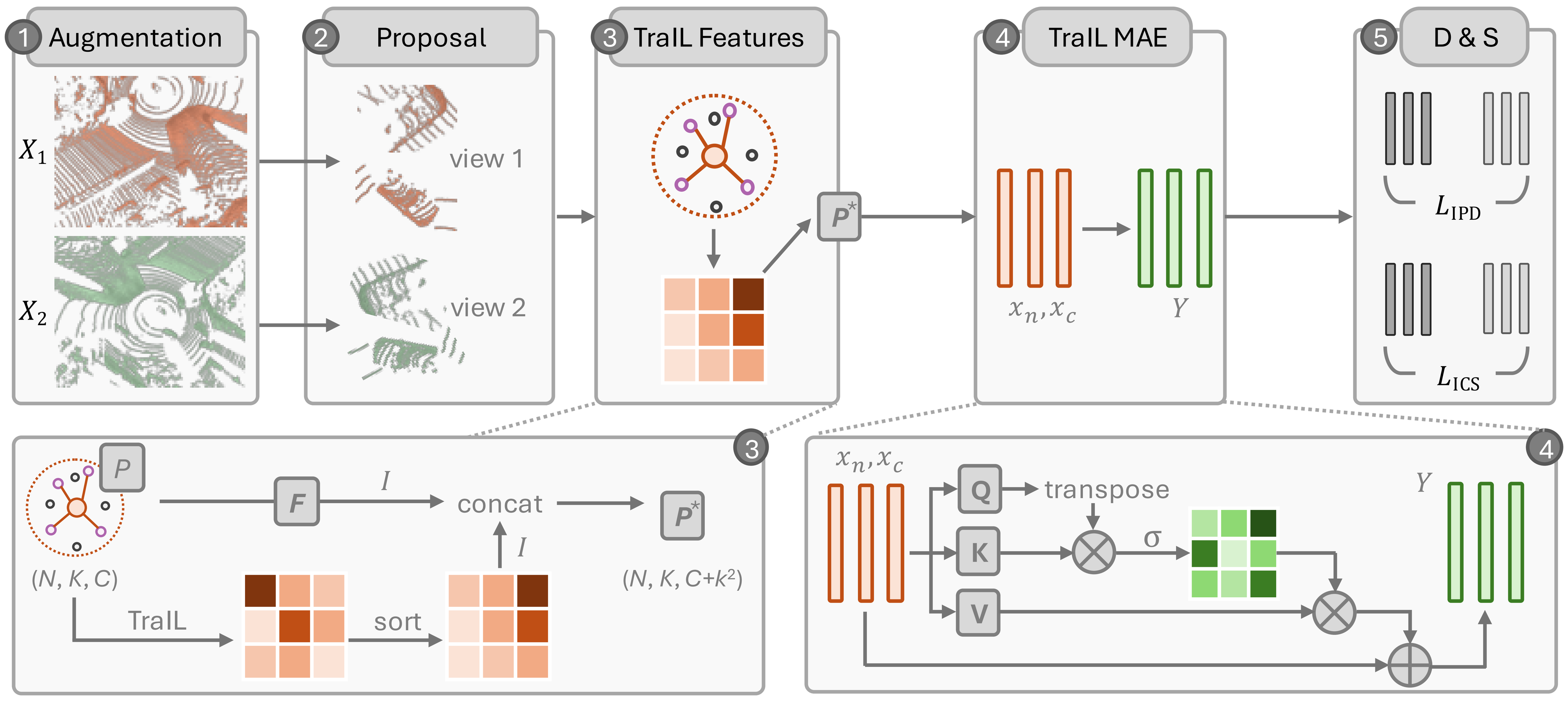
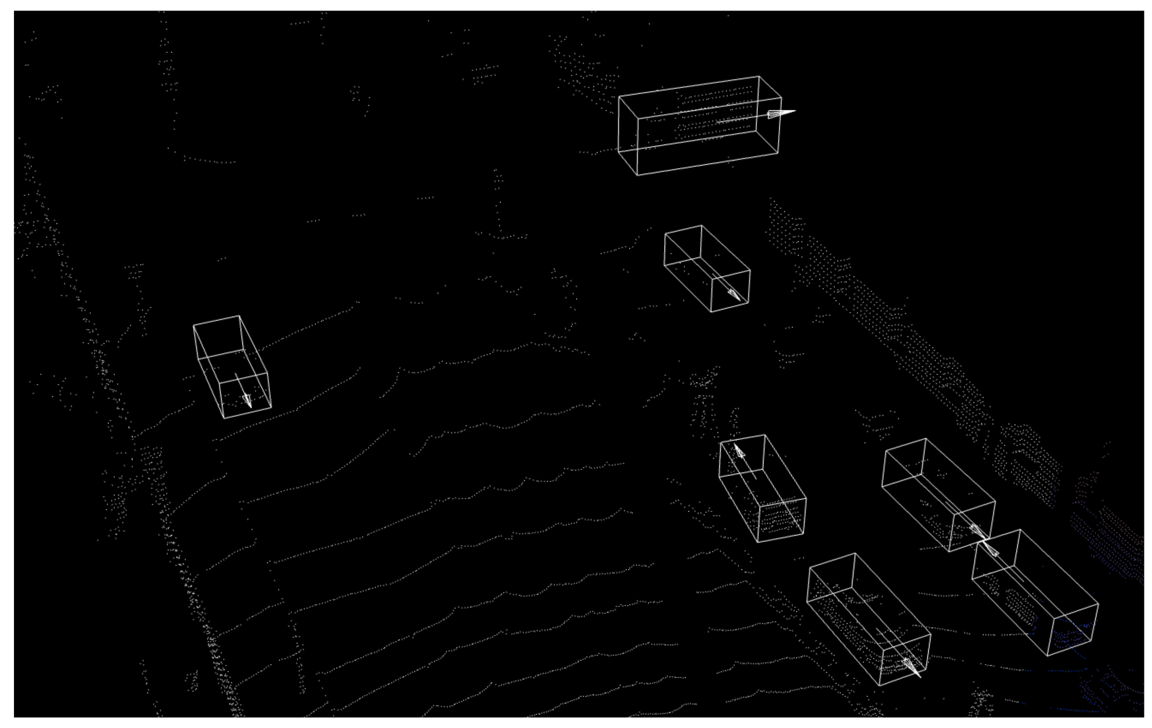
TraIL-Det: Transformation-Invariant Local Feature Networks for 3D LiDAR Object Detection with Unsupervised Pre-Training

BMVC 2024

Li Li, Tanqiu Qiao, Hubert P. H. Shum, and Toby P. Breckon
Durham University

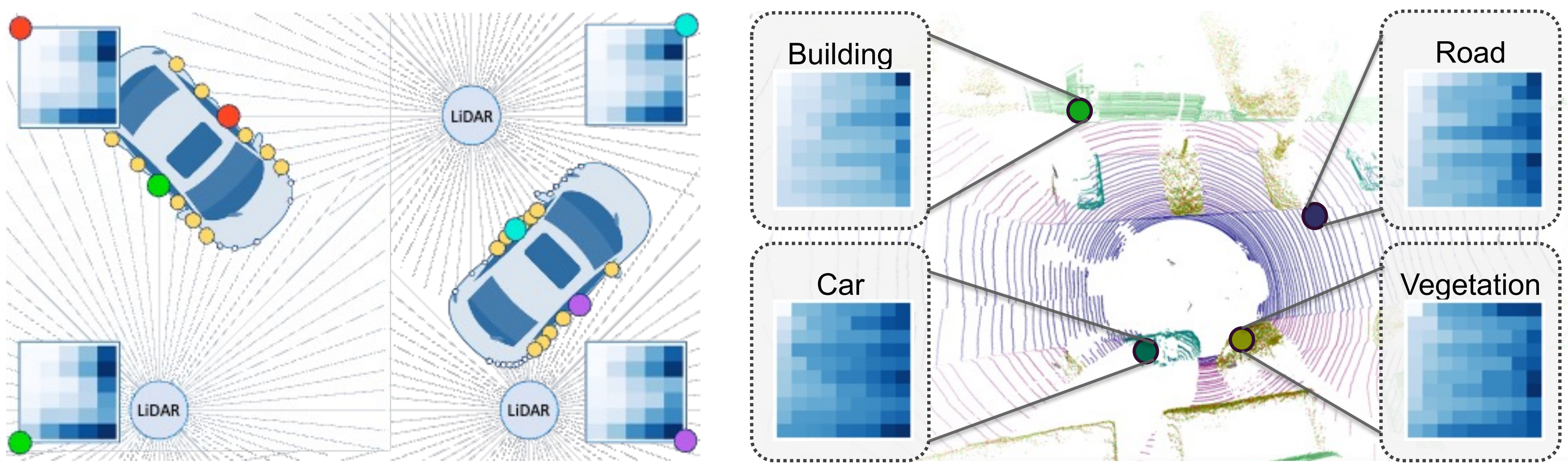




Motivations & Contributions

TraIL-Det Architecture

- To Enhance local geometry representation and detection accuracy, our TraIL features and TraIL-Det architecture significantly improve 3D LiDAR object detection pre-training in autonomous driving.



- A transformation-invariant local feature (TraIL) for 3D object detection, ensuring robustness to rigid transformations.
- An embedding method using Multi-head Self-Attention Encoder (MAE) for capturing geometric relations between points.
- A novel pre-training architecture (TraIL-Det) for 3D object detection that surpasses recent approaches.

Results & Conclusion

We use the standard SSL framework - pretrain a backbone network on large unlabeled data, then fine-tune it on downstream tasks with limited labeled data.

Table: Data-efficient 3D Object Detection on KITTI.

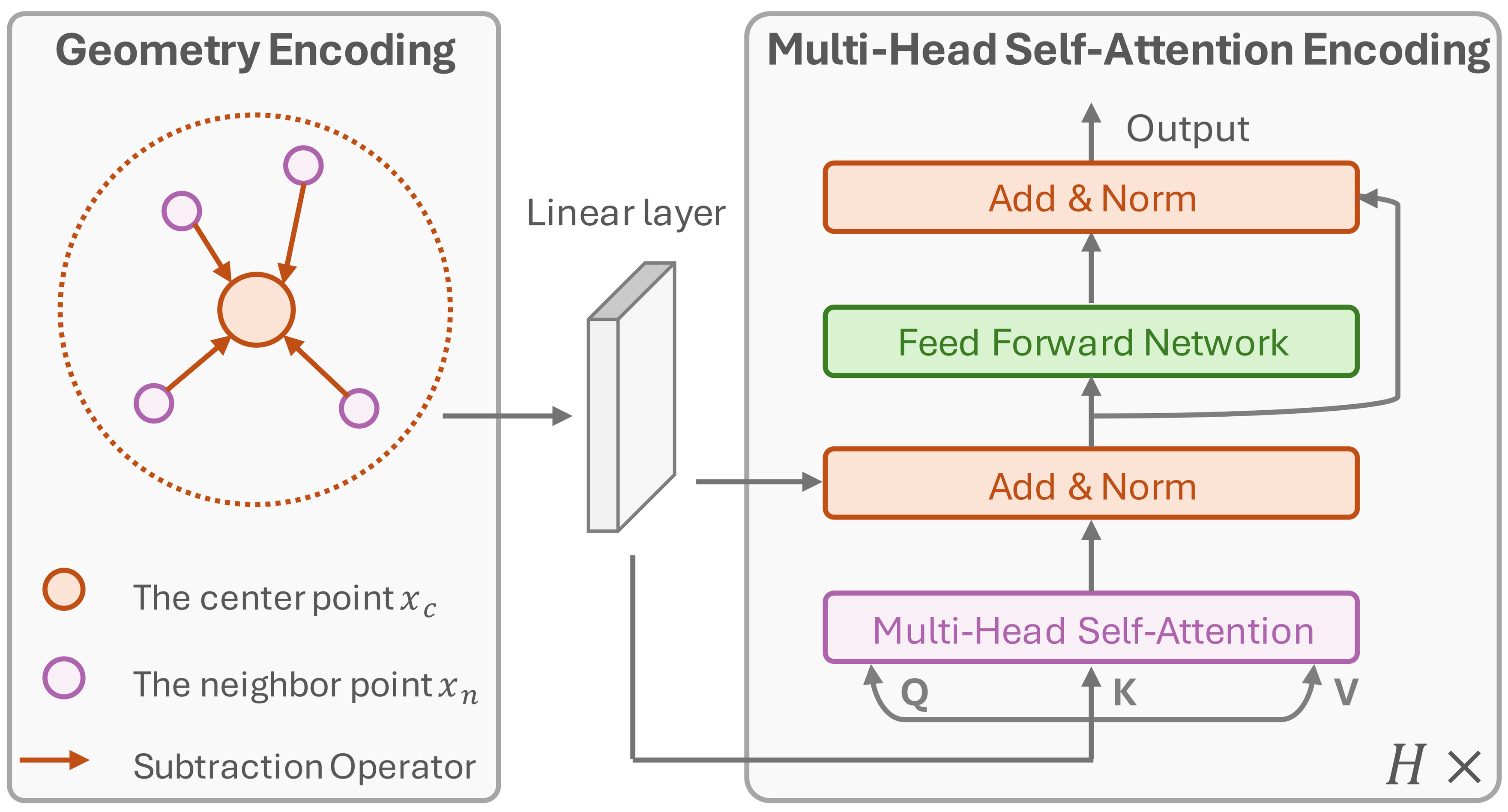
Fine-tuning with various label ratios	Detector	Pre-train. Schedule	mAP (Mod.)	Car Easy	Car Mod.	Car Hard	Pedestrian Easy	Pedestrian Mod.	Pedestrian Hard	Cyclist Easy	Cyclist Mod.	Cyclist Hard
20% (~ 0.7k frames)	PointRCNN	Scratch	63.51	88.64	75.23	72.47	55.49	48.90	42.23	85.41	66.39	61.74
		Prop.Con. [52]	66.20	88.52	77.02	72.56	58.66	51.90	44.98	90.27	69.67	65.05
		★ Ours	**67.80**	**89.07**	**78.86**	**73.63**	**59.12**	**53.37**	**46.11**	**92.95**	**71.16**	**66.12**
	PV-RCNN	Scratch	66.71	91.81	82.52	80.11	58.78	53.33	47.61	86.74	64.28	59.53
		Prop.Con. [52]	68.13	**91.96**	82.65	80.15	**62.58**	55.05	**50.06**	**88.58**	66.68	**62.32**
		★ Ours	**69.30**	91.88	**82.73**	**80.39**	62.22	**56.94**	49.85	88.43	**68.24**	61.19
50% (~ 1.8k frames)	PointRCNN	Scratch	66.73	89.12	77.85	75.36	61.82	54.58	**47.90**	86.30	67.76	63.26
		Prop.Con. [52]	69.23	89.32	79.97	**77.39**	62.19	54.47	46.49	**92.26**	73.25	68.51
		★ Ours	**69.77**	**90.47**	**81.23**	76.82	**64.15**	**54.79**	47.28	91.16	**73.29**	**71.13**
	PV-RCNN	Scratch	69.63	91.77	82.68	81.90	63.70	57.10	52.77	89.77	69.12	64.61
		Prop.Con. [52]	71.76	**92.29**	82.92	82.09	**65.82**	59.92	55.06	91.87	72.45	**67.53**
		★ Ours	**73.24**	90.15	**84.20**	**85.01**	64.28	**61.43**	**56.09**	**92.42**	**74.10**	66.23
100% (~ 3.7k frames)	PointRCNN	Scratch	69.45	90.02	80.56	**78.02**	62.59	55.66	48.69	89.87	72.12	67.52
		DepthCon. [55]	70.26	89.38	80.32	77.92	65.55	57.62	50.98	90.52	72.84	68.22
		Prop.Con. [52]	70.71	89.51	80.23	77.96	66.15	**58.82**	52.00	91.28	73.08	68.45
		★ Ours	**71.41**	**90.82**	**81.95**	77.85	**66.28**	58.73	**53.96**	**92.41**	**73.55**	**71.53**
	PV-RCNN	Scratch	70.57	-	84.50	-	-	57.06	-	-	70.14	-
		GCC-3D [24]	71.26	-	-	-	-	-	-	-	-	-
		STRL [14]	71.46	-	84.70	-	-	57.80	-	-	71.88	-
		PointCon. [46]	71.55	91.40	84.18	82.25	65.73	57.74	52.46	91.47	72.72	67.95
		Prop.Con. [52]	72.92	**92.45**	84.72	82.47	**68.43**	60.36	**55.01**	92.77	73.69	69.51
		★ Ours	**73.89**	92.10	**85.39**	**84.12**	68.01	**61.25**	54.29	**93.46**	**75.04**	**72.49**

Methodology

TraIL Features

- TraIL is defined for a point cloud patch X with K points where $K > k$ and k is the count of the nearest neighbours of a point, forming an $K \times k$ matrix TraIL(X; k). Each row i of this matrix includes the ordered distances from the i-th point in X to its k nearest neighbours.

Multi-attention Geometric Encoding



We propose the multi-attention geometric encoding. It computes asymmetric geometric features from the proposal P^* using center and neighbor points via subtraction. These features are refined with a proposal-aware encoding module using multi-head self-attention.

$$\mathbf{Q} = \delta\left(\boldsymbol{x}_c\right), \quad \mathbf{K} = \theta\left(\boldsymbol{x}_n - \boldsymbol{x}_c\right), \quad \mathbf{V} = \gamma\left(\boldsymbol{x}_n - \boldsymbol{x}_c\right),$$

$$S^{(\mathrm{att})}\left(\mathbf{Q}, \mathbf{K}, \mathbf{V}\right) = \operatorname{softmax}\left(\frac{\mathbf{Q}_h \mathbf{K}_h^{\top}}{\sqrt{D'}}\right) \cdot \mathbf{V}.$$

Links & Connections

GitHub

Paper

More Publications