Labeling Panoramas with Spherical Hourglass Networks

Carlos Esteves[1,2], Kostas Daniilidis[1], Ameesh Makadia[2]

[1]GRASP Laboratory, University of Pennsylvania [2]Google
{machc,kostas}@seas.upenn.edu makadia@google.com

Introduction

- *Contribution:* a model for dense labeling of spherical panoramas in arbitrary orientation and no augmentation. This is the first Spherical CNN architecture that scales to challenging problems such as semantic segmentation.
- *Conclusion:* proper spherical convolutions mitigate sampling problems, achieve **SO**(3) equivariance, and scale well to deeper architectures.

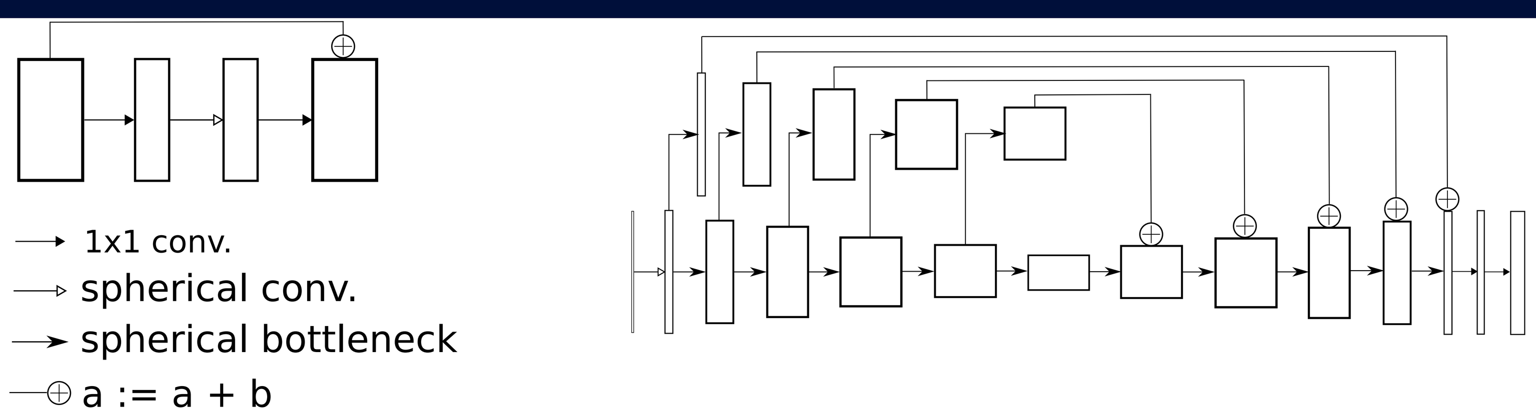
Spherical convolutions

- f, h are functions on the sphere S^2,
- **SO**(3) is the group of 3D rotations,
- η is the north pole.

$$(f \star h)(x) = \int_{g \in \mathbf{SO}(3)} f(g\eta) h(g^{-1}x)\, dg, \quad x \in S^2.$$

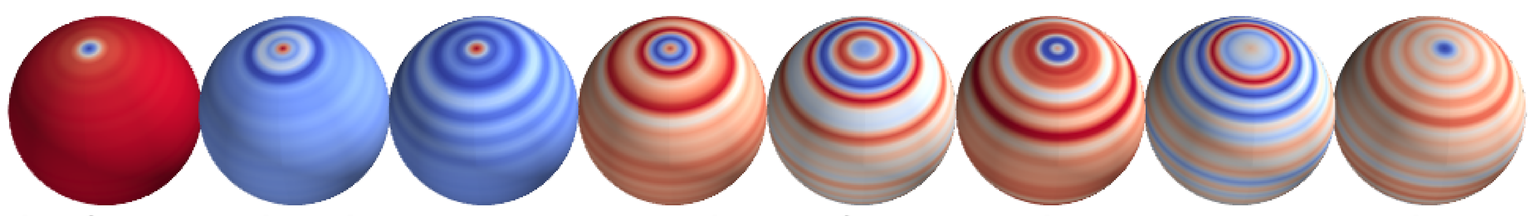
- Computed exactly for bandlimited inputs in the spectral domain (spherical harmonics expansion).
- Filters are zonal. Spherical correlation allows arbitrary filters, but is much more computationally expensive.

Method



Left: Spherical residual bottleneck block. *Right:* Spherical hourglass architecture.

- Spherical convolution implementation based on Esteves *et al.* ECCV 2018 (oral session 4A, Sep 13 8:30am).
- Filter localization is important, and can be tuned in the spectral parametrization; filters can be dilated as in "Dilated Convolutions" (Yu *et al.* 2016).



Left to right: Increasing number of spectral parameters enlarge the filter receptive field.

- Separation of variables method allows for an efficient convolution even at higher resolutions e.g. 256×256.
- Spherical residual block, inspired by He *et al.* 2016, allows more expressivity with faster operations.
- Hourglass architecture inspired by Newell *et al.* 2016.
- Nonlinearities break bandlimit assumption, increasing equivariance error. Using a large lowest resolution inside the hourglass (32×32) mitigates the problem.

Results

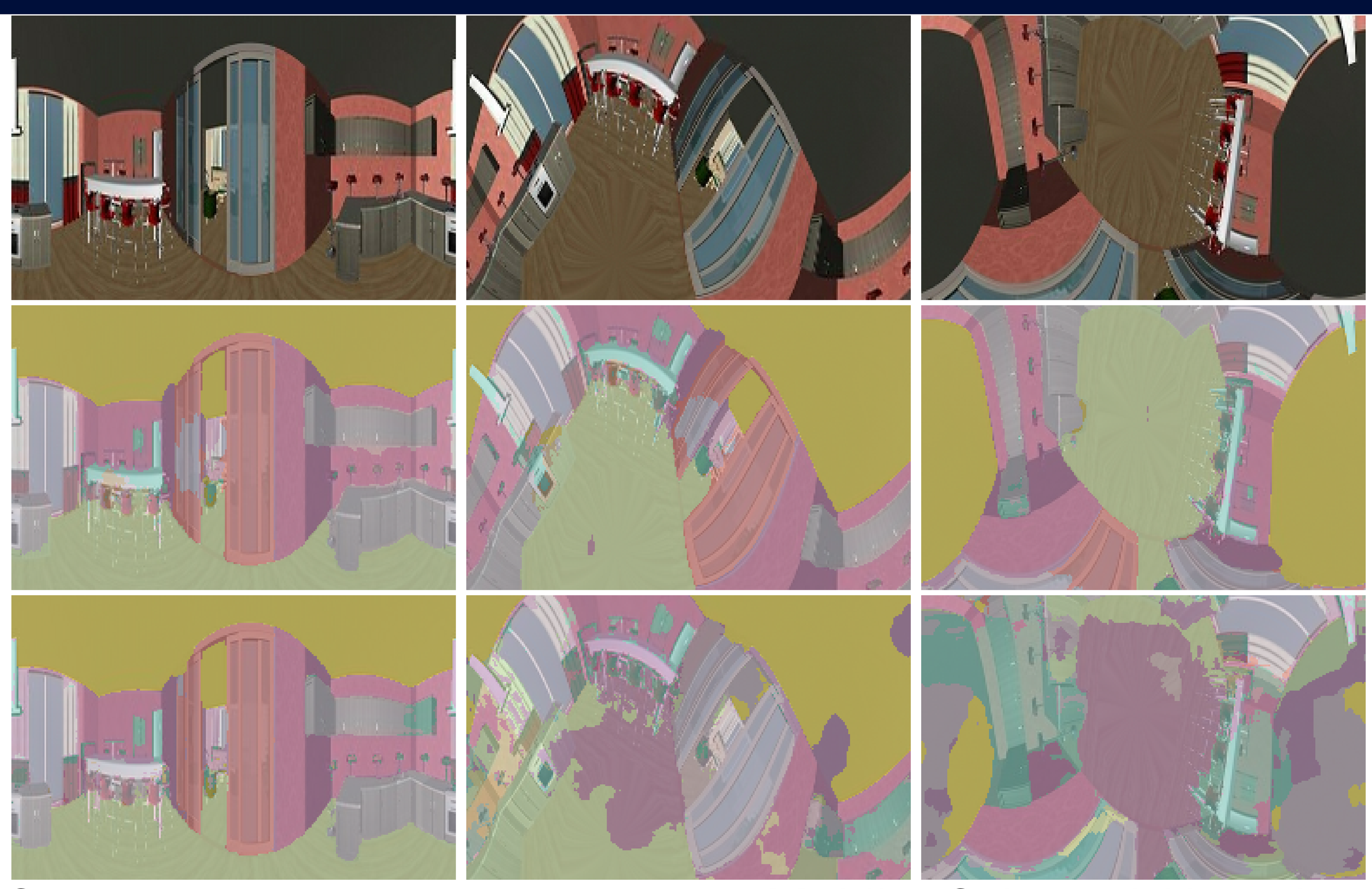
- Experiments on SunCG (Song *et al.* 2017) panoramas.
- Combinations of orientations seen in training/test time (canonical or random 3D rotation).
- Baseline replaces every spherical with a 2D convolution.

Mean IoU. *c* is canonical orientation, *3d* is arbitrary.

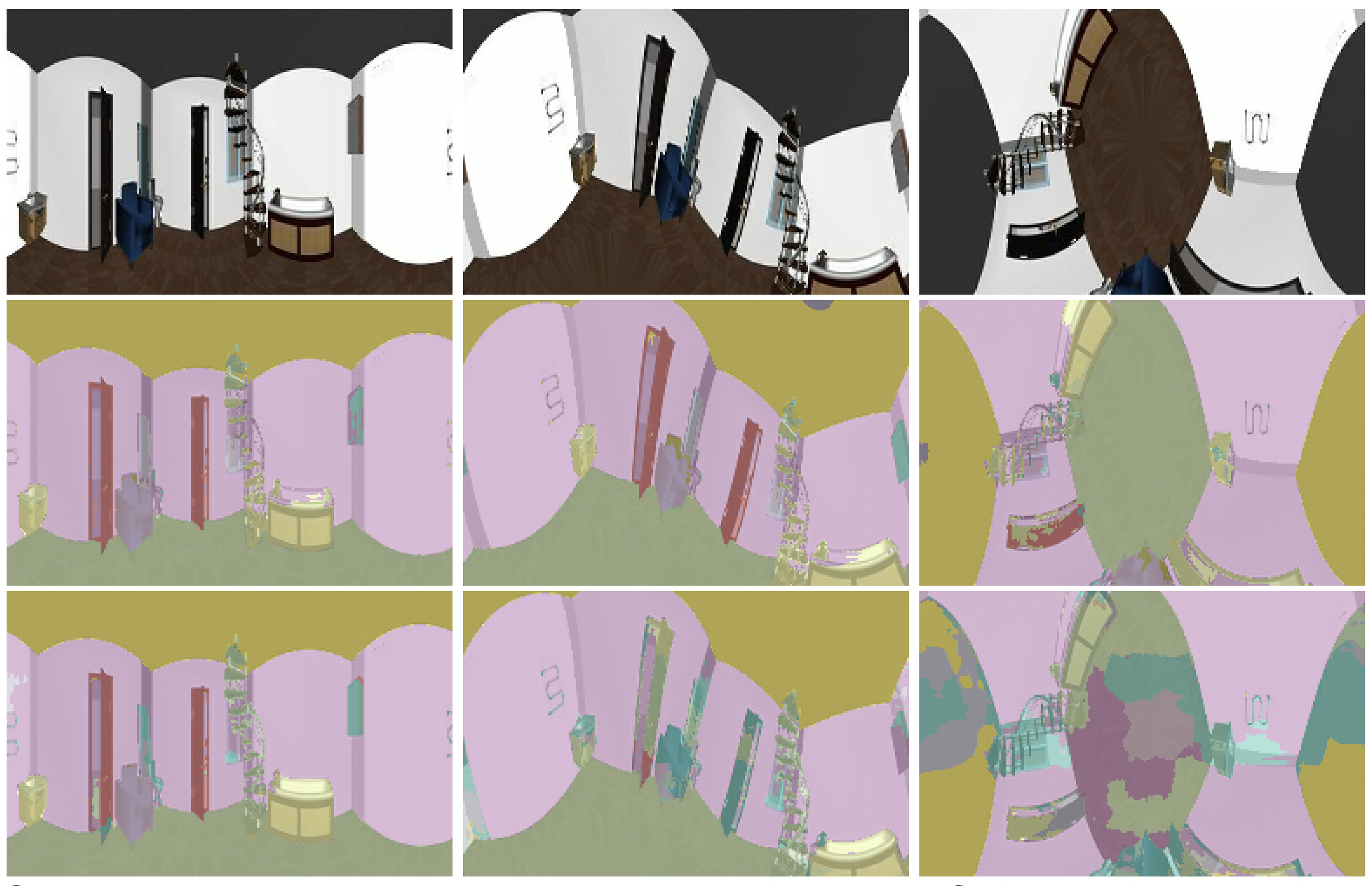
	train/test orientation		
	c/c	3d/3d	c/3d
SCHN (ours)	0.5683	**0.5582**	**0.5024**
2DHG	**0.6393**	0.5292	0.2237
SCHN/global	0.5343	0.5376	0.4758
SCHN/large	0.5983	0.5873	-

- Ours is superior when viewpoint is arbitrary, especially if there are unseen viewpoints in the test set.
- Lower performance on canonical orientation is due to radially symmetric nature of the filters.

Visualization



Scene 1. *Top:* input panoramas. *Middle:* Spherical Hourglass predictions. *Bottom:* 2D Hourglass predictions.



Scene 2. *Top:* input panoramas. *Middle:* Spherical Hourglass predictions. *Bottom:* 2D Hourglass predictions.

Conclusion

- Proper spherical convolutions are advantageous when dealing with spherical images, especiallly when no viewpoint assumption can be made.
- Deep Spherical CNN architectures are viable for more demanding tasks such as dense labeling.