Automatic Detection of Road Damage

Anca Dragulescu, Nathan Einstein, Nicholas Stern, Claire Stolz
Institute for Applied Computational Science | Daimler AG/Lab 1886

Problem Statement

In collaboration with Daimler AG's innovation lab, Lab 1886, we investigated the requirements for exploiting images from on-board vehicle cameras to automate the tedious process of detecting and evaluating the severity of road damage. Using paint damage as a case study, we explored alternative forms of data annotation, multiple segmentation models and pipelines, and prototyped interactive visual software. These severity classifications are intended to help German municipalities decide which roads to prioritize for repairs.

Data & Annotation

The data we received were obtained from smartphone cameras mounted on a car dashboard, taking snapshots of German roads. The data consist of 27,500 high-res (1080 x 1920) images from 38 distinct trips across varied landscapes.

To detect and assess road damage in an image, we need annotated examples with which to train a model. Due to the laborious nature of the task, we attempted to crowdsource the annotations through Amazon's Mechanical Turk platform.

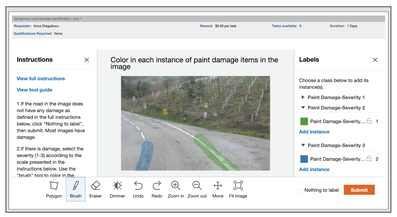


Figure 1: Mechanical Turk task interface.

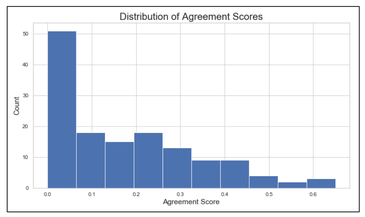
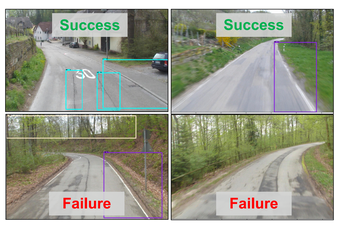


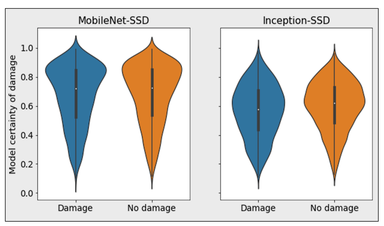
Figure 2: Distribution of annotation agreement.

Results point to extremely inconsistent annotations, therefore our team annotated approximately 1,300 images for training and evaluating our models.

Domain Adaptation

Previous work suggests that with sufficiently large, annotated datasets, deep neural networks can perform well for this task. In particular, Maeda et al.[1] achieved 75% accuracy on a multi-class dataset of road damage in Japan. We first applied the authors' pretrained models to our dataset, with mixed results:





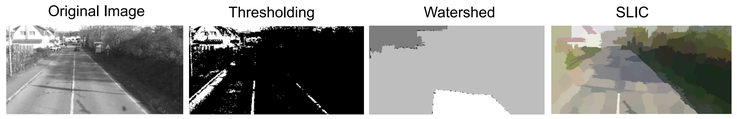
Unfortunately, the Japanese dataset was too different from the German dataset to perform sufficiently well. The violin plots above illustrate that the model was equally confident in its correct and incorrect predictions.

Multi-Stage Approach

1. Segmentation

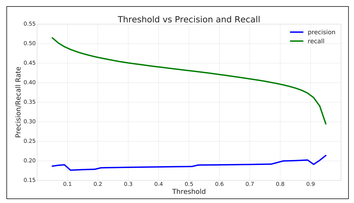
Traditional Computer Vision

- **Thresholding:** segment image into contiguous pixel groups that fall within a certain pixel value range (grayscale or color)
- **Watershed algorithm:** finds connected elements and sort by depth (pixel value)
- **Simple Linear Interactive Clustering (SLIC):** finds representative pixels for particular regions and enforces boundaries



Deep learning

Going beyond traditional computer vision techniques, we trained a UNet[2] semantic segmentation model to identify all paint damage (of any severity).



The model is only moderately successful at identifying damaged paint, as indicated by the low pixel precision (< 20%). Adjusting the threshold of the model improved precision slightly at the cost of recall.

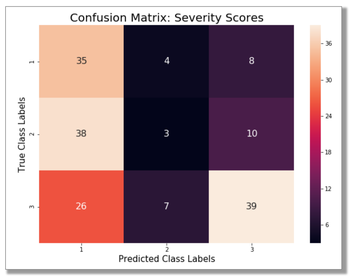
2. Severity Classification

After segmenting out the relevant area of a raw image, we built a CNN to classify the severity of paint damage on a scale of 1 to 3 representing low to high.

Figure 3: Example of how input was generated

Results reveal that severity class 1 and 2 are nearly indistinguishable. The scarcity of data and noisy labels limit model performance, even with pretrained convolutional weights.

Sev. 1 accuracy	Sev. 2 accuracy	Sev. 3 accuracy
74.5%	5.9%	54.2%

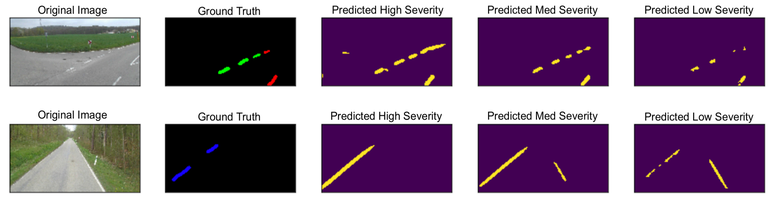


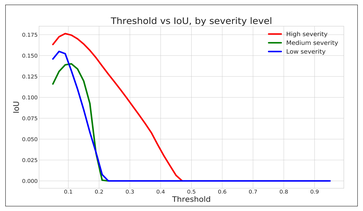
Unified Approach (Multiclass Segmentation)

A *multiclass segmentation* model classifies each pixel in an image. We considered multiclass segmentation as an end-to-end model that combines damage detection and severity classification. We trained two different model architectures, UNet and PSPNet, to classify pixels as one of four classes, shown in the key on the right.

Pixel Key

- background
- low severity
- medium severity
- high severity





The adjacent figure illustrates the high degree of uncertainty in the predictions of the multiclass segmentation model, especially for low- and medium-severity damage. Above a 20% certainty threshold, the model ceases to predict any damage overlapping with the ground-truth labels for those two classes.

Predicted Severity*

Ground Truth Severity	High only	Med only	Low only	Multiple severities	No damage predicted	Total
High	138,153	4	26,397	400,856	793,524	1,358,934
Med	75,518	3	19,040	283,999	338,639	717,199
Low	50,533	9	43,769	323,907	615,815	1,034,033
None	397,376	5	126,599	590,219	166,693,523	167,807,722
Total	661,580	21	215,805	1,598,981	168,441,501	

*Prediction threshold of 0.1

Conclusions

1. Annotating road damage severity consistently is complicated by the inherent subjectivity of the task. This is evidenced by how both the multi-stage and unified models can discern low- and high-severity classes, but rarely predict the medium-severity class.
2. More data is needed to conclusively determine whether multi-class segmentation or multi-stage classification performs better, but preliminary results are promising in that our models do learn to detect paint and classify low/high severity damage.
3. Quantitative evaluation of segmentation models is nuanced. First, comparison against ground-truth masks must be qualified by the amount of unintentional noise and the intentional context provided (i.e. buffer around true damage) in the annotations. Furthermore, the relative importance of precision and recall must be weighted by how conservative the end-user wants the model to be.

References

1. H. Maeda, Y. Sekimoto, T. Seto, et al., Road damage detection using deep neural networks with images captured through a smartphone, *Comput. Aided Civ. Infrastruct. Eng.*, 33 (2018), 1127-1141.
2. O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, *MICCAI*, (2015), 234-241.