PRISM-CAFO: Prior-conditioned Remote-sensing Infrastructure Segmentation and Mapping for CAFOs

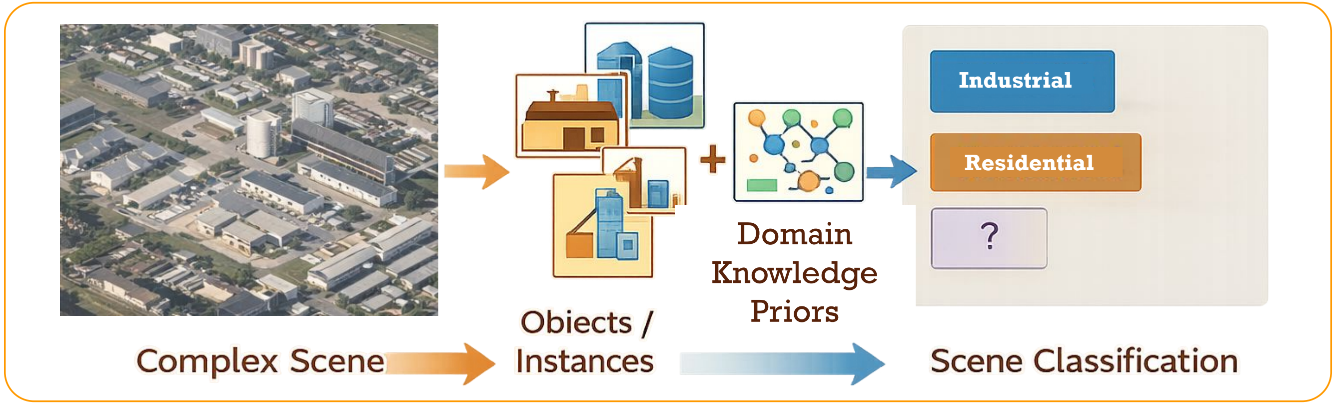
Oishee Bintey Hoque*; Nibir Chandra Mandal*; Kyle Luong; Amanda Wilson; Samarth Swarup; Madhav Marathe; Abhijin Adiga

Presented By: Oishee Bintey Hoque

Funded by: NSF and USDA-NIFA under the AI Institute: Agricultural AI for Transforming Workforce and Decision Support (AgAID) award No. 2021-67021-35344, and by University of Virginia Strategic Investment Fund award number SIF160.

Problem Statement

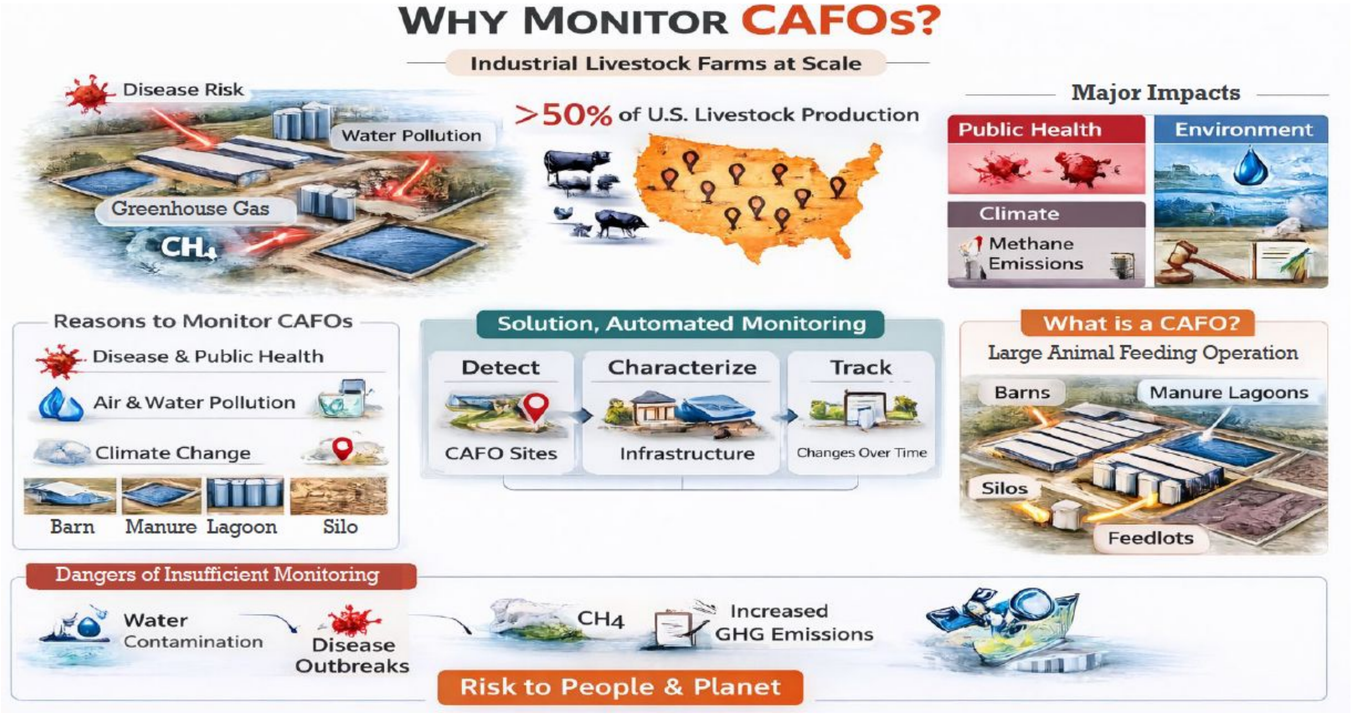
- Can object-centric representations with domain priors yield more accurate and interpretable scene classification than vision-only models?



Studied Domain

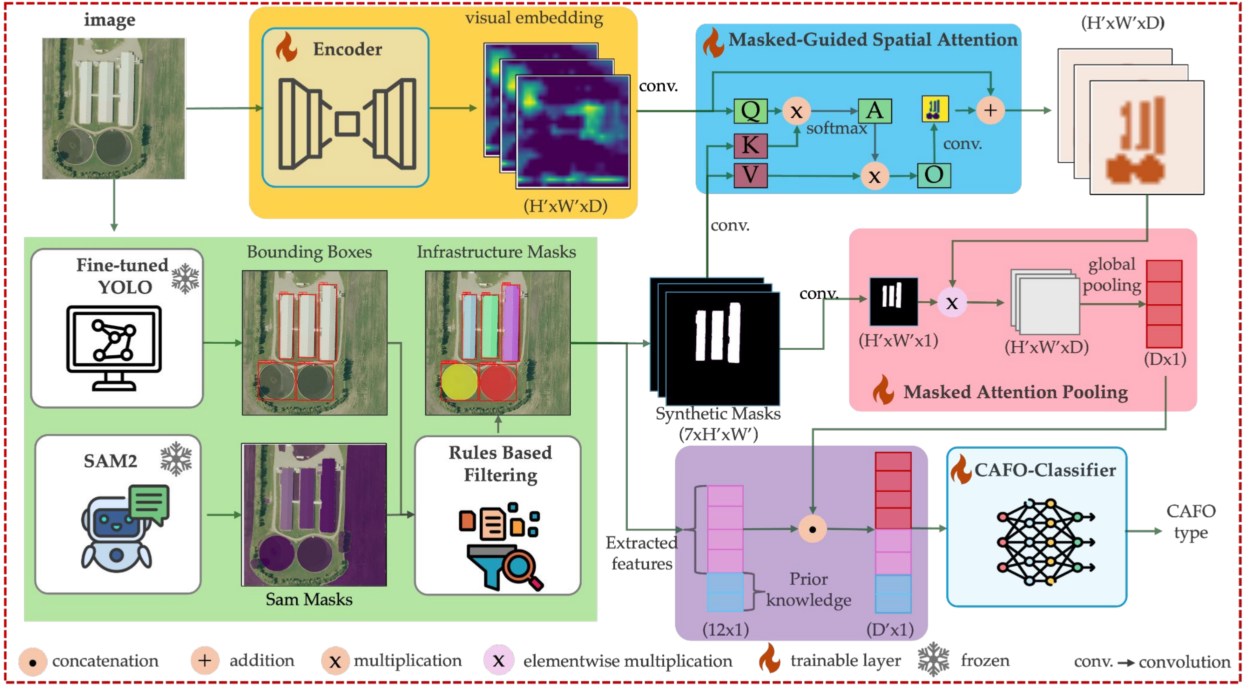
- Rural U.S. agricultural landscapes
- NAIP aerial imagery
- **Focus**: Concentrated Animal Feeding Operation (CAFO) infrastructure detection & categorization
  - High structural variability in CAFO infrastructure (size, composition, layout)
  - Strong visual overlap with non-CAFO agricultural/industrial sites
  - Lack of component-level annotations
- **Goal**: Accurate, infrastructure-aware, interpretable CAFO type classification

Why Monitor CAFOs?



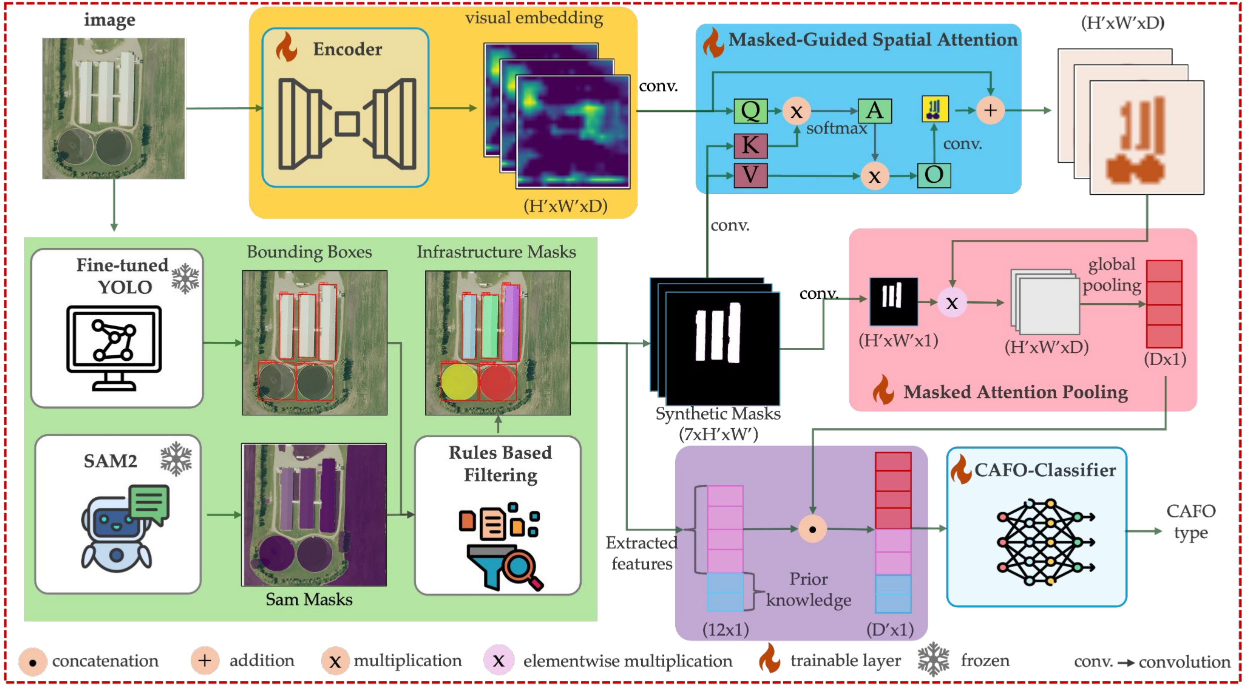
PRISM-CAFO Pipeline

Image → Encoder → Spatial Embeddings



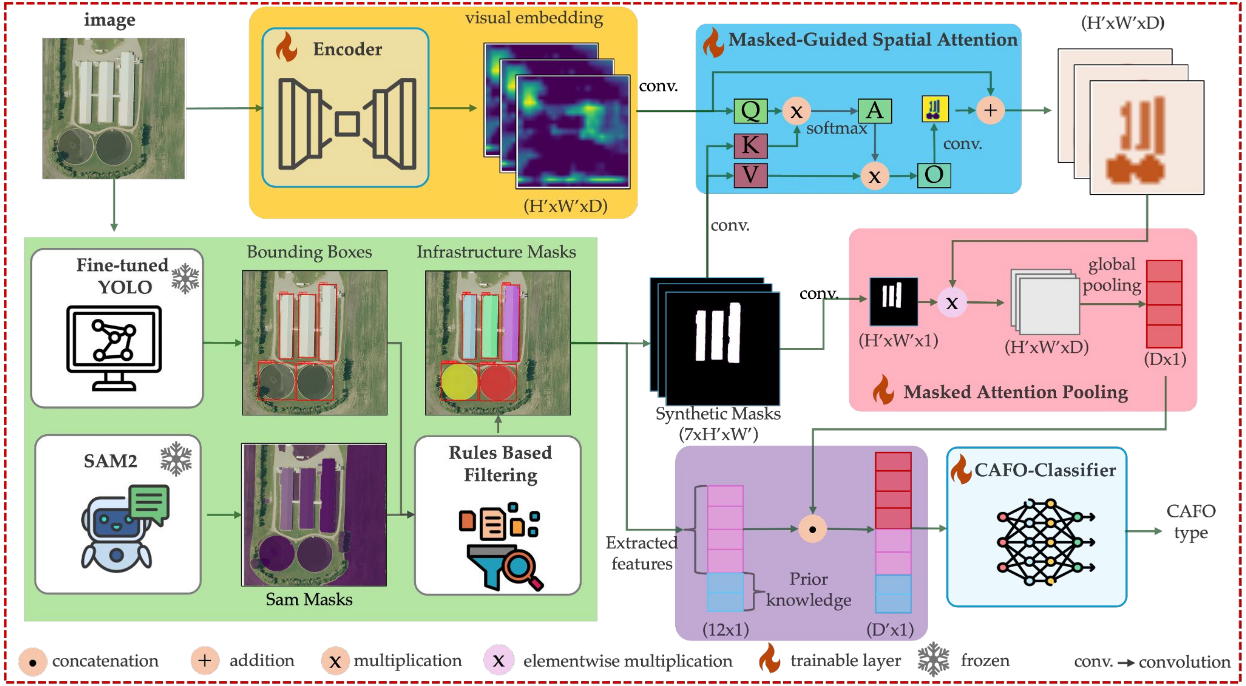
PRISM-CAFO Pipeline

Image → Encoder → Spatial Embeddings → Mask-Guided Attention (using detected infrastructure masks)



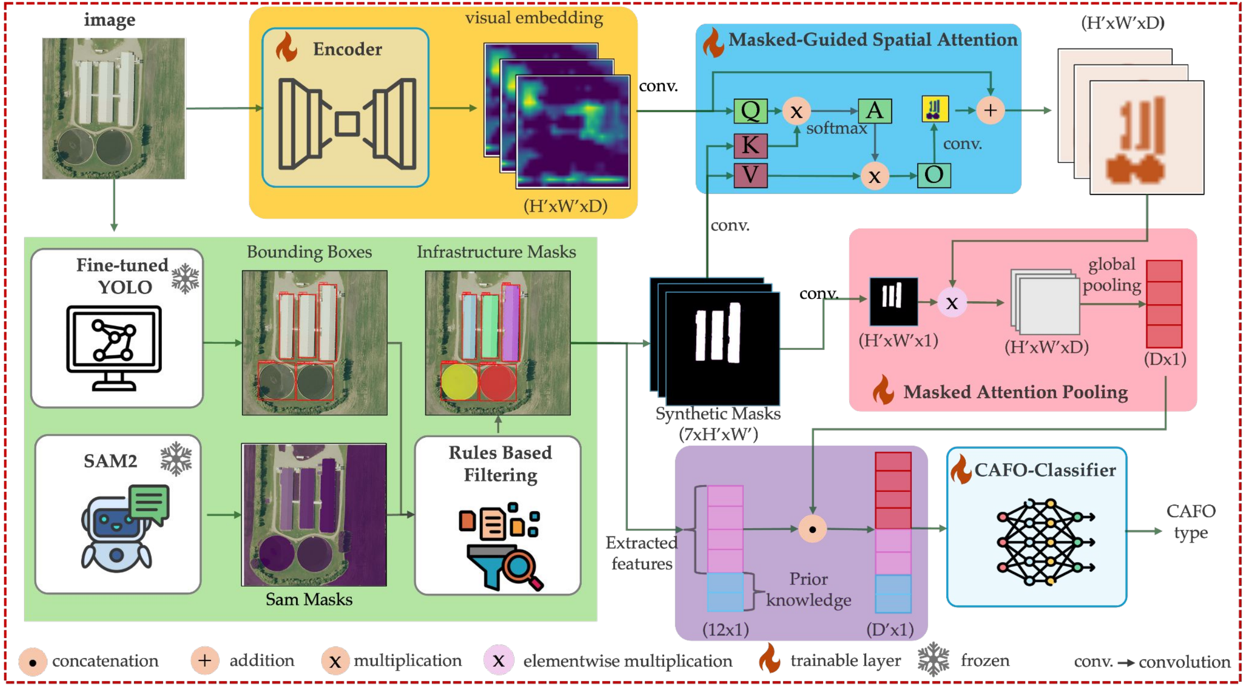
PRISM-CAFO Pipeline

Image → Encoder → Spatial Embeddings → Mask-Guided Attention (using detected infrastructure masks) → Masked Attention Pooling

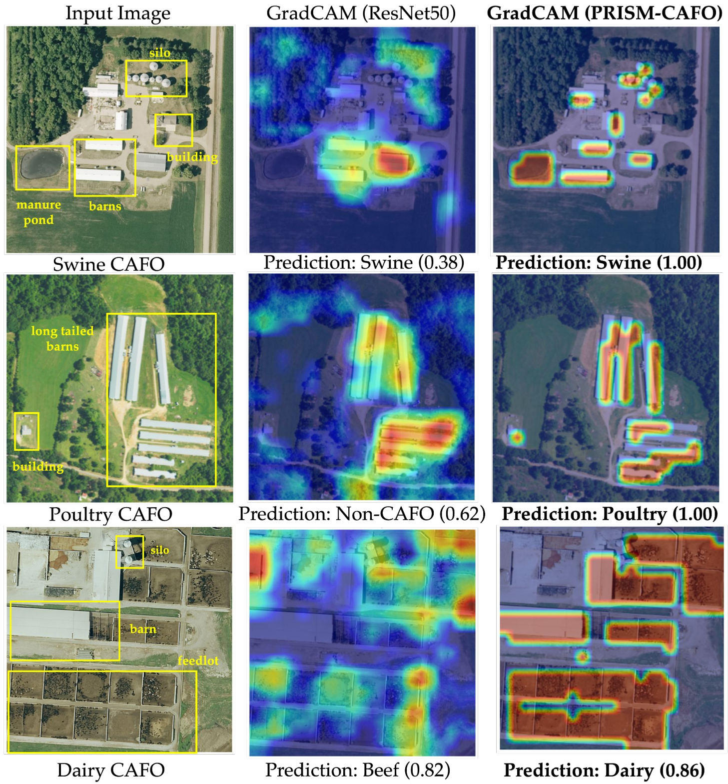


PRISM-CAFO Pipeline

Image → Encoder → Spatial Embeddings → Mask-Guided Attention (using detected infrastructure masks) → Masked Attention Pooling → Domain Prior Fusion → Classification



Interpretability



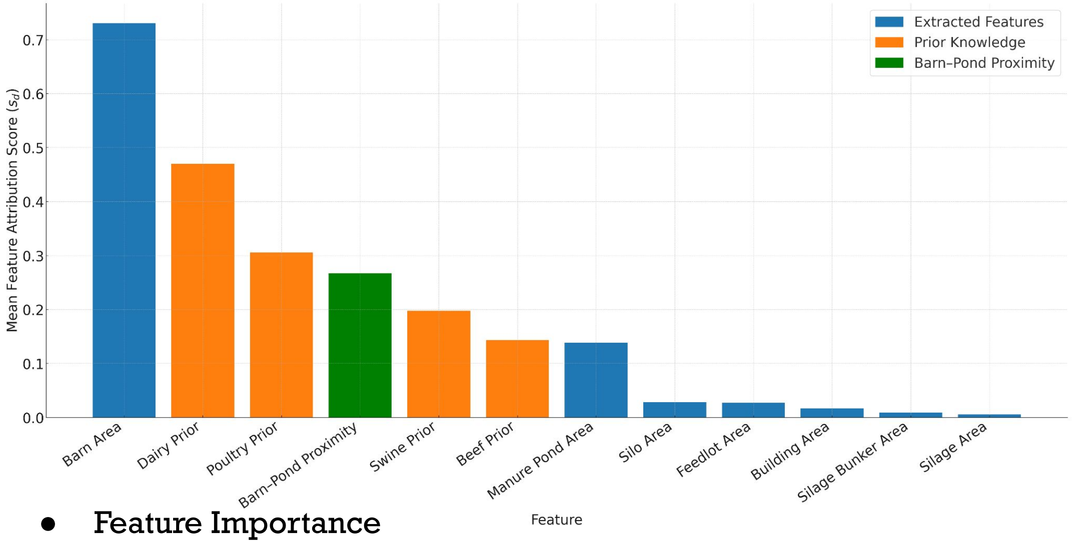
Result Summary

✔ Improves baseline (**15%** for spatial)

✔ Maximum improvement for CNN based models

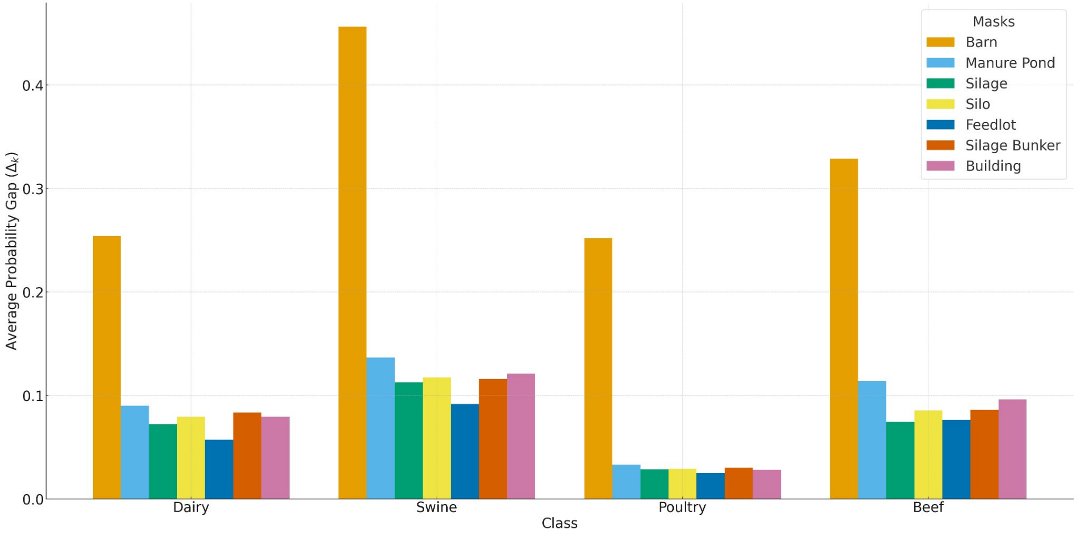
Model	Random (80%-20%) Split					Spatial Split				
	Swine	Poultry	Dairy	Beef	Neg	Swine	Poultry	Dairy	Beef	Neg
CLIP (ViT-B/32)	0.869	0.719	0.449	0.482	0.952	0.673	0.636	0.280	0.143	0.951
CLIP+PRISM-CAFO	0.854↓	0.707↓	0.418↓	0.428↓	0.962↑	0.550↓	0.554↓	0.341↑	0.150↑	0.968↑
DINOv2 ViT-B/16	0.900	0.793	0.577	0.624	0.963	0.758	0.737	0.406	0.182	0.958
DINOv2+PRISM-CAFO	0.899↓	0.808↑	0.579↑	0.547↓	0.981↑	0.735↓	0.747↑	0.446↑	0.159↓	0.975↑
RemoteCLIP (ViT-B/32)	0.875	0.755	0.466	0.489	0.960	0.700	0.708	0.354	0.188	0.958
RemoteCLIP+PRISM-CAFO	0.850↓	0.673↓	0.499↑	0.353↓	0.960=	0.603↓	0.664↓	0.311↓	0.111↓	0.971↑
Swin-B	0.881	0.772	0.573	0.556	0.958	0.692	0.706	0.446	0.159	0.951
Swin-B+PRISM-CAFO	0.928↑	0.848↑	0.678↑	0.689↑	0.981↑	0.874↑	0.758↑	0.584↑	0.164↑	0.986↑
ViT-B/16	0.895	0.775	0.578	0.596	0.962	0.769	0.730	0.439	0.249	0.955
ViT-B/16+PRISM-CAFO	0.917↑	0.839↑	0.612↑	0.621↑	0.989↑	0.715↓	0.735↑	0.313↓	0.222↓	0.979↑
EfficientNet-B0	0.863	0.744	0.466	0.573	0.942	0.712	0.695	0.200	0.190	0.941
EfficientNet-B0+PRISM-CAFO	0.929↑	0.864↑	0.650↑	0.680↑	0.980↑	0.855↑	0.800↑	0.540↑	0.322↑	0.967↑
EfficientNet-B3	0.859	0.722	0.519	0.548	0.936	0.642	0.737	0.350	0.205	0.936
EfficientNet-B3+PRISM-CAFO	0.939↑	0.872↑	0.629↑	0.739↑	0.798↑	0.780↑	0.539↑	0.539↑	0.306 ↑	0.975↑
ResNet18	0.865	0.702	0.452	0.544	0.933	0.608	0.657	0.220	0.169	0.921
ResNet18+PRISM-CAFO	0.916↑	0.829↑	0.657↑	0.709↑	0.984↑	0.852↑	0.868↑	0.522↑	0.253↑	0.991↑
ResNet50	0.866	0.735	0.486	0.551	0.937	0.680	0.685	0.301	0.124	0.932
ResNet50+PRISM-CAFO	0.876↑	0.811↑	0.642↑	0.684↑	0.963↑	0.833↑	0.850↑	0.531↑	0.090↓	0.987↑

Analysis: Significance of Components



- Feature Importance
  - Gradient–activation analysis
  - Barn area dominates feature importance (**0.74**)
  - Domain priors contribute strongly (**0.15–0.47**)
  - Barn–pond proximity adds discriminative signal (**0.27**)

Analysis: Significance of Components



- Domain knowledge Importance
  - Barn masks highest probability impact (**0.45** swine)
  - Manure lagoons critical for swine/beef classification
  - Multi-component diversity improves dairy recognition

Summary

✔ Detector-anchored infrastructure extraction using **YOLOv8** + **SAM2** + **geometric filtering**

✔ Prior-conditioned, mask-guided **spatial attention** for object-centric classification

✔ **Infrastructure** (object) -aware scene reasoning

✔ Consistent gains across backbones; up to **15%** F1 improvement (Swin-B)

✔ **Object-level interpretability** via mask attributions & Grad-CAM

✔ Systematic analysis of feature & mask importance

Thank you!!

Paper Link:
https://arxiv.org/pdf/2601.11451