3D Conv Nets for Robotic Manipulation

Chad DeChant, Joaquín Ruales, Jake Varley

Outline

- 3D Spatial CNNs
 - Extended 2D to 3D
- Grasping
 - Heatmap Approach
 - Shape Completion Approach
 - Comparison

2D Conv Filters





*

0

0

0

x filter

-1

-2

-1







*



3D Conv Filters



*

x filter



y filter



*

3D Filters in Theano

- We Implemented the following for the Keras Theano library
 - 3DConvolution
 - Theano conv3d2d (BZCXY format)
 - 3DMax pooling
 - Created max_pool_3d using max_pool_2d



3D Spatial CNN Test Data

- Random size geometric shapes







sphere axis-aligned cube rotated cube



Trained Layer 0 Filters



1 2 3 4 5 6 7 8

8 5x5x5 filters

Grasping Heatmaps



Grasping Heatmaps











Finger 3



Grasping Heatmaps

Patch: 32x32x32 Label: 1x1x1x nheatmaps





Trained Layer 0 Filters

常信		1.1		192	18				322			312		RIZ.							15	1.2						99.		的位
		12121	42.85	17 - 4. 20 - 4 - 4	17.00		1015	Million .			10.11		122	10.23	1.1.1		1000 C	10.12 30 10		10.00	TTP:	155	1	1100			1	2223	4.1	
A.	i.	100	4.5	30	100	and a second sec		投资	E C	医2				15.2	12/2	修行	ALC: N	1.15	1	22.11		1.1	1.5	a a a a	時這	1	6.15		12	
	1		1.1	22		38		國語		區的	1				2.5			194				E	2.5	間包			20 A			
2011							23.	1000	1000	100.00		1080 1011-10		12213			31.14	1000	10.00	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1			1000			10.00	100.0			
10	1.1	S	12.5	100		1.1					1	815				1		1.1			1.15	1213	1	02		11	100			100
24		88) 80)	1944											54	61			1.1					1					源		
			30	101	and										1.4.1	1			18	21			52					Щ.		1.1
		133						a trans			10.004								1			C.C.		111		2.5				1.14
		100	1.1.1	1977 I 1	1.1.1	2.0.0	100	100	12.1	1.00	4175	-	1000	1,25	Section of	1.1	10000 10007	14		The Local	1.24	And all a	1,75	133C-1		R. 43	21115		100	1.0
1		15	117	21	6.5		0.4														1.2	5	12				A Parts		392	164
						ġ.,					1.14			1.10	1					No. Carlos					22.5		1.0			
			Period Da					1.11	12.20						10133			14				CIR.	100	115				1.5		1.11
			10.00	113	1.08			See.		dian.						1.			2.14		12	20	Sec. alex	100	20	T		22	10.14	
					13								内容		11						自己	1			1					
	Cy and	2 13						12			1.74		80			100			1			1								
	- 7-	1.1.1	CRE				.36		1										100	265		15	the second		Carlos Para					1
		1000	1.00			110.0	2.25	17.100	1.00	1000			111.0	L FILLE	1	1.124		<u>1919</u>		Store Speci		Alexandra a	100	100	12142		200		Sec. 12	251
- 5	Sale P				1.1	12/5	1		20	4	1.54				ALC: N	1	1.1			La state					10.2				100	
	10					1	1.3		100					and and	5				1.20		and a state		1	1		1.4				and the
	1.46	0.5	-115																							t.				

1 16x16x16 filter

Another Approach...

Shape Completion

Complete the shape, then plan grasps on it



Shape Completion

Is shape completion feasible?Use MNIST









Reconstruction layer

Hidden layer 150 or 1000 units

Convolutional layers (70 3x3x3)

Convolutional layers (30 3x3x3)

INPUT

Cost function





Reconstruction layer

- One reconstruction output unit for each desired pixel
- Correct output binarized
- Sigmoid units encouraged to be close to either 0 or 1

```
def cross_entropy_error(self, y):
cost = - T.sum(y * T.log(self.output) + (1 - y) * T.log(1 - self.output), axis=1)
return T.mean(cost)
```

Cost function

Target output











3D digit completion

MNIST in three dimensions









Target Output



Model Output



















Training Data

- Household objects, Modelnet, BigBird









Shape Completion Data



Voxelization





3D Shape completion



Model Input

Model Output

Target Output

3D Shape completion



Model Input

Model Output

Target Output

Grasp Comparisons

Heatmap vs Shape Completion Grasps

1) generate grasps



2) compare ground truth grasp qualities



Conclusion

- 3D Spatial CNNs
 - Extended 2d to 3d
 - Keras Library
- Grasping
 - Heatmap Approach
 - Shape Completion Approach
 - Comparison

Thank you

