Dendritic Integration Inspired Artificial Neural Networks Capture Data Correlation

Chongming Liu, Jingyang Ma, Songting Li and Douglas Zhou

Institute of Natural Sciences & School of Mathematical Sciences

Shanghai Jiao Tong University

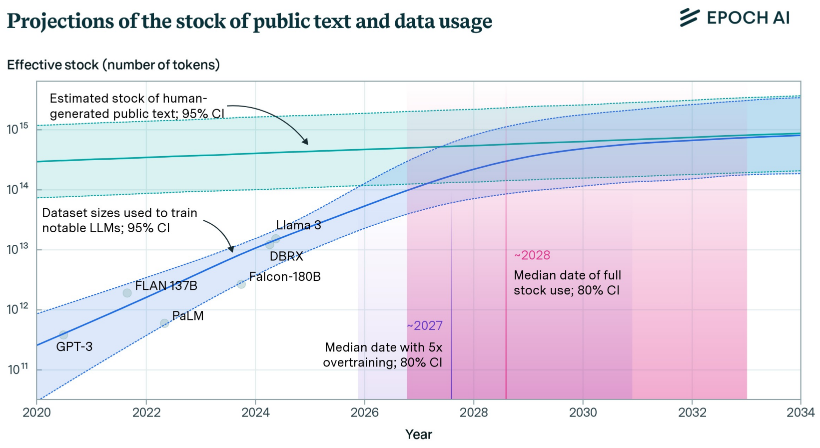
NEURAL INFORMATION PROCESSING SYSTEMS

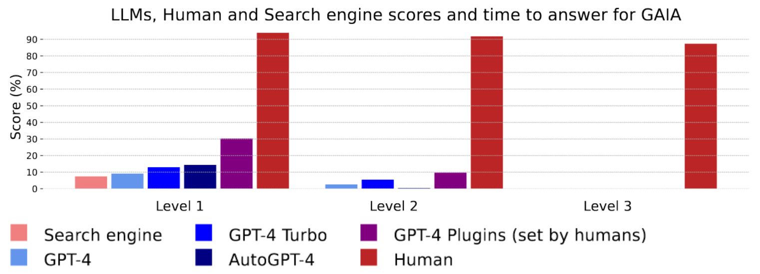
Has AI become powerful enough compare to the human brain?

"We will run out of data between 2026 to 2032"

Huge amount of training data

Villalobos et al. (2024)





GAIA: a benchmark for General AI Assistants

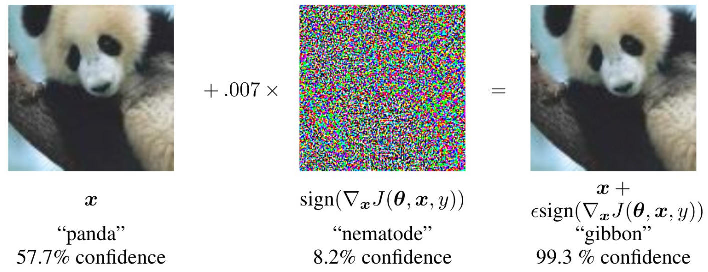
Bad Generalization Capability

Yann Lecun et al. (2023)

No!

Low robustness

Adversarial attacks



Goodfellow et al. (2015)

ChatGPT

GPT-4o~300,000W

High energy comsuption

Human brain~20W

NEURAL INFORMATION PROCESSING SYSTEMS

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

How to solve these problems?

Back to the brain!

nature communications

Perspective https://doi.org/10.1038/s41467-023-37180-x

Catalyzing next-generation Artificial Intelligence through NeuroAI

Received: 11 September 2022

Accepted: 3 March 2023

Published online: 22 March 2023

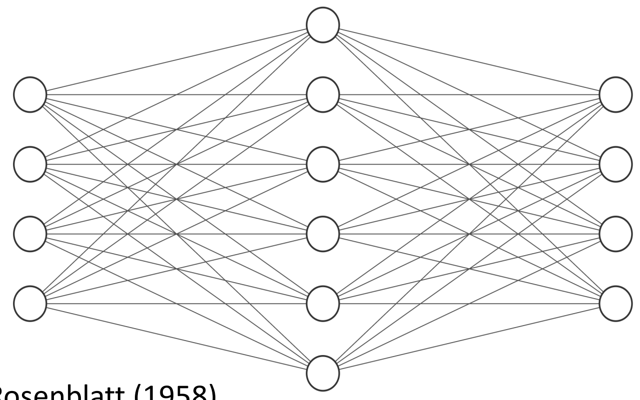
Check for updates

Anthony Zador[1,29], Sean Escola[2,29], Blake Richards[3,4,5,6,7], Bence Ölveczky[8], Yoshua Bengio[3], Kwabena Boahen[9], Matthew Botvinick[10], Dmitri Chklovskii[11], Anne Churchland[12], Claudia Clopath[13], James DiCarlo[14], Surya Ganguli[15], Jeff Hawkins[16], Konrad Körding[17], Alexei Koulakov[1], Yann LeCun[18,19], Timothy Lillicrap[10], Adam Marblestone[20], Bruno Olshausen[21], Alexandre Pouget[22], Cristina Savin[23], Terrence Sejnowski[24], Eero Simoncelli[25], Sara Solla[26], David Sussillo[18,27], Andreas S. Tolias[28] & Doris Tsao[21]

Dendrite: An important part for nonlinear integration

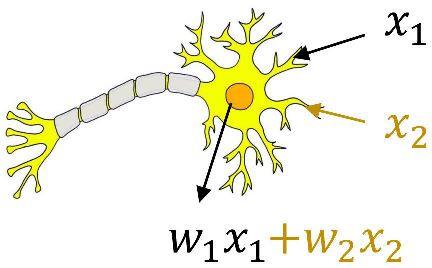
Multilayer-perceptron (MLP)

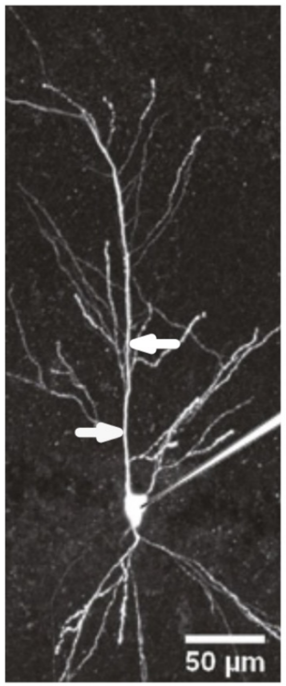
$f(\boldsymbol{x}) = \boldsymbol{w}_L \sigma(\boldsymbol{w}_{L-1} \sigma(\cdots \sigma(\boldsymbol{w}_2 \sigma(\boldsymbol{w}_1 \boldsymbol{x} + \boldsymbol{b}_1) + \boldsymbol{b}_2) \cdots) + \boldsymbol{b}_{L-1})$



Frank Rosenblatt (1958)

Biological interpretation





Dendrite has powerful nonlinearity!

A single neuron with dendrite can do:

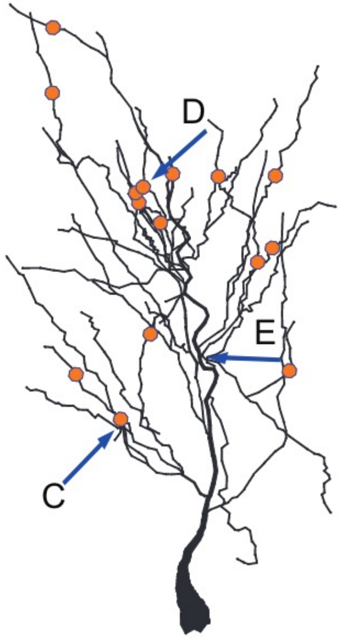
- Logical operation
- Signal amplification
- Parallel nonlinear processing
- …

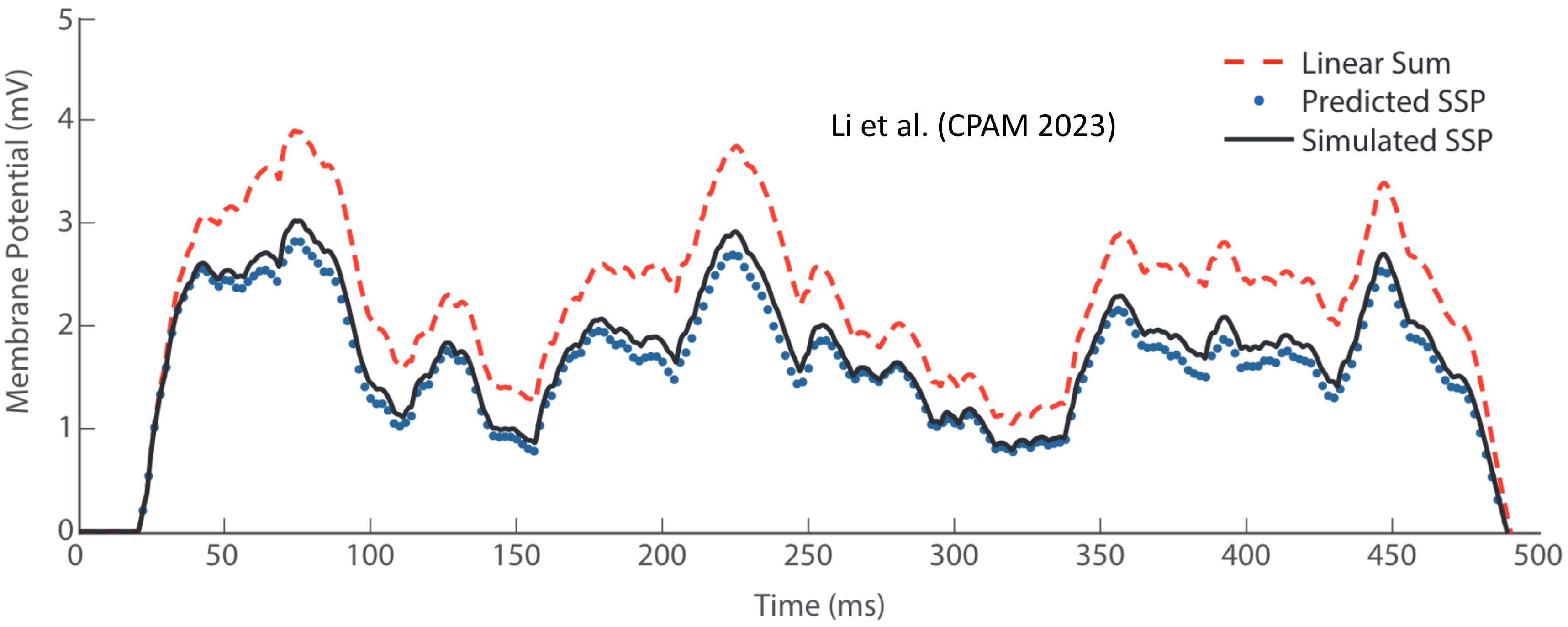
Chavlis, Poirazi (2021)

Questions: Is it possible to incorporate dendritic nonlinearity into ANN?

Need to quantify the dendritic nonlinearity at first.

Dendritic bilinear integration rule





$$Linear\ Sum = \sum_p V_E^p(t) + \sum_q V_I^q(t)$$

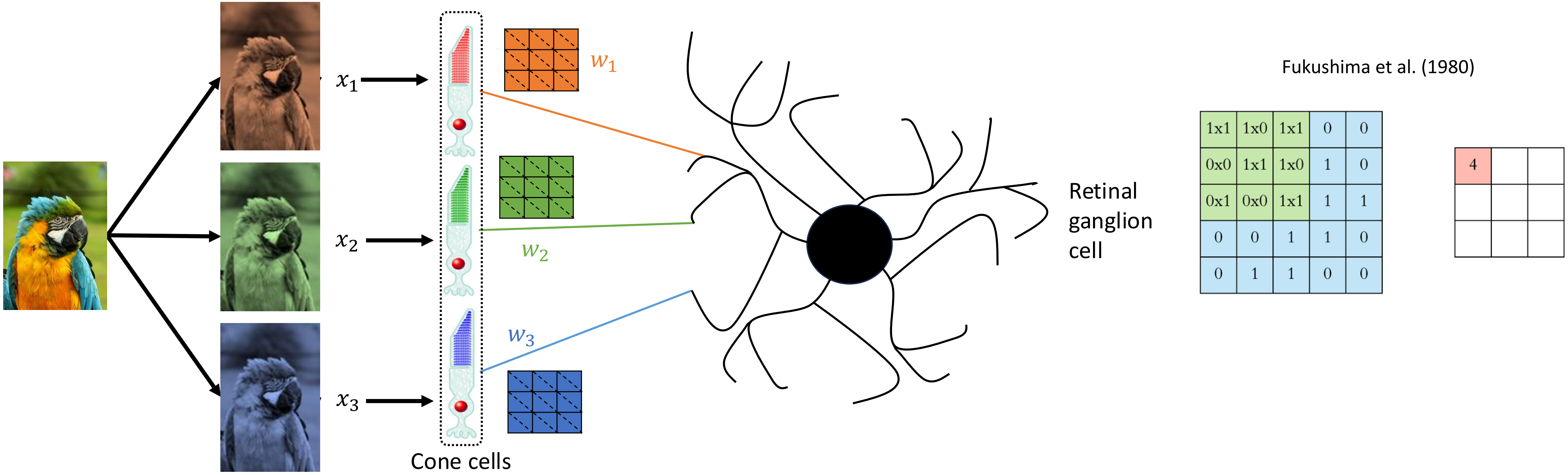
$$Bilinear\ Term = \sum_{i,j} \kappa_{EI}^{ij} V_E^i(t) V_I^j(t) + \sum_{k,l} \kappa_{EE}^{kl} V_E^k(t) V_I^l(t) + \sum_{m,n} \kappa_{II}^{mn} V_I^m(t) V_I^n(t)$$

$$Predicted\ SSP = Linear\ Sum + Bilinear\ Term$$

Given input $\boldsymbol{x}$, Dendritic bilinear integration rule can be simplified as:

$$\mathrm{soma}(\boldsymbol{x}) = (\boldsymbol{w} \odot \boldsymbol{x})^T \boldsymbol{K} (\boldsymbol{w} \odot \boldsymbol{x}) + \boldsymbol{w} \cdot \boldsymbol{x} + \boldsymbol{b}$$

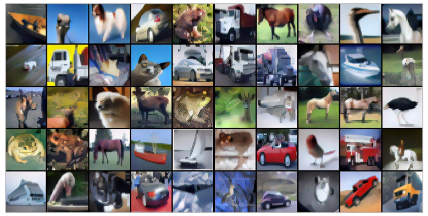
Dit-CNN: Dendritic integration inspired CNN



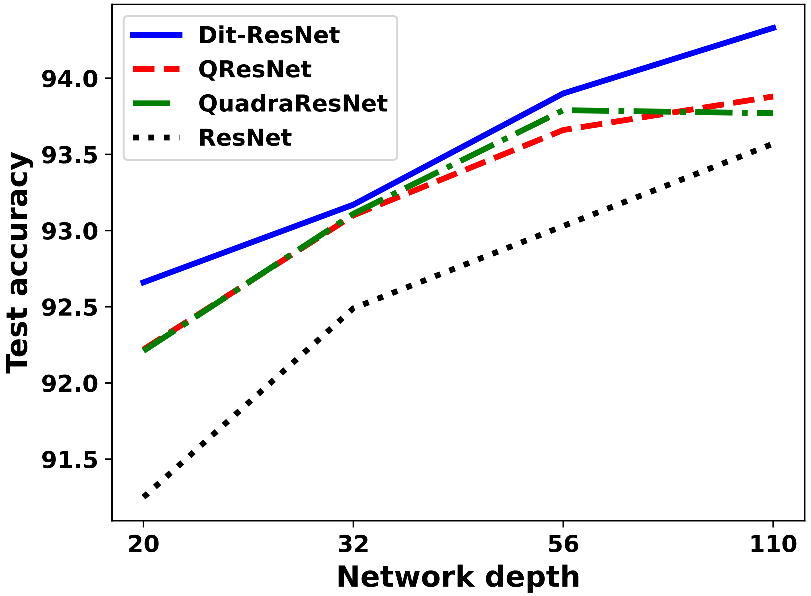
$$quadratic\ neuron(x_1, x_2, x_3) = \sigma(w_1 * x_1 + w_2 * x_2 + w_3 * x_3 + \sum_{i,j=1}^{3} \kappa_{ij}\,(w_i * x_i) \cdot (w_j * x_j))$$

Performance on CIFAR: Dit-CNNs shows significant enhancements

Model	# Param. (CIFAR10)	Acc. (CIFAR10)	# Param. (CIFAR100)	Acc. (CIFAR100)
ResNet-20[16]	0.27M	91.25%	0.30M	67.26±0.68%
QResNet-20[12]	0.81M	92.22%	0.84M	67.82±0.52%
QuadraResNet-20[53]	0.81M	92.21%	0.84M	68.02±0.44%
Dit-ResNet-20	0.30M	**92.66%**	0.33M	**68.66±0.34%**
ResNet-32[16]	0.46M	92.49%	0.49M	68.52±0.55%
QResNet-32[12]	1.39M	93.10%	1.42M	69.41±0.48%
QuadraResNet-32[53]	1.39M	93.11%	1.42M	69.54±0.44%
Dit-ResNet-32	0.49M	**93.17%**	0.52M	**69.68±0.32%**
ResNet-56[16]	0.86M	93.03%	0.89M	70.17±0.67%
QResNet-56[12]	2.55M	93.66%	2.58M	71.21±0.44%
QuadraResNet-56[53]	2.55M	93.79%	2.58M	70.98±0.76%
Dit-ResNet-56	0.89M	**93.90%**	0.92M	**71.40±0.35%**
ResNet-110[16]	1.73M	93.57%	1.76M	70.84±0.76%
QResNet-110[12]	5.17M	93.88%	5.20M	71.58±0.87%
QuadraResNet-110[53]	5.17M	93.77%	5.20M	71.72±0.81%
Dit-ResNet-110	1.76M	**94.33%**	1.79M	**72.40±0.85%**



CIFAR dataset: 60000 32×32 training images, 10000 test images

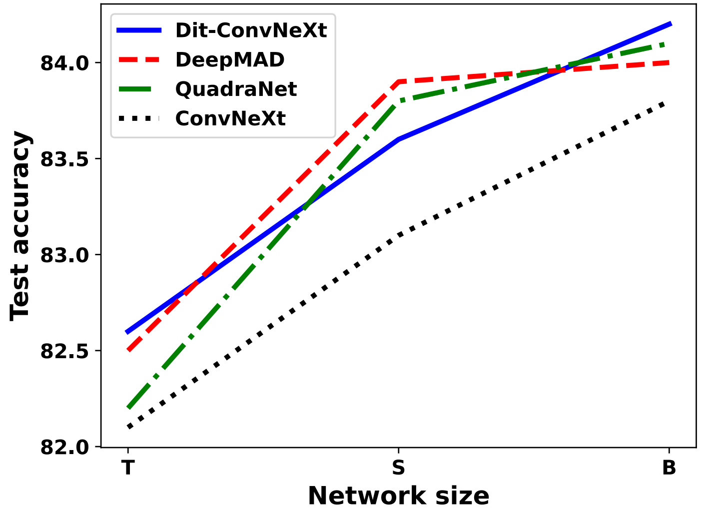


Performance on ImageNet-1K: Dit-CNNs compete favorably with state-of-the-art models

Table 3: Dit-ConvNeXts versus state-of-the-art (SOTA) models on ImageNet-1K. All models listed in the table are trained and validated at a resolution of 224 × 224.

Arch.	Model	# Param.	FLOPs	Top-1 acc. (%)
Transformers	Swin-T [32]	29M	4.5G	81.3
	DeiT-S [45]	22M	4.6G	79.8
State Space Models	VMamba-T [31]	22M	5.6G	82.2
	VideoMamba-S [26]	26M	4.3G	81.2
CNNs	ResNet-50 [50]	26M	4.1G	80.4
	SLaK-T [30]	30M	5.0G	82.5
	QuadraNet36-T [52]	24M	4.1G	82.2
	DeepMAD-29M [41]	29M	4.5G	82.5
	ConvNeXt-T [33]	29M	4.5G	82.1
	Dit-ConvNeXt-T	29M	5.0G	**82.6**
Transformers	Swin-S [32]	50M	8.7G	83.0
State Space Models	VMamba-S [31]	44M	11.2G	83.5
	VideoMamba-M [26]	74M	12.7G	82.8
CNNs	ResNet-101 [50]	45M	7.8G	81.5
	ResNet-152 [50]	60M	11.5G	82.0
	SLaK-S [30]	55M	9.8G	83.8
	QuadraNet36-S [52]	50M	8.9G	83.8
	DeepMAD-50M [41]	50M	8.7G	**83.9**
	ConvNeXt-S [33]	50M	8.7G	83.1
	Dit-ConvNeXt-S	50M	9.2G	83.6
Transformers	Swin-B [32]	88M	15.4G	83.5
	DeiT-B [45]	87M	17.6G	81.8
State Space Models	VMamba-B [31]	75M	18.0G	83.7
CNNs	SLaK-B [30]	95M	17.1G	84.0
	QuadraNet36-B [52]	90M	15.8G	84.1
	DeepMAD-89M [41]	89M	15.4G	84.0
	ConvNeXt-B [33]	89M	15.4G	83.8
	Dit-ConvNeXt-B	90M	16.7G	**84.2**

ImageNet: Large scale high resolution image (224×224) dataset with 1.28M training images and 50K test images from 1000 classes.



Theorem: Quadratic neuron always achieves optimal solution by capture data correlation

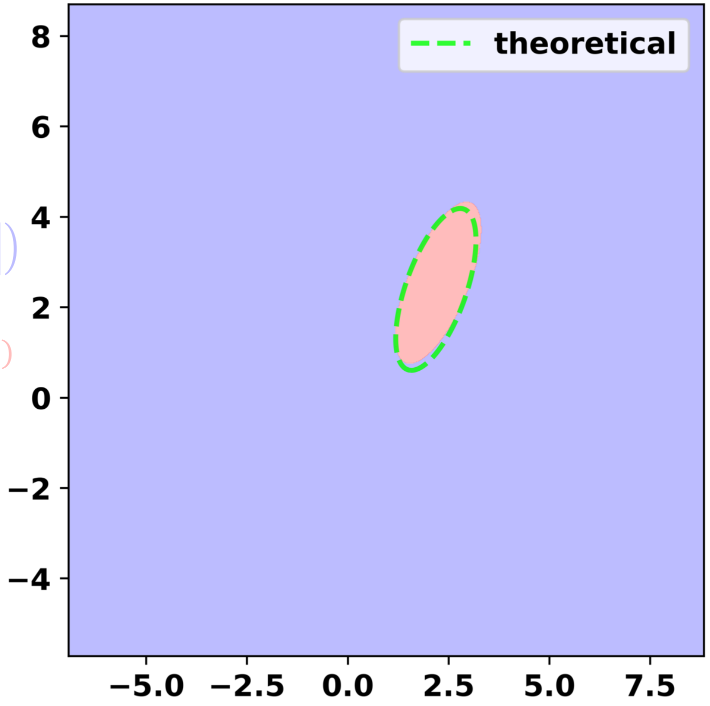
$\text{Class}_1 \sim N(\mu_1, \Sigma_1)$ $\text{Class}_2 \sim N(\mu_2, \Sigma_2)$

Theoretically optimal solution:

$$y_{opt}(x) = argmax_{j\in\{1,2\}} p_j(x)$$

$N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}2.1 & 0.4\\0.4 & 2.0\end{bmatrix}\right)$

$N\left(\begin{bmatrix}2\\2\end{bmatrix}, \begin{bmatrix}0.1 & 0.1\\0.1 & 0.3\end{bmatrix}\right)$



Theorem: Quadratic neuron always achieves optimal solution by capture data correlation

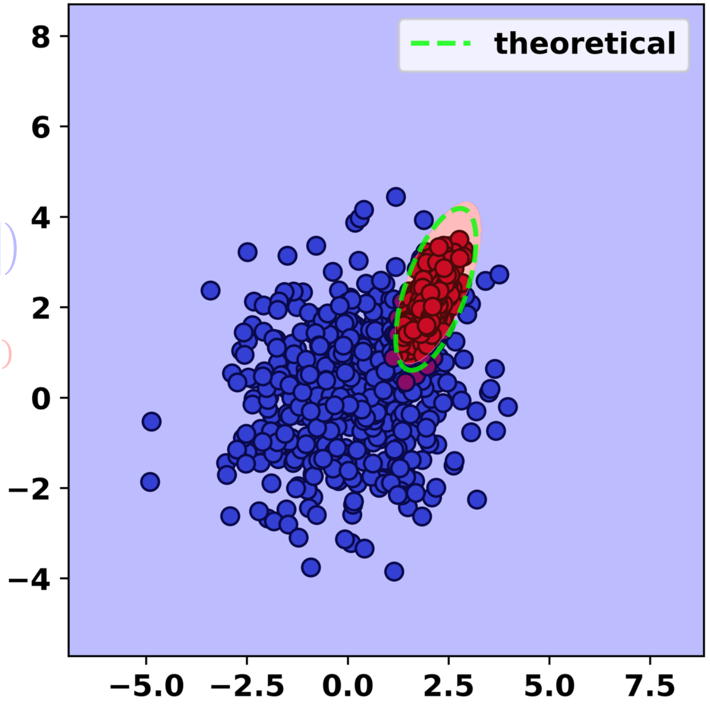
$Class_1 \sim N(\mu_1, \Sigma_1)$ $Class_2 \sim N(\mu_2, \Sigma_2)$

Theoretically optimal solution:

$$y_{opt}(x) = argmax_{j \in \{1,2\}} p_j(x)$$

$$N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}2.1 & 0.4\\0.4 & 2.0\end{bmatrix}\right)$$

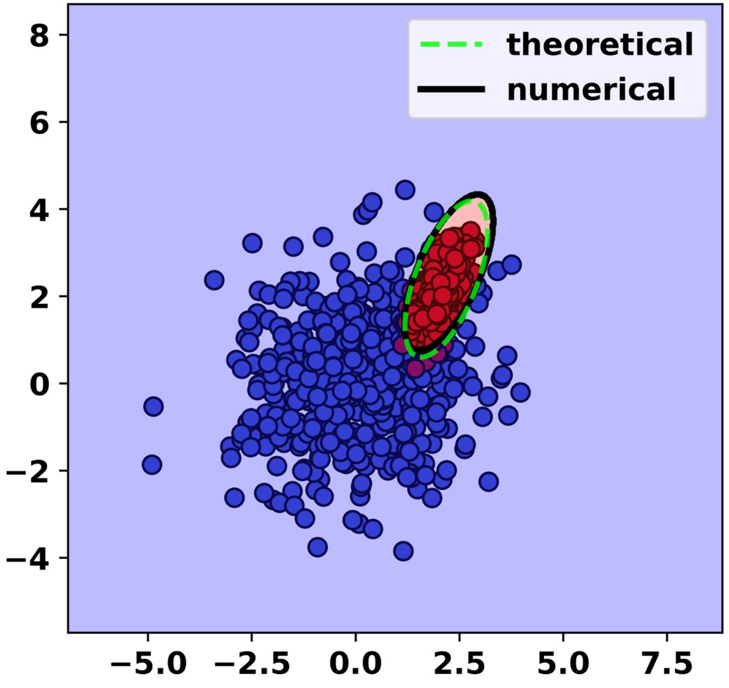
$$N\left(\begin{bmatrix}2\\2\end{bmatrix}, \begin{bmatrix}0.1 & 0.1\\0.1 & 0.3\end{bmatrix}\right)$$



Theorem (informal, Liu et al 2024). A single quadratic neuron can always achieve the theoretically optimal solution.

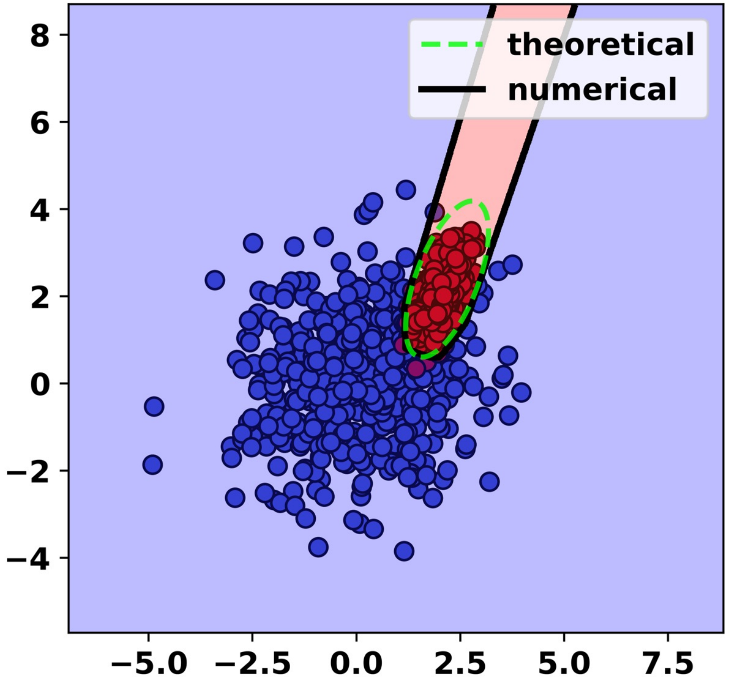
A quadratic neuron

training samples: 1000



MLP with two-layer

training samples: 1000



Theorem: Quadratic neuron always achieves optimal solution by capture data correlation

Theorem 5. *(Existence) The critical points with respect to the cross-entropy loss $L(\theta)$ are given as follows:*

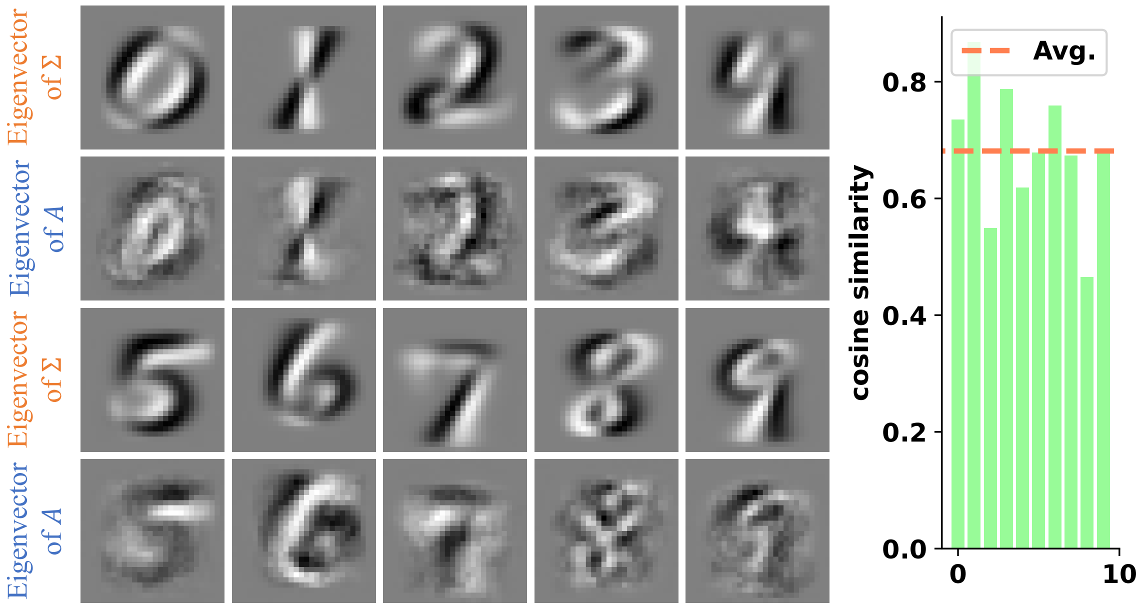
$$A_j^* = \Sigma_j^{-1},\ w_j^* = -2\Sigma_j^{-1}\mu_j,\ b^* = \mu_j^T \Sigma_j^{-1}\mu_j + \log(|\Sigma_j|),\ j \in [k],$$

where

$$L(\theta) = \frac{1}{k}\sum_{j=1}^{k} E_{x\sim class_j}\left[\log\left(1 + \sum_{i=1, i\neq j}^{k} e^{f_i(x) - f_j(x)}\right)\right].$$

Moreover, the corresponding classier generated by this formula is the same as the theoretically optimal classifier as defined in Equation (6).

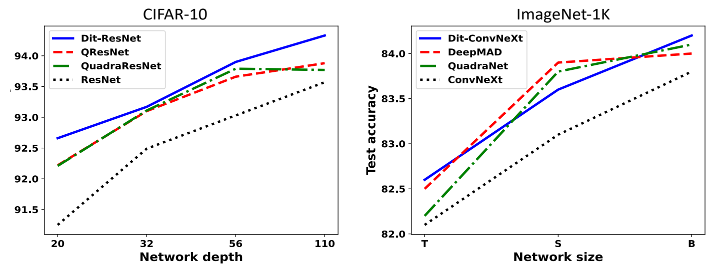
Quadratic coefficients equal to the covariance matrice of data

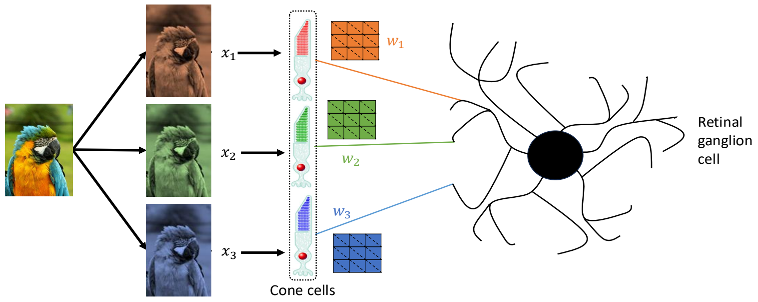


Similarity between quadratic coefficient and covariance matrice of data shows the capability of quadratic neurons to capture correlations.

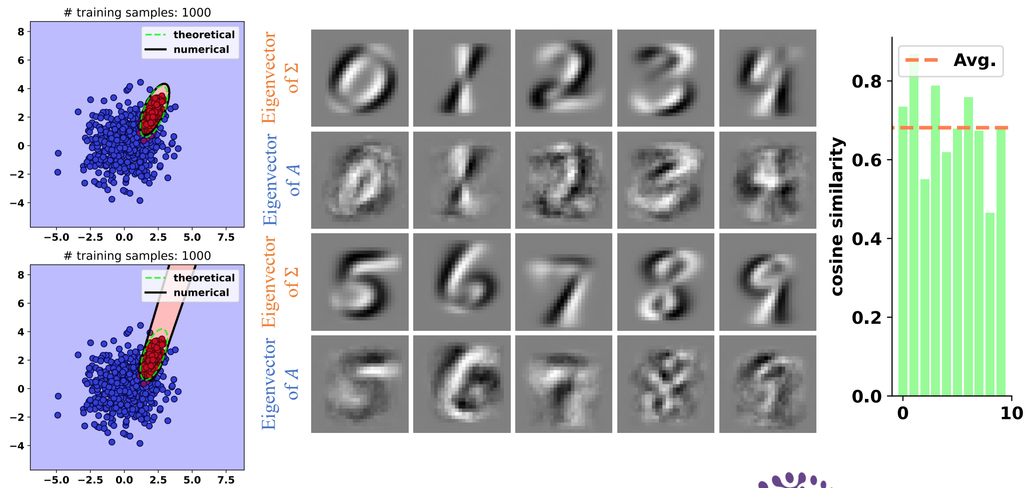
Summary

- We propose a new type of brain-inspired artificial neural network by incorporating dendritic bilinear integration rule into CNNs, which compete favorably with state-of-the-art models.





- Our analysis shows that the superior generalization capability of quadratic neurons stems from their inherent ability to capture data correlations.



Thanks!