Detecting, Mapping, and Grading Sidewalks using Street View Images and Secondary Sources for the city of Dallas

Andrew Abbott1, Alex Deshowitz1, Dennis Murray1, Dr. Eric Larson1

1 Southern Methodist University, Dallas, Texas

**Abstract.** Sidewalks provide visitors and residents of municipalities with a means of accessing the resources in the area. However, for individuals with mobility challenges, the mere presence of a sidewalk is often not enough information to determine whether they are able to easily access a point of interest. Features such as curb ramps are often much more important in the planning of a route for the mobility challenged. In this research, machine learning methods were used to identify curb ramps in “Street View” images. The data developed from this task would create another layer of geo-coded data that could be added to maps and their metadata. .The image dataset consists of XX panoramic street view images, with a “matching bounding box” dataset that identifies the location of curb cuts. In total, the data consists of more than 40,000 examples of the actual locations of curb ramps as well as locations where curb ramps are expected. . We used a convolutional neural network to identify the presence of curb ramps in the Project Sidewalk images, and benchmark the performance against a logistic regression model.

1 Introduction

For individuals with mobility limitations, the availability of sidewalks with quality curb ramps represents a necessity in navigating a city or urban environment. In the absence of this information wheelchair users are forced to rely on their own experience and knowledge of a specific area to successfully navigate to their destination in an efficient and safe manner. Without this knowledge, a wheelchair user may be required to take much longer routes, may sustain serious injuries, or may choose not to travel at all. Previous attempts at identifying curb cuts have been carried out through in-person (subjective) “Neighborhood Audits” or through information cataloged in Geographic Information Systems [18]. The completeness, timeliness, and quality of the information gathered in these methods is often lower than what is needed for a comprehensive view of the availability of a feature like curb ramps or sidewalks across a city. These approaches also tend to come with a very high cost of acquisition.

The city of Dallas approached SMU with the previously stated concerns regarding their infrastructure. There is a myriad of reasons for the city of Dallas to focus a portion of spending on sidewalk creation and repair. For the mobility impaired, availability of sidewalks is an essential requirement for moving about the metropolitan area. Additionally, sidewalks allow citizens to move in a safe manner around the city without the risks associated with walking in the streets amongst cars driven by ever-increasingly distracted drivers. Sidewalks also provide health benefits since they provide a means of pedestrian travel to near-home destinations for those who may otherwise choose to not travel. Lastly, sidewalks in disrepair tend to contribute to the “broken-windows” theory surrounding many under-privileged areas of the metroplex. Overall, sidewalks represent a worthwhile investment for the city and provide a means of safe and healthy travel for those who choose pedestrian transportation.

Despite the obvious benefits of accessibility to a city and its citizens, the problem of documenting and mapping current levels of accessibility is persistent. This represents a problem for the city in allocation of resources to the proper parts of the city. It also represents a problem to the citizens and visitors: difficulty navigating the city despite gaps in the availability of sidewalks and curb ramps for wheelchair users.

In the fiscal year 2015, Dallas had a proposed budget for infrastructure projects, such as sidewalks, of $7,135,208. That number has grown each of the past several years as the city has increased in size. The city of Dallas has grown in both its inhabited sprawl and in the density of the population at double-digit rates over the past 5 years. This growth can be explained by the city’s pro-business mentality and the relative value proposition that such a large city provides. Dallas boasts a centrally located position in the country, and the city’s status as a transportation hub allows businesses to send employees to any destination in the world with relative ease and efficiency. The City and State have pro-business tax and incentive policies, which have encouraged this growth. With this growth, the city must acquire a better way of allocating the budget for sidewalks and other infrastructure. Today, the city of Dallas essentially responds to complaints about sidewalks through its street services program. As inquiries come in, the city will generally put that particular inquiry on a list for assessment. There is no priority granted for severity of the situation. Additionally, maintenance in suburban areas is the responsibility of the home or property owner and this is not necessarily considered in the ranking process. The current process takes 2-3 months in order to get an assessment and cost estimate for each incident or property owner. Once the assessment has been done, either the city or the property owner will plan and fund the project. One program in Dallas allows the city to reimburse homeowners up to $500 or 50% of the repair cost, whichever is less.

Under these circumstances, the city spends significant time and resources just completing the assessments. Using the algorithmic approach that this paper describes, the city would be able to feed images of these incidents into the model and immediately receive a grading of the sidewalk in question. The city officials could then compare this grading to the grading of previous works to know whether this sidewalk was an immediate issue, who owns the sidewalk, and how much the repair may cost. This would allow the city to prioritize project desires into bins such as: critical, severe, moderate, and low-risk. Therefore, Dallas could dispatch crews to the areas where their services will be the most impactful to the safety and health of the public.

To rectify the lack of city-wide mapping of sidewalks and curb cuts, this paper demonstrates a machine learning based solution to the problem. Using a training set of images from Project Sidewalk at the University of Maryland, several machine learning methods were evaluated for accuracy in detecting curb cuts in a sidewalk. A method to extend the trained model to the full geography of the city of Dallas is also outlined. Additionally, we suggest an approach to leveraging this model to make sound recommendations regarding areas of focus for the city planning commission.

2 Previous Research

The current reach of accessibility features in the urban landscape is central to the research that was performed for this paper. Bennett, Kirby, and MacDonald [17] surveyed 79 intersections in Halifax, Nova Scotia. Their scoring methodology asked 8 different questions that addressed both the presence and quality of curb ramps at these intersections. Each question required a binary response. Several of the questions would appear to be answerable from the research we propose – the presence of curb ramps, accessibility from the line of travel (that is, the chair user can access the ramp without exiting the crosswalk), that the ramp is “free from irregularities”, and free from drainage grates. Four additional questions address the question of slopes and dimensions of the curb ramp. Their findings in the limited scope of the survey was that 98.7% of intersections had curb ramps, but just more than half, 53.8, had a direct line of travel from the crosswalk. All of the ramps were free from drainage grates, and 85.9% were free from irregularity. The average intersection scored 5.6. The researchers proposed that wheelchair users must adapt to the lack of infrastructure by increasing their skill and dexterity in maneuvering the chair. While improved skill among wheelchair users is desirable, it may also lie beyond the physical capabilities of the individual user.

It is important to view the context of provisions for access as not special accommodations for persons with disability, but instead bringing the world to be equally accessible to all people. Bromley et al [18] noted in review of legislation in the United Kingdom seeks to provide access to goods and services to all persons, but not necessarily the facilities containing goods and services. It is a fine distinction between the two, and within this context it could be judged that this is the granular difference that describes how accessibility isn’t a special accommodation but provides equal access to all. Respondents in this survey-based study in Swansea, Wales found 60% thought that lack of curb ramps were a “major” or “prohibitive” obstacle to access. As a result, respondents had to use domain knowledge of the city to navigate around obstacles, and sometimes take much longer paths to access. Among the respondents, 60.8% agreed that “the way places are designed” is the major problem for wheelchair users. This attitude was somewhat more evident among younger users of wheelchairs than their older cohort. Wheelchair users recommended “more dropped kerbs” more often than any other improvement to the center city shopping experience.

Another important piece of research was Clarke et al’s [13] audit of Streetview images as compared with an individual’s in-person audit. The study involved researchers in neighborhoods in Chicago walking each block from the inside to the outside, essentially walking the block twice, and assessing the quality of the sidewalks. This study found that subjective measures like sidewalk quality have much lower consistency between observation via Streetview and in-person observation and grading. Essentially, the conclusion is that features requiring high levels of precision can be hard to attain via Streetview images. This poses an interesting aspect to our research. For instance, our model will need to stay informed of areas that have been treated in the previous time periods. Therefore, if a database of previously updated sidewalks does not exist, we will need to provide a means of storing projects within a database that allows for those items or coordinates to be referenced. This will prevent outdated Streetview images from being used in the classification and scoring process.

While it is important to see that our research problem can be solved via machine learning techniques, it is also important to see that there is indeed a reason for the application of these techniques to solve the task at hand. Therefore, it is important to see that improving sidewalk quality, coupled with other factors can lead to better health for society overall. In Haina et al [11], the researchers looked at signal data such as walkability of neighborhoods in relation to the overall health of the individuals in the area. The evidence used to provide insight into the improved environment of an area was sanitation practices and tobacco sales restriction. This coupled with increased walkability of an area leads to higher levels of physical activity and better health over time.

The fourth area of research for this project focused on the general health benefits of neighborhood walkability. Deehr and Shumann [7] provided work for five different neighborhoods in the Seattle area. Their research considered the incidence of pedestrian strikes by motorists, the health factors of walking, and the current modes of transportation that pedestrians were using. Their research led to the city adding additional walking paths, and trails. Additionally, much of the research sparked additional community involvement in the design of multi-model transportation infrastructure. Additionally, in Richardson,Troxel et al [10], the authors sought to understand whether factors such as green space and walkability resulted in “moderate to vigorous physical activity” for the residents of randomly selected neighborhoods in Pittsburg, PA. When controlling for factors such as crime, green space, and walkability in the selected targets, it was discovered that variables such as gender, age, education, and overall walkability of the neighborhood did play significant roles in the levels of physical activity for an area. This research helps us reaffirm that there is immense potential for identifying areas that need this sort of infrastructure. Ultimately the goals of helping people lead healthier and safer lives are potential outcomes of the modeling exercise laid out in this paper.

Image recognition is not a new field. The use of machines to recognize images has been around for decades. As early as 1963, the electrical engineering department at MIT began using computers to recognize 3D images. While these initial applications were somewhat simple compared to those that we currently use today, they paved the way for what has now become a commonplace practice across industries.

Bahlman, Zhu, and Pelkofer’s work [3] provided meaningful advancements in image element detection and classification. In this paper, the authors built upon their previous research involving shape and color recognition to help classify street signs and traffic signals. Their work is interesting in that it has a 2-step approach where if the model fails on the first classification step, the image is thrown out. This model is important to our work, because it shows how an algorithm such as Adaboost can be used to detect both anomalous and important features for an image-based problem proposal.

Another important piece of research is Perona’s “A Bayesian Hierarchical Model for Learning Natural Scene Categories” [4]. In this paper, the authors provided an approach that allowed for very hands-off model building. This model will potentially provide a structure for our model should we encounter any issues with sparse image objects that are hard to classify correctly. In the model, the computer attempts to use human-based rules to classify image objects. Essentially, each image is broken down into a series of codebook images and reoccurring elements are scanned and classified. Each of these codebook images is additionally clustered using k-means clustering. This portion of the model is used to eliminate features that occur with low frequencies in the training data. The remarkable aspect of this paper is that the model was able to achieve a 78% accuracy rate with such a low amount of supervision. While there are many applications of image classification models, the models outlined above provide a solid basis for our understanding of the evolution of image recognition and model application. Our next area of concentration has been on the specific use of neural networks for problem solutions in the image recognition and classification space.

Dean, Corrado and a group of Google researchers in [20] created the predecessor for the modern open source library, TensorFlow in a model labeled as “DistBelief.” DistBelief allows for parallel processing of training data both within a machine and across a network of machines. Likewise, the DistrBelief process also allows “data parallelism” allowing for “multiple replicas of a model to optimize a single objective.” The effect, the researchers conclude, is a method for training moderate sized models more quickly than before, and giving capability to training very large data set models.

DistBelief was the basis for the 2015 release of the open source TensorFlow machine learning system, documented in Abadi,, Barham et al in [21].

An aspect of previous work that is of high importance for this work is the use and application of convolutional neural networks. As this is our model of choice, it was important for us to research the application of convolutional neural networks and their potential pitfalls. Goodfellow and a team from Google [5] showed an application of neural networks for image recognition. In this work, Goodfellow applied the DistBelief method for neural networks combined with Google Streetview images to recognize multi-digit numbers, namely street addresses. In the model, the researchers first addressed training the model to identify house numbers. This was a very important step as many variables come into play with these image captures. For instance, lighting, obstructions, and changing conditions can provide potential issues when identifying numbers from images. Additionally, varying font sizes, colors, and styles can impact the ability of the algorithms to correctly identify an image. An important aspect of this type of recognition is that if a single digit is misidentified, the entire interpretation is irrelevant and meaningless. Once the model was trained on house numbers, a more complete Streetview dataset was used. The final approach involved subtracting the mean from each image. In the end, the researcher’s models were able to achieve a 97.84% accuracy with this approach, which was just short of the human benchmark of 98% that was the target of the project. This piece of research and the approach acted as an important catalyst for our approach to identifying sidewalk obstructions and sidewalk grading.

Convolutional neural networks have also been used to improve the solutions submitted in the ImageNet Large-Scale Visual Recognition Challenge. In the work of Simonyan, Karen, and Zisserman [6], the team used convolutional neural networks combined with several other approaches to achieve one of the highest accuracy levels seen in the competition. Their application of multiple models to solve the problem provides a solid reference point for the problem that we solve in this model.

Logistic Regression and Artificial Neural Networks have become benchmarks in classification tasks across problem types. Dreiseitl and Ohno-Machado [19] researched the methodology of machine learning methods across more than 70 papers. The authors state that the two methods, Logistic Regression and Artificial Neural Networks, both have similar basis: statistical pattern recognition in large data sets. The authors reviewed 72 papers that compared outcomes of implementations of both logistic regression, and artificial neural networks. Artificial neural networks outperformed logistic regression in 51% of the studies, but 42% of the studies provided no difference in outcome between the methods. The underlying context, and understanding the authors are seeking is the ability to implement a model to a medical context. They label logistic regression and several other methods as a “White box” method, where the parameters are clearly stated and the method that the model uses to assign importance and come to a conclusion are clearer. In contrast, Artificial Neural Network and support vector machines are labeled as “black box” methods that do not provide interpretable markers of importance or provide methods to be verifiable.

Overall, our research helped us layout the precedent for image recognition, understand the application of the specific type of model that we are attempting to build, evaluate the effectiveness of proper infrastructure, and provide statistical affirmation of the health and societal benefits of proper infrastructure. With this knowledge established, we can make the case for our model to be used for the stated application problem.

4 Algorithm Design and Solution

The approach selected implemented a convolutional neural network as the final solution. The model employs a simple 4-layer approach and the implementation of relu as the activation function with the popular and efficient Adam optimization algorithm. In the interest of processing time, the images received from the Google StreetView were rescaled to 100 by 100 pixels with the color vector retained in the image. Each image was also scaled prior to being fed into the model. While there are many means for strong cross-validation procedures in existence, the chosen approach in this instance was a simple 80/20 split. This provided the model with plenty of data from which to learn and a modest amount of data for testing. Additionally, the model used the Keras sequential model with Google’s TensorFlow backend. The model also used 30 filters for each image and the size of each scanning filter was 3 by 3. This approach, while standard proved to be fairly robust and prevented the model from over-learning. Additionally, max pooling was also employed as a measure of overfitting, and a pool size of 2 by 2 was used. In order to test the ability of the model to improve upon its accuracy, a GridSearch of the filter parameter was performed. Upon further evaluation, of the filters, it was ultimately decided that 30 filters was optimal. Since adding additional filters did not tend to increase accuracy, and greatly increased compute time, it was decided that the 30-filter approach was indeed a sound model parameter.

The resulting model identifies curb cuts correctly 80% of the time. As a baseline approach, a Support Vector Machine and a Logistic Regression were both trained on the same data and evaluated for accuracy. Both models only resulted in an accuracy score of 67%. Therefore, the Convolutional Neural Network did indeed provide a level of precision that could not be attained in more basic machine learning approaches. It was also noticed that the time needed to train these models was also considerably more than the amount of time needed to train the neural network.

The output of this approach leads to the implementation of the QSI model that can then be used for municipalities to identify areas of focus for accessibility investment. The QSI is simple. The score is essentially just the number of curb ramps divided by the number of intersections for each 1 mile of street. It is recommended that streets be broken into 1-mile blocks for assessment purposes. This ensures that city grids do not become diluted in the overall assessment of the city streets. (NEXT STEPS)

Currently, we are working on a way to ingest images of 1-mile street blocks from areas of Dallas in order to score those streets with our model. We will then be able to rank 1-mile street sections for Dallas, see how Dallas compares to other major metropolitan areas, and recommend areas for attention to the city planning committee. This recommendation will be the ultimate output of our presentation

5 Project Plan

Domain Knowledge Research

Understand Social and Public Health considerations of sidewalks

Pedestrian Safety concerns

Persons with Physical Difficulties

Other Approaches to the Problem – Neighborhood Audits

Python Image Processing Knowledge

Google/Bing Street View API Knowledge

Machine Learning/Neural Networks with Images

Best Practices for improving performance in image classification problems

GIS (Geographic Information System) file formats and interchanges

Development of Machine Learning/Neural Network, Cross Validation, Testing

Documentation

Final Poster Presentation

6 Github

The address for this project’s Github page is: [github.com/dpmurraygt/CapstoneProject](file:///C:\SMU%20Data%20science\capstone\AppData\Local\Temp\Temp1_CapstoneProject-master.zip\CapstoneProject-master\github.com\dpmurraygt\CapstoneProject)

References

1. Kotaro Hara, Shiri Azenkot, Megan Campbell, Cynthia L. Bennett, Vicki Le, Sean Pannella, Robert Moore, Kelly Minckler, Rochelle H. Ng, and Jon E. Froehlich. 2015. Improving public transit accessibility for blindriders by crowdsourcing bus stop landmark locations with Google Street View: An extended analysis. ACM Trans. Access. Comput. 6, 2, Article 5 (March 2015), 23 pages.
2. Kotaro Hara, Jin Sun, Robert Moore, David Jacobs, Jon E. Froelich. 2014. Tohme: Detecting Curb Ramps in Google Stret View Using Crowdsourcing, Computer Vision, and Machine Learning. UIST ’14, October 5-8 2014. Honolulu, HI.
3. Claus Bahlmann, Ying Zhu, Visvanathan Ramesh, Martin Pelkofer, Thorsten Koehler. “A System for Traffic Sign Detection, Tracking, and Recognition Using Color, Shape, and Motion.” IEEE Xplore, July 2005.
4. Fei-Fei, Li, and Pietro Perona. "A bayesian hierarchical model for learning natural scene categories." *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. Vol. 2. IEEE, 2005.
5. Goodfellow, Ian J., et al. "Multi-digit number recognition from street view imagery using deep convolutional neural networks." *arXiv preprint arXiv:1312.6082* (2013).
6. Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).
7. Deehr, Rebecca C., and Amy Shumann. "Active Seattle: achieving walkability in diverse neighborhoods." *American journal of preventive medicine* 37.6 (2009): S403-S411.
8. Knopp, Jan, Josef Sivic, and Tomas Pajdla. "Avoiding confusing features in place recognition." *Computer Vision–ECCV 2010* (2010): 748-761.
9. Le, Quoc V. "Building high-level features using large scale unsupervised learning." *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013.
10. Richardson, Andrea S., et al. "One size doesn’t fit all: cross-sectional associations between neighborhood walkability, crime and physical activity depends on age and sex of residents." *BMC public health* 17.1 (2017): 97.
11. Hajna, Samantha, et al. "Neighborhood walkability: field validation of geographic information system measures." *American journal of preventive medicine* 44.6 (2013): e55-e59.
12. Hosler, Akiko S., et al. "Relationship between objectively measured walkability and exercise walking among adults with diabetes." *Journal of environmental and public health* 2014 (2014).
13. Clarke, Philippa, et al. "Using Google Earth to conduct a neighborhood audit: reliability of a virtual audit instrument." *Health & place* 16.6 (2010): 1224-1229.
14. Kim, Hyungjin et al. “A Probabilistic Feature Map-Based Localization System Using a Monocular Camera.” Sensors 2015
15. Kelly, Cheryl et al. “The Assoication of Sidewalk Walkability and Physical Disorder with Area-Level Race and Poverty.”
16. Sheppard, Stephen R.J., Cizek, Petr. “The Ethics of Google Earth: From Spatial Data to Landscape Visualization.”
17. Bennett, Sean & Kirby, Ronald & Macdonald, Blair. (2009). Wheelchair accessibility: Descriptive survey of curb ramps in an urban area. Disability and rehabilitation. Assistive technology. 4. 17-23. 10.1080/17483100802542603.
18. Bromley, Rosemary DF, David L. Matthews, and Colin J. Thomas. "City centre accessibility for wheelchair users: The consumer perspective and the planning implications." *Cities* 24.3 (2007): 229-241.
19. Dreiseitl, Stephan, and Lucila Ohno-Machado. "Logistic regression and artificial neural network classification models: a methodology review." *Journal of biomedical informatics* 35.5 (2002): 352-359.
20. Dean, Jeffrey, Greg Corrado, Rajat Monga, Kai Chen, Matthieu Devin, Mark Mao, Andrew Senior et al. "Large scale distributed deep networks." *Advances in neural information processing systems*, pp. 1223-1231. 2012.
21. Abadi, Martín, et al. "TensorFlow: A System for Large-Scale Machine Learning." *OSDI*. Vol. 16. 2016.