Graffiti Detection Efficient Design with Comprehensive Evaluation

Analytics Everywhere Lab AT THE UNIVERSITY OF NEW BRUNSWICK

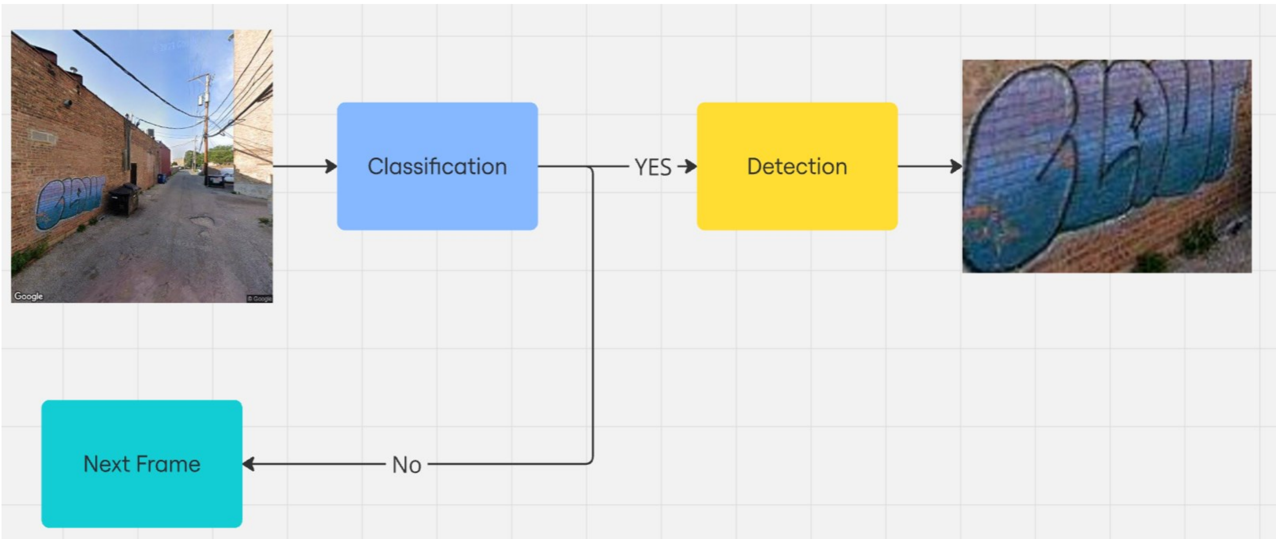
Maxwell Jianzhou Wang[1], Hung Nguyen[1], Rene Richard[1], Hung Cao[1]

[1]Analytics Everywhere Lab, University of New Brunswick

Abstract

Graffiti is not only an eyesore and costly to remove, it presents a hazard to enthusiasts, obstructs signage, causing liability for property owners and transportation agencies. This work enhance the safety and well-being of individuals in public spaces by designing a handful of efficient graffiti detection tools with less computing. We also provide a comprehensive evaluation to select recent popular lightweight detectors based on our needs.

Design



Input: Image frames of street scenes

1. Go to Classification Stage if
   1.1 Output is Yes, go to Detection Stage
   1.2 Output is No, go to next Frame
2. Go to Detection Stage
   2.1 Locate Graffiti
   2.2 Project detection bbox on the pointcloud to obtain depth information

Graffiti Dataset

Dataset Name	Number of Graffiti Imgs	Number of Non-Graffiti Imgs
17kGraffiti	8693	0
Graffiti.v1i.voc	3056	0
graffiti_streetview	799	19839
INDIGO	460 (Dissimilar)	0
STORM_Graffiti	1021	0

Classification Model Training and Evaluation

Dataset	Train Images	Val Images	Test Images
Number of Graffiti Imgs	7572	2163	1083
Number of Non-Graffiti Imgs	13886	3967	1985

Model	Accuracy	Total Process plus Inference Time on Acer
Optimized Xception	0.9971	0.1134s
MobileNet V4 (MNv4)	0.9867	0.0965s

Detection Model Training and Evalution

Num of train imgs	Num of val imgs	Num of test imgs	Img size
2250	540	204	640 x 640

Model	Precision	Recall	mAP@0.5	mAP@ 0.5:0.95	Preprocess Time	Inference Time
NanoDet-m			0.529	0.251	9ms	13.5ms
YOLOv7m	0.6991	0.4786	0.5181	0.2946	8ms	27ms
YOLOv7-tiny	0.4578	0.2641	0.2556	0.1176	2.1ms	5.9ms
YOLOv8s	0.581	0.345	0.376	0.195	0.9ms	3.3ms
YOLOv8n	0.472	0.345	0.34	0.164	1.0ms	1.2ms
YOLOv9s	0.578	0.35	0.376	0.184	0.3ms	4.6ms
YOLOv9t	0.446	0.334	0.32	0.158	0.8ms	2.5ms
YOLOv10s	0.522	0.37	0.375	0.194	0.9ms	4.6ms
YOLOv10n	0.416	0.335	0.397	0.146	1.1ms	1.7ms
YOLO11n	0.416	0.335	0.397	0.146	1.1ms	1.7ms
YOLO11s	0.55	0.339	0.355	0.182	0.5ms	3.2ms
YOLO12n	0.468	0.321	0.299	0.144	1.0ms	2.5ms
YOLO12s	0.509	0.352	0.364	0.175	0.4ms	5.1ms

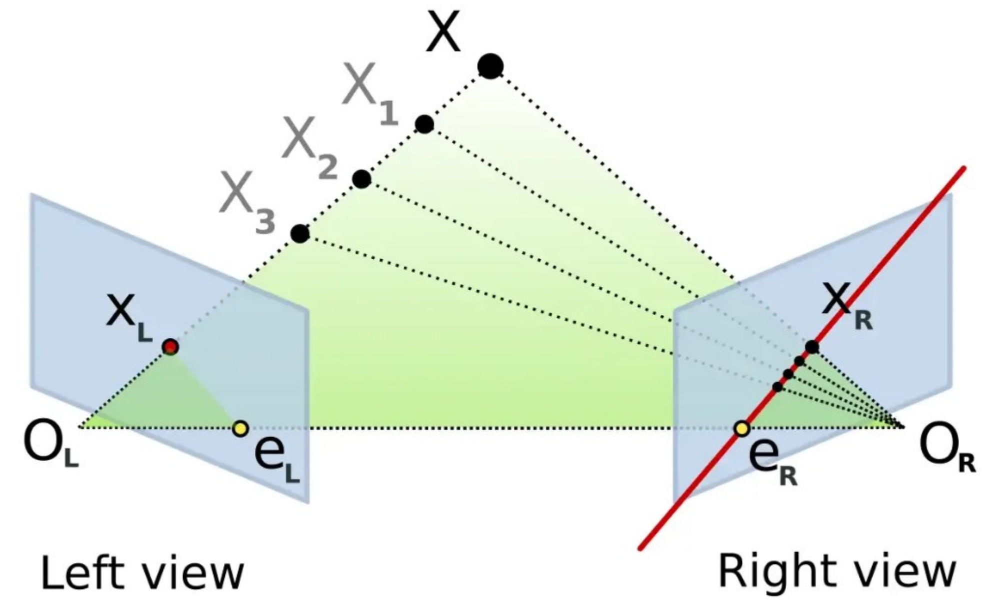
Evaluation Samples

Summary of Classification and Detection

- Optimized Xception [1] is a lightweight model created by Atah Nuh Mih et al. from AELab in 2024, which can sufficiently detect PCB defects and small objects. MobileNet V4[2] (MNv4) is also a lightweight model that can efficiently run on a small edge device and also provides high-accuracy performance. From the table on the left, we can tell the Optimized Xception has roughly 1% higher accuracy than MNv4, but the total process time, including inference time, is 16ms longer.

- NanoDet [3] is a Fully Convolutional One-Stage Object Detector (FCOS)-style anchor-free object detection which uses Generalized Focal Loss as classification and regression loss. YOLO v7 [4] and YOLO v8, v9, v10, v11, and v12 [5] are one-stage anchor-free detectors that have been developed rapidly especially become lightweight, well performed on both speed and accuracy also easily be deployed on small edge devices. As we can see, higher mAP requires more inference time (NanoDet-m(medium), YOLOv7m(medium)), while others (YOLO v8 --- v12) are either the Nano version or the Tiny version. Even these Nano and Tiny version detectors provide a few milliseconds inference times, but the mAP@0.5 is at least 12% lower than NanoDet and YOLOv7m.

Logic of Depth Estimation



Step of project 2D detection into 3D space and obtain depth information:

1. Obtain the camera's intrinsic and extrinsic.
2. Run inference of graffiti detectors.
3. Depth map alignment: using camera SDK to align depth and RGB images.
4. Retrieve Depth Information: for each detection result, prefer the bounding box center and extract the depth value from the depth map.
5. 2D to 3D projection: use camera intrinsics and depth value to project the 2D point into 3D space.

Problems and Limitations

1. Bounding box annotations of the dataset are very causal, i.e. bounding boxes are not aligned to graffiti, which makes the detection model hard to regress precisely.
2. Lack of diversity: only 1 out of 5 graffiti public datasets have annotated bounding boxes.
3. Insufficient number of images: we only have roughly 400 original labeled graffiti images, then applied data augmentation techniques to expand to 2250 training images.
4. Maximum epoch of training is set to 100 for all the detection training due to lack of time and device, so many of them are not converged. Since all the hyperparameters are set equals, the training and evaluation results still can reveal their performance objectively.

References

[1] Mih, Atah Nuh, Alireza Rahimi, Asfia Kawnine, Francis Palma, Monica Wachowicz, Rickey Dubay, and Hung Cao. "Achieving Pareto Optimality using Efficient Parameter Reduction for DNNs in Resource-Constrained Edge Environment." The 37th Canadian AI Conference, Guelph, ON (2024).

[2] Qin, D., Leichner, C., Delakis, M., Fornoni, M., Luo, S., Yang, F., ... & Howard, A. (2024, September). MobileNetV4: universal models for the mobile ecosystem. In European Conference on Computer Vision (pp. 78-96). Cham: Springer Nature Switzerland.

[3] R. Lyu, "Nanodet-plus: Super fast and high accuracy lightweight anchor-free object detection model," URL: https://github.com/RangiLyu/nanodet, 2021

[4] Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 7464-7475).

[5] https://github.com/ultralytics/ultralytics

Supported by the UNB-FCS Startup Fund (22–23 START UP/H CAO)

hcao3@unb.ca