Classification Drives Geographic Bias in Street Scene Segmentation

Supplementary Material

A. Dataset Comparision

We compared some of the widely used driving datasets in Table 3. It is evident that Mapillary Vistas (Vistas) [16] is the only dataset with global coverage and instance-level annotations.

B. Additional Details on Vistas

In Table 4, we show the number of images in each continent in Vistas. Table 5 shows the number of instances of each class in each continent. Tables 4 and 5 confirm that we had sufficient instances across continents (for most classes) to conduct robust geo-bias evaluations.

C. Class-merging on additional models

In addition to Swin-L (shown in Table 2), we applied class-merging on the following instance segmentation models: Mask-RCNN, ConvNext-L, and ConvNext-XL. Table 7 shows the $\Delta\text{CV}_{\text{det-det-corrected}}$ and the $\Delta\text{CV}_{\text{seg-seg-corrected}}$ values for these models. Similar to Swin-L, for the geo-biased classes, the CV values for detection and segmentation reduced for all models (Mask-RCNN, ConvNext-L, ConvNext-XL) after class-merging. This shows that classification errors significantly contributed to geo-biases in instance segmentation models.

To measure the impact of classification errors on geo-biases in detection models, we also applied class-merging to well-known object detection models (trained on Cityscapes): YOLOv7 [23] and Faster-RCNN [18]. Table 8 shows the $\Delta\text{CV}_{\text{det-det-corrected}}$ values for these models. Similar to the segmentation models (shown in Tables 2 and 7), the CV values for detection reduced for YOLOv7 and Faster-RCNN after class-merging.

Interestingly, in Faster-RCNN, the CV increased for buses (32%) and slightly for motorcycles (9%). To explain the increase in CV, we show the continent-box-IoUs for buses and motorcycles before and after class-merging in Table 6.

For buses, the detection performance of Faster-RCNN before class-merging was similar across continents. After class-merging, performance improved significantly in non-European continents like Africa (0.36 to 0.61), North America (0.30 to 0.51), and Oceania (0.37 to 0.54), compared to a smaller improvement in Europe (0.30 to 0.43). This led to an increase in CV. Similarly, for motorcycles, non-European continents like North America (0.40 to 0.45), South America (0.39 to 0.48), and Asia (0.42 to 0.46) saw larger gains than Europe (0.40 to 0.42).

Despite the increased CV in buses (and marginally in motorcycles) for Faster-RCNN, the results show that classification errors significantly impacted performance in non-European continents.

Table 3. A comparison of well-known driving datasets.

Driving Datasets	Continents Covered						Instance Segmentation
	Asia	Europe	N. America	S. America	Africa	Oceania	
IDD [22]	✓	✗	✗	✗	✗	✗	✗
BDD100K [26]	✗	✗	✓	✗	✗	✗	✓
Cityscapes [5]	✗	✓	✗	✗	✗	✗	✓
nuScenes [3]	✓	✗	✓	✗	✗	✗	✓
ApolloScapes [12]	✓	✗	✗	✗	✗	✗	✗
Mapillary Vistas [16]	✓	✓	✓	✓	✓	✓	✓

Table 4. Number of images per continent in the Vistas dataset.

Continent	Asia	Europe	N. America	S. America	Africa	Oceania	Total
Number of Images	2157	3752	3090	712	154	682	10547

Table 5. Number of instances of each class in each continent in the Vistas dataset.

Class	Asia	Europe	N. America	S. America	Africa	Oceania
person	1892	10106	4265	2197	833	1109
car	9514	21848	22068	4820	1090	4256
bus	459	943	435	350	80	94
truck	1205	886	1057	275	84	165
bicycle	411	1691	380	160	24	69
motorcycle	743	1022	149	388	128	37
rider	771	1178	322	469	127	62

Table 6. Detection performance on Faster-RCNN for buses and motorcycles before and after class-merging. Values on the left of the arrow are continent-box-IoUs, while values on the right are corrected-continent-box-IoUs.

	Continents					
Class	Europe	Africa	N. America	S. America	Asia	Oceania
bus	$0.30 \rightarrow 0.43$	$0.36 \rightarrow 0.61$	$0.30 \rightarrow 0.51$	$0.33 \rightarrow 0.49$	$0.29 \rightarrow 0.44$	$0.37 \rightarrow 0.54$
motorcycle	$0.40 \rightarrow 0.42$	$0.36 \rightarrow 0.43$	$0.40 \rightarrow 0.45$	$0.39 \rightarrow 0.48$	$0.42 \rightarrow 0.46$	$0.37 \rightarrow 0.40$

Table 7. Percentage change in CV for various instance segmentation models. Column 2 shows the % change in CV in detection performance before and after class-merging ($\Delta CV_{\text{det-det-corrected}}$). Column 3 shows the % change in CV in segmentation performance before and after class-merging ($\Delta CV_{\text{seg-seg-corrected}}$).

Model Name	Class	$\Delta CV_{\text{det-det-corrected}}$	$\Delta CV_{\text{seg-seg-corrected}}$
Mask-RCNN	person	4.05%	2.72%
	car	−5.97%	−6.97%
	bus	−29.06%	−29.18%
	truck	−28.84%	−24.96%
	bicycle	−15.41%	−11.45%
	motorcycle	−26.83%	−17.03%
	rider	−15.64%	−19.35%
ConvNext-L	person	−0.65%	1.34%
	car	4.87%	1.44%
	bus	−66.61%	−72.34%
	truck	−29.97%	−42.22%
	bicycle	−7.99%	−1.10%
	motorcycle	−30.42%	−40.02%
	rider	−34.12%	−46.19%
ConvNext-XL	person	−2.76%	−0.23%
	car	−10.43%	−14.04%
	bus	−74.10%	−72.78%
	truck	−29.14%	−34.48%
	bicycle	−16.60%	−3.39%
	motorcycle	−30.96%	−37.36%
	rider	−27.12%	−44.82%

Table 8. Percentage change in CV for various detection models. Column 2 shows the % change in CV in detection performance before and after class-merging ($\Delta CV_{\text{det-det-corrected}}$)

Model Name	Class	$\Delta CV_{\text{det-det-corrected}}$
YOLOv7	person	−5.81%
	car	2.59%
	bus	−55.49%
	truck	−34.21%
	bicycle	−0.47%
	motorcycle	−33.13%
	rider	−14.23%
Faster-RCNN	person	−2.66%
	car	−6.46%
	bus	32.55%
	truck	−51.98%
	bicycle	−1.85%
	motorcycle	9.41%
	rider	−6.79%