DETECTION AND COUNTING OF GRAZING CATTLE FROM AERIAL IMAGES USING CNN
/
CNN АШИГЛАН АГААРЫН ЗУРГААС БЭЛЧЭЭРИЙН МАЛЫГ ИЛРҮҮЛЭХ, ТООЛОХ

Hu GEJILETU[1)], Hishigdalai ULAANKHUU[2)], Bat-Erdene NYANDAG[3*)]

[1,3)] Department of Computer Science, School of Applied Sciences, Mongolian University of Life Sciences, 17024, Ulaanbaatar, Mongolia

[2)] School of Business, Mongolian University of the Humanities, 14200-0031 Ulaanbaatar, Mongolia

**Corresponding author, Email: baterdene@muls.edu.mn*

DOI: https://doi.org/10.35633/inmateh-75-39

Keywords: *deep learning, drone, object detection, yolov8*

ABSTRACT

This study explores the use of deep neural networks for detecting and quantifying the cattle population in Mongolia using drone imagery, addressing the limitations of traditional methods that are labor-intensive and time-consuming. A custom dataset of aerial images featuring grazing cattle in Mongolia was developed, focusing on winter and spring seasons, to train and validate a model based on state-of-the-art object detection algorithms. Specifically, the You Only Look Once (YOLOv8) architecture was employed to detect cattle across diverse environmental conditions. Model performance was evaluated using widely accepted metrics, including precision, recall, F1 score, and the mean average precision (mAP). The findings demonstrate the effectiveness of the proposed approach, with the YOLOv8 model achieving a mAP of 97.3% at an IoU threshold of 0.5, highlighting its potential for efficient cattle detection and monitoring in Mongolia's unique environmental contexts.

ХУРААНГУЙ

Энэхүү судалгаа нь дроны зураг ашиглан Монгол дахь үхрийн тоо толгойг илрүүлэх, тоо хэмжээг тогтооход гүн мэдрэлийн сүлжээг ашигласан. Энэхүү ажил нь хөдөлмөр, цаг хугацаа их шаарддаг уламжлалт аргуудаас татгалзах боломжийг судалсан. Хамгийн сүүлийн үеийн объект илрүүлэх алгоритмд суурилсан загварыг сургаж, турших зорилгоор өвөл, хаврын улиралд анхаарлаа хандуулж, Монголын бэлчээрийн үхрийг харуулсан агаарын зургийн зорилтод өгөгдлийн багцыг боловсруулсан. Үүндээ YOLOv8 архитектурыг байгаль орчны янз бүрийн нөхцөлд үхэр илрүүлэхэд ашигласан. Загварын гүйцэтгэлийг нарийвчлал, санах ой, F1 оноо, дундаж нарийвчлал (mAP) зэрэг нийтээр хүлээн зөвшөөрөгдсөн хэмжигдэхүүнүүдийг ашиглан үнэлэв. Судалгааны үр дүн нь санал болгож буй аргын үр дүнг харуулж байгаа бөгөөд YOLOv8 загвар нь IoU-ийн босго 0.5-д 97.3%-ийн mAP-д хүрсэн нь Монгол орны байгаль орчны өвөрмөц нөхцөлд үхрийг үр дүнтэй илрүүлэх, хянах, тоолох боломжийг харуулсан.

INTRODUCTION

According to the preliminary results of the annual livestock census, by the end of 2023, Mongolia had 64.7 million head of livestock, including 5.4 million cattle *(National Statistics Office of Mongolia).* In the traditional setting, cattle farming typically occurs in a natural environment in Mongolia, where challenges like accidental drownings in rivers, snowbound or landslides may cause substantial damage and create serious challenges for cattle management *(Xu et al., 2020).* Consequently, monitoring cattle in Mongolia, including their behaviors and health, has emerged as a critical research area. The rapid progress in deep learning, particularly in methods involving convolutional neural networks (CNNs), offers effective solutions for detecting and classifying animals *(Radovic et al., 2017).*

Spanning the years 1980 to 2023 across all continents, a notable 71.4% of studies have been published since 2019, reflecting an increased focus on livestock detection in recent years *(Ocholla et al., 2024).* Furthermore, 69.2% of the studies relied on drones for detection and counting purposes, with alternative methods such as manned aircraft, satellites, and camera traps being used less frequently *(Ocholla et al., 2024).* The increasing popularity of drones is a result of innovations in embedded systems and electronic communication, which have reduced costs and enhanced their availability *(Olson and Anderson, 2021).*

Fig. 1 presents the distribution of studies on livestock detection and counting analyzed across various countries (Ocholla *et al., 2024).*

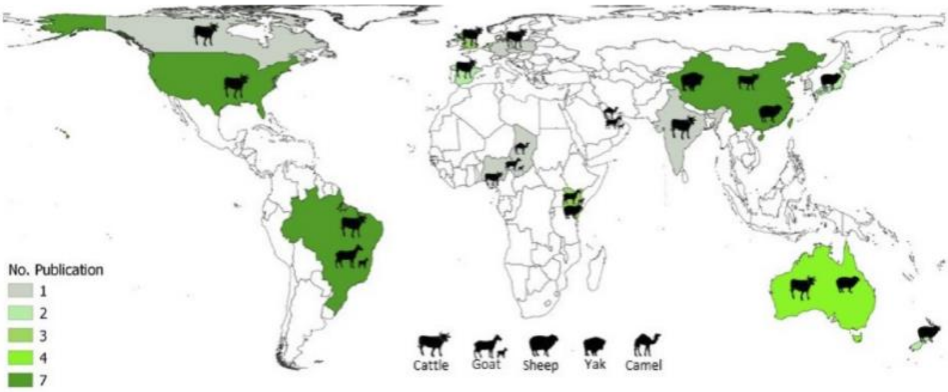

Fig. 1 - Country-wise distribution of studies on livestock detection and counting

Our research aims to leverage a cutting-edge object detection framework, specifically the YOLOv8 algorithm *(Redmon et al., 2016)*, to accurately detect and count grazing cattle in Mongolian grasslands using aerial images captured by drones.

Animal monitoring using remote technologies

Innovations in remote monitoring technologies have significantly improved data collection on animal behavior, enabling farmers to enhance meat quality, boost production efficiency, and ensure better health for livestock *(Xu et al., 2020).* Farmers are employing wearable technologies, including RFID, GPS collars *(*Derek *et al., 2018),* and smart ear tags *(Kumar et al., 2016)*, to track movement, behavior patterns, heart rate, body temperature, and other physiological indicators. These technologies play a crucial role in preventing illnesses and deaths among animals, thereby minimizing production losses *(Xu et al., 2020).*

Motion-sensor cameras, commonly referred to as camera traps, offer a cost-effective method for recording animals' presence, locations, and activities *(Verma and Gupta, 2018).* These technologies collectively contribute to a more efficient and sustainable approach to animal monitoring and management in agricultural settings.

One significant drawback of existing ground-based techniques, such as smart ear tags and motion-sense cameras, is their constraint within large geographic areas and complex terrains, making it challenging to identify and track animals effectively *(Gonzalez et al., 2016).* Combining drone technology with advanced machine learning models offers a promising approach to improving the management of livestock. Compared to traditional methods, drones demonstrate several unique advantages *(Windrim et al., 2019)*:

1. They are capable of operating at both low and extremely low altitudes.
2. They can acquire detailed, high-quality images even under varying weather conditions.
3. They can rapidly capture imagery across vast areas and challenging terrains that are otherwise difficult to access.

However, accurately and reliably counting animals in drone-captured imagery remains a crucial yet challenging task in intelligent livestock management *(Wang et al., 2021).* The primary aim of this research is to explore deep learning approaches for the detection and quantification of animals, focusing on the automated analysis of cattle populations in drone-based images. Additionally, it highlights the potential of drone vision and object detection techniques to improve livestock management practices.

Drone-based animal detection and counting

CNNs *(Lecun et al., 2015)* have proven to be an effective and reliable method for detection and counting tasks in image recognition, owing to its processing speed and accuracy. In recent years, advanced deep learning models such as Faster R-CNN *(Ren et al., 2015)*, Mask R-CNN *(He et al., 2017)*, and YOLO *(Redmon et al., 2016)* have shown great promise in detecting and classifying objects across vast datasets, achieving higher accuracy, precision, and faster processing speeds.

Although these results are impressive, livestock monitoring still faces challenges in diverse and complex scenarios, including visual clutter (e.g., vegetation), strong lighting contrasts and shadows from farm structures, low target resolution, and high animal densities (e.g., tightly packed herds or feedlots) *(Xu et al., 2020).* Therefore, a thorough evaluation of cattle-counting algorithms across various livestock farming environments in Mongolia is essential.

Deep learning algorithms for object detection are predominantly built on CNNs. These networks excel in image processing due to their capacity to extract hierarchical features directly from pixel-level data. These algorithms are specifically designed to identify and locate objects within images or videos *(Lecun et al., 2015).* Deep learning-based object detection methods can be classified into two categories: Two-Stage Detectors and One-Stage Detectors.

Two-Stage Detectors operate by initially generating potential regions of interest (RoIs) within an image using techniques such as selective search or region proposal networks (RPNs). In the subsequent step, these RoIs are classified, and their bounding boxes are adjusted for precision. Prominent examples of this category include R-CNN (Region-based Convolutional Neural Networks) *(Girshick et al., 2014)*, Faster R-CNN *(Ren et al., 2015)*, and Mask R-CNN *(He et al., 2017)*. In contrast, One-Stage Detectors eliminate the need for a separate region proposal stage and directly predict bounding boxes along with class probabilities. Notable examples of this approach include YOLO (You Only Look Once) *(Redmon et al., 2016)* and SSD (Single Shot Multibox Detector) *(Liu et al., 2016)*.

R-CNN and its variants, such as Faster R-CNN, are two-stage object detection frameworks known for their accuracy and detailed object localization but are slower and more computationally demanding. In contrast, YOLO is a single-stage algorithm optimized for real-time performance, making it ideal for applications requiring fast and efficient object detection, such as drone-based monitoring or video analysis *(Xu et al., 2020).*

Recent advancements in object detection methods have significantly enhanced the accuracy of monitoring livestock in pastures and open areas using drone-captured data.

The research done by *João Vitor de Andrade Porto et al., (2021),* highlights the application of Faster R-CNN for counting cattle in feedlots through aerial imagery, reporting an average precision of 89.7%.

Based on YOLOX, detection performance for counting cattle was improved through input resolution optimization, reaching a precision of 95.7% *(Wang et al., 2023)*.

Mask R-CNN was applied to extract features and trained on drone-based imagery, achieving a classification accuracy of 96% for livestock *(Xu et al., 2020).*

Among the approaches discussed in the literature, YOLO stands out as a particularly promising method for cattle monitoring in real-time drone scenarios. The newer iteration, YOLOv8 *(*https://u*ltralytics.com)*, demonstrates significant improvements in speed, accuracy, and efficiency, making it a suitable benchmark for this study. This research aims to evaluate the feasibility and effectiveness of using YOLOv8 to detect and quantify cattle populations in diverse environmental conditions. The objectives of the study are as follows:

1. Apply the YOLOv8 model, pre-trained on the MS COCO dataset, to detect cattle in drone-captured images across various seasons.
2. Collect drone imagery of cattle during the winter and spring seasons to develop a custom dataset.
3. Train a custom YOLOv8 model using the collected dataset and evaluate its performance using standard metrics such as mAP, precision, recall, and F1 score.
4. Analyze cattle detection results across different seasons using the pre-trained and custom-trained models to validate the correctness and effectiveness of this research.
5. Offer insights into the applicability of deep learning object detection techniques for livestock monitoring and propose directions for future research in this domain.

Object detection techniques

Object detection is essential for this task, as it identifies and categorizes objects of interest by enclosing them within bounding boxes. The process typically involves two or three stages: generating bounding boxes to identify regions of interest, extracting features from these regions, and classifying objects to determine their categories. This study focuses on detecting a single class: cattle. In addition to detection, accurate localization of animals within images is crucial for effective counting and monitoring of livestock behavior.

Modern object detection models typically comprise two primary components: a backbone for feature extraction and a head for object classification and localization. The backbone processes input images and generates a feature map that supports the remaining network operations. The choice of backbone (e.g., VGG *(Simonyan and Zisserman, 2015)*, ResNet *(He et al., 2016)*, DenseNet *(Huang et al., 2017)*, EfficientNet *(Tan and Le, 2019)*) depends on the detector’s operating platform (CPU or GPU).

Object detection methods can be broadly categorized into two approaches: the two-stage approach, which employs region proposal algorithms, and the one-stage approach, which directly combines detection and classification tasks into a single step for real-time processing. This study focuses on the one-stage method for two key reasons. First, one-stage detectors are faster, as they perform both detection and classification simultaneously, making them ideal for real-time applications and well-suited to the objectives of this project. Second, one-stage detectors are more straightforward and demand lower computational resources compared to two-stage detectors, making them a practical choice given the project's resource constraint.

The YOLO algorithm, introduced in 2015, gained popularity due to its simplicity, speed, and efficiency as a one-stage approach, requiring fewer computational resources and memory while achieving high detection speeds. This makes it ideal for real-time deployment on drones for object detection.

In this study, the YOLOv8 model, pre-trained on the MS COCO dataset, was trained on a custom dataset of 640 × 640-pixel images to optimize it for cattle detection in drone-captured images. The training was conducted using a learning rate of 0.001, a decay rate of 0.0005, and a momentum value of 0.937. These parameters were carefully chosen to enhance training efficiency and improve YOLOv8's accuracy in detecting cattle from drone-captured images.

MATERIALS AND METHODS

Datasets preparation and preprocessing

Currently, numerous large-scale open-source datasets, such as MS COCO *(Tsung-Yi Lin et al., 2014)*, are widely available for various computer vision tasks. However, these datasets predominantly consist of ground-based photographs and lack drone-captured images. Additionally, most cattle images within these datasets represent tropical and subtropical cattle, often depicted against green backgrounds in hot summer conditions. Initially, this study employed YOLOv8, pre-trained on the default MS COCO model, to predict cattle in drone-captured images from three seasons in Mongolia: summer, winter, and spring. The results indicated that YOLOv8, trained on the COCO dataset, achieved high accuracy in detecting cattle during the summer (see Fig. 2) but exhibited poor performance in winter and spring (Fig. 3). This finding highlights the absence of features representing Mongolian cattle during winter and spring in the MS COCO dataset. To address this limitation, additional drone images of cattle in Mongolia were collected during these seasons, and a custom model was trained specifically to detect Mongolian cattle in aerial images with high accuracy.

Fig. 2 - High accuracy of YOLOv8 pre-trained on the COCO dataset in detecting cattle during the summer

Fig. 3 - Low accuracy of YOLOv8 pre-trained on the COCO dataset in detecting cattle during the winter

The datasets utilized in this study were collected from Ulaanbaatar and its surrounding regions in Mongolia during the winter and spring seasons. Examples of cattle in winter are shown in Fig. 4, while examples for spring are depicted in Fig. 5.

Fig. 4 - Drone-captured image of cattle in winter

Fig. 5 - Drone-captured image of cattle in spring

Many flight activities were conducted using the Potensic drone *(www.potensic.com/atom.html)*, equipped with an integrated Sony camera, as shown in Fig. 6. The camera features a 1/2.3-inch CMOS image sensor capable of lateral and vertical rotation. It captures 4K videos and 12-megapixel photos with stabilization. Considering factors such as pixel resolution, photo capture delay, and operational convenience, videos were selected to create datasets of cattle in diverse scenes. The videos were recorded at 30 frames per second and saved in MOV format.

Fig. 6 - Image of the Potensic drone used for data collection

The drone performed 360-degree rotations above the cattle herd at altitudes ranging from 15 to 35 meters, capturing their posture from multiple angles. This approach significantly improved the accuracy of deep learning-based detection. Videos of varying lengths, up to 30 minutes, were recorded by the drone. However, as the YOLO training model requires image files, the OpenCV library was utilized in this study to extract frames from the videos. The most effective images for training were then carefully selected.

The winter dataset comprises 100 images containing a total of 1,563 cattle, with 70 images (1,376 cattle) allocated for training and 30 images (187 cattle) for testing. The spring dataset consists of 138 training images with 1,359 cattle, 17 testing images with 172 cattle, and 18 validation images containing 180 cattle. To ensure accurate labeling of ground truth data, the widely recognized image annotation and analysis platform Roboflow was utilized.

Our research is planned to continue over several years, with the aim of creating and sharing a benchmark drone-captured dataset of Mongolian livestock across all four seasons for scholarly purposes.

The Algorithm for detection and counting cattle

YOLOv8, developed by Ultralytics *(https://ultralytics.com)*, is the newer version of the widely recognized model for real-time object detection and image segmentation. Leveraging advancements in deep learning and computer vision, YOLOv8 achieves exceptional performance in both speed and accuracy. Its optimized architecture enhances adaptability, enabling deployment across diverse applications and ensuring compatibility with a range of hardware environments, including edge devices and cloud-based systems.

YOLOv8 incorporates several key features that contribute to its high performance:

1. Mosaic Data Augmentation: This training technique merges four images into a single composite image, enabling the model to learn from varied object placements and contexts.
2. Anchor-Free Detection: This approach eliminates reliance on predefined anchor boxes by directly predicting object centers and bounding box dimensions, enabling more accurate and efficient object localization.
3. C2f Module: The C2f module enhances feature extraction by combining convolutional layers with feature fusion, thereby improving the model's accuracy in object detection.
4. Decoupled Head: YOLOv8 incorporates a decoupled head design that separates classification and localization tasks, enhancing both the model's efficiency and accuracy.

5. Loss Function: The model employs a modified loss function to optimize the training process and enhance object detection performance.

These features collectively enhance YOLOv8's advanced performance in fast and accurate object detection and image segmentation.

The implementation details

The YOLOv8 model was set up in a Python environment and initialized with a pre-trained model from the MS COCO dataset. Following the organization, formatting, and annotation of the collected cattle data, the model was trained on this dataset. The trained model was then tested and evaluated to measure its performance.

The training process was conducted on a 64-bit Windows 11 system equipped with an Intel Core i7-13620U CPU running at 2.4 GHz, 16 GB of RAM, and an NVIDIA RTX4050 GPU with 6 GB of dedicated memory.

To evaluate the performance of the proposed approach, precision, average precision (AP), and recall were utilized as evaluation metrics. Precision quantifies the ratio of true positive predictions to the total number of positive predictions (Equation 1), while recall measures the proportion of true positives relative to all actual positive instances (Equation 2). The precision-recall curve provides a comprehensive evaluation of the model's performance by analyzing the area under the curve across different IoU thresholds. Average precision, as defined in Equation 3, summarizes the precision-recall trade-off. IoU (Intersection over Union) quantifies the overlap between the predicted and ground-truth bounding boxes, expressed as the ratio of their intersection to their union, as specified in Equation 4.

$$P = \frac{\text{True Positives}}{\text{True Positives}+\text{False Positives}} \tag{1}$$

$$R = \frac{\text{True Positives}}{\text{True Positives}+\text{False Negatives}} \tag{2}$$

$$AP = \sum_{n=1}^{N}[R(n) - R(n-1)] \times \max P(n) \tag{3}$$

where:

N - the number of Precision-Recall (PR) points calculated;
R(n) - the recall value at the n th point;
P(n) - the precision value at the n th point.

$$IoU = \frac{A \cap B}{A \cup B} \tag{4}$$

RESULTS and DISCUSSIONS

The pre-trained Ultralytics YOLOv8 model is derived from the Microsoft COCO dataset, which predominantly features ground-level imagery. However, this model demonstrates limited suitability for detecting and counting objects in aerial images. An initial evaluation of the COCO-based pre-trained YOLOv8 model revealed inaccuracies, including misclassifications of cattle as birds in aerial imagery. As depicted in Fig. 8, these inaccuracies encompassed additional detections of shadows, missed cattle, and erroneous bird predictions, as further illustrated in Fig. 7.

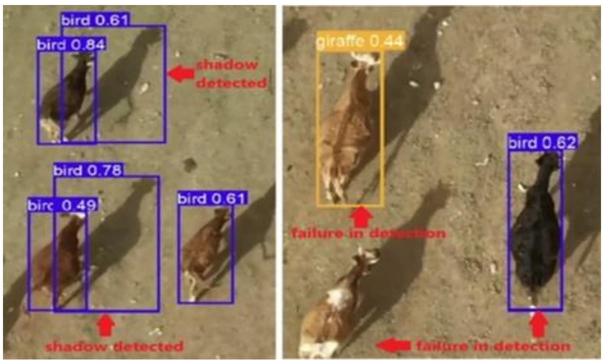

Fig. 7 - Examples of shadow and failure detections

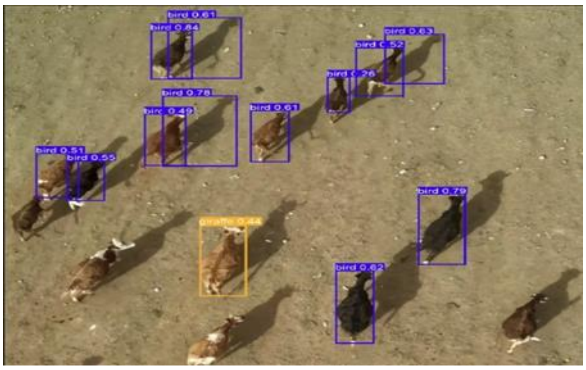

Fig. 8 - Predictions from the model pre-trained on the MS COCO dataset

The cattle detection model was trained on a custom-labeled dataset, resulting in high prediction accuracy. As demonstrated in Fig. 9, the custom-trained model accurately detected all cattle in the image with high confidence scores and no false positives, such as misclassifications involving shadows.

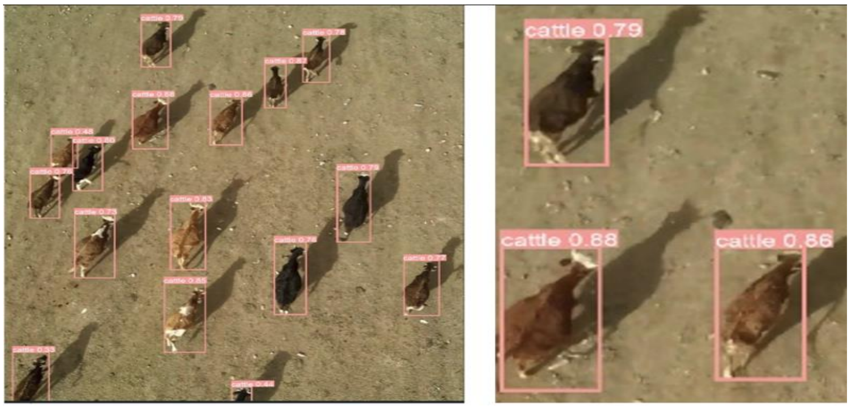

Fig. 9 - Prediction from the trained model with high confidence scores

Despite the blurriness of the images captured during winter, the trained model successfully detected cattle with high accuracy, as illustrated in Fig. 10.

Fig. 10 – Winter cattle detection results using the trained model

Comprehensively capturing a wide range of grazing cattle postures in natural environments is essential for deep learning-based object detection. In this study, low-altitude drones were employed to record 360-degree videos of cattle from multiple angles, ensuring the collection of diverse postural data. This approach was critical in achieving high model accuracy and reliable validation metrics. Unlike horses, which exhibit greater posture variability, cattle tend to graze in a more dispersed manner and display fewer postural variations due to their shorter necks, shorter legs, and robust bodies. These traits contributed to a more streamlined process for annotation, training, and validation in this research.

Accurate and comprehensive labeling of the full body of each cattle in the images, in addition to capturing diverse postures, is crucial for achieving effective object detection. This meticulous annotation process substantially contributes to improving the model's detection accuracy.

The performance of the trained model was assessed using evaluation metrics including mAP, precision, recall, and the F1 score. A comprehensive analysis of the results is presented in the following section.

1. Fig. 11 illustrates the Precision-Confidence Curve for the trained model. The plot indicates that the model achieves high accuracy beginning at a relatively low confidence threshold of 0.2 (20%), with precision generally improving as the confidence threshold increases. Precision attains its maximum value of 1.00 (100%) at a confidence threshold of 0.74 (74.0%), indicating a minimal occurrence of false positives at this level.

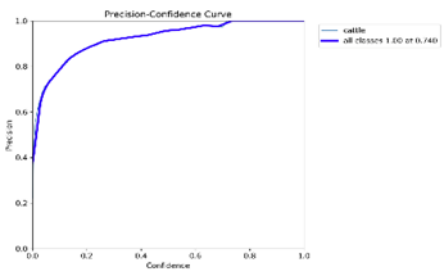

Fig. 11 - Precision-Confidence

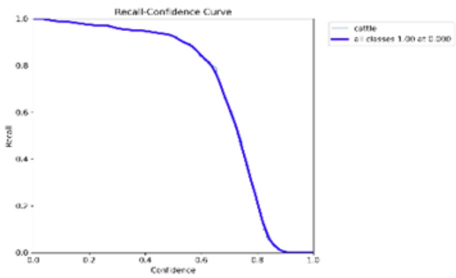

Fig. 12 - Recall-Confidence

2. Fig. 12 presents the Recall-Confidence Curve for the trained model, highlighting its robust capability. Notably, the model achieves and sustains high recall levels at a confidence threshold of 0.7 (70%).
3. Fig. 13 depicts the Recall-Confidence Curve for the trained model. The metric 'All classes 0.973 mAP@0.5' indicates that the model achieves a mAP of 97.3% across all classes at an IoU threshold of 0.5 (50%), reflecting its overall detection performance.

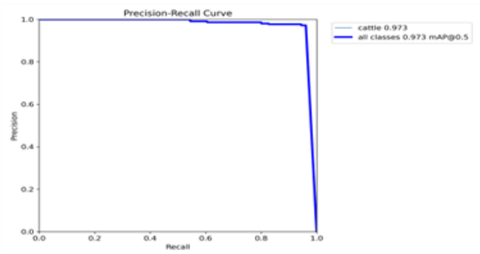

Fig. 13 - Precision-Recall

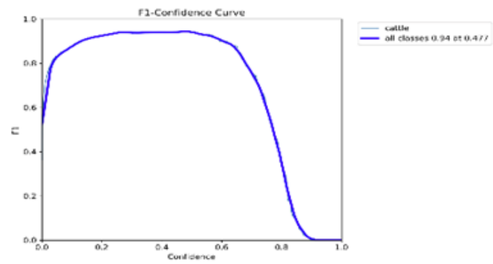

Fig. 14 - F1-Confidence

4. Fig. 14 presents the F1-Confidence Curve for the trained model, highlighting its performance. An F1 score of 0.94 (94%) is achieved at a confidence threshold of 0.477 (47.7%), reflecting a well-balanced trade-off between precision and recall.

Upon completing the training phase, the YOLO validation function was employed to evaluate the model's performance. This validation process was conducted using 180 pre-annotated instances of cattle from 18 images in the dataset's validation set. The corresponding results are summarized in Table 1 below:

Table 1

Evaluation metrics of the trained model during validation

Class	Images [Number]	Instances [Number]	Precision [%]	Recall [%]	mAP@0.5 [%]	mAP@0.5:0.95 [%]
cattle	18	180	97.2%	96.1%	97.3%	59.9%

These metrics indicate excellent model performance, highlighted by high precision, recall, and an impressive mAP@0.5 score. The findings validate the model’s effectiveness in accurately detecting and counting cattle in aerial images.

The confusion matrix for the single-class object detection model is depicted in the following figures. Fig. 15 presents the standard confusion matrix, while Fig. 16 illustrates the normalized version.

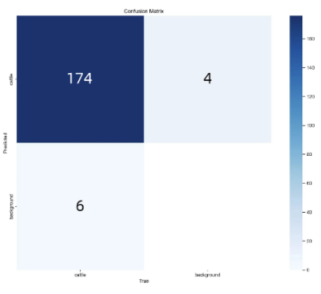

Fig. 15 - Standard confusion matrix

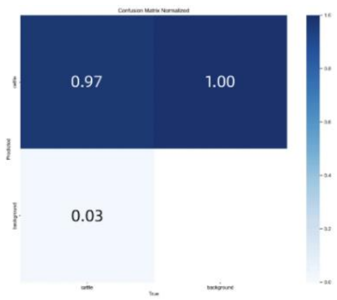

Fig. 16 - Normalized confusion matrix

CONCLUSIONS

The trained model achieved high performance metrics, validating the effectiveness of the proposed research methods and approach. This study demonstrates the feasibility of utilizing YOLOv8 for cattle detection and counting under diverse environmental conditions, contributing to advancements in livestock monitoring in Mongolia. Specifically, for counting large herds of livestock, this research provides robust evidence supporting a novel method that leverages deep learning-based object detection for high-precision identification in drone-captured images and accurate total counting.

Model predictions in this study indicate that the MS COCO-based model effectively learned the features of well-fed cattle during the summer but struggled to detect leaner Mongolian cattle in winter and spring. In contrast, the custom-trained model achieved high accuracy in identifying cattle across these seasons. This finding highlights the ability of deep learning approaches to capture significant feature variations associated with seasonal changes in cattle. Additionally, it underscores the potential of computer vision and deep learning to analyze the distinct characteristics of livestock under varying environmental conditions, offering valuable insights and methodologies to advance livestock research in Mongolia.

Mongolia's distinct four seasons-characterized by snow-covered winters, golden autumns, and lush green summers-combined with its varied terrain of rolling grasslands, alpine forests, and hills, pose unique challenges for cattle detection. Currently, the dataset comprises images collected and labeled during the winter and spring seasons. Over the next year, the dataset will be expanded through the collection of drone imagery capturing diverse cattle populations across all seasons and terrains in Mongolia. This initiative aims to develop a comprehensive aerial dataset of Mongolian cattle, facilitating the training of a high-accuracy detection and counting model optimized for diverse environmental conditions and landscapes throughout the country.

REFERENCES

[1] Bailey, D., Trotter, M., Knight, C., Thomas, M. (2018). Use of GPS tracking collars and accelerometers for rangeland livestock production research. *Translational Animal Science*, Vol. 2, pp. 81–88, Oxford University Press US, https://doi.org/10.1093/tas/txx006.

[2] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580-587, CA/USA, https://doi.org/10.1109/CVPR.2014.81.

[3] Gonzalez, L.F., Montes, G.A., Puig, E., Johnson, S., Mengersen, K., Gaston, K.J. (2016). Unmanned Aerial Vehicles (UAVs) and Artificial Intelligence Revolutionizing Wildlife Monitoring and Conservation. *Sensors*, Vol. 16 (1), pp. 97, https://doi.org/10.3390/s16010097.

[4] He, K., Zhang, X., Ren, S., Sun, J., (2016). Deep residual learning for image recognition. *In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 770–778, CA/USA.

[5] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. *In Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961–2969, CA/USA.

[6] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *In Proceedings of the IEEE conference on computer vision and pattern recognition,* pp. 4700-4708, CA/USA.

[7] João Vitor de Andrade Porto., Fábio Prestes Cesar Rezende., Gilberto Astolfi., Vanessa Aparecida de Moraes Weber., Marcio Carneiro Brito Pache., Hemerson Pistori. (2021). Automatic counting of cattle with Faster R-CNN on UAV images. In *Workshop de Visão Computacional (WVC)*, pp. 1-6, SBC, Português/Brasil, https://doi.org/10.5753/wvc.2021.18880.

[8] Kumar, S., and S. K. Singh. (2018). Monitoring of pet animal in smart cities using animal biometrics. *Future Generation Computer Systems*, Vol. 83, pp. 553-563, doi.org/10.1016/j.future.2016.12.006.

[9] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, Vol. 521, pp. 436–444, https://doi.org/10.1038/nature14539.

[10] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single shot multibox detector. *In Computer Vision–ECCV 2016: 14th European Conference,* pp. 21-37, Springer International Publishing.

[11] Olson, D., & Anderson, J. (2021). Review on unmanned aerial vehicles, remote sensors, imagery processing, and their applications in agriculture. *Agronomy Journal*, Vol. 113, pp. 971-992, https://doi.org/10.1002/agj2.20595.

[12] Ocholla, I. A., Pellikka, P., Karanja, F. N., Vuorinne, I., Odipo, V., & Heiskanen, J. (2024). Livestock detection in African rangelands: Potential of high-resolution remote sensing data. *Remote Sensing Applications: Society and Environment*, Vol. 33, 101139, https://doi.org/10.1016/j.rsase.2024.101139.

[13] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster RCNN: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, Vol. 28, MA/USA.

[14] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788, CA/USA.

[15] Radovic, M., Adarkwa, O., & Wang, Q. (2017). Object recognition in aerial images using convolutional neural networks. *Journal of Imaging,* Vol. 3, pp. 21. https://doi.org/10.3390/jimaging3020021.

[16] Simonyan, K., Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *arXiv preprint,* arXiv:1409.1556, https://doi.org/10.48550/arXiv.1409.1556.

[17] Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. *In International conference on machine learning,* pp. 6105-6114, PMLR.

[18] Tsung-Yi Lin., Michael Maire., Serge Belongie., James Hays., Pietro Perona., Deva Ramanan., Piotr Dollár & C. Lawrence Zitnick. (2014). Microsoft coco: Common objects in context. In *Computer vision–ECCV 2014: 13th European conference,* Zurich, Switzerland, pp. 740-755, Springer International Publishing.

[19] Verma, G. K., & Gupta, P. (2018). Wild Animal Detection Using Deep Convolutional Neural Network. *In Proceedings of 2nd International Conference on Computer Vision & Image Processing*, Vol. 2, pp. 327-338, Springer Singapore.

[20] Windrim, L., Bryson, M., McLean, M., Randle, J., & Stone, C. (2019). Automated Mapping of Woody Debris over Harvested Forest Plantations Using UAVs, High-Resolution Imagery, and Machine Learning. *Remote Sensing,* Vol. 11, pp. 733, https://doi.org/10.3390/rs11060733.

[21] Wang, D. L., Liao, X. H., Zhang, Y. J., Cong, N., Ye, H. P., Shao, Q. Q., & Xin, X. P. (2021). Real-time detection and weight estimation of grassland livestock based on unmanned aircraft system video streams. *Chinese Journal of Ecology*, Vol. 40, pp. 4099, https://doi.org/10.13292/j.1000-4890.202111.008.

[22] Wang, Y., Ma, L., Wang, Q., Wang, N., Wang, D., Wang, X., Zheng, Q., Hou, X., Ouyang, G. (2023). A lightweight and high-accuracy deep learning method for grassland grazing livestock detection using UAV imagery. *Remote sensing*, Vol. 15, pp. 1593, https://doi.org/10.3390/rs15061593.

[23] Xu, B., Wang, W., Falzon, G., Kwan, P., Guo, L., Chen, G., ... & Schneider, D. (2020). Automated cattle counting using Mask R-CNN in quadcopter vision system. *Computers and Electronics in Agriculture*, Vol. 171, 105300, https://doi.org/10.1016/j.compag.2020.105300.

[24] Xu, B., Wang, W., Falzon, G., Kwan, P., Guo, L., Sun, Z., & Li, C. (2020). Livestock classification and counting in quadcopter aerial images using Mask R-CNN. *International Journal of Remote Sensing*, Vol. 41, pp. 8121-8142, United Kingdom, https://doi.org/10.1080/01431161.2020.1734245.

[25] *** National Statistics Office of Mongolia, (2024), The total number of livestock reached 64.7 million decreased by 9.1% from the previous year, *NSO Livestock census*, M.Oyunjargal, B.Altansuvd, Mongolia.