Analysing the social network dynamics of dairy cows with computer vision

Laura Ozella[1], Davide Vitturini[2], Elena Diaz Vicuna[1], Marco Bovo[3], Claudio Forte[1], Mario Giacobini[1], Marco Grangetto[2]
[1]*Department of Veterinary Sciences, University of Torino, Grugliasco (TO), Italy*
[2]*Department of Computer Science, University of Torino, Torino, Italy*
[3]*Department of Agricultural and Food Sciences, University of Bologna, Bologna, Italy*
laura.ozella@unito.it

Abstract

The structured social hierarchies and individual social bonds within dairy cows' groups significantly influence their welfare and productivity. Traditional monitoring methods are labour-intensive and often lacking insights into precise spatial positioning or the nature of behaviour. Here, a computer vision system was proposed, employing YOLOv8, for automated identification and tracking of dairy cows in barns equipped with Automated Milking Systems. The study pioneers the analysis of social interactions among 240 dairy cows using computer vision. Through comprehensive dataset preparation and model training, robust performance in cow detection and tracking was achieved. Evaluation metrics demonstrate the model's effectiveness in real-world settings, with high precision and recall values. The construction of temporal social networks was enabled, revealing insights into cow interactions. The results indicate the model's suitability for practical use, paving the way for future real-time analyses of cow social networks and promising advancements in dairy farming management.

Keywords: Dairy Cows, Precision Livestock Farming, Computer Vision System, Convolutional Neural Networks, Social Interactions

Introduction

Cattle, like many other gregarious species, form structured groups characterized by well-defined social hierarchies, and within these stable social units, cows develop particular social bonds (Gygax et al., 2010). The social environment and the ability to express specific social behaviours hold significant implications for both the well-being and productivity of cows (Bouissou et al., 2001). Traditional methods for monitoring social interactions in dairy cows often rely on manual observation, which is both time-consuming and prone to inaccuracies. Consequently, conventional cow behaviour detection requires substantial staffing and resource allocation. Embedded sensor technology, such as spatial proximity loggers (Fielding et al., 2021; Leso et al., 2023) and Ultra-wideband (UWB) technology (Rocha et al., 2020), is commonly employed to explore social dynamics among animals. However, these sensors are limited to detecting social behaviour solely through spatial proximity assessment, lacking insights into precise spatial positioning or the nature of behaviour (positive or negative). Additionally, the use of wearable sensors poses challenges, necessitating attachment to the animal and management of battery life.
Computer vision technology is currently acknowledged as a pivotal tool for detecting cows' behaviours in a non-intrusive and cost-efficient manner (Tassinari et al., 2021; Zhang et al., 2023). Convolutional Neural Networks (CNNs) are algorithms widely used

in computer vision for livestock detection and recognition through the automatic extraction of both low-level features, such as edges and textures, and complex features (Wang et al., 2024). CNNs are used for cows' identity recognition (Zhang et al., 2023) as well as for the recognition of different behaviours such as lying, standing, walking (McDonagh et al., 2021). However, only a limited number of studies have focused on analysing social interactions among dairy cows using these technologies (Guzhva et al., 2016; Ren et al., 2021).
In our research, we developed and assessed the reliability of a computer vision system utilizing deep learning techniques, for the automated identification of 240 individual dairy cows in a barn equipped with Automated Milking Systems (AMSs). Specifically, we explored the application of YOLOv8 (YOLOv8), a widely used framework based on CNNs, for detecting and tracking cows in a barn environment (e.g., Zheng et al., 2022; Wang et al., 2024). For this study, we utilized meticulously annotated real data collected from the farm under investigation. To our knowledge, this is the first endeavour to analyse social interactions among a sizable population of dairy cows using computer vision. Our main goal is to monitor the cows and establish their social networks to understand how the social context influences both their welfare and production.

Material and Methods

Animals and Housing

The study was conducted in a cubicle free-stall barn situated on a commercial dairy farm in the North-West region of Italy. The structure of the barn consisted of two rectangular enclosed spaces, each measuring 45x30 meters. Each area hosted around 120 Holstein-Friesian lactating cows and was equipped with two AMSs (Lely Astronaut A4); the first was allocated for primiparous cows, whereas the second area was dedicated to multiparous cows. The cows were housed indoors for the duration of their entire production cycle.

Cameras configuration

Eight Super Wide Angle Fixed Bullet Network Cameras (Hikvision DS-2CD2T45G0P-I) were positioned to ensure a comprehensive view of both the barn area and the milking robots (Figure 1). Cameras were positioned to eliminate blind spots and ensure a comprehensive coverage of the barn. Each camera angle overlapped with others, providing multiple viewing perspectives, and maximizing visibility of cows regardless of their location within the barn. The cameras were installed outside the area where the animals were housed to prevent disturbances during routine checks and cleaning operations.

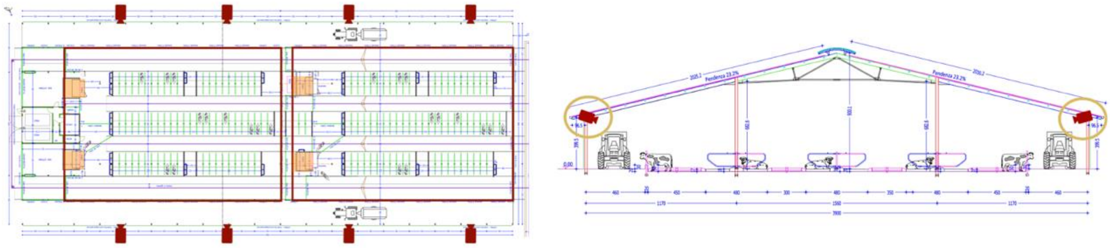

Figure 1. Left panel: upper view of the barn and location of the cameras (orange squares represent the milking robots). Right panel: side view of the barn.

Frame Selection and Dataset Composition

The data acquisition process was partitioned into multiple videos and time slots to ensure the dataset's maximum variability. The selection of the various time intervals aimed to capture all available lighting conditions, both natural and artificial, to enhance the dataset's robustness. Finally, to mitigate instances of low variance in the dataset, the videos were stored with low frame rate, i.e. 6 frames-per-second (FPS). From 144,000 frames, 400 images were selected as detailed in the following section, encompassing contexts differing in cow numbers and body positions. The resulting dataset covers 24 hours, with an average of 20.8 bounding boxes (displayed as a rectangular outline drawn around an object or a region of interest within an image) per image (Figure 2). Image analysis and management was conducted through Roboflow, an online platform with tools for managing, annotating, and preparing data for training artificial intelligence models, computer vision ones in particular (Roboflow). On this platform, the images were labelled and categorized into three separate sets (training, validation and test) based on a 70%, 20% and 10% split, respectively. The 10% allocated to the test set was essential for rigorously evaluating the model's generalization capabilities and ensuring its robust performance on unseen data. To further mitigate False Positives (FP), the test set was augmented with 1% of empty barn images and 1% of images from other barns (Ultralytics.a).

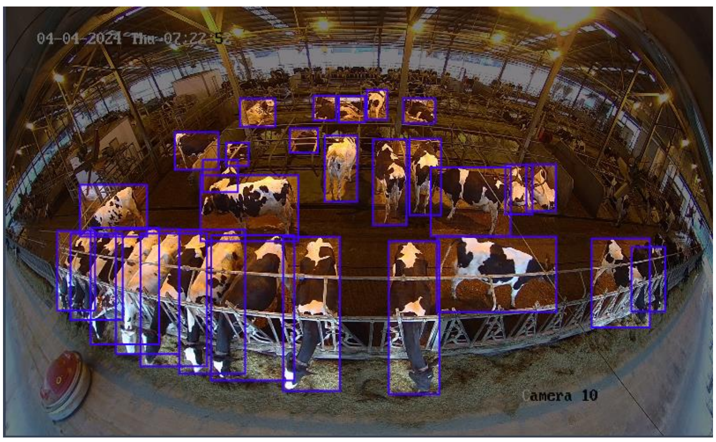

Figure 2. Example of image selected for the training dataset. The blue squares represent the ground truths used in the training process.

Pipeline Structure for Video Pre-processing

A structured pipeline of operations was devised to prepare the video dataset (V) for subsequent analysis. The objective was to extract a subset of key frames (K) that best represented the progression of the video. To achieve this, the pipeline prioritized both computational efficiency and the selection of informative frames capturing significant changes. The pipeline took a video V of length T frames (in this case, T = 3600 for a 10-minute video) as input and output a subset of key frames (K) using the following steps:

1. Canny Edge Detection with Experimental Thresholds (Canny, 1986). This facilitated the precise identification of edges E_i within each video frame (denoted as F_i, where $i = 1$ to T), highlighting the foreground subjects. Mathematically, the Canny algorithm output a binary edge map (E_i) for each frame F_i, where E_i(x, y) = 1 indicated the presence of an edge at pixel location (x, y) and Ei(x, y) = 0 in the opposite case.
2. Edges employment as Masks for the extraction of 3-Channel (RGB) Edges. The detected edges (E_i) were used to create a binary mask (M_i) for each frame F_i. This mask allowed us to isolate relevant edge features, focusing on areas of potential change, while mitigating the influence of noise and irrelevant background elements. The resulting Frame (denoted as F_i^r) was obtained by assigning the original pixel value (RGB) from frame F_i only to pixels where the corresponding location in the edge map E_i(x, y) had a value of 1. All other pixels in F_i^r were set to black.
3. Absolute Pixel-wise Difference Calculation for Top-K Frame Extraction. For each frame F_i^r , a difference metric D(F_i^r, F_{i-1}^r) was calculated by comparing each pixel of the current frame (F_i^r) and the previous frame (F_{i-1}^r). The top-k frames with the highest difference metric values were selected to enrich subset (K) with diverse and informative visual representations of the video content.

By structuring the pre-processing pipeline in this manner, we aimed to streamline computational efforts while enhancing the interpretability and richness of the resultant dataset for downstream computer vision tasks.

Object Detection model

The state-of-the-art Ultralytics YOLOv8 (Ultralytics.b) model was used for the detection and tracking of cows from annotated image datasets. Leveraging the successes of its predecessors, YOLOv8 introduces novel features and enhancements designed to elevate its performance and versatility.

Model training

The training procedure involved using annotated data to train the model. Various techniques were employed to enhance the model's performance, including *data augmentation*, *layer freezing* applied on the backbone, and *hyperparameter tuning*. Specifically, hyperparameters fine-tuning was performed with genetic algorithms embedded in YOLOv8 and involved 100 iterations of 30 epochs each. Additionally, k-fold cross-validation was applied to the validation set to reduce overfitting. Layer freezing was employed to expedite training and reduce costs by preserving pre-trained layers. Additionally, parameter adjustment for optimization was conducted during the training phase. Hyperparameters such as learning rate, momentum, optimizer settings, and the number of epochs were adjusted to find the best combination in order to achieve optimal model's performance. This iterative process involved experimenting with different configurations to maximize the effectiveness of the training process and to improve the model's accuracy and generalization capabilities.

Model Evaluation

Following the training process, the model's performance was thoroughly evaluated using various metrics and techniques. Evaluation metrics demonstrated consistent improvement in the model's performance over epochs, with decreasing training and validation losses, high precision and recall values, and consistently high mean average precision (mAP) scores at different Intersection over Union (IOU) thresholds. Additionally, holdout cross-validation techniques were utilized to evaluate the model's generalization ability and robustness, ensuring consistent performance across diverse datasets. The evaluation process provided insights into the model's effectiveness in accurately detecting and tracking cows in a barn environment, allowing for informed decisions on model deployment and further refinement if necessary.

Results and Discussion

Different models were tried out to assert the most suitable for our purpose. After various runs, it was decided to utilize the "YOLOv8*l*" version due to its trade-off between robustness and great inference speed (Table 1).

Table 1. Results and characteristics comparison over different YOLOv8 models. The mAPval50-95 was calculated during validation of the best model achieved with the specified size on our dataset. The speed refers to the inference time on a single image with dimensions of 1344 × 760 pixels.

Model	**mAPval 50-95**	**Speed (ms)**	**Params (M)**	**FLOPs (B)**
YOLOv8m	0.70355	23.859	25.9	78.9
YOLOv8l	0.71561	32.164	43.7	165.2
YOLOv8x	0.718	52.816	68.2	257.8

The speed was determined as the average inference time measured on our system over 100 inferences. Our system was equipped with an Intel(R) Core(TM) i7-6800K CPU @ 3.40GHz, 2 NVIDIA GeForce RTX 3060 12GB GPUs, and 32 GB of RAM. Various key metrics were assessed during both the training and evaluation phases (Figure 3).

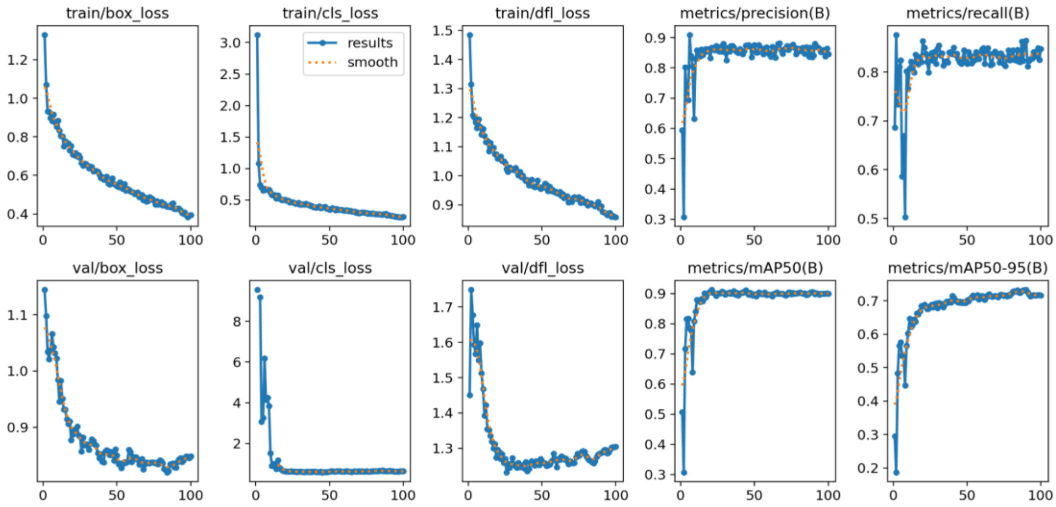

Figure 3. Key metrics achieved during the training and evaluation process.

- Precision and Recall: the model may struggle to balance high precision and high recall, possibly due to dense clusters of objects in the peripheral areas of the frames interfering with our primary goal.
- Precision and Recall Fluctuations: this may indicate occasional model overfitting to background noise, resulting in false positives. They are likely influenced by the presence of distant objects and heavy occlusion settings.
- Validation Loss: the values of validation loss are generally comparable with the training loss which suggest that the model is generalizing (the model's capacity to perform effectively on unseen instances or data points, extending beyond the specific examples it was trained on) and not overfitting the training data.
- mAP50 and mAP50-95: our approach consistently achieved a high mAP50 score of approximately 0.901, indicating excellent accuracy in "easy" detection tasks. In terms of mAP50-95, the model achieved a score of 0.733, confirming its readiness for real-world applications.

Projection of the Detected Subjects

In this study, we employed a point projection technique described by Ozella et al. (2024) to translate the detected cow centroids onto a simplified top-down barn layout. This method utilizes a calibration matrix (Wang et al., 2010), facilitating the conversion of image coordinates into real-world barn coordinates. The position (i.e., the real-world coordinates) represents the corresponding point on the barn floor map of a cow, expressed in centimetres (Figure 4).

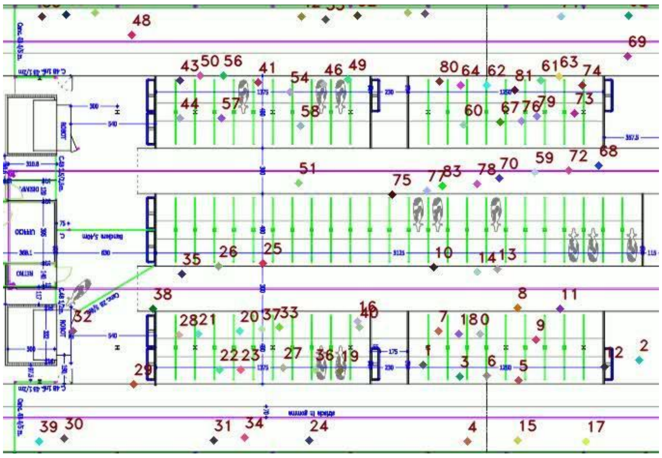

Figure 4. Results of the projection process. Each diamond shape represents the centroid of a detected cow, mapped onto the barn layout.

Using these real-world coordinates, the spatial proximity between each pair of cows can be computed to identify social interactions within the herd. This enables the construction of the temporal social network of the cows, also known as a time-varying network, where links are active only at certain points in time. The nodes of the network correspond to the cows, and the links correspond to the social distance between two cows.
To the best of the authors' knowledge, this study represents the first attempt to utilize computer vision for evaluating interactions across multiple subjects in various areas of the barn. While Guzhva et al. (2016) developed a video surveillance system for monitoring social interactions, their focus was solely on the waiting area of automatic milking stations, neglecting other aspects of dairy cattle behaviour or interactions occurring elsewhere in the barn. Additionally, Ren et al. (2021) aimed to implement a monitoring system using different technologies, including a computer vision system with a single camera, on a sample group of seven dairy cows. However, due to the restricted field of view of the camera, not all interactions among the animals were captured, limiting the system's effectiveness in discerning the nature of their interactions.
Instead, in our study, the use of multiple wide-angle cameras limited the presence of blind spots or areas of restricted visibility, allowing for a comprehensive view of interactions and analysis of the complete social network. The employment of cameras, positioned external to the cow's resting area, enabled an evaluation of the animals' interactions free from any source of disturbance. Additionally, our study required a generic setup and low-computational resources, further enhancing its replicability.

Conclusions

The model demonstrated strong performance in both validation and real-world scenarios, achieving a balance between precision and recall. While occasional overfitting to background noise occurred, its stability in later epochs indicated a successful trade-off. Consistency between validation and training loss values suggested effective generalization without overfitting. Furthermore, its high performance in common scenarios confirmed its suitability for practical use. Overall, the model met the authors' expectations, demonstrating its ability to generalize and reliably perform in diverse environments. In conclusion, the data obtained from this study will be essential for the development of future real-time analyses of the cows' social network.

Acknowledgements

The authors wish to thank Davide Vanzetti and colleagues for the precious collaboration.

References

Bouissou, M. F., Boissy, A., Neindre, P. L., and Veissier, I. 2001. *The social behaviour of cattle.* In Social behaviour in farm animals (pp. 113-145). Wallingford UK: cabi Pub-lishing.

Canny, J. 1986. "A Computational Approach to Edge Detection," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679-698, Nov. 1986

Fielding, H. R., Silk, M. J., McKinley, T. J., Delahay, R. J., Wilson-Aggarwal, J. K., Gauvin, L., Ozella, L., Cattuto, C., and McDonald, R. A. 2021. Spatial and temporal variation in proximity networks of commercial dairy cattle in Great Britain. *Preventive Veterinary Medicine*, 194, 105443.

Guzhva, O., Ardö, H., Herlin, A., Nilsson, M., Åström, K., and Bergsten, C. 2016. Feasibility study for the implementation of an automatic system for the detection of social interactions in the waiting area of automatic milking stations by using a video surveillance system. *Computers and Electronics in Agriculture*, 127, 506-509.

Gygax, L., Neisen, G., and Wechsler, B. 2010. Socio-spatial relationships in dairy cows. *Ethology*, 116(1), 10-23.

Leso, L., Ozella, L., Camiciottoli, S., Becciolini, V., and Barbari, M. 2022. Validation of proximity sensors to monitor social proximity in dairy cows: A pilot study. In Precision livestock farming '22 (pp. 41-47). *European Conference on Precision Livestock Farming*. Vienna, Austria.

McDonagh, J., Tzimiropoulos, G., Slinger, K. R., Huggett, Z. J., Down, P. M., and Bell, M. J. 2021. Detecting dairy cow behavior using vision technology. *Agriculture*, 11(7), 675.

Ozella, L., Magliola, A., Simone, V., Ghigo, M., Francesco, B., Grangetto, M., Forte, C., Montrucchio, G., and Giacobini, M. D. L. 2024. A computer vision approach for the automatic detection of social interactions of dairy cows in automatic milking systems. *ACTA Imeko*. In press

Ren, K., Bernes, G., Hetta, M., and Karlsson, J. 2021. Tracking and analysing social interactions in dairy cattle with real-time locating system and machine learning. *Journal of Systems Architecture*, 116, 102139.

Roboflow, https://roboflow.com/ (accessed February 2024).

Rocha, L. E., Terenius, O., Veissier, I., Meunier, B., and Nielsen, P. P. 2020. Persistence of sociality in group dynamics of dairy cattle. *Applied Animal Behaviour Science*, 223, 104921.

Tassinari, P., Bovo, M., Benni, S., Franzoni, S., Poggi, M., Mammi, L. M. E., Mattoccia, S., Di Stefano, L., Bonora, F., Barbanesi, A., Santolini, E., and Torreggiani, D. 2021. A computer vision approach based on deep learning for the detection of

dairy cows in free stall barn. *Computers and Electronics in Agriculture*, 182, 106030.
Ultralytics.a - Tips for Best Training Results. https://docs.ultralytics.com/yolov5/tutorials/tips_for_best_training_results/#dataset (Accessed April 2024).
Ultralytics.b - YOLOv8 model. https://docs.ultralytics.com/models/yolov8/ (Accessed April 2024).
Ultralytics.c - GitHub Repository Issue 189. https://github.com/ultralytics/ultralytics/issues/189 (Accessed April 2024).
Wang, Y. M., Li, Y., and Zheng, J. B. 2010. A camera calibration technique based on OpenCV. The 3rd *International Conference on Information Sciences and Interaction Sciences*, Chengdu, China, pp. 403-406.
Wang, Z., Hua, Z., Wen, Y., Zhang, S., Xu, X., and Song, H. 2024. E-YOLO: Recognition of estrus cow based on improved YOLOv8n model. *Expert Systems with Applications*, 238, 122212.
YOLOv8, https://yolov8.com/ (accessed February 2024).
Zhang, Y., Zhang, Q., Zhang, L., Li, J., Li, M., Liu, Y., and Shi, Y. 2023. Progress of machine vision technologies in intelligent dairy farming. *Applied Sciences*, 13(12), 7052.