remote sensing

Article

Agricultural Field Crop Type Semantic Segmentation and Boundary Extraction from Sentinel-2 Time Series Using Multitask Learning with Directional Feature-Sharing

Milena Atanasova [1,2], Luca Bergamasco [2] and Francesca Bovolo [2,*]

[1] Department of Information Engineering and Computer Science, University of Trento, 38123 Trento, Italy; matanasova@fbk.eu
[2] Fondazione Bruno Kessler, 38123 Trento, Italy; lbergamasco@fbk.eu
[*] Correspondence: bovolo@fbk.eu

Highlights

What are the main findings?
- We modeled the correlation between crop field semantic type and boundaries for agricultural field characterization improvement.
- Selectively and directionally sharing information between tasks is beneficial.

What are the implications of the main findings?
- Features from various agricultural tasks should be used in conjunction to better optimize the several correlated objectives.
- Careful design of shared and task-specific parts of the network is needed to ensure the tasks interfere with one another in a constructive way.

Abstract

Precise analysis of crop fields is essential for agricultural management. Various remote sensing tasks comprise the characterization of agricultural lands. The automatic extraction of boundaries and the crop type semantic segmentation tasks, both benefiting from satellite image time-series analysis, are crucial for continuous field monitoring. Recent models explore the multitask setting for boundary detection, but they do not incorporate class information that is crucial when several kinds of agricultural crops co-exist. Inspired by the idea that crop type segmentation and boundary extraction are closely connected, this study exploits a multitask learning framework for both agricultural segmentation and boundary detection tasks. The model uses a shared encoder with 3D convolutional blocks and task-specific decoders with cross-task feature exchange to simultaneously solve the two main tasks of crop type segmentation and boundary extraction. The model also leverages on estimating the distance to the closest border as an auxiliary task to improve training. The model effectively shares features across tasks and solves them simultaneously, achieving detailed crop-field characterization. To validate the method performance and examine inter-task relationships, two datasets composed of time series of Sentinel-2 images acquired from two agricultural areas in Austria, spanning two different years, were used.

Keywords: multitask learning; cross-task feature-sharing; crop type mapping; boundary extraction; satellite image time series; deep learning

Academic Editor: Jaromir Krzyszczak

Received: 16 June 2026
Revised: 23 July 2026
Accepted: 7 August 2026
Published: 12 August 2026

1. Introduction

Agricultural practices require continuous monitoring and extensive descriptions of agricultural land. The shift to effective and efficient agriculture requires advanced crop management and characterization techniques. Remote sensing images, which provide timely and objective information about crops [1], enable the development of algorithms that provide insights into the land by solving various tasks related to agricultural fields. Some of these tasks are crop type mapping, parcel identification, and border delineation. Obtaining diverse information across various tasks is important for agricultural practices, such as planning resources, structures, and optimizing land resources [2]. Thus, agricultural statistics are useful for understanding the distribution of crop types and ensuring food production [3], and describing the fields becomes of great significance also in terms of the socio-economic importance of agriculture [4].

Satellite image time series provide information in the spatial, spectral, and temporal domain [5]. They are a valuable tool for semantic segmentation of crop types [5]. Crop type mapping relies on the crop behaviors and can be performed at the pixel, parcel, or image level. The task of identifying the agricultural class from RS images benefits from the use of time series to examine the spectral behavior of plants throughout their phenological stages and distinguish them based on their growth stages, trends, peaks, etc. Several methods are proposed that use mechanisms for extracting spatial and temporal information. The work in [6] applies the convolutional operation in the temporal dimension to learn temporal and spectral features and consequently perform classification of satellite image time series, which is validated for a crop type classification task. The model in [7] uses a 3D CNN for multitemporal crop classification and shows that the simultaneous analysis of time and space is beneficial for crop type mapping. The method in [8] stacks image time series and explores band combinations for performing crop type classification. Some methods classify a particular type of crop [9,10], others deal with the task of segmentation of active and abandoned fields [11], or Integrated Crop–Livestock Systems [12,13]. However, these methods are designed for specific crop types and struggle to generalize.

Another noteworthy task in describing agricultural land is delineating borders and locating agricultural fields [14]. Since boundary extraction is influenced by the field size and their phenology [15], using image time series and semantic information is useful. Automated boundary delineation is usually based on image segmentation [16]. It is commonly performed with region-based or edge-based techniques [17]. The use of deep learning techniques enables more stable border delineation [18]. Field boundary detection has been often studied as a preliminary step necessary to perform parcel identification. Some methodologies follow a process where they first extract field boundaries and after that perform classification [19]. In such cases, the boundaries are seen as an auxiliary task, while the field classification is the main one. Most methods in the literature rely on single temporal images instead of time series [20–23]. This approach omits an important observation: during certain seasons, adjacent fields may appear the same, even if they differ (e.g., raw soil before planting), leading to poor boundary identification. Therefore, some models [3,19,24–26] use time series for boundary extraction. The model in [27] integrates multitemporal imagery into a segmentation-based workflow for enhancing the border delineation of land parcels. These methods are effective, but they do not include classification information—the crop type. And indeed, the crop type is a factor that influences the performance of both the border delineation and field extraction tasks [28].

Overall, the literature models have not considered semantic information when performing border or field delineation, and vice versa (i.e., not explicitly solving segmentation of crop types guided by boundaries), even though these agricultural tasks are strongly related and influence each other.

Correlated objectives can be analyzed using multitask learning (MTL) frameworks. MTL enables combining multiple objectives within a single approach. It learns tasks in parallel using a shared representation and fosters the reciprocal benefits of task learning [29]. MTL allows for representation learning across multiple related tasks [28]. MTL models can achieve either positive or negative transfer between objectives. Positive transfer refers to when tasks are helping each other, while negative transfer refers to when tasks have sub-optimal performance compared to when trained separately [30]. It is important to carefully examine which parts of the neural networks are shared, how feature sharing is implemented, and in what terms the loss function is defined. Single-task models may overlook potential synergies of the relevant tasks in agricultural fields (boundary extraction and semantic segmentation). The training signals for these tasks may indeed be beneficial to each other and thus boost the performance of each. In agricultural remote sensing, MTL has several uses. The model in [31] combines field survey data, Sentinel-2 imagery, and meteorological records to support dynamic early-season yield prediction. The work in [32] combines classification and regression of yield at the crop level. The model incorporates shared layers from multi-task learning to indirectly improve the prediction results of the regression task while optimizing the classification one. The study in [10] performed classification of rice fields on a large-scale spatio-temporal SAR dataset using MTL in a region-specific way, thus exploring common and region-specific features simultaneously.

Several multitask models exist for boundary detection. Since the boundary task is closely related to the regression task on distance to boundaries and the field segmentation task, some methods jointly solve them to leverage their mutual representations. The model in [21] integrates three tasks: mask prediction, edge prediction, and distance map estimation to obtain closed-boundary agricultural parcels. The work in [20] uses high-resolution images and three parallel decoders, assigning as the main task the agricultural field identification and, as auxiliary tasks, field boundary prediction and distance estimation. These tasks are combined using a Unet [18]. The method in [26] learns the tasks of crop-field extent, boundary, and distance from multitemporal data simultaneously, but relies on local standard deviation and does not explicitly model temporal patterns, which are essential for crop-type identification.

However, none of these methods consider the information regarding the crop type. Since different crops have different behaviour, different fields may have stronger or weaker field boundaries. The different crop types may also define the relative sizes of the fields (e.g., wheat and maize are usually grown in larger fields, while fruits or legumes are typically grown in smaller fields [28]). In fact, the heterogeneity in both field form and class type are some of the main motivations why automatic algorithms (rather than field work research or manual labeling which is labour-intensive and costly [33,34]) is beneficial in agricultural field characterization.

In this paper, we propose a study on agricultural field characterization using a multitask deep learning strategy that handles crop-type segmentation, field boundary delineation, and distance to boundary estimation (as an auxiliary task). In this study, we adapt an MTL framework [35] to solve the agriculturally related tasks and examine the optimal way to perform information sharing across three tasks (where [35] deals with two-task oprions only). Since the tasks differ in characteristics, it is valuable to consider their importance and relationships with one another to construct the ideal setting.

The model takes as input a multispectral remote sensing image time series. At first, the time series is passed through an encoder that is shared for all the downstream tasks. It extracts relevant spatio-temporal features using 3D convolutional encoder blocks. Having a single encoder encloses the information useful for all subsequent tasks, allowing the learning of a rich and robust representation of the input image time series. After the encoder, we use three decoding branches, one for each task, to perform the tasks. They use 3D convolutions and attention modules. During the decoding stage, information sharing is occurring between the features of the different tasks. Specific blocks using gating mechanisms are applied to selectively transfer information. The information is shared directionally across the three tasks, because they are not symmetrically related to one another: from the boundary extraction task to the segmentation and distance, and from the segmentation and distance to the boundaries. This novel three-task designed directional cross-task sharing mechanism helps to control the flow of feature sharing during the decoding phase of the network. By combining tasks in one architecture, the objective is to achieve positive influence between the selected tasks, better performance and perform a more thorough description of the agricultural land.

The attention of this work is on dealing with small and thin shaped crop fields and strong variability in the crop-type maps. We explore the utilization of multitask learning with time series data for the joint solving of three tasks of semantic segmentation, boundary extraction (and auxiliary distance to boundary estimation) for benefitting the agricultural applications.

To evaluate the proposed scheme, we provide an extensive experimental analysis on two study areas from two different years in Austria.

The contributions of this paper are the following:

- On the basis of a specific MTL framework [35], a novel realization for three agricultural downstream tasks are identified that benefit from mutual solving. This work studies the optimal configuration of the model and the cross-task sharing across the three tasks.
- The three cross-task relationships of segmentation, border delineation, and distance to boundary estimation are thoroughly examined in an agricultural setting.
- Positive transfer between three agricultural tasks is demonstrated, such as the inclusion of crop type information, which is beneficial for extraction of the boundaries of agricultural fields.

This paper is organized as follows: Section 2 explains the materials and data used for this study. Section 3 explains the methodology. Section 4 demonstrates the results and the experiments performed. Section 5 provides a discussion, and Section 6 draws a conclusion about the work conducted.

2. Materials

This section introduces the areas of study and the data preparation and pre-processing for analysis.

The two study areas are located in Austria. The first area, denoted as Austria I (the green rectangle in Figure 1) is located in the North-East of the country and concerns the year 2021. It is characterized by its agricultural land-cover, consisting of small and sometimes quite thin crop fields. The average size of the fields is 1.4 ha (with a standard deviation of 2.5 ha), while the median is 0.6 ha. The total area analyzed is around 3000 km^2. The second area, denoted as Austria II (in orange in Figure 1) is about the year 2022, it has irregularly shaped and diversely oriented fields, and the average crop field size is 1.1 ha (with a standard deviation of 1.6 ha), and the median is 0.6 ha.

Annual crop types were obtained from the available maps coming from farmers' declarations [36]. The sites show substantial variability in crop types and spatial distribution—often, fields of the same crop type are spread across the area and are not concentrated in a particular zone. From the available multipolygon file of the field segmentation, we obtained and rasterized reference maps of the field segmentation into their agricultural crop types. The minority classes were eliminated and the spectrally similar declared classes were grouped in order to have classes that can be discriminated using time series of remote sensing images [37,38]. This resulted in a challenging classification problem of 17 detailed crop type classes, with various statistical distributions (see Figure 2). We define the reference segmentation map as Y_s.

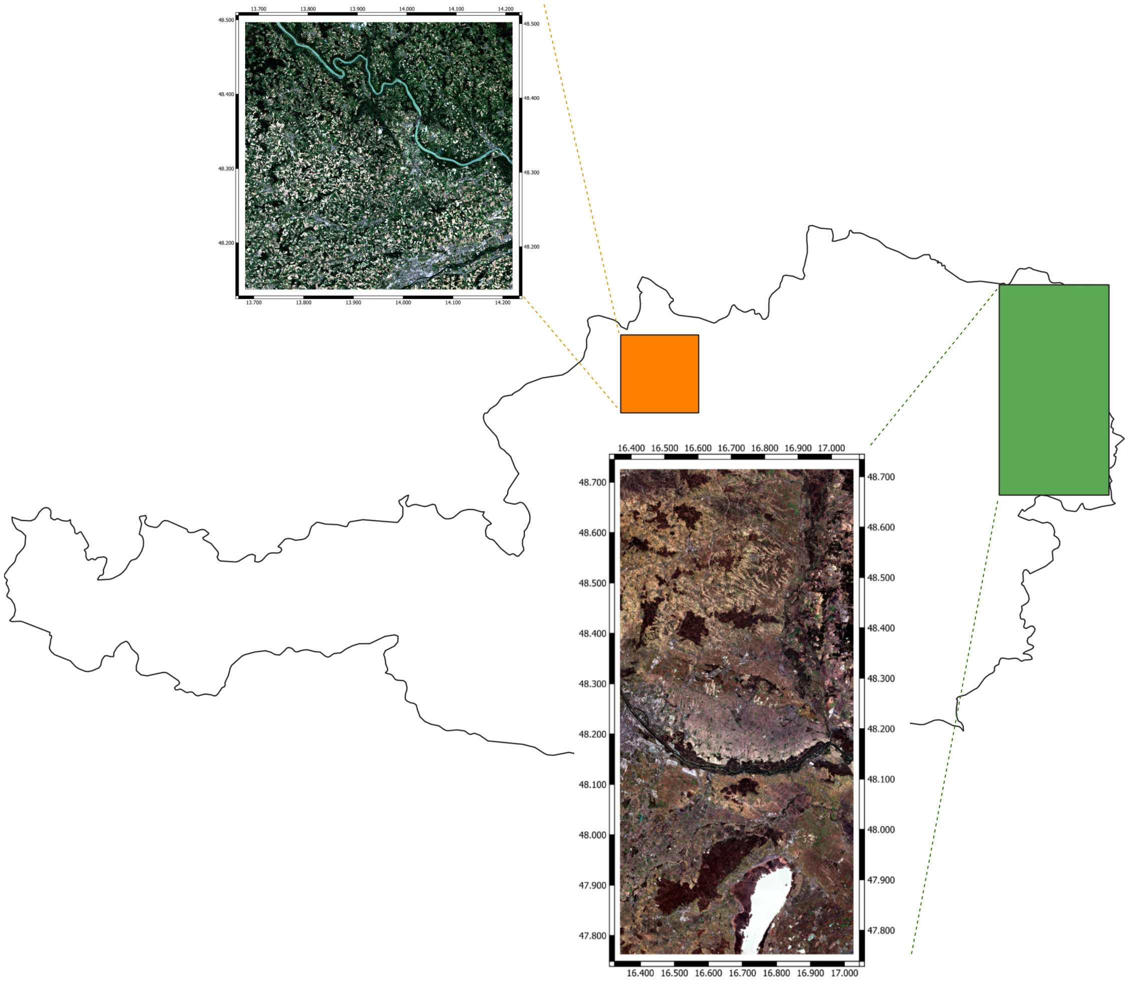


Figure 1. Spatial location of the two study areas in the Austrian map.

From the vector crop-type map, we extracted the boundaries of all fields using the QGIS software (QGIS Development Team. QGIS Geographic Information System, version 3.34; Open Source Geospatial Foundation: Beaverton, OR, USA, 2024. Available online: https://qgis.org) and the function "Boundary" that returns the topological boundary of the geometry. The boundaries were rasterized. The conversion into a raster achieves one-pixel raster representation of the vector line. The rasterization process also attaches a spatial resolution to the map. In this case, the resolution is 10m as we are using Sentinel 2 images. This operation introduces uncertainty that can be of around 1 pixel. To account for these uncertainties, and for the inherent uncertainties in the data creation, we include a buffer of one pixel in the reference boundary map. We denote the boundary map by Y_b. This map is binary, containing only two classes: boundary ω_b and non-boundary ω_{nb}.

For the distance to boundary map, we calculated the distance of each pixel to the nearest border, calculating its shortest distance to the background (the boundaries) using a function from the Python package *scipy* 1.18.0 (SciPy Developers. *scipy.ndimage.distance_transform_edt*; SciPy Documentation. Available online: https://docs.scipy.org/doc/scipy/reference/generated/scipy.ndimage.distance_transform_edt.html, accessed on 3 March 2026). We denote this reference as Y_d, which contains integer values.

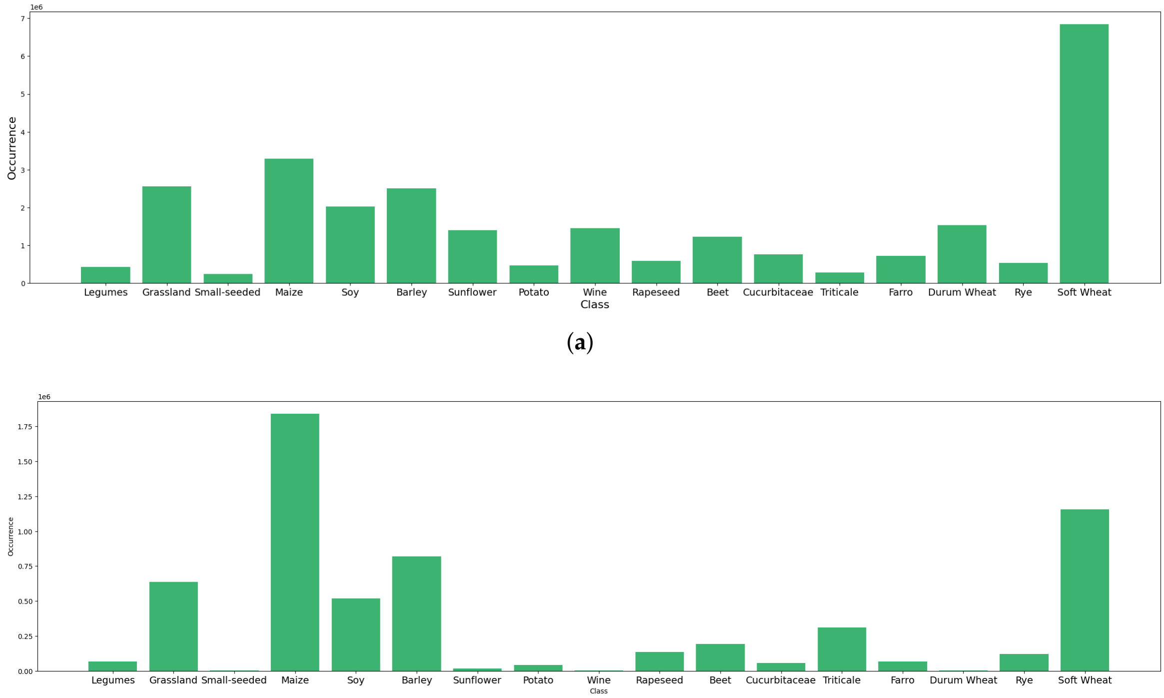


Figure 2. Distribution of the crop type classes in the reference maps: (**a**) Austria I. (**b**) Austria II.

The data source was Sentinel-2 image time series. Let us define as $\mathcal{X} = \{X_1, X_2, ...X_N\}$ the input image time series. $\mathcal{X}$ includes $N = 12$ images over the farming year (September to August) for 2021 and 2022 for Austria I and Austria II, respectively (Table 1). The choice was influenced by the availability of cloud-free data, as we selected samples that contain less than 10% cloudy observations throughout the time series. The usage of 12 dates per year allows for evenly distributed data throughout the year, and captures the growing period of the major types of crops. Still, the proposed methodology is general and can be used with a different number of images N in the time series. Ten of the available spectral bands were kept for analysis, excluding those at 60 m, as they do not provide information on vegetation characteristics. We rescaled all bands to 10 m/pixel spatial resolution. We normalized the data using the 2nd and 98th percentiles of the image time series.

$$X_i^{norm} = \frac{X_i - \mathcal{X}^{perc2}}{\mathcal{X}^{perc98} - \mathcal{X}^{perc2}} \quad (1)$$

Table 1. Dates of the Sentinel-2 images used to construct the time series dataset for analysis.

t	Austria I	Austria II
1	14/09/2020	05/09/2021
2	09/10/2020	25/09/2021
3	22/10/2020	20/10/2021
4	18/11/2020	31/12/2021
5	01/02/2021	24/02/2022
6	24/02/2021	11/03/2022
7	26/03/2021	26/03/2022
8	12/04/2021	13/04/2022
9	10/05/2021	15/05/2022
10	24/06/2021	19/06/2022
11	26/07/2021	19/07/2022
12	03/08/2021	11/08/2022

In order to construct the datasets for the model, the image time series were divided in patches of size $N \times C \times h \times w$, where $N = 12$ is the number of images in the time series, $C = 10$ is the number of input channels (spectral bands), and $h = 64$ and $w = 64$ are the height and width of the patches. An input time series patch x_i thus consists of multispectral images $\{x_i^1, x_i^2, ...x_i^N\}$. The reference labels related to x_i are the patches $y_{i,s} \in Y_s$, $y_{i,b} \in Y_b$, and $y_{i,d} \in Y_d$ (Figure 3). Thus, the training set is defined as:

$$TR = \left\{ \{x_i, (y_{i,s}, y_{i,b}, y_{i,d}) / x_i = \left\{x_i^1, x_i^2, ...x_i^N\right\}, y_{i,s} \in Y_s, y_{i,b} \in Y_b, y_{i,d} \in Y_d\} \right\}. \quad (2)$$

Working in patches allows us to process areas of arbitrary size and account for computational resources. Example time-series patches with corresponding reference maps for segmentation, boundary extraction, and distance to boundary estimation are shown in Figure 3.

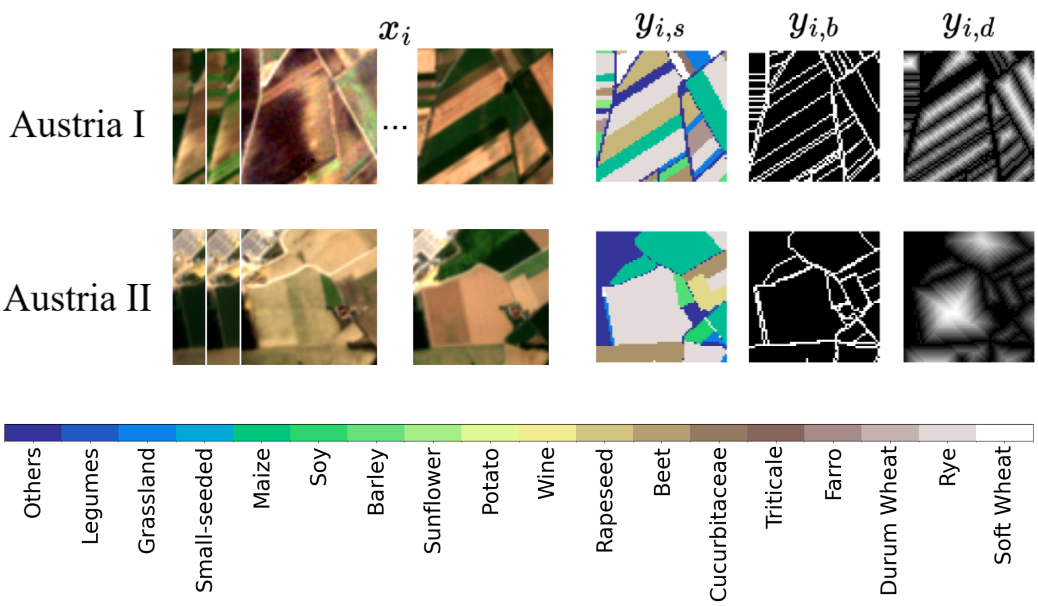


Figure 3. Example time series patches x_i and corresponding reference patches $y_{i,s}$, $y_{i,b}$ and $y_{i,d}$. The figure also presents the colorbar with the color mapping of the classes for the crop type segmentation, which is used throughout the whole article.

3. Method

3.1. Multitask Learning Model for Agricultural Field Characterization

The proposed method performs agricultural crop field characterization. It focuses on two tasks of equal relevance: crop type semantic segmentation and crop boundary extraction, and on a third auxiliary task of distance to boundary calculation. The optimal approach to tailor a multitask deep learning model, based on the framework proposed in [35], was studied to properly design a three task architecture and set the agricultural tasks of interest. This work builds on the framework in [35] by examining a three-task problem (instead of two) and thus focusing on optimal layer definition and cross-task feature-sharing to deal with the three tasks. The model solves the tasks tailored to agricultural parcel optimization, strengthening the usefulness of multitask learning in agricultural analysis with remote sensing time series, and examining the relationship across network features from the tasks.

The model architecture comprises a shared encoder and task-specific decoders enriched with spatio-temporal attention. The encoder obtains a rich representation of the input time series, exploiting spatial and temporal information, while the decoders handle the specific prediction tasks. The model is supervised and produces the following segmentation outputs: a semantic segmentation crop-type map, a crop field boundary map, and a distance-to-boundaries map (Figure 4). It leverages the multitemporal information contained in the image time series.

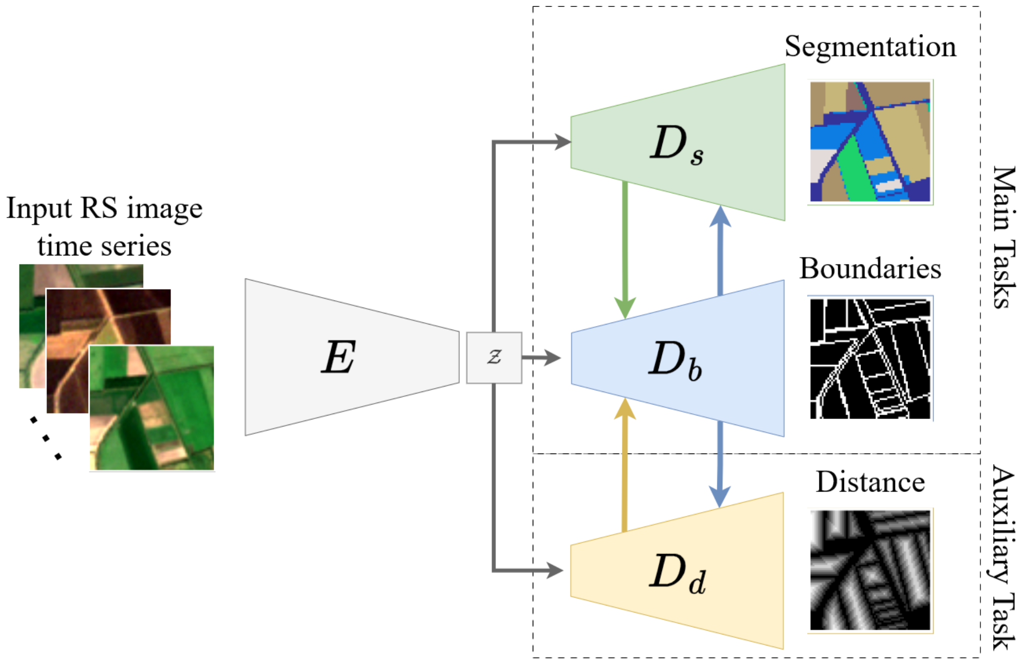


Figure 4. The algorithm block scheme. The input image time series is forwarded to an MTL neural network which extracts common relevant features with a shared encoder, after which the network divides into three decoder branches: for the semantic segmentation, boundary delineation and distance estimation tasks. The figure also outlines the directionality of the information exchange between the decoding parts.

An encoder E is shared among tasks and is composed of stacked 3D convolutions with dilation, batch normalizations and ReLU activation functions. The formula of the 3D convolution is the following:

$$F_l = \phi(W_l * F_{l-1} + b_l) \tag{3}$$

where F_l are the features at layer l, $*$ represents the convolution operation, W_l and b_l are the weights and biases, and $\phi(\cdot)$ is the activation function. E extracts and analyzes simultaneously spatial and temporal information, exploiting correlations among these dimensions. The encoder extracts representative common features $\mathcal{Z}$ for all the tasks from the input time series $\mathcal{X}$: $\mathcal{Z} = E(\mathcal{X})$.

The model divides into three decoding branches D_s, D_b, and D_d that produce the semantic segmentation, boundary, and distance maps, respectively. The decoders use 3D convolutions, spatio-temporal attention mechanisms (CBAM blocks [39]), and share cross-task spatio-temporal features between the layers. The distance to boundaries auxiliary task is only used during training, and thus we do not perform inference on it.

3.2. Feature Sharing

In the decoding part of the network, task-specific layers handle the prediction tasks. Selective cross-task sharing is performed (Figure 5) that allows to exchange features selectively and directionally between decoders, thereby benefiting from positive transfer between tasks. This strategy enables, on the one hand, considering the features of each task and, on the other hand, enhancing them with selective information from the other tasks.

To do this, gates G are designed that control the amount of shared information. Information is not inter-exchanged in an all-to-all fashion. This is because all-to-all task exchanges may hurt performance due to differences in task similarity and thus their correlation. To enforce the positive and limit the negative interference between these tasks, we propose using directional cross-task feature-sharing. This sharing occurs between convolutional blocks, with information continuously exchanged at various extraction levels. The gates operate at each decoder level independently. The directionality stays consistent in all the blocks. Therefore, the decoding parts consist of a convolutional block, followed by spatio-temporal CBAM attention [39] and the cross-task selective feature-sharing module, and these are repeated L times, where L is the number of decoding layers. Thus the network uses $L-1$ cross-task feature exchange blocks.

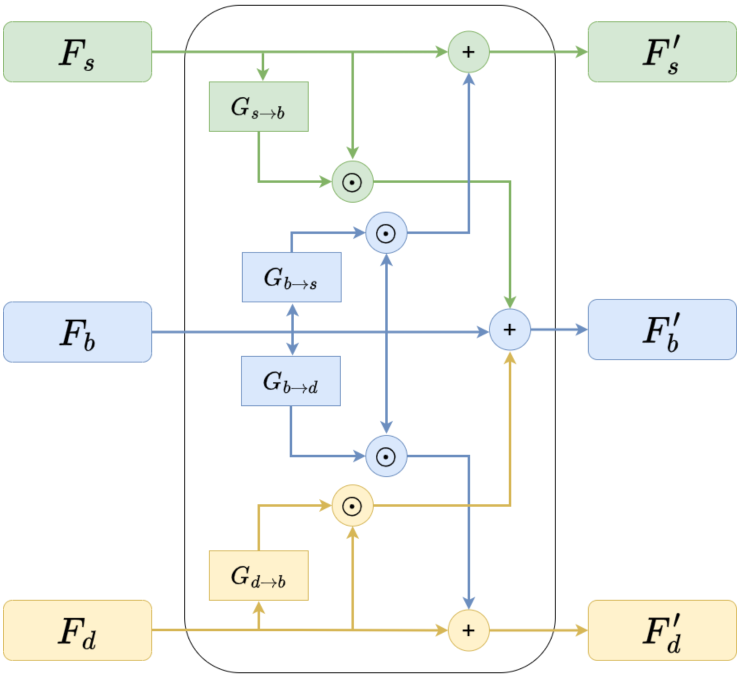


Figure 5. Direction feature exchange module.

We define the following gates:

$$\begin{aligned} G_{s\to b} &= \sigma(W_{s\to b} * F_s) \\ G_{d\to b} &= \sigma(W_{d\to b} * F_d) \\ G_{b\to s} &= \sigma(W_{b\to s} * F_b) \\ G_{b\to d} &= \sigma(W_{b\to d} * F_b) \end{aligned} \tag{4}$$

where s, b, and d define the semantic crop type segmentation, boundary extraction, and distance tasks, and F_s, F_b, and F_d are the corresponding features for these tasks, respectively; W are weights of a 3D convolutional operation, and σ is the sigmoid operation. The gates determine the weight assigned to features across tasks. In order to use the information provided by the gates, element-wise multiplication $\odot$ with the task features is applied to obtain the enhanced features after the cross-task exchange.

In the proposed method, we determine the direction of the gates based on considerations of the relatedness of the tasks. For the boundary detection task, features are transferred from both the semantic segmentation and the distance to boundaries tasks since they are closely related to the boundaries and provide benefits. The distance to boundaries has already been used as an auxiliary task in previous research [20,21], highlighting its usefulness in obtaining better boundary delineation. Meanwhile, semantic segmentation helps with boundary identification by providing information about crop type. It enables improved boundary detection in cases of crop spectral similarities. Different crops exhibit distinct spatial structures and spatio-temporal patterns which directly affect the boundary separability [28]. The enhanced features for the boundary detection F_b' become:

$$F_b' = F_b + G_{s\to b} \odot F_s + G_{d\to b} \odot F_d \tag{5}$$

Regarding the segmentation and distance tasks, feature sharing occurs from boundary extraction to each of them. Boundary extraction is closely related to both segmentation and distance to boundaries. The enhanced features for the semantic segmentation F_s' and distance to boundaries tasks F_d' become:

$$\begin{aligned} F_s' &= F_s + G_{b\to s} \odot F_b \\ F_d' &= F_d + G_{b\to d} \odot F_d \end{aligned} \tag{6}$$

Information in the decoder is not shared between the semantic segmentation task and the distance to boundaries task, since neither is closely influenced by the other (and vice versa), and such a combination leads to negative transfer. The regression task of distance estimation is concerning the geometrical details, while the crop type segmentation strongly concerns the spectral-temporal patterns. After these cross-task information exchanges, the features are forwarded in the convolutional and CBAM modules in a feed-forward manner in the three decoders.

3.3. Multitask Loss Function Optimization

To optimize the model, we define a multitask loss function that combines the task-specific losses and assigns weights to control the relative task importance during training while still optimizing everything together. The total loss to minimize is thus defined as:

$$\mathcal{L}_{total} = \lambda_s \mathcal{L}_s + \lambda_b \mathcal{L}_b + \lambda_d \mathcal{L}_d \tag{7}$$

where λ_s, λ_b, λ_d are the three loss weights for the segmentation, boundary, and distance tasks and: $\lambda_s + \lambda_b + \lambda_d = 1$, assuming the three losses have similar ranges. As for the losses: $\mathcal{L}_s$ is the crop type semantic segmentation loss–categorical cross entropy to account for the multiclass nature of the problem; $\mathcal{L}_b$ is the boundary extraction loss which is a binary cross entropy; and $\mathcal{L}_d$ is the distance to boundary loss, mean square error since it is a regression task. After performing empirical studies on the weights of the losses, the weights were set to: $\lambda_s = 0.45$, $\lambda_b = 0.45$, $\lambda_d = 0.1$. This weighting strategy gives greater importance to border detection and semantic segmentation (the two primary tasks), while the distance to boundary task serves as an auxiliary supporting task for the other two.

4. Results

This section presents the experimental setup and the quantitative and qualitative results.

4.1. Experimental Settings

The network uses a channel size of 16, 32, and 64 for the 3D convolutional blocks. The number of blocks L was set to 3, so the network has three encoding layers (shared across tasks), and three decoding layers (per task). The convolutional layer kernels are of size (3, 3, 3). Dilated convolution was used to capture broader details, thus the spatial dimensions of the features is constant throughout the architecture. The decoder feature maps across tasks have the same spatial dimension of the features which is a requirement for the proposed design of the cross-task sharing. Data augmentation included flipping and rotation. The batch size was set to 16 for hardware constraints. For training, an Adam optimizer [40] with an initial learning rate of 1×10^{-4} was used. The experiments were performed using PyTorch, 2.2.0 and the models were trained for a maximum of 50 epochs with an early-stopping criterion based on validation loss on an NVIDIA Ampere A40.

To validate the algorithm, we report results on the two datasets presented in Section 2. They are discussed in detail in Section 4.2. Experiments were performed as follows:

- *Experiment 1. The relevance of the multitask network.* In this experiment, we analyzed the contribution that the multitask learning provided by comparing the performance and efficiency of single and multitask networks.
- *Experiment 2. Analysis of the influence of the three agricultural tasks during their mutual solving.* In this study, we examined how the performance changes when we use different combinations of tasks during training. In this way, we could check the combination that benefits all the tasks and enables their mutual solution. We trained pairs tasks, exploring the three possible combinations (segmentation and boundaries, segmentation and distance, and boundaries and distance), with features exchanged between decoders. In these setups, we used the same network with a shared encoder and two task-specific decoders. The feature sharing in the decoding part was performed between the 3D convolutional decoding blocks. The information cross-task flow happens using the blocks that exchange gated features in the two directions (between the two tasks). We trained a multitask network with all three tasks and with all-to-all feature-sharing. This consisted of designing separate cross-task selective-feature-sharing blocks that share information between each pair of tasks: crop type segmentation and boundary extraction, boundary extraction and distance to boundary estimation, and crop type segmentation and distance to boundary estimation. The sharing blocks still use the gating mechanisms and the convolution operations. Such an all-to-all feature-sharing does not prioritize any task interconnectivity explicitly.
- *Experiment 3. The effect of the weight loss factors on each task.* For this experiment, we varied the relative importance put in each task by changing the loss weights λ_s, λ_b and λ_d, thus controlling the importance of each task during the training of the MTL neural network. The four sets of weights chosen were $\{0.33, 0.33, 0.33\}$, $\{0.40, 0.40, 0.20\}$, $\{0.45, 0.45, 0.10\}$ and $\{0.475, 0.475, 0.05\}$. We also provide the results of a dynamic weighting strategy [41]. The choice was made in order to check the intuition, coming also from a literature review, about the significance of the three tasks and their relative importance. The equal weight variant is a base on which to perform comparison of the other three cases, in which we examined the relative weight of the auxiliary distance to boundary task, w.r.t. the other two main tasks.
- *Experiment 4. Analysis of agricultural task interaction inside the directional feature-sharing blocks.* This experiment focused on the cross-task communication inside the network. We explored how the gating mechanisms behave in the trained model when given test examples in order to assess the interactions. To also provide a quantitative measure of task interaction, subsamples of the test set were used and statistics were computed on cross-task activations. These statistics are the mean, standard deviation, and saturation. The mean represents, on average, how much information the gates keep for a couple of tasks (a higher mean implies that more information is shared for the specific task directional couple); the standard deviation reports the variation across the values, while the saturation reports the fraction of values that fall above a threshold, which was set to 0.8. Extremely high saturation implies that, even when the mean is high and much information is shared between the specific directional couple of tasks, there is little variation in the gate values across test samples, suggesting more trivial solutions that do not depend on the input samples. The results are based on cross-task

directional feature-sharing blocks at the second level of the decoder, where more specific interactions occurred than in the first block.

- *Evaluation w.r.t. segmentation and boundary delineation algorithms.* In this study, we compared the performance of the proposed method with other literature models for single task segmentation and boundary extraction.

Evaluation Metrics for the Tasks

Various evaluation metrics were exploited to validate the algorithm performance across the tasks. For the crop-type semantic segmentation, the main metrics were Cohen's Kappa Score (κ) for multi-class segmentation, as well as the F1 score, precision, and recall for the class-specific metrics.

Regarding the boundary detection task, validation takes into account boundaries that separate one field from another or from non-agricultural land. Due to the spatial resolution of the sensor and the rasterization process during the reference map creation, boundary reference information is affected by uncertainty limited positional accuracy resulting in pixel misalignment. This issue reflects as the proposed model, and as comparison models. In both cases there are some pixel misalignments (e.g., slightly shifted or of different width, either thicker or thinner). To account for this aspect in the performance validation of the boundary detection task, the following quantitative metrics are considered. The F1 score was reported. To further evaluate the performance of the proposed method in border detection, Boundary F1 score (BF1) [42] was used. It gives some tolerance distance when calculating the F1 score (the tolerance set to 1 and to 2 pixels). BF1 is more concerned about evaluating contour proximity rather than filled area agreement. By reporting these metrics, a more comprehensive understanding is achieved of the performance of the boundary delineation results. This allows to mitigate the inherent problems due to the abovementioned uncertainties, while at the same time still quantifying critical errors (like missed boundaries between fields, or falsely identified borders inside one field).

4.2. Results on Field Crop Type Semantic Segmentation, Boundary Extraction and Distance to Boundary Estimation

This section presents the performance of the proposed model on the two datasets.

4.2.1. Austria I

For the Austria I dataset, the proposed method obtained $\kappa = 83.52\%$ for the semantic segmentation of the crop field classes. Class-based performance (Table 2) confirmed the good performance of the model, showing class-specific metrics to be high for the majority of the classes. The method exhibited only some lower performance for minority classes, like Triticale and Small-seeded grass, which are also similar in spectral-temporal pattern to Grassland class. The minority classes imbalance is an issue also for other comparative models for crop type segmentation, as it is an imminent problem with the 17 classes present (see Section 4.7).

For the boundary extraction task, the model achieved a good identification of the borders (F1 = 77.77%). When looking to the Boundary F1 with tolerance of 1 or of 2 pixels, we can see that the borders are well extracted (BF1 (tol.1) = 94.37% and BF1 (tol.2) = 98.74%). This result points out to good boundary detection extraction, while the majority of false positives and false negatives occurred due to the discussed pixel misalignment.

Example qualitative results are shown in Figure 6. From these maps, we saw that the classes are identified well, and boundaries even for thin fields were detected with good accuracy.

Table 2. Class-based performance of the proposed method (Austria I dataset).

Class	Precision	Recall	F1-Score	Support
Legumes	74.39%	74.61%	74.50%	1.56%
Grassland	82.92%	89.41%	86.04%	9.86%
Small-seeded grass	87.86%	42.95%	57.70%	0.92%
Maize	95.15%	87.90%	91.38%	12.20%
Soy	90.87%	90.11%	90.49%	8.22%
Barley	84.35%	88.60%	86.42%	8.93%
Sunflower	85.12%	87.92%	86.50%	5.33%
Potato	87.82%	80.87%	84.20%	2.10%
Wine	86.80%	91.43%	89.05%	5.57%
Rapeseed	75.12%	97.70%	84.94%	1.97%
Beet	90.54%	92.44%	91.48%	4.22%
Cucurbitaceae	71.96%	91.33%	80.50%	2.53%
Triticale	64.60%	22.25%	33.10%	1.07%
Farro	72.00%	30.33%	42.68%	2.43%
Durum Wheat	89.05%	62.13%	73.19%	5.97%
Rye	75.82%	66.25%	70.71%	1.82%
Soft Wheat	82.81%	91.71%	87.03%	25.30%

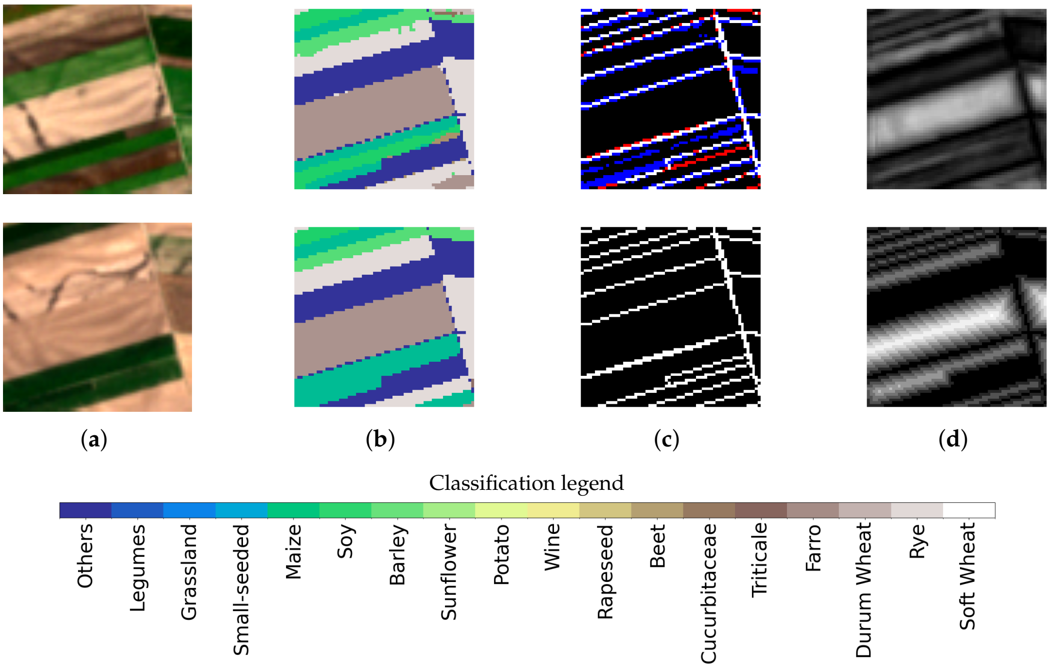


Figure 6. Austria I dataset. In (**a**), we see two example images from the time series, then first row model predictions and second row reference maps for: (**b**) crop type semantic segmentation, (**c**) boundary extraction; first row shows in black—true negatives, in white—true positives, in blue—false positives, and in red—false negatives, and (**d**) distance to boundary estimation.

4.2.2. Austria II

For the second Austria dataset, the performance was similarly strong to the previous dataset for the semantic field crop type segmentation, with $\kappa = 91.25\%$. Table 3 reveals class-specific performance, where most of the classes had high metrics, and the errors came from minority classes. In fact, this dataset presented more class imbalance with respect to the previous one, and that is why major classes had an overall better performance than in Austria I (classes like Maize and Soft Wheat). The imbalanced dataset influenced the minority classes negatively.

For the boundary extraction task for Austria II, the model showed a result similar to Austria I ($F1 = 73.84\%$). The method achieved Boundary F1 scores BF1 (tol.1) = 94.22% and BF1 (tol.2) = 99.07%. These results confirmed that boundaries were overall accurate (Figure 7).

The shapes of the fields in this dataset are not usually as thin as just several pixels, as in the Austria I dataset, but are irregularry oriented, having diverse shapes. Yet, the model managed to identify both the classes and the boundaries well, as can be noted from Figure 7.

Table 3. Class-based performance of the proposed method (Austria II dataset).

Class	Precision	Recall	F1-Score	Support
Legumes	92.92%	86.88%	89.80%	1.02%
Grassland	93.73%	87.23%	90.36%	11.67%
Small-seeded grass	0.00%	0.00%	0.00%	0.06%
Maize	96.48%	97.90%	97.18%	30.58%
Soy	95.75%	95.29%	95.52%	10.28%
Barley	96.92%	94.99%	95.95%	12.95%
Sunflower	77.97%	52.77%	62.94%	0.13%
Potato	63.00%	65.01%	63.99%	0.56%
Wine	0.00%	0.00%	0.00%	0.01%
Rapeseed	94.94%	97.89%	96.39%	2.49%
Beet	94.05%	98.85%	96.39%	2.87%
Cucurbitaceae	85.95%	83.10%	84.50%	1.05%
Triticale	78.66%	70.37%	74.29%	5.01%
Farro	56.90%	42.49%	48.65%	0.97%
Durum Wheat	0.00%	0.00%	0.00%	0.02%
Rye	59.01%	84.54%	69.50%	1.95%
Soft Wheat	92.27%	95.25%	93.74%	18.38%

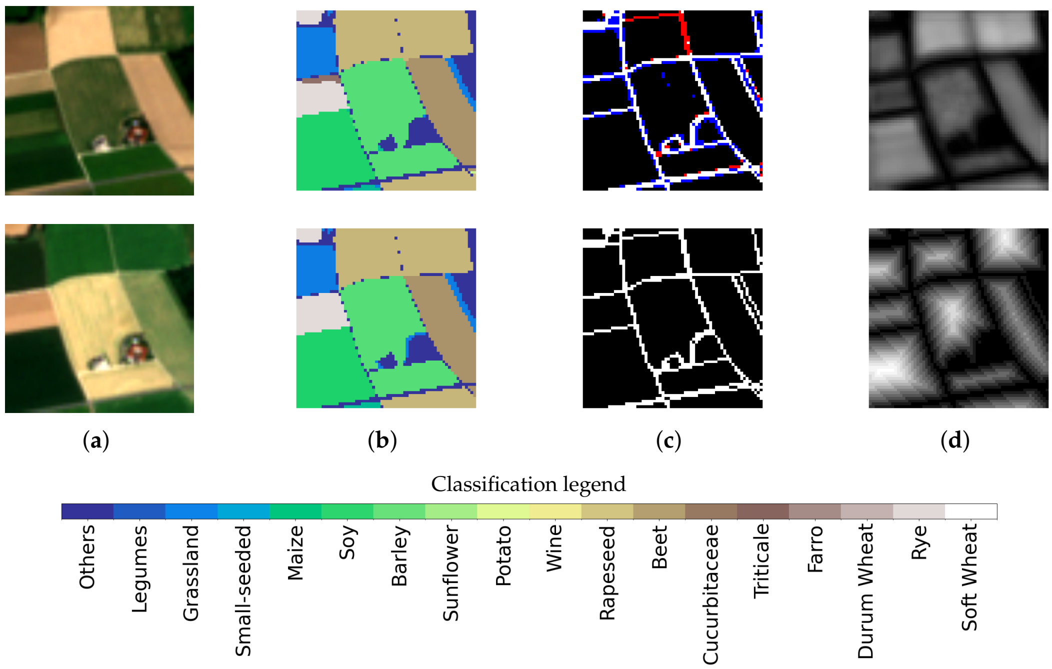


Figure 7. Austria II dataset. In (**a**), we see two example images from the time series, then first row model predictions and second row reference maps for: (**b**) crop type semantic segmentation, (**c**) boundary extraction first row shows in black—true negatives, in white—true positives, in blue—false positives, and in red—false negatives, and (**d**) distance to boundaries estimation.

4.3. Experiment 1. Analysis of the Relevance of the Multitask Framework

This study reports a comparison on the performance of the network when trained and tested on single tasks only (crop type segmentation or boundary extraction) to the proposed multitask model. The single-task (ST) models were trained separately. This focused the attention of the network on only one goal at a time. It is worth noting that the single task architecture shows the same number of parameters as the corresponding branch in the multi-task architecture. This choice allows us to fairly evaluate the added value of the multitask setup with respect to the single task one. Looking at the MTL performance of the proposed model in Table 4, we can note an improvement in both the crop type segmentation task (augment of almost 3% in the κ score), and in the boundary extraction task for Austria I w.r.t. the single tasks. Overall, even if the boundary extraction metric BF1 (tol. 1) is a bit lower in both Austria I and Austria II (Tables 4 and 5), when looking at the F1 score for this task, there is an improvement (of more than 2.5% for Austria I and of around 1.4% for Austria II). At a tolerance of two pixels, the results in all cases were comparable and very high. Therefore, with a multitask network, several agricultural outputs are obtained simultaneously, but also a boost in the performance is achieved. Even if the single task model is more specialized, it does not have the benefits of achieving richer deep representation guided by multiple objectives, positive transfer from feature share.

Table 4. Performance, parameter count and per patch inference time of the single-task models (ST) trained separately on each task and of the proposed method with directional feature-sharing in the decoders (Austria I dataset).

Model	Segmentation κ	Boundary Extraction F1	BF1 (tol.1)	BF1 (tol.2)	Params	Time (ms)
Segmentation	80.44%	-	-	-	242 k	13.52
Boundary	-	75.13%	95.34%	98.70%	242 k	13.47
Proposed	83.03%	77.77%	94.37%	98.74%	599 k	35.81

Table 5. Performance, parameter count and per patch inference time of the single-task models (ST) trained separately on each task and of the proposed method with directional feature-sharing in the decoders (Austria II dataset).

Model	Segmentation κ	Boundary Extraction F1	BF1 (tol.1)	BF1 (tol.2)	Params	Time (ms)
Segmentation	91.38%	-	-	-	242 k	12.85
Boundary	-	72.45%	95.71%	99.19%	242 k	12.68
Proposed	91.25%	73.84%	94.22%	99.07%	599 k	36.17

4.4. Experiment 2. Analysis of the Influence of the Three Agricultural Tasks During Their Mutual Solving

In order to study the task interdependencies, this experiment combined couples of tasks from the three-task problem and their performance was compared. All these setups provided information regarding the relative importance and the possible interconnections. Plots in Figure 8 show how the cross-task sharing is performed in the dual-task setups. The sharing happens using the blocks that exchange gated features in the two directions (between the two tasks and in line with the three-task set up).

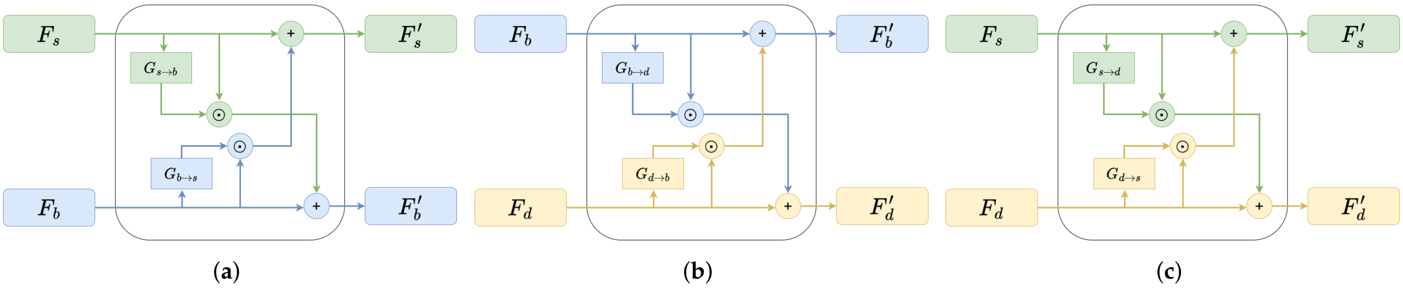


Figure 8. Cross-task selective feature-sharing mechanisms in the dual-task setups: (**a**) semantic segmentation and boundary detection, (**b**) boundary detection and distance estimation, (**c**) semantic segmentation and distance estimation.

Table 6 outlines the results for Austria I. The table clearly shows the improvement of the boundary extraction results when the distance estimation task was added to the analysis, with F1 score growing of around 2% in all the cases including the distance task. On the other hand though, the addition of the distance task was quite harmful for the segmentation of the crop type. The performance of the dual-task model trained only on segmentation and distance provided the worst semantic segmentation performance with κ being 8% lower w.r.t. other cases, and a similar situation was in the case when inter-exchange features happen in an all-to-all manner. This allowed us to deduce that the best type of feature-sharing is between crop type segmentation and boundary extraction, as well as between boundary extraction and distance to boundary estimation, thereby preventing strong negative task interference. The proposed method with directional feature-sharing achieved the best combination and performance.

Table 6. Performance, parameter count and per patch inference time of the dual-task (DT) models trained separately on the various couples of tasks (S—crop type segmentation, B—boundary detection, D—distance estimation), of the multitask models with all-to-all task feature-sharing (MTL a2a), and of the proposed method with directional feature-sharing in the decoders (Austria I dataset).

Model	Segmentation κ	Boundary Extraction F1	BF1 (tol.1)	BF1 (tol.2)	Params	Time (ms)
DT (S+B)	84.97%	75.36%	94.06%	98.51%	420 k	23.35
DT (S+D)	76.31%	-	-	-	420 k	23.57
DT (B+D)	-	77.49%	91.82%	98.06%	420 k	26.48
MTL (a2a)	76.13%	77.06%	94.06%	98.51%	609 k	37.92
Proposed	83.03%	77.77%	94.37%	98.74%	599 k	35.81

Table 7 examines the behaviour of the models in the Austria II dataset. More irregular agricultural field shapes had distances to boundaries higher, and showed higher standard deviation, leading to more uncertainty. We could still infer the strong relationship between the semantic segmentation and boundary extraction tasks. Even if the performance did not show the same oscillations as in the Austria I results (e.g., for the boundary extraction task, the models performed comparably, with F1 scores staying within 1.32 percentage points across setups), we still noted the importance of the proposed directional feature-sharing that achieved a comparable trade-off with the case performing the semantic segmentation and boundary detection tasks and outperformed the all-to-all task feature sharing.

Table 7. Performance, parameter count and per patch inference time of the dual-task (DT) models trained separately on the various couples of tasks (S—crop type segmentation, B—boundary detection, D—distance estimation), of the multitask models with all-to-all task feature-sharing (MTL a2a), and of the proposed method with directional feature-sharing in the decoders (Austria II dataset).

Model	Segmentation	Boundary Extraction			Params	Time
	κ	F1	BF1 (tol. 1)	BF1 (tol. 2)		(ms)
DT (S+B)	92.15%	73.69%	94.66%	99.04%	420 k	23.42
DT (S+D)	89.71%	-	-	-	420 k	23.64
DT (B+D)	-	72.97%	94.32%	99.05%	420 k	23.43
MTL (a2a)	90.60%	74.29%	94.00%	98.99%	609 k	38.14
Proposed	91.25%	73.84%	94.22%	99.07%	599 k	36.17

4.5. Experiment 3. The Effect of the Loss Weight Factors for Each Task

In multitask learning, loss calculation and minimization are crucial steps in designing a well-performing model. Loss weighting schemes are one of the tools to regulate the task importance by controlling the contribution of each task to the total loss during training.

A set of weights to test the model performance was identified. The set of weights was chosen to capture the dynamics and relative importance of tasks. In particular, crop type segmentation and boundary extraction are the two main tasks, while distance to boundary estimation is an auxiliary task. For this reason, the first two tasks were given greater weight than the latter. In this way, negative transfer effect diminished. Empirically, the losses did not differ much in scale during the training process, and thus neither task dominated convergence. In addition, experiments were performed using a dynamic weighting strategy [41]. With this strategy, in the final models the losses were weighted with the following sets of weights: $\{0.26, 0.57, 0.17\}$ (Austria I), and $\{0.33, 0.36, 0.31\}$ (Austria II) for segmentation, boundaries, and distance. The automatic weighting confirms the reasoning that the distance task is auxiliary to the others, giving it less relative weight w.r.t. the other two tasks.

From Table 8, we inferred that the best performing set of weights was $\{0.45, 0.45, 0.10\}$, since it gave the best results on the boundary extraction task across all the metrics, and a comparatively good crop type segmentation score. Decreasing λ_d brought a small improvement to crop type segmentation, but it was harmful to the boundary extraction task. The automatic weighting did not improve the performance compared to the weights chosen. Indeed, as losses have already quite similar scales, there is no one clearly dominating others. Furthermore, still the weighting has given larger weight to the distance task than in our setup, which has resulted in deteriorated segmentation performance.

Table 8. Performance of the proposed method using different loss-weighting schemes (Austria I dataset).

λ_s	λ_b	λ_d	Segmentation	Boundary Extraction		
			κ	F1	BF1 (tol. 1)	BF1 (tol. 2)
0.33	0.33	0.33	80.50%	77.63%	93.09%	98.33%
0.40	0.40	0.20	82.38%	77.17%	93.97%	98.62%
0.475	0.475	0.05	83.52%	77.23%	93.91%	98.58%
0.26	0.57	0.17	81.97%	77.56%	94.43%	98.70%
0.45	0.45	0.10	83.03%	77.77%	94.37%	98.74%

Table 9 reports very similar results to Austria I on the F1 score of the crop type segmentation score for Austria II, where again an increase in the distance to boundary estimation task weight led to a small performance drop. Also in this case, the set of weights $\{0.45, 0.45, 0.10\}$ provided good results. For this reason, this set was kept for our evaluation and visual analysis. However, we would like to stress again that we concentrate here on both crop type segmentation and boundary extraction with equal relevance. Yet, it is possible to adjust the weight values based on proper considerations on the specific tasks. As an example, practitioners who prioritize the crop type segmentation over the boundary extraction would prefer better segmentation results, and to give more importance to this task rather than the boundary extraction task, they could use a loss balanced toward this task, giving it more weight, while still benefiting from multitask learning and feature sharing.

Table 9. Performance of the proposed method using different loss-weighting schemes (Austria II dataset).

λ_s	λ_b	λ_d	Segmentation	Boundary Extraction		
			κ	F1	BF1 (tol. 1)	BF1 (tol. 2)
0.33	0.33	0.33	90.76%	74.09%	94.41%	99.10%
0.40	0.40	0.20	91.26%	73.00%	94.54%	99.10%
0.475	0.475	0.05	91.95%	71.89%	94.52%	99.04%
0.33	0.36	0.31	91.01%	72.48%	94.31%	99.02%
0.45	0.45	0.10	91.25%	73.84%	94.22%	99.07%

4.6. Experiment 4. Analysis of Agricultural Task Interaction Inside the Directional Feature-Sharing Blocks

This experiment examines the cross-task selective feature-sharing blocks and the gating mechanisms. As described in Section 3.2, the defined gates are $G_{s\to b}$, $G_{d\to b}$, $G_{b\to s}$ and $G_{b\to d}$, and we examined their statistics on test examples from the two datasets.

The crop type segmentation and boundary extraction tasks exhibited very strong and consistently high gate activations in both directions with $G_{s\to b}$ and $G_{b\to s}$ having mean activations of 0.8930 and 0.9475 with standard deviations of 0.1869 and 0.1721 for Austria I and 0.7871 and 0.8479 with standard deviations 0.1917 and 0.1851 for Austria II (Figure 9). This has lead to high feature-sharing in both directions. The saturation was high, but there was still some variation across the test samples. As can be seen in Figure 10, in the model there were meaningful interactions by selectively sharing or withholding information, depending on the model input. The same situation is valid for the distance estimation to boundary extraction task interaction which had high mean (0.8981 and 0.7913 for Austria I and Austria II) of the gate values. On the other hand, the gate with the lowest strength is the boundary to distance one. This means that less information is shared in these directional couple of tasks. The results are consistent across the two study areas, and in both cases the strongest gate was $G_{b\to s}$, indicating the importance of the boundary extraction task for the crop type segmentation. This finding indicates that the features for crop field boundaries provide spatial information on the segmentation, connected to the spatial extent of fields. Boundary information guides the segmentation to provide consistent labels inside fields. This strong link suggests the complementarity of the geometrical information for accurate crop-type mapping.

This study provided a deeper understanding of task interactions in the already-trained model and further confirmed the intuition about the paired task coupling: a contribution between the tasks was exhibited and the directional and selective fusion mechanism managed to perform structured task interaction.

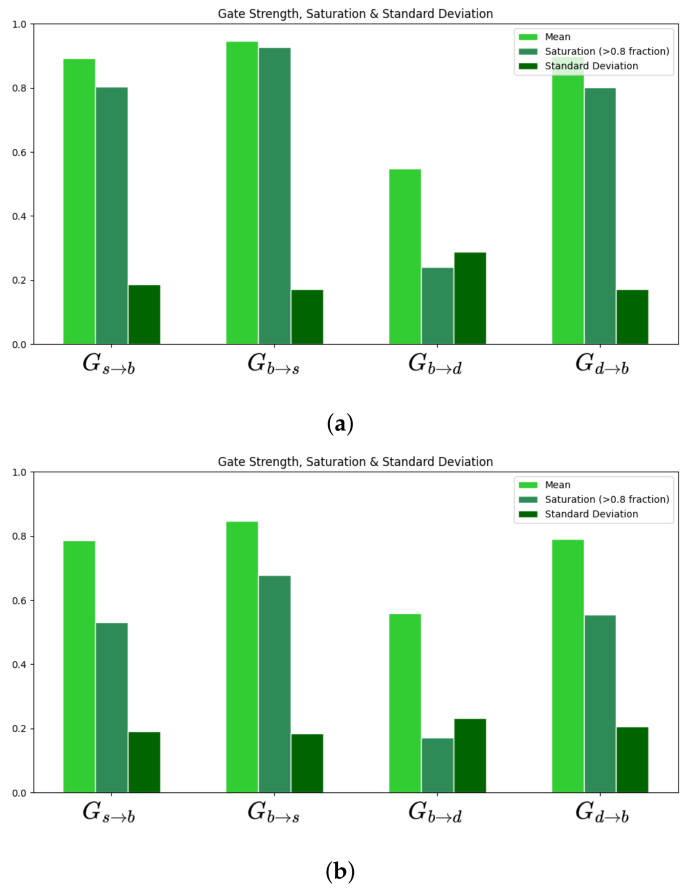


Figure 9. Bar plots showing mean values, saturation (fraction of values above the chosen threshold of 0.8) and standard deviation of the different gates for test subsets of (**a**) Austria I and (**b**) Austria II.

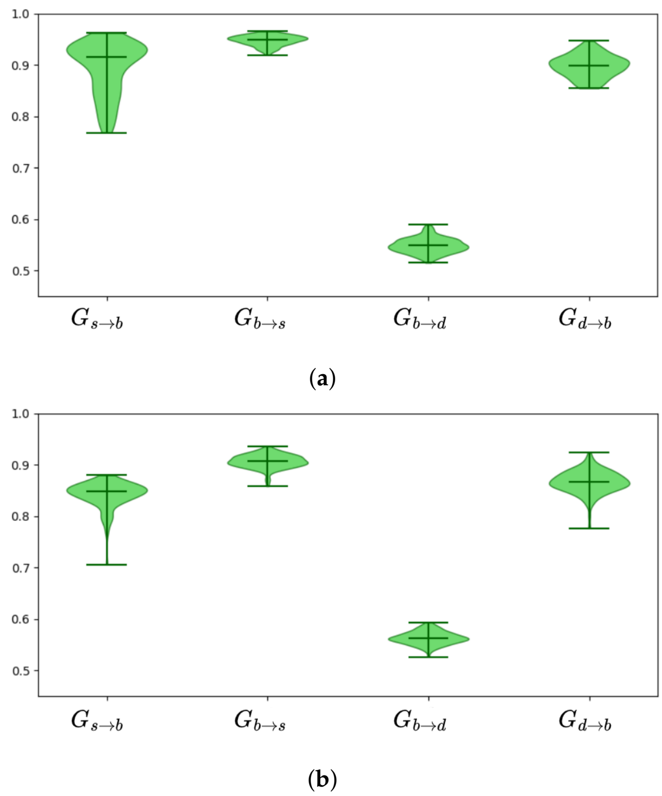


Figure 10. Violin plots showing the distribution of the different gates for test subsets of (**a**) Austria I and (**b**) Austria II.

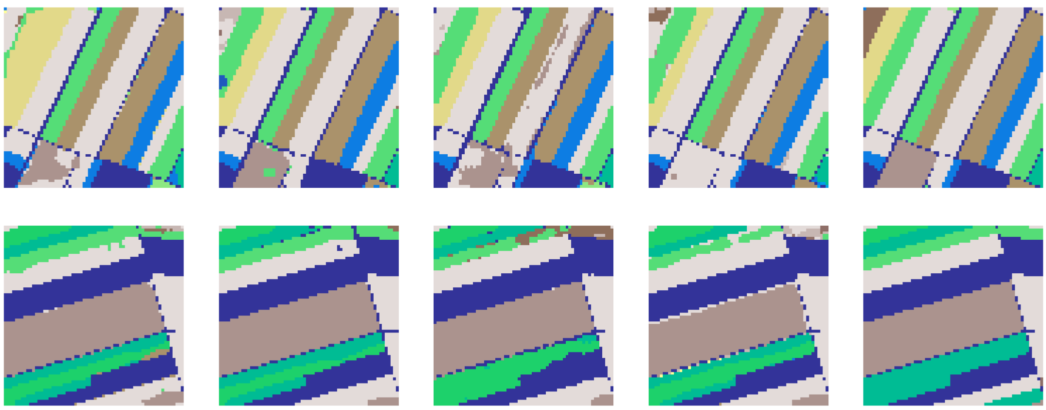
4.7. Evaluation w.r.t. Other Literature Algorithms

We compared the algorithm with methods for the crop type segmentation task, and for the boundary delineation task. For semantic segmentation of crop type, selected methods were U-tae [43] as well as a 3D convolutional Unet model [44], and a SITS mamba model [45] since they are established models for segmentation. For boundary detection, two supervised 3D Unet models (in [24,25]) were used. We also compared the proposed model with another multitask model that studies time series data [26] and outputs boundary extraction, parcel identification (rather than assigning the fields a class label), and distance to boundary estimation. This allowed for a comparison with a multitask method specific to the boundary extraction task.

For the crop-type semantic segmentation task, the proposed model outperformed the others significantly on both datasets, as shown in Table 10. The other models are single-task, and thus failed to capture some of the border details, leading to some errors around the fields (Figure 11). Even though all methods showed confusion of some classes (e.g., rye and farro or maize and soy), the proposed method was more spatially consistent, guided by the additional boundaries task. A clear example is in Figure 12, where in the second row, the biggest field (in brown) in the example patch was classified correctly, while in comparison models, it had some parts misclassified. In fact, the class imbalance was a problem for all the methods, and can be examined by the macro average and weighted average F1 scores per classes. For Austria I, the proposed model showed a macro average of 77.05%, while the weighted average was 84.46%. Similar but still lower performance was seen for UTAE and 3DUnet models (72.01% and 70.18% macro average, and 82.47% and 82.21% weighted average scores). For Austria II, the proposed model had a macro average of 68.19%, and a weighted one of 92.69% (and the comparison models showed worse performance). The discrepancy here is higher than the first dataset due to the more significant class imbalance.

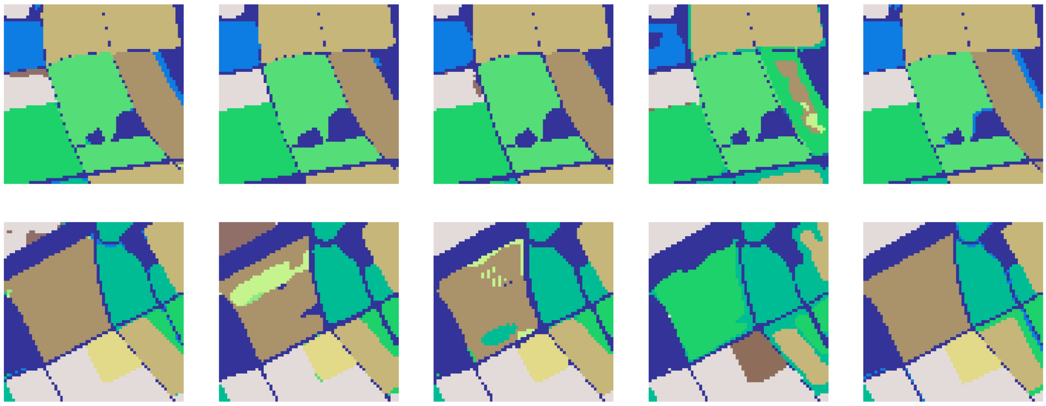
Table 10. Performance of the models on the crop type segmentation task in terms of Cohen Kappa score (Austria I and Austria II datasets).

Model	Austria I	Austria II
3D Unet [44]	76.82%	77.97%
Utae [43]	76.87%	79.51%
SITSMamba [45]	75.42 %	73.61%
Proposed	83.03%	91.25%



(a) (b) (c) (d) (e)

Figure 11. Austria I dataset comparison of results for the crop type segmentation task. Column (**a**) is the proposed model, (**b**) - Utae, (**c**) 3D unet, (**d**) SITS Mamba, and (**e**) reference maps.



(a) (b) (c) (d) (e)

Figure 12. Austria II dataset comparison of results for the crop type segmentation task. Column (**a**) is the proposed model, (**b**) - Utae, (**c**) 3D unet, (**d**) SITS Mamba, and (**d**) reference maps.

For the boundary extraction task, the most direct comparison was with the multitask model [26]. As shown in Tables 11 and 12, this model achieved lower F1 score than the proposed method. This indicates that adding the class type (which is not present in [26]) aids boundary distinction in multitask problems and that the task dependence between segmentation and boundaries is bidirectional. This can be seen in Figures 13 and 14a, from showing less false positive boundaries when the crop type information is present, which further helps divide nearby fields. Instead, in the MTL case where no crop field type is provided, some errors inside specific fields occur (Figures 13 and 14d, suggesting that having only information about the existence of an agricultural parcel is seldom sufficient, especially for densely situated fields. Thus, we could infer the benefits of having semantic information, which helped improve consistency between the semantic crop type and boundary extraction output. Instead, the proposed method differs in F1 score from the single-task literature method (3D Unet and 3D Unet Attn) performance. However, by looking at the BF1 score with tolerance in Tables 11 and 12, it appears that the examined models are comparable when evaluating the algorithm with a small pixel tolerance. Therefore, the models performed similarly. By visually examining some patches in Figures 13 and 14, we can see that the boundaries in both these cases are accurately extracted, fields delineated precisely. The models showed similar wrong predictions, like the false negatives in the Austria II samples (upper bound of the upper field in the first row of Figure 14 and the field on the right in the second row of the same figure).

Table 11. Performance of the models on the boundary extraction task (Austria I dataset).

Model	F1	BF1 (tol. 1)	BF1 (tol. 2)
3D Unet [24]	84.23%	94.80%	98.58%
3D Unet Attn [25]	80.22%	93.77%	98.23%
MTL Efficient Net UNet [26]	68.44%	85.47%	96.33%
Proposed	77.77%	94.37%	98.74%

Table 12. Performance of the models on the boundary extraction task (Austria II dataset).

Model	BF1	BF1 (tol. 1)	BF1 (tol. 2)
3D Unet [24]	82.72%	94.35%	98.81%
3D Unet Attn [25]	83.46%	94.94%	98.99%
MTL Efficient Net UNet [26]	65.75%	85.84%	97.52%
Proposed	73.84%	94.22%	99.07%

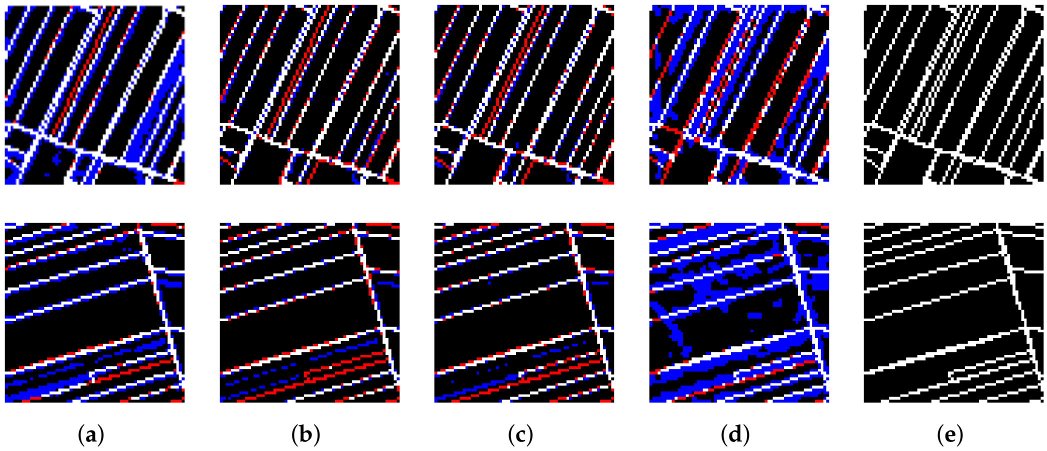


Figure 13. Austria I dataset comparison of results on the boundary extraction task (in black—true negatives, in white—true positives, in blue—false positives, and in red—false negatives). Column (**a**) is the proposed model, (**b**) 3D unet, (**c**) 3D unet with Attention, (**d**) MTL Efficient Net Unet, and (**e**) reference maps.

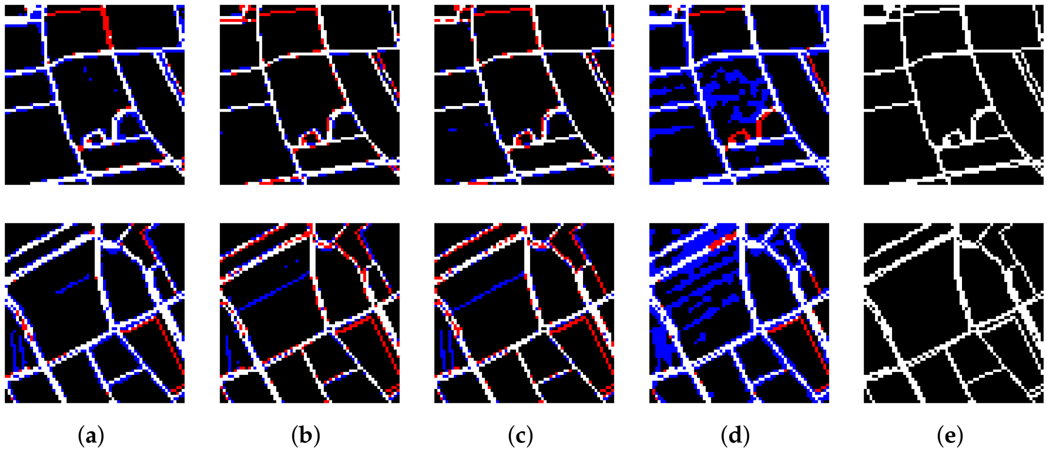


Figure 14. Austria II dataset comparison of results on the boundary extraction task (in black—true negatives, in white—true positives, in blue—false positives, and in red—false negatives). Column (**a**) is the proposed model, (**b**) 3D unet, (**c**) 3D unet with Attention, (**d**) MTL Efficient Net Unet, and (**e**) reference maps.

5. Discussion

The simultaneous solution of related agricultural tasks of crop type semantic segmentation and boundary extraction is a valuable methodology that enables crop field characterization from multiple perspectives. Having access to multiple outputs is a useful tool for downstream applications and end users. In the current literature around boundary extraction, MTL is often used, but in a setting where the classes of the fields are not available. Within this study, crop type semantic information was used in an MTL model, showing benefits. In fact, the positive transfer was not in a unique task-to-task direction, since the boundary extraction was also informative for the crop type segmentation, by providing relevant spatial information.

One of the intriguing aspects of this study, concerned with the design and analysis of task interaction, is the definition of the time series directional cross-task selective feature-sharing blocks. It required an understanding of the underlying tasks to refine an existing general multitask framework for the specific agricultural domain and outline the structural details.

The model still has some limitations. The cross-task selective sharing blocks have been examined to define the best combination, which requires effort as the number of tasks increases. Adding tasks can increase the number of features and parameters, potentially leading to instability during training. A future research direction would be to develop a way to effectively interactively combine features across many tasks, while avoiding complexity explosion. In addition, we acknowledge the fact that the crop field boundaries are not as thin as possible, as noted in the qualitative results. There is an inherent mismatch between the reference definition coming from the vector map and the model output which is pixel-based. Even if this is not an issue when we look for identifying the specific fields and the borders between them, it may be an issue when the application is on an area with smallholder farms. In those cases, it would be reasonable to include very-high-resolution data to better identify small fields and their borders.

This work on agricultural field characterization is not concerned around dealing with the eminent class imbalance in the crop type maps, and thus a limitation is that minority class identification showed lower performance than the rest of the classes. The poor performance on minority classes is acknowledged, as we have a complex 17 class types problem. Looking at macro average and weighted average F1 scores confirms this. The results do state that class imbalance and performance on minority classes is a challenge. Still, it is not a central issue of this work on agricultural parcel characterization with multitask learning. Thus, future directions of research connected to this study, that concern agricultural users, should focus on class-specific phenological patterns and how they influence both crop type segmentation and boundary delineation. Different crop classes with distinct spectral signatures over time would be more distinguishable than those with similar spectral signatures. An investigation is required to determine whether, in these cases, there is the possibility of early crop type identification and field boundary extraction. Our study analyzed a time series of fixed length; however, some cultures may require different length, seasonal timing or frequency of the time series for better (or early) identification. Since the proposed model does not operate on time series of variable length, it presents a possible point of improvement.

Overall, this study showed how we can utilize an existing MTL framework and redesign it in a reasonable and justifiable manner for three tasks in a specific agricultural MTL model. Since the objectives of the study all benefit from multitemporal images, we performed the analysis using a satellite image time-series model, tailoring it to the agricultural objectives and amending the feature-sharing blocks to account for the tasks of interest. By effectively leveraging tools and incorporating knowledge of the nature of the problems to be solved, we constructed a well-performing multitask model for agriculturally oriented crop field description.

6. Conclusions

This paper proposed a way to perform agricultural crop field characterization with Sentinel-2 image time series. A novel 3 task learning approach extracts agricultural field boundaries and classifies field crop types. Solving these tasks simultaneously reduces the burden of a two-stage model that first identifies field borders and only afterward performs object-based classification. Instead, with the proposed method, the combination the two main objectives of crop type segmentation and boundary extraction, as well as an auxiliary distance to boundary estimation, brings mutual benefits and the possibility to have a detailed characterization of the agricultural land. The agriculturally oriented multitask setup provided the benefits of positive task transfer. The tasks interdependencies in this agricultural setting were examined carefully, exploring performance and task interaction. The tasks of crop type segmentation and boundary extraction strongly influence each other.

This is true in both directions: boundary extraction helps semantic segmentation; but also semantics (crop type) is informative for the boundaries detection and it is a stronger relationship than having only the information of the existence of agricultural parcels (without any crop type label). With the directional selective cross task sharing mechanisms, the model managed to take advantage of this interrelatedness.

Performance validation was done on time-series datasets across two agricultural areas in Austria and obtained good results for both main tasks, underscoring the value of multitask learning with remote sensing image time series, especially for agricultural applications. For crop type segmentation task, the proposed time series model managed to achieve elevated performance, and for boundary extraction, the results showed that even if output boundaries are not as thin, the proposed method managed to identify them properly.

As future works, on the methodological part, we will focus on ways of automatic cross-task sharing between tasks. The optimal usage of the auxiliary task of distance is an interesting geometrical problem, which is aso a possibility for better exploration in future work. Furthermore, we plan to expand to other agricultural time-series related tasks, such as those concerning crop yield. Crop yield can be modeled as a regression problem and is influenced by several factors, including crop type; therefore, a multitask learning scheme is a valuable tool for this setting. We will examine the phenology patterns across the different culture classes in the time series for better tracking.

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

Funding: This research received no external funding.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

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

References

1. Yli-Heikkilä, M.; Wittke, S.; Luotamo, M.; Puttonen, E.; Sulkava, M.; Pellikka, P.; Heiskanen, J.; Klami, A. Scalable Crop Yield Prediction with Sentinel-2 Time Series and Temporal Convolutional Network. *Remote Sens.* **2022**, *14*, 4193. [CrossRef]
2. Chang, Z.; Li, H.; Chen, D.; Liu, Y.; Zou, C.; Chen, J.; Han, W.; Liu, S.; Zhang, N. Crop Type Identification Using High-Resolution Remote Sensing Images Based on an Improved DeepLabV3+ Network. *Remote Sens.* **2023**, *15*, 5088. [CrossRef]
3. North, H.C.; Pairman, D.; Belliss, S.E. Boundary Delineation of Agricultural Fields in Multitemporal Satellite Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 237–251. [CrossRef]
4. Trung, T.H.; Ky, N.V.; Phan, D.C.; Minh, D.B.; Nguyen, H.; Nasahara, K.N. First Agriculture Land Use Map in Vietnam Using an Adaptive Weighted Combined Loss Function for UNET++. *Remote Sens.* **2026**, *18*, 430. [CrossRef]
5. Miller, L.; Pelletier, C.; Webb, G.I. Deep Learning for Satellite Image Time-Series Analysis: A review. *IEEE Geosci. Remote Sens. Mag.* **2024**, *12*, 81–124. [CrossRef]
6. Pelletier, C.; Webb, G.I.; Petitjean, F. Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series. *Remote Sens.* **2019**, *11*, 523. [CrossRef]
7. Ji, S.; Zhang, C.; Xu, A.; Shi, Y.; Duan, Y. 3D Convolutional Neural Networks for Crop Classification with Multi-Temporal Remote Sensing Images. *Remote Sens.* **2018**, *10*, 75. [CrossRef]
8. Khan, A.H.; Zafar, Z.; Shahzad, M.; Berns, K.; Fraz, M.M. Crop Type Classification using Multi-temporal Sentinel-2 Satellite Imagery: A Deep Semantic Segmentation Approach. In Proceedings of the 2023 International Conference on Robotics and Automation in Industry (ICRAI), Peshawar, Pakistan, 3–5 March 2023; pp. 1–6. [CrossRef]
9. Zhao, Z.; Liu, Z.; Luo, H.; Yang, H.; Wang, B.; Jiang, Y.; Liu, Y.; Wu, Y. GF-2 Remote Sensing-Based Winter Wheat Extraction With Multitask Learning Vision Transformer. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 12454–12469. [CrossRef]

10. Lin, Z.; Zhong, R.; Xiong, X.; Guo, C.; Xu, J.; Zhu, Y.; Xu, J.; Ying, Y.; Ting, K.C.; Huang, J.; et al. Large-Scale Rice Mapping Using Multi-Task Spatiotemporal Deep Learning and Sentinel-1 SAR Time Series. *Remote Sens.* **2022**, *14*, 699. [CrossRef]
11. Zimmer-Dauphinee, J.; Wernke, S.A. Semantic Segmentation and Classification of Active and Abandoned Agricultural Fields through Deep Learning in the Southern Peruvian Andes. *Remote Sens.* **2024**, *16*, 3546. [CrossRef]
12. Werner, J.P.S.; Belgiu, M.; Bueno, I.T.; Dos Reis, A.A.; Toro, A.P.S.G.D.; Antunes, J.F.G.; Stein, A.; Lamparelli, R.A.C.; Magalhães, P.S.G.; Coutinho, A.C.; et al. Mapping Integrated Crop–Livestock Systems Using Fused Sentinel-2 and PlanetScope Time Series and Deep Learning. *Remote Sens.* **2024**, *16*, 1421. [CrossRef]
13. Kuchler, P.C.; Simões, M.; Ferraz, R.; Arvor, D.; Machado, P.L.O.D.A.; Rosa, M.; Gaetano, R.; Bégué, A. Monitoring Complex Integrated Crop–Livestock Systems at Regional Scale in Brazil: A Big Earth Observation Data Approach. *Remote Sens.* **2022**, *14*, 1648. [CrossRef]
14. Xu, W.; Wang, J.; Wang, C.; Li, Z.; Zhang, J.; Su, H.; Wu, S. Utilizing Dual-Stream Encoding and Transformer for Boundary-Aware Agricultural Parcel Extraction in Remote Sensing Images. *Remote Sens.* **2024**, *16*, 2637. [CrossRef]
15. Celik, M.F.; Persello, C.; Nelson, A.; Conchedda, G.; Tubiello, F.; Paris, C. Edge-Aware Crop Field Boundary Mapping From Multimodal Sentinel-1 and Sentinel-2 Data. *IEEE Geosci. Remote Sens. Lett.* **2026**, *23*, 2502105. [CrossRef]
16. García-Pedrero, A.; Gonzalo-Martín, C.; Lillo-Saavedra, M. A machine learning approach for agricultural parcel delineation through agglomerative segmentation. *Int. J. Remote Sens.* **2017**, *38*, 1809–1819. [CrossRef]
17. Teng, F.; Wu, L.; Liu, S. Extraction of Agricultural Parcels Using Vector Contour Segmentation Network with Hybrid Backbone and Multiscale Edge Feature Extraction. *Remote Sens.* **2025**, *17*, 2556. [CrossRef]
18. Waldner, F.; Diakogiannis, F.I. Deep learning on edge: Extracting field boundaries from satellite images with a convolutional neural network. *Remote Sens. Environ.* **2020**, *245*, 111741. [CrossRef]
19. Solano-Correa, Y.T.; Bovolo, F.; Bruzzone, L.; Fernández-Prieto, D. A Method for the Analysis of Small Crop Fields in Sentinel-2 Dense Time Series. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 2150–2164. [CrossRef]
20. Long, J.; Li, M.; Wang, X.; Stein, A. Delineation of agricultural fields using multi-task BsiNet from high-resolution satellite images. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *112*, 102871. [CrossRef]
21. Li, M.; Long, J.; Stein, A.; Wang, X. Using a semantic edge-aware multi-task neural network to delineate agricultural parcels from remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2023**, *200*, 24–40. [CrossRef]
22. Luo, W.; Zhang, C.; Li, Y.; Yan, Y. MLGNet: Multi-Task Learning Network with Attention-Guided Mechanism for Segmenting Agricultural Fields. *Remote Sens.* **2023**, *15*, 3934. [CrossRef]
23. Persello, C.; Tolpekin, V.; Bergado, J.; De By, R. Delineation of agricultural fields in smallholder farms from satellite images using fully convolutional networks and combinatorial grouping. *Remote Sens. Environ.* **2019**, *231*, 111253. [CrossRef] [PubMed]
24. Meshkini, K.; Doktor, D.; Bovolo, F. Crop Field Boundary Detection Using 3d Convolutions in Multi-Spectral Multi-Temporal Hr Satellite Images. In Proceedings of the IGARSS 2024—2024 IEEE International Geoscience and Remote Sensing Symposium, Athens, Greece, 7–12 July 2024; pp. 11486–11490. [CrossRef]
25. Meshkini, K.; Bovolo, F.; Doktor, D. Attention-based 3D convolutional neural network for crop boundary detection in high-resolution satellite image time series. In *Proceedings of the Artificial Intelligence and Image and Signal Processing for Remote Sensing XXX, Edinburgh, UK, 16–18 September 2024*; Bruzzone, L., Bovolo, F., Eds.; SPIE Digital Library: Bellingham, WA, USA, 2024. [CrossRef]
26. Chaudhury, B.; Sahadevan, A.S.; Mitra, P. Agricultural Field Boundary Delineation From Multi-Temporal IRS P-6 LISS IV Images Using Multi-Task Learning. In Proceedings of the 2023 International Conference on Machine Intelligence for GeoAnalytics and Remote Sensing (MIGARS), Hyderabad, India, 27–29 January 2023; pp. 1–4. [CrossRef]
27. Obialero, M.; Boccardo, P. Segmentation of Multitemporal PlanetScope Data to Improve the Land Parcel Identification System (LPIS). *Remote Sens.* **2025**, *17*, 1962. [CrossRef]
28. Zheng, J.; Ye, Z.; Wen, Y.; Huang, J.; Zhang, Z.; Li, Q.; Hu, Q.; Xu, B.; Zhao, L.; Fu, H. A Comprehensive Review of Agricultural Parcel and Boundary Delineation From Remote Sensing Images: Recent progress and future perspectives. *IEEE Geosci. Remote Sens. Mag.* **2026**, *14*, 2–33. [CrossRef]
29. Caruana, R. Multitask Learning. *Mach. Learn.* **1997**, *28*, 41–75. [CrossRef]
30. Malhotra, A.; Vatsa, M.; Singh, R. Dropped Scheduled Task: Mitigating Negative Transfer in Multi-task Learning using Dynamic Task Dropping. *Trans. Mach. Learn. Res.* **2023**, *2023*, 376.
31. Wang, H.; Dai, Y.; Yao, Q.; Ma, L.; Zhang, Z.; Lv, X. Multi-task learning model driven by climate and remote sensing data collaboration for mid-season cotton yield prediction. *Field Crop. Res.* **2025**, *333*, 110070. [CrossRef]
32. Chang, C.H.; Lin, J.; Chang, J.W.; Huang, Y.S.; Lai, M.H.; Chang, Y.J. Hybrid Deep Neural Networks with Multi-Tasking for Rice Yield Prediction Using Remote Sensing Data. *Agriculture* **2024**, *14*, 513. [CrossRef]
33. Zhu, Y.; Pan, Y.; Hu, T.; Zhang, D.; Dai, J. An Integrated Sample-Free Method for Agricultural Field Delineation From High-Resolution Remote Sensing Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 13112–13134. [CrossRef]

34. Xia, X.; Persello, C.; Koeva, M. Deep Fully Convolutional Networks for Cadastral Boundary Detection from UAV Images. *Remote Sens.* **2019**, *11*, 1725. [CrossRef]
35. Atanasova, M.; Bergamasco, L.; Bovolo, F. A Multitask Framework with Cross-task Selective Feature Sharing for Remote Sensing Image Time Series Analysis. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2026**, *19*, 10308–10323. [CrossRef]
36. Agrarmarkt Austria (AMA). Invekos Schläge Österreich 2021. 2021. Available online: https://www.data.gv.at/datasets/89e7d5e0-385f-40b2-9e42-fb3bb29e52a9 (accessed on 6 August 2026).
37. Weikmann, G.; Paris, C.; Bruzzone, L. TimeSen2Crop: A Million Labeled Samples Dataset of Sentinel 2 Image Time Series for Crop-Type Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 4699–4708. [CrossRef]
38. Weikmann, G.; Marinelli, D.; Paris, C.; Migdall, S.; Gleisberg, E.; Appel, F.; Bach, H.; Dowling, J.; Bruzzone, L. Multiyear Mapping of Water Demand at Crop Level: An End-to-End Workflow Based on High-Resolution Crop Type Maps and Meteorological Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 6758–6775. [CrossRef]
39. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
40. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
41. Kendall, A.; Gal, Y.; Cipolla, R. Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics. *arXiv* **2018**, arXiv:1705.07115. [CrossRef]
42. Martin, D.; Fowlkes, C.; Malik, J. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 530–549. [CrossRef] [PubMed]
43. Fare Garnot, V.S.; Landrieu, L. Panoptic Segmentation of Satellite Image Time Series with Convolutional Temporal Attention Networks. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 11–17 October 2021; pp. 4852–4861. [CrossRef]
44. Rustowicz, R.M.; Cheong, R.; Wang, L.; Ermon, S.; Burke, M.; Lobell, D. Semantic Segmentation of Crop Type in Africa: A Novel Dataset and Analysis of Deep Learning Methods. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Long Beach, CA, USA, 16–20 June 2019; pp. 75–82.
45. Qin, X.; Su, X.; Zhang, L. SITSMamba for Crop Classification based on Satellite Image Time Series. In Proceedings of the IGARSS 2025—2025 IEEE International Geoscience and Remote Sensing Symposium, Brisbane, Australia, 3–8 August 2025; pp. 7096–7099. [CrossRef]