MACHINE LEARNING
WHAT IS ML?

• A subfield of Artificial Intelligence
• On developing algorithms or systems for automated prediction and classification
• Examples:
  • Robotics, visual perception
  • Recommendation engine
  • Stock trading
  • Medical image analysis
  • Fraud: credit card, mortgage
  • and many others …
WHAT IS ML?

Image

[201, 185, 125, ..., 233]

Classification Task

- Classes:
  - 82% cat
  - 15% dog
  - 2% hat
  - 1% mug

Algorithms

Everyday operations

‘+’, ‘-’, ‘*’, ‘/’

from http://cs231n.github.io/classification/
WHAT IS ML?

What the computer sees

82% cat
15% dog
2% hat
1% mug

from http://cs231n.github.io/classification/
NEURAL NETWORKS

Image

[201, 185, 125, ..., 233]

Classes:
- 82% cat
- 15% dog
- 2% hat
- 1% mug

from http://neuralnetworksanddeeplearning.com/chap1.html
DEEP NEURAL NETWORKS

Image

[201, 185, 125, ..., 233]

Classes:
82% cat
15% dog
2% hat
1% mug

from http://neuralnetworksanddeeplearning.com/chap1.html
NN STUDY LINKS

- https://www.coursera.org/course/neuralnets
- http://cs231n.stanford.edu/
- http://karpathy.github.io/neuralnets/
DEEP LEARNING

With massive amounts of computational power, machines can now recognize objects and translate speech in real time. Artificial intelligence is finally getting smart.

Messages that quickly self-destruct could enhance the privacy of online communications and make people freer to be spontaneous.

Reading the DNA of fetuses will be the next frontier of the genomic revolution. But do you really want to know about the genetic problems or musical aptitude of your unborn child?

A maverick neuroscientist believes he has deciphered the code by which the brain works.

Doubling the efficiency of a solar cell would completely...
Discovering Structure

Dramatic increase in both computational power and the amount of data available from web, video cameras, laboratory measurements.

• Develop statistical models that can discover underlying structure, semantic relations, constraints, or invariances from data.

• Robust, adaptive models that can deal with missing measurements, nonstationary distributions, multimodal data.

• Deep Learning can also discover structure at multiple features levels: from low level pixels to high level semantic features.
Speech Recognition is Improved

According to Microsoft’s speech group:

Using DL

Word error rate on Switchboard

100%

10%

4%

2%

1%

1990  2000  2010
From their blog:

To put these algorithms to use, we had to work to overcome some limitations, for instance that they were built to handle 100 million ratings, instead of the more than 5 billion that we have, and that they were not built to adapt as members added more ratings. But once we overcame those challenges, we put the two algorithms into production, where they are still used as part of our recommendation engine.

(Salakhutdinov et. al. ICML, 2007, Salakhutdinov and Mnih, 2008)
Key Computational Challenges

Scaling up our deep learning algorithms:

- Learning from billions of (unlabeled) data points
- Developing new parallel algorithms
- Scaling up Computation using clusters of GPUs and FPGAs

Building bigger models using more data improves performance of deep learning algorithms!
Deep Learning in Action

- Achieves state-of-the-art on many object recognition tasks!
  Try it at deeplearning.cs.toronto.edu!
Deep Convolutional Nets

- Convolution
- Pooling
- Normalization
- Densely connected

Very deep network

High-level feature space

Prediction
Deep Convolutional Nets
Deep Convolutional Nets

Convolution

Pooling

Convolution
Deep Convolutional Nets

Pooling

Convolution
Deep Convolutional Nets

Convolution

Pooling

Convolution

Additional Sensor Modalities: Depth, Laser Scans, Optical Flow (for video)
Multimodal Deep Convolutional Nets

Pooling

Convolution
Visual Search

Query Image

0.39 tiger shark
0.26 killer whale
0.10 snorkel
0.08 scuba diver
0.08 coral reef
Visual Search

Query Image

0.93 agaric
0.07 mushroom
Learning Structured Outputs

- More challenging problem.
- How can we generate complete descriptions of images?

**Input**

![A man skiing down the snow covered mountain with a dark sky in the background.](image)

**Output**

A man skiing down the snow covered mountain with a dark sky in the background.
An Image-Text Encoder-Decoder

Feature space

Encoder: convolutional neural net

Decoder: neural language model

Steam ship in the water

Use the image features to additively bias the prediction of the next word representation.

Multimodal Neural Language Models (Kiros, et.al., ICML 2014)
An Image-Text Encoder-Decoder

Joint Feature space

Encoder: ConvNet

Decoder: neural language model

- Learn a joint embedding space of images and text:
  - Can condition on anything (images, words, phrases, etc)
  - Natural definition of a scoring function (inner products in the joint space)
  - Use a new language model that incorporates additional structure
Tagging and Retrieval

mosque, tower, building, cathedral, dome, castle

kitchen, stove, oven, refrigerator, microwave

ski, skiing, skiers, skiiers, snowmobile

bowl, cup, soup, cups, coffee

beach

snow
Caption Generation

- A car is parked in the middle of nowhere.
- A ferry boat on a marina with a group of people.
- A little boy with a bunch of friends on the street.
- There is a cat sitting on a shelf.
- A plate with a fork and a piece of cake.
- The two birds are trying to be seen in the water.

(Kiros, Salakhutdinov, Zemel, NIPS 2014)
Understanding Images

cat, ferret, hamster, weasel, puppy
cute, furry, cuddly, adorable, naughty

cup, bowl, coffee, soup, cups
yummy, delicious, plastic, foamy, savory

(Kiros, Salakhutdinov, Zemel, NIPS 2014)
Checkout demo at

deeplearning.cs.toronto.edu
Computer Vision Features

SIFT

Spin image

HoG

RIFT

Textons

GLOH
Computer Vision Features

- SIFT
- Spin image
- HoG
- RIFT
- Textons
- GLOH
- Deep Learning
Audio Features

Spectrogram

MFCC

Flux

ZCR

Rolloff
Audio Features

- ZCR
- Spectrogram
- MFCC
- Rolloff
- Flux
- ZCR
- Rolloff

Deep Learning
Expression Recognition

7 classes: neural, happy, anger, sad, surprised, disgust, and fear

• Surpassed human judge performances by 2%
Expression Recognition

Expression Categories

More layers

Pooling

Convolution

Expression Categories

...
Facial Keypoints

Kaggle Competition:
predict points locations

https://www.kaggle.com/c/facial-keypoints-detection/
Illumination Variations

Multilinear

One-shot

Latent codes: **identity**

Latent codes: **lighting**
Introduce a probabilistic framework that uses visual attention to learn generative models of objects of interest.
Deep Generative Face Model

• Gaussian Deep Belief Net is a type of deep generative model

• Hybrid undirected-directed graphical model

• Good generative model of objects and faces
Human Activity Analytics

optical flow

Deep Learning

Robbery

background
Human Activity Analytics

optical flow

Deep Learning

background

Theft
Brest Cancer - mitotic cells

(a) Scanner A           (b) Scanner H
CSC2431 report: Mitosis Detection in Breast Cancer Histological Images

Input

128   128   3
64
2
1

Convolve

Maxpool

Output

Input Image

128   128
64
2
1

Detection Algorithm

Input Image

Output Probability Map

Fig. 2. Convolutional net applied to a histological image.

1. More likely that the model will detect the mitosis at some scale which it would have otherwise missed at the canonical scale.

2. Better loss functions - The convolutional net in [3] used a cross entropy loss. We instead use a max-margin based loss function such as the hinge loss. This amounts to using a linear kernel SVM. We found that this loss functions further helps in reducing the error.

3. More context. Instead of 101×101 patches, we used bigger input patches 128×128. This provides more contextual information to the convolutional net.

3 EXPERIMENTAL RESULTS

3.1 Datasets and Evaluation

Recently, there were 3 contest or challenges related to mitosis detection [1]. However, for ease of comparison against a plethora of existing results, we will be using the dataset from the 2012 competition: http://ipal.cnrs.fr/ICPR2012.

The competition contains 5 biopsy slides imaged by 3 different hardware scanners. Each slide is scanned into 10 high-power fields which cover 512×512 µm² area. We will focus on the images from one scanner, an Aperio XT scanner. This scanner has resolution of 0.2456 microns, resulting in 50 images of size 2084 pixels by 2084 pixels. The 50 images contain around 300 mitosis. We will split the 50 images into 35 training and 15 for evaluation, the same protocol as the competition.

For every image in the dataset there is a corresponding csv files where each row contains a set of (x,y) point pairs indicating which pixel of the image is a mitotic cell. The number of rows in the csv files are the number of labeled mitotic cells of the corresponding image. We use these labels to give us positive training samples and all of the rest of the patches can be considered as negative training samples.

For evaluation, we evaluate the performance of our detector in the same way as the other contestants of the challenge. The test set consists of 15 images and was not part of the training process. A detection is deemed as correct if it is within 5 µm or 20 pixels of the ground truth mitosis. Aggregating over all 15 images, we counted the number of true positives (TP), false negatives (FN), and false positives (FP). Based on these statistics, the three important performance measures are recall $r = \frac{TP}{TP + FN}$, precision $p = \frac{TP}{TP + FP}$, and F-measure $f = \frac{2rp}{r + p}$.

3.2 Model Training

The general approach of our solution is to train a 2 way classifier on RGB images of size 128 by 128 pixels. The 2 classes are mitotic and non-mitotic. From the csv label files of our 35 training images, we obtained 67K positive training samples. These samples are 128x128x3 image patches which are labeled by experts as having a mitotic cell in the center of the patch. For the negative training samples...

Detection Algorithm

Input

128   128   3
64
2
1

Convolve

Maxpool

Output

Input Image

128   128
64
2
1

2048

Output Probability Map
**Algorithm**

**Fig. 3.** The detection process. **Left:** Convolutional Net outputs a raw probability map. **Middle:** Gaussian low-pass smoothing prepares the map for non-maximal suppression. **Right:** Non-maximal suppression suppress the pixel of the map which are not local maximum. Zoom in to see the "white dots".

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>This paper</td>
<td>0.75</td>
<td>0.85</td>
<td>0.800</td>
</tr>
<tr>
<td>DNN [3]</td>
<td>0.88</td>
<td>0.70</td>
<td>0.782</td>
</tr>
<tr>
<td>IPAL [11]</td>
<td>0.69</td>
<td>0.74</td>
<td>0.718</td>
</tr>
<tr>
<td>SUTECH</td>
<td>0.70</td>
<td>0.72</td>
<td>0.709</td>
</tr>
<tr>
<td>NEC [15]</td>
<td>0.74</td>
<td>0.59</td>
<td>0.659</td>
</tr>
<tr>
<td>UTRECHT [22]</td>
<td>0.51</td>
<td>0.68</td>
<td>0.583</td>
</tr>
<tr>
<td>WARWICK [13]</td>
<td>0.46</td>
<td>0.57</td>
<td>0.513</td>
</tr>
<tr>
<td>NUS</td>
<td>0.63</td>
<td>0.40</td>
<td>0.490</td>
</tr>
<tr>
<td>ISIK [21]</td>
<td>0.28</td>
<td>0.68</td>
<td>0.397</td>
</tr>
<tr>
<td>ETH-HEILDERBERG [20]</td>
<td>0.14</td>
<td>0.80</td>
<td>0.374</td>
</tr>
<tr>
<td>OKAN-IRISA-LIAMA</td>
<td>0.78</td>
<td>0.22</td>
<td>0.343</td>
</tr>
<tr>
<td>IITG</td>
<td>0.17</td>
<td>0.46</td>
<td>0.255</td>
</tr>
<tr>
<td>DREXEL</td>
<td>0.14</td>
<td>0.21</td>
<td>0.172</td>
</tr>
</tbody>
</table>

Table 2. Detection performance on the test set images. Comparison with previously published results on the exact same test set. Our result is state-of-the-art.

Training and detections uses the model GeForce Titan graphics processing cards from nVidia. They have 6GB of global memory and up to 3,000 GFLOPS/sec for single precision matrix multiplications. This allows training to be completed in as few as 6 hours. This is very efficient considering that our model contains 20 million tunable parameters.
Results

- Learning performs well even with a small dataset
- 50 images with 300 mitosis cells
- Better than untrained human
- State-of-the-art across all literature

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>This paper</td>
<td>0.75</td>
<td>0.85</td>
<td><strong>0.800</strong></td>
</tr>
<tr>
<td>DNN [3]</td>
<td>0.88</td>
<td>0.70</td>
<td>0.782</td>
</tr>
<tr>
<td>IPAL [11]</td>
<td>0.69</td>
<td>0.74</td>
<td>0.718</td>
</tr>
<tr>
<td>SUTECHE [7]</td>
<td>0.70</td>
<td>0.72</td>
<td>0.709</td>
</tr>
<tr>
<td>NEC [15]</td>
<td>0.74</td>
<td>0.59</td>
<td>0.659</td>
</tr>
<tr>
<td>UTRECHT [22]</td>
<td>0.51</td>
<td>0.68</td>
<td>0.583</td>
</tr>
<tr>
<td>WARWICK [13]</td>
<td>0.46</td>
<td>0.57</td>
<td>0.513</td>
</tr>
<tr>
<td>NUS</td>
<td>0.63</td>
<td>0.40</td>
<td>0.490</td>
</tr>
<tr>
<td>ISIK [21]</td>
<td>0.28</td>
<td>0.68</td>
<td>0.397</td>
</tr>
<tr>
<td>ETH-HEILDERBERG [20]</td>
<td>0.14</td>
<td>0.80</td>
<td>0.374</td>
</tr>
<tr>
<td>OKAN-IRISA-LIAMA</td>
<td>0.78</td>
<td>0.22</td>
<td>0.343</td>
</tr>
<tr>
<td>IITG</td>
<td>0.17</td>
<td>0.46</td>
<td>0.255</td>
</tr>
<tr>
<td>DREXEL</td>
<td>0.14</td>
<td>0.21</td>
<td>0.172</td>
</tr>
</tbody>
</table>

Table 2. Detection performance on the test set images. Comparison with previously published results on the exact same test set. Our result is state-of-the-art.