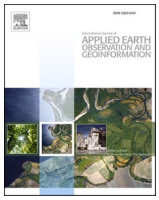
International Journal of Applied Earth Observations and Geoinformation 108 (2022) 102690

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Semantic segmentation with labeling uncertainty and class imbalance applied to vegetation mapping

Patrik Olã Bressan [a,b], José Marcato Junior [c], José Augusto Correa Martins [c], Maximilian Jaderson de Melo [a], Diogo Nunes Gonçalves [a], Daniel Matte Freitas [a], Ana Paula Marques Ramos [d,e], Michelle Taís Garcia Furuya [e], Lucas Prado Osco [f], Jonathan de Andrade Silva [a], Zhipeng Luo [g], Raymundo Cordero Garcia [c], Lingfei Ma [h,*], Jonathan Li [i], Wesley Nunes Gonçalves [a,c]

[a] Faculty of Computer Science, Federal University of Mato Grosso do Sul, Av. Costa e Silva, Campo Grande 79070-900, MS, Brazil
[b] Federal Institute of Mato Grosso do Sul, Jardim 79240-000, MS, Brazil
[c] Faculty of Engineering, Architecture, and Urbanism and Geography, Federal University of Mato Grosso do Sul, Av. Costa e Silva, Campo Grande 79070-900, MS, Brazil
[d] Agronomy Program, University of Western São Paulo, Rod. Raposo Tavares, km 572 - Limoeiro, Pres. Prudente 19067-175, SP, Brazil
[e] Environment and Regional Development Program, University of Western São Paulo, Rod. Raposo Tavares, km 572 - Limoeiro, Pres. Prudente 19067-175, SP, Brazil
[f] Faculty of Engineering and Architecture and Urbanism, University of Western São Paulo, Rod. Raposo Tavares, km 572 - Limoeiro, Pres. Prudente 19067-175, SP, Brazil
[g] School of Informatics, Xiamen University, Xiamen, FJ 361005, China
[h] Engineering Research Center of State Financial Security, Ministry of Education, Central University of Finance and Economics, Beijing 102206, China
[i] Department of Geography and Environmental Management, University of Waterloo, Waterloo ON N2L 3G1, Canada

ARTICLE INFO

Keywords:
Semantic segmentation
Labeling uncertainty
Class weighting
Loss function

ABSTRACT

Recently, Convolutional Neural Networks (CNN) methods achieved impressive success in semantic segmentation tasks. However, challenges like class imbalance around samples and the uncertainty in human pixel-labeling are not completely addressed. Here we present an approach that calculates a weight for each pixel considering its class and uncertainty during the labeling process. The pixel-wise weights are used at the training phase to increase or decrease the importance of the pixels accordingly. Experimental results were conducted adapting well-known CNN methods FCN and SegNet; however, this strategy can be applied to any segmentation method. We evaluated the experiments for semantic segmentation of urban trees in aerial imageries. The robustness of the approach was assessed using a dataset with terrestrial images from vegetation with a drastic imbalance condition. We achieved significant improvements in the tasks compared to the baseline methods. We also verified that the proposed strategy proved to be more invariant to noise. The approach presented in this paper could be used within a wide range of semantic segmentation methods to improve their robustness.

1. Introduction

Semantic segmentation is an image processing task that aims to establish a known class for each pixel. This task is crucial to infer knowledge of a scene in computer vision systems, as shown in recent studies of tree species segmentation (Lobo Torres et al., 2020b). In this field, significant advances have been achieved through Convolutional Neural Networks (CNNs) based methods, including ones such as SegNet (Badrinarayanan et al., 2017; Dowden et al., 2021), Fully Convolutional Network (FCN) (Long et al., 2015), and DeepLabv3+ (Chen et al., 2018). Even with the development of novel methods, the segmentation accuracy in many remote sensing applications is far from the expectation (Tian et al., 2021). In this context, strategies that can improve and can be integrated into any semantic segmentation method become of great

* Corresponding author.
E-mail addresses: patrik.bressan@ifms.edu.br (P.O. Bressan), jose.marcato@ufms.br (J.M. Junior), jose.a@ufms.br (J.A. Correa Martins), maximilian.melo@ufms.br (M.J. de Melo), daniel.freitas@ufms.br (D.M. Freitas), anaramos@unoeste.br (A.P. Marques Ramos), lucasosco@unoeste.br (L.P. Osco), jonathan.andrade@ufms.br (J. de Andrade Silva), zpluo@stu.xmu.edu.cn (Z. Luo), raymundo.garcia@ufms.br (R.C. Garcia), l53ma@cufe.edu.cn (L. Ma), junli@uwaterloo.ca (J. Li), wesley.goncalves@ufms.br (W.N. Gonçalves).

https://doi.org/10.1016/j.jag.2022.102690
Received 22 November 2021; Received in revised form 10 January 2022; Accepted 13 January 2022
Available online 17 February 2022

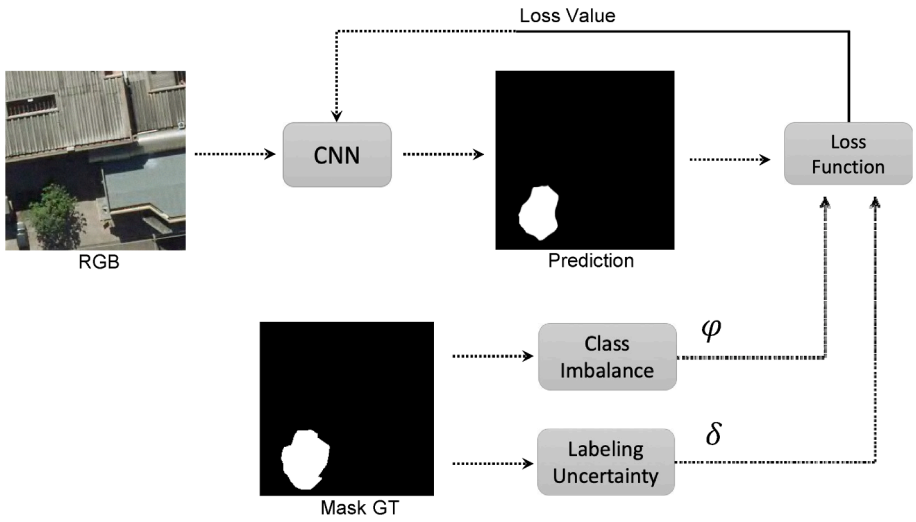


Fig. 1. The segmentation method receives the RGB image and provides the prediction. The GT mask is used to calculate the unbalance of the classes and the uncertainty in the annotation. All this information is combined into the new loss function, which calculates the loss value to guide learning the segmentation method.

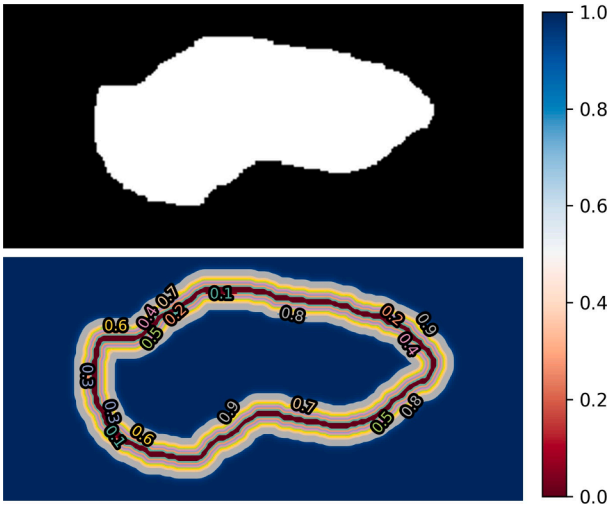


Fig. 2. Example of calculating the uncertainty $\delta(x)$ of each pixel x. As a pixel approaches the edge, the greater its uncertainty. The top figure represents the labeled mask of the object, while the bottom image corresponds with the uncertainty calculated.

interest.

The combination of two factors has 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 process (López et al., 2013). One way to minimize imbalance is by uniformly sampling data and collecting images (such as well-known datasets, ImageNet (Deng et al., 2009; Chrabaszcz et al., 2017), MNIST (Modified National Institute of Standards and Technology) (Lecun et al., 1998) and CIFAR 10/100), under-sampling the majority classes (Liu and Tsoumakas, 2019; Tsai et al., 2019; Sun et al., 2018; Ha and Lee, 2016), or over-sampling the minority classes (Fernández et al., 2018; Li et al., 2017; Nekooeimehr and Lai-Yuen, 2016; Castellanos et al., 2018). However, these approaches change the distribution of data and can affect learning and inference in a significant manner (Dal Pozzolo et al., 2015).

The second factor, much less explored in the literature, is related to the uncertainty in the image labeling (Bulò et al., 2017; Bischke et al., 2018). In low resolution or noisy images, the edges of objects become inaccurate, and even expert labeling may include annotation errors that affect the training of a network. Even in high-resolution images, some objects (e.g., trees (Lobo Torres et al., 2020b)) have complex edges that make them difficult to annotate.

In this study, we propose an approach to deal with class unbalance and uncertainty in the labeling process for image segmentation tasks to overcome the aforementioned issues. Specifically, we introduce a loss function where the contribution of each pixel is weighted. First, pixels belonging to minority classes have their importance increased. Second, since pixels near the edges of the object generally have greater uncertainty on labeling, their importance is diminished during training. These two pixel-wise weights are then combined and produce a satisfactory impact during training and inference of the segmentation methods.

Experiments were mainly conducted to segment urban trees in high-resolution aerial imageries. Urban trees benefit to the population, and their monitoring is relevant in multiple urban planning tasks. The adopted strategy significantly reduced the confusion between trees and undergrowth vegetation, improving the mapping of trees in urban environments. This is the first approach that overcomes both challenges using these pixel-wise weights during training to the best of our knowledge.

In summary, our original contributions are described as follows:

1. Development of a novel loss function to deal with both class unbalance and uncertainty issue in the labeling process for remote sensing image segmentation task;
2. Assessment in two very distinct datasets to show the strengthening of the proposed approach;
3. Significant reduction in the confusion between vegetation and background classes. We also verified that the proposed strategy proved to be more invariant to noise considering both datasets.

2. Related works

2.1. Imbalance Data

In semantic segmentation, approaches have already been proposed to deal with class imbalance. Traditional approaches can use resampling (e.g., oversampling and undersampling) and rebalancing schemes via statistic analysis, such as inverse or median frequency (Chan et al., 2019; Xu et al., 2015; Caesar et al., 2015). Despite correcting the imbalance, these approaches include several disadvantages on both oversampling

Fig. 3. Sample images from Urban Tree (UT) dataset. The top images correspond with the RGB input dataset while the bottom images correspond with the labeled example.

Fig. 4. Sample images from Soybean Disease (SD) dataset. The top images correspond with the RGB input dataset while the bottom images correspond with the labeled example.

and undersampling methods. Oversampling methods increase computational cost and may be more prone to overfitting due to the inclusion of duplicated data. On the other hand, undersampling methods can discard important data for learning, reducing accuracy in the prediction.

Approaches are also based on constraints during training, such as restricting the number of pixels contributing to the loss function during backpropagation at random (Bansal et al., 2016), based on the k highest loss of the pixels (Wu et al., 2016) or hard samples (Dong et al., 2019).

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

Method	Urban Tree		SD	
	PA (%)	IoU (%)	PA (%)	IoU (%)
SegNet	74.4	67.6	35.0	32.4
SegNet $+ \sigma = 1$	81.2	70.0	68.7	51.0
SegNet $+ \sigma = 2$	**83.8**	**70.5**	**77.7**	**56.7**
SegNet $+ \sigma = 3$	80.5	69.8	66.8	50.9

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

Method	Urban Tree		SD	
	PA (%)	IoU (%)	PA (%)	IoU (%)
FCN	82.0	73.0	75.0	61.1
FCN $+ \sigma = 1$	89.2	75.4	98.9	36.8
FCN $+ \sigma = 2$	**90.0**	76.0	**98.9**	37.1
FCN $+ \sigma = 3$	89.6	72.9	98.2	42.5

Huang et al. (2016) 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. (2018) 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 (Lin et al., 2020) was proposed 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 Johnson and Khoshgoftaar (2019).

2.2. Labeling Uncertainty

Labeling uncertainty is related to image resolution and object-edge complexity. As of recently, Bischke et al. (2018) 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 research. Bulò et al. (2017) 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. (2019) proposed learning boundary objects as an additional class to increase the feature similarity of the same object. Similarly, Shen et al. (2015) 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. (2017) proposed a new CNN architecture to predict segmentation labels at several resolutions. At each stage (scale), a loss function 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. (2018) proposed a novel architecture called local feature extraction, which aggregates local

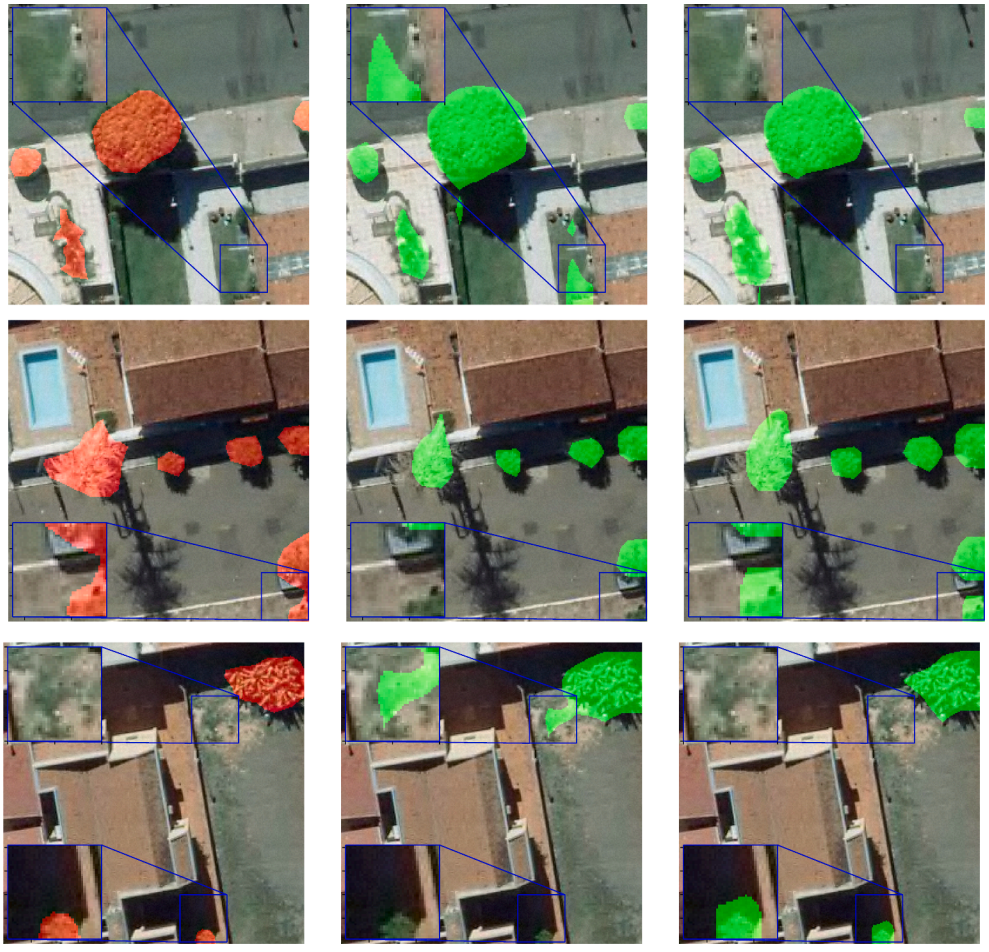


Fig. 5. Example of ground-truth (in the left - a), FCN (in the middle - b), and proposed approach (in the right - c) from Urban Tree dataset.

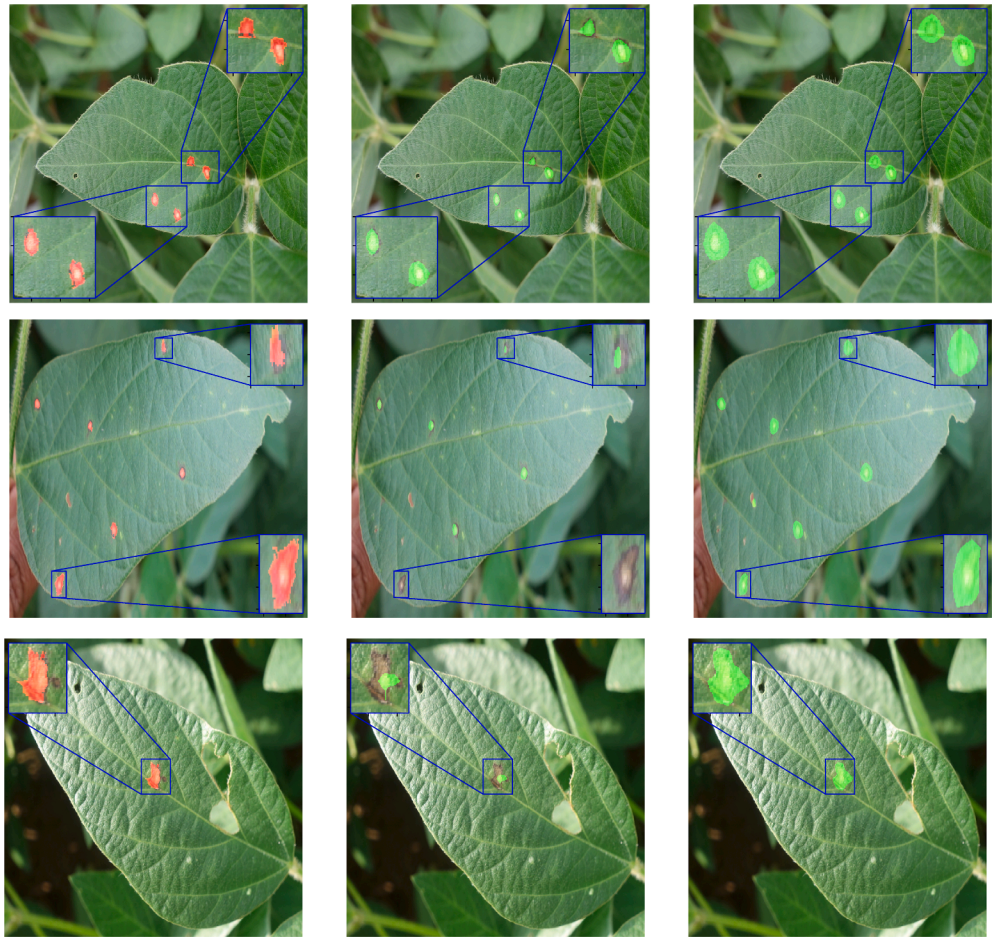


Fig. 6. Example of ground-truth (in the left - a), FCN (in the middle - b), and proposed approach (in the right - c) from Soybean Disease dataset.

features with decreasing dilation factor to segment small objects in remote sensing imagery.

2.3. Semantic segmentation applied to vegetation mapping

The mapping and monitoring of vegetation are crucial for applications in urban and rural environments. Semantic segmentation methods based on CNN have been employed for this task, providing total vegetation coverage throughout the study area.

Osco et al. (2021) investigated the use of FCN, U-Net, SegNet, DDCN, and DeepLabV3 + for the segmentation of citrus trees. The authors verified that all the methods performed equally for this task. Lobo Torres et al. (2020a) assessed SegNet, DeepLabv3+, U-Net, and FC-DenseNet for the segmentation of tree species. Minor differences occurred between the methods.

In the context of urban tree segmentation, Martins et al. (2021) also assessed most of the previously mentioned methods and also verified minor differences among them. The authors verified that most errors occurred in the edges of the canopies, and also, there were confusions with grassland.

In general, we verified that minor differences occur between semantic segmentation deep learning-based methods for segmenting the vegetation. However, it is still necessary to develop tools to maximize the segmentation accuracy (Tian et al., 2021). Here, we addressed this, proposing an approach that deals with class unbalance and uncertainty in the labeling process.

3. Methods

3.1. Proposed Approach

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 $\widehat{M}(x)$. The most common methods for this task are based on CNNs composed of convolution, pooling, and upsampling layers (Long et al., 2015; Badrinarayanan et al., 2017). Accordingly, the pixel-level mask $\widehat{M}$ is obtained through a CNN f_θ with layer parameters $\theta, \widehat{M} = f_\theta(I)$. The dominant loss function used to train a CNN takes the following equation:

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

where (I, M) is an example consisting of an image I and a ground-truth mask M of the training set $T, \widehat{M} = f_\theta(I)$ 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.

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

where n is the number of pixels.

The consequence of class imbalance is a bias towards the dominant classes over those that occupy smaller parts of the image. This occurs in

Fig. 7. Original images and their respective noisy images.

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	77.6	68.6	12.2	12.2
Ours	87.5	69.7	84.7	56.9

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 the models' learning task, making filter convergence and learning even more difficult for small objects.

3.2. 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 as shown in Fig. 1. A weight for each pixel $w(x)$ is used in the loss function according to Eq. 3.

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

Unlike other approaches (e.g., focal loss (Lin et al., 2020)), 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 imbalance, 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.

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

3.3. Dealing with Class Imbalance

The first characteristic takes the unbalance of classes into account. 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 $\varphi(c)$ 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 c.

$$\varphi\left(c\right) = \frac{m}{C * n^{c}} \tag{5}$$

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.

3.4. 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 of the object, the greater the uncertainty of the class label 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.

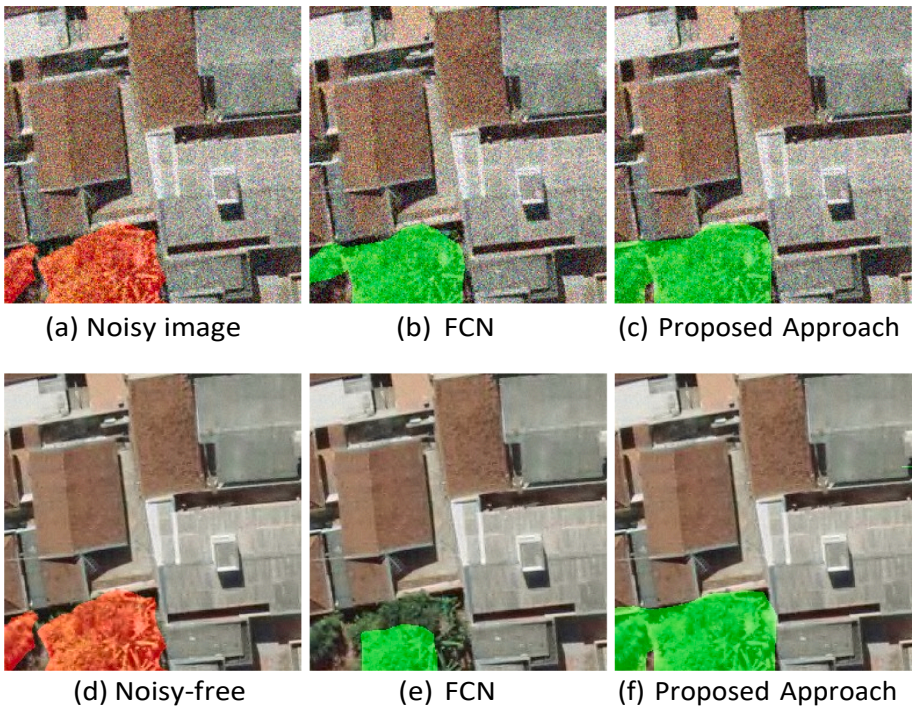


Fig. 8. Comparative results of the proposed approach and FCN trained in noisy images. The first row of images shows the segmentation using a noisy test image, while the second row of images shows the results using a noise-free test image.

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

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.

Fig. 2 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.

To evaluate the proposed approach, we used two well-known semantic segmentation methods: SegNet (Badrinarayanan et al., 2017; Dowden et al., 2021) and FCN (Long et al., 2015). SegNet (Badrinarayanan et al., 2017) is a CNN with encoder and decoder networks, with a final pixel-wise classification layer. The encoder provides a low-resolution activation map representing the most important features for each input. In this study, the encoder is composed of the convolutional and max-pooling layers of VGG16 (Simonyan and Zisserman, 2014). 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.

The FCN (Long et al., 2015) extends the standard classification CNN (VGG16 (Simonyan and Zisserman, 2014)) 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.

It is important to highlight that the proposed strategy can be adopted considering any semantic segmentation method. As previous studies showed that even some traditional deep learning methods outperformed state-of-the-art methods, here we focused only on showing the benefits of adopting the proposed approach compared to the baselines (method not adopting the strategy).

4. Experiments and Results

4.1. Image Datasets

Initially, we considered a dataset for semantic segmentation of urban trees. This dataset has the challenges of class imbalance and labeling uncertainty. Fig. 3 presents examples illustrating the challenges of semantic segmentation methods. The trees in Fig. 3 show that 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. 3 also illustrates the labeling challenge, in which some parts of the object are not visible in the image due to noise when capturing images.

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. The pixels of this dataset were labeled in two classes: trees and background. Examples of the Urban Tree dataset in Fig. 3 show that the boundaries of the trees are difficult to label. This dataset is composed of 966 non-overlapping patches of 256×256 pixels. In the experiments, 580, 193, and 193 patches were randomly used for training, validation, and testing, respectively.

Although this work focuses on tree segmentation from aerial images, an additional experiment considering a more drastic imbalance situation was conducted using terrestrial imagery. This experiment assesses the robustness of the approach among other types of images and challenging scenarios. The dataset is described as follows.

Soybean Disease (SD). The images from this dataset were obtained through PlantVillage (Hughes and Salathé, 2015), which contains several photographs taken by cell phones in soybean plantations. To compose the image dataset, 201 images with the frog-eye disease were identified and manually annotated as shown in Fig. 4. Thus, this dataset is composed of two classes: frog-eye disease and background. 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.

4.2. Experimental Setup

For the Urban Tree (UT) and Soybean Disease (SD) datasets, the images were resized to 256 × 256 and 1024 × 1024 pixels, respectively. We chose 1024x1024 pixels for the SD dataset due to the high resolution of the original images. Also, the soybean disease class occupies a small area in the original image, and resizing to 1024 × 1024 pixels ensures that the class occupies a reasonable amount of pixels (see Fig. 4).

For all segmentation methods, we use Stochastic Gradient Descent (SGD) optimizer with a learning rate of 0.001, momentum of 0.9 and weight decay of 0.0005. The number of epochs was 100 with a batch size equal to 4 for UT dataset and 2 for SD dataset. Due to the higher resolution of the SD dataset images, the batch size has been reduced to fit the GPU memory. The number of epochs, learning rate, momentum, and weight decay (same for both datasets) were defined after empirical experiments with the validation set that presented the best learning convergence. The backbone weights of the segmentation methods started with pre-trained weights on ImageNet.

We use the following popular segmentation metrics to evaluate the proposed approach and baselines: pixel accuracy (PA) and intersection over union (IoU). In semantic segmentation, these two metrics are consolidated and used in most works. 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., trees).

4.3. 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, which 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 74.4 to 83.8% in Urban Tree dataset, and 3.5 to 77.7% in Soybean Disease dataset). The proposed approach also showed superior IoU results, especially in Urban Tree and SD datasets, where IoU improved from 67.6 to 70.5%, and from 32.4 to 56.7%, respectively. Further, it is found that using $\sigma = 2$ provided the best result in both the Urban Tree and SD datasets. A lower value of σ for Urban Tree and SD datasets is expected due to the size of the foreground.

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 and SD datasets, considerable increases of 8% and 23.9% were obtained in the pixel accuracy, respectively. On the other hand, IoU obtained by the proposed approach was slightly higher in the UT dataset and lower in the SD dataset. Hence, the approach described here has proven to be effective for two datasets that include challenges of class imbalance and labeling uncertainty and for two semantic segmentation methods.

4.4. Discussion and Qualitative Results

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

Urban tree dataset. Fig. 5 presents two examples that show the advantages 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.

Soybean disease dataset. As shown in Fig. 6, the proposed approach was able to segment soybean diseases with high 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.

4.5. Noise Invariance

Noise invariance of semantic segmentation methods was assessed on the Urban Tree dataset. Gaussian noise with a standard deviation of 0.02 was added to the images. We chose this value after empirical testing in order to obtain images with a medium severity, 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. For the proposed method, we used the configuration that obtained the best results (see Tables 1 and 2), i.e., with a loss function considering the unbalance and the labeling uncertainty ($\sigma = 2$).

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., 87.5% versus 77.6% and 69.7% versus 68.6%).

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 the 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 87.5% and 84.7% in test images with and without noise, a drop of only 0.2%.

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).

5. Conclusions

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. The robustness of the proposed approach can be ascertained for the two datasets considered, which presented different characteristics and challenges.

The results showed that the proposed approach obtains superior metrics regardless of the segmentation method adopted (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 in segmenting the datasets. Our approach also proved to be more invariant to noise, even when training was performed on noisy images and tested on noise-free images. Further research should include the application of the proposed approach to segmentation problems with several classes in other situations.

Funding

This research was funded by CNPq (p: 433783/2018–4, 310517/2020–6, 314902/2018–0, 304052/2019–1 and 303559/2019–5), FUNDECT (p: 59/300.066/2015) and CAPES PrInt (p: 88881.311850/2018–01). The authors acknowledge the support of the UFMS (Federal University of Mato Grosso do Sul) and CAPES (Finance Code 001). This research was also partially supported by the Emerging Interdisciplinary Project of Central University of Finance and Economics.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to acknowledge Nvidia Corporation for the donation of the Titan X graphics card.

References

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39, 2481–2495.

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

Bischke, B., Helber, P., Borth, D., Dengel, A., 2018. Segmentation of imbalanced classes in satellite imagery using adaptive uncertainty weighted class loss, in: IGARSS, pp. 6191–6194.

Bulò, S.R., Neuhold, G., Kontschieder, P., 2017. Loss max-pooling for semantic image segmentation, in: CVPR, pp. 7082–7091.

Caesar, H., Uijlings, J., Ferrari, V., 2015. Joint calibration for semantic segmentation, in: BMVC, BMVA Press. pp. 29.1–29.13.

Castellanos, F.J., Valero-Mas, J.J., Calvo-Zaragoza, J., Rico-Juan, J.R., 2018. Oversampling imbalanced data in the string space. Pattern Recogn. Lett. 103, 32–38.

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

Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation, in: ECCV, Springer International Publishing. pp. 833–851.

Chrabaszcz, P., Loshchilov, I., Hutter, F., 2017. A downsampled variant of imagenet as an alternative to the CIFAR datasets. CoRR abs/1707.08819. arXiv:1707.08819.

Dal Pozzolo, A., Caelen, O., Bontempi, G., 2015. When is undersampling effective in unbalanced classification tasks?, in: Appice, A., Rodrigues, P.P., Santos Costa, V., Soares, C., Gama, J., Jorge, A. (Eds.), Machine Learning and Knowledge Discovery in Databases, Cham. pp. 200–215.

Deng, J., Dong, W., Socher, R., Li, L., Kai Li, Li Fei-Fei, 2009. Imagenet: A large-scale hierarchical image database, in: CVPR, pp. 248–255.

Ding, H., Jiang, X., Liu, A.Q., Thalmann, N.M., Wang, G., 2019. Boundary-aware feature propagation for scene segmentation, in: ICCV, pp. 6819–6829.

Dong, Q., Gong, S., Zhu, X., 2019. Imbalanced deep learning by minority class incremental rectification. IEEE Trans. Pattern Anal. Mach. Intell. 41, 1367–1381.

Dowden, B., De Silva, O., Huang, W., Oldford, D., 2021. Sea ice classification via deep neural network semantic segmentation. IEEE Sens. J. 21, 11879–11888. https://doi.org/10.1109/JSEN.2020.3031475.

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

Ha, J., Lee, J.S., 2016. A new under-sampling method using genetic algorithm for imbalanced data classification, in: International Conference on Ubiquitous Information Management and Communication, ACM, New York, NY, USA. pp. 95:1–95:6.

Hamaguchi, R., Fujita, A., Nemoto, K., Imaizumi, T., Hikosaka, S., 2018. Effective use of dilated convolutions for segmenting small object instances in remote sensing imagery, in: WACV, pp. 1442–1450.

Huang, C., Li, Y., Loy, C.C., Tang, X., 2016. Learning deep representation for imbalanced classification, in: CVPR, pp. 5375–5384.

Hughes, D.P., Salathé, M., 2015. 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. arXiv:1511.08060.

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

Johnson, J.M., Khoshgoftaar, T.M., 2019. Survey on deep learning with class imbalance. Journal of Big Data 6, 27.

Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86, 2278–2324.

Li, J., Liu, L.s., Fong, S., Wong, R.K., Mohammed, S., Fiaidhi, J., Sung, Y., Wong, K.K.L., 2017. Adaptive swarm balancing algorithms for rare-event prediction in imbalanced healthcare data. PLOS ONE 12, 1–25.

Lin, T., Goyal, P., Girshick, R., He, K., Dollár, P., 2020. Focal loss for dense object detection. IEEE Trans. Pattern Anal. Mach. Intell. 42, 318–327.

Liu, B., Tsoumakas, G., 2019. Dealing with class imbalance in classifier chains via random undersampling. Knowl.-Based Syst. 105292.

Lobo Torres, D., Queiroz Feitosa, R., Nigri Happ, P., Elena Cué La Rosa, L., Marcato Junior, J., Martins, J., Olã Bressan, P., Gonçalves, W.N., Liesenberg, V., 2020a. 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. URL: https://www.mdpi.com/1424-8220/20/2/563.

Lobo Torres, D., Queiroz Feitosa, R., Nigri Happ, P., Elena Cué La Rosa, L., Marcato Junior, J., Martins, J., Olã Bressan, P., Gonçalves, W.N., Liesenberg, V., 2020b. Applying fully convolutional architectures for semantic segmentation of a single tree species in urban environment on high resolution uav optical imagery. Sensors 20.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation, in: CVPR, pp. 3431–3440.

López, V., Fernández, A., García, S., Palade, V., Herrera, F., 2013. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. Inf. Sci. 250, 113–141.

Martins, J.A.C., Nogueira, K., Osco, L.P., Gomes, F.D.G., Furuya, D.E.G., Gonçalves, W.N., Sant'Ana, D.A., Ramos, A.P.M., Liesenberg, V., dos Santos, J.A., de Oliveira, P.T.S., Junior, J.M., 2021. Semantic segmentation of tree-canopy in urban environment with pixel-wise deep learning. Remote Sensing 13. https://doi.org/10.3390/rs13163054. URL: https://www.mdpi.com/2072-4292/13/16/3054.

Nekooeimehr, I., Lai-Yuen, S.K., 2016. Adaptive semi-unsupervised weighted oversampling (a-suwo) for imbalanced datasets. Expert Syst. Appl. 46, 405–416.

Osco, L.P., Nogueira, K., Marques Ramos, A.P., Faita Pinheiro, M.M., Furuya, D.E.G., Gonçalves, W.N., de Castro Jorge, L.A., Marcato Junior, J., dos Santos, J.A., 2021. Semantic segmentation of citrus-orchard using deep neural networks and multispectral uav-based imagery. Precision Agric. 22, 1171–1188. https://doi.org/10.1007/s11119-020-09777-5.

Ren, M., Zeng, W., Yang, B., Urtasun, R., 2018. Learning to reweight examples for robust deep learning, in: ICML, pp. 4331–4340.

Shen, W., Wang, X., Wang, Y., Bai, X., Zhang, Z., 2015. Deepcontour: A deep convolutional feature learned by positive-sharing loss for contour detection, in: CVPR, pp. 3982–3991.

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

Sun, B., Chen, H., Wang, J., Xie, H., 2018. Evolutionary under-sampling based bagging ensemble method for imbalanced data classification. Frontiers of Computer Science 12, 331–350.

Tian, T., Chu, Z., Hu, Q., Ma, L., 2021. Class-wise fully convolutional network for semantic segmentation of remote sensing images. Remote Sensing 13. https://doi.org/10.3390/rs13163211. URL: https://www.mdpi.com/2072-4292/13/16/3211.

Tsai, C.F., Lin, W.C., Hu, Y.H., Yao, G.T., 2019. Under-sampling class imbalanced datasets by combining clustering analysis and instance selection. Inf. Sci. 477, 47–54.

Wu, Z., Shen, C., van den Hengel, A., 2016. High-performance semantic segmentation using very deep fully convolutional networks. CoRR abs/1604.04339. arXiv:1604.04339.

Xu, J., Schwing, A.G., Urtasun, R., 2015. Learning to segment under various forms of weak supervision, in: CVPR, pp. 3781–3790.