CS230

Context-to-Image CNN Approach to Predict Soybean Yields in Illinois and Rural Areas

Christopher Yu
Stanford University
cyu95@stanford.edu

Benjamin Liu
Department of Computer Science
bencliu@stanford.edu

Abstract

Previous research has shown that MODIS and Landsat remote-sensing imagery can be used to successfully predict crop yields in the United States with CNN and RNN architectures. This study will present a novel dataset pipeline with the Planet Labs Dove Constellation to predict soybean yields in Illinois. Benchmarking prominent reference architectures, this study additionally explores a novel context-to-image deep learning model that leverages vegetation-constructed features [1, 2, 3, 4]. Barring collectively low performances induced by systemic dataset challenges, our novel architecture performs comparably to state-of-the-art models. Experiments applying transfer learning to leverage U.S. crop belt data on prediction tasks in rural, data-lacking areas provide evidence that models pretrained on data-rich agricultural data can be applied with reasonable confidence to areas with data shortages. Findings were leveraged in a comprehensive analysis of limitations surrounding Planet Labs-derived data for crop yield prediction tasks, and evidence was derived supporting the use of NDVI and NIR data in similar problems.

1 Introduction

Crop yield is a key agricultural metric that helps assess sustainable food production, governmental allocation of resources, food insecurity and famine, and other key components in farm communities. Traditional crop yield prediction techniques utilize locally surveyed data, which is suboptimal on a large scale due to high costs, human error, and inconsistency in data collection techniques. The pitfalls of these traditional collection techniques are further exacerbated in developing regions that lack sophisticated agricultural infrastructure. [5, 6].

The advent of widely available remote sensing data allows for a more robust approach to assess crop yields. Recent research has demonstrated that applying machine learning techniques to remote sensing data is a viable method to obtain accurate crop yield predictions [7]. However, key challenges that machine learning researchers have faced in this field include lack of labelled ground truth data and a primitive understanding of how climatology and location data can affect crop yield predictions [2, 8, 1, 9]. Additionally, previous studies have traditionally used MODIS or Landsat for input satellite imagery, services with multispectral pixel resolutions of 500m and 30m respectively. This characteristic is problematic due to the relative sizes of farmland. In example, a 500m resolution equates to around 4 pixels per farmland.

In this study, we take a novel approach to leveraging the new Planet Labs dove constellation, which boasts a pixel resolution of 3m. Using engineered vegetation indices in tandem with the collected four-band multispectral satellite imagery, we designed and tested a metadata-enhanced neural network architecture to output predicted county-level soybean yields for the state of Illinois and the country of Brazil. We further benchmarked our model against state-of-the-art models such as VGG-16 and Resnet-50. Following these tests, we utilized transfer learning to extrapolate our model's performance to the task of predicting soybean yields in select rural municipalities in Mato Grosso, Brazil, areas with scarce ground truth data.

2 Related work

Deep learning approaches to crop yield predictions have experienced unprecedented advances in recent years. In 2017, You et al. successfully demonstrate a deep learning approach to crop yields that excluded the use of

CS230: Deep Learning, Winter 2021, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

hand-crafted features, which was previously commonplace in the previous remote-sensing studies [1]. By assuming permutation invariance, You et al. were able to synthesize the 11-band MODIS imagery into 3-D pixel histograms by which custom CNN and LSTM models were trained to predict soy bean yields in United State. Sabini et al. expanded on the CNN architecture presented by You et al. to include additional hidden layers, which resulted in improved model performance [2]. Furthermore, the study concluded that the derived features from the infrared and temperature bands are the largest contributors to crop discrimination.

As our study seeks to deviate from the histogram-mapping approach explored by You et al., promising results achieved leveraging derived vegetation features and context metadata have served as valuable inspirations for experimental design considerations. In 2014, Johnson et al. was able to use vegetation indices and land surface temperature measurements derived from MODIS data to predict corn and soybean yields using a regression tree-based model [10]. Similarly, Lobell et al. demonstrated the benefits of including engineered features and indices in generalized remote-sensing crop yield applications [11]. Similar to these newly derived vegetation features, context metadata has been leveraged in image classification tasks with promising improvements to model performance. In example, Tang et al. demonstrated significant improvements to general image classification with GPS location context data [12]. Similar findings were achieved in studies applying oceanography and temperature metadata for image classification, which we stipulate to hold promise in crop yield tasks [13].

Crop yield machine learning research primarily features regions from the United States "Crop Belt", due to the widely available ground truth data and robust agricultural infrastructure; only a fraction of studies focus on developing rural regions. In 2018, Wang [6] et al. developed an LSTM model to predict soybean crop yields in Argentina by reducing 11-band MODIS satellite data into histograms. The study was able to successfully apply transfer learning to predict soybean yields in areas lacking sufficient ground-truth data, which inspired one application explored in our study.

3 Dataset Generation Pipeline

A majority of the crop-yield remote sensing studies use MODIS multi-spectral imagery from Terra and Aqua, which became operational in 1999 and 2002, respectively, with a 6-year design life. Thus, it is evident that higher-performing image constellations, such as Landsat 8/9 or PlanetScope, may hold more promise for contemporary remote-sensing studies. The pipeline used to generate our dataset is largely original work with minor references to PlanetScope's API tutorials. Due to the lack of remote sensing research that interfaces directly with PlanetScope's satellite imagery, substantial time was spent developing methods to identify, process, label, and store image data to serve our objectives. Refer to section 8.1 for further detail on the data pipeline and vegetation indices.

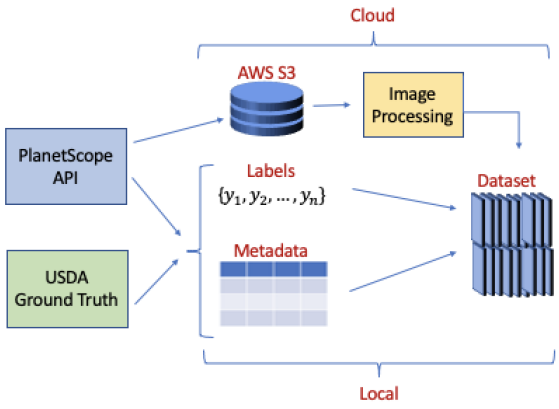

Figure 1: Diagram illustrating satellite imagery and ground truth-integrated data pipeline.

The training set consists of PlanetScope satellite imagery geometrically filtered by Illinois counties. The county area-of-interest polygons were derived from cartographic boundary shapefiles obtained from the US Census Bureau [14]. Additionally, a cloud-cover data filter is applied to the raw satellite imagery the cloud-cover threshold set to 25%. For the purposes of the initial model, the yield data and satellite imagery were obtained for the harvest seasons of 2017-2019, where the soybean truth yields were obtained from obtained from the National Agricultural Statistics Service (NASS) [15]. For the transfer learning application, the area-of-interest polygons were derived from cartographic boundary shapefiles of select municipalities of the state of Mato Grosso, Brazil which has the highest concentration of soy bean farms [16]. The Visvalingam-Whyatt algorithm with a weighted-area scale

factor of 0.07 was applied to the complex boundary polygons to simply the resulting GeoJSON file [17]. The crop yield ground truth for the Brazil data was obtained from "Instituto Brasileiro de Geografia e Estatistica" (IBGE) Municipal Agricultural Production [18]. In finality, the data set consists of 1,750,000 pixels per county/yield pair ($\approx$11295 images $\times$ 500 pixels tall $\times$ 500 pixels wide $\times$ 7 bands). Additionally, the train, validation, and test set distribution is 8:1:1, respectively. For the Brazil transfer learning application, there were 2056 total images from select municipalities in the Mato Grosso with available ground truth data.

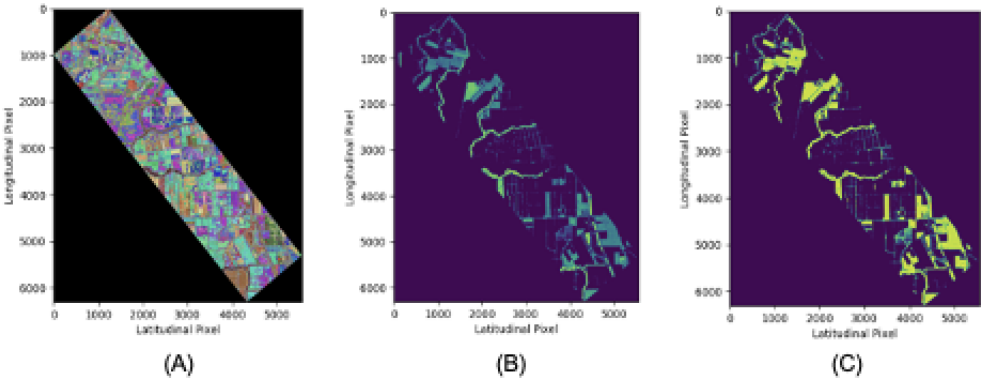

Figure 2: Representation of sampled image in (A) RGB, (B) NDVI, and (C) EVI pixelated forms. As seen, both representations displayed in (B) and (C) differentiate between more vegetative, lush areas in the original image (A).

The crop-yield prediction problem was re-vectored into a simplified 10-bin classification problem, for the purposes of selecting the optimal architecture for the Brazil transfer learning application. For constructed the labels, the bushel/acre per county pairs were scaled by $\text{Area}_{\text{county}}/\text{Area}_{\text{image}}$ to maintain the spatial dimensionality. A quantile-cut was then performed on the labels, while undersampling the labels in the smallest bin by half. Our research objective from a yield prediction perspective is to use a satellite image obtained during any period of a region's harvest period and accurately predict the reported soybean yield obtained from farmland in the specified image following harvest completion.

4 Methods

The various model architectures are listed in table1. To establish a performance baseline, we ran Resnet-50 [3] and VGG-16 [4] using the RGB subset of bands from our images. Additionally, we ran the dataset against the 5-band "Deep-1" architecture that was described in reference crop yield papers from Sabini et al. [2] and You et al. [1]. Lastly, our novel architecture deemed as the "concatenated-arm model" features a 5-band image input into a CNN arm, which follows a similar structure to that of Deep-1. In parallel, a dense network takes in a corresponding 4-band metadata vector which captures the anomalous pixel count, cloud-cover, and the x/y origin in the UXL coordinate system of the associated image. The outputs of each of these parallel network components are flattened, concatenated, and then passed through two ReLU fully-connected layers, followed by batch normalization and dropout layers to reduce over-fitting (See fig. 3).

For training, we utilized the Adam optimization algorithm [19] with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. Hyperparameter tuning on the optimal learning rate and mini-batch size was performed with instances of the Keras tuner hyperband class, in which randomized search was performed. Specifically, we optimized among 3 learning rates and mini-batch sizes each with evenly distributed exponential and linear intervals from (1^{-2}, 1^{-4}) and (24, 64) respectively. Additionally, the regularization parameter for the fully connected were trained on the exponential interval of (1^{-2}, 1^{-4}). The model was trained on a categorical cross-entropy loss $\text{Loss}_{\text{train}} = -\sum_{i=1}^{N} y_i \cdot \log \hat{y}_i$ for 200 epochs. The validation error metric equation was computed as $\text{Error}_{\text{val}} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)$.

In performing our experiments, we commenced by extracting images from a limited time window and geographic region of Illinois to verify our data pipeline functionality. Following promising classification results performed by baseline models, we expanded our dataset by surveying a broader population of Illinois counties and range of time. Using this enlarged dataset, we ran a comprehensive set of experiments to benchmark our novel context-to-image architecture against baseline models in both regression and classification settings. After elucidating and troubleshooting performance issues, we applied our optimal model to a transfer learning task in predicting yields in rural Brazil to assess performances on limited data. Comparisons in relative model performance were completed to provide more insight in the context of limited results.

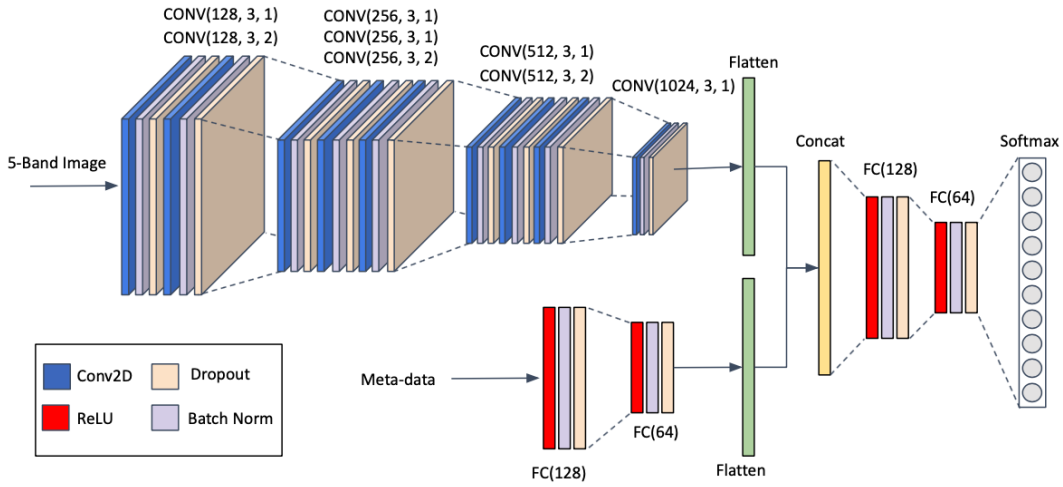

Figure 3: Diagram illustrating the context-to-image, concatenated-arms neural network architecture.

5 Results and Discussion

Our first set of experiments were conducted to obtain performance benchmarks from references architectures such as Resnet-50 and VGG-16. As seen in table 1, the low performance of the baseline models was an indication that there was an underlying issue with our dataset. Our original approach for collecting data was to extract the intersections of the county boundaries and PlanetScope "strips," as capturing the entire county in a singular image was impractical due to size constraints. As observed from the histogram comparing the county area to that of areas of interest (Fig. 5), we concluded that our model's learning process was likely inhibited by skewed area counts in favor of sparsity. In example, the images that aren't farm land would be assigned non-zero yield values, which would negatively impact training accuracy.

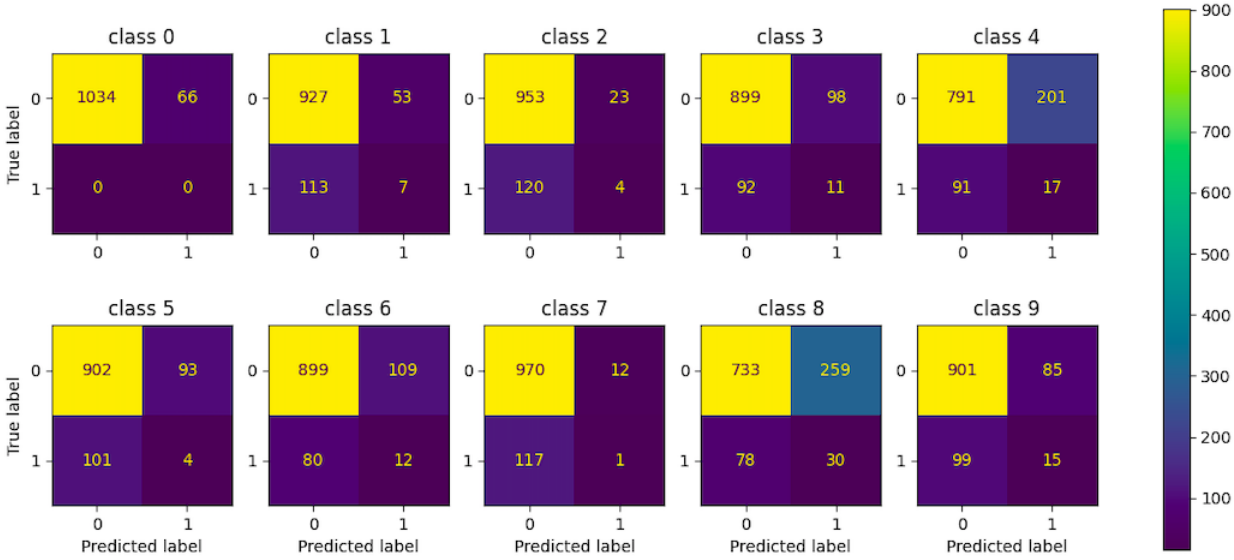

Figure 4: Confusion matrix of Deep-1 classification performance on Illinois data set. As observed, the model displays notably low levels of sensitivity with notable limits of predictions of class 3, 4, and 8, targets of future analysis.

Experiments conducted during our milestone period revealed similar limitations. For example, in running models with 7-band inputs, we observed that the inclusion of MSAVI and EVI bands induced severe over-fitting and inhibited model generalization, which led to a decision to exclude these indices in future experiments. Additionally,

poor class distribution assignments and manual assignments skewed the results of initial experiments. To circumvent this issue, we used a linear class assignment rule and under-sampled from skewed classes. Finally, we altered our labeling calculation metric based on relative image area rather than the number of unique satellite scenes. This metric captures less naive assumptions about our dataset but is still hypothesized to be limited due to the naive assumption that image area corresponds directly to crop yield. Thus, our method of image labelling is deemed to be a major limitation in our research and model performance. Further evidence of this limitation is seen in Fig. 4.

Despite lower performances by the baseline models, there are still notable insights derived from making direct comparisons between our model performances. Firstly, 5-band Deep-1 was the highest performing model evaluated under the Illinois dataset. One explanation for this is that it captures substantially more valuable data from vegetation indices and NIR bands compared to the baseline models. This observation was confirmed with subsequent experiments that revealed more limited deep-1 performance in 3-band and 4-band input scenarios. Secondly, in applying Deep-1 to our transfer learning application, we saw notable differences in model performance in the presence of pretrained weights. Specifically, the model with pretrained Illinois weights produced a training accuracy slightly higher than that of the model without pretrained weights. In comparing the validation performances between these models, the model pretrained on U.S. crop belt data achieved an accuracy of 0.354 while the alternate model achieved 0.254, displaying a significant difference. Despite our difficulties in evaluating test performances, our results support the notion that transfer learning can be successfully used to enhance crop yield prediction in rural, less data-rich areas. Inconsistencies observed in associated test metrics are hypothesized to arise from the difference in climate zones between the pretrained and trained regions. Specifically, qualitative analysis suggests that satellite scenes that do not encapsulate farmland could drastically differ between the two regions in character, thus posing learning challenges.

		Accuracy			Test Metrics		
Architecture	Region	Train	Val	Test	Precision	Recall	F1
Baseline: Resnet-50 3-bands [3]	Illinois	0.354	0.407	0.087	0.087	0.087	0.074
VGG-16 3-bands [4]	Illinois	0.365	0.400	0.095	0.097	0.095	0.075
Deep-1 5-bands [2, 1]	Illinois	0.434	0.436	0.092	0.093	0.092	0.079
Context-to-Image Deep-1 5-bands	Illinois	0.396	0.364	0.052	0.042	0.052	0.041
Context-to-Image Deep-1 5-bands	Brazil	0.254	0.140	0.000	0.000	0.000	0.000
Context-to-Image Transfer Learning	Brazil	0.278	0.354	0.000	0.000	0.000	0.000

Table 1: Table comparing classification performance across state-of-the-art and novel model architectures. As observed, dataset limitations resulted in severe overfitting across all models; however, relative comparisons indicate comparable results between the context-to-image and state-of-the-art models. Although challenges were faced in testing the Brazil-trained models, relative comparisons reveal that the transfer-learning approach is promising.

6 Conclusion and Future Work

This study features four novel contributions to the field of machine learning in crop yield predictions. Firstly, a novel dataset generation pipeline was created and tested to leverage the 3m resolution of satellite imagery for the Planet Labs Dove Constellation for machine learning applications. Secondly, evidence was discovered to support the use of NDVI and NIR image bands in parallel with traditional image pixel information that aligns with previous studies. Thirdly, a image-to-context model was designed and tested, displaying comparable performances to state of the art models in Illinois. This model supports the notion that image metadata can be leveraged to improve crop yield prediction tasks and build on previous studies that leverage metadata for standard classification tasks. Finally, promising yet limited results from the study's transfer learning application shows that pretraining on robust crop yield image data can boost prediction tasks in rural, data-lacking areas. Through our experiments, some key limitations that emerged included skewed image class distributions, naive assumptions in crop yield labeling, and general dataset construction challenges that arised from the Planet API. In future work, we hope to perform more detailed investigations on these limitations by firstly exploring more streamlined approaches to yield labeling to address model learning challenges. In addition, we plan to evaluate our models and approaches under more traditional sources of satellite imagery. With insights from these experiments, we hope to explore a more comprehensive set of time-series machine learning architectures and achieve a more consistent application of transfer learning across many rural regions for crop yield prediction.

7 Contributions

Ben:

- Led development of data pipeline using Planet Labs API
- Constructed video presentation
- Developed code for NN architectures
- Ran experiments and tuned model parameters
- Post-processing and data analysis
- Contributed to writing final paper and milestone

Chris:

- Assisted in development of data API. Extracted and processed ground truth data for Illinois and Brazil application.
- Developed code for NN architectures
- Ran experiments and tuned model parameters
- Post-processing and data analysis
- Contributed to writing final paper and milestone

8 Appendix

8.1 Data Pipeline

In our dataset, images from the state of Illinois were collected from raw satellite imagery acquired by the PS2 instrument from the PlanetScope Constellation. We constructed our dataset by firstly transferring ≈15,000 images into AWS S3 buckets using PlanetScope's Orders API. In the process, we created a log of image metadata including information such as image acquisition time, satellite scene ID, and snow cover. Leveraging this data with our collected ground truth, we were able to construct crop yield labels for our images based on scene ID and time statistics. We then conducted image processing, labeling, and compression of our raw image files into readable .npy files that were codified in a partition for our model data generator (Fig. 3). During image processing, we computed three additional bands by deriving the following vegetation indices: NDVI (Normalized Difference Vegetation Index), EVI (Enhanced Vegetation Index), and MSAVI (Modified Soil Adjusted Vegetation Index) as projected indicators of soil health and green vegetation in our images. Preceding these computations, we scaled our images by a factor of 0.04 using a bilinear resampling algorithm to standardize input data and reduce dimensionality.

Vegetation Index	Formula
Normalized difference vegetation index (NDVI)	$(NIR - R)/(NIR + R)$
Modified Soil adjusted vegetation index (MSAVI)	$(2NIR + 1 - ((2NIR + 1)^2 - 8(NIR - R))^{1/2})/2$
Enhanced vegetation index (EVI)	$2.5 * ((NIR - R)/(NIR + 6R - 7.5B + 1)$

Table 2: Equations for computing vegetation indices [9]

The input to our data pipeline consists of raw satellite imagery from the PlanetScope Constellation from Planet Labs. The processed data products acquired by the PS2 instrument consists of analytic orthotiles with four spectral bands (blue, green, red, NIR) with a nadir ground sampling distance of 3m. The orthotile product are radiometrically, geomtrically, and sensor-corrected and aligned to a cartographic UTM map projection. The PlanetScope constellation consists of more than 180 satellites in LEO orbit, allowing imagery of an area-of-interest to be sampled on a daily basis [20]. The dimensions of the raw PS2 orthotile data products ranged up to approximately 10,000 $\times$ 10,000 pixels. Due to the constraints of these large dimensions, we applied imaging scaling by a factor of 0.04 using a bilinear resampling algorithm. In an effort to standardize the input size, each raw image was centered in a 500 $\times$ 500 blank reference matrix.

8.2 Additional Figures

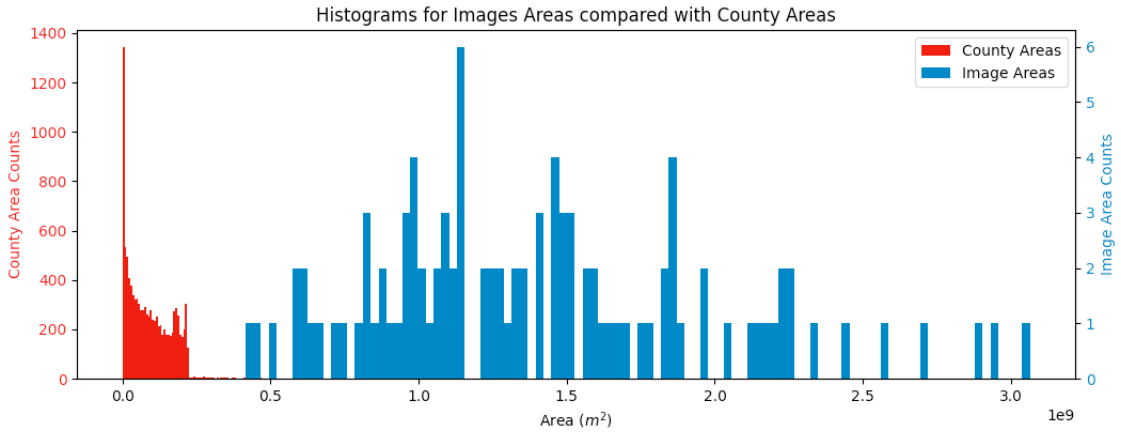

Figure 5: Histogram comparison between the county area and the image area of the Illinois data set. The large ratio between $\text{Area}_{\text{county}}/\text{Area}_{\text{image}}$ is one of the reasons the low performance of the baseline models.

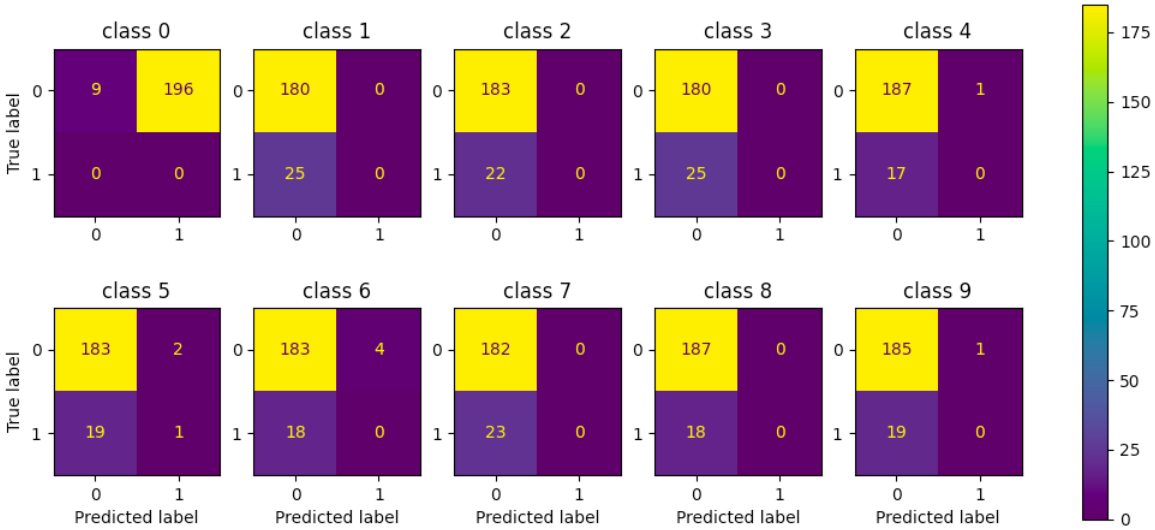

Figure 6: Confusion matrix for the context-to-image model with 5-band image arm (Deep-1) and metadata arm for the Brazil dataset. Major class imbalances were observed in class 0 due to labeling limitations.

References

[1] Jiaxuan You, Xiaocheng Li, Melvin Low, David Lobell, and Stefano Ermon. Deep gaussian process for crop yield prediction based on remote sensing data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

[2] Mark Sabini, Gili Rusak, and Brad Ross. Understanding satellite-imagery-based crop yield predictions. *Stanford*, 2017.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[4] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[5] Andreas Kamilaris and Francesc X Prenafeta-Boldú. Deep learning in agriculture: A survey. *Computers and electronics in agriculture*, 147:70–90, 2018.

[6] Anna X Wang, Caelin Tran, Nikhil Desai, David Lobell, and Stefano Ermon. Deep transfer learning for crop yield prediction with remote sensing data. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, pages 1–5, 2018.

[7] Thomas van Klompenburg, Ayalew Kassahun, and Cagatay Catal. Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177:105709, 2020.

[8] Saeed Khaki and Lizhi Wang. Crop yield prediction using deep neural networks. *Frontiers in plant science*, 10:621, 2019.

[9] Gaolong Zhu, Weimin Ju, JM Chen, and Yibo Liu. A novel moisture adjusted vegetation index (mavi) to reduce background reflectance and topographical effects on lai retrieval. *PloS one*, 9(7):e102560, 2014.

[10] David M Johnson. An assessment of pre-and within-season remotely sensed variables for forecasting corn and soybean yields in the united states. *Remote Sensing of Environment*, 141:116–128, 2014.

[11] David B Lobell. The use of satellite data for crop yield gap analysis. *Field Crops Research*, 143:56–64, 2013.

[12] Kevin Tang, Manohar Paluri, Li Fei-Fei, Rob Fergus, and Lubomir Bourdev. Improving image classification with location context. In *Proceedings of the IEEE international conference on computer vision*, pages 1008–1016, 2015.

[13] Jeffrey S Ellen, Casey A Graff, and Mark D Ohman. Improving plankton image classification using context metadata. *Limnology and Oceanography: Methods*, 17(8):439–461, 2019.

[14] U.S. Census Bureau. Illinois counties shp and kml files, 2019 cartographic boundary files. retrieved from https://www.census.gov/geographies/mapping-files/time-series/geo/cartographic-boundary.html. 2019.

[15] Usda national agricultural statistics service. *USDA National Agricultural Statistics Service*, 2019.

[16] Instituto Brasileiro de Geografia e Estatística. Municipal boundaries: Brasil, 1991. retrived from https://purl.stanford.edu/gk118zk6545. 2011.

[17] Maheswari Visvalingam and James D Whyatt. Line generalisation by repeated elimination of points. *The cartographic journal*, 30(1):46–51, 1993.

[18] Instituto Brasileiro de Geografia e Estatistica Brasil Sistema IBGE de Recuperacao Automatica. Producao agricola municipal: producao das lavouras temporarias.. retrived from https://sidra.ibge.gov.br/tabela/1612. 2011.

[19] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[20] Planet application program interface: In space for life on earth. 2017.