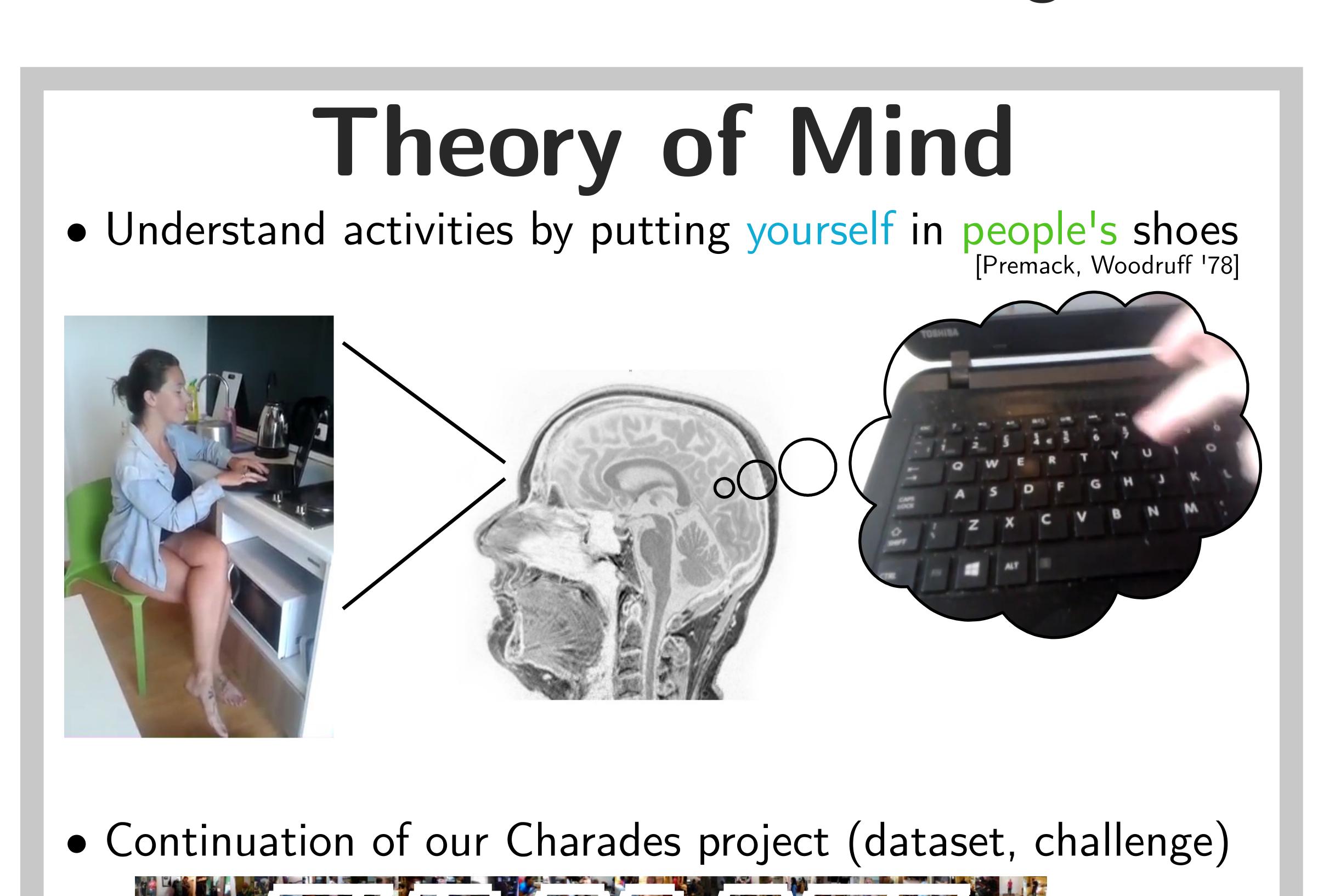


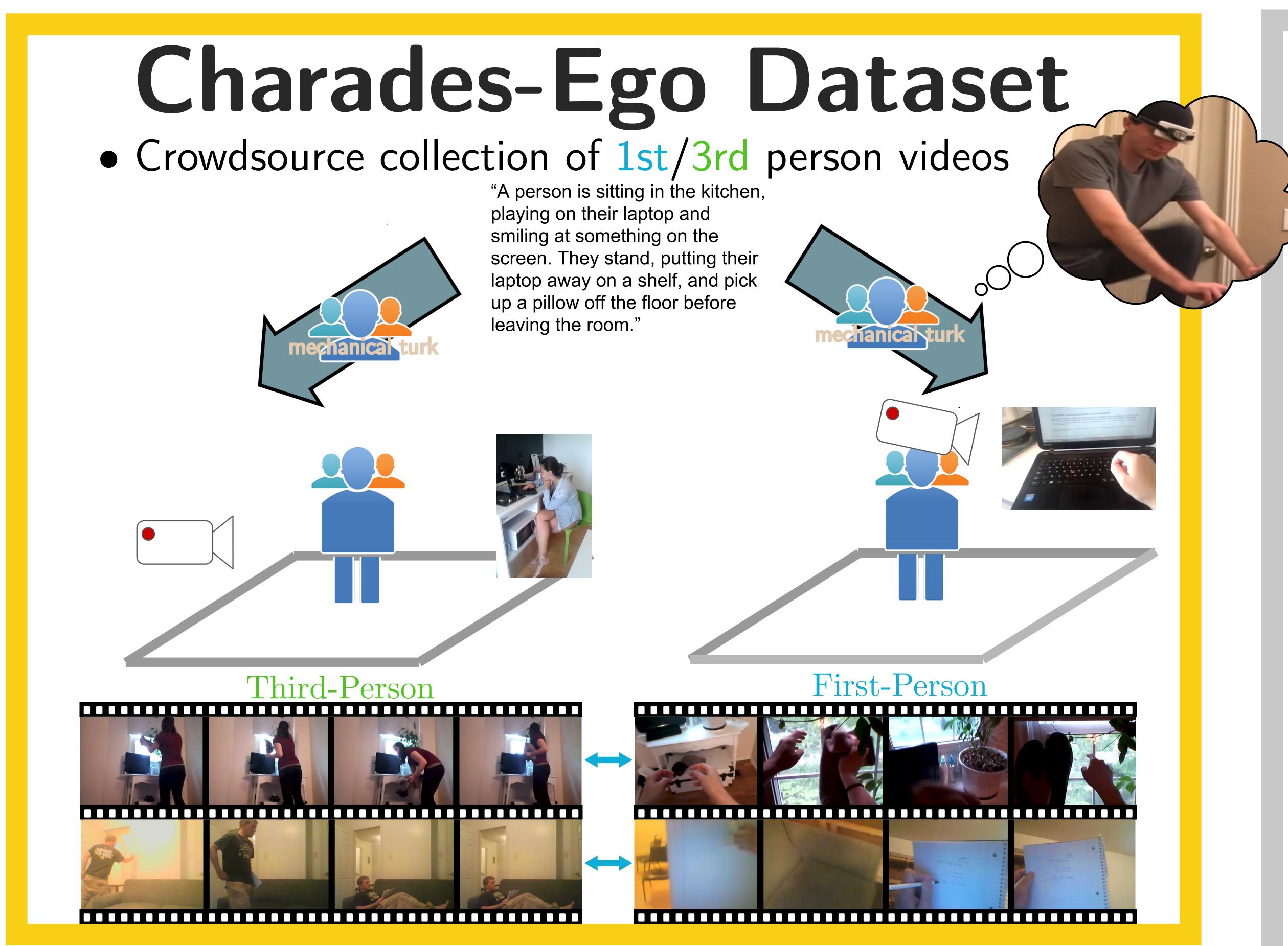
Actor and Observer:

Joint Modeling of First and Third-Person Videos



Gunnar A. Sigurdsson, Abhinav Gupta, Cordelia Schmid, Ali Farhadi, Karteek Alahari



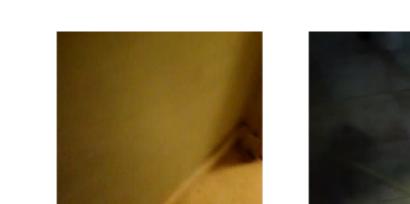


Learn to select data to learn a triplet ranking loss

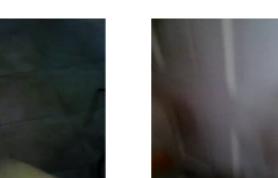
ActorObserverNet

CVPR'17

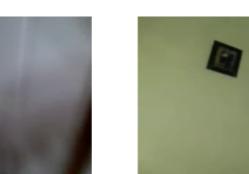
Before learning, many bad views:

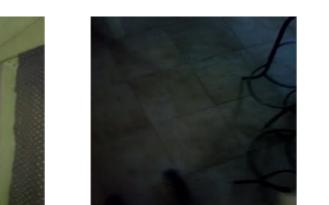
















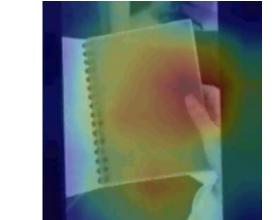


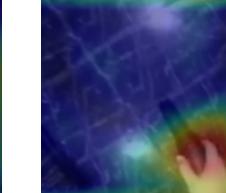


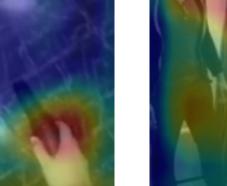








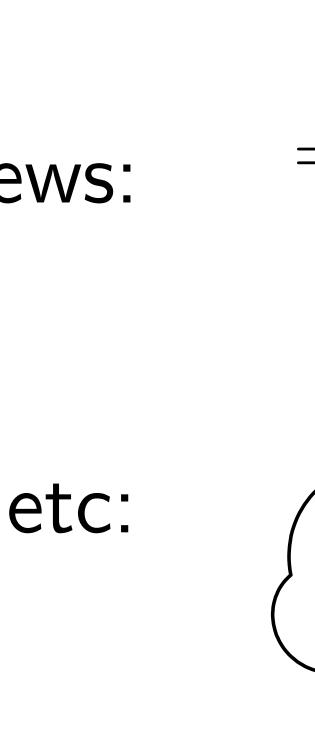


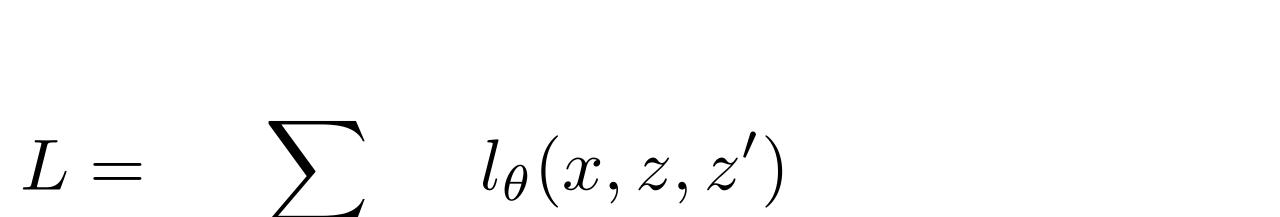


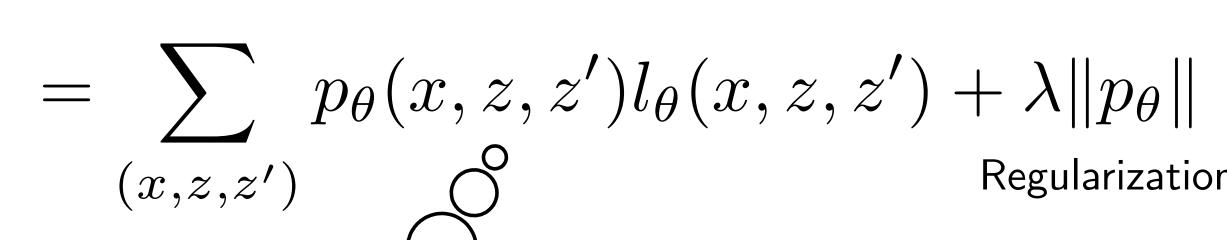


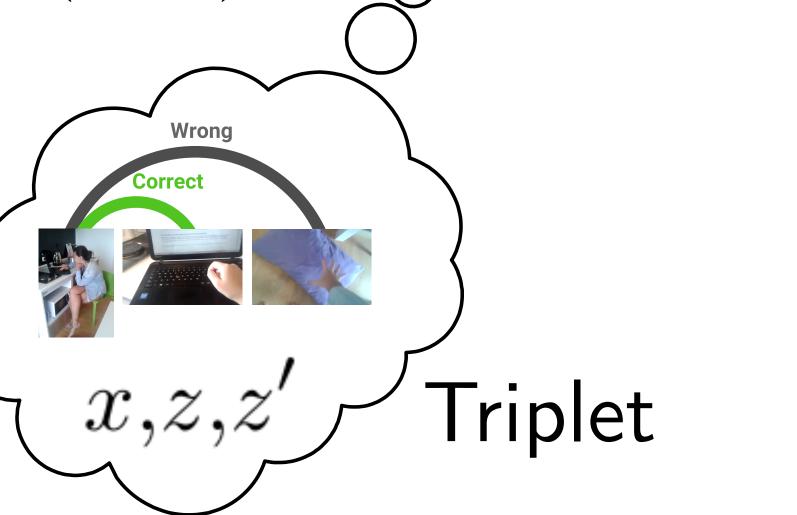








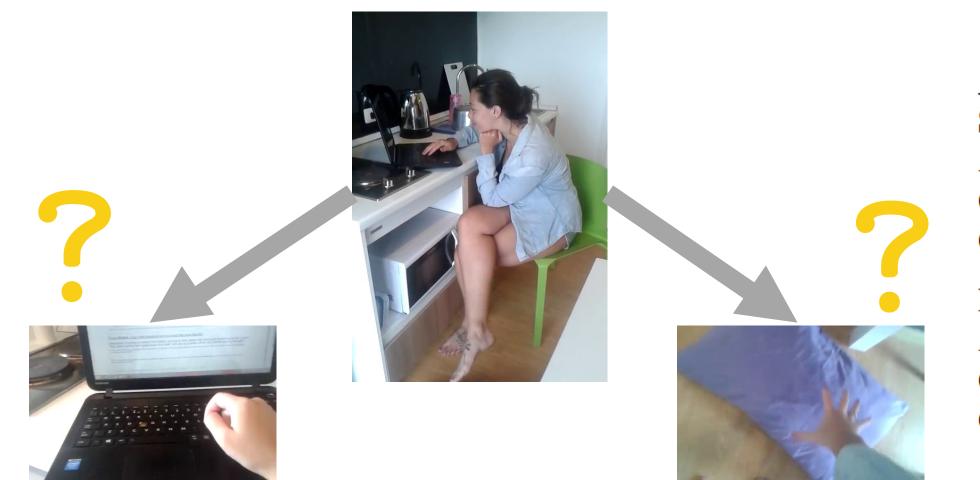




	$L_{ heta}(x,z,z')$	Final Loss		$l_{ heta}(x,z,z')$
	$p_{\theta}(x,z,z')$	X		Triplet Loss
	Video Softmax	VideoSoftmax	VideoSoftmax	
on	FC	FC	FC	
	ResNet-152	ResNet-152	ResNet-152	
	\boldsymbol{x}	z	z'	
	$Third ext{-}person$	$First ext{-}person \ Corresponding$	First-person Non -corresponding	

Experiments

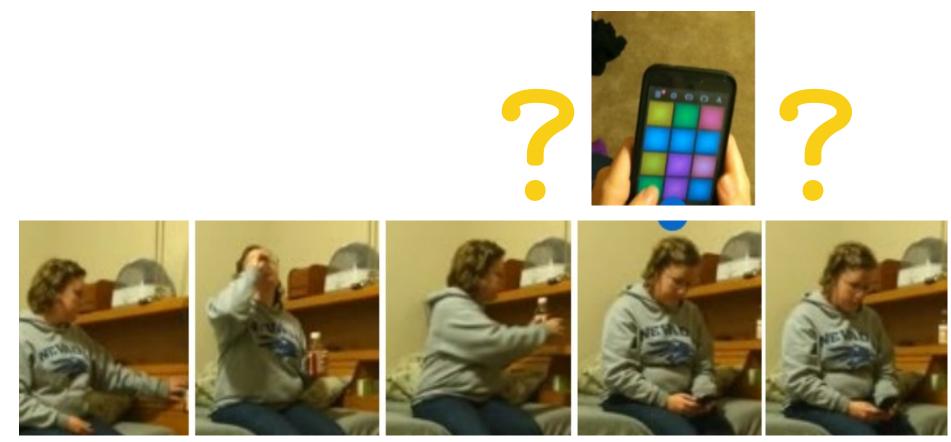
- Retrieve 1st person view given a 3rd person view:
- ullet Use $p_{ heta}$ to rank test samples



	Random	ImageNet ResNet-152	Charades Two-Stream	ActorObserverNe
Same person				
All test data	50.0	53.6	55.5	51.7
Choose 50% of test data	50.0	55.7	60.2	73.9
Choose 10% of test data	50.0	57.9	68.8	97.2
Different persons				
All test data	50.0	50.6	51.7	50.4
Choose 50% of test data	50.0	50.4	51.6	76.3
Choose 10% of test data	50.0	49.6	50.8	98.8

(% Classification Accuracy)

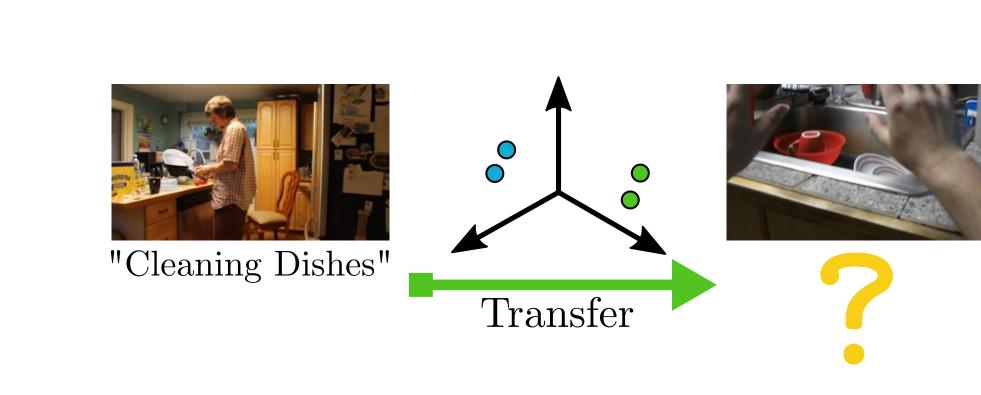
• Align a 1sec 1st person video to a 3rd person video:



	Random Chance	Human	ImageNet ResNet-152	Charades Two-Stream	ActorObserverNe
Same person	11.0	1.3	8.3	6.5	5.2
Different manageme	11.0	1.2	0.7	7.0	<i>(</i> 1

(Alignment Error in seconds)

• 3rd person supervision + 1st person zero-shot recognition:



	First-Person Test
Random	7.1
Zero-Shot	
VGG-16 Charades	15.6
ResNet-152 Charades	19.5
ResNet-152 Transfer	22.2
ActorObserverNet	25.2
Supervised on Charades-Ego	0
ResNet-152 Charades-Ego	28.2

- 8000 videos available! 3rd and 1st person
- 30 second videos 157 Charades classes
- Same sequence, same or different people!

