applied sciences

MDPI

Article

Deep-Learning-Based Land Cover Mapping in Franciacorta Wine Growing Area

Girma Tariku [1,*], Isabella Ghiglieno [2,*], Andres Sanchez Morchio [2], Luca Facciano [2], Celine Birolleau [2], Anna Simonetto [2], Ivan Serina [1] and Gianni Gilioli [2]

[1] Department of Information Engineering (DII), University of Brescia, 38 Via Branze, 25123 Brescia, Italy
[2] Agrofood Research Hub, Department of Civil, Environmental, Architectural Engineering, and Mathematics, University of Brescia, 43 Via Branze, 25123 Brescia, Italy
* Correspondence: g.tariku@unibs.it (G.T.); isabella.ghiglieno@unibs.it (I.G.)

Abstract: Land cover mapping is essential to understanding global land-use patterns and studying biodiversity composition and the functioning of eco-systems. The introduction of remote sensing technologies and artificial intelligence models made it possible to base land cover mapping on satellite imagery in order to monitor changes, assess ecosystem health, support conservation efforts, and reduce monitoring time. However, significant challenges remain in managing large, complex satellite imagery datasets, acquiring specialized datasets due to high costs and labor intensity, including a lack of comparative studies for the selection of optimal deep learning models. No less important is the scarcity of aerial datasets specifically tailored for agricultural areas. This study addresses these gaps by presenting a methodology for semantic segmentation of land covers in agricultural areas using satellite images and deep learning models with pre-trained backbones. We introduce an efficient methodology for preparing semantic segmentation datasets and contribute the "Land Cover Aerial Imagery" (LICAI) dataset for semantic segmentation. The study focuses on the Franciacorta area, Lombardy Region, leveraging the rich diversity of the dataset to effectively train and evaluate the models. We conducted a comparative study, using cutting-edge deep-learning-based segmentation models (U-Net, SegNet, DeepLabV3) with various pre-trained backbones (ResNet, Inception, DenseNet, EfficientNet) on our dataset acquired from Google Earth Pro. Through meticulous data acquisition, preprocessing, model selection, and evaluation, we demonstrate the effectiveness of these techniques in accurately identifying land cover classes. Integrating pre-trained feature extraction networks significantly improves performance across various metrics. Additionally, addressing challenges such as data availability, computational resources, and model interpretability is essential for advancing the field of remote sensing, in support of biodiversity conservation and the provision of ecosystem services and sustainable agriculture.

Keywords: land cover mapping; semantic segmentation; deep learning; satellite imagery; pre-trained backbone

Academic Editor: Borja Velazquez -Marti

Received: 17 December 2024
Revised: 7 January 2025
Accepted: 15 January 2025
Published: 17 January 2025

Citation: Tariku, G.; Ghiglieno, I.; Sanchez Morchio, A.; Facciano, L.; Birolleau, C.; Simonetto, A.; Serina, I.; Gilioli, G. Deep-Learning-Based Land Cover Mapping in Franciacorta Wine Growing Area. *Appl. Sci.* **2025**, *15*, 871. https://doi.org/10.3390/app15020871

1. Introduction

1.1. Background

Land cover mapping is a fundamental tool for understanding complex land-use patterns across the globe. It provides key information for such critical issues as eco-system health and biodiversity conservation, and it monitors changes over time. By accurately classifying various types of land covers, like forests, agricultural lands, wetlands, and

urban areas, land cover maps offer valuable insights. Advancements in remote sensing technology have increased the availability of satellite imagery with diverse characteristics, making it a cornerstone in land cover mapping [1].

Vineyards contribute significantly to biodiversity by supporting habitats that provide critical ecosystem services [2]. Grasslands, herbaceous zones, and forested areas surrounding vineyards act as ecological corridors, promoting species richness and fostering pollinators and natural predators essential for pest regulation and crop production [3,4]. These habitats also improve soil health, reduce erosion, and support water retention, thereby enhancing overall ecosystem resilience [5]. By maintaining a diverse landscape, vineyards can help preserve local flora and fauna, thereby supporting ecological stability and productivity [6]. Mapping surrounding land cover types is indispensable for biodiversity conservation in vineyard ecosystems. High-resolution maps generated through advanced technologies, such as satellite imagery and UAVs, allow the precise identification and monitoring of diverse habitats. These maps facilitate the understanding of spatial patterns, enabling informed decisions to optimize agricultural practices while conserving biodiversity [7]. For instance, detailed land cover maps guide the management of cover crops, monitor pest-prone zones, and assess the ecological benefits of natural habitats within and around vineyards [8].

Traditionally, classical machine learning techniques, such as SVM [9], decision trees [10], and random forest [11] have been used for land cover mapping from satellite images. However, these methods have limitations, such as the need for expert analysis [12] and challenges with high-resolution imagery [13]. An example is maximum likelihood classification (MLC), which is based on pixel probabilities but struggles with resolution limitations and expert input [14].

1.2. Problem Statement

The emphasis on using high-resolution satellite imagery for land cover mapping highlights the limitations of traditional classification methods [8]. While studies showcase the effectiveness of random forests and support vector machines with high-resolution data, such challenges as sensitivity to parameters and class imbalance persist, which requires careful optimization [15,16]. Land cover data is crucial in urban planning to address global challenges, but limitations in satellite image analysis, such as managing large data volumes and complex classifications, remain [17]. To address these challenges using very high-resolution (VHR) imagery, integrated approaches combining object-based image analysis (OBIA) and random forests have been proposed [18]. However, even these approaches face issues like spectral redundancy and the computational demands of incorporating temporal features [18].

1.3. Recent Advancements and Challenges in Deep Learning

Recent technological advancements have driven the evolution of image analysis, particularly with the introduction of deep learning algorithms [19]. Deep learning offers the advantage of automatic feature extraction, exemplified by the TASSEL framework, which combines CNN-based feature extraction with OBIA classification [20]. Recent advancements in deep learning, particularly convolutional neural networks (CNNs), hold significant potential for land cover classification and object detection using high-resolution imagery [21]. However, challenges remain in enhancing the generalization and robustness of these models, necessitating the use of diverse training datasets [22]. Additionally, studies have presented techniques using pre-trained neural networks for land cover mapping on Landsat 5/7 images, potentially accelerating map production but limiting generalizability [23].

In land cover mapping with deep learning, semantic segmentation excels by dividing images into meaningful segments, offering superior accuracy, especially in vegetation mapping [24–27]. To address the limitations of single-scale convolution kernels, advancements like multi-scale fully convolutional networks (MSFCNs) have been introduced [28]. Similarly, the U-Net temporal attention encoder (U-TAE) and the remote sensing segmentation transformer (RSSFormer) approaches enhance semantic segmentation with temporal attention or are specifically tailored for high-resolution remote sensing applications, respectively [29,30]. While CNNs are explored for per-pixel classification in high-resolution remote sensing, leveraging detailed data for precise object classification [31], challenges persist in acquiring specialized datasets due to high costs associated with data collection, demanding significant time, labor, and financial resources for activities such as field surveys and manual annotation of vegetation classes [29–31].

1.4. Study Objective and Research Gap

While significant progress has been made in developing deep-learning-based methods for land cover classification [24–27], a notable gap remains in the availability of remote sensing datasets tailored to agricultural regions [32–34]. To address this gap, this study presents a cost-effective methodology for preparing semantic segmentation datasets, focusing on reducing the time and costs traditionally required for data collection. This methodology enhances reproducibility and provides an effective approach to generating datasets for land cover mapping in vineyard ecosystems.

Our study focuses on mapping different land cover classes, including grasslands, herb-dominated habitats, and tree-dominated areas in the Franciacorta region of Lombardy, Italy. By using remote sensing and artificial intelligence, we introduce a methodology for mapping large areas, laying the groundwork for future studies on how surrounding land cover can contribute to the ecological balance and sustainability of vineyard ecosystems.

While significant progress has been made in developing land cover classification using semantic segmentation, a notable gap persists in creating comprehensive remote sensing datasets tailored to agricultural regions. To bridge this gap, this study presents a cost-effective methodology for preparing a semantic segmentation dataset to ensure reproducibility for other researchers. We introduce a novel land cover dataset comprising seven manually annotated classes, specifically designed for large-scale semantic segmentation, with a particular focus on the wine cultivation area. These essential classes include grasslands, arable land, herb-dominated areas, hedgerows, vineyards, tree-dominated man-made habitats, and olive groves.

The dataset was created through the analysis of a large agricultural area, specifically the Franciacorta region of Lombardy, Italy. Landsat satellite RGB images with a 30 m resolution, obtained using Google Earth Pro 7.3.6 software, served as the image source for in-depth analysis.

In this paper, we conducted a rigorous comparative analysis of state-of-the-art deep-learning-based segmentation models to determine the optimal approach. We employed a variety of models, including UNet, SegNet, and DeepLabV3, combined with different backbone convolutional neural networks.

2. Materials and Methods

The research methodology consisted of four sequential steps, as illustrated in Figure 1. Firstly, the study focused on the Franciacorta area in Lombardy, Northern Italy, defining the geographical context. Secondly, satellite imagery was acquired from Google Earth Pro, georeferenced with ArcGIS, and segmented using eCognition to create homogeneous image regions. In the third step, these segmented images were validated and classified

by a plant expert using QGISto ensure accurate identification of land cover types such as grasslands and vineyards. Lastly, the fourth step involved training semantic segmentation models U-Net, SegNet, and DeepLabV3—on smaller image patches derived from the large original images, with and without pre-trained backbones. Models were assessed based on performance metrics such as accuracy and the Jaccard Index, thereby completing a thorough approach to land cover analysis and classification.

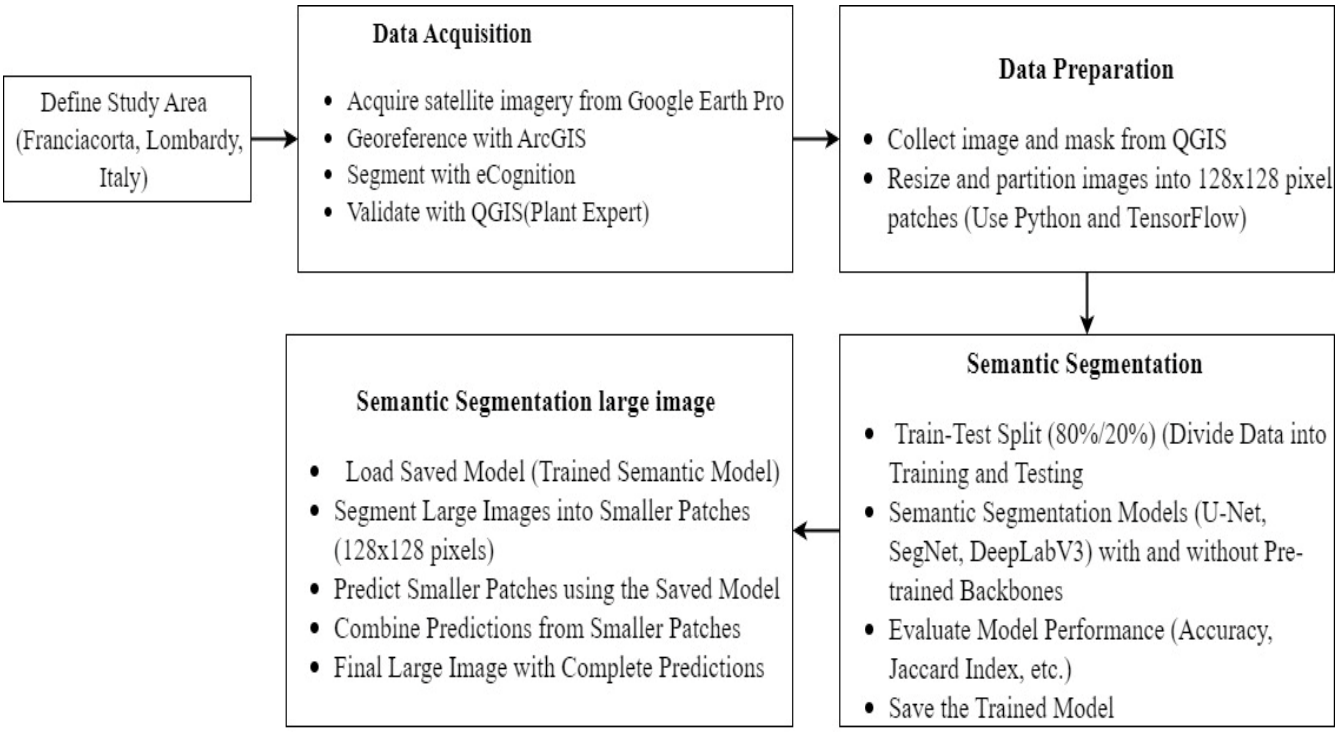

Figure 1. General workflow of semantic segmentation.

2.1. Study Site

This study focuses on the Franciacorta area, a famous Italian wine growing region located in Lombardy, Northern Italy (as shown in Figure 2). Nestled in the picturesque province of Brescia, Franciacorta is renowned for its exquisite landscape, rich history, and world-class wine production. This area encompasses approximately 195 square kilometers (77 square miles), and the study area is located at approximately 45°37′31.05″ N and 9°55′52.33″ E.

2.2. Data Acquisition

The total study area covers 2981.4 hectares. We divided this area into 18 image tiles. Among these tiles, the smallest covers 14 hectares with a 1192 × 700-pixel picture, and the largest 172 hectares with a 2392 × 1402-pixel picture. The following steps were taken to prepare the picture and the corresponding mask image:

Step 1: The process initiated with the acquisition of satellite imagery from Google Earth Pro by capturing both the visual image and its accompanying shape file in .kmz format as shown in Figure 3. This imagery encompasses a broad spatial scope and provides a comprehensive view of the target area. Subsequently, the acquired images were georeferenced using ArcGIS 10.3 software, a crucial step in aligning the satellite image with geographic coordinates for accurate spatial analysis.

Figure 2. Geographic coverage of the study area in Franciacorta, Lombardy, Northern Italy. Here Italy is outlined in red, Lombardy in yellow, and the study area in blue.

Figure 3. Satellite imagery acquisition and georeferencing process.

Step 2: After georeferencing, multiresolution segmentation (MRS) in eCognition partitioned the imagery into homogeneous regions based on spectral and spatial characteristics. The scale parameter was set to 100, compactness to 0.5, and shape to 0.4; these values optimized the balance between spatial coherence and object delineation. A georeferenced shapefile was then generated. This facilitated expert masking in QGIS to remove unnecessary fragmented objects.

Step 3: Following segmentation, the resulting segmented shape file and image were subject to meticulous examination and validation by a knowledgeable plant expert as shown

in Figure 4. Leveraging the capabilities of QGIS software 3.24.0, the expert meticulously inspected each segmented polygon and assigned appropriate plant names based on their botanical characteristics and distribution within the landscape.

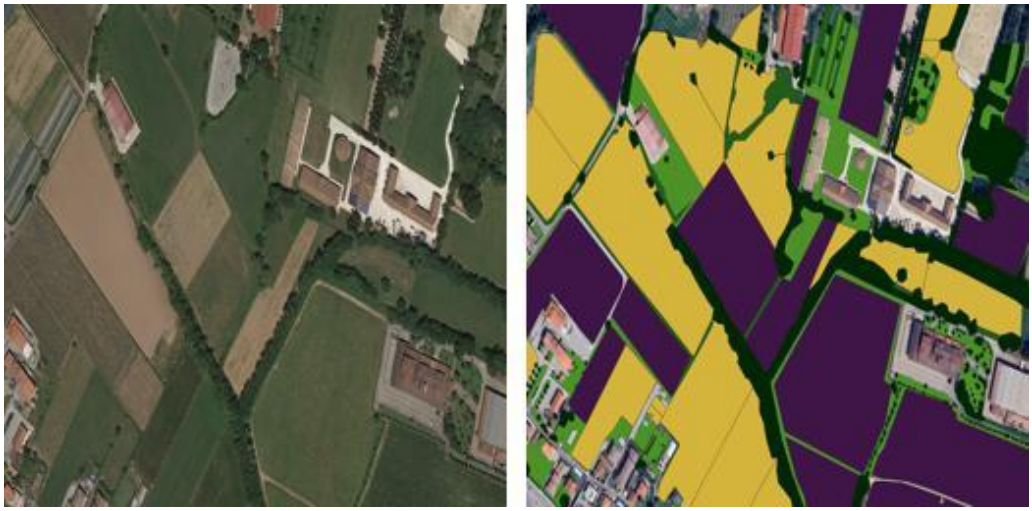

Figure 4. Validation process of segmented shape file in QGIS: original image (**left**); segmented classes (**right**). Grasslands are shown in green, arable lands in yellow, herb-dominated habitats in medium-dark green, vineyards in purple, tree-dominated man-made habitats in black. Additionally, *Olea europaea* groves are shown in brown.

Classes:

1. Grassland: grasslands are crucial for maintaining biodiversity in vineyard ecosystems in that they support various pollinators, herbivores, and predators [3]. These areas provide habitat for beneficial insects, such as bees and butterflies, which aid in pollination, a key service for vineyards. Additionally, grasslands can foster natural pest control by harboring insectivorous species that help keep pest populations in check [2].
2. Arboreal land: tree-dominated habitats, such as forests or woodlands, are vital for fostering biodiversity as they offer shelter and food resources for a variety of species, including birds, mammals, and insects. These areas also play a critical role in climate regulation and contribute to ecosystem connectivity. In vineyard landscapes, arboreal land acts as a buffer zone that protects against wind erosion, provides shade, and enhances the resilience of the ecosystem as a whole.
3. Herbaceous habitats: herbs and grasses are important to foster biodiversity in that they support herbivores, pollinators, and decomposers. These areas typically host plant species that help improve soil health and nutrient cycling, and hence the sustainability of surrounding agricultural areas. In vineyards, herb-dominated habitats can play a role in reducing soil erosion, supporting beneficial micro fauna, and providing a diverse range of species that contribute to the overall ecosystem.
4. Hedgerows: these are critical for maintaining ecosystem connectivity in agricultural landscapes. These linear strips of vegetation provide a corridor for wildlife movement, as they offer shelter, food, and protection for pollinators, birds, and small mammals. In vineyards, hedgerows can also act as natural pest control systems by supporting predatory species that reduce the abundance of vineyard pests, such as aphids or grapevine moths.
5. Vineyards: as the primary agricultural feature in this study, vineyards play a dual role in biodiversity. While they are managed for grape production, vineyards can still contribute to biodiversity by supporting specialized plants and animals, such

as birds and insects that thrive in cultivated areas. Effective vineyard management practices, such as cover crops and organic farming methods, can enhance biodiversity by providing pollinator habitats, improving soil health, and promoting sustainable farming practices.

6. Tree-dominated man-made habitats: these areas, including urban parks, orchards, and landscaped gardens, offer refuges for a range of fauna and flora, thereby promoting biodiversity within human-modified environments. While these habitats may not be as diverse as natural forests, they support species that are adapted to urbanized landscapes. In relation to vineyards, tree-dominated habitats can provide important wildlife corridors that help maintain genetic diversity among species and support beneficial pollinators and pest controllers.
7. *Olea europaea* groves: olive groves, primarily found in Mediterranean regions, offer a specialized habitat for certain plant and animal species. These groves support a range of native vegetation and provide food and shelter for various wildlife species, including birds, insects, and small mammals. In vineyards, the presence of olive groves can enhance biodiversity by fostering species interactions across adjacent agricultural landscapes, thereby contributing to sustainability and ecosystem services like pollination and pest control.

3. Experiment

In this work, the original images were large in pixel size. This posed a challenge for deep learning models, in that large image sizes require significant memory and can slow down the training process. This challenge was addressed by setting up a Python code to efficiently process large images. The code started by iterating through the directories and subdirectories within a specified root directory, which contained both image and mask files. For each image and mask file found, the code resized them to the nearest size divisible by a predefined 128 $\times$ 128-pixel patch size. The resized images and masks were then divided into non-overlapping patches using the patchify library from the TensorFlow function, making sure each patch matches the specified dimensions. These patches were saved as individual image files, preserving their association with their corresponding masks as shown in Figure 5. Additionally, the code utilized the split folders library from TensorFlow to split the patched images and masks into training, validation, and testing datasets at 70%, 15%, and 15% ratios, respectively. After splitting, the training dataset was comprised of 536 images, the validation dataset contained 115 images, and the testing dataset included 115 images. This streamlined process facilitated efficient model training, validation, and evaluation.

3.1. Backbones

In our study, we performed two semantic segmentation approaches: with and without leveraging a backbone. Semantic segmentation without a backbone involves training models from scratch, without the use of pre-trained networks. This method relies solely on the model's architecture to extract features directly from input data. While this approach offers independence from pre-existing biases and data patterns embedded in pre-trained networks, it typically requires more extensive training and larger datasets to achieve performance comparable with pre-trained backbone models.

Conversely, leveraging a pre-trained backbone is a widely adopted strategy aimed at enhancing model performance in semantic segmentation tasks. Backbones, such as ResNet, InceptionV2, DenseNet, and EfficientNet, are well-established architectures in CNN deep-learning. These models are pre-trained on large-scale datasets (such as ImageNet) and have

learned to extract meaningful features from images by encompassing spatial relationships and contextual information crucial for accurate segmentation.

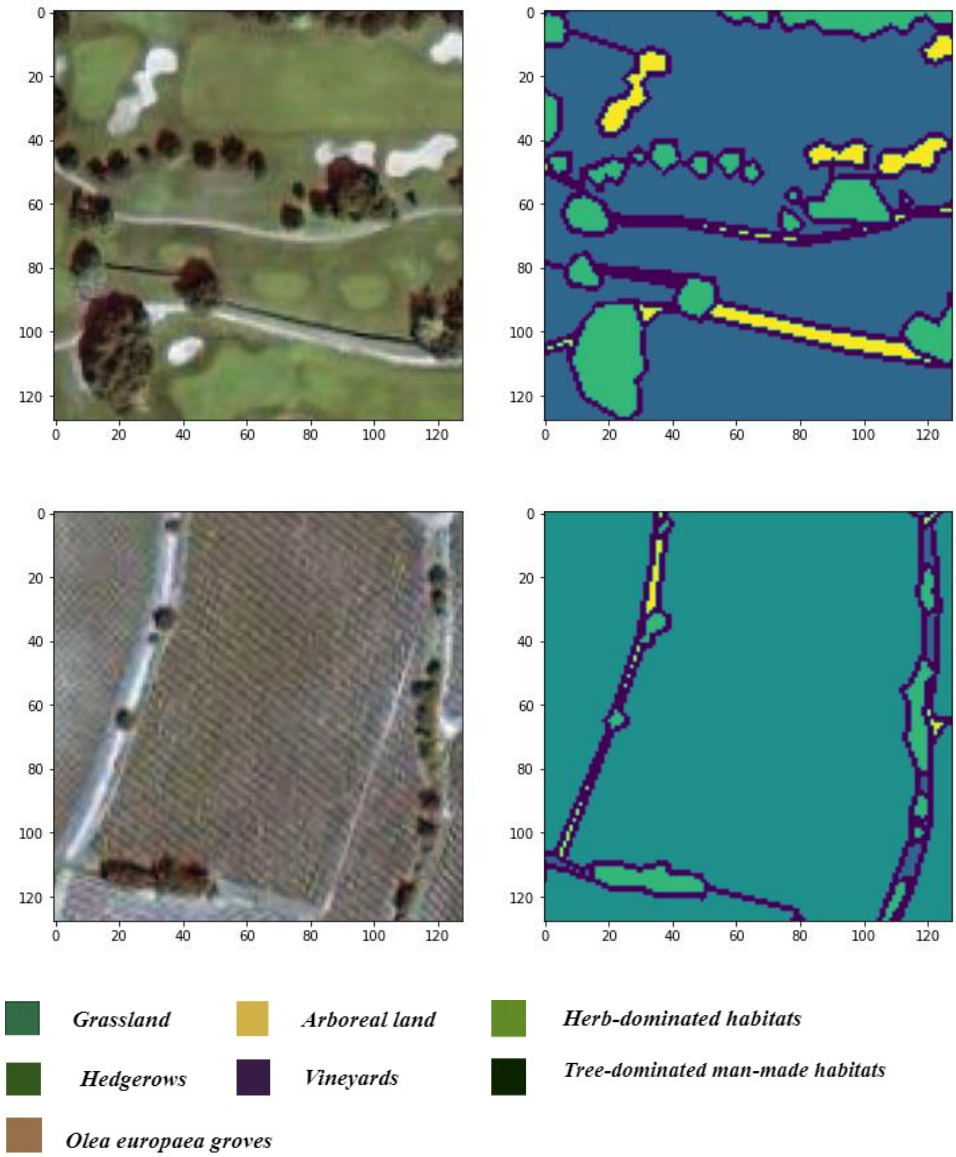

Figure 5. Illustration of two randomly selected 128 × 128-pixel image patches, extracted from the original images along with their corresponding ground truth masks. These patches are used to train and evaluate a deep learning model.

By fine-tuning these pre-trained backbones for specific segmentation tasks, models can efficiently adapt to new datasets and extract relevant features tailored to the nuances in land cover classification. This process leverages transfer learning, where knowledge acquired from previous tasks (e.g., image classification) is transferred to improve performance on current segmentation tasks. This approach not only accelerates model convergence during training but also enhances segmentation accuracy by leveraging the rich feature representations learned by backbone networks.

The choice of ResNet, InceptionV2, DenseNet, and EfficientNet as backbone architectures was motivated by their proven performance in various computer vision tasks and their distinct architectural strengths:

A. ResNet

ResNet [35], or Residual Network, is a pioneering convolutional neural network architecture devised to tackle the challenge of vanishing gradients in deep networks. Developed by Microsoft Research, ResNet introduces residual connections enabling the training of exceptionally deep networks by learning residual functions. These connections address the degradation problem associated with increasing network depth, resulting in state-of-the-art performance across various computer vision tasks.

B. Inception

An evolution of Google's Inception architecture, InceptionV3 [36] is renowned for its computational efficiency and high accuracy. It introduces several novel features, including the inception model, which enables the network to efficiently capture features at multiple scales through parallel convolutions. InceptionV3 employs factorized convolutions, which reduces computational cost while maintaining expressive power. Additionally, it incorporates batch normalization and auxiliary classifiers, thus facilitating convergence speed and regularization.

C. DensNet

DenseNet [37], short for "densely connected convolutional networks", is a convolutional neural network architecture known for its densely connected layers. Unlike traditional architecture where each layer is connected only to the subsequent layers, DenseNet introduces dense connections, where each layer receives direct inputs from all preceding layers. This design fosters feature reuse and facilitates the flow of gradients, leading to improved parameter efficiency and better gradient propagation.

D. EfficientNet

EfficientNet is a family of convolutional neural network architectures developed by [38]. It is specifically designed to achieve state-of-the-art accuracy while simultaneously being highly efficient in terms of computational resources. The key innovation behind EfficientNet is the compound scaling method, which uniformly scales network depth, width, and resolution in a systematic manner. This approach ensures that the network's parameters and computational cost are optimized for a given resource constraint, resulting in models that are both accurate and efficient across a wide range of tasks. In our study, we specifically utilized EfficientNetB0, which is the smallest variant in the EfficientNet family. EfficientNetB0 strikes a balance between model complexity and performance, making it suitable for various computer vision tasks, including image classification, object detection, and semantic segmentation.

3.2. Semantic Segmentation Models

We selected UNet [39], SegNet [40], and DeepLabV3+ [41] for our land cover segmentation due to their proven performance and distinct architectural strengths. By incorporating skip connections, UNet's encoder–decoder structure effectively captures both local and global features, which are essential for precise boundary delineation [39]. SegNet strikes a balance between memory efficiency and high-resolution feature preservation, making it suitable for handling complex satellite imagery [40]. DeepLabV3+, which utilizes atrous convolution and atrous spatial pyramid pooling (ASPP) modules, excels in capturing multi-scale contextual information, which is crucial for intricate segmentation tasks [41].

A. UNet

The UNet [39] semantic segmentation model is a deep learning architecture designed for pixel-wise classification tasks, particularly in image segmentation. In this sample, we

utilized the UNet model to perform semantic segmentation on UAV RGB images. The model employs a symmetrical encoder–decoder structure, with skip connections between corresponding encoder and decoder layers to preserve spatial information. This aids in capturing both local and global features, making it particularly effective for such tasks as object detection and boundary delineation.

Our UNet semantic segmentation model (Figure 6) is designed to process 128 × 128-pixel RGB images with a structured architecture aimed at capturing intricate spatial details. The model architecture is structured into two main sections: a contraction path (c1 to c5) and an expansive path (u6 to u9).

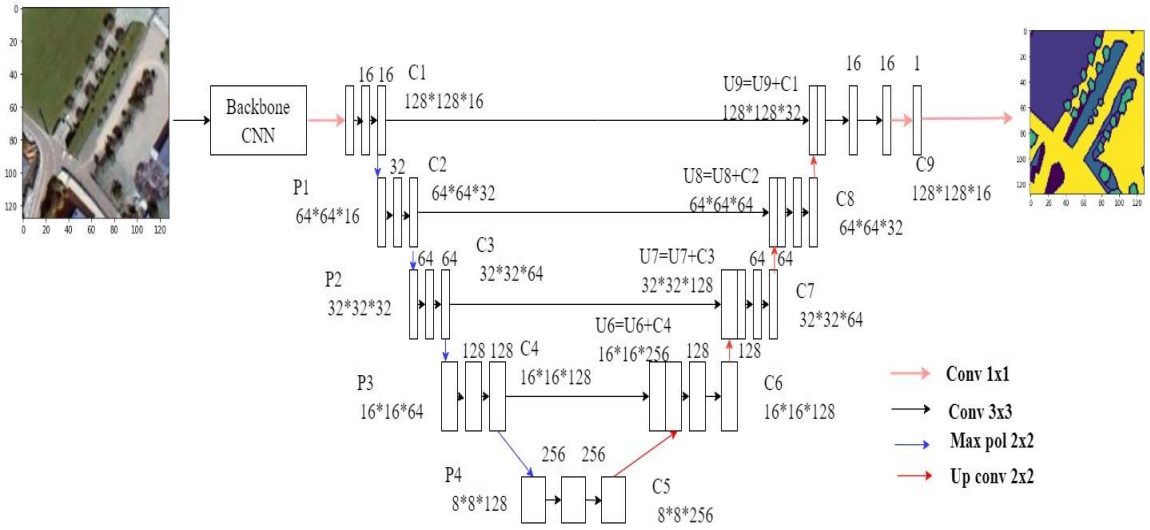

Figure 6. Architecture of the UNet semantic segmentation backbone model.

In the contraction path, successive convolutional blocks progressively increase the number of filters to extract and encode features from the input image. Max-pooling layers (p1 to p4) are strategically placed to down sample the spatial dimensions, thereby reducing computational load while preserving essential features. At the bottleneck layer (c5), the model captures high-level context and semantic information, which is crucial for accurate segmentation.

Conversely, the expansive path of the UNet model focuses on upsampling the feature maps using transposed convolutional layers. These layers help recover spatial resolution lost during the contraction path and enable the model to reconstruct detailed information. Each block in the expansive path (c6 to c9) sequentially reduces the number of filters, allowing the model to refine segmentation boundaries and enhance localization accuracy.

The final layer of the UNet model (outputs) employs a 1 × 1 convolutional layer with a sigmoid activation function, facilitating pixel-wise predictions for each class in the segmentation task. This setup enables the model to output probability maps where each pixel indicates the likelihood of belonging to a specific land cover class.

During training, the model is compiled using the Adam optimizer, which efficiently adjusts learning rates for each parameter during training. The choice of binary cross-entropy loss function aligns with the pixel-wise nature of the segmentation task, thereby optimizing the model to produce accurate binary predictions for each pixel. The accuracy metric provides insights into the model's performance by measuring the proportion of correctly classified pixels compared to the total number of pixels.

B. SegNet

SegNet [40] is a convolutional neural network (CNN) architecture specifically tailored for semantic segmentation tasks, where the goal is to assign a class label to each pixel in

an image. It features an encoder–decoder structure, which comprises an encoding path to extract hierarchical features from the input image and a decoding path to generate a pixel-wise segmentation map. Our SegNet semantic segmentation model, depicted in Figure 7, is optimized with the EfficientNetB0 architecture serving as its foundational backbone, with connected layers fully omitted to streamline feature extraction from input images. The model is structured around an encoder–decoder architecture, each fulfilling distinct roles in the segmentation process.

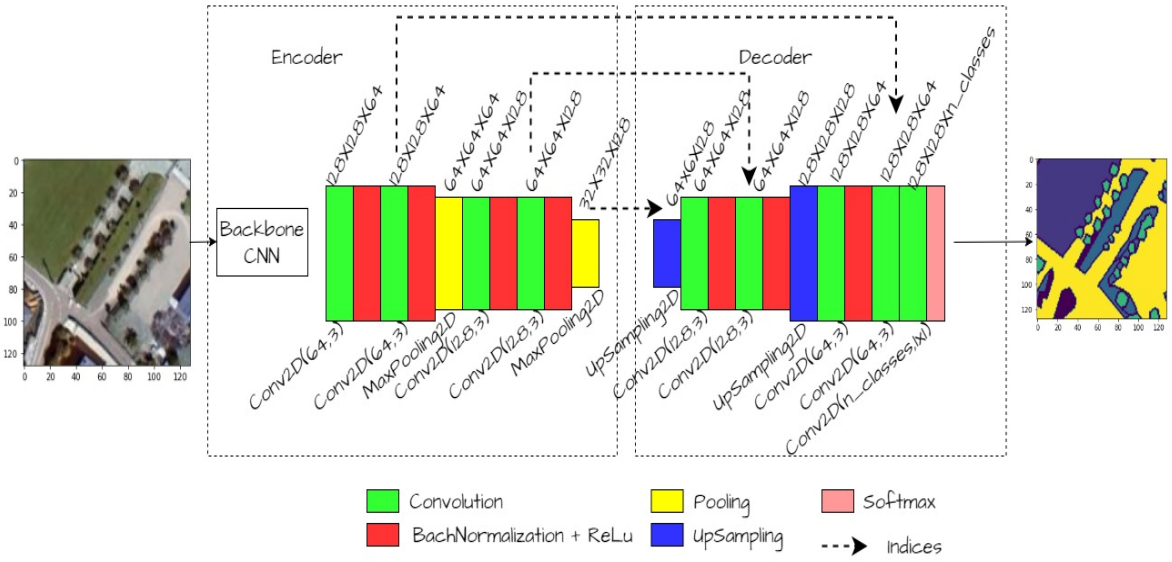

Figure 7. SegNet block diagram with backbone encoder–decoder CNN architecture.

The encoder, or contracting path, initiates with convolutional layers followed by batch normalization and max-pooling operations. This sequence progressively condenses spatial dimensions while effectively capturing essential features from the input images. By utilizing max-pooling layers, the encoder enhances computational efficiency by reducing the complexity of feature maps without compromising significant visual information.

In contrast, the decoder, or expansive path, employs upsampling layers to restore the spatial resolution of feature maps. These layers are crucial for recovering fine-grained details lost during the downsampling process in the encoder. Furthermore, the decoder integrates features extracted from the EfficientNetB0 backbone with the up-sampled feature maps to refine segmentation accuracy. This fusion of hierarchical features ensures that the model can leverage both high-level semantic information and detailed spatial context for precise pixel-wise predictions. Subsequent convolutional layers further refined the features, and the output layer generates pixel-wise predictions through a softmax activation function. The model constructed with the backbone input and output layer is designed to classify pixels in images.

C. DeepLab

DeepLabV3 [41] is a cutting-edge convolutional neural network architecture tailored for semantic image segmentation. It leverages atrous convolution to capture multi-scale contextual information efficiently. Key features include a pyramid network for hierarchical feature extraction, atrous spatial pyramid pooling for multi-scale feature aggregation, and efficient up-sampling methods for high-resolution segmentation maps. Our DeepLabV3+ model (Figure 8) utilized an EfficientNetB0 backbone for feature extraction without fully connected layers, serving as the encoder. This encoder processes input images and extracts features, which are then passed on to the decoder. The decoder consists of an atrous spatial pyramid pooling (ASPP) module followed by global average pooling and low-level feature

concatenation. The ASPP module employs convolutional layers with varying dilation rates to capture multi-scale contextual information. Additionally, global average pooling is performed to capture global context information. The decoder then combines ASPP outputs, global context, and low-level features using concatenation. Further convolutional layers refine the features, followed by up-sampling layers to restore spatial information. Finally, a 1×1 convolutional layer with SoftMax activation generates pixel-wise predictions for semantic segmentation. The model is compiled using the Adam optimizer and categorical cross-entropy loss and accuracy function for metrics.

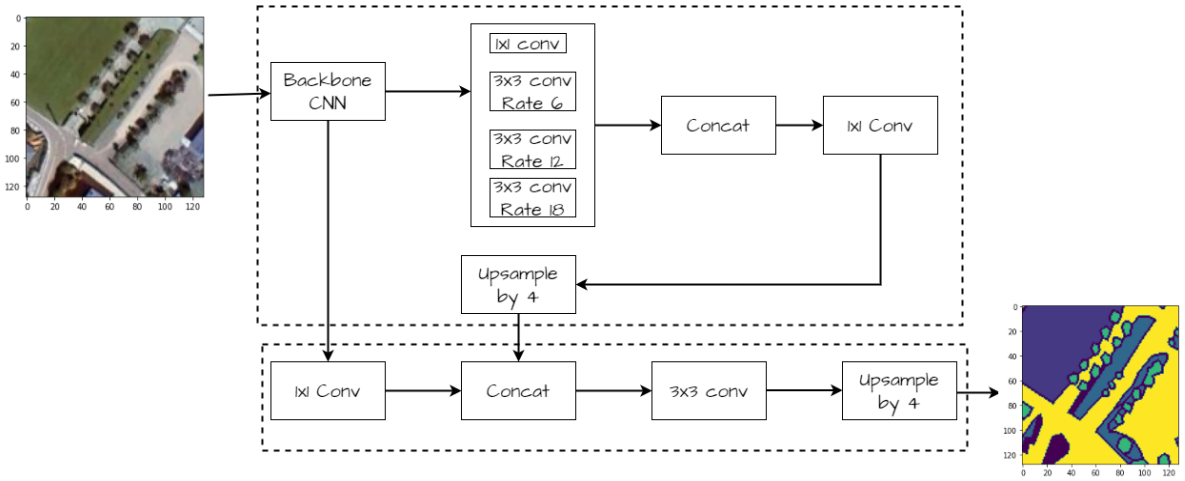

Figure 8. DeeplabV3 block diagram with backbone encoder–decoder CNN architecture.

We implemented U-Net, SegNet, and DeepLabV3 segmentation models for the semantic segmentation of satellite images. Initially, the image and mask datasets underwent preprocessing by subdividing them into smaller patches to facilitate training. Each patch was then scaled using min–max scaling, and the RGB masks were converted into integer labels based on predefined RGB values. These models were constructed using the EfficientNetB0 backbone, with and without encoder and decoder paths. A SoftMax activation function was applied to the output layer for multi-class segmentation. Following compilation with the Adam optimizer and categorical cross-entropy loss function, the model was trained for 100 epochs with batch size 16. The evaluation was made using accuracy and IoU metrics on both training and validation datasets, and the trained model was then saved for future reference. Performance metrics, including accuracy, precision, recall, F1 score, and Jaccard coefficient, were computed, and the segmentation performance was visually demonstrated on randomly selected test images.

3.3. Performance Metrics

In this study, several metrics were used to evaluate the performance of the semantic segmentation models:

- Accuracy—It measures the overall correctness of the segmentation by calculating the ratio of correctly predicted pixels to the total number of pixels. It is computed as

$$\text{Accuracy} = \frac{\text{True positives} + \text{True Negatives}}{\text{Total Pixels}}$$

- Precision—It quantifies the model's ability to correctly identify positive predictions among all predicted positives. It is computed as

$$\text{Precision} = \frac{\text{True positive}}{\text{True Positive} + \text{False Positive}}$$

- Recall (Sensitivity)—It measures the model's ability to detect all relevant instances of a class in the image. It is computed as

$$\text{Recall} = \frac{\text{True positive}}{\text{True Positive} + \text{False Negative}}$$

- F1 Score—It provides a balanced measure of precision and recall. It is calculated as the harmonic means of precision and recall:

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

- Jaccard Index (IoU)—Also referred to as intersection over union (IoU), it measures the overlap between the predicted and ground truth regions:

$$\text{IoU} = \frac{\text{Intersection of Predicted and Ground Truth Regions}}{\text{Union of Predicted and Ground Truth Regions}}$$

- Mean IoU (mIoU)—It calculates the average IoU across all classes and provides an overall measure of segmentation accuracy across different classes.

These metrics were computed using the trained segmentation models to evaluate their performance on the test dataset. Additionally, class weighting based on Dice scores was applied during training to mitigate the impact of class imbalance.

4. Result

Our semantic segmentation models (Unet, SegNet, and DeepLabV3) utilized TensorFlow and the segmentation-models library to perform image segmentation. The dataset consisted of a total of 765 images, divided into training and testing sets with a ratio of 0.80 and 0.20, respectively. Each image was annotated with one of seven classes: grasslands, arable land, herb-dominated habitats, hedgerows, vineyards, tree-dominated man-made habitats, and *Olea europea* groves. Initially, the images and their corresponding masks were loaded from the specified directories and pre-processed accordingly. The masks were encoded using label encoding and then split into training and testing datasets. Next, the model was configured with specific parameters, including the EfficientNet backbone architecture and a combination of Dice loss and Categorical Focal loss as optimization objectives. To address the class imbalance inherent in the dataset, we implemented class weighting based on Dice scores. This technique assigns higher weights to classes with lower coverage, ensuring that the model learns effectively from all classes. The model was trained using hyperparameters, as shown in Table 1. After training, the model was saved for future reference. Performance metrics such as accuracy, precision, recall, F1 score, Jaccard score, and mean IoU were computed using the trained model on the test dataset. Finally, the results were printed to evaluate the segmentation model's performance.

Table 2 shows a comparative analysis of performance metrics for the three Unet, SegNet, and DeeplabV3 semantic segmentation models under two distinct conditions: with and without backbone integration. These metrics serve as quantitative indicators of model effectiveness in accurately segmenting images. Across all models, the incorporation of backbone architecture consistently leads to improvements in performance metrics. Specifically, models equipped with backbone demonstrate enhanced accuracy, precision, recall, and F1 scores compared to their counterparts lacking backbone integration. Accuracy, which

measures correct pixel classification, ranges from 0.638 to 0.763 with backbone and from 0.667 to 0.762 without it. Similarly, precision, recall, and F1 scores exhibit higher values when using backbone architecture. Additionally, metrics like the Jaccard coefficient (IOU) and mean IOU, which evaluate the overlap between predicted and ground truth masks, show substantial improvements with backbone integration. These findings underscore the importance of incorporating pre-trained feature extraction networks, or backbones, to enhance the performance of semantic segmentation models, thereby advancing their applicability in diverse image analysis tasks.

Table 1. Hyperparameters and configurations used in the semantic segmentation model training with pre-trained backbone for land cover segmentation.

Hyperparameter	Value
Number of epochs	100
Batch size	16
Patch size	128
Dice loss weights	[0.1428, 0.1428, 0.1428, 0.1428, 0.1428, 0.1428, 0.1428]
Focal loss	1.0 (weight in total loss calculation)
Loss function	Total loss (Dice loss + focal loss)
Optimizer	Adam
Input shape	(128, 128, 3)
Number of classes	7

Table 2. Performance metrics comparison of UNet, SegNet, and DeeplabV3 models with and without a backbone.

Performance Metrics	UNet		SegNet		DeeplabV3	
	Without Backbone	with Backbone	Without Backbone	With Backbone	Without Backbone	with Backbone
Accuracy	0.574	0.653	0.564	0.673	0.681	0.763
Precision	0.590	0.657	0.563	0.678	0.6855	0.761
Recall	0.574	0.653	0.566	0.673	0.681	0.763
F1 score	0.573	0.646	0.559	0.672	0.674	0.756
Jaccard coefficient (IOU)	0.411	0.500	0.399	0.528	0.5308	0.626
Mean IOU	0.306	0.323	0.290	0.407	0.410	0.520

Based on the performance metrics provided, it can be observed that the DeeplabV3 model consistently outperformed the Unet and SegNet models across the metrics, especially when backbone integration was considered. With backbone integration, DeeplabV3 achieved the highest values for accuracy, precision, recall, F1 score, Jaccard coefficient (IOU), and mean IOU among the three models. This indicates that DeeplabV3, especially when equipped with backbone architecture, offers superior image segmentation capabilities compared to Unet and SegNet. Therefore, in terms of overall performance and effectiveness in segmenting images, DeeplabV3 stands out as the preferred model choice.

Our investigation also evaluated the effectiveness of various pre-trained backbones for semantic segmentation of land cover using DeepLabV3. As shown in Table 3, Efficient NetB0 emerged as the leader, achieving the highest overall accuracy (76.35%) and F1-score (75.60%). This indicates its superior ability to accurately segment land cover classes in the imagery. Resnet-34 followed closely with an accuracy of 70.83% and F1-score of 70.77%. While InceptionV3 and DenseNet exhibited lower overall accuracy (around 71–75%), they maintained good precision and recall values for land cover class segmentation (refer to

Table 2 for details). Mean intersection-over-union (mIOU) and Jaccard score mirrored these trends, with EfficientNetB0 achieving the highest values (52.05% and 62.60%, respectively). These findings highlight the importance of selecting an appropriate pre-trained backbone for DeepLabV3 in land cover segmentation tasks. The optimal choice can significantly impact the model's ability to differentiate and delineate land cover classes.

Table 3. Comparison of semantic segmentation performance using DeepLabV3 with different pre-trained backbones.

Backbone	Accuracy	Precision	Recall	F1 Score	Mean IOU	Jaccard Coefficient (IOU)
Resnet 34	70.83	73.29	70.84	70.77	46.32	56.90
EfficientNetB0	76.33	76.10	76.30	75.60	52.00	62.60
InceptionV3	71.46	74.00	71.46	72.01	47.50	61.40
Densest	75.07	76.18	75.05	74.66	50.70	61.75

We analyzed the receiver operating characteristic (ROC) curves, as shown in Figure 9, to evaluate our multi-class classification model's performance. The curves visualize the trade-off between true positive rate (TPR) and false positive rate (FPR) for each class, while the micro-average ROC curve summarizes the overall performance. The area under curve (AUC) quantifies the model's discriminative power. High AUC values, with the micro-average reaching 0.95 and individual class values ranging from 0.93 to 0.94, indicate that our model effectively differentiates between the various land cover classes.

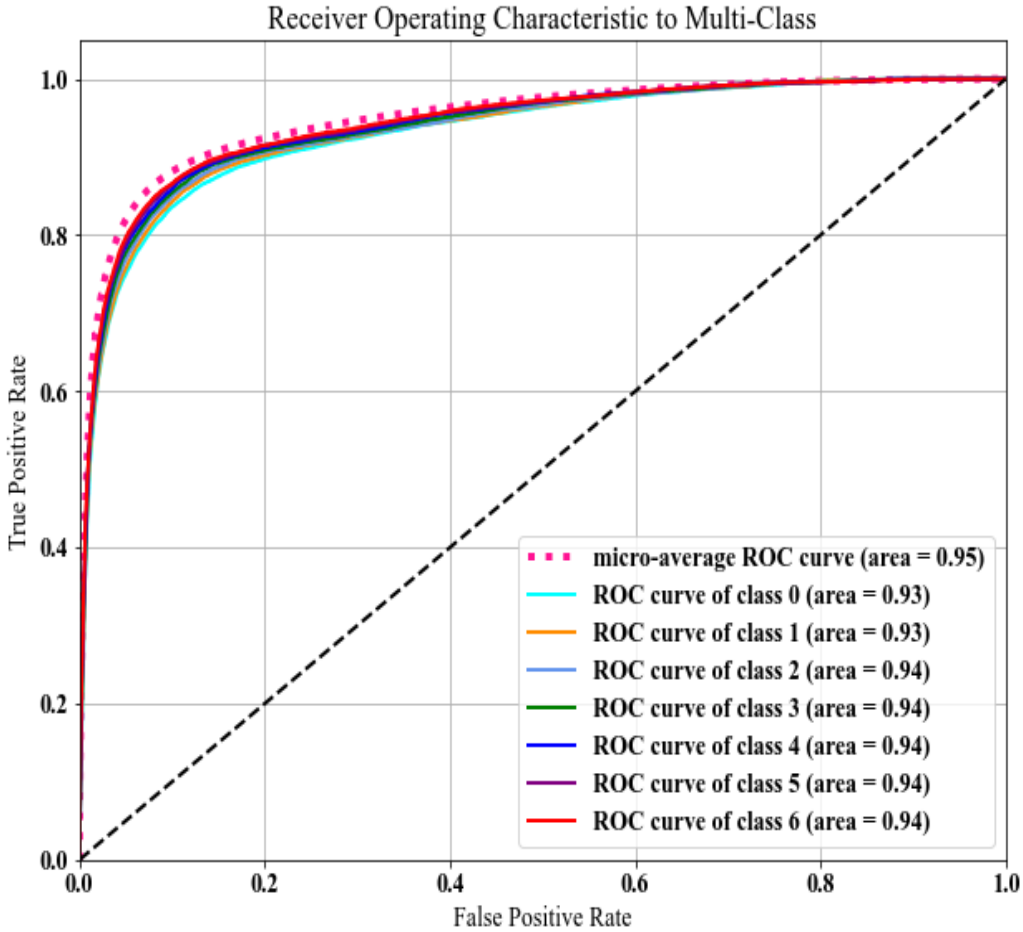

Figure 9. Receiver operating characteristic (ROC) curves show the performance of the multi-class land cover classification model. The micro-average ROC curve and individual class ROC curves demonstrate the model's strong ability to discriminate between different land cover classes.

After loading the trained model, we generated a batch of test images and their corresponding masks using the validation data generator. Next, we used the loaded model to compute predictions for the test images. Then we converted the predicted masks from categorical format to integer format for visualization and IoU calculation. Finally, in order to qualitatively assess the model's performance, we visualized a randomly selected test image along with its ground truth mask and predicted mask, as shown in Figure 10.

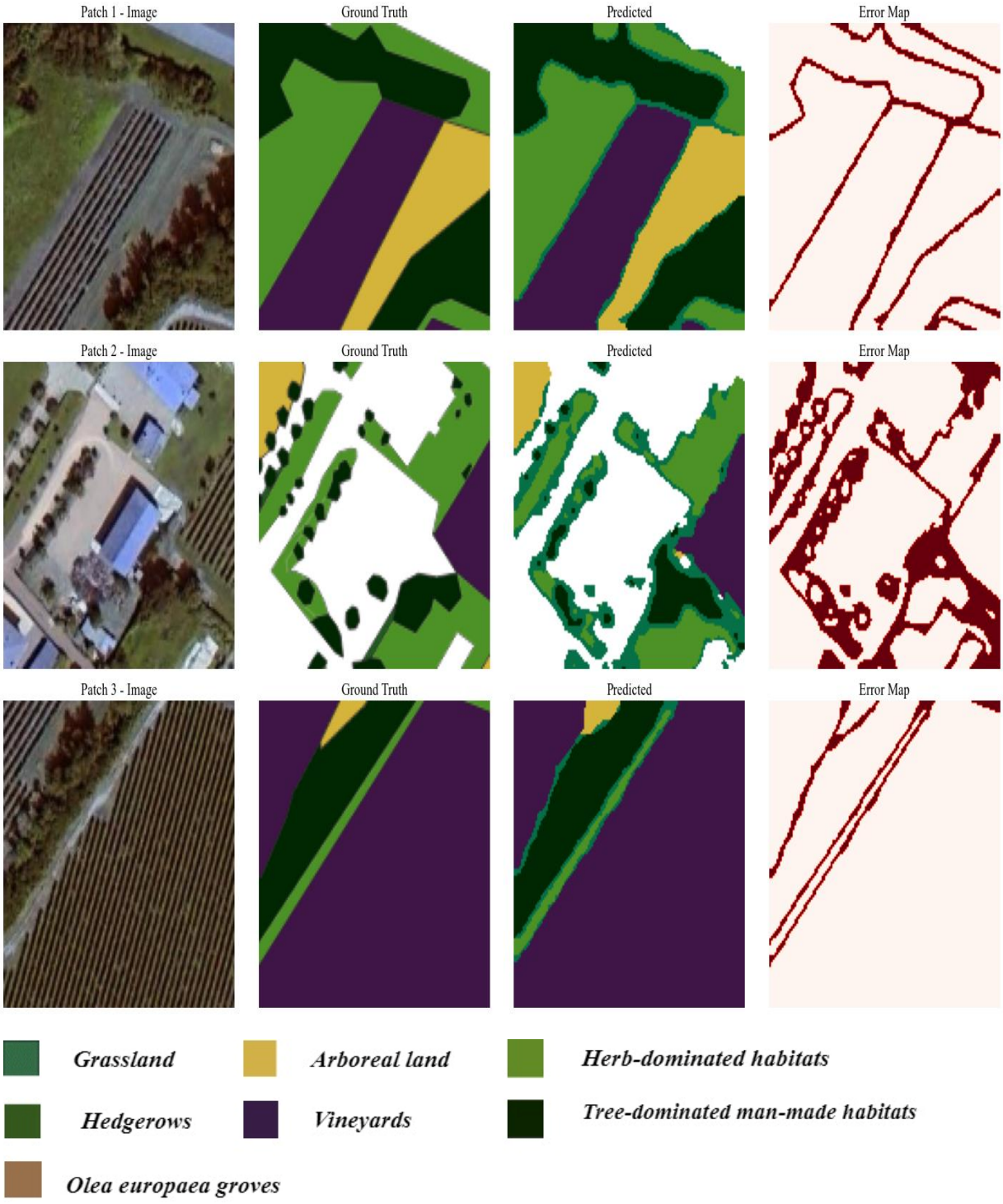

Figure 10. Display of a randomly selected test image, ground truth mask, predicted mask, and error map. The image illustrates the qualitative assessment of the model's performance, showing the alignment between the original test image, its corresponding ground truth mask, and the predicted mask generated by the model.

Finally, we applied the trained model to a large image (as shown in Figure 11). To achieve this, we segmented the image into patches of an appropriate size for processing. The prediction process was then carried out on each patch. Finally, the resulting segmented patches were stitched together to reconstruct the predicted mask for the entire large image. This approach facilitates the application of the model to images beyond the validation dataset's size.

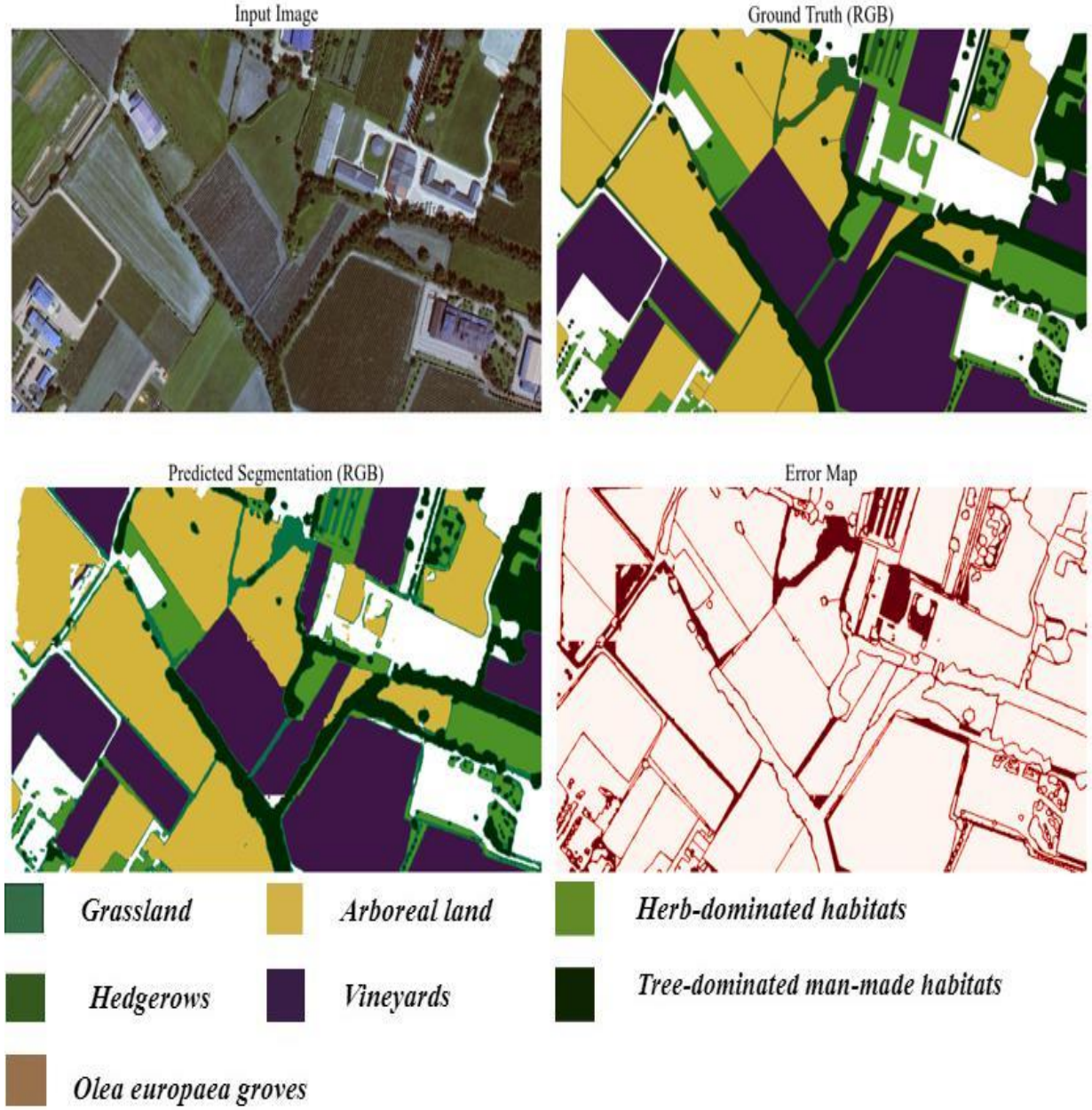

Figure 11. This figure illustrates the original image, the ground truth mask, the final segmented image mask obtained by applying a trained model, and the error map. The model segments the large image into patches, generates predictions for each patch, and seamlessly stitches the segmented patches together to reconstruct the predicted mask for the entire large image. The colors represent different land cover classes: green for grasslands, yellow for arable land, medium-dark green for herb-dominated habitats, purple for vineyards, black for tree-dominated man-made habitats, and brown for *Olea europaea* groves.

5. Discussion

In this study, we addressed key challenges and contributed significantly to the field of land cover mapping in agricultural areas. By focusing on semantic segmentation and using satellite imagery and deep learning models with a transfer learning backbone, we provided insights into the effectiveness of advanced techniques in accurately delineating land cover classes.

We introduced a land cover dataset annotated for seven distinct classes: grasslands, arable lands, herb-dominated habitats, hedgerows, vineyards, tree-dominated man-made habitats, and *Olea europea* groves. This dataset addresses a critical gap in the existing literature [32–34] by enabling comprehensive analysis of multiple land cover types within agricultural landscapes. Using this dataset, we developed a methodology for mapping the distribution of key land cover classes (grasslands, herb-dominated, and tree-dominated areas) across large areas of Franciacorta in Lombardy, Northern Italy. This mapping provides a foundational dataset for future research that studies the impact of surrounding land cover on vineyard ecosystem biodiversity and sustainability.

However, our approach also addresses a drawback highlighted in previous studies [17–19]. While traditional land cover mapping techniques have utilized classical machine learning algorithms and object-based image analysis (OBIA), they often struggle with spectral signature similarities, class heterogeneity, and limited generalizability. Additionally, acquiring specialized datasets for semantic segmentation has been challenging due to the high costs associated with data collection and manual annotation.

With careful data gathering, preparation, and model selection, we proved that deep learning models like UNet, SegNet, and DeepLabV3 excel in accurately mapping different land cover types from satellite imagery. By integrating pre-trained feature extraction networks, or backbones, into these models, we observed significant improvements in segmentation performance across various metrics. Our findings underscore the importance of leveraging advanced methodologies to achieve more accurate and reliable land cover mapping results. These advanced methodologies offer scalable and efficient solutions for environmental monitoring, conservation, and land management. By harnessing the power of satellite imagery and deep learning models, we can gain valuable insights into ecosystem health, biodiversity conservation, and land use dynamics.

The performance metrics of various segmentation models, as presented in Table 2, clearly demonstrate the superiority of the DeepLabv3 model with a backbone in achieving higher segmentation accuracy compared to UNet and SegNet. More specifically, DeepLabv3 with a backbone attained an accuracy of 0.763, a precision of 0.761, a recall of 0.763, an F1 score of 0.756, a Jaccard coefficient (IoU) of 0.626, and a mean IoU of 0.520. These metrics consistently outperform those of the other models tested with and without backbones. The enhanced performance of DeepLabv3 can be attributed to its advanced architecture, which integrates atrous spatial pyramid pooling (ASPP) and robust backbone networks, thereby facilitating multi-scale context aggregation and improved handling of complex spatial patterns [41]. This enables DeepLabv3 to excel in capturing fine-grained details and context, resulting in more accurate and reliable segmentation outcomes.

Among the evaluated backbones presented in Table 3, EfficientNetB0 demonstrates the highest performance across most metrics, including accuracy, precision, recall, F1 score, mean IOU, and Jaccard coefficient. This superior performance is due to EfficientNetB0's efficient use of computational resources while maintaining high accuracy and robustness in feature extraction. EfficientNetB0 leverages a compound scaling method that uniformly scales depth, width, and resolution, enabling it to achieve a balanced trade-off between model complexity and performance [36]. In contrast, while DenseNet and InceptionV3 also perform well, EfficientNetB0's advanced architectural optimizations provide it with a

more effective ability to capture detailed and diverse features, resulting in higher overall segmentation accuracy and more reliable performance metrics. ResNet34, while robust, lags in comparison, primarily due to its less efficient scaling and feature extraction capabilities.

Looking ahead, future researchers should focus on improving semantic segmentation models for land cover mapping, expanding the dataset to include a more extensive number of land cover classes beyond the seven considered in this study. Additionally, incorporating ground truth data collected through field surveys would enhance the accuracy and reliability of land cover mapping, particularly for complex and heterogeneous agricultural landscapes.

6. Conclusions

This study demonstrates the feasibility of employing satellite imagery and deep learning to accurately map land cover in agricultural regions. By developing a cost-effective and reproducible method, we created a novel dataset specifically tailored to wine-growing areas within the Franciacorta area. This dataset comprises seven manually annotated land cover classes, providing a valuable resource for agricultural research. Through rigorous comparison of state-of-the-art deep learning models, our research contributes significantly to advancing land cover mapping techniques.

The results demonstrate the effectiveness of deep learning models, particularly DeepLabv3, in accurately segmenting land cover classes from satellite imagery. DeepLabv3 with an EfficientNetB0 backbone outperformed other models in all performance metrics, making it the most reliable for land cover mapping. Its advanced architecture integrates atrous spatial pyramid pooling (ASPP) and efficient multi-scale context aggregation, enabling it to capture fine-grained details and handle complex spatial patterns effectively. This study highlights the potential of these advanced methods for supporting sustainable land management, environmental monitoring, and agricultural decision making. Furthermore, by analyzing different land cover types such as grasslands, herb-dominated habitats, and tree-dominated areas, we emphasize the importance of understanding how these surrounding habitats contribute to vineyard sustainability, fostering biodiversity, and enhancing ecosystem services like pollination and pest control.

Compared to traditional manual field-based mapping approaches, our methodology reduces the time required for data preparation by automating annotation tasks and utilizing open-access satellite imagery. Performance metrics show that DeepLabv3 with EfficientNetB0 improves segmentation accuracy by 10.9–13.3% and F1 score by 11.0–12.4% over other tested models. Additionally, it significantly lowers costs by avoiding the need for expensive high-resolution proprietary imagery and specialized equipment. While this research offers valuable insights, further investigations are necessary to enhance the generalizability of the findings to diverse agricultural regions. Expanding the dataset to include additional land cover classes and incorporating ground-truth data through field surveys are essential for refining model performance and expanding the application of these techniques. Moreover, this approach investigates the application of more advanced deep learning architectures.

Author Contributions: Conceptualization, G.T., G.G., I.S. and I.G.; methodology, G.T.; software, G.T., A.S.M., L.F. and C.B.; validation, I.G., G.G., I.S. and A.S.; formal analysis, G.T., I.G., G.G. and I.S.; investigation, G.T.; resources, G.T., A.S.M. and C.B.; data curation, G.T. and C.B.; writing—original draft preparation, G.T.; writing—review and editing, G.T., I.G., G.G. and I.S.; visualization, I.G. and G.G.; supervision, G.G.; project administration, G.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been partially supported by "Fondazione Cariplo" (Italy) and "the Lombardy Regional Government Authority" (Italy) under the project entitled: Progetto "Biodiversità, suolo e servizi ecosistemici: Metodi e tecniche per food system robusti, resilienti e sostenibili"—Bando Emblematici Maggiori 2020 and Climate Change AI project (No. IG-2023-174).

Data Availability Statement: The dataset presented in this study is available at the link https://doi.org/10.5281/zenodo.12189081 (accessed on 20 June 2024).

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

References

1. Olofsson, P.; Foody, G.M.; Herold, M.; Stehman, S.V.; Woodcock, C.E.; Wulder, M.A. Good practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* **2014**, *148*, 42–57. [CrossRef]
2. Williams, J.N.; Morandé, J.A.; Vaghti, M.G.; Medellín-Azuara, J.; Viers, J.H. Ecosystem Services in Vineyard Landscapes: A Focus on Aboveground Carbon Storage and Accumulation. *Carbon Balance Manag.* **2020**, *15*, 23. [CrossRef]
3. Giffard, B.; Winter, S.; Guidoni, S.; Nicolai, A.; Castaldini, M.; Cluzeau, D.; Coll, P.; Cortet, J.; Le Cadre, E.; D'errico, G.; et al. Vineyard Management and Its Impacts on Soil Biodiversity, Functions, and Ecosystem Services. *Front. Ecol. Evol.* **2022**, *10*, 850272. [CrossRef]
4. The Regenerative Viticulture Foundation. Biodiversity. Available online: https://www.regenerativeviticulture.org/toolkit/biodiversity/ (accessed on 9 October 2024).
5. Abad, J.; de Mendoza, I.H.; Marín, D.; Orcaray, L.; Santesteban, L.G. Cover crops in viticulture. A systematic review (1): Implications on soil characteristics and biodiversity in vineyard. *OENO One* **2021**, *55*, 295–3121. [CrossRef]
6. Hurajová, E.; Barroso, P.M.; Děkanovský, I.; Lumbantobing, Y.R.; Jiroušek, M.; Mugutdinov, A.; Havel, L.; Winkler, J. Biodiversity and Vegetation Succession in Vineyards, Moravia (Czech Republic). *Agriculture* **2024**, *14*, 1036. [CrossRef]
7. Paiola, A.; Assandri, G.; Brambilla, M.; Zottini, M.; Pedrini, P.; Nascimbene, J. Exploring the Potential of Vineyards for Biodiversity Conservation and Delivery of Biodiversity-Mediated Ecosystem Services: A Global-Scale Systematic Review. *Sci. Total Environ.* **2020**, *706*, 135839. [CrossRef] [PubMed]
8. Candiago, S.; Winkler, K.J.; Giombini, V.; Giupponi, C.; Egarter Vigl, L. An Ecosystem Service Approach to the Study of Vineyard Landscapes in the Context of Climate Change: A Review. *Sustain. Sci.* **2023**, *18*, 997–1013. [CrossRef] [PubMed]
9. Pal, M.; Mather, P.M. Support vector machines for classification in remote sensing. *Int. J. Remote Sens.* **2005**, *26*, 1007–1011. [CrossRef]
10. Pal, M.; Mather, P.M. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sens. Environ.* **2003**, *86*, 554–565. [CrossRef]
11. Cutler, D.R.; Edwards, T.C.; Beard, K.H.; Cutler, A.; Hess, K.T.; Gibson, J.; Lawler, J.J. Random forests for classification in ecology. *Ecology* **2007**, *88*, 2783–2792. [CrossRef] [PubMed]
12. Laban, N.; Abdellatif, B.; Ebeid, H.M.; Shedeed, H.A.; Tolba, M.F. Machine Learning for Enhancement Land Cover and Crop Types Classification. In *Machine Learning Paradigms: Theory and Application*; Hassanien, A.E., Ed.; Springer International Publishing: Cham, Switzerland, 2019; pp. 71–87. [CrossRef]
13. Phiri, D.; Simwanda, M.; Salekin, S.; Nyirenda, V.R.; Murayama, Y.; Ranagalage, M. Sentinel-2 Data for Land Cover/Use Mapping: A Review. *Remote Sens.* **2020**, *12*, 2291. [CrossRef]
14. Gauci, A.; Abela, J.; Austad, M.; Cassar, L.F.; Zarb Adami, K. A Machine Learning approach for automatic land cover mapping from DSLR images over the Maltese Islands. *Environ. Model. Softw.* **2018**, *99*, 1–10. [CrossRef]
15. Mardani, M.; Mardani, H.; De Simone, L.; Varas, S.; Kita, N.; Saito, T. Integration of Machine Learning and Open Access Geospatial Data for Land Cover Mapping. *Remote Sens.* **2019**, *11*, 1907. [CrossRef]
16. Pelletier, C.; Valero, S.; Inglada, J.; Champion, N.; Dedieu, G. Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sens. Environ.* **2016**, *187*, 156–168. [CrossRef]
17. Ouchra, H.; Belangour, A.; Erraissi, A. Machine Learning Algorithms for Satellite Image Classification Using Google Earth Engine and Landsat Satellite Data: Morocco Case Study. *IEEE Access* **2023**, *11*, 71127–71142. [CrossRef]
18. Han, R.; Liu, P.; Wang, G.; Zhang, H.; Wu, X. Advantage of Combining OBIA and Classifier Ensemble Method for Very High-Resolution Satellite Imagery Classification. *J. Sens.* **2020**, *2020*, 8855509. [CrossRef]
19. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [CrossRef]
20. Zaabar, N.; Niculescu, S.; Kamel, M.M. Application of Convolutional Neural Networks With Object-Based Image Analysis for Land Cover and Land Use Mapping in Coastal Areas: A Case Study in Ain Témouchent, Algeria. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 5177–5189. [CrossRef]

21. Zhang, X.; Han, L.; Han, L.; Zhu, L. How Well Do Deep Learning-Based Methods for Land Cover Classification and Object Detection Perform on High Resolution Remote Sensing Imagery? *Remote Sens.* **2020**, *12*, 417. [CrossRef]
22. Chi, M.; Plaza, A.; Benediktsson, J.A.; Sun, Z.; Shen, J.; Zhu, Y. Big Data for Remote Sensing: Challenges and Opportunities. *Proc. IEEE* **2016**, *104*, 2207–2219. [CrossRef]
23. Ienco, D.; Gbodjo, Y.J.E.; Gaetano, R.; Interdonato, R. Weakly Supervised Learning for Land Cover Mapping of Satellite Image Time Series via Attention-Based CNN. *IEEE Access* **2020**, *8*, 179547–179560. [CrossRef]
24. Yang, N.; Tang, H. Semantic Segmentation of Satellite Images: A Deep Learning Approach Integrated with Geospatial Hash Codes. *Remote Sens.* **2021**, *13*, 2723. [CrossRef]
25. 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]
26. Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Garcia-Rodriguez, J. A Review on Deep Learning Techniques Applied to Semantic Segmentation. *arXiv* **2017**, arXiv:1704.06857. [CrossRef]
27. Marmanis, D.; Wegner, J.D.; Galliani, S.; Schindler, K.; Datcu, M.; Stilla, U. Semantic Segmentation of Aerial Images with an Ensemble of CNNs. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* **2016**, *3*, 473–480. [CrossRef]
28. Li, R.; Zheng, S.; Duan, C.; Wang, L.; Zhang, C. Land Cover Classification from Remote Sensing Images Based on Multi-Scale Fully Convolutional Network. *Geo-Spat. Inform. Sci.* **2022**, *25*, 278–294. [CrossRef]
29. Tzepkenlis, A.; Marthoglou, K.; Grammalidis, N. Efficient Deep Semantic Segmentation for Land Cover Classification Using Sentinel Imagery. *Remote Sens.* **2023**, *15*, 2027. [CrossRef]
30. Xu, R.; Wang, C.; Zhang, J.; Xu, S.; Meng, W.; Zhang, X. RSSFormer: Foreground Saliency Enhancement for Remote Sensing Land-Cover Segmentation. *IEEE Trans. Image Process.* **2023**, *32*, 1052–1064. [CrossRef]
31. Längkvist, M.; Kiselev, A.; Alirezaie, M.; Loutfi, A. Classification and Segmentation of Satellite Orthoimagery Using Convolutional Neural Networks. *Remote Sens.* **2016**, *8*, 329. [CrossRef]
32. Vali, A.; Comai, S.; Matteucci, M. Deep Learning for Land Use and Land Cover Classification Based on Hyperspectral and Multispectral Earth Observation Data: A Review. *Remote Sens.* **2020**, *12*, 2495. [CrossRef]
33. Digra, M.; Dhir, R.; Sharma, N. Land Use Land Cover Classification of Remote Sensing Images Based on Deep Learning Approaches: A Statistical Analysis and Review. *Arab. J. Geosci.* **2022**, *15*, 1003. [CrossRef]
34. Zhao, S.; Tu, K.; Ye, S.; Tang, H.; Hu, Y.; Xie, C. Land Use and Land Cover Classification Meets Deep Learning: A Review. *Sensors* **2023**, *23*, 8966. [CrossRef] [PubMed]
35. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
36. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
37. Weinberger, K.Q. Densely Connected Convolutional Networks. *arXiv* **2018**, arXiv:1608.06993. [CrossRef]
38. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv* **2020**, arXiv:1905.11946. [CrossRef]
39. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv* **2015**, arXiv:1505.04597. [CrossRef]
40. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef] [PubMed]
41. 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. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [CrossRef] [PubMed]