

# CSCI 1430 Final Project Report:

## Urban Canopy Cover

Elizabeth Zhang, Jenny Tan, Filip Kierzenka, Alexander Kamper  
Brown University

### Abstract

*Expanding urban tree canopy cover is an important strategy for mitigating climate change, but estimating green space is an unresolved problem for spatial researchers. We aimed to reimplement Cai et al.'s approach using convolutional neural networks and compared it with a simpler architecture, DeepGreen. We also mapped and visualized the results for Providence's canopy cover, and interpreted the model using the LIME explainer and Grad-CAM. We achieved a mean-absolute error for the tree canopy cover estimation of 0.0909 for Cai et al.'s approach and 0.0969 for the DeepGreen approach.*

### 1. Introduction

Expanding urban tree canopy cover is one strategy for mitigating climate change, and a robust body of research indicates that green space is associated with a diversity of positive psychological and physiological effects for urban residents. Aside from carbon capture, enhancing wellbeing, and providing other ecosystem services, green spaces are beautiful and enhance the experience of urban dwelling. One method for estimating urban green space is by measuring tree canopy cover. However, quantifying urban tree canopy cover at scale is an unresolved problem for spatial researchers. Our goal was to replicate Cai et al.'s (2019) [1] novel approach and apply it to the city of Providence. Cai et al. (2019) applied two neural networks to street-level rather than aerial imagery in order to perform image segmentation and calculate greenness scores point-by-point along a city's road network. The details and background research supporting this work are discussed in the method section. Assuming our process would be streamlined by referencing code published by the authors and MIT's Senseable City Treepedia project, we defined several intermediate and stretch goals. First, we wanted to map and visualize our results in a way that would be intuitive for our audience and potential researchers interested in Providence's urban canopy cover. Second, we wanted to compare the model's performance against slightly different architectures. Finally, we wanted to interpret our

model using the LIME explainer from homework 5 and Gradient-weighted Class Activation Mapping (Grad-CAM).

### 2. Related Work

As we concretized our project aims, we conducted an informal literature review on computer vision techniques for quantifying urban canopy cover. Based on our sampling, this body of research can be split primarily into two frameworks. First, there are approaches which attempt to segment canopy cover from satellite imagery and calculate it as a proportion of urban area. This idea is straightforward and falls within the standard set of techniques for geographic inquiry since the advent of geographic information systems (GIS) and the widespread availability of up-to-date, high resolution satellite images. The second approach proposes the deployment of UAVs, or drones, so that researchers can capture their own image data across whatever wavelengths and at whatever resolution their sensory equipment can capture (Seiferling, 2017) [5]. Research in this area is often coupled with research into machine learning approaches to autonomous flight so that data collection may be streamlined and standardized (Moreno-Armendáriz, Marco A et al., 2019) [4]. This line of research breaks with the tradition of much modern geographic research and intrigues with its seeming orientation towards citizen science. Ignoring differences in granularity and access, however, these two approaches pose fundamentally the same engineering challenges for computer vision.

### 3. Method

We utilized a combination of images from Google Street View and the Cityscapes dataset ( 8000 images total). We used the original paper's library to get the Google Street View images from five cities: Cambridge, Johannesburg, Oslo, Sao Paolo, and Singapore. The Cityscapes dataset originally had multi-class labels, and we converted the dataset to have binary labels for vegetation and non-vegetation. 80% of the total images were used for training, and the other 20% were used for validation. The images' labels were their Green View Index - a pixel level average of canopy cover.

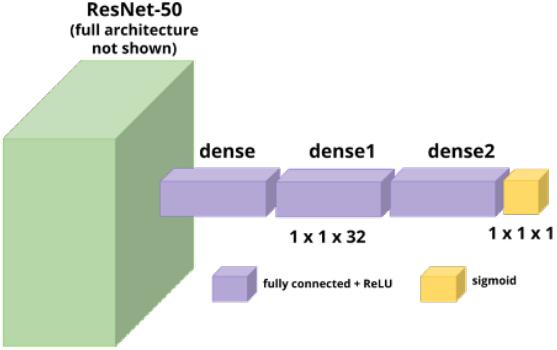


Figure 1. Deep Convolutional Neural Net (DCNN) Architecture

Hyperparameter	Value
Batch size	50
Learning rate	0.0001
Momentum	0.01
Optimizer	RMSProp

Table 1. Hyperparameters referenced from Cai et al.’s methodology (DCNN)

Hyperparameter	Value
Batch size	50
Learning rate	0.01
Optimizer	Adam

Table 2. Hyperparameters referenced from DeepGreen

GVI labels for Cityscapes were calculated using the “vegetation” mask in that dataset and the labels for GSV calculated via the original paper’s GVI library. GSV images and metadata for Providence were collected using our own modified version of the original paper’s methods.

We re-implemented the convolutional neural network that was used in Cai’s approach. To reduce training time and computational power required, we used a ResNet50 architecture pre-trained on ImageNet and a dense classification head (Figure 1 and Table 1). For the Deep Green Model, we used 5 convolutional blocks with a dense classification head that were not pretrained (Figure 2 and Table 2). We decided on this model to determine how a decrease in model complexity would impact overall results.

We trained the model on the Google Street View images supplemented with the Cityscapes dataset. For gradient descent, we used mean absolute error because we wanted to minimize the difference between the predictions and the labels.

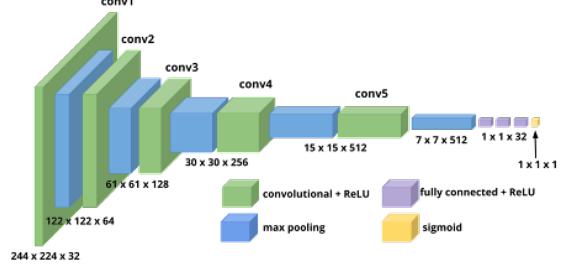


Figure 2. DeepGreen Architecture

Model	Lowest val. MAE	Epoch Achieved
DCNN	0.0909	13
DeepGreen	0.0969	18

Table 3. Training results for both of our models.

## 4. Results

We reimplemented both the DCNN model and the DeepGreen model, and were able to achieve a mean-absolute error of 0.0909 and 0.0969, respectively (Table 3). We found that the DeepGreen model’s performance was similar in terms of mean-absolute error, and that the DeepGreen model was generally better than the DCNN model at distinguishing between vegetation and non-vegetation in the LIME explainer images. We have shown that a smaller, simpler architecture can still be effective at estimating the GVI. We did succeed in applying our model to Providence, but due to unknown errors with our code for sampling images from the Google Street View API, we couldn’t calculate a GVI for every point we wanted. We tried reducing the number of calls to the API by placing points every 200m on the Providence street network instead of every 20m, which decreased the total number of points we needed to collect data for by about 80%. We still failed to acquire a GVI for each point, but the points we could get data for were arranged in 4 bands running East to West across the city (Figure 4). This at least allows us to see that GVI values for Providence accord with what we’d expect in particular areas, such as Blackstone Park in the North East corner of the city being especially green, and downtown Providence in the center of the middle band being especially red.

### 4.1. Technical Discussion

Jun Yang and colleagues pioneered a novel approach to the quantification of urban canopy cover in 2009, proposing that instead of a top-down view of the city, researchers consider a pedestrian’s perspective to calculate what they termed a “Green View Index” (GVI) (Yang et al., 2009) [6]. Their approach was laborious, requiring researchers to manually collect and analyze images of the city on foot. Later,

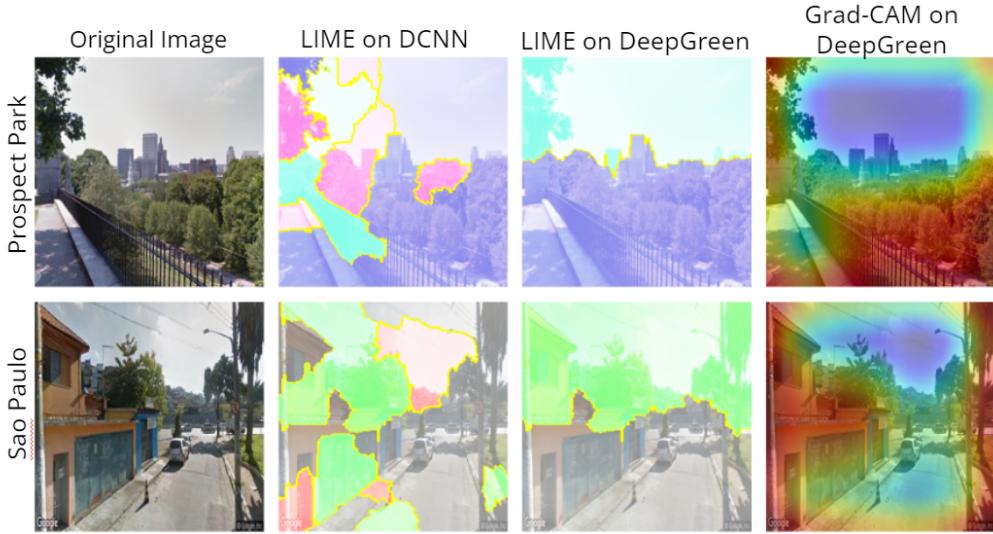


Figure 3. LIME and GradCAM results for the DCNN and DeepGreen model

Xiaojiang Li et al., optimized this process by developing a program which could sample data from Google Street View (GSV). However, their automated technique for finding green vegetation was simplistic, simply calculating differences in the red, blue, and green bands of their dataset (Li et al., 2015) [3]. Following an explosion of research in deep learning, Bill Cai collaborated with Li and a colleague in MIT’s Senseable City Lab to refine the GVI methodology using deep learning. They deployed a deep convolutional neural net (DCNN) to perform semantic segmentation on vegetation in GSV images, then used a second CNN with modified ResNet architecture to perform end-to-end learning on its outputs and calculate a GVI score for each coordinate at which the image was sampled.

We were able to achieve a mean-absolute error of 0.0909 from Cai et. al.’s model after 13 epochs. This was higher than the mean-absolute error described in the paper, which was 0.0467. This could be due to a difference in hyperparameters used in training, since the paper did not specify them. Additionally, we were able to achieve a mean-absolute error of 0.0969 using the DeepGreen model after 18 epochs (Table 3).

The DeepGreen model was better at classifying vegetation. In the images generated by the LIME explainer, we can see that the vegetation is marked, but other non-vegetation objects that were green in color were ignored. This did not seem to be the case in the DCNN model, as the DCNN highlighted a green garage door in the Sao Paulo test image (Figure 3). However, for the DeepGreen model, there was a significant difference between the LIME and GradCAM results. The LIME image for the Sao Paolo image shows the correct classification of vegetation, but the GradCAM image

marks the buildings as more important than the vegetation. We were surprised by this disagreement, but it may be a matter of interpretation. GradCAM’s “hot zones” refer to regions which, in the last CNN layer, increase the outputted value - it’s possible that the overall values were generally too high before the final layer, and so the final CNN layer was reducing the value because of the vegetation. This would explain why the region with visible vegetation were blue in the GradCAM results.

#### 4.2. Societal Discussion

Research in political ecology and related disciplines has consistently shown how urban green spaces are distributed inequitably according to socioeconomic factors and legacies of structural racism. Heynens et al., (2006), performed a case study in Milwaukee, Illinois and found that canopy cover was correlated positively with both median household income and non-Hispanic white residents, demonstrating how richer, white residents had greater access to the psychological and physiological benefits of green space as compared to poorer, nonwhite residents (Heynens et al., 2006) [2]. Urban tree canopy hence falls under the larger issue of urban environmental inequality and environmental justice more broadly.

We received five points of feedback on the societal implications of our project through the SRC peer critique. They are outlined below, accompanied by our response to each.

1. The Street View car does not pass through all city streets, so that there are some remote places that may be ignored, which is unfair to the people who live there.

This is true, but we also establish a sampling method of every 20 meters for computational efficiency so we acknowl-

edge at the outset that this is not the most high resolution method possible for calculating urban canopy cover. One possible avenue for future research would be to set up our own street view data collection system with a mounted camera on a bike or car. Alternatively, cross-referencing our output with models using satellite or drone imagery could help address these blindspots in our data.

2. The privacy of others may be violated when getting the street view, because people generally do not notice the street view car passing by them.

GSV is a legal project and takes measures to blur personally identifiable information like faces and license plates. Because the model outputs a greenness score which can then be mapped back onto the city, images of people will never show up in the outputs of our project.

3. The cities chosen are among the more developed areas. Calculating tree canopy area only in developed cities may affect the equity of development in different areas.

Our aim is actually to apply this model to Providence, a relatively small city which wasn't included in the set of global cities the Treepedia project included in their publications. We've discussed the possibility of continuing this work outside the scope of this class and considered creating a web app or some documentation that will make it easier for people without a programming background to create urban canopy cover maps of their own cities.

4. The data obtained by using only street view vehicles can only obtain data on the tree canopy around the driveways, but some walkways or large green areas where there are no driveways, these data will not be obtained, thus giving a bias to the overall data correctness.

We sample images from a city (in this case Providence) by placing points every 20m on all roads, then querying the GSV API for panoramic images at those points. So its true that our method is somewhat constrained to roads. However, many parks and other green spaces are bordered by streets so we'd expect them to contribute to the greenscore of the city. Once again, cross-referencing our output with models using satellite or drone imagery could help remedy this issue.

5. Some cities will have a higher percentage of evergreen trees in all seasons, and some cities will have trees that lose all their leaves in the fall. If these differences are not differentiated, they may bias the correctness of the results.

During preprocessing we select images only from seasons where trees are green, which resolves this issue. However, there is merit to the fact that some urban areas are greener for longer periods of the year than others. This temporal component is not something we consider in the green score but perhaps could be an avenue for future research.

## 5. Conclusion

In the future, more research can be done to experiment with various kinds of architectures in order to improve the

accuracy of estimating the GVI. Additionally, creating more mappings of the GVI in various geographic locations could be of interest to potential researchers who are interested in urban tree canopy cover. In the early stages of our project we discussed potentially converting our work into a web application that researchers and interested individuals could use to easily enter their location of interest and extract a set of green view indices for research. Visually this could be made more beautiful and engaging for non-academic users by embedding interactive web maps that would visualize GVI for the area of interest and have multiple toggle-able layers so the user could reference GVI against environmental justice factors like household income, race, and relevant health outcomes. Ultimately, we didn't have time for this component and there remained a few bottlenecks that would need to be resolved before moving forward. Namely, due to lack of experience with the Google Streetview API, our code for collecting images and metadata behaved erratically for no apparent reason. Addressing this kink in the city choice, to point creation, to green view calculation and mapping pipeline would be a critical first step toward making this project into a potentially useful research tool.

## References

- [1] Bill Yang Cai, Xiaojiang Li, and Carlo Ratti. Quantifying urban canopy cover with deep convolutional neural networks. *ArXiv*, abs/1912.02109, 2019. [1](#)
- [2] Nik Heynen, Harold A. Perkins, and Parama Roy. The political ecology of uneven urban green space. *Urban Affairs Review*, 42:25 – 3, 2006. [3](#)
- [3] Xiaojiang Li, Chuanrong Zhang, Wei Li, Robert M. Ricard, Qingyan Meng, and Weixing Zhang. Assessing street-level urban greenery using google street view and a modified green view index. *Urban Forestry & Urban Greening*, 14:675–685, 2015. [3](#)
- [4] Marco A. Moreno-Armendáriz, Hiram Calvo, Carlos A. Duchanoy, Anayantzin P. López-Juárez, Israel A. Vargas-Monroy, and Miguel Santiago Suárez-Castañón. Deep green diagnostics: Urban green space analysis using deep learning and drone images. *Sensors*, 19(23), 2019. [1](#)
- [5] I. Seiferling, N. Naik, C. Ratti, and R. Proulx. Green streets quantifying and mapping urban trees with street-level imagery and computer vision. *Landscape and Urban Planning*, 164:93–101, 2017. [1](#)
- [6] Jun Yang, J. Yang, Linsen Zhao, Joe Rayl McBride, and Peng Gong. Can you see green? assessing the visibility of urban forests in cities. *Landscape and Urban Planning*, 91:97–104, 2009. [2](#)

## Appendix

### Team contributions

Please describe in one paragraph per team member what each of you contributed to the project.

**Elizabeth** I was responsible for hooking up the LIME explainer to our model. I was also responsible for implementing the GradCAM visualizer, which required integrating a Sequential model version of our DeepGreen model into our code, as well as producing superimposed images for the final figures. I was responsible along with Filip for getting the data loaded into the program, which included looking into the Tensorflow documentation to figure out how Dataset objects worked, how to shuffle and partition the data, etc. Additionally, I designed the model architecture figures, helped with the layout and wording of our poster, as well as cleaned up the code for the final submission.

**Jenny** I was mainly responsible for implementing the model architecture for the DCNN and DeepGreen model, and I read the research papers for both models to determine the correct hyperparameters and loss function. I trained the DeepGreen model and modified the code so that it would run. Additionally, I acquired the Cityscapes data, and preprocessed it so that the multi-class labels were converted to binary labels for vegetation and non-vegetation. I also wrote the Abstract, part of the Method and Results section, and added all of the citations to the final paper. I also worked on the Introduction and Discussion section of the poster.

**Alexander** I was mainly responsible for acquiring, processing, and visualizing the data for Providence. I was able to reference some code online but ultimately decided to work with the Google Street View API python package with help from Filip. This entailed manipulating shapefiles, reading from them to make calls to the GSV API, applying the model to its output, then storing the GVI values with their coordinates as a CSV file so they could be projected as points in GIS with a graduated symbology. I wrote sections 1, 2, and 4.2 of this write-up and contributed to others. I also contributed to the poster and took a lead role in the research/literature review phase of this project.

**Filip** I was responsible for getting the model training functionality working. I worked with Lizzy to write the training code, figure out how to best setup our datasets, and connect our model to that training code. I also did some work on the data side with Alex and Jenny; mainly helping to collect the Providence images via the GSV API, helping format the data for prediction with our model, and format the models' output. I collected the images and did some writing for the results and methodology sections of the poster.

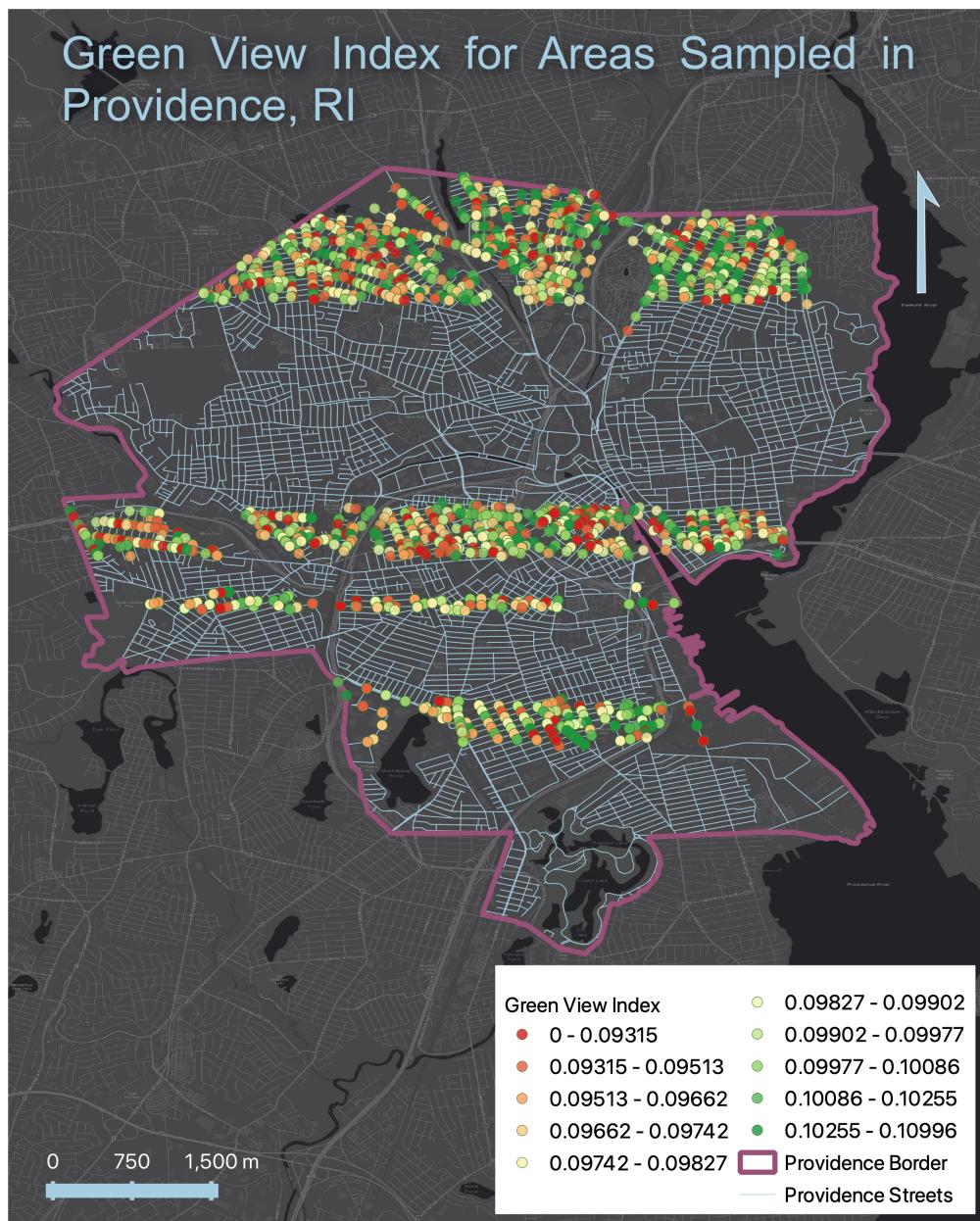


Figure 4. GVI map for Providence