Polar Transformer Networks

Carlos Esteves, Christine Allen-Blanchette, Xiaowei Zhou, Kostas Daniilidis

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

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

Introduction

- We consider the problem of image classification under arbitrary orientation and scale.
- Our CNN architecture learns an image representation invariant to translation and equivariant to rotation and dilation.
- Main contribution is the polar transformer module, which performs a differentiable log-polar transform, where the transform origin is a latent variable.

Log-polar properties

- Rotations around the origin become vertical shifts, and dilations around the origin become horizontal shifts

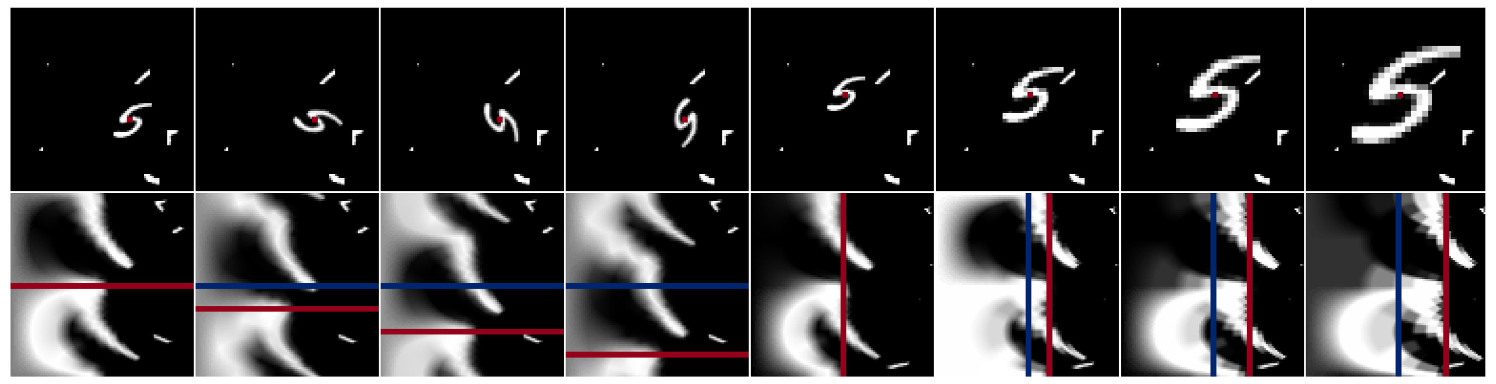

Figure 1: Distance between lines is proportional to rotation angle/scale

Method

- Fully convolutional polar origin predictor outputs a heatmap.
- Heatmap centroid (two coordinates) and input image go into the polar transformer module, which performs a polar transform around given origin coordinates.
- Obtained polar representation is invariant to translation.
- Rotations and dilations become shifts, which are handled equivariantly by a conventional classifier CNN.

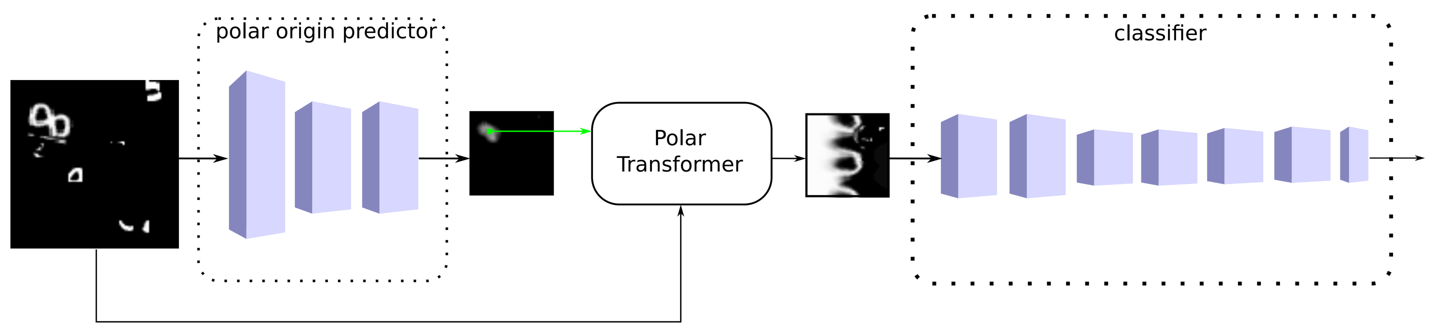

Figure 2: Network architecture

- Differentiable sampling from Spatial Transformer Networks (Jaderberg et al. [2015]) combined with log-polar transform.
- Wrap-around padding to account for angle periodicity.

MNIST variants

Table 1: Rotated MNIST

Model	err %	pars.	time
PTN-B+	1.14	129k	4.38s
PTN-B++	0.95	129k	4.38s
PTN-C-B+	1.01	254k	7.36s
PTN-C-B++	**0.89**	254k	7.36s
PCNN-B+	1.37	124k	3.30s
CCNN-B+	1.53	124k	2.98s
STN-B+	1.31	146k	4.57s
OR-TIPooling [1]	1.54	1M	-
TI-Pooling [2]	1.2	1M	42.9s
RotEqNet [3]	1.01	100k	-

Table 2: SIM2MNIST

Model	err %	pars.	time
PTN-S+	5.44	35k	11.92s
PTN-B+	**5.03**	134k	12.02s
PCNN-B+	15.46	129k	5.33s
CCNN-B+	11.73	129k	5.28s
STN-B+	12.35	150k	10.41s
HNet [4]	9.28	44k	31.42s

[1] Zhou et al. [2017]
[2] Laptev et al. [2016]
[3] Marcos et al. [2016]
[4] Worrall et al. [2016]

- State of the art on rot. MNIST and SIM2MNIST at submission.
- Compare PTN with PCNN (polar transform with fixed origin) to verify the advantages of learning the origin.
- Compare with STN (Spatial Transformer) to verify the advantages over regressing all transformation parameters.

Feature map visualization

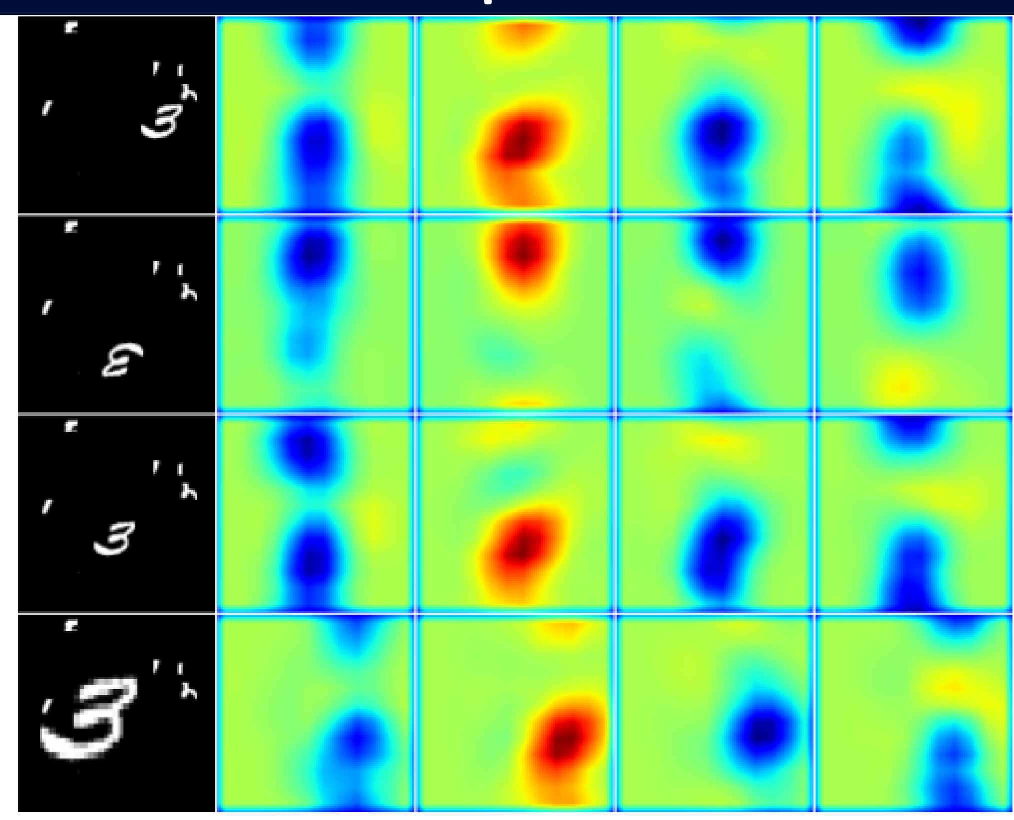

Figure 3: Feature maps on the last convolutional layer.

- 1st and 2nd rows: 180° rotation results in a half-height vertical shift.
- 3rd and 4th rows: dilation results in a right shift.
- 1st and 3rd rows show invariance to translation.

Rotated Street View House Numbers (SVHN)

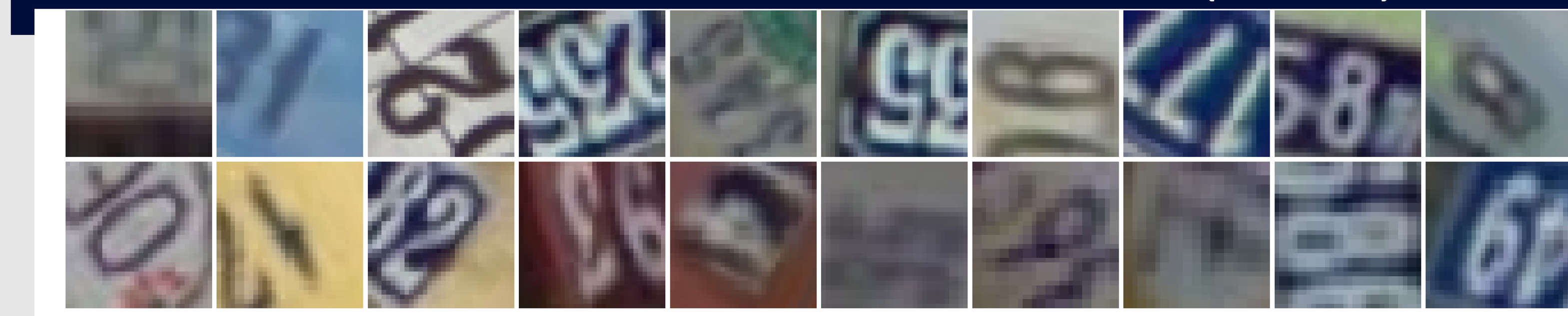

Figure 4: ROTSVHN samples.

Table 3: SVHN classification error. Minus suffix: 6 and 9 removed.

	SVHN	ROTSVHN	SVHN-	ROTSVHN-
PTN-ResNet32 (Ours)	2.82%	**7.90%**	2.85%	**3.96%**
ResNet32	**2.25%**	9.83%	**2.09%**	5.39%

Cylindrical Transformer Network and ModelNet40

- Axis of a cylindrical transform is learned by fixing the orientation, slicing subspace along it in channels and convolving to obtain a heatmap; the centroid determine the axis.
- Channel-wise polar transform is then applied.
- Representation is equivariant to rotations around the family of parallel axis determined by input orientation.

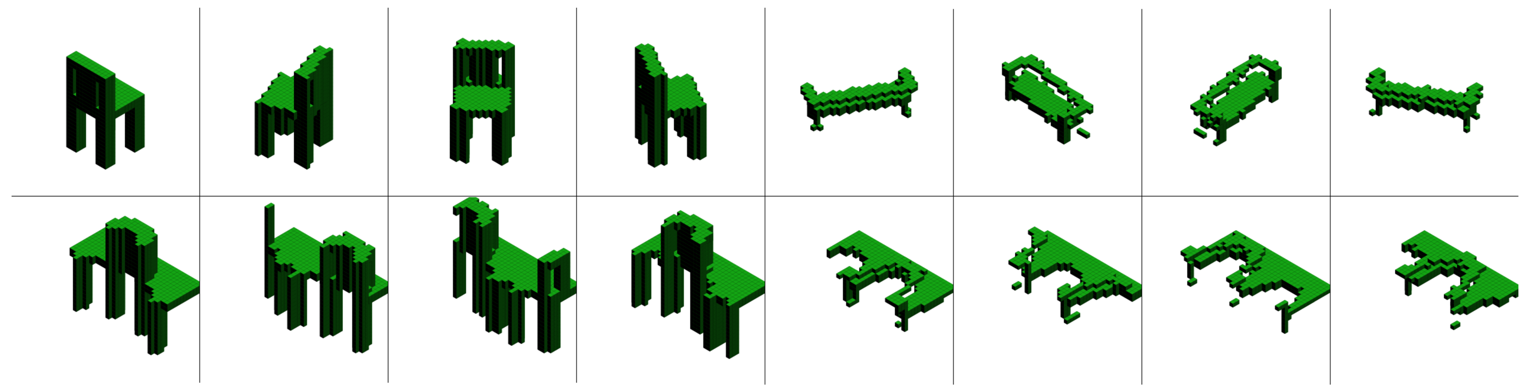

Figure 5: Occupancy grids and corresponding cylindrical representations.

Table 4: ModelNet40 classification (only voxel-based methods).

Model	class acc. %	inst. acc. %
Cylindrical Transformer (Ours)	**86.5**	**89.9**
3D ShapeNets (Wu et al. [2015])	77.3	-
VoxNet (Maturana and Scherer [2015])	83	-
MO-SubvolumeSup (Qi et al. [2016])	86.0	89.2
MO-Aniprobing (Qi et al. [2016])	85.6	**89.9**

Conclusion and final remarks

- Equivariant representations allow high accuracy with fewer parameters, faster training time and less data augmentation.
- Refer to the paper for theoretical background on group convolutions and canonical coordinates.
- Check out our work on SO(3) equivariance with Spherical CNNs.