

AIPR Workshop 2025

Imagery-related Challenges in GenAI and LLMs

October 13, 2025

George Washington University, SEH

Paper: 1345

**Examination of How an Adversarial or Naive Change to an Image may be
Tracked Visually, Layer by Layer, Using both CNNs and AMMs**

presented by

James P. LaRue, PhD

Jadco Signals

Joining Analog and Digital Communications

James@JadcoSignals.com
www.DataPlasticity.com

Abstract— In the AIPR 2017 Workshop we presented a paper that looked at “The Anatomy of a Neural Network”. The main purpose was to show how to take a simple LeCun-based neural network, with its multiple internal operations, and extract layer-to-layer input/output pairs to form a series of bidirectional association memory matrices (AMMs) from CNN Layers.

From these transformations we showed how to achieve a practical solution that met the expectations of the Universal Approximation Theorem; how we could to collapse a six-layer system into a single layer.

For the AIPR 2025 Workshop we will examine how a change in the image, be it adversarial or just a naïve-rotation of the image, can be physically tracked across each of the CNN/AMMs, when each is examined in vector form, i.e., we will visually exploit the error from the first layer on to the last layer.

There are many approaches to measuring layer vulnerabilities, such as autoencoders, t-SNEs, histograms, and residuals. In contrast to these methods, we do not offer a measure, we are simply demonstrating that we can take a step back and look at how the error propagates to make a more informed decision on designing a measure.

Paper currently up for review in IEEE Transactions on Neural Networks and Learning Systems

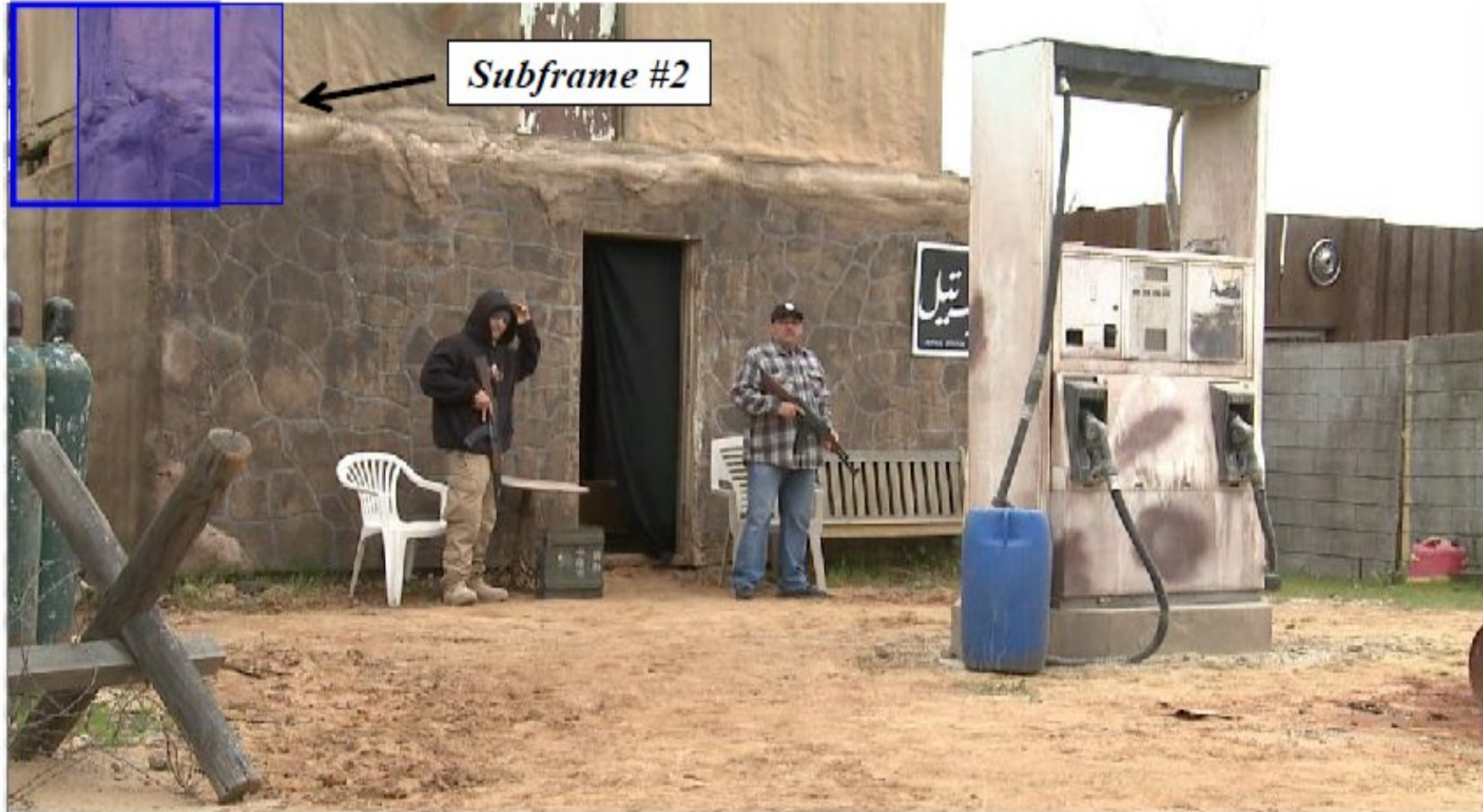
TITLE: Explaining Neural Networks: Hierarchical Backpropagated Ensemble Learning

Manuscript Type: Regular Paper

ABSTRACT: Deep models, characterized by complex structures and end-to-end optimization, proved effective in providing decision support based on real-world data. However, the lack of transparency in their decision-making process and the difficulty in interpreting the role of individual neurons limited their practical applicability in many critical and sensitive domains. Inspired by the parallels between neural networks and ensemble models, where performance was achieved through the collaboration of multiple weak learners, this paper presents a novel perspective that reframes neural networks as hierarchical ensembles. We propose the Hierarchical Backpropagated Ensemble model, wherein each neuron functions both as a base learner and as part of an ensemble of preceding neurons. This framework applies ensemble learning techniques to neural networks, allowing each neuron to focus on specific sub-tasks while progressively constructing a network that meets global objectives. Experimental results on real-world data show that this hierarchical structure enhances the effectiveness of traditional ensemble models, and the ensemble-based explanations offer improved initialization and dynamically adjustable network structures, leading to more efficient training.

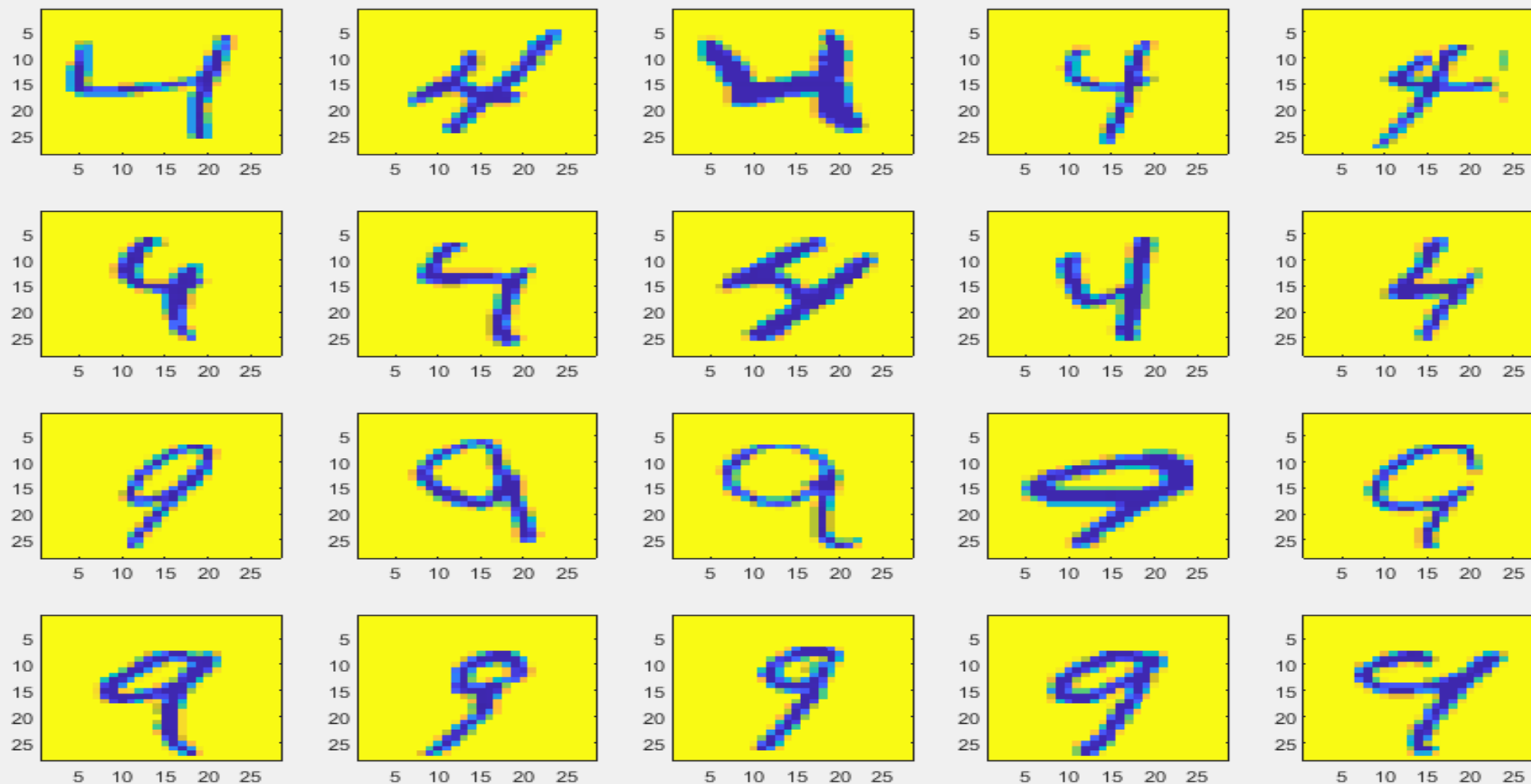
AFRL Information Directorate to Darpa Innovation House 2011-2012

Looking for AK-47's, algorithm found 2, and then found 2 more...



Adversarial POVs:

What do you see?, What does the machine see?, BUT, what does the Labeler see??



8 years ago

The CNN Black Box



1. Neural Network - NN

Alexander Bain (1873) and William James (1890)
McCulloch and Pitts (1943)
Rosenblatt (1958)
Werbos (1975)
Rumelhart, Hinton, Williams, McClelland (1984)

2. Convolutional Neural Network - CNN

Fukushima (1980)
LeCun and Bengio (1995)

3. Bi-directional Associative Matrix - BAM

Kosko (with Kohonen and Grossberg) (1988)
stability issue.

4. Scale-Invariant Feature Transform - SIFT

David Lowe (1999-2012)

5. CNN + GPU (The fast computer brought it back)

Krizhevsky, Sutskever, Hinton (2012)

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.*

**IEEE Applied Imagery Pattern Recognition
Workshop**

**Session II, Deep Learning, Session Chair: Moody
October 10, 2017**

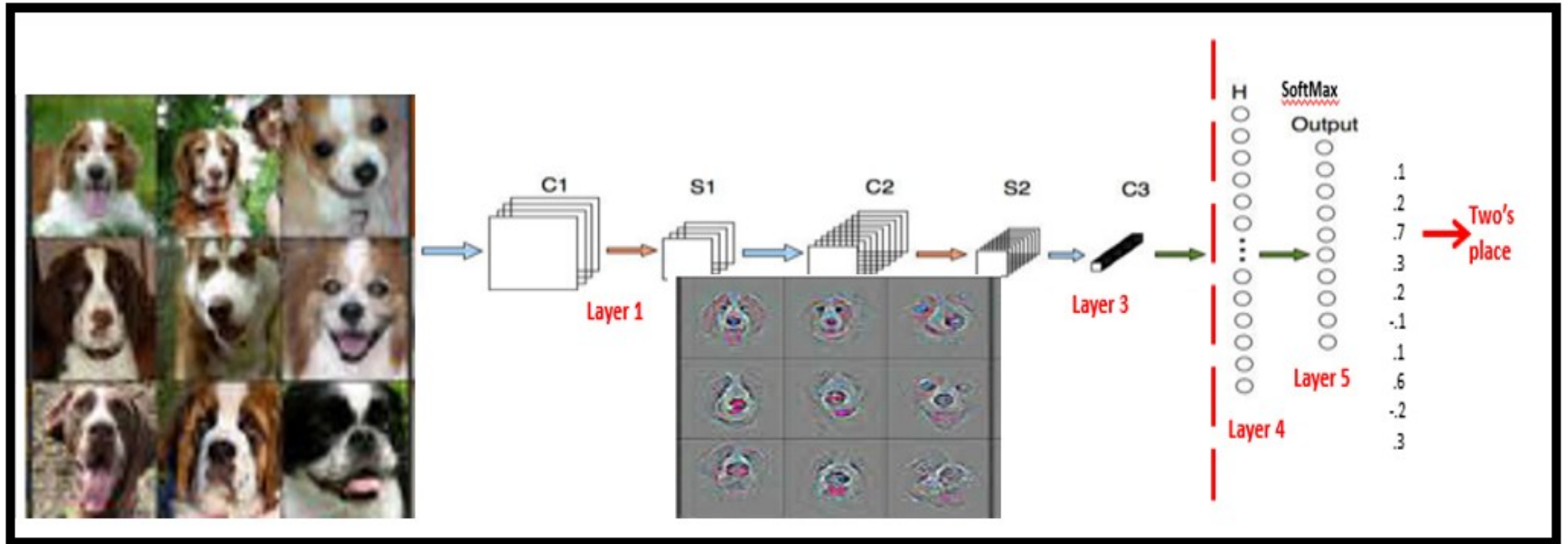
The Cosmos Club, Washington, DC

TITLE: *The Anatomy of a Neural Network*

J. LaRue, R. Tutwiler, D. LaRue

CNN System

Images to Separation to Decision



input
matrices

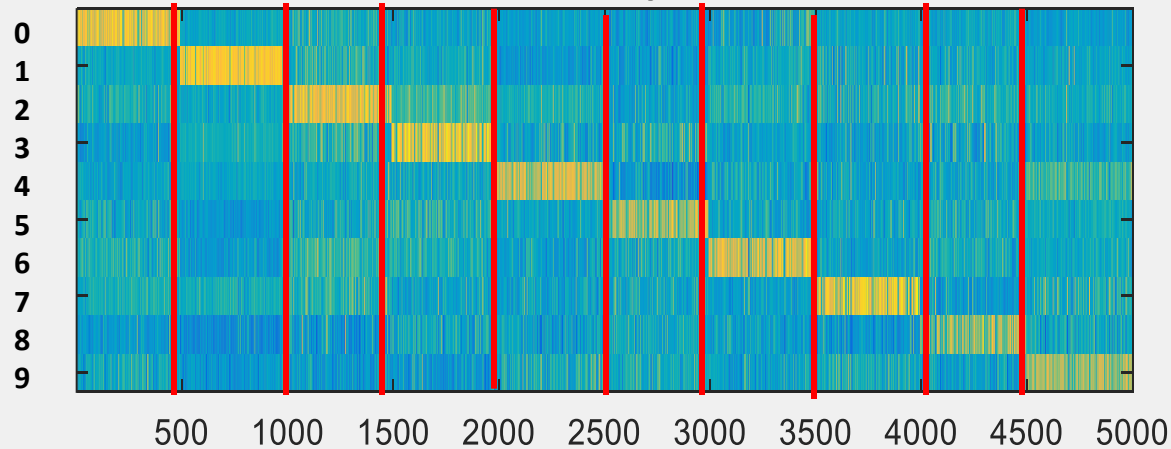


Feature mapping
matrices

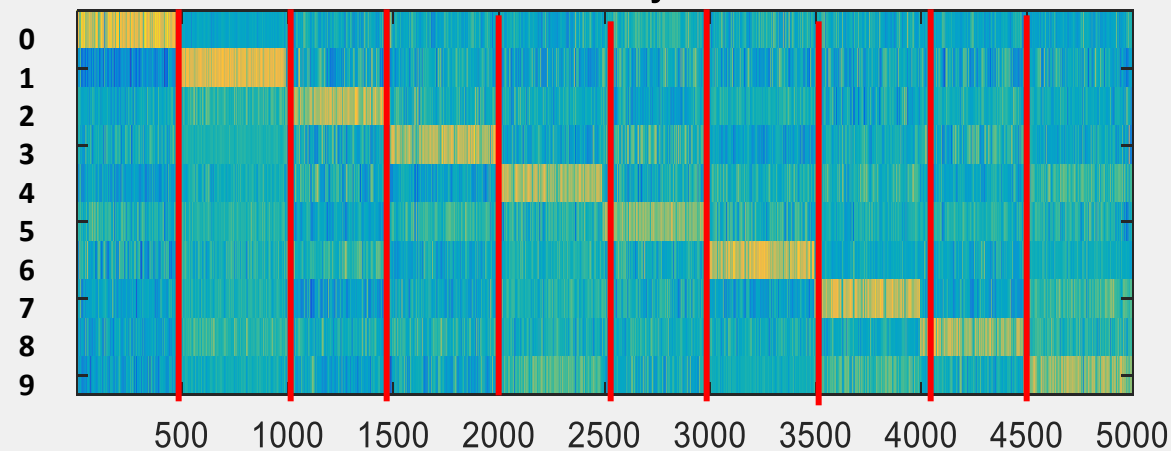


Flattened
Vectors

CNN Accuracy 0.8382 (in 60 minutes)



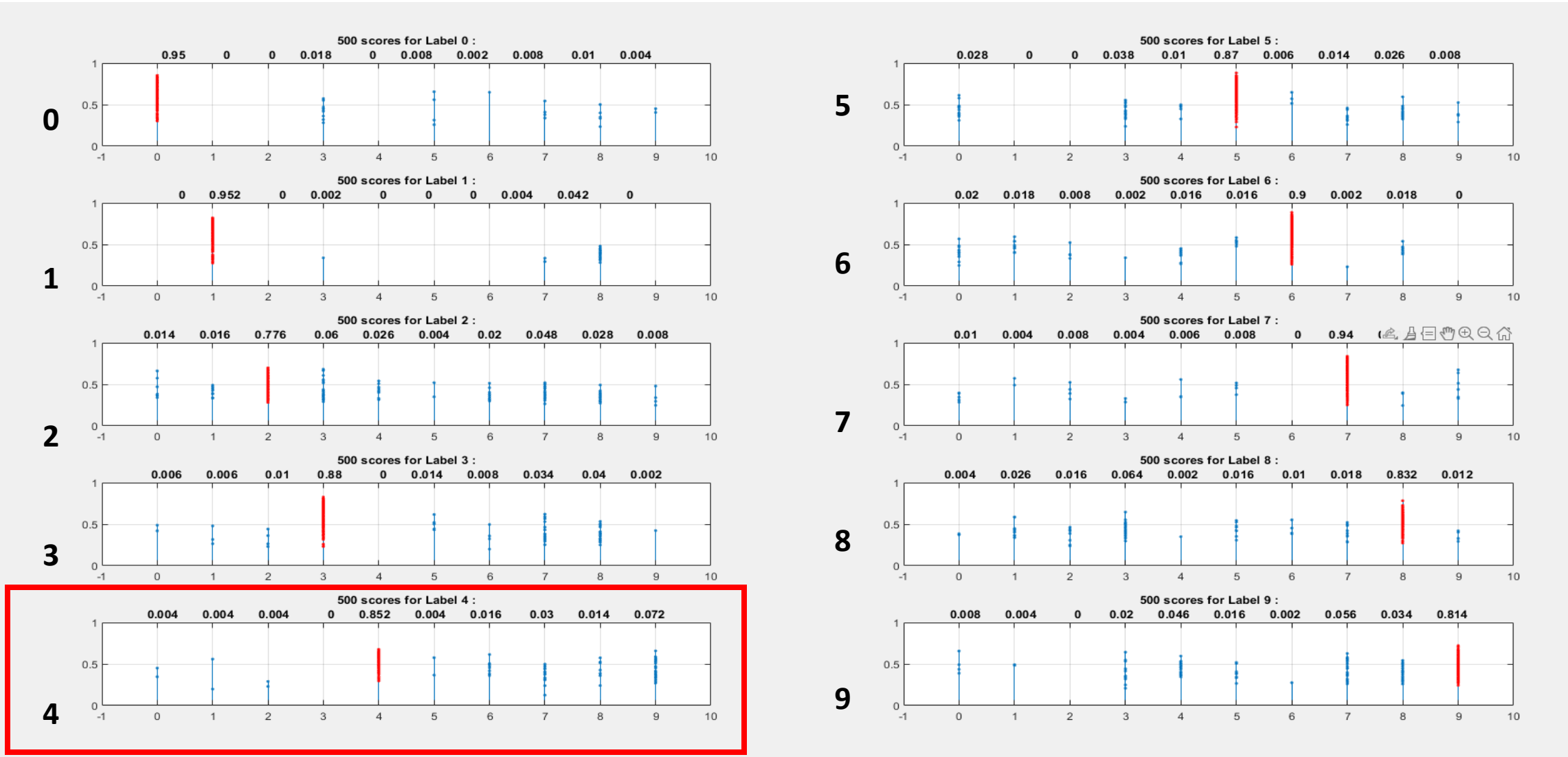
AMM Accuracy 0.8026 (in 2 minutes)



Presented a result to contrast CNN vs AMM using 500 training examples from each of the 10 classes. Yes, more training, better accuracy, but in looking at the MNIST LABELED data set, started thinking, what do those ~ 20% fails look like? Then I looked at the scores for each class, and then stated thinking, in the context of adversarial, are we are own worst enemy,???...

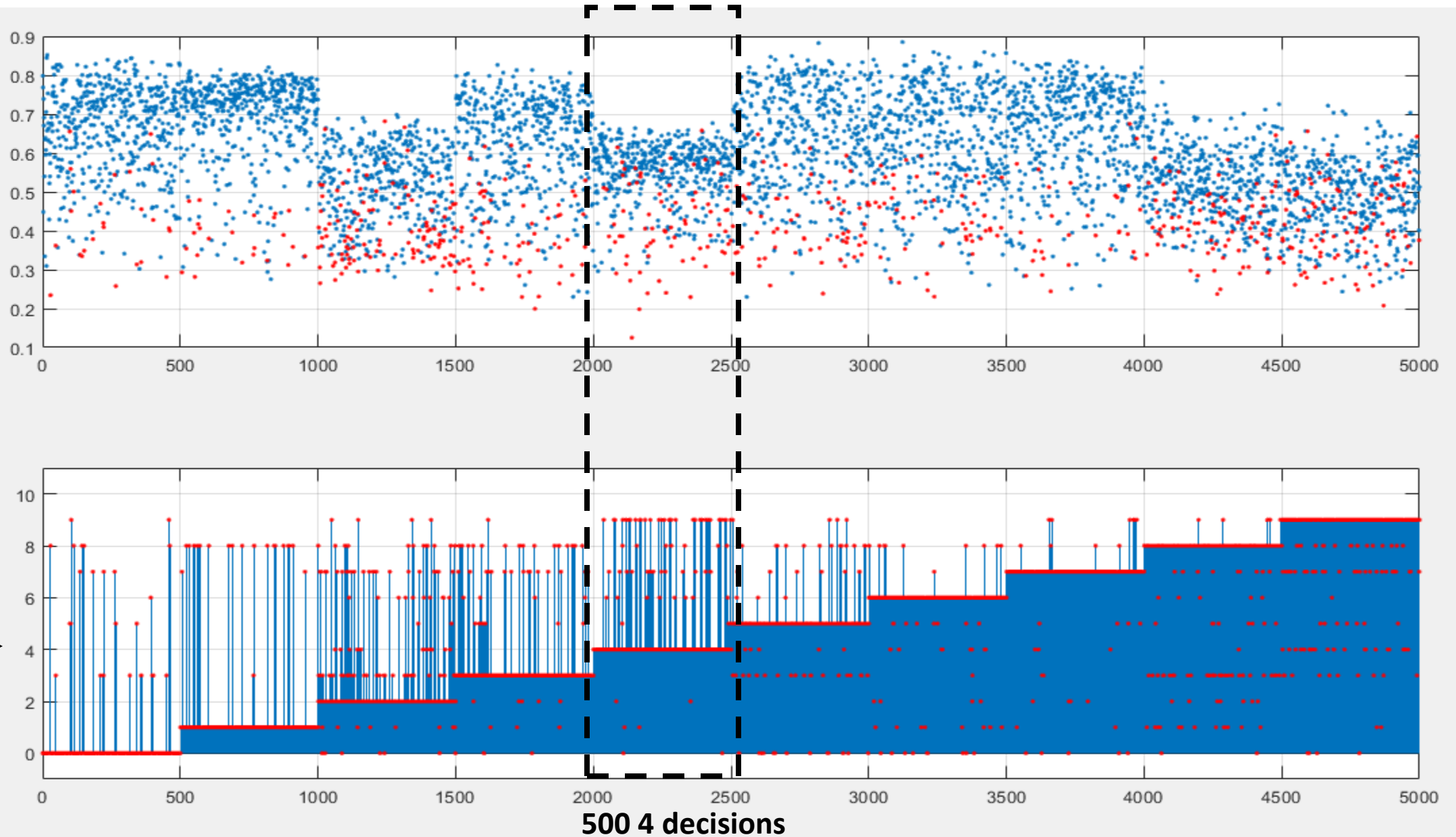


From the 500 each test, the scores, (these are not probabilities)

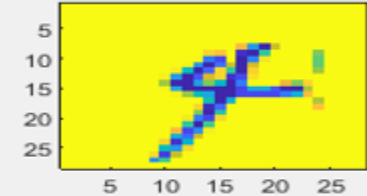
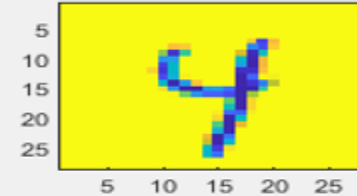
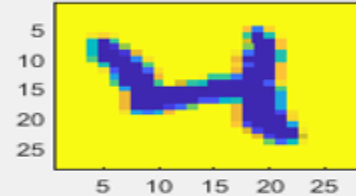
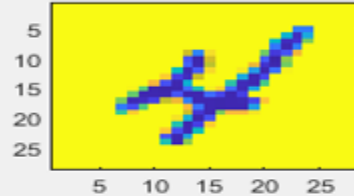
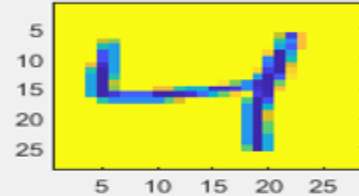


85 of the 4's decided correctly, 7% decided as 9's,...but what were the scores that led to the decision?

Identifying the scores and the winners for each class

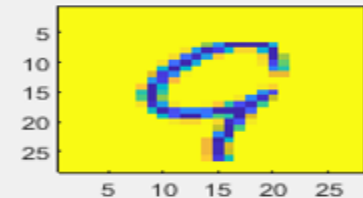
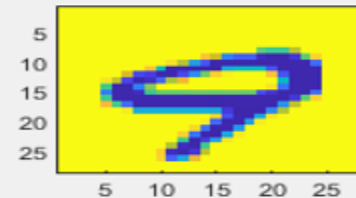
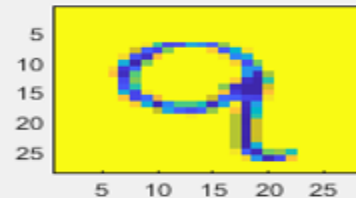
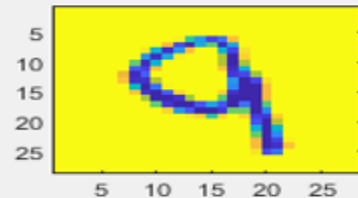
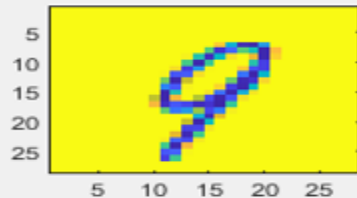


Scores for each of 10 Labeled 4's

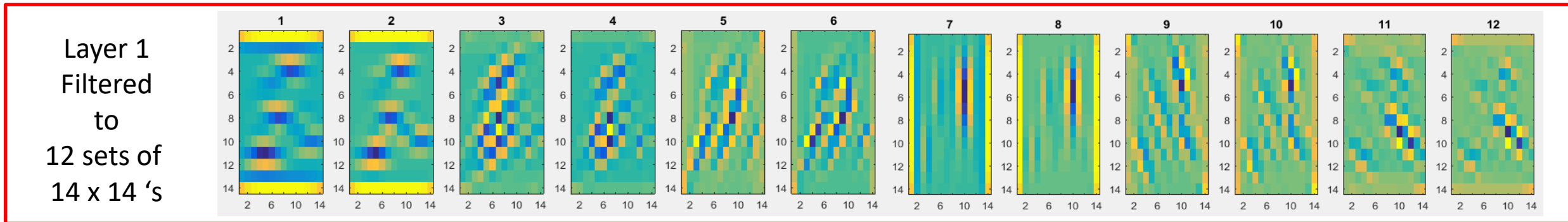
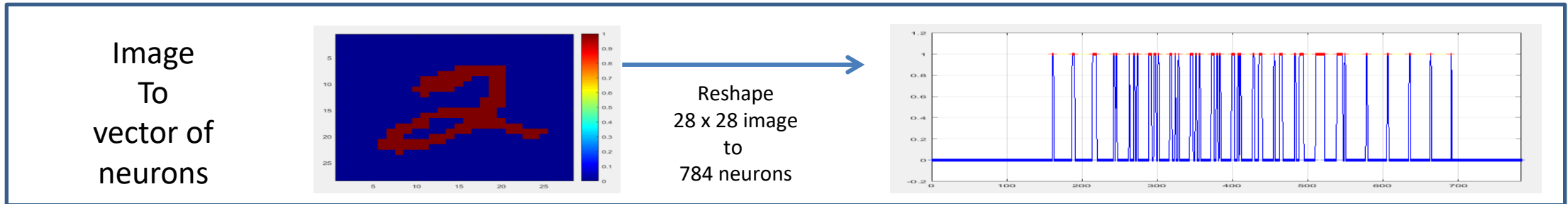


0	0.1498	-0.1319	-0.1376	0.1806	-0.0158
1	0.1636	0.1085	0.1779	0.0992	0.0110
2	-0.0861	-0.0357	-0.1479	-0.0152	-0.0554
3	0.1412	0.0257	0.1101	0.0802	0.0668
4	0.5790	0.6126	0.6040	0.5563	0.3089
5	-0.0430	0.1395	-0.1046	0.0186	0.1069
6	-0.1345	-0.1017	-0.1118	-0.0801	0.1510
7	0.0251	-0.0148	-0.0393	0.1115	0.2594
8	0.1374	0.0520	-0.2076	0.2402	0.2301
9	0.3181	0.1134	0.2813	0.5007	0.2688

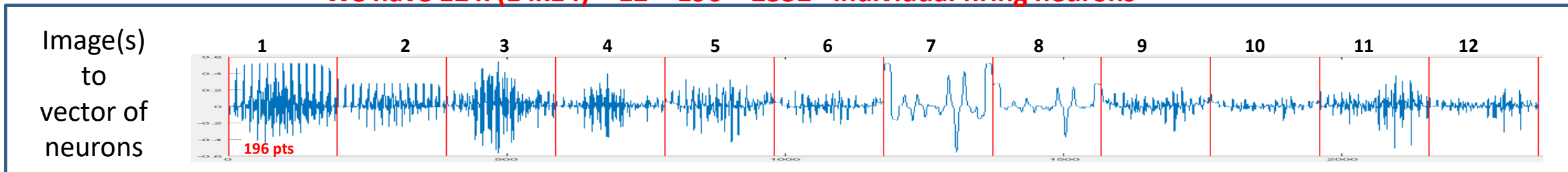
0	0.0037	-0.0608	0.0687	0.0556	-0.1889
1	0.0828	0.0395	-0.1721	-0.0847	0.0175
2	-0.0398	-0.1658	-0.0384	-0.0814	-0.0653
3	0.0105	-0.0398	0.0915	-0.0101	0.1052
4	0.3936	0.3597	0.0780	0.4299	0.3141
5	0.1624	0.3789	0.1724	0.0360	0.1458
6	0.0061	0.0418	0.1929	-0.1262	0.0036
7	0.1028	0.0912	0.2541	0.2934	0.2152
8	0.3280	0.2387	0.2411	0.1882	0.2000
9	0.5106	0.4165	0.3711	0.5428	0.3142



Here: we simply flatten the Feature mapping matrices
(easier to monitor over time)



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

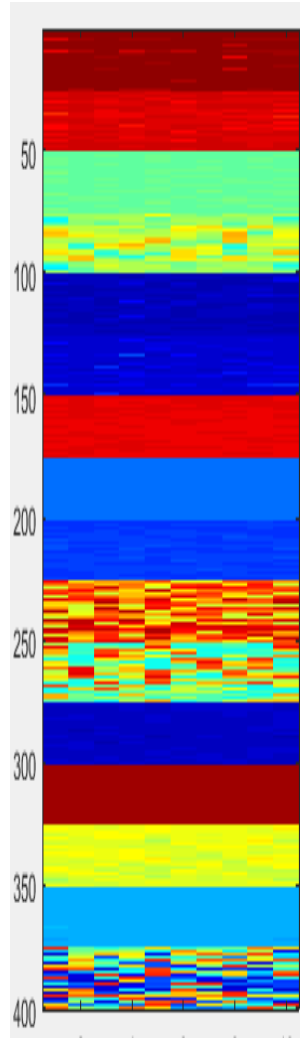
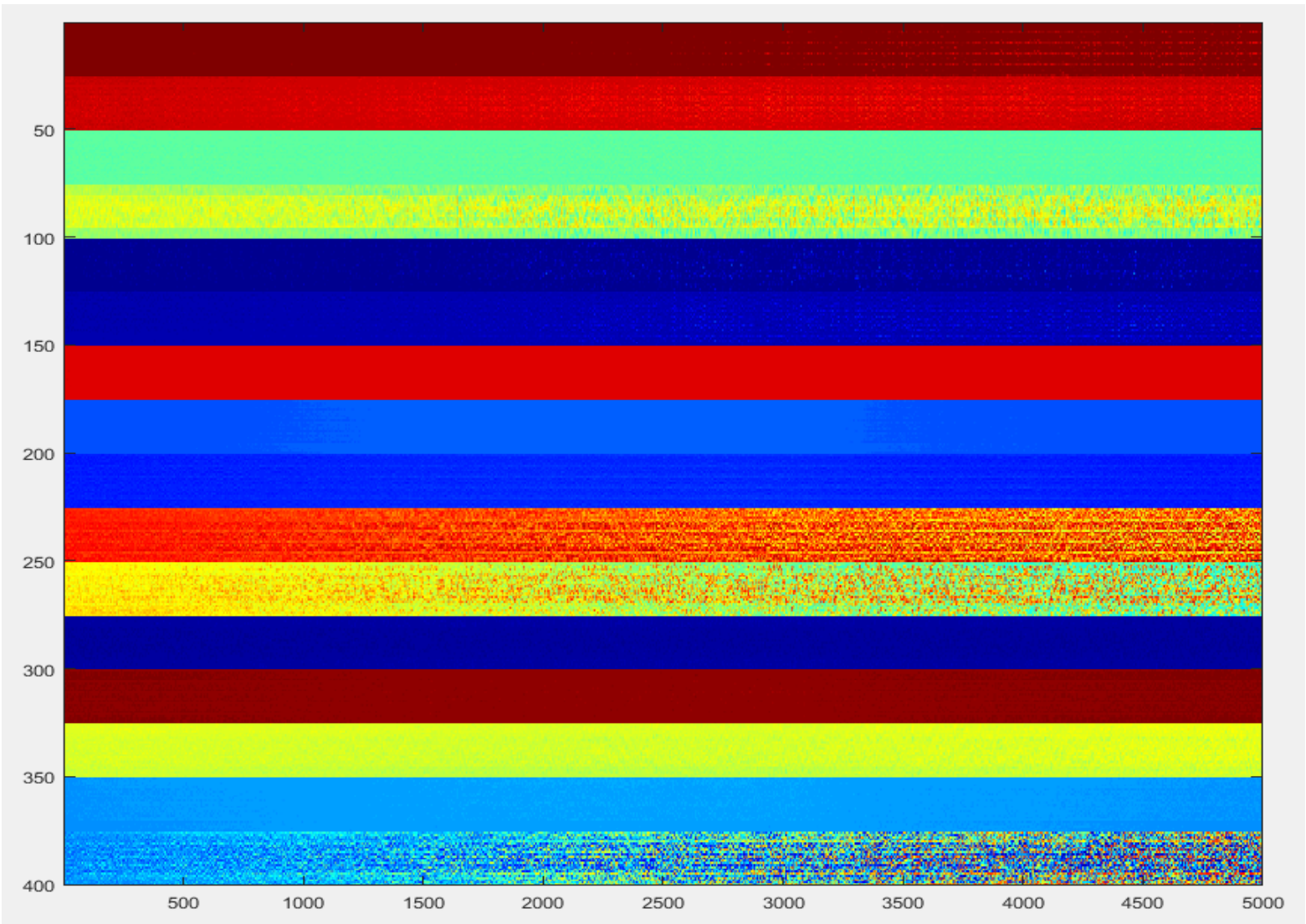
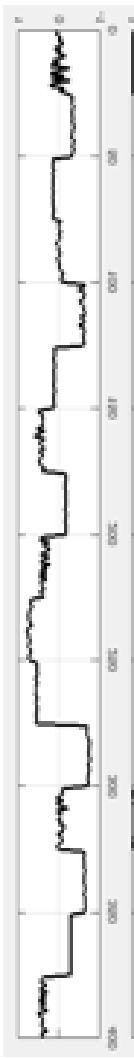


1 →

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

→ 2352

Monitoring the Training of Layer 4 Layer from beginning to ending of training

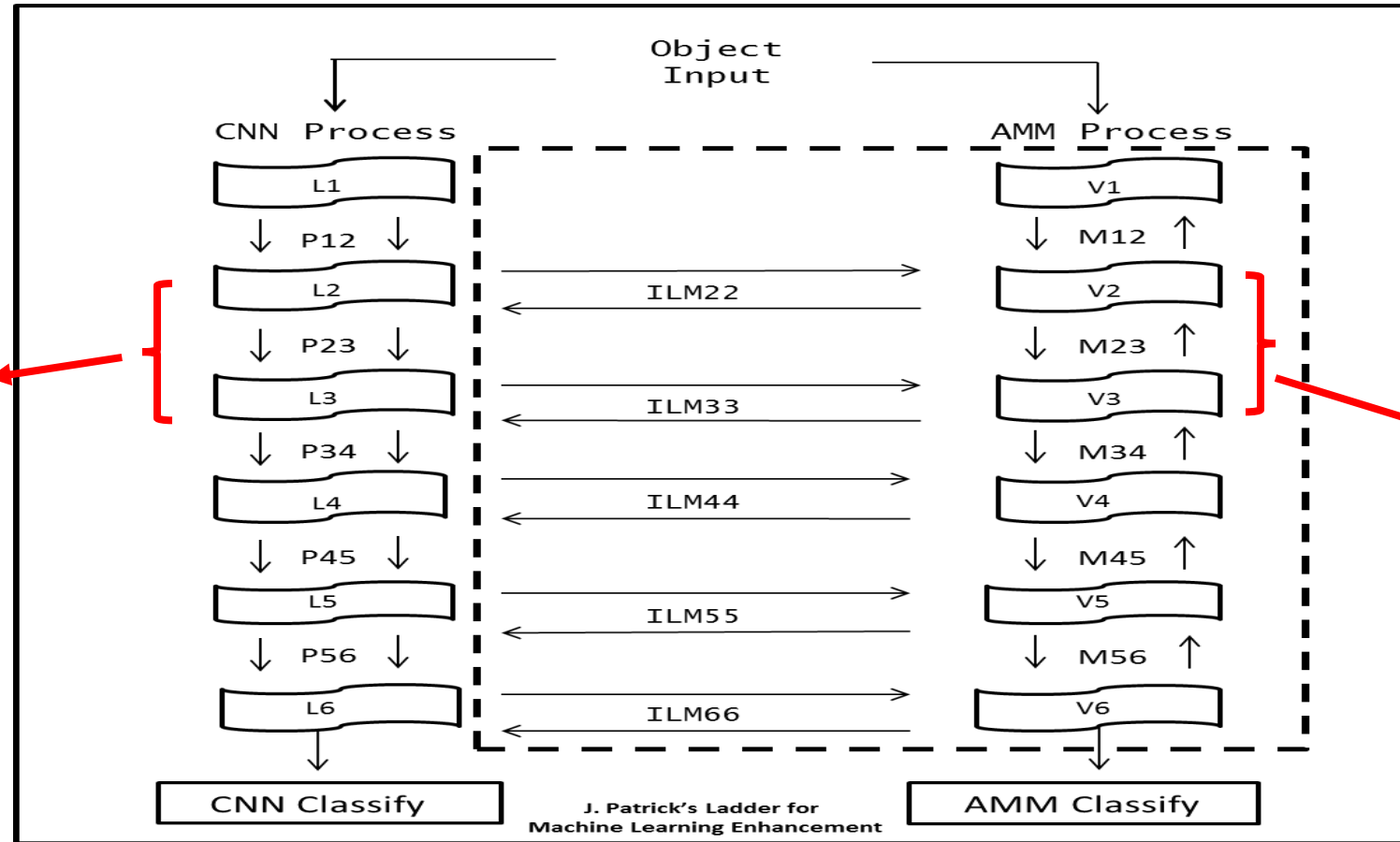


CNN vs AMM

Convolutional Neural Net vs Association Memory Matrix (need to work together)

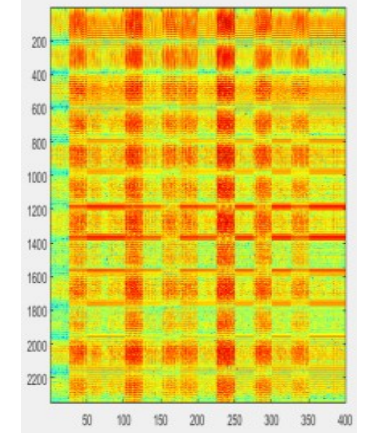
On this side there are several CNN steps per layer, e.g. subsampling, tanh applied, to produce a “convolutional” output

```
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 ;
    %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
```



On this side there one matrix-based step per layer to produce an “associated” AMM output which by-passes internal CNN steps

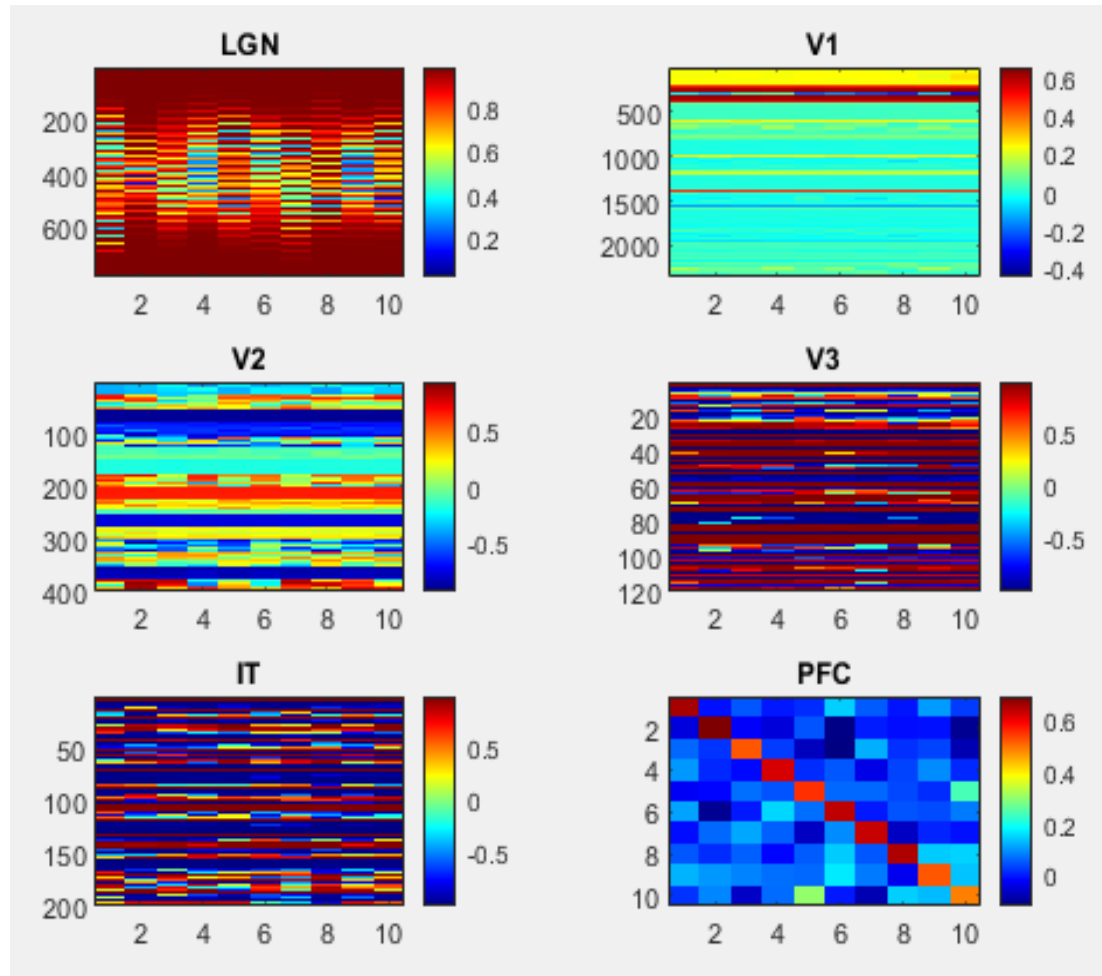
2352 x 400



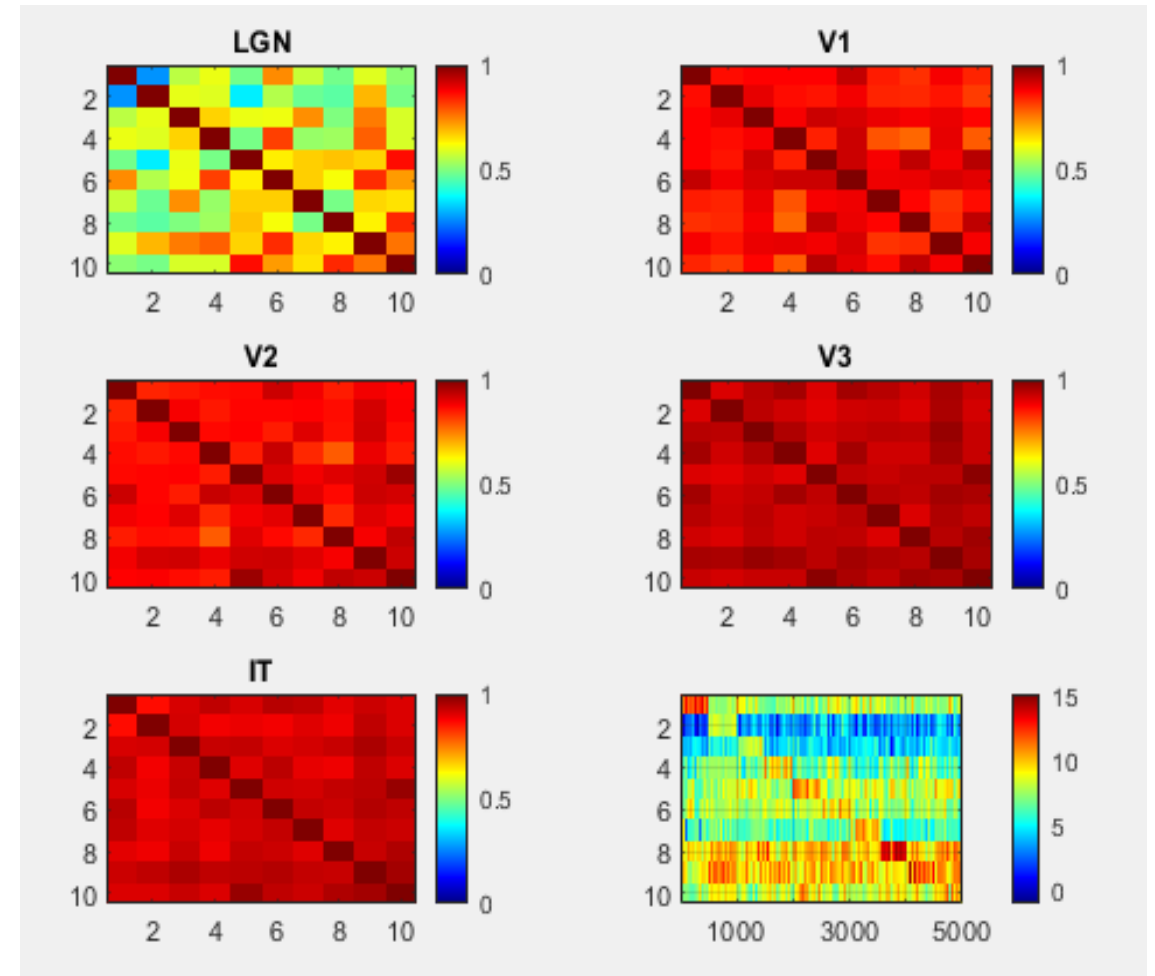
Patent No.: **US 10,242,313 B2**
Date of Patent: **Mar. 26, 2019**

Exploiting Layer Representatives using CNNs and AMMs

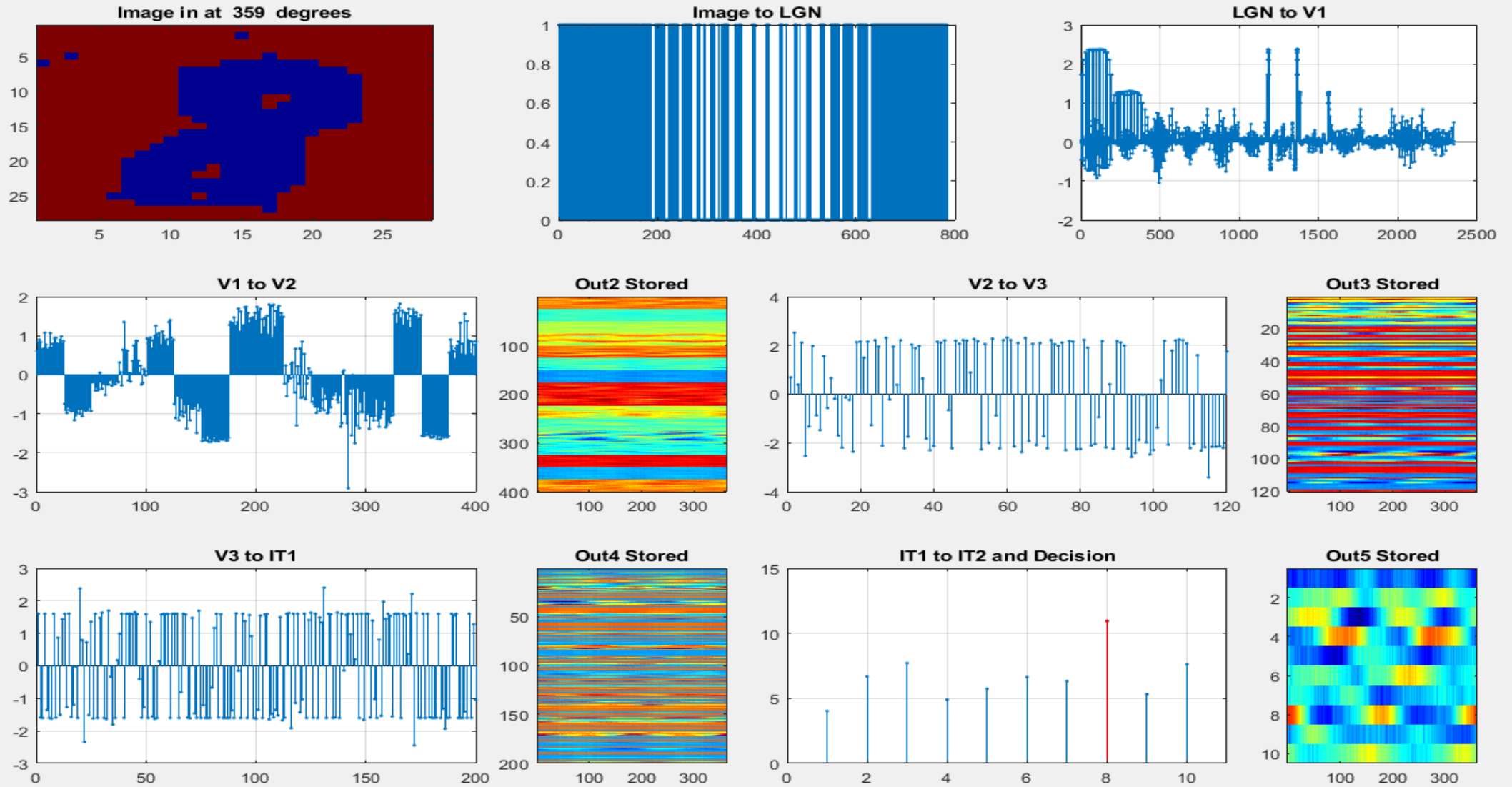
Form 10 Class representatives from each Layer



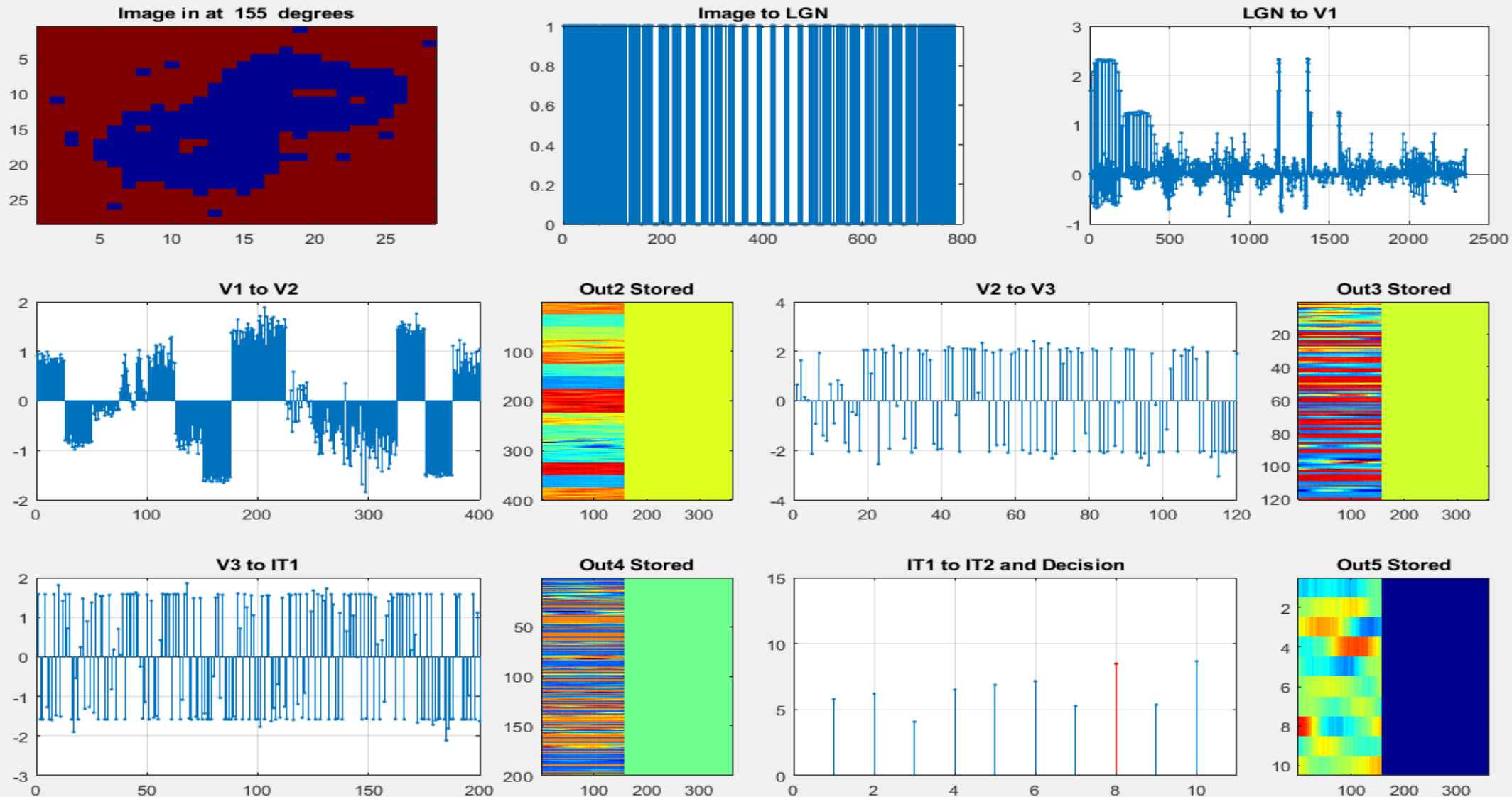
Looking for Diagonal Correlations on the Diagonal



8 Wins!

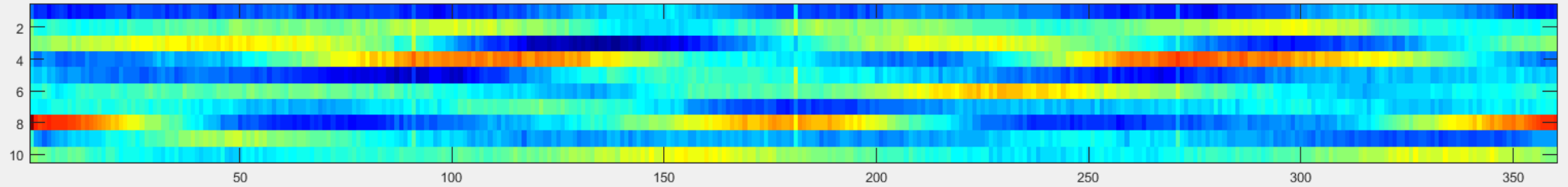


It's Close



Crazy 8's

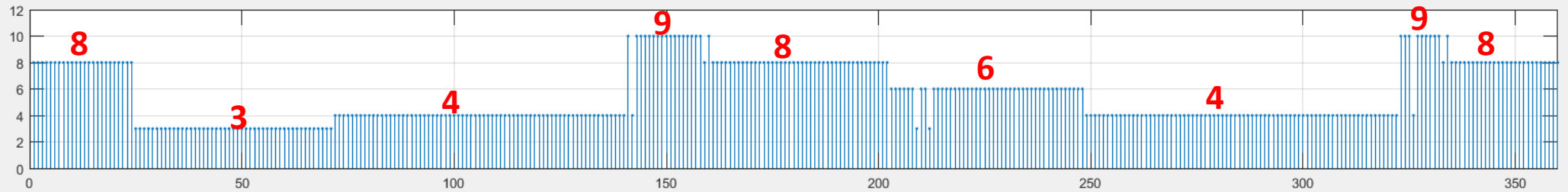
10
Class
Scores
(0-9)



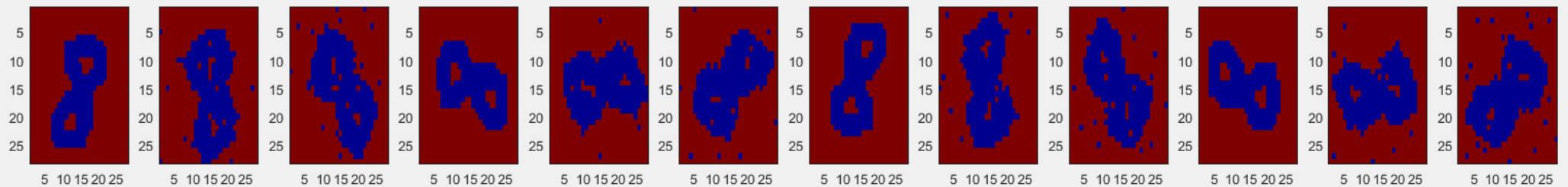
The
winning
score



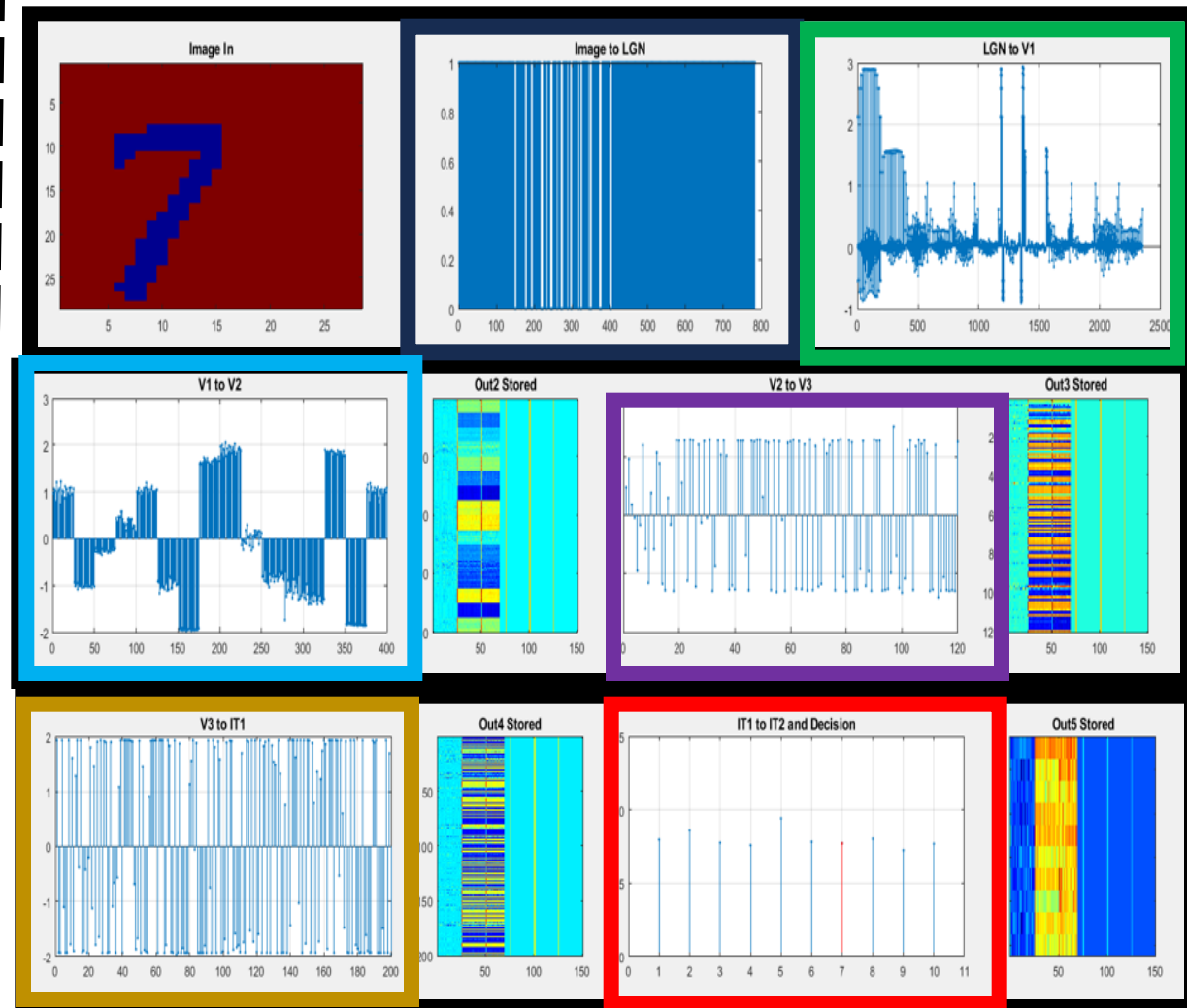
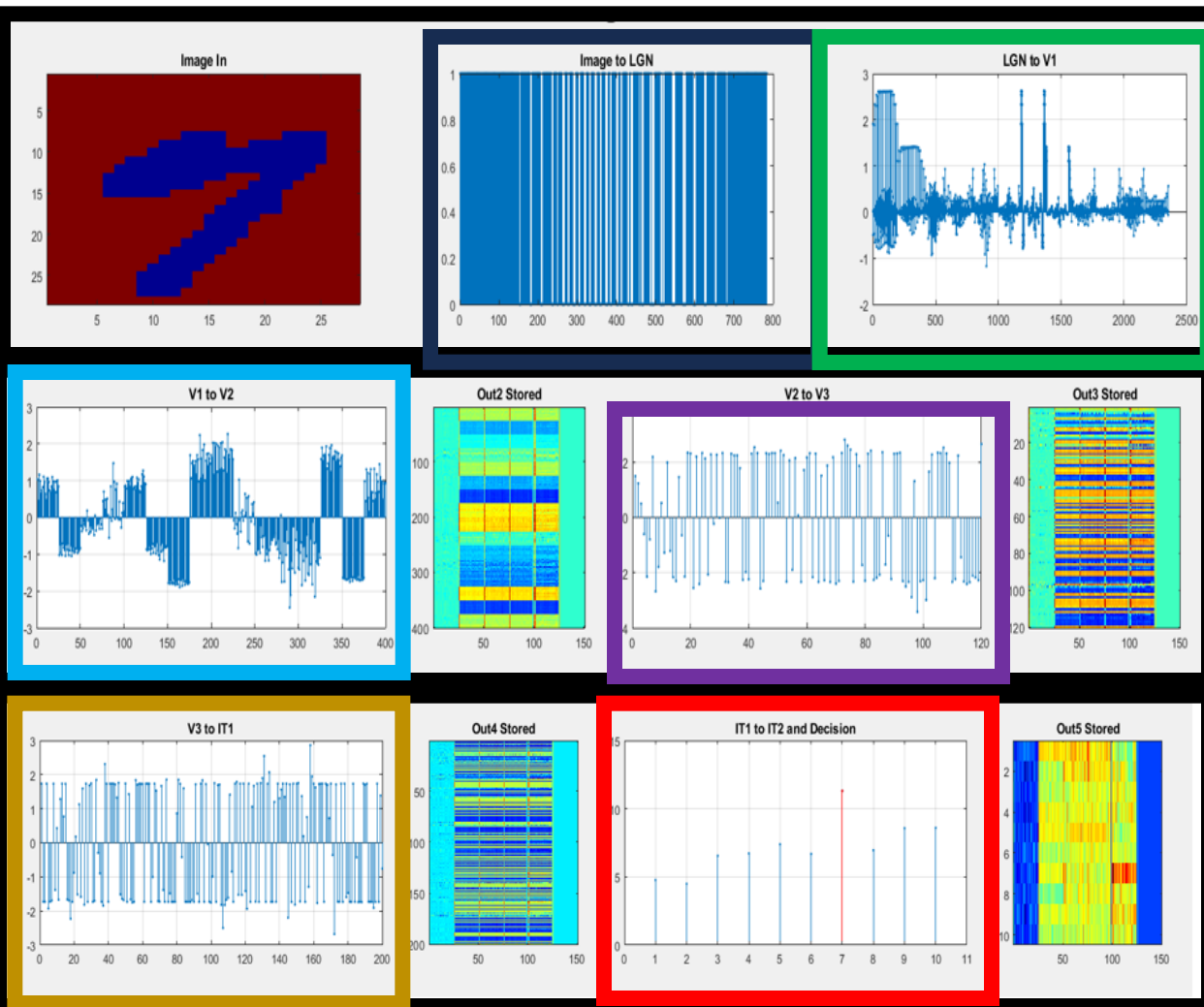
The
decision



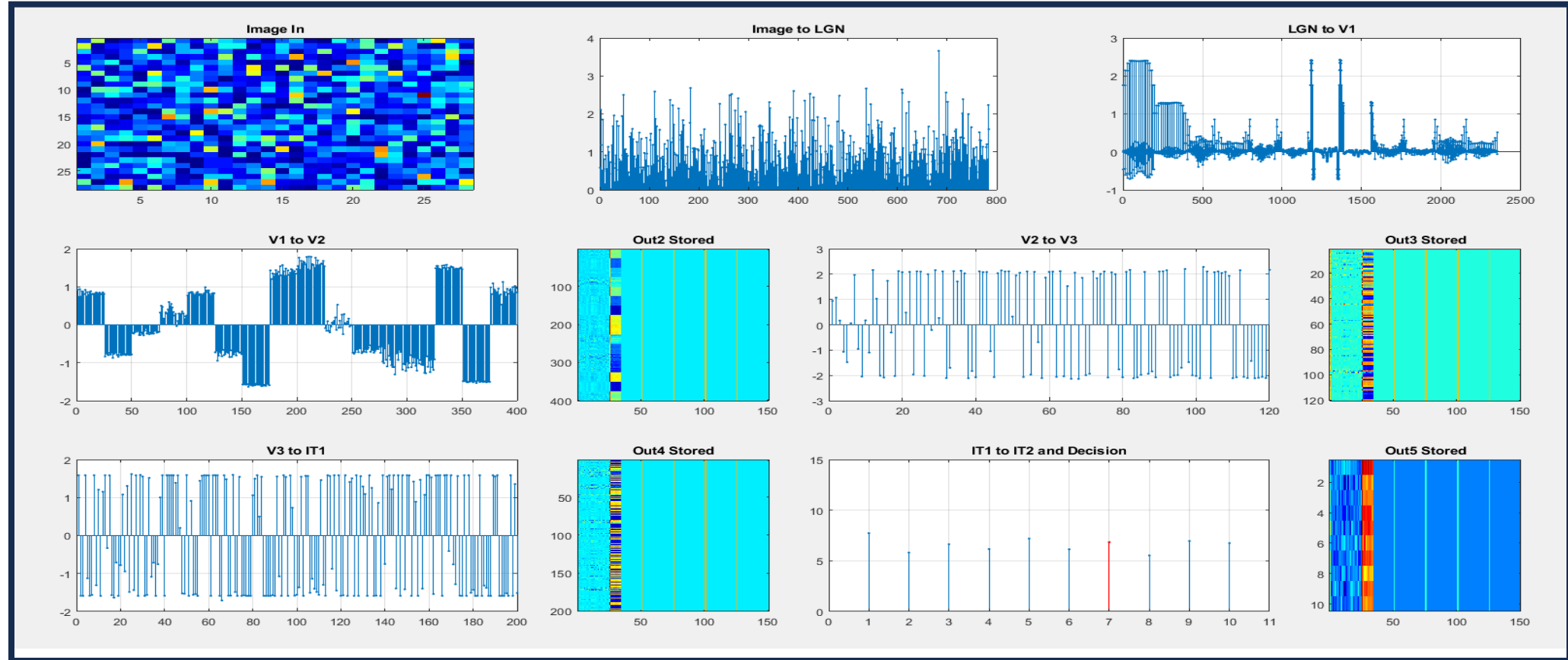
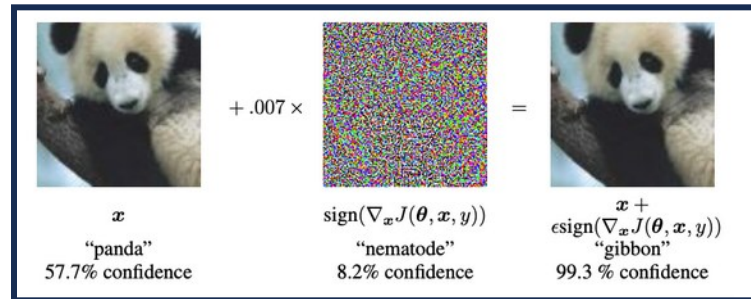
Rotation



Observing Layer by Layer Changes (adversarial or unintended)



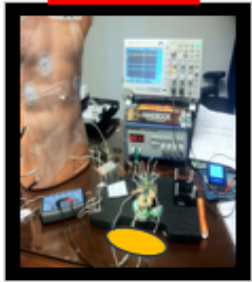
BLUF: Panda or No Panda, Random numbers are destined to be classified



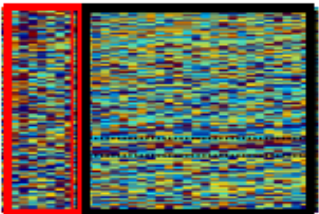
Final Bigger Picture, or Just the Beginning

Bridging the RAINNEs with the Joint Proximity Association Templates (JPAT)

Heart



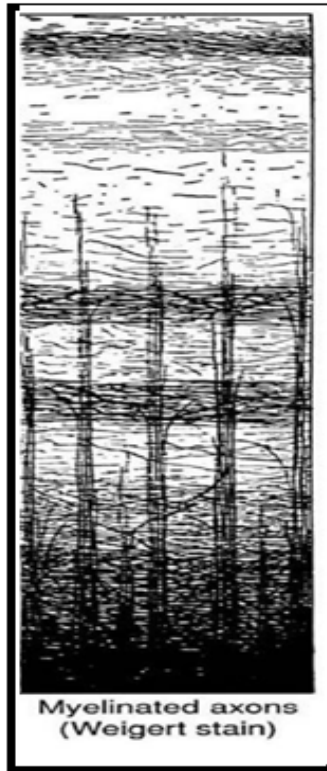
NC002695
Genomes 'talk'



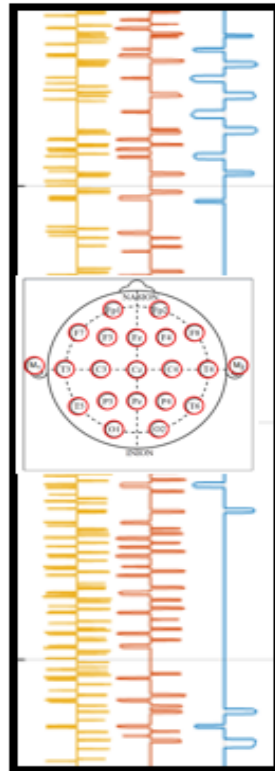
inspired by
John Butler
DNA Forensics

“Real” neurons produce beat patterns
Analyzing EEG with Single-Period Single-Frequency Sinusoids

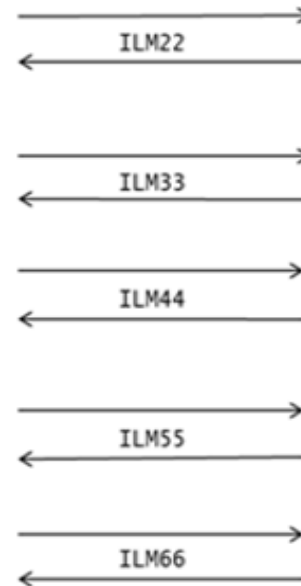
Patent No.: US 11,179,085 B2
Date of Patent: Nov. 23, 2021



Myelinated axons
(Weigert stain)



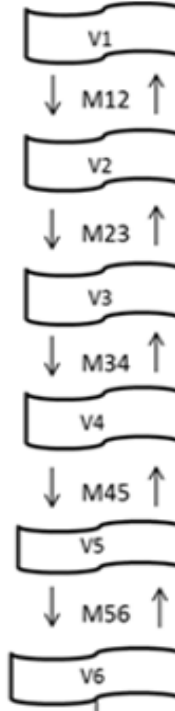
Bridging
Real2Imag



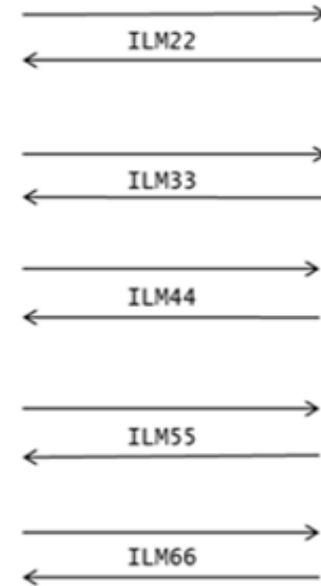
“Imaginary” neurons produce classes
(sight, sound, touch, taste, smell)

JPAT for Neural Networks
Patent No.: US 10,242,313 B2
Date of Patent: Mar. 26, 2019

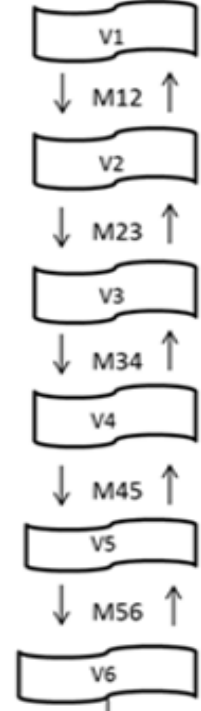
Left Eye



Bridging
CNN2CNN



Right Eye



Thank you AI PR and GWU

Questions?

James P. LaRue
James@JadcoSignals.com
www.DataPlasticity.com