Introduction to Deep Learning

Zhujin Liang
Intelligent Media Computing (IMC) Laboratory
Sun Yat-Sen University

Some slides are from X. Wang, R. Fergus, H. Lee, B. Póczos and A. Singh.

10 BREAKTHROUGH TECHNOLOGIES 2013

MIT Technology Review

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?

Additive Manufacturing

Skeptical about 3-D printing? GE, the world's largest manufacturer, is on the verge of using the technology to make jet parts.

Baxter: The Blue-Collar Robot

Rodney Brooks's newest creation is easy to interact with, but the complex innovations behind the robot show just how hard it is to get along with people.

Memory Implants

A maverick neuroscientist believes he has deciphered the code by which the brain forms long-term memories. Next: testing a prosthetic implant for people suffering from long-term memory loss.

Smart Watches

The designers of the Pebble watch realized that a mobile phone is more useful if you don't have to take it out of your pocket.

Ultra-Efficient Solar Power

Doubling the efficiency of a solar cell would completely change the economics of renewable energy. Nanotechnology just might make it possible.

Big Data from Cheap Phones

Collecting and analyzing information from simple cell phones can provide surprising insights into how people move about and behave – and even help us understand the spread of diseases.

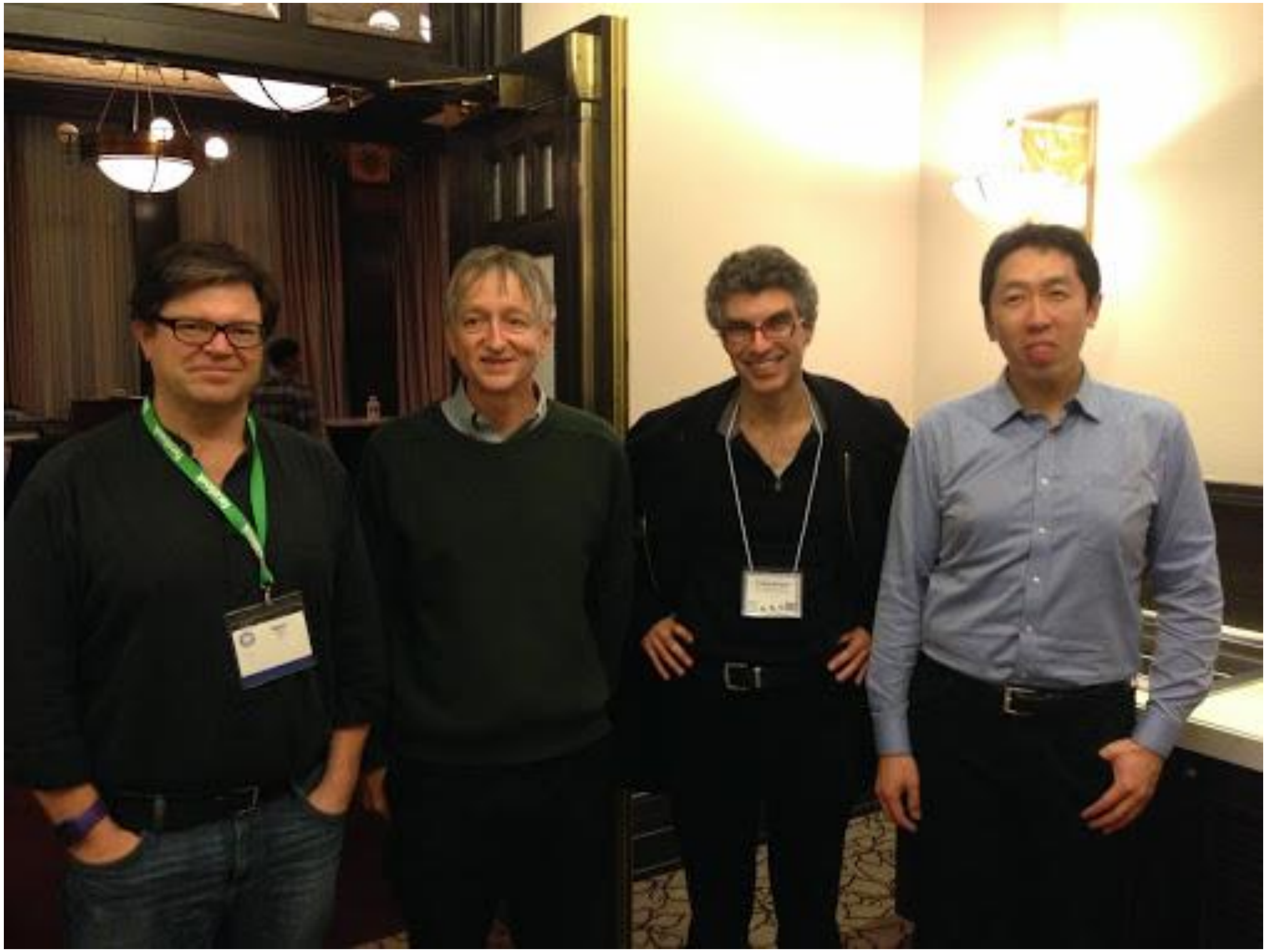
Supergrids

A new high-power circuit breaker could finally make highly efficient DC power grids practical.

News on Deep Learning

Baidu established Institute of Deep Learning	2012
Hinton's group won ImageNet Contest	Oct. 2012
Hinton joined Google	March 2013
Google announced deep learning based visual search engine	March 2013
Baidu announced deep learning based visual search engine	June 2013
Yahoo acquired startup LookFlow working on deep learning	Oct. 2013
Facebook established a new AI lab in NewYork and recruited Yann LeCun	Dec. 2013
Google Acquires DeepMind for USD 400 Million	January 2014
Baidu established a new lab at Shenzhen, China	2014
Baidu established a new lab at silicon valley and Andrwe Ng is the director	May 2014
Deep learning reached human performance on face verification on LFW	June 2014

Deep Learning Gurus



Examples from ImageNet

Result of ImageNet 2012

Rank	Name	Error rate	Description
1	**U. Toronto**	0.15315	Deep learning
2	U. Tokyo	0.26172	Hand-crafted features and learning models. Bottleneck.
3	U. Oxford	0.26979	
4	Xerox/INRIA	0.27058	

Object recognition over 1,000,000 images and 1,000 categories (2 GPU)

A. Krizhevsky, L. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," NIPS, 2012.

Result of ImageNet 2013

- ImageNet 2013 – image classification challenge

Rank	Name	Error rate	Description
1	NYU	0.11197	Deep learning
2	NUS	0.12535	Deep learning
3	Oxford	0.13555	Deep learning

MSRA, IBM, Adobe, NEC, Clarifai, Berkley, U. Tokyo, UCLA, UIUC, Toronto .... Top 20 groups all used deep learning

- ImageNet 2013 – object detection challenge

Rank	Name	Mean Average Precision	Description
1	UvA-Euvision	0.22581	Hand-crafted features
2	NEC-MU	0.20895	Hand-crafted features
3	NYU	0.19400	Deep learning

Result of ImageNet 2014

- ImageNet 2014 – Image classification challenge

Rank	Name	Error rate	Description
1	Google	0.06656	Deep learning
2	Oxford	0.07325	Deep learning
3	MSRA	0.08062	Deep learning

- ImageNet 2014 – object detection challenge

Rank	Name	Mean Average Precision	Description
1	Google	0.43933	Deep learning
2	CUHK	0.40656	Deep learning
3	DeepInsight	0.40452	Deep learning
4	UvA-Euvision	0.35421	Deep learning
5	Berkley Vision	0.34521	Deep learning

News on Deep Learning

Baidu established Institute of Deep Learning	2012
Hinton's group won ImageNet Contest	Oct. 2012
Hinton joined Google	March 2013
Google announced deep learning based visual search engine	March 2013
Baidu announced deep learning based visual search engine	June 2013
Yahoo acquired startup LookFlow working on deep learning	Oct. 2013
Facebook established a new AI lab in NewYork and recruited Yann LeCun	Dec. 2013
Google Acquires DeepMind for USD 400 Million	January 2014
Baidu established a new lab at Shenzhen, China	2014
Baidu established a new lab at silicon valley and Andrwe Ng is the director	May 2014
Deep learning reached human performance on face verification on LFW	June 2014

Machine Learning

$$x \longrightarrow \boxed{F(x)} \longrightarrow y \begin{cases} \text{Class label (Classification)} \\ \text{Vector (Estimation)} \end{cases}$$

Object recognition

{dog, cat, horse, flower, …}

Low-resolution image

Super resolution

High-resolution image

Traditional Recognition Approach

Features are not learned

Input data (pixels) → feature representation (hand-crafted) → Learning Algorithm (e.g., SVM)

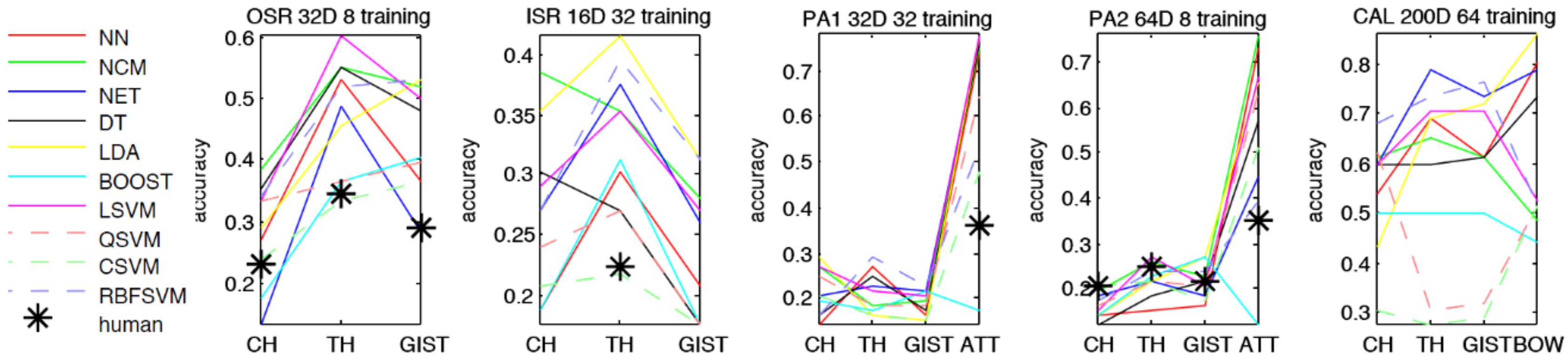
Image

Low-level vision features (edges, SIFT, HOG, etc.)

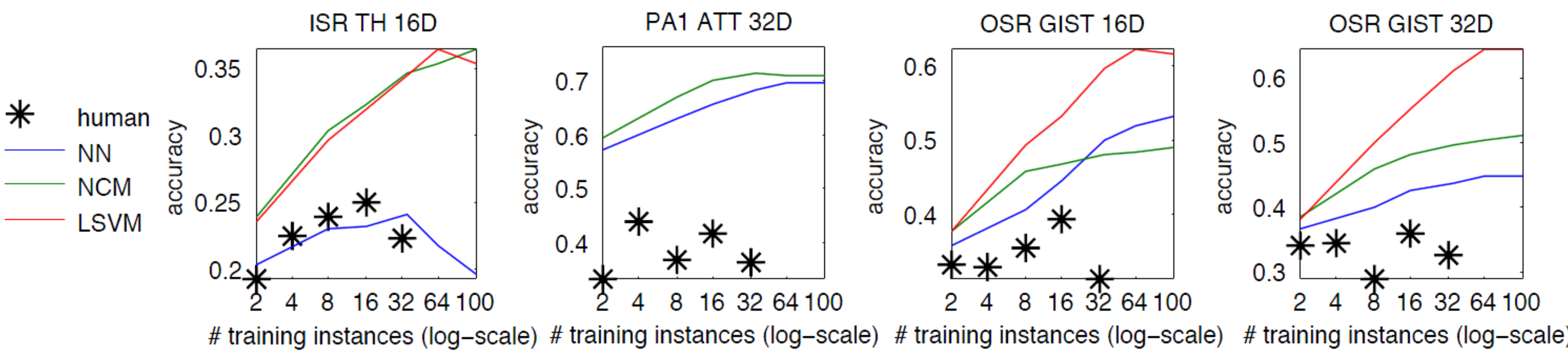
Object detection / classification

What Limits Current Performance

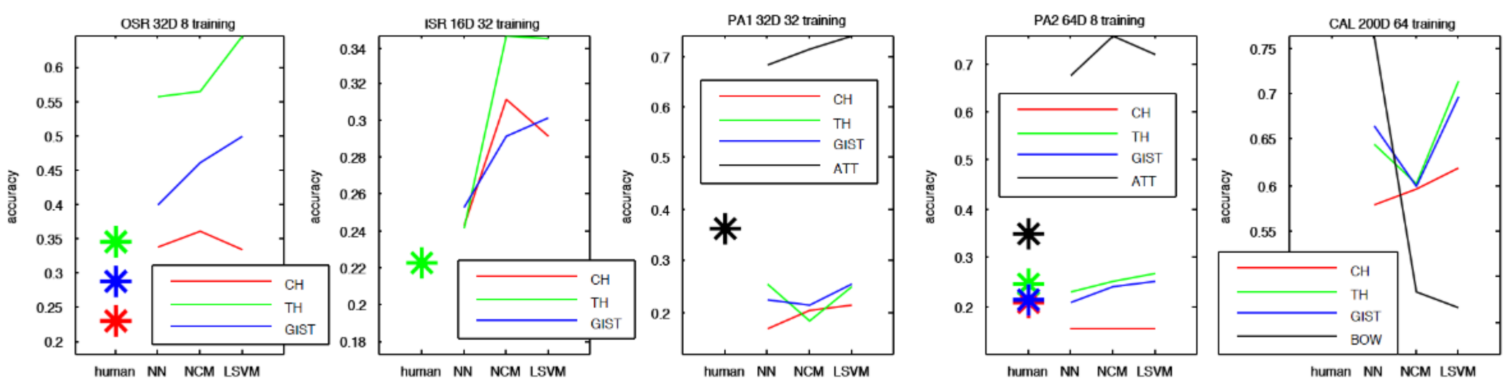
Role of Learning Algorithms



Role of Data

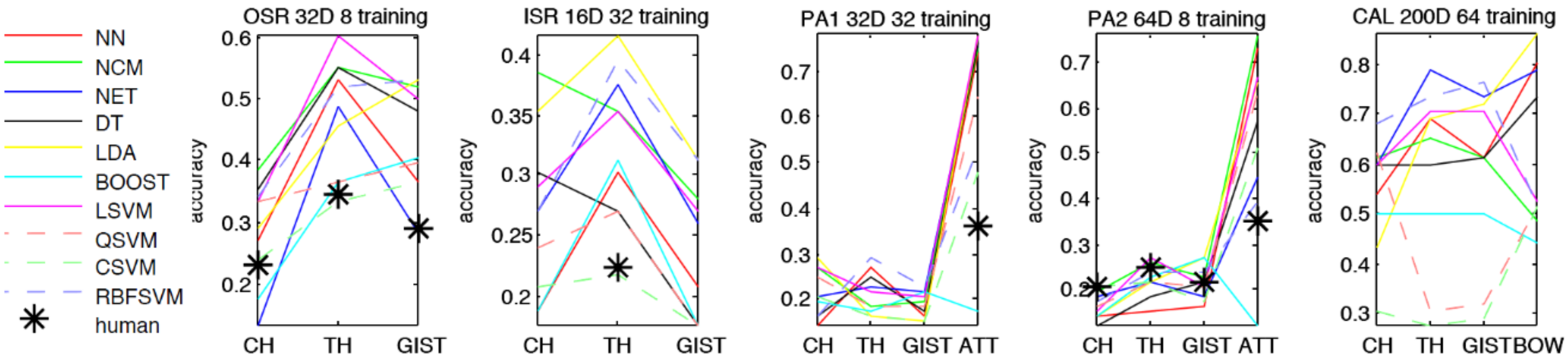


Role of Features

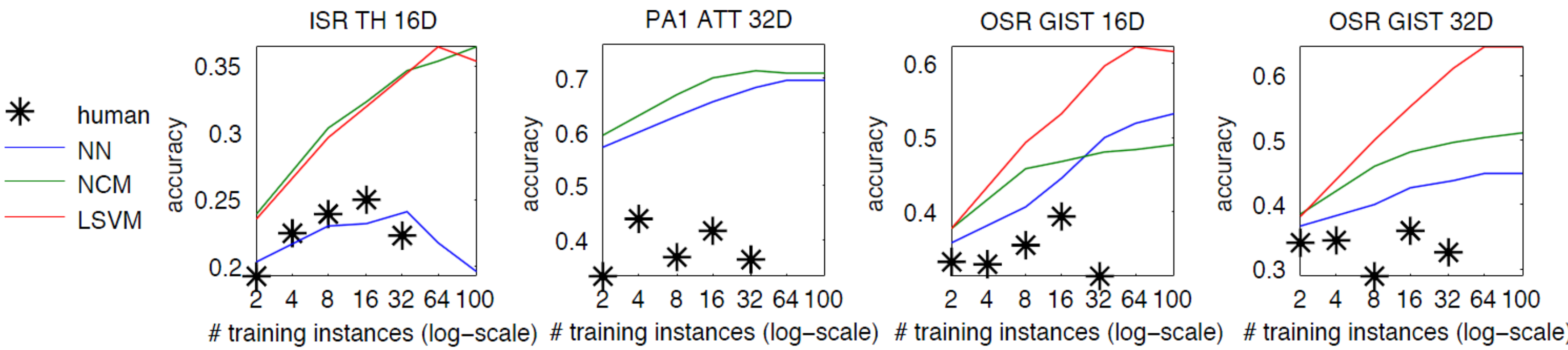


What Limits Current Performance

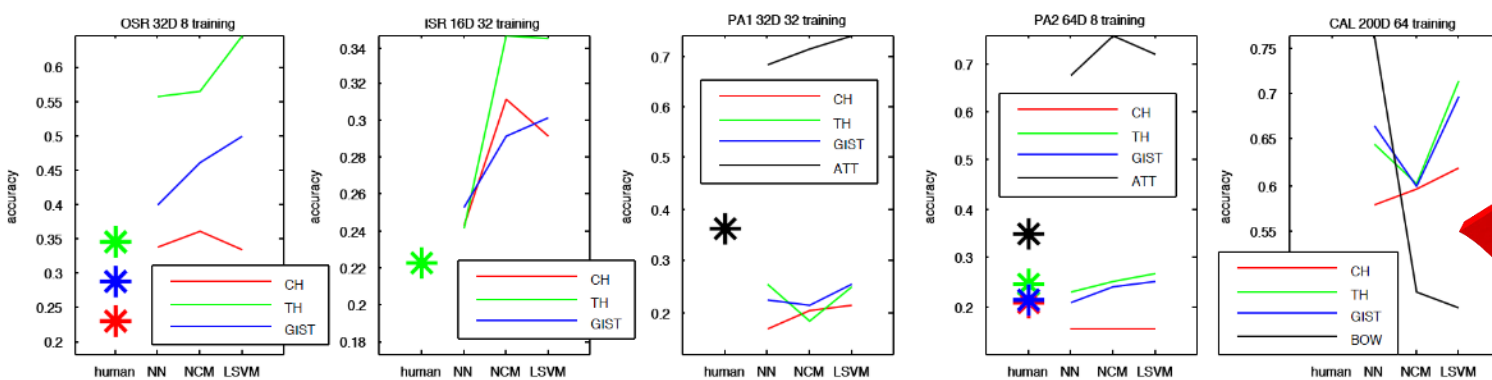
Role of Learning Algorithms



Role of Data

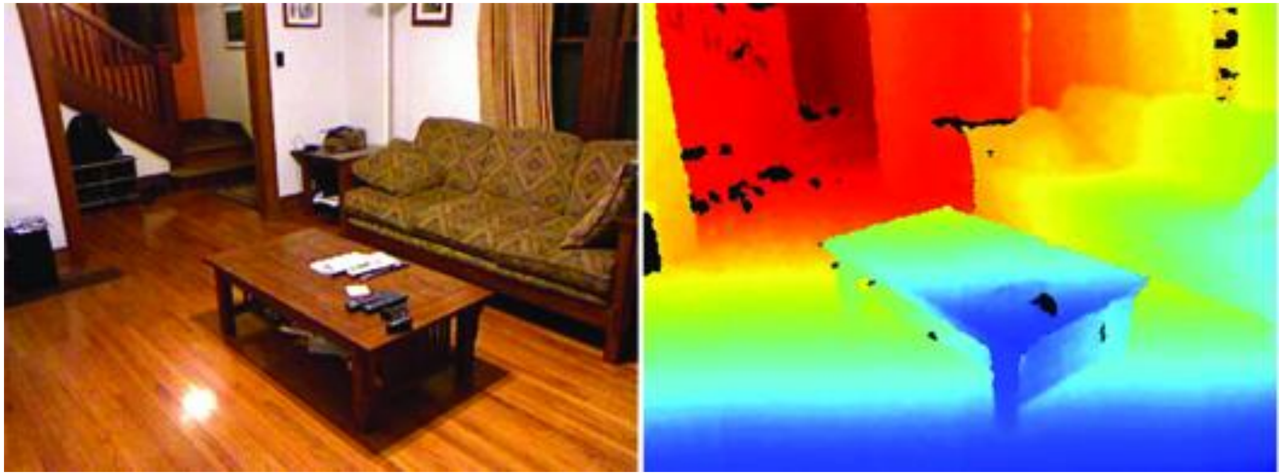


Role of Features



Motivation

- Features are the main factor impacting accuracies.
- Multitude of hand-designed features currently in use.
  - Reply on human domain knowledge much more than data
  - Feature design is separate from training the classifier
  - If handcrafted features have multiple parameters, it is hard to manually tune them
  - Developing effective features for new applications is slow



Motivation

- Features are the main factor impacting accuracies.
- Multitude of hand-designed features currently in use.
  - Reply on human domain knowledge much more than data
  - Feature design is separate from training the classifier
  - If handcrafted features have multiple parameters, it is hard to manually tune them
  - Developing effective features for new applications is slow

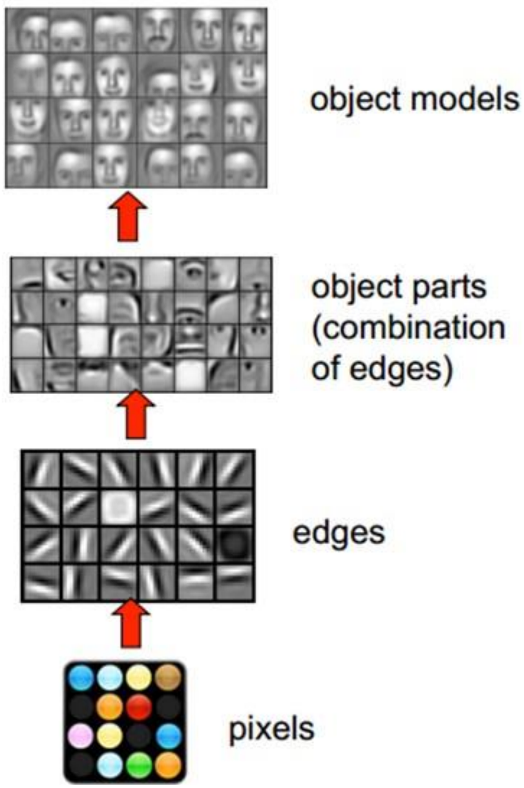
Learning features directly from data may be another way to **improve** the **quality** of features.

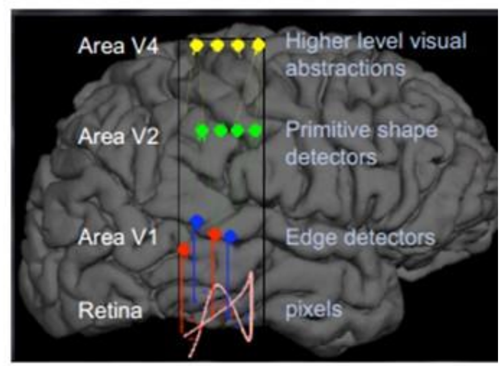
Feature Learning

- Learning transformations of the data that make it easier to extract useful information when building classifiers or predictors
  - Jointly learning feature transformations and classifiers makes their integration optimal
  - Learn the values of a huge number of parameters in feature representations, **which dramatically increase the capacity of deep models**
  - Make better use of big data
  - Faster to get feature representations for new applications

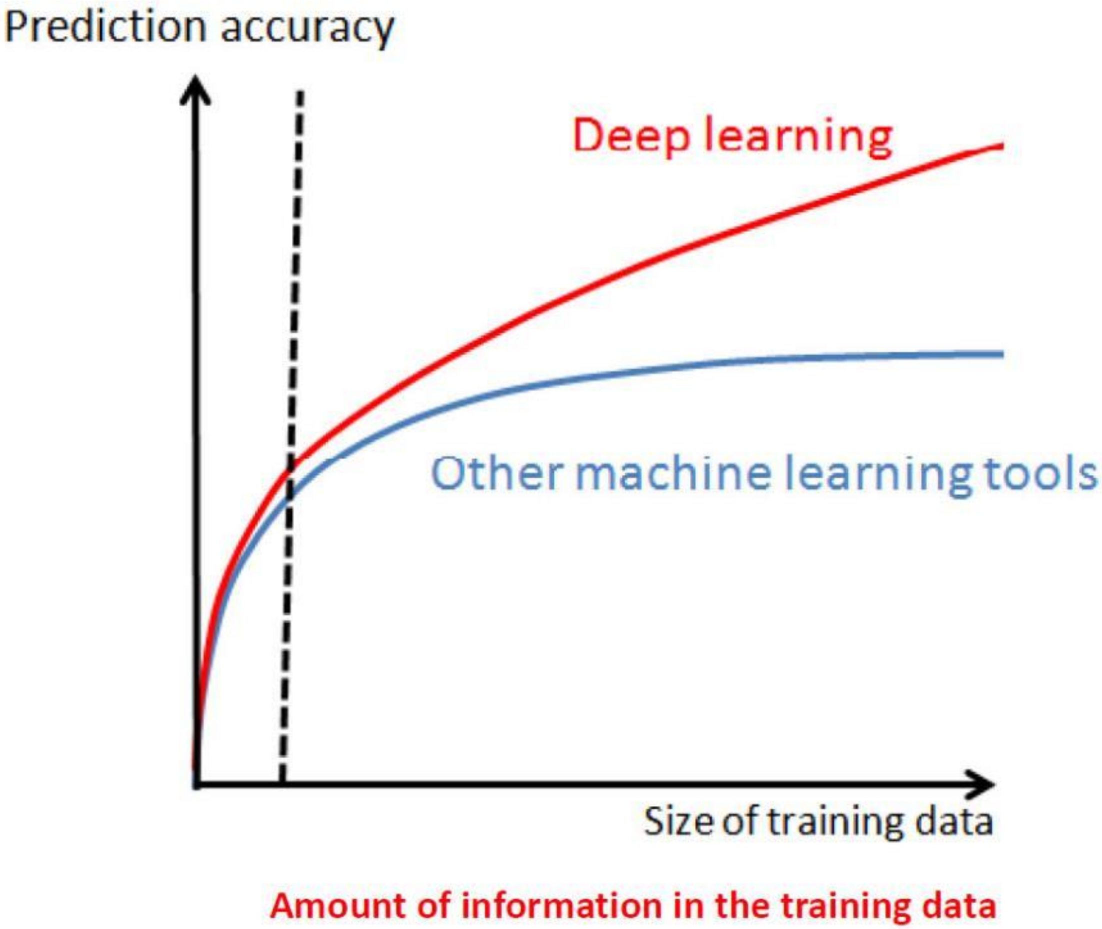
Deep Learning Means Feature Learning

- Deep learning methods aim to
  - Learning feature hierarchies
  - where features from higher levels of the hierarchy are formed by lower level features



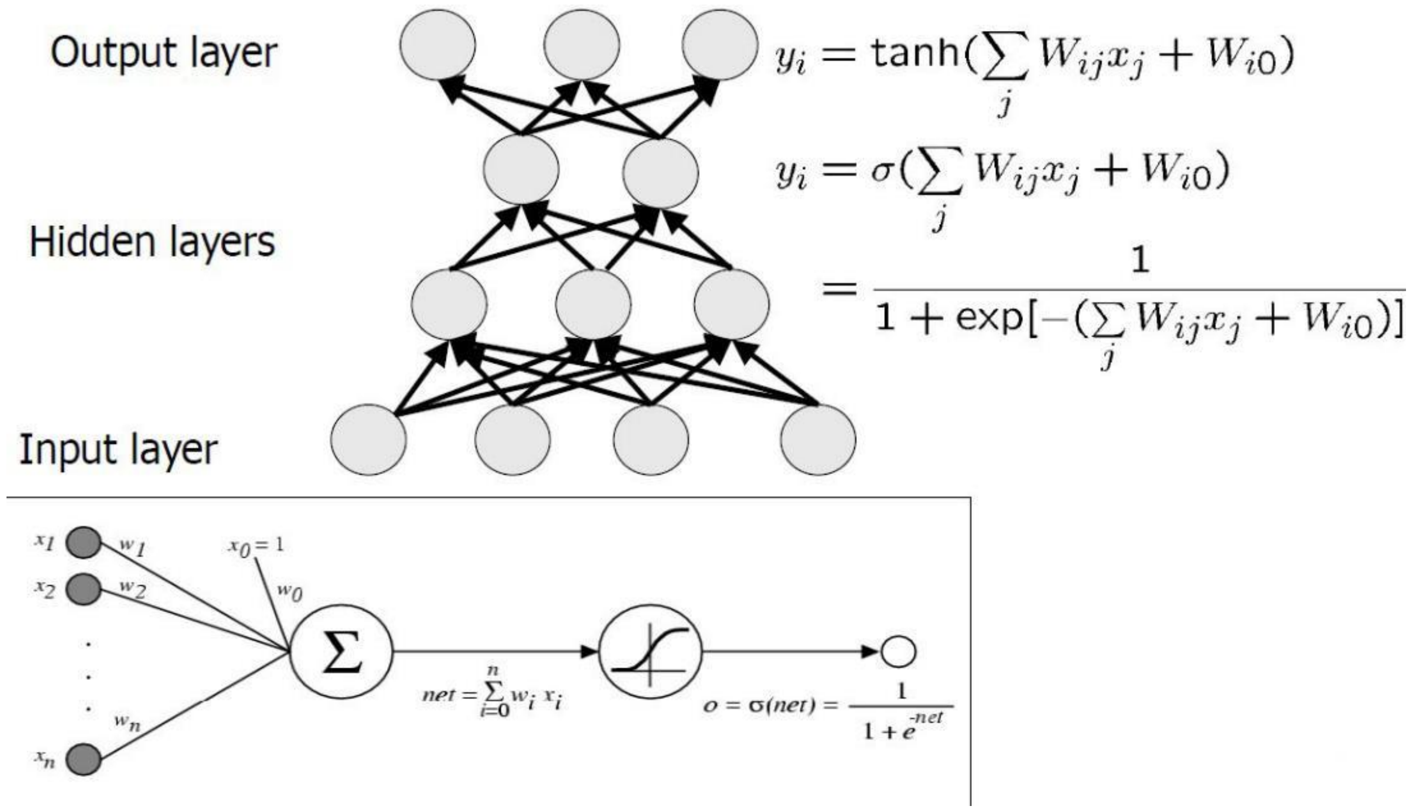


Deep Learning vs others



Deep architectures

- **Definition**: Deep architectures are composed of multiple levels of non-linear operations, such as neural nets with many hidden layers.



Introduction on Classical Deep Models

- Deep Belief Net (DBN)
  - G. E. Hinton, S. Osindero, and Y. Teh, "A Fast Learning Algorithm for Deep Belief Nets," Neural Computation, Vol. 18, pp. 1527-1544, 2006.
- Auto-encoder
  - G. E. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," Science, Vol. 313, pp. 504-507, July 2006.
- Convolutional Neural Networks (CNN)
  - Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based Learning Applied to Document Recognition," Proceedings of the IEEE, Vol. 86, pp. 2278-2324, 1998.

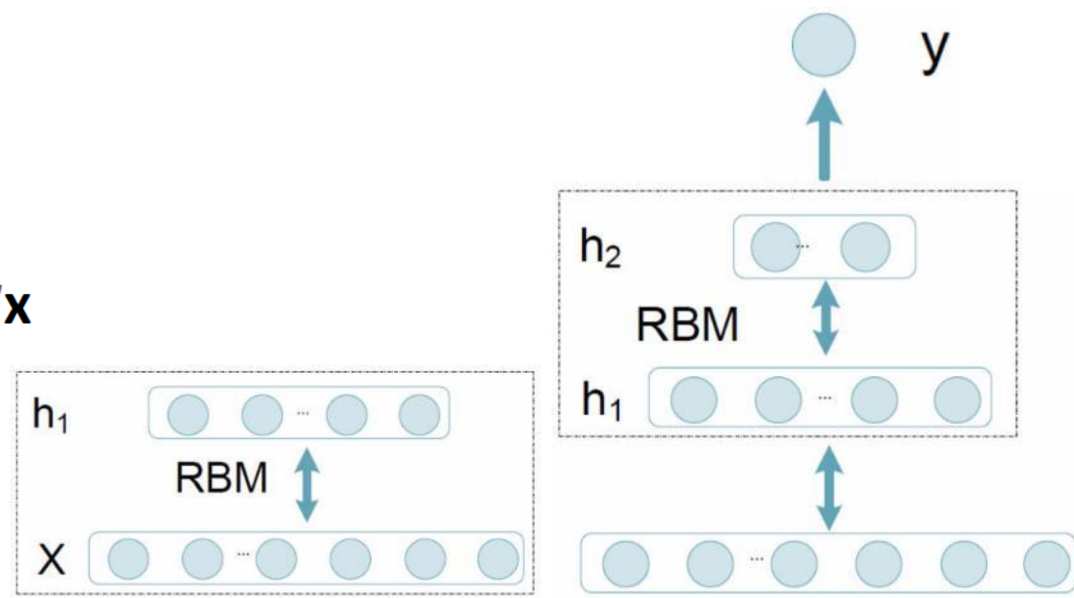
Classical Deep Models

- Deep Belief Net
  - Hinton'06

$$\mathrm{P}(\mathbf{x},\mathbf{h}_1,\mathbf{h}_2) = \mathrm{p}(\mathbf{x}|\mathbf{h}_1)\ \mathrm{p}(\mathbf{h}_1,\mathbf{h}_2)$$

$$P(\mathbf{x},\mathbf{h}_1) = \frac{e^{-E(\mathbf{x},\mathbf{h}_1)}}{\sum_{\mathbf{x},\mathbf{h}_1} e^{-E(\mathbf{x},\mathbf{h}_1)}}$$

$$\mathrm{E}(\mathbf{x},\mathbf{h}_1)=\mathbf{b}'\,\mathbf{x}+\mathbf{c}'\,\mathbf{h}_1+\mathbf{h}_1'\,\mathbf{W}\mathbf{x}$$



Classical Deep Models

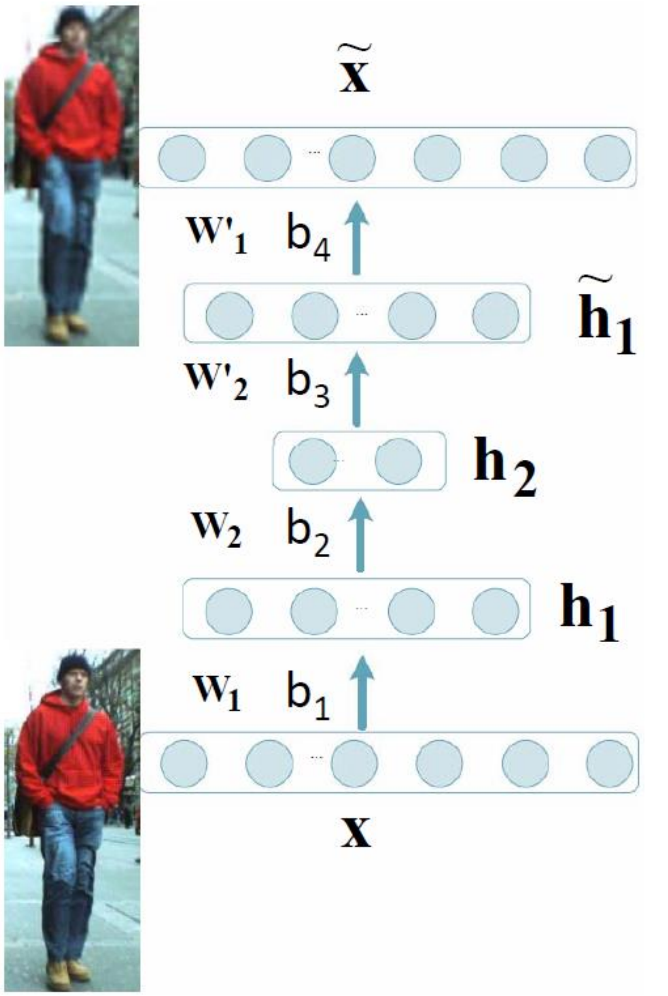
- Auto-encoder
  - Hinton and Salakhutdinov 2006

Encoding: $\mathbf{h}_1 = \sigma(\mathbf{W}_1\mathbf{x}+b_1)$

$\mathbf{h}_2 = \sigma(\mathbf{W}_2\mathbf{h}_1+b_2)$

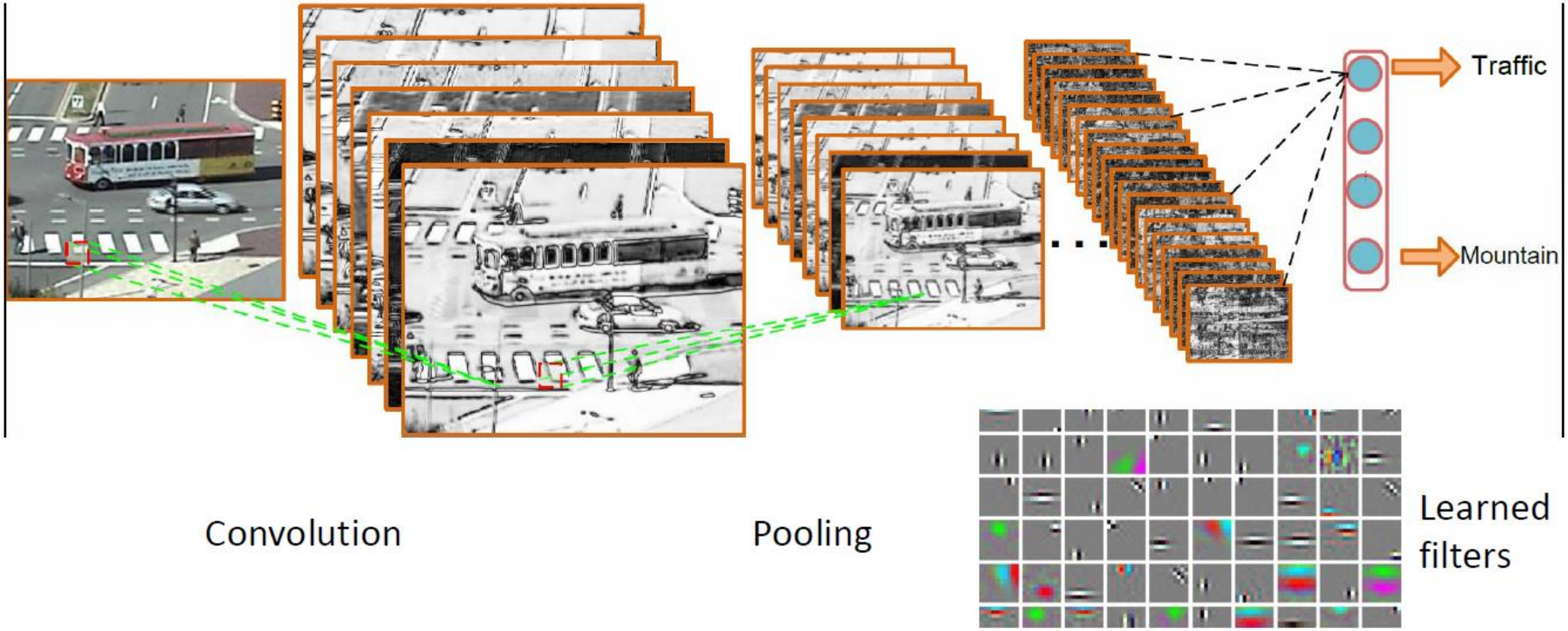
Decoding: $\tilde{\mathbf{h}}_1 = \sigma(\mathbf{W}'_2\mathbf{h}_2+b_3)$

$\tilde{\mathbf{x}} = \sigma(\mathbf{W}'_1\mathbf{h}_1+b_4)$



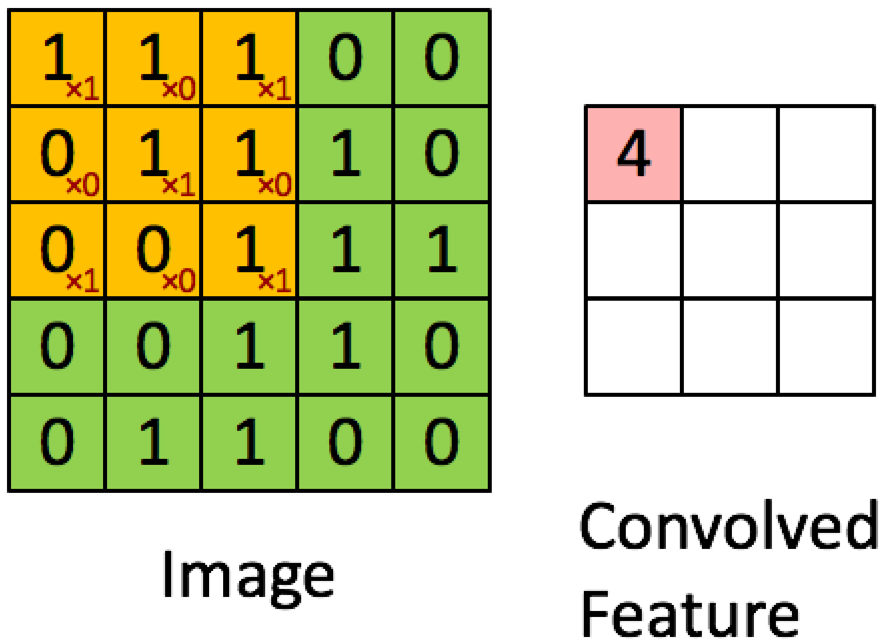
Classical Deep Models

- Convolutional Neural Networks (CNN)
  - First proposed by Fukushima in 1980
  - Improved by LeCun, Bottou, Bengio and Haffner in 1998



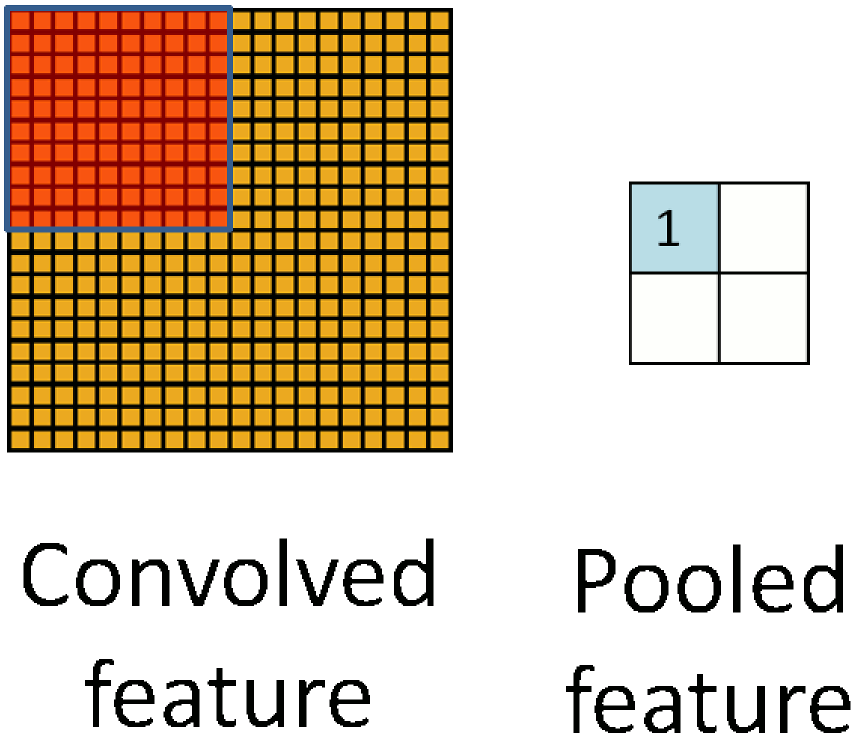
Convolution (Filter)

- 图像可以看作一个二维空间的离散函数，表示为$f(x,y)$，假设有对于二维卷积操作函数$C(u,v)$，则会产生输出图像$g(x,y) = f(x,y) * C(u,v)$



Ref: http://t.cn/zYmWQGa , access this for viewing gif.

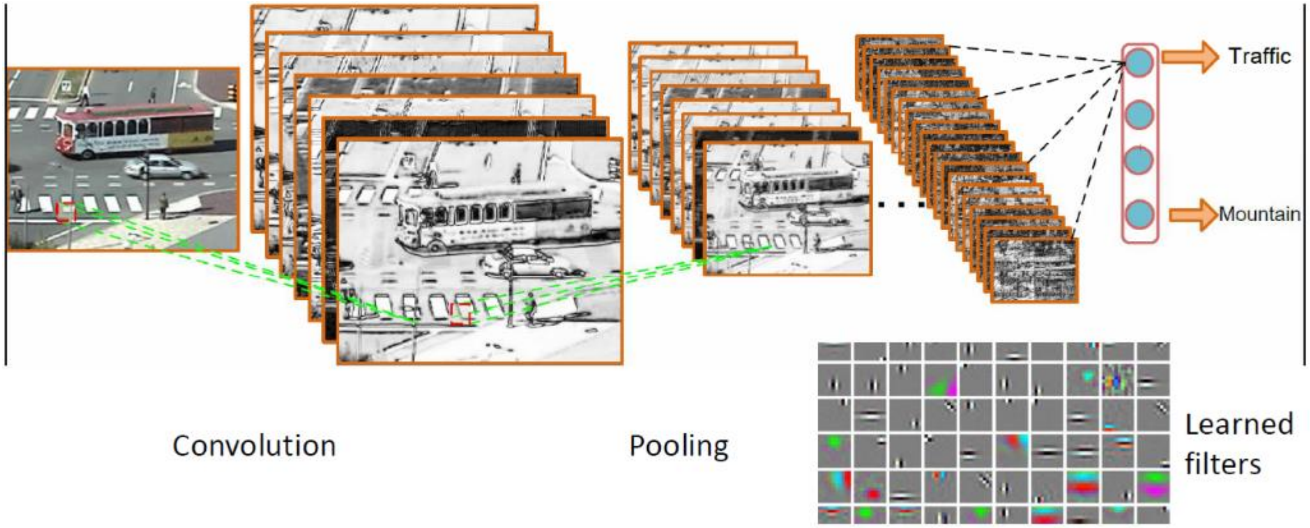
Pooling



Ref: http://t.cn/zYmWQGa , access this for viewing gif

Convolutional Neural Networks

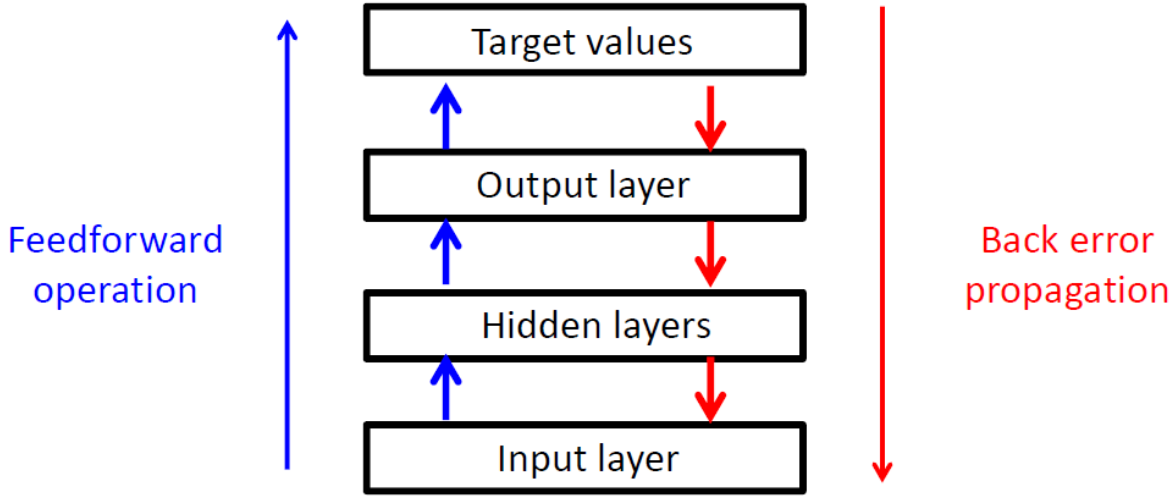
- There are four key ideas behind CNNs that take advantage of the properties of natural signals:
  - Local connection
  - Shared weights
  - Pooling
  - The use of many layers



Backpropagation

$$\mathbf{W} \leftarrow \mathbf{W} - \eta \bigtriangledown J(\mathbf{W})$$

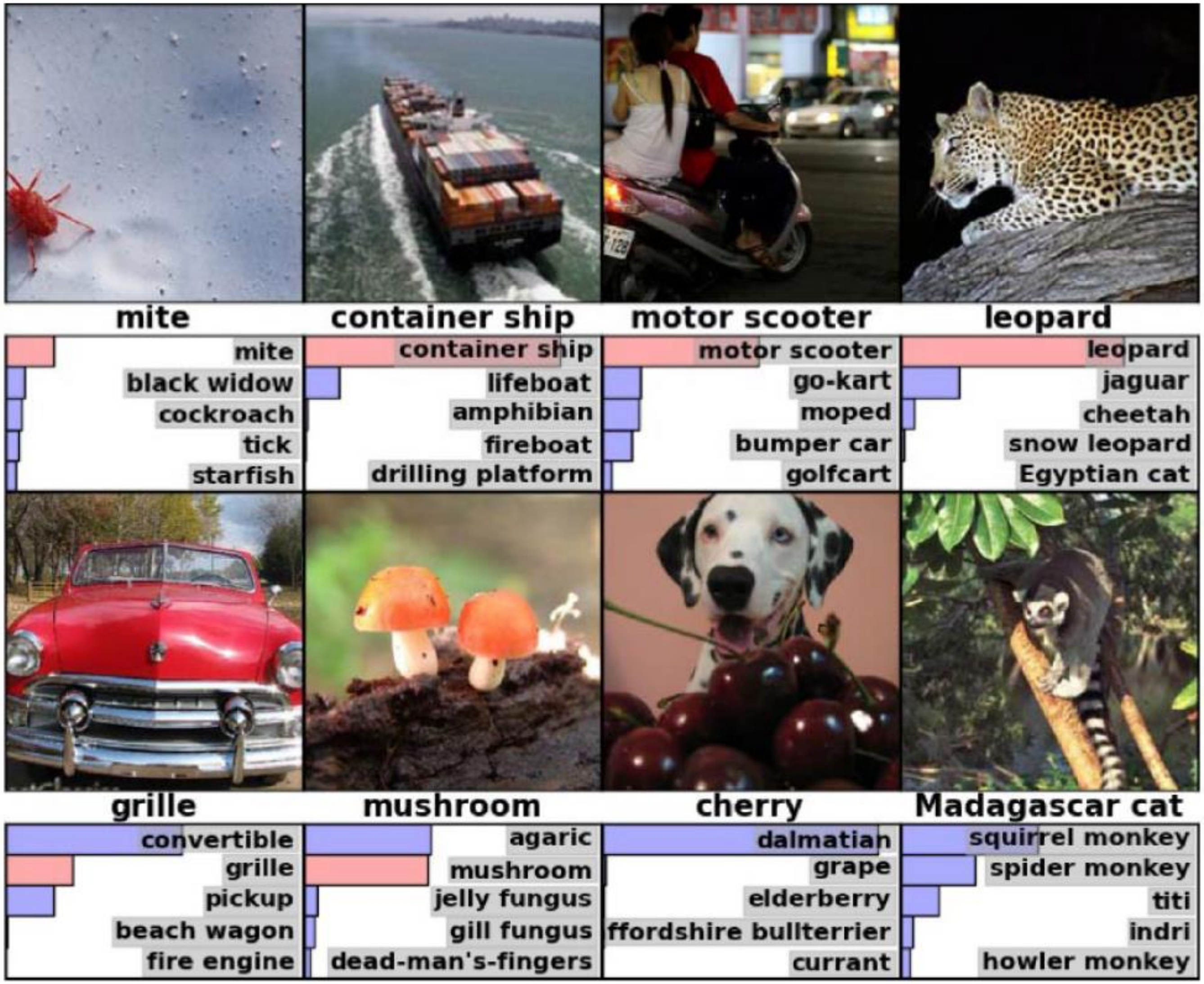
$\mathbf{W}$ is the parameter of the network; J is the objective function



D. E. Rumelhart, G. E. Hinton, R. J. Williams, "Learning Representations by Back-propagation Errors," Nature, Vol. 323, pp. 533-536, 1986.

Applications of Deep Learning

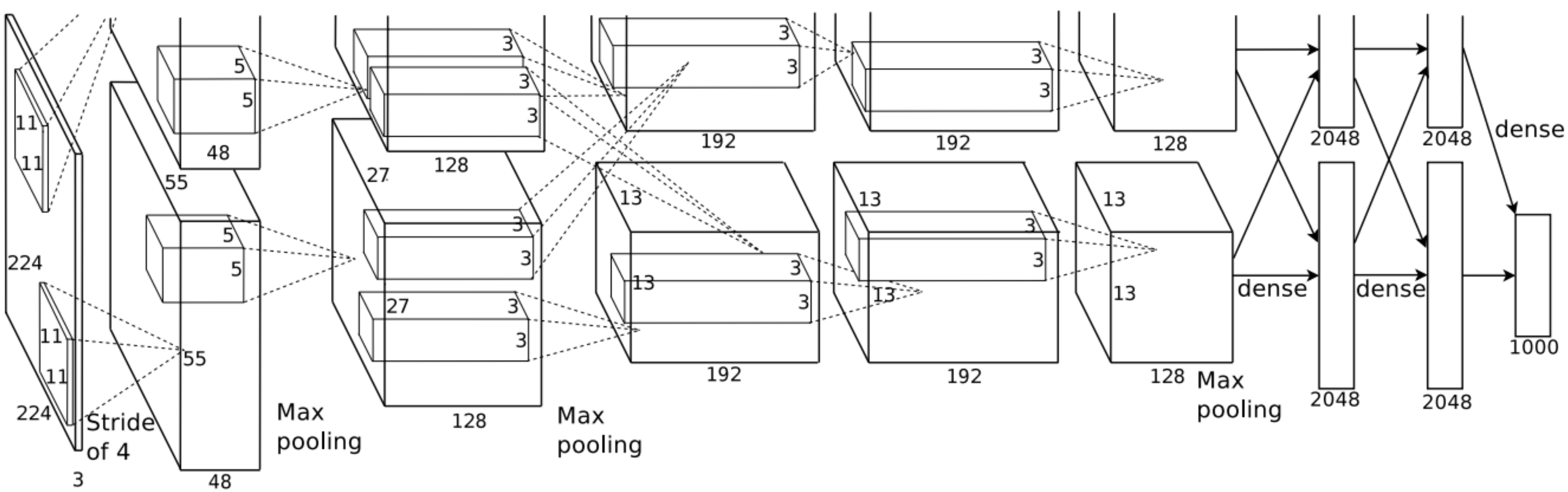
Image Classification



Ref: ImageNet Classification with Deep Convolutional Neural Networks

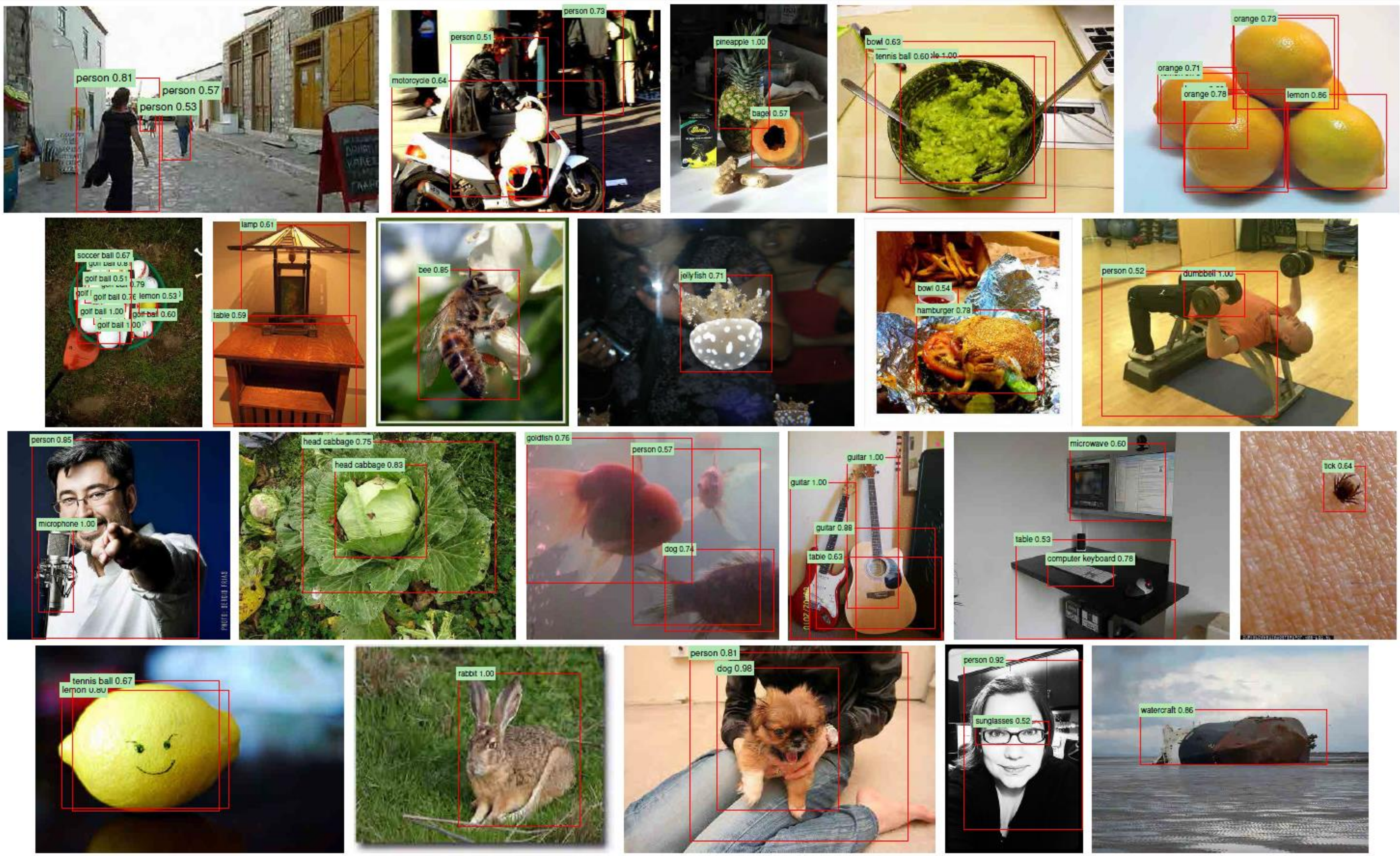
Image Classification

- Hinton group's groundbreaking work on ImageNet
  - They did not have much experience on general image classification on ImageNet
  - It took one week to train the network with 60 Million parameters



Ref: ImageNet Classification with Deep Convolutional Neural Networks

Object Detection

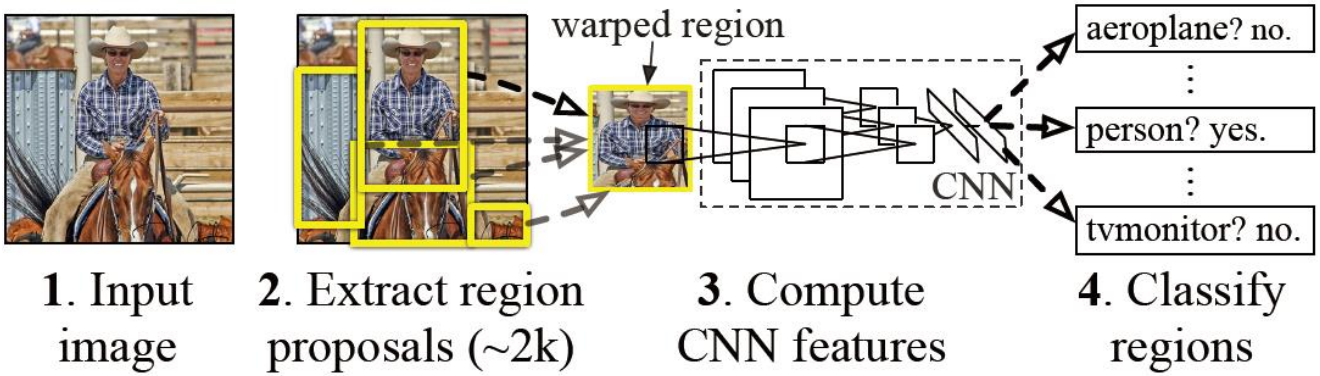


Ref: Rich feature hierarchies for accurate object detection and semantic segmentation

Object Detection

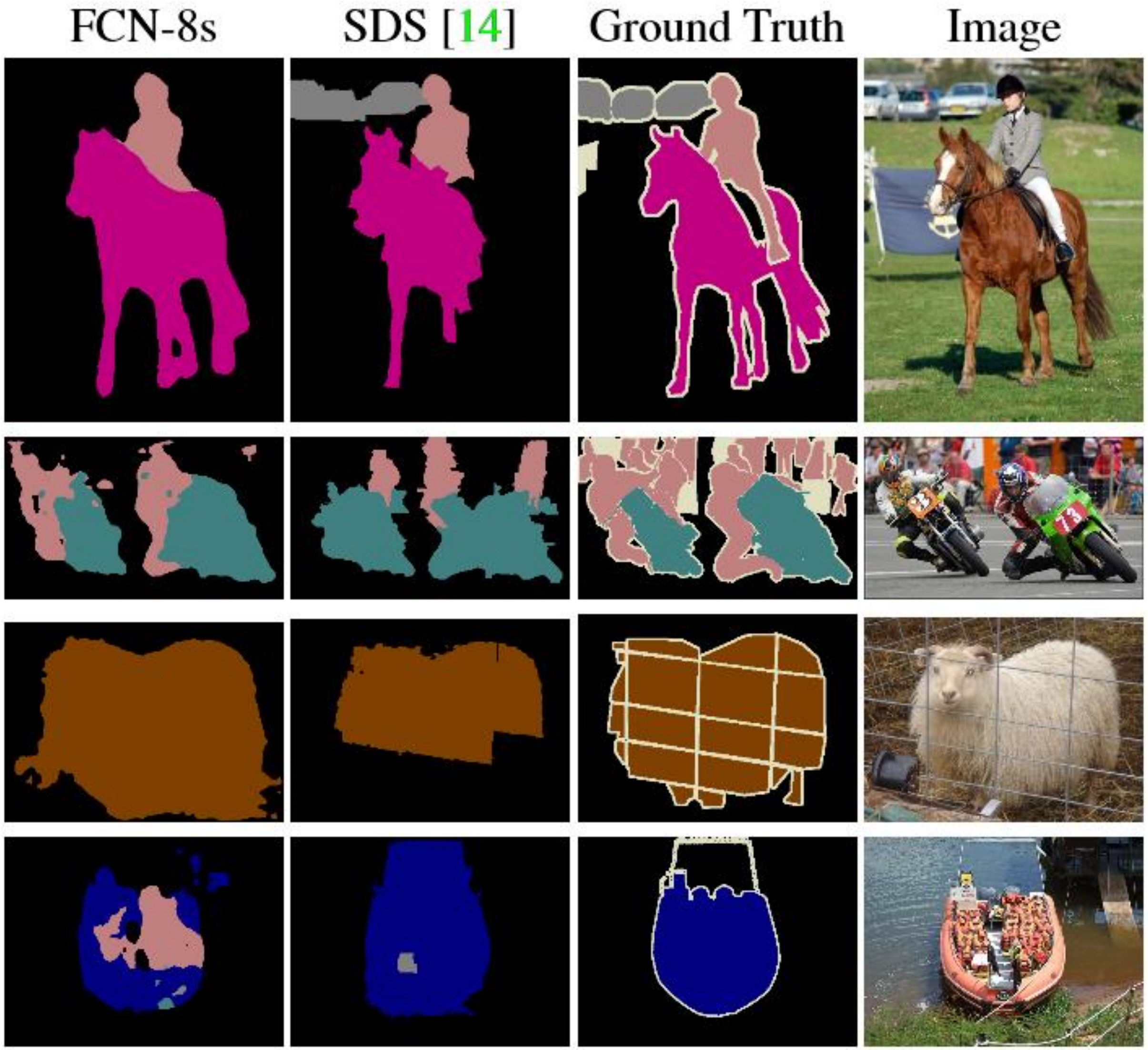
- R-CNN
  - Step 1. Input an image
  - Step 2. Use selective search to obtain ~2k proposals
  - Step 3. Warp each proposal and apply CNN to extract its features
  - Step 4. Adopt class-specified SVM to score each proposal
  - Step 5. Rank the proposals and use NMS to get the bboxes.
  - Step 6. Use class-specified regressors to refine the bboxes' positions.

R-CNN: *Regions with CNN features*

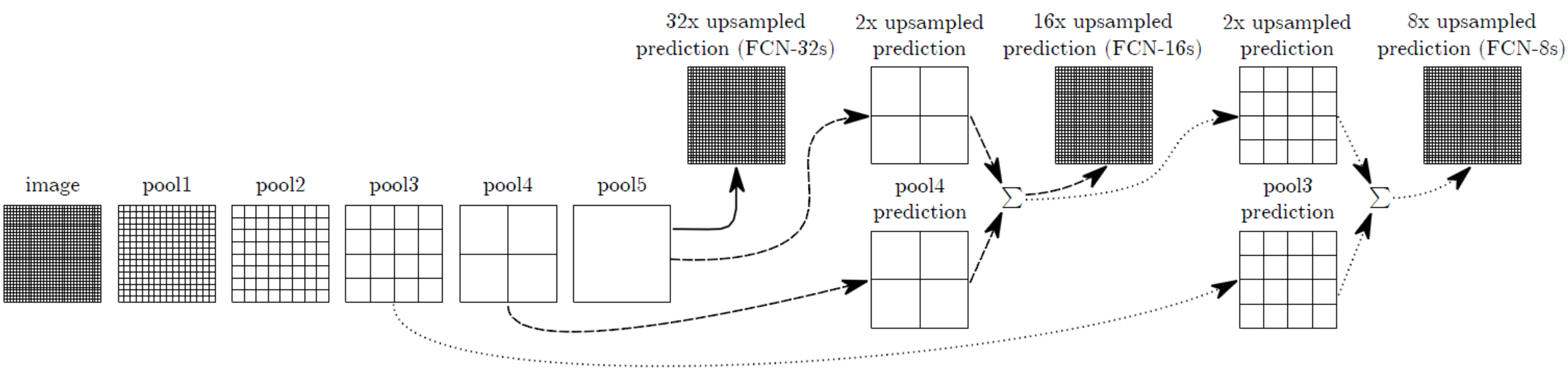


Ref: Rich feature hierarchies for accurate object detection and semantic segmentation

Segmentation

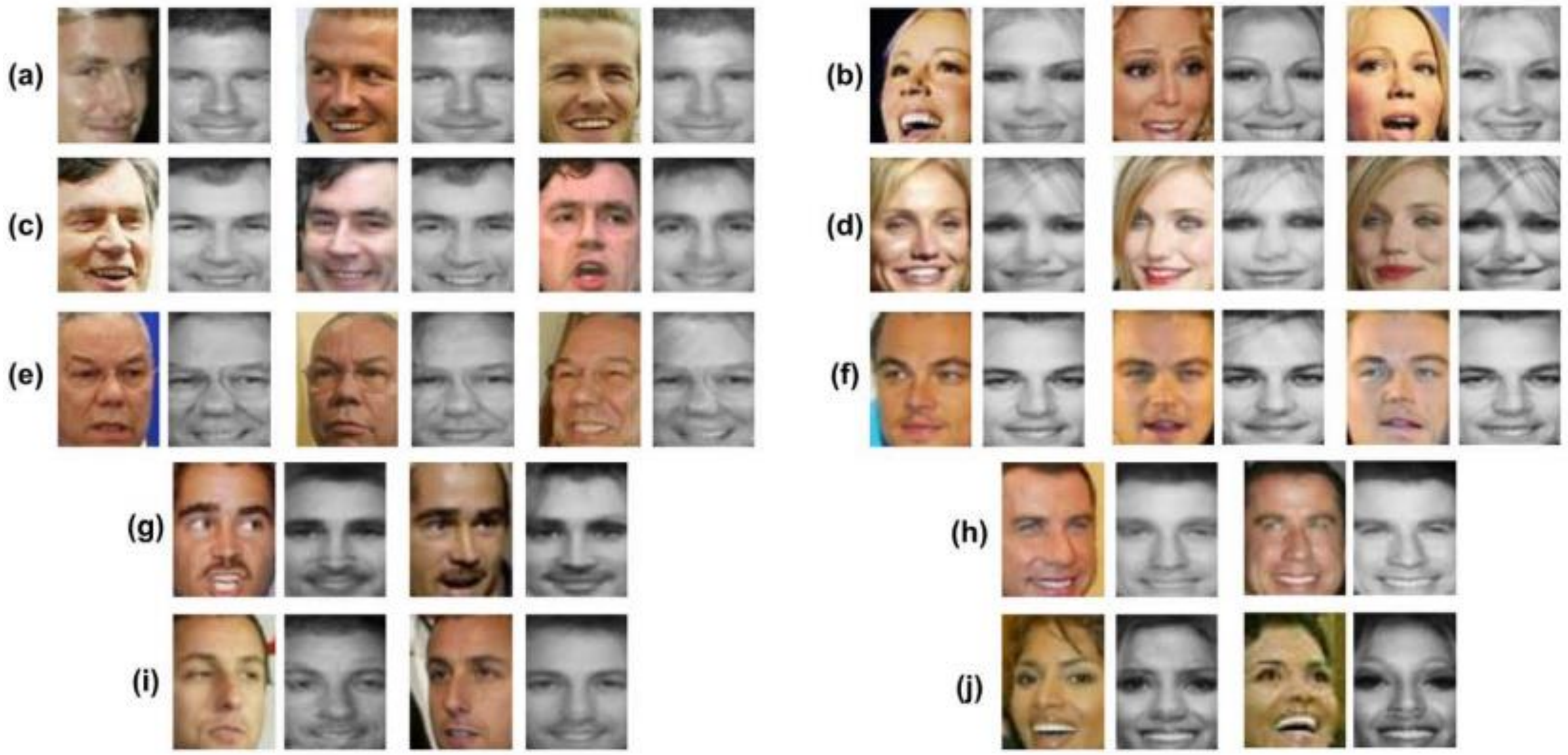


Segmentation



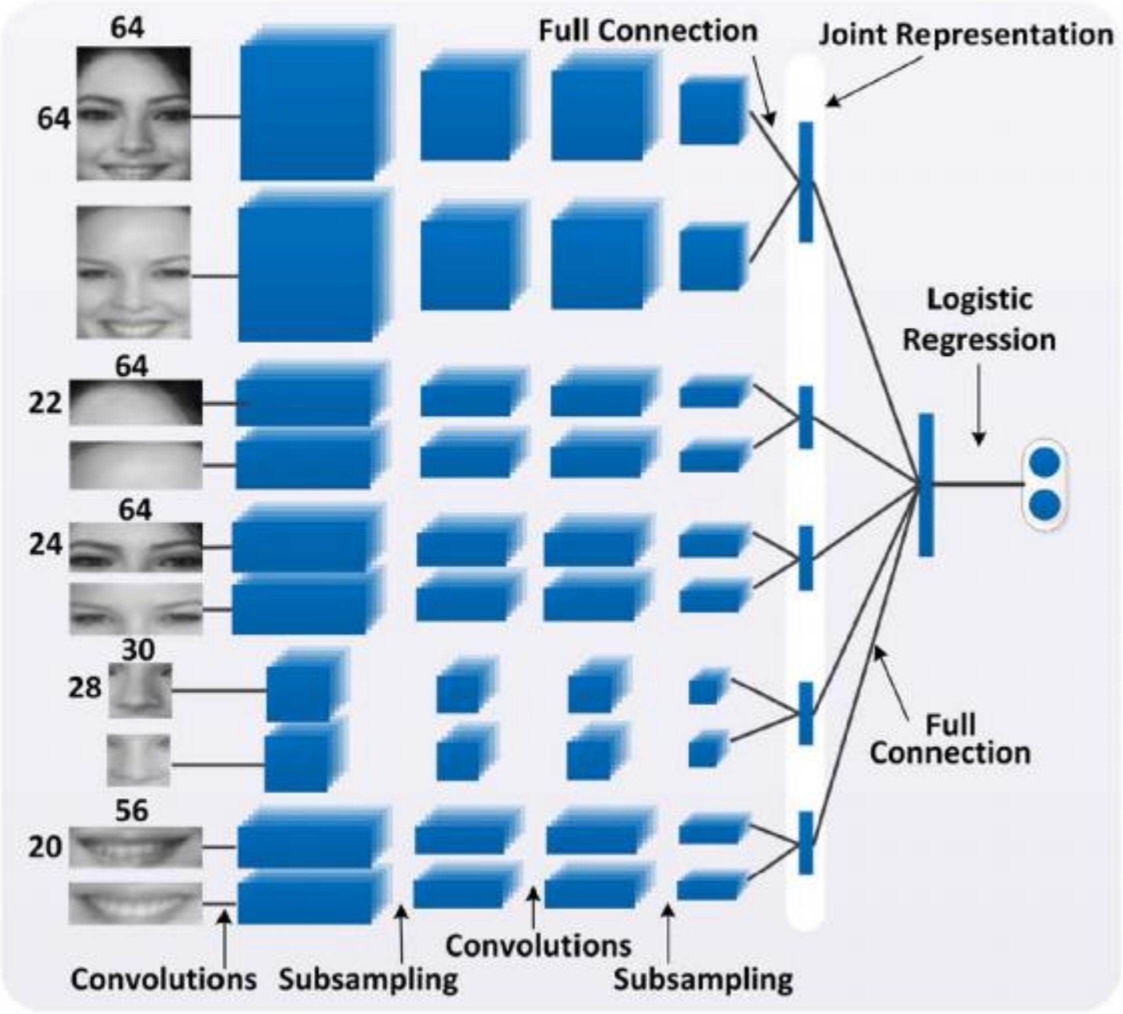
Ref: Fully Convolutional Networks for Semantic Segmentation

Recover Canonical-View Faces



Ref: Recover Canonical-View Faces in the wild with Deep Neural Networks

Recover Canonical-View Faces

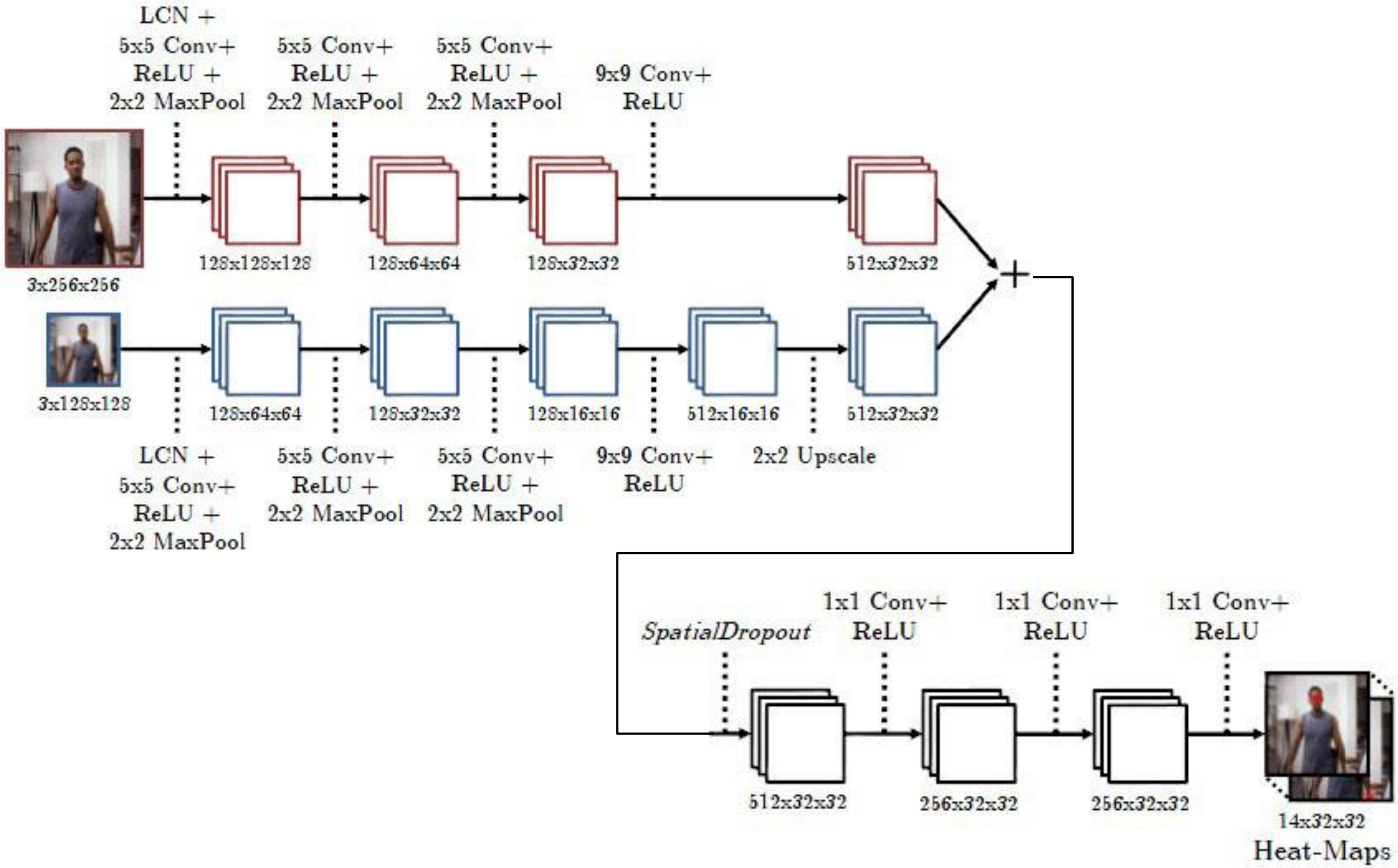


Ref: Recover Canonical-View Faces in the wild with Deep Neural Networks

Pose Estimation

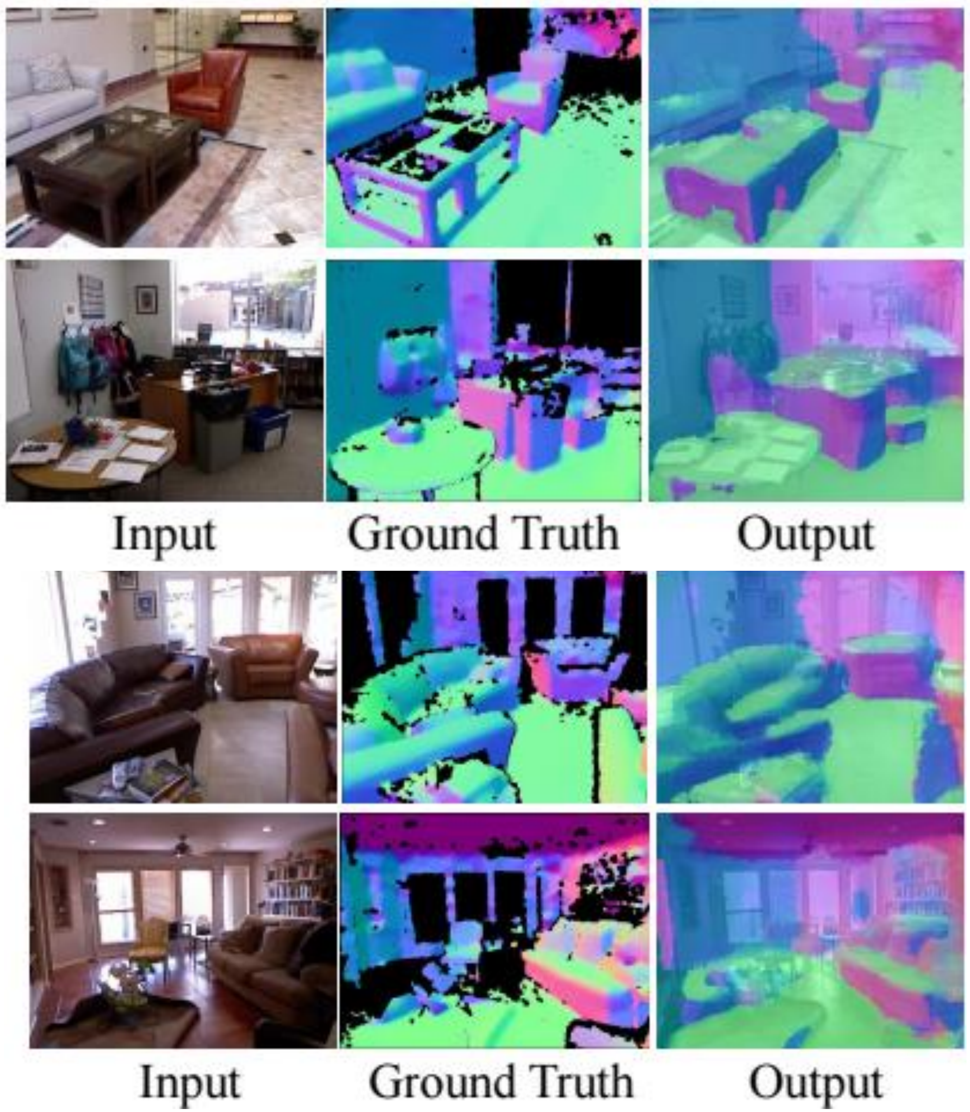
Ref: Efficient Object Localization Using Convolutional Networks

Pose Estimation



Ref: Efficient Object Localization Using Convolutional Networks

Surface Normal Estimation



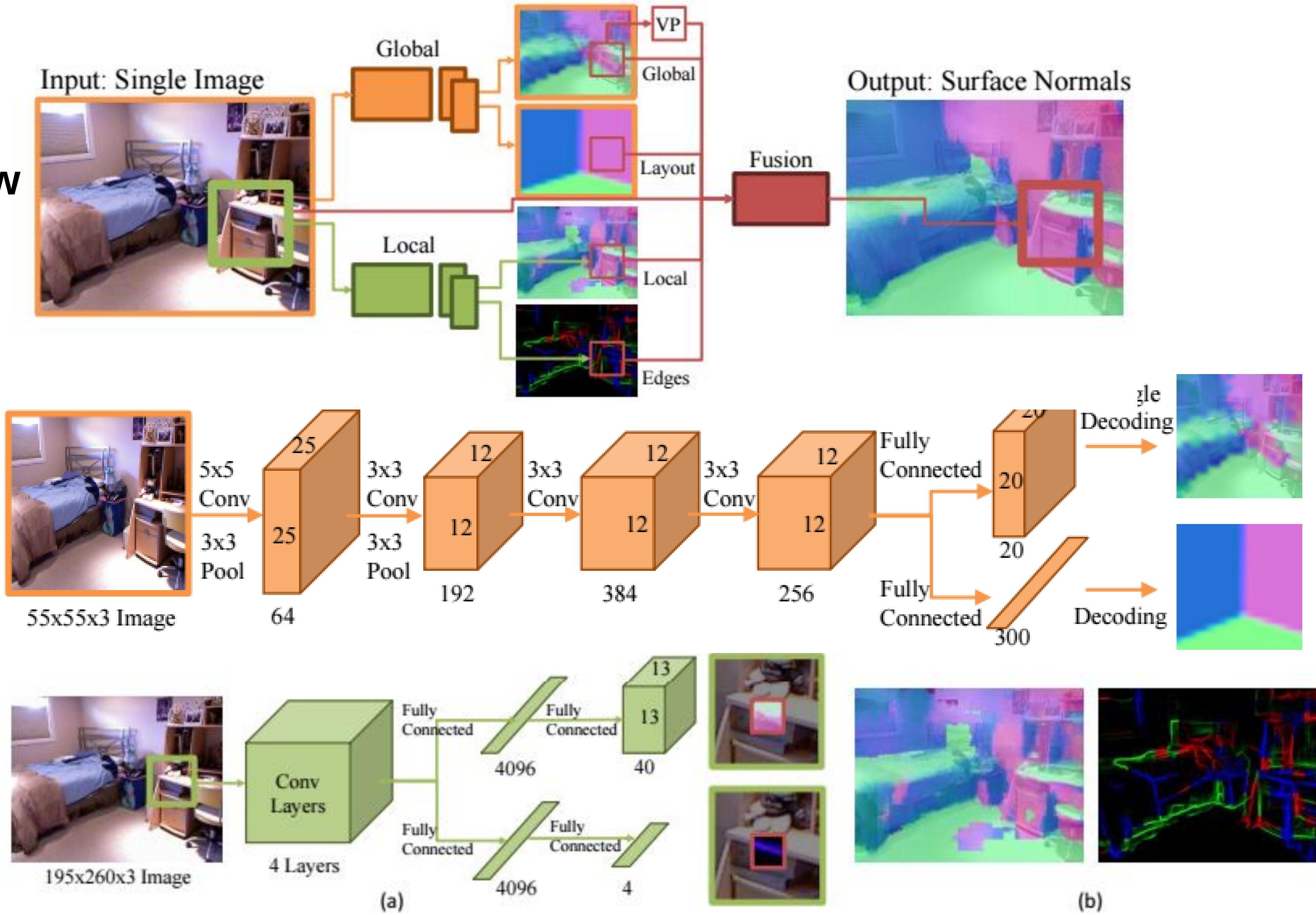
Ref: Designing Deep Networks for Surface Normal Estimation

Surface Normal Estimation

Overview

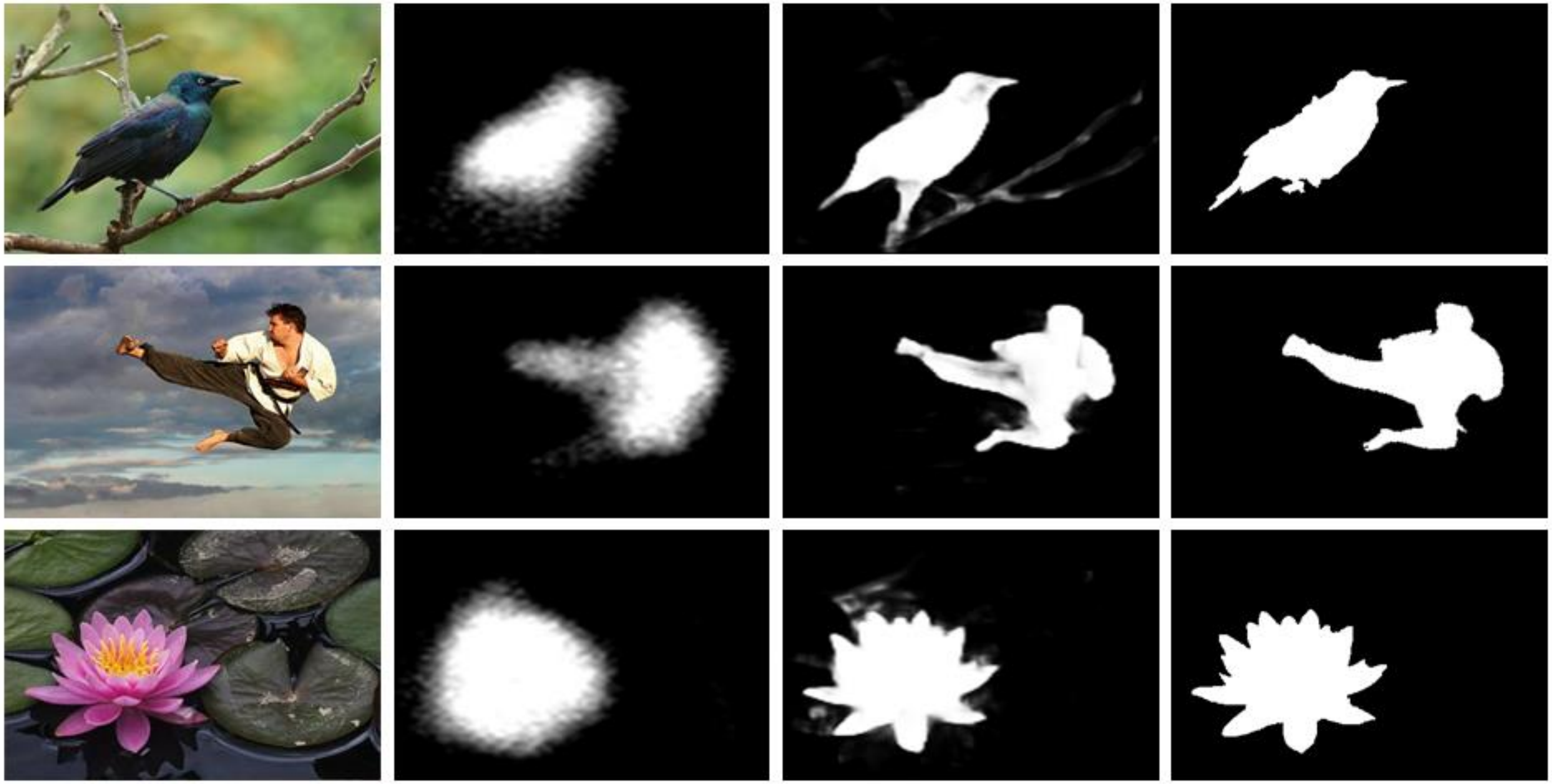
Global

Local



Ref: Designing Deep Networks for Surface Normal Estimation

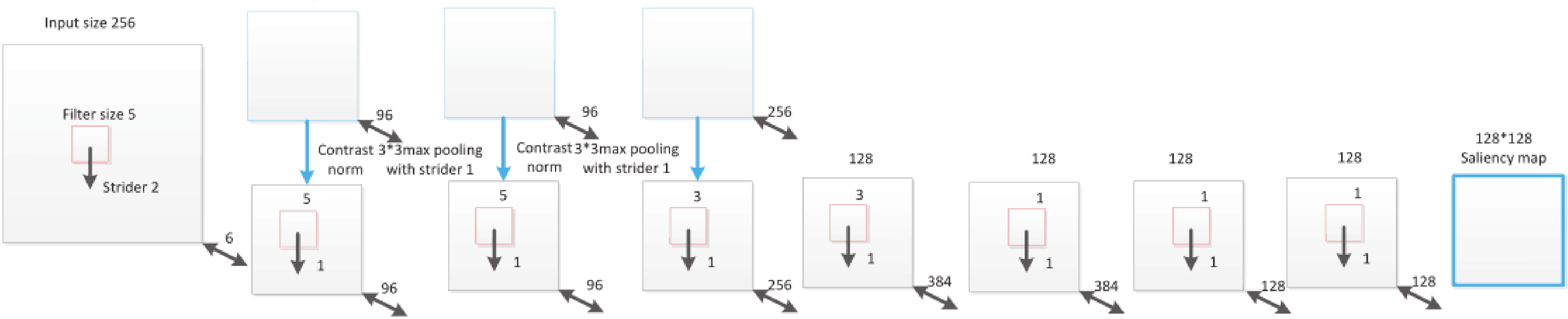
Saliency Detection



(a) Original image (b) Coarse saliency map (c) Fine saliency map (d) Ground truth

Ref: Deep Image Saliency Computing via Progressive Representation Learning

Saliency Detection



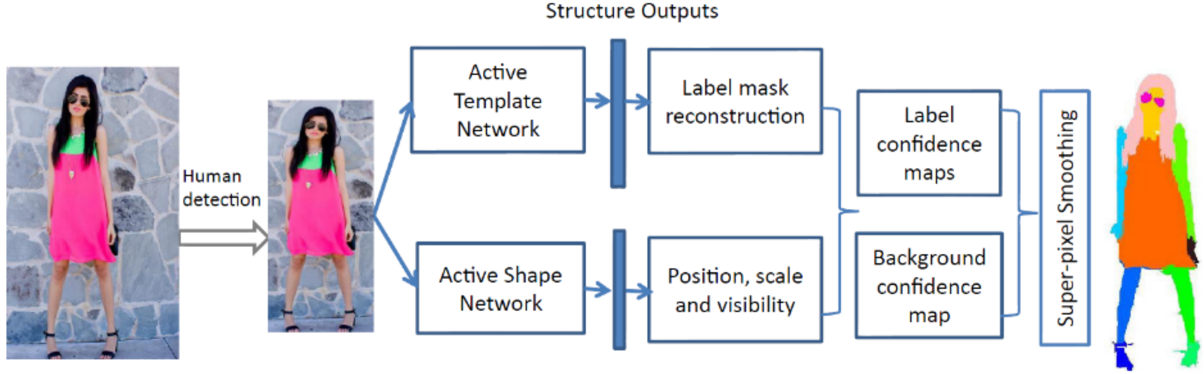
Ref: Deep Image Saliency Computing via Progressive Representation Learning

Human Parsing

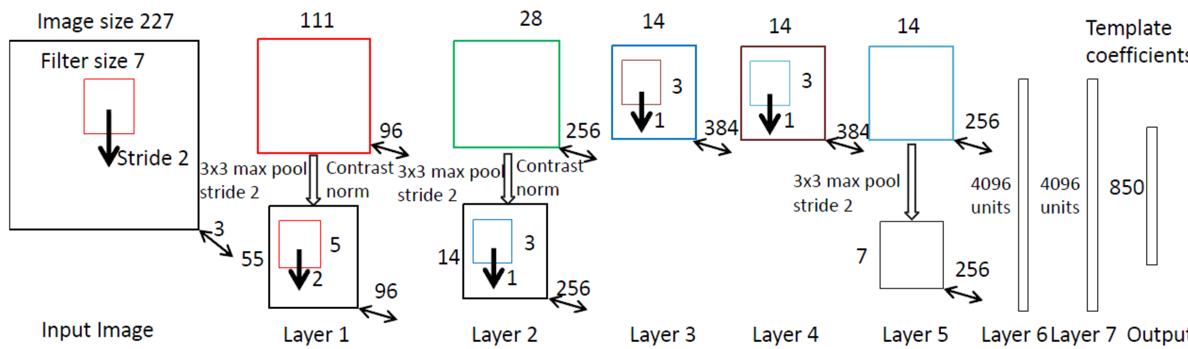
Ref: Deep Human Parsing with Active Template Regression

Human Parsing

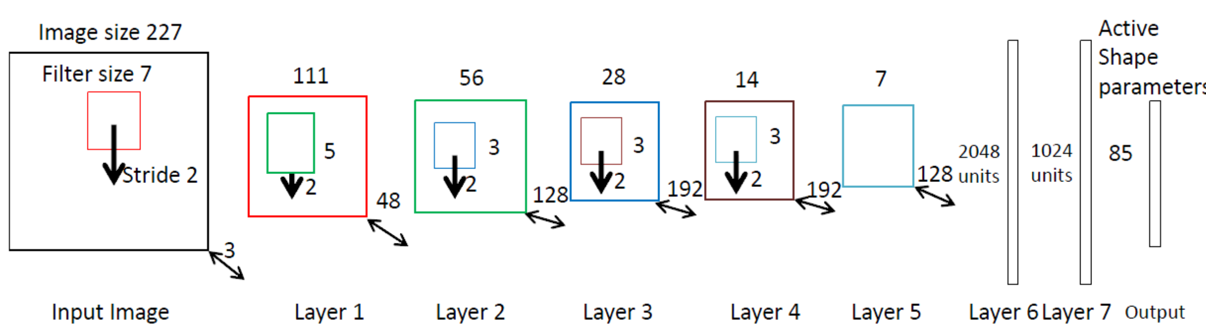
Framework



Active Shape Network



Active Template Network

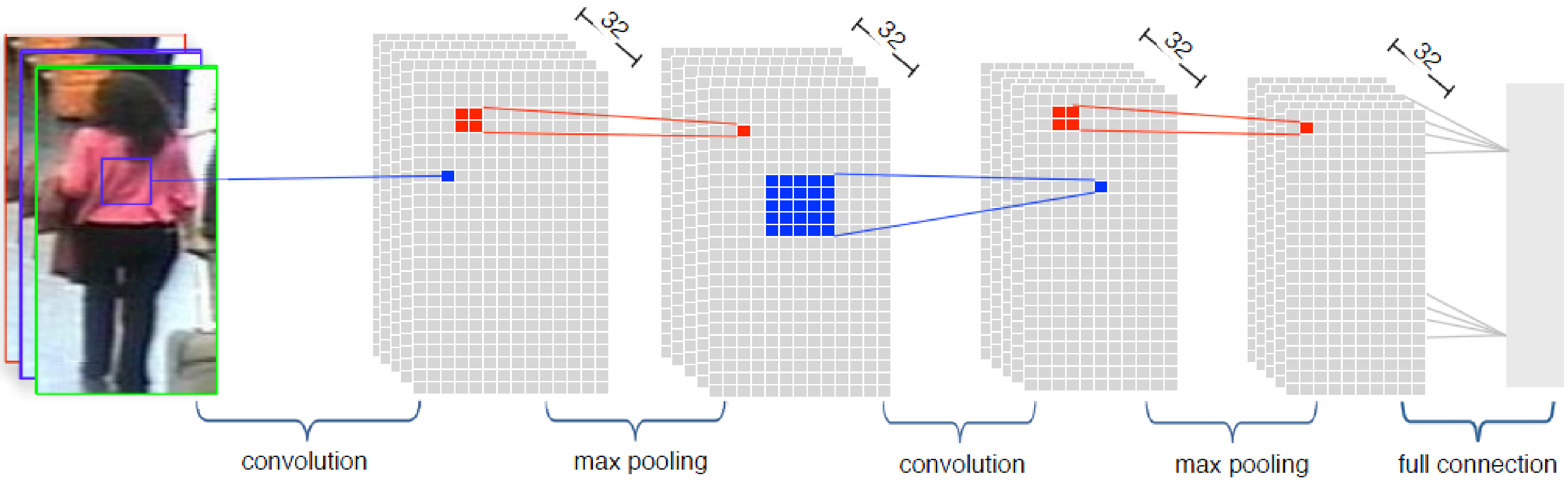


Ref: Deep Human Parsing with Active Template Regression

Person Re-identification

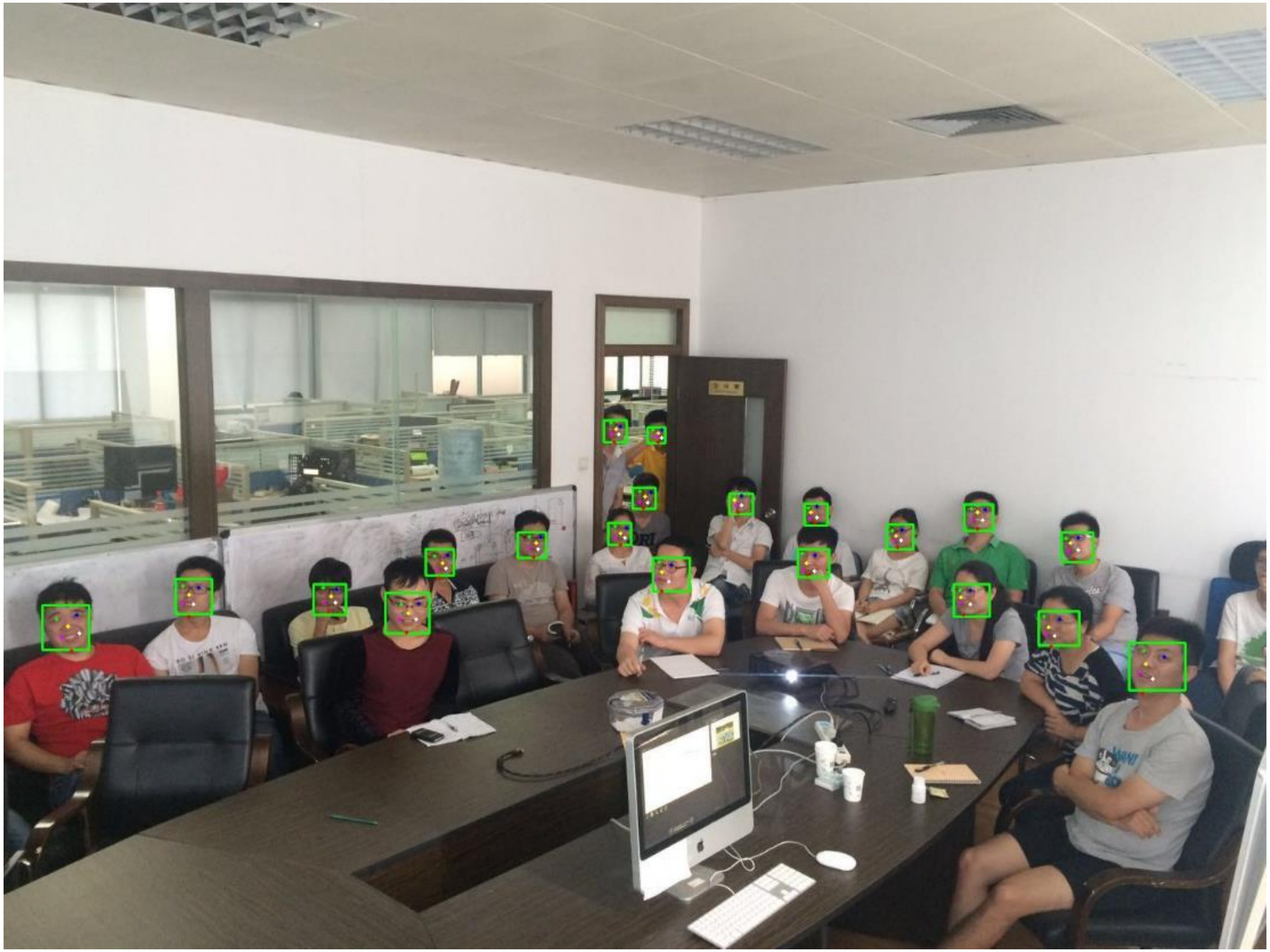
Ref: Deep Feature Learning with Relative Distance Comparison for Person Re-identication

Person Re-identification



Ref: Deep feature learning with relative distance comparison for person re-identification

Face Detection and Landmark Localization

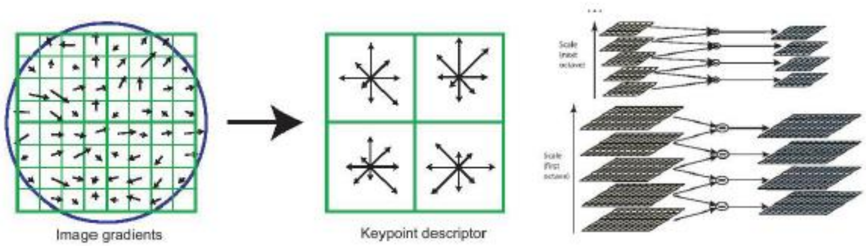


Thank you!

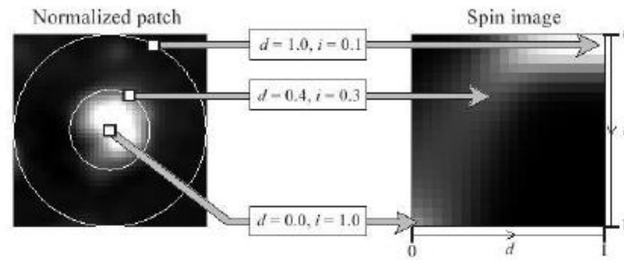
Slides can be downloaded from

Email: alfredtofu@gmail.com
Homepage: http://cvmarcher.com

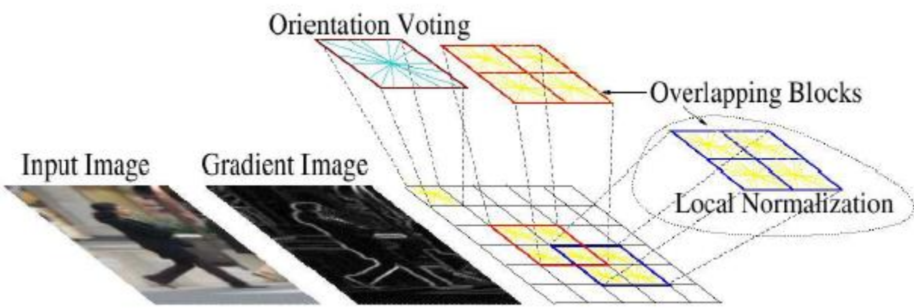
Computer vision features



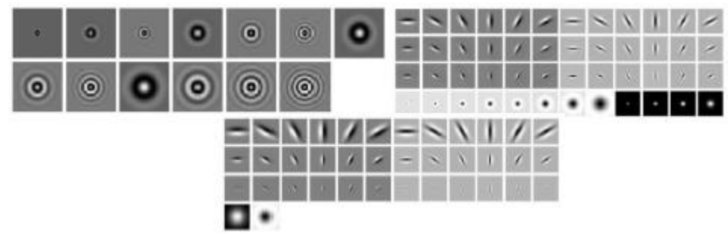
SIFT



Spin image



HoG



Textons

and many others:

SURF, MSER, LBP, Color-SIFT, Color histogram, GLOH, …..

Live Demo

Updated after my talk

Live Demo

Updated after my talk

Live Demo

Updated after my talk

Live Demo

Updated after my talk