Robust Category-Level 6D Pose Estimation with Feature-level Render and Compare (Supplementary Materials)

Wufei Ma[1], Angtian Wang[1], Alan Yuille[1], and Adam Kortylewski[1,2,3]

[1] Johns Hopkins University, Baltimore MD 21218, USA
{wma27,angtianwang,ayuille1,akortyl1}@jhu.edu
[2] Max Planck Institute for Informatics, Saarbrücken, Germany
[3] University of Freiburg, Germany

1 Implementation of Generative 6D Proposals

With the scale-invariant contrastive features, the pose optimization has a clear global minimum near the ground truth location. The loss landscapes are smooth, with decent gradients around the global minimum. This nice property allows us to search for generative 6D proposals from a sparse sampling over six dimensions.

Our generative 6D proposals are predicted in two steps. In the first step we locate the object principal points based on shape-agnostic reconstruction losses. We relax the pose optimization problem by ignoring the interior structure of the 3D model $\mathcal{M}$ and predict the object principal points by maximizing

$$\max_{m} \sum_{1\leq i\leq h, 1\leq j\leq w} \max_{1\leq k\leq N} \|\mathbf{F}_{i,j} - C_k\|^2 \tag{1}$$

We solve this optimization problem efficiently by computing and caching the keypoint correlation scores $\|\mathbf{F}_{i,j} - C_k\|^2$ for $1 \leq i \leq h$, $1 \leq j \leq w$, and $1 \leq k \leq N$. In the second step, we predict an initial 6D pose with sparse pose sampling centered at the coarse 2D locations predicted in the first step. Instead of rendering a feature map for each sampled pose, we compute vertex-level reconstruction losses based on the 2D coordinates of the vertices in the 3D model $\mathcal{M}$. Since the structure of the 3D model $\mathcal{M}$ and the sampled poses are invariant across all testing samples, the 2D coordinates and visibility of the vertices can be precomputed and cached. This allows us to predict initial 6D poses that are robust to partial occlusion and truncation and easy to optimize with a small complexity overhead.

2 Loss Landscapes of Different 6D Proposals

Previous methods for 3D object detection or 6D pose optimization are built on top of a 2D region proposal network or refine predictions from a separate pose estimation network. However, the first-stage networks are unreliable for objects with out-of-distribution textures or shapes or even missing objects if an

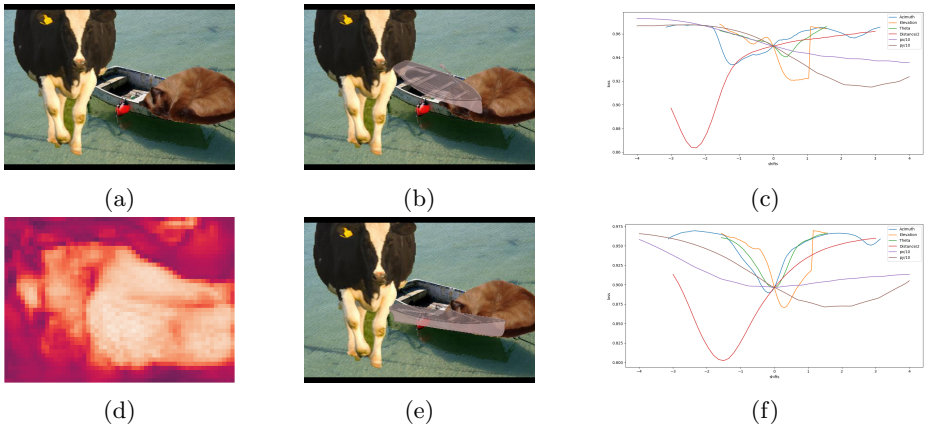

Fig. 1: Our object locating algorithm efficiently locates objects in the image and is robust to partial occlusion and truncation. Moreover, the generative 6D proposals are easier to optimize than using the Faster R-CNN prediction as the initial 6D pose. (b) and (c) shows the Faster R-CNN predicted pose and the loss landscapes over six dimensions centered at the Faster R-CNN prediction. (d) shows the reconstruction loss $\max_r \|C_r - f_i\|^2$ for each pixel. (e) and (f) shows the generative 6D proposals and the loss landscapes around the proposed 6D poses. In this example with partial occlusion, the initial pose predicted by our object locating algorithm is located at regions with good gradients and hence easy to optimize.

object is partially occluded or truncated. This essentially limits the performance of the hybrid model. Therefore, we propose to search generative 6D proposals from the predicted feature maps using the learned 3D model $\mathfrak{N}$. The generative 6D proposals are robust to partial occlusion and truncation and are easy to optimize. We compare the loss landscapes of the Faster R-CNN proposals and the generative 6D proposals in Figure 1. The loss landscapes are centered at the proposed 6D poses and show the loss landscapes around the 6D proposals. As we can see, our proposed 6D proposals are located in regions with good gradients pointing to the global minimum, which is easy to optimize.

3 Quantitative Results

3.1 Quantitative Results on the PASCAL3D+ dataset and the Occluded PASCAL3D+ dataset

Table 1: Quantitative results of 6D pose estimation on the first 6 categories of PASCAL3D+ dataset.

Subset	Level	Category	Method	Pose Acc ($\frac{\pi}{6}$) ↑	Pose Acc ($\frac{\pi}{18}$) ↑	Median Pose Error ↓	Median ADD ↓
ImageNet	0	Aeroplane	FRCNN+Cls	69.66	22.60	0.31	0.57
ImageNet	0	Aeroplane	FRCNN+NeMo	58.20	13.83	0.44	1.46
ImageNet	0	Aeroplane	RTM3DExt	49.02	18.68	0.54	1.46
ImageNet	0	Aeroplane	Ours	**87.10**	**46.13**	**0.19**	**0.58**
ImageNet	0	Bicycle	FRCNN+Cls	66.51	15.35	0.36	0.79
ImageNet	0	Bicycle	FRCNN+NeMo	50.08	10.08	0.52	2.82
ImageNet	0	Bicycle	RTM3DExt	61.86	11.78	0.43	1.61
ImageNet	0	Bicycle	Ours	**68.06**	**22.95**	**0.33**	**0.49**
ImageNet	0	Boat	FRCNN+Cls	57.79	16.62	0.40	1.22
ImageNet	0	Boat	FRCNN+NeMo	44.85	7.08	0.60	2.34
ImageNet	0	Boat	RTM3DExt	29.75	10.76	1.13	2.91
ImageNet	0	Boat	Ours	**60.43**	**18.22**	**0.40**	**1.03**
ImageNet	0	Bottle	FRCNN+Cls	**84.34**	**42.70**	**0.20**	1.22
ImageNet	0	Bottle	FRCNN+NeMo	62.38	20.48	0.36	2.32
ImageNet	0	Bottle	RTM3DExt	80.59	43.91	0.21	**0.74**
ImageNet	0	Bottle	Ours	76.57	8.97	0.34	0.77
ImageNet	0	Bus	FRCNN+Cls	**93.61**	**70.68**	0.14	0.49
ImageNet	0	Bus	FRCNN+NeMo	88.91	71.05	0.11	1.32
ImageNet	0	Bus	RTM3DExt	89.66	79.14	**0.08**	0.43
ImageNet	0	Bus	Ours	82.33	68.05	0.09	**0.38**
ImageNet	0	Car	FRCNN+Cls	91.74	57.30	0.16	0.48
ImageNet	0	Car	FRCNN+NeMo	77.65	48.19	0.19	1.26
ImageNet	0	Car	RTM3DExt	93.44	65.93	0.13	0.48
ImageNet	0	Car	Ours	**96.13**	**80.68**	**0.08**	**0.30**

Table 2: Quantitative results of 6D pose estimation on the second 6 categories of PASCAL3D+ dataset.

Subset	Level	Category	Method	Pose Acc ($\frac{\pi}{6}$) ↑	Pose Acc ($\frac{\pi}{18}$) ↑	Median Pose Error ↓	Median ADD ↓
ImageNet	0	Chair	FRCNN+Cls	80.67	27.42	0.25	1.33
ImageNet	0	Chair	FRCNN+NeMo	48.92	5.92	0.53	2.80
ImageNet	0	Chair	RTM3DExt	83.23	30.97	0.26	0.93
ImageNet	0	Chair	Ours	**89.55**	**51.08**	**0.17**	**0.59**
ImageNet	0	Diningtable	FRCNN+Cls	65.39	27.84	0.29	1.07
ImageNet	0	Diningtable	FRCNN+NeMo	59.15	17.87	0.40	3.02
ImageNet	0	Diningtable	RTM3DExt	72.68	28.97	0.28	1.85
ImageNet	0	Diningtable	Ours	**82.83**	**59.32**	**0.13**	**0.76**
ImageNet	0	Motorbike	FRCNN+Cls	62.08	10.74	0.39	1.01
ImageNet	0	Motorbike	FRCNN+NeMo	51.51	11.91	0.51	2.43
ImageNet	0	Motorbike	RTM3DExt	62.75	13.42	0.41	2.10
ImageNet	0	Motorbike	Ours	**71.98**	**21.14**	**0.31**	**0.58**
ImageNet	0	Sofa	FRCNN+Cls	**91.51**	37.82	0.20	0.74
ImageNet	0	Sofa	FRCNN+NeMo	82.53	37.82	0.23	1.57
ImageNet	0	Sofa	RTM3DExt	91.99	**48.08**	0.18	0.56
ImageNet	0	Sofa	Ours	90.06	47.60	**0.18**	**0.45**
ImageNet	0	Train	FRCNN+Cls	**94.27**	**61.61**	**0.15**	**0.70**
ImageNet	0	Train	FRCNN+NeMo	87.77	48.45	0.18	1.03
ImageNet	0	Train	RTM3DExt	90.09	54.95	0.16	1.25
ImageNet	0	Train	Ours	58.36	30.50	0.36	1.24
ImageNet	0	Tvmonitor	FRCNN+Cls	78.14	22.03	0.30	1.00
ImageNet	0	Tvmonitor	FRCNN+NeMo	66.88	17.20	0.37	2.60
ImageNet	0	Tvmonitor	RTM3DExt	**82.15**	22.99	0.28	0.66
ImageNet	0	Tvmonitor	Ours	78.62	**30.06**	**0.27**	**0.62**
ImageNet	0	Mean	FRCNN+Cls	78.90	37.35	0.22	0.74
ImageNet	0	Mean	FRCNN+NeMo	66.06	28.44	0.33	1.84
ImageNet	0	Mean	RTM3DExt	74.94	39.56	0.23	0.92
ImageNet	0	Mean	Ours	**81.45**	**47.68**	**0.19**	**0.53**

Table 3: Quantitative results of 6D pose estimation on the first 6 categories of Occluded PASCAL3D+ dataset (occlusion level 1).

Subset	Level	Category	Method	Pose Acc ($\frac{\pi}{6}$) ↑	Pose Acc ($\frac{\pi}{18}$) ↑	Median Pose Error ↓	Median ADD ↓
ImageNet	1	Aeroplane	FRCNN+Cls	50.89	12.83	0.51	0.90
ImageNet	1	Aeroplane	FRCNN+NeMo	37.33	7.89	0.74	1.61
ImageNet	1	Aeroplane	RTM3DExt	28.71	6.31	2.64	4.01
ImageNet	1	Aeroplane	Ours	**61.62**	**18.09**	**0.39**	**0.90**
ImageNet	1	Bicycle	FRCNN+Cls	39.75	6.47	0.76	1.10
ImageNet	1	Bicycle	FRCNN+NeMo	24.29	5.21	1.37	2.53
ImageNet	1	Bicycle	RTM3DExt	20.50	3.63	3.14	8.03
ImageNet	1	Bicycle	Ours	**53.00**	**13.88**	**0.49**	**0.72**
ImageNet	1	Boat	FRCNN+Cls	34.20	9.01	1.08	1.66
ImageNet	1	Boat	FRCNN+NeMo	25.95	4.56	1.39	2.43
ImageNet	1	Boat	RTM3DExt	10.21	2.93	3.11	4.87
ImageNet	1	Boat	Ours	**39.52**	**10.10**	**0.72**	**1.51**
ImageNet	1	Bottle	FRCNN+Cls	**67.49**	30.60	**0.28**	1.74
ImageNet	1	Bottle	FRCNN+NeMo	45.77	14.62	0.74	2.72
ImageNet	1	Bottle	RTM3DExt	65.98	**32.10**	0.29	1.60
ImageNet	1	Bottle	Ours	62.98	7.92	0.42	**0.95**
ImageNet	1	Bus	FRCNN+Cls	**86.34**	**54.46**	**0.17**	0.82
ImageNet	1	Bus	FRCNN+NeMo	78.37	41.18	0.22	1.36
ImageNet	1	Bus	RTM3DExt	55.03	34.16	0.38	3.42
ImageNet	1	Bus	Ours	74.76	49.53	0.18	**0.58**
ImageNet	1	Car	FRCNN+Cls	78.65	40.59	0.20	0.69
ImageNet	1	Car	FRCNN+NeMo	66.93	33.45	0.28	1.33
ImageNet	1	Car	RTM3DExt	56.05	28.04	0.38	2.20
ImageNet	1	Car	Ours	**84.69**	**60.20**	**0.14**	**0.41**

Table 4: Quantitative results of 6D pose estimation on the second 6 categories of Occluded PASCAL3D+ dataset (occlusion level 1).

Subset	Level	Category	Method	Pose Acc ($\frac{\pi}{6}$) ↑	Pose Acc ($\frac{\pi}{18}$) ↑	Median Pose Error ↓	Median ADD ↓
ImageNet	1	Chair	FRCNN+Cls	54.03	16.73	0.47	1.97
ImageNet	1	Chair	FRCNN+NeMo	24.19	2.62	1.01	3.30
ImageNet	1	Chair	RTM3DExt	38.10	8.67	0.76	3.29
ImageNet	1	Chair	Ours	**66.33**	**21.77**	**0.36**	**0.93**
ImageNet	1	Diningtable	FRCNN+Cls	47.60	21.05	0.60	1.44
ImageNet	1	Diningtable	FRCNN+NeMo	34.24	6.64	1.02	2.80
ImageNet	1	Diningtable	RTM3DExt	35.37	10.39	1.64	4.54
ImageNet	1	Diningtable	Ours	**65.68**	**32.58**	**0.28**	**1.07**
ImageNet	1	Motorbike	FRCNN+Cls	36.51	5.36	0.84	1.20
ImageNet	1	Motorbike	FRCNN+NeMo	26.64	2.42	1.34	2.47
ImageNet	1	Motorbike	RTM3DExt	15.05	1.38	3.14	9.17
ImageNet	1	Motorbike	Ours	**53.29**	**12.11**	**0.48**	**0.90**
ImageNet	1	Sofa	FRCNN+Cls	81.18	**31.73**	**0.25**	0.99
ImageNet	1	Sofa	FRCNN+NeMo	69.19	19.37	0.33	1.61
ImageNet	1	Sofa	RTM3DExt	62.18	17.16	0.37	1.92
ImageNet	1	Sofa	Ours	**82.10**	27.49	0.28	**0.79**
ImageNet	1	Train	FRCNN+Cls	**80.09**	**42.34**	**0.20**	**1.11**
ImageNet	1	Train	FRCNN+NeMo	73.78	32.39	0.25	1.37
ImageNet	1	Train	RTM3DExt	69.98	37.60	0.25	1.88
ImageNet	1	Train	Ours	50.39	23.38	0.51	1.61
ImageNet	1	Tvmonitor	FRCNN+Cls	57.63	15.42	0.45	1.55
ImageNet	1	Tvmonitor	FRCNN+NeMo	42.69	7.63	0.60	2.69
ImageNet	1	Tvmonitor	RTM3DExt	52.60	12.01	0.50	1.75
ImageNet	1	Tvmonitor	Ours	**66.40**	**16.40**	**0.40**	**0.93**
ImageNet	1	Mean	FRCNN+Cls	61.48	26.11	0.33	1.07
ImageNet	1	Mean	FRCNN+NeMo	48.34	17.46	0.55	1.90
ImageNet	1	Mean	RTM3DExt	43.55	17.68	0.82	3.29
ImageNet	1	Mean	Ours	**66.63**	**30.84**	**0.31**	**0.77**

Table 5: Quantitative results of 6D pose estimation on the first 6 categories of Occluded PASCAL3D+ dataset (occlusion level 2).

Subset	Level	Category	Method	Pose Acc ($\frac{\pi}{6}$) ↑	Pose Acc ($\frac{\pi}{18}$) ↑	Median Pose Error ↓	Median ADD ↓
ImageNet	2	Aeroplane	FRCNN+Cls	26.68	3.74	1.07	1.26
ImageNet	2	Aeroplane	FRCNN+NeMo	22.52	3.09	1.21	1.68
ImageNet	2	Aeroplane	RTM3DExt	10.67	1.49	3.14	7.47
ImageNet	2	Aeroplane	Ours	**35.86**	**5.98**	**0.83**	**1.35**
ImageNet	2	Bicycle	FRCNN+Cls	18.58	2.26	1.69	1.67
ImageNet	2	Bicycle	FRCNN+NeMo	15.35	1.29	2.28	2.52
ImageNet	2	Bicycle	RTM3DExt	6.14	0.65	3.14	18.65
ImageNet	2	Bicycle	Ours	**36.03**	**8.40**	**0.73**	**1.04**
ImageNet	2	Boat	FRCNN+Cls	20.00	**3.87**	1.98	2.16
ImageNet	2	Boat	FRCNN+NeMo	14.81	1.44	2.17	2.64
ImageNet	2	Boat	RTM3DExt	4.75	1.66	3.14	12.19
ImageNet	2	Boat	Ours	**27.18**	3.76	**1.40**	**2.16**
ImageNet	2	Bottle	FRCNN+Cls	**39.31**	**13.38**	2.80	3.09
ImageNet	2	Bottle	FRCNN+NeMo	24.41	5.66	3.03	3.87
ImageNet	2	Bottle	RTM3DExt	33.24	15.31	3.14	7.90
ImageNet	2	Bottle	Ours	38.21	4.14	**0.65**	**1.88**
ImageNet	2	Bus	FRCNN+Cls	**70.48**	**33.52**	**0.27**	0.88
ImageNet	2	Bus	FRCNN+NeMo	60.95	20.76	0.38	1.27
ImageNet	2	Bus	RTM3DExt	20.38	10.29	3.14	32.98
ImageNet	2	Bus	Ours	55.43	25.14	0.41	**0.86**
ImageNet	2	Car	FRCNN+Cls	58.31	23.47	0.36	0.88
ImageNet	2	Car	FRCNN+NeMo	51.30	19.33	0.50	1.31
ImageNet	2	Car	RTM3DExt	25.46	9.49	3.14	8.43
ImageNet	2	Car	Ours	**70.02**	**35.76**	**0.26**	**0.68**

Table 6: Quantitative results of 6D pose estimation on the second 6 categories of Occluded PASCAL3D+ dataset (occlusion level 2).

Subset	Level	Category	Method	Pose Acc ($\frac{\pi}{6}$) ↑	Pose Acc ($\frac{\pi}{18}$) ↑	Median Pose Error ↓	Median ADD ↓
ImageNet	2	Chair	FRCNN+Cls	31.06	8.28	1.08	2.60
ImageNet	2	Chair	FRCNN+NeMo	13.66	1.66	2.45	3.79
ImageNet	2	Chair	RTM3DExt	22.36	3.93	3.14	7.38
ImageNet	2	Chair	Ours	**42.03**	**9.52**	**0.64**	**1.50**
ImageNet	2	Diningtable	FRCNN+Cls	31.98	**14.63**	0.99	1.92
ImageNet	2	Diningtable	FRCNN+NeMo	25.38	5.06	1.32	2.76
ImageNet	2	Diningtable	RTM3DExt	15.90	4.34	3.14	7.85
ImageNet	2	Diningtable	Ours	**43.45**	13.91	**0.64**	**1.57**
ImageNet	2	Motorbike	FRCNN+Cls	18.60	1.58	1.72	1.60
ImageNet	2	Motorbike	FRCNN+NeMo	13.16	1.40	2.06	2.48
ImageNet	2	Motorbike	RTM3DExt	3.51	0.18	3.14	30.67
ImageNet	2	Motorbike	Ours	**34.56**	**5.44**	**0.77**	**1.35**
ImageNet	2	Sofa	FRCNN+Cls	**64.68**	**23.22**	**0.34**	1.35
ImageNet	2	Sofa	FRCNN+NeMo	55.85	11.52	0.47	1.63
ImageNet	2	Sofa	RTM3DExt	38.39	8.25	0.98	4.12
ImageNet	2	Sofa	Ours	56.05	11.90	0.45	**1.20**
ImageNet	2	Train	FRCNN+Cls	**60.03**	**23.98**	**0.36**	**1.86**
ImageNet	2	Train	FRCNN+NeMo	56.12	19.09	0.43	1.93
ImageNet	2	Train	RTM3DExt	46.82	23.00	0.68	4.59
ImageNet	2	Train	Ours	39.15	12.56	1.75	2.39
ImageNet	2	Tvmonitor	FRCNN+Cls	41.03	**9.47**	0.71	2.12
ImageNet	2	Tvmonitor	FRCNN+NeMo	28.74	5.15	1.19	2.84
ImageNet	2	Tvmonitor	RTM3DExt	31.40	6.98	3.14	5.03
ImageNet	2	Tvmonitor	Ours	**47.34**	8.80	**0.55**	**1.47**
ImageNet	2	Mean	FRCNN+Cls	41.94	14.74	0.75	1.47
ImageNet	2	Mean	FRCNN+NeMo	34.32	9.64	1.05	2.03
ImageNet	2	Mean	RTM3DExt	21.27	7.24	3.14	9.00
ImageNet	2	Mean	Ours	**47.95**	**16.25**	**0.56**	**1.22**

Table 7: Quantitative results of 6D pose estimation on the first 6 categories of Occluded PASCAL3D+ dataset (occlusion level 3).

Subset	Level	Category	Method	Pose Acc ($\frac{\pi}{6}$) ↑	Pose Acc ($\frac{\pi}{18}$) ↑	Median Pose Error ↓	Median ADD ↓
ImageNet	3	Aeroplane	FRCNN+Cls	10.41	**2.33**	1.68	**1.41**
ImageNet	3	Aeroplane	FRCNN+NeMo	8.64	0.78	1.82	1.72
ImageNet	3	Aeroplane	RTM3DExt	3.32	0.11	3.14	12.39
ImageNet	3	Aeroplane	Ours	**14.40**	1.66	**1.46**	1.61
ImageNet	3	Bicycle	FRCNN+Cls	7.49	1.33	2.06	2.20
ImageNet	3	Bicycle	FRCNN+NeMo	6.16	0.67	2.64	2.68
ImageNet	3	Bicycle	RTM3DExt	0.83	0.00	3.14	83.21
ImageNet	3	Bicycle	Ours	**19.80**	**2.83**	**1.28**	**1.63**
ImageNet	3	Boat	FRCNN+Cls	11.27	**3.04**	2.53	**2.70**
ImageNet	3	Boat	FRCNN+NeMo	7.78	0.79	2.83	2.96
ImageNet	3	Boat	RTM3DExt	2.03	0.34	3.14	49.79
ImageNet	3	Boat	Ours	**12.40**	1.24	**2.10**	3.13
ImageNet	3	Bottle	FRCNN+Cls	14.31	4.03	3.10	3.86
ImageNet	3	Bottle	FRCNN+NeMo	9.03	1.81	3.10	4.33
ImageNet	3	Bottle	RTM3DExt	20.14	**8.89**	3.14	15.41
ImageNet	3	Bottle	Ours	**20.97**	0.83	**1.06**	**3.33**
ImageNet	3	Bus	FRCNN+Cls	46.45	12.67	0.57	1.34
ImageNet	3	Bus	FRCNN+NeMo	37.62	8.64	0.79	1.54
ImageNet	3	Bus	RTM3DExt	1.92	0.58	3.14	38.54
ImageNet	3	Bus	Ours	33.21	7.68	1.04	1.82
ImageNet	3	Car	FRCNN+Cls	30.86	6.84	1.07	**1.14**
ImageNet	3	Car	FRCNN+NeMo	26.89	5.52	1.28	1.52
ImageNet	3	Car	RTM3DExt	7.42	2.18	3.14	56.89
ImageNet	3	Car	Ours	**42.21**	**11.08**	**0.74**	1.20

Table 8: Quantitative results of 6D pose estimation on the second 6 categories of Occluded PASCAL3D+ dataset (occlusion level 3).

Subset	Level	Category	Method	Pose Acc ($\frac{\pi}{6}$) ↑	Pose Acc ($\frac{\pi}{18}$) ↑	Median Pose Error ↓	Median ADD ↓
ImageNet	3	Chair	FRCNN+Cls	15.04	3.60	2.61	3.29
ImageNet	3	Chair	FRCNN+NeMo	6.36	0.64	2.91	4.32
ImageNet	3	Chair	RTM3DExt	13.98	2.12	3.14	9.18
ImageNet	3	Chair	Ours	**24.58**	**3.60**	**1.02**	**2.36**
ImageNet	3	Diningtable	FRCNN+Cls	19.91	**7.44**	1.69	**2.15**
ImageNet	3	Diningtable	FRCNN+NeMo	14.70	2.14	2.60	3.02
ImageNet	3	Diningtable	RTM3DExt	11.53	3.26	3.14	8.83
ImageNet	3	Diningtable	Ours	**24.37**	4.37	**1.06**	2.17
ImageNet	3	Motorbike	FRCNN+Cls	6.13	0.36	2.39	2.47
ImageNet	3	Motorbike	FRCNN+NeMo	6.31	0.72	2.77	3.06
ImageNet	3	Motorbike	RTM3DExt	0.90	0.18	3.14	67.59
ImageNet	3	Motorbike	Ours	**14.95**	**1.26**	**1.46**	**2.41**
ImageNet	3	Sofa	FRCNN+Cls	**38.07**	**8.48**	0.95	1.75
ImageNet	3	Sofa	FRCNN+NeMo	30.37	5.13	1.11	1.79
ImageNet	3	Sofa	RTM3DExt	27.42	6.31	3.14	6.09
ImageNet	3	Sofa	Ours	34.91	4.14	**0.62**	**1.48**
ImageNet	3	Train	FRCNN+Cls	**37.74**	9.43	2.55	**2.35**
ImageNet	3	Train	FRCNN+NeMo	34.31	7.72	2.56	2.40
ImageNet	3	Train	RTM3DExt	28.13	**13.72**	3.14	9.78
ImageNet	3	Train	Ours	23.50	6.69	**2.46**	3.56
ImageNet	3	Tvmonitor	FRCNN+Cls	21.84	**5.63**	3.00	3.00
ImageNet	3	Tvmonitor	FRCNN+NeMo	17.06	1.71	2.98	3.53
ImageNet	3	Tvmonitor	RTM3DExt	20.14	4.27	3.14	8.43
ImageNet	3	Tvmonitor	Ours	**33.11**	4.10	**0.77**	**2.35**
ImageNet	3	Mean	FRCNN+Cls	22.42	**5.58**	2.01	1.95
ImageNet	3	Mean	FRCNN+NeMo	18.17	5.30	2.40	2.35
ImageNet	3	Mean	RTM3DExt	10.17	3.11	3.14	19.92
ImageNet	3	Mean	Ours	**27.43**	5.30	**1.07**	**1.94**