Hierarchical Approaches

Pawnesh Gautam, 07-05-2022

Referenced Papers:

1. Training of Convolutional Networks on Multiple Heterogeneous Datasets for Street Scene Semantic Segmentation

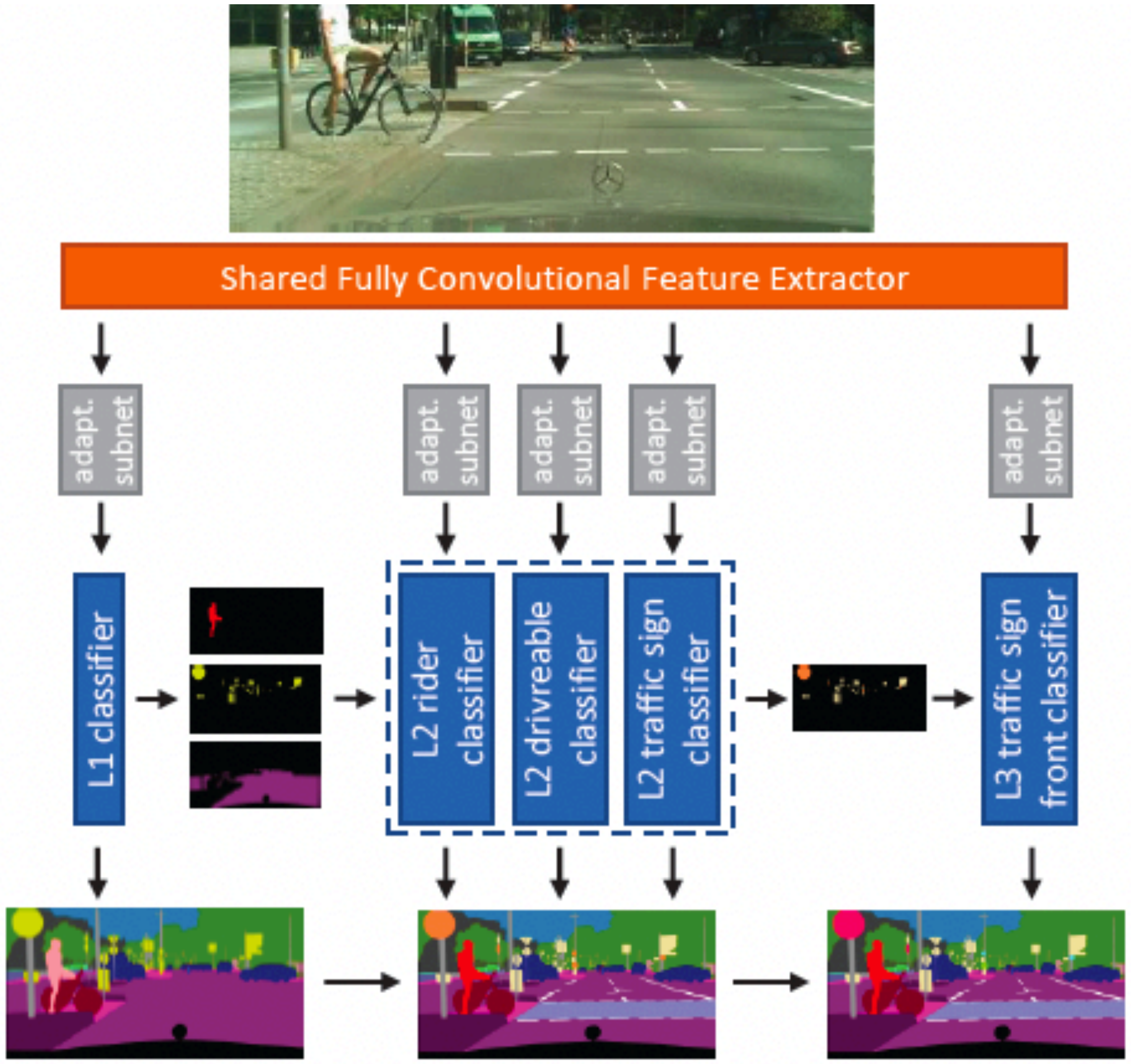
Meletis and Dubbelman: https://arxiv.org/pdf/1803.05675.pdf

- Datasets:
  - Training:
    - Cityscapes
    - GTSDB (traffic signs) non-pixel annotation
    - Mapillary Vistas

1. Training of Convolutional Networks on Multiple Heterogeneous Datasets for Street Scene Semantic Segmentation

Meletis and Dubbelman: https://arxiv.org/pdf/1803.05675.pdf

- Backbone: ResNet-50
- Same adaptation subnetworks
- Hybrid upsampling
- (2x2 learnable fractional strides convolutional layer, bilinear upsampling)
- Batch: 4 Cityscapes:Vistas:GTSDB = 1:2:1
- Input: 512x706
- Inference Time: 58 ms per frame
- Loss: Hierarchical loss



1. Training of Convolutional Networks on Multiple Heterogeneous Datasets for Street Scene Semantic Segmentation

Meletis and Dubbelman: https://arxiv.org/pdf/1803.05675.pdf

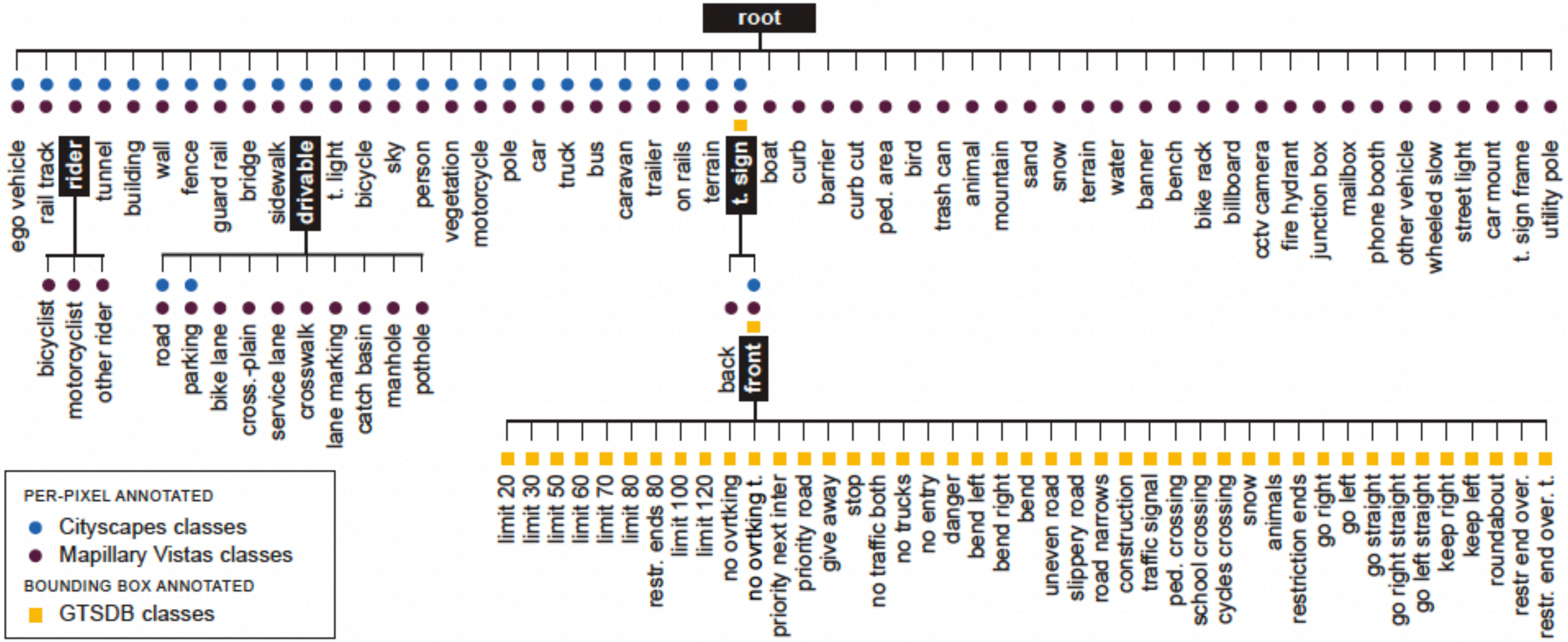
- 

Fig. 2. Three-level semantic label hierarchy combining 108 classes from Cityscapes, Mapillary Vistas and GTSDB dataset. Classes marked in black correspond to the L1, L2, and L3 classifiers of Fig. 1.

1. Training of Convolutional Networks on Multiple Heterogeneous Datasets for Street Scene Semantic Segmentation

Meletis and Dubbelman: https://arxiv.org/pdf/1803.05675.pdf

- Both losses are accumulated per classifier to the so called hierarchical loss:

$$L^j = -\frac{1}{\left|P_1^j\right|} \sum_{p \in P_1^j} \log \sigma^{j,p}_{y^j,p} - \frac{1}{\left|P_2^j\right|} \sum_{p \in P_2^j} \log \sigma^{j,p}_{y^j,p} \,, \quad (1)$$

$$L^{total} = \sum_j \lambda^j \cdot L^j + regularizer \,.$$

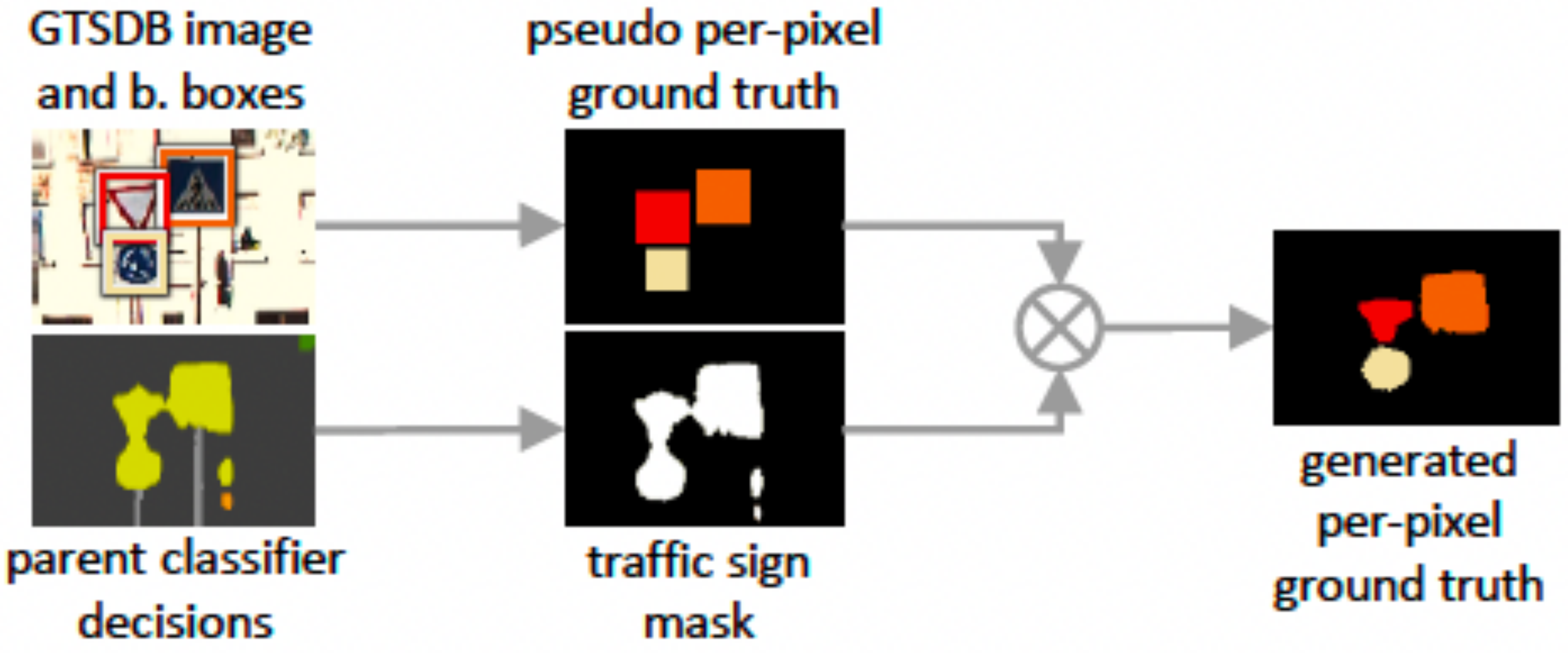


Fig. 3. Online procedure during training for generating per-pixel ground truth from bounding box labels.

1. Training of Convolutional Networks on Multiple Heterogeneous Datasets for Street Scene Semantic Segmentation

Meletis and Dubbelman: https://arxiv.org/pdf/1803.05675.pdf

- TABLE II

FLAT CLASSIFICATION PERFORMANCE BASELINES ON PER-PIXEL ANNOTATED DATASETS.

Tested on	Same dataset			Cross-dataset
	Cityscapes	Vistas	GTSDB	Cityscapes Extended traffic sign classes
mPA (%)	53.6	36.5	25.4	19.1
mIoU (%)	46.2	29.6	17.2	3.0
Trained on	Cityscapes	Vistas	GTSDB	Cityscapes + GTSDB

TABLE III

PERFORMANCE OF OUR COMPLETE HIERARCHICAL CLASSIFICATION APPROACH ON 4 DATASETS.

Tested on	Same dataset			Cross-dataset
	Cityscapes classes	Vistas classes	GTSDB classes	Cityscapes Extended traffic sign classes
mPA (%)	66.6	38.9	57.7	29.7
mIoU (%)	57.3	31.9	41.5	8.3
Trained on	Cityscapes + Vistas + GTSDB			

2. MSeg: A composite Dataset for Multi-domain Semantic Segmentation

Lambert and Liu: https://arxiv.org/pdf/2112.13762.pdf

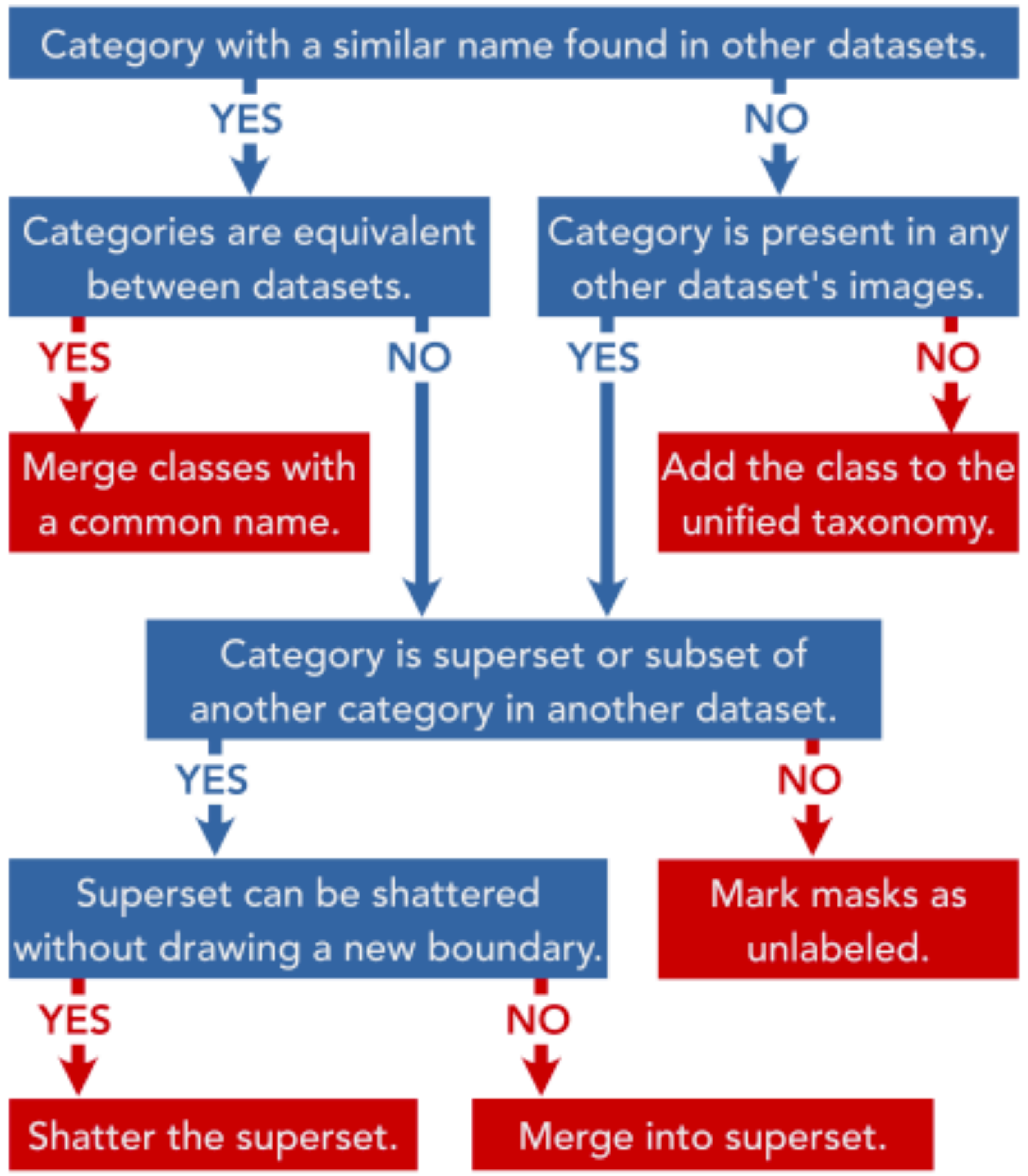


Fig. 2: Procedure for determining the set of categories in the MSeg taxonomy.

TABLE 1: Component datasets in MSeg.

Dataset name	Origin domain	# Images
Training & Validation		
COCO [10] + COCO Stuff [50]	Everyday objects	123,287
ADE20K [11]	Everyday objects	22,210
Mapillary [9]	Driving (Worldwide)	20,000
IDD [13]	Driving (India)	7,974
BDD [14]	Driving (United States)	8,000
Cityscapes [8]	Driving (Germany)	3,475
SUN RGBD [15]	Indoor	5,285
Test		
PASCAL VOC [51]	Everyday objects	1,449
PASCAL Context [52]	Everyday objects	5,105
CamVid [53]	Driving (U.K.)	101
WildDash-v1[1] [21]	Driving (Worldwide)	70
KITTI [54]	Driving (Germany)	200
ScanNet-20 [55]	Indoor	5,436

2. MSeg: A composite Dataset for Multi-domain Semantic Segmentation

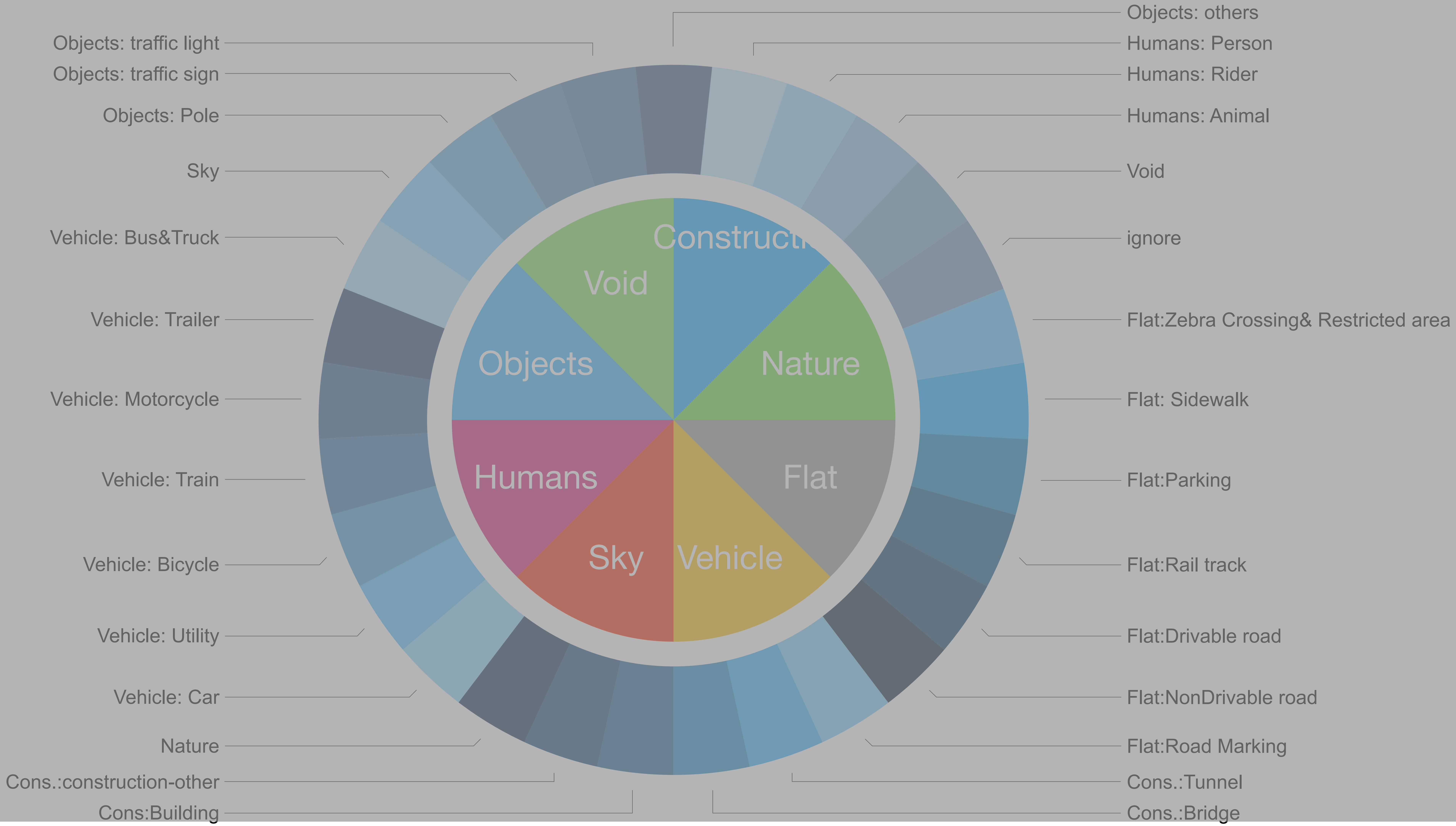
Lambert and Liu: https://arxiv.org/pdf/2112.13762.pdf

-

3. Autonomous driving datasets

- Cityscapes
- nuImages
- CamVid
- Mapillary
- IDD
- A2D2
- GTA-5

3. Autonomous driving datasets:
CityScape+CamVid+GTA5+nuScene+Mapillary+IDD+A2D2



Multi Domain Heterogeneous Dataset Hierarchical Semantic Segmentation Methods

Flat Network with hierarchical loss

- Loss function:
  - Loss of superclass (level 1)
  - Loss of level 2 classes
  - Loss of level 3 classes
  - Total loss = loss_level_1 + loss_level2 + loss_level3

Class Hierarchy

Vehicles	Small vehicles	Car	Pickup truck	
		Van		
		ego_vehicles		
	Large Vehicles	Truck	Train	
		Bus		
	two_wheelers	Motorcycle	Bicycle	
	other_vehicels	Caravan	trailer	Boat
		wheeled_slow	other_vehicles	Plane
Flat	normal_road	road	parking	rail_track
		sidewalk	curb	Pedestrian area
		Dirt track	runway	Crosswalk plain
		Bikelane	Service lane	Path
		Curb cut		
	road_marking	General marking	Zebra marking	manhole
		catch+basin	Pothole	all_road_marking
Construction	building_infra	Building	Grandstand	Stairs
		house	step_stair	stage
		stairway	Skyscraper	Column
		Wall	Bridge	tunnel
		fence	guard_rail	
	other_infra	Pier-dock	Canopy	Fountain
		Bannister	Barrier	awning-sunshade
		Hovel-hut		
Traffic_objects	pole_all	Pole	Utility pole	tower
	Sign	traffic_sign	Billboard	banner
		traffic_sign_frame	traffic_sign back	trade-brand
		picture	Poster	
	glowing_obj	street_light	Traffic light	Light-source

Level 1 super classes

Fine grained classes

	Other_obj	trashcan	phone_booth	cctv_camera
		junction_box	fyre_hydrant	mailbox
		bikerack		
Nature Objects	sky_ Vegetation	Tree	palm_tree	Sky
		plant	Vegetation	
	ground	terrain	Mountain	earth-ground
		soil-ground	hill	Field
		Rock		
	other_entities	Water	Snow	sand
		river	Lake	sea
		Water-fall	Swimming-pool	
Vulnerable_road_user VRU	Animal	ground_animal	Bird	
	Human	Motorcyclist	Bicyclist	other_rider
		Person		
Indoor_obj	furniture	Bed	cabinet	door
		Table	Chair	sofa
		Shelf	Armchair	desk
		drawers	Counter	Ward robe
		Coffee table	Ottoman	stool
		other_furniture	Book-case	box
		Swivel-chair	cradle	Case-showcase
		Seat	Bench	Window
	Bedroom obj	Cushion	Pillow	lamp
		rug	Mirror	curtain
		Flower	Blind	
		Blanket-cover	Apparel	Fireplace
				Ceiling
		Floor	Sconce	
	Bathroom obj	Bathtub	toilet	Towel
		Shower	shower_curtain	
	other_obj	Escalator	Flag	Clock

Level 1 super classes

Fine grained classes

		Radiator	Glass-drinking	plate
		Fan	tray	Vase
		pot	Bag	Tent
		Basket	barrel	Playing-toy
		Conveyerbelt	Chandelier	Book
		pool-billard	ball	Tank
		Sculpture	base	
	Electronic	Monitor	CRT-screen	screen-projector
		Television	Arcade machine	computer
		Screen		
	Kitchen_obj	hood_exhaust	Dishwasher	
		microwave	Food	Oven
		Washer	Buffet	Bottle
		Bar	Kitchen-island	Stove
		counter-top	Refrigerator	Sink

Level 1 super classes

Fine grained classes

Multi Heads Architecture: example

Level 1 Head

- Ignore
- Vehicles
- Flat
- Construction
- Traffic objects
- Nature objects
- VRU objects
- Indoor objects

Level 2 Head: Vehicles

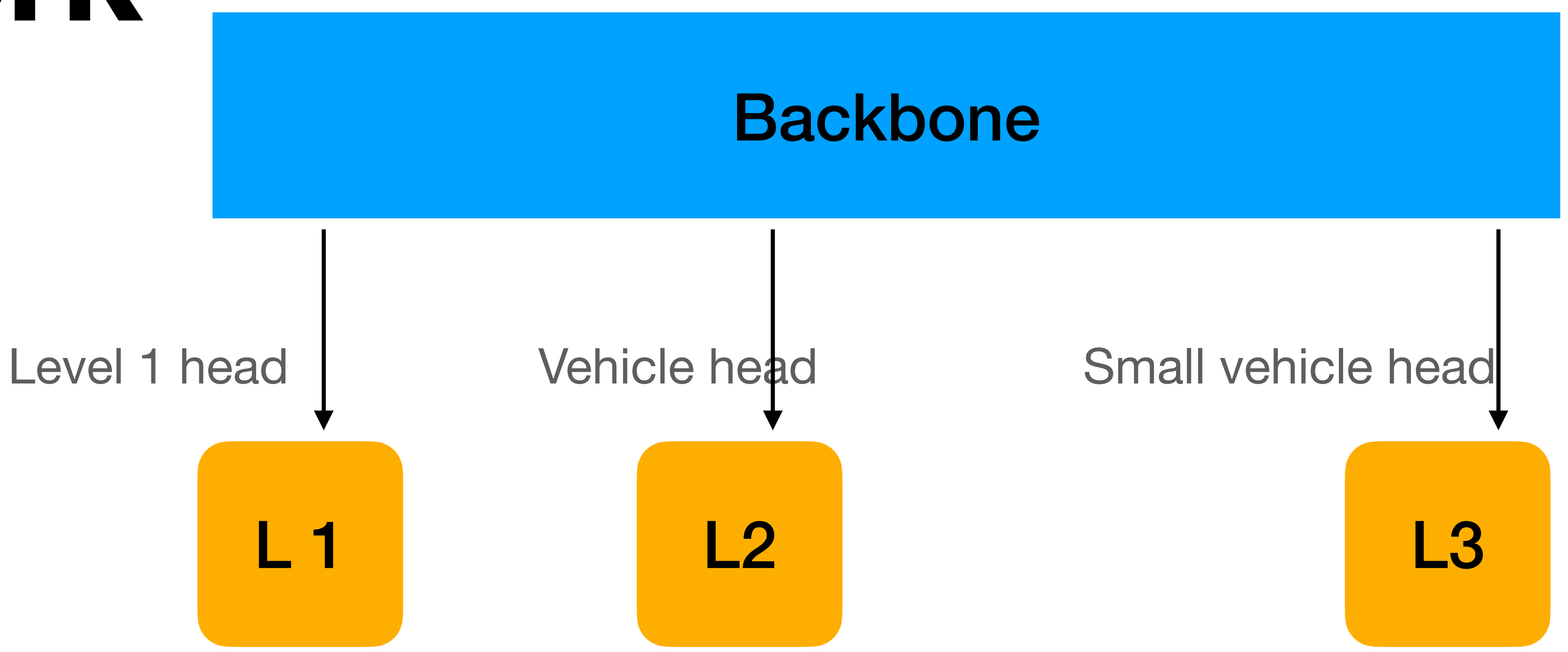
- Ignore
- Small vehicles
- Large vehicles
- Two wheelers vehicles
- Other vehicles

Level 3 Head: Small Vehicles

- Ignore
- Car
- pickup_truck
- Van
- ego_vehicle

Hierarchical Network

- Overview



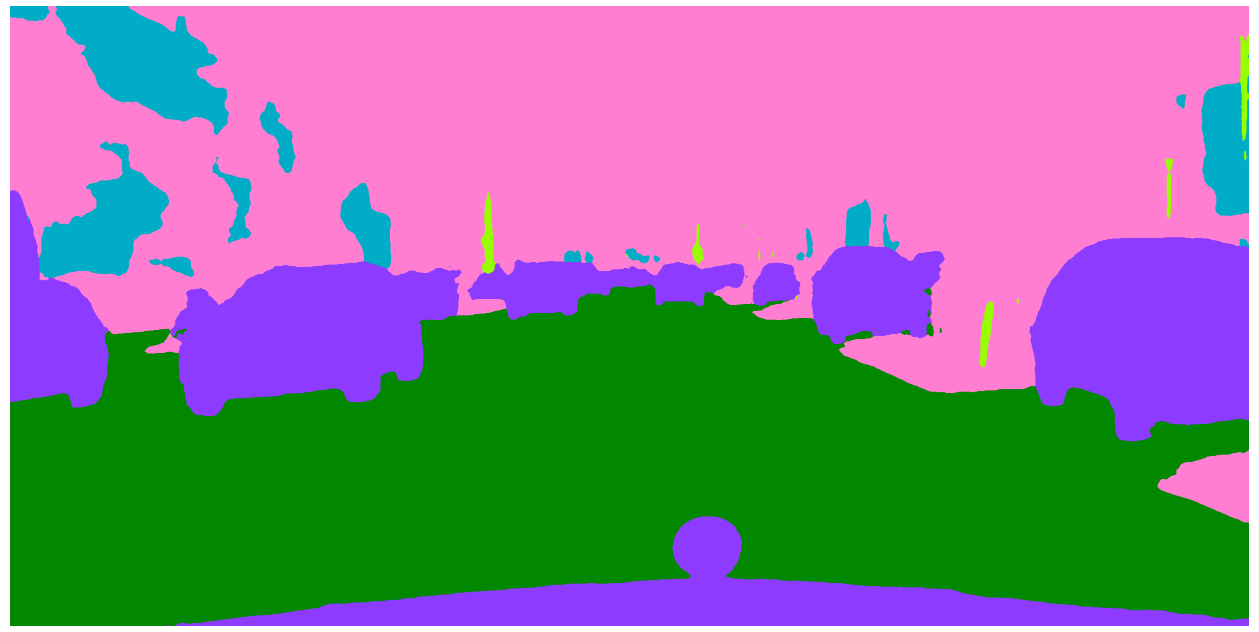
Example for level 1 vehicle class

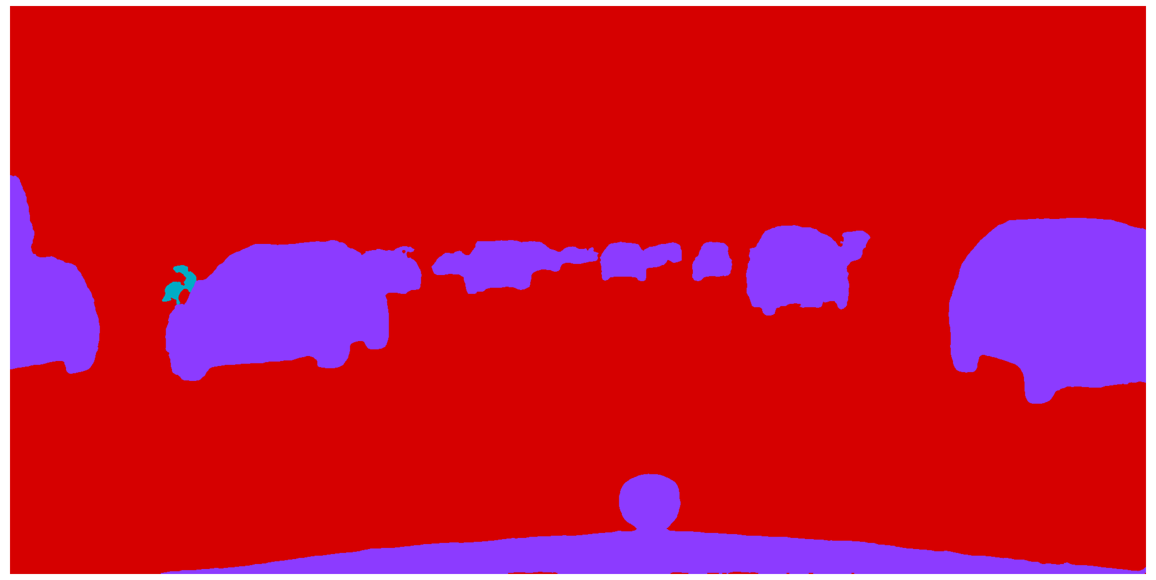
Prediction ->

…Vehicles …

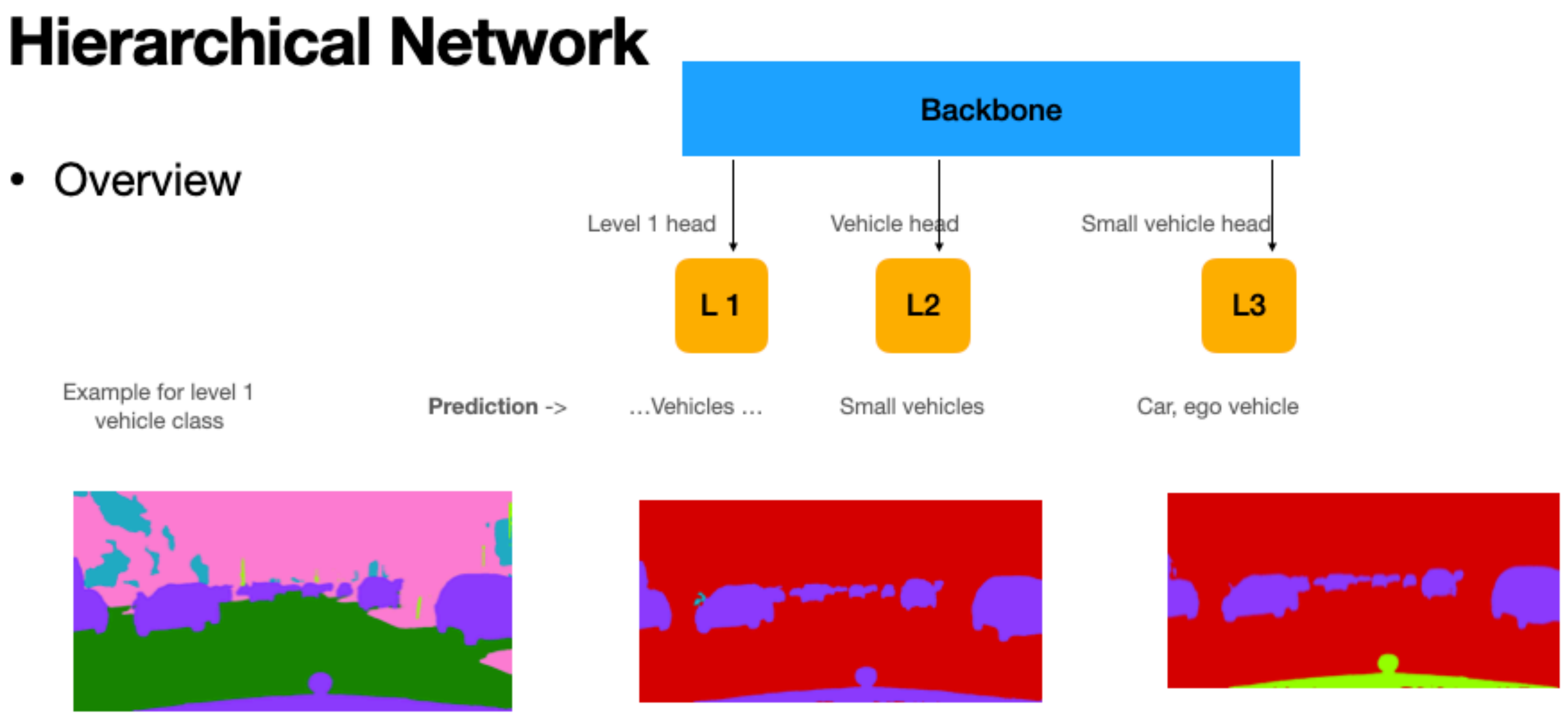
Small vehicles

Car, ego vehicle





Hierarchical N/W

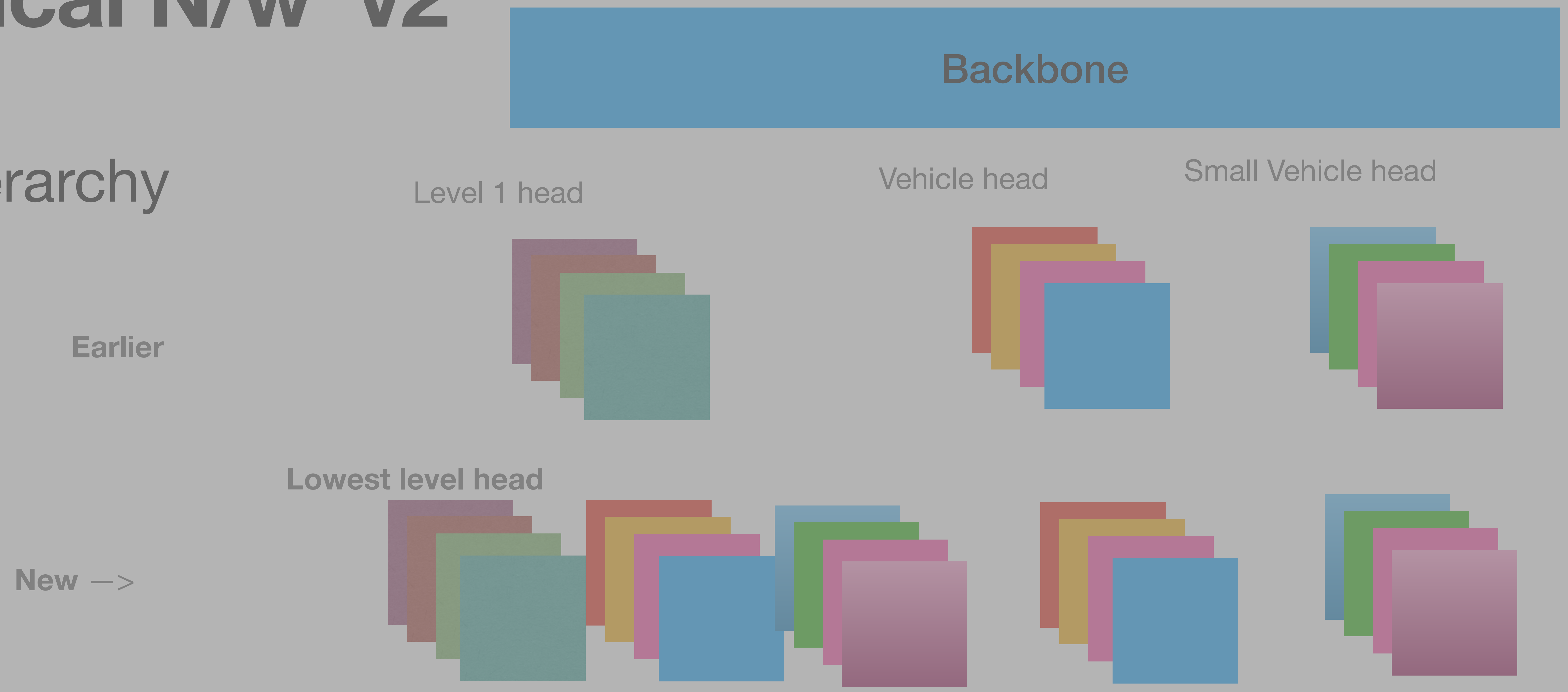


- Loss function: total loss
  - c1 * L1_loss + c2 * L2_loss + c3 * L3loss
- Discussion:
  - Everything relies on L1_head
    - solution: to improve mIOU -> class weights at L1 head
  - Loss function : other possibilites to connect all heads through loss function !!!
  - If L1 predicted negative: output is negative

- .
- .

Hierarchical N/w v2

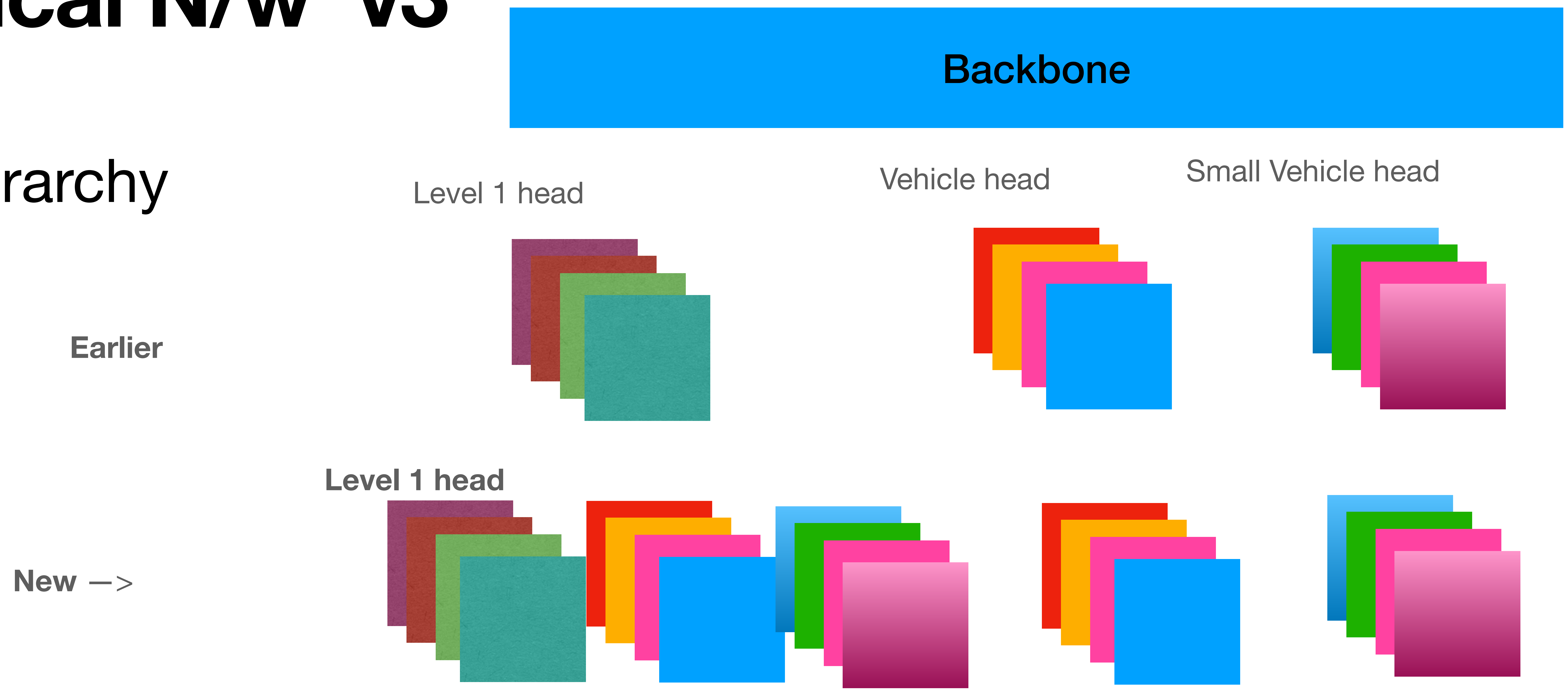
- Inverse Hierarchy



- Advantage:
    - Dependency over L1 head is reduced

Hierarchical N/w v3

- Inverse Hierarchy



- Advantage:
  - Level 1 is enriched with features!!

Results:

Comparison

	RVC2018_Best	RVC2020_1st	RVC2020_2nd	SingleDataset- mIoU	FlatMultiDataset-mIoU-1	FlatMultiDataset-mIoU-2	HierarchicalMultiDataset- mIoU	ImprovedHMDataset-mIoU
Evaluated on:		Test data	Test data	Validation data	Validation data			
Cityscapes	80.2	74.7	**80.7**	0.7399	0.7263		0	
Vistas		**40.4**	34.2	0.382	0.367		0	
Viper		**62.5**	40.7	0.496	**0.655**		0	
Wilddash	**59.1**	**45.4**	35.2	0.452	**0.554***		0	
Ade		31.1	**33.2**	0.386	0.304		0	
Scannet	48.0	**54.6**	48.5	-	**0.596**		0	
KiTTI	69.6	63.9	62.6					