IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 13 January 2024, accepted 22 January 2024, date of publication 26 January 2024, date of current version 5 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3359048

APPLIED RESEARCH

Intelligent Wood Inspection Approach Utilizing Enhanced Swin Transformer

ZHIGANG DING[1], FUCHENG FU[2], JISHI ZHENG[3], HAIYAN YANG[4], FUMIN ZOU[2], AND KONG LINGHUA[1]

[1]School of Mechanical and Automotive Engineering, Fujian University of Technology, Fuzhou 350118, China
[2]School of Electrical and Electronics and Physics, Fujian University of Technology, Fuzhou 350118, China
[3]Intelligent Transportation System Research Center, Fujian University of Technology, Fuzhou 350118, China
[4]School of Computer Science and Mathematics, Fujian University of Technology, Fuzhou 350118, China

Corresponding author: Fucheng Fu (849089522@qq.com)

This work was supported in part by the School-Enterprise Cooperation Project of Fujian Jinsen Forestry Company Ltd., under Grant GY-H-20154; in part by the Forestry Technology Project of Fujian Province under Grant 2021FKJO6; and in part by the Natural Foundation of Fujian Science and Technology Department under Grant 2018J01619.

ABSTRACT Wood diameter needs to be measured in the process of production, sales and import and export. In order to solve the problem that it is difficult to accurately measure the densely stacked and irregularly arranged vehicle wood manually, this paper proposes a timber segmentation methodology that leverages a Swin Transformer model mechanism to enhance the performance of the target detection model. The method automatically learns and calculates distinct regions in the input image, assigning varying weights to different sizes and shapes of wood. This approach achieves finer detection of densely stacked logs, thereby promoting intelligent inspection and enhancing inspection efficiency.This study optimizes the backbone network by refining its modules and incorporating the operation of the log-space bias module. Additionally, improvements are made to the feature fusion network and loss function to further enhance network performance. The instance segmentation model parameters are also modified, encompassing multi-scale training, an increased number of training samples, improved image input size, and effective data widening techniques, all of which enhance log measurement accuracy and resolve the issue of partially occluded logs.This study conducts multiple control experiments to evaluate various scale metrics, such as mean average precision (mAP), log true detection rate, false detection rate, as well as comparing the root count and volume of logs through prediction. The experiments demonstrate that the mAP of this methodology reaches 0.685, and the true detection rate reaches 0.96 when compared with mainstream neural networks of similar scale, highlighting the advantages of this paper's approach in wood segmentation detection. The model exhibits a strong detection effect on dense wood, effectively overcoming occlusion challenges, leading to more accurate measurement data. Moreover, the algorithm demonstrates robustness and migration ability, rendering it highly applicable to the task of detecting and segmenting dense wood of all sizes.

INDEX TERMS Dense wood detection, Swin transformer, obscured targets, deep learning.

I. INTRODUCTION

Wood, a renewable and biodegradable resource, holds a prominent position in the contemporary world as a sustainable, eco-friendly material. It finds extensive application across diverse construction projects, standing as a pivotal driver of economic advancement. Moreover, fostering intelligent and information-driven growth constitutes a vital objective within the global forestry sector [1]. Ensuring timber accuracy stands as a vital factor affecting both the quality and economic efficiency of the timber industry, which is governed by a spectrum of regulations and standards to ensure product quality and safety. Accurate timber sizing constitutes one of the most crucial steps in adhering to

The associate editor coordinating the review of this manuscript and approving it for publication was Prakasam Periasamy.

FIGURE 1. Manual on-site detection.

these regulations and standards. Deviations from the required dimensions and specifications can lead to product rejections or penalties, directly impacting sales and market position. Additionally, inspectors are entrusted with precisely measuring and documenting parameters such as timber dimensions, length, and width, as the accuracy of their inspections directly affects their remuneration.

However, the timber gauging process entails inspecting logs of all sizes in a complex environment to ascertain their volume and grade for industrial processing. Currently, manual inspection remains prevalent, as depicted in Figure 1. This approach not only necessitates significant human and material resources but also suffers from diminished precision. Addressing this issue bears immense importance for the advancement of the timber industry.

A. IMAGE DETECTION SEGMENTATION

With the advancement of visual technology, image detection and segmentation techniques in the realm of visual processing can be broadly categorized into two directions.

The first involves conventional image detection. For instance, Yella et al. [2] employed multiple color spaces alongside geometric operators to segment timber images, extracting edge details. This facilitated the segmentation and computation of log quantities and diameters on truck beds. However, this algorithm exhibits limited adaptability to diverse and intricate environments, as its accuracy diminishes in the presence of log occlusion. Budiman et al. [3] devised a portable measuring tool to assess the minimum diameter of logs. They harnessed edge detection algorithms to discern edge pixels in separated images, effectively halving the measurement time for logs and achieving a measurement precision of 97%. Nevertheless, this method is tailored exclusively for individual logs, is sensitive to lighting conditions, and struggles with image quality constraints and intricate scenarios.

The second direction revolves around deep learning-based image detection. Tang et al. [4], for instance, utilized the SSD [5]framework for training. The resultant model proficiently detects and identifies log endpoints in natural surroundings, even in cases involving overlapping logs, external debris, and the interference of log cross-sections. Nonetheless, this approach demands high data quality, substantial datasets, and annotations, and its inference time warrants enhancement. Samdangdech et al. [6], on the other hand, merged the SSD network architecture with the FCN [7] fully convolutional network to extract pixel regions for segmenting log endpoints. They achieved a segmentation method for onboard eucalyptus tree images, boasting an average accuracy of 94.45%. This led to reduced estimation time for log quantities and lowered human costs. However, inaccuracies in localization and segmentation may arise when log endpoints exhibit cracks or are occluded by other objects, potentially resulting in misidentifications or false positives.

B. DISCUSSION AND ANALYSIS

The aforementioned approaches underscore the extensive exploration undertaken in log detection, encompassing both traditional image processing methods and sophisticated deep learning models. Despite these efforts, challenges persist in terms of limited robustness and suboptimal log detection performance.

To refine the precision of log detection, the author ventures into the realm of deep learning, delving into the viability of employing an instance segmentation model for segmenting wood end faces. This pursuit culminates in the proposition of a novel methodology designed to assess the count of logs within an entire truckload, capitalizing on an enhanced Swin Transformer [8]. This augmentation is achieved through a holistic approach, refining the backbone network, trimming model parameters, and extracting more dynamic and efficacious features. Furthermore, the method integrates the BFP [9] feature fusion network to bolster the network's feature extraction capacity, specifically catering to the detection of diminutive target entities. The loss function for bounding box regression is devised using the CIOU [10], serving as the algorithm for filtering

FIGURE 2. Dataset samples.

prediction frames.Experiments show that the improved model introduced in this paper has a good detection effect on the logs in the log scene of the whole vehicle, which significantly improves the detection efficiency of the logs, with the mAP reaching 0.685 and the true detection rate reaching 0.96. Experiments and practical application in forest farms verify the effectiveness of the proposed method and model, and provide important and unique contributions and progress for the research and practice in related fields. At the same time, this paper provides valuable insights and solutions for the future research direction, and further contributes to the wood detection algorithm under complex background.

As of the present moment, this model has been actively applied in the daily timber transportation operations of Fujian Jinsen Co., Ltd. The successful integration of the improved model has replaced manual operations, streamlining the intricate task that typically required collaboration in timber inspection to a single-person operation. This not only mitigates the safety risks associated with personnel climbing trucks during inspection but also reduces the time required for manual timber inspection per truck from approximately ten minutes to a matter of seconds. This results in heightened speed and precision, significantly diminishing inspection costs, simplifying workflow, and making a substantial contribution to industrial production.

II. MATERIALS AND METHODS

A. DATASET

The dataset utilized in this study was collected from our team's forestry site, as illustrated in Figure 2. It consists of 500 images capturing entire truckloads of timber. The number of timbers per truckload varies, ranging from 40 to 200. The timber logs were randomly distributed in size, with dimensions spanning between 5 cm to 45 cm, and the majority of the timbers falling within a medium-size range. The dataset was carefully curated to include images captured under diverse lighting conditions, with various wood end backgrounds and camera angles. After a thorough screening process, 150 high-quality images were retained as raw data, following the removal of blurred images.

TABLE 1. Annotated image dataset statistics table.

Dataset Information	All	Small	Medium	Large
Dataset	15533	3432	11753	348
Train	9320	2241	7114	247
Validation	3107	613	2296	48
Test	3106	578	2343	53

For annotation, the entire wooden logs in the images were marked using the polygon annotation tool in the Labelme software. However, the bark of the logs was not annotated. This annotation process enables the model to learn the contour features of the wood during training, allowing for accurate timber counting based on the contour mask map. With this comprehensive dataset and meticulous annotations, the model can be trained effectively to perform timber counting tasks in real-world scenarios.

All the annotated images were partitioned into training, validation, and test sets at a ratio of 6:2:2. The annotated dataset was statistically divided based on the target size categories defined in the COCO dataset, namely small targets (pixel area $< 32 \times 32$), medium targets ($32 \times 32 <$ pixel area $< 96 \times 96$), and large targets (pixel area $> 96 \times 96$). The distribution of the dataset according to these target size categories is presented in Table 1.

B. IMPROVED SWIN TRANSFORMER NETWORK

Currently, prevalent models for timber detection predominantly employ the ResNet [11] architecture. However, with the continual advancement of technology, achieving a notable enhancement in the precision and robustness of this model has become a challenging endeavor. Confronted with intricate detection tasks, there is an urgent necessity to introduce novel models for the augmentation of detection capabilities.

Recently, the Swin Transformer model, derived from the foundational architecture of the Transformer [12] model,

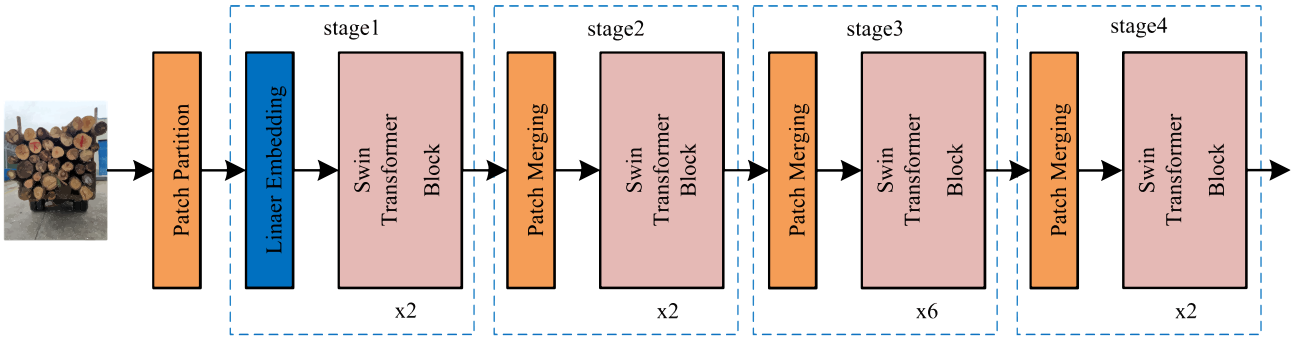


FIGURE 3. Architecture of the swin transformer model.

has recently demonstrated exceptional performance across various computer vision applications, establishing itself as a proficient solution for tasks related to target detection.This research leverages the target detection model based on the Swin Transformer, adopting the central structure of the Swin Transformer and integrating it with the Mask R-CNN [13] framework. The entire architecture of the Swin Transformer model is depicted in Figure 3, The input image undergoes a patch mapping process, initially segmented into 4×4 blocks, which are then expanded into pixel points on the channels using convolutional layers.

However, initially in practical application scenarios, the Swin Transformer did not demonstrate the expected robust performance in wood detection, revealing certain limitations associated with specific tasks. This could be attributed to various factors such as model parameters, training data, or specific application environments, primarily presenting the following issues:

1) When loading logs onto the truck using a hooker, there is a likelihood of carrying along unwanted elements like weeds, dirt, and other attachments. These foreign objects may obstruct the ends of the logs, leading to missed detections by the model.

2) The inspection process for the entire truckload of logs can be affected by external factors such as outdoor lighting conditions and the dense arrangement of the timber with variations in diameter grades. As a result, the accuracy of the final inspection may be somewhat compromised.

This suggests that, in order to enhance the performance of the Swin Transformer in wood detection tasks, further optimization and adjustments to the model may be necessary.

1) OPTIMISATION OF THE BACKBONE NETWORK

Swin Transformer V2 [14] is a model that has been partially optimized based on the Swin Transformer model. Some modules have been modified to further enhance the model's performance. However, when applied to the target detection task of wood segmentation, it does not yield satisfactory results in practical scenarios involving small detection objects, large detection targets, and low image resolutions. Despite this, certain modules from the Swin Transformer V2 model still offer valuable insights.

In this paper, we utilize and optimize certain modules from both the Swin Transformer and Swin Transformer V2 models, which are based on the attention mechanism and have demonstrated promising results. These selected modules serve as the framework for our backbone network, which is then integrated into an optimized Mask R-CNN segmentation model. By leveraging the strengths of these modules and their respective models, we aim to enhance the overall performance of the wood segmentation task in the Mask R-CNN model.

During optimization, the research observed that improvements such as cosine attention and LayerNorm posterior were not very effective when applied to the wood segmentation task in practical scenarios involving small detection objects, an increased number of detection targets, and low image resolution. This is because these enhancements tend to expand the model's capacity, which may not be suitable for models with small datasets. However, the logarithmic spatial continuous position bias addresses the challenge of window size inconsistency, thereby enhancing performance during training on practical production wood dataset images of varying sizes. It effectively handles the migration problem arising due to inconsistencies in window sizes. The utilization of logarithmic space is adopted to improve the model's performance in training on wood detection datasets of different sizes. This adjustment helps address the challenges associated with window size inconsistency and is illustrated in Figure 4, outlining the potential impact of these bias improvements on the overall performance of the model in wood segmentation tasks.

2) FEATURE FUSION BASED ON BFP NETWORKS

The Feature Pyramid Network (FPN) [15], employed commonly in detection tasks, achieves information fusion between various feature maps through a top-down pathway, constructing a feature pyramid. However, its predominant utilization of upsampling and downsampling to merge features across different levels proves ineffective in addressing the issue of indistinct features due to widespread occlusion challenges among wood. Moreover, despite the FPN's pyramid structure providing multi-scale feature maps, it still grapples with the problem of scale mismatch, leading to

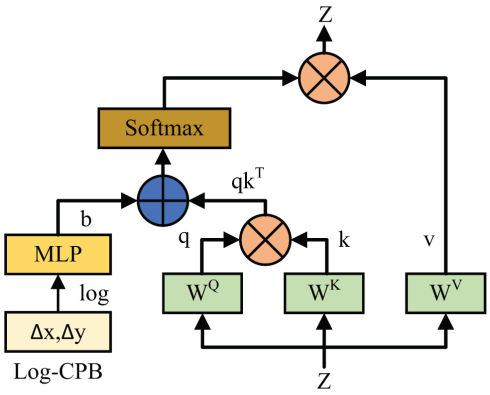


FIGURE 4. Continuous position deviation in logarithmic space.

suboptimal detection results for both small and large wooden targets.

In order to enhance the characteristic information of wood and elevate the performance of target detection, this study adopts the Balanced Feature Pyramid Network (BFP) as a replacement for the original FPN network. The BFP network innovatively addresses the limitations of FPN, offering improved feature clarity amidst occlusion challenges and enhancing detection efficacy for targets of various sizes. The BFP network enhances the expressive capacity of each level feature map by leveraging information from multiple hierarchical feature maps. It achieves a balance in the fused feature information for different scales of wood, ensuring that semantic information from non-adjacent levels is preserved throughout the information propagation process without dilution. Simultaneously, the BFP feature fusion network considers cross-level and cross-scale feature fusion to comprehensively capture the characteristics of wood. This design effectively elevates the performance of wood detection tasks, enabling the network to more accurately identify wooden targets under various scales, shapes, and occlusion scenarios. The structural depiction of the BFP feature fusion network is illustrated in Figure 5.

3) LOSS FUNCTION BASED ON CIOU NETWORK

The Generalized Intersection over Union (GIOU) [16] stands as a prevalent loss function in wood detection models, formulated as follows:

$$L_{GIOU} = 1 - IOU + \frac{C - (A \cup B)}{|C|} \tag{1}$$

$$IOU = \frac{|A \cap B|}{|A \cup B|} \tag{2}$$

However, in practical wood detection tasks, sensitivity to variations in bounding box dimensions gradually becomes apparent, especially when dealing with a multitude of wood elements of disparate scales. This results in training instability and inconsistent performance across targets of varying sizes. Additionally, GIOU lacks the incorporation of learnable parameters, thereby rendering it incapable of enhancing performance on wood detection tasks through the acquisition of adaptive parameters. To address these issues and elevate detection accuracy, this study introduces the Complete Intersection over Union (CIOU) as follows:

$$L_{CIOU} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha\nu \tag{3}$$

$$\alpha = \frac{\nu}{1 - IOU + \nu} \tag{4}$$

$$\nu = \frac{4}{\pi^2}(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h})^2 \tag{5}$$

By introducing penalties for diagonal distance and aspect ratio, CIOU exhibits heightened robustness when handling elliptical-shaped wood target boxes, mitigating oscillations and instability during training. Simultaneously, the introduction of learnable parameters allows the model to adapt more effectively to specific wood detection tasks. When evaluating the matching degree of target boxes, consideration of the shape of the target box facilitates a closer alignment between predicted and actual boxes, enhancing the precision of object detection models.

C. IMAGE PROCESSING

For the image dataset, this study employs Albumentations, a Python-based image augmentation library primarily designed for deep learning and computer vision tasks to enhance the quality of training models. Leveraging the highly optimized OpenCV library, Albumentations rapidly augments image data, exhibiting superior performance compared to most data augmentation libraries, as illustrated in Table 2.

The augmentation techniques involve blur processing, manipulation of image color channels, and the introduction of interference noise to simulate the generation of images under various outdoor weather conditions, contributing to image augmentation and expanding the dataset. This strategy aims to enhance the generalization of deep learning models, consequently improving the segmentation accuracy of wood images in diverse weather environments, as depicted in Figure 6.

Furthermore, this study enhances the original wood dataset's resolution from 1600×1200 to 2000×1200. During data training, a multi-scale training approach is implemented by varying the input size of images to multiple scales. In the training process, each image is randomly assigned a scale for input into the model, enabling the model to learn features of objects at different sizes. This proves particularly effective in extracting features of small targets, significantly enhancing the model's detection capabilities.

III. RESULTS AND ANALYSIS

A. TRAINING ENVIRONMENT AND HYPERPARAMETER SETUP

The algorithm is built upon PyTorch, a prominent open-source framework for deep learning. To expedite the training and inference of the neural network, NVIDIA CUDA GPUs are employed. Table 3 provides details about the specific

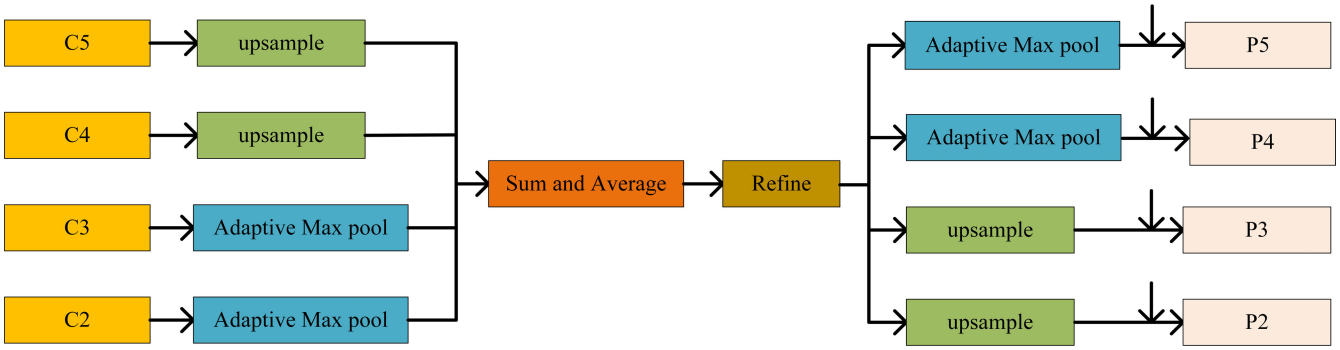


FIGURE 5. Schematic diagram of BFP network.

TABLE 2. Model accuracy changes after using albumentations.

Model	Base augmentations(%)	AutoAugment augmentations(%)
ResNet-50	76.3	77.6
ResNet-200	78.5	80.0
AmoebaNet-B(6,190)	82.2	82.8
AmoebaNet-C(6,228)	83.1	83.5

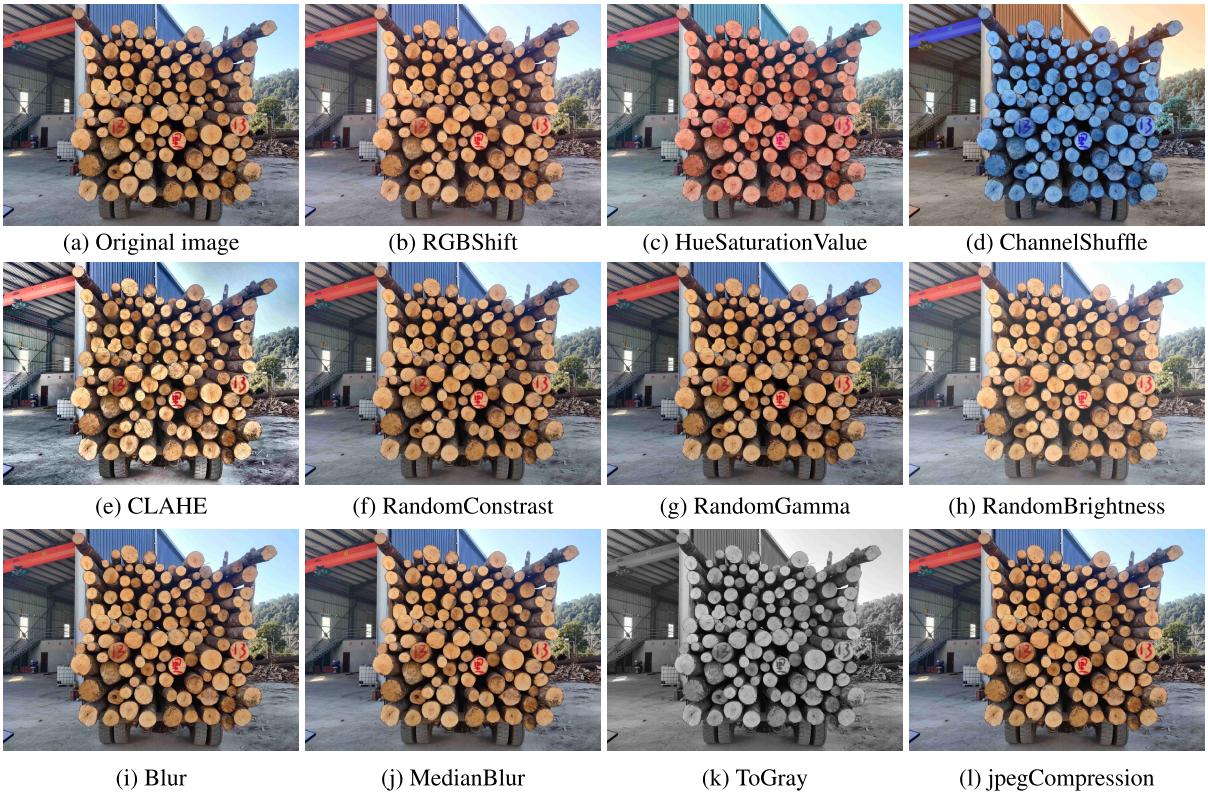


FIGURE 6. Data augmentation with the albumentations library.

experimental environment, including information related to the hardware and software configuration.

Based on the detection requirements of this experiment, the general training configuration parameters are as follows: the number of target categories is 1, with the detection category specified as "wood." To optimize network operation efficiency, the Rectified Linear Unit (ReLU) serves as the activation function for the model. The BatchSize is set to 1,

TABLE 3. Experimental environment.

Projects	Hardware indicators
Operating systems	Windows 11
CPU	i5-12600KF 3.7GHz
GPU	NVIDIA 3060 12GB
Accelerated environment	CUDA 11.1
PyTorch Versions	1.11
Python version	3.8.13

TABLE 4. Key evaluation metrics for the MS COCO dataset.

Evaluation indicators	Meanings
AP	IOU=0.50:0.05:0.95
mAP	Mean average precision
mAP_s	Small targets
mAP_m	Medium-sized targets
mAP_l	Large targets

and there are 2 data loading threads. The number of training Epochs is 36, with an initial learning rate of 0.01. A linearly varying warm-up is applied during the first 500 iterations to stabilize parameter gradients in the initial stages of training, and Stochastic Gradient Descent (SGD) is employed for gradient optimization. Additionally, a descent strategy is used, where the learning rate is multiplied by a factor of 0.1 at Epoch numbers equal to 12, 20, and 28, respectively, further refining the training process and enhancing the model's performance.

B. MODEL EVALUATION

To comprehensively assess the model's performance, this study employs several evaluation metrics from the MS COCO dataset, as presented in Table 4.

The evaluation procedure involves the utilization of ten distinct IOU thresholds, ranging from 0.5 to 0.95 in intervals of 0.05. These thresholds facilitate the calculation of Average Precisions (AP) for small (mAP_s), medium (mAP_m), and large (mAP_l) target sizes, respectively. This approach offers a nuanced evaluation that accounts for targets of varying scales. The cumulative assessment metric, termed mAP (mean Average Precision), is derived by averaging the APs across all IOU thresholds. This comprehensive metric, mAP, encapsulates the statistical mean of the evaluation outcomes, encapsulating the algorithm's performance holistically.

C. MODEL IMPROVEMENT EXPERIMENT AND RESULT ANALYSIS

To validate the enhanced performance of the optimized Swin Transformer algorithm, a series of three meticulously designed experiments were conducted within the same experimental environment and under consistent training data conditions. These experiments encompassed a performance assessment of the predominant detection model, an ablation study, and a performance evaluation of the dominant target detection model in the context of real log detection.

The ablation study, an integral component of the evaluation process, sought to analyze the impact of distinct enhancements within the same network framework on overall network performance. This step allowed for a granular understanding of the contributions of different modifications to the network's efficacy.Subsequently, the performance of the proposed research method was benchmarked against that of mainstream detection networks through two crucial experiments: a performance comparison with mainstream detection models and an evaluation of its true detection performance in log detection tasks. These comparative analyses provided a comprehensive perspective on the strengths and weaknesses of the research approach.

Given that the proposed algorithm's backbone network is an improvement built upon the Swin-T architecture, the choice of comparison models was purposefully aligned with networks of similar dimensions. This strategy ensured a fair and accurate assessment of the algorithm's advancements within a relevant context.

1) PERFORMANCE COMPARISON OF MAINSTREAM DETECTION MODELS

In order to facilitate a comprehensive comparative analysis of various prominent models with respect to log detection performance, this research conducted a systematic comparative experiment. As detailed in Table 5, the enhanced model presented in this study achieved an mAP (mean average precision) of 0.685, representing a notable improvement of 0.024 when contrasted with the performance of the Swin-T algorithm. Additionally, it exhibited an extra enhancement in IOU (Intersection over Union) of 0.015. Furthermore, in comparison to peer models operating within the same framework, such as ResNet-50, the mAP demonstrated a superiority margin of 0.029, while the IOU registered a 0.017 increase. Across diverse frameworks, when measured against models of similar scale, including Cascade RCNN, HTC, and TOOD, the mAP outperformed them by margins of 0.029, 0.027, and 0.017, respectively. Although there was a slight IOU decrease compared to HTC by 0.021, it nevertheless showcased overall improvements in comparison to other models.

To further highlight the disparity between the model's performance before and after enhancement, we direct the output of identical channels from the same layer, as illustrated in Figure 7. The refined model demonstrates heightened focus on intricate details within the images, resulting in a more pronounced delineation of wood contours and an augmented capacity for feature extraction compared to its predecessor. Concurrently, for empirical validation of the model's efficacy, field tests were conducted in a forestry setting, the visual representation of the pre and post-enhancement detection performance is depicted in Figure 8, illustrating a substantial improvement in the model's ability to detect occluded wood and small target timber.

TABLE 5. Performance comparison of state-of-the-art detection models.

Models	Backbone	mAP (%)	mAP_s (%)	mAP_m (%)	mAP_l (%)	IOU (%)
Mask R-CNN	Swin-T	66.1	44.8	70.5	83.3	89.8
Mask R-CNN	ResNet-50	65.6	53.3	81.5	89.3	89.6
Cascade RCNN	ResNet-50	68.5	58.3	81.2	86.7	87.5
HTC	ResNeXt-50	64.7	48.1	71.3	91.6	93.4
TOOD	Swin-T	66.8	54.1	70.4	89.2	63.9
Algorithms in this paper	Improvements to Swin-T	68.5	65.6	84.7	92.8	91.3

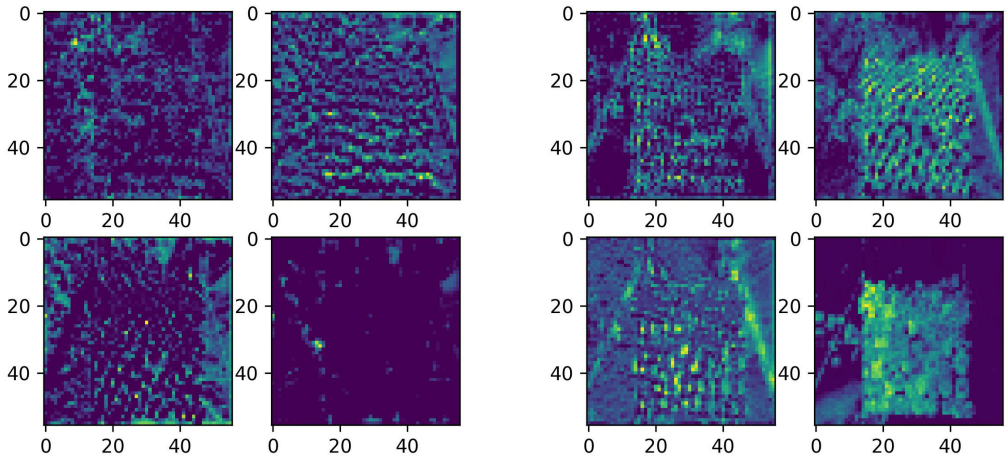
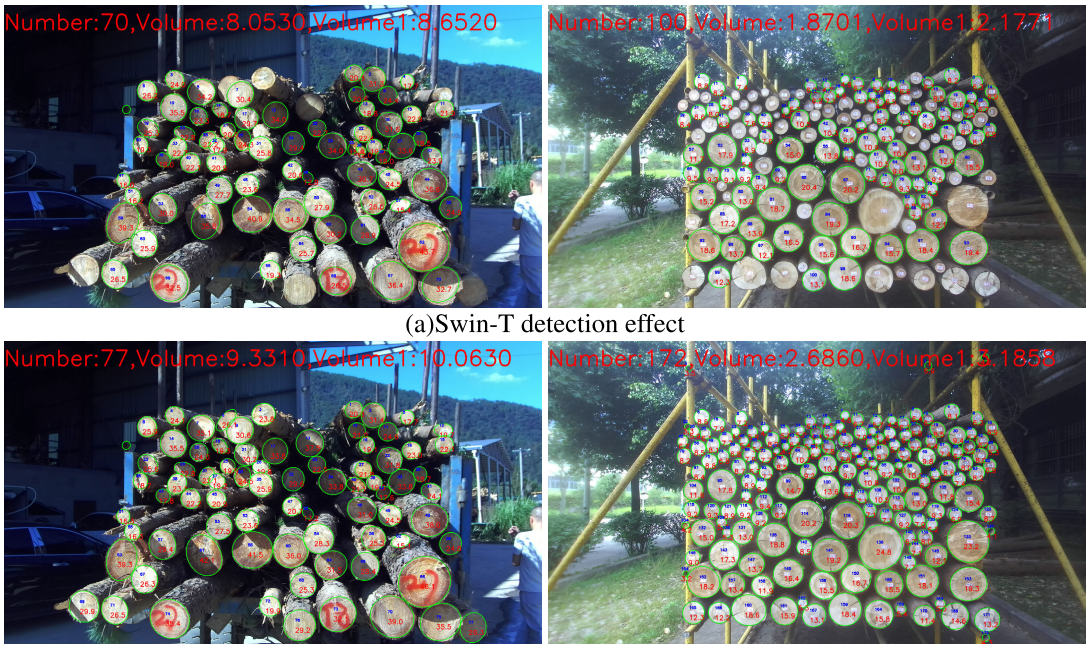


FIGURE 7. Comparison of feature maps before(left) and after(right) improvement.



(a)Swin-T detection effect

(b)Improved Swin-T detection

FIGURE 8. Comparison of effects before and after improvement.

2) ABLATION TESTS

To comprehensively assess the impact of the proposed improvement methods on the performance of the Swin Transformer algorithm, four meticulously designed scenarios were employed, each analyzing distinct aspects of enhancement. All scenarios maintained uniform training parameters,

TABLE 6. Ablation experiment.

Models	Backbone Network	Feature fusion	Loss function	Data enhancement	mAP(%)	IOU(%)
Swin Transformer	×	×	×	×	66.1	88.6
Scheme 1	✓	×	×	×	67.6	89.6
Scheme 2	✓	✓	×	×	67.8	89.7
Scheme 3	✓	✓	✓	×	68.2	92.7
Scheme 4	✓	✓	✓	✓	68.5	91.3

TABLE 7. Comparison of model performance in detecting logs in real-world scenarios.

Models	Backbone	Actual number of logs	Number of model tests	Number of model misdetections	False detection rate(%)	Number of model true checks	Number of model misses	Model log authenticity rate(%)
Mask R-CNN	Swin-T	1783	1685	22	1.321	1653	130	92.753
Mask R-CNN	ResNet-50	1783	1653	22	1.36	1592	191	89.325
Cascade RCNN	ResNet-50	1783	1657	21	1.28	1606	177	90.108
HTC	ResNeXt-50	1783	1703	13	0.82	1691	92	94.884
TOOD	Swin-T	1783	1644	25	1.52	1609	174	90.282
Algorithms in this paper	Improvements to Swin-T	1783	1739	8	0.46	1716	67	96.255

ensuring consistent evaluation. The outcomes, elucidating the effects of these methods on the model's detection performance, are consolidated in Table 6, where the symbols "✓" and "×" denote the integration and omission of respective improvement strategies within the network model.

Scheme 1, which employs continuous relative position bias in logarithmic space within the backbone network instead of the previous parametric learnable relative position bias method, effectively addresses the suboptimal performance of timber dataset images during actual production training. This improvement is particularly evident in the training phase, resulting in a substantial mAP enhancement of 0.015 compared to the baseline. Scheme 2, tailored to datasets featuring numerous small and medium-sized targets, introduces a novel feature fusion structure. By fusing the bottom layer feature map with the top layer feature map, the model's feature extraction capability is enhanced, leading to modest increments in both mAP and IoU. In Scheme 3, the integration of the CIOU loss function aims to enhance the model's accuracy in locating target frames. This results in improved sample robustness and learnability, fostering better alignment with target frames and thereby significantly increasing both mAP and IoU by 0.004 and 0.03, respectively, over Scheme 2.Finally, Scheme 4 introduces the Albumentations data enhancement module, effectively enhancing training speed and contributing to an additional mAP boost of 0.003. However, this improvement is accompanied by a slight IoU reduction of 0.014.Comparing the mAP values before and after the proposed enhancements, the model in this study achieves an mAP of 0.685, surpassing the pre-improvement mAP of Swin Transformer. Moreover, this mAP improvement of 0.024 demonstrates the efficacy of the proposed enhancements. Additionally, the IoU reaches 0.913, marking a substantial enhancement of 0.027 over the pre-improvement value.

3) COMPARISON OF MODEL PERFORMANCE IN DETECTING LOGS IN REAL-WORLD SCENARIOS

To substantiate the effectiveness of the improved model in real-world scenarios, professional personnel conducted manual measurements in an actual forest setting, resulting in an actual count of 1783 logs. Subsequently, various state-of-the-art models were employed to estimate the number of logs. Table 7 reveals that, when compared to the pre-improved Swin Transformer model, the true detection rate for logs has increased from 92.753% to 96.255%, marking a significant gain of 3.502%. The false detection rate has reduced from 1.32% to 0.46%, a decrease of 0.86%. In comparison to other leading detection models, this model demonstrates a superior performance in both true detection rate and false detection rate for logs.

Meanwhile, considering the comprehensive data presented in Tables 5 and Tables 6, it can be concluded that the model achieves the most favorable evaluation metrics while also attaining the highest true detection rate in practical log detection. This serves as compelling evidence that the model has indeed made substantial improvements in log detection performance.

IV. CONCLUSION

The efficiency and effectiveness of log handling play a pivotal role in automating the timber industry, underscoring the significance of accurate and streamlined log segmentation. This paper tackles the challenge of detecting and segmenting complete logs by harnessing the power of the Swin Transformer algorithm. The study delves into four primary dimensions, each systematically revamped to elevate network performance and bolster generalization.

Primarily, the integration of the log-space continuous positional bias method addresses the migration issue arising from inconsistent window sizes. This innovation offers

a pragmatic solution to real-world production challenges. Subsequently, the BFP feature fusion module is harnessed to harmonize data from varying resolutions, thereby amplifying the network's overall efficacy. Furthermore, the incorporation of the CIOU loss function enriches the model's precision in pinpointing target frames, fostering a higher level of detection accuracy. Concluding the suite of improvements, the application of the Albumentations data enhancement technique adapts to the distinctive attributes and complexities of log datasets. Through operations like rotation, cropping, scaling, flipping, and noise injection applied to the original images, dataset size and diversity are expanded, effectively enhancing the network's resilience and adaptability.

To validate the prowess and supremacy of the proposed algorithm, an extensive comparative analysis is conducted, encompassing classical target detection, segmentation, and log segmentation algorithms. This thorough assessment employs the same experimental platform environment and evaluates against metrics of detection accuracy and effectiveness. The outcomes convincingly portray the superiority of the algorithm presented in this paper across all metrics, establishing its prowess in efficiently detecting and segmenting complete logs. Moreover, ablation comparison experiments are undertaken to dissect the impact and contribution of the introduced enhancements. The results conclusively demonstrate the enhancements' positive influence on detection performance, affirming their indispensability and efficacy. In summary, the algorithms formulated within this study yield substantial achievements in the realm of detecting and segmenting entire logs. This research presents a promising and innovative avenue for streamlining log processing efficiency while preserving accuracy.

REFERENCES

[1] L. Tang, G. Shao, and L. Dai, "Roles of digital technology in China's sustainable forestry development," *Int. J. Sustain. Develop. World Ecol.*, vol. 16, no. 2, pp. 94–101, May 2009, doi: 10.1080/13504500902794000.

[2] S. Yella and M. Dougherty, "Automatically detecting the number of logs on a timber truck," *J. Intell. Syst.*, vol. 22, no. 4, pp. 417–435, Dec. 2013, doi: 10.1515/jisys-2013-0026.

[3] F. Budiman, R. Mardiyanto, and R. Rachmat, "A handy and accurate device to measure smallest diameter of log to reduce measurement errors," in *Proc. Int. Seminar Intell. Technol. Appl.*, 2016, pp. 423–428.

[4] H. Tang, K. Wang, J. Gu, X. Li, and W. Jian, "Application of SSD framework model in detection of logs end," *J. Phys., Conf. Ser.*, vol. 1486, no. 7, Apr. 2020, Art. no. 072051, doi: 10.1088/1742-6596/1486/7/072051.

[5] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Comput. Vis. (ECCV)*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham, Switzerland: Springer, 2016, pp. 21–37.

[6] N. Samdangdech and S. Phiphobmongkol, "Log-end cut-area detection in images taken from rear end of eucalyptus timber trucks," in *Proc. 15th Int. Joint Conf. Comput. Sci. Softw. Eng. (JCSSE)*, Jul. 2018, pp. 1–6.

[7] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[8] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 10012–10022.

[9] J. Pang, K. Chen, J. Shi, H. Feng, W. Ouyang, and D. Lin, "Libra R-CNN: Towards balanced learning for object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 821–830.

[10] Z. Zheng, P. Wang, D. Ren, W. Liu, R. Ye, Q. Hu, and W. Zuo, "Enhancing geometric factors in model learning and inference for object detection and instance segmentation," *IEEE Trans. Cybern.*, vol. 52, no. 8, pp. 8574–8586, Aug. 2022.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30. Red Hook, NY, USA: Curran Associates, 2017, pp. 1–11.

[13] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[14] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong, F. Wei, and B. Guo, "Swin transformer V2: Scaling up capacity and resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 12009–12019.

[15] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 936–944.

[16] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 658–666.

ZHIGANG DING received the master's degree from Jilin University, China, in 2007. He is with the Fujian University of Technology. He is also a Master's Supervisor and currently the Head of the Vehicle Engineering Laboratory. He has participated in National 863 Program and more than ten provincial and municipal scientific research projects. His main research interest includes application of artificial intelligence in the vehicle industry.

FUCHENG FU was born in Xuchang, Henan, China, in 1997. He is currently pursuing the M.A.Eng. degree with the School of Transportation, Fujian University of Technology, China. His research interests include image processing and machine learning.

JISHI ZHENG received the Ph.D. degree from Central South University, China, in 2015. From January to July 2019, he was a Visiting Scholar with the Robotics Laboratory, Computer and Electronic Engineering Department, University of Essex, U.K. He is with the Fujian University of Technology, where he is currently the Head of the Internet of Things major, the Director of teaching and research with the Department of Traffic Information and Control, and the Executive Director of the Fujian Society of Aeronautics. He has presided over and participated in more than ten provincial and municipal scientific research projects. His main research interests include application of artificial intelligence in the industry and research of UAV flight control algorithm.

HAIYAN YANG received the Ph.D. degree in computer applied technology from Central South University, China, in 2015. She is currently with the Fujian University of Technology. She possesses extensive teaching experience, responsible for instructing several core computer courses. Her primary research interests encompass image processing and pattern recognition.

FUMIN ZOU received the Ph.D. degree from Central South University, in 2009. He currently holds the positions of a Professor and the Dean with the School of Electronics, Electrical, and Physical Science, Fujian University of Technology. Additionally, he is also a part-time Professor and the Doctoral Supervisor with the College of Computer and Big Data, Fuzhou University. He has led and participated in over 50 provincial and ministerial-level research projects, authored more than 100 academic articles, and applied for more than 100 national invention patents.

KONG LINGHUA received the Ph.D. degree in mechanical engineering from McGill University, Canada, in 2004. He has been with the Fujian University of Technology, since 2015. He has designed and developed a variety of new products and equipment and received 15 patents. As the lead author, he has published 20 influential articles included in SCIEI.