A Machine Learning Approach to Recognize Environmental Features Associated with Social Factors

Jonathan T. Diaz-Ramos

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
in
Computer Engineering

Creed F. Jones, Chair
Julia M. Gohlke
A. Lynn Abbott

May 7, 2024
Blacksburg, Virginia

Keywords: computer vision, object detection, image segmentation, deep learning

A Machine Learning Approach to Recognize Environmental Features Associated with Social Factors

Jonathan T. Diaz-Ramos

(ABSTRACT)

In this thesis we aim to supplement the Climate and Economic Justice Screening Tool (CEJST), which assists federal agencies in identifying disadvantaged census tracts, by extracting five environmental features from Google Street View (GSV) images. The five environmental features are garbage bags, greenery, and three distinct road damage types (longitudinal, transverse, and alligator cracks), which were identified using image classification, object detection, and image segmentation. We evaluate three cities using this developed feature space in order to distinguish between disadvantaged and non-disadvantaged census tracts. The results of the analysis reveal the significance of the feature space and demonstrate the time efficiency, detail, and cost-effectiveness of the proposed methodology.

A Machine Learning Approach to Recognize Environmental Features Associated with Social Factors

Jonathan T. Diaz-Ramos

(GENERAL AUDIENCE ABSTRACT)

In this thesis we aim to supplement the Climate and Economic Justice Screening Tool (CEJST), which assists federal agencies in identifying disadvantaged census tracts, by extracting five environmental features from Google Street View (GSV) images. The five environmental features are garbage bags, greenery, and three distinct road damage types (longitudinal, transverse, and alligator cracks), which were identified using image classification, object detection, and image segmentation. We evaluate three cities using this developed feature space in order to distinguish between disadvantaged and non-disadvantaged census tracts. The results of the analysis reveal the significance of the feature space and demonstrate the time efficiency, detail, and cost-effectiveness of the proposed methodology.

Dedication

This thesis is dedicated to Lola, Tía Beatriz, and Tío Edwin who demonstrated examples of enjoyment, compassion, and resilience. Your memories live on in the pages of this work.

Acknowledgments

My first thank you is extended to my committee members who supported me at every step of this journey. I owe special thanks to Dr. Jones who helped me refine my initial concept into the work you will find in the chapters of this thesis. Dr. Gohlke, whose distinctive insights from beyond the field of machine learning, shaped the main goal of this thesis. I am also grateful to Dr. Abbott for introducing me to the world of computer vision through his engaging class lectures. Thanks to my Mother, whose unparalleled work ethic, discipline, and perseverance inspired me to push forward in challenging times. My brother for the guaranteed laughter after every phone call. My sister's supportive talks gave me self-awareness which was invaluable in this journey. Thanks to my close friends for believing in me and providing me with the ability to see different perspectives of my work. A final thanks to my middle school AVID teacher, who first introduced me to a college degree. And to all the First-Generation and/or Low-Income (FGLI) students, may the work of this thesis inspire you to earn and pursue a graduate degree.

Contents

List of Figures

List of Tables

List of Abbreviations

γ Aspect Ratio of the Elliptical Gaussian Envelope for the Gabor Filter

λ Wavelength of the Sinusoidal Component for the Gabor Filter

ϕ Phase Relationship in Response to Texture Pattern for the Gabor Filter

σ Receptive Field and Localization of the Gabor Filter

θ Orientation of the Kernel Window for the Gabor Filter

CCA Connected Component Analysis

CEJST Climate and Economic Justice Screening Tool

CEQ Council of Environmental Quality

CNN Convolutional Neural Network

CRDDC Crowd sensing-based Road Damage Detection Challenge

GSV Google Street View

LOOCV Leave-One-Out Cross-Validation

mAP Mean Average Precision

MRLC Multi-Resolution Land Characteristics

MSA Metropolitan Statistical Area

NLCD National Land Cover Database

PCA Principal Component Analysis

SAM Segment Anything Model

SVM Support Vector Machine

YOLO You Only Look Once

Chapter 1

Introduction

1.1 Motivation

The Council of Environmental Quality (CEQ) was initiated in 1969 by the National Environmental Policy Act [1]. The responsibility of the council is to advise the President of the United States in developing policies in efforts regarding climate change, public lands, oceans, and other areas. In order to advise the President, the council works closely with Federal environmental efforts and agencies. One of its main responsibilities is to review infrastructure projects and other federal actions that reflect the input given by the public and local communities [2].

In 2022 the CEQ received an executive order to develop CEJST helping Federal agencies identify disadvantaged census tracts [3]. Census tracts, typically containing between 1,200 and 8,000 residents, are statistical subdivisions of a county [4]. The executive order was to ensure the benefits of federal programs were reaching census tracts that have been identified with pollution and historic underinvestment. These inequalities were restated by the CEQ Chair Brenda Mallory, "the Climate Economic Justice Screening Tool identifies communities that have faced historic injustices and have borne the brunt of pollution so we ensure they're some of the first to see the benefits of climate action."[3]

The executive order established the Justice40 initiative, providing an overview of the impact

key decisions have on disadvantaged communities. The Justice40 initiative stated that 40% of the overall benefits of certain federal investments flow to disadvantaged communities. These investments focus on key areas such as climate change mitigation, clean energy, clean transit, affordable and sustainable housing, workforce development, pollution remediation, and critical water infrastructure. Federal agencies are tasked with regulating and implementing investments [28]. Additionally, federal funding supports institutions like schools, hospitals, and fire departments, which rely on this funding to perform their daily tasks [9]. Ineffective flows of investments can affect these institutions and communities then having a negative impact on the residents that reside within census tracts.

CEJST holds valuable information about individual census tracts across the United States. The tool splits datasets into categories of burdens which include climate change, energy, water and wastewater, health, housing, legacy pollution, transportation, and workforce development. The categories provide more detailed information. For example, the housing category considers the lack of green space in a census tract, while workforce development addresses issues such as poverty and unemployment experienced by residents. Also, climate change notes the expected agriculture loss rate for specific time periods.

Using these categories of burdens census tracts are categorized as disadvantaged based on two factors. The first is whether they meet the criteria for a variation of environmental, climate, or other burdens. The second is if a community meets a socioeconomic burden. The threshold for defining a disadvantaged census tract is shown in the table below:

Categories	**Indicators at or Above the 90th Percentile**
Climate Change	Expected agriculture loss rate, population loss rate, and building loss rate
Health	Low life expectancy, asthma, diabetes
Housing	Lack of green spaces, housing costs
Workplace Development	Linguistic isolation, low median income, poverty, unemployment
Legacy Pollution	Proximity to hazardous waste, super-fund sites, risk management, and risk management plan facilities
Energy	Energy Costs
Transportation	Diesel particulate matter exposure, transportation barriers, traffic proximity and volume
Water and Wastewater	Underground storage tanks and releases, wastewater discharge

Table 1.1: Table shows the key descriptors for defining a disadvantaged census tract. CEJST uses the details above in defining a disadvantaged census tract. Once an environmental and socioeconomic descriptor is met a census tract is considered to be disadvantaged.

The datasets used by CEJST are gathered to construct the categories shown. For example, the housing category includes data on green spaces within a census tract, which is derived from satellite imagery.

The CEQ acquired the National Land Cover Database (NLCD) provided by the Multi-Resolution Land Characteristics (MRLC) consortium which uses machine learning algorithms to note green space from satellite imagery. NLCD highlights multiple areas present in satellite imagery which include agriculture, forest grassland, shrubland, and more. The database tends to be updated on a yearly basis with some exceptions. For example, from 2016 to 2019 there was no update provided by the MRLC consortium [6]. This begins to show the challenges CEJST faces, as its sources for gathering information about census tracts can take well over a year.

Time inefficiencies are again shown by the United States Census Bureau. CEJST uses the socioeconomic indicators identified by the U.S. Census Bureau to assess the well-being of communities and identify census tracts that are of concern [7]. To gather the socioeconomic indicators an extensive process taking 10 years is conducted. The most recent survey which began in 2010 will be concluded in 2024, showing that some surveys can well exceed the 10-year intended time period.

Figure 1.1: The Google Street View images show four distinct roads across the United States. In each image, we observe an object detection model making a prediction on the different types of road damage present. The predictions are tagged by a category followed by a confidence score. D00 represents longitudinal cracks and D10 transverse cracks

Another challenging factor CEJST experiences is missing information that can potentially guide federal agencies into making prompt funding decisions for specific census tracts. As

an example, in 2019 the Government Accountability Office (GAO) observed that most road pavement conditions vary widely across different states; though, they recognized a consistent trend in observing poor pavement in census tracts with higher family poverty rates [8]. Some examples of road damage can be seen in *Figure 1.1.* This investigation led to an informative finding but required a vast amount of time and effort. The study begins with a discussion of whether environmental features can be recognized in a more time-efficient manner.

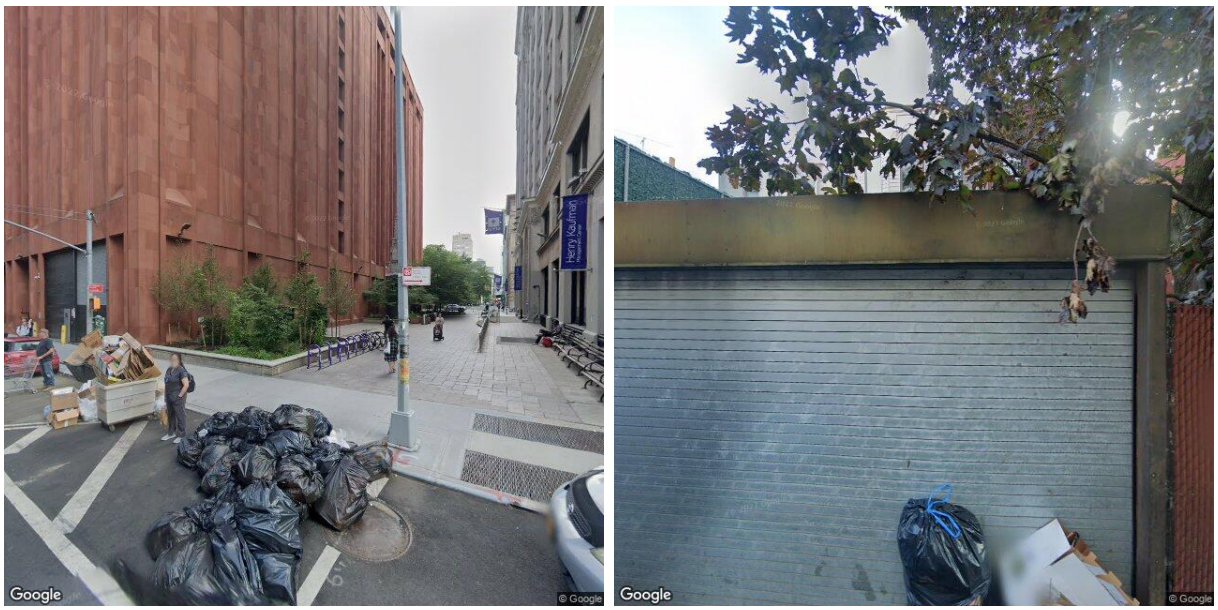

Figure 1.2: The two Google Street View images show garbage bags present in an urban environment. Both images are from Manhattan, NY. The image on the left gives a pile of black garbage bags sitting on the side of a city street. The right image contains a garbage bag along with a cardboard box outside a garage door.

An additional feature provided by the U.S. Census Bureau that aids in federal funding decisions was identified in the U.S. Census Bureau's American Housing Survey [10]. The survey recognized a link between higher levels of litter in low-income neighborhoods, which contributes to discussions on funding decisions. Also, the presence of any form of litter, like garbage bags, can give insights into the effectiveness of trash collection services in census tracts, potentially aiding federal agencies in observing disparities in waste management infrastructures (*See Figure 1.2*). Though this feature may be provided by the U.S. Census

Bureau, the time period to acquire this valuable information is extensive.

The extensiveness introduces challenges since the need to classify disadvantaged census tracts is dire and important. An opportunity to help federal agencies make prompt funding decisions can be found using Google Street View (GSV) images. Just as the NLCD acquires green space from satellite imagery, GSV images give one the ability to potentially view differing environmental features between census tracts. This holds true since road damage, garbage bags, and green space can be recognized and evaluated from these images.

Not only can they be acquired, they can provide more accurate and timely efficient data. The time efficiency in GSV images is valuable. In general, major cities are updated approximately once every year, while less populated areas see updates every three years or so [29]. The increased frequency of acquiring information on census tracts challenges that of the NLCD and the U.S. Census Bureau, as these resources require more time to collect data.

With time efficiency being improved, the level of detail and accuracy of GSV images offer more comprehensive information, enabling a more effective near real-time evaluation of public policies aimed at supporting disadvantaged census tracts. Green spaces offer a use-case scenario, illustrating a practical application of this concept. Satellite imagery gives insight into green spaces, but it falls short when recognizing bushes, shrubs, and other small ground covers (*See Figure 1.3*) which contribute significantly as one expands the evaluation of specific census tracts. By offering the pedestrian perspective of green spaces, GSV images introduce a vertical dimension of vegetation, unlike satellite imagery which provides a top-down view. GSV images account for the height, density, and immediate visibility of green space, including any physical blockages [30].

While both GSV and satellite images offer valuable information about green spaces, understanding the perspective of residents in census tracts is extremely important. Both sources

of information can complement each other; GSV images capture seasonal variations and detailed compositions of green space, making them a valuable complement to the broader, landscape-level data provided by satellite images. Also, the granularity GSV images offer for green spaces gives insight into enhancement initiatives. The limiting factors experienced by CEJST and the benefits GSV images offer present an initiative to utilize GSV images supplementing CEJST thereby helping federal agencies make accurate and timely efficient decisions regarding funding allocations.

Figure 1.3: Both images displayed are from Maricopa County in Phoenix, AZ. The satellite image on the left was given by the Enhanced Visualization Analysis (EVA) tool provided by the National Land Cover Database (NLCD). The right image is provided by Google Street View. Both images give insight into their own level of detail of vegetation.

Applying computer vision techniques to GSV images such as binary image classifiers, texture segmentation, and object detection have been used to study garbage bags, road damage, and greenery in images. Binary image classifiers like trained convolutional neural networks work by gradually adjusting a series of filters that can correctly classify an object of interest. Texture segmentation identifies regions in images where textures are unique and grouped together. Object detection can involve the ability to identify whether a number of objects

of interest exist in an image - a counting mechanism can help in use cases.

In extracting features from disadvantaged and non-disadvantaged communities through a series of GSV images, these computer vision techniques were utilized to focus on specific aspects. Texture segmentation measures the percentage of greenery by focusing on its unique texture. Object detection is used to measure the quantity of different degrees of road damage. Lastly, a binary image classifier determines if garbage exists in an environment. A comparative analysis using these three aspects of the environment was conducted to observe potential differences between disadvantaged and non-disadvantaged communities. The study explores how additional computer vision techniques can supplement the CEJST tool for the benefit of federal agencies and census tracts.

1.2 Main Contributions

In an aim to supplement the Climate and Economic Justice Screening Tool by providing environmental features extracted from Google Street View images, the main contributions of this work are as follows:

1. Provide access to more detailed and up-to-date quantitative data, which has the potential to support CEJST and decision-making processes regarding public funding.
2. Introduce five new environmental features not previously accounted for in CEJST.
3. Innovate a method to accurately extract greenery from street-level images by utilizing a filter bank, binary classifier, and Connected Component Analysis.

Chapter 2

Road Damage Detection

2.1 Introduction

A counting mechanism for differing degrees of road damage present in Google Street View (GSV) images was needed to evaluate census tracts. Road damages like longitudinal, transverse, and alligator cracks hold similar features making them difficult to distinguish. There are a number of factors that contribute to the complexity of a crack which include geometrical shape, background noises, differing luminations, and other irregularities. The geometrical shape of a crack can change with numbers, length, width, and size. Background noises such as varying textures of roads can interfere with a model in distinguishing deformities. Different illuminations change a model's ability to observe cracks [14]. With these challenges, an adaptive image recognition method was needed.

In the Crowd Sensing-based Road Damage Detection Challenge (CRDDC) of 2022, object detection was favored for handling the challenges associated with road damage detection [13]. A dataset consisting of 53,429 cracks present in images taken across six different countries was offered to the contestants. An object detection model known as YOLOv7 was used within the challenge. YOLOv7 serves as a useful counting mechanism for road damages present in GSV images. With this approach, one can obtain the total number of road damages from each category for a given census tract.

This model was chosen due to its capability to adapt to the challenges in identifying road damage, specifically because of its high speed and accuracy in detecting multiple objects. YOLOv7's architecture allows for efficient processing of the higher-resolution images present in the dataset, making it ideal for analyzing what the CRDDC provides. Additionally, YOLOv7 incorporates advanced techniques such as multi-scale predictions and a deeper network architecture, which enhance its ability to detect small and large cracks accurately [15].

The model and data set were used in the *Road Damages Detection and Classification with YOLOv7* paper[16]. The work introduces the perspective of using GSV images of roads from the United States for validation. This approach is expanded upon and further investigated by tuning particular parameters introduced by the team. Here we strive to achieve a similar performance which was an average of 81.7% for F1 scores of GSV images in the United States.

2.2 Data Description and Methodology

The data set the participants of the challenge used consisted of satellite images, GSV images, and images captured on portable devices amongst others on which roads are present. Transverse, alligator, longitudinal, and potholes on roads were boxed in each image. The pictures come from different countries which include the United States, India, Czech Republic, Japan, China, and Norway. The total number of cracks present in images equated to 53,429. From the dataset, there is a significantly large magnitude in the number of longitudinal cracks present – the initial finding gave insight into the imbalance present in the data set.

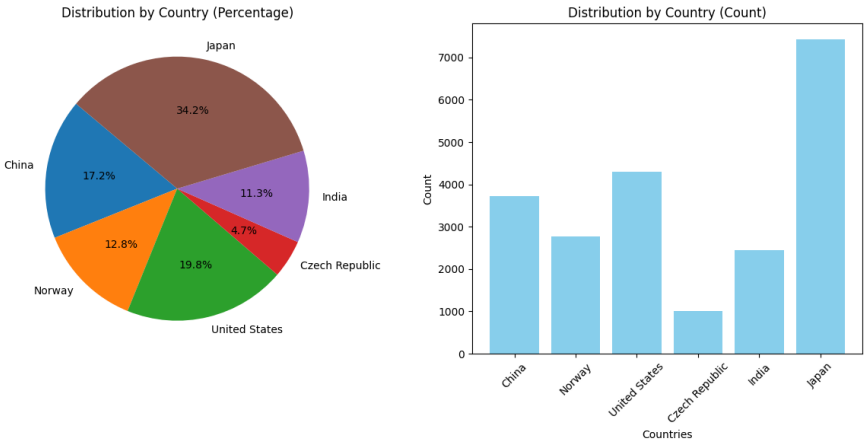

Figure 2.1: The diagrams show the distribution of the data used for training to recognize road damages. The training set consisted of five countries: Japan, China, Norway, the Czech Republic, India, and the United States. The pie chart on the left displays the distribution in terms of percentages while the bar graph on the right gives us the count per country.

21,690 images were used for training. Of the total training count, 34.2% of the images came from Japan (*See Figure 2.1*). The imbalance of the images caused concern because the primary analysis is solely focused on GSV images within the United States. However, the image imbalance introduced more variability in terms of how damages may look across different roads. This led to a higher probability of achieving generalizability since different road damages from all countries were considered.

The number of road damages presented a unique challenge that led to different approaches when experimenting. For longitudinal cracks, there were a total of 24,943 labels on images resulting in a high bias for these damages. Only 11 labels of potholes were made available for testing; after experimentations, a removal of the category was best since other objects like those holding similar characteristics were misclassified as potholes. Moreover, a total of 53,429 ground truth labels were used for the training set (*See Figure 2.2*). With the removal of potholes, the total was reduced to 46,962.

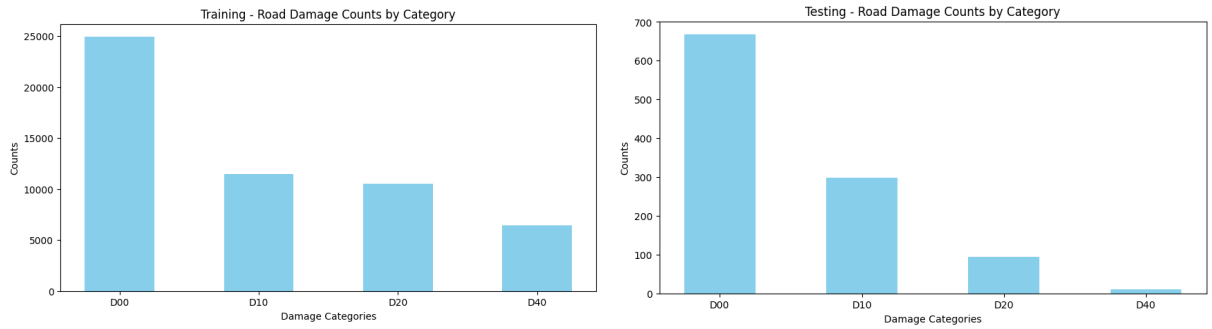

Figure 2.2: The bar graphs represent the total count of road damages for training and testing. The labels D00, D10, D20, and D40 are respectively longitudinal cracks, transverse cracks, alligator cracks, and potholes.

For validation, the total number of ground truth labels was 1,072 and 1,061 without potholes. The validation set consisted of GSV images solely from the United States; though, with this approach only 478 GSV images were available. The initial analysis of the dataset led to two different approaches to finding the most optimal model for analyzing census tracts. The first of which considers all classes, and the second removes a class and introduces tuned parameters that exist within YOLOv7.

To evaluate the different approaches the F1-score, precision, recall, mean average precision (mAP), and confusion matrix were taken into account. Understanding the F1-Score through the series of trials gave the understanding of the balance between precision and recall, which was then adjusted by different confidence thresholds during the model's evaluation. Observing the precision and recall were useful since they were not affected by the skewed distribution present in the data set. They also contribute to the calculation of the mAP for all classes. Finally, the confusion matrix provided gave a straightforward view of the model's predictions in comparison to the actual true values.

2.3 Experimental Results

Two different experiments were conducted to find the best approach when evaluating roads in GSV images. The first approach serves as an initial analysis of the algorithm used. This analysis gave a brief overview of how the algorithm performs with the given data set. The second approach incorporates the findings of the first by removing a class category which previously hurt the overall performance of the method. Additionally, the second approach tunes specific parameters present in the algorithm to achieve the best performance given the available data. In both approaches, we train for optimal performance until both methods converge.

2.3.1 Initial Exploration

In this experiment, we run the algorithm for 60 epochs. The data set for this approach contains a training set with 21,690 images along with the original 24,943 ground truth labels. The validation set had a total of 487 images tagged with 1,071 ground truth labels. No parameters were further tuned for this experiment.

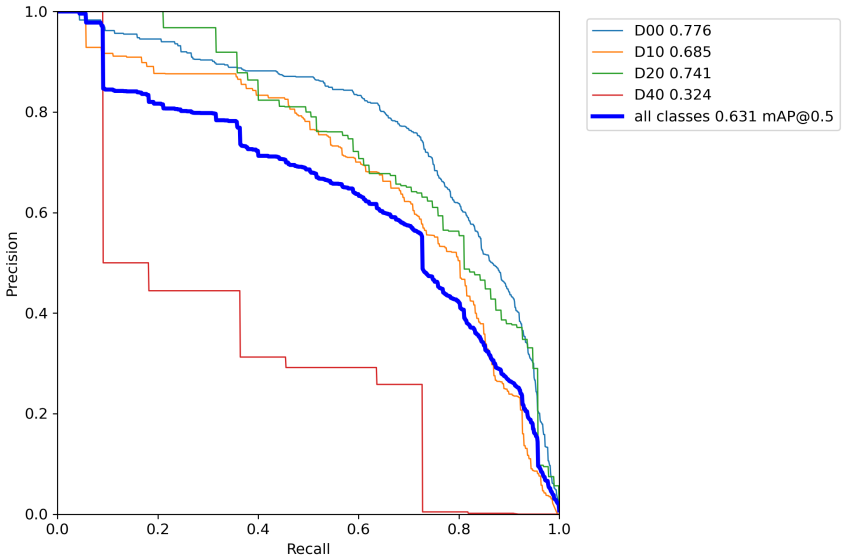

Figure 2.3: The precision and recall during the model's evaluation for 60 epochs are displayed. The labels D00, D10, D20, and D40 are respectively longitudinal cracks, transverse cracks, alligator cracks, and potholes. The curve highlighted in blue marks the overall mAP@0.5 achieved for all classes. Each individual mAP@0.5 for each class is also shown to the right of the plot.

One of the initial remarks made when evaluating the initial approach was the lack of performance in recognizing potholes - labeled as D40. We can see that as the confidence threshold is re-calibrated throughout the model's evaluation, D40 lags behind the other three categories. The other categories display a far better balance between precision and recall as they are much closer to the top right of the graph (*See Figure 2.3*). mAP at the set standard threshold of 0.5 is also noted for evaluation.

mAP@0.5 for longitudinal, transverse, and alligator cracks were respectively 0.77, 0.68, and 0.74. The performance achieved was accepted, but potholes again seem to be the outlier of the group with a mAP@0.5 of 0.32. The overall mAP@0.5 of all classes was 0.631. Observing these results made it obvious that the lack of data for potholes hurt the performance of the method – leading to acknowledgment that potholes were difficult to recognize in GSV images.

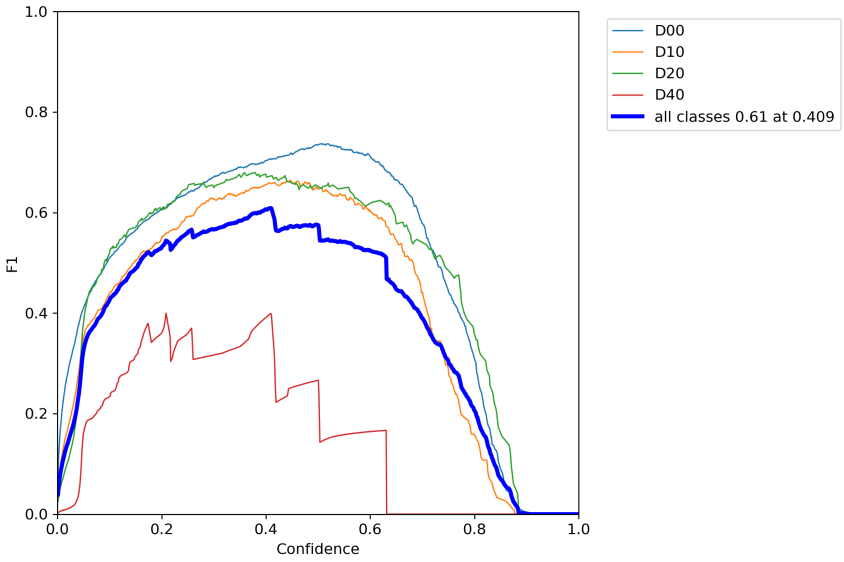

Figure 2.4: F1-Score at different confidence thresholds for all classes are shown - the number of epochs for this experiment was 60. The labels displayed to the right of the graph D00, D10, D20, and D40 are respectively longitudinal cracks, transverse cracks, alligator cracks, and potholes.

As the confidence threshold is tuned the F1-Score is being taken into account. Again, the algorithm seems to struggle to recognize potholes in GSV images. The consistency shown in the other three classes is not present when observing potholes. This is noticeable since at a confidence score of 0.4 the F1-Score of potholes lags vigorously compared to its peers. The evaluation displayed on the right entails that the best F1-Score achieved while taking into account all classes is held at 0.641 with a confidence score of 0.409 (*See Figure 2.4*).

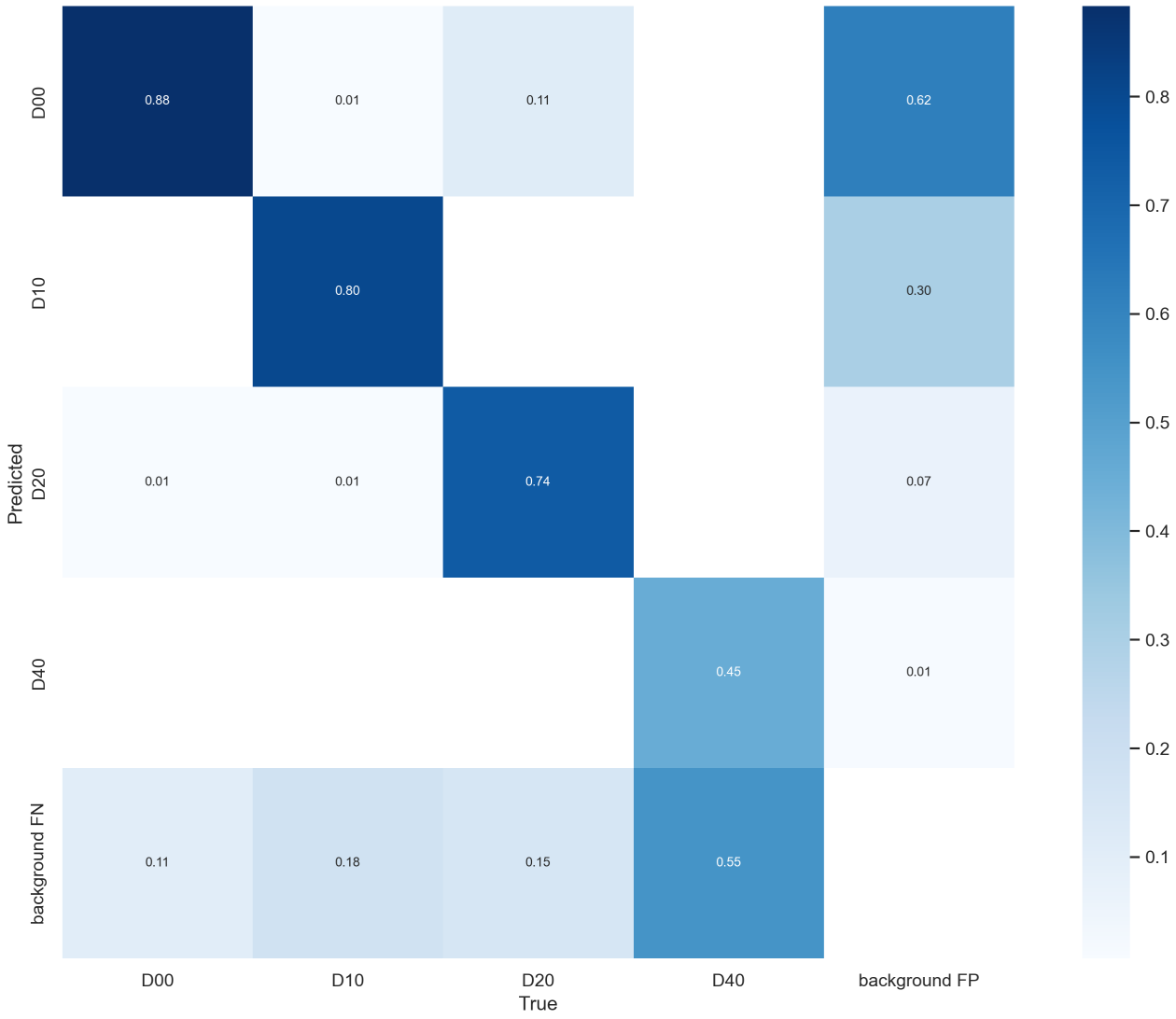

Figure 2.5: The displayed is a confusion matrix which was given after running for 60 epochs. The left labels show the predictions made for D00, D10, D20, and D40 which are respectively longitudinal cracks, transverse cracks, alligator cracks, and potholes. Also, "background FN" gives insight into the false negatives for each class. The X-axis follows the same format this time indicating true labels for each subgroup. Next to the plot we have a gradient measure giving more clarity to the values present.

The values displayed on the diagonal (0.88 for longitudinal cracks, 0.8 for transverse cracks, 0.74 for alligator cracks, and 0.45 for potholes) show the proportions of labels that were correctly classified by the algorithm. We see that good performance is given by the first three road damages. For example, 88% of longitudinal crack instances were correctly classified by YOLOv7. However, 11% of alligator cracks were predicted as longitudinal cracks – shown in the third column on the first row. The fifth row, Background FN, shows false negatives for each class where 0.55 displayed on the fourth column indicates many actual pothole instances were missed by the model (*See Figure 2.5*). In conclusion, the confusion matrix gave further insight into the complexity of classifying potholes and the performance gap that existed.

Figure 2.6: The images showcase two Google Street View (GSV) images. Each image was evaluated by a trained YOLOv7 model. The predictions are tagged by a category followed by a confidence score. The labels displayed D00, D10, D20, and D40 are respectively longitudinal cracks, transverse cracks, alligator cracks, and potholes.

The images show two examples of the approach to making a prediction (*See Figure 2.6*). On the left, we see a pothole (D40), and a longitudinal crack (D00) being made. Their respective confidence scores are 30% and 80%. The classified pothole seems too reflective. The reflection

seen on the supposed pothole suggests that the predicted object is a misclassification. In the right image, we have longitudinal cracks, transverse cracks (D10), and alligator cracks (D20). Each prediction made in the right image has low confidence scores of less than 50% except for the alligator crack which is 80%.

2.3.2 Advanced Tuning

In this experiment, we adjust the original data set by removing potholes. The data set now has 21,690 images with 46,962 ground truth labels for training. 487 images with 1,061 labels were used to validate the algorithm. Two parameters were tuned for this experiment. Paste-in, which randomly pastes objects from one image to another, was reduced to a probability of 0. Rotation rotates an image in the training set and is tuned to 10 degrees. Lastly, we train the model for 90 epochs.

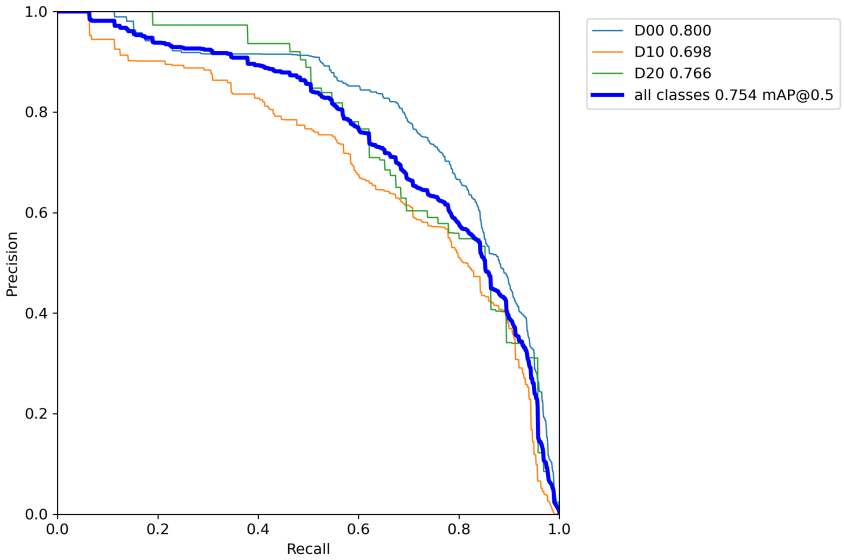

Figure 2.7: The precision and recall during the model's evaluation for 90 epochs are displayed. The labels D00, D10, D20, and D40 are respectively longitudinal cracks, transverse cracks, alligator cracks, and potholes. The curve highlighted in blue marks the overall mAP@0.5 achieved for all classes. Each individual mAP@0.5 for each class is also shown to the right of the plot.

A noticeable note made was that the overall performance of the classifications was improved (*See Figure 2.7*). The mAP@0.5 of 0.754 suggests good performance across all classes. The best performances coming from longitudinal cracks (D00) of 0.8 were expected since there was more of a variety of labels for these damages. One of the more challenging damages to recognize in this experiment was transverse cracks (D10). The drop-off in precision as there is an increase in recall suggests room for improvement, but a mAP@0.5 of 0.698 denotes a modest achievement.

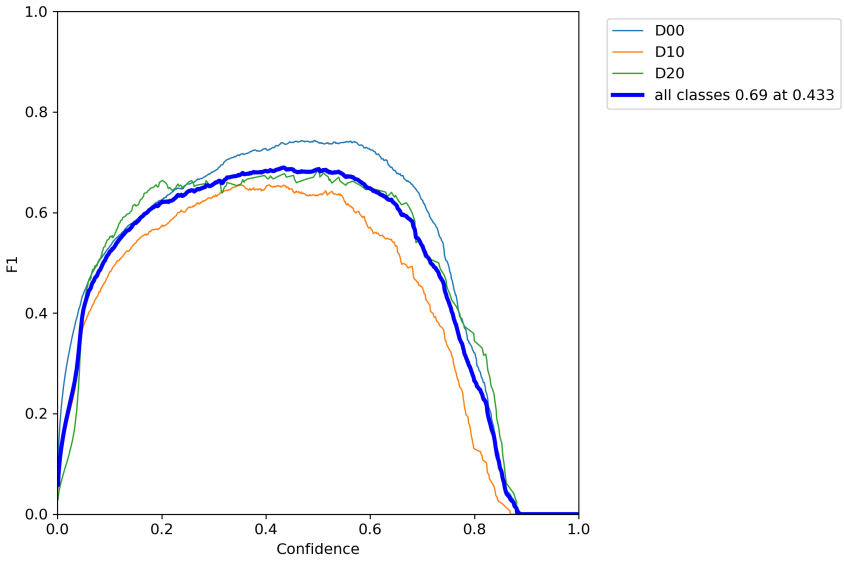

Figure 2.8: F1-Score at different confidence thresholds for all classes are shown - the number of epochs for this experiment was 90. The labels displayed to the right of the graph D00, D10, D20, and D40 are respectively longitudinal cracks, transverse cracks, alligator cracks, and potholes.

Each road damage type seems to have differing peaks of F1 scores with their corresponding confidence scores. Evaluating all classes (blue curve), the F1-Score achieved for all classes is 0.69 at a confidence threshold of 0.433 suggesting that, when the threshold is set to 0.433, the balance between precision and recall is at its best (*See Figure 2.8*). The performance achieved shows an overall good balance of precision and recall across the different road damage types.

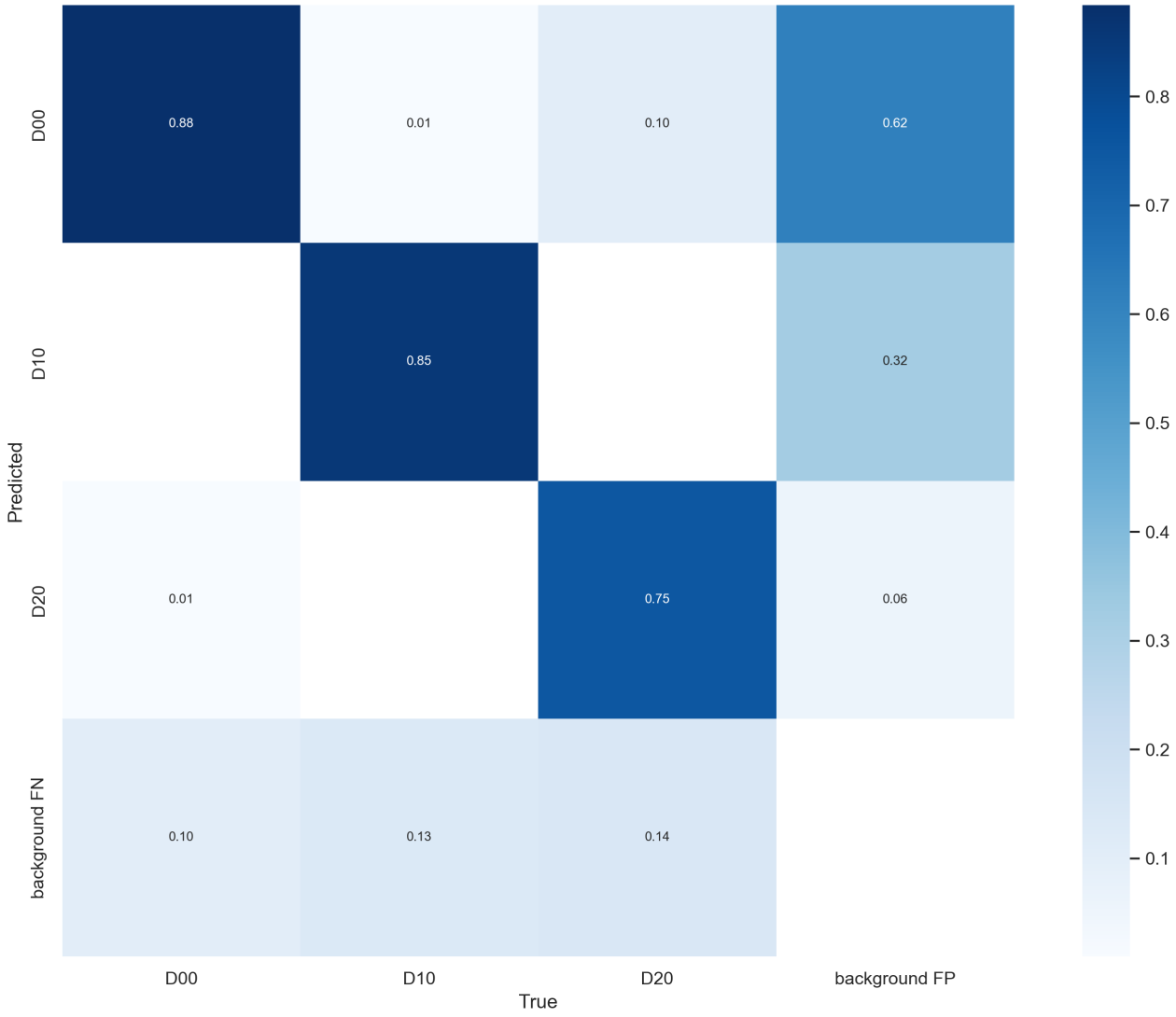

Figure 2.9: The displayed is a confusion matrix which was given after running for 90 epochs. The left labels show the predictions made for D00, D10, and D20 which are respectively longitudinal cracks, transverse cracks, and alligator cracks. The X-axis follows the same format – indicating true labels for each subgroup. Next to the plot we have a gradient measure giving clarity to the values present.

A relatively high percentage of 62% was associated with classifying the background of a GSV image as D00 (*See Figure 2.9*). The model also misclassified the background as transverse (D10) and alligator (D20) cracks with errors of 32% and 6% respectively. This gave a perspective that the trained model may be overly sensitive to the texture or characteristics of backgrounds present in GSV images. The overall performance was accepted since correct predictions for each class were at or above 75%.

Figure 2.10: The images showcase two Google Street View (GSV) images. Each image was evaluated by a trained YOLOv7 model. The predictions are tagged by a category. The labels displayed D00, D10, and D20 are respectively longitudinal, transverse, and alligator cracks.

The two images in *Figure 2.10* have been processed by the trained model. In the left image, we see a transverse crack (D10) and a longitudinal crack (D00) being recognized. Surprisingly no detection was made on the right side of the road near the fences where some road damage appears to be present. In the right image, we have a longitudinal and alligator (D20) crack being identified. These damages are relatively large compared to the overall size of the Google Street View (GSV) image.

2.4 Discussion

Achieving a mAP@0.5 of 0.754 for all classes was acceptable for evaluating road damages across different census tracts. The caveat to the final approach was that many false classifications were being made on the background of GSV images. Conducting an investigation into the similar features shared with backgrounds and roads would be fruitful in fine-tuning the current approach. Some initial expectations may be that the misclassifications come from differing textures seen on roads. For example, roads that hold rocks or other types of textured surfaces may lead to incorrect classifications. Introducing more of a variety of roads in the data set can help mitigate the false predictions. Also, minimizing misidentifications of potholes was a fruitful decision. Over-identification, or false positives, in a series of images, should be at best avoided since the misclassification of a census tract as disadvantaged carries significant implications and requires careful consideration.

Chapter 3

Garbage Bag Classifier

3.1 Introduction

Identifying garbage bags which can vary in color, size, and geometric shape can be challenging. An approach to combat the unique shapes of garbage bags is to implement a Convolutional Neural Network (CNN) where the model learns how to manipulate filters to identify whether a garbage bag is present within an image i.e. a binary CNN. Some commonalities garbage bags share are the unique edges they form as they sit on the sides of city streets. Another defining characteristic is their shading as a bag distorts from different settings. One will need to take into account the Google Street View (GSV) vehicle as it travels from street to street. Inclines are present when the vehicle travels due to hills. Making the garbage bag positioned in different quadrants of a given GSV image. Images are distorted due to image stitching that Google implements to take a 360° view of their environments; thus, creating an even more unique shape of garbage bags. Lighting due to differing weathering conditions creates a more complex scene to distinguish garbage bags from other similarly shaped objects in an image.

A CNN proved to be the best tool in identifying garbage bags due to their adaptability in identifying unique features shared across garbage bags. This method has the ability to learn shared features among objects of interest through adjusting its weights, and filters, learned from a batch of images. Input images run through a series of convolutional and pooling

layers thereby constructing a hierarchical representation of features. These features are then used to make a prediction [18].

Previous work introduced the use of an object detector model known as YOLOv3. The dataset used to train the model consisted of 1027 boxed garbage bags; among other highlighted objects such as cardboard and garbage containers. The mean average percussion (mAP) for garbage bags was 0.875 [17]. The performance achieved by the group was notable, but the question of whether the process could be simplified to a simple CNN network was explored. In this chapter, we explored the architecture of a binary CNN to classify the existence of garbage bags present within GSV images.

3.2 Data Collection and Model Architecture

Figure 3.1: The Google Street View (GSV) image is from Bronx County, New York. In this image, one sees two piles of garbage bags on the sides of one city street. Both piles have their own unique shapes.

Data sets that contain single and piles of garbage bags on the sides of streets in urban cities within the United States were rare (*See Figure 3.1*); therefore, data augmentations had to be implemented in order to expand the training set. GSV images containing garbage bags were collected from Bronx County, NY, Skid Row, LA, and the city of San Francisco. A total of 270 garbage bags were found and expanded upon. The augmentations implemented were combinations of Gaussian blur, horizontal flip, rotation, and distortions.

Figure 3.2: Image augmentations were applied to a Google Street View (GSV) image containing two piles of garbage bags from Bronx County, New York. From left to right, the first image augmentation applied was a rotation of 15 degrees. The second image is the result of applying Gaussian blur. The next image is a horizontal flip. Finally, the last image contains a distortion.

Gaussian blur was used to accommodate differing qualities of the downloaded images which stemmed from bad weather or a low-quality camera. Garbage bags can be found on either side of a city street, horizontal flips help identify garbage bags on both sides of a given street and introduce differing shapes. Rotations of approximately 15° were used to introduce inclines specific GSV images contain. Google offers a 360° perspective of a particular point which then introduces distortions due to image stitching; adding distortions to the training set helped with this issue (*See Figure 3.2*). The augmentations changed the distinct features of the garbage bags which were their shape, edges, and size giving the model more of a variety of how garbage bags look in an image. With the augmentations, our training set amounted to a total of 5,400 images. 2,700 images obtaining garbage within a scene and the rest where we see no garbage in city streets. Implementing data augmentations also helped

limit over-fitting.

Even with the augmentations in place, the performance of the model was not improving. Different parameters were tuned which included more filters, and linear layers, working with different optimizers such as Stochastic Gradient Descent (SGD) and Adam, but not many proved to aid the classifier in any way like the learning rate experimented with in this study. The learning rate was initially static at 1×10^{-3} and until a cyclical learning rate was introduced. The performance of the model achieved a new level of accuracy that was then used to analyze images for given census tracts.

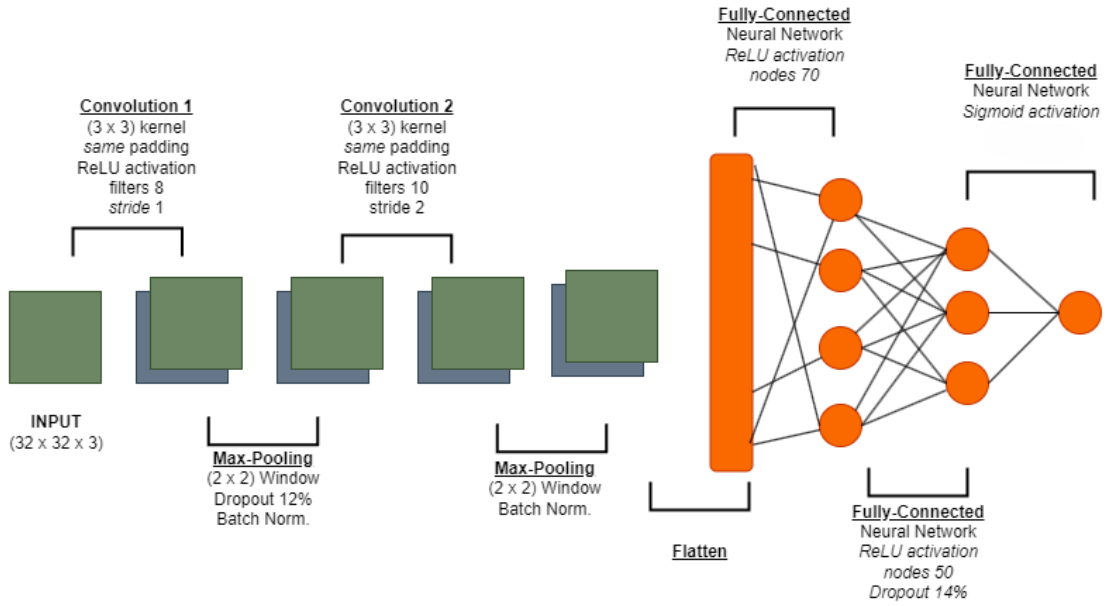

Figure 3.3: A diagram of the binary convolutional neural network architecture used to identify garbage bags in city streets.

The final bone structure of the model begins with the input of 32 by 32 pixel images with its RGB channels. Reducing the image dimensions helped with the limiting computational resources available at the time. This input layer contains eight filters of dimensions (3×3), stride of one, the activation function used is known as ReLU. The stride chosen was to extract every possible feature found in the input images; thus, evaluating every single pixel.

After the input layer, to conserve the dimensions of the feature maps "same" padding was used which adds borders of zeros to the input image. The layer after comes with a (2×2) max pooling operation. Due to over-fitting, a drop out of 12% was implemented; reducing the complexity of the network. Batch normalization was also added to help with the model's overfitting issue. The next convolution layer holds 10 filters of size (3×3), strides of (2×2), same padding, and the activation function ReLU. The strides were increased to accommodate for flattening. The filter dimensions needed to be reduced to limit the number of weights being held matching hardware availability. The following two layers contain 70 and 50 nodes respectively. The numbers were found after trial and error using a varying number of nodes. A 14% dropout was implemented afterwards. Lastly, the classification segment where we have one node and the activation function known as Sigmoid.

To compute the adaptive learning rates for any given weight, the optimizer Adam was utilized. After implementing a more simplistic network, which helped with over-fitting, the issue still affected the accuracy with a 15% difference in validation and training. A cyclical learning rate with min 4.43×10^{-5} and a max of 1.83×10^{-4} was applied. This reduced the gap between the validation and training accuracy to approximately 3%.

3.3 Experimental Results

3.3.1 Cyclical Learning Rate

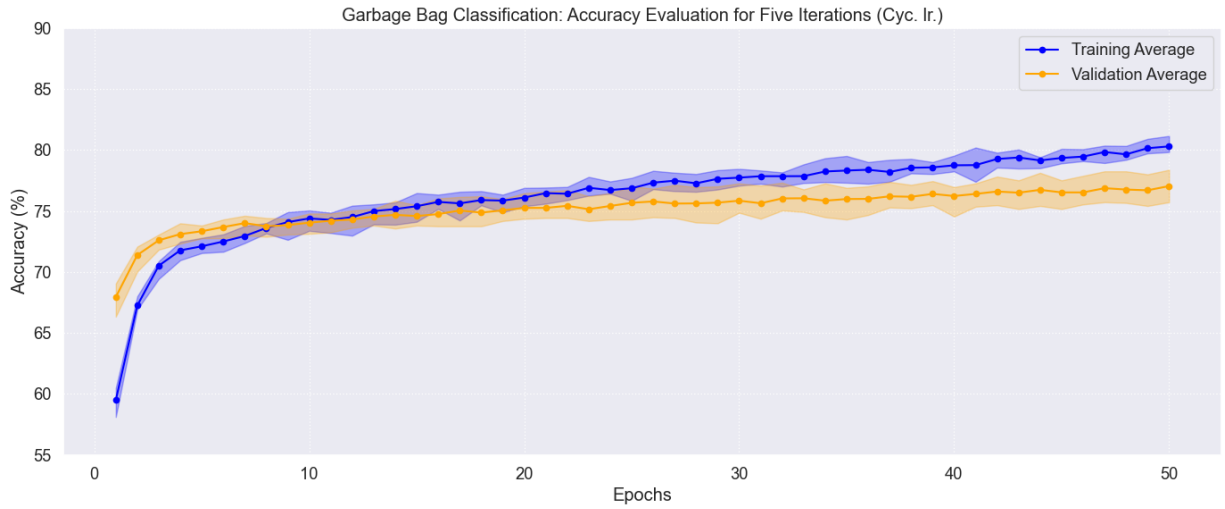

Figure 3.4: The average performance of the binary convolutional neural network for five different tests; using a cyclical learning rate. Blue highlights training accuracy while light orange is the validation accuracy.

Applying the cyclical learning rate with a min of 4.43×10^{-5} and a max of 1.83×10^{-4} helped with accuracy. The boundaries provided an estimate of where the optimum learning rate exists. Throughout training the near optimum learning rate was being used resulting in the best performance given the current dataset and architecture of the neural network [19]. This prevented extreme over-fitting and reduced the gap in the validation and training performance to approximately 3.25%. On average the training and validation accuracy were respectively 80.30% and 77.05% (*See Figure 3.4*). Additionally, training the model for more than 50 epochs resulted in more over-fitting.

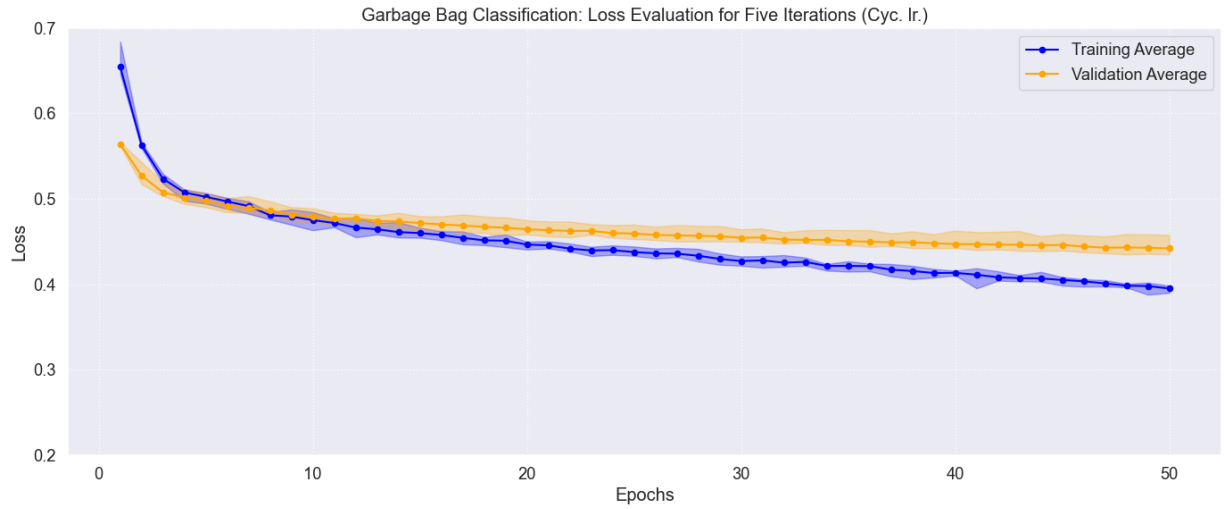

Figure 3.5: The average loss performance of the binary convolutional neural network for five different tests; using a cyclical learning rate. Blue highlights training loss. Light orange is the validation loss.

Over-fitting is still present when applying the cyclical learning rate. The loss over epochs was still relatively high after training the model. On average the validation and training loss were respectively 0.44 and 0.39. This suggests that the model is still experiencing poor performance after observing the high loss (*See Figure 3.5*). The over-fitting could be caused by the complexity of the model, learning rate issues, or noise present in the training data. Tuning the learning rates further led to more subpar performance - the changes led to more serious over-fitting issues. The complexity of the model was tuned during the experimentation phase. Adding regularizers or probing the dropout led to further sub-optimal performances compared to the one shown.

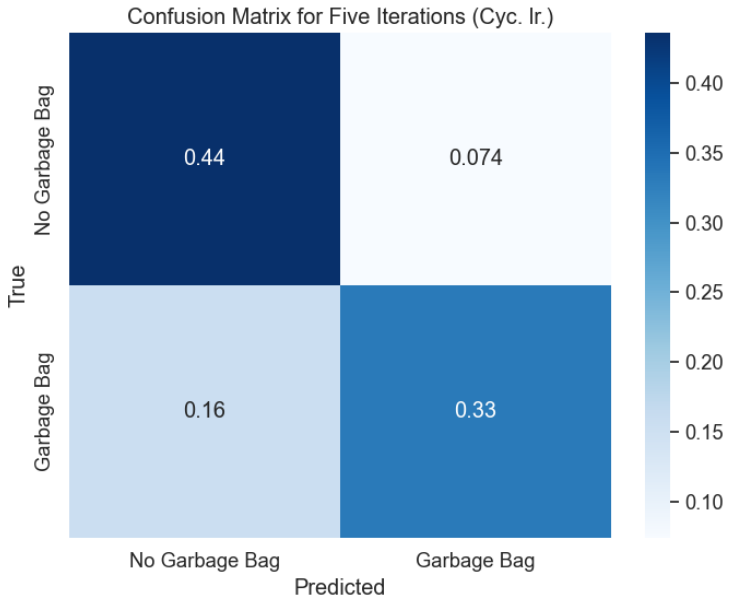

Figure 3.6: The performance of the binary convolutional neural network using a cyclical learning rate. The matrix is the aggregation of each confusion matrix for five different test iterations. Values are normalized and displayed. Intensity level is given by the blue color spectrum.

The confusion matrix after using a cyclical learning rate experiment tells us that e.g. 44% of the test set was correctly predicted to have no garbage bag present. Garbage bags were recognized by the model since 33% had a garbage bag which held to be true (*See Figure 3.6*). Overall, the model was still challenged when classifying whether a garbage bag was present within an image. The model favored images that held no garbage bag shown by the false positives.

3.3.2 Static Learning Rate

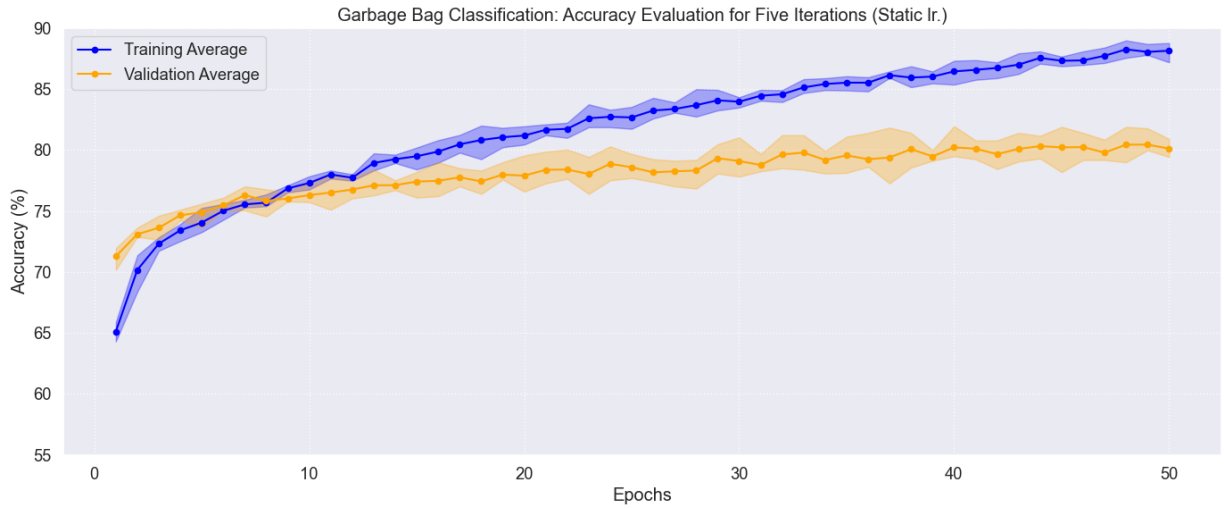

Figure 3.7: The average performance of the binary convolutional neural network for five different tests; using a static learning rate. Blue highlights training accuracy while light orange is the validation accuracy.

Using the same model architecture, a static learning rate of 1.85×10^{-4} was chosen which then led to extreme over-fitting. After training for 50 epochs, the validation and training accuracy was 80.12% and 88.15% (*See Figure 3.7*). The static learning rate was too high and low for different stages of training. Having a fixed learning rate hinders the model's ability to make meaningful adjustments to its parameters [19]. The poor adjustments challenged the model's ability to adapt to the intricacies of classifying garbage bags.

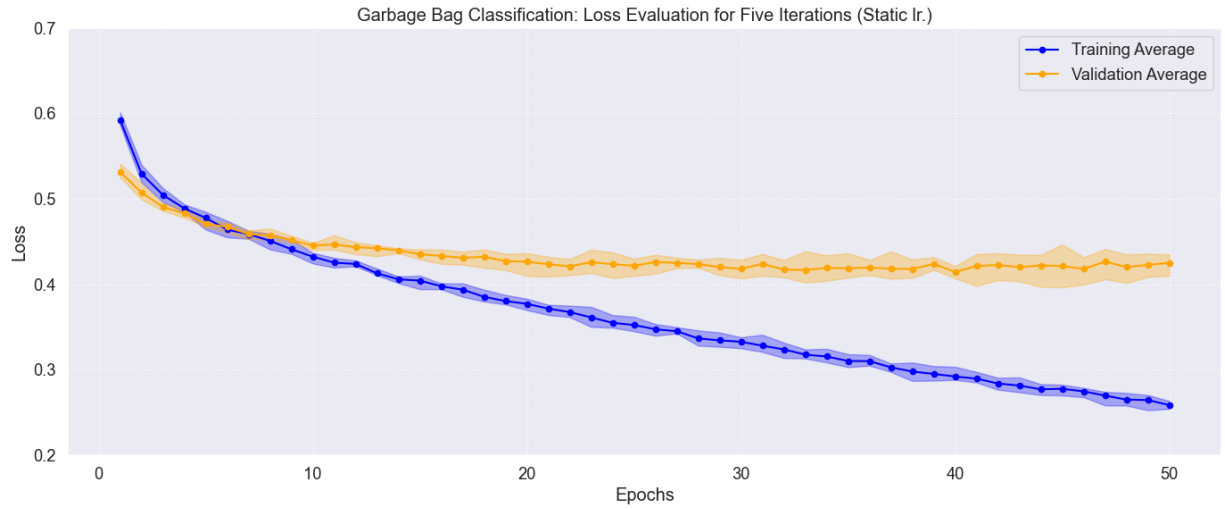

Figure 3.8: The average loss performance of the binary convolutional neural network for five different tests; using a static learning rate. Blue highlights training loss. Light orange is the validation loss.

Over-fitting is clear when observing the loss over epochs for the static learning rate experiment. After 50 epochs, the average validation loss was 0.42, and the average training loss was 0.26 (*See Figure 3.8*). The significant difference in the validation and training loss tells us that the model was struggling to adjust its weights. Decreasing the learning rate even further was the next step when the model was over-fitting; though, the decrease did not resolve the issue. Adaptive learning rates proved to be more useful.

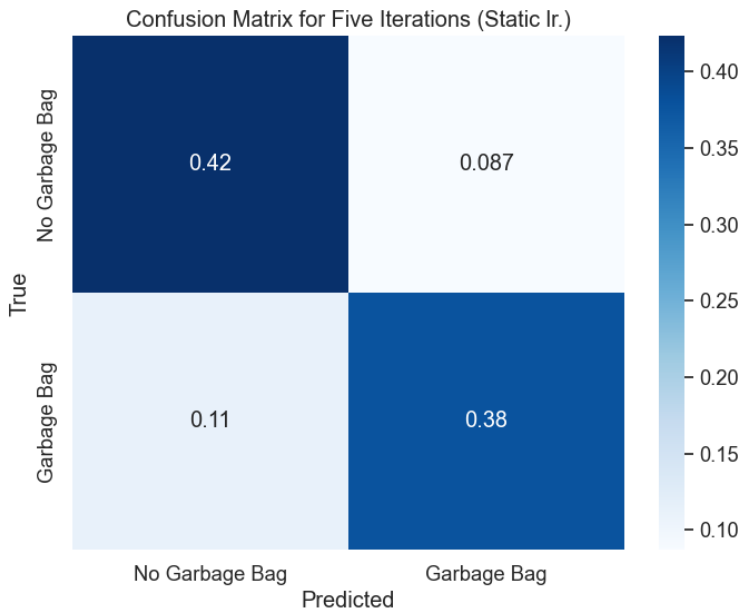

Figure 3.9: The performance of the binary convolutional neural network using a static learning rate. The matrix is the aggregation of each confusion matrix for five different test iterations. Values are normalized and displayed. Intensity level is given by the blue color spectrum.

The confusion matrix seems to be fairly balanced in predicting the presence and absence of garbage bags (*See Figure 3.9*). However, this provides an example of how crucial it is to examine other performance metrics when evaluating a binary CNN. Initially, it may be implying satisfactory performance but a closer analysis reveals its flaws. As previously described, the 8% difference in validation and training accuracy of five iterations reveals that the approach led to over-fitting. The gap highlights that arriving at conclusions about the performance with the confusion matrix alone is not sufficient.

3.4 Discussion

In this chapter, we highlighted an approach for identifying garbage bags on the sides of city streets. A unique binary convolutional neural network was experimented with. The

model's performance was stagnant and insufficient when using a fixed learning rate. After applying the cyclical learning rate, the performance achieved by the garbage bag classifier was approximately 77% in accuracy when testing. Using the same model and applying a static learning rate the accuracy for testing was 80%, but the performance showed that the model was fitted more towards the training set making it not well adaptive to new data. Using the cyclical learning rate to count the number of garbage bags in census tracts proved to be the most effective due to its adaptability. This approach showed that the method did not lead to extreme over-fitting. The generalizability of the method is crucial particularly when analyzing census tracts from different cities. An adaptive model is better equipped to handle the subtle changes in census tracts as one analyzes different urban environments.

Though the developed method proved generalizability to some degree it still lacks in predicting the correct feature due to the premeditated adjustments. Firstly, the dimensionality reduction from the original (640×640) GSV image to (32×32) significantly affected the visibility of garbage bags. Once the reduction was done the model can no longer effectively recognize garbage bags since they only took up a small portion of GSV images. The second adjustment involved distortions to expand the training set. To account for the distortions, the borders of the image were removed and replaced with black pixels. This specific alteration no longer matched a GSV image since no image contains disfigured black borders.

Future improvements could involve methods such as image slicing. Training a model to identify garbage bags of size (32×32) and then analyzing sliced GSV images of the same dimensionality can prove beneficial. The training set for the proposed approach will need to contain sliced GSV images of sidewalks, roads, buildings, skies, and more to account for the wide variability the method will introduce. In summary, improving the accuracy and relevance of the model's predictions will need to be prioritized for the next phase.

Chapter 4

Greenery Detection

4.1 Introduction

Detecting greenery, such as bushes, grass, leaves, and shrubs in images poses challenges. For one, analyzing the sole color green is not sufficient. Painted murals on buildings, colored cars, and other highlighted pieces within a Google Street View (GSV) image share this feature. Finding the differentiating feature all vegetation share throughout the seasons of the year was found to be texture. Identifying the unique texture greenery shares is achievable.

Though with texture other obstacles emerge, not all vegetation share the same texture. For example, freshly caught lawns and forest areas carry their own unique textures. Cut lawns are more smooth while trees, bushes, and shrubs have more creased edges and lines. Establishing a textured filter bank helped tackle these hurdles. The filter bank can contain varying filters meant to handle varying textures. By filtering an image through a textured focus filter bank one has the potential to obtain filters that highlight smooth lawns and detailed vegetation.

To account for varying textures a filter with tunable parameters is needed. The Gabor filter was a primary candidate and was utilized in this study. This filter holds six adjustable parameters within OpenCV. The k-size, or kernel window, parameter is determined by the size of the feature needed to be recognized. σ was used to define the receptive field and

localization of the filter. θ determined the orientation of the kernel window. The wavelength of the sinusoidal component is specified by λ. γ controls the aspect ratio of the elliptical Gaussian envelope. Lastly, ϕ determines the phase relationship between the filter's response and the texture pattern in the image [24].

A series of differing parameters were used for Gabor in the texture filter bank. Many parameters were experimented with through grid search; the experimental runs led to a set that best fits the task. The ultimate decision for finding the best approach given a limited data set depended upon an extension. Once an image was passed through the unique filters, a binary classifier, in this case Random Forest, decided whether a pixel represented a region containing greenery. After the classification, two extensions were explored as an additional post-processing technique.

The two that gave the most insight were the Segment Anything Model (SAM) and Connected Component Analysis (CCA). SAM was powerful in segmenting images into accurate regions. One critical caveat to the tool was that no tunable parameter existed to specify greenery in images. To combat this hurdle we used the developed classifier to help specify regions of interests.

The second extension explored, CCA, helped in diffusing an image. With the classifier and texture filter bank, some non-greenery regions in images were highlighted. CCA, which had 8 connected components, proved useful in eliminating these regions by focusing on others that had more pixels. Given the limited data set, these two approaches provided the most insight.

4.2 Data Description and Methodology

A limited amount of labeled GSV images of greenery were available for training. Also, the amount of computing power necessary was far too grand to consider a data set that had little variability. A total of 42 distinct images of greenery were extracted from GSV and labeled; each image was accompanied by 224 individual Gabor filters. All filtered images deliver a unique perspective to the training process. For example, many images offered variability of the texture of greenery. Others contained textures that challenged the method. Textures that exist in bricks and gravel are similar to some forms of vegetation.

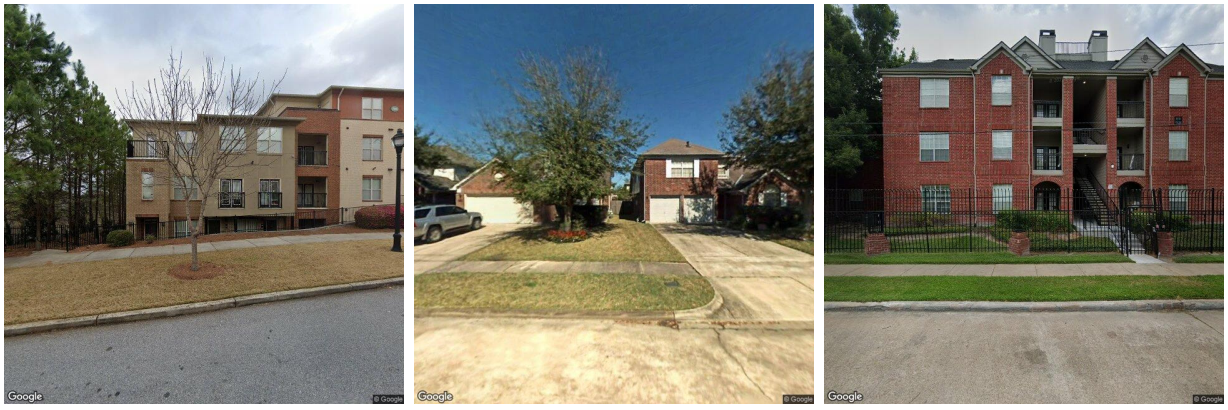

Figure 4.1: Three Google Street View (GSV) images are displayed. All three images were part of the training phase for all approaches to detecting greenery. The display offers insight into features that challenge the methods e.g. textured bricks. They also offer different variations of greenery throughout the seasons of the year.

To examine the outputs of the explored approaches a total of 18 test images were gathered and hand labeled. The 18 images brought similar characteristics that challenge a method which include competing textures (*See Figure 4.1*). Three evaluation metrics were considered for each approach which were F1-score, pixel accuracy, and intersection over union (IoU). The F1 score helped with the imbalance taking place given that a batch of GSV images contained little to no greenery. Pixel accuracy provided a more simplistic and intuitive performance evaluation since its task is to account for correctly classified pixels. IoU gave

the amount of overlap of the prediction and actual regions of interest.

After each evaluation, it was important to note that not all regions were highlighted precisely. A tool utilized to label the images was provided by Roboflow. Roboflow gives users the ability to label images quickly, but in return, some pixels may fall within non-greenery regions [25]. This became apparent after a closer inspection of the labeled images.

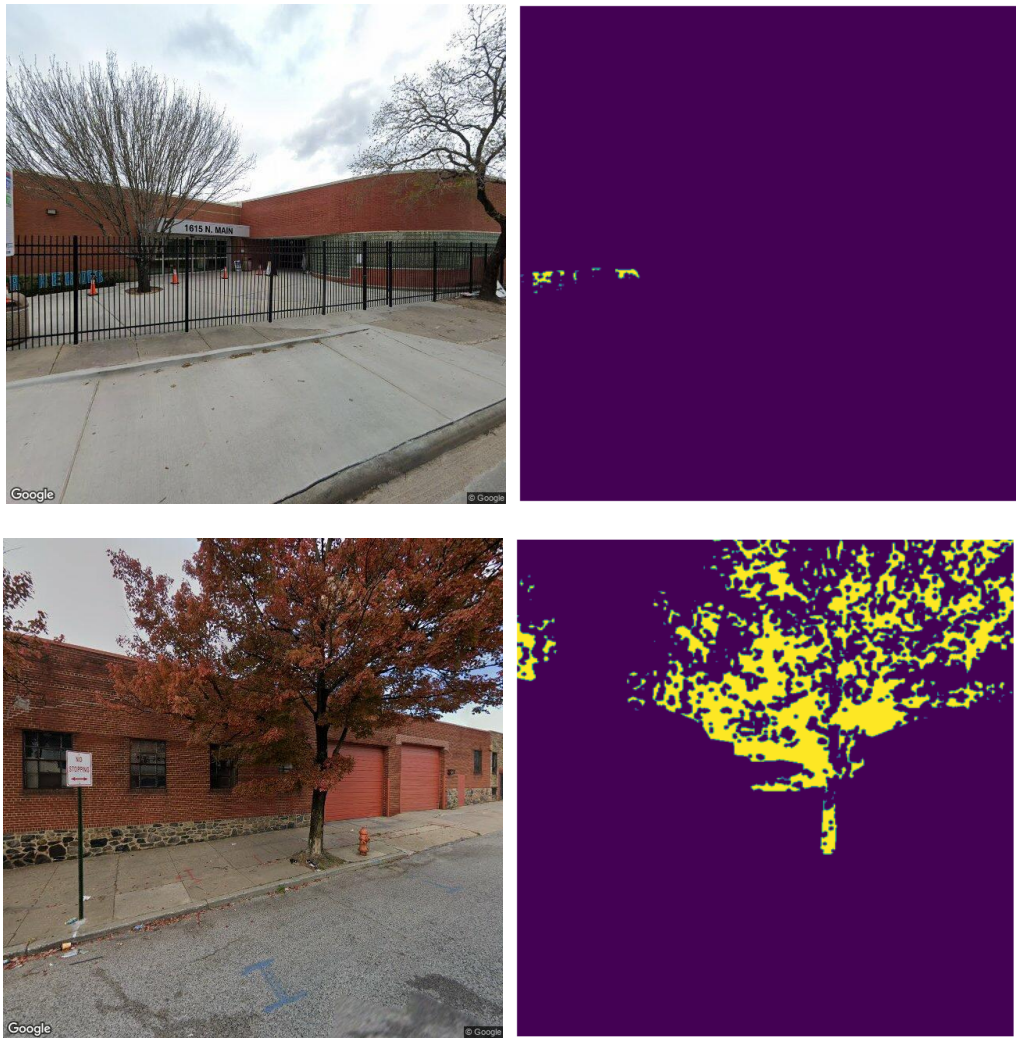

Figure 4.2: Two Google Street View (GSV) images and their masks are displayed. Both were used to evaluate the performances of the greenery detection approaches. We can see that some of the labeled masks are not precisely defining greenery in the image. The top-right image captures only a piece of bush behind the fence. The bottom-right label shows how the bark of a tree is being used to define greenery.

Each approach shares the same initial process in extracting greenery - involving a threshold. We begin by reading the image in HSV color space. Shade (hue), the vibrance of color (saturation), and brightness of color (value) were particularly useful in color segmentation. Color segmentation was important because the next step was to remove colors that did not represent vegetation. We preserve specific values of green, yellow, and brown color ranges. Green as hue 25 to 75, saturation 40 to 255, and value 50 to 255, captured a wide spectrum of color. For yellow, hue 30, saturation 255, and value 255 aimed to detect a specific condition of yellow. Lastly, Brown's range was set from hue 10 to 20 with saturation and value starting at 30, targeting darker shades of brown. This mask acts as a filter for all images including labels (*See Figure 4.2*).

The initial filtered mask is then passed through the Gabor feature extractor. Four of the six tunable parameters were optimized. Orientation is defined by $\frac{\pi}{4}$ and $\frac{3\pi}{4}$. σ was narrowed down to the value of 3. Our wavelength, or frequency, was an array from 0.2 to 0.55, with a step of 0.05. γ was defined by values from 0.2 to 0.85, with a step size of 0.05. Lastly, the kernel window was of size (31×31) pixels and phi held at 0.

Each image was associated with 224 distinct Gabor filters. With the number of GSV images being 42, the data frame extracted from the filter bank was of size (42×226). Two additional columns were included to account for the original image and its label. All filtered images and labels were then collapsed into a one-dimensional array for classification.

A scaling technique known as min-max scaling was applied to the data frame which scaled the features from -1 to 1. After scaling, Random Forest was used to classify the pixel values at 1 or 0 - where 1 represents greenery. For this classifier, we had a total of 100 different decision trees and a random state of 42. Once the data frame was passed through the classifier we were given a predicted mask for greenery in a GSV image.

4.3 Experimental Results

We investigate the best method to post-process the output of the filter bank and the binary classifier. First, we evaluate the performance of the current method. We explore the implementation of SAM as an additional post-processing technique. Also, experiment with the use case of CCA to eliminate unpopular regions in a predicted mask. The goal of these experiments is to understand whether an additional image post-processing technique would bear fruitful to the current approach.

4.3.1 Gabor Feature Extractor and Random Forest

This experiment follows the same methodology in first taking the input image and extracting relevant pixel values via thresholding. Next, we pass the image through the Gabor filter bank. Once the filter bank has created a data frame a scaling technique known as min-max scaling is implemented. The resulting data frame is then given to a classifier, known as Random Forest, for a prediction. The classifier was configured to have 100 decision trees and a random state of 42.

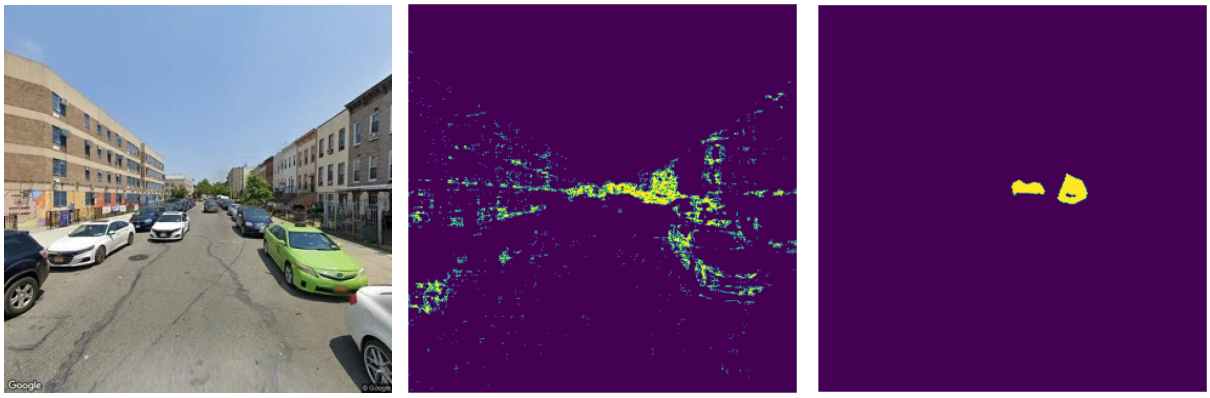

Figure 4.3: A Google Street View (GSV) image is shown with the Gabor Bank and classifier prediction. The final image on the right is the labeled image highlighting greenery within the GSV image.

The initial approach can successfully extract greenery from the GSV image with a few imperfections. First, the approach gets conflicted with highlighting parts of buildings. This makes sense since the method gets challenged when presented with images of bricks or a form of cement. It seems that greenery and brick-textured surfaces seen in images share some similarities. However, this assumption gets tested when we focus on the car. Individual pixels of the side and front of the green car are highlighted. Further investigation into the similarities of the recognized regions is needed (*See Figure 4.3*).

The average pixel accuracy across all 18 different images resulted in 92.19%. Pixel accuracy gives relatively high feedback on performance since a majority of the classification for an image is on non-greenery regions; as an example, only a small portion of this image contains greenery. The average IoU was found to be 0.45. IoU showed that the method gave moderate predictions for its images. Also, the average F1-score for this approach was 0.58; again, revealing a modest performance by the first approach.

4.3.2 Gabor Feature Extractor and Random Forest with SAM

The foundation of SAM by Meta AI begins with a transformer-based architecture which gives it the ability to handle sequences of data. The core of what SAM uses is a technique called Masked Autoencoders (MAE) where parts of a particular image are augmented – making the model learn to predict these missing pieces. Reconstructing the missed segments helps the model understand object shapes and boundaries [26].

During the segmentation process, we see SAM utilize a combination of visual tokens generated from the image. Its positional encodings provide spatial inferences which are then processed through the transformers layers. The final output is a series of predictions of which parts of an image belong to which segments [26]. In this experiment, we utilize the

regions depicted by SAM to highlight green spaces. With the predicted image provided by the filter bank and classifier, we use a constant N, where N represents the number of pixels that exist in a region highlighted by SAM, to determine which regions should be considered as greenery. SAM's role is to help highlight a green space that the initial method is trying to recognize.

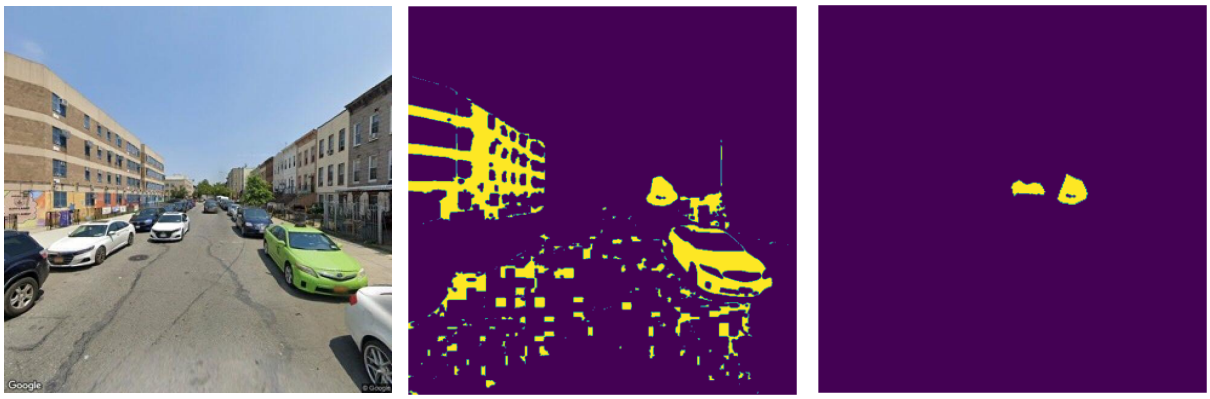

Figure 4.4: The display contains one Google Street View (GSV) image and two segmented images. A segmented image located in the middle is provided by an additional image post-processing technique using SAM. The rightmost image labels the green space that exists in the GSV image.

In this experiment, we set N to 500. A total of 500 pixels of one's needed to fall within a region recognized by SAM in order for the region to be recognized as green space. From the results, we can tell that this approach had faults. Since the Gabor feature extractor and classifier recognized pieces of the car, SAM took note of the entire region since the criteria of N points were met – this goes for the building and pieces of the road as well. However, a tree was successfully recognized in the first quadrant of the GSV image (*See Figure 4.4*).

The average F1-score, pixel accuracy and IoU were respectively 0.53, 87.53, and 0.42. From the performance, this approach was modest. Using SAM to further post-process the initial method proved to be harmful to the performance. Pixel accuracy still tends to be relatively high since we also consider non-greenery pixels – taking up a majority of images.

4.3.3 Gabor Feature Extractor and Random Forest with CCA

Connected Component Analysis (CCA) is a tool that recognizes groups of adjacent pixels sharing the same pixel value – typical regions contain ones surrounded by zeros. The image processing technique scans an image pixel by pixel and groups together connected pixels based on predefined connectivity criteria – 8-connectivity was used in this experiment [27]. An additional step was taken to best remove potential noise, a minimum area of 900 pixels was set. The area focuses on connected components smaller than the size of 900 pixels – if a connected component is smaller than the criteria then it is removed.

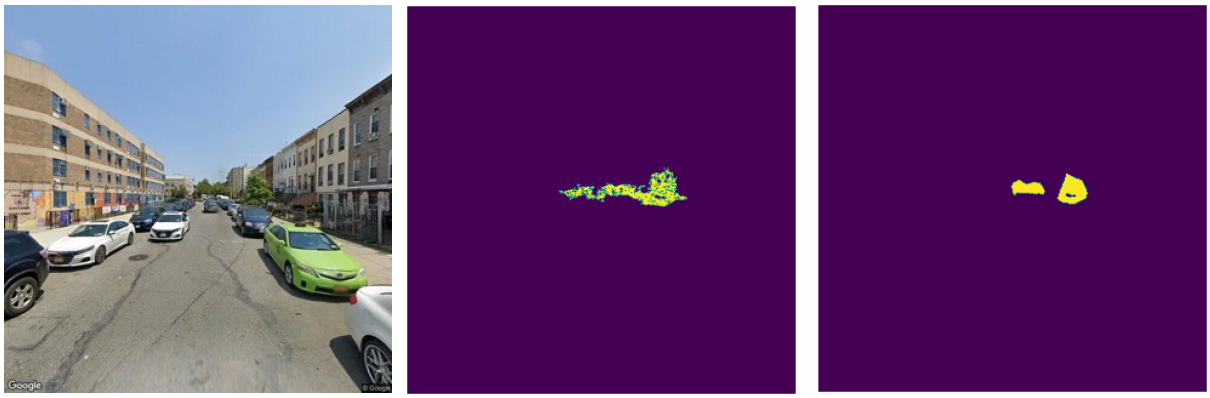

Figure 4.5: In the display we see the Gabor filter bank, Random Forest, and Connected Component Analysis working to highlight the green space of the Google Street View (GSV) image. The image in the middle is the predicted mask while the far right is the labeled image.

This additional post-processing technique to the initial approach proved to be useful in eliminating irrelevant regions. Challenging textures like those that exist in bricks and other surfaces seem to have been successfully removed in the provided example. Still, some pieces of a building that exists in the first quadrant of the image remain, though small it shows that the method can benefit from further improvements (*See Figure 4.5*). Overall, using CCA proved to be the most beneficial after evaluating the test images.

The average F1-score, pixel accuracy and IoU were respectively 0.60, 93.47%, and 0.48. The

mixed outcome suggests that the approach had a moderate ability in distinguishing between green spaces and the background across the test set – the approach was reliable in the broader task of segmenting greenery from GSV images. IoU being on the lower end indicates that the method struggles with precisely outlining object boundaries as less than half of predictions seem to overlap with true green spaces. With the F1-score being somewhat low it points towards the potential issue of achieving a balance between precision and recall. In this experiment, we still witness the approach still struggling to capture all relevant object pixels without including excessive background.

Approach	**F1 Score**	**IoU**	**Pixel Accuracy**
Base	0.58	0.45	92.16
Base and SAM	0.53	0.42	87.53
Base and CCA	0.60	0.48	93.47

Table 4.1: The table encapsulates the F1 Score, IoU, and Pixel Accuracy of all three approaches explored in extracting greenery from GSV images. Base refers to the first approach which includes constructing a Gabor feature extractor and adding a classifier known as Random Forest for predictions.

In summary, using the Gabor feature extractor and Random Forest with CCA proved to be the most effective method (*See Table 4.1*). Although pixel accuracy for each method was high, we must keep in mind that during the image segmentation process, we are also classifying non-greenery pixels, which make up a majority of the images, leading to high accuracy. Including CCA in the first approach slightly increased the F1-Score and IoU, making it the preferred method for evaluating census tracts. Overall, we observed modest performance improvements with each approach.

4.4 Discussion

In this chapter, we investigate three methods in order to extract green spaces using image segmentation from GSV images. The first approach used a Gabor filter bank and a classifier known as Random Forest to extract greenery. The second method adds a powerful image segmentation model known as SAM as an additional image post-processing solution. The third approach applies CCA to the output of the Gabor filter bank and classifier. From the three methods, CCA proved to be the best in terms of performance. Future improvements suggest acquiring a more accurately labeled dataset. Some of the labeled images highlight non-greenery regions such as walls, barks on trees, and other pieces that were not intended to be labeled. Overall, the results were accepted and the third approach was used to evaluate census tracts.

Chapter 5

Evaluating Census Tracts

5.1 Introduction

One of the main tasks that the CEJST tool provides is determining which census tracts are disadvantaged. Thereby helping federal agencies make funding allocations for communities in need. In the aim to supplement the tool, we investigate the predictive power in determining disadvantaged and non-disadvantaged census tracts through the features provided by the developed models. In this chapter, we investigate each of the features impact and introduce the ceiling met when conducting the analysis, followed by future implementations to further the project. We begin this investigation by selecting cities that are of most interest.

The three cities chosen for this study were from different regions of the United States and each offered unique characteristics, including environmental challenges, socioeconomic conditions, and urban infrastructures, which may provide informative findings in our results. This diversity is important because the fewer similarities the cities share, the more informative our results will be, as we are observing different features that may influence our outcomes. The chosen cities were Phoenix, Chicago, and the city of Atlanta. Atlanta holds one of the highest Gini coefficients in the United States. Potentially observing environmental discrepancies in the census tracts of Atlanta can be informative. Phoenix poses environmental challenges. The city has very little greenery due to the hot weather. This city introduces the fact that not all features may contribute to an accurate prediction – in this case, green spaces are

limited in all areas due to heat. Lastly, Chicago resides in the northern area of the United States. The city of Chicago introduces significant variations of infrastructure and brings a diverse urban landscape in terms of socioeconomic status and environmental characteristics. The varied characteristics the cities introduced highlighted the value of our results.

5.2 Census Tract Random Sampling

Due to low budgeting, only 36 census tracts were extracted from Phoenix and Chicago, and 38 census tracts were extracted from the city of Atlanta. To remove biases, census tracts were selected randomly from counties. The counties were determined by a drawn perimeter from the center of the city. This circular perimeter has a radius of 15 miles and a program was developed so that any county that resides within this circle was chosen for census tract random sampling.

Figure 5.1: A Metropolitan Statistical Area (MSA) of Atlanta, GA is shown. The map displays the counties encompassed by Atlanta.

For Atlanta, six different counties were noted which were Cobb, Dekalb, Fulton, Clayton, and Douglas. Fulton County was the only one encompassed by Atlanta's city limits - the others are considered to be the Metropolitan Statistical Area (MSA) of Atlanta (*See Figure 5.1*). This is important to recognize since more green spaces tend to be seen in suburban neighborhoods, located in MSA's, than urban living spaces. For the cities of Phoenix and Chicago, only one county (Maricopa and Cook County) fell within the generated perimeter; both counties are within city limits.

Once the counties were defined, 160 randomly selected points within each census tract were used to extract GSV images. At each point, we extract four GSV images since Google captures 360° scenes of any given space. However, because of randomizing the points in the census tracts, some points fell in spaces where no GSV image exists. The result of this was

that some census tracts hold more or less images compared to others. We accounted for the minimum number of images a census tract held in a city and used it as a baseline for the said city. The number of images per census tract for Phoenix, Chicago, and Atlanta were respectively 144, 192, and 164.

After downloading the images, each census tract is analyzed by all three models. The analysis provided the average number of garbage bags observed in a census tract, the average count of road damage per damage type, and the average percentage of greenery highlighted in an image. The model's output assigned five unique features to each census tract. Once the features were formulated into a data frame, we conducted additional analysis to evaluate the predictive strength of these features.

5.3 Assessment of the Current Feature Space

Achieving peak performance given the data set and task of identifying disadvantaged and non-disadvantaged census tracts proved to be challenging, but after exploring different methods one was deemed acceptable. Two base models were utilized for an ensemble approach. The first base model was a Support Vector Machine (SVM). SVM's proved to be useful since the classifier works by finding a hyperplane that best separates both classes in the provided feature space [20]. The second base model was Naive Bayes. This classifier calculates the probability of each class based on our input features and classifies data by selecting the class with the highest probability [21].

The majority vote is done by scikit-learn's Voting Classifier. Using the "soft" voting method the probabilities for each class, provided by our base models, are averaged and the final prediction is chosen by the highest average probability [22]. The probabilities for each class are stored and evaluated as confidence scores. Due to the limited data available, we

use Leave-One-Out Cross-Validation (LOOCV) to analyze our found method. Additionally, we assessed generalizability by training the method on all available data and then testing it on the same data set. This process was done individually for Phoenix, Chicago, and Atlanta, and also collectively for all three cities when the data was merged into a single data set. To visualize the predictions we implement Principal Component Analysis (PCA) on the data frame after classifications. The points are plotted into a two-dimensional feature space provided by PCA where we have N-components being two. Lastly, we study feature importance using Logistic Regression.

5.3.1 Evaluating Phoenix, AZ

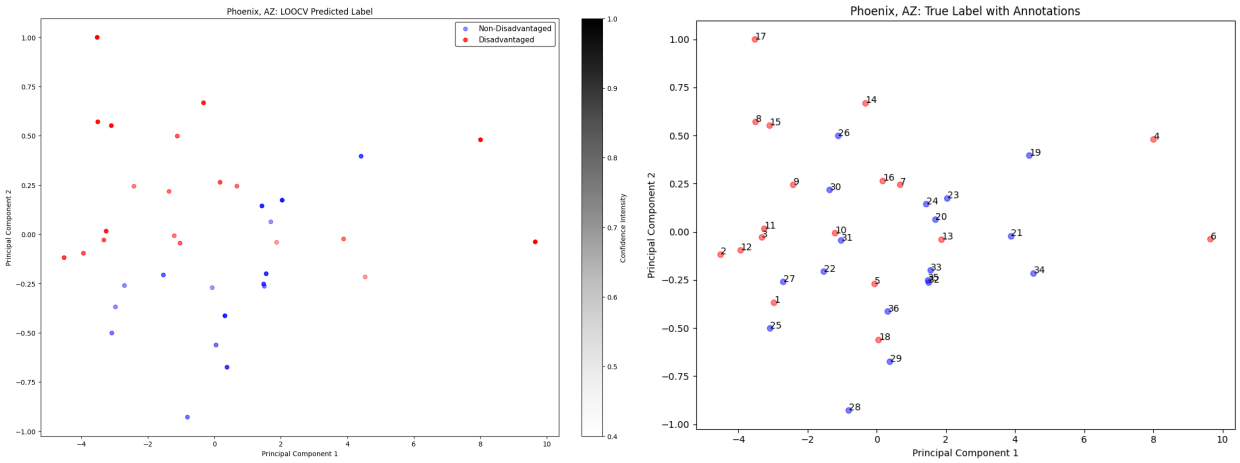

Figure 5.2: The left plot shows the developed ensemble method evaluating Phoenix, AZ via LOOCV – visualized by using PCA. The gradient shows the confidence scores of the predictions made for each census tract. Our right plot reveals the true labels for each census tract - each assigned an annotation.

In this experiment, we use the ensemble method to evaluate Phoenix, AZ. With the developed approach, we achieved an accuracy of 78% under LOOCV. Using the entire data set as a single training and testing sample led to an accuracy of 81%. The performances achieved exceeded expectations. Some census tracts were on the verge of being classified correctly.

As an example, census tracts 34 (*See Figure 5.2*) had a low confidence score since its true label is non-disadvantaged.

5.3.2 Evaluating Chicago, IL

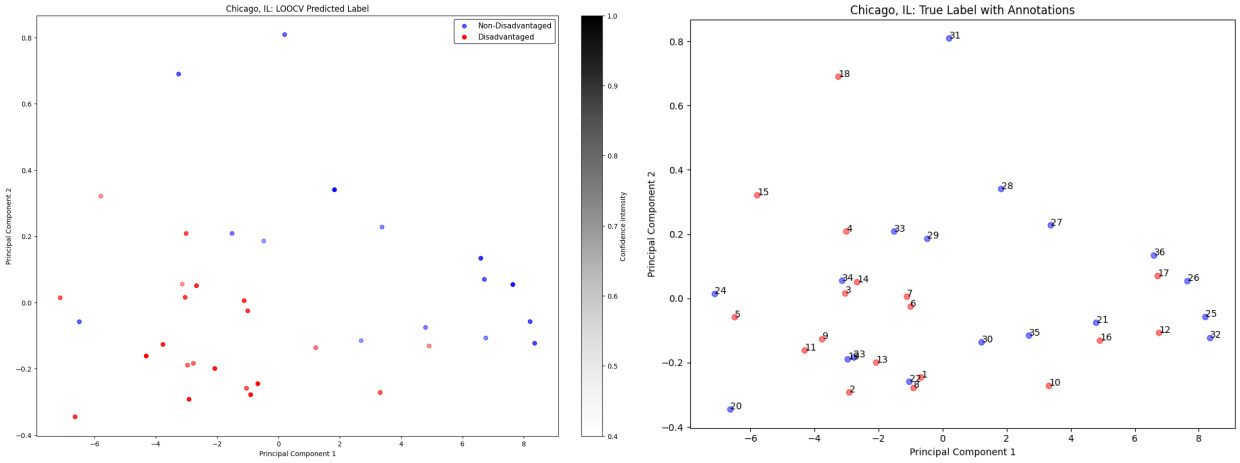

Figure 5.3: The first graph displays the ensemble method's evaluation of Chicago IL with LOOCV – visualized by using PCA. The gradient bar presents the confidence scores of the predictions made for each census tract. The second graph displays the actual labels assigned to each census tract tagged with annotations.

For Chicago, the ensemble method achieved an accuracy of 69% via LOOCV. Using the entire data set as one training and testing phase we achieved an accuracy of 75%. The moderate performance shows the ceiling of the predictive power the current features have on the census tracts, which then highlights areas for potential improvement. For instance, broadening the array of base models within the ensemble method could help capture more complex patterns (*See Figure 5.3*).

5.3.3 Evaluating Atlanta, GA

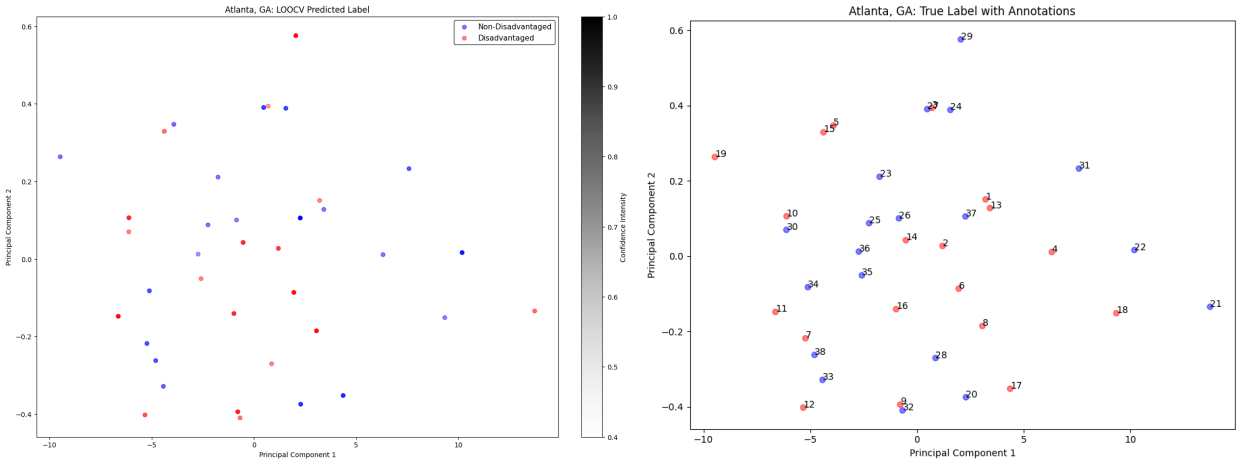

Figure 5.4: The left graph shows the ensemble method's evaluation of Chicago IL by LOOCV. Both graphs are visualized by applying PCA to Chicago. The gradient bar gives the confidence scores of the predictions made for each census tract. The second graph displays the actual labels assigned to each census tract.

Atlanta proved to be the most challenging in terms of improving performance. The accuracy, using the ensemble method, was 66% under LOOCV. The accuracy in which no data splitting occurs was also 66%. The non-existent gap suggests that the acquired data lacks some degree of variability. With this approach, we were relatively stable and consistent in our predictions across the small dataset – especially during each iteration of LOOCV (*See Figure 5.4*).

5.3.4 Evaluating All Cities

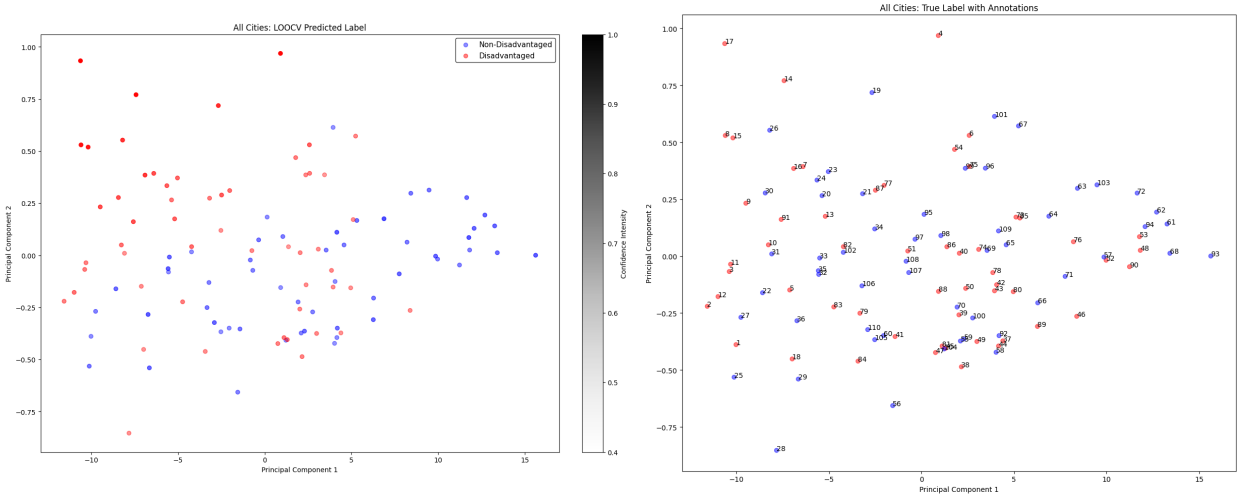

Figure 5.5: The ensemble method's evaluation by LOOCV of all cities configured into one data frame is shown. Both graphs are visualized by applying PCA. Confidence scores of the predictions made for each census tract are shown by the gradient bar. The second graph displays the actual labels assigned to each census tract.

After normalization, we combined all data sets into one singular data frame for a total of 110 census tracts. Applying the ensemble method via LOOCV gave an accuracy of 76% and the no-split approach gave 67%. At first, the high performance by LOOCV was unexpected, but given that more data was available the ensemble method was able to capture complex patterns hidden in the feature space. The roughly 9% gap in both exams suggests that extreme outliers are involved in this combined data set. We recognized that obtaining more data can give further insights into finding a better approach to classify disadvantaged and non-disadvantaged census tracts with the current feature space (*See Figure 5.5*).

5.3.5 Feature Importance with Logistic Regression

Feature	**Coefficient**	**P-Value**
Longitudinal Cracks	0.015	0.341
Transverse Cracks	0.016	0.180
Alligator Cracks	0.011	0.092
Greenery	-0.007	0.189
Garbage	8.0×10^{-4}	0.652

Table 5.1: The table above details the Logistic Regression coefficients for each feature along with their P-Values. This experiment followed the evaluation of 110 census tracts collectively from the cities of Atlanta, Phoenix, and Chicago. Half of these were disadvantaged census tracts, while the other half were non-disadvantaged census tracts.

A quantitative assessment for feature importance of the all cities data frame was explored. With the use of Logistic Regression (LR) and LOOCV, we achieved an accuracy of 56%. The performance achieved by the classifier was poor but it gave insights into the relevancy of the features had in the predictions. The LR coefficients were small but the directions of the values agreed with the initial assumption. Longitudinal, transverse, alligator cracks, and garbage bags point toward disadvantaged census tracts while additional greenery suggests a non-disadvantaged census tract. The magnitude of the P-Values points toward statistical coincidence, the features do not suggest strong associations with the census tracts

5.4 Variability Across Different Landscapes

In this section, we conduct a qualitative analysis of the results found in the experimental phase. Having an understanding of performance from an instinctive perspective rather than a quantitative view can help provide federal agencies with an idea of how this tool may be used. The results of the analysis led to two findings. Chosen features require careful consideration and a much more robust rationale for the selections made. The second suggests that different cities may need their own unique model – finding a universal model that is applicable to all cities in the United States will be much more difficult to develop.

5.4.1 Feature Importance

After conducting a quantitative assessment of the chosen environmental features, we realized a more in-depth qualitative assessment was needed. Important insights were found when conducting an analysis on Phoenix, AZ. Throughout the year the city faces extreme weather conditions. Heat can generally be very extreme and introduce complications to communities in the state of Arizona. Due to the heat, there exists a lack of greenery. Choosing a feature that may not be present in all census tracts can harm a model from making an accurate prediction or providing a perspective of the current status of a census tract.

Figure 5.6: Four Google Street View (GSV) images were chosen to display Phoenix's greenery. The top two images give insights into the lack of greenery some places in the city may experience due to the weather, but not all places in Phoenix experience a loss in vegetation. The bottom images provide examples of how green spaces still exist in the chosen census tracts.

The average percentage of green space observed in Phoenix for disadvantaged and non-disadvantaged census tracts were respectively 8.25% and 9.55%. Phoenix had half the percentage of greenery compared to Chicago and Atlanta which were both over 15% for each class of census tracts. Observing which environmental features are more prevalent in both classes of census tracts could bear fruit for a more informative analysis (*See Figure 5.6*). A similar intuitive conclusion came from the city of Atlanta.

Figure 5.7: Four Google Street View (GSV) images from Atlanta, GA are shown. They give perspective into the different types of locations gathered for the analysis. All images seem to be coming from a suburban environment within the MSA of Atlanta.

Since we extracted census tracts from outside the city limits of Atlanta, parts of suburban

environments were chosen (*See Figure 5.7*). In suburban living spaces, we see more greenery. Urban environments tend to have less greenery due to buildings, frequent sidewalks, streets, and other areas that limit the amount of green spaces. Because greenery tends to be very common in suburban environments, using it as a feature to learn from census tracts may not prove so useful. The results led to the conclusion that, as one changes the environment, an investigator will need to place more careful thought into which environmental features can reveal relevant information about a given census tract.

5.4.2 Tailored Models for Individual Cities

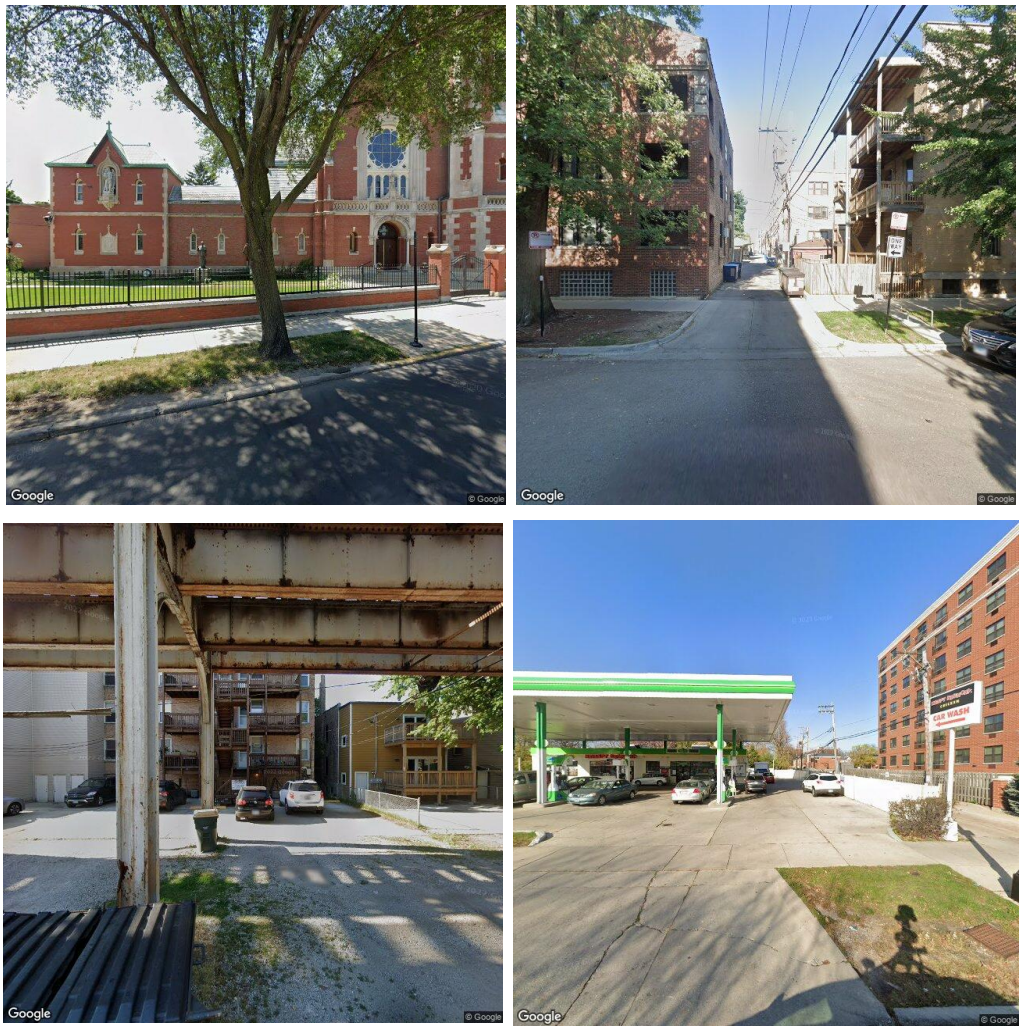

Figure 5.8: Four Google Street View (GSV) images are presented from Chicago, IL. The images show additional features of the city including the city's architectural style.

After observing images from Chicago, additional environmental features present in GSV images could be considered. Some of Chicago's infrastructure showed degrees of aging in terms of corrosion or other sorts of anomalies (*See Figure 5.8*). Distinctive building infrastructures built in previous decades were also noticeable. This led to an assessment as to whether specific cities will benefit from their own tailored model.

The tailored model is to account for unique features only present in individual cities. For example, if degradation in infrastructure is prominent in the city of Chicago, but only to

certain communities, the feature becomes even more powerful in informing an investigator of a census tract. Tailoring models that can use these more prominent features can be more useful than using a feature space that is only relevant to another city. Ultimately, the decision of whether to develop a new model for an individual city is determined by their chosen feature space.

5.5 Future Directions in Census Tract Analysis

For future work, we start by helping the community comprehend the analysis results. First, detailing the project's budgeting costs. After data collection, we acquired a total of 18,820 images from a total of 110 census tracts across three cities. Google charges $7 USD for every 1,000 GSV API requests [23]. Every downloaded image is considered to be a request. The total project cost amounted to approximately $131 USD. This cost is significantly lower compared to conducting a structured survey evaluating environmental features of the same census tracts. This sheds light on the financial efficiency of analyzing environmental characteristics via GSV.

GSV provides a variety of images at the same locations; images that are captured across different months and years. Access to these varied images can give a better perspective of how a census tract changes over time. In this analysis, we were able to download the first image Google offered of each location – the dates were randomized and could not be controlled. This access enables a further analysis of how a census tract evolves through different periods of time.

The second piece to consider for future implementations is to develop a method that accurately identifies the important environmental features that may define a census tract's status. Once key features are recognized they can provide valuable insights about census

tracts. Relevant features influence the accuracy of a model aimed at distinguishing these communities. It is also important because the focus of the model is to correctly classify census tracts, thereby offering new perspectives into census tracts as they are analyzed.

A potential method to recognize the relevant environmental characteristics in a census tract can begin by collecting a large number of images from a city. We then use these images to perform a binary classification to identify disadvantaged and non-disadvantaged census tracts with the help of a Convolutional Neural Network (CNN). After achieving an acceptable accuracy, one can analyze the feature maps provided by the CNN to potentially acquire more information on relevant environmental features that require closer attention. This analysis can lead to a valuable perspective on potential overlooked features.

5.6 Discussion

In this chapter, we discussed the evaluation method developed to distinguish the different classes of census tracts. After achieving the reported performance, we conduct a qualitative analysis of the results found. We arrive at a conclusion which suggests that choosing the relevant features for a city that is strongly associated with the different classes of the census tract is important. One's chosen feature space determines the reliability of a developed model. With correctly chosen environmental features present in GSV images one can be more accurately informed about the status of a census tract.

Chapter 6

Conclusions

In this study, we aim to supplement the Climate and Economic Justice Screening Tool by providing five additional environmental features of census tracts. Through Google Street View (GSV) images, we were able to recognize garbage bags, road damages (transverse, longitudinal, and alligator cracks), and greenery using image classification, object detection, and image segmentation. After developing the models, the tools were used to analyze 110 census tracts across the cities of Phoenix, Chicago, and Atlanta. The analysis of the results led to informative findings that provide federal agencies insights into the potential use case of GSV images.

GSV images are more accessible, less expensive, and offer granularity compared to the current methodology of evaluating census tracts. The accessibility allows federal agencies to monitor any proposed and ongoing enhancement initiatives in census tracts. Also, for downloading 1,000 GSV images, Google charges $7 USD [23]. Google also takes multiple 360° scenes from locations on a monthly basis. They hold images that span over a decade across many locations. Using GSV images to extract environmental features from census tracts is less expensive, more time efficient, and gives more variability in terms of available data.

After conducting a qualitative analysis to distinguish the different classes of census tracts using the developed feature space, it was concluded that selecting the appropriate features may significantly enhance performance. This became apparent after evaluating the city of Phoenix. Phoenix's extreme heat waves throughout the year affect its green space no

matter the class of a census tract. Using greenery as a distinguishing feature requires careful thought. If all census tracts lack greenery, using it to gather information about the status of a census tract may not prove to be very informative.

Lastly, for federal agencies to use this methodology, additional research is required to determine the best feature space for a given city. The unique characteristics of a city like architectural design, urban planning, or other environmental features present in GSV images can be used to evaluate census tracts. The defining characteristics of a city introduce complexities in determining the best feature space. Though, once a feature space is appropriately determined, it can provide a reliable methodology for evaluating census tracts; thus, supplementing the Climate and Economic Justice Screening Tool.

Bibliography

[1] White House. The Council on Environmental Quality - About. National Archives and Records Administration. https://obamawhitehouse.archives.gov/administration/eop/ceq/about

[2] White House.(2022, August 9). Council on Environmental Quality. The White House. https://www.whitehouse.gov/ceq/#:~:text=The%20Council%20on%20Environmental%20Quality,America's%20public%20health%20and%20environment.

[3] White House. "Biden-Harris Administration Launches Version 1.0 of Climate and Economic Justice Screening Tool, Key Step in Implementing President Biden's Justice40 Initiative." The White House, The United States Government, 22 Nov. 2022.

[4] United States Census Bureau. Census tracts. Census.gov. https://www2.census.gov/geo/pdfs/education/CensusTracts.pdf

[5] Council of Environmental Quality. "Climate and Economic Justice Screening Tool ." Nov. 2022, screening tool.https://screeningtool.geoplatform.gov/

[6] Multi-Resolution Land Characteristics (MRLC) Consortium. "NLCD 2019 Percent Developed Imperviousness (CONUS)." NLCD 2019 Percent Developed Imperviousness (CONUS) | Multi-Resolution Land Characteristics (MRLC) Consortium, 1 Jan. 2023.

[7] United States Environmental Protection Agency. (2024, January 3). EJScreen Map Descriptions. EJScreen: Environmental Justice Screening and Mapping Tool. www.epa.gov/ejscreen/ejscreen-map-descriptions

[8] U.S. Government Accountability Office. "National Highways: Analysis of Available Data Could Better Ensure Equitable Pavement Condition." National Highways: Analysis of Available Data Could Better Ensure Equitable Pavement Condition | U.S. GAO, United States , 28 July 2022.

[9] "U.S. Census Data." Usa.gov, United States Government, 27 July 2023, www.usa.gov/census-data

[10] U.S. Census Bureau. "2019 American Housing Survey Data Now Available." Census.Gov, United States, 8 Oct. 2021, www.census.gov/newsroom/press-releases/2020/2019-american-housing-survey.html

[11] Multi-Resolution Land Characteristics (MRLC) Consortium. (2021, March 25). MRLC NLCD Eva Tool. EVA Tool. www.mrlc.gov/eva/

[12] Google. How street view works and where we will collect images next. Google Maps Street View. www.google.com/streetview/how-it-works/

[13] Arya, Deeksha, et al. "RDD2020: An annotated image dataset for Automatic Road Damage Detection Using Deep Learning." Data in Brief, vol. 36, 18 Sept. 2022, pp. 1–16.

[14] Ai, Dihao, et al. "Computer Vision Framework for Crack Detection of Civil Infrastructure-A Review." Engineering Applications of Artificial Intelligence, Pergamon, 21 Oct. 2022

[15] Wang, Chien-Yao, et al. "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors." Computer Vision Foundation. 6 July, 2022. https://arxiv.org/pdf/2207.02696

[16] Pham, V., Nguyen, D., Donan, C. (2022, October 31). Road Damages Detection and Classification with Yolov7. arXiv. https://arxiv.org/abs/2211.00091

[17] Sukel, Maarten. "Garbage Object Detection Using PYTORCH and Yolov3." Medium, Maarten Sukel, 10 Oct. 2019, www.medium.com/maarten-sukel/garbage-object-detection-using-pytorch-and-yolov3-d6c4e0424a10

[18] IBM. "What Are Convolutional Neural Networks?" IBM, www.ibm.com/topics/convolutional-neural-networks.

[19] Smith, Leslie N. Cyclical Learning Rates for Training Neural Networks, U.S. Naval Research Laboratory, 4 Apr. 2017, www.arxiv.org/abs/1506.01186.

[20] MathWorks. "Support Vector Machines for Binary Classification." MathWorks.

[21] Scikit-Learn. "Naive Bayes — scikit-learn 0.24.2 documentation." Scikit, www.scikit-learn.org/stable/modules/naive_bayes.html.

[22] Scikit-Learn. "Sklearn.Ensemble.VotingClassifier." Scikit, www.scikit-learn.org/stable/modules/generated/sklearn.ensemble.VotingClassifier.html.

[23] "Platform Pricing & API Costs." Google Maps Platform, Google.

[24] Scikit-Image. "Gabor filter — scikit-image v0.19.0 docs." Scikit.www.scikit-image.org/docs/stable/auto_examples/features_detection/plot_gabor.html/.

[25] Dwyer, B., Nelson, J., Hansen, T., et. al. (2024). Roboflow (Version 1.0) [Software]. Available from https://roboflow.com. computer vision.

[26] Kirillov, Alexander, et al. "Segment Anything." arXiv preprint arXiv:2304.02643 (2023)

[27] OpenCV. "Structural Analysis and Shape Descriptors — OpenCV 3.0.0-dev documentation." OpenCV. https://docs.opencv.org/3.0-beta/modules/imgproc/doc/structural_analysis_and_shape_descriptors.html#connectedcomponents

[28] The White House. Justice40 initiative: A Whole-Of-Government Initiative. https://www.whitehouse.gov/environmentaljustice/justice40/

[29] Antonelli, W. It can take years for google maps to update certain features - here's how they get the data to update Street View, traffic, and more. Business Insider.https://www.businessinsider.com/guides/tech/how-often-does-google-maps-update

[30] Larkin, A., Hystad, P. Evaluating street view exposure measures of visible green space for health research. J Expo Sci Environ Epidemiol 29, 447–456 (2019). https://doi.org/10.1038/s41370-018-0017-1