

Article

VEPL-Net: A Deep Learning Ensemble for Automatic Segmentation of Vegetation Encroachment in Power Line Corridors Using UAV Imagery

Mateo Cano-Solis ^{1,*} , John R. Ballesteros ¹ and German Sanchez-Torres ² 

¹ Facultad de Minas, Universidad Nacional de Colombia, Medellín 050041, Colombia; jballes@unal.edu.co

² Facultad de Ingeniería, Universidad del Magdalena, Santa Marta 470004, Colombia; gsanchez@unimagdalena.edu.co

* Correspondence: macanoso@unal.edu.co

Abstract: Vegetation encroachment in power line corridors remains a major challenge for modern energy-dependent societies, as it can cause power outages and lead to significant financial losses. Unmanned Aerial Vehicles (UAVs) have emerged as a promising solution for monitoring infrastructure, owing to their ability to acquire high-resolution overhead images of these areas quickly and affordably. However, accurate segmentation of the vegetation encroachment in this imagery is a challenging task, due to the complexity of the scene and the high pixel imbalance between the power lines, the vegetation and the background classes. In this paper, we propose a deep learning-based approach to tackle this problem caused by the original and different geometry of the objects. Specifically, we use DeepLabV3, U-Net and a modified version of the U-Net architecture with VGG-16 weights to train two separate models. One of them segments the dominant classes, the vegetation from the background, achieving an IoU of 0.77. The other one segments power line corridors from the background, obtaining an IoU of 0.64. Finally, ensembling both models into one creates an “encroachment” zone, where power lines and vegetation are intersected. We train our models using the Vegetation Encroachment in Power Line Corridors dataset (VEPL), which includes RGB orthomosaics and multi-label masks for segmentation. Experimental results demonstrate that our approach outperforms individual networks and original prominent architectures when applied to this specific problem. This approach has the potential to significantly improve the efficiency and accuracy of vegetation encroachment monitoring using UAV, thus helping to ensure the reliability and sustainability of power supply.

Keywords: deep learning; vegetation encroachment monitoring; semantic segmentation; power line corridors; UAV; GeoAI



Citation: Cano-Solis, M.; Ballesteros, J.R.; Sanchez-Torres, G. VEPL-Net: A Deep Learning Ensemble for Automatic Segmentation of Vegetation Encroachment in Power Line Corridors Using UAV Imagery. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 454. <https://doi.org/10.3390/ijgi12110454>

Academic Editors:
Diego González-Aguilera,
Pablo Rodríguez-Gonzálvez and
Wolfgang Kainz

Received: 11 August 2023
Revised: 13 October 2023
Accepted: 25 October 2023
Published: 6 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Vegetation encroachment is a common situation in power line corridors, causing flashovers in transmission and distribution overhead lines. Overgrown trees near transmission lines can disrupt power circuits, generating short circuits and tripping circuit breakers, which may result in power outages. These disruptions lead to economic losses for utility companies and widespread blackouts for consumers, especially impacting businesses heavily reliant on electricity these days. It is crucial for electric utility companies to monitor and manage vegetation encroachment near transmission lines to ensure continuous power supply and prevent damage to conductors, in alignment with electricity regulations [1]. In 2003, in the United States and Canada there was a power outage that affected close to 50 million users (11% of the population of both countries). Similar incidents also occurred in Italy and Switzerland during the same year, affecting more than 60 million users. In tropical countries like Malaysia, where approximately 66% of the country is covered with forest, vegetation encroachment is the third biggest cause of electricity supply interruption,

representing about 18% of total failures in power cuts [1]. In other studies such as in [2], the authors account for the monetary losses in United States companies due to vegetation encroachment, and they calculated that a 30 min outage in a medium company can cost up to USD 16,500, reaching up to USD 94,000 for an 8 h suspension.

Despite the importance of monitoring vegetation encroachment in transmission lines, the most used technique is human-based field trips, known to be a time-consuming task and logistically intensive due to the long journeys and harsh natural conditions that must be faced. Consequently, this technique suffers from poor scalability, as reported in [2–5]. To overcome these limitations, refs. [2–6] proposed applying deep learning combined with Unmanned Aerial Vehicle (UAV) images of transmission lines. Furthermore, optical imagery captured by UAVs are easy to collect, analyze and store. However, there is a lack of UAV-based datasets that can be used to train neural networks (NNs), and specifically for the purpose of segmenting vegetation encroachment in infrastructure. This limitation was first addressed in the Vegetation Encroachment in Power Line Corridors (VEPL) dataset, a collection of UAV images of vegetation encroachment in power line corridors acquired in Colombia, South America [6]. This dataset includes three folders of paired image-masks: the original imagery, geometric-augmented imagery and one with spectral augmentation. The images are RGB, and the multi-class masks have three classes that represent the vegetation, the power lines and the background. Some of the challenges imposed by this dataset are:

- Class imbalance at pixel level between the background, the vegetation and the power line classes [7].
- The vegetation class can incorporate an extensive variety of trees and shrubbery, with different colors, shapes and heights.
- In some cases, power lines overlap vegetation when modeled in two dimensions, making semantic segmentation inaccurate.

According to [8], semantic segmentation is the classification of images at the pixel level, which makes the class imbalance become an inherent problem. In [9], it is mentioned that this task faces a tension between semantics and location: global information resolves the “what”, while local information resolves the “where”. To test their hypothesis, those authors used a Fully Convolutional Network pre-trained with different classification networks like AlexNet and VGGNet. In [10], the authors developed one of the most widely used state-of-the-art algorithms in semantic segmentation, called the DeepLab. It is a Deep Convolutional Network that includes the *atrous convolution*, upsampled filters for dense feature extraction that use extended spatial pyramid pooling (ASPP). ASPP enhances the network’s ability to capture multi-scale features from the input image by utilizing dilated convolutions at different rates. ASPP is particularly effective in capturing context at various scales, which is crucial for accurately segmenting objects of different sizes. Authors in [11] proposed the USPP, a CNN framework introduced to generate segmentation on high-resolution remote sensing images. The USPP is a novel model that showcases the effectiveness of integrating the encoder–decoder and spatial pyramid pooling module for segmentation purposes. This pooling method encodes objects and image context at multiple scales. Also, some solutions are developed for mobile devices, like MobileNet [12], that used segmentation decoder Lite Reduced Atrous Spatial Pyramid Pooling (LR-ASPP). Authors in [9] presented a module with dilated convolutions to systematically aggregate multi-scale contextual information without losing resolution for Road Extraction over aerial images. This network is built with residual units, addressing the issue of class imbalance on a similar architecture to the U-Net.

Different authors have faced high imbalanced classes in remote sensing. In [13], authors proposed using a pre-trained encoder as backbone for generating faster and more precise predictions, showing that transfer learning is a useful technique for remote sensing imbalance classes, like road extraction from satellite images. In [14], authors also dealt with class imbalance when using semantic segmentation and remote sensing for mapping land cover in urban areas, and they made use of customized loss functions and combined CNN

architectures to obtain the best overall performance. In a survey developed by [15], multiple solutions for improving segmentation in remote sensing were described. One of the most used was fusion-based strategies, so when the input data had a different structure or geometry, a separate network was applied to handle each data type, and fusion happened at the classification stage. Also, the importance of custom loss functions to handle imbalanced datasets is highlighted. Lastly, different authors found out that the use of Transfer Learning helps to tackle the scarcity of data and models to gain more prediction power.

In this paper, the semantic segmentation of a highly imbalanced classes problem, exhibited by the vegetation encroachment in power lines, is tackled by the combination of a pre-trained Vgg-16, U-Net, and DeepLab3 networks, with the customized Tversky loss function, which is coined as the VEPL-Net [16]. Employed architectures have previously shown good performance, low complexity, computational efficiency and good results with UAV-based imagery [15]. The Tversky loss function, based on the Tversky index, has demonstrated its ability to achieve a better trade-off between precision and recall in convolutional deep neural networks (CNNs) [17–23]. The objective of this study is to propose an alternative approach that helps to automatically segment vegetation encroachment in power lines using UAV images.

The rest of the paper is organized as follows. In Section 2, we describe the VEPL dataset and our methodology for semantic segmentation on this dataset. The details of the VEPL-Net architecture and the pre-trained weights used are presented, followed by the ensemble composed of two separate neural networks for addressing the multi-class segmentation, as well as the data augmentation techniques used, the description of the loss function and the training methodology.

In Section 3, the experiment results, the performance evaluation and an analysis of the VEPL-Net are provided. In Section 4, future research in this field is presented. Finally, the contributions of this study are summarized.

2. Materials and Methods

Semantic segmentation for vegetation encroachment in power lines still has multiple unsolved challenges, including dataset unavailability and highly imbalanced classes. This section describes the materials and methodology employed to carry out the study.

2.1. VEPL Dataset

The VEPL dataset comprises orthomosaics that have been tessellated to generate pairs of images and masks, representing three distinct classes: vegetation, power lines and background. This dataset was specifically developed for the semantic segmentation of vegetation encroachment in power lines, offering a significant advantage in training Deep Learning models for monitoring several kilometers in less time, and at lower cost compared to the conventional field trips approach [6].

The VEPL dataset was acquired through autonomous drone flights along a secondary road located in Envigado, Colombia, South America. The dataset covers approximately 2.4 km of roads, resulting in a total of 532 pairs of image-mask chunks. This number was expanded to 3724 chunks using geometric augmentation, and to 3192 image-mask pairs using spectral augmentation [6]. Figure 1 shows an example of the VEPL dataset.

Geometric augmentation involves altering the geometric properties of images to ensure that neural networks can effectively handle changes in object position and orientation. This encompasses random rotations, grid distortions, horizontal flips, scale shifts and elastic deformations. Spectral augmentation aims to enhance model robustness against variations in lighting and color within images. The employed techniques were Random Brightness and Contrast, Hue Saturation Changes, Gaussian Blur Filter, Gamma Correction and CLAHE (Contrast Limited Adaptive Histogram Equalization). Figure 2 shows examples of the geometric and spectral augmentation in the VEPL dataset [6].



Figure 1. The VEPL dataset. (a) RGB tessellated image, with a size of 256×256 pixels. (b) Multi-label mask with *vegetation class* (green), *power line class* (gray) and *background class* (black).

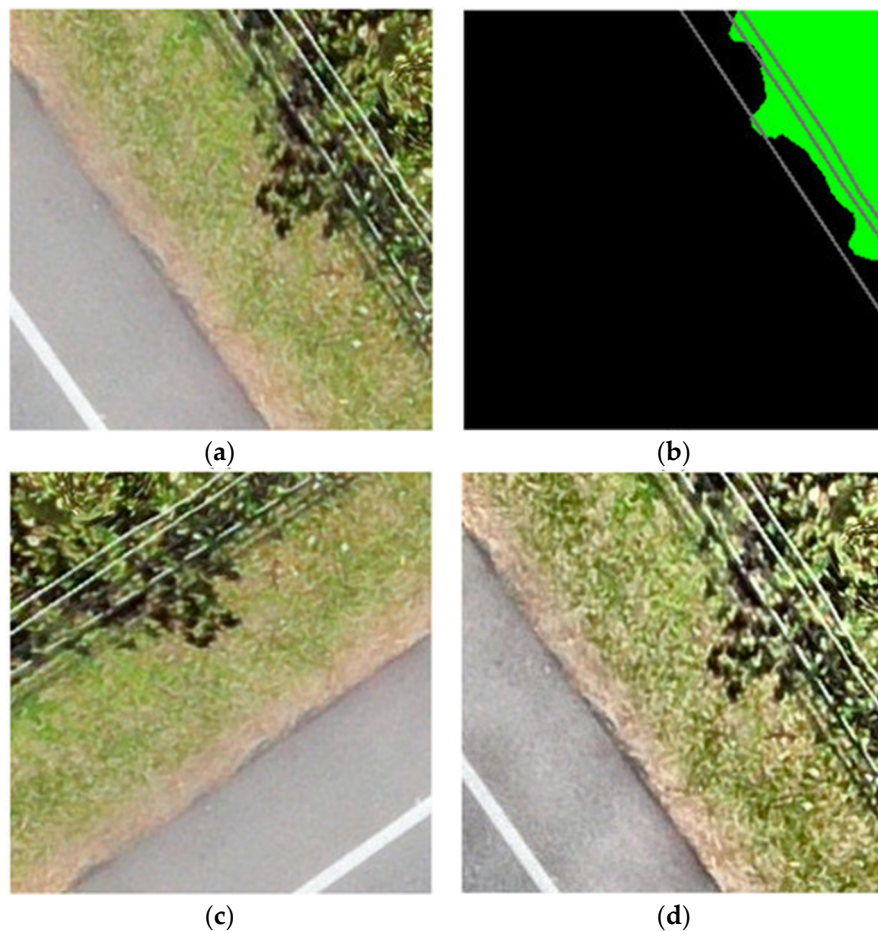


Figure 2. Geometric and spectral augmentation in the VEPL dataset. (a) Example of an RGB image, with a size of 256×256 pixels. (b) Corresponding multi-label mask with vegetation class (green), power line class (gray) and background class (black). (c) RGB image with geometric data augmentation applied (RandomRotate90). (d) RGB image with spectral data augmentation applied (CLAHE, Apply Contrast Limited Adaptive Histogram Equalization). Adapted and modified from [6].

Despite executing imbalance checks and data augmentation, the VEPL dataset still presents a high imbalance between classes, with the power line class being particularly scarce due to its geometry (linear) compared to vegetation and background classes (polygons). The VEPL dataset is freely available in Zenodo at <https://doi.org/10.5281/zenodo.7800234> (accessed on 1 August 2023).

2.2. Deep Learning Architectures

The selection of an appropriate architecture is crucial for semantic segmentation. In this section, we delve into two prominent deep learning architectures tailored for that purpose. The first architecture explored is the U-Net, which represents one of the pioneering solutions for addressing semantic segmentation. The other one is that proposed in DeepLab [10]. After that, the key aspects of the VEPL Net are explained.

2.2.1. U-Net

The U-Net architecture was originally proposed in 2015 in [24], and originally the research was focused on biomedical image segmentation. It consists of an encoder and decoder path, where the encoder, or contracting part, is a convolutional neural network (CNN) that captures context, and the decoder, the expanding part, enables the precise location of features. The typical use of a CNN is the classification task, but in many challenges of deep learning, the desired output should include a pixel-localization of each class [24]. In the VEPL-Net, classification and localization are mandatory due to the necessity of identifying and locating the vegetation and the power line pixels. The encoder in the U-Net involves applying two consecutive 3×3 convolutions, each followed by a rectified linear unit (ReLU) activation function, and a 2×2 max pooling operation with a stride of 2. This downsampling process doubles the number of feature channels. On the other hand, the decoder step includes an upsampling operation on the feature map, followed by a 2×2 convolution that reduces the number of feature channels by half. The resulting feature map is then concatenated with the corresponding cropped feature map from the contracting path. Subsequently, two 3×3 convolutions are applied, each followed by a ReLU activation. The cropping step is necessary to account for the loss of border pixels that occurs during each convolution. Finally, a 1×1 convolutional layer is employed to map each 64-component feature vector to the desired number of classes. Altogether, the network consists of a total of 23 convolutional layers [24]. Figure 3 presents the U-Net architecture.

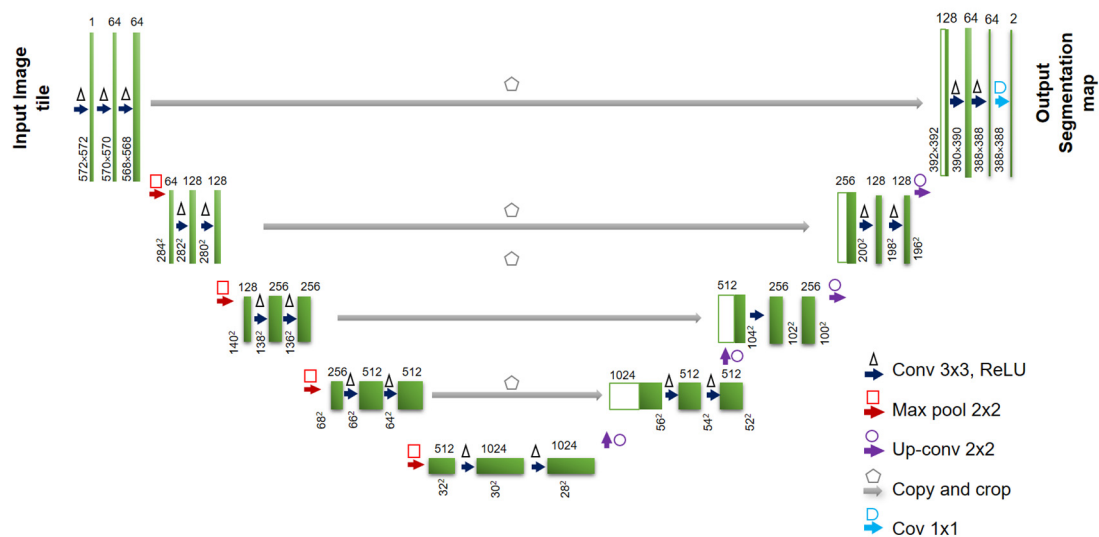


Figure 3. U-Net architecture (example for 32×32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations. Source: adapted from [24].

Multiple implementations of a U-Net-like backbone are conducted when dealing with satellite imagery. In [17], the authors proposed ResUNet-a, which combines a novel loss function based on the Dice loss function for the task of semantic segmentation of high-resolution aerial images. ResUNet incorporates a UNet encoder/decoder backbone

with residual connections, atrous convolutions, pyramid scene parsing pooling and multi-tasking inference. Additionally, in [18], a novel end-to-end change detection method is proposed based on the UNet++ architecture for semantic segmentation. This approach leverages an effective encoder–decoder structure to generate high-accuracy change maps from co-registered image pairs. The fusion of multiple side outputs at different semantic levels produces a final change map with superior accuracy, outperforming other state-of-the-art methods on very high-resolution satellite image datasets. Finally, in [19], the authors proposed RCNN-UNet, an end-to-end deep learning model for road information extraction from aerial images. This model mitigates propagation errors, leverages spatial context and low-level features and employs multi-task learning for road detection and centerline extraction.

2.2.2. DeepLab

In [10], the authors emphasize the effectiveness of atrous convolution, which involves convolution with upsampled filters. This technique allows the explicit control of the resolution at which features are computed within Deep Convolutional Neural Networks (DCNNs). Atrous convolution enables the incorporation of larger context without increasing the number of parameters or computational complexity. The authors also propose atrous spatial pyramid pooling (ASPP) to achieve robust object segmentation at multiple scales. ASPP applies filters with varying sampling rates and effective fields-of-view to the incoming convolutional feature layers, thereby capturing objects and image context at different scales. Third, the authors enhance the localization of object boundaries by combining methods from DCNNs and probabilistic graphical models. While the common combination of max-pooling and downsampling layers in DCNNs provides invariance, it often compromises localization accuracy. To address this issue, they integrate the responses from the final DCNN layer with a fully connected Conditional Random Field (CRF). This fusion approach demonstrated through qualitative and quantitative analysis a significant improvement of the localization performance.

2.2.3. Transfer Learning

Data dependence is one of the most cited problems in deep learning, and at the same time, data collection is a complex and expensive task, which makes it extremely difficult to build a large-scale, high-quality annotated dataset [25]. Transfer learning is a valuable technique for addressing the challenge of limited training data. It aims to leverage knowledge from a source domain to a target domain, overcoming the assumption that training and test data must be independent and identically distributed. This approach has proven highly beneficial for domains with insufficient training data, enabling notable improvements in model performance [25]. The authors in [20] have used VGG-16 and inception to incorporate accurate U-Net models for image segmentation, specifically as encoders. Additionally, the authors in [26] use transfer learning capabilities of FCNs in various satellite imagery datasets, retrieving information on small-scale urban structures even at decametric geospatial resolution. In other studies, such as in [27], VGG16 is used without the top layer as the encoder for crack detection in pavements and bridges. Similarly, in [28], the authors implement transfer learning between different crop types, reducing training times up to 80%.

2.3. Loss Function

The loss function can deeply affect the learning process of a model, and the right loss function selection is key to account for the imbalance problems that are frequent in semantic segmentation. For classification models, the most used loss function is the Cross Entropy loss, and for the regression models, the L1 and L2 are the most common, whereas for semantic segmentation authors lean on the Categorical Cross Entropy loss and Dice similarity [29]. In semantic segmentation, the usual technique combines loss functions and weighted schemas, trying to handle the imbalance of the minor class and the high

presence of the background pixels, penalizing the dominant and giving more attention to the minority class [29].

The Focal Loss is one of the recent solutions for class imbalance when conducting semantic segmentation; it reshapes the cross entropy loss such that it down-weights the loss to well-classified examples, reducing the relative loss for well-classified pixels and putting more focus on hard, misclassified ones [29,30]. The authors in [29] performed a review on the loss functions of existing architectures for semantic segmentation, including some like multimodal logistic loss in the FCN. DeepLab authors use as a loss function the sum of cross-entropy terms for each spatial position in the CNN output map. PSPNet implemented two loss functions, one for the main branch using cross-entropy loss to train the final classifier and another one after the fourth stage, both with a weighted balancing. Finally, the Gradient Difference Loss was used in SegmPred, which is designed to sharpen results by penalizing high-frequency mismatches such as errors along the object boundaries. One of these functions is known as Tversky loss [16], which is specifically designed for image segmentation in medical imaging applications, particularly for segmenting lesions. In such cases, the number of voxels representing lesions is often significantly lower than the number of voxels representing non-lesions. The authors of the study suggested a generalized loss function based on the Tversky index, which achieved a more optimal balance between precision and recall during the training of 3D fully convolutional deep neural networks.

2.4. Training Strategy

We initially attempted to train the network using all three classes of the VEPL dataset. It was challenging to accurately identify the power lines due to their linear geometry, which resulted in a relatively small number of pixels compared to the other classes. Consequently, we then adopted the following strategy:

- All three classes must be present in each image-mask pair. Therefore, we decided to create two datasets based on the VEPL dataset. One dataset contained the vegetation and background classes, while the other contained the power line class. This approach transformed the problem from a multi-class to a binary classification problem.
- We made use of a custom loss function for training, specifically the Tversky loss function. This yielded better results compared to traditional loss functions since it provided more emphasis on the minority classes. Both the vegetation and power line classes have a lower presence than the background class.
- We employed the U-Net and DeepLab architectures, considering pre-trained weights to enhance the neural network's generalization capabilities.
- The evaluation metrics were specific for semantic segmentation tasks.

3. Results and Discussion

Three distinct neural network architectures were explored: U-Net [24], U-Net with a VGG-16 encoder [20,24,27] and DeepLab [10]. Each architecture introduced unique attributes for semantic segmentation. The VEPL dataset was used with and without augmentation. By modifying the training dataset composition, the loss function and the neural network architecture, up to 36 different neural networks were trained in total—18 for the vegetation-background dataset and 18 for the power line-background dataset. The following are the variations chosen:

- Dataset composition: augmentations encompassed geometric and spectral transformations, thereby introducing comprehensive diversity into the training process.
- Neural network architecture: U-Net, U-Net with VGG-16 encoder and DeepLab.
- Loss function: Tversky loss and binary cross-entropy loss.

The results are presented in two subsections. The first presents the results obtained from all the trained neural networks, with a focus on the best solution for each dataset. The performance of all 36 trained neural networks and evaluations of their accuracy, among

other metrics related to semantic segmentation, are presented. Additionally, comparative results obtained using different datasets, augmentation techniques, neural network architectures and loss functions are highlighted. This analysis provides a comprehensive understanding of the effectiveness of various configurations in addressing the segmentation task.

The second one demonstrates the implementation of VEPL-Net, a strategy that combines the best solution for the power lines and the vegetation classes and identifies the vegetation encroachment. Quantitative metrics to demonstrate the effectiveness of VEPL-Net in addressing this specific problem are highlighted.

3.1. Performance Evaluation of Trained Neural Networks for Semantic Segmentation

All neural networks were trained using a GPU NVIDIA T4(×2) in Kaggle notebooks. In Tables 1 and 2, we present the results of training and validation accuracy for the power line and vegetation classes.

Table 1. Maximum training accuracy for all neural networks.

Class	Data Trained	Train Accuracy					
		DeepLab		Unet		Unet + vgg Encoder	
		BCE *	Tversky	BCE *	Tversky	BCE *	Tversky
Power Line	Without augmentation	-	-	0.820	0.674	0.689	0.511
	Geometric augmentation	0.976	0.969	0.968	0.820	0.969	0.861
	Spectral augmentation	0.946	0.906	0.965	0.749	0.972	0.975
Vegetation	Without augmentation	-	-	0.850	0.801	0.876	0.833
	Geometric augmentation	0.950	0.907	0.879	0.850	0.897	0.858
	Spectral augmentation	0.987	0.955	0.846	0.770	0.911	0.835

* BCE—Binary cross-entropy.

Table 2. Maximum validation accuracy for all neural networks.

Class	Data Trained	Validation Accuracy					
		DeepLab		Unet		Unet + vgg Encoder	
		BCE *	Tversky	BCE *	Tversky	BCE *	Tversky
Power Line	Without augmentation	-	-	0.914	0.776	0.738	0.616
	Geometric augmentation	0.981	0.977	0.981	0.945	0.981	0.948
	Spectral augmentation	0.968	0.938	0.982	0.890	0.982	0.981
Vegetation	Without augmentation	-	-	0.861	0.841	0.905	0.867
	Geometric augmentation	0.879	0.902	0.873	0.832	0.872	0.880
	Spectral augmentation	0.874	0.873	0.833	0.799	0.882	0.858

* BCE—Binary cross-entropy.

As can be seen in Tables 1 and 2, the use of augmentation leads to better performances in both the training and validation datasets. For the original dataset, without augmentation, DeepLab does not have results due to the low number of examples and the complexity of the task. For the vegetation class, it obtained a better performance in the dataset with geometric augmentation, showing consistency in both the training and validation sets. For the geometric augmentation the best architecture was the U-Net with the VGG-16 encoder, and the best loss function was Tversky, obtaining a good performance and similar results in both training and validation and showing no over or underfitting. For the power line class, the best dataset is the one that has geometric augmentation, but in this case with the DeepLab architecture. The results of the intersection over union (IoU), a popular metric for semantic segmentation [31], are included. mIoU is the average value of the intersection of the prediction and ground truth divided by their union, applied to the whole dataset [32]. When it comes to semantic segmentation, it is generally preferable to use Intersection over Union as an evaluation metric rather than for accuracy. IoU is a more appropriate metric for semantic segmentation because it accounts for spatial accuracy and

object localization, providing a better evaluation of the model's performance. Tables 3 and 4 present the results of the training and validation datasets using IoU for both classes. A better performance is obtained for the vegetation class using spectral augmentation, but with a little overfitting for some configurations. On the other hand, an overall equal performance is obtained for geometric augmentation in both the training and validation datasets. IoU has a drastic descensus for the power line class when compared to the accuracy metric; for this reason, the use of IoU is suggested for semantic segmentation. The power line class obtained better results for both augmentation methods when using DeepLab and the U-Net with the VGG-16 encoder.

Table 3. Maximum IoU values for the training dataset, all neural networks.

Train IoU							
Class	Data Trained	DeepLab		Unet		Unet + vgg Encoder	
		BCE	Tversky	BCE	Tversky	BCE	Tversky
Power Line	Without augmentation	-	-	0.054	0.061	0.040	0.059
	Geometric augmentation	0.310	0.450	0.069	0.107	0.077	0.129
	Spectral augmentation	0.062	0.175	0.060	0.090	0.083	0.506
Vegetation	Without augmentation	-	-	0.605	0.597	0.619	0.562
	Geometric augmentation	0.842	0.805	0.681	0.718	0.712	0.743
	Spectral augmentation	0.950	0.890	0.621	0.654	0.730	0.723

Table 4. Maximum IoU values for the validation dataset, all neural networks.

Val IoU							
Class	Data Trained	DeepLab		Unet		Unet + vgg Encoder	
		BCE *	Tversky	BCE *	Tversky	BCE *	Tversky
Power Line	Without augmentation	-	-	0.046	0.060	0.035	0.050
	Geometric augmentation	0.140	0.221	0.024	0.065	0.029	0.073
	Spectral augmentation	0.025	0.070	0.025	0.055	0.028	0.235
Vegetation	Without augmentation	-	-	0.754	0.680	0.816	0.649
	Geometric augmentation	0.764	0.824	0.716	0.768	0.719	0.779
	Spectral augmentation	0.815	0.795	0.666	0.754	0.758	0.800

* BCE—Binary cross-entropy.

Then all models were evaluated using the IoU over the original test dataset to select the best combination of augmentation, architecture and loss function for both classes. Table 5 shows the results.

Table 5. IoU for all models in the test dataset.

IoU Metric							
Class	Data Trained	DeepLab		Unet		Unet + vgg Encoder	
		BCE *	Tversky	BCE *	Tversky	BCE *	Tversky
Power Line	Without augmentation	-	-	0.504	0.427	0.344	0.295
	Geometric augmentation	0.593	0.639	0.515	0.490	0.575	0.523
	Spectral augmentation	0.514	0.548	0.517	0.494	0.534	0.625
Vegetation	Without augmentation	-	-	0.687	0.624	0.740	0.318
	Geometric augmentation	0.759	0.726	0.733	0.632	0.752	0.698
	Spectral augmentation	0.737	0.759	0.717	0.607	0.771	0.656

* BCE—Binary cross-entropy.

Data augmentation gave the model a better performance in IoU for both classes, and in general the use of Tversky loss [16] is more important for the power line class due to its high imbalance. For the vegetation class, the use of binary cross-entropy or Tversky loss does not have a significant impact. The best models for both classes were:

- Vegetation class: model trained with spectral augmentation dataset using U-Net with VGG-16 encoder and binary cross-entropy loss, obtaining up to 0.77 in IoU.
- Power line class: model trained with geometric augmentation dataset using DeepLab and Tversky loss, obtaining up to 0.64 in IoU.

The following Figure 4 show the prediction capacity of both models for predicting their respective class.

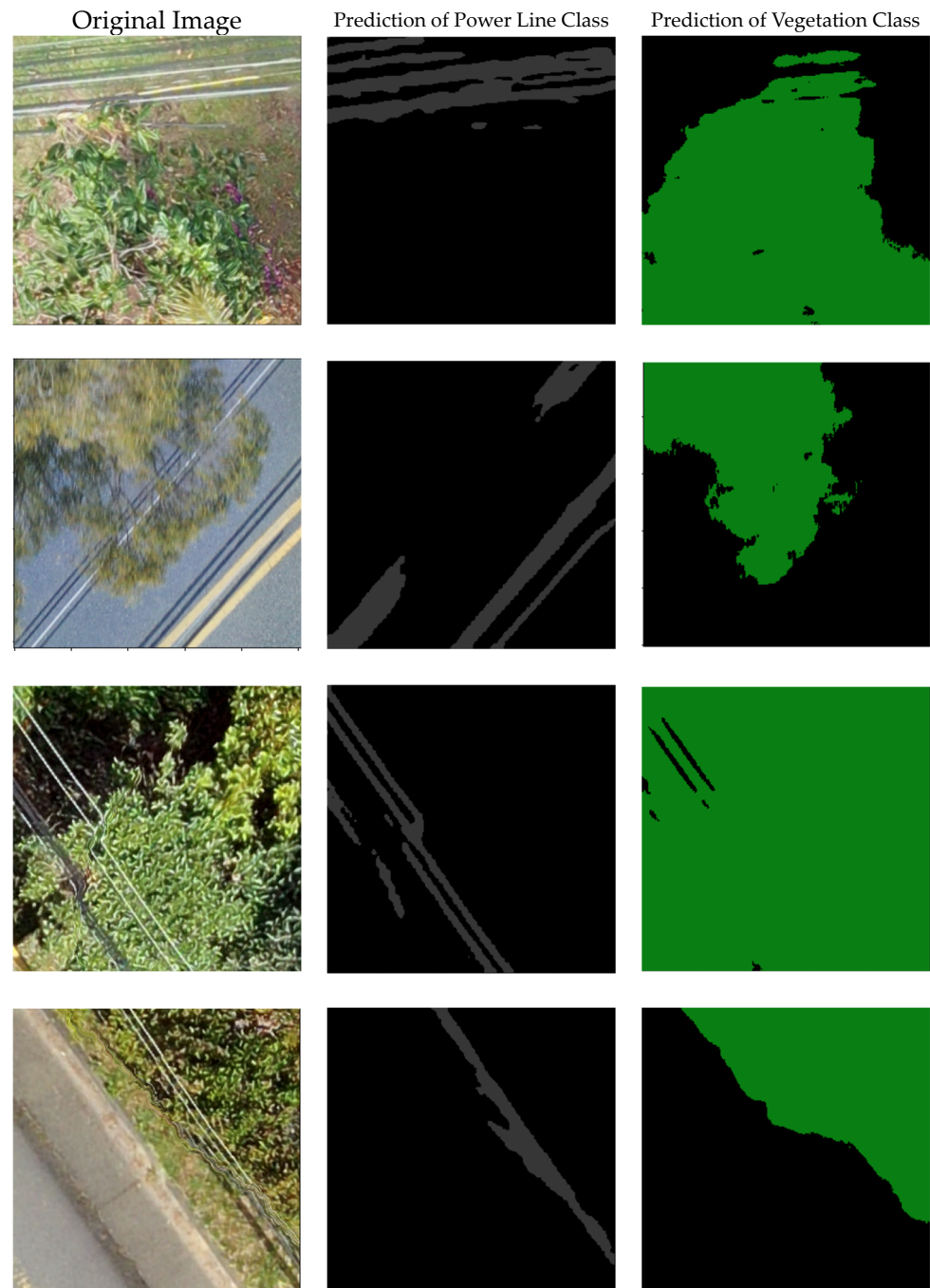


Figure 4. Examples of input images and the predictions obtained for best model for each class.

3.2. VEPL-Net: A Fusion of Neural Networks for Enhanced Segmentation of Vegetation Encroachment in Power Lines

The best performing neural networks for both classes according to previous results were chosen. The workflow for the predictions is shown in Figure 5.

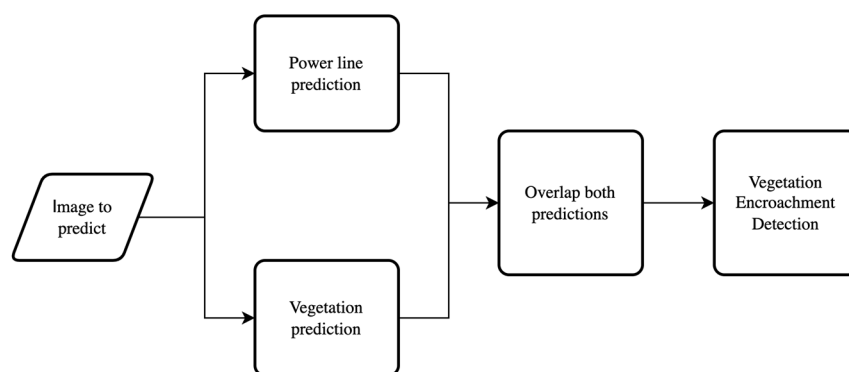


Figure 5. Proposed workflow for VEPL-Net.

The proposed workflow takes an image as input to predict vegetation invasion on power lines. Utilizing the best architecture obtained for each class, separate predictions are made for each class. Subsequently, intersection is performed between the two prediction masks. The final step generates an alert indicating the areas where vegetation invasion occurs. As mentioned, the advantage of this approach lies in the ability to generate more accurate predictions for each class, considering their significant geometric variations. Figure 6 shows multiple examples of input images and the final output with predictions and warning of vegetation encroachment.

The obtained predictions show a good performance of the model in segmenting both classes, in particular the power line class, despite its high imbalance, which also leads to a good prediction of vegetation encroachment. This study deals with the semantic segmentation of vegetation encroachment in power line corridors using deep learning. A customized network architecture and loss function called VEPL-Net is proposed. In the obtained results, data augmentation plays a crucial role in enhancing the performance of the model for vegetation and power line classes. For the vegetation class, the model trained with the spectral augmentation dataset, employing a U-Net with VGG-16 encoder and binary cross-entropy loss, and achieved the highest IoU of 0.77. This result suggests that utilizing spectral augmentation helps in capturing the variability in vegetation types, leading to more accurate segmentation. On the other hand, the power line class benefits significantly from the use of the Tversky loss function, the geometric augmentation and the DeepLab modules, obtaining an IoU of up to 0.64. The Tversky loss effectively addresses the class imbalance issue, resulting in an improved segmentation performance for power line corridors. Similar studies, like that developed by [33], used UAV and transformers to detect zones of vegetation encroachment, and the prediction for the zone of encroachment (mixed both power lines and vegetation) obtained a Jaccard index of 0.87. Others studies like [34] combined UAVs and deep learning, but in this case to detect transmission towers, reaching an accuracy of 98.6% for a DenseNet. Transformers were shown to be more effective in tackling the problem when a huge amount of training images was available or when the computing power was not a limitation, like in the case of this current research.

These findings indicate that different strategies are needed to optimize the segmentation models for each class separately, showing the effectiveness of the proposed VEPL-Net architecture and the significance of the penalizing loss function.

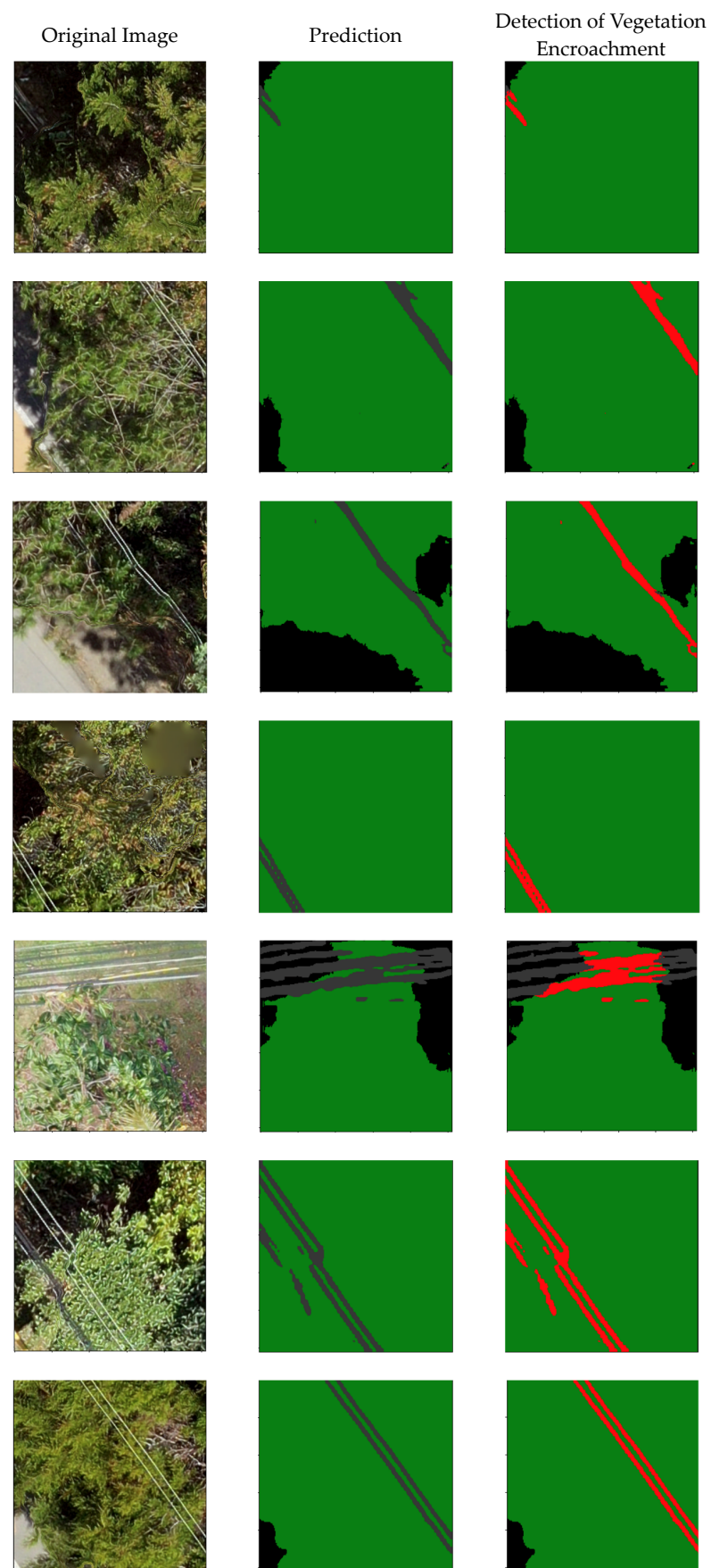


Figure 6. Examples of input images and the final output mask with predictions and warning of vegetation encroachment.

4. Conclusions

This work proposes the VEPL-NET as an alternative to automatically monitoring vegetation encroachment in power line corridors with the integration of deep learning techniques and UAV imagery.

The VEPL-Net is an effective approach, demonstrating its ability to handle data imbalance resulting when the target classes come from objects of a different geometry. This was possible due to the integration of an appropriate loss function and pre-trained weights coming from prominent segmentation architectures like the U-Net and DeepLab.

In the vegetation class, the application of spectral augmentation alongside a U-Net with VGG-16 encoder and binary cross-entropy loss achieved an IoU score of 0.77. Meanwhile, the power lines class reaped significant benefits from the Tversky loss function, which effectively addressed the pixel-level class imbalance inherent in this category. The use of a geometric augmentation dataset in conjunction with DeepLab and the Tversky loss led to an IoU of up to 0.64.

Experiments validate the effectiveness of the proposed VEPL-Net architecture and strategies but also emphasize the necessity of individualized approaches for addressing the complexities inherent in power line corridor monitoring. The synergy between machine learning techniques and UAVs has unlocked new possibilities for efficient, accurate and cost-effective monitoring, which holds the potential to benefit regions globally facing similar challenges.

Further research and refinement of the proposed method hold immense promise. The ongoing development of advanced neural network architectures, coupled with the continuous evolution of UAV technology, open doors to advancements in power line corridor monitoring, enhancing power supply reliability, reducing economic losses and achieving a sustainable energy infrastructure.

Author Contributions: Conceptualization, Mateo Cano-Solis, John R. Ballesteros and German Sanchez-Torres; data curation, Mateo Cano-Solis and John R. Ballesteros; formal analysis, Mateo Cano-Solis; investigation, Mateo Cano-Solis and John R. Ballesteros; methodology, Mateo Cano-Solis, John R. Ballesteros and German Sanchez-Torres; resources, German Sanchez-Torres; software, Mateo Cano-Solis and John R. Ballesteros; supervision, John R. Ballesteros and German Sanchez-Torres; validation, Mateo Cano-Solis, John R. Ballesteros and German Sanchez-Torres; visualization, Mateo Cano-Solis and John R. Ballesteros; writing—original draft, Mateo Cano-Solis; writing—review and editing, Mateo Cano-Solis, John R. Ballesteros and German Sanchez-Torres. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are openly available in Zenodo at <https://doi.org/10.5281/zenodo.7800234> (access on 1 August 2023).

Acknowledgments: We would like to thank GIDIA (Research group in Artificial Intelligence of the Universidad Nacional de Colombia) for providing guidelines, reviews to the work and the use of facilities.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ahmad, J.; Malik, A.S.; Xia, L.; Ashikin, N. Vegetation encroachment monitoring for transmission lines right-of-ways: A survey. *Electr. Power Syst. Res.* **2013**, *95*, 339–352. [[CrossRef](#)]
2. Nguyen, V.N.; Jenssen, R.; Roverso, D. Automatic autonomous vision-based power line inspection: A review of current status and the potential role of deep learning. *Int. J. Electr. Power Energy Syst.* **2018**, *99*, 107–120. [[CrossRef](#)]
3. Song, J.; Qian, J.; Li, Y.; Liu, Z.; Chen, Y.; Chen, J. Automatic Extraction of Power Lines from Aerial Images of Unmanned Aerial Vehicles. *Sensors* **2022**, *22*, 6431. [[CrossRef](#)] [[PubMed](#)]
4. Han, G.; Zhang, M.; Li, Q.; Liu, X.; Li, T.; Zhao, L.; Liu, K.; Qin, L. A Lightweight Aerial Power Line Segmentation Algorithm Based on Attention Mechanism. *Machines* **2022**, *10*, 881. [[CrossRef](#)]

5. Cheng, X. Research on the Application of Computer Vision Technology in Power System UAV Line Inspection. *E3S Web Conf.* **2022**, *358*, 01030. [\[CrossRef\]](#)
6. Cano-Solis, M.; Ballesteros, J.R.; Branch-Bedoya, J.W. VEPL Dataset: A Vegetation Encroachment in Power Line Corridors Dataset for Semantic Segmentation of Drone Aerial Orthomosaics. *Data* **2023**, *8*, 128. [\[CrossRef\]](#)
7. Wang, J.; Yang, Y.; Mao, J.; Huang, Z.; Huang, C.; Xu, W. CNN-RNN: A Unified Framework for Multi-label Image Classification. *arXiv* **2016**, arXiv:1604.04573.
8. Thoma, M. A Survey of Semantic Segmentation. *arXiv* **2016**, arXiv:1602.06541.
9. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *arXiv* **2015**, arXiv:1411.4038.
10. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *arXiv* **2017**, arXiv:1606.00915. [\[CrossRef\]](#)
11. Liu, Y.; Gross, L.; Li, Z.; Li, X.; Fan, X.; Qi, W. Automatic Building Extraction on High-Resolution Remote Sensing Imagery Using Deep Convolutional Encoder-Decoder With Spatial Pyramid Pooling. *IEEE Access* **2019**, *7*, 128774–128786. [\[CrossRef\]](#)
12. Howard, A.; Sandler, M.; Chen, B.; Wang, W.; Chen, L.-C.; Tan, M.; Chu, G.; Vasudevan, V.; Zhu, Y.; Pang, R.; et al. Searching for MobileNetV3. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1314–1324.
13. Zhou, L.; Zhang, C.; Wu, M. D-LinkNet: LinkNet with Pretrained Encoder and Dilated Convolution for High Resolution Satellite Imagery Road Extraction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Salt Lake City, UT, USA, 18–23 June 2018; pp. 192–1924. [\[CrossRef\]](#)
14. Kampffmeyer, M.; Salberg, A.-B.; Jenssen, R. Semantic Segmentation of Small Objects and Modeling of Uncertainty in Urban Remote Sensing Images Using Deep Convolutional Neural Networks. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 680–688. [\[CrossRef\]](#)
15. Yuan, X.; Shi, J.; Gu, L. A review of deep learning methods for semantic segmentation of remote sensing imagery. *Expert Syst. Appl.* **2021**, *169*, 114417. [\[CrossRef\]](#)
16. Salehi, S.S.M.; Erdogmus, D.; Gholipour, A. Tversky loss function for image segmentation using 3D fully convolutional deep networks. *arXiv* **2017**, arXiv:1706.05721.
17. Diakogiannis, F.I.; Waldner, F.; Caccetta, P.; Wu, C. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS J. Photogramm. Remote Sens.* **2020**, *162*, 94–114. [\[CrossRef\]](#)
18. Peng, D.; Zhang, Y.; Guan, H. End-to-end change detection for high resolution satellite images using improved UNet++. *Remote Sens.* **2019**, *11*, 1382. [\[CrossRef\]](#)
19. Yang, X.; Li, X.; Ye, Y.; Lau, R.Y.K.; Zhang, X.; Huang, X. Road Detection and Centerline Extraction Via Deep Recurrent Convolutional Neural Network U-Net. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 7209–7220. [\[CrossRef\]](#)
20. Giorgiani do Nascimento, R.; Viana, F. Satellite Image Classification and Segmentation with Transfer Learning. In *AIAA Scitech 2020 Forum*; AIAA SciTech Forum; American Institute of Aeronautics and Astronautics: Reston, VI, USA, 2020.
21. Lee, M.-G.; Cho, H.-B.; Youm, S.-K.; Kim, S.-W. Detection of Pine Wilt Disease Using Time Series UAV Imagery and Deep Learning Semantic Segmentation. *Forests* **2023**, *14*, 1576. [\[CrossRef\]](#)
22. Shi, T.; Guo, Z.; Li, C.; Lan, X.; Gao, X.; Yan, X. Improvement of deep learning Method for water body segmentation of remote sensing images based on attention modules. *Earth Sci. Inform.* **2023**, *16*, 2865–2876. [\[CrossRef\]](#)
23. Yu, Z.; Wan, F.; Lei, G.; Xiong, Y.; Xu, L.; Ye, Z.; Liu, W.; Zhou, W.; Xu, C. RSLC-Deeplab: A Ground Object Classification Method for High-Resolution Remote Sensing Images. *Electron. Switz.* **2023**, *12*, 3653. [\[CrossRef\]](#)
24. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv* **2015**, arXiv:1505.04597.
25. Tan, C.; Sun, F.; Kong, T.; Zhang, W.; Yang, C.; Liu, C. A Survey on Deep Transfer Learning. In *Artificial Neural Networks and Machine Learning—ICANN 2018*; Kůrková, V., Manolopoulos, Y., Hammer, B., Iliadis, L., Maglogiannis, I., Eds.; In Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2018; pp. 270–279. [\[CrossRef\]](#)
26. Wurm, M.; Stark, T.; Zhu, X.X.; Weigand, M.; Taubenböck, H. Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* **2019**, *150*, 59–69. [\[CrossRef\]](#)
27. Chen, T.; Cai, Z.; Zhao, X.; Chen, C.; Liang, X.; Zou, T.; Wang, P. Pavement crack detection and recognition using the architecture of segNet. *J. Ind. Inf. Integr.* **2020**, *18*, 100144. [\[CrossRef\]](#)
28. Bosilj, P.; Aptoula, E.; Duckett, T.; Cielniak, G. Transfer learning between crop types for semantic segmentation of crops versus weeds in precision agriculture. *J. Field Robot.* **2020**, *37*, 7–19. [\[CrossRef\]](#)
29. Garcia-Garcia, A.; Orts-Escobedo, S.; Oprea, S.; Villena-Martinez, V.; Martinez-Gonzalez, P.; Garcia-Rodriguez, J. A survey on deep learning techniques for image and video semantic segmentation. *Appl. Soft Comput.* **2018**, *70*, 41–65. [\[CrossRef\]](#)
30. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. *arXiv* **2018**, arXiv:1708.02002.
31. Yang, H.L.; Yuan, J.; Lunga, D.; Laverdiere, M.; Rose, A.; Bhaduri, B. Building Extraction at Scale Using Convolutional Neural Network: Mapping of the United States. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 2600–2614. [\[CrossRef\]](#)
32. Zhu, Q.; Liao, C.; Hu, H.; Mei, X.; Li, H. MAP-Net: Multiple Attending Path Neural Network for Building Footprint Extraction from Remote Sensed Imagery. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 6169–6181. [\[CrossRef\]](#)

33. Vemula, S.; Frye, M. Multi-head Attention Based Transformers for Vegetation Encroachment Over Powerline Corridors using UAV. In Proceedings of the 2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC), San Antonio, TX, USA, 3–7 October 2021; pp. 1–5. [[CrossRef](#)]
34. Damodaran, S.; Shanmugam, L.; Swaroopan, N.M.J. Extraction of Overhead Transmission Towers from UAV Images. In Proceedings of the 2023 5th International Conference on Electrical, Computer and Communication Technologies, ICECCT 2023, Erode, India, 22–24 February 2023. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.