**Review of Existing Research**

**Image Classification Models**

In (3) the authors describe a method for making an assessment of the presence, and then classification to the correct sign, for street signs that would be detected in the environment in real-time. This type of system would be required for an autonomous or assistive vehicle to operate in the environment with other vehicles. They built upon prior research, which focused around shape and color applied to specific areas of the image and sign. Their assessment asserts that when an image fails the first step, either color or shape recognition, it would not proceed to the second step (either shape or color). The authors in this paper propose an AdaBoost based detection system, paired with a Bayesian classification system. The AdaBoost part of the model includes both a color and shape. The model delivered low rates of missing signs (1.4%) and a very low rate of false positives (detecting signs when none was present, 0.03%). Signs were correctly classified in 94% of cases.

In (4) the authors are attempting to classify different images into classes that represent the overall composition of the image, instead of the items within the image. In this case, an image could be a label of rooms in a house (Kitchen, Bedroom, Dining Room), instead of the items in the room (Bed, Refrigerator, Table, Chair). The model described attempts to follow a similar process to a human labeling an image based on the items comprising the scene. In this case, the image is divided into a codebook of parts of the images. The codebook is a random selection of regions of the original image that provide insight into the reoccurring elements that comprise images in different classes. The codebook is further reduced by using k-means clustering to remove low frequency arrangements from the codebook. The advantage of the model shown by the authors is the low degree of supervision required, and the ability to learn classes that might be composites of other classes. The model achieved a 76% accuracy rate when tested with 100 images.

In (8) Knopp et al. attempted to recognize the location depicted in a geo-coded image from a street view type database. Their model uses the geocoding information to reduce the domain of potential classifications of the scene depicted in the image. The database of images consisted of 17,000 images harvested from Google Street View, combined with 8,000 images from the photosharing website Panoramio. Images were concentrated in Paris, with an objective of classifying the place shown in the image. The model used a “Visual bag of words” approach to describe features within the images, with a second step that considers the layout of those features within the image. A confusion score was developed to help show the amount of confusing features that, while recognizable, do not help discriminate the location of an image. The best result provided 47.9% accuracy of classifying locations within Paris.

In (9), Le et al seek to create a classification model without the pretense of a set of labeled objective data. The authors characterized this type of learning as “self-taught” learning that has previously been used to find basic features like blobs and edges. Data in this study was comprised of single frames of video from 10 million different YouTube videos. The best performing neuron achieved 81% accuracy in facial recognition in the sample data. The model incorporates stochastic gradient descent to allow for scaling the processing of images across a cluster of hardware.

In (14) Kim et al describe the use of an alternative to assigning location, instead of use of a Global Positioning System (GPS), instead using image-based localization. Prior research detailed in the paper reflects feature extraction methods based on edge detection and then mapping the features. However, this method was reliant on a high volume of features, and a high accuracy of mapping that would necessitate views from a variety of angles and heights to provide a high certainty of correct assignment of location. This method would not be usable for a robot with a limited singular viewpoint and possibly with distortion from motion. Kim proposes a probabilistic method, based on a data set that has inadequate or inaccurate features. The algorithm described in the paper found lower mean difference from actually position on all three coordinate planes, as well as smaller standard deviations than the previously defined P3P and OPnP.

**Convolutional Neural Networks for Images**

In (5) Goodfellow and a team from Google used the DistBelief method of deep neural networks to train a model to recognize multi-digit numbers from Google Street View. The problem presented is very broad – that street address numbers are represented in a variety of different formats, locations, and with changing conditions from lighting and obstructions. The paper addressed first using the model to recognize numbers from the Google Street View House Number dataset, and then from a more complete Street View image dataset. The recognition of street numbers in images represents a special problem – in order to be useful, all digits in the sequence must be correctly recognized. Images were pre-processed for this paper by “subtracting the mean of each image.” Working in isolation for the Street View House Number dataset, the model was trained over 6 days and achieved 97.84% accuracy (just short of their benchmark of 98% for human recognition.) The Street View House Number data set does have additional pre-processing work to isolate the house numbers and bound the area that the numbers exist within. The more complete Street View data set carried additional challenges of a larger image field to locate the numbers, with more variation in the type and layout of the numbers as well as blurring of some images. The Google Street View data set yielded 91% accuracy. The full image data set used five convolutional layers. For Google, this method has provided scalability to the translation of Street View images to maps: more than 100 million addresses at time of the writing of the paper had been translated to images to mapping. (As a sidenote to this paper, DistBelief was a closed-source effort that was a predecessor to the well-known and open source “TensorFlow” library.)

In (6) Simonyan and Zisserman used convolutional neural networks to classify images for the ImageNet Large-Scale Visual Recognition Challenge. Their neural networks were configured to use a small 3x3 receptive field, versus prior years where fields as large as 11x11 were used. Images were also randomly scaled for training, so that object identification became independent of the size of the object within the image. Training was carried out using logistic regression and gradient descent. Initially, the learning rate was set at 0.01 and was reduced by a factor of 10 as the error rate was not changing further. Their final model used the combined outputs of two neural networks, labeled as multi-crop and dense evaluation and achieved an error rate of 23.7%, better than the comparison set of models from prior years.

**Neighborhood Quality/Infrastructure Surveys**

In (11) Hajna et al. investigated the relationship between information collected and compiled in Geographic Information Systems (GIS) used by government relates to the neighborhood walkability as compiled through participant survey. The study found a good level of statistical correlation (R=0.7) between the GIS and audit based measures of walkability. However, correlation between GIS and participant survey measures of walkability were low (R=0.2). The study was conducted in 200 Montreal, Canada neighborhoods. The authors of the study used the precedent of sanitation and tobacco sales restrictions as proof that changes in environmental design provide significant public health outcomes. In this case, increasing walkability of the neighborhood will lead to higher levels of physical activity and better health over time.

In (13) Clarke et al used a neighborhood audit perfomed via the images on Google Street View in comparison with a neighborhood audit performed by a person physically in the neighborhood surveyed. In this research, City of Chicago Community Adult Health Study collected information via observation in person within the neigbborhoods studied. The reviewer walked each block twice – once on the inside, and once on the outside of the block to draw a complete study of the street and it’s attributes. Subjective measures (like sidewalk quality) had a much lower consistency between the Street View and Physical observation than objective measures (binary attributes of presence of an object). Measures that could be temporary – like the presence of garbage, broken glass and litter on the sidewalk had low consistency between Street View and physical observations, showing that measurement that requires high levels of precision could be hard to obtain with Street View.

In (16) Sheppard and Cizek discuss the ethical ramifications of the use of “virtual globes” as a tool in the process of governing and planning communities. The author references the work of MacEachern to reveal a difference between the use of a visual image for “Visualization”, which may be for exploring or analyzing, and for communication. The authors discuss the benefits to scientists, including helping build engagement with the public through more visually immersive demonstration. The risk of the use of these systems can be the production of misleading visualizations, that may not properly present uncertainty, as well as the limits of the visualization to convey the actual data. The visualization may lack the resolution needed to show the detail.

**Walkability/Public Health Concerns**

In (7) Deehr and Shumann considered walkability of five neighborhoods in Seattle with populations that are usually underrepresented in the city processes. Data employed in their selection method considered health factors, occurrence of motorists striking pedestrians, and modes of transportation used by the area’s residents. Policy changes that were recommended by the project include consideration of walking as a mode of transportation equal to car, bicycle or public transit. The team also recommended changes in design for accommodation of pedestrians. The project also helped start community-led projects to add walking trails to several neighborhoods in Seattle. Finally, the project also led to promotional efforts oriented towards citizens to identify walking trails, and tie together walking with other modes of transit. A master plan for making Seattle the most walkable city in the country was adapted in 2009.

In (10) Richardson, Troxel et al. conducted statistical research to understand whether neighborhood factors associated with green space and walkability had positive results for “Moderate to vigorous physical activity” for the residents of the neighborhood. The research was based on medical recommendations of 150 weekly iminutes of moderate intensity activity, or 75 minutes of vigorous activity. Random selection was used to select Pittsburgh households for participation in the study. The study had 1,015 households included in the study using accelerometers to quantify physical activity over a seven day timeframe. Factors of neighborhood crime, green space, and walkability were controlled for, as well as individual participants’ demographic data. Walkability data was gained from physical audit of the neighborhood. Walkability components included: “traffic signs, pedestrian crossings, sidewalks, lighting, transit and mixed use.” The study found significant differences at the level of p<0.05 for Gender (Female participants walking more), and the inclusion of “Any physical limitation walking 1 block”. Several other variables were found with p-values that placed them just beyond significance, including Walkability index, Education, and Age in Years. There were also significant interactions between walkability, age and gender.

In (12) Hosler and Gallant, et al., researched the relationship between objective walkability scores and walking for exercise among adults. The research was conducted in Albany, NY and Greene and Columbia Counties, New York. Participants in the study were from “medically underserved” populations. Walkability studies were conducted that included scoring of the sidewalk based on criteria of coverage of the street, material of the sidewalk, width of the sidewalk, buffer to street material and size, height of the curb, and presence of curb ramps. Other amenities of the street, its safety, and the amenities of the neighborhood were also considered. Scoring of sidewalks did not have a statistical effect (p=0.198) on “Any Walking”, controlling for demographics of the participants. However, walking greater than 150 minutes per week had a significant influence from Sidewalk scoring (p=0.013). Traffic safety (p=0.011) was the only significant influence for “Any walking”.

In (15) Kelly, Schootman, et al searched for correlation between area-wide demographic measures of poverty and race, with the levels of walkability. The project was conducted in the St. Louis area, with random selection of blocks that identified as either white or African American. The selected areas were audited by physical inspection of the sidewalks, measuring a binary outcome of sidewalks or no sidewalks, and if sidewalks are present on both sides of the streets. Sidewalk quality was assessed then considering levelness,, and condition of the sidewalk (considering the presence of cracks or broken sections. Temporary obstructions were also considered in grading sidewalk quality. Other measures of community quality were also considered, including vacant buildings, graffiti and trash. The research found that more than 1/3 of block groups surveyed had no sidewalks on either side of the street. Of the remainder of streets surveyed (those with sidewalks on at least one side of the street), about one in eight have a “lot of unevenness.” Block groups that were “predominantly” African-American in resident groups were almost 30 times more likely to have sidewalks that had a lot of unevenness than block groups with mostly white residents. Poverty rate was also considered as a variable, but the statistical analysis did not find a large difference in outcome when the model included this variable and did not include this variable.

**Project Sidewalk Predecessors**

Project Sidewalk, a broad scale research project at the University of Maryland, has established several lines of research that will provide a body of work and basis for this project’s extension to machine learning and image classification.

Broadly, Project Sidewalk has used crowd sourced labor to label features on Google Street View related to the accessibility issues. In (1), the researchers had an objective of labeling bus stop attributes by site, using crowd- sourced labor. First the researchers first worked to define requirements for cataloging transit access points for persons with disabilities. After defining the requirements for cataloging site attributes, the team crowd sourced labored to help define all of the physical attributes of each transit access point. The visuals within Google Street View were compared in an audit of 179 sites to ensure that the images in Google Street View were an accurate representation of the actual physical environment. The auditors found that 29/179 surveyed sites were missing from Google Maps data but were found in the physical survey.

Crowd sourced labor for labeling was provisioned through the Amazon Mechanical Turk service using minimally trained users. The training was incorporated for the user as part of the introduction to the work. The interface allowed users to label attributes, specific to bus transit, included signage, shelters, seating and trash disposal bins. The crowd sourced labor was found to have an accuracy of 82.5% in labeling features. Errors tended more in the direction of omission of features. The sample size of sites was 150.

A similar Project Sidewalk team in (2) applied a similar process in another research work. Again, images were sourced from Google Street View to attempt to catalog physical attributes of accessibility to disabled individuals. Again, the images were presented to crowd sourced labor for purposes of cataloging the attributes. The crowd sourced labor this time achieved a high score on a Spearman Rank Correlation. The process was further expanded in this paper with a discussion of a machine learning extension to the project.

A four step process was described. First, images were scraped from Google. Then crowd sourced labeling provided wire framing of locations of curb cuts, and other crowd sourced labeling verified the correctness of the labels. Using the correctly labeled dataset, another process called svDetect uses a “Deformable Part Model” to attempt to identify the features of a training set in another set of images. Another step of a support vector machine was also used to help refine the performance of the model. The performance of the model was limited, reaching less than 80% in recall.

**Sources**

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.”