

A Flexible and Generalizable Pipeline for Predicting Concentrated Animal Feeding Operation Populations from Satellite Imagery

Anya Fries
Stanford University
afries@stanford.edu

Deepan Shah
Stanford University
deepans@stanford.edu

William Zhang
Stanford University
wxyz@stanford.edu

Abstract

Predicting concentrated animal feeding operation (CAFO) populations is a necessary component of estimating United States agricultural greenhouse gas emissions and their associated downstream pollution effects. In this paper, we predict dairy, poultry, and beef CAFO populations from National Agriculture Imagery Program (NAIP) images of known locations in California. Utilizing semantic segmentation, deep learning models, and simple features extracted from the images, we build a versatile ensemble model to predict populations for each farm type—achieving an r^2 of 0.47, 0.31, and 0.30 on the test sets for dairy, poultry, and beef farms respectively. Despite being trained only on California NAIP imagery, we find that the model performs well on out-of-sample data. In leveraging advanced methodologies for predicting CAFO populations, our paper moves toward refining estimates of agricultural greenhouse gas emissions among other environmental metrics, thereby enlightening political actors to implement needed policy decisions.

1. Introduction

Livestock agricultural operations have become increasingly industrialized, and more than 40% of US livestock are raised in concentrated animal feeding operations (CAFOs) [3]. CAFOs have effects on many areas of environmental health, including water quality, air pollution, pathogens, and greenhouse gas emissions [8]. These effects are directly related to the size of the population at these facilities. Accurately assessing CAFO populations is a critical component in monitoring their environmental impact. However, the current practice of determining their populations through manual counting is labor-intensive and time-consuming. These issues motivate the need for a low-cost, scalable method of assessing CAFO populations.

As agricultural activities contribute significantly to greenhouse gas emissions, representing approximately 10%

of the total emissions in the United States and 24% globally, the need for fast quantification to achieve estimates on a large scale becomes evident [13]. Many of these emissions come from livestock; CAFOs contributed an estimated 7% of US greenhouse gas emissions and 18% globally [8].

In light of the urgency to address climate change and achieve the Sustainable Development Goals, an accurate measurement of agricultural greenhouse gas emissions is imperative, serving as a crucial metric for tracking emission reduction efforts. This paper contributes to this goal by creating a generalizable pipeline to predict dairy, poultry, and beef CAFO populations from publicly available satellite images from the National Agriculture Imagery Program (NAIP). By quantifying these populations, estimates of greenhouse gas emissions attributable to livestock can be refined, thereby advancing our understanding of broader agricultural impact.

The practical implications of this research are far-reaching: CAFO population predictions provide a foundation for enhanced monitoring and accountability within the agricultural sector. Estimating livestock headcounts where they are not available can inform policy decisions and assist political action to mitigate the environmental impact of CAFOs. Pinpointing and measuring CAFOs and their populations offers valuable insights into potential sources of water and air pollution, as well as issues related to odors and pathogens. This knowledge enables swift action to be taken, addressing these concerns promptly.

While previous work has always focused on a single CAFO type, we aim to provide a low-cost, scalable, and flexible machine-learning pipeline to predict CAFO populations from satellite imagery of various facility types (dairy, poultry, beef) that can be applied at any time (i.e., does not rely on Census data). We aim to validate this model out-of-sample to show its general utility.

2. Related Work

Locating CAFOs Locating a CAFO from satellite imagery is a necessary first step that precedes predicting facil-

ity counts. Prevailing methodologies in this domain primarily hinge on image classification, identifying the presence of a farm within an image, and semantic segmentation, distinguishing pixels corresponding to farm structures. Handan-Nader et al. [6] employed an image classification approach in which they utilized two Convolutional Neural Networks to categorize images based on the presence of either hog or poultry facilities (one for each facility type) and took the higher of the two scores as the predicted label. Robinson et al. [15] performed semantic segmentation, utilizing a U-Net model for binary semantic segmentation of poultry barns. They highlighted the importance of rotation augmentation in training to improve out-of-sample performance and used a lengthy post-processing procedure grounded in domain knowledge to increase model precision further, showcasing the potential for domain-specific strategies to enhance model performance. The U-Net architecture has generally demonstrated versatility in detecting buildings in satellite imagery, even with limited training data [16, 19].

Predicting population from farm imagery Two primary approaches exist for predicting facility counts: focusing on the facility itself or counting the animals within it. In the former, the facility’s characteristics and size offer insights into the animal population. Intuitively, the area of barns and/or feedlots should scale with the number of animals, among other physical indicators. Jeong et al. [9] used a U-Net model on NAIP aerial imagery to estimate dairy facility areas in California’s San Joaquin Valley, then used these areas to downscale county-level dairy population estimates. This approach builds on the Vista-CA project [11], which identified dairy farms in California using satellite imagery and herd data from the State Water Resources Control Board’s 2019 CAF fees list. While Jeong et al. achieved >95% correlation with Vista-CA estimates, this method relies on county-level data, only available every 5 years from the U.S. Census of Agriculture (AgCensus). This differs from our approach, where we aim to create a model that does not rely on AgCensus data - allowing for greater flexibility, granularity and scope.

Counting based methods Counting animal populations directly from imagery has been applied to aerial and satellite imagery of open areas [18, 20]. When buildings do not obscure animals, these methods can be used to monitor these livestock populations efficiently, and it has been applied to cattle ranches and pastures [14]. These methods require very high-resolution aerial or satellite imagery, making them difficult to scale due to the cost of acquiring data. Most work has been applied to UAV imagery [1, 17], which typically has a resolution of less than 5cm GSD, far higher than what is available through commercial satellites. One work has demonstrated cattle counting from satellites

using 40cm resolution imagery but shows declining performance as the density of animals increases [10]. Counting-based methods also work best on larger animals like cattle in open fields. For these reasons, counting-based methods are difficult to apply to densely populated CAFOs and smaller animals like chickens. We aim to utilize publicly available satellite imagery, which tends to be lower resolution than commercially sourced imagery.

3. Method

We aim to predict CAFO population sizes from satellite imagery of known farm locations. The datasets (described in further detail in Section 4: Experiments) are relatively small. This presents a unique challenge due to the many features in the satellite imagery. Direct application of standard supervised learning techniques, especially training deep models on full-resolution images, is likely to overfit and generalize poorly. To tackle this challenge effectively, our approach involves feature extraction from the images (to reduce their complexity), which is input into a simple model to generate the final predictions. This strategy allows us to circumvent the computational demands of working directly on full-resolution images while aiming to achieve accurate and meaningful results. We extract simple features, facility area, and deep features, which are input into an ensemble model (Figure 1).

Simple features The simple features extracted from the images are classical computer vision methods, including the histogram of oriented gradients (HOG) [4] and statistics calculated from the distribution of pixel values in each band. As a pre-processing step, we center-crop each image to 2048x2048 pixels. HOG is a feature extraction technique typically used in computer vision for object detection. It counts occurrences of gradient orientation in localized portions of an image. We use 3 orientation bins and calculate each bin’s mean, standard deviation, median, min, max, upper and lower quartiles, used as the features we use in our downstream model. We also calculate the mean, standard deviation, median, min, max, upper and lower quartiles for each image band and use these features to represent the distribution of colors in each image.

Facility area We utilize two pre-trained U-Net models to estimate the facility area within each image. The U-Net model developed by Jeong et al. [9] is applied to dairy and beef farms. This model is trained on California dairy farms and classifies image pixels into the farm (barn and feedlots) vs not. The farm area is estimated by counting pixels. The model has a standard U-Net architecture with a ResNet-50 pre-trained encoder. To utilize the model, we center-cropped our images to 2016x2016 pixels and then

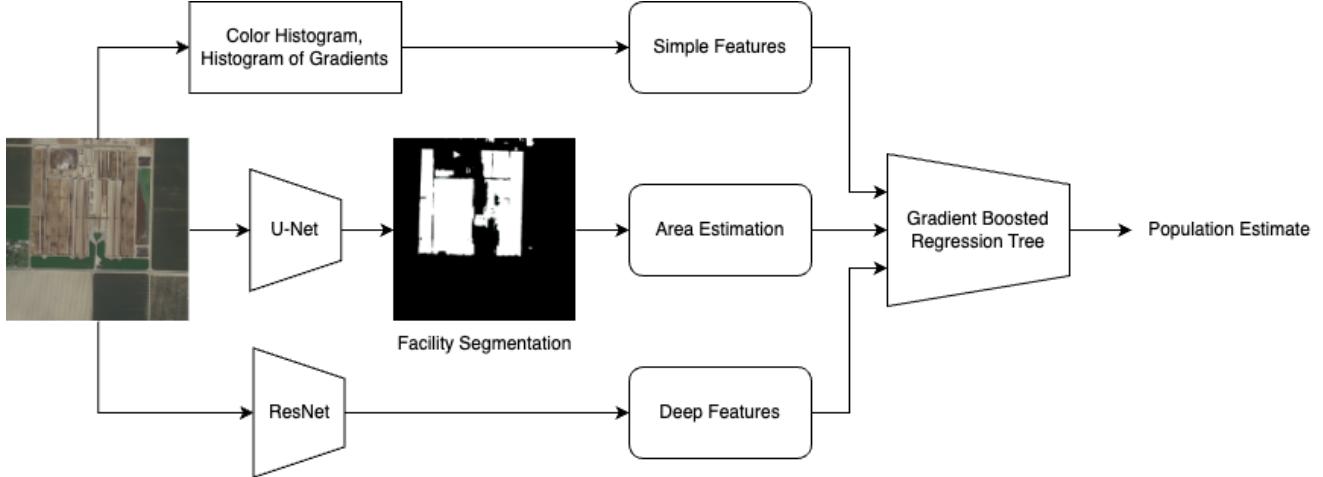


Figure 1. Our proposed pipeline.

partitioned this into 224x224 pixel chunks, on which the model is applied. Post-processing involves stitching the model outputs together, followed by pixel counting of areas with values exceeding 0.5.

We use the U-Net model from Robinson et al. [15] for poultry farms. This model is trained on Delmarva Peninsula poultry farms and classifies only the barn areas. Similar to the dairy and beef model, it is a standard U-Net architecture with a ResNet-18 pre-trained encoder. The application of this model follows a similar process to the first, involving center-cropping each image to 2048x2048 pixels, splitting them into 256x256 pixel chunks, applying the model to these chunks, and subsequently stitching the outputs. Post-processing includes applying a softmax operation to the two-channel output of the U-Net, with one channel identifying negative pixels and the other positive "farm" pixels. The pixels with values greater than 0.2 (threshold chosen via visual inspection) in the second layer of the softmax, corresponding to positive farm pixels, are then counted to refine the segmentation outcome.

Deep features We use a fine-tuned ResNet-50 [7] encoder as a deep feature extractor. We alter the final layer to be a fully connected layer comprising 2 linear layers with a ReLU in between. The final two layers were retrained using Mean Square Error (MSE) loss on population count with rotation and colour augmentation. This leads to a feature vector of size 256 (the final features before the output), representing the most essential input information.

Ensemble model We aggregate the simple features, facility area estimation, and ResNet features using a gradient-boosted decision tree. This method iteratively trains decision trees to predict the residuals of the previous decision

tree. The final model outputs the sum of the predictions of each tree. We use a learning rate of 0.1, a maximum of 100 trees, and the Friedman MSE criterion.

We also tested other models for the final prediction ensemble model, including linear regression, lasso regression, ridge regression, elastic net [21], decision tree, random forest, and AdaBoost [5].

4. Experiments

To predict the CAFO population, we aim to find an optimal model structure for three facility types simultaneously (dairy, poultry, and beef). We first present the training and model-building data. Then, the results of various experiments for predicting facility area and CAFO population using this data are shown. Finally, out-of-sample testing and an ablation study are presented.

4.1. Data

Satellite imagery We use RGB satellite imagery from the National Agriculture Imagery Program (NAIP) with 0.6m resolution. We select a region centred at each farm covering a 2 km by 2 km area.

CAFO data The California Integrated Water Quality System provides data [2] from site visits of facilities that discharge wastewater, from which we obtain facility types, locations, and population counts. A histogram of population sizes for each facility type is shown in Figure 2 and the locations are given in Figure 3, largely in central California.

We obtained further historical data with updated farm population counts over time through email correspondence with the CA Waterboards and so considered two training data pipelines: (a) using only the most recent image and population count for farms, and (b) using up to 3 (2)

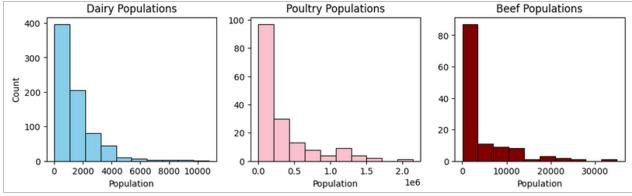


Figure 2. Histogram of population sizes for different facilities. The population has a heavy-tailed distribution.

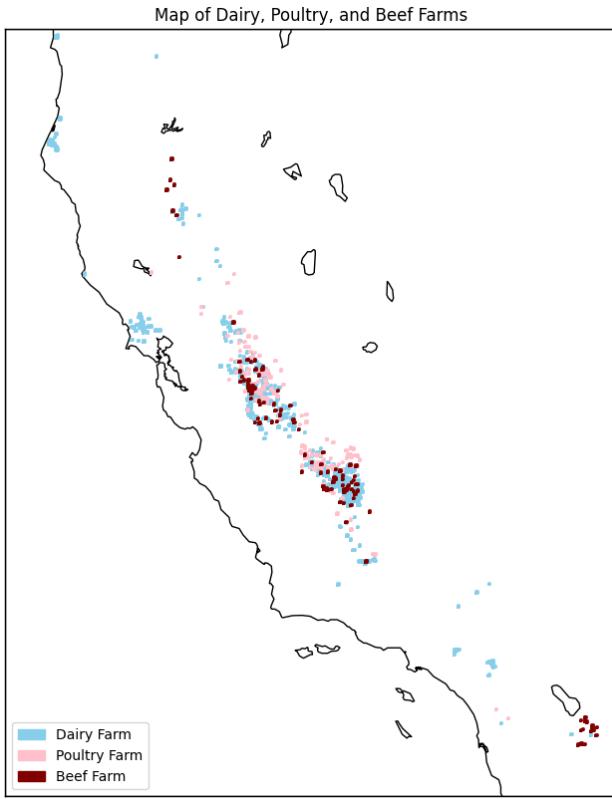


Figure 3. Map of our dataset’s dairy, poultry, and beef farms. All of these farms reside in California, mostly in the Central Valley.

most recent population counts and corresponding images for dairies (poultry, beef). We restricted the number of images to avoid over-fitting to certain farms. Pipeline (b) was proposed as training with multiple images per farm should increase generalizability since more data points are included. This can be viewed as a natural form of data augmentation.

We removed entries with populations less than 5 and beef with greater than 100,000 as these were deemed outliers, likely to bias our results. Our final datasets comprised 756 dairies, 168 poutries, and 122 beef facilities. The total images for pipeline (b) are 1903, 210, and 122 for dairy, poultry, and beef facilities. We used a 70%, 15%, and 15% split at the farm level for training, testing, and validation.

Given the temporal aspect of the population data, we also explored predicting CAFO populations over time (Appendix 6.1), but the data was not granular enough for this task.

4.2. Results

Evaluation metrics To assess the quality of our predictions, we report standard metrics: root mean squared error (RMSE), mean absolute error (MAE), and r^2 .

Facility area estimation To assess the performance of the U-Net models, we provide histograms of the distribution of farm sizes derived from the pixel counts in Figure 4.

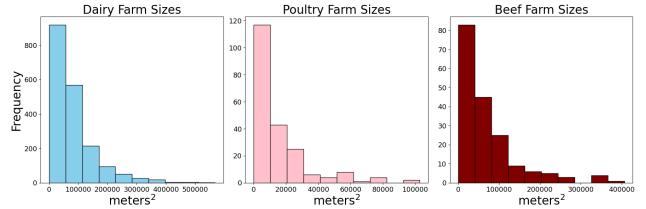


Figure 4. Farm sizes predicted from U-Net.

We see that the distributions of farm sizes each follow a similar heavy-tailed distribution. The resemblance of these histograms to the population distributions depicted in 2 suggests a positive indication that the predicted farm sizes generated by the U-Net models are linked to corresponding population sizes.

As mentioned in 3, we find that the U-Net from Jeoung et al. performs well on dairy and beef farms, even though it was trained only on dairy farms. We explored the poultry U-Net from Robinson et al. but found that it performed well on only the poultry data, as expected. Figure 5, 6, and 7 show examples of the U-Net outputs and the final predicted population from the gradient-boosted decision tree.



Figure 5. Dairy farm image and U-Net output mask. The gradient-boosted decision tree predicted a population of 783.

Deep features The ResNet was finally trained with MSE loss with an Adam optimizer, a learning rate 0.02, and a



Figure 6. Poultry farm image and U-Net output mask. The gradient-boosted decision tree predicted a population of 1,159,624.



Figure 7. Beef farm image and U-Net output mask. The gradient-boosted decision tree predicted a population of 604.

scheduler with step size 10 and gamma 0.8 for 30 epochs. As shown in the ablation study (Section 4.4), the model performs well alone, nearly comparable to the ensemble model.

Interestingly, for dairy and beef, the barns are not the most important features of the image - the surrounding land appears more important. The saliency maps are shown in the Appendix (Figure 14 and Figure 15). This motivated us to include the U-Net and ResNet as they provide complementary information. However, the barns are often the most important predictor for poultry facilities (Appendix, Figure 16), highlighting that the different CAFO types have heterogeneous feature importance.

Model selection To determine the ensemble model for the final population prediction, we tested several regression models on our validation data, with results shown in Table 1. We chose gradient-boosted decision trees for the final model since it performed best on two of three datasets.

Multiple training images The results of using multiple images and population counts per farm in our training dataset instead of just the most recent data point are displayed in Table 2. Performance slightly increased for the dairy dataset and slightly decreased for the poultry and beef datasets. Since the dairy dataset is larger than the poultry

Model	Dairy	Poultry	Beef
Linear Regression	563.37	786584.24	9363.32
Lasso	548.50	271741.98	3240.84
Ridge	560.74	245211.25	3262.90
Decision Tree	544.46	246081.36	3876.28
Gradient Boosting	554.73	215940.41	3043.60
Random Forest	542.07	238185.73	3109.90
AdaBoost	555.15	276117.61	3111.72
Elastic Net	549.41	234503.04	3164.52

Table 1. Ensemble model results reported on each of the validation datasets. The metric reported here is MAE.

and beef datasets, we believe the benefit is more apparent when the dataset size is larger. When the dataset size is smaller, the model tends to become more biased towards duplicates in the training set. We choose to run our final model with multiple training images per farm because it should increase the generalizability of our method if more data is available and the current difference in performance is not marked.

# imgs	Dataset	RMSE	MAE	r^2
Single	Dairy	975.90	570.87	0.53
	Poultry	293120.97	201965.24	0.48
	Beef	3254.02	2153.77	0.88
Multiple	Dairy	951.35	554.73	0.56
	Poultry	302772.78	215940.41	0.44
	Beef	4418.42	3043.60	0.76

Table 2. Performance on the validation set using single vs multiple images per farm in our training dataset. Gradient-boosted decision tree is the model used here.

Dataset	RMSE	MAE	r^2
Dairy (val)	951.35	554.73	0.56
Dairy (test)	825.97	580.56	0.49
Poultry (val)	302772.78	215940.41	0.44
Poultry (test)	391583.49	221836.81	0.43
Beef (val)	4418.42	3043.60	0.78
Beef (test)	6282.51	3697.27	0.30

Table 3. Final model results on the validation and test set using all features and gradient-boosted regression tree.

Discussion We see that we obtained the highest test r^2 on the dairy dataset, and decreased performance on the poultry and beef datasets. The dairy dataset size is significantly

larger, with poultry coming in second and beef coming in third. We believe that this indicates that a larger dataset size would improve the model’s performance. Due to the small size of our beef dataset, we see a larger decrease in performance between validation and test data. This is because we are more likely to overfit validation data when performing model selection, which generalizes poorly to the test data. For dairy, the model has a lower RMSE on the test data compared to the validation data, indicating that the larger dataset size aids the model’s generalizability.

4.3. Out of sample data validation

Since our facilities are all located in California, we test our model on out-of-state data to assess the generalizability of our method.

Dairy We use NAIP images of dairy farms in Minnesota, with population labels given as ranges (0-49, 50-99, 100-199, 200-299, 300-399, 400-499, 500-1000, >1000). It is worth noting that the differences in adjacent classes are very small compared to the distribution in our training dataset. Since there is a distribution shift in the population labels, and we only have population ranges, we formulate the problem as a multi-class classification and refit the gradient-boosted decision tree on a subset of the data. We have 1442 total data points and use 70% for training and 30% for testing. We compare the performance of using all features against using only U-Net features to test whether adding simple and ResNet features aids generalization.

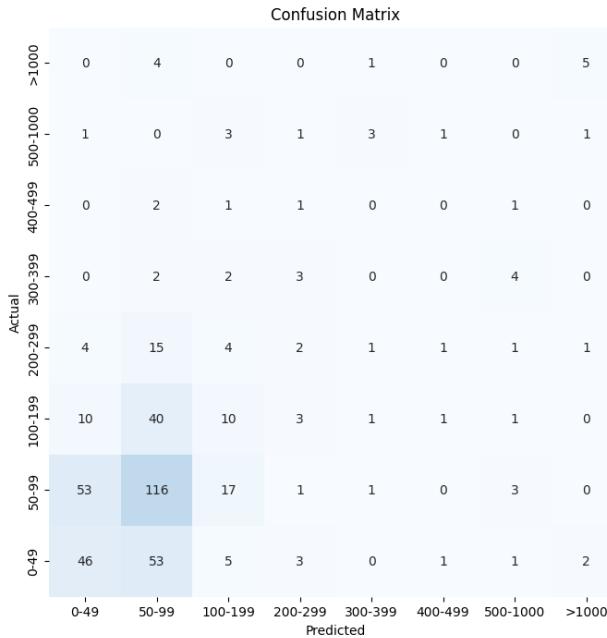


Figure 8. Confusion matrix for dairy out of sample predictions.

Metric	All Features	Only U-Net
Accuracy	41.3%	36.9%
Accuracy within 1 class	82.2%	74.1%
Accuracy within 2 classes	92.1%	85.4%
Accuracy within 3 classes	95.6%	90.5%

Table 4. Dairy out of sample prediction accuracy.

The confusion matrix for the predictions using all features is shown in Figure 8. We also report accuracy and accuracy within n classes in Table 4 for the model with all features and the model with only U-Net features. The model with all features is able to successfully classify over 80% of farms within 1 class, and over 90% within 2 classes. When we only use U-Net features, classification performance decreases significantly. These results show that our method could be adapted to different regions by keeping the feature extraction techniques unchanged and refitting only the final layer. They also indicate that the simple and ResNet feature extraction techniques improve the generalizability of our model.

Poultry We can test our poultry model out-of-sample using the nationwide poultry facility map (PFM) [15] to estimate farm-level populations and aggregate at the county level, compared to the 2017 USA AgCensus [12]. As the PFM is at the barn level and our models are to be run on farms, we perform hierarchical clustering until the size of an NAIP image tile to generate ‘farms’. We also note that the PFM is generated via a machine-learning algorithm. Thus, certain farms identified are not actual facilities, and some true facilities may be missing. Figure 9 shows failure cases of the PFM, where barns are incorrectly found (false positives).



Figure 9. Failure cases of the PFM in three states. Roads, bridges, docks, sand banks, and snow have all been found to be common false positives.

We pull data for Kentucky, South Carolina, and Mississippi and made predictions using the most recent NAIP image before 2017. A few farms were not retrieved; in this case, we re-weighted county totals according to how many farms we could not pull satellite imagery for.

The model obtained an r^2 of 0.77 without fine-tuning in Mississippi, and 0.56 and 0.64 in Kentucky and South Carolina. Results are promising given that we are building on another ML pipeline and still achieving this result: the pipeline can generalize relatively well in a different geographic region. Figure 10 shows predictions vs truth for Mississippi. The other states are in the Appendix 6.4. Many of the counties for which we are over-predicting have no facilities. This is likely a result of the PFM incorrectly attributing barns in those regions. At the same time, we cannot attribute enough to the counties with a huge amount of operations. This could be a limitation of our model (e.g., smaller CAFOs in training data) or a result of the PFM not finding sufficient CAFOs.

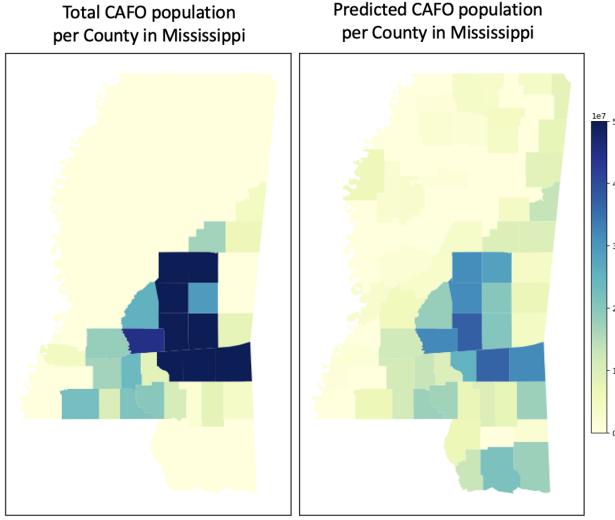


Figure 10. True and predicted county aggregates for poultry out-of-sample in Mississippi. Predictions were made at 2321 farms in 59 counties, achieving an r^2 of 0.77.

4.4. Ablation study

We perform an ablation study on the different feature extraction techniques in our model. We remove one of the three features (simple, U-Net, ResNet) and refit the final ensemble model. We also evaluate the standalone performance of each of the three features. We use r^2 as the metric of comparison for the ablation study, as it can be directly compared across all the datasets.

The ablation study results are shown in Figure 11. The full model performed the same or better than any other models. Simple features performed the worst on its own, while ResNet performed the best on its own. We see that combining simple and ResNet almost reached the performance of the full model and was the best-performing model with one removed feature. However, the full model performs the best overall, indicating that all feature extraction techniques

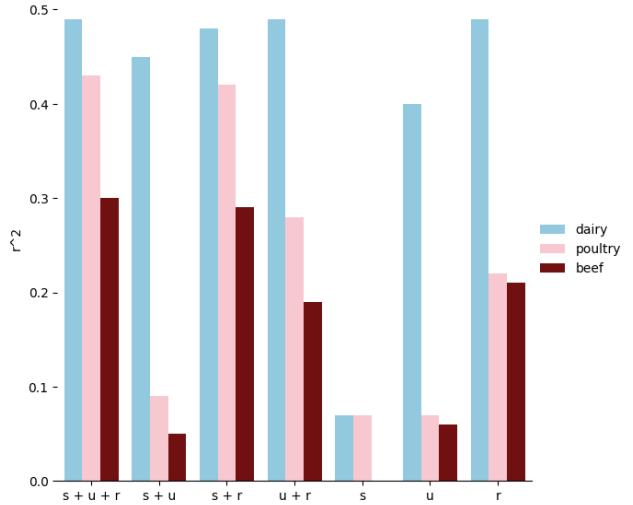


Figure 11. Ablation study results on the test datasets. r^2 is plotted for each model configuration. The label indicates the included features (s=simple, u=U-Net, r=ResNet).

contribute to the model’s performance.

5. Conclusion

We have demonstrated a generalizable method for predicting CAFO populations from publicly available satellite imagery, using an ensemble of simple features and deep learning features. Our results show that the ensemble approach improves the performance of our model on out-of-sample data. Further work is needed to create a model trained on nationwide and global data, as we were limited to locations in California. We have shown that the method can generalize to other locations without retraining the deep learning models. Further work can be done to apply this method to estimate populations of other types of animal facilities beyond dairy, poultry, and beef CAFOs.

Future work can explore how we can use estimated CAFO populations to predict environmental impacts and if it can be used to monitor important environmental indicators, such as methane emissions, water pollution, and air pollution.

Another possible area of exploration is in multi-task learning and self-supervised learning. Our method only uses training data from the same CAFO type, and a different model is needed for each type of farm. However, future work can experiment with self-supervised learning techniques like contrastive learning that can use the data from other farm types to improve its predictions. It would be beneficial to create a classifier that determines the CAFO type. Attaching this to the head of our model would enable it to be run on any farm without prior knowledge of its type, creating a unified pipeline.

Code for this project can be found at <https://github.com/anyaafries/animal-facilities>

References

- [1] Jayme Garcia Arnal Barbedo, Luciano Vieira Koenigkan, Patricia Menezes Santos, and Andrea Roberto Bueno Ribeiro. Counting cattle in uav images—dealing with clustered animals and animal/background contrast changes. *Sensors*, 20(7):2126, 2020. [2](#)
- [2] California State Water Resources Control Board. Surface water - water quality regulated facility information: Confined animal facilities regulated by the water board. <https://data.ca.gov/dataset/surface-water-water-quality-regulated-facility-information/resource/c16335af-f2dc-41e6-a429-f19edba5b957>, 2023. [Online; accessed 31/07/2023]. [3](#)
- [3] Claudia Copeland. Animal waste and water quality: Epa regulation of concentrated animal feeding operations (cafos). Congressional Research Service, the Library of Congress Washington, DC, USA, 2006. [1](#)
- [4] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 886–893. Ieee, 2005. [2](#)
- [5] Yoav Freund and Robert E Schapire. A desicion-theoretic generalization of on-line learning and an application to boosting. In *European conference on computational learning theory*, pages 23–37. Springer, 1995. [3](#)
- [6] Cassandra Handan-Nader and Daniel E Ho. Deep learning to map concentrated animal feeding operations. *Nature Sustainability*, 2(4):298–306, 2019. [2](#)
- [7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. [3](#)
- [8] Carrie Hribar. Understanding concentrated animal feeding operations and their impact on communities. 2010. [1](#)
- [9] Seongeon Jeong, Marc L. Fischer, Hanna Breunig, Alison R. Marklein, Francesca M. Hopkins, and Sebastien C. Biraud. Artificial intelligence approach for estimating dairy methane emissions. *Environmental Science & Technology*, 56(8):4849–4858, 2022. PMID: 35363471. [2](#)
- [10] Issam Laradji, Pau Rodriguez, Freddie Kalaitzis, David Vazquez, Ross Young, Ed Davey, and Alexandre La-coste. Counting cows: Tracking illegal cattle ranching from high-resolution satellite imagery. *arXiv preprint arXiv:2011.07369*, 2020. [2](#)
- [11] A. R. Marklein, D. Meyer, M. L. Fischer, S. Jeong, T. Rafiq, M. Carr, and F. M. Hopkins. Facility-scale inventory of dairy methane emissions in california: implications for mitigation. *Earth System Science Data*, 13(3):1151–1166, 2021. [2](#)
- [12] USDA NASS: Census of Agriculture. <http://www.nass.usda.gov/AgCensus>, 2017. [Online; accessed 15/08/2023]. [6](#)
- [13] United States. Environmental Protection Agency. Office of Policy. *Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2021*. The Agency, 2021. [1](#)
- [14] Maryam Rahnemoonfar, Dugan Dobbs, Masoud Yari, and Michael J Starek. Discountnet: Discriminating and counting network for real-time counting and localization of sparse objects in high-resolution uav imagery. *Remote Sensing*, 11(9):1128, 2019. [2](#)
- [15] Caleb Robinson, Ben Chugg, Brandon Anderson, Juan M. Lavista Ferres, and Daniel E. Ho. Mapping industrial poultry operations at scale with deep learning and aerial imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:7458–7471, 2022. [2](#), [3](#), [6](#)
- [16] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III* 18, pages 234–241. Springer, 2015. [2](#)
- [17] Wen Shao, Rei Kawakami, Ryota Yoshihashi, Shaodi You, Hidemichi Kawase, and Takeshi Naemura. Cattle detection and counting in uav images based on convolutional neural networks. *International Journal of Remote Sensing*, 41(1):31–52, 2020. [2](#)
- [18] Dongliang Wang, Qingjie Song, Xiaohan Liao, Huping Ye, Quanqin Shao, Jiangwen Fan, Nan Cong, Xiaoping Xin, Huanyin Yue, and Haiyan Zhang. Integrating satellite and unmanned aircraft system (uas) imagery to model livestock population dynamics in the longbao wetland national nature reserve, china. *Science of The Total Environment*, 746:140327, 2020. [2](#)
- [19] Eric Wang and Dali Wang. Using u-net to detect buildings in satellite images. *Journal of Computer and Communications*, 10(06):132–138, 2022. [2](#)
- [20] Yifei Xue, Tiejun Wang, and Andrew K Skidmore. Automatic counting of large mammals from very high resolution panchromatic satellite imagery. *Remote sensing*, 9(9):878, 2017. [2](#)
- [21] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(2):301–320, 2005. [3](#)

6. Appendix

6.1. A Temporal Element

We explored the temporal dimension of farm populations to see if we could predict CAFO populations over time. We examined the subset of farms with multiple populations to determine whether the model effectively captures changes in these populations over time. We were looking for a binary outcome to predict if a population change of over 10% occurred. The model achieved a training accuracy of 59%. However, it's important to note that the testing accuracy was considerably lower, at 44%. The confusion matrix is in Figure Figure 12. While this could suggest potential, we examined certain farms, for example Figure 13. Notably, the visual clues are insufficient compared to the variation in population.

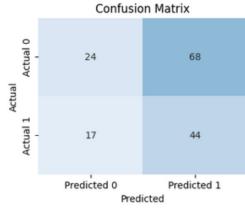


Figure 12. Confusion matrix for the temporal task of predicting whether a dairy had a population change of $\geq 10\%$ or not. Results on the test set.

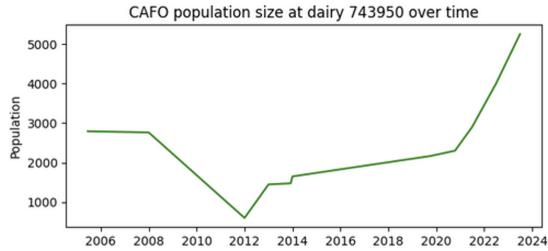


Figure 13. An example at population progression at a dairy with satellite imagery changes.

6.2. ResNet Saliency Maps

Here we show the important features for the different CAFO types: Barns do not seem important for ResNet predictions for beef and dairy, but they do for poultry.

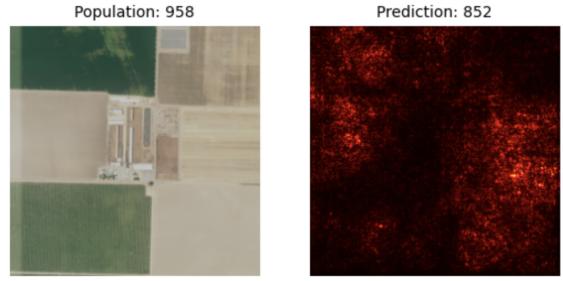


Figure 14. Dairy prediction from ResNet with saliency mask.

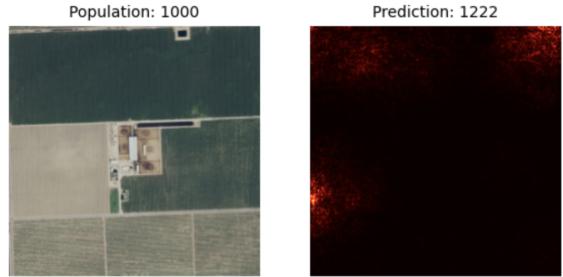


Figure 15. Beef prediction from ResNet with saliency mask.

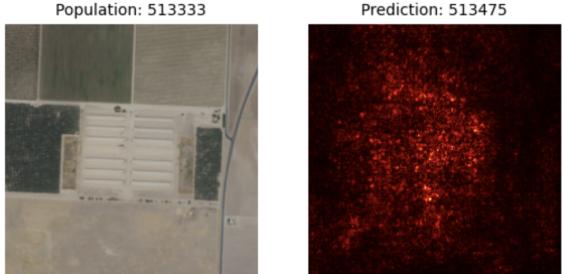


Figure 16. Poultry prediction from ResNet with saliency mask.

6.3. Failure Cases

We examine failure cases for the model for each different farm type. For dairies, of the performers with bigger difference than test MAE, 90% were "Mature dairy cattle" and 10% were "Heifers (non-dairy affiliated)." The example with the biggest difference is Figure 17, a "Heifers (non-dairy affiliated)" farm, where the true population is 5072 and the model predicts 1610.7. In this example, the U-Net prediction is not perfect, but the error likely comes from the lack of images in this population range in the dataset (the model is perhaps biased to predict lower population sizes due to the heavy-tailed nature of the populations in the dataset).

For poultry farms, of the performers with bigger difference than test MAE, 78.6% were of subtype "Non-layers (other than liquid manure system)," 14.3% were "Turkeys," and 7.1% were unknown. The example with the biggest dif-

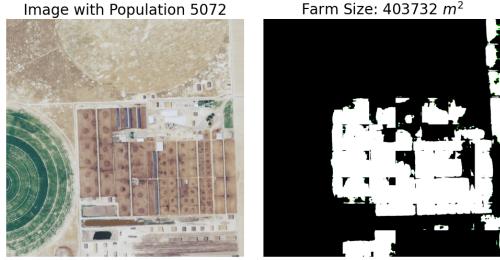


Figure 17. Dairy test set failure example. Model predicts a population of 1610.7

ference is Figure 18, a “Non-layers (other than liquid manure system)” farm, where the true population is 1,611,440 and the predicted population is 27,901.7. In this example, the U-Net doesn’t pick up many pixels, but upon visual inspection, it’s hard to see any poultry farms in this image; this could be a case where the pre-processing image cropping/augmentation cut out the farm.



Figure 18. Poultry test set failure example. Model predicts a population of 27,901.7

For beef farms, of the performers with bigger difference than test MAE, 50% are “Cattle or cow/calf pairs,” and 50% are “Heifers (non dairy affiliated).” The example with the biggest difference is Figure 19, a “Cattle or calf/cow pairs” farm, where the true population is 12,000 and the predicted population is 31,304.8. In this example, the U-Net incorrectly segments the farm, picking up pixels that don’t constitute the beef farm.



Figure 19. Beef test set failure example. Model predicts a population of 31,304.8.7

6.4. Further poultry out-of-sample

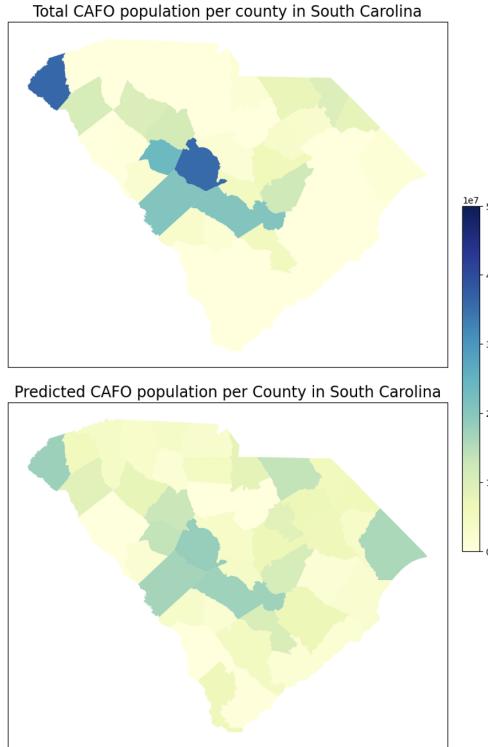


Figure 20. True and predicted county aggregates for poultry out-of-sample in South Carolina. Predictions were made at 1307 farms in 41 counties, achieving an r^2 of 0.64.

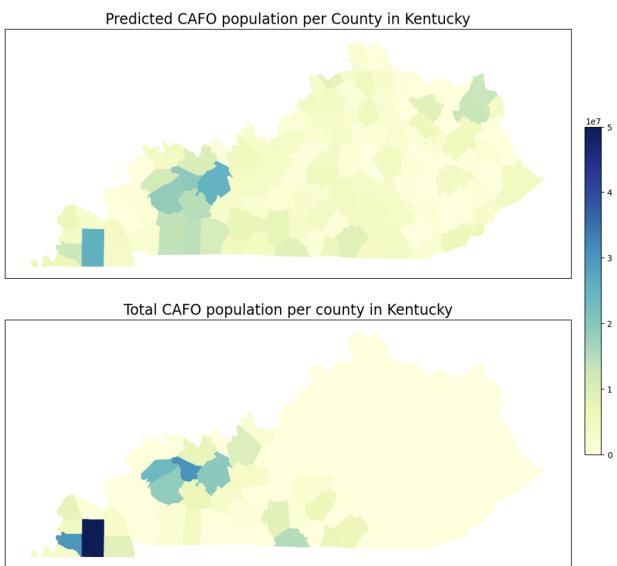


Figure 21. True and predicted county aggregates for poultry out-of-sample in Kentucky. Predictions were made for 1044 farms in 102 counties, achieving an r^2 of 0.56.