A Semi-Supervised Model for Non-Cellular Elements Segmentation in Microscopy Images of Wood

Zihao Xu*, Peter Nordström†, Sina Sheikholeslami*, Ahmad Al-Shishtawy*†, Vladimir Vlassov*
*Department of Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden
Email: zihxu@kth.se, sinash@kth.se, ahmadas@kth.se, vladv@kth.se
†RISE Research Institutes of Sweden
Email: peter.nordstrom@ri.se, ahmad.al-shishtawy@ri.se

***Abstract*—In wood science, accurate segmentation of non-cellular elements in microscopy images is critical for assessing wood quality and understanding growth patterns. Yet, it is challenging due to the complex morphology of wood components. This work explores the development of a semi-supervised deep learning model for segmenting non-cellular elements in wood microscopy images of Norway spruce, an essential source for construction materials in Europe, addressing the challenge of manual annotation's labor intensity and expertise requirement. The segmentation model employs advanced deep learning architectures, including Convolutional Neural Networks and a Vision Transformer, to capture the intrinsic patterns embedded in wood structures. We proposed a Pixel-level Guided Mean-Teacher (PG-MT) framework as an improvement to the Mean-Teacher semi-supervised learning technique. Our framework enables pixel-level guided correction to enhance segmentation accuracy and model robustness with limited labeled datasets. Our experimental evaluations show that the proposed PG-MT framework improved the Dice score for medullary ray segmentation by 0.95% and the IoU score by 1.14% over the Uncertainty-Aware Mean-Teacher (UA-MT) framework. Additionally, the integration with laboratory instruments emphasizes the model's effectiveness in accurately estimating cross-sectional cell wall thickness, demonstrating a strong correlation with X-ray measurements. This result validates the model's practical applicability in laboratory settings, enhancing the analysis of wood properties. This work provides a robust semi-supervised DL framework for segmenting non-cellular elements in wood microscopy images, significantly reducing the annotation burden and paving the way for more automated and precise wood property analysis.**

***Index Terms*—Computer vision, Deep learning, Microscopy image segmentation, Semi-supervised learning, Mean Teacher, Wood Science**

I. Introduction

Understanding the intricate structure of wood is crucial across scientific and industrial domains, from material science to forestry and wood processing. Accurately identifying and analyzing wood components in microscopy images is challenging due to their complex morphology, necessitating precise segmentation tools. Evaluating wood and fiber characteristics in laboratory settings involves examining samples to assess properties such as density and wall thickness, providing valuable insights into wood quality and variability. Precise segmentation tools enhance wood analysis, aiding in revealing growth patterns, identifying defects, and optimizing processing techniques. Accurately identifying wood's microscopic structures, including cellular and non-cellular elements, is essential for improving material performance and developing new products. While methods like StarDist [1] segment cellular components effectively, few focus on non-cellular elements. Traditionally, segmenting these elements has been a manual task prone to human error.

In recent years, Machine Learning (ML) and Deep Learning (DL) have emerged as promising approaches for the automated segmentation of complex structures in microscopy images. Convolutional Neural Networks (CNNs) have shown remarkable success in various image segmentation tasks by learning hierarchical representations directly from raw data [2], [3]. In wood microscopy images, DL-based methods offer significant advantages in effectiveness, precision, and robustness over traditional approaches. These models can learn intricate patterns in wood structures without needing handcrafted features or extensive preprocessing. However, image labeling for training DL models is time-consuming and requires expertise, posing a challenge to developing accurate models. Using semi-supervised learning to leverage unlabeled images can reduce the cost and effort of data annotation.

Based on the challenges in the previous works, this paper addresses the following research questions: Can DL models effectively automate the segmentation of non-cellular elements in wood microscopy images? Can semi-supervised learning improve segmentation performance with limited labeled images by effectively using unlabeled images?

The main contributions of this paper are as follows:

- We selected, implemented, and evaluated a Deep Learning model for precise segmentation of non-cellular elements in wood microscopy images.
- We designed the PG-MT (Pixel-level Guided Mean-Teacher) framework, a semi-supervised learning frame-

This work was supported by Bio4Energy, a Swedish strategic research environment that focuses on biorefineries for the sustainable production of renewable energy carriers, chemicals, and materials (https://bio4energy.se/). The authors would also like to acknowledge funding from Vinnova for the Digital Cellulose Competence Center (DCC), diary number 2016–05193.

work for wood microscopy image segmentation with better performance than the baseline UA-MT (Uncertainty-Aware Mean-Teacher) framework.
- We applied the proposed model to image-only cell wall thickness estimation, which proved its practicability to broader application scenarios.

The remainder of the paper is structured as follows: Section II reviews the background and related work on semantic segmentation in botanical and wood microscopy, labeled data challenges, and the potential of semi-supervised learning to improve performance. Section III describes the datasets, data preparation, and model selection. Section IV outlines the deep learning model implementations, including model parameters, loss functions, and a sliding window post-processing technique for generating segmentation maps on high-resolution wood microscopy images. Section V compares model performance, examines training methods, evaluates post-processing, and assesses cell wall thickness estimation accuracy. Section VI concludes with a summary and future work.

II. Background and Related Work

Semantic segmentation has applications across various scientific disciplines, including medical image analysis, environmental sciences, material science, and agricultural research [4].

In botanical studies, Ergun [5] applied the U-Net model to segment rays in wood microscopy images. A study on Arabidopsis thaliana used deep learning to segment hypocotyl sections for tissue morphogenesis analysis [6]. Another study introduced a large-scale microscopy dataset of potato tubers, enabling precise cell microstructure assessment with CNNs and semantic segmentation [7]. Similarly, research on Japanese hardwoods combines polarized optical microscopy with deep learning to analyze wood cell anatomy and cellulose microfibril angle [8]. CNNs have also been used to segment wood structures from micro-X-ray tomography, aiding the study of mechanical properties and moisture swelling [9].

Supervised DL methods rely on an abundance of labeled data for training a well-performing model. However, in many domains, such as medical image analysis, the acquisition of labeled data is usually quite expensive and time-consuming [10], [11]. Semi-supervised learning [12] can alleviate this challenge by utilizing both labeled and unlabeled samples.

Mean Teacher is a well-known framework in semi-supervised learning [13]. In [14], a confidence-aware mean teacher scheme was proposed. The confidence-aware module learns under the guidance of the true class probability, appended after the segmentation network. For unlabeled images, the confidence is used to regularize the consistency loss between teacher and student prediction in a finer level, by giving unconfident pixels less weights.

In [15], probability mask map segmentation and signed distance map regression tasks are jointly learned. Segmentation mask and distance map can easily transform to each other by customized sigmoid function. However, the limitation of this method is that it may only be applicable on images whose segmentation boundaries are mostly smooth and regular, otherwise the distance map may mislead the training.

The semi-supervised methods can be used in microscopy image segmentation. In the work of Dawoud *et al.*, a semi-supervised method was proposed to enhance cell segmentation in microscopy images [16]. This method significantly improved the model's performance by leveraging unlabeled data through edge detection tasks, even when only a few labeled examples were available.

The successful application of semi-supervised learning in tasks where the labeled data are scarce, in particular in the medical image analysis domain, indicates that it is possible to apply this approach to wood microscopy image analysis, where the same labeled/annotated data scarcity problems hold. By blending both labeled and unlabeled data, this approach enhances model generalization, resting on foundational assumptions like smoothness and low density to ensure effective learning from limited annotations.

III. Method

A. Data Preparation and Split

The data we are concerned with in this work are a subset from a dataset of 8170 microscopy image samples of Norway spruce (Picea abies). Each spruce image sample, collected with the SilviScan cell scanner, is in a size of 1392×1040 pixels with a bit depth of 16. Figure 1 demonstrates one sample image. Initially, 250 raw spruce samples are selected randomly from 8170 spruce samples available in the dataset.

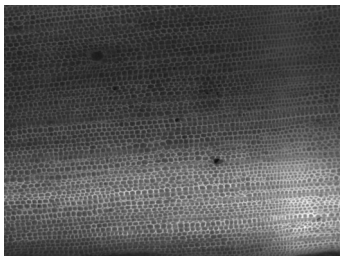

Fig. 1: Spruce sample collected with SilviScan image analyzer.

After removing images with large stains, poor lighting, or focus issues, 113 samples were retained for annotation as training and validation sets for the DL model. For each selected sample, a 256×256-pixel patch is randomly cropped to match the requirements of the model interface (see Figure 2). Annotation is done using the *ISAT-SAM* tool [17]. The set is then split into training (93 samples) and validation (20 samples) folds, with 5 distinct random seeds used to create varied training and validation set configurations.

Across the 5 training folds, background pixels average $63.42 \pm 0.39\%$, cell wall pixels $33.60 \pm 0.36\%$, and medullary ray pixels only $2.98 \pm 0.06\%$, indicating an imbalance in class proportions in the dataset.

An additional dataset is used for cell wall thickness estimation. It includes 50 stitched images, making a whole spruce sample from pith to bark, and CSV data containing

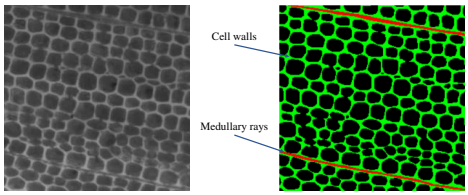

Fig. 2: A sample 256×256 image patch with corresponding annotation. The green pixels represent cell wall structure, while the red pixels represent medullary ray structure.

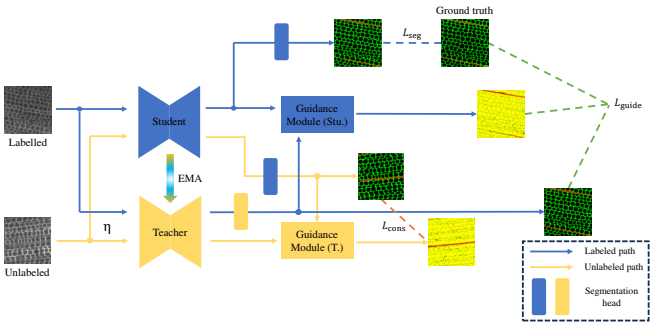

Fig. 3: An overview of the PG-MT framework.

the average cell wall thickness every 0.025mm at the cross section measured using the SilviScan method [18].

B. DL Model Selection and Justification

Three candidate frameworks, U-Net [19], Attention U-Net [20] and TransUNet [21] are used for training on annotated images for fully supervised learning.

1) **U-Net**: Known for its efficiency and effectiveness, especially in medical imaging with limited data. Its simple CNN architecture is well-suited for datasets with a distinct semantic structure and serves as a baseline for model comparisons.
2) **Attention U-Net**: Enhances U-Net with an attention gate at the skip connection, helping the model focus on relevant regions and ignore irrelevant ones, making it effective for noisy or complex images.
3) **TransUNet**: Incorporates a Transformer as the encoder, capturing global context, and follows the U-Net structure for precise segmentation. The Transformer's ability to capture long-range dependencies may lead to more accurate segmentation.

The core of the semi-supervised training framework used in this work is Mean-Teacher [13], in which the student learns from labeled data, and the teacher is the Exponential Moving Average (EMA) of the student, constrained by an additional consistency loss. We propose Pixel-level Guided Mean-Teacher (PG-MT), which adds a guidance module to generate a pixel-level guidance map that helps refine the segmentation model's prediction. Figure 3 shows an overview of PG-MT.

During training, the framework does not compute the consistency loss directly from the segmentation maps generated by the teacher and student models. Instead, it uses the guidance map created by the guidance module to instruct the segmentation backbone on which pixels have been mispredicted. This approach helps prevent scenarios where the teacher model might provide incorrect guidance to the student model, preventing model deterioration in terms of segmentation performance.

The guidance module, as shown in Figure 4, is an isolated module independent from the segmentation backbone. It takes the segmentation map and image feature representation as

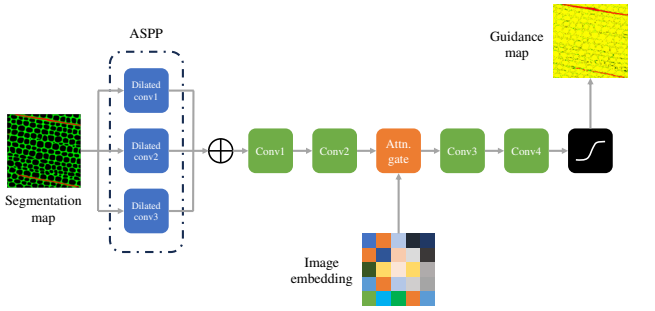

Fig. 4: Schematics of the guidance module.

input to capture the embedded information in both the image and the model, in order to provide progressive auto-correction guidance. The Atrous Spatial Pyramid Pooling (ASPP) module takes advantage of dilated convolution, which helps better capture the information with different receptive fields. The attention gate incorporates the image embedding to provide a deeper level of guidance. Finally, a sigmoid function is applied at the end of the module to produce a guidance map.

The guidance map, matching the segmentation map's spatial dimensions with an additional channel, indicates desired pixel classifications and correctness. The extra channel, when activated, preserves the segmentation; otherwise, updates occur based on a probability threshold p_{th}.

The semi-supervised framework's total loss (1) comprises supervised segmentation loss (a hybrid of focal [22] and Dice losses), guidance loss, and consistency loss. The segmentation model is trained using paired labeled data.

$$\mathcal{L} = \lambda_{\text{seg}}\mathcal{L}_{\text{seg}} + \lambda_{\text{guide}}\mathcal{L}_{\text{guide}} + \lambda_{\text{cons}}\mathcal{L}_{\text{cons}} \tag{1}$$

The guidance module is trained in a supervised fashion, where a paired data consisting of an segmentation map/image embedding and a consistency map is used as input and output. A consistency map is generated by simply comparing the segmentation map and the ground truth. Suppose in the original segmentation task, there are C semantic classes in total. Then the consistency map can be expressed as:

$$m_{\text{cons}}[i,j] = \begin{cases} \text{gt}[i,j], & \text{if } m_{\text{seg}}[i,j] \neq \text{gt}[i,j] \\ C+1, & \text{if } m_{\text{seg}}[i,j] = \text{gt}[i,j] \end{cases} \tag{2}$$

The guidance map generation task can be converted to a $C+1$-class classification task. Here, a weighted cross-entropy loss, as seen in (3), is used in the guidance loss because most pixels are valued $C+1$, which means most segmentation is right. Therefore, the weight given to this class should be reduced properly, otherwise the module will give biased prediction towards this $C+1$-th class to reduce loss.

$$\mathcal{L}_{\text{guide}} = \mathcal{L}_{\text{W-CE}}(m_{\text{guide}}, m_{\text{cons}}) \tag{3}$$

In the PG-MT framework, the consistency loss is influenced by a guidance map generated by the guidance module. This modifies the vanilla MT, where only the teacher's prediction was considered. If the guidance map strongly believes in its corrected class prediction, it will create a new segmentation map based on this correction, using (4).

$$m'_{\text{seg}}[i,j] = \begin{cases} \arg\max m_{\text{guide}}[i,j], & \text{if } \arg\max m_{\text{guide}}[i,j] \neq C+1 \\ & \text{and } \max m_{\text{guide}}[i,j] \geq p_{\text{th}} \\ m_{\text{seg}}[i,j], & \text{otherwise} \end{cases} \tag{4}$$

Apart from providing guidance to the segmentation task, the guidance map can also serve as a confidence term in the final consistency loss. The term $m_{\text{guide}}[i,j]$ is a $C+1$-length vector. Therefore, the value $m_{\text{guide}}[i,j,c]$ can be regarded as the confidence of the correction of class c at pixel (i,j).

The consistency loss is a weighted average of the cross-entropy loss between two slightly different segmentation maps:

$$\mathcal{L}_{\text{cons}} = m^s_{\text{guide}} \cdot \mathcal{L}_{\text{CE}}(m_{\text{seg}}, m'_{\text{seg}}) \tag{5}$$

Besides, for the hyperparameter before the consistency loss λ_{cons}, a Gaussian ramp-up function is utilized to increase the proportion of consistency loss from 0 to 1, to alleviate the wrong guidance of unlabeled data at the early stage of training [13]. The Gaussian ramp-up function is expressed as:

$$\lambda_{\text{cons}} = \lambda_{\text{cons-max}} \cdot \exp(-\beta(1-t/T)^2) \tag{6}$$

where β is a tunable hyperparameter that controls the shape of the Gaussian curve, and t and T represent the current training step and the total number of steps, respectively.

Specifically, the baseline Mean-Teacher used in this work is the Uncertainty-Aware Mean-Teacher (UA-MT) framework, which is proposed by Yu *et al.* for semi-supervised left atrium segmentatation [23]. It uses an uncertainty-aware loss. In order to prevent unreliable prediction produced by the teacher model from disturbing the learning process, the prediction with high uncertainty will be given less weight in the consistency loss. The uncertainty simply follows the definition of entropy, which measures the degree of disorder in the probability distribution:

$$u = -\sum_{c}^{C} \mathbf{p}_c \log \mathbf{p}_c / \log C \tag{7}$$

where C represents the number of classes, p denotes the probability of prediction, and $\log C$ serves as a scaling factor that normalizes uncertainty within the range of 0 to 1.

Then the uncertainty map is applied to the original consistency loss to produce a weighted average:

$$\mathcal{L}_{\text{cons}} = \frac{\sum_i \mathbb{1}(u_i < H)||f'_i - f_i||^2}{\sum_i \mathbb{1}(u_i < H)} \tag{8}$$

where f'_i and f_i are prediction of the teacher model and student model, respectively, $\mathbb{1}$ is the indicator function and H is the threshold of uncertainty.

IV. Implementation

The U-Net and Attention U-Net models were built from scratch [19], [24]. U-Net uses feature map resolutions of $[256, 128, 64, 32]$ and feature channels of $[64, 128, 256, 512]$, with two convolution operations (kernel size 3), batch normalization [25], and ReLU activation at each layer. Attention U-Net adds an attention gate with a resampler, utilizing bilinear upsampling with a scale factor of 2 for improved localization. TransUNet integrates a Vision Transformer (ViT) in the encoder, using a pre-trained DeiT-small model [26], pre-trained on ImageNet-1k [27], with a patch size of 16 and embedding dimension of 384. A dropout layer (0.3 probability) is included to prevent overfitting.

The supervised loss $\mathcal{L}_{\text{seg}}$ combines focal loss ($\lambda_1 = 0.7$) and Dice loss ($\lambda_2 = 0.3$). For focal loss, the class weights are $\alpha = 0.25$ for background, $\alpha = 0.75$ for cell walls, and $\alpha = 1.0$ for medullary rays, with a focusing parameter $\gamma = 2.0$ to address class imbalance.

For both UA-MT and PG-MT frameworks, the maximum consistency loss weight λ_{cons} is capped at 0.1, while the guidance loss weight λ_{guide} in the pixel-level guided version is increased to 2.

In the UA-MT framework, the uncertainty threshold H gradually increases from 0.25 to 0.75 throughout training, following the Gaussian ramp-up function in (6). In the PG-MT framework, the weighted cross-entropy loss for guidance as shown in (3) assigns class weights of $[1, 1, 1, 0.2]$, reducing the weight for the $C+1$-th class to avoid biased predictions. The smoothing factor s in the consistency loss (5) is set at 1.5 to moderate the influence of uncertain guidance, and the correction probability threshold p_{th} is set at 0.7. The ramp-up function coefficient β in (6) is set to 5. The teacher's model uses a smoothing coefficient α of 0.99.

Since the current DL model produces segmentation maps for 256×256 images, it can't directly handle higher-resolution images. This section introduces a simple post-processing approach to generate a universal segmentation map for wood microscopy images of any size.

The post-processing procedure can be dissected into three steps (see Figure 5):

1) **Sliding Window Implementation**: Begin by applying a 256x256 pixel sliding window to traverse the original high-resolution image. This window extracts segments of the image sequentially, allowing localized processing on each segment to generate predictions.
2) **Aggregating Overlapping Predictions**: As the sliding windows overlap, each pixel will generally appear in multiple windows and have multiple predictions. To

address this, calculate a weighted average of the predictions for each pixel across the overlapping windows. Ensure the overlap amount is less than 256; a recommended value is 100.

3) **Weight Distribution for Smoother Transitions**: Assign a higher weight to predictions from the central region of each window compared to those closer to the edges. This approach prioritizes the most reliable part of each window's prediction, reducing discontinuities and artifacts at the boundaries between window segments, resulting in smoother transitions in the final composite image.

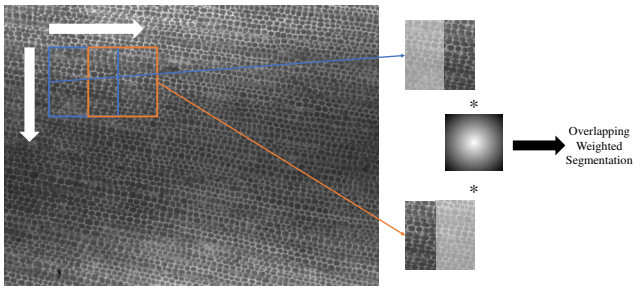

Fig. 5: Illustration of image post-processing procedure on a larger size image.

To demonstrate the impact of the sliding window boundaries on the robustness of predictions, the weight matrix is designed such that the weights in the center are larger and decrease towards the edges. The weights are proportionally distributed based on the distance from the center of the window to its four corners. The specific expression is as follows:

$$W[i][j] = \max\left(1 - \frac{\sqrt{(i-h/2)^2 + (j-h/2)^2}}{h/\sqrt{2}}, 0.001\right) \quad (9)$$

V. Results and Analysis

The performance of the three candidate DL models, U-Net, Attention U-Net, and TransUNet, are compared comprehensively using fully supervised training. Based on the comparison results, a proper model will be chosen as the segmentation model in the later semi-supervised training to obtain the best possible experimental performance. All training procedures are repeated 3 times with different random seeds on the same configuration to ensure the reliability of the results.

Table I presents the segmentation results using the aforementioned methods. Each method is assessed in terms of average Dice score and IoU score over 5 validation sets, and the scores are provided for the weighted average (Avg.), cell wall, and medullary ray segments. Adaptive Thresholding (A.T.) is an example of a traditional image analysis method.

The results in Table I are consistent with the visualized segmentation results as shown in Figure 6: Traditional method is not comparable to DL-based methods in wood microscopy image segmentation. Considering all metrics, TransUNet achieves the best average performance across the 5 validation sets and will be selected for subsequent experiments on semisupervised training.

TABLE I: Comparison of segmentation result of different supervised methods and adaptive thresholding (A.T.).

Method	Dice Score			IoU Score		
	Avg.	Cell Wall	Ray	Avg.	Cell Wall	Ray
A.T.	/	0.8115	/	/	0.6830	/
U-Net	0.8617	0.8888	0.7580	0.7655	0.8000	0.6106
A. U-Net	0.8711	0.8982	0.7714	0.7794	0.8152	0.6280
TransUNet	**0.8747**	**0.9017**	**0.7761**	**0.7849**	**0.8210**	**0.6344**

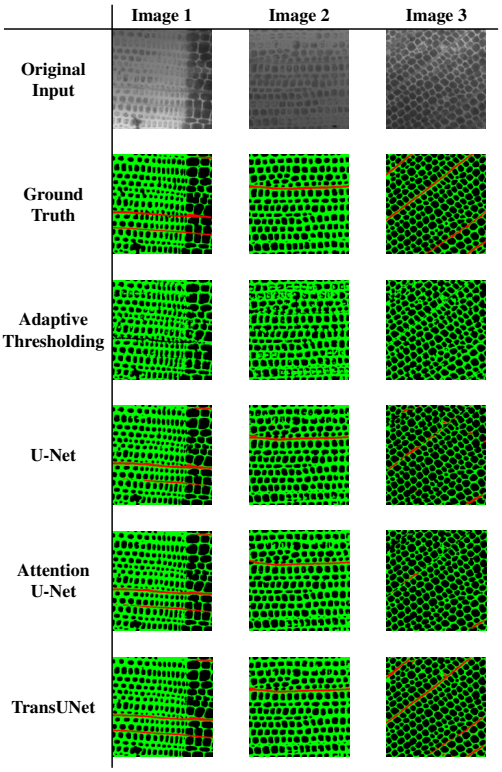

Fig. 6: Visualization of segmentation result by different methods. The adaptive thresholding method is only able to segment cell walls. TransUNet provides the most precise segmentation.

To demonstrate the performance of semi-supervised methods over purely supervised methods in the absence of sufficient labeled data, three groups of experiments are carried out: (i) supervised training using only 20% labeled data, (ii) UA-MT with a mix of 20% labeled and 80% unlabeled data, and (iii) applying PG-MT under the same data conditions. The results of the experiments are presented in Table II.

TABLE II: Comparison of segmentation result of supervised (Sup.) and semi-supervised training methods.

Method	Dice Score			IoU Score		
	Avg.	Cell Wall	Ray	Avg.	Cell Wall	Ray
Sup. (20%)	0.8342	0.8818	0.6861	0.7301	0.7886	0.5229
UA-MT (20%)	0.8525	**0.8911**	0.7218	0.7546	**0.8036**	0.5656
PG-MT (20%)	**0.8546**	0.8906	**0.7313**	**0.7568**	0.8028	**0.5770**
Sup. (100%)	0.8747	0.9017	0.7761	0.7849	0.8210	0.6344

The quantitative results in II show that, with the same amount of labeled data, the performance of the two Mean-Teacher frameworks improves significantly with additional unlabeled images compared to fully supervised methods. This indicates that, when labeling time is limited, incorporating raw unlabeled data in training can enhance performance, saving time and labor costs, though a gap exists compared to fully supervised training with all labeled images.

Between two Mean-Teacher frameworks, it can be observed that PG-MT has better performance. The advantage of PG-MT comes from its better segmentation in medullary ray structure. As mentioned before, it is because the guidance module is able to provide a finer level of correction, which corrects wrong pixels to medullary ray pixels.

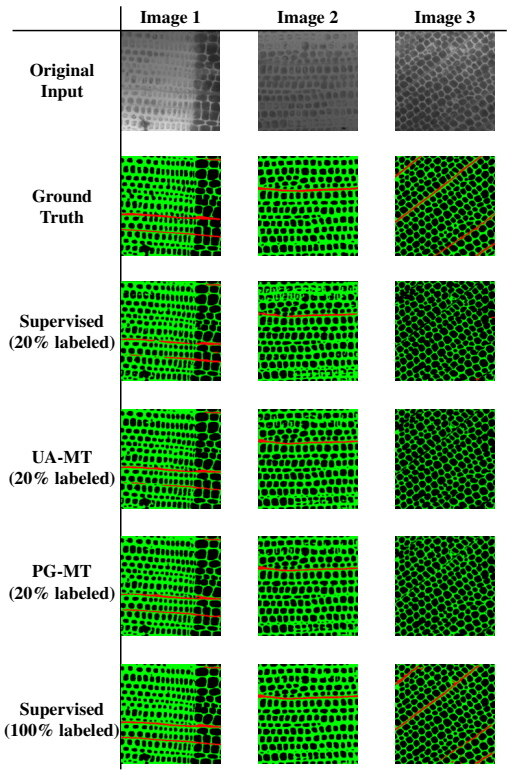

Fig. 7: Visualization of segmentation result using supervised and semi-supervised methods.

Looking at Figure 7, the model trained with only 20% labeled data in a supervised approach appears to have some inaccuracies and wrongly identifies lots of air bubbles as cell walls, which indicates that the model may suffer from insufficient training data so that it is overfitted. UA-MT framework alleviates the issue of overfitting by introducing the information of unlabeled data to the training procedure. The result shows a better alignment with the ground truth compared to the fully supervised model trained on the same amount of labeled data. Similarly, PG-MT framework also improves the segmentation quality with less noisy prediction. In Image 1, the segmentation of medullary ray is in higher accuracy, probably owing to the guidance module's ability to correct the prediction at a pixel to a specific class. However, in general, there is a limited improvement over the UA-MT framework since medullary rays in Image 3 are still not identified.

To better understand the behavior of the guidance module in the PG-MT framework, the produced guidance map is visualized in Figure 8, with consistency map as comparison.

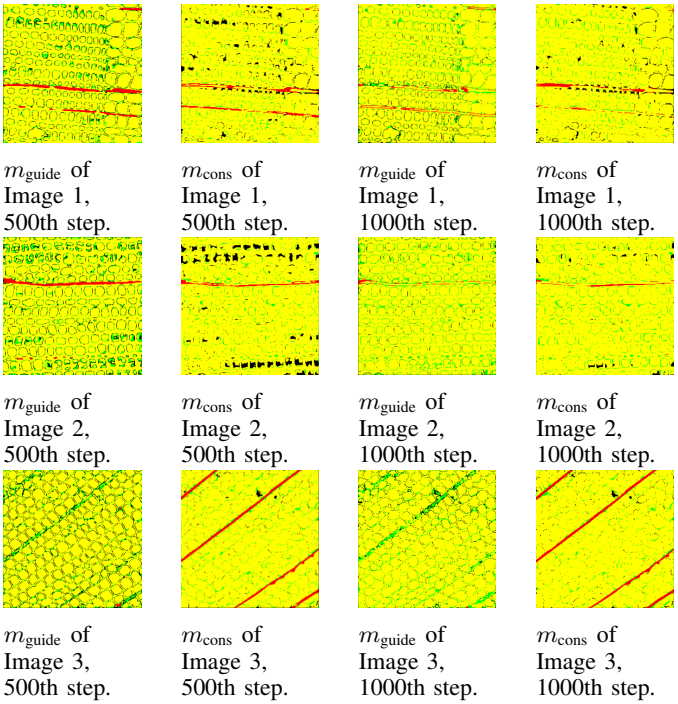

Fig. 8: Evolution of guidance and consistency maps in pixel-level guided Mean-Teacher framework throughout the training process. Yellow pixels indicate correct prediction; other colors represent guidance of correction for corresponding classes.

Figure 8 shows that the guidance module adapts to the evolving consistency map during training, ultimately identifying most pixels as "correct". However, two limitations persist: (i) corrections focus mainly on cell boundaries, often neglecting errors within cells, and (ii) the guidance map quality depends on labeled data, as evidenced by the absence of medullary ray predictions throughout training. This suggests both the segmentation model and guidance module may lack understanding of certain patterns.

To further evaluate our proposed method, we apply the previously mentioned models and methods to large-sized wood microscopy images. The goal is to estimate the thickness of the cell walls in the cross-sections, which will then be compared with ground truth measurements obtained from X-ray techniques. This allows for a comparison of the performance of different models/methods.

Figure 9 and Figure 10 demonstrate the comparison of predicted segmentation with/without post-processing. In Figure 10, there are discontinuities in the segmentation of the medullary ray regions. The proposed post-processing methods effectively address this issue by combining the predictions of multiple overlapping windows for a single pixel using

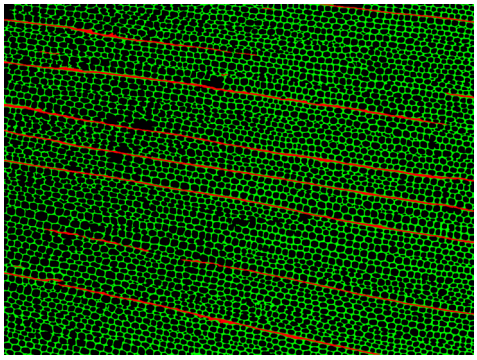

Fig. 9: Segmentation of image in Fig. 1 with post-processing.

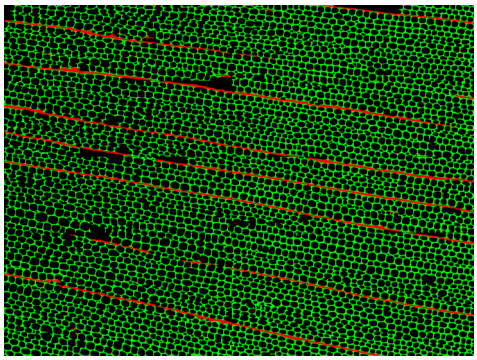

Fig. 10: Segmentation of image in Fig. 1 without post-processing. Segmentation of medullary ray is discontinuous.

a sliding window and a weight matrix. This process also considers the continuity of the image, thereby producing reliable segmentation results. Also, the application of this post-processing technique allows DL models to perform larger-scale tasks, especially for the subsequent cell wall thickness estimation experiments.

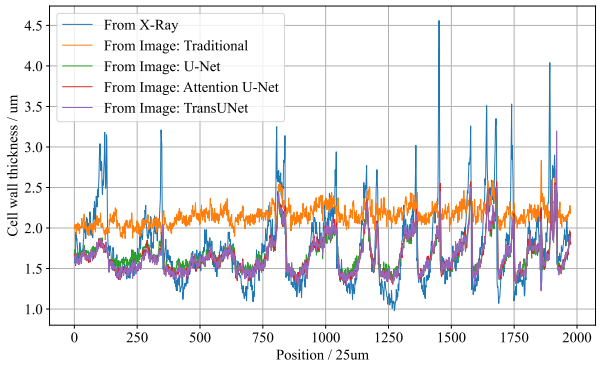

Fig. 11: Estimation of cell wall thickness along the cross section of stitched image #1 by different methods.

Figure 11 shows cell wall thickness estimation along a stitched image's cross-section. The traditional image analysis method deviates significantly from the ground truth, making it unsuitable compared to DL-based methods. While all DL models provide similar estimates, the red and purple curves align more closely with the ground truth than the green curve, indicating that Attention U-Net and TransUNet perform best.

Figure 12 compares the performance of different methods/models using a box plot. DL models outperform the adaptive threshold method in segmentation precision, with a slight performance gap between Attention U-Net and TransUNet.

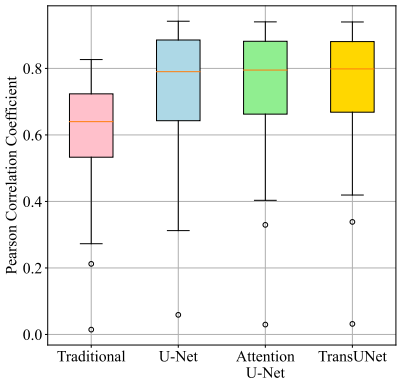

Fig. 12: Pearson correlation between the estimations and ground truth for the cell wall thickness estimation task.

Similar analysis can also be conducted for comparison between supervised methods and semi-supervised methods. Figure 13 demonstrates estimation of cell wall thickness using semi-supervised methods.

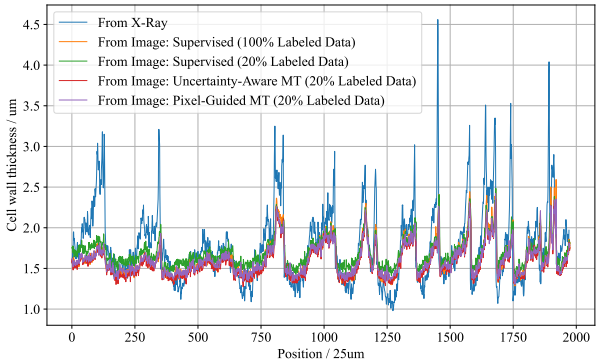

Fig. 13: Estimation of cell wall thickness along the cross section of stitched image #1 using semi-supervised methods.

It can be observed that when lacking labeled data, the estimation of cell wall thickness deviates from the ground truth, having a performance gap between the fully labeled situation. And it is also clear that two semi-supervised frameworks are reducing this gap by utilizing information of unlabeled images.

Figure 14 compares the performance between the supervised and semi-supervised methods using box plot.

It is interesting to see in 14, both two semi-supervised methods have improvement over supervised method, but there's no significant difference between the two. The reason is that Pearson correlation only focuses on the trend of variation of thickness along the path, it is not sensitive to minor changes in individual wrongly-classified pixels.

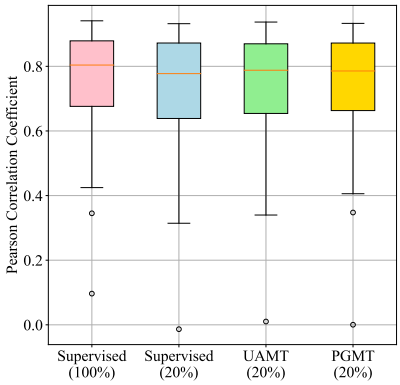

Fig. 14: Pearson correlation between the estimations and ground truth for the cell wall thickness estimation task.

VI. Conclusion and Future Work

This work tackles the automation of segmenting of non-cellular elements in wood microscopy images by integrating DL models with semi-supervised learning techniques. Experiments with Norway spruce datasets demonstrate the effectiveness of the developed DL models in enhancing wood cell analysis, highlighting their real-world relevance and impact in wood science. Key to these results was the use of advanced DL architectures like TransUNet and the Mean-Teacher framework. We propose a Pixel-level Guided Mean-Teacher (PG-MT) framework to improve segmentation accuracy and model robustness with limited labeled datasets. The semi-supervised learning component effectively leveraged unlabeled data, enhancing model performance and generalization to unseen images, reducing reliance on extensive labeled datasets and lowering annotation costs. In conclusion, integrating DL architectures with semi-supervised learning techniques improves segmentation accuracy and efficiency, enhancing wood science research without needing large annotated datasets.

Future work should expand the dataset to include more wood species beyond Norway spruce, enhancing model robustness and generalizability. Improving annotation quality and consistency through a standardized protocol or multiple annotators is also essential to reduce bias. Additionally, integrating data from different measurement techniques, such as combining X-ray density measurements with image analysis, could improve segmentation accuracy, deepen understanding of wood properties, and address technological limitations by providing additional contextual information.

References

[1] U. Schmidt, M. Weigert, C. Broaddus, and G. Myers, *Cell Detection with Star-Convex Polygons*. Springer International Publishing, 2018.

[2] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[3] Z. Li, W. Yang, S. Peng, and F. Liu, "A survey of convolutional neural networks: Analysis, applications, and prospects," 2020.

[4] S. Minaee *et al.*, "Image segmentation using deep learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3523–3542, 2022.

[5] H. Ergun, "Segmentation of rays in wood microscopy images using the u-net model," *BioResources*, vol. 16, pp. 721–728, 12 2020.

[6] A. Zakieva, L. Cerrone, and T. Greb, "Deep machine learning for cell segmentation and quantitative analysis of radial plant growth," *Cells & Development*, vol. 174, p. 203842, 2023.

[7] S. Biswas and S. Barma, "A large-scale optical microscopy image dataset of potato tuber for deep learning based plant cell assessment," *Scientific Data*, vol. 7, p. 371, 10 2020.

[8] Y. Kita, T. Setiyobudi, T. Awano, A. Yoshinaga, and J. Sugiyama, "Simultaneous cell-by-cell recognition and microfibril angle determination in japanese hardwoods by polarized optical microscopy combined with semantic segmentation," *Cellulose*, vol. 30, pp. 1–12, 07 2023.

[9] X. Arzola-Villegas *et al.*, "Convolutional neural network for segmenting micro-x-ray computed tomography images of wood cellular structures," *Applied Sciences*, vol. 13, no. 14, 2023.

[10] R. Jiao, Y. Zhang, L. Ding, R. Cai, and J. Zhang, "Learning with limited annotations: A survey on deep semi-supervised learning for medical image segmentation," 2023.

[11] X. Luo, M. Hu, T. Song, G. Wang, and S. Zhang, "Semi-supervised medical image segmentation via cross teaching between cnn and transformer," 2022.

[12] O. Chapelle, B. Schölkopf, and A. Zien, "Semi-supervised learning. adaptive computation and machine learning series," 2006.

[13] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," 2018.

[14] Z. Xie, E. Tu, H. Zheng, Y. Gu, and J. Yang, "Semi-supervised skin lesion segmentation with learning model confidence," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 1135–1139.

[15] Y. Zhang and J. Zhang, "Dual-task mutual learning for semi-supervised medical image segmentation," 2021.

[16] Y. Dawoud, K. Ernst, G. Carneiro, and V. Belagiannis, "Edge-based self-supervision for semi-supervised few-shot microscopy image cell segmentation," 2022.

[17] yatengLG, "Isat with segment anything: An interactive semi-automatic annotation tool based on segment anything," 2024. [Online]. Available: https://github.com/yatengLG/ISAT_with_segment_anything

[18] A. Scallan and H. Green, "A technique for determining the transverse dimensions of the fibres in wood," *Wood and Fiber Science*, pp. 323–333, 1974.

[19] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015. [Online]. Available: https://arxiv.org/abs/1505.04597

[20] O. Oktay *et al.*, "Attention u-net: Learning where to look for the pancreas," 2018. [Online]. Available: https://arxiv.org/abs/1804.03999

[21] J. Chen *et al.*, "Transunet: Transformers make strong encoders for medical image segmentation," 2021. [Online]. Available: https://arxiv.org/abs/2102.04306

[22] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318–327, 2020.

[23] L. Yu, S. Wang, X. Li, C.-W. Fu, and P.-A. Heng, "Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation," 2019.

[24] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention u-net: Learning where to look for the pancreas," 2018.

[25] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015.

[26] faceboookresearch, "deit," https://github.com/facebookresearch/deit, 2024.

[27] J. Deng *et al.*, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.