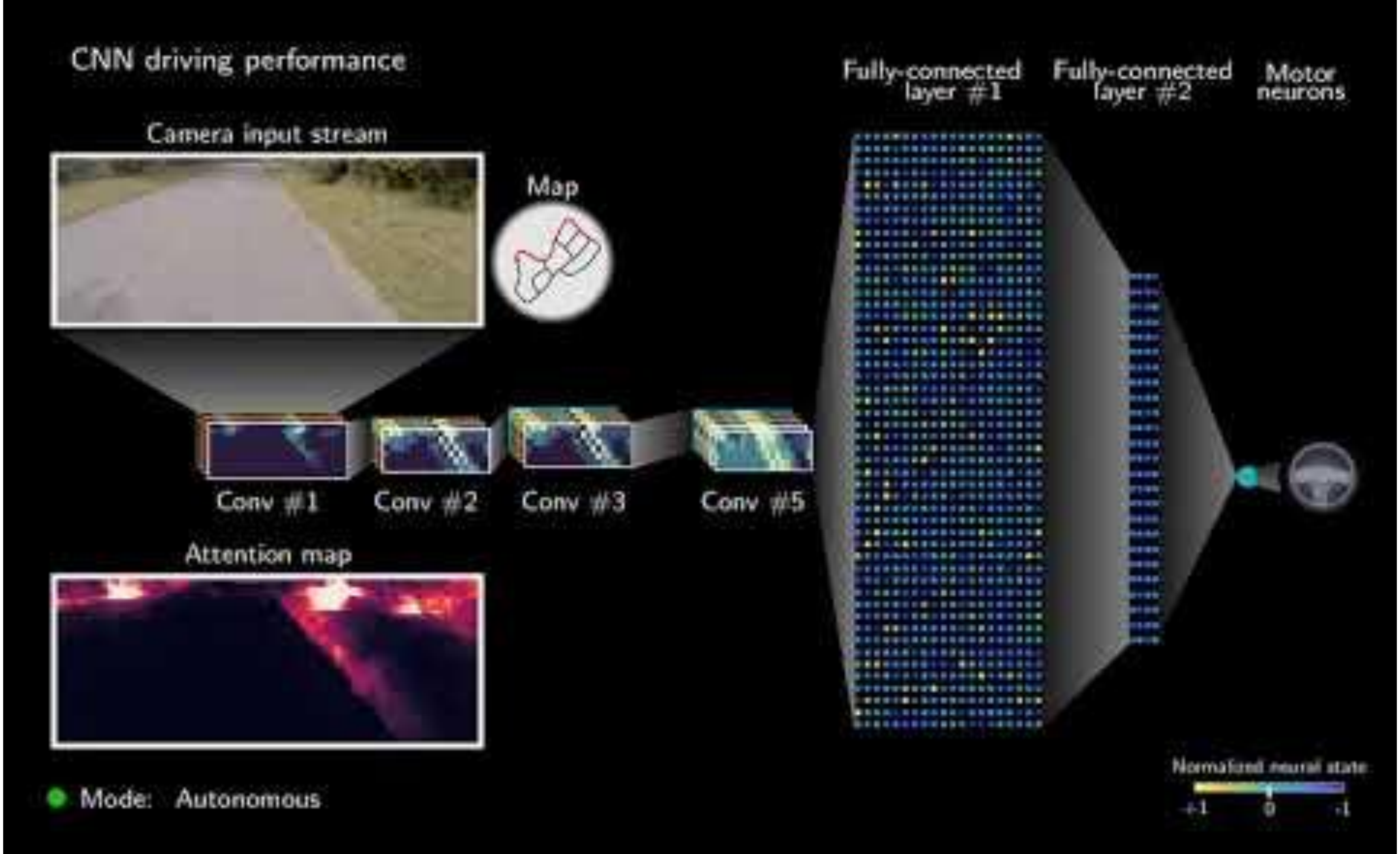
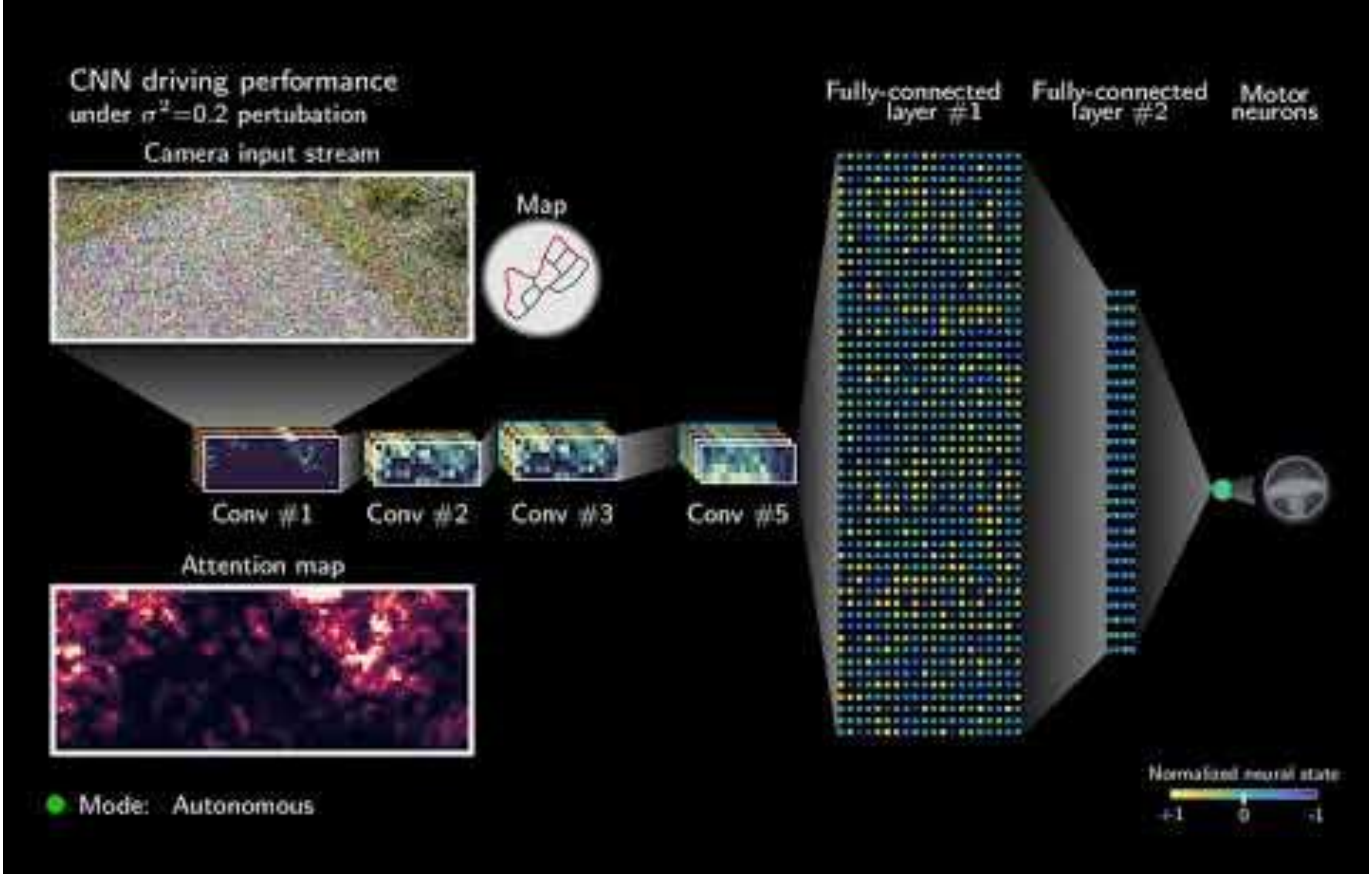
Neural Circuit Policies
Machine Learning Inspired by the *C. elegans* Nervous System

Mathias Lechner
ISTA & MIT

Institute of Science and Technology Austria

Massachusetts Institute of Technology


CNN driving performance
Camera input stream
Map
Conv #1
Conv #2
Conv #3
Conv #5
Attention map
Mode: Autonomous
Fully-connected layer #1
Fully-connected layer #2
Motor neurons
Normalized neural state
+1
0
-1


CNN driving performance
under σ^2=0.2 pertubation
Camera input stream
Map
Conv #1
Conv #2
Conv #3
Conv #5
Attention map
Fully-connected layer #1
Fully-connected layer #2
Motor neurons
Normalized neural state
+1
0
-1
Mode: Autonomous

Artificial neural networks

Artificial neural networks

- High level abstraction of biological nervous system

Artificial neural networks

- High level abstraction of biological nervous system
- Neuron = sum and activation function

$$y = f\left(\sum_{i=1}^{n} x_i w_i + b\right)$$

Artificial neural networks

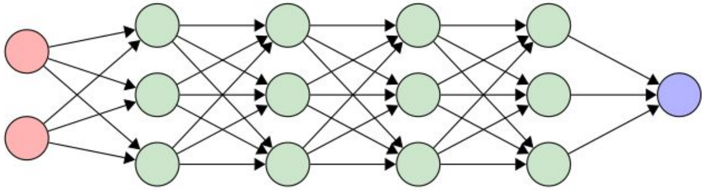
- High level abstraction of biological nervous system
- Neuron = sum and activation function
- Synapse = multiplication with a constant

$$y = f\left(\sum_{i=1}^{n} x_i w_i + b\right)$$

Artificial neural networks

- High level abstraction of biological nervous system
- Neuron = sum and activation function
- Synapse = multiplication with a constant
- Connectivity = Feedforward

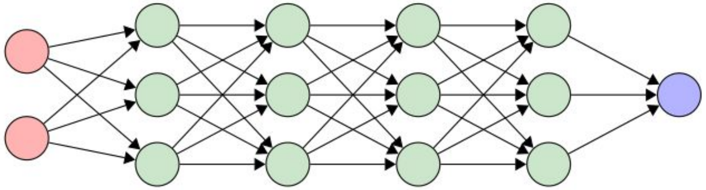
$$y = f\left(\sum_{i=1}^{n} x_i w_i + b\right)$$



Artificial neural networks

- High level abstraction of biological nervous system
- Neuron = sum and activation function
- Synapse = multiplication with a constant
- Connectivity = Feedforward

$$y = f\left(\sum_{i=1}^{n} x_i w_i + b\right)$$



What do we gain if we move a bit closer to biology?

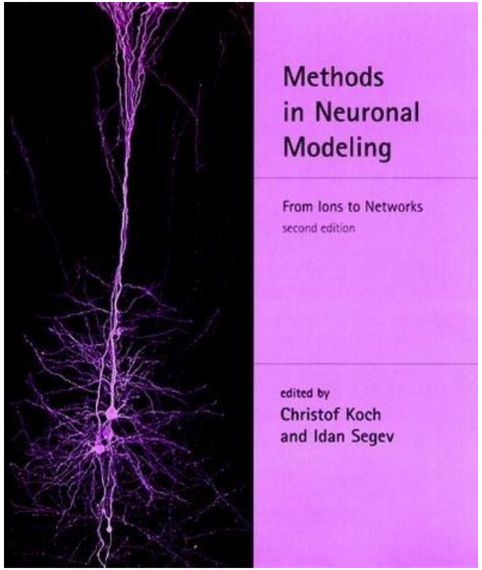
C. elegans

- 1 mm long
- Transparent body
- 959 cells (adult)
- 95 muscle cells
- 302 neurons (non-spiking)
- ~8000 synapses
- Social behavior
- Learning
- Complex search behavior
- Multi-modal sensory processing

Neuron model

[Koch and Segev 1989, Wicks et al. 1996]

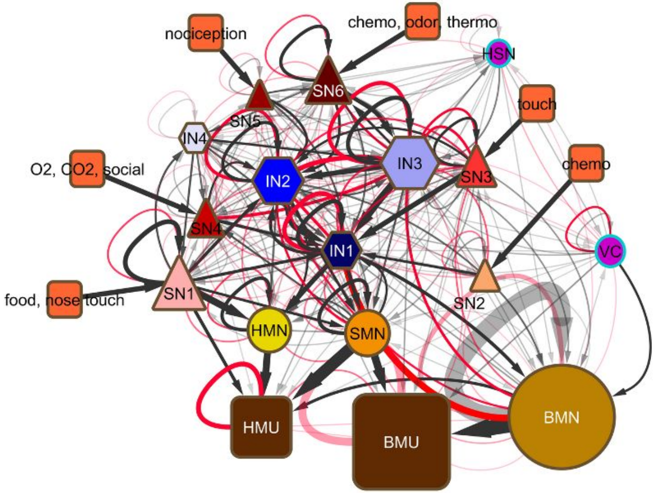
- Neuron membrane = Capacitor
- Ion-channels



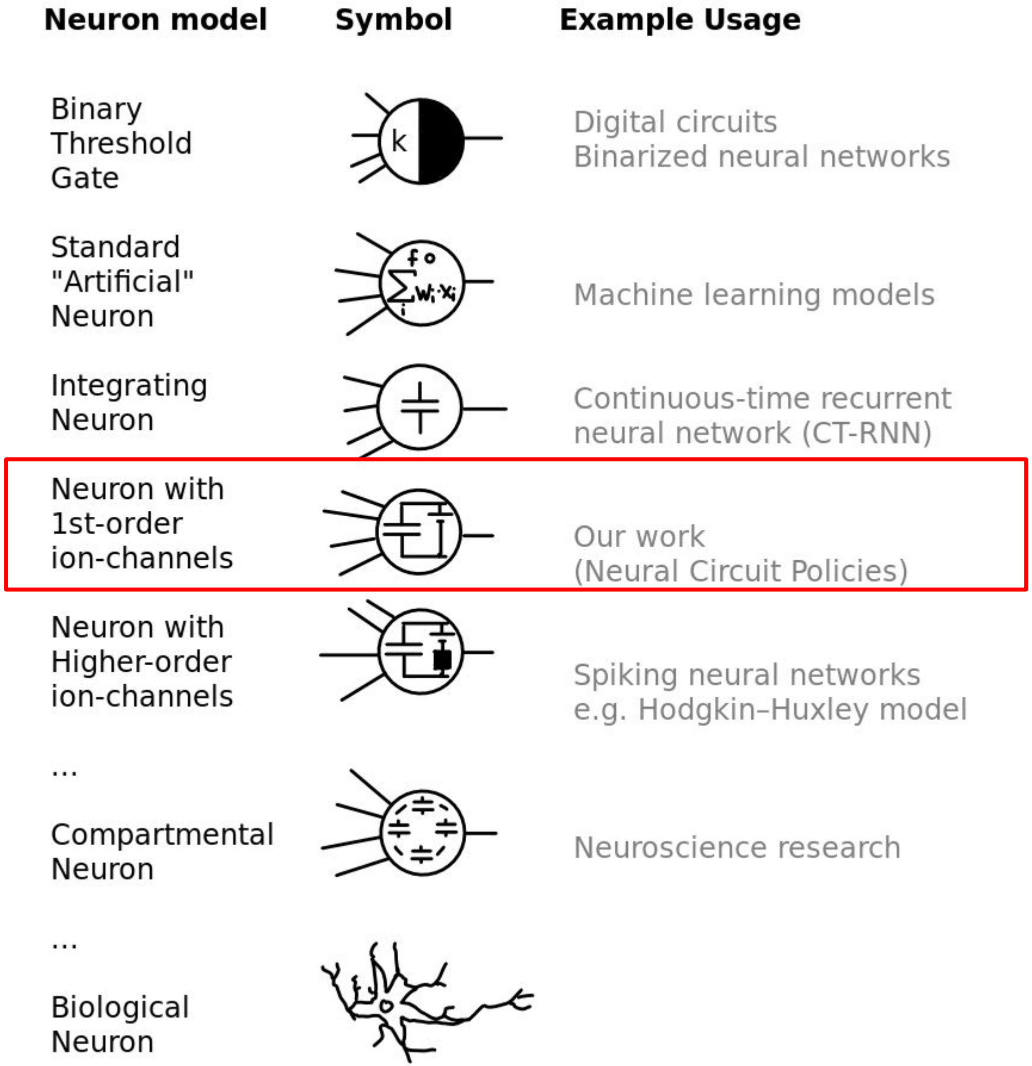
Connection model

[White et al. 1986, Cook et al. 2019]

- Sparsity
- Recurrency



Subset of the C. elegans connectome (Copyright Emmons Lab/wormwiring.org)



Neuron model
Symbol
Example Usage
Binary Threshold Gate
k
Digital circuits
Binarized neural networks
Standard "Artificial" Neuron
Machine learning models
Integrating Neuron
Continuous-time recurrent neural network (CT-RNN)
Neuron with 1st-order ion-channels
Our work (Neural Circuit Policies)
Neuron with Higher-order ion-channels
Spiking neural networks e.g. Hodgkin–Huxley model
...
Compartmental Neuron
Neuroscience research
...
Biological Neuron

A closer look at the neuron model

Standard Recurrent
Neural Network (RNN)
[Hopfield 1982]

$$x(t+1) = f_\theta\Big(x(t), I(t), t\Big)$$

Hasani*, Lechner*, Amini, Rus, Grosu. *Liquid Time-constant Networks*. AAAI , 2021

A closer look at the neuron model

Standard Recurrent
Neural Network (RNN)
[Hopfield 1982]

$$x(t+1) = f_\theta\Big(x(t), I(t), t\Big)$$

Neural ODE
[Chen et al. NeurIPS 2018]

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = f_\theta\Big(x(t), I(t), t\Big)$$

Hasani*, Lechner*, Amini, Rus, Grosu. *Liquid Time-constant Networks*. AAAI , 2021

A closer look at the neuron model

Standard Recurrent Neural Network (RNN)
[Hopfield 1982]

$$x(t+1) = f_\theta\Big(x(t), I(t), t\Big)$$

Neural ODE
[Chen et al. NeurIPS 2018]

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = f_\theta\Big(x(t), I(t), t\Big)$$

Continuous-time (CT) RNN
[Funahashi et al. 1993]

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = -\frac{x(t)}{\tau} + f_\theta\Big(x(t), I(t), t\Big)$$

Hasani*, Lechner*, Amini, Rus, Grosu. *Liquid Time-constant Networks*. AAAI , 2021

A closer look at the neuron model

Standard Recurrent Neural Network (RNN)
[Hopfield 1982]

$$x(t+1) = f_\theta\Big(x(t), I(t), t\Big)$$

Neural ODE
[Chen et al. NeurIPS 2018]

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = f_\theta\Big(x(t), I(t), t\Big)$$

Continuous-time (CT) RNN
[Funahashi et al. 1993]

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = -\frac{x(t)}{\tau} + f_\theta\Big(x(t), I(t), t\Big)$$

Liquid Time-Constant Network (LTC)

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = \underbrace{-\frac{x(t)}{\tau}}_{\text{“Leaky-integrator”}} + f_\theta\Big(x(t), I(t), t\Big)\underbrace{\Big(A - x(t)\Big)}_{\text{Conductance-based synapse model}}$$

Hasani*, Lechner*, Amini, Rus, Grosu. *Liquid Time-constant Networks*. AAAI , 2021

Some properties of LTC

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = -\frac{x(t)}{\tau} + f_\theta\Big(x(t), I(t), t\Big)\Big(A - x(t))\Big)$$

Theorem 1: "Time-constant" is bounded
(i.e., in x(t)*g(x) the values of g(x) are bounded)

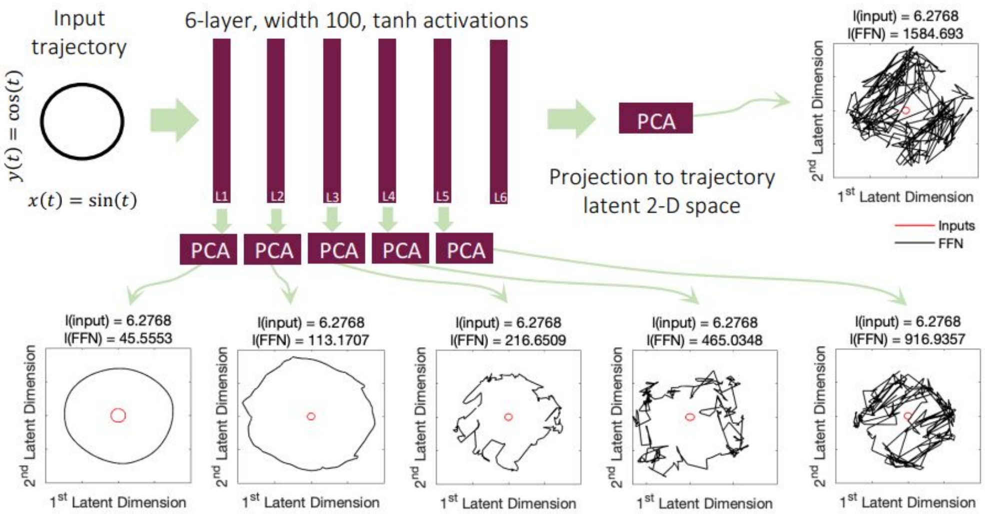
Theorem 2: The state of each neuron is bounded

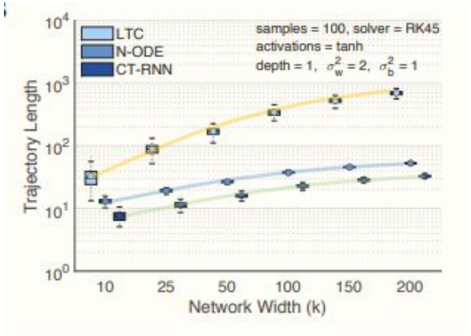
Theorem 3: LTCs are universal approximator

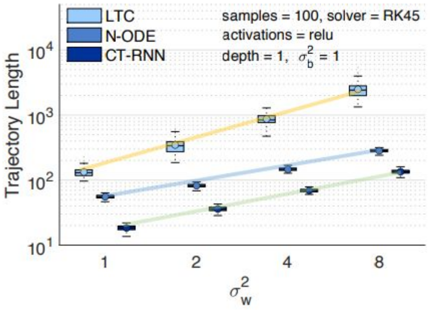
Hasani*, Lechner*, Amini, Rus, Grosu. *Liquid Time-constant Networks*. AAAI , 2021

Trajectory length

- Empirical metric for modelling capacity [Raghu et al. 2017]







Hasani*, Lechner*, Amini, Rus, Grosu. *Liquid Time-constant Networks*. AAAI , 2021

Experiments

- Fully-connected (“all-to-all”) LTC
- Compare to baseline RNNs

Time series prediction Mean and standard deviation, n=5

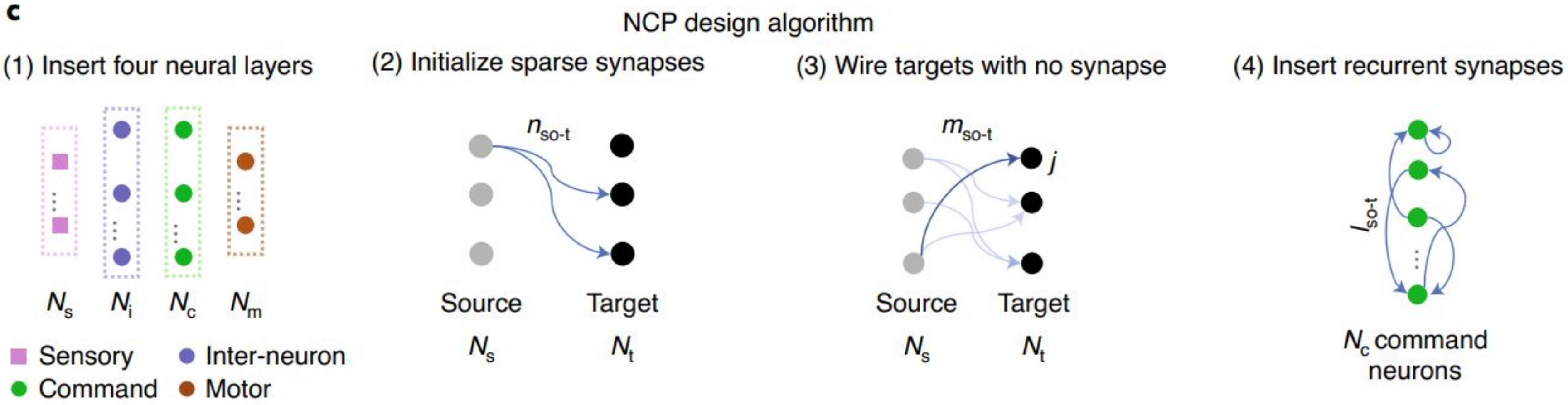
Dataset	Metric	LSTM [28]	CT-RNN [47]	Neural ODE [6]	CT-GRU [38]	LTC (ours)
Gesture	(accuracy)	64.57% ± 0.59	59.01% ± 1.22	46.97% ± 3.03	68.31% ± 1.78	**69.55%** ± **1.13**
Occupancy	(accuracy)	93.18% ± 1.66	94.54% ± 0.54	90.15% ± 1.71	91.44% ± 1.67	**94.63%** ± **0.17**
Activity recognition	(accuracy)	95.85% ± 0.29	95.73% ± 0.47	**97.26**% ± 0.10	96.16% ± 0.39	95.67% ± 0.575
Sequential MNIST	(accuracy)	**98.41**% ± 0.12	96.73% ± 0.19	97.61% ± 0.14	98.27% ± 0.14	97.57% ± 0.18
Traffic	(squared error)	0.169 ± 0.004	0.224 ± 0.008	1.512 ± 0.179	0.389 ± 0.076	**0.099** ± 0.0095
Power	(squared-error)	0.628 ± 0.003	0.742 ± 0.005	1.254 ± 0.149	**0.586** ± 0.003	0.642 ± 0.021
Ozone	(F1-score)	0.284 ± 0.025	0.236 ± 0.011	0.168 ± 0.006	0.260 ± 0.024	**0.302** ± 0.0155

[28] Hochreiter et al. 1997
[47] Rubanova et al. NeurIPS 2019
[6] Chen et al. NeurIPS, 2018
[38] Moser et al. Arxiv, 2017

Hasani*, Lechner*, Amini, Rus, Grosu. *Liquid Time-constant Networks*. AAAI , 2021

Combining the LTC model with structured sparsity:

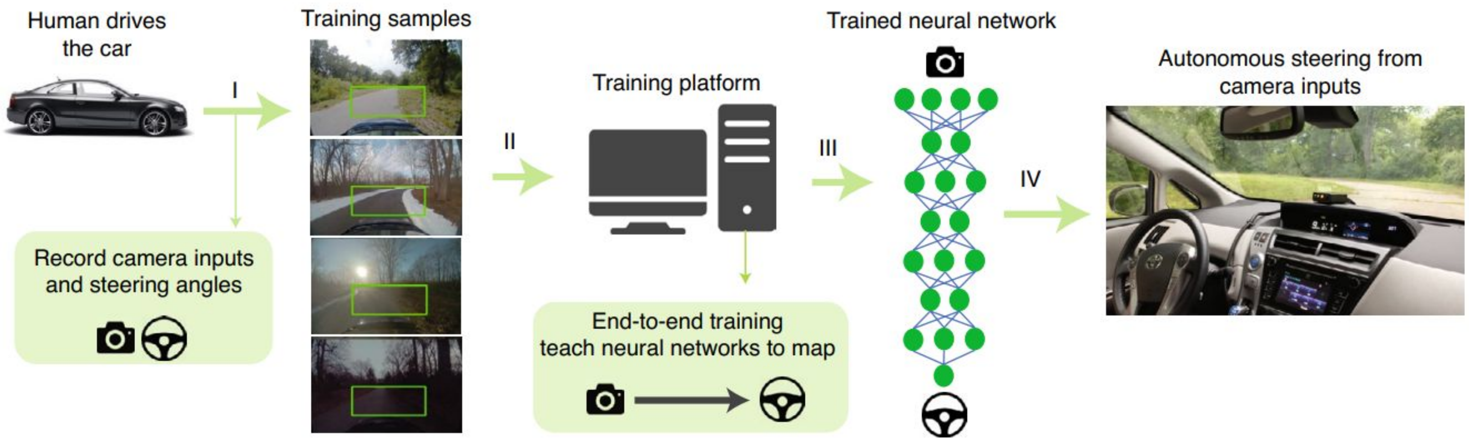
Neural Circuit Policies (NCP)



Lechner*, Hasani*, Amini, Henzinger, Rus, Grosu. *Neural circuit policies enabling auditable autonomy*. Nature Machine Intelligence, 2020

Experiment setup

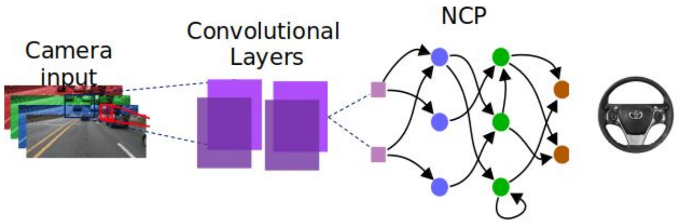
- End-to-end autonomous driving
- Offline open-loop training (supervised learning)
- Online closed-loop test on real car



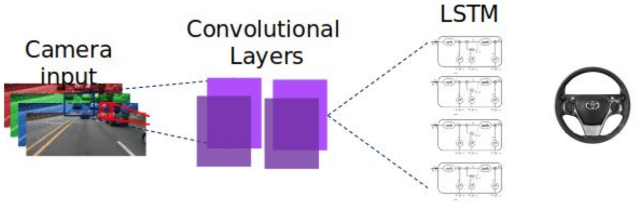
Lechner*, Hasani*, Amini, Henzinger, Rus, Grosu. *Neural circuit policies enabling auditable autonomy*. Nature Machine Intelligence, 2020

Experiment setup

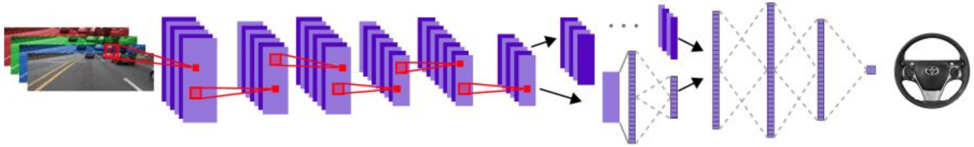
NCPs



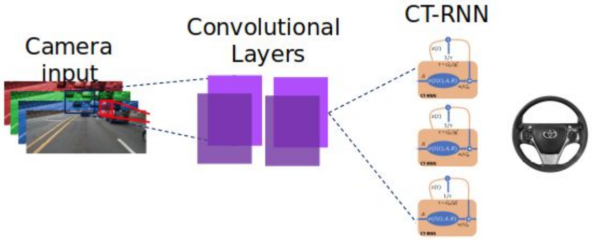
LSTMs



Convolutional Neural Networks (CNNs)

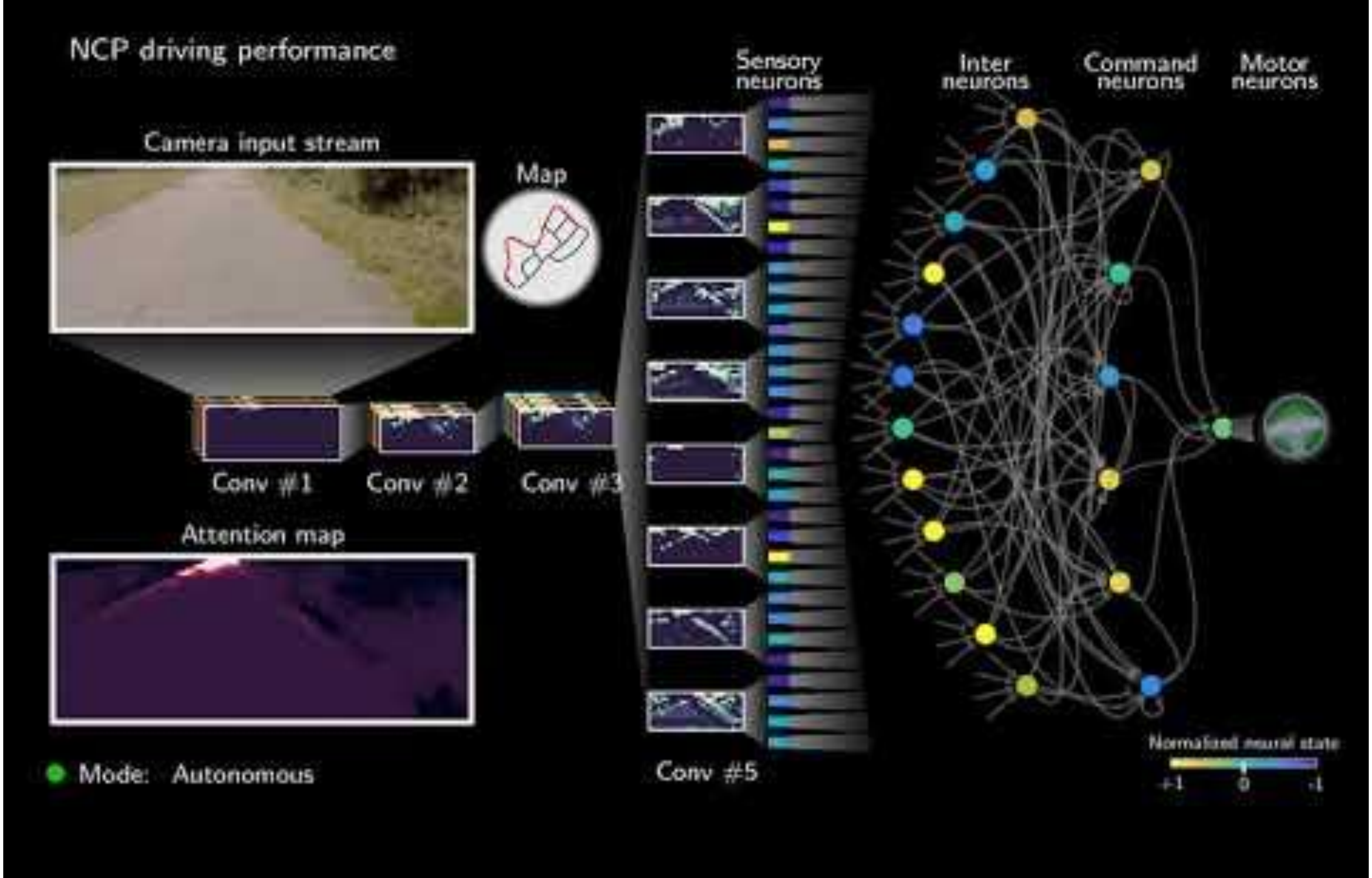


CT-RNNs

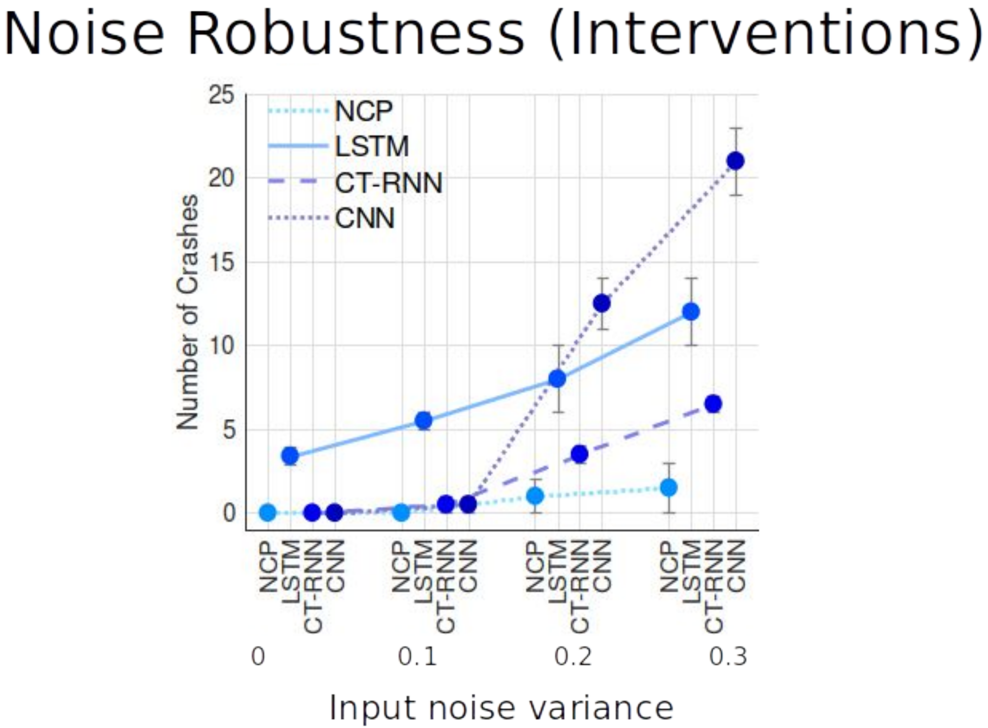


Model	Conv layers Param	RNN neurons	RNN synapses	RNN trainable param
CNN	5,068,900	-	-	-
CT-RNN	79,420	64	6112	6273
LSTM	79,420	64	24640	24897
NCP	79,420	**19**	**253**	**1065**

Lechner*, Hasani*, Amini, Henzinger, Rus, Grosu. *Neural circuit policies enabling auditable autonomy*. Nature Machine Intelligence, 2020

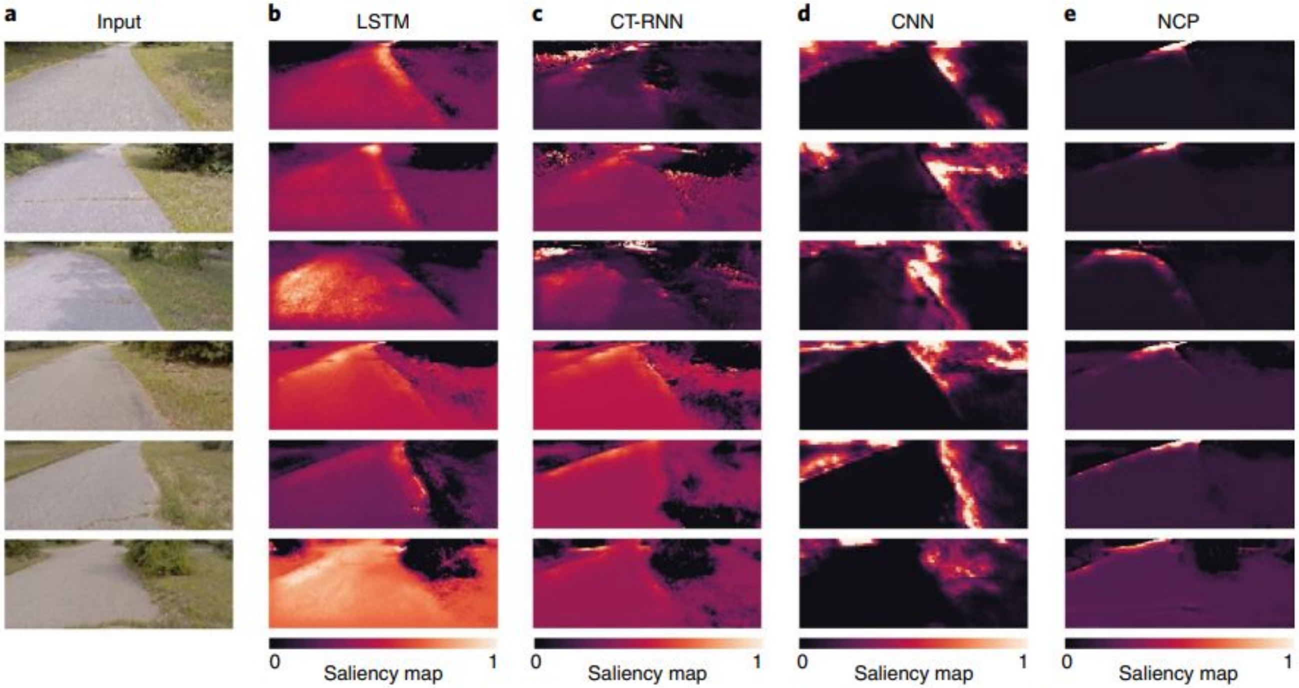

NCP driving performance
Camera input stream
Map
Conv #1
Conv #2
Conv #3
Attention map
Sensory neurons
Conv #5
Inter neurons
Command neurons
Motor neurons
Normalized neural state
+1
0
-1
Mode: Autonomous

Results



Lechner*, Hasani*, Amini, Henzinger, Rus, Grosu. *Neural circuit policies enabling auditable autonomy*. Nature Machine Intelligence, 2020

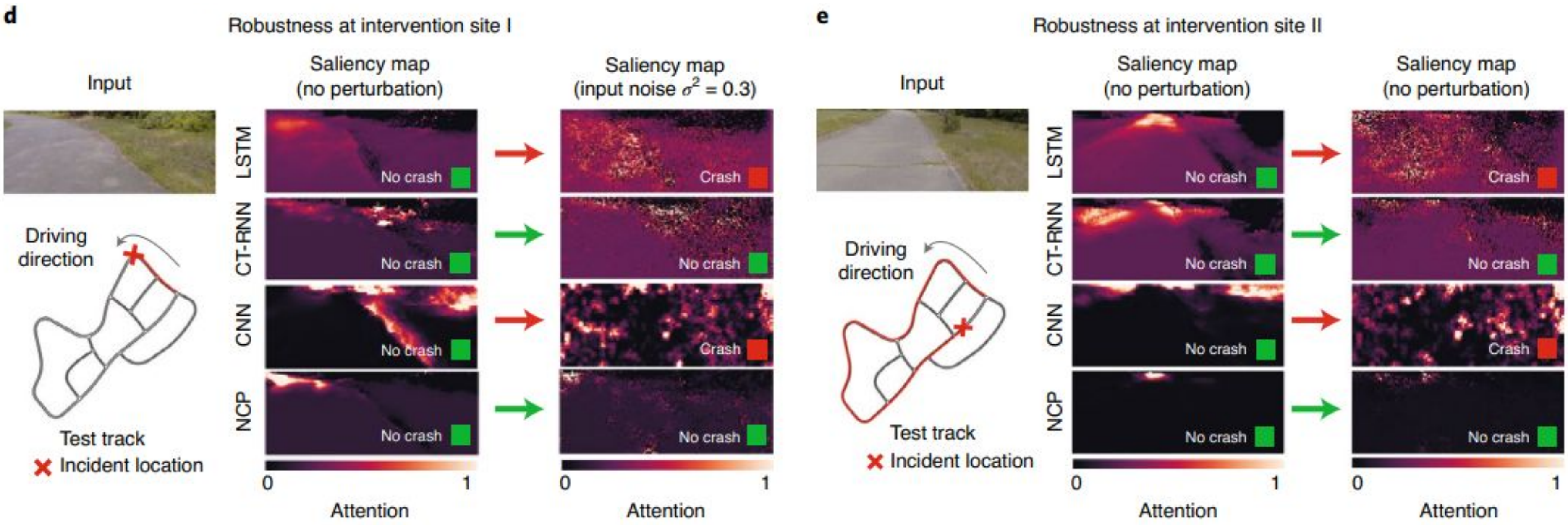
Saliency maps



Saliency maps show where each network is learned to attend while driving

Lechner*, Hasani*, Amini, Henzinger, Rus, Grosu. *Neural circuit policies enabling auditable autonomy*. Nature Machine Intelligence, 2020

Saliency maps



Lechner*, Hasani*, Amini, Henzinger, Rus, Grosu. *Neural circuit policies enabling auditable autonomy*. Nature Machine Intelligence, 2020

Why did the NCP learn a more robust behavior?

Dynamical Causal Models (DCM) [Friston et al., 2003]

Dynamical Causal Models (DCM) [Friston et al., 2003]

- Bilinear approximation of a dynamical system

$$d\mathbf{x}/dt = (A + \mathbf{I}(t)B)\mathbf{x}(t) + C\mathbf{I}(t)$$

$$A = \frac{\partial F}{\partial \mathbf{x}(t)}\bigg|_{I=0}, \quad B = \frac{\partial^2 F}{\partial \mathbf{x}(t)\partial \mathbf{I}(t)}, \quad C = \frac{\partial F}{\partial \mathbf{I}(t)}\bigg|_{x=0},$$

Vorbach*, Hasani*, Amini, Lechner, Rus. *Causal Navigation by Continuous-time Neural Networks*. NeurIPS, 2021

Dynamical Causal Models (DCM) [Friston et al., 2003]

- Bilinear approximation of a dynamical system
- Shown to learn causal structures of brain regions and sequential tasks
  [Breakspear, 2017, Ju and Bassett, 2020, Penny et al., 2005]

$$d\mathbf{x}/dt = (A + \mathbf{I}(t)B)\mathbf{x}(t) + C\mathbf{I}(t)$$

$$A = \frac{\partial F}{\partial \mathbf{x}(t)}\Big|_{I=0}, \quad B = \frac{\partial^2 F}{\partial \mathbf{x}(t)\partial \mathbf{I}(t)}, \quad C = \frac{\partial F}{\partial \mathbf{I}(t)}\Big|_{x=0},$$

Proposition 1: LTCs are dynamical causal models

Vorbach*, Hasani*, Amini, Lechner, Rus. *Causal Navigation by Continuous-time Neural Networks*. NeurIPS, 2021

Experiment setup

- Drone navigation

Vorbach*, Hasani*, Amini, Lechner, Rus. *Causal Navigation by Continuous-time Neural Networks*. NeurIPS, 2021

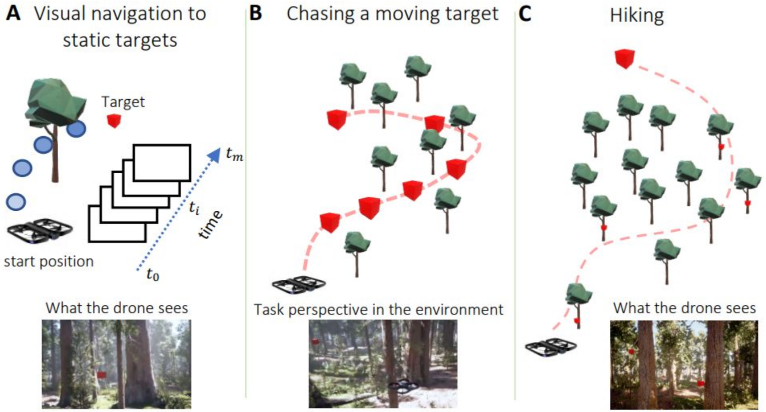
Experiment setup

- Drone navigation
- Photorealistic simulation (Airsim)

Vorbach*, Hasani*, Amini, Lechner, Rus. *Causal Navigation by Continuous-time Neural Networks*. NeurIPS, 2021

Experiment setup

- Drone navigation
- Photorealistic simulation (Airsim)
- 3 tasks



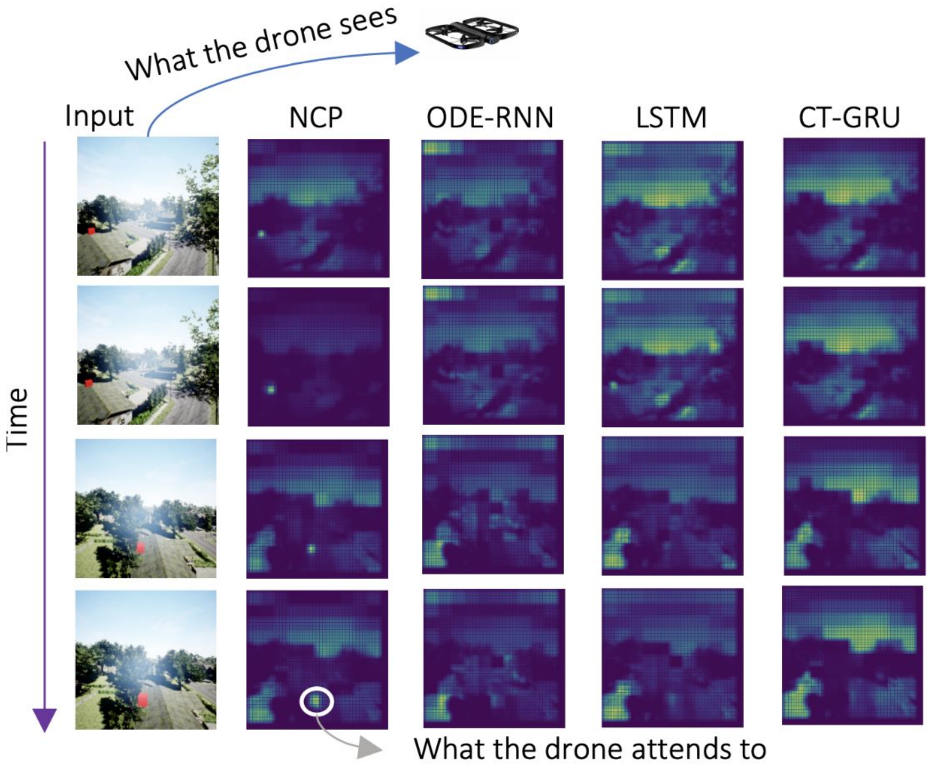
Vorbach*, Hasani*, Amini, Lechner, Rus. *Causal Navigation by Continuous-time Neural Networks*. NeurIPS, 2021

Results

	Static Target					Chasing				Hiking
Model	**Clear**	**Fog**	**Light Rain**	**Heavy Rain**	**Occlusion**	**Clear**	**Fog**	**Light Rain**	**Heavy Rain**	**Clear**
CNN	36%	6%	32%	2%	4%	50%	42%	54%	28%	0%
LSTM	24%	22%	22%	4%	20%	66%	**62%**	56%	44%	2%
ODE-RNN	18%	10%	18%	2%	24%	52%	42%	62%	44%	4%
CT-GRU	40%	8%	**60%**	32%	28%	38%	36%	48%	42%	0%
NCP (ours)	**48%**	**40%**	52%	**60%**	**32%**	**78%**	52%	**84%**	**54%**	**30%**

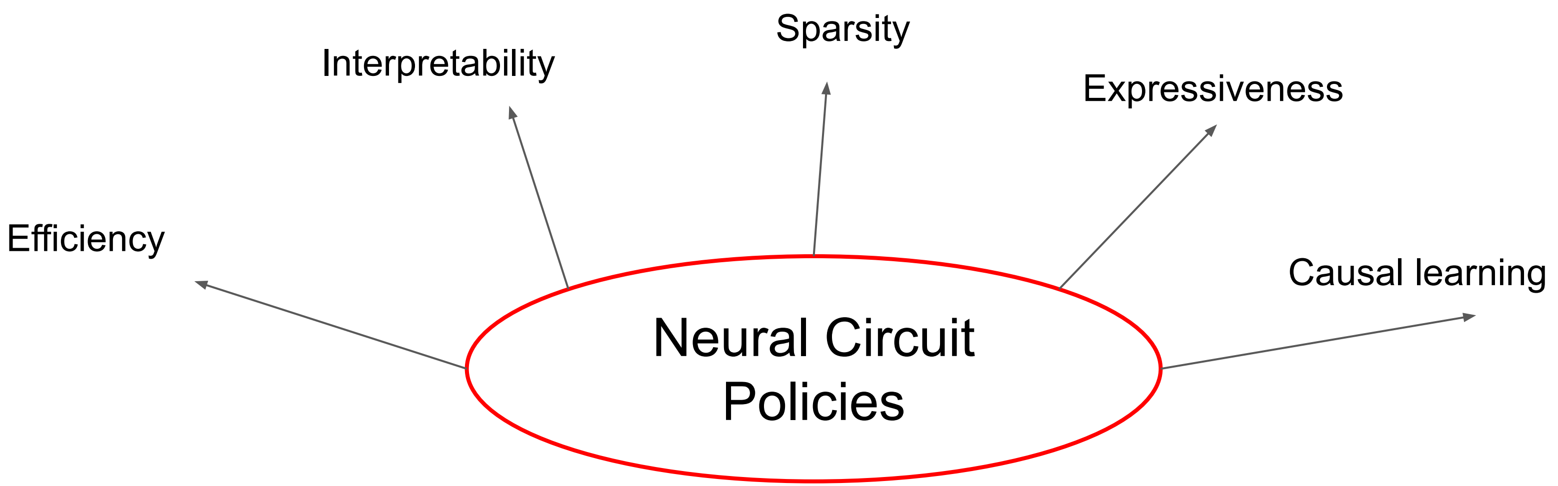
Vorbach*, Hasani*, Amini, Lechner, Rus. *Causal Navigation by Continuous-time Neural Networks*. NeurIPS, 2021

Attention maps



Vorbach*, Hasani*, Amini, Lechner, Rus. *Causal Navigation by Continuous-time Neural Networks*. NeurIPS, 2021

Summary


Sparsity
Interpretability
Expressiveness
Efficiency
Causal learning
Neural Circuit Policies