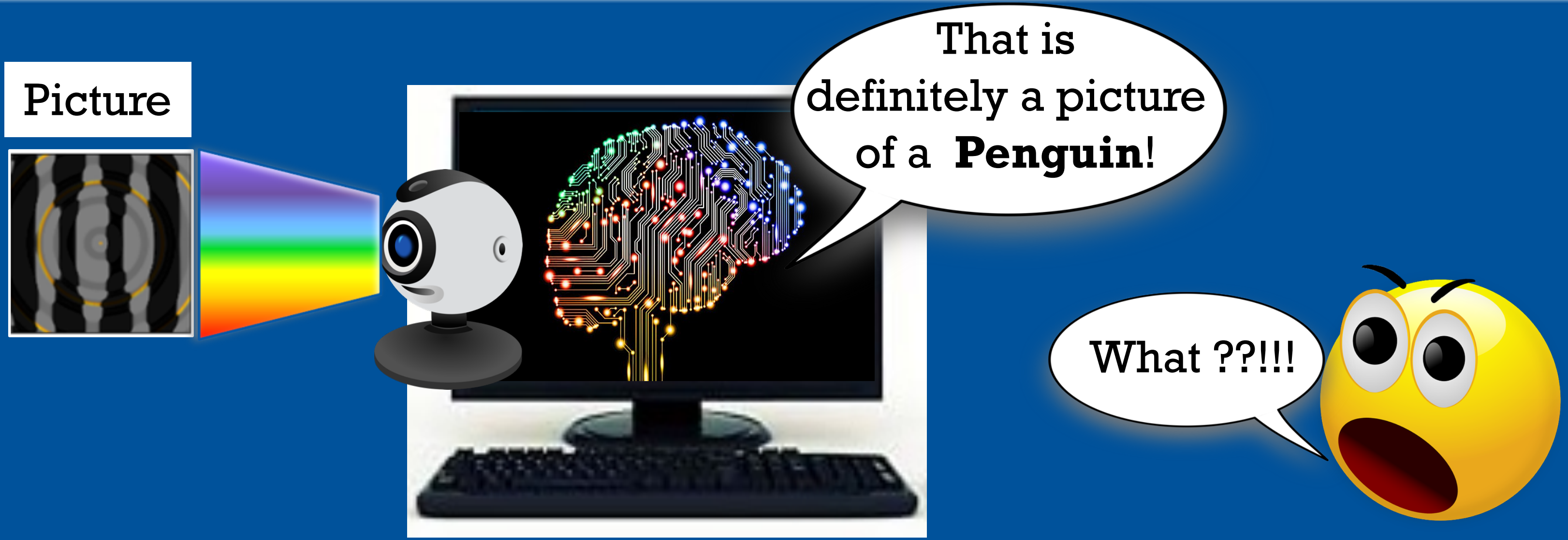


Troubles with Deep Learning

Paul Gader

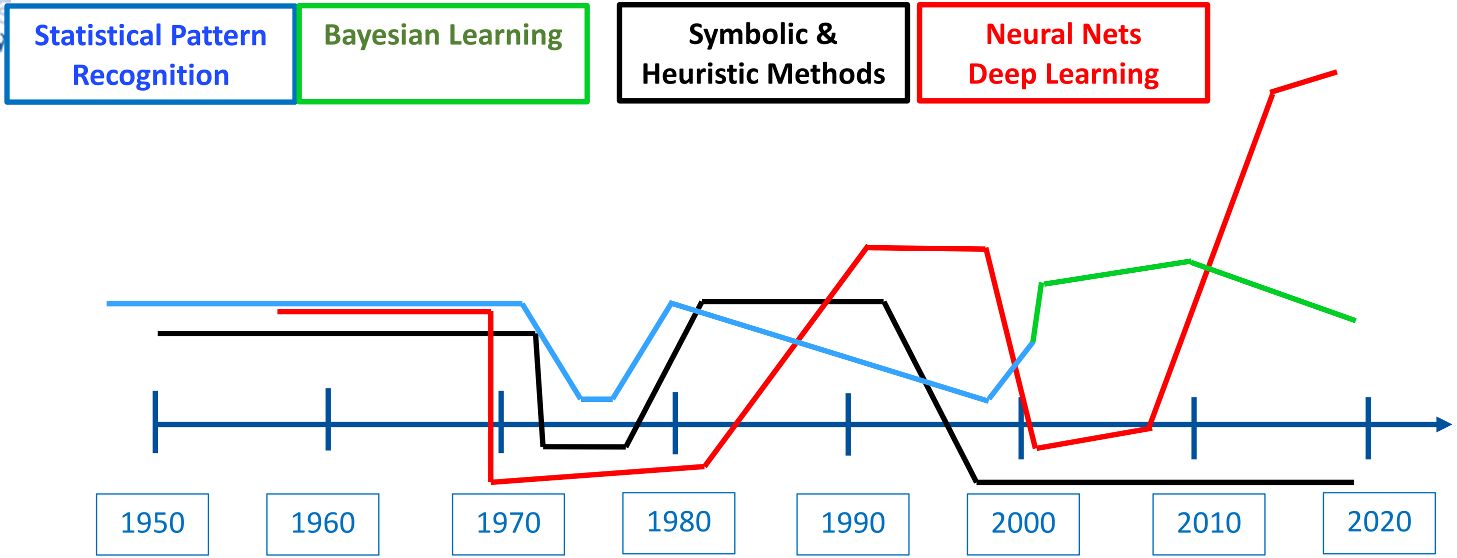
Artificial Intelligence

Computer Systems that demonstrate intelligent behavior or perform tasks requiring perception

- Medical Diagnosis,
- Speech & Handwriting Recognition,
- Natural Language Processing,
- Human-Computer Interaction,
- Big Data Sets

- Computer Vision,
- Sensor Networks,
- Signal/Image Analysis
- Remote Sensing
- Playing Games,
- Robotics

Simple Illustration of History of AI



Artificial Neural Networks (ANNs) driving field currently

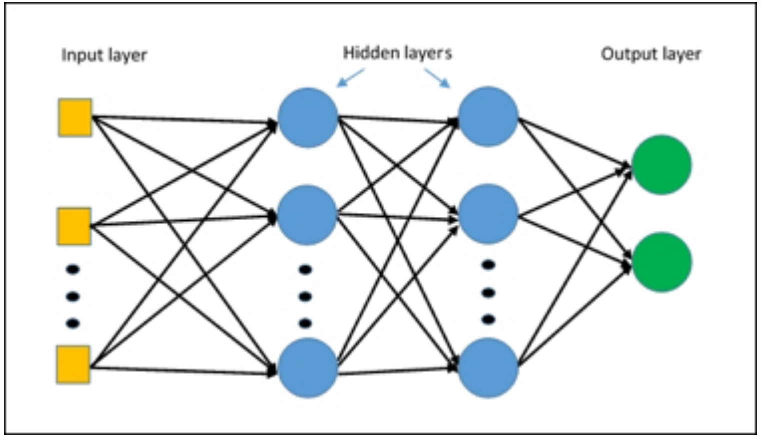
Three waves of ANN investigations

Everyone ”knows” ANNswith Deep Learning is Great

- Deep Learning = many layers
  - Weights
  - Connections

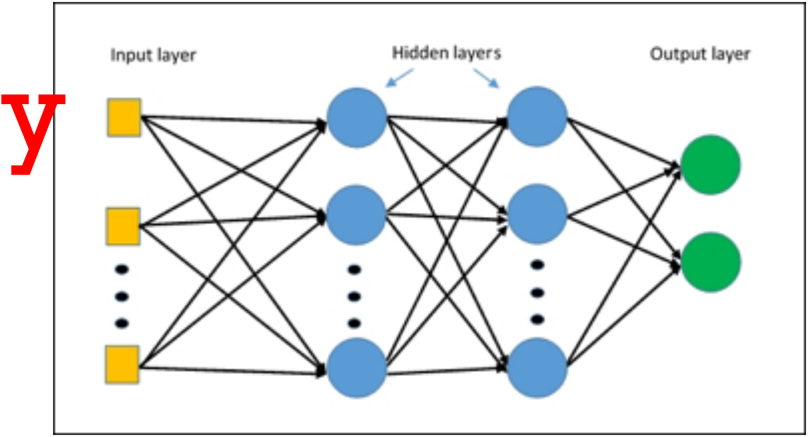
- Some Astounding Acccomplishments

Reported on in Press…



Everyone ”knows” ANNswith Deep Learning is Great

So
What Could Possibly
Go Wrong?



Three Waves of Neural Network Computation

- Wave 1
  - 1950s and 1960s
  - Single Layer Networks called Perceptrons
  - Ended When Minsky & Papert showed Single Layer Networks couldn’t do XOR
- Wave 2
  - 1980s and 1990s
  - Associative Networks, CNNs, Multi-Layer Perceptrons (2 hidden layers max)
  - Started with Hopfield and Rumelhart and McClellan
  - Faded away and primarily replaced with SVM and Bayesian techniques
- Wave 3 is Now

Basic Drivers of Three Waves

- High Performance Processors
- New Architectures (e.g. GPUs)
- Better Memory Management (e.g. Caching)
- Large Data Sets (Internet, Sensors, Hard Drives)

Problem

- ANNs Learn Complicated Functions in High-Dimensional Space
  - Dimension = the Number of Weights in Network
  - Often Millions of Dimensions!
  - Hard to control

Problem (cont.)

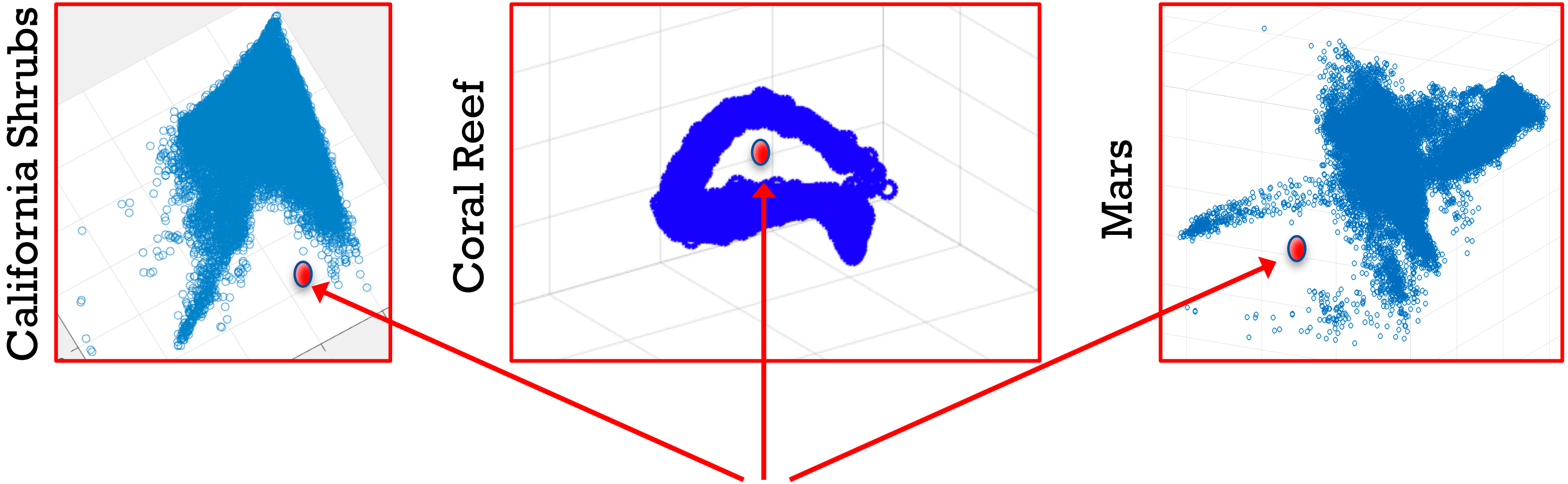
- Neural Networks are not Robust
  - Large changes to inputs can produce no changes in outputs
  - Small changes to inputs can produce large changes in outputs

We Want Competency Aware ANNs

- **Competency Aware ANN** can identify when
  - Input is **valid or invalid**
- Many ANNs can't identify invalid inputs
  - Why not just look at distance to training set?

We Want Competency Aware ANNs

Consider these 3D Projections of High-Dimensional Hyperspectral Data



California Shrubs

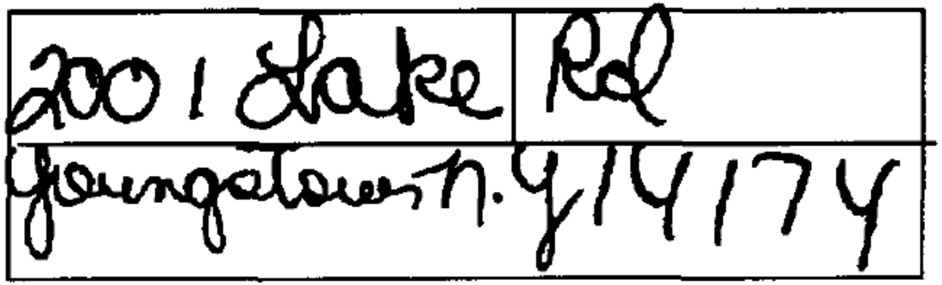
Coral Reef

Mars

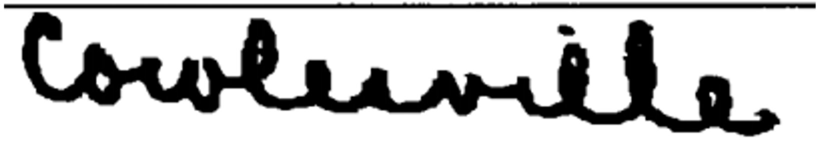
Are these points close?

Examples

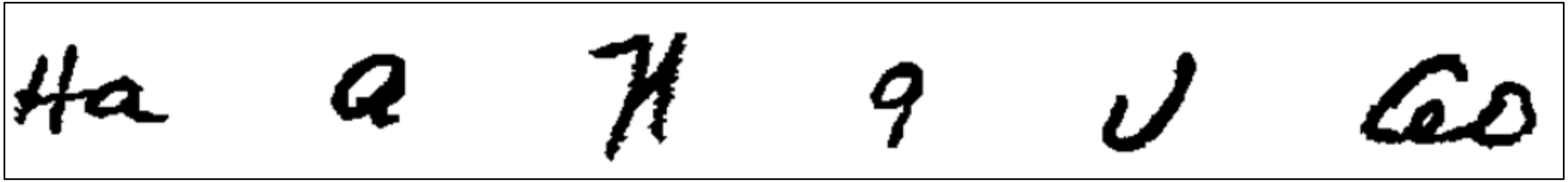
The Early Days
Handwritten Address Recognition
~ 1990-1993



2001 Lake Rd
Youngstown N.Y. 14174

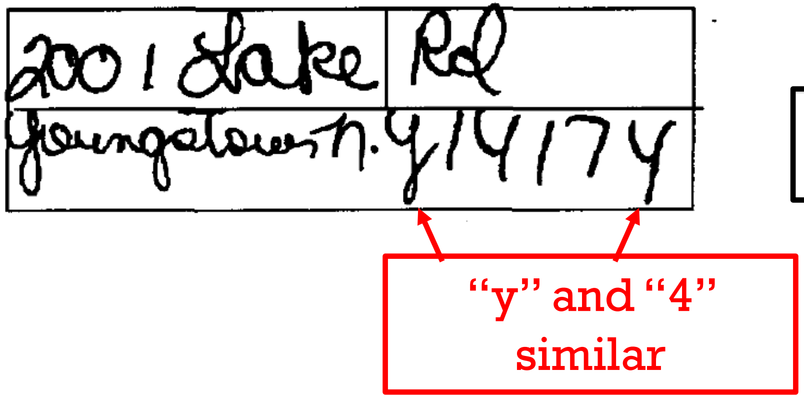


Cowlesville



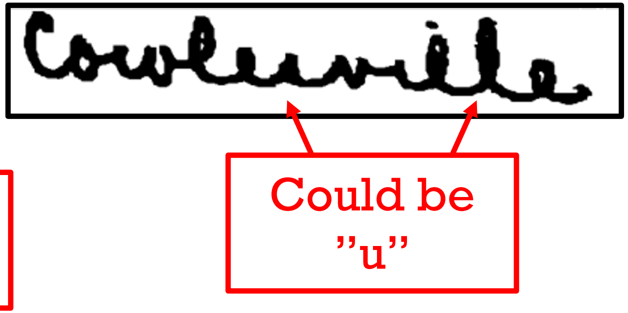
The Early Days Handwritten Address Recognition ~ 1990-1995

Handwritten Characters are Ambiguous. Difficult to determine number of characters in an image



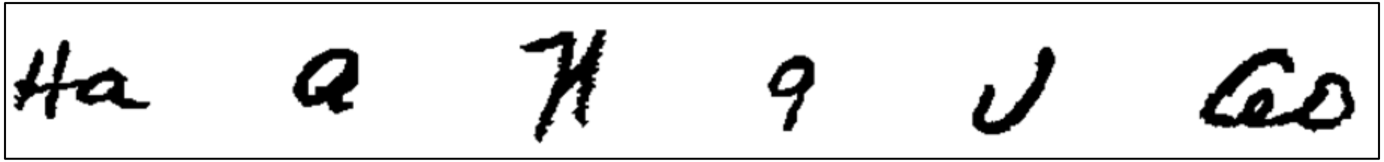
"y" and "4" similar

"Cowlesville"

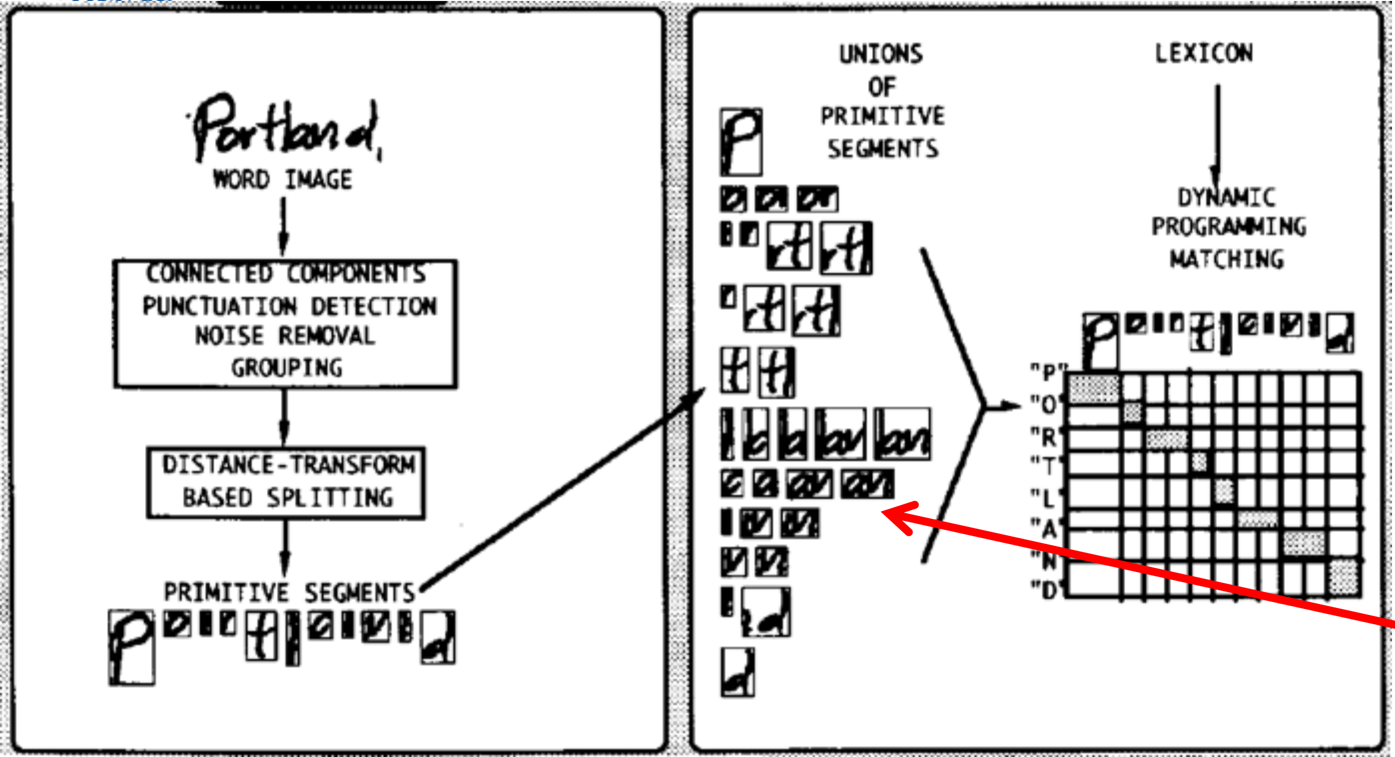


Could be "u"

What are these?



The Early Days
Handwritten Word Recognition
~ 1990-1995



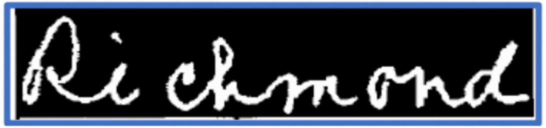
Oversegment & Match Word Images to Strings

Find best combination of image segments to match string

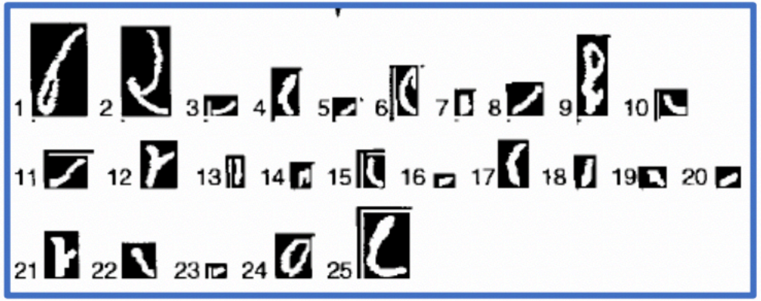
Many Non-Characters = Invalid Inputs

Another

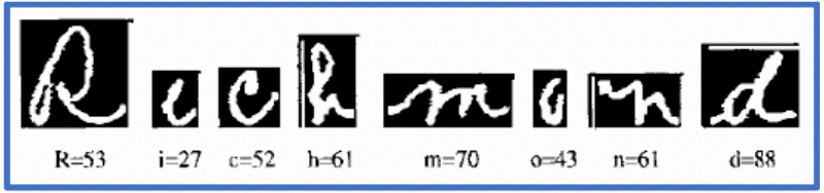
Look



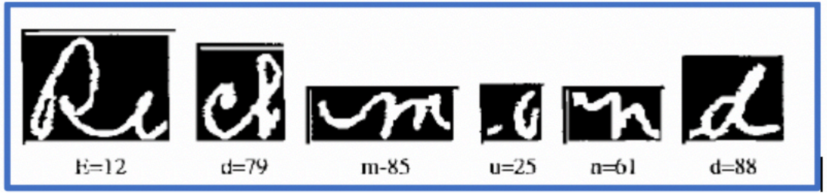
Primitive Segments



Best Unions of Primitives to Match “Richmond”



Best Unions of Primitives to Match “Richmond”



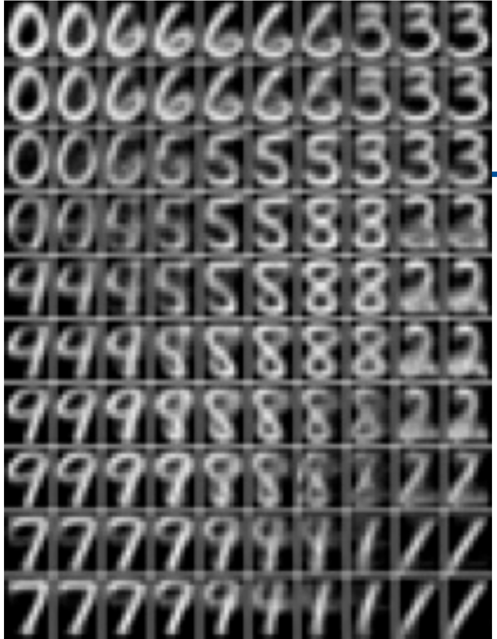
The Early Days

Handwritten Word Recognition

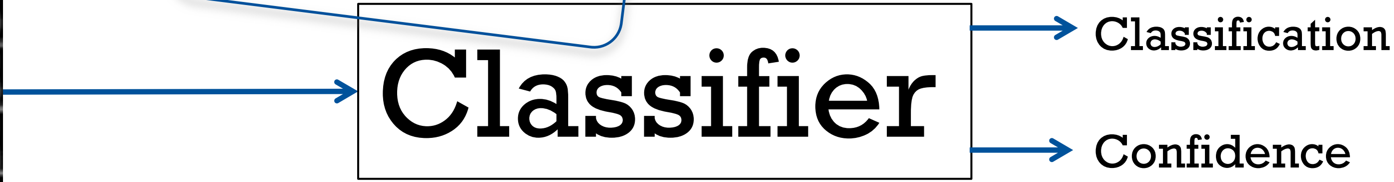
~ 1990-1993

Self-Organizing Map

(SOM)



Do you really look like them?



Better for Word Recognition
Worse for Character Recognition

Progression to LM / IED

Fielded in 2008

TV: “Bomb Hunters: Afghanistan”

From the US Army on 19 Aug 2008: "Bottom line: The team has accomplished more than was expected. Training is complete and the unit feels ready to conduct operations. From soldier to commanding general, everyone that has seen the system is confident that it is much better than anything they currently have and they want it. NVESD and NIITEK can be very proud of the outstanding effort of the NIITEK team."

Paul, I consider your group as part of the NIITEK team. On 28 Aug 2008, I received a call from Afghanistan. The call stated that the soldiers are elated because the system found exactly what it had set out to find and it found these multiple threats before anyone hit them.

Extensive Algorithm Testing

SOM-based Method Best

and

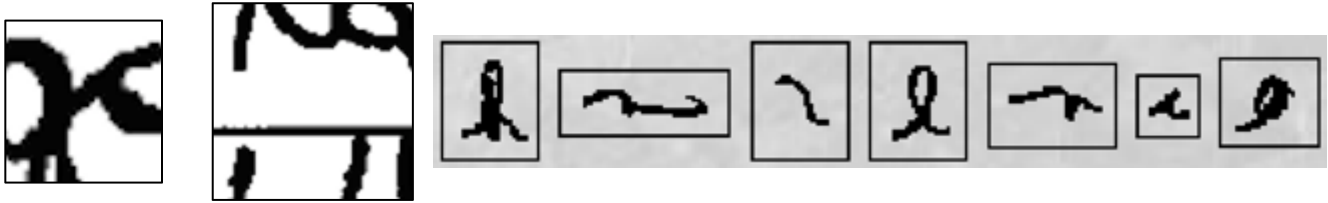
was first to be fielded

(by my former post-doc and colleague Hichem Frigui)

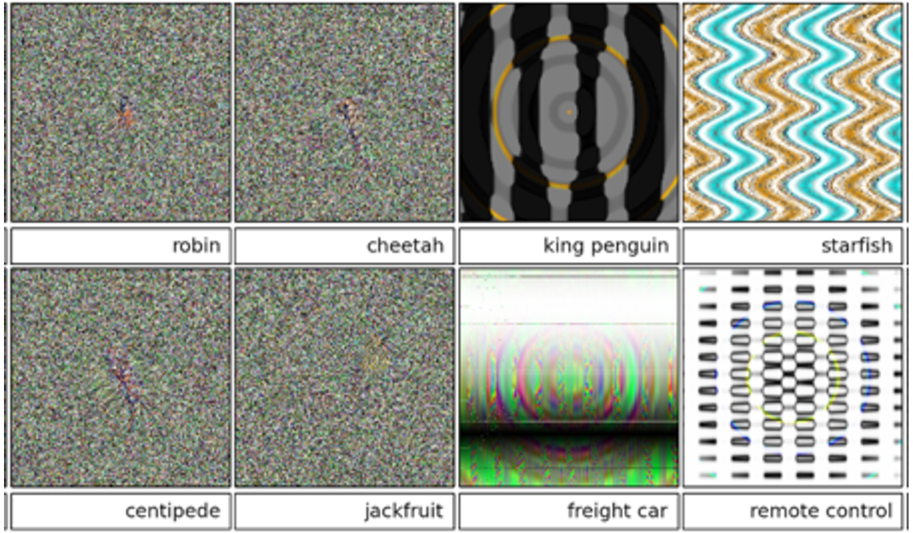
Invalid Input Examples

Neural Networks can make confident classification decisions on “Garbage” Inputs

- Natural



- Concocted



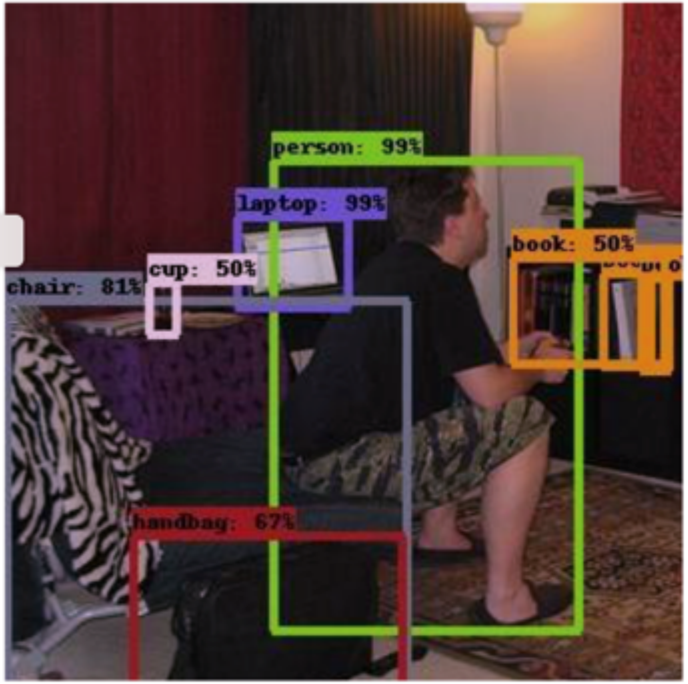
The Elephant in the Room

Amir Rosenfeld1, Richard Zemel2, and John K. Tsotsos1

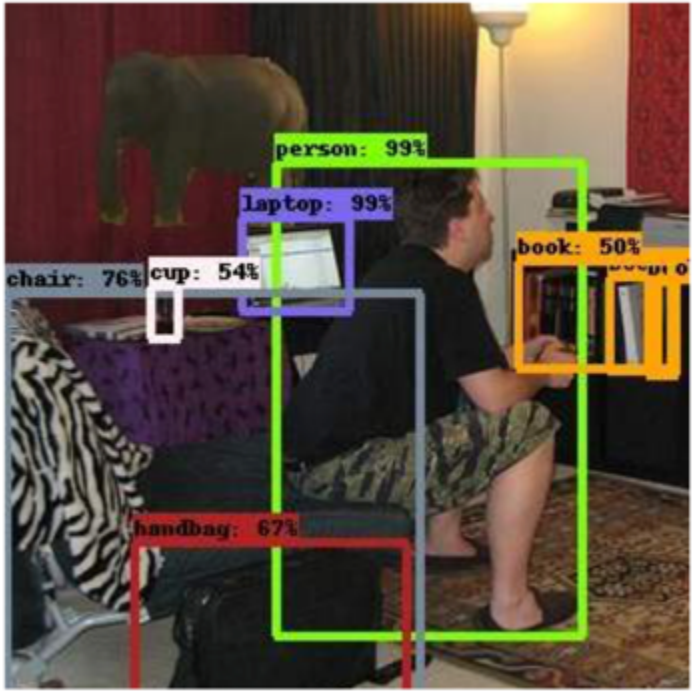
- A state-of-the-art object detector detects multiple images in a room
- An image of an elephant can be put in the room and cause many problems

Elephant in Room Examples

Original Image

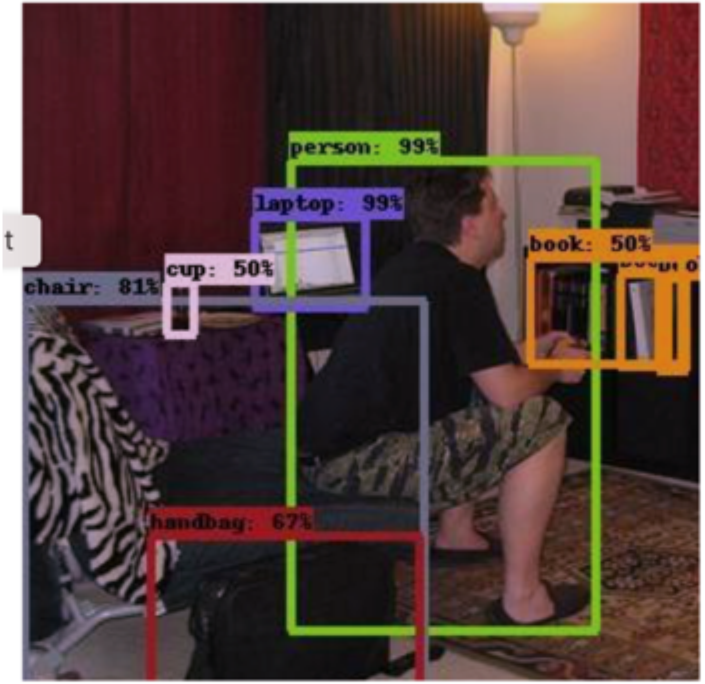


Elephant 1 – No Change

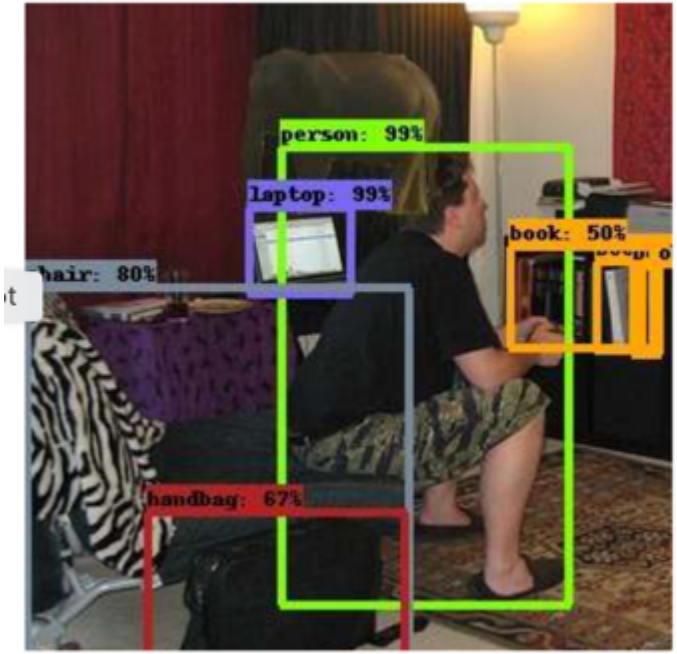


Elephant in Room Examples

Original Image

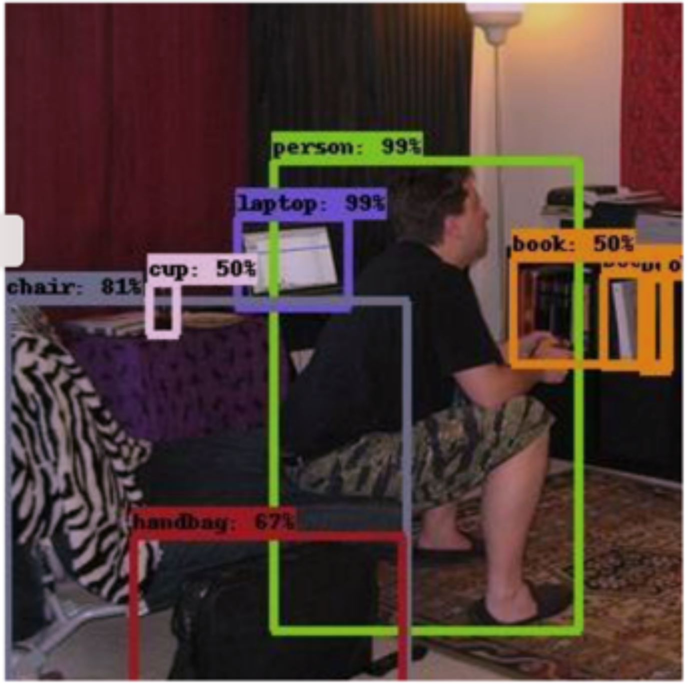


Elephant 2 – Cup Disappears

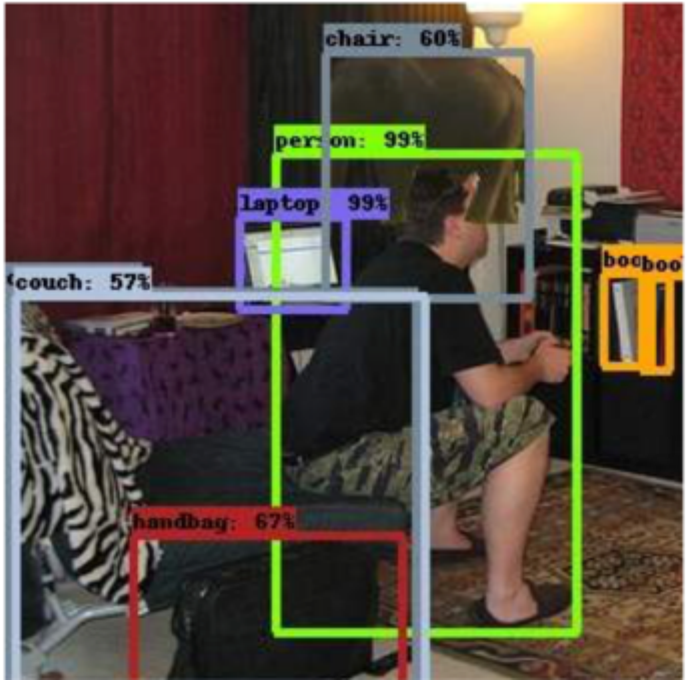


Elephant in Room Examples

Original Image

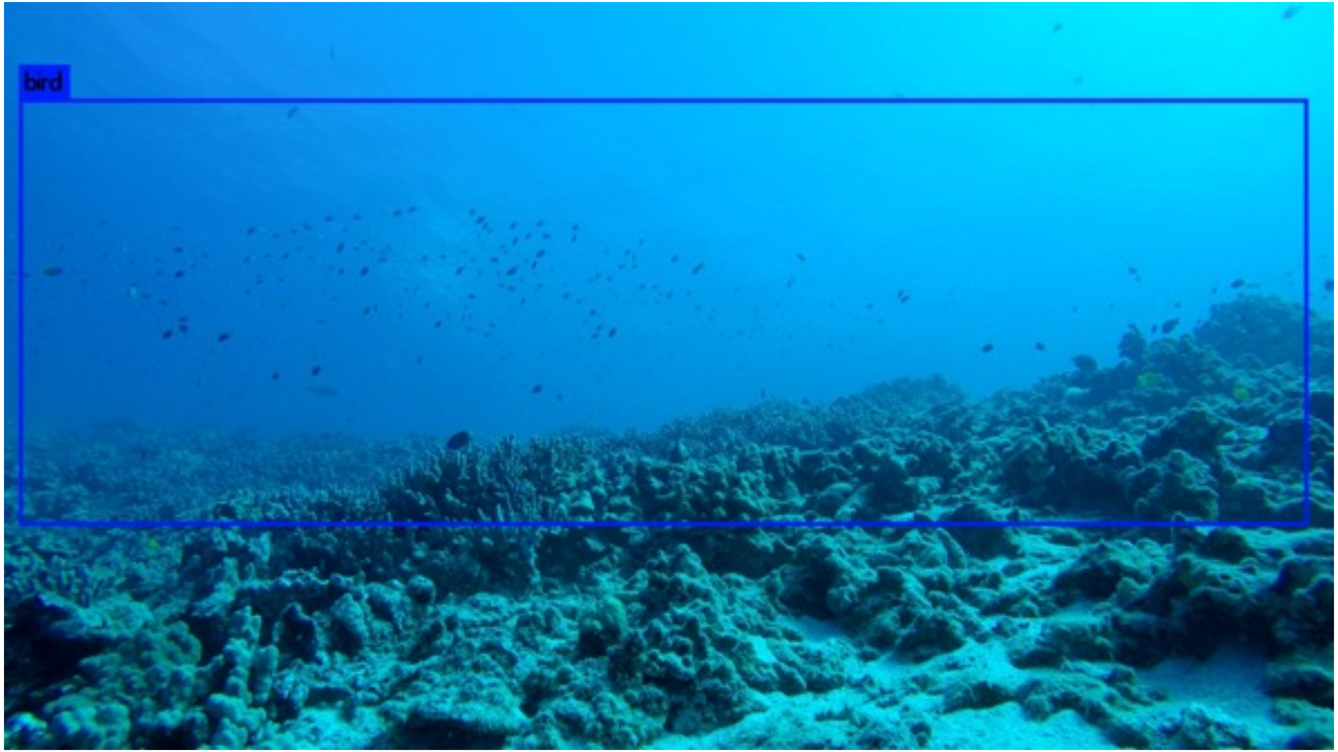


Elephant 3
Elephant becomes Chair
Chair becomes Couch
Cup Disappears



My PhD student Ron Fick ran
Yolo ANN (Well-known) on Underwater Video

Blue Box Called a Bird



So, what should be done?

- Fix it
  - Large Data Sets.
  - Train many networks: Ensembles.
  - Train with Outliers

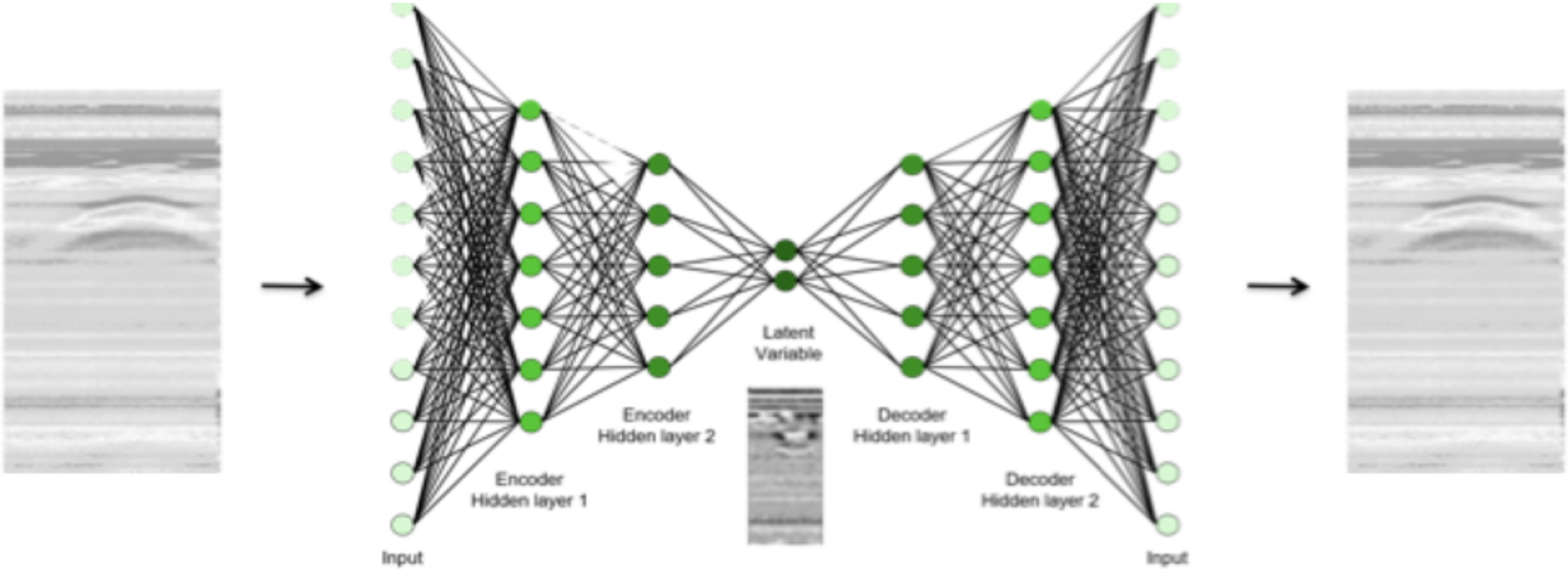
So, what should be done?

- Fix it
  - Large Data Sets.
  - Train many networks: Ensembles.
  - Train with Outliers

Auto-Encoders
Target Output = Input

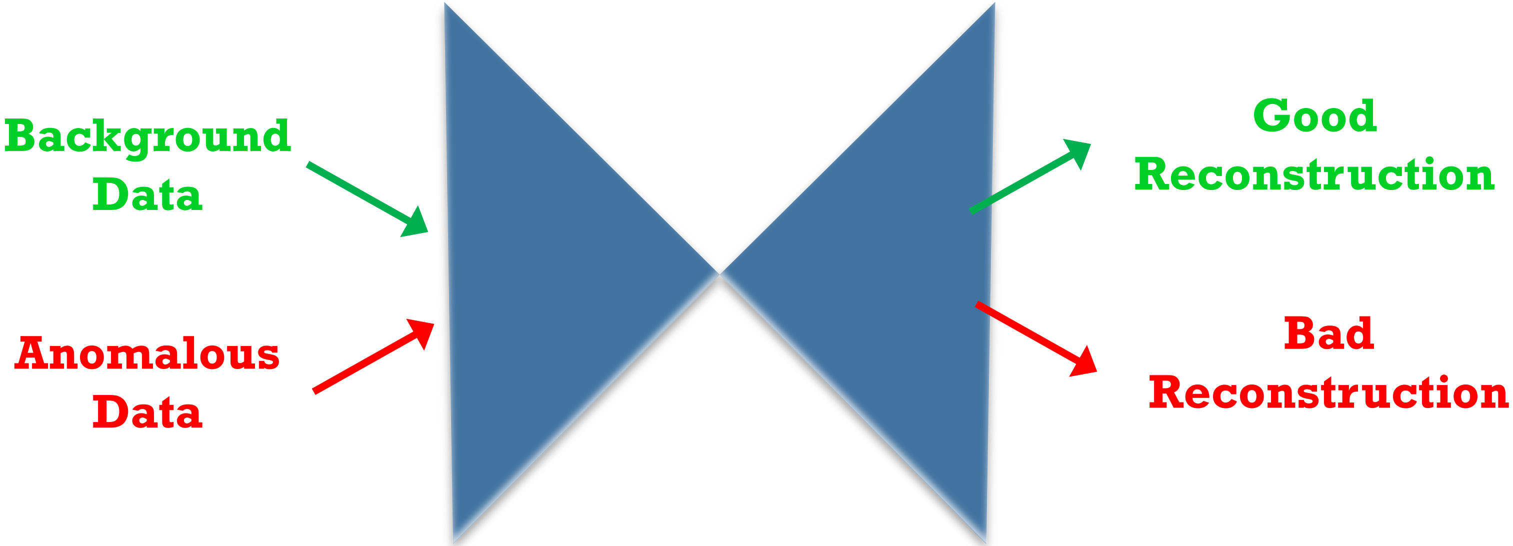
Reconstructs training data well but not other data (sometimes)

Good for Anomaly Detection



Auto-Encoders

Learns to reconstruct training data well but not other data (sometimes)



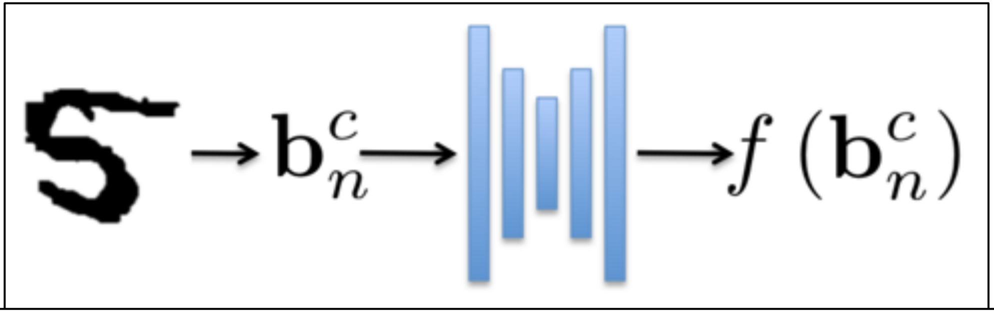
Background Data

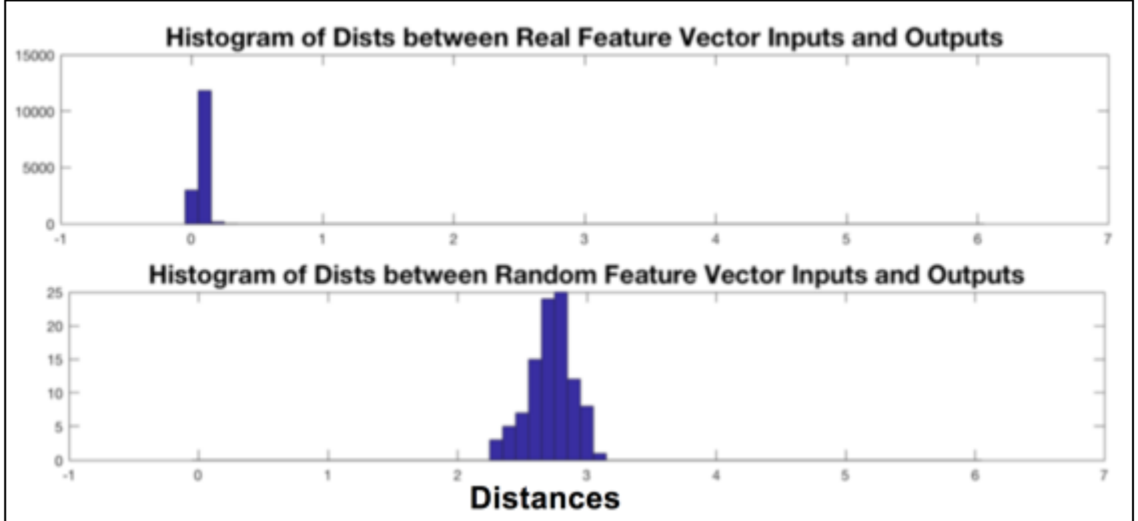
Anomalous Data

Good Reconstruction

Bad Reconstruction

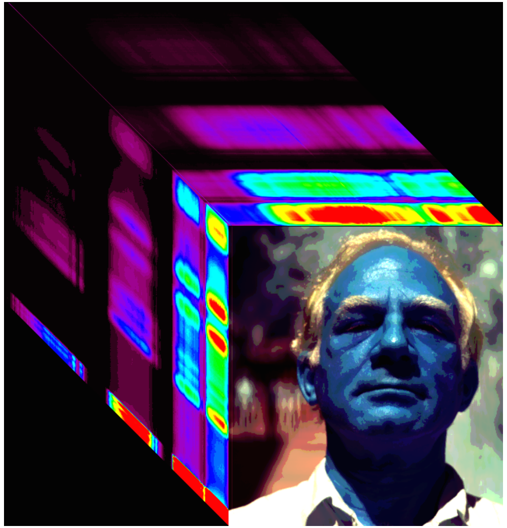
Auto-Encoders Experiment



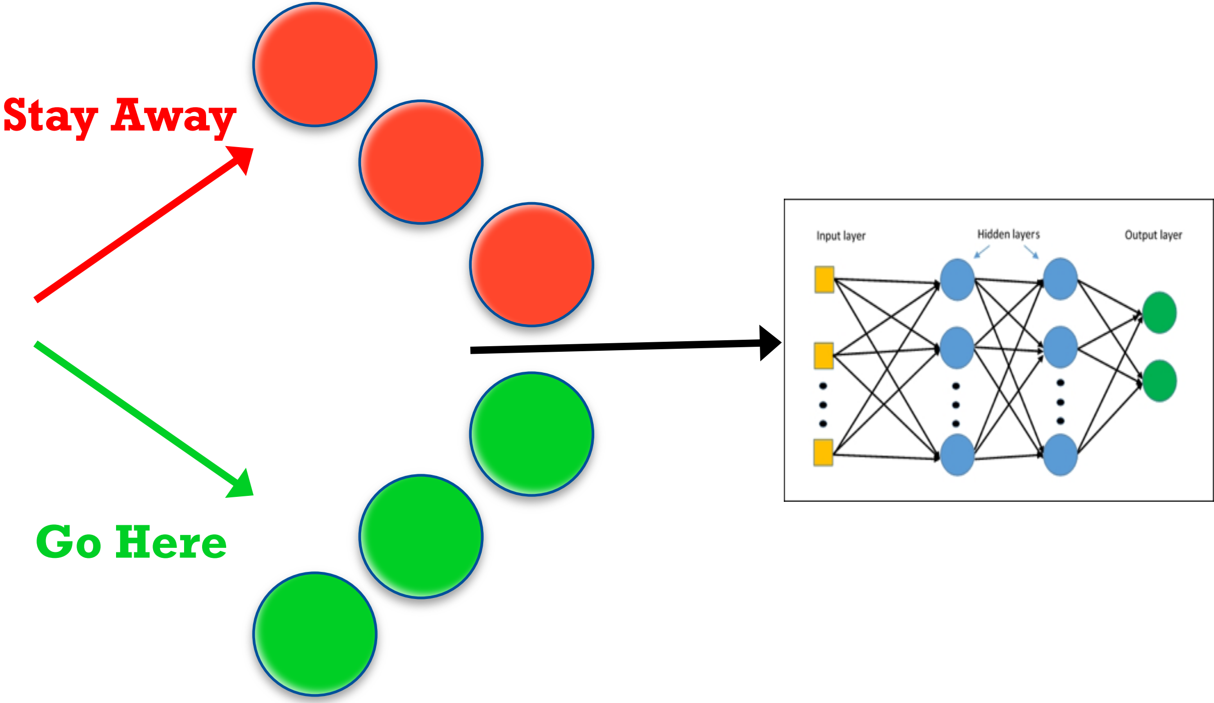


- Can construct Anomaly Detectors from background only data
- Learn to reconstruct back

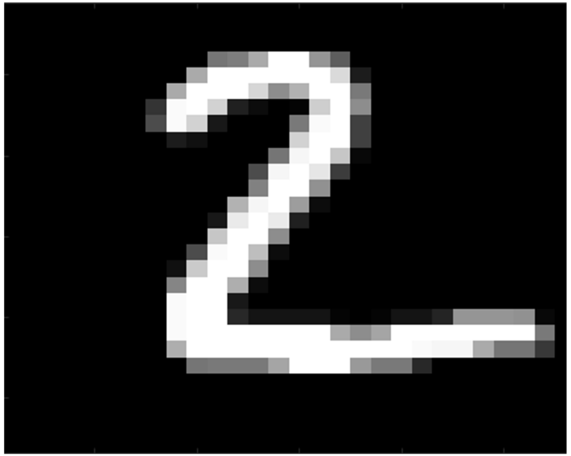
Keep away from bad places and go to good places

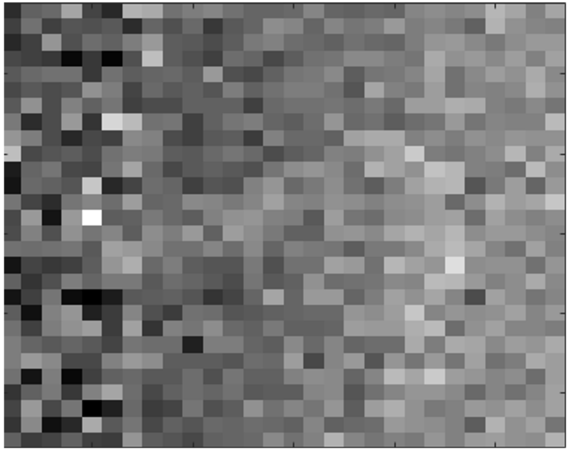


Input Data



NULL SPACE IMAGES

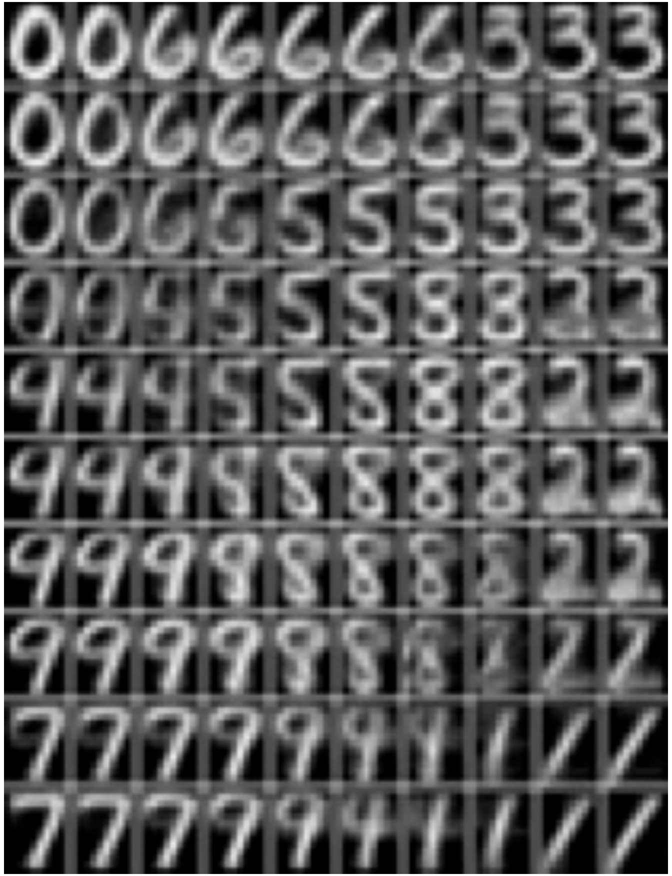




Original 2 and Original 2 + Null Space Image

Produce same output!!

Self Organizing Map and Manifolds



These patterns represent the Good Places.

If you are not close to one of them, then you are in a Bad Place

Summary

ANNs are not Robust

Have to be careful

There are several methods for mitigating this problem

Working on some hyperspectral now, results soon