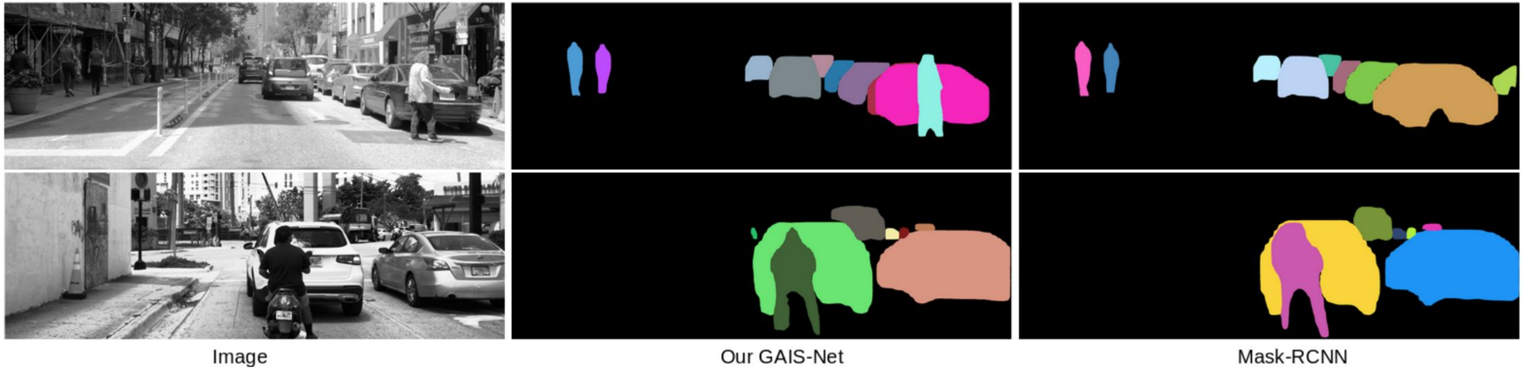
Geometry-Aware Instance Segmentation with Disparity Maps

Cho-Ying Wu[1], Xiaoyan Hu[2], Michael Happold[2], Qiangeng Xu[1], Ulrich Neumann[1]

[1]University of Southern California
[2]Argo AI

Contribution

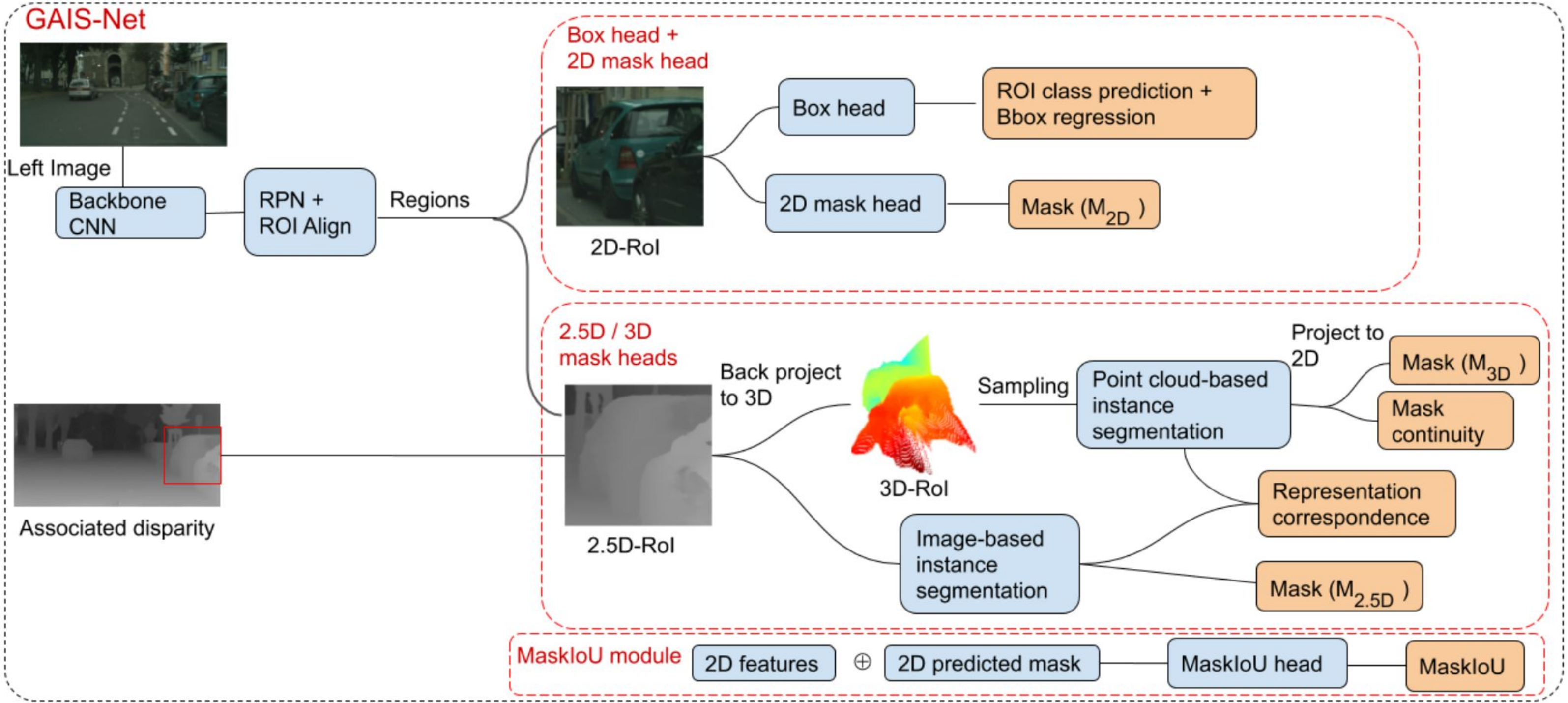
- The first to perform instance segmentation and on imagery by fusing **images** and **disparity** information to regress object masks.
- We collect High-Quality Driving Stereo (HQDS) with f x b 4 times larger than the current best Cityscapes



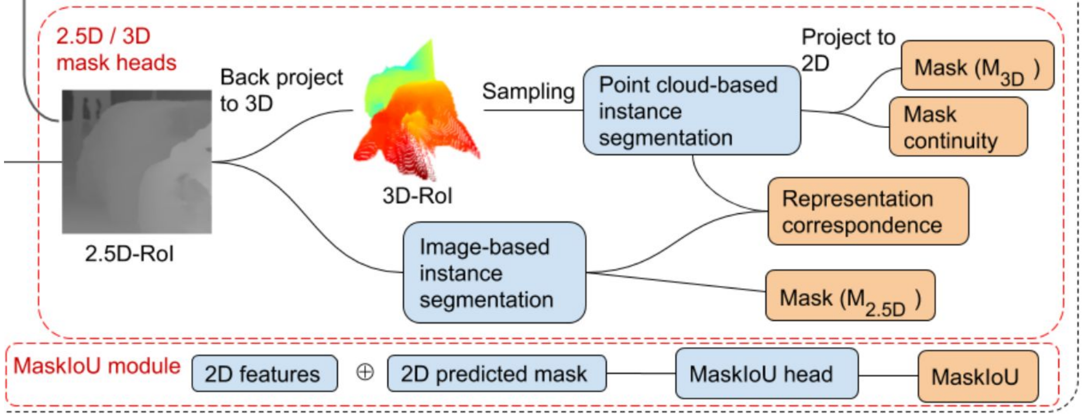
Method

Our Geometry-Aware Instance Segmentation Network (GAIS-Net) pipeline.

Based on Mask-RCNN, we introduce geometry at ROI heads.



Method



Representations:

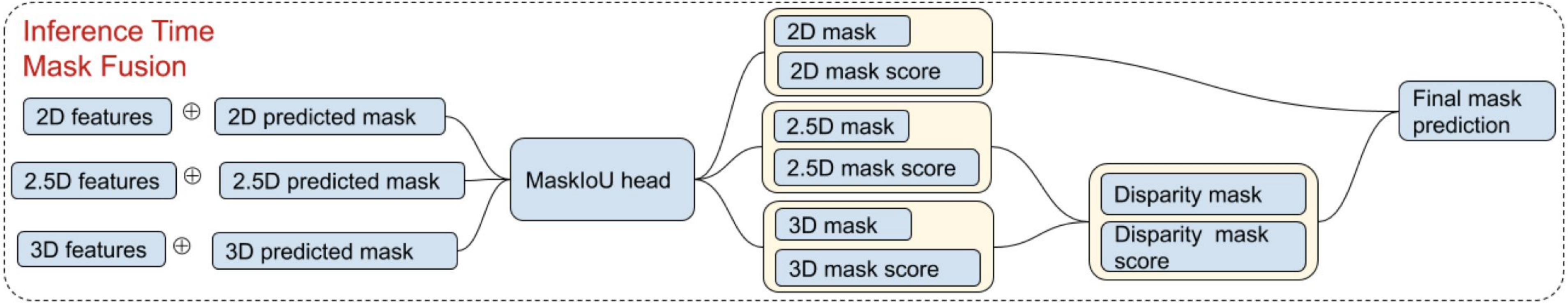
- 2D images from cameras
- 2.5D disparity from stereo cameras
- 3D pseudo-lidar representation

Mask Fusion:

- Scoring-based methods
- MaskIoU scores to evaluate mask quality.

Method

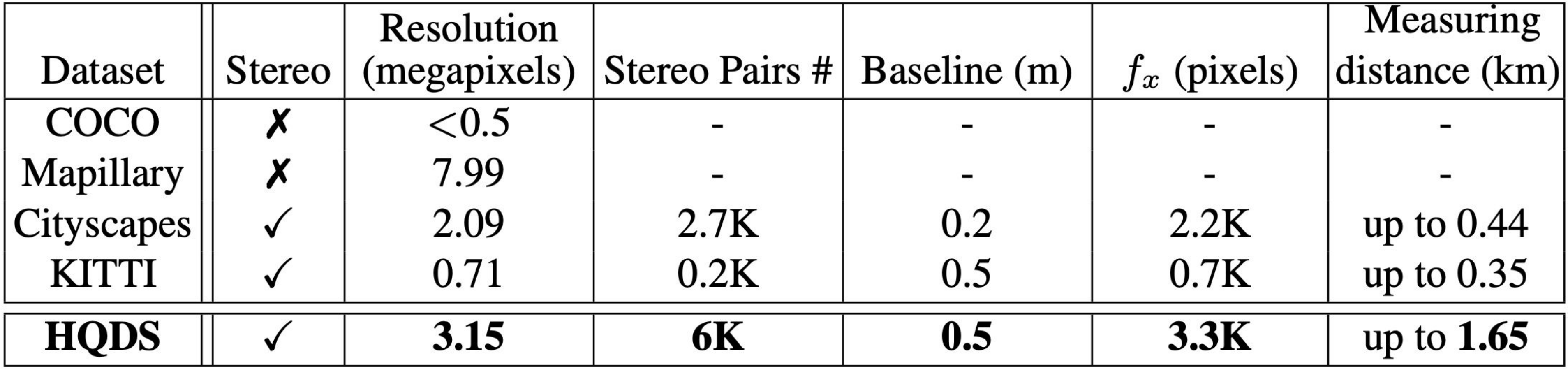
Inference time mask fusion



- 2.5D and 3D masks are combined first.
- The final prediction is by combining images and geometric masks.

Dataset

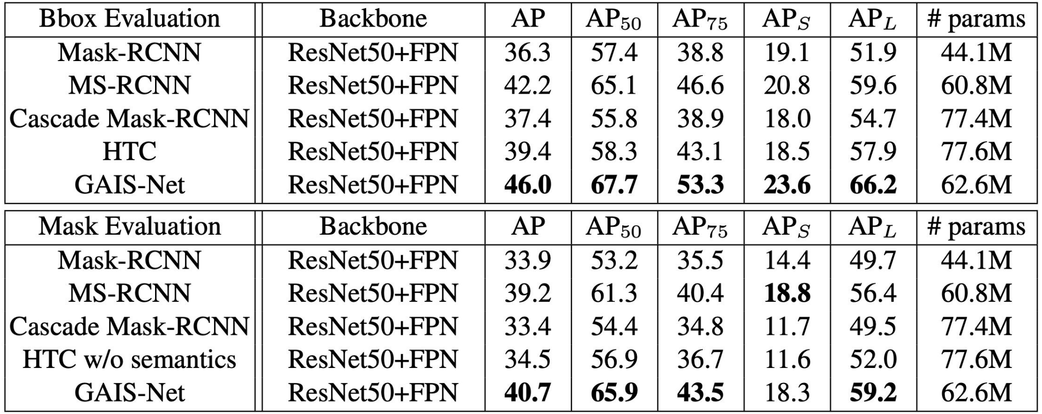
We collect High Quality Driving Stereo. The comparison with other dataset is as follows.



Dataset	Stereo	Resolution (megapixels)	Stereo Pairs #	Baseline (m)	f_x (pixels)	Measuring distance (km)
COCO	✗	<0.5	-	-	-	-
Mapillary	✗	7.99	-	-	-	-
Cityscapes	✓	2.09	2.7K	0.2	2.2K	up to 0.44
KITTI	✓	0.71	0.2K	0.5	0.7K	up to 0.35
HQDS	✓	**3.15**	**6K**	**0.5**	**3.3K**	up to **1.65**

Results

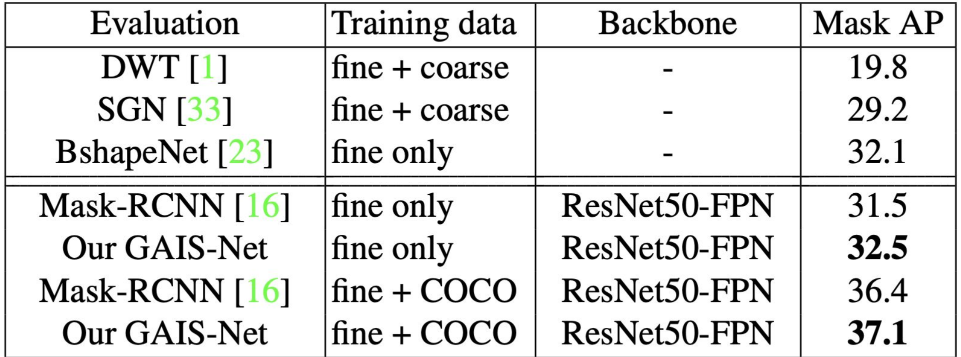
1. HQDS



Bbox Evaluation	Backbone	AP	AP_{50}	AP_{75}	AP_S	AP_L	# params
Mask-RCNN	ResNet50+FPN	36.3	57.4	38.8	19.1	51.9	44.1M
MS-RCNN	ResNet50+FPN	42.2	65.1	46.6	20.8	59.6	60.8M
Cascade Mask-RCNN	ResNet50+FPN	37.4	55.8	38.9	18.0	54.7	77.4M
HTC	ResNet50+FPN	39.4	58.3	43.1	18.5	57.9	77.6M
GAIS-Net	ResNet50+FPN	**46.0**	**67.7**	**53.3**	**23.6**	**66.2**	62.6M

Mask Evaluation	Backbone	AP	AP_{50}	AP_{75}	AP_S	AP_L	# params
Mask-RCNN	ResNet50+FPN	33.9	53.2	35.5	14.4	49.7	44.1M
MS-RCNN	ResNet50+FPN	39.2	61.3	40.4	**18.8**	56.4	60.8M
Cascade Mask-RCNN	ResNet50+FPN	33.4	54.4	34.8	11.7	49.5	77.4M
HTC w/o semantics	ResNet50+FPN	34.5	56.9	36.7	11.6	52.0	77.6M
GAIS-Net	ResNet50+FPN	**40.7**	**65.9**	**43.5**	18.3	**59.2**	62.6M

2. Cityscapes



Evaluation	Training data	Backbone	Mask AP
DWT [1]	fine + coarse	-	19.8
SGN [33]	fine + coarse	-	29.2
BshapeNet [23]	fine only	-	32.1
Mask-RCNN [16]	fine only	ResNet50-FPN	31.5
Our GAIS-Net	fine only	ResNet50-FPN	**32.5**
Mask-RCNN [16]	fine + COCO	ResNet50-FPN	36.4
Our GAIS-Net	fine + COCO	ResNet50-FPN	**37.1**

Results

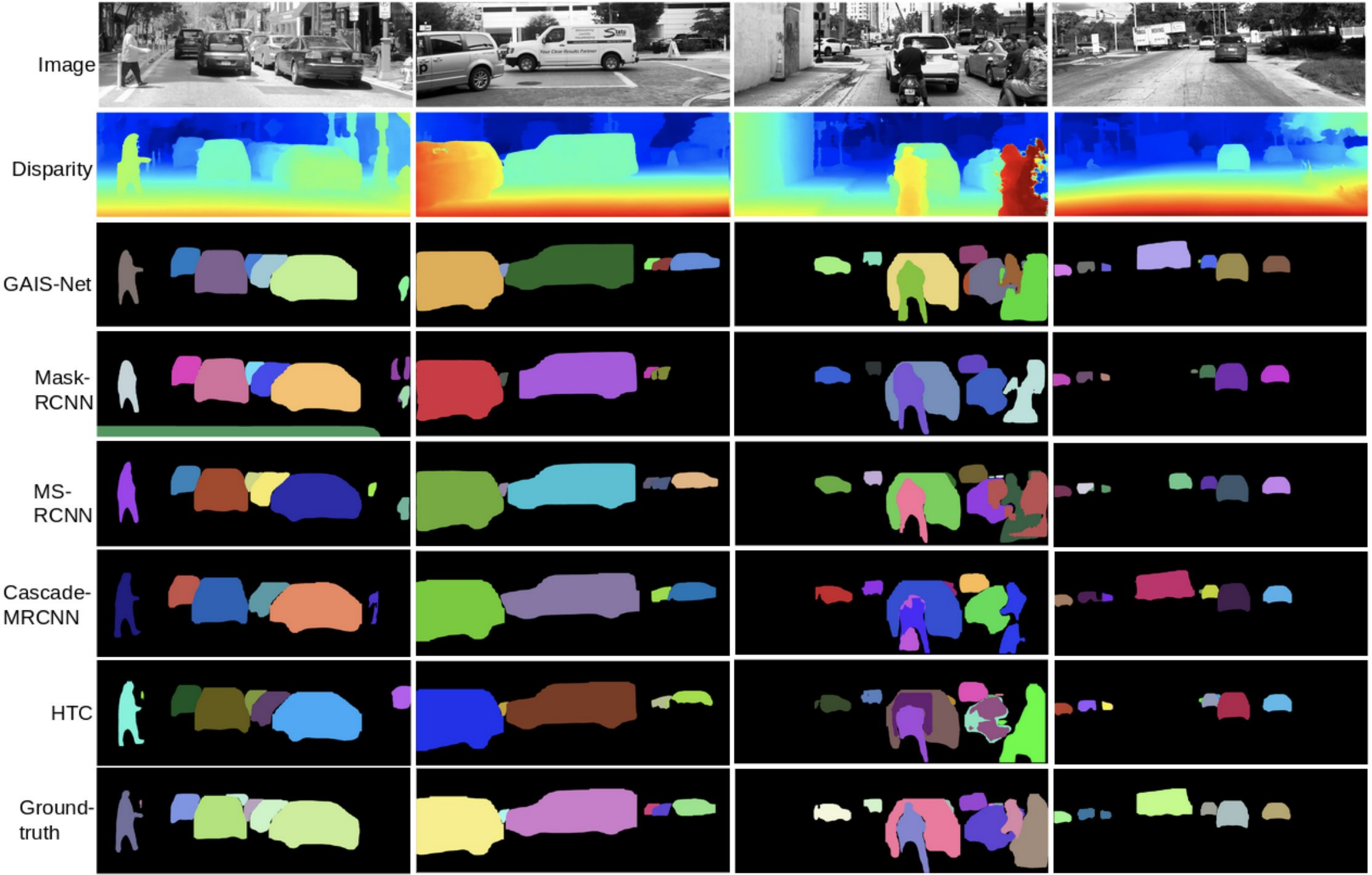


Image
Disparity
GAIS-Net
Mask-RCNN
MS-RCNN
Cascade-MRCNN
HTC
Ground-truth