ShapeY: Measuring Shape Recognition Capacity Using Nearest Neighbor Matching

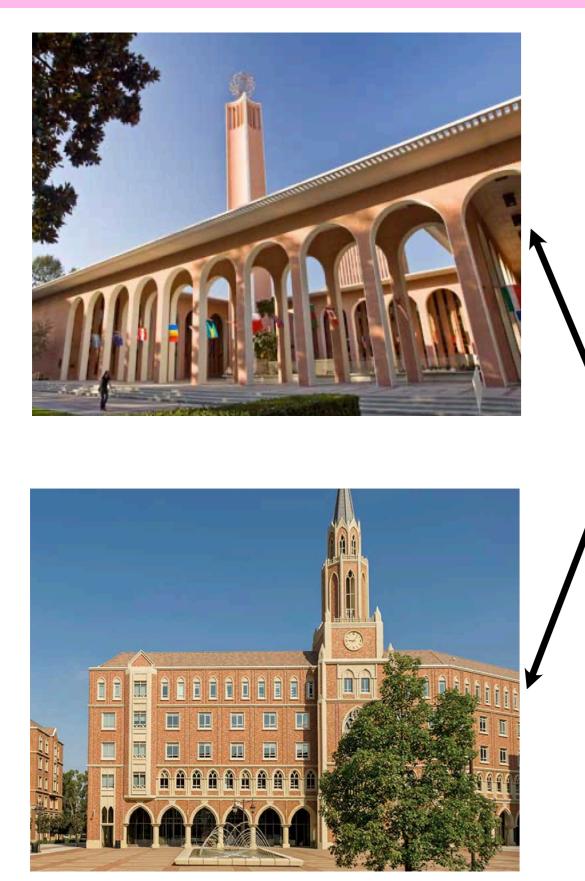
Jong Woo Nam¹, Amanda S. Rios¹, Bartlett W. Mel²

¹ Neuroscience Graduate Program, University of Southern California ² Biomedical Engineering Department, University of Southern California

1. The Core Capability of a Recognizing System: Knowing What Looks Like What

Reference





Question: Which image looks <u>most similar to</u> the reference?

The main idea behind the ShapeY benchmarking system:

For a competent recognizing system, a reference view should match most closely to another view of the same object, regardless of the number of distractor object views.

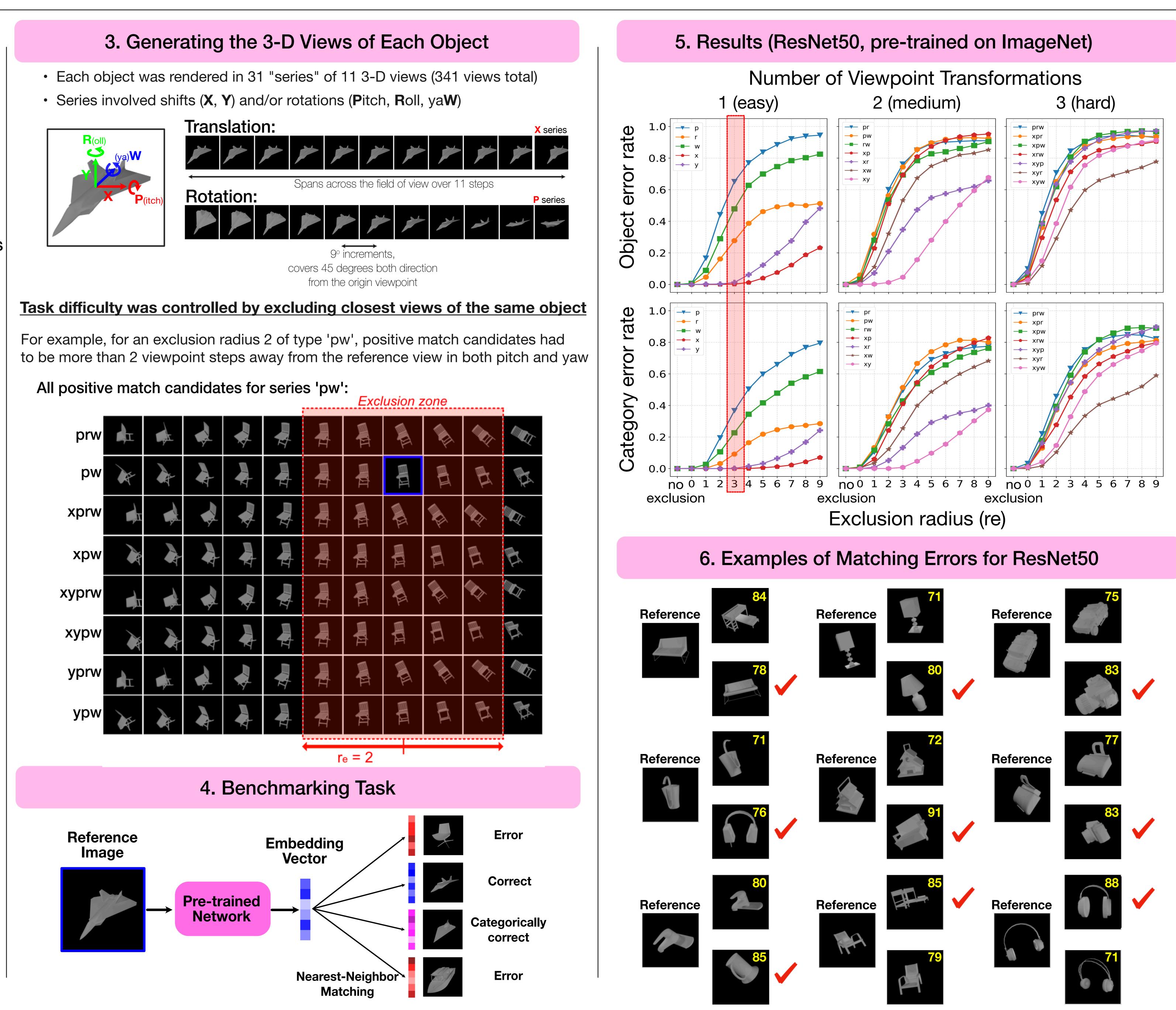
2. The ShapeY Dataset

categories

xample

<i>¥</i> -		A	RA T	\$1		×	5		6	W.	Ý		жқ.	Ť	,	Y		1
×.				-			$\boldsymbol{\mathcal{F}}$	Ĩ	\bigcirc		ġ.		19	and and	4	$\not\vdash$	-	
		Ņ					2	A A		1	Natio		A	No.		/	-	Ø
*		Ø		-		H	ñ			۲	ø			¥	-	/	,	
*		ð		-		T.	6	×	0	•	ð	0		藩	*	*		1
*	0	M					1	1	\mathbf{S}	-	all a	4		ALL		×	-	×
*		ø	-	2			1	1	5		6	1	-	A.	T	ø		st.
¥		I		0		T	ス		6		¥.		4		4	×		1
*				-	Ø	E	Ч	5			S		#	¥	•	1		1
*	$\boldsymbol{\bigtriangleup}$	Ø		6	ð		4		•		3			衜		1		1

- 20 different object categories (Airplane, Chair, Lamp, etc)
- 10 instances in each category
- 341 3-D views of each object
- Total of 68,200 images





<figure>

8. Conclusion

- ShapeY measures shape recognition capacity using simple nearest-neighbor matching in the embedding space
- ResNet50's embedding space is badly tangled for purposes of shape recognition, leading to poor matching performance even with mild 3-D viewpoint transformations
- Other training schemes such as the SimCLRv1 showed even worse performance
- ShapeY currently contains ~1% of the number of shape categories that humans easily master (including children); we expect performance of systems like ResNet to deteriorate to near zero as the number of object categories is increased to human levels. **Find ShapeY here:**

https://github.com/njw0709/ShapeY