Capturing Active Transportation Counts at Intersections Using Ultralytics YOLOv8 Image and Video Tagging

Work by Alicia Hopper, Tammy Lee, and Sirisha Kothuri
Presented by Alicia Hopper

Introduction

"State of the Art" - How are things done today?

- Manual Counts
- Counter Devices
- Some machine-learning

Sirisha Kothuri, Krista Nordback, Andrew Schrope, Taylor Phillips, and Miguel Figliozzi. Bicycle and pedestrian counts at signalized intersections using existing infrastructure: Opportunities and challenges. 2644(1):11–18.

Manual Counts

- Most common method
- Needs:
  - People
  - Lots of time
  - Ideally a video camera

Mara Chagas Diogenes, Ryan Greene-Roesel, Lindsay S. Arnold, and David R. Ragland. Pedestrian counting methods at intersections: A comparative study. 2002(1):26–30.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
8	Time	Using Crosswalk	Right	Thru	Left	Using Crosswalk	Right	Thru	Left	Using Crosswalk	Right	Thru	Left	Using Crosswalk	Right	Thru	Left
9	6:00 AM	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
10	6:05 AM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	6:10 AM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	6:15 AM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	6:20 AM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	6:25 AM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	6:30 AM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	6:35 AM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	6:40 AM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	6:45 AM	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
19	6:50 AM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	6:55 AM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	7:00 AM	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
22	7:05 AM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	7:10 AM	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
24	7:15 AM	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0
25	7:20 AM	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
26	7:25 AM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	7:30 AM	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
28	7:35 AM	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
29	7:40 AM	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
30	7:45 AM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	7:50 AM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
32	7:55 AM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33	8:00 AM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
34	8:05 AM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
35	8:10 AM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
36	8:15 AM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
37	8:20 AM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
38	8:25 AM	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
39	8:30 AM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
40	8:35 AM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
41	8:40 AM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Images from Tammy Lee

Counting Devices

- $$$
- Most counting devices have focused on cars
- Can be inexact
- Need:
    - Purchase and install device in each location
        - $$$, people, time

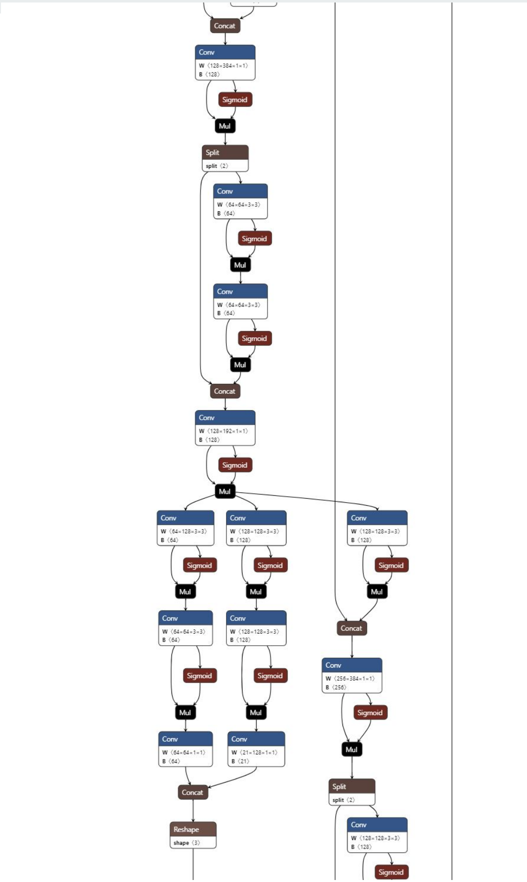
Ryan Greene-Roesel, Mara Chagas Diogenes, David R. Ragland, and Luis Antonio Lindau. Effectiveness of a commercially available automated pedestrian counting device in urban environments: Comparison with manual counts.

Machine Learning - Computer Vision

- Offers a promising alternative
- Lots of options available
- Research focuses on self-driving vehicles
- Needs:
    - Video Camera
    - Trained Model
    - Computer

Sanjukta Ghosh, Peter Amon, Andreas Hutter, and André Kaup. Reliable pedestrian detection using a deep neural network trained on pedestrian counts. In 2017 IEEE International Conference on Image Processing (ICIP), pages 685–689. ISSN: 2381-8549.

WikileaksIntern, CC0, via Wikimedia Commons



netron.app used for model visualization

Choosing a Model

- YOLO models feature in research
- Latest model: yolov8
- sizes: n, s, m, l, x
- Model summary: 225 layers, 11,143,727 parameters, 11,143,711 gradients, 28.7 GFLOPs

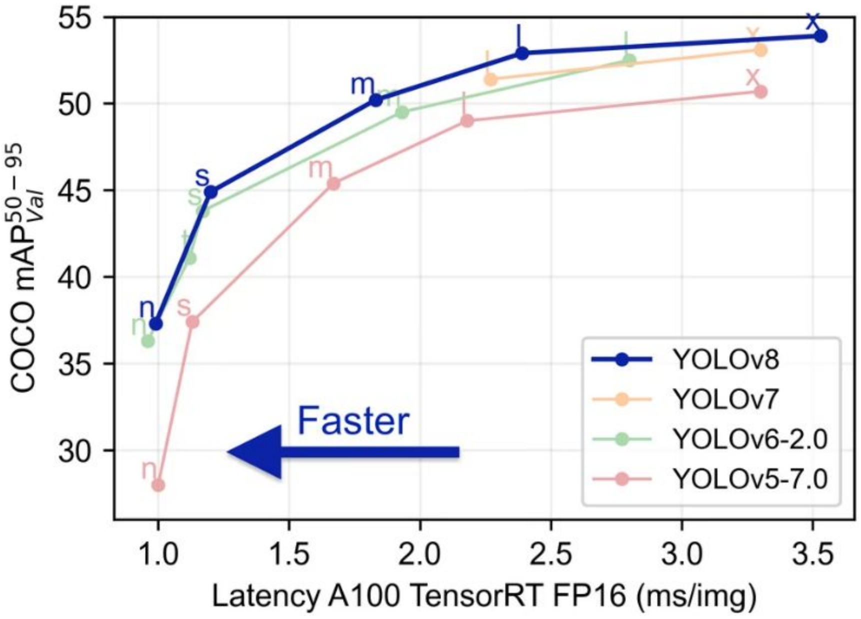
Yuhan Wang and Han Yang. Multi-target pedestrian tracking based on YOLOv5 and DeepSORT. In 2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), pages 508–514.

Performance

Detection (COCO) Detection (Open Images V7) Segmentation (COCO) C

See Detection Docs for usage examples with these models trained on COCO, which include

Model	size (pixels)	mAP^{val} 50-95	Speed CPU ONNX (ms)	Speed A100 TensorRT (ms)
YOLOv8n	640	37.3	80.4	0.99
YOLOv8s	640	44.9	128.4	1.20
YOLOv8m	640	50.2	234.7	1.83
YOLOv8l	640	52.9	375.2	2.39
YOLOv8x	640	53.9	479.1	3.53



Images from https://docs.ultralytics.com/models/yolov8/#performance-metrics

Training a YOLO model

- Python
- Training Data
    - Images
    - Annotations
- GPU

EuroCity Persons Dataset

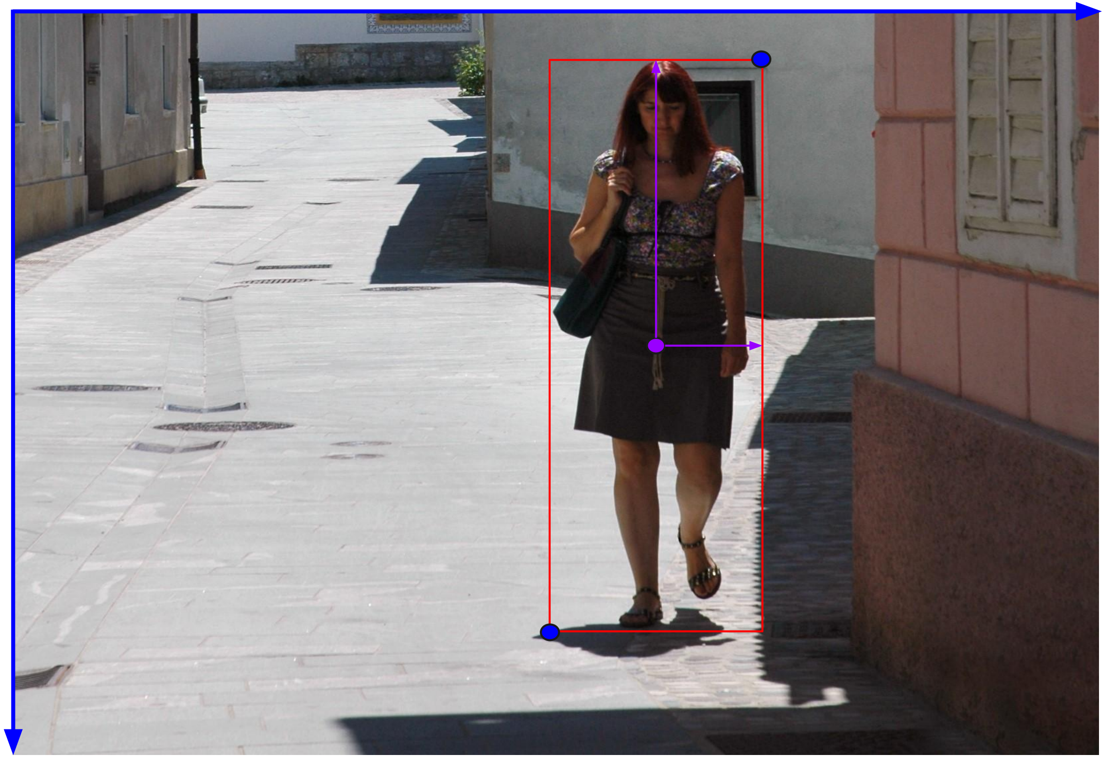
- 40217 total daytime images
- 183004 total daytime pedestrians found in images
- 18216 total daytime riders found in images
- Pedestrian, rider + bicycle, scooter, wheelchair, tricycle, motorbike, buggy, co-rider

Markus Braun, Sebastian Krebs, Fabian B. Flohr, and Dariu M. Gavrila. Eurocity persons: A novel benchmark for person detection in traffic scenes. IEEE Transactions on Pattern Analysis and Machine Intelligence, pages 1–1, 2019.

Converting the Data

```
{
 "tags": [],
 "imageheight": height,
 "imagewidth": width,
 "children": [
  {
   "tags": [],
   "x0": x0,
   "y1": y1,
   "y0": y0,
   "x1": x1,
   "children": [],
   "identity": "pedestrian"
  },
```

```
0 x_center y_center x_dif
y_dif
```

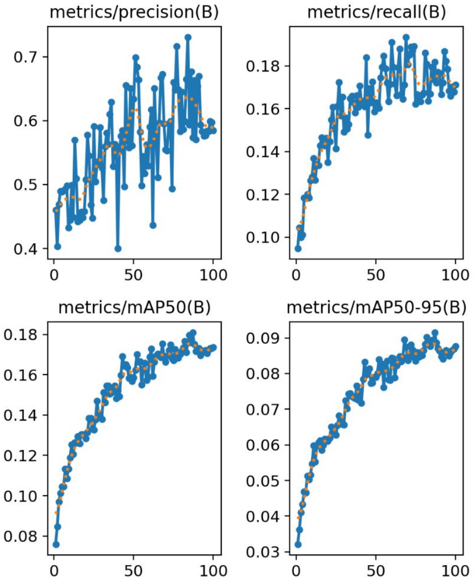


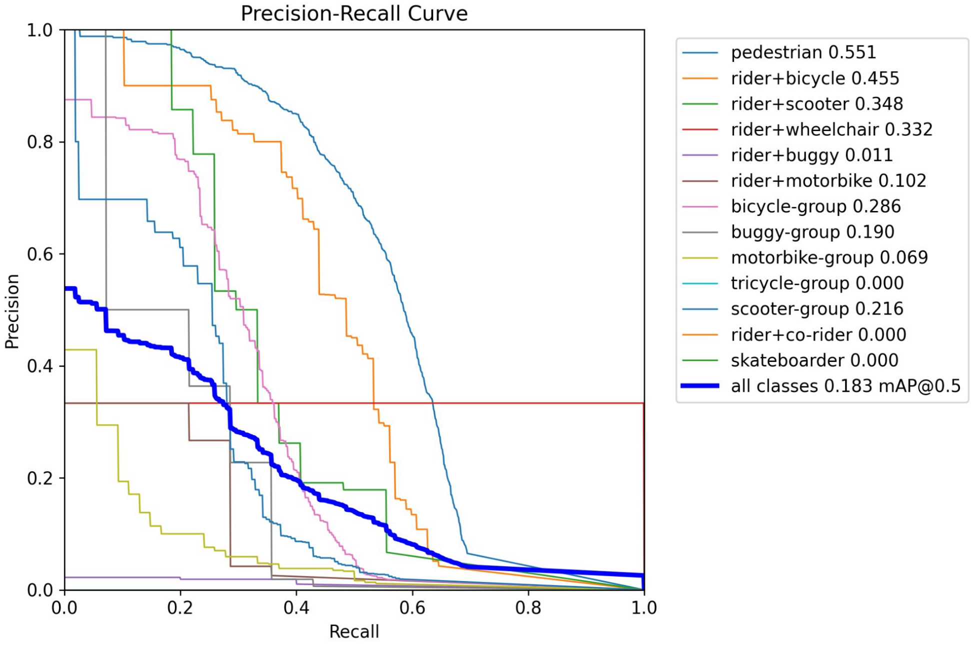
DimiTalen, CC0, via Wikimedia Commons. Annotations added by Alicia Hopper

Images used to train

- Daytime
- Not “far-away”
- Not occluded
- 6 batches

Training

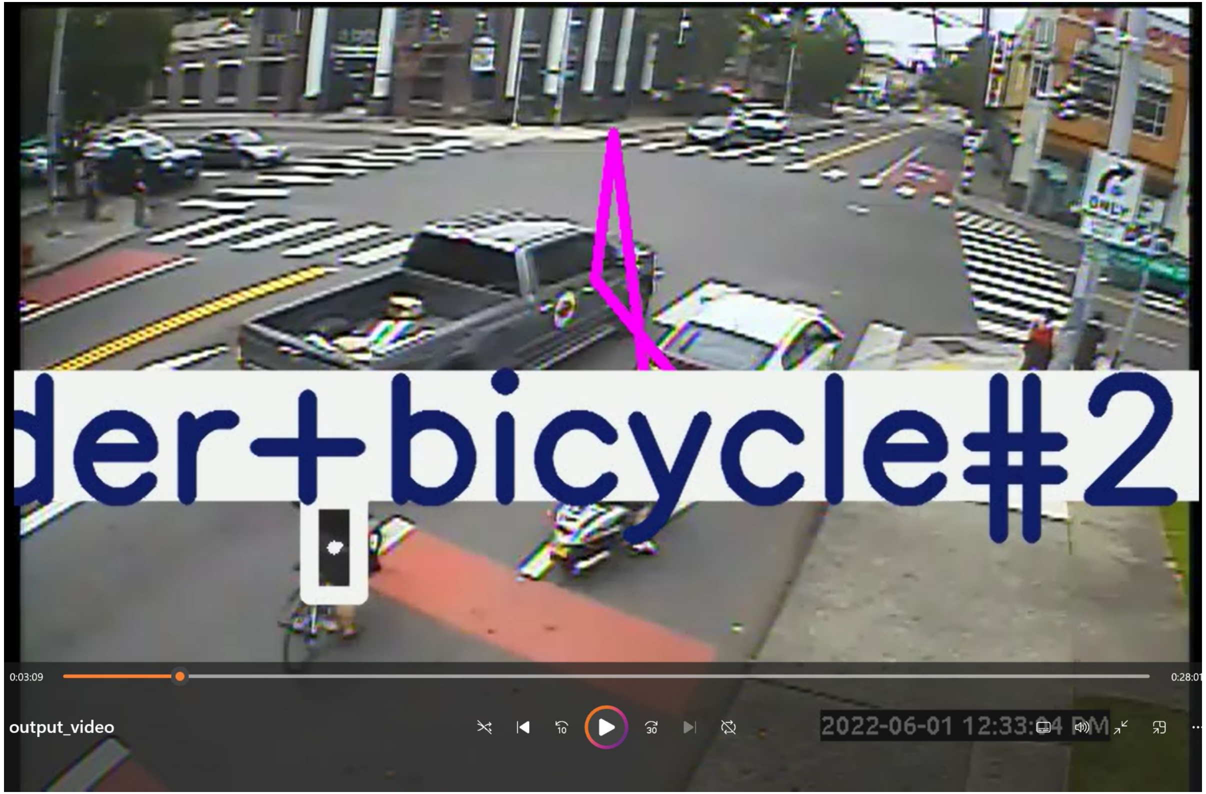

metrics/precision(B)
metrics/recall(B)
metrics/mAP50(B)
metrics/mAP50-95(B)


Precision-Recall Curve
Precision
Recall
pedestrian 0.551
rider+bicycle 0.455
rider+scooter 0.348
rider+wheelchair 0.332
rider+buggy 0.011
rider+motorbike 0.102
bicycle-group 0.286
buggy-group 0.190
motorbike-group 0.069
tricycle-group 0.000
scooter-group 0.216
rider+co-rider 0.000
skateboarder 0.000
all classes 0.183 mAP@0.5

Using the Model

- Input video
- Select regions to count in (crosswalks)
- Python Program
- Output

Demo



Results and Conclusion

- Pedestrians at 0.551 mAP50
- Rider and Bicycle at 0.455 mAP50
- Using video data, expecting higher level of accuracy

Further Work

- Train on more difficult data
- More testing and validation
- Other models?
- Outside of Intersections
- https://github.com/a-hopps/2024-PSU-REU

Thanks for Listening!
Any Questions?