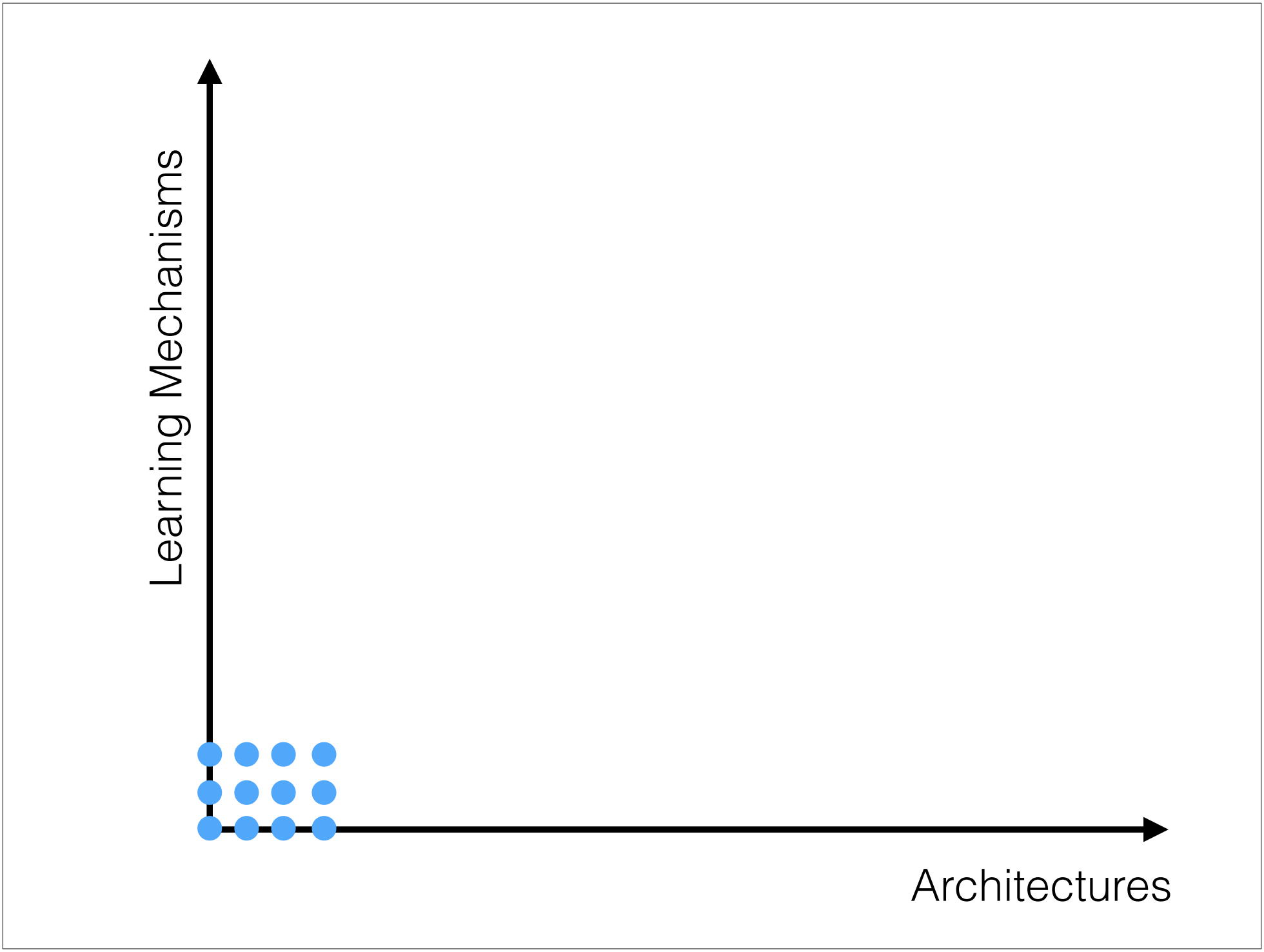
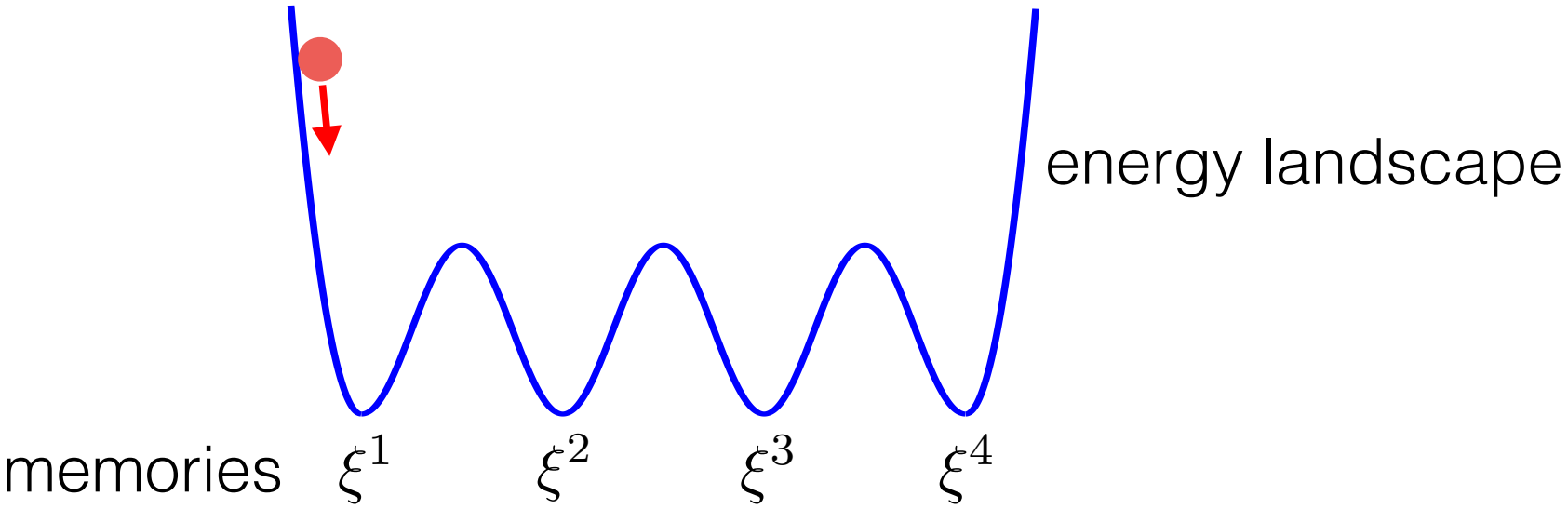
Dense Associative Memories and Deep Learning

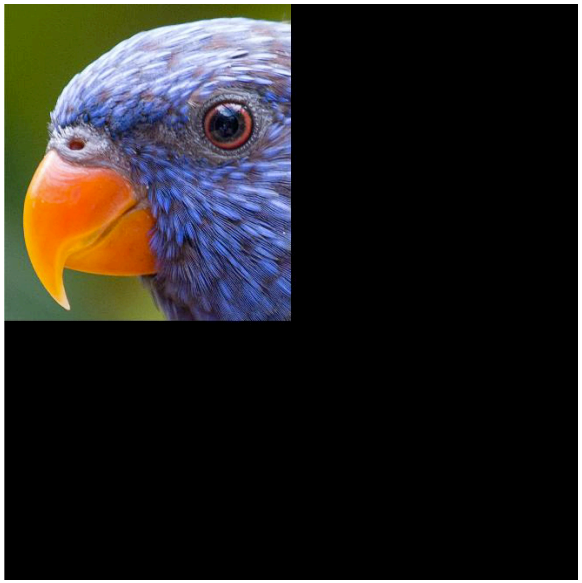
Dmitry Krotov

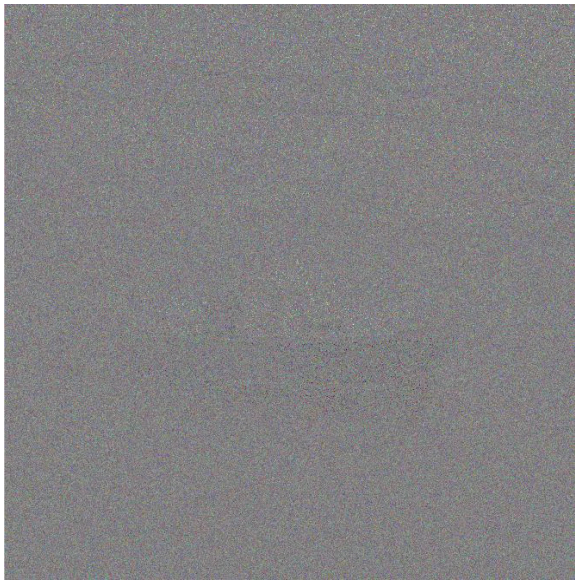
IBM Research
MIT-IBM Watson AI Lab
Institute for Advanced Study


Learning Mechanisms
Architectures

What is associative memory?







Standard Associative Memory

$$E = -\sum_{i,j=1}^{N} \sigma_i T_{ij} \sigma_j \qquad T_{ij} = \sum_{\mu=1}^{K} \xi_i^\mu \xi_j^\mu$$

σ_i -dynamical variables

ξ_i^μ -memorized patterns

N -number of neurons

K -number of memories

$$E = -\sum_{\mu=1}^{K} \left(\sum_{i=1}^{N} \xi_i^\mu \sigma_i \right)^2$$

$$K^{\max} \approx 0.14N$$

Dense Associative Memory

$$E = -\sum_{\mu=1}^{K} \left(\sum_{i=1}^{N} \xi_i^\mu \sigma_i \right)^n$$

$n \geq 2$

power of the interaction vertex

$$K^{\max} \approx \alpha_n N^{n-1}$$

$$\sigma_i^{(t+1)} = Sign\Big[\sum_{\mu=1}^{K}\Big(F\Big(\xi_i^\mu + \sum_{j\neq i}\xi_j^\mu\sigma_j^{(t)}\Big) - F\Big(-\xi_i^\mu + \sum_{j\neq i}\xi_j^\mu\sigma_j^{(t)}\Big)\Big)\Big]$$

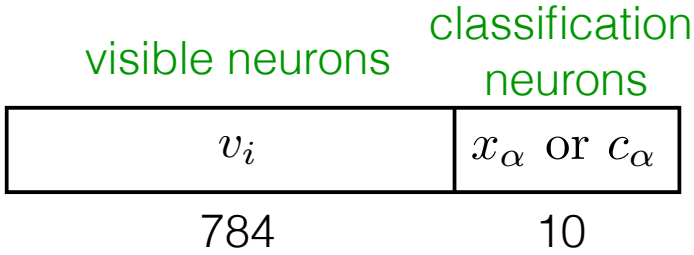
$$\langle \xi_i^\mu \rangle = 0$$

$$\langle \xi_i^\mu \xi_j^\nu \rangle = \delta^{\mu\nu}\delta_{ij}$$

Pattern recognition with DAM

$v_i =$  28

28



$$\sigma_i^{(t+1)} = Sign\Big[\sum_{\mu=1}^{K}\Big(F\Big(\xi_i^\mu + \sum_{j\neq i}\xi_j^\mu\sigma_j^{(t)}\Big) - F\Big(-\xi_i^\mu + \sum_{j\neq i}\xi_j^\mu\sigma_j^{(t)}\Big)\Big)\Big]$$

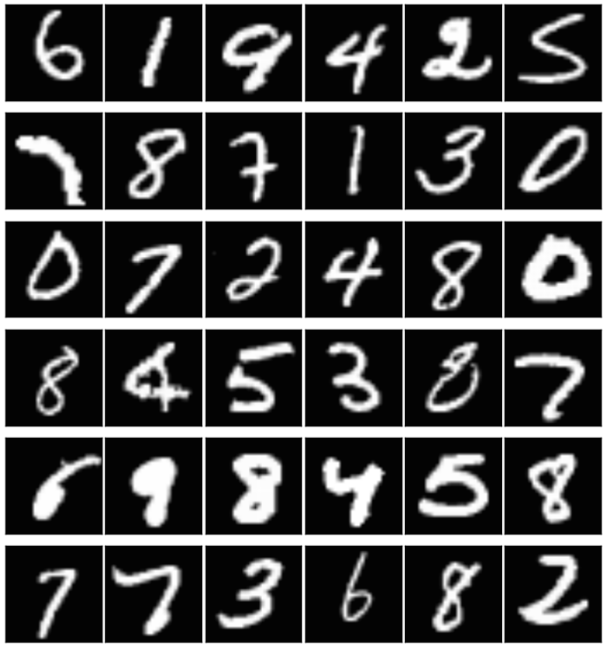
$$c_\alpha = g\Big[\beta\sum_{\mu=1}^{K}\Big(F\Big(-\xi_\alpha^\mu x_\alpha + \sum_{\gamma\neq\alpha}\xi_\gamma^\mu x_\gamma + \sum_{i=1}^{N}\xi_i^\mu v_i\Big) - F\Big(\xi_\alpha^\mu x_\alpha + \sum_{\gamma\neq\alpha}\xi_\gamma^\mu x_\gamma + \sum_{i=1}^{N}\xi_i^\mu v_i\Big)\Big)\Big]$$

$$g(x) = \tanh(x)$$

training

random memories

$\xi_i^\mu \in \mathcal{N}(0, 0.1)$



constructed memory vectors

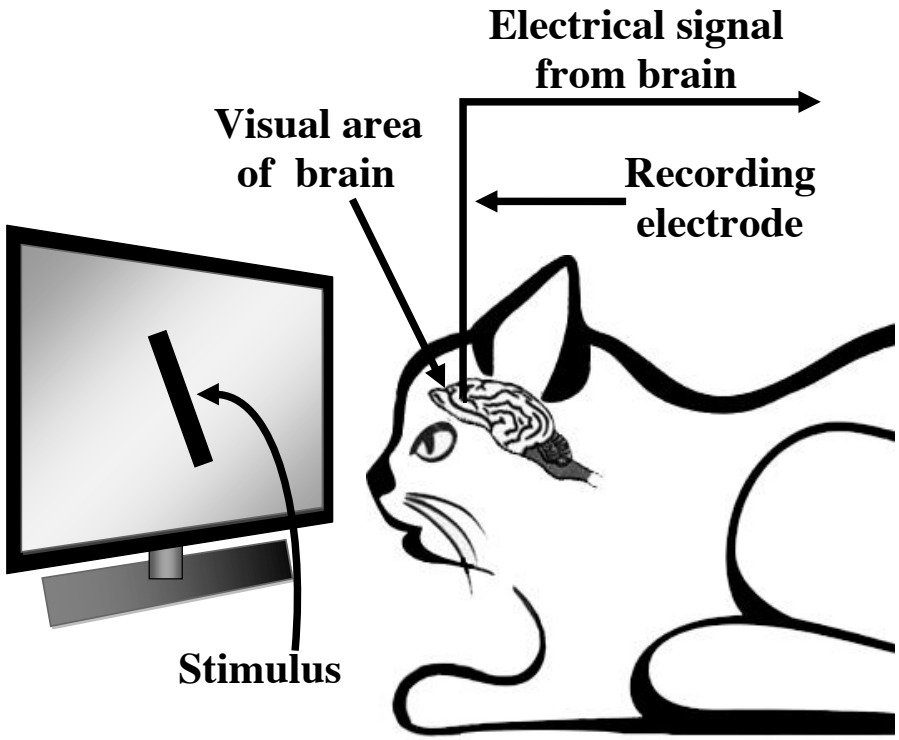
MNIST Dataset

Main question:

What kind of representation of the data has the neural network learned?

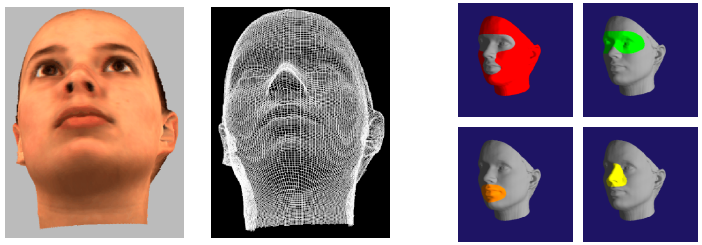
Features vs. prototypes in psychology and neuroscience

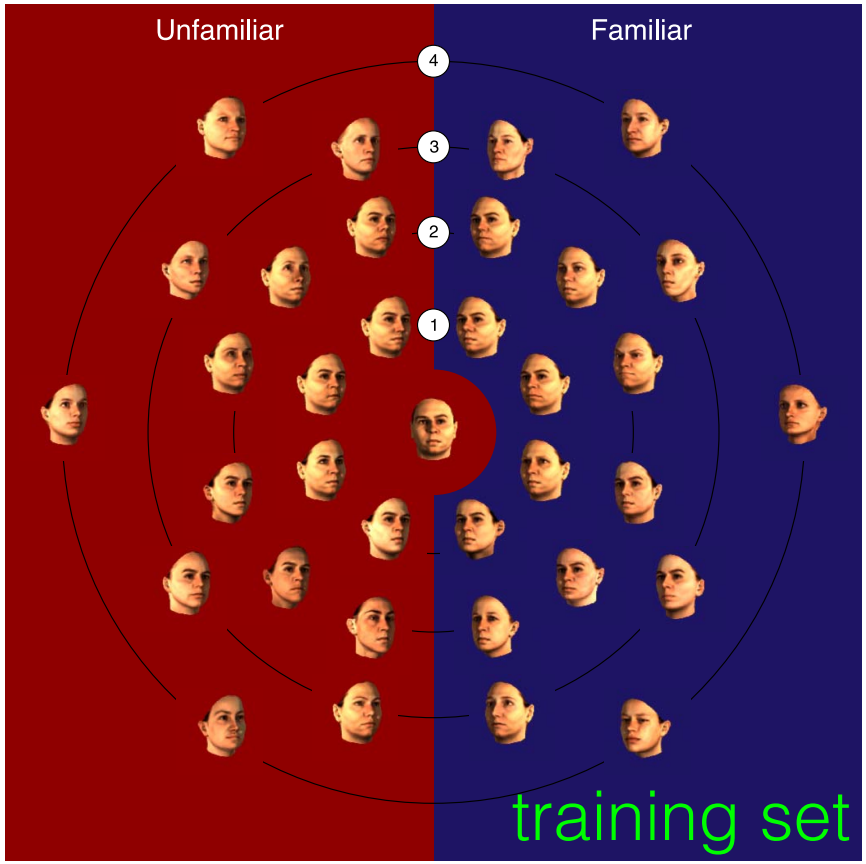
Feature-matching theory



Hubel,Wiesel, 1959

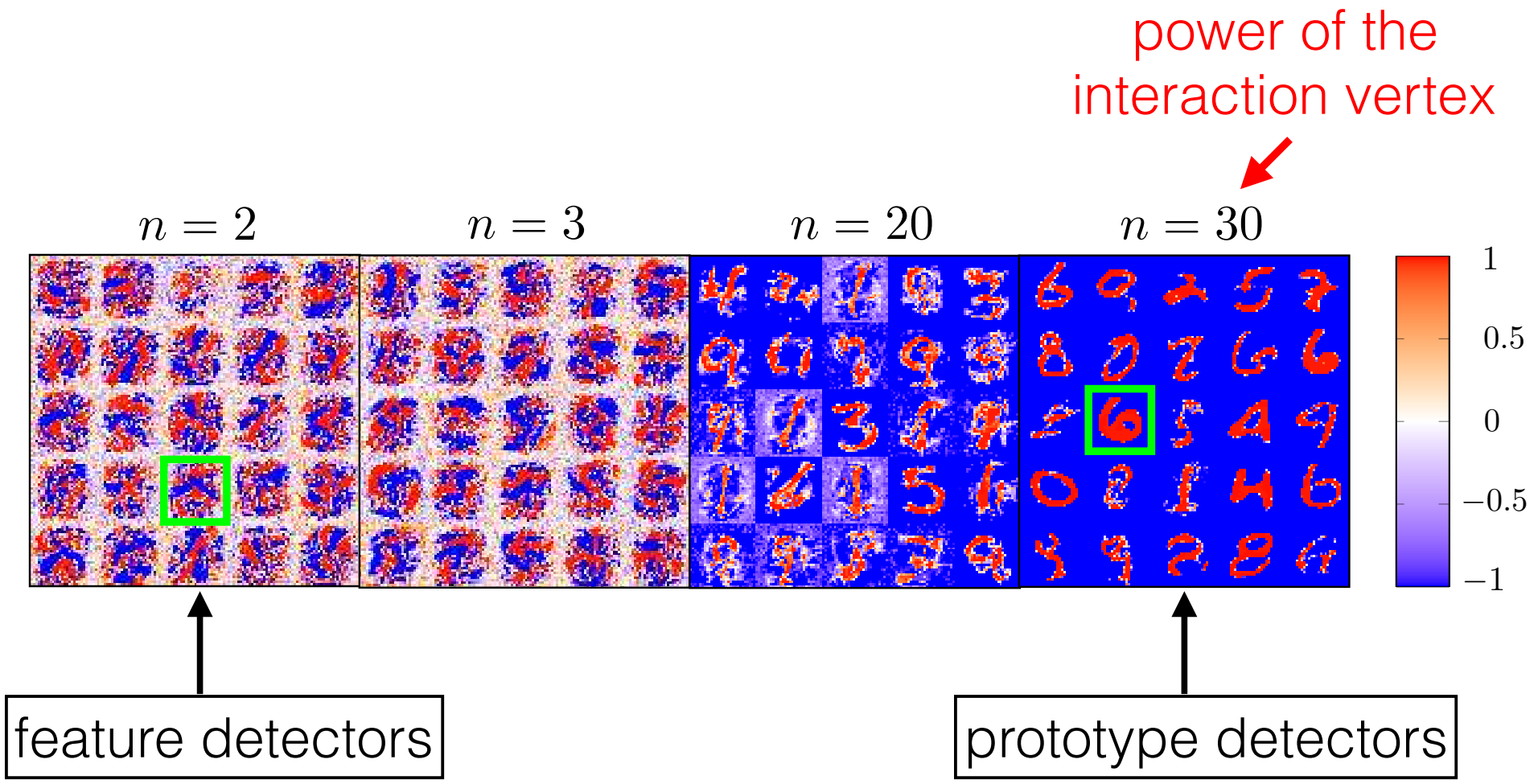
Prototype theory





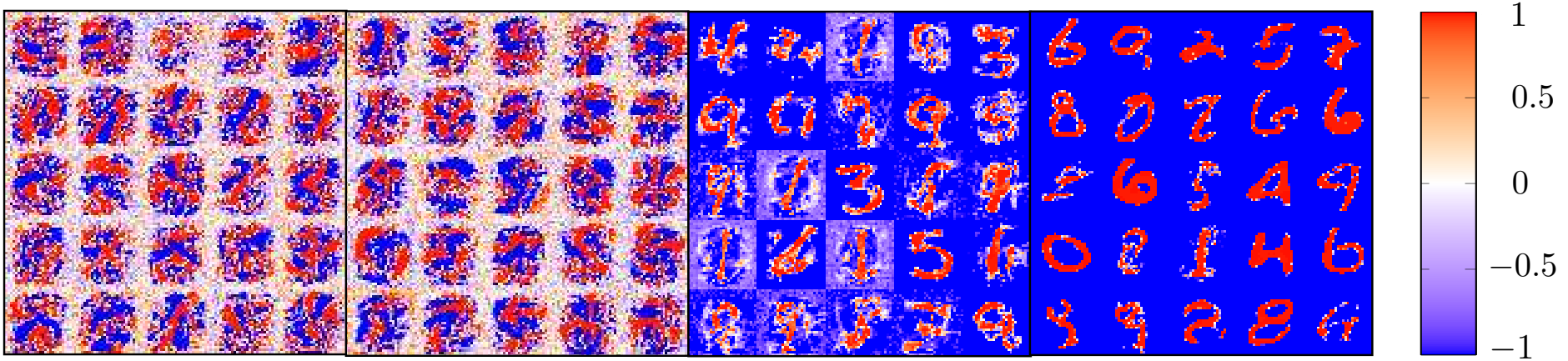
Solso, McCarthy,1981
Wallis, et al., Journal of Vision,2008

Feature to prototype transition



Feature to prototype transition

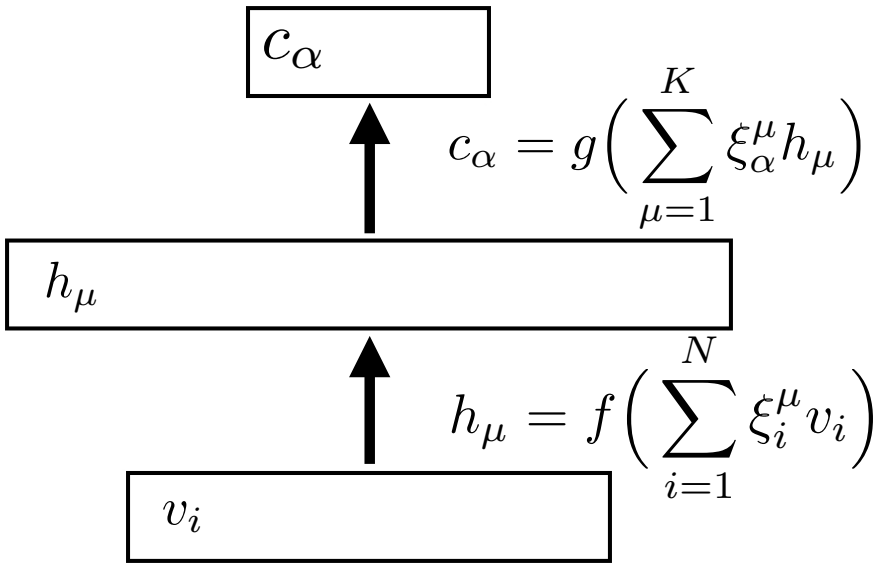
power of the interaction vertex

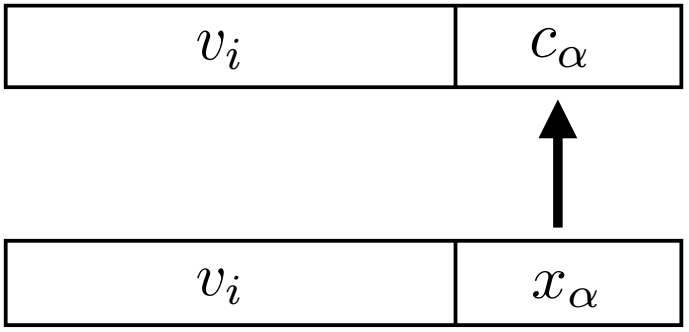


1.51% 1.44% 1.61% 1.80%

1.6% Simard, Steinkraus, Platt, 2003

Duality with feed-forward nets





$$E = -\sum_{\mu=1}^{K} F\Big(\sum_{i=1}^{N} \xi_i^\mu v_i + \sum_{\alpha=1}^{10} \xi_\alpha^\mu c_\alpha\Big)$$

Duality rule:

$$f(x) = F'(x)$$

activation function

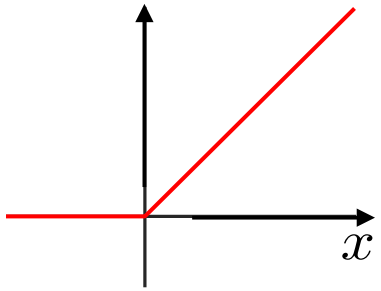
energy function

Commonly used activation functions

$n = 2$

standard Hopfield net

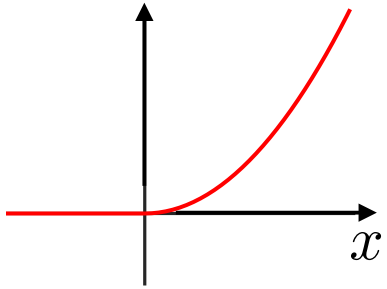
$f(x) = \mathrm{ReLU}$



n

DAM

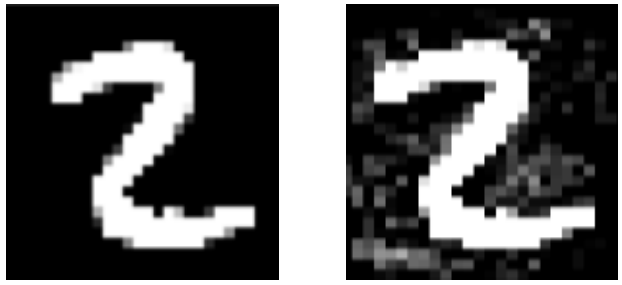
$f(x) = \mathrm{ReP}_{n-1}$



Question:

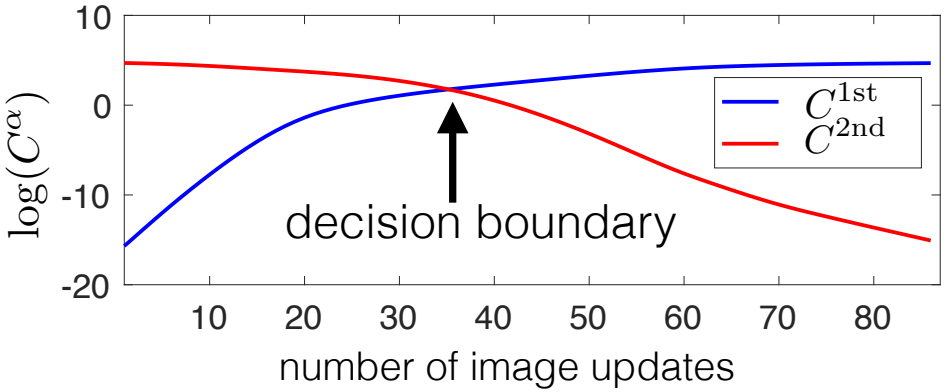
Are there any tasks for which models with higher order interactions perform better than models with quadratic interactions?

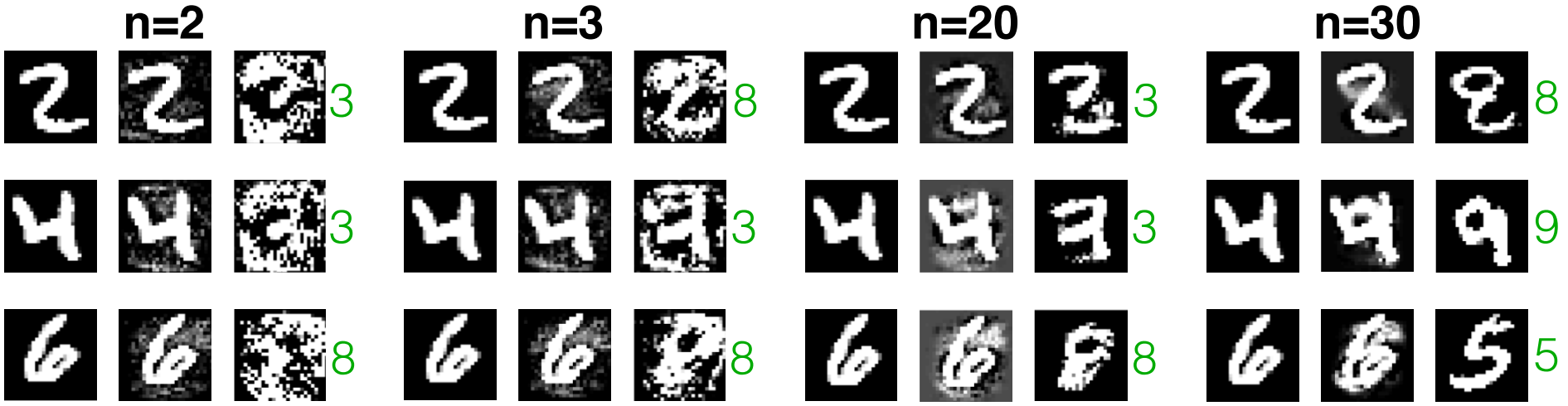
Adversarial Inputs

2  3

$$v_i \rightarrow v_i - \frac{\partial C}{\partial v_i}$$

Adversarial Deformations in DAM

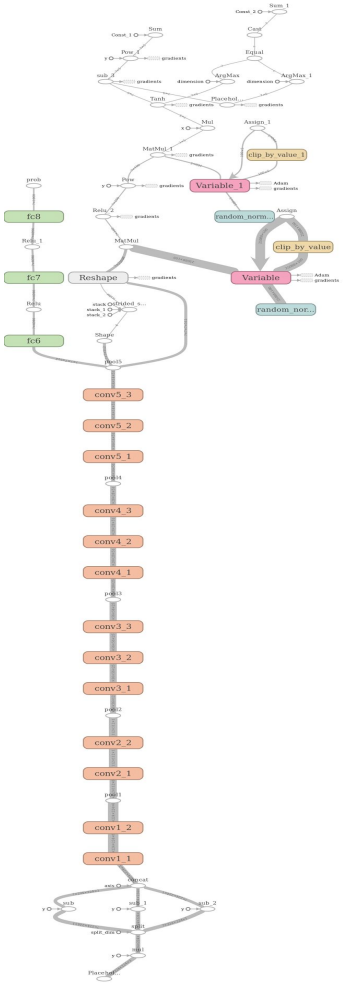




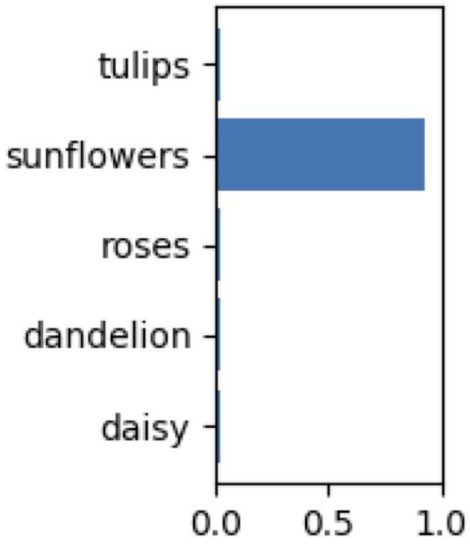
Question:

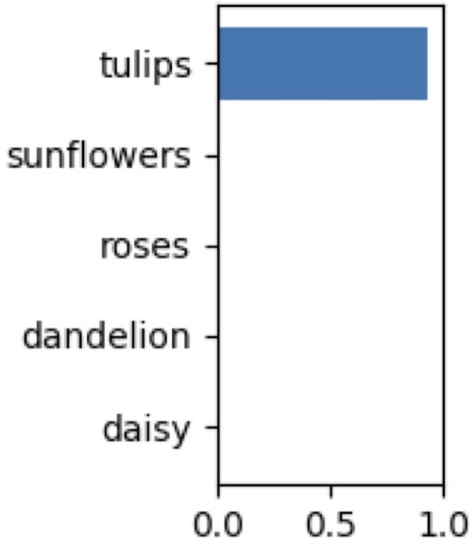
Can we use Dense Associative Memories for classification of high resolution images?

VGG16 coupled to DAM



Adversarial Inputs in the Image Domain





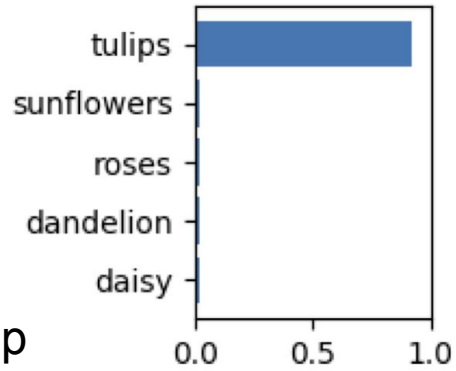
Input transfer

Initial Image

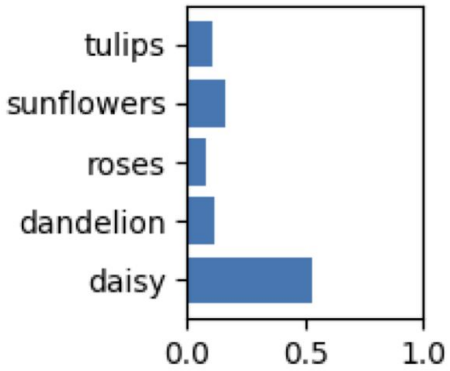
made with n=2

Target class: tulip

classified by n=2



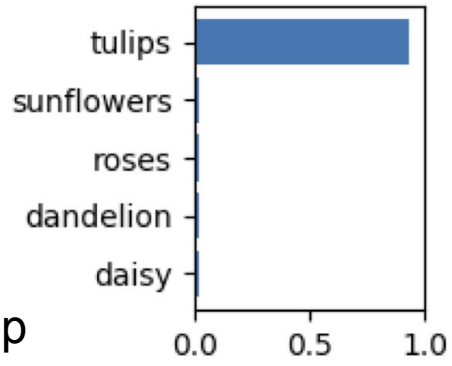
classified by n=8



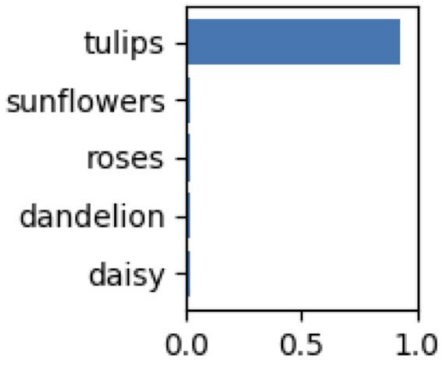
made with n=8

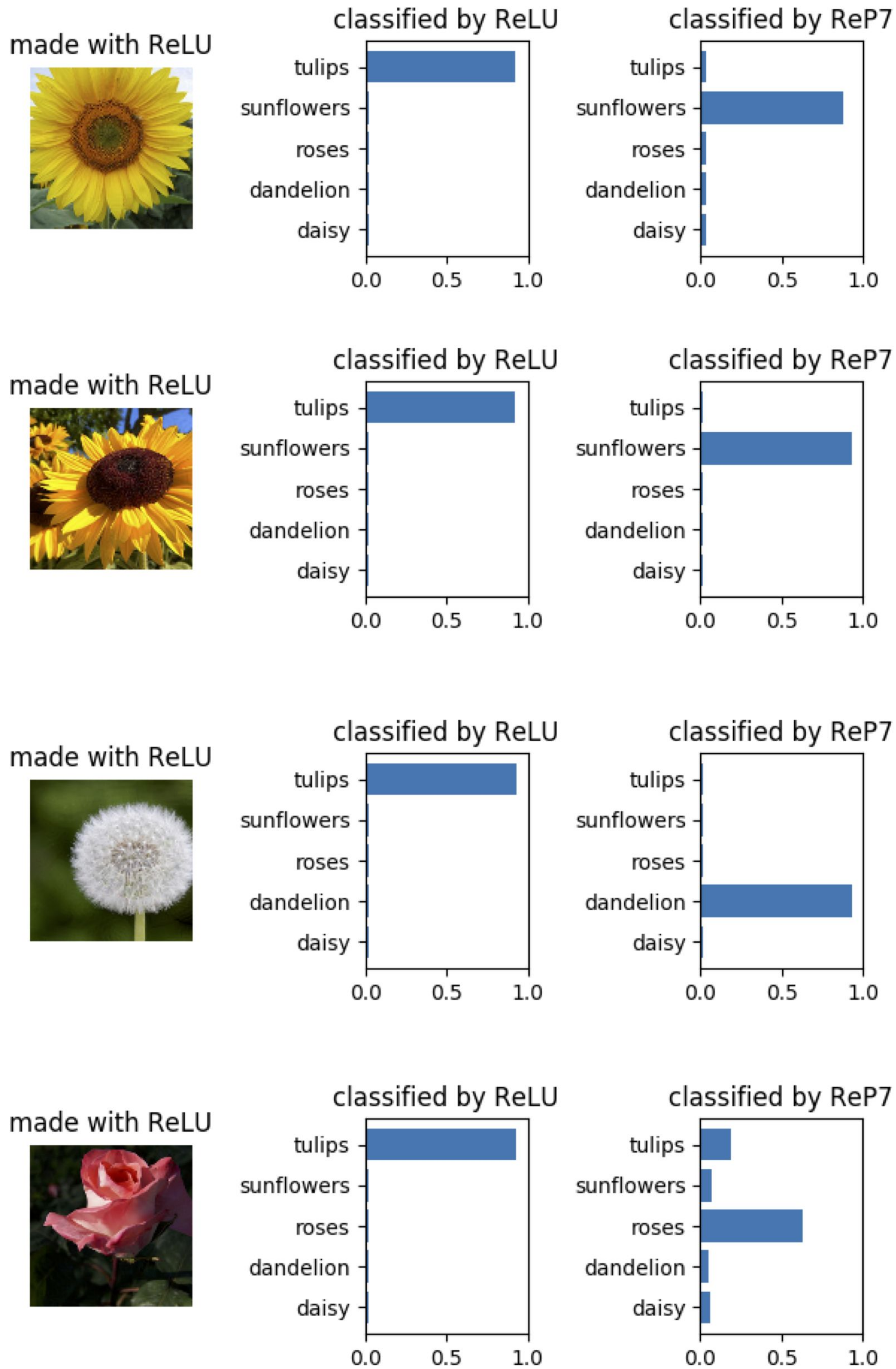
Target class: tulip

classified by n=2



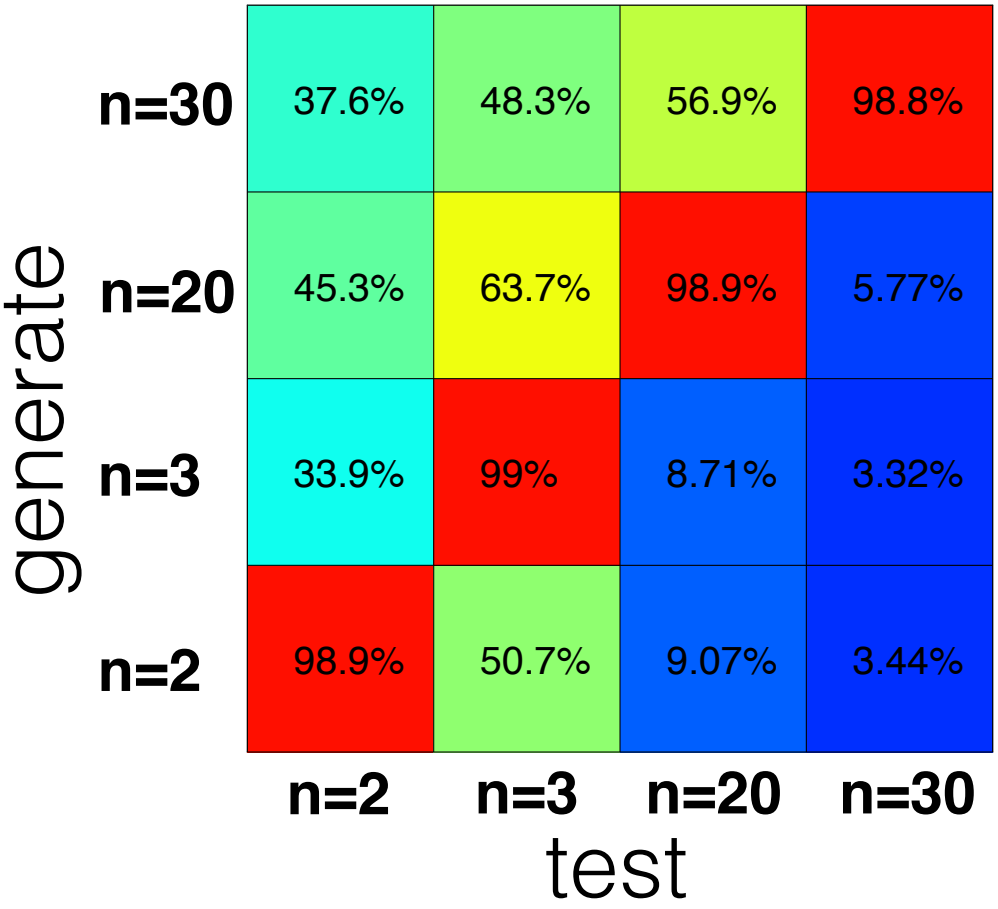
classified by n=8




made with ReLU
classified by ReLU
classified by ReP7
tulips
sunflowers
roses
dandelion
daisy
0.0
0.5
1.0
made with ReLU
classified by ReLU
classified by ReP7
tulips
sunflowers
roses
dandelion
daisy
0.0
0.5
1.0
made with ReLU
classified by ReLU
classified by ReP7
tulips
sunflowers
roses
dandelion
daisy
0.0
0.5
1.0
made with ReLU
classified by ReLU
classified by ReP7
tulips
sunflowers
roses
dandelion
daisy
0.0
0.5
1.0

Error rate of misclassification

Generate \ Classify	n=2	n=8
n=2	100%	32%
n=8	57%	100%


generate
n=30
n=20
n=3
n=2
37.6%
48.3%
56.9%
98.8%
45.3%
63.7%
98.9%
5.77%
33.9%
99%
8.71%
3.32%
98.9%
50.7%
9.07%
3.44%
n=2
n=3
n=20
n=30
test

Results on ImageNet

Accuracy: 69%

lorikeet PredL:91 TrueL:91

Model T PredL:662 TrueL:662

dowitcher PredL:143 TrueL:143

suspension bridge PredL:840 TrueL:840

wing PredL:909 TrueL:909

toyshop PredL:866 TrueL:866

folding chair PredL:560 TrueL:560

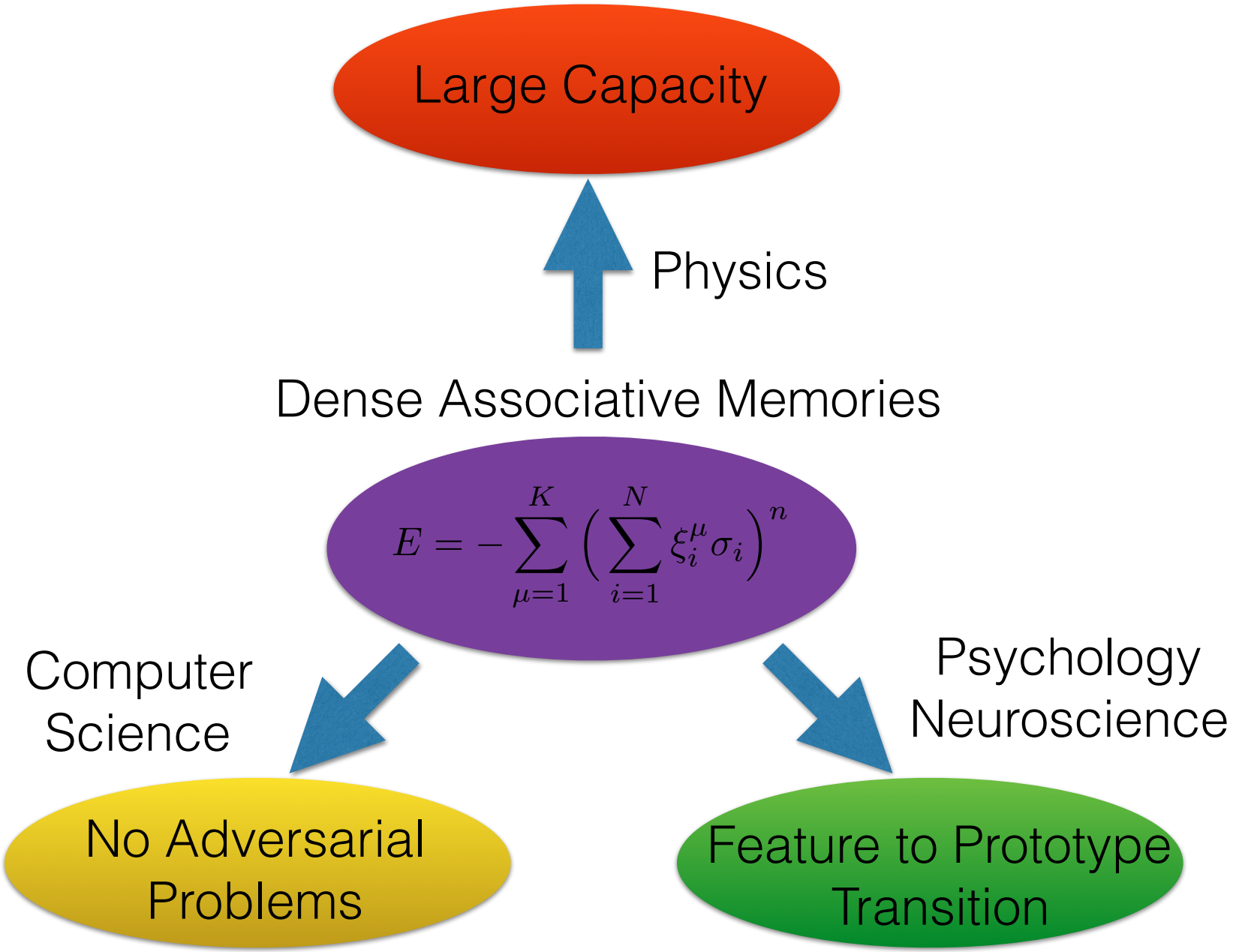
ImageNet errors

moving van PredL:676 TrueL:735

police van, police wagon, paddy wagon, patrol wagon, wagon, black Maria

guillotine PredL:584 TrueL:443

bell cote, bell cot


Large Capacity
Physics
Dense Associative Memories
$E = -\sum_{\mu=1}^{K} \left(\sum_{i=1}^{N} \xi_i^{\mu} \sigma_i \right)^n$
Computer Science
Psychology Neuroscience
No Adversarial Problems
Feature to Prototype Transition