ICML 2021

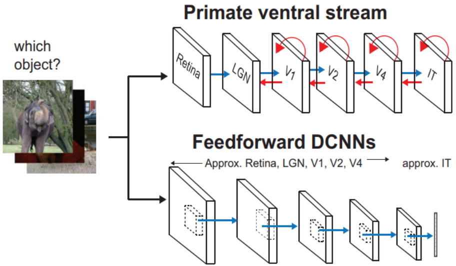
Deep Continuous Networks

Nergis Tömen, Silvia L. Pintea, Jan van Gemert

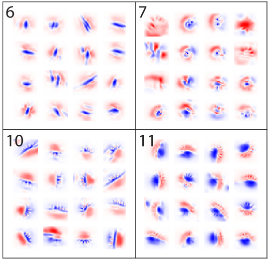
n.tomen@tudelft.nl

Computer Vision Lab, TU Delft

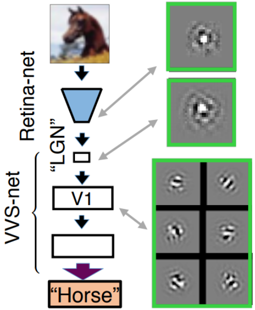
TU Delft Delft University of Technology



Kar et al., Nature Neuroscience, 2019 [9]



Ecker et al., ICLR, 2019 [6]



Lindsey et al., ICLR, 2019 [11]

Motivation

Conventional feed-forward CNNs

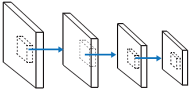
- **Spatially discrete**: use discretized, typically 3×3 kernels 
- cannot learn the receptive field size during training

Conventional feed-forward CNNs

- **Spatially discrete**: use discretized, typically 3×3 kernels

- cannot learn the receptive field size during training

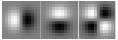
- **Temporally discrete**: use discrete, sequential layers



- cannot model the continuous evolution of neuronal activations

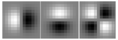
Deep Continuous Networks

We present **Deep Continuous Networks (DCNs)**:

- **spatially continuous** filter descriptions 
- can learn the kernel size and receptive field size during training

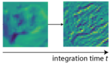
We present **Deep Continuous Networks (DCNs)**:

- **spatially continuous** filter descriptions



- can learn the kernel size and receptive field size during training

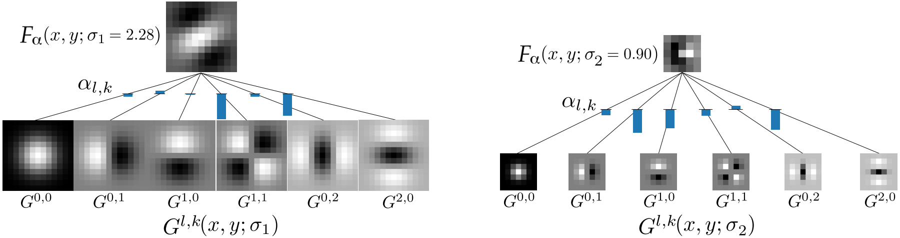
- **depthwise continuous** evolution of feature maps



- can model temporal dynamics of neuronal activations in response to images

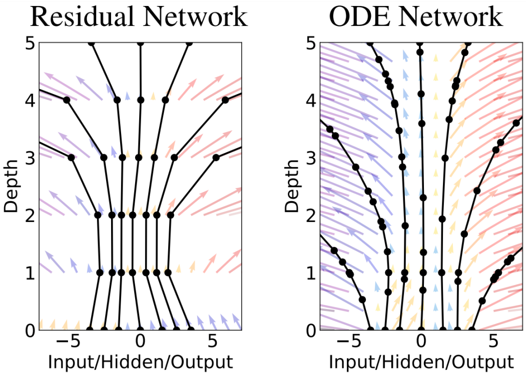
Spatially Continuous Filters

- **Spatially continuous filters**: weighted sum of basis functions
- Gaussian N-jet basis [7, 10, 8] with trainable scale (σ) parameter



Depthwise Continuous Layers

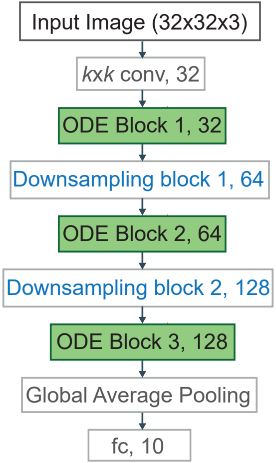
- **Continuous evolution of neural activations** via neural ODEs [4, 12] with continuous depth t



Chen et al., NeurIPS, 2018 [4]

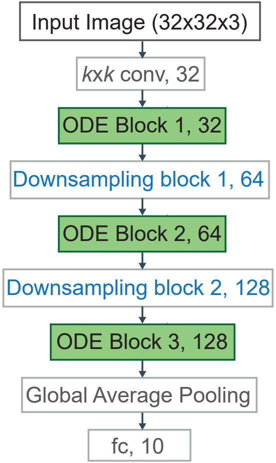
Deep Continuous Networks

- DCN Architecture: Cascade of continuous ODE Blocks
- **ODE Block:**
  - Convolutional filters: Gaussian N-jet
  - Feature maps: Computed by an ODE solver



Deep Continuous Networks

- DCN Architecture: Cascade of continuous ODE Blocks
- **ODE Block:**
  - Convolutional filters: Gaussian N-jet
  - Feature maps: Computed by an ODE solver
- Baseline models
  - ODE-Net (spatially discrete)
  - ResNet-blocks (spatially and depthwise discrete)



DCNs are **parameter efficient** due to the structured filter definitions.

Model	Continuity		Accuracy (%)	Parameters
	Spatial	Temporal		
ODE-Net	x	✓	89.6 ± 0.3	560K
ResNet-blocks	x	x	89.0 ± 0.2	555K
ResNet-SRF-blocks	✓	x	88.3 ± 0.03	426K
ResNet-SRF-full	✓	x	89.3 ± 0.4	323K
DCN-ODE	✓	✓	89.5 ± 0.2	**429K**
DCN-full	✓	✓	89.2 ± 0.3	**326K**
DCN σ^{ji}	✓	✓	89.7 ± 0.3	**472K**

CIFAR-10 Classification: Small-data regime

DCNs are **data efficient**.

Model	# images per class										
	2	4	8	16	32	52	64	103	128	512	1024
ResNet34†	17.5±2.5	19.5±1.4	23.3±1.6	28.3±1.4	33.2±1.2	–	41.7±1.1	–	49.1±1.3	–	–
CNTK†	**18.8**±2.1	**21.3**±1.9	25.5±1.9	30.5±1.2	36.6±0.9	–	42.6±0.7	–	48.9±0.7	–	–
ResNet-blocks	16.7±0.8	19.6±1.0	22.0±1.3	28.1±1.7	35.4±0.9	39.8±0.6	41.6±1.5	49.0±0.2	50.9±0.6	70.4±1.2	76.8±0.7
ODE-Net	16.8±2.8	20.5±0.8	23.1±2.5	29.8±0.8	36.4±1.0	41.7±1.2	42.3±0.2	48.6±0.5	50.7±0.7	71.7±1.5	77.4±0.5
DCN-ODE	16.4±1.6	19.8±0.7	**26.5**±0.9	**31.2**±0.6	**37.7**±0.6	**44.5**±0.8	**48.0**±1.3	**54.2**±0.8	**58.2**±0.7	**75.5**±0.8	**79.7**±0.3

Baseline results † from [2].

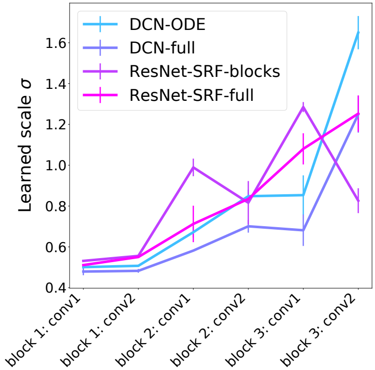
CIFAR-10 Classification: Filter meta-parametrization

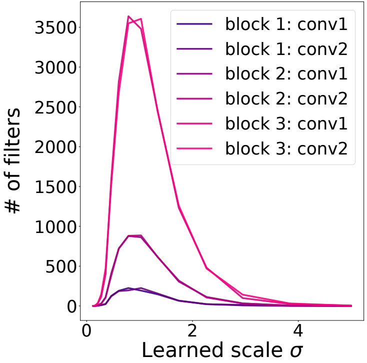
DCNs allow for **meta-parametrization** of filters as a function of depth t.

Model	Parametrization	Accuracy (%)
DCN-ODE	σ, α	89.46 ± 0.16
DCN $\sigma(t)$	$\sigma = 2^{at+b}$, α	89.97 ± 0.30
DCN $\sigma(t^2)$	$\sigma = 2^{at^2+bt+c}$, α	89.93 ± 0.28
DCN $\sigma(t)$, $\alpha(t)$	$\sigma = 2^{a_s t+b_s}$, $\alpha = a_\alpha t + b_\alpha$	89.88 ± 0.25

Learning the receptive field size

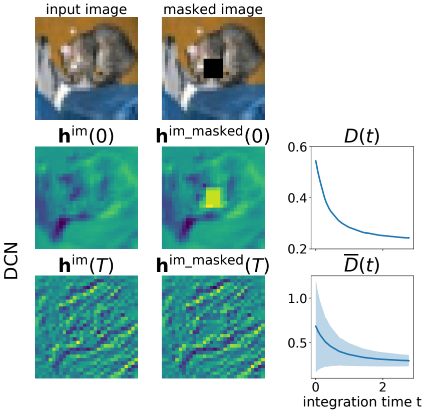
σ distributions after training are **consistent with biological observations** [13].

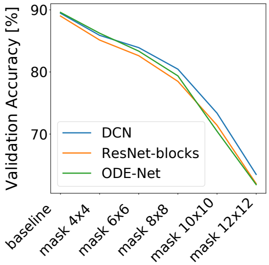




Similar to rate-based, continuous-receptive-field population models of biological vision [3, 1, 5],

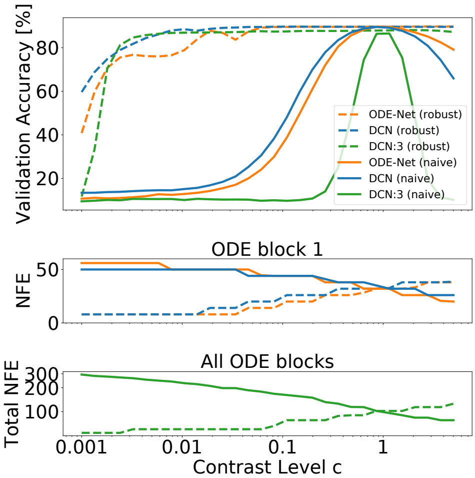
DCNs display **emergent pattern completion**.





Efficient ODE computation via contrast robustness

- ODE-based dynamics are **sensitive to contrast** changes.
- **Contrast robustness** can be improved by scaling the numerical ODE integration time proportionately to input contrast at test time.
- Contrast-robust networks can **cut down the computational cost**.



Conclusion

We present spatially and depthwise-continuous DCN models:

- Data efficient
- Learn biologically plausible RF sizes
- Links to biological models from computational neuroscience
- Computational efficiency via contrast-robustness

Thanks!

Please see the paper for more!

References

[1] Shunichi Amari. “Dynamics of pattern formation in lateral-inhibition type neural fields”. In: *Biological Cybernetics* 27.2 (1977), pp. 77–87.

[2] Sanjeev Arora et al. “Harnessing the Power of Infinitely Wide Deep Nets on Small-data Tasks”. In: *International Conference on Learning Representations (ICLR)*. 2020.

[3] Paul C Bressloff et al. “Geometric visual hallucinations, Euclidean symmetry and the functional architecture of striate cortex”. In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 356.1407 (2001), pp. 299–330.

[4] Ricky T. Q. Chen et al. “Neural Ordinary Differential Equations”. In: *NeurIPS* (2018).

[5] Stephen Coombes. “Waves, bumps, and patterns in neural field theories”. In: *Biological Cybernetics* 93.2 (2005), pp. 91–108.

[6] Alexander S Ecker et al. “A rotation-equivariant convolutional neural network model of primary visual cortex”. In: *International Conference on Learning Representations (ICLR)*. 2019.

[7] Luc Florack et al. “The Gaussian scale-space paradigm and the multiscale local jet”. In: *IJCV* 18.1 (1996), pp. 61–75.

[8] Jorn-Henrik Jacobsen et al. “Structured receptive fields in CNNs”. In: *CVPR*. 2016, pp. 2610–2619.

[9] Kohitij Kar et al. “Evidence that recurrent circuits are critical to the ventral stream’s execution of core object recognition behavior”. In: *Nature Neuroscience* 22.6 (2019), pp. 974–983.

[10] Tony Lindeberg. *Scale-space theory in computer vision*. Vol. 256. Springer Science & Business Media, 2013.

[11] Jack Lindsey et al. “A Unified Theory of Early Visual Representations from Retina to Cortex through Anatomically Constrained Deep CNNs”. In: *International Conference on Learning Representations (ICLR)*. 2019.

[12] Lars Ruthotto and Eldad Haber. “Deep neural networks motivated by partial differential equations”. In: *Journal of Mathematical Imaging and Vision* (2019), pp. 1–13.

[13] Hsin-Hao Yu et al. “Spatial and temporal frequency tuning in striate cortex: functional uniformity and specializations related to receptive field eccentricity”. In: *European Journal of Neuroscience* 31.6 (2010), pp. 1043–1062.