Using Computer Vision to Streamline City Code Enforcement

Ben Harris

VRNE.co

ben@vrne.co

Abstract

This paper proposes an approach to streamline reactive and proactive maintenance of an entire city. Over five weeks, data from two census tracts were captured weekly. Using a web application, two city code enforcement officers reviewed each reactive or proactive data point. The reviewed data points were then used to measure the efficacy of our approach.

1. Introduction

Cities are required to manage large land areas and customer populations. These customers will typically report tens of thousands of issues a year. Some of the top complaints reported are tall grass and weeds, abandoned vehicles, dilapidated structures, and trash. Using computer vision is helpful for reactively and proactively managing these issues.

For this study, two census tracts were selected based on recommendations from our trial city, Norman, OK. This resulted in 34 lane miles being driven weekly (Fig 1.) One to nine data points were captured per parcel (home). As a result, 3,000-10,000 data points were captured and processed each week. Each data point was processed through a computer vision algorithm with the results stored in a database.

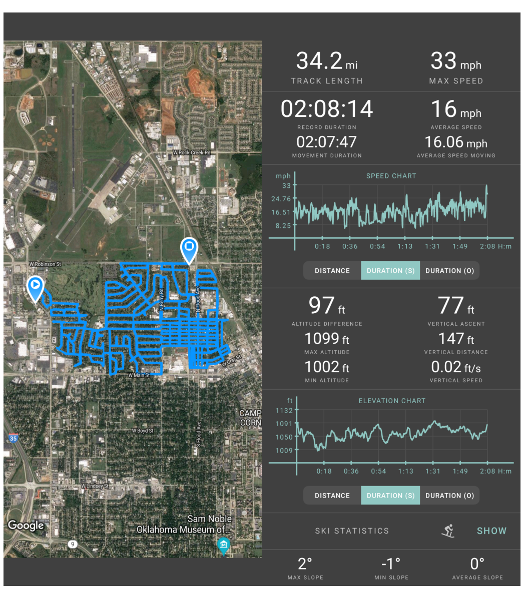


Figure 1. Driving Statistics

On top of the proactive data captured through the computer vision algorithm, the City of Norman provided a daily feed of reactive issues (customer complaints) by a CSV each morning. An automated workflow processed, filtered, and inserted the data into the database.

With both proactive and reactive data being stored in the database, the first objective was to measure how well the captured data worked for reactive issues. This removed the computer vision noise and gave a baseline efficacy. To capture qualitative data, a custom application was developed to allow the review of each data point. The application was provided to the City of Norman through two tablets.

For the first two weeks of the study, two city code enforcement officers used the application to review each reactive complaint. Once enough reactive data had been collected, the proactive data was added to the application for the remaining two weeks. To minimize added workload, a limit of thirty-six data points were added each week. A weight based on issue frequency was added to each proactive data point and used to select the top 36 data points.

Through the application, the following qualitative questions were answered by the code enforcement officers. *Was the data helpful? Is the vision algorithm correct? Did the application provide enough information to make a correct decision?* In addition, the data allowed quantitative measurement of the questions: *What percentage of reactive issues were automatically identified by the computer vision? How long did it take to complete the tasks? How long did it take to review an issue?* Both sets of questions helped clarify the utility of the application.

2. Field Data Collection

The data was collected using two 8 MP Sony cameras connected to an embedded system. Every image captured was downsized from 4k to 1080p resolution. Custom code was written to collect the images, direction, speed, latitude, longitude, and timestamp. Each data point was sent to a real time queue, which allowed an application to process each data point, determine which parcel the data point was within, and save the data to a database. This allowed each device to be tracked real time, as well as to verify the data is collected for each parcel.

Due to feedback from Norman code enforcement officers, multiple data collection approaches were used throughout the trial. During the first two weeks, a 60 degree field of view lens was used to save one image per parcel. In the third week, service roads were added and three images were saved per parcel. For the fourth and fifth weeks, a 175 degree field of view lens was used. In addition to the new lens, three photos per road bordering a parcel were saved.

3. Computer Vision

After data was collected each week, the selected images were processed through a computer vision model using a rented GPU. The models were trained using various data preparation techniques and vision architectures. See [3] for the paper on training the model and Fig 2. for our current true positive rates. Note: Fig 2. reports findings with the background predictions removed to provide a clearer interpretation. The number of photos being processed per parcel increases the likelihood of a detection, leading the background detections to decrease.

Due to the imbalanced nature of the code enforcement violations, there are very few examples of violations when compared to normal classes. One class example would be abandoned cars versus normal cars. This imbalance leads to high false negative rates. A solution to this problem is to train a separate model to classify the difference between a vehicle and an abandoned vehicle. Another solution would be collecting and annotating much more data. In the course of this pilot, option one greatly improved the results. However, this solution wasn't applied to Fig 2. results.

Architecture	Date	Tents	Trash	Graffiti	Broken Window	Tall Grass	Abandoned Car
Mask r-cnn[1]	2021	88.9	66.7	96.6	15.9	36.1	1.7
Mask r-cnn[1]	2/19/2022	0	84.3	0	12.3	55.9	0
CBnet2-SWIN [2]	2/21/2022	0	83	0	38	62.68	40
CBnet2-SWIN [2]	2/22/2022	0	80	0	57	71	49
CBnet2-SWIN [2]	2/27/2022	54	91	86	40	69	48
CBnet2-SWIN [2]	3/1/2022	82	92	94	37	78	43

Figure 2. Class True Positive Rates

4. Qualitative Data Collection

The goal of this technology is to augment human ability, not to fully automate code enforcement. To account for human interpretation, it was important for code enforcement officers to provide feedback. An app was designed to merge the human perspective with technology.

The application was designed to allow a code enforcement officer to reactively or proactively review each parcel. As reactive data was logged, the new issues were added to the app for review the following morning. The final two weeks of the trial, the proactive tasks were added Monday morning.

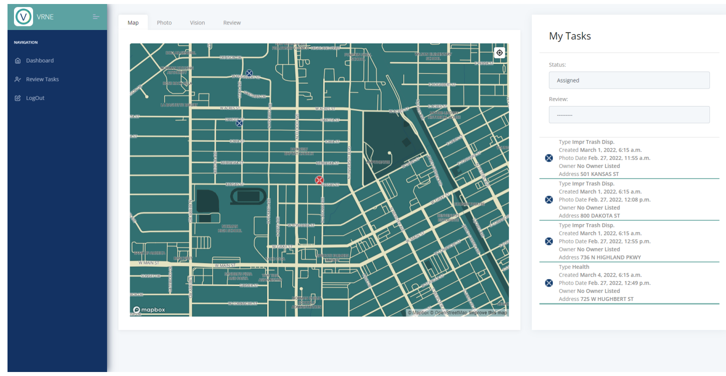


Figure 3. Application

Once the task was added, a code enforcement officer selected an individual assigned task. The officer then reviewed each photo of the property under the Photo tab. Next, the officer reviewed the Vision tab to check if the computer vision model identified an issue correctly. Second to last, a form with three qualitative questions was answered under the Review tab (Fig 4). Finally, once in the field (i.e.: at a parcel/home), the officer opened the app and selected the In Progress tasks. A final form was completed to close the task (Fig 5).

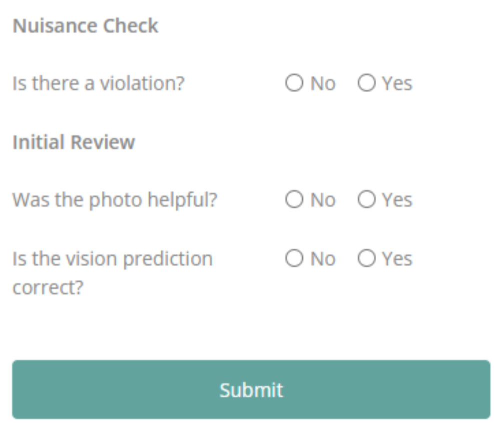
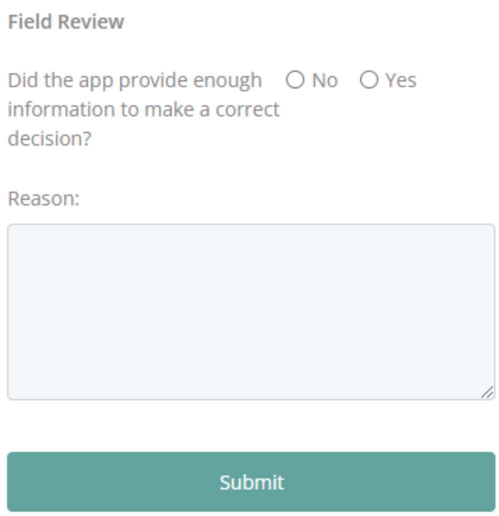


Figure 4. Initial Review Form


Field Review

Did the app provide enough information to make a correct decision? No Yes

Reason:

Submit

Figure 5. Field Review Form

Each qualitative question was designed to measure a specific outcome. “Is there a violation?” measured the percentage of tasks containing a true violation and also determined how many of the tasks were false violations. “Was the photo helpful?” provided feedback on whether photos provided enough information to make a decision. The final question in the first form, “Is the vision prediction correct?” measured whether the computer vision was working correctly. In the second form, the question: “Did the application provide enough information to make a correct decision?” answered if the app provided enough information to make the correct decision. The reasoning behind their answer could be entered through a free text field.

Varied perspectives were expected due to the qualitative nature of the questions. As part of onboarding, it was communicated there was no wrong answer and officers were asked to be as honest as possible. It was also communicated that the goal of the trial was to perform research and to learn from their feedback.

Finally, a weekly meeting was scheduled to elicit feedback and allow for real time problem solving. Each week, their suggestions were noted and reviewed. If a suggestion was a simple addition or fix, it was completed before the next weekly meeting. This resulted in several changes and improvements throughout the trial.

5. Results

Over the five week trial, 86 tasks were completed; 44 proactive and 42 reactive. Forty-two tasks were not completed or incomplete. This resulted in a significant percentage of the data being excluded.

Using the completed data, the percentage of endorsement of each qualitative question was calculated. The first level reviewed was overall effectiveness. Per completed ratings, the data within the application could be used to make a complete, correct decision 65% of the time. The computer vision algorithm predicted a violation correctly 43% of the time. Finally, 42% of the tasks were true violations. See Fig 6.

Tasks	% Correct Decision	% Valid Prediction	% Violation
86	65.00%	43%	42%

Figure 6. Overall Effectiveness

It is clear there is a difference between reactive and proactive task outcomes. Once again, the additional level of information is helpful, but doesn’t tell the whole story. Further analysis of proactive and reactive task groups provides additional information. See Fig 7.

Type	Tasks	% Correct Decision	% Valid Prediction	% Violation
Proactive	44	86%	34%	23%
Reactive	42	43%	52%	62%

Figure 7. Type Effectiveness

The final analysis includes the adjustments made each week of the trial. This micro-level data allows for better interpretation (Fig 9.). For the first two weeks of the trial, one photo per parcel led to an overall lower app decision effectiveness (31%). Adding more photos per parcel in week three led to a significant increase (~32%) in app decision effectiveness. This increase remained consistent throughout the remainder of the trial.

When a chi-squared test is applied to the correct decision column, a p-value of .044 is obtained, reinforcing that there is an improvement between the one and multiple photos approach. See Fig 8. for the comparison between one and multiple photos.

Photos Saved	Type	Tasks	% Correct Decision	% Valid Prediction	% Violation
One	Reactive	26	31%	46%	58%
Multiple	Reactive	16	63%	63%	69%

Figure 8. Multiple Photos Effectiveness

Week	Type	Tasks	% Correct Decision	% Valid Prediction	% Violation	Changes
5	Reactive	17	29%	35%	47%	60 Degree Lens
6	Reactive	9	33%	67%	78%	60 Degree Lens
7	Reactive	11	64%	55%	64%	Added more photos
8	Proactive	25	80%	40%	20%	175 Degree Lens
8	Reactive	1	100%	100%	100%	175 Degree Lens
9	Proactive	19	95%	21%	26%	175 Degree

						Lens
9	Reactive	4	50%	75%	75%	175 Degree Lens

Figure 9. Weekly Change Effectiveness

Another difference between one and multiple photos is the valid prediction percentage. While it appears to increase by 17%, when applying the chi-squared test, a *p*-value > .05 was obtained. This shows there is not a significant difference between collecting one or multiple photos and a valid prediction percentage.

The quantitative data captured by the computer vision model and application allow further review. Each week the computer vision model flagged parcels for review. This flag allowed the determination of the percentage of the reactive cases flagged correctly by the algorithm. Over the trial, 58% of the parcels were correctly flagged by the computer vision model. See Fig 10.

Parcels Predicted Correctly by Algorithm	Parcels reported by Community	% Flagged Correctly
28	48	58%

Figure 10. Flagged Correctly

The remaining quantitative data captured relates to efficiency. It shows the code enforcement officers took almost three days to complete a task once assigned. Also, once the initial review form [Fig. 4] had been completed, it took an additional two minutes to complete the final form. From testing, a total of three minutes per task should be expected. If the qualitative components used for the purposes of this trial were removed, it is estimated to take around 15-30 seconds to complete a task.

Statistic	Result
Median Days to Complete Task	2.625
Median Minutes to Complete Review	2

Figure 11. Time Statistics

For reference, the code enforcement officers reported that they each spend four to five hours in the field each day and, using their current system, it takes ten minutes to fill out a violation.

6. Conclusion

With most medium and large cities using a reactive approach to code enforcement, this approach should save these cities significant resources. The results show that 63% of reactive issues can be accurately and completely reviewed using our technology, greatly reducing the number of field visits. It could save even more resources if the application was provided to the city call center to screen each customer issue.

While the results prove the proactive technology automatically identifies most of the reactive issues, the technology will need time and additional data to become more accurate. It is an iterative process that will only improve with time.

However, this technology is valuable to most medium and large cities due to the ability to quickly review each flagged parcel. Using VRNE's application, a task takes seconds instead of minutes to review. It is estimated this approach is 20 to 40 times (10 minutes/(.25-.5 minutes)) faster than current solutions. On top of the task time savings, there will be a significant reduction in the costs associated with vehicles, gas, emissions, and driving time in the field.

References

[1] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).

[2] Liang, T., Chu, X., Liu, Y., Wang, Y., Tang, Z., Chu, W., ... & Ling, H. (2021). Cbnetv2: A composite backbone network architecture for object detection. *arXiv preprint arXiv:2107.00420*.

[3] Harris, J. (2022). Instance Segmentation Techniques for City Code Enforcement.