Penn Engineering
GRASP LABORATORY
General Robotics, Automation, Sensing & Perception Lab

Learning SO(3) Equivariant Representations with Spherical CNNs

Carlos Esteves[1], Christine Allen-Blanchette[1], Ameesh Makadia[2], Kostas Daniilidis[1]

[1]GRASP Laboratory, University of Pennsylvania
{machc,allec,kostas}@seas.upenn.edu

[2]Google
makadia@google.com

Research at Google

Introduction

- *Motivation:* SO(3) equivariant representations are desirable for shape and spherical image analysis in arbitrary orientations.
- *Contributions:* (1) First neural network based on spherical convolutions, (2) novel pooling and filter parameterization methods, (3) lower capacity than competing networks applied on 3D data without sacrificing performance.

Spherical convolution ($*$) and correlation ($\star$)

$$(f * h)(x) = \int_{R \in \mathbf{SO}(3)} f(R\eta)h(R^{-1}x)\, dR$$

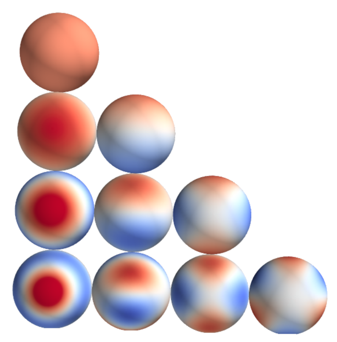
$$(f \star h)(R) = \int_{x \in S^2} f(x)h(Rx)\, dx$$

$$\hat{f}^\ell_m = \int_{S^2} f(x)\overline{Y^\ell_m}(x)dx.$$

$$\widehat{(f * h)}^\ell_m = \alpha_\ell \hat{f}^\ell_m \hat{h}^\ell_0$$

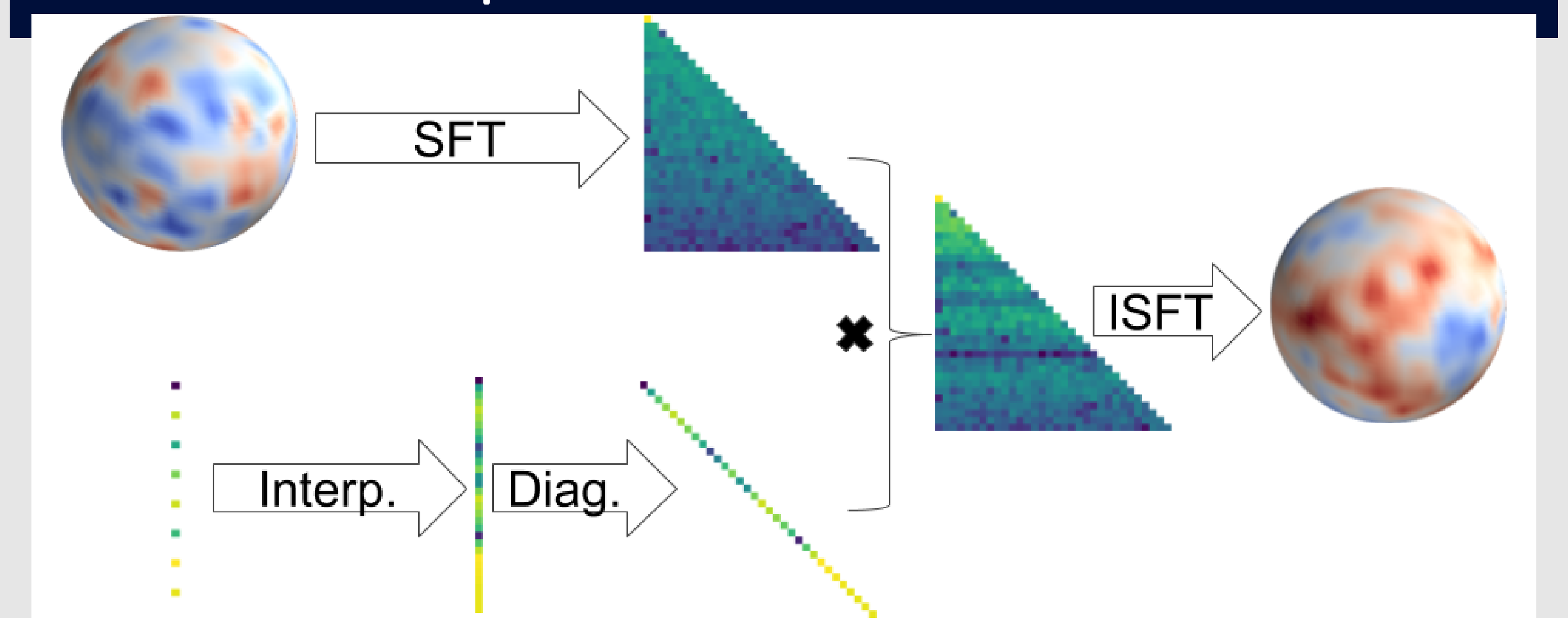
Convolution, compared with correlation:

- filters and outputs in S^2,
- faster and memory efficient,
- parameter efficient, but less discriminative filters.



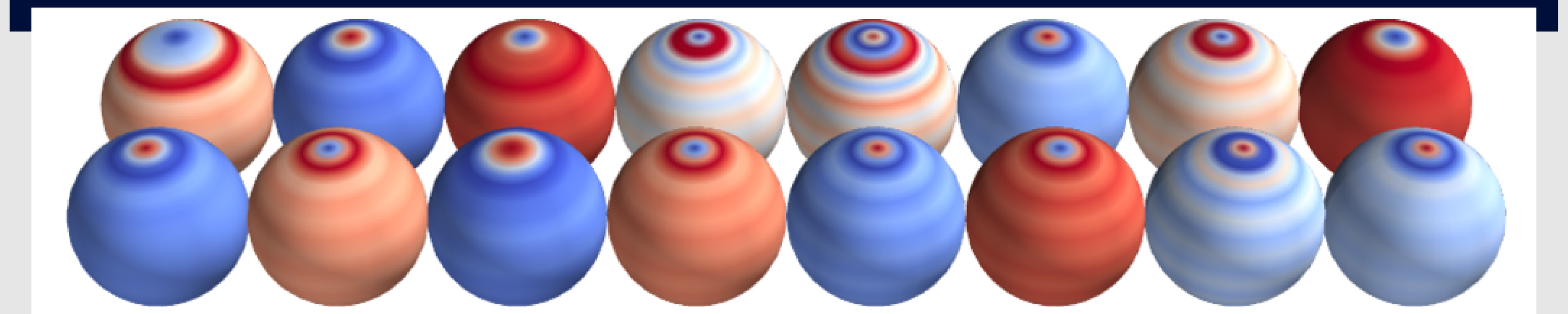
Spherical harmonics

Spherical convolutional block

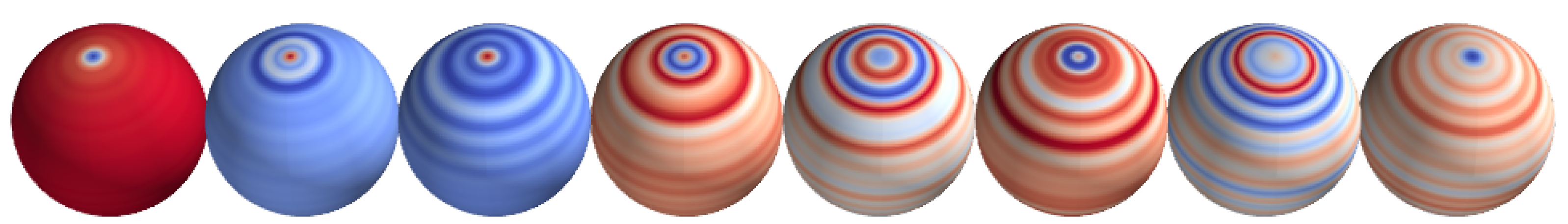


(I)SFT: (inverse) spherical Fourier transform

Localized filters

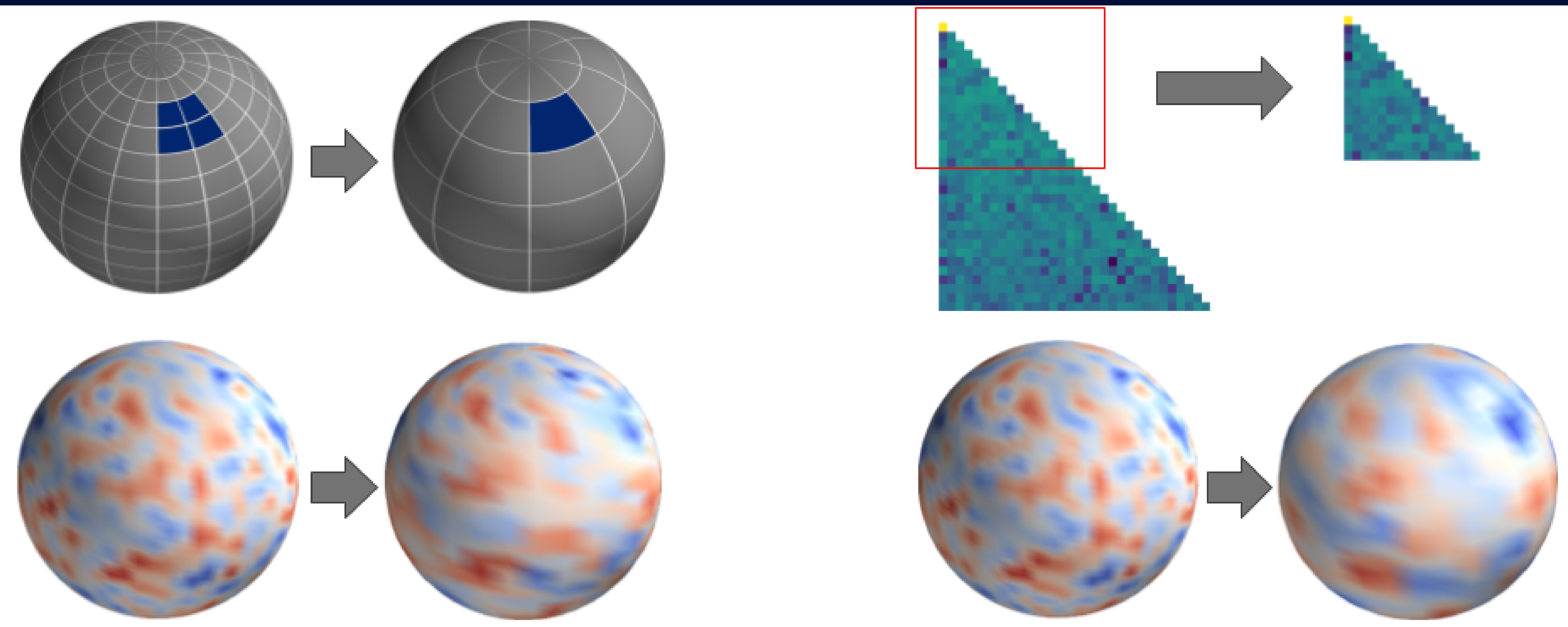


Learned filters with 8 anchors



Left to right: increasing number of anchors

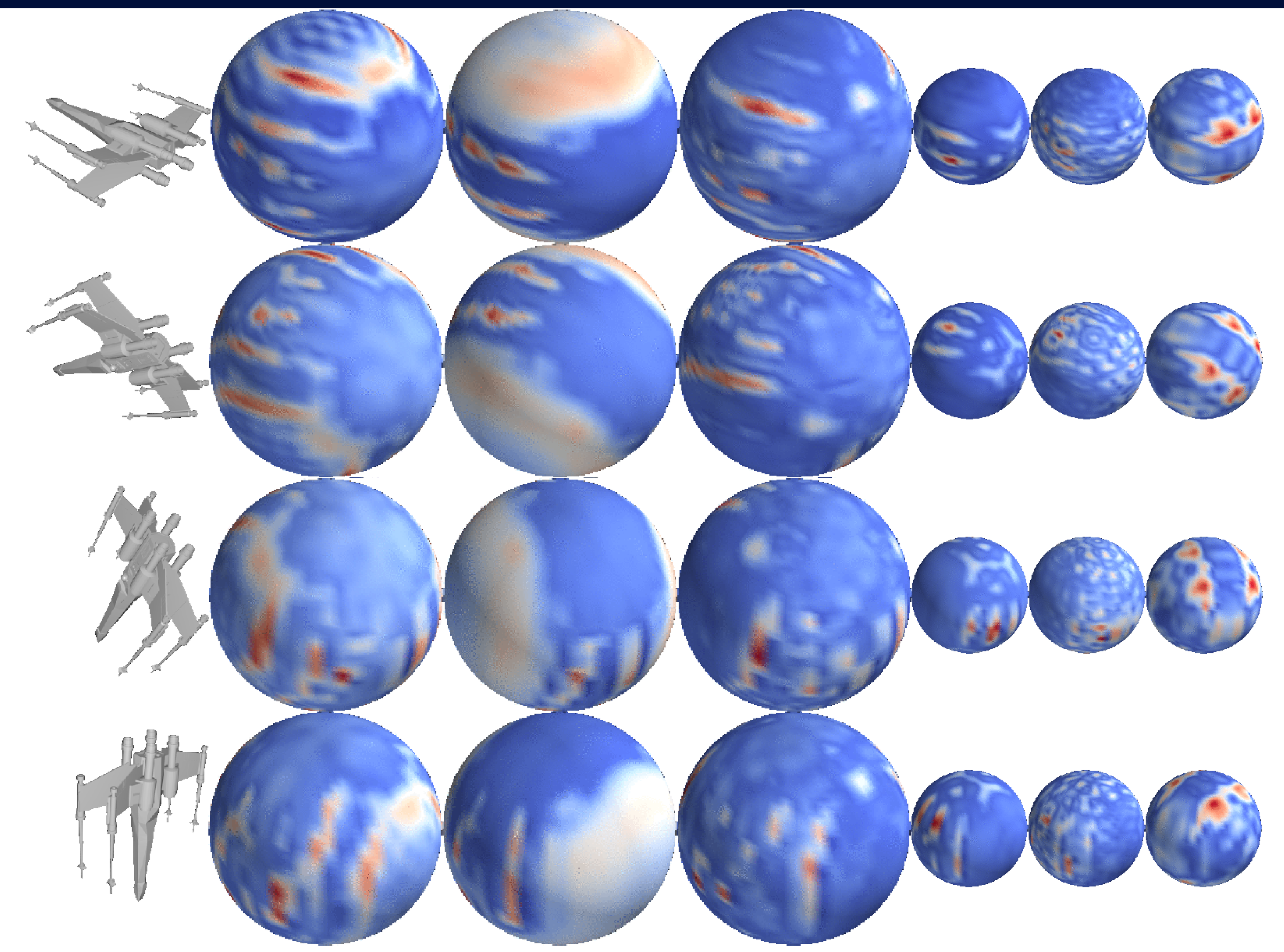
Pooling



Weighted average pooling

Spectral pooling

Equivariant feature maps



Rows show feature maps obtained from leftmost input.
Each column shows a single channel from a different layer.

SHREC'17 perturbed shape retrieval

We match the state of the art with 16x fewer parameters.

	micro			macro			total		
	P@N	R@N	mAP	P@N	R@N	mAP	score	input size	params
Furuya[1]	**0.814**	0.683	0.656	**0.607**	0.539	**0.476**	**1.13**	126×10^3	8.4M
Ours	*0.717*	*0.737*	*0.685*	*0.450*	*0.550*	*0.444*	**1.13**	$\mathbf{2 \times 64^2}$	**0.5M**
Tatsuma[2]	0.705	**0.769**	**0.696**	0.424	**0.563**	0.418	*1.11*	38×224^2	3M
Cohen[3]	0.701	0.711	0.676	-	-	-	-	6×128^2	*1.4M*
Zhou[4]	0.660	0.650	0.567	0.443	0.508	0.406	0.97	50×224^2	36M

[1]Furuya *et al.* BMVC'16, [2]Tatsuma *et al.* TVC'09, [3]Cohen *et al.* ICLR'18, [4]Zhou *et al.* CVPR'16

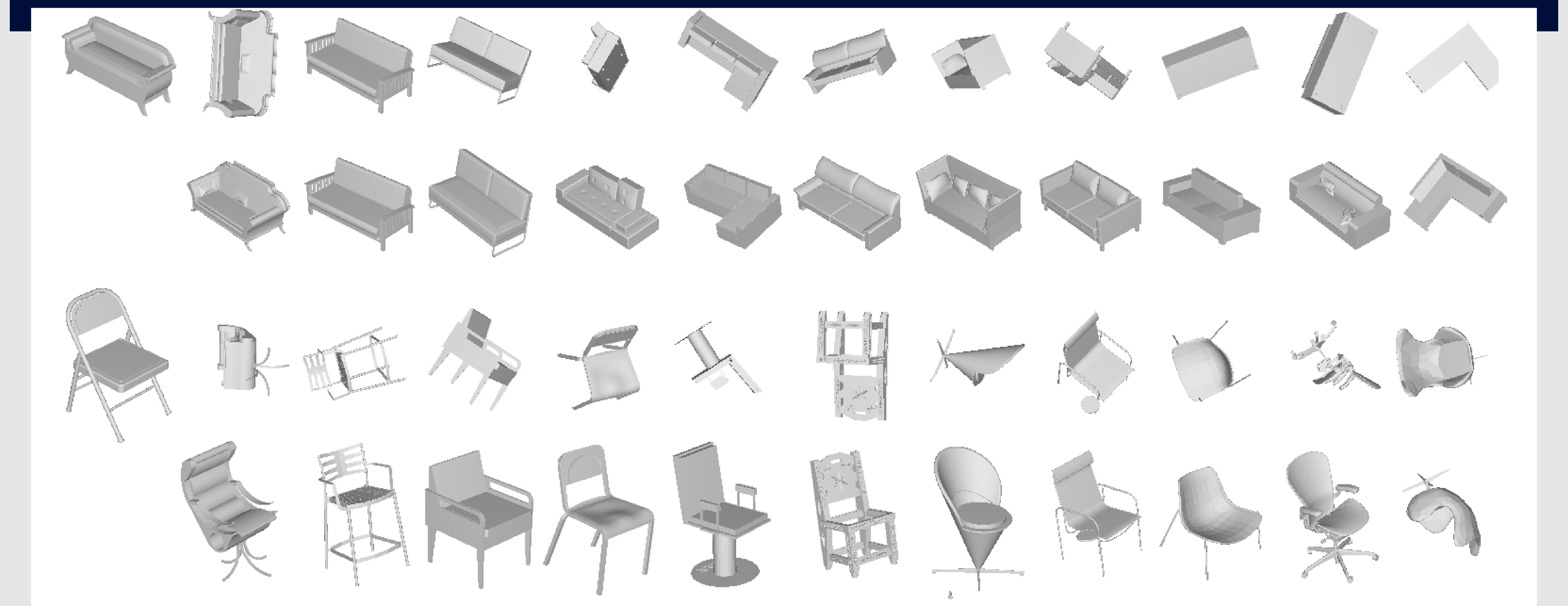
ModelNet40 shape classification

We outperform all other methods under arbitrary orientations.

Method	z/z	SO3/SO3	z/SO3	params	inp. size
PointNet[1]	89.2	83.6	14.7	3.5M	2048 x 3
PointNet++[2]	89.3	85.0	28.6	1.7M	1024 x 3
VoxNet[3]	83.0	73.0	-	0.9M	30^3
SubVolSup[4]	88.5	82.7	36.6	17M	30^3
SubVolSup MO[4]	89.5	85.0	45.5	17M	20×30^3
MVCNN[5] 12x	89.5	77.6	70.1	99M	12×224^2
MVCNN[5] 80x	**90.2**	86.0	-	99M	80×224^2
RotationNet[6] 20x	**92.4**	80.0	20.2	58.9M	20×224^2
Ours	88.9	**86.9**	**78.6**	**0.5M**	$\mathbf{2 \times 64^2}$

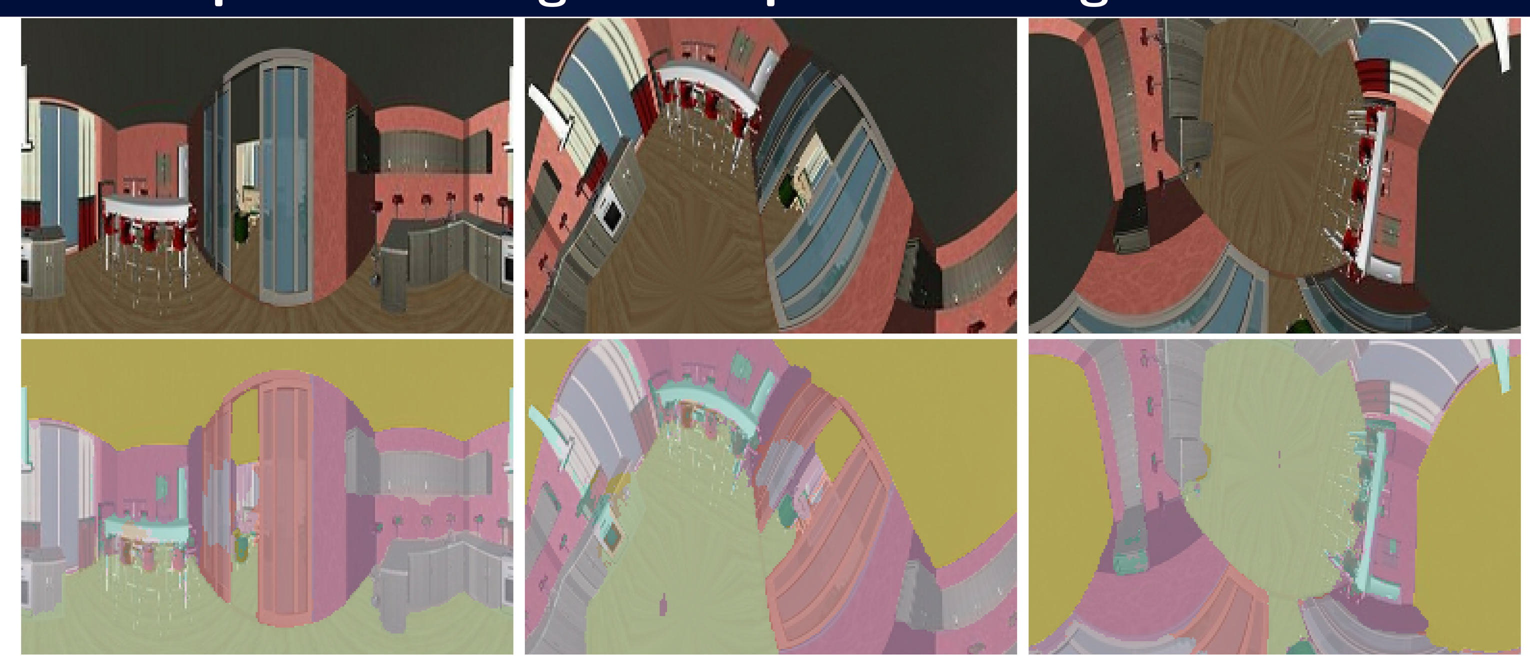
[1]Qi, Su *et al.* CVPR'17, [2]Qi *et al.* NIPS'17, [3]Maturana *et al.* IROS'15, [4]Qi, Su *et al.* CVPR'16, [5]Su *et al.* ICCV'15, [6]Kanezaki *et al.* CVPR'18

Shape alignment



Alignment using spherical correlation of feature maps.

Spherical hourglass for panorama segmentation



Results on SunCG (Song *et al.* CVPR'17)

Conclusion

- Spherical convolutions and localized filters (enforced by smooth spectrum) bring good performance and scalability.
- We match or surpass competing methods with orders of magnitude fewer parameters under SO(3) perturbations.