This study's object is the process of determining the fractional composition of winter wheat grain mixtures using computer vision and deep learning methods. The basic task that needs to be solved is the high complexity, subjectivity, and speed of determining the fractional composition of grain by conventional methods.

The results obtained demonstrate the successful training and comparative analysis of several YOLO11-seg instance segmentation models on a specialized dataset deployed on the NVIDIA Jetson Orin platform. In particular, the YOLO11m-seg model with an image size of 640 × 640 pixels achieved the optimal compromise between accuracy and speed, achieving a Mask mAP50-95 index of 0.558 at an output speed of 62.5 ms/image. Training the YOLO11n-seg 1280 × 1280 model provided the best average segmentation accuracy (Mask mAP50-95 0.640) by increasing the performance of identifying objects of "complex" classes, which are key for accurate determination of the fractional composition.

The results have made it possible to solve the problem under consideration through the empirically justified choice of architecture. Unlike hypothetical approaches, the study provides specific quantitative data on the performance of different YOLO11-seg architectures. That allowed for a reasonable selection of the model that best meets the requirements for accuracy and speed for practical deployment, solving the problem of uncertainty in the choice of architecture.

The findings create the basis for automating grain quality control, increasing its efficiency, objectivity, and also significantly reducing labor intensity by automating routine operations. For practical use of the system, it is necessary to ensure stable lighting conditions, as well as the presence of a digital camera, a computer, and appropriate software

Keywords: machine vision, image segmentation, YOLO, grain mixtures, winter wheat, fractional composition

UDC 633.1:681.78

DOI: 10.15587/1729-4061.2025.338124

SUBSTANTIATING THE YOLO11 ARCHITECTURE FOR DETERMINING THE FRACTIONAL COMPOSITION OF WINTER WHEAT GRAIN MIXTURES

Serhii Stepanenko
Corresponding author
Doctor of Technical Sciences, Senior Researcher*
E-mail: stepanenko_s@ukr.net

Alvian Kuzmych
PhD, Senior Researcher*

Andrii Borys
PhD, Senior Researcher
Department of Agricultural Navigation and Automation of Mobile Processes**

Viktor Dnes
PhD, Senior Researcher
Department of Simulation of Technological Processes in Crop Production**

Serhii Kharchenko
Doctor of Technical Sciences, Professor
Department of Fundamentals of Production Engineering
Lublin University of Technology
Nadbystrzycka str., 36, Lublin, Poland, 20-618

Ivan Rogovskii
Doctor of Technical Sciences, Professor***

Gennadii Golub
Doctor of Technical Sciences, Professor***

Mykola Berezovyi
PhD, Associate Professor
Department of Mechanics****

Andrii Lutsiuk
PhD Student
Department of Electronics and Information Technology
Institute of Information and Communication Technologies and Electronic Engineering
Lviv Polytechnic National University
S. Bandery str., 12, Lviv, Ukraine, 79013

*Department of Mechanical and Technological Problems of Harvesting and Post-Harvest Processing of Grain and Oilseed Crops**

**Institute of Mechanics and Automatics of Agroindustrial Production of the National Academy of Agrarian Sciences of Ukraine
Vokzalna str., 11/1, Hlevakha vil., Ukraine, 08631

Department of Technical Service and Engineering Management named after M. P. Momotenko*

****National University of Life and Environmental Sciences of Ukraine
Heroiv Oborony str., 15, Kyiv, Ukraine, 03041

Received 03.06.2025
Received in revised form 02.07.2025
Accepted 21.07.2025
Published 31.07.2025

How to Cite: *Stepanenko, S., Kuzmych, A., Borys, A., Dnes, V., Kharchenko, S., Rogovskii, I., Golub, G., Berezovyi, M., Lutsiuk, A. (2025). Substantiating the YOLO11 architecture for determining the fractional composition of winter wheat grain mixtures. Eastern-European Journal of Enterprise Technologies, 4 (2 (136)), 81–92. https://doi.org/10.15587/1729-4061.2025.338124*

1. Introduction

Winter wheat is one of the most important grain crops, which forms the basis of the diet of a significant part of humanity and is an important raw material for many branches in the food industry. The quality of wheat directly affects the quality of final products, such as bread, pasta, confectionery, and animal feed [1]. One of the most important indicators of the quality of wheat grain is its fractional composition, i.e., the ratio of the mass of the components of the grain mixture.

Analysis of the fractional composition of wheat grain is carried out at all stages of agricultural production and processing [2]. During harvesting and post-harvest processing of the crop, constant quality control over technological operations is executed. The presence of a large number of small, thin, or damaged grains indicates low crop quality [3]. Control over the fractional composition of grain implies determining its technological properties for the production of flour and cereals, the formation of commodity batches of the crop. Control over the fractional composition is important for optimizing post-harvest processing and storage processes [4]. The efficiency of grain cleaning and sorting machines directly depends on knowledge of the fractional composition of raw materials. This allows one to properly adjust the equipment and minimize the loss of wholesome grain. Fractional composition analysis is used in breeding to design new wheat varieties with improved grain quality indicators, in particular, high uniformity and optimal size. Fractional analysis makes it possible to detect the presence of weed seeds, damaged grains, mineral and organic impurities, which is critically important for food safety.

The speed of fractional composition analysis is important in modern agricultural production: when receiving grain at elevators or grain receiving points, during control in post-harvest processing processes [5]. The ability to quickly assess changes in the fractional composition of materials during technological separation operations allows for timely adjustment of the parameters of the corresponding equipment.

The accuracy of determining the fractional composition is important for compliance with quality standards, increasing production efficiency, and product safety for consumers.

Conventional methods of manual analysis of grain materials, although standard, have significant drawbacks: they are subjective, laborious, and slow [6]. This can lead to delays in assessing grain quality and making important technological decisions. Modern technologies, such as computer vision and methods based on image analysis, could significantly speed up the process of determining the fractional composition and increase its accuracy.

Deep learning methods have become the object of research in many industries, including the agricultural sector. Methods based on deep learning are able to directly learn features from training data [7]. These methods are used for crop monitoring, disease detection, yield prediction, and plant species recognition [8].

The use of computer vision and deep learning technologies for analyzing the fractional composition of wheat opens up a number of significant advantages. These are automation of the analysis process, minimizing the influence of the human factor, ensuring a high degree of objectivity and reproducibility of results, speed, and the ability to receive analysis results in real time. Therefore, the implementation of such methods in the practice of agricultural enterprises is an urgent task.

2. Literature review and problem statement

Studies demonstrate the effectiveness of convolutional neural networks (CNNs) for classifying sunflower diseases [9] and hybrid models for extracting features from small datasets [10].

For effective selective weed control in agricultural fields where lighting conditions and background frequently change, the authors of [11] used a new deep learning graph architecture – Graph Weeds Net (GWN). According to their results, this system is able to recognize different types of weeds in ordinary RGB images, achieving a high categorization accuracy of 98.1%. This system was designed for the recognition and classification of objects (weeds) represented in the form of graph structures. To analyze the fractional composition of grain mixtures, it is necessary to solve the problem of quantitative determination of particle distribution. The use of graph architecture of a neural network could complicate the detection and segmentation of individual particles, especially with a high density of objects that often overlap in the image, typical of grain mixtures.

The authors of study [12] use a neural network based on the MaskRCNN model for an automated system for scanning and recognizing plants, which provides a detection speed of 0.2 s. with an accuracy rate of 89%. The MaskRCNN model employed is characterized by high accuracy of object segmentation. However, the two-stage approach to image processing reduces the performance of Mask R-CNN, which is especially critical for real-time tasks. It also requires more significant computing resources, in particular the use of powerful graphics processors (GPUs).

Among neural network models, the YOLO series is popular due to its speed and efficiency in object detection tasks, which is important for deployment at the periphery [13]. They require prepared image datasets for training. Comparative studies of 25 latest YOLO detectors of seven versions for weed detection [14] allowed the authors to report the significant potential of using YOLO family models for real-time relationships.

In [15], the results of research on a system that analyzes grain quality and the presence of impurities in grain material are described. It is shown that neural networks (NNs) are able to perform the tasks of detecting impurities in grain material, assessing the level of grain quality (good, poor, and average quality). Categorization is performed by color, shape, and size. However, the issue of quantitative assessment of impurities remains unresolved. The reason for this may be the limited computational characteristics of early neural networks. A way to overcome these difficulties may be the implementation of convolutional neural networks (CNNs), which automatically learn features.

Study [16] reports the results of a study on a system for sorting durum wheat grains. The system is based on an artificial neural network (ANN), one of the machine learning methods. The authors propose a method that, using a combined analysis of the shape, color, and complex texture of the grain surface, is able to automatically identify glassy grains. This creates a basis for the design of industrial sorting systems that could operate in real time and significantly increase the efficiency of quality control at elevators and processing plants. However, unresolved issues related to the presence of a significant number of fractions of grain materials remained. The reason for this may be the limited computational capabilities of grain sorting systems. An option to overcome these difficulties may be the use of modern deep learning models. This is the approach used in [17].

In [17], the results of research on the construction and improvement of the YOLO-wheat model for automatic determination of wheat grain quality are given. The authors of the study focused on the detection of whole damaged and moldy wheat grains. As a base, the YOLOv8n model was chosen for automatic detection of grain appearance. Experimental results showed that YOLO-wheat achieved a mAP value of 91.3% in determining the exact position and recognizing the appearance of wheat grains. This makes it possible to categorize and determine the number of objects of the corresponding classes. However, using the model built does not make it possible to estimate the fractional composition of the grain mixture. The reason for this is the use of YOLO detection models, which provide bounding boxes around each detected object and assign a class (label) to this object. The model does not provide information about the exact shape or boundaries of the object within the bounding box, only its approximate location. An option to overcome these difficulties may be the use of segmentation models, which provide much more detailed information than detection, since they work at the pixel level.

In study [18], a method for detecting wheat impurities by combining terahertz spectral imaging and a convolutional neural network (CNN) was proposed. The results of the study show that this method can effectively recognize various impurities in wheat with an accuracy of 97.83%. However, training and using complex CNNs can require significant computational resources, which may be a problem for some applications. In addition, the significant cost of terahertz equipment may limit its widespread use. An option to overcome these difficulties may be the use of model optimization, as well as the transfer of more complex calculations to cloud systems.

For real-time wheat seed detection, the authors of study [19] used the YOLOv8-HD model, which outperforms the main models in terms of accuracy, speed, and model size. However, when processing images in which objects overlap, the accuracy of wheat grain detection was only 77.6%, indicating the need for further improvement. The proposed model provided detection and counting of wheat grains, straw, and husk in different scenarios. However, to determine the fractional composition of grain mixtures, counting the number of individual objects is not enough since they can have different dimensional characteristics and mass. In the public domain, there are resources such as GrainSpace, which contain millions of images of wheat, corn and rice, designed for fine recognition [20]. This dataset includes images of individual grains cut out using the YOLOv5 model. Other datasets focus on wheat heads (e.g., the Global Wheat Head Detection dataset [21]) or on specific grain states (ideal, sprouted, diseased, and damaged grains [20]).

Some sources emphasize the difficulty of building accurate deep learning models due to the difficulty of constructing large datasets under uncontrolled conditions with uneven illumination and grain overlap [22]. Other studies emphasize the importance of building high-quality datasets with high resolution and detailed image annotations [13]. Controlled and clean datasets (e.g., individual grains on a black background [22]) allow for high accuracy in specific tasks. However, they often lack the diversity needed to generalize real-world conditions with different environments and observation conditions. Achieving reliable fractional composition analysis across a variety of agricultural settings requires either significantly larger or more diverse datasets.

Our review of the literature reveals numerous challenges that remain in the field of grain quality control. Most machine learning applications often focus on categorization, such as distinguishing healthy grains from diseased or contaminated grains. However, a more detailed approach is needed to determine the fractional composition, which involves not only identifying grain types but also quantifying their proportions in the mixture. The need to obtain accurate values of the areas of the corresponding classes of objects in images [23] indicates that simple object detection using bounding boxes is insufficient for determining the fractional composition of grain mixtures and goes beyond simple categorization or detection. Such an approach is crucial for accurately determining the proportion of each component in the mixture and requires models capable of understanding objects at the pixel level, such as Mask R-CNN or improved YOLO models for segmentation. In addition, the use of high-precision but expensive technologies, such as terahertz equipment [18], remains economically impractical for widespread implementation, which emphasizes the need to devise more accessible and scalable solutions based on conventional optical images.

The above allows us to argue that it is advisable to conduct a study aimed at substantiating the architecture of neural network models for determining the fractional composition of winter wheat grain mixtures.

3. The aim and objectives of the study

The aim of our study is to determine the parameters of a neural network for assessing the fractional composition of winter wheat grain samples by analyzing their digital images. This will enable fast and effective grain quality control and will be a prerequisite for designing systems that automate separation processes.

To achieve this aim, the following objectives were accomplished:

– to develop an algorithm for determining the fractional composition of winter wheat grain mixtures by analyzing their digital images;

– to collect a data set and train models from the YOLO11-seg family for categorizing and segmenting objects in images of grain materials;

– to test the trained models on images of winter wheat grain samples.

4. The study materials and methods

The object of our study is the process of determining the fractional composition of winter wheat grain mixtures using computer vision and deep learning methods. The principal hypothesis of the study assumes that modern neural network architectures (in particular, YOLO-Segmentation) could allow for accurate and fast determination of the fractional composition of winter wheat grain mixtures. This will be sufficient for practical application in real time.

A number of assumptions were adopted to conduct the study that simplify the conditions and allow us to focus on the main goal. It is assumed that the built dataset, although limited, is sufficiently representative for training models for recognizing the main wheat fractions, and the annotation is performed with reasonable accuracy. It is assumed that the YOLO11-seg architectures are one of the most effective and suitable for this task of segmenting instances under the conditions of boundary calculations at the time of our study.

The following simplifications were used to conduct the study. Images for the dataset were acquired under laboratory, controlled conditions. This simplification allowed us to minimize the influence of external factors and focus on training the model to recognize the fractions themselves.

Only the main fractions characteristic of wheat analysis were considered. Other possible impurities that may occur in practice were not included in the recognition object.

To develop an algorithm for determining the fractional composition of grain mixtures, real samples of winter wheat samples taken during the 2024 harvesting season from the combine harvester hopper were photographed. The contamination of the test samples was 1.2–1.5%, the content of crushed grain was 1.4–1.8%; the content of weak grain was within 2.0–2.5%.

For photographing, a sample of material weighing 40–50 g was distributed in a single layer on a sheet of white office paper in A4 format inside a bounding box measuring 200 × 200 mm. Photographing was carried out using cameras of different smartphones under natural lighting under an auto focus and an auto flash mode. The next step was to crop the edges of the image and save it in a 1:1 ratio. The width and height of the grain contour in pixels were measured. The intermediate image was compressed to sizes 640 × 640; 800 × 800 and 1280 × 1280 pixels. For each compressed image variant, the width and height of the grain contour were measured.

To form the dataset, winter wheat grain material from the 2024 harvest was used. Sampling was carried out according to ISO 24333:2009 "Cereals and cereal products – Sampling" using a grain sampler. The fractional composition of wheat grain materials was studied according to the ISO 7970:2021 "Wheat (*Triticum aestivum* L.) – Specification" method.

During the analysis, samples of winter wheat grain material were divided into the following five fractions: whole grain, crushed grain, thin grain, straw impurities, inflorescence particles (field). Thin grain was separated from the sample by sieving on a sieve measuring 2.0 × 20.0 mm using a laboratory sieve RLU-3. Other fractions were separated manually.

A bench was made for photographing samples of winter wheat grain materials (Fig. 1). The bench consisted of an adjustable stand, a platform for installing a smartphone, and a panel with two contours of ring LED lighting.

For photography, individual samples of grain material components were placed in a single layer on a sheet of white A4 paper.

The photographs were taken using digital smartphone cameras with a resolution of 10 to 50 megapixels under an autofocus mode. The distance from the camera lens to the material sample varied within 250–350 mm, and the illumination level was within 300–6000 lux.

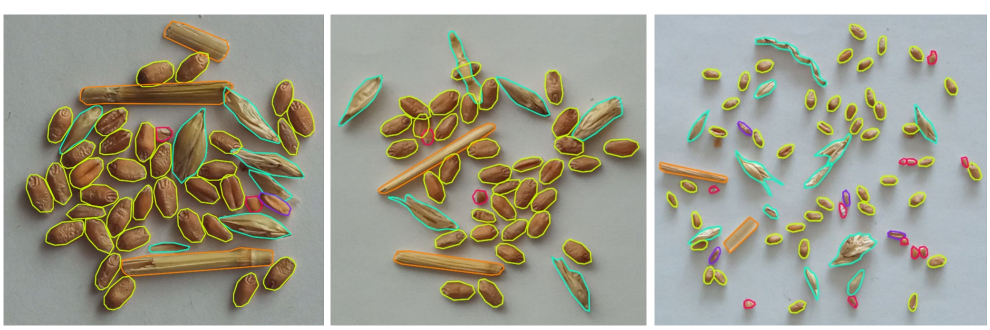
To form the training data set, 280 photographs of winter wheat grain material components were taken. The resulting images were reduced to a size of 1280 × 1280 pixels by cutting off the edges.

Fig. 1. Laboratory bench for photographing grain samples

Image annotation was performed using the Roboflow Annotate interface (Fig. 2). The Smart Polygon tool was used to annotate the images. This is an AI-supported function that significantly speeds up the creation of image annotations. After automatic mask generation, each segmented object was manually inspected and corrected to ensure high annotation accuracy. Special attention was paid to annotating partially overlapping objects using overlapping segmentation masks. This ensures accurate representation of the data at the pixel level. No preprocessing or data augmentation was performed directly in the Roboflow interface.

The dataset was divided into training and validation sets, consisting of 234 and 46 images, respectively (Table 1). The annotated data were exported as TXT annotation files and YAML configuration files for use with YOLO11.

The platform for training the models was the NVIDIA Jetson AGX Orin 64GB developer kit (China), which provides up to 275 TOPS of AI performance with a 2048-core NVIDIA Ampere architecture GPU and a 12-core Arm Cortex-A78AE processor [24] (Fig. 3).



a *b* *c*

Fig. 2. Annotating images using the Roboflow Annotate interface: *a* – with dense placement of objects; *b* – with partial overlap of objects; *c* – with a variable distance from the camera lens to the material sample

Table 1

Number of annotations for each dataset feature class

Class name	Number of class objects		
	Entire dataset	Training set	Validation set
Ground grain (BG)	1407	1168	239
Chaff (Ch)	2166	1827	339
Whole grain (NG)	10801	9003	1798
Shredded straw (St)	1163	986	177
Thin grain (TG)	1579	1329	250
Together (all)	17116	14313	2803

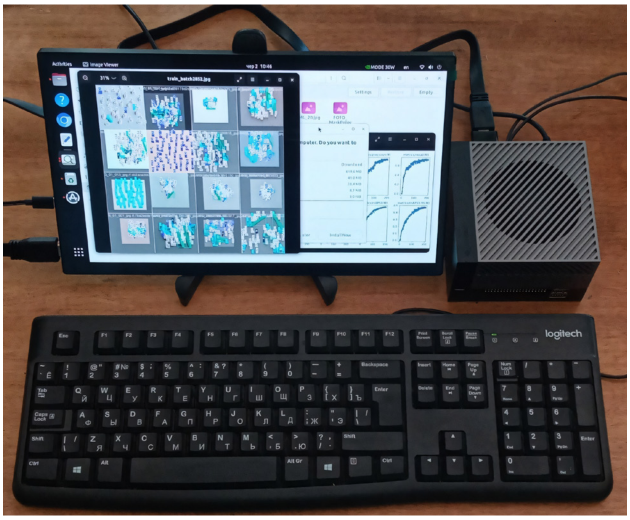


Fig. 3. Training models on the NVIDIA Jetson AGX Orin 64GB development kit

The platform is equipped with two dedicated deep learning accelerators (DLA v2.0) and a hardware computer vision and image processing accelerator (PVA v2.0). The JetPack software platform version 6.2 includes the Linux kernel 5.15, CUDA 12.6 for GPUs, the cuDNN 9.3 deep neural network library, and the TensorRT10.3 deep learning environment for optimizing and deploying AI models.

To study the fractional composition of winter wheat grain mixtures, deep learning models from the YOLO11-Segmentation family were selected. These models represent a compromise between speed, accuracy, and model size, which makes them suitable for deployment on peripheral devices. The characteristics of the models and the results of their tests on the COCO val2017 dataset and the NVIDIA T4 architecture are given in Table 2.

Testing these models will allow us to analyze their performance and accuracy characteristics on a real dataset, which will make it possible to choose the most suitable model for specific deployment requirements.

Standard hyperparameters optimized for this task were used to train the YOLO11-Segmentation model. In particular, the learning rate was set to 0.01 with subsequent reduction using Cosine Annealing, which provided fast convergence at the beginning and fine-tuning at the final stages. The batch size of 16 was chosen as a compromise between performance and available GPU memory. Training lasted up to 200 epochs using the Early Stopping technique (patience = 10) to avoid overfitting on the training dataset. The AdamW optimizer was chosen due to its high efficiency and stable convergence.

The input image size for the YOLO11 models was set at three levels: 640 × 640; 800 × 800; and 1280 × 1280 pixels. Data augmentation techniques were not used in this study.

When evaluating the performance of deep learning models for object detection and segmentation, a set of standard metrics was used: detection precision (P); detection recall (R); average accuracy calculated at an IoU threshold of 0.5 (mAP50); average accuracy at different IoU thresholds ranging from 0.5 to 0.95 with a step of 0.05 (mAP50–95). They help establish how well the model performs the task of identifying objects and accurately delineating their boundaries.

The practical effectiveness of the models was evaluated by output speed and GPU memory usage, which are critical metrics for assessing the real-world performance of the model, especially when deployed on limited resources.

The trained models were tested by detecting objects and segmenting them on images of real grain samples. The models were deployed on CPU (Intel Core(TM), i5-9400, 2.90GHz) and GPU (NVIDIA Jetson AGX Orin 64GB) to compare computational efficiency.

5. Results of research on the architecture of YOLO11 models for determining the fractional composition of winter wheat grain mixtures

5. 1. Development of an algorithm for determining the fractional composition of winter wheat grain mixtures

To develop an algorithm for determining the fractional composition of grain mixtures, photographs of real samples of winter wheat samples were taken. The results of processing digital photos obtained from cameras of various smartphones after edge clipping and compression are given in Table 3.

When compressing intermediate images to a size of 640 × 640, the smaller contour size (width) of the grain is only 11–13 pixels. This is the minimum value for an acceptable quality of object categorization. In this case, the segmentation will be quite rough, and the contour accuracy may be unsatisfactory.

To improve the quality of segmentation, it is proposed to use the method of dividing the intermediate image into a set of n^2 elements, and sequential analysis of each of the image elements.

The algorithm for implementing the proposed method is shown in Fig. 4.

Table 2

Characteristics of deep learning models from the YOLO11-Seg family [25]

Model	Size, pixel	mAP 50–95 (box)	mAP 50–95 (mask)	Speed T4^{TensorRT10}, (ms)	Number of parameters, million	GFLOPs
YOLO11n-seg	640	38.9	32.0	1.8 ± 0.0	2.9	10.4
YOLO11s-seg	640	46.6	37.8	2.9 ± 0.0	10.1	35.5
YOLO11m-seg	640	51.5	41.5	6.3 ± 0.1	22.4	123.3
YOLO11l-seg	640	53.4	42.9	7.8 ± 0.2	27.6	142.2

Table 3

Processing parameters for digital photos taken from cameras of various smartphones

No.	Camera brand	Original photo settings		Photo parameters after cropping the edges		Grain outline size in the image, pixel			
		size, pixel	file size, MB	size, pixel	file size, MB	original	after compression to size		
							640	800	1280
1	Nokia G22	4080 × 3060	3.48	2375 × 2375	2.01	73 × 50	20 × 13	25 × 17	39 × 27
2	OPPO A17	3072 × 4080	3.82	2560 × 2560	2.46	43 × 78	11 × 20	13 × 24	22 × 39
3	Xiaomi M2004/19C	3120 × 4160	2.9	2700 × 2700	2.02	47 × 90	11 × 21	14 × 27	22 × 43
4	OPPO A17K	2448 × 3264	2.67	2100 × 2100	1.64	71 × 39	22 × 12	27 × 15	43 × 24
5	Samsung A506FM	4032 × 3024	2.61	2700 × 2700	2.13	85 × 51	20 × 12	25 × 15	40 × 24
6	Xiaomi M2103 K19G	2992 × 2992	2.40	1940 × 1940	1.34	33 × 62	11 × 20	14 × 26	22 × 41
7	Samsung Galaxy A24	4080 × 1884	2.18	1500 × 1500	824 kb	47 × 30	20 × 13	25 × 16	40 × 25

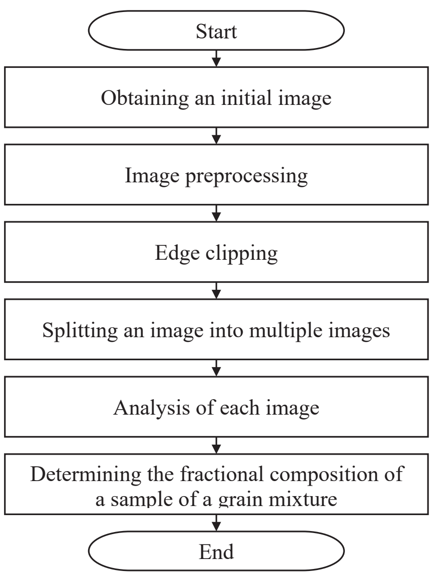


Fig. 4. Algorithm for determining the fractional composition of a sample of a winter wheat grain mixture

The algorithm for determining the fractional composition of a sample of a winter wheat grain mixture (Fig. 4) consists of the following stages:

Stage 1 – obtaining the initial image. At this stage, either direct photography of the grain mixture or work with a previously saved image is assumed. The requirement for the image is to place the grain mixture sample in the thinnest possible layer on a uniform white background.

Stage 2 – pre-processing of the image. At this stage, the area containing the grain mixture sample is automatically determined and reduced to a standard size for further analysis. For this purpose, a sequence of computer vision methods based on the contrast between the grain and the surface on which it is placed is used.

To select the area containing the grain, image binarization is used. By setting a brightness threshold, the image is converted to black and white, where the grains and the background become clearly delimited. Morphological operations erosion and dilation are applied to the binary image. This helps remove small noise and gaps in the object mask, making its contour more coherent. The FindContours contour search algorithm from the OpenCV library is used to find all connected components in the binary image. Among all the found contours, the largest in area is selected. It is assumed that it corresponds to the area with the grain sample. A bounding rectangle is built around this contour, which determines the coordinates of the upper left corner, as well as its width and height.

If the resulting area differs from a square, it is reduced to a square while preserving the center of the area, and the side of the square is chosen equal to the larger side of the rectangle.

Stage 3 – clipping the edges of the image according to the obtained square area. Based on the coordinates and dimensions of the bounding square determined in Stage 2, the original image is cropped. The image is cropped to a size of $L \times L$ pixels, which makes it possible to obtain an image that contains only the area with a grain sample, eliminating excess background and other extraneous objects that may affect further analysis by the neural network.

Stage 4 – dividing the image into a set of images. To bring the image size into line with the required input parameters of the neural network, the division factor is determined (rounded to an integer)

$$n = \frac{L}{k}, \tag{1}$$

where L is the size of the image side, pixel; k is the required image size for processing by the neural network, pixel.

The image is divided by width and height into n parts, that is, as a result we obtain a set of n^2 images. Each of the n^2 obtained images is scaled to the input size of the neural network.

Stage 5 – analysis of each image from the obtained set using a neural network. At this stage, the following takes place:

1) categorization and segmentation of grain mixture sample objects in each image;

2) determination of the total area of masks for each type of grain mixture sample objects in each image.

Stage 6 – determination of the fractional composition of the grain mixture sample. For this purpose, the values of the total area of masks for each type of grain mixture sample objects are summed up and, taking into account the weight coefficients, the fractional composition of the grain mixture sample is determined.

5. 2. Results of training YOLO11 models on the constructed dataset

The dependences of the loss function for bounding boxes, which is responsible for the accuracy of predicting the coordinates and sizes of the boxes, as well as the loss function for segmentation, which is responsible for the accuracy of predicting binary segmentation masks, are shown in Fig. 5.

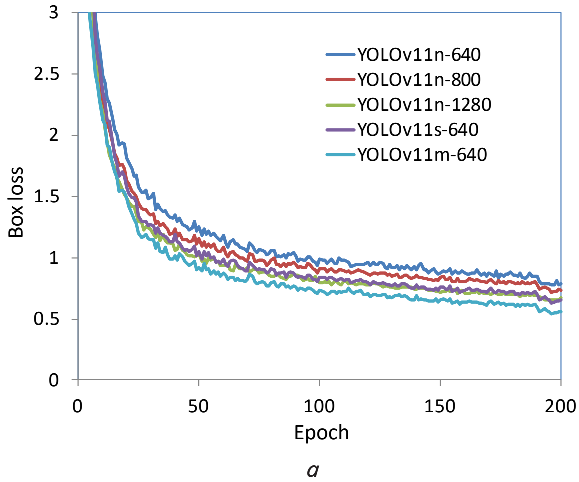


Fig. 5. Loss curves during model training: *a* – localization; *b* – segmentation

The dependences of changes in standard metrics for assessing the quality of model segmentation: detection precision (P); detection recall (R); average segmentation accuracy mAP50 and mAP50–95, are shown in Fig. 6.

Analysis of the dependences of loss functions reveals a similar nature of the learning process for all models. The decrease in the loss value over 200 training epochs indicates the possibility of continuing to train the models, which is a reserve for increasing their performance. Table 4 summarizes the key metrics, output time, and memory usage during training for each of the experimental models.

The YOLO11m-seg 640 model achieved the highest average segmentation accuracy mAP50 – 0.877. In contrast, the YOLO11n-seg 1280 model is characterized by the highest average accuracy at increased levels of complexity mAP50-95 – 0.640. The results of analysis of the performance of these models with a distribution by object classes are given in Tables 5, 6.

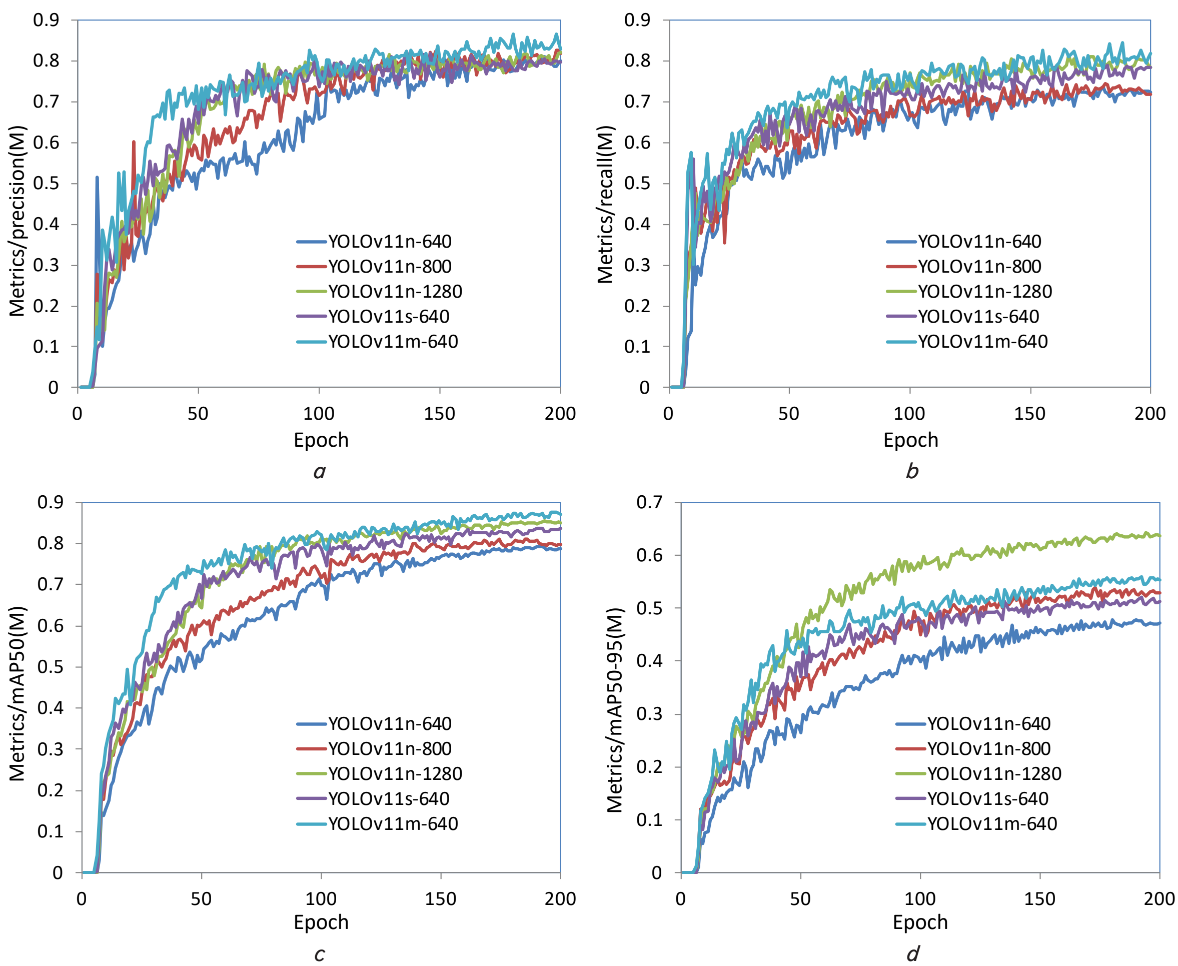


Fig. 6. Model training performance metrics: *a* – segmentation precision; *b* – segmentation recall; *c* – average accuracy mAP50; *d* – average accuracy at increased IoU thresholds mAP50-95

Table 4

Performance of YOLO models on authentic constructed dataset

Model/size	Box mAP50	Box mAP50-95	Mask mAP50	Mask mAP50-95	Output time (ms)	GPU memory (GB)
YOLO11n-seg 640	0.816	0.613	0.792	0.478	17.1	9.48
YOLO11n-seg 800	0.811	0.623	0.806	0.536	22.1	15.3
YOLO11n-seg 1280	0.864	0.695	0.850	0.640	61.4	29.1
YOLO11s-seg 640	0.852	0.668	0.838	0.519	29.0	12.1
YOLO11m-seg 640	0.889	0.727	0.877	0.558	62.5	14.3
YOLO11l-seg 640	0.877	0.723	0.861	0.549	74.9	16.5

Table 5

Analysis of performance indicators of the YOLO11m-seg 640 model by classes

Class	Box (P)	Box (R)	Box mAP50	Box mAP50–95	Mask (P)	Mask (R)	Mask mAP50	Mask mAP50-95
all	0.826	0.839	0.889	0.727	0.835	0.824	0.877	0.558
BG	0.786	0.741	0.824	0.654	0.782	0.707	0.780	0.355
Ch	0.853	0.870	0.919	0.731	0.876	0.854	0.910	0.677
NG	0.929	0.978	0.975	0.876	0.932	0.974	0.975	0.690
St	0.815	0.847	0.880	0.688	0.816	0.842	0.874	0.612
TG	0.749	0.760	0.846	0.688	0.768	0.742	0.846	0.455

Table 6

Analysis of performance indicators of the YOLO11n-seg 1280 model by classes

Class	Box (P)	Box (R)	Box mAP50	Box mAP50–95	Mask (P)	Mask (R)	Mask mAP50	Mask mAP50-95
all	0.813	0.812	0.864	0.695	0.806	0.796	0.850	0.640
BG	0.755	0.746	0.788	0.614	0.756	0.738	0.788	0.565
Ch	0.810	0.850	0.894	0.691	0.808	0.838	0.881	0.681
NG	0.924	0.971	0.973	0.860	0.926	0.970	0.973	0.821
St	0.807	0.808	0.844	0.635	0.768	0.763	0.790	0.516
TG	0.770	0.684	0.819	0.672	0.774	0.672	0.819	0.619

Analysis of the results reveals the presence of two classes of objects with the highest evaluation metrics: full grain (NG) and chaff (Ch).

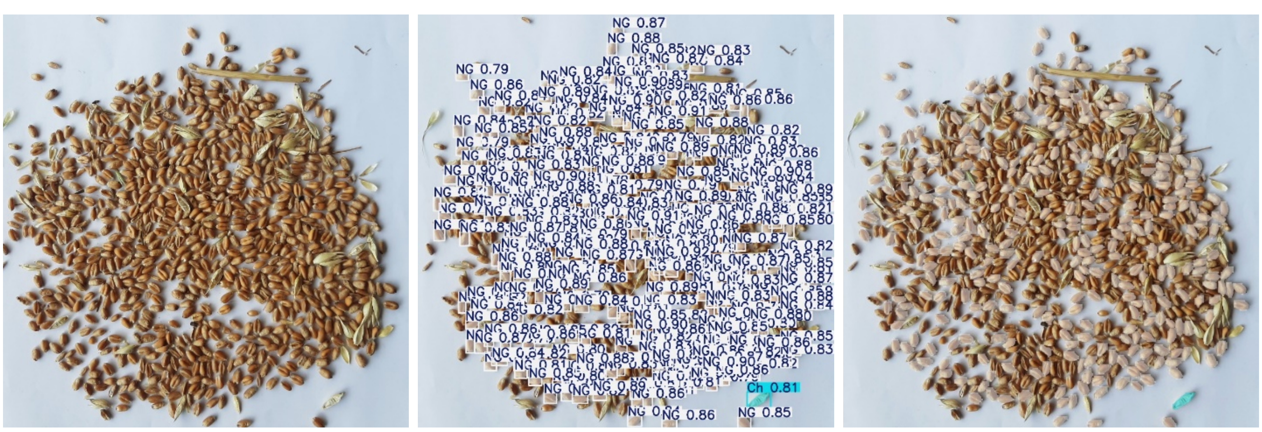
5. 3. Results of testing models on real images of winter wheat grain mixtures

Testing the trained models on images of real samples of grain material after edge clipping (Table 3) did not provide acceptable levels of detection and segmentation of objects. The YOLO11n-seg 1280 model was able to detect and segment only about 35% of objects of the NG class (whole grain) in an image with a size of 2700 × 2700 pixels (Fig. 7). At the same time, the model was unable to detect objects of the BG (broken grain), St (straw), and TG (thin grain) classes.

Testing the models on images divided into four elements showed a high level of detection of grain mixture objects (Table 7).

According to the results of testing the models on the CPU (Intel Core(TM) i5-9400 2.90GHz), the processing time for an image of size 1350 × 1350 was 115–340 ms. Testing the corresponding models on the GPU (Jetson AGX Orin 64 GB) provided an increase in performance. The processing time was reduced to 40–127 ms, depending on the architecture of the model used (from the smallest YOLOv11n-seg to the largest YOLOv11l-seg, respectively).

Analysis of the test results revealed the highest quality and accuracy of object segmentation when increasing the size of the model images to 1280 pixels (Fig. 8).



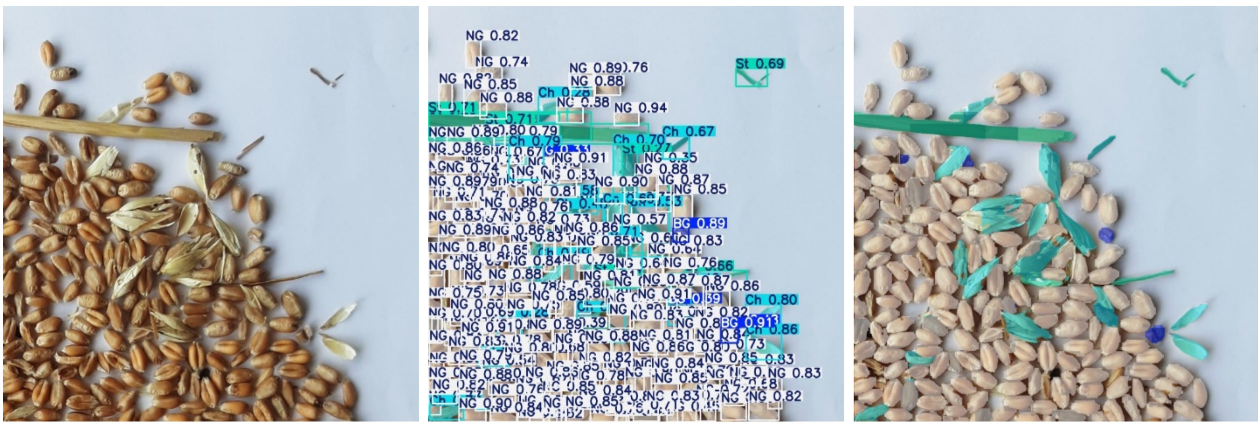
a *b* *c*

Fig. 7. Results of testing the YOLO11n-seg 1280 model on an image of a grain sample with a size of 2700 × 2700: *a* — original image; *b* — image with bounding boxes applied; *c* — image with object masks applied

Table 7

Results of testing experimental models on images of grain material of size 1350 × 1350, deployed on GPU (Jetson AGX Orin 64 GB) and CPU (Intel Core(TM) i5-9400 2.90GHz)

Model/size	Number of detected objects by class					Output time, (ms)	
	BG	Ch	NG	St	TG	GPU (Jetson)	CPU (Intel i5)
YOLO11n-seg 640	4	20	181	6	1	39.4	115.4
YOLO11n-seg 800	7	26	167	9	–	59.9	139.1
YOLO11n-seg 1280	7	22	169	8	1	90.6	324.1
YOLO11s-seg 640	8	21	168	7	1	62.0	265.4
YOLO11m-seg 640	7	24	166	6	1	105.3	447.4
YOLO11l-seg 640	13	31	170	5	1	126.8	539.5



a *b* *c*

Fig. 8. Results of testing the YOLO11n-seg 1280 model on an image of a grain sample with a size of 1350 × 1350: *a* — original image; *b* — image with bounding boxes applied; *c* — image with object masks applied

Reducing the size of model images is accompanied by an increase in the graininess of binary object masks, and therefore a decrease in segmentation accuracy.

6. Discussion of results based on the architecture of YOLO11 models for determining the fractional composition of winter wheat grain mixtures

Analysis of our results of processing digital photos obtained from the cameras of various smartphones (Table 3) reveals that after cropping and compressing them to a size of 1280 × 1280 pixels, the smallest size (width) of full grains is 22–27 pixels. For satisfactory quality of object segmentation, with a localization accuracy threshold of $IoU \geq 0.7$, objects should usually have at least 30–50 pixels on the smallest side. Therefore, when developing an algorithm for determining the fractional composition of a grain mixture sample, the tiling principle was used – dividing the image into a set of images. Thus, when dividing the input image into 4 elements ($n = 2$), on each of the elements reduced to a size of 1280 × 1280 pixels, the width of the grains will be 44–54 pixels. This is enough to create high-quality segmentation masks.

The Inference Time of the YOLO11n-seg model at 640 × 640, which is 17.1 ms on validation images and 39.4 ms on real images, makes this model suitable for real-time applications. Overall performance: mAP50 (Box: 0.816, Mask: 0.792) is satisfactory for detecting objects in grain mixtures. However, as can be seen from Table 4, the mAP50–95 (Mask: 0.478) is low, indicating that the YOLO11n-seg model has lower object segmentation accuracy at higher IoU thresholds. This will adversely affect the accuracy of determining the fractional composition of grain materials.

The YOLO11s-seg model at 640 × 640 provides an increase in mAP50-95 indicators compared to YOLO11n-seg: from 0.613 to 0.668 (Box), and from 0.478 to 0.519 (Mask). This is due to the increased power and complexity of the "s" model, which is characterized by 10.1 million parameters (weights and biases that can be trained), compared to 2.9 million parameters of the "n" model (Table 2). At the same time, the output speed on the valid set is 29.0 ms/image, which will provide a performance of 34–35 FPS for real-time applications. The output speed on test images was 62.0 ms and 265.4 ms when deployed on GPU and CPU, respectively (Table 7). Additional acceleration, without loss of accuracy, can be provided by further exporting the model to TensorRT format for deployment on Jetson Orin.

Using YOLO11m-seg at 640 × 640 demonstrates a significant improvement in accuracy due to the increased power and complexity of the model (22.3 million parameters, 123.0 GFLOPs). The loss values for YOLO11m-seg are the lowest among the previous models (Fig. 5). This means that this model converges better and generally distinguishes objects and their boundaries better than the smaller versions. The average accuracy values of Box mAP50–95 (0.727) and Mask mAP50–95 (0.558) have increased significantly compared to the previous models. This means that this model performs accurate detection and segmentation even at high IoU thresholds. The average accuracy values of mAP50 are also very high (Box 0.889, Mask 0.877), confirming its high performance. Although the accuracy has increased, the output time to the valid set has also increased to 62.5 ms, which corresponds to approximately 16 FPS. Testing the model on real grain images showed Inference Time values of 105.3 ms and 447.4 ms on GPU and CPU, respectively. This may be acceptable for many grain quality control tasks, but not for fast real-time applications.

When usi ng the YOLO11l-seg model at 640 × 640, the overall accuracy of the model slightly decreased compared to that of YOLO11m-seg (Table 4), despite the increased model power. The inference time of 74.9 ms is the highest among all previous models. This will significantly limit its use for real-time applications without further optimization. This experiment shows that increasing the complexity of the model does not always lead to a linear improvement in accuracy for a given dataset and image size, especially when it is accompanied by a significant decrease in speed.

Training the YOLO11n-seg model on the 800 × 800 size resulted in a slight improvement in accuracy in some aspects, but with the expected increase in processing time and memory usage (Table 4). The average mAP50 accuracy value remained at about the same high level (Box 0.811, Mask 0.806) as at 640 × 640. The average mAP50–95 accuracy value improved noticeably (Fig. 6) especially for Mask (from 0.478 to 0.536). This is a positive point, as it indicates better localization and accuracy of object masks. Training the YOLO11n-seg model on the 800 × 800 size resulted in an increase in the inference time from 17.1 ms to 22.1 ms, which is acceptable for real-time applications. At the same time, the GPU memory usage increased from 9.48 GB to 15.3 GB.

Training the model at 1280 × 1280 provided the best accuracy indicators for the YOLO11n-seg model, especially for the mAP50–95 metrics, which are key for accurate localization and segmentation. The Mask mAP50–95 value reached 0.640, which is a significant improvement compared to 0.536 at 800 × 800 and 0.478 at 640 × 640. This is an indicator of high segmentation quality even at very strict thresholds, which will provide increased accuracy in determining the fractional composition of grain materials. The output speed slowed down significantly compared to previous versions (from 17.1 ms to 61.4 ms at 640 × 640, and from 22.1 ms to 61.4 ms at 800 × 800). This is the expected cost of working with high-resolution images.

Increasing the size of images for training models results in each object occupying a larger number of pixels in the input image for the model. This is especially true for small objects, such as components of a grain mixture. At low resolution, objects can lose important features. As the resolution increases, small objects become "bigger" for the model, providing more pixel data and context for their detection and localization. This directly affects mAP50–95, as small objects are often "difficult" to detect with high IoU. Improved detection and more accurate localization of small objects contributes to an increase in mAP at higher IoU thresholds.

Analysis of model performance indicators (Tables 5, 6) reveals the presence of two classes of objects characterized by increased detection and segmentation accuracy: full grain (NG), and chaff (Ch) in all models, during the research. The value of segmentation accuracy mAP50 for the class (NG) reaches 0.975. This is obviously explained by the relatively simple shape of the whole grain and the significant number of representations of objects of this class (over 10 thousand) in the training dataset.

The most difficult to segment were the classes BG (broken grain) and TG (thin grain). The objects in question belong to what we define as "complex classes" for the neural network, which creates specific problems for segmentation. Apparently, this is primarily due to the reduced size of these objects compared to objects of the NG class. In addition, broken grain has significant heterogeneity of shape, which also complicates its segmentation. Many objects of the TG class are visually similar to objects of the whole grain. When categorizing grain materials, the method of sieving it on a sieve with rectangular holes is used. In this case, two visually similar grains that differ in thickness by only 0.1 mm will be assigned to two different classes. This is accompanied by a decrease in the accuracy of detecting objects of the TG class.

Increasing the image size of the models significantly affected the improvement of the segmentation performance of complex class objects (Table 6). The Mask mAP50–95 index for the BG class increased from 0.399 to 0.565 when the model image size increased from 640 × 640 to 1280 × 1280, which indicates a significantly better segmentation of crushed grains. At the same time, the detection recall (R) also increased significantly to 0.746 Box and 0.738 Mask. Similarly, for the TG class, the Mask mAP50–95 index improved from 0.494 to 0.619.

The use of more powerful models is also accompanied by an increase in the segmentation performance of complex class objects, primarily the Box mAP50–95 index (Table 5). Thus, for the BG class, the Box mAP50–95 index value increased from 0.505 (YOLO11n-seg 640 model) to 0.654 (YOLO11m-seg 640 model). Similarly, for the TG class, the Box mAP50–95 index improved from 0.595 to 0.688. At the same time, the increase in the Mask mAP50–95 index was less noticeable.

So, according to the results of our research, it was found that the YOLO11m-seg 640 × 640 model provides the best accuracy of object localization at a speed of about 16 FPS. The YOLO11s-seg 640 × 640 model provides a processing speed of 34–35 FPS with quite decent accuracy. Training the YOLO11n-seg model at 1280 × 1280 provided the best average segmentation accuracy by improving the performance of detecting objects of "complex" classes, which are key for accurately determining the fractional composition of grain mixtures.

It should be noted that there is a trade-off between the speed and accuracy of the models. Increasing both the input image size and the complexity of the models allowed the model to see more detail, which led to better localization of objects, but this leads to an increase in image processing time.

Unlike [16, 18], in which the use of YOLO models makes it possible to categorize and determine the number of objects of the corresponding classes, the use of YOLO-Segmentation models makes it possible to measure the area or shape of each object. This becomes possible due to the definition of the exact boundaries of each individual grain and at the pixel level even if they overlap.

Constructing a custom dataset specifically designed for the analysis of wheat fractional analysis, taking into account the features of overlap and the diversity of fractions, ensures that the training data is suitable for the task. This is critically important for high-accuracy segmentation of grain mixture objects.

Thus, our study converts the intuitive choice of neural network architecture into data based on empirical data. Recommendations are given that make it possible to reasonably choose the YOLO-seg architecture for determining the fractional composition of wheat on computing devices, ensuring the optimal balance between accuracy, speed, and efficient use of resources. This significantly reduces the risks of ineffective deployment and accelerates the implementation of quality control systems in agriculture.

A limitation of our study is the completeness of fractions when training the models. Under actual conditions, grain samples may contain other types of impurities or damage:

sprouted grains, weed seeds, grains of other crops, clods of earth. The model will not be able to recognize objects that were not included in the dataset.

Among the shortcomings, one should note the reduced performance for the "complex" classes: crushed and thin grains. This indicates that the models have significant difficulties with the accurate segmentation of these heterogeneous or small objects. This limits the accuracy of the full fractional analysis.

Further research will involve collecting additional, diverse data for complex classes, careful annotation of complex cases of overlapping or defects of objects, conducting training with data augmentation, which simulates conditions that complicate their recognition. It is also planned to optimize the model for deployment – converting the model to TensorRT, quantizing the model to accelerate the inference, integration into separation systems or quality control of grain materials.

7. Conclusions

1. Our analysis of images of grain material samples reveals that even with an increase in the size of the neural model to 1280×1280 pixels, it is not possible to achieve acceptable object sizes for satisfactory segmentation quality with a localization accuracy threshold of $\mathrm{Io}U \geq 0.7$. Given this, the algorithm for determining the fractional composition of a grain mixture sample provides for the use of the method for dividing the intermediate image into a set of n^2 elements, and sequential analysis of each of the image elements.

2. Based on the results of training models from the YOLO11-seg family on the collected data set, which contained 280 images from over 17 thousand objects of five classes of winter wheat grain materials, it was established that there is a trade-off between the speed and accuracy of different models. An increase in both the input image size and the complexity of the models leads to better localization of objects, but this is accompanied by an increase in image processing time. It was found that the YOLO11m-seg 640×640 model provides the best accuracy of object localization at a speed of about 16 FPS. The YOLO11s-seg 640×640 model provides a processing speed of 34–35 FPS with quite decent accuracy. Training the YOLO11n-seg 1280×1280 model provided the best average segmentation accuracy by increasing the performance of identifying objects of "complex" classes, which are key for accurate determination of the fractional composition of grain mixtures.

3. Analysis of our results from testing the trained models on images of real samples of grain materials revealed the highest quality and accuracy of object segmentation when increasing the model size to 1280 pixels and dividing the image into four elements. Reducing the size of the model images was accompanied by an increase in the graininess of the binary masks of the objects. According to the results of testing the models on the CPU (Intel Core(TM) i5-9400 2.90 GHz), the processing time of an image of size 1350×1350 was 115–340 ms. Testing the corresponding models on the GPU (Jetson AGX Orin 64 GB) provided an increase in performance. The processing time was reduced to 40–127 ms.

Conflicts of interest

The authors declare that they have no conflicts of interest in relation to the current study, including financial, personal, authorship, or any other, that could affect the study, as well as the results reported in this paper.

Funding

The research was carried out as part of the project "Design of a system for analyzing and recognizing grain materials using machine vision" within the framework of grant topic 2023.04/0040, funded by the National Research Foundation of Ukraine.

Data availability

The data will be provided upon reasonable request.

Use of artificial intelligence

The authors used artificial intelligence technologies within acceptable limits to provide their own verified data, which is described in the research methodology section.

References

1. Walker, C. K., Assadzadeh, S., Wallace, A. J., Delahunty, A. J., Clancy, A. B., McDonald, L. S. et al. (2023). Technologies and Data Analytics to Manage Grain Quality On-Farm – A Review. Agronomy, 13 (4), 1129. https://doi.org/10.3390/agronomy13041129
2. Barbedo, J. G. A. (2025). A Review of Artificial Intelligence Techniques for Wheat Crop Monitoring and Management. Agronomy, 15 (5), 1157. https://doi.org/10.3390/agronomy15051157
3. Looh, G. A., Xie, F., Wang, X., Looh, A. N., Hind, H. (2025). Grain kernel damage during threshing: a comprehensive review of theories and models. Journal of Agricultural Engineering. https://doi.org/10.4081/jae.2025.1674
4. Stepanenko, S., Kotov, B., Kuzmych, A., Kalinichenko, R., Hryshchenko, V. (2023). Research of the process of air separation of grain material in a vertical zigzag channel. Journal of Central European Agriculture, 24 (1), 225–235. https://doi.org/10.5513/jcea01/24.1.3732
5. Zhao, W., Liu, S., Li, X., Han, X., Yang, H. (2022). Fast and accurate wheat grain quality detection based on improved YOLOv5. Computers and Electronics in Agriculture, 202, 107426. https://doi.org/10.1016/j.compag.2022.107426
6. Saini, M., Singh, J., Prakash, N. R. (2014). Analysis of Wheat Grain Varieties Using Image Processing: A Review. International Journal of Science and Research (IJSR), 3 (6), 490–495. Available at: https://www.ijsr.net/getabstract.php?paperid=2014254
7. Rong, D., Wang, H., Xie, L., Ying, Y., Zhang, Y. (2020). Impurity detection of juglans using deep learning and machine vision. Computers and Electronics in Agriculture, 178, 105764. https://doi.org/10.1016/j.compag.2020.105764
8. Gupta, P. K., Waoo, A. A. (2024). Analyzing machine learning methods to enhance grain quality assessment and evolution. ShodhKosh: Journal of Visual and Performing Arts, 5 (1). https://doi.org/10.29121/shodhkosh.v5.i1.2024.1876

9. Sirohi, A., Malik, A. (2021). A Hybrid Model for the Classification of Sunflower Diseases Using Deep Learning. 2021 2nd International Conference on Intelligent Engineering and Management (ICIEM), 58–62. https://doi.org/10.1109/iciem51511.2021.9445342
10. Ünal, Y., Dudak, M. N. (2024). Deep Learning Approaches for Sunflower Disease Classification: A Study of Convolutional Neural Networks with Squeeze and Excitation Attention Blocks. Bitlis Eren Üniversitesi Fen Bilimleri Dergisi, 13 (1), 247–258. https://doi.org/10.17798/bitlisfen.1380995
11. Hu, K., Coleman, G., Zeng, S., Wang, Z., Walsh, M. (2020). Graph weeds net: A graph-based deep learning method for weed recognition. Computers and Electronics in Agriculture, 174, 105520. https://doi.org/10.1016/j.compag.2020.105520
12. Kaldarova, M., Akanova, A., Nazyrova, A., Mukanova, A., Tynykulova, A. (2023). Identification of weeds in fields based on computer vision technology. Eastern-European Journal of Enterprise Technologies, 4 (2 (124)), 44–52. https://doi.org/10.15587/1729-4061.2023.284600
13. Qiu, Z., Wang, F., Wang, W., Li, T., Jin, X., Qing, S., Shi, Y. (2024). YOLO-SDL: a lightweight wheat grain detection technology based on an improved YOLOv8n model. Frontiers in Plant Science, 15. https://doi.org/10.3389/fpls.2024.1495222
14. Dang, F., Chen, D., Lu, Y., Li, Z. (2023). YOLOWeeds: A novel benchmark of YOLO object detectors for multi-class weed detection in cotton production systems. Computers and Electronics in Agriculture, 205, 107655. https://doi.org/10.1016/j.compag.2023.107655
15. Sharma, D., Sawant, S. D. (2017). Grain quality detection by using image processing for public distribution. 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), 1118–1122. https://doi.org/10.1109/iccons.2017.8250640
16. Kaya, E., Saritas, İ. (2019). Towards a real-time sorting system: Identification of vitreous durum wheat kernels using ANN based on their morphological, colour, wavelet and gaborlet features. Computers and Electronics in Agriculture, 166, 105016. https://doi.org/10.1016/j.compag.2019.105016
17. Kong, Q., Ma, N. (2025). Wheat grain appearance quality detection based on improved YOLOv8n. INMATEH Agricultural Engineering, 356–365. https://doi.org/10.35633/inmateh-75-30
18. Shen, Y., Yin, Y., Li, B., Zhao, C., Li, G. (2021). Detection of impurities in wheat using terahertz spectral imaging and convolutional neural networks. Computers and Electronics in Agriculture, 181, 105931. https://doi.org/10.1016/j.compag.2020.105931
19. Ma, N., Su, Y., Yang, L., Li, Z., Yan, H. (2024). Wheat Seed Detection and Counting Method Based on Improved YOLOv8 Model. Sensors, 24 (5), 1654. https://doi.org/10.3390/s24051654
20. Fan, L., Ding, Y., Fan, D., Di, D., Pagnucco, M., Song, Y. (2022). GrainSpace: A Large-scale Dataset for Fine-grained and Domain-adaptive Recognition of Cereal Grains. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 21084–21093. https://doi.org/10.1109/cvpr52688.2022.02044
21. Zhang, J., Min, A., Steffenson, B. J., Su, W.-H., Hirsch, C. D., Anderson, J. et al. (2022). Wheat-Net: An Automatic Dense Wheat Spike Segmentation Method Based on an Optimized Hybrid Task Cascade Model. Frontiers in Plant Science, 13. https://doi.org/10.3389/fpls.2022.834938
22. Guo, X., Wang, J., Gao, G., Cheng, Z., Qiao, Z., Zhang, R. et al. (2025). LWheatNet: a lightweight convolutional neural network with mixed attention mechanism for wheat seed classification. Frontiers in Plant Science, 15. https://doi.org/10.3389/fpls.2024.1509656
23. Stepanenko, S., Kuzmych, A., Kharchenko, S., Andrey, B., Dnes, V., Volyk, D., Kalinichenko, R. (2025). A Machine Vision Approach for Grain Quality Control During Separation. Journal of Engineering Sciences, 12 (1), E9–E17. https://doi.org/10.21272/jes.2025.12(1).e2
24. NVIDIA® Jetson AGX Orin™ 64GB Developer Kit. Stereolabs.
25. Segmentation Models. Ultralytics. Available at: https://docs.ultralytics.com/tasks/segment/#models