

SPIE Defense and Security

Disruptive Technologies in Information Sciences

Conference 10652-40 Session Chair: Jon R. Williams

April 18, 2018

Orlando, Florida

Title: ***Joint proximity association template for neural networks***

James P. LaRue PhD, Jadco Signals (contact for today's presentation: James@JadcoSignals.com)
Tech Director, ICF Cyber Security Services (contact James.LaRue@ICF.com)

Abstract: A technical solution is described for implementing a computer-executed system of association memory matrices to replace the proximal layers of a convolutional neural network (CNN). An example method includes configuring one Associative Memory Matrix (AMM) for each configured layer in the CNN. This one-to-one conversion method motivates the name to the product: the Joint Proximity Association Template (JPAT) for Neural Networks.

The invention is a numerically stable soft-ware based implementation that (1) reduces the long training times, (2) reduces the execution time, and (3) produces bidirectional intra-layer connections and potentially, inter-layer connections as well. The method further includes, potentially, forming a single AMM, from the multiple AMMs corresponding to the multiple and proximal layers of the CNN, in anticipation of the well-known (Cybenko/Hornik) Universal Approximation Theorem.

Historical Path to a Convolutional, and Bi-directional, NN

1. Neural Network - NN

Alexander Bain (1873) and William James (1890)

Neurons interact.

McCulloch and Pitts (1943)

1st computational model.

Rosenblatt (1958)

Feed-forward perceptron, convergence issue.

Werbos (1975)

Fixes the Rosenblatt problem, goes unrecognized.

Rumelhart, Hinton, Williams, McClelland (1984)

Recognize Werbos work.

2. Convolutional Neural Network - CNN

Fukushima (1980)

Neocognitron – hidden layer visual pattern recognition.

LeCun and Bengio (1995)

Convolutional ‘Neocognitron’ (CNN), long training.

3. Bi-directional Associative Matrix - BAM

Kosko (with Kohonen and Grossberg) (1988)

Bi-directional I/O matrix, no hidden layers, stability issue.

4. Scale-Invariant Feature Transform - SIFT

David Lowe (1999-2012)

SIFT, fast alternative to CNN, got the job done fast, Patented.

5. CNN + GPU

Krizhevsky, Sutskever, Hinton (2012)

Deep Learning with GPUs (ImageNet, TensorFlow).

Couple of Comments

Minsky and Papert (1969)

Need at least one hidden layer between I/O to be meaningful.

Cybenko/Hornik (1989)

Universal Approximation Theorem (UAT), one single hidden layer.

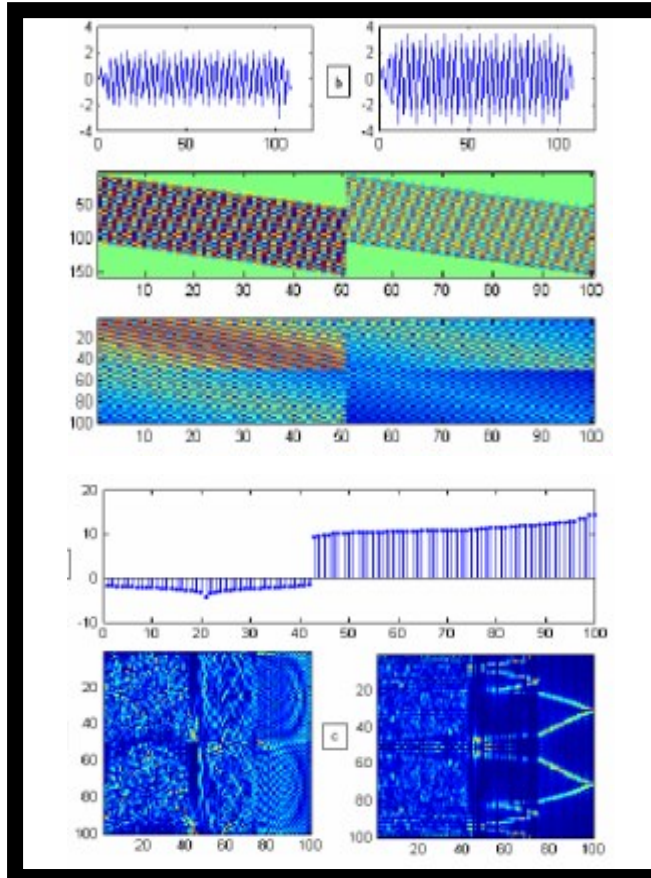
My background

AAS, BA, BEd, MS, PhD, and a Post-Doc at AFRL under Dr. Andy Noga
(throw in Cleveland Wrecking, Avondale Shipyards as RADAR Technician, and ECG Tech Certificate)

Hands on Technician/Signal Processor
Here, working on Wireless Heart Monitor
(interest in tying together ECG, EEG, DNA)



Group theoretic vector associations
among structures of SVD and Nullspace
from convolutional multipath



Listened to advice on the subtlety
of whitening Co-channel components
after Singular Value Decomposition

◆ AMUSE and SOBI Algorithm Using SOS:

Step 1: Prewhitening by Eigenvalue Decomposition (EVD)

$$\mathbf{R}_x = E\{\mathbf{x}[n]\mathbf{x}^H[n]\} \quad (P \times P \text{ matrix})$$

$$\text{EVD} \Rightarrow \begin{cases} \beta_1, \beta_2, \dots, \beta_K : \text{largest } K \text{ eigenvalues of } \mathbf{R}_x \\ \mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_K : \text{associated } K \text{ eigenvectors of } \mathbf{R}_x \end{cases}$$

$$\hat{\sigma}_w^2 : \text{average of the other (smallest) } P-K \text{ eigenvalues of } \mathbf{R}_x \\ (\text{assuming that } \mathbf{R}_w = E\{\mathbf{w}[n]\mathbf{w}^H[n]\} = \sigma_w^2 \mathbf{I})$$

$$\hat{\mathbf{D}} = \left[\frac{\mathbf{f}_1}{\sqrt{\beta_1 - \hat{\sigma}_w^2}}, \dots, \frac{\mathbf{f}_K}{\sqrt{\beta_K - \hat{\sigma}_w^2}} \right] \quad (\text{whitening matrix})$$

$$\mathbf{z}[n] = \hat{\mathbf{D}}\mathbf{x}[n] = \mathbf{U}\mathbf{s}[n] + \hat{\mathbf{D}}\mathbf{w}[n] \quad (\text{dimension-reduced whitening spatial processing})$$

$$\mathbf{U} : K \times K \text{ unitary matrix}$$

6. Bidirectional Neural Network – JPAT with UAT and Cross-Domain capabilities

LaRue/Luzanov 2011 - AFRL /AFOSR

Initiated BAM model of CNN.

LaRue/Copeland 2012 – DARPA VMR Innovation House

Finished BAM model, came across UAT.

LaRue/LaRue 2013 - JADCO Signals

Patent app. 2014, Recorded 2016, Allowed 2018.

James P. LaRue (Jadco Signals) and Yuriy Luzanov (AFRL-RIGC), Poster 8745-76: “*Stabilizing bidirectional associative memory with principles in independent component analysis and null space*”, SPIE Defense, Security, and Sensing 29 April - 3 May 2013, Baltimore, Maryland

IEEE Applied Imagery Pattern Recognition Workshop

Session II, Deep Learning, Session Chair: Moody

October 10, 2017

The Cosmos Club, Washington, DC

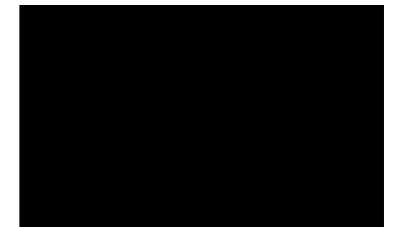
TITLE: *The Anatomy of a Neural Network*

with

Richard L. Tutwiler, PhD, ARL- Penn State University(Professor Emeritus)/CEO LiveMotion3D LLC

Dennison J. LaRue, Biomedical Engineering, Clemson (2017), Freshman, University of South Carolina Law School

Looking inside
the

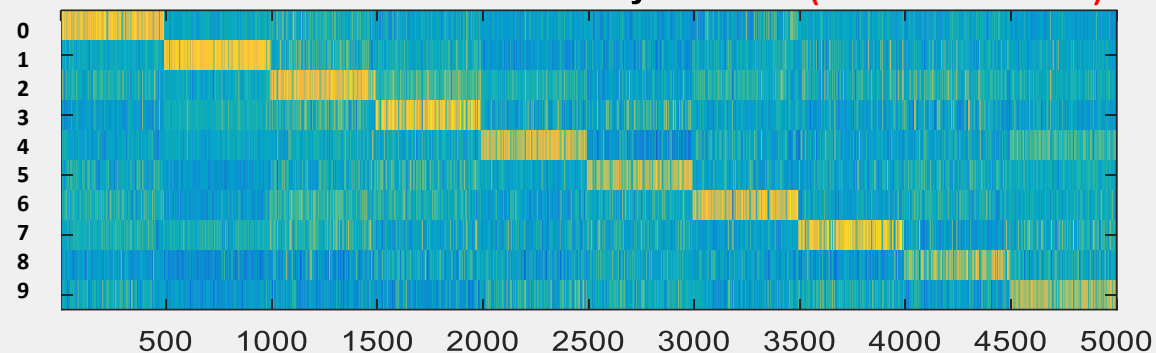


A result to think about using CNN and MNIST data

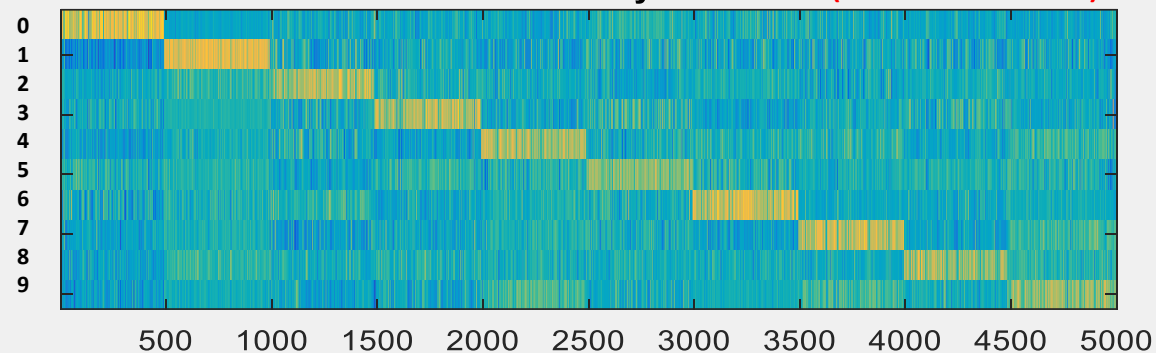
5,000 entries of LeCun's
MNIST Numerals 0 - 9



CNN Accuracy 0.8382 (in 60 minutes)

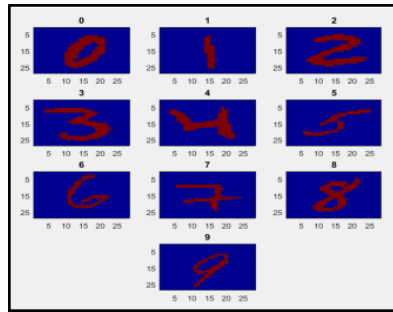


AMM Accuracy 0.8026 (in 2 minutes)



Note: $8026/8382 = .9575$

Visualizing a sequence of layers: L0, L1, L2, L3, L4, L5



LeCun's MNIST
0,1,2,3,4,5,6,7,8,9

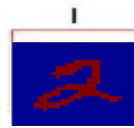
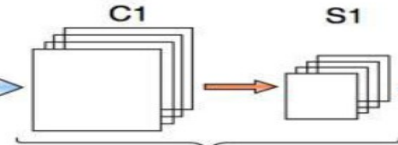
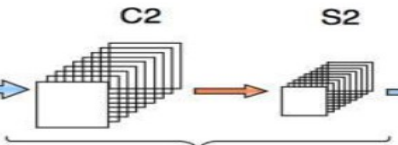


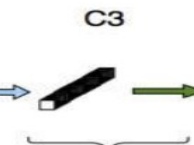
Image
Layer 0



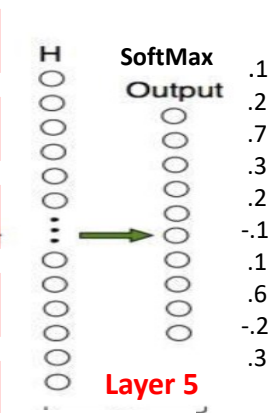
Layer 1



Layer 2

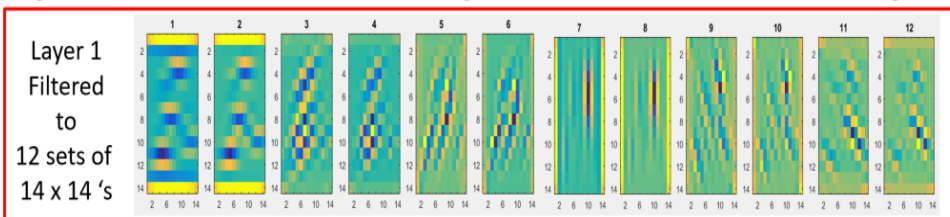
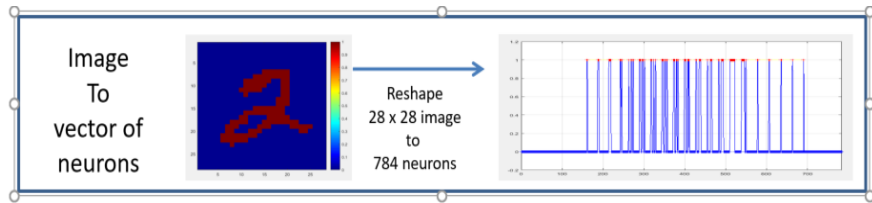


Layer 3

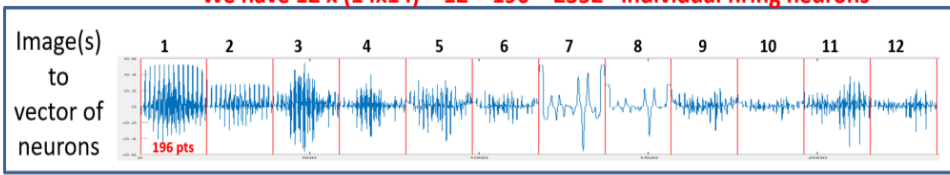


Layer 4

Two's place



We have $12 \times (14 \times 14) = 12 \times 196 = 2352$ 'individual firing neurons'



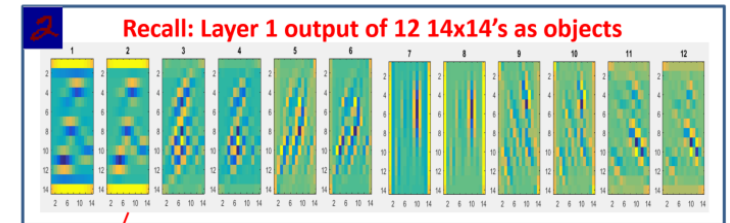
1 $\longrightarrow$ $14 \times 14 \times 12 = 196 \times 12 = 2352$ pts. $\longrightarrow$ 2352

Layer 1
OUT

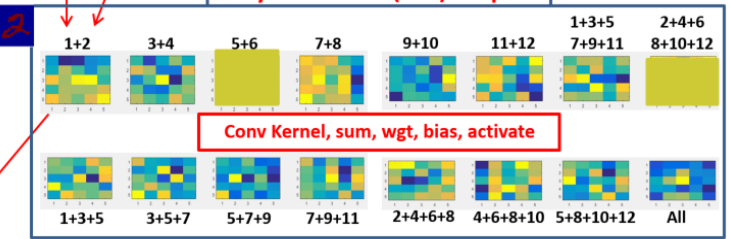
Layer 2
OUT

Detail Processes

```
for i=1:nOMsV2 % 16 neurons here
    for j=1:connlistV2
        a5=OMV1(i).list(j)) % from Layer 1
        a6=KV2(connlistV2(i).list(j))
        convResults=temp + conv2(a5,a6);
    end
    % Apply the activation function
    temp = tanh(convResults);
    %subsample #1
    a7=temp(1:2:end,:);
    a8=temp(2:2:end,:);
    temp2 = a7+a8; % sum
    %subsample #2
    x=temp2(1:1:2:end);
    y=temp2(1:2:2:end);
    temp3=(x+y)/2
    % bring in wgt, bias, activate
    temp4 = temp3*wgt2(i) + bias2(2,i)
    Layer2Out = tanh(temp4);
end
```

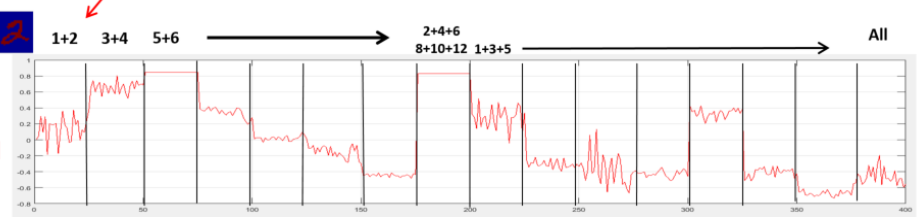


Layer 2 has 16 (5x5) outputs

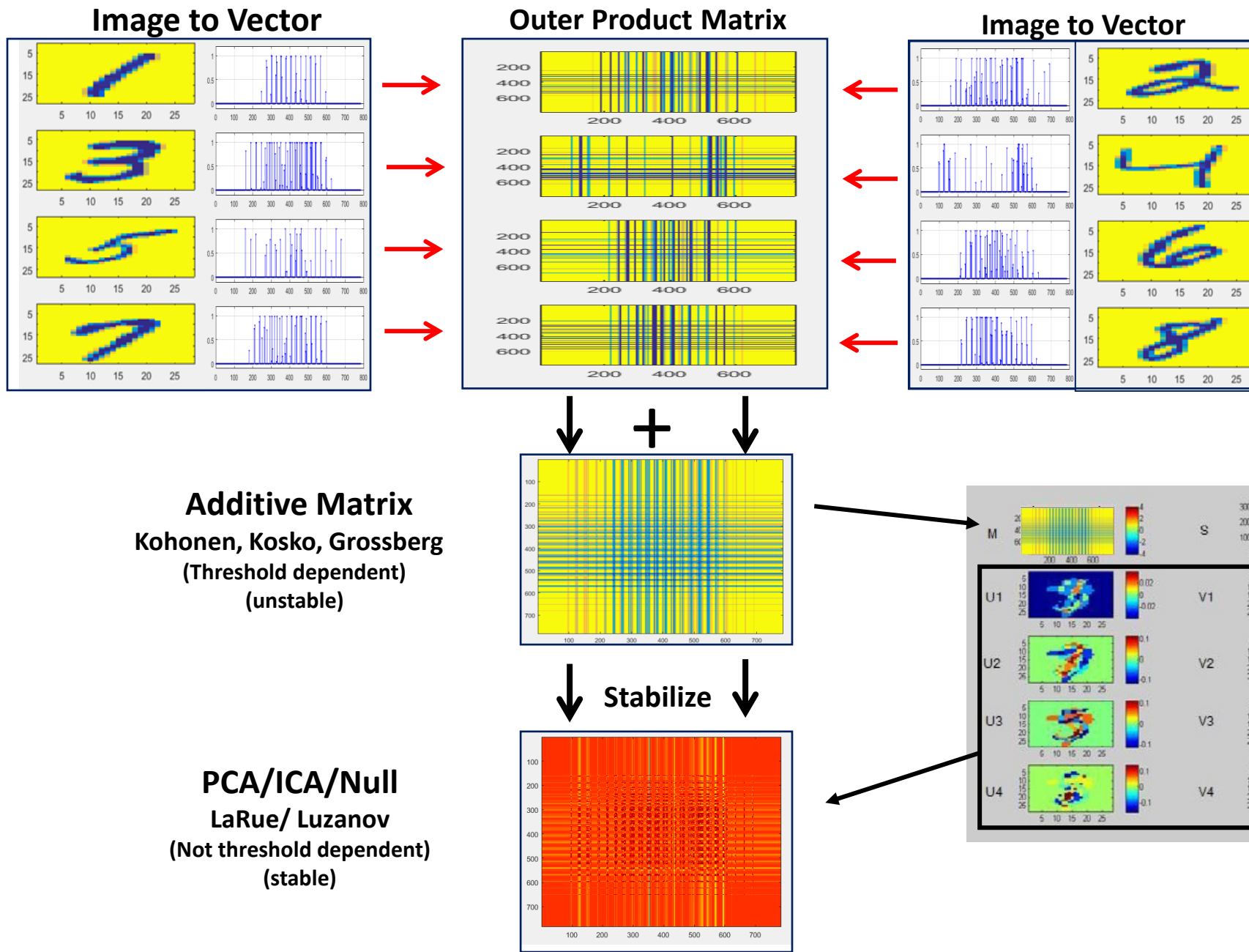


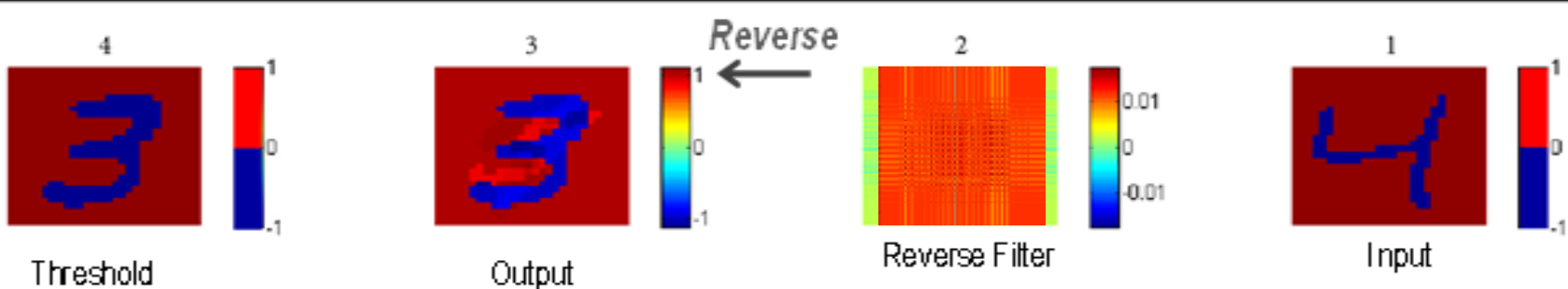
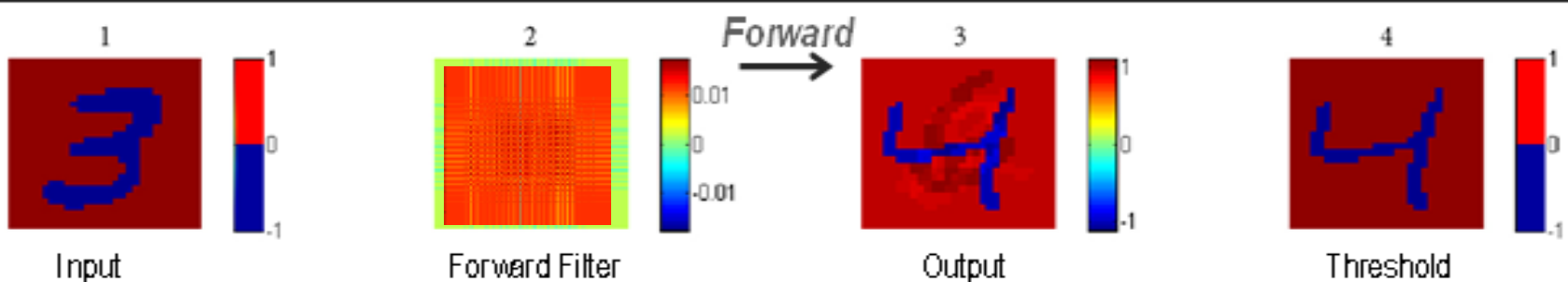
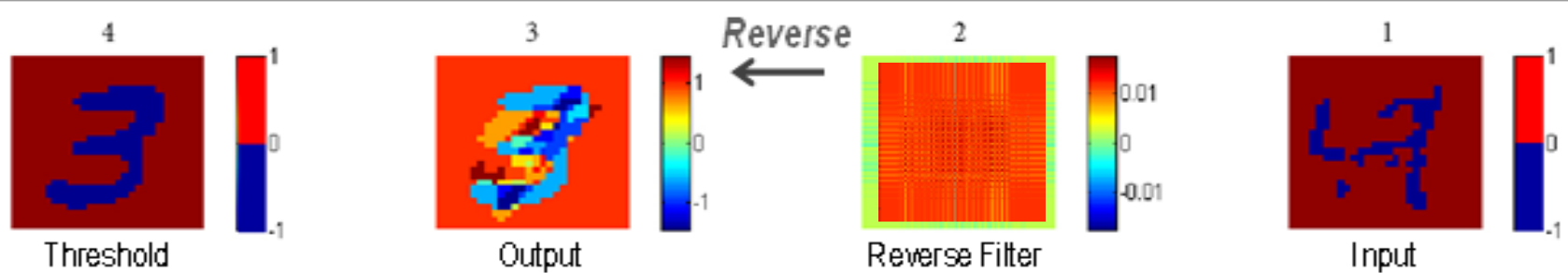
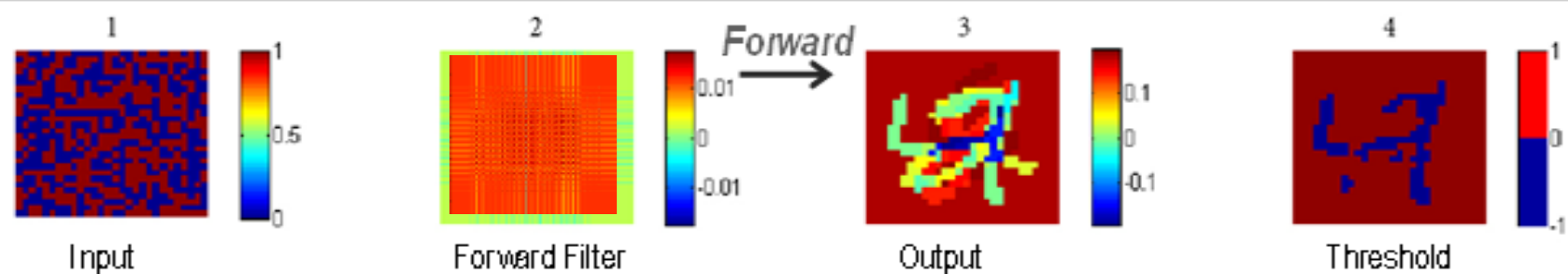
Conv Kernel, sum, wgt, bias, activate

16
5x5's
equals
400
individual
neurons
firing

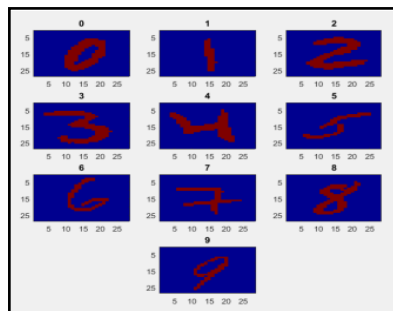


Part 4: Recipe for “Associating” and then Resonating





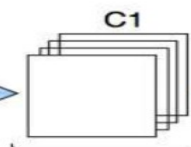
Visualizing a sequence of layers: L0, L1, L2, L3, L4, L5



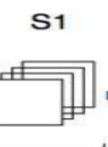
LeCun's MNIST
0,1,2,3,4,5,6,7,8,9



Image
Layer 0



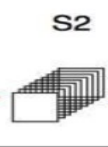
Layer 1



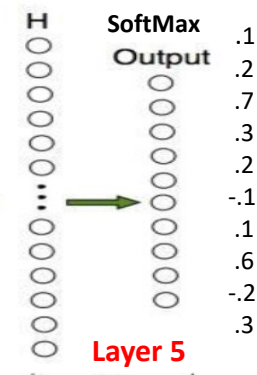
Layer 2



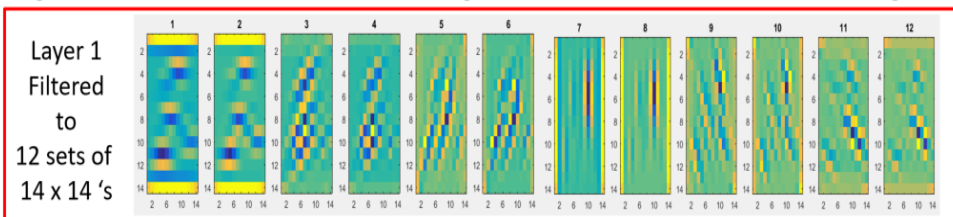
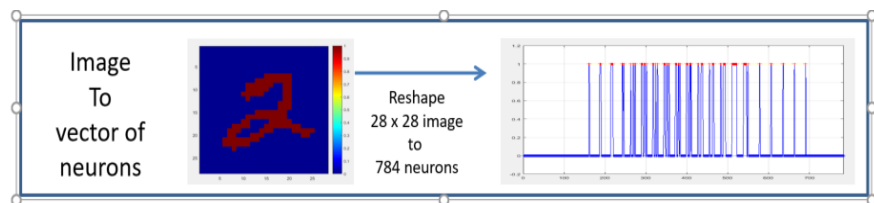
Layer 3



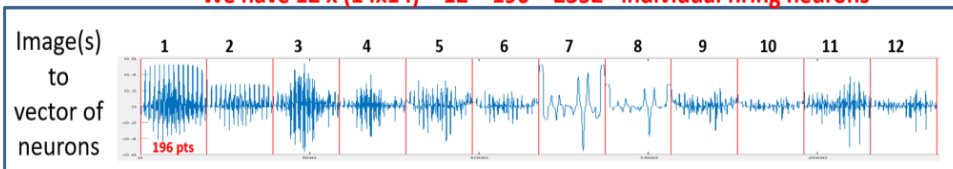
Layer 4



Two's place



We have $12 \times (14 \times 14) = 12 \times 196 = 2352$ 'individual firing neurons'



1 $\longrightarrow$ $14 \times 14 \times 12 = 196 \times 12 = 2352$ pts. $\longrightarrow$ 2352

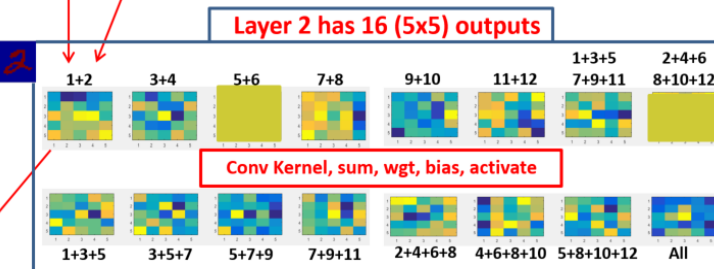
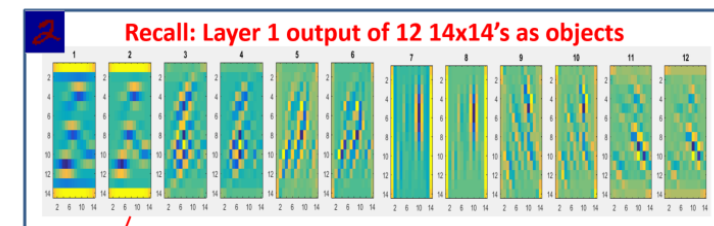
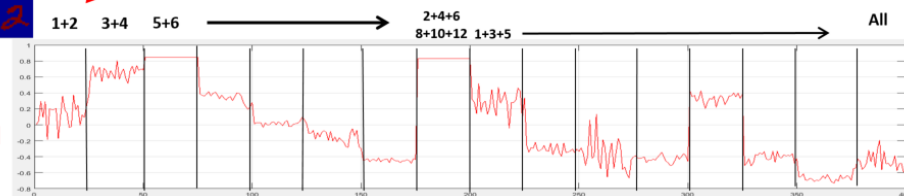
Layer 1
OUT

Layer 2
OUT

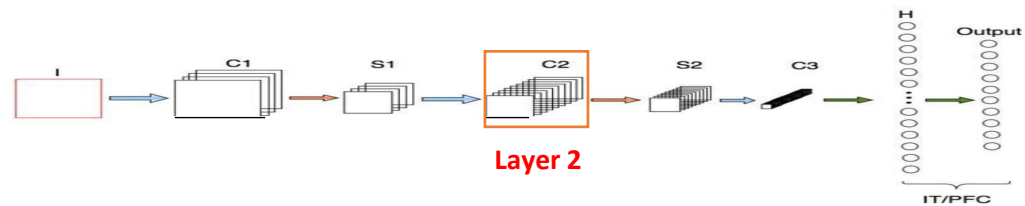
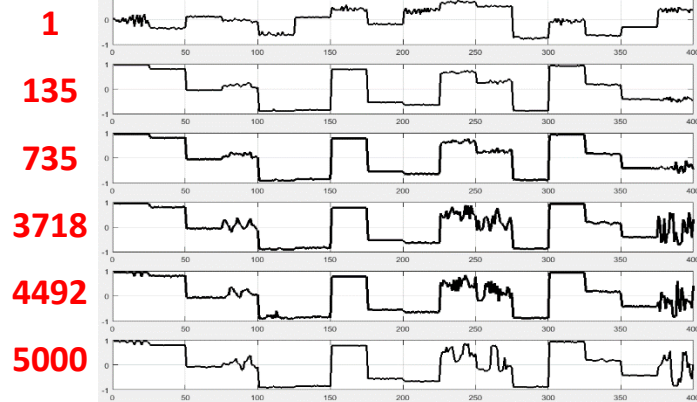
Detail Processes

```
for i=1:nOMsV2 % 16 neurons here
    for j=1:connlistV2
        a5=OMV1(i).list(j) % from Layer 1
        a6=KV2(connlistV2(i).list(j))
        convResults=temp + conv2(a5,a6);
    end
    % Apply the activation function
    temp = tanh(convResults);
    %subsample #1
    a7=temp(1:2:end,:);
    a8=temp(2:2:end,:);
    temp2 = a7+a8; % sum
    %subsample #2
    x=temp2(:,1:2:end);
    y=temp2(:,2:2:end);
    temp3=(x+y)/2
    % bring in wgt, bias activate
    temp4 = temp3*wgt2(i) + bias2(2,i)
    Layer2Out = tanh(temp4);
end
```

16
5x5's
equals
400
individual
neurons
firing



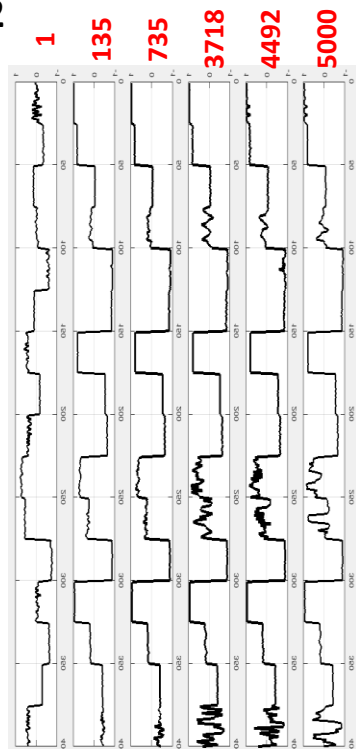
Step 1



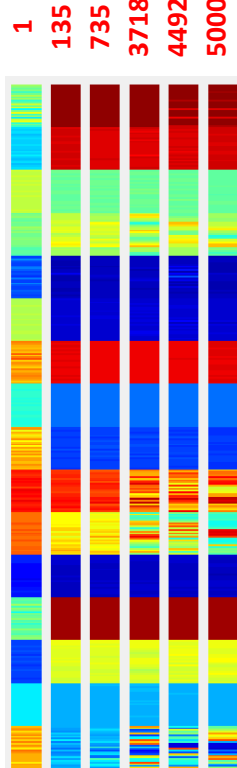
How to see **Layer 2** convergence over 5,000 iterations

1. Plots of iterations: **1, 135, 735, 3718, 4492, 5000**
2. Turn on side (rotation, reflection, (dihedral group))
3. Rotate out as matrix columns, color scale normalized to [- 1, 1]
4. Put all 5,000 together, and ***observe convergence left to right***

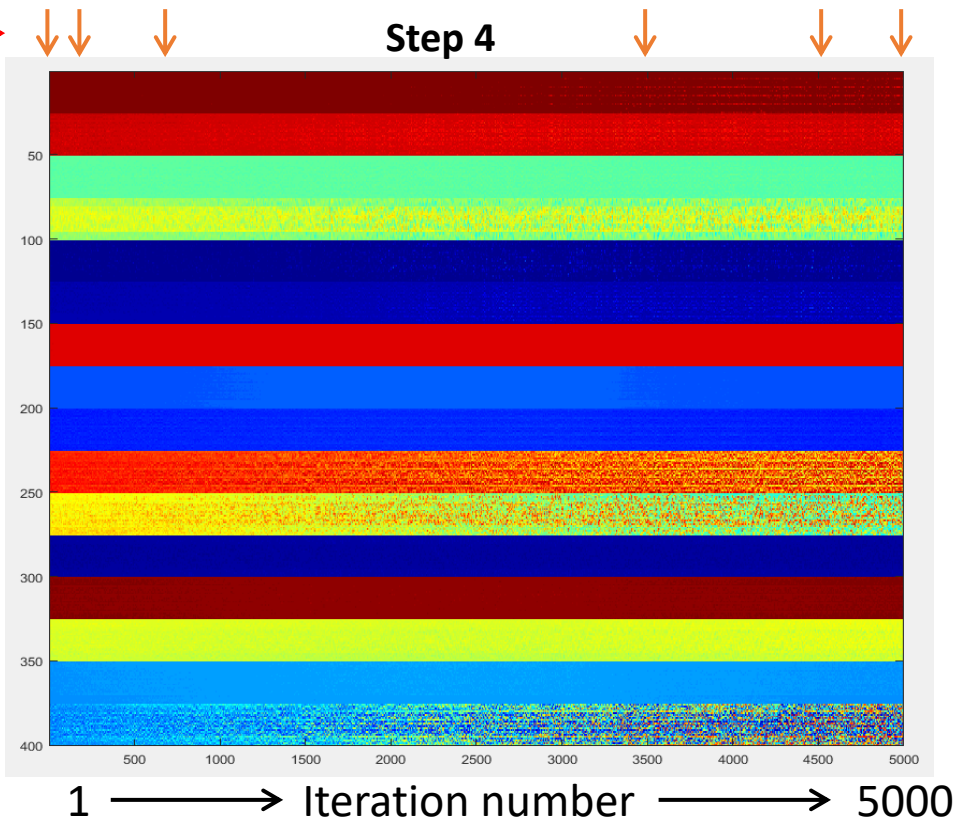
Step 2



Step 3



Step 4



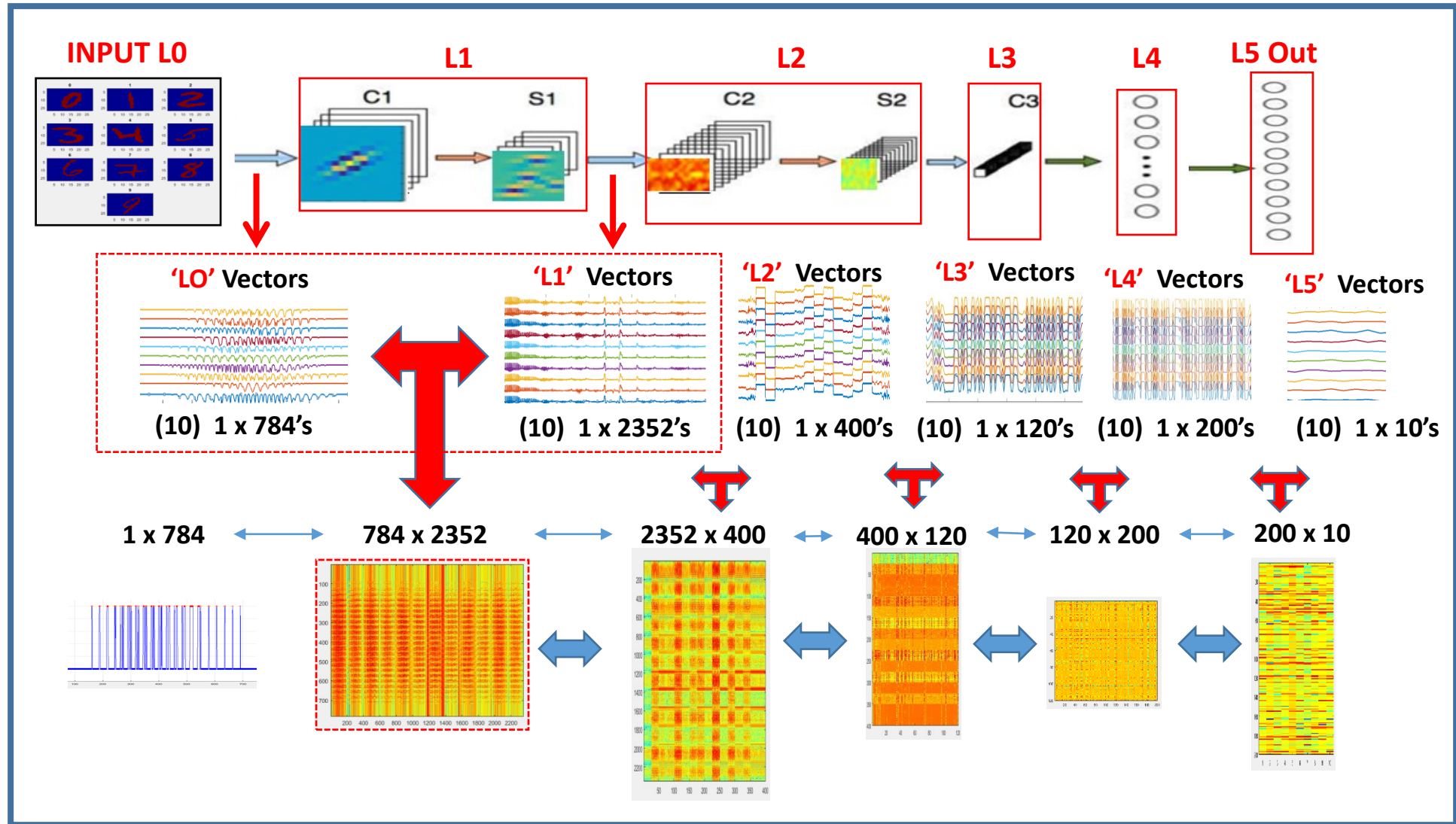
Connecting Layer 1 to Layer 2
Feed Forward processes vs. Bi-directional Matrix

Multiple Processes

Connecting two layers			
CNN		BNN	
Image IN	L1	Image IN	V1
	28x28		1 x 784
1. 12 KV1 Filters		1. Matrix	784x2352
2. 2D Convolution			
3. Tanh compress			
4. Down sample by 4			
5. Multiply by weights			
6. Add biases			
7. Tanh compress			
L2 out	(12) 14x14	V2 out	2352x1

Single Operation
(memory hog)

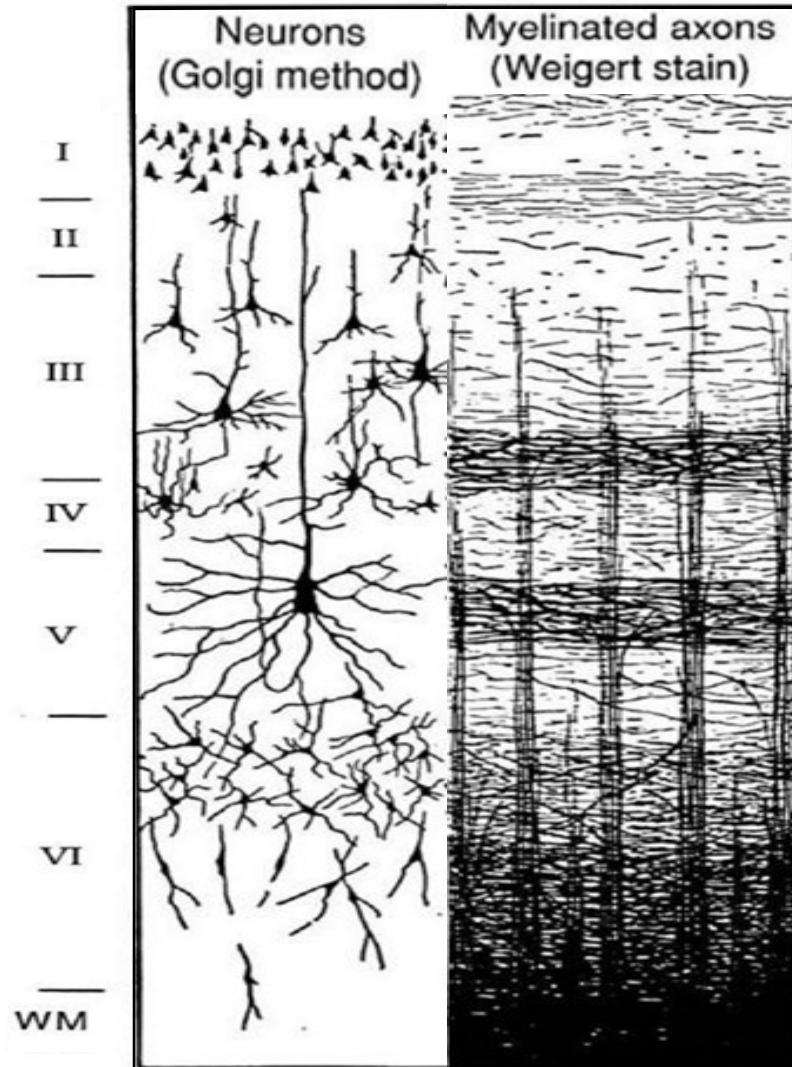
How we convert CNN I/O pairs into BNN I/O pairs



Note: $(784 \times 2352) \times (2352 \times 400) \times (400 \times 120) \times (120 \times 200) \times (200 \times 10) = (784 \times 10)$
 $(A \times B) \times (B \times C) \times (C \times D) \times (D \times E) \times (E \times F) = (A \times F)$

Bridging the gap with RAINNE and JPAT

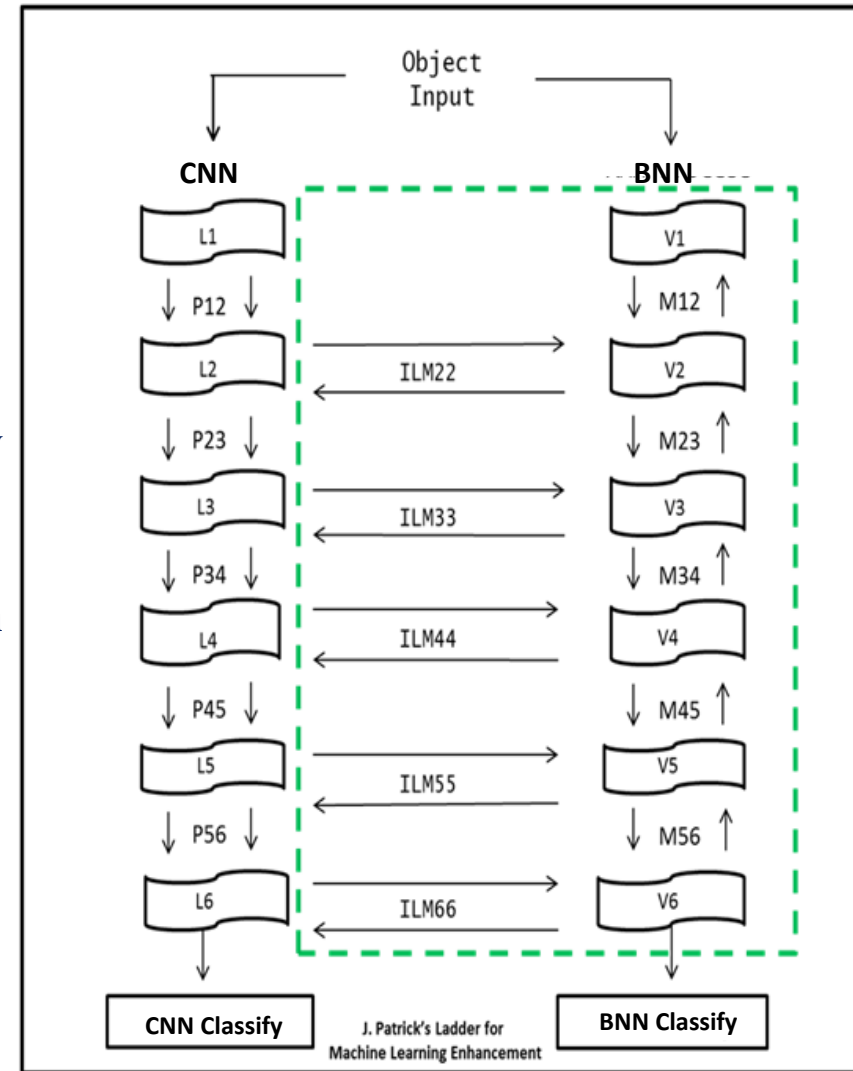
and back to Alexander Bain (1873) and William James (1890)



Vertical and horizontal connections
of the Ventral Pathways

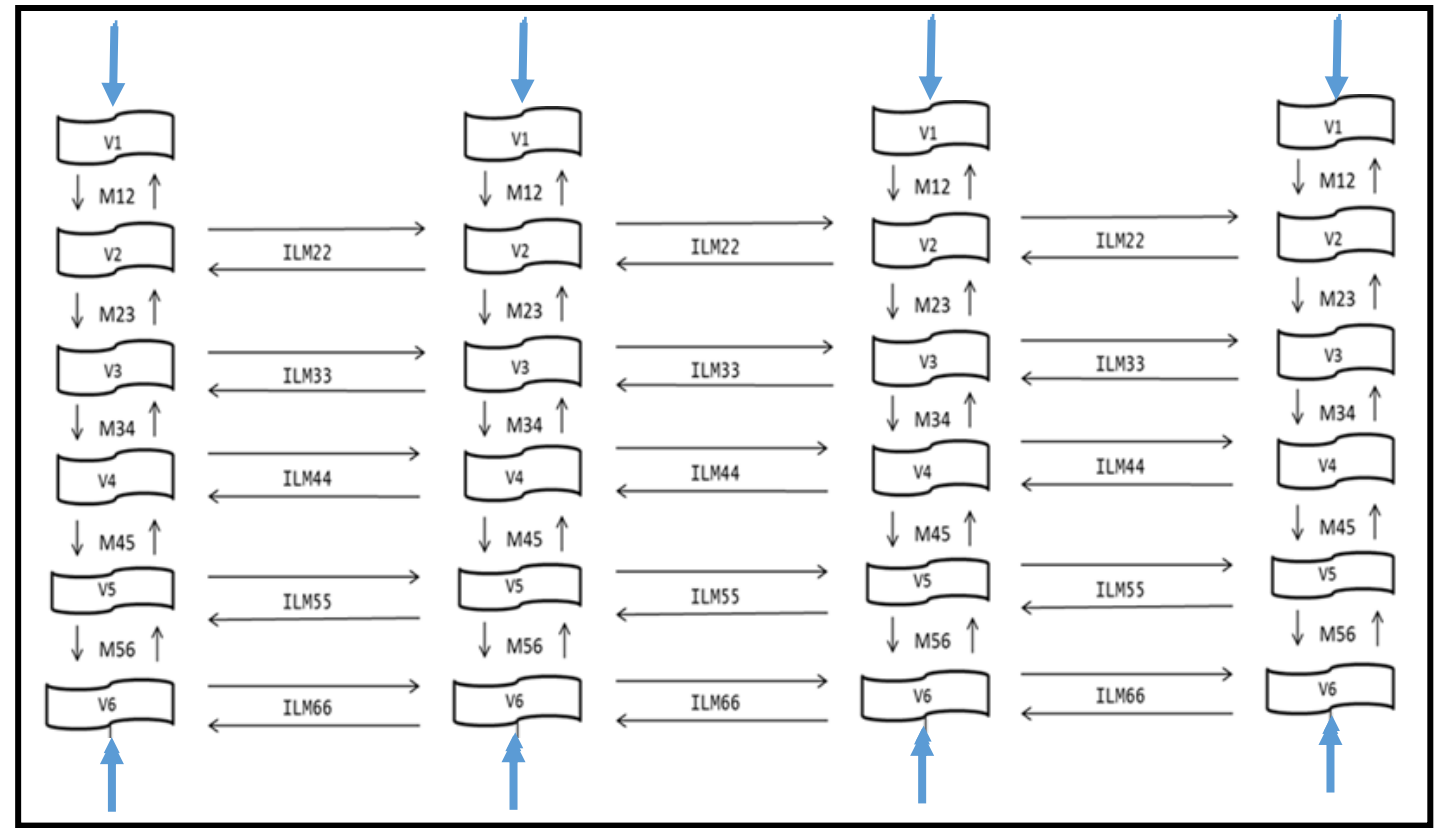
**Real
and
Imaginary
Neural
Network
Emulation**

L-Layer
V-Vector
P-process
M-Matrix
IL-Inter-layer



Joint Proximity Association Template
for (quadri)-directional Neural Networks

Brainstorm applications
 For a
Joint Proximity Association Template
 in the context of a
Group Assured System
 architecture of
Distributed Intelligent Gateways
 we must associate vectors of 'number'

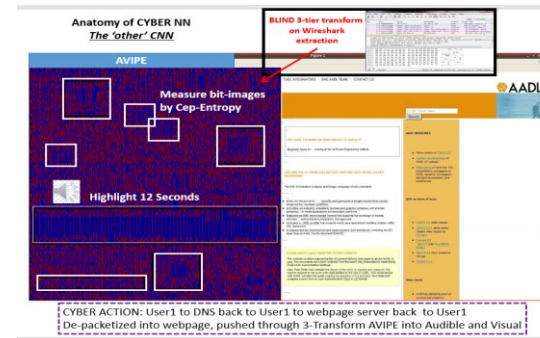


Sensor integrated
Quadri-directional
classification
 EO/IR/Temperature
 adding Recurrent NN
 to modulate and update
 (GEOSS)
 (IoT and BlockChain)

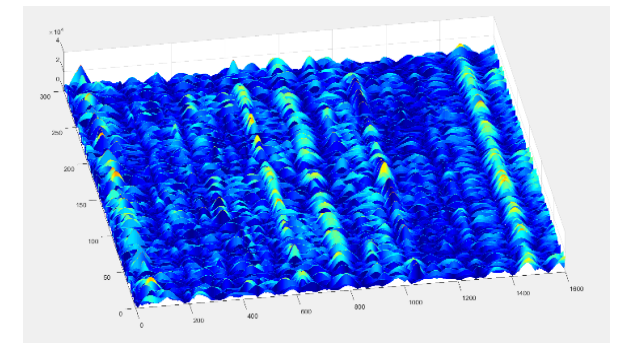
Revisit Kanerva's
Sparse Distributed Memory
 for
rapid search and association

- **Memristor-based NN**
with cross-talk mitigation
- Adiabatic Quantum NN
- HPC gearing and CDMA

Countering Deception
 in cyber space with
 multiple sensors



Collective Behaviors
 over operational
 Human Computer Interfaces



Thank You SPIE

and conference chairs:

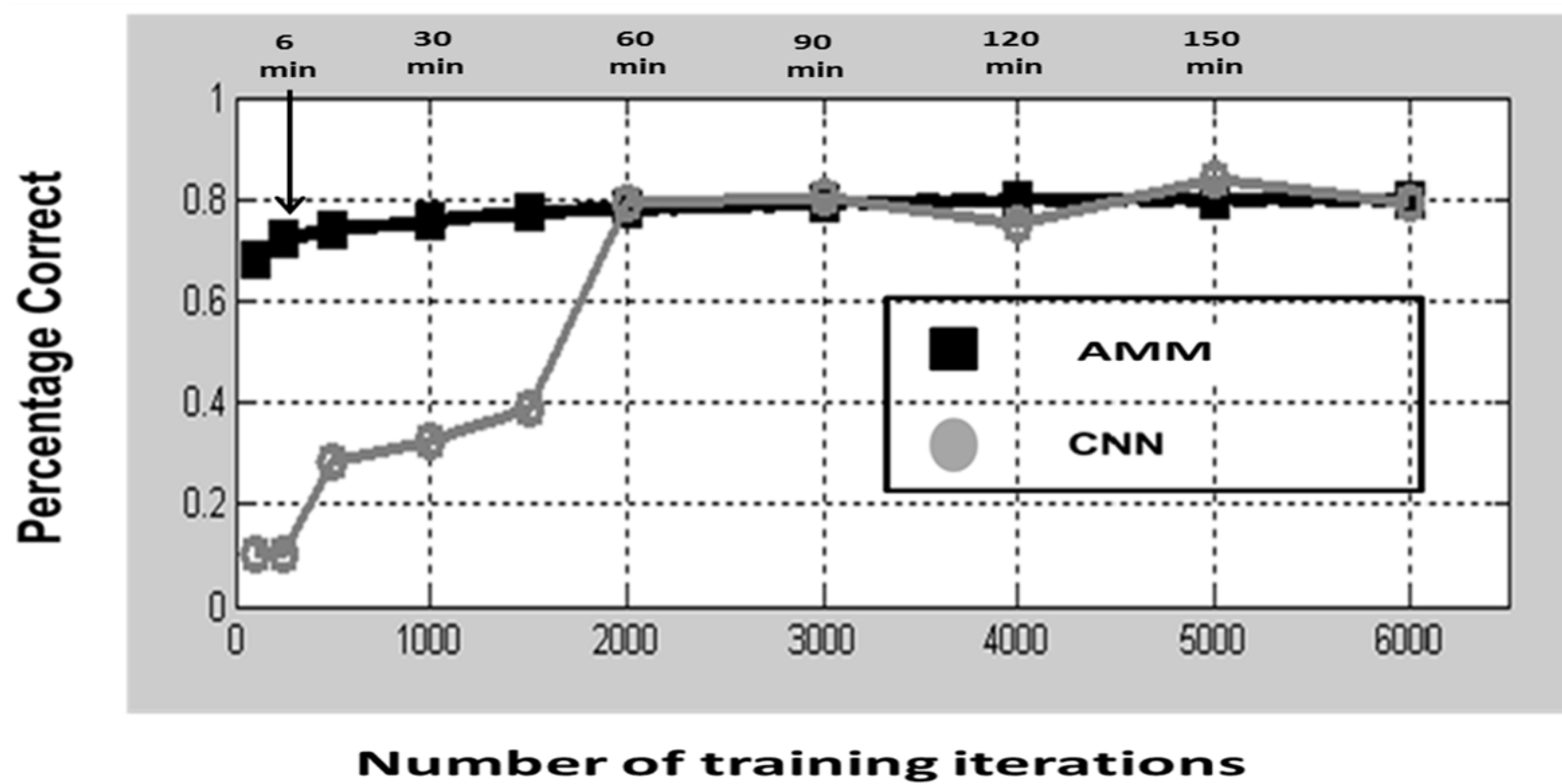
Misty Blowers, Russel D. Hall, and Venkateswara R. Dasari

and also

AFRL, AFOSR, DARPA Innovation House,
Rick Tutwiler ARL Penn State, USPTO, AIPR, Denise

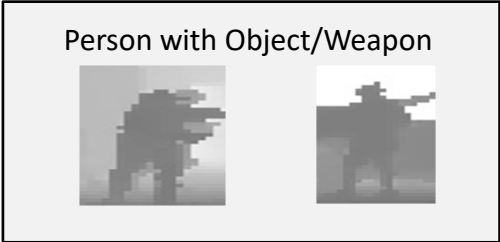
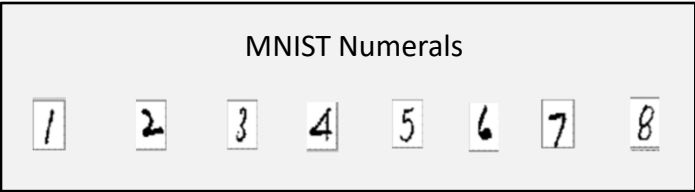
James P. LaRue PhD, Jadco Signals (contact for today's presentation: James@JadcoSignals.com)
Tech Director, ICF Cyber Security Services (contact James.LaRue@ICF.com)

Back-Up



Other applications

Customer	Data Type	Learning Process
Air Force Research Laboratory & Air Force Office Scientific Research	Hand Written Numerals MNIST Data Set	CNN + Perceptron
Neurotechnology	Biometric Data Fingerprints	Perceptron Only
Defense Advanced Research Projects Agency	Image Data Armed Personnel	CNN + Perceptron
Pennsylvania State & Army Research Laboratory	Video Data Person with Object/Weapon	CNN + Perceptron



J. Patrick's Ladder Results			
CNN + Perceptron Cases	Hand Written Numerals	Training	Execution
	Armed Personnel		
	Person with Object/Weapon	10X	20X
Perceptron Only Case	Fingerprints	Training	Execution
		10X	3X

Chartier, S. and Boukadoum, 2006, *A bidirectional heteroassociative memory for binary and grey-level patterns*. *IEEE Trans. Neural Netw.*, 2006, 17, 2, 385–396. (follow on paper in 2011)

Acevedo-Mosqueda et al., *Bidirectional Associative Memories: Different Approaches*, *ACM Computing Surveys*, Vol. 45, No. 2, Article 18, February 2013.