Emanuelly De Souza Salles[1], Carlos Henrique Wachholz de Souza[1], Rodrigo Clemente Thom de Souza[1].

[1]Agricultural Engineering, Federal University of Paraná (UFPR) – Jandaia do Sul Campus, Brazil.

APPLICATION OF CNNS IN CATTLE COUNTING USING RPAS

A paper from the Proceedings of the

17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture

13-15 July 2026

Porto Alegre, Brazil

Abstract

Precision livestock farming demands technological solutions capable of optimizing the management of biological assets, aiming to increase productive efficiency and reduce operational errors. This work proposes the application of Convolutional Neural Networks (CNNs) for the automated counting of cattle from aerial images captured by Remotely Piloted Aircraft (RPAs). The central objective was to perform a comparative performance analysis between two architectures, YOLOv8x and YOLOv11x, focusing on the accuracy of detection and the robustness of herd counting. The methodology encompassed the training and validation of the models on a public set of aerial images of cattle, followed by the inference step on the validation set to estimate the count per image, using performance metrics such as Average Precision (mAP), inference time, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). The results showed that, although the YOLOv8x exhibits a slight superiority in frames per second (FPS) rate, the YOLOv11x demonstrated greater effectiveness in the inventory task, reducing the mean absolute error from 0.84 to 0.66 cattle per image. It is concluded that the YOLOv11x showed better overall performance for the studied scenario, especially regarding the reduction of counting error, establishing itself as a promising alternative to support herd inventory in Precision Livestock Farming applications.

Keywords: *Computer Vision; YOLO Architecture; Productive Efficiency.*

1 INTRODUCTION

The growing demand for animal-derived foods, coupled with the need for increased production efficiency and sustainability, has driven the adoption of advanced technologies in the livestock sector. Precision Livestock Farming has emerged as an innovative approach that utilizes technological tools for monitoring, management, and improved decision-making, aiming to optimize production systems. Historically, Brazil has been one of the world's most important producers and exporters of animal protein, possessing the largest cattle herd. The Brazilian cattle herd was estimated at 197 million head in 2023. Among the challenges faced in extensive livestock farming, the counting and monitoring of cattle stand out; these activities have traditionally been performed manually, making them prone to errors and associated with high labor costs and operational risks. With the advancement of Remotely Piloted Aircraft (RPAs), it has become possible to obtain high-resolution aerial imagery, enabling a new approach to animal monitoring over large areas. To overcome the limitations of manual methods, Deep Learning algorithms—particularly Convolutional Neural Networks (CNNs)—have played a prominent role by autonomously learning data characteristics and patterns. The YOLO (You Only Look Once) architecture established itself by framing object detection as a unified regression task. This study aims to comparatively evaluate the performance of the YOLOv8x and YOLOv11x architectures in the automated detection and counting of cattle in aerial images, seeking to identify which architecture offers the best balance between detection precision, counting robustness, and computational performance for application in herd inventory.

2 MATERIALS AND METHODS

2.1 PRECISION LIVESTOCK FARMING AND THE USE OF REMOTELY PILOTED AIRCRAFT

Precision Livestock Farming has evolved into Livestock 4.0, characterized by the advancement of digital technologies aimed at greater production efficiency, animal welfare, and environmental sustainability. Remotely Piloted Aircraft (RPAs) are establishing themselves as a key tool, offering advantages such as reduced operating costs and time optimization when monitoring vast areas. RGB cameras capture true-color images, enabling animal counts to be performed directly from the air.

Figure 1 - Example of an RGB image

SOURCE: Luciano Vieira Koenigkan (2024).

2.2 CONVOLUTIONAL NEURAL NETWORKS (CNNS) AND THE YOLO ARCHITECTURE

CNNs are designed to exploit spatial correlations within images using convolutional layers. YOLO is a one-stage approach that processes images in real time and directly generates bounding boxes.

YOLOv8 formalized the anchor-free strategy by predicting parameters directly for candidate points and implemented a decoupled detection head. Meanwhile, the YOLOv11 architecture integrates state-of-the-art technical modules, such as C3K2 for deep hierarchical feature extraction and the C2PSA module, which applies spatial attention to reduce background interference in complex scenes.

Figure 2 - Aerial visualization using RPAS, indicating cattle numbering and counts via computer vision.

SOURCE: Arruda et al. (2024).

3 MATERIALS AND METHODS

3.1 DATA PREPARATION AND PROCESSING

The dataset used in this study was CowSpots, obtained from a public computer vision repository; it contains aerial images of pasture and feedlot environments with pre-defined bounding boxes. The dataset was partitioned into 726 images for training and 210 images for validation. The process focused strictly on a single class of interest (cattle).

3.2 PARAMETERIZATION AND TRAINING

Development was conducted in a cloud environment (Google Colab) using NVIDIA A100-SXM4 GPU hardware acceleration. For training, image dimensions increased to 1024x1024 pixels, enabling the network to detect small-sized cattle from a high-altitude nadir view.

TABLE 1 - Experimental Configuration Parameters

Parameter	Technical Details
Model Architecture	YOLOv8x
Framework	Ultralytics 8.4.71
Initial Weights	Yolo8x.pt
Image Size (imgsz)	1024

Optimizer	Auto
Learning Rate (lr0)	0.01
Epochs	150
Random Seed	0
Confidence Threshold	0.4
IoU Threshold (NMS)	0.70
Augmentation Strategy	Ultralytics Default
Parameter	Technical Details
Model Architecture	YOLOv8x
Framework	Ultralytics 8.4.71

SOURCE: THE AUTHOR (2026).

3.3 PARAMETERIZATION AND TRAINING

Predictive performance was analyzed based on global and count metrics. Spatial precision depended on the Intersection over Union (IoU), calculated as the ratio between the Intersection Area and the Union Area of the boxes.

The classification of hits and errors generated Precision (P) and Recall (R), consolidated in the F1-Score:

$$F1 = 2 \times \frac{P \times R}{P + R}$$

$P = \text{Precisão}\ (Precision)$
$R = \text{Revocação}\ (Recall)$

For the agricultural analysis (actual counts), the main error metrics used were the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE), defined by the equations:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

$\hat{y}_i = \text{contagem predita na imagem i}$
$y_i = \text{contagem real de bovinos na imagem i}$
$n = \text{número total de imagens no conjunto de validação}$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$$

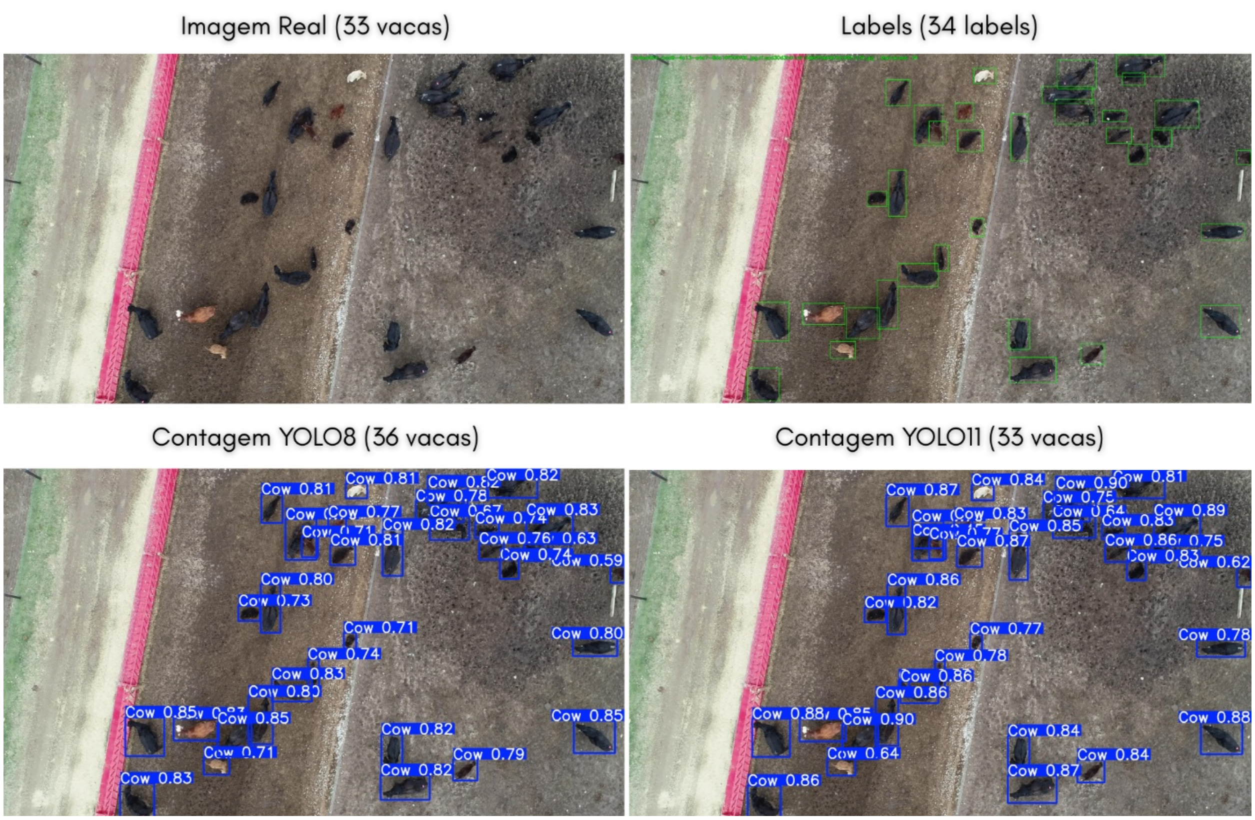
$\hat{y}_i$ = contagem predita na imagem i
y_i = contagem real de bovinos na imagem i
n = número total de imagens avaliadas

4 RESULTS AND DISCUSSION

4.1 QUALITATIVE AND SPATIAL DETECTION ANALYSIS

The models were validated by comparing real images (with labels) against automatic detections. In high-density scenarios, YOLOv8x demonstrated greater sensitivity to noise, whereas YOLOv11x yielded more refined detections.

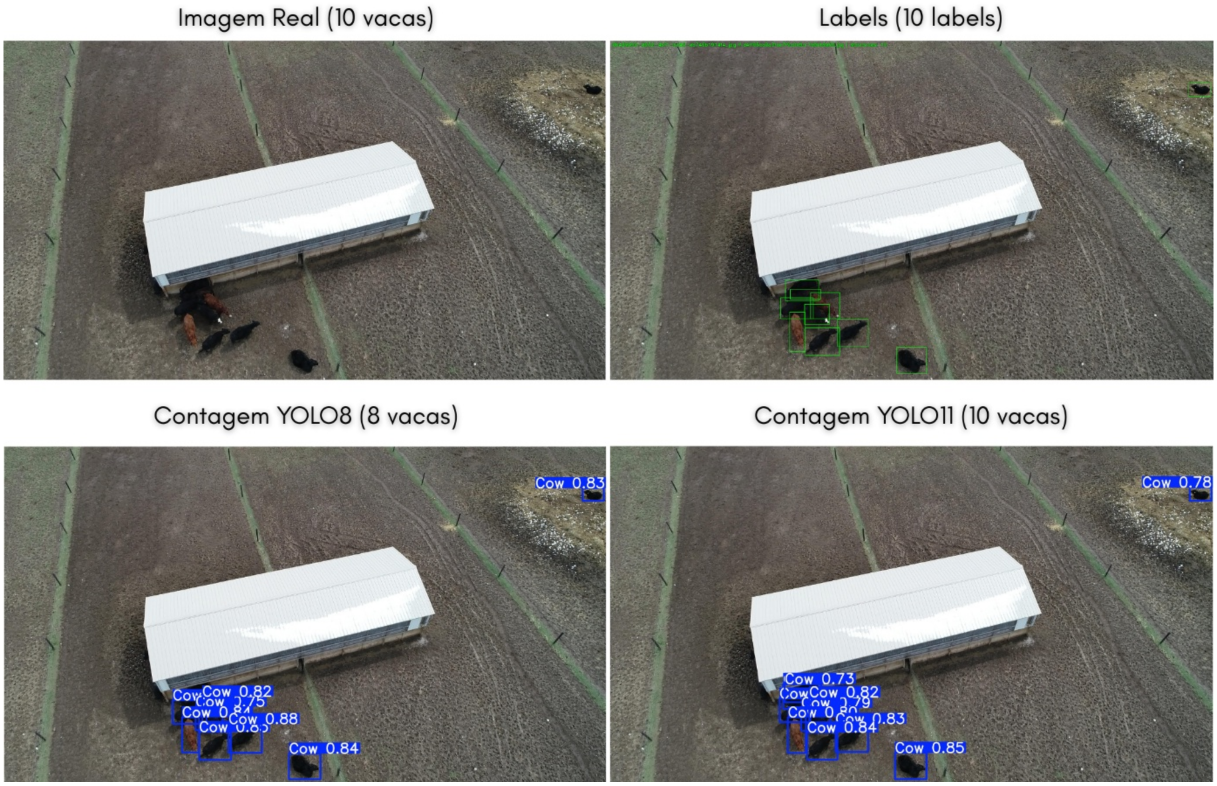
Figure 3 - Comparison image 1



SOURCE: THE AUTHOR (2026).

As observed in Figure 3, YOLOv8x overestimated the count (36 cows) due to the density, whereas YOLOv11x counted 33 cows, better distinguishing the instances.

Figure 4 - Comparison image 2

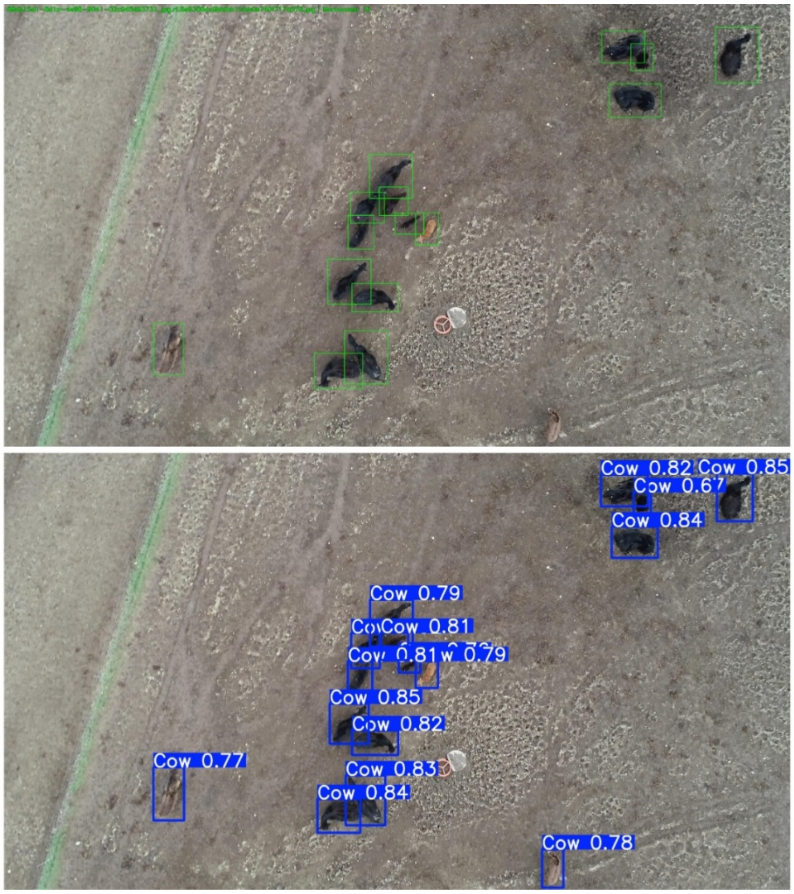


SOURCE: THE AUTHOR (2026).

In Figure 4, YOLOv8x counted 8 and YOLOv11 counted 10, demonstrating superiority under partial occlusion and shadows.

Additionally, it was detected that part of the error was not a network failure, but a lack of labeling in the database itself:

Figure 5 - Example of lack of labeling



SOURCE: THE AUTHOR (2026).

Some of the model's detections were classified as false positives, but the animals were visually present in the image and had been overlooked during annotation.

4.2 MÉTRICAS QUANTITATIVAS E DESEMPENHO DE CONTAGEM

Consolidation of data from the 210 images revealed the superiority of YOLOv11x in reducing herd-related logistical discrepancies.

TABLE 2 - Comparison of Metrics.

Metrics	YOLOv8x	YOLOv11x
mAP@0.5:0.95	0.5663	0.5703
mAP@0.5	0.9448	0.9455
Precision	0.9281	0.9268
Recall	0.9281	0.9270
F1-score	0.9232	0.9269

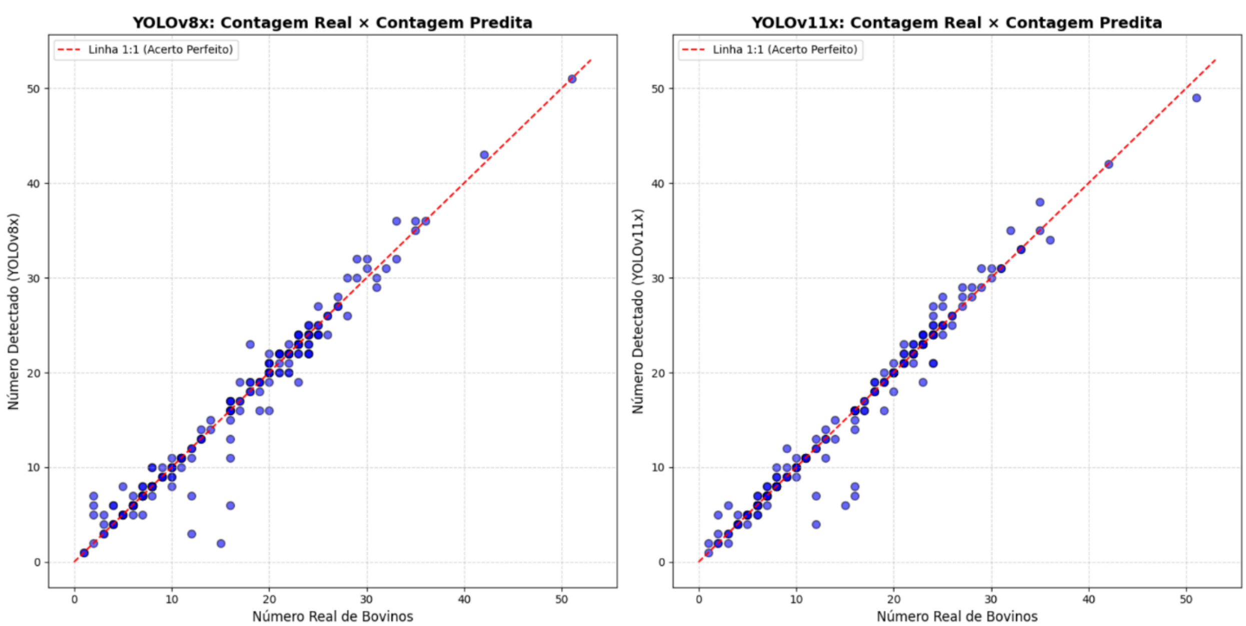
Inference Time (ms/img)	10.13 ms	9.99 ms
Total Time (ms/img)	15.93 ms	17.38 ms
FPS (Approximate)	62.8 frames/segundo	57.5 frames/segundo
Count MAE	0.838 bovinos/ imagem	0.657 bovinos/imagem
Total Detected	3292 bovinos	3302 bovinos
Total Actual	3314	3314
Cumulative net difference	22 bovinos	12 bovinos
Sum of absolute errors per image	176	138
Absolute error per 100 cattle	5,31	4,16
RMSE	1,807	1,555

SOURCE: THE AUTHOR (2026).

Although YOLOv8x holds a slight advantage in processing speed (62.8 FPS versus 57.5 FPS), YOLOv11x reduced the MAE from 0.838 to 0.657 cattle per image (a 21.6% drop) and the RMSE from 1.807 to 1.555 (a 13.9% reduction).

The dispersion of counting errors visually demonstrates this improvement:

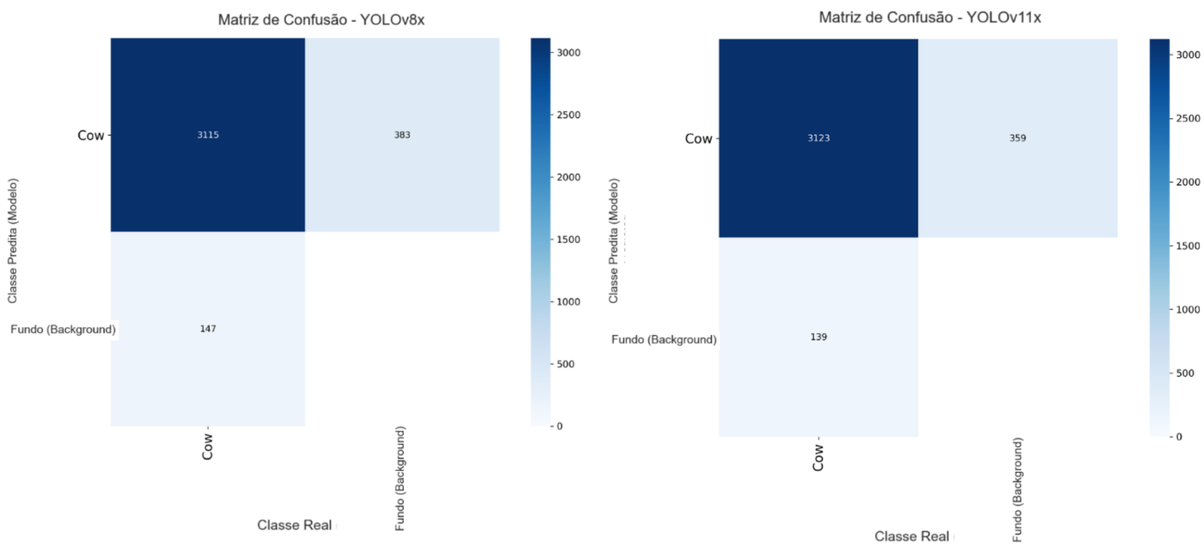
GRAPHIC 1 - Discrepancy between actual count and count



SOURCE: THE AUTHOR (2026).

Graph 1 shows that YOLOv11 exhibits much less dispersion around the zero-error line. The absolute confusion matrices confirm the robustness in rejecting background elements:

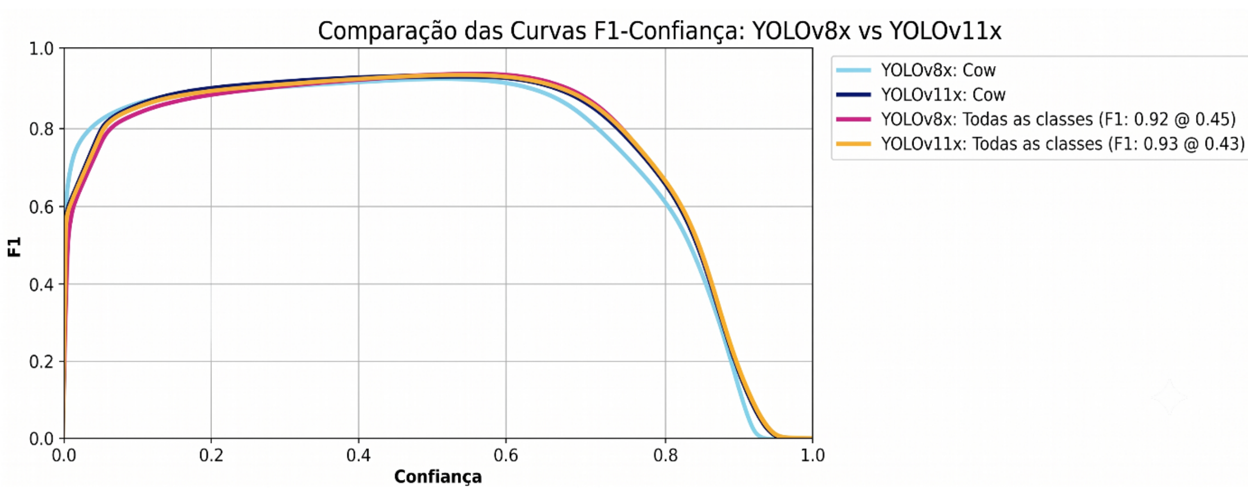
GRAPHIC 2 - Confusion matrix



SOURCE: THE AUTHOR (2026).

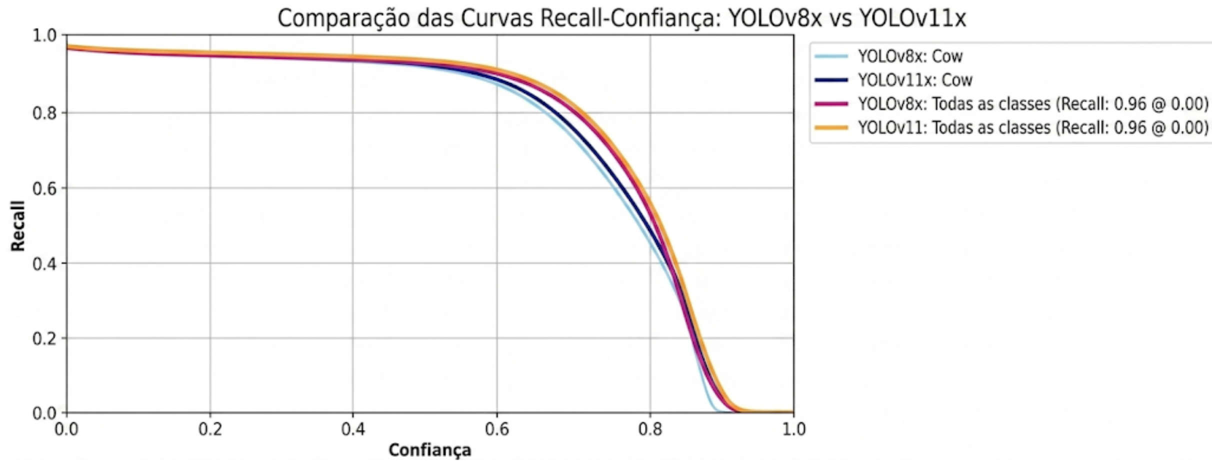
Finally, an analysis of the algorithm's performance curves showed that the F1-Score demonstrated the optimal balance between the metrics:

Figure 6 - F1-score vs. confidence curve



SOURCE: THE AUTHOR (2026).

Figure 7 - Recall vs. Confidence Curve



SOURCE: THE AUTHOR (2026).

In Figures 6 and 7, YOLOv11x achieves a slightly higher F1 score (0.93) with an excellent area under the Precision-Recall curve.

5 CONCLUSION

The application of CNNs to images captured by UAVs proved highly effective for counting cattle herds. Comparative analysis revealed that the YOLOv11x architecture outperformed YOLOv8x in the crucial aspect of livestock inventory. YOLOv11x demonstrated superior discriminative capability, mitigating false positives caused by soil textures and shadows, which resulted in a 21.6% reduction in Mean Absolute Error (MAE). It is concluded that the latest version of the algorithm reinforces the viability of autonomous systems in precision livestock farming. Future work should focus on validating these architectures in unstructured outdoor environments and investigating instance segmentation to estimate the biomass and body condition of grazing herds.

6 REFERENCES

ABIEC – Brazilian Beef Exporters Association. Beef Report 2024: Profile of cattle farming in Brazil. São Paulo: ABIEC, 2024.

BARBEDO, J. G. A.; KOENIGKAN, L. V.; SANTOS, P. M. Counting cattle in UAV images—Dealing with clustered animals and animal/background contrast changes. Sensors, v. 20, n. 7, p. 2126, 2020.

COWSPOTSCALPOLY. Cowspots Dataset. Available at: https://universe.roboflow.com/cowspotscalpoly/cowspots. Accessed on: Oct. 18, 2025.

HE, K. et al. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.

LI, Z. et al. Method for dairy cow target detection and tracking based on lightweight YOLO v11. Animals, v. 15, n. 16, p. 2439, 2025.

REDMON, J. et al. You only look once: Unified, real-time object detection. Proceedings of IEEE CVPR, 779–788, 2016.

SAPKOTA, R.; KARKEE, M. Ultralytics YOLO Evolution: An Overview of YOLO 26, YOLO 11, YOLO v8 and YOLO v5 Object Detectors. arXiv, 2025.