

The Mapillary Vistas Dataset for Semantic Understanding of Street Scenes

Supplementary Material

Gerhard Neuhold, Tobias Ollmann, Samuel Rota Bulò, Peter Kontschieder
Mapillary Research
research@mapillary.com

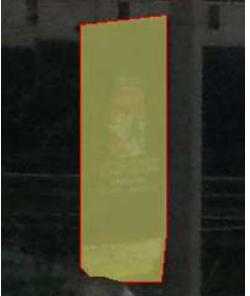
Abstract

We provide the following additional contributions to our ICCV submission:

- *definition of the main categories in the dataset together with an illustrative image;*
- *detailed results on semantic segmentation and instance-specific semantic segmentation;*
- *the taxonomy of classes.*

Class-Label Definitions

In the following table we provide a description of the categories in the dataset. Each entry provides the category's name together with its root-level and macro-level groups, the definition of the category including some instructions for the annotator, an illustrative RGB image and the corresponding segmentation. The latter segmentation is shown only for a single object instance, also in the case multiple instances are depicted.

Root-level	Macro-level	Category	Definition	RGB Image	Example
object	object	banner	Typically a flag, made of textile or plastic.		
		bench	A place to sit down.		

bike rack A construction to place and lock a bike.



billboard Advertisements placed in public space and also those mounted on buildings.



catch-basin Sewage water drainage, typically with a grid on top or inlet of a curb.



cctv-camera Surveillance camera mounted in public space.



fire-hydrant Water outlet typically used by fire-fighters.



junction-box

An electrical junction box can contain junctions of electric wires and/or cables.



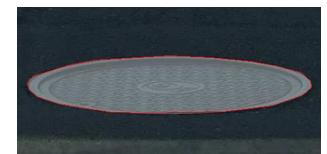
mailbox

A box to put mail into.



manhole

Opening in road to access canal system. Has to be big enough to fit a person and includes squared ones.



phone booth

A public call box typically equipped with a payphone.



pothole

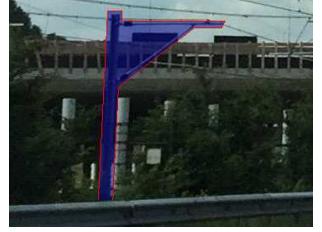
A depression or hole in a road caused by wear or subsidence.



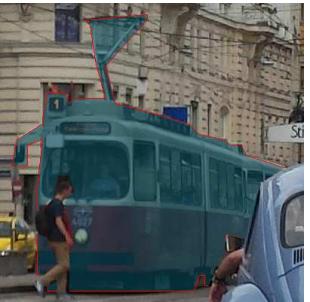
street-light

Lights mounted on poles to illuminate the road or public places.



support	pole	Thin and elongated, typically vertically oriented poles, <i>e.g.</i> sign or traffic light poles. Does not include objects mounted on the pole. Poles holding only street lights should also be labeled as pole.		
	traffic-sign-frame	Frame construction holding large signs, where frame can be made of round or square poles. Can be typically found on highways.		
	utility-pole	Utility poles are supposed to carry electrical wires/cables, could be used to mount mobile phone network receivers, and are used in case both, cables <i>and</i> street lights are mounted.		
object	traffic light	Traffic lights without poles in different orientations (upright, horizontal, side, back, front) and for all types of traffic participants, <i>e.g.</i> regular traffic light, bus traffic light, train traffic light. If the type can be determined but does not match any of the defined categories (<i>e.g.</i> traffic lights for trains), then <i>traffic light - other</i> is used. If the orientation or type of traffic light cannot be determined, <i>traffic light - ambiguous</i> is used.		

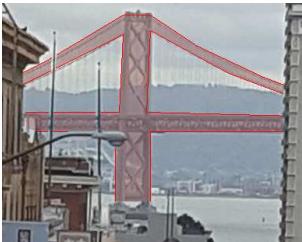
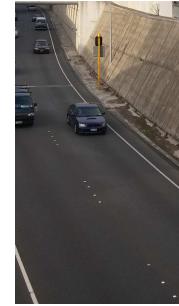
traffic-sign	front	<p>Signs with the purpose of conveying information to drivers/cyclists/pedestrians, e.g. traffic signs or warning reflector posts. If the type of a front-viewed traffic sign cannot be determined, e.g. because of unknown language, <i>traffic sign - unknown</i> is used as fallback.</p>		
	back	Back side of traffic signs.		
object	trash-can	Recycling facility for waste collection.		
vehicle	bicycle	<p>Includes bicycles without the cyclist or possibly other passengers. The cyclist and passengers receive label <i>cyclist</i>.</p>		
	boat	Vehicle suitable for riding/sailing on water.		

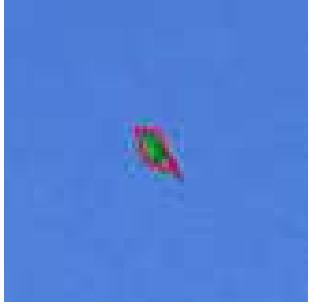
bus	Includes buses that are intended for 9+ persons for public or long-distance transport.		
car	This includes cars, jeeps, SUVs, vans with a continuous body shape (<i>i.e.</i> the driver cabin and cargo compartment are one).		
caravan	Vehicles that (appear to) contain living quarters. This also includes trailers that are used for living and has priority over the trailer class.		
motorcycle	This includes motorcycles, mopeds without the driver or other passengers. Both driver and passengers receive the label motorcyclist.		
on rails	All vehicles moving on rails, <i>e.g.</i> trams, trains.		

other-vehicle	Fallback category for vehicles not explicitly defined otherwise in meta category level vehicle.				
trailer	Trailers that can be attached to any vehicle, but excludes trailers attached to trucks. The latter belong to class truck.				
truck	This includes trucks, vans with a body that is separate from the driver cabin, pickup trucks, as well as their attached trailers.				
wheeled-slow	Covers all slow moving wheeled objects like wheelchairs, mobility-scooters, scooters, prams or rickshaw. All of which can potentially have riders (other rider) sitting on them, similar to cycles are driven by cyclists.				
con- struc- tion	flat	bike lane	Paved areas intended for bicycle use only. Does not include regions meant to be used by both cyclists and pedestrians. Regions of road intended for mixed cycle and car use are labeled as road. Contains only regions of ground specifically designated for cyclists where pedestrians and cars should not enter, e.g. separated by curb or lane markings (dashed or driveable).		

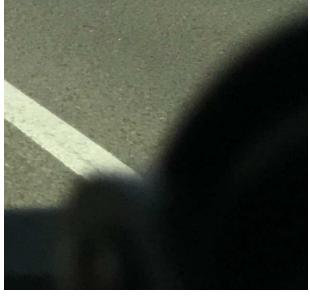
crosswalk plain	Region of a road which is meant to cross the street but not marked with a zebra pattern, instead two parallel lines indicate the crosswalk. The road surface is marked as plain crosswalk, the two lines indicating the border are marked as general markings.		
curb cut	Lowered curbs ($\approx 3\text{cm}$) or curb cuts ($<3\text{cm}$) which can be driven over, <i>e.g.</i> by wheelchairs.		
parking	Surfaces that are intended for parking and separated from the road either via elevation or different texture/material or lane markings. Only the area which is available for parking has to be annotated. Flat areas used to get on a specific parking slot need to be labeled as road.		
pedestrian area	Areas solely to be used by pedestrians. <i>E.g.</i> in inner city areas or parks, shopping zones, <i>etc.</i>		
rail track	Horizontal surface on which rail cars typically drive. If rail-tracks are embedded in a conventional road block, the rail track is drawn on top of a road label.		
road	Surface on which cars usually drive on. Typically delimited by curbs, parking areas or bicycle lanes.		

service lane	Road shoulders, typically on highways, which offer enough space for a car to drive and stop in emergency cases. Usually on the right side of drivable lanes (viewed in driving direction, in case of right-hand traffic), but can also be on the left side.		
sidewalk	Horizontal surface designated for pedestrians or cyclists. Delimited from the road by some obstacles, e.g. curbs (raised or lowered) or poles, but not only by marking. Often elevated compared to the road and often located at the side of a road. Also includes walkable parts of traffic islands, but not pedestrian areas.		
barrier	curb		
fence	Raised (>3cm) formation typically between road and sidewalk or terrain. Includes the brick's horizontal and vertical surface. If there is no clear border between sidewalk/terrain and curb, only the vertical surface is labeled.		
guard rail	Freestanding structure with holes that separate two (or more) outdoor areas, sometimes temporary ones.		
other barrier	Fallback category for other forms of barriers, e.g. lane separator, jersey barrier.		

	wall	Individually standing walls that separate two (or more) outdoor areas, and do not provide support for a building. Includes transparent walls.		
structure	bridge	Bridges (on which the ego-vehicle is not driving) including everything (fences, guard rails) permanently attached to them.		
	building	Includes structures that house/shelter humans, <i>e.g.</i> low-rises, skyscrapers, bus stops, carports. Translucent buildings made of glass still receive the label building. Also includes scaffolding attached to buildings.		
	tunnel	Tunnel walls and the dark space encased by the tunnel, but excluding vehicles.		
marking	crosswalk zebra	Crosswalk indicated by a zebra pattern.		
	general	Other lane markings including dashed and solid lane separation markings, stop lines, text, arrows or restricted areas.		

human	person	Persons moving by their own legs, sitting or in any unusual pose but not meant to drive a vehicle.		
rider	bicyclist	Humans driving a bicycle. Does not include the bicycle or persons moving a bike while walking.		
motorcyclist		Person riding a motorcycle without the motorcycle, includes passengers.		
other rider		People riding other types of vehicles like rickshaw, wheeled chair, scooter, mobility-scooter.		
animal	bird	Includes birds on the ground.		

ground animal	Animals which move on the ground.			
nature	mountain	Distant elevations of the ground.		
	sand	Includes surfaces consisting of sand or sand heaps, e.g. at construction areas.		
	sky	Sky including fog or heavy precipitation.		
	snow	Snow covered surfaces.		
	terrain	Grass and all other types of horizontal spreading vegetation, includes soil.		
	vegetation	Trees and other forms of vertical growing vegetation.		
	water	Ground surfaces covered by water.		

void	car mount	Mount used to hold the camera while capturing.		
ego vehicle	ego vehicle	Ego-vehicle, e.g. car hood, dashboard, driving mirror.		
unlabeled	unlabeled	Regions of images without an assigned category. Categories which are not included in the research version are mapped to unlabeled.		

Detailed results on semantic segmentation and instance-specific semantic segmentation

In Tab. 1 we report the per-category AP and AP@0.5 scores obtained for the Mapillary Vistas Dataset’s instance-specific semantic segmentation task on both, validation set and test set for the Mask R-CNN variant of the UCenter team that won the LSUN challenge. In the table we distinguish results obtained with a single model and an ensemble of models (see the main paper for details about their method).

In Tab. 2 we provide per-category IoU scores obtained for the Mapillary Vistas Dataset’s semantic segmentation task on both, the validation set and test set for PSPNet and Wider Resnet based approaches described in the main paper. We denote PSPNet + Cityscapes pre-training as *PSPNet⁺*, PSPNet⁺ + multi-scale testing as *PSPNet⁺⁺*, and PSPNet⁺⁺ with 4 model ensembles as *PSPNet⁺⁺⁺*. Similarly, we denote the basic Wider Resnet with uniform sampling and cross-entropy loss as *WiderNet*, WiderNet + loss max-pooling and uniform sampling as *WiderNet⁺*, and WiderNet + loss max-pooling and balanced class sampling as *WiderNet⁺⁺*.

Class-Labels Taxonomy

In Fig. 1 we provide the taxonomy for all class labels. On the left, we provide the taxonomy for the root-level category *object*, where we omitted macro-level category *object* for presentation clarity (*i.e.*, we connected its children to the root-level category directly such as *e.g. Bench*). On the right, we see the taxonomies for the rest of the labels.

	validation set				test set	
	UCenter single model		UCenter ensemble		UCenter ensemble	
	AP	AP@0.5	AP	AP@0.5	AP	AP@0.5
Bird	7.10	17.50	10.00	25.80	10.00	26.40
Ground Animal	13.80	27.10	20.30	43.10	21.00	40.40
Crosswalk - Plain	16.10	30.90	11.30	26.30	12.80	28.30
Person	38.50	70.90	35.90	68.80	37.00	69.50
Bicyclist	27.70	63.60	23.00	53.20	23.60	54.60
Motorcyclist	20.70	52.30	16.90	49.70	19.30	53.10
Other Rider	2.70	8.50	0.60	3.20	0.70	2.50
Lane Marking - Crosswalk	31.00	58.70	26.20	52.40	27.30	53.80
Banner	21.60	32.60	20.70	32.80	21.10	33.90
Bench	12.10	25.80	11.10	26.10	11.60	28.10
Bike Rack	2.80	8.40	1.80	4.80	3.30	8.40
Billboard	28.50	45.70	27.10	44.90	27.60	45.20
Catch Basin	21.30	40.20	21.70	43.00	22.70	43.70
CCTV Camera	15.00	35.40	10.10	26.40	10.50	28.10
Fire Hydrant	49.00	78.10	47.70	78.20	48.30	78.20
Junction Box	34.90	53.00	27.90	43.20	28.40	43.40
Mailbox	11.30	20.50	18.00	32.10	18.40	32.60
Manhole	37.70	58.70	36.40	58.90	37.00	59.60
Phone Booth	5.70	12.90	7.80	14.30	8.50	14.50
Street Light	45.10	78.60	37.10	71.90	37.60	72.20
Pole	18.60	50.20	15.40	44.50	16.10	45.50
Traffic Sign Frame	7.90	27.40	4.60	16.80	5.60	22.10
Utility Pole	17.10	51.30	15.50	48.30	16.50	50.00
Traffic Light	52.00	83.60	45.00	77.90	45.80	78.20
Traffic Sign (Back)	31.30	57.30	24.00	50.50	25.10	51.00
Traffic Sign (Front)	58.50	79.40	50.80	74.10	51.20	74.20
Trash Can	36.60	56.40	35.60	56.40	36.20	56.90
Bicycle	22.90	54.00	19.60	48.20	20.40	49.20
Boat	1.00	2.90	2.90	7.50	3.20	9.10
Bus	51.90	66.90	46.60	62.90	47.40	63.00
Car	63.50	85.50	58.90	82.40	59.70	82.70
Caravan	2.40	3.40	6.60	9.50	7.50	9.50
Motorcycle	30.20	64.30	24.80	56.80	26.70	59.00
Other Vehicle	16.70	29.60	11.20	24.20	12.10	25.80
Trailer	11.90	20.00	16.80	24.70	18.00	24.50
Truck	44.50	61.40	42.00	58.40	42.90	59.00
Wheeled Slow	9.60	23.50	12.90	30.20	15.20	33.40
Avg	24.80	44.20	22.80	42.50	23.70	43.50

Table 1. Detailed results on the instance-specific semantic segmentation task of Mapillary Vistas Dataset in [%].

	validation set						test set			
	PSPNet ⁺	PSPNet ⁺⁺	PSPNet ⁺⁺⁺	WiderNet	WiderNet ⁺	WiderNet ⁺⁺	PSPNet ⁺⁺⁺	WiderNet	WiderNet ⁺	WiderNet ⁺⁺
Bird	15.80	17.11	9.04	0.00	0.00	14.45	1.48	0.00	0.00	6.98
Ground Animal	24.52	33.80	45.61	0.00	0.00	41.15	16.52	0.00	0.00	32.42
Curb	61.51	63.82	64.90	57.67	58.98	58.42	66.60	58.31	60.08	58.56
Fence	63.99	65.85	66.97	59.22	60.53	59.33	66.86	55.94	57.12	56.55
Guard Rail	63.33	65.20	67.01	62.91	62.83	63.87	67.25	63.39	62.96	62.70
Barrier	57.82	59.36	61.16	52.17	54.34	52.38	67.21	57.80	59.41	58.49
Wall	51.73	53.32	53.28	44.88	47.91	45.43	53.30	41.27	42.28	41.98
Bike Lane	38.17	39.21	41.05	21.18	21.96	24.51	27.24	17.00	16.40	18.70
Crosswalk - Plain	36.51	42.33	36.17	9.86	12.72	16.87	39.25	13.74	15.33	16.60
Curb Cut	21.65	23.10	25.24	8.70	13.58	14.09	24.83	8.30	12.90	14.44
Parking	25.69	27.00	28.35	6.51	10.22	12.14	21.58	8.47	10.30	9.66
Pedestrian Area	40.88	38.40	43.97	16.27	16.80	21.55	20.35	12.28	11.29	12.77
Rail Track	53.75	54.35	57.95	44.06	39.95	46.58	68.00	53.16	49.63	50.98
Road	87.90	88.71	89.29	84.15	84.17	84.43	80.27	85.60	85.83	85.13
Service Lane	45.67	46.53	45.85	23.35	25.69	37.31	46.79	27.47	27.56	33.92
Sidewalk	68.24	68.95	72.07	58.60	58.59	56.68	72.40	58.93	58.79	56.00
Bridge	75.77	77.30	78.60	67.63	68.80	67.72	74.23	61.30	62.59	59.91
Building	88.60	89.18	89.20	86.81	87.52	86.78	88.32	84.00	84.41	83.79
Tunnel	32.61	32.91	27.59	14.51	20.68	19.83	45.86	19.59	21.48	18.60
Person	77.75	79.68	80.11	74.90	75.23	77.76	78.54	73.69	74.58	74.78
Bicyclist	64.07	66.06	66.39	45.38	42.42	52.38	54.60	48.88	49.23	45.96
Motorcyclist	67.19	69.59	69.48	50.88	51.25	64.89	69.38	52.17	54.73	64.40
Other Rider	1.32	1.34	1.35	0.00	0.00	1.77	2.73	0.00	0.00	4.58
Lane Marking - Crosswalk	68.79	70.84	72.29	61.42	62.94	61.75	77.55	65.39	66.51	64.74
Lane Marking - General	58.44	60.79	61.42	53.48	53.89	53.13	61.98	53.74	54.25	53.43
Mountain	51.39	52.80	55.49	42.19	47.98	50.14	61.70	52.33	55.89	54.84
Sand	35.25	34.60	29.26	0.50	8.46	17.03	30.13	0.05	13.39	15.50
Sky	89.98	90.65	90.83	98.05	98.11	98.02	80.74	98.15	98.21	98.09
Snow	78.15	78.02	81.84	76.26	77.90	78.91	78.57	66.02	67.10	65.29
Terrain	71.48	72.67	73.43	69.57	69.97	69.22	72.37	66.56	66.94	66.19
Vegetation	90.34	90.86	91.05	90.00	90.19	89.93	81.23	90.63	90.82	90.50
Water	78.85	80.64	81.57	68.18	71.49	70.15	67.88	57.40	59.21	50.22
Banner	34.80	37.23	37.22	25.18	29.26	29.66	39.05	23.33	26.88	29.46
Bench	41.38	44.90	47.44	44.44	48.81	49.93	45.12	30.03	34.94	33.58
Bike Rack	21.14	23.47	22.60	6.13	12.04	12.09	13.59	4.30	8.10	11.26
Billboard	51.51	53.85	54.49	46.11	46.84	46.70	56.35	47.82	47.92	47.73
Catch Basin	44.84	47.97	47.59	36.19	37.34	41.56	45.96	29.96	33.25	34.97
CCTV Camera	21.32	24.21	12.32	0.00	0.00	9.17	12.26	0.00	0.01	7.51
Fire Hydrant	59.09	63.99	64.46	15.16	36.53	55.25	56.88	19.31	42.43	54.53
Junction Box	44.94	50.33	57.96	34.12	36.23	41.21	52.97	37.82	41.04	44.80
Mailbox	13.95	17.77	14.80	0.05	1.10	19.03	11.97	0.00	0.89	11.55
Manhole	52.12	55.04	56.34	46.56	47.33	48.37	49.62	39.25	39.24	42.64
Phone Booth	12.16	15.43	16.78	0.08	10.43	13.90	9.78	0.01	1.58	5.57
Pothole	12.94	13.35	9.09	0.00	0.00	5.62	3.15	0.00	0.00	5.07
Street Light	46.82	49.99	48.53	46.88	50.02	49.26	51.26	49.09	52.16	51.89
Pole	50.46	54.16	54.01	50.04	51.50	50.92	56.60	51.55	53.43	52.53
Traffic Sign Frame	56.69	57.71	57.23	47.04	50.29	52.26	59.81	46.90	50.20	50.80
Utility Pole	52.76	55.64	55.98	48.18	49.75	49.10	57.53	47.62	50.17	49.01
Traffic Light	69.63	72.47	73.02	66.20	69.07	69.00	77.52	71.87	74.52	74.29
Traffic Sign (Back)	48.58	52.08	52.51	39.97	44.97	45.76	54.15	40.62	46.67	47.83
Traffic Sign (Front)	75.27	76.78	76.70	73.63	74.73	74.39	80.75	76.60	78.14	77.47
Trash Can	55.32	57.90	59.33	49.06	51.25	53.98	61.78	49.66	53.05	54.74
Bicycle	63.92	66.11	68.99	63.17	64.53	67.16	65.33	56.67	59.96	60.93
Boat	41.69	42.80	38.98	12.28	24.04	55.28	20.04	6.95	11.25	13.64
Bus	79.48	81.45	82.88	67.06	68.48	69.33	82.78	63.25	64.42	60.66
Car	92.18	93.05	93.08	89.94	90.54	90.72	92.20	89.20	89.82	89.21
Caravan	0.03	0.00	0.00	0.00	2.94	19.95	0.00	0.00	0.00	10.35
Motorcycle	66.16	68.14	70.36	60.60	64.44	66.23	73.21	62.74	64.48	68.92
On Rails	55.06	57.00	53.33	20.00	24.91	46.67	51.32	13.57	14.31	16.94
Other Vehicle	28.37	30.65	33.81	17.34	22.28	27.18	39.19	18.57	23.00	29.15
Trailer	13.43	12.92	12.44	0.00	0.00	1.58	15.96	0.00	0.00	7.33
Truck	71.13	72.14	75.40	61.52	65.21	63.74	72.73	57.49	58.57	56.89
Wheeled Slow	24.39	28.55	26.80	1.69	21.56	36.26	35.37	1.32	16.63	29.35
Car Mount	71.95	72.96	74.78	65.80	67.37	60.10	81.98	73.16	75.47	67.06
Ego Vehicle	92.93	93.19	93.32	89.16	89.80	88.82	95.05	91.38	92.17	90.32
mean IoU	51.59	53.51	53.85	41.12	43.78	47.72	52.41	40.79	42.98	44.84

Table 2. Detailed results on the semantic segmentation task of Mapillary Vistas Dataset in [%].

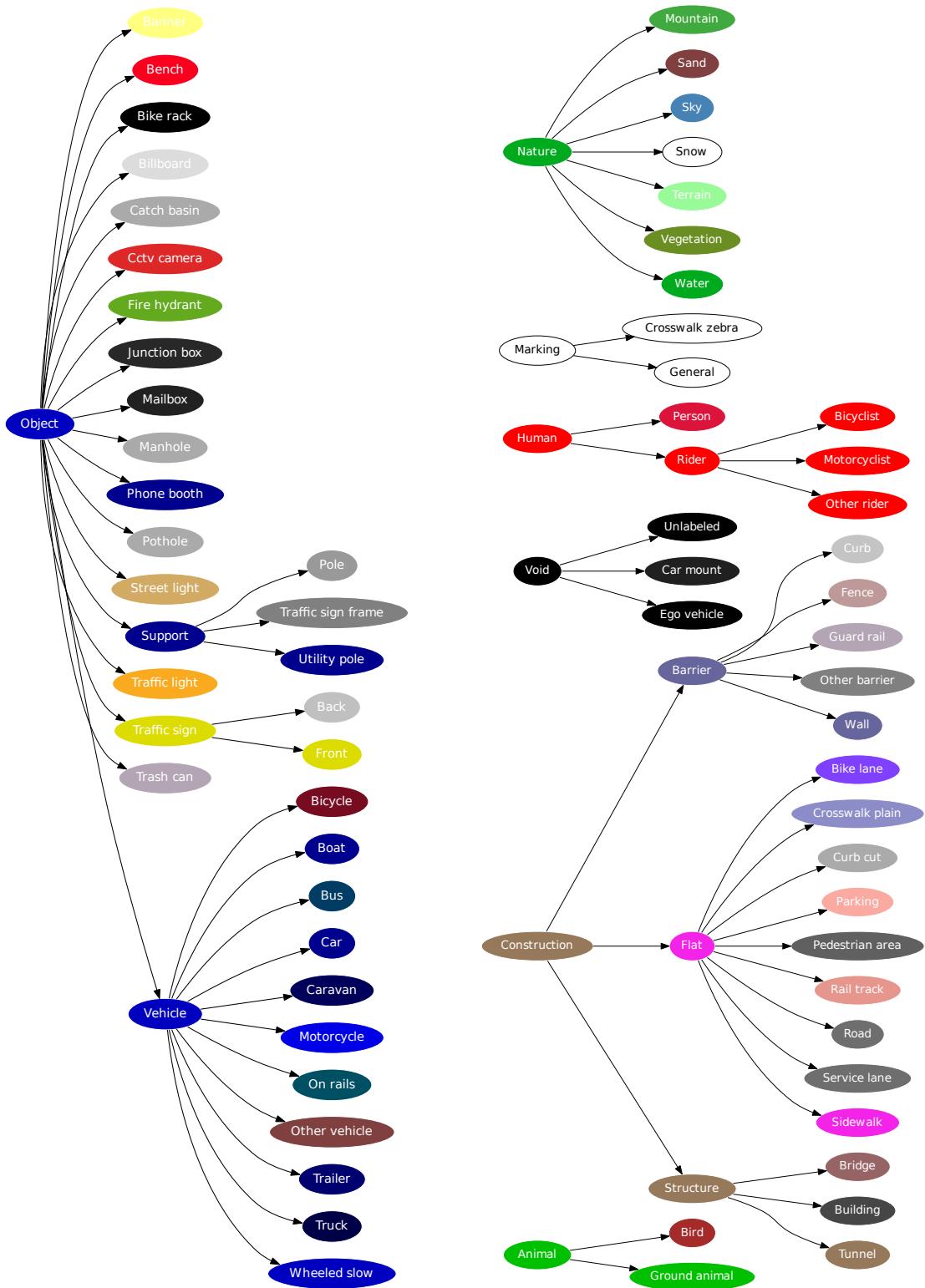


Figure 1. Taxonomy of class labels