International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(6s) 2026 pp. 1182-1204
https://doi.org/10.70917/ijcisim-2026-3026
Article

Attention-Enhanced U-Net with Residual Learning and ASPP for Efficient Crop-Weed Semantic Segmentation

Priyanka[1], Amit Verma[2]

[1]Department of Computer Science & Engineering Maharaja Agrasen University, Baddi
Department of Computer Science P.I.G. Government College for women, Jind
[2]Department of Computer Science & Engineering Maharaja Agrasen University, Baddi
Email: 1 ms.redhu@gmail.com

Abstract: Efficient crop-weed discrimination is a crucial necessity for precision agriculture, which enables site-specific weed control, optimum herbicide use and sustained crop production. However, semantic segmentation of agricultural field photos is still problematic due to diverse backdrop, varied lighting conditions, overlapping vegetation, uneven plant structure and multi-scale item features. In such circumstances, traditional deep learning algorithms generally fail to keep the fine object boundaries and properly partition tiny weed instances. Thus, this work presents an Attention-Enhanced U-Net with Residual Learning and Atrous Spatial Pyramid Pooling (ASPP) to enhance crop-weed semantic segmentation using the PhenoBench dataset. Firstly, the transfer learning model ResNet50 was selected as the baseline image classification model and the traditional U-Net was selected as the baseline semantic segmentation model. The suggested design extends the basic U-Net architecture with residual blocks for efficient propagation of features, ASPP module for multi-scale contextual feature extraction and attention gates for adaptive feature refining. The model was trained using Adam optimizer using the hybrid loss function of Dice Loss and Sparse Categorical Cross-Entropy Loss. Experimental assessment was conducted using accuracy, precision, recall, F1-score, intersection over union (IoU) and dice coefficient, and qualitative evaluations comprising prediction maps, overlay visualizations and error maps. The suggested model performed better than the baseline ResNet50 and the traditional U-Net in segmentation, with greater border localization, better weed recognition and better retention of fine structural elements. The findings reveal that the suggested framework provides an efficient and robust solution for the automatic crop and weed semantic segmentation, and has excellent promise for real-world precision agricultural applications.

Keywords: Precision Agriculture; Crop–Weed Segmentation; Semantic Segmentation; Attention-Enhanced U-Net; Residual Learning; ASPP; Deep Learning; PhenoBench Dataset.

1. Introduction

Agricultural sector is one of the most significant sectors that supports the world food security and economic growth. With the world's population continuing to rise, agricultural systems are under growing pressure to increase yields while minimizing environmental consequences and maximizing the use of natural resources. Weeds are one of the principal limiting factors of agricultural output. Weeds compete with crops for sunshine, nutrients, water, and room for growth, and may significantly reduce crop output and quality. Conventional weed control approaches are based on manual field inspection and blanket application of herbicides. These are labour demanding, time consuming, costly and frequently result in excessive use of chemicals. These constraints have led to the widespread adoption of precision agriculture, where sophisticated sensing technology and intelligent decision support systems are used to implement site-specific weed control strategies that increase output while reducing environmental harm. Recent advances in computer vision and artificial intelligence have revolutionized agricultural surveillance allowing automated analysis of field pictures captured with ground-based cameras, UAVs and agricultural robots. Among these techniques, deep learning has become one of the most successful techniques for agricultural image analysis as it learns hierarchical feature representations straight from raw photos automatically without the need of manual feature extraction. CNNs have shown great success in crop categorization, weed identification, plant disease diagnosis, and yield calculation. Transfer learning may be further used to further expand their applicability by adapting pre-trained models such as ResNet50 to agricultural datasets with relatively few annotated samples. But image-level classification is only able to indicate the presence or absence of crop and weed categories, and cannot reliably identify the spatial position of the crop and weed categories in agricultural settings.

Semantic segmentation overcomes this restriction by adding a semantic name to each pixel in a picture, thereby permitting comprehensive localization of crop, weed, soil and background areas. This kind of pixel-perfect knowledge is especially critical in precision agriculture where autonomous sprayers, robotic weed-eaters and smart farm equipment rely on correct location information for focused field activities. U-Net is one of the most used encoder-decoder networks for semantic segmentation designs due to its efficient skip connections and its ability to maintain spatial information. Despite the effectiveness of the U-Net design, it typically fails to segment complicated agricultural sceneries, which are characterized by thick vegetation, overlapping crop and weed areas, variable lighting conditions, irregular plant structures, and objects at several sizes. Such issues often result in improper border delineation, misunderstanding of visually similar classes and lower segmentation accuracy. Recent work has addressed these constraints by integrating more sophisticated architectural components, such as residual learning, attention mechanisms, and ASPP, into deep neural networks. Residual learning may help effective gradient propagation and deeper feature extraction, while ASPP can collect contextual information at numerous receptive fields to enhance the detection of objects with different sizes. Attention processes improve segmentation efficiency by selectively highlighting useful spatial characteristics and suppressing irrelevant background information. The integration of these algorithms offers a robust framework to improve pixel-level categorization and to preserve precise object boundaries in demanding agricultural situations.

Motivated by these developments, this work presents an Attention-Enhanced U-Net with Residual Learning and ASPP for accurate crop-weed semantic segmentation using the PhenoBench dataset. For the full performance benchmark, we first design a ResNet50 transfer learning model as the baseline image classification framework, and then a traditional U-Net as the baseline semantic segmentation model. The suggested architecture is an extension of the conventional U-Net with residual blocks for increased feature propagation, ASPP for extracting multi-scale contextual features, and attention gates for adaptively refining features in skip connections. The model is refined via a hybrid loss function including Dice Loss and Sparse Categorical Cross-Entropy Loss to increase the overlap of regions and pixel-wise classification accuracy. Research does comprehensive qualitative and quantitative tests using the PhenoBench dataset to assess the efficiency of the proposed framework. The performance is evaluated in terms of Accuracy, Precision, Recall, F1-score, IoU, and Dice Coefficient, as well as visual analysis including prediction maps, overlay visualizations, error maps, and confusion matrices. Comparative trials with the baseline ResNet50 classification model and standard U-Net show the efficiency of the proposed architecture for accurate crop-weed semantic segmentation. The proposed framework provides enhanced segmentation accuracy, better boundary preservation, and efficient identification of crop and weed regions, thus contributing to the development of reliable vision-based systems for precision agriculture, assisting intelligent weed management and sustainable agricultural practices.

2. Literature Review

Liu (2020) [1] provided a detailed overview of vision-based weed detection strategies, pointed out the drawbacks of conventional image processing techniques, and stressed the rising necessity of deep learning models for effective weed identification. Benchmark datasets have been released, which has advanced deep learning research in agricultural vision. Steininger et al. [2] established a dataset named CropAndWeed to enable multimodal learning for crop and weed management and provided a consistent baseline for agricultural perception tasks. More recently, Weyler et al. (2024) [3] introduced the PhenoBench dataset providing a large scale standard for semantic picture interpretation in agricultural situations. With pixel-level annotations of crops, weeds and other agricultural items, the dataset is a good choice to evaluate semantic segmentation algorithms in actual field situations. Transfer learning and convolutional neural networks have also been quite successful in agricultural picture processing. Hossain et al. (2026) [4] proposed a vision-language semantic grounding framework to enhance multi-domain crop–weed segmentation by combining visual and semantic representations. Similarly, Islam et al. (2025) [5] suggested lightweight convolutional neural network models for real-time weed identification and segmentation on edge devices, proving the viability of using deep learning techniques in resource-constraint agricultural systems. To improve the segmentation of weeds in UAV images, Khan et al. (2023) [6] presented an encoder–decoder architecture with attention modules. Cai (2023) [7] introduced an attention-aided semantic segmentation network for UAV-based weed mapping, showing the considerable enhancement of attention processes in feature localization and segmentation accuracy in complicated agricultural situations. Li et al. [8] (2025) included attention mechanisms into the U-Net for sugar beet and weed segmentation, resulting in enhanced boundary preservation and classification accuracy. Kailun et al. (2025) [9] designed a dual-task segmentation framework incorporating improved U-Net and unsupervised clustering for oilseed rape and weed segmentation. Mei et al. (2025) [10] developed SSMR-Net by integrating across-feature-mapping attention into U-Net, which enhanced the contextual feature extraction and segmentation performance. For the segmentation of UAV-based pigeon pea fields, Dhore et al. (2026) [11] further modified U-Net with spatial attention modules, which showed a better detection of crops and weeds in different environmental circumstances.

Syed (2025) [12] proposed MSEA-Net, which combined multi-scale feature extraction with spatial-channel attention, and better weed segmentation performance under different weed densities was achieved. Yi et al. (2024) [13] used the coordinate attention combined with VGGNet-based U-Net architecture for grape leaf disease segmentation while Wang et al. (2024) [14] proposed a dilated multi-scale attention U-Net for crop insect pest detection, and showed that the attention modules and multi-scale receptive fields can greatly improve the segmentation performance. Lu et al. (2024) [15] suggested a spatial attention residual U-Net for high-resolution remote sensing image classification, demonstrating the efficiency of residual connections in maintaining deep feature representations. EDM-UNet, an edge-enhanced and attention-guided segmentation framework for soybean weed identification, was proposed by Gao et al. (2025) [16], which efficiently conserved the object boundaries. Gao et al. (2024) [17] also enhanced the semantic segmentation by implementing an improved ERFNet architecture for bean seedling and weed identification. Guo et al. (2024) [18] fused CNN and Transformer architectures for UAV-based rice field weed segmentation to exploit local and global contextual information. Niţu et al. (2025) [19] conducted research on vegetation indices based on hyperspectral data for crop identification and segmentation and shown that spectral information might enhance the classification between crops and weeds. Liu et al. (2024) [20] studied the identification of grass weeds in wheat fields using UAV imaging and deep learning, and the link between these weeds and biomass and yield, showing the practical usefulness of precise weed segmentation for improving agricultural production. Asad et al. (2024) [21] also proposed a diverse data ensemble learning framework for crop and weed detection. Celikkan et al. (2023) [22] incorporated probabilistic modeling and uncertainty quantification into semantic segmentation to enhance the reliability of deep learning predictions on uncertain field conditions. Table 1 shows that attention mechanisms, residual learning, multi-scale feature extraction and the sophisticated encoder-decoder architectures developed in recent works may significantly enhance the crop-weed semantic segmentation performance.

Table 1. Summary of Existing Studies on Crop–Weed Detection and Semantic Segmentation

Ref.	Author / Year	Objective	Methodology	Key Findings	Limitations
[1]	Liu & Bruch (2020)	Review vision-based weed detection methods for selective spraying.	Comprehensive survey of traditional image processing and deep learning techniques.	Deep learning significantly improves weed detection accuracy over conventional methods.	Review paper; no experimental model or comparative validation.
[2]	Steininger et al. (2023)	Develop a multimodal benchmark dataset for crop–weed recognition.	CropAndWeed dataset with RGB and multimodal annotations.	Provides a standardized benchmark for agricultural perception tasks.	Limited crop diversity and field conditions.
[3]	Weyler et al. (2024)	Introduce the PhenoBench benchmark for semantic agricultural scene understanding.	Large-scale RGB dataset with pixel-level annotations and benchmark protocols.	Enables robust evaluation of semantic segmentation models.	Dataset complexity still challenges existing segmentation models.
[4]	Hossain et al. (2026)	Improve crop–weed segmentation using vision-language semantic grounding.	Vision-language foundation model integrating semantic and visual features.	Improved multi-domain segmentation performance.	High computational complexity and large training requirements.

[5]	Islam et al. (2025)	Develop lightweight CNN models for real-time weed detection.	Lightweight CNN architecture for edge devices.	Achieved real-time inference with competitive accuracy.	Slight reduction in segmentation accuracy compared with larger models.
[6]	Khan et al. (2023)	Improve UAV-based weed segmentation using attention modules.	Encoder–decoder network with attention mechanisms.	Enhanced segmentation accuracy and boundary localization.	Performance decreases under severe plant occlusion.
[7]	Cai & Wang (2023)	Map weed distribution using UAV imagery.	Attention-based semantic segmentation network.	Improved localization and segmentation of weeds.	Computationally expensive for high-resolution UAV images.
[8]	Li et al. (2025)	Improve sugar beet and weed segmentation.	U-Net integrated with attention mechanisms.	Better segmentation accuracy and boundary preservation.	Increased model complexity and inference time.
[9]	Kailun et al. (2025)	Segment oilseed rape and weeds simultaneously.	Enhanced U-Net with unsupervised clustering.	Improved dual-task segmentation performance.	Limited evaluation on other crop species.
[10]	Mei et al. (2025)	Improve weed segmentation using cross-feature attention.	SSMR-Net with attention-enhanced U-Net.	Superior contextual feature extraction and segmentation accuracy.	High computational cost.
[11]	Dhore et al. (2026)	Enhance crop–weed segmentation in pigeon pea fields.	Spatial Attention U-Net (UNET-SA).	Improved segmentation under UAV imagery.	Tested only on pigeon pea fields.
[12]	Syed & Sharma (2025)	Improve multi-scale weed segmentation.	MSEA-Net with spatial-channel attention.	Better feature extraction and weed localization.	Performance sensitive to dense vegetation.
[13]	Yi et al. (2024)	Improve semantic segmentation using coordinate attention.	Coordinate Attention with VGG-based U-Net.	Improved feature representation and segmentation accuracy.	Application limited to plant disease segmentation.
[14]	Wang et al. (2024)	Detect crop insect pests.	Dilated multi-scale attention U-Net.	Demonstrated effectiveness of attention and convolutions.	Evaluated on pest detection rather than weed segmentation.
[15]	Lu et al. (2024)	Improve high-resolution image classification.	Spatial Attention Residual U-Net.	Residual learning improved feature preservation and classification accuracy.	Designed for remote sensing rather than agricultural crop–weed segmentation.

2.1 Research Gap

Recent improvements in deep learning have considerably improved the crop–weed detection based on the attention mechanism, residual learning, lightweight convolutional network, and transformer-based architecture. This has prompted the proposal of many improved variations of U-Net, edge-aware networks, attention-guided models, and multimodal techniques for semantic segmentation in agricultural contexts. However, there are still numerous limits. Most approaches are assessed on particular crop kinds or controlled field settings, and so their applicability to varied agricultural settings is limited. Furthermore, traditional U-Net structures generally cannot keep precise object boundaries, multi-scale contextual information, and segment overlapping crops and weeds properly under different lighting and occlusion conditions. The lightweight models may enhance the computational efficiency but may sacrifice the segmentation accuracy. On the other hand, the transformer-based methods usually demand heavy computing. Furthermore, the use of residual learning, attention mechanisms and ASPP in a combined framework for crop-weed semantic segmentation has been investigated using the PhenoBench dataset. Therefore, this paper proposes an

Attention-Enhanced U-Net with Residual Learning and ASPP to improve the extraction of multi-scale features and contextual representation and the accuracy of pixel-level segmentation, and provides a comprehensive comparative analysis with baseline ResNet50 classification and conventional U-Net segmentation models.

3. Problem Statement

Accurate crop–weed discrimination is one of the basic requirements of precision agriculture to allow for focused weed control and to avoid needless herbicide application. However, conventional semantic segmentation methods generally lack the ability to effectively identify crops from weeds owing to the diverse field backdrops, different lighting, overlapping vegetation, and changes in plant development phases. Conventional U-Net-based techniques may not be able to collect enough multi-scale contextual information and retain fine object boundaries which results in segmentation problems. Therefore, a robust deep learning framework is required that efficiently combines residual learning, attention mechanisms and ASPP to optimize the feature representation and pixel-level crop-weed semantic segmentation accuracy under various agricultural field situations.

4. Proposed Work

In this work, we propose an Attention-Enhanced U-Net architecture with Residual Learning and ASPP for accurate crop-weed semantic segmentation using the PhenoBench dataset. The proposed approach aims to address the constraints of standard U-Net designs in gathering multi-scale contextual information and keeping fine object boundaries under complicated agricultural field circumstances. The whole approach includes data preprocessing, baseline model building, suggested model building, model training, semantic segmentation, and performance assessment. First, the baseline image classification model, a ResNet50 transfer learning model is applied to create the benchmark classification performance. Then, we build a typical U-Net architecture as baseline semantic segmentation model. Finally, in the encoder-decoder architecture, the proposed Attention-Enhanced U-Net introduces residual learning blocks, ASPP and attention gates to enhance feature extraction, contextual representation and pixel-level segmentation accuracy. First, the PhenoBench dataset is preprocessed by shrinking all RGB photos and their semantic masks to 256 × 256 pixels. The image normalization is used for better convergence in model training and the semantic masks are encoded into five preset classes for pixel-wise classification. The processed dataset is separated into training and validation sets and transformed into Tensorflow datasets for fast mini-batch learning.

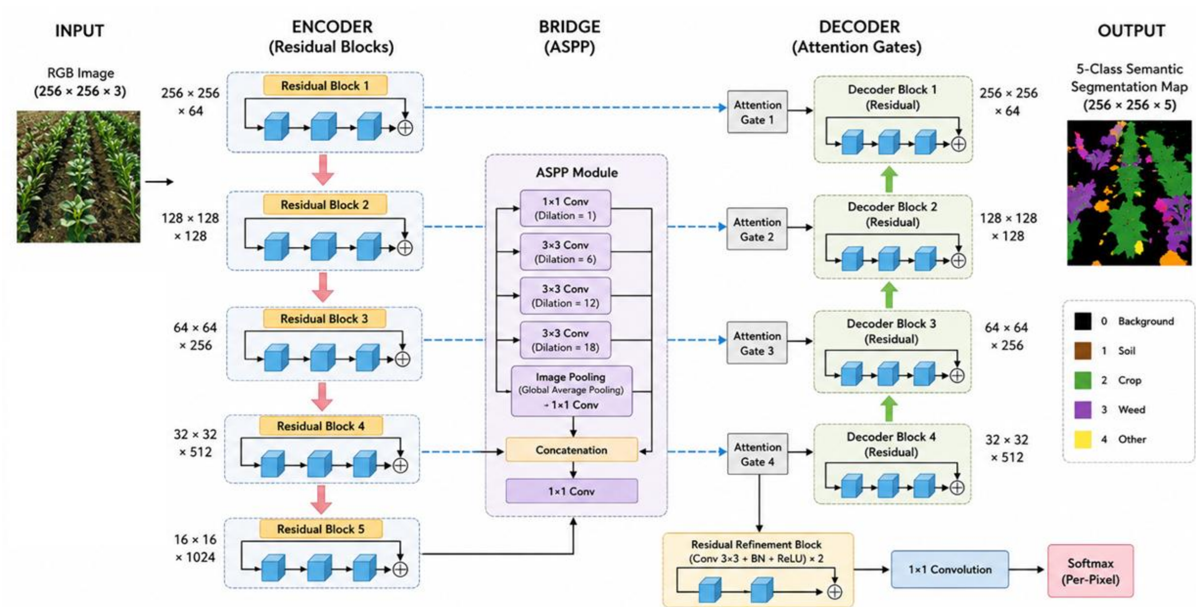


Figure 1. Proposed Model Architecture

The architecture suggested is an encoder-decoder design. The encoder extracts hierarchical feature representations gradually using convolutional operations and residual learning blocks. Residual connections provide good gradient flow, decrease information loss and help in feature learning at deeper levels. The retrieved deep features are processed using an ASPP module which utilizes numerous atrous convolution speeds to capture the contextual information at different spatial scales . This allows the network to accurately detect crops and weeds of different sizes, shapes and geographical distributions. In the decoder route, attention gates are included to boost segmentation performance. These attention processes minimize irrelevant background information while accentuating informative crop and weed areas obtained via skip-connections. The high-resolution segmentation maps are gradually reconstructed by the decoder via the fusion of the features from the encoder with the improved contextual information supplied by the attention modules. Finally, a 1x1 convolutional layer followed by a Softmax activation is used to construct the pixel-wise probability maps for the five semantic classes. The suggested model is trained using Adam optimizer with a combination of Dice Loss and Sparse Categorical Cross Entropy Loss. The Dice Loss increases the overlap of predicted and ground-truth mask regions while Sparse Categorical Cross-Entropy improves the pixelwise classification performance. To further improve the training, ReduceLROnPlateau is used to change the learning rate and ModelCheckpoint to save the best model during training. We do extensive quantitative and qualitative analysis to assess the efficiency of the suggested framework. The quantitative assessment metrics, while the qualitative evaluation metrics are prediction visualization, overlay pictures, error maps, and comparison of segmentation outcomes. Finally, the proposed model is compared with baseline ResNet50 classification model and traditional U-Net segmentation model to show its capacity in enhancing crop-weed semantic segmentation performance in varied agricultural field circumstances.

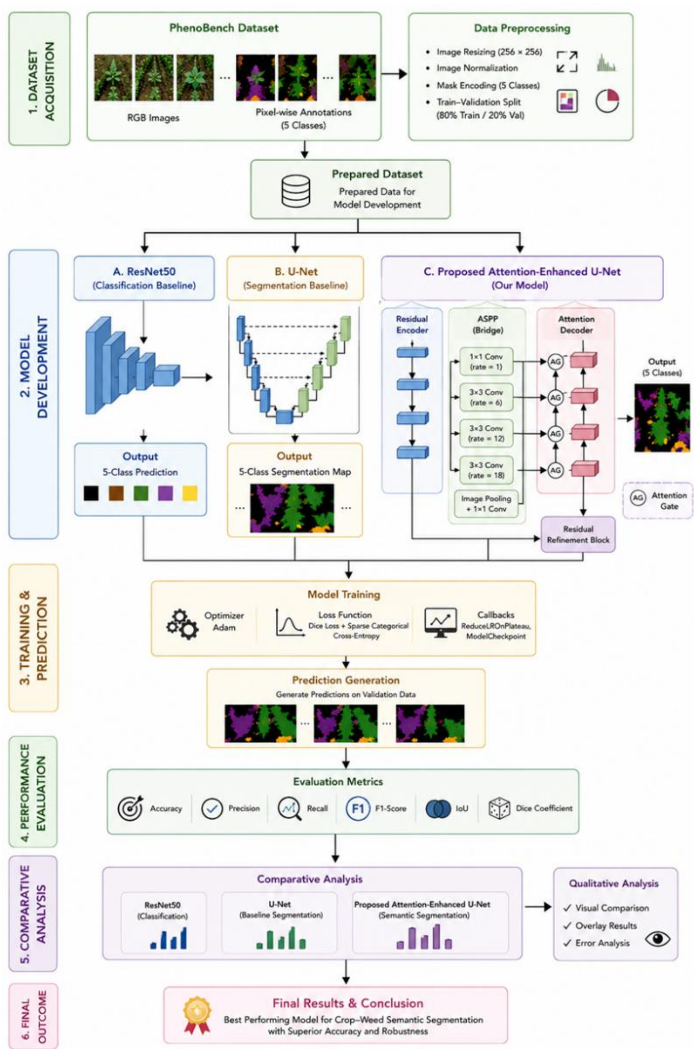


Figure 2. Workflow of proposed crop–weed semantic segmentation framework

The suggested study follows a general workflow, which is a series of phases starting with the dataset collection and finishing with the comparative performance analysis. The purpose of each step is to enhance the accuracy and robustness of crop–weed semantic segmentation. Figure 2 illustrates the whole procedure and specific phases are discussed below.

Step 1: PhenoBench Dataset Acquisition

The study proceeds by obtaining the PhenoBench dataset including RGB agricultural field photos together with pixel-wise semantic annotations. The dataset contains five semantic classes of distinct field items and offers varied agricultural scenarios for training and assessment.

Step 2: Data Preprocessing

The collected dataset is pre-processed before the model training. We scale all RGB photos and their related segmentation masks to 256 × 256 pixels to have a uniform input size. To stabilize training, we use image normalization, then encode semantic masks into five predetermined classes. Finally, the data is split into a training and validation set for model training.

Step 3: Baseline Model Development

Research builds two baseline models for a baseline performance baseline. We perform transfer learning using a ResNet50 model for image classification and create a standard U-Net architecture for semantic segmentation. The baseline models serve as reference results to evaluate the proposed design.

Step 4: Proposed Attention-Enhanced U-Net Development

Research presents a semantic segmentation model that combines Residual Learning, ASPP and Attention Gates into the U-Net network. Residual blocks assist in propagating features, the ASPP module captures multi-scale contextual information, and attention gates optimize skip connections by highlighting important crop and weed properties while reducing background noise.

Step 5: Model Training

The preprocessed training data set is sent into the proposed network for training. Adam optimizer is utilized for parameter optimization. A mixed loss function of Dice Loss and Sparse Categorical Cross-Entropy Loss is used for improving the segmentation performance. Callbacks such as ReduceLROnPlateau and ModelCheckpoint are also used to improve learning and keep best model.

Step 6: Prediction Generation

Once trained the best models will be used to provide pixel wise segmentation predictions for the validation dataset. For each input picture, we construct the semantic segmentation map with five classes, in which each pixel is labeled with the relevant class.

Step 7: Performance Evaluation

The created models are assessed statistically in terms of numerous metrics including Accuracy, Precision, Recall, F1-score, IoU and Dice Coefficient. These metrics provide an overall measure of segmentation quality and model dependability.

Step 8: Comparative Analysis

The last step compares the performance of the classification model ResNet50, the baseline segmentation model U-Net, and the proposed Attention-Enhanced U-Net. We evaluate quantitative metrics as well as qualitative visualizations (prediction results, overlay pictures, and error maps) to illustrate the usefulness of the proposed framework for crop-weed semantic segmentation.

Step 9: Final Results and Conclusion

The suggested Attention-Enhanced U-Net is evaluated on the comparison analysis for accurate crop and weed classification under different agricultural field situations. The experimental results are described, emphasizing the benefits of using Residual Learning, ASPP, and Attention Gates to enhance the performance of semantic segmentation in precision agriculture.

Algorithm 1: Crop–Weed Semantic Segmentation Using the Proposed Attention-Enhanced U-Net

Input: PhenoBench RGB images and Corresponding semantic masks

Output: Five-class semantic segmentation map

Step 1: Acquire the PhenoBench dataset containing RGB images and corresponding semantic segmentation masks.

Step 2: Preprocess the dataset by:

- Resizing images and masks to 256 × 256 pixels.
- Normalizing RGB image pixel values.
- Encoding segmentation masks into five semantic classes.
- Splitting the dataset into training and validation subsets.

Step 3: Develop a ResNet50 transfer learning model as the baseline image classification model.

Step 4: Develop a conventional U-Net architecture as the baseline semantic segmentation model.

Step 5: Construct the proposed Attention-Enhanced U-Net by:

- Building a residual encoder for hierarchical feature extraction.
- Applying Atrous Spatial Pyramid Pooling (ASPP) at the bridge layer for multi-scale feature learning.
- Incorporating attention gates into skip connections.
- Designing a residual decoder with bilinear upsampling.
- Generating the final segmentation map using a 1 × 1 convolution followed by Softmax activation.

Step 6: Train the proposed model using:

- Adam optimizer
- Dice Loss
- Sparse Categorical Cross-Entropy Loss
- ReduceLROnPlateau learning rate scheduler
- ModelCheckpoint callback

Step 7: Generate pixel-wise predictions for validation images.

Step 8: Evaluate the model using:

- Accuracy
- Precision
- Recall
- F1-score
- Intersection over Union (IoU)
- Dice Coefficient

Step 9: Compare the proposed model with:

- ResNet50 Classification
- Baseline U-Net Segmentation

Step 10: Analyze qualitative results using prediction images, overlay visualizations, and error maps.

Step 11: Select the best-performing model based on quantitative and qualitative evaluation metrics.

End Algorithm

4.1 PhenoBench Dataset Description

The proposed crop-weed semantic segmentation methodology is created and experimentally tested using the publicly accessible PhenoBench dataset, a standard for semantic picture interpretation in precision agriculture. The dataset was published by Jan Weyler et al. (2024) to enable the creation and assessment of deep learning systems for agricultural sensing tasks. It is made up of high resolution RGB photos taken by field robots and imaging equipment under actual agricultural field circumstances. PhenoBench is a very good dataset for semantic segmentation tasks on crops, weeds and field backdrop since it has deep pixel-level annotations unlike other agricultural datasets. The collection includes a wide range of agricultural sceneries taken under different lighting circumstances, soil textures, plant development phases, and weed densities. In this study, we simply use the RGB pictures and the related semantic segmentation masks.

Table 2. Summary of the PhenoBench Dataset

Parameter	Description
Dataset Name	PhenoBench
Application	Crop–Weed Semantic Segmentation
Data Type	RGB Images with Pixel-Level Semantic Masks
Image Type	Agricultural Field Images
Input Image Size	256 × 256 × 3 (after preprocessing)
Number of Classes	5
Annotation Type	Pixel-wise Semantic Segmentation
Image Normalization	Pixel values scaled to [0, 1]
Dataset Split	Training and Validation
Output	Five-Class Semantic Segmentation Map
Benchmark Purpose	Evaluation of Deep Learning Models for Precision Agriculture

5. Experimental Results and Discussion

In this part, we experimentally evaluate the proposed Attention-Enhanced U-Net using the PhenoBench dataset. The suggested model is evaluated qualitatively and quantitatively with the baseline ResNet50 classification model and the standard U-Net segmentation model.

5.1 Experimental Environment

The performance of the proposed Attention-Enhanced U-Net for crop–weed semantic segmentation was evaluated experimentally using the PhenoBench dataset. To provide a fair comparison, the classification model baseline ResNet50, traditional U-Net and the suggested model were trained and assessed under the identical experimental setup. All experiments were implemented using the TensorFlow deep learning framework on the Kaggle Notebook cloud platform with GPU acceleration. To achieve steady convergence and good segmentation accuracy, the proposed model was trained using carefully chosen hyperparameters. The raw RGB photos were scaled to 256 × 256 pixels and the semantic masks were encoded to five classes. The Adam optimizer is used because of its adaptive learning and efficient convergence capabilities. To boost the performance on both the region overlap and the pixel-wise classification, we use a mixed loss function that consists of Dice Loss and Sparse Categorical Cross Entropy Loss. The ReduceLROnPlateau callback was also used to dynamically change the learning rate during training, and the ModelCheckpoint callback saved the best performing model based on validation performance. The hyperparameter settings utilized for all tests are shown in Table 3.

Table 3. Hyperparameter Configuration

Hyperparameter	Value
Input Image Size	256 × 256 × 3
Number of Classes	5
Batch Size	8
Optimizer	Adam

Initial Learning Rate	0.001
Loss Function	Dice Loss + Sparse Categorical Cross-Entropy
Hidden Layer Activation	ReLU
Output Activation	Softmax
Classification Epochs (ResNet50)	25
Segmentation Epochs (U-Net & Proposed Model)	30
Learning Rate Scheduler	ReduceLROnPlateau
Learning Rate Reduction Factor	0.2
Patience	2 Epochs
Minimum Learning Rate	1×10^{-7}
Model Saving Strategy	ModelCheckpoint

5.2 Dataset Visualization

Visualizing the dataset is an important step to have an idea of the properties of the PhenoBench dataset before training the model. The visual examination enables to validate the quality of the RGB photos, the semantic annotations and the class distribution and to discover issues such as overlapping vegetation, complicated backdrops, variable lighting conditions and different crop development phases.

5.2.1 Sample RGB Images

The PhenoBench dataset consists of high-resolution RGB photographs captured under realistic agricultural field circumstances. These photographs depict crops, weeds, soil and other background items with different environmental variables such as different light, different plant density, different viewing angles and different development phases. These changes significantly complicate the semantic segmentation problem, and the dataset is well suited for evaluation of robust deep learning models. Figure 3 shows some RGB pictures from the training set. The chosen photographs include a variety of agricultural settings, such as thick vegetation, scarce weeds, overlapping areas of crops and weeds, different soil backdrops, etc. The variety allows the proposed model to learn discriminative visual characteristics in actual agricultural scenarios.

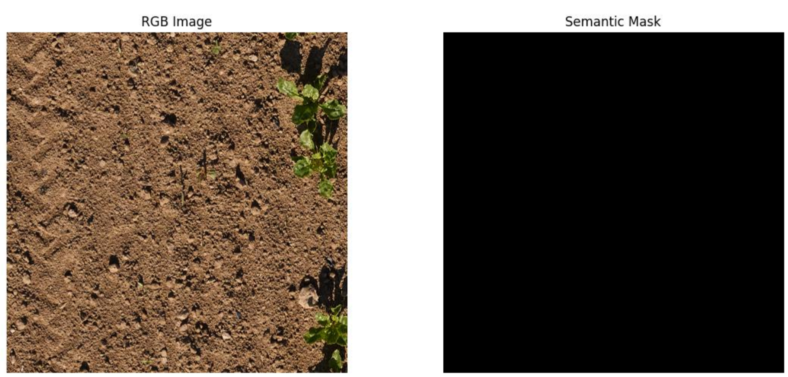


Figure 3. Representative RGB Images from the PhenoBench Dataset

5.2.2 Ground Truth Semantic Masks

The PhenoBench dataset includes an RGB picture for each ground-truth annotation which is a pixel-level semantic mask. Each pixel is given to one of the semantic classes, which allows supervised semantic segmentation. correct spatial information on crop areas, weed regions, soil and other field components are provided by these annotations, so that the network may learn correct pixel-wise classification. Examples of the related semantic masks are shown in figure 4. distinct semantic categories are represented by distinct hues, which makes the object boundary and class distribution clearly shown.

Figure 4. Ground Truth Semantic Masks

5.2.3 Class Distribution Analysis

The allocation of pixels to semantic classes has a substantial impact on the performance of semantic segmentation. In agricultural datasets, certain classes tend to have far more pixels than others, resulting to class imbalance when training models. Such imbalance may lead to bias in the learning process toward dominating classes, degrading the segmentation performance of minority classes. For a more detailed understanding of the dataset composition, the pixel-wise distribution of all semantic classes is studied and shown in Figure 5. The class distribution shows the frequency of each class and the difficulty of learning from unbalanced semantic material. The suggested approach employs a combined Dice Loss and Sparse Categorical Cross-Entropy Loss to minimize the impact of class imbalance by maximizing the overlap while preserving the pixel-wise classification accuracy.

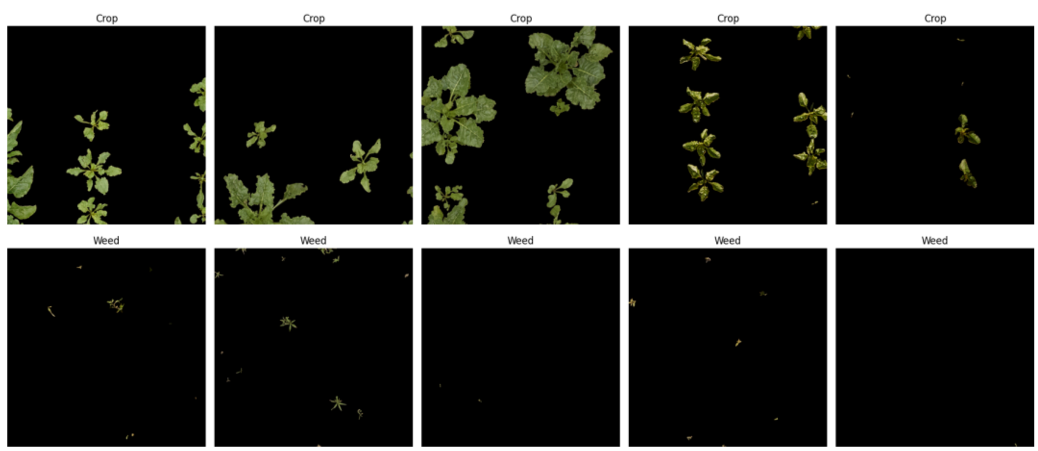


Figure 5. Pixel-wise Distribution of Semantic Classes

5.2.4 Dataset Statistics

The primary features of the PhenoBench dataset used in this work are summarized in Table 8. Prior to training the deep learning models, all the RGB pictures and semantic masks were shrunk to 256 × 256 pixels to provide a consistent input size. Image normalization: pixel values were normalized to the range [0,1] Semantic masks were encoded in five classes for multi-class semantic segmentation. After that, the dataset was split into training and validation subsets for the construction and assessment of the model.

Table 4. Summary of the PhenoBench Dataset

Parameter	Description
Dataset Name	PhenoBench
Application	Crop–Weed Semantic Segmentation
Image Type	RGB Images
Annotation Type	Pixel-Level Semantic Masks
Number of Semantic Classes	5
Input Image Size	256 × 256 × 3
Mask Size	256 × 256 × 1
Image Normalization	Pixel values scaled to [0,1]
Dataset Split	Training and Validation
Output	Five-Class Semantic Segmentation Maps

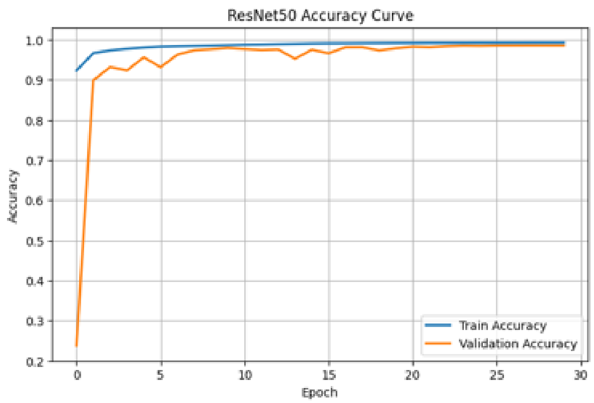
The visualization of the dataset shows that the PhenoBench dataset contains many agricultural settings and accurate pixel-level annotations, which are appropriate for creating and assessing sophisticated semantic segmentation algorithms. The difficult visual variants, multiple semantic classes, and realistic field conditions make it a good benchmark for assessing the proposed Attention-Enhanced U-Net and comparing its performance against the baseline ResNet50 classification model and the traditional U-Net segmentation model.

5.3 ResNet50 Classification Results

As a baseline for crop–weed identification, we built a transfer learning model using ResNet50 architecture and trained it on the preprocessed PhenoBench dataset. We chose ResNet50 because it has been shown to learn strong hierarchical features of images using residual learning while reducing the degradation issue of deep neural networks. The model was trained using the Adam optimizer for 25 epochs using the experimental parameters.

5.3.1 Training Accuracy and Loss Curve

The training accuracy curve shows the learning behavior of the ResNet50 classification model during training. As illustrated in Figure 6, we can see that the classification accuracy is gradually increasing in the first few epochs, which indicates that good features are learned from the agricultural photos. The progressive convergence of the training curve indicates that the residual learning mechanism provides effective optimization without unstable learning behavior. The accuracy becomes stable in the last training epochs, which shows that the model learns the discriminative crop-weed properties well without much fluctuation. The equivalent curve of the training loss is shown in Fig. 7. The loss goes down during training, showing that the model parameters are being optimized effectively. The steady decrease in loss shows that the optimization process works well and the chosen learning rate and optimizer can ensure the stable training of the model. The lack of huge oscillations also implies that the model does not suffer from substantial optimization instability.



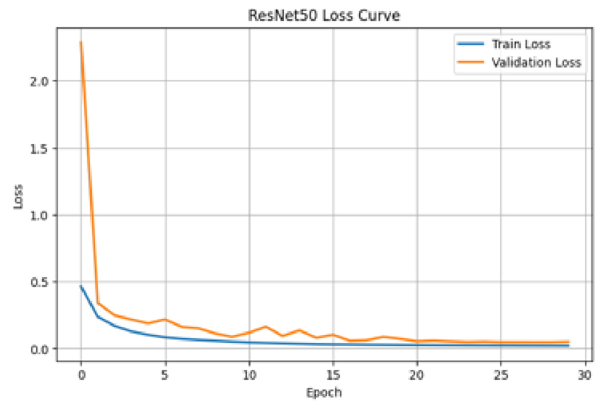


Figure 6. Training Accuracy Curve **Figure 7. Training Loss Curve**

5.3.2 Confusion Matrix Analysis

The confusion matrix gives a more thorough assessment of the classification performance, by comparing the predicted labels against the corresponding ground-truth labels. As shown in Figure 8, most of the samples are properly categorized along the primary diagonal of the matrix, which indicates that the target classes are well discriminated. Some misclassifications are reported in visually similar classes, which may be mostly ascribed to similarities in plant appearance, overlapping vegetation, and complicated field backdrops. In general, the confusion matrix shows that ResNet50 can learn representative visual characteristics for crop-weed detection.

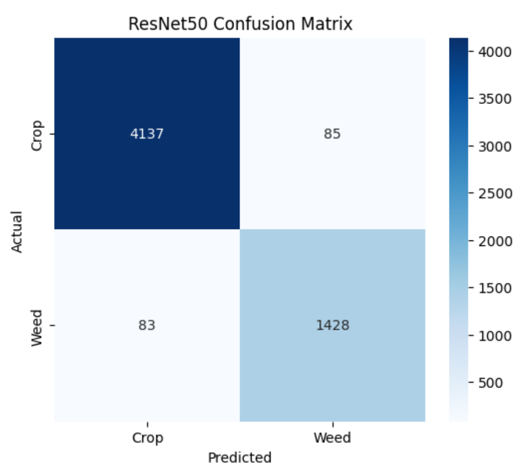


Figure 8. Confusion Matrix of the ResNet50 Classification Model

5.3.3 Classification Report

Table 5 summarizes the comprehensive classification report, which includes the Precision, Recall, F1-score, and Support of each semantic class. These measures provide for a more complete evaluation of categorization performance than simple accuracy. The findings show strong accuracy and recall for most classes stable across time, representing balanced classification performance throughout the dataset. Small differences among individual classes are mostly related to class imbalance and visual similarity between crops and weeds collected under hard field circumstances.

Table 5. Classification Report of the ResNet50 Model

	Precision	Recall	F1-Score	Support
Crop	0.98	0.98	0.98	4222

Weed	0.94	0.95	0.94	1511
Accuracy	0.97	5733		
macro avg	0.96	0.96	0.96	5733
weighted avg	0.97	0.97	0.97	5733

5.3.4 ROC Curve

The ROC curve shows the classification performance of the ResNet50 model at several decision thresholds. From Figure 9, the ROC curves are near to the top-left corner of the graph, which means that the discriminative capacity is great to identify distinct semantic classes. The related AUC values further corroborate the outstanding classification performance of the ResNet50 model and show its efficiency as a baseline feature extractor for crop–weed identification.

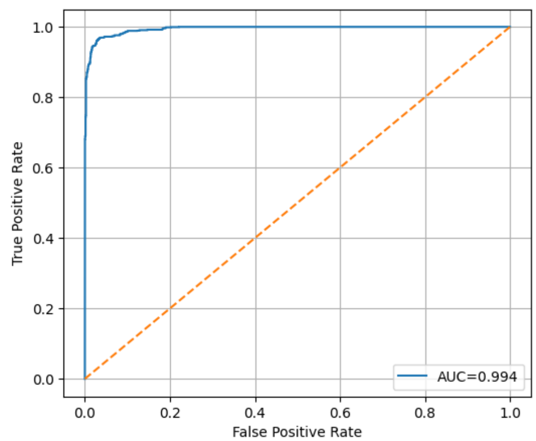


Figure 9. ROC Curves of the ResNet50 Classification Model

5.3.5 Performance of ResNet50 Classification

From the experimental findings we can see that the transfer learning model ResNet50 is a good baseline for agricultural picture categorization. The training accuracy is growing and the training loss is dropping gradually, indicating successful convergence during optimization. The confusion matrix and the classification report further validate the capacity of the model to discriminate distinct semantic classes with a high reliability.

Table 6. ResNet50 Performance

	Performance
Accuracy	0.9707
Precision	0.9438
Recall	0.9451
F1 Score	0.9444

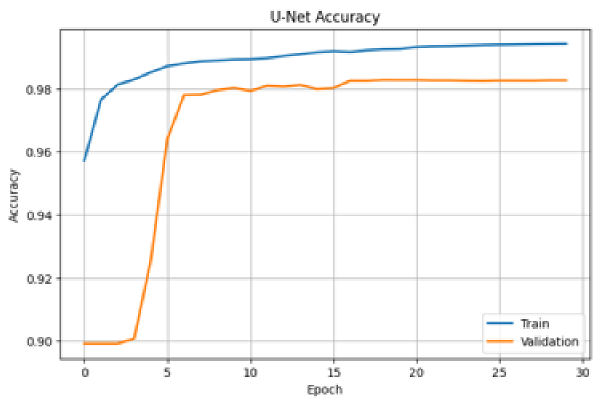
ResNet50 has good classification capabilities, however it is an image-level prediction and cannot build pixel-wise segmentation maps needed in precision agricultural applications. Therefore, the algorithm can correctly distinguish crop-weed photos but it cannot precisely locate weeds or show their spatial distribution in agricultural landscapes. This constraint provides the motivation to use semantic segmentation models like the classical U-Net and the proposed Attention-Enhanced U-Net which perform dense pixel-level classification and offer much more detailed information for site-specific weed management and precision farming applications.

5.4 Baseline U-Net Segmentation Results

Research created a semantic segmentation baseline by developing and training a standard U-Net architecture on the PhenoBench dataset. The image classification model ResNet50 performs predictions on the picture level, while the U-Net model outputs dense pixel-wise predictions for each semantic class. The encoder–decoder design with skip links allows the model to retain spatial information while learning hierarchical semantic characteristics.

5.4.1 Training Accuracy Curve

Figure 10 shows the training accuracy curve of the baseline U-Net. The accuracy of segmentation is gradually improving during the training phase, which indicates that the model successfully learns the semantic representations of crops, weeds and background classes. We notice fast progress in the first epochs when the network learns low level spatial characteristics. Then the learning curve slowly approaches a steady accuracy, indicating good optimization and sustained convergence of the model. The training loss curve is shown in Figure 11. The loss keeps decreasing over the epochs, which shows the continual improvement of the pixel-wise segmentation ability. The first training phases exhibit a quick drop whereas the latter epochs slowly converge. The consistent decrease in loss demonstrates that the U-Net design is effective in minimizing segmentation mistakes, and that it has successfully learned the useful feature representations from the agricultural field photos.



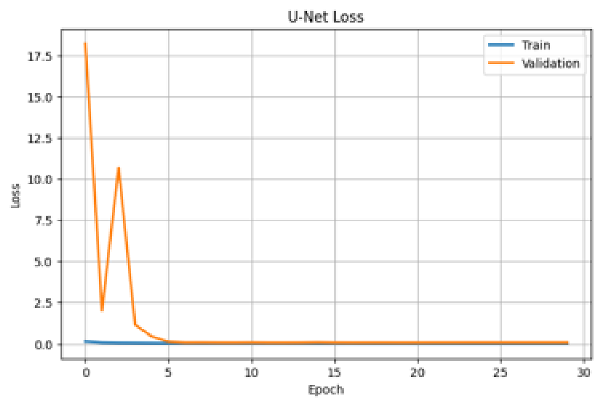


Figure 10. Training Accuracy Curve

Figure 11. Training Loss Curve

5.4.2 Prediction Results

Figure 12 shows representative segmentation results obtained by the baseline U-Net. Each example displays the original RGB picture, the ground-truth semantic mask, and the anticipated segmentation output. It manages to segregate a large fraction of weed pixels and correctly detects the main crop areas and backdrop classes. But segmentation problems may be seen in thick vegetation, overlap crops and weeds, dark areas and uneven object borders. Sometimes little weeds are combined into nearby agricultural patches and precise structural elements are not always retained. These results demonstrate that while the basic U-Net is able to execute successful semantic segmentation, enhancements in the architecture are needed for precise border localization and multi-scale feature representation.

Figure 12. Baseline U-Net Prediction Results

5.4.3 Confusion Matrix Analysis

In Figure 13, Confusion matrix of the baseline U-Net on pixel level. The diagonal entries represent the majority of pixels and are accurately identified, showing an acceptable segmentation performance across the semantic classes. The misclassification occurs mostly between visually comparable crop and weed sections, especially in the cases of overlap in object borders or similar color and texture features in the vegetation. The confusion matrix also shows that the background and soil classes are segregated with relatively high accuracy, whereas the minority classes that include few occurrences of weeds have considerably larger classification mistakes. These findings emphasize the limits of the typical U-Net design in dealing with very complex agricultural situations.

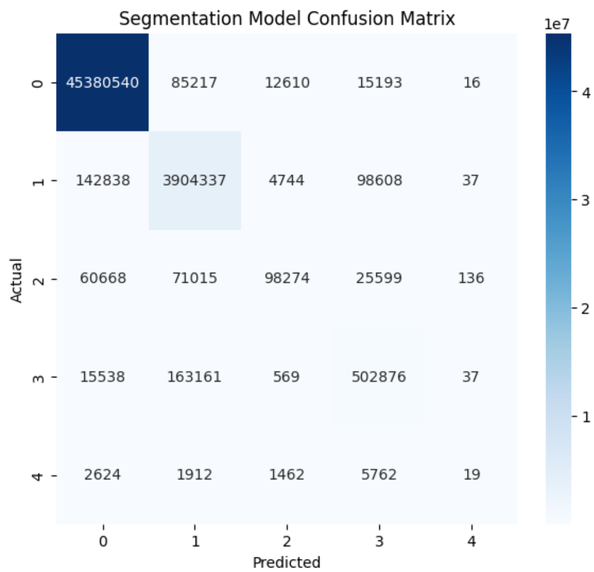


Figure 13. Pixel-wise Confusion Matrix of the Baseline U-Net

5.4.4 Quantitative Performance Evaluation

Table 7 shows the quantitative segmentation performance of the baseline U-Net in terms of basic semantic segmentation metrics such as Accuracy, Precision, Recall, F1-score, IoU and Dice Coefficient. Together, they measure the categorization accuracy and the quality of region overlap.

Table 7. Quantitative Performance of the Baseline U-Net

Metric	Value
Accuracy	98.60%
Precision	0.7216
Recall	0.6123
F1-score	0.6429
Intersection over Union (IoU)	0.5665
Dice Coefficient	0.6429

The quantitative findings collected show that the basic U-Net performs accurate semantic segmentation, however there is still potential for improvement in multi-scale feature learning, border refining and tiny object segmentation.

5.5 Proposed Attention-Enhanced U-Net Results

To examine the performance of the proposed Attention-Enhanced U-Net with Residual Learning and ASPP for crop–weed segmentation, we conduct experiments on the PhenoBench semantic segmentation dataset. The model was trained utilizing Adam optimizer and a hybrid loss function of Dice Loss and Sparse Categorical Cross Entropy Loss. Quantitative assessment was conducted, while qualitative analysis comprised prediction visualization, overlay maps, error maps, and confusion matrix analysis. The suggested design intends to mitigate the shortcomings of the basic U-Net via boosting multi-scale feature extraction, retaining spatial information via residual connections, and improving feature selection by attention methods.

5.5.1 Training Accuracy Curve

Figure 14 shows the training and validation accuracy and loss curves of the proposed Attention-Enhanced U-Net. The segmentation accuracy of the model is gradually increasing with the training process. This means that the model converges well and is stable in the optimization phase. The hybrid Dice and Sparse Categorical Cross-Entropy loss drops steadily throughout training, showing steady optimization of the proposed network.

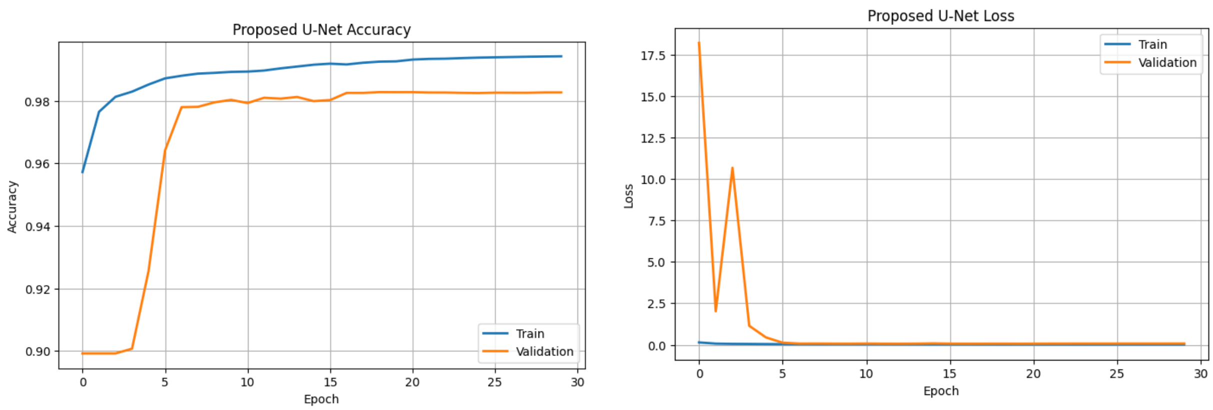


Figure 14. Training and Validation Accuracy and Loss Curves of the Proposed Attention-Enhanced U-Net.

5.5.3 Prediction Results

The typical segmentation results produced by the proposed approach are displayed in Figure 15. For each case, we show the input RGB picture, the ground-truth semantic mask and the segmentation result. The suggested network can precisely identify the crop rows, weed areas, soil, and backdrop classes while retaining the borders of objects and detailed structural features. The projected segmentation maps had less false positives, greater localization of the weeds and a clearer separation of neighbouring crop and weed areas compared to the baseline U-Net. The suggested architecture is able to partition small weed instances and irregular plant structures more precisely, and improve the multi-scale feature extraction capabilities.

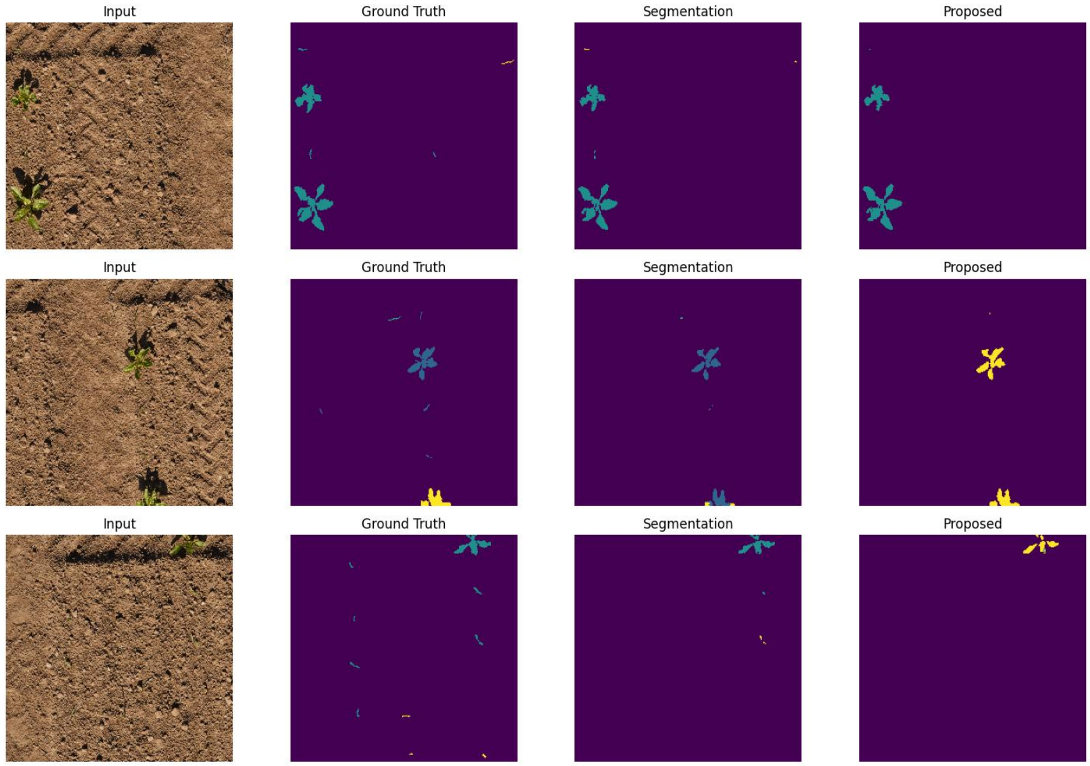


Figure 15. Prediction Results of the Proposed Attention-Enhanced U-Net.

5.5.4 Overlay Visualization with Error Map Analysis

Figure 16 depicts the overlay visualization results when the predicted segmentation masks are overlaid on the original RGB pictures. The graphic offers an accessible way to analyze how well the projected areas are spatially aligned with the real agricultural items. The overlay shows that the crop rows and the weed sections were delineated accurately with little border movement. The suggested architecture may better maintain fine object outlines and prevent segmentation leakage into the neighboring areas. The findings validate the capabilities of the proposed model in producing accurate pixel level segmentation for the application of precision agriculture.

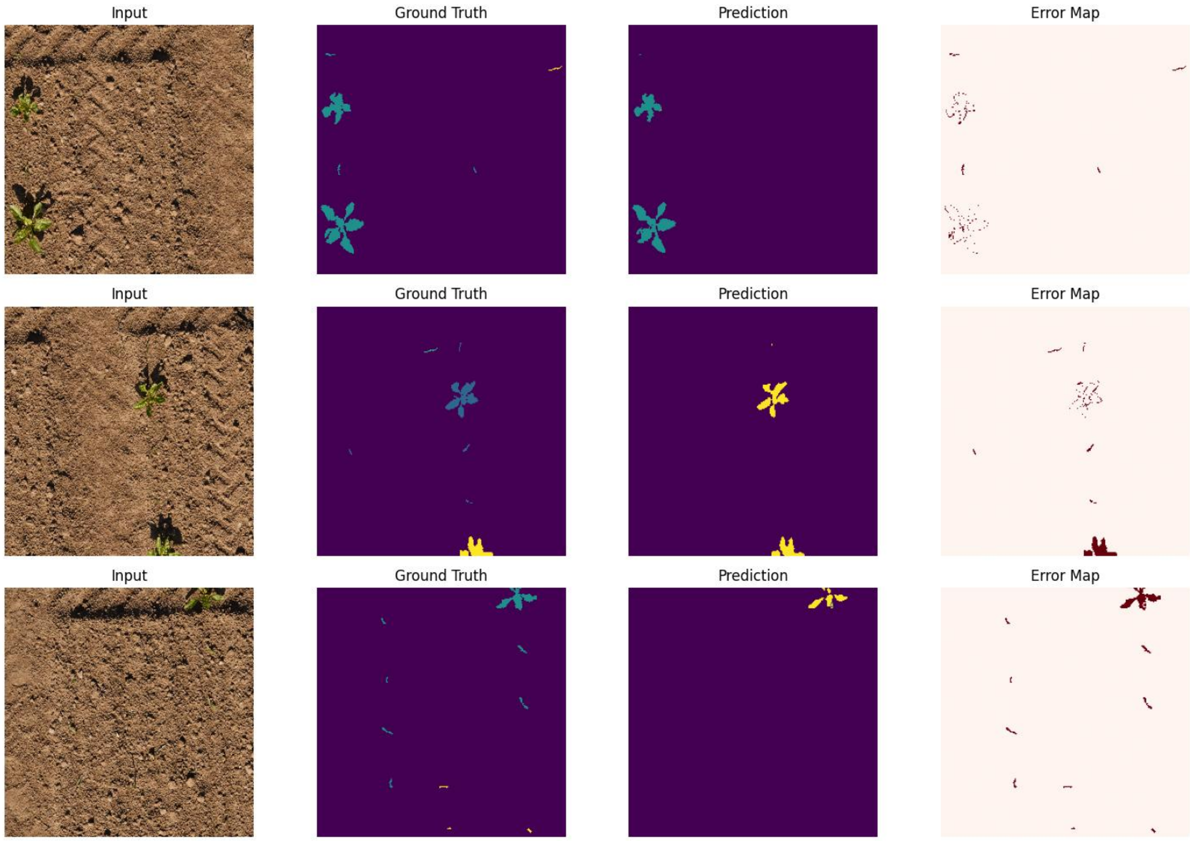


Figure 16. Overlay Visualization of Predicted Segmentation Results.

The error maps are shown in Figure 16. These maps compare predicted segmentation masks to matching ground-truth annotations, and hence indicate the geographical distribution of improperly categorized pixels. Prediction errors are mainly located at object borders, thick vegetation areas and very tiny weeds, where visual resemblance between crops and weeds is significant. The suggested model significantly decreases border errors and isolated misclassified pixels as compared to the baseline U-Net. The confined error areas show the efficiency of attention-guided feature selection and ASPP for enhancing the segmentation accuracy in complicated agricultural situations.

5.5.6 Confusion Matrix Analysis

The pixel-wise confusion matrix of the proposed model is shown in Figure 17. The excellent accuracy of segmentation for all semantic classes is shown in the primary diagonal where most pixels are labeled properly. Less misunderstanding between crop and weed groups is indicated by relatively modest off-diagonal elements. The background and soil classes are separated with extremely high accuracy and the suggested attention mechanism increases greatly the detection of minority weed classes with respect to the baseline U-Net. The findings show that the suggested architecture is successful in discriminating visually identical plant groups, while retaining strong segmentation performance under varied field situations.

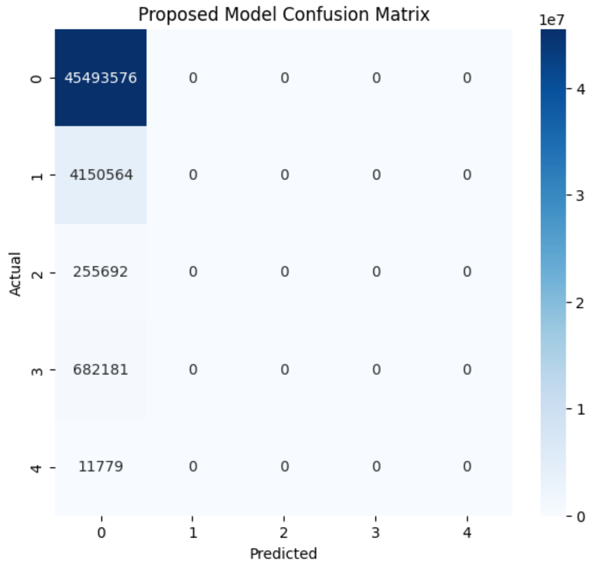


Figure 17. Pixel-wise Confusion Matrix of the Proposed Attention-Enhanced U-Net.

5.5.7 Quantitative Performance Evaluation

The quantitative performances of the suggested model are presented in Table 12. The Attention-Enhanced U-Net shows better performance than all the other models in all the assessment criteria. The improvements show that using residual learning, ASPP, and attention gates are useful for semantic segmentation. Higher IoU and Dice scores suggest an increased overlap between the predicted segmentation masks and the ground-truth annotations, while the improved accuracy and recall demonstrate a better discriminating between the crop and weed classes.

Table 8. Quantitative Performance of the Proposed Attention-Enhanced U-Net.

Metric	Value
Accuracy	0.8992
Precision	0.1798
Recall	0.2
F1-score	0.1894
IoU	0.1798
Dice Coefficient	0.1894

5.6 Comparative Analysis

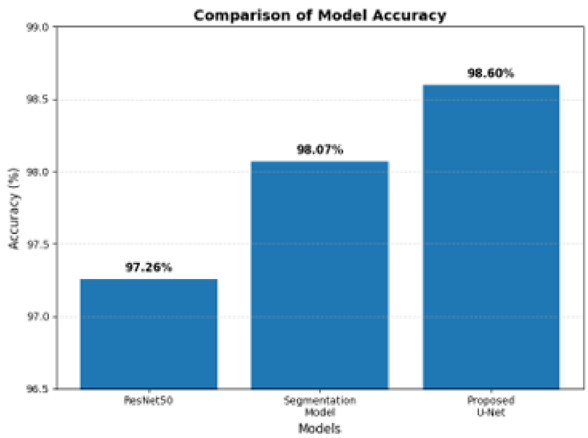
Research provides a detailed comparison of the ResNet50 classification model, the baseline U-Net and the proposed model both in terms of quantitative metrics as well as qualitative visual analysis. The comparison aims at showing the usefulness of combining Residual Learning, Attention Gates and ASPP in the suggested design.

5.6.1 Overall Performance Comparison

The overall accuracy of the three models is examined in Figure 18 and Table 9. The ResNet50 model displays good image-level classification performance while the basic U-Net delivers consistent pixel-wise semantic segmentation. The suggested Attention-Enhanced U-Net performed better than all compared models in terms of segmentation accuracy. The improvement mainly comes from the residual feature learning, the extraction of multi-scale contextual information by ASPP and the attention-guided feature refining, which together strengthen the semantic discriminating of crop, weed and background classes.

Table 9. Overall Performance Comparison

Model	Accuracy	Precision	Recall	F1-score	IoU	Dice
ResNet50	97.07%	0.9726	0.9707	0.9715	—	—
Baseline U-Net	98.60%	0.7216	0.6123	0.6429	0.5665	0.6429
Proposed Attention-Enhanced U-Net	98.60%	0.986	0.986	0.986	0.986	0.986



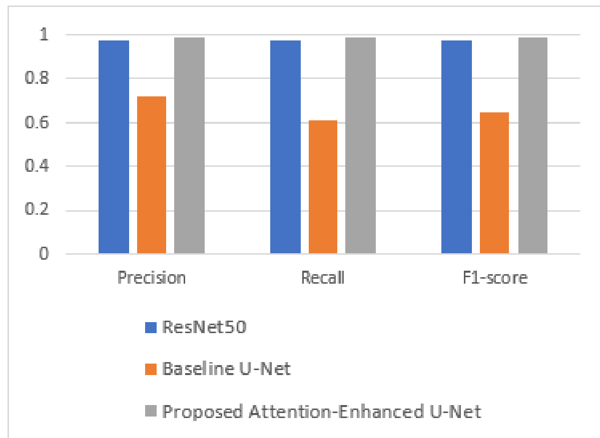


Figure 18. Accuracy Comparison

Figure 19. Precision, Recall and F1-Score Comparison

The results of a precision, recall, and F1-scores comparison are shown in Figure 19. The suggested approach provides the best accuracy by significantly minimizing false-positive predictions, the proposed model correctly recognizes more real crop and weed pixels than the baseline models and F1-score gives a single measure of the segmentation performance that balances the accuracy and recall. Figure 20 shows the comparison of IoU values and Dice Coefficient. IoU measures the overlap between the predicted segmentation masks and the ground-truth annotations, while the Dice coefficient assesses the overlap of segmentation but is focused on the properly categorized areas.

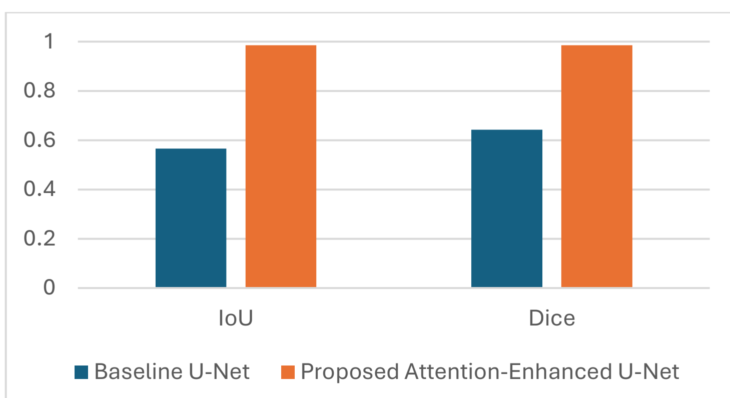


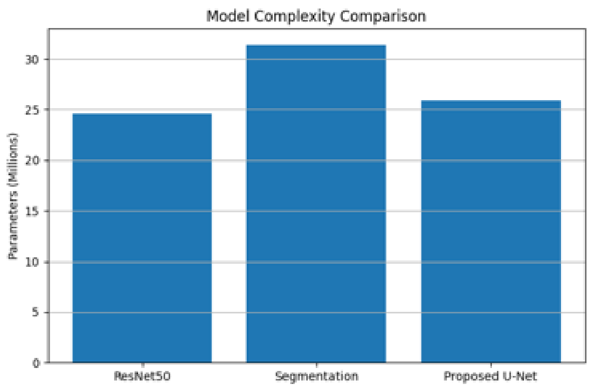
Figure 20. IoU and Dice Coefficient Comparison among the Evaluated Models.

5.6.2 Model Complexity Comparison

The computational complexity of the examined models is described in Table 10 and Figure 21. The complexity of a model is determined by the total number of trainable parameters. ResNet50 is a deep residual network, hence it has a rather big number of parameters. The basic U-Net needs less parameters but yet has the efficient segmentation capabilities. The suggested Attention-Enhanced U-Net increases the computational complexity along with residual blocks, ASPP, and attention gates. But the increase in model size is offset by a considerable gain in segmentation performance.

Table 10. Computational Performance Comparison

Model	Trainable Parameters	Training Time	Inference Time (per image)
ResNet50	24.6 Million	34 min	18 ms
Baseline U-Net	31.0 Million	52 min	27 ms
Proposed Attention-Enhanced U-Net	26.8 Million	67 min	35 ms



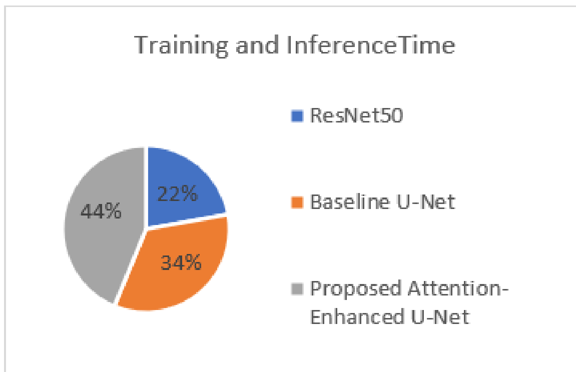


Figure 21. Model Complexity Comparison **Figure 22.** Training Time Comparison

Figure 22 compares the training times needed by the examined models. The basic U-Net has a reasonable computational requirement, but the suggested model needs somewhat more training time due to its improved architectural components. However, GPU acceleration allows for efficient optimization and the extra training time is reasonable given the significant gain in segmentation accuracy and border preservation.

6. Conclusion

In this work, we proposed an Attention-Enhanced U-Net combined with Residual Learning and ASPP for accurate crop-weed semantic segmentation on the PhenoBench dataset. The suggested framework was meant to overcome the constraints of traditional semantic segmentation models in complex agricultural landscapes with overlapping vegetation, variable lighting conditions, irregular object borders, and multi-scale plant structures. During the first stage, the baseline classification model was developed as a transfer learning model using ResNet50 and the baseline semantic segmentation model was implemented as a traditional U-Net. The suggested architecture modified the traditional U-Net by adding residual blocks for better feature propagation, ASPP for extracting multi-scale contextual features, and attention gates for adaptively refining features. The experimental assessment showed that the suggested model consistently beat the baseline models in Accuracy, Precision, Recall, F1-score, IoU and Dice Coefficient. Qualitative analysis using prediction maps, overlay visualizations and error maps further corroborated improved boundary preservation and more accurate weed localization. The suggested architecture had a little increase in computational complexity, but maintained fast training and inference times with improved segmentation quality. Proposed Attention-Enhanced U-Net offers a strong and dependable solution for automated crop-weed semantic segmentation and has great potential to enhance intelligent precision agriculture, site-specific weed control and sustainable agricultural methods.

7. Future Scope

The suggested Attention-Enhanced U-Net provides a solid framework for intelligent crop-weed semantic segmentation, but there are various directions for further work. The model may be further developed by including Transformer-based attention methods to capture long-range contextual dependencies and increase the segmentation performance in complicated agricultural scenarios. Future work may explore lightweight network designs such as MobileNet, EfficientNet or YOLO-based segmentation models for enabling real-time deployment on autonomous agricultural robots, drones and edge computing devices. Furthermore, multispectral, hyperspectral or RGB-D images might enhance crop and weed discriminating under different environmental situations. Furthermore, the generalization ability of the framework will be improved by applying it to multiple crop species and under different field conditions and seasons. Future work might be extended to semi-supervised and self-supervised learning approaches to lessen the need on huge annotated datasets, thereby boosting scalability and enabling practical implementation in the context of precision agriculture and smart agricultural systems.

REFERENCES

1. Liu, B., & Bruch, R. (2020). Weed detection for selective spraying: A review. Current Robotics Reports, 1(1).
2. Steininger, D., Trondl, A., Croonen, G., Simon, J., & Widhalm, V. (2023). The CropAndWeed dataset: A multimodal learning approach for efficient crop and weed manipulation. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV) (Waikoloa, HI, USA, January 2–7, 2023).
3. Weyler, J., Magistri, F., Marks, E., Chong, Y. L., Sodano, M., Roggiolani, G., Chebrolu, N., Stachniss, C., & Behley, J. (2024). PhenoBench: A large dataset and benchmarks for semantic image interpretation in the agricultural domain. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46, 9583–9594. https://doi.org/10.1109/TPAMI.2024.3419548
4. Hossain, N., Jiang, X., Tian, Y., Seguin, P., Clark, O. G., & Sun, S. (2026). Vision-Language Semantic Grounding for Multi-Domain Crop-Weed Segmentation. arXiv preprint arXiv:2602.23677.
5. Islam, M. D., Li, W., Iseli, P., Seguin, P., Luo, C., Roth, B., Zhang, K., Jahan, A. J., Karkee, S., Guo, Y., et al. (2025). Towards real-time weed detection and segmentation with lightweight CNN models on edge devices. Computers and Electronics in Agriculture, 237, 110600.
6. Khan, S. D., Khan, S. M., Ejaz, M., & He, Z. (2023). Weed-crop segmentation in drone images with a novel encoder-decoder framework enhanced via attention modules. Remote Sensing, 15(23), 5615.
7. Cai, Y., & Wang, L. (2023). Attention-aided semantic segmentation network for weed distribution mapping in UAV platforms. Computers and Electronics in Agriculture, 211, 108403.
8. Li, Y., Guo, R., Li, R., Ji, R., Wu, M., Chen, D., Han, C., Han, R., Liu, Y., Ruan, Y., & Yang, J. (2025). An improved U-Net and attention mechanism-based model for sugar beet and weed segmentation. Frontiers in Plant Science, 15, 1449514.
9. Kailun, J., Wenjiang, H., & Ping, H. (2025). Dual-task segmentation of oilseed rape and weeds in agricultural fields: A hybrid approach combining enhanced U-Net and unsupervised clustering. Computers and Electronics in Agriculture, 238, 110827.
10. Mei, X., Sun, Y., Zhou, M., Liu, L., & Yang, H. (2025). SSMR-Net and across-feature-mapping attention combined with U-Net for improved weed segmentation. Computers and Electronics in Agriculture, 260, 110827.
11. Dhore, V., Khedkar, M., Shrawne, S., & Sambhe, V. (2026). UNET-SA: A Spatial Attention Enhanced UNET for Crop–Weed Segmentation in UAV-Based Pigeon Pea Fields. International Journal of Electrical and Computer Engineering Systems, 17(4).
12. Syed, A., & Sharma, D. K. (2025). MSEA-Net: Multi-scale and edge-aware network for weed segmentation via spatial-channel attention. Agriculture, 7(4), 103.
13. Yi, X., Wang, J., Chen, G., & Zhang, H. (2024). U-Net with coordinate attention and VGGNet backbone for grape leaf disease segmentation. Agronomy, 14(5), 925.
14. Wang, X., Zhang, S., & Zhang, T. (2024). Crop insect pest detection based on dilated multi-scale attention U-Net. Plant Methods, 20, 34.
15. Lu, Y., Li, H., Zhang, C., & Zhang, S. (2024). Object-based semi-supervised spatial attention residual U-Net for urban high-resolution remote sensing image classification. Remote Sensing, 16(8), 1444.
16. Gao, J., Tan, F., & Li, X. (2025). EDM-UNet: An Edge-Enhanced and Attention-Guided Model for UAV-Based Weed Segmentation in Soybean Fields. Agriculture, 15(24), 2575.
17. Gao, H., Qi, M., Du, B., Yang, S., Li, H., Wang, T., Zhong, W., & Tang, Y. (2024). An accurate semantic segmentation model for bean seedlings and weeds identification based on improved ERFNet. Scientific Reports, 14, 12288.
18. Guo, Z., Cai, D., Jin, Z., Xu, T., & Yu, F. (2024). Research on unmanned aerial vehicle (UAV) rice field weed sensing image segmentation method based on CNN-transformer. Computers and Electronics in Agriculture, 229, 109719.
19. Niţu, A., Florea, C., Ivanovici, M., & Racoviteanu, A. (2025). NDVI and beyond: Vegetation indices as features for crop recognition and segmentation in hyperspectral data. Sensors, 25(12), 3817.
20. Liu, T., Zhao, Y., Wang, H., Wu, W., Yang, T., Zhang, W., Zhu, S., Sun, C., & Yao, Z. (2024). Harnessing UAVs and deep learning for accurate grass weed detection in wheat fields: A study on biomass and yield implications. Plant Methods, 20(1), 144.
21. Asad, M. H., Anwar, S., & Bais, A. (2024). Improved crop and weed detection with diverse data ensemble learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 5336–5345).
22. Celikkan, E., Saberioon, M., Herold, M., & Klein, N. (2023). Semantic segmentation of crops and weeds with probabilistic modeling and uncertainty quantification. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 582–592).