IRSTI 28.23.25 https://doi.org/10.26577/jpcsit2024020106

A. Yeshmukhametov

Nazarbayev University, Astana, Kazakhstan
e-mail: azamat.yeshmukhametov@nu.edu.kz

UNLOCKING AGRICULTURAL AUTOMATION: INTEGRATING SLAM AND COMPUTER VISION FOR LIVESTOCK MANAGEMENT

Abstract. The use of autonomous livestock detection is crucial in modern agriculture, providing efficient control and management of animals. This article explores the use of the SLAM (Simultaneous Localization and Mapping) algorithm in conjunction with computer vision to address various challenges in enhancing the capabilities of autonomous robots and detecting livestock. Integrating computer vision and SLAM technology allows autonomous robots to successfully navigate complex conditions, adapt to dynamic environments, and accurately determine the location of livestock in real-time. This research also presents a method for simultaneously estimating the agent's position in space and mapping the surrounding environment. This approach enables robots to adapt to different lighting and weather conditions, ensuring reliable operation in various agricultural environments. Computer vision enables autonomous robots to accurately detect livestock based on visual data, enabling them to effectively monitor and manage animals. We discuss various issues that can be addressed using this combination of technologies, including navigation in unknown or changing environments, creating three-dimensional models of the surrounding environment, as well as autonomous control of robots and unmanned vehicles. This article also provides an overview of existing approaches and techniques used to address these issues, evaluating their advantages and limitations. In conclusion, we discuss the prospects for the development of this field and potential directions for future research.

Key words: simultaneous localization and mapping, computer vision, agriculture, modeling and simulation.

1 Introduction

In the modern world, robotics and automation are rapidly advancing, offering unique opportunities for innovative solutions. Mobile robots are becoming an integral part of everyday life, penetrating various fields such as home cleaning, delivery, medical, and military services. A key aspect of their operation is the ability to autonomously navigate in the surrounding environment, emphasizing the importance of developing efficient SLAM methods and computer vision.

Upon reviewing research [1], it presents a proposition of an innovative mapping and navigation system in farmland, based on computer vision and Internet of Things (IoT) technologies. The system consists of three levels of subsystems: robotic vehicles, edge computing nodes, and a cloud server. The application of the Mesh-SLAM algorithm allows for rapidly and effectively creating a three-dimensional map of the farm with high accuracy. Experimental results demonstrate improved performance when using a distributed IoT architecture compared to a centralized cloud approach. In conclusion, the significance of this approach for implementing autonomous agricultural systems in real-world conditions is emphasized.

The article [2] describes the development and proposes an innovative agricultural robot called FaRo (FArming RObot), designed for autonomous cultivation of crops without human intervention. It highlights the importance of agriculture in meeting the needs of humanity, especially in the face of global population growth and a decrease in the number of agricultural workers. The main difference between FaRo and other agricultural platforms lies in its ability to autonomously perform the cultivation process from seeding to harvesting. Special attention is given to the harvesting tool, which is also explained and demonstrated. Overall, the article emphasizes the significance of FaRo in addressing the challenges of modern agriculture and its potential to revolutionize farming practices through automation and autonomy.

The result of the research [3] is the development and implementation of an autonomous robot with a weed detection system, which allows for effective weed control without damaging crops, as well as addressing the issue of labor shortage in agriculture, which can lead to increased productivity and improved sustainability of agricultural production.

The work [4] presents the optimization of the You Only Look Once (YOLO) object detection algorithm for detecting traffic lights and controlling the steering angle for automatic lane keeping in a simulated environment. The research compares the performance of YOLO models for traffic light detection and proposes a convolutional neural network (CNN) for steering angle control. The proposed algorithms are tested on an autonomous model in the simulated environment of Gazebo, ROS.

A systematic review [5] was conducted on the application of computer vision systems based on convolutional neural networks (CNNs) in animal husbandry. Five computer vision tasks were considered: image classification, object detection, semantic/instance segmentation, pose estimation, and tracking. The preparation of the system, selection of CNN architectures, algorithm development strategies, and performance evaluation of models were analyzed. The applications of these systems in animal husbandry were discussed, and future research directions were proposed, including the development of lightweight mobile systems, creation of specialized datasets, and advancement of CNN architectures for pose estimation, anomaly detection, and animal condition assessment.

The implementation [6] of robotic milking on dairy farms was discussed. It is noted that robotic milking reduces labor costs on farms of all sizes and offers a more flexible lifestyle for families milking up to 250 cows. Various aspects of barn layout and herd management were discussed, including the impact on milking frequency, prevention of lameness, cow routing, and grouping. It is suggested that free cow movement may be preferable with good management. The importance of lameness prevention is also emphasized, along with the need for further research in this area.

2 SLAM

SLAM (Simultaneous Localization and Mapping) is an actively researched area in robotics. It updates maps in unknown environments while retaining information about the robot's location[7]. Challenges include sensor cumulative errors, data matching complexities, and environmental changes. The algorithm involves scanning, data matching, map updating, and resampling to create an accurate environment representation (detailed in Figure 1). Termination conditions include failed scanning or reaching the endpoint.

In work [8], a review of SLAM technology, particularly its visual version (vSLAM), which uses cameras for motion estimation and map building, was proposed. The article provided an overview of recent vSLAM algorithms on technical and historical aspects. Algorithms were categorized based on image processing approaches: feature-based, direct, and using RGB-D cameras.

In research [9], a new SLAM framework based on a combination of inexpensive LiDAR and video sensor was proposed. The framework introduced a new cost function considering both scanning data and images, and applied a Bag of Words model with visual features for loop closure detection. A new way of representing a 2.5D map was introduced, displaying both obstacles and visual features, along with a fast map transition method. The results of the proposed method demonstrated better performance compared to using only LiDAR or only cameras, and the transition speed using the 2.5D map was significantly higher than using a traditional grid map.

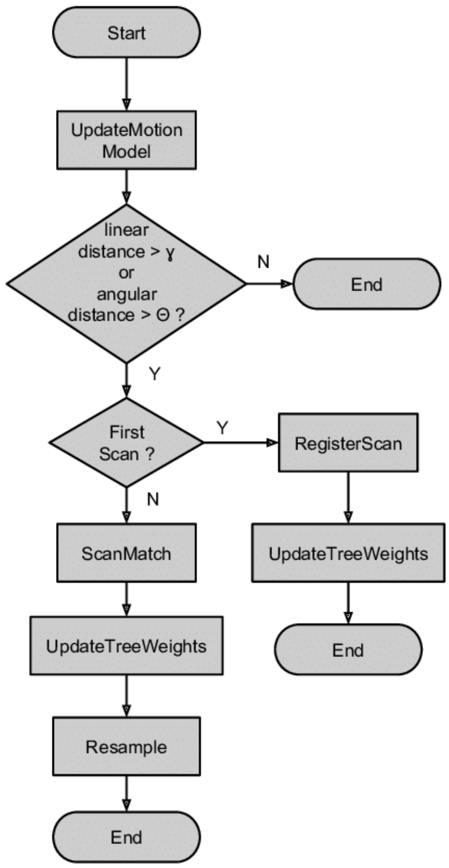

Figure 1 – Block diagram of SLAM algorithm execution

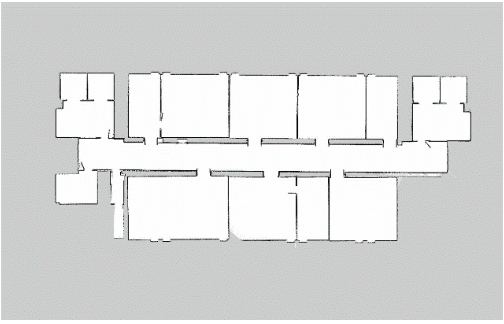

Figure 2 – Data collection about the terrain

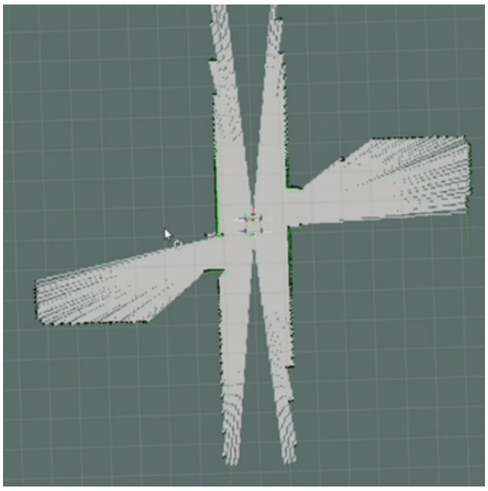

Figure 3 – Final map result

Testing of the SLAM system on mobile robots in an indoor environment using ROS was conducted [10]. The experiments were carried out in the USAR environment created in the Gazebo simulator, using the gmapping, karto SLAM, and hector SLAM algorithms for room mapping. The results of the experiments showed that the SLAM system on mobile robots is feasible, and high-precision maps can be created.

3. Creating a virtual model of the surrounding environment in Gazebo and ROS

Management of 3D objects is implemented in the Gazebo environment. Gazebo is a real-time robot simulation and modeling environment that provides a convenient toolkit for creating virtual environments for testing and developing various robotic applications [11]. Gazebo is widely used in academic and industrial fields for testing control algorithms, navigation, visualization, and other aspects of robotics.

The robot will move in a virtual environment created in Gazebo using ROS [12] control, which already has a map of the surrounding environment. The robot's body is modeled in Blender, and its sensors, including LiDAR, camera, and IMU, are integrated using URDF. The key features of Gazebo include realistic physical modeling of objects, support for various sensors, powerful visualization tools, flexible extensibility through plugins, integration with ROS, and the ability to create various testing scenarios. Exporting 3D models to Gazebo can be done from various 3D modeling programs such as Blender or Autodesk Maya to formats supported by Gazebo, such as COLLADA or OBJ, followed by creating a model description in URDF format for use in simulation.

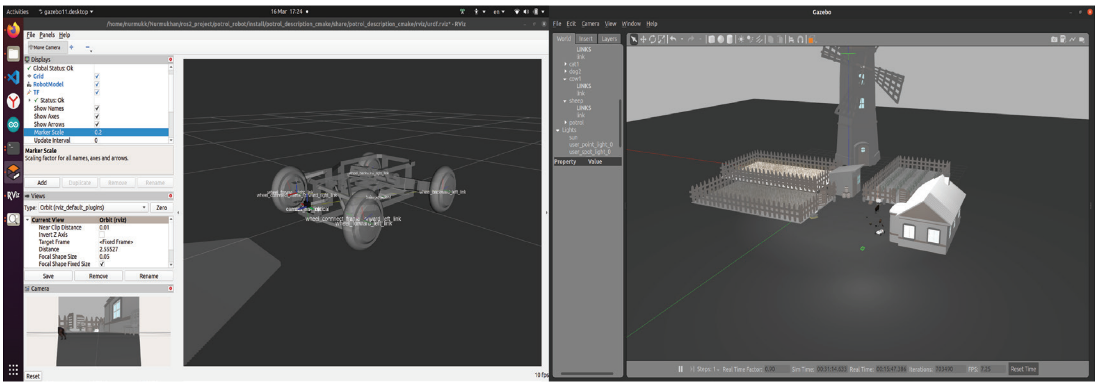

Figure 4 – 3D model of the robot and environment in Gazebo and ROS

The use of professional free and open-source software for creating three-dimensional computer graphics in Gazebo is widely used worldwide among researchers. For example, in work [13], the evolution of intelligent wheelchairs with new control systems aimed at assisting the user in increasing their independence is described. At the beginning of the work, a 3D model of a motorized wheelchair with robotic tools for use in simulation environments is proposed. The model is based on the Robotic Operating System (ROS) and Gazebo, which facilitates the addition of sensors and actuators. The proposed approach also allows using one controller for both simulation and the real system. The developed model can be used to test new approaches in simulations before real-world implementation.

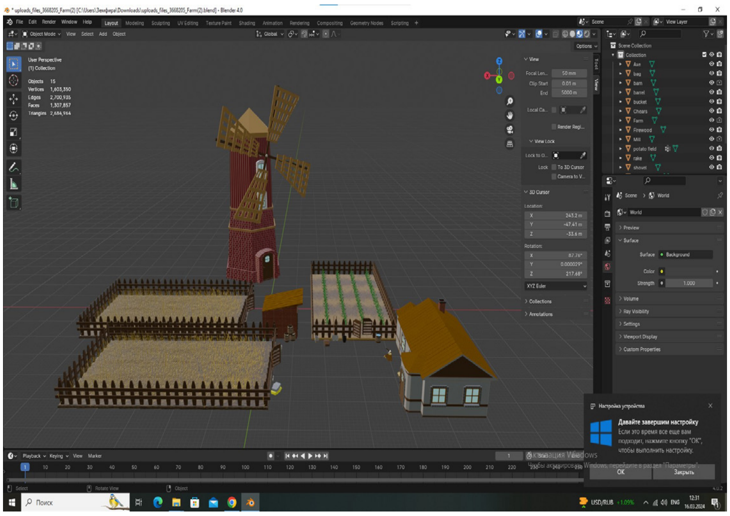

Figure 5 – Modeling the Gazebo world in Blender

4. Study of Computer Vision Technologies in the Context of Robotics and Autonomous Systems

Key Concepts and Principles of Computer Vision: Computer vision, integrated into the field of artificial intelligence, deals with processing and analyzing visual information using computers. This process involves extracting key features from images or videos, such as edges, corners, colors, and textures. Image segmentation, dividing them into separate components, facilitates subsequent analysis of objects. Object classification in images or videos using trained models determines their categories. Also within computer vision, detection and highlighting of specific objects or regions of interest in the image occur.

The Importance of Computer Vision in Robotics and Autonomous Systems: Computer vision plays a crucial role in robotics, allowing machines to interpret visual information from the environment and interact with it. Integrating computer vision into robotics expands the capabilities of robots and facilitates a wide range of applications. Computer vision enables robots to perceive the surrounding environment, which is critical for navigation and interaction with objects and humans. Robots can use computer vision to detect objects, analyze their properties, and make appropriate decisions, such as object tracking or obstacle avoidance. Overall, computer vision opens up broad prospects for improving the functionality and efficiency of robots and autonomous systems, making them more capable of interacting with the surrounding world.

Improving Performance and Efficiency: Automating animal monitoring processes with YOLOv5 can increase farm productivity and efficiency. For example, in work [14], an efficient and autonomous system based on computer vision is presented for detecting wild animals in border areas. The method uses the YOLO object detection model to recognize six types of entities: humans and five different animal species. After detecting the animal, it is tracked to determine its intentions, and notifications are sent to the relevant authorities based on the analysis results. A prototype of the system based on Raspberry Pi devices with cameras is also described. The system demonstrates high accuracy in detecting and identifying animals (98.8%) and humans (99.8%) and can be easily expanded to detect other animal species with sufficient training data."

"Farmers can quickly detect problems or anomalies in the behavior or condition of animals, allowing them to take measures to improve housing conditions and care.

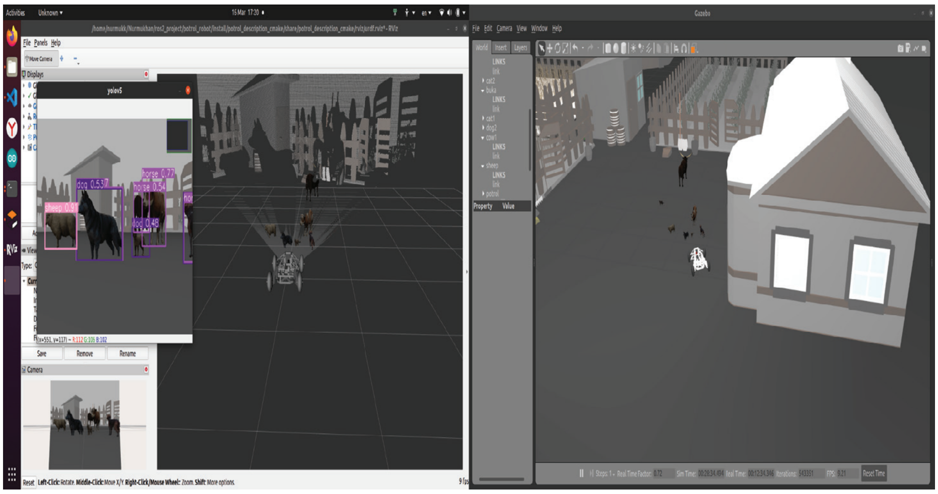

Figure 6 – Recognition of animals in a 3D simulation

YOLOv5 was used for real-time object recognition [15]. YOLO (You Only Look Once) is widely used in computer vision for object detection and classification in images and videos.

Architecture of the YOLO-v5 Model. In YOLOv5, these components are used together to create a complete neural network architecture capable of detecting objects in images with high accuracy and efficiency.

Backbone: This part of the neural network architecture is responsible for extracting features from input images. Pre-trained convolutional neural networks such as ResNet, Darknet, or EfficientNet are commonly used.

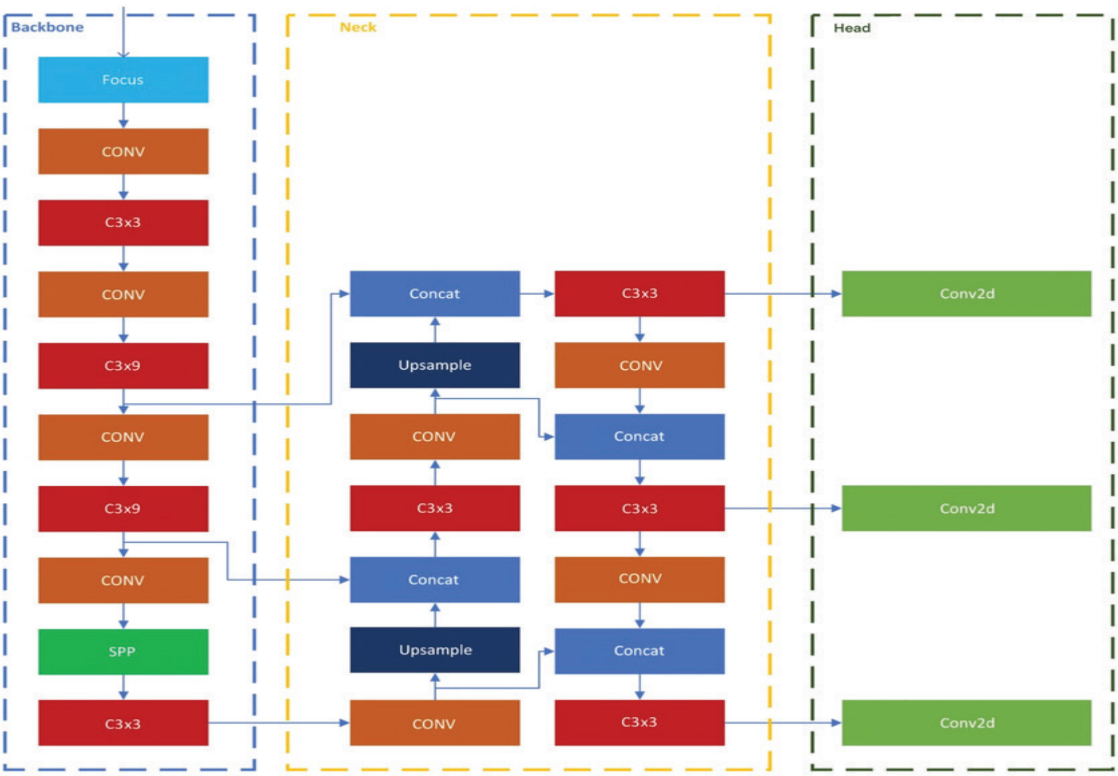

Figure 7 – The architecture of YOLOv5

Neck: This part of the architecture processes features obtained from the backbone to make them more suitable for further processing stages.

Head: This is the final part of the architecture responsible for predicting the coordinates and classes of objects. Typically, the head consists of various layers such as convolutional layers, pooling layers, and fully connected layers.

C3, FOCUS, CONV, SPP: These are types of layers or modules used in the YOLOv5 architecture:

C3: Convolutional layer that performs convolution with multiple kernels (usually 3x3), allowing the model to extract more complex spatial features.

FOCUS: A special type of convolutional layer that helps reduce computation volume, speeding up image processing.

CONV: Ordinary convolutional layer that performs convolution operation on input data.

SPP (Spatial Pyramid Pooling): A layer that allows the model to work with objects of different sizes in the image by creating a pyramid of features of different scales.

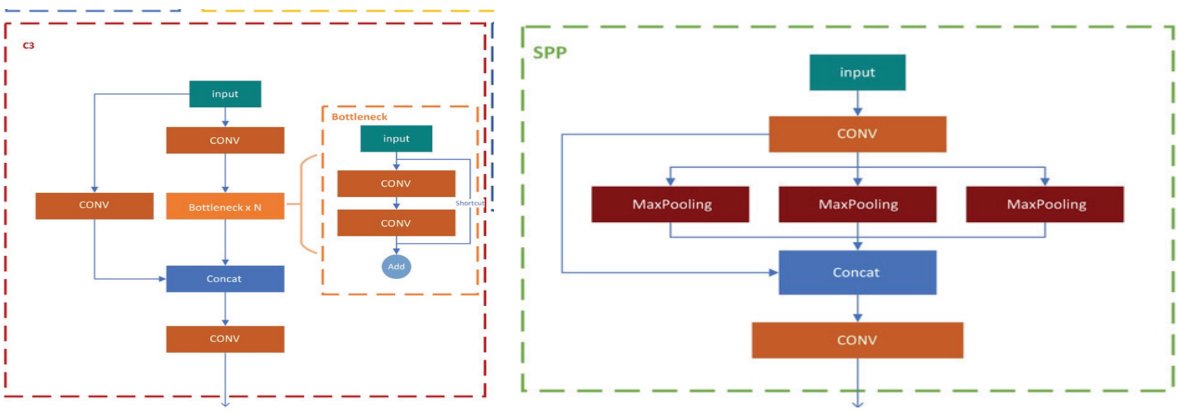

Figure 8 – Components of the YOLOv5 Architecture

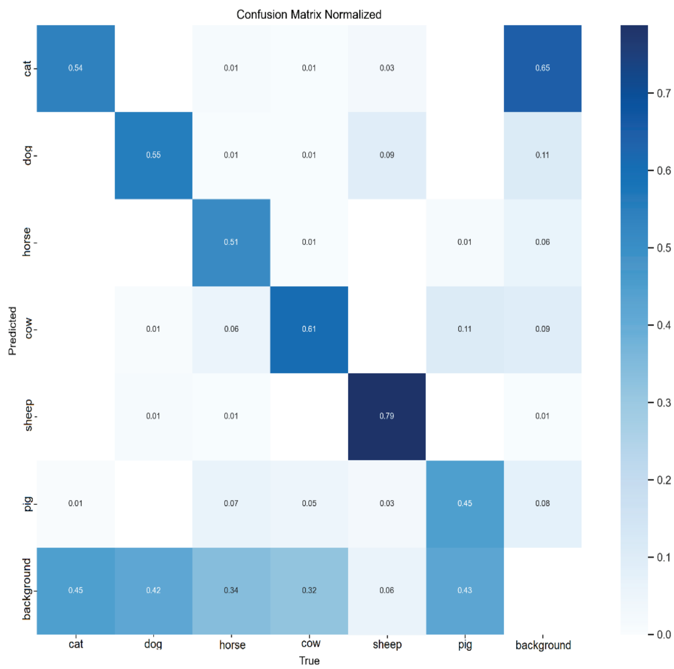

Figure 9 – Confusion Matrix Normalized

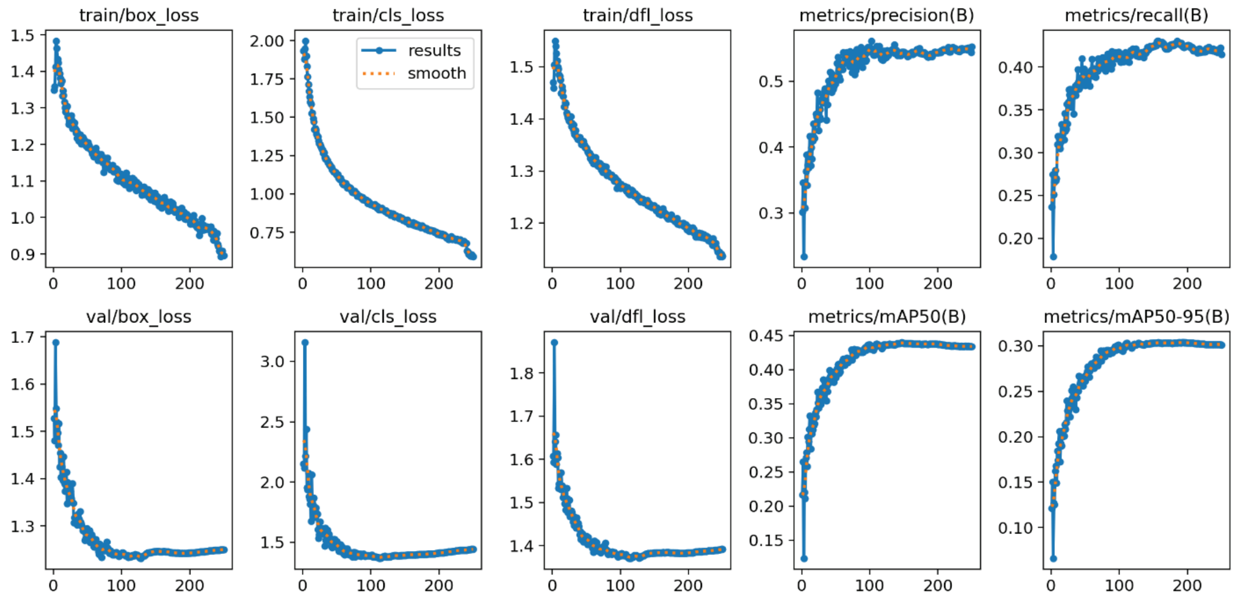

Figure 10 – The overall training results

Training Results of the YOLOv5 Model.

1. train/box_loss: This is the loss function associated with the error in predicting bounding box coordinates of objects in the training dataset.
2. train/cls_loss: This is the loss function associated with the error in classifying objects in the training dataset.
3. train/dfl_loss: This is the loss function associated with the error in predicting additional features (e.g., orientations or sizes) of objects in the training dataset.
4. metrics/precision(B): Precision is a metric that measures the proportion of objects predicted as positive that are truly positive. B in this case likely indicates the batch of data.
5. metrics/recall(B): Recall is a metric that measures the proportion of true positive objects that were correctly detected by the model. B in this case likely indicates the batch of data.
6. val/box_loss: Same as train/box_loss but for validation (test) data.
7. val/cls_loss: Same as train/cls_loss but for validation (test) data.
8. val/dfl_loss: Same as train/dfl_loss but for validation (test) data.
9. metrics/mAP50(B): Mean Average Precision (mAP) at a specified IoU (Intersection over Union) threshold of 50%. B in this case likely indicates the batch of data.
10. metrics/mAP50-95(B): Same as metrics/mAP50 but for IoU ranging from 50% to 95%.

5 Conclusion

In the article, we discussed the importance of applying the SLAM algorithm in combination with computer vision for livestock detection in agriculture. The use of these technologies opens up new prospects for effective animal management and enhancing the functionality of autonomous robots in agricultural settings. Decent results in computer vision have been presented, and in the future, we will focus on improving recognition accuracy.

References

1. Zhao, Wei, et al. "Ground-level mapping and navigating for agriculture based on IoT and computer vision." IEEE Access 8 (2020): 221975-221985.
2. Yeshmukhametov, Azamat, et al. "Designing of CNC based agricultural robot with a novel tomato harvesting continuum manipulator tool." International Journal of Mechanical Engineering and Robotics Research 9.6 (2020): 876-881.
3. Yeshmukhametov, Azamat, et al. "Development of Mobile Robot with Autonomous Mobile Robot Weeding and Weed Recognition by Using Computer Vision." 2023 23rd International Conference on Control, Automation and Systems (ICCAS). IEEE, 2023.
4. Hoang, Tran Ngoc, et al. "Optimizing YOLO Performance for Traffic Light Detection and End-to-End Steering Control for Autonomous Vehicles in Gazebo-ROS2." International Journal of Advanced Computer Science and Applications 14.7 (2023).
5. Li, Guoming, et al. "Practices and applications of convolutional neural network-based computer vision systems in animal farming: A review." Sensors 21.4 (2021): 1492.
6. Rodenburg, Jack. "Robotic milking: Technology, farm design, and effects on work flow." Journal of Dairy Science 100.9 (2017): 7729-7738.
7. Durrant-Whyte, Hugh, and Tim Bailey. "Simultaneous localization and mapping: part I." IEEE robotics & automation magazine 13.2 (2006): 99-110.
8. Taketomi, Takafumi, Hideaki Uchiyama, and Sei Ikeda. "Visual SLAM algorithms: A survey from 2010 to 2016." IPSJ transactions on computer vision and applications 9 (2017): 1-11.
9. Zheng, Shuran, et al. "Simultaneous Localization and Mapping (SLAM) for Autonomous Driving: Concept and Analysis." Remote Sensing 15.4 (2023): 1156.
10. Koenig, Nathan, and Andrew Howard. "Design and use paradigms for gazebo, an open-source multi-robot simulator." 2004 IEEE/RSJ international conference on intelligent robots and systems (IROS) (IEEE Cat. No. 04CH37566). Vol. 3. Ieee, 2004.
11. Xuexi, Zhang, et al. "SLAM algorithm analysis of mobile robot based on lidar." 2019 Chinese Control Conference (CCC). IEEE, 2019.
12. Quigley, Morgan, et al. "ROS: an open-source Robot Operating System." ICRA workshop on open source software. Vol. 3. No. 3.2. 2009.
13. Cruz, Ana Beatriz, Armando Sousa, and Luìs Paulo Reis. "Controller for real and simulated wheelchair with a multimodal interface using gazebo and ROS." 2020 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC). IEEE, 2020.
14. Sayagavi, Ashwini V., T. S. B. Sudarshan, and Prashanth C. Ravoor. "Deep learning methods for animal recognition and tracking to detect intrusions." Information and Communication Technology for Intelligent Systems: Proceedings of ICTIS 2020, Volume 2. Springer Singapore, 2021.
15. Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

Information about authors:

Azamat Yeshmukhametov – PhD, Postdoctoral Scholar, School of Engineering and Digital Sciences, Department of Robotics and Mechatronics of Nazarbayev University(Astana, Kazakhstan, e-mail: azamat.yeshmukhametov@nu.edu.kz).

Submission received: 19 March, 2024.
Revised: 21 March, 2024.
Accepted: 21 March, 2024.