UAV-Enabled Semantic Segmentation for Precision Farming Using Deep Learning

Master of Science (Research) Thesis

by

Aditya Kanade

Under the supervision of

Prof. Somnath Dey and Dr. Ayan Mondal

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

July 2025

UAV-Enabled Semantic Segmentation for Precision Farming Using Deep Learning

A THESIS

Submitted in partial fulfillment of the

requirements for the award of the degree

of

Master of Science (Research)

by

Aditya Kanade

2304101002

Under the supervision of

Prof. Somnath Dey and Dr. Ayan Mondal

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

July 2025

INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **UAV-Enabled Semantic Segmentation for Precision Farming Using Deep Learning** in the partial fulfillment of the requirements for the award of the degree of **Master of Science (Research)** and submitted in the **Department of Computer Science and Engineering, Indian Institute of Technology Indore,** is an authentic record of my own work carried out during the period from July 2023 to July 2025 under the supervision of Prof. Somnath Dey and Dr. Ayan Mondal, Indian Institute of Technology Indore, India.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

AKanade

20/July/2025
Signature of the Student with Date

(Aditya Kanade)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Somnath Dey

AMondal

Signature of Thesis Supervisors with Date

(Prof. Somnath Dey) **(Dr. Ayan Mondal)**

ACKNOWLEDGEMENTS

I would like to express my heartfelt gratitude to my supervisor, **Prof. Somnath Dey**, Professor, Department of Computer Science and Engineering, IIT Indore, and **Dr. Ayan Mondal**, Assistant Professor, Department of Computer Science and Engineering, IIT Indore, for his unwavering support, expert guidance, and constant encouragement throughout the course of my M.S.(R) thesis titled UAV- Enabled Semantic Segmentation for Precision Farming Using Deep Learning. His deep knowledge, clear vision, and insightful feedback were invaluable at every stage of this research. I am truly thankful for the time and effort he dedicated to mentoring me and helping me grow as a researcher.

I would like to sincerely thank **Prof. Suhas S. Joshi**, Director, IIT Indore, for providing an excellent research environment and all the necessary resources that enabled me to carry out this work. His leadership and commitment to academic excellence have been truly inspiring.

I also extend my appreciation to all the faculty members of the CSE Department at IIT Indore for their valuable guidance, academic mentorship, and support during my MS program. The knowledge and skills I gained through their courses and interactions have greatly contributed to the successful completion of this thesis.

Lastly, I am thankful to everyone who supported me, directly or indirectly, during this journey. Your encouragement and belief in me have been a great source of motivation.

Aditya Kanade
Master of Science (Research)
Department of Computer Science and Engineering
Indian Institute of Technology Indore

Dedicated to My Family

ABSTRACT

In this work, we study the semantic segmentation of captured UAV (Unmanned Aerial Vehicle) images for enhanced crop and weed segmentation in precision agriculture. In the existing literature, researchers studied segmentation techniques; however, there is a need for deep feature extraction to capture the spatial and contextual information, especially for the complex agriculture domain, which involves the overlapping of crop, weed, and background pixels. To address this, we present VResUNet++ architecture that combines VGG16 and ResNet50 in the backbone of UNet for deep semantic feature extraction and helps in improving the segmentation accuracy and performance. This improved segmentation method helps in weed detection, crop health monitoring, and early disease detection. Our hybrid model has outperformed the state-of-the-art models like UNet, UNetResNet50, and UNetVGG16. The outcome of the extensive experiment shows a significant improvement in the precision of 99.83%, recall of 98.65%, and accuracy of 98.69% on the Weedmap dataset.

Contents

List of Figures

List of Tables

List of Abbreviations and Acronyms

UAV Unmanned Aerial Vehicle

DNN Deep Neural Network

CNN Convolutional Neural Network

VGG Visual Geometry Group 16-layer Network

RGB Red Green Blue

HSI Hue Saturation Intensity

ACCF Advanced Comprehensive Color Feature

SVD Single Value Decomposition

CEI Contrast Enhanced Image

MLT Multilevel Thresholding

R-MCE Recursive Minimum Cross-Entropy

CSA Cuckoo Search Algorithm

MRF Markov Random Field

KNN K-Nearest Neighbors

RF Random Forest

VI Vegetation Index

FCN Fully Convolution Network

AISA Automatic Image Segmentation Algorithm

CRF Conditional Random Field

CBAM Concentration based attention Module

DIC Depth Identity Convolution

SPA Spatial Attention

CHA Channel Attention

DAFPN Double Attention Mechanism Feature Pyramid Network

VHR Very High Resolution

DCPM Double Cross Pooling Method

Respath Residual Path

ReLU Rectified Linear Unit

ResNet Residual Network

ViT Vision Transformer

VIA VGG Image Annotator

SAM Segment Anything Model

TP True Positive

FP False Positive

FN False Negative

TN True Negative

IoU Intersection over Union

SOTA State-of-the-Art

Chapter 1

Introduction

1.1 Background

The agriculture sector is significantly transforming with the help of advanced technologies that increase productivity, sustainability, and efficiency. One of the growing innovations is Unmanned Aerial Vehicle (UAVs), commonly called drones, which are expanding as an essential tool in the emergence of precision agriculture. It provides high-resolution, real-time data, which helps farmers and agronomists in monitoring crops with remarkable accuracy and assists management strategies based on the site, which were previously unreachable with traditional methods.

(a) Soil Loosening

(b) Spraying Pesticides

Figure 1.1: Traditional Farming Methods [1]

1.2 Precision Agriculture and Its Components

Precision agriculture includes the collection and analysis of detailed spatial and temporal data to efficiently use the agricultural inputs such as water, fertilizers, and pesticides. The main purpose is to maximize the crop yield while minimizing the effects on the environment and reducing the wastage of resources. UAVs play a crucial role in this major shift by providing flexible, cost-effective, and reliable data collection over large areas. In contrast, satellite imagery is limited to lower spatial resolution and cloud cover. UAVs can capture detailed images as per the user's convenience, irrespective of the weather conditions. This ability is beneficial for monitoring the health of the crops, detecting diseases, weed detection, and yield estimation.

(a) Soil Loosening

(b) Spraying Pesticides

Figure 1.2: Modern Farming Methods [2]

1.3 Importance of Crop Classification

The rapid growth of UAVs in agriculture has been increasing with machine learning and deep learning techniques. Deep learning algorithms can learn from large datasets for complex representation. They have shown significant success in computer vision tasks such as image classification, object detection, and semantic segmentation. Semantic segmentation is important in precision agriculture because it allows the pixel-level classification of UAV-captured images. This assessment is critical for differentiating between crops, weeds, soil, and other classes that support multiple applications such as yield detection, weed mapping, disease detection, and crop health monitoring.

1.4 Challenges in Crop Classification

Although there are several advances, however, there exist multiple challenges in the deployment of deep learning models for semantic segmentation in precision agriculture. The fundamental challenge is diverse field conditions, which include different lighting, soil type, crop variety, and multiple growth stages. These differences greatly affect the appearance of crops and weeds in UAV images, which makes it difficult for the model to train on these datasets and later generalize to new environments. Moreover, the collection and annotation of a high-quality dataset are manpower-demanding and time-consuming, which results in a limited training dataset for specific crops or regions.

Another common challenge in precision agriculture is class imbalance. For example, an image contains a wide area of background or a crop with a few pixels of weeds. This results in the favor of the majority of classes, which leads to poor detection and segmentation of minority classes that are of great interest to the farmers. Additionally, many state-of-the-art deep learning models, like UNet and DeeplabV3+, rely on an encoder network that can lose the fine spatial details during the downsample phase. This loss of details is unfavorable in precision agriculture, where fine details of small or overlapping regions are important for determining individuals such as crops, weeds, or backgrounds.

The deployment of deep learning models on UAVs is computationally intensive. Since the hardware of UAVs is lightweight and power-efficient, it is not capable of running complex models in real time. This limitation demands the need for the deployment of an efficient architecture that balances accuracy with computational efficiency and real-time image processing of the field.

1.5 Research Gap

To overcome these challenges, recent researchers are focusing on agricultural applications based on deep learning models. One potential solution is the use of a hybrid encoder backbone, which blends the strengths of different convolutional neural network (CNN) architectures to improve both semantic recognition and spatial detail preservation. For instance, the integration of ResNet50 brings the residual learning capabilities, while VGG16 brings the simplicity and effectiveness in acquiring the spatial features. These settings can make the segmentation model more robust and accurate. Furthermore, the overfitting can be improved with the help of some regularization techniques like dropout and L2 regularization, along with the transfer learning and fine-tuning of the model, which can generalize the model to new datasets.

1.6 Objectives of the Work

In our thesis, we propose a modified UNet architecture in which we are modifying the encoder part of the UNet by combining the features of ResNet50 and VGG16 to address the challenges of segmentation in precision agriculture. This model improves pixel-level classification by enhancing the extraction of both high-level semantic features and fine spatial details. For further performance improvement, we are incorporating advanced regularization methods and training and validating on diverse UAV-captured datasets.

1.7 Contribution of the Thesis

This thesis makes the following key contributions:

- Proposes a novel hybrid UNet architecture that integrates ResNet50 and VGG16 as encoder backbones to improve segmentation accuracy and spatial detail preservation.
- Employs regularization methods such as dropout and L2 regularization, along with fine-tuning strategies, to enhance model generalization and prevent overfitting.
- Demonstrates significant improvements in multi-class segmentation performance, particularly for minority classes like weeds, using standard metrics such as Dice coefficient, Intersection over Union (IoU), precision, and recall.
- Provides a comprehensive evaluation of the model on diverse UAV datasets, showcasing its potential for real-world deployment in precision agriculture.

1.8 Organization of the Thesis

This thesis is structured into five distinct chapters to present a coherent flow of the research work. The first chapter introduces the background, importance, and challenges of crop classification in the context of precision agriculture. The subsequent chapters of this thesis are organized as follows:

- **Chapter 2: Literature Survey**

  This chapter provides an in-depth survey of related articles based on precision agriculture. It focuses on the methodology and limitations of every study in real-world scenarios. It contains three approaches — vision-based approach, machine learning-based approach, and deep learning-based approach.

- **Chapter 3: Proposed Methodology**

  This chapter explains the methodology adopted in this thesis, beginning with detailed preprocessing steps including image resizing and augmentation techniques like rotation, flipping, cropping, and scaling. It further describes the architectures that are UNet, UNetResNet50, UNetVGG16, and the proposed VResUNet++ architecture.

- **Chapter 4: Experimental Results**

  This chapter outlines the performance of the proposed models. It includes a thorough description of the dataset used and elaborates on the evaluation parameters, namely precision, recall, accuracy, dice coefficient, and intersection of union. It also discusses the training and validation outcomes over multiple epochs for all models, supported by relevant tables.

- **Chapter 5: Conclusions and Future Work**

  The concluding chapter wraps up the thesis by highlighting the main findings and contributions. It reflects on the overall model performance, highlights the strengths and challenges observed during implementation, and proposes future research directions, including model optimization, real-time deployment, and expanding the dataset for broader generalizability.

Chapter 2

Literature Review

In this chapter, we have reviewed the existing work done in the same area of precision farming. The existing approach can be divided into 3 categories: vision-based approaches, machine learning-based approaches, and deep learning-based approaches.

2.1 Vision-Based Approaches

Ikonomakis et al. [6] introduced a seeded region growing and merging algorithm for segmenting both grayscale and color images, starting with seed pixels and growing regions based on homogeneity functions. The method effectively partitions images into homogeneous regions, with adjustable parameters that allow fine control over the segmentation process, making it a valuable tool for image analysis tasks. In another research, Jiang et al. [7] presented a color-based image processing method for agricultural image segmentation, where crops are separated from solid backgrounds by analyzing their color differences. Among several tested transformations in RGB and HSI color spaces, the excess green index (2G-R-B) proved to be the most effective, offering fast segmentation, good quality, and robustness to varying sunlight, making it suitable for real-time agriculture navigation and target extraction. Guijarro et al. [8] proposed a combined image segmentation strategy for automatically identifying green plants, soil, and sky textures in barley and corn images by integrating multiple thresholding methods and supervised fuzzy clustering. The approach improves segmentation accuracy under varying illumination, aiding tasks like site-specific treatments and autonomous robot navigation in agriculture. Sarkate et al. [9] developed another color-based image segmentation method for precise yield prediction of gerbera flowers by applying thresholding in the HSV color space and histogram analysis. Flowers were effectively extracted and counted from polyhouse images despite some errors caused by field conditions like illumination and overlapping. Jothiaruna et al. [10] have proposed a segmentation approach for disease spots on leaves, using Advanced Comprehensive Color Feature(ACCF) and Region Growing to address challenges like cluttered backgrounds and uneven illumination. The ACCF utilizes multiple color spaces and singular value decomposition (SVD) for enhanced disease spot detection. At the same time, the region-growing method eliminates background noise, achieving an average segmentation accuracy of 87% under real field conditions. Tan et al. [11] present an algorithm for accurately segmenting and counting touching hybrid rice grains, integrating watershed segmentation, an improved corner point detection algorithm, and a BP neural network classifier. The method obtained an average accuracy of 94.63%, demonstrating its potential for enhancing automated rice grain evaluation in agricultural production. Xue et al. [12] presented an improved watershed segmentation algorithm for extracting cultivated land boundaries from high-resolution remote sensing imagery, incorporating pre-contrast enhancement and post-region merging in CIE color spaces (Lab and Luv). The results demonstrate improved performance in terms of time efficiency, accuracy, and transferability in comparison to traditional RGB-based methods, with significant improvements in segmentation accuracy, making it a valuable tool for automated agricultural land boundary extraction. Lu et al. [13] introduced a contrast-optimization method for robust plant segmentation from RGB images by creating contrast-enhanced images (CEIs) through an optimized linear combination of RGB channels, followed by automatic thresholding. Evaluated across five diverse datasets, the method achieved an average F1 score of 95%, outperforming traditional color indices by 4%, demonstrating greater robustness under varying imaging conditions, and offering a simple, adaptable approach for plant image analysis in precision agriculture. Kumar et al. [14] present an efficient multilevel thresholding (MLT) approach in the segmentation of crop images, combining the recursive minimum cross-entropy (R-MCE) technique with the cuckoo search algorithm (CSA) based on Lévy flight. The presented method outperformed alternative thresholding methods in the context of accuracy and segmentation standards, achieving better results in PSNR, SSIM, FSIM, and MSE for crop images with complex backgrounds. Cerruto et al. [15] evaluate the impact of image segmentation thresholding on droplet dimensional analysis, particularly in pesticide application. By analyzing images of droplets under different threshold values, it was found that variations in volumetric and mean diameters were linearly correlated with threshold changes, but the absolute errors remained small (less than 1.0%). The study concludes that using an automatic threshold selection tool, depending on the average gray level of the image, can significantly reduce operator dependency, accelerating the segmentation processes and making it more objective.

2.2 Machine Learning-Based Approaches

There are some conventional supervised machine learning techniques employed for precision agriculture. Han et al. [16] have proposed a Support Vector Machine (SVM) for farmland image segmentation, specifically identifying crop rows and minimizing noise influence. Yue et al. [17] presented a Markov Random Field (MRF) approach for rice planthopper image segmentation, which incorporates fractional order differential and multi-feature fusion to enhance texture and color information that helps in demonstrating better robustness in varying field conditions and for real-time pest identification in precision agriculture. Rangel et al. [18] have focused on the early detection of crop health issues like nutritional deficiencies, diseases, and pests by using K-Nearest Neighbors (KNN) for grapevine leaf classification. Tian et al. [19] implemented an improved K-means algorithm that automatically determines the clustering number using the Davies-Boulding index and prevents local optima, which helps in more accurate and efficient segmentation of tomato leaf images. Dávila-Rodríguez et al. [20] proposed decision-tree-based real-time citrus segmentation with 97.1% accuracy. Ramos et al. [21] demonstrated the effectiveness of Random Forest (RF) in precision agriculture applications for yield production by integrating a ranking-based approach using vegetation indices (VIs) to enhance RF's accuracy for maize yield production.

2.3 Deep Learning-Based Approaches

Milioto et al. [22] have proposed a lightweight encoder-decoder CNN for real-time semantic segmentation of crops, weeds, and soil using only RGB data enriched with vegetation indices, achieving fast inference (20+ Hz) suitable for field deployment. The network enhances generalization and training efficiency by integrating task-specific features like ExG, CIVE, and NDI into a 14-channel input. Khan et al. [23] present a segmentation-driven approach using correlation coefficient-based techniques to isolate diseased regions in fruit crops. The infected areas are further processed using deep CNN features from VGG16 and AlexNet for high-accuracy classification. Zhuang et al. [24] have proposed an end-to-end deep semantic segmentation method using fully convolutional networks (FCN) to efficiently segment green vegetation in complex field conditions. By leveraging multi-scale features and vegetation indices, the method achieves real-time segmentation suitable for agricultural automation. Qiao et al. [25] have proposed a Mask R-CNN-based method for cattle segmentation and contour extraction in feedlot environments. Fawakherji et al. [26] proposed a modified U-Net for pixel-wise vegetation/soil segmentation, with a VGG-16 encoder and a custom decoder. Xiong et al. [27] have proposed an Automatic Image Segmentation Algorithm (AISA) based on GrabCut to eliminate background noise and enhance disease feature extraction in crop images. Integrating AISA with the MobileNet CNN improved classification accuracy, especially in complex real-world scenarios. Yue et al. [28] propose an enhanced SegNet model integrated with conditional random fields (CRFs) for improved crop disease image segmentation. By refining the initial segmentation output with CRFs, the method achieves higher segmentation accuracy and faster processing time. Peng et al. [29] proposed a DeepLabV3+ based semantic segmentation model integrated with the Xception_65 backbone for accurate detection of litchi branches. Their approach leveraged transfer learning, atrous spatial pyramid pooling, and shallow-deep feature fusion to enhance robustness and accuracy, particularly for small target detection in agricultural automation. Zhang et al. [30] proposed a modified Mask-RCNN with SE attention for UAV-based litchi segmentation and improved detection in complex environments. Similarly, DAFPN integrated spatial and channel attention into Mask-RCNN for accurate small farmland segmentation, while Mask-R-FCN fused FCN and Mask-RCNN to better handle small object segmentation in remote sensing imagery. Jia et al. [31] proposed an optimized Mask R-CNN model using a ResNet-DenseNet backbone for detecting and segmenting overlapped apples in orchard environments. The method effectively addressed occlusion challenges, enabling accurate real-time fruit recognition for apple harvesting robots. Li et al. [32] proposed the PSPNet model, incorporating MobileNetV2 as the backbone for semantic segmentation of landslides in Tibet's Nyingchi region, offering faster convergence and reduced computational load while maintaining accuracy. You et al. [33] proposed a DNN-based semantic segmentation network for weed/crop recognition, integrating four key components: hybrid dilated convolution with DropBlock, a universal function approximation block for optimized RGB-NIR indices, a bridge attention block for capturing long-range contextual information, and a spatial pyramid refinement block for precise multi-scale feature fusion. Wang et al. [34] proposed an encoder-decoder deep learning model for crop–weed segmentation, emphasizing integrating NIR data and image enhancement techniques to improve accuracy and robustness under varying lighting. Wu et al. [35] used DeepLabV3+ with various backbones like Xception-65, Xception-71, ResNet-50, and ResNet-101 to segment abnormal (yellow, withered, decayed) leaves in hydroponic lettuce for robotic sorting. The ResNet-101 model with uniform weight assignment achieved the best performance. Lv et al. [36] proposed an improved PSPNet model for accurately identifying cotton boll growth stages in complex field backgrounds. The model integrates context coding modules during encoding and shallow feature fusion during decoding. presents an enhanced semantic segmentation model, m-DeepLabV3+, for rapid and accurate detection of straw return in conservation tillage. By introducing MixConv, channel shuffling, a concentration-based attention module (CBAM), and squeeze and extraction networks (SE), the model improves segmentation performance, reduces parameter count, and mitigates overfitting through R-drop. Anand et al. [37] developed AgriSegNet by integrating multi-scale attention modules for semantic segmentation of aerial images. Zou et al. [38] proposed a modified U-Net architecture for weed segmentation in complex field environments, incorporating a data augmentation method based on foreground and background combination. Roy et al. [39] propose En-UNet, an enhanced UNet model for real-time semantic segmentation of rotten and fresh apples using RGB images. Tassis et al. [40] proposed an integrated deep learning framework combining Mask R-CNN for instance segmentation, UNet and PSPNet for semantic segmentation, and ResNet for classification to detect pests and diseases in coffee leaves from in-field smartphone images. Osco et al. [41] compared FCN, U-Net, SegNet, DDCN, and DeepLabV3+ for the semantic segmentation of citrus orchards using UAV-based multispectral imagery. Mishra et al. [42] have proposed the use of deep convolutional neural networks (CNNs), specifically Inception V4, to estimate weed density in soybean crops by performing robust vegetation segmentation. The method effectively distinguishes weeds from crops and shows high potential for use in smart agriculture. Zhang et al. [43] proposed PSPNet for PolSAR image semantic segmentation in agricultural areas, integrating multiscale feature fusion, a polarimetric channel attention module, and an edge-aware loss to enhance boundary accuracy and spatial consistency. Sun et al. [44] proposed a novel semantic segmentation network, SADNet, to segment UAV-based orchard images, incorporating modules like depthwise identity convolution (DIC), ASPP, and scSE to improve feature extraction, attention, and receptive field. Wirawan et al. [45] applied the UNet model for semantic segmentation of rice field bunds from RGB images. They optimized the model's performance by using a dataset configuration without binarization, achieving a high accuracy. He et al. [46] proposed a two-phase method for detecting farmland boundary lines: MobileV2-UNet for accurate farmland area segmentation and a multi-boundary detection method using frame correlation and RANSAC for boundary line extraction. Raei et al. [47] proposed a U-Net architecture with a ResNet-34 backbone for semantic segmentation of irrigation systems using very high-resolution remote sensing imagery. Niu et al. [48] proposed HSI-TransUNet, a transformer-based semantic segmentation model for accurate crop mapping from UAV hyperspectral imagery. The model incorporates four key improvements: a spectral-feature attention module, residual connections in Transformer layers, sub-pixel convolutions in the decoder to avoid the chessboard effect, and a hybrid loss function for boundary refinement. Jin et al. [49] proposed an improved Mask R-CNN model integrating CBAM attention and depthwise separable convolutions for accurate weed segmentation in complex field environments. Peng et al. [50] developed an improved DeepLabV3+ model with a novel ResDense backbone—integrating ResNet and DenseNet—and employed focal loss to enhance the segmentation accuracy of litchi branches,outperforming baseline models across varying image complexities and demonstrating superior robustness in complex orchard conditions. Zhu et al. [51] introduced LD-DeepLabv3+, a two-stage DeepLabv3+ framework with adaptive loss for apple leaf disease segmentation under complex scenes. By combining reverse attention, RFB modules, and channel attention with adaptive loss. Baravdish et al. [52] explore two techniques, marginal loss, and background masking, to perform semantic segmentation with partially annotated data, addressing the challenge of manually labeling each pixel. Cao et al. [53] propose a method for accurate instance segmentation of small farmland using a Mask R-CNN model integrated with a double attention mechanism feature pyramid network (DAFPN). The DAFPN consists of two attention modules: spatial attention (SPA) and channel attention (CHA), which enhance feature extraction by focusing on spatial relationships and adaptive channel merging. The model was tested on very high-resolution (VHR) satellite images, showing a significant performance improvement over the standard Mask R-CNN. Cai et al. [54] propose an attention-aided semantic segmentation approach for weed identification in pineapple fields using UAV imagery. Building on PSPNet, three attention modules, SE, ECA, and CBAM, are evaluated, with the ECA module integrated into the SPP layer yielding the best performance. Feng et al. [55] introduced a new pooling method, ”double cross pooling” (DCPM), combining twill and strip pooling to improve the accuracy of apple leaf lesion segmentation. DCPM significantly improved performance when added to existing segmentation networks like DeepLabV3+, PSPNet, and U-Net. Sun et al. [56] proposed RL-DeepLabv3+, a lightweight semantic segmentation model for rice lodging detection in unmanned rice harvesters. Singh et al. [57] proposed a modified version of U-Net, DeepUNet, for semantic segmentation of satellite images, focusing on pre-processing using FAAGKFCM and SLIC superpixel techniques. Yu et al. [58] proposed ASE-UNet, an improved U-Net-based semantic segmentation model for accurate orange fruit segmentation in complex agricultural environments. ASE-UNet enhances the backbone architecture to preserve spatial details, integrates a Shape Feature Extraction Module (SFEM) for better distinction between fruits and background, and utilizes an attention mechanism to refine feature fusion. Wang et al. [59] proposed MDE-UNet, a hybrid model combining Multitask Deformable UNet and an enhanced lightweight UNet with residual attention, for accurate farmland boundary segmentation from GF-2 remote sensing images. Zhang et al. [60] proposed a modified U-Net (MU-Net) for plant-diseased leaf image segmentation by introducing residual blocks (Resblock) and residual paths (Respath) to enhance feature transformation, increase network depth, and improve accuracy and efficiency. Diao et al. [61] proposed an improved UNet-based algorithm, ASPP-UNet, for recognizing maize crop row centerlines in complex farmland environments. This method enhanced segmentation accuracy by replacing UNet's traditional convolution structure with multi-scale expansive convolution (ASPP) and integrated an improved vertical projection method with the least squares method for precise centerline fitting, achieving better real-time performance and accuracy in comparison to traditional methods. Gupta et al. [62] proposed a U-Net-based semantic segmentation approach using Inception-ResNetV2 for multiclass weed identification in brinjal fields. Zhang et al. [63] proposed an improved Mask R-CNN model with feature fusion and hybrid attention mechanisms. Guo et al. [64] proposed InstaCropNet, a dual-branch U-Net-based architecture for precise maize crop row detection. Liu et al. [65] proposed a high-throughput rice seedling measurement method based on an improved UNet model enhanced with Coordinate Attention (CA) and Vision Transformer (ViT) blocks for more accurate segmentation and trait evaluation. Purohit et al. [66] present a ResNet-UNet model for segmenting crops and weeds in UAV images. Zhang et al. [67] proposed Fast-UNet, an optimized UNet model with reduced convolutional kernels and an integrated ASPP module for efficient, real-time navigation path recognition between fruit tree rows. It enhances accuracy and speed while enabling generalization across different fruit datasets using transfer learning. Bhatti et al. [68] proposed using a CNN-based U-Net architecture to precisely segment forest images from satellite data, emphasizing the impact of training data size on segmentation accuracy. Yang et al. [69] proposed FRPNet, a lightweight real-time segmentation model using FasterNet with residuals, PASPP for multi-scale feature extraction, and a novel Ohd-Loss to improve segmentation in unstructured field environments. Shen et al. [70] proposed weed and disease detection in crops using improved PSPNet architectures, attention modules, and semi-supervised segmentation and density estimation strategies. Islam et al. [71] proposed an improved Mask R-CNN with ResNet-101 and CB-Net was used to detect and segment lettuce seedlings from tray images, showing a strong correlation between manual and model-estimated leaf areas, enabling accurate seedling size estimation and supporting automated seedling sorting. Deka et al. [72] introduce a modified Mask-RCNN model with SE (Squeeze-and-Excitation) channel attention for litchi fruit segmentation using UAV imagery. The proposed model improves detection accuracy by addressing challenges such as occlusion and uneven color in complex environments. Cai et al. [73] introduce an improved DeepLabV3+ semantic segmentation model for accurately extracting cultivated land parcels from high-resolution UAV imagery. By incorporating MobileNetV2 as the backbone network, along with the CBAM attention mechanism and DFF dynamic feature fusion module, the model enhances segmentation accuracy, offering precise extraction of irregularly shaped cropland parcels, making it a valuable tool for modern agricultural management and land planning. Wu et al. [74] combine terahertz imaging with semantic segmentation models (SegNet and DeepLab V3+) to assess thin-shelled watermelon seeds, achieving precise segmentation of seed coats and kernels. Wang et al. [75] propose a lightweight segmentation model, WE-DeepLabV3+, for P. notoginseng leaf disease detection, integrating MobileNetV2 as the backbone and introducing the Window Attention-ASPP and Efficient Channel Attention modules. Experimental results demonstrate that WE-DeepLabV3+ outperforms other models in segmentation accuracy while significantly reducing the model's parameters, making it suitable for mobile deployment. DeepLab V3+ outperforms SegNet, offering superior speed and accuracy, making it highly effective for non-destructive phenotypic analysis of seeds. Lu et al. [76] utilized DeepLabV3+ with a ResNet backbone for accurate segmentation of corn, tobacco, and barley from remote sensing images. Using XGBRegressor, the method predicted corn yield with a low error, highlighting its potential for yield forecasting and agricultural monitoring. Habib et al. [77] have proposed a novel segmentation model for weed detection in agricultural images, which leverages advances in architecture design to provide accurate and fast real-time performance. The model outperforms existing approaches by demonstrating strong generalization capabilities and efficiency, making it a valuable tool for precision agriculture and sustainable weed management. Peng et al. [78] have proposed a DRFG framework that fuses 3D-CNN and GATs with fuzzy C-means clustering and PCA for enhanced spatial-spectral feature extraction. This approach improves segmentation performance in hyperspectral data by leveraging deep learning and graph-based methods for better accuracy under limited labeled data. Miao et al. [79] proposed VGG16-UNet for the semantic segmentation of foxtail millet seed CT images. Wang et al. [80] presented an enhanced semantic segmentation model, m-DeepLabV3+, for rapid and accurate detection of straw return in conservation tillage. By introducing MixConv, channel shuffling, a concentration-based attention module (CBAM), and squeeze and extraction networks (SE), the model improves segmentation performance, reduces parameter count, and mitigates overfitting through R-drop. Xu et al. [81] present a two-stage lightweight segmentation model based on DeepLab V3+ for detecting cotton verticillium wilt, utilizing advanced feature fusion, attention mechanisms, and lightweight strategies like channel pruning. The model achieves superior segmentation accuracy and is optimized for mobile deployment with minimal parameters and fast inference. Yao et al. [82] proposed a real-time crop lodging recognition system using a modified DeepLabV3+ with Xception for training and MobileNetV2 for deployment, optimized for use on Jetson Nano. Their integrated approach with ROS and RealSense depth-sensing offers high accuracy with a low computational cost for harvester deployment.

Chapter 3

Proposed Methodology

In this chapter, we discuss the adopted methodology with detailed preprocessing steps including image resizing and augmentation techniques like rotation, flipping, cropping, and scaling. It further describes the architectures that are UNet, UNetResNet50, UNetVGG16, and the proposed VResUNet++ architecture.

3.1 Data Preprocessing

We consider the original UAV RGB images. which are ortho-mosaic images. The size of the images ranges from [4000-6000]×[4000-6000] that are divided into multiple patches. The size of each patch is 256 x 256. The patches are augmented while flipping and rotation are considered.

3.2 Model Architecture

3.2.1 Base Architecture

In this work, we use UNet [83] as the base architecture, as shown in Fig. 3.2. It is a convolutional neural network designed specifically for the semantic segmentation

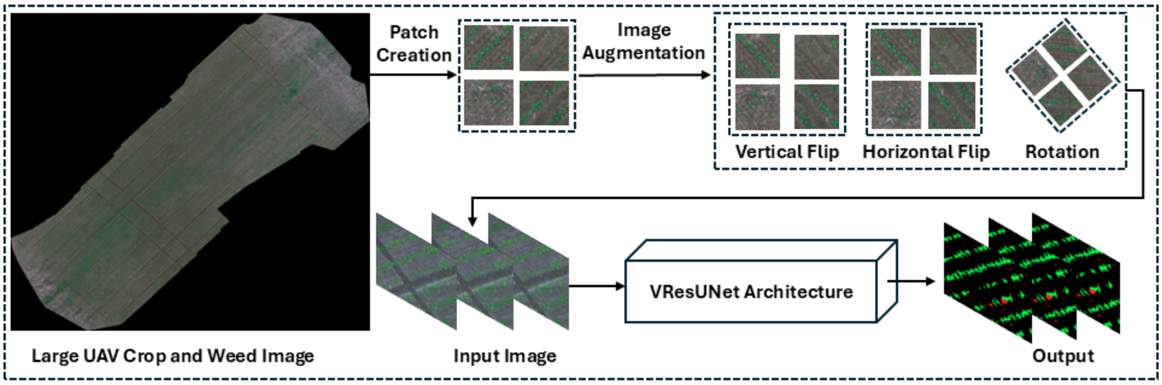


Figure 3.1: Block Diagram

task. Due to its symmetric encoder-decoder design with a skip connection that helps in pixel-wise classification, we consider the architecture as our base architecture.

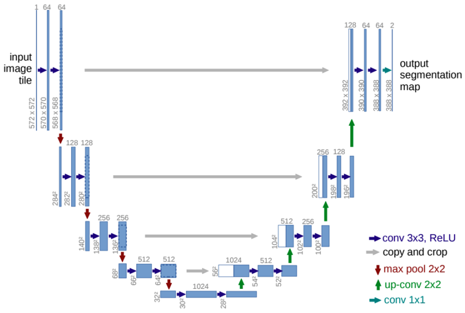


Figure 3.2: U-Net Architecture

The U-Net architecture consists of two blocks: the encoder and decoder blocks. The encoder block is called the contracting path, and the decoder block is called the expansive path. The encoder block follows the structure of a typical convolutional network. It consists of multiple blocks, each of which has two consecutive 3 x 3 convolutional layers with a ReLU activation function with the same padding, followed by a 2 x 2 max-pooling operation with a stride of 2. The pooling operation is responsible for reducing the spatial resolution of the feature map by half while increasing the number of feature channels. The network learns on high-level semantics as it goes deeper and captures more features. The connecting part of the encoder block and decoder block is called as bottleneck. It has 3 x 3 convolutional layers followed by an

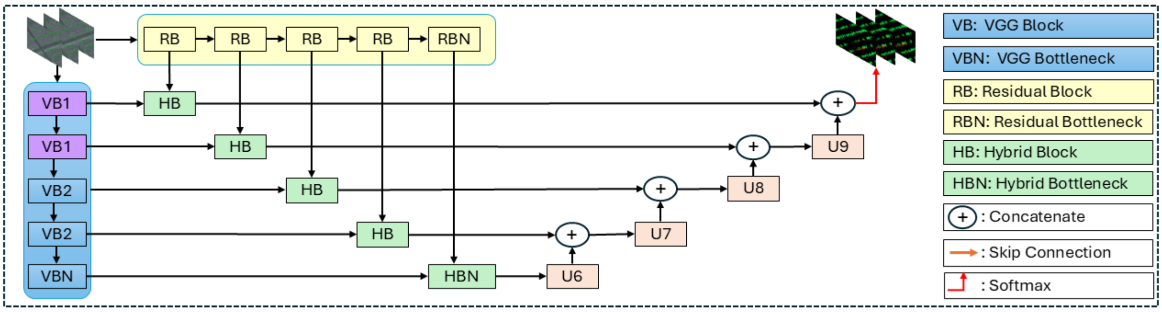


Figure 3.3: VResUNet++ Architecture

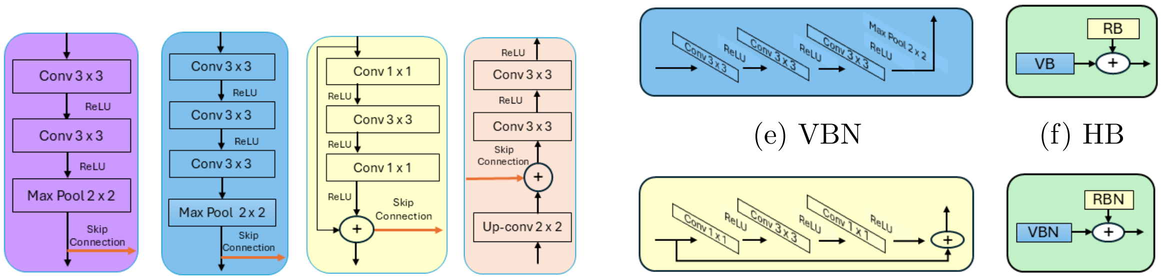
upsampling operation. It is the most compressed representation of the input, and it plays an important role in capturing the features of the entire image.

The decoder block restores the original spatial resolution. It has an upsampling operation that doubles the spatial dimension of the feature maps. After the upsampling, it gets concatenated with the corresponding feature map from the encoder block, which passes via skip connection, and then it is followed by two 3 x 3 convolutional layers and a ReLU activation function. The skip connection is responsible for recovering the spatial information that might be lost during downsampling and ensures that fine-grained details are preserved.

The final layer contains the 1 x 1 convolution that maps the feature vector at each pixel location to the desired number of classes, followed by a softmax (multi-class) or sigmoid (binary class) activation function for segmentation. The training of the network is done by pixel-wise loss functions such as categorical cross-entropy or dice loss. The advantage of U-Net is that it uses the data and memory efficiently via symmetric skip connections and a compact architecture. That makes it suitable for agricultural image segmentation, where precise boundary detection is essential.

3.2.2 Proposed VResUNet++ Architecture

We present a mixed encoder-decoder structure, as shown in Fig. 3.3, for semantic segmentation jobs in precision agriculture. The model combines the benefits of VGG16



(a) VB1 (b) VB2 (c) RB (d) Decoder (e) VBN (f) HB (g) RBN (h) HBN

Figure 3.4: Combined figure: where (a) is VGG Block_1, (b) is VGG Block_2,(c) is Residual Block, (d) is Decoder Block, (e) is VGG Bottleneck Block, (f) is Hybrid Block, (g) is Residual Bottleneck, and (h) is Hybrid Bottleneck

and ResNet50 into one plan that improves feature understanding at many spatial levels. The goal is to find and segment regions of interest, such as crops and weeds, from pictures taken using drones and other aerial means. The model takes an RGB input image of 256 x 256 x 3. This image is processed using the VGG16 and the ResNet50 models at the same time. Those models are trained using the ImageNet dataset, and the top fully connected layers are removed. The convolutional layers of VGG16 keep detailed spatial features. Mathematically, the output feature map y [84] after applying convolution, bias, and the ReLU activation function is as follows:

$$\mathbf{y} = \sigma(W * \mathbf{x} + b) \tag{3.1}$$

where x is the input image or feature map, W is the learnable convolutional filter/kernel weights, b denotes the bias term added to each output channel after convolution and σ represents the ReLU activation function.

Residual blocks in ResNet50 help the model to learn more features by using the skip connection that carries the information without change, which eases the network to learn and understand more complex features, as shown in Fig. 3.5.

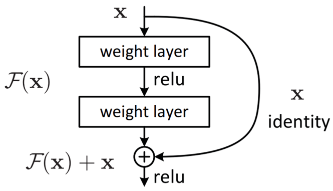

x
weight layer
relu
weight layer
$\mathcal{F}(\mathbf{x})$
x
identity
$\mathcal{F}(\mathbf{x}) + \mathbf{x}$
relu

Figure 3.5: Skip Connection in ResNet Architecture [3]

In residual learning [3], the output of the residual block y is evaluated as —

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, W_i) + \mathbf{x} \tag{3.2}$$

where x represents the input to the residual block. It is the output from the previous layer. W_i is the weights (parameters) of the convolutional layers within the residual block. These are trainable during backpropagation. $F(x, W_i)$ denotes the residual mapping function.

At the encoder layer, VGG16 and ResNet50 are utilized to obtain five feature maps from each of their layers. The features of VGG16 are selected from block1_conv2, block2_conv2, block3_conv3, block4_conv3, and block5_conv3 layers, with spatial resolutions of 256 × 256, 128 × 128, 64 × 64, 32 × 32, and 16 × 16, respectively. The first two blocks of VGG contain two convolutions of 3 x 3, followed by ReLU, and max-pooling at the end, as shown in Fig. 3.4, while the next two blocks of VGG contain three convolutions of 3 x 3 followed by ReLU and max-pooling at the end, as shown in Fig. 3.4. The max-pooling is responsible for reducing the spatial dimension. The main reason for this configuration is that early layers operate on a large feature map, while later layers have a small feature map. Hence, adding more convolutions at a later layer reduces the computation cost. Additionally, the initial blocks extract low-level features like edges and textures, whereas the later blocks extract higher-level features like shapes and objects having complex patterns. Hence, we argue that we require more layers.

In the case of ResNet50, the feature layers chosen are conv1_relu, conv2_block3_out, conv3_block4_out, conv4_block6_out, and conv5_block3_out. The residual block consists of one 1 x 1 convolution and ReLU, followed by one 3 x 3 convolution and ReLU, followed by one 1 x 1 convolution and ReLU. In each residual block, the input is added to the output of the convolutional layers, allowing the network to learn the residual function. This added skip connection preserves the important features and supports the gradient flow effectively through deep layers, which makes training more effective and stable, as depicted in Equation (3.2).

In the final part, we use the combination of VGG and ResNet. The final part of the VGG block is called the VGG bottleneck, i.e., the last encoder part of the VGG block. It consists of three 3 x 3 convolutions and ReLU, followed by max-pooling with a stride of 2. Similarly, the final part of the residual block is a residual bottleneck. It consists of one 1 x 1 convolution and ReLU, followed by one 3 x 3 convolution and ReLU, followed by one 1 x 1 convolution and ReLU, and the final output is concatenated with the initial input, as shown in Fig. 3.4.

The feature maps of VGG16 and ResNet50 at each level are concatenated along the channel axis, as shown in Equation 3.3. The first concatenation of the VGG block and the residual block is called a hybrid block, as shown in Fig. 3.4, and the concatenation of the VGG bottleneck and the residual bottleneck is called a hybrid bottleneck, as shown in Fig. 3.4. This is accomplished by forming hybrid feature representations that enhance the encoding capability of the encoder in representing low-level texture features and high-level semantic features thoroughly. The resulting feature maps are referred to as hybrid blocks. The feature concatenation H_l for the VResUNet++ hybrid block is as follows:

$$H_l = \text{Concat}\left(F_l^{\text{VGG}},\ F_l^{\text{ResNet}}\right) \tag{3.3}$$

where F_l^{VGG} is the output of the VGG block, F_l^{ResNet} is the output of the ResNet block, and Concat is the concatenation of the output of the VGG and ResNet blocks. The hybrid blocks are now employed as an input to the decoder path constructed based on the U-Net structure, as shown below.

$$H_l' = \sigma\left(W_H * H_l + b_H\right) \tag{3.4}$$

where H_l is hybrid feature map before convolution, W_H is learnable weights, b_H is bias, and H_l' hybrid map after convolution.

The decoder path comprises a series of Conv2DTranspose layers which gradually upsampling feature maps to the original 256×256 resolution, as shown in Equation (3.5).

$$Y = W^T * X \tag{3.5}$$

where X is the input feature map, W^T is the transposed convolution filter, $*$ is the convolution operation, and Y is the output feature map.

Skip connections are used for each upsampling step to merge the decoder output with the hybrid encoder block having the same depth, preserving spatial information and context as shown in Fig. 3.4d. It is followed by two subsequent 3×3 convolutional layers with ReLU activation, L2 regularization, and dropout for enhanced generalization and lesser overfitting.

The final output layer is a 1×1 convolution with softmax activation [85] and provides a segmentation mask of the shape 256×256×C, with C being the desired number of classes. In this work, we have 3 classes: crop, weed, and background. Here, we have —

$$\hat{y}_{i,j,c} = \frac{e^{z_{i,j,c}}}{\sum\limits_{k=1}^{C} e^{z_{i,j,k}}} \tag{3.6}$$

where $\hat{y}_{i,j,c}$ is the softmax probability of class c at pixel location (i,j), $z_{i,j,c}$ is logit(raw score) for class c at pixel (i, j), C is the total number of classes, $\sum_{k=1}^{C} e^{z_{i,j,k}}$ is the normalization factor, that is the sum of the exponential of all class logits at pixel (i, j). Training the model is possible using a combination of categorical cross-entropy loss with optional Dice or (Intersection over Union) IoU-based losses, and evaluation metrics for the model include accuracy, precision, recall, Dice coefficient, and IoU.

This hybrid approach, suggested here, efficiently utilizes the strengths of the complementarity of VGG16 and ResNet50 backbones and enhances the accuracy of segmentation by preserving spatial detail and semantic context. The architecture of the model is perfectly adaptable for use in agricultural image analysis, where precision boundary detection and representation of multi-scales of features are a requirement.

Chapter 4

Experimental Results

This chapter outlines the performance of the proposed models. It includes a thorough description of the dataset used and elaborates on the evaluation parameters, namely precision, recall, accuracy, dice coefficient, and intersection of union. It also discusses the training and validation outcomes over multiple epochs for all models, supported by relevant tables.

4.1 Dataset

- **Weedmap dataset** [5]: The Weedmap dataset includes large-scale semantic weed mapping images in precision agriculture using aerial multispectral imaging. The images are captured from sugar beet fields in Eschikon, Switzerland, and Rheinbach, Germany, over five months, using commercial UAV platforms equipped with MicaSense RedEdge-M and Sequoia multispectral cameras.

  Thereafter, the information is grouped into two broad categories — Orthomosaic and Tiled images. The orthomosaic information comprises *eight* composite images of vast regions with pixel-level ground truth, and the tiled information comprises 18,746 small-sized images (tiles) clipped from orthomosaics in a sliding window fashion along with their respective segmentation masks. Every 480 × 360 pixel tile enables patch-level high-resolution training and emulates real-world field complexities such as crop density, illumination variability, and weeds.

All images within the dataset possess multispectral data on 12 channels (RedEdge-M) and 8 channels (Sequoia), including RGB, CIR, and NDVI views. Color-coded and indexed ground truth annotations are given for crop, weed, and background at pixel resolution.

The data set is approximately 5.36 GB, has an area of fields of 16,554 m^2, and comprises more than 1.76 billion pixels, with 1.39 billion for the training set and 367 million for the test set, a huge figure by any measure compared to other available open-source multispectral annotated datasets to identify agricultural weeds. The acquired Ground Sampling Distance (GSD) is approximately 1 cm, which offers enough resolution to identify fine-grained crop and weed details, such as crops: 15–20 pixels and weeds: 5–10 pixels.

The image annotations use a standard naming convention with numbered samples (000–007) and metadata in subdirectories for orthomosaics, reflectance, ground truth masks, and binary masks. Color ground truth displays class labels as — background (black), crop (green), and weed (red); accordingly, the index maps specify these as — background = 0, crop = 1, and weed = 2.

This large and heterogeneous dataset facilitates effective training and testing of semantic segmentation models in actual agricultural conditions and is specifically useful for further research in precision agriculture, crop-weed separation, and image processing with UAVs.

- **GobhiSet** [4]: The GobhiSet dataset is an open RGB high-resolution aerial image dataset with secondary annotation of Brassica oleracea var. botrytis (cauliflower) crops throughout their phenological development cycle. Photograph acquisition was done through the use of a DJI Phantom 4 Pro Obsidian unmanned aerial system (UAS) platform capturing photographs along a linear grid flight path on six days over an experimental farm in the vicinity of Portici, Italy. All flights were carried out within solar noon times (12:00-12:25), from an altitude of between 4.275 m and 4.749 m, considering wind turbulence. Data consists of 244 raw original RGB images at a spatial resolution of 5472 × 3648 pixels for high spatial resolution for big-scale crop examination. October 8, 21, and 29 images were manually labeled using the VGG Image Annotator (VIA) v1.0.6.

Bounded boxes for crop structures are tagged and kept in the COCO segmentation description within JSON files that were marked as region- and shape-describing. All three tagged sets were utilized in training a pipeline of automated tagging that consisted of Grounding DINO as well as the Segment Anything Model (SAM). The automated annotations generated Pascal VOC-style binary segmentation masks and saved them as PNG images of full aerial views, individual crop rows, and orthomosaic reconstructions. Seven classes (Row 1 to Row 7) are defined by the dataset, which corresponds to separate crop rows.

A single crop ID is also annotated on every object in a row, and crop repetition and location data are provided in CSV format. Benchmarking was conducted utilizing 21 October annotations for testing annotation quality. The orthomosaics computed by date from raw images allow for longitudinal plant growth observation. The data set is capable of supporting a wide range of applications, from semantic segmentation, phenological monitoring, object detection, and growth modeling. Python code that is utilized to process the annotations, create masks, and organize the dataset is made available so that it can be replicated and incorporated into deep learning pipelines. Overall, GobhiSet is a better resource to support automatic crop inspection in precision agriculture and computer vision.

4.2 Evaluation Metrics

We evaluated the performance of the proposed scheme considering the following parameters.

- **Accuracy**: It is the measure of the proportion of correct predictions out of all predictions made.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.1}$$

  where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative, respectively.

- **Precision**: It is the ratio of true positives to all predicted positives. It measures how many predicted positives are correct.

$$Precision = \frac{TP}{TP + FP} \tag{4.2}$$

- **Recall (Sensitivity)**: It is the ratio of true positives (TP) to all actual positives (TP + FN), measuring how many real positive cases the model correctly identifies.

$$Recall = \frac{TP}{TP + FN} \tag{4.3}$$

- **Dice Coefficient**: The Dice Coefficient measures the overlap between predicted and true positive regions, calculated as twice the true positives (2TP) divided by the sum of twice the true positives plus false positives and false negatives. It ranges from 0 (no overlap) to 1 (perfect overlap).

$$Dice = \frac{2TP}{2TP + FP + FN} \tag{4.4}$$

- **Intersection over Union (IoU)**: IoU measures the overlap between the predicted and true positive regions, calculated as the ratio of true positives (TP) to the union of predicted and actual positives (TP + FP + FN). It quantifies how well the predicted area matches the ground truth.

$$IoU = \frac{TP}{TP + FP + FN} \tag{4.5}$$

4.3 Experimental Setup

We conducted our study on Ubuntu 22.04.1 LTS. The equipment used in the experiment is listed in Table 4.1, consisting of NVIDIA Corporation GA102 [GeForce RTX 3090] as a GPU with 128.0 GiB memory. The hardware model is Dell Inc. Precision 7920 Tower with processor Intel Xeon(R) Silver 4214 CPU @ 2.20GHz × 48 that has a disk capacity of 8.5 TB. The environment is TensorFlow for the deep learning framework. In the experiment, both training and testing are conducted in this environment using the same equipment.

Table 4.1: Experiment environment setup

Configuration	**Type**
Hardware Model	Dell Inc. Precision 7920 Tower
Operating System	Ubuntu 22.04.1 LTS
CPU	Intel® Xeon(R) Silver 4214 CPU @ 2.20GHz × 48
GPU	NVIDIA Corporation GA102 [GeForce RTX 3090]
Memory	128.0 GiB

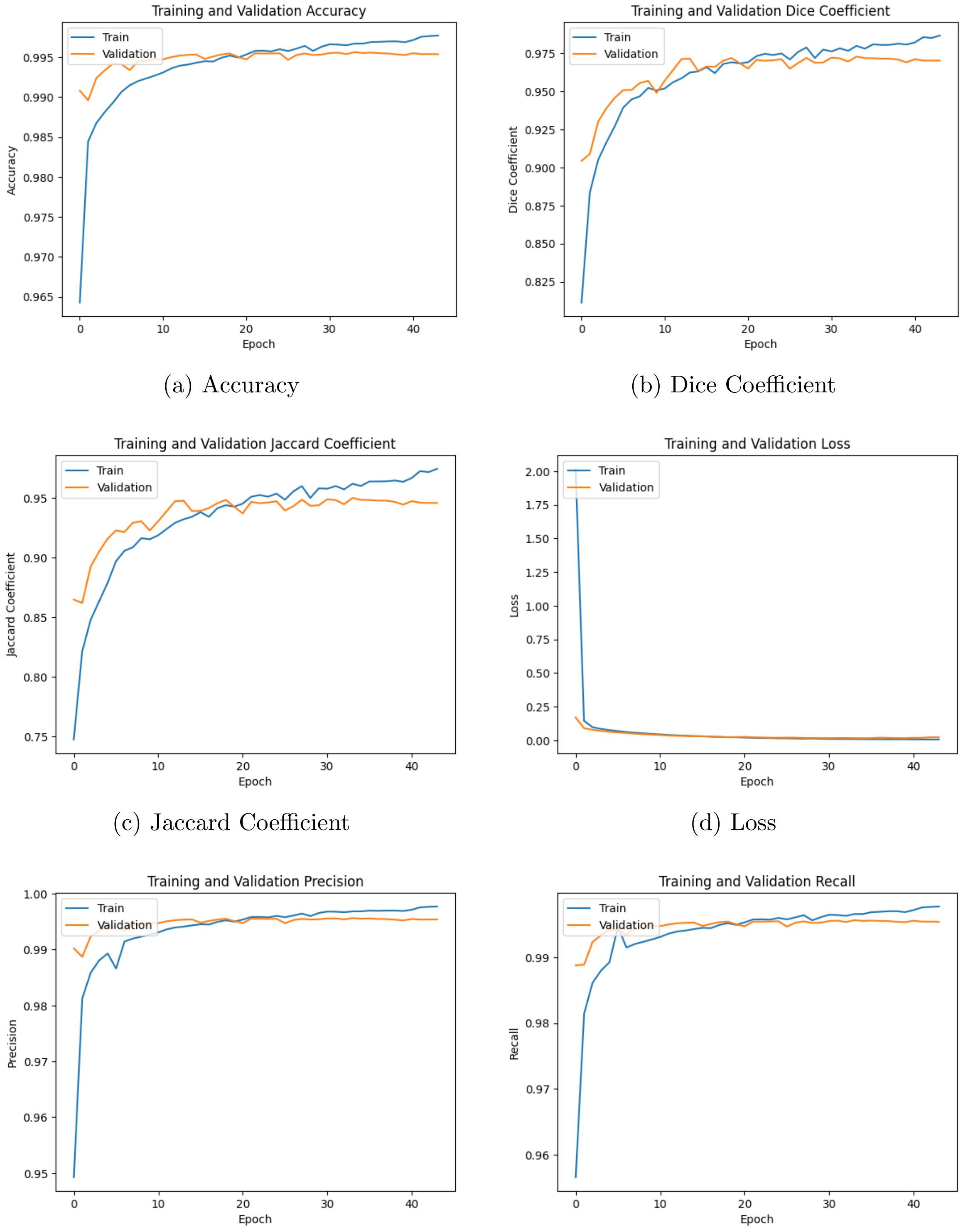
4.4 Model Training

In the model training part, we considered a VResUNet++, a hybrid U-Net model, which is constructed by integrating the VGG16 and ResNet50 at the encoder part of the UNet, which is trained on the ImageNet to leverage their strength in spatial feature retention in the case of VGG16 and deep semantic abstraction in the case of ResNet50. For training, we are using the UAV-acquired aerial crop images, which have been resized to 256×256×3. Both layers were responsible for the intermediate feature map extraction via concatenation. This fusion formed hybrid skip connections at each stage of the encoder. The decoder followed a U-Net-inspired design transpose convolution for upsampling, followed by two Conv2D layers per block, each regularized with L2 ($\lambda = 1e - 4$) regularization and Dropout (rate = 0.1) to reduce overfitting on limited agriculture datasets.

We train with the Adam optimizer with an initial learning rate of $1e - 4$, and categorical cross-entropy is used for the loss function due to the multi-class segmentation setting. The batch size is fixed at 16, and early stopping is applied based on validation loss to avoid overfitting. The performance is evaluated using metrics, such as accuracy, precision, recall, dice coefficient, and IoU, which together measure both pixel-level agreement and classwise region overlap. The strategy improves the segmentation performance across the heterogeneous crop and weed regions in the UAV imagery.

Training and Validation Results

The performance of the proposed VResUNet++ architecture has been shown in Figure 4.1. It illustrates the model's convergence and generalization behavior. It shows robust performance across all metrics. The smooth convergence shows that the model is stable for precise image segmentation tasks in precision agriculture. The smooth converging curves show the capability of the model, especially for the complex agriculture domain, which involves the overlapping of crop, weed, and background pixels. The close alignment between training and validation values confirms minimal overfitting and good generalization.



(a) Accuracy

(b) Dice Coefficient

(c) Jaccard Coefficient

(d) Loss

(e) Precision

(f) Recall

Figure 4.1: Training and validation plots for the VResUNet++ architecture

4.5 Benchmarks

We compared the proposed VResUNet++ model with three widely used encoder-decoder architectures — UNet, UNet with ResNet50, and UNet with VGG — in semantic segmentation, specifically, in the precision agriculture domain. Each of the benchmark models is based on the UNet architecture with varying backbones.

4.5.1 UNet

The UNet architecture [86] follows the encoder-decoder structure. The encoder is responsible for the downsampling of the original image and captures the features, while the decoder is responsible for the restoration of the resolution by upsampling. The skip connection between the encoder and decoder is responsible for preserving the fine spatial information that helps the network to segment the small objects and boundaries. Due to the simplicity of the UNet model, it is widely adopted in the precision agriculture sector for segmentation. However, the standard UNet lacks sufficient depth in the encoder, which hinders semantic feature extraction.

4.5.2 UNet with ResNet50

The researchers have explored various enhancements in the UNet for upgrading its capability. Diakogiannis *et al.* [87] modified the encoder part of the UNet with the ResNet50 backbone, which enables the training of a deeper network by eliminating the vanishing gradient problem. The modification of the encoder part of the UNet with ResNet50 significantly improves the ability to extract the high-level semantic feature, especially in large UAV images. However, the aggressive downsampling of ResNet reduces the spatial resolution, resulting in blurred object boundary detection, which is a critical issue in crop and weed segmentation.

4.5.3 UNet with VGG

Ullah *et al.* [88] compared varIoUs segmentation architectures having different backbone networks, e.g., UNet with a VGG16 backbone. In the comparison, he upgraded the encoder part of the UNet with VGG16 as a backbone to enhance its feature extraction capabilities. This modification takes advantage of the hierarchical feature learning capability of VGG at the encoder block, which results in deeper semantic feature extraction while preserving the spatial information, which enhances the segmentation performance.

4.6 Ablation Study

Initially, we started with the original U-Net model [83], which is a groundbreaking architecture for image segmentation in the medical as well as agricultural domains. However, this architecture has some shortcomings, as mentioned below.

1. This model with agricultural UAV images, which are large and complex, results in poor generalization.

2. Using the scratch-trained encoder results in limited feature extraction ability.

3. Another major problem is the vanishing gradient problem in the deep layers, in which the gradient becomes extremely small as they are propagated backward through the network during backpropagation. This results in slow or negligible learning in the initial layers.

4. Finally, there is a difference in training and validation due to the overfitting of the model.

- **UNetResNet50 Architecture** We are updating the backbone of the U-Net architecture with a ResNet50 encoder backbone, which is inspired by the UNetResNet50 architecture [87]. This architecture has semantic feature extraction with precise spatial reconstruction. The encoder is based on the pre-trained ResNet50 model, which is trained on ImageNet and replaces the standard UNet encoder layer. This update in the backbone allows the model to leverage the hierarchical feature representation, which it learned from large-scale data. The residual block of the encoder helps in eliminating the vanishing gradient problem and allows deeper network training. However, the aggressive downsampling in ResNet caused the loss of spatial resolution, which affects the precise boundary detection.

- **UNetVGG16 Architecture** Another update that we have made is updating the encoder layer of the UNet architecture with the VGG16 model, pre-trained on the ImageNet dataset, which is inspired by the UNetVGG16 architecture [88]. Unlike residual networks (like ResNet), VGG16 uses a stack of 3x3 convolutional layers without skip connections, which preserves the spatial locality and is better at boundary segmentation and small object detection in crop fields. However, there is a lack of residual depth that makes shallow semantic representation. Also, in the heterogeneous crop regions, there is less context awareness.

4.7 Results and Discussions

The proposed VResUNet++ has been trained and tested on the Gobhiset and Weedmap datasets. Key metrics include precision, recall, accuracy, dice coefficient, and intersection over union (IoU). The model shows phenomenal results.

Table 4.2 describes the overall performance of VResUNet++ on the Gobhiset and Weedmap datasets. The VResUNet++ model has achieved precision 99.74%, recall 99.74%, accuracy 99.74%, dice coefficient 99.66%, and IoU 95.54% on Gobhiset dataset. Similarly, on Weedmap dataset VResUNet++ has achieved precison 98.83%, recall 98.65%, accuracy 98.69%, dice coefficient 72.93% and IoU 65.00%.

Table 4.2: Performance of VResUNet++ on Gobhiset and Weedmap dataset

	Dataset	
Metrics	**Gobhiset**	**Weedmap**
Precision	99.74%	98.83%
Recall	99.74%	98.65%
Accuracy	99.74%	98.69%
Dice Coefficient	99.66%	72.93%
IoU	95.54%	65.00%

Table 4.3: Class-wise Performance of VResUNet++ on Gobhiset and Weedmap dataset

Dataset	**Class**	**IoU per Class**	**Dice Coefficient per Class**
Gobhiset	Crop	89.38%	94.39%
	Background	99.66%	99.83%
Weedmap	Crop	74.57%	85.83%
	Weed	36.29%	52.26%
	Background	98.57%	99.28%

Table 4.3 describes the class-wise performance of the VResUNet++ model on the Gobhiset and weedmap datasets. VResUNet++ has achieved IoU for crop 89.39%, background 99.66%, dice efficiency for crop 94.39% and for background 99.83% on the Gobhiset dataset. Similarly, in the case of weedmap dataset we have multiclass such as crop, weed, and background. The VResUNet++ has achieved IoU for crop 74.57%, weed 36.29% and background 98.57%, similarly dice coefficient for crop 85.83%, weed 52.26% and background 99.28%. These results show that our model has achieved exceptional results on different datasets. which demonstrates the robust performance in the precision agriculture environment. Additionally, we have achieved great class-wise performance which help us in differentiating between crop, weed and background. Such performance is crucial for real-time crop segmentation in precision agriculture.

Table 4.4: Comparison of overall metrics across different models for Gobhiset dataset [4]

Model	**Test Loss**	**Accuracy**	**Precision**	**Recall**	**Dice Coefficient**	**IoU**
UNet [83]	02.85%	99.21%	99.21%	99.21%	91.55%	85.81%
UNetResNet50 [87]	02.08%	99.66%	99.66%	99.66%	96.78%	93.95%
UNetXception [89]	05.33%	99.46%	99.47%	99.45%	94.10%	89.42%
UNetVGG16 [88]	02.67%	99.71%	99.72%	99.71%	94.36%	94.98%
VResUNet++ (ResNet50 + VGG16)	**01.20%**	99.66%	99.66%	99.66%	96.88%	94.12%
VResUNet++ (ResNet50 + VGG16) (Dropout, L2 regularization and Finetune)	01.48%	**99.74%**	**99.74%**	**99.74%**	**97.66%**	**95.54%**

Table 4.5: IoU and Dice Coefficient per class across different models for Gobhiset dataset [4]

	IoU Per Class		**Dice Coefficient**	
Metrics	**Background**	**Crop**	**Background**	**Crop**
UNet [83]	99.03%	72.87%	99.51%	84.31%
UNetResNet50 [87]	99.59%	87.38%	99.79%	93.26%
UNetXception [89]	99.38%	81.44%	99.69%	89.77%
UNetVGG16 [88]	99.64%	88.94%	99.82%	94.14%
VResUNet++ (ResNet50 + VGG16)	99.58%	87.30%	99.79%	93.22%
VResUNet++ (ResNet50 + VGG16) (Dropout, L2 Regularization, and Finetune)	**99.66%**	**89.38%**	**99.83%**	**94.39%**

Table 4.6: Comparison of overall metrics across different Weedmaps for the Weedmaps dataset [5]

Overall Metrics						
Model	**Test Loss**	**Accuracy**	**Precision**	**Recall**	**Dice Coefficient**	**IoU**
UNet [83]	05.67%	96.76%	96.90%	96.66%	68.34%	60.13%
UNetResNet50 [87]	05.77%	97.76%	97.90%	97.66%	69.34%	61.13%
UNetXception [89]	06.15%	98.31%	98.47%	98.16%	67.57%	59.55%
UNetVGG16 [88]	06.70%	98.66%	98.76%	98.57%	72.48%	64.33%
VResUNet++ (ResNet50 + VGG16) (Dropout, L2 Regularization, and Finetune)	**04.40%**	**98.69%**	**98.83%**	**98.65%**	**72.93%**	**65.00%**

Table 4.7: IoU and Dice Coefficient per class across different models for Weedmap dataset [5]

	IoU Per Class			**Dice Coefficient**		
Metrics	**Crop**	**Weed**	**Background**	**Crop**	**Weed**	**Background**
UNet [83]	62.89%	20.67%	96.23%	76.45%	34.64%	96.78%
UNetResNet50 [87]	63.89%	21.97%	97.65%	77.97%	36.02%	98.81%
UNetVGG16 [88]	74.26%	31.99%	98.52%	85.23%	48.48%	99.25%
VResUNet++ (ResNet50 + VGG16) (Dropout, L2 Regularization, and Finetune)	**74.57%**	**36.29%**	**98.57%**	**85.83%**	**53.26%**	**99.28%**

4.8 Comparison with SoTA Methods

To evaluate the effectiveness of the proposed VResUNet++ model, we compare its performance with the state-of-the-art segmentation approaches, including UNet, UNetResNet50, and UNetVGG16, on UAV datasets that are Gobhiset and Weedmap. As summarized in Table 4.4, VResUNet++ achieved the highest dice coefficient of 97.66% on the Gobhiset dataset, outperforming UNetVGG16 by 3.3%, UNetResNet50 by 0.88%, and UNet by 6.11%. In terms of IoU, VResUNet++ achieved the highest performance of 95.54%, which is outperforming the UNetVGG16 by 0.56%, UNetResNet50 by 1.59%, and UNet by 9.73%. This clearly shows that the VResUNet++ model has the best performance over the rest of the models.

In Table 4.5, it is summarized that VResUNet++ performed the best on the Gobhiset dataset. It achieved the highest IoU per class for the crop of 89.38%, outperforming UNetVGG16 by 0.44%, UNetResNet50 by 2%, and UNet by 16.51%. Similarly, the dice coefficient for the crop in VResUNet++ is 94.39%, which outperforms the UNetVGG16 by 0.25%, UNetResNet50 by 1.13%, and UNet by 10.08%, which shows that VResUNet++ has the best classwise performance in terms of IoU and dice coefficient on the Gobhiset dataset.

The performance of the VResUNet++ model on the Weedmap dataset is summarized in Table 4.6. The VResUNet++ model achieved a dice coefficient of 72.93%, outperforming UNetVGG16 by 0.45%, UNetResNet50 by 3.59%, and UNet by 4.59%. Similarly, VResUNet++ has achieved the IoU of 65.00%, which surpasses UNetVGG16 by 0.67%, UNetResNet50 by 3.87%, and UNet by 4.87%. This indicates that VResUNet++ has the highest performance relative to the rest of the state-of-the-art models.

Table 4.7 presents the comparison of VResUNet++ with the rest of the state-of-the-art models on the Weedmap dataset, which is a multiclass dataset containing 3 classes – crop, weed, and background. The IoU of the VResUNet++ model for crop is 74.57%, outperforming UNetVGG16 by 0.31%, UNetResNet50 by 10.68%, and UNet by 11.68%. The IoU of VResUNet++ for weed is 36.29%, which is outperforming UNetVGG16 by 4.3%, UNetResNet50 by 14.32%, and UNet by 15.62%. Similarly, in the case of the dice coefficient of the VResUNet++ model for crop is 85.83%, which outperforms UNetVGG16 by 0.6%, UNetResNet50 by 7.86%, and UNet by 9.38%. The dice coefficient of the VResUNet++ model in weed is 53.26%, which outperforms the UNetVGG16 by 4.78%, UNetResNet50 by 17.24%, and UNet by 18.62%.

Chapter 5

Conclusions

In this paper, we have proposed a novel VResUNet++ architecture that combines the features of ResNet50 and VGG16 for feature extraction for semantic segmentation in precision agriculture. The hybrid model has helped us in extracting the low-level spatial feature and deep semantic context for improving the segmentation performance in UAV-based crop and weed images.

Through experiments, our model has outperformed state-of-the-art architectures such as UNet, UNetResNet50, and UNetVGG16 in terms of precision, recall, dice coefficient, and IoU. The integration of regularization techniques such as dropout, L2, and loss function helped us in minimizing the overfitting and generalizing in the agriculture scenario.

This work highlights the performance of the hybrid encoder design for high-resolution UAV-based image segmentation in agriculture. In the future, we aim to explore the real-time deployment of our proposed work for crop and weed image segmentation.

Bibliography

[1] Terra Drone Agri, "Traditional and modern farming: What you need to know," https://terra-droneagri.com/traditional-and-modern-farming-what-you-need-to-know/, 2023, accessed: May 23, 2025.

[2] Agri Farming, "Soil preparation in agriculture – methods and tips," https://www.agrifarming.in/soil-preparation-in-agriculture-methods-and-tips, 2024, accessed: 2025-06-04. [Online]. Available: https://www.agrifarming.in/soil-preparation-in-agriculture-methods-and-tips

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015. [Online]. Available: https://arxiv.org/abs/1512.03385

[4] S. Rana, M. Crimaldi, D. Barretta, P. Carillo, V. Cirillo, A. Maggio, F. Sarghini, and S. Gerbino, "Gobhiset: Dataset of raw, manually, and automatically annotated rgb images across phenology of brassica oleracea var. botrytis," *Data in Brief*, vol. 54, p. 110506, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S235234092400475X

[5] R. K. Z. C. P. L. F. L. J. N. C. S. A. W. I. Sa, M. Popovic and R. Siegwart, "Weedmap: A large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming," *MDPI Remote Sensing*, vol. 10, no. 9, Aug 2018.

[6] N. Ikonomatakis, K. Plataniotis, M. Zervakis, and A. Venetsanopoulos, "Region growing and region merging image segmentation," in *Proceedings of 13th International Conference on Digital Signal Processing*, vol. 1, July 1997, pp. 299–302 vol.1.

[7] G.-Q. Jiang, C.-J. Zhao, and J.-Y. Qi, "The research of image segmentation based on color characteristic," in *2011 International Conference on Machine Learning and Cybernetics*, vol. 4, July 2011, pp. 1851–1855.

[8] M. Guijarro, G. Pajares, I. Riomoros, P. Herrera, X. Burgos-Artizzu, and A. Ribeiro, "Automatic segmentation of relevant textures in agricultural images," *Computers and Electronics in Agriculture*, vol. 75, no. 1, pp. 75–83, 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169910001924

[9] R. S. Sarkate, N. V. Kalyankar, and P. B. Khanale, "Application of computer vision and color image segmentation for yield prediction precision," in *2013 International Conference on Information Systems and Computer Networks*, March 2013, pp. 9–13.

[10] N. Jothiaruna, K. Joseph Abraham Sundar, and B. Karthikeyan, "A segmentation method for disease spot images incorporating chrominance in comprehensive color feature and region growing," *Computers and Electronics in Agriculture*, vol. 165, p. 104934, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169919300730

[11] S. Tan, X. Ma, Z. Mai, L. Qi, and Y. Wang, "Segmentation and counting algorithm for touching hybrid rice grains," *Computers and Electronics in Agriculture*, vol. 162, pp. 493–504, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016816991830913X

[12] Y. Xue, J. Zhao, and M. Zhang, “A watershed-segmentation-based improved algorithm for extracting cultivated land boundaries,” *Remote Sensing*, vol. 13, no. 5, 2021. [Online]. Available: https://www.mdpi.com/2072-4292/13/5/939

[13] Y. Lu, S. Young, H. Wang, and N. Wijewardane, “Robust plant segmentation of color images based on image contrast optimization,” *Computers and Electronics in Agriculture*, vol. 193, p. 106711, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016816992200028X

[14] A. Kumar, A. Kumar, A. Vishwakarma, and G. K. Singh, “Multilevel thresholding for crop image segmentation based on recursive minimum cross entropy using a swarm-based technique,” *Computers and Electronics in Agriculture*, vol. 203, p. 107488, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169922007967

[15] E. Cerruto, G. Manetto, S. Privitera, R. Papa, and D. Longo, “Effect of image segmentation thresholding on droplet size measurement,” *Agronomy*, vol. 12, no. 7, 2022. [Online]. Available: https://www.mdpi.com/2073-4395/12/7/1677

[16] Y. Han, Y. Wang, and Y. Zhao, “Support vector machine-based image segmentation approach for automatic agriculture vehicle,” in *2012 International Conference on Image Analysis and Signal Processing*, 2012, pp. 1–5.

[17] H. Yue, K. Cai, H. Lin, H. Man, and Z. Zeng, “A markov random field model for image segmentation of rice planthopper in rice fields,” *Journal of Engineering Science and Technology Review*, vol. 9, pp. 31–38, 04 2016.

[18] B. M. S. Rangel, M. A. A. Fernández, J. C. Murillo, J. C. Pedraza Ortega, and J. M. R. Arreguín, “Knn-based image segmentation for grapevine potassium deficiency diagnosis,” in *2016 International Conference on Electronics, Communications and Computers (CONIELECOMP)*, Feb 2016, pp. 48–53.

[19] K. Tian, J. Li, J. Zeng, A. Evans, and L. Zhang, "Segmentation of tomato leaf images based on adaptive clustering number of k-means algorithm," *Computers and Electronics in Agriculture*, vol. 165, p. 104962, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169919305538

[20] I.-A. Dávila-Rodríguez, M.-A. Nuño-Maganda, Y. Hernández-Mier, and S. Polanco-Martagón, "Decision-tree based pixel classification for real-time citrus segmentation on fpga," in *2019 International Conference on ReConFigurable Computing and FPGAs (ReConFig)*, Dec 2019, pp. 1–8.

[21] A. P. Marques Ramos, L. Prado Osco, D. Elis Garcia Furuya, W. Nunes Gonçalves, D. Cordeiro Santana, L. Pereira Ribeiro Teodoro, C. Antonio da Silva Junior, G. Fernando Capristo-Silva, J. Li, F. Henrique Rojo Baio, J. Marcato Junior, P. Eduardo Teodoro, and H. Pistori, "A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices," *Computers and Electronics in Agriculture*, vol. 178, p. 105791, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169920319591

[22] A. Milioto, P. Lottes, and C. Stachniss, "Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in cnns," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 2229–2235.

[23] M. A. Khan, T. Akram, M. Sharif, M. Awais, K. Javed, H. Ali, and T. Saba, "Ccdf: Automatic system for segmentation and recognition of fruit crops diseases based on correlation coefficient and deep cnn features," *Computers and Electronics in Agriculture*, vol. 155, pp. 220–236, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169918303120

[24] S. Zhuang, P. Wang, and B. Jiang, “Segmentation of green vegetation in the field using deep neural networks,” in *2018 13th World Congress on Intelligent Control and Automation (WCICA)*, 2018, pp. 509–514.

[25] Y. Qiao, M. Truman, and S. Sukkarieh, “Cattle segmentation and contour extraction based on mask r-cnn for precision livestock farming,” *Computers and Electronics in Agriculture*, vol. 165, p. 104958, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169919304077

[26] M. Fawakherji, A. Youssef, D. Bloisi, A. Pretto, and D. Nardi, “Crop and weeds classification for precision agriculture using context-independent pixel-wise segmentation,” in *2019 Third IEEE International Conference on Robotic Computing (IRC)*, 2019, pp. 146–152.

[27] Y. Xiong, L. Liang, L. Wang, J. She, and M. Wu, “Identification of cash crop diseases using automatic image segmentation algorithm and deep learning with expanded dataset,” *Computers and Electronics in Agriculture*, vol. 177, p. 105712, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169920300284

[28] Y. Yue, X. Li, H. Zhao, and H. Wang, “Image segmentation method of crop diseases based on improved segnet neural network,” in *2020 IEEE International Conference on Mechatronics and Automation (ICMA)*, 2020, pp. 1986–1991.

[29] H. Peng, C. Xue, Y. Shao, K. Chen, J. Xiong, Z. Xie, and L. Zhang, “Semantic segmentation of litchi branches using deeplabv3+ model,” *IEEE Access*, vol. 8, pp. 164 546–164 555, 2020.

[30] Y. Zhang and M. Chi, “Mask-r-fcn: A deep fusion network for semantic segmentation,” *IEEE Access*, vol. 8, pp. 155 753–155 765, 2020.

[31] W. Jia, Y. Tian, R. Luo, Z. Zhang, J. Lian, and Y. Zheng, "Detection and segmentation of overlapped fruits based on optimized mask r-cnn application in apple harvesting robot," *Computers and Electronics in Agriculture*, vol. 172, p. 105380, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169919326274

[32] Z. Li and Y. Guo, "Semantic segmentation of landslide images in nyingchi region based on pspnet network," in *2020 7th International Conference on Information Science and Control Engineering (ICISCE)*, 2020, pp. 1269–1273.

[33] J. You, W. Liu, and J. Lee, "A dnn-based semantic segmentation for detecting weed and crop," *Computers and Electronics in Agriculture*, vol. 178, p. 105750, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169920305792

[34] A. Wang, Y. Xu, X. Wei, and B. Cui, "Semantic segmentation of crop and weed using an encoder-decoder network and image enhancement method under uncontrolled outdoor illumination," *IEEE Access*, vol. 8, pp. 81 724–81 734, 2020.

[35] Z. Wu, R. Yang, F. Gao, W. Wang, L. Fu, and R. Li, "Segmentation of abnormal leaves of hydroponic lettuce based on deeplabv3+ for robotic sorting," *Computers and Electronics in Agriculture*, vol. 190, p. 106443, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169921004609

[36] Q. Lv and H. Wang, "Cotton boll growth status recognition method under complex background based on semantic segmentation," in *2021 4th International Conference on Robotics, Control and Automation Engineering (RCAE)*, 2021, pp. 50–54.

[37] T. Anand, S. Sinha, M. Mandal, V. Chamola, and F. R. Yu, “Agrisegnet: Deep aerial semantic segmentation framework for iot-assisted precision agriculture,” *IEEE Sensors Journal*, vol. 21, no. 16, pp. 17 581–17 590, 2021.

[38] K. Zou, X. Chen, Y. Wang, C. Zhang, and F. Zhang, “A modified u-net with a specific data argumentation method for semantic segmentation of weed images in the field,” *Computers and Electronics in Agriculture*, vol. 187, p. 106242, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169921002593

[39] K. Roy, S. S. Chaudhuri, and S. Pramanik, “Deep learning based real-time industrial framework for rotten and fresh fruit detection using semantic segmentation,” *Microsystem Technologies*, vol. 27, no. 9, pp. 3365–3375, 2021. [Online]. Available: https://doi.org/10.1007/s00542-020-05123-x

[40] L. M. Tassis, J. E. Tozzi de Souza, and R. A. Krohling, “A deep learning approach combining instance and semantic segmentation to identify diseases and pests of coffee leaves from in-field images,” *Computers and Electronics in Agriculture*, vol. 186, p. 106191, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169921002088

[41] L. P. Osco, K. Nogueira, A. P. M. Ramos, M. M. F. Pinheiro, D. E. G. Furuya, W. N. Gonçalves, L. A. de Castro Jorge, J. M. Junior, and J. A. dos Santos, “Semantic segmentation of citrus-orchard using deep neural networks and multispectral uav-based imagery,” *Precision Agriculture*, vol. 22, no. 4, pp. 1171–1188, 2021. [Online]. Available: https://doi.org/10.1007/s11119-020-09777-5

[42] A. M. Mishra, S. Harnal, V. Gautam, R. Tiwari, and S. Upadhyay, “Weed density estimation in soya bean crop using deep convolutional neural networks in

smart agriculture," *Journal of Plant Diseases and Protection*, vol. 129, no. 3, pp. 593–604, 2022. [Online]. Available: https://doi.org/10.1007/s41348-022-00595-7

[43] R. Zhang, J. Chen, L. Feng, S. Li, W. Yang, and D. Guo, "A refined pyramid scene parsing network for polarimetric sar image semantic segmentation in agricultural areas," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[44] Q. Sun, R. Zhang, L. Chen, L. Zhang, H. Zhang, and C. Zhao, "Semantic segmentation and path planning for orchards based on uav images," *Computers and Electronics in Agriculture*, vol. 200, p. 107222, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169922005361

[45] I. B. M. Y. Wirawan, I. M. G. Sunarya, and I. M. D. Maysanjaya, "Semantic segmentation of rice field bund on unmanned aerial vehicle image using unet," in *2022 14th International Conference on Information Technology and Electrical Engineering (ICITEE)*, 2022, pp. 211–216.

[46] Y. He, X. Zhang, Z. Zhang, and H. Fang, "Automated detection of boundary line in paddy field using mobilev2-unet and ransac," *Computers and Electronics in Agriculture*, vol. 194, p. 106697, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016816992200014X

[47] E. Raei, A. Akbari Asanjan, M. R. Nikoo, M. Sadegh, S. Pourshahabi, and J. F. Adamowski, "A deep learning image segmentation model for agricultural irrigation system classification," *Computers and Electronics in Agriculture*, vol. 198, p. 106977, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169922002940

[48] B. Niu, Q. Feng, B. Chen, C. Ou, Y. Liu, and J. Yang, "Hsi-transunet: A transformer based semantic segmentation model for crop mapping from uav hyperspectral imagery," *Computers and Electronics*

in Agriculture, vol. 201, p. 107297, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169922006093

[49] S. Jin, H. Dai, J. Peng, Y. He, M. Zhu, W. Yu, and Q. Li, "An improved mask r-cnn method for weed segmentation," in *2022 IEEE 17th Conference on Industrial Electronics and Applications (ICIEA)*, 2022, pp. 1430–1435.

[50] H. Peng, J. Zhong, H. Liu, J. Li, M. Yao, and X. Zhang, "Resdense-focal-deeplabv3+ enabled litchi branch semantic segmentation for robotic harvesting," *Computers and Electronics in Agriculture*, vol. 206, p. 107691, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169923000790

[51] S. Zhu, W. Ma, J. Lu, B. Ren, C. Wang, and J. Wang, "A novel approach for apple leaf disease image segmentation in complex scenes based on two-stage deeplabv3+ with adaptive loss," *Computers and Electronics in Agriculture*, vol. 204, p. 107539, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016816992200847X

[52] G. Baravdish and P. Ranawaka, "Semantic segmentation of weed and crop with partially annotated data for automated agriculture," in *2023 IEEE International Conference on Agrosystem Engineering, Technology Applications (AGRETA)*, 2023, pp. 17–22.

[53] Y. Cao, Z. Zhao, Y. Huang, X. Lin, S. Luo, B. Xiang, and H. Yang, "Case instance segmentation of small farmland based on mask r-cnn of feature pyramid network with double attention mechanism in high resolution satellite images," *Computers and Electronics in Agriculture*, vol. 212, p. 108073, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169923004611

[54] Y. Cai, F. Zeng, J. Xiao, W. Ai, G. Kang, Y. Lin, Z. Cai, H. Shi, S. Zhong, and X. Yue, "Attention-aided semantic segmentation network for weed identification in pineapple field," *Computers and Electronics in Agriculture*, vol. 210, p. 107881, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169923002697

[55] J. Feng, X. Chao, Z. Zhang, D. He, J. Zhang, and Z. Ye, "A pooling module with multidirectional and multi-scale spatial information and its application on semantic segmentation of leaf lesions," *Precision Agriculture*, vol. 24, no. 6, pp. 2416–2437, 2023. [Online]. Available: https://doi.org/10.1007/s11119-023-10046-4

[56] J. Sun, J. Zhou, Y. He, H. Jia, and Z. Liang, "Rl-deeplabv3+: A lightweight rice lodging semantic segmentation model for unmanned rice harvester," *Computers and Electronics in Agriculture*, vol. 209, p. 107823, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169923002119

[57] N. J. Singh and K. Nongmeikapam, "Semantic segmentation of satellite images using deep-unet," *Arabian Journal for Science and Engineering*, vol. 48, no. 2, pp. 1193–1205, 2023. [Online]. Available: https://doi.org/10.1007/s13369-022-06734-4

[58] C. Yu, D. Lin, and C. He, "Ase-unet: An orange fruit segmentation model in an agricultural environment based on deep learning," *Optical Memory and Neural Networks*, vol. 32, no. 4, pp. 247–257, 2023. [Online]. Available: https://doi.org/10.3103/S1060992X23040045

[59] Y. Wang, L. Gu, T. Jiang, and F. Gao, "Mde-unet: A multitask deformable unet combined enhancement network for farmland boundary segmentation," *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1–5, 2023.

[60] S. Zhang and C. Zhang, "Modified u-net for plant diseased leaf image segmentation," *Computers and Electronics in Agriculture*, vol. 204, p. 107511, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169922008195

[61] Z. Diao, P. Guo, B. Zhang, D. Zhang, J. Yan, Z. He, S. Zhao, and C. Zhao, "Maize crop row recognition algorithm based on improved unet network," *Computers and Electronics in Agriculture*, vol. 210, p. 107940, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169923003289

[62] S. K. Gupta, S. K. Yadav, S. K. Soni, U. Shanker, and P. K. Singh, "Multiclass weed identification using semantic segmentation: An automated approach for precision agriculture," *Ecological Informatics*, vol. 78, p. 102366, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1574954123003953

[63] W. Zhang, Y. Wang, G. Shen, C. Li, M. Li, and Y. Guo, "Tobacco leaf segmentation based on improved mask rcnn algorithm and sam model," *IEEE Access*, vol. 11, pp. 103 102–103 114, 2023.

[64] Z. Guo, Y. Geng, C. Wang, Y. Xue, D. Sun, Z. Lou, T. Chen, T. Geng, and L. Quan, "Instacropnet: An efficient unet-based architecture for precise crop row detection in agricultural applications," *Artificial Intelligence in Agriculture*, vol. 12, pp. 85–96, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2589721724000175

[65] S. Liu, Z. Huang, Z. Xu, F. Zhao, D. Xiong, S. Peng, and J. Huang, "High-throughput measurement method for rice seedling based on improved unet model," *Computers and Electronics in Agriculture*, vol. 219, p. 108770,

2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169924001613

[66] A. Purohit, P. Pujari, K. Shah, and A. V. Nimkar, "Crop and weed segmentation using resnet-unet architecture," in *2024 First International Conference on Data, Computation and Communication (ICDCC)*, 2024, pp. 1–8.

[67] L. Zhang, M. Li, X. Zhu, Y. Chen, J. Huang, Z. Wang, T. Hu, Z. Wang, and K. Fang, "Navigation path recognition between rows of fruit trees based on semantic segmentation," *Computers and Electronics in Agriculture*, vol. 216, p. 108511, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169923008992

[68] M. A. Bhatti, M. Syam, H. Chen, Y. Hu, L. W. Keung, Z. Zeeshan, Y. A. Ali, and N. Sarhan, "Utilizing convolutional neural networks (cnn) and u-net architecture for precise crop and weed segmentation in agricultural imagery: A deep learning approach," *Big Data Research*, vol. 36, p. 100465, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214579624000418

[69] B. Yang, S. Yang, P. Wang, H. Wang, J. Jiang, R. Ni, and C. Yang, "Frpnet: An improved faster-resnet with paspp for real-time semantic segmentation in the unstructured field scene," *Computers and Electronics in Agriculture*, vol. 217, p. 108623, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169924000140

[70] Y. Shen, X. Sun, J. Cui, and Y. Lu, "Application of pyramid scene parsing network in leaf segmentation for wheat stripe rust," in *2024 5th International Conference on Computer Vision, Image and Deep Learning (CVIDL)*, 2024, pp. 926–930.

[71] S. Islam, M. N. Reza, M. Chowdhury, S. Ahmed, K.-H. Lee, M. Ali, Y. J. Cho, D. H. Noh, and S.-O. Chung, "Detection and segmentation of lettuce seedlings from seedling-growing tray imagery using an improved mask r-cnn method," *Smart Agricultural Technology*, vol. 8, p. 100455, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2772375524000601

[72] B. Deka and D. Chakraborty, "Uav sensing-based litchi segmentation using modified mask-rcnn for precision agriculture," *IEEE Transactions on AgriFood Electronics*, vol. 2, no. 2, pp. 509–517, 2024.

[73] L. Cai, C. Kang, D. Zhang, and S. Wang, "Cultivated land parcel recognition method based on deeplabv3+ semantic segmentation model for high-resolution unmanned aerial vehicle (uav) imagery," in *2024 4th International Conference on Electronic Information Engineering and Computer Communication (EIECC)*, 2024, pp. 662–666.

[74] J. Wu, X. Yuan, Y. Yang, T. Xia, Y. Li, J. hu Cheng, C. Yu, and C. Liu, "Research on terahertz image analysis of thin-shell seeds based on semantic segmentation," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 323, p. 124897, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1386142524010631

[75] Z. Wang, L. Yang, R. Wang, L. Lei, H. Ding, and Q. Yang, "We-deeplabv3+: A lightweight segmentation model for panax notoginseng leaf diseases," *Computers and Electronics in Agriculture*, vol. 227, p. 109612, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169924010032

[76] J. Lu, Y. Yang, H. Zhong, and T. Wu, "Agricultural field segmentation and corn yield prediction based on deeplabv3+ and xgbregressor," in *2024 9th International*

Conference on Electronic Technology and Information Science (ICETIS), 2024, pp. 694–698.

[77] M. Habib, S. Sekhra, A. Tannouche, and Y. Ounejjar, "New segmentation approach for effective weed management in agriculture," *Smart Agricultural Technology*, vol. 8, p. 100505, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2772375524001102

[78] M. Peng, Y. Liu, I. A. Qadri, U. A. Bhatti, B. Ahmed, N. M. Sarhan, and E. Awwad, "Advanced image segmentation for precision agriculture using cnn-gat fusion and fuzzy c-means clustering," *Computers and Electronics in Agriculture*, vol. 226, p. 109431, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169924008226

[79] Y. Miao, R. Wang, Z. Jing, K. Wang, M. Tan, F. Li, W. Zhang, J. Han, and Y. Han, "Ct image segmentation of foxtail millet seeds based on semantic segmentation model vgg16-unet," *Plant Methods*, vol. 20, no. 1, p. 169, 2024. [Online]. Available: https://doi.org/10.1186/s13007-024-01288-y

[80] Y. Wang, X. Gao, Y. Sun, Y. Liu, L. Wang, and M. Liu, "Semantic segmentation-based conservation tillage corn straw return cover type recognition," *Computers and Electronics in Agriculture*, vol. 229, p. 109792, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169924011839

[81] Y. Xu, B. Ma, G. Yu, R. Zhang, H. Tan, F. Dong, and H. Bian, "Accurate cotton verticillium wilt segmentation in field background based on the two-stage lightweight deeplabv3+ model," *Computers and Electronics in Agriculture*, vol. 229, p. 109814, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169924012055

[82] C. Yao, D. Lv, H. Li, J. Fu, C. Li, X. Gao, and D. Hong, "A real-time crop lodging recognition method for combine harvesters based on machine vision and modified deeplab v3+," *Smart Agricultural Technology*, vol. 11, p. 100926, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2772375525001595

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

[84] I. Hadji and R. P. Wildes, "What do we understand about convolutional networks?" 2018. [Online]. Available: https://arxiv.org/abs/1803.08834

[85] I. Kouretas and V. Paliouras, "Simplified hardware implementation of the softmax activation function," in *2019 8th International Conference on Modern Circuits and Systems Technologies (MOCAST)*, 2019, pp. 1–4.

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

[87] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 162, pp. 94–114, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271620300149

[88] M. Ullah, F. Islam, and A. Bais, "Quantifying consistency of crop establishment using a lightweight u-net deep learning architecture and image processing techniques," *Computers and Electronics in Agriculture*, vol. 217, p. 108617,

2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169924000085

[89] F. Moodi, F. K. Shoushtari, G. Valizadeh, D. Mazinani, H. M. Salari, and H. S. Rad, “Attention xception unet (axunet): A novel combination of cnn and self-attention for brain tumor segmentation,” 2025. [Online]. Available: https://arxiv.org/abs/2503.20446