Improved Yolov8 Algorithm for Cow Recognition Based on Soft NMS

Xiaoyang Liu, Zeyu Han

National Agricultural University, Yantai, Shandong 100083

Abstract: ***Currently, artificial intelligence and IoT technology are widely used in the agricultural field, and China's smart agriculture is steadily developing. The introduction of computer vision technology is gradually freeing farms and other animal recognition systems from the reliance on sensors in traditional monitoring systems. However, animal recognition in dense scenes has the characteristics of small space, large quantity, large individual volume, and serious occlusion and adhesion problems, making it difficult to accurately identify every animal in the animal population. This article uses the YOLOv8 model to implement cow recognition and replaces traditional non maximum suppression algorithms with the Soft NMS algorithm. By comparison, the individual recognition of large animals with severe occlusion has been solved, laying the foundation for the next stage of animal behavior recognition.***

Keywords: Object detection; Non maximum suppression algorithm; Yolo.

1. INTRODUCTION

1.1 Background and significance

With the popularization of agricultural IoT technology, it has accelerated the integration of agriculture, animal husbandry, and artificial intelligence technology in China. Using computer systems to assist farmers and ranchers in real-time detection and anomaly warning of animals and plants can greatly liberate labor and enable faster and better completion of traditional regulatory tasks. Real time monitoring and early warning of the behavior and physical condition of animals such as cattle and sheep in animal husbandry can achieve faster and higher quality production of meat and other related products.

For a long time, the regulation of animals in China's animal husbandry industry has relied on sensors, and a common method is to wear sensors that collect temperature, three-axis acceleration, and angular velocity on the neck or legs of animals. By receiving sensor data, the current animal's physical state and behavior can be determined. Especially in order to obtain more detailed data, some researchers inject sensors into animals, which to some extent achieves control over each animal. However, wearing or injecting sensors in animals themselves is a limitation on animal behavior, especially when injecting sensors often causes local inflammation that cannot be treated. Therefore, it is necessary to find better methods to complete such tasks.

In recent years, the development of computer vision technology has made rapid progress. In just ten years, researchers have gone from only 60% accuracy on ImageNet to achieving extremely high accuracy in tasks such as object detection, behavior recognition, object segmentation, and image generation. Thanks to this, China's smart agriculture is developing towards the integration of artificial intelligence and IoT technology. Installing high-precision cameras that integrate environmental parameter sensors and are equipped with deep learning models and processors in agricultural and pastoral areas can perfectly replace sensor methods and solve a series of problems associated with sensor methods. This article is based on the improvement of the computer vision model YOLOv8, making it more suitable for small space, large quantity, large volume, and high occlusion cow recognition tasks.

1.2 Literature Review

The object detection of deep learning methods can automatically complete feature extraction tasks in the first few layers of classification model training. The current common object detection task models are divided into two methods: one-stage (mainly YOLO and SSD) and two-stage (mainly Fast RCNN). Guo Yangyang integrated spatial attention mechanism with YOLOv4 model and proposed YOLOv4-SAM deep learning network. Based on extracting multi-scale features of cows, it highlights relevant features such as biological features, improves the representation ability of biological vision, and achieves high-precision detection of individual cows. Feng Ce proposed to define feeding behavior through the interaction between pig heads and feeding troughs in his research on pig feeding detection, and established a pig head recognition model using VGG16 to achieve accurate recognition of pig heads in structured scenes. WANG R et al. replaced the Conv of YOLOv5 with GhostConv and integrated a lightweight attention mechanism into the YOLOv5 model. Using YOLOv5s pre training, they achieved accurate recognition of cow riding behavior. Diao et al. [1] optimized Bi-LSTM networks for lung cancer detection, achieving improved diagnostic accuracy through enhanced network architecture. In urban analytics, Li et al. [2] proposed a user-centered framework for interactive smart city data exploration, emphasizing human-computer interaction in decision-making. Industrial applications have also benefited, as Zhao et al. [3] developed a deep learning-based approach to optimize steel production scheduling, reducing operational inefficiencies. Meanwhile, Yang et al. [4] introduced a big data-driven AI model for economic cycle prediction, offering improved forecasting precision. Transportation systems have seen similar innovations, with Tu [5] presenting a reliable vehicle platooning system using 5G link aggregation for smart road infrastructure. Computer vision research has progressed notably, exemplified by Ding et al. [6], who proposed a novel attention mechanism for clothing-changing person re-identification, addressing key challenges in feature decoupling and multimodal fusion. In biochemistry, Zhuang et al. [7] developed phosphinic acid-based inhibitors targeting tubulin polyglycylation, opening new therapeutic possibilities. Autonomous systems have advanced through Zhou et al. [8], who implemented LSTM-driven path planning for UAVs, enhancing navigation efficiency. Mechanical engineering applications include Wang et al. [9], who utilized machine learning for fatigue life evaluation of pump spindle assemblies with parametrized geometry. Finally, Gong et al. [10] applied machine learning to predict extreme financial market volatility using unstructured data, demonstrating cross-domain adaptability.

2. OVERVIEW OF RELATED TECHNOLOGIES

2.1 Livestock recognition framework based on machine vision

The livestock recognition technology based on machine vision has slightly different data collection devices, algorithms, and targets used in various studies, but its overall research framework is basically similar..

Firstly, there is data collection and annotation. By collecting video or image data of various livestock in different scenarios and using tools such as Labelimg or cvat to annotate them in YOLO format. Further divide the training set and validation set to form a trainable dataset.

Then there is the import and modification of the model. As object detection tasks become increasingly difficult, traditional models such as convolutional neural networks are unable to achieve good detection results. Therefore, most researchers are using YOLO series models to complete object detection tasks. Modify the native YOLO model to better adapt to specific tasks, taking into account the specificity of individual tasks for researchers.

Finally, after the model training and testing are completed, the model is loaded onto the camera to complete the subsequent tasks.

2.2 Overview of Related Technologies

2.2.1 Yolo series models and Yolov8

The YOLO (You Only Look Once) algorithm series is designed to handle real-time object detection tasks. In April 2023, Ultrlytics released the YOLOv8 model, representing a major breakthrough in the YOLO series of models. The YOLOv8 model performs well on various datasets and application fields, and can be adjusted slightly to meet specific task requirements. As shown in Figure 1, it is the YOLOv8 structural diagram. It can be seen that YOLOv8 replaces the previous coupling head with a decoupling head when performing different tasks. For different tasks, pass through different convolution modules before the final step.

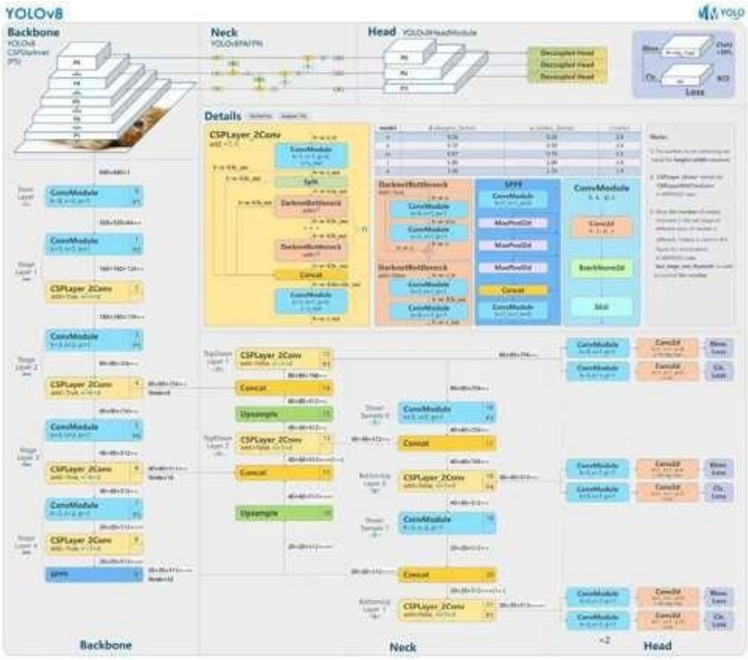

Figure 1: YOLOv8 model structure diagram

2.2.2 IOU

IoU(Intersection over Union), Compare and contrast. For two overlapping boxed regions, IoU is the result obtained by dividing the overlapping portion of the two regions by the union of the two regions. The calculation method is shown in Figure 2. Generally speaking, a score greater than 0.5 can be considered a good result.

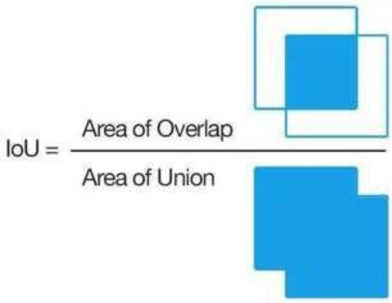

Figure 2: IOU (Intersection over Union) Diagram

2.2.3 NMS Algorithm

In object detection tasks, it is easy to miss detection due to object occlusion and adhesion. As shown in Figure 3, when locating an object, the algorithm finds many boxes and needs to retain the best box. Non maximum suppression (NMS) is the process of finding the optimal preservation among multiple annotated boxes for the target and calculating the intersection to union ratio.

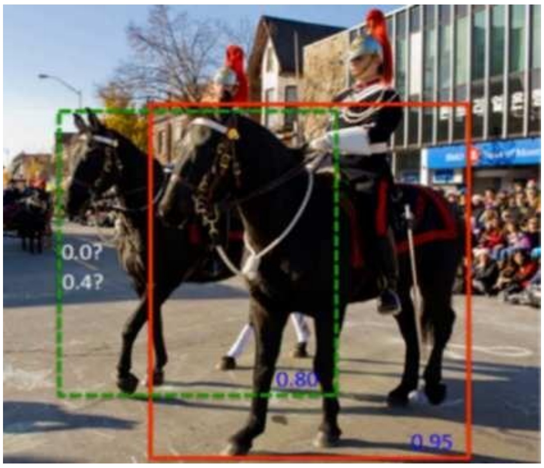

Figure 3: Example of Traditional NMS Algorithm

3. DATASET INTRODUCTION

The cow dataset used in this article was collected from various scenarios. This dataset only manually annotated the cow category, and the annotation file saved the number of categories in the corresponding image (labeled as 0, cow category), as well as the coordinates of the upper left and lower right diagonal points of each individual in this category in the annotation box. By using the coordinates of these two points, a rectangular box can be determined. The dairy cow detection and recognition dataset used in this experiment includes 3250 images in the training set and 860 images in the validation set. Some of the selected data, sample datasets, and annotations are shown in Figure 3. The comparative experiments in this article will also be based on examples. YOLOv8 can accept image data of any size and adjust the received image data to a size of 640x640 within the model, while maintaining the original aspect ratio. However, when using YOLOv5, we need to manually adjust the image before sending it into the model. In addition, in order to enhance the generalization ability and robustness of the model, this paper uses data augmentation techniques such as rotation, scaling, and color transformation to expand the dataset and reduce the risk of overfitting.

Figure 3: Example of Traditional NMS Algorithm

4. DATASET INTRODUCTION

The cow dataset used in this article was collected from various scenarios. This dataset only manually annotated the cow category, and the annotation file saved the number of categories in the corresponding image (labeled as 0, cow category), as well as the coordinates of the upper left and lower right diagonal points of each individual in this category in the annotation box. By using the coordinates of these two points, a rectangular box can be determined. The dairy cow detection and recognition dataset used in this experiment includes 3250 images in the training set and 860 images in the validation set. Some of the selected data, sample datasets, and annotations are shown in Figure 3. The comparative experiments in this article will also be based on examples. YOLOv8 can accept image data of any size and adjust the received image data to a size of 640x640 within the model, while maintaining the original aspect ratio. However, when using YOLOv5, we need to manually adjust the image before sending it into the model. In addition, in order to enhance the generalization ability and robustness of the model, this paper uses data augmentation techniques such as rotation, scaling, and color transformation to expand the dataset and reduce the risk of overfitting.

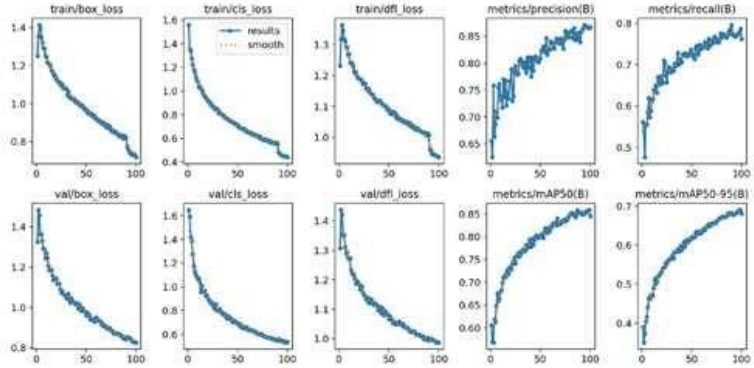

Figure 5: Loss and accuracy on the training and validation sets during the training process

Intuitively speaking, this article uses a trained model to make predictions on the validation set shown in Figure 3. As shown in Figure 6, it can be clearly seen that the occluded targets were still well labeled even in the presence of dense cows. Overall, The YOLOv8 model performs well on the dataset, with high detection accuracy and robustness

Figure 6: Prediction results on the labeled validation set in Figure 3

5. CONCLUSION AND OUTLOOK

In summary, introducing the Soft NMS algorithm into the YOLOv8 model has greatly improved the recognition of images with severe occlusion and dense cow classes. In some images, the detection of small and adhesive targets is poor, and the overall single image detection time of the model is 0.17-0.22 seconds, which is relatively slow for most frame rate cameras at present. This is largely related to the YOLOv8 model. So the follow-up work can be carried out by adding a lightweight attention mechanism and replacing the convolution module with a lightweight convolution module.

REFERENCES

[1] Diao, Su, et al. "Optimizing Bi-LSTM networks for improved lung cancer detection accuracy." PloS one 20.2 (2025): e0316136.
[2] X. Li, L. Evans, and X. Zhang, “Interactive data exploration for smart city analytics: A user-centered perspective,” 01 2025.
[3] Zhao, H., Chen, Y., Dang, B., & Jian, X. (2024). Research on Steel Production Scheduling Optimization Based on Deep Learning.
[4] Yang, W., Zhang, B., & Wang, J. (2025). Research on AI Economic Cycle Prediction Method Based on Big Data.
[5] Tu, Tongwei. "Reliable Vehicle Platooning via Redundant 5G Link Aggregation in Smart Roads." (2025).
[6] Ding, Y., Wang, X., Yuan, H., Qu, M., & Jian, X. (2025). Decoupling feature-driven and multimodal fusion attention for clothing-changing person re-identification. Artificial Intelligence Review, 58(8), 1-26.

[7] Zhuang, Zaile, et al. "Phosphinic acid-based inhibitors of tubulin polyglycylation." Chemical Communications 58.45 (2022): 6530-6533.
[8] Zhou, Dianyi, et al. "Research on LSTM-driven UAV path planning." Fourth International Conference on Advanced Algorithms and Neural Networks (AANN 2024). Vol. 13416. SPIE, 2024.
[9] Wang, Lizhe, et al. "Machine Learning-Based Fatigue Life Evaluation of the Pump Spindle Assembly With Parametrized Geometry." ASME International Mechanical Engineering Congress and Exposition. Vol. 87684. American Society of Mechanical Engineers, 2023.
[10] Gong, Chenwei, et al. "Application of Machine Learning in Predicting Extreme Volatility in Financial Markets: Based on Unstructured Data." The 1st International scientific and practical conference "Technologies for improving old methods, theories and hypotheses"(January 07–10, 2025) Sofia, Bulgaria. International Science Group. 2025. 405 p.. 2025.