Supplementary Material for ImGeoNet: Image-induced Geometry-aware Voxel Representation for Multi-view 3D Object Detection

A. Further Experiment

A.1. Object Sizes in ScanNet200

Table 1. **Mean object sizes (m^3) across different groups in ScanNet200**, which are determined from the training data.

	ScanNet200		
	Head	Common	Tail
Object size	0.92	0.14	0.03

We report the mean object sizes of different category groups within ScanNet200 [2]. The results presented in Table. 1 suggest that objects within the common and tail groups tend to have smaller sizes when compared to those within the head group.

A.2. Visualization

This study presents the qualitative results of 3D object detection on three distinct datasets, namely ARKitScenes (as depicted in Fig. 1 and Fig. 2), ScanNetV2 (as illustrated in Fig. 3), and ScanNet200 (as shown in Fig. 4). The effectiveness of our proposed geometry-aware representation can be verified by ImGeoNet's superior outcomes, in comparison to ImVoxelNet [3], across all datasets. Additionally, ImGeoNet generates more precise results than the seminal point cloud-based approach, VoteNet [1], in practical scenarios that involve sparse and noisy point clouds or diverse object classes.

A.3. Per-class Evaluation

We present the 3D object detection scores per class across different datasets. The outcomes obtained from ARKitScenes are provided in Table. 2. Similarly, the outcomes obtained from ScanNetV2 are presented in Table. 3 Finally, the results for ScanNet200 are provided separately for three category groups: head, common, and tail. They are presented in Table. 4, Table. 5, and Table. 6, respectively.

References

[1] Charles R Qi, Or Litany, Kaiming He, and Leonidas J Guibas. Deep hough voting for 3d object detection in point clouds. In *ICCV*, 2019. 1

[2] David Rozenberszki, Or Litany, and Angela Dai. Language-grounded indoor 3d semantic segmentation in the wild. In *ECCV*, 2022. 1

[3] Danila Rukhovich, Anna Vorontsova, and Anton Konushin. Imvoxelnet: Image to voxels projection for monocular and multi-view general-purpose 3d object detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022. 1

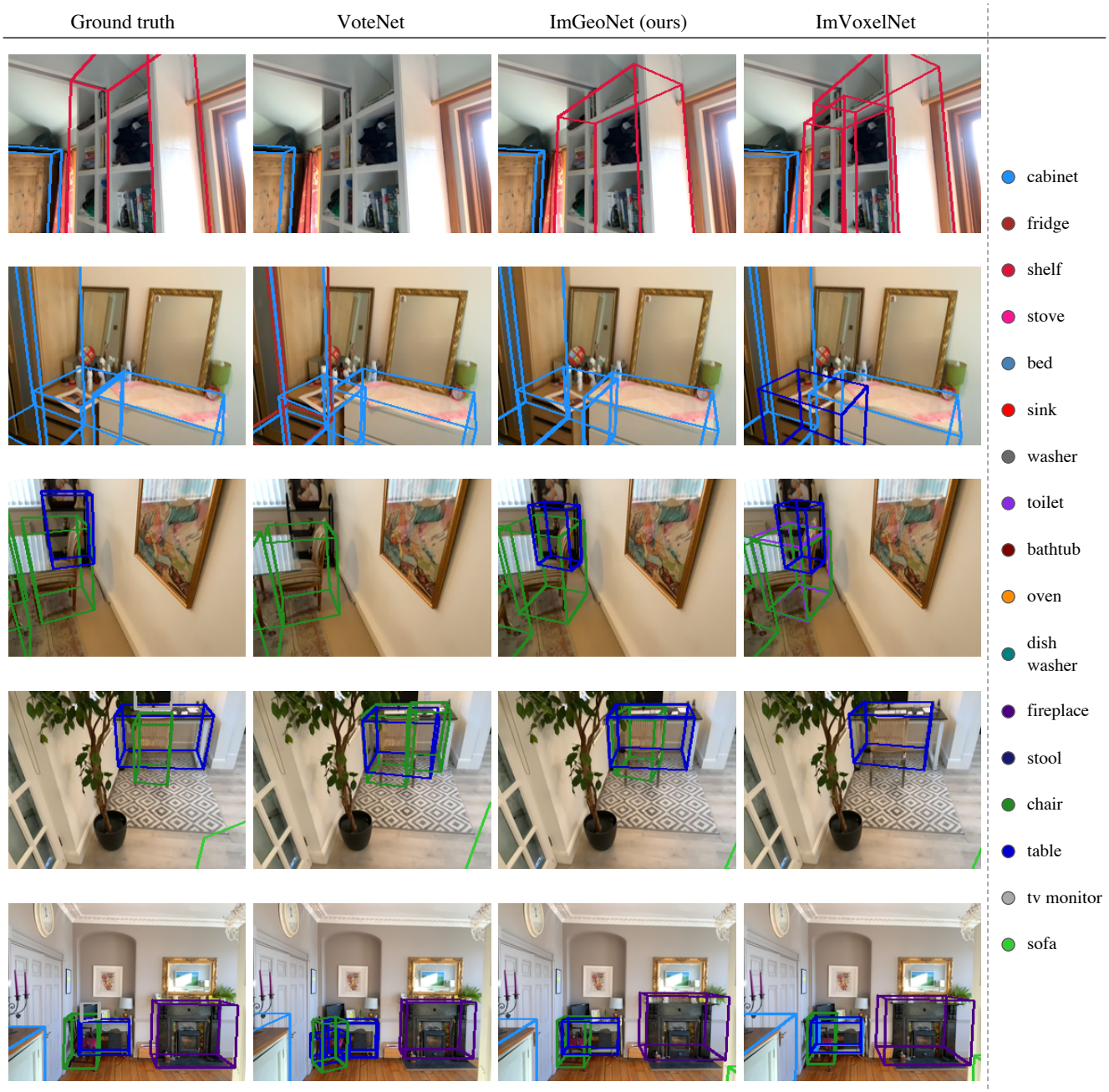

Figure 1. Qualitative results on ARKitScenes.

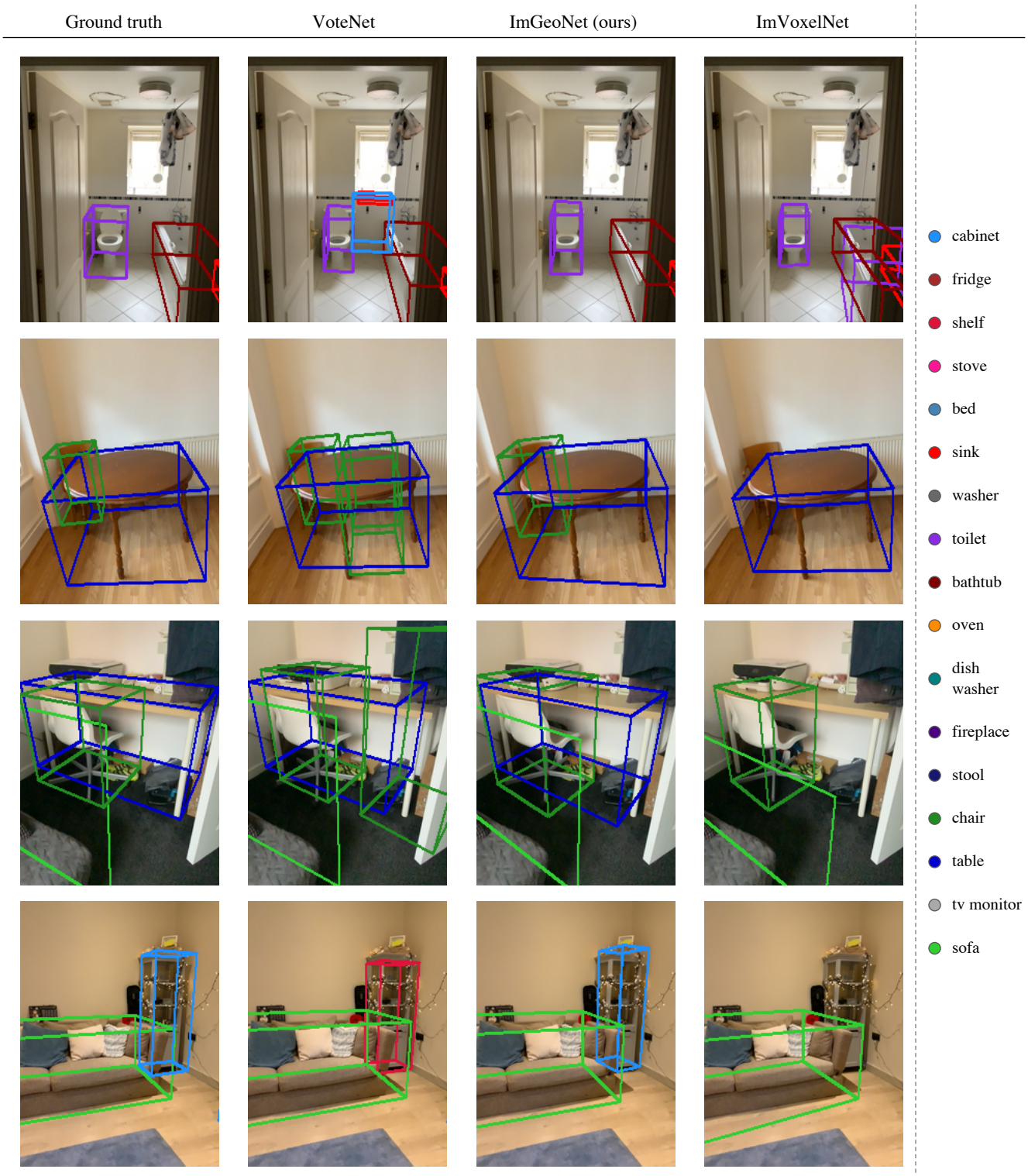

Figure 2. More qualitative results on ARKitScenes.

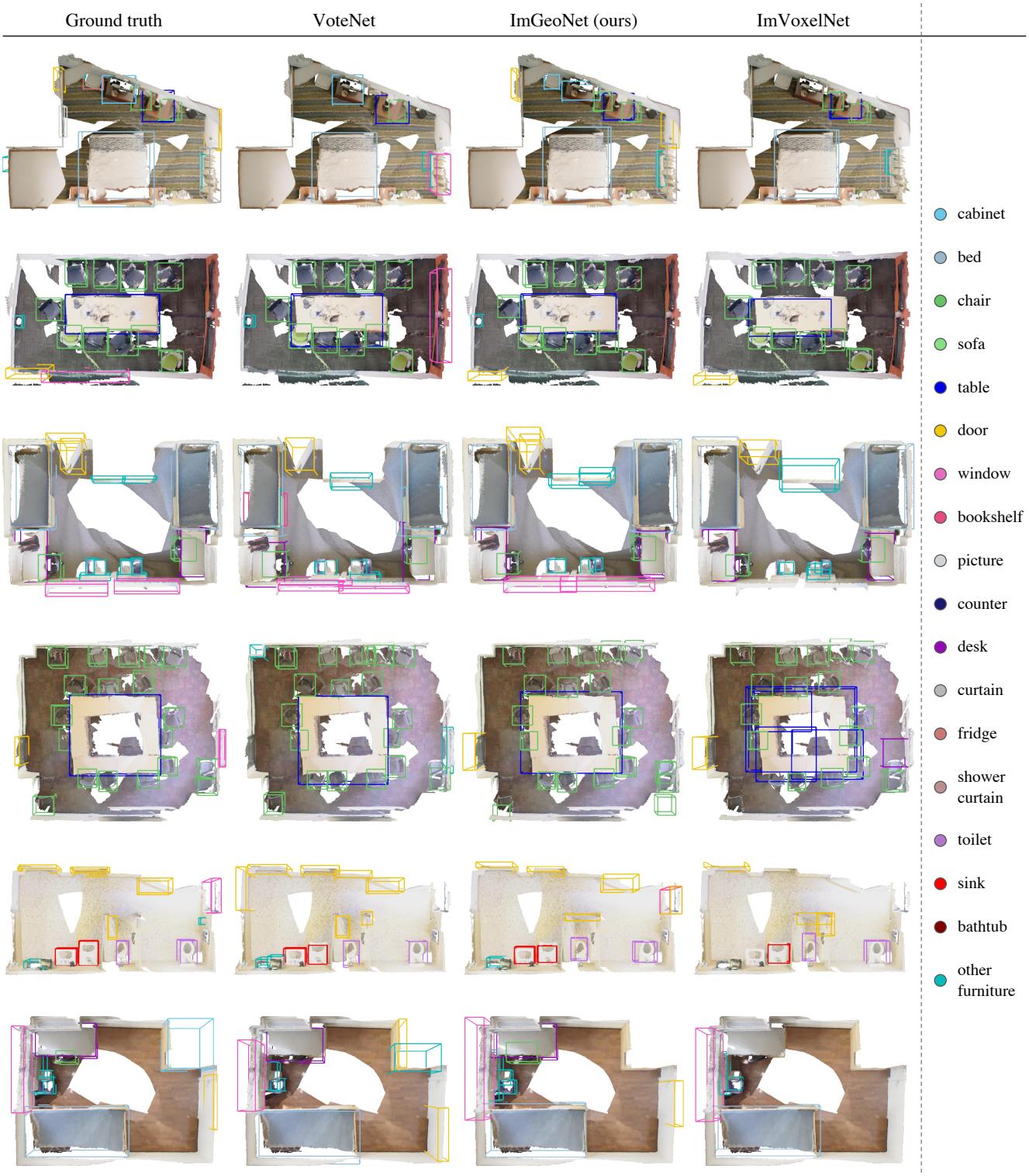

Figure 3. Qualitative results on ScanNetV2.

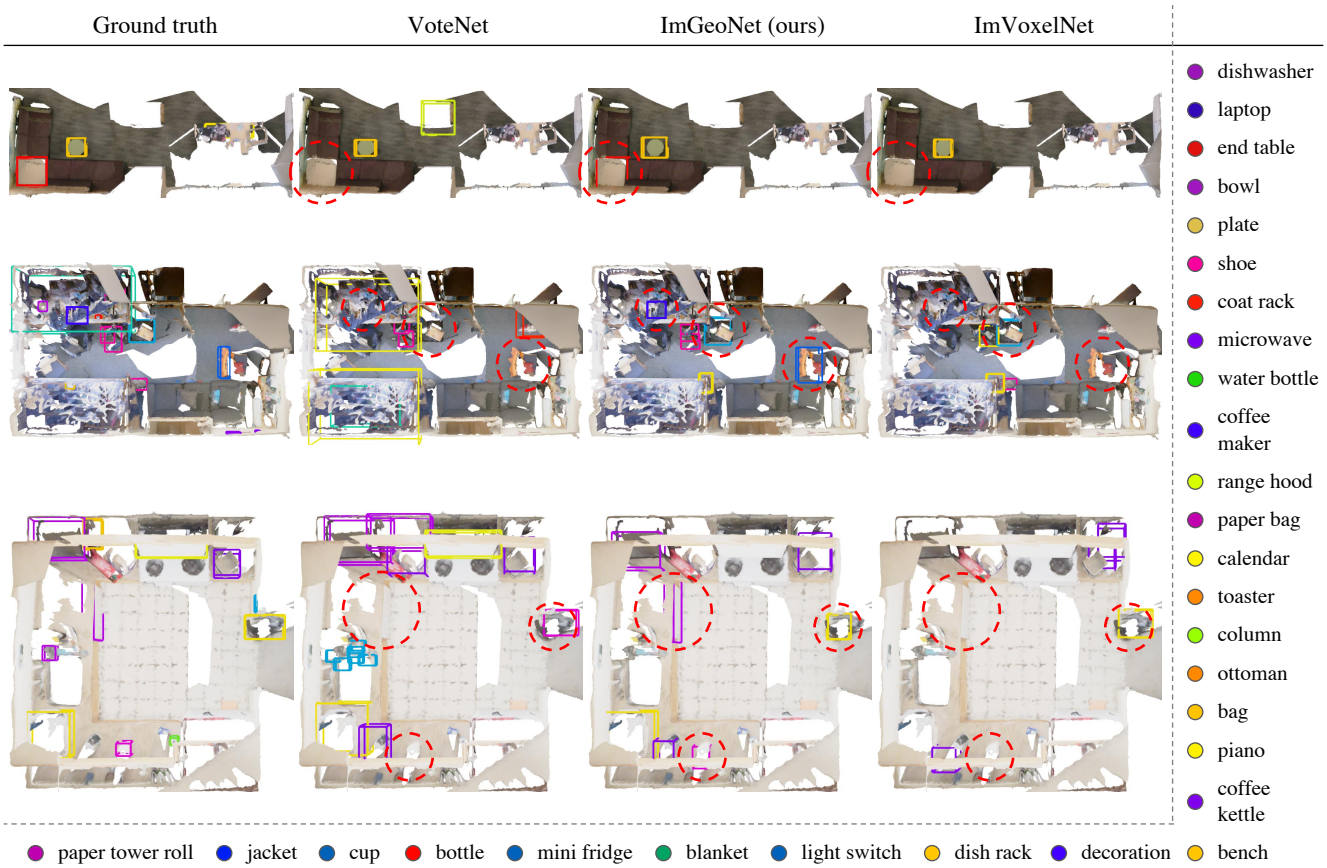

Figure 4. Qualitative results on ScanNet200. Only the common and tail category groups are presented.

Table 2. 3D object detection scores per class on ARKitScenes.

	cab	refrig	shelf	stove	bed	sink	washr	toil	bath	oven	dwashr	firepl	stool	chair	tabl	tvmnt	sofa	mAP@0.25
ImGeoNet (ours)	55.8	82.6	48.4	20.4	89.3	52.8	80.0	92.5	94.7	66.0	18.1	68.8	30.6	72.3	70.3	2.2	79.0	**60.2**
ImVoxelNet	53.1	82.3	42.4	13.5	87.5	50.4	77.4	92.3	94.6	64.1	28.6	55.1	25.1	72.9	66.8	2.0	77.7	58.0
VoteNet	62.1	79.6	26.0	4.9	92.9	43.5	56.5	91.9	94.1	37.4	6.4	67.8	18.6	57.4	62.8	21.6	83.2	53.3
	cab	refrig	shelf	stove	bed	sink	washr	toil	bath	oven	dwashr	firepl	stool	chair	tabl	tvmnt	sofa	mAP@0.5
ImGeoNet (ours)	31.8	72.5	21.7	3.9	83.3	19.9	71.2	84.8	91.0	44.4	15.9	23.1	13.3	49.3	45.1	0.1	67.2	**43.4**
ImVoxelNet	25.2	68.8	11.3	3.9	75.9	10.9	61.2	82.5	89.6	38.3	24.7	13.4	11.0	42.3	36.0	0.0	64.2	38.8
VoteNet	37.3	75.5	15.2	0.5	89.7	20.7	50.3	70.6	93.7	15.4	2.8	28.7	10.4	30.3	40.2	1.3	71.7	38.5

Table 3. 3D object detection scores per class on ScanNetV2.

	cab	bed	chair	sofa	tabl	door	wind	bkshf	pic	cntr	desk	curt	fridg	showr	toil	sink	bath	ofurn	mAP@0.25
ImGeoNet (ours)	38.7	86.5	76.6	75.7	59.3	42.0	28.1	59.2	4.3	42.8	71.5	36.9	51.8	44.1	95.2	58.0	79.6	36.8	**54.8**
ImVoxelNet	34.4	81.9	74.6	71.9	55.1	32.0	16.9	38.2	2.0	37.0	62.2	26.5	54.1	39.0	90.7	52.1	74.9	32.8	48.7
VoteNet	36.3	87.9	88.7	89.6	58.8	47.3	38.1	44.6	7.8	56.1	71.7	47.2	45.4	57.1	94.9	54.7	92.1	37.2	58.7
	cab	bed	chair	sofa	tabl	door	wind	bkshf	pic	cntr	desk	curt	fridg	showr	toil	sink	bath	ofurn	mAP@0.5
ImGeoNet (ours)	14.3	74.2	47.4	46.9	41.0	8.1	2.0	26.9	0.5	6.6	44.7	4.4	28.2	3.9	71.0	25.9	48.3	17.2	**28.4**
ImVoxelNet	11.0	69.3	42.3	30.1	32.2	4.8	0.9	11.3	0.2	2.7	39.2	4.0	23.1	3.2	67.5	20.3	55.5	10.9	23.8
VoteNet	8.1	76.1	67.2	68.8	42.4	15.3	6.4	28.0	1.3	9.5	37.5	11.6	27.8	10.0	86.5	16.8	78.9	11.7	33.5

Table 4. 3D object detection scores per class on ScanNet200's **head** group (IoU threshold = 0.25).

	wall	chair	floor	table	door	couch	cabinet	shelf
ImGeoNet (ours)	28.4	65.9	41.7	42.0	42.0	67.6	22.5	25.9
ImVoxelNet	18.9	64.0	43.5	39.0	31.1	69.4	21.8	19.9
VoteNet	46.5	79.0	96.6	50.3	53.9	89.5	34.1	35.7

	desk	office chair	bed	pillow	sink	picture	window	toilet
ImGeoNet (ours)	70.5	24.0	96.0	61.5	61.9	10.9	25.2	92.9
ImVoxelNet	64.4	20.7	91.5	45.4	56.2	4.2	12.8	93.0
VoteNet	71.9	26.8	93.2	27.2	47.2	6.1	34.0	96.7

	bookshelf	monitor	curtain	book	armchair	coffee table	box	refrigerator
ImGeoNet (ours)	52.2	68.0	30.6	33.7	57.7	74.3	24.9	49.5
ImVoxelNet	53.6	58.1	20.8	21.9	54.8	69.5	18.7	52.7
VoteNet	34.2	56.3	35.5	6.8	51.1	78.4	5.2	63.5

	lamp	kitchen cabinet	towel	clothes	tv	nightstand	counter	dresser
ImGeoNet (ours)	41.2	24.4	19.5	11.7	57.5	76.5	24.0	41.5
ImVoxelNet	35.3	23.3	11.8	5.5	43.3	71.0	9.2	44.3
VoteNet	34.1	33.8	15.1	9.0	49.2	71.1	24.8	22.5

	stool	plant	ceiling	bathtub	backpack	tv stand	whiteboard	shower curtain
ImGeoNet (ours)	15.2	45.5	0.0	71.2	62.9	52.3	27.9	38.3
ImVoxelNet	13.6	29.3	0.0	77.0	62.2	48.1	14.1	32.0
VoteNet	21.9	22.7	36.1	86.8	32.6	35.0	24.2	63.2

	trash can	closet	stairs	stove	board	washing machine	mirror	copier
ImGeoNet (ours)	59.4	10.1	43.3	70.6	5.1	48.8	12.4	79.8
ImVoxelNet	53.2	7.8	32.8	64.4	1.8	56.7	6.3	72.2
VoteNet	41.7	3.9	15.1	75.7	4.2	48.5	14.3	91.7

	sofa chair	file cabinet	shower	blinds	blackboard	radiator	recycling bin	wardrobe
ImGeoNet (ours)	2.6	39.7	6.3	0.0	45.4	33.0	60.3	12.6
ImVoxelNet	5.7	33.8	16.8	0.0	44.4	25.8	51.4	9.4
VoteNet	10.1	16.1	26.9	1.4	51.1	37.8	7.3	38.4

	clothes dryer	bathroom stall	shower wall	kitchen counter	doorframe	mailbox	object	bathroom vanity
ImGeoNet (ours)	0.0	47.0	56.8	31.8	18.2	7.0	1.2	44.3
ImVoxelNet	0.0	48.2	58.1	24.6	16.6	4.2	0.5	41.7
VoteNet	0.4	43.5	60.1	20.6	30.1	12.5	0.3	57.4

	closet wall	stair rail	**mAP**
ImGeoNet (ours)	0.1	1.1	**38.1**
ImVoxelNet	4.5	1.9	34.1
VoteNet	17.2	10.9	38.5

Table 5. 3D object detection scores per class on ScanNet200's **common** group (IoU threshold = 0.25).

	cushion	end table	dining table	keyboard	bag	toilet paper	printer
ImGeoNet (ours)	0.0	19.9	7.8	18.4	12.8	17.7	41.7
ImVoxelNet	0.0	17.8	3.3	5.5	10.4	12.1	29.4
VoteNet	0.1	17.2	5.7	0.4	2.7	2.1	29.6

	blanket	microwave	shoe	computer tower	bottle	bin	ottoman
ImGeoNet (ours)	19.3	51.5	36.3	21.3	1.1	0.5	51.2
ImVoxelNet	6.5	31.5	27.0	17.8	2.3	0.2	54.6
VoteNet	0.7	23.4	4.6	45.0	0.3	1.7	56.6

	bench	basket	fan	laptop	person	paper towel dispenser	oven
ImGeoNet (ours)	8.5	5.3	12.3	76.0	8.7	72.2	17.7
ImVoxelNet	8.0	13.3	7.4	62.4	9.9	81.6	6.6
VoteNet	21.2	0.1	3.0	0.4	14.5	63.3	0.8

	rack	piano	suitcase	rail	container	telephone	stand
ImGeoNet (ours)	0.0	30.6	48.2	0.4	2.1	24.4	0.0
ImVoxelNet	0.0	13.8	43.3	0.1	1.3	21.8	0.0
VoteNet	0.0	33.9	49.2	2.0	0.5	1.5	0.0

	light	laundry basket	pipe	seat	column	ladder	jacket
ImGeoNet (ours)	0.0	3.7	0.1	0.0	0.0	5.0	17.7
ImVoxelNet	0.0	4.8	0.0	0.0	0.0	14.0	14.6
VoteNet	1.7	1.1	0.0	1.4	30.9	1.1	0.2

	storage bin	coffee maker	dishwasher	machine	mat	windowsill	bulletin board
ImGeoNet (ours)	9.4	56.8	48.1	0.0	19.6	3.9	19.9
ImVoxelNet	5.5	26.5	27.9	0.0	9.8	0.3	11.8
VoteNet	0.9	12.9	4.8	0.4	0.0	7.5	2.5

	fireplace	mini fridge	water cooler	shower door	pillar	ledge	furniture
ImGeoNet (ours)	3.9	43.5	0.0	7.0	0.0	1.2	0.0
ImVoxelNet	9.6	36.1	1.7	18.3	0.0	1.0	0.0
VoteNet	0.3	22.8	2.3	20.4	5.7	0.0	0.1

	cart	decoration	closet door	vacuum cleaner	dish rack	range hood	projector screen
ImGeoNet (ours)	23.6	0.7	2.9	53.4	96.1	20.7	0.0
ImVoxelNet	18.4	0.0	0.6	16.9	84.2	14.0	0.0
VoteNet	31.2	20.0	0.3	19.1	60.8	4.0	6.0

	divider	bathroom counter	laundry hamper	bathroom stall door	ceiling light	trash bin	bathroom cabinet
ImGeoNet (ours)	19.5	0.0	8.9	1.2	0.0	18.9	35.8
ImVoxelNet	10.0	0.8	7.9	0.0	0.0	33.2	37.8
VoteNet	33.9	1.1	15.9	64.1	3.3	55.0	65.9

	potted plant	mattress	**mAP**
ImGeoNet (ours)	0.0	0.0	**17.3**
ImVoxelNet	10.0	0.0	14.0
VoteNet	100.0	60.4	16.0

Table 6. 3D object detection scores per class on ScanNet200's **tail** group (IoU threshold = 0.25).

	paper	plate	soap dispenser	bucket	clock	guitar	toilet paper holder
ImGeoNet (ours)	3.0	0.0	15.8	9.4	0.3	74.6	0.0
ImVoxelNet	0.5	0.0	20.3	8.3	0.7	98.3	0.0
VoteNet	0.0	0.0	0.1	2.4	0.1	0.0	0.0

	speaker	cup	paper towel roll	bar	toaster	ironing board	soap dish
ImGeoNet (ours)	0.0	0.2	26.1	1.7	0.0	1.7	1.2
ImVoxelNet	0.0	0.2	23.0	1.8	0.0	0.0	0.0
VoteNet	0.0	0.0	0.3	0.2	0.0	0.1	0.0

	toilet paper dispenser	fire extinguisher	ball	hat	shower curtain rod	paper cutter	tray
ImGeoNet (ours)	47.3	0.0	25.1	0.0	0.0	59.9	1.0
ImVoxelNet	20.5	0.0	54.7	0.0	0.0	51.7	0.9
VoteNet	11.3	0.0	2.6	0.0	0.0	0.0	0.0

	toaster oven	mouse	toilet seat cover dispenser	scale	tissue box	light switch	crate
ImGeoNet (ours)	1.0	0.0	6.4	90.7	7.7	0.0	5.0
ImVoxelNet	17.3	0.0	0.0	60.6	12.2	0.0	10.0
VoteNet	24.4	0.0	6.2	0.0	0.0	0.0	0.3

	power outlet	sign	projector	plunger	stuffed animal	headphones	broom
ImGeoNet (ours)	0.0	12.4	0.0	0.0	0.0	0.0	0.0
ImVoxelNet	0.0	8.4	0.0	0.0	0.0	0.0	0.0
VoteNet	0.0	0.0	0.0	0.0	0.0	0.0	0.0

	dustpan	hair dryer	water bottle	handicap bar	vent	shower floor	water pitcher
ImGeoNet (ours)	0.0	0.0	0.0	0.0	0.0	50.0	0.0
ImVoxelNet	0.0	0.0	0.0	0.0	0.0	0.0	0.0
VoteNet	0.0	0.0	0.0	0.0	0.0	100.0	0.0

	bowl	paper bag	laundry detergent	dumbbell	tube	closet rod	coffee kettle
ImGeoNet (ours)	8.6	0.2	0.0	50.0	0.1	0.0	0.0
ImVoxelNet	2.9	0.0	0.0	30.0	0.1	0.0	0.0
VoteNet	0.0	0.7	0.0	0.0	0.0	0.0	0.0

	shower head	keyboard piano	case of water bottles	coat rack	folded chair	fire alarm	power strip
ImGeoNet (ours)	0.0	60.0	3.0	0.0	0.0	0.0	0.0
ImVoxelNet	0.0	13.2	0.0	0.0	0.0	0.0	0.0
VoteNet	0.0	0.3	4.8	11.1	1.0	0.0	0.0

	calendar	poster	**mAP**
ImGeoNet (ours)	0.0	0.0	**9.7**
ImVoxelNet	6.7	2.9	7.7
VoteNet	0.0	0.0	2.9