Cow Body Part Detection using Faster R-CNN with Transfer Learning and COCO Evaluation

Bárbara Paola Alcántara Vega - A01799609

Abstract

This project presents a deep learning approach to detecting specific cow body parts, namely the rump and udder, using a fine-tuned Faster R-CNN model. The dataset consists of labeled images annotated with bounding boxes in COCO format, collected and prepared through Roboflow. To enhance model performance, several optimizations were introduced, including early stopping, longer training epochs, and model checkpoint saving. The model's performance is evaluated using COCO metrics, specifically Average Precision (AP) and Average Recall (AR) across different IoU thresholds. Preliminary results with only using the rump and udder indicate promising detection accuracy even with limited training epochs, and future iterations aim to refine bounding box precision for other parts of the body such as legs and feet in addition to overall body characteristics such as body depth, strength and dairy form along with the expansion of the dataset with lower resolution images, different lighting and cow positions order to increase robustness across varying image conditions.

Introduction

Automated detection of animal body parts holds significant potential in livestock management, enabling improved monitoring of health, posture, and productivity. In dairy cattle, identifying anatomical regions such as the rump and udder can aid in health diagnostics, milk yield estimation, and behavior analysis.

UDDER COMPOSITE INDEX	
Udder Depth	Front Teat Placement
Fore Udder Attachment	Rear Teat Placement
Udder Cleft	Teat Length
Rear Udder Height	Stature
Rear Udder Width	
DAIRY CAPACITY COMPOSITE INDEX	
Dairy Form	
Strength	
FEET & LEGS COMPOSITE INDEX	
Foot Angle	Feet & Legs Score
Rear Legs - Side View	Stature
Rear Legs - Rear View	
BODY WEIGHT COMPOSITE INDEX	
Stature	Rump Width
Body Depth	Dairy Form
Strength	

Figure 1. Frequently used composites in linear trait evaluations (Holstein Association USA, Inc., 2025).

Udder Composite and Feet & Legs Composite were developed in order to increase individual lactation averages, productive life, and lifetime production of fat and protein. Adittionally, the composites take into account a negative weight on stature results in the decoupling of the composites from stature, allowing breeders to improve udders and feet & legs without making their cows taller.

Each evaluation is made based on the standards indicated in the Linear Descriptive Traits manual from the Holstein Association USA, Inc., 2024, by a specialist in dairy cattle evaluation, however, the evaluations require many good quality photos and/or formal visits in order to score an animal, and the costs originating from scoring each animal in order to do a phenotypic evaluation for most of the herd with a breeding program in mind, can be quite expensive.

With recent advances in technology, the alternative of automating most of the scoring process by visual machine learning detection presents a great opportunity to facilitate evaluations in smaller farms in order to begin breeding programs that increase overall herd health and productivity. Such detection poses challenges due to lighting variations, occlusions, and differing animal poses.

This project explores the use of a region-based convolutional neural network (Faster R-CNN) to identify cow body parts from images. Unlike pre-trained YOLO models, this implementation is built entirely from scratch in PyTorch, providing greater control over the model architecture, dataset processing, and evaluation methods.

Related Work

Modern object detection techniques such as YOLO, SSD, and Faster R-CNN have demonstrated remarkable success in a wide range of applications. YOLO (You Only Look Once) models offer fast real-time detection but rely heavily on pre-defined architectures. SSD (Single Shot Multibox Detector) performs detection in a single forward pass, trading off precision for speed. In contrast, Faster R-CNN provides a more robust region proposal mechanism, offering higher accuracy for tasks where spatial precision is critical.

This project leverages the Faster R-CNN architecture as a balance between accuracy and flexibility, allowing fine-tuning for specific use cases such as detecting anatomical cow regions to allow future incorporation of scoring for each body part. The model was trained using a custom dataset labeled in COCO format using RoboFlow, departing from reliance on YOLO-specific frameworks.

Methodology

The dataset used in this project, 'vacas.v1i.coco', consists of cow images labeled with bounding boxes for distinct anatomical regions, primarily the rump and udder. The dataset was preprocessed into COCO JSON annotations and split into training, validation, and testing sets (80/10/10 ratio). The images were then normalized and converted into tensors before being passed into the model.

The Faster R-CNN model, implemented with a ResNet-50 backbone, was fine-tuned to predict bounding boxes for the target classes. The model's output includes the predicted bounding box coordinates, class labels, and confidence scores. An AdamW optimizer was used with a learning rate of 5e-4, and training was carried out over several epochs with early stopping to prevent overfitting.

Checkpoint saving was integrated to preserve the best-performing model weights based on validation performance. First tests demonstrated some bounding boxes were incorrectly positioned with a low percentage of certainty corresponding to a body part, so to discard low-confidence detections, a 0.75 detection threshold was added, ensuring only robust bounding boxes are retained for inference.

Model evaluation is performed using the COCO evaluation toolkit, reporting Average Precision (AP) and Average Recall (AR) at different Intersections over Union (IoU) thresholds. These metrics provide a comprehensive assessment of the model's detection precision and completeness.

Results

After initial tests, the epochs were set to 50 in order to locate the epoch with best results, ending up with an early stopping at epoch 29. The best model performance was reached on epoch 9, with its preliminary COCO evaluation results demonstrating successful object detection capability:

```
Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.619
Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=100 ] = 0.952
Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=100 ] = 0.752
Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.323
Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.600
Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.657
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=  1 ] = 0.665
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets= 10 ] = 0.681
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.681
Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.333
Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.664
Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.704
```

Validation mAP: 0.6193

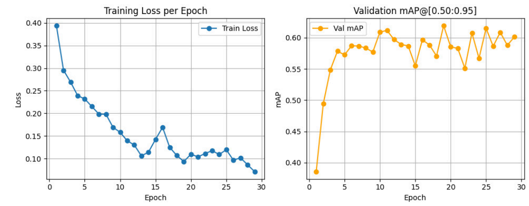

Figure 2. Training loss per epoch and validation mAP across epochs.

After importing the already trained model, visualizations are included after inserting the image we wish to segmentate, illustrating bounding box predictions over test images and the model's ability to differentiate between cow body parts.

Discussion

Initial experiments indicate that the Faster R-CNN approach effectively learns to localize anatomical cow regions, even with minimal training data. The implementation of early stopping helps mitigate overfitting, ensuring that the model's performance generalizes across unseen data. Model checkpointing ensures reproducibility and recovery of the best-performing weights, allowing consistent inference results.

The integration of a 0.75 confidence threshold eliminated most of the false positives and improved overall detection reliability, where the false positives that were detected are related to cow individuals in background settings from which the dataset did not contain elements for a more robust training, such as cows on grass.

Conclusion and Future Work

This study successfully establishes a foundational object detection framework for identifying cow body parts using Faster R-CNN. By introducing early stopping, checkpoint saving, and a confidence threshold, the model demonstrates potential for high-precision agricultural image analysis.

Future experiments will focus on more body parts, scoring for each body part and parameter tuning to achieve stable convergence and improved bounding box accuracy and body part scoring in order to help farmers identify a cow's linear trait overall score and individual body part score with an addition of the comparison of performance between epochs and analyze convergence trends in loss and mAP.

Future improvements will include hyperparameter tuning, advanced data augmentation, and model deployment via a separate inference pipeline. The inference system will enable users to upload images and automatically obtain visualized detections with bounding boxes drawn around the detected regions of interest.

References

Holstein Association USA, Inc. (2025). Linear type evaluations. https://www.holsteinusa.com/genetic_evaluations/ss_linear.html}