JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

MV-JAR: Masked Voxel Jigsaw and Reconstruction for LiDAR-Based Self-Supervised Pre-Training







Runsen Xu[1,2] Tai Wang[1,2] Wenwei Zhang[3,2] Runjian Chen[4] Jinkun Cao[5] Jiangmiao Pang[2] Dahua Lin[1,2]

[1]The Chinese University of Hong Kong [2]Shanghai AI Laboratory [3]S-Lab, NTU

[4]The University of Hong Kong [5]Carnegie Mellon University

Poster: WED-PM-104

Paper: https://arxiv.org/abs/2303.13510

Project: https://github.com/SmartBot-PJLab/MV-JAR

香港中文大學
The Chinese University of Hong Kong

上海人工智能实验室
Shanghai Artificial Intelligence Laboratory

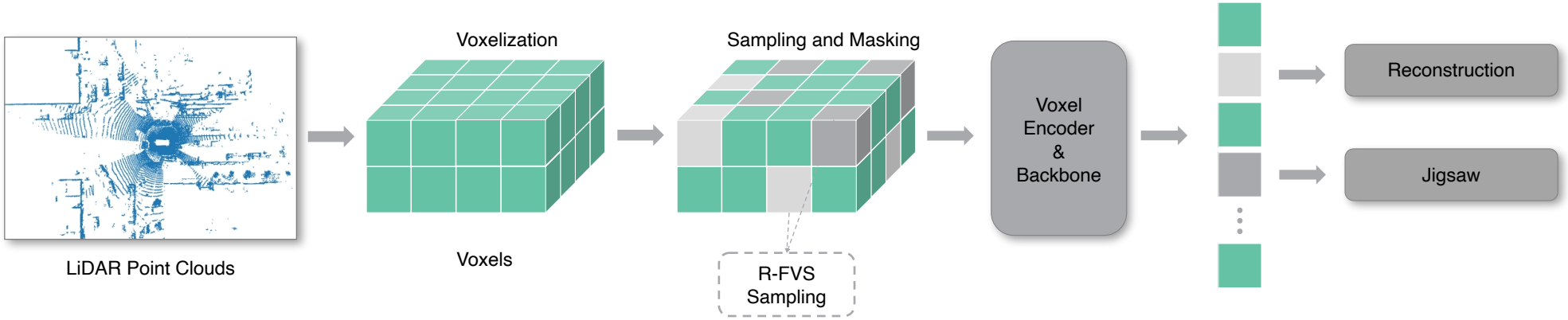
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

香港大學
THE UNIVERSITY OF HONG KONG

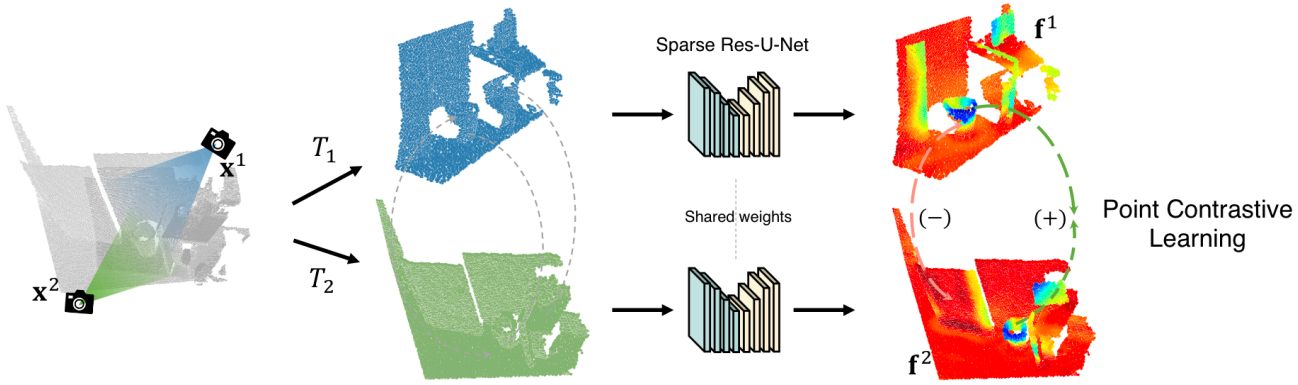
CARNEGIE MELLON UNIVERSITY
PITTSBURGH PENNSYLVANIA 1900

Overview

- MV-JAR
  - Reduces demand for labeled data by self-supervised pre-training.
  - Models voxel distributions in the scene by masked voxel jigsaw.
  - Models point distributions in the voxel by masked voxel reconstruction.
  - Reversed-Furthest-Voxel-Sampling to address the uneven distributions of LiDAR points.
  - Accelerates model convergence and improves performances across various data scales.
- A carefully designed data-efficient 3D object detection benchmark on Waymo.



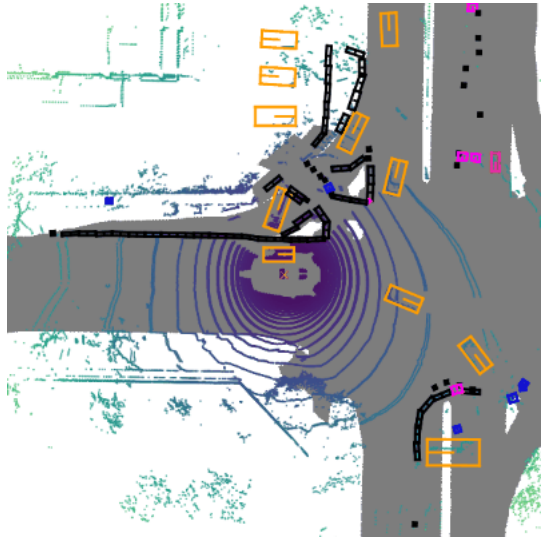
Motivations



Indoor points: static partial views & matching pairs[1]

Contrastive Learning

V.S.

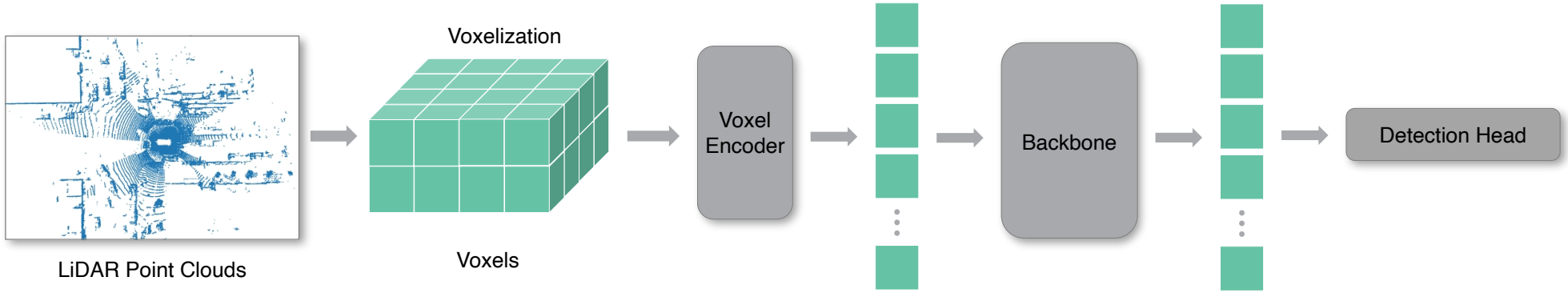


LiDAR points: dynamic & challenging pair formation

Masked Voxel Modeling

[1]Saining Xie et al. Pointcontrast: Unsupervised pre-training for 3d point cloud understanding. In ECCV, 2020.

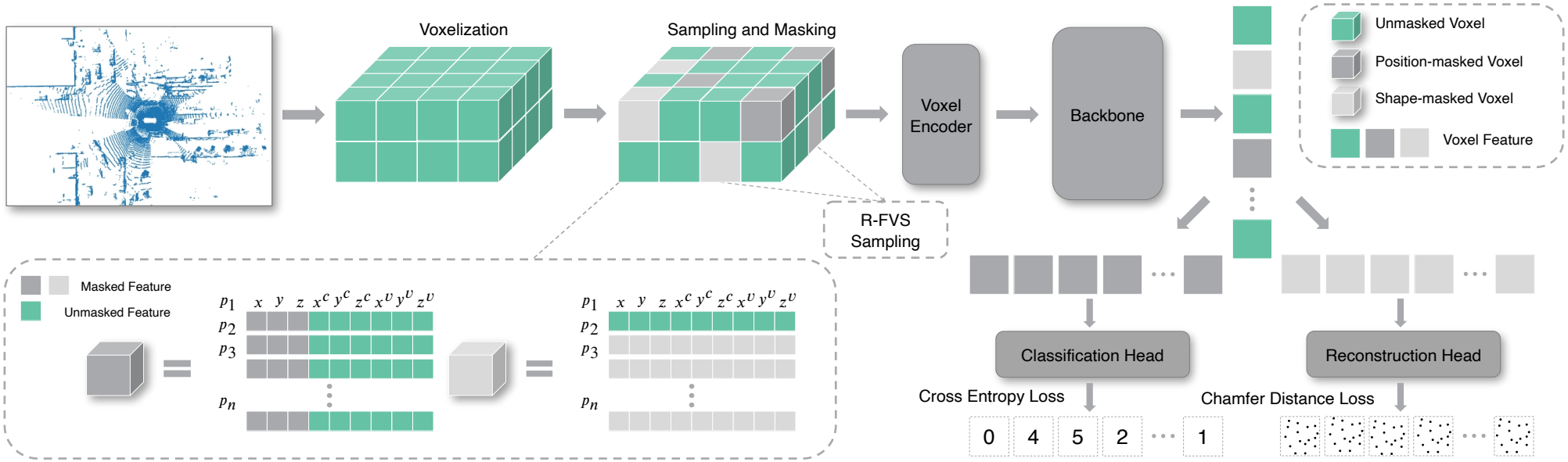
Motivations



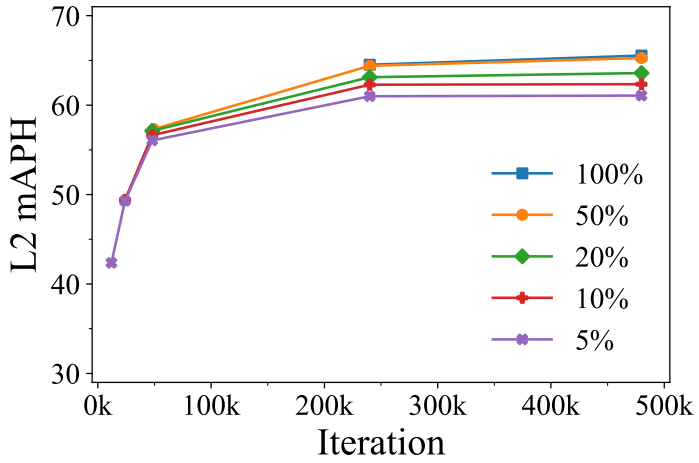
Scene-Voxel-Point Hierarchy in LiDAR Detectors

Methodology

- Masked Voxel Jigsaw: Mask positions and put back.
- Masked Voxel Reconstruction: Mask shapes and reconstruct.
- Reversed-Furthest-Voxel-Sampling: Keep the furthest voxels (sampled by FPS), while mask the rest.



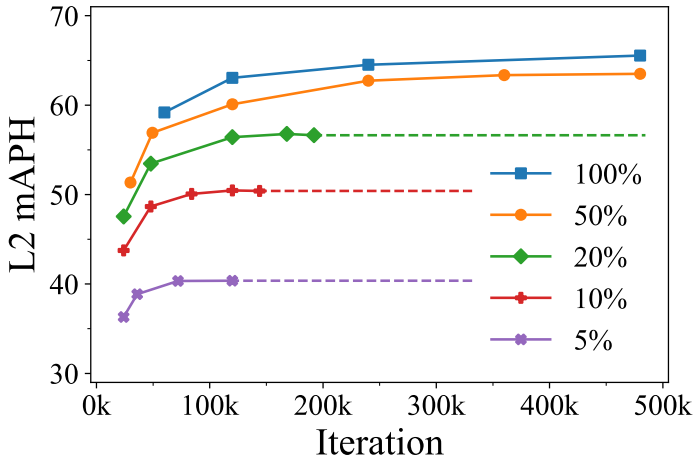
The New Benchmark on Waymo



Existing Data-efficient Experiments

Observation: Incomplete model convergence

Issue: Fine-tuning splits share similar data diversities

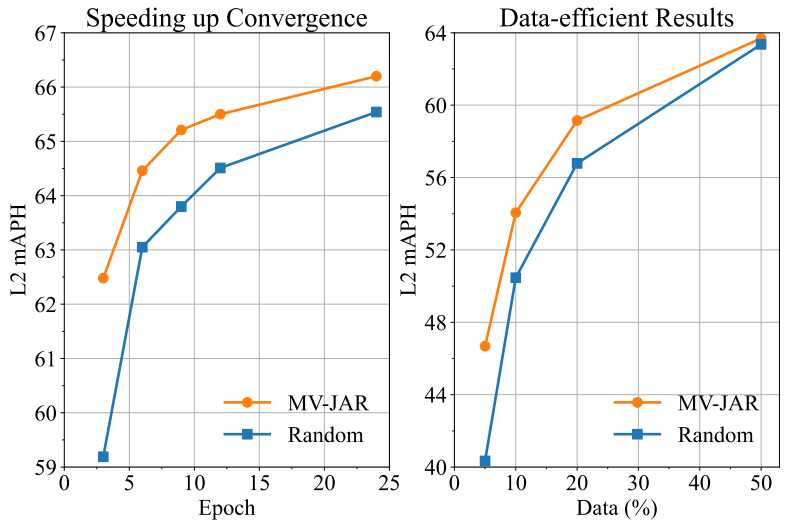


Our New Benchmark

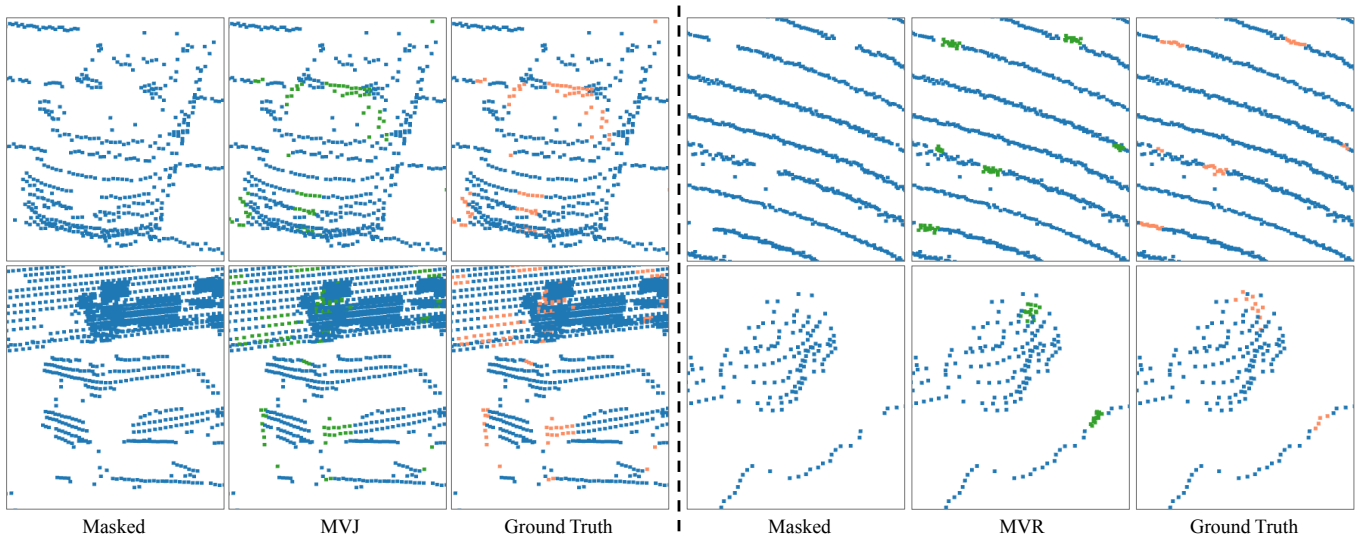
Solution: Sequence-based data sampling

Solution: Ensure model convergence

Experimental Results



SST[1] Detection Performance on Waymo



Visualization

[1]Lue Fan et al. Embracing single stride 3d object detector with sparse transformer. In CVPR, 2022.

Experimental Results

Fine-tuned with 5% Data on Waymo

Data Amount	Initialization	Overall	
		L2 mAP	L2 mAPH
5%	Random	44.41	40.34
	PointContrast[1]	45.32	41.30
	ProposalContrast[2]	46.62	42.58
	MV-JAR (Ours)	**50.52**$^{+6.11}$	**46.68**$^{+6.34}$

Transferring from Waymo to KITTI

Initialization	Overall		
	Easy	Mod.	Hard
Random	74.71	64.43	60.00
PointContrast[1]	73.35	62.53	59.01
ProposalContrast[2]	73.63	63.34	59.40
MV-JAR (Ours)	**75.22**	**63.80**	**60.35**

See our paper for comprehensive experimental results.

[1]Saining Xie et al. Pointcontrast: Unsupervised pre-training for 3d point cloud understanding. In ECCV, 2020.
[2]Junbo Yin et al. Proposalcontrast: Unsupervised pre-training for lidar-based 3d objectdetection. In ECCV, 2022.

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

MV-JAR: Masked Voxel Jigsaw and Reconstruction for LiDAR-Based Self-Supervised Pre-Training







Runsen Xu[1,2] Tai Wang[1,2] Wenwei Zhang[3,2] Runjian Chen[4] Jinkun Cao[5] Jiangmiao Pang[2] Dahua Lin[1,2]

[1]The Chinese University of Hong Kong [2]Shanghai AI Laboratory [3]S-Lab, NTU

[4]The University of Hong Kong [5]Carnegie Mellon University

Poster: WED-PM-104

Paper: https://arxiv.org/abs/2303.13510

Project: https://github.com/SmartBot-PJLab/MV-JAR

香港中文大學
The Chinese University of Hong Kong

上海人工智能实验室
Shanghai Artificial Intelligence Laboratory

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

香港大學
THE UNIVERSITY OF HONG KONG

CARNEGIE MELLON UNIVERSITY
PITTSBURGH PENNSYLVANIA 1900