SMAR 2019 – Fifth Conference on Smart Monitoring,
Assessment and Rehabilitation of Civil Structures

SMAR 2019

Moving Beyond the Romans: Deep Learning and Road Maintenance

Mark DESANTIS [1] and Christoph MERTZ [1]

[1] RoadBotics, Pittsburgh, USA

Contact e-mail: mark@roadbotics.com, christoph@roadbotics.com

ABSTRACT: A central challenge to maintaining roads that has existed since the Romans built the Appian Way over 2,000 years ago is regular, thorough inspection of the roads. Visual inspection was the inspection mode in ancient Rome and it is the most common method now around the world and not just for roads but for all large infrastructure. Fortunately, recent advances in both AI/deep learning and inexpensive but precise sensors has created the opportunity to transform the way infrastructure is monitored and repaired, with the result of lower maintenance costs. Roadbotics uses standard smartphones and computer vision with deep learning to assess road surfaces and roadways. The images from the smartphones are analyzed to find cracks, potholes, and patches and give the road an overall score between 1 and 5. The final results can be loaded into any standard GIS asset management application or displayed to the end-user on an interactive map through a web browser.

1 INTRODUCTION

Roadway maintenance presents a significant issue worldwide. A 2014 report estimated that the United States would need to spend 45.2 billion dollars per year to maintain the existing infrastructure in good conditions, nearly three times the existing expenditure of 16.2 billion dollars per year (Jaffe, 2015). Pavement inspection and monitoring is a critical part of successfully maintaining roadway infrastructure that often places a significant cost and labor burden on public works departments and roadway managers. The predominant current practice is not unlike what the Romans did 2000 years ago: visual surveys, where the Roman chariot has been replaced by a pickup truck. A second type of practice is less common: high-tech, high-cost solutions that require specialized vehicles.

Technological advances have enabled very precise measurements of roadway distresses. For example, ARAN systems can be outfitted with numerous sensors including video, longitudinal laser profiling, transverse laser profiling, ground penetrating radar, and more (ARAN - survey subsystems, 2018). Unfortunately, systems like this require substantial capital investment. This typically means that smaller public works departments, municipalities, and even cities with substantial road networks must get inspections from contracting firms where the costs are passed on by charging hundreds of dollars per kilometer. While the level of precision provided by these technologies is impressive, it is debatable whether it is critical for all applications. Furthermore, the cost burden often forces those municipalities that do decide to get these inspections into a 2- to 5-year inspection cycle, where either the entire network is done every few years or portions of the road network are done each year.

Often the financial burden of technological solutions deters roadway managers who then opt to perform visual surveys. This is especially true of local governments with small public works departments overseeing road networks on the order of 200 km in length. Visual surveys, either done in-house or by contracting with engineering firms, require field crews to drive the entire network to make visual judgements and record their findings about the road condition. Furthermore, these inspections are often done at speeds below the speed limit, with crew member potentially getting out of the vehicle for closer inspection on an active roadway, potentially posing a safety hazard. These inspections are often the most practical for smaller road networks, since they require little to no upfront cost and existing resources (e.g. employees and public works vehicles) can be deployed.

Visual inspections typically follow an established rating methodology such as PASER (Walker, 2002) or Pavement Condition Index (PCI) (Pavement Condition Index 101, 2009). For asphalt surfaces, PASER involves breaking the road network into 1-2 km sections in rural areas and 1-4 block sections in urban areas. Each section is given a score from 1-10 based on the presence and severity of various asphalt distresses, such as cracking, potholes, raveling, bleeding, and rutting. Similarly, PCI typically involves dividing the road network into 1 block sections in urban areas and up to 10 km sections and evaluating the presence and severity of various distresses. Once the presence and severity of each distress type is determined, they are all given a weighted contribution to a 0-100 rating. One issue with visual inspections is that while they attempt to introduce objectivity in determining the overall score, there is a certain level of subjectivity that goes into assessing the severity and overall impact of different distresses. For example, recognizing a large pothole is something that can be done by almost anyone with little to no training, whereas distinguishing five levels of severity in raveling or rutting by visual appearance can introduce disagreement between raters.

Recent advances in computer vision and machine learning, especially with respect to deep neural networks, have empowered numerous applications. Billions of dollars are currently being invested into the sensing capabilities of autonomous vehicles as well as safety systems such as Subaru's EyeSight technology. Similar technologies can be leveraged for automated visual assessments using video data as the basis of a low-cost sensor platform. Varadharajan et al (2014) originally proposed this idea in a paper showing the feasibility of detecting cracks in asphalt using computer vision and video data collected with smartphones. While some degree of capital investment is needed in developing these technologies, their utility can scale extremely well so that the research and development costs can be amortized across many road inspections, ultimately minimizing the costs passed on to roadway managers. In addition to advances in software, consumer demand has driven significant developments in hardware. Now, a typical smartphone contains a powerful processor, high quality camera, GPS sensor, accelerometer, gyroscope, magnetometer, large storage, an intuitive user interface and more.

All these capabilities are available for prices ranging from $200-$800. These technological advancements can be harnessed either as is or as cheaply built, application specific hardware. These hardware devices can be deployed by professional drivers, fleet vehicles (e.g. garbage trucks, or package delivery vehicles), or by existing public works vehicles. Overall, the combination of low-cost data acquisition platforms with the capability and scalability of machine learning poses an opportunity to provide efficient, affordable solutions that strike a middle ground between the precision of current high-tech solutions and lower-cost visual inspections.

2 LOW-COST SENSOR PLATFORMS

Hardware advances in the past few years have gained huge strides, primarily driven by consumer applications, such as social media. Consumer devices like a typical smartphone, dash-cameras such as those made by BlackVue, or various action cams such as the GoPro Hero line all offer relatively high-quality camera sensors, accelerometers, and global positioning for only a few hundred dollars. Smart phones in particular not only come with sensors such as gyroscope and magnetometer, but are also programmable, typically with robust software development kits (SDK). Furthermore, some models have been equipped with stereo cameras or even built-in thermal imaging. These capabilities are packaged in a small form-factor, capable of storing an entire day's worth of data collection on an SD card. Overall, these devices have formed the foundations of a low-cost, mobile sensor platform capable of acquiring the necessary data for road inspection. Data acquisition pipelines can be built by deploying professionally trained drivers using custom applications, free or paid crowd-sourcing, attaching hardware to existing municipally owned vehicles, or leveraging existing vehicle fleets. The latter poses an interesting opportunity, since there are garbage trucks, street sweepers, mail delivery vehicles, school buses, public transportation, and delivery trucks that already drive over nearly every street at predictable time intervals. In addition, many of the newer consumer vehicles are equipped with video and GPS capabilities that are paired with data connections. Overall, there are ample opportunities to build a low-cost sensor platform adequate for road inspection.

3 DETECTING DISTRESSES WITH DEEP LEARNING

Applications of convolutional neural networks (CNN) for computer vision have exploded recently, including systems for autonomous vehicles (e.g. Huval et al, 2015), processing satellite imagery (e. g. Chen et al, 2014), automating medical diagnoses (e. g. Esteva et al, 2017), and much more. Applications are typically classified as semantic segmentations or object identification. Semantic segmentation involves classifying each pixel of the image as one label in predetermined set of classes (e.g. sidewalk, person, pavement, etc). Object identification involves detecting the presence of a predetermined set of objects and additionally giving their location in the form of a bounding box. More recent approaches have combined the two approaches to identify the presence of objects and output their pixel level segmentation, such as with mask R-CNN (He et al, 2017).

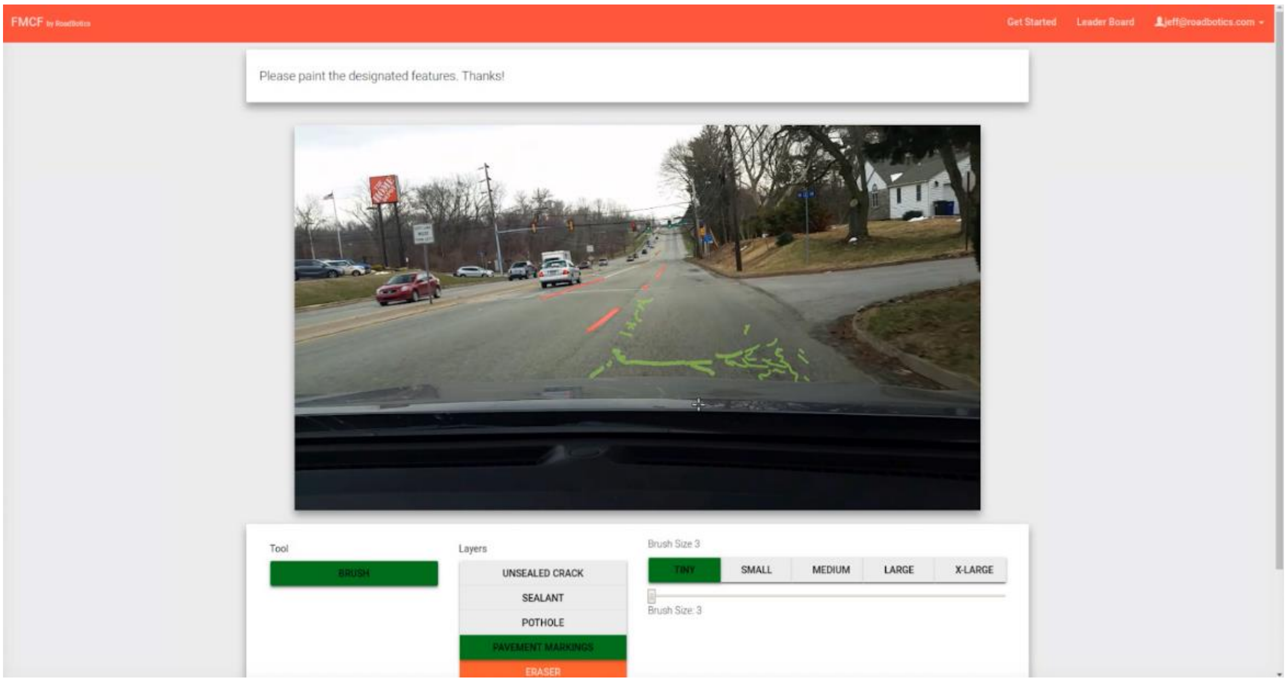

Figure 1. Annotation of distresses by painting over them using custom web-based labeling platform.

We have developed a semantic segmentation model to identify pavement distresses. This type of model is appropriate, since in most cases the pavement distresses do not make up discrete objects (e.g. raveling, cracking). Using a custom-built labeling platform (see Figure 1), trained labelers annotated selected distresses for video frames by "painting" over top of the distresses with a brush tool. Training the model to identify distresses then essentially becomes a process of teaching the model to recreate the "paintings" of the human labelers. In more technical terms, the annotations by human labelers are converted to binary masks (arrays of 0s and 1s) for each distress type with values of 1 denoting the presence of that particular distress for a given pixel. The model then learns to map RGB color images (MxNx3 array where MxN is the image resolution and the last dimension is the 3 colors) to a probability map for each class (MxNxP array where P is the number of classes). Figures 2 and 3 show example model predictions for various classes.

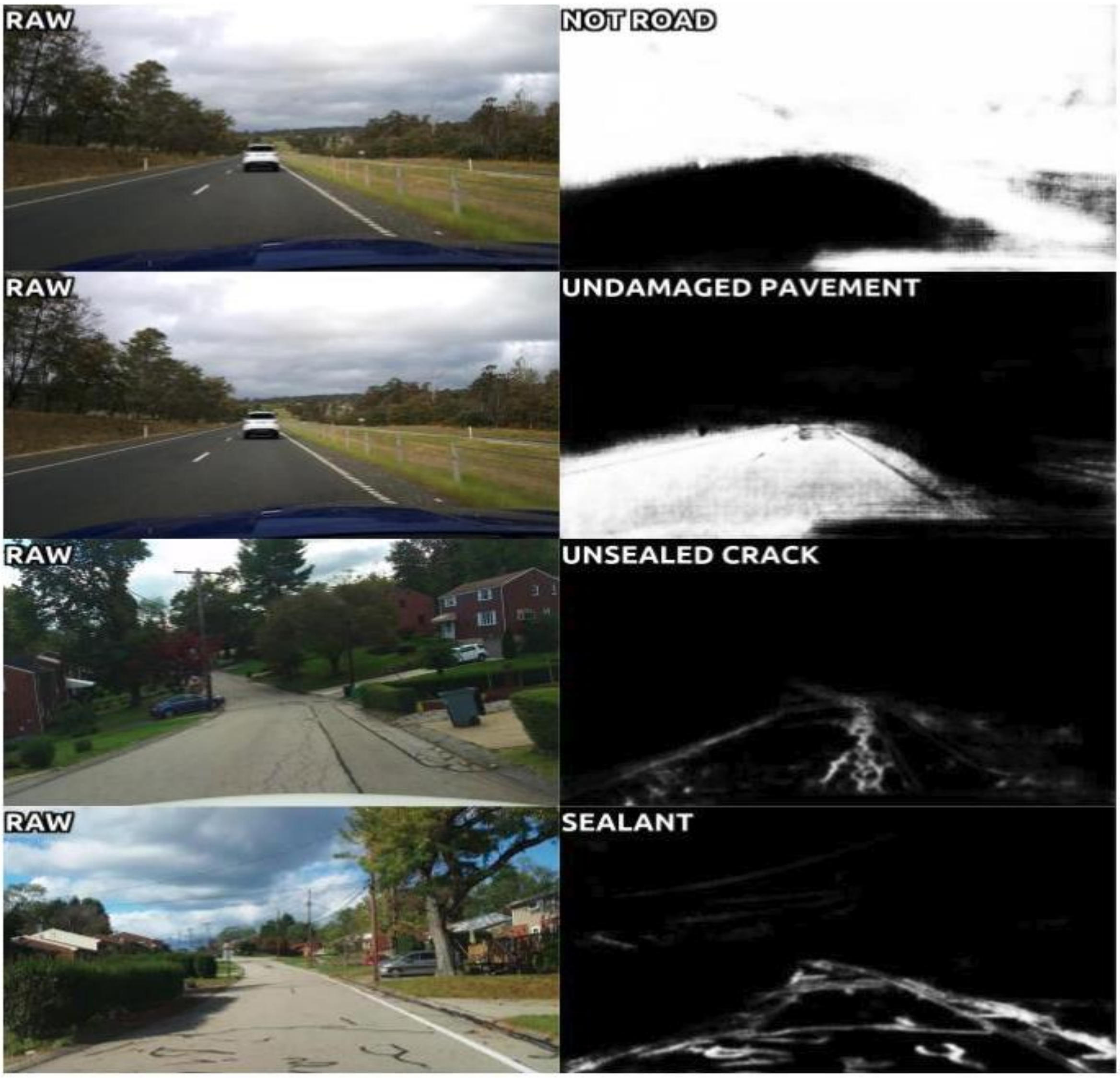

Figure 2. Selected segmentations from our deep learning model. Left column contains raw video frames. Right column contains model output where grayscale value indicates probability (i.e. black = 0%; white=100%).

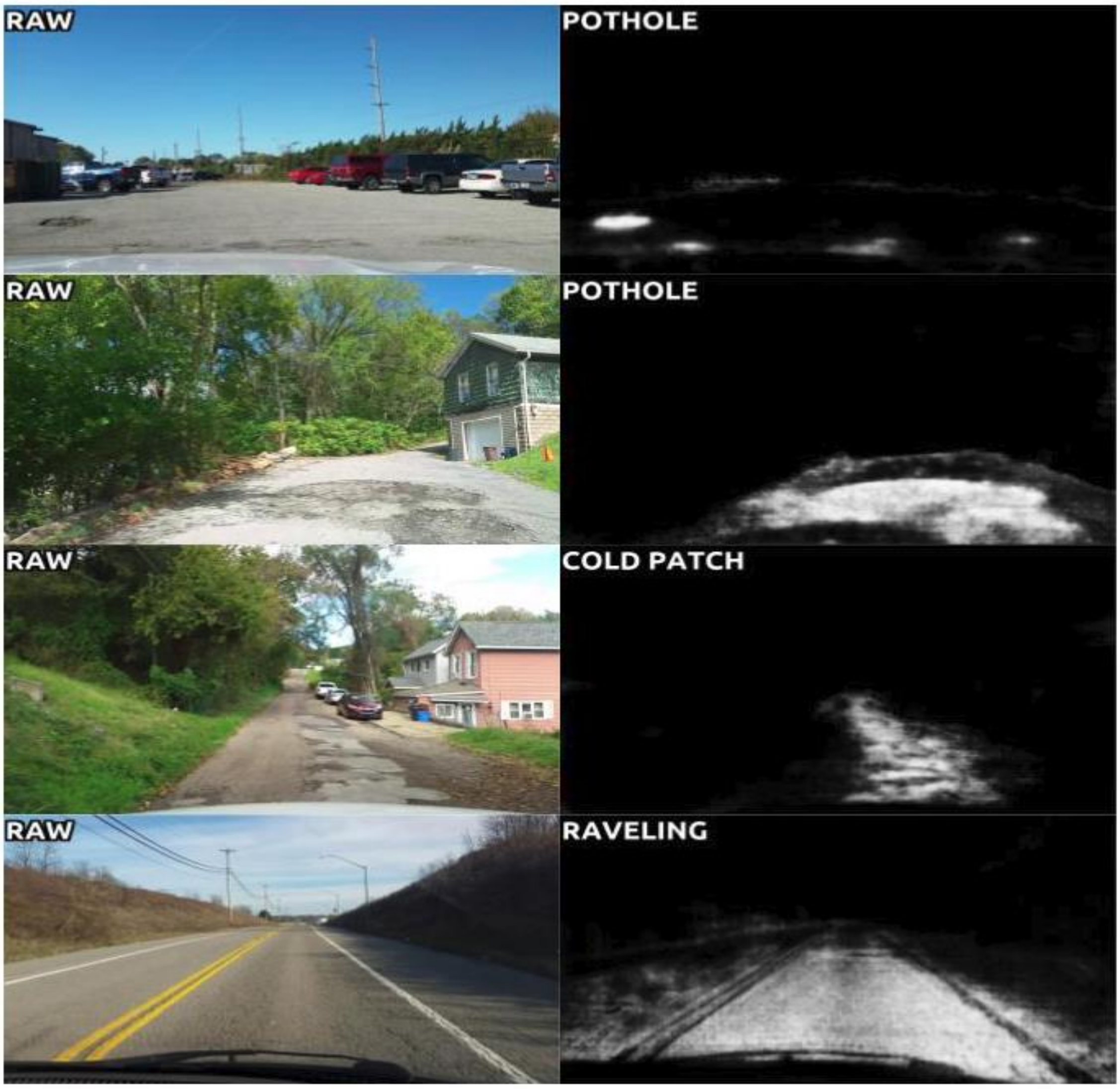

Figure 3. Selected segmentations from our deep learning model. Left column contains raw video frames. Right column contains model output where grayscale value indicates probability (i.e. black = 0%; white=100%).

4 AUTOMATED ROAD INSPECTION SYSTEM

We have developed a commercially available automated road inspection system. Using a custom developed application, we collect video and GPS data using Android smartphones mounted to the windshield with standard suction cup mounts. This setup is depicted in Figure 4. A second phone is used to navigate prepared turn-by-turn routes to ensure coverage of the desired region. Professionally trained drivers are deployed to drive the prepared routes. After collection, our application uploads video and GPS data to Google Cloud Platform (GCP), where individual frames are extracted at 3-meter intervals and indexed by time and spatial coordinates for efficient lookup.

Figure 4. An android phone running custom software records video, GPS, and accelerometer data. A second phone is used for turn-by-turn navigation (not shown).

Once the data have been uploaded and processed, scaled deployment of our deep learning model starts. Frames are processed in parallel by up to 96000 computational workers (1000 server instances with 96 computational threads each). We typically use fewer workers (~1000), in which an entire municipality (around 200 km) can be processed in a couple hours. Aggregate statistics are calculated on the segmentation masks per frame, which are combined into an overall 1-5 score, similar to the calculation of PCI. Example images for each rating are shown in Figure 5.

Figure 5. Example images demonstrating our overall 1-5 ratings.

After all required frames have been rated, the output enters the final stage of the processing pipeline where all the point data (video frame with GPS coordinates, rating, and other metadata) are localized to Open Street Maps roads that have been divided into intersection-to-intersection segments. Average ratings are calculated on a per segment basis, and the final results for point and segment data are sent to our web-browser-based visualization platform for delivery to the end-user. Screen-shots of network level visualization are shown in Figure 6.

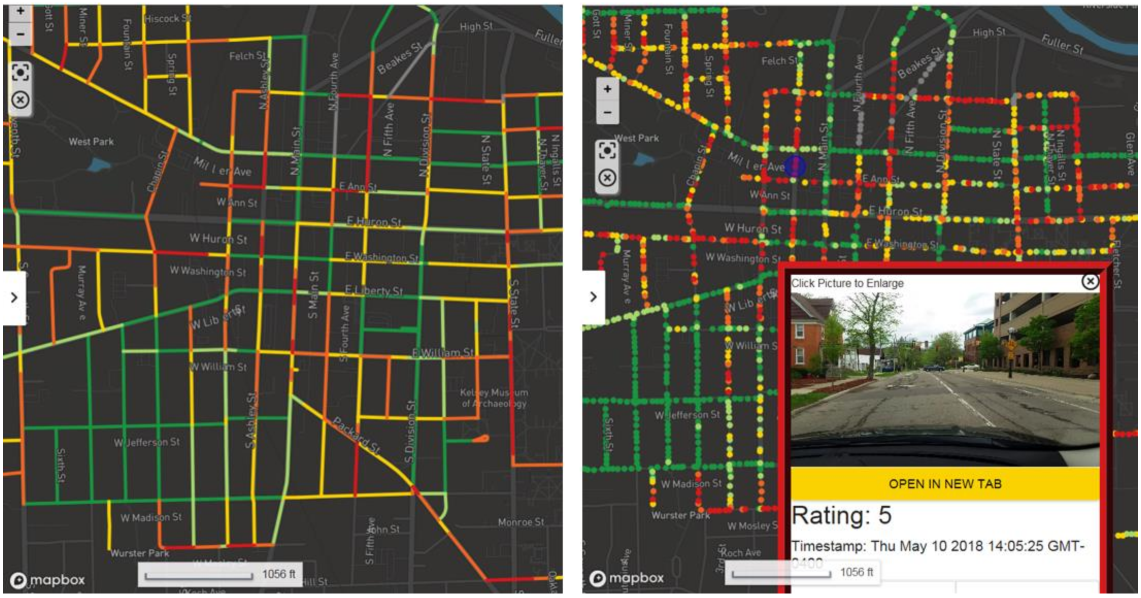

Figure 6. Network level display of ratings. The left image shows ratings that have been averaged on intersection-to-intersection road segments. The right image shows individual images that are sampled every 3 meters along the road. Users can click the points or segments to display additional information, such as to view the associated image.

5 CONCLUDING REMARKS

We have demonstrated the successful implementation of a fully automated, AI-based pavement inspection system. This system requires minimal capital investment in hardware by taking advantage of low-cost, readily available smartphones as our sensor platform. The use of cloud computing means that the system can scale efficiently. There are, however, ample opportunities to expand capabilities. For example, consider two roadway types which are not paved surfaces but are nevertheless commonplace: brick roads and gravel roads. The former are common in residential neighborhoods, as well as scenic downtowns and departments of public works do spend money to assess and maintain them. They pose a further level of challenge to AI-based assessment as they are more difficult to visually assess in the first place and visual features that could be indicative of damage in a paved surface (e.g. long cracks), may actually be normal features (e. g. the indentation separating each brick). Gravel roads, by contrast, tend to be more frequent in rural areas, but often of equal or greater importance, because gravel roads are more often than not the roads by which important capital assets are serviced. Examples include wind farms, power plants, and factories. Future efforts will focus on improving the current capabilities as well expanding them to other surface types, such as gravel and brick roads.

6 CITATIONS AND REFERENCES

ARAN - Survey Subsystems (2018). Fugro Roadware. Retrieved from: http://www.roadware.com/products/survey_equipment/aran_subsystems/

Chen, X., Xiang, S., Liu, C. L., & Pan, C. H. (2014). Vehicle detection in satellite images by hybrid deep convolutional neural networks. *IEEE Geoscience and remote sensing letters*, 11(10), 1797-1801.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017, October). Mask r-cnn. In Computer Vision (ICCV), 2017 *IEEE International Conference on (pp. 2980-2988). IEEE.*

Huval, B., Wang, T., Tandon, S., Kiske, J., Song, W., Pazhayampallil, J., ... & Mujica, F. (2015). An empirical evaluation of deep learning on highway driving. *arXiv preprint arXiv:*1504.01716.

Jaffe Eric (2015, February 12) America's Infrastructure Crisis is Really a Maintenance Crisis. Retrieved from: https://www.citylab.com/solutions/2015/02/americas-infrastructure-crisis-is-really-a-maintenance-crisis/385452/

Pavement Condition Index 101 (2009, December). *OGRA's Milestones.* Retrieved from: http://hawaiiasphalt.org/wp/wp-content/uploads/PCI-101.pdf

Varadharajan, S., Jose, S., Sharma, K., Wander, L., & Mertz, C. (2014, March). Vision for road inspection. In *Applications of Computer Vision (WACV), 2014 IEEE Winter Conference* on(pp. 115-122). IEEE.

Walker, D., Kummer, S., & Entine, L. (2002). Pavement Surface Evaluation and Rating; PASER Asphalt Roads Manual. *Transportation Information Center University of Wisc*