Light U-Net with a New Morphological Attention Gate Model Application to Analyse Wood Sections

Rémi Decelle[1], Phuc Ngo[1], Isabelle Debled-Rennesson[1], Frédéric Mothe[2] and Fleur Longuetaud[2]
[1]*Université de Lorraine, CNRS, LORIA, UMR 7503, Vandoeuvre-lès-Nancy, F-54506, France*
[2]*Université de Lorraine, AgroParisTech, INRAE, SILVA, F-54000 Nancy, France*

Keywords: Wood Analysis, Mathematical Morphology, Depthwise Separable Convolution, Attention Model.

Abstract: This article focuses on heartwood segmentation from cross-section RGB images (see Fig.1). In this context, we propose a novel attention gate (AG) model for both improving performance and making light convolutional neural networks (CNNs). Our proposed AG is based on mathematical morphology operators. Our light CNN is based on the U-Net architecture and called Light U-net (LU-Net). Experimental results show that AGs consistently improve the prediction performance of LU-Net across different wood cross-section datasets. Our proposed morphological AG achieves better performance than original U-Net with 10 times less parameters.

1 INTRODUCTION

In this paper, we focus on neural networks (NNs) to segment heartwood in wood cross-section (CS) images. There are few publications on raw wood CS image analysis captured by a RGB camera. CS analysis of RGB image is relevant to estimate wood quality. More precisely, the wood quality can be defined by several properties (Barnett and Jeronimidis, 2003) among which: mechanical resistance, dimensional stability, durability and aesthetic.

All of these characteristics are unfortunately not directly measurable on CS images. However, they can be estimated by considering intermediate characteristics visible on the images. In this paper, the characteristic studied is the amount of heartwood (see Fig.1), which is related to the durability properties of Douglas-fir wood. In the Fig.1, for a better visualization, only the contour of the heartwood is marked with the blue line. In addition to a high segmentation accuracy, the time performance is also an important criterion for real world applications (both industry or scientific applications).

For segmentation from CS images, only few methods have been assessed (Decelle and Jalilian, 2020; Wimmer et al., 2021). Different convolutional neural networks (CNNs) are used in (Decelle and Jalilian, 2020) for segmenting the wood logs. They compared different CNNs for such a task on six different datasets. (Wimmer et al., 2021) also proposed a method based on CNNs. They proceed twice a CNN to increase performance. None of these studies focused on heartwood segmentation from CS images.

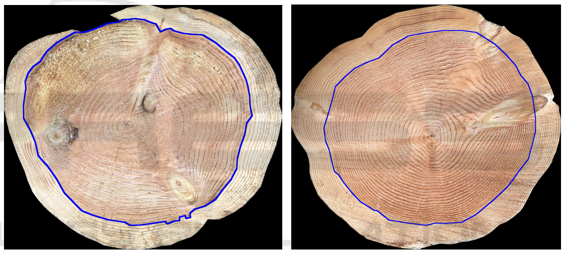


Figure 1: RGB images of CS with manually delineated heartwood contour in blue.

For the best of our knowledge, there is only one publication which focuses on heartwood segmentation on raw CS images. Raatevaara et al. (Raatevaara et al., 2020) developed a method based on region growing techniques followed by a post-processing.

In this work, we focus on heartwood segmentation. We propose a novel attention gate (AG) to evaluate CNNs with less parameters. Indeed, NNs can compute fastly the segmentation which is an important criterion in sawmill environment. Moreover, they have shown their performances in other similar tasks.

2 RELATED WORK

In this section, we recall different techniques: reducing parameters, attention mechanism and mathematical morphology for CNNs.

Decelle, R., Ngo, P., Debled-Rennesson, I., Mothe, F. and Longuetaud, F.
Light U-Net with a New Morphological Attention Gate Model Application to Analyse Wood Sections.
DOI: 10.5220/0011626800003411
In *Proceedings of the 12th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2023)*, pages 759-766
ISBN: 978-989-758-626-2; ISSN: 2184-4313

2.1 Reducing Parameters

Increasing the depth of CNNs has been regarded as an intuitive way to boost performance of the networks for different learning tasks. However, for some applications, a large CNN is not necessarily the one offering the best performance, in particular when the available dataset of training is limited. In this paper, we address the specific problem of heartwood segmentation with a small dataset, having a CNN with many parameters seems not relevant and could lead to redundancy in the features learned, which are not necessary. Many model compression techniques have been proposed to reduce parameters, delete redundancy, and/or computation time once the training is done. Moreover, having a network with a great amount of parameters increases the risk of overfitting. This is especially true when the amount of data is limited.

Network quantization is a technique for reducing parameters. It consists in quantizing filter kernels in convolution layers and weights in fully connected layers (Liang et al., 2021). Other method is knowledge distillation which focuses on transferring knowledge from a large model to a smaller one (Gou et al., 2021). Pruning technique works by removing weights whose contribution to the network performance is not significant (Luo et al., 2017).

Another approach is to design new layers. In this article, we focuses on the Depthwise separable convolutional (DSC) layer which is similar to a convolution with less parameters. More precisely, DSC consists of first performing a depthwise spatial convolution which acts on each input channel separately, and following by a pointwise convolution which mixes the resulting output channels. Such layers have been used for weather forecasting in order to obtain a lighter network (Trebing et al., 2021).

2.2 Attention Mechanism

Reducing parameters can lead to poorer performance. Adding attention gate (AG) would help to compensate for this decrease in performance. Attention mechanism is a key-role in human perception and computer vision tasks. Indeed, AGs can allocate the available resources to selectively focus on processing particular parts instead of the whole scene. Generally, there are two types of attention mechanism: spatial and channel attentions (Woo et al., 2018).

Multiple AGs are used to address a well-known weakness in convolution. Hu et al. (Hu et al., 2018) proposed the squeeze-and-excitation module and used global average-pooled features to compute channel-wise attention. Woo et al. (Woo et al., 2018) combined the spatial and channel attentions to propose a convolutional block attention module (CBAM). Their module sequentially infers attention maps along two separate paths, channel and spatial, then attention maps are multiplied to the input feature map for adaptive feature refinement, which increases the accuracy of image recognition. Oktay et al. (Oktay et al., 2018) developed a new spatial attention module (named AAG) by adding lower-level features, even though it is computationally more expensive than other AGs. Yang et al. (Yang et al., 2020) integrated channel attention and wavelet transform so that output feature maps contain frequency features. Zhu et al. (Zhu et al., 2021) highlighted the limitations by attentional activations-based models when spatial and channel features are separated. They developed a new attention module to address these limitations. Finally, Misra et al. (Misra et al., 2021) proposed to rotate an input tensor in order to capture cross- dimension interaction by using a three-branch structure. For that, the triplet attention module builds inter-dimensional dependencies by the rotation operation followed by residual transformations and encodes inter-channel and spatial information. Their module added a negligible computational time. In the experiments, we will compare our proposed AG with CBAM module, AAG module and Triplet module.

2.3 Mathematical Morphology

AGs use operators that highlight important features. Mathematical Morphology (MM) applies specific operations on images to recover or filter out different structures. MM has led to important successes in many computer vision tasks, such as filtering, segmentation, feature extraction, and so on. In this work, we will use MM operations in our AG.

Mondal et al. (Mondal et al., 2019) used morphological layer in order to emphasise or remove different structures of an image. They applied their method for de-raining images. Mellouli et al. (Mellouli et al., 2019) incorporated morphological operations in convolutional layers in order to generate enhanced feature maps. They used the method for digit recognition. Franchi et al. (Franchi et al., 2020) proposed to replace the standard max-pooling with a learned morphological pooling. Their results proved to be experimentally beneficial on MNIST dataset.

3 PROPOSED METHOD

In this section, we first recall the two MM operators: dilation and erosion. Then, we describe the proposed

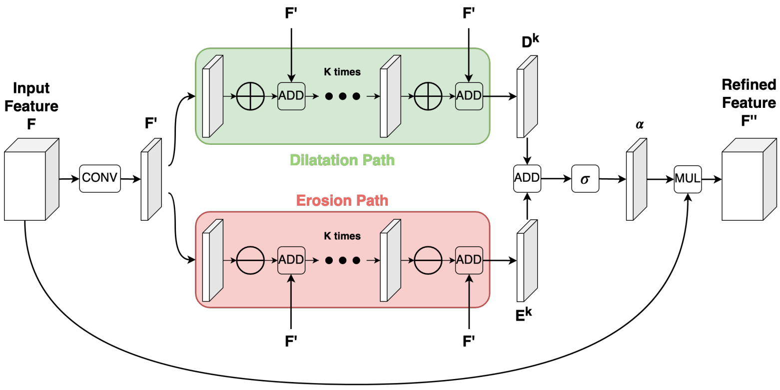


Figure 2: Diagram of proposed Morphological Attention Gate

AG. Afterwards, we present the proposed light U-Net (LU-Net) using DSC layers.

3.1 Morphological Layer

Basic MM operators are dilation and erosion. Other morphological filtering can be defined by combining these operators. In this work, we borrow the morphological layers introduced in (Mondal et al., 2019).

Let I be the input gray-scale image. Dilation $\oplus$ and erosion $\ominus$ operations for a pixel (x,y) of I are defined as follows :

$$(I \oplus \mathbf{W_d})(x,y) = \max_{i \in U, j \in V} (I(x-i, y-j) + \mathbf{W_d}(i,j)) \tag{1}$$

$$(I \ominus \mathbf{W_e})(x,y) = \min_{i \in U, j \in V} (I(x-i, y-j) + \mathbf{W_e}(i,j)) \tag{2}$$

where $\mathbf{W_d} \in \mathbb{R}^{a \times b}$, $\mathbf{W_e} \in \mathbb{R}^{a \times b}$, $U = \{1,2,\ldots,a\}$, $V = \{1,2,\ldots,b\}$ and $a,b \in \mathbb{N}$. Both W_d and W_e are respectively dilation and erosion kernel of size $a \times b$.

3.2 Morphological Attention Gate (MAG)

Our proposed AG focuses on spatial information but not channel AG. Indeed, heartwood generally is of the same colour that varies according to the species. For instance, douglas fir heartwood is in red tones. Then, AG for channel seems not very relevant.

Given an input feature map $\mathbf{F} \in \mathbb{R}^{H \times W \times C}$, where H, W and C are integers, our morphological attention gate (MAG) first infers a 2D spatial attention map $\mathbf{F'} \in \mathbb{R}^{H \times W \times 1}$ as illustrated in Fig.2. It results that $\mathbf{F'}$ is equal to:

$$\mathbf{F'} = \mathbf{W_s} * \mathbf{F}$$

where $\mathbf{W_s}$ contains the weights of a channel-wise 1×1 convolution and $*$ denotes the convolution.

We have considered two paths inside the AG. The first one uses $k \in \mathbb{N}$ dilatation layers, and the second one uses k erosion layers. We have considered an erosion (or dilation) sequence using different weights in order to remove noise or enhance information. Multiple dilation and erosion maps are useful because it may have noise in the input features that could not be remove by a single operation.

The overall dilatation path can be summarised as:

$$\forall i \in [\![0,\ldots,k-1]\!], \quad \begin{cases} \mathbf{D}^0 = & \mathbf{F'} \\ \mathbf{D}^i = & (\mathbf{D}^{i-1} \oplus \mathbf{W}^i_\mathbf{d}) + \mathbf{F'} \end{cases} \tag{3}$$

The spatial attention map $\mathbf{F'}$ is also passed in an erosion path, given an eroded map $\mathbf{E}^{k-1}$, where the dilation layer $\oplus$ is replaced by an erosion layer $\ominus$.

Since we can not know which path is more effective for noise removal in a particular situation, we further combine both to a single feature map using by a pixel-wise addition followed by a sigmoid activation σ. It results a 2D map α. Then, the refined intermediary feature α is pixel-wise multiplicate $\odot$ by the input features $\mathbf{F}$ channel by channel:

$$\mathbf{F''} = \underbrace{\sigma(\mathbf{D}^{k-1} + \mathbf{E}^{k-1})}_{\alpha} \odot \mathbf{F}$$

3.3 Network Architecture

In this section, we detail our light CNN based on U-Net (Ronneberger et al., 2015). U-Net has been widely used on small datasets and provides fine performance.

3.3.1 U-Net

U-Net is an encoder-decoder structure. The encoder part applies twice a convolution, followed by a batch

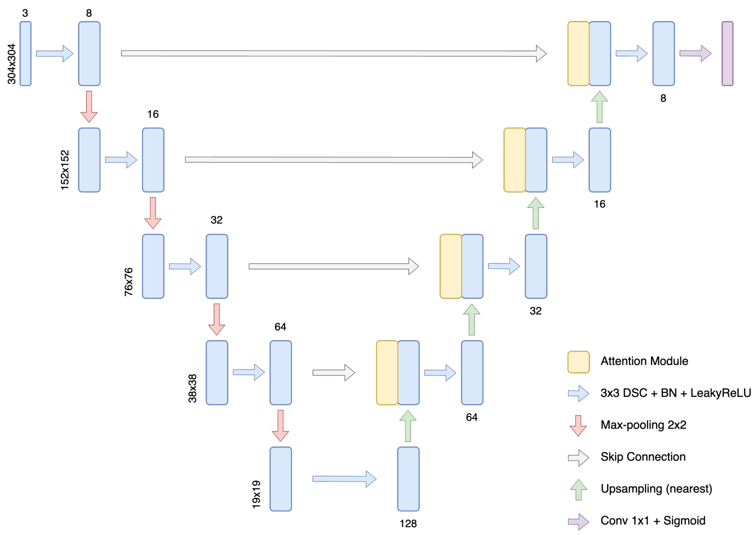


Figure 3: The proposed Light U-Net architecture.

normalization and an activation function (ReLU). Then, a max-pooling layer downsamples the image size and doubles the number of features map. The decoder part concatenates features from the encoder part with an upsampled version of lower features. As in the encoder part, the concatenation is passed in a double convolution, a batch normalization and a ReLU activation. Finally, a 1×1 convolution is applied to one output image.

3.3.2 Light U-Net (LU-Net)

Instead of performing convolution twice, we have reduced to one time. We replace each convolutional layer by DSC and change ReLU activation to Leaky ReLU. Furthermore, shakeout, a generalized dropout (Kang et al., 2018), is added to each convolutional layer. Max-pooling are used for downsampling features and nearest interpolation are applied for upsampling. The last layer is kept. LU-Net's architecture is shown in the Fig.3

3.3.3 Other Architectures

For comparison, we trained other U-Net architectures similar to LU-Net but with different AGs. We compare our module with CBAM (Woo et al., 2018), AAG (Oktay et al., 2018) and Triplet module (Misra et al., 2021). In addition, we trained the standard U-Net architecture (shakeout included). Each model has 8 features map for the first convolution.

Table 1 highlights a comparison of the models' parameters. The standard U-Net architecture has parameters that increase quadratically with the number of filters in the first layer. As it can be seen, our proposed architecture has significantly fewer parameters than the latter.

4 EXPERIMENTAL RESULTS

In this section, we describe the used datasets and then we provide implementation details. Afterwards, we compare the proposed method with the four other models[1].

4.1 Dataset

For the experimentations, two datasets (*logyard* and *sawmill*) of wood log ends CS of Douglas fir are used. These two datasets are from (Longuetaud et al., 2022). Since removing the background in order to have only the CS can be done automatically (Schraml and Uhl, 2014), (Decelle and Jalilian, 2020), (Wimmer et al., 2021), we decide to keep only the CS on the image. Images have been segmented manually to remove background. Ground truths have been done manually. The first one, called *logyard*, consists of 208 images. The second dataset, called *sawmill*, consists of 150 images of the same logs. Figure 4 shows five examples of the same logs in both datasets.

4.2 Training

All models were trained for a maximum of 100 epochs. The input size is fixed at 304×304. We used data augmentation each time. Random deformations

[1]Source code: https://gitlab.com/Ryukhaan/treetrace/-/tree/master/heartwood/deeplearning

Table 1: Cross validation MCC of the models on both datasets for the considered 8 folds.

	Model	Parameters	Relative Size	F1	F2	F3	F4	F5	F6	F7	F8	Mean	Std
logyard	**U-Net**	378,321	1 ×	0.931	0.911	0.932	**0.926**	0.915	0.900	0.933	0.925	0.921	**0.011**
	LU-Net	33,428	0.08 ×	0.883	0.906	0.926	0.881	0.927	0.926	0.888	0.920	0.907	0.019
	LU-Net + AAG	99,672	0.26 ×	0.903	**0.930**	0.939	0.911	0.906	0.942	**0.936**	0.911	0.922	0.015
	LU-Net + CBAM	44,940	0.12 ×	0.897	0.913	**0.949**	0.915	0.899	0.941	0.904	0.916	0.917	0.018
	LU-Net + Triplet	34,652	0,09 ×	0.925	0.926	0.908	0.924	0.874	0.923	0.911	**0.929**	0.915	0.018
	LU-Net + MAG	34,332	0.09 ×	**0.936**	0.923	0.924	0.925	**0.939**	**0.952**	0.934	0.907	**0.930**	0.013
sawmill	**U-Net**	378,321	1 ×	0.892	0.882	0.907	0.943	0.790	0.699	0.888	0.894	0.862	0.074
	LU-Net	33,428	0.08 ×	0.932	0.928	0.926	0.931	0.890	0.931	0.929	0.939	0.926	0.014
	LU-Net + AAG	99,672	0.26 ×	0.934	0.925	0.920	0.929	**0.931**	0.917	0.941	0.938	0.929	**0.008**
	LU-Net + CBAM	44,940	0.12 ×	0.932	**0.932**	0.916	0.931	0.928	**0.932**	**0.945**	0.920	0.930	**0.008**
	LU-Net + Triplet	34,652	0,09 ×	0.923	0.930	**0.934**	0.866	0.923	**0.932**	0.931	0.933	0.923	0.021
	LU-Net + MAG	34,332	0.09 ×	**0.941**	0.925	0.921	**0.946**	0.926	0.928	0.942	**0.940**	**0.934**	0.009

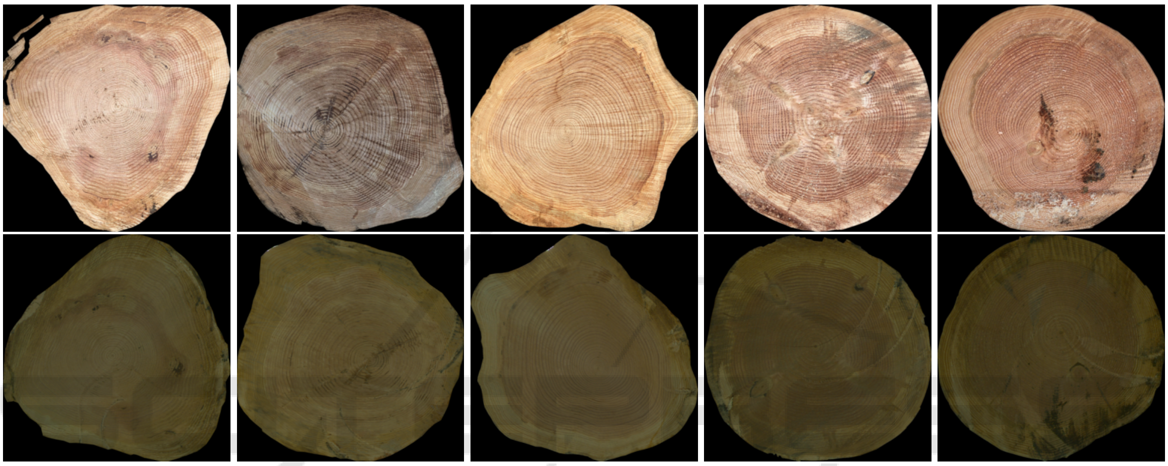


Figure 4: On the first row, images from *logyard* dataset. On the second row, images from *sawmill* dataset.

are proceeded: scaling, rotation, vertical and horizontal shift, zooming and shearing. The data augmentation is done for each batch passed to the network. The augmentation is also done on the validation set. The initial learning rate was set to 0.001 and Adam optimizer was used with default values ($\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\varepsilon = 1e{-}7$). Shakeout's parameters are $\tau = 0.1$ and $c = 0.1$.

The training was done on a single NVidia 2070 Super with 8Gb of VRAM. The used loss function is the Matthews coefficient correlation (MCC) introduced in (Abhishek and Hamarneh, 2021):

$$\mathcal{L}_{MCC} = 1 - \frac{\sum \hat{y}_i y_i - \frac{\sum \hat{y}_i \sum y_i}{N}}{\sqrt{F}}$$

$$F = \sum \hat{y}_i y_i - \frac{\sum \hat{y}_i (\sum y_i)^2}{N} - \frac{(\sum \hat{y}_i)^2 \sum y_i}{N} + \left(\frac{\sum \hat{y}_i \sum y_i}{N}\right)^2$$

where N is the number of samples, y_i is the value of the ground truth and $\hat{y}_i$ is the value of the prediction. The output is a mask representing the area of heartwood. This loss tackles the class imbalance problem. It has been shown to improve performance.

4.3 Results

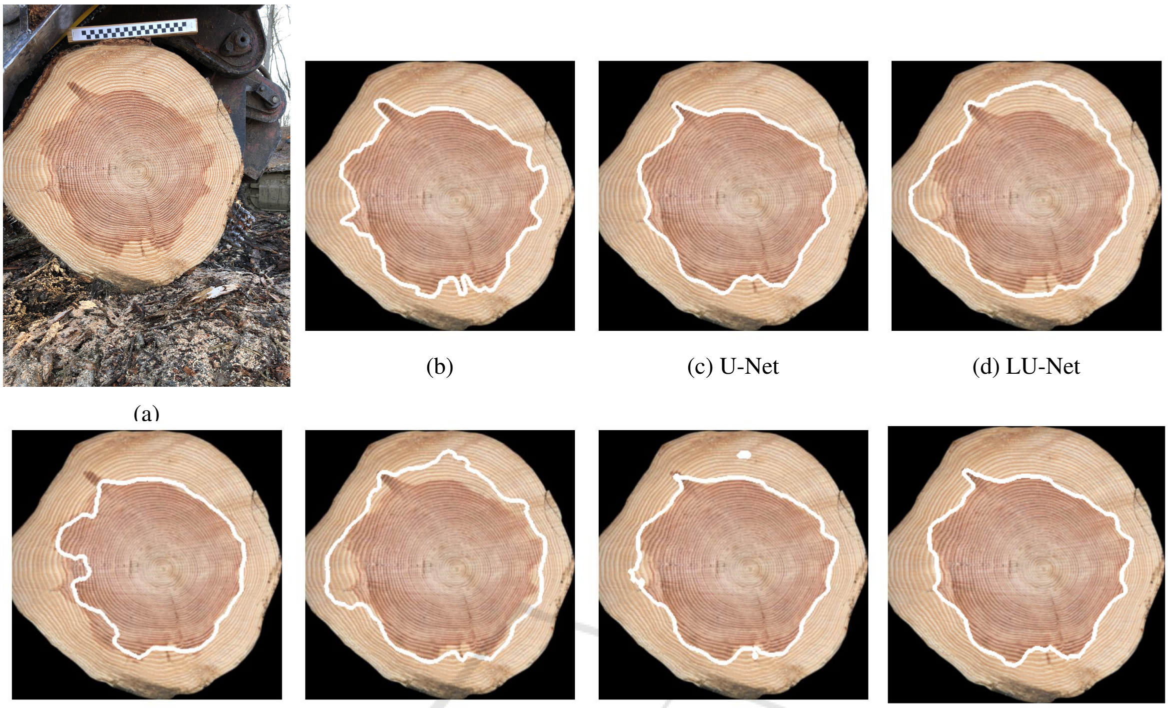
Experimental results have been proceeded using a 8-fold cross validation on both datasets. We take 6 fold for the training set (respectively 156 images for *logyard* dataset and 113 images for *sawmill* dataset), one for the validation (resp. 26 images and 19 images) and one for testing (resp. 26 images and 18 images).

The best results have been obtained with $k = 3$ (see Eq.3) and kernel of size 7×7 for both erosion and dilation layers. In addition to the MCC loss, we calculate the MCC score after thresholding the predicted image:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

where TP is True Positive, TN is True Negative, FP is False Positive and FN is False Negative.

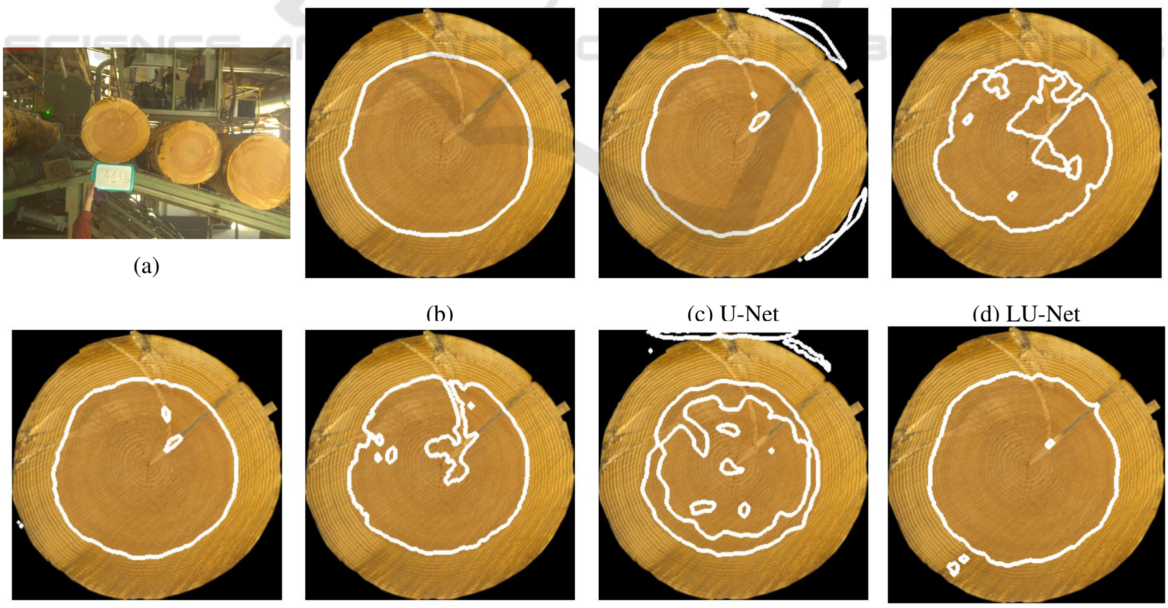
Table 1 shows the MCC score for each fold of the cross validation for both datasets. For *logyard* dataset, LU-Net is less accurate than the original one. However, when AGs are added, LU-Net shows better results. U-Net is more stable than other models, it



(a) (b) (c) U-Net (d) LU-Net

(e) AAG (f) CBAM (g) Triplet (h) Our

Figure 5: One example of an image from *logyard* dataset. (a) Original image. (b) Input image with removed background and contour of the ground truth. (c) Output from U-Net. (d) Output from LU-Net. (e)-(g) Output from LU-Net with additional attention gate : (e) AAG (Oktay et al., 2018), (f) CBAM (Woo et al., 2018) and (g) Triplet (Misra et al., 2021). (h) LU-Net with our proposed attention gate.



(a) (b) (c) U-Net (d) LU-Net

(e) AAG (f) CBAM (g) Triplet (h) Our

Figure 6: One example of an image from *sawmill* dataset. (a) Original image. (b) Input image with removed background and contour of the ground truth. (c) Output from U-Net. (d) Output from LU-Net. (e)-(g) Output from LU-Net with attention gate : (e) AAG (Oktay et al., 2018), (f) CBAM (Woo et al., 2018) and (g) Triplet (Misra et al., 2021). (h) LU-Net with our proposed attention gate.

Table 2: Mean computation time (in ms) to proceed one image for each network.

U-Net	LU-Net	+AAG	+CBAM	+Triplet	+MAG
188	67	76	97	81	89

provided the lowest standard deviation over the folds. Our proposed AG gives the best average score and outperforms on 3 of the folds.

For the *sawmill* dataset, we observe that U-Net gives low performance. LU-Net outperforms U-Net. A possible explanation for these results is that images are of very low contrast. Furthermore, in RGB images each channel is correlated to each other. As a result, convolution filters fail to highlight the heartwood. Instead, the DSCs applies a convolution filter on each channel then merge them by a pointwise operator. They act like a colour deconvolution (Ruifrok et al., 2001).

Fig.5 and Fig.6 highlight the outputs provided by the different networks for the same image in each dataset (from the testing set). For a better visualisation, only the countour of the heartwood is shown. In the Fig.5 we see that U-Net provides a good segmentation. As excepted, the lighted one is less precise. Adding an attention module increasing the precision of the light U-Net, as execpted. However, AAG underestimates the heartwood in this case, CBAM overestimates, Triplet proposes an unconnex heartwood (a small part of sapwood is considered as heartwood). But our proposed module performs as good as U-Net and even better.

In the Fig.6, the heartwood is not as coloured as in the Fig.5. We can see U-Net has considered a part of the background as the heartwood. LU-Net is clearly worse with many holes inside the heartwood, despite that the heartwood in the background has been removed. AAG is goo. CBAM performs as well as LU-Net. Triplet module is the worst output for this image, the center of the heartwood is missing. Finally our module is as good as AAG module, but our module consider a part of the sapwood as heartwood.

Table 2 shows the mean computation time for each network. The first thing we notice is that the light version of U-Net (LU-Net) is faster. It's expected since the convolution has been simplified (by using seperable depthwise convolution instead). On the contrary, the computation time increases when an attention module is added. The LU-Net with CBAM attention takes the longest execution time. Our attention module has the same time as the classical version of U-Net. In the end, taking into account the previous results, our attention module offers better results, for a minimal addition of parameters and a very small increase in computation time.

5 CONCLUSION

This paper introduced a light U-Net architecture for single-class image segmentation. Besides, we introduced an attention gate based on morphological operators (erosion and dilatation). The key is that our spatial morphological attention gate performs better than some of other attention gates used in a light network. Lightening the network leads to a significant reduction in the number of its parameters. Adding an attention gate slightly increases the number of parameters but allows to compensate the less good performances of such a light network. Erosion and dilatation are time-consuming operations. Thus, our AG is more time-consuming than usual convolution, but it marginally increases the number of network parameters. However, it provides the best results for our two datasets for heartwood segmentation of Douglas fir.

ACKNOWLEDGEMENTS

This research was made possible by support from the French National Research Agency, in the framework of the project TreeTrace, ANR-17-CE10-0016.

REFERENCES

Abhishek, K. and Hamarneh, G. (2021). Matthews correlation coefficient loss for deep convolutional networks: Application to skin lesion segmentation. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 225–229. IEEE.

Barnett, J. R. and Jeronimidis, G. (2003). *Wood quality and its biological basis*. CRC Press.

Decelle, R. and Jalilian, E. (2020). Neural networks for cross-section segmentation in raw images of log ends. In *2020 IEEE 4th International Conference on Image Processing, Applications and Systems (IPAS)*, pages 131–137. IEEE.

Franchi, G., Fehri, A., and Yao, A. (2020). Deep morphological networks. *Pattern Recognition*, 102:107246.

Gou, J., Yu, B., Maybank, S. J., and Tao, D. (2021). Knowledge distillation: A survey. *International Journal of Computer Vision*, 129(6):1789–1819.

Hu, J., Shen, L., and Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141.

Kang, G., Li, J., and Tao, D. (2018). Shakeout: A new approach to regularized deep neural network training. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(5):1245–1258.

Liang, T., Glossner, J., Wang, L., Shi, S., and Zhang, X. (2021). Pruning and quantization for deep neural network acceleration: A survey. *Neurocomputing*, 461:370–403.

Longuetaud, F., Pot, G., Mothe, F., Barthelemy, A., Decelle, R., Delconte, F., Ge, X., Guillaume, G., Mancini, T., Ravoajanahary, T., et al. (2022). Traceability and quality assessment of douglas fir (pseudotsuga menziesii (mirb.) franco) logs: the treetrace_douglas database. *Annals of Forest Science*, 79(1):1–21.

Luo, J.-H., Wu, J., and Lin, W. (2017). Thinet: A filter level pruning method for deep neural network compression. In *Proceedings of the IEEE international conference on computer vision*, pages 5058–5066.

Mellouli, D., Hamdani, T. M., Sanchez-Medina, J. J., Ayed, M. B., and Alimi, A. M. (2019). Morphological convolutional neural network architecture for digit recognition. *IEEE transactions on neural networks and learning systems*, 30(9):2876–2885.

Misra, D., Nalamada, T., Arasanipalai, A. U., and Hou, Q. (2021). Rotate to attend: Convolutional triplet attention module. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3139–3148.

Mondal, R., Purkait, P., Santra, S., and Chanda, B. (2019). Morphological networks for image deraining. In *International Conference on Discrete Geometry for Computer Imagery*, pages 262–275. Springer.

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., et al. (2018). Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*.

Raatevaara, A., Korpunen, H., Tiitta, M., Tomppo, L., Kulju, S., Antikainen, J., and Uusitalo, J. (2020). Electrical impedance and image analysis methods in detecting and measuring scots pine heartwood from a log end during tree harvesting. *Computers and Electronics in Agriculture*, 177:105690.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.

Ruifrok, A. C., Johnston, D. A., et al. (2001). Quantification of histochemical staining by color deconvolution. *Analytical and quantitative cytology and histology*, 23(4):291–299.

Schraml, R. and Uhl, A. (2014). Similarity based cross-section segmentation in rough log end images. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pages 614–623. Springer.

Trebing, K., Stanczyk, T., and Mehrkanoon, S. (2021). Smaat-unet: Precipitation nowcasting using a small attention-unet architecture. *Pattern Recognition Letters*, 145:178–186.

Wimmer, G., Schraml, R., Hofbauer, H., Petutschnigg, A., and Uhl, A. (2021). Two-stage cnn-based wood log recognition. *arXiv preprint arXiv:2101.04450*.

Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S. (2018). Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19.

Yang, H.-H., Yang, C.-H. H., and Wang, Y.-C. F. (2020). Wavelet channel attention module with a fusion network for single image deraining. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 883–887. IEEE.

Zhu, B., Hofstee, P., Lee, J., and Al-Ars, Z. (2021). An attention module for convolutional neural networks. In *International Conference on Artificial Neural Networks*, pages 167–178. Springer.