POLITECNICO DI TORINO
Repository ISTITUZIONALE

KAN You See It? KANs and Sentinel for Effective and Explainable Crop Field Segmentation

Original
KAN You See It? KANs and Sentinel for Effective and Explainable Crop Field Segmentation / Rege Cambrin, D., Poeta, E., Pastor, E., Cerquitelli, T., Baralis, E., Garza, P.. - 15625:(2025), pp. 115-131. (European Conference on Computer Vision Milan (ITA) September 29–October 4, 2024) [10.1007/978-3-031-91835-3_8].

Availability:
This version is available at: 11583/2992545 since: 2025-07-07T08:49:59Z

Publisher:
Springer

Published
DOI:10.1007/978-3-031-91835-3_8

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
Springer postprint/Author's Accepted Manuscript

This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: http://dx.doi.org/10.1007/978-3-031-91835-3_8

(Article begins on next page)

18 November 2026

KAN You See It? KANs and Sentinel for Effective and Explainable Crop Field Segmentation

Daniele Rege Cambrin[1], Eleonora Poeta[1], Eliana Pastor[1], Tania Cerquitelli[1], Elena Baralis[1], and Paolo Garza[1]

Politecnico di Torino, Turin, Italy
{daniele.regecambrin,eleonora.poeta,eliana.pastor,
tania.cerquitelli,elena.baralis,paolo.garza}@polito.it

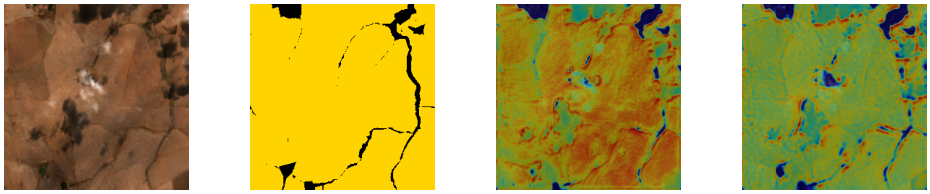
Abstract. Segmentation of crop fields is essential for enhancing agricultural productivity, monitoring crop health, and promoting sustainable practices. Deep learning models adopted for this task must ensure accurate and reliable predictions to avoid economic losses and environmental impact. The newly proposed Kolmogorov-Arnold networks (KANs) offer promising advancements in the performance of neural networks. This paper analyzes the integration of KAN layers into the U-Net architecture (U-KAN) to segment crop fields using Sentinel-2 and Sentinel-1 satellite images and provides an analysis of the performance and explainability of these networks. U-KAN performs comparable to or better than the full-convolutional U-Net in half of the GFLOPs. Furthermore, gradient-based explanation techniques show that U-KAN predictions are highly plausible and that the network has a very high ability to focus on the boundaries of cultivated areas rather than on the areas themselves. The per-channel relevance analysis also reveals which channels are critical for explaining model behavior and which have little to no impact, thus providing insights into the features the model relies on for the segmentation task.

Keywords: Remote Sensing · Explainable AI · Crop Field Segmentation

1 Introduction

In recent years, remote sensing and deep neural networks have revolutionized how we approach agricultural management, environmental monitoring, and many earth-observation-related tasks. Their combination proved to be effective in a wide range of tasks, such as emergency management [19] and land cover [31]. One task related to land cover is the segmentation of crop fields, which is crucial for optimizing agricultural productivity, assessing crop health, and planning sustainable farming practices [7].

The accuracy and interpretability of the neural networks used in this process are fundamental to ensure reliable and actionable insights. Accurate segmentation of crop fields enables precise calculations of area coverage, assessment of



(a) RGB Image **(b)** Ground Truth **(c)** U-Net **(d)** U-KAN

Fig. 1: Example image from the South Africa Crop Type dataset [45], Sentinel-2. (a) displays the image from Sentinel-2 in RGB, and (b) shows the corresponding ground truth, with crop field areas for segmentation highlighted in yellow. (c) and (d) present the saliency maps generated by U-Net and U-KAN, respectively, where red pixels indicate the areas of highest network focus.

crop types, and monitoring of agronomic factors such as plant health and soil conditions [6]. This information is critical for making informed decisions on irrigation, fertilization, and crop rotation, which are essential for enhancing yield and sustainability [10]. Furthermore, accuracy in semantic segmentation tasks directly influences economic planning and policy-making at various governmental and institutional levels. Providing accurate decisions is essential, yet model understandability and accessibility are also vital to allow practitioners to validate them and adhere to institutional regulations [44]. These factors are crucial as they significantly impact both economies and the environment [15]. Deep learning models can achieve high accuracy, but they are often considered "black boxes" due to their intricate architectures composed of numerous layers and parameters that are difficult to interpret. This complexity poses significant challenges in understanding the decision-making processes underlying these models. In the context of remote sensing, the interpretability of such models is further complicated by the nature of the data, which includes various spectral bands, temporal sequences, and spatial resolutions. Additionally, factors such as noise, occlusions, and atmospheric effects can obscure the models' decision-making processes.

Consequently, explainability in deep learning for remote sensing is crucial, as it ensures that humans can understand the decisions and outputs of these models. Developing techniques to elucidate the logic behind the model outputs is essential for validating its results and establishing confidence in its practical applications. A widely adopted approach to enable model explainability is to provide explanations for individual predictions of a model in a post-hoc fashion, allowing its interpretability while not affecting its accuracy. This solution finds application in the domain of earth observation, where explanations are presented as saliency maps (or heat maps), highlighting which parts of the satellite image influence the model prediction [15, 18] (see Figure 1 (c) and (d) as examples).

The recent introduction of Kolgomorov-Arnold Networks (KANs) [23] posed a new paradigm for neural networks as an alternative to Multi-Layer Perceptrons (MLPs). Inspired by the Kolmogorov-Arnold representation theorem [2,20], KANs allow learning custom activations of the edge of the network. In this way,

it is possible to analyze the contribution of individual components of the input data, thus providing a more transparent view of the network's decision-making process. Considering their potential in improving vision tasks, KANs were recently integrated [22] into the U-Net architecture [35], which is a well-known segmentation architecture. The resulting network, denoted as U-KAN, was tested for medical image segmentation and demonstrated superior accuracy and efficiency. In this paper, we are the first to explore the adoption of U-KAN for crop field segmentation, analyzing U-KAN and U-Net from the performance and explainability perspectives. We consider the following research questions (RQs): **RQ1.** How does U-KAN perform compared to U-Net for the task of crop field segmentation? **RQ2.** Which parts of the satellite image influence the models' predictions the most? Do U-Net and U-KAN prioritize different aspects of the image?

To address the first research question, we evaluate U-KAN and U-Net on the South Africa Crop Type dataset [45] for crop field segmentation on Sentinel-2 [14] and Sentinel-1 [43] satellite images. Our findings indicate that U-KANs are more accurate and efficient than U-Nets regarding Intersection-Over-Union (IoU) and Giga Floating Point Operations Per Second (GFLOPs), respectively.

For the second research question, we analyze the explainability of both models. We leverage post-hoc explainability techniques to identify which parts of an image influence the identification of the model. We compare these importance scores, denoted as saliency maps, for U-Net and U-KAN to analyze differences in the identification behavior. Our results reveal that U-Net and U-KAN indeed consider different aspects of the image for predictions. U-KANs generally focus on the edges of the crops, while U-Nets concentrate more on the internals, as exemplified in Figure 1 (c) and (d). Moreover, we quantitatively evaluate the quality of the saliency maps, revealing that U-KANs are more faithful and plausible. Our contributions can be summarized as follows:

- We are the first to explore the application of U-KAN for crop field segmentation.
- We perform a comparative analysis of U-KAN and U-Net for the crop field segmentation on satellite images.
- We analyze the explainability of U-KAN and U-Net, utilizing post-hoc explainability techniques to generate and evaluate saliency maps.
- We reveal that U-KAN offers superior accuracy and efficiency, as well as more faithful and plausible saliency maps compared to U-Net.

The code to reproduce the experiment is available at `https://github.com/DarthReca/crop-field-segmentation-ukan`.

2 Related Work

In this section, we provide an overview of the advancements in remote sensing for agriculture, the explainability of neural networks, and their intersection.

2.1 Remote Sensing

Remote sensing technologies have been extensively applied in agriculture to enhance crop monitoring, management, and productivity. Early studies focused on utilizing satellite imagery to assess crop health and estimate yields [5]. With advancements in sensor technology and data processing techniques, the resolution and accuracy of remote sensing data have significantly improved, enabling more detailed analysis of agricultural landscapes [33]. One application of remote sensing in agriculture is crop field segmentation, which involves identifying cultivated areas. The introduction of Convolutional Neural Networks (CNNs) and the U-Net architecture has further enhanced crop field segmentation [4, 47]. Although recent advancements have proposed other architectures, it still remains one of the most effective baselines in remote sensing, thanks to its design [12]. Integrating multispectral and hyperspectral imaging has also contributed to more accurate crop field segmentation. These images capture data across various wavelengths, providing richer information about crop characteristics [42].

2.2 Explainable AI

Explainable artificial intelligence (XAI) is a branch of AI research dedicated to making machine learning models interpretable and understandable to humans [1, 3, 32]. In recent years, there has been a significant interest in applying XAI techniques to Earth observation tasks, driven by the necessity to interpret complex AI models applied in remote sensing [15,18]. Solutions in this domain follow a standard categorization of XAI approaches: interpretable by design and post-hoc explainability methods [32]. By-design approaches, such as [21, 28, 29, 39], integrate interpretability intrinsically into the design of the model algorithm or its architecture. However, they often fail to explain individual model predictions, raising doubts about their actual ability to help humans understand the process [15]. Additionally, these approaches tend to be less accurate than black-box models. To address these limitations, many works focus on post-hoc explanations [16–18,26,41], which aim to explain trained black-box models while preserving their accuracy and enhancing transparency.

Saliency maps are one of the most adopted post-hoc methods for visualizing which parts of an input image influence the model's prediction. Saliency maps (or heat maps) are overlayed pixel-based importance scores over the input image, highlighting how much each pixel contributes to the prediction. These maps have been widely adopted for semantic segmentation in tasks such as medical diagnosis [16–18, 26]. The urgency of understanding the decision process of the models in remote sensing has led to works investigating their application to segment satellite imagery and agricultural fields [18].

Among these works, Kakogeorgiou and Karantzalos [18] proposed a systematic evaluation of saliency maps for multi-label deep learning classification tasks in remote sensing. Their work tested ten explainable AI methods for deriving saliency maps and proposed a systematic analysis from both qualitative and quantitative perspectives. Our approach aligns with this systematic evaluation;

however, rather than comparing multiple XAI approaches to explain the same model, we systematically evaluate explanations by comparing two models using the same explainability technique to derive saliency maps. Their study identified Grad-CAM [37] as a reliable and interpretable method while being less computationally expensive [18]. We leverage the result of this analysis, and we adopt Grad-CAM as the explainability method.

In line with the demand for interpretability, the recently proposed KANs [23] offer their own level of interpretability by allowing interactions with the network through pruning and by visualizing the learnable activation functions. The work [22] demonstrated the integration of KANs into the U-NET architecture [35], enhancing performance and efficiency for medical diagnosis tasks. Our study is the first to apply U-KANs for crop field segmentation. Furthermore, we compare the U-NET and U-KAN architectures from an explainability perspective, proposing a systematic evaluation of their explanations. To the best of our knowledge, we are also the first to evaluate U-KAN for the post-hoc explainability perspective, analyzing the individual explanations provided by Grad-CAM.

3 Methodology

In this section, we formalize the problem by first detailing the crop field segmentation task, followed by the explainability part, and finally, we present the models.

3.1 Problem statement

This work addresses the crop field segmentation problem based on radiometric or hyperspectral images. The problem can be formulated as follows: Let I be an arbitrary satellite image of size $W \times H \times D$, where W and H are the width and height of the images in pixels, respectively, while D is the depth of the images (i.e., the number of features per pixel). The objective is to automatically create a binary mask M represented by a matrix of size $W \times H$ associated with I, where the value 1 in a cell indicates the associated pixel contains cultivated area. In contrast, 0 is related to any non-cultivated area.

Explainability Statement We aim to help users fully understand how the model achieves effective segmentation by providing them with visual explanations of model predictions. From an XAI perspective, the problem can be formulated as follows: Given an image I and its binary mask M, we want to produce a saliency map (or heat map) S of size $W \times H$ to highlight the regions of I that are important for the model's prediction. Each element s_i of S is the importance score associated with pixel p_i of image I. Each value s_i represents the influence p_i on the prediction of cultivated area. Visualizing the saliency map eases the interpretation of how the model arrives at its decisions.

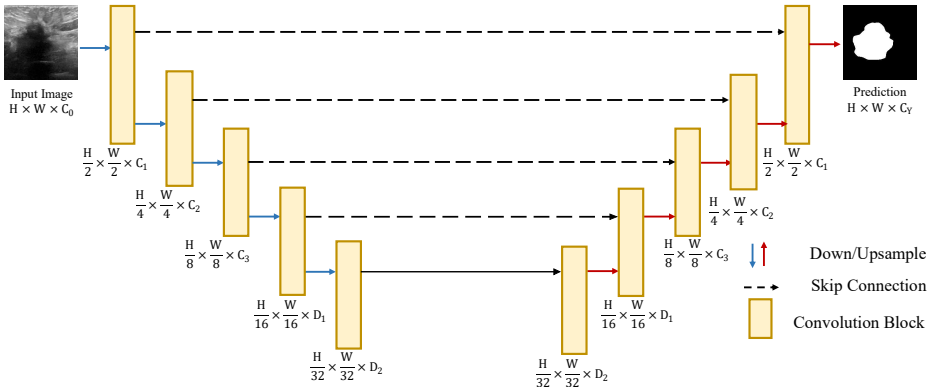


Fig. 2: U-Net architecture [35]. It is characterized by a U-shape structure with a contracting path (left side) to capture context and a symmetric expanding path (right side) to enable precise localization.

3.2 Models

In this study, we compare the well-known U-Net [35] with a modified version [22] which integrates KAN [23] layers into the architecture. In the following, we first outline the U-Net architecture. We then outline the KAN neural networks and, finally, its integration into the U-KAN architecture.

U-Net U-Net is a convolutional neural network architecture designed primarily for biomedical image segmentation [35]. Its architecture is characterized by a U-shaped structure (as seen in Figure 2), with a contracting path to capture context and a symmetric expanding path to enable precise localization. The contracting path consists of repeated application of convolutions followed by max-pooling operations, while the expanding path involves upsampling and convolutional layers to reconstruct the image resolution. This design allows U-Net to effectively learn from relatively few training images and produce high-quality segmentations, making it a popular choice for segmentation beyond medical imaging.

KAN Kolmogorov–Arnold Networks [23] (KANs) are a novel type of neural network inspired by the Kolmogorov-Arnold representation theorem [2, 20], which states that every multivariate continuous function $f : [0, 1]^n \to \mathbb{R}$ can be represented as a superposition of the two-argument addition of continuous functions of one variable:

$$f(x) = f(x_1, x_2, ..., x_n) = \sum_{q=0}^{2n} \Phi_q \Big(\sum_{p=1}^{n} \phi_{q,p}(x_p) \Big) \tag{1}$$

where $\phi_{q,p} : [0, 1] \to \mathbb{R}$ and $\Phi_q : \mathbb{R} \to \mathbb{R}$. Unlike traditional Multi-Layer Perceptrons (MLPs) that have fixed activation functions on nodes, KANs employ

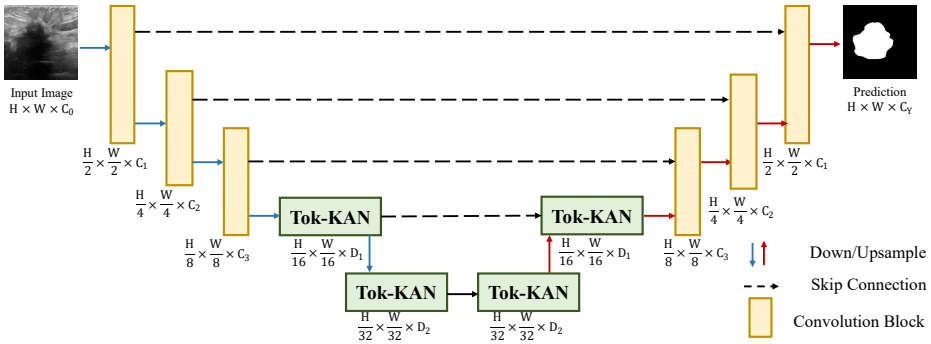


Fig. 3: U-KAN [22]. It has the same U-shape as U-Net, but the deepest layers are implemented as Tok-KAN blocks. Tok-KAN blocks are composed of a Tokenization layer, a KAN layer, a Downsampling layer, and a Normalization layer.

learnable activation functions on edges. This is achieved by replacing every linear weight parameter with a univariate function parameterized as a spline. The activations change step-by-step to better approximate the desired target during the training, and KANs offer the possibility of visualizing the learnable activation functions. In this way, KANs can be more transparent and efficient in learning more complex relations than MLPs, offering promising alternatives to traditional deep learning models. The learned activations can be low-cost functions (e.g., constant or linear) when there is no need for complex non-linearities. This also grants the possibility of understanding the salient parts of the input.

U-KAN U-KAN [22] proposes to implement the deepest layer of the U-Net using KANs. These layers are composed of a tokenization layer, a KAN layer, a downsampling layer, and a final normalization layer. In Figure 3, it is possible to see the key difference stands only in how the deepest representations are processed by the network. The main features of the U-Net, such as downsampling and skip-connection, remain invariant, sharing the same benefits. The modification in the encoder's last layers and the decoder's firsts allows the network to learn custom activation functions instead of fixed ones, potentially improving the representativity of the embeddings and reducing the required computational resources by learning simple activations when needed.

4 Experimental Setup

This section describes the adopted dataset, the experimental setting, and the adopted evaluation metrics for the crop segmentation performance and quality of the derived explanations.

4.1 Dataset

We employed the South Africa Crop Type dataset [45], which contains images taken from both Sentinel-2 and Sentinel-1 covering a wide region of South Africa. The dataset includes small and irregular-shaped [25] crop fields, making it more challenging, and provides higher resolution imagery (of size 256×256) than other datasets covering the area. The annotations contain the mask of the areas covered by a certain crop. For our analysis, we limit the scope to distinguish between cultivated and non-cultivated areas. We analyzed the results using both types of imagery from Sentinel-2 and Sentinel-1. Sentinel-1 [43] is a satellite mission under the Copernicus program, comprising two identical satellites equipped with C-band Synthetic Aperture Radar (SAR). It provides all-weather, day-and-night radar imaging. The satellite can work in both single-polarization and double-polarization modes. On land, it mainly works collecting VV and VH polarizations.Sentinel-2 [14] is part of the Copernicus program and consists of two satellites. These satellites carry high-resolution multispectral imaging instruments with 13 spectral bands ranging from Ultra-Blue, Visible, Near Infrared (NIR), and Short Wave infrared (SWIR). It is particularly sensitive to vegetation due to the presence of instruments that work in the infrared spectrum.

While Sentinel-1 images can cover different atmospheric situations due to the radiometric nature of the imagery, Sentinel-2 is affected by clouds and similar atmospheric disturbances. Since the provided cloud masks are often not accurate, we computed the masks with the s2cloudless algorithm [38]. We exclude low-quality Sentinel-2 images with a strong overlap between the cloud mask and the areas containing crops (intersection over 0.7).

Since no splits were provided, we randomly divided the dataset into a training set containing 2019 training, 267 validation, and 364 test samples. The three splits are similar ($p > 0.9$) according to the chi-square test when measuring the class frequencies. In this way, we created two datasets with three splits each due to the fact the dates of Sentinel-1 and Sentinel-2 do not exactly match because of the different revisit times. Figure 4 shows a sample from the test set in both Sentinel-1 VV and Sentinel-2 RGB.

4.2 Experimental Setting

Crop Field Segmentation The images have size $256 \times 256 \times 2$ for Sentinel-1 data and shape $256 \times 256 \times 12$ for Sentinel-2. We train all networks with an AdamW [24] optimizer and a learning rate scheduler with a reduction on plateau of factor 0.2 and patience of 5. The initial learning rate was set to 1e-4, and the batch size to 16. We trained the models for 60 epochs. We apply random horizontal and vertical flipping as augmentations. The loss function is generalized dice loss [40], which takes into account the class imbalance in the images. We compared the two networks with the same encoder (and so decoder) embedding sizes to better understand how they exploit the same representation space. We evaluated the networks under Intersection-Over-Union (IoU), F1-Score (F1), Precision (Prec), and Recall (Rec) for the positive class. We also evaluated GFLOPs to measure the efficiency of the network.



Fig. 4: South Africa Crop Type dataset example. Sentinel-1 shows the grayscale of the VV polarization, while Sentinel-2 shows the RGB of the same area. The mask shows the cultivated area in white.

Explainability We use Grad-CAM [37] as a visual post-hoc explanation method for this task because of its proven effectiveness in previous XAI studies in remote sensing [18]. Grad-CAM is particularly valuable because it helps answer the critical question, *"Where does the model focus when segmenting crops?"*.

For each image, we generate a single saliency map to quantify the influence of each pixel on the prediction of the positive class (i.e., cultivated area) made by a model (i.e., U-NET or U-KAN). In GradCAM, the generation process first involves computing the gradients of the positive class score with respect to feature maps of a selected convolutional layer. These gradients are then averaged globally to obtain the importance weights for each feature map. Next, a weighted sum of these feature maps is performed using the calculated weights. This yields a coarse location map that highlights the regions of the input image that are most influential in model segmentation decisions. Finally, we apply a ReLU activation to this weighted sum to ensure that only positive influences are considered, producing the final Grad-CAM heat map. In our experiments, we use Sentinel-2 data, which provides multispectral images over 12 channels, and generate the explanations for the test set images.

We assessed the *Plausibility*, *Sufficiency*, and *Per-channel Relevance* of the generated Grad-CAM heatmaps. Below, we provide a detailed description of each metric.

Plausibility Plausibility refers to the degree to which the explanations align with human understanding and domain-specific knowledge [13,18,36]. Assessing plausibility is crucial to ensure that models not only perform well but also align with human expectations and knowledge. In this study, we want to assess to which extent each obtained heat map aligns with the relative ground truth. We evaluate the plausibility of saliency maps by computing the IoU, F1, precision, and recall scores between the generated saliency map and the corresponding ground truth mask. Higher scores indicate that the saliency maps are more plausible, as they better align with the ground truth.

Since our saliency maps provide continuous explanations where each pixel has a value of importance, we established a threshold of importance to define which pixels are considered important for the segmentation with Otsu method [34]. This method segments the saliency map into distinct regions, creating a binary mask directly comparable to the binary truth mask.

Sufficiency Sufficiency is an aspect of faithfulness that evaluates whether an explanation indeed captures the key factors contributing to the segmentation and, therefore, is sufficient to explain it [18].

To assess the sufficiency, we retain only the important pixels identified by the explanation and mask the others. Specifically, we use the Otsu method to threshold the binary saliency maps, preserve only the image's relevant pixels, and set the others to zero. We then evaluate the performance metrics (IoU, F1-score, Precision, and Recall) for the positive class on this altered image. Sufficiency is computed as the difference in these metrics between the original and altered images. A smaller drop in performance indicates a more sufficient explanation.

Per-channel Relevance We assess the impact of image channels on the model explanations. For this evaluation, we extend the concept of occlusion sensitivity [8]—typically used to mask parts of an image—to occlude entire input data channels This evaluation aligns better with the nature of our data, where each channel represents distinct features with specific importance and classification roles. By systematically occluding an entire channel, we assess how its absence affects the model's explanation, providing insights into each channel's role in the classification process.

In our experiments, we occlude one channel at a time by zeroing it (which means no data for this imagery) and then compute the saliency map. Next, we measure the IoU between the saliency map obtained by occluding one channel and the saliency map obtained using all channels. A high IoU indicates that the excluded channel is less critical for the model's predictions as its omission has little impact on the model behavior. Conversely, a lower IoU suggests that the channel plays a significant role in the model's explanation and, thus, in its segmentation process.

5 Experimental Results

In this section, we present the results obtained for the analyzed dataset. First, we outline crop segmentation performance, addressing our research question RQ1. We then analyze the explanations of U-Net and U-KAN predictions from the qualitative and quantitative perspectives, addressing RQ2.

5.1 Task Performances

In Table 1, we report the results obtained when employing U-Net and U-KAN on Sentinel-1 and Sentinel-2 data. U-KAN provides the best performance in terms

of IoU on Sentinel-2, proving its adaptability in dealing with complex relations. On Sentinel-1, U-KAN gets comparable performance in terms of IoU to U-Net. In terms of precision, the KAN variant is more performant, scoring $\approx +3\%$ than U-Net. Although Sentinel-1 imagery is less affected by atmospheric events, Sentinel-2 bands provide a better understanding of the area with both networks. U-KAN better exploits the numerous channels of Sentinel-2, while the 2 channels of SAR are exploited in a similar manner by the two networks. U-KAN proves to be more computationally efficient than a standard U-Net when looking at GFLOPS: it consumes half the one of U-Net.

The KAN variant proves to be a preferable solution in both cases, providing better or comparable performance in fewer GFLOPs. Moreover, it proves to be more precise in any case.

Table 1: Results of U-Net and U-KAN using Sentinel-1 (S1) and Sentinel-2 (S2) data.

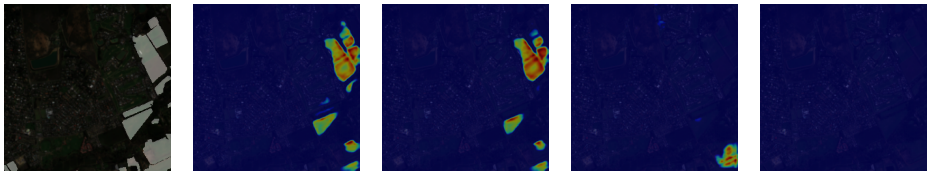
	Model	GFLOPs ↓	IoU ↑	F1 ↑	Prec ↑	Rec ↑
S1	U-Net	79.89	**65.59**	**79.21**	74.56	**85.56**
	U-KAN	**44.89**	65.36	79.03	**77.40**	77.50
S2	U-Net	80.65	72.95	84.35	72.57	**94.33**
	U-KAN	**45.65**	**74.82**	**85.59**	**75.24**	93.31

5.2 Analysis of explanations

We analyze the explainability results of U-KAN and U-Net networks on the Sentinel-2 dataset through qualitative and quantitative assessments.

Qualitative Evaluation We examine the saliency maps obtained from both networks. This analysis provides insight into the areas where each model focuses its attention, offering a deeper understanding of their segmentation behaviors.

Figure 1 shows examples of saliency maps generated by the U-Net and U-KAN models. The red pixels indicate the points where the network's attention is most focused, highlighting the differences in the behavior of the two models. Figure 1 (c) reveals that the U-Net model focuses on a significantly larger area than the U-KAN model. This observation suggests that, regardless of the effectiveness of the segmentation task, the U-Net model tends to distribute its focus over larger regions. In contrast, the U-KAN model has an interesting feature in its approach to the segmentation task. The network focuses predominantly on the boundaries of cultivated areas rather than within the areas themselves. This focus on boundaries suggests that the U-KAN model prioritizes delineating the edges of areas of interest. This last feature opens up the potential use of U-KAN networks in boundary delimitation and mapping tasks, where precise edge detection and delineation are crucial [27, 46]. These considerations generally apply



(a) RGB and GT **(b)** All Channels **(c)** No B01 **(d)** No B06 **(e)** No B11

Fig. 5: Per-channel relevance examples of U-KAN. The figure shows the ground truth over the original RGB image (a) and the saliency map of all 12 channels (b). Images (c), (d), and (e) display saliency maps generated by obscuring channels corresponding to B01, B06, and B11, respectively.

to the images in the dataset. We included multiple examples of saliency maps in our repository.

Figure 5 illustrates the change in saliency maps when specific channels related to bands B01, B06, and B11 are obscured for the U-KAN. From Figure 5(c), we can infer that by obscuring the channel relative to band B01, the resulting saliency map is very similar to the original one (Figure 5(b)), indicating that this channel has little impact on model focus. When the channel corresponding to band B06 is obscured, we observe a change in the saliency map, suggesting an influence of this channel on network focus. The most significant change occurs when we obscure the channel relative to B11 (Figure 5(e)). The resulting saliency map shows that the model does not focus on any specific area, implying that the network cannot detect segmentation traits without this channel. This result highlights the critical importance of band B11 in the model's ability to perform the segmentation task. Similar observations apply to the U-Net. We further investigate the impact of channels on saliency maps in the following quantitive analysis.

Table 2: Plausibility of Explanations for Sentinel-2.

Model	IoU	F1	Prec	Rec
U-KAN	**73.52**	80.77	**84.49**	77.74
U-Net	68.19	**84.50**	82.23	**87.13**

Quantitative Evaluation Table 2 presents the evaluation results of *Plausibility*. Regarding the plausibility of explanations, U-KAN demonstrates higher IoU and Precision than its competitor, U-Net. This indicates that U-KAN provides more accurate and reliable explanations, aligning more closely with human understanding. On the other hand, U-Net has a higher Recall and F1 score of plausibility. This could be because U-Net is more sensitive and captures more features in saliency maps, although it includes more false positives.

Table 3: Sufficiency of Explanations for Sentinel-2.

Model	IoU	F1	Prec	Rec
U-KAN	67.94 (-6.88)	80.59 (-5.0)	84.46 (+9.22)	77.45 (-15.86)
U-Net	68.85 (-4.1)	81.26 (-3.09)	84.10 (+11.53)	78.95 (-15.38)

In Table 3, we present the results assessing the *Sufficiency* of explanations. Sufficiency is quantified as the difference in metrics between masked images and their original counterparts. An interesting observation is the variation in the Precision metric. While there is a decrease in other metrics, which aligns with removing less critical pixels, there is an improvement in Precision for both U-KAN and U-Net. This increase in precision is noteworthy because by excluding less important pixels, both networks demonstrate an enhanced ability to delineate pixels belonging to crop class.

Table 4: Per-channel relevance (lower is better) based on IoU for Sentinel-2.

	B01	B02	B03	B04	B05	B06	B07	B08	B8A	B09	B11	B12
U-KAN	72.92	74.62	46.90	45.90	**0.11**	47.49	38.51	15.61	**0.00**	42.48	**0.19**	47.03
U-Net	68.08	68.90	46.24	46.26	**0.13**	47.19	36.91	15.54	**0.00**	42.55	**0.18**	48.30

Table 4 reports the *Per-channel relevance* results. For each model, we report the IoU score between the saliency map of all channels and the one obtained by obscuring the specific channel related to the band. In this context, a lower IoU signifies the higher importance of a channel. Specifically, if removing a channel yields a lower or zero IoU, it indicates that the channel plays an important role in the final segmentation task. For both U-KAN and U-Net models, we obtain that the channels corresponding to bands B05 (705 nm - Red Edge), B8A (865 nm - Narrow Near-Infrared), and B11 (1610 nm - Shortwave Infrared 1) are identified as the most significant for crop segmentation task due to their specific sensitivities. Specifically, B05 and B8A are sensitive to chlorophyll content and vegetation biomass and B11 to moisture content in soil and vegetation [9,11,30]. This quantitative evaluation across all test samples aligns with the qualitative results displayed in Figure 5, confirming the discussed insights. Analyzing the relevance of each channel in Sentinel-2 images opens the possibility of reducing the number of channels used by focusing on only the most important ones. This optimization can enhance efficiency and reduce computational costs while maintaining the quality of the analysis.

5.3 Analysis of the trained models

As previously mentioned, KANs are designed with a level of interpretability through the possibility of visualizing learnable activation functions. Although

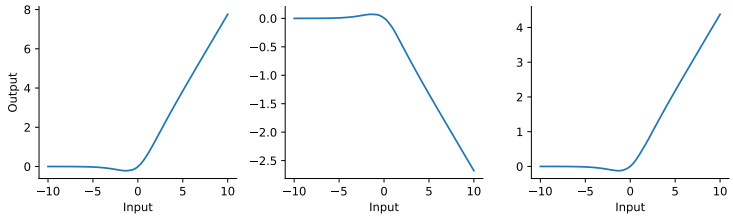


Fig. 6: Activation functions for the first element of the embeddings of the KAN decoder layer.

our main goal is to explore the post-hoc explainability of the U-KAN network with respect to the U-Net, we also sought to analyze the behavior of learnable activation functions in a decoding layer. This is also particularly relevant when analyzing the resource consumption of the network since the learned function can be simple to compute (e.g., linear) or complex when necessary, potentially helping both GFLOPs and performance, as shown before. In Figure 6, we report the learned activations for an element of the embeddings in a decoder layer. We can see substantial differences between the base function (SiLU) and ReLU, commonly employed by U-Net. U-KAN effectively represented more complex relations in the deep embeddings. The second activation is reversed along the y-axis compared to SiLU. The first and the third are similar but have a different slope (the first function is steeper). Moreover, each element of the embedding learns different activations with different complexity. Some of them could be constant activations for certain elements in the embeddings. The learned functions with a variance < 1 are $\approx 26\%$ the total, while the ones with a variance < 0.1 are $\approx 8\%$. This indicates the irrelevant parts of the embedding because every input is mapped to the same value every time.

6 Conclusions

In this work, we have shown how the new KANs can improve well-known architecture applied to the agricultural field, particularly in terms of efficiency, using only half the resources of full CNN architecture. Our study indicates that U-KAN offers superior performance when dealing with the numerous spectral channels of Sentinel-2 compared to U-Net, while they perform similarly with only two SAR channels. The explainability analysis also reveals two significant insights. First, the U-KAN network's emphasis on boundary details makes it particularly effective for tasks such as boundary delimitation and mapping. Second, not all the channels are useful for the segmentation task. So, users can decide to rely only on the most important reducing computational costs of the models. In future work, we intend to implement the insights from our network explainability analysis to enhance performance and reduce computational costs.

References

1. Adadi, A., Berrada, M.: Peeking inside the black-box: a survey on explainable artificial intelligence (xai). IEEE access **6**, 52138–52160 (2018)
2. Arnold, V.: On functions of three variables. In: Proceedings of the USSR Academy of Sciences. pp. 678–681 (1957)
3. Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., et al.: Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. Information fusion **58**, 82–115 (2020)
4. Ayushi, Buttar, P.K.: Satellite imagery analysis for crop type segmentation using u-net architecture. Procedia Computer Science **235**, 3418–3427 (2024). https://doi.org/10.1016/j.procs.2024.04.322, http://dx.doi.org/10.1016/j.procs.2024.04.322
5. Bauer, M.: Spectral inputs to crop identification and condition assessment. Proceedings of the IEEE **73**(6), 1071–1085 (1985). https://doi.org/10.1109/PROC.1985.13238
6. Behmann, J., Mahlein, A.K., Rumpf, T., Römer, C., Plümer, L.: A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. Precision Agriculture **16**(3), 239–260 (Aug 2014). https://doi.org/10.1007/s11119-014-9372-7, http://dx.doi.org/10.1007/s11119-014-9372-7
7. Belal, A.A., EL-Ramady, H., Jalhoum, M., Gad, A., Mohamed, E.S.: Precision Farming Technologies to Increase Soil and Crop Productivity, p. 117–154. Springer International Publishing (2021). https://doi.org/10.1007/978-3-030-78574-1_6, http://dx.doi.org/10.1007/978-3-030-78574-1_6
8. Borys, K., Schmitt, Y.A., Nauta, M., Seifert, C., Krämer, N., Friedrich, C.M., Nensa, F.: Explainable ai in medical imaging: An overview for clinical practitioners–saliency-based xai approaches. European journal of radiology **162**, 110787 (2023)
9. Broge, N.H., Leblanc, E.: Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. Remote sensing of environment **76**(2), 156–172 (2001)
10. Buttar, P.K., Sachan, M.K.: Semantic segmentation of satellite images for crop type identification in smallholder farms. The Journal of Supercomputing **80**(2), 1367–1395 (Jul 2023). https://doi.org/10.1007/s11227-023-05533-4, http://dx.doi.org/10.1007/s11227-023-05533-4
11. Castaldi, F., Hueni, A., Chabrillat, S., Ward, K., Buttafuoco, G., Bomans, B., Vreys, K., Brell, M., van Wesemael, B.: Evaluating the capability of the sentinel 2 data for soil organic carbon prediction in croplands. ISPRS Journal of Photogrammetry and Remote Sensing **147**, 267–282 (2019)
12. Corley, I., Robinson, C., Ortiz, A.: A change detection reality check. arXiv preprint arXiv:2402.06994 (2024)
13. Doshi-Velez, F., Kim, B.: Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608 (2017)
14. Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., et al.: Sentinel-2: Esa's optical high-resolution mission for gmes operational services. Remote sensing of Environment **120**, 25–36 (2012)
15. Gevaert, C.M.: Explainable ai for earth observation: A review including societal and regulatory perspectives. International Journal of Applied Earth Observation and Geoinformation **112**, 102869 (2022)

16. Huang, X., Sun, Y., Feng, S., Ye, Y., Li, X.: Better visual interpretation for remote sensing scene classification. IEEE Geoscience and Remote Sensing Letters **19**, 1–5 (2021)
17. Hung, S.C., Wu, H.C., Tseng, M.H.: Integrating image quality enhancement methods and deep learning techniques for remote sensing scene classification. Applied Sciences **11**(24), 11659 (2021)
18. Kakogeorgiou, I., Karantzalos, K.: Evaluating explainable artificial intelligence methods for multi-label deep learning classification tasks in remote sensing. International Journal of Applied Earth Observation and Geoinformation **103**, 102520 (2021)
19. Khan, S.M., Shafi, I., Butt, W.H., Diez, I.d.l.T., Flores, M.A.L., Galán, J.C., Ashraf, I.: A systematic review of disaster management systems: approaches, challenges, and future directions. Land **12**(8), 1514 (2023)
20. Kolmogorov, A.N.: On the representation of continuous functions of several variables by superpositions of continuous functions of a smaller number of variables. In: Proceedings of the USSR Academy of Sciences (1956)
21. Levering, A., Marcos, D., Lobry, S., Tuia, D.: Interpretable scenicness from sentinel-2 imagery. In: IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium. pp. 3983–3986. IEEE (2020)
22. Li, C., Liu, X., Li, W., Wang, C., Liu, H., Yuan, Y.: U-kan makes strong backbone for medical image segmentation and generation. arXiv preprint arXiv:2406.02918 (2024)
23. Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T.Y., Tegmark, M.: Kan: Kolmogorov-arnold networks (2024), https://arxiv.org/abs/2404.19756
24. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations (2019)
25. M Rustowicz, R., 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 (June 2019)
26. Maddy, E.S., Boukabara, S.A.: Miidaps-ai: An explainable machine-learning algorithm for infrared and microwave remote sensing and data assimilation preprocessing-application to leo and geo sensors. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **14**, 8566–8576 (2021)
27. Mahdianpari, M., Salehi, B., Rezaee, M., Mohammadimanesh, F., Zhang, Y.: Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery. Remote Sensing **10**(7), 1119 (2018)
28. Martínez-Ferrer, L., Piles, M., Camps-Valls, G.: Crop yield estimation and interpretability with gaussian processes. IEEE Geoscience and Remote Sensing Letters **18**(12), 2043–2047 (2020)
29. Mateo-Sanchis, A., Piles, M., Amorós-López, J., Muñoz-Marí, J., Adsuara, J.E., Moreno-Martínez, Á., Camps-Valls, G.: Learning main drivers of crop progress and failure in europe with interpretable machine learning. International Journal of Applied Earth Observation and Geoinformation **104**, 102574 (2021)
30. Van der Meer, F., Van der Werff, H., Van Ruitenbeek, F.: Potential of esa's sentinel-2 for geological applications. Remote sensing of environment **148**, 124–133 (2014)
31. Moharram, M.A., Sundaram, D.M.: Land use and land cover classification with hyperspectral data: A comprehensive review of methods, challenges and future directions. Neurocomputing **536**, 90–113 (Jun 2023). https://doi.org/10.1016/j.neucom.2023.03.025, http://dx.doi.org/10.1016/j.neucom.2023.03.025

32. Molnar, C.: Interpretable machine learning. Lulu. com (2020)
33. Mulla, D.J.: Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. Biosystems Engineering **114**(4), 358–371 (Apr 2013). https://doi.org/10.1016/j.biosystemseng.2012.08.009, http://dx.doi.org/10.1016/j.biosystemseng.2012.08.009
34. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics **9**(1), 62–66 (1979). https://doi.org/10.1109/TSMC.1979.4310076
35. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. pp. 234–241. Springer (2015)
36. Samek, W., Wiegand, T., Müller, K.R.: Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. arXiv preprint arXiv:1708.08296 (2017)
37. Selvaraju, R.R., Das, A., Vedantam, R., Cogswell, M., Parikh, D., Batra, D.: Grad-cam: Why did you say that? arXiv preprint arXiv:1611.07450 (2016)
38. Skakun, S., Wevers, J., Brockmann, C., Doxani, G., Aleksandrov, M., Batič, M., Frantz, D., Gascon, F., Gómez-Chova, L., Hagolle, O., López-Puigdollers, D., Louis, J., Lubej, M., Mateo-García, G., Osman, J., Peressutti, D., Pflug, B., Puc, J., Richter, R., Roger, J.C., Scaramuzza, P., Vermote, E., Vesel, N., Zupanc, A., Žust, L.: Cloud mask intercomparison exercise (cmix): An evaluation of cloud masking algorithms for landsat 8 and sentinel-2. Remote Sensing of Environment **274**, 112990 (Jun 2022). https://doi.org/10.1016/j.rse.2022.112990, http://dx.doi.org/10.1016/j.rse.2022.112990
39. Stomberg, T., Weber, I., Schmitt, M., Roscher, R.: Jungle-net: Using explainable machine learning to gain new insights into the appearance of wilderness in satellite imagery. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences **3**, 317–324 (2021)
40. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Jorge Cardoso, M.: Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations, p. 240–248. Springer International Publishing (2017). https://doi.org/10.1007/978-3-319-67558-9_28, http://dx.doi.org/10.1007/978-3-319-67558-9_28
41. Temenos, A., Temenos, N., Kaselimi, M., Doulamis, A., Doulamis, N.: Interpretable deep learning framework for land use and land cover classification in remote sensing using shap. IEEE Geoscience and Remote Sensing Letters **20**, 1–5 (2023)
42. Thenkabail, P.S., Enclona, E.A., Ashton, M.S., Van Der Meer, B.: Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. Remote Sens. Environ. **91**(3-4), 354–376 (Jun 2004)
43. Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M., et al.: Gmes sentinel-1 mission. Remote sensing of environment **120**, 9–24 (2012)
44. UNESCO, C.: Recommendation on the ethics of artificial intelligence (2021)
45. Western Cape Department of Agriculture, Radiant Earth Foundation: Crop type classification dataset for western cape, south africa (2021)
46. Xie, S., Tu, Z.: Holistically-nested edge detection. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (December 2015)
47. Zhong, L., Hu, L., Zhou, H.: Deep learning based multi-temporal crop classification. Remote Sensing of Environment **221**, 430–443 (Feb 2019). https://doi.org/10.1016/j.rse.2018.11.032, http://dx.doi.org/10.1016/j.rse.2018.11.032