A Mixed Classification-Regression Framework for 3D Pose Estimation from 2D Images

Siddharth Mahendran[1]
siddharthm@jhu.edu
Haider Ali[2]
hali@jhu.edu
René Vidal[1]
rvidal@cis.jhu.edu

[1] Center for Imaging Science, Mathematical Institute for Data Science, Johns Hopkins University, Baltimore, MD, USA

[2] Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA

We now present more results §1 and implementation details §2 that we were unable to include in the main paper due to space constraints.

1 Results

In Tables 2 and 3, we present the results of our eight models and three baselines compared to current state-of-the-art under the two metric of $MedErr$ and $Acc_{\frac{\pi}{6}}$ respectively. As we mentioned in the paper, we run all experiments three times and report the mean and standard deviation (in brackets) across these three trials. We also show figures of images where we obtain the least pose estimation error and the most pose estimation error for every object category using one run of model $\mathcal{M}_G+$. As can be seen from Figs. [1-12], we make the most error under three conditions: (i) when the objects are really blurry (very small in pixel size in the original image), (ii) the shape of the object is uncommon (possibly very few examples seen during training) and (iii) the pose of a test image is very different from common poses observed during training. The first condition is best observed in the bad cases for categories aeroplane and car where almost all the images shown are very blurry. The second condition is best observed in categories boat and chair where the bad cases contain uncommon boats and chairs. The third condition is best observed in categories bottle and tvmonitor where the bad images are in very different poses compared to the best images.

We also present the performance of our models $\mathcal{M}_G$ and $\mathcal{M}_G+$ across different object categories of the Pascal3D+ dataset during ablation experiments in Tables 4-11. These are detailed tables for the results shown in Tables 3 and 4 of the main paper and an overview of the experiments is shown below.

2 Implementation details

We use the ResNet-50 upto layer4 (2048-dim feature output) as our feature network. The pose networks are of the form Input-FC-BN-ReLU-FC-BN-ReLU-FC-Output where FC is a

Expt.	$\mathcal{M}_G$		$\mathcal{M}_G$+	
	MedErr	$Acc_{\frac{\pi}{6}}$	*MedErr*	$Acc_{\frac{\pi}{6}}$
K-Means dictionary size K	Table 4	Table 5	Table 6	Table 7
Weighting parameter α	Table 8	Table 9	Table 10	Table 11

Table 1: Overview of Tables of results showing ablation analysis

Method	aero	bike	boat	bottle	bus	car	chair	dtable	mbike	sofa	train	tv	Mean
[4]	13.8	17.7	21.3	12.9	5.8	9.1	14.8	15.2	14.7	13.7	8.7	15.4	13.59
[6]	15.4	14.8	25.6	9.3	3.6	6.0	9.7	10.8	16.7	9.5	6.1	12.6	11.68
[2]	13.6	12.5	22.8	8.3	3.1	5.8	11.9	12.5	12.3	12.8	6.3	11.9	11.15
[1]	10.0	15.6	19.1	8.6	3.3	5.1	13.7	11.8	12.2	13.5	6.7	11.0	10.88
$\mathcal{R}_E$	14.5	17.7	39.3	7.4	4.0	7.8	15.2	26.6	17.5	10.5	11.5	14.1	15.50
	(1.2)	(0.4)	(3.4)	(0.3)	(0.4)	(0.5)	(0.4)	(2.0)	(0.2)	(0.7)	(2.0)	(0.6)	(0.34)
$\mathcal{R}_G$	11.8	15.9	27.2	7.2	2.9	5.2	11.6	15.0	14.3	10.8	5.4	12.4	11.63
	(0.6)	(0.5)	(2.0)	(0.6)	(0.1)	(0.1)	(0.4)	(3.9)	(0.6)	(0.1)	(0.3)	(0.7)	(0.51)
$\mathcal{C}$	11.7	15.3	21.5	9.3	4.1	7.4	11.2	17.8	17.0	11.0	7.0	13.1	12.20
	(0.7)	(0.7)	(2.1)	(0.3)	(0.1)	(0.3)	(0.2)	(3.9)	(0.7)	(0.4)	(0.3)	(0.3)	(0.44)
$\mathcal{M}_S$	11.0	15.5	21.0	8.8	3.8	7.0	10.8	21.0	16.6	10.7	6.5	13.1	12.14
	(0.9)	(0.8)	(0.8)	(0.5)	(0.2)	(0.0)	(0.2)	(5.8)	(0.2)	(1.0)	(0.3)	(0.7)	(0.37)
$\mathcal{M}_G$	12.1	16.0	19.9	8.9	3.4	6.5	10.8	15.2	16.4	9.6	5.9	13.0	11.48
	(0.6)	(0.5)	(1.1)	(0.2)	(0.2)	(0.3)	(0.0)	(6.0)	(0.5)	(0.6)	(0.9)	(0.7)	(0.23)
$\mathcal{M}_{LE}$	12.8	15.2	23.4	9.0	4.0	7.4	11.1	16.8	16.1	10.7	6.6	12.3	12.11
	(1.0)	(0.8)	(3.3)	(0.2)	(0.1)	(0.2)	(0.1)	(3.5)	(0.7)	(1.4)	(0.1)	(0.8)	(0.32)
$\mathcal{M}_P$	11.4	16.3	25.6	7.0	2.6	5.1	11.3	16.0	13.6	10.2	5.5	12.0	11.38
	(0.4)	(0.6)	(0.6)	(0.4)	(0.1)	(0.1)	(0.4)	(2.1)	(1.0)	(0.7)	(0.3)	(0.3)	(0.16)
$\mathcal{M}_S$+	12.2	15.7	24.4	9.9	3.6	6.5	12.0	14.8	14.4	11.9	6.4	11.6	11.95
	(0.2)	(1.0)	(0.6)	(0.2)	(0.2)	(0.2)	(0.4)	(0.8)	(0.8)	(0.7)	(0.2)	(0.3)	(0.26)
$\mathcal{M}_G$+	8.5	14.8	20.5	7.0	3.1	5.1	9.3	11.3	14.2	10.2	5.6	11.7	10.10
	(0.0)	(1.2)	(1.4)	(0.4)	(0.2)	(0.1)	(0.1)	(2.0)	(0.6)	(0.3)	(0.1)	(0.2)	(0.38)
$\mathcal{M}_{LE}$+	12.3	16.7	24.7	7.5	3.6	6.5	11.5	15.5	15.1	11.1	7.3	12.1	11.99
	(0.4)	(0.4)	(1.9)	(0.2)	(0.1)	(0.1)	(0.5)	(0.4)	(0.5)	(0.5)	(0.7)	(0.2)	(0.15)
$\mathcal{M}_P$+	10.6	15.0	23.9	6.7	2.7	4.7	9.8	12.6	13.9	9.7	5.3	11.7	10.54
	(0.2)	(0.2)	(1.2)	(0.2)	(0.1)	(0.1)	(0.1)	(0.5)	(0.9)	(0.1)	(0.1)	(0.5)	(0.16)

Table 2: Performance of our models under the *MedErr* metric (lower is better).

fully connected layer, BN is a batch normalization layer and ReLU is the standard rectified linear unit non-linearity. The pose networks for models $\mathcal{R}_E$ and $\mathcal{R}_G$ are of size 2048-1000-500-3. The pose network of model $\mathcal{C}$ is of size 2048-1000-500-100 where 100 is the size of the K-Means dictionary we use to discretize the pose space. The bin and delta networks of models $\mathcal{M}_S$, $\mathcal{M}_G$, $\mathcal{M}_{LE}$ and $\mathcal{M}_P$ are of sizes 2048-1000-500-100 and 2048-1000-500-3 respectively. For models $\mathcal{M}_S$+, $\mathcal{M}_G$+, $\mathcal{M}_{LE}$+ and $\mathcal{M}_P$+ where we have one delta network per pose-bin per object category, our bin network is of size 2048-1000-500-16 (corresponding to 16 pose-bins) and we use a 2-layer delta network of size 2048-100-3.

For the models, $\mathcal{M}_G$ and $\mathcal{M}_G$+, we initialize the network weights with 1 epoch of training over the models $\mathcal{M}_S$ and $\mathcal{M}_S$+. All other models are initialized using pre-trained networks on the ImageNet image classification problem. The models $\mathcal{M}_S$, $\mathcal{M}_G$, $\mathcal{M}_P$, $\mathcal{M}_S$+ and $\mathcal{M}_P$+ were trained with $\alpha = 1$. For the model $\mathcal{M}_G$+, we use a value of $\alpha = 10$ and for the models $\mathcal{M}_{LE}$ and $\mathcal{M}_{LE}$+, we use $\alpha = 0.1$.

Method	aero	bike	boat	bottle	bus	car	chair	dtable	mbike	sofa	train	tv	Mean
[4]	0.81	0.77	0.59	0.93	0.98	0.89	0.80	0.62	0.88	0.82	0.80	0.80	0.8075
[3]	0.74	0.83	0.52	0.91	0.91	0.88	0.86	0.73	0.78	0.90	0.86	0.92	0.8200
[1]	0.78	0.83	0.57	0.93	0.94	0.90	0.80	0.68	0.86	0.82	0.82	0.85	0.8103
[1]	0.83	0.82	0.64	0.95	0.97	0.94	0.80	0.71	0.88	0.87	0.80	0.86	0.8392
$\mathcal{R}_E$	0.77	0.75	0.41	0.96	0.91	0.83	0.72	0.56	0.75	0.90	0.75	0.87	0.7656
	(0.01)	(0.01)	(0.03)	(0.00)	(0.01)	(0.01)	(0.02)	(0.04)	(0.01)	(0.02)	(0.03)	(0.00)	(0.0015)
$\mathcal{R}_G$	0.80	0.78	0.54	0.97	0.95	0.93	0.83	0.59	0.82	0.91	0.81	0.86	0.8166
	(0.01)	(0.00)	(0.03)	(0.00)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.0041)
$\mathcal{C}$	0.84	0.77	0.60	0.95	0.97	0.95	0.90	0.63	0.78	0.94	0.81	0.87	0.8350
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.06)	(0.02)	(0.01)	(0.01)	(0.01)	(0.0045)
$\mathcal{M}_S$	0.83	0.78	0.61	0.96	0.96	0.94	0.90	0.56	0.79	0.95	0.82	0.87	0.8303
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.04)	(0.01)	(0.02)	(0.00)	(0.02)	(0.0014)
$\mathcal{M}_G$	0.83	0.76	0.63	0.96	0.97	0.93	0.91	0.57	0.78	0.95	0.82	0.88	0.8335
	(0.02)	(0.03)	(0.01)	(0.01)	(0.02)	(0.00)	(0.00)	(0.04)	(0.02)	(0.02)	(0.01)	(0.01)	(0.0067)
$\mathcal{M}_{LE}$	0.83	0.77	0.58	0.96	0.96	0.94	0.91	0.71	0.81	0.93	0.81	0.87	0.8410
	(0.02)	(0.02)	(0.03)	(0.00)	(0.01)	(0.01)	(0.00)	(0.10)	(0.00)	(0.01)	(0.01)	(0.01)	(0.0025)
$\mathcal{M}_P$	0.80	0.77	0.56	0.97	0.97	0.93	0.82	0.57	0.81	0.92	0.82	0.88	0.8185
	(0.01)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.04)	(0.03)	(0.00)	(0.01)	(0.01)	(0.0035)
$\mathcal{M}_S+$	0.82	0.80	0.59	0.94	0.97	0.94	0.91	0.63	0.81	0.97	0.83	0.87	0.8387
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.04)	(0.00)	(0.00)	(0.01)	(0.00)	(0.0044)
$\mathcal{M}_G+$	0.87	0.81	0.64	0.96	0.97	0.95	0.92	0.67	0.85	0.97	0.82	0.88	0.8588
	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.10)	(0.01)	(0.01)	(0.01)	(0.00)	(0.0111)
$\mathcal{M}_{LE}+$	0.81	0.77	0.56	0.96	0.97	0.92	0.86	0.73	0.79	0.93	0.80	0.89	0.8329
	(0.02)	(0.02)	(0.03)	(0.01)	(0.01)	(0.00)	(0.01)	(0.06)	(0.02)	(0.02)	(0.01)	(0.01)	(0.0032)
$\mathcal{M}_P+$	0.84	0.82	0.59	0.97	0.97	0.95	0.88	0.68	0.84	0.93	0.81	0.89	0.8470
	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.04)	(0.01)	(0.01)	(0.01)	(0.01)	(0.0044)

Table 3: Performance of our models under the $Acc_{\frac{\pi}{6}}$ metric (higher is better).

Method	aero	bike	boat	bottle	bus	car	chair	dtable	mbike	sofa	train	tv	Mean
K=24	12.4	16.3	23.5	8.7	2.7	5.4	11.6	17.4	16.3	13.7	6.2	15.6	12.48
	(0.3)	(0.1)	(0.3)	(0.1)	(0.0)	(0.3)	(0.1)	(2.6)	(0.4)	(1.4)	(0.5)	(0.5)	(0.35)
K=50	12.9	16.0	21.1	8.3	3.3	6.1	11.2	22.2	17.7	11.8	5.8	13.9	12.53
	(0.9)	(0.9)	(1.3)	(0.2)	(0.2)	(0.1)	(0.4)	(5.1)	(1.1)	(0.9)	(0.2)	(0.2)	(0.45)
K=100	12.1	16.0	19.9	8.9	3.4	6.5	10.8	15.2	16.4	9.6	5.9	13.0	11.48
	(0.6)	(0.5)	(1.1)	(0.2)	(0.2)	(0.3)	(0.0)	(6.0)	(0.5)	(0.6)	(0.9)	(0.7)	(0.23)
K=200	10.6	16.4	21.6	8.1	3.2	6.0	9.9	14.6	16.0	11.1	6.3	13.4	11.44
	(0.2)	(1.4)	(0.2)	(0.4)	(0.1)	(0.2)	(0.2)	(1.2)	(0.8)	(0.6)	(0.9)	(0.4)	(0.20)

Table 4: Ablation analysis of the size of K-Means dictionary in model $\mathcal{M}_G$ under the *MedErr* metric (lower is better).

Method	aero	bike	boat	bottle	bus	car	chair	dtable	mbike	sofa	train	tv	Mean
K=24	0.84	0.82	0.58	0.95	0.97	0.94	0.89	0.57	0.77	0.91	0.81	0.86	0.8266
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.04)	(0.01)	(0.02)	(0.01)	(0.01)	(0.0047)
K=50	0.82	0.80	0.59	0.95	0.97	0.94	0.92	0.54	0.78	0.91	0.83	0.89	0.8281
	(0.01)	(0.02)	(0.03)	(0.00)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.0063)
K=100	0.83	0.76	0.63	0.96	0.97	0.93	0.91	0.57	0.78	0.95	0.82	0.88	0.8335
	(0.02)	(0.03)	(0.01)	(0.01)	(0.02)	(0.00)	(0.00)	(0.04)	(0.02)	(0.02)	(0.01)	(0.01)	(0.0067)
K=200	0.84	0.76	0.62	0.96	0.98	0.94	0.92	0.65	0.80	0.96	0.82	0.87	0.8439
	(0.01)	(0.02)	(0.00)	(0.00)	(0.00)	(0.01)	(0.02)	(0.04)	(0.02)	(0.02)	(0.00)	(0.01)	(0.0054)

Table 5: Ablation analysis of the size of K-Means dictionary in model $\mathcal{M}_G$ under the $Acc_{\frac{\pi}{6}}$ metric (higher is better).

Method	aero	bike	boat	bottle	bus	car	chair	dtable	mbike	sofa	train	tv	Mean
K=4	10.4	13.3	21.9	7.2	2.9	5.3	9.9	16.3	14.1	10.4	5.0	12.5	10.78
	(0.5)	(0.4)	(1.4)	(0.3)	(0.2)	(0.1)	(0.1)	(1.9)	(0.0)	(0.7)	(0.3)	(0.5)	(0.35)
K=8	10.5	14.8	21.5	6.8	2.7	4.9	9.7	16.1	14.9	10.2	5.6	12.3	10.85
	(0.5)	(0.5)	(1.5)	(0.4)	(0.1)	(0.1)	(0.4)	(2.9)	(0.1)	(0.3)	(0.3)	(0.5)	(0.38)
K=16	9.9	14.3	21.3	7.3	2.7	4.9	9.6	13.0	14.7	10.8	5.2	11.7	10.46
	(0.4)	(0.5)	(0.7)	(0.1)	(0.1)	(0.1)	(0.3)	(3.8)	(0.5)	(1.2)	(0.3)	(0.3)	(0.28)
K=24	9.7	15.3	23.5	7.1	2.9	5.0	10.0	13.3	14.4	11.3	5.3	13.1	10.91
	(0.2)	(0.8)	(1.0)	(0.4)	(0.1)	(0.2)	(0.5)	(2.6)	(0.3)	(0.5)	(0.5)	(0.4)	(0.22)

Table 6: Ablation analysis of the size of K-Means dictionary in model $\mathcal{M}_G+$ under the *MedErr* metric (lower is better).

Method	aero	bike	boat	bottle	bus	car	chair	dtable	mbike	sofa	train	tv	Mean
K=4	0.85	0.80	0.61	0.97	0.97	0.95	0.87	0.67	0.84	0.93	0.83	0.86	0.8453
	(0.00)	(0.00)	(0.02)	(0.00)	(0.01)	(0.00)	(0.02)	(0.08)	(0.01)	(0.02)	(0.00)	(0.01)	(0.0085)
K=8	0.83	0.79	0.60	0.97	0.96	0.95	0.91	0.62	0.81	0.95	0.83	0.89	0.8427
	(0.01)	(0.02)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.04)	(0.02)	(0.00)	(0.00)	(0.00)	(0.0017)
K=16	0.84	0.82	0.61	0.96	0.98	0.96	0.92	0.67	0.82	0.97	0.82	0.90	0.8553
	(0.01)	(0.02)	(0.02)	(0.00)	(0.00)	(0.00)	(0.01)	(0.07)	(0.01)	(0.01)	(0.01)	(0.02)	(0.0035)
K=24	0.87	0.80	0.60	0.96	0.97	0.95	0.90	0.65	0.83	0.94	0.82	0.87	0.8467
	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.04)	(0.01)	(0.01)	(0.01)	(0.01)	(0.0054)

Table 7: Ablation analysis of the size of K-Means dictionary in model $\mathcal{M}_G+$ under the $Acc_{\frac{\pi}{6}}$ metric (higher is better).

Method	aero	bike	boat	bottle	bus	car	chair	dtable	mbike	sofa	train	tv	Mean
$\alpha = 0.1$	11.8	16.1	20.8	8.4	3.3	6.4	10.6	28.5	15.0	11.2	6.0	12.3	12.53
	(0.1)	(0.8)	(0.5)	(0.2)	(0.1)	(0.1)	(0.2)	(11.7)	(1.0)	(0.7)	(0.1)	(0.7)	(1.00)
$\alpha = 1$	12.1	16.0	19.9	8.9	3.4	6.5	10.8	15.2	16.4	9.6	5.9	13.0	11.48
	(0.6)	(0.5)	(1.1)	(0.2)	(0.2)	(0.3)	(0.0)	(6.0)	(0.5)	(0.6)	(0.9)	(0.7)	(0.23)
$\alpha = 10$	12.1	14.5	22.8	8.7	3.1	6.5	10.9	15.1	16.3	10.6	6.0	13.0	11.63
	(0.8)	(0.1)	(0.5)	(0.3)	(0.0)	(0.3)	(0.4)	(0.5)	(0.8)	(0.5)	(0.5)	(0.3)	(0.24)

Table 8: Ablation analysis of the weighting parameter α in model $\mathcal{M}_G$ under the *MedErr* metric (lower is better).

Method	aero	bike	boat	bottle	bus	car	chair	dtable	mbike	sofa	train	tv	Mean
$\alpha = 0.1$	0.84	0.77	0.62	0.96	0.96	0.94	0.91	0.51	0.82	0.96	0.81	0.88	0.8306
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.04)	(0.01)	(0.02)	(0.01)	(0.01)	(0.0049)
$\alpha = 1$	0.83	0.76	0.63	0.96	0.97	0.93	0.91	0.57	0.78	0.95	0.82	0.88	0.8335
	(0.02)	(0.03)	(0.01)	(0.01)	(0.02)	(0.00)	(0.00)	(0.04)	(0.02)	(0.02)	(0.01)	(0.01)	(0.0067)
$\alpha = 10$	0.82	0.79	0.59	0.96	0.97	0.94	0.91	0.67	0.81	0.95	0.82	0.88	0.8424
	(0.02)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)	(0.02)	(0.02)	(0.00)	(0.01)	(0.0060)

Table 9: Ablation analysis of the weighting parameter α in model $\mathcal{M}_G$ under the $Acc_{\frac{\pi}{6}}$ metric (higher is better).

Method	aero	bike	boat	bottle	bus	car	chair	dtable	mbike	sofa	train	tv	Mean
$\alpha = 0.1$	10.3	16.0	24.0	7.1	3.2	5.5	10.3	11.8	15.2	10.6	6.0	12.3	11.01
	(0.6)	(2.0)	(2.1)	(0.1)	(0.5)	(0.8)	(1.3)	(0.9)	(0.9)	(1.1)	(0.9)	(0.8)	(0.92)
$\alpha = 1$	9.9	14.3	21.3	7.3	2.7	4.9	9.6	13.0	14.7	10.8	5.2	11.7	10.46
	(0.4)	(0.5)	(0.7)	(0.1)	(0.1)	(0.1)	(0.3)	(3.8)	(0.5)	(1.2)	(0.3)	(0.3)	(0.28)
$\alpha = 10$	8.5	14.8	20.5	7.0	3.1	5.1	9.3	11.3	14.2	10.2	5.6	11.7	10.10
	(0.0)	(1.2)	(1.4)	(0.4)	(0.2)	(0.1)	(0.1)	(2.0)	(0.6)	(0.3)	(0.1)	(0.2)	(0.38)

Table 10: Ablation analysis of the weighting parameter α in model $\mathcal{M}_G+$ under the *MedErr* metric (lower is better).

Method	aero	bike	boat	bottle	bus	car	chair	dtable	mbike	sofa	train	tv	Mean
$\alpha = 0.1$	0.86 (0.02)	0.81 (0.02)	0.59 (0.04)	0.96 (0.01)	0.98 (0.01)	0.95 (0.00)	0.90 (0.03)	0.65 (0.04)	0.80 (0.01)	0.96 (0.01)	0.82 (0.01)	0.89 (0.00)	0.8473 (0.0102)
$\alpha = 1$	0.84 (0.01)	0.82 (0.02)	0.61 (0.02)	0.96 (0.00)	0.98 (0.00)	0.96 (0.00)	0.92 (0.01)	0.67 (0.07)	0.82 (0.01)	0.97 (0.01)	0.82 (0.01)	0.90 (0.02)	0.8553 (0.0035)
$\alpha = 10$	0.87 (0.01)	0.81 (0.01)	0.64 (0.01)	0.96 (0.00)	0.97 (0.01)	0.95 (0.01)	0.92 (0.01)	0.67 (0.10)	0.85 (0.01)	0.97 (0.01)	0.82 (0.01)	0.88 (0.00)	0.8588 (0.0111)

Table 11: Ablation analysis of the weighting parameter α in model $\mathcal{M}_G+$ under the $Acc_{\frac{\pi}{6}}$ metric (higher is better).

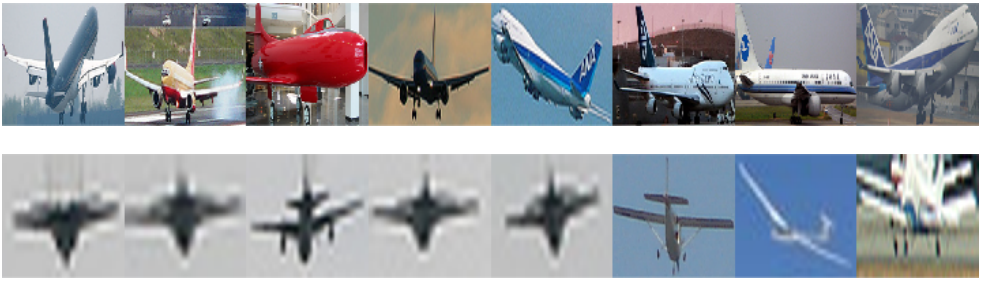

Figure 1: Best (top row) and Worst (bottom row) images for Category: Aeroplane

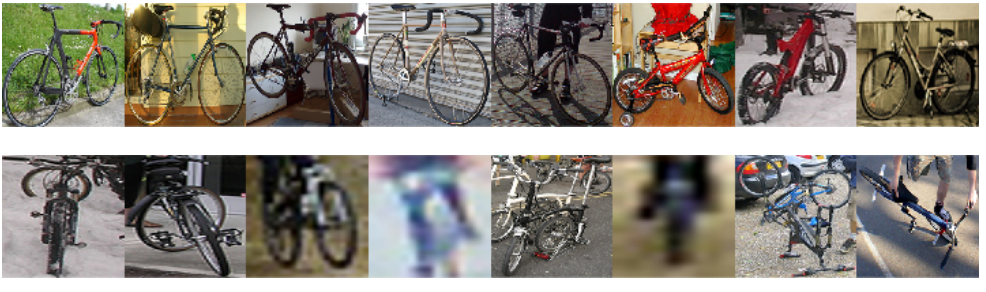

Figure 2: Best (top row) and Worst (bottom row) images for Category: Bicycle

Figure 3: Best (top row) and Worst (bottom row) images for Category: Boat

Figure 4: Best (top row) and Worst (bottom row) images for Category: Bottle

Figure 5: Best (top row) and Worst (bottom row) images for Category: Bus

Figure 6: Best (top row) and Worst (bottom row) images for Category: Car

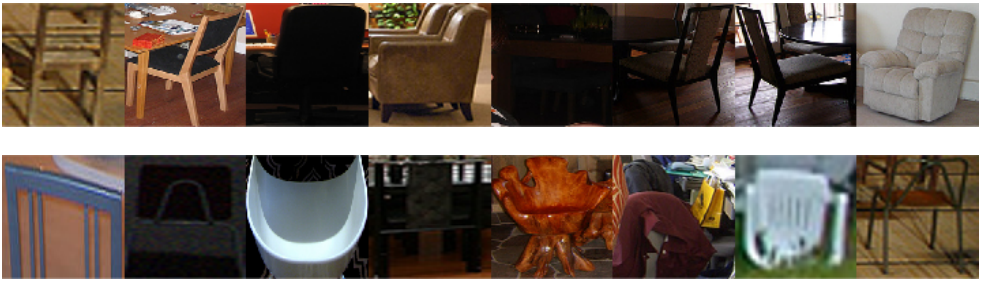

Figure 7: Best (top row) and Worst (bottom row) images for Category: Chair

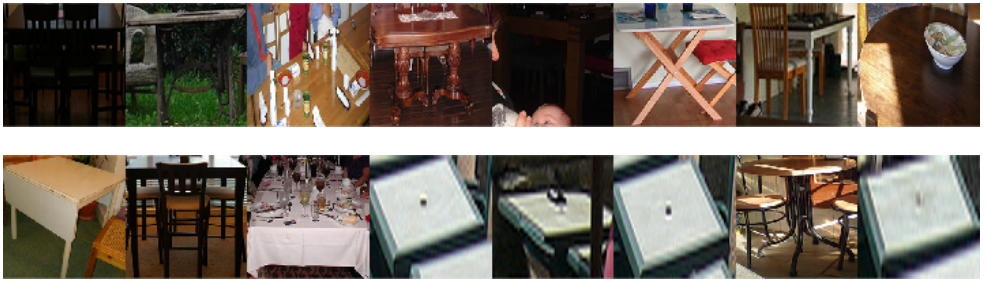

Figure 8: Best (top row) and Worst (bottom row) images for Category: Diningtable

Figure 9: Best (top row) and Worst (bottom row) images for Category: Motorbike

Figure 10: Best (top row) and Worst (bottom row) images for Category: Sofa

Figure 11: Best (top row) and Worst (bottom row) images for Category: Train

Figure 12: Best (top row) and Worst (bottom row) images for Category: Tvmonitor

References

[1] Alexander Grabner, Peter M. Roth, and Vincent Lepetit. 3d pose estimation and 3d model retrieval for objects in the wild. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[2] Arsalan Mousavian, Dragomir Anguelov, John Flynn, and Jana Kosecka. 3d bounding box estimation using deep learning and geometry. In *CVPR*, 2017.

[3] Hao Su, Charles R. Qi, Yangyan Li, and Leonidas J. Guibas. Render for cnn: Viewpoint estimation in images using cnns trained with rendered 3d model views. In *IEEE International Conference on Computer Vision*, 2015.

[4] Shubham Tulsiani and Jitendra Malik. Viewpoints and keypoints. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.