Supplementary Material for "Multi-Output Learning for Camera Relocalization"

Abner Guzman-Rivera[‡*] Pushmeet Kohli[†] Ben Glocker[♭*] Jamie Shotton[†]
Toby Sharp[†] Andrew Fitzgibbon[†] Shahram Izadi[†]

Microsoft Research[†] University of Illinois[‡] Imperial College London[♭]

1 Model Distortion

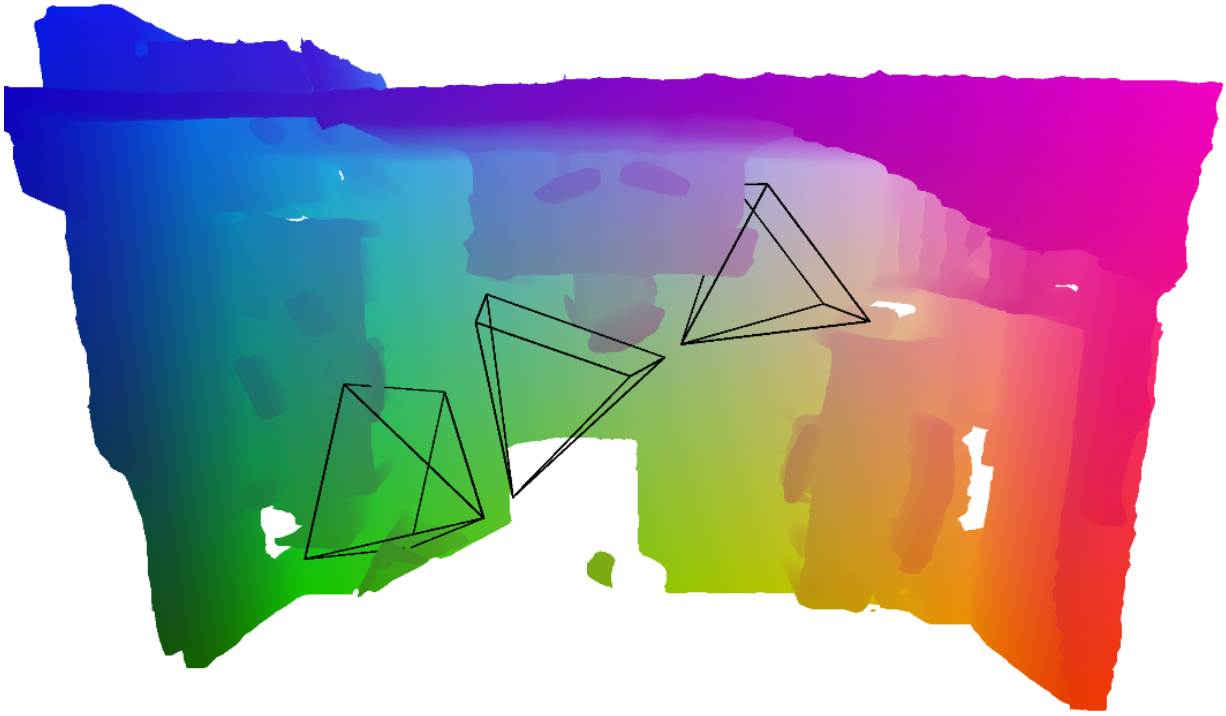


Figure 1: View from above for the scene shown in Fig. 2. Difficult case for the L1 reconstruction error due to model distortion. The distortion is evident in the obliqueness of the two side-walls.

Fig. 1 is an instance of 3D model distortion resulting from camera drift at the time of model reconstruction. We see that the side walls of the room are not parallel as they should be. The deformation is also

[*] Work done while author was at Microsoft Research.

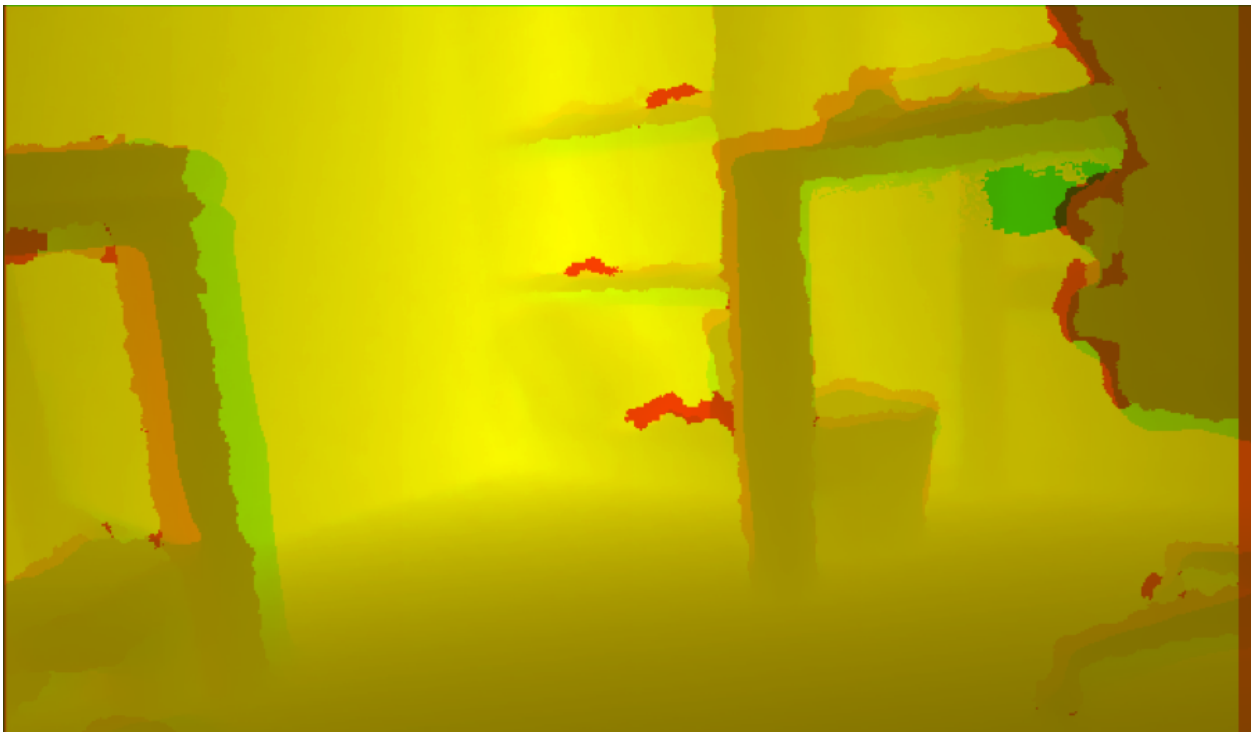
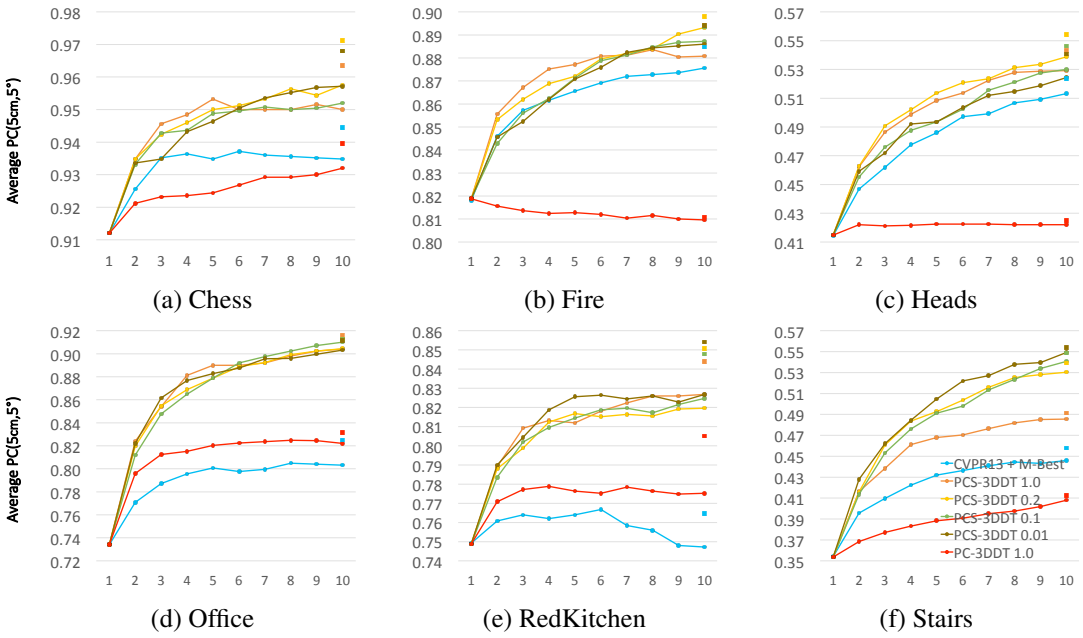


Figure 2: Difficult case for the L1 reconstruction error due to model distortion. Input depth (red channel) and depth raycast from ground-truth (green channel) are shown superimposed. Observe how it would be impossible to align the legs of both desks simultaneously.

evident in the view shown in Fig. 2. There, the rendered view and the input frame are superimposed revealing that it would be impossible to correctly align both desks (simultaneously). For the L1 reconstruction error this translates into an unrealistically high error.

2 Results on Individual Scenes



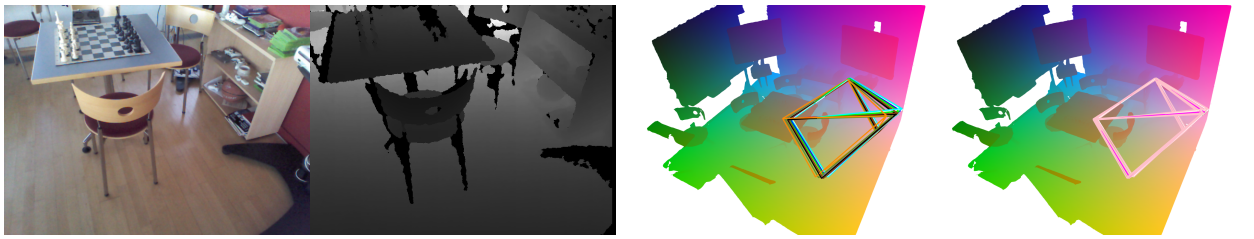
(a) Chess (b) Fire (c) Heads

(d) Office (e) RedKitchen (f) Stairs

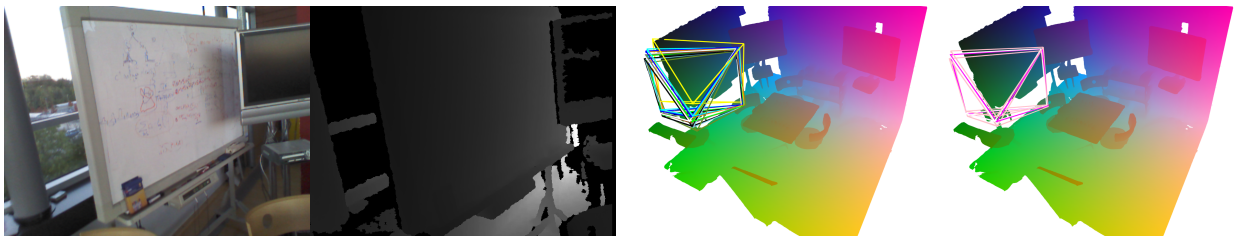
Figure 3: Average $\mathrm{PC}(5\mathrm{cm}, 5^\circ)$ (y-axis) (5 runs per scene) vs. training iteration t (x-axis). Comparison of multi-output models and baselines. The legends indicate the loss, selector and σ used during training. Squares correspond to poses resulting from aggregation.

In Fig. 3 we show results for individual scenes which reveal different trends on a per scene basis. For instance, we see different trends for the effect of parameter σ of the weight update rule. In general, this parameter controls the diversity of the learned predictors. Higher values of σ enable more variability in the example weights – of course, training data also has a direct influence on the obtainable diversity. Note, *e.g.*, how on Fig. 3f (Stairs) the *lowest* value of σ clearly outperforms other settings. This is to contrast with, *e.g.*, Fig. 3d (Office) where the *highest* value of σ is the best performing.

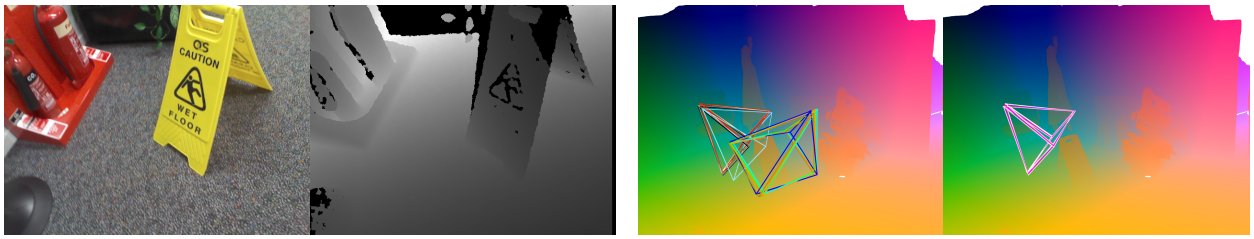
3 More Qualitative Results



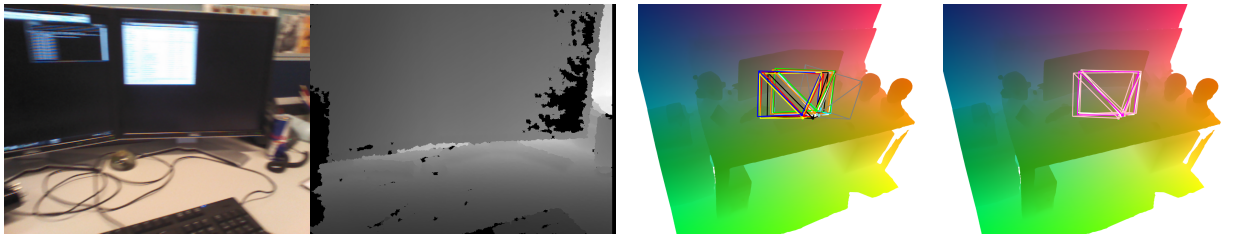
(a) Chess.



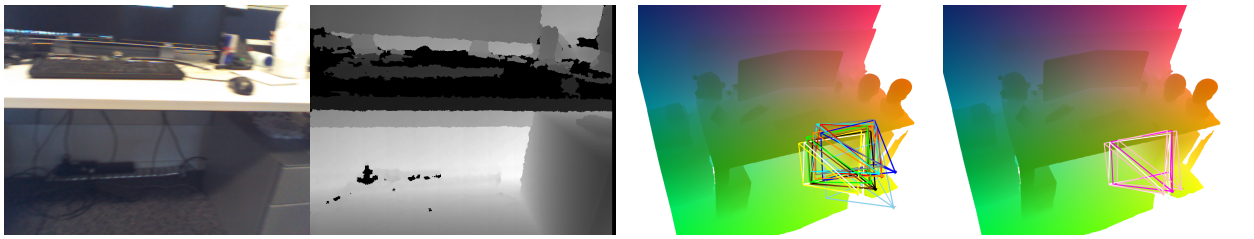
(b) Chess.



(c) Fire.

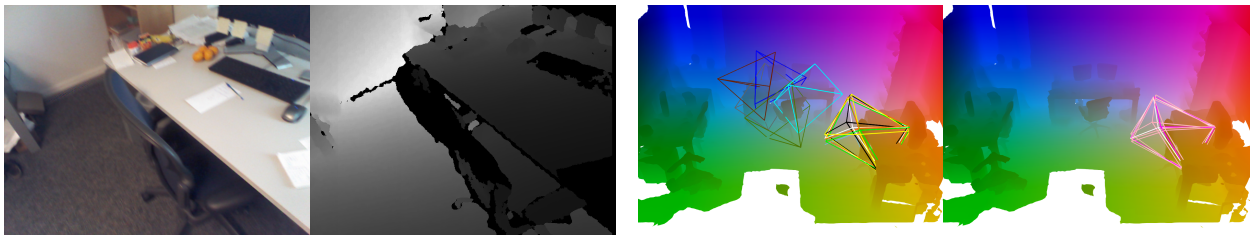


(d) Heads.

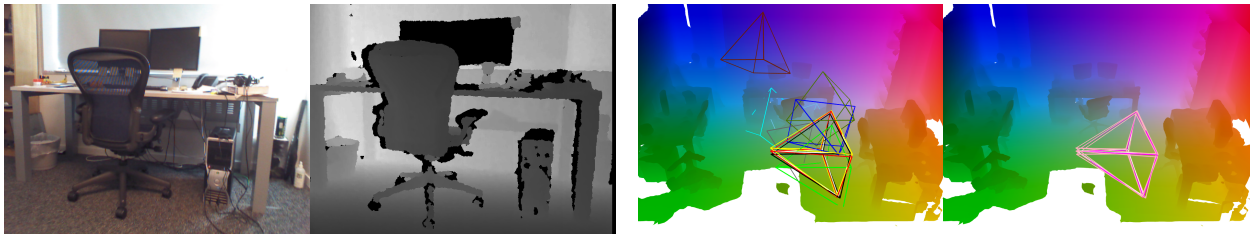


(e) Heads.

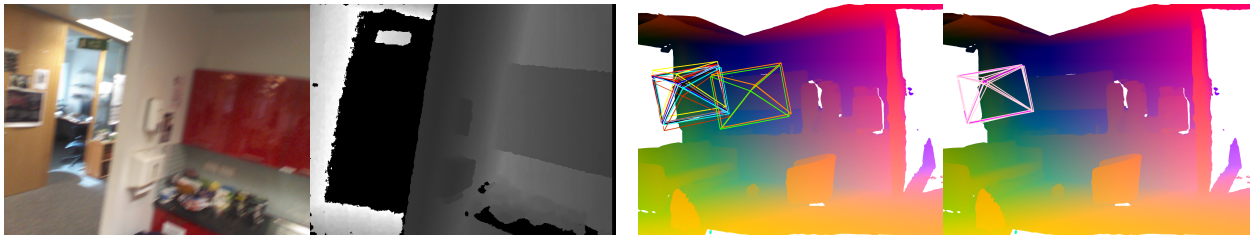
Figure 4: Qualitative camera relocalization results. Left-pair is the input RGB-D frame. Right-pair left: M predictions (colors); ground-truth (white); and selector's pick (black). Right-pair right: Poses in best-scoring cluster (pink); cluster mean (magenta); and ground-truth (white).



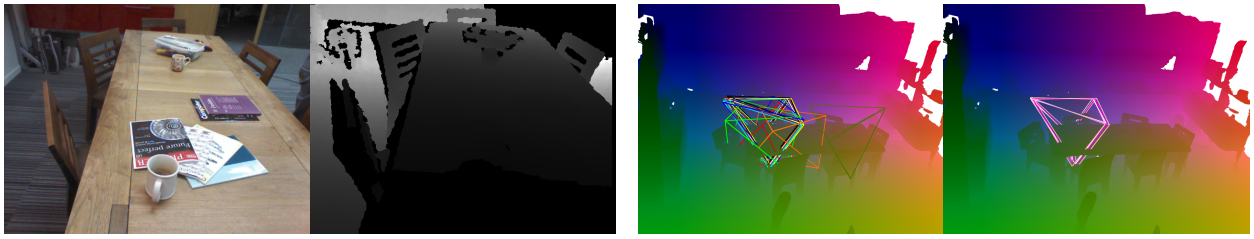
(a) Office.



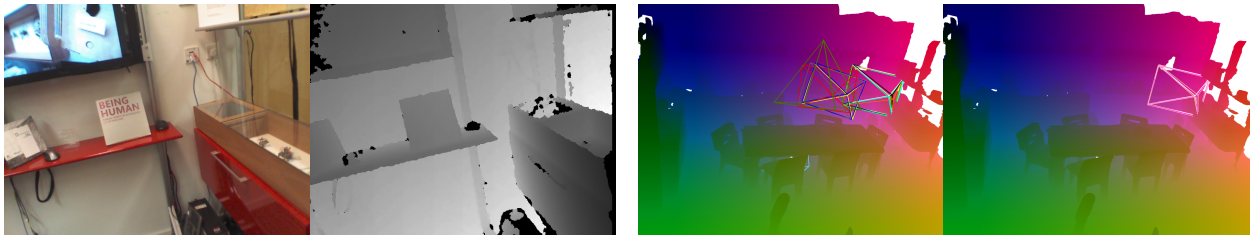
(b) Office.



(c) Pumpkin.

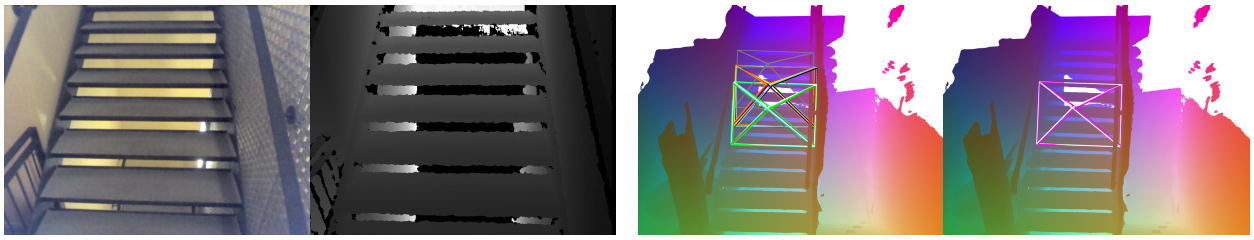


(d) RedKitchen.



(e) RedKitchen.

Figure 5: Qualitative camera relocalization results. Left-pair is the input RGB-D frame. Right-pair left: M predictions (colors); ground-truth (white); and selector's pick (black). Right-pair right: Poses in best-scoring cluster (pink); cluster mean (magenta); and ground-truth (white).



(a) Stairs.

Figure 6: Qualitative camera relocalization results. Left-pair is the input RGB-D frame. Right-pair left: M predictions (colors); ground-truth (white); and selector's pick (black). Right-pair right: Poses in best-scoring cluster (pink); cluster mean (magenta); and ground-truth (white).