LION: Linear Group RNN for 3D Object Detection in Point Clouds

Zhe Liu* · Jinghua Hou* · Xinyu Wang* · Xiaoqing Ye · Jingdong Wang · Hengshuang Zhao · Xiang Bai

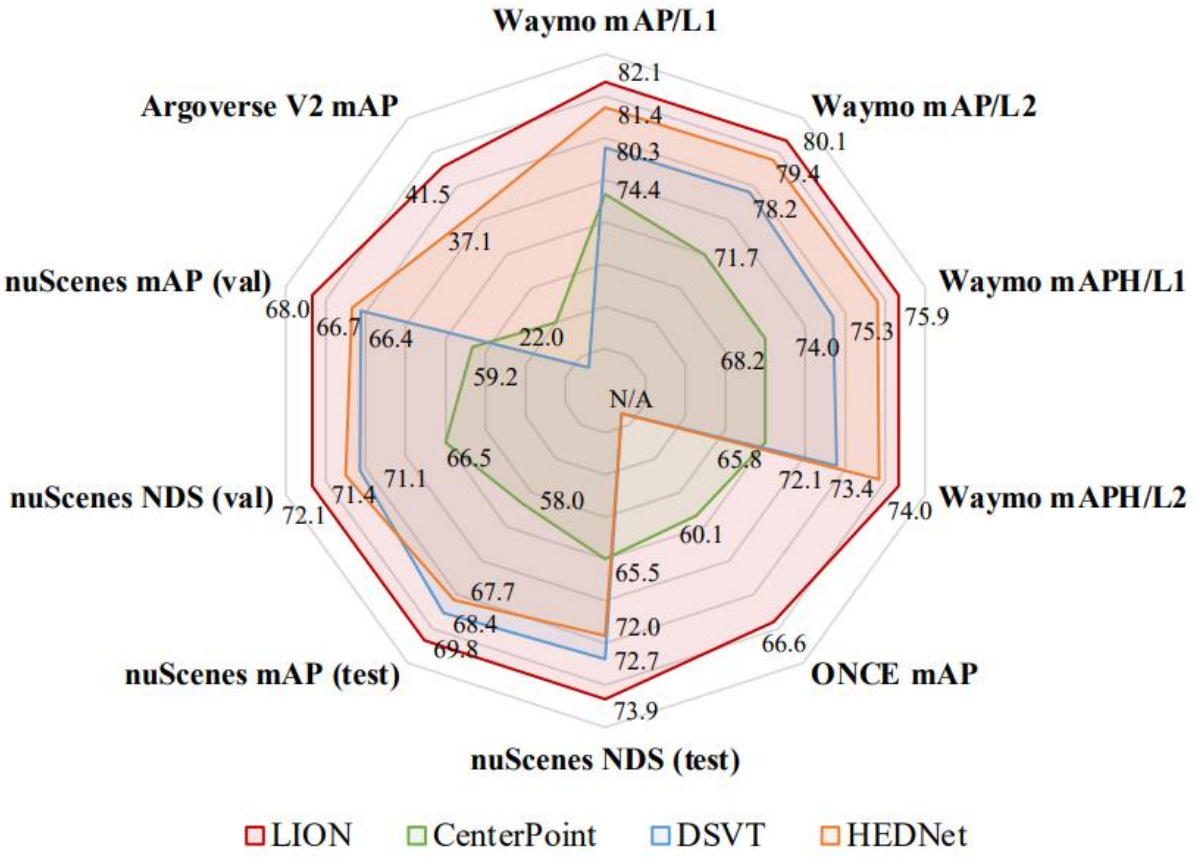
Speaker： Zhe Liu

Code is available at: https://github.com/happinesslz/LION

QR Code

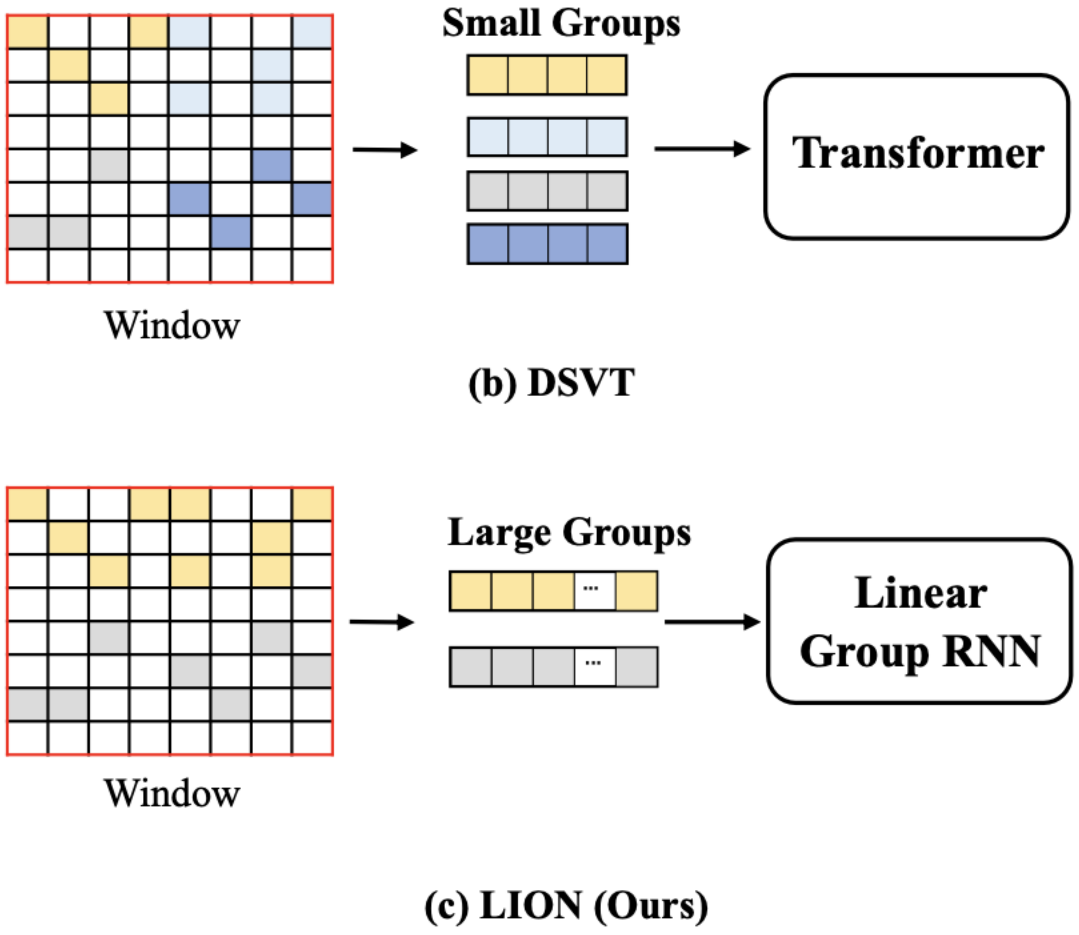
Overview

- **Strong Performance.** LION achieves state-of-the-art performance on Waymo, nuScenes, Argoverse V2, and ONCE datasets. 💪
- **Strong Generalization.** LION can support most of representative linear RNN operators .😀
- **More Friendly.** LION can train all models on less 24G GPU memory~(i.e., RTX 3090, RTX4090, V100 and A100 are enough to train our LION).😎



Motivation

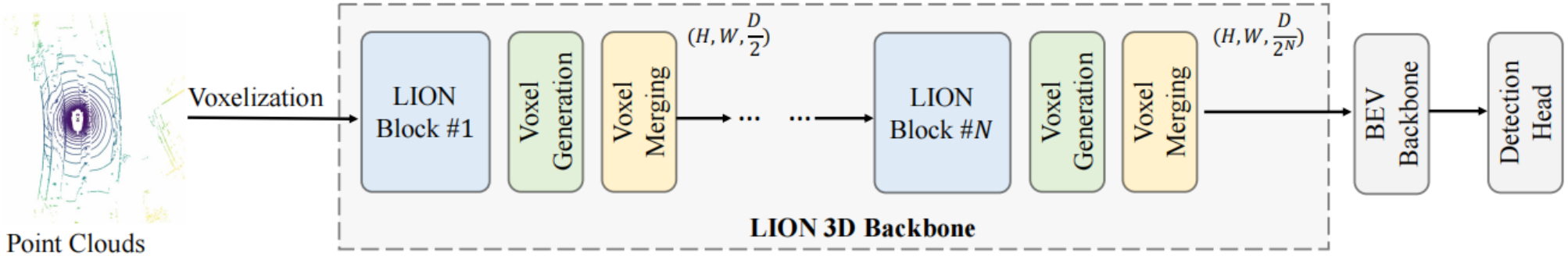
- Perform self-attention with only **a small group size due to computation limitations**.
- Some linear RNN operators **with linear computational complexity** have achieved competitive performance with transformers.
- Perform **long-range feature interaction in larger groups at a lower computation cost**



Contributions

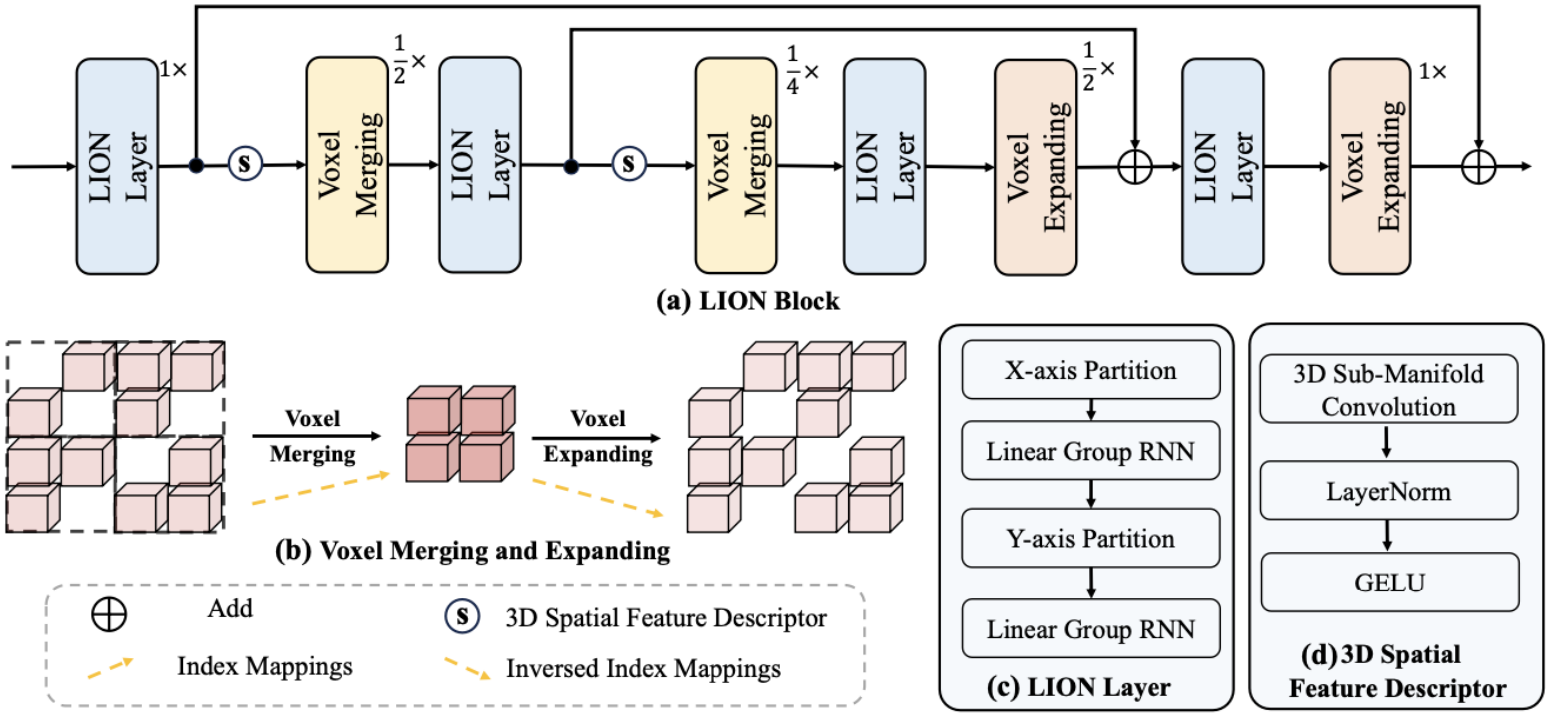
- Propose a **simple and effective window-based 3D backbone based on the linear group RNN**
- Introduce a simple **3D spatial feature descriptor** and integrate it with the linear group RNN
- Propose a new **3D voxel generation strategy** to densify foreground features.
- Verify the **generalization of our LION** with different linear group RNN mechanisms (e.g., **Mamba, RWKV, RetNet**).

Method



The pipeline of LION

Method



The pipeline of LION block

Experiments

Methods	Present at	Operator	*Vehicle* 3D AP/APH L1	*Vehicle* 3D AP/APH L2	*Pedestrian* 3D AP/APH L1	*Pedestrian* 3D AP/APH L2	*Cyclist* 3D AP/APH L1	*Cyclist* 3D AP/APH L2	mAP/mAPH L2
SECOND [60]	Sensors 18	Sparse Conv	72.3/71.7	63.9/63.3	68.7/58.2	60.7/51.3	60.6/59.3	58.3/57.0	61.0/57.2
PointPillars [26]	CVPR 19		72.1/71.5	63.6/63.1	70.6/56.7	62.8/50.3	64.4/62.3	61.9/59.9	62.8/57.8
CenterPoint [66]	CVPR 21		74.2/73.6	66.2/65.7	76.6/70.5	68.8/63.2	72.3/71.1	69.7/68.5	68.2/65.8
PV-RCNN‡ [45]	CVPR 20		78.0/77.5	69.4/69.0	79.2/73.0	70.4/64.7	71.5/70.3	69.0/67.8	69.6/67.2
PillarNet-34 [44]	ECCV 22		79.1/78.6	70.9/70.5	80.6/74.0	72.3/66.2	72.3/71.2	69.7/68.7	71.0/68.5
FSD‡ [17]	NeurIPS 22		79.2/78.8	70.5/70.1	82.6/77.3	73.9/69.1	77.1/76.0	74.4/73.3	72.9/70.8
AFDetV2 [25]	AAAI 22		77.6/77.1	69.7/69.2	80.2/74.6	72.2/67.0	73.7/72.7	71.0/70.1	71.0/68.8
PillarNeXt [27]	CVPR 23		78.4/77.9	70.3/69.8	82.5/77.1	74.9/69.8	73.2/72.2	70.6/69.6	71.9/69.7
VoxelNext [9]	CVPR 23		78.2/77.7	69.9/69.4	81.5/76.3	73.5/68.6	76.1/74.9	73.3/72.2	72.2/70.1
CenterFormer[72]	ECCV 22		75.0/74.4	69.9/69.4	78.6/73.0	73.6/68.3	72.3/71.3	69.8/68.8	71.1/68.9
PV-RCNN++‡ [46]	IJCV 22		79.3/78.8	70.6/70.2	81.3/76.3	73.2/68.0	73.7/72.7	71.2/70.2	71.7/69.5
TransFusion [2]	CVPR 22		–/–	–/65.1	–/–	–/63.7	–/–	–/65.9	–/64.9
ConQueR [74]	CVPR 23		76.1/75.6	68.7/68.2	79.0/72.3	70.9/64.7	73.9/72.5	71.4/70.1	70.3/67.7
FocalFormer3D [7]	ICCV 23		–/–	68.1/67.6	–/–	72.7/66.8	–/–	73.7/72.6	71.5/69.0
HEDNet [68]	NeurIPS 23		81.1/80.6	73.2/72.7	84.4/80.0	76.8/72.6	78.7/77.7	75.8/74.9	75.3/73.4
SST_TS‡ [15]	CVPR 22	Transformer	76.2/75.8	68.0/67.6	81.4/74.0	72.8/65.9	–/–	–/–	–/–
SWFormer [50]	ECCV 22		77.8/77.3	69.2/68.8	80.9/72.7	72.5/64.9	–/–	–/–	–/–
OcTr [70]	CVPR 23		78.1/77.6	69.8/69.3	80.8/74.4	72.5/66.5	72.6/71.5	69.9/68.9	70.7/68.2
DSVT-Pillar [57]	CVPR 23		79.3/78.8	70.9/70.5	82.8/77.0	75.2/69.8	76.4/75.4	73.6/72.7	73.2/71.0
DSVT-Voxel [57]	CVPR 23		79.7/79.3	71.4/71.0	83.7/78.9	76.1/71.5	77.5/76.5	74.6/73.7	74.0/72.1
LION-RetNet (Ours)	–	RNN	79.0/78.5	70.6/70.2	84.6/80.0	77.2/72.8	79.0/78.0	76.1/75.1	74.6/72.7
LION-RWKV (Ours)	–		79.7/79.3	71.3/71.0	84.6/80.0	77.1/72.7	78.7/77.7	75.8/74.8	74.7/72.8
LION-Mamba (Ours)	–		79.5/79.1	71.1/70.7	84.9/80.4	77.5/73.2	79.7/78.7	76.7/75.8	75.1/73.2
LION-Mamba-L (Ours)	–		**80.3/79.9**	**72.0/71.6**	**85.8/81.4**	**78.5/74.3**	**80.1/79.0**	**77.2/76.2**	**75.9/74.0**

Table 1: Performance on the Waymo Open Dataset *validation* set (train with 100% training data)

Experiments

Methods	Present at	Frames	*Vehicle* 3D AP/APH L1	*Vehicle* 3D AP/APH L2	*Pedestrian* 3D AP/APH L1	*Pedestrian* 3D AP/APH L2	*Cyclist* 3D AP/APH L1	*Cyclist* 3D AP/APH L2	mAP/mAPH L2
PV-RCNN++‡ [49]	IJCV 22	1	81.6/81.2	73.9/73.5	80.4/75.0	74.1/69.0	71.9/70.8	69.3/68.2	72.4/70.2
AFDetV2 [26]	AAAI 22	1	–/–	–/–	–/–	–/–	–/–	–/–	72.2/70.3
PillarNeXt [28]	CVPR 23	1	–/–	–/–	–/–	–/–	–/–	–/–	72.2/69.6
FSD‡ [17]	NeurIPS 22	1	82.7/82.3	74.4/74.1	82.9/77.9	75.9/71.3	75.6/74.4	72.9/71.8	74.4/72.4
PillarNeXt [28]	CVPR 23	1	–/–	–/–	–/–	–/–	–/–	–/–	–/72.0
CenterPoint++ [69]	CVPR 21	3	82.8/82.3	75.5/75.1	81.1/78.2	75.1/72.4	74.4/73.3	72.0/71.0	74.2/72.8
PillarNeXt [28]	CVPR 23	3	83.3/82.8	76.2/75.8	84.4/81.4	78.8/76.0	73.7/72.7	71.6/70.6	75.5/74.1
LION-Mamba-L (ours)	–	3	84.7/84.3	77.2/76.9	87.2/84.5	82.0/79.3	79.2/78.3	76.8/75.9	78.7/77.4

Table 2: Results of our LION with multiple frames as input on the Waymo Open Dataset *test* set

Experiments

Table 2: Performances on the nuScenes *validation* and *test* set. 'T.L.', 'C.V.', 'Ped.', 'M.T.', 'T.C.', and 'B.R.' are short for trailer, construction vehicle, pedestrian, motor, traffic cone, and barrier, respectively. All results are reported without any test-time augmentation and model ensembling.

Method	Present at	NDS	mAP	Car	Truck	Bus	T.L.	C.V.	Ped.	M.T.	Bike	T.C.	B.R.
Performances on the *validation* set													
CenterPoint [66]	CVPR 21	66.5	59.2	84.9	57.4	70.7	38.1	16.9	85.1	59.0	42.0	69.8	68.3
VoxelNeXt [9]	CVPR 23	66.7	60.5	83.9	55.5	70.5	38.1	21.1	84.6	62.8	50.0	69.4	69.4
Uni3DETR [58]	NeurIPS 23	68.5	61.7	–	–	–	–	–	–	–	–	–	–
TransFusion-LiDAR [2]	CVPR 22	70.1	65.5	86.9	60.8	73.1	43.4	25.2	87.5	72.9	57.3	77.2	70.3
DSVT [57]	CVPR 23	71.1	66.4	87.4	62.6	75.9	42.1	25.3	88.2	74.8	58.7	77.9	71.0
HEDNet [68]	NeurIPS 23	71.4	66.7	87.7	60.6	77.8	**50.7**	**28.9**	87.1	74.3	56.8	76.3	66.9
LION-RetNet (Ours)	–	71.9	67.3	87.9	64.3	**78.7**	44.6	27.6	88.9	73.5	56.6	79.2	**72.1**
LION-RWKV (Ours)	–	71.7	66.8	**88.1**	59.0	77.6	46.6	28.0	**89.7**	74.3	56.2	80.1	68.3
LION-Mamba (Ours)	–	**72.1**	**68.0**	87.9	**64.9**	77.6	44.4	28.5	89.6	**75.6**	**59.4**	**80.8**	71.6
Performances on the *test* set													
TransFusion-LiDAR [2]	CVPR 22	70.2	65.5	86.2	56.7	66.3	58.8	28.2	86.1	68.3	44.2	82.0	78.2
DSVT [57]	CVPR 23	72.7	68.4	86.8	58.4	67.3	63.1	**37.1**	88.0	73.0	47.2	84.9	78.4
HEDNet [68]	NeurIPS 23	72.0	67.7	87.1	56.5	**70.4**	63.5	33.6	87.9	70.4	44.8	85.1	78.1
LION-Mamba (Ours)	–	**73.9**	**69.8**	**87.2**	**61.1**	68.9	**65.0**	36.3	**90.0**	**74.0**	**49.2**	**87.3**	**79.5**

Table 3: Performance on the nuScenes *validation* and *test* set

Experiments

Method	mAP	Vehicle	Bus	Pedestrian	Stop Sign	Box Truck	Bollard	C-Barrel	Motorcyclist	MPC-Sign	Motorcycle	Bicycle	A-Bus	School Bus	Truck Cab	C-Cone	V-Trailer	Sign	Large Vehicle	Stroller	Bicyclist	Truck	MBT	Dog	Wheelchair	W-Device	W-Rider
CenterPoint [69]	22.0	67.6	38.9	46.5	16.9	37.4	40.1	32.2	28.6	27.4	33.4	24.5	8.7	25.8	22.6	29.5	22.4	6.3	3.9	0.5	20.1	22.1	0.0	3.9	0.5	10.9	4.2
HEDNet [71]	37.1	78.2	47.7	67.6	46.4	45.9	56.9	67.0	48.7	46.5	58.2	47.5	23.3	40.9	27.5	46.8	27.9	20.6	6.9	27.2	38.7	21.6	0.0	30.7	9.5	28.5	8.7
VoxelNeXt [9]	30.7	72.7	38.8	63.2	40.2	40.1	53.9	64.9	44.7	39.4	42.4	40.6	20.1	25.2	19.9	44.9	20.9	14.9	6.8	15.7	32.4	16.9	0.0	14.4	0.1	17.4	6.6
FSDv1 [17]	28.2	68.1	40.9	59.0	29.0	38.5	41.8	42.6	39.7	26.2	49.0	38.6	20.4	30.5	14.8	41.2	26.9	11.9	5.9	13.8	33.4	21.1	0.0	9.5	7.1	14.0	9.2
FSDv2 [18]	37.6	**77.0**	47.6	70.5	43.6	41.5	53.9	58.5	56.8	39.0	60.7	49.4	28.4	41.9	**30.2**	44.9	**33.4**	16.6	7.3	32.5	45.9	**24.0**	**1.0**	12.6	**17.1**	26.3	17.2
SAFDNet [70]	39.7	78.5	**49.4**	70.7	51.5	44.7	65.7	72.3	54.3	**49.7**	60.8	50.0	**31.3**	**44.9**	24.7	55.4	31.4	22.1	7.1	31.1	42.7	23.6	0.0	26.1	1.4	30.2	11.5
LION-RetNet	40.7	74.7	41.0	72.7	47.5	44.2	66.9	**77.0**	57.1	48.3	63.7	55.1	27.0	42.5	25.2	57.9	29.7	22.0	6.9	39.3	47.3	19.9	0.0	28.8	12.8	**37.7**	**12.6**
LION-RWKV	41.1	76.3	44.6	**74.0**	52.1	**46.0**	**68.1**	75.8	55.8	49.4	62.8	55.3	27.1	42.9	25.9	**60.1**	30.9	22.2	**9.3**	36.5	**55.3**	23.2	0.0	27.8	7.1	37.6	11.4
LION-Mamba	**41.5**	75.1	43.6	73.9	**53.9**	45.1	66.4	74.7	**61.3**	48.7	**65.1**	**56.2**	21.7	42.7	25.3	58.4	28.9	**23.6**	8.3	**49.5**	47.3	19.0	0.0	**31.4**	8.7	37.6	11.8

Table 4: Performance on the Argoverse V2 *validation* set

Experiments

Method	Vehicle				Pedestrian				Cyclist				mAP
	overall	0-30m	30-50m	50m-inf	overall	0-30m	30-50m	50m-inf	overall	0-30m	30-50m	50m-inf	
PointRCNN [50]	52.1	74.5	40.9	16.8	4.3	6.2	2.4	0.9	29.8	46.0	20.9	5.5	28.7
PointPillars [27]	68.6	80.9	62.1	47.0	17.6	19.7	15.2	10.2	46.8	58.3	40.3	25.9	44.3
SECOND [63]	71.2	84.0	63.0	47.3	26.4	29.3	24.1	18.1	58.0	70.0	52.4	34.6	51.9
PV-RCNN [48]	77.8	**89.4**	**72.6**	58.6	23.5	25.6	22.8	17.3	59.4	71.7	52.6	36.2	53.6
CenterPoint [69]	66.8	80.1	59.6	43.4	49.9	56.2	42.6	**26.3**	63.5	74.3	57.9	41.5	60.1
PointPainting [57]	66.2	80.3	59.8	42.3	44.8	52.6	36.6	22.5	62.3	73.6	57.2	40.4	57.8
LION-RetNet	78.1	88.7	72.4	**58.5**	52.4	60.5	43.6	**26.3**	68.3	**79.4**	62.9	46.1	66.3
LION-RWKV	**78.3**	89.2	**72.6**	56.7	50.6	60.0	40.4	24.2	68.4	**79.4**	**63.2**	45.7	65.8
LION-Mamba	78.2	89.1	**72.6**	57.5	**53.2**	**62.4**	**44.0**	24.5	**68.5**	79.2	**63.2**	**47.1**	**66.6**

Table 5: Performance on the ONCE *validation* set

Experiments

Method	*Car*			*Pedestrian*			*Cyclist*			mAP
	Easy	Moderate	Hard	Easy	Moderate	Hard	Easy	Moderate	Hard	
VoxelNet [75]	77.5	65.1	57.7	39.5	33.7	31.5	61.2	48.4	44.4	51.0
SECOND [63]	83.1	73.7	66.2	51.1	42.6	37.3	70.5	53.9	46.9	58.4
PointPillars [27]	79.1	75.0	68.3	52.1	43.5	41.5	75.8	59.1	52.9	60.8
PointRCNN [50]	85.9	75.8	68.3	49.4	41.8	38.6	73.9	59.6	53.6	60.8
TANet [34]	83.8	75.4	67.7	54.9	46.7	42.4	73.8	59.9	53.5	62.0
DSVT-Pillar* [60]	87.3	77.4	76.2	61.4	56.8	51.8	82.3	67.1	63.7	69.3
DSVT-Voxel* [60]	87.8	77.8	76.8	66.1	59.7	55.2	83.5	66.7	63.2	70.8
LION-TTT	87.9	78.0	76.7	63.4	58.6	53.7	84.0	69.6	64.5	70.7
LION-xLSTM	87.7	77.9	76.8	66.6	59.3	54.0	82.4	67.4	63.4	70.6
LION-RetNet	88.0	77.9	76.7	67.4	60.2	55.8	83.6	69.6	64.6	71.5
LION-Mamba	**88.6**	**78.3**	77.2	67.2	60.2	55.6	83.0	68.6	63.9	71.4
LION-RWKV	88.5	**78.3**	**77.1**	**68.9**	**62.2**	**58.1**	**89.6**	**71.2**	**66.9**	**73.4**

Table 5: Performance on the KITTI *validation* set

Ablation Study

Large Group Size	3D Spatial Feature Descriptor	Voxel Generation	3D AP/APH (L2)			mAP/mAPH (L2)
			Vehicle	*Pedestrian*	*Cyclist*	
–	–	–	65.6/65.2	72.3/65.0	68.3/67.2	68.8/65.8
✓	–	–	66.2/65.7	73.7/67.2	68.7/67.6	69.5/66.9
✓	✓	–	66.5/66.1	74.8/69.6	70.9/70.0	70.8/68.6
✓	–	✓	66.4/66.0	73.5/67.4	70.4/69.3	70.1/67.6
✓	✓	✓	**67.0/66.6**	**75.4/70.2**	**71.9/71.0**	**71.4/69.3**

Thanks for your listening!

Open-sourced Code

Page : https://happinesslz.github.io/projects/LION/

Code: https://github.com/happinesslz/LION