ORA3D: Overlap Region Aware Multi-view 3D Object Detection

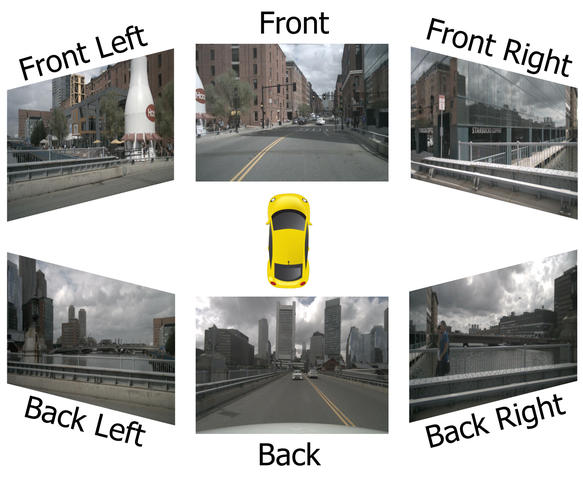
Wonseok Roh[1], Gyusam Chang[1], Seokha Moon[1], Giljoo Nam[2], Chanyoung Kim[1], Younghyun Kim[3], Sanpil Kim[1*], and Jinkyu Kim[1*]

[1]Korea University, [2]KAIST, [3]Hyundai Motor Company

BMVC 2022

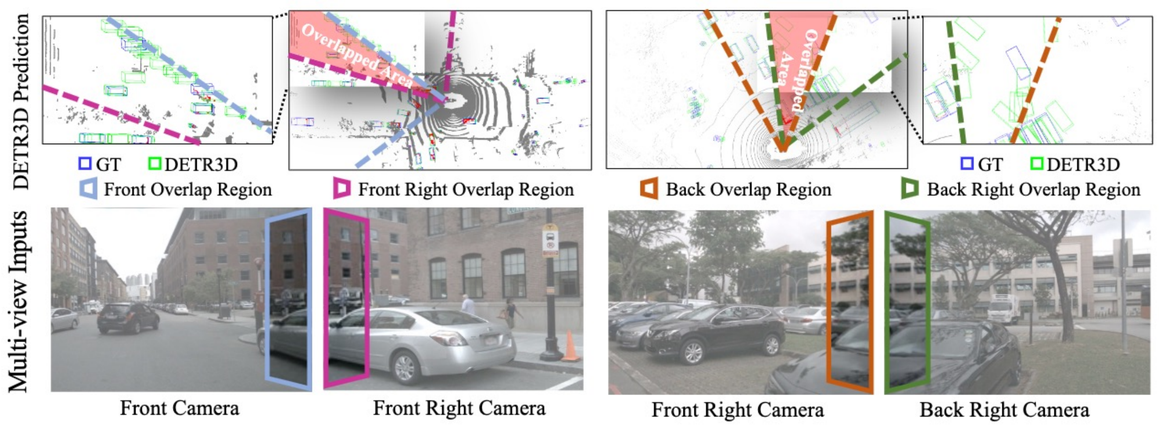
Project Page : https://kuai-lab.github.io/bmvc2022ora3d

Multi-view Camera System



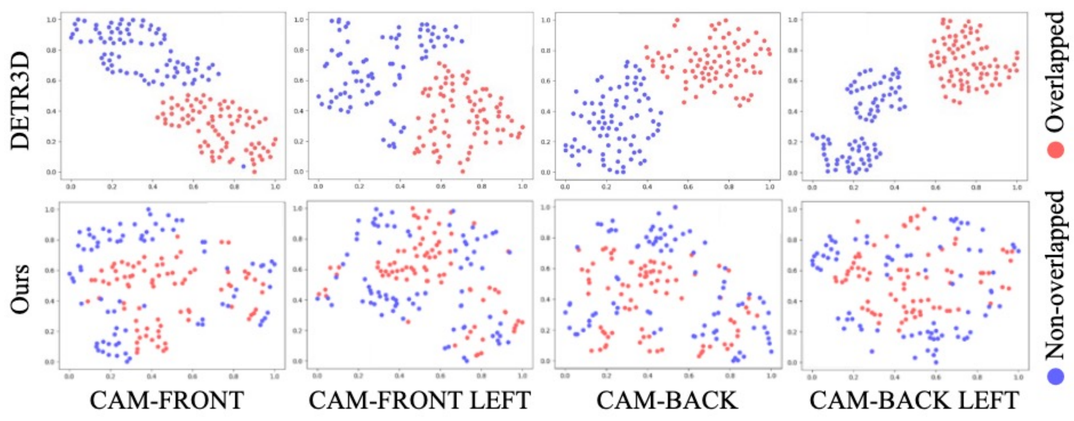
- Alternative balanced option of the monocular and stereo vision systems.
- Potentially replace LiDAR.
- Adjacent cameras have a strong association. (**Overlap Region**)

Limitations of Previous Work (DETR3D)



- We found that the network (**without explicit guidance**) does not totally use the **geometric potentials** of multi-view camera systems.
- **Multiple false positives** in the overlapped regions.

Domain Shift Issue (Non-Overlap vs. Overlap)



- We empirically observe that DNN has a **strong inductive bias.**
- As we clearly see in the first row (DETR3D), the features of each two groups form **distinguishable clusters**.

Experimental Results

Performance Comparison with SOTA on the nuScenes dataset

Model	Mono|Multi	Backbone	NDS(↑)	mAP(↑)
CenterNet [54]	Mono	DLA34	0.328	0.306
FCOS3D [29]	Mono	ResNet101	0.415	0.343
PGD [30]	Mono	ResNet101	0.428	0.369
DETR3D [43]	Multi	ResNet101	0.425	0.346
DETR3D† [43]	Multi	ResNet101	0.434	0.349
Ours	Multi	ResNet101	**0.445**	**0.367**

(a) Validation set

Model	Mono|Multi	Backbone	NDS(↑)	mAP(↑)
FCOS3D [29]	Mono	ResNet101	0.428	0.358
PGD [30]	Mono	ResNet101	0.448	0.386
DD3D* [27]	Mono	V2-99	0.477	0.418
DETR3D‡ [43]	Multi	V2-99	0.479	0.412
BEVDet‡ [10]	Multi	V2-99	0.482	0.422
Ours‡	Multi	V2-99	**0.489**	**0.423**

(b) Test set

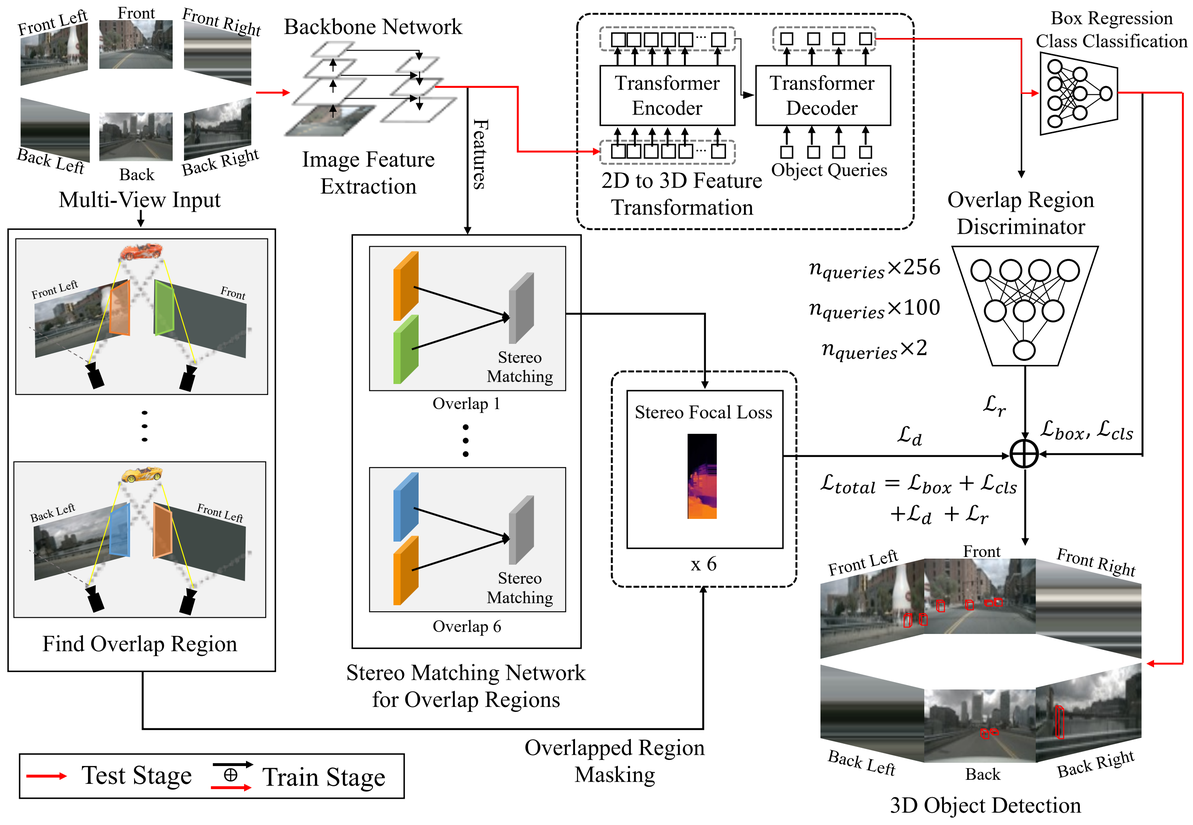
Average Precision for each object class on the nuScenes test set

Model	Car	Truck	Bus	Trailer	C.V	Ped.	Motor.	Bicycle	T.C	Barrier	mAP
DETR3D [43]	0.603	0.333	0.290	**0.358**	0.170	0.455	0.413	0.308	0.627	0.565	0.412
Ours	**0.609**	**0.338**	**0.323**	0.347	**0.174**	**0.467**	**0.420**	**0.311**	**0.649**	**0.589**	**0.423**

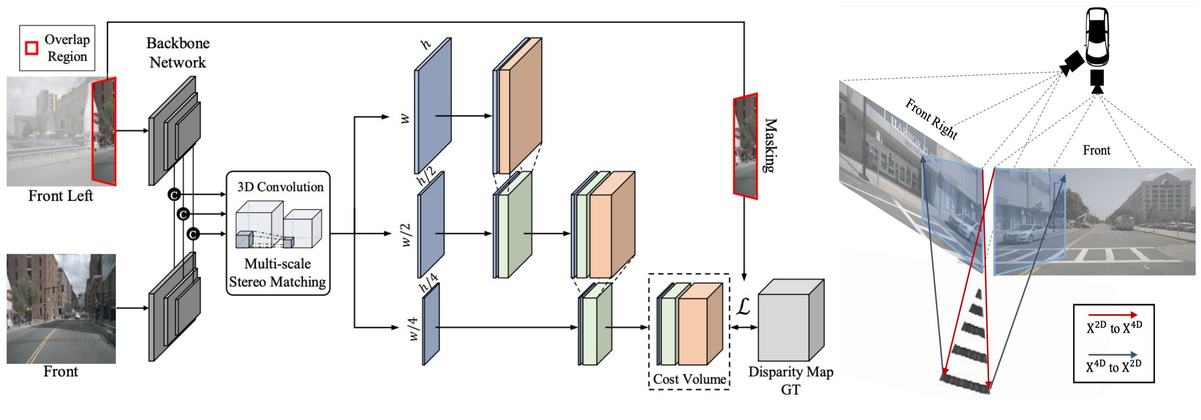
Performance Comparison in Overlap Region

Model	Mono|Multi	NDS (↑)	mAP (↑)	mATE (↓)	mASE (↓)	mAOE (↓)	mAVE (↓)	mAAE (↓)
FCOS3D [29]	Mono	0.317	0.213	0.841	**0.276**	0.604	1.122	**0.173**
DETR3D [43]	Multi	0.356	0.231	0.825	0.280	0.400	0.863	0.223
Ours	Multi	**0.408**	**0.264**	**0.677**	0.280	**0.361**	**0.746**	0.181

Main Architecture



(1) Stereo Matching Network for Weak Depth Supervision



- **Stereo Disparity Estimation** head reconstructs a dense disparity map with **overlap region pairs** of neighboring cameras.
- Although this region is relatively small, it serves as a **geometric link** between two images.

(2) Adversarial Overlap Region Discriminator

- We propose to use a regularizer to explicitly minimize the representational gap between **non-overlap regions vs. overlap regions**.
- We constrain the object detection head from learning region-invariant information via an adversarial framework using **Gradient Reversal Layer**.

(3) Total Loss

$$\mathcal{L}_{total} = \lambda_{cls}\mathcal{L}_{cls} + \lambda_{box}\mathcal{L}_{box} + \lambda_{d}\mathcal{L}_{d} - \lambda_{r}\mathcal{L}_{r}$$

Summary

- We report that existing works often neglect properly dealing with **objects in the overlap region**, which limits fully using the **geometric potentials** of multi-view camera systems, causing performance degradation.
- We propose to use outputs from the traditional **stereo disparity estimation** model on the **overlap region** and apply them as **weak supervision** to improve the detection accuracy over the overlap region. We empirically find that this supervision significantly improves the overall detection accuracy.
- We introduce an **overlap region discriminator** that adversarially learns to minimize the **covariate shift** between objects from non-overlap regions vs. from overlap regions.