Recurrent Scene Parsing with Perspective Understanding in the Loop

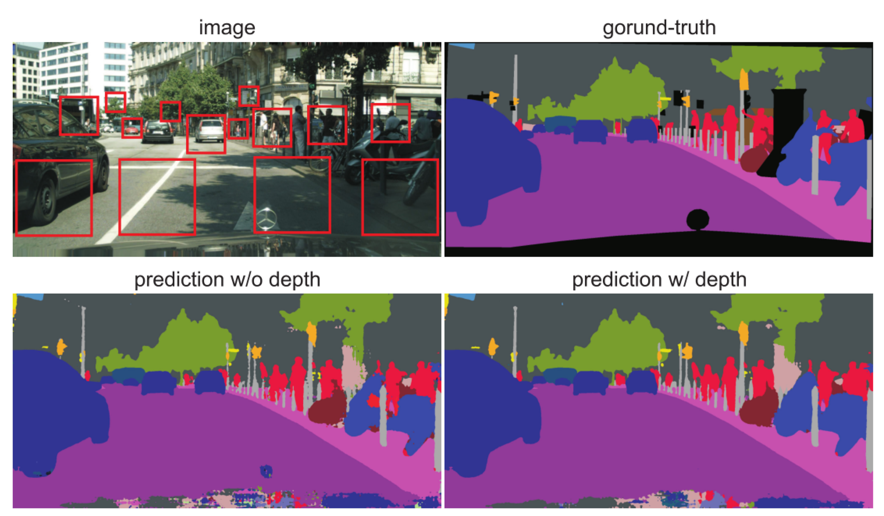
Shu Kong, Charless Fowlkes

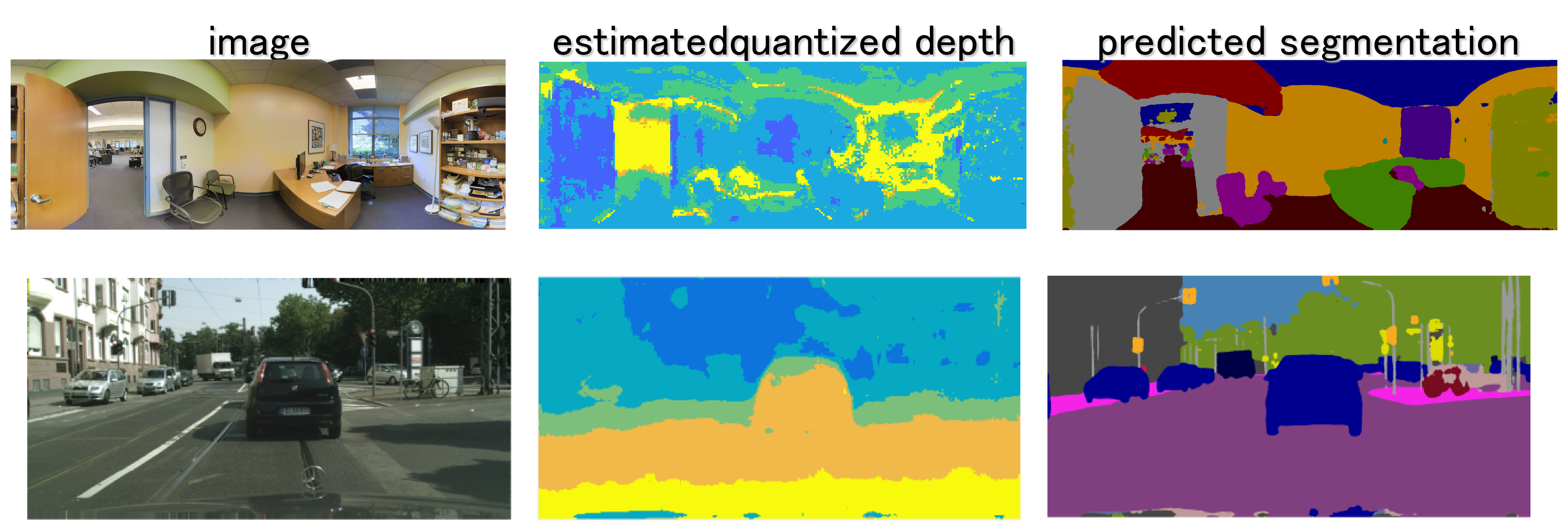
Department of Computer Science, University of California, Irvine, California, USA

Abstract

Depth-aware Gating Module

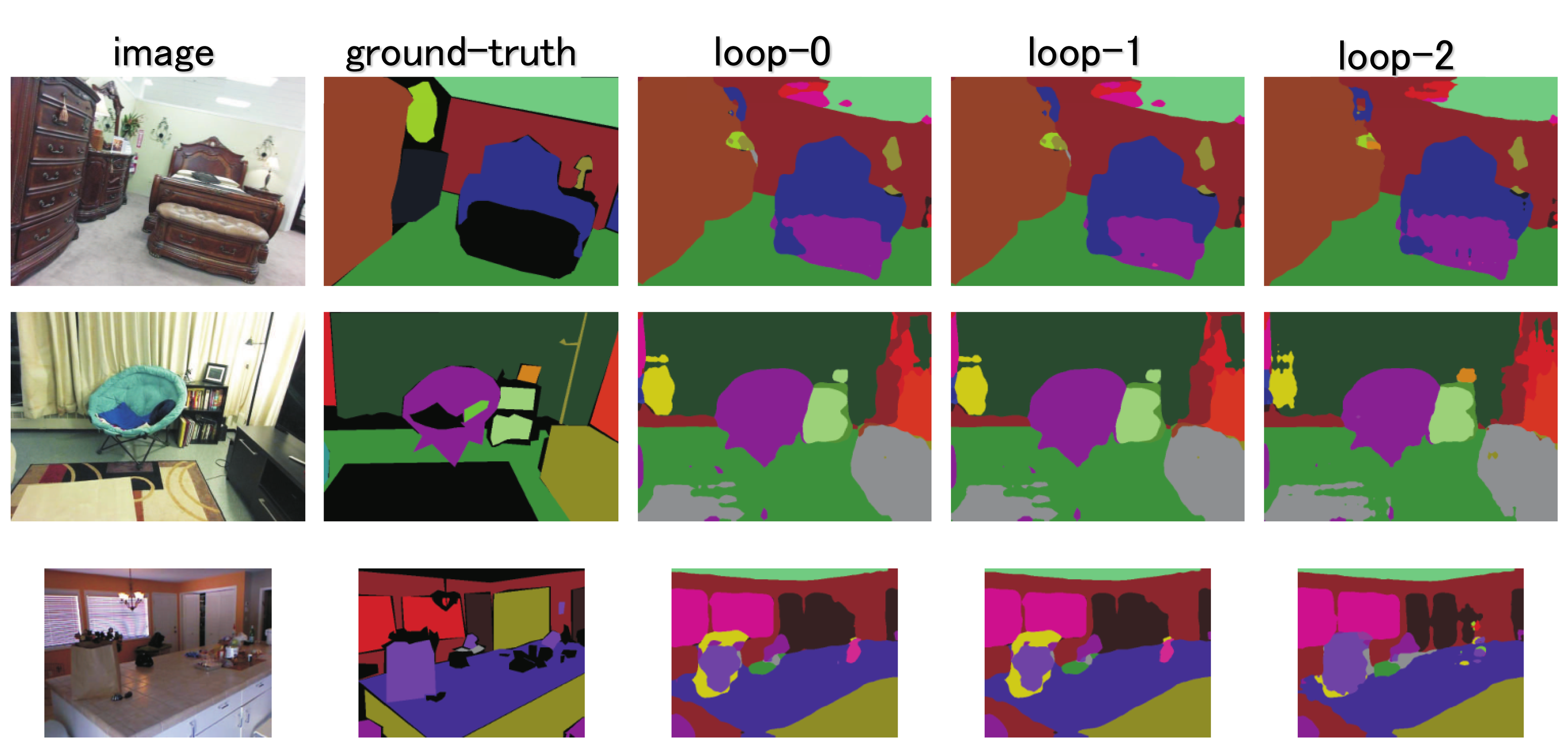
We propose adaptively choosing the pooling field size according to the object scale (inversely proportional to the depth), so that small details can be preserved for objects at distance and a larger receptive field can be used for objects nearer to the camera.





Recurrent Refinement Module

We iteratively refine the segmentation results, leveraging the depth estimate and output prediction from the previous loop.



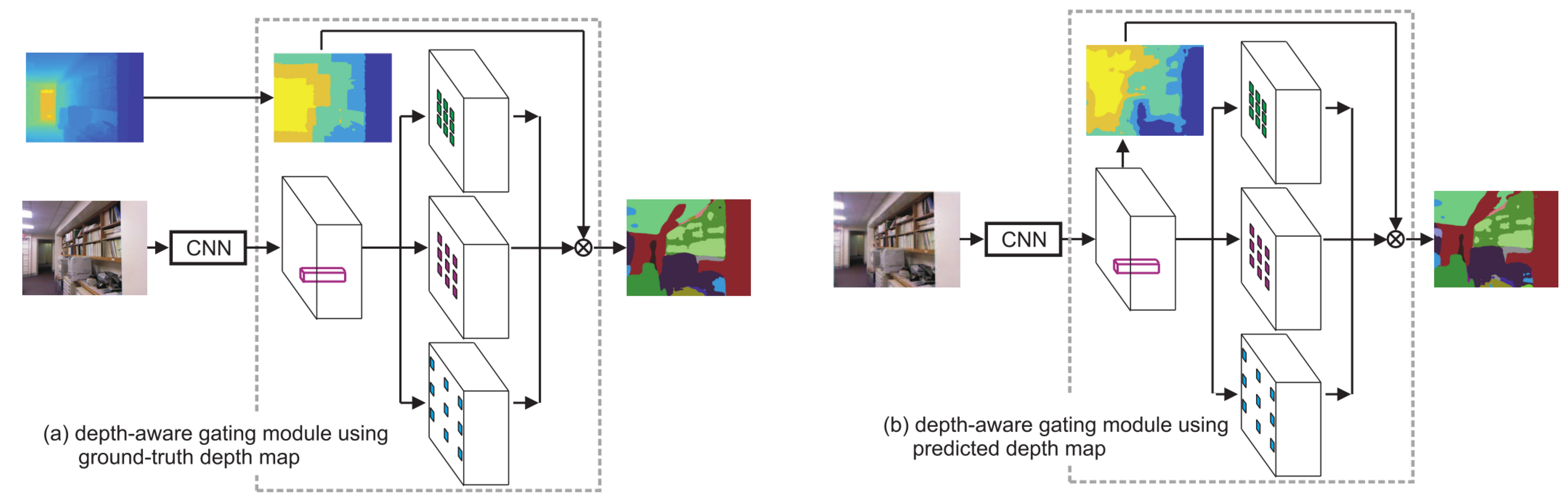
Broad Impact and Potential Application

1. Parsing panoramas and estimating real-world scale for AR and VR, (e.g. panorama navigation)
2. Parsing street scene for autonomous vehicles
3. Parsing and navigating indoor scenes for assistive robotics

Demo, code and model can be found at the project webpage under author's webpage
http://www.ics.uci.edu/~skong2/recurrentDepthSeg

Acknowledgements: This work was supported by NSF grants DBI-1262547, IIS-1253538, hardware donation from NVIDIA.

Depth-Aware Gating Module



- Depth map is quantized into five discrete scales in our experiments. Disparity/depth is either provided as input or estimated directly from the input RGB image if stereo is not available.
- Using depth-gated pooling allows for more accurate segment label predictions by avoiding pooling across small multiple distant objects while simultaneously allowing sufficiently large pooling fields for nearby objects.
- Depth prediction branch achieves starte-of-the-art performance in monocular depth estimation (shown below)

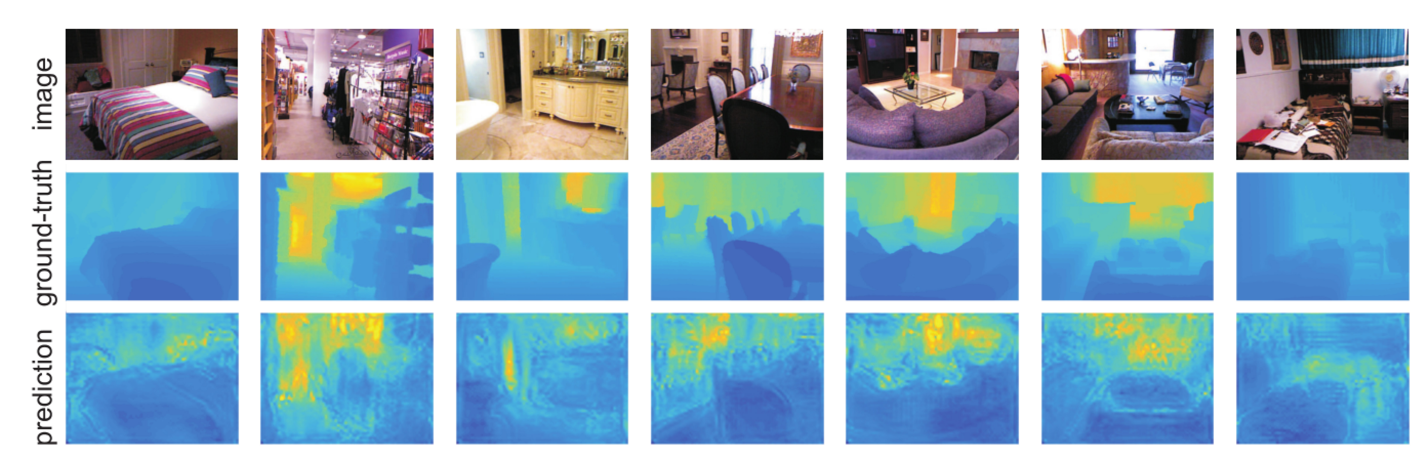


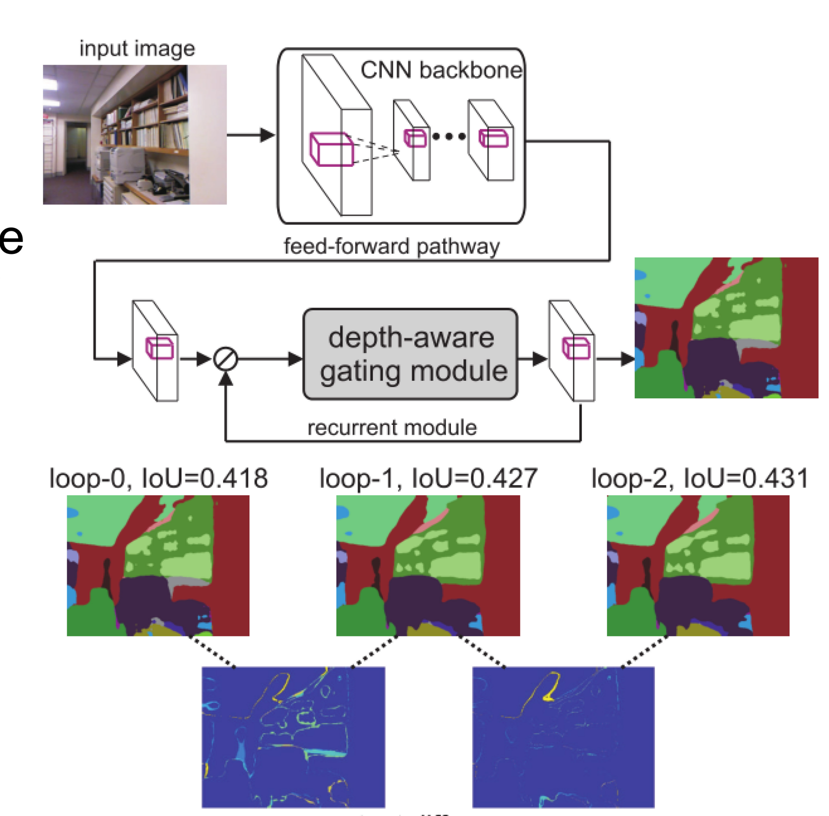
Table 1: Depth prediction on NYU-depth-v2 dataset.

	Ladicky [17]	Liu [22]	Eigen [8]	Eigen [7]	Ours	Ours-Blur
$\delta < 1.25$	0.542	0.614	0.614	0.769	0.809	0.816
$\delta < 1.25^2$	0.829	0.883	0.888	0.950	0.945	0.950
$\delta < 1.25^3$	0.940	0.971	0.972	0.988	0.986	0.989

Recurrent Refinement Module

Updated depth predictions at each iteration gate pooling fields used for semantic segmentation, increasing the flexibility and representation power of our system and yielding improved segmentation.

The figure at right shows the segmentation prediction prior to, and after two recurrent iterations for a particular image, as well as the difference in predictions between consecutive iterations.



Experiment

- Four datasets are used for evaluation where our models achieves state-of-the-art or competitative performance.
- Experiments validate the effectiveness of depth-aware gating module and recurrent refinement module.
- Performance is improved remarkably on datasets with large perspective variations, e.g. Stanford-2D-3D and Cityscapes.

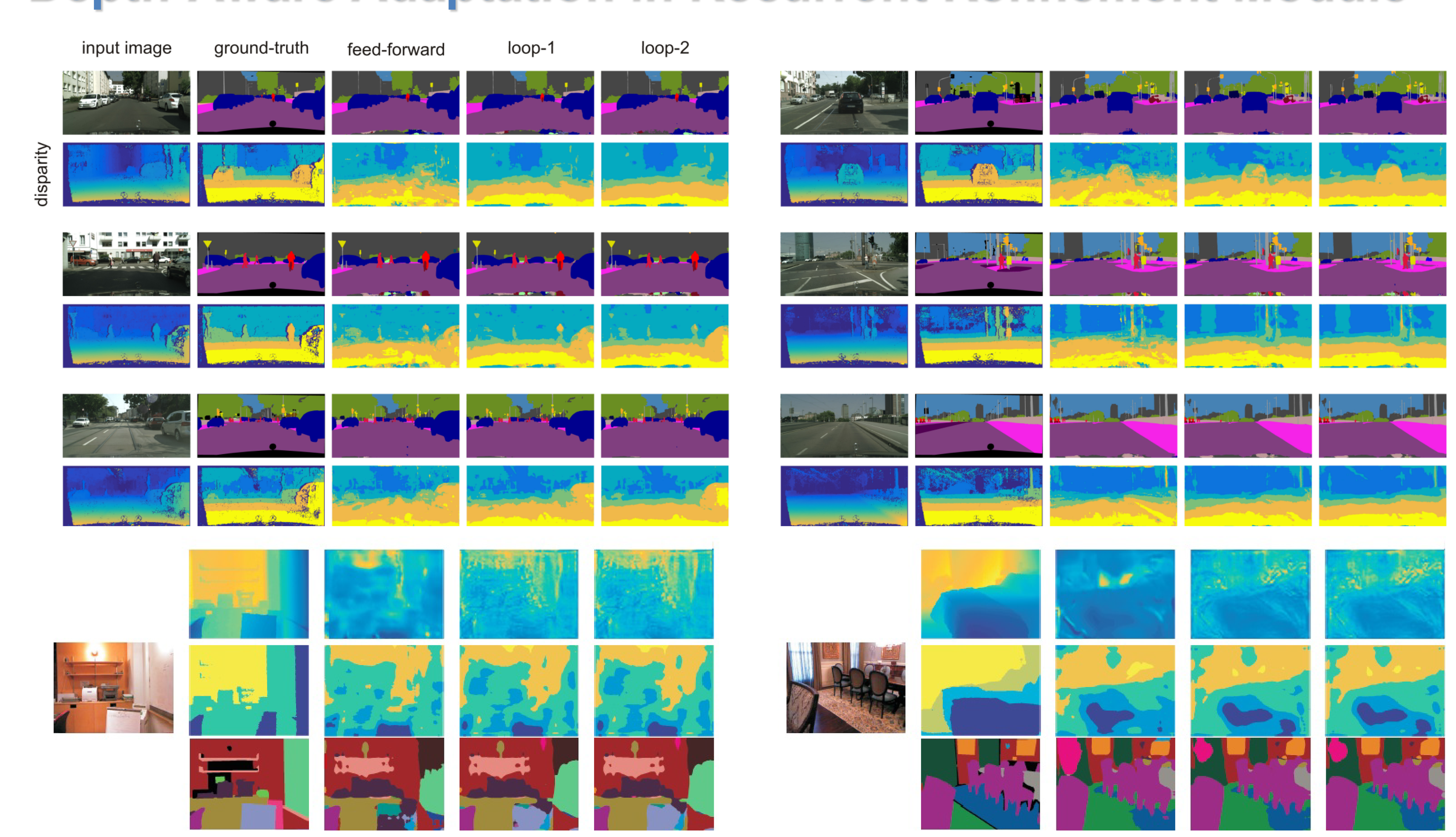
	NYU-depth-v2		SUN-RGBD		Stanford-2D-3D		Cityscapes
	IoU	pixel acc.	IoU	pixel acc.	IoU	pixel acc.	IoU
baseline	0.406	0.703	0.402	0.776	0.644	0.866	0.738
w/ gt-depth	0.413	0.708	0.422	0.787	0.730	0.897	0.753
w/ pred-depth	0.418	0.711	0.423	0.789	0.742	0.900	0.759
loop1 w/o depth	0.419	0.706	0.432	0.793	0.744	0.901	0.762
loop1 w/ gt-depth	0.425	0.711	0.439	0.798	0.747	0.902	0.769
loop1 w/ pred-depth	0.427	0.712	0.440	0.798	0.753	0.906	0.772
loop2	0.431	0.713	0.443	0.799	0.760	0.908	0.776
loop2 (test-aug)	0.445	0.721	0.451	0.803	0.765	0.910	0.791 / 0.782*
DeepLab [2]	-	-	-	-	0.698†	0.880†	0.704 / 0.704*
LRR [4]	-	-	-	-	-	-	0.700 / 0.697*
Context [5]	0.406	0.700	0.423	0.784	-	-	- / 0.716*
PSPNet [7]	-	-	-	-	0.674†	0.876†	- / 0.784*
RefineNet-Res50 [6]	0.438	-	-	-	-	-	- / -
RefineNet-Res101 [6]	0.447	-	0.457	0.804	-	-	- / 0.736*
RefineNet-Res152 [6]	0.465	0.736	0.459	0.806	-	-	- / -

- Training an gating module without using quantized depth as a supervisory signal also improves over baseline significantly, but doesn't fully utilize all branches.

	baseline		tied, avg.		gt-depth, tied, gating		gt-depth, untied, gating		attention, untied, gating		pred-depth, untied, gating	
	IoU	nIoU	IoU	nIoU	IoU	nIoU	IoU	nIoU	IoU	nIoU	IoU	nIoU
Score Avg.	**0.738**	**0.547**	**0.747**	**0.554**	**0.748**	**0.556**	**0.753**	**0.561**	**0.755**	**0.562**	**0.759**	**0.571**

Visualization

Depth-Aware Adaptation in Recurrent Refinement Module



Visualization of Attention Model