# Land cover mapping with Sentinel-2 imagery using deep learning semantic segmentation models

Viktoriia Hnatushenko[1,2*,†], Oleksandr Honcharov[1, †]

[1] *Ukrainian State University of Science and Technologies, Gagarina Ave. 4, 49600, Dnipro, Ukraine*
[2] *Gottfried Wilhelm Leibniz University Hannover,NienburgerStr. 1D, 30167 Hannover, Germany*

**Abstract**

Land cover mapping is essential for environmental monitoring and evaluating the effects of human activities. Recent studies have demonstrated the effective application of particular deep learning models for tasks such as wetland mapping. Nonetheless, it is still ambiguous which advanced models developed for natural images are most appropriate for remote sensing data. This study focuses on the segmentation of agricultural fields using satellite imagery to distinguish between cultivated and non-cultivated areas. We employed Sentinel-2 imagery obtained during the summer of 2023 in Ukraine, illustrating the nation's varied land cover. The models were trained to differentiate among three principal categories: water, fields, and background.

We chose and optimised five advanced semantic segmentation models, each embodying distinct methodological methods derived from U-Net. Upon examination, all models exhibited robust performance, with total accuracy spanning from 80% to 89.2%. The highest-performing models were U-Net with Residual Blocks and U-Net with Residual Blocks and Batch Normalisation, whereas U-Net with LeakyReLU Activation exhibited much quicker inference times.

The findings suggest that semantic segmentation algorithms are highly effective for efficient land cover mapping utilising multispectral satellite images and establish a dependable benchmark for assessing future advancements in this domain.

**Keywords**

Semantic segmentation, agricultural lands, satellite images, deep learning, U-Net architecture.

## 1. Introduction

Land cover (LC) changes play a crucial role in assessing the current state of the environment. Human activities or regional climate variations can drive these changes. LC is considered one of the essential climate variables [1], making its timely assessment a key application in satellite remote sensing. Annually, thematic maps are essential for the purpose of addressing a wide range of environmental and land management needs. For medium-resolution mapping (approximately 250 m), the measurement uncertainty should be kept below 15%, while for high-resolution mapping (10–30 m), it should be kept below 5% [34].

Satellite optical imagery is the primary source of information for modern land cover mapping techniques. This process is significantly influenced by Landsat data, which is frequently supplemented by images from MODIS or SPOT-5 [12]. Additionally, high-resolution imagery and digital elevation models (DEMs) are employed as supplementary information sources for land cover mapping [14]. A contemporary source of optical imagery is the Copernicus Sentinel-2. It has become an additional critical data source due to its five-day revisit interval [13].

The Copernicus program [15], which is implemented by the European Space Agency (ESA) with the assistance of Sentinel satellites, is an example of an international initiative that actively contributes to the provision of unrestricted access to Earth observation (EO) data for a diverse array

*Information Technology and Implementation (IT&I-2024), November 20-21, 2024, Kyiv, Ukraine*
* Corresponding author.
† These authors contributed equally.
: vvitagnat@gmail.com (Vik. Hnatushenko); alexgoncharov06@gmail.com (O. Honcharov)
0000-0001-5304-4144 (Vik. Hnatushenko); 0009-0002-4349-4859 (O. Honcharov)

of commercial and non-commercial applications. The EU and ESA's Copernicus program is a multibillion-euro investment initiative that is designed to provide critical services using satellite data that is both extremely accurate and timely. The primary goals of the program are to enhance environmental management, mitigate the consequences of climate change, and promote the creation of new applications and services, including urban planning support and environmental monitoring.

The practical application of sophisticated methodologies, such as deep learning, that were previously untenable due to the necessity of extensive, representative datasets, is now facilitated by the availability of complimentary satellite data for mapping through initiatives such as Copernicus. Deep learning has led to significant progress in the fields of computer vision and pattern recognition in recent years [16, 17]. The intricate, multilevel architecture of deep learning models is responsible for their efficacy, as it enables the extraction of hierarchical feature sets from data and generates non-linear functions. Furthermore, the comprehensive learning framework of the system enables the concurrent acquisition of features from raw inputs and the forecasting of the objective task, thereby eliminating the need for heuristic feature design. This provides an advantage over traditional machine learning methods, including support vector machines (SVM) and random forests (RF), which operate on a multi-step feature engineering process. This procedure is replaced by a streamlined, end-to-end workflow in deep learning [18]. The availability of a vast array of datasets is a critical prerequisite for the effective application of deep learning techniques, as it allows the model to autonomously derive representative features for the predictive tasks.

Conventional supervised classification algorithms [19] have been the primary method employed by the majority of land cover mapping systems, regardless of the imagery type employed. Support vector machines (SVM), decision trees, random forests (RF), and maximum likelihood classifiers (MLC) are the most frequently employed classifiers. The development and improvement of segmentation models necessitates a substantial investment of time and professional effort in feature engineering, the process of acquiring the numerous features necessary for classification.

The growing need for accurate monitoring and management of agricultural land through satellite technologies presents challenges in the processing and analysis of geospatial data [26, 27]. To make good use of this kind of data, precise semantic segmentation algorithms must be created that can quickly and accurately tell the difference between different types of surfaces. Traditional methods often lack sufficient accuracy or demand significant computational resources, limiting their application in real-time and over large areas. Integrating deep learning into geospatial data analysis holds promise for addressing these issues, but the selection and optimisation of models for specific agricultural segmentation tasks remain unresolved [25,28]. The lack of a universal approach for choosing a neural network architecture that performs segmentation tasks efficiently with minimal computational time highlights the need for further research and adaptation of existing deep learning models.

## 2. Related Work

Advances in deep learning methods, notably convolutional neural networks (CNNs), have had a substantial influence on computer vision disciplines such as self-driving vehicles, image search engines, medical diagnostics, and augmented reality [19]. These innovations are also seeing increased use in agriculture and remote sensing.

Zhu et al. [2] emphasised the unique characteristics of remote sensing imagery, which provide significant issues when compared to typical RGB imagery. These problems include the georeferenced nature of the data, its multimodal composition, specialised imaging geometries, and interpretation complications. The challenge is exacerbated by a lack of adequate ground truth or labelled data for training deep learning models. Furthermore, most cutting-edge CNNs are intended for three-channel RGB pictures, demanding changes for optimal performance with remote sensing data.

Despite these constraints, recent research has looked at the use of deep learning in remote sensing imaging, focussing on applications such images preprocessing [29], target identification [30], classification [9], and semantic feature extraction and scene interpretation [16]. Deep learning

approaches for land cover (LC) or land use mapping have mostly focused on optical satellite, aerial, and multispectral data because to their similarity to RGB images typically utilised in computer vision research.

One key problem for academics is the lack of consistent, countrywide labelled data that spans both geographical and temporal dimensions. The European Union's Common Agricultural Policy (CAP) requires each member state to create paid agencies to handle this problem. These organisations use the Land Parcel Identification System (LPIS) [4] to gather data on parcel geometry and crop types for each farmer. This approach assures that data is collected consistently across nations, but national-scale ground truth data has been limited for years owing to access restrictions to these statements. Since 2019, nations such as France, Catalonia, Estonia, Croatia, Slovenia, Slovakia, and Luxembourg have gradually improved public access to these datasets, opening up significant prospects for creative agricultural applications within the Earth observation community.

Existing datasets, such as BigEarthNet [5], are largely concerned with land use/land cover categorisation from open data sources. BigEarthNet is one of the first large-scale benchmark archives, spanning 10 European nations and 125 Sentinel-2 tiles. Similarly, the Eurosat dataset [6] provides multiclass annotations for all 13 Sentinel-2 spectral bands, totalling 27,000 geo-referenced and labelled picture segments. Another dataset, So2Sat [7], focusses on metropolitan regions throughout the globe utilising Sentinel-1 and Sentinel-2 picture segments, with hand labelling by domain professionals.

However, the time component of satellite picture gathering is often disregarded in most available datasets, which prioritise annotated tags above segmentation masks. This restricts their usefulness to simple classification tasks and renders them unsuitable for more complicated applications such as object identification, picture segmentation, and parcel counting. The absence of the temporal component also limits deep learning models' capacity to capture seasonal patterns across different land cover classes.

For the classification of crop types using optical satellite images, a variety of deep learning methods have been devised, occasionally surpassing conventional computer vision or machine learning methods [8]. For example, [9] uses a hybrid technique of 1-D and 2-D CNNs to categorise 11 land cover types in Ukraine using Sentinel-2 and Landsat-8 data. Extra data, like area borders and statistical data, were added to the model to make the forecasts more accurate. Similarly, by including an additional branch for independent pixel-wise categorisation, the FG-Unet architecture [10] improves upon the popular U-Net model [11] and allows for more precise polygon borders.

The potential of hybrid feature selection for semantic crop and weed segmentation was established in [23]. This methodology may serve as a basis for creating more precise systems for identifying crops and weeds in fields, which is critical for successful agromanagement. Furthermore, in [24], three different U-Net topologies were tested for inventorying inland water bodies. The findings showed the benefits of attention processes and pre-trained networks in improving segmentation accuracy, which might be used to the identification of different agricultural land types.

Recent studies have significantly advanced semantic segmentation models by introducing modifications to the U-Net architecture to enhance feature extraction and accuracy. CM-UNet incorporates a Mamba-based decoder with a Channel and Spatial Mamba (CSMamba) block and a Multi-Scale Attention Aggregation (MSAA) module, achieving superior segmentation metrics across multiple remote sensing datasets [35]. Another approach combines DenseNet with U-Net, dilated convolutions, and DeconvNet, leading to an 11.1% increase in Pixel Accuracy and a 13.5% improvement in mean Intersection over Union (mIoU) while reducing parameters by 59% compared to traditional U-Net models on the Potsdam dataset [36].

Attention mechanisms are increasingly employed to capture multi-scale information and enhance segmentation accuracy. HAssNet utilizes a spatial attention mechanism for global correlation and channel attention to improve task-related channel focus, achieving a 6.7% mIoU improvement over prior models on remote sensing data [37]. Additionally, Deep Attention U-Net enhances global feature extraction by incorporating channel self-attention, showing a 2.48% improvement in mIoU over baseline U-Net models, particularly in handling occlusions [38].

Hybrid models that integrate CNNs and Transformers have also shown promising results. MFTransNet, a CNN-Transformer hybrid, demonstrates efficient segmentation across high-resolution remote sensing data, balancing accuracy with resource utilization [39]. Similarly, HST-UNet combines Shunted Transformer embedding with a Multi-Scale Convolutional Attention Network (MSCAN), achieving high F1 scores on ISPRS datasets [40].

Furthermore, models like AMMUNet introduce Granular Multi-Head Self-Attention (GMSA) and Attention Map Merging Mechanism (AMMM) to enhance segmentation precision, achieving mIoU scores of 75.48% on Vaihingen and 77.90% on Potsdam datasets, thus showing advantages in handling fine-grained details in agricultural and remote sensing contexts [41].

These recent studies underscore the critical role of architectural modifications, attention mechanisms, and hybrid CNN-Transformer architectures in overcoming challenges like class imbalance and limited segmentation precision, paving the way for more robust applications in agricultural monitoring and environmental management.

## 3. Research Objectives

The aim of this research is to develop and conduct a comparative analysis of U-Net architecture modifications for the task of semantic segmentation of agricultural lands based on satellite imagery. The study seeks to identify optimal architectural and training approaches that ensure high segmentation accuracy with minimal computational costs, with the goal of improving the efficiency of agricultural land monitoring and management. Special attention is given to analysing the impact of residual blocks, normalisation methods, and regularisation techniques on overall model performance, in order to establish best practices for processing geospatial data in the agricultural sector.

## 4. Methodology

### 4.1. Neural network

The main structure of this work is based on U-Net model [3], which is well known as an encoder-decoder configuration with skip connections to enable accurate pixel-wise segmentation. The encoder has multiple convolutional and pooling layers to obtain spatial information, while the decoder restores back original resolution by using upsampling and concatenation with corresponding encoder layers for contextillation.

In order to increase the performance of segmentation, some improvements were made: for better flow of at gradients, residual blocks were included, and Batch Normalisation was introduced to stabilise training while Dropout was used in an attempt to reduce overfitting. Activation functions ReLU and LeakyReLU were used for adding non-linearity helping the model to learn complex patterns. Softmax activation was then used on the output layer to classify the pixels between agricultural and non-agricultural classes.

For segmentation accuracy and visual consistency, we tested for various models trained with Adam Optimiser to determine the best architecture for the effective way of supporting agricultural field monitoring.

### 4.2. Data Source

In order to accomplish the goals of this study, we used satellite images acquired on June 5, 2023, from the Sentinel-2 mission (scene identifier: S2A_MSIL2A_20230605T083601_N0509_R064_T36TWS_20230605T125758.SAFE) in the Copernicus HUB archive. In order to assess the state of agricultural areas before the disastrous Kakhovka Hydroelectric Power Plant, this date was chosen. The chosen images include data from the following spectrum bands:

- **B03 (Blue band, 490 nm)**: In order to recognise surface water and differentiate it from vegetation, this band is crucial since it helps differentiate between water bodies and plants.
- **B04 (Red band, 665 nm):** It is vital to analyse plant health and identify regions of stressed or unhealthy vegetation using B04 (Red band, 665 nm), which is mostly used for measuring chlorophyll levels in plants.
- **B8A (Near-infrared, 865 nm)**: When measuring plant biomass and health, the near-infrared band (B8A, 865 nm) is essential. Its sensitivity to plant structure makes it a valuable tool for crop productivity estimation and field monitoring in agriculture.
- **B11 (Shortwave infrared, 1610 nm):** Soil moisture levels and plant water content may be effectively analysed using the B11 (shortwave infrared, 1610 nm) band. As a result, it can shed light on irrigation requirements and drought situations by differentiating between dry regions, healthy flora, and bare soil.

For model training, a 2560x2560 pixel area (51.2 thousand $m^2$) was extracted from the larger image, allowing for the generation of 400 image fragments, each sized 128x128 pixels (Figure 1). In the Figure 1 red box indicates the region selected for data collection and model training, featuring a variety of land covers such as agricultural fields, water bodies, and urban areas. A manual annotation of the data was performed for this image with three classes (Figure 2): agricultural fields (field), water bodies (water), and others (other), which includes roads, forests, urban and rural development areas, and more. In the figure 2 each image tile was manually annotated into three distinct classes: Agricultural fields ( green), Water bodies (blue), and Other (brown).

**Figure 1:** Study area map.

## 4.3. Data preprocessing

In order to guarantee the efficacy of the models' learning process, the following comprehensive data preprocessing protocol was implemented:

**Image Fragmentation**

A total of 400 individual image tiles were produced by dividing the selected region into smaller fragments, each of which measured 128x128 pixels. This fragmentation process enabled the model to concentrate on smaller areas, thereby improving its segmentation performance and capturing detailed features across a variety of land cover types. The model was able to more effectively understand the subtleties of various surface characteristics by dividing the larger image into smaller segments.

**Manual Annotation**

Three distinct classifications were carefully annotated onto each of the image tiles (Figure 2):

- **Agricultural fields (field)**: Areas that are cultivated for the purpose of agriculture.
- **Water bodies (water)**: This category encompasses rivers, lakes, and other substantial bodies of water.

- **Other (other)**: Including urban areas, forests, roads, and non-agricultural land cover.

This manual annotation procedure guaranteed the production of precise, high-quality labels, which are essential for the training of deep learning models in semantic segmentation.

**Pixel Intensity Normalisation**

The pixel intensity values of the images were normalised to the range [0, 1] to aid in the quicker and more stable convergence required for model training. This normalisation step allowed the model to process the input data more effectively. To guarantee that the models were trained solely on the original data, no additional data augmentation techniques were implemented, as the dataset was a sufficient size.

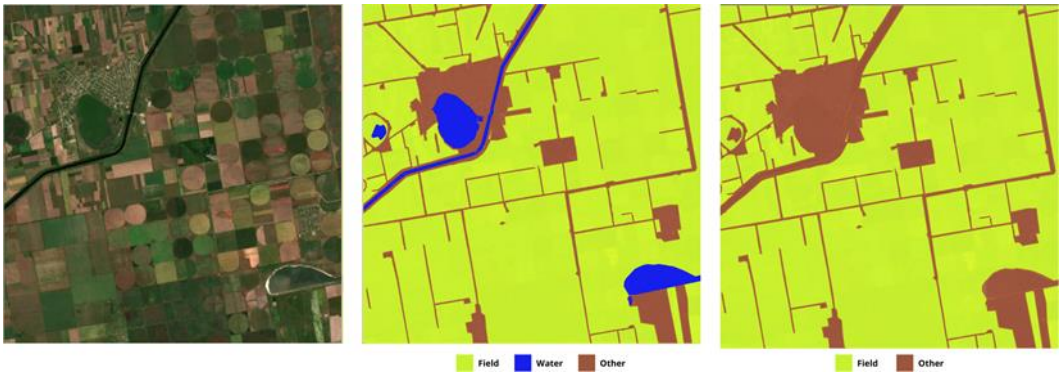


**Figure 2.** Data annotation based on satellite images.

### 4.4. Evaluation methods and analytical tools

**Evaluation Methods:**

The primary strategy for evaluating segmentation results was based on the confusion matrix and visual analysis, aimed at assessing the quality of the extracted agricultural fields. Future work will expand the evaluation tools by incorporating additional metrics, allowing for a more comprehensive analysis of the models' performance.

**Analytical Tools and Software:**

In this study, Python was utilised as the foundation for developing the deep learning models. The TensorFlow and Keras libraries were employed to build and train the neural networks, while NumPy and Rasterio were used for satellite image processing, providing powerful tools for data manipulation. GeoPandas were crucial for handling geospatial data and facilitating efficient spatial analysis. Manual data annotation for training the models was performed using the GroundWork tool, which ensured high-quality training datasets and accurate identification of target objects in the images.

This methodology reflects a comprehensive approach to analysing agricultural lands using satellite imagery and deep learning technologies. It provides a robust foundation for assessing the potential of various models in segmentation tasks.

## 5. Experiments

In this study, five different variations of the U-Net architecture were designed and implemented, each with specific enhancements aimed at improving segmentation accuracy and generalisation:

### Model #1: U-Net with ReLU Activation (3x3 Kernel => 2x2 Max Pooling)

The spatial resolution of the input image (128x128 pixels with 4 channels) is gradually reduced by three encoder blocks in this model (Figure 3), which employ max-pooling and convolutional operations. The encoder's filter count increases from 32 to 128. The "Bottleneck" layer, located at the

core of the network, employs 128 filters to process the data, deriving more detailed features and preserving critical information.

The decoder is intended to recapture the spatial resolution to its original scale by integrating feature maps from corresponding encoder blocks and performing upsampling and concatenation operations. This allows for the preservation of spatial context. For each decoding phase, UpSampling2D is implemented, followed by a convolutional layer that employs ReLU activation. This enables the network to more precisely reconstruct the segmentation map.

Each pixel is classified into one of the two target classes: agricultural fields and non-agricultural areas, using softmax activation in the output layer. This architecture serves as a benchmark for assessing the effects of more sophisticated enhancements, offering a simple U-Net design that requires no further modifications.

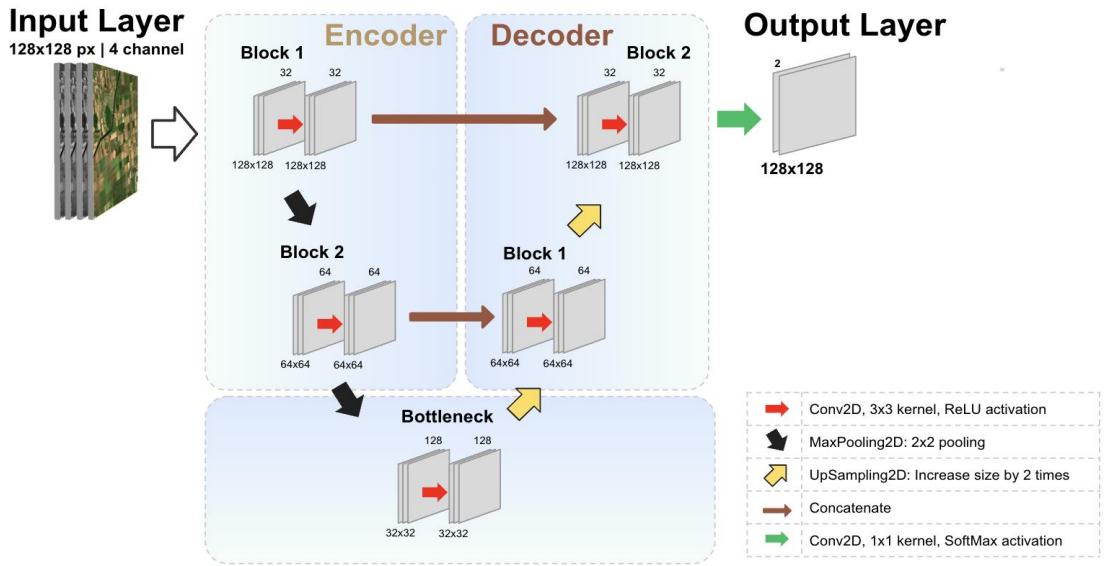


**Figure 3:** Architecture of Model #1 (U-Net with ReLU Activation, 3x3 Kernel => 2x2 Max Pooling).

### Model #2: U-Net with LeakyReLU Activation and Batch Normalisation (3x3 Kernel => 2x2 Max Pooling)

The spatial resolution of the input image (128x128 pixels) is reduced by four encoder blocks, each of which employs a combination of 3x3 convolutions and 2x2 max-pooling operations, as seen in Figure 4. At each block, the number of filters increases, going from 64 to 512, allowing the model to capture complicated patterns at numerous scales.

Positioned at the centre, the bottleneck layer employs 1024 filters to extract deep semantic features and efficiently encode contextual information. Each convolutional layer is succeeded by Batch Normalisation, which normalises activations to stabilise and expedite the training process.
The decoder component of the network reconstructs the spatial resolution while maintaining the context from the encoder by concatenating and upsampling the corresponding encoder feature maps. In order to incorporate non-linearity, a LeakyReLU-activated convolutional layer is included after each upsampling step. In an effort to guarantee semantic segmentation that is both precise and seamless, the output layer implements Softmax to generate pixel-wise classification into two classes.

### Model #3: U-Net with Residual Blocks

The design of Model #3 (Figure 5) integrates the traditional U-Net framework with residual blocks, so augmenting the model's efficacy by facilitating improved gradient flow during training. The model starts with an input layer that handles pictures of 128x128 pixels and including 4 channels. Subsequently, there are four encoder blocks with residual connections: the initial block has 32 filters,

while the subsequent block comprises 64 filters. Each block has two convolutional layers utilising ReLU activation, succeeded by max-pooling, which diminishes the picture dimensions by fifty percent. The core or "bottleneck" layer analyses data using 128 filters at a resolution of 32x32 pixels.

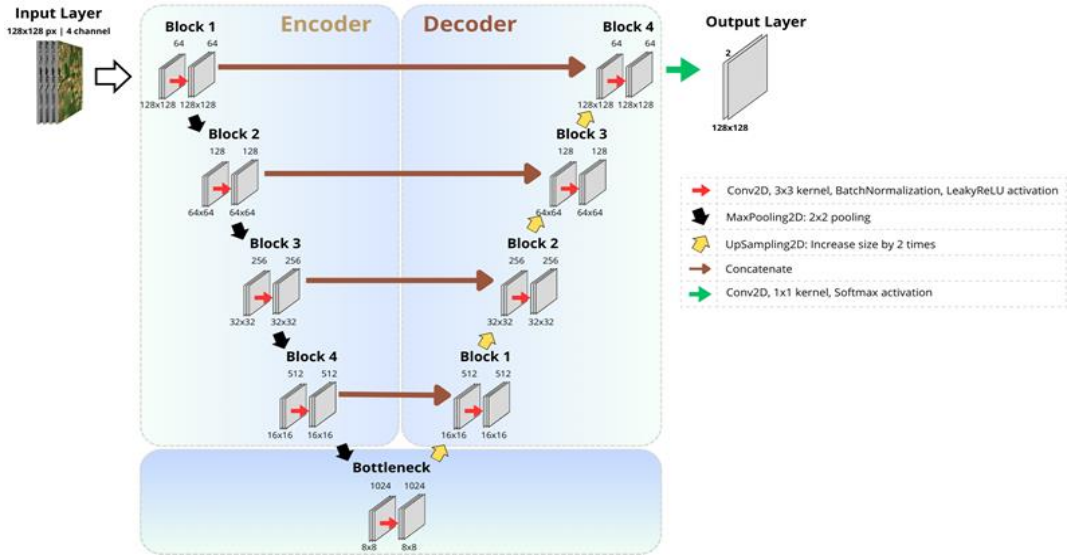


**Figure 4:** Architecture of Model #2 (U-Net with LeakyReLU Activation and Batch Normalisation, 3x3 Kernel => 2x2 Max Pooling).

The decoder has two blocks that sequentially restore the picture to its original dimensions: the first block, utilising 64 filters, upsamples the image to 64x64 pixels, while the second block, employing 32 filters, restores the image to 128x128 pixels. Both decoder blocks include upsampling, concatenation with corresponding encoder blocks, and two convolutional procedures using ReLU activation and residual connections. The output layer utilises softmax to categorise each pixel into one of two distinct groups, hence facilitating successful semantic segmentation of the image.

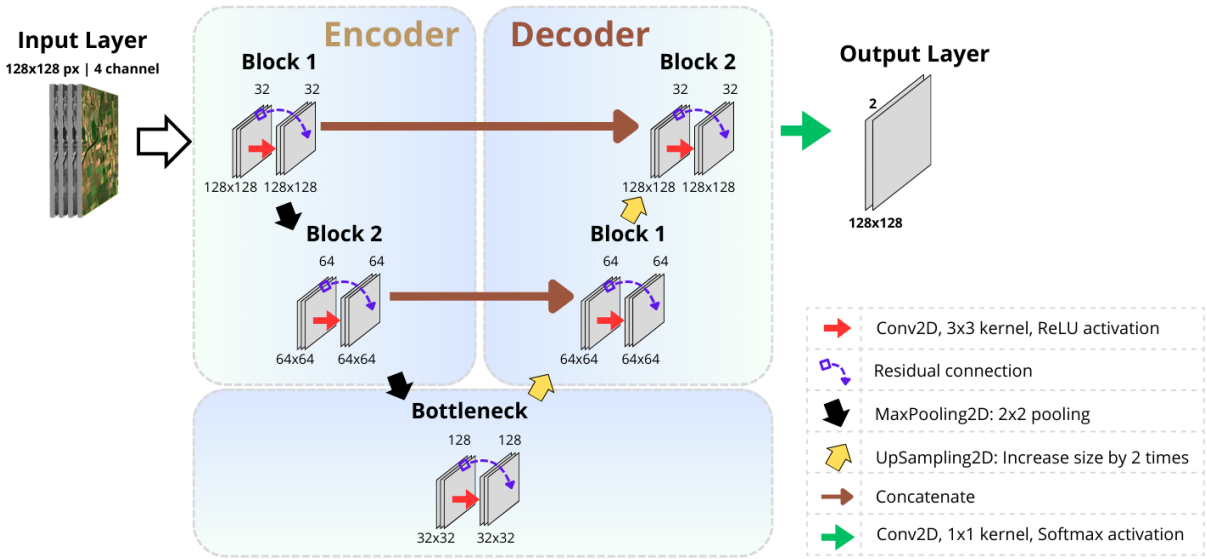


**Figure 5:** Architecture of Model #3 (U-Net with Residual Blocks).

### Model #4: U-Net with Residual Blocks, Batch Normalisation, and Dropout

By adding Batch Normalisation and Dropout layers to each residual block, Model #4's architecture (Figure 7) builds on the design concepts of Model #2 and emphasises enhanced efficiency. By normalising the input data prior to activation, batch normalisation helps to reduce the internal covariate shift issue and increases training speed and stability overall. Dropout, when applied at a rate of 0.5, reduces overfitting by arbitrarily turning off neurones during training, which enhances generalisation by gaining more robust patterns of data. With these improvements, Model #3 is more robust against overfitting problems and performs better in semantic segmentation tasks, especially on complicated and varied datasets.

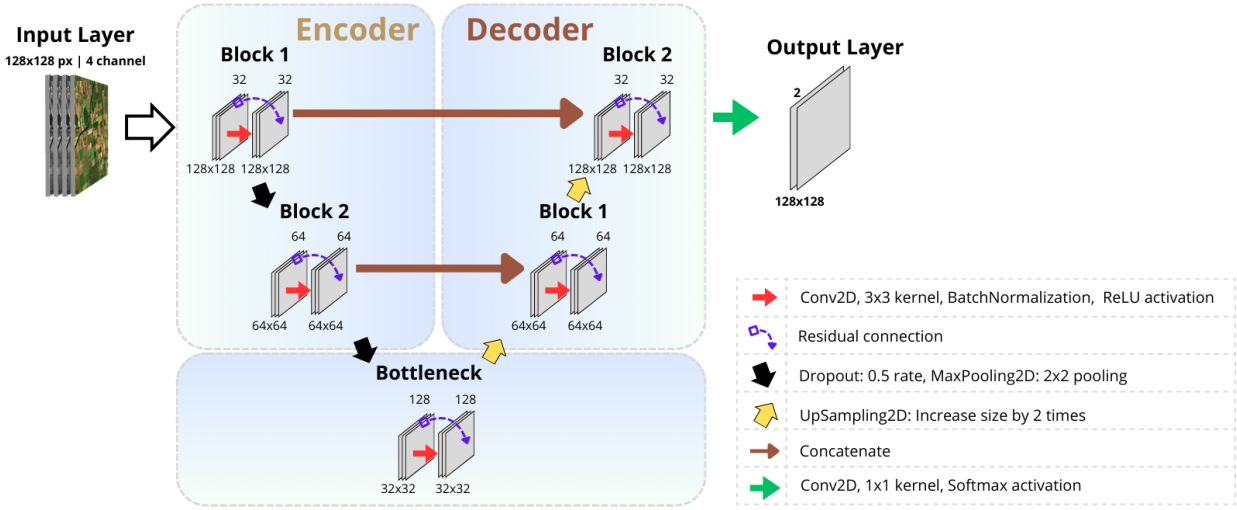


**Figure 6:** Architecture of Model #4 (U-Net with Residual Blocks, Batch Normalisation, and Dropout).

### Model #5: U-Net with Residual Blocks, Batch Normalisation, Dropout, and LeakyReLU

The architecture of Model #5 is intended to capitalise on the advantages of a U-Net style (residual connections) in conjunction with regularisation techniques, resulting in superior segmentation accuracy. It employs four channels to compute 128×128-pixel images. The encoder is constructed by layering three residual blocks, each of which contains two 3x3 convolutional layers with Batch Normalisation and LeakyReLU activation. In this case, the number of filters is doubled to 64 and a MaxPooling2D is added to reduce spatial dimensions. The total number of filters is 128, which is then used with a MaxPooling2D.

The decoder employs two residual blocks for upsampling (residual_upconv_before and residual_upconv_after) in order to align with the encoder structure. This is followed by feature concatenation with the corresponding encoder block. The fourth block employs UpSampling2D, concatenates the second encoder block, and employs two 64-layer convolutional filters. The fifth block then collapses with the first encoder block and applies 32 filters to restore the image dimensions to their original size.

In order to prevent overfitting, dropout layers (with a 0.5 rate) are incorporated after each residual block. In order to classify each pixel into an agricultural field or non-agricultural area, the final result is predicted after an 11-layer convolution with softmax activation in the final output layer. This model is capable of effectively capturing more complex patterns while assuring training stability by utilising leaky ReLU activation, Batch Normalisation, Dropout, and residual blocks.

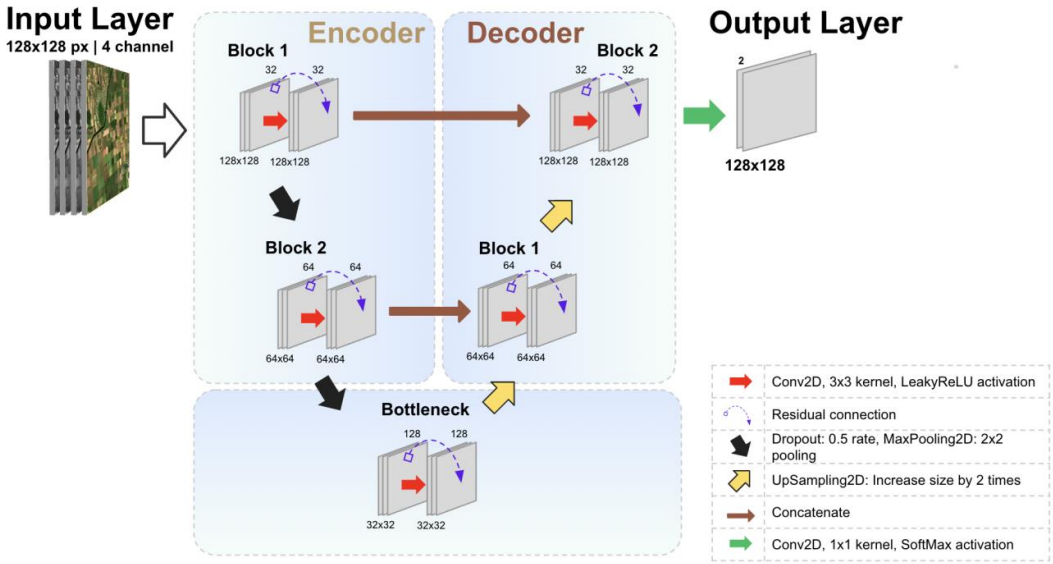


**Figure 7:** Architecture of Model #5 (U-Net with Residual Blocks, Batch Normalisation, Dropout, and LeakyReLU).

## 6. Results and discussion

The performance of the five U-Net model variations, each with distinct architectural enhancements, was evaluated based on multiple metrics, including validation accuracy, Intersection over Union (IoU), F1 Score, approximate training time, and model stability notes. These metrics provide insights into each model's segmentation performance, computational efficiency, and generalization capability. Table 1 summarizes these metrics for each model, highlighting the strengths and trade-offs of different architectural approaches.

**Table 1**: Performance Comparison of U-Net Model Variations

| Model | Accuracy | IoU | F1 Score | Approx. Training Time (hours) | Model Stability Notes |
|---|---|---|---|---|---|
| Model #1: U-Net with ReLU Activation | 79.67% | 65.15% | 78.90% | 0.06 | Stable |
| Model #2: U-Net with LeakyReLU and Batch Normalization | 89.23% | 82.79% | 90.58% | 0.43 | Overfitting |
| Model #3: U-Net with Residual Blocks | 80.59% | 80.13% | 88.97% | 0.06 | Stable |
| Model #4: U-Net with Residual Blocks, Batch Normalization, and Dropout | 87.65% | 78.39% | 87.89% | 0.08 | Overfitting |
| Model #5: U-Net with Residual Blocks, Batch Normalization, Dropout, and LeakyReLU | 86.73% | 78.38% | 87.88% | 0.08 | Stable |

### Model #1

**Accuracy**: The model showed a significant improvement in accuracy over the initial epochs, increasing from 70.06% in the first epoch to 79.67 and IoU 65.15% on the validation set by the tenth epoch. The model demonstrates stability throughout training, with a relatively high F1 Score (78.90%) and minimal overfitting.. This steady progression suggests that the model effectively learned basic feature representations but plateaued early, indicating a need for additional enhancements.
**Visual Analysis**:

The model demonstrates strong performance in segmenting large water bodies from satellite imagery (Fig.8), successfully delineating water resources from other landscape elements. However, in more complex regions with mixed terrain, the model exhibited occasional misclassifications, especially around areas with similar spectral characteristics. These errors indicate that while the model is capable of general segmentation tasks, it struggles with fine-grained differentiation, suggesting the need for further refinement.

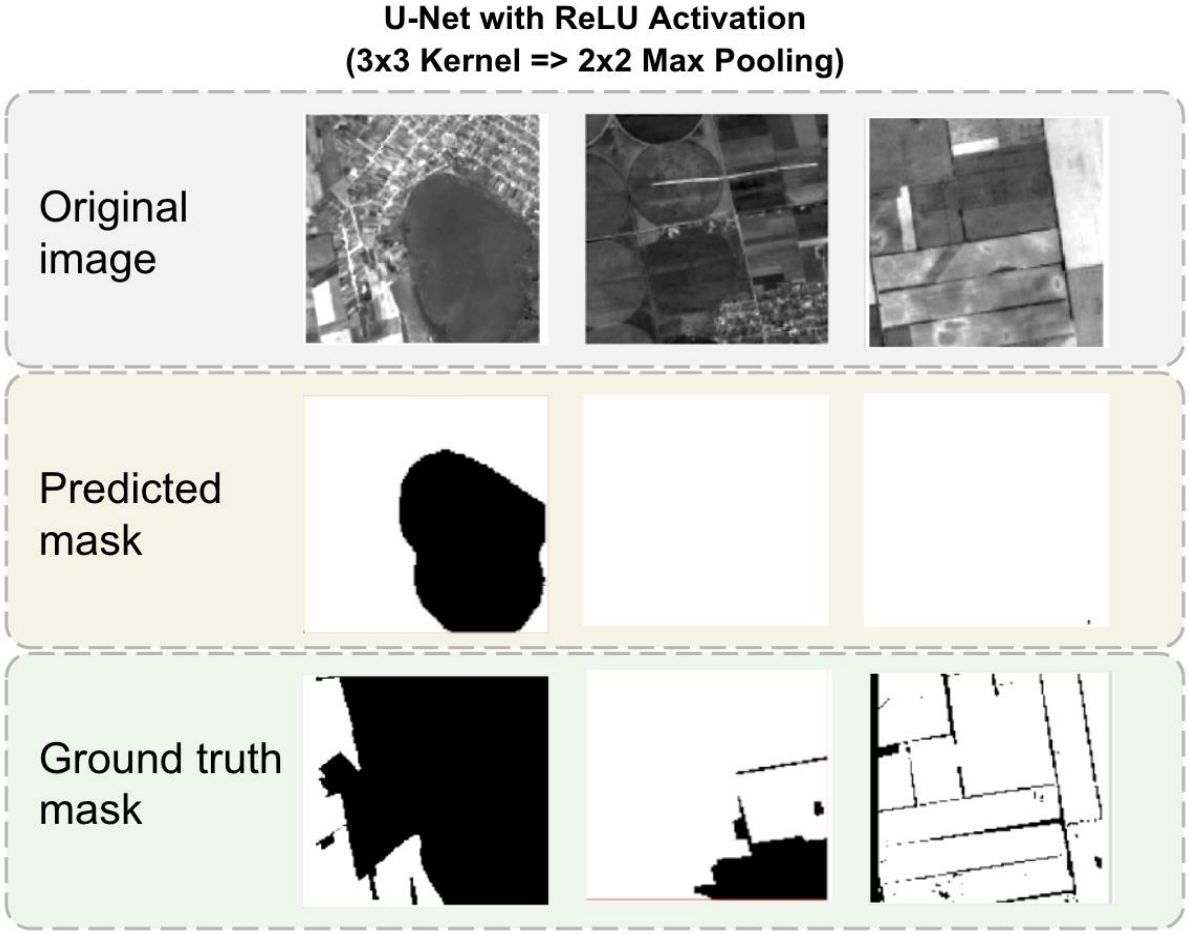


**Figure 8:** Results from Model #1: Original image; Predicted mask; Ground truth mask.

### Model #2

**Accuracy**: Improved validation accuracy from 81.24% to 89.23% and achieving the highes IoU 82.79% among the models, along with an F1 Score of 90.58%

**Visual Analysis:**

- Agricultural Fields: The model successfully segments fields but makes errors when terrain and textures of fields resemble other natural elements.
- Water Bodies: The model accurately identifies water bodies, showing clear boundaries around water features.
- Urban Areas: The model occasionally misclassifies buildings and other structures as fields, indicating a need for further refinement to improve class distinction.

### Model #3

**Accuracy:** The highest accuracy on the training data was achieved by the final epoch, reaching approximately 80.59%. with an IoU of 80.13% and an F1 Score of 88.97%.

With a quick training time (0.06 hours), Model #3 offers a balance between accuracy and stability, making it suitable for applications where both are valued.

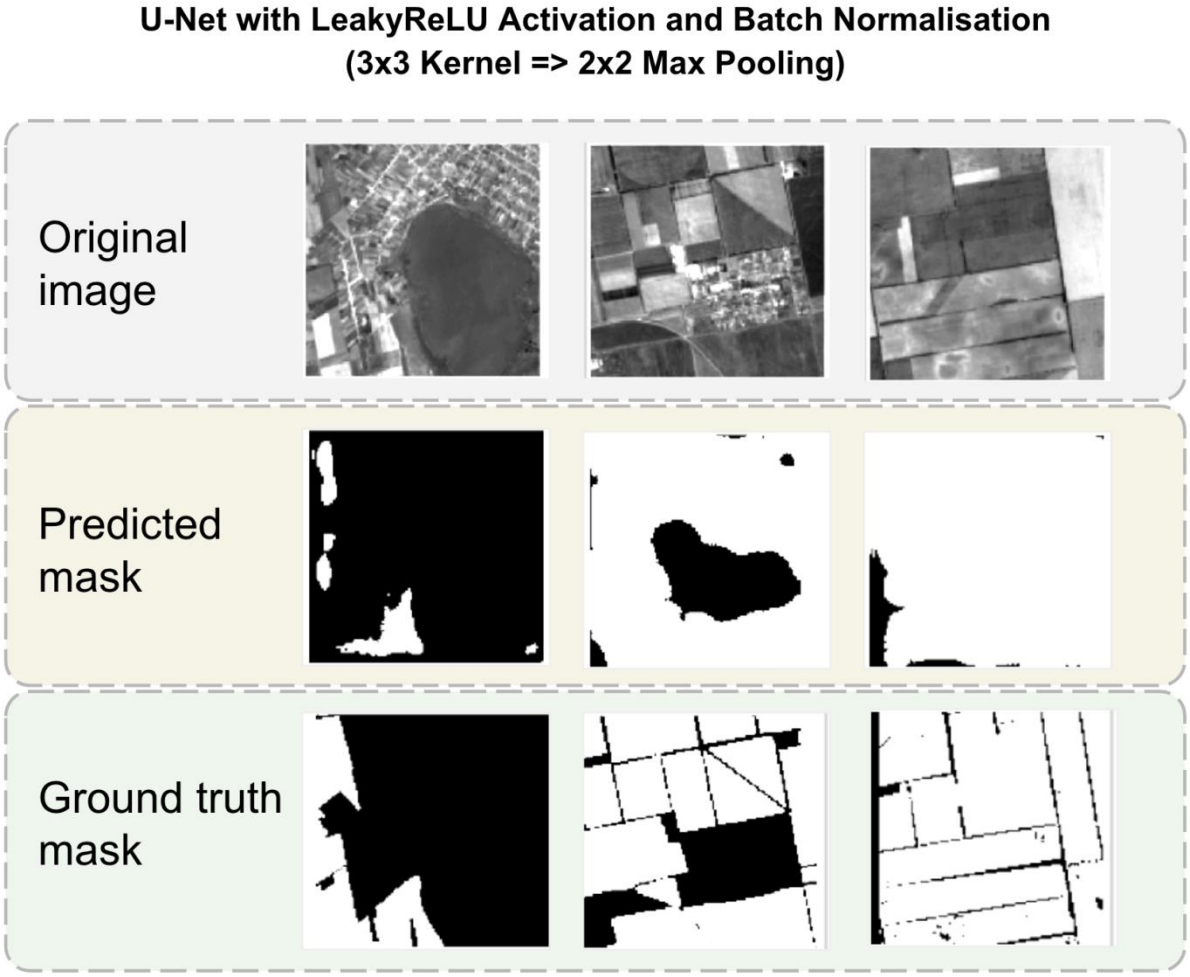


**Figure 9:** Results from Model #2: Original image; Predicted mask; Ground truth mask.

**Precision:** While the model demonstrates the ability to distinguish between classes, there is significant room for improvement. A high number of false positives indicate that the model often misclassifies non-cultivated areas as cultivated. Conversely, the number of false negatives, although lower, suggests that the model tends to miss some cultivated areas.

**Visual Analysis:**

With the inclusion of residual blocks, the model exhibits improved segmentation capabilities. However, certain regions still lack precision compared to the ground truth masks, indicating the need for further refinement. This analysis shows that Model #2 is capable of identifying segmented zones, but errors persist, particularly in classifying non-agricultural areas. To improve performance, deeper architectures or advanced training methods, such as transfer learning or additional data augmentation, may be required.

### Model #4

**Accuracy**: The highest training accuracy was achieved in the final epoch, reaching approximately 87.65% and an IoU of 78.39%, with an F1 Score of 87.89%.

**Precision**: Shows improvements in class recognition compared to previous models. However, a high number of false positives and false negatives indicate the need for further model optimization.

**Visual Analysis:**

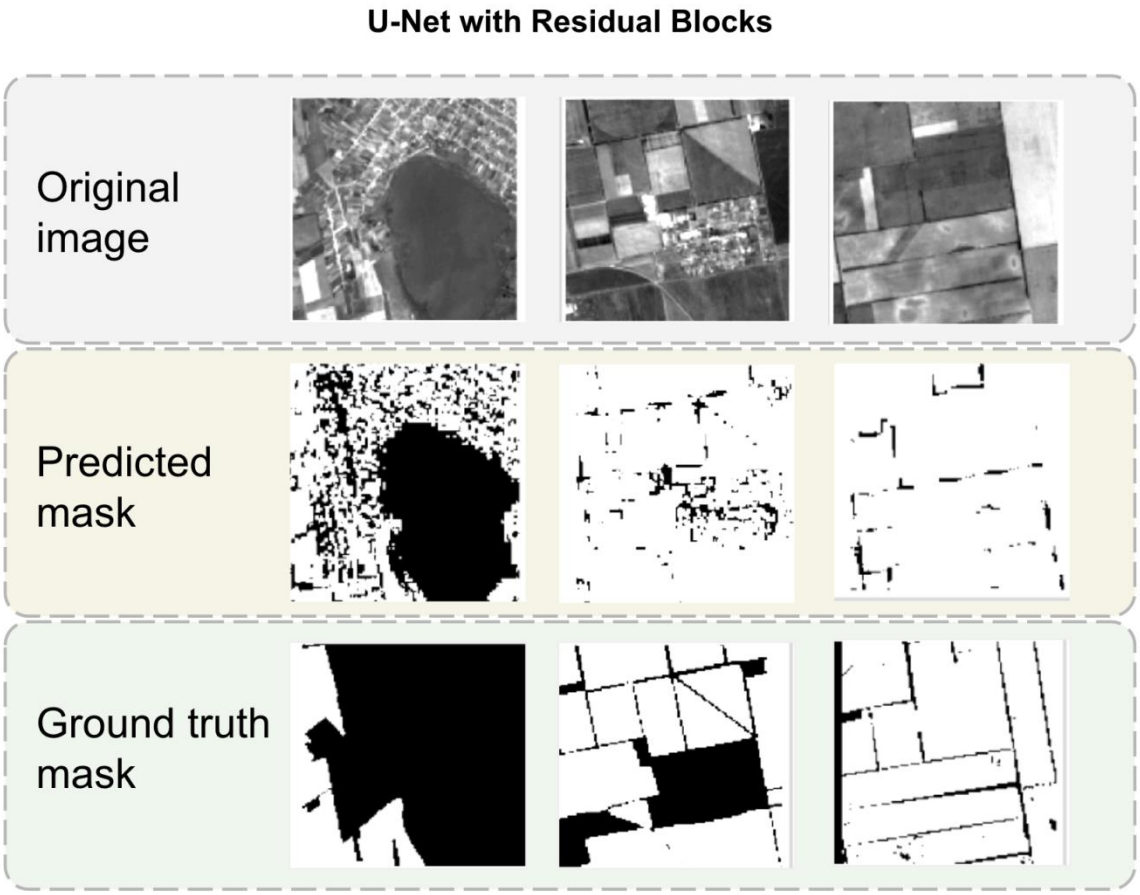


**Figure 10:** Results from Model #3: Original image; Predicted mask; Ground truth mask.

Predicted masks show that the model can detect objects, but there are inaccuracies in processing edges and finer details, especially in complex image regions. Model #3, which employs residual blocks with Batch Normalisation and Dropout, demonstrates better overall classification accuracy compared to previous iterations. The inclusion of Dropout helps prevent overfitting, and Batch Normalisation stabilizes and accelerates the training process. However, the high number of false positives and false negatives suggests that further improvements are required, particularly in boundary detection. Based on the visual analysis of the predicted masks, the model requires enhanced segmentation accuracy to achieve clearer and more precise object detection, with minimized classification errors.

### Model #5

**Accuracy**: The model achieved a peak training accuracy of 86.73% with an IoU of 78.38% and an F1 Score of 87.88% in one of the final epochs.. While this indicates reasonable performance, the model showed variability across epochs, suggesting potential overfitting.
**Precision**: The model demonstrates adequate classification capability, but there is significant room for improvement, particularly in reducing false positive predictions. The model tends to overestimate the presence of agricultural fields, which affects overall precision.
**Visual Analysis:**

The visual assessment of predicted segmentation masks indicates that the model effectively identifies large objects, such as extensive agricultural fields, but struggles with smaller and more detailed objects, resulting in misclassification. This limitation is particularly evident in regions with complex textures or mixed land cover. Despite employing residual blocks and Dropout, the model still produces inconsistencies along object boundaries, leading to misinterpretation of finer details. This suggests that further refinement is necessary, such as through hyperparameter tuning,

additional data augmentation techniques, or integrating advanced learning strategies like transfer learning.

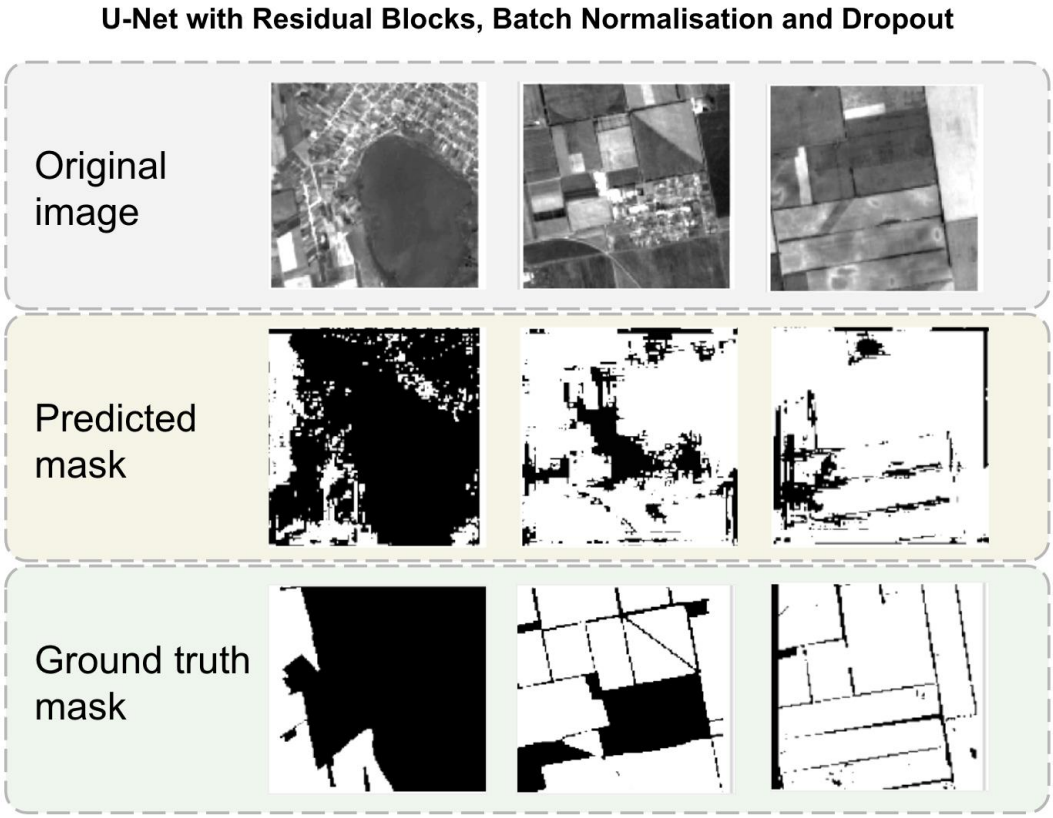


**Figure 11. Results from Model #4: Original image; Predicted mask; Ground truth mask**

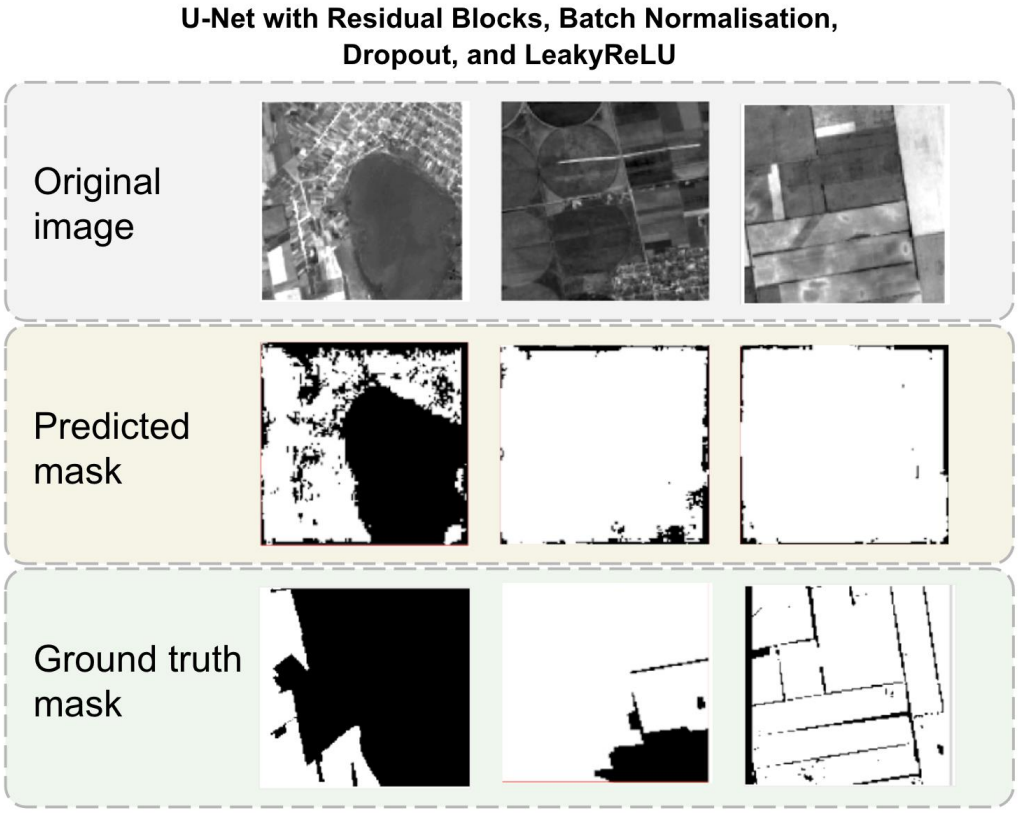


**Figure 12:** Results from Model #5: Original image; Predicted mask; Ground truth mask.

### 6.1. Discussion

The assessment of the five U-Net models revealed discrepancies in segmentation efficacy based on the architectural alterations used. Although residual blocks and Batch Normalisation enhanced training stability and mitigated gradient problems, the models had difficulties in attaining high accuracy in intricate areas, especially when textures were analogous across classes.

Models #2 and #3 saw improved feature propagation due to residual connections; nonetheless, the elevated incidence of false positives in uncultivated regions indicates that managing spatial context continues to pose a difficulty. Model #4 attained the best overall accuracy; yet, boundary misclassifications remained, highlighting the need for more sophisticated techniques to enhance edge detection and fine-grained segmentation.

Model #5, including all changes, demonstrated significant training robustness but inadequately captured tiny and intricate objects. This indicates a more extensive constraint in the models' capacity to manage complex spatial patterns and border areas, despite gradual architectural enhancements.

The findings indicate that future endeavours should prioritise the integration of sophisticated approaches such as attention mechanisms, which are more effective in capturing spatial relationships. Recent research indicates that attention-based models significantly enhance segmentation performance by enabling the model to concentrate on pertinent characteristics while disregarding extraneous ones [32]. Furthermore, hybrid architectures that combine CNNs with transformer-based models for enhanced contextual comprehension are becoming more common in the domain [30].

Investigating multi-scale feature extraction and using temporal information from satellite time series may improve the models' capacity to distinguish between classes with greater precision. Kussul et al. [31] showed that the use of temporal data significantly enhanced land cover categorisation accuracy in dynamic conditions.

Furthermore, the integration of generative adversarial networks (GANs) for data augmentation has shown potential in augmenting model resilience to overfitting and boosting generalisation capacities [33]. This may be especially advantageous in agricultural contexts where labelled data is often limited.

Although these U-Net versions provide a robust basis for agricultural land segmentation, more research is required to create more advanced models capable of addressing intricate segmentation challenges and enhancing practical applicability.

## 7. Conclusions

This research assessed five variants of U-Net architectures for the semantic segmentation of agricultural landscapes using high-resolution satellite images. The models included existing architectural improvements, including residual blocks, Batch Normalisation, and Dropout, to tackle issues associated with various agricultural landscapes. Despite these alterations enhancing training stability and segmentation performance, persistent limits across all models underscore the need for more refinement and the investigation of more sophisticated approaches.

U-Net with Residual Blocks, Batch Normalisation, and Dropout attained the best accuracy of 87.65%. The incorporation of Dropout markedly minimised overfitting, but Batch Normalisation enhanced stability and expedited training. Nonetheless, false positives and false negatives, especially in areas with unclear borders, continued to pose a concern. Incorporating multi-scale feature extraction or hierarchical network topologies may enhance the model's capacity to manage complicated regions. The examination across all models identified a persistent problem in precisely segmenting intricate borders and differentiating between classes with similar textures. Although the models shown significant promise for extensive segmentation tasks, their limitations suggest that more study is necessary to attain enhanced accuracy and resilience. Future research may concentrate on a more comprehensive investigation of hyperparameter optimisation, the incorporation of

attention processes, and the use of multi-temporal or multi-spectral data to enhance model flexibility and segmentation efficacy.

Furthermore, integrating supplementary assessment measures, like Intersection over Union (IoU), F1-score, and Precision-Recall curves, may provide a more thorough comprehension of model performance. Considering that accuracy alone may not encompass the intricacies of segmentation tasks, particularly with unbalanced class distributions, using these measures would provide more profound insights into the advantages and disadvantages of each model.

In conclusion, although the U-Net variations evaluated in this study demonstrate encouraging outcomes for agricultural land segmentation, future endeavours should prioritise the incorporation of more complex architectural elements, the exploration of advanced loss functions, and the utilisation of multi-modal data to enhance overall efficacy and relevance to practical land monitoring contexts. This study establishes a significant basis for creating more efficient and dependable instruments for agricultural land management, enhancing sustainable farming practices, and optimising land resource utilisation.

## Declaration on Generative AI

The authors have not employed any Generative AI tools.

## References


[1] National Report on the State of the Environment in Ukraine. - URL: https://mepr.gov.ua/diyalnist/napryamky/ekologichnyj-monitoryng/natsionalni-dopovidi-pro-stan-navkolyshnogo-pryrodnogo-seredovyshha-v-ukrayini/

[2] X. X. Zhu et al., "Deep learning in remote sensing: A review", 2017.

[3] O. Ronneberger, P. Fischer and T. Brox, "U-net: Convolutional networks for biomedical image segmentation", Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention, pp. 234-241, 2015.

[4] P. W. Owen et al., "The land parcel identification system: A useful tool to determine the eligibility of agricultural land—But its management could be further improved.", Oct. 2016.

[5] G. Sumbul, M. Charfuelan, B. Demir and V. Markl, Bigearthnet: A large-scale benchmark archive for remote sensing image understanding, Proc. IEEE Int. Geosci. Remote Sens. Symp., pp. 5901-5904, Jul. 2019, doi.org/10.1109/IGARSS.2019.8900532

[6] P. Helber, B. Bischke, A. Dengel and D. Borth, "EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification", IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 12, no. 7, pp. 2217-2226, 2019.

[7] X. X. Zhu et al., "So2Sat LCZ42: A benchmark data set for the classification of global local climate zones [Software and Data Sets]", IEEE Geosci. Remote Sens. Mag., vol. 8, no. 3, pp. 76-89, Sep. 2020.

[8] V. Sitokonstantinou, I. Papoutsis, C. Kontoes, A. Lafarga Arnal, A. P. Armesto Andrés and J. A. Garraza Zurbano, "Scalable parcel-based crop identification scheme using sentinel-2 data time-series for the monitoring of the common agricultural policy", Remote Sens., vol. 10, no. 6, 2018.

[9] N. Kussul, M. Lavreniuk, S. Skakun and A. Shelestov, "Deep learning classification of land cover and crop types using remote sensing data", IEEE Geosci. Remote Sens. Lett., vol. 14, no. 5, pp. 778-782, May 2017.

[10] A. Stoian, V. Poulain, J. Inglada, V. Poughon and D. Derksen, "Land cover maps production with high resolution satellite image time series and convolutional neural networks: Adaptations and limits for operational systems", Remote Sens., vol. 11, no. 17, 2019, URL: https://www.mdpi.com/2072-4292/11/17/1986.

[11] Zhang, C., & Wang, Y. (2020). Remote Sensing Object Detection in the Deep Learning Era—A Review. Remote Sensing, 12(6), 953. doi:10.3390/rs12060953.

[12] Y. He, E. Lee and T. A. Warner, Continuous annual land use and land cover mapping using AVHRR GIMMS NDVI3g and MODIS MCD12Q1 datasets over China from 1982 to 2012, 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 2016, pp. 5470-5472, doi: 10.1109/IGARSS.2016.7730425.
[13] K. Karra, C. Kontgis, Z. Statman-Weil, J. C. Mazzariello, M. Mathis and S. P. Brumby, "Global land use / land cover with Sentinel 2 and deep learning," 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 2021, pp. 4704-4707, doi: 10.1109/IGARSS47720.2021.9553499.
[14] P. Wang, L. Wang, H. Leung and G. Zhang, Super-Resolution Mapping Based on Spatial–Spectral Correlation for Spectral Imagery, IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 3, pp. 2256-2268, March 2021, doi: 10.1109/TGRS.2020.3004353.
[15] M. Drusch, U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort, A. Meygret, F. Spoto, O. Sy, F. Marchese, P. Bargellini, Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services, Remote Sensing of Environment, Volume 120, 2012, Pages 25-36, https://doi.org/10.1016/j.rse.2011.11.026.
[16] J. Long, E. Shelhamer and T. Darrell, Fully convolutional networks for semantic segmentation, 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 3431-3440, doi: 10.1109/CVPR.2015.7298965.
[17] J. Donahue et al., Long-term recurrent convolutional networks for visual recognition and description, 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 2625-2634, doi: 10.1109/CVPR.2015.7298878.
[18] I. Goodfellow, Y. Bengio and A. Courville, Deep Learning, Cambridge, MA, USA:MIT Press, 2016.
[19] V. P. Yele, R. R. Sedamkar and S. Alegavi, Systematic Analysis of Effective Segmentation and Classification for Land Use Land Cover in Hyperspectral Image using Deep Learning Methods: A Review of the State of the Art : Reviewing Deep Learning Techniques for Land Use and Cover in Hyperspectral Images, 2024 20th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP), Babol, Iran, Islamic Republic of, 2024, pp. 1-8, doi: 10.1109/AISP61396.2024.10475229.
[20] A.Jamali, S.K.Roy, L.Hashemi Beni, B.Pradhan, J.Li, , P.Ghamisi, Residual wave vision U-Net for flood mapping using dual polarization Sentinel-1 SAR imagery. International Journal of Applied Earth Observation and Geoinformation 127, March 2024, 103662. https://doi.org/10.1016/j.jag.2024.103662 .
[21] S.A.Yoganathan,et al., Generating synthetic images from cone beam computed tomography using self-attention residual UNet for head and neck radiotherapy. Physics and Imaging in Radiation Oncology VOLUME 28, 100512, OCTOBER 2023. https://doi.org/10.1016/j.phro.2023.100512.
[22] L.L. Janneh, Z. Youngjun, M. Hydara, Z. Cui, Deep learning-based hybrid feature selection for the semantic segmentation of crops and weeds. ICT Express 10 (2024) 118–124. https://doi.org/10.1016/j.icte.2023.07.008.
[23] A. Ghaznavi, M. Saberioon, J. Brom, S. Itzerott, Comparative performance analysis of simple U-Net, residual attention U-Net, and VGG16-U-Net for inventory inland water bodies. Applied Computing and Geosciences 21 (2024) 100150. https://doi.org/10.1016/j.acags.2023.100150 .
[24] Vik. Hnatushenko, Vo. Hnatushenko, V. Kashtan, C. Heipke, 2023: Detection of Forest Fire Consequences on Satellite Images using a Neural Network. In: Kersten T., Tilly N. (Eds.), 43. Wissenschaftlich-Technische Jahrestagung der DGPF e.V. - München, Publikations DGPF, Vol. 31
https://www.dgpf.de/src/tagung/jt2023/proceedings/paper/15_dgpf2023_Hnatushenko_et_al.pdf
[25] C.Heipke & F.Rottensteiner, Deep learning for geometric and semantic tasks in photogrammetry and remote sensing. Geo-spatial Information Science, 23(1), 2020, pp.10-19, https://doi.org/10.1080/10095020.2020.1718003

[26] G. Kaplan & U. Avdan, Sentinel-1 and Sentinel-2 Data fusion for wetlands mapping: Balikdami, Turkey. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., 42(3), 2018, 729-734, https://doi.org/10.5194/isprs-archives-XLII-3-729-2018.
[27] V. Kashtan & V. Hnatushenko, Deep Learning Technology for Automatic Burnt Area Extraction Using Satellite High Spatial Resolution Images. In: Babichev, S., Lytvynenko, V. (eds) Lecture Notes in Data Engineering, Computational Intelligence, and Decision Making. ISDMCI 2022. Lecture Notes on Data Engineering and Communications Technologies, 149. Springer, Cham. https://doi.org/10.1007/978-3-031-16203-9_37.
[28] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning," Cambridge, MA: MIT Press, 2016. https://doi.org/10.5555/3298689.
[29] A. Kamilaris & F. X. Prenafeta-Boldú, Deep Learning in Agricultural Remote Sensing Applications: A Meta-Analysis and Review. Remote Sensing, 10(12), 2018, 1954. doi:10.3390/rs10121954.
[30] J. Chen, Y. Zhang & L. Wang, Transformer-based Deep Learning for Remote Sensing Image Classification: A Review. Remote Sensing, 14(5), 2022, 1123. doi:10.3390/rs14051123.
[31] N. Kussul, M. Lavreniuk, S. Skakun, Temporal Deep Learning Models for Land Cover Classification Using Satellite Data. IEEE Transactions on Geoscience and Remote Sensing, 60, 2022, 1-12. doi:10.1109/TGRS.2021.3088254.
[32] X. Zhang, Y. Liu, Y. Wang, Attention U-Net for Medical Image Segmentation. Medical Image Analysis, 67, 2021, 101855. doi:10.1016/j.media.2020.101855.
[33] P. Isola, J.-Y. Zhu, T. Zhou, A. A. Efros, Image-to-Image Translation with Conditional Adversarial Networks. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 5967-5976. doi:10.1109/CVPR.2017.632.
[34] V. Hnatushenko, Vik. Hnatushenko, Recognition of High Dimensional Multi-Sensor Remote Sensing Data of Various Spatial Resolution. IEEE Third International Conferenceon Data Stream Mining&Processing (DSMP), Lviv, Ukraine, 2020, pp. 262-265, doi: 10.1109/DSMP47368.2020.9204186
[35] Liu. Mushui, Dan. Jun, Lu. Ziqian, Yu. Yunlong, Li. Yingming, Li. Xi, CM-UNet: Hybrid CNN-Mamba UNet for Remote Sensing Image Semantic Segmentation (2024). doi:10.48550/arXiv.2405.10530.
[36] Su. Zhongbin, Li. Wei, Ma. Zheng, Gao. Rui, An improved U-Net method for the semantic segmentation of remote sensing images. Applied Intelligence(2022) 52. 1-13. doi:10.1007/s10489-021-02542-9.
[37] Yan. Chen, Quan. Dong, Xiaofeng. Wang, Qianchuan. Zhang, Menglei. Kang, Wenxiang. Jiang, Mengyuan. Wang, Lixiang. Xu, Chen. Zhang, A Hybrid-Attention Semantic Segmentation Network for Remote Sensing. Cognitive Computation (2022). doi:10.1007/s13042-022-01517-7.
[38] Jiacheng. Li, Deep Attention U-Net: A Network Model with Global Feature Perception Ability (2023). doi: 10.48550/arXiv.2304.10829.
[39] He. Shumeng, Houqun. Yang, Xiaoying. Zhang, Xuanyu. Li. MFTransNet: A Multi-Modal Fusion with CNN-Transformer Network for Semantic Segmentation of HSR Remote Sensing Images. Mathematics (2023) 11(3):722. doi:10.3390/math11030722.
[40] Huacong. Zhou, Xiangling. Xiao, Huihui. Li, Xiaoyong. Liu, Peng. Liang, Hybrid Shunted Transformer Embedding U-Net for Remote Sensing Image Semantic Segmentation, Neural Computing and Applications (2024). doi:10.1007/s00521-024-09888-4.
[41] Yang. Yang, Shunyi. Zheng, AMMUNet: Multi-Scale Attention Map Merging for Remote Sensing Image Segmentation (2024). doi: 10.48550/arXiv.2404.13408.