JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

LinK: Linear Kernel for LiDAR-based 3D Perception

Tao Lu [1] Xiang Ding [1] Haisong Liu [1] Gangshan Wu [1] Limin Wang [1,2] *

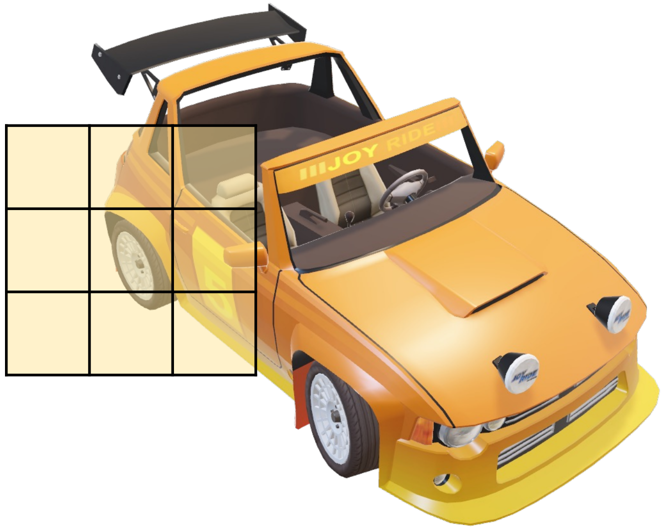
[1]State Key Laboratory for Novel Software Technology, Nanjing University [2]Shanghai AI Lab

{taolu,xding,liuhs}@smail.nju.edu.cn, {gswu,lmwang}@nju.edu.cn

Problem

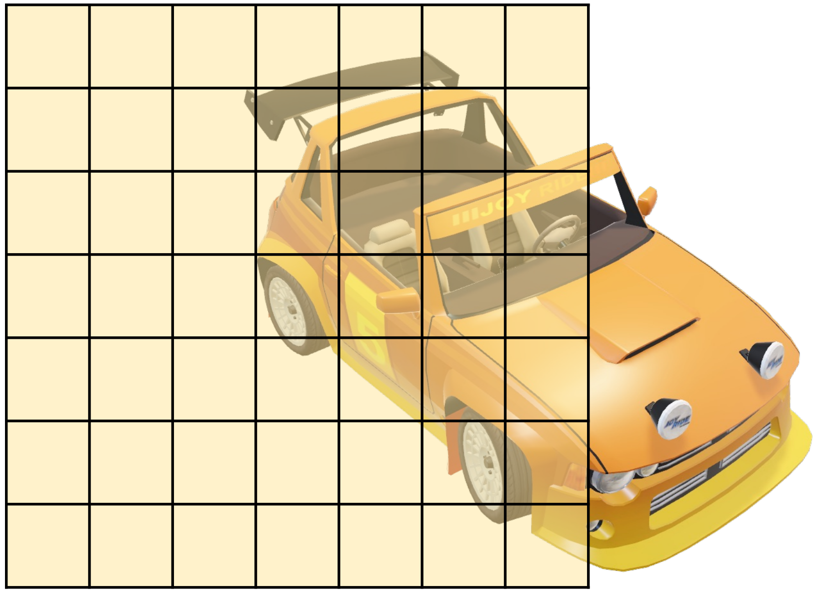
❖How to scale up kernels in 3D?

Small kernel



Large kernel

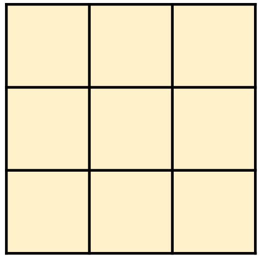
- More informative context
- Better shape prior
- …

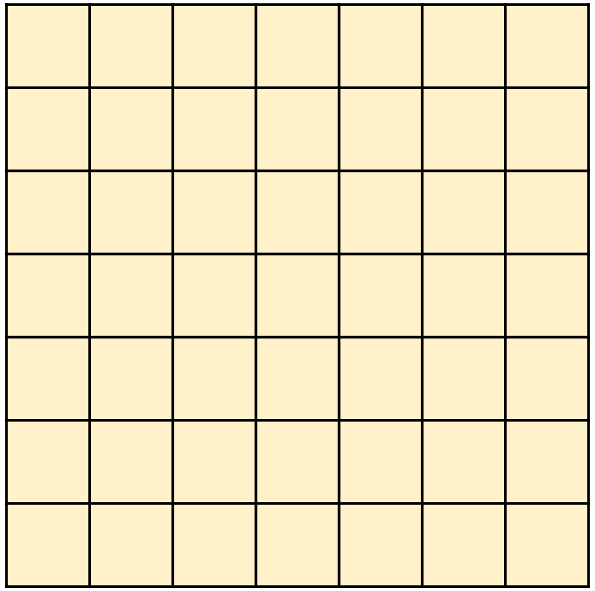


MULTIMEDIA COMPUTING GROUP
媒体计算研究组

Difficulties

- Cubically increasing overhead





Larger kernel

......

3×3×3

7×7×7

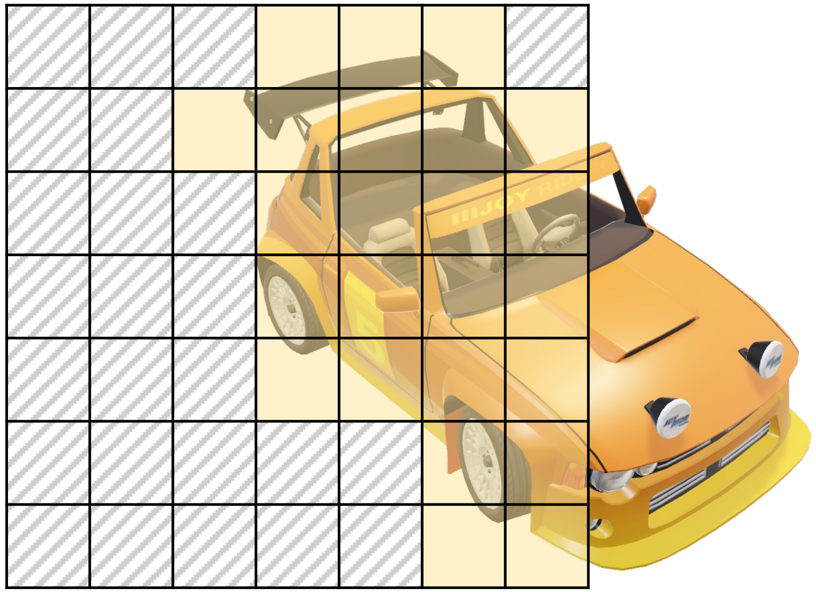
21×21×21

$$\left(\frac{7}{3}\right)^3 \approx \mathbf{12.7}\uparrow$$

$$\left(\frac{21}{3}\right)^3 = \mathbf{343}\uparrow$$

Difficulties

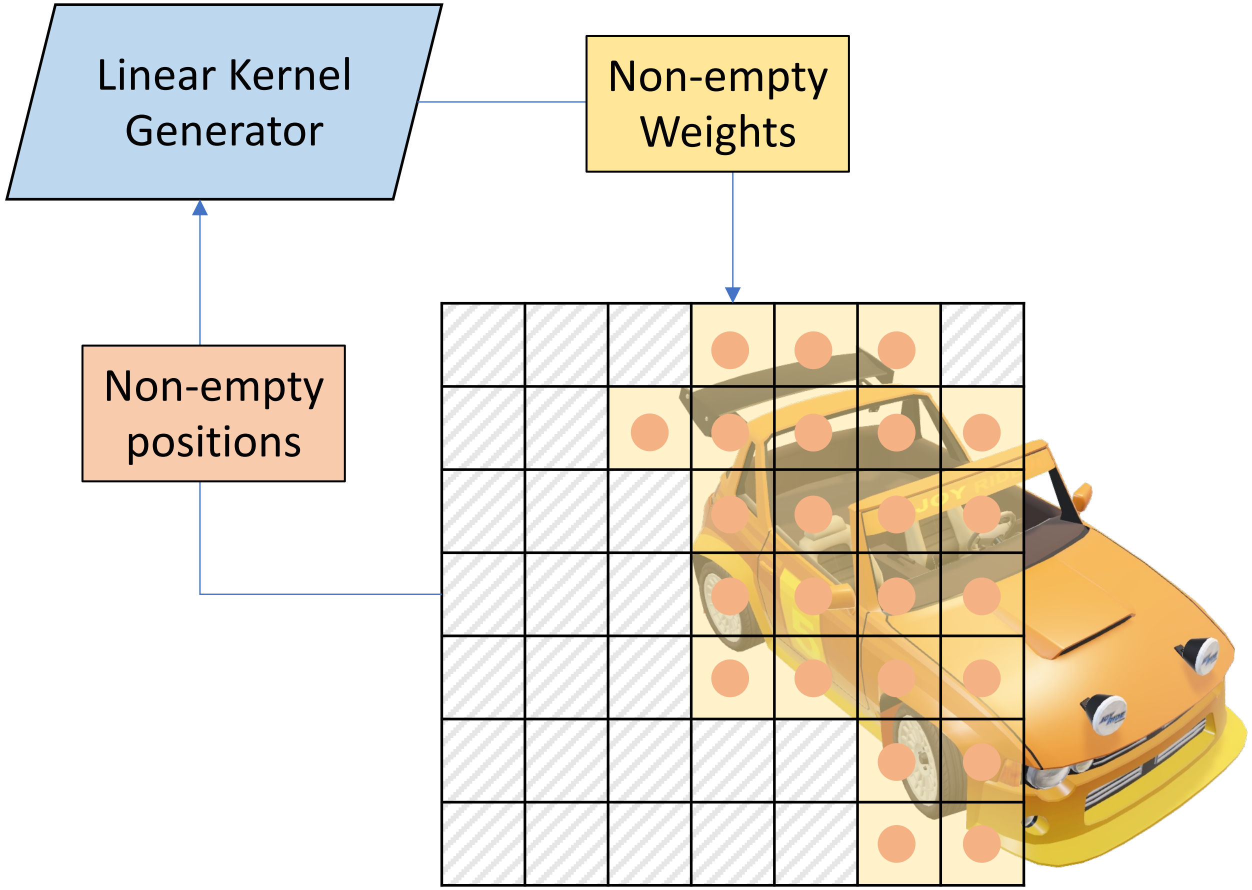
- ➢Sparsity slows down the optimization

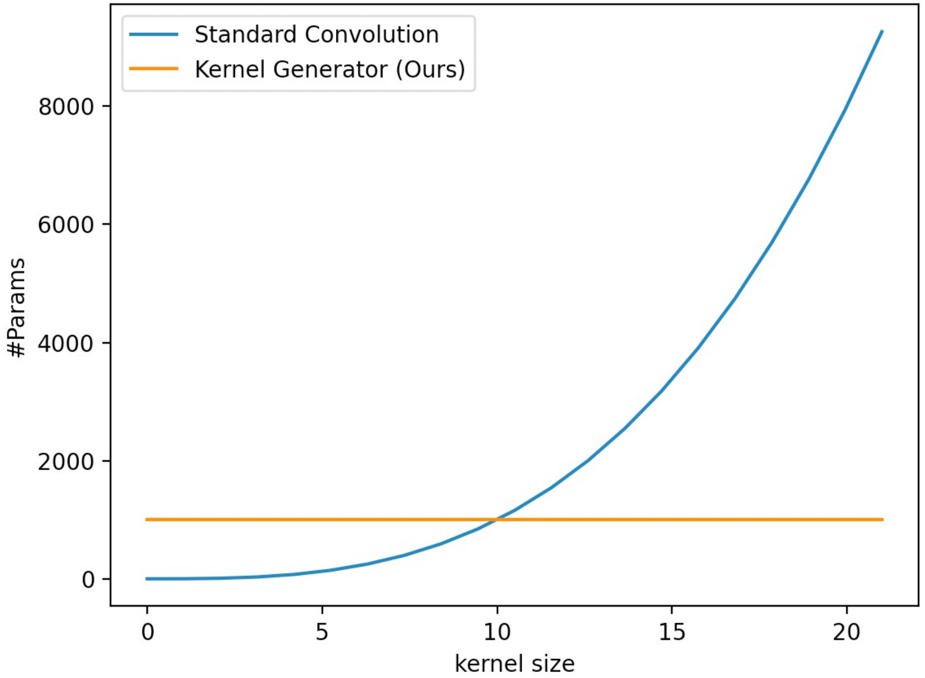


Empty area fails to be updated in backward process

Our Solution

➢ Linear Kernel Generator



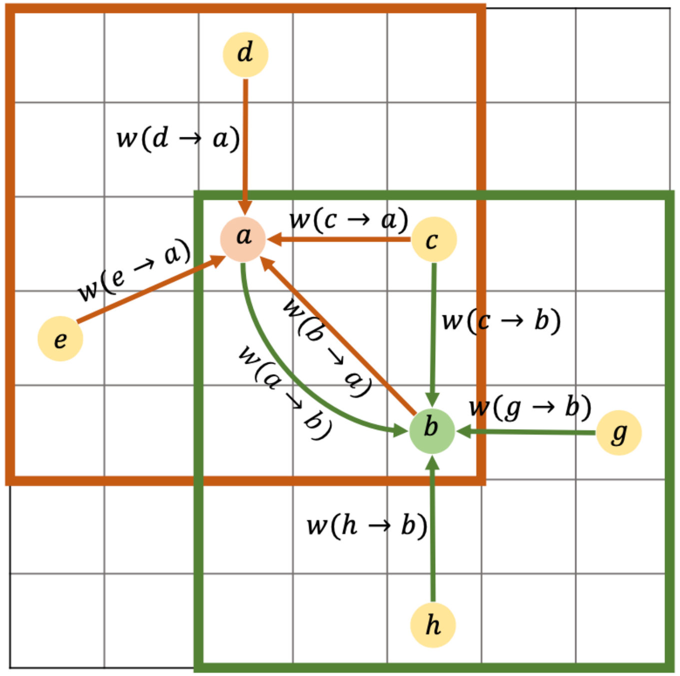


- ✓ Constant amount of learnable params, not increase along with the kernel size;
- ✓ Layer-wise sharing generator makes it friendly to optimization process.

MULTIMEDIA COMPUTING GROUP
媒体计算研究组

Our Solution

➢ Pre-aggregation



Local offset

$\{a, b, c\}$

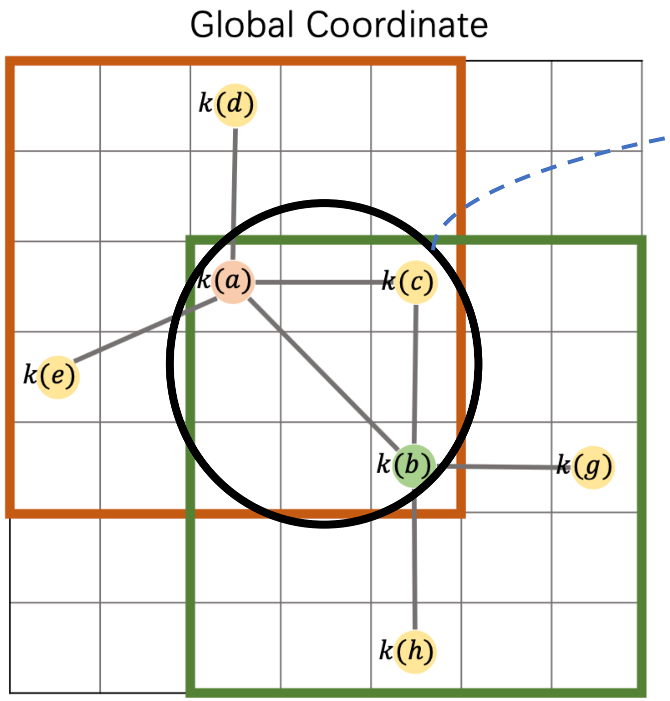
$$a = w(a-a)\cdot f_a + w(b-a)\cdot f_b + w(c-a)\cdot f_c$$

$$b = w(a-b)\cdot f_a + w(b-b)\cdot f_b + w(c-b)\cdot f_c$$

The overlap area is processed repeatedly!

Our Solution

➢Pre-aggregation



$$f_O = k(a)\cdot f_a + k(b)\cdot f_b + k(c)\cdot f_c.$$

$$w(x-y) = k(x)k(-y)$$

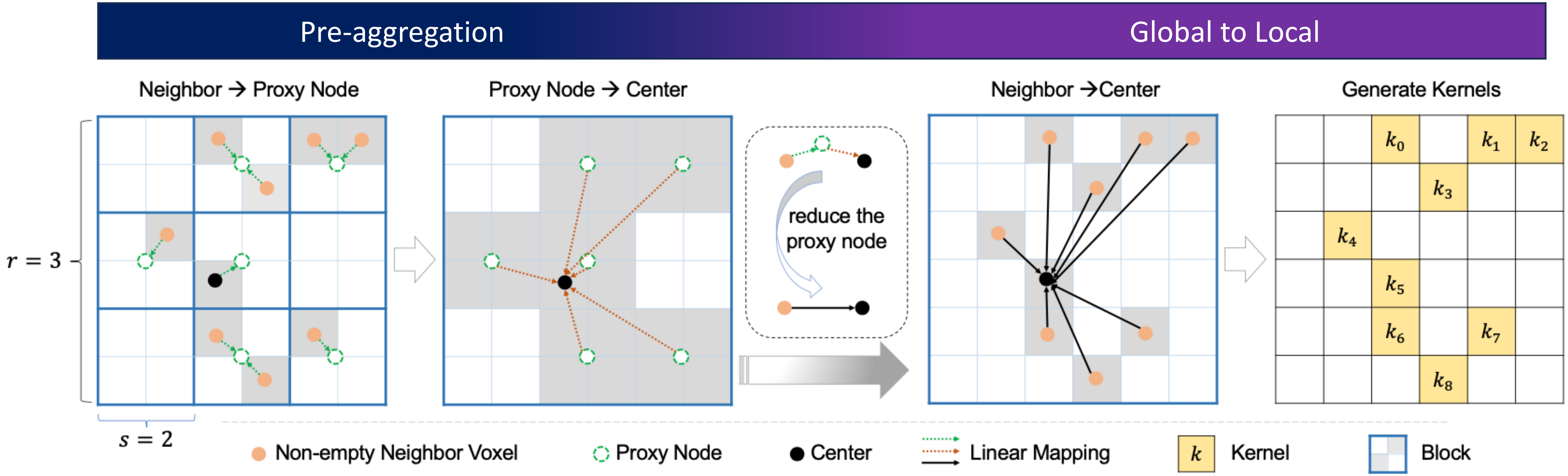
$$\begin{cases} f_{O\to a} = f_O \cdot k(-a) = \sum_{p\in O} w(p-a)\cdot f_{a+(p-a)}, \\ f_{O\to b} = f_O \cdot k(-b) = \sum_{p\in O} w(p-b)\cdot f_{b+(p-b)}. \end{cases}$$

Global coordinate

Pre-aggregation with global coordinate makes the overlap area reusable!

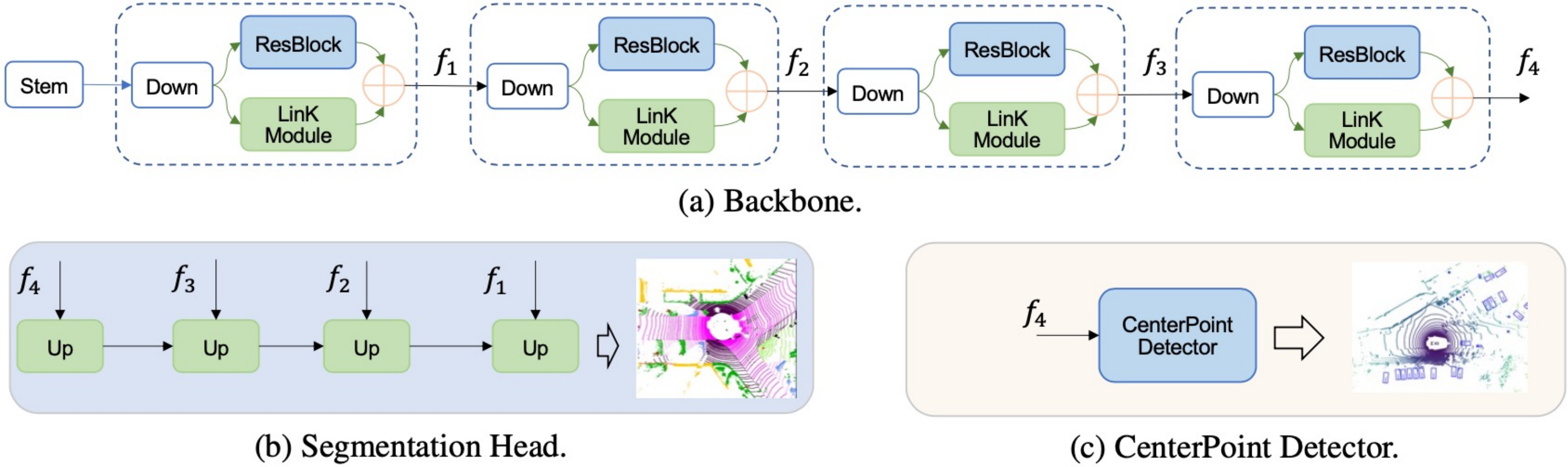
Our Solution

➢Full pipeline of LinK



Our Solution

➢Network Architecture



(a) Architecture of the LinK-based backbone; (b) the constructed network for 3D semantic segmentation; (c) the constructed network for 3D object detection.

Experiment: Detection

Table 1. Results on the test phase of nuScenes Detection. **Bold**: best results. * denotes using TTA.

Methods	Source	NDS	mAP	car	truck	bus	trailer	construction _vehicle	pedestrian	motorcycle	bicycle	traffic_cone	barrier
PointPillars [31]	*CVPR19*	45.3	30.5	68.4	23.0	28.2	23.4	4.1	59.7	27.4	1.1	30.8	38.9
3DSSD [47]	*CVPR20*	56.4	42.6	81.2	47.2	61.4	30.5	12.6	70.2	36.0	8.6	31.1	47.9
CenterPoint [36]	*CVPR21*	65.5	58.0	84.6	51.0	60.2	53.2	17.5	83.4	53.7	28.7	76.7	70.9
HotSpotNet [48]	*ECCV20*	66.0	59.3	83.1	50.9	56.4	53.3	23.0	81.3	63.5	36.6	73.0	71.6
TransFusion-L [39]	*CVPR22*	70.2	65.5	86.2	**56.7**	66.3	58.8	28.2	86.1	68.3	44.2	**82.0**	**78.2**
Focals Conv [49]	*CVPR22*	70.0	63.8	**86.7**	56.3	**67.7**	59.5	23.8	**87.5**	64.5	36.3	81.4	74.1
LargeKernel [1]	*arXiv22*	70.5	65.3	85.9	55.3	66.2	60.2	26.8	85.6	72.5	46.6	80.0	74.3
LinK	*Ours*	**71.0**	**66.3**	86.1	55.7	65.7	**62.1**	**30.9**	85.8	**73.5**	**47.5**	80.4	75.5
VISTA* [50]	*CVPR22*	70.4	63.7	84.7	54.2	64.0	55.0	29.1	83.6	71.0	45.2	78.6	71.8
UVTR-LiDAR* [51]	*NeurIPS22*	69.7	63.9	86.3	52.2	62.8	59.7	33.7	84.5	68.8	41.1	74.7	74.9
MDRNet* [52]	*arXiv22*	72.8	68.4	**87.9**	58.5	67.3	64.1	30.2	**89.0**	77.0	50.7	**85.0**	74.7
LargeKernel3D* [1]	*arXiv22*	72.8	68.8	87.3	59.1	68.5	65.6	30.2	88.3	77.8	53.5	82.4	75.0
LinK*	*Ours*	**73.4**	**69.8**	87.3	**60.2**	**69.8**	**65.9**	**34.0**	88.2	**78.8**	**54.3**	83.0	**76.8**

Experiment: Segmentation

Table 2. SemanticKITTI test results. Red: surpassing the baseline; **bold**: best results; 'P': point cloud; 'R': range map; 'V': voxel.

Method	Input	mIoU	Car	Bicycle	Motorcycle	Truck	Other-vehicle	Person	Bicyclist	Motorcyclist	Road	Parking	Sidewalk	Other-ground	Building	Fence	Vegetation	Trunk	Terrain	Pole	Traffic-sign
RandLA-Net [41]	P	53.9	94.2	26.0	25.8	40.1	38.9	49.2	48.2	7.2	90.7	60.3	73.7	20.4	86.9	56.3	81.4	61.3	66.8	49.2	47.7
RangeNet++ [60]	R	52.2	91.4	25.7	34.4	25.7	23.0	38.3	38.8	4.8	91.8	65.0	75.2	27.8	87.4	58.6	80.5	55.1	64.6	47.9	55.9
SqueezeSegV3 [61]	R	55.9	92.5	38.7	36.5	29.6	33.0	45.6	46.2	20.1	91.7	63.4	74.8	26.4	89.0	59.4	82.0	58.7	65.4	49.6	58.9
SalsaNext [62]	R	59.5	91.9	48.3	38.6	38.9	31.9	60.2	59.0	19.4	91.7	63.7	75.8	29.1	90.2	64.2	81.8	63.6	66.5	54.3	62.1
SPVNAS [42]	P+V	67.0	97.2	50.6	50.4	56.6	58.0	67.4	67.1	50.3	90.2	67.6	75.4	21.8	91.6	66.9	86.1	73.4	71.0	64.3	67.3
Cylinder3D [43]	V	67.8	97.1	67.6	64.0	59.0	58.6	73.9	67.9	36.0	91.4	65.1	75.5	32.3	91.0	66.5	85.4	71.8	68.5	62.6	65.6
(AF)2-S3Net [63]	V	69.7	94.5	65.4	**86.8**	39.2	41.1	**80.7**	**80.4**	**74.3**	91.3	68.8	72.5	**53.5**	87.9	63.2	70.2	68.5	53.7	61.5	**71.0**
DRINet [64]	P+V	67.5	96.9	57.0	56.0	43.3	54.5	69.4	75.1	58.9	90.7	65.0	75.2	26.2	91.5	67.3	85.2	72.6	68.8	63.5	66.0
RPVNet [44]	R+P+V	70.3	**97.6**	**68.4**	68.7	44.2	61.1	75.9	74.4	73.4	**93.4**	**70.3**	**80.7**	33.3	**93.5**	**72.1**	**86.5**	**75.1**	**71.7**	**64.8**	61.4
Mink(baseline) [15]	V	68.0	97.1	51.8	56.4	43.3	56.8	70.2	75.7	51.8	89.9	67.8	74.8	32.9	91.5	66.5	86.2	74.6	71.0	63.5	70.0
LinK(Ours)	V	**70.7**	97.4	58.4	56.6	52.9	**64.2**	72.3	77.0	69.1	90.6	68.2	76.2	34.5	92.0	68.8	85.7	74.3	70.5	**64.8**	69.5

Experiment: Ablations

- How does large kernel work?
  - ✓ Large objects benefit greatly.

Table 5. Performance on different scale objects.

Category	Size(m^3)	Detection		Segmentation	
		Center Point	+LinK	Mink	+LinK
Truck	$6 \times 2 \times 2$	51.0	(+4.7)55.7	43.3	(+9.6)52.9
Person	$0.4 \times 0.4 \times 2$	83.4	(+2.4)85.8	70.2	(+2.1)72.3

- The influence of kernel size

Table 1. Different kernel sizes for segmentation. Without TTA.

$r \times s$	mIoU(%)@SemKITTI val
3×2	66.9
3×3	67.3
3×5	67.5
3×7	67.2

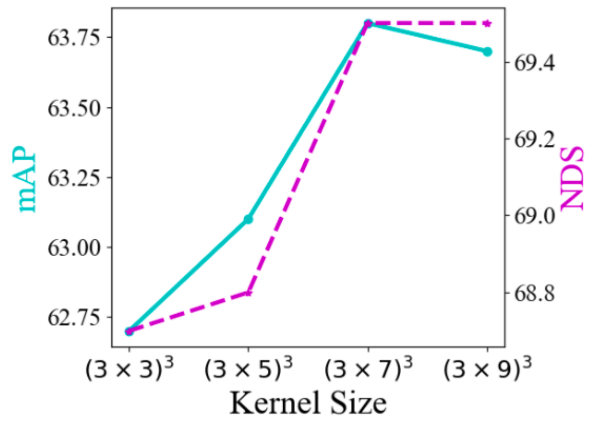


Figure 7. Detection performance with different kernel sizes.

Visualizations

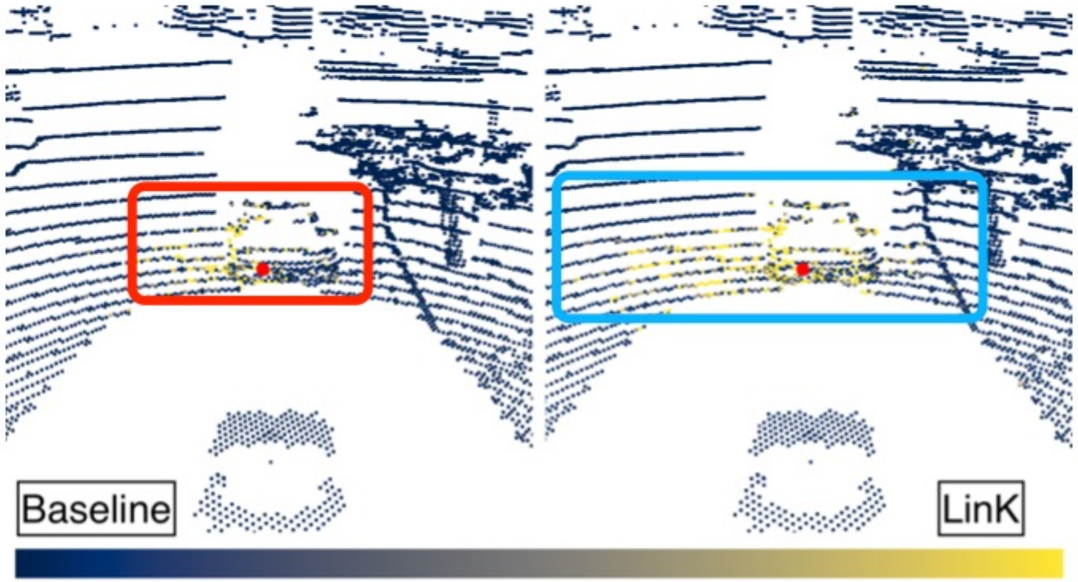
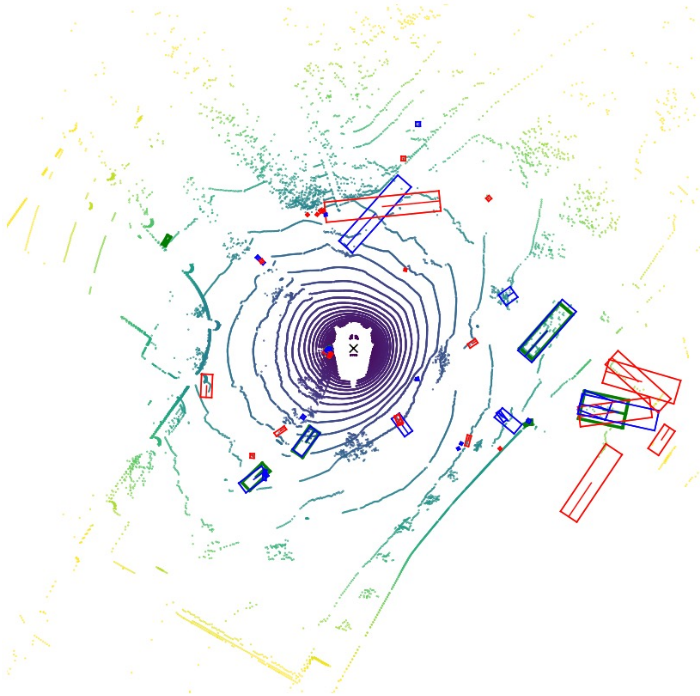
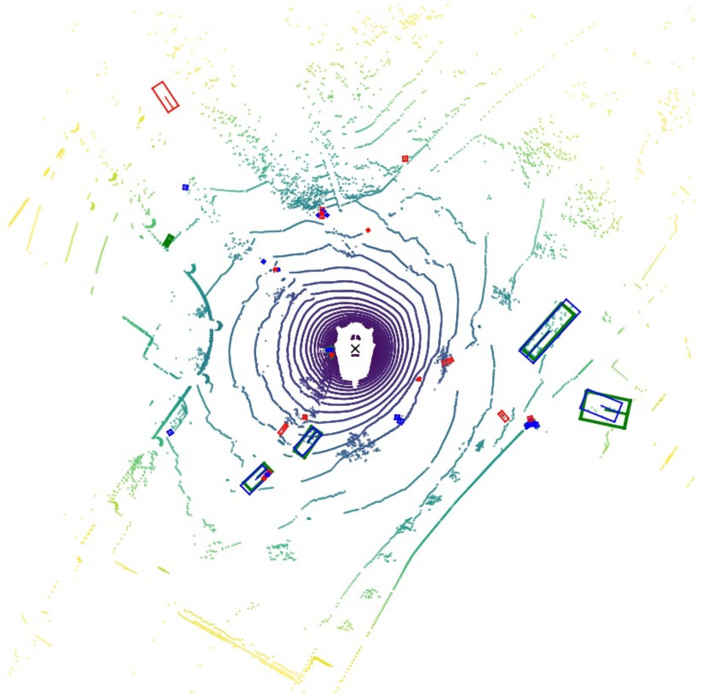


Figure 6. The effective receptive field (ERF) of the detection. The brightness indicates the degree of activation. LinK enjoys a wider-range perception.

Visualizations



(a) Baseline



(b) LinK

Thanks!

Code

Contact

- Tao Lu: taolu@smail.nju.edu.cn
- Xiang Ding: xding@smail.nju.edu.cn
- Haisong Liu: liuhs@smail.nju.edu.cn

Paper