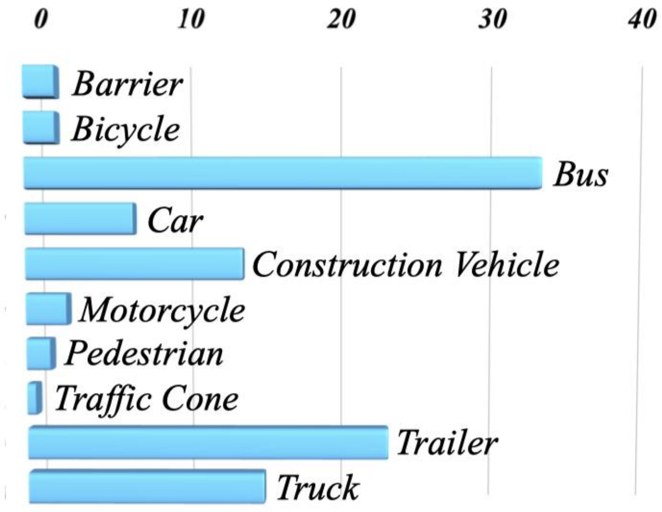
OctreeOcc: Efficient and Multi-Granularity Occupancy Prediction Using Octree Queries

OctreeOcc: Efficient and Multi-Granularity Occupancy Prediction Using Octree Queries

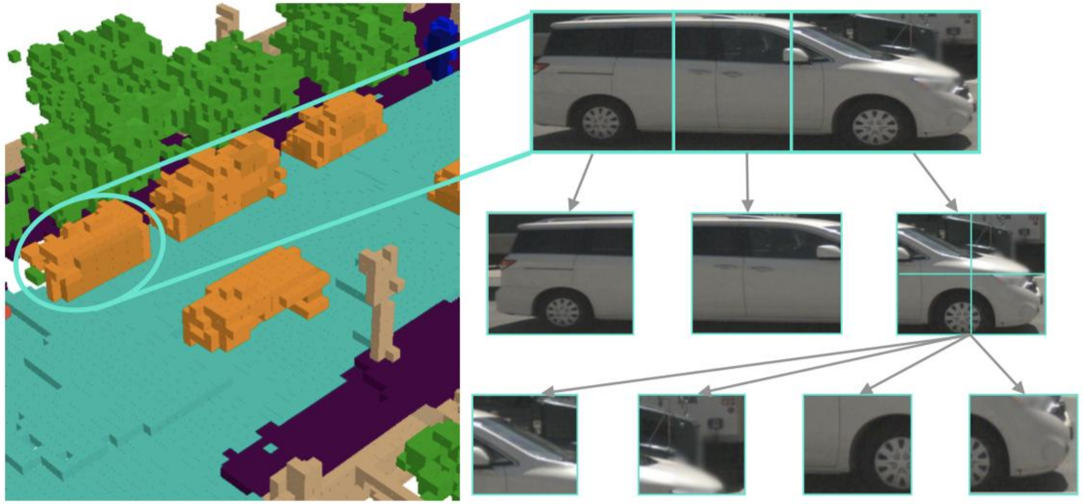
1. Traditional approaches often depend on dense, regular grid representations, resulting in **high computational costs**.

2. Different objects in the scene require **different granularity**.

3. Utilizing octree to **assign different voxel granularities to distinct semantic regions** reduces computational overhead while maintaining performance.



(a) Instance Size Inbalance



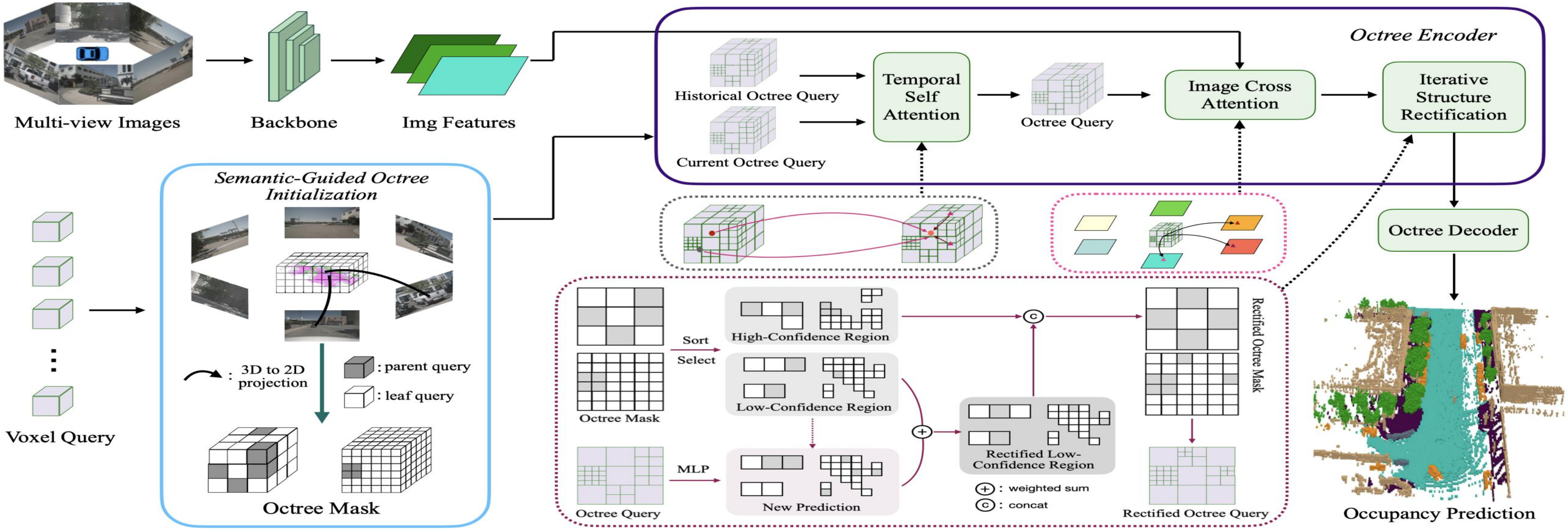
(b) 3D and 2D Visualization of Octree Structures

OctreeOcc: Efficient and Multi-Granularity Occupancy Prediction Using Octree Queries

1. **Semantic priors** are used to initialize the octree structure effectively.

2. After converting the dense query into an octree query, its features are updated via temporal and spatial attention.

3. The confidence of the octree structure is recalculated during query encoding, allowing **dynamic adjustments to the structure**.



Method Pipeline

OctreeOcc: Efficient and Multi-Granularity Occupancy Prediction Using Octree Queries

Method	Image Backbone	Reference	mIoU	others	barrier	bicycle	bus	car	const. veh.	motorcycle	pedestrain	traffic cone	trailer	truck	drive. suf.	other flat	sidewalk	terrain	manmade	vegetation
MonoScene [3]	ResNet101	CVPR'22	6.06	1.75	7.23	4.26	4.93	9.38	5.67	3.98	3.01	5.90	4.45	7.17	14.91	6.32	7.92	7.43	1.01	7.65
BEVDet [9]	ResNet101	arxiv'21	11.73	2.09	15.29	0.0	4.18	12.97	1.35	0.0	0.43	0.13	6.59	6.66	52.72	19.04	26.45	21.78	14.51	15.26
BEVFormer [21]	ResNet101	ECCV'22	23.67	5.03	38.79	9.98	34.41	41.09	13.24	16.50	18.15	17.83	18.66	27.70	48.95	27.73	29.08	25.38	15.41	14.46
BEVStereo [19]	ResNet101	AAAI'23	24.51	5.73	38.41	7.88	38.70	41.20	17.56	17.33	14.69	10.31	16.84	29.62	54.08	28.92	32.68	26.54	18.74	17.49
TPVFormer [11]	ResNet101	CVPR'23	28.34	6.67	39.20	14.24	41.54	46.98	19.21	22.64	17.87	14.54	30.20	35.51	56.18	33.65	35.69	31.61	19.97	16.12
OccFormer [56]	ResNet101	ICCV'23	21.93	5.94	30.29	12.32	34.40	39.17	14.44	16.45	17.22	9.27	13.90	26.36	50.99	30.96	34.66	22.73	6.76	6.97
CTF-Occ [43]	ResNet101	NeurIPS'23	28.53	8.09	39.33	20.56	38.29	42.24	16.93	24.52	22.72	21.05	22.98	31.11	53.33	33.84	37.98	33.23	20.79	18.00
RenderOcc [32]	ResNet101	ICRA'24	26.11	4.84	31.72	10.72	27.67	26.45	13.87	18.2	17.67	17.84	21.19	23.25	63.20	36.42	46.21	44.26	19.58	20.72
BEVDet4D [8]*	Swin-B	arxiv'22	42.02	**12.15**	49.63	25.1	52.02	54.46	27.87	27.99	28.94	27.23	36.43	42.22	82.31	43.29	54.46	57.9	48.61	43.55
PanoOcc [50]*	ResNet101	CVPR'24	42.13	11.67	50.48	29.64	49.44	55.52	23.29	**33.26**	30.55	30.99	34.43	42.57	83.31	44.23	54.40	56.04	45.94	40.40
FB-OCC [22]*	ResNet101	ICCV'23	43.41	12.10	50.23	**32.31**	48.55	52.89	**31.20**	31.25	**30.78**	**32.33**	37.06	40.22	**83.34**	**49.27**	**57.13**	**59.88**	47.67	41.76
Ours*	ResNet101	N/A	**44.02**	11.96	**51.70**	29.93	**53.52**	**56.77**	30.83	33.17	30.65	29.99	**37.76**	**43.87**	83.17	44.52	55.45	58.86	**49.52**	**46.33**

3D Occupancy prediction performance on Occ3D-nuScenes dataset

Methods	Query Form	mIoU	Latency	Memory
BEVFormer [21]	2D BEV	23.67	302ms	25100M
TPVFormer [11]	2D tri-plane	28.34	341ms	29000M
PanoOcc [50]	3D voxel	42.13	502ms	35000M
FBOCC [22]	3D voxels & 2D BEV	43.41	463ms	31000M
Ours	Octree Query	**44.02**	386ms	26500M

Comparison of **query form and efficiency** with SOTA methods on the Occ3D-nuScenes