Semantic Segmentation With Labeling Uncertainty and Class Imbalance

Patrik Olã Bressan
Federal University of Mato Grosso do Sul
José Marcato Junior
Federal University of Mato Grosso do Sul
José Augusto Correa Martins
Federal University of Mato Grosso do Sul
Maximilian Jaderson Melo
Federal University of Mato Grosso do Sul
Diogo Nunes Gonçalves
Federal University of Mato Grosso do Sul
Daniel Matte Freitas
Federal University of Mato Grosso do Sul
Ana Paula Marques Ramos (anaramos@unoeste.br)
University of Western S ˜ ao Paulo (UNOESTE)
Lucas Prado Osco
University of Western S ˜ ao Paulo (UNOESTE)
Jonathan Andrade Silva
Federal University of Mato Grosso do Sul
Zhipeng Luo
Xiamen University
Jonathan Li
University of Waterloo (UW)
Raymundo Cordero Garcia
Federal University of Mato Grosso do Sul
Wesley Nunes Gonçalves
Federal University of Mato Grosso do Sul

Research Article

Keywords: Convolutional Neural Networks (CNN) , pixel-labeling process, aerial, terrestrial

Posted Date: April 14th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-409625/v1

Semantic Segmentation With Labeling Uncertainty and Class Imbalance

Patrik Olã Bressan[1,2], José Marcato Junior[3], José Augusto Correa Martins[3], Maximilian Jaderson de Melo[1], Diogo Nunes Gonçalves[1], Daniel Matte Freitas[1], Ana Paula Marques Ramos[4,*], Lucas Prado Osco[5], Jonathan de Andrade Silva[1], Zhipeng Luo[6], Jonathan Li[7], Raymundo Cordero Garcia[3], and Wesley Nunes Gonçalves[1,3,*]

[1]Faculty of Computer Science, Federal University of Mato Grosso do Sul, Av. Costa e Silva, Campo Grande 79070-900, Brazil
[2]Federal Institute of Mato Grosso do Sul, Jardim 79240-000, Brazil
[3]Faculty of Engineering, Architecture, and Urbanism and Geography, Federal University of Mato Grosso do Sul, Av. Costa e Silva, Campo Grande 79070-900, Brazil
[4]Post-Graduate Program in Environment and Regional Development, University of Western São Paulo (UNOESTE), Rodovia Raposo Tavares, km 572-Limoeiro, 19067-175, Presidente Prudente, São Paulo, Brazil
[5]Faculty of Engineering and Architecture and Urbanism, University of Western São Paulo (UNOESTE), Rodovia Raposo Tavares, km 572-Limoeiro, 19067-175, Presidente Prudente, São Paulo, Brazil
[6]School of Informatics, Xiamen University, Xiamen FJ 361005, China
[7]Department of Geography and Environmental Management, University of Waterloo (UW), Waterloo, ON N2L 3G1, Canada
[*]wesley.goncalves@ufms.br, anaramos@unoeste.br

ABSTRACT

Recently, methods based on Convolutional Neural Networks (CNN) achieved impressive success in semantic segmentation tasks. However, challenges such as the class imbalance and the uncertainty in the pixel-labeling process are not completely addressed. As such, we present a new approach that calculates a weight for each pixel considering its class and uncertainty during the labeling process. The pixel-wise weights are used during training to increase or decrease the importance of the pixels. Experimental results using different datasets like aerial, terrestrial, and ultrasound images show that the proposed approach leads to significant improvements in three challenging segmentation tasks in comparison to baseline methods. It was also proved to be more invariant to noise. The approach presented here may be used within a wide range of semantic segmentation methods to improve their robustness.

Introduction

Semantic segmentation is the task of dividing an image into regions whose pixels in the same area have similar properties, whether in color, texture or belonging to the same object. This task is crucial to infer knowledge of a scene in computer vision systems, as shown in tasks of tree species segmentation[1]. Recently, significant advances in semantic segmentation have been achieved through Convolutional Neural Networks (CNN), including methods such as SegNet[2], Fully Convolutional Network (FCN) [3], and DeepLabv3+ [4].

Despite recent advances, two factors have been little explored in the literature during the training of CNNs for semantic segmentation. The first factor is the unbalance of class distribution, where dominant portions of the data are assigned to a few classes while many classes have little representation in the data. As a consequence, semantic segmentation methods are biased to the dominant classes during the inference[5]. One way to minimize imbalance is by uniformly sampling data and collecting images (such as well-known image datasets, ImageNet[6,7], MNIST (Modified National Institute of Standards and Technology)[8] and CIFAR 10/100), under-sampling the majority classes[9–12], or over-sampling the minority classes[13–16]. However, these approaches change the distribution of data and can affect learning and inference[17].

The second factor, much less explored in the literature, is related to uncertainty in image labeling[18,19]. In low resolution or noisy images, the edges of objects become inaccurate and even expert labeling may include annotation errors that impact training. Even in high-resolution images, some objects (e.g., trees[1]) have complex edges that make them difficult to annotate.

To overcome the aforementioned issues, we propose an approach to deal with class unbalance and uncertainty in the labeling process for image segmentation tasks. For this approach, we introduce a loss function where the contribution of each pixel is

weighted. First, pixels belonging to minority classes have their importance increased. Second, pixels near the edges of the object generally have greater uncertainty during labeling and thus have their importance diminished during training. These two pixel-wise weights are combined and produce a great impact during training and inference of the segmentation methods. We evaluate our approach in three datasets which contains the challenges mentioned above. Experimental results show that the proposed approach provides significant increments when compared to the baseline and state-of-the-art methods. The following sections present the revision of the related works (section 2), the methodological approach proposed (section 3), being followed by experiments and results (section 4), and the conclusion (section 5).

Related works

Imbalance Data

In semantic segmentation, some approaches have been proposed to deal with class imbalance. Traditional approaches use resampling (e.g. oversampling and undersampling) and rebalancing schemes via statistic analysis, such as inverse or median frequency[20–22]. Despite correcting the imbalance, these approaches include several disadvantages. Oversampling methods increase computational cost and may be more prone to overfitting due to the inclusion of duplicate data. On the other hand, undersampling methods can discard important data for learning. Approaches are also based on constraints during training, such as restricting the number of pixels contributing to the loss function during backpropagation at random[23], based on the k highest loss of the pixels[24] or hard samples[25]. Huang et al.[26] reduced the effect of class imbalance by enforcing inter-cluster and inter-class margins in standard deep learning frameworks. These margins can be applied through quintuplet instance sampling and the associated triple-header hinge loss. Ren et al.[27] proposed a meta-learning framework that assigns weights to training examples based on their gradient directions to reduce class imbalance and corrupted label problems. Recently, focal loss[28] was proposed in order to penalize hard samples assuming that they belong to the minority class. However, this does not happen when minority classes are well defined and may not have their participation in training effectively. A survey on deep learning with class imbalance can be found in[29].

Labeling Uncertainty

Labeling uncertainty is related to image resolution and object-edge complexity. Similar to this work, Bischke et al.[19] applied an adaptive uncertainty weighted class loss to segment satellite imagery. However, only the uncertainty of the class is considered and not the uncertainty of every single pixel, as proposed in this work. Bulò et al.[18] proposed a max-pooling loss that adaptively re-weights the contributions of each pixel based on their observed losses. However, this method does not consider objects whose edges are not well defined and therefore present uncertainties during labeling. Ding et al.[30] proposed learning boundary objects as an additional class to increase the feature similarity of the same object. Similarly, Shen et al.[31] addressed the contour detection problem by combining a loss function for contour versus non-contour samples. The labeling uncertainty problem is also related to the size of the object in the image since small objects are harder to label. Islam et al.[32] proposed a new CNN architecture to predict segmentation labels at several resolutions. A loss function at each stage (scale) provides supervision to improve detail on segmentation labels. Although it improves the segmentation of object edges, labeling uncertainty is still a problem that degrades the result. Hamaguchi et al.[33] proposed a novel architecture called local feature extraction which aggregates local features with decreasing dilation factor to segment small objects in remote sensing imagery.

Experiments and Results

Image Dataset

Three image datasets were used in this study to demonstrate the robustness of our method. These datasets, described below, have the challenges of class imbalance and labeling uncertainty.

Urban Tree (UT). This dataset is composed of aerial RGB orthoimages generated with a GSD (Ground Sample Distance) of 10 cm from Campo Grande municipality in Brazil. Examples of Urban Tree dataset in Fig. 1 show that the boundaries of the trees are difficult to label. This dataset is composed by 966 non-overlapping patches of 256×256 pixels. In the experiments, 580, 193 and 193 patches were used for training, validation and testing, respectively.

Rib Eye Area (REA). This image dataset consists of ultrasound images of the Longissimus dorsi muscle, between the 11th and 13th ribs of cattle. The goal is to automatically calculate the rib eye area (REA), an important region for decision making during cattle breeding. The main challenge is the uncertainty in the REA annotation, since the image is noisy and even experts have difficulty in delimiting the borders of this region. Fig. 2 presents examples of images and the annotation made by a specialist. We can observe that some borders are absent and depend on the subjectivity and knowledge of the annotator. To evaluate the segmentation methods, 76 images with 309×213 resolution were obtained and labeled by an expert. Due to the number of images, the division of the images in training and testing followed 5-fold cross-validation.

Figure 1. Sample images from Urban Tree (UT) dataset.

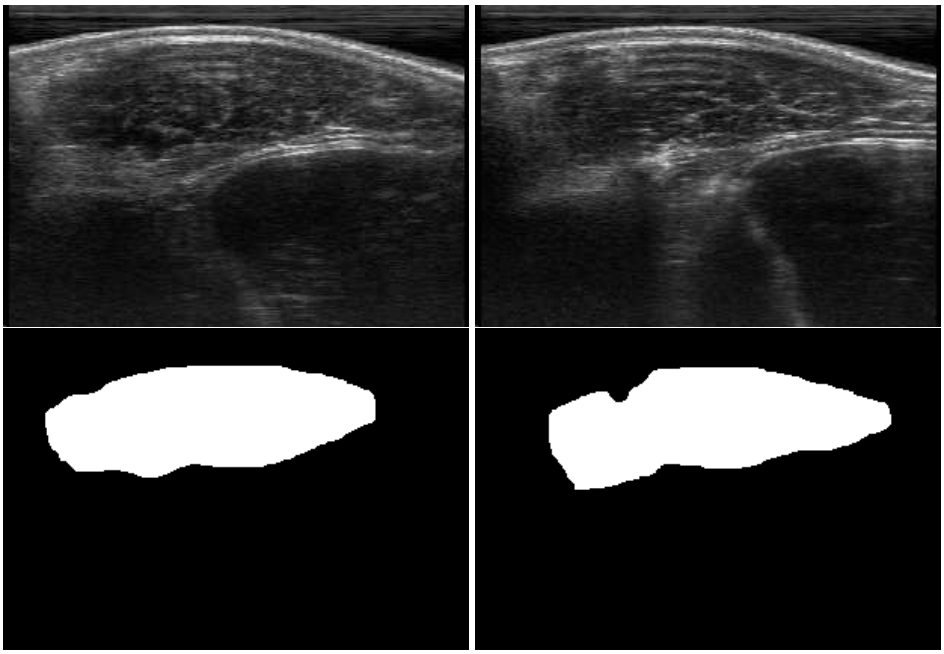


Figure 2. Sample images from Rib Eye Area (REA) dataset.

Soybean Disease (SD). The images from this dataset was obtained through PlantVillage[34], which contains several photographs taken by cell phone in soybean plantations. To compose the image dataset, 201 images with frog-eye disease were identified and manually annotated as shown in Fig. 3. It is important to emphasize that the images were taken in the field, and present several lighting challenges. The images were randomly divided into three sets: 121 for training, 40 for validation and 40 for testing.

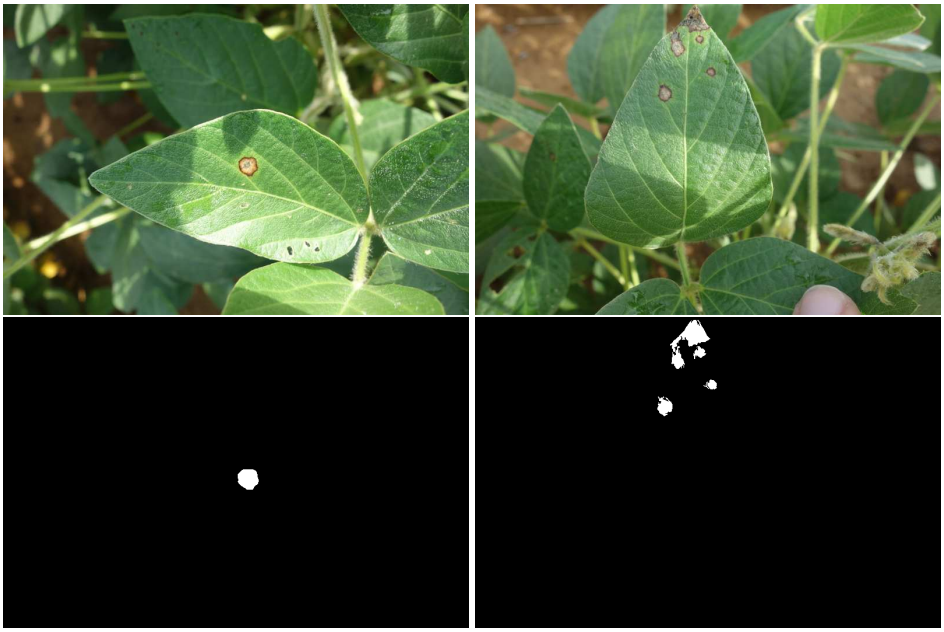


Figure 3. Sample images from Soybean Disease (SD) dataset.

Experimental setup

For REA and Urban Tree datasets, the images have been resized to 256×256 pixels. For Soybean Disease dataset, we resized the images to 1024×1024 pixels because the original ones have high resolution. For all segmentation methods, we use Stochastic Gradient Descent (SGD) optimizer with learning rate of 0.001, momentum of 0.9 and weight decay of 0.0005. The backbone weights of the segmentation methods started with pre-trained weights on ImageNet.

To evaluate the proposed approach and baselines, we use the following popular segmentation metrics: pixel accuracy (PA) and intersection over union (IoU). PA is the percentage of pixels correctly classified for each class. On the other hand, IoU is given by dividing the intersection area by the union area between prediction and ground-truth. Since the background is dominant in most images, we report the PA and IoU results only for the class of interest (e.g., REA, tree and diseases).

Results

In Tables 1 and 2, we compare the baseline methods and the proposed approach using SegNet and FCN, respectively. The main parameter of the proposed approach is σ that corresponds to the spread of uncertainty used in the loss function. Therefore, results for different values of σ were also reported.

For SegNet (Table 1), the proposed approach improved pixel accuracy (e.g., from 0.744 to 0.838 in Urban Tree dataset, 0.888 to 0.927 in REA dataset, and 0.35 to 0.777 in Soybean Disease dataset). The proposed approach also showed superior IoU results, especially in Urban Tree and SD datasets, where IoU improved from 0.676 to 0.705, and from 0.324 to 0.567, respectively. Further, it is found that using $\sigma = 2$ provided the best result in Urban Tree and SD datasets, while $\sigma = 3$ provided the best one in REA dataset. A lower value of σ for Urban Tree and SD datasets is expected due to the size of the foreground (tree and disease), which generally occupies a smaller area than the background compared to REA.

The proposed approach also provided better results using the FCN. From Table 2 it is observed that the results increase with the inclusion of the proposed approach. In Urban Tree, REA, and SD datasets, considerable increases of 8%, 0.5% and 23.9% were obtained in the pixel accuracy, respectively. On the other hand, IoU obtained by the proposed approach was slightly higher in Urban Tree and REA datasets and lower in SD dataset. Hence, the approach described here has proven to be effective for three datasets that include challenges of class imbalance and labelling uncertainty and for two semantic segmentation methods.

Table 1. Comparative results between the proposed approach using SegNet and baseline in the three image datasets.

Method	Urban Tree		REA		SD	
	PA	IoU	PA	IoU	PA	IoU
SegNet	0.744	0.676	0.888	0.841	0.350	0.324
SegNet + $\sigma = 1$	0.812	0.700	0.918	0.852	0.687	0.510
SegNet + $\sigma = 2$	**0.838**	**0.705**	0.918	0.852	**0.777**	**0.567**
SegNet + $\sigma = 3$	0.805	0.698	**0.927**	**0.853**	0.668	0.509

Table 2. Comparative results between the proposed approach using FCN and baseline in the three image datasets.

Method	Urban Tree		REA		SD	
	PA	IoU	PA	IoU	PA	IoU
FCN	0.820	0.730	0.966	0.861	0.750	**0.611**
FCN + $\sigma = 1$	0.892	0.754	0.967	**0.866**	**0.989**	0.368
FCN + $\sigma = 2$	**0.900**	**0.760**	0.967	0.863	**0.989**	0.371
FCN + $\sigma = 3$	0.896	0.729	**0.971**	0.865	0.982	0.425

Discussion and qualitative results

As shown in the previous section, FCN achieved better results than SegNet in the three image datasets. Therefore we discuss and present visual results of the FCN baseline and FCN using the proposed approach.

Urban tree dataset. Fig. 4 presents three examples that show the advantage of the proposed approach. The first column shows the ground-truth while the second and third columns present the result of the segmentation using the baseline and the proposed approach. The first example (first row) shows that the baseline incorrectly segments grass as a tree. On the other hand, the proposed approach can correctly segment the grass as a background, even though the colors are similar. The second example shows that the proposed approach is capable of correctly segmenting small foreground regions. This is because the importance of these pixels is increased during training and the weights of the convolutional layers tend to adjust better for these regions. Finally, the third example also shows small regions correctly segmented by the proposed approach. Also, it is possible to observe that the tree edge is better defined when compared to the baseline. This is possible due to the uncertainty included in tree-border regions, which are hardly labeled correctly. Concerning the border of objects, the proposed method decreases the importance of pixels, making CNN weights take this into account.

REA dataset. This image dataset has high uncertainty during labeling due to noise from the ultrasound image. In some cases, the border of REA is not completely visible and must be estimated by the specialist. Therefore, the proposed approach becomes essential to obtain accurate segmentation at the edges. The segmentation examples in Fig. 5 show that the baseline was not able to define the REA correctly due to the uncertainty of the labeling. On the other hand, the proposed approach presents results close to the specialist in regions that the border needs to be estimated.

Soybean disease dataset. As shown in Fig. 6, the proposed approach was able to segment soybean diseases with excellent pixel accuracy. It detects regions of disease that the baseline was not capable of, as illustrated in the second example. The proposed approach also segments the disease pixels more accurately compared to the baseline (see the third example). However, the proposed approach generally segments a region larger than the ground-truth, which explains the lower IoU compared to the baseline. In this task, it is important to have a low false-negative (as in the proposed approach) to detect diseases early and reduce losses.

Noise Invariance

Noise invariance of semantic segmentation methods was assessed on the Urban Tree dataset. Gaussian noise with $\sigma = 0.02$ was added to the images as illustrated in Fig. 7. We trained the proposed approach and the FCN baseline using the noisy images. Then, we evaluated them in the test set with and without noise. The results using noisy images in the training of both approaches are shown in Table 3. The second column of the table presents the results using noisy test images. As expected, both approaches still provided good results as they were trained and tested on noisy images. Our approach has achieved superior pixel accuracy and IoU compared to the baseline (e.g., 0.875 versus 0.776 and 0.697 versus 0.686).

Although these results are promising, it is not possible to guarantee that the methods discarded noise in training, since the test images were also noisy. To effectively assess the noise invariance, the third column of Table 3 shows the results using noisy images in training and noise-free images in the test. The baseline FCN presented weak results, showing that the noise had great interference in its training. On the other hand, the proposed approach showed consistent results, which demonstrates its robustness to noise. Our approach obtained pixel accuracy of 0.875 and 0.847 in test images with and without noise, a drop of

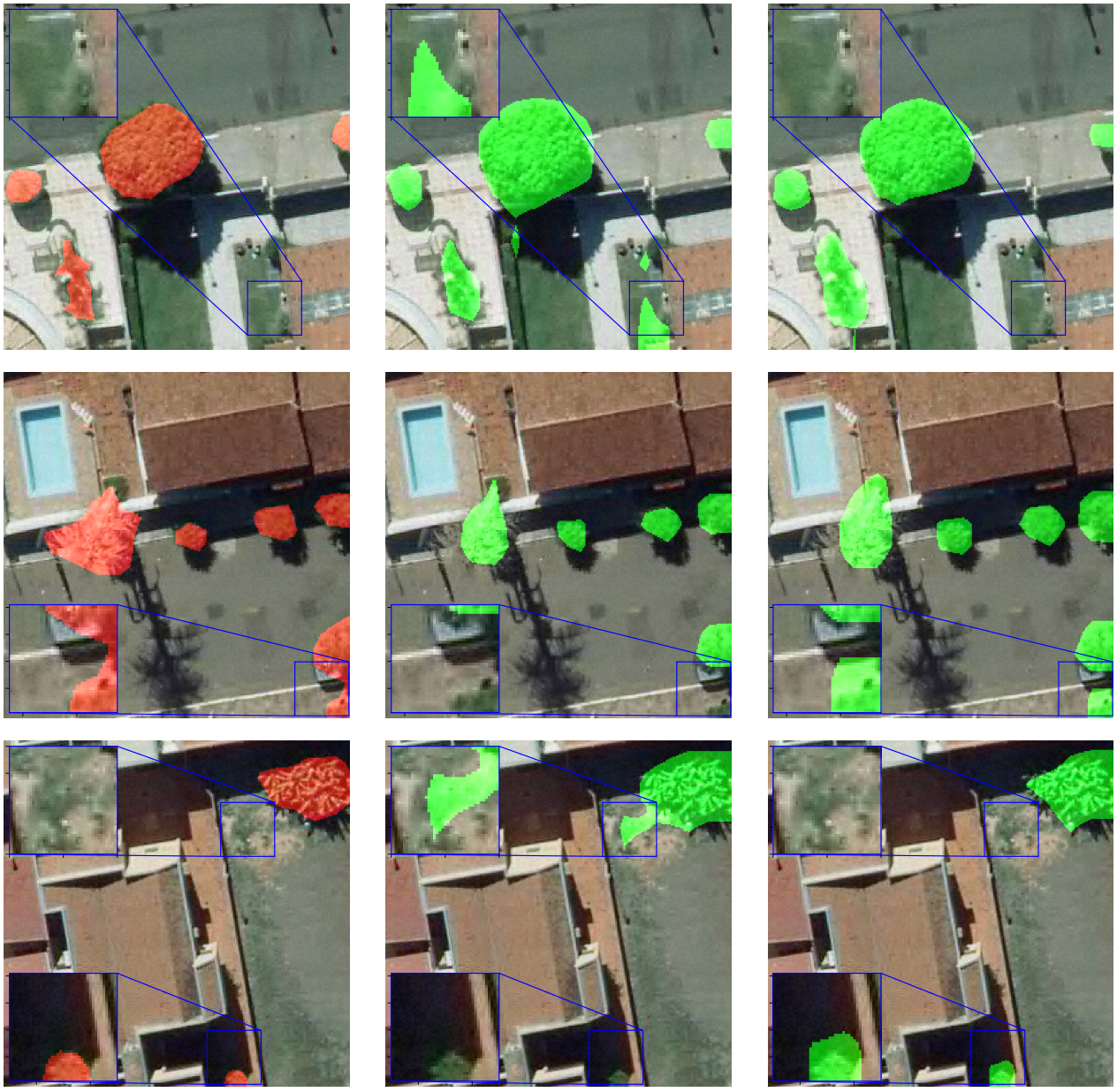


Figure 4. Example of ground-truth (in the left - a) , FCN (in the midle - b), and proposed approach (in the right - c) from Urban Tree dataset.

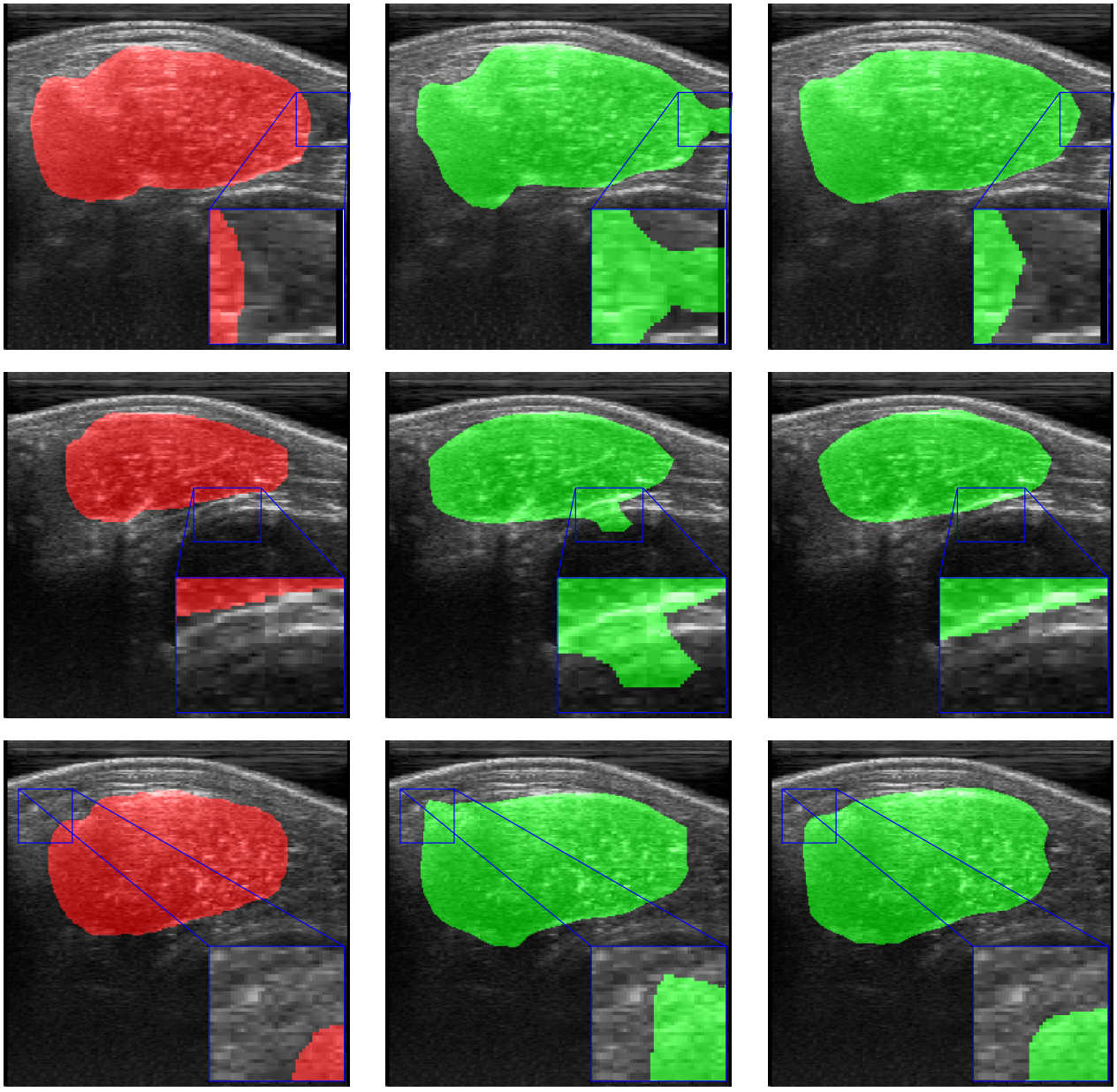


Figure 5. Example of ground-truth (in the left - a) , FCN (in the midle - b), and proposed approach (in the right - c) from REA dataset.

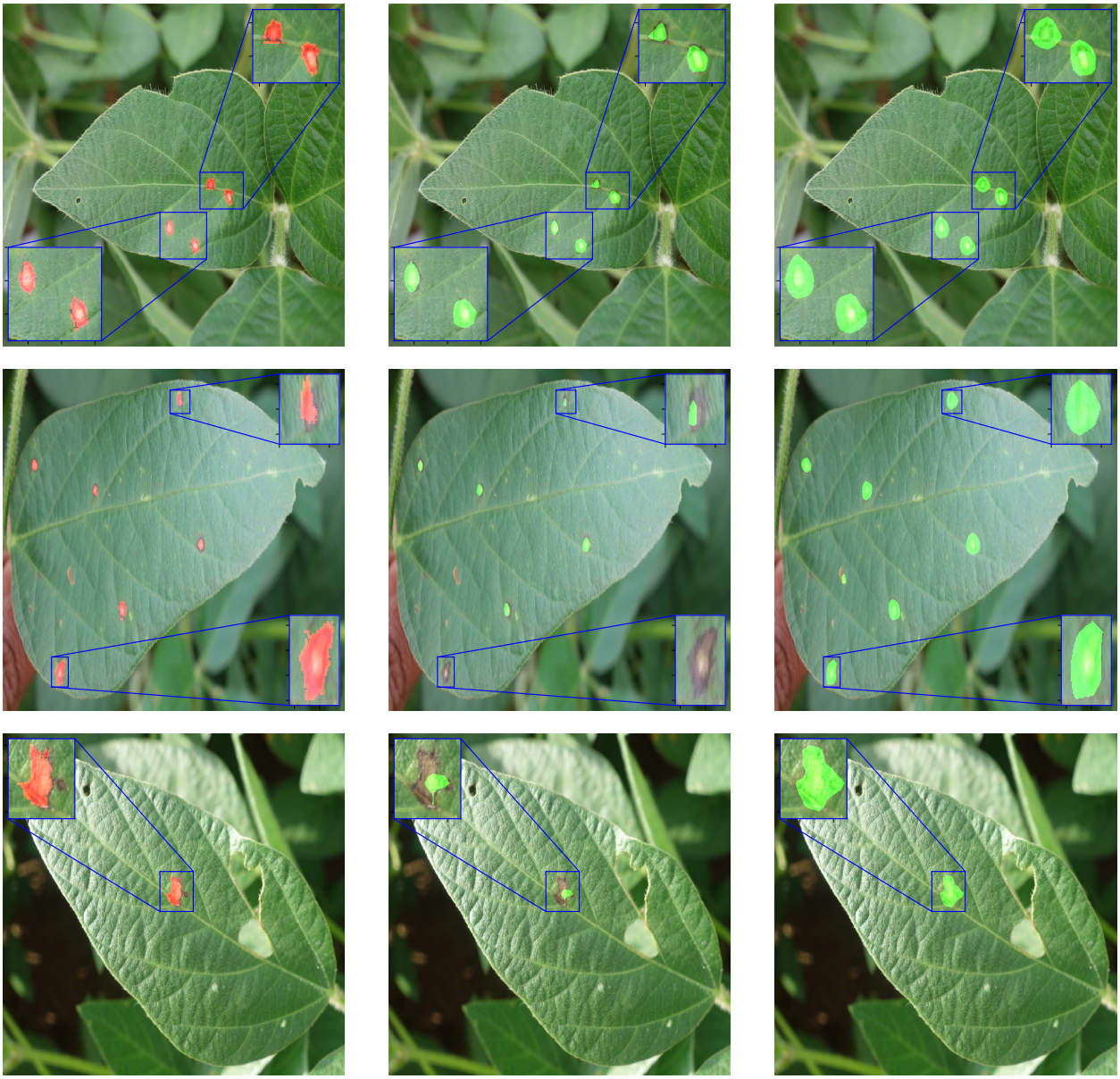
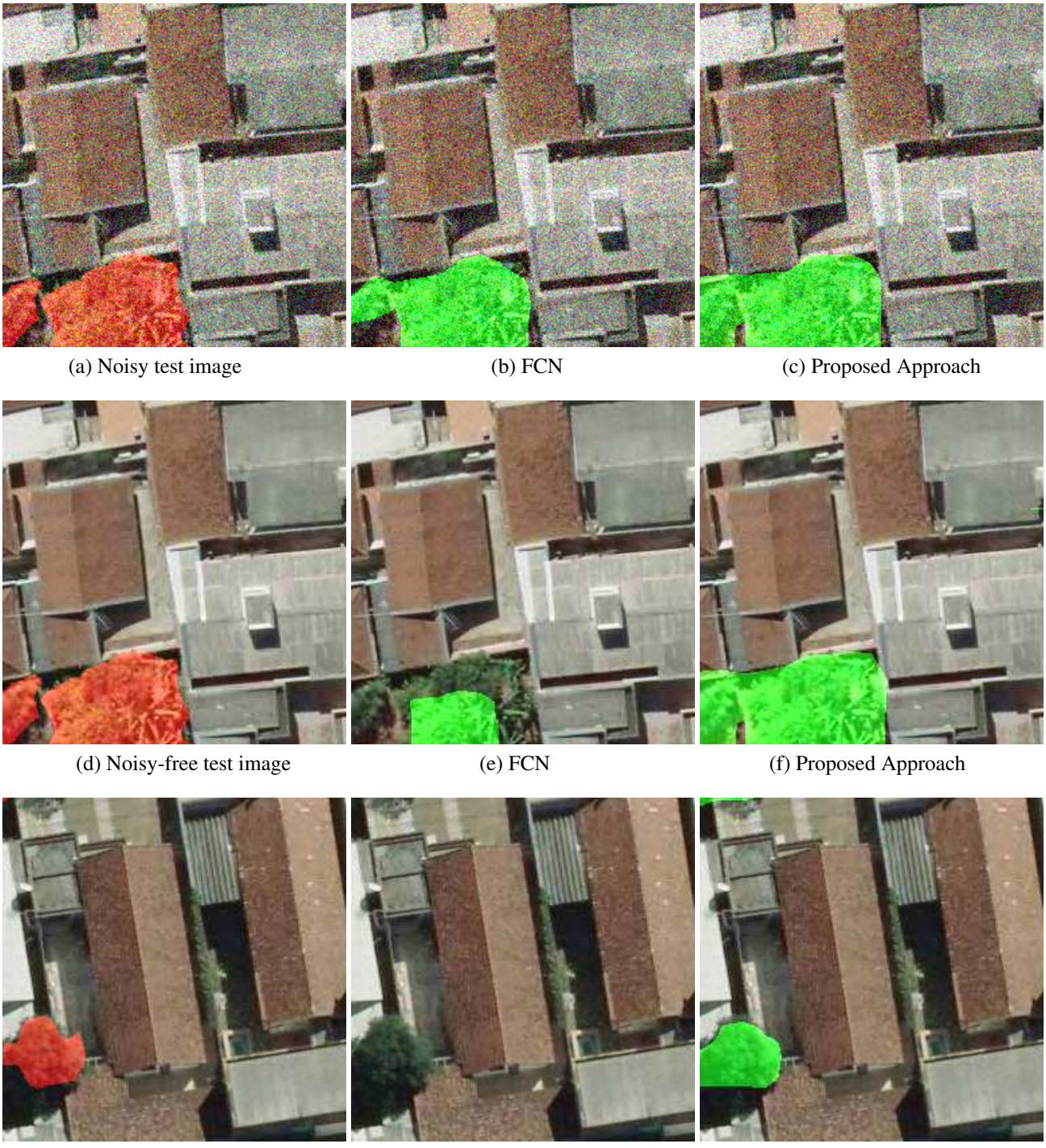


Figure 6. Example of ground-truth (in the left - a) , FCN (in the midle - b), and proposed approach (in the right - c) from Soybean Disease dataset.

Figure 7. Original images and their respective noisy images.



(a) Noisy test image (b) FCN (c) Proposed Approach

(d) Noisy-free test image (e) FCN (f) Proposed Approach

(g) Noisy-free test image (h) FCN (i) Proposed Approach

Figure 8. Comparative results of the proposed approach and FCN trained in noisy images. The first row of images shows the segmentation using a noisy test image while the second row of images shows the results using a noisy-free test image.

only 0.028.

Table 3. Comparative results between our method and the baseline FCN using noisy images to train.

Method	Noisy Images		Noise-free Images	
	PA	IoU	PA	IoU
FCN R-CNN	0.776	0.686	0.122	0.122
Ours	0.875	0.697	0.847	0.569

Fig. 8 shows visual segmentation results of both methods in test images with and without noise. The results of the baseline FCN and the proposed approach in a noisy test image (Fig. 8(a) are shown in Figs. 8(b) and 8(c), respectively. As the methods were trained on noisy images, they achieved satisfactory results despite the apparent noise. However, when a noisy-free image is used in testing methods trained with noisy images, the results of the proposed approach are superior to FCN as shown in Figs. 8(d)- 8(i).

Conclusion

A correctly weighting loss is important for semantic segmentation methods, mainly in datasets with imbalanced classes and labeling uncertainty. This paper shows how these challenges can be considered in a new loss function. The proposed approach combines two weights: i) the importance of the class given its occurrence and ii) the uncertainty in the labeling of pixels close to the edges. To the best of our knowledge, this is the first approach that overcomes both challenges using pixel-wise weights during training.

The robustness of the proposed approach can be ascertained for the three datasets considered; which presented different characteristics and challenges. The results showed that the proposed approach obtains superior metrics regardless of the segmentation method (e.g., SegNet and FCN). Significant results with an increase of up to 40% in accuracy were achieved by the proposed approach, which clearly shows its relevance. Our approach also proved to be more invariant to noise, even when training was performed on noisy images and tested on noise-free images.

As future work, we intend to evaluate new semantic segmentation methods. Further research also includes the application of the proposed approach to segmentation problems with several classes.

Methods

Improving semantic segmentation with labeling uncertainty and class imbalance

The purpose of semantic segmentation methods is to assign a label to each pixel x of an image $I(x)$, providing a pixel-level mask $\hat{M}(x)$. The most common methods for this task are based on CNNs composed of convolution, pooling, and upsampling layers[2,3]. This way, the pixel-level mask $\hat{M}$ is obtained through a CNN f_θ with layer parameters θ, $\hat{M} = f_\theta(I)$. The dominant loss function used to train a CNN takes the following form:

$$\min_{\theta \in \Theta} \sum_{(I,M) \in T} L(\hat{M}, M) + \lambda R(\theta) \tag{1}$$

where (I,M) is an example consisting of an image I and a ground-truth mask M of the training set T,$\hat{M} = f_\theta(I)$ is the predicted mask, L is a loss function (e.g., cross-entropy) that penalizes the wrong labels, and R is a regularizer.

In semantic segmentation tasks, the loss function L is usually decomposed into a sum of pixel losses according to Eq. 2. The weight of each pixel contributes uniformly during training.

$$L(\hat{M}, M) = \frac{1}{n} \sum_{x=1}^{n} L(\hat{M}(x), M(x)) \tag{2}$$

where n is the number of pixels.

There are two main issues within this approach: i) class imbalance; and ii) uncertainty in the annotation. The consequence of class imbalance is a bias towards the dominant classes over the classes that occupy smaller parts in the image. This occurs in most real-world image segmentation problems, where few classes dominate most images. Also, some classes do not have well-defined borders (e.g., trees), resulting in uncertainly labeled pixels. An incorrectly labeled pixel influences model learning, making filter convergence and learning even more difficult for small objects.

Figs. 1 and 2 present examples that illustrate the challenges of semantic segmentation methods. The trees in Fig. 1 show that in most images the foreground covers fewer pixels than the background (class imbalance). Besides, trees have edges that are difficult to label, and some pixels may be incorrectly labeled. Fig. 1 also illustrates the labeling challenge, in which some parts of the object are not visible in the image due to noise when capturing images.

Proposed loss function

To improve these issues, we propose to weight the contribution of each pixel based on its labeled class importance and uncertainty of its labeling. A weight for each pixel $w(x)$ is used in the loss function according to Eq. 3.

$$L(\hat{M},M) = \frac{1}{n}\sum_{x=1}^{n} \omega(x) \cdot L(\hat{M}(x),M(x)) \tag{3}$$

Unlike other approaches (e.g., focal loss[28]), the weight $\omega(x)$ of the pixel x is calculated by considering two important characteristics as shown in Eq. 4. The first part $\varphi^{c(x)}$ considers class unbalance, where $c(x)$ is the class labeled for pixel x. The second part $\delta(x)$ considers the labeling uncertainty of the pixel x. Both parts are described in detail in the sections below.

$$\omega(x) = \varphi^{c(x)} \cdot \delta(x) \tag{4}$$

Dealing with class imbalance

The first characteristic takes into account the unbalance of classes. To determine the weight of each class c, we use the training set according to Eq. 5. The lower the number of pixels in a given class, the higher the weight so that CNN layer filters fit evenly. When ω^c equals 1 for all classes, training is performed as traditionally. It is important to note that this weight is the same for all pixels in the same class.

$$\omega^c = \frac{m}{C * n^c} \tag{5}$$

where m is the number of pixels of all training images, C is the number of classes, and n^c is the number of pixels that belong to class c.

Dealing with labeling uncertainty

The second characteristic considers labeling uncertainty and is calculated for each pixel in the image. This is especially true for objects with poorly defined edges or low-resolution images. We consider that the closer to the edge, the greater the uncertainty of the class labeled for a given pixel. On the other hand, pixels near the center of objects are labeled more accurately. This feature can be modeled by Eq. 6 considering the distance of a pixel to the edges. The main parameter σ determines the spread of uncertainty around the edge.

$$\delta(x) = 1 - e^{-\frac{d(x)^2}{2\sigma^2}} \tag{6}$$

where $d(x)$ is the distance from the pixel x to the nearest edge pixel (can be calculated efficiently using the Euclidean Distance Transform) and σ is the standard deviation.

Figure 9 illustrates the process of calculating $\delta(x)$ for each pixel x. It is possible to observe that the closer to the object's edge, the lower the value of $\delta(x)$ and therefore it is considered as a pixel with high uncertainty. As a given pixel moves away from the edge, its uncertainty in the labeling is reduced.

Semantic Segmentation Methods

To evaluate the proposed approach, we used two well-known semantic segmentation methods: SegNet[2] and FCN[3]. SegNet[2] is a CNN with encoder and decoder networks, with a final pixel-wise classification layer. For each input, the encoder provides a low-resolution activation map representing the most important features. In this work, the encoder is composed of the convolutional and max-pooling layers of VGG16[35]. Then, the segmented image is reconstructed by the decoder. The decoder network is composed of convolutional and upsampling layers that use the corresponding max-pooling indices from the encoder to upsample the low-resolution feature map. In the last layer, a softmax classifier receives the feature map from the decoder for pixel-wise classification.

FCN[3] extended classification CNN (VGG16[35]) by transforming it into fully convolutional, where the fully connected layers were replaced by convolutional layers. In this way, the first part produces a feature map with low-resolution from the image, which is upsampled to produce pixel-wise predictions for segmentation.

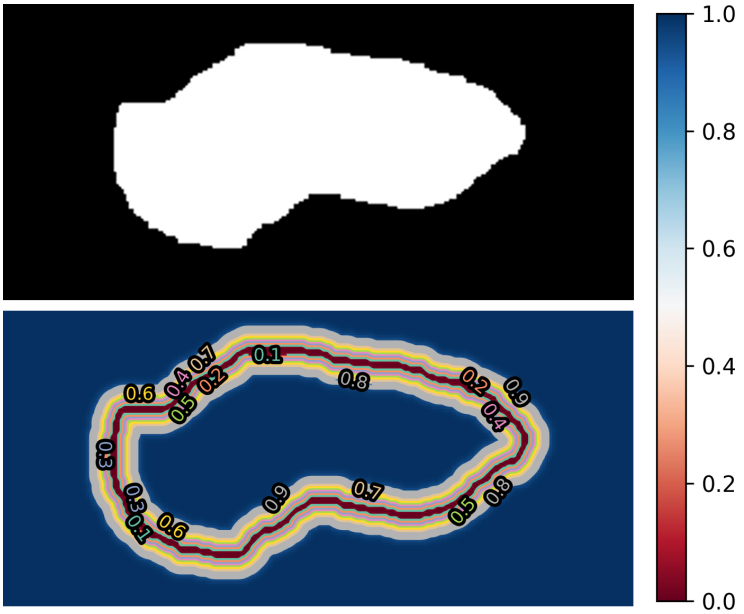


Figure 9. Example of calculating the uncertainty $\delta(x)$ of each pixel x. As a pixel approaches the edge, the greater its uncertainty.

References

1. Lobo Torres, D. *et al.* Applying fully convolutional architectures for semantic segmentation of a single tree species in urban environment on high resolution uav optical imagery. *Sensors* **20**, DOI: 10.3390/s20020563 (2020).
2. Badrinarayanan, V., Kendall, A. & Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis Mach. Intell.* **39**, 2481–2495, DOI: 10.1109/TPAMI.2016.2644615 (2017).
3. Long, J., Shelhamer, E. & Darrell, T. Fully convolutional networks for semantic segmentation. In *CVPR*, 3431–3440 (2015).
4. Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F. & Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *ECCV*, 833–851 (Springer International Publishing, 2018).
5. López, V., Fernández, A., García, S., Palade, V. & Herrera, F. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Inf. Sci.* **250**, 113 – 141, DOI: https://doi.org/10.1016/j.ins.2013.07.007 (2013).
6. Deng, J. *et al.* Imagenet: A large-scale hierarchical image database. In *CVPR*, 248–255, DOI: 10.1109/CVPR.2009.5206848 (2009).
7. Chrabaszcz, P., Loshchilov, I. & Hutter, F. A downsampled variant of imagenet as an alternative to the CIFAR datasets. *CoRR* **abs/1707.08819** (2017). 1707.08819.
8. Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2324, DOI: 10.1109/5.726791 (1998).
9. Liu, B. & Tsoumakas, G. Dealing with class imbalance in classifier chains via random undersampling. *Knowledge-Based Syst.* 105292, DOI: https://doi.org/10.1016/j.knosys.2019.105292 (2019).
10. Tsai, C.-F., Lin, W.-C., Hu, Y.-H. & Yao, G.-T. Under-sampling class imbalanced datasets by combining clustering analysis and instance selection. *Inf. Sci.* **477**, 47 – 54, DOI: https://doi.org/10.1016/j.ins.2018.10.029 (2019).
11. Sun, B., Chen, H., Wang, J. & Xie, H. Evolutionary under-sampling based bagging ensemble method for imbalanced data classification. *Front. Comput. Sci.* **12**, 331–350, DOI: 10.1007/s11704-016-5306-z (2018).
12. Ha, J. & Lee, J.-S. A new under-sampling method using genetic algorithm for imbalanced data classification. In *International Conference on Ubiquitous Information Management and Communication*, IMCOM '16, 95:1–95:6, DOI: 10.1145/2857546.2857643 (ACM, New York, NY, USA, 2016).

13. Fernández, A., García, S., Herrera, F. & Chawla, N. V. Smote for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary. *J. Artif. Int. Res.* **61**, 863–905 (2018).

14. Li, J. *et al.* Adaptive swarm balancing algorithms for rare-event prediction in imbalanced healthcare data. *PLOS ONE* **12**, 1–25, DOI: 10.1371/journal.pone.0180830 (2017).

15. Nekooeimehr, I. & Lai-Yuen, S. K. Adaptive semi-unsupervised weighted oversampling (a-suwo) for imbalanced datasets. *Expert. Syst. with Appl.* **46**, 405 – 416, DOI: https://doi.org/10.1016/j.eswa.2015.10.031 (2016).

16. Castellanos, F. J., Valero-Mas, J. J., Calvo-Zaragoza, J. & Rico-Juan, J. R. Oversampling imbalanced data in the string space. *Pattern Recognit. Lett.* **103**, 32 – 38, DOI: https://doi.org/10.1016/j.patrec.2018.01.003 (2018).

17. Dal Pozzolo, A., Caelen, O. & Bontempi, G. When is undersampling effective in unbalanced classification tasks? In Appice, A. *et al.* (eds.) *Machine Learning and Knowledge Discovery in Databases*, 200–215 (Cham, 2015).

18. Bulò, S. R., Neuhold, G. & Kontschieder, P. Loss max-pooling for semantic image segmentation. In *CVPR*, 7082–7091, DOI: 10.1109/CVPR.2017.749 (2017).

19. Bischke, B., Helber, P., Borth, D. & Dengel, A. Segmentation of imbalanced classes in satellite imagery using adaptive uncertainty weighted class loss. In *IGARSS*, 6191–6194, DOI: 10.1109/IGARSS.2018.8517836 (2018).

20. Chan, R., Rottmann, M., Hüger, F., Schlicht, P. & Gottschalk, H. Application of decision rules for handling class imbalance in semantic segmentation. *CoRR* **abs/1901.08394** (2019). 1901.08394.

21. Xu, J., Schwing, A. G. & Urtasun, R. Learning to segment under various forms of weak supervision. In *CVPR*, 3781–3790, DOI: 10.1109/CVPR.2015.7299002 (2015).

22. Caesar, H., Uijlings, J. & Ferrari, V. Joint calibration for semantic segmentation. In *BMVC*, 29.1–29.13, DOI: 10.5244/C.29.29 (BMVA Press, 2015).

23. Bansal, A., Chen, X., Russell, B. C., Gupta, A. & Ramanan, D. Pixelnet: Towards a general pixel-level architecture. *CoRR* **abs/1609.06694** (2016). 1609.06694.

24. Wu, Z., Shen, C. & van den Hengel, A. High-performance semantic segmentation using very deep fully convolutional networks. *CoRR* **abs/1604.04339** (2016). 1604.04339.

25. Dong, Q., Gong, S. & Zhu, X. Imbalanced deep learning by minority class incremental rectification. *IEEE Transactions on Pattern Analysis Mach. Intell.* **41**, 1367–1381, DOI: 10.1109/TPAMI.2018.2832629 (2019).

26. Huang, C., Li, Y., Loy, C. C. & Tang, X. Learning deep representation for imbalanced classification. In *CVPR*, 5375–5384, DOI: 10.1109/CVPR.2016.580 (2016).

27. Ren, M., Zeng, W., Yang, B. & Urtasun, R. Learning to reweight examples for robust deep learning. In *ICML*, 4331–4340 (2018).

28. Lin, T., Goyal, P., Girshick, R., He, K. & Dollár, P. Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis Mach. Intell.* **42**, 318–327 (2020).

29. Johnson, J. M. & Khoshgoftaar, T. M. Survey on deep learning with class imbalance. *J. Big Data* **6**, 27, DOI: 10.1186/s40537-019-0192-5 (2019).

30. Ding, H., Jiang, X., Liu, A. Q., Thalmann, N. M. & Wang, G. Boundary-aware feature propagation for scene segmentation. In *ICCV*, 6819–6829 (2019).

31. Shen, W., Wang, X., Wang, Y., Bai, X. & Zhang, Z. Deepcontour: A deep convolutional feature learned by positive-sharing loss for contour detection. In *CVPR*, 3982–3991, DOI: 10.1109/CVPR.2015.7299024 (2015).

32. Islam, M. A., Naha, S., Rochan, M., Bruce, N. & Wang, Y. Label refinement network for coarse-to-fine semantic segmentation (2017). 1703.00551.

33. Hamaguchi, R., Fujita, A., Nemoto, K., Imaizumi, T. & Hikosaka, S. Effective use of dilated convolutions for segmenting small object instances in remote sensing imagery. In *WACV*, 1442–1450, DOI: 10.1109/WACV.2018.00162 (2018).

34. Hughes, D. P. & Salathé, M. An open access repository of images on plant health to enable the development of mobile disease diagnostics through machine learning and crowdsourcing. *CoRR* **abs/1511.08060** (2015). 1511.08060.

35. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. *CoRR* **abs/1409.1556** (2014).

Author contributions statement

J.M.J., W.N.G., J.A.S. and L.P.O conceived the experiment, P.O.B and W.N.G. conducted the experiment(s) and produced the figures and tables, J.A.S., Z. L., and J.L. evaluated the results, J.A.C.M., M.J.M., L.P.O., A.P.M.R., and W.N.G. wrote the main manuscript text, R.C.G., D. M. F., and D.N.G helped validating the discussion and conclusion. All authors reviewed the manuscript.

Acknowledgements

The authors acknowledge the support of the UFMS (Federal University of Mato Grosso do Sul) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (Finance Code 001).

Competing interests

This research was partially funded by CNPq (p: 433783/2018-4, 303559/2019-5, 304052/2019-1, and 310517/2020-6) and CAPES Print (p: 88881.311850/2018-01). The authors acknowledge the support of the UFMS (Federal Universityof Mato Grosso do Sul) and CAPES (Finance Code 001), and NVIDIA© for the donation of the Titan X graphics card used in the experiments.

Conflicts of interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Figures

Figure 1

Sample images from Urban Tree (UT) dataset.

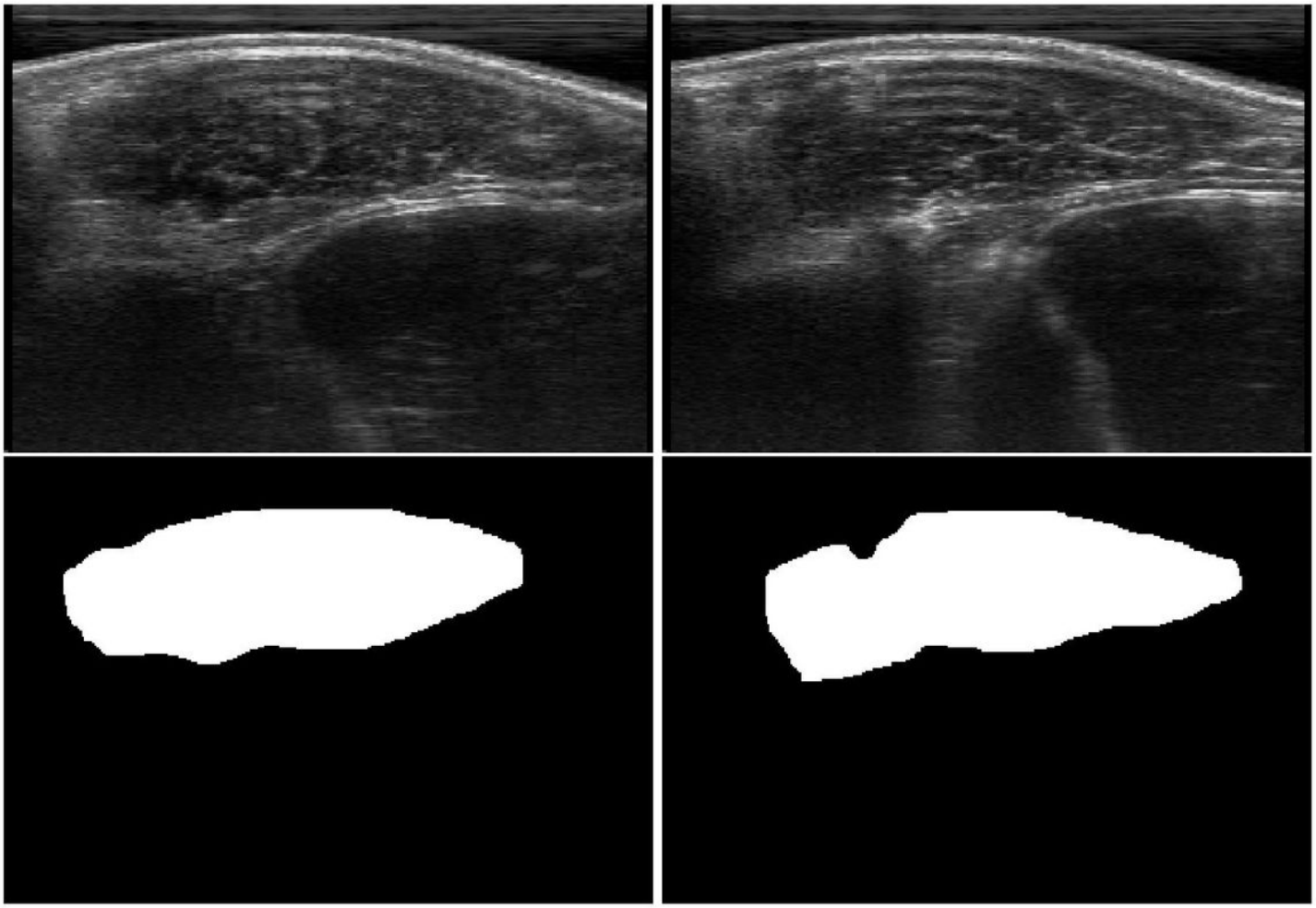


Figure 2

Sample images from Rib Eye Area (REA) dataset

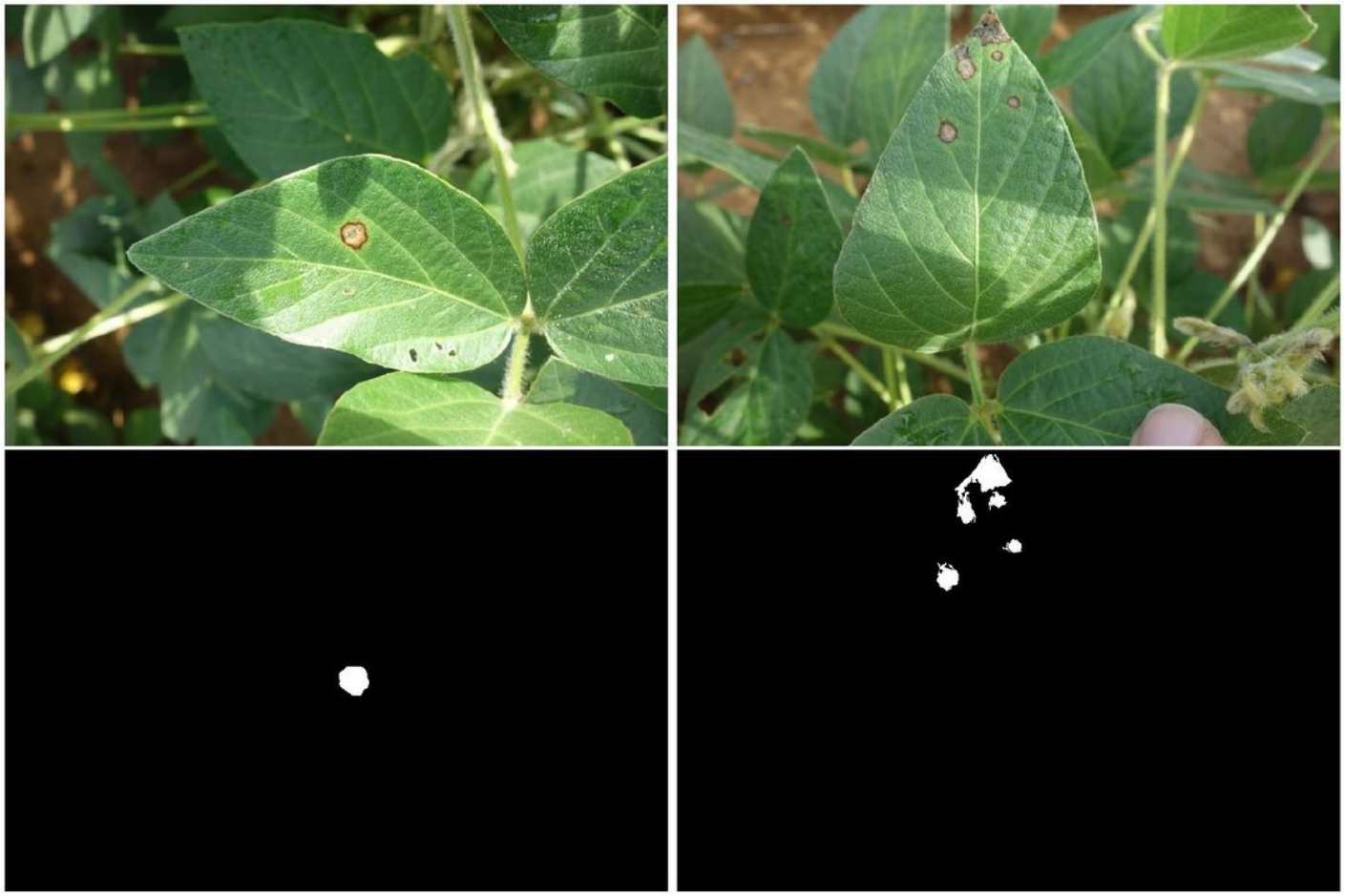


Figure 3

Sample images from Soybean Disease (SD) dataset.

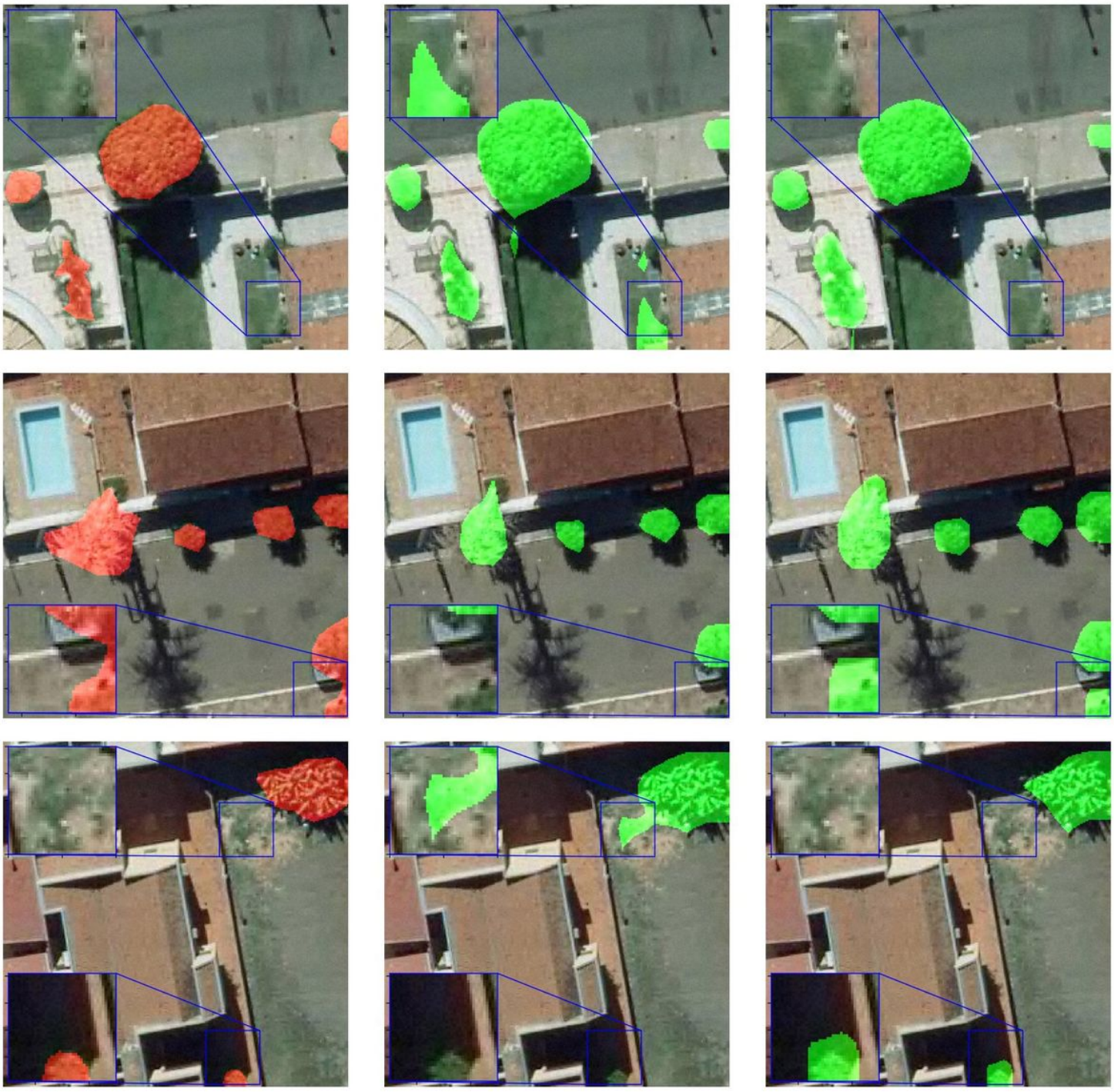


Figure 4

Example of ground-truth (in the left - a) , FCN (in the midle - b), and proposed approach (in the right - c) from Urban Tree dataset.

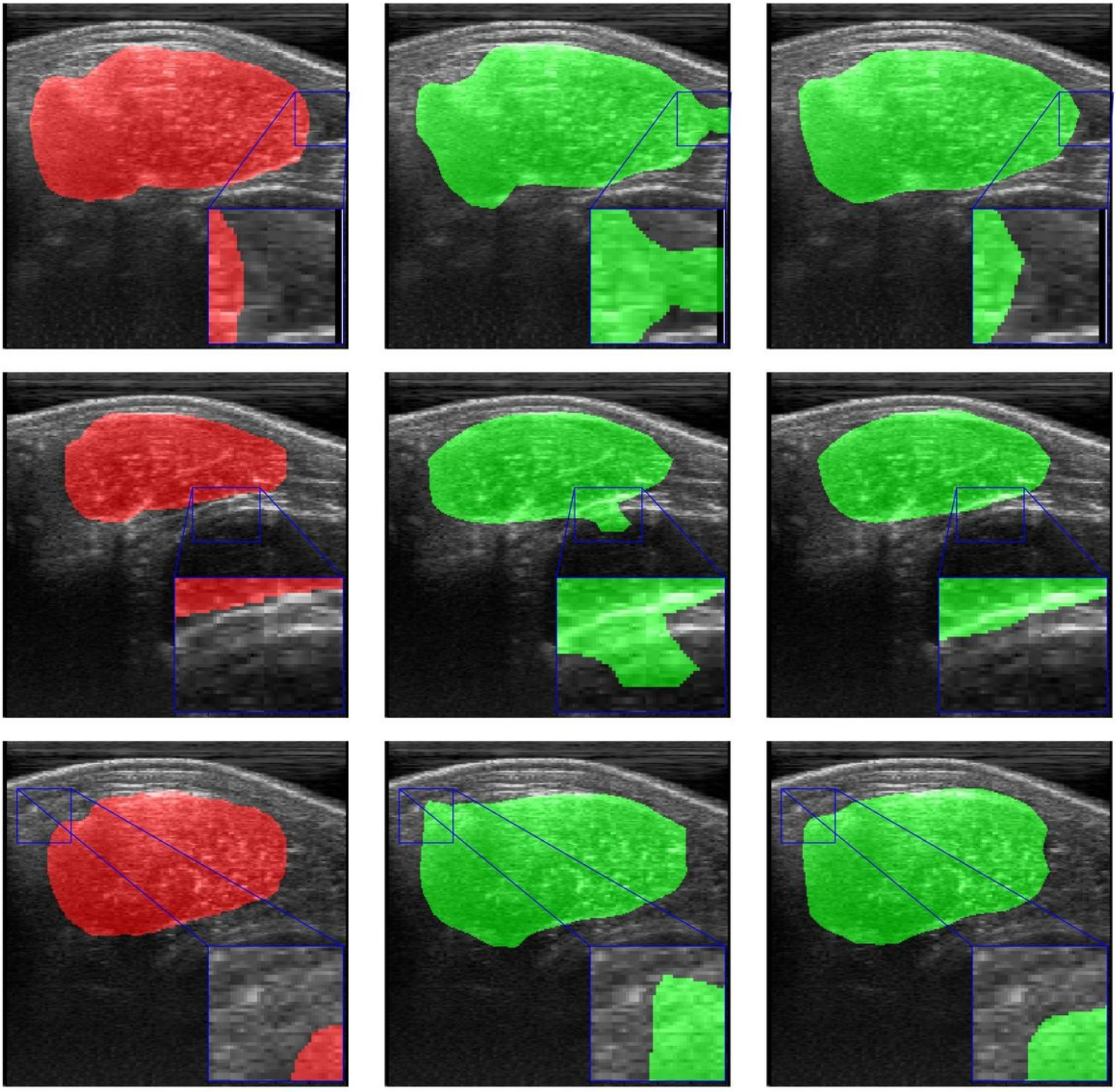


Figure 5

Example of ground-truth (in the left - a) , FCN (in the midle - b), and proposed approach (in the right - c) from REA dataset.

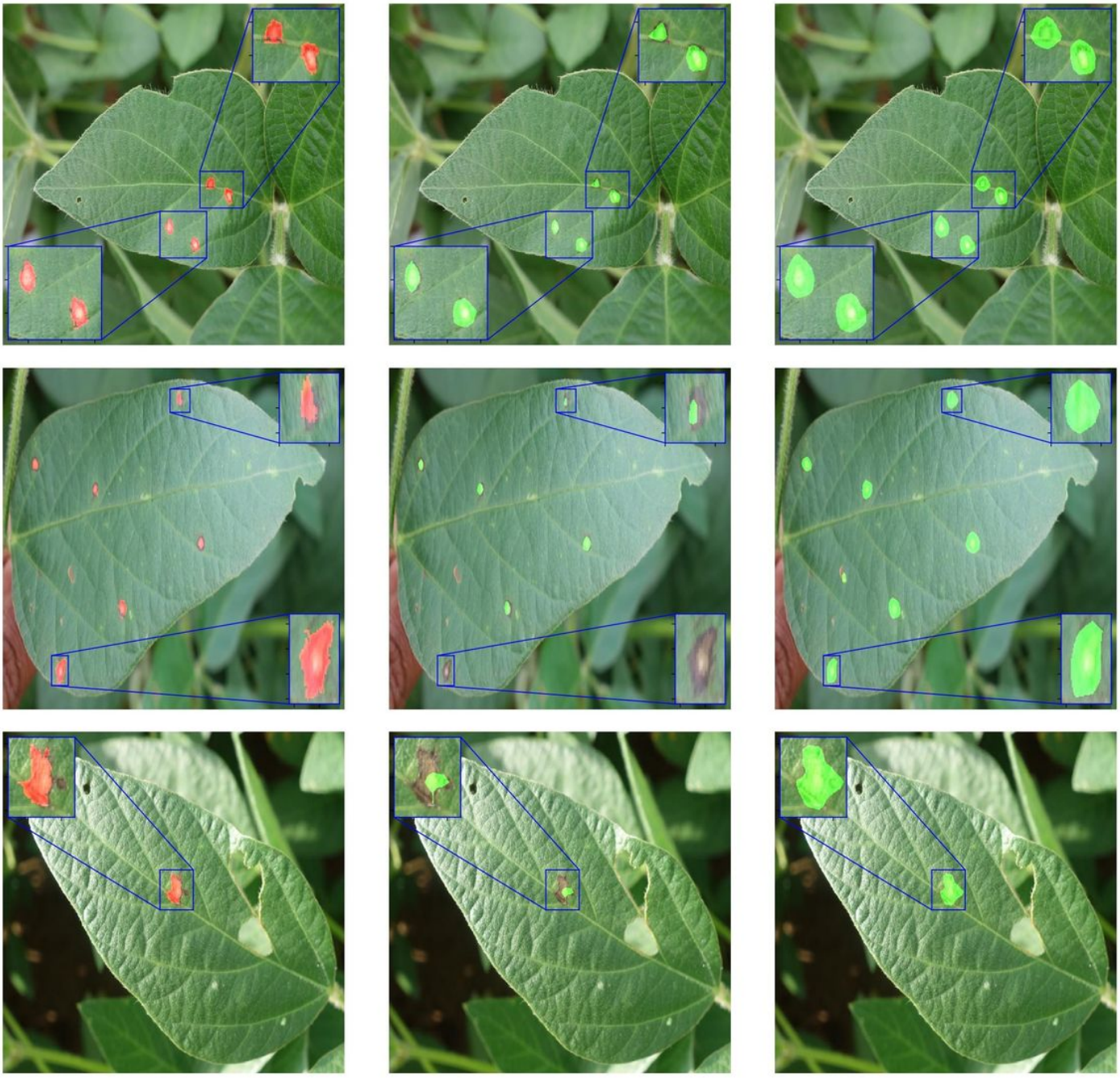


Figure 6

Example of ground-truth (in the left - a) , FCN (in the midle - b), and proposed approach (in the right - c) from Soybean Disease dataset.

Figure 7

Original images and their respective noisy images.

Figure 8

Comparative results of the proposed approach and FCN trained in noisy images. The first row of images shows the segmentation using a noisy test image while the second row of images shows the results using a noisy-free test image.

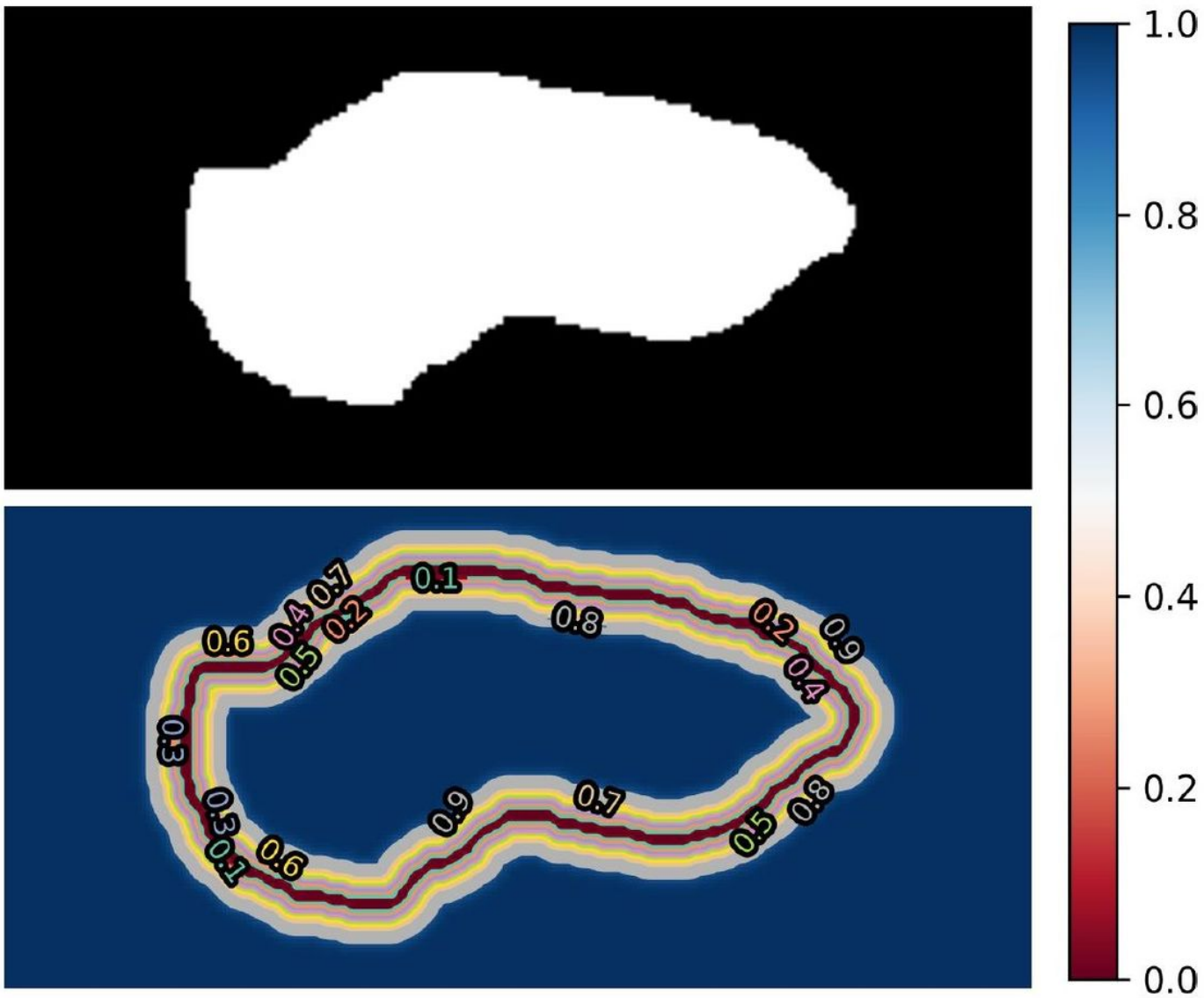


Figure 9

Example of calculating the uncertainty δ(x) of each pixel x. As a pixel approaches the edge, the greater its uncertainty.