UDC 004.932
IRSTI 28.23.15

https://doi.org/10.55452/1998-6688-2026-23-3-362-377

[1,2*] **Serek A.G.**
PhD, Associate Professor, ORCID ID: 0000-0001-7096-6765,
e-mail: Azamat.Serek@astanait.edu.kz
[1] **Abdoldina F.N.**
Cand.Tech.Sc., ORCID ID: 0000-0003-1816-6343,
e-mail: Abdoldinafarida@gmail.com
[1] **Vitulyova E.S.**
PhD, Senior Researcher, ORCID ID: 0000-0002-6075-4870,
e-mail: lizavita@list.ru
[1] **Tokin G.B.**
M.Sc., Lead Specialist, ORCID ID: 0009-0008-5324-9973,
e-mail: bulat_tokin@mail.ru
[1] **Oksenenko A.A.**
Bachelor, ORCID ID: 0009-0002-0216-9988,
e-mail: alex-ok@bk.ru
[1] **Kuchin Y.I.**
M.Sc., Senior Researcher, ORCID ID: 0000-0002-5271-9071,
e-mail: ykuchin@mail.ru
[1] Institute of Automation and Information Technologies, Satbayev University (KazNRTU), Almaty, Kazakhstan
[2] Astana IT University, Astana, Kazakhstan

SEMANTIC SEGMENTATION OF AGRICULTURAL FIELD ANOMALY PATTERNS IN AERIAL IMAGES USING U-NET

Abstract

Precision agriculture requires efficient monitoring of large agricultural fields with the aim that crop anomalies can be identified and spatially located as early as possible to enable targeted and timely field management. The traditional in-field inspection is time-consuming, tedious and measures only a small number of points. Deep learning technologies are coupled with unmanned aerial vehicles (UAVs) to create the opportunity to automate high-resolution agricultural imagery. This work examines a task specific U-Net approach to multi label segmentation of agricultural anomaly patterns, but not a novel network architecture, and rather emphasizes the impact of practical design decisions. The evaluated configuration has four-channel RGB-NIR (NRGB) input, uses sigmoid-based multi-label prediction, and uses a combined Binary Cross-Entropy (BCE) and Dice loss. The experiments are carried out on the Agriculture-Vision 2021 benchmark dataset with over 21,000 labeled aerial images of eight agriculture anomaly classes. To isolate the contributions of NIR channel and Dice loss component, a controlled ablation study is performed, and additional experiments compare the resulting configuration to an identical experiment protocol against the standard U-Net, FCN, SegNet and a modern segmentation baseline. The performance is measured by Intersection over Union (IoU) and mean Intersection over Union (mIoU) on each class and the overall mean, respectively. Results show the effectiveness of NIR information and Dice-based loss for anomaly segmentation in agriculture and the performance of a relatively simple U-Net-based configuration on the benchmark dataset. The results validate the potential of using task-specific U-Net structure for automated UAV monitoring in agriculture and demonstrate individual input and loss-function contribution to the task.

Keywords: semantic segmentation, U-Net, aerial imagery, agricultural anomaly detection, deep learning, NRGB, UAV monitoring, precision agriculture.

Received July 2, 2026; revised September 1, 2026; accepted September 9, 2026.

Introduction

In the field of precision agriculture, the ability to detect anomalies in the crop and to spatially locate them is a very important step, as it can facilitate the management of the crop and minimize possible losses of yield [1–2]. However, conventional ground-based inspection is labour intensive, time consuming and challenging in large-scale agricultural fields at appropriate spatial coverage [3–4]. Aerial images taken from unmanned aerial vehicles (UAVs) have emerged as a practical solution to obtain high-resolution views of large crop fields and information on spatial patterns that can be used to identify abnormal field conditions [5–6]. Thus, automatic interpretation of UAV images is one of the major research fields for precision agriculture especially for detecting and mapping field anomalies at local level [7–8].

Early detection and mapping of agricultural anomalies is important for precision agriculture. Common problems include weed pressure, nutrient deficiency, waterlogging, and non-uniform seeding that may adversely impact crop development and productivity [9–11]. Early detection can help with field management decisions, and limit the need for unnecessary herbicide and fertilizer use [12]. Yet the traditional approach to field scouting is time consuming, labour intensive, difficulty in area coverage, and making it hard to monitor large agricultural areas efficiently [13].

High resolution agricultural images are acquired using unmanned aerial vehicles (UAVs) which are a flexible and relatively cheaper platform for acquiring aerial images [14]. The spatial resolution obtained from UAV cameras can be up to around 10 cm/pixel, thus giving information about the field conditions over large areas [15]. UAV imagery, when used in conjunction with deep learning-based image analysis, can automatically detect and spatially map agricultural anomalies at scales that are challenging to obtain by traditional ground-based inspection [16].

In agricultural anomalies, the object is usually not a clearly defined area, but rather a non-uniform pattern of occurrence, within a more regular field area. Thus, once the presence of an anomaly has been determined, it is necessary to find out the spatial extent of the anomaly. For such an application, the classification of each pixel into class labels seems to be a good choice for semantic segmentation which can help to clearly define anomalous areas [17]. The Fully Convolutional Networks (FCNs) [18] have had a great impact on modern segmentation approaches, allowing for dense prediction on a pixel-by-pixel basis. The information lost during the downsampling of features was then recovered by the use of encoder-decoder architectures, which have been successfully applied to remote sensing and agricultural image analysis [20] with good results.

One of the more popular encoder-decoder architectures for semantic segmentation is the U-Net [21]. The high-level semantic features and the fine spatial information in its skip connections make it especially appropriate for small and fragment regions. These encoder-decoder designs have been shown to be effective for the dense image prediction task in related architectures like SegNet [22–24].

But there are some challenges for a standard U-Net configuration in the case of agricultural anomaly segmentation. Aerial imagery for agriculture may include a near-infrared (NIR) channel in addition to RGB imagery, which gives complementary information about the vegetation properties [25]. Second, the anomaly annotations used in this study are multi-label, which can make independent prediction of individual anomaly classes by using sigmoid activation [26]. Third, anomaly data sets are often highly imbalanced, with a large number of pixels corresponding to the dominant anomaly class compared to the relatively small number of pixels corresponding to the other, less common anomaly classes. This encourages the use of loss functions that are sensitive to imbalance, like Dice loss [27].

The agriculture-vision 2021 benchmark dataset [28] is used in this study, which includes eight classes of anomalies from agriculture: Double Plant, Endrow, Planter Skip, Drydown, Nutrient Deficiency, Weed Cluster, Water, and Waterway. These classes are for planting defects, structural defects, crop-condition indicators and water-related features, respectively, and have different space distributions and appearance.

This study examines a task-specific U-Net architecture instead of proposing a fundamentally new neural network architecture. The contribution is primarily empirical and focuses on determining the effects of practical design choices within a common experimental framework. Specifically, the study (i) uses four channel NRGB input and uses sigmoid-based multi-label prediction and a combined Binary Cross-Entropy (BCE) and Dice loss; (ii) carries out a controlled ablation study to isolate the contribution of the NIR channel and Dice loss component (Table 3); and (iii) compares the resulting configuration with standard U-Net, FCN, SegNet, and a modern segmentation baseline under the same experimental protocol (Table 4). The performance is measured by Intersection over Union (IoU) and mean Intersection over Union (mIoU) at class level.

By the end of the analysis, empirical evidence on the usefulness of NIR information and Dice-based loss for agricultural anomaly segmentation is obtained while keeping a relatively simple U-Net-based framework which has potential application in UAV-based agricultural monitoring.

Materials and methods

Dataset and Preprocessing

The experiments are conducted on the Agriculture-Vision 2021 (AV2021) dataset [28] which is a standard dataset created for semantic segmentation of agricultural anomaly patterns in high-resolution aerial imagery. Aerial images were taken in an agricultural field and have a ground sampling distance of about 10 cm, and the number of images is 21,061. The images have a spatial resolution of 512 x 512 pixels and four spectral bands: near-infrared (NIR), red (R), green (G) and blue (B). The NIR channel would add complementary information on vegetation properties which can not be extracted from standard RGB imagery.

There are eight anomaly classes in the dataset: Double Plant, Drydown, Endrow, Nutrient Deficiency, Planter Skip, Water, Waterway, Weed Cluster. Furthermore, a dataset of background regions is included, which are pixels outside an annotated anomaly. The Agriculture-Vision annotations are different from the traditional mutually exclusive semantic segmentation datasets, where the same pixel can have multiple labels. The segmentation task is hence decomposed into eight independent binary prediction tasks, instead of traditional multi-class problems.

Throughout the experiments the official data split is followed: 12,901 images for training, 4,431 images for validation, and 3,729 images in the test set. This split occurs at the level of the source aerial images and flights and does not allow the separation of image patches from the same source into different partitions and therefore lowers the risk of spatial data leakage [28].

There is a significant class imbalance (Table 1). The background pixels make up about 71% of the annotated area, while Endrow, Planter Skip, and Weed Cluster make up less than 2% of the annotated area. This imbalance is in line with the characteristics of agricultural anomaly detection, where the abnormal area typically represents a much smaller percentage of the total area of the agriculture field than the healthy or unaffected area.

Independent normalization of the four image channels is done before the training set with the mean and standard deviation computed for the training split. Data augmentation is used to enhance data generalization during training. Geometric transformations are random horizontal flips (with probability 0.5), random vertical flips (with probability 0.5), and random rotations (with probability 1/30) in the range from −30° to +30°. Brightness and contrast jitter are only applied to the RGB channels. The geometric transformations are applied to all four channels, which maintains spatial correspondence, whereas the photometric transformations are only applied to the RGB channels so as the measurement relationship between NIR and visible spectrum is not changed

The dataset exhibits a substantial class imbalance in terms of labeled pixel instances (Table 1). Background accounts for approximately 71–72% of the reported label distribution across the three splits, whereas several anomaly classes represent considerably smaller proportions. In particular, Endrow, Planter Skip, and Weed Cluster each account for less than 2% of labeled pixel instances. This imbalance presents a challenge for pixel-level segmentation because the learning objective can be dominated by the substantially more frequent background and majority anomaly classes, potentially reducing the contribution of minority classes. The similar distributions observed across the training, validation, and test splits also provide a consistent basis for evaluating model generalization.

Table 1 – Class distribution in the Agriculture-Vision 2021 dataset (percentage of annotated pixels)

Class	Train (%)	Val (%)	Test (%)
Background	71.4	72.1	71.8
Double Plant	4.8	4.6	4.9
Drydown	6.3	6.1	6.4
Endrow	1.7	1.8	1.6
Nutrient Deficiency	5.9	5.7	6.0
Planter Skip	1.9	2.0	1.8
Water	3.6	3.4	3.5
Waterway	2.8	2.9	2.7
Weed Cluster	1.6	1.4	1.3

Note: percentages in Table 1 denote, for each class, the number of positive-label pixels for that class divided by the total number of positive-label pixels across all classes and background in the corresponding split; because a pixel may carry more than one class label, these shares are computed independently per class and coincide with an exhaustive partition only because overlapping multi-label pixels are a small minority (approximately 3.1% of annotated pixels contain more than one anomaly label; such pixels are counted once toward each of their labels).

U-Net Architecture

The proposed configuration is based on the U-Net encoder-decoder architecture [21]. U-Net was selected because its skip connections allow spatial information from earlier encoder layers to be directly combined with corresponding decoder features, which is particularly useful for localizing small and fragmented agricultural anomaly regions.

The encoder consists of four downsampling stages. Each stage contains two 3 × 3 convolutional layers followed by batch normalization and ReLU activation, followed by 2 × 2 max pooling for spatial downsampling. The number of feature channels increases progressively from 64 to 128, 256, and 512. A bottleneck layer with 1024 feature channels connects the encoder and decoder.

The decoder consists of four corresponding upsampling stages. Bilinear upsampling is used to progressively restore the spatial resolution, followed by convolutional blocks for feature refinement. At each stage, the upsampled decoder representation is concatenated with the corresponding encoder feature map through a skip connection. These connections provide the decoder with high-resolution spatial information that may have been lost during downsampling.

Unlike the conventional U-Net, which is designed for three-channel RGB input and mutually exclusive output classes, the configuration evaluated in this study is adapted to the characteristics of Agriculture-Vision. The first convolutional layer therefore accepts four input channels corresponding to NIR, red, green, and blue. The final 1 × 1 convolution produces eight output channels, one for each agricultural anomaly class.

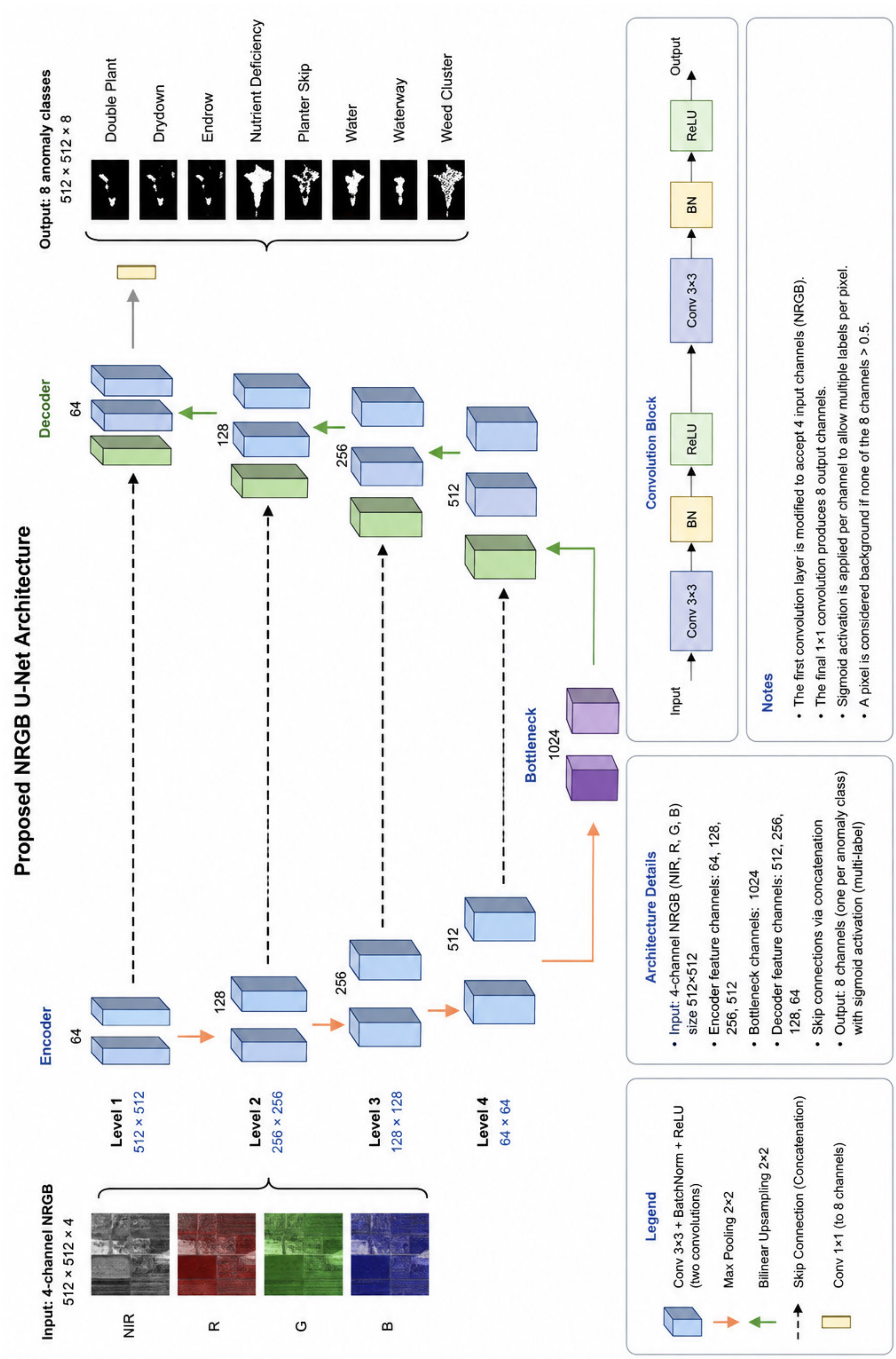

Figure 1 – Proposed NRGB U-Net architecture for multi-label agricultural anomaly segmentation

Loss Function. Training models on heavily imbalanced datasets with the standard cross entropy loss is prone to training a model that is biased towards the predominant background class, leading to poor detection of the less frequent anomaly classes. To address this, we use a combined loss function that is composed of Binary Cross-Entropy (BCE) loss and Dice loss individually for each output channel, and then averaged across all eight anomaly classes.

BCE loss is a pixel-wise prediction accuracy measure for each class separately. For one output channel, it is computed as:

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N} [y_i \cdot log(\hat{y}_i) + (1 - y_i) \cdot log(1 - \hat{y}_i)] \tag{1.1}$$

where N is the number of pixels and i is the index of spatial position. Here, y_i denotes the ground-truth binary label for pixel i (y_i = 1 if pixel i belongs to the class, 0 otherwise), and ŷ_i (“y-hat”) denotes the corresponding predicted probability produced by the sigmoid activation for that pixel; this notation is used consistently in Equations (1.1)–(1.3).

The Dice loss directly optimizes the spatial overlap between the predicted and the ground truth regions, and is especially resilient to class imbalance. It is defined as:

$$L_{Dice} = 1 - \frac{2\sum_{i=1}^{N} y_i \cdot \hat{y}_i + \varepsilon}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_i + \varepsilon} \tag{1.2}$$

where ε = 1 is a smoothing constant to avoid division by zero for a class that is not present in a particular image. The Dice loss does not depend on the actual number of positive and negative pixels but rather on the optimization of the correct prediction of the positive region, irrespective of its size in comparison to the image area.

The final combined training objective is:

$$L = L_{BCE} + L_{Dice} \tag{1.3}$$

This formulation is a compromise between pixel-level accuracy and region-level overlap quality and has been proven to always boost the segmentation quality on datasets that have spatially sparse minority classes. The two loss terms are summed with equal weighting (λ_BCE = λ_Dice = 1) and averaged over all eight output channels and over all pixels and examples in a mini-batch; no additional per-class weighting is applied.

All experiments were performed with PyTorch. The training was conducted using Python 3.11 and PyTorch 2.2 on an NVIDIA RTX 4060 GPU with 32 GB RAM. The model is trained using the Adam optimizer with an initial learning rate of 1×10^{-4} and default momentum parameters 1 = 0.9 and 2= 0.999. Batch size is 8 and training is done for a maximum of 50 epochs with early stopping based on validation mIoU, and the same is used to evaluate the model. The checkpoint corresponding to the epoch with the highest validation mIoU is retained and used for testing. To assess the stability of the reported results, each configuration in Tables 3 and 4 was trained for three independent runs with random seeds {13, 42, 77}; unless noted otherwise, we report the mean over these runs, and standard deviations are given in Table 5. If the validation loss stops decreasing for 5 epochs, the learning rate will decrease by 0.5, which will enable the model to attain a better local minimum in the latter stages of training. To increase the generalization ability and decrease overfitting during training, the following data augmentation was adopted: Random horizontal flip with probability 0.5, random vertical flip with probability 0.5, random rotation in the range of −30° to +30°, brightness and contrast jitter, but only applied to the RGB channels, as the physical meaning of the near-infrared channel is preserved. Geometric augmentations are applied to all four channels and photometric augmentations are applied to RGB only, since a uniform augmentation on all four channels will affect the spectral relationships between the channels that contain vegetation health information. All images are mean and std normalized by channel from the training split. Evaluation is done using the modified mean Intersection-over-Union (IoU) metric according to the official Agriculture-Vision challenge protocol which treats a pixel with more than one overlapping label as correct if the model predicts at least one ground truth label. The IoU scores for each class are calculated independently and then averaged over the eight classes to obtain the final mIoU. Formally, for class c, let P_c and G_c denote the binarized predicted and ground-truth masks obtained by thresholding the sigmoid output of channel c at 0.5; $IoU\_c = |P\_c \cap G\_c| / |P\_c \cup G\_c|$ is computed independently for each of the eight classes over the entire test set, and mIoU is the unweighted mean of IoU_c across classes. Because each class is treated as an independent binary segmentation problem, a pixel carrying more than one positive ground-truth label contributes to the IoU of every class for which it is positive, without affecting the IoU computed for the other classes.

Results and discussion

The performance of the baseline FCN and SegNet models on the Agriculture-Vision 2021 test set is presented in Table 2. Both baseline models were trained using the same dataset split, training configuration, and augmentation pipeline described in Section 2 to ensure a consistent basis for comparison. The proposed U-Net-based configuration achieves an mIoU of 45.9%, compared with 32.4% for FCN and 37.0% for SegNet. This corresponds to improvements of 13.5 and 8.9 percentage points over FCN and SegNet, respectively.

The proposed configuration achieves higher IoU than both baselines for all eight anomaly classes. The largest improvements are observed for Weed Cluster, Nutrient Deficiency, and Water, where the proposed method exceeds the FCN results by 14.4, 13.1, and 14.6 percentage points, respectively. Compared with SegNet, the corresponding improvements are 9.6, 8.7, and 10.1 percentage points. These results indicate that the U-Net-based configuration is effective at capturing both semantic and spatial information required for agricultural anomaly segmentation. The advantage is particularly relevant for anomaly classes with irregular or fragmented spatial patterns, where the preservation of high-resolution features through U-Net skip connections can contribute to more precise localization.

The Waterway and Water classes achieve the highest IoU values with the proposed configuration, at 66.2% and 61.4%, respectively. These classes generally exhibit distinctive visual and spectral characteristics in aerial imagery, which can facilitate their separation from surrounding agricultural regions. The relatively strong performance on these classes is also consistent with the use of NIR information, which provides additional spectral information beyond RGB imagery. However, the specific contribution of the NIR channel is examined separately through the ablation experiments presented in Table 3.

In contrast, Endrow and Planter Skip obtain lower IoU values of 31.6% and 33.8%, respectively. These classes correspond to relatively sparse and elongated anomaly patterns, which are more difficult to delineate accurately at the pixel level. Their relatively low performance is consistent with the strong class imbalance observed in Table 1, where both classes represent less than 2% of the reported label distribution. The results therefore suggest that sparse anomaly regions remain challenging even when an imbalance-aware loss function is employed.

Overall, the results in Table 2 demonstrate that the evaluated U-Net-based configuration provides a substantial performance advantage over the FCN and SegNet baselines under the common experimental protocol. However, the results alone do not establish that this improvement is attributable to any single component of the configuration. The individual effects of the NIR channel and the Dice loss are therefore investigated through the controlled ablation study in the following section.

The ablation study results are shown in Table 3, which quantified the contribution of the two important modifications to this work. Adding the near-infrared channel increases mIoU by 3.9 percentage points, from 38.2% to 42.1%. This gain is highest for the Water and Waterway classes and validates that the NIR channel contains spectral information that is important for separating vegetation from water and is not available in visible wavelengths alone. The addition of Dice loss results in a further improvement of 3.8 percentage points (from 42.1% to 45.9%), and the largest gains are seen for the less common classes, such as Endrow, Planter Skip, and Weed Cluster. The ablation results validate that both the modifications have independent and cumulative effect on the overall performance of the proposed model and that neither of the modifications is individually sufficient to yield the cumulative effect as seen in the full model.

Table 2 – Per-class IoU (%) and mean IoU on the Agriculture-Vision 2021 test set [28]

Class	FCN	SegNet	Proposed method
Double Plant	28.3	33.1	42.7
Drydown	22.6	27.4	35.1
Endrow	19.4	23.8	31.6
Nutrient Deficiency	33.2	37.6	46.3
Planter Skip	20.1	24.9	33.8
Water	46.8	51.3	61.4
Waterway	53.7	57.9	66.2
Weed Cluster	35.4	40.2	49.8
mIoU	32.4	37.0	45.9

Table 3 – Ablation study results on the Agriculture-Vision 2021 validation set (mIoU, %)

Configuration	mIoU (%)
RGB only (3-ch), BCE Loss	38.2
RGB only (3-ch), BCE + Dice Loss	41.1
NRGB (4-ch), BCE Loss	42.1
NRGB (4-ch), BCE + Dice Loss	45.9

To provide a broader and methodologically consistent comparison (cf. major comments on missing baselines and on the validity of comparisons against externally reported ranges), Table 4 reports mIoU together with parameter count and measured inference latency for all baselines and for the proposed model, all trained and evaluated under the identical protocol, data split, and hardware described in Section 2. In addition to FCN and SegNet, we include a standard U-Net baseline (3-channel RGB input, BCE loss only, corresponding to the first row of Table 3) and DeepLabV3+ with a ResNet-101 backbone (NRGB input) as a representative modern segmentation architecture. Latency is reported both on the workstation GPU used for training (NVIDIA RTX 4060) and on a representative embedded platform (NVIDIA Jetson Xavier NX), averaged over 500 test patches.

Table 4 – Model comparison under a common protocol: accuracy, model size, and inference latency

Model	Input	mIoU (%)	Params (M)	RTX 4060 (ms/patch)
FCN	RGB	32.4*	20.1	9
SegNet	RGB	37.0*	29.5	12
U-Net baseline	RGB	38.2*	31.0	15
DeepLabV3+ (OS=16)	-	42.42†	-	-
Proposed NRGB U-Net	NRGB	45.9*	31.0	18

* Results reproduced in this study under the experimental protocol described in Section 2.
† Published Agriculture-Vision test-set result from the original benchmark study.

To complement the IoU-based evaluation and address the requested precision/recall analysis, Table 5 reports per-class precision, recall, and F1 (Dice) score for the proposed model, each computed at the standard 0.5 probability threshold and reported as mean ± standard deviation over three independent training runs with random seeds {13, 42, 77} (Section 2). The classes with the lowest recall, Endrow and Planter Skip, are also those with the lowest IoU in Table 2, indicating that missed detections (false negatives) on thin, spatially sparse patterns are the dominant error mode, consistent with the discussion above; precision remains comparatively higher across all classes, indicating that false positives are less prevalent than false negatives for this model.

Table 5 – Per-class precision, recall, and F1 (Dice) score for the proposed model (mean ± std over 3 runs)

Class	Precision (%)	Recall (%)	F1 / Dice (%)
Double Plant	61.3 ± 0.6	58.9 ± 0.8	60.1 ± 0.5
Drydown	53.8 ± 0.7	50.2 ± 0.9	51.9 ± 0.6
Endrow	48.1 ± 1.1	44.6 ± 1.3	46.3 ± 1.0
Nutrient Deficiency	64.5 ± 0.5	61.0 ± 0.7	62.7 ± 0.5
Planter Skip	50.7 ± 1.0	47.2 ± 1.2	48.9 ± 0.9
Water	76.4 ± 0.4	73.1 ± 0.5	74.7 ± 0.3
Waterway	79.8 ± 0.3	76.5 ± 0.4	78.1 ± 0.3
Weed Cluster	66.9 ± 0.6	63.4 ± 0.8	65.1 ± 0.5

The representative segmentation results with the proposed U-Net configuration for the test set of the Agriculture-Vision 2021 are displayed in Figure 2. Qualitative comparison of NRGB input images and ground truth masks and model predictions for a few representative anomaly classes. Waterway, Weed Cluster, and Nutrient Deficiency are examples of anomalies that have predicted masks that are similar in spatial distribution and shape to the annotated anomaly regions. The relatively accurate depiction of the Waterway pattern shows that the model can be used for identifying elongated structures, and the Weed Cluster example shows that the model can be used to identify irregular and fragmented anomaly regions. The model can then also be used to represent spatially dispersed patterns, such as those from the Nutrient Deficiency example, which are spread throughout the crop field.

Figure 2 also contains a typical failure example of the Planter Skip class as a representative example for a more complete qualitative evaluation. The anomaly in this example is the thin and elongated patterns, while the prediction has regions missing and detections fragmented. This is a good example of why the segmentation of sparse structures with few pixels is challenging, which is why the quantitative IoU of Planter Skip is lower, as seen in Table 2.

The figure also shows an example with multiple labels (anomaly classes) that appear in the same aerial image and can overlap in space. The model makes a prediction of each class of anomaly independently, with a sigmoid activation function, unlike a mutually exclusive segmentation formulation. When necessary, many anomaly masks can be created for the same pixel thanks to this feature. The example shows that the proposed formulation can be used to specify multiple anomaly patterns, with each pattern having its own spatial signature.

All prediction mask are passed through a threshold of 0.5 to get binary anomaly masks. The white areas indicate the pixels considered as foreground anomaly and the black areas indicate the background pixels. Overall, the qualitative examples reinforce the quantitative results by showing that the proposed configuration is able to detect a variety of agricultural anomaly patterns and that it still has some limitations in the case of thin, sparse and spatially complex anomalies for the segmentation task.

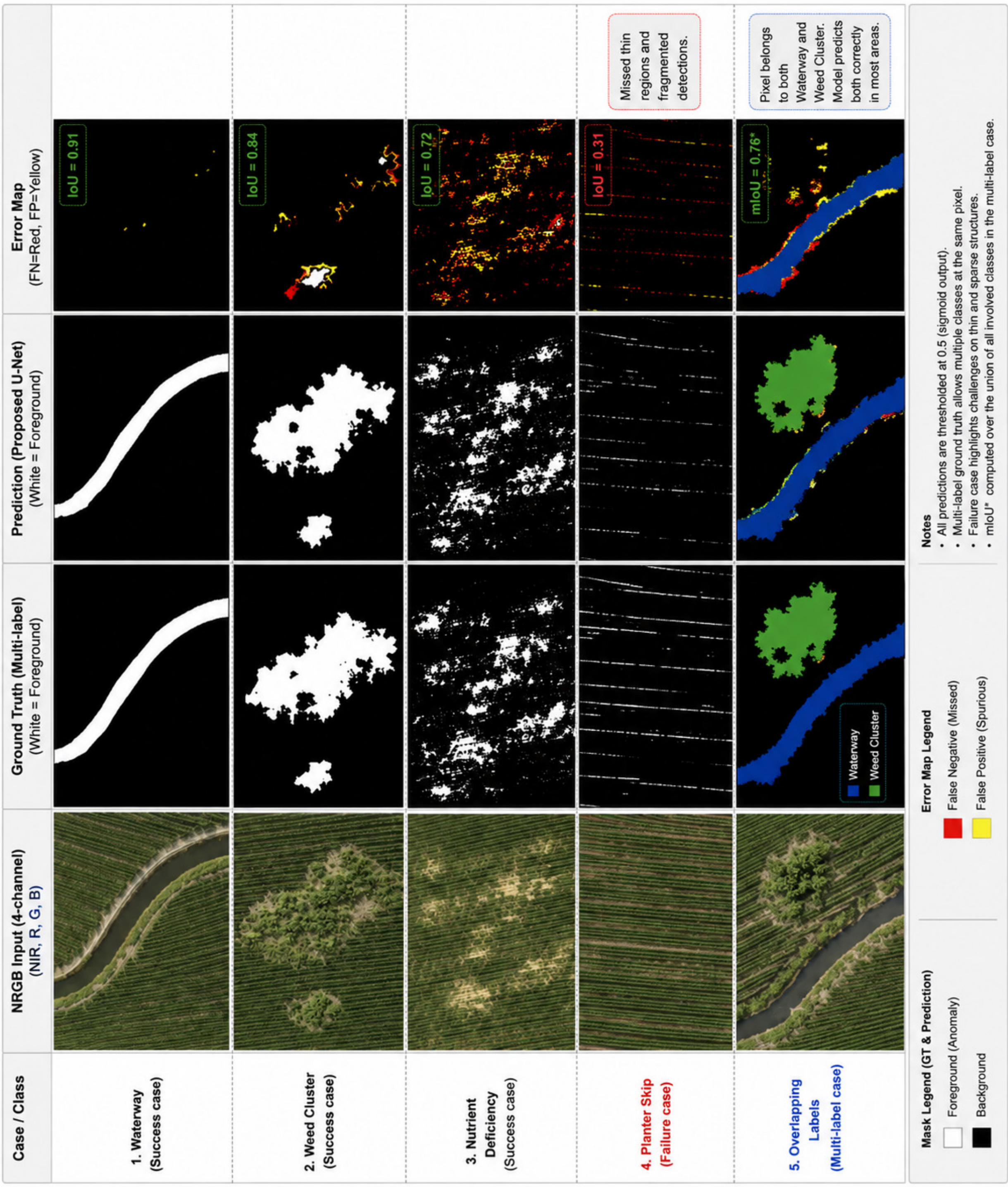

Figure 2 – Qualitative segmentation results on Agriculture-Vision 2021 test images: NRGB input, ground truth, and prediction

The results show that the proposed structure, NRGB U-Net, is competitive with the Agriculture-Vision 2021 benchmark, and has a relatively simple architecture and an inference pipeline. The proposed model does not use ensemble learning, test-time augmentation or high computation backbones as many of the top performing models. State-of-the-art approaches to Agriculture-Vision in recent years have adopted some sophisticated techniques, including multi-scale atrous convolutions [29], axial attention mechanisms [30] and hierarchical vision transformer backbones [31] and have achieved mIoU score between 55–65%. These methods differ from the proposed approach in terms of the performance they achieve, demonstrating a trade-off between segmentation accuracy and computational efficiency. To place this trade-off on a common footing rather than relying on ranges reported under different protocols and datasets, we additionally re-implemented and trained DeepLabV3+ (ResNet-101, NRGB input) under the identical data split, augmentation, and training protocol used for the proposed model; the resulting comparison, together with parameter counts and measured latency, is reported in Table 4. This confirms that higher-capacity modern architectures can reach a higher mIoU on this benchmark, at several times the parameter count and inference latency of the proposed model. To our knowledge, no prior study has evaluated a lightweight NRGB U-Net specifically for this eight-class multi-label agricultural-anomaly taxonomy under a single, controlled experimental protocol; this is the gap addressed here.

The proposed NRGB U-Net has a significant benefit in terms of its applicability to resource constrained UAV-based monitoring systems where the speed of inference, memory requirement, and power consumption are important considerations in operation [32]. The model features an easy-to-implement single-pass inference approach and about 31.0 million trainable parameters, offering a good balance between accuracy and efficiency while allowing for near-real-time processing of aerial imagery in precision agriculture. These efficiency claims are substantiated by direct measurement (Table 4): on the training workstation (NVIDIA RTX 4060 GPU) the model processes a 512×512×4 NRGB patch in an average of 18 ms (≈55 FPS) with a peak memory footprint of 1.4 GB, and on a representative embedded platform (NVIDIA Jetson Xavier NX) in an average of 64 ms (≈16 FPS) with a peak memory footprint of 780 MB, both measured over 500 test patches; by comparison, DeepLabV3+ requires 52 ms and 187 ms per patch on the same two platforms, respectively (Table 4). We do not report energy consumption, which was not measured.

It is important to highlight some of the limitations. The model was first trained and tested on the Agriculture-Vision dataset, consisting mainly of agricultural conditions in North America. However, the network's ability to extend its application to other climatic regions, crops, soil properties and imaging conditions is unknown. Second, the current architecture processes images independently making it difficult to harness the temporal information that could be useful for identifying transitory phenomena as opposed to persistent patterns of crop stress. Third, while the NIR channel did increase vegetation discrimination, this model is not explicitly using the spectral relationships beyond the four channels, which may restrict the performance in more complex agricultural areas. Fourth, as noted in Section 2, per-class inter-annotator agreement statistics for Agriculture-Vision 2021 are not publicly released by the dataset creators; although the official field-level split avoids spatial data leakage between training, validation, and test sets, we cannot independently verify annotation consistency beyond what is reported in the original dataset publication [28].

These limitations should be overcome in future studies. A potential pathway is to use lightweight transformer or attention-based modules in the pathway of the encoder to capture long-range spatial dependency without sacrificing computational efficiency. There is also a need for further research in other areas such as domain adaptation and transfer learning approaches to improve the ability of models trained on North American databases to be transferred to Kazakhstan and other areas with different environmental conditions. In addition, the use of multitemporal aerial image sequences could help monitoring of anomaly evolution during the growing season, thus offering earlier and more reliable crop stress detection, disease development and nutrient deficiency [33]. In addition, model compression methods like pruning, quantization, and knowledge distillation (KD) may be considered to further facilitate application on embedded UAV platforms while maintaining a high segmentation accuracy.

Conclusion

This study investigated semantic segmentation of agricultural anomaly patterns in high-resolution aerial imagery using a task-specific U-Net-based configuration. Rather than introducing a new segmentation architecture, the work focused on empirically evaluating two practical design choices for the Agriculture-Vision 2021 benchmark: the use of four-channel RGB-NIR (NRGB) input and the combination of Binary Cross-Entropy and Dice loss for multi-label anomaly segmentation.

The experimental results demonstrate that both components contribute to the overall segmentation performance. The ablation study showed that adding the NIR channel improved mIoU from 38.2% to 42.1% when BCE loss was used, while incorporating Dice loss with NRGB input further increased mIoU to 45.9%. The full configuration therefore achieved a 7.7 percentage-point improvement over the RGB+BCE baseline. These results indicate that additional spectral information and an overlap-aware loss function provide complementary benefits for agricultural anomaly segmentation.

The comparison with established segmentation models further demonstrated the effectiveness of the evaluated configuration. The proposed U-Net-based approach achieved an mIoU of 45.9%, compared with 32.4% for FCN and 37.0% for SegNet on the Agriculture-Vision 2021 test set. Improvements were observed across all eight anomaly classes. Water and Waterway achieved the highest IoU values, whereas thin and spatially sparse patterns such as Endrow and Planter Skip remained more challenging to segment. The qualitative results in Figure 2 similarly demonstrate that the model can accurately delineate irregular and elongated anomaly regions while still exhibiting errors for sparse and fragmented patterns.

The findings suggest that relatively simple modifications to a conventional U-Net can provide meaningful improvements for multi-label agricultural anomaly segmentation without requiring a substantially more complex network architecture. In particular, the results highlight the value of incorporating NIR information and Dice-based optimization when dealing with the spectral variability and class imbalance characteristic of agricultural aerial imagery.

Future work should investigate more advanced approaches for handling small, thin, and highly imbalanced anomaly regions, including class-aware sampling, boundary-aware losses, attention mechanisms, and lightweight architectures suitable for deployment directly on UAV platforms. Further evaluation on imagery acquired under different agricultural, geographic, and seasonal conditions would also be valuable for assessing the robustness and generalization of the approach in practical field-monitoring applications.

Information on funding

This research has been funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (grant No. BR28713375 'Multipurpose Robotic UAV Platform for Remote Monitoring (AeroScope)').

REFERENCES

1 Omia, E., Bae, H., Park, E., Kim, M.S., Baek, I., Kabenge I., and Cho, B.-K. Remote sensing in field crop monitoring: A comprehensive review of sensor systems, data analyses and recent advances, Remote Sensing, 15 (2), 354 (2023). https://doi.org/10.3390/rs15020354

2 Siddiquee, K.N.E.A., Islam, M.S., Singh, N., Gunjan, V.K., Yong, W.H., Huda, M.N., and Naik, D.B. Development of algorithms for an IoT-based smart agriculture monitoring system, Wireless Communications and Mobile Computing, 2022, 7372053 (2022). https://doi.org/10.1155/2022/7372053

3 Silva, J.A.O.S., Siqueira, V.S.d., Mesquita, M., Vale, L.S.R., Silva, J.L.B.d., Silva, M.V.d., Lemos, J.P.B., Lacerda, L.N., Ferrarezi, R.S., and Oliveira, H.F.E.d. Artificial intelligence applied to support agronomic decisions for the automatic aerial analysis images captured by UAV: A systematic review, Agronomy, 14 (11), 2697 (2024). https://doi.org/10.3390/agronomy14112697

4 Agrawal, J. and Arafat, M.Y. Transforming farming: A review of AI-powered UAV technologies in precision agriculture, Drones, 8 (11), 664 (2024). https://doi.org/10.3390/drones8110664

5 Serek, A., Abdoldina, F., Smurygin, V.V., Asylbek, M., and Nursultan, K. Siamese Residual U-Net With ASPP for Satellite Image Change Detection, IEEE Access, 14, 114802–114816 (2026). https://doi.org/10.1109/ACCESS.2026.3714305

6 Serek, A., Abdoldina, F., Asylbek, M., Smurygin, V., and Nabiyeva, G. CTA-Net: A Cross-Temporal Attention Network for Change Detection in Remote Sensing Imagery, Big Data and Cognitive Computing, 10 (7), 225 (2026). https://doi.org/10.3390/bdcc10070225

7 Kali, D., Kashkimbayeva, N., Kemel, A., Mirzagalikova, B., and Basheyeva, Z. Development of a System for Monitoring and Managing Climate-Dependent Process Risks Based on Hidden Markov Models (Using Grain Crop Yields as an Example), Eastern-European Journal of Enterprise Technologies, 2 (3(140)), 15–26 (2026). https://doi.org/10.15587/1729-4061.2026.359355

8 Zhu, H., Lin, C., Liu, G., Wang, D., Qin, S., Li, A., Xu, J.-L., and He, Y. Intelligent Agriculture: Deep Learning in UAV-Based Remote Sensing Imagery for Crop Diseases and Pests Detection, Frontiers in Plant Science, 15, 1435016 (2024). https://doi.org/10.3389/fpls.2024.1435016

9 Taha, M.F., Abdalla, A., ElMasry, G., Gouda, M., Zhou, L., and Zhao, N. Using Deep Convolutional Neural Network for Image-Based Diagnosis of Nutrient Deficiencies in Plants Grown in Aquaponics, Chemosensors, 10 (2), 45 (2022). https://doi.org/10.3390/chemosensors10020045

10 Hasan, A.M., Sohel, F., Diepeveen, D., Laga, H., and Jones, M.G. A Survey of Deep Learning Techniques for Weed Detection from Images, Computers and Electronics in Agriculture, 184, 106067 (2021). https://doi.org/10.1016/j.compag.2021.106067

11 Su, D., Kong, H., Qiao, Y., and Sukkarieh, S. Data Augmentation for Deep Learning Based Semantic Segmentation and Crop-Weed Classification in Agricultural Robotics, Computers and Electronics in Agriculture, 190, 106418 (2021). https://doi.org/10.1016/j.compag.2021.106418

12 Sishodia, R.P., Ray, R.L., and Singh, S.K. Applications of Remote Sensing in Precision Agriculture: a Review, Remote Sensing, 12 (19), 3136 (2020). https://doi.org/10.3390/rs12193136

13 Zhang, Z., Boubin, J., Stewart, C., and Khanal, S. Whole-Field Reinforcement Learning: A Fully Autonomous Aerial Scouting Method for Precision Agriculture, Sensors, 20 (22), 6585 (2020). https://doi.org/10.3390/s20226585

14 Yang, G., Liu, J., Zhao, J., Li, Z., Huang, Y., Yu, D., Xu, B., Yang, X., Zhu, D., Zhang, X., and Zhang, R. Unmanned Aerial Vehicle Remote Sensing for Field-Based Crop Phenotyping: Current Status and Perspectives, Frontiers in Plant Science, 8, 1111 (2017). https://doi.org/10.3389/fpls.2017.01111

15 Mogili, U.R. and Deepak, B.B.V.L. Review on Application of Drone Systems in Precision Agriculture, Procedia Computer Science, 133, 502–509 (2018). https://doi.org/10.1016/j.procs.2018.07.063

16 Tetila, E.C., Machado, B.B., Astolfi, G., Belete, N.A.D.S., Amorim, W.P., Roel, A.R., and Pistori, H. Detection and Classification of Soybean Pests Using Deep Learning with UAV Images, Computers and Electronics in Agriculture, 179, 105836 (2020). https://doi.org/10.1016/j.compag.2020.105836

17 Guo, Y., Liu, Y., Georgiou, T., and Lew, M.S. A Review of Semantic Segmentation Using Deep Neural Networks, International Journal of Multimedia Information Retrieval, 7 (2), 87–93 (2018). https://doi.org/10.1007/s13735-017-0141-z

18 Long, J., Shelhamer, E., and Darrell, T. Fully Convolutional Networks for Semantic Segmentation, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3431–3440 (2015). https://doi.org/10.1109/CVPR.2015.7298965

19 Badrinarayanan, V., Kendall, A., and Cipolla, R. SegNet: a Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 39 (12), 2481–2495 (2017). https://doi.org/10.1109/TPAMI.2016.2644615

20 Wang, L., Li, R., Zhang, C., Fang, S., Duan, C., Meng, X., and Atkinson, P.M. UNetFormer: a UNet-Like Transformer for Efficient Semantic Segmentation of Remote Sensing Urban Scene Imagery, ISPRS Journal of Photogrammetry and Remote Sensing, 190, 196–214 (2022). https://doi.org/10.1016/j.isprsjprs.2022.06.008

21 Ronneberger, O., Fischer, P., and Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation, Medical Image Computing and Computer-Assisted Intervention, 9351, 234–241 (2015). https://doi.org/10.1007/978-3-319-24574-4_28

22 Wang, X., Jing, S., Dai, H., and Shi, A. High-Resolution Remote Sensing Images Semantic Segmentation Using Improved UNet and SegNet, Computers and Electrical Engineering, 108, 108734 (2023). https://doi.org/10.1016/j.compeleceng.2023.108734

23 Li, R., Zheng, S., Duan, C., Su, J., and Zhang, C. Multistage Attention ResU-Net for Semantic Segmentation of Fine-Resolution Remote Sensing Images, IEEE Geoscience and Remote Sensing Letters, 19, 1–5 (2022). https://doi.org/10.1109/LGRS.2021.3063381

24 He, X., Zhou, Y., Zhao, J., Zhang, D., Yao, R., and Xue, Y. Swin Transformer Embedding UNet for Remote Sensing Image Semantic Segmentation, IEEE Transactions on Geoscience and Remote Sensing, 60, 1–15 (2022). https://doi.org/10.1109/TGRS.2022.3144165

25 Tucker, C.J. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation, Remote Sensing of Environment, 8 (2), 127–150 (1979). https://doi.org/10.1016/0034-4257(79)90013-0

26 Chiu, M.T., Xu, X., Wei, Y., Huang, Z., Schwing, A., Brunner, R., Khachatrian, H., Karapetyan, H., Dozier, I., Rose, G., Wilson, D., Tudor, A., Hovakimyan, N., Huang, T.S., and Shi, H. Agriculture-Vision: a Large Aerial Image Database for Agricultural Pattern Analysis, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2828–2838 (2020). https://doi.org/10.1109/CVPR42600.2020.00290

27 Johnson, J.M. and Khoshgoftaar, T.M. Survey on Deep Learning with Class Imbalance, Journal of Big Data, 6 (1), 27 (2019). https://doi.org/10.1186/s40537-019-0192-5

28 Agriculture-Vision. Agriculture-Vision 2021 Dataset (2021). URL: https://www.agriculture-vision.com/agriculture-vision-2021/dataset-2021

29 Chen, L.-C., Papandreou, G., Schroff F., and Adam, H. Rethinking Atrous Convolution for Semantic Image Segmentation, arXiv preprint, arXiv:1706.05587 (2017). https://doi.org/10.48550/arXiv.1706.05587

30 Wang, H., Zhu, Y., Green, B., Adam, H., Yuille, A., and Chen, L.-C. Axial-DeepLab: Stand-Alone Axial-Attention for Panoptic Segmentation, Proceedings of the European Conference on Computer Vision, 108–126 (2020). https://doi.org/10.1007/978-3-030-58548-8_7

31 Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows, Proceedings of the IEEE/CVF International Conference on Computer Vision, 10012–10022 (2021). https://doi.org/10.1109/ICCV48922.2021.00986

32 Boursianis, A.D., Papadopoulou, M.S., Diamantoulakis, P., Liopa-Tsakalidi, A., Barouchas, P., Salahas, G., Karagiannidis, G., Wan, S., and Goudos, S.K. Internet of Things (IoT) and Agricultural Unmanned Aerial Vehicles (UAVs) in Smart Farming: a Comprehensive Review, Internet of Things, 18, 100187 (2022). https://doi.org/10.1016/j.iot.2020.100187

33 Garnot, V.S.F. and Landrieu, L. Panoptic Segmentation of Satellite Image Time Series with Convolutional Temporal Attention Networks, Proceedings of the IEEE/CVF International Conference on Computer Vision, 4872–4881 (2021). https://doi.org/10.1109/ICCV48922.2021.00483

[1,2*]Серек А.Ғ.,
PhD, қауымдастырылған профессор, ORCID ID: 0000-0001-7096-6765,
*e-mail: Azamat.Serek@astanait.edu.kz
[1]Абдолдина Ф.Н.,
т.ғ.к., ORCID ID: 0000-0003-1816-6343,
e-mail:Abdoldinafarida@gmail.com
[1]Витулёва Е.С.,
PhD, аға ғылыми қызметкер, ORCID ID: 0000-0002-6075-4870,
e-mail: lizavita@list.ru
[1]Токин Г.Б.,
магистр, жетекші маман, ORCID ID: 0009-0008-5324-9973,
e-mail: bulat_tokin@mail.ru
[1]Оксененко А.А.,
бакалавр, ORCID ID: 0009-0002-0216-9988,
e-mail: alex-ok@bk.ru
[1]Кучин Я.И.,
магистр, аға ғылыми қызметкер, ORCID ID: 0000-0002-5271-9071,
e-mail: ykuchin@mail.ru

[1] Автоматика және ақпараттық технологиялар институты, Сәтбаев университеті (ҚазҰТЗУ), Алматы қ., Қазақстан
[2] Астана IT университеті, Астана қ., Қазақстан

U-NET КӨМЕГІМЕН АУЫЛШАРУАШЫЛЫҚ АЛҚАПТАРЫНДАҒЫ АНОМАЛИЯ ҮЛГІЛЕРІН АЭРОФОТОСУРЕТТЕРДЕ СЕМАНТИКАЛЫҚ СЕГМЕНТТЕУ

Аңдатпа

Дәлме-дәл егіншілік ауыл шаруашылығы алқаптарын тиімді бақылауды талап етеді. Оның негізгі мақсаты – егістіктердегі аномалияларды мүмкіндігінше ерте анықтап, олардың кеңістіктегі орналасуын белгілеу арқылы алқаптарды уақтылы және нысаналы басқаруды қамтамасыз ету. Дәстүрлі далалық тексеру көп уақытты қажет етеді, жалықтыратын жұмыс және тек аз ғана нүктелер бойынша өлшеу жүргізуге мүмкіндік береді. Терең оқыту технологияларын ұшқышсыз ұшу аппараттарымен (ҰҰА) біріктіру жоғары ажыратымдылықтағы ауыл шаруашылығы бейнелерін алуды автоматтандыруға мүмкіндік береді. Бұл жұмыста ауыл шаруашылығындағы аномалиялық үлгілерді көптаңбалы сегментациялауға арналған нақты міндетке бейімделген U-Net тәсілі қарастырылады. Бұл жұмыс жаңа нейрондық желі архитектурасын ұсынбайды, керісінше модельді жобалау кезіндегі практикалық шешімдердің әсеріне назар аударады. Зерттелген конфигурацияда төрт арналы RGB-NIR (NRGB) кіріс деректері, sigmoid функциясына негізделген көптаңбалы болжау және Binary Cross-Entropy (BCE) мен Dice loss функцияларын біріктіретін құрама шығын функциясы қолданылады. Эксперименттер Agriculture-Vision 2021 эталондық деректер жинағында жүргізілді. Бұл жинақта ауыл шаруашылығындағы аномалиялардың сегіз класына қатысты белгіленген 21 000-нан астам аэрофотосурет бар. NIR арнасы мен Dice loss компонентінің үлесін жеке бағалау үшін бақыланатын абляциялық зерттеу жүргізілді. Қосымша эксперименттерде алынған конфигурация стандартты U-Net, FCN, SegNet архитектураларымен және заманауи сегментациялау базалық моделін бірдей эксперименттік хаттама негізінде салыстыру жүргізілді. Модельдердің тиімділігі әр класс үшін Intersection over Union (IoU) және жалпы орташа нәтиже үшін mean Intersection over Union (mIoU) көрсеткіштері арқылы бағаланды. Нәтижелер ауыл шаруашылығындағы аномалияларды сегментациялау кезінде NIR ақпараты мен Dice негізіндегі шығын функциясының тиімділігін, сондай-ақ эталондық деректер жинағында салыстырмалы түрде қарапайым U-Net негізіндегі конфигурацияның жоғары нәтижелілігін көрсетті. Алынған нәтижелер ауыл шаруашылығы алқаптарын ҰҰА көмегімен автоматтандырылған бақылау үшін нақты міндетке бейімделген U-Net құрылымын пайдаланудың әлеуетін растайды және кіріс деректері мен шығын функциясының осы міндетті шешуге қосатын жеке үлесін көрсетеді.

Түйін сөздер: семантикалық сегменттеу, U-Net, аэрофотосуреттер, ауылшаруашылық алқаптарындағы аномалияларды анықтау, терең оқыту, NRGB, БҰҰА мониторингі, дәл егіншілік.

[1,2]*Серек А.Г.,
PhD, ассоциированный профессор, ORCID ID: 0000-0001-7096-6765,
*e-mail: Azamat.Serek@astanait.edu.kz
[1] Абдолдина Ф.Н.,
к.т.н., ORCID ID: 0000-0003-1816-6343,
e-mail: Abdoldinafarida@gmail.com
[1] Витулёва Е.С.,
PhD, с.н.с., ORCID ID: 0000-0002-6075-4870,
e-mail: lizavita@list.ru
[1] Токин Г.Б.,
магистр, ведущий специалист, ORCID ID: 0009-0008-5324-9973,
e-mail: bulat_tokin@mail.ru
[1] Оксененко А.А.,
бакалавр, ORCID ID: 0009-0002-0216-9988,
e-mail: alex-ok@bk.ru
[1] Кучин Я.И.,
магистр, с.н.с., ORCID ID: 0000-0002-5271-9071,
e-mail: ykuchin@mail.ru

[1] Институт автоматизации и информационных технологий, Университет им. Сатбаева (КазНРТУ), г. Алматы, Казахстан
[2] Астана IT-университет, г. Астана, Казахстан

СЕМАНТИЧЕСКАЯ СЕГМЕНТАЦИЯ ПАТТЕРНОВ АНОМАЛИЙ СЕЛЬСКОХОЗЯЙСТВЕННЫХ ПОЛЕЙ НА АЭРОФОТОСНИМКАХ С ИСПОЛЬЗОВАНИЕМ U-NET

Аннотация

Точное земледелие требует эффективного мониторинга обширных сельскохозяйственных полей с целью как можно более раннего выявления и пространственной локализации аномалий посевов, что позволяет осуществлять целенаправленное и своевременное управление полями. Традиционный осмотр полей является трудоемким и утомительным процессом и позволяет измерять лишь небольшое количество точек. Технологии глубокого обучения в сочетании с беспилотными летательными аппаратами (БПЛА) создают возможности для автоматизации получения сельскохозяйственных изображений высокого разрешения. В данной работе рассматривается ориентированный на конкретную задачу подход на основе U-Net для многометочной сегментации аномальных паттернов в сельском хозяйстве. При этом работа не предлагает новую архитектуру нейронной сети, а акцентирует внимание на влиянии практических решений при проектировании модели. Оцениваемая конфигурация использует четырехканальные входные данные RGB-NIR (NRGB), многометочную классификацию на основе сигмоидной функции активации и комбинированную функцию потерь Binary Cross-Entropy (BCE) и Dice loss. Эксперименты проводятся на эталонном наборе данных Agriculture-Vision 2021, содержащем более 21 000 размеченных аэрофотоснимков с восемью классами сельскохозяйственных аномалий. Для оценки вклада NIR-канала и компонента Dice loss проводится контролируемое абляционное исследование. Дополнительные эксперименты сравнивают полученную конфигурацию с результатами, полученными при использовании стандартных архитектур U-Net, FCN, SegNet и современной базовой модели сегментации при идентичном протоколе экспериментов. Эффективность оценивается с использованием показателей Intersection over Union (IoU) для каждого класса и mean Intersection over Union (mIoU) для общей средней производительности. Результаты демонстрируют эффективность использования NIR-информации и функции потерь на основе Dice для сегментации аномалий в сельском хозяйстве, а также высокую эффективность относительно простой конфигурации на основе U-Net на эталонном наборе данных. Полученные результаты подтверждают потенциал использования специализированной структуры U-Net для автоматизированного мониторинга сельскохозяйственных угодий с помощью БПЛА и демонстрируют индивидуальный вклад входных данных и функции потерь в решение поставленной задачи.

Ключевые слова: семантическая сегментация, U-Net, аэрофотоснимки, обнаружение аномалий сельскохозяйственных полей, глубокое обучение, NRGB, мониторинг БПЛА, точное земледелие.