Cow Face Detection for Precision Livestock Management using YOLOv8

Umair Ali[1], Wasif Muhammad[2]

[1]Intelligent Systems Laboratory, Department of Electrical Engineering, University of Gujrat, Gujrat, Pakistan

[2]Intelligent Systems Laboratory, Department of Electrical Engineering, University of Gujrat, Gujrat, Pakistan

* Correspondence: Wasif Muhammad, syed.wasif@uog.edu.pk

Abstract. Precision livestock management is revolutionising traditional agricultural practices by enhancing productivity, yield, and automation while minimising labour and inaccuracies. Traditional animal recognition methods are often unreliable, leading to a growing trend of using cameras for animal identification, health monitoring, data management, and maintaining cattle records. However, small-scale farms with limited livestock, such as cows and goats, often encounter overfitting issues in conventional machine-learning models due to insufficient training data. Individual cow identification using facial features is better possible after detection of cow face. This study addresses these challenges by fine-tuning YOLOv8, a pretrained model, using a combination of self-captured images and publicly available datasets to detect cow faces in a complex environment. Incorporating publically available data and adding a pre-trained COCO model has significantly generalized the model to detect cow faces in complex environments.. Yolov8 with COCO pretrained model is detecting nearly all kinds of cow faces which can later be used to classify the individual cow.

Introduction

Agriculture industry is revolutionizing by adding advanced technologies to conventional methods yielding high efficiency and productivity. Data driven approaches in agriculture have introduced a new field called Precision agriculture. These approaches are significant in livestock management to compete world demand for Food [1]. Countries dependent on livestock products have been facing storage of food and agricultural resources that can be beneficial to mankind. However there are severe challenges related to animal health, labour and management [2]. Where there is need for increase in animal production there is also need for implementation of smart and efficient systems that could increase the production as well as ensure their sustainability [3].

Traditional livestock management is insufficient in the modern world because of their consumption of time, chances of error and require manpower. Studies have shown poor labour can be a problem for livestock health and there is a good chance of spreading diseases due to lack of care and poor management [4]. There is a good chance of dispute among farmers because of the mixing of animals and their inability to track individual cows accurately [5]. To address their issues people have been using various for animal identification and tracking their performance. In general there are three main kinds of methodologies which have been implemented for recognition of cows that are permanent, semi-permanent and temporary identification methods.

Permanent identification methods are used when animals are marked physically. These methods are commonly used [6] but highly unreliable especially for large herds. Semi-permanent identification methods such as adding tags and collars are widely adopted to keep track of cows but there is a good chance of loss of tags, wear and tear. labour is involved so there is a good chance of inaccuracies in keeping records [7]. Temporary Methods such as RFID tags have gained popularity due to their ability to provide automated identification; however, their durability is often questioned, and farmers report frequent maintenance issues. However, RFID systems can disturb animals due to unfamiliar sounds, overhead machinery, and the stress caused by handling or installation.

There is a need for advanced solutions to keep the history of cows and camera based systems are suitable for these tasks. Until traditional and conventional methods computer vision based systems are widely adopted and accurate for recognition [8]. Cows normally identify each other using facial features so it is quite useful to recognize cows using their facial features. Computer vision approach of facial based cow recognition is quite helpful for individual cow identification without needing physical tags and making it more sustainable and scalable for large herds [9].

For facial recognition convolutional neural networks (CNNs) are widely being used due to their ability to extract complex features. However there is a good chance of overfitting due to the limited number of dataset [10]. These challenges are crucial in livestock management as there are always variations in lighting, angles, and animal movement in capturing cow images [11]. Preprocessing of data and its augmentation can improve the performance of cow recognition but require diverse datasets for generalization. Ultimately for high accuracy there is first need to detect the cow face and then classify cows based on their facial features. A supervised learning based approach with manual annotations to localize cow faces before proceeding to classification can be useful for accuracy [12].

Current study proposed a transfer learning-based approach in which Yolov8 used a pre-trained model on the COCO dataset for the first detection of faces accurately. The pretrained model can be fine tuned using a small dataset for cow face localization. This involves using YOLO (You Only Look Once), a state-of-the-art object detection model known for its speed and accuracy in detecting objects in real-time [13].

The objective of this research is to use a pretrained model for cow face detection in variational background conditions at different angles and lighting. By fine-tuning YOLOv8, a state-of-the-art pretrained object detection model, with a dataset of annotated frames from self collected videos and publicly available sources, the aim of the study is to improve accuracy in cow face detection. The proposed approach combines the precision of YOLOv8 with self collected and publicly available dataset, ensuring reliable cow face detection in a complex environment.

Material and Methods

Investigation site. The dataset for cow facial recognition is captured by recording videos of 37 different cows. Every 10th frame is extracted from video to make diverse variational environment based images and annotated cow faces manually using Roboflow tool. For generalization further annotated web based dataset is corporated for having high accuracy. The pretrained yolov8 model is fine-tuned in a simulation environment by feeding the dataset.
Material and methods. The methodology for this study consists of multiple steps as shown in Figure 1.
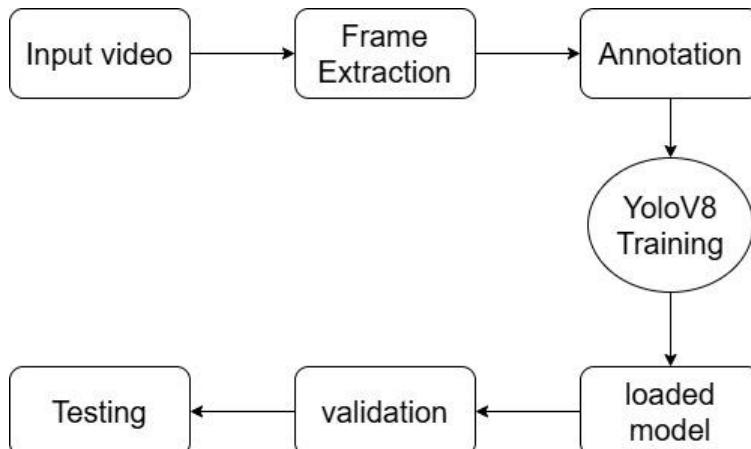

Figure 1.1 Workflow for Cow Face Detection using YOLOv8: Input videos are processed through frame extraction and annotated for training. The YOLOv8 model is fine-tuned and loaded for validation and detection tasks

Dataset

Videos of 37 individual cows were recorded from various angles under different lighting conditions and movements. From these recordings, 1,078 frames were extracted at regular intervals of every 10th frame to create the initial dataset. To further expand the dataset and improve diversity, 1,410 annotated images were added from a publicly available dataset on Roboflow. This brought the total dataset size to 2,488 images before preprocessing, with three different classes. The classes from the external dataset were adjusted and mapped into a single class, CowFace, to maintain consistency across all images.
The dataset was then preprocessed to prepare it for training. All images were resized to 640×640 pixels to standardize dimensions while retaining important features. After resizing,

the dataset was split into three subsets: 1,800 images for training, 560 images for validation, and 86 images for testing. Data augmentation techniques were applied to increase variability in the training set, including horizontal flips, zoom cropping (0–20%), rotations (-15° to +15°), and shear transformations (±10° horizontally and vertically). These augmentations created three additional variations for each image, resulting in a total dataset size of 5,096 images after preprocessing and augmentation. This process ensured that the dataset was diverse and ready for effective model training.

YOLOv8 architecture

The fundamental Architecture of the first version of Yolo is shown in Figure 1.2
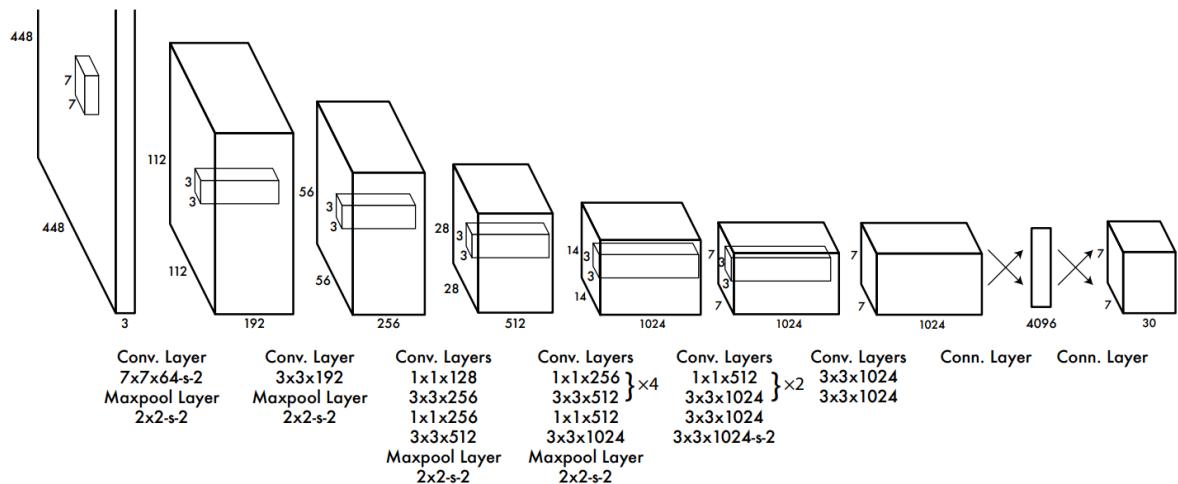
Figure 1.2 The YOLOv1 architecture [13] processes an input image through sequential convolutional layers for feature extraction reduce dimensions with max-pooling, and outputs object detection predictions through fully connected layers.

The original YOLOv1 architecture [13] was proposed by Joseph Redmon et al. It begins with a 7×7 convolutional layer (stride 2) for large feature extraction, followed by max-pooling (2×2, stride 2). The network then employs alternating 1×1 and 3 × 3 convolutional layers to balance dimensionality reduction and feature extraction, using 24 convolutional layers in total. Finally, 2 fully connected layers are added, with the last layer producing a 7 × 7 × 30 output for bounding boxes, confidence scores, and class probabilities. This design integrates detection into a single neural network for efficiency.

After several improvements, YOLOv8 was proposed as a highly stable and advanced version, incorporating robust community-driven ideas and modern concepts for enhanced performance and versatility. Any Official Paper is not yet published for Yolov8 but the architecture we find is available [14] as shown in Figure 1.3
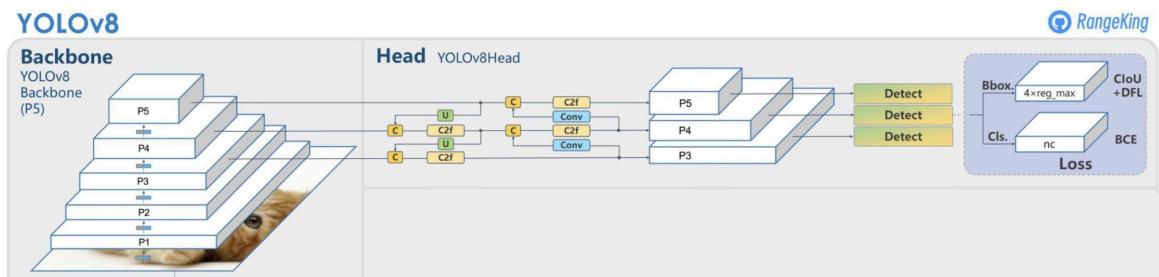
Figure 1.3 YOLOv8 predicts object centers directly (anchor-free), making it faster and simpler. It uses a new C2f module for better feature extraction and reduces model size with efficient design tweaks. Mosaic data mixing is used early in training but stopped later for accuracy.

YOLOv8 introduces several key improvements that enhance its performance and efficiency. It predicts object centres directly without the need for anchor boxes, simplifying the model and making it faster and more efficient. YOLOv8 incorporated a C2f module which can better extract the feature from frames under a complex environment and its design is optimized. YOLOv8 is scalable and can be used for both high performance GPUs and edge devices.

Implementation of YOLOv8

YOLOv8 has demonstrated exceptional performance in detecting cow faces across diverse test datasets and real-world web images. It achieves high precision through a combination of transfer learning with help of COCO weights and fine-tuning on a custom dataset. This dataset, built from video frames and augmented Roboflow images, captures varying lighting, angles, and environmental conditions, ensuring robust generalization. Trained over 50 epochs at a 640×640 resolution with a batch size of 16, the model leverages single-class detection to streamline focus on cow faces. It has previously outperformed SSD, Faster R-CNN, and earlier YOLO iterations in both speed and accuracy. Its real-time capabilities, supported by efficient computational performance, make it ideal for deployment in livestock monitoring systems. By balancing augmented data diversity with optimized architecture YOLOv8 delivers scalable solutions. The implementation of YOLO for cow face detection involves three key mathematical principles. These three principles include loss optimization, weight updates, and bounding box prediction. The YOLO loss function is a combination of bounding box regression, object confidence, and classification loss to ensure accurate detection and classification.

Result and discussion

Yolov8-based facial recognition is successfully detecting cow faces from test datasets and it is also tested and confirmed by detecting cow faces from the web images. The model was fine-tuned using a custom dataset, which included both the captured frames from videos and additional images sourced from a publicly available dataset on Roboflow. This dataset was chosen for its diversity and it contains varied angles, lighting conditions and cow face features. The fine-tuning process involved 50 training epochs, a batch size of 16, and a resized image resolution of 640×640 pixels to align with the preprocessing steps. YOLOv8's flexibility in handling single-class detection and its compatibility with augmented datasets made it ideal for this research. Compared to earlier YOLO versions and other object detection frameworks like SSD or Faster R-CNN, YOLOv8 offered significant advantages in terms of inference speed.

Comparison of Label and Prediction

Label Image   Prediction Image

Figure 2.1 Left: Manually labelled image of cow faces; Right: YOLOv8 prediction of the same image. The model, fine-tuned on a diverse custom dataset, accurately detects cow faces under varying conditions, as shown in the comparison.

The trained model is also tested with multiple pictures available online. Figure 2.2 below describes the working of the model and as a generalization, any cow face can be detected.

Original and Prediction Comparison

Raw Image from Web   *Yolov8 based Prediction Image*
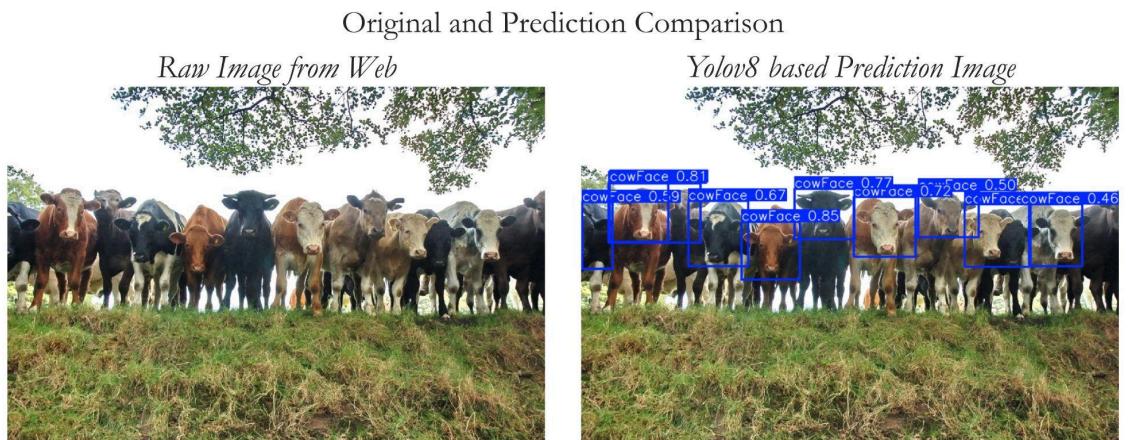
Figure 2.2 A raw image sourced from the web is displayed on the left side as the original picture. After processing, the tuned model accurately detects and labels the cow face class, with the processed image shown on the right side.

The ability to generalize on smaller datasets without overfitting makes it more useful. Moreover, the model's streamlined deployment and its ability to deliver real-time predictions make it a powerful tool for practical applications in livestock monitoring. The combination of a robust architecture, transfer learning from COCO weights, and a fine-tuned dataset ensured that the YOLOv8 model delivered high performance. The Precision in detecting cow faces, proving its superiority over conventional methods.

Figure 2.3 illustrates the progression of key performance metrics throughout training. Precision stabilizes around 0.976, signifying a low false positive rate. Recall converges near 0.955, indicating the model's ability to detect the most relevant objects. The mAP50 metric reaches an average of 0.981, showcasing the model's strong detection performance at an IoU threshold of 0.5. These metrics collectively highlight the YOLO model's high reliability in object detection tasks.
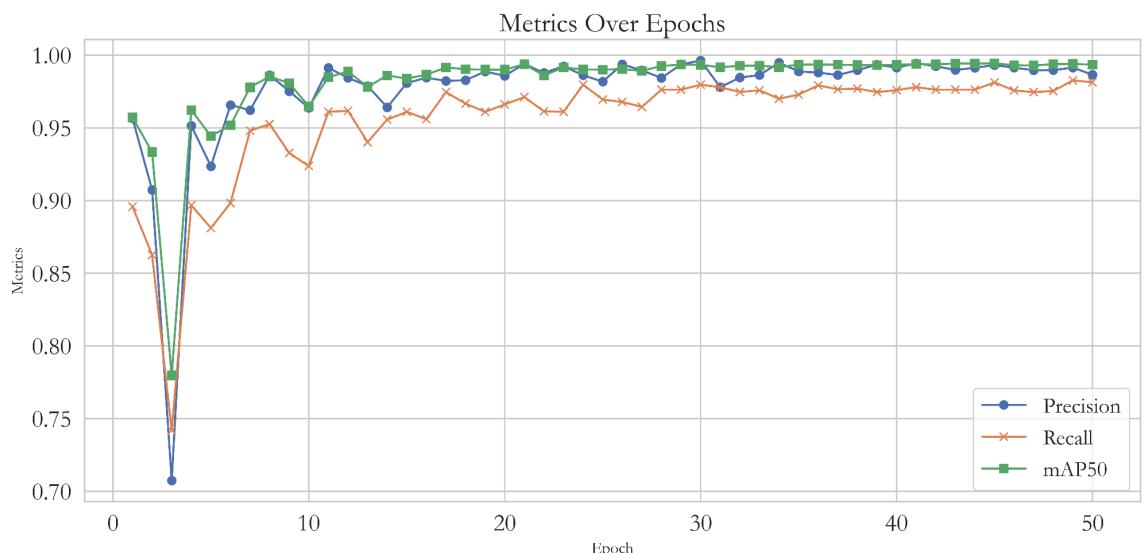
Figure 2.3 . Precision, recall, and mAP50 metrics stabilize at high values, demonstrating reliable and balanced object detection performance.

Figure 2.4 depicts the trends of box, classification, and distributional focal losses for both training and validation datasets. The steady decrease in training losses reflects the model's ability to adapt to the dataset, while the slightly higher validation losses indicate some generalization challenges. However, the alignment of the two curves suggests minimal overfitting, demonstrating that the model effectively captures essential features.
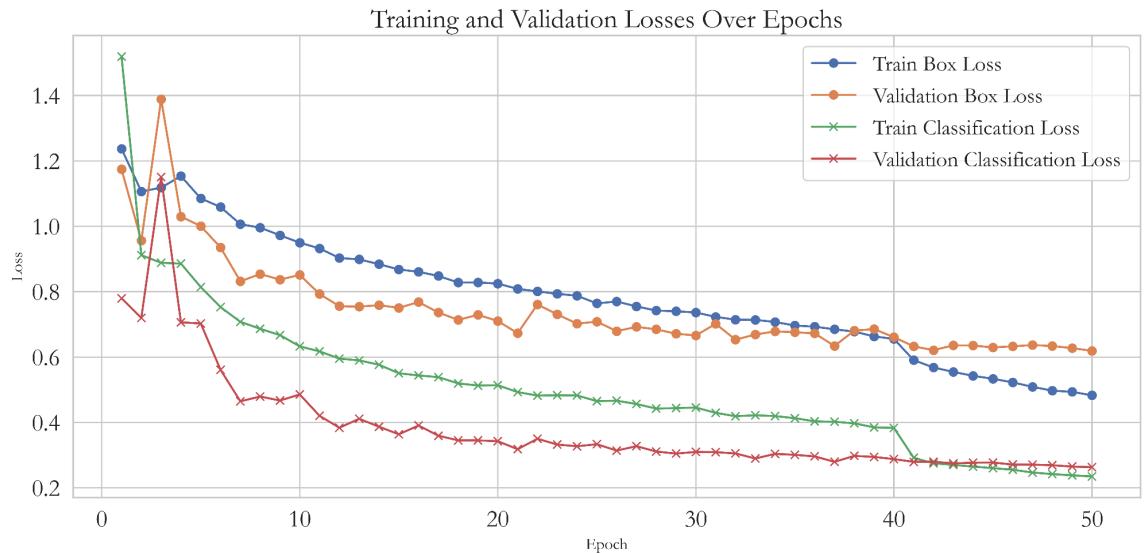
Figure 2.4 Training and validation losses show a steady decline, indicating effective learning with minimal overfitting, despite slightly higher validation losses
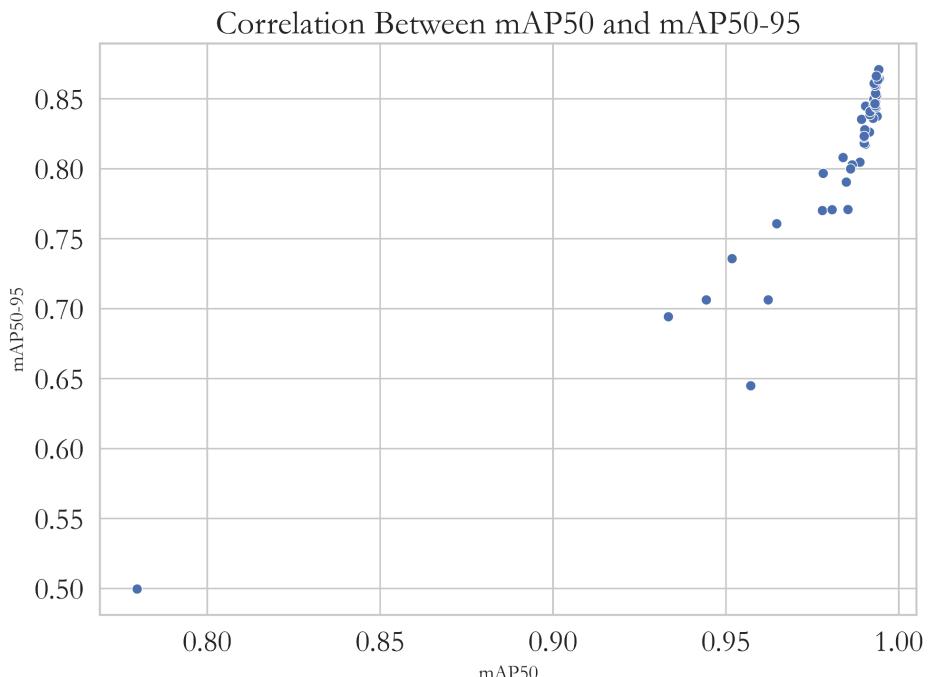

Figure 2.5 A strong linear correlation between mAP50 and mAP50-95 confirms consistent detection accuracy across varying IoU thresholds.

The scatterplot shown in Figure 2.5 establishes a strong linear correlation (≈ 0.98) between mAP50 and mAP50-95. This relationship demonstrates that improvements in detection accuracy at a 0.5 IoU threshold positively impact performance across stricter IoU thresholds.

The near-linear trend confirms the robustness of the YOLO model in accommodating varying levels of localization strictness. Performance metrics are computed after each epoch to evaluate the model's detection capability. These include precision (the ratio of true positives to all predicted positives), recall (the ratio of true positives to actual positives), and mAP50 (mean average precision at a 0.5 IoU threshold). The mAP50-95 metric extends this evaluation across a range of IoU thresholds (0.5 to 0.95), providing a comprehensive performance measure. Validation losses, calculated on unseen data, assess the generalization capability of the trained model.
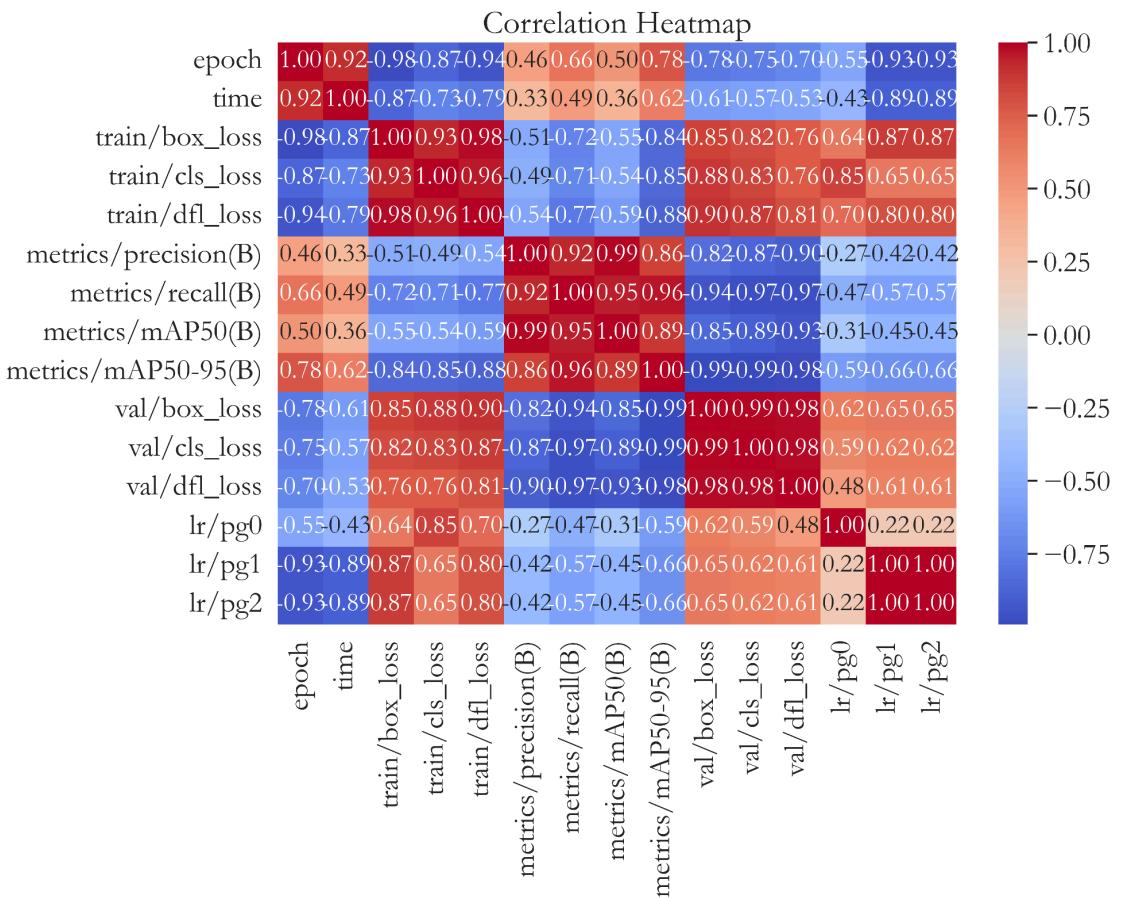

Figure 2.6 The correlation heatmap reveals key metric interdependencies, highlighting the impact of reduced losses on detection accuracy and the optimization behavior of learning rates.

The heatmap in Figure 2.6 highlights the relationships between key metrics, losses, and learning rates. Strong negative correlations are observed between box loss and mAP50/mAP50-95, indicating that lower localization errors lead to higher detection accuracy. Positive correlations between precision and recall emphasize their interdependence, where improvements in one directly enhance the other. Learning rate parameters (e.g., lr/pg0, lr/pg1, and lr/pg2) exhibit weak direct correlations with precision and recall, reflecting their role in optimizing model weights rather than directly influencing evaluation metrics. The heatmap provides critical insights into the optimization behavior and performance interdependencies of the model.

The results validate the YOLO model's high precision and overall robustness. The strong correlation between mAP50 and mAP50-95 demonstrates the model's capability to perform consistently across varying IoU thresholds. This makes it suitable for applications requiring high detection accuracy. Validation losses deviation highlights potential areas for improvement, such as introducing advanced augmentation techniques or further fine-tuning hyperparameters. These findings underscore applicability of the model in real-world scenarios where precise and reliable object detection is critical.

Table 4.1 shows an analysis of training and validation losses and performance metrics. It highlights the model's effectiveness in object detection. Training box loss decreased steadily across epochs, with validation box loss following a similar trend. Its slightly higher values indicate good generalization with minimal overfitting. Classification losses for both training and validation consistently declined, showcasing the model's improving accuracy in object classification. Performance metrics further emphasize robustness with precision averaging 0.976, recall at 0.955, and mAP50 achieving an impressive 0.981 that reflects high detection accuracy at an IoU threshold of 0.5. Additionally, the mAP50-95 average of 0.812 demonstrates reliable performance across varying IoU thresholds. Gradual reductions in learning rates across parameter groups (lr/pg0, lr/pg1, lr/pg2) facilitated smooth convergence and prevented loss oscillations to ensure stable training progress.

.
Table 4.1. analyzes training and validation losses, showing steady improvement in object localization and classification accuracy. The metrics like precision (0.976), recall (0.955), and mAP50 (0.981) reflect a good detection accuracy. Gradual learning rate reductions ensured smooth convergence and stable training progress.

epoch	train /box _loss	train /cls_ loss	metrics/prec ision(B)	metrics /recall(B)	metrics /mAP5 0(B)	metrics /mAP5 0-95(B)	val/bo x_loss	lr/pg 0	lr/pg1	lr/pg2
1	1.237	1.519	0.956	0.89581	0.95719	0.64492	1.1744	0.0700	0.003324	0.003324
2	1.106	0.911	0.907	0.86255	0.93337	0.69425	0.9560	0.0399	0.006525	0.006525
3	1.118	0.888	0.707	0.74189	0.77968	0.4997	1.3886	0.0096	0.009595	0.009595
4	1.153	0.885	0.951	0.89676	0.96222	0.70639	1.0297	0.0094	0.009406	0.009406
-	-	-	-	-	-	-	-	-	-	
48	0.497	0.242	0.989	0.97536	0.99387	0.86499	0.6339	0.0006	0.000694	0.000694
49	0.493	0.238	0.991	0.98266	0.99399	0.86363	0.6274	0.0004	0.000496	0.000496
50	0.483	0.235	0.986	0.98136	0.9935	0.86628	0.6191	0.0002	0.000298	0.000298

Conclusion. This study successfully implemented a transfer learning-based approach for detecting cow faces in diverse farm environments using YOLOv8. By fine-tuning the model with a carefully selected dataset, which combined frames captured from video footage and

publicly available annotated images. The system demonstrated excellent performance in both precision and recall. The results highlighted in YOLOv8 performance are improving in each epoch and reducing box losses.

The model achieved a precision of 0.976 having recall of 0.955 and mAP50 of 0.981 that indicated its capability to detect cow faces with high accuracy. The integration of data augmentation and fine-tuning on a diverse dataset played a key role in optimizing generalization. The implementation of early stopping helped avoid overfitting. Additionally, the strong correlation between mAP50 and mAP50-95 demonstrated the model's consistency across varying levels of object localization.

The findings confirm that YOLOv8 is a highly effective tool for real-time face detection in livestock localization applications and livestock monitoring. This approach not only lays the foundation for accurate cow face detection but also sets the stage for further advancements in classification tasks. This makes it a valuable contribution to the livestock domain for cow face detection. Future work may explore additional model refinements and integration with other technologies for cow face detection and recognizing individual cows.

REFERENCES

[1] P. J. Zarco-Tejada, N. Hubbard, and P. Loudjani, "Precision agriculture: an opportunity for EU farmers—potential support with the CAP 2014–2020," *Joint Research Centre (JRC) of the European Commission*, 2014.

[2] S. K. Seelan, S. Laguette, G. M. Casady, and G. A. Seielstad, "Remote sensing applications for precision agriculture: A learning community approach," *Remote Sensing of Environment*, vol. 88, no. 1-2, pp. 157–169, 2003.

[3] A. Monteiro, S. Santos, and P. Gonçalves, "Precision agriculture for crop and livestock farming—brief review," *Animals*, vol. 11, no. 8, p. 2345, 2021.

[4] A. I. Awad, "From classical methods to animal biometrics: A review on cattle identification and tracking," *Computers and Electronics in Agriculture*, vol. 123, pp. 423–435, 2016.

[5] L. Bo, L. Yuefeng, B. Xiang, W. Yue, L. Haofeng, and L. Xuan, "Research on dairy cow identification methods in dairy farm," *Indian Journal of Animal Research*, vol. 57, no. 12, pp. 1733–1739, 2023.

[6] A. Johnston and D. Edwards, "Welfare implications of identification of cattle by ear tags," *Veterinary Record*, vol. 138, no. 25, pp. 612–614, 1996.

[7] D. Wardrope, "Problems with the use of ear tags in cattle," *Veterinary Record*, vol. 137, no. 26, p. 675, 1995.

[8] A. Ruchay, K. Dorofeev, V. Kalschikov, V. Kolpakov, and K. Dzhulamanov, "A depth camera-based system for automatic measurement of live cattle body parameters," *IOP Conference Series: Earth and Environmental Science*, vol. 341, no. 1, p. 012148, 2019.

[9] M. Vlad, R. A. Parvulet, M. S. Vlad, et al., "A survey of livestock identification systems," in *Proceedings of the 13th WSEAS International Conference on Automation and Information (ICAI12)*, WSEAS Press Iasi, Romania, pp. 165–170, 2012.

[10] P. Thanapol, K. Lavangnananda, P. Bouvry, F. Pinel, and F. Leprévost, "Reducing overfitting and improving generalization in training convolutional neural networks (CNN) under limited sample sizes in image recognition," in *2020-5th International Conference on Information Technology (InCIT)*, IEEE, pp. 300–305, 2020.

[11] I. K. Mirani, C. Tianhua, M. A. A. Khan, S. M. Aamir, and W. Menhaj, "Object recognition in different lighting conditions at various angles by deep learning method," *arXiv preprint arXiv:2210.09618*, 2022.

[12] X. Lei, X. Wen, and Z. Li, "A multi-target cow face detection model in complex scenes," *The Visual Computer*, pp. 1–22, 2024.

[13] J. Redmon, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[14] J. Solawetz and Francesco, "What is YOLOv8? The ultimate guide," *https://blog.roboflow.com/what-is-yolov8/*, 2023, accessed: 2023-04-30.

Check List

1. **Abstract is designed as advised in the start of this template**

2. **Objectives are clearly mentioned**

3. **Novelty statement is provided**

4. **Diagram of flow of methodology is available**

5. **Results and discussion sections are provided separately and comprised of more than 70% of the complete length of manuscript.**

6. **References are provided as per latest literature of last 2-3 years**

7. **Avoid to use old literature which has become obsolete**