Transferring Labels to Solve Annotation Mismatches Across Object Detection Datasets

Yuan-Hong Liao[2], David Acuna[1], Rafid Mahmood[1,3], James Lucas[1], Viraj Prabhu[4], Sanja Fidler[1,2]

[1] NVIDIA, [2] University of Toronto & Vector Institute, [3] University of Ottawa, [4] Georgia Institute of Technology

Check here for more details!

Contributions

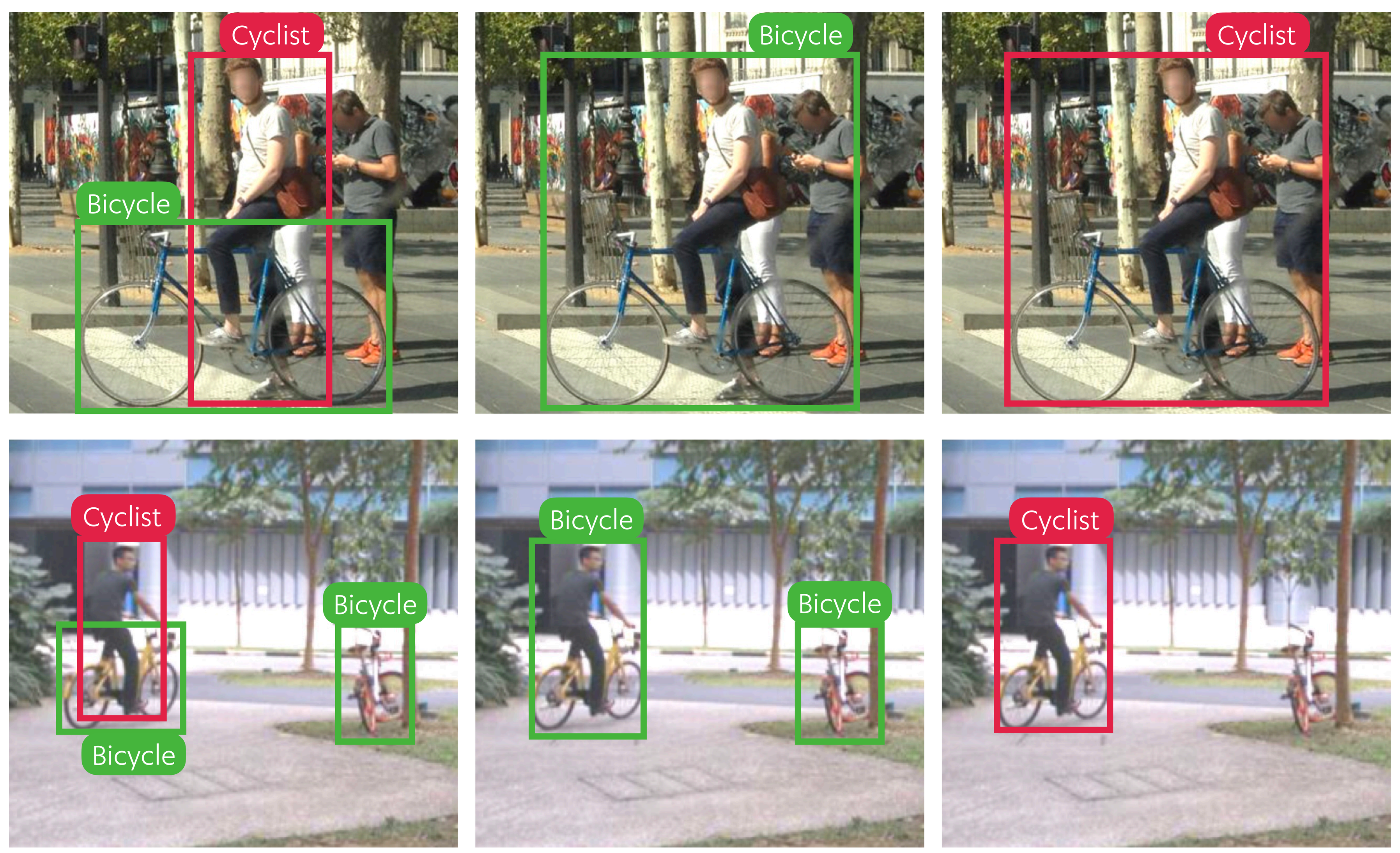
Keywords: Data-centric, Annotation mismatches, Label transfer

1. A prevalent but under-explored label issue: We characterize **annotation mismatches** in object detection datasets.
2. A **data-centric** framework: We formulate **Label Transfer** that performs transfer in the label space.
3. Our approach: **Label Guided -Labeling** (LGPL) that consistently improves downstream detectors across four transferring scenarios and three object detectors, on average by 1.88 mAP and 2.65 AP^{75}

What are Annotation Mismatches?

Annotation mismatches stem from differences in annotation protocols, including class taxonomies, instructions, and label post-processing, etc.

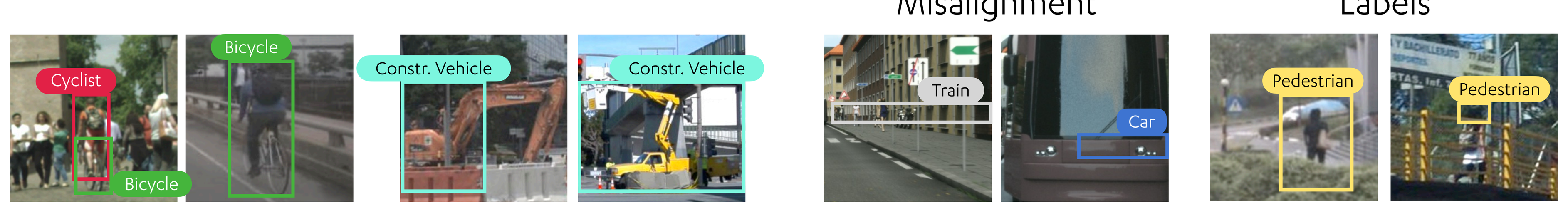
Find the detection label errors in the following images (**answers at the bottom) Bicycle Cyclist



Taxonomy of Annotation Mismatches

We pinpoint four fundamental types of annotation mismatches: class semantics, annotation instructions, human-machine misalignment, and cross-modality labels.

Class Semantics | Annotation Instructions | Human-machine Misalignment | Cross-modality Labels

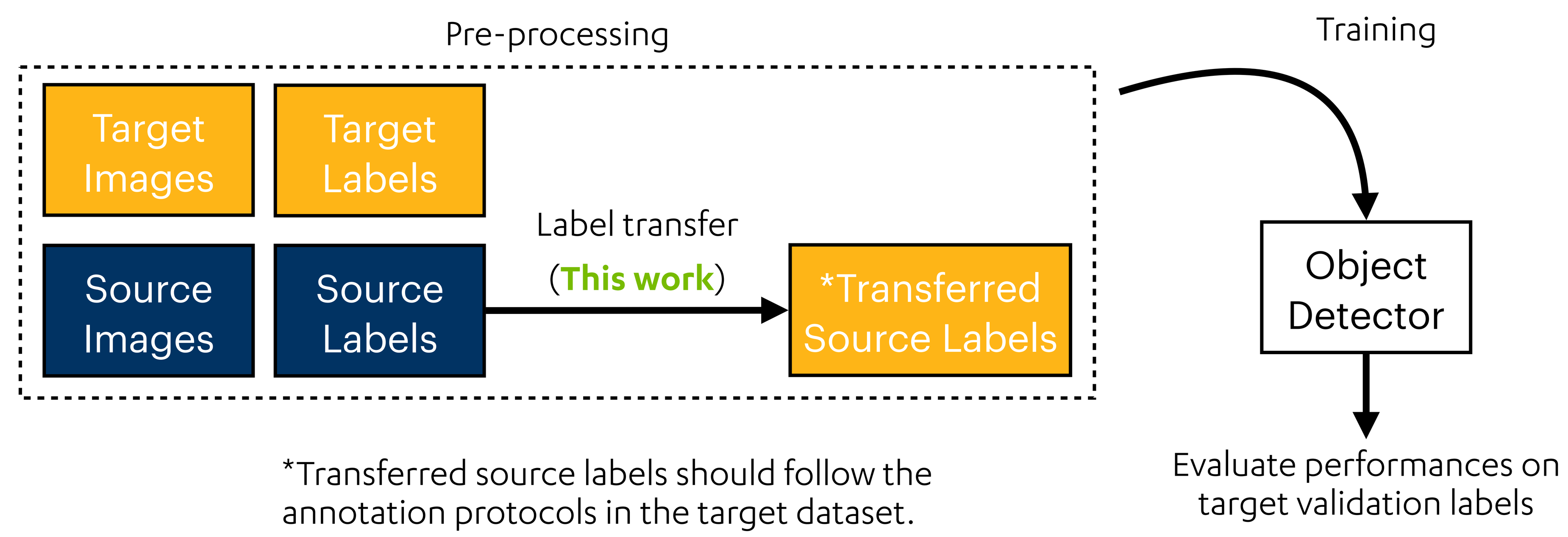


Cityscapes, nuImages, nuImages, Waymo, Synscapes, nuScenes, Mapilliary Vistas

Label Transfer — a data-centric framework

We formulate a data-centric framework to mitigate annotation mismatches — **Label transfer**. Label transfer is a transfer learning framework that explicitly adjusts the labels.

Label transfer can be considered as a pre-processing step before the detector training and can be applied in a plug-and-play fashion.

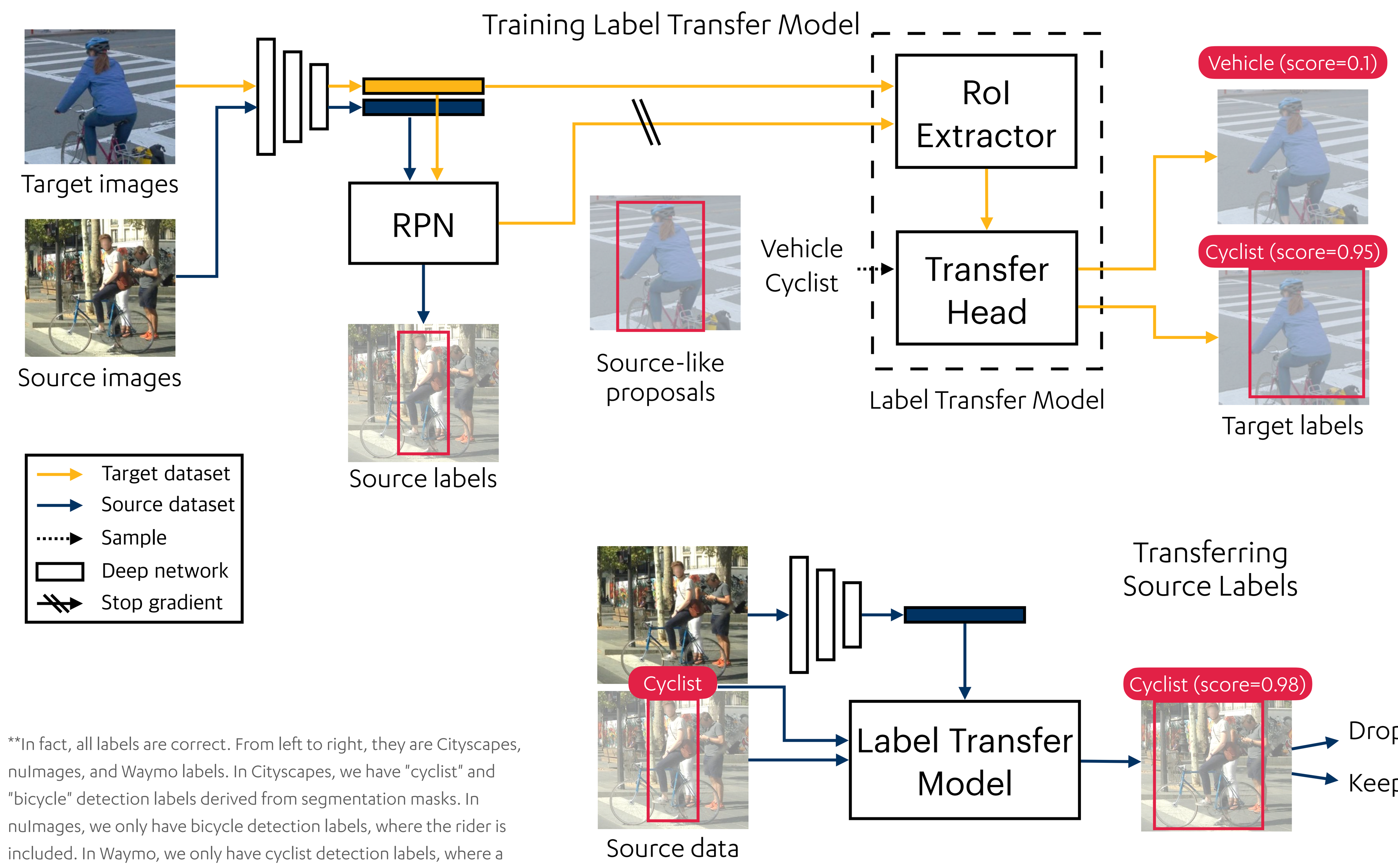


*Transferred source labels should follow the annotation protocols in the target dataset.

Label Guided Pseudo-Labeling

Challenges: No paired labels on the same images.

Motivations: With the modest assumptions, we identify that a label transfer model is secretly in your two-stage object detectors.



**In fact, all labels are correct. From left to right, they are Cityscapes, nuImages, and Waymo labels. In Cityscapes, we have "cyclist" and "bicycle" detection labels derived from segmentation masks. In nuImages, we only have bicycle detection labels, where the rider is included. In Waymo, we only have cyclist detection labels, where a rider is included but parked bicycles are not.

Experimental Results

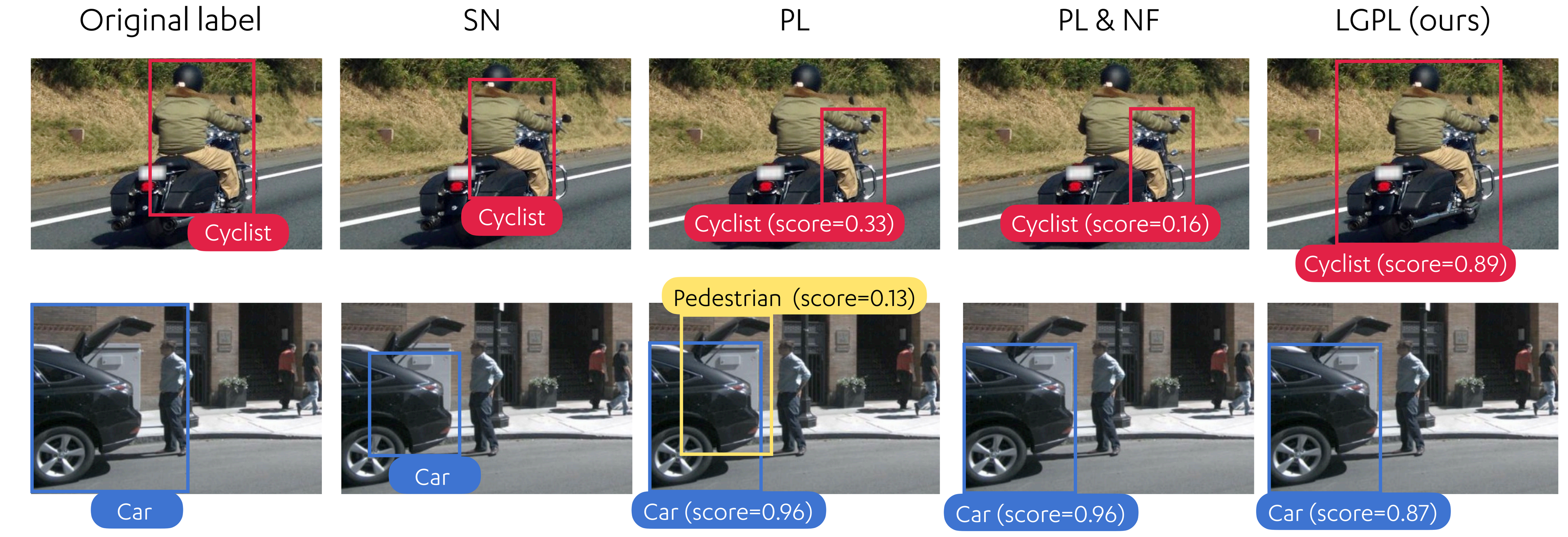
Transferring scenarios
1. nuScenes → nuImages
2. Synscapes → Cityscapes
3. Internal Dataset → nuImages
4. MVD-🚲 + nuImages-🚲 → Waymo-🚲

Baselines
1. No transfer
2. Statistical normalization (SN) [1]
3. Pseudo-labeling (PL) [2]
4. Pseudo-labeling + noise filtering (PL&NF) [3]
5. SAM-transfer [4]

	Label transfer model	YOLOv3	Def-DETR	Faster-RCNN
Source = nuScenes Target = nuImages	No transfer	31.24	39.65	41.25
	SN	31.95	39.59	40.79
	PL	28.67	39.12	40.49
	PL & NF	33.26	40.97	40.68
	LGPL (Ours)	**34.8** +3.56	**41.52** +1.87	**42.6** +1.35
Source = Synscapes Target = Cityscapes	No transfer	26.87	32.93	38.74
	SN	25.53	32.7	36.91
	PL	28.86	30.67	37.88
	PL & NF	28.27	33.04	39.05
	LGPL (Ours)	**29.29** +2.42	**34.45** +1.58	**39.71** +0.97
Source = Internal-Dataset Target = nuImages†	No transfer	39.17	46.79	47.91
	SN	39.07	47.05	48.05
	PL	37.87	47.41	48.5
	PL & NF	39.85	47.67	48.2
	LGPL (Ours)	**41.17** +2	**48.4** +1.61	**48.89** +0.98

Qualitative Results (MVD→Waymo, nuScenes→nuImages)

Original label | SN | PL | PL & NF | LGPL (ours)



References

[1] Wang et al., Train in germany, test in the usa: Making 3d object detectors generalize, CVPR'20

[2] Lee et al., Pseudo-label : The simple and efficient semi-supervised learning method for deep neural networks.

[3] Mao et al., Noisy localization annotation refinement for object detection. ICIP'20

[4] Kirillov et al., Segment anything. ICCV'23

Waymo-bicycle nuImages-bicycle