Journal of the Maharaja Sayajirao University of Baroda
ISSN :0025-0422

EFFECTIVE DEEP LEARNING ARCHITECTURES TO IMPROVE AGRICULTURAL PRODUCTION QUALITY

Dr. R. Jayaprakash, Assistant Professor, Department of Computer Science, NGM College,Pollachi.
Ms. A. Kalaivani, Assistant Professor, Department of Computer Science, NGM College, Pollachi.
jpinfosoft@gmail.com

Abstract: Agriculture, a powerful word that requires no definition, contributes the most profitable share of the Indian Economy. On a wider view, intensifying productivity requires very basic things like improving the quality and quantity of the productivity and dwindling the agricultural expenses. There are certain cases that affect productivity, among which the presence of weeds and pests in crop fields stands top. Traditional methods, such as manual removal of weeds and spraying of agrochemical products like herbicides, and pesticides have chances of harming the crops and also add to the expenses. Selective treatment against weeds and pests is a cost-effective method that reduces manpower and usage of the agrochemical, at the same time it requires an effective computer vision system to identify issues and should be smaller in size to run in the resource-constrained device. Machine learning-based activities for identifying weed portions and pest detections are tedious and time-consuming processes. This research work presents a method of effective segmentation in the agricultural field, to accomplish this; a convolution neural network is proposed named Reduced Residual U-Net Convolution (RRUNC) network. It is an encoder-decoder based architecture. This network employs semantic process to analyse the crop field images pixel-wise. In order to reduce the parameter size, the deep convolution technique is used which curtails the number of parameters generated by the model with a very negligible in the accuracy. The experimental findings show that the proposed deep learning-based RRUNC model outperformed well in terms of segmentation results with good accuracy and less error-rate.

Key words: Agriculture, Convolution, Pest, Residual, Weeds.

I.INTRODUCTION

The agricultural sector plays a crucial role in the economy and is considered to be one of the cornerstones of the economy of developing countries, especially India. Because food and raw materials are not the only things that agriculture provides: it also gives employment for a huge number of the population in India. According to the Indian Agriculture and Allied Industry report published in November 2021, about 58% of India’s population relies on agriculture as its primary source of income. A total of 17.8% of India's Gross Value Added (GVA) in FY20 came from agriculture and related sectors. Globally, India is ranked fifth in terms of production, consumption, export, and expected growth.

Traditional agricultural activities spend a lot of time and expenditure on their agricultural work due to the lack of use of technological advancements. To preserve their production from weeds, farmers utilized physical weeding and spraying of herbicides in the past. Manual weed removal increases expenditures and takes more time and effort, thus increasing labour costs. While spraying herbicides over agricultural land may spoil the soil, it also has the potential to turn plants deadly, posing a risk to human health. Farmers are losing interest in their agricultural labour as a result of this, and are leaving the industry. To address these difficulties and boost production with existing agricultural land, agricultural activities should utilize modern research and technology development. This research focuses on finding a more efficient strategy to deal with weeds and pests.

Weeds are the most important biotic constraint impacting agricultural production. Weeds are invasive plants. The presence of weeds in a crop field impairs the yield, quantity, and quality. It degrades the crop yield invisibly. The weeds absorb all the nutrients required by the crops from the soil, water, and fertilizer. In the absence of adequate nutrients, crops may become of poor quality and hence susceptible to crop diseases. Weeds are not susceptible to pests and diseases since these are normally controlled by their natural habits. As a result, it quickly spreads across a wide area taking up more space on agricultural land which decreases crop yields.

Pests are another major threat to agriculture productivity. Pests infest agriculture fields, causing a variety of diseases and wreaking havoc on plants. The presence of pests in agricultural land prevents farmers from getting good yields from their harvests. Consequently, identifying pests earlier in their life cycle is an essential and profitable practice for farmers, as they can avoid most of the crop diseases that affect their crops. But in reality, it is difficult to identify these diseases at an early stage.

For pest identification, they rely on visual assessments done by the agronomist force (Yadav et al. 2021). Identifying the right agronomists at right time is not practically possible. Most pest diseases have identical characteristics, making it difficult to identify correctly even for experienced farmers. As a result, it may lead to inadequate treatment, which is equally damaging to plants. Pest can be controlled a little easier in earlier stages rather than in posterior stages. But, many pests in their early stages are worm-like, and so it's more difficult to classify them. Some of the sample pest images of earlier and later stage are shown in Figure 1. Therefore, farmers might fail to take the right action at the earliest stage, and before implementing the necessary measures the pest may spread to half of the field. Utilizing current technology and automating a few of their activities could allow them to overcome these challenges, like weed detection, and pest detection that cannot be solved using traditional methods.

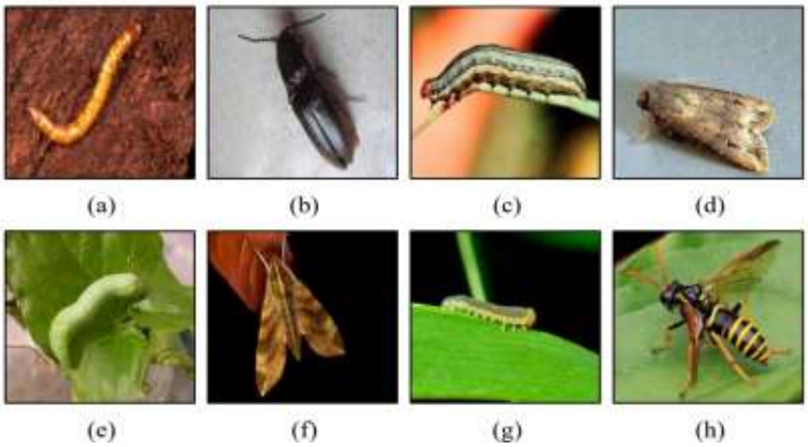


Figure 1. Some of the sample images of pests that are in worm type in its earlier stage. a) Wireworm – earlier stage, b) Wireworm – later stage, c) Armyworm – earlier stage, d) Armyworm – later stage, e) Ampelophaga – earlier stage, f) Ampelophaga – later stage, g) Wheat sawfly – earlier stage, h) Wheat sawfly – later stage

II. LITERATURE SURVEY

This section provides an overview of the contributions of computer vision applications to agricultural activities against weeds and pests. In the real world, especially in the agriculture field, a deep learning model that is both computationally and accurately efficient is required. The majority of research work in the deep learning field is concerned with developing new deep learning architecture to improve efficiency, while a few studies look at parameter and model size reduction.

Simonyan & Zisserman (2014) proposed a way of constructing a deep learning model. A constant and small-sized kernel 33 is used in all the convolution layers of their proposed CNN model, where the depth of this model was 16 -19 convolution layers. Here, each layer is equipped with ReLU activation and Batch Normalization function. As we descended deeper into the model, the number of filters doubled, such as 64, 128, 256, and 512.

For autonomous driving car applications, Badrinarayanan et al. (2017) presented an encoder-decoder based SegNet architecture. Here, the proposed model was organized in such a way that every block/layer in the encoder has a corresponding block/layer on the decoder side. In this work, the usage of pixel-wise labeling to segment road scene objects like vehicles, buildings, pedestrian trees, and other objects for smooth driving of a car, is considered. Both encoder and decoder have 13 convolution layers on each side. Yasrab et al. (2017) proposed a CNN for Semantic Segmentation for driver Assistance system (CSSA), which employed a semantic segmentation technique. The structure of the CSSA architecture is based on SegNet architecture along with the dropout layer. The model size of CSSA was ~30 MB and Class Average Accuracy achieved (CAA) was 60.2%.

The U-net architecture, which is an encoder-decoder type model, was proposed by Ronneberger et al. (2015). A contracting path and a symmetric expanding path are the two major components of this model. The significance of the proposed work lies in combining the contracting path with expanding path by creating a concatenation path that integrates the feature map generated in the contracting path with the expanding path. Since this model used the unpadded convolution, there is a possibility of losing corner features.

For segmenting the head-shoulder section of pedestrians, Xie et al. (2019) used a modified version of the UNet architecture. Because it used a large number of filters in the convolution layers of its design, it created a significant number of parameters. Augustaukas and Lipnickas (2019) proposed a pixel-wise road pavement defect detection method. The underlying model structure was based on the U-Net architecture and the convolution layers in this model were using padded convolution operations. This was used to detect the pixel-level cracks in the road which helped in the maintenance and monitoring of roads.

Naqvi et al. (2020) introduced the OcularNet model for accurately detecting ocular areas like the iris and sclera. The suggested OcularNet model was a SegNet architecture with lite-residual encoders and decoders. The key functionality of this work was using the residual skip-connection between the convolution layers of the SegNet model. It helps to transfer the high-frequency information among the convolution layers, which were used to attain precise ocular regions segmentation with the true boundary of ocular regions.

The DenseNet model was developed by Huang et al. (2017), in which each layer is connected to every other layer in the model using the concatenation operation in a feed-forward way. As a result of this, feature maps generated by all the preceding layers could be given as to a layer and its own feature map was used as input to its subsequent layers. Hence the vanishing gradient problem was reduced, and feature maps were better utilized with dense connectivity.

For efficient object identification, Tan et al. (2020) developed EfficientDet, a one-stage object detector that used a weighted bidirectional feature network and a tailored compound scaling algorithm. The proposed EfficientDet model was delivered as different variants, such as EfficientDet D0 to D7. All these variants were evaluated with the COCO dataset and achieved better accuracy with fewer parameters.

Yankun et al. (2021) introduced the Large Motion Trend fusion (LMT-CH) vehicle tracking algorithm, which was based on the big motion trend coupled with the colour histogram. The EfficientDet model was employed as the underlying model in this vehicle tracking algorithm, which increased the accuracy of the real-time applications. The approach's shortcomings were that accuracy might be marginally reduced if vehicles suffered significant changes during the drive.

The motivation behind the work proposed by Chen et al. (2021) was guiding model selection for the critical task of Vehicle and pedestrian detection in autonomous driving applications. In this work, the authors analyzed several object detection models, including Faster R-CNN, R-FCN, and SSD, along with several typical feature extractors, such as ResNet50, ResNet101, MobileNet_V1, MobileNet_V2, Inception_V2, and Inception_ ResNet_V2. Further, all the above models were pre-trained with the COCO dataset for improving their performance. Later, various experiments were conducted on these pre-trained models using the benchmark KITTI dataset. Finally, the authors demonstrated that Faster R-CNN ResNet50 obtained the highest AP and SSD MobileNet_V2 was the fastest model.

A CNN architecture without anchors was built by Liu et al. in 2020. He also made a frame-by-frame technique incorporating a lightweight stacked hourglass network to predict the heatmap at the center point of a surgical tool for real-time surgical tool detection during robot-assisted surgery. On the VisDrone2019 dataset, Pailla et al. (2019) employed CenterNet with the HourGlass-104 backbone network for real-time object detection, outperforming other significant object detection models.

Chen et al. (2019) suggested a separable convolution-based CNN model with greater parameter efficiency to adaptively extract forensics-related features from picture patches. To achieve parameter efficiency, the suggested model uses a separable convolution technique rather than the traditional convolution technique. The proposed model was used with a smaller size image like 32×32.

This section also discussed various existing research work carried out on the process of detecting weeds, and pests in the agricultural field, which includes both machine learning, and deep learningbased approaches. Machine Learning-based approaches struggle hard to produce better results because of their manual feature extraction process. Deep Learning-based approaches simplified the feature extraction process that helps to improve the model's performance. As a result of this, efficient deep learning models were developed to aid in the improvement of Smart farming performance in terms of crop-weed segmentation, and pest detection in agricultural activities.

III. MATERIALS & METHODS

To build a computer vision application for segmenting pest and weed portions in the agricultural field images, an encoder-decoder based on the deep learning model is built and trained using a large number of images and their corresponding target label images. In these images, the pixels can be categorized into three classes: background, pests, and weeds. Accurate pest-weed section segmentation is made feasible by identifying every pixel in the image rather than just one whole image (Long et al. 2015). Later, this trained model can be utilized to segment pest-weed portions from the unseen images. This work proposes a pest weed segmentation model based on Unet (Xie et al. 2018), DeepUnet (Li et al. 2018), and Residual Blocks (Naqvi et al. 2020), where the strength of residual and skip connections is incorporated into the encoder-decoder model.

The primary objective of the proposed approach is to build an effective computer vision system using a deep learning model to facilitate pixel-wise labeling in-order to classify the crops and weeds accurately in the agricultural land. This can be used to aid robotics to do selective spraying and mechanical weed removal. The proposed model is effective in every way, including obtaining a high segmentation accuracy, being less computationally complex, reduction in the size of the model, and having a lower error rate. The system framework of the proposed work is shown in Figure 2.

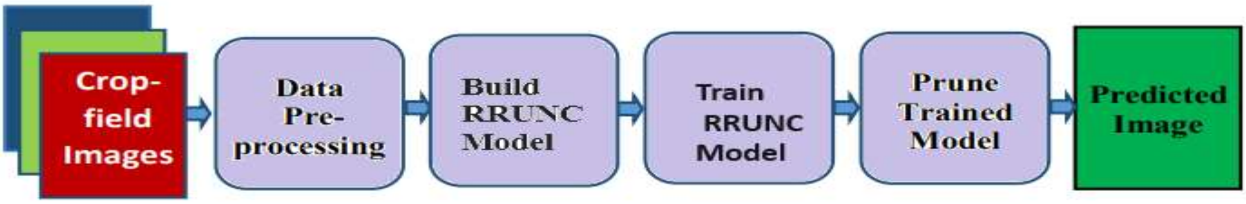


Figure 2. The system framework of the proposed crop-weed

3.1 Data Pre-processing

The dataset used here is the benchmark dataset, Crop/Weed Field Image Dataset (CWFID) proposed in the work (Haug&Ostermann 2014). This dataset contains 60 RGB images with their corresponding annotated label images. The annotation process of the CWFID dataset was done by the experts (Haug&Ostermann 2014). The portions of the actual crop field images are categorized into three classes, namely crop, weed, and background. In the label images, the crop field portions are shaded with different colours such as red, yellow, and black for the crop, weed, and background portions, respectively.

Dataset is fuel for building a deep learning model for any application. The success of building an efficient deep learning model requires a large number of images to train the model otherwise model is prone to overfit. An effective way to address this issue is data augmentation. Data augmentation is used to create new transformed images from existing original images available in the dataset. With this, a larger dataset can be created from the existing smaller dataset. This helps to upgrade the ability of the model to generalize and improve its performance since it creates more variance in training images.

Initially, the existing 60 RGB images and annotated images are divided into a 5:1 ratio for the training and testing process. At this rate, both the actual crop-field and label images are applied for various augmentation operations like rotation, horizontal flipping, vertical flipping, height shifting, width shifting, zooming, and shearing. In order to create a valid pair of actual and label images during the augmentation process, these augmentation operations must be applied in the same order on both, the actual field images and the label images. This is an extremely crucial task when augmenting the training images of encoder-decoder based segmentation model. The effect of each augmentation operations on a sample actual and label image, with this augmentation process, a new dataset containing 750 training images and 100 testing images is created, and every image is resized to 224×224×3.

3.2 Proposed Model Architecture

To achieve the precise crop-weed segmentation, a model called Reduced Residual U-Net Convolution (RRUNC) is proposed in this work, which is an Encoder-Decoder based U-Net architecture (Ronneberger et al. 2015) where depth-wise separable convolution is used as core convolution operation. It contains two major sections, namely the contracting path and expanding path. This architecture has 19 convolution layers in total, which is fewer than some of the top architectures, including VGG19 (Simonyan & Zisserman 2014), SegNet (Badrinarayanan et al. 2017), and Deep UNet (Li et al. 2018). The structure of the proposed RRUNC architecture is illustrated in Figure 3.

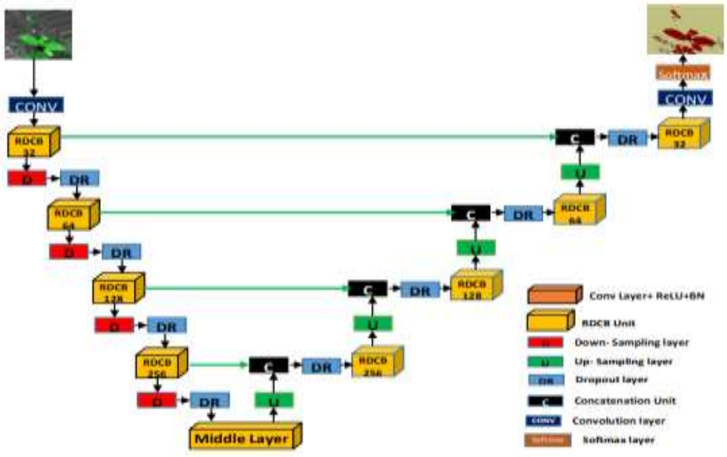


Figure 3. The Structure of proposed Reduced Residual U-Net Using Convolution (RRUNC) Model Architecture

In the proposed RRUNC, a Residual Deep Convolution Block (R-DCB) is introduced. Among the 19 convolution layers, 16 convolution layers are organized as 8 R-DCB units (in which 4 units are located in the contracting path; 4 units are located in the expanding path), one-layer acts as the middle layer, and two layers are used to input and output the image. Additionally, a down-sampling layer and a dropout layer are added in the contracting path following each R-DCB unit, and an up-sampling layer is added corresponding to the down-sampling layer in the expanding path. Going deeper into the contracting path, the number of filters used in the R-DCB unit is increased as 32, 64, 128, and 256 which is lesser than the number of filters used in the modified U-Net network (Xie et al. 2018), and the reversed order is used in expanding path. The middle layer contains 512 numbers of filters.

Here, the output of each R-DCB unit in the contracting path is integrated with the same level of R-DCB unit in the expanding path using a concatenation path via the concatenation unit. The output of each contracting path R-DCB unit is combined with the feature map from the appropriate up-sampling layer using the concatenation path and concatenation unit, which is then fed into the R-DCB unit. Hence, the location information from the contractive path is integrated with contextual information in the expanding path to form generalized information that helps to obtain excellent segmentation accuracy.

3.3 Residual Deep Convolution Block (R-DCB)

The structure of the R-DCB unit is shown in Figure 4. Each unit contains two convolution layers and one residual connection. Each convolution layer is followed by the ReLU activation layer and Batch Normalization (BN) layer to increase the segmentation accuracy and its outputs are represented by y1 and y2. Assume x, y are the input and output of an R-DCB unit, the ReLU activation function and Batch Normalization are represented by σ, the convolution operation is represented by Mci and the function of the RDCB unit is represented by MR-DCB (x) respectively. The mathematical form of y1 and y2 are mentioned in Equations (3.1) and (3.2) respectively.

$$y1 = \sigma\,(Mc1\,(x)) \tag{3.1}$$

$$y2 = \sigma\,(Mc2\,(y1)) \tag{3.2}$$

$$= \sigma\,(Mc2\,(\sigma\,(Mc1\,(x)))) \tag{3.3}$$

The residual connection layer adds both outputs y1 and y2 and then it is fed into ReLU activation layer and BN layer whose output is represented in the Equation (3.4).

$$y = MRDCB\,(x) = \sigma\,(y1 + y2) = \sigma\,(\sigma\,(Mc1\,(x)) + \sigma(Mc2\,(\sigma\,(Mc1\,(x))))) \tag{3.4}$$

The residual connection path in the RDCB unit facilitates local residual learning to improve the information flow among the architecture blocks. It will improve the representation capacity of the model and will curtail the error rate of the segmentation with the smaller dataset. So, the corner features of the crops and weeds can be predicted accurately.

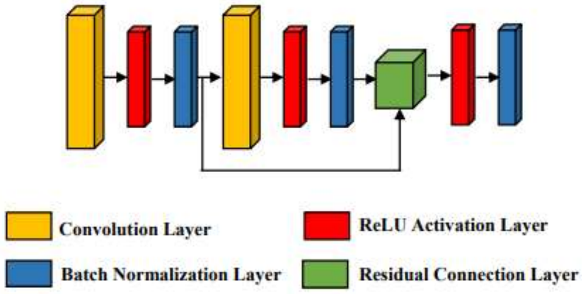


Figure 4. The Structure of Residual Deep distinguishable Convolution Block (RDCB) Unit

3.4 Pruning the Model
Usually, the machine/deep learning model is the core component of computer vision applications. Since the target environment (agricultural land) includes resource-constrained devices, building a real-time computer vision application to aid farmers requires a smaller size model. Although larger deep learning models may perform better in a lab environment, they are more difficult to implement in devices with limited resources. Even though it has been deployed, it fails to give better results. For this, the proposed deep learning model has been compressed using filter pruning.

After building the final model, the model will undergo a training process till achieving better validation accuracy and test accuracy. Subsequently, the model will be exposed to the pruning process to reduce the size of the model (Gaikwad & El-Sharkawy 2018, Mao et al. 2019). During the pruning process, the redundant and insignificant connections and their corresponding lowest magnitude weights are iteratively eliminated. So, the connection between the channels will become sparse. This pruning process achieves significant size reduction in the final model and it can be delivered by less than ~3MB in size without significant recession in performance.

IV. EXPERIMENTAL ENVIRONMENT DESIGN
The design of the experimental environment for implementing the proposed model and the evaluation metrics used to measure the performance of the proposed model are discussed here.

4.1 Experimental Setup
All the models utilized in this work are based on Tensor Flow and Keras framework. These models have been experimented on the online Google Colab platform which provides NVIDIA TeslaT4 GPU with CUDA Version 11.2, RAM of 13 GB, and Disk Space of 68 GB. Google Colab is a Google product that allows python code to be run on a cloud environment via a web browser, and hence there is no hardware restriction for it. Simple computers with an internet connection are capable of running it.

The proposed network is trained from the scratch using a dataset of 750 RGB images and its equivalent annotated images or labelled images; each image is a size of 224×224×3. For tracking the performance of the model during training epochs, the training dataset is split into training and validation datasets using the cross-validation method. Since the validation_split = 0.2 is used here, 20% of the images from the training dataset are brought into the validation set. Hence, 600 images are used for the training process and 150 images are used for the validation process in every epoch. When cross-validation is used, images from training and validation datasets are shuffled between them on each training epoch. Hence, all the images in both the training and validation dataset are utilized for training the model. Even though the training and validation split reduces the training samples, it will not degrade the model performance because of the cross-validation method.

Later, for testing the trained model, a separate test data set is built using a holdout method, which contains 100 crop field images and their equivalent label images. The holdout method for training and testing split was chosen because it helps to evaluate the model's ability in segmenting the weed and pest portions on unseen images.

The proposed model is trained with 80 training epochs with the batch size as 5 images per batch and each epoch contains 120 steps. The loss function used here is categorical_cross_entropy. Since, it has three classes which are crop, weed, and soil, this loss function is quite suitable. This measures the cross-entropy loss between the label and prediction. For this, labelled images are transformed into one-hot representations for facilitating the comparison between prediction and labels. During the training, the loss between the prediction and labels is optimized utilizing an ADAM optimizer with a learning rate of 0.001. It is a stochastic gradient descent method that depends on the adaptive estimation of first-order and second-order moments.

After pruning the model, two more training epochs were implemented for fine-tuning the model. The other significant deep learning architecture like SegNet (Badrinarayanan et al. 2017), U-Net (Xie et al.2018), and Residual U-Net (Yang et al. 2019) are implemented and their architecture configuration details are given in Figure 5. These architectures are trained with the same augmented CWFID dataset in the same experimental setup and the performance is evaluated for comparison.

Layer	Kernel Size	No of Layers	No of Filters	Output Shape
Conv_IP	3×3	1	64	224×224×64
Conv _1C	3×3	2	64	224×224×64
R_ConC1	-	-	-	224×224×64
Down_sampl	2×2	1	-	112×112×64
Conv _2C	3×3	2	128	112×112×128
R_ConC2	-	-	-	112×112×128
Down_sampl	2×2	1	-	56×56×128
Conv _3C	3×3	2	256	56×56×256
R_ConC3	-	-	-	56×56×256
Down_sampl	2×2	1	-	28×28×256
Conv _4C	3×3	1	512	28×28×512
R_ConC4				28×28×512
Down_sampl	2×2	1	-	14×14×512
Conv _Mid	3×3	1	512	14×14×512
Up_samp1	2×2	1	-	28×28×512
Concat_1	-	-	-	28×28×1024
Conv _1E	3×3	1	512	28×28×512
R_ConE1	-	-	-	28×28×512
Up_samp2	2×2	1	-	56×56×512
Concat_2	-	-	-	56×56×768
Conv _2E	3×3	2	256	56×56×256
R_ConE2	-	-	-	56×56×256
Up_samp3	2×2	1	-	112×112×256
Concat_3	-	-	-	112×112×384
Conv _3E	3×3	2	128	112×112×128
R_ConE3	-	-	-	112×112×128
Up_samp4	2×2	1	-	224×224×128
Concat_4	-	-	-	224×224×192
Conv _4E	3×3	2	64	224×224×64
R_ConE4	-	-	-	224×224×64
Conv_OUT	3×3	1	3	224×224×3
Softmax	-	-	1	224×224×3

Figure 5. The configuration of Residual U-Net architecture

4.2 Evaluation Metrics
In this work, six metrics are used to evaluate the performance of the model, which are

- Accuracy (A),
- Error-rate (E),
- Overall Precision (OP),
- Overall Recall (OR),
- Overall F1 score (OF),
- Number of Parameters (NoP)

Accuracy (A) states that the percent of the pixels in the prediction image is predicted correctly. Error-rate (E) is expressed as the rate of pixels in the prediction image is predicted wrongly, when compared to label images. Overall Precision (OP) is expressed as the ratio of correctly predicted positive observations to all predicted positive observations. Over Recall (OR) is the ratio between the number of correctly predicted positive observations and the total number of observations in the actual class. Overall F1 score is a metric that considers both OP and OR and it is a harmonic mean of OP and OR.
The mathematical forms of the above metrics are mentioned in Equations 4.5 to 4.9, respectively. The Number of Parameters (NoP) generated by the model is a total number of the learnable elements of all the convolution layers in CNN for deep convolution respectively.

Accuracy (A) = (OTP+OTN) / (OTP+OTN+OFP+OFN) (4.5)
Error-rate (E) = $-$ Yj$*$log(Pj) (4.6)
Overall Precision (OP) = OTP/ (OT P+OFP) (4.7)
Overall Recall (OR) = OTP OTP+OFN (4.8)
$OveralF1 = 2 \times OverallPrecison \times OverallRecallOP{+}OR$ (4.9)

where OTP is Overall True Positive, OTN is Overall True Negative, OFP is Overall False Positive, OFN is Overall False Negative, t is the total number of samples, Pj represents the jth predicted output of the model, and Yj represents the corresponding j-th target pixel value.

Accuracy (A) = (OTP+OTN) (OTP+OTN+OFP+OFN) (4.10)

V. RESULTS AND DISCUSSION

At first, the deep learning architectures SegNet (Badrinarayanan et al. 2017), U-Net (Xie et al. 2018), Residual U-Net (Yang et al. 2019), and proposed RRUSC (Reduced Residual U-Net using Standard Convolution) networks are implemented with a standard convolution technique. These models are trained using the training parameters mentioned in Table 1. The metrics of the proposed architecture RRUSC are considerably good, achieving a segmentation accuracy of 96.14% at validation time and 95.37% at testing time, which is higher than that of others. The overall F1 score achieved by the proposed RRUNC is also good and better than other state of the art models. So, it is computationally very expensive since it is implemented using a standard convolution technique.

Table 1. Training Parameters

Training Parameters	Values
Batch size	5
Steps per epoch	120
Leaning rate	0.001
Optimizer	ADAM optimizer
Loss function	categorical cross_entropy

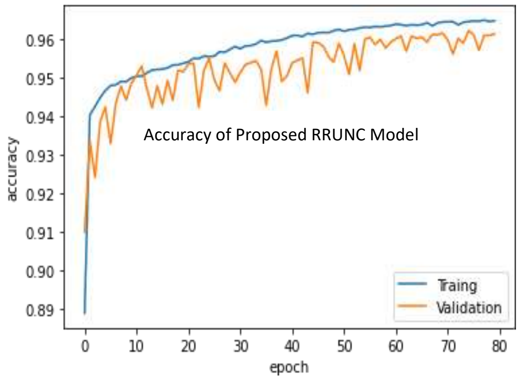
As part of the crop-weed segmentation task on the CWFID dataset, Hashemi-Beni et al. employed FCN-8s, FCN-16s, FCN-32s, and U-Net deep learning models (Hashemi-Beni et al. 2022). The performance comparison of the proposed RRUNC model with the existing work is presented in Table 2. From Table 2, the proposed Pruned RRUNC achieved a better segmentation accuracy of 96.06%, when compared to that of the existing models.

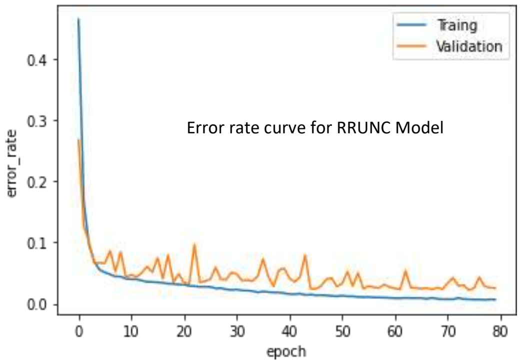
Table 2. Comparison with existing work

ModelName	A (%)	NoP (Million)	Model Size (MB)
FCN-8s	81.1	-	-

FCN-16s	77.2	-	-
FCN-32s	68.4	-	-
U-Net	77.9	-	-
ProposedPruned RRUNC	96.06	0.655	2.34

A-Accuracy, NoP–Number of Parameters, -Not given in the paper





(a) (b)

Figure6. Learning curves of the proposed RRUNC Model,

a) Accuracy curve, b) Error rate curve

VI. CONCLUSION

An encoder-decoder based deep learning model, named Reduced Residual U-Net using Convolution (RRUNC) network is proposed in this work for crop-weed segmentation. Residual Deep separable Convolution Block (R-DCB) was introduced. The proposed model is constructed in a computationally effective manner without degrading the performance of the model. The performance of this proposed model was evaluated using the benchmark dataset Crop Weed Field Image Dataset CWFID). The experimental findings show that the proposed deep learning-based RRUNC model outperformed well in terms of segmentation results with good accuracy and less error-rate. This proposed smaller model size is more suitable to build a lightweight computer vision application for weed portions and pest detection on agricultural land. It is also easily deployed on resource-constrained portable devices used by the farmers in the agricultural field.

REFERENCES

1. Yadav, S, Sengar, N, Singh, A, Singh, A & Dutta, MK 2021, ‘Identification of disease using deep learning and evaluation of bacteriosis in peach leaf’, Ecological Informatics, vol. 61, pp. 1-12.

2. Simonyan, K & Zisserman, A 2014, ‘Very deep convolutional networks for large-scale image recognition’, arXiv preprint arXiv:1409.1556.

3. Badrinarayanan, V, Handa, A &Cipolla, R 2015, ‘Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixel wise labelling’, arXiv preprint arXiv:1505.

4. Yasrab, R, Gu, N & Zhang, X 2017, ‘An encoder-decoder based convolution neural network (CNN) for future advanced driver assistance system (ADAS)’, Applied Sciences, vol. 7, no. 4, pp. 1-21.

5. Ronneberger, O, Fischer, P &Brox, T 2015, ‘U-net: Convolutional networks for biomedical image segmentation’, in International Conference on Medical image computing and computer-assisted intervention, Springer, Cham, pp. 234-241.

6. Duan, K, Bai, S, Xie, L, Qi, H, Huang, Q & Tian, Q 2019, ‘Centernet: Keypoint triplets for object detection’, in Proceedings of IEEE/CVF international conference on computer vision, pp. 6569-6578.

7. Augustaukas, R &Lipnickas, A 2019, ‘Pixel-wise road pavement defects detection using U-net deep neural network’, in 10th IEEE International Conference on intelligent data acquisition and advanced computing systems: Technology and applications (IDAACS), vol. 1, pp. 468-471.

8. Naqvi, RA, Lee, SW &Loh, WK 2020, ‘Ocular-net: Lite-residual encoder decoder network for accurate ocular regions segmentation in various sensor images’, in IEEE International Conference on Big Data and Smart Computing (BigComp), pp. 121-124.

9. Xue, Y, Chen, S, Qin, J, Liu, Y, Huang, B & Chen, H 2017, ‘Application of deep learning in automated analysis of molecular images in cancer: a survey’, Contrast media & molecular imaging, vol. 2017, pp. 1- 10.

10. Tan, M, Pang, R & Le, QV 2020, ‘Efficient det: Scalable and efficient object detection’, in Proceedings of IEEE/CVF Conference on computer vision and pattern recognition, pp. 10781-10790.

11. Yankun, Y, Xiaoping, D, Wenbo, C &Qiqige, W 2021, ‘A Color Histogram Based Large Motion Trend Fusion Algorithm for Vehicle Tracking’, IEEE Access, vol. 9, pp. 83394-83401.

12. Zhao, Y, Chen, Z, Gao, X, Song, W, Xiong, Q, Hu, J & Zhang, Z 2021, ‘Plant disease detection using generated leaves based on doubleGAN’, IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 19, no. 3, pp. 1817–1826.

13. Liu, Y, Zhao, Z, Chang, F & Hu, S 2020, ‘An anchor-free convolutional neural network for real-time surgical tool detection in robot-assisted surgery’, IEEE Access, vol. 8, pp. 78193-78201.

14. Wu, X, Zhan, C, Lai, YK, Cheng, MM & Yang, J 2019, ‘Ip102: A large-scale benchmark dataset for insect pest recognition’, in Proceedings of IEEE/CVF Conference on computer vision and pattern recognition, pp. 8787-8796.

15. Long, J, Shelhamer, E & Darrell, T 2015, ‘Fully convolutional networks for semantic segmentation’, in Proceedings of IEEE Conference on computer vision and pattern recognition, pp. 3431-3440.

16. Xie, HX, Lin, CY, Zheng, H & Lin, PY 2018, ‘An Unet-based head shoulder segmentation network’, in IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), pp. 1-2.

17. Li, R, Liu, W, Yang, L, Sun, S, Hu, W, Zhang, F & Li, W 2018, ‘DeepUNet: A deep fully convolutional network for pixel-level sea land segmentation’, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 11, no. 11, pp. 3954-3962.

18. Haug, S & Ostermann, J 2014, ‘A crop/weed field image dataset for the evaluation of computer vision based precision agriculture tasks’, in European conference on computer vision, Springer, pp. 105-116.

19. Gaikwad, AS & El-Sharkawy, M 2018, ‘Pruning convolution neural network (squeezenet) using taylor expansion-based criterion’, in IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), pp. 1-5.

20. Hashemi-Beni, L, Gebrehiwot, A, Karimoddini, A, Shahbazi, A &Dorbu, F 2022, ‘Deep convolutional neural networks for weeds and crops discrimination from UAS imagery’, Front. Remote Sensing, vol. 3, pp. 1-22.