Assessing the Robustness of Neuro-Symbolic Modelling

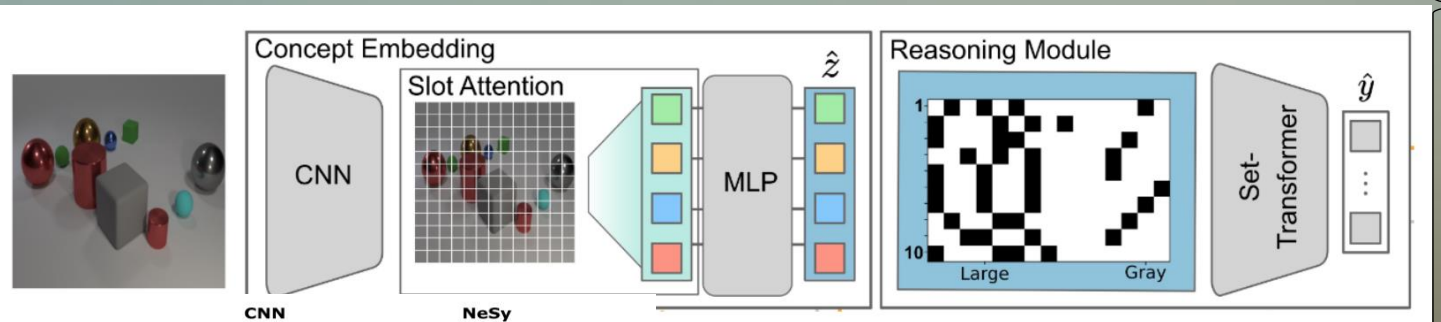
Jingxuan Chen (Shino) – MSc in AI

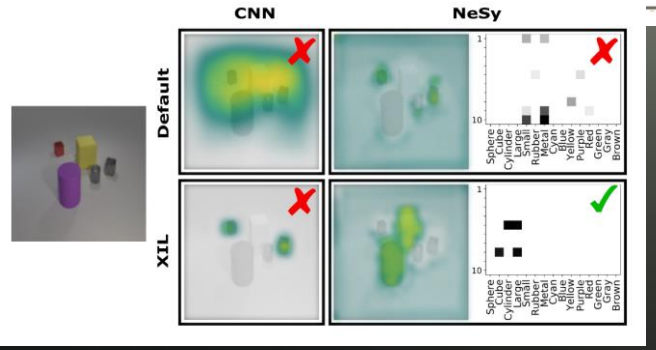
Xuelong An – MA in Cognitive Science

Paolo Cassina – MA in Cognitive Science

1. Model comparison is not comprehensive

Model	Validation (confounded)	Test (non-confounded)
CLEVR-Hans3		
CNN (Default)	99.55 ± 0.10	70.34 ± 0.30
CNN (XIL)	99.69 ± 0.08	70.77 ± 0.37
NeSy (Default)	98.55 ± 0.27	○ $\mathbf{81.71 \pm 3.09}$
NeSy XIL	100.00 ± 0.00	• $\mathbf{91.31 \pm 3.13}$





Siri are you better than Alexa

Tap to Edit

That's like comparing apples and... not apples.


Now it's your turn!
Deep learning
Hurricane Harvey
Pre-disaster
Post-disaster
Labeled imagery (ground truth)
Damage assessment from xView 2 model
Santa Rosa Wildfire
Pre-disaster
Post-disaster
Labeled imagery (ground truth)
Damage assessment from xView 2 model
Black Box
Explainable AI
Output
Input
NeSy
0 1 2 3 4
5 6 7 8 9
#114
SRGRAFO

Methodology

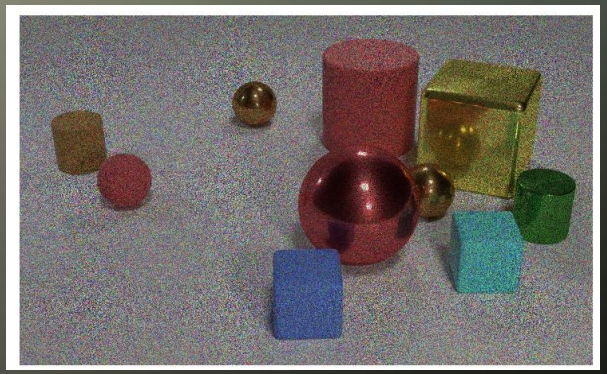
Models

- 1. ResNet18 (pretrained)
- 2. Perceptual ResNet18: ResNet18 + Set Transformer
- 3. Reasoning ResNet18: Slot Attention + ResNet18
- 4. Concept Learner: Slot Attention + Set Transformer

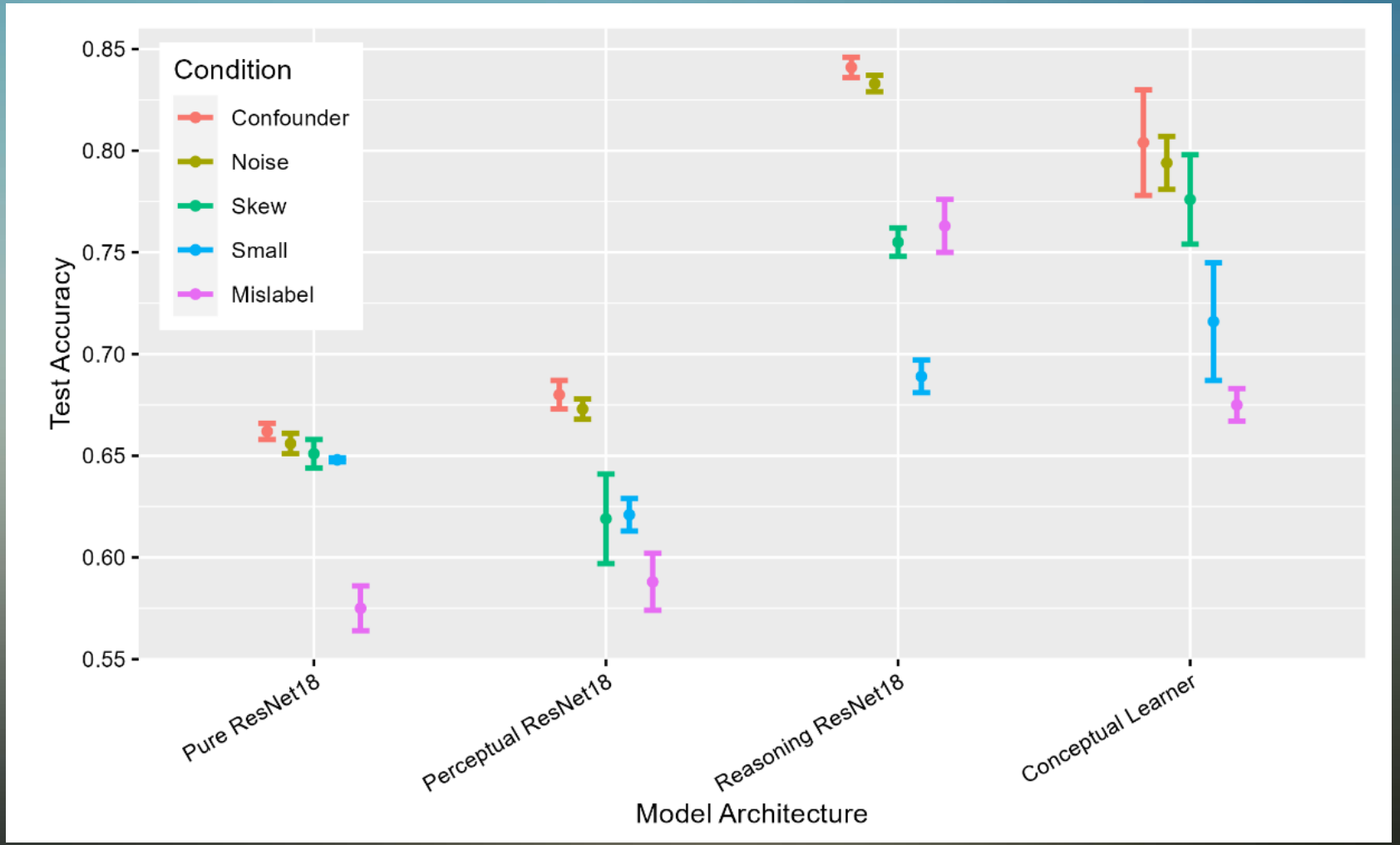
Model Architecture			Name	Number of Parameters	
Slot Attention	ResNet18	Set Transformer		Overall	Trainable
	✓		Pure ResNet18	11.2M	11.2M
	✓	✓	Perceptual ResNet18	11.4M	11.4M
✓	✓		Reasoning ResNet18	11.6M	11.2M
✓		✓	Concept Learner	539K	158K

Data Conditions

- Confounder base condition
- Noise to 1/3 of images
- Class 1 Skewness
- Overall small training data by keeping 1/3 of all images
- Mislabels by shifting label to 1/3 of images
- Robustness as a comparative metric between **test accuracies between models**



Results & Discussion



Detailed table of results

Model Architecture	Confounder (Base Condition)			Noise (1/3)			Skew (1/3 of Class 1)			Small (1/3)			Mislabel (1/3)		
	Train Acc	Val Acc	Test Acc	Train Acc	Val Acc	Test Acc	Train Acc	Val Acc	Test Acc	Train Acc	Val Acc	Test Acc	Train Acc	Val Acc	Test Acc
Pure ResNet18	1±0	0.971±0.003	0.662±0.004	1±0	0.968±0.003	0.656±0.005	1±0	0.954±0.004	0.651±0.007	1±0	0.947±0.013	0.648±0.01	1±0	0.819±0.03	0.575±0.011
Perceptual ResNet1	1±0	0.968±0.005	0.68±0.007	1±0	0.96±0.004	0.673±0.005	0.957±0.055	0.894±0.043	0.619±0.022	0.986±0.015	0.868±0.007	0.621±0.008	0.809±0.067	0.833±0.017	0.588±0.014
Reasoning ResNet1	0.928±0.003	0.925±0.005	**0.841±0.005**	0.925±0.003	0.927±0.007	**0.833±0.004**	0.882±0.005	0.865±0.002	0.775±0.007	0.807±0.004	0.79±0.008	0.689±0.008	0.733±0.004	0.873±0.007	**0.763±0.013**
Concept Learner	0.984±0.004	0.98±0.006	0.804±0.026	0.983±0.002	0.98±0.005	0.794±0.013	0.977±0.005	0.973±0.004	**0.776±0.022**	0.968±0.005	0.962±0.006	**0.716±0.029**	0.784±0.005	0.958±0.003	0.675±0.008

Discussion

- Our results corroborate to growing interest of NeSy: it is robust to different dataset complications
- Fully connected layers arguably can’t learn to reason, but only identify statistical correlation