Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/180204

Neural network methods for the detection of farm animals in dense dynamic groups on images

Olga Ivashchuk[1], *Lyazzat Atymtayeva* [2], *Alexei Zhigalov* [1], *Bagdat Yagalieva* [3,*], *Oleg Ivashchuk* [4], *and Vyacheslav Fedorov* [1]

[1]Department of Information and Robotic Systems, Belgorod State National Research University, Belgorod, and Russia
[2]Department of Computer Science, Faculty of Engineering and Science, Suleyman Demirel University, Kaskelen, Kazakhstan
[3]Department of Cybersecurity, Information Processing and Storage, Institute of Automation and Information Technologies, Satbayev University, Almaty, Kazakhstan
[4]Department of Computer Science, Faculty of Science and Technology, Yessenov University, Aktau, Kazakhstan

Received: 27 Sep. 2023, Revised: 29 Dec. 2023, Accepted: 3 Jan. 2024
Published online: 1 Mar. 2024

Abstract: Currently, there is an urgent need for non-invasive monitoring of farm animals' health status, enabling swift responses to adverse situations such as morbidity, feeding disorders, and aggression. Globally, technologies for video monitoring of animals are being developed, which include image processing using intelligent methods, especially artificial neural networks. This paper presents the results of developing and investigating methods and models for detecting (individually identifying) farm animals, with a focus on pigs as a case study. These animals are located in dense, dynamic groups within agricultural complexes where traditional identification methods are less effective. To overcome this challenge, advanced neural network architectures, specifically Faster R-CNN and YOLOv5, were selected, finely tuned, and trained. The application of the YOLOv5 network achieved a detection accuracy with a mean Average Precision (mAP) of 94.05%, surpassing the accuracy demonstrated in comparable studies. These results provide a foundation for a hardware-software complex designed for non-invasive, automated monitoring of animal conditions, integrating intelligent data analysis. This system offers crucial support for science-based decision-making in the fields of animal husbandry and food security management.

Keywords: artificial neural network (ANN), detection, tracking, face recognition, animal recognition and identification, non-invasive automated monitoring, YOLOv5

1 Introduction

In the contemporary world, knowledge-intensive technologies based on artificial neural networks (ANNs) are rapidly evolving. Their applications extend to monitoring the state and movement of living objects in agricultural production. It is crucial to emphasize that key indicators of technological efficiency in animal husbandry include productivity, labor efficiency, and material intensity [1,2]. Notably, improvements in productivity significantly depend on the ability to refine livestock control and ensure animal well-being. Labor intensity can be reduced by introducing automation, while material intensity can be decreased by moving away from the currently used radio frequency identification (RFID) animal tags, which are known for their substantial costs. The use of video surveillance followed by neural network processing of the information presents an effective solution to these challenges. Specifically, this approach can greatly improve the accuracy of detecting abnormal animal behavior, such as disease, aggression, feeding disorders, etc., eliminating the need for invasive methods like tags, ear tags, or body-mounted devices like "trackers." Moreover, it facilitates automated monitoring of growth, weight gain, and animal counting [3,4]. Researchers worldwide are actively developing neural network methods in this area, which can be broadly categorized into the following main classes:

Detection (isolating individuals) of live objects in photo or video images;

Identification of live objects (determining the identity of an unknown object relative to a known one);

Tracking of movements ("tracking");

Determining the qualitative condition of live objects.

* Corresponding author e-mail: b.yagaliyeva@satbayev.university

The evolution of Artificial Neural Networks (ANNs) has progressed to a point where specific architectures (i.e., configurations of layers) offer effective solutions for a range of tasks, from image processing to text and sound analysis. However, to achieve optimal results in specialized tasks, fine-tuning of ANNs is often required. This process involves selecting appropriate ANN configurations, modifying layers, adjusting the number and order of layers, selecting hyperparameters (parameters that remain constant during network training or change according to a predetermined rule), and more.

In the field of neural network development for animal observation, certain unique aspects must be considered. For example, individual animals of certain species exhibit minimal differences, and animals often congregate in dense, dynamic groups. This paper introduces a method for detecting farm animals in images, focusing on pigs as a case study. The challenge of recognizing pigs arises from their lack of distinctive patterns, particularly on their muzzles, and the potential for contamination to hinder accurate identification. The tendency of pigs to cluster tightly, whether in motion or at rest, complicates individual identification and movement tracking.

The specific issue addressed in this paper is the detection of pig faces in group settings using deep neural networks. While the term "face" is used analogously to human face detection, developing neural networks for animal detection is a distinct task due to the unique external characteristics and behaviors of animals. By employing well-established object detection architectures such as YOLOv5 [5], Faster R-CNN based on ResNet-50 [6,7,8], and MobileNetV3 [9], the networks are fine-tuned to accommodate the behavioral patterns of pigs. Trained on a dataset comprising 1200 images, including those with densely grouped animals, a comparative analysis of the networks' detection quality and speed is conducted. The results demonstrate accuracy (mAP - Mean Average Precision) significantly surpassing that of similar studies [10], particularly in detecting pig faces, with the YOLOv5 network showing the highest accuracy [11].

Few studies focus on animal recognition methods, especially in the context of detecting pig faces. A comprehensive review of both domestic and international literature reveals that this research achieves the most accurate results in detecting pig faces within group images. The scientific novelty lies in the fine-tuning of current neural networks for object detection, leading to high accuracy and speed in detecting pig faces.

This study on pig face detection initiates a series of research projects aimed at developing a comprehensive hardware and software system for the automatic identification, tracking, and classification of pig behavioral patterns. The ultimate goal is the early detection of diseases and aggressive behavior, optimization of climate control, and establishment of appropriate feeding regimes in pig farming.

2 Methods for detection and identification of animals in images

Object detection in an image involves identifying bounding boxes around objects within the pixel coordinate system of the original image and classifying these objects into predefined categories. The challenge is heightened by the variability in the number of objects present within a single image. Modern solutions predominantly utilize neural networks to tackle this problem effectively. Typically, the objective is to define a rectangular bounding box, with sides parallel to the dimensions of the original image. Neural networks process images represented as arrays of size CxHxW, where 'C' denotes the number of channels, 'H' the height, and 'W' the width of the image.

In addressing the problem of object identification, locating objects within the image is the initial step. Interestingly, several studies on pig identification lack detailed methods for isolating individual animals in photographs, focusing instead on identification algorithms using images where objects, particularly pig "faces," have already been segregated [12,13,14,15,16,17].

Existing research on pig detection predominantly focuses on the comprehensive detection of entire pig entities. For instance, [18] proposes a system that discerns pigs' posture, average speed, and distance traveled to diagnose overall health conditions, including pain, musculoskeletal issues, and digestion problems. The authors employ a neural network capable of simultaneously detecting individual pigs and classifying their behavior. The study compares the performance of the YOLO9000 [19] and Faster R-CNN [6] networks, both based on the ResNet-50 architecture [8] for feature extraction. It concludes that the YOLO9000 network excels in speed and accuracy. Another study [20] applies the YOLO9000 network to detect pig bodies under varied lighting conditions. Furthermore, [21] explores pig detection using the same architecture, introducing an innovative algorithm to address the challenge of occlusion among closely situated individuals. In [22], pig detection using Faster R-CNN is combined with pig pose classification. Subsequent studies [23,**?**] extend detection to whole pig bodies, focusing on identifying key points that accurately represent each animal's pose, thereby determining their location within the pen. The first of these studies employs a neural network based on the SegNet architecture [25], while the second utilizes a Mask R-CNN image segmentation network [26].

There is a notable scarcity of research addressing methodologies for detecting pig "faces." In [15,**?**], pig "face" detection is implemented as the initial step in addressing the broader identification challenge. It's important to note that the datasets in these studies predominantly feature images capturing single individuals, significantly simplifying the detection process. The first paper uses the Faster R-CNN network for face detection, while the second explores various popular ANN architectures designed for detection. Despite acknowledging the high accuracy of YOLOv3, the study favors the EfficientDet-D0 network [28] for its balance of detection quality and speed.

In [11], the authors modified the YOLOv3 network to detect pig "faces" in images of group settings and achieved notable results in Mean Average Precision (mAP): 90.18 % for the Intersection over Union (IoU) threshold $T_{IoU} = 0.5$. A detailed explanation of mAP and IoU is provided in Section 4 of this paper.

Thus, the main conclusions are as follows:

There are few studies focusing on the detection of pig "faces," with most concentrating on single individuals in images;

Pig "faces" in groups are detected in [11], achieving an mAP of 90.18% at an IoU threshold $T_{IoU} = 0.5$, which, to the best of the authors' knowledge, is the only study of its kind available in popular scientific sources online;

Several papers on pig detection have employed neural networks from the R-CNN and YOLO series.

The authors suggest that, similar to human identification methods, utilizing pig "faces" may yield more precise results than using other body parts. Therefore, for effective and non-invasive monitoring of pigs on farms through video cameras, it is crucial to tackle the challenge of detecting pig "faces" within dense, dynamic groups. This facilitates subsequent identification and tracking. Previous studies have demonstrated the effectiveness of R-CNN and YOLO series networks in pig detection. This study employs contemporary adaptations of these networks to achieve robust results.

3 The choice of neural network architecture for pig 'face' detection

The goal of the detection task in this study is to enable the neural network to accurately locate pig "faces" in images and delineate them with rectangular frames.

Over the past few years, two primary approaches to neural network-based detection have become prominent:

The R-CNN series of algorithms, representing a two-stage detection methodology. Initially, these algorithms identify "regions of interest" where the target object is likely to be located. In the subsequent stage, objects within these regions are classified, and the coordinates of the bounding boxes are refined.

The YOLO series of algorithms, which adopt a one-stage approach. Here, the network simultaneously predicts the coordinates of a set number of bounding boxes, alongside classifying the objects and estimating the probability of their presence. The coordinates of these boxes are iteratively refined during the training process.

As noted in [29], the YOLO series is renowned for its speed, while the R-CNN series is recognized for its accuracy in pinpointing the coordinates of intricately shaped and closely spaced small objects.

In this research, we explored, fine-tuned, and trained Artificial Neural Networks (ANNs) for pig face detection using two distinct architectural approaches: YOLOv5 [5] and Faster R-CNN [6] enhanced with the Feature Pyramid Network (FPN) mechanism [30]. The FPN mechanism is known for improving the detection of small objects, as evidenced in [30]. For the Faster R-CNN, we employed two different backbone types: ResNet-50 [8] and MobileNetV3 [9]. In convolutional neural networks, the "backbone" refers to the deep convolutional network that extracts feature maps from an image, where a "feature map" is an array representing the crucial features of the processed image. The choice of YOLOv5 and Faster R-CNN with FPN (hereafter referred to as Faster R-CNN-FPN) is motivated by the former's current status as a leading network in object detection and the latter's potential efficacy in detecting pigs, particularly in dense group settings.

The implementation of the Faster R-CNN-based ANN was done using the torchvision.models.detection library [31], employing the class faster_rcnn.FasterRCNN [32]. Functions like fasterrcnn_resnet50_fpn [33] and fasterrcnn_mobilenet_v3_large__fpn [34] were used to implement the ResNet-50 and MobileNetV3-large backbones with FPN, respectively. For YOLOv5 [5], we used the PyTorch library.

We applied a transfer learning strategy to each variant of the ANNs. This approach uses neuronal weights pre-trained on the MS COCO dataset [35] for initialization, enhancing the learning process. The MS COCO dataset, known for its extensive collection of over 200 thousand annotated images with more than 1.5 million objects, is widely utilized for training neural networks. By leveraging these pre-trained weights, our focus was on refining the last few layers of the network, often termed "unfrozen" layers, rather than training the entire network from scratch on the target dataset.

4 The results of different types of neural networks for the detection of pig 'faces' in images

For the sake of generality, the "faces" of the pigs are hereafter referred to as relevant objects. It should be noted that neural network detection algorithms can be used to select objects of one class or several at the same time, but in this case only objects of one relevant class are searched.

We used 1200 and 130 (respectively) open-source pictures of pigs as training and evaluation datasets for the trained network. The training dataset is referred to as training dataset and the validation dataset is referred to as validation dataset in Russian technical literature.

The quality of the trained ANNs was evaluated on a validation dataset using traditional object detection quality metrics [10]. The following is a brief description of them.

In the validation process, the ANN outputs the coordinates of the bounding rectangles, and for each rectangle, the probability (in articles in English, the confidence score) that the rectangle B_p contains an object of a particular class, in this case, the "face" of a pig. When calculating the quality metrics of the detection methods for each predicted rectangle one must determine, first, whether the object in it is relevant, and, second, whether the rectangle's coordinates are found accurately enough. The accuracy of finding rectangle coordinates is characterized by the *IoU (Intersection over Union)* score. For each relevant object from the marked-up dataset found by the neural network, the coordinates of the bounding rectangle B_{gt}, determined at marking *(the original rectangle or Ground-truth bounding box)*, and the rectangle B_p, predicted by the neural network (hereafter, *the predicted rectangle or Predicted bounding box*) are known. $IoU = \frac{S(B_p \bigcap B_{gt})}{S(B_p \bigcup B_{gt})}$ - is the ratio of the intersection area to the union area (respectively) of *the original rectangle* and the *predicted rectangle* (respectively). If $IoU > T_{IoU}$ (where T_{IoU} is some threshold for IoU), it is assumed that the rectangle coordinates are correctly defined.

Predicted objects, depending on the "correctness" of the prediction of their class and location, are categorised into groups: "*True Positive, False Positive and False Negative*".

TP, FP and is the number of *True Positive*, *False Positive* and *False Negative* objects (respectively) on all images in the dataset (in this case the validation dataset).

Let T_{score} be some threshold for *confidence score*. A predicted object is *True Positive* if three conditions are met:$confidence\ score > T_{score}$, $IoU > T_{IoU}$ and the object class is defined correctly (as stated above, in our case there is only one class, the "face" of the pig). A predicted object is False Positive if $confidence\ score > T_{score}$, but either the object $IoU \leq T_{IoU}$ class is not correctly defined.

Let N be the total number of relevant objects in all the photos in the dataset (i.e. the number of rectangles B_{gt}). Then the number of objects is *False Positive*: $FN = N - TP$. After calculating TP, FP and FN we can find the following quality metrics: $Precision = \frac{TP}{TP+FP}$ and $Recall = \frac{TP}{TP+FN} = \frac{TP}{N}$.

An important indicator of the detection quality is AP *"Average Precision"* – a certain integral of the function representing the dependence *Precision* on *Recall*, (within [0, 1], since the values of *Precision* and *Recall* are from 0 to 1) and at different values of T_{score}. In order to calculate AP the sequence of values T_{score}: $\{T_{score,0}, T_{score,2}, \dots, T_{score,i}, \dots, T_{score,k}\}$ is put and for each value T_{score} of this sequence we find the values of TP, FP and FN, and we calculate *Precision* and *Recall*, then we approximate calculate the definite integral of the function $Precision = F(Recall)$ between 0 and 1, i.e. the area under the graph of this function. Note that the curves$Precision = F(Recall)$ will generally be different at different thresholds T_{IoU}, because TP and FP will also be different T_{IoU} and, and, consequently, the values AP will be different.

Let $\{r_0, r_1, \dots, r_i, \dots, r_k\}$ be a sequence of values *Recall* and $\{p_0, p_2, \dots, p_i, \dots, p_k\}$ – be a sequence of values of *Precision*: $p_i = Recall(r_i)$ $(i = 0, \dots, k)$. To approximate the definite integral for the purpose of estimation AP, for each one r_i $(i = 0, \dots, k)$ we find: $\hat{p}(r_i) = \max(p_j = Precision(r_j)\big|_{r_j \geq r_i})$. Then the estimate$AP$ is calculated as:

$$\widehat{AP} = \sum_{i=0}^{k-1} (r_{i+1} - r_i)\hat{p}(r_{i+1}). \tag{1}$$

Two popular choices$\{r_0, r_1, \dots, r_i, \dots, r_k\}$ for computing $\widehat{AP}$ [10] are the methods originally implemented for PASCAL VOC 2010-2012 [36,37,38] and MS COCO datasets [35] (respectively). In the first case, in the sequence $\{r_0, r_1, \dots, r_i, \dots, r_k\}$ $r_i = Recall(T_{score,i})$ (let us call: the PASCAL VOC method), in the second case

$$\{r_i|_{i=0,\dots,100}\} = \{0.0 + i \cdot 0.01 \ \ (i = 0, \dots, 100)\} \tag{2}$$

points in number $k = 101$, evenly spaced on the axis (called: MS COCO method). The PASCAL VOC method has one more peculiarity: if several B_p correspond to oneB_{gt} (that is for them $IoU > T_{IoU}$), then B_p the one with the highest *confidence score* is considered to be *True Positive*, and the others are False Positive. Depending on these methods the sets of estimates AP given in the articles are slightly different.

Table 1, containing the results of this work, gives the following estimates of the indicatorAP (the first one is calculated by the PASCAL VOC method, the others by MS COCO): $AP^{PASCAL\ VOC} = \widehat{AP}$, where $\widehat{AP}$ calculated by formula (1) and $r_i = Recall(T_{score,i})$ where $T_{IoU} = 0.5$; $AP^{[0.50:.05:.95]}$, calculated by formula:

$$\{AP^{@[0.50:.05:.95]} = \frac{1}{10} \sum_{\substack{T_{IoU} = 0.5 + 0.05 \cdot j \\ j = 0, \dots, 9}} \widehat{AP_j}\} \tag{3}$$

where $\widehat{AP_j}$ is by formula (1), where r_i are the elements of the sequence defined by formula (2); $AP^{@.50}$ and $AP^{@.75}$, representing the values of $\widehat{AP}$, calculated by formula (1), where r_i are the elements of the sequence defined by formula (2), at $T_{IoU} = 0.5$ and $T_{IoU} = 0.75$ (respectively); AP^{small}, AP^{medium}, AP^{large}, representing the values of $\widehat{AP}$, calculated by formula (3) for objects with the area of the rectangle S, bounded by B_p, is within: $S \leq 32^2$, $32^2 < S < 96^2$, $S \geq 96^2$ (in pixels) respectively.

Note. If a dataset contains objects of more than one class, the average for all classes is usually denoted: *mAP* (Mean average precision). Since a single class is considered here, then $mAP^{PASCAL\ VOC} = AP^{PASCAL\ VOC}$.

For the MS COCO method the original website [35] indicates that *mAP* the notation the notation, and by $AP^{[0.50:.05:.95]}$, $AP^{@.50}$, $AP^{@.75}$, AP^{small}, AP^{medium}, AP^{large} the mean value of these indicators for all classes is understood.

Table 1 shows the results of the evaluation of the quality of face detection of pigs in images using different types of ANN (in %), as well as the average face detection time per photo (in milliseconds).

The information in Table 1 shows that the detection quality of the "faces" of the pigs by the YOLOv5x6 network, is significantly higher than that of the Faster R-CNN-FPN with ResNet-50 and with MobileNet-v3-large as backbone. The average face detection time per photo with Faster R-CNN-FPN with MobileNet-v3-large as backbone is slightly less than that of YOLOv5x6, but the difference is not significant.

Figures 1 and 2 show the results of face detection in pigs using YOLOv5 in the modern modification of YOLOv5x6 (Figures 1-left, 2-left) and Faster R-CNN-FPN with ResNet-50 backbone (Figures 1-right, 2-right).

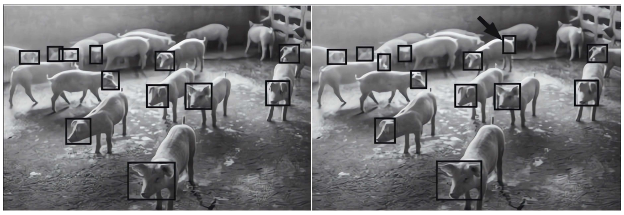


Fig. 1: Results of pig 'face' detection for YOLOv5 and Faster R-CNN-FPN with backbone ResNet-50 (paddock photo)

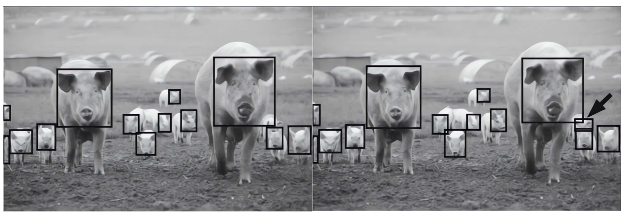


Fig. 2: Results of pig 'face' detection for YOLOv5 and Faster R-CNN-FPN with backbone ResNet-50 (photo in the field)

It is visible that both neural networks predicted almost all the 'faces' of the pigs in the pictures. Figure 1-right shows with an arrow that the Faster R-CNN-FPN neural network incorrectly predicted one rectangle. Figure 2-right shows an arrow with a rectangle at the "face" of a pig (small size, in a dense group), which Faster R-CN-FPN detected, but YOLOv5x6 did not.

Thus, it can be concluded from the experimental results that YOLOv5x6 is the best network (out of those discussed in the article) with respect to quality and detection speed.

Each of the neural networks considered was trained over 100 *epochs*. During one epoch the whole training dataset, divided into *batches*, "passes through" (i.e. all neural network layers process in turn) the neural network. In one iteration of training one *batch* is fed to the input of the neural network, and then it is sequentially transformed by all neural network layers. At the output of the last layer ANN receives a batch of output signals and calculates the value of some error function (*loss function*) that characterizes the difference between the output signals of ANN and the desired result. Then the trained ANN parameters are adjusted so as to reduce the value of the error function. Various modifications of the gradient descent method (*optimizers*) are used to minimize the error function. Figure 3 shows graphs characterizing the training process of the YOLOv5x6 neural network, which showed the highest quality of detection among the considered networks. During training the following parameters were used: the size of the data batch - 12, the number of object classes - 1, the initial image was converted to the size 400x400. The values of other parameters were chosen "by default" [5]. Figures 3-a, 3-c show the change in the loss functions used to estimate the discrepancy between the coordinates of the original and predicted rectangles during training and validation respectively, and Figures 3-b, 3-d show the change in the loss functions used to estimate the discrepancy between the true and predicted object classes during training and validation respectively.

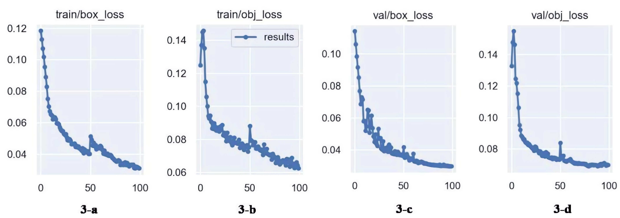


Fig. 3: Change in loss functions during training and validation of the YOLOv5x6 neural network

The values of the lossfunctions decrease from epoch to epoch, including for the validation set, indicating that the neural network is learning effectively.

For the calculations we used a computer with the following components: Intel Code i5-10600K processor with characteristics: 12 cores, frequency - 4.10 GHz; memory (RAM): 64GB; NVIDIA GeForce RTX 3090 GPU with: 24GB GDDR6X video memory, 1755/19500MHz GPU/memory, 10496 MPUs. Development environment: Jupiter Notebook.

Table 1: Assessment of the quality of the detection of pig ”faces” in the images (in %) and the average time to detect objects in one photo (ms)

The type of neural network	**Backbone**	$mAP^{PASCAL\ VOC}$	$AP^{[0.50:.05:.95]}$	$AP^{@.50}$	$AP^{@.75}$	AP^{small}	AP^{medium}	AP^{large}	**Average detection time of pig ”faces” per photo, ms**
Faster R-CNN-FPN [30]	ResNet-50	88,18	65,06	87,41	75,58	37,03	46,71	71,49	40,1
	MobileNet-v3-large	87,22	66,02	87,20	76,52	11,83	40,25	75,48	21,5
YOLOv5x6 [30]	YOLOv5x6	**94,05**	78,80	**93,45**	86,84	51,58	59,68	85,67	25,3

Calculations were performed on the GPU using CUDA - Compute Unified Device Architecture - a hardware-software architecture for parallel computing. The CUDA compiler driver (nvcc): NVIDIA (R) Cuda compiler driver, Cuda compilation tools, release 11.7, V11.7.64. Programming language: python 3.9 using machine learning framework PyTorch (v1.11.0) and software libraries: NVIDIA CUDA® Deep Neural Network (cuDNN v8400), torchvision (v0.12.0), numpy (v1.22.4), matplotlib (v3.5.2), albuptions (v1.0.3), OpenCV (v4.5.5).

5 Conclusion

In this research, we addressed the following key tasks:

Selection of promising methods for detecting pig ”faces” in dense groups within images, based on a review of scientific and technical literature. Notably, we focused on the Faster R-CNN-FPN neural network architectures, which include deep neural networks from the ResNet and MobileNet series as their basis, and YOLOv5.

Acquisition of a dataset comprising images of pig ”faces” for ANN training.

Fine-tuning of the ANNs: This involved selecting the backbone, a deep convolutional network responsible for feature map extraction, determining the optimal number of ”unfrozen” layers in the backbone for training, and choosing and tuning optimizer parameters. These parameters included the learning rate and its adjustment during training, batch size, the number of training epochs, and applying data augmentation techniques to subtly vary images across different training epochs.

Training of the ANNs.

Analysis of the detection results.

Applying the YOLOv5 network in its YOLOv5x6 modification led to a detection accuracy of $mAP^{PASCAL\ VOC} = 94,05\%$ at an Intersection over Union (IoU) threshold of 0.5. This accuracy significantly surpasses the results shown in [11]: $mAP^{PASCAL\ VOC} = 90,18\%$. To the best of the authors' knowledge, at the time of writing, no other works have been published on the detection of pig ”faces” in group images. Among the neural networks evaluated in this study, YOLOv5x6 demonstrated the most impressive results in terms of detection quality, albeit slightly trailing behind the Faster R-CNN-FPN network with the MobileNet-v3-large backbone.

Directions for Further Research:

Conducting experiments under real-world conditions at pig farms, which includes installing and configuring video cameras and computers for neural network data processing.

Developing methods for identifying pigs by their ”faces.”

Creating methods for tracking pigs, particularly when identifying them by their ”faces.”

Classifying pig behavior to hypothesize about diseases, suboptimal room temperatures, and other environmental or health factors.

Acknowledgement

The research work was financially supported by the Russian Science Foundation (project no. 22-11-20016) on

the subject of “Development and research of an intelligent decision support system for the adaptation” of agricultural areas in terms of the greenhouse effect dynamics.
The authors express their sincere gratitude to the anonymous referee for their meticulous review and valuable comments, which have significantly contributed to the enhancement of this paper.

References

[1] K.R. Shanmugam and A. Venkataramani, Technical Efficiency in Agricultural Production and Its Determinants: An Exploratory Study at the District Level, *Ind. Jn. of Agri. Econ.*, **61**, (2006).

[2] G.V. Fedotova, I.F. Gorlov, A.V. Glushchenko, M.I. Slozhenkina and A.K. Natyrov, Trends of Scientific and Technical Development of Agriculture in Russia, *Digital Economy: Complexity and Variety vs. Rationality*, **87**, (2020).

[3] S. Neethirajan, The role of sensors, big data and machine learning in modern animal farming, *Sensing and Bio-Sensing Research*, **29** 100367 (2020).

[4] Q. Yang and D. Xiao, A review of video-based pig behavior recognition Applied Animal Behaviour Science, *Applied Animal Behaviour Science*, 105146 (2020).

[5] B. Mahaur and K.K. Mishra, Small-object detection based on YOLOv5 in autonomous driving systems, *Pattern Recognition Letters*, **168** 115-122 (2023).

[6] S. Ren, R. Girshick and J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, *IEEE Trans. Pattern Anal. Mach. Intell*, **39**, 1137-1149 (2017).

[7] X. Zhang, S. Ren and J. Sun, Deep Residual Learning for Image Recognition *IEEE Conference on Computer Vision and Pattern Recognition*, 770-778 (2016).

[8] B.Li and D. Lima, Facial expression recognition via ResNet-50, *International Journal of Cognitive Computing in Engineering*, **2**, 57-64 (2021).

[9] S. Qian, C. Ning and Y. Hu, MobileNetV3 for Image Classification, *IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Nanchang, China*, 490-497 (2021).

[10] R. Padilla, W. Passos, B. Dias, S. Netto and E.B. da Silva, A Comparative Analysis of Object Detection Metrics with a Companion Open-Source Toolkit, *IEEE Trans. Pattern Anal. Mach. Intell*, **10(3)**, 1137-1149 (2021).

[11] G. Li, J. Jiao, G. Shi, H. Ma, L. Gu and L. Tao, Fast Recognition of Pig Faces Based on Improved Yolov3, *Journal of Physics: Conference Series*, **2171**, (2022).

[12] M.F. Hansen, M.L. Smith, L.N. Smith, M.G. Salter, E.M. Baxter and M. Farish, Towards on-farm pig face recognition using convolutional neural networks, *Computers in Industry*, **98**, 145–152 (2018).

[13] M. Marsot, X. Shan, L.Ye, P. Feng, X. Yan, C. Li and Y. Zhao, An adaptive pig face recognition approach using Convolutional Neural Networks, *J. Computers and Electronics in Agriculture*, **173**, (2020).

[14] K. Wang, C. Chen and Y. He, Research on pig face recognition model based on keras convolutional neural network, *J. IOP Conference Series: Earth and Environmental Science*, **474**, (2020).

[15] R. Wang, Z. Shi, Q. Li, R. Gao, C. Zhao and L. Feng, Pig face recognition model based on a cascaded network, *Applied Engineering in Agriculture*, **37(5)**, 879-890 (2021).

[16] W. Shigang, W. Jian, C. Meimei and W. Jinyang, A pig face recognition method for distinguishing features, *IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)* (2021).

[17] Z. Wang and T. Liu, Two-stage method based on triplet margin loss for pig face recognition, *J. Computers and Electronics in Agriculture*, **194**, (2022).

[18] A. Alameer I. Kyriazakis and J. Bacardit, Automated recognition of postures and drinking behaviour for the detection of compromised health in pigs, *J. Scientific Reports*, **10:13665**, (2020).

[19] J. Redmo and A. Farhadi, YOLO9000: Better, Faster, Stronger, *IEEE Conference on Computer Vision and Pattern Recognition*, 7263–7271 (2017).

[20] J. Sa, Y. Choi, H. Lee, Y. Chung, D. Park and J. Cho, Fast Pig Detection with a Top-View Camera under Various Illumination Conditions, *J. Symmetry*, **11**, 266 (2019).

[21] M. Ju, Y. Choi and J. Seo, A Kinect-Based Segmentation of Touching-Pigs for Real-Time Monitoring, *Sensors*, **18**, 1746 (2018).

[22] M. Riekert, A. Klein, F. Adrion, C. Hoffmann and E. Gallmann, Automatically detecting pig position and posture by 2D camera imaging and deep learning, *J. Computers and Electronics in Agriculture*, **174**, (2020).

[23] E.T. Psota, M. Mittek, L.C. Perez, T. Schmidt and B. Mote, Multi-Pig Part Detection and Association with a Fully-Convolutional Network, *J. Sensors*, **19(4)**, 852 (2019).

[24] J. Xu, S. Zhou, A. Xu, J. Ye and A. Zhao, Automatic scoring of postures in grouped pigs using depth image and CNN-SVM, *Computers and Electronics in Agriculture*, **194**, 106746 (2022).

[25] V. Badrinarayanan, A. Kendall and R. Cipolla, SegNet: A Deep Convolutional Encoder-Decoder Architecture for Scene Segmentation, *IEEE Trans. Pattern Anal. Mach. Intell*, **39**, 2481-2495 (2017).

[26] K. He, G. Gkioxari, P. Dollar and R. Girshick, Mask R-CNN, *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2961-2969 (2017).

[27] Z. Wang, W. Chen, S. Gu, Y. Wang and J. Wang, Evaluation of trunk borer infestation duration using MOS E-nose combined with different feature extraction methods and GS-SVM, *Computers and Electronics in Agriculture*, **170**, 105293 (2020).

[28] M.R. Tan and Q.V. Le, EfficientDet: Scalable and Efficient Object Detection, *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit*, 10778–10787 (2020).

[29] J. Low, BIoU: An Improved Bounding Box Regression for Object Detection, *Power Electron. Appl.*, **12(4)** 51 (2022).

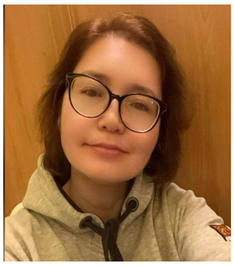
[30] Y. Gong, X. Yu, Y. Ding, X. Peng, J. Zhao and Z. Han, Effective Fusion Factor in FPN for Tiny Object Detection, *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 1160-1168 (2022).

[31] S. Kim and J. Huh, Consistency of Medical Data Using Intelligent Neuron Faster R-CNN Algorithm for Smart Health Care Application, *Healthcare*, **8(2)** 185 (2020).

[32] A. Salvador, X. Giro-i-Nieto, F. Marques and S. Satoh, Faster R-CNN Features for Instance Search, *Proceedings*

of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 9-16 (2016).
[33] J. Hou, C. Yang, Y. He and B. Hou, Detecting diseases in apple tree leaves using FPN–ISResNet–Faster RCNN, *European Journal of Remote Sensing*, **56:1**, (2023).
[34] I. Patel and S. Patel, An Optimized Deep Learning Model For Flower Classification Using NAS-FPN And Faster RCNN, *International journal of scientific and technology research volume*, **9(3)**, (2020).
[35] K. Tong and Y. Wu, Rethinking PASCAL-VOC and MS-COCO dataset for small object detection, *Journal of Visual Communication and Image Representation*, **93**, 103830 (2023).
[36] M. Everingham, L.V. Gool, C.K. Williams, J. Winn and A. Zisserman, The Pascal Visual Object Classes (VOC) Challenge, *Int. J. Comput.*, **88** 303–338 (2010).
[37] L. Atymtayeva, M. Kanatov, A.M. Musleh and G. Tulemissova, Facial Expression Recognition System: Selection of Models, *Applied Mathematics and Information Sciences*, **17(2)**, 375-383 (2023).
[38] M. Kanatov, L. Atymtayeva, M. Mendes, Improved facial expression recognition with xception deep net and preprocessed images, *Applied Mathematics and Information Sciences*, **13(5)**, 859-865 (2019).

Olga Ivashchuk, Doctor of Technical Sciences, Professor, Head of Physics and Technology Department, Head of the Department of Information and Robotics Systems, Belgorod State National Research University, Russia. Author of more than 200 scientific and educational-methodical works, including those published in international scientific journals. Chairman of Dissertation Council. Leader of large scientific projects. She is an expert of Russian academic and scientific foundations, a member of editorial board of journals. Her research interests include the construction of complex automated monitoring and control systems, artificial intelligence and machine learning, as well as mathematical and situational modeling to solve interdisciplinary problems.



Lyazzat Atymtayeva received the Ph.D and Doctor of Science degree in Mechanics, Mathematics and Computer Science at Al-Farabi National University, Kazakhstan. Now she is working as associate professor in Information Systems at SDU University. Her research interests are in the areas of mechanics, applied mathematics and computer science including the numerical and rigorous mathematical methods and models for mechanical engineering and computer science, intelligent and expert systems in Information Security, Artificial Intelligence and Machine Learning, Project Management and Human-Computer Interaction. She has published research papers in reputed international journals of mathematical and computer sciences. She is a reviewer and an editor of international journals in mathematics and information sciences.

Alexey Zhigalov, having more than 5 years of experience in the field of machine learning and computer vision, both in industry and in academic circles, successfully implemented many projects from idea to product as an engineer and manager. He leads a team of engineers and researchers who work on cutting-edge technologies such as natural language processing, computer vision and deep learning. He published a number of papers in journals and conferences and received several patents in the field of computer vision.

Bagdat Yagaliyeva received the Ph.D degree in Engineer Robotics at Satbayev Technical National University, Mathematics and Physics at Al-Farabi National University, Kazakhstan. Now she is working as associate professor in department of ”Cybersecurity, information processing and storage” at Institute of Automation and Information Technologies. Her research interests are in the areas of mechanics, applied mathematics and computer science including the numerical and rigorous mathematical methods and models for mechanical engineering and computer science, intelligent and expert systems in Information Security, Artificial Intelligence and Machine Learning. She has 60 publications, 13 of them in highly rated peer-reviewed publications.

Oleg Ivashchuk received the candidate of physical and mathematical sciences degree at Belgorod State National Research University, Russian. Now he is working as associate professor in department Computer Science at Yessenov University. His research interests - data science, artificial intelligence, machine vision, miniature pyroelectric and piezoelectric X-Ray sources. He has over 50 publications, of which 20 are in highly rated peer-reviewed publications. Hirsch index in scientometric base Scopus - 5.

Vyacheslav Fedorov - Candidate of Technical Sciences, Associate Professor at the Department of Information and Robotics Systems, Belgorod State National Research University, Russia. He teaches students courses on programming and designing information systems. He is an academic secretary of the dissertation council, an author of scientific articles. He actively participates in scientific projects, develops original algorithms and software products for evaluating and predicting the results of interaction between various components of the biosphere and the technosphere.