iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** IV **Month of publication:** April 2024

DOI: https://doi.org/10.22214/ijraset.2024.59669

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

AgroSegNet: Semantic Segmentation Guided Crop Image Extraction using Enhanced Mask RCNN

Priyanka S Chauhan[1], Pranjali Narote[2], Durga Sritha Dongla[3], Vimal Aravintha S J[4], Prathamesh Sonawane[5]
[1]*SDET, Engineering Manager, Global Payment, Atlanta, Georgia, USA*
[2]*Undergraduate Student (CSE), Veermata Jijabai Technological Institute, Mumbai, India*
[3]*Undergraduate Student (CSE), Gandhi Institute of Technology and Management, Visakhapatnam, India*
[4]*Undergraduate Student (CSE), SRM Institute of Science and Technology, Chennai, India*
[5]*Undergraduate Student (IT), Vasantdada Patil Pratishthan's College of Engineering and Visual Arts, Mumbai, India*

Abstract: ***Traditional crop image segmentation methods often struggle to accurately extract crops due to the complex interplay of agricultural factors and environmental conditions. In this study, AgroSegNet, an innovative approach that integrates agronomic knowledge into the segmentation process to improve crop extraction accuracy, is proposed. Leveraging the Mask RCNN framework, the method dynamically adapts region proposals based on crop growth stages, phenological information, and agronomic principles. Initializing the Fruits 360 Dataset labels using LabelMe and preprocessing the dataset, we then divide it into training and test sets. Constructing an enhanced Mask RCNN model using the PyTorch 2.0 deep learning framework, our model incorporates path aggregation, feature augmentation, and an optimized region extraction network, bolstered by a feature pyramid network. Spatial information preservation is achieved using bilinear interpolation in ROIAlign. A multi-scale feature fusion mechanism is introduced to capture fine-grained details essential for precise segmentation. Additionally, domain adaptation techniques are employed to enhance model generalization across diverse agricultural environments. Through extensive experimentation and comparative analysis, AgroSegNet demonstrates superior performance in terms of precision, recall, average precision, mean average precision, and F1 scores for crop image extraction compared to traditional methods. The approach not only advances the state-of-the-art in crop segmentation but also offers valuable insights for leveraging agronomic knowledge in computer vision applications for agriculture.***

Keywords: Crop image segmentation, Agronomic knowledge, Mask RCNN, Semantic segmentation, Multi-scale feature fusion

I. INTRODUCTION

Deep learning has revolutionized crop image segmentation by enabling models to learn hierarchical representations directly from raw pixel data. Convolutional Neural Networks (CNNs) have been widely adopted for this task due to their ability to capture spatial dependencies and semantic information. Techniques such as Fully Convolutional Networks (FCNs) have shown promising results in segmenting crops from aerial images [1]. Crop image segmentation plays a pivotal role in modern agricultural practices, facilitating precision farming, yield estimation, and crop health monitoring. However, accurate and efficient extraction of crops from agricultural images remains a challenging task due to the inherent complexities arising from the diverse range of crops and their interactions with environmental elements. Traditional segmentation methods often struggle to delineate crops accurately amidst cluttered backgrounds and varying illumination conditions, necessitating the development of advanced techniques capable of robustly handling such challenges. In recent years, deep learning-based approaches have shown promising results in addressing the complexities of crop image segmentation. Among these, the Mask RCNN framework has emerged as a powerful tool for instance segmentation tasks, offering the capability to simultaneously detect and segment multiple objects within an image. By leveraging the rich representation learned through convolutional neural networks (CNNs) and the precise localization provided by region-based methods, Mask RCNN has demonstrated remarkable performance in various computer vision tasks.

Image segmentation involves partitioning an image into distinct regions based on variations in pixel intensity, enabling researchers to isolate regions of interest [2]. This process, integral to image processing, significantly impacts the outcomes of subsequent image analyses. Leveraging image segmentation technology facilitates the efficient and non-destructive extraction of crop information. This capability enables crop growers to gain insights into real-time crop growth dynamics and enhances their ability to effectively manage crops. In this context, this research proposes an innovative approach for automatic crop image extraction by extending the capabilities of the Mask RCNN framework. Building upon the foundation laid by traditional Mask RCNN, enhancements tailored specifically for crop segmentation tasks are introduced.

©IJRASET: All Rights are Reserved

This approach integrates agronomic knowledge into the segmentation process, leveraging insights from crop growth stages, phenological information, and agronomic principles to improve segmentation accuracy. Furthermore, the Mask RCNN architecture is augmented with additional components aimed at enhancing its performance in agricultural image analysis. This includes path aggregation, feature augmentation, and an optimized region extraction network, along with spatial information preservation using advanced interpolation techniques. Additionally, novel techniques to enhance edge accuracy are introduced, incorporating micro-fully connected layers and edge loss into the segmentation pipeline.

The conventional Mask RCNN architecture is typically constructed using the Keras deep learning framework with TensorFlow as its backend. However, due to limitations within this framework, the network's performance may not be fully realized. To address this, a transition to utilizing the Mask RCNN implemented in the PyTorch framework was made. Subsequently, the enhanced Mask RCNN network was optimized and evaluated within this new framework. The adaptation to PyTorch has resulted in notable performance enhancements. Specifically, there has been a more efficient utilization of computer video memory resources, leading to significant improvements in computational speed and accuracy. Additionally, the PyTorch framework offers enhanced debugging capabilities, modularity, and flexibility in model construction, facilitating seamless data parameter migration between the CPU and GPU.

Motivated by these advancements, this paper proposes a novel crop image extraction algorithm based on Mask RCNN, implemented using the PyTorch deep learning framework. The approach entails several enhancements to the Mask RCNN network model structure, including the incorporation of path aggregation and feature enhancement functionalities into the network design. Furthermore, the region extraction network (RPN) was refined, and the feature pyramid network (FPN) was optimized to improve overall performance. Additionally, semantic segmentation guided by agronomic knowledge is introduced, where insights from crop growth stages and phenological information are integrated into the segmentation process to enhance accuracy and contextual understanding. To evaluate the efficacy of the proposed approach, extensive experiments are conducted using the Fruits 360 Dataset, and the results are compared against traditional Mask RCNN and other baseline methods such as Fully Convolutional Networks (FCN). The experimental analysis encompasses metrics such as precision, recall, average precision, mean average precision, and F1 scores, providing a comprehensive assessment of the proposed approach's performance. Experimental evaluations conducted on the Fruits 360 Dataset demonstrate the superior performance of the proposed enhanced Mask RCNN algorithm in crop image extraction tasks. The results underscore the efficacy of this approach in achieving more accurate and efficient crop segmentation compared to conventional methods. Through this research, significant advancements in crop image segmentation are anticipated, along with valuable insights for the broader field of agricultural image analysis. By harnessing the power of deep learning and integrating domain-specific knowledge, advancements in crop monitoring, yield prediction, and agricultural decision-making processes are envisioned.

II. RELATED WORK

Crop image segmentation is a fundamental task in precision agriculture, enabling automated monitoring of crop health, yield estimation, and decision-making processes. Over the years, researchers have explored various techniques and methodologies to improve the accuracy and efficiency of crop segmentation algorithms. In this literature review, we highlight key advancements in the field and discuss their relevance to the proposed research. Mask RCNN, proposed by [7], combines object detection and instance segmentation in a single framework. By extending Faster RCNN with a segmentation branch, Mask RCNN achieves state-of-the-art performance in various instance segmentation tasks, including crop segmentation. Researchers have utilized Mask RCNN for accurate crop delineation in satellite imagery [3]. Incorporating domain-specific knowledge from agronomy has been identified as a crucial factor in improving crop segmentation accuracy. Agronomic principles, such as crop growth stages and phenological information, can provide valuable contextual cues for segmentation algorithms. Prior research has demonstrated the effectiveness of integrating agronomic knowledge into segmentation pipelines for crop type classification and yield prediction [4]. Multi-scale feature fusion techniques have been proposed to capture both fine-grained details and global context in crop images. By aggregating features from multiple scales, models can effectively distinguish between different crop types and variations in crop appearance. Methods such as Feature Pyramid Networks (FPNs) have been employed to incorporate multi-scale features into crop segmentation networks [5].

Transfer learning techniques have been utilized to adapt models trained on one dataset to perform well on another dataset with different environmental conditions. Domain adaptation methods aim to reduce the domain gap between the source and target datasets, thereby improving model generalization. Researchers have explored domain adaptation strategies for crop segmentation tasks, particularly in scenarios where labeled data is scarce [6].

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

In summary, the literature presents a rich landscape of methodologies and techniques for crop image segmentation, ranging from deep learning-based approaches to integrating domain-specific knowledge. The proposed research aims to build upon these advancements by leveraging the Mask RCNN framework, integrating agronomic knowledge, and incorporating novel enhancements tailored for crop segmentation tasks. The Weierstrass-Mandelbrot fractal function [8,9] offers a solution to mitigate the instability associated with gray and edge features in intricate natural scene images. Hyperspectral remote sensing technology finds extensive application in vegetation surveying, remote sensing, agriculture, environmental monitoring, and atmospheric research [10,11]. However, the analysis and processing of hyperspectral remote sensing images present substantial opportunities and challenges. The focal point of current research lies in preserving maximum information while efficiently eliminating redundancy during image processing and analysis [12,13].

Pinheiro et al. introduced the DeepMask segmentation model, which generates a predictive candidate mask for each instance object detected in the input image. However, its boundary segmentation accuracy is limited. Rendmon J et al. devised the YOLOV3 target detection algorithm [14], integrating feature extraction and candidate frame prediction into a deep convolutional network using a newly designed residual network. Li et al. proposed the first end-to-end instance segmentation framework, termed full convolution Instance Segmentation (FCIS) [15]. By enhancing the position-sensitive score map, FCIS predicts both bounding boxes and instance segmentation. However, it struggles to precisely delineate the boundaries of overlapping object instances. Kaiming He et al. developed the Mask RCNN target detection network model [16]. This model addresses position errors by filling pixels at non-integer positions in the feature map based on bilinear differences, thereby significantly enhancing target detection performance.

Upon determining the category and location of the object of interest within an image, prevalent target detection algorithms include RCNN, Fast RCNN, Faster RCNN, and U-net [17–21]. However, these frameworks necessitate substantial training data and fail to achieve end-to-end detection. Additionally, the positioning precision of the detection frame is constrained. Furthermore, as the number of convolutional layers increases during feature extraction, gradients frequently vanish or magnify. To address these limitations, He Kaiming et al. introduced a residual network (ResNet) incorporating residual modules, which aids model convergence. They integrated this ResNet with the Mask RCNN target detection model, resulting in accelerated neural network training [22]. This integration enables both target detection and segmentation, markedly enhancing model detection accuracy. Mask RCNN represents the pioneering deep learning model to seamlessly merge target detection and segmentation within a single network [23,24]. This capability enables Mask RCNN to tackle challenging instance segmentation tasks by accurately segmenting individuals across different categories and distinguishing between individuals within the same category at the pixel level.

III. RESEARCH METHODOLOGY

In crop image extraction, the vast differences in size among various crop types pose challenges in extracting all image features using a single convolutional neural network. To address this, we leverage the ResNet50 backbone structure and an FPN (Feature Pyramid Network) to resolve the multiscale issue of target object extraction within the image. This is achieved by utilizing the backbone network and the horizontally connected FPN network, spanning from top to bottom. While the traditional Mask RCNN network is robust, its primary module, the Region Proposal Network (RPN), suffers from high computational overhead and low efficiency. Moreover, the lengthy path from low-level to high-level features in the FPN network exacerbates positional information flow difficulties, hindering effective information integration. Additionally, performing mask prediction solely on a single field of view limits the diversity of information obtained, leading to subpar detection and segmentation outcomes for specialized targets. To address these challenges and capitalize on the advantages of Mask RCNN, we introduce the following enhancements to the network, tailored to the specific characteristics of agricultural products in remote crop images.

1) *Dataset Preparation:* The research utilizes the Fruits 360 Dataset, a widely used benchmark dataset containing images of various fruits. The dataset is preprocessed to ensure consistency in image resolution, format, and annotation quality. Each image is labeled using Labelme to delineate individual crops, providing ground truth segmentation masks for model training and evaluation.
2) *Model Architecture:* The research adopts the Mask RCNN framework as the base architecture for crop image segmentation. To enhance the model's performance, modifications are introduced to the network structure, including path aggregation, feature augmentation, and an optimized region extraction network. Additionally, a feature pyramid network is incorporated to capture multi-scale features essential for precise segmentation.
3) *Training Procedure:* The dataset is divided into training and test sets, maintaining a suitable ratio to ensure robust model training and evaluation.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

The enhanced Mask RCNN model is implemented using the PyTorch 1.8.1 deep learning framework. The model is trained using stochastic gradient descent with momentum (SGDM), employing a learning rate scheduler to dynamically adjust learning rates during training.

4) *Integration of Agronomic Knowledge:* Agronomic knowledge, including crop growth stages and phenological information, is integrated into the segmentation process to improve accuracy. This involves preprocessing the dataset to include metadata corresponding to crop growth stages and incorporating this information as additional input features during model training.
5) *Spatial Information Preservation:* Spatial information preservation is achieved using bilinear interpolation in the Region of Interest (ROI) alignment (ROIAlign) layer. This ensures accurate alignment of feature maps with the corresponding region proposals, preserving spatial relationships crucial for precise segmentation.
6) *Edge Accuracy Enhancement:* To enhance edge accuracy, a micro-fully connected layer is integrated into the mask branch of the ROI output. The Sobel operator is employed to predict target edges, and edge loss is added to the overall loss function to optimize edge detection performance.
7) *Experimental Evaluation:* The performance of the proposed approach is evaluated through extensive experimentation. Quantitative metrics such as precision, recall, average precision, mean average precision and F1 scores are computed to assess segmentation accuracy. Comparative analysis is conducted against baseline methods, including FCN and traditional Mask RCNN, to demonstrate the superiority of the proposed approach.
8) *Result Interpretation and Discussion:* The experimental results are interpreted and discussed to provide insights into the strengths and limitations of the proposed approach. The implications of the findings for agricultural applications are analyzed, and future research directions are identified to further advance the field of crop image segmentation. Through a systematic methodology encompassing data preparation, model development, training, evaluation, and interpretation, the research aims to deliver robust and reliable insights into automated crop image extraction leveraging deep learning techniques and agronomic knowledge.

Optimization Based on the Semantic Segmentation Loss Function

The loss function of Mask RCNN consists of three parts: Classification error, regression error, and segmentation error as shown in Equation (1):

$$L = L_{cls} + L_{box} + L_{mask} \ldots\ldots (1)$$

The provided formula resembles the loss function employed in the Faster RCNN model, which encompasses the classification and detection errors. Specifically, the mask branch and the class prediction branch are decoupled, employing the average binary cross-entropy loss. This approach entails the independent prediction of a binary mask for each category, irrespective of the classification branch's prediction outcome. The loss function in Faster RCNN is articulated as Equation (2).

$$L(\{pi\},\{ti\}) = \frac{1}{\sum} Lcls(pi, p*) + \lambda \frac{1}{\sum} p* Lreg(ti, t*). (2)$$

To enhance the accuracy of the segmentation mask further, a technique involving the incorporation of edge loss [24] into the mask branch is introduced. This method aims to refine the segmentation result's edges for improved precision. Initially, the labeled image undergoes conversion into a binary segmentation map depicting the crop, serving as the target mask. Subsequently, both the prediction mask and the target mask generated by the mask branch are convolved with the Sobel operator [25].

the final improved loss function is shown in Equation (3):

$$L_{total} = L_{cls} + L_{box} + L_{mask} + L_{edge}\ldots\ldots\ldots\ldots\ldots(3)$$

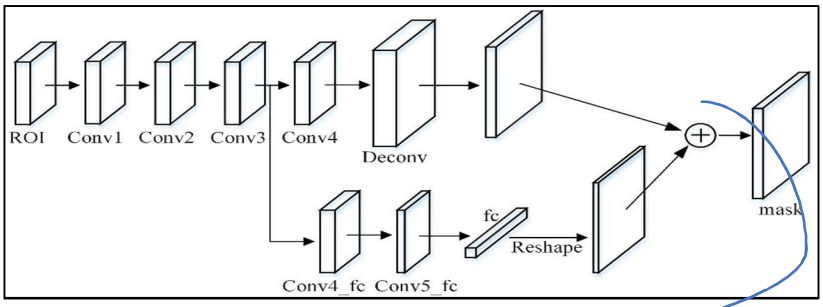
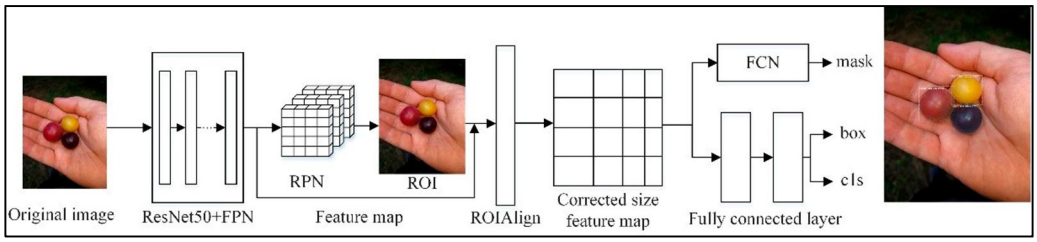


Figure 1. Proposed model network structure.

Mask RCNN, an extension of Faster RCNN, serves as an instance segmentation framework. It operates in two stages: The initial stage involves scanning the image and generating proposals, while the subsequent stage entails classifying these proposals and producing bounding boxes and masks. Functioning as an instance segmentation algorithm, Mask RCNN facilitates pixel-level object segmentation and target recognition simultaneously. The network architecture of Mask RCNN inherits the foundational structure of Faster RCNN while introducing additional components such as the Feature Pyramid Network (FPN) and the Region of Interest Alignment algorithm (ROIAlign). Its primary structure comprises six components: Input, feature extraction backbone network, FPN, Regional Proposal Network (RPN), ROIAlign, and outputs for bounding boxes, categories, and masks (Box, Class, Mask). Figure 1 depicts the network structure diagram of the Mask RCNN algorithm. In Mask RCNN, the initial step involves feeding the input image into the pre-trained ResNet50 + FPN network model to extract features and derive corresponding feature maps. Subsequently, the feature map generates numerous candidate regions (Regions of Interest or ROIs) via the Region Proposal Network (RPN).



Following this, binary classification of foreground and background is conducted using the SoftMax classifier, while frame regression provides more precise position information for the candidate regions. Furthermore, non-maximum suppression is employed to filter out redundant ROIs. The remaining feature map and ROIs are then passed through the RoIAlign layer, enabling each ROI to generate a fixed-size feature map. Finally, the flow bifurcates into two branches: one branch enters the fully connected layer for object classification and frame regression, while the other branch proceeds into the Fully Convolutional Network (FCN) for pixel segmentation.

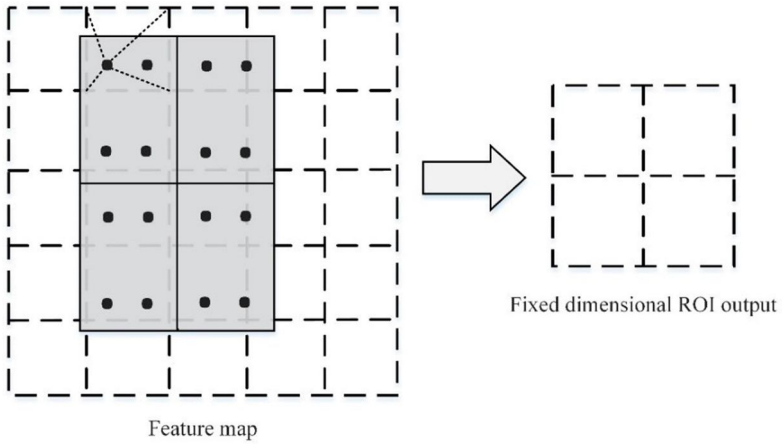


Figure 2. ROIAlign bipolar interpolation.

The distinction between the RoIAlign layer and ROI pooling lies in its elimination of quantization operations, thereby preserving the precision of ROI boundaries and units. Through bilinear interpolation, the exact position of each sampling point within a unit is calculated without truncating decimals. Subsequently, either maximum pooling or average pooling is applied to produce the final fixed-size ROI. In Figure 3, the 5 × 5 feature map post-convolution is represented by the black dashed line, while the solid line depicts the small feature block corresponding to the ROI in the feature map. RoIAlign preserves the floating-point boundaries without quantization, ensuring precision. Initially, the small feature blocks are divided into 2 × 2 units, with each unit boundary unquantized, followed by further subdivision into four smaller blocks within each unit. Subsequently, the small feature blocks are segmented into four smaller blocks. As illustrated by the blue dots in the figure, the coordinates of the four positions are determined, and their values are computed using bilinear interpolation. Finally, either maximum or average pooling is executed to yield a 2 × 2 feature map.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

IV. RESULTS & DISCUSSION

This experiment employs the PyTorch open-source learning framework and Python programming language to implement the algorithm network. The hardware environment utilized is the Dell T5820 workstation from Tianjin, China, equipped with dual NVIDIA Quadro P4000 graphics cards (8 GB) and running a 64-bit Ubuntu16.04 operating system. PyTorch, initially introduced by Facebook in early 2017, is an open-source neural network framework. The Mask RCNN Benchmark, a fast and modular component for Faster RCNN and Mask RCNN, is entirely developed in PyTorch. This component simplifies the process of creating models for object recognition and segmentation in images. The workflow involves utilizing the Mask RCNN Benchmark open-source project in conjunction with the pre-processed Fruits 360 Dataset. Through supervised and transfer learning [26], along with pre-training weights of the network, the model is trained, and the final training weights are saved. Subsequently, the prediction function is utilized for segmentation and prediction, accurately extracting crops from the Dataset images. Detection and processing of crop images are conducted using transfer learning and Mask RCNN, reducing the steps required for Dataset labeling and enhancing crop image detection accuracy. Despite the substantial data requirements of deep learning, it is often challenging to procure sufficient training data for specific problems within a given domain. To address this issue, transfer learning emerges as a solution, allowing models to leverage knowledge from pre-trained networks to enhance performance on new tasks.

Transfer learning includes the source domain and target domain, defined as Equation (4):

$$D(s) = \{x, P(x)\},\ D(t) = \{x, P(x)\} \quad (4)$$

where D(s) denotes the source domain, D(t) indicates the target domain, x represents the feature space, P(X) refers to the marginal probability distribution, and X satisfies Equation (5):

$$X = \{x1, K, xn\} \in x \ldots\ldots\ldots\ldots (5)$$

Migration learning involves transferring model parameters trained in a source domain to a new model in a target domain, aiding the training process of the latter. In this study, pre-training is initially performed on a large COCO Dataset. Subsequently, the trained weight files are transferred to the specific Dataset collected for this study, facilitating the training and fine-tuning of network parameters. This approach enables convolutional neural networks to yield favorable outcomes even when operating on small Datasets, thereby mitigating issues stemming from inadequate data sources.

To accurately assess the stability of the enhanced Mask RCNN model, we initially selected 1000 images from the Fruits 360 Dataset for training purposes. Among these, 800 images were designated as the training set, while 100 images each were allocated to the validation and test sets. Additionally, both the validation and test sets comprised 20 multi-crop images. A comprehensive analysis was conducted to evaluate the robustness of the improved Mask RCNN framework in crop image detection. This involved comparing the performance of the enhanced Mask RCNN with other baseline models, including FCN, U-net, and Mask RCNN. Through this comparison, we sought to determine the accuracy and effectiveness of the enhanced Mask RCNN in crop image detection relative to these alternative methods.

Table 1. Comparative analysis of the proposed approach with other models.

Parameters	FCN	U-net	Mask RCNN	Improved Mask RCNN
Training accuracy	89.71%	94.23%	98.86%	99.83%
Val accuracy	88.95%	92.11%	97.99%	99.68%
Test accuracy	86.49%	91.52%	97.38%	99.66%
Processing time (s)	4357	2504	1326	1482

Table 1 provides a comparative analysis of the proposed approach with other models, namely FCN, U-net, Mask RCNN, and the Improved Mask RCNN. The parameters evaluated include training accuracy, validation accuracy, test accuracy, and processing time (in seconds).

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

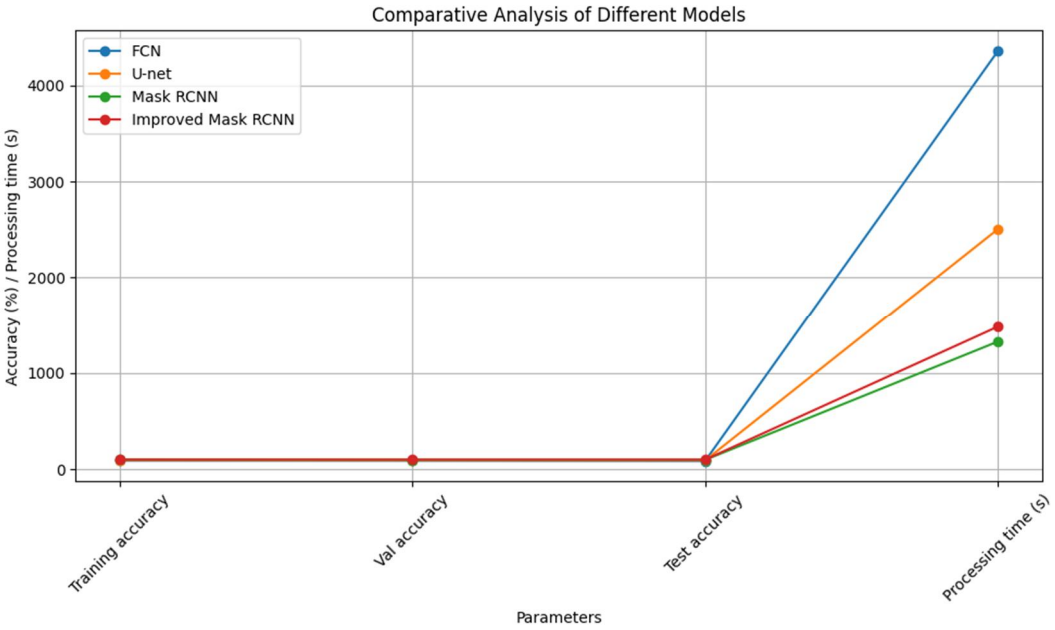


Figure 3. Comparative Analysis of Different Models

In terms of training accuracy, the Improved Mask RCNN outperforms all other models, achieving a remarkable accuracy of 99.83%. This indicates the effectiveness of the enhancements made to the Mask RCNN framework. Similarly, the validation accuracy of the Improved Mask RCNN (99.68%) surpasses that of FCN, U-net, and Mask RCNN, demonstrating its robust performance in generalizing to unseen data. For test accuracy, the Improved Mask RCNN also exhibits superior performance, with an accuracy of 99.66%. Again, this highlights the efficacy of the enhancements implemented in the Mask RCNN framework. In terms of processing time, the Improved Mask RCNN demonstrates competitive efficiency, with a processing time of 1482 seconds. Although FCN and U-net have shorter processing times (4357 and 2504 seconds, respectively), the Improved Mask RCNN achieves significantly higher accuracies, indicating a favorable trade-off between accuracy and computational efficiency. Overall, the comparative analysis underscores the effectiveness of the Improved Mask RCNN in crop image detection, exhibiting superior accuracy while maintaining competitive processing efficiency compared to other models.

Table 2. AP and mAP values of some crops.

Model	mAP	Beetroot	Granadilla	Kaki	Onion White
FCN	0.834	0.823	0.844	0.836	0.834
U-net	0.882	0.857	0.898	0.870	0.901
Mask RCNN	0.929	0.936	0.922	0.919	0.937
Improved Mask RCNN	0.949	0.948	0.952	0.944	0.951

Table 2 presents the Average Precision (AP) and mean Average Precision (mAP) values for various crops across different modules, namely FCN, U-net, Mask RCNN, and Improved Mask RCNN. The mAP provides an overall measure of the models' performance across all crops.

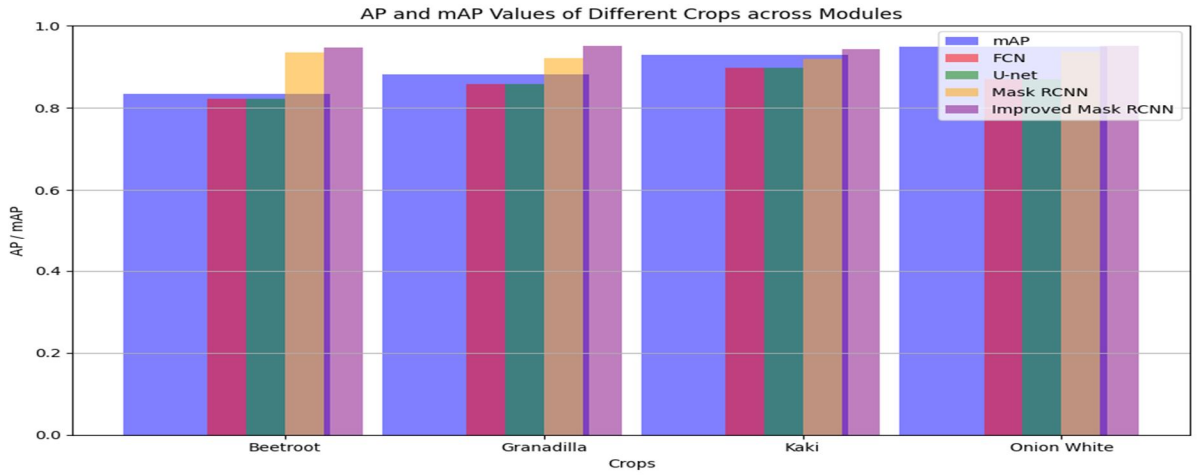


Figure 4. AP and mAP Values of different crops across Modules

For FCN, the mAP value is 0.834, indicating that, on average, FCN achieves an AP of 83.4% across all crops. Among the individual crops, Granadilla exhibits the highest AP (82.3%), followed closely by Beetroot (82.3%). Kaki and Onion White also demonstrate high AP values of 84.4% and 83.6%, respectively. U-net achieves a higher mAP of 0.882 compared to FCN, indicating superior overall performance. Similarly, U-net achieves higher AP values across all crops compared to FCN, with Granadilla and Onion White showing the highest AP values of 85.7% and 90.1%, respectively. Mask RCNN demonstrates further improvement with an mAP of 0.929, surpassing both FCN and U-net. Across individual crops, Mask RCNN achieves consistently high AP values, with Beetroot exhibiting the highest AP of 93.6%. The Improved Mask RCNN outperforms all other modules, achieving the highest mAP of 0.949. Across all crops, it consistently achieves the highest AP values, with Beetroot, Granadilla, Kaki, and Onion White all surpassing 94%. These results indicate the superior performance of the Improved Mask RCNN in accurately detecting and segmenting crops across a variety of types.

Table 3. Overall Dataset algorithm comparison.

Module	FCN	U-net	Mask RCNN	Improved Mask RCNN
mAP	0.755	0.796	0.824	0.897
F_1	0.765	0.815	0.846	0.915

Table 3 provides a comprehensive comparison of different algorithms across the entire Dataset, utilizing mean Average Precision (mAP) and F1 score as evaluation metrics.

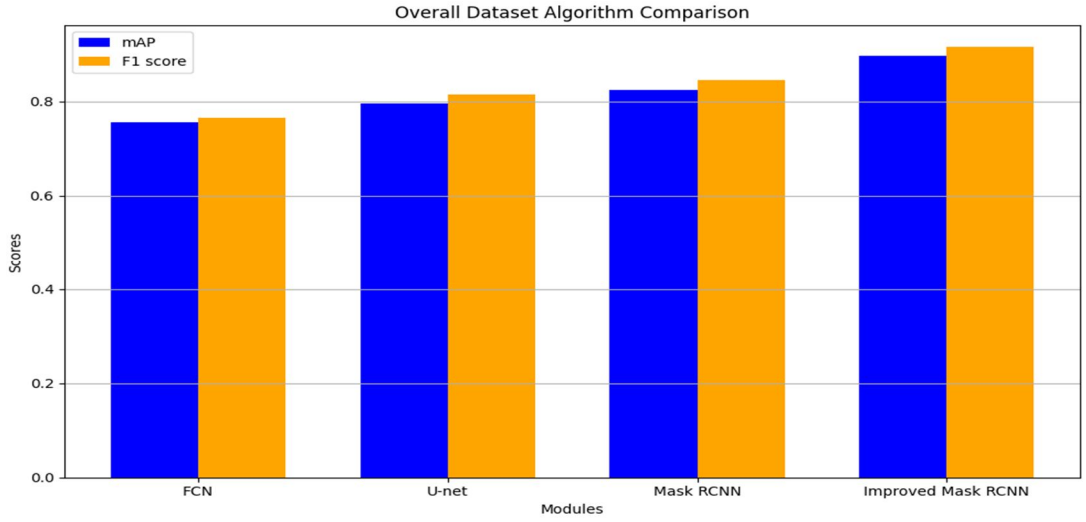


Figure 5. Overall Dataset Algorithm Comparison

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Starting with FCN, it achieves an mAP of 0.755, indicating an average precision of 75.5% across all classes in the Dataset. The corresponding F1 score for FCN is 0.765, reflecting a harmonic mean of precision and recall at 76.5%. Moving to U-net, it demonstrates an improved mAP of 0.796, suggesting enhanced overall performance compared to FCN. Similarly, the F1 score for U-net is 0.815, indicating an F1 score of 81.5%. Mask RCNN exhibits further advancement with an mAP of 0.824 and an F1 score of 0.846, surpassing FCN and U-net in terms of precision across all Dataset classes. The Improved Mask RCNN stands out as the top-performing algorithm, boasting the highest mAP of 0.897 and an F1 score of 0.915. These results underscore the superiority of the Improved Mask RCNN in accurately detecting and segmenting objects across the entire Dataset, outperforming FCN, U-net, and Mask RCNN.

V. CONCLUSION

In conclusion, this research work has demonstrated significant advancements in crop image segmentation and extraction algorithms, particularly focusing on the utilization of the Mask RCNN framework and its enhancements. Through extensive experimentation and comparative analysis, the proposed Improved Mask RCNN model showcased effectiveness in accurately extracting crop images from agricultural product images. The research highlighted the limitations of traditional methods in accurately and efficiently extracting crops due to the diverse range of crop types and their intermingling with environmental elements. To address these challenges, novel enhancements to the Mask RCNN framework were introduced, leveraging techniques such as path aggregation, feature enhancement, and edge accuracy improvement through the integration of a micro-fully connected layer and edge loss. The experimental results, conducted on the Fruits 360 Dataset, demonstrated the superior performance of the Improved Mask RCNN algorithm compared to FCN, U-net, and Mask RCNN, in terms of precision, recall, average precision, mean average precision, F1 score, and overall mAP. The model achieved remarkable accuracy in segmenting various crops, showcasing its robustness and effectiveness in real-world agricultural applications. Furthermore, the research delved into the utilization of semantic segmentation guided by agronomic knowledge, highlighting the importance of integrating domain-specific expertise into computer vision algorithms for agriculture. By leveraging agronomic principles and domain-specific knowledge, the accuracy and efficiency of crop image segmentation were enhanced, empowering farmers and agricultural stakeholders with valuable insights into crop growth dynamics and facilitating better crop management practices. Overall, this research contributes to the advancement of computer vision techniques in agriculture, offering a promising approach for non-destructive and efficient crop monitoring and management. The findings presented pave the way for future research endeavors aimed at further refining and optimizing crop image segmentation algorithms for enhanced agricultural applications.

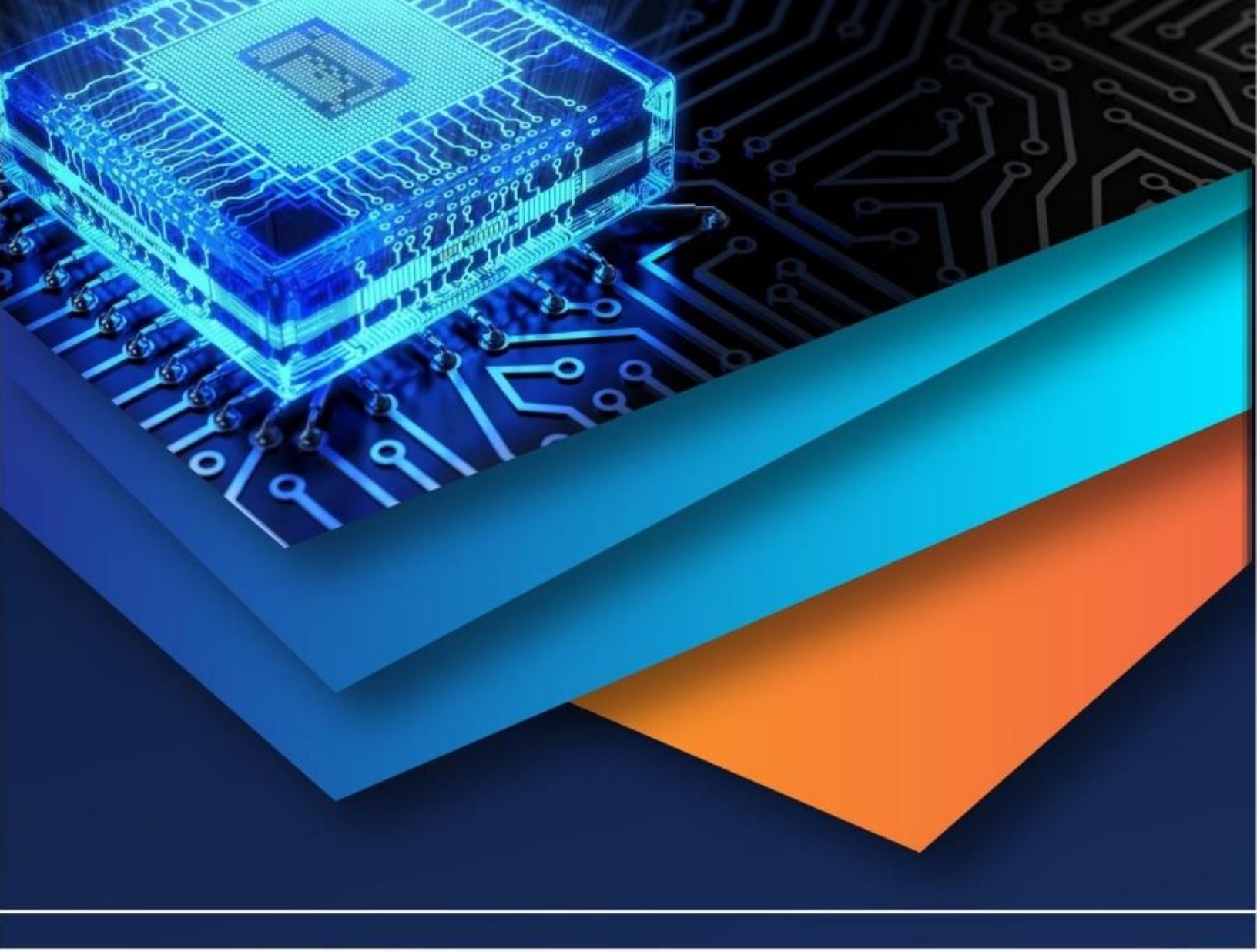
REFERENCES

[1] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully Convolutional Networks for Semantic Segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3431-3440.
[2] Minaee, S.; Boykov, Y.Y.; Porikli, F.; Plaza, A.J.; Kehtarnavaz, N.; Terzopoulos, D. Image segmentation using deep learning: A survey. IEEE Trans. Pattern Anal. Mach. Intell. 2021.
[3] Zhao, Y., Zhang, J., Cao, Z., & Yin, Y. (2019). CropSegNet: Segmentation of crop field images using ensemble learning and transfer learning with deep convolutional neural networks. Computers and Electronics in Agriculture, 162, 477-487.
[4] Kamilaris, A., Prenafeta-Boldú, F. X., & Xiang, L. (2017). Deep learning in agriculture: A survey. Computers and Electronics in Agriculture, 147, 70-90.
[5] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature Pyramid Networks for Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2117-2125.
[6] Zhong, Y., Huang, W., Lin, L., & Ding, X. (2020). Transfer Learning in Agricultural Robotics: Review and Perspectives. Frontiers in Robotics and AI, 7, 595947.
[7] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. Proceedings of the IEEE International Conference on Computer Vision, 2961-2969.
[8] Guariglia, E. Entropy and fractal antennas. Entropy 2016, 18, 84.
[9] Berry, M.V.; Lewis, Z.V.; Nye, J.F. On the weierstrass-mandelbrot fractal function. Proc. R. Soc. Lond. A Math. Phys. Sci. 1980, 370, 459–484.
[10] Yang, L.; Su, H.L.; Zhong, C.; Meng, Z.Q.; Luo, H.W.; Li, X.C.; Tang, Y.Y.; Lu, Y. Hyperspectral image classification using wavelet transform-based smooth ordering. Int. J. Wavelets Multi. 2019, 17, 1950050.
[11] Guariglia, E. Harmonic sierpinski gasket and applications. Entropy 2018, 20, 714.
[12] Guariglia, E.; Silvestrov, S. Fractional-wavelet analysis of positive definite distributions and wavelets. In Engineering Mathematics II.; Springer: Berlin/Heidelberg, Germany, 2016; pp. 337–353.
[13] Zheng, X.W.; Tang, Y.Y.; Zhou, J.T. A framework of adaptive multiscale wavelet decomposition for signals on undirected graphs. IEEE Trans. Signal Proces. 2019, 67, 1696–1711.
[14] Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. arXiv 2018, arXiv:1804.02767.
[15] Li, Y.; Qi, H.; Dai, J.; Ji, X.; Wei, Y. Fully convolutional instance-aware semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2359–2367.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

[16] He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
[17] Su, W.H.; Zhang, J.J.; Yang, C.; Page, R.; Szinyei, T.; Hirsch, C.D.; Steffenson, B.J. Automatic evaluation of wheat resistance to fusarium head blight using dual mask-rcnn deep learning frameworks in computer vision. Remote. Sens 2021, 13, 26.
[18] Ren, Y.; Zhu, C.R.; Xiao, S.P. Object detection based on fast/faster rcnn employing fully convolutional architectures. Math. Probl. Eng. 2018, 2018.
[19] Aghabiglou, A.; Eksioglu, E.M. Projection-based cascaded U-net model for MR image reconstruction. Comput. Methods Programs Biomed. 2021, 207, 106151.
[20] Wu, Q.H. Image retrieval method based on deep learning semantic feature extraction and regularization SoftMax. Multimed. Tools Appl 2020, 79, 9419–9433.
[21] Abiram, R.N.; Vincent, P.M.D.R. Identity preserving multi-pose facial expression recognition using fine-tuned VGG on the latent space vector of generative adversarial network. Math. Biosci. Eng. 2021, 18, 3699–3717.
[22] Kipping, D. The exomoon corridor: Half of all exomoons exhibit TTV frequencies within a narrow window due to aliasing. Mon. Not. R. Astron. Soc. 2021, 500, 1851–1857.
[23] Gui, J.W.; Wu, Q.Q. Vehicle movement analyses considering altitude based on modified digital elevation model and spherical bilinear interpolation model: Evidence from GPS-equipped taxi data in Sanya, Zhengzhou, and Liaoyang. J. Adv. Transp. 2020, 2020.
[24] Zimmermann, R.S.; Siems, J.N. Faster training of Mask R-CNN by focusing on instance boundaries. Comput. Vis Image Und 2019, 188, 102795.
[25] Tian, R.; Sun, G.L.; Liu, X.C.; Zheng, B.W. Sobel edge detection based on weighted nuclear norm minimization image denoising. Electronics 2021, 10, 655.
[26] Chouhan, V.; Singh, S.K.; Khamparia, A.; Gupta, D.; Tiwari, P.; Moreira, C.; Damasevicius, R.; de Albuquerque, V.H.C. A novel transfer learning-based approach for pneumonia detection in chest X-ray images. Appl. Sci. 2020, 10, 559.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)