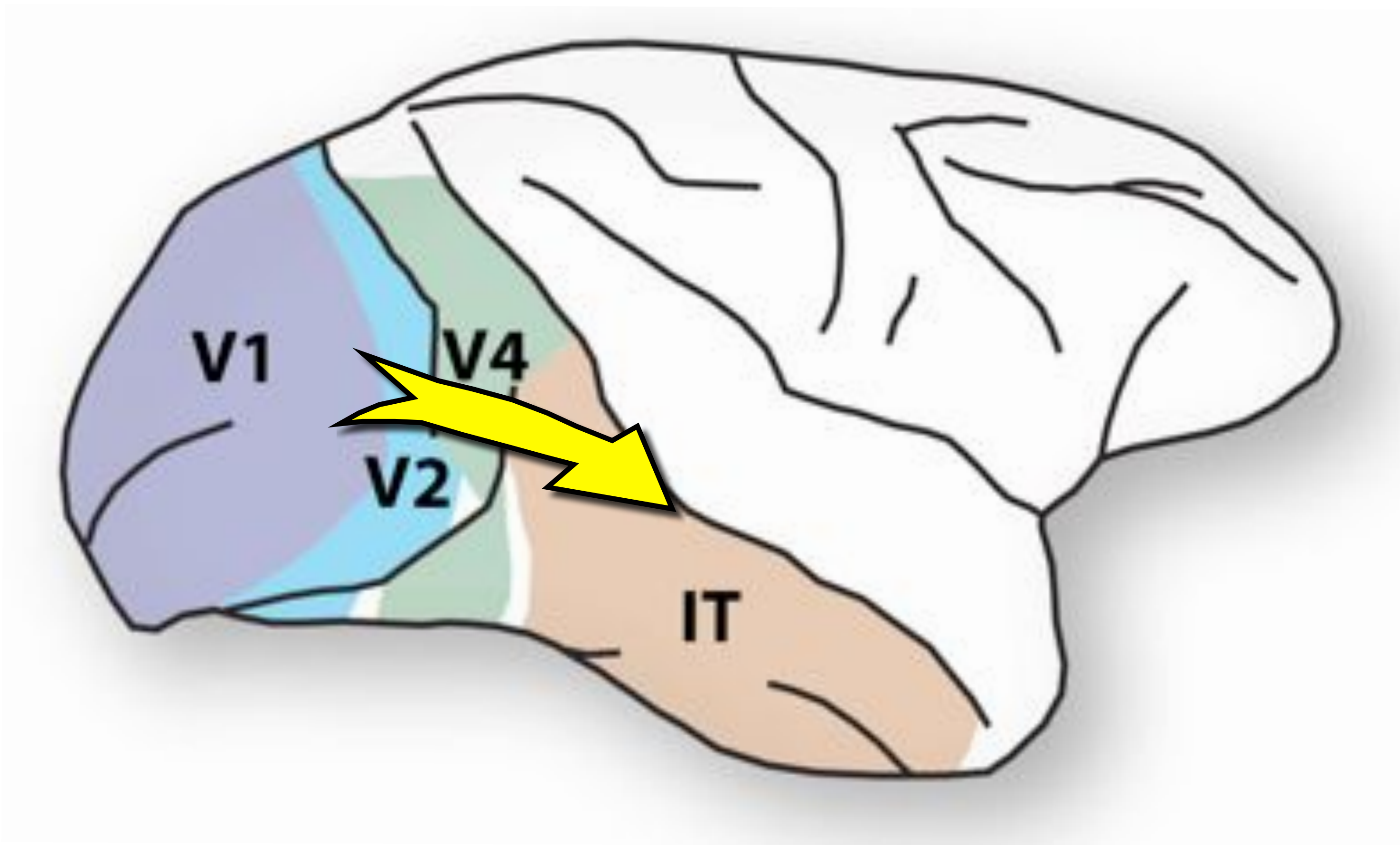
Focus technique: Outils avancés de Computer Vision

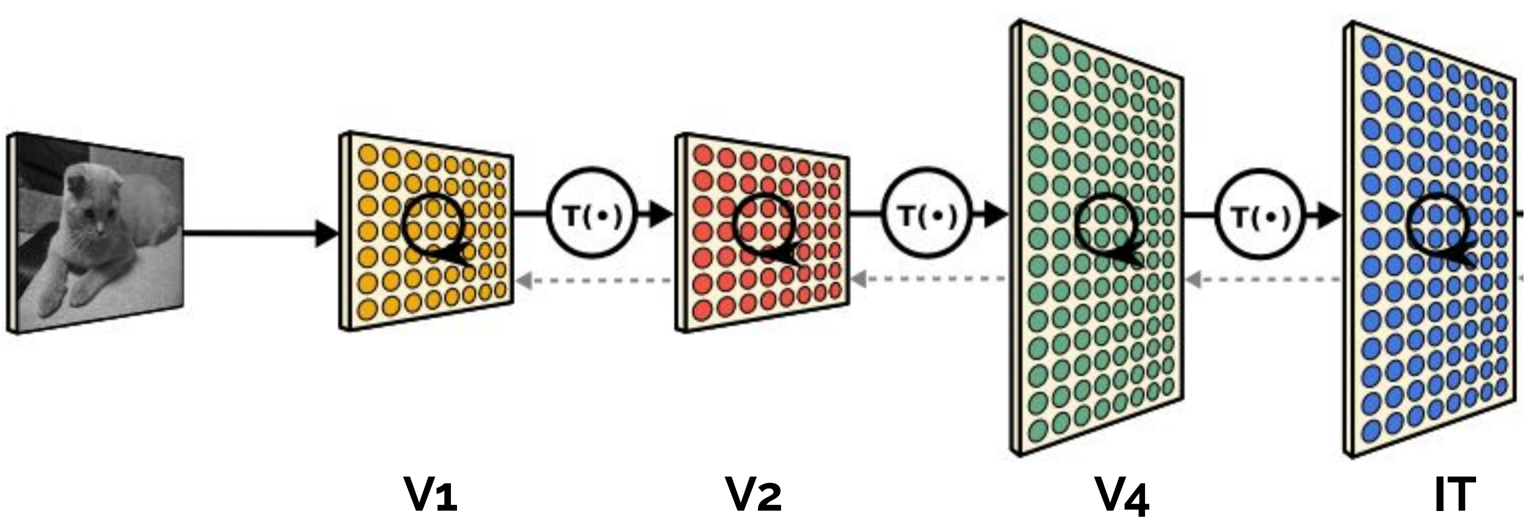
Thomas Serre and Sabine Muzellec

ANITI ANR-3IA Artificial and Natural Intelligence Toulouse Institute

Brown University
Cognitive Linguistic & Psychological Sciences
Carney Institute for Brain Science

Feedback mechanisms

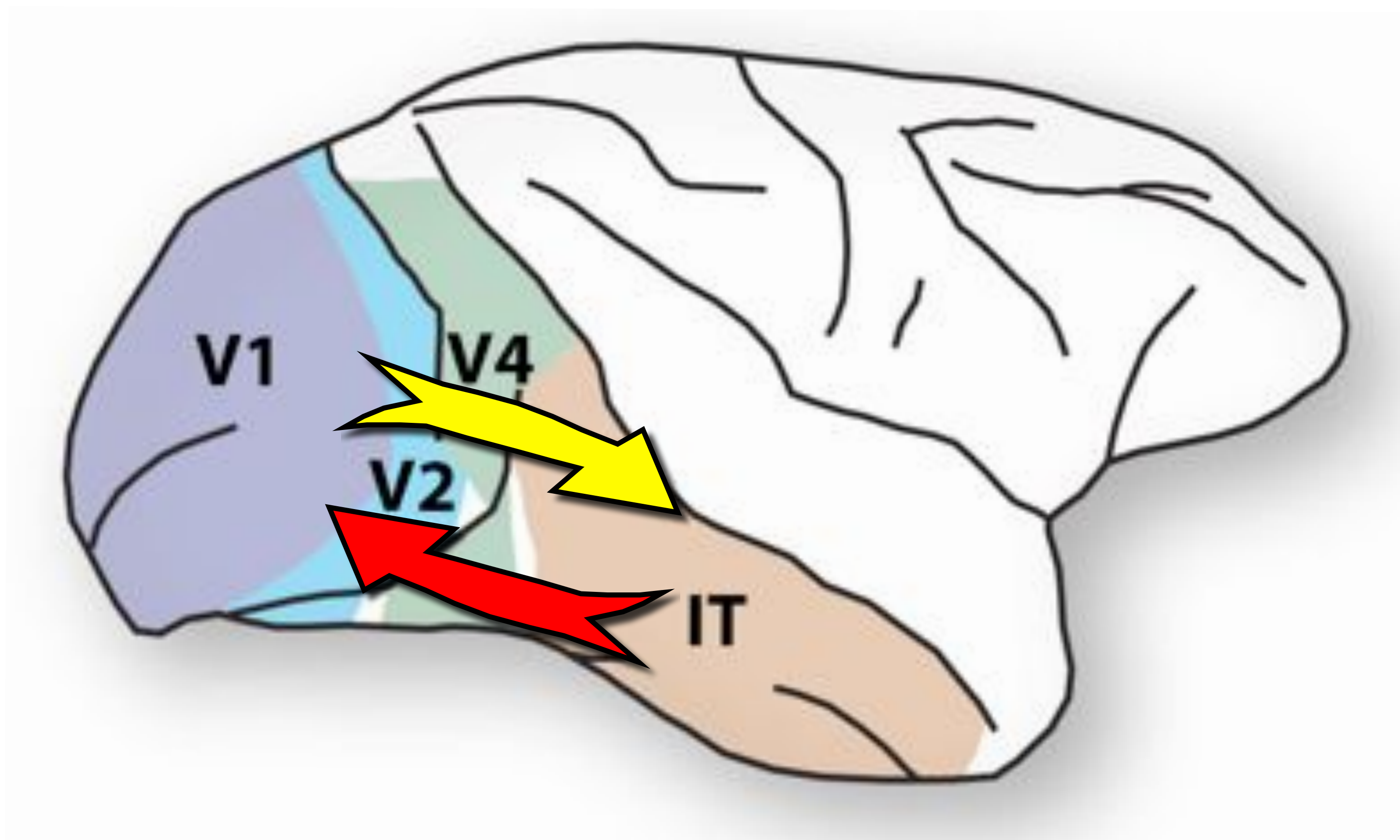


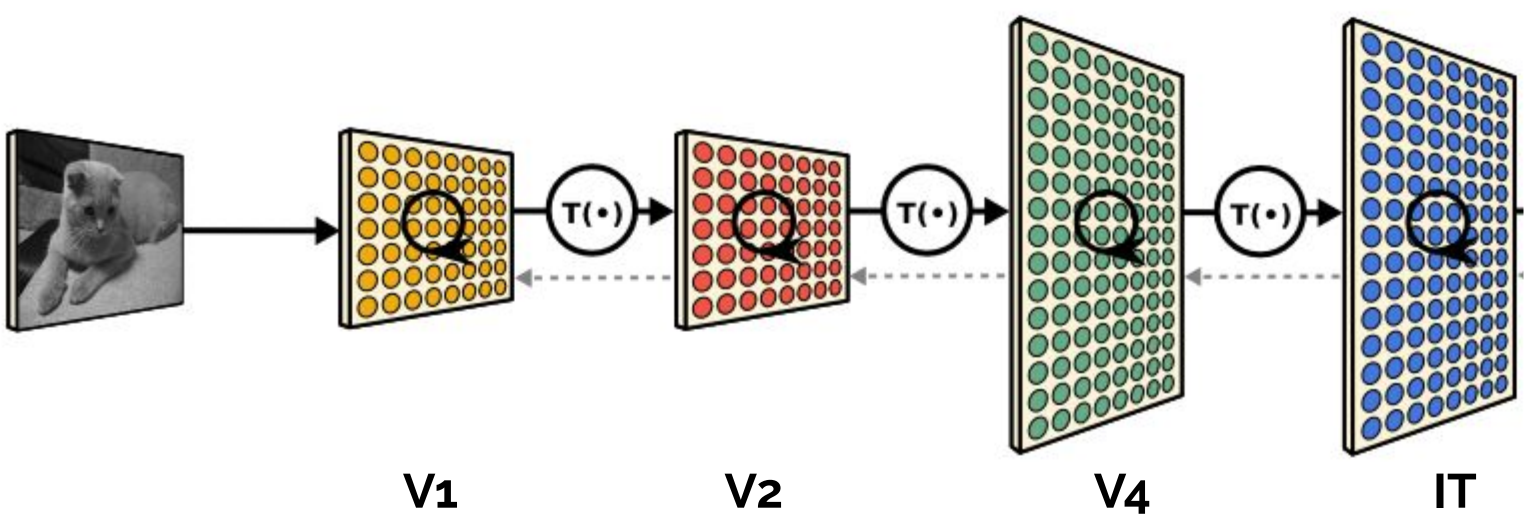


- Deep feedforward nets (CNNs) have become orders of magnitude deeper than our visual system
- Limited generalization beyond the training data

source: https://twitter.com/tserre/status/1400579406083596296

Feedback
mechanisms

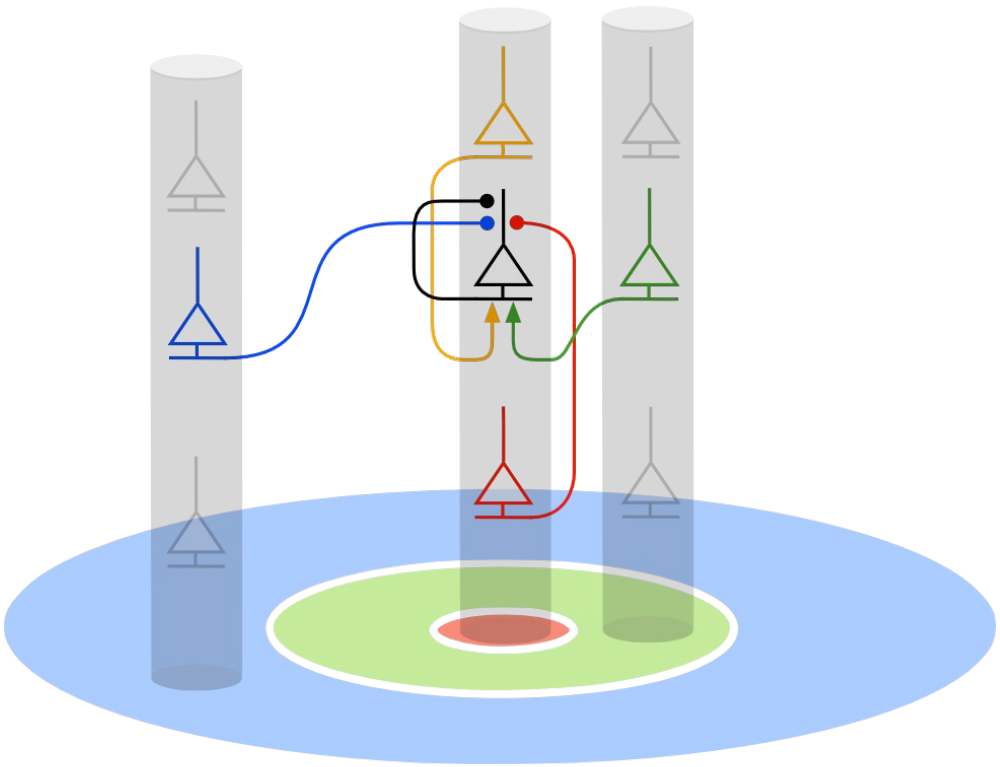




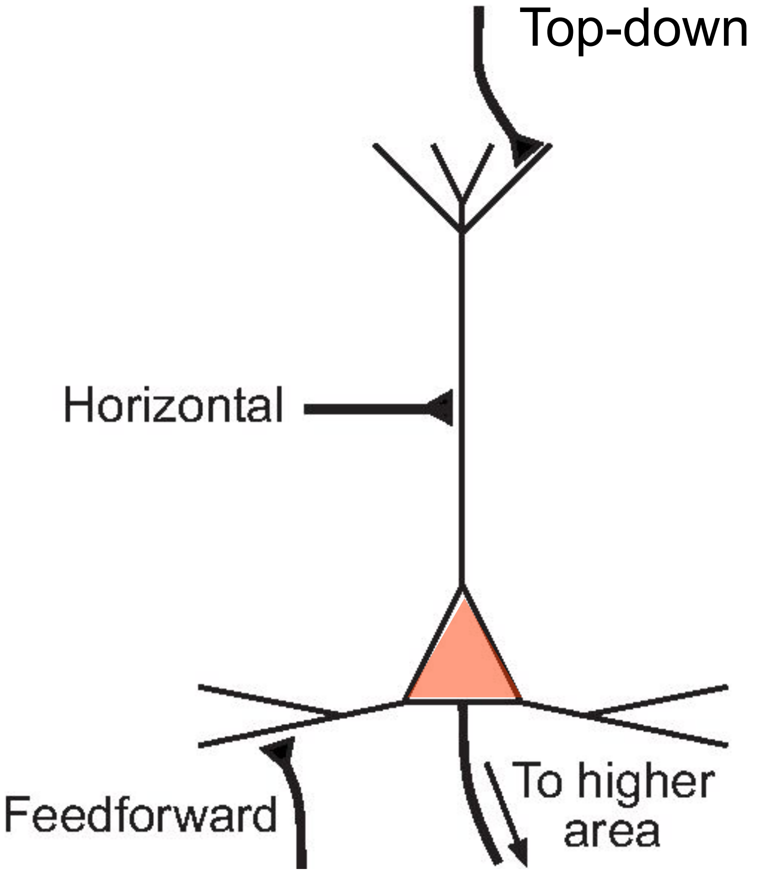
- Feedback plays key role in visual perception beyond feedforward sweep
- Visual processing builds “depth” dynamically through recurrent connections

Kreiman & Serre Year in Cognitive Neurosci. 2020

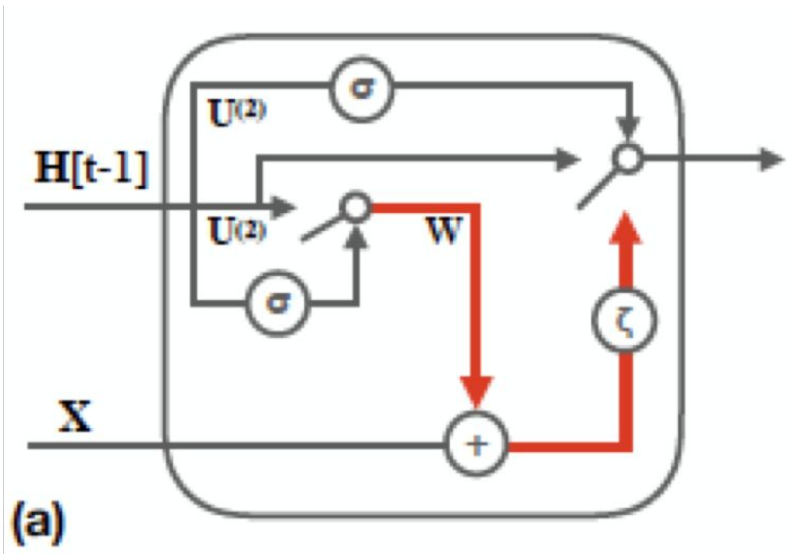
Recurrent neural circuit model(s)

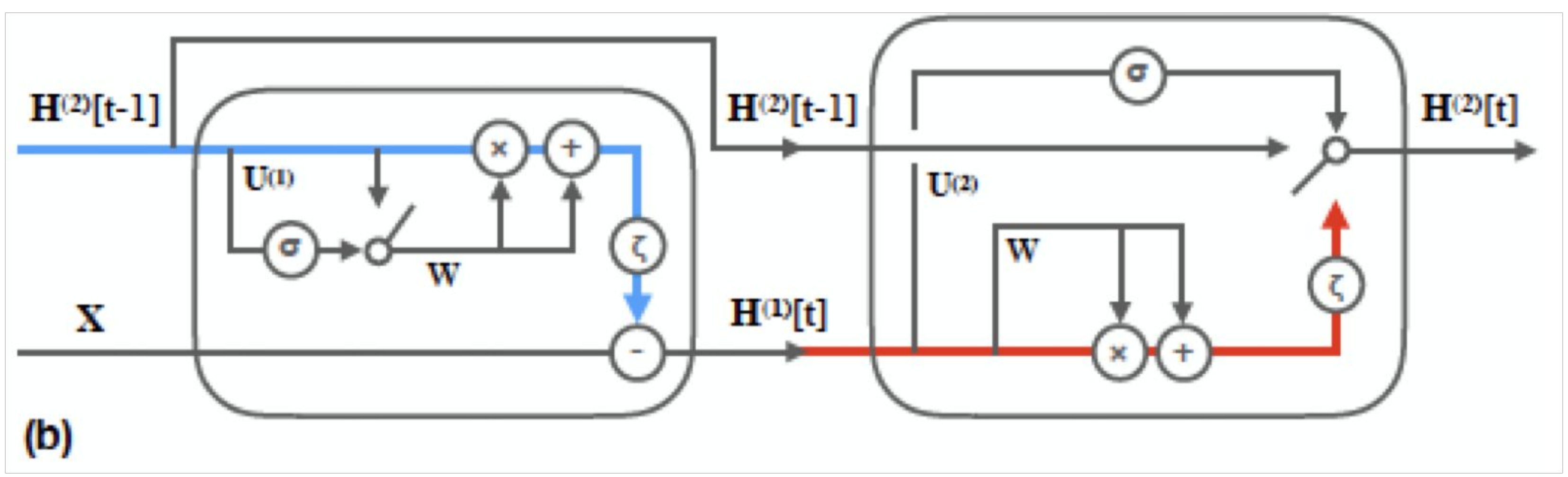


$$C^{(1)}_{xyk} = (\mathbf{W}^I * \mathbf{H}^{(2)})_{xyk}$$
$$C^{(2)}_{xyk} = (\mathbf{W}^E * \mathbf{H}^{(1)})_{xyk}$$
$$\eta \dot{H}^{(1)}_{xyk} + \epsilon^2 H^{(1)}_{xyk} = \left[\xi X_{xyk} - (\alpha H^{(1)}_{xyk} + \mu)\, C^{(1)}_{xyk}\right]_+$$
$$\tau \dot{H}^{(2)}_{xyk} + \sigma^2 H^{(2)}_{xyk} = \left[\gamma C^{(2)}_{xyk}\right]_+ .$$

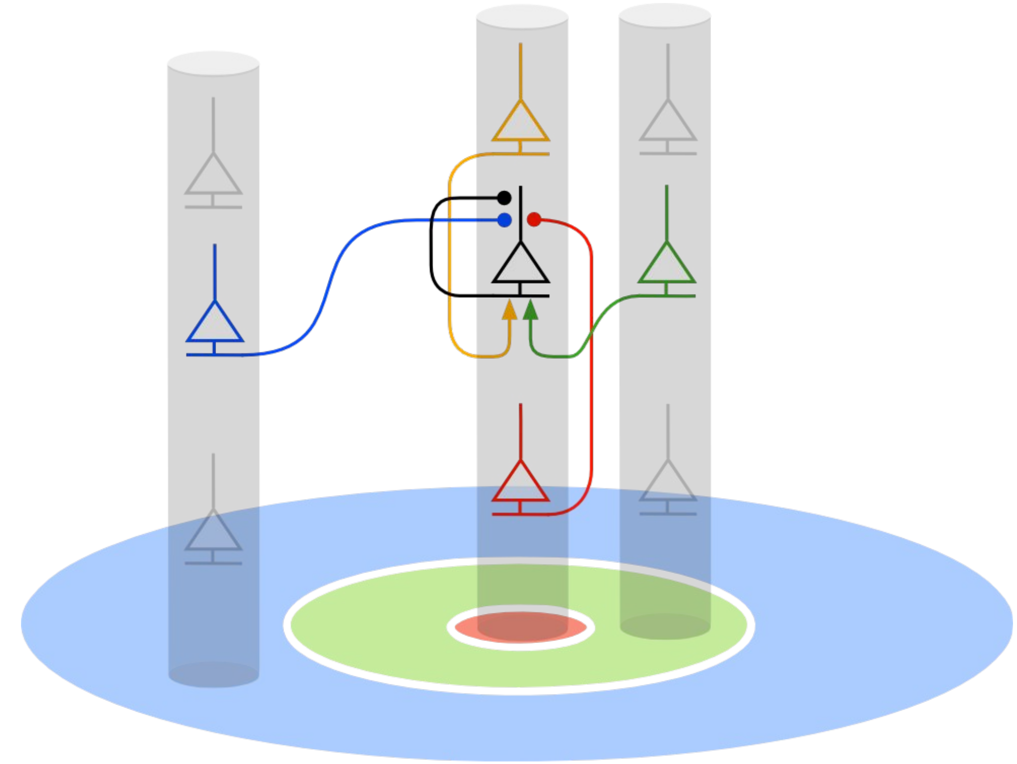


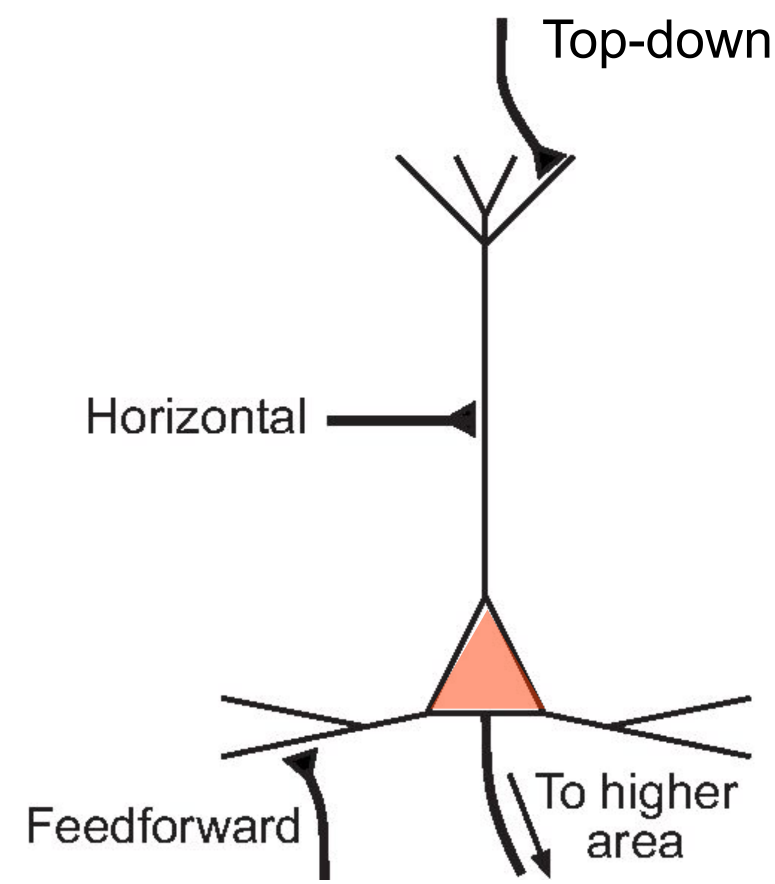
Recurrent neural circuit model(s)





GRU (a) vs hGRU (b)

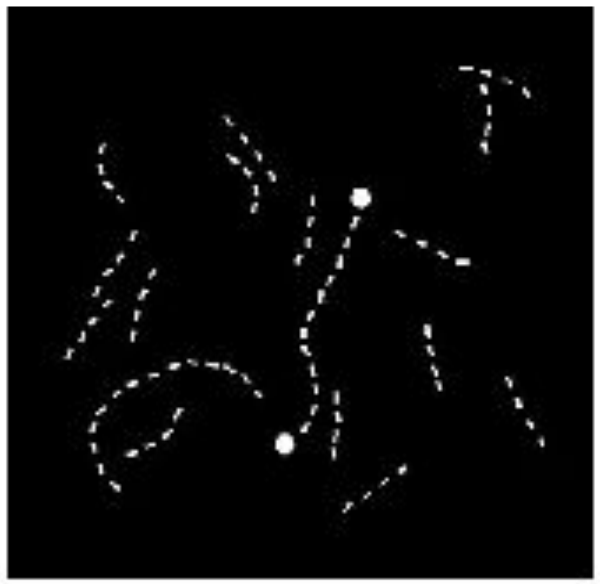




Linsley Kim Veerabadran Windolf & Serre NeurIPS 2018

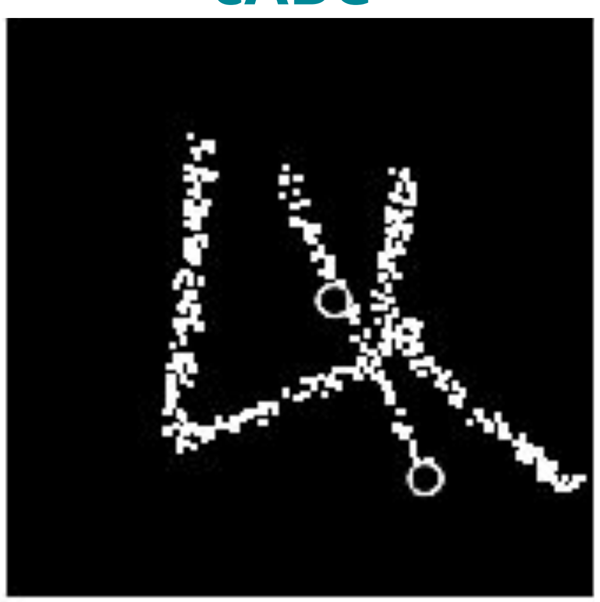
Vision beyond
image categorization

PathFinder



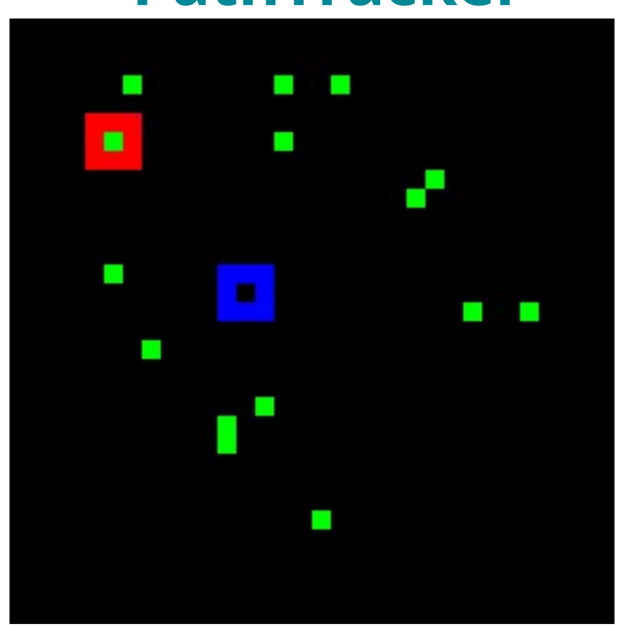
Linsley et al NeurIPS 2018

cABC

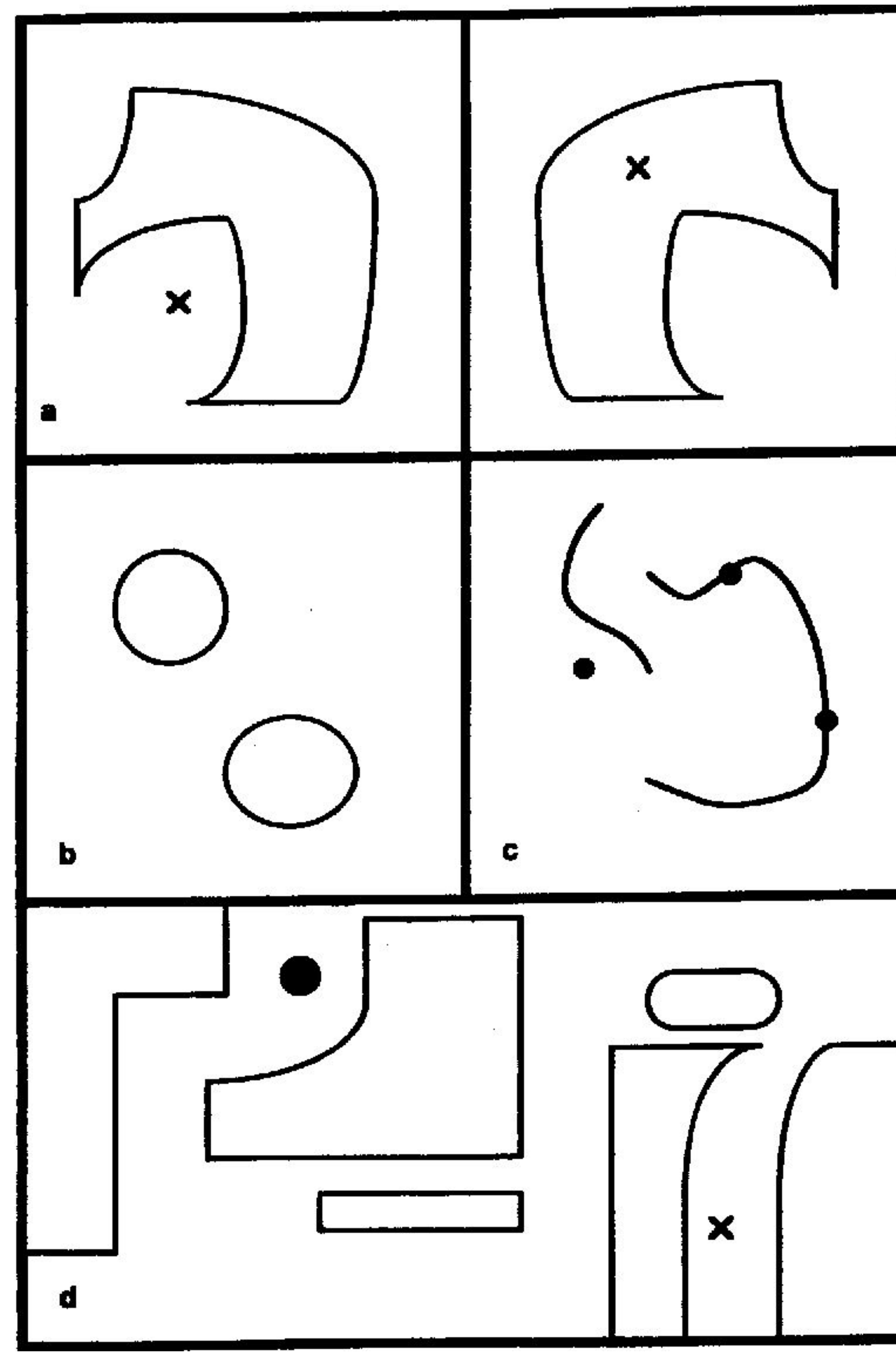


Kim et al ICLR 2020

PathTracker



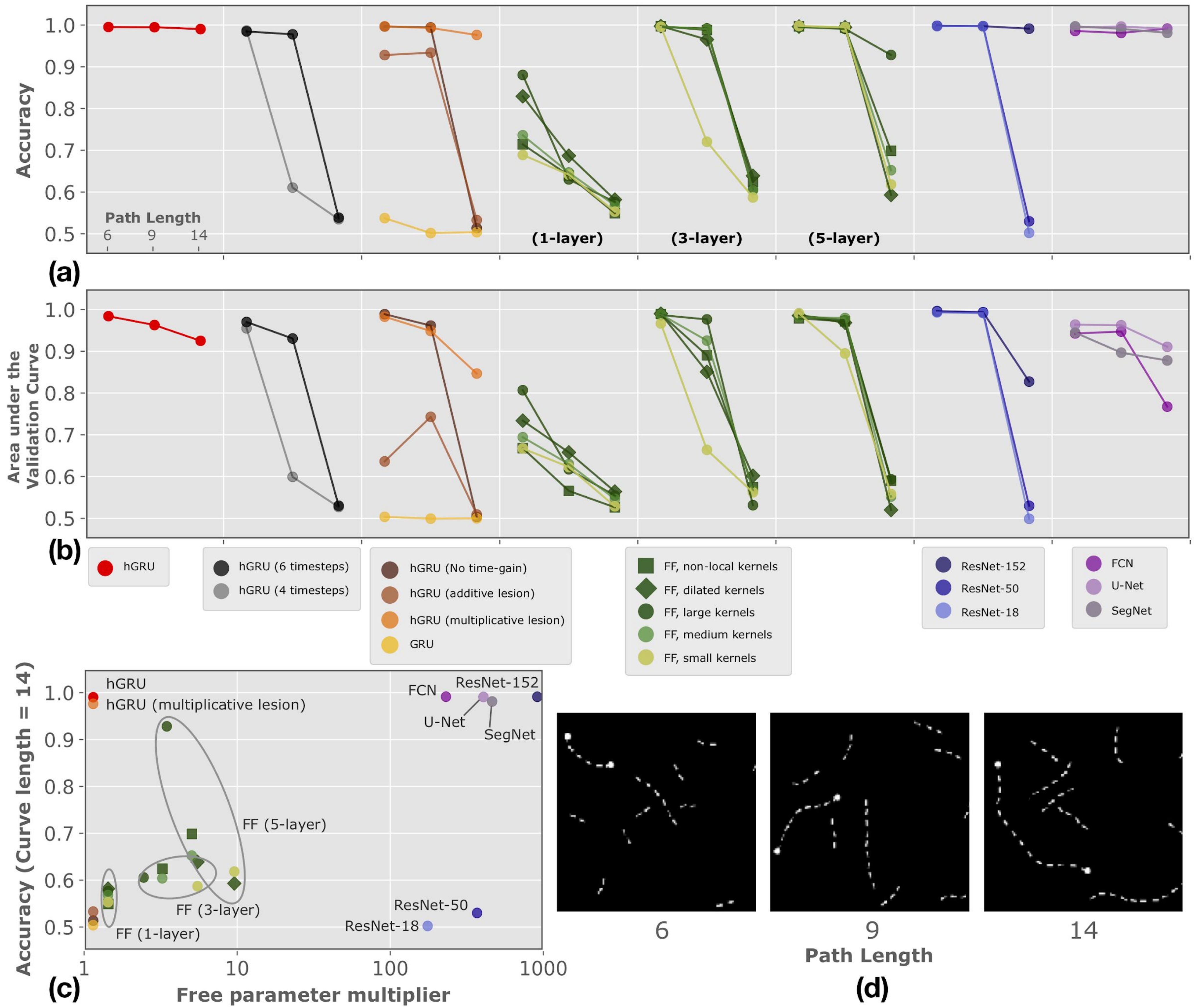
Linsley et al NeurIPS 2021



Ullman 1996

PathFinder

hGRU



PathFinder

hGRU

This is a low-res PathFinder (32x32)

Model	ListOps	Text	Retrieval	Image	Pathfinder	Path-X	Avg
Transformer	36.37	64.27	57.46	42.44	71.40	FAIL	54.39
Local Attention	15.82	52.98	53.39	41.46	66.63	FAIL	46.06
Sparse Trans.	17.07	63.58	**59.59**	**44.24**	71.71	FAIL	51.24
Longformer	35.63	62.85	56.89	42.22	69.71	FAIL	53.46
Linformer	35.70	53.94	52.27	38.56	76.34	FAIL	51.36
Reformer	**37.27**	56.10	53.40	38.07	68.50	FAIL	50.67
Sinkhorn Trans.	33.67	61.20	53.83	41.23	67.45	FAIL	51.39
Synthesizer	36.99	61.68	54.67	41.61	69.45	FAIL	52.88
BigBird	36.05	64.02	59.29	40.83	74.87	FAIL	**55.01**
Linear Trans.	16.13	**65.90**	53.09	42.34	75.30	FAIL	50.55
Performer	18.01	65.40	53.82	42.77	**77.05**	FAIL	51.41
Task Avg (Std)	29 (9.7)	61 (4.6)	55 (2.6)	41 (1.8)	72 (3.7)	FAIL	52 (2.4)

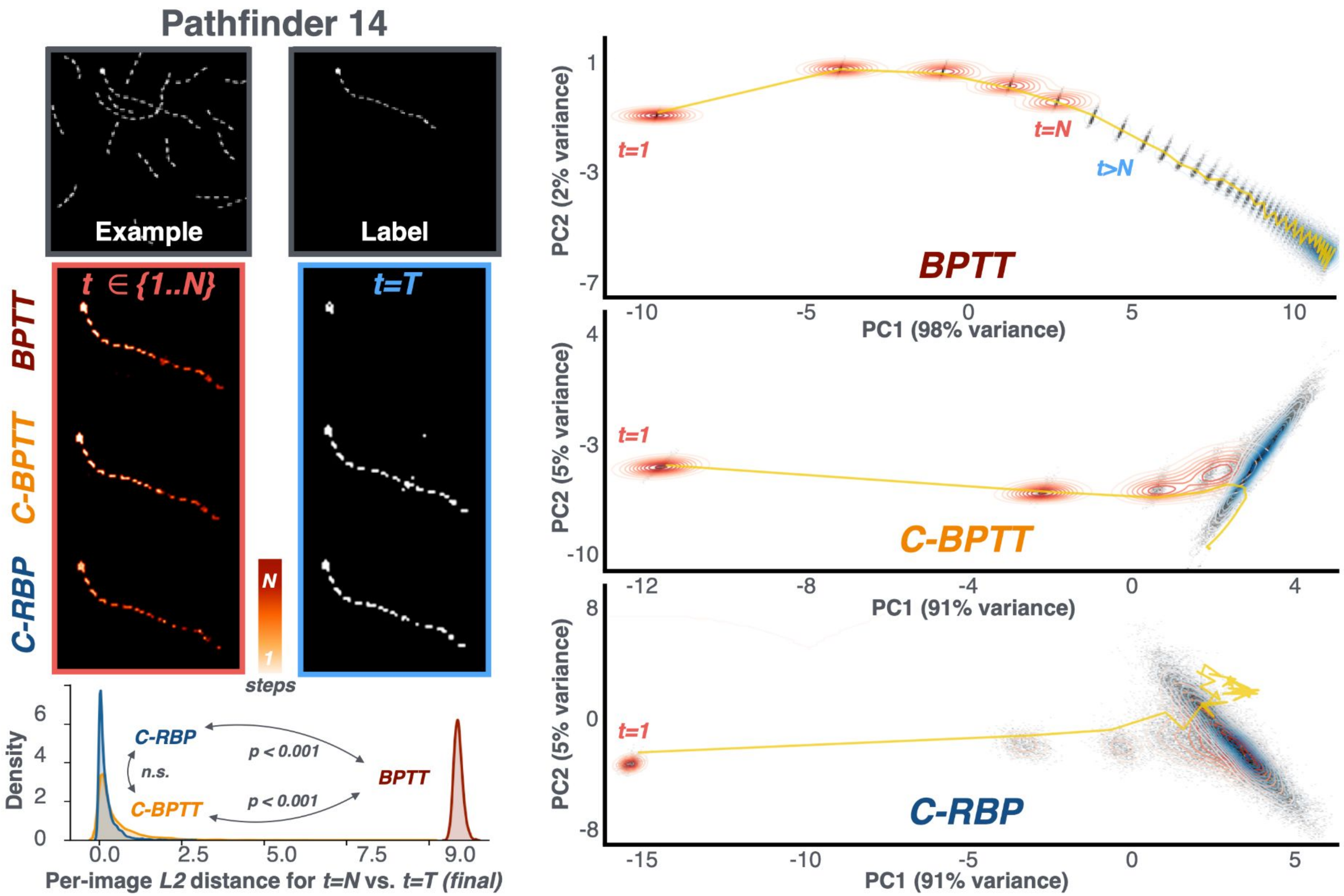
This is the real PathFinder (128x128)

PathFinder

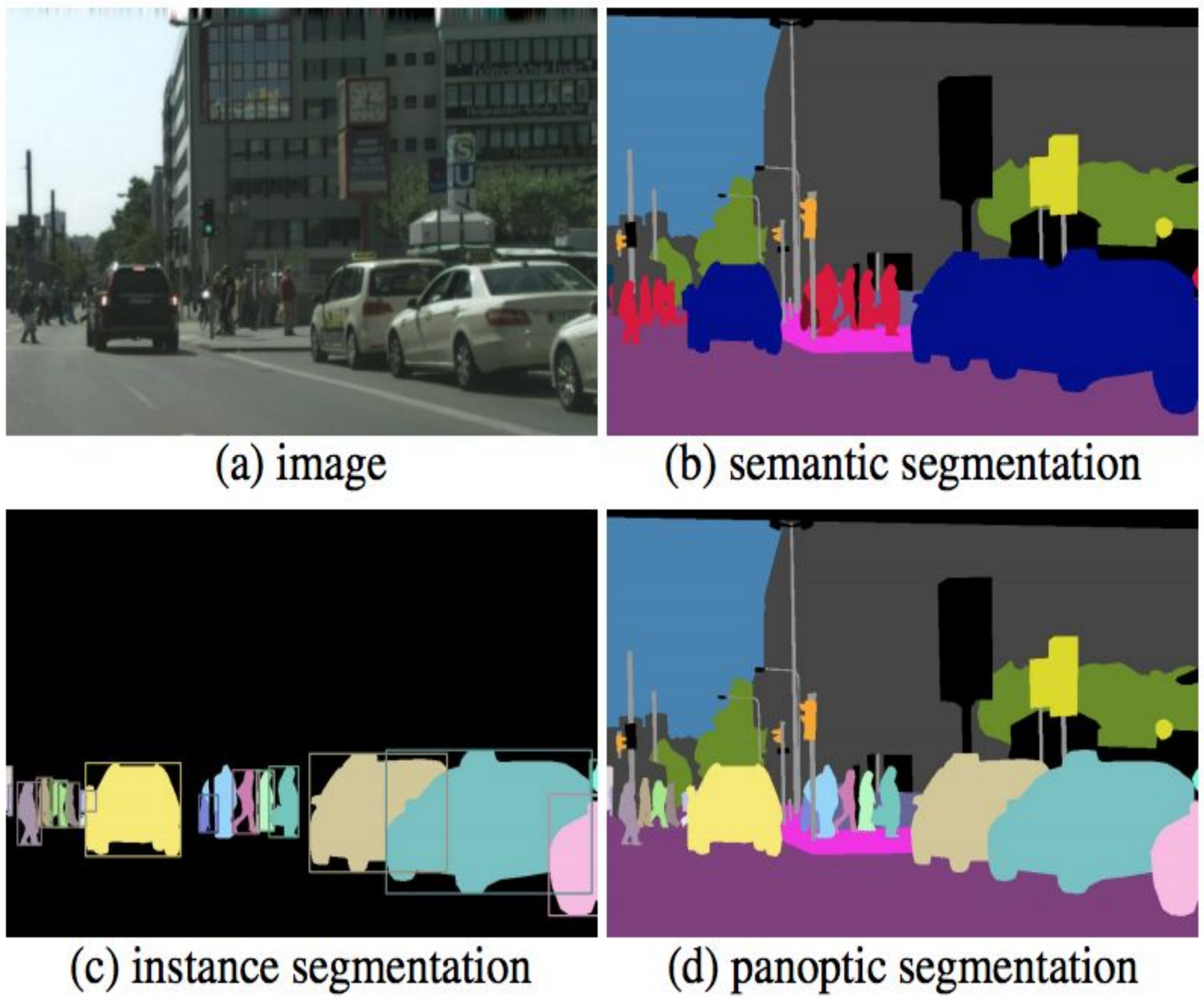
hGRU

Google [Colab]
https://tinyurl.com/tinyhgru

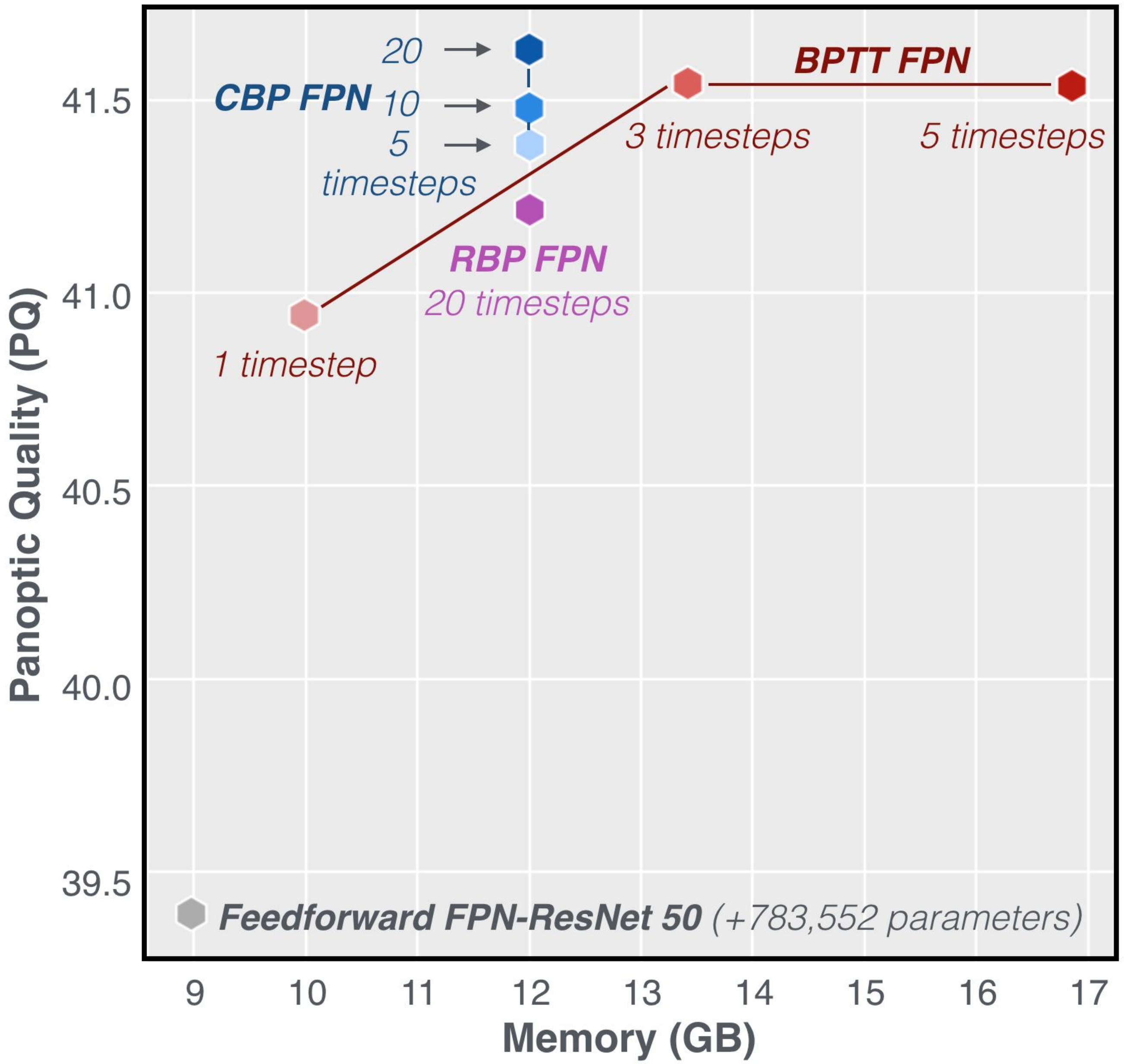
Stabilizing RNNs to help generalize beyond training distrib.



MS COCO
hGRU



(a) image
(b) semantic segmentation
(c) instance segmentation
(d) panoptic segmentation



20
CBP FPN 10
5
timesteps
BPTT FPN
3 timesteps
5 timesteps
RBP FPN
20 timesteps
1 timestep
Feedforward FPN-ResNet 50 (+783,552 parameters)
Panoptic Quality (PQ)
41.5
41.0
40.5
40.0
39.5
9
10
11
12
13
14
15
16
17
Memory (GB)

Feedforward

C-RBP

Linsley* Ashok* Govindarajan* Liu & Serre NeurIPS 2020

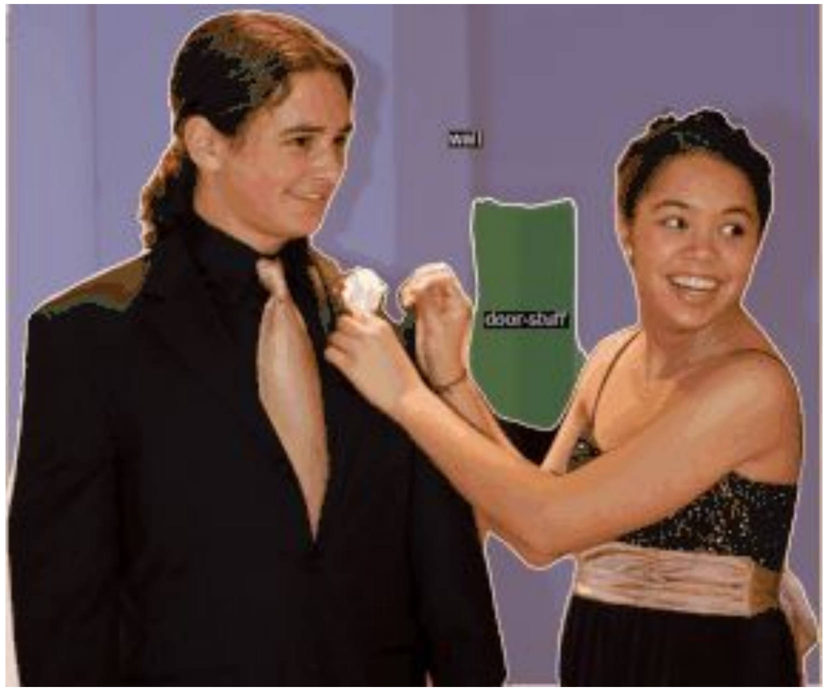
MS COCO

hGRU


wall


banner
wall
fence
playing field


wall-tile
cabinet
wall


wall
door-stuff


wall
door-stuff
fence


dirt
grass

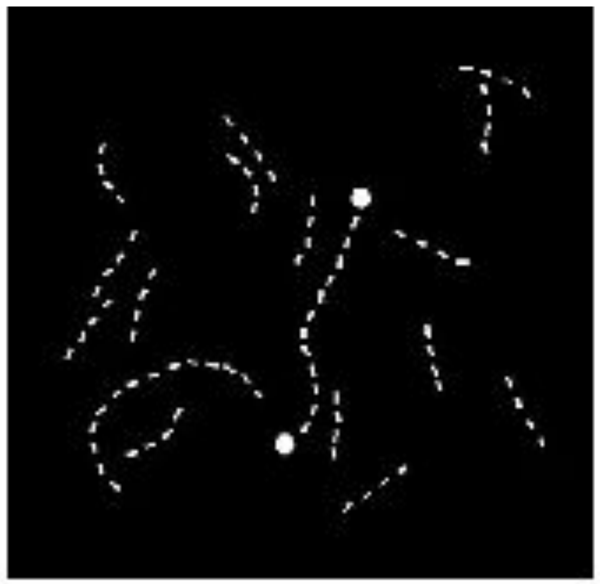
MS COCO

hGRU

Google colab
https://tinyurl.com/cocohgru

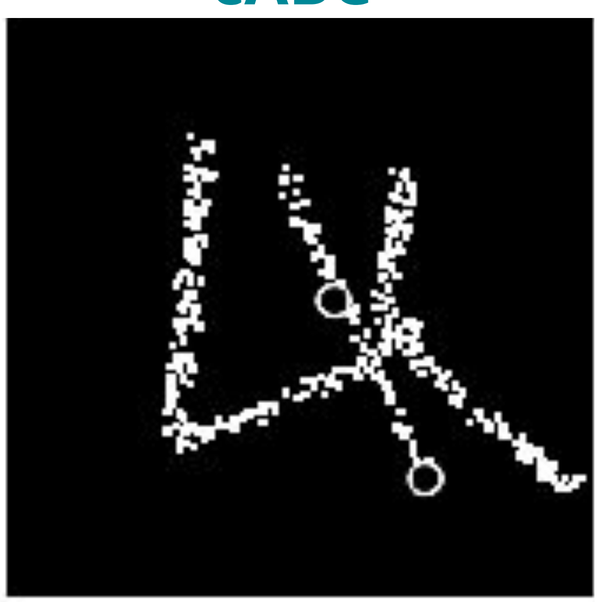
Vision beyond image categorization

PathFinder



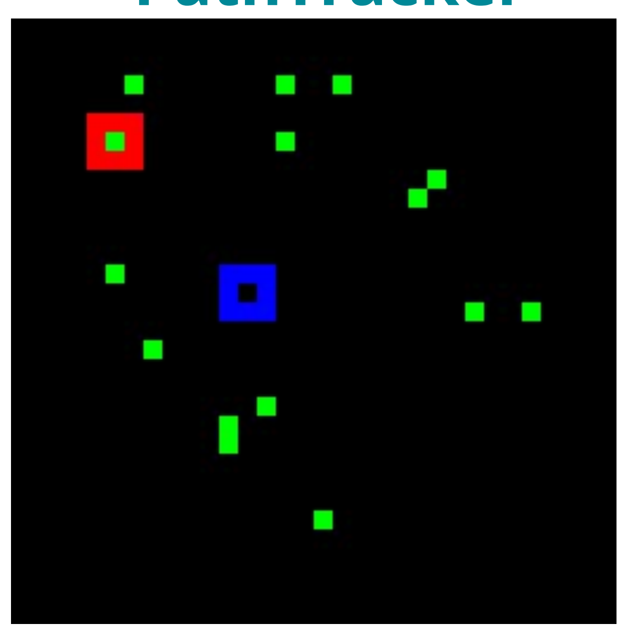
Linsley et al NeurIPS 2018

cABC

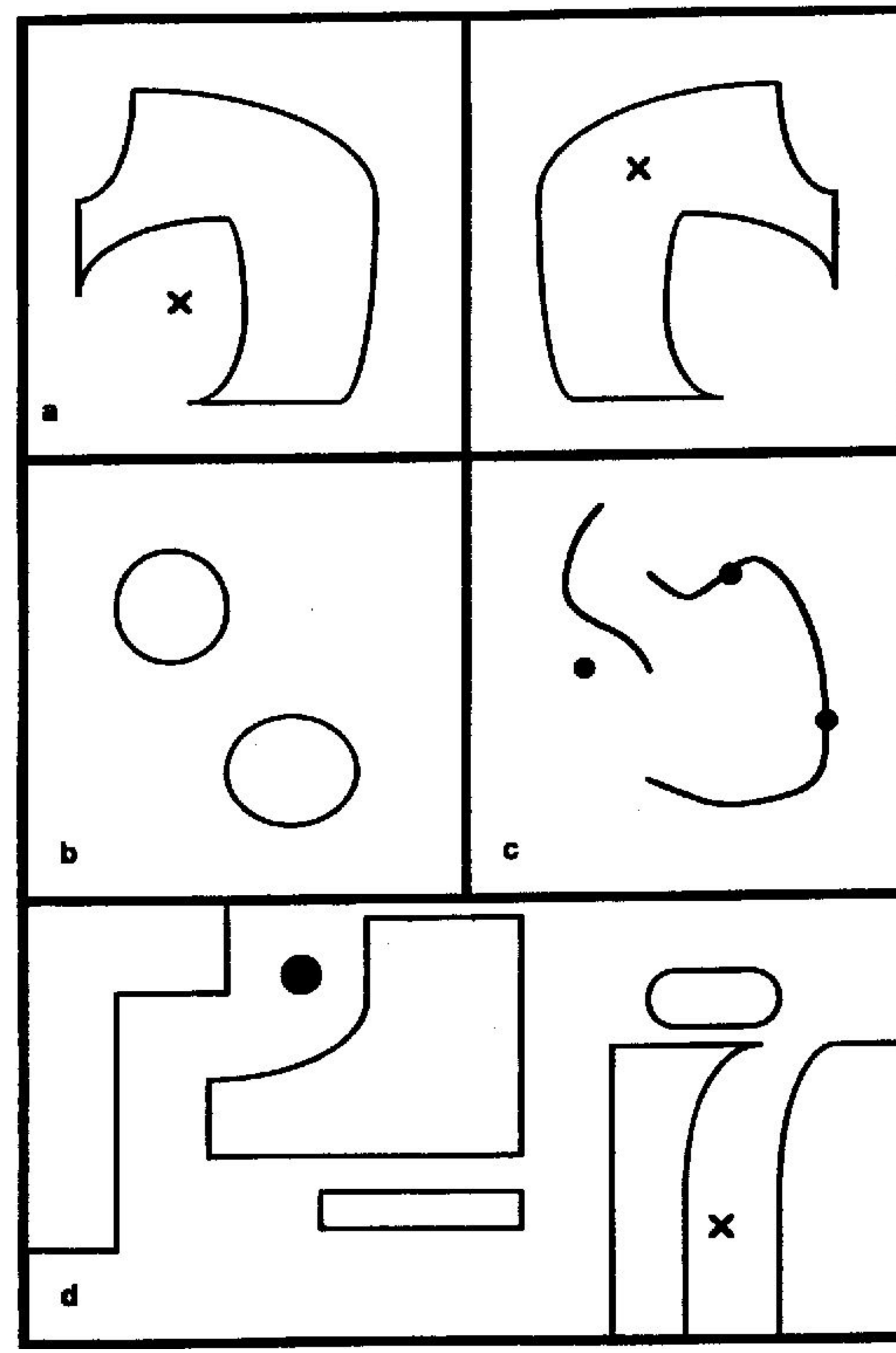


Kim et al ICLR 2020

PathTracker

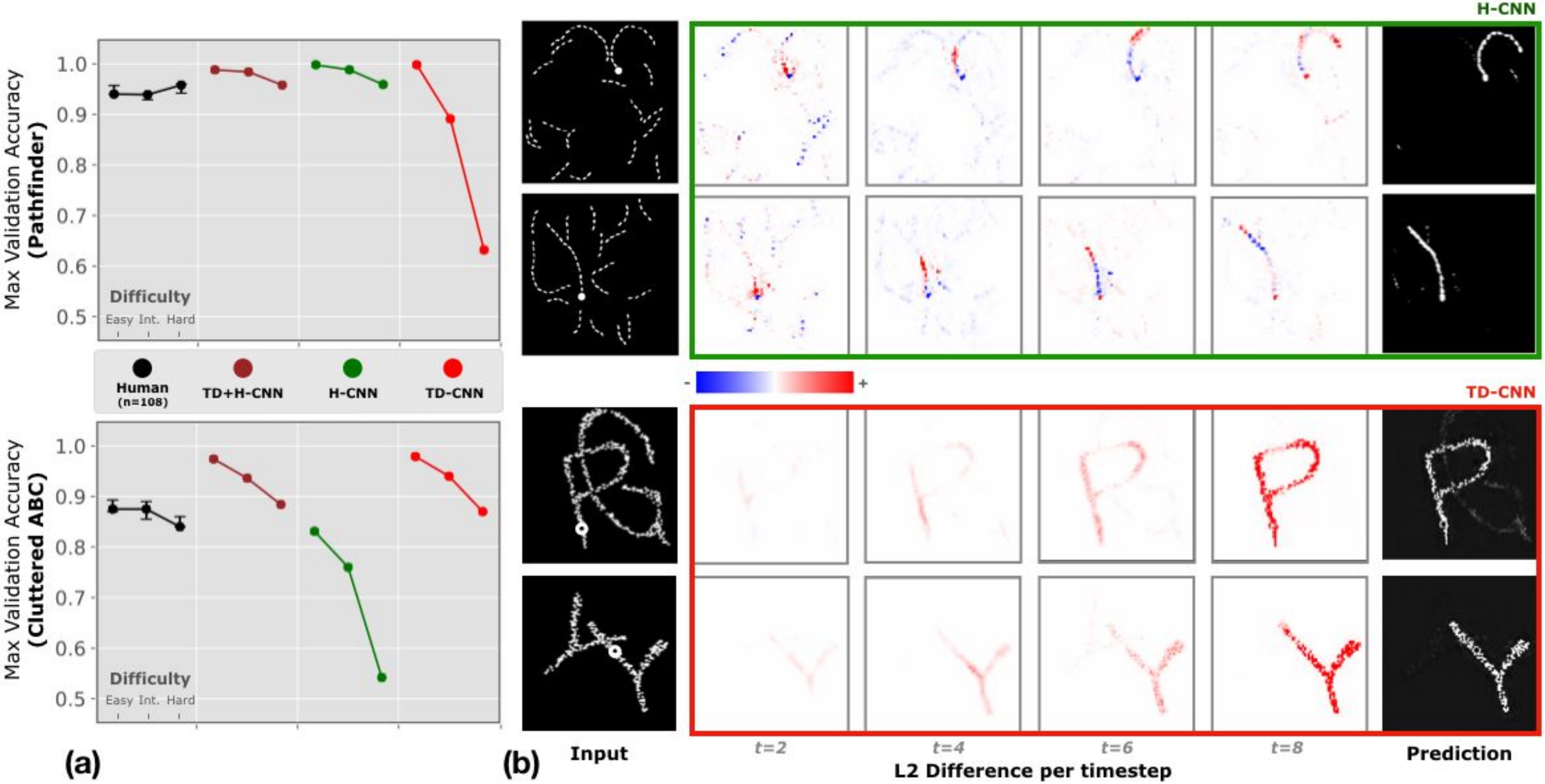


Linsley et al NeurIPS 2021



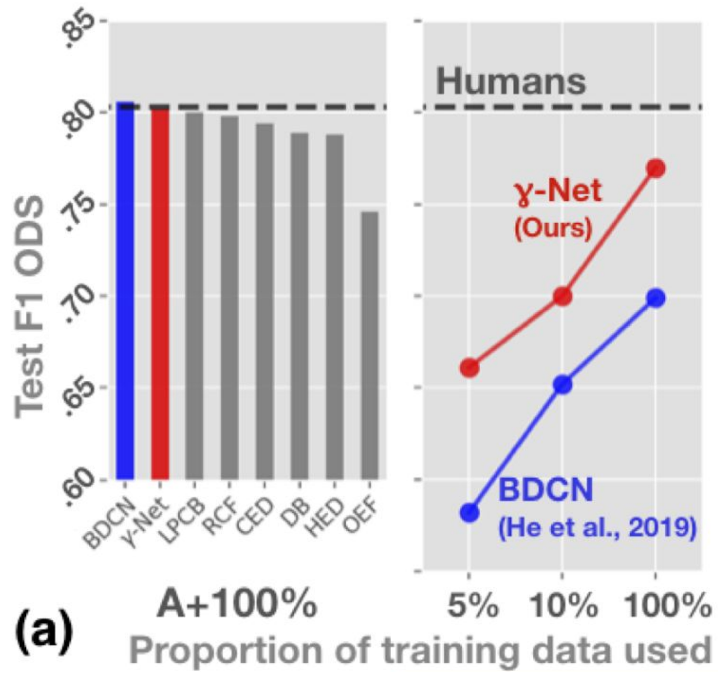
Ullman 1996

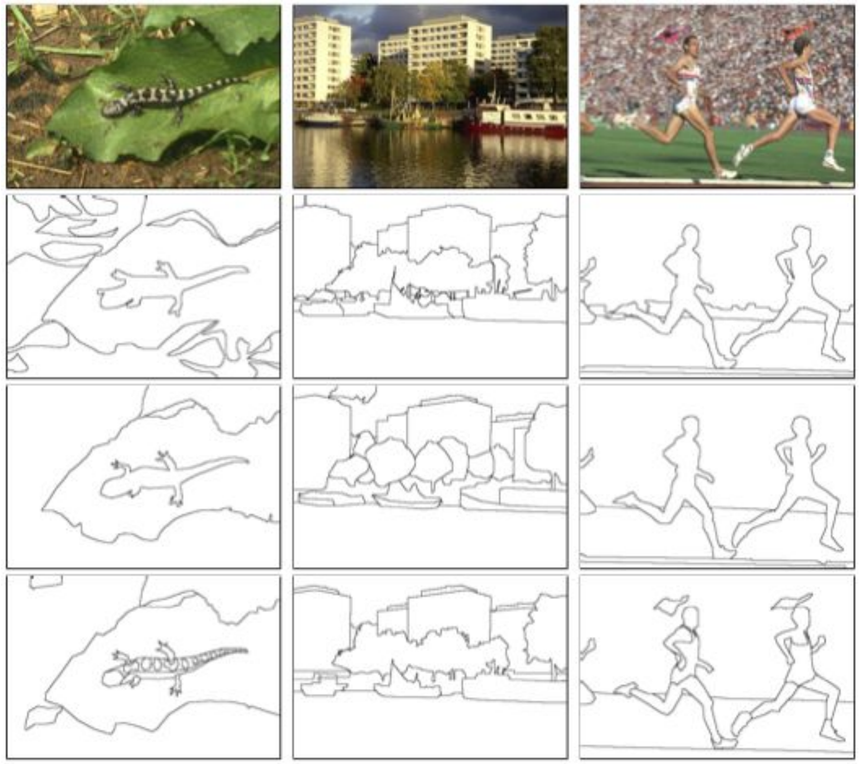
cABC
fGRU



Max Validation Accuracy (Pathfinder)
Max Validation Accuracy (Cluttered ABC)
1.0
0.9
0.8
0.7
0.6
0.5
Difficulty
Easy Int. Hard
Human (n=108)
TD+H-CNN
H-CNN
TD-CNN
(a)
(b)
H-CNN
TD-CNN
Input
t=2
t=4
t=6
t=8
L2 Difference per timestep
Prediction

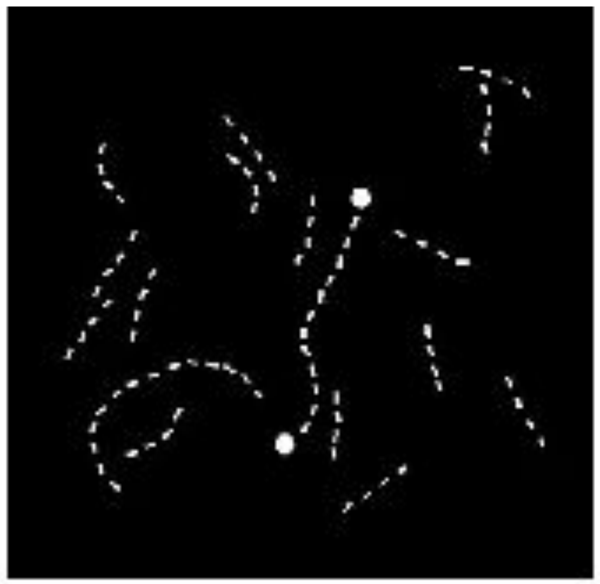
Contours detection
GammaNet





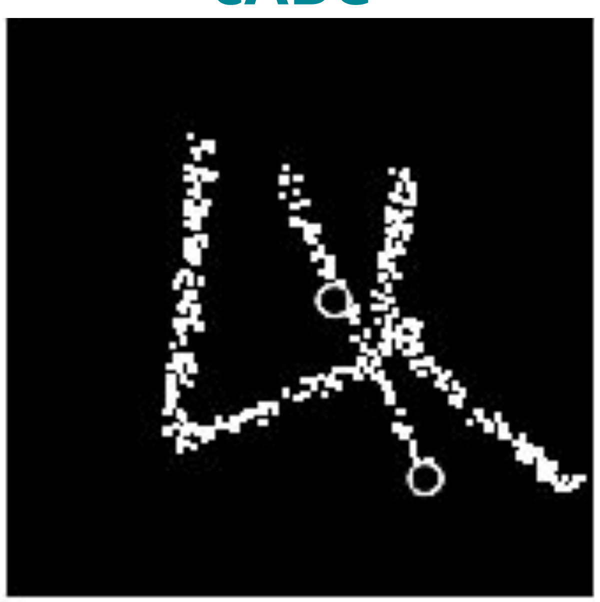
Vision beyond
image categorization

PathFinder



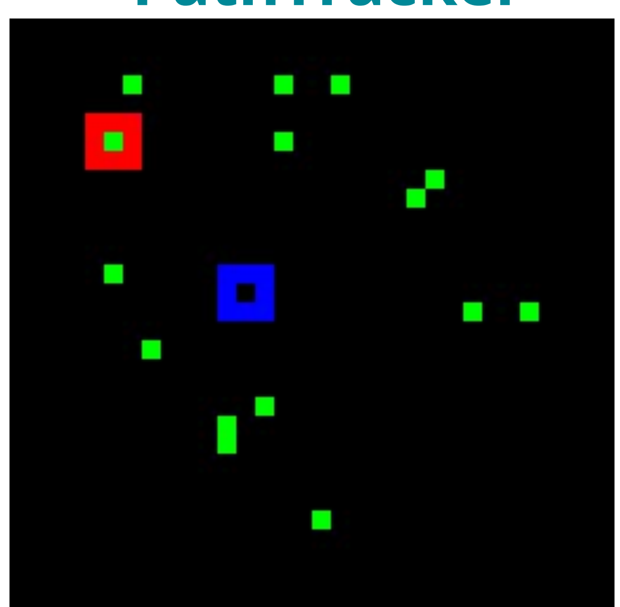
Linsley et al NeurIPS 2018

cABC

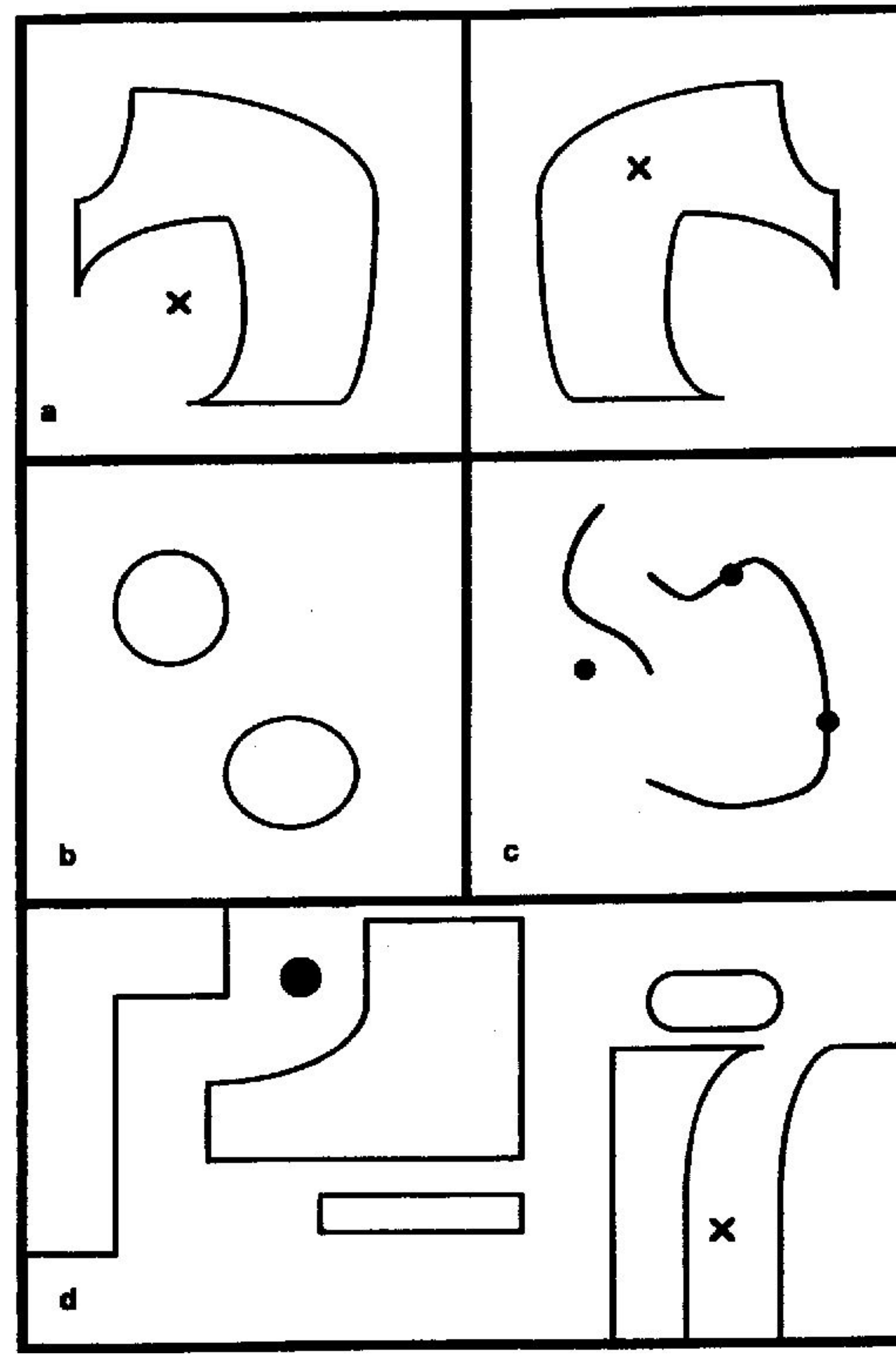


Kim et al ICLR 2020

PathTracker



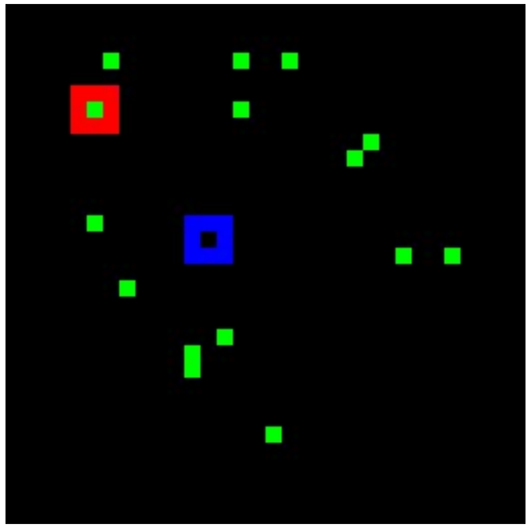
Linsley et al NeurIPS 2021

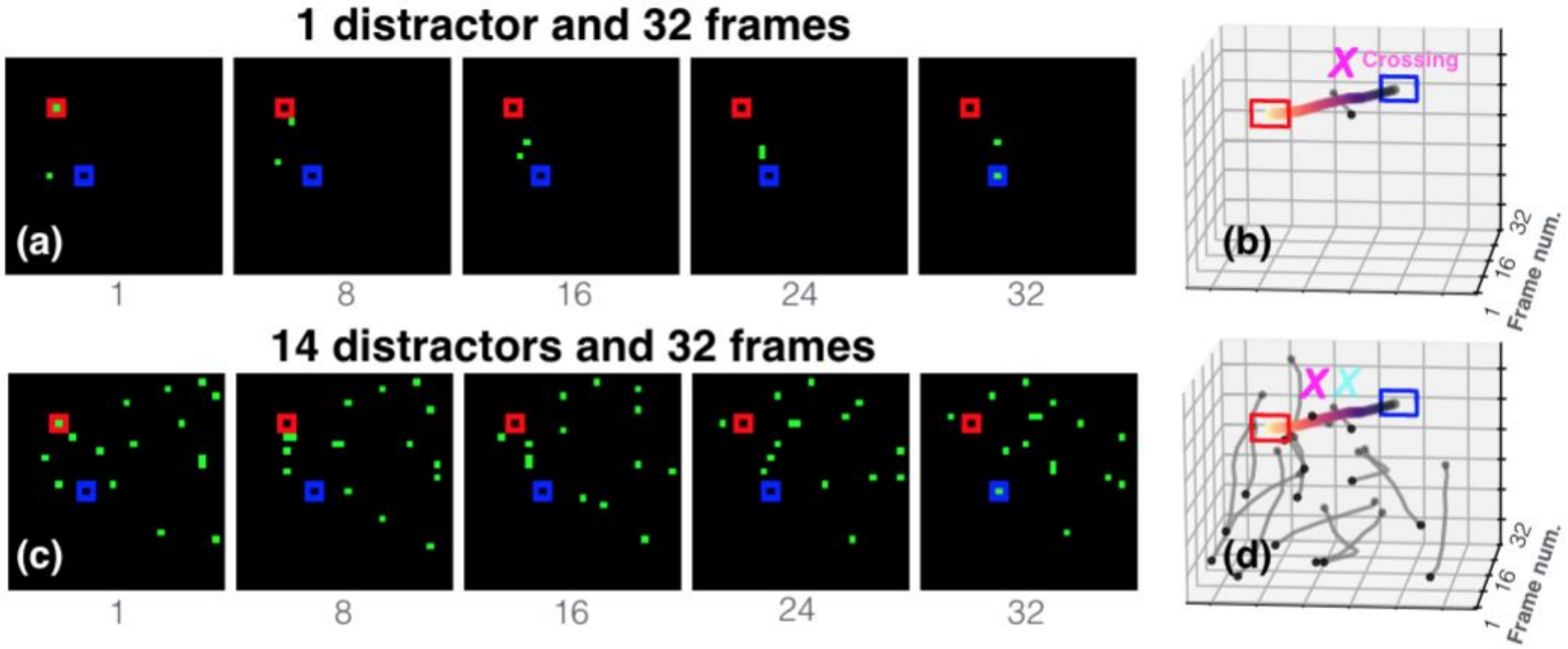


Ullman 1996

PathTracker

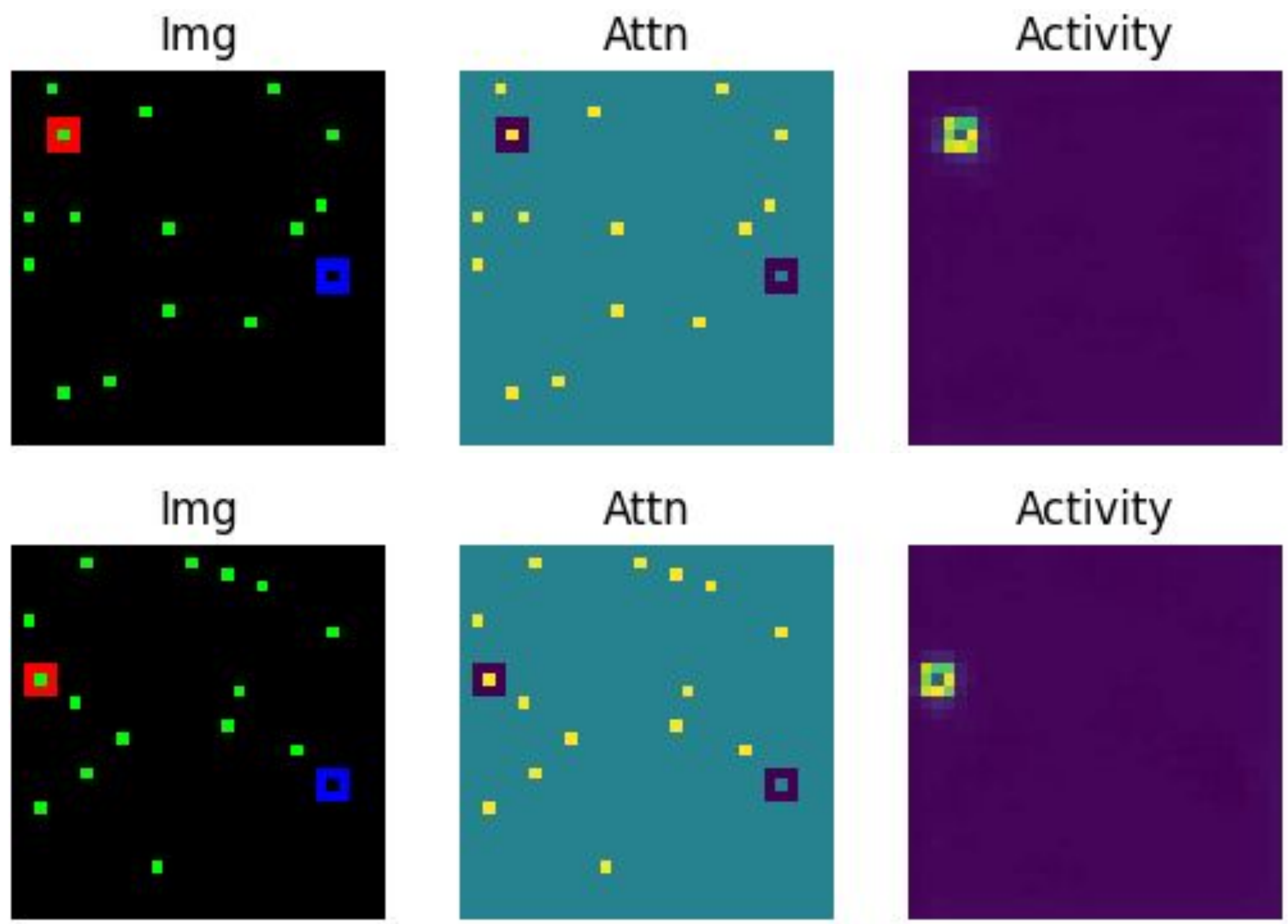
InT

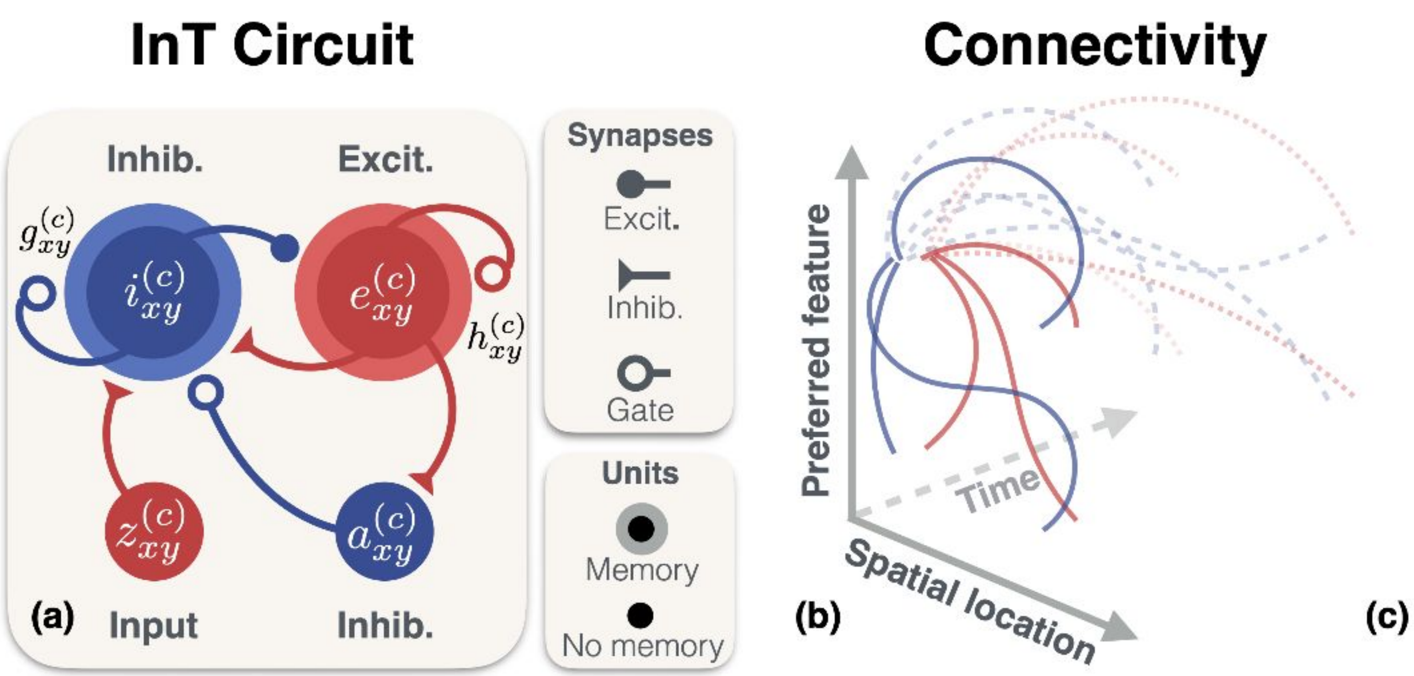




PathTracker

InT





Description

(c)

$$i[t] = g\, i[t-1] + (1-g)\,[z[t] - (\underbrace{\gamma i[t] a[t]}_{\text{Div. } (\div)} + \underbrace{\beta}_{\text{Sub. } (-)})\, m[t] - i[t-1]]_+$$

$$e[t] = h\, e[t-1] + (1-h)\,[i[t] + (\underbrace{\nu i[t]}_{\text{Mult. } (\mathrm{x})} + \underbrace{\mu}_{\text{Add. } (+)})\, n[t] - e[t-1]]_+$$

where

$$M_{e,i} = \left[m_{xy}^{(c)}\right] = \mathbf{W}_{e,i} * (E \odot A)$$

$$E = \left[e_{xy}^{(c)}\right] \qquad N_{i,e} = \left[n_{xy}^{(c)}\right] = \mathbf{W}_{i,e} * I$$

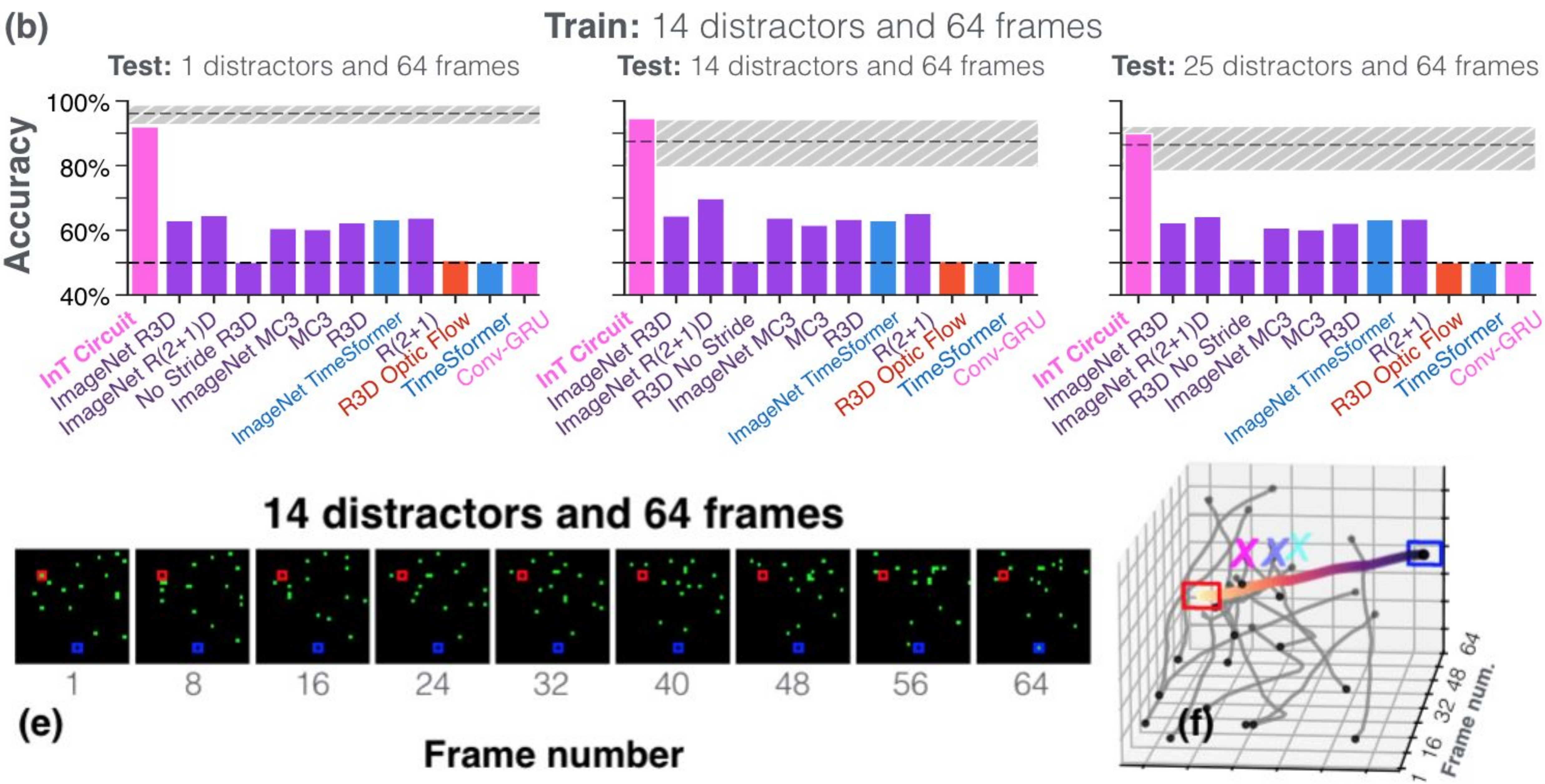
$$I = \left[i_{xy}^{(c)}\right] \qquad G = \left[g_{xy}^{(c)}\right] = \sigma(\mathbf{W}_g * I + \mathbf{U}_g * Z)$$

$$A = \left[a_{xy}^{(c)}\right] = \sigma(\mathbf{W}_a * E) \qquad H = \left[h_{xy}^{(c)}\right] = \sigma(\mathbf{W}_h * E + \mathbf{U}_h * I)$$

Linsley Malik* Kim Govindarajan Mingolla Serre NeurIPS 2021*

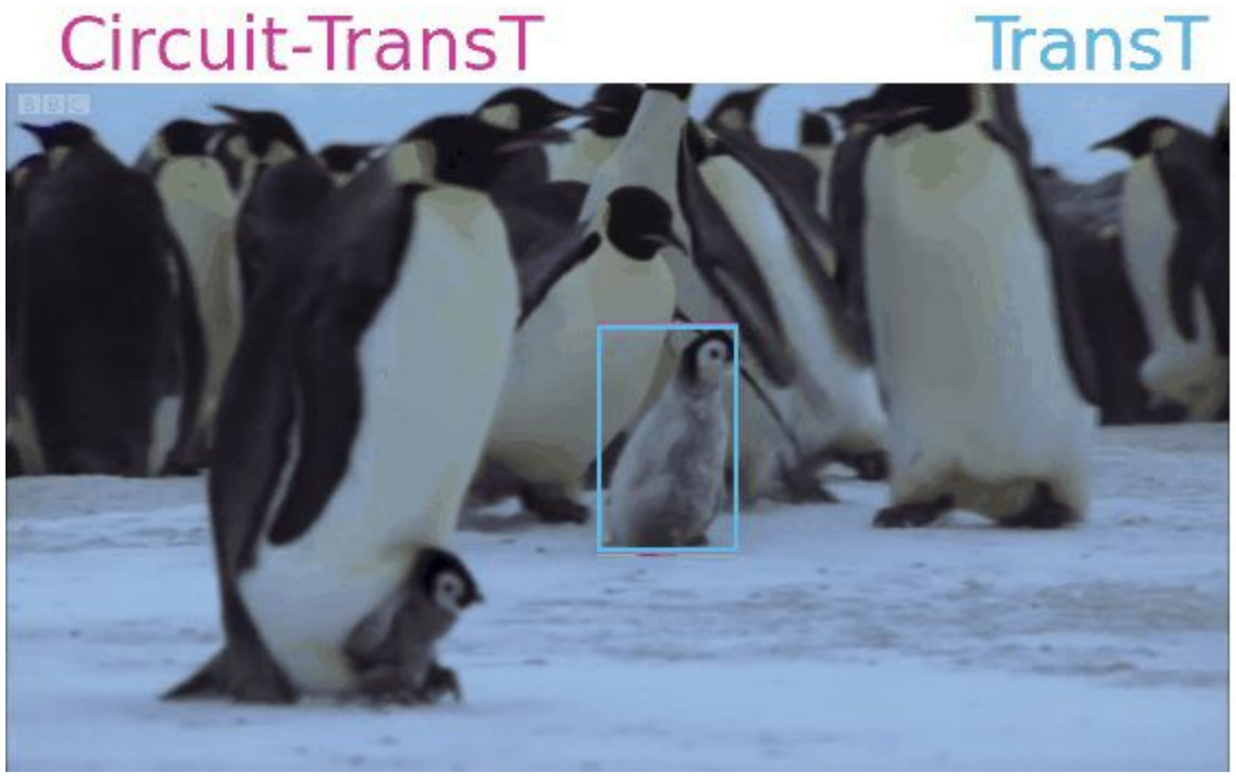
PathTracker

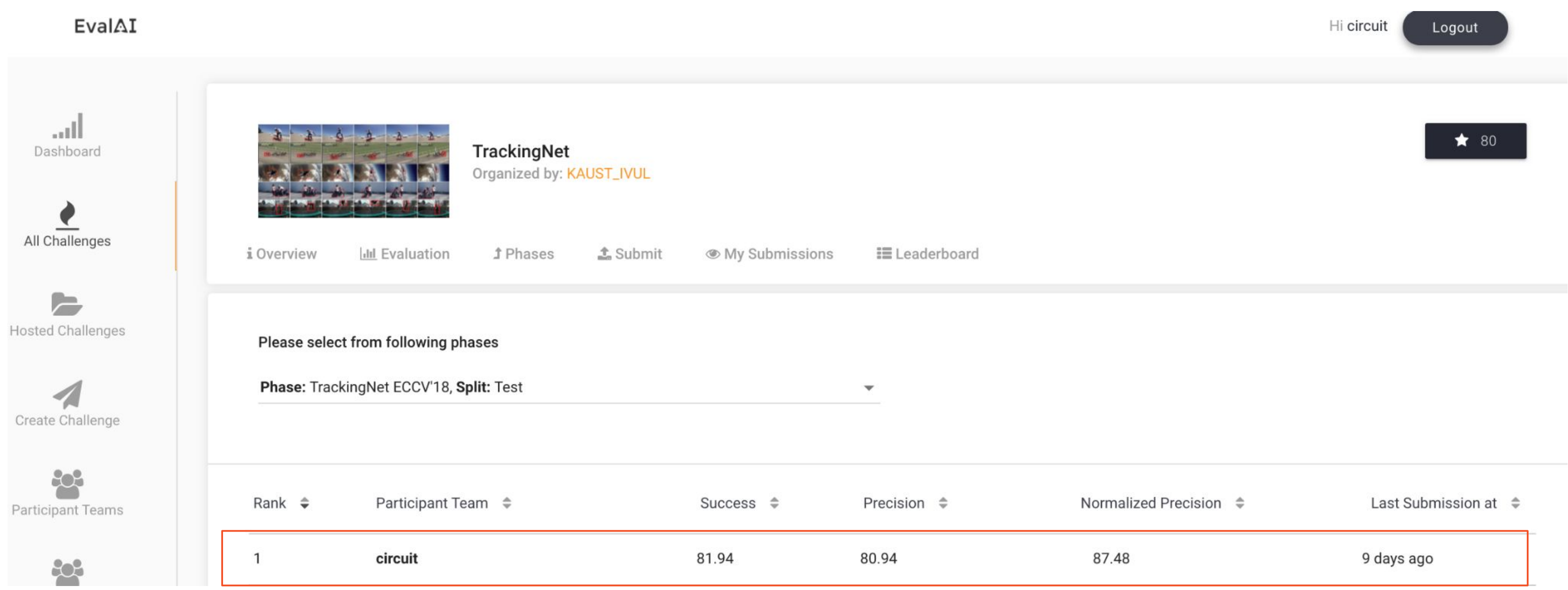
InT



Currently ~~#1~~ (#2)

TrackingNet leaderboard





PathTracker

InT

PathTracker

InT

PathTracker Website:

https://pathtracker.github.io/

PathTracker Code:

https://github.com/pathtracker-code/

Serre Lab Models
(in a PyTorch plug-and-play format)

Github repo

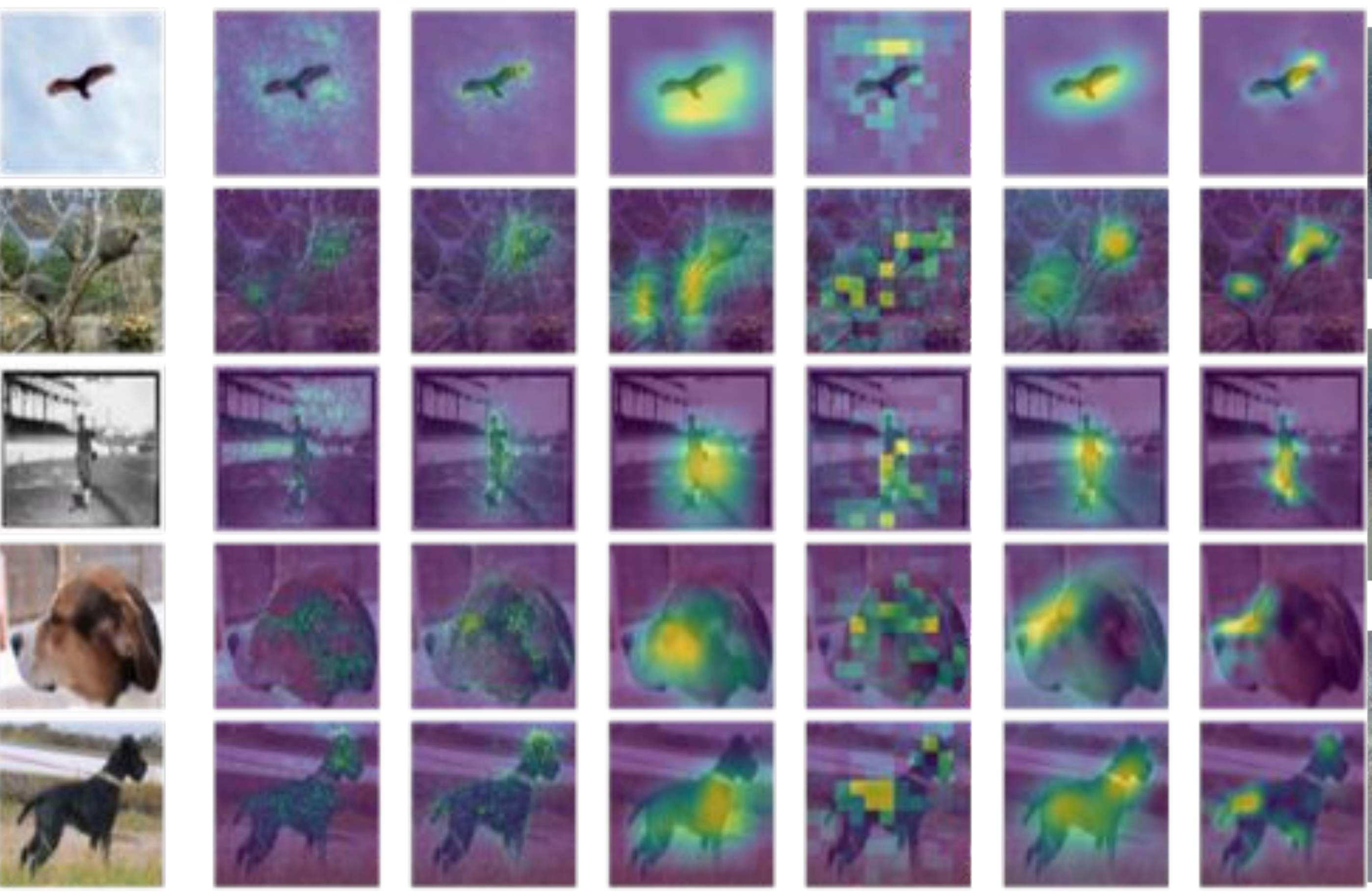
Additional resources

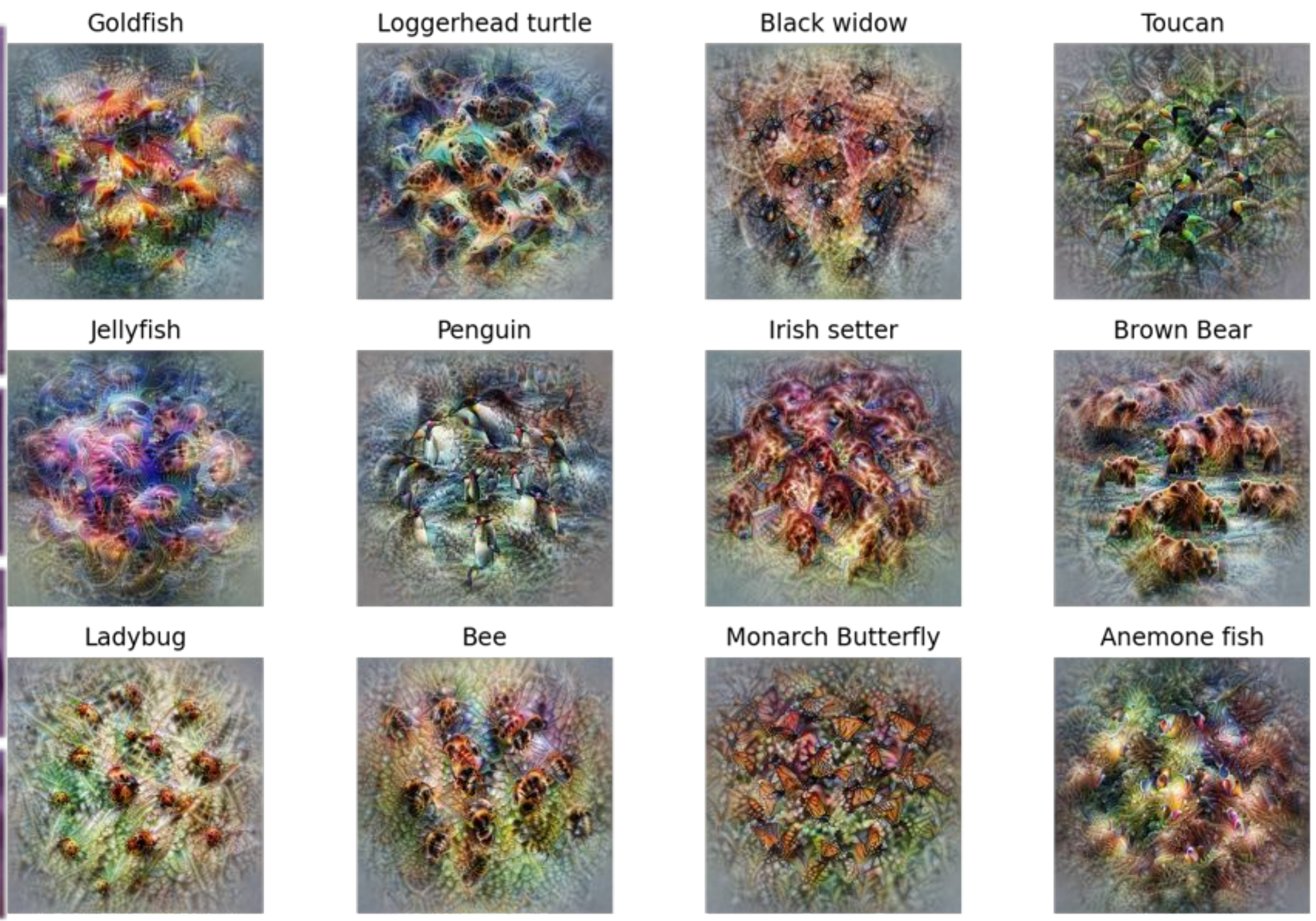
- Serre lab website (including links to all publications)
- Thomas Serre thomas_serre@brown.edu
- Sabine Muzellec sabine_muzellec@brown.edu

Xplique

👋 A Neural Networks Explainability Toolbox

https://github.com/deel-ai/xplique





Thomas FEL and all the DEEL Team

thomas_fel@brown.edu