# Digital Twin-Driven Teat Localization and Shape Identification for Dairy Cow (Student Poster)

**Aarushi Gupta *, Yuexing Hao *†, Yuting Yang *, Tiancheng Yuan *,**
**Matthias Wieland, Parminder S. Basran, Ken Birman**
[1]Cornell University
Ithaca NY, USA
{ag2288, yh727, yy354, ty373, mjw248, psb92, kpb3}@cornell.edu

## Abstract

Dairy owners invest heavily to keep their animals healthy. There is good reason to hope that technologies such as computer vision and artificial intelligence (AI) could reduce costs, yet obstacles arise when adapting these advanced tools to farming environments. In this work, we applied AI tools to dairy cow teat localization and teat shape classification, obtaining a model that achieves a mean average precision of 0.783. This could then be used as a first step towards automating and accelerating the detection and treatment of hyperkeratosis, mastitis, and other medical conditions that significantly burden the dairy industry.

## Introduction

The condition of teats in dairy cows is widely used as predictor of not just animal health but also milk quality (A J Seykora 1985). Studies have shown that teat-end shape is related to a cow's mastitis resistance (Lojda and Matouskova. 1976), the somatic cell count and percent 2-min milk (Seykora and McDaniel 1985).

Traditional teat shape assessment requires close visual observation. While veterinarians routinely perform this task, dairy workers in small farms find the task burdensome, reducing accessibility of a potentially valuable predictive tool. On large farms, the task is simply impractical: thousands of cows might be present, and managed by a limited number of dairy workers. This precludes daily examination of cows' teats to detect changes that might be early precursors of animal health issues. Our work focuses on two questions: Automated and accurate teat shape assessment, but also deployment: How can computer vision tools of this kind actually be used on modern dairy farms?

Machine learning (ML) fits well in addressing the dairy owners' need for precise and automatic teat-shape related metrics analysis. While there are existing works about cow teat classification and identification from a veterinary perspective (Mein et al. 2001), no prior works used machine learning models to identify teat shape. Our project provides a machine learning solution that includes data collection, labeled with expertise knowledge, and trained ML models. We experiment with a selection of models to perform the cows' teat localization and teat shape classification. Our model can achieve 0.783 Average Precision, and provide results with graphs like Figure 1 for the users to view the cow's teat image with the classification results attached.

## Proposed Methodology

To collect training data, we installed a video camera in a rotary milking parlor. The parlor consists of a series of stalls which the cow enters to be washed and milked, then released into the main barn. Our camera continuously records video of the cows' teat and udder as the parlor rotates. We then extract *keyframes*: ML-selected still images in which the teats are clearly visible. Using keyframe timestamps, we match the cow's stall ID and ground truth scoring against the collected images.

Next we train ML models to perform automated teat localization, and then perform teat shape classification within the selected sub-images. We explore several ML models that can perform both functionalities within a single model. Faster-RCNN (Ren et al. 2015) is a two-stage detector, which relies on a Regional Proposal Networks (RPN) to propose many potential regions of interest (RoI), and then applies a classifier backbone. YOLO-F (Chen et al. 2021), a modified version of YOLO, is a single-stage detector. DINO (Zhang et al. 2022), a modified version DETR (Carion et al. 2020), uses transformer architecture.

## Experimental Evaluation

### Datasets

We collected the video datasets from an Upstate New York dairy farm on February 21st, 2022. The video streams are collected through a GoPro camera at a perpendicular angle from the teats. A veterinarian scored the teat shape manually, following the Seykora and Daniel (Seykora and McDaniel 1985) guidelines. We extract a subset of 2-3 keyframes for certain cows to obtain a more distinct and detailed representation of all their teats in the images.

---

*These authors contributed equally, ordered alphabetically by last name.

†corresponding author


Figure 1: (Left) Using LabelMe to draw bounding boxes and precisely annotate cow teats with the scores provided by the expert veterinarian. (Middle) ML predictions of each teat shape's score. (Right) A plot of the bounding box train loss vs. training iterations

LabelMe, by default, generates one separate JSON file with the annotation information for each image in our dataset. For conversion to the standard COCO object detection dataset format, we adopted a script for aggregating all annotation information into two JSON files, which align with the images from the training set and validation set. We applied a train test split of 85% and 15%.

## Experimental Results

We run all our experiments on a platform that contains an AMD Ryzen 7 5800X3D and a NVIDIA RTX 4090. We use mean average precision (mAP) as our performance metric. For COCO datasets, mAP is calculated for Intersection over Union(IoU) values. The IoU is given by the area of overlap divided by the area of union in between the ground truth bounding box and the predicted bounding box. Our dataset consists of only small-scale objects, whose areas are often smaller than $32 \times 32$ pixels. So during training, we focus on the mAPs. We follow the aforementioned scoring system and find a total of 4 distinct class labels $[1, 3, 7, 8]$ from worst to best teat shape conditions.

As seen from Table1, DINO delivers the best performance and only consumes around 110% in runtime, compared to the baselines.

| model name | validation mAPs | avg inference time |
|---|---|---|
| DINO | 0.783 | 628 ms |
| YOLO-F | 0.634 | 598 ms |
| Faster RCNN | 0.573 | 576 ms |

Table 1: List of model performance, mAPs stands for the bounding boxes mean average precision for small objects

## Conclusion

We explored teat localization and shape classification using ML models using a preliminary dataset of 348 images with 968 objects from 4 distinct classes. We explored different object detectors across various architectures and find DINO performs the best overall. Our automated digital-twin approach was shown to yield accurate classifications.

## Acknowledgments

This project is sponsored by the Cornell Institute for Digital Agriculture (CIDA).

## References

A J Seykora, B. T. M. 1985. Udder and teat morphology related to mastitis resistance: a review. *Journal of dairy science vol. 68,8*.

Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; and Zagoruyko, S. 2020. End-to-end object detection with transformers. In *European conference on computer vision*, 213–229. Springer.

Chen, Q.; Wang, Y.; Yang, T.; Zhang, X.; Cheng, J.; and Sun, J. 2021. You only look one-level feature. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 13039–13048.

Lojda, M. S., L.; and Matouskova., O. 1976. The shape of the teat and teat-end and the location of the teat canal orifice in relation to subclinical mastitis in cattle. *Aeta Bet. Brno.*, 181–185.

Mein, G.; Neijenhuis, F.; Morgan, W.; Reinemann, D.; Hillerton, E.; Baines, J.; Ohnstad, I.; Rasmussen, M.; Timms, L.; Britt, J.; Farnsworth, R.; and Cook, N. 2001. Evaluation of Bovine Teat Condition in Commercial Dairy Herds: 1. Non-Infectious Factors.

Ren, S.; He, K.; Girshick, R.; and Sun, J. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28.

Seykora, A. J.; and McDaniel, B. T. 1985. Heritabilities of Teat Traits and their Relationships with Milk Yield, Somatic Cell Count, and Percent Two-Minute Milk1. *Journal of Dairy Science*, 68(10): 2670–2683.

Zhang, H.; Li, F.; Liu, S.; Zhang, L.; Su, H.; Zhu, J.; Ni, L. M.; and Shum, H.-Y. 2022. Dino: Detr with improved denoising anchor boxes for end-to-end object detection. *arXiv preprint arXiv:2203.03605*.