NEURAL NETWORK MAGIC

Training data

Random network architecture

Learning algorithm

Excellent results
THE REAL MAGIC OF DEEP LEARNING

Let’s us solve problems we don’t know how to program
NEURAL NETWORKS RESEARCH AT BELL LABS, HOLMDEL 1985 - 1995
BELL LABS BUILDING IN HOLMDEL
Now called Bell Works

Around 1990:
6,000 employees in Holmdel
~300 in Research
~30 in machine learning, including:

Larry Jackel, Yann LeCun, Leon Bottou, John Denker, Vladimir Vapnik, Yoshua Bengio, Hans Peter Graf, Patrice Simard, Corinna Cortes, and many others
ORIGINAL DATABASE ~300 DIGITS

Demonstrated need for large database for benchmarking

→ USPS database

→ MNIST database (60,000 digits)
USING PRIOR KNOWLEDGE

What class is point “X”?

If representation is north, south, east, west, then choose green

If better representation is elevation above sea level choose red

Using prior knowledge beats requiring tons of training data

Example: use convolutions for OCR
CONVOLUTIONAL NEURAL NETWORK (CNN)
HOW WE CAME TO VIEW LEARNING

Largely Vladimir Vapnik’s influence

Choose the right structure – “Structural Risk Minimization”

→ Bring prior knowledge to the learning machine

Capacity control – matching learning machine complexity to available data

→ Examine learning curves
**LEARNING CURVES**

If test error $>>$ training error: get more training examples or decrease capacity

If test error $\sim$ training error: increase capacity

Smart structure allows low errors with lower capacity

Cortes, Jackel, Wan-Ping Chiang: Limits on Learning Machine Accuracy Imposed by Data Quality
PILOTNET LEARNING CURVE

MSE

Number of images (patches) used in training

Test set error

Train set error
THE (FRIENDLY) RECOGNIZER WAR

Who can read the digits best?

Comparison of Classifier Methods: A Case Study in Handwritten Digit Recognition

Léon Bottou*, Corinna Cortes, John S. Denker, Harris Drucker, Isabelle Guyon, L. D. Jackel, Yann LeCun, Urs A. Müller†, Eduard Säckinger, Patrice Simard, and Vladimir Vapnik

AT&T Bell Laboratories, Holmdel, NJ 07733
*Neuristique, 24 rue des Petites Ecuries, 75010 Paris, France
†Electronics Laboratory, Swiss Federal Institute of Technology, ETH Zentrum, CH-8092 Zürich, Switzerland

Year 1994

http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=576879
Larry Jackel: “In hindsight, the similar test errors are not so surprising since all the researchers were using variants of the same learning principles (everyone talked to Vladimir and studied the learning curves).”

The Optimal Margin (SVM) was remarkable because, unlike the other top performers, it did not include knowledge about the geometry of the problem.
Classification error on test set

Inference time in sec on a Sparc 10

Training time in days on a Sparc 10

Year 1994
LESSONS LEARNED

Look at the data during all processing steps
Solid debugging tools are critical
Validate the training data
The work is experimental in nature
Work with real data
EPILOG

1995 Deployed Holmdel neural nets at Wachovia Bank – Eventually our technology was processing 20% of the checks written in the US

1995 AT&T Bell Labs fractured

2002 AT&T Labs - mass layoffs

2003 Bell Labs work led to new programs at DARPA – LAGR, Learning Locomotion, Deep Learning, Challenges

2012 Deep Learning becomes popular triggered by availability of data, compute power and ready commercial applications
LEARNED AUTONOMY
DRIVE-AV
A full self-driving car stack

DRIVE AV

DRIVEWORKS SDK

DRIVE OS

DRIVE PX – AI CAR COMPUTER
NEW JERSEY TEAM GOAL

Solve the hard and unsolved autonomous vehicles (AV) problems using learning

To provide enhanced safety through algorithmic diversity

To create functionality that may not be achievable otherwise, such as perception-based turns or merges onto a busy highway

Additional AV labs in California, Boulder, Seattle, and Europe
An ideal location for autonomous driving R&D

Private roads for easy testing

Quick access to diverse public roads

Space to store and work on the cars
CONVENIENT LOCATION

Close to New York City (45 miles)

Accessible by commuter train, ferry, bus, and car

Family friendly, yet near a major metropolitan area

Close to popular beaches along the Jersey Shore
WHERE DO RULES FAIL?

Hand crafted features and rules work well when the problem domain is well understood.

Simple rules, such as “follow the lane lines” fail quickly

→ Can you list every feature you look at while driving?

It is extremely hard to write down every single rule required for safe driving – it is relatively easy to collect lots of training data (just drive)
LEARNING TO PREDICT THE PATH

Record data from lots of humans driving their cars:
- Sensor data
- Human driven path

Training data

Sensor data

CNN

CNN predicted path

Error (training) signal

Human driven path
EXAMPLE ARCHITECTURE

Single frame input with 200x66 pixel resolution

Input is from a single camera with a small patch cropped and scaled from the full image

~250,000 neuron weights, not a huge network

The network grows as we use more sensors and handle harder tasks
Meet NVIDIA BB8

https://www.youtube.com/watch?v=-96BEoXJMs0&t=39s
DRIVING WITH LEARNED PATH PREDICTION
LEARNED LANE CHANGES
LEARNED TURNS
https://www.youtube.com/watch?v=Sm-NBdSzP6E
VISUALIZATION

Where does the network look?
Visualization
Where does the network look?
ATYPICAL VEHICLE CLASS
OPEN CHALLENGES

Deal with ambiguous situations
There is often more than one correct answer

Learn from imperfect behavior
Several observations in the same situation, most are correct, some are not correct
DRIVING IN THE SNOW
Thank You