Appendix - LocalBins: Improving Depth Estimation by Learning Local Distributions

1 Visualizing bin predictions

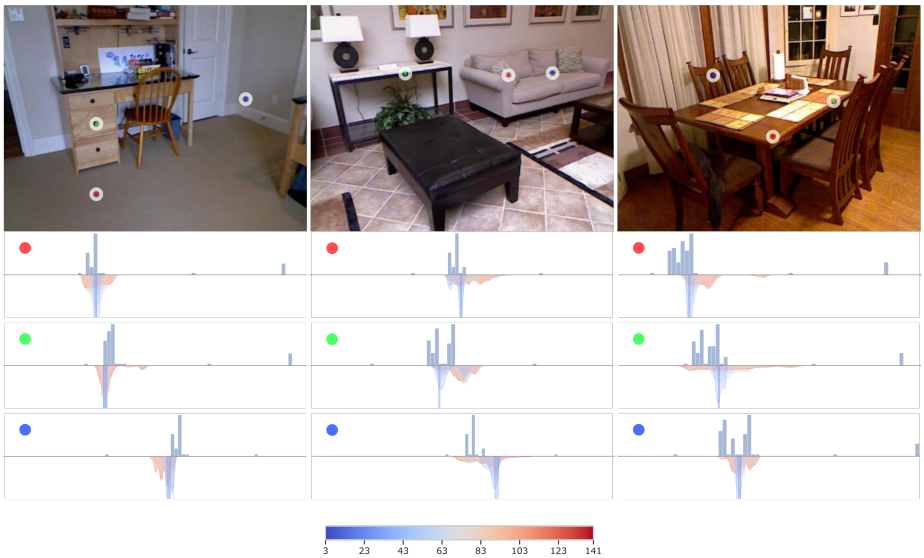

Fig. 1. Visualization of bin predictions, showing input RGB (first row) and density plots of local depth distributions at selected locations across the depth range interval. Density plots consisting of density of bin centers predicted by LocalBins module at the selected pixel location (**top**) and density of ground truth depth values (**bottom**) for various window sizes (indicated by the colorbar).

Fig. 1 shows a qualitative comparison of the bin predictions of the LocalBins module against the ground truth depth distribution in the local neighborhoods of various sizes.

Our main goal in this work was to have the LocalBins module predict the local depth distribution at each pixel. However, in Query-Response training, windows of various sizes are generated, and regularization is imposed on the bin predictions of the entire window together rather than for each individual pixel location separately. While being vital (refer to 'Query-Response training' section in the main paper), this renders the 'size' of the 'local neighborhood' rather indeterministic. The model may either end up using a fixed size and

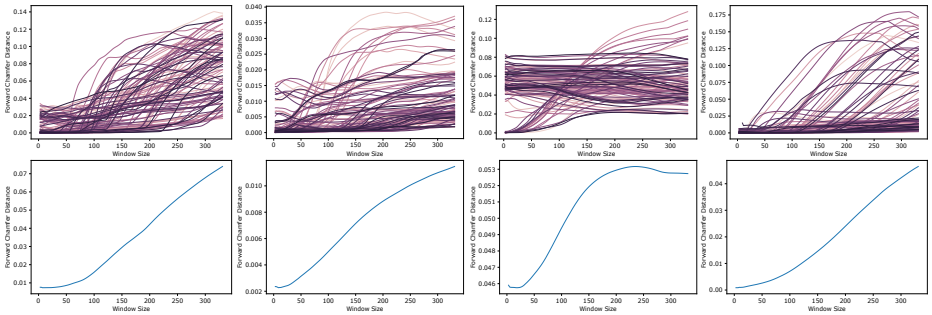

Fig. 2. How *"local"* are the bin predictions of the LocalBins module? Plots show the average absolute difference between the ground truth depth values and their nearest bin centers predicted by the LocalBins module. Top row shows the average differences plotted for 100 random locations for four randomly selected input images (columns). GT depth values are taken from the 'concentric' bounding boxes and compared against the bins predicted at the center. Bottom row shows the mean across the 100 locations for each image.

always predict the bins at each pixel that reflect the density of depth values within a fixed-sized neighborhood, or the model may as well choose to cover different sized neighborhoods depending upon the context. Visualizations (See Fig. 2) indicate the latter case.

2 Different Backbones

Backbone	#params(M)	d1	d2	d3	REL	RMS	log10
MobileNetV2-100	20.26	0.812	0.963	0.992	0.141	0.480	0.060
DenseNet-161	102.4	0.898	0.978	0.995	0.105	0.369	0.045
EfficientNetV2-S	38.71	0.894	0.981	0.995	0.106	0.376	0.045
EfficientNetV2-M	71.53	0.898	0.983	0.996	0.102	0.365	0.044
EfficientNetV2-L	136.3	0.910	0.986	0.997	0.098	0.351	0.042

Table 1. Performance of LocalBins with various backbone encoders

We switch the EfficientNet-b5 encoder in our default model with various other backbones and list the performance in Table. 1