

Image2Mesh: A Learning Framework for Single Image 3D Reconstruction

Jhony K. Pontes[†], Chen Kong^{*}, Sridha Sridharan[†], Simon Lucey^{*}, Anders Eriksson[†], Clinton Fookes[†]

Queensland University of Technology[†] Carnegie Mellon University^{*}

In this document, we provide additional results to supplement our main submission. We provide an ablation study and additional qualitative results for the 3D reconstruction from a single image using both synthetic and real world data.

1 Ablation Study

An extensive ablation study was performed to show how sensitive the proposed method is for each step - Model selection (IDX), free-form deformation (FFD) and linear combination (LC). Table 1 clearly shows the importance of the FFD and LC steps to improve the performance of the 3D reconstruction. Table 2 shows the ablation study considering the index for the 3D model selected from the graph to be the ground truth (IDX_{GT}). One can notice it slightly outperforms Table 1 for most categories as expected. Table 3 shows the results when considering the IDX , the FFD and the LC to be the ground truths, in different combinations. Note that we did not perform the combinations $IDX + LC_{GT}$, $IDX + FFD_{GT} + LC_{GT}$ and $IDX + FFD + LC_{GT}$ as the IDX is not the ground truth and it might end up selecting a different subgraph to perform the linear combination.

2 3D Reconstruction from a Single Image

2.1 3D Reconstruction from Synthetic Images

Qualitative results for the 3D reconstruction from synthetic images are shown in Figure 1 and 2. Failure cases are shown in Figure 3 where in most cases an “incorrect” model was selected from the graph. For the chair and sofa examples, due to the image perspective, it is impossible to say whether there are more details in the chair or not, as shown in the ground-truth model, or say whether the sofa is of one or more places. However, our framework made fair estimations for both cases considering the image perspective. For the diningtable instance, a good table was first selected but in the deformation process part of its top was removed due to some bad linear interpolation between the 3D models.

2.2 3D Reconstruction from Real World Images

Qualitative results for the 3D reconstruction from real world images are shown in Figure 4. One can see that our proposed method performed well real data.

An interesting example is the airplane where in the image there is a jet fighter and to start the deformation process our method selected a four-engine aircraft. However, with the estimated FFD and the sparse linear combination parameters, our method deformed the selected airplane to approximate it to a fighter jet.

	IDX		IDX+FFD		IDX+LC		IDX+FFD+LC	
	$dist_{3D}$	IoU	$dist_{3D}$	IoU	$dist_{3D}$	IoU	$dist_{3D}$	IoU
car	0.027	0.391	0.023	0.410	0.007	0.622	0.006	0.664
bicycle	0.064	0.494	0.052	0.483	0.022	0.708	0.025	0.795
motorbike	0.036	0.348	0.028	0.383	0.006	0.675	0.007	0.679
aeroplane	0.066	0.388	0.057	0.329	0.023	0.507	0.023	0.551
bus	0.025	0.454	0.020	0.432	0.003	0.773	0.003	0.776
chair	0.074	0.188	0.085	0.182	0.027	0.360	0.034	0.403
diningtable	0.186	0.158	0.183	0.163	0.241	0.337	0.165	0.332
sofa	0.130	0.242	0.145	0.258	0.072	0.352	0.063	0.402
Mean	0.076	0.333	0.074	0.330	0.050	0.542	0.041	0.575

Table 1: Quantitative results for our proposed method evaluated on the synthetic test set. We show the performance of every step - Model selection (*IDX*), FFD and linear combination (*LC*). One can clearly notice the importance of the FFD and *LC* steps to improve the performance of the 3D reconstruction.

	IDX _{GT}		IDX _{GT} +FFD		IDX _{GT} +LC		IDX _{GT} +FFD+LC	
	$dist_{3D}$	IoU	$dist_{3D}$	IoU	$dist_{3D}$	IoU	$dist_{3D}$	IoU
car	0.025	0.398	0.020	0.452	0.005	0.614	0.004	0.673
bicycle	0.058	0.499	0.046	0.489	0.014	0.679	0.003	0.799
motorbike	0.036	0.348	0.028	0.382	0.007	0.666	0.004	0.702
aeroplane	0.061	0.399	0.060	0.334	0.013	0.500	0.008	0.591
bus	0.024	0.454	0.018	0.433	0.002	0.802	0.002	0.814
chair	0.072	0.192	0.076	0.182	0.019	0.409	0.018	0.390
diningtable	0.191	0.160	0.178	0.172	0.321	0.364	0.223	0.305
sofa	0.092	0.258	0.103	0.254	0.061	0.375	0.073	0.381
Mean	0.070	0.339	0.066	0.337	0.055	0.551	0.042	0.582

Table 2: Quantitative results for our method evaluated on the synthetic test set. We show the performance of every step - Model selection (*IDX*), FFD and linear combination (*LC*). We consider (*IDX*) to be the ground truth (*IDX_{GT}*).

	IDX _{GT} +FFD _{GT}		IDX+FFD _{GT} +LC		IDX _{GT} +FFD+LC _{GT}		IDX+FFD _{GT}		IDX _{GT} +FFD _{GT} +LC	
	$dist_{3D}$	IoU	$dist_{3D}$	IoU	$dist_{3D}$	IoU	$dist_{3D}$	IoU	$dist_{3D}$	IoU
car	0.018	0.459	2.62e-3	0.908	0.003	0.735	0.018	0.429	2.50e-3	0.932
bicycle	0.043	0.511	2.16e-5	0.971	0.003	0.814	0.050	0.511	2.05e-5	0.977
motorbike	0.025	0.389	3.59e-5	0.958	0.002	0.757	0.033	0.389	3.46e-5	0.968
aeroplane	0.051	0.421	0.001	0.849	0.019	0.625	0.062	0.421	9.62e-3	0.852
bus	0.014	0.442	2.58e-5	0.943	0.001	0.830	0.024	0.389	2.23e-5	0.959
chair	0.053	0.289	0.002	0.872	0.015	0.428	0.057	0.264	0.002	0.871
diningtable	0.155	0.238	0.002	0.865	0.011	0.498	0.262	0.238	0.002	0.876
sofa	0.047	0.394	0.003	0.802	0.060	0.418	0.054	0.374	0.002	0.812
Mean	0.051	0.393	1.34e-3	0.896	0.014	0.638	0.070	0.377	2.24e-3	0.906

Table 3: Quantitative results for our proposed method evaluated on the synthetic test set when considering the selected model (*IDX*), the FFD and the linear combination (*LC*) to be the ground truths, in different combinations.

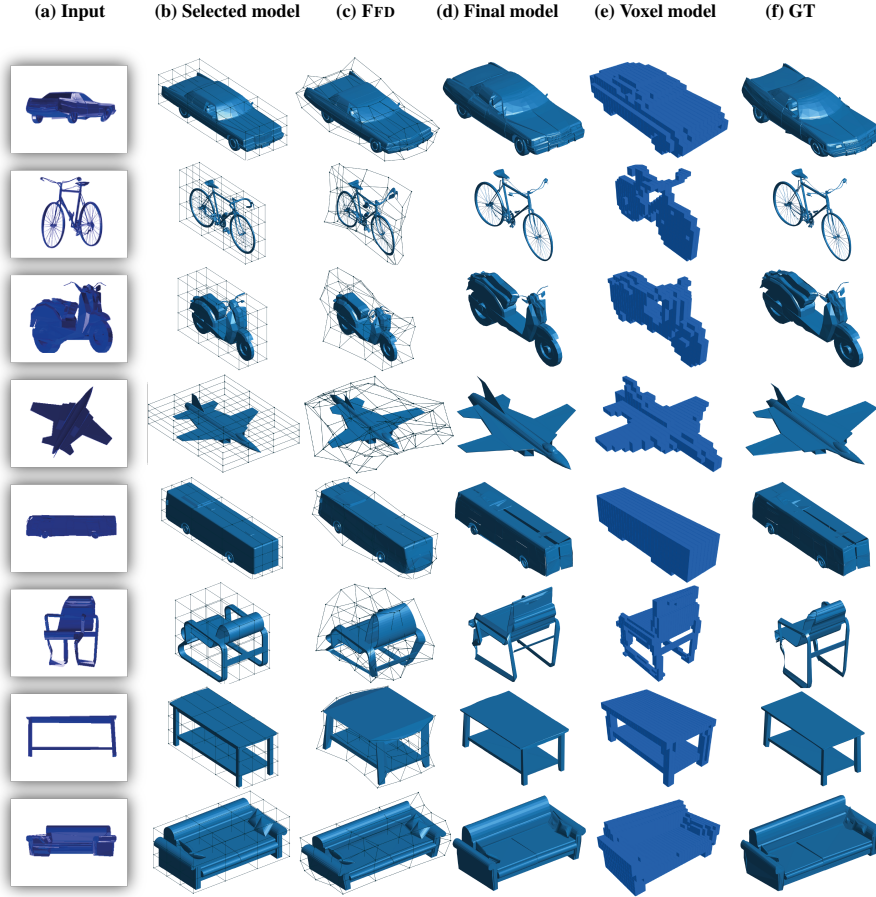


Fig. 1: Visual results on synthetic data. (a) shows the input image; (b) the selected model from the graph; (c) the selected model deformed by FFD. The final 3D model deformed by linear combination is shown in (d). The voxelized final model is shown in (e) and the ground truth in (f).



Fig. 2: Visual results on synthetic data. (a) shows the input image; (b) the selected model from the graph; (c) the selected model deformed by FFD. The final 3D model deformed by linear combination is shown in (d). The voxelized final model is shown in (e) and the ground truth in (f).



Fig. 3: Failure cases. (a) shows the input image; (b) the selected model from the graph; (c) the selected model deformed by FFD. The final 3D model deformed by linear combination is shown in (d). The voxelized final model is shown in (e) and the ground truth in (f).



Fig. 4: Visual results on real-world data. (a) shows the input image; (b) the selected model; (c) the selected model deformed by FFD. The final 3D model reconstructed by linear combination is shown in (d). We compare with [1] in (e) and the ground truth is shown in (f).

References

1. Pontes, J.K., Kong, C., Eriksson, A., Fookes, C., Sridharan, S., Lucey, S.: Compact model representation for 3D reconstruction. In: 3DV. (2017) 6