

An Efficient Deep Learning Based AgriResUpNet Architecture for Semantic Segmentation of Crop and Weed Images



Ali Asgar Hussain*^{ID}, Pramod Sekharan Nair^{ID}

Department of Computer Science and Engineering, Medi-Caps University, Rau Indore 453331, India

Corresponding Author Email: er.ali.asgar.hussain@gmail.com

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.290516>

ABSTRACT

Received: 28 December 2023

Revised: 14 July 2024

Accepted: 1 August 2024

Available online: 24 October 2024

Keywords:

crop and weeds, semantic segmentation, deep learning, AgriResUpNet, U-Net, residual learning

The environment and agricultural output are being challenged by weeds and crops. Innovative advancements in alternative weed control techniques that seek to reduce the need on herbicides have been spurred by the growing demand for sustainable weed control methods. Having weed recognition that is sufficiently successful is a hurdle to the implementation of these methods for selective in-crop application. Deep learning has shown remarkable promise in a variety of vision tasks, leading to the development of several effective image-based weeds and crops identification systems. This study looks at the newest developments in deep learning techniques for pixel-wise semantic segmentation of identifying crops and weeds. The hardest problem is semantic segmentation-based weeds and crops recognition, which needs to be solved for smart farming to work well. The goal is to give each pixel of a picture its own class name. Deep learning for smart farming has a number of useful applications, one of the most important of which is identifying the exact position of crops and weeds on farms. There are already quite complex systems for separating weeds and crops, with millions of factors that need more time to train. To get around these problems, we suggest an AgriResUpNet design based on deep learning. It is a carefully put together mix of the U-Net and residual learning frameworks. To see how well the suggested model works on the crop and weed GitHub dataset in terms of pixel accuracy, precision, f1-score, and IoU measures. The suggested network is tested and compared with other cutting-edge networks using a crop and weed dataset that is open to the public. Also, the AgriResUpNet model gave an IoU of 97.58% for weeds and 94.82% for crops. In addition, the data show that the AgriResUpNet can easily find both crops and weeds. This suggests that this design is better for finding weeds early in the growing season. It was shown in experiments and comparisons that the suggested network does better than existing designs in intersection over union (IoU) and F1-score.

1. INTRODUCTION

As the global population continues to rapidly expand, there is a corresponding rise in the prevalence of food insecurity around the globe. It is necessary to raise agricultural yields in order to address the shortage of food that is now occurring. An increase in crop output is being achieved by the utilisation of smart technology in contemporary agricultural methods such as precise agriculture, smart farming, food technology, plant breeding, and others. Smart agriculture is a form of agriculture that incorporates artificially intelligent technologies with the purpose of making intelligent decisions in order to maximise crop productivity [1, 2]. A number of factors, including but not limited to plant diseases, irrigation systems, the use of agrochemicals, insect infestations, and weeds, are among the most significant factors that influence crop output. Over the duration of the period between 2003 and 2014, weeds alone were responsible for an economic loss of around 11 billion USD across 18 states in India. The yield loss can be reduced by up to fifty percent or more with the use of automated weed control and management technologies [3-5].

For any agricultural crop, the most significant sources of damage are weeds and pests [6, 7]. According to the study by Fennimore et al. [8], there are a number of conventional ways that are utilised in order to manage the growth of weeds and pests in order to achieve good yields. These approaches have a number of drawbacks, the most significant being that are the pollution of the surroundings and the poisoning of the crops, both of which have negative consequences for human health. As a result of the development of more advanced technology, robots have lately been utilised for selective spraying, which targets just weeds and does not cause any damage to crops [9-11]. According to the study of Lottes et al. [12], the most significant obstacle that these autonomous platforms must overcome is the identification of the precise position of crops as well as weeds. The ability of these robots to identify weeds and separate them from crops is one of the most important uses of deep learning in the field of smart farming [13-15]. Nevertheless, in order to automate the agricultural machinery [16, 17], researchers must first find solutions to a number of issues, such as classification, tracking, detection, as well as segmentation. In order to do this, we provide here a model for

the identification of crops and weeds that is oriented on semantic segmentation [18].

Semantic segmentation focuses on identifying each pixel's classification by pixel-by-pixel (dense) predictions of a picture. Semantic segmentation often makes use of fully convolution networks (FCNs) [19, 20], that were initially proposed by Long et al. [21]. These investigations frequently used an encoder-decoder approach, in which the encoder employs convolutions to create a latent representation of an image, while the decoder concentrates on upsampling the latent representation in relation to the original picture size to make dense predictions. With favourable outcomes for medical pictures, U-Net was suggested by expanding the decoder's capability [22]. Furthermore, SegNet [23], used pooled indices for its decoder, whereas other encoder-decoder architectures simply used the pooled values for non-linear upsampling to maintain boundary data. Using global context as a better foundation for pixel-level predictions, pyramid scene parsing network (PSPNet) as well as different-region-based contextual aggregation via a pyramid pooling module [24] took use of this capacity. DeepLab models used atrous convolutions in place of a traditional encoder-decoder method in order to minimise downsampling operations while maintaining a wide receptive field [25].

Through the use of smart agriculture farming, weed control is of utmost significance in order to enhance production and decrease the amount of herbicide pollution. The main contribution of this paper is to show how to improve performance in the presence of crop and weed images features by introducing different deep learning algorithms. In this respect, algorithms based on deep learning have come to attract an increasing amount of interest for the purpose of crop and weed segmentation in agricultural settings. These algorithms have shown promising outcomes. This research presents the implementation of the U-Net with Residual learning called AgriResUpNet' network, a state-of-the-art convolutional neural network (CNN) technique that has only been employed in precision agriculture on a very seldom basis. The purpose of this implementation was to perform semantic segmentation of weed photos. After that, we evaluated performance of the model in comparison to that of the U-Net method using a number of different criteria like accuracy, precision, recall, F1-score and Loss. This research contribution of this work as:

- Data was collected using the GitHub dataset, which included both the original photos and crop/weed masks. Using OpenCV, color-based segmentation was used to define accurate thresholds, and grayscale masks were created. To assess the accuracy of the segmentation, Exploratory Data Analysis (EDA) made use of bounding box samples and overlay masks.

- A unique color-based segmentation approach was presented for the purpose of accurately identifying crop and weed regions. Implemented efficient procedures for the preparation of data, such as preprocessing, normalization, and splitting the dataset.

- An architecture called AgriResUpNet was developed, which combines residual learning with U-Net for the purpose of performing robust agricultural image segmentation.

- A combination of U-Net and residual learning is used to extract features in order to model AgriResUpNet. The encoder is responsible for collecting low-level details, the bridge is responsible for boosting expression, and the decoder is responsible for recovering spatial information in order to

achieve correct segmentation.

- Accomplished exceptional performance metrics, which contributed to the development of applications for accuracy growing. AgriResUpNet achieved excellent levels of accuracy, IoU, F1 Score, and Precision in its weed and crop segmentation, demonstrating remarkable performance in this area.

The rapid progress in agricultural technology has required the creation of strong techniques for precise and efficient crop and weed picture segmentation. Conventional methods frequently prove inadequate in managing the intricacies of varied agricultural settings. To address these difficulties, we provide AgriResUpNet, a cutting-edge deep learning framework specifically developed for the semantic segmentation of crop and weed pictures.

AgriResUpNet utilises a residual learning framework in conjunction with an upsampling network to greatly improve the accuracy of segmentation and processing efficiency. Our architecture combines advanced convolutional neural networks (CNNs) with residual blocks, resulting in enhanced feature extraction and preservation, unlike traditional techniques. The upsampling components enhance the segmentation outputs, allowing for accurate demarcation of boundaries between crops and weeds. AgriResUpNet's main advantages lie in its capacity to handle high-resolution photos and its resilience in various environmental conditions, making it crucial for practical agricultural uses. Furthermore, AgriResUpNet has exceptional efficiency in terms of processing speed and precision when compared to current models, rendering it a great asset for contemporary precision agriculture. This research highlights the capacity of deep learning to transform agricultural practices by offering a scalable and efficient solution for crop management and weed control.

2. LITERATURE SURVEY

An area of study that has been receiving a lot of attention is the classification and detection of smart agricultural farming areas for the purpose of recognizing crops and weeds. In this part, we have mentioned a few of the research that have been conducted about the identification and categorization of weeds through the use of AI-based approaches such as deep learning, machine learning, computer vision, robotics, and other similar methods.

For the purpose of expanding upon this study article, a number of works in the disciplines of neural networks, artificial intelligence, as well as precision farming were examined and analyzed. In the field of agriculture, CNNs have been utilized to find solutions to a wide range of issues. A model that is based on deep learning and has the ability of recognizing 26 distinct illnesses in 14 different crop species was proposed by Mohanty et al. [26]. This model was developed with the purpose of distinguishing healthy plants from sick ones. An accuracy of classification that was more than 99% was achieved by authors through the use of pre-trained AlexNet and GoogleNet on a dataset consisting of 54,306 individuals. An approach that makes use of pre-trained Inception-v3 architecture was given by Teimouri et al. [27], for the purpose of estimating the species of weeds and the phases of their growth. Using their suggested model, they are able to estimate the number of leaves with a level of accuracy that is seventy percent.

According to the study of Dyrmann et al. [28], DetectNet was utilized in order to determine the locations of weeds in leaf-occluded crops. For the purpose of identifying weeds in cereal fields, their network was trained using 17,000 annotations of photos of weeds. Although the system has a detection accuracy of 46% when it comes to weeds, it is unable to identify overlapping weeds or weeds that are tiny. A CNN-based approach was presented by dos Santos Ferreira et al. [29] in order to recognize weeds as well as categorize them as either grass or broadleaf. This was done in order to allow for the selection of herbicides for soybean crops. In the study by Lottes et al. [12], stem identification was accomplished through the use of a sliding window-based technique. Each local window gives information about the position of the stem or an area that does not include a stem. In order to collect location data regarding weeds for site-specific weed management, Ma et al. [30] presented a dataset and carried out tests on a SegNet-based encoder-decoder network (via transfer learning) for semantic segmentation. The results of these studies reached a mean average accuracy of up to 92.7%.

For the semantic segment [31], researchers compare two deep learning frameworks—the ResNet versus the fully convolutional network—to determine how accurate the weed as well as crop segment is. As part of the case study, forty plant as well as weed photos were culled from an open repository. The findings demonstrate that both structures achieve a global accuracy of over 90% in the verification phase package.

In the study by Zhuang et al. [32], research has been planned to identify seedlings of broadleaf weeds in wheat fields. Because their recall remains at 58% or lower, they have determined that FR-CNN, YOLOv3, VFNet, TridentNet, as well as CenterNet are unfit for identification. In contrast, F1-

scores more than 95% have been achieved with classification using VGGNet as well as AlexNet. The models utilised for training in this work were trained using a dataset with a relatively tiny image resolution of 200×200 px.

Saqib et al. [5] have put out a deep learning weed identification algorithm that can be utilised effectively for agricultural weed control. You Only Look Once (YOLO), an object identification technique centred on Convolutional Neural Networks, is used for both training and prediction in the model recommend. The gathered information includes RGB pictures of four distinct weed species: California poppy, Creeping thistle, bindweed, as well as grass. With an average loss of 1.8 and a mean average accuracy value of 73.1%, it has successfully forecasted 98.88% of weeds.

For training deep learning model [33], a publicly available, open-source Sugarbeets dataset was utilised. To test method, picked 1,300 pictures of plants and weeds. They obtained a 93% accuracy rate in our model between crops and weeds using a hybrid technique that utilises semantic segmentation as well as the YOLO (You Only Look Once) algorithm.

In the study of Fathipoor et al. [34], semantic segmentation of weed photos was accomplished by the implementation of the U-Net++ network, that is a state-of-the-art convolutional neural network (CNN) method. This technique has only been employed in precision farming on a very seldom basis. In terms of overall accuracy, intersection over union (IoU), recall, as well as F1-score measures, the results demonstrate that the U-Net++ performs better than the standard U-Net. In addition, the U-Net++ model offered a cannabis IoU measurement of 65%, whereas the U-Net model offered a weed IoU measurement of 56%.

Table 1. Summary of related work

Reference	Methodology	Key Findings	Limitations	Future Work
[26]	Pre-trained AlexNet, GoogleNet	Detecting 26 illnesses in 14 different crop species with a 99% success rate	Only identified 26 diseases in 14 crop species	In future used various number of crops images, also new Transfer learning model
[27]	Pre-trained Inception-v3	Accuracy in weed species and development stage estimations of 70%	Only identified 26 diseases in 14 crop species	In future used various number of crops images, also new Transfer learning model
[28]	DetectNet	Concerns with overlaps and a 46% success rate in weed detection	Unable to identify weeds that are both small and overlapping	Improve detection accuracy, handle overlapping weeds
[30]	SegNet based encoder-decoder network	Automatic semantic SSWM segmentation with an average accuracy of 92.7%	Limited to the provided dataset	Recognise weeds more accurately and deal with overlapping ones
[31]	Fully convolutional network, ResNet	Global weed and crop segment accuracy of 90% or higher	achieved only 90% accuracy	Use novel transfer learning methods; investigate various datasets; evaluate adaptability
[32]	AlexNet, VGGNet	Achieving classification F1-scores > 95% for broadleaf weed seedling identification	Small image resolution dataset used (200×200 px)	Increase recall by delving into higher-resolution datasets
[5]	YOLO (You Only Look Once)	Accuracy in weed detection at 98.88% and average precision of 73.1 percent	Precision measure perform very poor	In future add enhance version of YOLO Model
[33]	Semantic segmentation, YOLO	An impressive 93% success rate in differentiating between crops and weeds	YOLO model not perform well	Improve the YOLOv5 Model in the future
[34]	U-Net++	Uses accuracy metrics better than conventional U-Net	Only used single model	Additional investigation of U-Net++ for precision farming
[35]	FCN, Unet, Fast-SCNN, SegNet	Differentiating between weeds and crops using segmentation techniques	Range of acc, Unet is at 87.5% and SegNet at 92.08%)	Revise models to achieve higher precision
[18]	CED-Net	Beats DeepLabv3, FCN-8s, SegNet, U-Net, and SegNet	Compared to others, the proportion of total parameters utilised is lower	Investigate ways to optimise parameters for pre-existing models

In the study by Kumar et al. [35], a technique called semantic segmentation was used to distinguish between weeds and crops, employing models like FCN, Unet, Fast-SCNN, and SegNet. Unet, an improved form of CNN, achieved an accuracy rate of 87.5%. Meanwhile, SegNet reached an average accuracy rate of 92.08%.

Khan et al. [18] distinguished between crops and weeds using a cascaded encoder-decoder network, including CED-Net, as a semantic segmentation approach. In terms of the intersection over union (IoU), F1-score, sensitivity, true detection rate, and average precision contrast metrics, the suggested network surpasses state-of-the-art architectures like U-Net, SegNet, FCN-8s, and DeepLabv3. This is achieved by using fractions of total variables that are only $1/5.74 \times$ U-Net, $1/5.77 \times$ SegNet, $1/3.04 \times$ FCN-8s, and $1/3.24 \times$ DeepLabv3, respectively.

Table 1 shows the summary of related work for classification and detection of weed and crop.

3. SECTION HEADINGS

The following Figure 1 shows the proposed methodology flowchart for the classification and semantic segmentation of crop and weeds using deep learning models. The suggested methodology begins with the collecting of agricultural photographs that have been tagged from a particular dataset. This is then followed by a thorough preprocessing step that makes use of a color-based segmentation method to create binary masks that differentiate between crop and weed regions. As part of the Exploratory Data Analysis (EDA), masks are superimposed on the source pictures, and bounding box samples are generated for visual examination. Data preparation for training includes fundamental operations like establishing input dimensions, converting data to NumPy arrays, and constructing consistent binary masks. Each of these procedures is necessary for the training process. The AgriResUpNet architecture is the core of the technique. It is a carefully created combine of the U-Net as well as residual learning structures, and it was developed to provide greater accuracy in feature extraction and semantic segmentation. To ensure that there is no confusion, the hyperparameters and architecture details have been stated unambiguously. The model is trained and tested, and it produces remarkable results in the segmentation of crops and weeds, demonstrating its potential for use in precision agricultural applications.

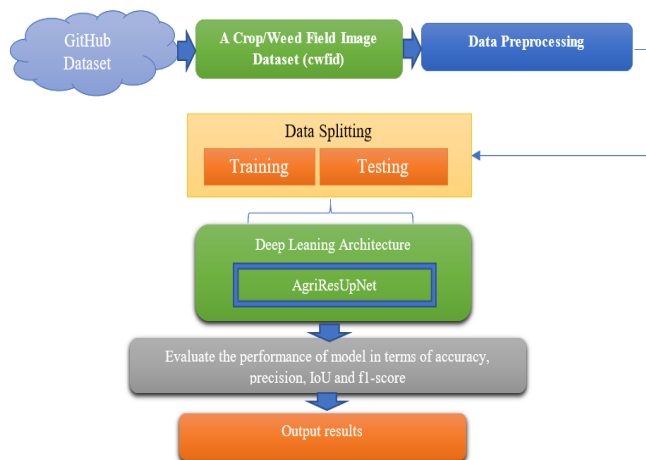


Figure 1. Proposed flowchart

3.1 Data collection

For the purpose of data collecting, a dataset that is accessible at <https://github.com/cwfid/dataset> is utilised. This dataset is comprised of agricultural photos that have been annotated. The original photos are 3755808 pixels in size and contain a data type of uint8, which means they have three colour channels. Their dimensions are (966, 1296, 3). Furthermore, the photos' crop and weed masks are identical in size, shape, and data type. There are 60 photos in the dataset; 40 will be used for training and 20 will be used for testing. For validation reasons, four photos are reserved in the training set. The next phases in the deep learning-based crop and weed segmentation process are built upon this extensive dataset. The following Figures 2 and 3 shows the input images of weed and crop dataset with sample of images and training images.

The dataset used for training AgriResUpNet comprises a diverse collection of crops and weed images sourced from various agricultural fields, ensuring a wide range of environmental conditions and crop types. It includes high-resolution RGB images of different crops and weeds, captured under varying lighting and weather conditions to enhance the model's robustness. The dataset also contains annotated examples with pixel-wise labels for both crops and weeds, highlighting different growth stages and overlapping scenarios. This variability ensures the model can generalize well across different field conditions and accurately segment crops and weeds in complex agricultural landscapes.

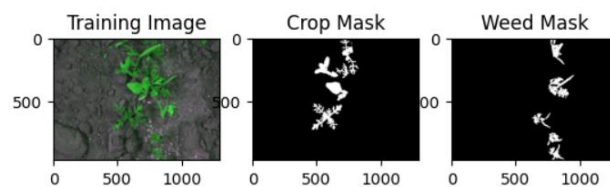


Figure 2. Input images of training, and crop/weed mask

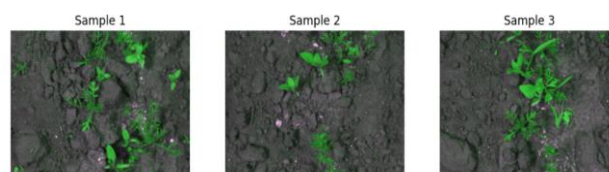


Figure 3. Sample images of input dataset

3.2 Data preprocessing

One crucial step in our technique during the pre-processing stage is to use a color-based segmentation technique. This technique is used to generate binary masks for crop and weed regions from annotated agricultural photos. At this crucial phase, it is imperative to isolate and identify precise regions of interest within the images. We utilise the Open CV library to import both the image and its related annotation mask. The binary matrices, crop mask and weed mask, are initialised to indicate the lack of crop and presence of weed, respectively.

Advancing to our color-based segmentation utilizes distinct color channels for accurate discrimination between crop and weed regions. We assume that crop regions are predominantly represented by the blue channel, while weed regions are characterized by the green channel. Parameters such as B Threshold and G Threshold are introduced to establish precise

color thresholds for effective segmentation. Pixels surpassing the blue threshold and falling below the green threshold are identified as crop, while those surpassing the green threshold and falling below the blue threshold are recognized as weed. This meticulous approach ensures the creation of highly accurate binary masks, serving as a foundational step for subsequent image analysis tasks. The flexibility to adjust threshold values enhances the adaptability of our agricultural image segmentation methodology to diverse color representations in annotation masks, thereby improving robustness and versatility.

The convert to grayscale function, integral to our methodology, transforms annotated color masks into a binary format, aiding agricultural image analysis. Utilizing Open CV, it efficiently converts multi-channel masks to grayscale, simplifying subsequent processing. This step enhances mask interpretability, contributing to the robustness of our segmentation methodology.

3.3 Exploratory Data Analysis (EDA)

To better understand the agricultural images identified with metadata, we use two essential methods in the Exploratory Data Analysis (EDA) stage. To begin, we impose masks on top of the original images, giving a perspective from above that emphasises the spatial relationship between the marked masks and the real items in the farm settings. To evaluate the precision and efficacy of the segmentation technique, this visual aid is crucial.

Figure 4 shows the bar graph of classes distribution with number of pixels. In figure x-axis shows the number of classes that is crop and weed while y-axis shows the number of pixels.

Figure 5 displays a graph that illustrates the frequency of width and height distribution and shows the width and height distribution graph with frequency. In figure x-axis shows the width and height of images and while y-axis shows the frequency of images. Blue bar shows the width and green bar shows the Height.

Figure 6 displays exemplary photos that illustrate overlay masks used in the study.

Figure 7 shows some examples of bounding boxes that have been placed around crops and weeds.

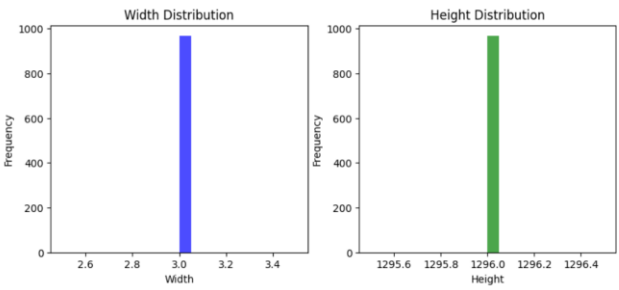


Figure 5. Width and height distribution graph with frequency



Figure 6. Sample images of overlay masks

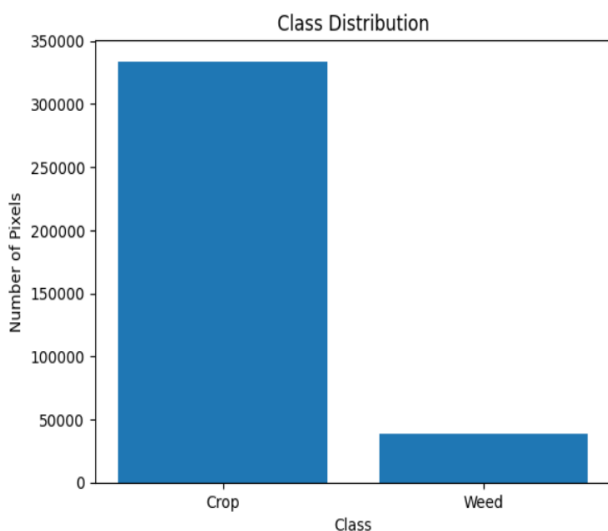


Figure 4. Bar graph of classes distribution with pixels

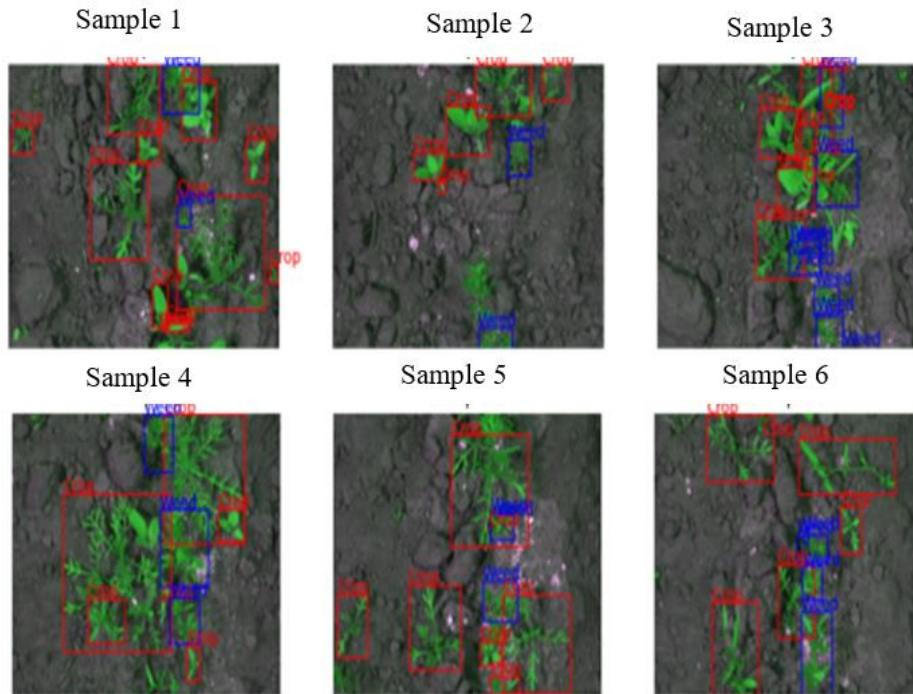


Figure 7. Sample images of bounding box from crop and weed

3.4 Preparing data for training

In the preprocessing stage, the code performs vital operations to prepare the data for effective model training. By examining the dimensions of first image in training dataset, input dimensions for subsequent model configurations are determined. This essential step ensures compatibility between the model architecture and the input data.

Subsequently, code converts lists containing image and mask data into NumPy arrays. This conversion enhances computational efficiency during training, as NumPy arrays are optimized for numerical operations. Notably, the pixel values of the images are normalized to the range $[0, 1]$, a common practice in DL to facilitate convergence.

The masks representing crop and weed regions are processed to create a unified binary mask. This binary mask serves as the ground truth for the model, allowing it to learn to differentiate between crop and weed areas in the agricultural images.

To evaluate the model's performance during training, the dataset is split into training and validation sets using the `train_test_split` function data into 90:10 ratio. This ensures that the model is trained on a diverse subset of the data and evaluated on a separate subset, providing a reliable metric of its generalization capabilities. Overall, these preprocessing steps lay the foundation for a robust and effective training process, crucial for the success of the subsequent semantic segmentation model.

3.5 Modeling

The AgriResUpNet architecture is meticulously crafted to address the challenges posed by agricultural image segmentation tasks. It combines the strengths of both the U-Net and residual learning frameworks to achieve superior feature extraction and semantic segmentation accuracy.

In the encoder, the module initiates the processing by applying a convolutional operation followed by a residual

block. This initial step sets the tone for the subsequent feature extraction process, effectively capturing low-level features. The encoder further refines the feature hierarchy through four residual blocks, each reducing spatial dimensions via strided convolutions. This multi-scale feature representation ensures the model's ability to discern objects of varying sizes in the input image.

The bridge module acts as a bottleneck, enhancing feature expression before transitioning to decoder. It consists of two convolutional blocks, consolidating both high and low-level features. The decoder then employs upsample and concatenation blocks to recover spatial information and combine features from the encoder. This hierarchical integration is crucial for accurate segmentation, particularly when dealing with complex agricultural landscapes.

The residual blocks within the decoder refine the feature maps, facilitating the precise localization of objects. The incorporation of skip connections ensures seamless transfer of information between encoder and decoder, aiding in the reconstruction of high-resolution semantic maps. Batch normalization and rectified linear unit (ReLU) activation functions are applied throughout the architecture to stabilize and activate feature representations, respectively.

The final convolutional layer, with a sigmoid activation function, produces a binary mask indicating the probability of the presence of the target objects. This pixel-wise classification allows the model to discern between crop and weed regions, enabling fine-grained segmentation.

The AgriResUpNet architecture is tailored for input images of shape $(966, 1296, 3)$. The choice of these s aligns with the requirements of agricultural images, where the larger spatial resolution demands an intricate understanding of the field's composition. The model is trained with a binary cross-entropy loss function and optimized using the Adam optimizer.

This architecture excels in capturing contextual information, enabling it to make informed decisions about object boundaries and locations. The synergy of residual connections, skip connections, and multi-scale feature extraction positions

AgriResUpNet as a powerful tool for precise agricultural image segmentation. Its technical robustness and efficiency make it well-suited for deployment in precision farming applications, contributing to advancements in sustainable agriculture through improved crop monitoring and weed management. The following Table 1 shows the hyperparameter tuning of proposed models AgriResUpNet.

3.5.1 Implementation details of AgriResUpNet

Encoder Module

The encoder consists of a series of convolutional layers followed by residual blocks. Each block reduces spatial dimensions and increases the depth of feature maps.

Components

(1) Convolutional Layer: Initial convolution to capture low-level features.

(2) Residual Blocks: Each block contains convolutional layers, batch normalization, and ReLU activations. These blocks also have skip connections to add input directly to the output.

```

Pseudocode for Encoder:
def encoder(input_tensor):
    x = Conv2D(filters=16, kernel_size=(3, 3),
padding='same')(input_tensor)
    x = BatchNormalization()(x)
    x = ReLU()(x)

    filters = [32, 64, 128, 256]
    for f in filters:
        x = residual_block(x, f)
        x = Conv2D(f, (3, 3), strides=(2, 2),
padding='same')(x) # Reducing spatial dimensions
        x = BatchNormalization()(x)
        x = ReLU()(x)
    return x

def res_block(x, filters):
    shortcut = x
    x = Conv2D(filters, (3, 3), padding='same')(x)
    x = BatchNormalization()(x)
    x = ReLU()(x)
    x = Conv2D(filters, (3, 3), padding='same')(x)
    x = BatchNormalization()(x)
    x = Add()(shortcut, x)
    x = ReLU()(x)

Pseudocode for Decoder:
def decoder(input_layer, encoder_features):
    filter_sizes = [256, 128, 64, 32, 16]

    for i, size in enumerate(filter_sizes):
        input_layer = UpSampling2D(size=(2,
2))(input_layer)
        input_layer = Concatenate()(input_layer,
encoder_features[i])
        input_layer = residual_block(input_layer, size)

    return input_layer.

```

In order to enhance the description of the experimental design, we included numerous essential particulars. The dataset was partitioned into three subsets: training, validation, and testing, with a split ratio of 70:15:15, respectively. This

guarantees a thorough assessment of the model's performance and aids in avoiding overfitting. Data augmentation techniques were extensively utilised to improve the model's ability to generalise. The employed approaches encompassed random rotations, horizontal and vertical flips, zooming, and colour jittering, effectively simulating diverse real-world settings and augmenting the variety of training data. In order to tackle the possible imbalance in class distribution, particularly due to the different frequencies of crops and weeds in the dataset, we employed a variety of methods. Initially, the technique of oversampling was applied to the minority classes while under sampling was applied to the dominant classes in order to achieve a more equitable distribution. In addition, class weights were incorporated into the loss function to impose a greater penalty on misclassifications of the minority class. This encourages the model to prioritise the underrepresented classes and allocate more attention to them. These techniques prevented the model from exhibiting bias towards the majority class, hence enhancing its capability to reliably distinguish between crops and weeds in various settings. The implementation of an all-encompassing strategy for managing and enhancing datasets played a crucial role in attaining exceptional performance metrics for AgriResUpNet.

Table 2. Hyper-parameters details of AgriResUpNet

Architecture	Modified U-Net with Residual Blocks (AgriResUpNet)
Input Shape	(966, 1296, 3)-Image dimensions (height, width, channels)
Encoder Filters	[16, 32, 64, 128, 256]
Activation Function	ReLU (Rectified Linear Unit)
Batch Normalization	Applied after each convolutional layer
Skip Connections	Implemented through residual blocks for both encoding and decoding paths
Output Activation	Sigmoid (for binary segmentation)
Loss Function	Binary Cross entropy
Optimizer	Adam
Learning Rate	Default value used by Adam optimizer
Metrics	Binary Accuracy
Upsampling Method	Bilinear interpolation used in the upsampling blocks
Input Normalization	Image pixel values scaled between 0 and 1
Epochs	100
Batch size	1

Table 2 provides a comprehensive overview of the hyperparameters used in AgriResUpNet. These include a modified U-Net architecture with residual blocks, ReLU activation function, Adam optimiser, and binary cross-entropy loss function.

4. PROPOSED SEMANTIC SEGMENTATION MODELS

The AgriResUpNet architecture is meticulously crafted to address the challenges posed by agricultural image segmentation tasks. It combines the strengths of both the U-Net and residual learning frameworks to achieve superior feature extraction and semantic segmentation accuracy.

4.1 U-Net

For semantic segmentation, the U-Net architecture—a kind of Fully Convolutional Networks (FCN)—is the most common technique. For biological image segmentation and localization, Ronneberger et al. [22] developed the original U-Net architecture. The capacity to offer localization and categorization in U-Nets' output makes them superior to conventional CNNs [36-38]. Here, "localization" refers to assigning a specific class to each pixel in a picture. Another advantage is that U-Nets can produce more accurate segmentations with less training photos than FCNs can. Making use of upsampling layers with several feature channels allows for the transfer of contextual data to higher resolution layers, which is how this is accomplished. For this specific application, U-Net segmentation networks are the best option since they are large and easily scalable, they are not dependent on localization or contextual data, and they're employed by cutting-edge approaches for semantic segmentation of plant images.

Two sections make up the U-net network, as shown already: The first is a contracting route that makes use of a standard CNN [39] design. Two 3×3 convolutions, an activation unit called ReLU, and a max-pooling layer make up each block of the contracting route. This pattern is iterated several times. Part two, the extended route, is where U-net really shines; at every level, it uses 2×2 up-convolution to upsample the feature map. After that, the upsampled feature map is superimposed with the cropped and joined feature map from the equivalent layer in the contracting route. After that, ReLU activation and two 3×3 convolutions occur in quick succession. In the last step, we apply one more 1×1 convolution to get the feature map down to the necessary number of channels and then we have the segmented picture. Edge pixel characteristics contain the least amount of contextual information as well as must be deleted, which is why cropping is essential. In addition to creating a network with a u-shape appearance, this method also propagates contextual data across the network, enabling it to partition items in a given region by drawing on context from a broader overlapping area. Figure 8 shows the general design of the U-net.

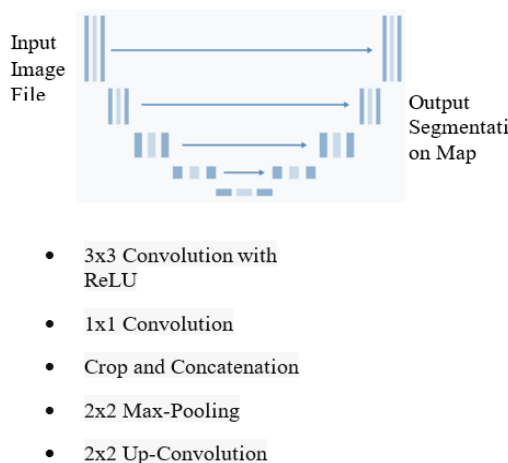


Figure 8. U-net architecture [40]

Design of the U-net, seen in Figure 2. These arrows are meant to indicate the various operations, the blue boxes are meant to represent the feature map at each layer, and the grey boxes are meant to represent the cropped feature maps that

were obtained via the contracting path.

Given is the energy function that is associated with the network:

$$E = \sum w(x) \log(p_{k(x)}(x)) \quad (1)$$

where, p_k is pixel-wise SoftMax function used over final feature map, definite as:

$$P_{ik} = \frac{\exp(a_{ik}(x))}{\sum_{k'} \exp(a_{ik'}(x))} \quad (2)$$

P_{ik} represents the probability of activation in channel k at pixel location i . $a_{ik}(x)$ denotes the activation in channel k at pixel location i for the input x . The denominator is the sum of the exponential function of the product of a constant "a" and a function "k" evaluated at "x", over the sum of the variable "k". The expression is written as $\exp(a_{ik})$. The value of (x) is obtained by summing the exponential activations of all channels k' at pixel location i , and then normalising the output to create a probability distribution across the channels.

4.2 Residual learning

Deep neural networks have been shown to have a very high level of performance on picture classification tasks, despite the fact that they are more challenging to train. When it comes to training deeper neural networks, it often requires a significant amount of time and more computer capacity. This is because of the number of variables as well as the vanishing gradient problem. Deep residual networks, also known as ResNets, have the potential to accelerate the process of training and achieve higher levels of accuracy in comparison to its counterpart neural networks. This enhancement is accomplished by ResNets with the incorporation of a straightforward skip link that is parallel to the layers of convolutional neural networks [41, 42]. In contrast to traditional neural networks, ResNets train residual functions while keeping the layer inputs in mind. Residual nets allow the stacked layers to match a residual mapping instead of hoping they directly fit a desired underlying mapping. We allow the stacked layers to suit a residual mapping instead of expecting they immediately match a specified underlying mapping. We allow the stacked nonlinear layers to fit an additional mapping of $F(x) := H(x) \rightarrow x$, where $H(x)$ is the required underlying mapping. By adding x to the original mapping, we get $F(x) + x$. Optimising the residual mapping, rather than the original, unreferenced mapping, should be easier, according to our hypothesis. It would be more convenient to just reduce the residual to zero rather than fit an optimum identity mapping using a stack of nonlinear layers, if that were the case. Using "shortcut connections" in feedforward neural networks, the $F(x) + x$ formula may be accomplished (Figure 9).

The Figure 9 illustrates the notion of "residual learning" in a neural network. The display features a residual block consisting of two layers. The initial layer applies a weight and a Rectified Linear Unit (ReLU) activation function to the input, whereas the subsequent layer applies another weight. A shortcut connection is created by adding the output of the second layer to the original input. This strategy effectively mitigates the issue of the vanishing gradient problem in deep networks. If identity mappings are successful, techniques may adjust weights towards zero in order to achieve mappings that are close to the identity function.

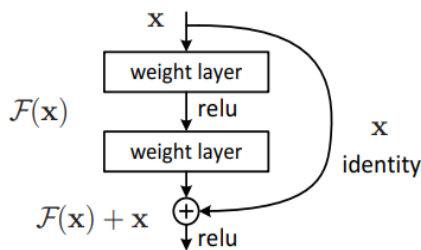


Figure 9. Residual learning: A fundamental component [43]

4.3 Hyperparameters tuning

In deep learning models, hyperparameters are the user-defined variables that regulate the learning procedure. The values of these hyperparameters are established before the model starts learning, and they are utilised to improve the model's learning. Internal configuration variables known as model parameters are learned by the model autonomously. Expertise and a lot of trial and error are needed for the hyperparameter tuning procedure. Setting hyper-parameters (e.g., optimizer, epochs, batch size, activation function, loss function, etc.) is not a straightforward or easy task.

4.3.1 Activation function (ReLU (Rectified Linear Unit) and Sigmoid) of the output layer

There are a few other names for this function: squashing, threshold, transfer, and more. Using the network input and threshold value as inputs, it performs a function and outputs the values that result from the neuron's activation. The neuron's prior activity status and the input from outside the cell determine this. Sigmoid and ReLU are two popular activation functions, while there are others that work better for certain types of problems. Any value between zero and infinity can be taken on by the ReLU activation function. Remaining as an equation, it is:

$$\text{ReLU}(x) = \max(0, x) \tag{3}$$

With a large range of positive values for the input observations xi, this choice is good. Naturally, the ReLU is a weak pick along with the identity function is superior if the input xi may take on negative values.

When doing a binary classification, this activation function is applied at the output layer. The likelihood of an input becoming a member of a class is the result it produces. The sigmoid function σ is capable of taking on any value between 0 and 1. Its formula is the same as before. From a mathematical perspective, it seems like:

$$\text{Sigmoid } \sigma(x) = \frac{1}{1 + e^{-x}} \tag{4}$$

Use of this activation function is conditional on the input observations xi falling within the range of 0, 1 or having been normalised to that range.

4.3.2 Loss function (Binary Cross entropy)

When comparing the actual output with the output predicted by a machine learning model, a loss function may be used to quantify the mistake. throughout model construction, it is possible to provide this parameter, which in turn influences the model's performance throughout training. The task at hand and

the data type at hand dictate the loss function to be employed. For many classification and segmentation activities, the baseline loss function is binary cross entropy, often known as log loss. Binary cross-entropy may be expressed as:

$$L_{BCE} = \sum_x -(T_x \log(P_k) + (1 - T_x) \log(1 - P_x)) \tag{5}$$

Using T as the truth data, T_x as an element of T, and P_x as an element of the network's output prediction mask [44].

4.3.3 Optimizer (Adam)

Adam is a kind of adaptive learning that takes advantage of rate optimisation to make things better. By adding a momentum term to the equation and preserving the element-wise squared value moving average and parameter gradients, it becomes available whenever the variables are altered. To train a deep neural network, one must tweak the model's parameters (such as the learning rate and weights) until one gets the best possible outcome with the least amount of loss (the gap between the actual and expected outputs). This procedure for fine-tuning is known as optimisation. The optimisation functions are the procedures or algorithms that are employed while optimising. For this task, Adam optimiser was utilised to construct the neural network. The steps of Adam's optimisation computation are represented by the equation below.

$$Sdw = \beta_2 - Sdw(1 - \beta_2) * dw^2 \tag{6}$$

$$w = w - \alpha \frac{dw}{\sqrt{Sdw + \epsilon}} \tag{7}$$

Both binary and multi-class optimisation can benefit from this optimizer's ability to reduce loss. Among the optimizer functions, it possesses the quickest convergence time. This is due to the fact that Adam updates the parameters during training at each iteration using three parameters: a weighted average of gradient, a weighted average of squared gradient, and a learning rate.

4.3.4 Batch size

The training dataset in a single batch is taken into consideration by this. The dataset can be separated into many batches, which dictates the number of iterations, because feeding the computer a whole epoch would be too much for the system to handle.

4.3.5 Epoch 100

When all data is input and output from the neural network simultaneously, one epoch is reached. The neural network's weights may be improved by continuously giving it training data. After every loop, the parameter is updated. Typically, accuracy and loss may be improved by increasing the number of epochs.

5. RESULT AND DISCUSSION

In this section provide the experimental results of proposed modes. For this research experiment used Weed and Crop dataset from GitHub that the split into train and test with 60:40 ratio. For the classification and segmentation of crop and weed

data images using DL techniques (AgriResUpNet). Following section provide the implemented results of model evaluation in terms of Accuracy, IOU, F1-score, Precision, and Loss Measures, also provide the experimental setup, experimental results of crop and weed dataset with plotting graph, and tables, and then last provide the comparative analysis between base and proposed models.

5.1 Experimental setup

The experimental results perform on HP workstation equipped with 32GB of RAM, a 1TB hard drive, the Windows 10 operating system, a 24GB Nvidia graphics processing unit (GPU), and an Intel Core i7 CPU. By using this hardware, we used python programming language [45] and Jupyter notebook [46] with some python packages like NumPy, pandas, Sk-Learn, and matplotlib etc. [47].

5.2 Evaluation metrics

This study proposes a segmentation algorithm and uses experimental results to compare its performance to that of other algorithms evaluated using identical evaluation indices. Various assessment metrics, including F1-score, precision, average precision (AP), and intersection over Union (IoU), were used to quantify and evaluate the suggested network's performance. True positive (TP), true negative (TN), false positive (FP), and false negative (FN) were the variables identified by computing the confusion matrix between the prediction and the ground truth in order to compute these metrics. From these bitmaps, TP, TN, FP, and FN rates were determined.

- TP (True positive)-no. of pixels that were incorrectly classified into class.
- TN (True negative)-no. of pixels that should have been classified to specified class but were not.
- FP (False positive)-no. of pixels for which the class was incorrectly classified.
- FN (False negative)-no. of pixels for which the class was not assigned and which ought not to have been.

Consequently, the initial metric employed is the Intersection over Union (IoU) ratio, which yields a value equivalent to dividing the overlap region of the prediction mask by the union region encompassing the target mask. On the basis of the confusion matrix, the following Eq. (8) can be derived:

$$IoU = \frac{TP}{TP + FP + FN} \tag{8}$$

Pixel Accuracy (PA): Pixel Accuracy (PA) is an additional metric utilised to assess the model under consideration. It signifies the proportion of predicted pixels in relation to the total number of pixels. On the basis of the confusion matrix, the Eq. (9) as follows can be derived:

$$PA = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

Precision: Precision indicates the performance of a categorization model during evaluation. Calculates the proportion of accurately classified images relative to the total number of images predicted to be members of a particular class; this is mathematically represented as Eq. (10):

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

Recall: In reference to positive values, this metric is employed to evaluate the predictive efficacy of the model. Eq. (11) mathematically expresses the metric by dividing the number of accurately classified images by the total number of images belonging to a particular class:

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

F1-score: F1-score is the harmonic mean of recall and precision. By utilising the harmonic mean, it becomes highly advantageous to consider both recall and precision. This facilitates the evaluation of both recall and precision in terms of the F1-score. The formula used in mathematics for the score is Eq. (12):

$$F1 \text{ measure} = 2 * \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{12}$$

5.3 Results of weed segmentation using the AgriResUpNet model are presented

The Table 2 shows the proposed AgriResUpNet Model results on Weed Segmentation in terms of accuracy, precision, f1score and IoU measure. Also Figure 10 shows the bar graph of parameter performance of AgriResUpNet model on weed segmentation. AgriResUpNet Model obtain 99.96% accuracy, IoU of 97.58%, F1-score of 98.77%, precision of 98.80% and loss of 0.009%, respectively, on weed Segmentation.

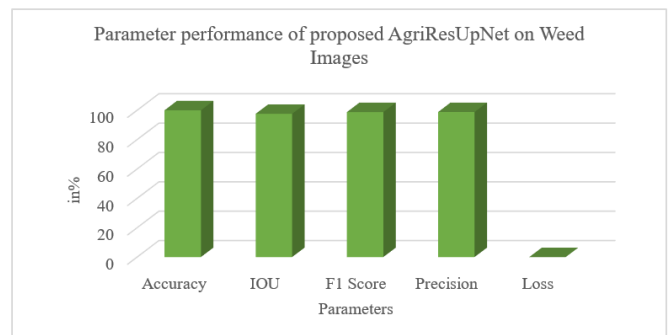


Figure 10. Bar graph of parameter performance of AgriResUpNet model on weed segmentation

The AgriResUpNet model's training and validation curves for weed segmentation are shown in Figure 11. After 100 iterations, the model's accuracy and loss graphs reveal its remarkable convergence, with a validation loss of 0.0552 and a final training loss of 9.6810e-04. The model's impressive accuracy levels—99.17% in validation and 99.96% in training—demonstrate its advanced understanding of the segmentation task for weed.

Further evidence that the model is successful in collecting nuanced segmentation information is provided by the IoU plot, the precision plot, and the F1-score plot, shows in Figure 11. Particularly, the AgriResUpNet model obtains high values for IoU and F1-score, with values of 0.9758 and 0.9877, accordingly, on the training set. Additionally, it obtains a precision of 0.9880. On the validation set, the model continues

to demonstrate competitive performance, obtaining values of 0.5837 for the IoU, 0.7213 for the accuracy, and 0.7072 for the F1-score, correspondingly.

5.4 Results on crop segmentation using AgriResUpNet model

The Table 3 shows the proposed AgriResUpNet Model results on Crop Segmentation in terms of accuracy, precision, f1score and IoU measure. Table 3 presents the performance of the AgriResUpNet model for weed segmentation, showing high accuracy, IOU, F1 Score, precision, and minimal loss.

Table 4 showcases the AgriResUpNet model's crop segmentation performance, indicating high accuracy, IOU, F1 Score, precision, and a low loss value.

Figure 11 shows the training and validation curves for loss and accuracy of the AgriResUpNet model, highlighting its performance during weed segmentation, with convergence indicating effective learning.

Figure 12 illustrates the training and validation curves for IoU, precision, and F1-score of the AgriResUpNet model, demonstrating its effectiveness in accurately segmenting weeds by tracking these key metrics.

Figure 13 shows the bar graph of parameter performance of AgriResUpNet model on Crop segmentation. AgriResUpNet Model obtain 99.84% accuracy, IoU of 94.82%, F1-score of 96.01%, precision of 95.92% and loss of 0.0039%, respectively, on Crop Segmentation.

Figure 14 shows the AgriResUpNet model's training as well

as validation curves for crop segmentation. Having a validation loss of 0.0492 after 100 epochs and a final training loss of 0.0039, the accuracy as well as loss graphs show that the model was well trained. With a training accuracy of 99.84% and a validation accuracy of 99.08%, the model clearly excels at learning the finer points of crop segmentation.

Additional insights may be seen in Figure 15, which displays graphs of IoU, precision, and F1-score. On the training set, the AgriResUpNet model achieves a remarkable 0.9592 accuracy, an IoU of 0.9482, and an F1-score of 0.9601, respectively. Maintaining competitive performance on the validation set, the model achieves IoU of 0.8393, accuracy of 0.8618, and F1-score of 0.9098. These measures have shown steady progress across the training epochs, which shows that AgriResUpNet is very good at properly outlining crop borders and that it is resilient and can generalise well to crop segmentation tasks.

Table 3. Weed segmentation using AgriResUpNet model

Model	Accuracy	IOU	F1 Score	Precision	Loss
AgriResUp Net	99.96	97.58	98.77	98.80	0.0009

Table 4. Crop segmentation using AgriResUpNet model

Model	Accuracy	IOU	F1 Score	Precision	Loss
AgriResUp Net	99.84	94.82	96.01	95.92	0.0039

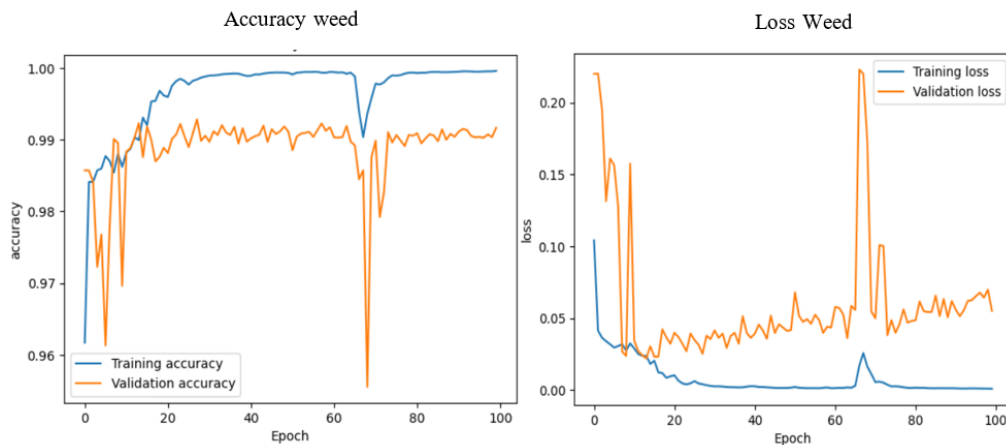
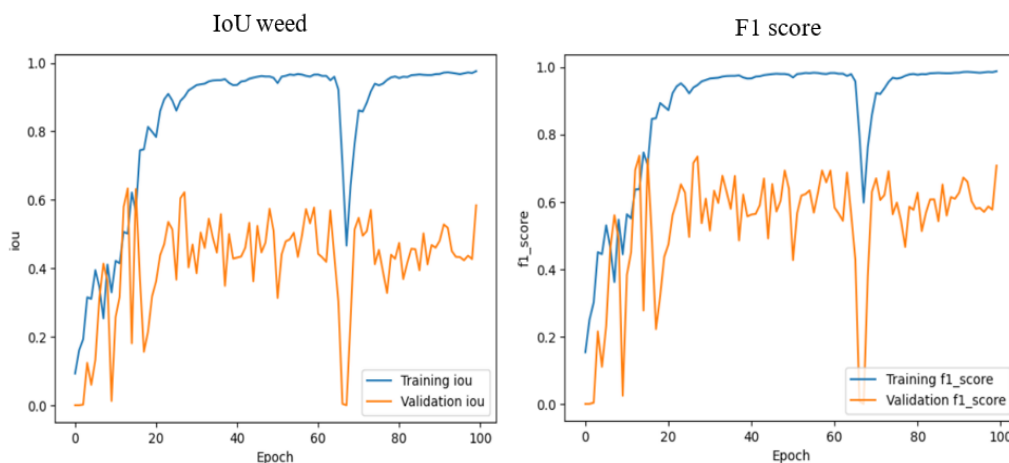


Figure 11. Train/Val plotting curve of loss and accuracy of AgriResUpNet model on weed segmentation



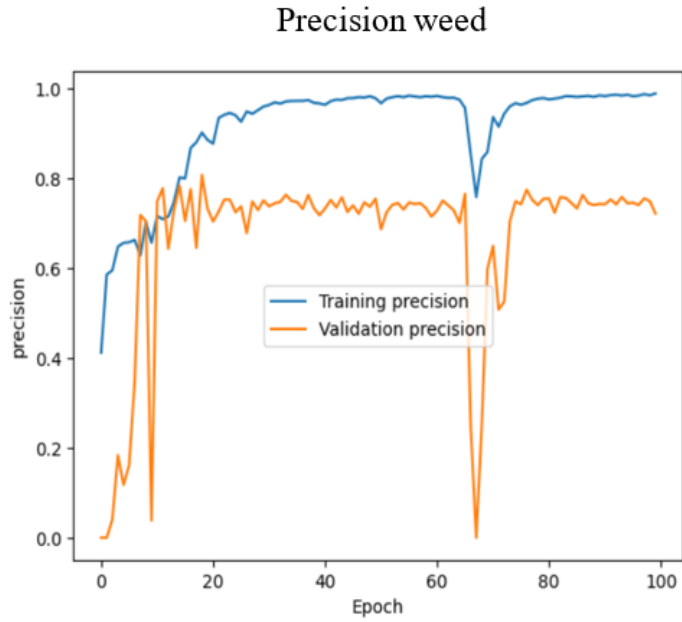


Figure 12. Train/Val plotting curve of IoU, precision and F1-score of AgriResUpNet model on weed segmentation

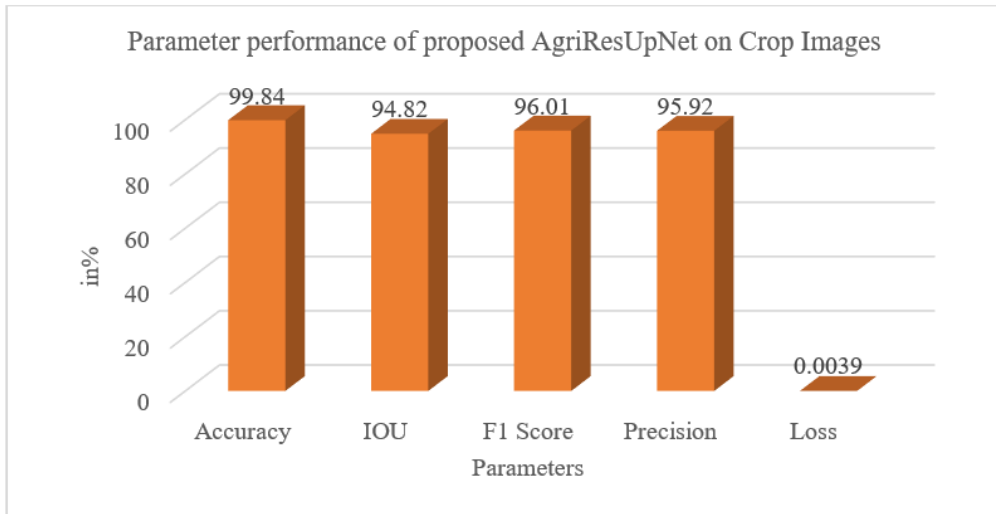


Figure 13. Bar graph of parameter performance of AgriResUpNet model on crop segmentation

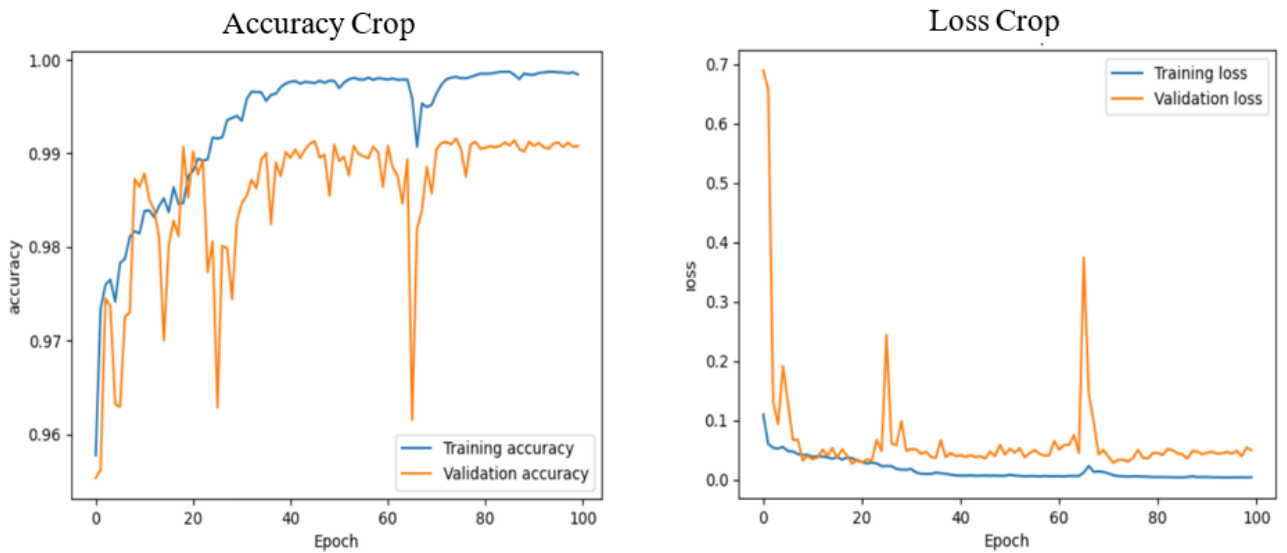


Figure 14. Train/Val plotting curve of loss and accuracy of AgriResUpNet model on crop segmentation

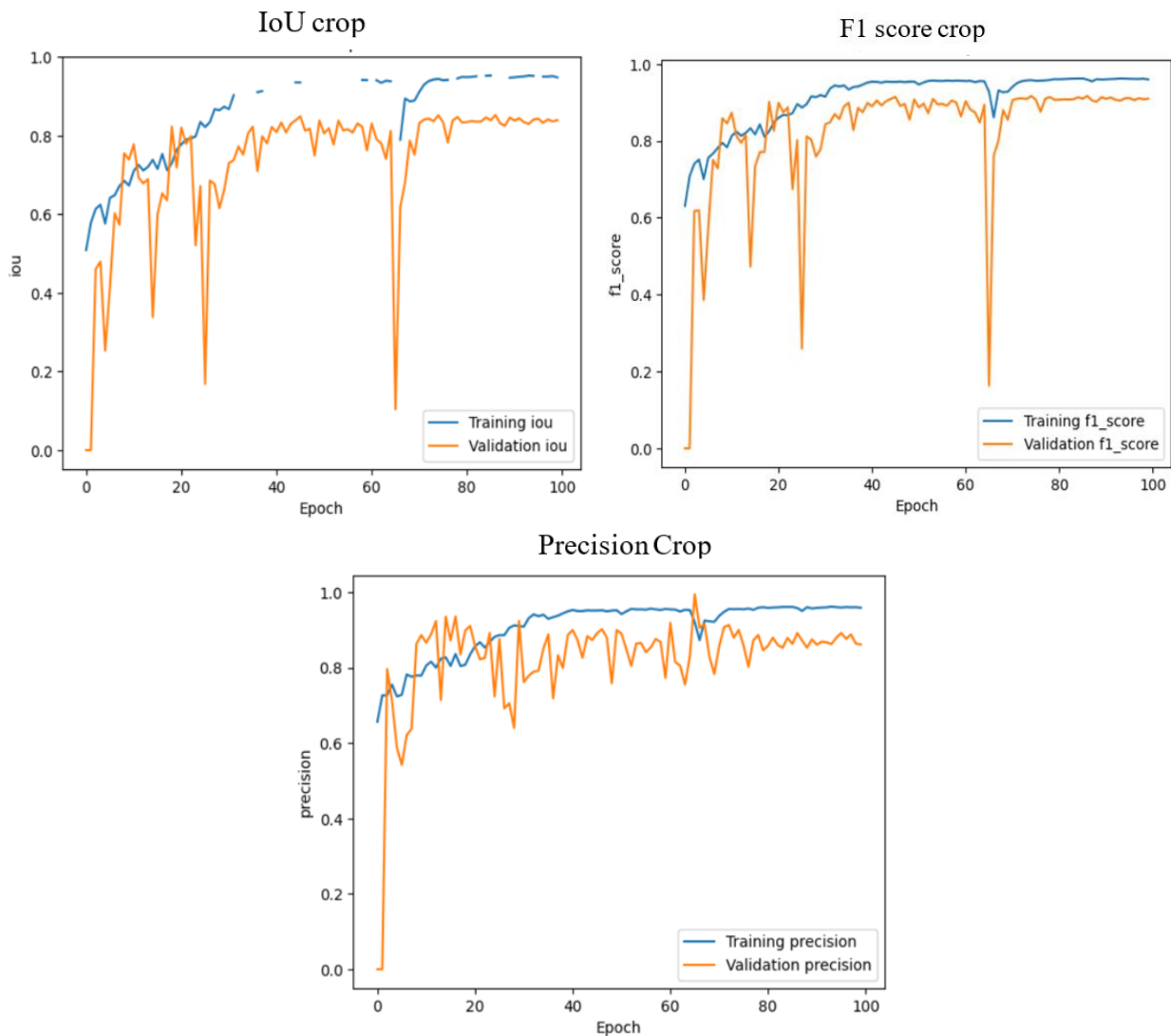


Figure 15. Train/Val plotting curve of IoU, precision and F1-score of AgriResUpNet model on crop segmentation

5.5 Comparative analysis and discussion

Different metrics, such as intersection over union (IoU) and F1-score, were generated in order to conduct a quantitative investigation for the purpose of comparing the proposed AgriResUpNet to other networks on the crop and weed dataset. The performance of our proposed AgriResUpNet is superior to that of other networks with unique margins for each and every assessment index. In Table 4, we provide a summary of the segmentation performance of our proposed architecture in comparison to all other networks and their respective assessment metrics.

The proposed AgriResUpNet for crop and weed segmentation demonstrates significant improvements in performance measures when compared to the basis models (U-Net, SegNet, FCN-8s, DeepLabv3, and CED-Net) for crop as well as weed classification, shows in Figures 16 and 17. With considerable increases in IoU and F1 Score for both crop and weed segmentation tasks, AgriResUpNet consistently beats the basic models across all assessed measures. With regard to crop and weed segmentation, in particular, AgriResUpNet earns excellent IoU scores of 94.82% and 97.58%, respectively, suggesting that it is superior in its capacity to properly identify borders. AgriResUpNet achieved a classification accuracy of 96.01% for crop segmentation as

well as 98.77% for weed segmentation, respectively, according to the F1 Score findings, which further emphasise the model's precision and recall balance. On the other hand, U-Net exhibits IoU scores of 77.75% (for crops) and 66.61% (for weed), with F1 Scores that approximate to 85.10%. In terms of crop/weed segmentation, SegNet and FCN-8s have IoU scores of 52.76%/57.17% and 62.08%/54.11%, respectively. Additionally, their F1 Scores are 70.08%/70.08% and 74.46%/74.46%, respectively. The results of DeepLabv3 indicate that the IoU scores for crop are 75.50% and for weed are 61.44%, with the corresponding F1-scores being 82.86%. The IoU scores that CED-Net obtains are 81.20% for crops and 70.16% for weeds, while the F1 scores that it achieves are 87.39%. These complete parameter values demonstrate that the AgriResUpNet architecture that was suggested is effective in improving the accuracy of segmentation for agricultural applications. This architecture also provides a potential approach for improving crop and weed identification in precision farming situations.

Table 5 displays a comparative examination of the Intersection over Union (IoU) scores for several models. The Intersection over Union (IoU) score for Convolutional Neural Networks (CNNs) is 65.74%, signifying a reasonable degree of accuracy. The TL-ResUNet model exhibits a substantial enhancement, achieving an Intersection over Union (IoU) of

81%, hence showcasing a more resilient and reliable performance. Nevertheless, the suggested model surpasses both alternatives, attaining an amazing Intersection over Union (IoU) score of 97.58%, indicating a remarkably precise capacity to execute segmentation. The comparison demonstrates the higher performance of the proposed model in terms of Intersection over Union (IoU), proving it as a more effective solution for the given task.

Table 5 presents a comparison of crop and weed segmentation models, demonstrating that AgriResUpNet outperforms other models in terms of both IOU and F1 Score measures.

Table 6 presents a comparison of IoU performance among

different models. It demonstrates that the suggested model exhibits a substantial improvement (97.58) compared to CNNs (65.74) and TL-ResUNet (81).

Figure 18 depicts the visual representation illustrates a bar chart labelled as "Comparative Analysis." The performance of three models, namely AgriResUpNet (Proposed), TL-ResUNet mode, and CNNs, is compared based on their Intersection over Union (IoU) performance. AgriResUpNet demonstrates superior performance compared to the other models, achieving an IoU score of 97.58. TL-ResUNet follows with a score of 81, while CNNs achieve a score of 65.74. A higher IoU indicates a greater level of accuracy in the model.

Table 5. Comparison between base and proposed models for crop and weed segmentation

Models	Crop Segmentation		Weed Segmentation	
	IOU	F1 Score	IOU	F1 Score
U-Net	77.75	85.10	66.61	85.10
SegNet	52.76	70.08	57.17	70.08
FCN-8s	62.08	74.46	54.11	74.46
DeepLabv3	75.50	82.86	61.44	82.86
CED-Net	81.20	87.39	70.16	87.39
AgriResUpNet (Proposed)	94.82	96.01	97.58	98.77

Table 6. Comparative Analysis of Existing Model and Proposed Model

Models	IoU	References
CNNs	65.74	[48]
TL-ResUNet mode	81	[49]
Proposed model	97.58	--

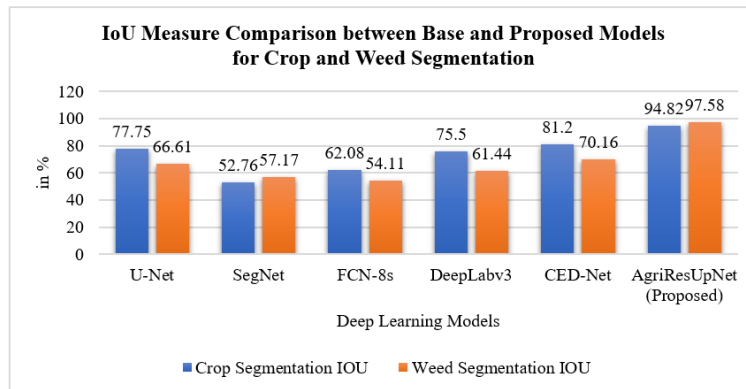


Figure 16. Bar graph of IoU measure comparison between base and proposed models for crop and weed segmentation

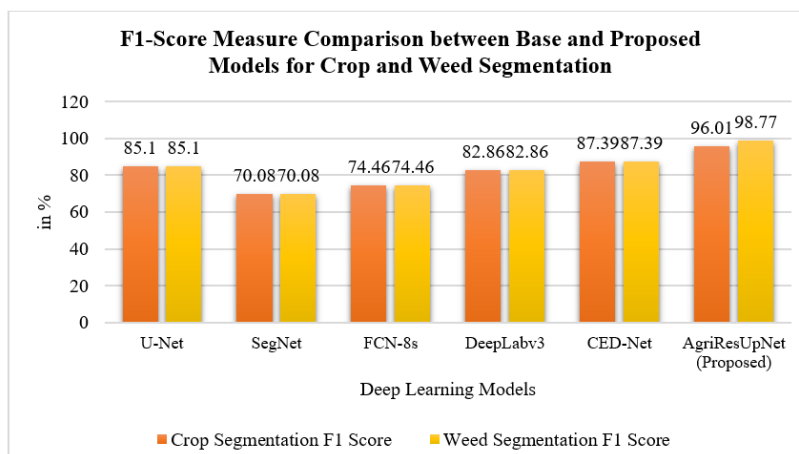


Figure 17. Bar graph of F1-score measure comparison between base and proposed models for crop and weed segmentation

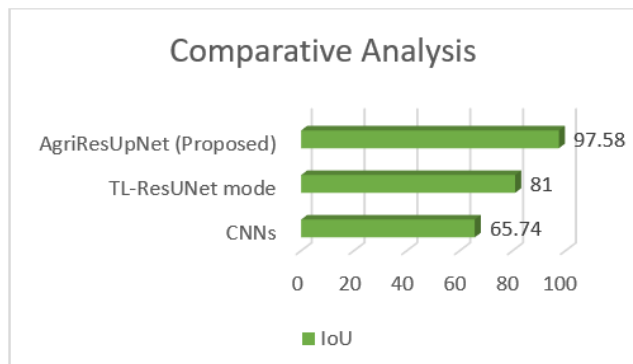


Figure 18. Comparative analysis graph

6. CONCLUSION

Semantic segmentation is a crucial component of image processing as well as machine vision. To accomplish precise segmentation, it identifies and assesses each pixel of an image. AgriResUpNet is a straightforward and effective segmentation model that can be trained on big data sets. We employ the AgriResUpNet networks in this study to classify and segment crop and weed photos. Simulations demonstrated that the AgriResUpNet suggested in this work outperforms previous algorithms in terms of IoU and F1-score when compared with public dataset crop as well as weed datasets. For the weed segmentation AgriResUpNet model obtain 94.82% IoU and 96.01% F1-score, while on Weed dataset proposed model get 97.58% IoU and 98.77% F1-score, respectively. We believe that the revised AgriResUpNet architecture suggested in this study will be more suited for crop picture segmentation. The comparison demonstrated that our methodology increased the stability and objectivity of counting findings while saving researchers time, and the technique is already being used to analyse pictures of radiation-injured bone marrow cell alterations. Our long-term goals include improving the model's segmentation performance by fine-tuning its parameters and then applying them to even better base models; expanding the model's applicability to various forms of medical data with varying segmentation aims; enhancing AI's capacity to detect and isolate weeds and crops; and assisting medical professionals in their work.

7. FUTURE WORK DIRECTIONS

(1) Enhancement through advanced attention mechanisms:

Integrate advanced attention mechanisms like self-attention or spatial attention modules into AgriResUpNet. These mechanisms can help the model focus on important regions in the image, improving the precision of segmentation. Attention mechanisms can enhance the model's ability to distinguish between similar-looking crops and weeds, especially in complex agricultural landscapes.

(2) Multi-modal data integration:

Expand the model's capabilities by incorporating multi-modal data, such as combining RGB images with depth or thermal imaging data. This integration can provide additional contextual information, helping to improve segmentation accuracy. For example, depth data can help differentiate crops and weeds based on height, while thermal imaging can highlight differences in temperature, which may correspond to

different plant types.

(3) Transfer learning and domain adaptation:

Apply transfer learning techniques by pre-training the AgriResUpNet model on large, diverse datasets before fine-tuning it on specific agricultural datasets. Additionally, explore domain adaptation strategies to make the model robust across different environments and crop types. This approach can significantly improve the model's performance when applied to new or unseen datasets, ensuring its versatility and generalizability.

Expanding the model's applicability to various forms of medical data with varying segmentation aims could open up new avenues for AgriResUpNet. For instance, fine-tuning the model for different medical imaging modalities such as MRI, CT scans, or X-rays can help in accurate segmentation of tissues, tumors, and other anatomical structures. By doing so, the model could assist medical professionals in diagnostics and treatment planning, ultimately improving patient outcomes.

REFERENCES

- [1] Fatma, S., Dash, P.P. (2019). Moment invariant based weed/crop discrimination for smart farming. In 2019 International Conference on Computer, Electrical & Communication Engineering (ICCECE), Kolkata, India, pp. 1-5. <https://doi.org/10.1109/ICCECE44727.2019.9001903>
- [2] Kamath, R., Balachandra, M., Prabhu, S. (202). Classification of crop and weed from digital images: A review. In 2021 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER), Nitte, India, pp. 12-17. <https://doi.org/10.1109/DISCOVER52564.2021.9663729>
- [3] Maram, B., Das, S., Daniya, T., Cristin, R. (2022). A framework for weed detection in agricultural fields using image processing and machine learning algorithms. In 2022 International Conference on Intelligent Controller and Computing for Smart Power (ICICCSP), Hyderabad, India, pp. 1-6. <https://doi.org/10.1109/ICICCSP53532.2022.9862451>
- [4] Liu, J., Xiang, J., Jin, Y., Liu, R., Yan, J., Wang, L. (2021). Boost precision agriculture with unmanned aerial vehicle remote sensing and edge intelligence: A survey. *Remote Sensing*, 13(21): 4387. <https://doi.org/10.3390/rs13214387>
- [5] Saqib, M.A., Aqib, M., Tahir, M.N., Hafeez, Y. (2023). Towards deep learning based smart farming for intelligent weeds management in crops. *Frontiers in Plant Science*, 14: 1211235. <https://doi.org/10.3389/fpls.2023.1211235>
- [6] Panati, H.S., Gopika, P., Andrushia, D., Neebha, M. (2023). Weeds and crop image classification using deep learning technique. In 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 1: 117-122. <https://doi.org/10.1109/ICACCS57279.2023.10112958>
- [7] Tejada, A.I., Castro, R.C. (2019). Algorithm of weed detection in crops by computational vision. In 2019 International Conference on Electronics, Communications and Computers (CONIELECOMP), Cholula, Mexico, pp. 124-128.

- <https://doi.org/10.1109/CONIELECOMP.2019.8673182>
- [8] Fennimore, S.A., Slaughter, D.C., Siemens, M.C., Leon, R.G., Saber, M.N. (2016). Technology for automation of weed control in specialty crops. *Weed Technology*, 30(4): 823-837. <https://doi.org/10.1614/wt-d-16-00070.1>
- [9] Shaner, D.L., Beckie, H.J. (2014). The future for weed control and technology. *Pest Management Science*, 70(9): 1329-1339. <https://doi.org/10.1002/ps.3706>
- [10] Kerimkhulle, S., Kerimkulov, Z., Bakhtiyarov, D., Turtayeva, N., Kim, J. (2021). In-field crop-weed classification using remote sensing and neural network. In 2021 IEEE International Conference on Smart Information Systems and Technologies (SIST), Nur-Sultan, Kazakhstan, pp. 1-6. <https://doi.org/10.1109/SIST50301.2021.9465970>
- [11] Farooq, A., Hu, J., Jia, X. (201). Weed classification in hyperspectral remote sensing images via deep convolutional neural network. In IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, pp. 3816-3819. <https://doi.org/10.1109/IGARSS.2018.8518541>
- [12] Lottes, P., Behley, J., Chebrolu, N., Milioto, A., Stachniss, C. (2018). Joint stem detection and crop-weed classification for plant-specific treatment in precision farming. In 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, pp. 8233-8238. <https://doi.org/10.1109/IROS.2018.8593678>
- [13] Moazzam, S.I., Khan, U.S., Nawaz, T., Qureshi, W.S. (2022). Crop and weeds classification in aerial imagery of sesame crop fields using a patch-based deep learning model-ensembling method. In 2022 2nd International Conference on Digital Futures and Transformative Technologies (ICoDT2), Rawalpindi, Pakistan, pp. 1-7. <https://doi.org/10.1109/ICoDT255437.2022.9787455>
- [14] Ravindaran, R., Kasthuri, N., Preethi, S., Adithya, B., Sp, G., Dharanidharan, K., Aravinth, S. (2023). Performance analysis of a VGG based deep learning model for classification of weeds and crops. In 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, pp. 1-5. <https://doi.org/10.1109/ICCCNT56998.2023.10307169>
- [15] Reddy, L.U.K., Rohitharun, S., Sujana, S. (2022). Weed detection using AlexNet architecture in the farming fields. In 2022 3rd International Conference for Emerging Technology (INCET), Belgaum, India, pp. 1-6. <https://doi.org/10.1109/INCET54531.2022.9824586>
- [16] Farooq, A., Jia, X., Hu, J., Zhou, J. (2019). Knowledge transfer via convolution neural networks for multi-resolution lawn weed classification. In 2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), Amsterdam, Netherlands, pp. 1-5. <https://doi.org/10.1109/WHISPERS.2019.8920832>
- [17] Gayathri, U., Praveena, V. (2023). A survey paper on weed identification using deep learning techniques. In 2023 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, pp. 1-4. <https://doi.org/10.1109/ICCCI56745.2023.10128610>
- [18] Khan, A., Ilyas, T., Umraiz, M., Mannan, Z.I., Kim, H. (2020). Ced-net: Crops and weeds segmentation for smart farming using a small cascaded encoder-decoder architecture. *Electronics*, 9(10): 1602. <https://doi.org/10.3390/ELECTRONICS9101602>
- [19] Gupta, N., Jalal, A.S. (2020). Text or non-text image classification using fully convolution network (FCN). In 2020 international conference on contemporary computing and applications (IC3A), Lucknow, India, pp. 150-153. <https://doi.org/10.1109/IC3A48958.2020.233287>
- [20] Jiang, Z. (2019). A novel crop weed recognition method based on transfer learning from VGG16 implemented by Keras. In IOP Conference Series: Materials Science and Engineering, 677(3): 032073. <https://doi.org/10.1088/1757-899X/677/3/032073>
- [21] Long, J., Shelhamer, E., Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431-3440. <https://doi.org/10.1109/TPAMI.2016.2572683>
- [22] Ronneberger, O., Fischer, P., Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, pp. 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [23] Badrinarayanan, V., Kendall, A., Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12): 2481-2495. <https://doi.org/10.1109/TPAMI.2016.2644615>
- [24] Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J. (2017). Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, pp. 2881-2890. <https://doi.org/10.1109/CVPR.2017.660>
- [25] Yang, Z., Peng, X., Yin, Z. (2020). Deeplab_v3_plus-net for image semantic segmentation with channel compression. In 2020 IEEE 20th International Conference on Communication Technology (ICCT), Nanning, China, pp. 1320-1324. <https://doi.org/10.1109/ICCT50939.2020.9295748>
- [26] Mohanty, S.P., Hughes, D.P., Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, 7: 1419. <https://doi.org/10.3389/fpls.2016.01419>
- [27] Teimouri, N., Dyrmann, M., Nielsen, P.R., Mathiassen, S.K., Somerville, G.J., Jørgensen, R.N. (2018). Weed growth stage estimator using deep convolutional neural networks. *Sensors*, 18(5): 1580. <https://doi.org/10.3390/s18051580>
- [28] Dyrmann, M., Jørgensen, R.N., Midtby, H.S. (2017). RoboWeedSupport-Detection of weed locations in leaf occluded cereal crops using a fully convolutional neural network. *Advances in Animal Biosciences*, 8(2): 842-847. <https://doi.org/10.1017/s2040470017000206>
- [29] dos Santos Ferreira, A., Freitas, D.M., da Silva, G.G., Pistori, H., Folhes, M.T. (2017). Weed detection in soybean crops using ConvNets. *Computers and Electronics in Agriculture*, 143: 314-324. <https://doi.org/10.1016/j.compag.2017.10.027>
- [30] Ma, X., Deng, X., Qi, L., Jiang, Y., Li, H., Wang, Y., Xing, X. (2019). Fully convolutional network for rice seedling and weed image segmentation at the seedling stage in paddy fields. *PLoS One*, 14(4): e0215676.

- <https://doi.org/10.1371/journal.pone.0215676>
- [31] Kamal, S., Shende, V.G., Swaroopa, K., Bindhu Madhavi, P., Akram, P.S., Pant, K., Patil, S.D., Sahile, K. (2022). FCN network-based weed and crop segmentation for IoT-aided agriculture applications. *Wireless Communications and Mobile Computing*, 2022(1): 2770706. <https://doi.org/10.1155/2022/2770706>
- [32] Zhuang, J., Li, X., Bagavathiannan, M., Jin, X., Yang, J., Meng, W., Li, T., LI, L., Wang, Y., Chen, H., Yu, J. (2022). Evaluation of different deep convolutional neural networks for detection of broadleaf weed seedlings in wheat. *Pest Management Science*, 78(2): 521-529. <https://doi.org/10.1002/ps.6656>
- [33] Charania, S., Lendave, P., Borwankar, J., Kadge, S. (2023). A novel approach to weed detection using segmentation and image processing techniques. In 2023 World Conference on Communication & Computing (WCONF), RAIPUR, India, pp. 1-5. <https://doi.org/10.1109/WCONF58270.2023.10235132>
- [34] Fathipoor, H., Shah-Hosseini, R., Arefi, H. (2023). Crop and weed segmentation on ground-based images using deep convolutional neural network. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 10: 195-200. <https://doi.org/10.5194/isprs-annals-X-4-W1-2022-195-2023>
- [35] Kumar, A., Rajanand, A., Kujur, A.D., Rathore, Y., Janghel, R.R. (2022). Segmentation of rice seedling using deep learning algorithm. In 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT), Indore, India, pp. 164-168. <https://doi.org/10.1109/CSNT54456.2022.9787601>
- [36] Chitty-Venkata, K.T., Somani, A.K., Kothandaraman, S. (2021). Searching architecture and precision for U-net based image restoration tasks. In 2021 IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA, pp. 1989-1993. <https://doi.org/10.1109/ICIP42928.2021.9506777>
- [37] Miao, F., Zheng, S., Tao, B. (2019). Crop weed identification system based on convolutional neural network. In 2019 IEEE 2nd International Conference on Electronic Information and Communication Technology (ICEICT), Harbin, China, pp. 595-598. <https://doi.org/10.1109/ICEICT.2019.8846268>
- [38] Mishra, N., Jahan, I., Nadeem, M.R., Sharma, V. (2023). A comparative study of resnet50, efficientnetb7, inceptionv3, vgg16 models in crop and weed classification. In 2023 4th International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, pp. 1-5. <https://doi.org/10.1109/ICIEM59379.2023.10166032>
- [39] Tlebalidinova, A., Karmenova, M., Ponkina, E., Bondarovich, A. (2022). CNN-based approaches for weed detection. In 2022 10th International Scientific Conference on Computer Science (COMSCI), Sofia, Bulgaria, pp. 1-4. <https://doi.org/10.1109/COMSCI55378.2022.9912593>
- [40] Siddique, N. (2021). U-net based deep learning architectures for object segmentation in biomedical images (Master's thesis, Purdue University).
- [41] Ebrahimi, M.S., Abadi, H.K. (2021). Study of residual networks for image recognition. In *Intelligent Computing: Proceedings of the 2021 Computing Conference*, 2: 754-763. https://doi.org/10.1007/978-3-030-80126-7_53.
- [42] Li, W., Sun, W., Zhao, Y., Yuan, Z., Liu, Y. (2020). Deep image compression with residual learning. *Applied Sciences*, 10(11): 4023. <https://doi.org/10.3390/app10114023>
- [43] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [44] Montazerolghaem, M., Sun, Y., Sasso, G., Haworth, A. (2023). U-Net architecture for prostate segmentation: the impact of loss function on system performance. *Bioengineering*, 10(4): 412. <https://doi.org/10.3390/bioengineering10040412>
- [45] Butwall, M., Ranka, P., Shah, S. (2019). Python in field of data science: A review. *International Journal of Computer Applications*, 178(49). <https://doi.org/10.5120/ijca2019919404>
- [46] Johnson, J.W., Jin, K.H. (2020). Jupyter notebooks in education. *Journal of Computing Sciences in Colleges*.
- [47] Ranjani, J., Sheela, A., Meena, K.P. (2019). Combination of NumPy, SciPy and Matplotlib/PyLab-a good alternative methodology to MATLAB-A Comparative analysis. In 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT). <https://doi.org/10.1109/ICIICT1.2019.8741475>
- [48] Jiang, K., Afzaal, U., Lee, J. (2022). Transformer-based weed segmentation for grass management. *Sensors*, 23(1), 65. <https://doi.org/10.3390/s23010065>
- [49] Safarov, F., Temurbek, K., Jamoljon, D., Temur, O., Chedjou, J.C., Abdusalomov, A.B., Cho, Y.I. (2022). Improved agricultural field segmentation in satellite imagery using TL-ResUNet architecture. *Sensors*, 22(24): 9784. <https://doi.org/10.3390/s22249784>