

PowerBEV: A Powerful Yet Lightweight Framework for Instance Prediction in Bird's-Eye View

Peizheng Li^{*,1,3}, Shuxiao Ding^{*,1,2}, Xieyuanli Chen², Niklas Hanselmann^{1,3}, Marius Cordts¹, Juergen Gall^{2,4}

^{*}Equal contribution, ¹Mercedes-Benz AG, ²University of Bonn, ³University of Tübingen,

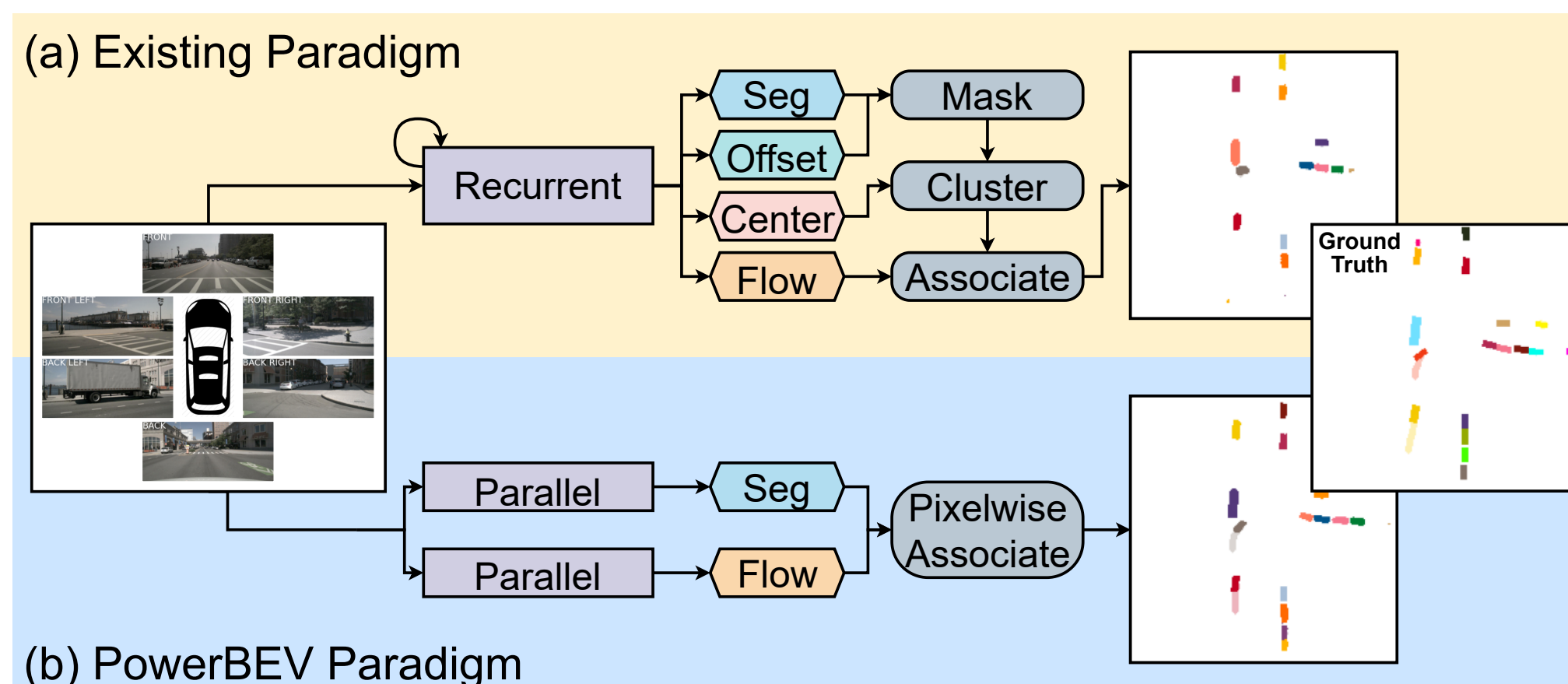
⁴Lamarr Institute for Machine Learning and Artificial Intelligence



Mercedes-Benz UNIVERSITÄT BONN

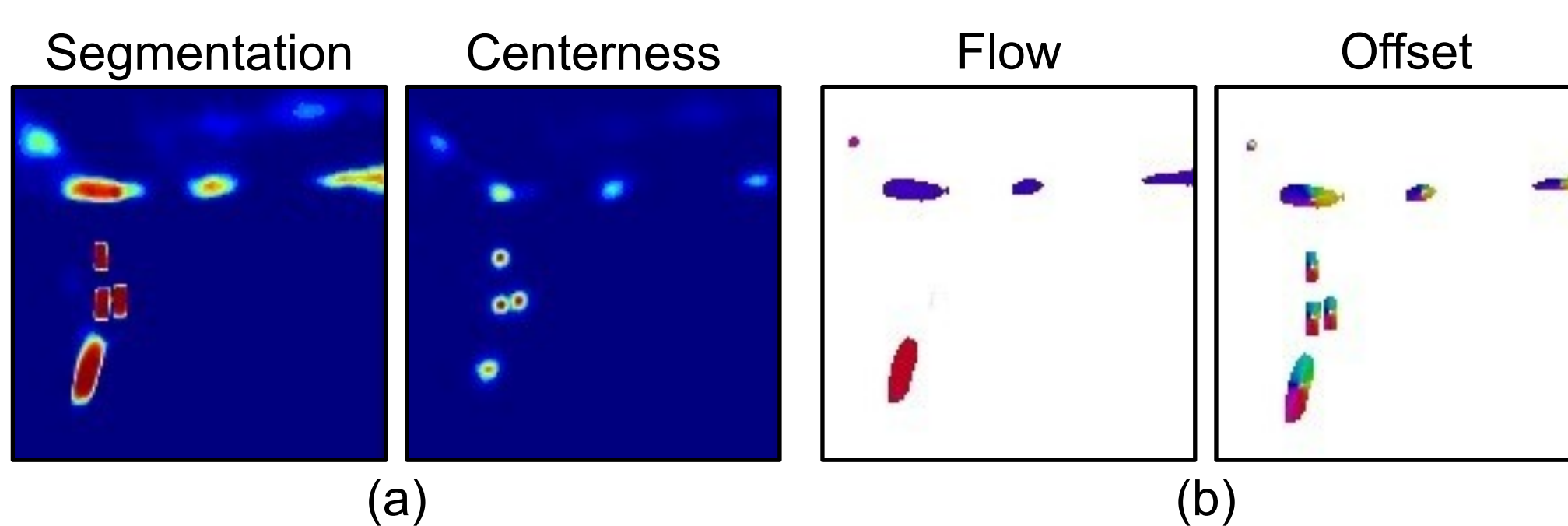


Motivation



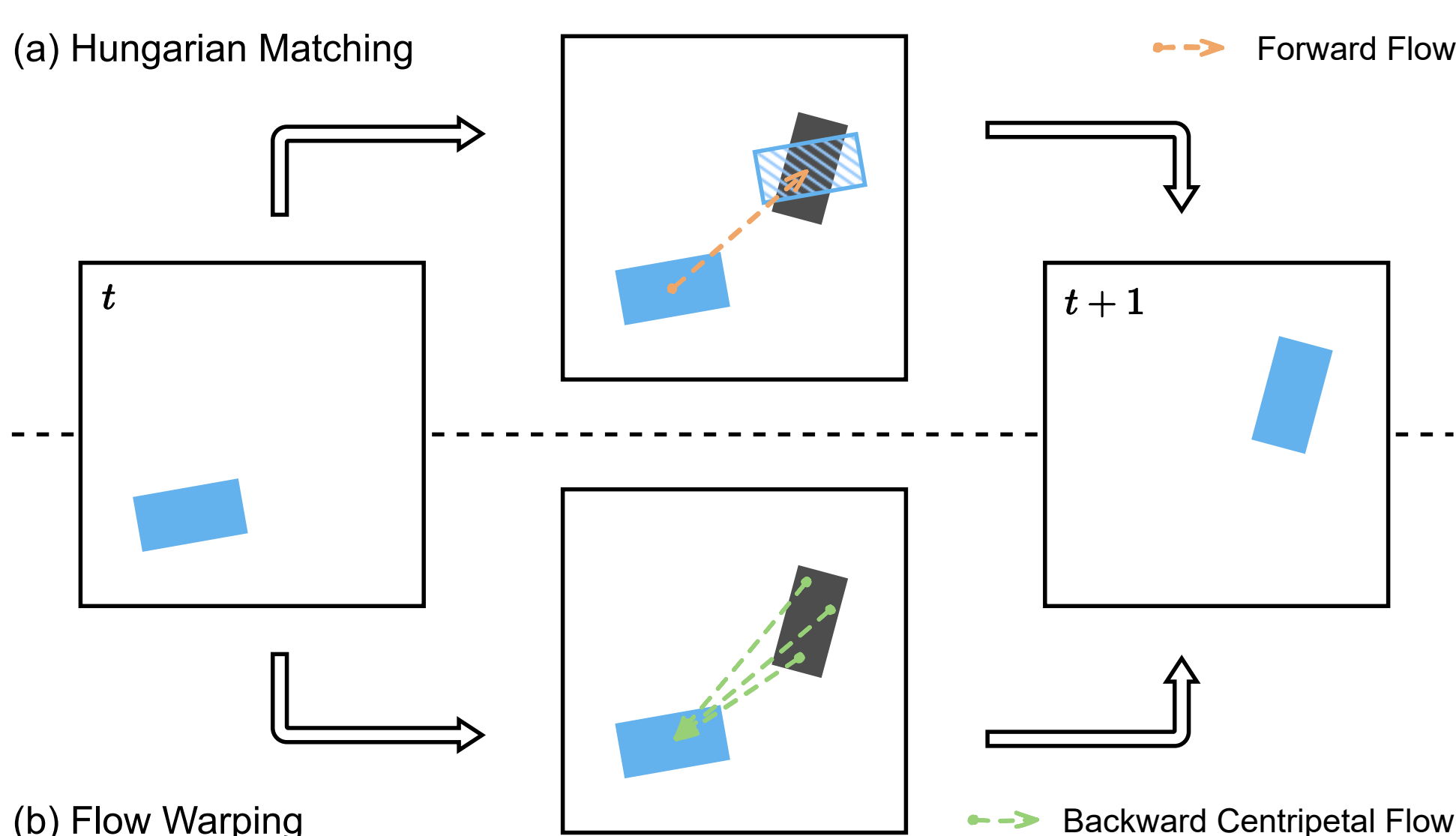
The existing BEV instance prediction paradigm (a), outputs 4 predictions per frame using spatial RNNs. To eliminate the framework redundancy, we propose a more lightweight yet powerful parallel prediction paradigm, namely PowerBEV (b). It consists entirely of 2D CNNs supplemented by flow warping post-processing based on only 2 outputs.

Task Similarities



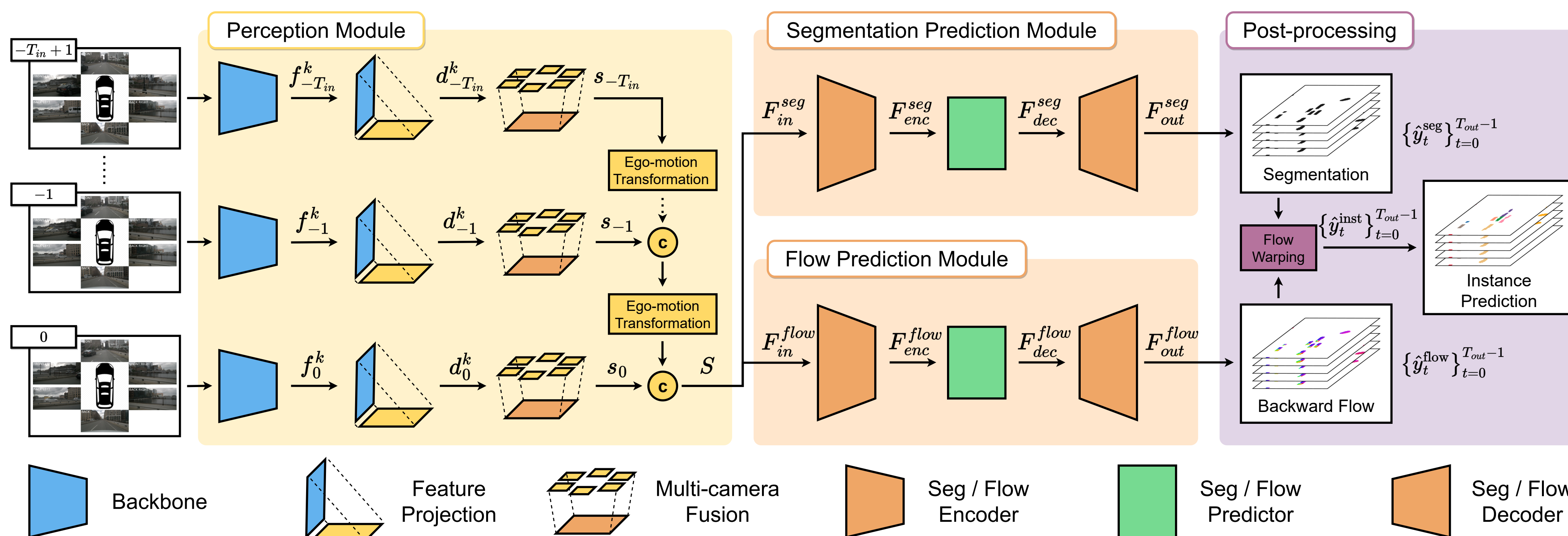
(a) The local maxima of the segmentation is highly similar to the centerness; (b) the flow can be understood as a motion offset.

Instance Matching



(a) Instance-level association using forward flow, (b) our proposed warping-based pixel-level association using backward centripetal flow.

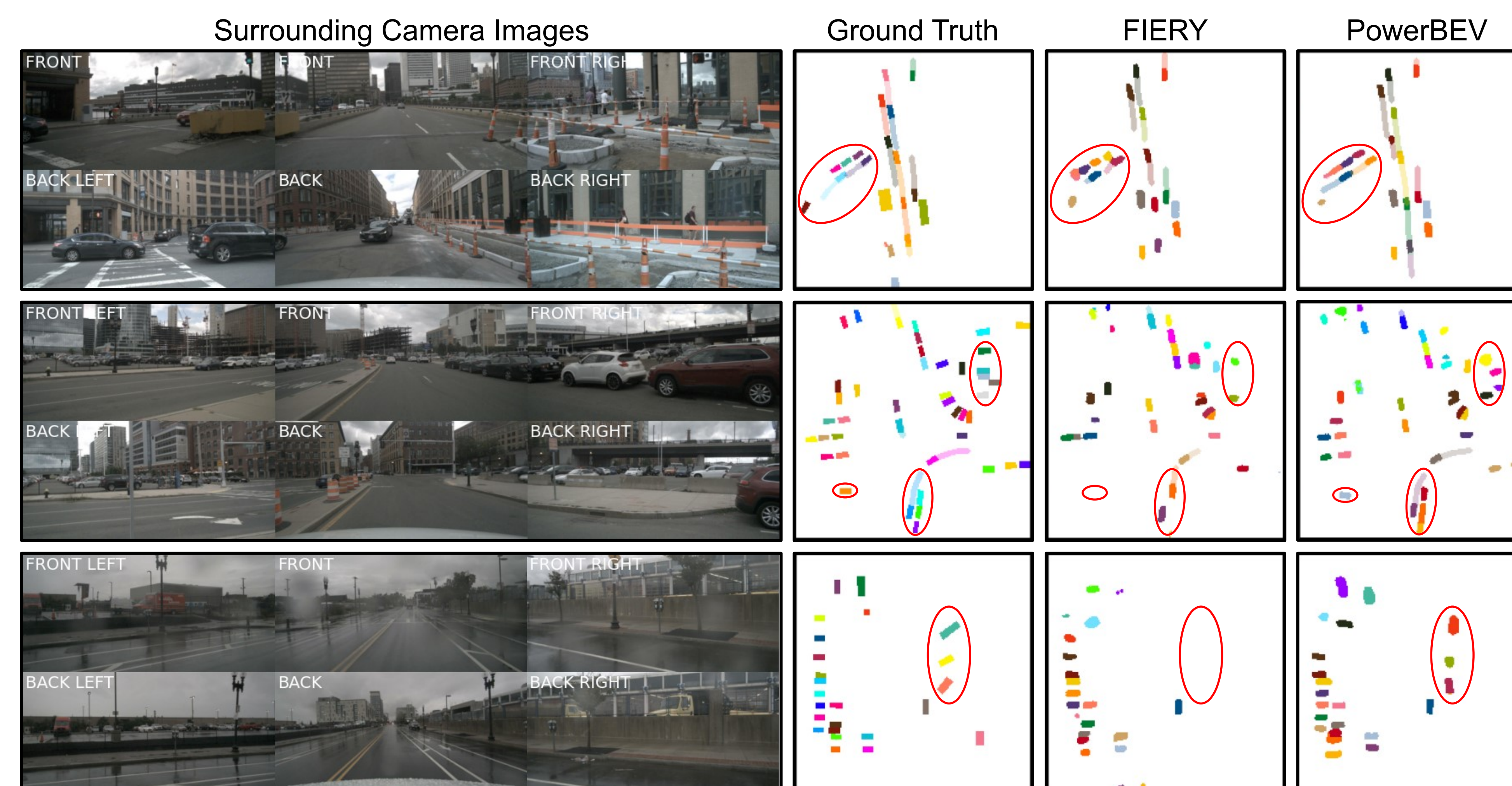
Overview



Runtime Analysis

Method	Time Horizon	#Param.	FLOPs	Runtime (ms)			Total
				Perception	Prediction	Post-proc.	
FIERY	2s	8.4M	206.8G	504	36	82	622
PowerBEV	2s	39.3M	92.6G	506	46	14	566
FIERY	8s	8.5M	709.5G	503	62	396	961
PowerBEV	8s	39.6M	108.4G	504	52	59	615

Visualization

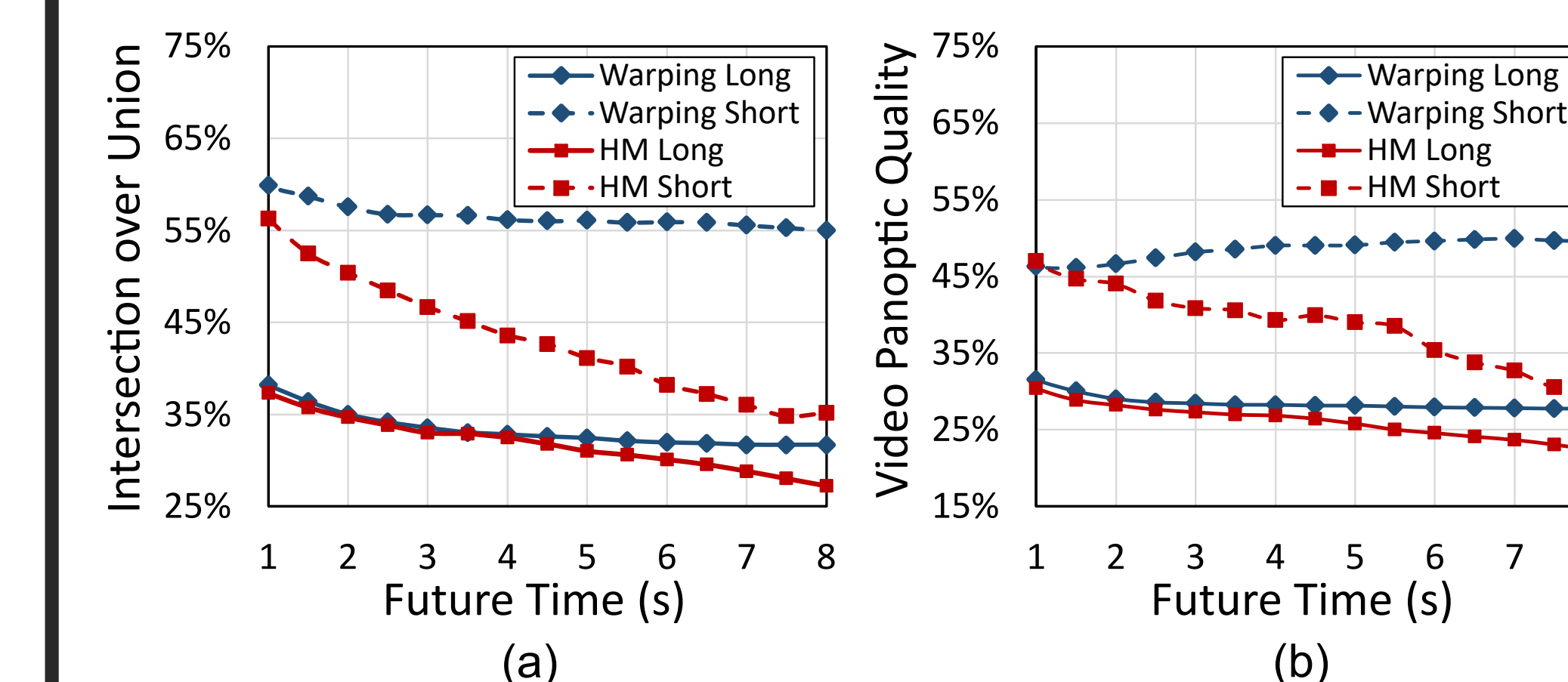


Main Results

Method	IoU		VPQ	
	Short	Long	Short	Long
FIERY [1]	59.4	36.7	50.2	29.9
FIERY [‡] (repr.) [1]	58.3	38.2	48.2	30.9
StretchBEV [2]	55.5	37.1	46.0	29.0
BEVerse [†] [3]	60.3	38.7	52.2	33.3
PowerBEV [‡]	62.5	39.3	55.5	33.8

Results on NuScenes validation split. [†] uses a larger image size of 704×256 , others use 480×224 . Models with [‡] use our optimized label generation.

Post-processing



Comparison of different post-processing methods for the 8s time horizon: (a) IoU and (b) VPQ.

References

- [1] Hu, Anthony, et al. "Fiery: Future instance prediction in bird's-eye view from surround monocular cameras." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.
- [2] Akan, Adil Kaan, and Fatma Güney. "Stretchbev: Stretching future instance prediction spatially and temporally." *European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 2022.
- [3] Zhang, Yunpeng, et al. "Beverse: Unified perception and prediction in birds-eye-view for vision-centric autonomous driving." *arXiv preprint arXiv:2205.09743* (2022).

Contact & Paper & Code

peizheng.li@mercedes-benz.com
 shuxiao.ding@mercedes-benz.com
 niklas.hanselmann@mercedes-benz.com
 Arxiv: <https://arxiv.org/abs/2306.10761>
 Code: <https://github.com/EdwardLeeLPZ/PowerBEV>

