STRIDE: Street View-based Environmental Feature Detection and Pedestrian Collision Prediction, Supplementary Material

Cristina González[1,2]* Nicolás Ayobi[1,2]* Felipe Escallón[1,2]
Laura Baldovino-Chiquillo[3] Maria Wilches-Mogollón[2] Donny Pasos[4] Nicole Ramírez[2]
Jose Pinzón[3,6] Olga Sarmiento[3] D. Alex Quistberg[3,5] Pablo Arbeláez[1,2]

[1]Center for Research and Formation in Artificial Intelligence, Universidad de los Andes, Colombia
[2]School of Engineering, Universidad de los Andes, Colombia [3]School of Medicine, Universidad de los Andes, Colombia
[4]School of Economics, Universidad de los Andes, Colombia [5]Dornsife School of Public Health, Drexel University, USA
[6]School of Architecture and Design, Pontificia Universidad Javeriana, Colombia

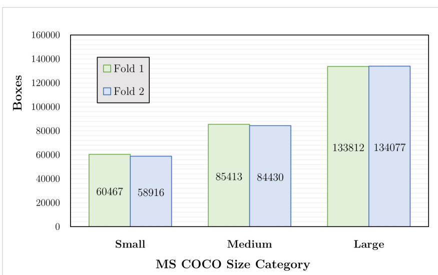

Figure 1: **Distribution of bounding box absolute areas according to MS COCO's standard.** Most of our boxes are considered large objects with MS COCO's standards. This distribution is due to frequent classes that naturally represent large street objects like trees, sidewalks, or curbs. Nevertheless, most instances have very small relative areas due to the large size of our images.

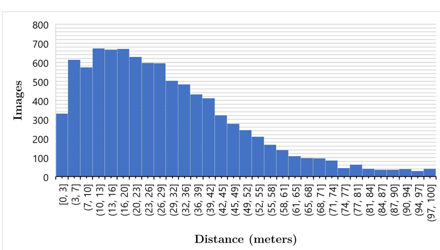

Figure 2: **Distribution of distances from sampled points to the closest crossing point in the training set.** More than 50% of the sampled images were matched to a crossing point within a 30-meter radius. The number of images decreases as the distance value increases.

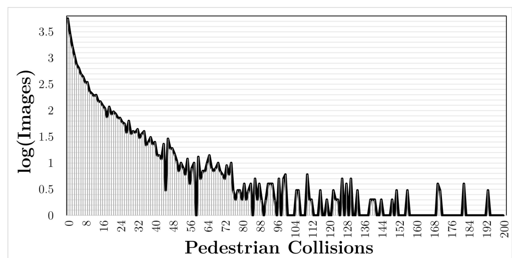

Figure 3: **Logarithmic distribution of pedestrian collisions in the entire dataset.** Our dataset presents an evident long-tail distribution as most sampled points have low collision incidence. We observe that most of our dataset has less than 100 collisions, and the majority of our images have between 0 and 10 collisions.

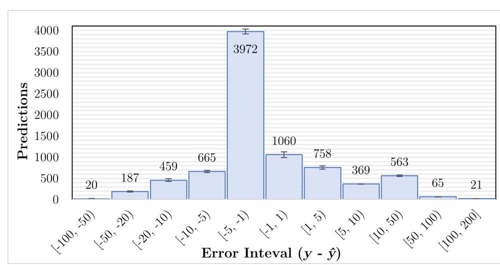

Figure 4: **Distribution of the Exact Error $(y - \hat{y})$ between ground truths (y) and predictions $(\hat{y})$ for the testing set.** The figure shows the number of predictions for each exact error range. The majority of predictions exhibit errors between -5 and 5. Furthermore, the model demonstrates a tendency towards slight overestimations, as the majority of the predictions exhibit negative errors relative to the ground truth values.

*Equal contribution

Images with <u>Zero</u> Collision Frequency

Figure 5: **Example annotations of images with zero pedestrian collisions.** The images shown correspond primarily to residential areas with small amounts of built environment objects.

Images with <u>Intermediate</u> Collision Frequency (20-60 Collisions)

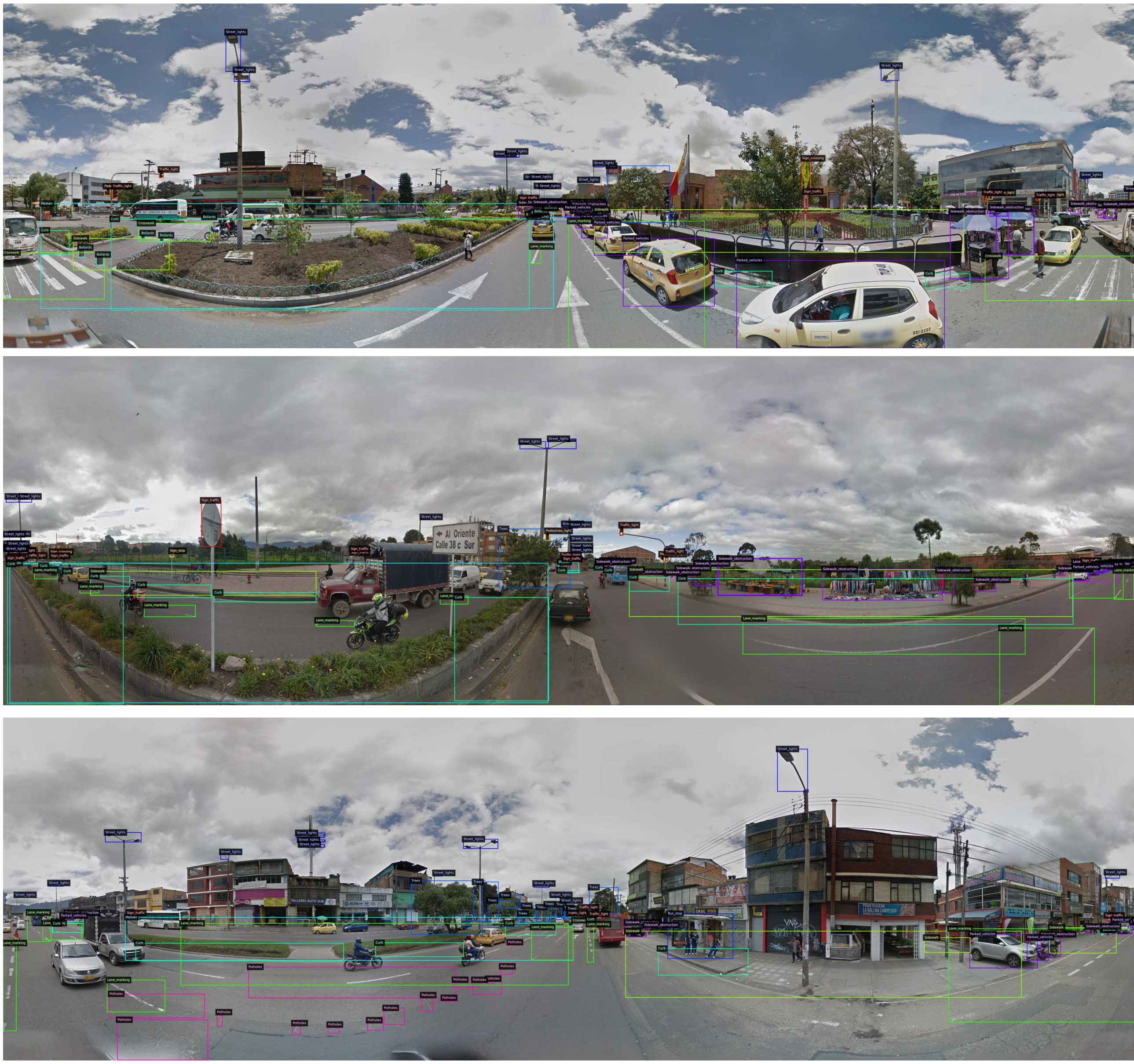

Figure 6: **Example annotations of images with intermediate amounts of pedestrian collisions.** The images shown correspond mostly to regular streets with multiple sidewalk obstructions and parked vehicles (top and middle), or potholes (bottom).

Images with High Collision Frequency (>100 collisions)

Figure 7: **Example annotations of images with high amounts of pedestrian collisions.** The images shown correspond to large highways with multiple visible objects (all), sidewalk obstructions (top and middle), and poor street markings (bottom).

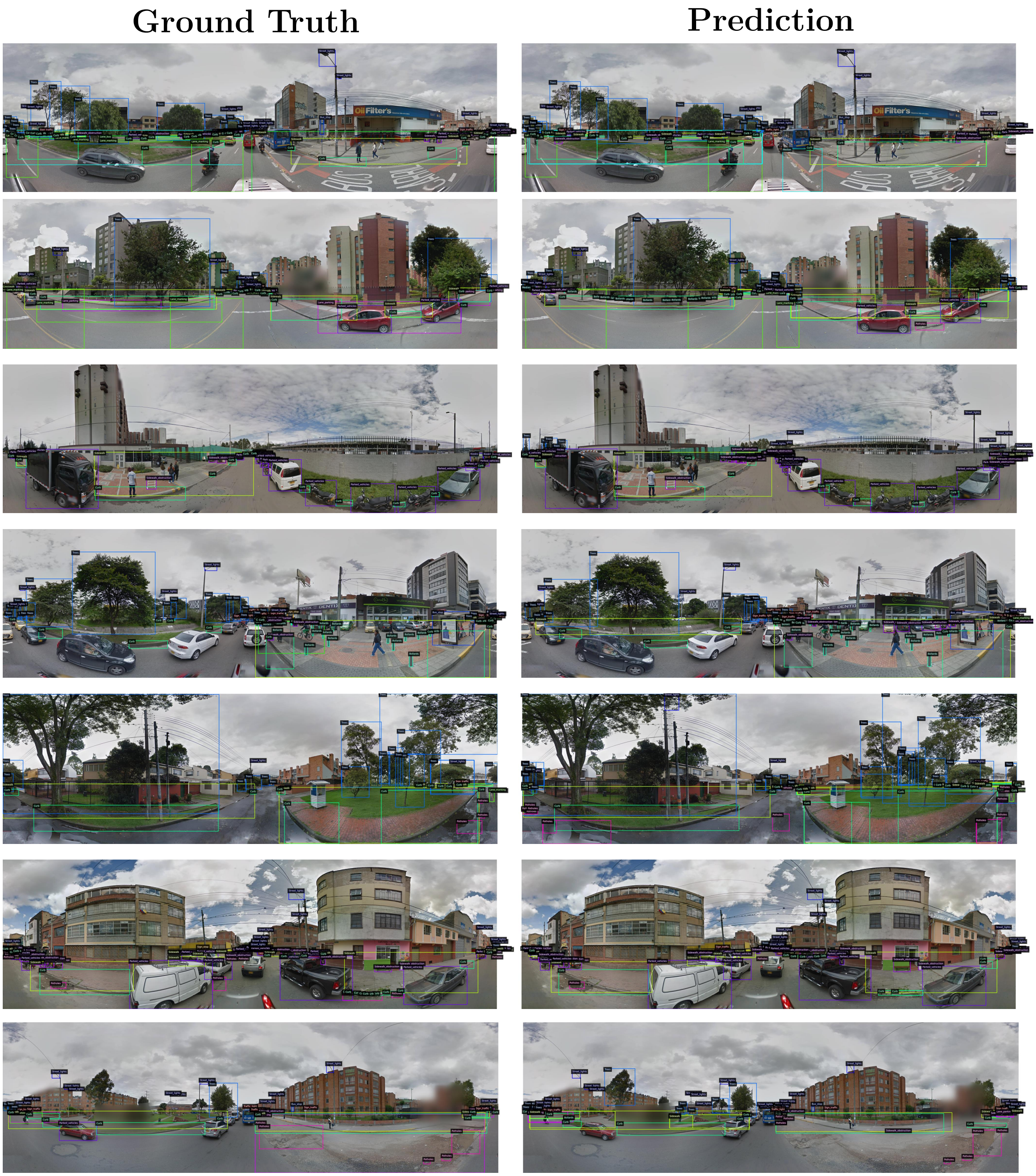

Figure 8: **Examples of qualitative detection results.** Our model correctly identifies most objects in the images with some false positives and some misses.

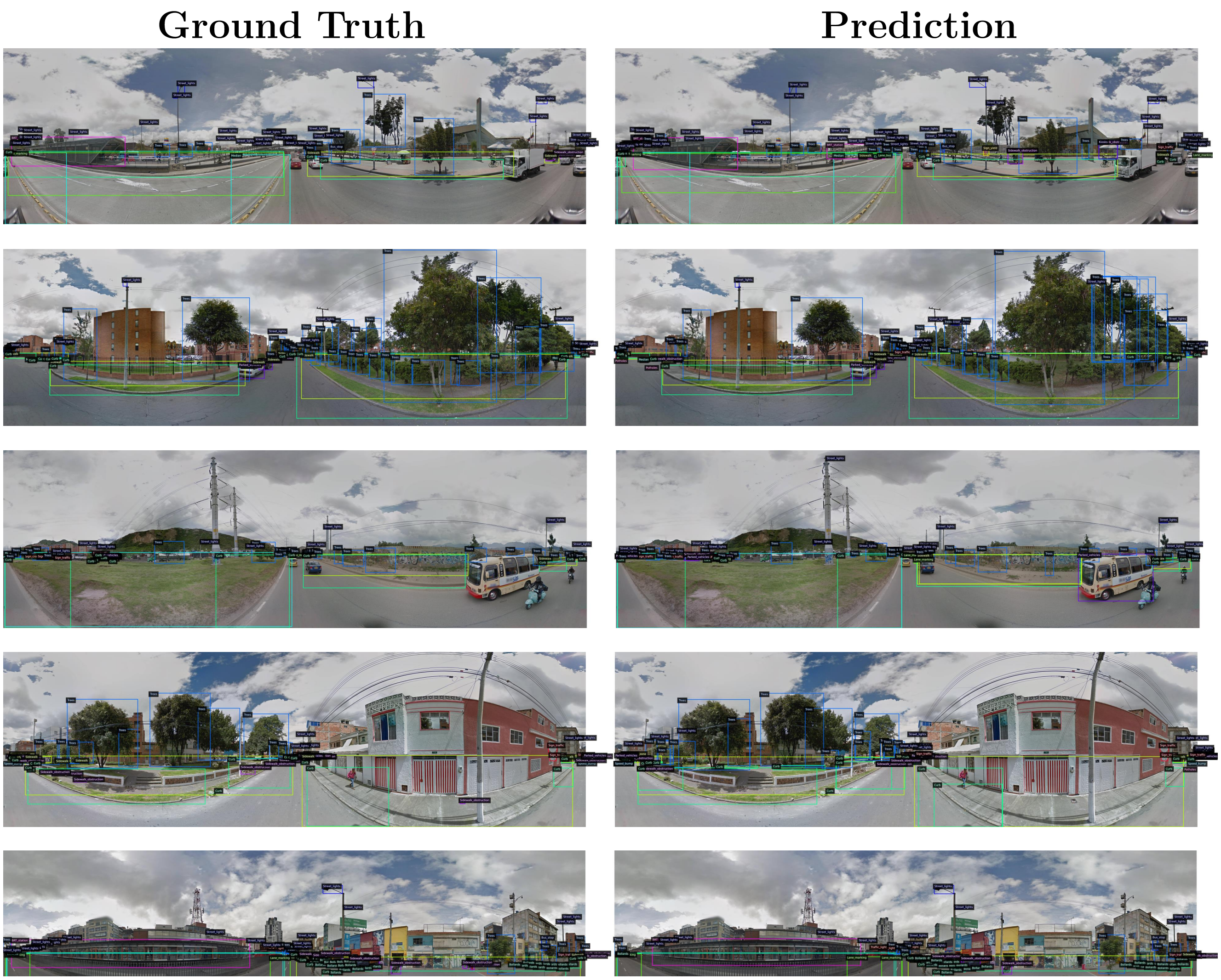

Figure 9: **Examples of qualitative detection results.** Our model correctly identifies most objects in the images with some false positives and some misses.

Class	DINO						Deformable DETR					
	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
School zone sign	59.82	80.33	67.79	46.56	86.96	-	58.22	81.69	66.15	45.76	84.15	-
Street lights	52.42	79.11	57.53	41.30	73.00	-	47.76	75.45	51.9	37.14	69.77	-
Stop sign	57.02	78.71	64.54	50.93	80.72	-	55.09	78.7	62.42	49.49	76.31	-
Perked vehicles	54.67	74.32	59.66	28.25	66.89	81.47	50.18	72.53	54.49	25.01	61.86	78.77
Traffic sign	49.64	72.80	55.39	41.28	78.16	-	46.95	72.22	51.7	38.85	74.45	-
Bus stop	47.61	71.23	53.33	36.75	65.76	-	44.52	70.7	48.41	34.94	60.5	-
Trees	40.83	65.86	41.57	7.89	42.41	80.64	35.5	60.74	34.93	5.37	36.93	76.78
Sidewalk	42.89	64.23	43.90	6.37	41.06	85.48	39.88	62.03	40.55	5.33	37.56	83.03
Lane marking	43.42	62.73	47.94	9.51	51.59	65.07	43.24	62.52	46.26	7.55	51.78	61.96
Traffic light	32.77	62.51	30.35	29.35	64.53	-	30.83	62.89	26.23	27.54	62.4	-
Curb	37.25	61.76	37.65	16.40	54.92	84.88	34.97	59.37	34.61	14.45	52.17	83.91
Bollards	34.89	61.49	35.21	29.93	58.61	67.75	31.35	58.56	30.53	26.82	53.83	66.75
Kiosks	39.56	59.08	42.71	13.15	45.43	–	34.77	54.7	36.75	10.12	40.49	–
Crossing sign	42.93	58.15	50.48	34.24	69.18	–	42.86	60.3	48.92	32.33	71.82	–
Crosswalk	31.40	51.85	32.75	5.77	40.92	28.70	29.02	53.05	29.06	5.74	37.41	24.57
Speed bump	26.84	51.41	25.53	9.14	42.82	–	25.19	50.54	22.41	9.02	39.81	–
Yield sign	30.83	50.43	32.29	23.55	69.88	–	30.21	49.85	30.45	23.44	67.32	–
Pedestrian light	22.15	46.53	18.18	18.29	55.53	–	21.13	47.26	15.7	17.37	51.41	–
Median	29.50	42.24	30.88	3.56	27.67	80.64	27.55	40.55	28.36	2.56	24.79	80.71
Sidewalk obstruction	25.85	41.18	27.05	11.35	34.43	60.32	22.29	38.35	22.35	9.22	30.03	53.58
BRT station	18.61	34.60	16.88	0.08	20.77	42.01	11.78	27.07	8.62	0	13.04	34.08
Bike lane	17.69	25.55	19.05	0.11	15.64	49.69	16.31	25.77	17.3	0.02	14.87	43.31
Bus lane	16.46	23.83	16.41	1.19	5.51	49.71	13.97	21.63	14.14	0.4	5.24	43.85
Pothholes	6.64	15.16	5.03	4.00	9.03	–	5.37	13.26	3.53	3.32	7.44	–
Parking lane	6.16	12.71	4.98	0.09	5.05	23.92	5.03	12.34	2.97	0.02	4.38	17.72
Roundabout	7.47	8.60	8.43	–	6.47	14.82	10.59	12.6	12.25	–	8.4	26.71
Median barrier	4.71	8.00	4.92	1.71	4.83	30.85	2.91	5.21	2.99	1.05	2.09	37.02
Total	32.59	50.53	34.46	18.11	45.10	56.40	30.28	49.25	31.26	16.65	42.23	54.18

Table 1: **Per class Detection Performance**. We compare DINO with Deformable DETR in our object detection task. DINO surpasses Deformable DETR in all metrics, and the detection performance on each AP subtype maintains constant among both models. – indicates the absence of annotations for specific object sizes and categories.