# Toward an Automated Pruning for Apple Trees Based on Computer Vision Techniques

Keming Hu and Oky Dicky Ardiansyah Prima

*Graduate School of Software and Information Science, Iwate Prefectural University*
152-52 Takizawa, Iwate, 020-0693, Japan
e-mail: g231t015@s.iwate-pu.ac.jp, prima@iwate-pu.ac.jp

***Abstract*—Effective pruning can contribute to the growth of plants. Similarly, pruning apple trees can help them absorb nutrients and grow stronger. However, Japan's apple farming industry is facing many challenges today, such as aging population, talent shortage, and reduced farmland area. Despite many studies attempting to solve the problems of aging population and talent shortage through automated pruning, but as a preliminary step for pruning, the process of identifying apple trees is too complex and difficult to achieve in real-world environments. In this paper, we propose a more straightforward apple trees recognition method based on computer vision to achieve pruning of apple trees in real environments. The method roughly consists of three steps: 1) segmenting apple trees through semantic segmentation, 2) skeletonizing the apple tree by segmentation image, 3) the representation of graph tree is done by applying breadth-first search. We tested 12 models for apple tree segmentation, and the Segfomer model achieved an accuracy of 76.72 and an intersection over union(IoU) of 64.29.**

***Keywords-Trees-Pruning; AI, semantic segmentation, thinning tree.***

## I. INTRODUCTION

One of the world's most famous apple-producing regions is in Japan. Due to the high quality and price of Japanese apples, Japan enjoys a high reputation and market share in the international market. However, in recent years, Japan's labor shortage, aging population, and young people's unwillingness to engage in agricultural labor have become one of the bottlenecks restricting the development of Japanese apple orchards. To address these issues, many traditional manual labor practices have been replaced by automation and mechanization. For example, Global Positioning System (GPS) technology is used during the planting stage to achieve automatic navigation, spraying, and fertilization of apple orchards. In the picking stage, apple orchards are also beginning to use automated picking robots. During the pruning stage of apple tree growth, traditional manual pruning is still mostly used. However, manual pruning has some negative effects, including: 1) low efficiency and higher time and labor costs, 2) requiring the operator to have certain skills and experience, otherwise it may affect the growth and yield of apple trees, 3) due to labor shortages, many apple orchards are unable to complete pruning work in a timely and effective manner, 4) manual pruning requires the operator to work on the tree for a long time, which can easily lead to physical fatigue and injury. To address these issues, it is necessary to use robots to replace manual operations.

Pruning refers to the process of removing unwanted branches from apple trees in order to promote better growth. Apple trees grow rapidly, so they require frequent pruning to control their growth and shape in order to obtain better sunlight exposure and nutrition. In addition, apple trees generally bear a lot of apples. If not pruned, too many apples will concentrate on the same branch, leading to excessive apple quantity and density, which can affect the growth and maturity of the apple and make it difficult to harvest.

Generally, the best time for pruning is after the dormant period of the apple tree ends and before new buds' sprout. This can avoid affecting the growth of the fruit tree. The specific timing varies depending on the region and climate conditions, usually between March and April in the spring. When pruning, certain rules should be followed: 1) preserve the main trunk and major lateral branches, and appropriately trim other branches to maintain the tree crown's ventilation, light penetration, and transparency. 2) in the case of a dense tree crown, loosen it appropriately to ensure that each fruit has sufficient sunlight and air. 3) branches should be smaller than the trunk. 4) the lowest branch should be 2 to 3 ft above the ground.

Traditional apple tree pruning is difficult, requires high precision, and has low efficiency. Therefore, the emergence of automated apple tree pruning technology can effectively solve these problems. Automated pruning robots can replace manual pruning of apple trees, which can improve pruning efficiency, ensure pruning precision and quality, and reduce the labor intensity and risks for operators.

In section II of this article, we will introduce some previous research and current problems encountered in apple tree pruning. In section III, we will introduce our proposed method. In section IV, we will evaluate and discuss our results. Finally, we will summarize our paper in section V.

## II. RELATED WORKS

Karkee et al. [1] attempted to extract branches of apple trees using a depth camera and estimate the length of branches and distance to the next branch to identify the branches that should be pruned. Majeed et al. [2] identified fruit trees and their backgrounds using depth information and extracted branches

Copyright (c) IARIA, 2023. ISBN: 978-1-68558-078-0 

(a) Original image (b) Anntotation image

Fig. 1. An example of apple trees for this study.

from image information using semantic segmentation. Zhang et al. [3] demonstrated that accurate branch extraction can be achieved by combining raw depth information with a pseudo-color image obtained by coloring the depth information. While these approaches have shown that the extraction of fruit tree branches using camera sensing is possible, the topology of the extracted branches has not yet been analyzed.

Fumey Damien et al. [4] believed that if the topological structure of fruit tree branches and trunks is known, the relationships between various parts of the tree can be determined, such as the branching of the trunk, the positions of leaves, etc. This allows for pruning while ensuring the growth and yield of the fruit tree. With the topological structure of fruit tree branches and trunks, pruning rules can be more systematic and precise, avoiding inconsistencies in pruning due to subjective factors and differences in experience, thereby improving the efficiency and quality of fruit tree pruning.

```
Input: Masked image (M)
Output: The topological structure of an apple tree
Function findConnection(M):
    process queue ← [root point];
    node link ← [root point];
    while process queue lenght > 0 do
        now point ← process queue.pop(0);
        for each in now point arround(M) do
            node link[now].nextAdd(each);
            process queue.add(each);
        end
    end
    return node link;
```

Fig. 2. Breadth-first search of an apple tree.

## III. Materials and Methods

The primary step for successful pruning of apple trees is to obtain the topological structure of the tree, which involves analyzing the intersections of the branches and using these points to determine the topology of the tree. This directed acyclic graph provides information on the distance between nodes, branch lengths, and directions. Obtaining the topological structure of the tree through this method allows for a comprehensive understanding of both the individual branch structure and the overall direction of the tree, thereby ensuring adequate ventilation, light transmission, and pruning effectiveness.

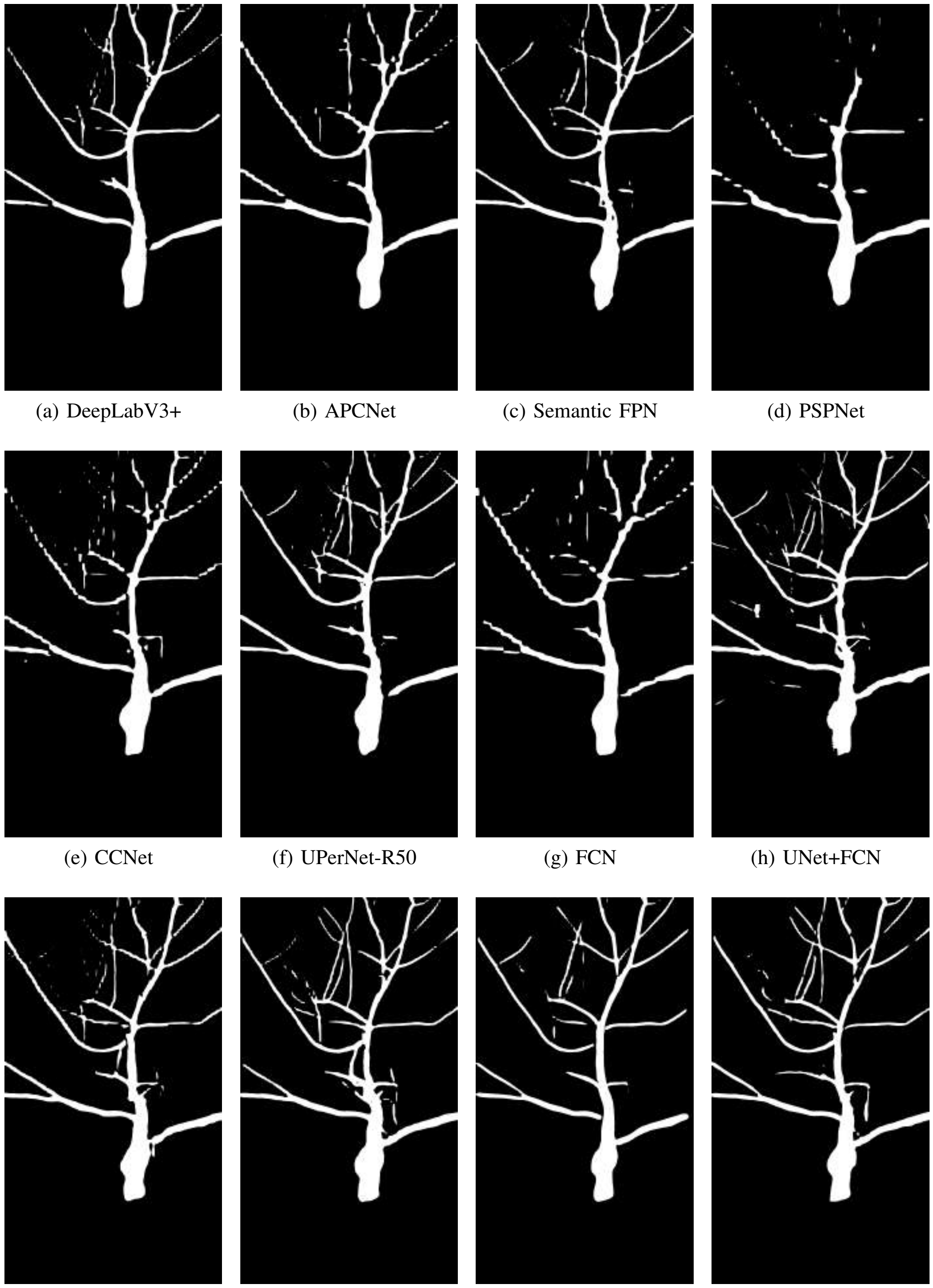
To be able to obtain the topology of the apple tree, here is the method we propose: first, the apple tree is segmented semantically using RGB cameras, and the mask of the tree is obtained. Then, dilation is applied to the mask to connect discontinuous branches. Next, the skeleton of the apple tree is extracted from the dilated mask, and finally, the topological structure of the tree is obtained through a breadth-first search from the root.

### A. Semantic Segmentation of Apple Trees

Supervised learning semantic segmentation requires pixel-level annotation of the original images. Due to the complex shape of fruit trees, it is very difficult to label the entire tree, so we chose to annotate by tree branches and trunks, and then integrate them together. We used the labelme [5] annotation tool to annotate the images, as shown in Figure 1, which shows the original dataset images and the annotated images. We use the open-source OpenMMLab [6] framework and initialized the network with the provided pre-trained weights for training. Due to the small size of the dataset, we applied image augmentation techniques, such as random cropping and flipping during training, and used the cosine annealing learning rate update strategy to further stabilize the network training. The network strategy and detailed parameters will be discussed in section IV of this paper.

### B. Analysis of Apple Tree Topology

To obtain the topological structure of fruit trees, we aim to refine the mask representing branches that the neural network obtained from semantic segmentation. First, we perform a graphic dilation on the segmented mask to connect any disconnected or disjointed branches, enabling further processing. Next, we perform skeletonization on the dilated image. In computer graphics, skeletonization is the process of converting the edges or curves of a Two-Dimensional (2D), or Three-Dimensional (3D) image or object into its skeleton or centerline. By skeletonizing the fruit tree, this study can extract the object's topological structure, geometric shape, size, and orientation. We refer to the masked image of the skeletonized tree as an array M. The bottom-most point of the array is considered as the root point, and we perform a breadth-first search on all pixels along the paths using the algorithm in Figure 2, while recording the parent-child relationships between nodes, until all nodes have been explored. These



(a) DeepLabV3+ (b) APCNet (c) Semantic FPN (d) PSPNet

(e) CCNet (f) UPerNet-R50 (g) FCN (h) UNet+FCN

(i) CGNet (j) UPerNet-swin (k) UPerNet ViT-b16-mln (l) SegFormer

Fig. 3. An example of segmentation results of an apple trees used in this study.

parent-child relationships between nodes can approximate the topological structure of the apple tree, and pruning analysis can be performed based on this structure.

## IV. Results and Discussion

Our data was obtained from an apple orchard in Hanamaki City, Iwate Prefecture, Japan. The trees in this orchard do not have a specific spindle-shaped structure and are planted separately, which is advantageous for our research. Figure 1 shows one of the images from our test set, which is part of a dataset that includes 160 images of trees taken from different directions. We used 140 images from five trees as the training set and 20 images from four trees as the test set.

We used OpenMMLab to build 12 semantic segmentation models to attempt to segment branches from apple tree, and completed the training on NVIDIA RTX2080Ti (11G). We evaluated our models using accuracy, IoU, and parameters of those models as shown in Table 1. Traditional models that used Convolution Neural Network (CNN) Resnet [7] as the backbone generally had an IoU of around 55 on the test set, while CNN with Feature Pyramid Networks (FPN) [8] layers could reach 62. Networks that used attention mechanism

TABLE I
THE RESULT OF SEGMENTATION

| Name | Backbone | Input Size | Batch Size | Iterations | Acc | IoU |
|---|---|---|---|---|---|---|
| DeepLabV3+ [9] | R50 | 480x480 | 6 | 10000 | 65.26 | 54.48 |
| APCNet [10] | R50 | 512x512 | 6 | 10000 | 67.09 | 58.39 |
| Semantic+FPN | R50 | 512x512 | 16 | 5000 | 72.1 | 62.35 |
| PSPNet [11] | R50 | 480x480 | 8 | 3500 | 53.29 | 46.44 |
| CCNet [12] | R50-D8 | 512x512 | 6 | 10000 | 68.98 | 59.19 |
| UPerNet [13] | R50 | 512x512 | 8 | 5000 | 69.99 | 60.44 |
| FCN [14] | R101 | 512x512 | 3 | 10000 | 64.48 | 55.27 |
| UNet [15]+FCN | UNet-S5-D16 | 64x64 | 128 | 1000 | 73.5 | 62.29 |
| CGNet [16] | M3N21 | 680x680 | 12 | 10000 | 74.87 | 65.13 |
| UPerNet | Swin-S [17] | 512x512 | 8 | 30000 | 76.98 | 64.24 |
| UPerNet | ViT-B [18]+LN+MLN | 512x512 | 4 | 60000 | 56.89 | 43.27 |
| SegFormer [19] | MIT-B5 | 512x512 | 4 | 60000 | 76.72 | 64.29 |

(a) Original (b) Mask (c) Skeleton

Fig. 4. Topological structure result of tree #1.

(a) Original (b) Mask (c) Skeleton

Fig. 5. Topological structure result of tree #2.

(a) Original (b) Mask (c) Skeleton

Fig. 6. Topological structure result of tree #3.

(a) Original (b) Mask (c) Skeleton

Fig. 7. Topological structure result of tree #4.

modules as the backbone achieved higher IoU on the test set and also achieved higher accuracy, but due to the more complex network model with more parameters, the batch size was limited, and more iterations and training time were required.

We tested semantic segmentation on 12 different networks using the image from Figure 1, and the results are shown in Figure 3. The model with ResNet as the backbone for feature extraction generates discontinuous masks for fine branches, while the model incorporating self-attention mechanism can effectively improve this problem. This phenomenon may be due to the fact that CNNs can only sense local structures by scanning images with convolutional kernels, leading to a poor grasp of the overall location and direction of apple trees and branches. By leveraging the attention mechanism in the encoder part of the Transformer, global features can be integrated and facilitate the segmentation of slender branches. Based on the model evaluation results and image testing results, we chose Segformer, which yields higher IoU, relatively coherent tree branches, and more accurate backgroundforeground segmentation, to explore the topological structure of apple trees.

We used the mask obtained from segmentation by Segformer to perform skeletonization, and the results of breadth-first search from the root of the apple tree on the skeletonized image are shown in Figures 4 to 7. Different colors were used to clearly indicate the different branches connected to each other. Due to the possibility of circular structures caused by too large mask areas in the skeletonized image, we used breadth-first search to break these circular structures and obtained a directed acyclic graph starting from the root node. This directed acyclic graph can represent the topological structure of the apple tree.

Currently, we are developing the automated pruning based on the topological structure of the apple tree by comparing images of the same apple tree before and after pruning. We will present the results in our next paper.

## V. Conclusion and Future work

In this study, we attempted to explore the topological structure of apple trees by segmenting their branches, skeletonizing them, and conducting exploration. Despite the limited number of training images available for the neural network, we were still able to extract the branches of apple trees from the images. Segformer was found to be the most effective neural network model for segmentation. Future tasks will include the development of apple tree pruning rules in conjunction with the topological structure of the tree. In addition, we will collect more image datasets and modify the hyper-parameters of our neural network to improve the segmentation results to provide a better topological structure of the tree.

## References

[1] M. Karkee et al., "Identification of pruning branches in tall spindle apple trees for automated pruning," Computers and Electronics in Agriculture, 103, pp. 127–135, 2014.

[2] Y. Majeed et al., "Apple Tree Trunk and Branch Segmentation for Automatic Trellis Training Using Convolutional Neural Network Based Semantic Segmentation," IFAC PapersOnLine, 51-17, pp. 75–80, 2018.

[3] J. Zhang et al., "Branch detection for apple trees trained in fruiting wall architecture using depth features and Regions-Convolutional Neural Network (R-CNN)," Computers and Electronics in Agriculture, 155, pp. 386–393, 2018.

[4] D. Fumey, P. Lauri, Y. Guédon, C. Codin, and E. Costes, "Effects of Pruning on the Apple Tree: from Tree Architecture to Modeling," Proc. IXth IS on Orchard Systems Ed.: T.L. Robinson Acta Hort. 903, ISHS, 2011.

[5] Labelme, https://github.com/wkentaro/labelme [retrieved at April 4, 2023]

[6] OpenMMLab, https://github.com/open-mmlab [retrieved at April 4, 2023]

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, 2016.

[8] A. Kirillov, R. Girshick, K. He, and P. Dollár, "Panoptic Feature Pyramid Networks," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, pp. 6392-6401, 2019.

[9] L. Chen et al., "Encoder-decoder with atrous separable convolution for semantic image segmentation," Proceedings of the European conference on computer vision (ECCV), 2018.

[10] J. He, Z. Deng, L. Zhou, Y. Wang, and Y. Qiao, "Adaptive Pyramid Context Network for Semantic Segmentation," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, pp. 7511-7520, 2019.

[11] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid Scene Parsing Network," IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, pp. 6230-6239, 2017.

[12] Z. Huang et al., "CCnet: Criss-cross attention for semantic segmentation," Proceedings of the IEEE/CVF international conference on computer vision, 2019.

[13] T. Xiao, Y. Liu, B. Zhou, Y. Jiang, and J. Sun, "Unified Perceptual Parsing for Scene Understanding," European Conference on Computer Vision, pp. 418-434, 2018.

[14] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, pp. 3431-3440, 2015.

[15] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convo- lutional networks for biomedical image segmentation," Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015, 2015.

[16] T. Wu, S. Tang, R. Zhang, J. Chao, and Y. Zhang, "CGNet: A Light-Weight Context Guided Network for Semantic Segmentation," in IEEE Transactions on Image Processing, 30, pp. 1169-1179, 2020.

[17] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," Proceedings of the IEEE/CVF international conference on computer vision, 2021.

[18] D. Alexey et al., "An image is worth 16x16 words: Transformers for image recognition at scale," ILCR Conference, pp. 1-21, 2021.

[19] E. Xie et al., "SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers," Neural Information Processing Systems, 2021.

Copyright (c) IARIA, 2023. ISBN: 978-1-68558-078-0 