data.

In this works, we extracted 96 × 96 sub-cuboid data. For training process, we used 264 sub-cuboids which were extracted from training boards. For testing, we used 132 sub-cuboids extracted from 6 testing wood boards. The data related to wood boards should be found at [4].

III. THE SEGMENTER

The architecture of the segmenter has been shown in Fig.1. The segmenter might be viewed as being made of two parts. Let the left part be called decision part or encoder, and the right part let it be called up sampling part or decoder.

The segmenter might use skip connection. The encoder is made of an input layer, one max-pooling layer and convolution layers. Each of these convolution layers is followed by ReLU. Whereas on the decoder side, there are two transposed convolution layers, five convolution layers, and

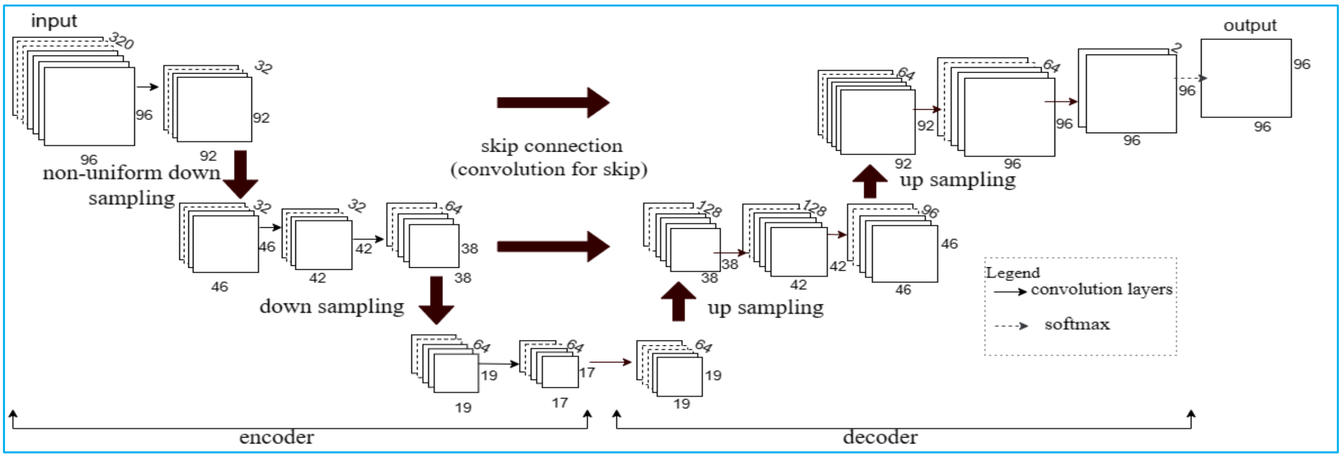

Fig. 1. Architecture of the used segmente

two depth concatenation layers. As for skip connection there are one convolution layer as per skip connection.

A. *Types of Down sampling at the encoder*

In this paper two types of down sampling were used. These two types are going to be called uniform down sampling and non-uniform down sampling.

The non-uniform down sampling is used to refer to the down sampling process of feature map using max pooling layer. The max pooling layer provide output feature map which correspond to the maximum values of the input feature map.

The uniform down sampling is used to refer to a down sampling process of the feature maps using convolution process. Let suppose that a feature map is to be down sampled by a factor of two horizontally, and by a factor of two vertically. In this case a filter of size 2 × 2 is made where except one element of the filter all the remaining elements of the filter are set to zero. By convolving the feature map with this filter and by using a stride of two horizontally, and stride of two vertically, a feature map gets generated which is a down sampled version of the input feature map.

To be noted that down sampling the features generated by the first convolution layer of the encoder is done by using non-uniform down sampling. On the other hand, to down sample the feature maps which have spatial size 38 × 38 the two types of down sampling were tested.

B. *Transposed Convolution layer*

Transposed convolution layers are used for up sampling of the input features map at the decoder part. In this paper, transposed 2-D convolution layer is used at the decoder to up sample feature maps. The up sampling network uses two transposed convolution layers. The height and width of the filters at the transposed convolution layers are 2 × 2 and the stride is 2. The first transposed convolution layer has 64 filters and the second has 96 filters. In Fig.1. up arrow represents transposed convolution.

C. *Skip Connection*

The skip connection uses one convolution layer to process features from the encoder side and generates features of same spatial size as the processed features. The generated features get concatenated with feature maps of same spatial size at the decoder side.

In this paper, two approaches of using skip connection were tested. In one approach there are two skip connections. The first skip connection forward the processed features which get generated by the first convolution layer at the encoder side. The second skip connection, on the other hand, forward the processed features which have spatial size 38 × 38. Only this skip connection is used in the second approach where only one skip connection is used.

IV. Experimental Results

This section describes the experiment results on the segmenter. Three types of segmenters were used in experiments. They are a segmenter without skip connection, a segmenter with one skip connection and a segmenter with two skip connections.

To be noted that we only trained the decoder part of the segmenter and the remaining part of the segmenter has not been retrained. This means that we reused the spatial classifier [1] weights and biases to reduce the training time. The training process has been done using sub-cuboids training data. Whereas, the testing was done using sub-cuboids training data, sub-cuboids testing data and whole wood board images.

To be noted that the classifier [1] which was used at the encoder side of the segmenter uses two max-pooling layers to down sample two sets of feature maps. In this paper the effect of replacing the second max-pooling layer with a convolution layer which performs uniform down sampling was tested. In reference to Fig. 1, this down sampling works on the feature maps which have spatial size 38 × 38.

Table I, shows the results of using the two types of down sampling to down sample the mentioned feature maps. By seeing this results, it seems that using uniform down-sampling to down sample the mentioned feature maps is better, in terms of accuracy, with respect to using max-pooling layer. To be noted that these experiments were carried out using the segmenter which uses one skip connection.

The remaining results, in this paper, are related to using non-uniform down sampling to down sample the feature maps which have spatial size 92 × 92, and uniform down sampling to down sample the feature maps which have spatial size 38 × 38.As for the use of skip connections in the segmenter it has been found that the accuracy on testing data when two skip connections are used is 91.61 %, whereas, when one skip connection is used is 92.59 %. On the other hand the accuracy of the segmenter without skip connection is 91.47 %. These results suggest that the segmenter with one skip connection provides the highest accuracy. To be noted that in these simulations, the weights and biases of the last two convolution layers in the upsampling part, or in other words in the decoder, are randomly initialized.

Table II and III shows the accuracy results of two type of segmenters on six testing wood boards. The results are calculated based on the number of pixels of sapwood and heartwood in each wood board. Table II shows the accuracy results of the network which does not use skip connection and Table III shows the results of the network which uses one-skip connection. By comparing the results in Table II and III, it is possible to note that especially for heartwood, the correctly classified pixels by using segmenter with one-skip connection are slightly lower than the segmenter without skip connection.However, for Sapwood the segmenter with one skip connection provides around 87% average accuracy against comparing results of non-uniform and uniform downsampling at the encoder of the one skip connection segmenter around 79% for the segmenter without skip connection. Thus the overall accuracy results when using of the one-skip connection segmenter is higher by 1.38% with respect to the segmenter without skip connection. The segmentation result of one skip connection segmenter is shown for one board in Fig.2.

TABLE I. COMPARING RESULTS OF NON-UNIFORM AND UNIFORM DOWNSAMPLING TO DOWN SAMPLE THE FEATURE MAPS WHICH HAVE SPATIAL SIZE 38 X 38. THE SEGMENTER IS THE ONE WHICH USES ONE SKIP CONNECTION.

No	**Down sampling**	**Accuracy**	
		Training data	**Testing data**
1	non-uniform	96.74	92.59
2	uniform	96.88	92.73

TABLE II. SEGMENTATION RESULTS ON WOOD BOARDS OF THE NETWORK WITHOUT SKIP CONNECTON

Testing wood boards	**Heartwood**		**Sapwood**	
	Ground truth	**Correctly segmented**	**Ground truth**	**Correctly segmented**
b_55_1_ch0	418249	400255	102939	40186
b_57_0_ch0	364854	343757	95393	53169
b_70_0_ch0	194126	109630	225919	200012
b_70_1_ch0	294624	254621	82089	70122
b_94_0_ch0	452738	445000	191761	168371
b_9_1_ch0	327658	244015	162077	152075
Total pixels	**2052249**	**1797278**	**860178**	**683935**
Average accuracy		87.58%		79.51%
Overall accuracy				**85.19%**

TABLE III. SEGMENTATION RESULTS ON WOOD BOARDS OF THE NETWORK WHICH USES ONE-SKIP CONNECTION

Testing wood boards	**Heartwood**		**Sapwood**	
	Ground truth	**Correctly segmented**	**Ground truth**	**Correctly segmented**
b_55_1_ch0	418249	397896	102939	67760
b_57_0_ch0	364854	354992	95393	78359
b_70_0_ch0	194126	85944	225919	199266
b_70_1_ch0	294624	262634	82089	62629
b_94_0_ch0	452738	448179	191761	180268
b_9_1_ch0	327658	221854	162077	161549
Total pixels	**2052249**	**1771499**	**860178**	**749831**
Average accuracy		86.32%		87.17%
Overall accuracy				**86.57%**

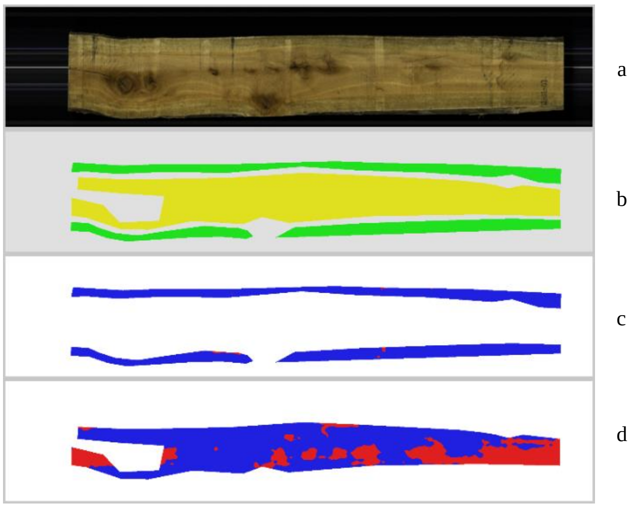

Fig. 2. Segmentation results of one skip connection segmenter. a) is RGB format of HSI wood board of the board named b_9_1_ch0, b) is ground truth, c) is the segmentation results for sapwood and d) is the segmentation result for heartwood. green color is sapwood area and yellow color is heartwood area, blue color is correctly segmented area and red color is incorrectly segmented area.

V. CONCLUSION

The proposed segmenter can predict heartwood and sapwood area on the wood board hyperspectral images with reasonable accuracy. We also proved that there is no need to train the encoder part of the segmenter. Furthermore, there is no need to train the segmenter with whole HSI wood images. Nevertheless, the proposed segmenter can segment the whole hyperspectral input image. This hopefully reduces the training complexity.

REFERENCES

[1] E.Sidorova, O.Hagman, M.Sehlstedt-Persson, O.Karlsson, and D.Sandberg, “Hyperspectral Imaging Surface Analysis for Dried and Thermally Modified Wood: An Exploratory Study”, Journal of spectraoscopy, Volume 2018 , Article ID 7423501 , https://doi.org/10.1155/2018/7423501

[2] H.Grahn and P.Geladi, Techniques and Applications of Hyperspectral Image Analysis, John Wiley & Sons, 2007

[3] P. P. Htun, M. Boschetti, A. Buriro, R. Confalonieri, B. Sun, A. Htwe, T. Tillo, “A lightweight approach for wood hyperspectral images classification”, in: 2021 IEEE International Conference on Multimedia & Expo Workshops (ICME), IEEE Computer Society, Los Alamitos, CA, USA, 2021, pp. 1–4. doi:10.1109/ ICMEW53276. 2021.9455943.URL https://doi.ieeecomputersociety.org /10.1109/ ICMEW53276.2021.9455943

[4] M. Boschetti, M. Matteo Caffini, S. Faccini, A. Buriro, R. Con falonieri, B. Sun, T. Tillo, Hyperspectral images for wood recognition, http://dx.doi.org/10.17632/2sfw446fht.2 (2022). doi:10.17632/2sfw446fht.2. URL http://dx.doi.org/10.17632/2sfw 446 fht.2.