CASNet: Common Attribute Support Network for image instance and panoptic segmentation

Xiaolong Liu, Yuqing Hou, Anbang Yao, Yurong Chen and Keqiang Li
Intel Labs China
School of Vehicle, Tsinghua University, Beijing, China

Hello!

I am Xiaolong Liu

Intel Labs China (ILC) && Tsinghua University, Beijing China

xiaolong.liu@intel.com

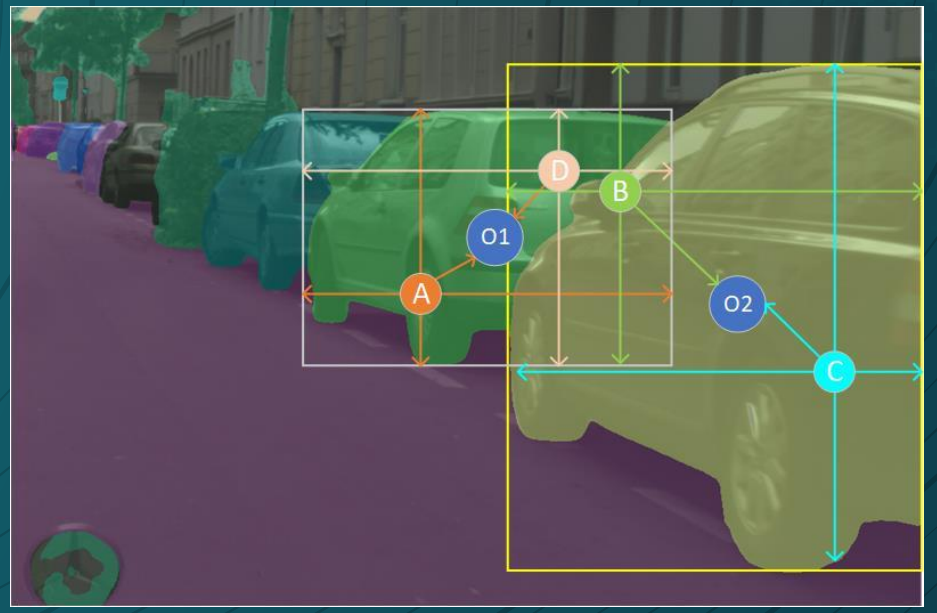
CASNet

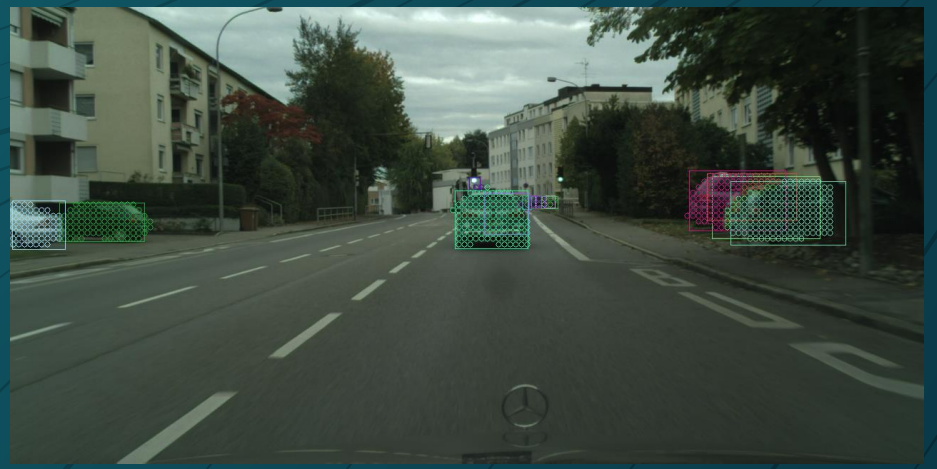
- Biologic vision can recognize an object from inference.
- For example, human can infer a vehicle only from part of the whole view considering most vehicles have attributes in common, such as tires, wheels, wind shield screens and so on.
- As of other objects in image, they also have attributes in common.
- Same instance with same bounding box, geometry, weight or other centers.

CASNet

Common Attribute Support Network

- Simple philosophy
- Work in fully convolutional way
- One-stage
- Non-overlap
- panoptic segmentation can be realized.
- High precision on par with Mask-RCNN
- Correct the bounding boxes
- one step forward from semantic to instance segmentation

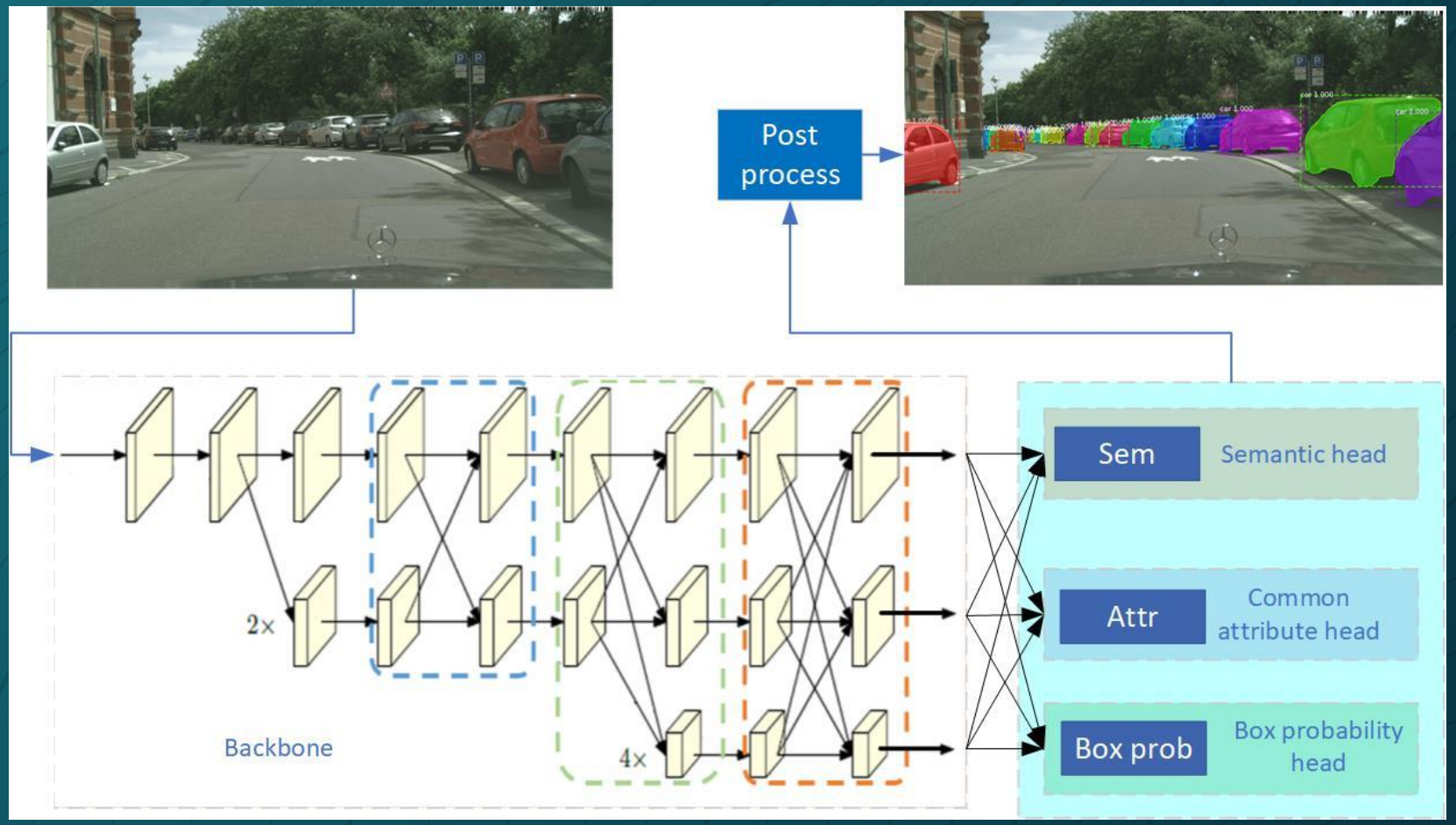




CASNet

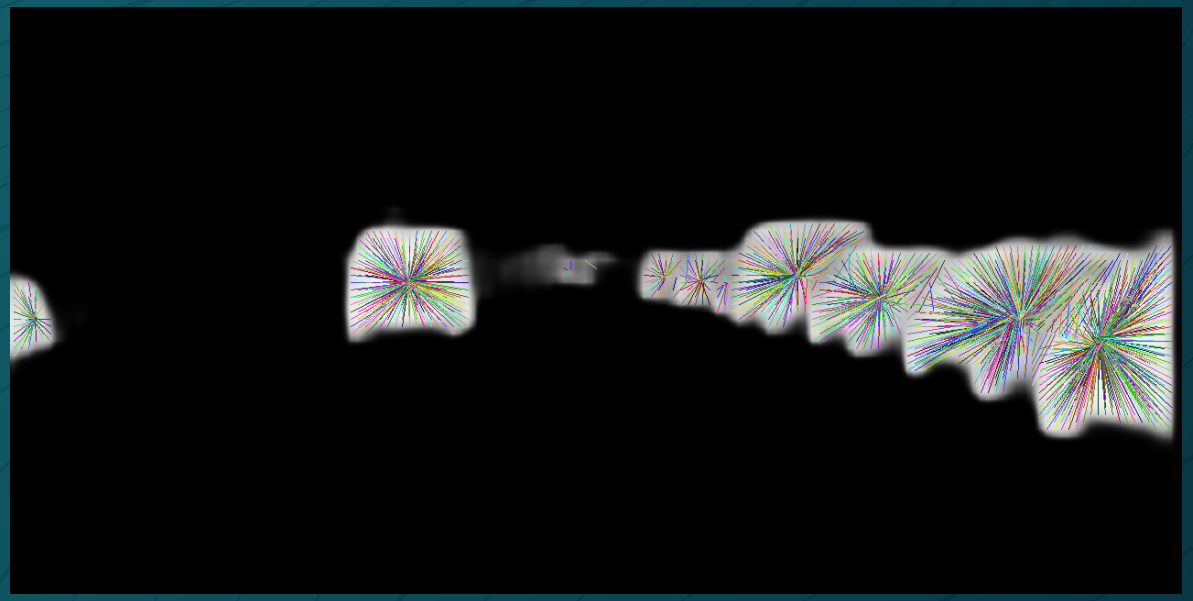
Network structure

- **backbone,**
- **Semantic head,**
- **Common attribute head,**
- **box probability head.**



Post processing

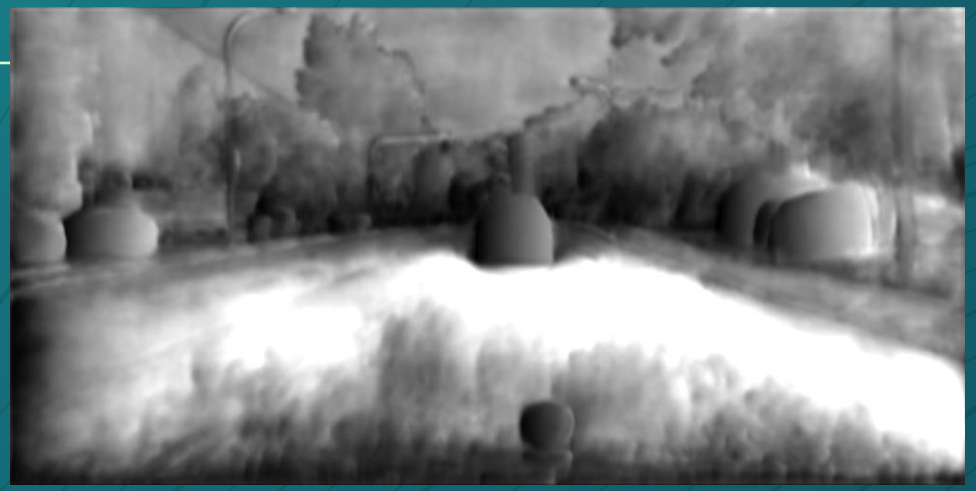
- Choose probability over a threshold from box probability head.
- Common attribution head calculates every pixels' bounding box and centers
- NMS operation, the primary seed boxes will be got.
- Then each pixel's bounding box will give a IOU ratio with all the seed boxes, the most prominent one is the support result.

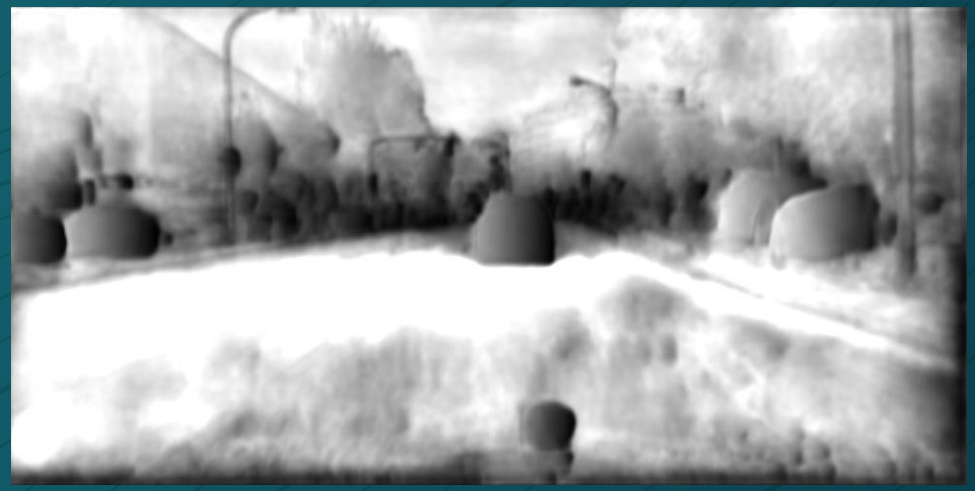


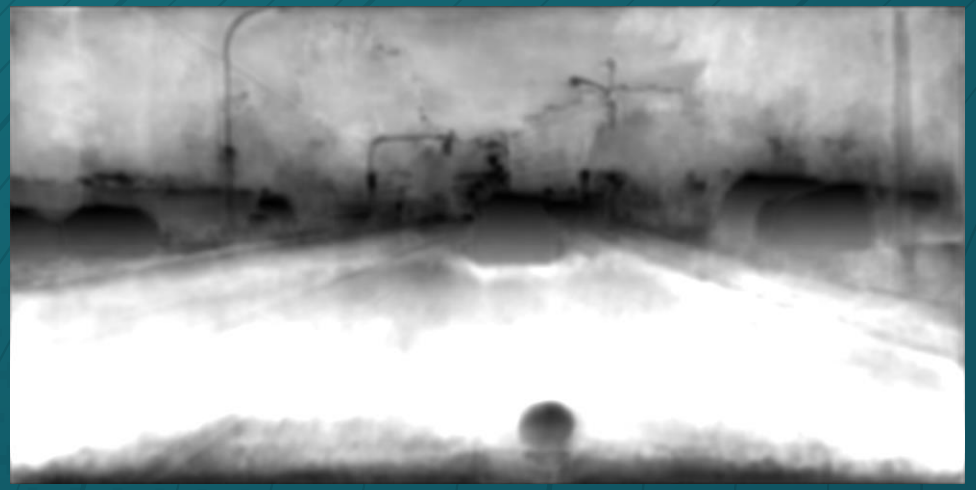
Experiments

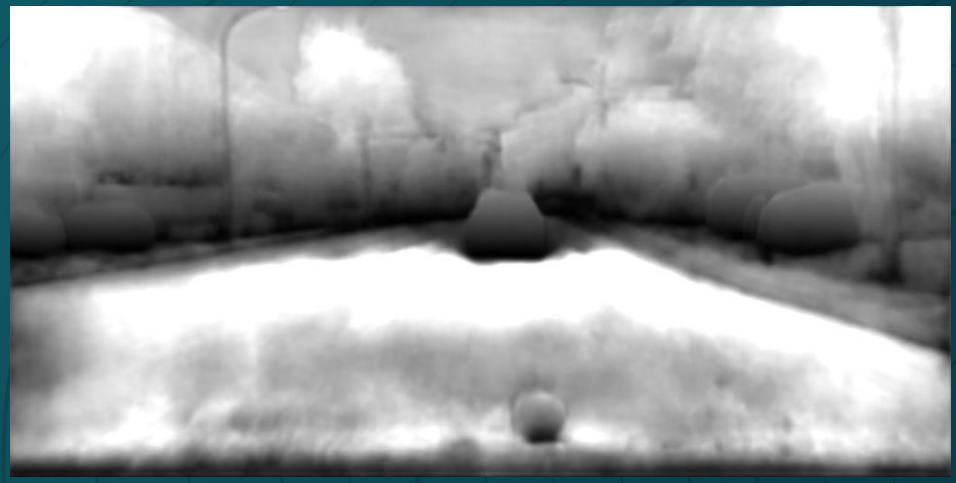
- Cityscapes dataset,
- 5000 fine-annotated split into 2975, 500 and 1525 images for training, validation and testing
- Cityscapes consists of 8 and 11 classes for things and stuff.
- 24 epochs and dropping learning rate by factor of 10 at 18 and 22 epochs.
- HRNet batch size 16, use 8 GPUs with batch size 8.

How well is the prediction of common attributes









Instance segmentation results

Method	mask AP	AP_{50}	mask AP(val)	AP_{50}(val)
SGN[28]	0.250	0.449	0.292	
Mask R-CNN R50[28]	0.262	0.499	0.315	
SegNet[28]	0.295	0.556		
PANet[fine-only][28]	0.318	0.571	0.365	
SAIS[3]	17.4	36.7		
Pixel Encoding[3]	8.9	21.1		
pixelwise[3]	20.0	38.8		
AdaptIS(R50)[33]	-	-	0.323	
AdaptIS(R101)[33]	-	-	0.339	
UPSNet(R50)[40]	-	-	0.339	
Ours(joint training)	0.2527	0.4780	0.328	0.566
Ours(separated training)			0.363	0.599

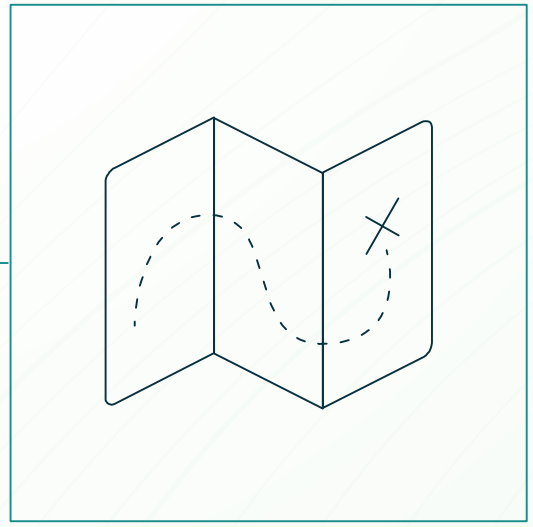
what	AP	AP_50
person	0.506	0.725
rider	0.663	0.778
car	0.572	0.793
truck	0.560	0.603
bus	0.708	0.773
train	0.635	0.645
motorcycle	0.458	0.627
bicycle	0.488	0.705
average	0.574	0.706

Panoptic segmentation results

Models	PQ	SQ	RQ	PQ^{Th}	PQ^{St}	mIoU	AP
CRF + PSPNet [24]	53.8	-	-	42.5	62.1	71.6	28.6
MR-CNN-PSP	58.0	79.2	71.8	52.3	62.2	75.2	32.8
TASCNet[23]	55.9	-	-	50.5	59.8	-	-
UPSNet-M-COCO[41]	61.8	81.3	74.8	57.6	64.8	79.2	39.0
Panoptic FPN[21]	61.2	80.9	74.4	54.0	66.4	80.9	36.4
DeeperLab[42] Xception-71	56.5	-	-	-	-	-	-
AdaptIS X101 [33]	62.0	-	-	64.4	58.7	79.2	36.3
Seamless[29]	60.3	-	-	-	-	77.5	33.6
UTIPS[25]	61.4	81.1	74.7	54.7	66.3	79.5	33.7
MS Net(joint)	59.0	81.0	71.5	47.8	67.2		32.8
MS Net(separate)	66.1	83.3	78.4	53.6	75.2		35.8

Category	All	Things	Stuff
PQ	0.848	0.673	0.974
SQ	0.946	0.907	0.974
RQ	0.891	0.742	1.00

Result for visualization



Thanks!

Any questions?

You can find me at xiaolong.liu@intel.com