CleverDistiller: Simple and Spatially Consistent Cross-modal Distillation

Hariprasath Govindarajan*[1,2], Maciej K. Wozniak*[1,3], Marvin Klingner[1],
Camille Maurice[1] , B Ravi Kiran[1] , Senthil Yogamani[1]

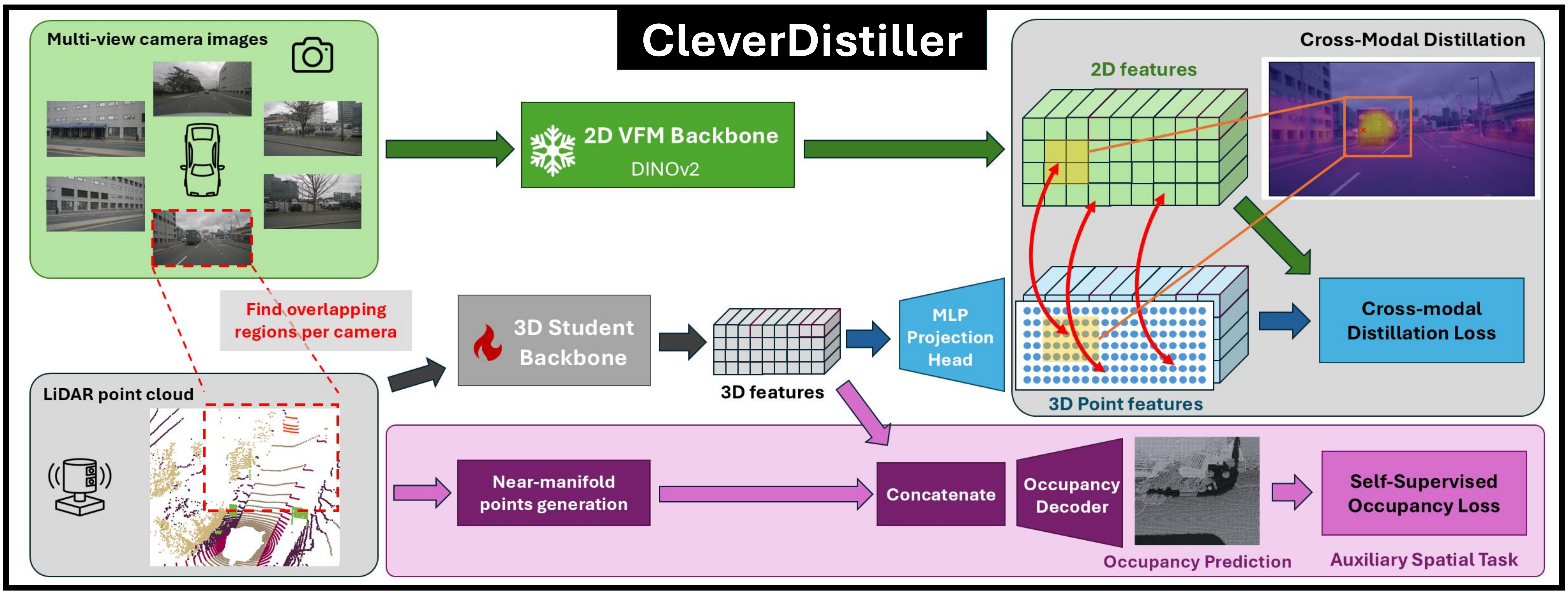
[1] Qualcomm Technologies, Inc. [2] Linköping University, Sweden [3] KTH Royal Institute of Technology, Sweden

Cross-modal Knowledge Distillation (CMKD): Vision → LiDAR

❶ Distilled models learn features common to both modalities.
❷ Misses out on features specific to the LiDAR modality.
❸ Use complicated training setups.

Simple CMKD methods can excel, with *clever* design choices

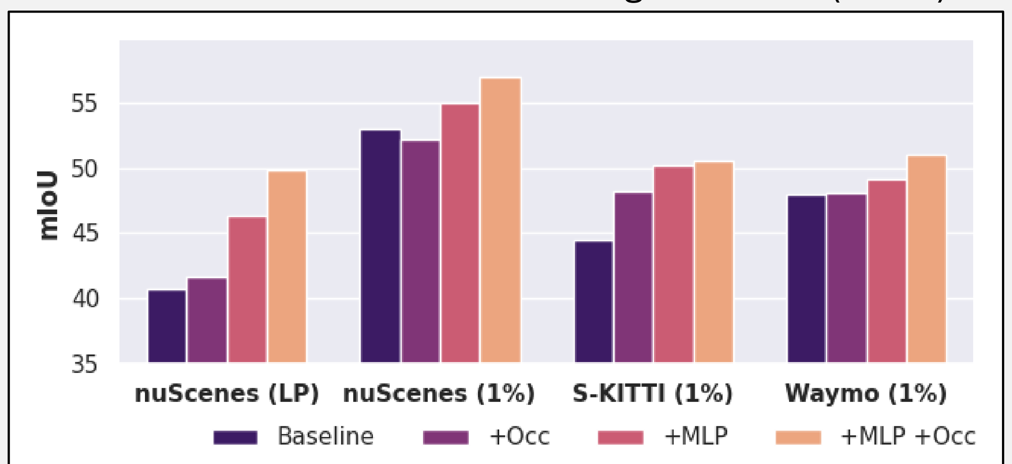
+ MLP projection head	Enable LiDAR backbone to learn additional information not contained in vision features.
+ LiDAR-specific auxiliary task	With self-supervised occupancy prediction, encourage LiDAR backbone to learn LiDAR-exclusive information.



Impact of adding MLP and Occupancy prediction to ScaLR

MinkUNet-34 distilled on nuScenes from DINO-v2-Small. Point cloud segmentation (mIoU)

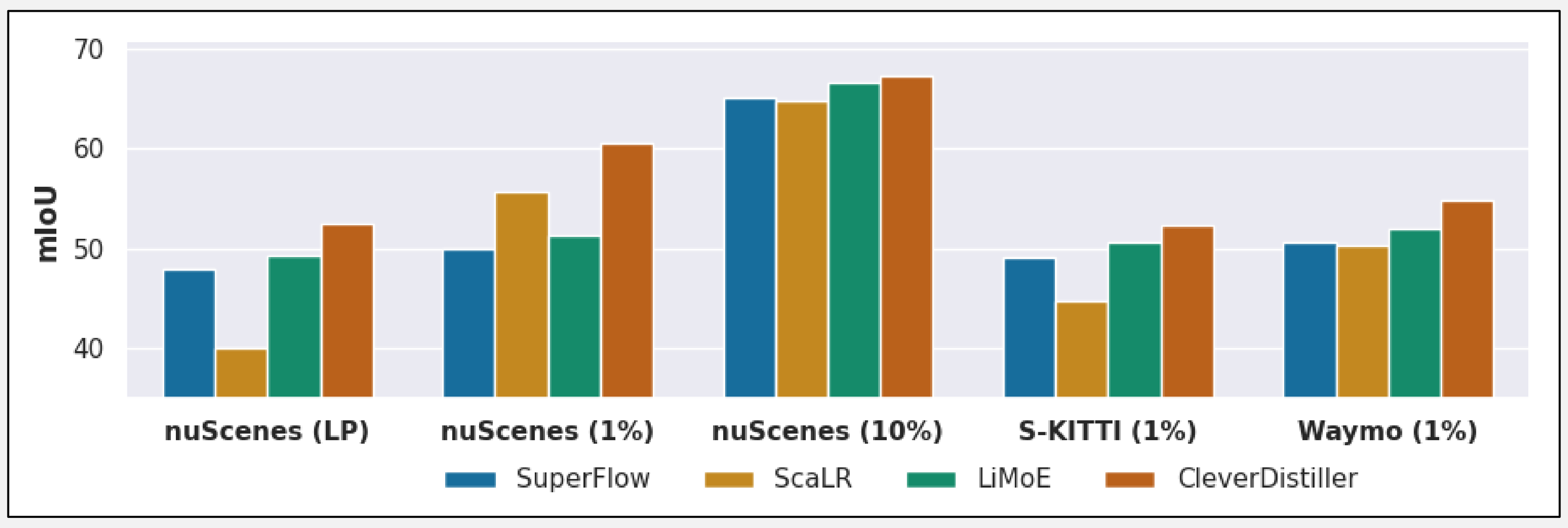
MLP	Occ	Matrix Rank	RankMe
✗	✗	22	1.89
✗	✓	25	2.33
✓	✗	33	4.24
✓	✓	41	5.27

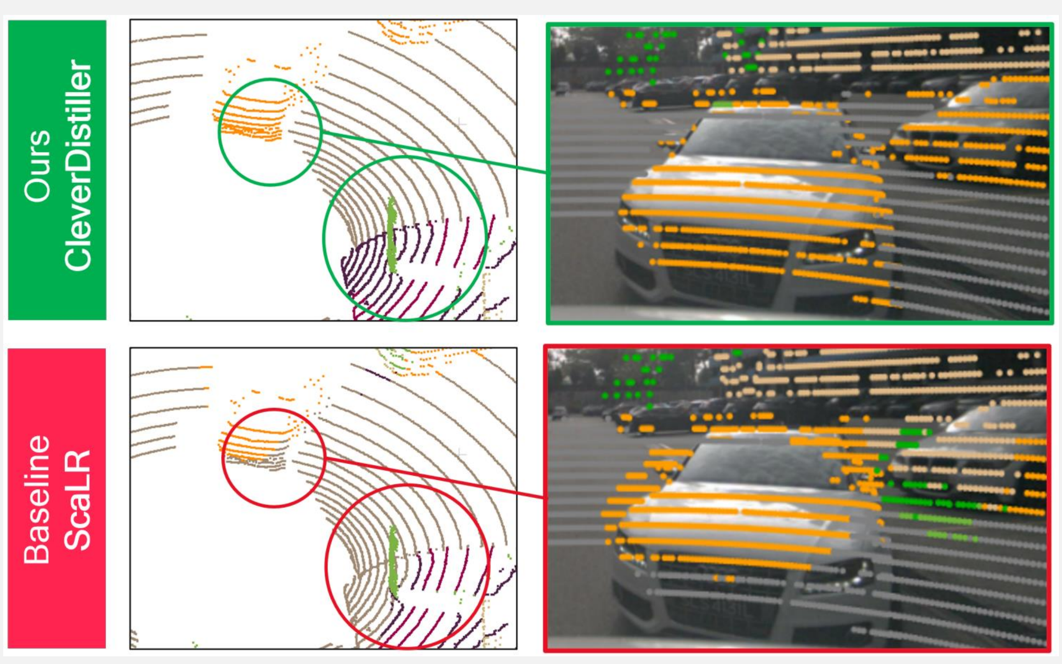


Occupancy prediction is more effective in the presence of an MLP projection head

Comparison with other CMKD methods

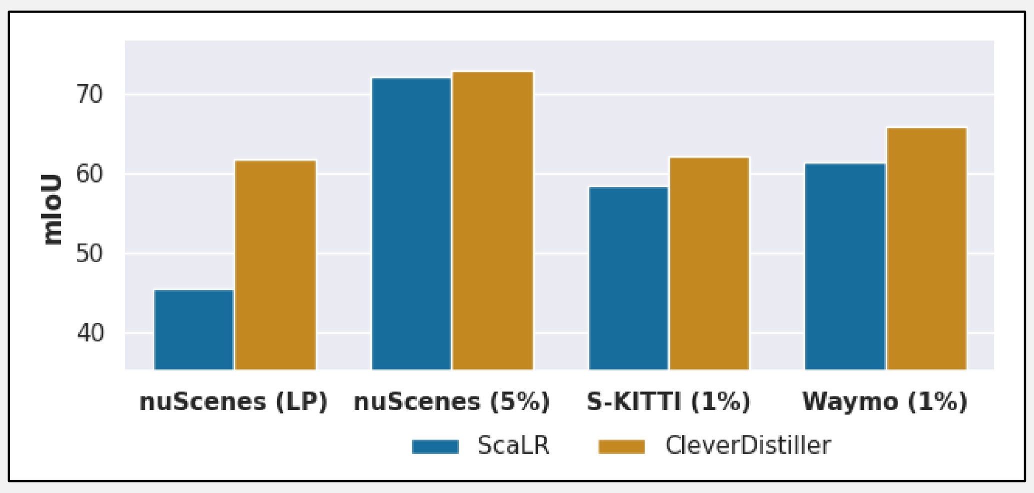
MinkUNet-34 distilled on nuScenes from DINO-v2-Large. Point cloud segmentation (mIoU)





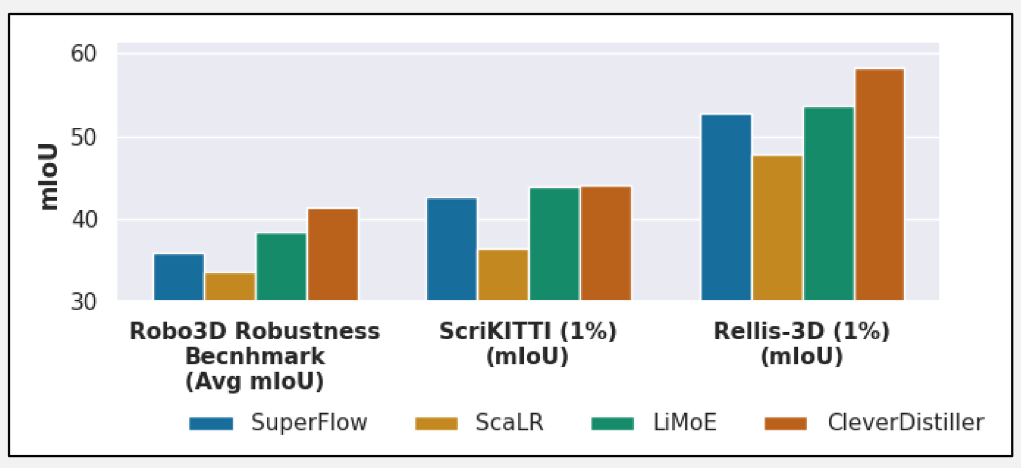
On SOTA LiDAR architecture

PointTransformerV3 distilled on nuScenes from DINO-v2-Large. Point cloud segmentation (mIoU)



Domain generalization and robustness

MinkUNet-34 distilled on nuScenes from DINO-v2-Large.



Paper