IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 15, 2021, accepted August 16, 2021, date of publication August 26, 2021, date of current version September 3, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3108003

DeepVeg: Deep Learning Model for Segmentation of Weed, Canola, and Canola Flea Beetle Damage

MOHANA DAS AND ABDUL BAIS, (Senior Member, IEEE)
Faculty of Engineering and Applied Science, University of Regina, Regina, SK S4S 0A2, Canada

Corresponding author: Abdul Bais (abdul.bais@uregina.ca)

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under Discovery Grant RGPIN-2021-04171 entitled Crop Stress Management using Multi-source Data Fusion, and in part by the Faculty of Graduate Studies and Research, University of Regina.

ABSTRACT Farmers around the world face the challenge of growing more food for the increasing world population. On top of that, external threats such as pests (weeds and insects) pose a threat to crop production and it is necessary to take early steps to reduce the risk. This paper presents semantic segmentation of canola field images collected under natural conditions. The dataset contains four unbalanced classes; background, crop, weeds, and damages in the crop. The damages to the crop leaves are small round shaped and share the same texture and colour as whitish stones from the background. We propose, DeepVeg, a deep learning segmentation model that focuses on the smallest (damage) class without affecting other classes to solve the class imbalance issue. Early stage canola field image dataset is utilized for training and testing the proposed model. Evaluation results show that the proposed method outperforms the benchmark deep learning models and effectively addresses the weed and damaged canola plants segmentation problem. The DeepVeg model demonstrates a superior mean intersection over union score greater than 0.76 and *accuracy* above 0.97 for four class segmentation. The model also shows robustness in detecting unlabelled, newly grown weeds and canola and is also able to distinguish the similar rounded structured canola plant and weed with small amounts of data for model training, which is suitable for early stage damage and weed segmentation.

INDEX TERMS Weed segmentation, semantic segmentation, canola, leave damage, flea beetle damage, deep learning.

I. INTRODUCTION

The ability of weed plants to compete with crops for water, sunlight, and nutrients impacts the crop health by reducing the required amount of water and nutrients [1], [2]. A number of research articles have reported that the presence of unwanted weed plants can reduce the crop yield up to 34% [3]–[5]. Similarly, canola plant damage in US and Canada due to flea beetle is economically important [6]. Initially, adult flea beetle feed from weed and migrate to canola leaves, stems, and flowers when canola crop emerges [7]. Canola crop damage by flea beetle can reduce the yield up to 5% [8] and overall canola yield reduction due to weed and flea beetle damage becomes around 39%. To protect the crop from unwanted weed plants and flea beetle damages, uniform chemical spray on the entire field is commonly used which can lead to environmental pollution [3]. For a targeted spray, automated identification of weed plants is required which can provide precise individual target location [9].

The associate editor coordinating the review of this manuscript and approving it for publication was Gulistan Raja.

Plant leaves are the directly observable structure containing the health information of plants that can be easily photographed [10]. Observation of canola leaves is studied for crop diagnostics and analysis of diseases like damaged segments, marssonia, curls, bacterial wilts, and kate blights [11]–[13]. Visual inspection of damaged plant portions like leaf structure, area, pattern, density, and severity can assist the plant pathologists in plant disease diagnostic process [14]–[18]. On-site analysis of leaf samples can estimate the disease severity. However, on-sight inspection is a difficult and time-consuming process that requires crop pathologists. Measurements taken in this way, are error-susceptible and may affect the analysis due to human bias factor and fatigue. Therefore, computer-aided solutions for damaged plant analysis are needed and image segmentation is a solution to detect the weed and canola plant damage. With colour field images as input, the objective of this computer-aided solution is to introduce an automatic system for generating the segmentation masks with near realistic agricultural pathology potential [19]–[21]. Similarly, detection or segmentation of the damages in canola plants helps in diagnoses

and early precautionary measures to keep the canola crop healthy. Visual aspects and types of damages depend on the disease whereas the damage by flea beetle in canola plant can be observed as small yellowing tissue damages and shallow pits [22]. Several computer-aided solutions are reported for detection or segmentation of weed and crop [23]–[27]. These segmentation approaches can be categorised into *classical image processing*, *supervised* machine learning and *unsupervised* machine learning. Unsupervised methods require parameter tuning and mainly include threshold-based methods. The supervised learning-based methods solve these problems differently and require large dataset as well as computation power to mine the hidden features and patterns of a particular subject.

Visual features of crop images help to segment the desired portions of plants and classical image processing techniques rely on these features which may include histograms, histogram peaks, edges, thresholds, colours, locality, and orientations. To extract these features, classical image processing techniques are employed in a range of studies that utilize edge-based or filter based feature extraction methods [28]–[30]. Extracted features can be categorized or grouped into respective classes on the basis of the similarity index. These classical image processing methods address a specific problem set at a time whereas the multi-task learning schemes can solve multiple tasks using a unified model.

Unsupervised machine learning methods for crop segmentation mainly include clustering methods or hybrid techniques using feature extraction and clustering methods. The k-medoids clustering method makes clusters of discriminative colour ranges. The clustering method is applicable for grayscale and colour images which makes it robust to image noise [31]. Otsu's method [32] and k-means clustering techniques [33] are widely used in segmentation tasks. In [16], a segmentation method based on unique fermi energy level is proposed which outperform the Otsu's method [32] and k-means clustering [33]. A hybrid approach is introduced in [34] for damaged plants segmentation. This hybrid technique utilize the saliency region threshold and k-mean clustering algorithms [34]. From the findings of the aforementioned techniques it can be concluded that the threshold-based techniques are sensitive to outliers and initial seed and may infer inaccurate segmentation of plant damage and lead to an erroneous solution [35].

Real time classification of crops and weeds is carried using deep learning method in [25]. A small amount of Red Green Blue (RGB) data is considered for semantic segmentation using SegNet [36] model. The dataset contains weed, sugar beets and soil classes. The images are captured in RGB and Near Infra-Red (NIR) modes considering different weather and lighting conditions. This model is able to achieve near real time inference but quantitative results show that the weed class suffer the most. A seminal work on segmentation of weed and crop plants using deep learning methods is reported in [26]. Two different architectures of Convolutional Neural Network (CNN) models are used jointly for RGB + NIR crop field images. First CNN model is used to extract the vegetation from background whereas a deeper CNN model is utilized to classify the vegetation into the crop and weed classes. The dataset is simple and contains clean soil without dead stems, dead leaves and overlapping plants. In a major recent advancement [37] a cascaded encoder decoder architecture for the segmentation of weeds and crops is proposed. Two staged architecture is proposed where each stage contains two U-Net [38] models. In the first stage, weed and crop classes are segmented using two U-Net [38] models. The second stage replicates the preceding one but trained for higher resolution images. As a result, four U-Net [38] models are used for segmentation of weed and crop from soil. A recent work on canola and weed segmentation under natural field condition using a modified U-Net is reported in [39]. This work achieves comparable performance using a quarter of trainable parameters.

These studies mainly focus on weed, crop and soil segmentation while addressing multiple challenges in vegetation segmentation. However, there are several challenges which are not addressed yet. These challenges include crop damages by insects, complex soil with dead stems and leaves, variable lighting conditions in RGB only images and presence of different kinds of weed. To the best of our knowledge, different shapes of weeds, canola and early canola flea beetle damage segmentation using a unified deep learning model is not reported yet.

The complexities and challenges of the segmentation task can be observed in Figure 1 and are summarized as:

- Simultaneous segmentation of weed, canola and canola plant damages by flea beetle.
- Canola and weed plants share the same colour and texture information in the given region of interest.
- Segmentation of highly imbalanced dataset is not a trivial task specifically for multi-class problems.
- Semantic segmentation of more than one plant type in different soil states is challenging where the soil states vary with the variable weather conditions.
- Presence of soil clusters, dead stems, and small whitish stones in the region of interest.

To address these problems, this research proposes a novel deep learning architecture for pixel-level semantic segmentation of weed plant, canola plant and canola flea beetle damages from the complex soil background. Model architecture and objective function are designed to learn discriminative features to segment the imbalanced classes. In order to analyse the performance of the proposed model, baseline models are also trained and evaluated in the same environment. Furthermore, data preprocessing is optimized to handle class imbalance problem. A four-class segmentation dataset is acquired and used for all experiments and evaluations. Contributions of this research are:

- A four class (weed, canola, flea beetle canola damage and complex background) crop field dataset and its complexity analysis.

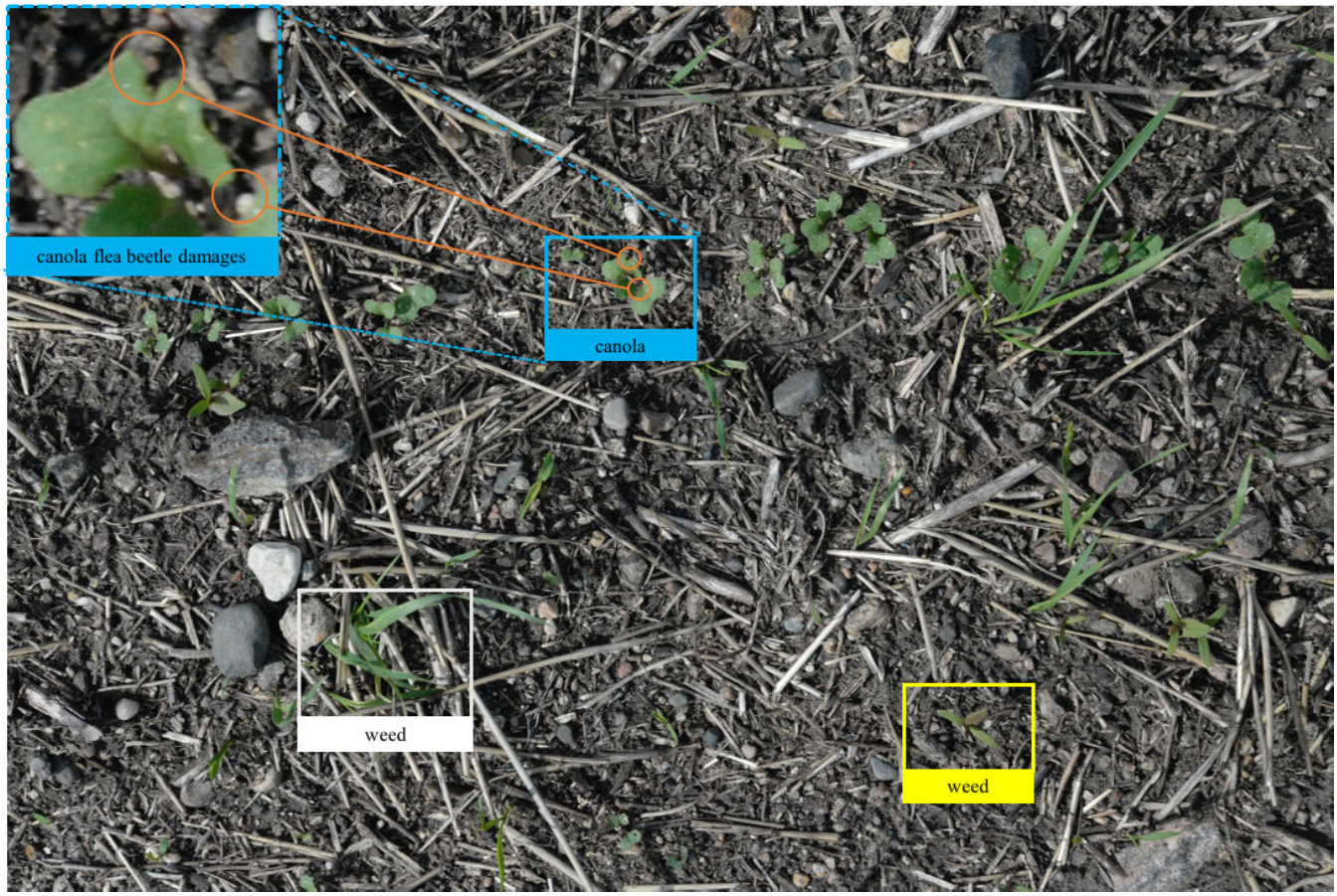


FIGURE 1. Exemplar image from the acquired dataset containing early stage canola flea beetle damage, two different types of early stage weed plants and complex background.

- A novel deep learning architecture with skip connections, residual blocks and pyramid scene parsing layers for segmentation of the highly imbalanced and complex dataset.
- Analysis of the proposed deep learning architecture and loss function to improve the segmentation performance by penalizing the least probable classes without compromising on the sharp boundaries.

The remainder of this paper is structured as follows. Section II provides the problem statement, objective function, model architecture, training, and optimization details. Results, experiments, and result analysis are discussed in Section III. Finally, Section IV presents the concluding remarks.

II. METHODOLOGY

To address the aforementioned segmentation problems, the proposed deep learning model (shown in Figure 2) aims to learn and discriminate such overlapped and imbalanced class features. The model takes an image and its corresponding segmentation masks of all four classes as input and learns the features that capture the appearance, edges, colour, size, and related information. Once it repeats the learning process on all the training images, it validates the model training in terms of evaluation metrics. The validation process takes a validation input image from the validation dataset and predicts corresponding masks. The measured distance between the validation ground truth and prediction is utilized for objective function optimization. Once the given objective function converges, the trained model is evaluated on the test dataset. Model architecture, model parameters, objective function, and learning schemes are responsible for the learning capability of a model. Design choices, model structure and the way it address the challenges is discussed in the following sections.

A. PROBLEM STATEMENT

The problem statement can be defined as given a set of unseen images, the model should focus on all the imbalanced classes while keeping the global context of the image. Having a crop field image x of size $(w \times h \times c)$ and label y as input to the model, the deep learning segmentation model can be defined as a learnable function of input which maps the input x to the output segmentation mask $\hat{y}$ as:

$$\hat{y}^{h,w,k} = \mathbb{G}(x^{h,w,c}) \tag{1}$$

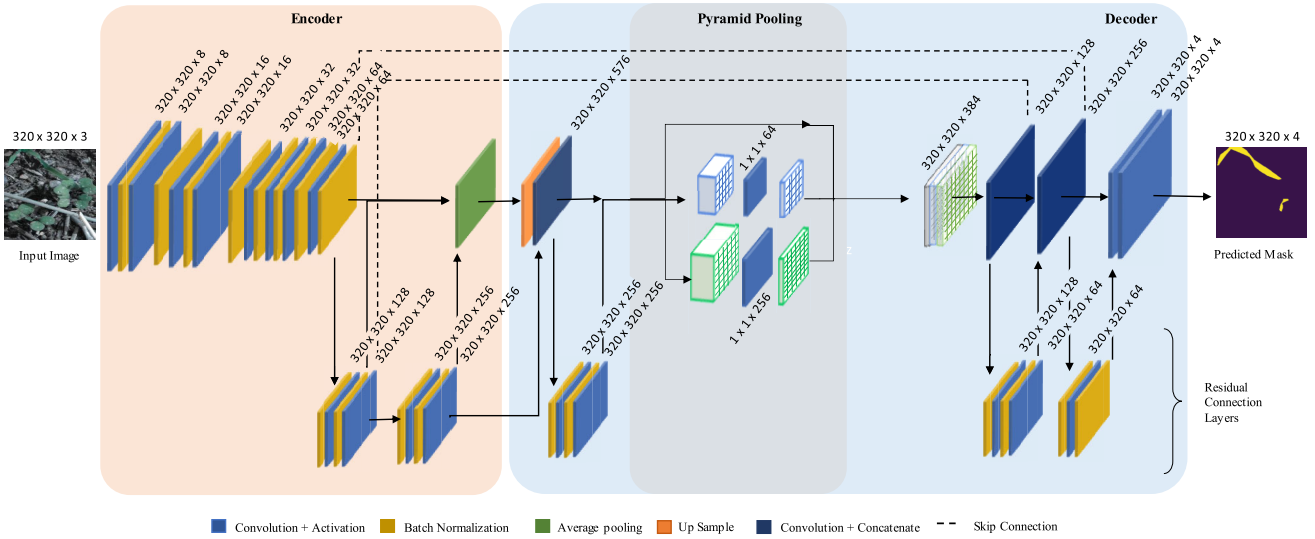


FIGURE 2. Encoder-decoder based deep learning architecture for semantic segmentation of weeds, canola, and canola leaf damage from the background.

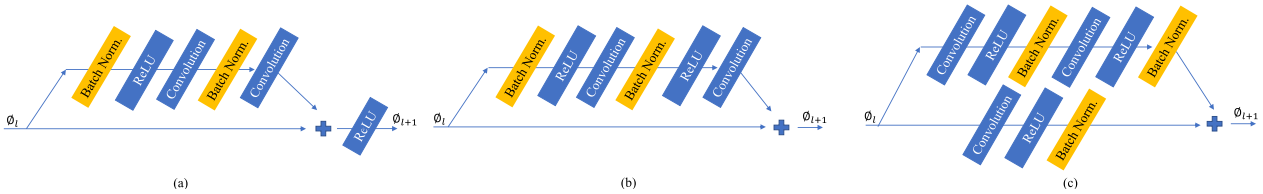


FIGURE 3. Residual block structure (a) Standard residual block [40], where activation is applied after branch addition. (b) In [41], the activation is applied before branch addition. (c) The proposed residual block contains convolutional, activation and batch normalization layers where the activation is applied before branch addition.

where h, w and c are image height, width and number of channels. k is the predicted number of segmentation masks (prediction channels) for the underline deep learning model.

B. DATASET ACQUISITION

A multi-class plant segmentation dataset is collected for segmentation of weed, flea beetle canola damages and canola plants. RGB images are collected from canola fields in Manitoba at early growth stage using quad mounted NIKON D610 camera. A total of 13000 images of dimensions $4016 \times 6016 \times 3$ are captured on different days with varying weather, soil, lighting, moisture and temperature values. Variability in image contrast, brightness and environmental factors ensure the dataset generalization and also removes environmental bias. Only 384 images out of 13000 images are used for this research work due to time constraint for annotation.

1) DATASET ANNOTATION

Manual annotation of all the four classes is performed while keeping the annotation quality same across all the images. LabelMe [42] annotation tool is selected due to its rich and user friendly interface, open-source, output format options either CSV or JSON file and re-annotation/correction options.

Polygons are applied on the boundary of each object by clicking on the edges of the objects. Three types of labels are assigned where each label present a particular class from weed, canola or canola damages. The remaining pixels are marked as the background. The entire dataset is labelled in similar fashion and downloaded in JSON format. The JSON architecture contain dictionaries for labels and corresponding multiple points of a polygon.

2) DATASET ANALYSIS

First analysis conducted for the acquired dataset is class frequency count. For segmentation, which is our case, the absolute count of pixels against each class can be achieved from the segmentation masks. For this, the number of pixels against each class are counted for all the images of the dataset. This analysis resulted in pixel-wise class count and a very high imbalance of segmentation classes is found as a result.

The obtained class distributions in terms of pixel wise percentage are 95.973%, 0.134%, 2.349%, 1.469% for background, flea beetle damages, weed and canola respectively. The class ratio depicts that the background class is prominent over all other classes. Similarly, the pixels belong to weed class are almost double in count as compared to the canola class. The number of pixels in the background class

are 716 times more than the number of pixels in the canola flea beetle damage class. This high imbalance dataset makes the segmentation task more challenging and requires class imbalance handling to reduce the segmentation bias.

C. PREPROCESSING

Data cleaning is applied in order to clean and handle the class imbalance to some extent in the preprocessing stage.

1) IMAGE TO PATCHES

It is not challenging to train a neural network on full high resolution image due to the memory constraint. For this reason, each image and corresponding segmentation masks are divided into small segments where each segment contains $803 \times 1203 \times 3$ image chunk. Image padding is applied at the boundary of image and corresponding mask if required. Each image resulted in 25 patches.

2) CLASS BALANCING

Since we divide each image into image patches, only background class in a patch is possible. We drop the image and corresponding image patches, which only contain the background class. This method is not a significant contribution to solving the class imbalance problem but dropped a small fraction of soil.

D. MODEL ARCHITECTURE

Deep learning architectures mainly define the approach to solve a particular problem or a set of problems. In this work we perform extensive comparison of well-known semantic models with different benchmark encoding modules. There are three main semantic models which are utilized under two different encoder networks as backend. We use U-Net [38], SegNet [36] and DeepLabV3+ [43] as our main benchmark segmentation models. Two encoder architectures ResNet [40] and VGG-19 are used to address weed and damaged canola segmentation. Our approach can be differentiated from U-Net [38] and SegNet [36] as we concatenated the skip connections and added the residual connections for feature sharing. ResNet [44] utilize the residual connections for deeper architectures to avoid several issues that are discussed later in this section.

Selection of the architecture, number of layers, activation functions, convolution methods and their sequences are selected empirically. Our model contains two networks: encoder and decoder network each of which contains residual blocks and skip connections. The structural information is an important feature to discriminate the near-similar classes. In middle layer, we use pyramid scene parsing layers which parse the middle layer features in two resolution types. Multi resolution scene parsing layers allow the model to learn the features at two levels.

1) ENCODER STRUCTURE

The encoder network contains the convolution, activation, batch-normalization and down-sample layers. The encoder network as shown in Figure 2 consists of 8 blocks where each block contains a specific number of convolution, activation and batch normalization layers. The number of filters in each convolution layer varies and is mentioned at the top of each layer. Each filter is 3×3 and we use *same* padding.

The average pooling layer squeezes the features spatially without affecting the depth of filters and forwards the down-sampled features to the batch normalization and convolutional layers. At the end of the encoder, sample remain at the lowest resolution so that they can capture high-level features regarding the shapes and structure.

We added two residual blocks in the encoder part to learn features without facing degradation problem. Each residual block consists of repetitive layer patterns where each pattern contains batch normalization, convolution and activation layers. Design choice of residual blocks and average pooling layer is discussed in Section II-E.

2) DECODER STRUCTURE

The second part of the proposed model is a decoder network (Figure 2) which learns the low-level features and gradually expands the features to the size of the masks. The decoder network consists of a defined combination of different layers: up-sampling, convolution with concatenation, batch normalization and convolution. The up-sampling layer expands the features in the spatial domain and feeds the features to the convolution and up-sampling layer. Convolutional layer extracts further features and concatenates these features with skip connection features. Further layers of batch normalization and convolution normalize the features and extract more low-level discriminant features. The decoder network consists of four blocks. The last layers of the decoder network contain only convolutional and a final *softmax* layer to map features to the generated masks. The decoder has three residual blocks. Two skip connections are being connected to the decoder. Concatenation skip connections are used to share the low level features with high level feature learning layers.

3) PYRAMID POOLING

To extract the most discriminative features, multi-level resolution features are also extracted by introducing the scene parsing layer. The scene parsing layer is added to the existing model with minimal modifications and shown in Figure 2. Middle part of the proposed model consists of pyramid scene parsing layers. The feature sets are provided to two pyramid channels of different kernels (see Figure 2). Each channel applies convolution and down sampling to extract features at different resolutions. With the idea that these channels learn the multi-resolution features for better generalization of the model, features from all three branches are mapped and concatenated in the decoder network.

E. NETWORK DESIGN CHOICES

The proposed DeepVeg model consists of a number of convolutional, activation, pooling, up/down-sampling and concatenation layers. The choice and composition of these layers

define a network. Architecture of deep learning networks may lead to several challenges which includes model complexity, learning capability, convergence of the objective function and performance. This section discuss the challenges and network design choices to address the underline challenges.

1) DIMINISHING FEATURE REUSE

Residual networks mainly contain identity skip connections to address the vanishing gradient problem [45]. However, identity skip connections can introduce the diminishing feature reuse problem and further increase the training time [46]. Diminishing feature reuse problem avoids the network to learn features in intermediate layers because skip connections provide input to intermediate layers which sometimes leads to forward pass vanishing gradients [41]. Sometimes, several residual blocks learn important and useful features whereas the other residual blocks only provide a small contribution to achieve the final goal [47]. To address these challenges, [48] proposed a Wide-ResNet model by incorporating more than one skip paths in each residual layer rather than focusing on the depth.

Thus, to avoid this problem, we propose residual connections without dropping the long skip connections (see Figure 3). The residual block in the proposed model enforces the weights to learn features in residual connections and shares these features later in the decoder part. Learning in this way allows the model to avoid diminishing feature reuse problem as each residual block learns separate features from the previous layer.

2) FEATURE SIZE REDUCTION

To get more discriminative or detailed features, multiple convolutional layers are connected sequentially. Multiple convolutional layers increase the number of parameters and hence the model complexity. To address this challenge, a non-linear down-sampling (pooling layer) method is used which reduces the feature space. We use 2×2 average pooling with stride 2. The input is partitioned into small rectangles with non-overlapping regions. Each small region output a single average value.

3) VANISHING GRADIENTS

Deep learning models can face vanishing gradient problem where the increment in the number of convolutional layers and activation decreases the gradients of the loss function, making the neural network hard to train [49]. To address the vanishing gradient problem, we use a Rectified Linear Unit (*ReLU*) activation in each convolution block. *ReLU* is a piece-wise activation function [50] and can be formulated as:

$$a_{i,j,k} = max(g_{i,j,k}, 0) \tag{2}$$

where $g_{i,j,k}$ is the input at channel k and $a_{i,j,k}$ is the activation output. The simplest operation $max(\cdot)$ in *ReLU* activation layer allows faster convergence as compared to *Sigmoid* [51] and *tanh* [52] activation layers. Back-propagation could suffer due to singularity at zero but *ReLU* outperforms the other activation functions (*tanh* and *sigmoid*) and allows the hidden states to incorporate sparsity and likelihood of vanishing gradients found to be reduced with *ReLU* [53].

4) COVARIATE SHIFT

In training, changing distributions at the input of each layer requires a low learning rate for convergence which slows down the training process due to internal covariate shift. To address this problem, we use batch normalization, as suggested in [54], after each convolutional layer where the distribution seems Gaussian and using activation (*ReLU*) results in a more stable output of the activation [55]. Batch normalization allows a higher learning rate which accelerates the network convergence. Further, batch normalization provides a soft edge to initialization and acts as regularization.

F. OBJECTIVE FUNCTION

This research focuses on the segmentation problem for weed, canola and canola flea beetle damage given the imbalanced class data with a large region of interest. The objective function can be defined in terms of loss function between predicted and ground truth masks $(y, \hat{y})$ where the aim is to minimize the loss term for the given dataset.

For imbalanced classes, Weighted Binary Cross Entropy (WBCE) loss is a modified version of Binary Cross Entropy (BCE) loss which assigns weights as per the class imbalance ratio. The imbalance ratio can be provided as an external parameter obtained from the entire dataset. WBCE loss can be formulated as [56]:

$$WBCE = -\frac{1}{M}\sum_{m=1}^{M}[w \times y_m \times log(h_\theta(x_m)) + (1 - y_m) \times log(1 - h_\theta(x_m))] \tag{3}$$

where w represents the positive class weight, M is the number of training samples, h_θ are the model weights, y_m and x_m are predictions and actual labels respectively.

To solve the imbalance class problem for more than two classes, categorical cross entropy with class weight assignment is required. Weighted Categorical Cross Entropy (WCCE) can be derived from WBCE function [56] as:

$$WCCE = -\frac{1}{M}\sum_{k=1}^{K}\sum_{m=1}^{M}[w_k \times y_m^k \times log(h_\theta(x_m, k))] \tag{4}$$

where y_m^k is the target label of example m for class k, w_k is the weight for class k, and K is the total number of classes.

III. EXPERIMENTATION

To address the segmentation problem, the proposed model is trained and optimized for the acquired dataset. For a fair evaluation of the proposed model, we also trained, optimized, and evaluated deep learning benchmark segmentation

models which include U-Net [38], U-Net ResNet-50 [40], SegNet [36], SegNet ResNet-50, DeepLabV3+ [43], DeepLabV3+ VGG-19.

Hyper-parameters of all experimented models are selected empirically by conducting multiple experiments for hyper-parameters optimization. To reduce the training time, pre-trained weights of VGG-19 and ResNet-50 are utilized and fine-tuned for the acquired dataset for all the models except the proposed model. The proposed model is trained from scratch without pretrained weights.

A. TRAINING MECHANISM

Initial weights of the proposed model are randomly initialized which follows a standard normal distribution. Each experiment conducted using the proposed model used the same seeds. The training process followed the following steps:

- Image and corresponding mask from the training dataset is divided into multiple non-overlapping chunks of dimensions $320 \times 320 \times 3$.
- The model learns the features against all the classes and after each epoch, it validates the learned features on validation dataset.
- Initially, the model generates shallow masks against the validation dataset and its loss logged as reference for the next epoch.
- The learning processes recurs multiple times and only stops if it meets the stopping criteria.

B. HYPER-PARAMETER OPTIMIZATION

The selection of hyper-parameters is a difficult task and requires several experiments. In this research, all the hyper-parameters are selected empirically after conducting several experiments to obtain the most optimal results.

Learning rate (η) determines the rate of model weights updates. Low learning rate can introduce local minima problem whereas larger learning rate can lead to model convergence issue. As a tradeoff between low and high learning values, experiments resulted ($\eta = 2 \times 10^{-4}$) as an optimal learning rate. In order to achieve global minima of the loss function, learning rate decay of value (75×10^{-2}) is utilized. This method decreases the learning rate value gradually as optimizer approaches to global minima of the loss function. We use Adam optimizer in our training. Furthermore, early stopping (ES) criteria selected by observing the training and validation loss are used to avoid over-fitting. The selected parameters for the proposed model and other benchmark models for are summarised in Table 1. We can see that the parameters of the proposed model are less compared to most of the benchmark models.

C. DATASET SPLITTING

The dataset contains 384 images with corresponding annotations. We divide the data into train, validation, and test sets using the split ratios of 70%, 10% and 20%, respectively. A uniform distribution is used to select random samples from the entire dataset for each set. The train and validation sets are used to train the models whereas the test set is never exposed in the training process. For evaluation and comparison of all the trained models, the test set is used to get predictions and evaluation metrics are computed on the test set.

D. SYSTEM CONFIGURATION

The proposed model and all other benchmark models are trained on a batch size of 2 image patches, which entirely depends on the GPU and system memory. To avoid noisy gradients, batch-wise upgrading of learnable parameters is performed. The models are trained using an NVIDIA 1080*Ti*, with CUDA 8.0 and cuDNN v5, using Keras with TensorFlow 2.2.0 backend.

E. EVALUATION METRICS

Intersection over Union (*IoU*), *Recall*, *Precision*, *Accuracy*, and $F1 - score$ metrics are used for evaluation and performance comparison of the proposed and benchmark models. To obtain results against these metrics, confusion matrix is computed for all four classes: weed, canola damage and background. A confusion matrix provides class-wise average quantitative measures in terms of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

TP provides the total count of correct predictions against the positive class and TN measures the total count of correct predictions against the negative class where positive and negative classes denote the target and remaining classes respectively. Similarly, FP provides the total count of wrong predictions against the positive class and FN measures the total count of wrong predictions against the negative class. *Precision* indicates the correct results given the total predictions by the model [57]. *Recall* can be defined as the total percentage of the correct results predicted by the model [57]. *Accuracy* is a measure of ratio between total count of correct predictions to the total number of pixels. Weighted average of *Precision* and *Recall* is used to calculate the $F1 - score$. *IoU* is obtained by dividing the overlap area with area of the union [58]. All five evaluation metrics are given below:

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$Accuracy = \frac{TP}{TP + FP} \tag{7}$$

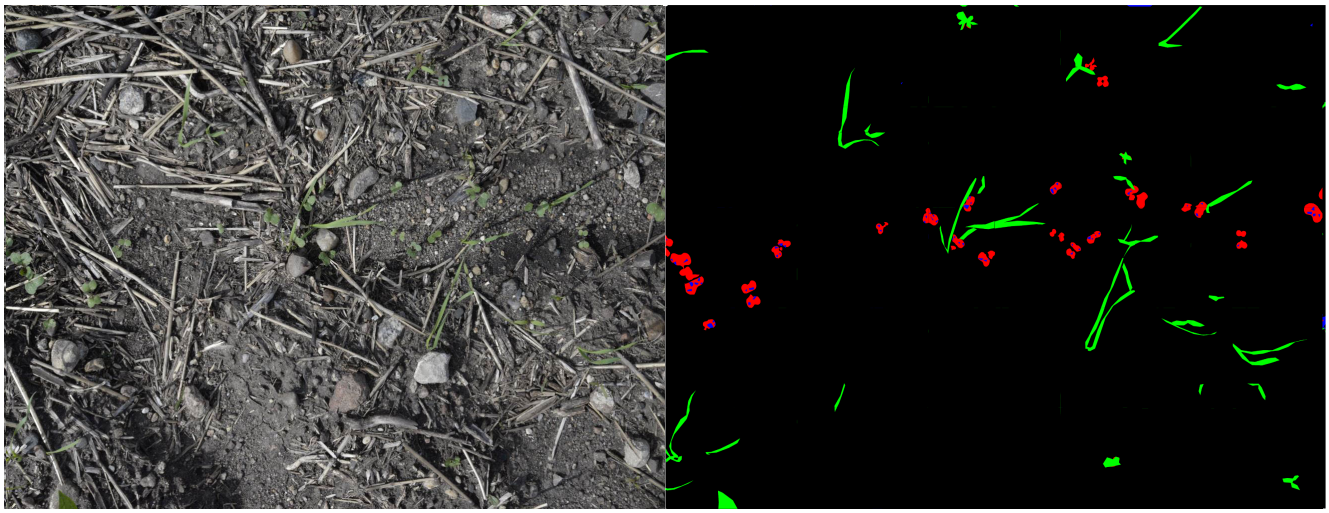
$$F1score = 2 \times \frac{Recall \times Precision}{Recall + Precision} \tag{8}$$

$$IoU = \frac{areaofintersection}{areaofunion} \tag{9}$$

Accuracy is a good measure for equally distributed classes but for the imbalanced classes, which is the case, $F1-score$ is comparatively a more useful measure. However, this research provides analysis based on all the aforementioned evaluation metrics obtained for test dataset.

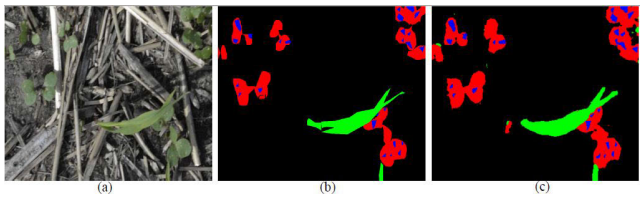
TABLE 1. Hyper-parameters selection for four class segmentation.

Model	Parameters						
	Epochs	Parameters	Batch size	Image size	Optimizer	Learning rate	Inference time (s)
U-Net	ES	4,472,960	2	800×1200	Adam	2×10^{-4}	0.77
U-Net ResNet-50	ES	16,374,596	2	800×1216	Adam	2×10^{-4}	0.93
SegNet	ES	3,698756	2	800×1200	Adam	2×10^{-4}	0.63
SegNet ResNet-50	ES	11,863,173	2	800×1216	Adam	2×10^{-4}	0.70
DeepLabV3+	ES	27,851,844	2	512×800	Adam	2×10^{-4}	0.66
DeepLabV3+ VGG19	ES	22,317,828	2	512×800	Adam	2×10^{-4}	1.08
Proposed	ES	4,250,720	2	320×320	Adam	2×10^{-4}	1.69



(a) (b)

FIGURE 4. (a) Test input image and (b) corresponding prediction of the proposed model. The predicted output includes weed (green), canola (red) and flea beetle (blue) damage classes. (Better view in color).



(a) (b) (c)

FIGURE 5. (a) Test input image (in chunk), (b) ground truth and (c) predicted output by the proposed model. The predicted output includes weed (green), canola (red) and flea beetle damage classes (blue). (Better view in color).

IV. RESULTS AND DISCUSSION

A. QUALITATIVE ANALYSIS

Segmentation of all the classes using the proposed model in a full image are shown in Figure 4, where, red colour represents canola plant, green colour represents weeds, blue dots are damages and black is background. In Figure 4 and Figure 5, it can be seen that the segmentation model is able to segment and differentiate the visual features of weed, damaged canola plants, and background.

The flea beetle damage can be categorized on the basis of segment appearance as the damage appear in dark or light small segments as shown in Figure 1. Combining these types of damages in a single class (damaged canola plants) and distinguishing these visual features from the background (soil) is a challenging task. The soil contains similar appearances in the form of small mud clusters making it difficult for the model to distinguish the damage.

One key point to look at is the shape of the damaged area. In early stages, the damage is of circular shape. With growth of damage the shape varies from circular to sometime oval shape or some arbitrary shape, but it is very rare that the damage will have rectangular or triangular shape. This raise the concern about the ground truth because the damage class labels are rough approximations. Figure 5 clearly depicts that when there are multiple damaged patches closer to each other and on a same leaf, the annotator collectively labels them as one blob. It is because manual labelling at this minimal level is hard. The cost of approximation appears in ground truth, it does not map and represents the exact area of interest(damage).

Looking at the predictions, the proposed model can draw boundary between separate spots and differentiate between two or multiple damaged area given the connected blob while training. Sharing the enriched features from encoder layers to layers in decoder using skip connections not only help in maintaining spatial stability but also assist in learning class

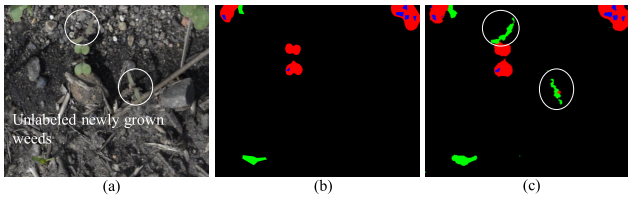


FIGURE 6. (a) Test input image (in chunk), (b) ground truth and (c) proposed model detected unlabeled newly grown weeds (green) correctly. (Better view in color).

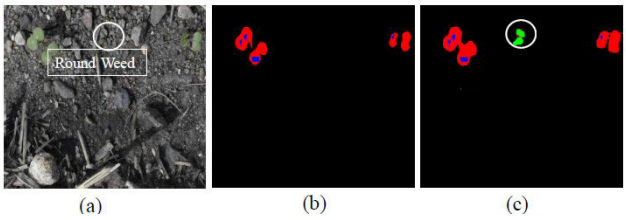


FIGURE 7. (a) Test input image (in chunk), (b) corresponding ground truth where weed labeling is missing and (c) proposed model detected the unlabeled weed as well as distinguished between the round weed and the same circular shaped canola. (Better view in color).

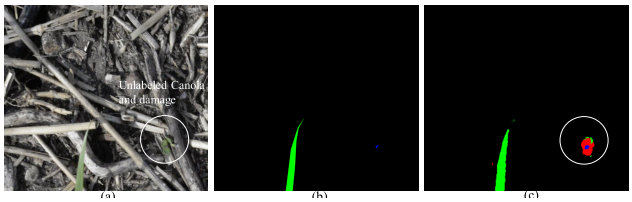


FIGURE 8. (a) Test input image (in chunk), (b) corresponding ground truth where canola and damage labeling is missing and (c) proposed model detected unlabeled canola (red) and damage (blue) properly. (Better view in color).

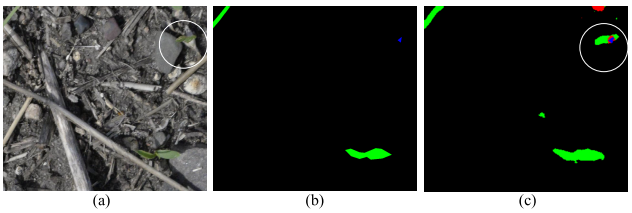


FIGURE 9. (a) Test input image (in chunk), (b) corresponding ground truth where canola is unlabeled and (c) proposed model detected unlabeled canola but mistaken canola as weed due to its long rectangular structure. (Better view in color).

specific features. The skip connection contains low and high level information which is helpful to differentiate the classes in a broader context. It is also helpful in stabilizing the visual cues.

Figure 5 and Figure 6 show that the weed and damage class are segmented as expected which ensures that the model has learned the features of all imbalanced classes without showing a tendency towards the most prominent class. It does not over-fit or show bias towards the most prominent class. The proposed model perform satisfactory and find the missing part of label of weed in Figure 6, where we can see that in the ground truth weeds are not labelled, but proposed model detected them correctly in the output prediction. It can also be seen that the proposed model predicts the newly grown weeds correctly which are not labelled in the ground truth. There are different shaped weeds in the dataset. The model shows outstanding performance in identifying different shaped weeds where it distinguishes between the circular shaped weed and the same shaped canola plant and detected the circular shaped weed perfectly as shown in Figure 7. In Figure 8 input image contains canola plant but the annotator missed these canola plant areas in the corresponding ground truth mask. Model segmented the non-annotated damaged canola plant correctly.

However, sometimes ground truth segmentation masks against the damaged canola plants overlaps on the healthy leaf area which can introduce learning uncertainty. As a result, this can increase the false positives which cause depreciation of class specific as well as overall segmentation performance. Sometimes there is reflection of light on canola plants. The folded leaves of canola showing resemblance to weed are also challenging for the model to classify.

In Figure 9 there is a canola plant which is predicted as weed which is the only case in total 384 images. The reason for this failure is that the canola plant leaves are showing resemblance to grassy weed (needle shape). Due to sameness with circular shape leaves the model confuses the crop with weed. Another possible argument can be the ratio of weeds available in dataset. As in most of the data samples the majority of weed is grassy, the model focuses more on grassy weed and learns better mapping. Nevertheless, the proposed model could detect the plants and damage although these were not labelled in the ground truth which proves that the proposed model learned the features in detail and can effectively detect weed, canola and damage from unlabelled or mislabelled ground truth. Figure 10 shows the predicted output of all the classes by the benchmark models and the proposed model.

B. QUANTITATIVE ANALYSIS

Given a complex background we aim to detect weed and damage in canola plants. The results and segmentation performance are evaluated for both weed and damage classes separately as well as collectively. The segmentation of weed and background seems not much challenging task as compared to the damage and canola plants leaves. The reason is the high class imbalance.

1) CLASS-WISE IoU PERFORMANCE EVALUATION

In terms of IoU, proposed model performs the best when compared to all other reported models for four class dataset. The proposed model also converges well during training and validation time as shown in Figure 11.

The IoU score on the test dataset with respect to each class are reported in Table 2. The proposed model reported class-wise values for weed, canola damage, canola plants

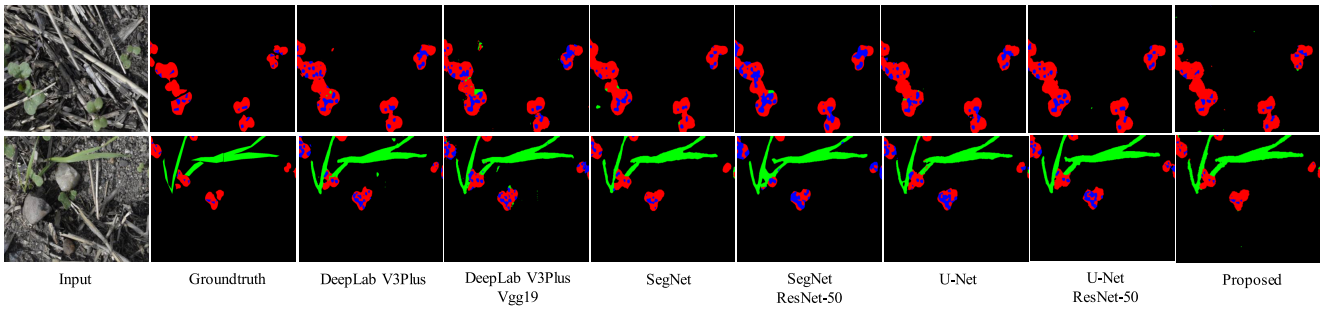


FIGURE 10. Predicted outputs by the benchmark models and the proposed model for 4 class (weed, canola, background and damage.

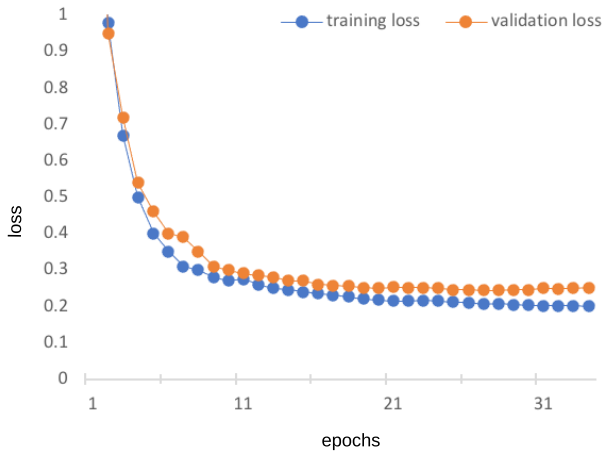


FIGURE 11. Training and validation losses for the proposed model.

and background are 0.657590, 0.689031, 0.743413 and 0.981655 accordingly. Also the proposed model reported the highest *IoU* score of 0.767922 compared to all the other benchmark models. SegNet performed the second best according to *IoU* score and class-wise values for canola damage, background and canola plant. The class-wise values for canola damage, canola plant and background reported by SegNet are 0.362581, 0.647704 and 0.982064 while the mean *IoU* score is 0.652096. SegNet ResNet-50 model reported second highest class-wise score for weed compared to all other models which is 0.643585 for weed and the *IoU* score is 0.593705 while at the same time it reported the least class-wise *IoU* score for canola plant and damages which are 0.607643 and 0.142007 accordingly. The promising encoder decoder U-Net and U-Net ResNet-50 however, reported a comparable mean *IoU* and class-wise *IoU* score. The mean *IoU* scores reported by both the models are 0.609552 and 0.633363. The DeepLabV3+ VGG19 reported the least scores for both class-wise *IoU* and mean *IoU*. The mean *IoU* score reported by this model is 0.564739. While DeepLabV3+ gave the third best class-wise *IoU* score for weed which is 0.636580 and this model reported comparable mean *IoU* value which is 0.616056.

The *IoU* score on the test dataset with respect to each class are reported in Table 2. The proposed model reported class-wise values for weed, canola damage, canola plants and background are 0.657590, 0.689031, 0.743413 and 0.981655 accordingly. Also the proposed model reported the highest *IoU* score of 0.767922 compared to all the other benchmark models. SegNet performed the second best according to *IoU* score and class-wise values for canola damage, background and canola plant. The class-wise values for canola damage, canola plant and background reported by SegNet are 0.362581, 0.647704 and 0.982064 while the mean *IoU* score is 0.652096. SegNet ResNet-50 model reported second highest class-wise score for weed compared to all other models which is 0.643585 for weed and the *IoU* score is 0.593705 while at the same time it reported the least class-wise *IoU* score for canola plant and damages which are 0.607643 and 0.142007 accordingly. The promising encoder decoder U-Net and U-Net ResNet-50 however, reported a comparable mean *IoU* and class-wise *IoU* score. The mean *IoU* scores reported by both the models are 0.609552 and 0.633363. The DeepLabV3+ VGG-19 reported the least scores for both class-wise *IoU* and mean *IoU*. The mean *IoU* score reported by this model is 0.564739. While DeepLabV3+ gave the third best class-wise *IoU* score for weed which is 0.636580 and this model reported comparable mean *IoU* value which is 0.616056.

The relatively low performance for damage is because of difficulty in labelling. It is very hard to label the damage at pixel level. We approximate the damage by drawing rough polygon around the area hoping that the model may understand the exact damage. Although DeepLabv3+ [43] is a benchmark model for semantic segmentation but for this task it is unable to achieve promising results. The reason for this is the last layer of DeepLabv3+ [43]. The last layer in DeepLabv3+ upsample the results by four times. In our dataset the size of damage segments is mostly very small in size. Upsampling the results four times will not help in achieving good performance for small objects. So, with DeepLabv3+ getting good results for small objects is difficult.

In case of damage detection, DeepLabv3+ [43] reports mean *IoU* score of 0.564739 with VGG-19 [59] as backend encoder. On the other hand, when ResNet-50 [40] is used as encoder it outperforms VGG-19 reporting 0.616056 mean *IoU* score. The possible reason is residual connections in

ResNet blocks which particularly targets the problem of vanishing gradient. SegNet with ResNet-50 outperforms DeepLabv3+ based model. Both SegNet based models reported *IoU* score for damage class of 0.362581 and 0.142007 for simple and ResNet-50 based models respectively. SegNet [36] based model is efficient in terms of memory but not in performance. The skip connections and the residual blocks in the proposed model are more enriched and share all the low and high level information from encoder to decoder part of the network. Learning of the spatial information is insufficient by U-Net network. Hence, only a small part of the damage can be segmented. SegNet based model only shares the pooling indices which may not be enough in maintaining spatial stability and useful information. As discussed earlier the visual cues for damage detection are very similar to mud clusters. The damage on canola plants also show similarities with dead plants and damage on the weedy plants. Under the given circumstances it is hard for the model to understand damage for the canola plant.

2) RESULT ANALYSIS BASED ON PRECISION-RECALL, $F1-score$ AND ACCURACY VALUE

Table 3 summarizes the test evaluation results for the proposed and the bench-mark models based on *Precision*, *Recall*, *Accuracy* and $F1-score$ class-wise.

For damage class, the proposed model scores the most for *Precision* which is 0.754461, which means it segmented most of the positive predictions. The *Recall* score is almost same as the *Precision* score for the proposed model which is 0.72199. The U-Net and SegNet with residual connections scored more for the *Recall* in damage class which are respectively 0.888542 and 0.886655. For canola plant class, the proposed model returned the highest *Recall* value which is 0.982860 whereas the *Precision* score is 0.702626.

For the weed class the U-Net ResNet-50 model resulted the highest *Recall* value which is 0.990417 and the second highest *Recall* value for the weed class is obtained by the proposed model which is 0.977805. The highest *Precision* value for the weed class is obtained by the SegNet ResNet-50 model which is 0.633927. The proposed model scored the highest $F1-score$ for the weed, damage and the canola plant class which are respectively 0.745532, 0.737873 and 0.819447. The *Accuracy* score for the proposed model is also the highest compared to other benchmark models which is 0.976163. From the scores we can clearly see that the proposed model outperformed the other benchmark models with every evaluation metrics.

As suggested, the $F1-score$ and *IoU* are good measures for segmentation evaluation of imbalanced classes. U-Net, U-Net ResNet-50, SegNet, SegNet ResNet-50 and DeepLabV3+ perform very close in terms of *Accuracy* by resulting 0.974470, 0.972529, 0.976519, 0.974506 and 0.972529 scores respectively. The DeepLabV3+ with VGG-19 pretrained weights performed worst as compared to other benchmark deep learning models and resulted in a 0.968581 *Accuracy* score. In contrast, the proposed model outperformed all the benchmark deep learning models and reported the highest *Accuracy* score. This class-wise result evaluation is not a good measure as compared to overall evaluation. *Recall* and *IoU* measures ensure that the proposed model is able to segment out damaged canola plants in a better way and can assist the pathologists to analyse and diagnose the plant disease at an extensive level.

TABLE 2. Class-wise evaluation for IoU score and mIoU for four class (background, damage, canola plant, and weed) of the proposed and benchmark models.

Model	Class	Scores	
		IoU	mIoU
U-Net	background	0.9806881	0.6095523
	damage	0.20736236	
	weed	0.6211705	
	canola plant	0.62899041	
U-Net ResNet-50	background	0.97683465	0.6333630
	damage	0.30433557	
	weed	0.61820849	
	canola plant	0.63407338	
DeepLabV3+ VGG19	background	0.97626462	0.5647390
	damage	0.13456267	
	weed	0.56724904	
	canola plant	0.58087994	
DeepLabV3+	background	0.98041654	0.6160568
	damage	0.22188479	
	weed	0.62534554	
	canola plant	0.63658048	
SegNet	background	0.98206439	0.6520968
	damage	0.36258132	
	weed	0.61603673	
	canola plant	0.64770493	
SegNet ResNet-50	background	0.98158478	0.5937053
	damage	0.14200783	
	weed	0.64358522	
	canola plant	0.60764341	
Proposed	background	0.98165512	0.7679226
	damage	0.68903114	
	weed	0.65759061	
	canola plant	0.74341364	

C. OVERALL PERFORMANCE EVALUATION

From the summarised results in Table 2 and Table 3, we can say that $F1-score$ and *IoU* metrics are suggested to evaluate the results for imbalanced classes. As compared to *Precision*, *Recall*, and *Accuracy*, the $F1-score$ and *IoU* score weigh more in our case. Assigning more weight to minority class helps the model to more effectively segment and understand the difficult examples. In our case we use the weighted loss to help the model to understand and detect the minority class better irrespective of the data distribution. The train dataset contains some unlabelled canola flea beetle damages. The so trained models when tested on different data, it points out most of the mislabelled classes. This shows that models are generalizing well to unknown and known data and can be helpful to correct the ground truths, this is shown in Figure 5. Assigning labels at pixel level for damaged area is very

TABLE 3. Class-wise evaluation score for four class (background, damage, canola plant, and weed) of proposed model and the benchmark models.

Model	Class	Scores			
		precision	recall	f1	Accuracy
U-Net	background	0.999520	0.977967	0.988626	0.974470
	damage	0.220527	0.816903	0.347299	
	weed	0.608626	0.947334	0.741115	
	canola plant	0.697537	0.853285	0.767590	
U-Net ResNet 50	background	0.999834	0.973689	0.986588	0.972529
	damage	0.338757	0.888542	0.490508	
	weed	0.587898	0.990417	0.737830	
	canola plant	0.667880	0.904800	0.768495	
DeepLabV3+ VGG19	background	0.999639	0.973589	0.986442	0.968581
	damage	0.135995	0.852610	0.234575	
	weed	0.593103	0.857350	0.701156	
	canola plant	0.618378	0.867795	0.722158	
DeepLabV3+	background	0.999486	0.978415	0.988838	0.975223
	damage	0.243619	0.813756	0.374978	
	weed	0.614604	0.957388	0.748623	
	canola plant	0.702883	0.858023	0.772743	
SegNet	background	0.999427	0.979761	0.989496	0.976519
	damage	0.407493	0.767056	0.532238	
	weed	0.601176	0.961891	0.739911	
	canola plant	0.717368	0.856323	0.780711	
SegNet ResNet-50	background	0.999488	0.979277	0.989279	0.974506
	damage	0.146862	0.886655	0.251986	
	weed	0.633927	0.948374	0.759905	
	canola plant	0.721711	0.790052	0.754336	
Proposed	background	0.999852	0.976295	0.987933	0.976563
	damage	0.754461	0.721999	0.737873	
	weed	0.602428	0.977805	0.745532	
	canola plant	0.702626	0.982860	0.819447	

challenging. The model trained on noisy labels can perform really promising and help us to fine tune the ground truth.

The proposed model trained on such data understands these challenges and works pretty well in removing discontinuities and detecting damage plants which are not in the ground truth. The proposed model could extract out the weed, canola and damage properly from background regardless of the class distribution. All models are able to identify the boundaries for each damage patch and classify them as separate patches. The proposed model outperforms in detecting the discrete boundary and indicating the mislabelled weed and canola as shown in Figure 6 and Figure 8. This proposed model is a single model to promisingly perform in the detection of early stage damages and different shapes/ types of weeds at the same time.

Other than the proposed model, all the models fail to detect the damage under occluded crop plant and mostly classify the crop plant as weed.

While analysing the model architecture, the proposed model is having less number of parameters and layers which make the model structure comparatively small than the other benchmark models and hence the computation time is also less.

In a broader context, the obtained results support the quantitative analysis and depict that the proposed model with pyramid pooling and residual blocks connected encoder decoder combined with weighted loss function can learn the least discriminant features. Furthermore, this analysis also makes certain that using WCCE loss is helpful to understand imbalanced class data.

D. COMPUTATIONAL COMPLEXITY

Computational complexity of deep learning models can be quantified in terms of training time and inference time. With the advancement in computation power and hardware accelerated devices, training time of deep learning models is not crucial. However, the inference time is an important factor. The number of model parameters and average inference time (based on 1475 images) for all the evaluated models are shown in Table 1. The system configuration (see Section III-D) is kept constant for all the models.

We can see from Table 1 that U-Net ResNet-50 has four times more parameters compared to U-Net but its inference time is only about 20% higher. On the other hand, DeepLabV3+ has more parameters but lower inference time compared to DeepLabV3+ VGG19. In terms of computational complexity, the proposed model and U-Net have almost equal number of parameters and the proposed model has higher inference time. However, the proposed model has better accuracy and mIoU score compared to U-Net. This analysis shows that computational complexity of deep learning

models is not only dependent on the number of parameters but also on the model architecture.

E. LIMITATION

Image segmentation is a very difficult process due to nature of objects imaged where they often do not have very well defined boundaries in images. Poor lighting, poorly contrasted background and limited information on true boundaries make it difficult for a segmentation algorithm to generate an accurate boundary. In such circumstances, segmentation errors are to be expected due to wrongly assigned pixels to different region which results in the imprecise definition of region in images.

V. CONCLUSION

Weed can reduce the crop yield in multiple ways and detection of weed plants is a crucial step for crop yield analysis. In canola plants, flea beetle is an another factor that impacts the canola yield by damaging the canola leaves in early stages. Segmentation of weed and canola flee beetle damage involves multiple challenges: highly imbalanced classes, overlapped visual features of weed and canola flee beetle damaged plants, and the complex soil background. We address these challenges by proposing a unified deep learning model which is also smaller in size and parameters in comparison to other benchmark models. The proposed model utilize multiple residual and skip connections to learn most discriminative features. We also trained and evaluated benchmark deep learning models (U-Net, U-Net ResNet-50, SegNet, SegNet ResNet-50, DeepLabV3+, DeepLabV3+ VGG-19). Obtained results on the acquired dataset showed that the proposed model is competitive with benchmark state-of-the-art deep learning techniques. designing of a single unique segmentation model which can accurately detect the early stage damages and different types of weeds simultaneously has not been carried out yet.

However, there is room for further improvements in segmentation of such complex and high imbalance dataset. Generative adversarial networks can be deployed for synthetic dataset generation. Labelling can be better with oval or circular shape as the damage which can improve the learning process. NIR images can be incorporated to make the system more generic. Group normalization can also be utilized to overcome small batch size limitations.

ACKNOWLEDGMENT

The authors would like to thank Mr. Kevin Shearer of SAIRS Ltd. (http://www.sairs.ca/) for his efforts in high-resolution data collection of canola fields at two different growth stages. The authors would also like to thank Mr. Brett Bauereiss of JB Farms for helping in this project by providing access to his canola fields.

REFERENCES

[1] W. S. Lee, V. Alchanatis, C. Yang, M. Hirafuji, D. Moshou, and C. Li, "Sensing technologies for precision specialty crop production," *Comput. Electron. Agricult.*, vol. 74, no. 1, pp. 2–33, Oct. 2010.

[2] A. Wang, W. Zhang, and X. Wei, "A review on weed detection using ground-based machine vision and image processing techniques," *Comput. Electron. Agricult.*, vol. 158, pp. 226–240, Mar. 2019.

[3] J. Gao, W. Liao, D. Nuyttens, P. Lootens, J. Vangeyte, A. Pižurica, Y. He, and J. G. Pieters, "Fusion of pixel and object-based features for weed mapping using unmanned aerial vehicle imagery," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 67, pp. 43–53, May 2018.

[4] E. Hamuda, M. Glavin, and E. Jones, "A survey of image processing techniques for plant extraction and segmentation in the field," *Comput. Electron. Agricult.*, vol. 125, pp. 184–199, Jul. 2016.

[5] D. C. Slaughter, D. K. Giles, and D. Downey, "Autonomous robotic weed control systems: A review," *Comput. Electron. Agricult.*, vol. 61, no. 1, pp. 63–78, 2008.

[6] J. Soroka, L. Grenkow, J. Otani, J. Gavloski, and O. Olfert, "Flea beetle (Coleoptera: Chrysomelidae) species in canola (Brassicaceae) on the northern Great Plains of North America," *Can. Entomol.*, vol. 150, no. 1, pp. 100–115, Feb. 2018.

[7] R. Lamb, "Assessing the susceptibility of crucifer seedlings to flea beetle (Phyllotreta spp.) damage," *Can. J. Plant Sci.*, vol. 68, no. 1, pp. 85–93, 1988.

[8] H. F. Abouziena and W. M. Haggag, "Weed control in clean agriculture: A review1," *Planta Daninha*, vol. 34, no. 2, pp. 377–392, Jun. 2016.

[9] P. Lottes, J. Behley, A. Milioto, and C. Stachniss, "Fully convolutional networks with sequential information for robust crop and weed detection in precision farming," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 2870–2877, Oct. 2018.

[10] M. Minervini, H. Scharr, and S. A. Tsaftaris, "Image analysis: The new bottleneck in plant phenotyping [applications corner]," *IEEE Signal Process. Mag.*, vol. 32, no. 4, pp. 126–131, Jul. 2015.

[11] D. James, A.-M. Schmidt, E. Wall, M. Green, and S. Masri, "Reliable detection and identification of genetically modified maize, soybean, and canola by multiplex PCR analysis," *J. Agricult. Food Chem.*, vol. 51, no. 20, pp. 5829–5834, Sep. 2003.

[12] V. N. T. Le, B. Apopei, and K. Alameh, "Effective plant discrimination based on the combination of local binary pattern operators and multiclass support vector machine methods," *Inf. Process. Agricult.*, vol. 6, pp. 116–131, Mar. 2019.

[13] B. Pommerrenig, A. Junker, I. Abreu, A. Bieber, J. Fuge, E. Willner, M. D. Bienert, T. Altmann, and G. P. Bienert, "Identification of rapeseed (Brassica napus) cultivars with a high tolerance to boron-deficient conditions," *Frontiers Plant Sci.*, vol. 9, p. 1142, Aug. 2018.

[14] S. Bagde, S. Patil, S. Patil, and P. Patil, "Artificial neural network based plant leaf disease detection," *Int. J. Comput. Sci. Mobile Comput.*, vol. 4, no. 4, pp. 900–905, 2015.

[15] S. Weizheng, W. Yachun, C. Zhanliang, and W. Hongda, "Grading method of leaf spot disease based on image processing," in *Proc. Int. Conf. Comput. Sci. Softw. Eng.*, vol. 6, 2008, pp. 491–494.

[16] S. Phadikar, J. Sil, and A. K. Das, "Rice diseases classification using feature selection and rule generation techniques," *Comput. Electron. Agricult.*, vol. 90, pp. 76–85, Jan. 2015.

[17] A. Rastogi, R. Arora, and S. Sharma, "Leaf disease detection and grading using computer vision technology & fuzzy logic," in *Proc. 2nd Int. Conf. Signal Process. Integr. Netw. (SPIN)*, Feb. 2015, pp. 500–505.

[18] S. Prasad, S. K. Peddoju, and D. Ghosh, "Multi-resolution mobile vision system for plant leaf disease diagnosis," *Signal, Image Video Process.*, vol. 10, no. 2, pp. 379–388, Feb. 2016.

[19] S. B. Dhaygude and N. P. Kumbhar, "Agricultural plant leaf disease detection using image processing," *Int. J. Adv. Res. Elect., Electron. Instrum. Eng.*, vol. 2, no. 1, pp. 599–602, 2013.

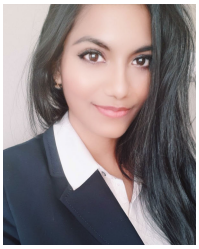
[20] S. Arivazhagan, R. N. Shebiah, S. Ananthi, and S. V. Varthini, "Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features," *Agricult. Eng. Int., CIGR J.*, vol. 15, no. 1, pp. 211–217, 2013.

[21] A. H. Kulkarni and A. Patil, "Applying image processing technique to detect plant diseases," *Int. J. Mod. Eng. Res.*, vol. 2, no. 5, pp. 3661–3664, 2012.

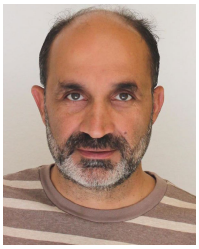
[22] O. Lundin, Å. Myrbeck, and R. Bommarco, "The effects of reduced tillage and earlier seeding on flea beetle (Phyllotreta spp.) crop damage in spring oilseed rape (Brassica napus L.)," *Crop Protection*, vol. 107, pp. 104–107, May 2018.

[23] P. Lottes, R. Khanna, J. Pfeifer, R. Siegwart, and C. Stachniss, "UAV-based crop and weed classification for smart farming," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2017, pp. 3024–3031.

[24] M. H. Asad and A. Bais, "Weed detection in canola fields using maximum likelihood classification and deep convolutional neural network," *Inf. Process. Agricult.*, vol. 7, no. 4, pp. 535–545, Dec. 2020.
[25] A. Milioto, P. Lottes, and C. Stachniss, "Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in CNNs," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2018, pp. 2229–2235.
[26] C. Potena, D. Nardi, and A. Pretto, "Fast and accurate crop and weed identification with summarized train sets for precision agriculture," in *Proc. Int. Conf. Intell. Auton. Syst.* Cham, Switzerland: Springer, 2016, pp. 105–121.
[27] M. H. Asad and A. Bais, "Crop and weed leaf area index mapping using multi-source remote and proximal sensing," *IEEE Access*, vol. 8, pp. 138179–138190, 2020.
[28] R. Shinde, C. Mathew, and C. Patil, "Segmentation technique for soybean leaves disease detection," *Int. J. Advance Res.*, vol. 3, no. 5, pp. 522–528, 2015.
[29] C. Xia, J.-M. Lee, Y. Li, Y.-H. Song, B.-K. Chung, and T.-S. Chon, "Plant leaf detection using modified active shape models," *Biosyst. Eng.*, vol. 116, no. 1, pp. 23–35, 2013.
[30] S. Bankar, A. Dube, P. Kadam, and S. Deokule, "Plant disease detection techniques using canny edge detection & color histogram in image processing," *Int. J. Comput. Sci. Inf. Technol.*, vol. 5, no. 2, pp. 1165–1168, 2014.
[31] A. Shire, U. Jawarkar, and M. Manmode, "A review paper on: Agricultural plant leaf disease detection using image processing," *Int. J. Innov. Sci. Eng. Technol.*, vol. 2, no. 1, pp. 282–285, 2015.
[32] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.
[33] X. Jin and J. Han, *K-Means Clustering*. Boston, MA, USA: Springer, 2016, pp. 1–3, doi: 10.1007/978-1-4899-7502-7_431-1.
[34] J. Gui, L. Hao, Q. Zhang, and X. Bao, "A new method for soybean leaf disease detection based on modified salient regions," *Int. J. Multimedia Ubiquitous Eng.*, vol. 10, no. 6, pp. 45–52, 2015.
[35] N. N. Kurniawati, S. N. H. S. Abdullah, S. Abdullah, and S. Abdullah, "Investigation on image processing techniques for diagnosing paddy diseases," in *Proc. Int. Conf. Soft Comput. Pattern Recognit.*, 2009, pp. 272–277.
[36] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
[37] A. Khan, T. Ilyas, M. Umraiz, Z. I. Mannan, and H. Kim, "CED-Net: Crops and weeds segmentation for smart farming using a small cascaded encoder-decoder architecture," *Electronics*, vol. 9, no. 10, p. 1602, Oct. 2020.
[38] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.
[39] H. S. Ullah, M. H. Asad, and A. Bais, "End to end segmentation of canola field images using dilated U-Net," *IEEE Access*, vol. 9, pp. 59741–59753, 2021.
[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[41] G. Huang, Y. Sun, Z. Liu, D. Sedra, and K. Q. Weinberger, "Deep networks with stochastic depth," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 646–661.
[42] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "LabelMe: A database and web-based tool for image annotation," *Int. J. Comput. Vis.*, vol. 77, nos. 1–3, pp. 157–173, 2008.
[43] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Computer Vision—ECCV*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Cham, Switzerland: Springer, 2018, pp. 833–851.
[44] M. Long, H. Zhu, J. Wang, and M. I. Jordan, "Unsupervised domain adaptation with residual transfer networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 136–144.
[45] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 630–645.
[46] N. Khurshid, M. Tharani, M. Taj, and F. Z. Qureshi, "A residual-dyad encoder discriminator network for remote sensing image matching," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 2001–2014, Mar. 2020.
[47] Z. Wu, C. Shen, and A. Van Den Hengel, "Wider or deeper: Revisiting the ResNet model for visual recognition," *Pattern Recognit.*, vol. 90, pp. 119–133, Jun. 2019.
[48] S. Zagoruyko and N. Komodakis, "Wide residual networks," 2016, *arXiv:1605.07146*. [Online]. Available: http://arxiv.org/abs/1605.07146
[49] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *Uncertain. Fuzziness Knowl.-Based Syst.*, vol. 6, no. 2, pp. 107–116, 1998.
[50] A. F. Agarap, "Deep learning using rectified linear units (ReLU)," 2018, *arXiv:1803.08375*. [Online]. Available: http://arxiv.org/abs/1803.08375
[51] J. Han and C. Moraga, "The influence of the sigmoid function parameters on the speed of backpropagation learning," in *Proc. Int. Workshop Artif. Neural Netw.* Berlin, Germany: Springer, 1995, pp. 195–201.
[52] B. L. Kalman and S. C. Kwasny, "Why tanh: Choosing a sigmoidal function," in *Proc. Int. Joint Conf. Neural Netw.*, vol. 4, Jun. 1992, pp. 578–581.
[53] S. Li, W. Li, C. Cook, C. Zhu, and Y. Gao, "Independently recurrent neural network (IndRNN): Building a longer and deeper RNN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5457–5466.
[54] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, "How does batch normalization help optimization?" in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 2483–2493.
[55] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
[56] Y. Ho and S. Wookey, "The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling," *IEEE Access*, vol. 8, pp. 4806–4813, 2020.
[57] K. M. Ting, *Precision and Recall*. Boston, MA, USA: Springer, 2010, p. 781, doi: 10.1007/978-0-387-30164-8_652.
[58] M. A. Rahman and Y. Wang, "Optimizing intersection-over-union in deep neural networks for image segmentation," in *Proc. Int. Symp. Vis. Comput.* Cham, Switzerland: Springer, 2016, pp. 234–244.
[59] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556



MOHANA DAS received the Bachelor of Technology degree in electrical engineering from the National Institute of Technology Rourkela and the M.A.Sc. degree from the University of Regina, SK, Canada. She worked on canola, crop damage, and weed detection for the M.A.Sc. thesis. For her undergraduate capstone project, she worked on the CORDIC algorithm and signal processing. Her research interests include deep learning, computer vision, and machine learning.



ABDUL BAIS (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from the University of Engineering and Technology, Peshawar, Pakistan, and the Ph.D. degree in electrical engineering and information technology from Vienna University of Technology Vienna, Austria. From 2010 to 2013, he was a Postdoctoral Fellow with the Faculty of Engineering and Applied Science, University of Regina, SK, Canada, where he is currently an Associate Professor with the Electronic Systems Engineering Program. His research interests include real-time data stream mining, deep learning/machine learning, image processing, and computer vision. His current research projects include multi-source data processing for precision agriculture, accent and language-based speaker recognition for forensic applications, and detection of cohesive communities in multimedia images. His research is supported by The Natural Sciences and Engineering Research Council of Canada (NSERC Alliance and Discovery Programs), Mitacs, and AgTech industry. He is a Certified Instructor with NVIDIA Deep Learning Institute (Fundamentals of Deep Learning and Fundamentals of Accelerated Data Science with RAPIDS). He is a Licensed Professional Engineer in SK.