Comparative research of bounding box based and image segmentation based neural network models for individual cattle identification

Denis Shilin*, *Ivan* Shubin, and *Dmitry* Shestov

Scientific and production association industrial capital Ltd, Tashkentskaya str., b. 17, Moscow, Russian Federation

Abstract. Software and hardware based tools for collecting information about animals are an important part of most large agricultural enterprises. Various information collection methods and technologies help farmers to receive data about the health, behavior and location of animals therefore becoming an important tool for monitoring the area of animal captivity or even open spaces. This work is devoted to a comparative study of the effectiveness of methods for identifying cattle based on neural network technologies and machine learning. In this paper, the effectiveness of the bounding box-based method and the image segmentation-based method were investigated. Images of cattle, which can be obtained using computer vision methods, are used as input images for identifiers. The authors are aimed to determine a representative set of features which can be used for reliable and accurate cow identification using neural network based methods, estimate efficiency of those methods using precision, recall and F1-score quality metrics and suggest certain improvements that may improve the quality of identification for neural network methods, if possible. Neural network methods, the effectiveness of which was verified by quality metrics, were also tested on a set of images that did not participate in the formation of the training sample set.

1 Introduction

There are many methods for collecting information about the current position, activity and health condition of cattle on farms and in other places of captivity. Most of these methods rely on a certain type of device that is attached directly to the animal's body: pedometers, thermometers, accelerometers, etc. However, attaching third-party devices to an animal's body can negatively affect their behavior, and health conditions [1]. Eliminating the negative aspects of this approach has become one of the main reasons for creating methods that provide an opportunity for collecting information about the cattle based on image analysis using neural network technologies and machine learning methods.

Convolutional neural network methods were chosen as an animal identification mechanism, since previously these methods were successfully used to identify cattle of

* Corresponding author: ya.shubin-ivan@yandex.ru

different breeds [2], and also demonstrated high accuracy of cattle recognition using cameras mounted on drones [7]. Consequently, it has been suggested that these methods can show high accuracy in identifying cattle, even if they belong to the same breed, as well as if there are features in the transmission of images.

In some parts of this article, the process of determining which particular cow is represented in the image is called "identification", and in some parts the same process is called "classification", since the identification process is a special case of the classification process in which the number of classes is equal to the number of unique objects in the sample set.

1.1 Body pattern based recognition

The method of identifying cows by hair color is based on the use of characteristics of color patterns on the body of animals for their identification and classification. Since most of the animal's body has a hairline, the necessary information for classification is easily perceived by most machine vision cameras with sufficient illumination. This approach has been used in classification before. [3] Since the position of hair fragments with pigmentation other than the predominant color in the animal's color varies from one individual to another, this approach demonstrates classification accuracy at the level of 0.90 – 0.95. However, the effectiveness of this approach decreases when animals with a monophonic hair colour are included in the sample. This is due to the reduced number of unique features that can be used for classification. Therefore, the classification method based on the colour of the animal's hair is not universal, and, therefore, will not be considered in this work as the main classification model.

1.2 Facial feature based recognition

This classification method uses the external characteristics of the animal's head. Such as: shape of the horns, distance between the eyes and nostrils, shape of the nose, color of the hair and scalp, as well as other present features: the presence of pigmentation features: birthmarks, albinism, etc. Researchers mainly use a neural network approach for solving the classification or identification problem based on this type of recognition [4], since it allows to automatically adjust the classifier to several features at once, requiring the developer only to specify the region of interest using annotation tools. Since the classification method based on the external features of an animal's head can include several features at the same time, it can be used to classify or identify an animal with a monophonic hair color, unlike the method described in paragraph 1.1.

1.3 Nose shape based recognition

This method is based on the external features of the animal's nose: the shape, the distance between the nostrils and the pattern of the epidermis on the animal's nose. The latter feature is the most promising, since researchers note [5] the similarity of the epidermis pattern on the cow's nose and human fingers. Given that human fingerprints have been used for identification in criminology and information security for a long time [6], this feature can be used as the main one to solve the classification problem.

To implement a classification method based on the pattern of the epidermis of the animal's nose, it is necessary to obtain data about this pattern from each animal, then create a database for classification using information about these patterns, finally it will be necessary to detect and segment the area containing the animal's nose during the classification procedure so that this area can be submitted click on the classifier and get the classification result. Since this approach requires a multi-stage approach to processing the input image, as well as the development of additional software tools, for the purposes of this research, we will limit ourselves to segmentation of the animal's nose area and subsequent identification of the animal using the shape of it.

2 Description of the model quality criteria

In order to determine the quality of classification, identification or recognition tasks in machine learning engineering are using quality metrics which are based on dividing all objects into 4 categories listed below:

1. TP (True Positive) – the method assigned the object to the class to which it actually belongs;
2. FP (False Positive) – the method assigned the object to a class to which it does not actually belong;
3. FN (False Negative) – the method did not assign the object to the class to which it actually belongs;
4. TN (True Negative) – the method did not assign the object to a class to which it does not actually belong.

Precision parameter is the ratio of true positive to and all the objects that the neural network algorithm classified as positive.

$$Precision = TP / (TP + FP) \tag{1}$$

Recall is the ratio of true positive and the sum of all positive objects.

$$Recall = TP / (TP + FN) \tag{2}$$

F1-score is the average harmonic value of precision and recall.

$$F1 = (2*Precision*Recall) / (Precision + Recall) \tag{3}$$

Comparing the values of these metrics is used to determine the quality of classification and choose the model that most effectively copes with the task and makes fewer mistakes.

3 Methods and experiment results

The YOLOv8 algorithm is based on the application of convolution and pooling operations which are used to determine regions of interest in an image. The model consists of two parts called ‘backbone’ and ‘head'. ‘Backbone’ is a modified version of the CSPDarknet53 architecture consisting of 53 convolutional layers and provides interconnection to ensure optimal information exchange between layers. The ‘Head’ consists of several convolutional layers, followed by fully connected layers, which are responsible for constructing bounding boxes to indicate regions of interest in images and indicating the probability of an object belonging to one of the deterministic classes. The model is designed to solve detection and classification problems. There is also a modified model for solving the segmentation problem: YOLOv8-SEG.

3.1 Sample set description

The training sample for YOLOv8 model was formed from 1339 images of 55 individual cows. For the case of this model, a monochrome filter was applied to the images to reduce the effect of animal's facial hair pigmentation on the identificator and therefore increase the influence of other features. Then, using the annotation tool "YOLO- Annotation-Tool", 1339 regions of interest were defined on the images using bounding boxes. In this case, the "class" of the object is the number of every animal in the training sample set. The number of classes is equal to the number of animals in the training sample set. Then an augmentation procedure was performed in order to increase the stability of the classifier to interference. To do this, up to 5% of the pixels in the image were replaced with white pixels.

After the annotation and augmentation procedure, the sample set was divided into training, validation and test sets. The training sample set consists of 937 images, the validation sample set consists of 268, and the test sample set consists of 134. Each of the samples contains objects of every single class listed above.

The training sample for YOLOv8-SEG model was formed from 1761 images that contained pictures of the lower part of the head of 32 individual cows. Then, using the "Super Annotate" annotation tool, 1761 regions of interest were defined on these images, indicating its clear boundaries, the region of interest was described by a closed polyline, which defines the perimeter of the object of interest: cow’s nose. After that, each region of interest on every image was assigned to its own class. In this case, the "class" of the object is the number of every animal in the training sample set. Then an augmentation procedure was performed in order to increase the stability of the segmentation algorithm to interference. To do this, up to 5% of the pixels in the image were replaced with white pixels.

After the annotation and augmentation procedure, the sample was divided into training, validation and test samples set. The training sample set consists of 1761 images, the validation sample set consists of 167, and the test sample set consists of 84. Each of the samples contains objects of all classes involved in the implementation of the learning process.

The following software tools and Internet resources were used to implement the training and testing process for YOLOv8 and YOLOv8-SEG models:

1. Python 3.9 is the programming language most compatible with YOLO models;
2. OpenCV is a library for implementing machine vision;
3. Numpy is a library for processing numerical arrays and matrices;
4. Roboflow is an online resource that provides tools for augmentation and distribution of training samples, as well as computing power for training neural network models;
5. Visual Studio Code is a software development environment.

Table 1. YOLOv8 configuration parameters.

Parameter Names	Value
Number of classes	54
Class Names	1, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 2, 20 21, 22, 23, 24, 25, 26, 27, 28, 29, 3, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 4, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 5, 50, 51, 52, 53, 54, 6, 7, 8, 9.
Resolution of input image, pixels	640 * 640
Training limit	310

Table 2. YOLOv8-SEG configuration parameters.

Parameter Name	Value
Number of classes	32
Class Names	10_cow, 11_cow, 12_cow, 13_cow, 14_cow, 15_cow, 16_cow, 17_cow, 18_cow, 19_cow, 1_cow, 20_cow, 21_cow, 22_cow, 23_cow, 24_cow, 25_cow, 26_cow, 27_cow, 28_cow, 29_cow, 2_cow, 30_cow, 31_cow, 32_cow, 3_cow, 4_cow, 5_cow, 6_cow, 7_cow, 8_cow, 9_cow.

Resolution of input image, pixels	640 * 640
Training limit	310

3.2 Experiment results for YOLOv8

After the experiment, a classifier based on the YOLOv8 neural network model was obtained. Graphs of the dependence of classification quality metrics on the training epoch are presented below:

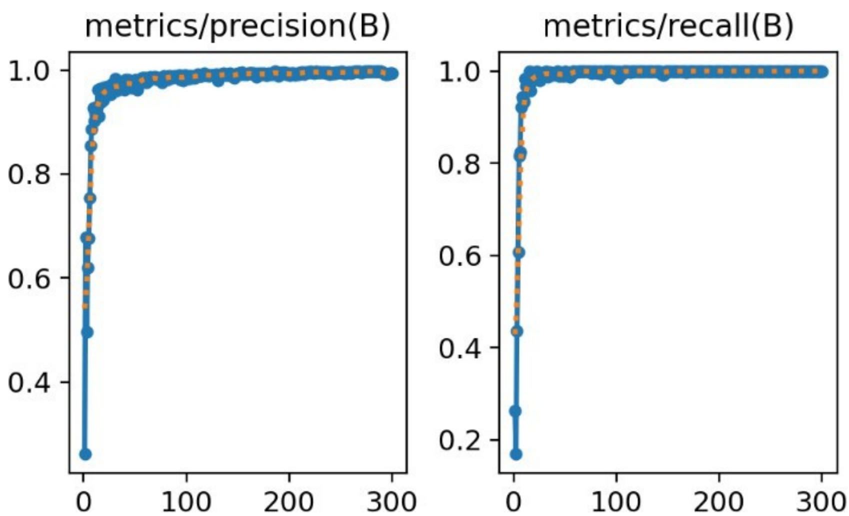

Fig. 1. Graph of the dependence of precision and recall metrics on the training epoch for YOLOv8 model.

The values obtained are used to calculate F1-score:

$$F1 = (2*1*0.994) / (1 + 0.994) \quad (4)$$

Examples of performance of the resulting model on a set of data that wasn't used in a training procedure are shown in the figure below:

Fig. 2. The example of YOLOv8 prefomance for images that were not used in training procedure.

Table 3. YOLOv8 resulting metrics.

Metric	**Value**
Precision	1
Recall	0.994
F1-score	0.997

3.3 Experiment results for YOLOv8-SEG

After the experiment, a classifier based on the YOLOv8- SEG neural network model was obtained. Graphs of the dependence of classification quality metrics on the training epoch are presented below:

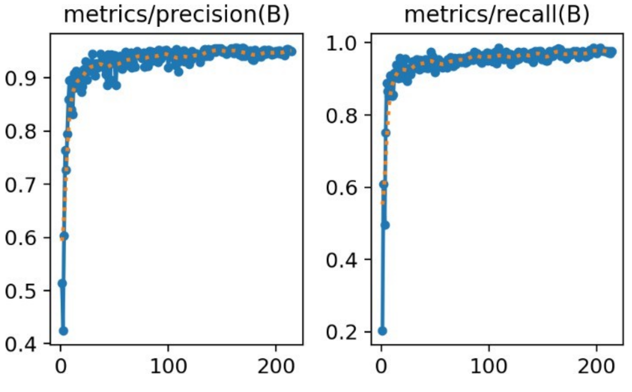

Fig. 3. Graph of the dependence of precision and recall metrics on the training epoch for YOLOv8-SEG model.

The values obtained are used to calculate F1-score:

$$F1 = (2*0.948*0.971) / (0.948+0.971) \quad (5)$$

Examples of performance of the resulting model on a set of data that wasn't used in a training procedure are shown in the figure below:

Fig. 4. The example of YOLOv8 performance for images that were not used in training procedure.

Table 4. YOLOv8-SEG resulting metrics.

Metric	**Value**
Precision	0.948
Recall	0.971
F1-score	0.959

4 Conclusion

As a results 2 operating neural network based identificators were obtained. F1-score resulting metric for both neural network models exceeds 0.95, therefore both identificators can be described as effective models that have a potential to be used for scientific and industrial application. The quality metrics for the identificator based on the YOLOv8 model turned out to be higher than the set of metrics for the classifier based on the YOLOv8-SEG model. Since the structure of these two models differs slightly, it can be concluded that identification based on the external features of the animal's head is more effective than identification based on the external features of the animal's nose using segmentation. However, the YOLOv8-SEG based method can be improved by combining it with identification based on cow nose pattern.

Thus, the identificator based on YOLOv8 model is more effective than identificator based on YOLOv8-SEG model.

The authors of the article express their gratitude to Peerayut Nilchuen from Kasetsart University Bangkok, Thailand for providing materials which became the base for formation of a training sample set.

References

1. F. Tuyttens, C. Molento, S. Benaissa, FVETS **9** 889623 (2022)
2. A. Rivas, A. G. Briones, J. M. C. Rodriguez, P. Chamoso, Sensors **18(7)** 18072048 (2018)
3. R. W. Bello, A. S. A. Mohamed, A.Z. Talib, D. A. Olubummo, IJACSA **11(3)**, 92-98 (2020)
4. Z. Weng, F. Meng, S. Liu, Y. Zhang, J.Compag **196(9)** 106871 (2022)
5. R. W. Bello, A. S. A. Mohamed, A.Z. Talib, GUJS **33(3)**, 831-844 (2020)
6. D. V. Isyutin-Fedotkov, *Fundamentals of the criminalistics study of personality* (Prospekt, 2018)
7. S. Manoj, V. Kanchana, S. Rakshith, *Identification of Cattle Breed using the Convolutional Neural Network*, in Proceedings of the 3rd International Conference on Signal Processing and Communication, ICPSC, 13-14 May 2021, Coimbatore, India (2021)