Pakistan Journal of Life and Social Sciences
www.pjlss.edu.pk

https://doi.org/10.57239/PJLSS-2026-24.1.0046

RESEARCH ARTICLE

Yield Monitoring Method Based on Neural Networks Vision Transformers

Rudoy Dmitry[1], Gapon Nikolay[1], Odabashyan Mary[1]*, Zhdanova Marina[1], Azhinov Alexander[1], Olshevskaya Anastasiya[1], Kryazhevskikh Maria[1], Marchenko Sergey[1]

[1]Don State Technical University, Rostov-on-Don, Russian Federation

ARTICLE INFO

Received: Apr 18, 2026

Accepted: May 18, 2026

Keywords

Soil
Agriculture
Neural Network
Vision Transformer
Agricultural Image
Segmentation
Digitalization
Agroecosystem Stability
Data Analysis
Agricultural Land

***Corresponding Author:**
modabashyan@donstu.ru

ABSTRACT

Crop area assessment is an important task in agriculture and can be used to obtain accurate in-formation on many issues such as crop yield assessment, food policy development, adjustment of planting patterns, which is of great importance to ensure national food security. This paper discusses crop yield monitoring based on an image segmentation method based on the Vision Transformer (ViT) neural network. The network divides the image into small fragments. These fragments are processed using transformer layers, and the self-attention mechanism helps the model focus on important areas of the image, even if these areas are far from each other. Due to this ability to identify long-term dependencies, ViT is especially effective when working with complex images with intricate details. The proposed method has shown high efficiency in com-plex agricultural image segmentation tasks. This method outperforms state-of-the-art methods based on deep neural networks by 10-15% on average. Soil analysis is carried out for the identi-fied unseeded areas to identify the causes of low crop germination in this area. The article exam-ines the most reliable and effective methods for determining agrochemical parameters of soil.

INTRODUCTION

Agriculture is one of the most widespread and rapidly developing sectors in the world, present in nearly every country [1]. Approximately 1.1 billion people are engaged in this industry. Land serves as the primary means of production in agriculture, without which it cannot be realized. Given this fact, it is concerning to observe the irrational use of land resources amid growing human demands. The increasing global population requires more food and additional territories for its cultivation. While industrial facilities, vehicle emissions, and electricity production were once considered the main sources of environmental pollution, agriculture has now joined their ranks. Awareness of the considerable damage that human activities in this sector inflict on ecosystems has been recognized for over 40 years. Since 1980, the United Nations has identified agricultural damage as one of the four most significant threats to the environment. [1]

Agricultural land constitutes approximately 13% of the Earth's total landmass, and there is a persistent effort to expand this area through methods such as wetland drainage, desert irrigation, and deforestation [1]. However, this pursuit of increasing agricultural territory often leads to substantial losses, as previously cultivated lands undergo degradation. Each year, approximately 7 million hectares of land become unsuitable for farming, resulting in a decrease in available arable land to about 2.5 billion hectares, down from 4.5 billion hectares prior to the rapid growth of the agricultural sector. It would be more beneficial to focus resources on the conservation of the most productive soils rather than attempting to improve less fertile areas.

Moreover, in the quest for higher yields and profits, there is often a neglect of soil diagnostics and the appropriate selection of agricultural equipment, as well as a disregard for crop rotation practices. This neglect ultimately diminishes the land's future viability for cultivation.

In today's world, where technology increasingly permeates various aspects of life, the automation and optimization of agricultural processes are becoming increasingly vital. A key focus area is the monitoring and control of crop yields. The application of artificial intelligence in agriculture can positively influence soil health. Soil condition should be constantly monitored as it is a very dynamic and open system and can be affected by the environment. Manually collecting and analyzing field condition data is often time consuming and not always an effective solution. Consequently, developing a software tool capable of automatically recognizing and classifying unplanted areas in agricultural fields has become a pressing necessity.

The relevance of this research lies in the development of an intelligent decision-making system for recognizing and classifying unplanted areas in agricultural fields, which has the potential to significantly increase crop yields. Currently, manual inspection and field monitoring remain prevalent practices, yet they are inefficient and associated with various challenges.

The manual monitoring of agricultural fields presents several limitations, particularly given the extensive areas involved. This method can be time-consuming, requiring significant effort from agronomists and laborers. It is also susceptible to subjectivity and human error, which can lead to inconsistencies in assessments. Furthermore, manual monitoring typically covers only limited sections of a field, leaving many areas unexamined and vulnerable to issues. These shortcomings highlight the pressing need for automated monitoring systems capable of enhancing both the efficiency and accuracy of crop field analysis. Unmanned aerial vehicle (UAV) platforms have recently garnered significant attention in various applications, particularly in precision agriculture, primarily due to their relatively low cost and high capability to capture areas with very high spatial resolution. UAV imagery is predominantly utilized to replace traditional visual inspections of agricultural landscapes, facilitating timely decision-making and enhancing both productivity and sustainability in agriculture. The type of techniques applied to remotely sensed data comes from artificial intelligence. Deep learning is rapidly gaining momentum as an image processing and data analysis technique. Deep learning is a type of machine learning technique that is built as a deeper type of artificial neural network that allows hierarchical representation of data. Despite the fact that deep neural networks require high computational resources and their effectiveness is largely related to the quality of the training data used, they have demonstrated impressive achievements in solving various tasks, including image classification, semantic segmentation, object detection, and several others. The deep learning method used in this study to address a specific task allows for the representation of various regions of the scene according to their target characteristics. These characteristics serve as the foundation for training deep learning models and facilitate the effective differentiation of detected vegetation from other objects in the image. Semantic segmentation can assign a class label to each pixel in an image and return as output another image, typically with the same input data size, where each pixel is associated with a single class. This output, commonly referred to as a topic map, can help in fully understanding the scene, which in turn can help many applications. In agriculture related tasks, most semantic segmentation processes with deep neural networks utilize RGB (red-green-blue) images or include a combination with other information to help solve a particular problem. Mask R-CNN (Region-based Convolutional Neural Network) [1] architectures have been proposed for detection and segmentation in combination with RGB and RGB+HSI (hue-saturation-intensity) images. SegNet architecture (Convolutional neural network architecture designed for semantic segmentation) [2] was also compared with FCN (Fully convolutional network) method in RGB dataset for rice canopy identification. Another study implementing RGB and near-infrared (NIR) information from a sensor embedded in a ground robot was able to transfer knowledge from a network trained on another crop to a semantic weed segment with the SegNet-Basic architecture. FCN combined with RGB-based images from a UAV was used to segment a winter wheat ear [3]. The assessment of field conditions in a remote sensing image provides important information about the agricultural area. Semantic field segmentation by deep networks is a new and improved computational approach to

accurately separate vegetation from other objects in an image-scene. As an advantage, it should provide accurate state estimation while requiring low human effort. CNNs rely on the use of convolutional layers and filters that sequentially analyze the image, focusing on local details. In contrast, ViT (the Vision Transformer) [4] is an approach that brings the Transformers architecture, widely used in natural language processing, to the field of image analysis. The key idea of ViT is to partition an image into small patches, which are then processed as sequences of tokens. This architecture is highly flexible and scalable, allowing models to be trained on large amounts of data with minimal changes to the structure. One of the main advantages of ViT is the model's ability to account for global dependencies in an image through attention mechanisms. Unlike convolutional neural networks (CNNs) [5], which focus on local features, ViT effectively integrates information from different parts of the image, which makes it particularly useful in semantic segmentation tasks where precise separation of objects and regions in the image is required. In this study, Agriculture-Vision 2020, which contains aerial images of agricultural land and annotations reflecting different anomalies and vegetation types, was used as the main dataset. This choice allows us to demonstrate the potential of ViT [6] for semantic segmentation of agricultural data. The use of ViT in agriculture opens up new opportunities for increasing crop yields and sustainable land use. With its ability to effectively incorporate global dependencies into images, ViT is becoming an indispensable tool in the arsenal of modern researchers and agrotechnologists.

MATERIALS AND METHODS

Dataset

The dataset [7] is designed for agricultural applications using computer vision. It includes nearly 95,000 high-resolution images from 3,432 farms in the United States, labeled to identify key agricultural problems such as drought or nutrient deficiencies.

Researchers marked up nine types of anomalies in the images, namely double cropping, drought, seed release, nutrient deficiency, planter skipping, storm damage, water, waterway and weed accumulation. All of these factors have a significant impact on field conditions and final yield. As a pilot experiment, the researchers tested state-of-the-art models for semantic segmentation on the data.

The dataset is designed for tasks aimed at improving semantic segmentation models, especially for solving problems in agriculture.

Agriculture-Vision dataset [7] includes 94,986 high-resolution field images (an example is shown in Figure 1) from 3,432 farms with RGB and NIR channels at 10 cm per pixel 512 by 512 pixels.

This dataset contains six types of annotations: Cloud shadow, Double plant, Planter skip, Standing Water, Waterway and Weed cluster. These types of field anomalies have great impacts on the potential yield of farmlands, therefore it is extremely important to accurately locate them. In the Agriculture-Vision dataset, these six patterns are stored separately as binary masks due to potential overlaps between patterns. Users are free to decide how to use these annotations.

Each field image has a file name in the format of (field id)_(x1)-(y1)-(x2)-(y2).(jpg/png). Each field id uniquely identifies the farmland that the image is cropped from, and (x1, y1, x2, y2) is a 4-tuple indicating the position in which the image is cropped. Please refer to our paper for more details regarding how we construct the dataset.

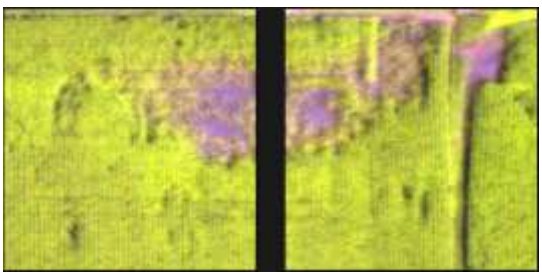

Figure 1: Example of images from Agriculture-Vision dataset in RGB format.

The annotations conducted by volunteers cover nine agricultural problems such as drought, nutrient deficiencies, and storm damage. This dataset is designed to enhance the application of computer vision in monitoring agricultural fields. Table 1 presents a comparison of the Agriculture-Vision dataset with similar datasets based on various criteria.

Table 1: Comparison of Agriculture-Vision with similar datasets.

Dataset	Image	Classes	Labels	Tasks	Image size	Pixels	Channels	Resolution (GSD)
Aerial images								
Inria Aerial Image [8]	180	2	180	seg.	5000×5000	4.5B	RGB	30 cm/px
DOTA [9]	2,806	14	188,282	det.	≤4000×4000	44.9B	RGB	various
iSAID [10]	2,806	15	655,451	seg.	≤4000×4000	44.9B	RGB	various
AID [11]	10,000	30	10,000	els.	600×600	3.6B	RGB	50-800 cm/px
DeepGlobe Building [12]	24,586	2	302,701	det./seg.	650×650	10.4B	9 bands	31-124 cm/px
EuroSAT [13]	27,000	10	27,000	els.	256×256	1.77B	13 Bands	30 cm/px
SAT-4 [14]	500,000	4	500,000	els.	28×28	0.39B	RGB, NIR	600 cm/px
SAT-6 [14]	405,000	6	405,000	els.	28×28	0.32B	RGB, NIR	600 cm/px
Agricultural images								
Crop/Weed discrimination [15]	60	2	494	seg.	1296×966	0.08B	RGB	N/A
Sensefly Crop Field [16]	5,260	N/A	N/A	N/A	N/A	N/A	NRG, Red edge	12.13 cm/px
DeepWeeds [17]	17,509	1	17,509	els.	1920×1200	40.3B	RGB	N/A
Agriculture-Vision [18]	**94,986**	**9**	**169,086**	**seg.**	**512×512**	**22.6B**	**RGB, NIR**	**10/15/20 cm/px**

The classes into which the dataset was divided and presented in Figure 2. From the diagram we can conclude that there is a strong imbalance of data, which can affect the final result of the neural network.

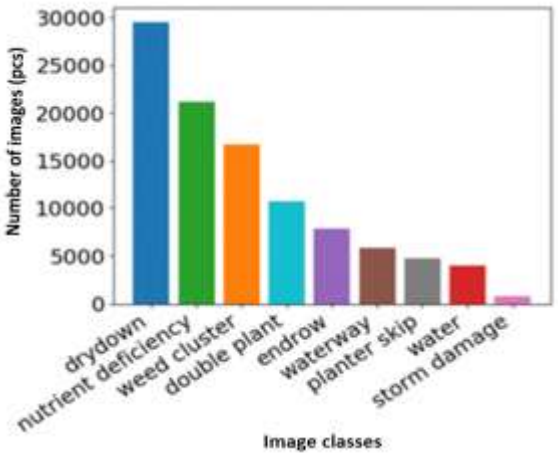

Figure 2: Number of images containing each of 9 classes.

Data Preparation

The Agriculture-Vision dataset comprises a substantial collection of labeled images; however, it exhibits a significant class imbalance due to the uneven distribution of various types of anomalies and normal regions. Figure 2 shows that some classes, such as "healthy vegetation", can occur much more frequently than "soil erosion" or "drought".

Unbalanced data is a common problem in datasets where one class is represented much more often than others. This can lead to serious problems when training machine learning models.

This creates difficulties in model training as algorithms tend to favor more frequent classes while ignoring rare ones, resulting in poor performance on the less represented class. In this case, standard metrics to account for the error will not work correctly, creating misleading data about the training and performance of the neural network.

To combat the imbalance, it is rational to use various methods such as:

1.data balancing;

2.data augmentation;

3.use of generative models;

4.metrics to evaluate the quality of the models.

Data balancing is an effective method. We can distinguish weighting of weights and sample reduction [19]. In the experiments, weighting of weights proved to be better, so in the future we used a combination of weights (0.04, 0.05, 0.06, 0.08, 0.10, 0.12, 0.15, 0.20) and the loss function CrossEntropyLoss.

Data augmentation and the use of generative models, such as a Generative Adversarial Network (GAN) models and its variants [20], was not applied in this work, as the total image volume was considered sufficient. At the same time, generating images from initially noisy data may entail a large amount of poor-quality training and test data. This will not allow to train the model efficiently and qualitatively. Another important factor is the computational power required for image generation. At the same time, the authors still believe that these methods can be extremely effective and admit the possibility of using them in future works.

ViT Architecture

ViT architecture [4] partitions an image into small fragments, in this paper a 16×16 pixel partitioning is applied. Each fragment is converted into a one-dimensional vector and a linear projection is applied to it to form a token, similar to the way words are processed in NLP transformers. These tokens are fed into the transformer model, where self-aware layers capture the relationships between fragments, allowing the model to focus on important features of the image. ViT operation concept and transformer unit in ViT presented on Figure 3.

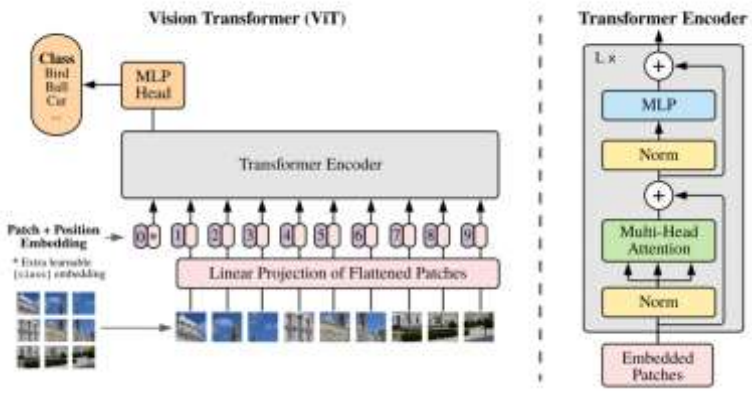

Figure 3: ViT operation concept and transformer unit in ViT [4].

Key Components of the ViT Architecture*:*

Patch Embedding: an image is divided into fragments, each of which is aligned and projected into the embedding space.

Patch Embedding converts images into sequences of tokens for transformer operation. The image is divided into patches of fixed size, each of which is projected into a vector of a given dimensionality. This is accomplished using convolution, which performs partitioning and projection simultaneously. The resulting tokens are arranged in a sequence compatible with the transform blocks. The module simplifies image processing and prepares the images for feature extraction.

Position coding: since Transformers do not have a built-in understanding of the spatial position of fragments, position coding is added to fragment embeddings for preservation.

The Transformer Block ViT includes two key modules: a self-awareness mechanism and a multilayer perseptron (MLP). These modules are combined with residual connectivity and normalization to ensure learning stability and efficiency.

1. Self-Attention: Allows the model to consider the relationships between all tokens in a sequence, which is important for analyzing global image features.

2. MLP: Consists of two linear layers with intermediate nonlinearity (e.g., GELU). The first layer increases the dimensionality of the hidden representation, while the second layer returns it to the original dimensionality.

3. Residual coupling and normalization: Each module (self-explanation and MLP) is augmented with residual coupling and normalization. This improves the convergence of the model and prevents gradient fading.

After processing by Transformer, the tokens are converted back to the spatial structure of the image, preserving the original resolution, to perform tasks like image segmentation or reconstruction.

The strength of ViT lies in Transformer's ability to model long-term dependencies, which allows the model to focus on both local and global features of the image. This allows ViT to achieve high performance, especially when training on large datasets.

RESULTS

In the early days of deep learning, CNNs were the gold standard for tasks such as image classification, object detection, and segmentation. However, the performance of these networks declined as they became deeper and more complex. ViT, on the other hand, has shown that an alternative approach that utilizes self-monitoring mechanisms can outperform CNNs, especially on large datasets. A key advantage of the ViT model is its ability to model relationships across the entire image, unlike CNNs that focus on local spatial hierarchies.

When training ViT models, images are divided into fragments and each fragment is processed as a token, similar to the way words are processed in natural language processing tasks. These fragments are processed using transform layers, and the self-awareness mechanism helps the model to focus on important regions of the image, even if these regions are far apart. This ability to detect long-term dependencies makes ViT particularly effective when dealing with complex images with intricate details.

An example of the neural network operation is shown in Figure 4. The figure shows frames from the Agriculture-Vision dataset [7]. Considering that the images from this dataset are 512 by 512 pixels in size, and each pixel represents 10 cm, the dimensions of the field area limited by the image are 51.2 by 51.2 meters.

After processing by the proposed method, the area of the image where the culture did not sprout is highlighted in red.

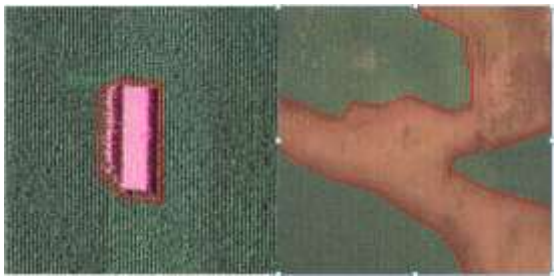

Figure 4: Processed images.

There are several key metrics for assessing the quality of image segmentation performance, each reflecting different aspects of the accuracy and efficiency of the method. These metrics allow an

objective comparison of different algorithms and determine their applicability in different environments, including agriculture.

1) Pixel Accuracy (PA) measures the proportion of pixels that were correctly classified by the model. The formula for calculating Pixel Accuracy is:

$$PA = \frac{\sum_{i=0}^{K} p_{ii}}{\sum_{i=0}^{K} \sum_{j=0}^{K} p_{ij}}$$

where p_{ii} — number of pixels correctly classified as class i, and p_{ij} — number of pixels belonging to class i, but classified as class j.

2) Mean Pixel Accuracy (MPA) represents the average accuracy across all classes, which allows for the imbalance in data between different classes to be accounted for:

$$MPA = \frac{1}{K+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij}}$$

where K — number of classes including background.

Intersection over Union (IoU), also known as the Jaccard index, is one of the most common metrics for assessing segmentation quality. It is calculated as the ratio of the intersection area of the predicted segmentation area and the true area to the area of their union:

$$IoU = \frac{|A \cap B|}{|A \cup B|}$$

where A and B — true and predicted segmentation areas, respectively. Value of IoU vary from 0 to 1, where 1 corresponds to a complete overlap of areas.

Mean Intersection over Union (mIoU) represents the average IoU across all classes:

$$mIoU = \frac{1}{K+1} \sum_{i=0}^{k} IoU_i$$

where IoU_i — IoU metric value for i-th class.

Precision and Recall are classical metrics for evaluating classification accuracy, including segmentation tasks. The formulas for their calculation are given below:

$$Precision = \frac{TP}{TP + FP}, \qquad Recall = \frac{TP}{TP + FN}$$

where TP — number of true positive predictions, FP — number of false positive predictions, FN — number of false negative predictions.

F1-Score is a harmonic average between Precision and Recall, and it is often used to evaluate models when both accuracy and completeness are important to consider:

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

There are several key metrics for assessing the quality of image segmentation performance, each reflecting different aspects of the accuracy and efficiency of the method. These metrics allow you to objectively compare different algorithms and determine their applicability in different environments, including agriculture.

Table 2 presents the evaluation results of state-of-the-art segmentation methods based on deep neural networks. These methods have shown high efficiency in complex agricultural image

segmentation tasks. The hardware used to conduct the experiment included a high-performance IBM PC-compatible personal computer equipped with an Intel Core i9 processor, which provides high computing power and multitasking. The NVidia GeForce GTX 4090 graphics card included in the hardware has advanced graphics capabilities, which allows for efficient processing of complex visual data and resource-intensive tasks such as rendering and machine learning. 32 GB of RAM ensures fast processing of large amounts of data and maintains stable system operation even when running multiple resource-intensive applications simultaneously.

Table 2: Results of evaluation of modern segmentation methods based on deep neural networks.

Method	Pixel Accuracy	Mean Pixel Accuracy	IoU	mIoU	Precision	Recall	F1-Score
The proposed method	98.01%	0.98	0.98	0.88	99.09	98.96	98.96
CNN [21]	88.51%	0.72	0.68	0.74	85.11	84.52	84.32
SegNet [22]	89.84%	0.71	0.71	0.71	87.88	87.81	87.85
DeepLab [23]	95.08%	0.94	0.94	0.81	96.01	96.96	96.98
PSPNet [24]	96.01%	0.96	0.96	0.76	97.21	97.55	96.01
Mask R-CNN [25]	93.01%	0.75	0.75	0.75	95.22	94.57	94.53
Autoencoder [26]	88.02%	0.81	0.74	0.73	87.86	87.57	83,25
CNN + CRF [27]	89.05%	0.71	0.71	0.61	85.45	83.58	84.17
CNN + Active Contours [28]	90.05%	0.77	0.75	0.75	89.51	88.52	88.84

This method outperforms modern methods based on deep neural networks by an average of 10-15%.

Soil Analysis

The segmentation of agricultural fields into areas with emerged crops and other objects in the imagery allows for the identification of regions where the crops have not germinated for various reasons. To optimize resources and increase the yield of these areas, it is necessary to conduct soil analysis, which will help determine the reasons for the lack of germination. Such analysis may include the investigation of factors such as soil composition, nutrient levels, moisture, acidity, and the presence of pests or diseases, enabling more effective management of agricultural resources and informed decision-making to improve yield.

Agrochemical analysis of soil is carried out to determine the degree of its provision with basic elements of mineral nutrition, to establish its mechanical composition, hydrogen index and the degree of saturation with organic matter, i.e. those elements that determine the level of fertility.

Agrochemical analysis provides a comprehensive assessment of soil properties that determine plant growth and development: saturation with macronutrients and organic matter, media reaction (pH), availability of nutrients (nitrogen, potassium, phosphorus). The study shows to what extent the balance of elements meets the needs of plants, how the soil reacts to fertilizer application and what changes are necessary to increase yields.

Agrochemical analysis of soil is carried out on 8 main indicators:

moisture, organic matter, hydrolytic acidity, pH of salt extract (for agrochemical characterization of soil); nitrate nitrogen, ammonium nitrogen, mobile forms of phosphorus and potassium (to determine the content of macroelements).

Based on the analysis, a conclusion is given about the soil condition, recommendations on its use are given, doses of ameliorants, mineral and organic fertilizers for the planned harvest are calculated.

Procedure of soil sampling for agrochemical analysis. The soil sample is taken from the top layer (20 cm for planting vegetables, berries, potatoes, etc., 10 cm for lawns). To obtain an average result for the plot, soil samples should be taken in different parts of the plot using two methods:

1) envelope method - sampling in 4 corners (before reaching the corner about 1/4 of the diagonal) and in the center of the plot;

2) along the plot diagonal, in 4 - 5 points, through equal distances between sampling points.

For sampling, make a rectangular hole 20 cm deep (10 cm for lawns) with a shovel and carefully cut a layer of soil equal in width and thickness to a matchbox with a long knife from top to bottom. If the soil is light, this layer can be carefully poured into the bottom of the hole, where a clean sheet of paper should be placed beforehand. Pour the selected samples into one new polyethylene bag.

If the soil on the plot differs significantly in color or density (clay or sand), or the plot itself has a strong slope, samples should be taken from different areas, e.g. upper, middle and lower part of the slope, etc. Soil from one part of the slope (soil color) is collected in one bag, from another part in a second bag, etc.

The package shall be provided with an accompanying document indicating:

- place of sampling: district, village, (No. of garden), street, house, place on the plot (top of slope, clay, etc.);
- Date And Time Of Sampling.

Soil should be delivered within 1-2 days from the moment of sampling, until that time the sample should be stored in a refrigerator without freezing.

Soil fertility of agro-ecosystems in a multi-year plan also depends on climatic, and for specific years – on weather conditions, phytosanitary, ecological-toxicological and radiological conditions. Integral indicator of effective soil fertility is crop yields, productivity of fodder lands, quality of crop production in compliance with normative environmental requirements. Tables 1-2 summarize the main agrochemical and biochemical indicators of soil and the methodology of their determination.

Table 3: Agrochemical indicators and research methods

№	Soil indicators	Research method
1	Determination of soil temperature	Pyrometer DT-810 "CEM". Measured temperature range from -50 to +800°C with an error of 1.5°C
2	Determination of soil moisture	Moisture meter with "Thetaprobe" sensor, which determines the volumetric moisture content in the soil
3	Determination of the reaction of the medium (pH)	Potentiometric method
4	Determination of electrical conductivity	Conductometric method
5	Determination of nitrate	Ionometric method
6	Determination of organic matter content	According to I.V. Tyurin in modification of V.V. Nikitin
7	Determination of particle size distribution	Sieve method
8	Determination of hydrolytic acidity	The Kappen method modified by CSRIASA

Table 4: Biochemical parameters and methods of research

№	Soil indicators	Research method
1	Catalase activity	According to A.Sh. Galstyan (1956)
2	Dehydrogenase activity	According to A.Sh. Galstyan (1956)
3	Invertase activity	Colorimetric method with Felling's reagent
4	Phosphatase activity	By A.S. Galstyan, E.A. Harutyunyan (1966)
5	Urease activity	According to A.Sh. Galstyan (1965)

Table 5: General soil indicators and methods of their research.

INDICATORS	REACTIVES	EQUIPMENT	REGULATORY DOCUMENT
pH of water extract	Distilled water	Mortar, pestle, sieve 1-2 mm, spatula (spoon), laboratory scales, measuring cylinder, 150 ml containers (flasks), shaker, pH-meter	GOST 26423-85
pH of salt extract	KCl, distilled water	Mortar, pestle, sieve 1-2 mm, spatula (spoon), laboratory scales, measuring cylinder, 150 ml containers (flasks), shaker, pH-meter, magnetic stirrer.	GOST 26483-85
Organic matter	Distilled water, sulfur-chromium mixture (prepared by dissolving 23.2 g of K2Cr2O7 in 400 ml of water, then carefully add 2 l of concentrated H2SO4 with a density of 1.84 g/cm^3).	Paper, pestle, tweezers, 1 mm sieve, scales, 50 ml heat-resistant glass test tubes, 25 ml cylinder, 30 cm glass rod, water bath, test tubes, 30 cm long glass sticks, rubber pear, glass tube, 100 ml conical flasks, funnels, spectrophotometer (1, 2, 4 cm cuvettes).	GOST 26213-91 According to the Tyurin method modified by CSRIASA

Active carbon	KMnO4, distilled water, CaCl2, HCl or KOH	Sieve 1-2 mm, 50 ml flasks, 20 ml tubes spectrophotometer, centrifuge, cuvettes 1, 2, 4 cm	Modified Blair method
Catalase activity	Distilled water, hydrogen peroxide	Rubber hose, burettes, twin flasks, rubber stoppers, pipettes, pipettes, pipettes, tripod	Method of A.Sh. Galstyan
Dehydrogenase activity	2,3,5-triphenyltetrazolium chloride - TTX, triphenylformazan - THF, glucose, toluene or iodonitrotetrazolium chloride (INT), N,N-dimethylformamide, distilled water, mixture of N,N-dimethylformamide and ethanolDistilled water, hydrogen peroxide	20 ml test tubes, thermostat, desiccator, ethyl alcohol or acetone, spectrophotometer, centrifuge, desiccator	Method of A.Sh. Galstyan
Invertase activity	Distilled water, sucrose, toluene, SegNet salt (potassium sodium tartaric acid), KOH or NaOH, CuSO4, glucose	50 ml flasks, thermostat, 20 ml test tubes, water bath, spectrophotometer, centrifuge, filters, funnels, cortical stoppers	Modified colorimetric method of F.H. Haziev
Phosphatase activity	Distilled water, toluene, sodium phenolphthalein phosphate, alum, NH4OH, phenolphthalein, ethanol or sodium p-nitrophenyl phosphate, CaCl2, NaOH, p-nitrophenyl sodium phosphate, tris (hydro-xymethyl) aminomethane (HOCH2)3CNH2, maleic acid, boric acid, citric acid, HCl, KOH.	50 ml flasks, cork stoppers, dense filter, funnels, FEC, cylinders, pipettes and pipettes	The modified method of A.Sh. Galstyan and E.A. Harutyunyan or the method of M. Tabatabai and J. Bremner. Tabatabai and J. Bremner method
"Breathing" soil	NaOH or KOH, HCl or H2SO4, phenolphthalein	Sieve 1-2 mm, gauze bags, 250 ml wide-mouth flasks, rubber stoppers, thermostat, burettes for titration	Galstyan method
Humidity	KCl	Beakers, tweezers, mortar, desiccator, desiccator, corks	GOST 28268-89
SOIL CHEMISTRY			
Carbonate and bicarbonate ions in aqueous extracts	Distilled water, sulfuric acid	Mortar, pestle, 1-2 mm sieve, spatula/spoon, electronic scales, 150 ml process vessels (cassettes), measuring cylinder, shaker, filters, pipette/pipette, 100 ml conical flasks, filters, chemical beaker, phenolphthalein, burettes for titration.	GOST 26424-85
Chloride ions		Mortar, pestle, sieve 1-2 mm, spatula/spoon, electronic scales, 150 ml process containers (cassettes), measuring cylinder, shaker, filters, dispenser/pipette, 100 ml conical flasks, filters, chemical beaker, phenolphthalein, burettes for titration	GOST 26425-85

* GOST – Russian National Standard

When improving the methodology of integrated monitoring of soil fertility of agricultural soils, along with the reflection of traditional provisions, it is necessary to take into account the need to: expand the set of controlled agrochemical and biochemical indicators of soil fertility (described in detail in Tables 3.4) for a more complete assessment and increase the efficiency of fertilizer use and other elements of farming systems; develop rational (optimal) levels of fertility of the main types, subtypes and varieties of soils on an expanded list of indicators for the leading agricultural crops.

Scientific approaches to the timing and technique of soil sampling, rational levels of indicators of properties of different types and varieties of soils taking into account the requirements of cultivated crops and types of crop rotations, integrated assessment of soil fertility, etc. need further improvement.

Adjustment of technologies of integrated crop protection against weeds, pests and diseases, timing and doses of fertilizers during top dressing, mechanical tillage, as well as making decisions on irrigation or drainage system regulation based on the results of operational monitoring will increase crop yields and improve its quality.

CONCLUSION

ViT has revolutionized the field of computer vision by offering a powerful alternative to traditional models based on convolutional neural networks. With their ability to identify global relationships in images and process data efficiently, ViT has opened up new possibilities for image classification and analysis. The application of ViT in agriculture, particularly through datasets such as Agriculture-Vision, demonstrates the potential of this technology to automate and improve crop monitoring, paving the way for precision agriculture.

VITs demonstrate high accuracy in segmenting crop areas into sown and bare areas. This method demonstrates the accuracy of segmentation of areas with unsown crops on the Agriculture-Vision dataset was 98.01%, which outperforms state-of-the-art methods based on deep neural networks by an average of 10-15%.

The automation of crop monitoring using deep learning and computer vision technologies offers several advantages over manual monitoring. Automated systems are capable of processing and analyzing large volumes of data, minimizing the likelihood of human errors and providing more reliable results while significantly reducing time expenditures. Modern technologies can collect data in real time, allowing for updated information on the condition of plants and soil, and enabling quick responses to emerging issues.

Once areas with non-emergent crops are identified, agronomists can promptly conduct soil diagnostics to determine the causes of the problems. This allows them to focus their efforts on restoring soil health and enhancing its viability. Timely identification and resolution of issues in the fields contribute to increased resilience of agricultural practices to changing climatic conditions and plant diseases.

The architecture and computer vision technologies presented in the study have significant potential for increasing agricultural productivity. In addition to monitoring and segmenting cultivated areas, these technologies can help address a multitude of tasks. ViT can be used to assess the state of ecosystems, including studying biodiversity and monitoring changes in vegetation cover. They can identify specific pest species on plants and detect signs of disease on leaves and stems, helping agronomists respond quickly and take actions to save the crops.

In combination with other sensors and technologies, ViTs can be employed to analyze soil quality and its properties, which is crucial for developing effective fertilization and crop rotation methods. ViT can also analyze various parameters such as crop status, climatic conditions, and soil data to predict future potential yields.

These technologies can be integrated into autonomous agricultural machines to ensure the efficient execution of agricultural operations. Furthermore, ViT can be utilized for automated classification and sorting of products such as fruits and vegetables by size, color, and quality, facilitating the packaging and sales processes.

These examples illustrate how the technologies enable comprehensive crop protection from weeds, pests, and diseases, allowing for precise adjustments in the timing and dosage of fertilizer application during top dressing. Additionally, these systems facilitate informed decision-making regarding the adjustment of irrigation or drainage systems based on operational monitoring results. By implementing these advanced practices, crop yields can potentially increase, while simultaneously reducing production costs and enhancing the overall quality of the crops.

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "Conceptualization, R.D. and G.N.; methodology, G.N and Zh.M.; software, O.A.; validation, T.S. and O.M.; formal analysis, T.S.; investigation, Sh.V.; resources, R.D.; data curation, R.D.; writing—original draft preparation, O.A.; writing—review and editing, O.M.; visualization, A.A., O.M. and T.S. All authors have read and agreed to the published version of the manuscript.

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

Funding: the study was supported by a grant in the form of a subsidy from the federal budget to educational organizations of higher education for the implementation of activities aimed at supporting student scientific communities.

REFERENCES

Yongliang Qiao, Matthew Truman, Salah Sukkarieh "Cattle segmentation and contour extraction based on Mask R-CNN for pre-cision livestock farming" In Computers and Electronics in Agriculture, 165, 104958 (2019)

Xu Ma, Xiangwu Deng, Long Qi, Yu Jiang, Hongwei Li, Yuwei Wang, Xupo Xing "Fully convolutional network for rice seedling and weed image segmentation at the seedling stage in paddy fields" In Plos One (2019)

Daoyong Wang, Yuanyuan Fu, Guijun Yang, Xiaodong Yang, Dong Liang, Chengquan Zhou "Combined Use of FCN and Harris Corner Detection for Counting Wheat Ears in Field Conditions" In Institute of Electrical and Electronics Engineers 7 (2019)

Aghamohammadesmaeilketabforoosh, K.; Nikan, S.; Antonini, G.; Pearce, J.M. Optimizing Strawberry Disease and Quality De-tection with Vision Transformers and Attention-Based Convolutional Neural Networks. Foods 2024, 13, 1869. https://doi.org/10.3390/foods13121869

Choi W-J, Jang S-H, Moon T, Seo K-S, Choi D-S, Oh M-M. Continuous Growth Monitoring and Prediction with 1D Convolutional Neural Network Using Generated Data with Vision Transformer. Plants. 2024; 13(21):3110. https://doi.org/10.3390/plants13213110

Salman Khan, Muzammal Naseer, Munawar Hayat, Syed Waqas Zamir, Fahad Shahbaz Khan, Mubarak Shah "Transformers in Vision: A Survey" In Acm Journals 54, 105, 1-41 (2022) https://dl.acm.org/doi/abs/10.1145/3505244

Chiu, M. T., Xu, X., Wei, Y., Huang, Z., Schwing, A. G., Brunner, R., ... & Shi, H. (2020). Agriculture-vision: A large aerial image database for agricultural pattern analysis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recogni-tion (pp. 2828-2838).

Maggiori, E., Tarabalka, Y., Charpiat, G., & Alliez, P. (2017, July). Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark. In 2017 IEEE International geoscience and remote sensing symposium (IGARSS) (pp. 3226-3229). IEEE.

Xia, G. S., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., Zhang, L. (2018). DOTA: A large-scale dataset for object detection in aerial images. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3974-3983).

Waqas Zamir, S., Arora, A., Gupta, A., Khan, S., Sun, G., Shahbaz Khan, F., Bai, X. (2019): A large-scale dataset for instance seg-mentation in aerial images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Work-shops (pp. 28-37).

Xia, G. S., Hu, J., Hu, F., Shi, B., Bai, X., Zhong, Y., Lu, X. (2017). AID: A benchmark data set for performance evaluation of aerial scene classification. IEEE Transactions on Geoscience and Remote Sensing, 55(7), 3965-3981.

Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Raskar, R. (2018). Deepglobe 2018: A challenge to parse the earth through satellite images. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 172-181).

Helber, P., Bischke, B., Dengel, A., & Borth, D. (2019). Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(7), 2217-2226.

Basu, S., Ganguly, S., Mukhopadhyay, S., DiBiano, R., Karki, M., Nemani, R. (2015, November). Deepsat: a learning framework for satellite imagery. In Proceedings of the 23rd SIGSPATIAL international conference on advances in geographic information systems (pp. 1-10).

Haug, S., & Ostermann, J. (2015). A crop/weed field image dataset for the evaluation of computer vision based precision agricul-ture tasks. In Computer Vision-ECCV 2014 Workshops: Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part IV 13 (pp. 105-116). Springer International Publishing.

Kallimani, C., Heidarian, R., van Evert, F. K., Rijk, B., & Kooistra, L. (2020). UAV-based Multispectral & Thermal dataset for ex-ploring the diurnal variability, radiometric & geometric accuracy for precision agriculture. Open Data Journal for Agricultural Research, 6, 1-7.

Olsen, A., Konovalov, D. A., Philippa, B., Ridd, P., Wood, J. C., Johns, J., White, R. D. (2019). DeepWeeds: A multiclass weed spe-cies image dataset for deep learning. Scientific reports, 9(1), 2058.

Chiu, M. T., Xu, X., Wei, Y., Huang, Z., Schwing, A. G., Brunner, R., Shi, H. (2020). Agriculture-vision: A large aerial image data-base for agricultural pattern analysis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2828-2838).

Thakkar, M., & Vanzara, R. (2024). Enhancing crop yield estimation from remote sensing data: a comparative study of the Quar-tile Clean Image method and vision transformer. Discover Applied Sciences, 6(11), 610.

Ren, M., Zeng, W., Yang, B., & Urtasun, R. (2018, July). Learning to reweight examples for robust deep learning. In International conference on machine learning (pp. 4334-4343). PMLR.

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). A review of the use of convolutional neural networks in agriculture. The Journal of Agricultural Science, 156(3), 312-322.

Boonpook, W., Tan, Y., Nardkulpat, A., Torsri, K., Torteeka, P., Kamsing, P., Jainaen, M. (2023). Deep learning semantic segmen-tation for land use and land cover types using Landsat 8 imagery. ISPRS International Journal of Geo-Information, 12(1), 14.

Li, K., Zhang, L., Li, B., Li, S., Ma, J. (2022). Attention-optimized DeepLab V3+ for automatic estimation of cucumber disease se-verity. Plant Methods, 18(1), 109.

Chen, S., Song, Y., Su, J., Fang, Y., Shen, L., Mi, Z., Su, B. (2021). Segmentation of field grape bunches via an improved pyramid scene parsing network. International journal of agricultural and biological engineering, 14(6), 185-194.

Huang, X., Peng, D., Qi, H., Zhou, L., & Zhang, C. (2024). Detection and Instance Segmentation of Grape Clusters in Orchard Envi-ronments Using an Improved Mask R-CNN Model. Agriculture, 14(6), 918.

Mujkic, E., Philipsen, M. P., Moeslund, T. B., Christiansen, M. P., Ravn, O. (2022). Anomaly detection for agricultural vehicles us-ing autoencoders. Sensors, 22(10), 3608.

Vemulapalli, R., Tuzel, O., Liu, M. Y., Chellapa, R. (2016). Gaussian conditional random field network for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3224-3233).

Hatamizadeh, A., Sengupta, D., Terzopoulos, D. (2019). End-to-end deep convolutional active contours for image segmentation. ar