Using Aerial Images to Predict Biden's Vote Share in the 2020 General Election

Michael Byun (mbyun@stanford.edu)
Catherine Chen (tche101@stanford.edu)
Dawson Verley (verley@stanford.edu)

Introduction

The American political system is characterized by a growing partisan divide. For instance, the debates featured by the Congress and news media convey more partisan messages (Prior, 2013); people from both parties report an increased level of hatred toward the opposite party (Iyengar et al., 2019); and voters' ideal position in policies become further away (Krasa & Polborn, 2014). At times, it seems that we are living in two Americas characterized by bitter ideological opposition. But is this polarization so entrenched that it's apparent in the physical environment of American towns and cities? In other words, are we literally living in two separate Americas? The current project applies deep learning to predict partisanship using aerial imagery.

The connection between built environments, physical space, and politics is not obvious, making this a challenging task. However, it's certainly possible that local politics could be reflected by physical features like building construction, infrastructure, public transportation, tree cover, and bike lanes. Likewise, features of the built environment like spatial segregation may create and reinforce partisan divides. We're interested in whether these effects are large and generalizable enough that aerial imagery can be used to predict partisanship reliably. If so, this would speak to the profound depth of the partisan divide in America and open up new lines of research in fields such as Urban Studies and Political Science.

Data

The current project uses aerial imagery to predict Biden's vote share in Florida in the 2020 General Election. The aerial images consist of 200k image tiles obtained from the National Agriculture Imagery Program (NAIP). The source was selected because their images were largely cloudless with a 1m resolution and had water masked out. We segmented each 3.75x3.75 minute JPEG2000 image into 49 tiles, resampled each tile into a 256x256 pixel square PNG, and converted the image from 4-channel RGBA to 3-channel RGB. We then reverse-engineered NAIP image tile file naming schema to recover coordinates for image centers.The output included granular precinct-level voting data from the 2020 US presidential election, sampled at the midpoint of each tile. See Figure 1 for example images.

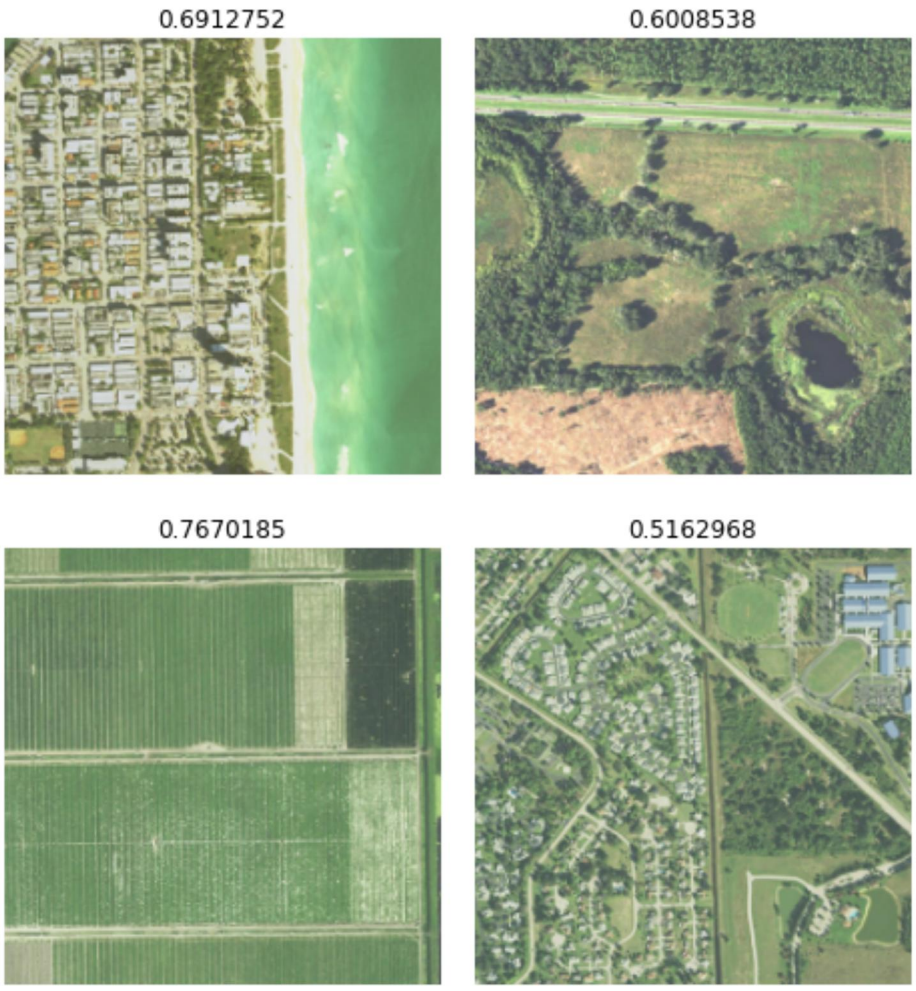

Figure 1. Random samples from the dataset. The numbers on top of each image are the Biden vote share of the election precinct that each image's center belongs to.

Data cleaning

We filtered out examples which had no available precinct election outcomes from our dataset (e.g. NAIP image segments taken over coastal or intertidal regions outside of established precinct boundaries). Then, we made several attempts to remedy the fact that an overwhelming majority of the dataset was uninhabited land, that was less explanatory of voting. We probabilistically filtered out examples with low Biden percentage, randomly keeping 3% of examples with Biden vote share below 30%, and 20% of examples with Biden vote share below 45%. We experimented with different thresholds and probabilities to maximize the proportion of populated/built-up land in the dataset.

Learning Methodology

We sliced the data into train/dev/test with a take-and-skip method. The training set consists of N = 20,000, dev set N = 5,000, and the remaining images were put into the test set. We batch-normed the dataset with batch size 32. We implemented transfer learning with InceptionV3 trained on ImageNet (see Figure 2). The input of our model was 256x256x3 RGB images, and the last layer was a sigmoid given we were predicting vote share in [0,1]. We trained our model in Google Colab, since our data was in Google Drive.

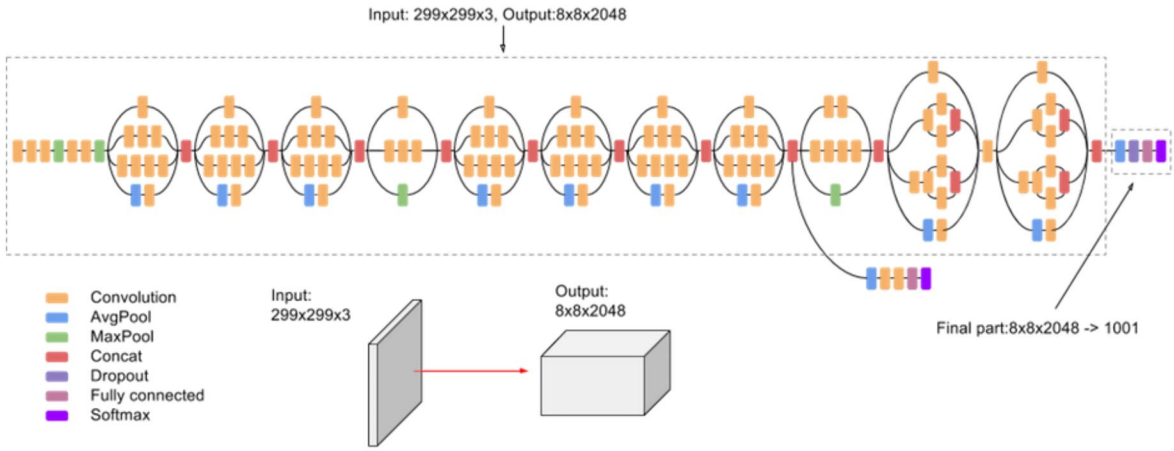

Figure 2. Illustration of the model.

Note: We replaced the top of the model with average pooling, a 32-unit fully connected layer, and a fully-connected output node with a sigmoid activation function. The input of our model was 256x256x3. Image from Google Cloud TPU Documentation under CC-BY-4.0.

Results

Using mean squared error as the evaluation metric, the train error was 0.16. The test error was 0.31 on 10 randomly sampled examples from the test set (evaluating on 100 sampled examples takes more than a couple hours; tried twice). For all examples, the model predicts 1, despite the input features being different (see Figure 3).

Figure 3. Examples of images of which the predicted Biden vote share was 1

We then spent some time analyzing why the model predicts 1 for many images despite their heterogeneous landscape features. We explored whether there were coding errors in our model implementation, but have not identified any so far. We also looked for problems in the labels of our training data, but none of the examples were labeled 1, and the mean/median label was definitely lower than 1.

The uniformity of the predictions may be a consequence of a vanishing gradient problem caused by the sigmoid output activation, the depth of the Inception V3 network, and the relative scarcity of training examples — a challenge that a different network architecture might address more effectively.

As a sanity check, we also trained a much simpler CNN on the data. Using only two convolutional layers with 32 and 64 filters, respectively, two pooling layers, a fully-connected sigmoid output layer, and Xavier initialization, we achieved slightly worse performance on the

train and test sets. This suggests that transfer learning was an appropriate technique. Moreover, the failure of both approaches suggest that this is an extremely difficult problem with a high signal-to-noise ratio.

Most of the data were random unpopulated bits of swamp or farmland. We tried to make the dataset more balanced by randomly excluding most examples below a Biden-percentage threshold, but this left us with a lot of random unpopulated bits of swamp or farmland that happened to have a high Biden percentage. A better dataset that contains more built-up area—e.g. one which samples image locations based on population density or land cover—may produce better results. Alternatively, a more complex neural network architecture that first identifies features like buildings and roads then uses attributes of those features to make predictions may also perform better. More time, data, and compute power may also produce better results. Finally, Colab and Google Drive have extremely cumbersome file I/O limitations which significantly hinder the speed of training, evaluation and iteration. This issue may be resolved through the use of memory-mapped HDF5 files in the data pipeline. While the implementation of this technique can be complicated and error-prone, it may improve the overall productivity of this deep learning project. We suggest these approaches for future work.

References

"Advanced Guide to Inception v3." 2022. Google Cloud TPU Documentation, CC BY 4.0.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, *22*, 129-146.

Krasa, S., & Polborn, M. (2014). Policy divergence and voter polarization in a structural model of elections. *The Journal of Law and Economics*, *57*(1), 31-76.

Park, A., Watkins, M. et al (2021). Presidential precinct data for the 2020 general election. *The New York Times*.

Prior, M. (2013). Media and political polarization. *Annual Review of Political Science*, *16*, 101-127.

Aerial Photography, The National Agriculture Imagery Program (2016). Farm Service Agency, U.S. Department of Agriculture.