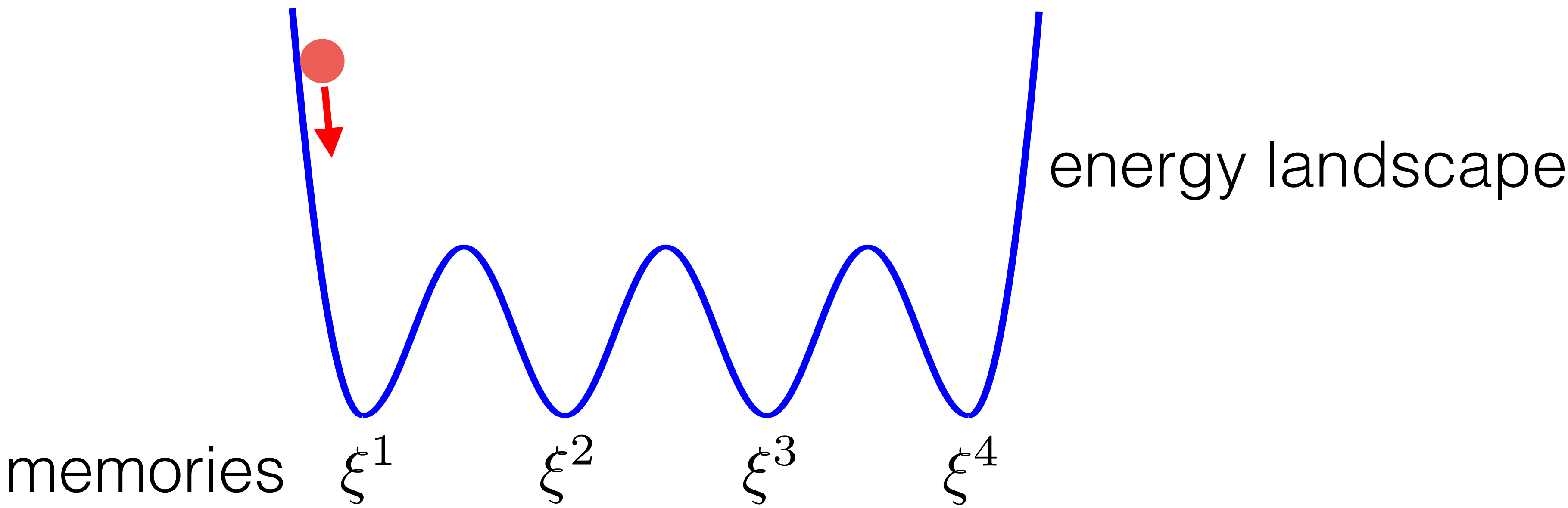
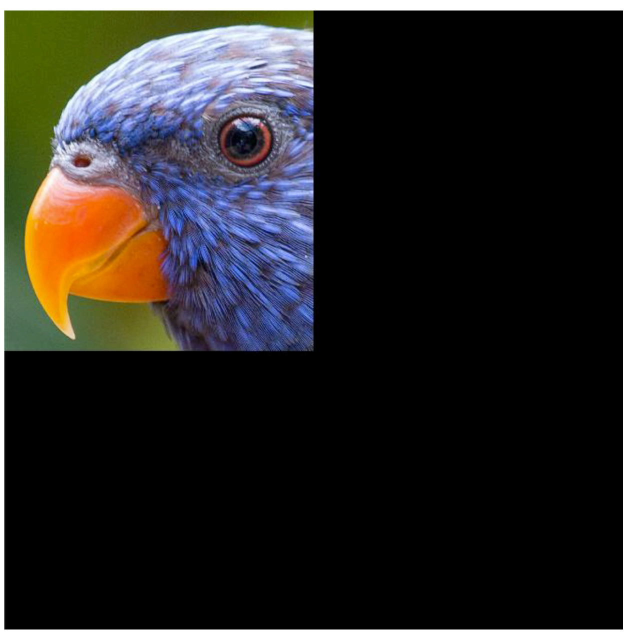
Dense Associative Memories and Deep Learning

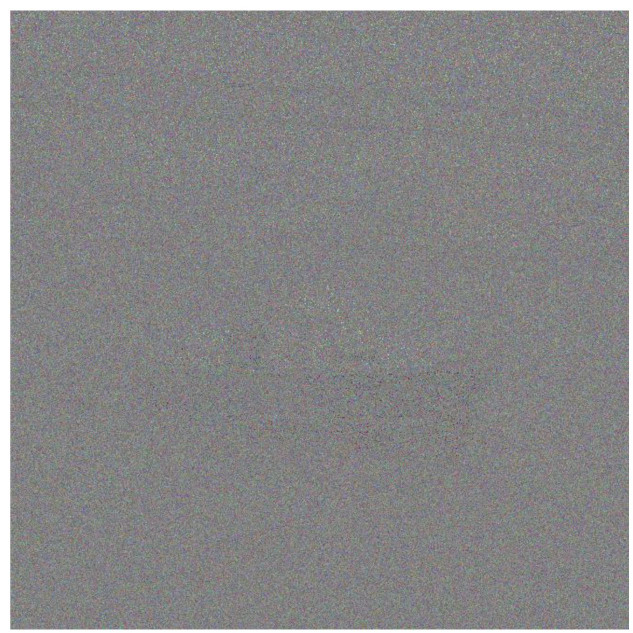
Dmitry Krotov
Institute for Advanced Study

IAS | INSTITUTE FOR ADVANCED STUDY

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

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

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

Dense Associative Memory

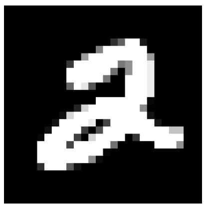
$$E = -\sum_{\mu=1}^{K} \Big(\sum_{i=1}^{N} \xi_i^\mu \sigma_i \Big)^n$$

$$n \geq 2$$

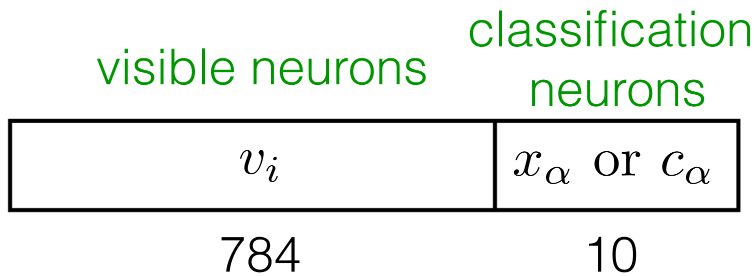
power of the interaction vertex

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

Pattern recognition with DAM

$v_i =$  28

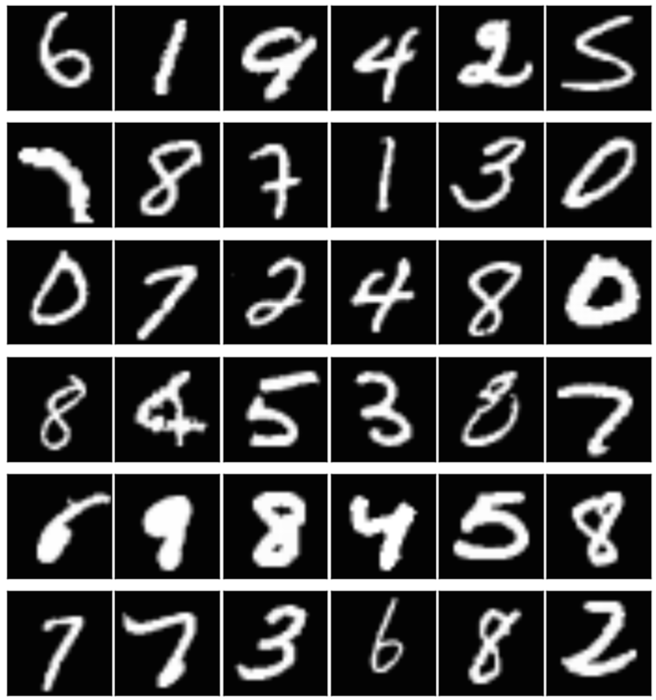
28



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

training

constructed memory vectors



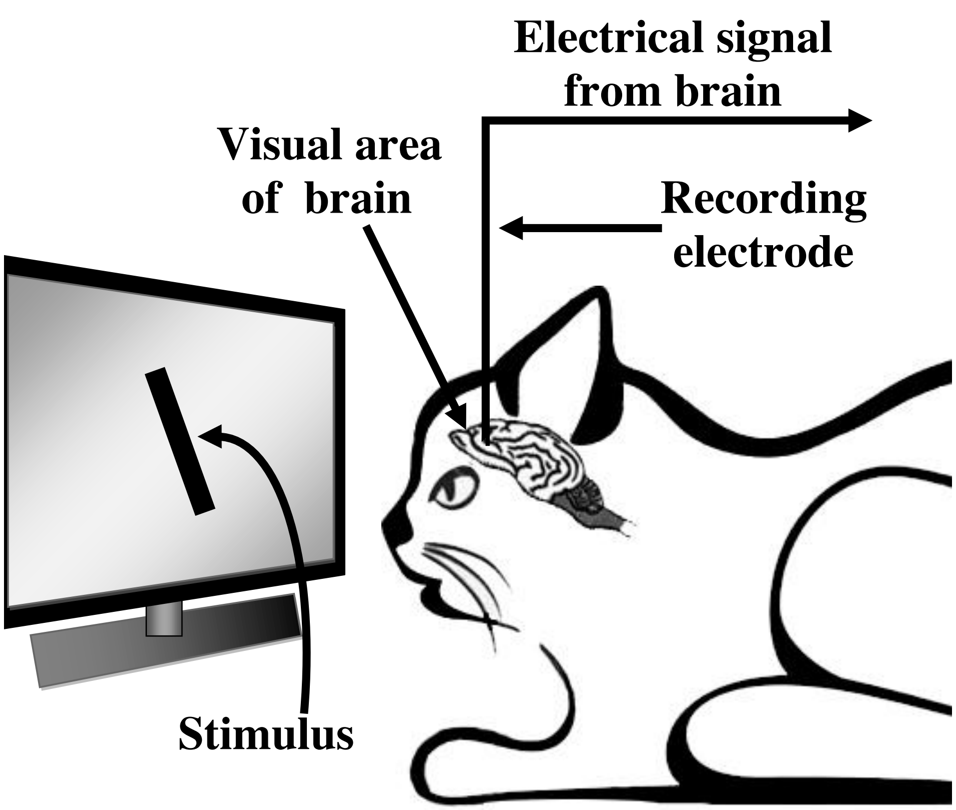
MNIST Dataset

Main question:

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

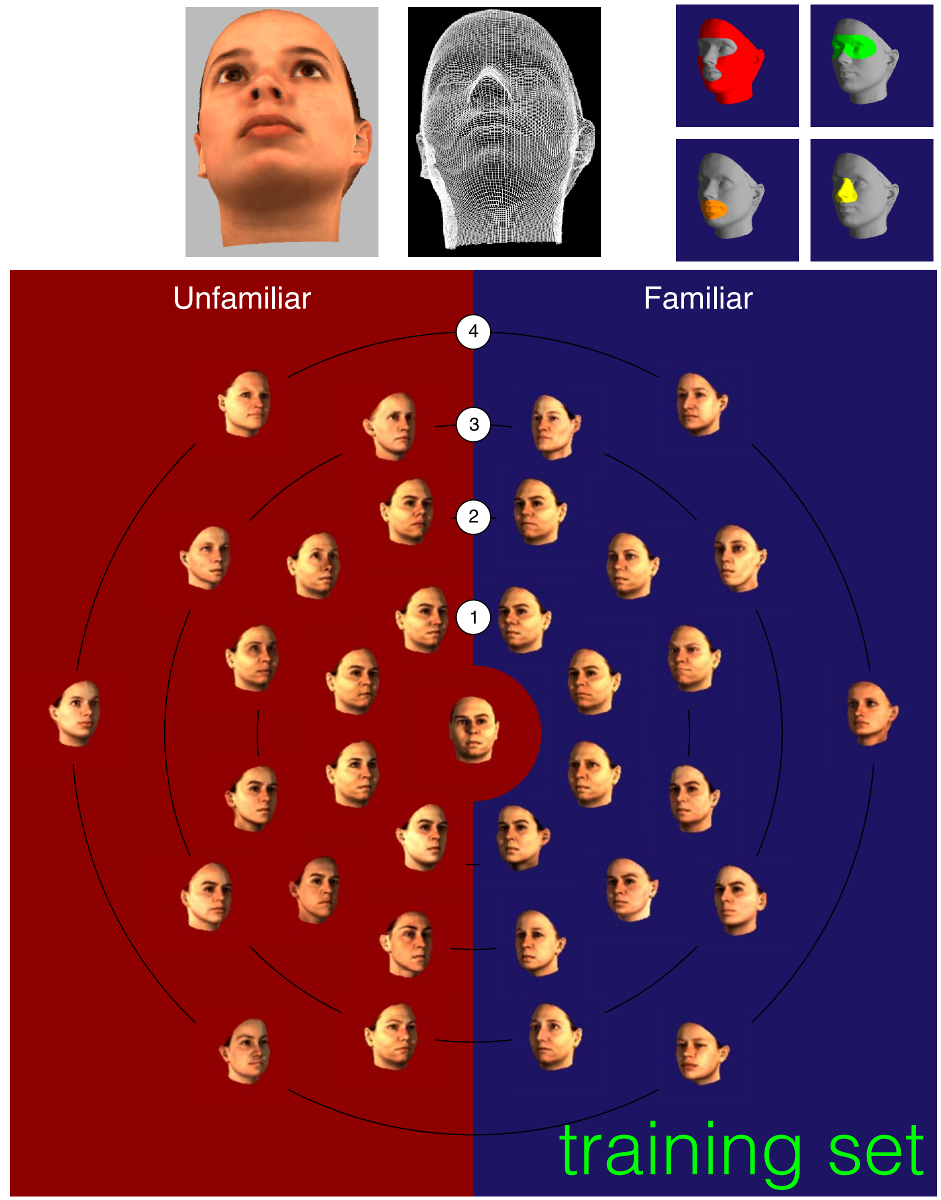
Features vs. prototypes in psychology and neuroscience

Feature-matching theory



Hubel,Wiesel, 1959

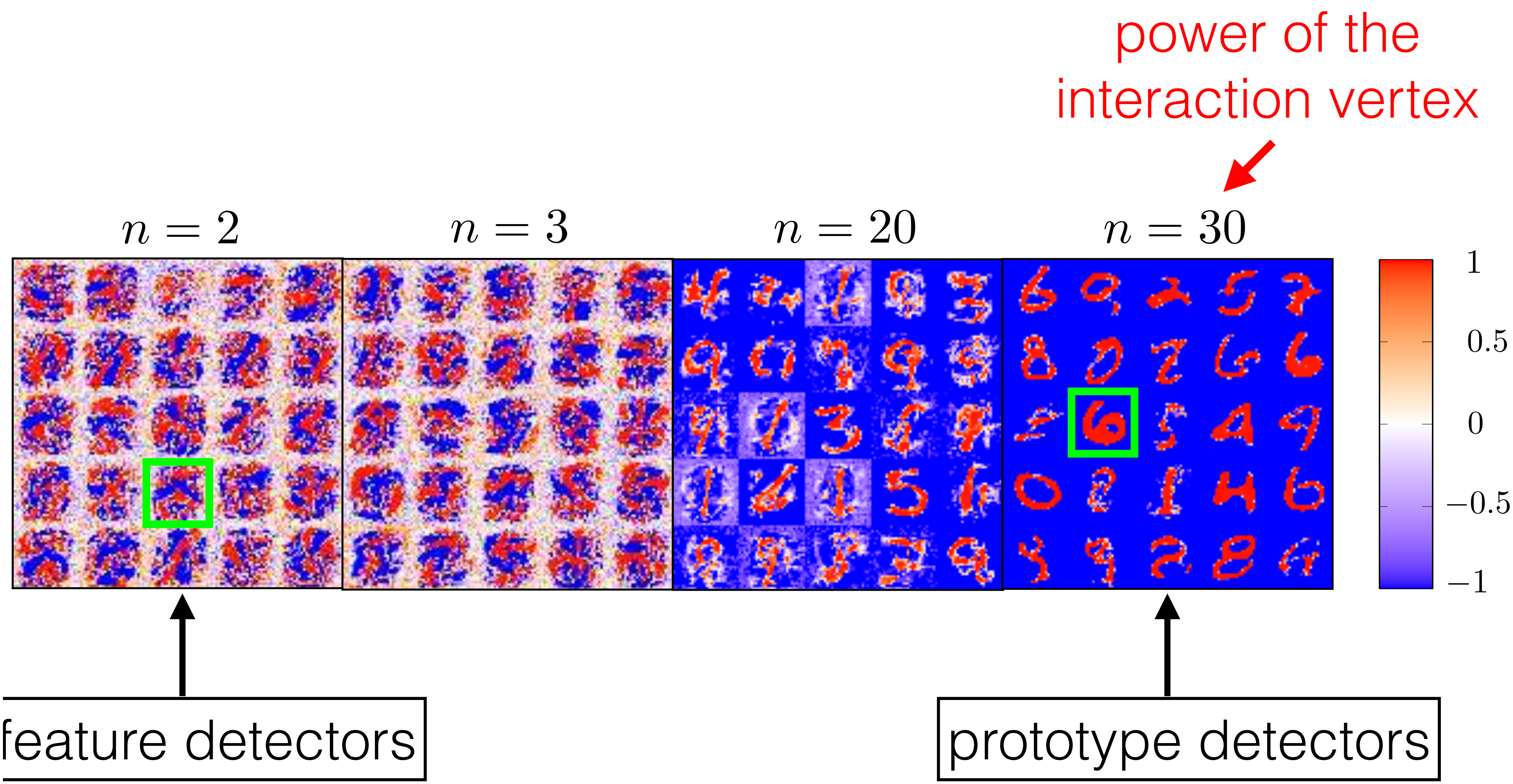
Prototype theory



Solso, McCarthy,1981

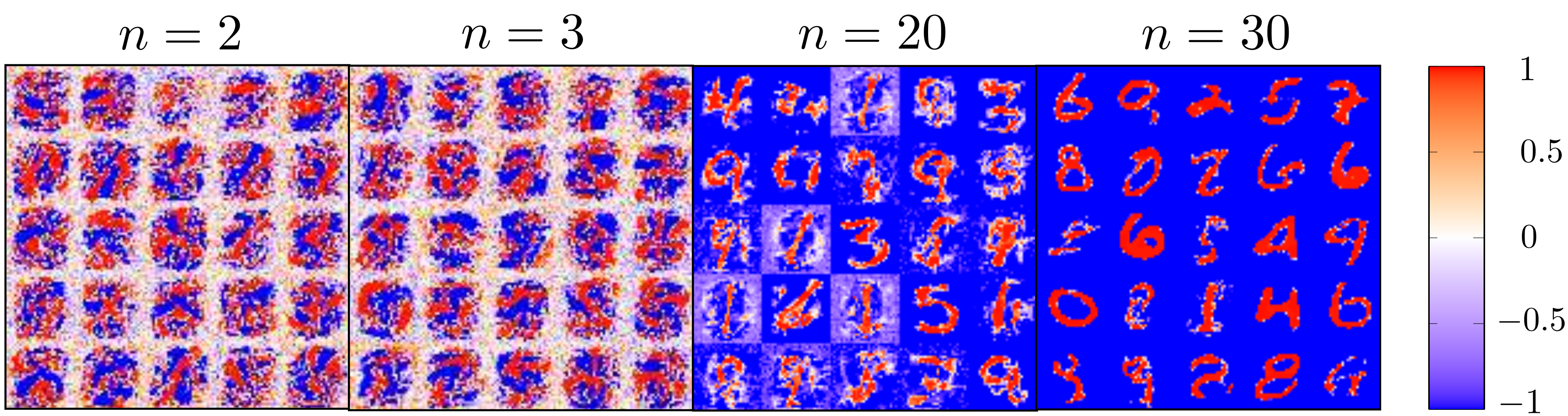
Wallis, et al., Journal of Vision,2008

Feature to prototype transition



Feature to prototype transition

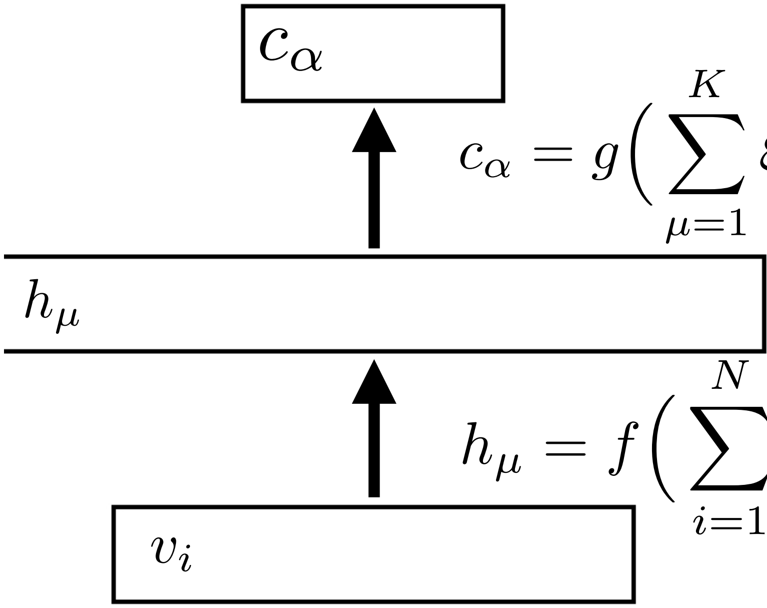
power of the
interaction vertex



1.51% 1.44% 1.61% 1.80%

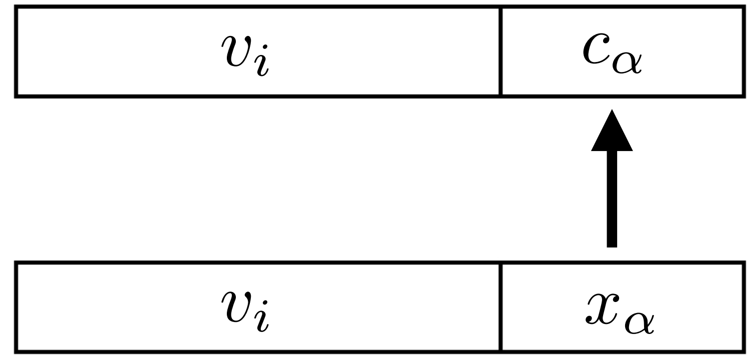
1.6% Simard, Steinkraus, Platt, 2003

Duality with feed-forward nets



$$c_\alpha = g\Big(\sum_{\mu=1}^{K} \xi_\alpha^\mu h_\mu\Big)$$

$$h_\mu = f\Big(\sum_{i=1}^{N} \xi_i^\mu v_i\Big)$$



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

Duality rule:

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

activation function

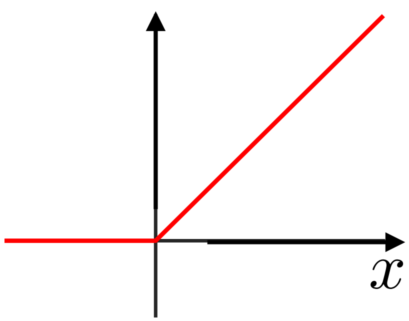
energy function

Commonly used activation functions

$n = 2$

standard
Hopfield net

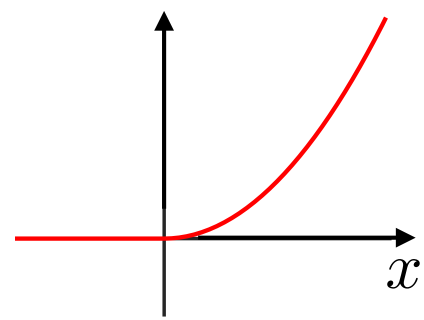
$f(x) = \mathrm{ReLU}$



n

DAM

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



Question:

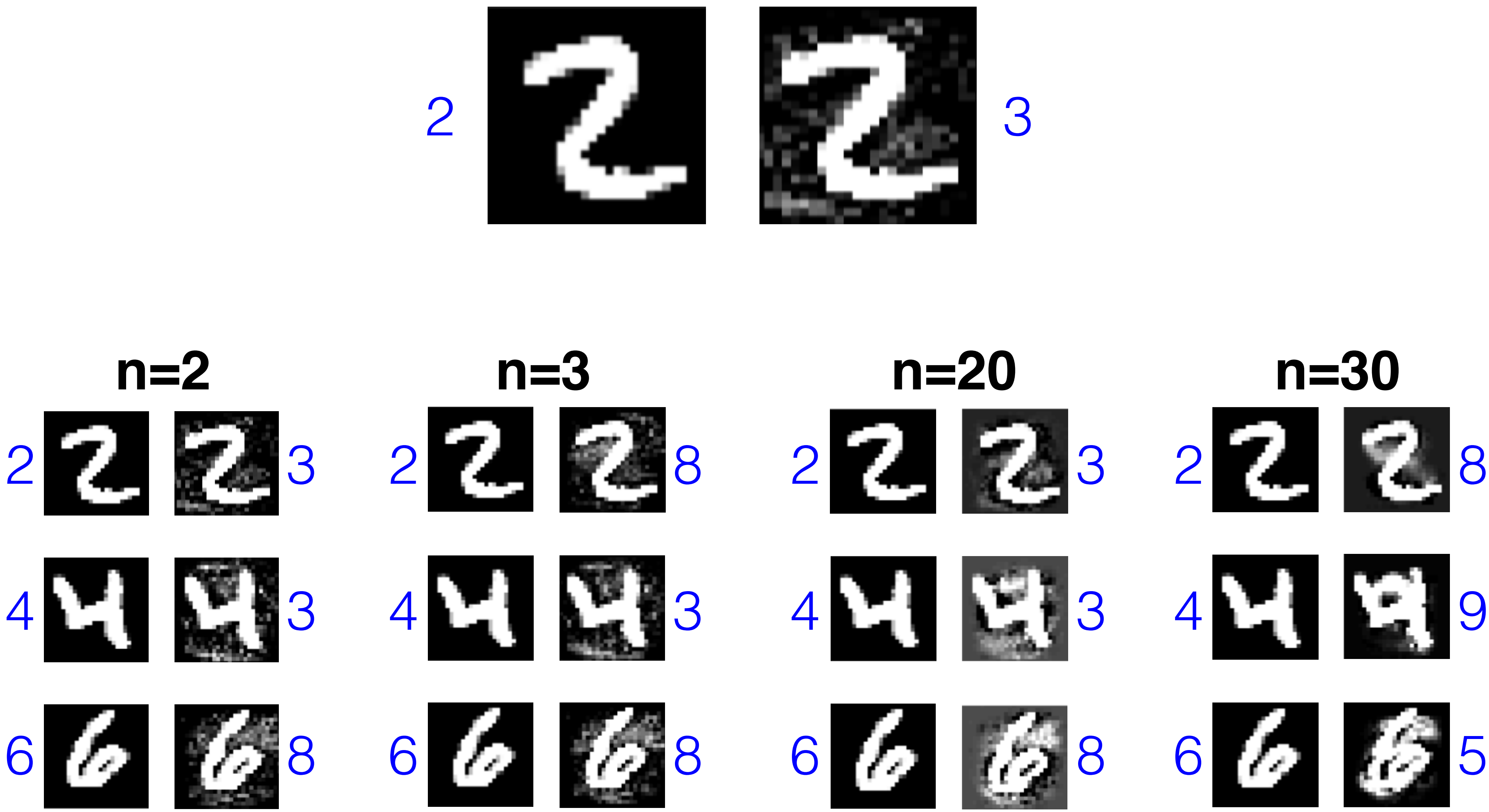
Are there any tasks for which models with higher order interactions perform better than models with quadratic interactions?

Adversarial Inputs

2   3

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

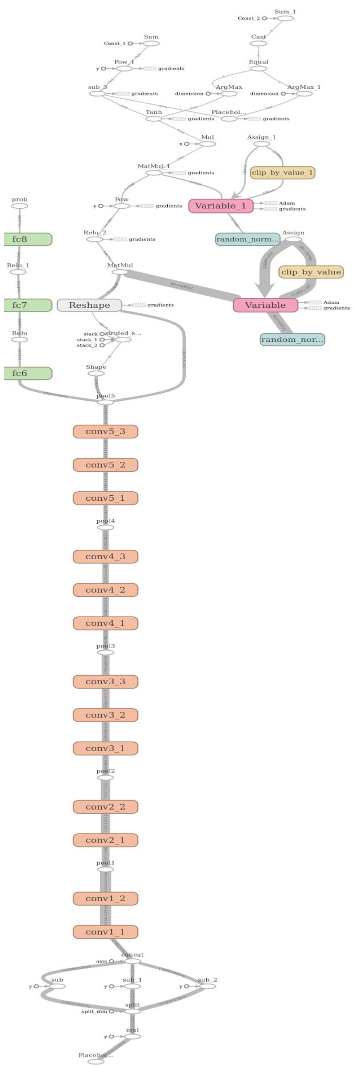
Adversarial Inputs



Question:

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

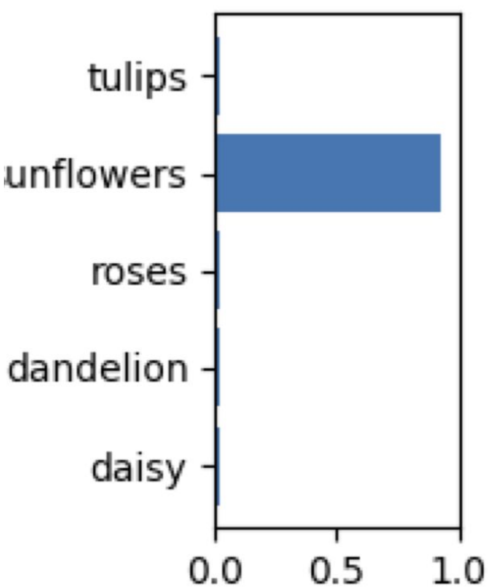
VGG16 coupled to DAM

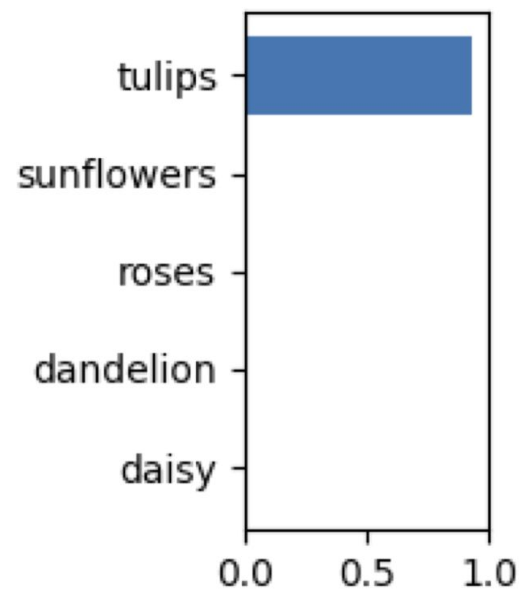


Adversarial Inputs in the Image Domain

+

=





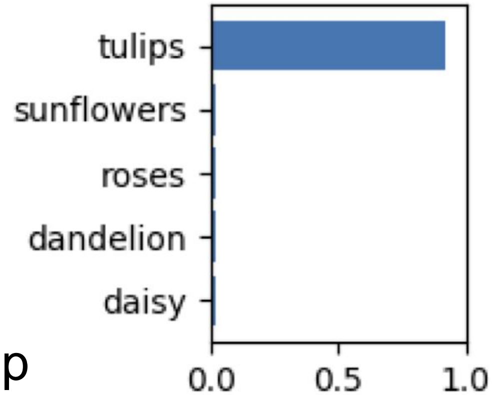
Input transfer

Initial Image

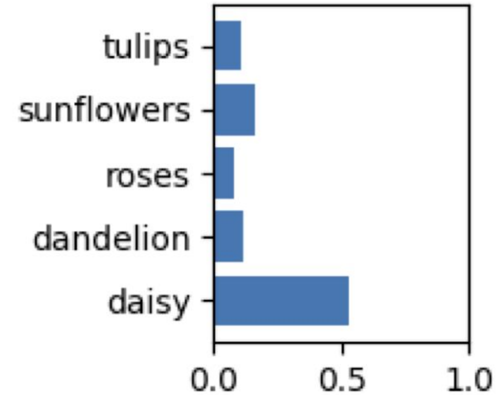
made with n=2

Target class: tulip

classified by n=2



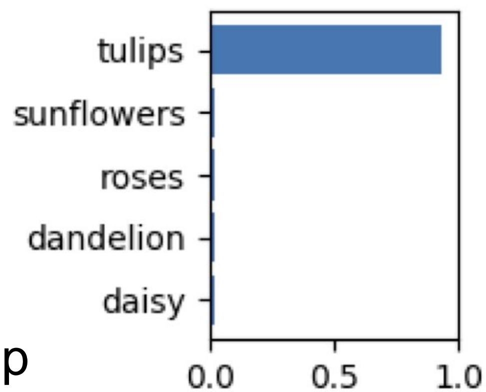
classified by n=8



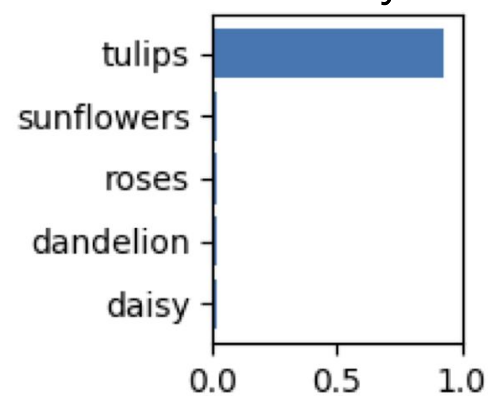
made with n=8

Target class: tulip

classified by n=2



classified by n=8



Error rate of misclassification

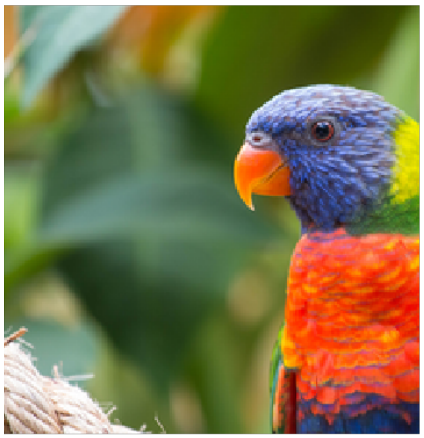
Classify

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

Results on ImageNet

Accuracy: 69%

rikeet PredL:91 TrueL:91



Model T PredL:662 TrueL:662

dowitcher PredL:143 TrueL:143

suspension bridge PredL:840 TrueL:840

ving PredL:909 TrueL:909

toyshop PredL:866 TrueL:866

folding chair PredL:560 TrueL:560

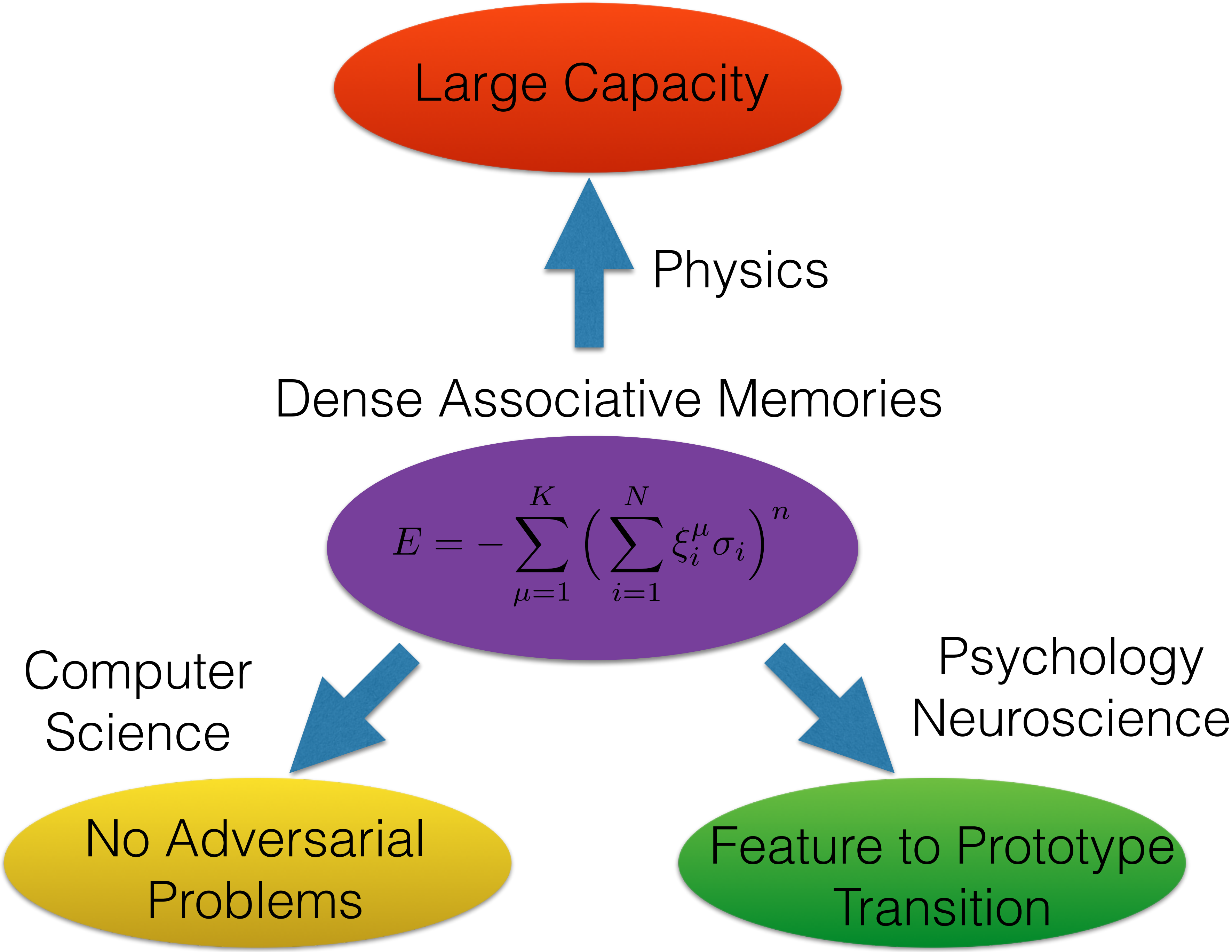
ImageNet errors

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
$E = -\sum_{\mu=1}^{K}\Big(\sum_{i=1}^{N}\xi_i^\mu \sigma_i\Big)^n$
Computer Science
Psychology Neuroscience
No Adversarial Problems
Feature to Prototype Transition

References:

- D.Krotov,J.Hopfield,”Dense Associative Memory for Pattern Recognition”, Advances in Neural Information Processing Systems, 1172-1180, 2016

- D.Krotov, J.Hopfield, “Dense Associative Memory is Robust to Adversarial Inputs”, arXiv:1701.00939