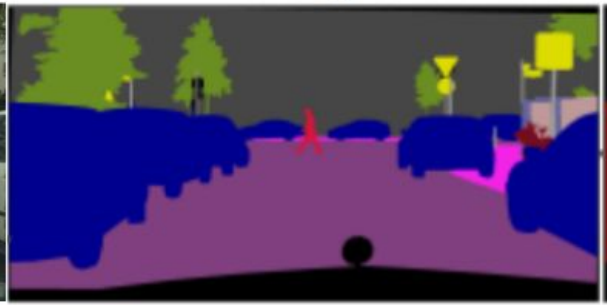
# Panoptic-Depth Color Map for Combination of Depth and Image Segmentation

Jia-Quan Yu(游家權), Soo-Chang Pei(貝蘇章)

# Introduction - Task Definition

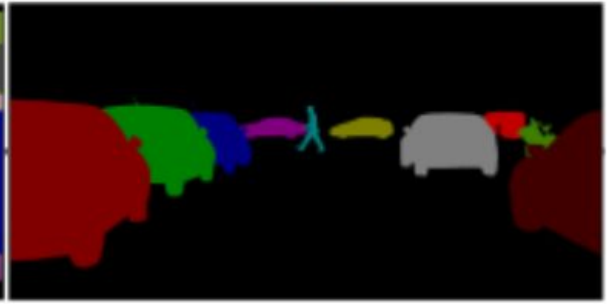
**Camera Image**

**Semantic Segmentation**



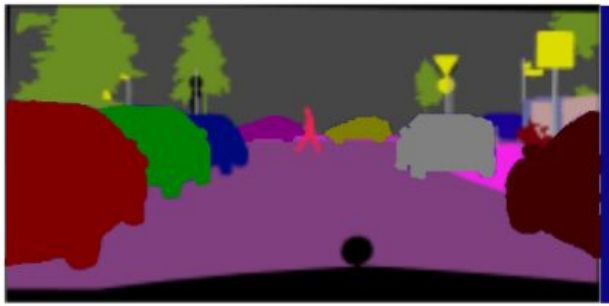
- pixel classification

**Instance Segmentation**



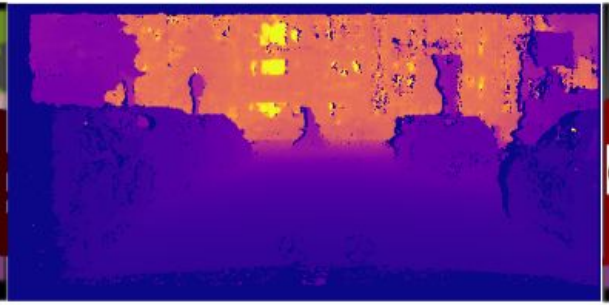
- object detection with mask

**Panoptic Segmentation**



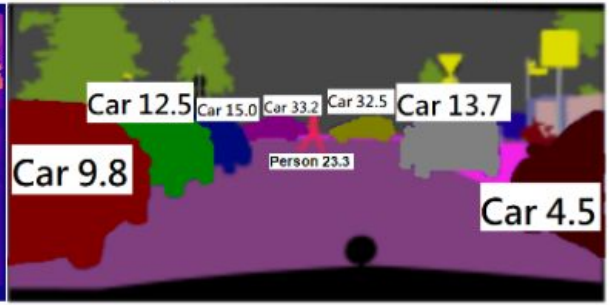
- semantic + instance seg.

**Depth Estimation**



- distance to camera center

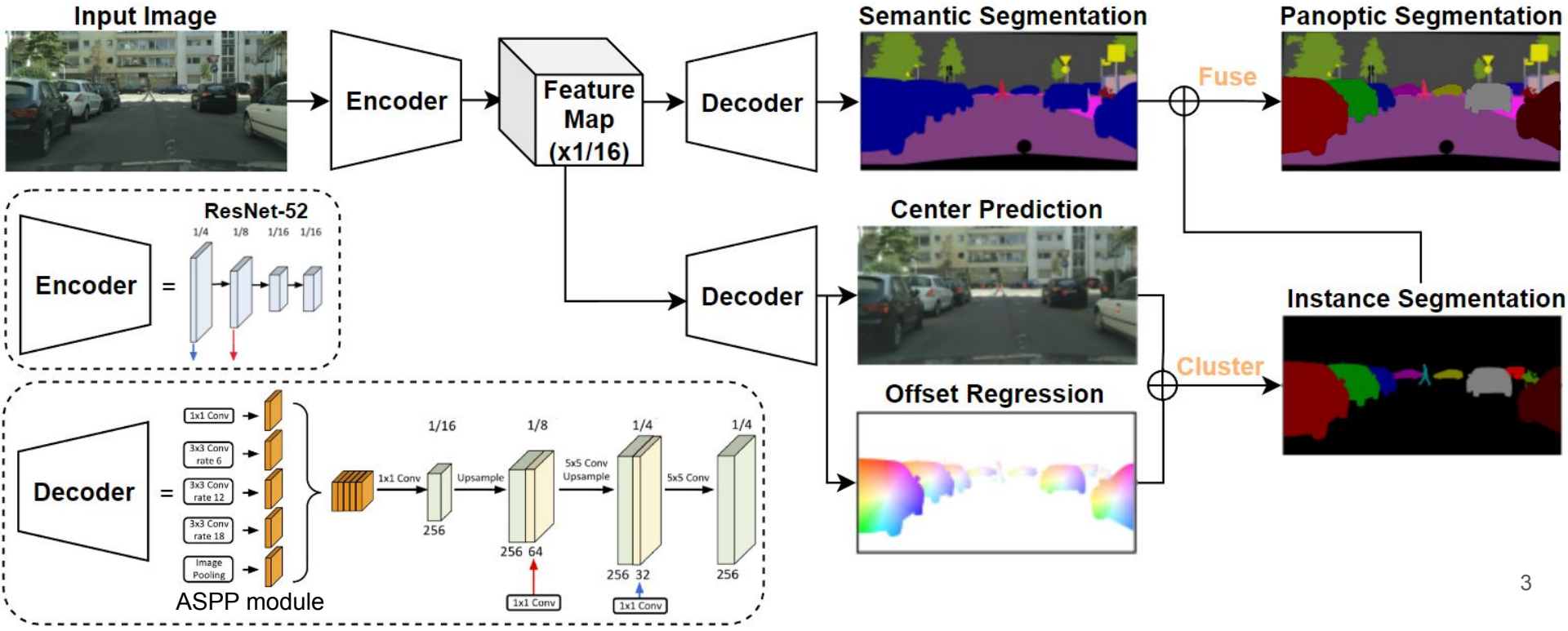
**Segmentation with Depth**




- panoptic seg. + depth estimation

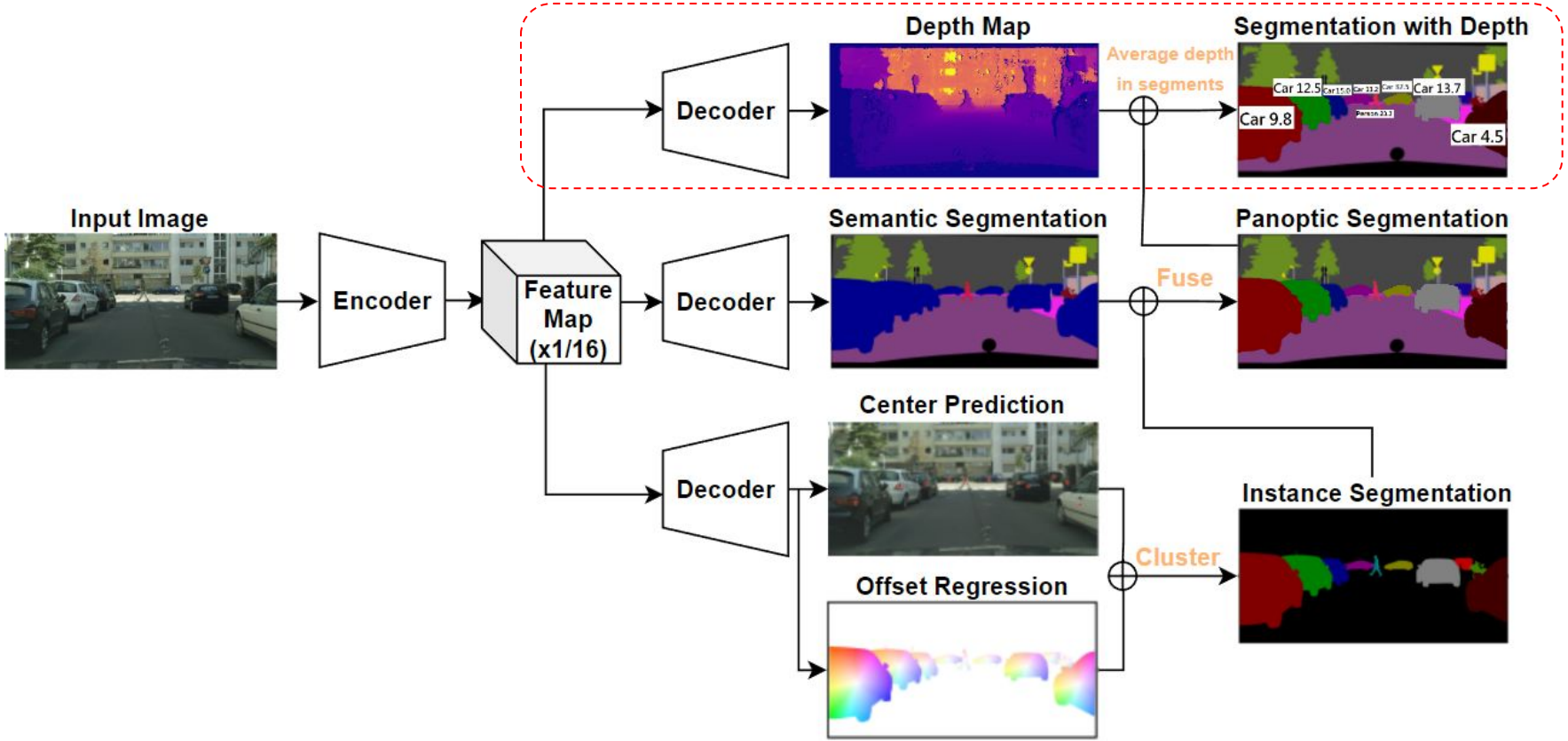
# Related Work - Panoptic Segmentation

- Network: Panoptic-FPN[1], UPSNet[2], and **Panoptic-DeepLab**[3]



# Proposed Method - Panoptic-DepthLab Architecture

Add depth estimation branch to Panoptic-DeepLab

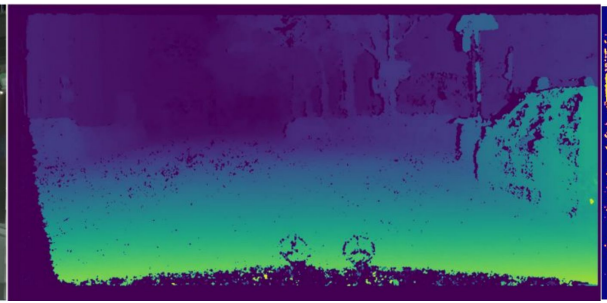


# Experiment Settings

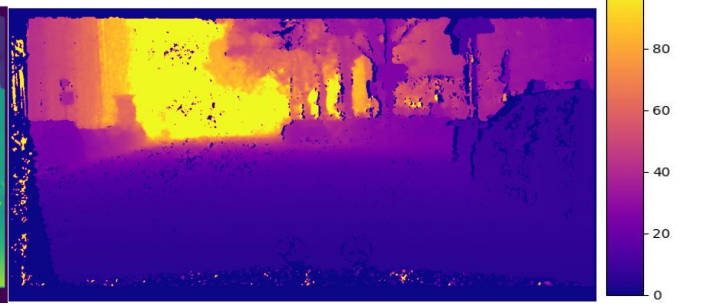
- Dataset: **Cityscape**[4] consists of 2,975 images in training dataset and 500 image in validation dataset.
- We use pre-train weight from Panoptic-DeepLab, which train for 90k iteration.
- We fine-tune for another 10k iteration with depth estimation ground truth.
- The depth estimation ground truth is derived from Cityscape disparity map.

CityScape Image

Disparity Map



Depth Estimation Ground Truth



# Experiment Result

- Panoptic-DepthLab performs slightly better than others works.
- For depth estimation branch, we try DORN[5] and L1 loss. L1 loss perform better in our experiment.

**Panoptic Segmentation Evaluation**

Panoptic Quality

$$PQ = \overbrace{\frac{\sum_{(p,g)TP} IoU(p,g)}{|TP|}}^{\text{Segmentation Quality(SQ)}} \times \overbrace{\frac{|TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}}^{\text{Recognition Quality(RQ)}}$$

| Methods | PQ | SQ | RQ |
|---|---|---|---|
| Panoptic-FPN | 55.4 | 77.9 | 69.3 |
| UPSNet | 60.1 | 80.3 | **73.5** |
| Panoptic-DepthLab | **60.3** | **81.5** | 72.9 |

**Depth Estimation Evaluation**

$$\text{absErr} = \frac{1}{N}\sum_{i=1}^{N} |\frac{d_i^* - d_i}{d_i}|$$

| Loss | sqErr | absErr | IRMSE | SILog |
|---|---|---|---|---|
| DORN | 0.72 | 0.69 | 44.64 | 22.31 |
| L1 | **0.63** | **0.60** | **34.57** | **18.58** |

# Experiment Result - 1

# Experiment Result - 2

# Experiment Result - 3

# Conclusion

- We are the first to combine the tasks of image segmentation and depth estimation into a unified task that can be essential to autonomous driving.
- We propose **Panoptic-DepthLab**, an extension of Panoptic-DeepLab that incorporates a depth estimation branch. Our architecture shows that depth and segmentation branch can share the same extracted feature map.
- Our method achieves high-quality **segmentation results with depth** on the Cityscape dataset, while we use **color map** to show the criticality of each object

# Reference

[1] Panoptic-FPN, https://arxiv.org/abs/1901.02446

[2] UPSNet: A Unified Panoptic Segmentation Network, https://arxiv.org/abs/1901.03784

[3] Panoptic-DeepLab: A Simple, Strong, and Fast Baseline for Bottom-Up Panoptic Segmentation, https://arxiv.org/abs/1911.10194

[4] The Cityscapes Dataset for Semantic Urban Scene Understanding, https://arxiv.org/abs/1604.01685

[5] Deep Ordinal Regression Network for Monocular Depth Estimation, https://arxiv.org/abs/1806.02446

[6] Panoptic-DepthLab code in Github, https://github.com/KenYu910645/detectron2