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]

Algorithms

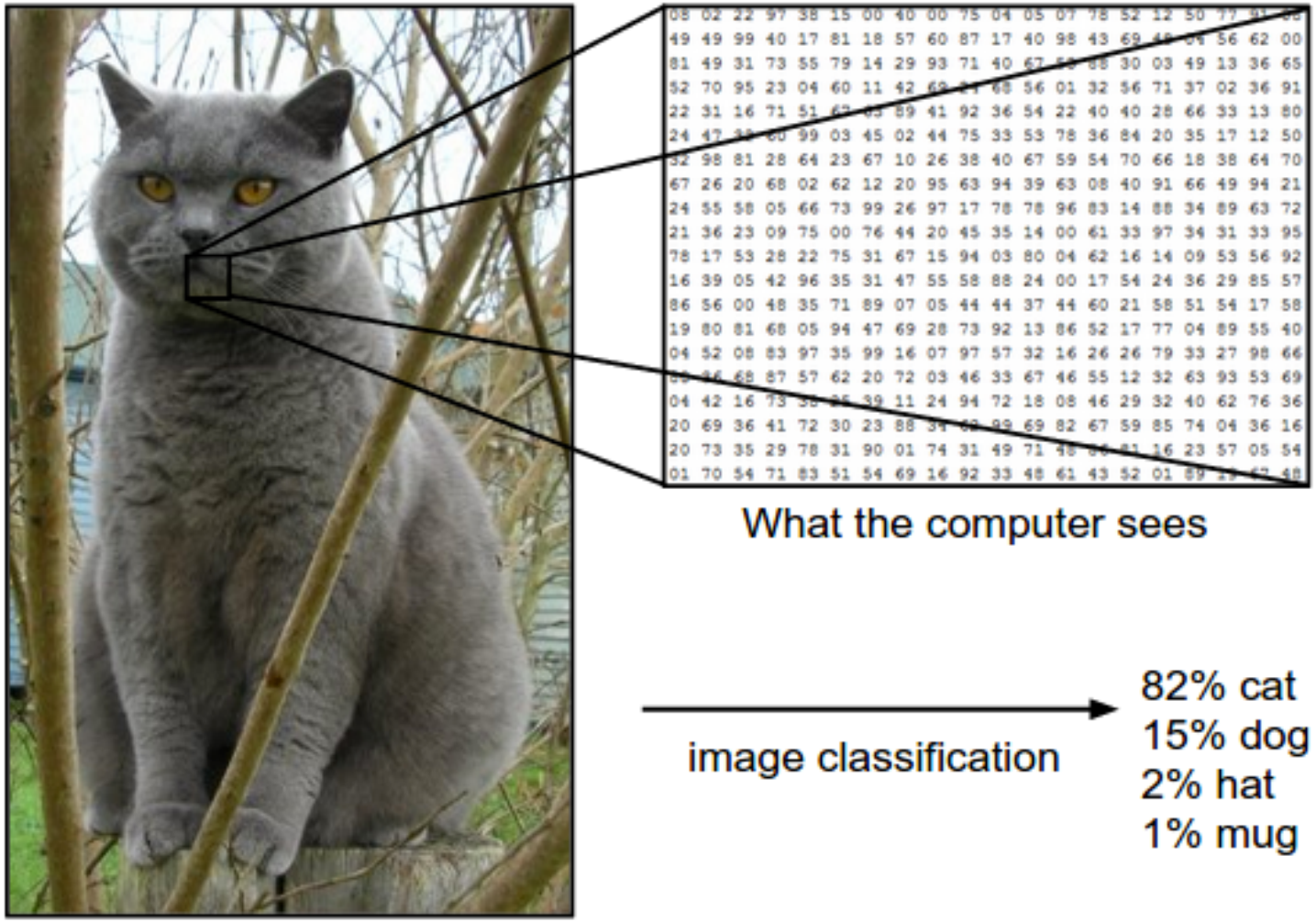
Everyday operations
'+', '-', '*', '/'

Classification Task

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

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

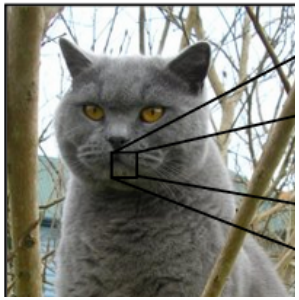
WHAT IS ML?



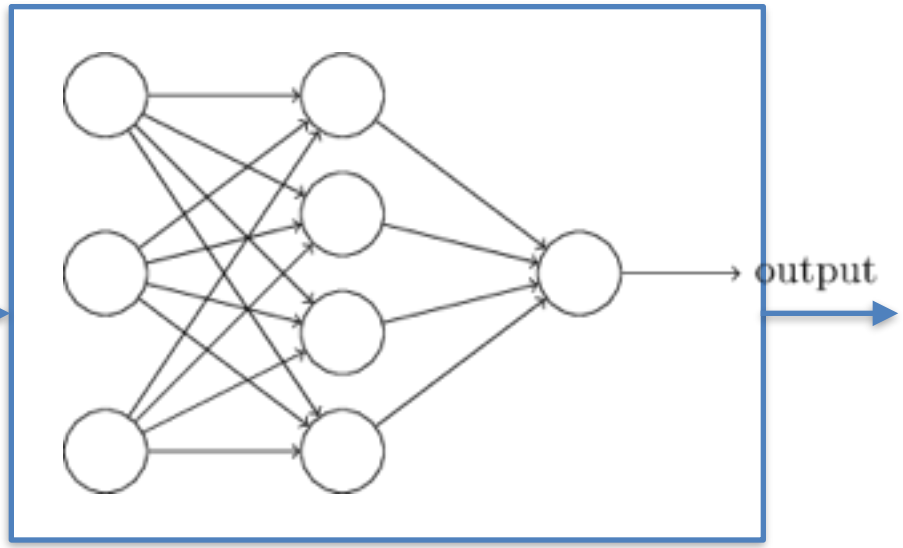
from http://cs231n.github.io/classification/

NEURAL NETWORKS

Image



[201, 185, 125, …, 233]

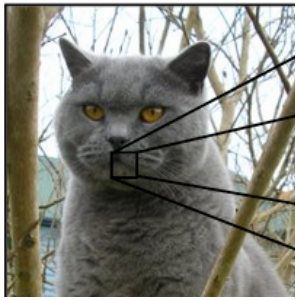


Classes:
82% cat
15% dog
2% hat
1% mug

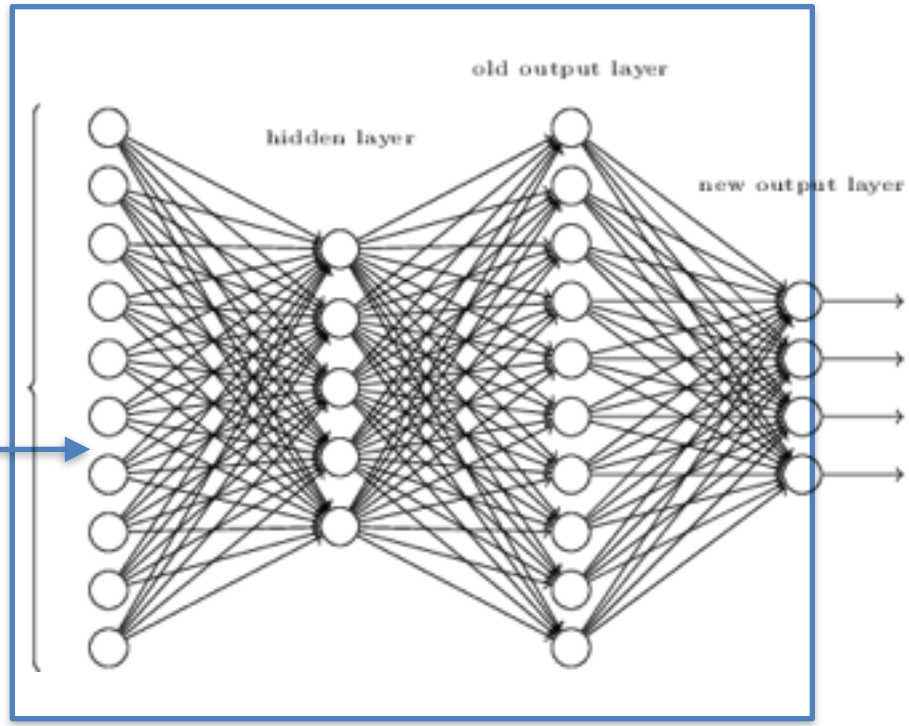
from http://neuralnetworksanddeeplearning.com/chap1.html

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/
- http://neuralnetworksanddeeplearning.com/chap1.html
- http://scikit-learn.org/stable/

DEEP LEARNING

MIT Technology Review

10 BREAKTHROUGH TECHNOLOGIES 2013

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.

Temporary Social Media

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

Prenatal DNA Sequencing

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?

Memory Implants

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

Smart Watches

Ultra-Efficient Solar Power

Doubling the efficiency of a solar cell would completely

Discovering Structure

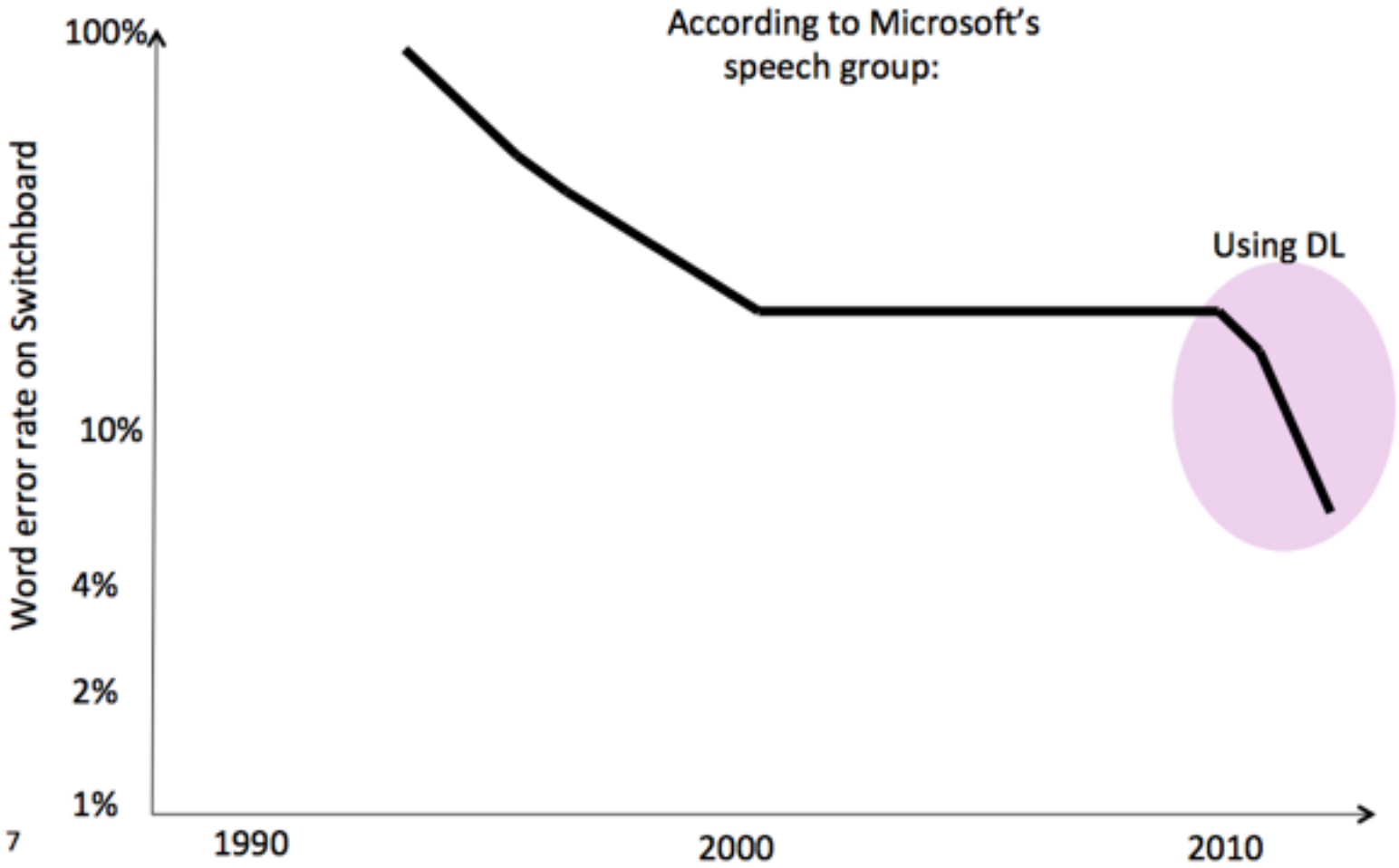
Dramatic increase in both computational power and the amount of data available from web, video cameras, laboratory measurements.

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

- Robust, adaptive models 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



NETFLIX The Netflix Tech Blog

Netflix uses:

- Restricted Boltzmann machines
- Probabilistic Matrix Factorization

- 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

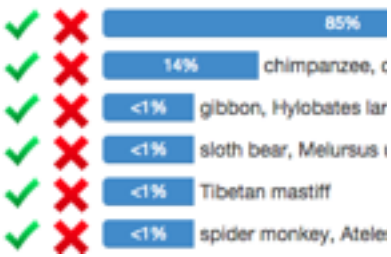
Building bigger models using more data improves performance of deep learning algorithms!

Deep Learning in Action

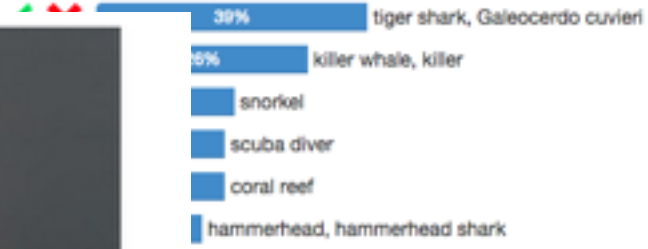
- Achieves state-of-the-art on many object recognition tasks!

Try it at deeplearning.cs.toronto.edu!

Possible tags:

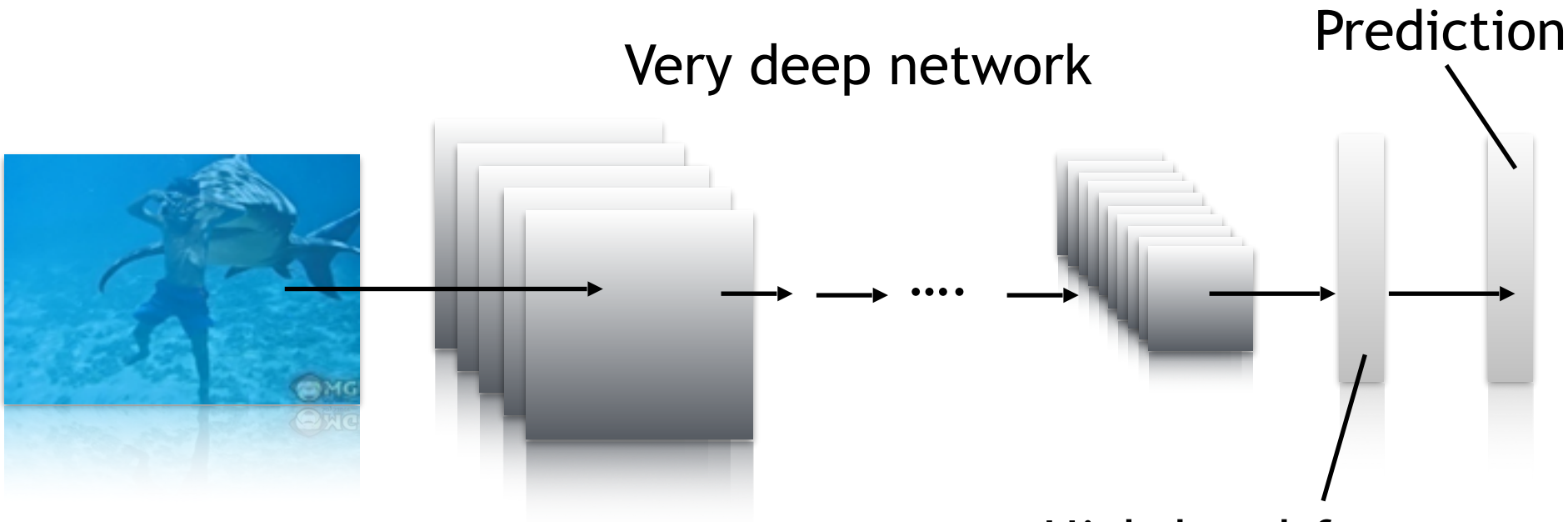


Possible tags:



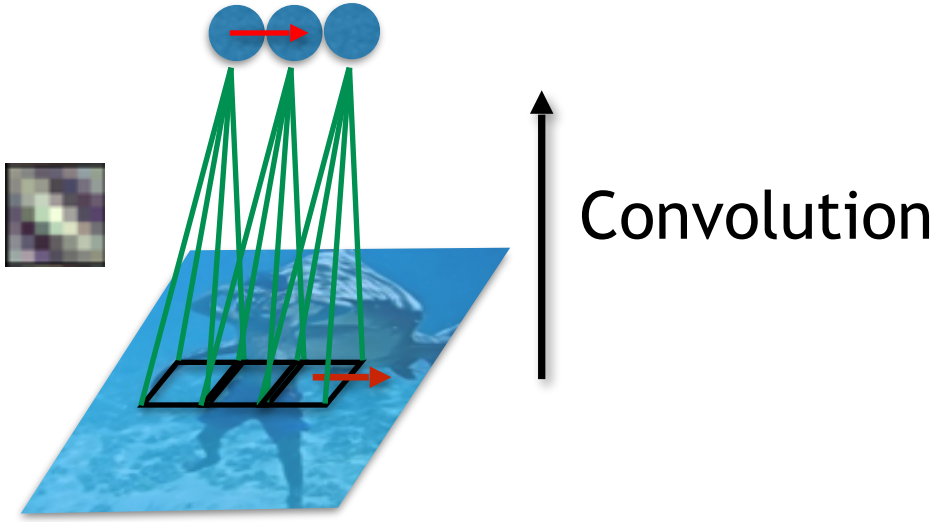
Possible tags:

Deep Convolutional Nets

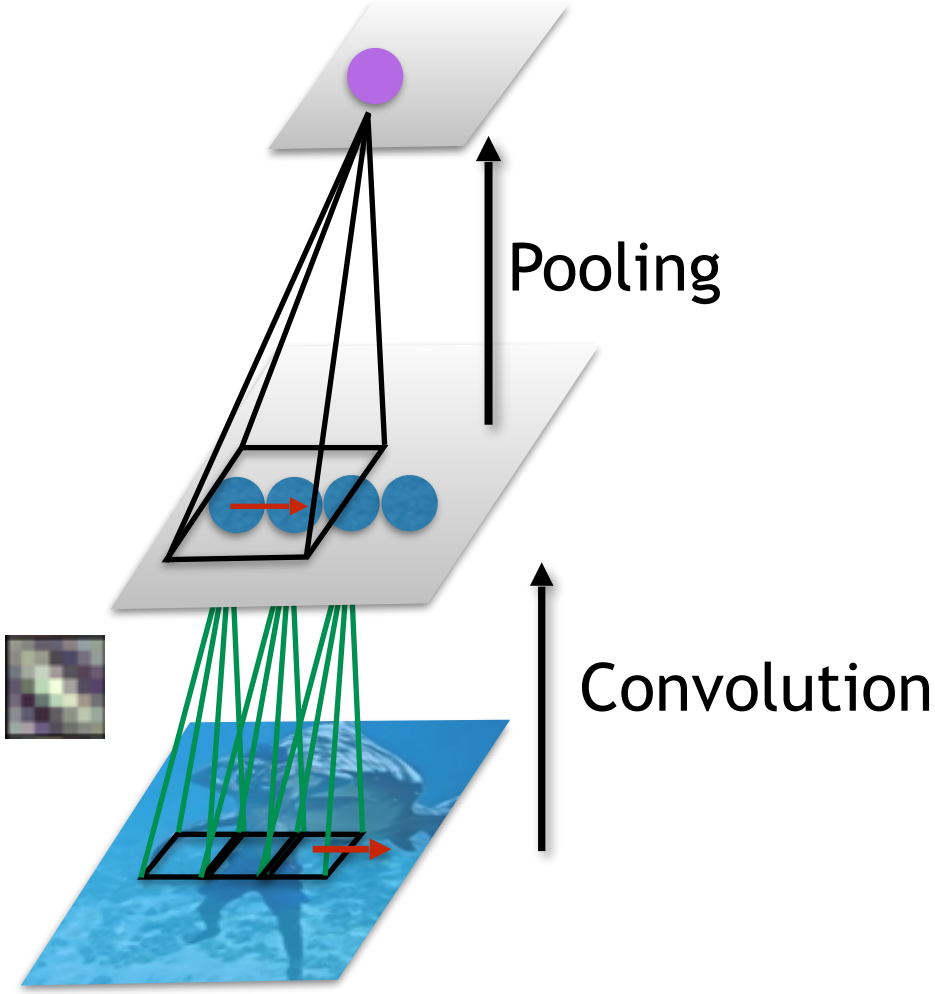


- Convolution
- Pooling
- Normalization
- Densely connected

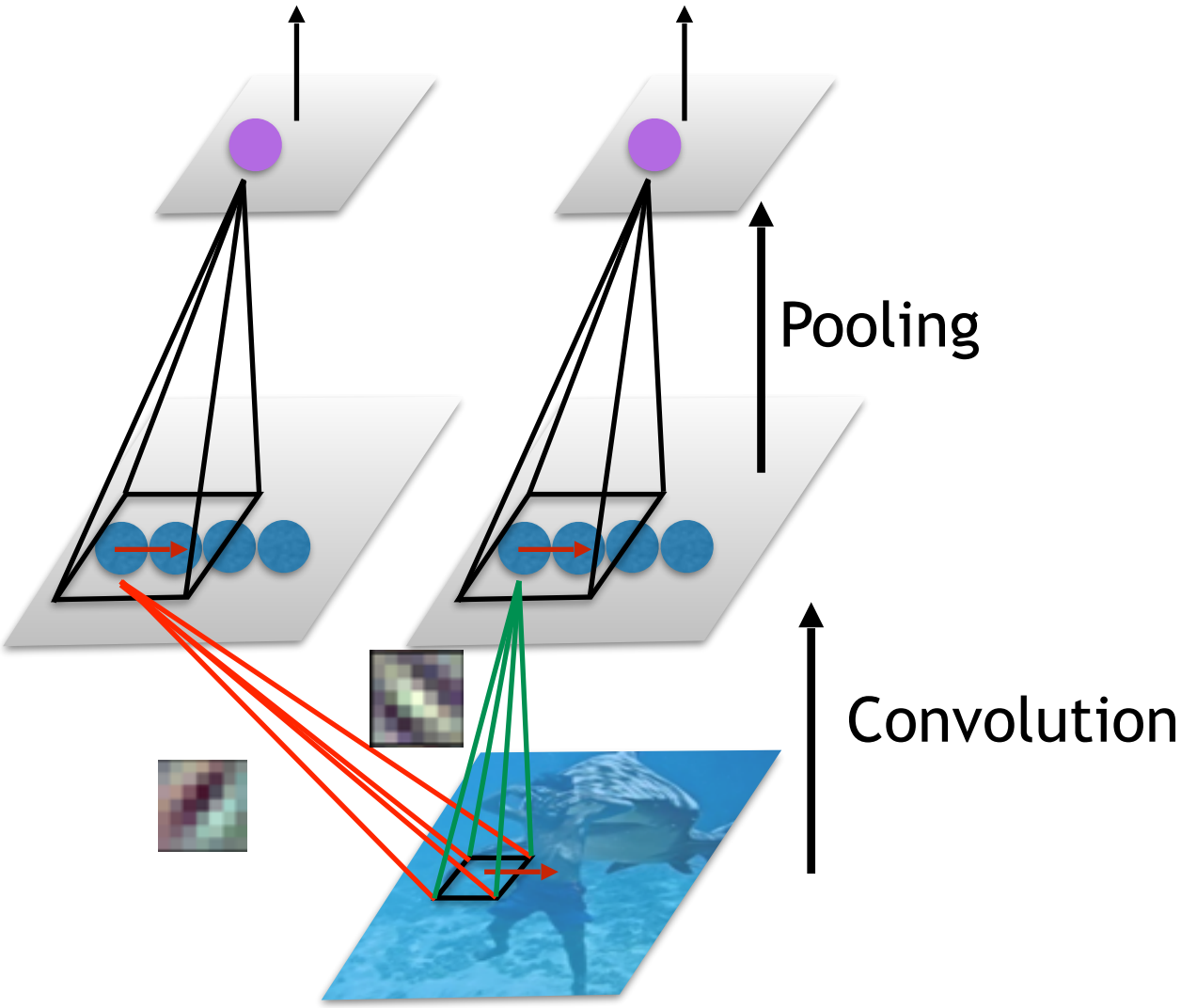
Deep Convolutional Nets



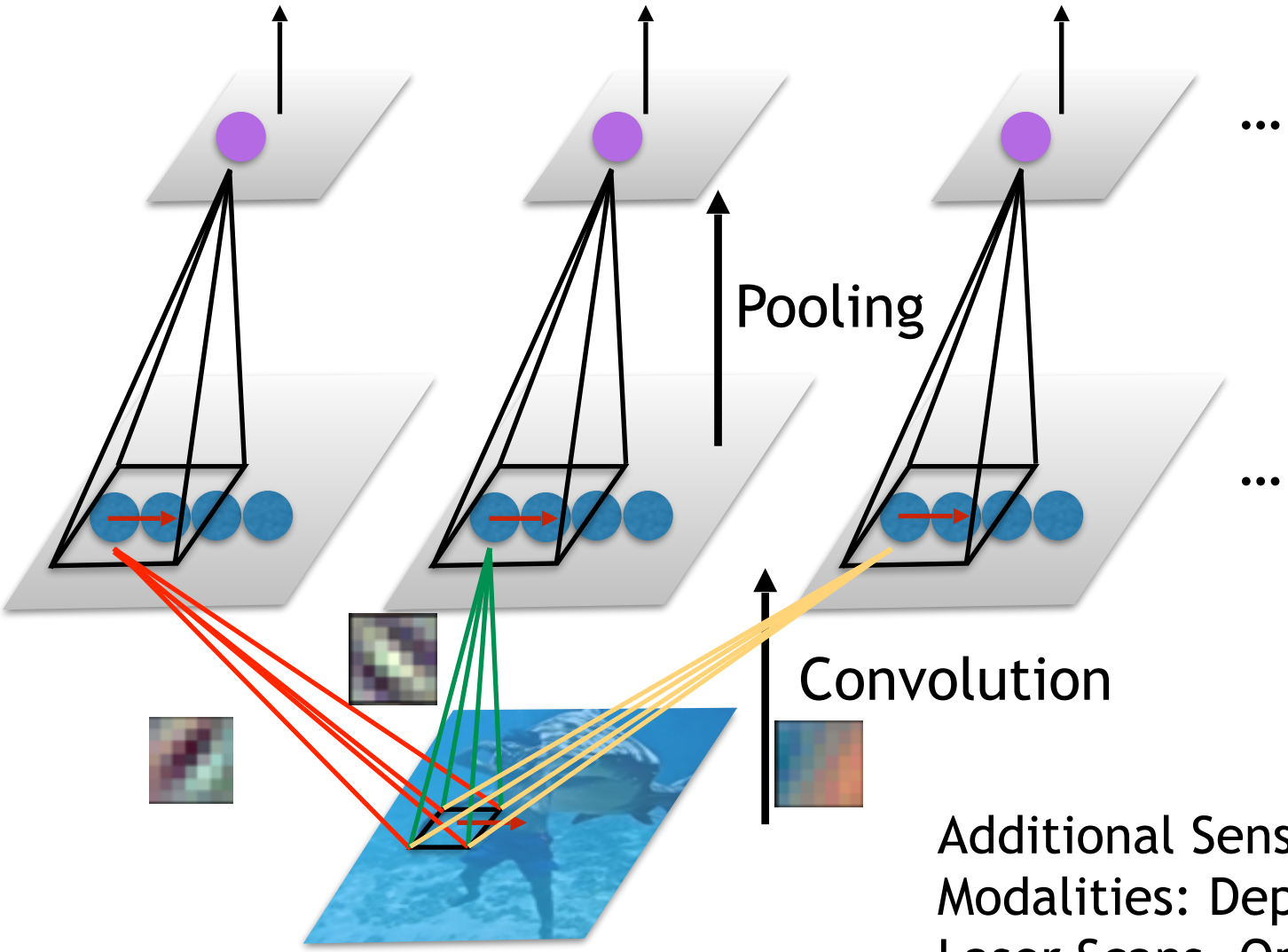
Deep Convolutional Nets



Deep Convolutional Nets

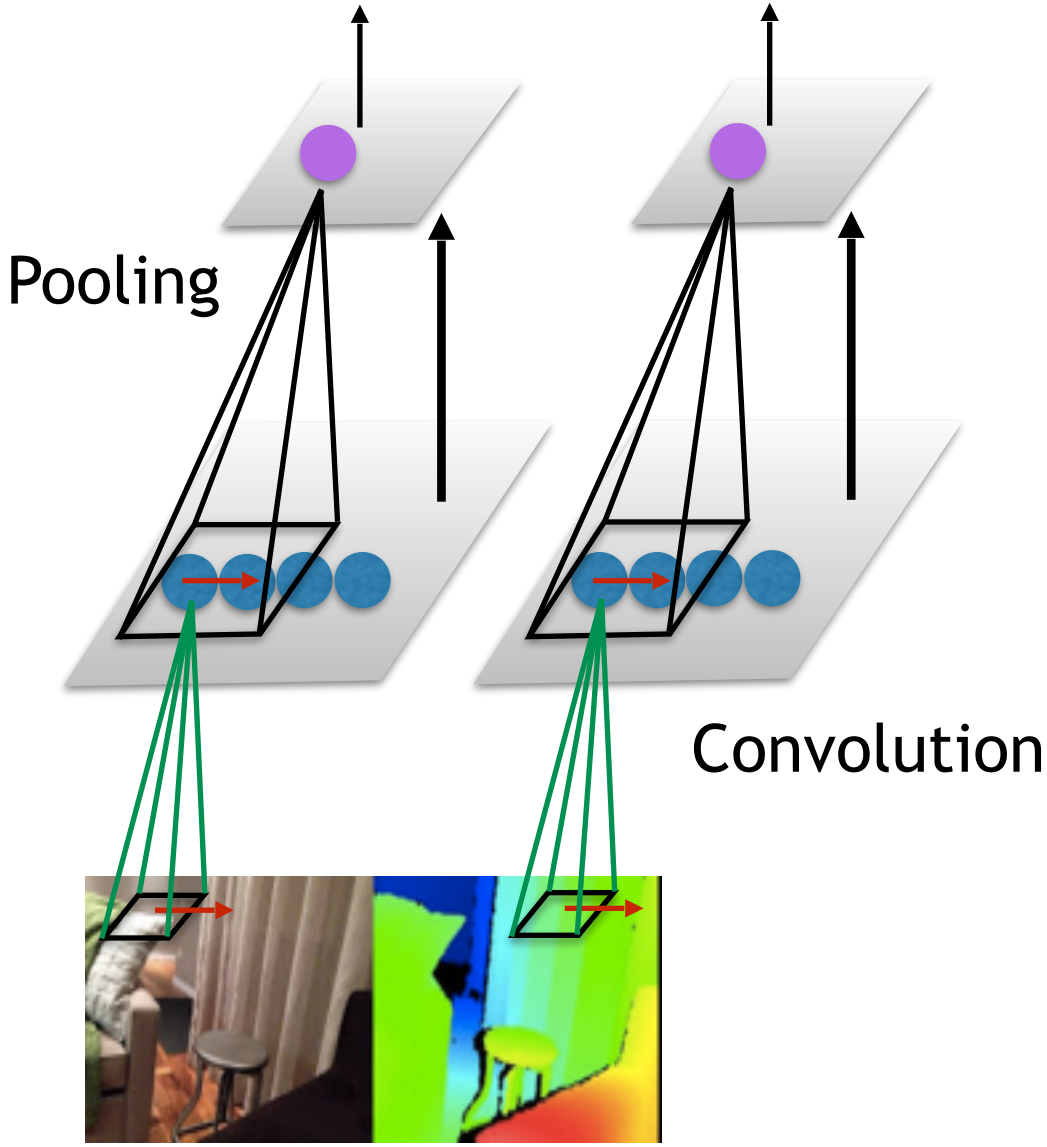


Deep Convolutional Nets



Additional Sensor Modalities: Depth, Laser Scans, Optical Flow (for video)

Multimodal Deep Convolutional Nets



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?

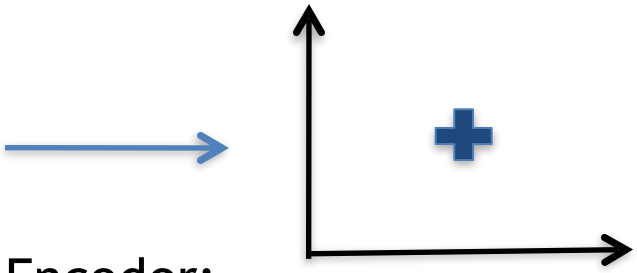
Input

Output

A man skiing down the snow covered mountain with a dark sky in the background.

An Image-Text Encoder-Decoder

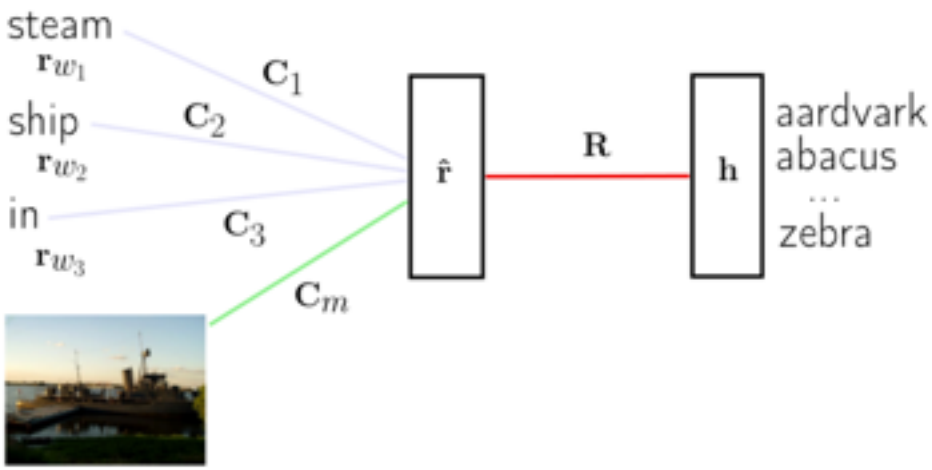
Feature space



steam ship in the water

Encoder: convolutional neural net

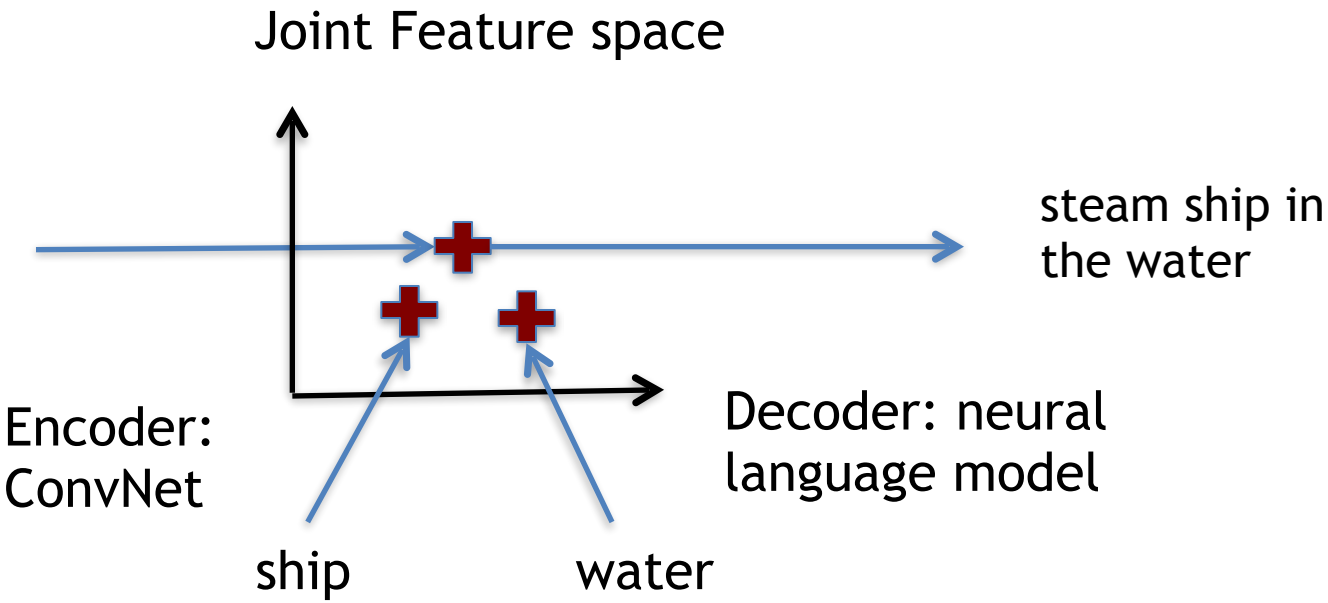
Decoder: neural language model



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



- 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

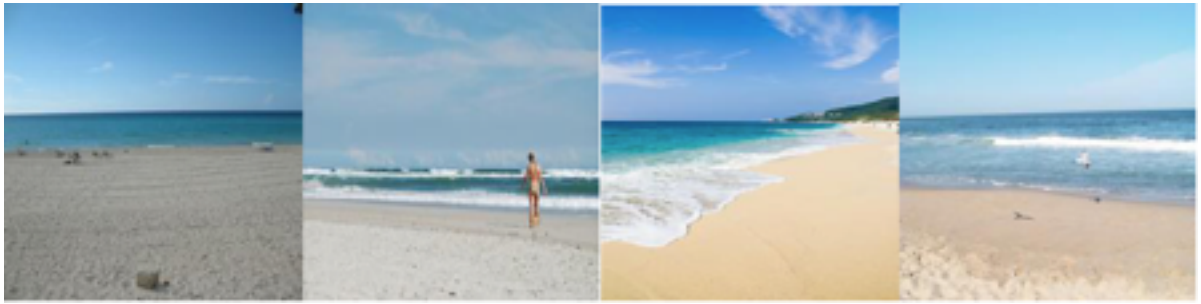
mosque, tower, building, cathedral, dome, castle

ski, skiing, skiers, skiiers, snowmobile

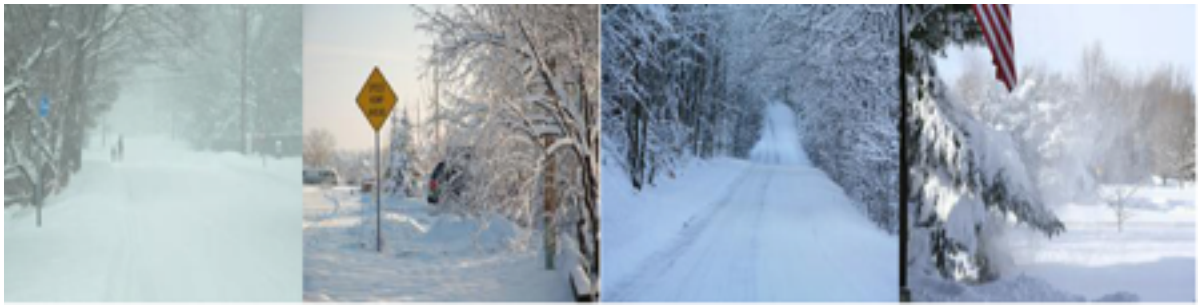
kitchen, stove, oven, refrigerator, microwave

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

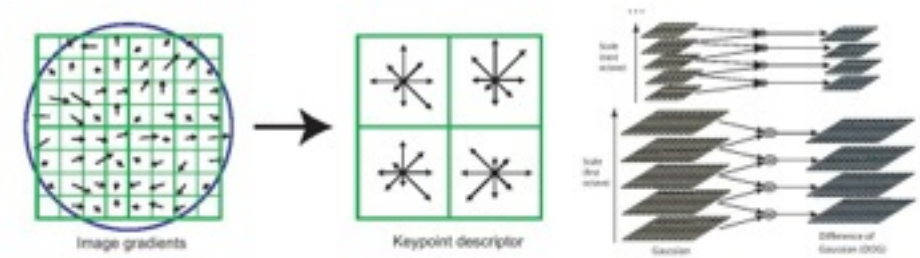
yummy, delicious, plastic, foamy, savory

(Kiros, Salakhutdinov, Zemel, NIPS 2014)

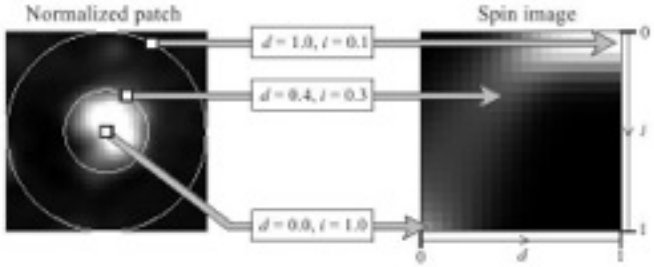
Checkout demo at

deeplearning.cs.toronto.edu

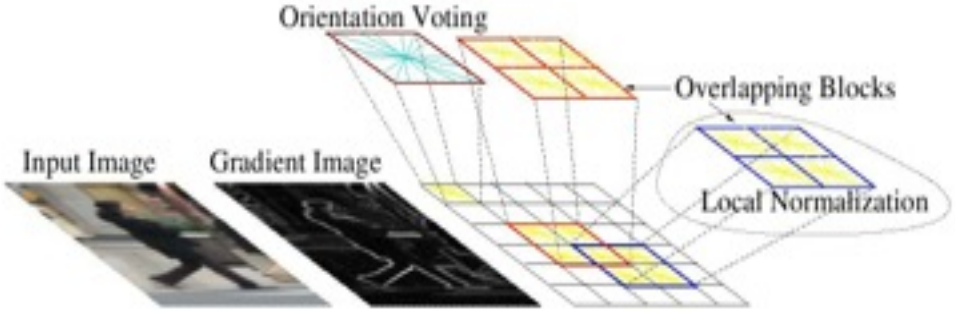
Computer Vision Features



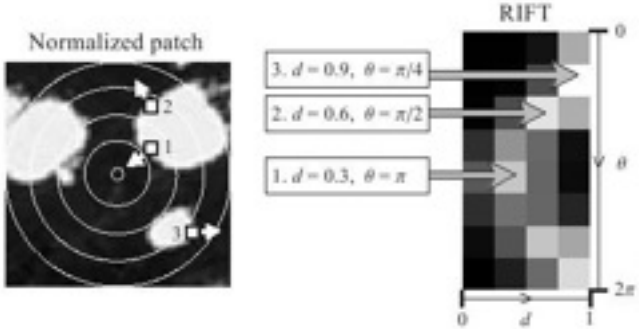
SIFT



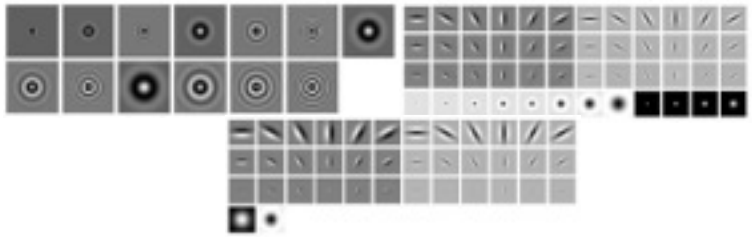
Spin image



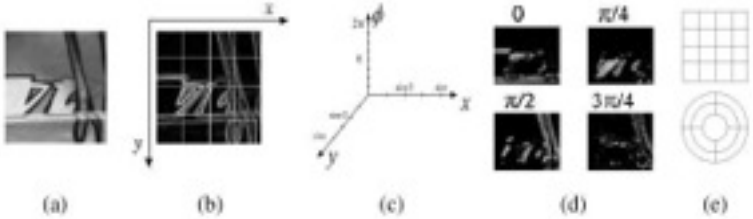
HoG



RIFT

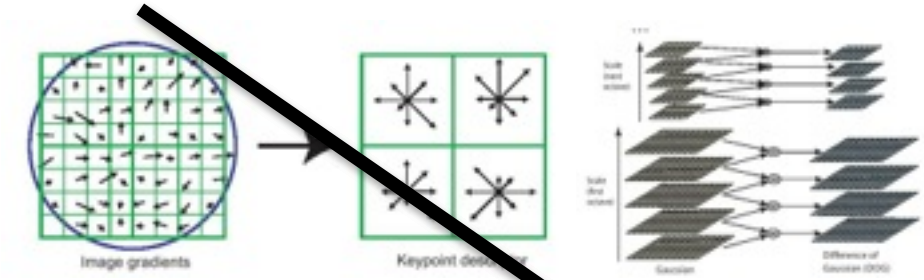


Textons

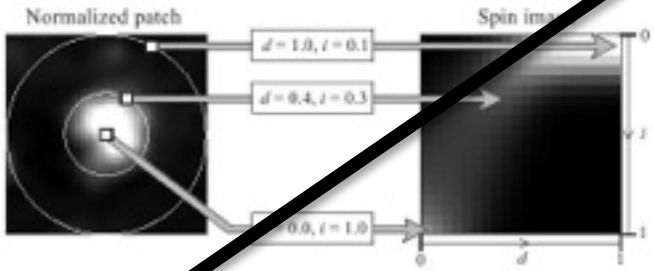


GLOH

Computer Vision Features

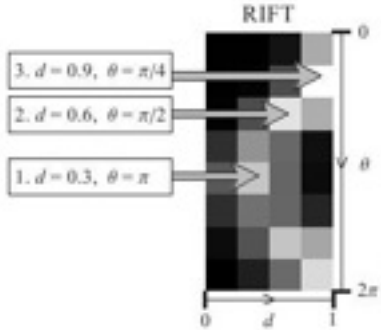


SIFT

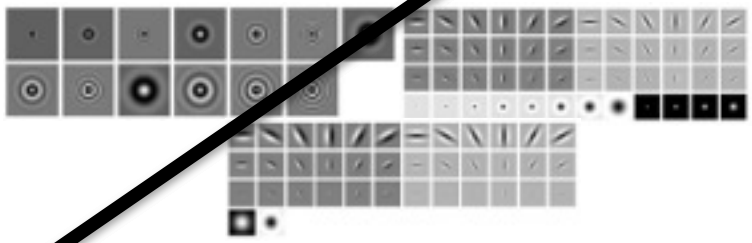


Spin image

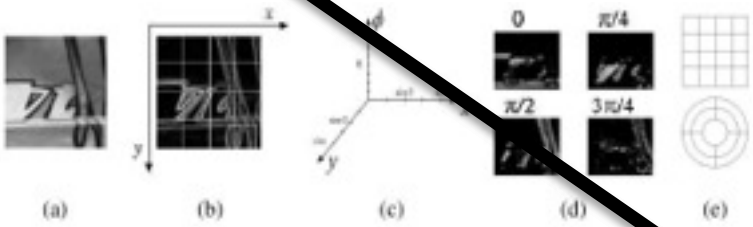
HoG



RIFT



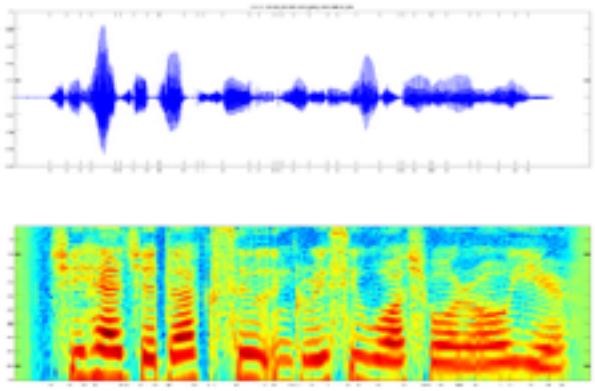
Textons



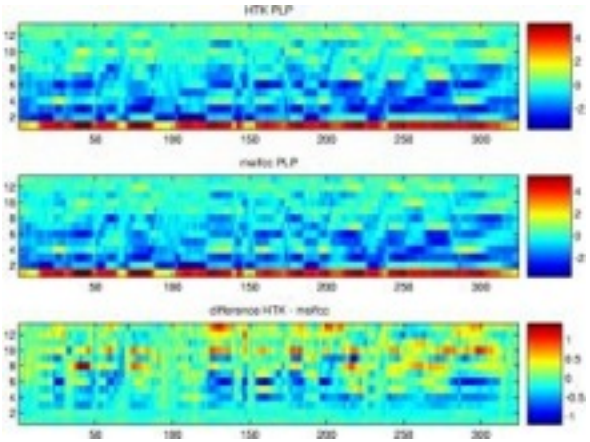
GLOH

Deep Learning

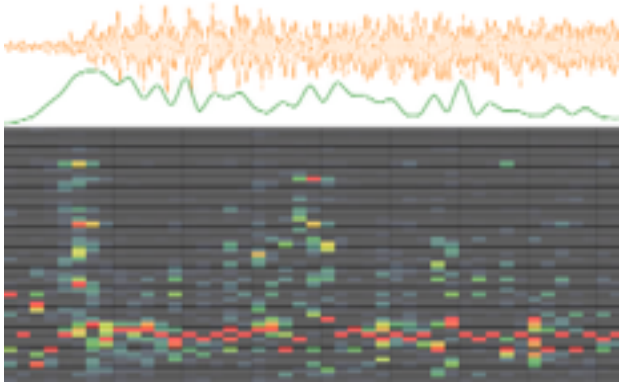
Audio Features



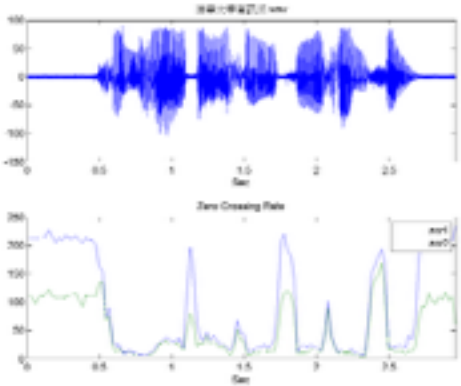
Spectrogram



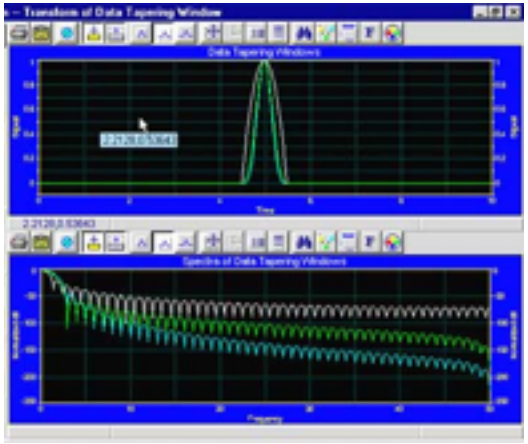
MFCC



Flux

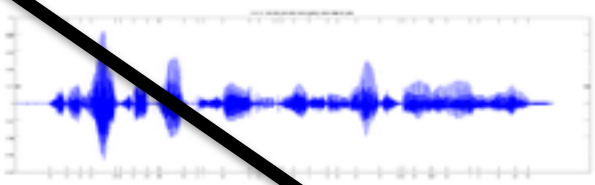


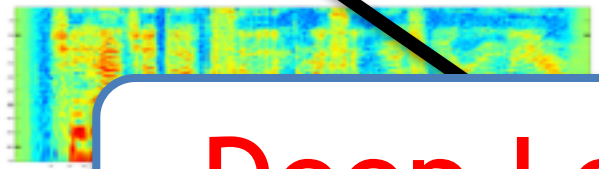
ZCR

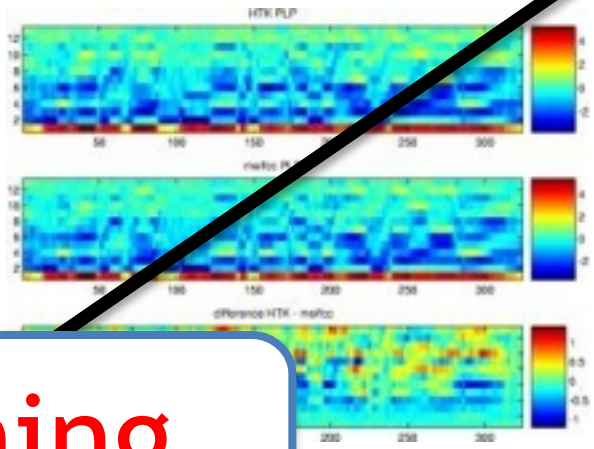


Rolloff

Audio Features

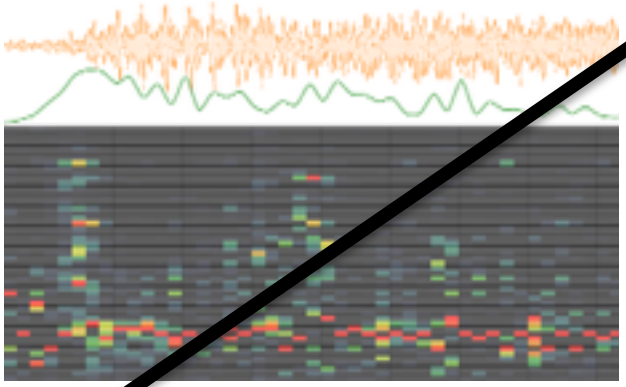




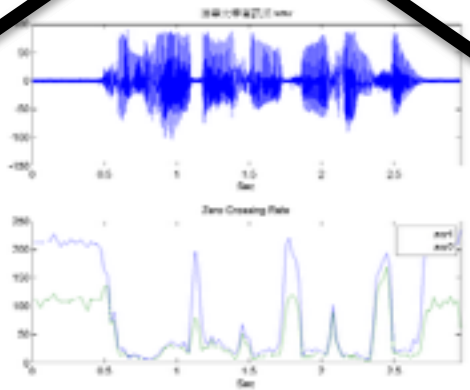


CC

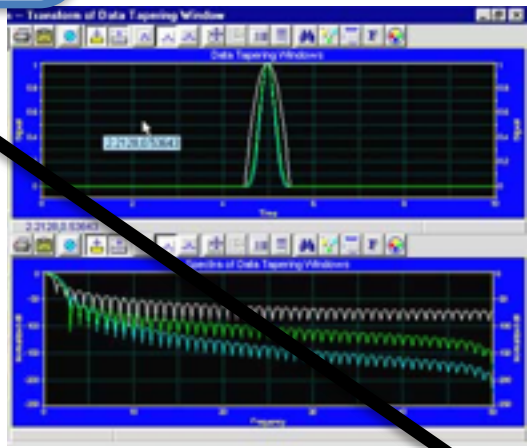
Deep Learning



Flux



ZCR



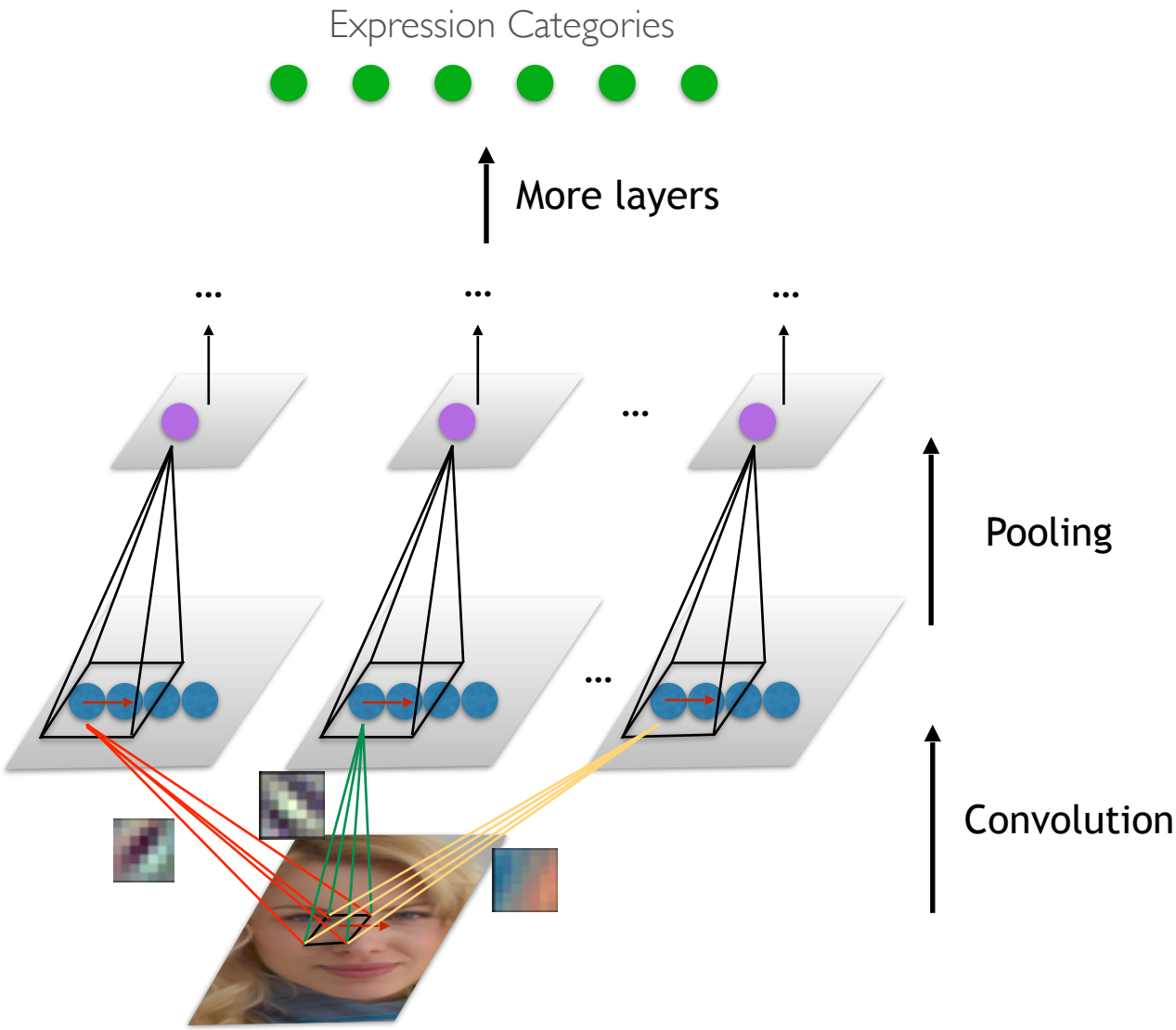
Rolloff

Expression Recognition

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

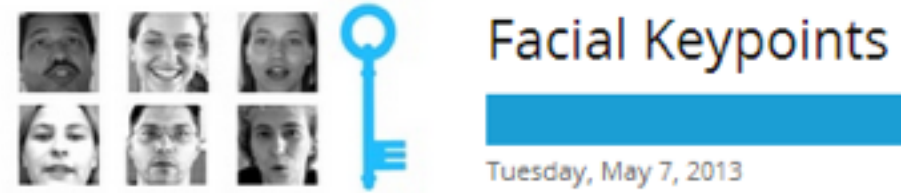
- **Surpassed human judge performances by 2%**

Expression Recognition



Facial Keypoints

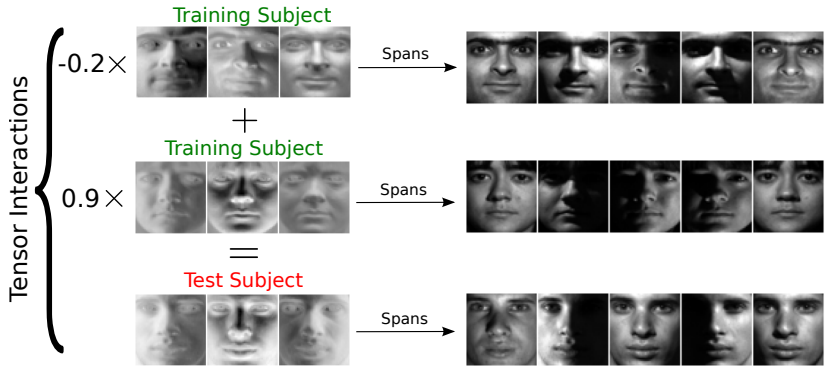




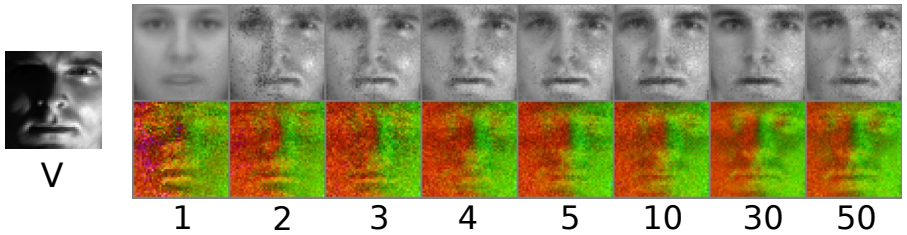
Kaggle Competition: predict points locations

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

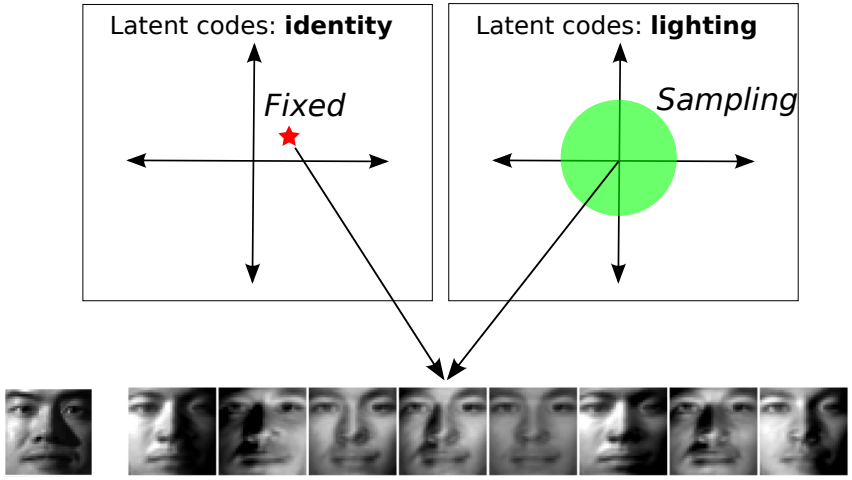
Illumination Variations

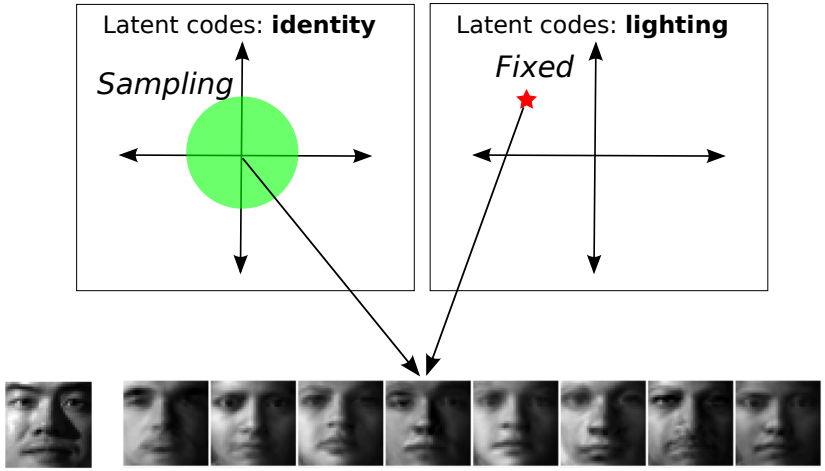


Multilinear



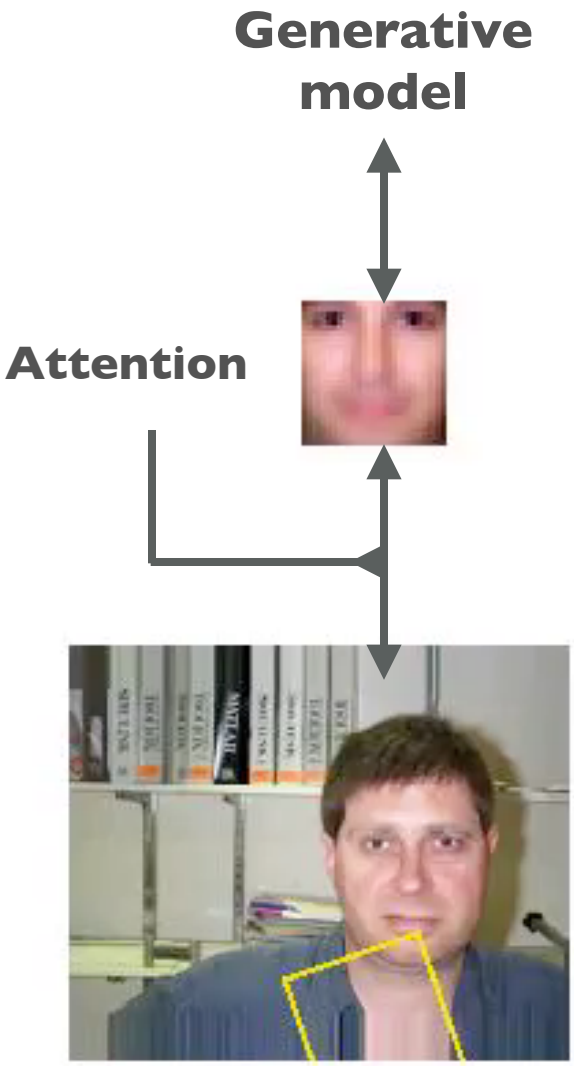
One-shot





Localization using Attention

Introduce a probabilistic framework that uses visual attention to learn generative models of objects of interest

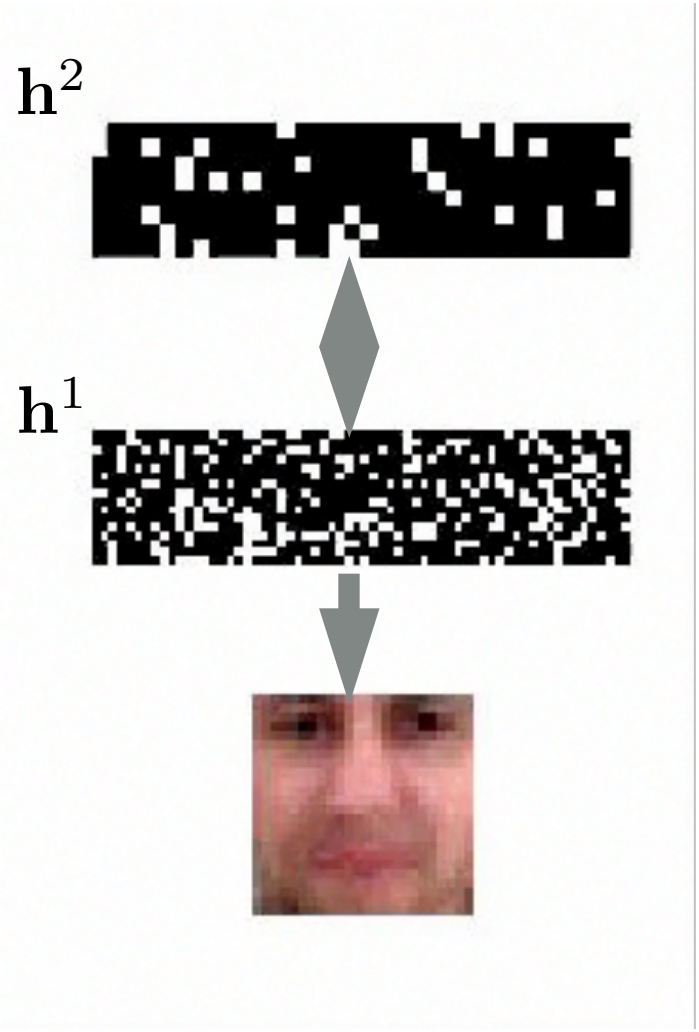


Deep Generative Face Model

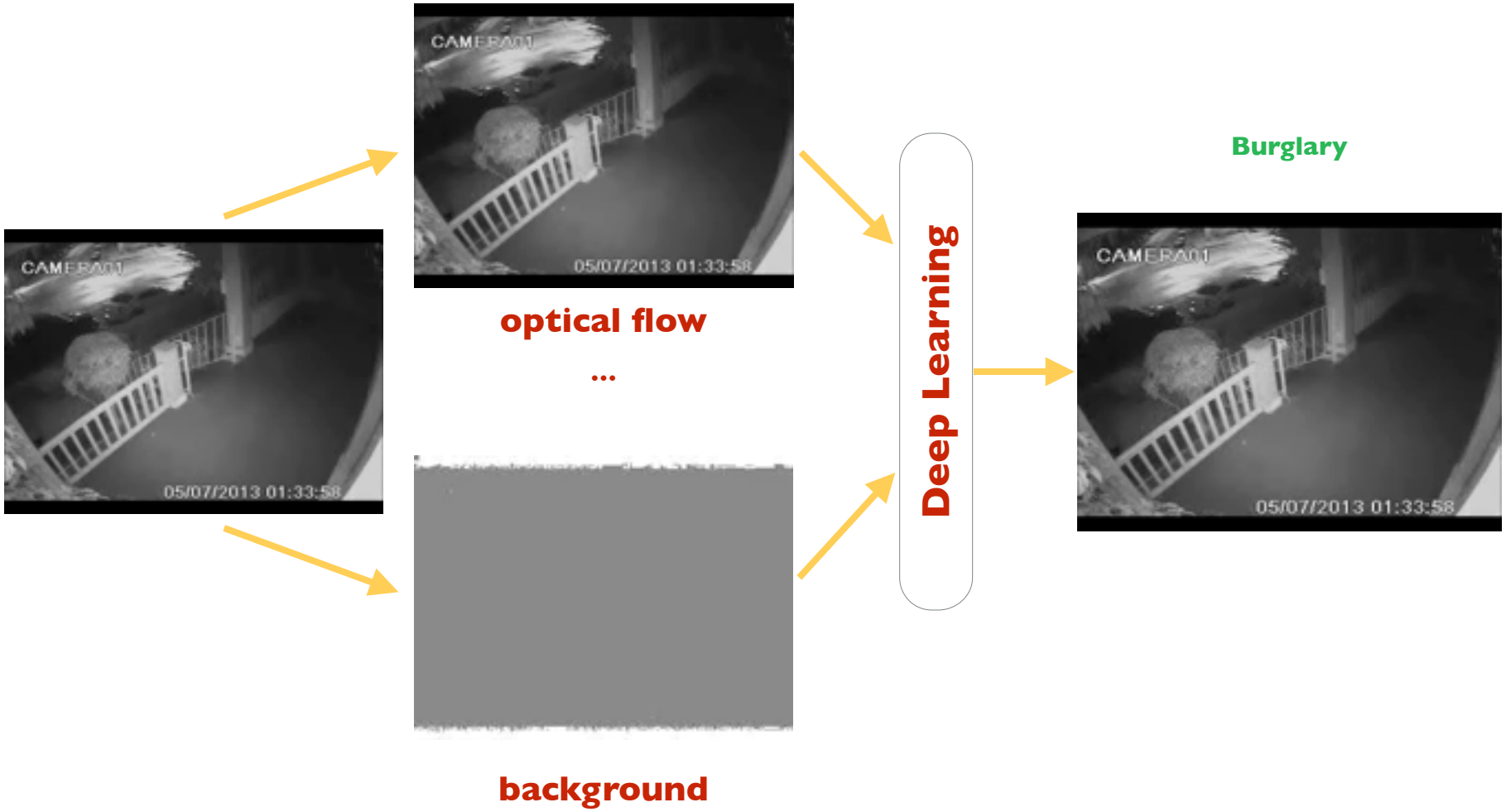
- Gaussian Deep Belief Net is a type of deep generative model

- Hybrid undirected-directed graphical model

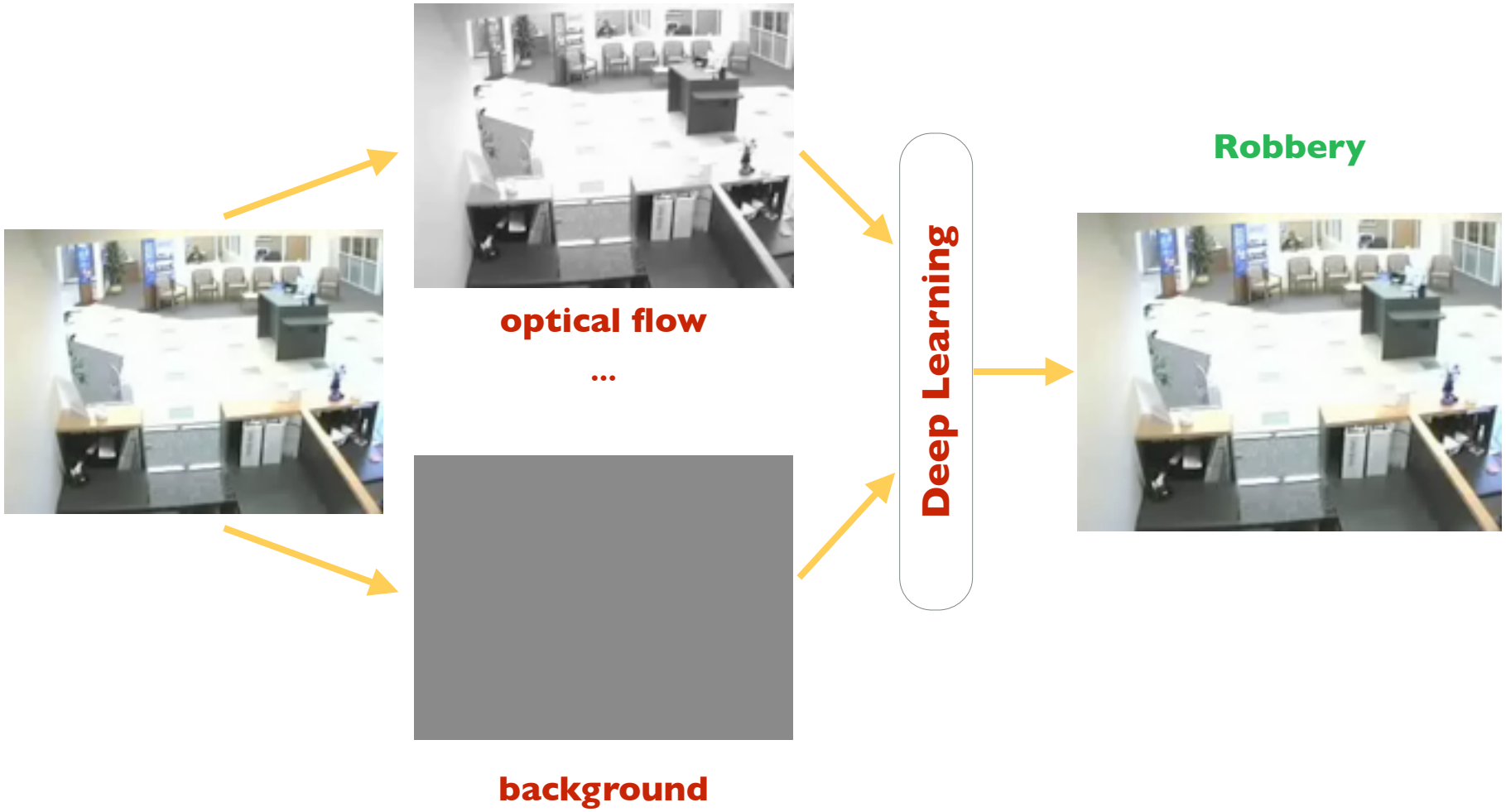
- Good generative model of objects and faces



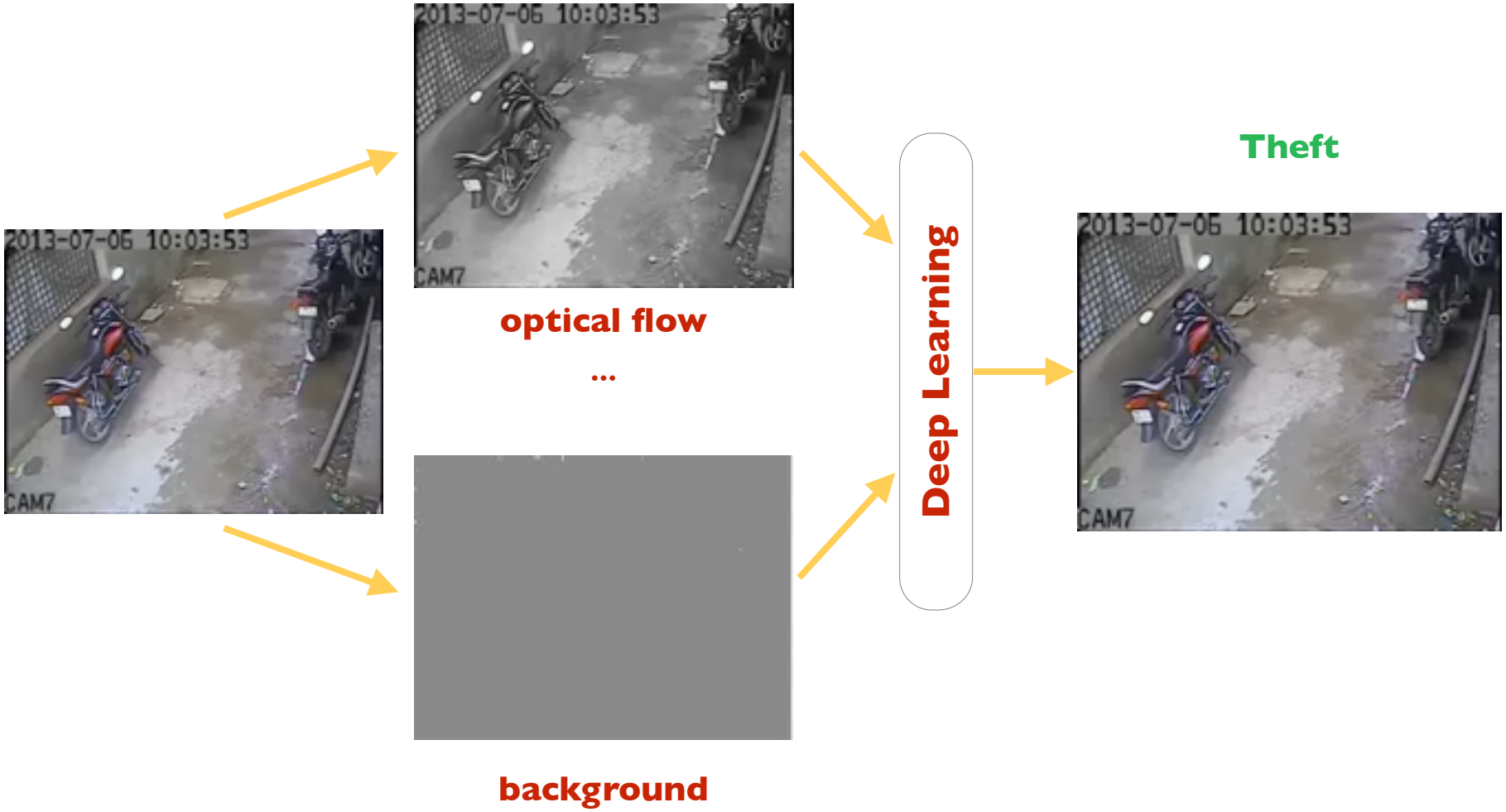
Human Activity Analytics



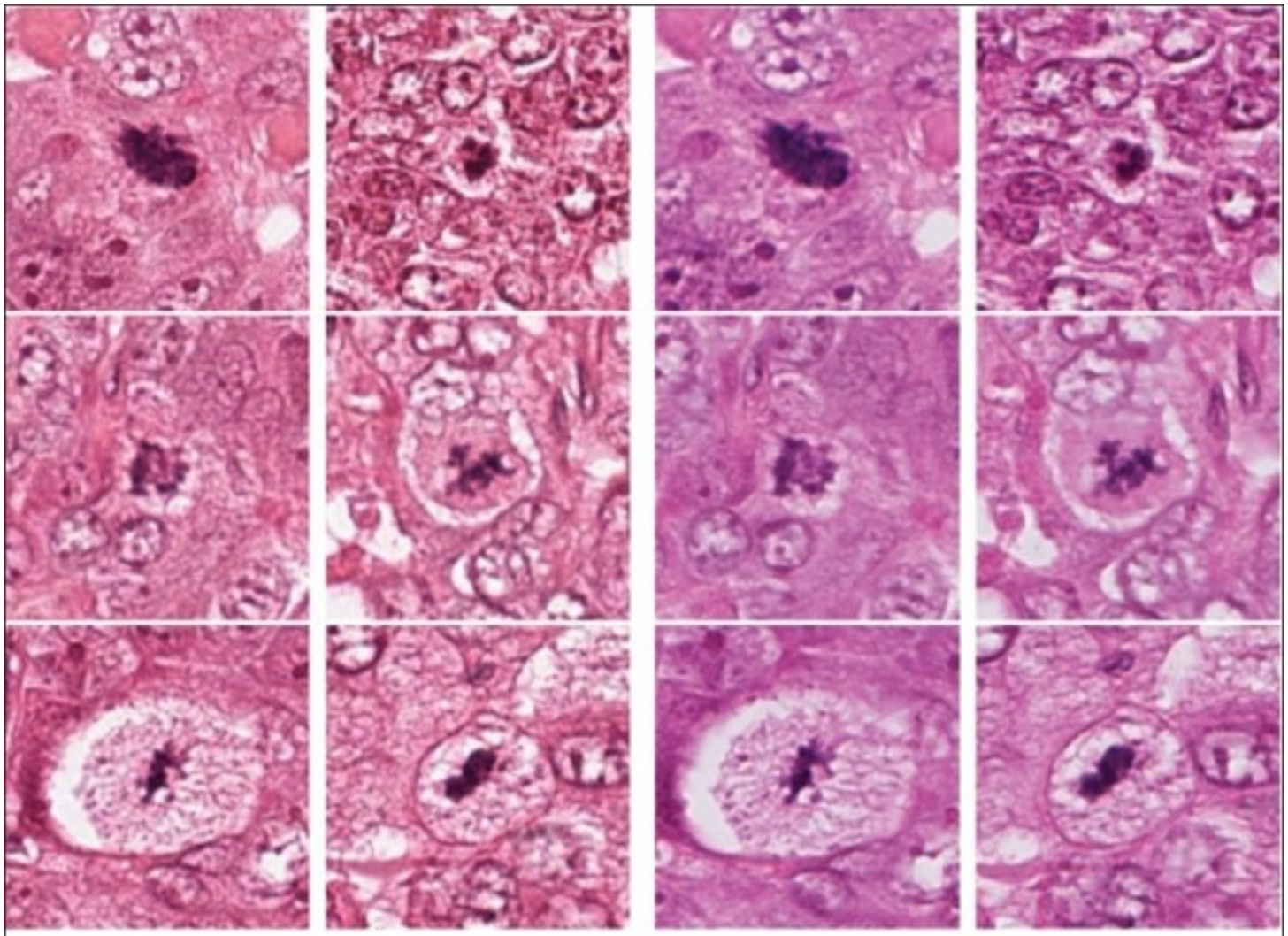
Human Activity Analytics



Human Activity Analytics

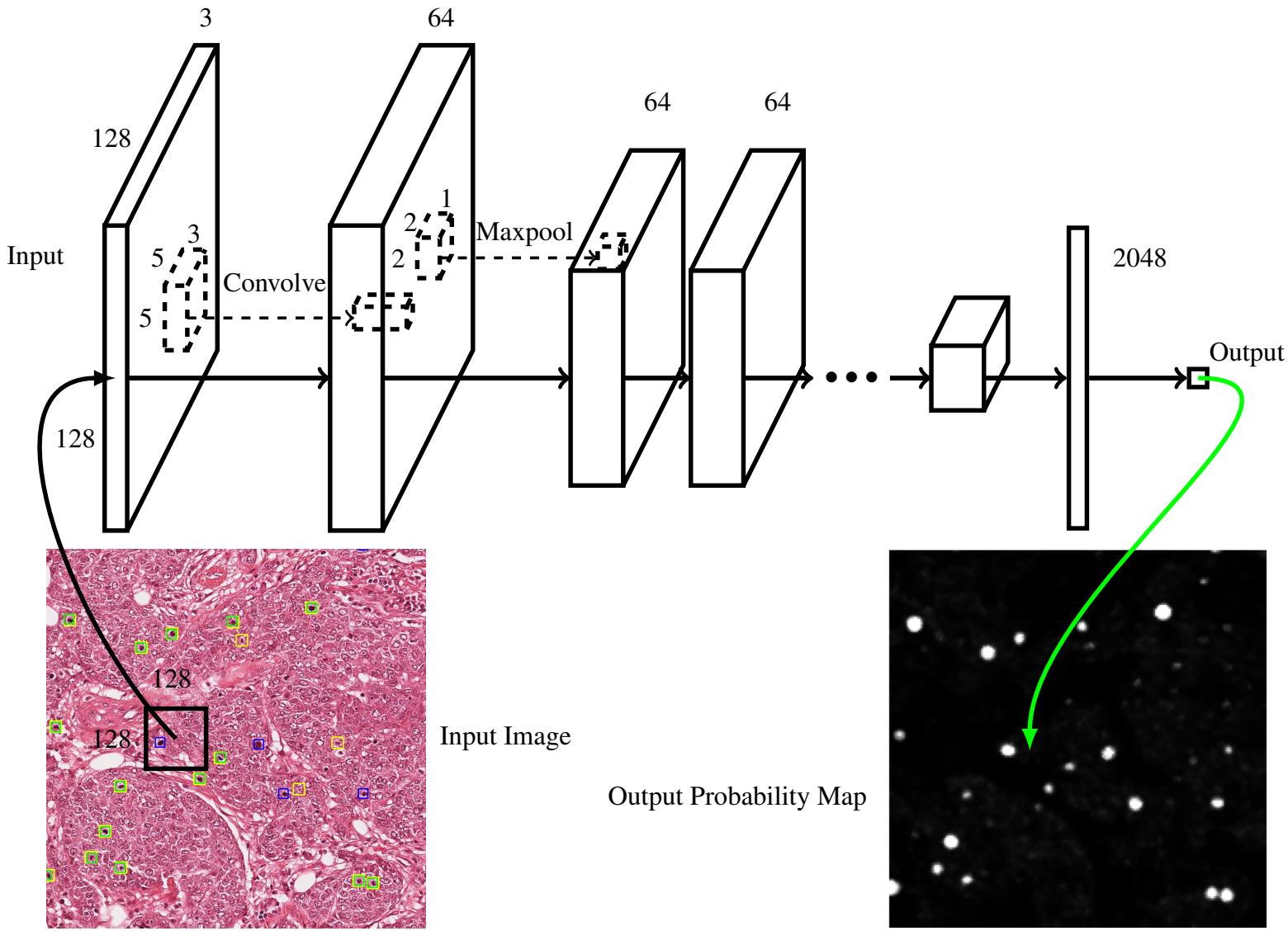


Brest Cancer - mitotic cells



(a) Scanner A (b) Scanner H

Detection Algorithm



Algorithm

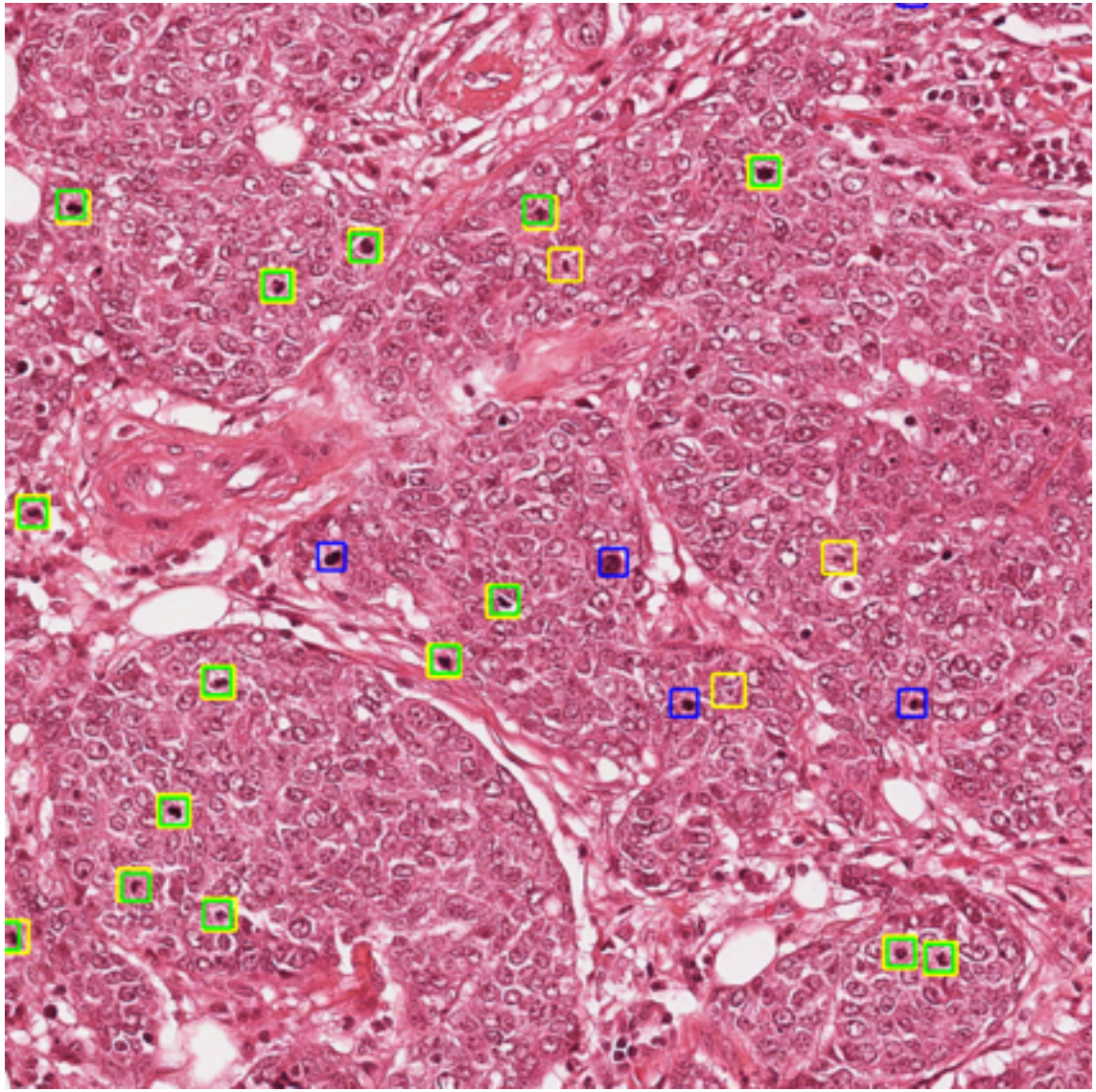
Fig. 3. *The detection process.* **Left***: Convolutional Net outputs a raw probability map.* **Middle***: Gaussian low-pass smoothing prepares the map for non-maximal suppression.* **Right***: Non-maximal suppression suppress the pixel of the map which are not local maximum. Zoom in to see the "white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

Results

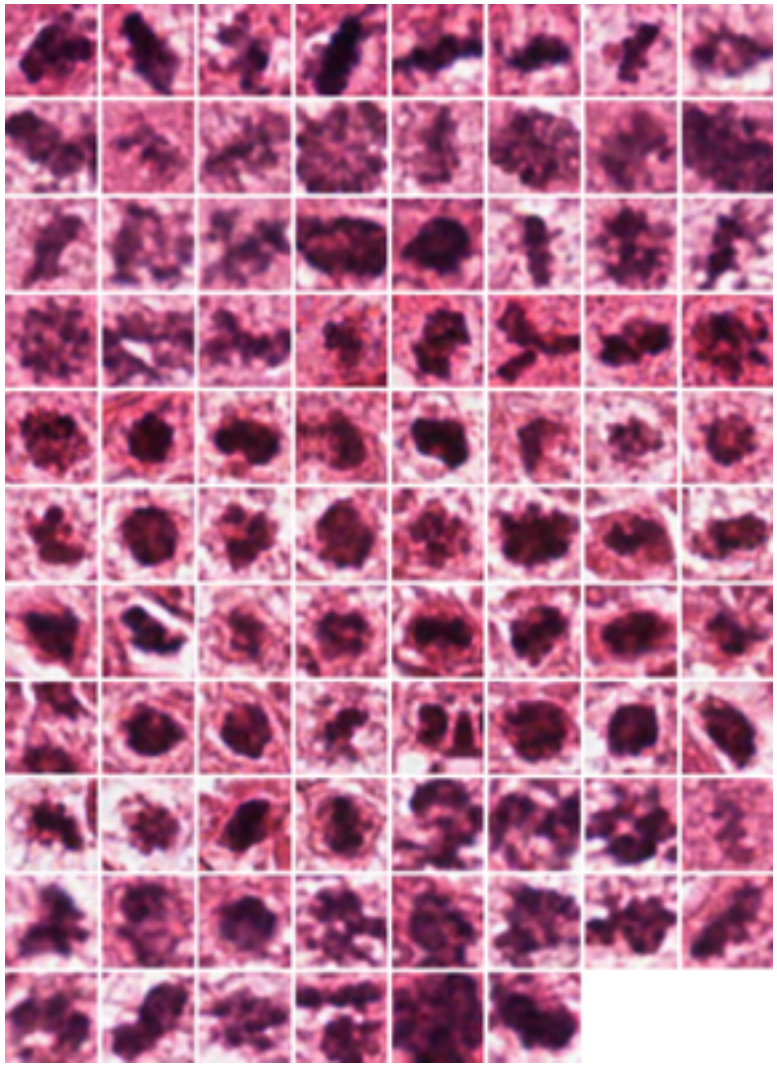
- Learning performs well even with a small dataset
- 50 images with 300 mitosis cells
- Better than untrained human
- State-of-the-art across all literature

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

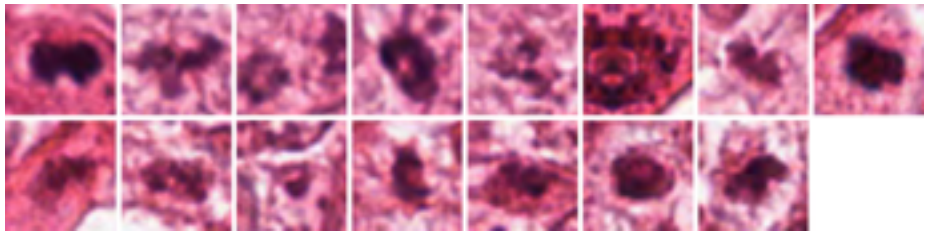
Table 2. Detection performance on the test set images. Comparison with previously published results on the exact same test set. Our result is state-of-the-art.



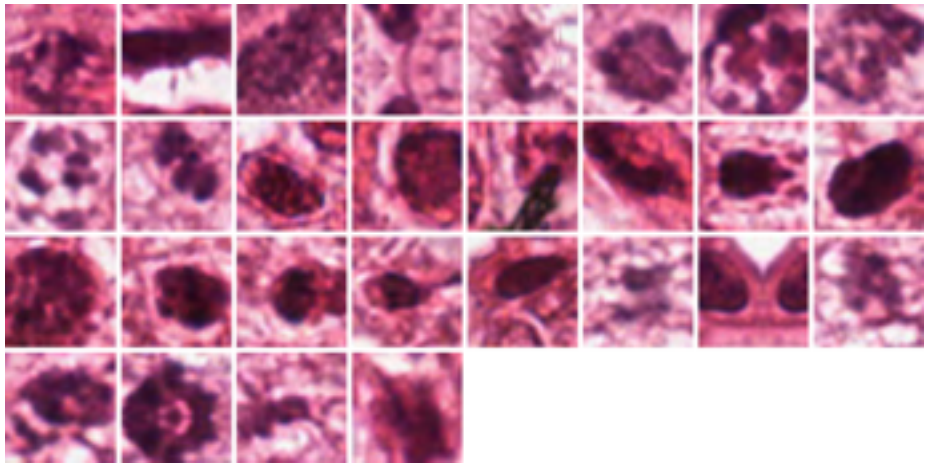
Results



True Positives



False Negatives



False Positives