Individual Cow Identification Using Non-Fixed Point-of-View Images and Deep Learning

Yordan Kalmukov[1], Boris Evstatiev[2], Seher Kadirova[3]
Department of Computer Systems and Technologies, University of Ruse "Angel Kanchev", Ruse, Bulgaria[1]
Department of Automatics and Electronics, University of Ruse "Angel Kanchev", Ruse, Bulgaria[2, 3]

***Abstract*—Monitoring and traceability are crucial for ensuring efficient and financially beneficial cattle breeding in contemporary animal husbandry. While most farmers rely mainly on ear tags, the development of computer vision and machine learning methods opened many new noninvasive opportunities for the identification, localization, and behavior recognition of cows. In this paper, a series of experimental analyses are presented aimed at investigating the possibility of identification of cows using non-fixed point-of-view images and deep learning. 14 objects were chosen and a photo session was made for each one, which provides training/validation images with different viewing angles of the animals. Next, a darknet-53-based convolutional neural network (CNN) was trained using YOLOv3, capable of identifying the investigated objects. The optimal model achieved 92.2% accuracy when photos of single or grouped non-overlapping animals were used. On the other hand, the trained CNN showed poor performance with group images, containing overlapping cows. The obtained results showed that cows could be reliably recognized using non-fixed point-of-view images, which is the main novelty of this study; however, certain limitations exist in the usage scenarios.**

Keywords—Cow identification; convolutional neural network; YOLOv3; non-fixed point-of-view

I. INTRODUCTION

The advancement in information and communication technologies and artificial intelligence created numerous opportunities in all spheres of human society. They became the backbone of precision agriculture, allowing the optimization of all processes in the agricultural sector including animal husbandry. Nowadays, the food industry worldwide is oriented towards animal products, including meat, milk and milk products, eggs, etc.

An important aspect of livestock farming is the identification of animals, which offers many management and production benefits. Previously livestock owners were marking and branding their cattle with the main concern of proving their ownership. Many additional benefits are offered by contemporary animal identification, such as traceability of the herd and origin, tracking the cattle performance, disease management, keeping accurate track of the animal count, age, status, etc., and as a result, supporting the decision-making process.

The main approach in cattle identification is the application of ear tags. Each country has its regulatory framework to handle this problem, but in many cases, cows must be tagged after their birth. Furthermore, a cattle passport is created for them, including the date of birth, ear tag number, breed, sex, etc. [1]. Recently Radio Frequency Identification (RFID) and Electronic Identification (EID) tags are also popular, as they allow faster and automated data collection about the animals. Other approaches in animal identification are based on nose rings, collars, and image recognition [2,3,4,5]. Furthermore, many studies have suggested using multifunctional Internet of Things devices, which not only identify the animals but also allow monitoring and analyzing their behavior [2,6,7], tracking their location [8], etc.

When it comes to image processing for monitoring and recognition of cattle, the available studies investigate a wide range of opportunities, such as animal counting [9,10], individual cow recognition [11], behavior recognition and monitoring [12,13], animal tracking [14], cattle body detection [15], etc. Furthermore, in animal identification, so far only two image spectra have proven themselves appropriate for the above activities - visible (RGB) [6] and infrared [16].

The application of computer vision for the identification of species and animals has been investigated in numerous previous studies. In [16] the problems and opportunities when recognizing deer and wild boars based on infrared images and the You Only Look Once (YOLO) v3 neural network were studied. Similarly, in [17] RGB and infrared images obtained from camera traps, and a convolutional neural network (CNN), were used to identify frogs, lizards, and snakes. The recognition achieved an F1 score between 82% and 96% for the different species. This approach was also used in [18] with other wild animals, such as Amur tigers, Amur leopards, wild boards, several deer breeds, Asian black bears, red foxes, and other animals. The trained YOLOv5 neural network achieved precision and recall equal to 0.987 and 0.975, respectively.

Many other studies tried to identify not only the species but also to identify the specific animal. For example, in [19] RGB images and YOLOv7 were used for recognizing horses. The obtained mean average precisions (mAP) at the 50% threshold were 99.5% for identifying an animal when using its face and 99.7% when using its nose. Similarly, in [20] sheep faces were recognized based on RGB images. They were preprocessed using noise removal, brightness/saturation/contrast adjustment and when necessary horizontal flipping, and were thereafter fed to different neural networks. The highest accuracy and F1 score were achieved with RepB-Sheepnet, reaching more than 99%. Face recognition for identifying cows with a dual 3D camera setup and the Iterative Closest Point method was used in [21]. The authors have chosen to use the face of the cows, because of the rigid structure of the skull. The reported identification rates varied between 88% and 99%, depending on the gallery point clouds per cow. In another study, the Siamese DB Capsule neural network was applied for the face recognition of cattle [22]. The used approach pairs the images, i.e. each image is compared with another one, and “pos” or “neg” is returned, depending on whether the images belong to different or the same categories. The study achieved recognition accuracy of up to 93% and an F1 score of up to 93.54%. Several other studies have also tried to identify cattle by their noses. In [23] cattle were recognized using their noses using RGB images as row data. The proposed methodology includes supplying grayscale images to a Deep belief network DBN for the recognition process, which was implemented using the Matlab R2019b toolkit. The authors reported that the recognition accuracy reached up to 99% with 400 training images, and it decreases when fewer images are used. Nose-based cattle identification was also used in [24], where the muzzle patterns of each animal were obtained with the help of numerous deep-learning algorithms. The highest achieved accuracy was 98.7% for a VGG16_BN-based model.

The recognition of the cows can also be classified depending on where the cameras are positioned and how many cameras are used. Common approaches are using topview, sideview, and backview cameras. In [25] the cows were photographed from above, with the idea of recognizing their skin pattern. The methodology includes background removal, image rotation, alignment according to the template, and pixelized binary image creation. The nearest neighbor approach and the Hammering distance measure were used for the recognition process. The authors reported a Top-1 accuracy of 61.5% and a Top-4 accuracy of 83%. Furthermore, they stated that such an approach does not require retraining the recognition algorithm (compared to deep learning) and is very fast. A similar approach was used in [26], where photographs of the cows’ bodies and a neural network were used to recognize their body patterns. The reported accuracy was more than 92% for the training data and 90% for the testing data.

In [27] 7069 topview images of 62 cows were used for training, 1801 images for validation, and 1104 images for testing different CNNs. The ResSTN model achieved the highest average recognition accuracy of 94.58% and slightly higher accuracy under daylight lighting conditions. In [28] daytime and nighttime topview images were used to track the cattle movement over a farm. The idea was to use many strategically placed cameras not only to identify the animal but also to classify its behavior into the categories “resting”, “standing”, “standing up” and “walking”. Different versions of YOLOv5 were used to identify the animals, which achieved mAP ranging from 92.7% to 95.3%.

Other studies used sideview cameras for cattle identification. In [29] RGB images of 13 cows were used for their recognition using a convolutional neural network with ResNet50 as the backbone. The authors reported more than 98% recognition accuracy of the investigated objects. Similarly, in [30], side RGB images of cows were used for their recognition. Different color spaces were used to obtain the most appropriate one for distinguishing the animals. The Euclidean distance of feature vectors of critical points was used as the criteria for identification, together with the Brute Force Matcher algorithm. The optimal accuracy of 99.31% was obtained for the Lαβ color component. Another approach for the identification of animals was presented in [31] that relies on rump RGB images. 2140 images of 195 cows were used for training a convolutional neural network and 917 for its validation. The CNN based on a Mobilenet v2 backbone returned the highest accuracy, reaching up to 99.76%.

In [32] three cameras were used for cow recognition - topview, frontview and sideview. An enhanced filter algorithm was proposed, combining the mean-shift and particle-Kalman filter algorithms. The image processing was implemented using Matlab, though no accuracy has been reported. Several cameras were also used in [33], where top and sideview RGB and depth images from a 15-fps video were applied for cow identification. The study uses Euclidean cluster extraction to select the largest 3D point cluster representing the cow, and then to estimate the average silhouette of the animal. Thereafter, the differences between the obtained silhouette and the probed one are evaluated. The achieved algorithm accuracy reached 75.6%.

A completely different approach was used in [34]. The YOLOv3 neural network was utilized to “read” the ear tag of cows, while they were near the drinker. The idea was to estimate how long the animals are drinking and as a result to estimate the approximate volume of water drunk. A mean average precision of 89% and an F1 score of 86% were obtained.

Other studies had more advanced goals, such as behavior identification. In [30] the calving time of cows was predicted via motion classification using a 360° overhead RGB camera. The methodology includes object identification, background subtraction, generation of the object contour, and principal component analysis (PCA) to extract features. An average accuracy of up to 95% was reported for detecting and classifying cow motions.

The analysis of previous animal recognition studies showed that deep learning provides acceptable results. Different barriers exist when using deep learning for cow identification, such as the limited number of images of each object, the fluctuations of the positions and angles of view, the appearance of numerous cows the network wasn’t trained for, etc. [22]. Furthermore, often it is hard for a human being to tell the difference between two separate cows, especially if they are single-colored, which might be a problem when preparing the training/validation data. Most previous studies have used images of the animals, taken from a specific point of view (side, back, frontal, etc.) or from above that are rotated to an appropriate orientation. However, there are almost no studies, dealing with the identification of cows with non-fixed point-of-views. The abovementioned shows the existence of a knowledge gap in this area, which should be addressed.

This study aims to investigate the possibility of real-time identification of cows using deep learning and images, which were obtained without limitations for the angle of view and orientation of the animals. To achieve this the neural network should be trained with numerous images, representing the animals from different points of view.

II. Materials and Methods

A. The Study Area

The experimental part of this study was conducted at a farm for outdoor cow breeding. It is located in the village of Trastenik, Ruse District, north-central region of Bulgaria (Fig. 1) and has coordinates 43.65830753698392, 25.845059022235304. 30 dairy cows from the Bulgarian Black and White cow breed and the Red and White Holstein cow breed are bred on the farm. All images and videos of cows were shot on a summer day (5 July 2023), late in the afternoon between 5 and 7 p.m.

Fig. 1. Location of the experimental pasture.

B. Methodology for Data Collection and Data Processing

The data collection and processing methodology, applied in this study is summarized in Fig. 2. It can be divided into five main steps, which are explained below.

1) Step 1. Data collection: This step begins with choosing the cows that are the object of the investigation; thereafter, numerous images of each object are made. The successful recognition of the object requires enough images representing the animal from different sides and in different circumstances (staying, grazing, etc.). This is achieved in two ways:

- By making numerous photos of the animal;
- By filming a video of the animal from all sides and extracting appropriate frames from it.

The described procedure is repeated for each cow that should be investigated. Taking photos from different distances and with different backgrounds should also be considered.

2) Step 2. Data selection: In this phase, the already collected data is analyzed, sorted, and filtered. Initially, all photos taken that contain a certain object are sorted in separate folders (per object). Furthermore, if a video was made of a certain animal, frames are extracted from it as images, representing the animal with different viewing angles. They are also sorted in the corresponding folders. Finally, several of the prepared training images are filtered out for the validation set so that they are not used during training.

3) Step 3. Data preparation: The goal of this step is to prepare the training and validation data. Each object is marked in a rectangle and classified using LabelImg or an alternative tool. This is repeated for each of the images and each of the objects. If a certain image contains more than one of the investigated objects, all of them could be marked and categorized in a different class, as shown in Fig. 3. In this step both the training and the validation images are classified.

4) Step 4. Deep learning: In the next data processing phase a machine learning model is trained. In this study, we use the YOLO v.3 object recognition system, and therefore all data should be prepared and sorted accordingly [35]. We chose version 3 because it has shown good results in previous studies. The preparation for this step includes:

- tuning up a config file corresponding to the graphical processing unit (GPU) characteristics, the number of classes, and the number of training iterations;
- setting up the training itself, i.e. selecting the training and testing data, the config file, the initial weights as well as the application of mean average precision (mAP).

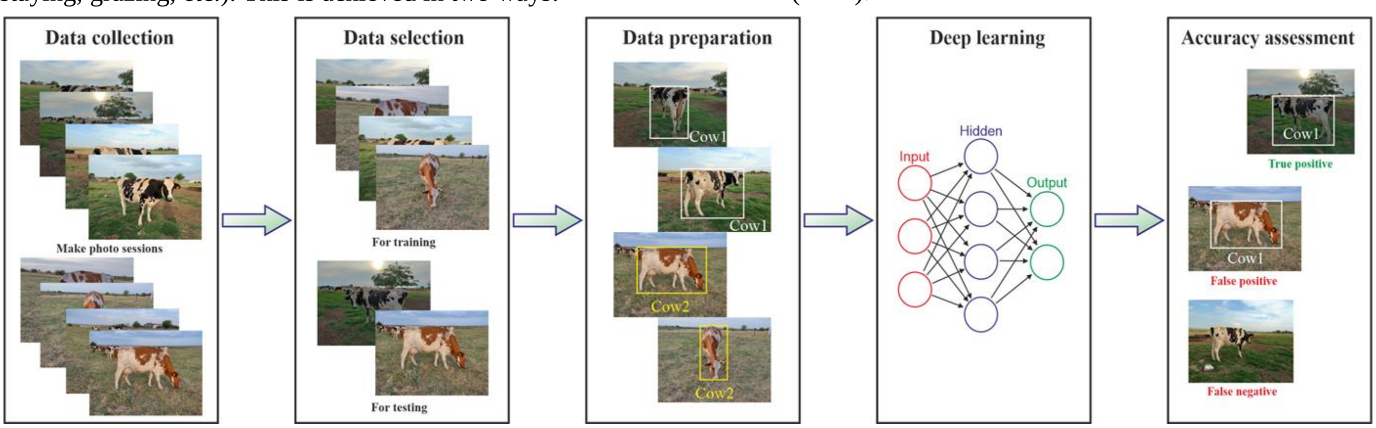


Fig. 2. Overview of the used methodology.

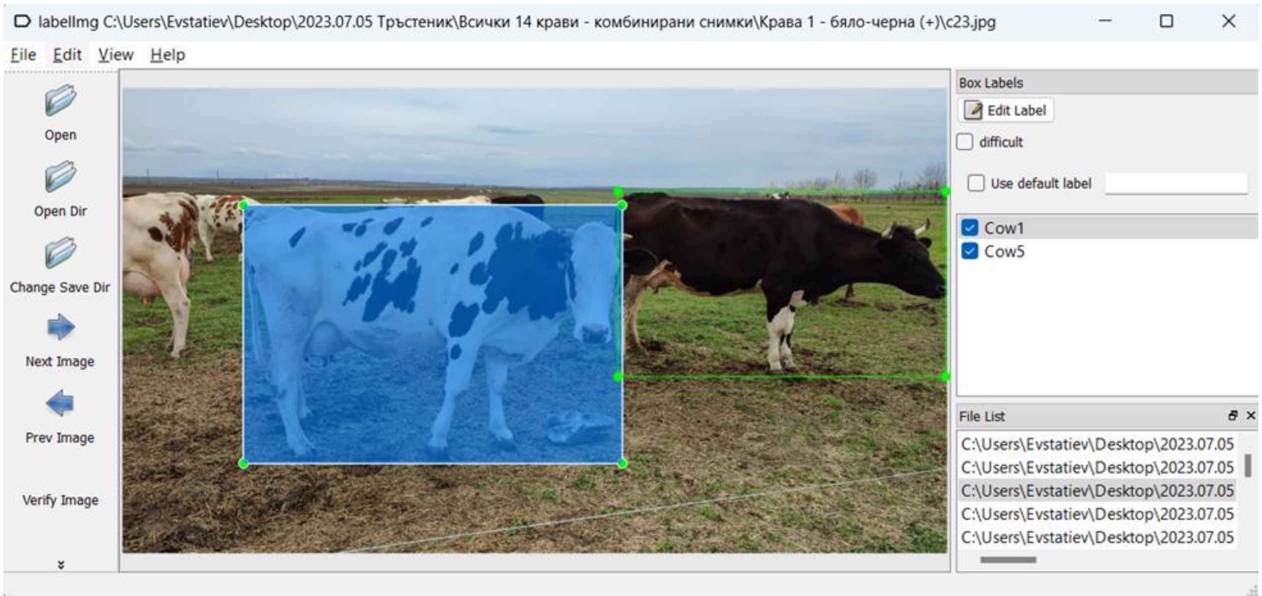


Fig. 3. Marking each object with a rectangle in the LabelImg tool.

5) Step 5. Accuracy assessment: The final step of the data processing is to estimate and evaluate the models' accuracies. In this study this is implemented from several perspectives:

- The first one is the automatic accuracy assessment during the training of the models, which is implemented by YOLO itself and is based on the mean average precision performance metric. It is achieved using validation images prepared in steps 2 and 3. The results from this assessment are taken into account when selecting the optimal classification model, i.e. the one with the highest mAP. The meaning of mAP metric is as follows:

$$mAP = \frac{1}{N}\sum_{n=1}^{N} AP(n) \tag{1}$$

where N is the total number of classes and AP(n) is the average precision for a given class n, which is calculated as the weighted mean of precision at each threshold.

- The second perspective is to use additional images and videos, which were not used for training and validation, and a human operator to confirm that the recognized objects (cows) are correct or incorrect. For images, this evaluation is straightforward. Videos could be processed in the following way:
  - Each video is resampled to 1 frame per second (FPS) framerate.
  - The video is analyzed with YOLO v.3 using the selected optimal model and the analysis results are saved as video files using screen recording software like OBS Studio or alternative.
  - The recorded video is observed by an operator, frame by frame, and a confusion matrix is created for each object.
  - The model performance is assessed in terms of Precision, Recall, and F1 Score, whose meaning is described below. The Precision metric assesses the accuracy of positive predictions and is defined with:

$$P = \frac{TP}{TP+FP} \tag{2}$$

The Recall metric gives the proportion of true positive (TP) predictions among all positives and is defined as:

$$R = \frac{TP}{TP+FN} \tag{3}$$

The F1 score metric balances the score of precision and recall according to:

$$F1 = 2 \times \frac{P \times R}{P+R} \tag{4}$$

Finally, the average accuracy for the whole testing dataset is estimated using the following equation:

$$Accuracy = \frac{Number\ of\ testing\ frames}{Number\ of\ testing\ frames + FP + FN} \tag{5}$$

Next, the obtained results are evaluated and analyzed using the created confusion matrix. Situations with false positives (FP) and false negatives (FN) are closely analyzed, to identify the reasons behind this.

III. Results and Discussion

A. Training of Convolutional Neural Networks

801 images of 14 different dairy cows were used for training. Some of the cows look very similar and are quite difficult to distinguish even for a human being (Fig. 4). For most cows 50-70 photos each were used, though there are also cows photographed over 100 times, while others were shot far less than that (just 10-15 images). All photos were taken under natural environment conditions, without separating the animals from the herd or placing them in an enclosure. An Mx-

M16TB-R079 RGB camera by Mobotix AG (Langmeil, Germany) in continuous video recording mode and three mobile phones were used for filming. The resulting videos and images have different resolutions and color saturation.

It should be noted that most of the cows look different from their two sides, which makes it important to select an appropriate collection of photos, representing the animals from all sides for training an adequate model. That is why for each cow an improvised photo session was made (Fig. 5).

Next, according to the methodology, a folder with images was made for each object. Furthermore, frames from the video recordings were chosen, extracted, and placed in the corresponding folders. On each image, the target objects were selected using the LabelImg tool. Thereafter, the images were divided for training, validation, and testing purposes, as shown in Table I. A total of 44 images were used for validation and 37 images of individual cows for testing. The table data shows that the datasets are imbalanced, which should be considered when interpreting the results.

The training of the CNN was performed using an NVidia RTX 3060 GPU with 12 GB dedicated video RAM, which allowed roughly 1000 training iterations per hour. According to the developers of YOLO, the recommended number of training iterations is the number of classes multiplied by 2000, but not less than the number of images. Therefore, initial training was conducted with a maximum number of 28000 iterations. With such a configuration, YOLO saves the calculated weights at every 10000 iterations, at the maximum number of iterations (28000), and at the maximum mAP value. The latter is calculated during the training process using the validation dataset. The highest obtained mAP is 100% and was achieved at 18500 iterations.

In previous training sessions with cow images, we have observed that the neural network could be easily overtrained and begin to miss (fail to recognize) cows that it was trained for, but which were viewed from a slightly different position than those in the training dataset. Experiments with different numbers of training iterations showed that more cows (especially in group photos) were recognized with weights obtained in fewer iterations. To avoid overtraining, we retrained the network with 9000 iterations, thus the YOLO system saved the weights every 1000 iterations. As a result, we obtained multiple alternative weight files that allow us to perform additional experimental analyses and determine the optimal number of iterations. When training with 9000 iterations, the maximum value for mAP is 98.81% and was reached at 6850 iterations. The training and validation results with 9000 iterations are shown in Fig. 6.

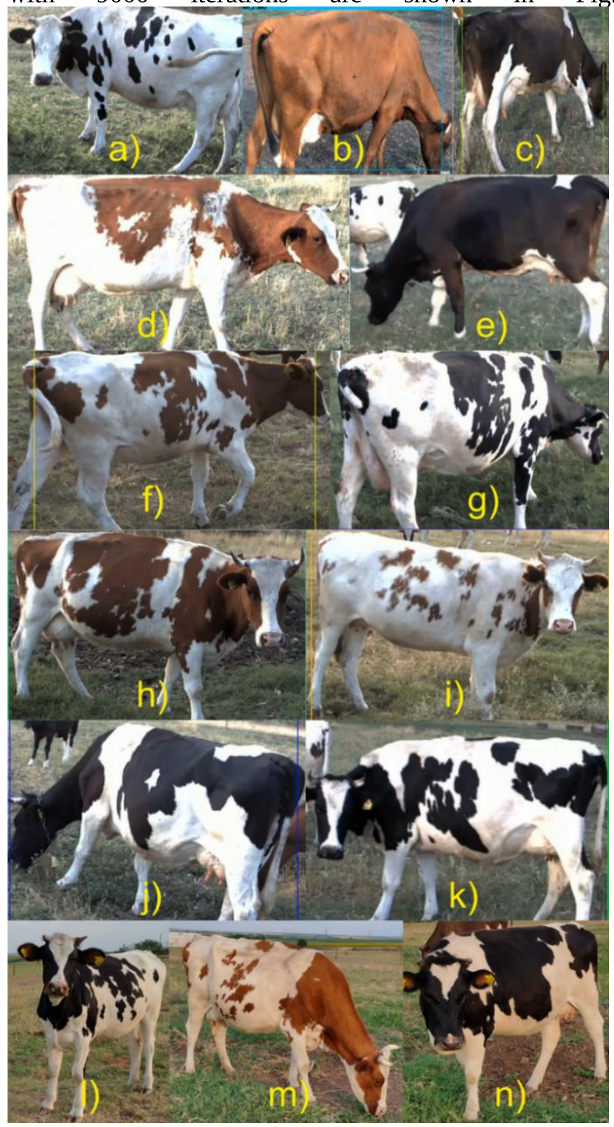


Fig. 4. The investigated cows.

After a series of experimental analyses, we found that for both group photos and videos, more cows are recognized when using the weights file obtained at 6850 iterations, which is why it is used for all subsequent experiments.

Fig. 5. Taking a photo session of cow №8 from different points of view: a) right side; b) back side; c) left side; d) front side.

TABLE I. SUMMARY OF THE NUMBER OF IMAGES OF EACH OBJECT USED FOR TRAINING THE CONVOLUTIONAL NEURAL NETWORK

Cow №	**1**		**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Training images	52		24	52	149	80	35	71	81	58	72	49	11	45	22
Validation images	3		1	2	4	4	2	4	3	5	5	4	1	4	2
Testing images	3		2	3	7	2	1	2	5	3	2	2	1	2	2
Testing video frames	89		117	166	139	187	58	93	91	111	165	232	/	/	/

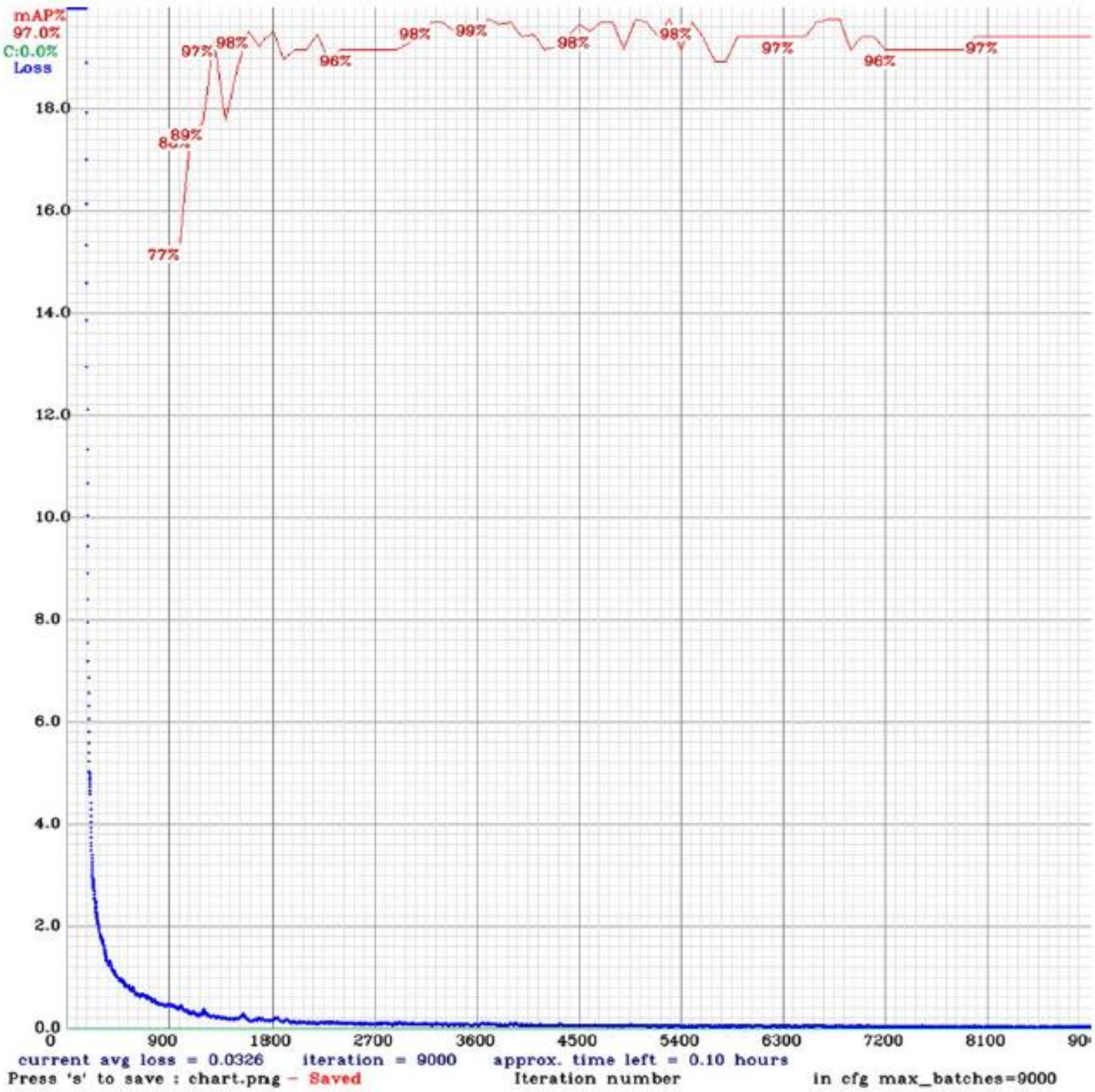


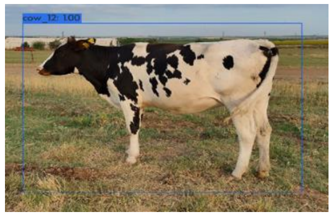
Fig. 6. YOLO training results with 801 images of 14 cows at 9000 iterations, with maximum mAP 98.81% at 6850 iterations.

It is interesting to note that at 18500 iterations mAP has a maximum value of 100% and cows in individual pictures are recognized with a maximum probability of 0.99 or 1. At 6850 iterations, despite the slightly lower mAP value of 98.81% and lower individual probabilities, more animals are recognized in group photos and videos. This suggests that at 18500 iterations, the neural network is already overtrained. Another possibility is that the validation dataset, which is 5.625% of the training dataset, is too small and could be increased in future studies.

B. *Assessment of the Accuracy*

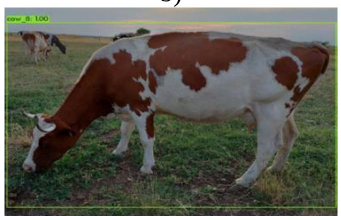
Initially, 37 testing images not applied in the training and validation process were used. Their distribution between the different objects of the study is shown in Table I. In all of them, only one cow is visible or the other cows are positioned behind the recognized object. In this scenario, the achieved recognition rate is 100% and this applies to all 14 investigated objects. Furthermore, the recognition is done with a probability between 98% and 100%. Fig. 7 presents several examples of the correctly recognized cows from the testing dataset.

a)



b)

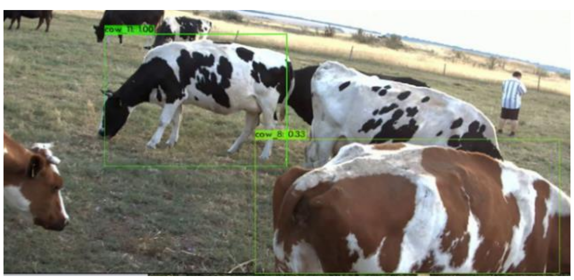
c)



d)

Fig. 7. Examples of recognized cows with high (typically 100%) probability in cases where the images contain only one cow and the animal is seen in sufficient detail.

A significantly different situation occurs when there are numerous cows in the image. In some cases, the model was able to recognize correctly two cows in a single image, as shown in Fig. 8. Nevertheless, it could be noticed that in Fig. 8, a cow_8 is identified with low probability (33%), which might be explained by the fact that it is partially visible. Similarly, in Fig. 8(b) again cow_8 is recognized with a probability of 31%, which might be caused by its overlapping with other similarly colored cows.



a)

b)

Fig. 8. Examples of correctly recognized two cows in a single image: a) one of the objects is partially visible; b) both objects are fully visible and overlapping with other cows.

However, that is not always the case. In many other situations, the trained model experienced difficulties recognizing multiple cows in group photos, although it was trained for many of them. In these cases, usually up to 1 and more rarely two cows in the front are recognized using the weights obtained at 6850 iterations. If the weights file, obtained at 18 500 iterations is used instead, the results are worse. This problem is demonstrated in Fig. 9, where cows numbered five and eight are available in the two photos, which were taken with a time difference of several seconds. Nevertheless, the trained model identifies either one or the other, but not both objects. In this example the probability rate is relatively high – 70% and 88%, respectively. It could be noticed that in Fig. 9(a) there are other cows behind cow_5, which might be influencing its recognition performance. Similarly, in Fig. 9(b) cow_8 is partially overlapping with cow_5, which once again might be the reason for such behavior.

More problems with the identification of cows in group photos are demonstrated in Fig. 10 and Fig. 11. The first one shows that two objects (cow_5 and cow_4) were not identified. The most probable reason is that the model fails to distinguish them as separate objects from the numerous surrounding animals. Cow_5 overlaps with two other cows and cow_4 with one in front and many others in the background. This suggestion is confirmed by the example, demonstrated in Fig. 11, where three cows are identified as a single animal (cow_1), even though only one of them is cow_1.

a)

b)

Fig. 9. An example of recognizing cow 8 (a) and cow 5 (b) in two photos of the two objects.

The reason behind the abovementioned problems might be the used datasets. In fact, no group images of cows were included on purpose in the training/validation datasets, which might have limited the identification ability of the CNN model under such circumstances.

Recognizing cows in videos is very similar to recognizing objects in still images, just the video should be preprocessed and, in most cases, converted to multiple still images (shots). In our case, the video of each cow was analyzed and converted to a video with 1 fps framerate. Thereafter, an operator manually classified each frame of the video as either true positive, false negative, or false positive to assess the accuracy of the CNN model. This way the performance of the model is evaluated individually for each object. In the following experimental analysis, we assessed the model performance only for the first 11 cows since video materials were only available for them. The total number of video frames used for testing is 1448.

Each object was analyzed from two perspectives:

Fig. 10. An example of 2 cows not recognized in a group photo with many others overlapping.

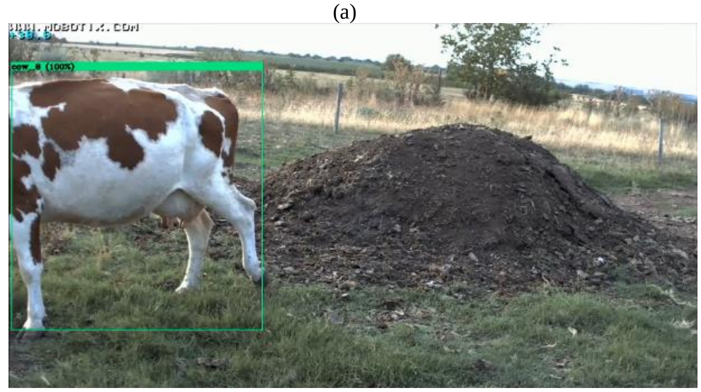
Fig. 11. An example of 3 overlapping cows recognized as a single object.

- The animal is fully visible by the camera [Fig. 12(a)] – it is assumed that at least the following is visible: full carcass, full head (unless it is behind the carcass), and at least half of the legs.
- The animal is only partially visible by the camera [Fig. 12(b)] – at least one of the abovementioned requirements is not met. Nevertheless, large enough part of the cow should be visible, so that the operator can recognize it.

The results from the analysis of the video frames, including true positives, false negatives, false positives, precision, recall, and F1 score are summarized in Table II. When the objects are fully visible the F1 score varies between 0.89 and 1.00 for the different objects and is above 0.95 for most of them. These results are very similar to those obtained in still images of single animals, which corresponds to the expectations. The only exceptions are cow 2 (0.89) and cow 6 (0.90). Yet, this is not unexpected because a limited number of images for these objects were used to train the CNN model - 24 and 35, respectively; while for all other objects, the available images are more than 50. The average accuracy of the trained optimal model was obtained to be 92.2% when used with non-group photos.

When the cows are not fully visible, the F1 score varies between 0.43 and 0.98 for the different objects. The lowest results were achieved for cow_6 (0.43), cow_5 (0.80), and cow_11 (0.84), which is caused mainly by the lower recall, which accounts for the influence of the false negatives. In most cases, the cropped animals were simply not recognized, which is an expected behavior in such situations.

(a)



(b)

Fig. 12. Examples for fully visible (a) and partially visible (b) cow, according to the accepted rules in this study.

TABLE II. A SAMPLE CONFUSION MATRIX

Situation	TP[c]	FN[d]	FP[e]	Precision	Recall	F1
Cow 1 (fully visible) [a]	31	0	2	0.94	1.00	0.97
Cow 1 (partially visible) [b]	40	8	0	1.00	0.83	0.91
Cow 2 (fully visible)	31	8	0	1.00	0.80	0.89
Cow 2 (partially visible)	58	3	0	1.00	0.95	0.98
Cow 3 (fully visible)	71	0	2	0.97	1.00	0.99
Cow 3 (partially visible)	57	4	2	0.97	0.93	0.95
Cow 4 (fully visible)	98	0	0	1.00	1.00	1.00
Cow 4 (partially visible)	26	5	2	0.93	0.84	0.88
Cow 5 (fully visible)	112	2	11	0.91	0.98	0.95
Cow 5 (partially visible)	37	5	13	0.74	0.88	0.80
Cow 6 (fully visible)	42	2	7	0.86	0.95	0.90
Cow 6 (partially visible)	3	6	2	0.60	0.33	0.43
Cow 7 (fully visible)	34	0	0	1.00	1.00	1.00
Cow 7 (partially visible)	30	9	1	0.97	0.77	0.86
Cow 8 (fully visible)	63	0	0	1.00	1.00	1.00
Cow 8 (partially visible)	22	1	0	1.00	0.96	0.98
Cow 9 (fully visible)	68	0	0	1.00	1.00	1.00
Cow 9 (partially visible)	35	1	1	0.97	0.97	0.97
Cow 10 (fully visible)	126	1	2	0.98	0.99	0.99
Cow 10 (partially visible)	24	0	4	0.86	1.00	0.92
Cow 11 (fully visible)	126	5	2	0.98	0.96	0.97
Cow 11 (partially visible)	28	11	0	1.00	0.72	0.84
Average (fully visible)	802	18	26	0.97	0.98	0.97
Average (partially visible)	360	53	25	0.94	0.87	0.90

[a.] The following parts of the animal are visible: the whole carcass, at least half of the legs, and the whole head. If part of the carcass or head are not visible because of the angle of view, the cow is also considered to be fully visible.

[b.] In all other cases the cow is considered to be only partially visible.

[c.] The object was correctly recognized

[d.] The object was not recognized or was incorrectly recognized.

[e.] Another object was recognized as this one.

C. Comparison with Previous Studies

The results obtained in this study cannot be directly compared to other studies, as to the best of our knowledge no previous studies have tried to identify cows using mixed images, showing them from different points of view. Nevertheless, we can compare our results with those obtained in studies, using topview, sideview, backview and faceview images, independently. In [27] outside topview images of cows were used, representing either the full body or randomly cropped body. The obtained accuracy was 95.23% and 90.85%, respectively, which generally corresponds to the average 92% accuracy, achieved in our study.

In [29] sideview full body images were used to achieve 98.58% accuracy. Other similar studies are [36] and [37], reaching average accuracies of 96.65% and 90.2%, respectively. In [37] an F1 score of 86% was estimated, which is lower than the one obtained in this study (97% and 90%, respectively for full body and cropped body). In [31] backside full body images of cows were used to achieve 99.76% accuracy, and in [22] head images for face recognition were used with 93% accuracy and 93.54% F1 score.

Table III summarizes the results from the performed comparison. It can be noticed that most studies assessed only the accuracy of the trained models, which is known to be misleading in the case of imbalanced datasets. That is why in our study we have also obtained precision, recall, and their average F1 score, which gives a more accurate evaluation of the performance of the trained CNN.

TABLE III. COMPARISON OF OUR RESULTS WITH THOSE, OBTAINED IN PREVIOUS STUDIES

Paper	CNN backbone	View	Location	Cropping	Accuracy	F1 score
Wang et al. [27]	ResSTN	Topview	Outside	Full body	95.23%	N/A
				Randomly cropped	90.85%	N/A
Fu et al. [29]	ResNet50	Sideview	Inside	Full body	98.58%	N/A
Hou et al. [31]	Mobilenet v2	Backside	Inside	Full body	99.76%	N/A
Shen et al. [36]	AlexNet	Sideview	Inside	Full body	96.65%	N/A
Wang et al. [37]	ResNet101	Sideview	Outside	Full body	90.2%	86%
Xu et al. [22]	Capsule network	Faceview	Inside	Head only	93.00%	93.54%
Ours	Darknet-53	All sides	Outside	Full body	92.2%	97%
				Randomly cropped		90%

IV. CONCLUSIONS

In this study, the possibility of training a convolutional neural network for recognizing cows regardless of the viewing angle was investigated. A dataset containing 801 images of 14 cows was used for training a YOLOv3 model. To prepare the training, validation, and testing datasets, photo sessions of each animal were made, so that they represent the cows from different sides. Out of the trained models, the one selected as optimal achieved a mean average precision of 98.81% after 6850 iterations.

The results from the evaluation of the trained CNN showed that its recognition rate greatly depends on the usage circumstances. The model showed excellent performance at recognizing objects on images/video frames, where there are no other animals, where the animal is in the front, or where the animal is not overlapping with other animals. This is confirmed by the obtained 97% F1 score for fully visible cows and 90% for partially visible cows, as well as by the achieved average accuracy of 92.2%. On the other hand, the trained YOLOv3 CNN does not perform well when numerous objects exist on the image/video frame, especially with overlapping objects. In such a situation, the model fails to identify all cows and rarely recognizes more than one.

The abovementioned indicates that the proposed approach for the recognition of cows with a non-fixed point-of-view is applicable but with certain limitations and under certain usage scenarios. Such limitations are: the object being identified should be alone on the image/frame, or should not be overlapping with other animals. The investigated approach has numerous possible applications, such as monitoring the animals' movement, localization, and behavior, or even as a replacement for ear tags.

In the present study, 801 images for training were used, which is 57 on average per animal. It is interesting to investigate if the performance of CNN models will increase when more training images are used, especially in group photos with overlapping cows. Furthermore, more group photos of animals should be included in the training/validation datasets, which might increase the recognition rate in such scenarios. If that doesn't happen, another option is to train an additional neural network, responsible for recognizing the extents of each object. With such an approach each cow could be extracted from the image and identified independently of the others, which should allow reducing the impact of overlapping animals. The abovementioned represent promising research topics, which are an object for future studies.

ACKNOWLEDGMENT

This research was funded by the Ministry of Education and Science of Bulgaria under the National Research Program "Intelligent Animal Husbandry", grant number Д01-62/18.03.2021.

REFERENCES

[1] M. Knock, G.A. Carroll, "The Potential of Post-Mortem Carcass Assessments in Reflecting the Welfare of Beef and Dairy Cattle," Animals, vol. 9, no. 959, 2019, doi: https://doi.org/10.3390/ani9110959

[2] P. Kumar, S. Luo, and K. Shaukat, "A Comprehensive Review of Deep Learning Approaches for Animal Detection on Video Data" International Journal of Advanced Computer Science and Applications(IJACSA), vol. 14, no. 11, 2023, doi: http://dx.doi.org/10.14569/IJACSA.2023.01411144

[3] T. Gao, D. Fan, H. Wu, X. Chen, S. Song, Y. Sun, J. Tian, "Research on the Vision-Based Dairy Cow Ear Tag Recognition Method," Sensors, vol. 24, no. 2194, 2024, doi: https://doi.org/10.3390/s24072194

[4] J. Li, Y. Liu, W. Zheng, X. Chen, Y. Ma, L. Guo, "Monitoring Cattle Ruminating Behavior Based on an Improved Keypoint Detection Model," Animals, vol. 14, no. 1791, 2024, doi: https://doi.org/10.3390/ani14121791

[5] Y. Kalmukov, B. Evstatiev, "Methods for Automated Remote Sensing and Counting of Animals," 2022 8th International Conference on Energy Efficiency and Agricultural Engineering (EE&AE), pp. 1-7, IEEE, 2022, doi: https://doi.org/10.1109/EEAE53789.2022.9831239

[6] S. Hu, A. Reverter, R. Arablouei, G. Bishop-Hurley, J. McNally, F. Alvarenga, A. Ingham, "Analyzing Cattle Activity Patterns with Ear Tag Accelerometer Data," Animals, vol. 14, no. 301, 2024, doi: https://doi.org/10.3390/ani14020301

[7] T. Mladenova, I. Valova, B. Evstatiev, N. Valov, I. Varlyakov, T. Markov, S. Stoycheva, L. Mondeshka, N. Markov, "Evaluation of the Efficiency of Machine Learning Algorithms for Identification of Cattle Behavior Using Accelerometer and Gyroscope Data," AgriEngineering, vol. 6, no. 3, pp. 2179-2197, 2024, doi: https://doi.org/10.3390/agriengineering6030128

[8] T.T. Zin, M.Z. Pwint, P.T. Seint, S. Thant, S. Misawa, K. Sumi, K, Yoshida, "Automatic Cow Location Tracking System Using Ear Tag Visual Analysis," Sensors, vol. 20, no. 3564, 2020, doi: https://doi.org/10.3390/s20123564

[9] K. Arai, T. Higuchi and T. Murakami, "Estimation Method of the Total Number of Wild Animals based on Modified Jolly's Method" International Journal of Advanced Computer Science and Applications(IJACSA), vol. 8, no. 1, 2017, doi: http://dx.doi.org/10.14569/IJACSA.2017.080143

[10] O. Ittoo and S. Pudaruth, "Automatic Recognition of Marine Creatures using Deep Learning" International Journal of Advanced Computer Science and Applications(IJACSA), vol. 15, no. 1, 2024, doi: http://dx.doi.org/10.14569/IJACSA.2024.0150106

[11] Z. Weng, S. Liu, Z. Zheng, Y. Zhang, C. Gong, "Cattle Facial Matching Recognition Algorithm Based on Multi-View Feature Fusion," Electronics, vol. 12, no. 156, 2023, doi: https://doi.org/10.3390/electronics12010156

[12] N. M. Arago, C. I. Alvarez, A. G. Mabale, C. G. Legista, N. E. Repiso, R. R. A. Robles, T. M. Amado, R. Jr. L. Jorda, A. C. Thio-ac, J. S. Velasco, and L. K. S. Tolentino, "Automated Estrus Detection for Dairy Cattle through Neural Networks and Bounding Box Corner Analysis" International Journal of Advanced Computer Science and Applications(IJACSA), vol. 11, no. 9, 2020, doi: http://dx.doi.org/10.14569/IJACSA.2020.0110935

[13] P. Houngue, R. Sagbo, G. Dahoue, and J. Komaclo, "Smart Monitoring System using Internet of Things: Application for Agricultural Management in Benin" International Journal of Advanced Computer Science and Applications(IJACSA), vol. 13, no. 12, 2022, doi: http://dx.doi.org/10.14569/IJACSA.2022.01312114

[14] S.C. Tun, T. Onizuka, P. Tin, M. Aikawa, I. Kobayashi, T.T. Zin, "Revolutionizing Cow Welfare Monitoring: A Novel Top-View Perspective with Depth Camera-Based Lameness Classification," J. Imaging, vol. 10, no. 67, 2024, doi: https://doi.org/10.3390/jimaging10030067

[15] W. Hao, C. Ren, M. Han, L. Zhang, F. Li, Z. Liu, "Cattle Body Detection Based on YOLOv5-EMA for Precision Livestock Farming," Animals, vol. 13, no. 3535, 2023, doi: https://doi.org/10.3390/ani13223535

[16] Ł. Popek, R.. Perz, G. Galiński, "Comparison of Different Methods of Animal Detection and Recognition on Thermal Camera Images," Electronics, vol. 12, no. 270, 2023, doi: https://doi.org/10.3390/electronics12020270

[17] S. Binta Islam, D. Valles, T.J. Hibbitts, W.A. Ryberg, D.K. Walkup, M.R.J. Forstner, "Animal Species Recognition with Deep Convolutional Neural Networks from Ecological Camera Trap Images," Animals, vol. 13, no. 1526, 2023, doi: https://doi.org/10.3390/ani13091526

[18] M. Tan, W. Chao, J.-K. Cheng, M. Zhou, Y. Ma, X. Jiang, J. Ge, L. Yu, L. Feng, "Animal Detection and Classification from Camera Trap Images Using Different Mainstream Object Detection Architectures," Animals, vol. 12, no. 1976, 2022, doi: https://doi.org/10.3390/ani12151976

[19] M. Ahmad, S. Abbas, A. Fatima, G.F. Issa, T.M. Ghazal, M.A. Khan, "Deep Transfer Learning-Based Animal Face Identification Model Empowered with Vision-Based Hybrid Approach," Appl. Sci., vol. 13, no. 1178, 2023, doi: https://doi.org/10.3390/app13021178

[20] Z. Wan, F. Tian, C. Zhang, "Sheep Face Recognition Model Based on Deep Learning and Bilinear Feature Fusion," Animals, vol. 13, no. 1957, 2023, doi: https://doi.org/10.3390/ani13121957

[21] D. Yeleshetty, L. Spreeuwers, Y. Li, “3D Face Recognition for Cows,” 2020 International Conference of the Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, 2020, pp. 1-5.

[22] F. Xu, J. Gao, X. Pan, “Cow face recognition for a small sample based on Siamese DB Capsule Network,” Ieee Access, vol. 10, 2022, pp. 63189-63198, doi: https://doi.org/10.1109/ACCESS.2022.3182806

[23] R.W. Bello, A.Z.H. Talib, A.S.A.B. Mohamed, “Deep learning-based architectures for recognition of cow using cow nose image pattern,” Gazi University Journal of Science, vol. 33, no. 3, 2020, pp. 831-844, doi: https://doi.org/10.35378/gujs.605631

[24] G. Li, G.E. Erickson, Y. Xiong, “Individual Beef Cattle Identification Using Muzzle Images and Deep Learning Techniques,” Animals, vol. 12, no. 1453, 2022, doi: https://doi.org/10.3390/ani12111453

[25] M. Ramesh, A.R. Reibman, J.P. Boerman, “Eidetic recognition of cattle using keypoint alignment,” Electronic Imaging, vol. 279, 2023, pp. 279-1 - 279-6, doi: https://doi.org/10.2352/EI.2023.35.7.IMAGE-279

[26] R.W. Bellom, A.Z. Talib, A.S. Mohamed, D.A. Olubummo, F.N. Otobo, “Image-based individual cow recognition using body patterns,” International Journal of Advanced Computer Science and Applications(IJACSA), vol. 11, no. 3, 2020, doi: http://dx.doi.org/10.14569/IJACSA.2020.0110311

[27] B. Wang, X. Li, X. An, W. Duan, Y. Wang, D. Wang, J. Qi, “Open-Set Recognition of Individual Cows Based on Spatial Feature Transformation and Metric Learning,” Animals, vol. 14, no. 1175, 2024, doi: https://doi.org/10.3390/ani14081175

[28] A. Fuentes, S. Han, M.F. Nasir, J. Park, S. Yoon, D.S. Park, “Multiview Monitoring of Individual Cattle Behavior Based on Action Recognition in Closed Barns Using Deep Learning,” Animals, vol. 13, 2020, doi: https://doi.org/10.3390/ani13122020

[29] L. Fu, S. Li, S. Kong, R. Ni, H. Pang, Y. Sun, T. Hu, Y. Mu, Y. Guo, H. Gong, “Lightweight individual cow identification based on Ghost combined with attention mechanism,” Plos one, vol. 17, no. 10, 2022, p.e0275435, doi: https://doi.org/10.1371/journal.pone.0275435

[30] F. Lv, C. Zhang, C. Lv, “Image recognition of individual cow based on SIFT in Lαβ color space,” MATEC Web of Conferences, vol. 176, no. 01023, 2018, doi: https://doi.org/10.1051/matecconf/201817601023

[31] H. Hou, W. Shi, J. Guo, Z. Zhang, W. Shen, S. Kou, “Cow Rump Identification Based on Lightweight Convolutional Neural Networks,” Information, vol. 12, no. 361, 2021, doi: https://doi.org/10.3390/info12090361

[32] R.W. Bello, A.Z. Talib, A.S.A. Mohamed, “Real-time cow detection and identification using enhanced particle filter,” IOP Conference Series: Materials Science and Engineering, vol. 1051, no. 1, 2021, doi: https://doi.org/10.1088/1757-899X/1051/1/012001

[33] F. Okura, S. Ikuma, Y. Makihara, D. Muramatsu, K. Nakada, Y. Yagi, “RGB-D video-based individual identification of dairy cows using gait and texture analyses,” Computers and Electronics in Agriculture, vol. 165, no. 104944, 2019, doi: https://doi.org/10.1016/j.compag.2019.104944

[34] A. Pretto, G. Savio, F. Gottardo, F. Uccheddu, G. Concheri, “A novel low-cost visual ear tag based identification system for precision beef cattle livestock farming,” Information Processing in Agriculture, vol. 11, no. 1, 2024, doi: https://doi.org/10.1016/j.inpa.2022.10.003

[35] J. Redmon, A. Farhadi, “YOLOv3: An incremental improvement,” arXiv:1804.02767, 2018, [online] Available: https://arxiv.org/abs/1804.02767

[36] W. Shen, H. Hu, B. Dai, et al., “Individual identification of dairy cows based on convolutional neural networks,” Multimed Tools Appl, vol. 79, 2020, pp. 14711-14724, doi: https://doi.org/10.1007/s11042-019-7344-7

[37] J. Wanget al., “Open Pose Mask R-CNN Network for Individual Cattle Recognition,” IEEE Access, vol. 11, 2023, pp. 113752-113768, doi: https://doi.org/10.1109/ACCESS.2023.3321152