

CAFOSat: A Strongly Annotated Dataset for Infrastructure-Aware CAFO Mapping Using High-Resolution Imagery

Supplementary Material

A. CAFOSat Overview.

An overview of the data processing pipeline is provided in Figure 2. The pipeline described in this section produces CAFOSat, a comprehensive, ML-ready dataset for CAFO detection and infrastructure mapping across the contiguous United States. The full dataset comprises **39,257 base image patches** (833×833 pixels, 0.6 m resolution) from 2023 NAIP imagery, spanning 20 U.S. states and six livestock categories (Swine, Poultry, Dairy, Beef, Horses, Sheep/Goats), plus 20,771 curated negative samples. An additional 6,454 synthetically augmented patches are included, bringing the total to approximately 45,000 patches. For each sample, CAFOSat provides:

- **Image patches:** High-resolution NAIP aerial imagery (833×833 px, 0.6 m/px) centered on CAFO or non-CAFO locations, organized by U.S. state.
- **Classification labels:** Facility-level CAFO type labels (Negative, Swine, Poultry, Dairy, Beef, Horses, Sheep/Goats) with human-verified and AI-annotated variants indicated via a `verified_label` flag.
- **Infrastructure-level annotations:** Binary flags for barn presence, manure pond presence, grazing area presence, and other infrastructure for 4,513 manually verified patches (see Table 2 for per-class counts).
- **Bounding box annotations:** Geospatial bounding boxes (`geom_bbox`) and polygon geometries (`geometry`) with associated coordinate reference systems for each patch.
- **Geospatial metadata:** Original weak coordinates (`weak_x`, `weak_y`) alongside refined coordinates (`refined_x`, `refined_y`), U.S. state, spatial resolution, and unique facility identifiers (`CAFO_UNIQUE_ID`).
- **Augmentation metadata:** For synthetic patches, the original source patch path, the inpainting prompt used, and the `image_type` flag distinguishing real from augmented samples.
- **Standardized splits:** Six pre-defined train/val/test split configurations (Verified, Augmented, Merged, Set 1, Set 2, All-Training) provided as boolean flags in the master metadata file (`CAFOSat.csv`), enabling reproducible benchmarking across different experimental setups (see Table 2).

The dataset is publicly available on HuggingFace¹ under a CC BY 4.0 license, with data loaders and processing scripts on GitHub². The remainder of this section details how each component is constructed.

B. Data Sources

Here, we provide a summary of data sources and acquisition. Details are in the supplement. Table S1 lists all data sources.

B.1. Satellite Imagery

We leverage aerial imagery from the National Agriculture Imagery Program (NAIP), accessed via the Microsoft Planetary Computer. For each geolocated CAFO point, the nearest cloud-free NAIP image is queried and downloaded using the STAC API. These images typically offer spatial resolutions between 60 cm and 1 m, depending on acquisition year and state. For this experimentation, we collected the most recent (2023) data for the studied states.

B.2. CAFO-Dataset

We collected CAFO location and corresponding cafo type primarily from two sources: (i) Department of Geographical and Sustainability Sciences of IOWA³ (denoted as IOWA-CAFO Inventory) and (ii) animal feeding operation report from states (denoted as State-CAFO Inventory).

IOWA-CAFO Inventory: This data inventory aggregates CAFO facility data from state environmental agencies across nine southeastern U.S. states: Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, and Texas. Data is collected from permit databases, nutrient management plans, and agency inspections. Each record includes geolocation, animal type (poultry, swine, beef, dairy), and manure management details. While some states (e.g., Georgia, North Carolina) provide structured digital datasets, others require manual extraction from reports. This dataset standardizes these sources to enable regional CAFO distribution and impact analysis.

State-CAFO Inventory Several U.S. states publish CAFO reports curated by official state agencies using per-

¹<https://huggingface.co/datasets/oishee3003/CAFOSat>

²<https://github.com/oishee-hoque/CAFOSat>

³<https://cafomaps.org/>

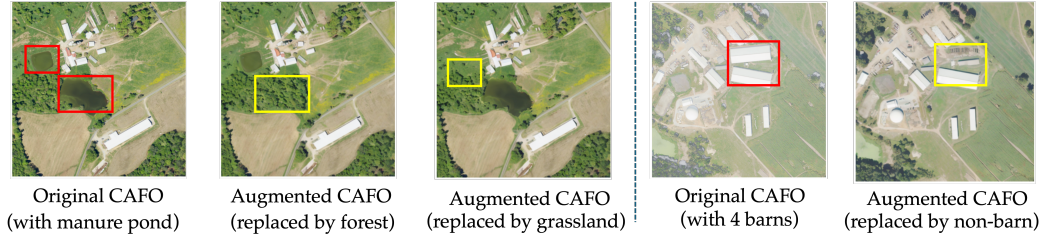


Figure S1. Examples from our augmented dataset. In the first example (left), the augmented images replace the manure pond with forest and grassland; in the second example (right), one barn is replaced with non-barn infrastructure. In both cases, the structural definition of the CAFO remains intact.

Table S1. Summary of Core CAFO Data Sources Used in Livestock Detection.

Source	Description	Use
NAIP Imagery (2023) [42]	High-resolution aerial imagery from the USDA National Agriculture Imagery Program	Visual input for CAFO patch extraction and model training
CAFOMaps [6]	Multi-state (i.e., Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, and Texas) labeled dataset containing 6,604 CAFOs with animal type annotations (e.g., poultry, swine, beef, dairy), curated by IOWA researchers	Ground-truth labels for training and validation of CAFO classification models
State-CAFO Inventory [4, 16, 23, 26, 29, 30, 32]	Six independent data sources from official state-level CAFO registries corresponding to Indiana, Iowa, Maryland, Michigan, Minnesota, New York, and Delaware, curated from permit records, inspections, and nutrient management plans	Ground truth labels for ML-ready dataset
Land Use Masks (NLCD) [25]	National Land Cover Database (NLCD) used for masking agricultural zones	Used to create negative samples

mit records, inspection data, and self-reported nutrient management plans. Delaware’s CAFO report, maintained by the Socially Responsible Agricultural Project (SRAP), provides spatial boundaries and regulatory attributes for permitted operations. Indiana’s CAFO report, compiled by the Indiana Department of Environmental Management (IDEM), is publicly available via an ArcGIS map and contains metadata on facility type and permit status. Iowa’s inventory is provided by the Iowa Department of Natural Resources (Iowa DNR), offering geolocated animal feeding operations with supporting regulatory information. Maryland’s dataset is curated by the Maryland Department of the Environment (MDE), detailing registered CAFOs with geographic and operational metadata. Michigan’s CAFO data is managed by the Michigan Department of Environment, Great Lakes, and Energy (EGLE), covering operations across different time periods with detailed facility-level information. Additional inventories are available from the FracTracker Alliance and the New York Department of State (NYDOS), including permit identifiers and operator details across various temporal spans.

B.3. Land Use Masks (NLCD)

We utilize national-scale raster products to identify and contextualize agricultural areas. The MRLC National Land Cover Database (NLCD) offers 30m-resolution land cover classifications across 16 categories, including cultivated cropland, grassland, barren land, and pasture (See Table S1). This dataset is further used to generate stratified negative samples based on land cover types and spatial extent.

C. Additional Details for Section 3

Manual Verification Setup. Figure S2 visualizes the verification setup we used to manually verify and annotate the data.

Prompt Guided Data Augmentation To improve model generalization in CAFO infrastructure analysis, we develop a prompt-guided augmentation pipeline that removes visually identifiable structures, such as barns, manure ponds, and other supporting facilities. We then replace them with semantically plausible non-infrastructure content. This results in label-preserving image variants that diversify structural configurations while retaining the associated CAFO



Figure S2. Several Manual Verification and Labeling Setup

types (e.g., swine, dairy, poultry) as defined in the metadata. In total, 6454 samples were augmented using the following prompt-guided procedure (3921 involving barns, 1344 manure ponds, and 2089 other infrastructure elements):

- Infrastructure Detection via Vision-Language Prompts:** In this step our objective is to localize and identify physical structures within satellite imagery. Using GroundingDINO, a vision-language object detector, we detect infrastructure based on natural language prompts that encode high-level visual priors observed in overhead imagery. We carefully designed 25 prompts that capture geometric shape (e.g., rectangular, circular), material appearance (e.g., white roofs, dark water), and spatial context (e.g., isolated placement, proximity to buildings). For example, we use *“a long white rectangular building located in an open, unobstructed area”* for barn detection (see the Appendix Table S2). To ensure that prompts are grounded in semantically valid regions and to reduce spurious detections, we apply infrastructure detection only to patches where metadata confirms the presence of the corresponding structure. Note that the prompts are not intended to infer the facility’s animal type; rather, they target visual infrastructure elements that may appear across multiple CAFO categories. The prompts are all listed in Table S2.
- Inpainting for Structure Removal:** For structure removal, we generate binary masks from subsets of predicted bounding boxes (up to five per image) and use these to guide the removal of detected infrastructure. The RGB image and corresponding mask are passed to a Stable Diffusion Inpainting model, conditioned on non-

infrastructure prompts (e.g., “grassland”, “a small water pool”, “cluster of trees”). The model synthesizes contextually consistent content to fill the masked regions. To preserve the validity of the CAFO type label (e.g., swine, dairy, poultry), we remove infrastructure only when multiple instances are present within a patch. For example, if ten barns are detected, a subset (e.g., two or three) are removed to ensure the structural identity of the facility remains intact. This constraint prevents label ambiguity and preserves the semantic consistency of the augmented data. We also provide detailed metadata (including the number and type of structures removed, the inpainting prompt used, and the bounding box configuration) for each sample. This ensures full traceability and allows verification that label integrity is maintained across all synthetically generated samples.

This augmentation strategy introduces visual variation within the same semantic label, enabling models to generalize beyond repeated infrastructure patterns. By selectively removing structures while preserving the CAFO type, the model is encouraged to learn broader contextual features (e.g., landscape, vegetation, layout) rather than overfitting to infrastructure-specific cues.

D. Additional Details for Section 4

Training Setup. We trained all models using PyTorch on an NVIDIA A40 GPU with 40GB of memory. Training was conducted for a maximum of 30 epochs with early stopping (patience = 5) based on validation loss. We used a batch size of 32, a learning rate of 0.0001, and mixed-precision training (FP16) for improved efficiency. Each experiment used 12 data loading workers and was distributed across 2 GPUs where applicable. All models were trained on 8-class classification using the curated dataset described in Section 4.

Baseline Models In our benchmarking framework, we incorporate a diverse set of deep learning models to evaluate their effectiveness in both binary and multiclass CAFO classification tasks. We use ResNet18 and ResNet50 as strong convolutional baselines due to their proven reliability and computational efficiency. To explore transformer-based models, we include ViT-B/16 and Swin-B — with ViT offering global context understanding and Swin providing better spatial inductive biases through hierarchical attention. We also add EfficientNet-B0 and B3 for their state-of-the-art accuracy-efficiency tradeoffs, making them suitable for scalable deployment. Beyond these supervised architectures, we evaluate DINOv2-ViT-B, a self-supervised

Table S2. Grouped prompts and rationales for agricultural infrastructure types, with rationales structured by geometric shape, material appearance, and spatial context.

Structure	Prompt(s)	Rationale
Barn	a long white rectangular building located in an open, unobstructed area	Rectangular in shape with elongated dimensions, featuring light-colored roofs, and positioned in isolated, open fields with minimal surrounding structures.
	a large rectangular structure placed alone in a spacious open area	
	a light gray rectangular structure with a bright roof placed in a cleared field	Rectangular buildings with reflective or light-colored roofs, typically situated in uniform, cleared agricultural areas.
	a rectangular building with a reflective or light-colored roof in a uniform open setting	
	a simple white rectangular building with a flat or sloped roof surrounded by open land	Wide, low-profile rectangular shapes with simple white roofs, located far from roads or residential structures.
	a wide, low-rise building with a simple roof and no nearby roads or houses	
	a pale rectangular building that is isolated from other nearby structures	Box-shaped buildings with pale or white surfaces, spatially detached from other infrastructure in cleared zones.
	a white box-shaped building standing by itself in a cleared landscape	
	a long rectangular building with a white or gray roof and no surrounding clutter	Linear rectangular structures with white or gray roofing, placed in bare or uncluttered terrain.
	a linear structure with strong roof edges, positioned in an empty ground space	
Manure Pond	a dark brown or black irregular pond surrounded by open land	Irregularly shaped, dark-colored water bodies commonly located in bare land with limited vegetation or structures.
	a muddy or black spot on the ground surrounded by bare land	
	a small black water pool next to a large building	Dark, compact pools with irregular or circular shapes, typically near barns or placed away from residential zones.
	a dark patch of water placed away from houses or roads	
	a dark green or brown pool of water near agricultural buildings	Rectangular or rounded shapes with dark tones and low reflectance, often adjacent to farm structures.
	a rectangular dark spot with no reflection near farm infrastructure	
	a dirty water pond placed in a fenced or open farm area	Shallow, flat pits with dark water content, commonly found in fenced or exposed areas within farm boundaries.
a small flat black or brown pit in open space		
Other	a tall round silver structure standing alone in an open area	Cylindrical or rounded forms with metallic or reflective surfaces, typically isolated or aligned along roads or access routes.
	a silver or white round tank placed horizontally on the ground	
	a dark or metallic cylinder standing upright near a road or dirt track	Small rectangular or vertical pipe-like forms, white or neutral in appearance, located near barns or within farm complexes.
	a small box-shaped white structure placed near larger buildings	
	a narrow vertical pipe-like structure standing on open soil	Flat or elongated infrastructure lacking roofing, with metal or white coloring, typically positioned next to buildings.
	a flat structure with no roof, made of metal or concrete, placed near a building	
	a narrow white object placed across open soil with no walls or roof	Curved or residential-style structures with white exteriors and driveways, indicative of greenhouses or farmhouses.
	a small building with a curved roof sitting in a field	
a white rectangular building with windows and a driveway nearby		

vision transformer known for its strong generalization on downstream tasks. Finally, we integrate CLIP and RemoteCLIP, two vision-language models that offer zero-shot capability and strong representation learning without task-specific fine-tuning.

Evaluation Metrics To assess model performance, we use three complementary metrics: Accuracy, F1 Macro, and mean Average Precision (mAP). **Accuracy** measures the overall proportion of correct predictions across all classes but may be biased toward dominant classes in imbalanced datasets. To address this, we report **F1 Macro**, which com-

putes the F1 score independently for each class and averages them, giving equal weight to all classes regardless of frequency—making it especially important for evaluating performance on underrepresented CAFO types. Finally, we include **mean Average Precision (mAP)** to capture the precision-recall trade-off across thresholds, providing a robust indicator of class separability and detection quality. Together, these metrics offer a comprehensive view of model performance across both dominant and minority classes.

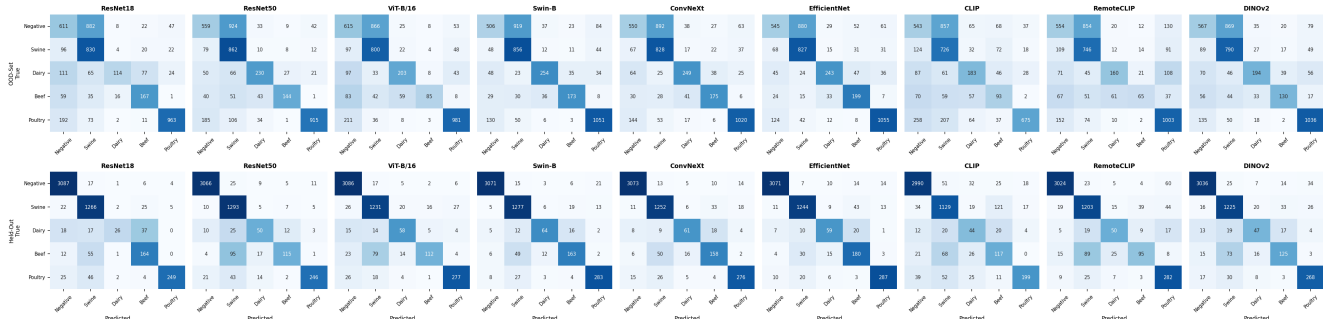


Figure S3. Confusion matrices for all nine models on the *Verified-Set* (*top row*) and *Held-Out Set* (*bottom row*), restricted to the five major classes. Cell values show raw counts; color intensity reflects row-normalized accuracy. Models perform substantially better on the in-distribution *Held-Out Set*. On the *Verified-Set*, *Negative* samples are frequently misclassified as *Swine*, and *Dairy* and *Beef* show persistent mutual confusion due to visual similarity. *Poultry* is the easiest class to recognize across both sets, owing to its visually distinct long-barn infrastructure.

Model	Negative	Barn	Manure Pond	Grazing Area	Other Infra.
CLIP	0.432	0.505	0.414	0.313	0.689
ConvNeXt-Tiny	0.406	0.589	0.494	0.563	0.825
DINOv2-ViT-B	0.461	0.573	0.511	0.542	0.808
EfficientNet-B0	0.463	0.593	0.618	0.593	0.863
RemoteCLIP	0.465	0.515	0.385	0.343	0.717
ResNet18	0.473	0.584	0.558	0.536	0.831
ResNet50	0.464	0.611	0.579	0.574	0.842
Swin-B	0.496	0.596	0.571	0.563	0.854
ViT-B/16	0.483	0.591	0.593	0.546	0.836

Table S3. Per-class F1 scores for infrastructure-level detection across all evaluated models, computed on the *Verified Set*. Each model is trained on the *Standard Training Set* and evaluated on binary presence/absence of barn, manure pond, grazing area, and other infrastructure within manually annotated CAFO patches. Bold values indicate the best per-column performance.