VIPER: Vision-based system to detect potential predators for herding with robots

Xiao Yang, Lidia Sánchez-González[0000−0002−0760−1170], and Francisco J. Rodríguez-Lera[0000−0002−8400−7079]

Departamento de Ingenierías Mecánica, Informática y Aeroespacial, Universidad de León, 24071, León, Spain
xiyang02@estudiantes.unileon.es,{lidia.sanchez,fjrodl}@unileon.es

Abstract. This paper presents a vision system that can be deployed in a ROS2 node that identifies potential predators such as wolves in order to herd autonomously a flock. A pre-trained YOLOv8 model is fine tunned so as to include more classes that are not supported such as wolves and the whole model is re-trained with the COCO dataset and the new one, keeping the recognition of the previous learned classes as well as the new ones. To do that a dataset of images of wolves in open spaces has been acquired and labelled for semantic segmentation purposes. The obtained vision model is wrapped in a ROS2 node, making it possible to implement it in a robotic device either a 4-legged robot or a rover robot. With this vision-based model, a robot can determine if a potential predator is detected and provide useful information to the herder to avoid area where the flock is in danger. Extensive livestock farming can benefit from the use of this technology to facilitate the tasks required on a daily basis.

Keywords: Precision Livestock Farming · Computer vision · Herding with robots · Predator detection

1 Introduction

In recent years, with the advancement of technology and the rapid development of artificial intelligence, many technologies have been applied in various fields, including agriculture [23]. Animal monitoring and management, which is the focus of this article, have particularly benefited from these advancements. For instance, drones equipped with high-definition cameras are useful for aerial surveillance [15], and infrastructures like walk-over weighing (WoW) systems [5] enable the monitoring of weight gain. Additionally, GPS collars can localize animals in the wild, creating virtual fences [7], among other applications.

Given that the benefits are often limited, it is crucial to leverage existing intelligent systems to enhance efficiency [24], make decisions based on valuable data [13,17,25], and contribute to the sustainability of farms [2].

Furthermore, these technological advancements have enabled early disease detection and real-time health monitoring through posture and behavioral analysis [3]. Currently, image processing is employed by automated feeding systems to customize portions based on user requirements, thereby reducing waste and maximizing growth [21]. Additionally, breeding and intelligent monitoring are enhanced through genetic data analysis, improving the sustainability and efficiency of farming practices [14].

Several technologies, such as wearable devices like accelerometers, are available for real-time monitoring of sheep activities [10]. However, these methods are costly, labor-intensive, and pose the risk of harming the sheep. Unquestionably, the use of Unmanned Aerial Vehicles (UAVs) [4,20] equipped with image processing technology seems the most suitable method although their noise can produce stress in the animals [1]. Autonomous UAVs can efficiently collect and analyze data, detect real-time anomalies, and notify the shepherd. The data collected by these UAVs can further aid in optimizing AI models to enhance animal detection, behavioral analysis, and disease detection.

The state-of-the-art YOLO model is employed due to its speed, making it suitable for real-time tasks and surpassing previous architectures [11]. Additionally, the YOLO model can be deployed in a robot with ROS2-compatible hardware [16], as there are existing solutions that can be integrated with the de facto standard in robotics, ROS2 [6]. The fine-tuning technique [22] is applied to tailor the dataset specifically for the herding task while retaining the model's existing knowledge.

The following sections detail the proposed pipeline. Section 2 describes the dataset collected for the herding task. Section 3 explains the model implementation. The obtained results are presented in Section 4. Finally, Section 6 summarizes the conclusions drawn from this study.

2 Data acquisition

Despite the fact that YOLO was trained with the COCO dataset, which contains many samples, the number of classes is limited to just 80 and does not include objects useful for herding, such as wolves. Therefore, it was necessary to create a dataset that includes this potential predator to enable the detection of possible threats to the flock.

To build this specialized dataset for the herding task, the first step involves recording videos of four classes: sheep, wolves, dogs, and people. These recordings are made from various viewpoints to create an enriched dataset, including close views, distant views, and indoor and outdoor settings. Additionally, animals with different coat colors are included to ensure diversity. Some samples are shown in Figure 1.

In total, we processed 38 videos of varying lengths, including 27 videos featuring different samples of sheep, dogs, and people, and 11 videos of wolves and people. By capturing 30 frames per second, we obtained over 80,000 images. After the segmentation annotation process, we have a total of 67,207 annotated images to build our database.

The dataset's annotation process utilized the powerful *X-Anylabeling* annotation tool [26], which combines state-of-the-art (SOTA) models with traditional

Fig. 1. Samples of the images were acquired from different viewpoints: close view (top left), distant view (top right), indoor (bottom left), and outdoor (bottom right)

labeling methods such as *Labelme* [19]. Following this, a manual review process was conducted to check all the annotations and correct them individually.

Currently, this vision dataset is published on the Hugging Face platform [8]. It is publicly available for anyone interested in utilizing it for research or development purposes.

3 Vision-model for animal segmentation

Comparing the YOLO models [18,9] to other vision models based on convolutional layers, YOLO stands out for its high performance, adaptability, and user-friendliness, making it one of the most efficient architectures in computer vision. Its main tasks include object detection, classification, tracking, and instance segmentation. We employed the segmentation technique for our dataset, which provides more detailed information than simple animal detection by applying a mask to each animal in addition to identifying its number and position. We used YOLOv8 and YOLOv9 in practice because the YOLOv10 segmentation version, although published, is not yet available.

To adapt the YOLO model, it is common to employ transfer learning, a technique that conserves computational resources and time by using a pretrained model. YOLO is pretrained on the COCO dataset [12]. The YOLOv8 model offers five different configurations based on the number of parameters: nano

(YOLOv8n), small (YOLOv8s), medium (YOLOv8m), large (YOLOv8l), and extra-large (YOLOv8x). Similarly, the YOLOv9 model provides five configurations, with only two supporting the segmentation task: large (YOLOv9c) and extra-large (YOLOv9e). As the number of model parameters increases, the performance in recognizing objects improves, but this comes at the cost of reduced training and inference speed. In this work, we use both a small model and a large model to provide flexibility depending on the hardware capabilities of the robot. Specifically, we use the YOLOv8s model and the YOLOv9c model.

4 Experimental results

For our experiments, we began the training process by randomly dividing the images into 70% for training, 20% for validation, and 10% for testing. Using the recommended hyperparameters for YOLO, we employed the SGD optimizer with a learning rate of 0.01, an image input size of 640×640, and an automatic batch size (14 for YOLOv9c, and 103 for YOLOv8s). Training was conducted over 50 epochs with an EarlyStopping function set to 10 epochs, which stops the training if performance does not improve for 10 consecutive epochs.

The training results are presented in Table 1 and the test results are summarized in Table 2. YOLOv9 obtains slightly higher results; however, it is important to note that these results are obtained for a dataset limited to the four aforementioned classes. After this first training process we observed that the model had lost recognition of 77 other COCO categories and now only recognizes sheep, people, dogs, and wolves, which was not the expected outcome.

Table 1. Results obtained for the validation set by YOLOv8 and YOLOv9

Model	Validation Set							
	Bounding Box				Semantic Mask			
	Precision	Recall	mAP50	mAP50-95	Precision	Recall	mAP50	mAP50-95
YOLOv8s	0.903	0.899	0.941	0.716	0.896	0.883	0.938	0.658
YOLOv9c	0.903	0.899	0.944	0.725	0.896	0.887	0.939	0.669

Table 2. Results obtained for the test set by YOLOv8 and YOLOv9

Model	Test Set							
	Bounding Box				Semantic Mask			
	Precision	Recall	mAP50	mAP50-95	Precision	Recall	mAP50	mAP50-95
YOLOv8s	0.903	0.913	0.947	0.724	0.901	0.880	0.935	0.638
YOLOv9c	0.904	0.913	0.949	0.732	0.899	0.889	0.936	0.654

We initially attempted to freeze 20 layers of the model while only adjusting the parameters of the final model output layer, but the results were deemed unacceptable. To address this issue, we augmented our dataset by incorporating 5000 COCO images and subsequently froze 10 layers of the model. During this training process, we organized the images into videos, with 29 videos (consisting of 55,734 frames, including 4000 COCO images) used for training, and 9 videos (comprising 11,473 frames where 1000 are from COCO dataset) used for validation. The training hyperparameters remained consistent with those used in the initial training phase. The results of this approach are presented in Table 3. As we can observe, this approach keeps the performance of the official model or is a bit less precise but they recognize properly the desire classes.

Table 3. Results obtained by YOLOv8 and YOLOv9 with the validation set of the dataset augmented with COCO

Model	Validation Set							
	Bounding Box				Semantic Mask			
	Precision	Recall	mAP50	mAP50-95	Precision	Recall	mAP50	mAP50-95
YOLOv8s	0.544	0.395	0.417	0.284	0.525	0.384	0.399	0.249
YOLOv9c	0.695	0.558	0.61	0.453	0.68	0.545	0.588	0.387
YOLOv8s-official				0.446				0.368

5 Discussion

Upon analyzing the results obtained in the previous section, it is evident that the outcomes for both the validation and test data are nearly identical, suggesting that the model has generalized well for this data type. The results are highly promising when using our dataset, demonstrating high accuracy in both bounding box determination and mask generation for each animal.

Furthermore, after incorporating images from the COCO dataset, YOLOv9 yields significantly better results than YOLOv8 due to its increased number of parameters and overall improvements. Although the metrics deteriorated upon enlarging the dataset with COCO images, it is understandable considering that YOLOv8, according to official documentation, achieves a mean Average Precision (mAP50-95) of 44.6% for bounding boxes and 36.8% for masks. Our primary objective is to maintain recognition accuracy for other classes while focusing on wolves and sheep for herding tasks. Thus, we consider the obtained results satisfactory. The results for these two classes are detailed in Table 4.

An additional experiment was conducted with the retrained models to assess its performance in detecting animals, and other classes like fruits or vehicles. The models demonstrated a high capacity to segment sheep, people, fruits, and vehicles, as well as wolves. However, it occasionally confused dogs with sheep due to their similar characteristics. Nonetheless, the model proved effective in warning shepherds when wolves approached the herd during herding events. Examples of inference results from some images are depicted in Figure 2.

Table 4. Results of validation sets for wolf and sheep classes obtained by YOLOv8 and YOLOv9

Model		Validation Set							
		Box				Mask			
		Precision	Recall	mAP50	mAP50-95	Precision	Recall	mAP50	mAP50-95
YOLOv8s	sheep	0.77	0.619	0.743	0.459	0.768	0.613	0.736	0.427
	wolf	0.853	0.67	0.798	0.569	0.855	0.669	0.799	0.548
YOLOv9c	sheep	0.821	0.695	0.816	0.548	0.816	0.687	0.808	0.51
	wolf	0.915	0.805	0.906	0.69	0.917	0.806	0.907	0.664

Fig. 2. Example of outputs detecting wolves, sheep, vehicles and fruits

In addition to the experiment we can look at the confusion matrix 3 to see how each class behaves individually; the resulting diagonal matrix indicates that most of the predictions of each class were accurate.

Finally, the retrained models are deployed to a node of The Robot Operating System 2 (ROS2) using the 'YOLOv8_ros' library published on GitHub [6]. It is capable of being used on many robots with camera sensors. All code is available at `https://github.com/shepherd-robot/VIPER`.

6 Conclusions

In conclusion, this study has successfully curated a publicly available dataset comprising images of sheep and wolves, which has been published on Hugging Face. The primary objective of this dataset is to streamline the automation

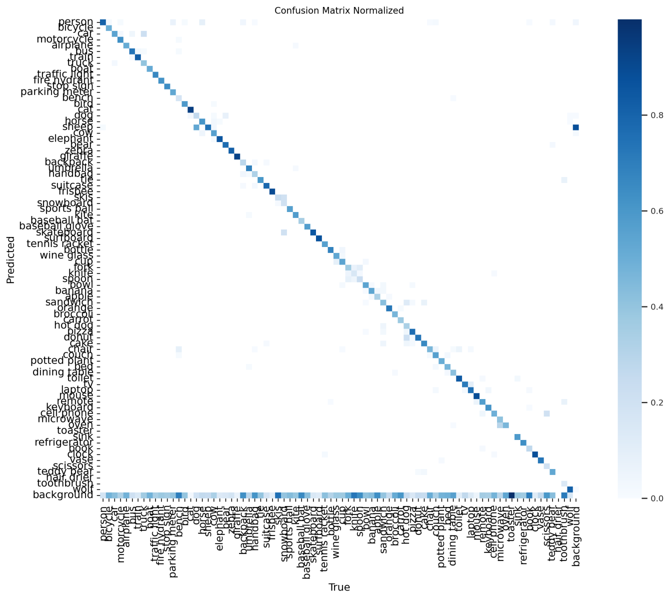

Fig. 3. Normalized confusion matrix obtained by the YOLOv9 model

of herding tasks using robots. Utilizing convolutional neural network models YOLOv8 and YOLOv9, we have implemented an animal recognition approach to detect and segment animals within the images. Fine-tuning techniques were employed to customize the models for the new dataset, integrating data from the COCO dataset while preserving the models' prior knowledge by freezing half of the layers. The obtained results underscore the effectiveness of the proposed system in accurately detecting and segmenting sheep and wolves.

This technological advancement paves the way for the development of versatile systems deployable on various robotic platforms, such as quadrupedal robots or rovers, utilizing a ROS2 node. By harnessing such capabilities, autonomous herding technologies can be further enhanced, facilitating the identification of potential predators like wolves and enabling precise localization of the herd.

Acknowledgements

We gratefully acknowledge the financial support of Grant TED2021-132356B-I00 funded by MCIN/AEI/10.13039/501100011033 and by the "European Union NextGenerationEU/PRTR".

References

1. Abdulai, G., Sama, M., Jackson, J.: A preliminary study of the physiological and behavioral response of beef cattle to unmanned aerial vehicles (uavs). Ap-

plied Animal Behaviour Science **241**, 105355 (2021). https://doi.org/10.1016/j.applanim.2021.105355
2. Aubert, B.A., Schroeder, A., Grimaudo, J.: IT as enabler of sustainable farming: An empirical analysis of farmers' adoption decision of precision agriculture technology. Decision Support Systems **54**(1), 510–520 (dec 2012). https://doi.org/10.1016/J.DSS.2012.07.002
3. Džermeikaitė, K., Bačėninaitė, D., Antanaitis, R.: Innovations in cattle farming: Application of innovative technologies and sensors in the diagnosis of diseases. Animals **13**(5) (2023). https://doi.org/10.3390/ani13050780
4. Gonzalez, L.F., Montes, G.A., Puig, E., Johnson, S., Mengersen, K., Gaston, K.J.: Unmanned aerial vehicles (uavs) and artificial intelligence revolutionizing wildlife monitoring and conservation. Sensors **16**(1) (2016). https://doi.org/10.3390/s16010097
5. González-García, E., Alhamada, M., Pradel, J., Douls, S., Parisot, S., Bocquier, F., Menassol, J., Llach, I., González, L.: A mobile and automated walk-over-weighing system for a close and remote monitoring of liveweight in sheep. Computers and Electronics in Agriculture **153**, 226–238 (2018). https://doi.org/https://doi.org/10.1016/j.compag.2018.08.022
6. Ángel González-Santamarta, M.: ROS2 wrap for YOLO, https://github.com/mgonzs13/yolov8_ros, Robotics Group, Universidad de León
7. Herlin, A., Brunberg, E., Hultgren, J., Högberg, N., Rydberg, A., Skarin, A.: Animal welfare implications of digital tools for monitoring and management of cattle and sheep on pasture. Animals **11**(3) (2021). https://doi.org/10.3390/ani11030829
8. Jain, S.M.: Hugging face. In: Introduction to transformers for NLP: With the hugging face library and models to solve problems, pp. 51–67. Springer (2022)
9. Jocher, G., Chaurasia, A., Qiu, J.: Ultralytics yolov8 (2023), https://github.com/ultralytics/ultralytics
10. Kleanthous, N., Hussain, A., Khan, W., Sneddon, J., Liatsis, P.: Deep transfer learning in sheep activity recognition using accelerometer data. Expert Systems with Applications **207**, 117925 (2022). https://doi.org/10.1016/j.eswa.2022.117925
11. Laroca, R., Severo, E., Zanlorensi, L.A., Oliveira, L.S., Gonçalves, G.R., Schwartz, W.R., Menotti, D.: A robust real-time automatic license plate recognition based on the yolo detector. In: 2018 International Joint Conference on Neural Networks (IJCNN). pp. 1–10. IEEE (2018)
12. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft COCO: Common objects in context. In: Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. pp. 740–755. Springer (2014)
13. Niloofar, P., Francis, D.P., Lazarova-Molnar, S., Vulpe, A., Vochin, M.C., Suciu, G., Balanescu, M., Anestis, V., Bartzanas, T.: Data-driven decision support in livestock farming for improved animal health, welfare and greenhouse gas emissions: Overview and challenges. Computers and Electronics in Agriculture **190**, 106406 (2021). https://doi.org/10.1016/j.compag.2021.106406
14. Pandey, D.K., Mishra, R.: Towards sustainable agriculture: Harnessing ai for global food security. Artificial Intelligence in Agriculture **12**, 72–84 (2024). https://doi.org/10.1016/j.aiia.2024.04.003
15. Puri, V., Nayyar, A., Raja, L.: Agriculture drones: A modern breakthrough in precision agriculture. Journal of Statistics and Management Systems **20**(4), 507–518 (2017)

16. Quigley, M., Gerkey, B., Smart, W.D.: Programming Robots with ROS: a practical introduction to the Robot Operating System. O'Reilly Media, Inc. (2015)
17. Recio, B., Rubio, F., Criado, J.A.: A decision support system for farm planning using AgriSupport II. Decision Support Systems **36**(2), 189–203 (2003). https://doi.org/10.1016/S0167-9236(02)00134-3
18. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection (2016)
19. Russell, B.C., Torralba, A., Murphy, K.P., Freeman, W.T.: Labelme: a database and web-based tool for image annotation. International journal of computer vision **77**, 157–173 (2008)
20. Schad, L., Fischer, J.: Opportunities and risks in the use of drones for studying animal behaviour. Methods in Ecology and Evolution **14**(8), 1864–1872 (2023). https://doi.org/10.1111/2041-210X.13922
21. Sinnott, A.M., Kennedy, E., Bokkers, E.: The effects of manual and automated milk feeding methods on group-housed calf health, behaviour, growth and labour. Livestock Science **244**, 104343 (2021). https://doi.org/https://doi.org/10.1016/j.livsci.2020.104343
22. Tajbakhsh, N., Shin, J.Y., Gurudu, S.R., Hurst, R.T., Kendall, C.B., Gotway, M.B., Liang, J.: Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE transactions on medical imaging **35**(5), 1299–1312 (2016)
23. Talaviya, T., Shah, D., Patel, N., Yagnik, H., Shah, M.: Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides. Artificial Intelligence in Agriculture **4**, 58–73 (2020)
24. Tullo, E., Finzi, A., Guarino, M.: Review: Environmental impact of livestock farming and Precision Livestock Farming as a mitigation strategy (feb 2019). https://doi.org/10.1016/j.scitotenv.2018.10.018
25. Van Meensel, J., Lauwers, L., Kempen, I., Dessein, J., Van Huylenbroeck, G.: Effect of a participatory approach on the successful development of agricultural decision support systems: The case of Pigs2win. Decision Support Systems **54**(1), 164–172 (dec 2012). https://doi.org/10.1016/J.DSS.2012.05.002
26. Wang, W.: Advanced auto labeling solution with added features. https://github.com/CVHub520/X-AnyLabeling (2023)