



CRIPAC

智能感知与计算研究中心  
Center for Research on Intelligent Perception and Computing



中国科学院自动化研究所  
Institute of Automation  
Chinese Academy of Sciences

2018

## “Deep Learning Lecture”

# Lecture 1 : Introduction

Wang Liang

Center for Research on Intelligent Perception and Computing (CRIPAC)  
National Laboratory of Pattern Recognition (NLPR)  
Institute of Automation, Chinese Academy of Science (CASIA)

# Outline

---

**1** Course Information

**2** What is Deep Learning

**3** Applications of Deep Learning

**4** Future Directions

# Course Goals

---

- Give you an opportunity for you to explore an interesting multivariate analysis problem of your choice in the context of a real-world data set
- How deep learning works
- How to frame tasks into learning problems
- How to use toolkits to implement designed models
- When and why specific deep learning techniques work for specific problems

# Course Prerequisites

---

- Calculus and Linear Algebra
- Probability and Statistics
- Machine Learning
- Advanced programming language (like Python and C)
- Time and patience

# Course Overview

---

- 1. Introduction
- 2. Mathematical Basics
- 3. Feedforward Network
- 4. Convolutional Neural Network 卷积神经网络
- 5. Recurrent Neural Network 递归神经网络
- 6. Regularization and Optimization 正则化与最优化
- 7. Deep Generative Models 深度生成模型
- 8. Reinforcement Learning 强化学习
- 9. Attention and Memory
- 10. RBM, DBN and DBM
- 11. Graphical Model
- 12. DL Application
- 13. Project Presentation

# Course Logistics

---

- Course: Theory (3 hours per week) + Course project
- The location of course room: 教1-101
- Final grade = 50% from the group project + 50% from the final exam
- The group project contains a technical report and project presentation
- The project presentation is decided by random sampling due to time limit
- Each team is composed of 5-10 members

# Course Logistics

---

- We strongly encourage students to form study groups. Students may discuss and work on project problems in groups.
- However, each student must write down the solutions independently, and without referring to written notes from the joint session. In other words, each student must understand the solution well enough in order to reconstruct it by him/herself.
- In addition, each student should write on the problem set the bunch of people with whom s/he collaborated.

# Course Logistics

---

Projects will be evaluated based on:

1. The technical quality of the work. (I.e., Does the technical material make sense? Are the things tried reasonable? Are the proposed algorithms or applications clever and interesting? Do the authors convey novel insight about the problem and/or algorithms?)
2. The completeness and novelty of the work, and the clarity of the write-up. (Spare enough time for the write-up since it may be harder than you imagine)

# Course Logistics

---

Your final report is expected to be a 4~8 page report. You should submit both an electronic and a hardcopy version for your final report. It should roughly have the following format:

- **Introduction - Motivation**
- **Problem definition**
- **Proposed method**
  - Intuition - why should it be better than other methods?
  - Description of its algorithms
- **Experiments**
  - Description of your testbed; list of questions your experiments are designed to answer
  - Details of the experiments; observations
- **Conclusions**

# Course Logistics

---

- Write final reports in the form of CVPR papers:  
[http://cvpr2017.thecvf.com/submission/main\\_conference/author\\_guidelines](http://cvpr2017.thecvf.com/submission/main_conference/author_guidelines) (using Latex, which is very useful for your writing papers in later research life)
- Provide corresponding codes, and a “readme.txt” file to illustrate how to run the codes

# People



Professor: Liang Wang

Email: [wangliang@nlpr.ia.ac.cn](mailto:wangliang@nlpr.ia.ac.cn)



Associate Professor: Wei Wang

Email: [wangwei@nlpr.ia.ac.cn](mailto:wangwei@nlpr.ia.ac.cn)



Teaching Assistant: Junbo Wang , PhD candidate

Email : [junbo.wang@nlpr.ia.ac.cn](mailto:junbo.wang@nlpr.ia.ac.cn)

Homepage: <http://www.cripac.ia.ac.cn/CN/column/column149.shtml>

# Our Group

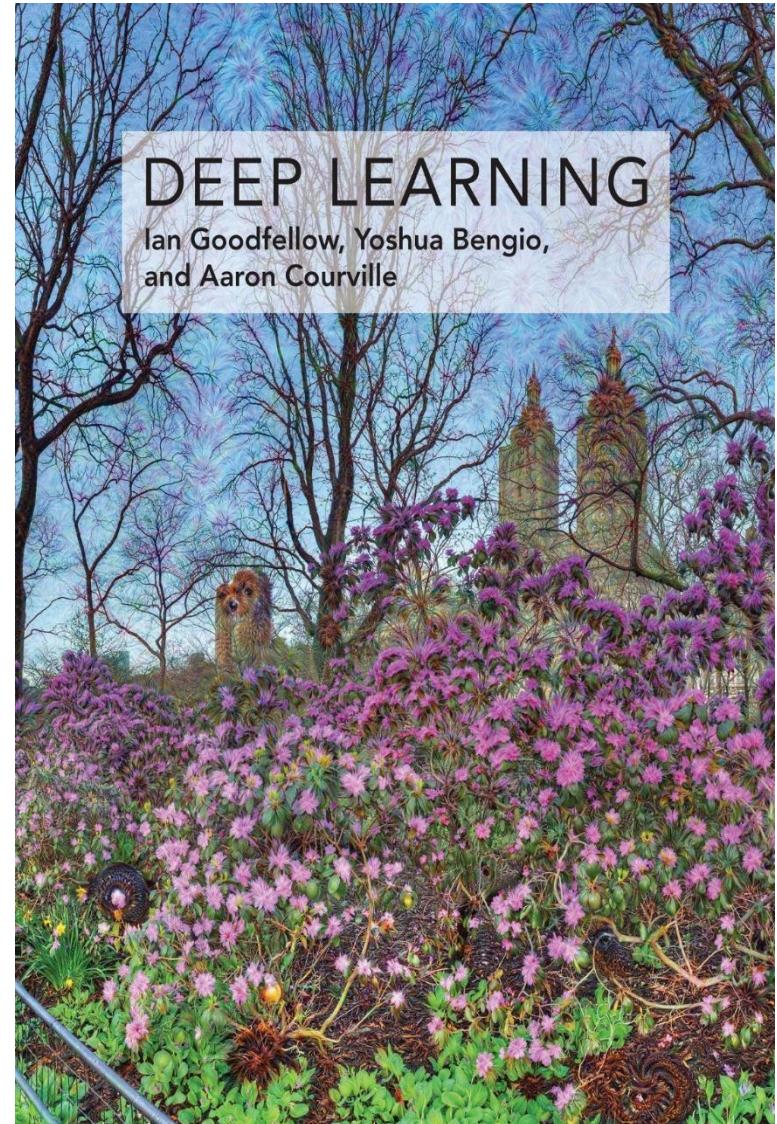


Email: [multimodal\\_comp@126.com](mailto:multimodal_comp@126.com)

# Reference book

---

- [http://www.deeplearningbook.org/lecture\\_slides.html](http://www.deeplearningbook.org/lecture_slides.html)
- [https://github.com/goodfeli/dlbook\\_exercises](https://github.com/goodfeli/dlbook_exercises)



# Some Resources: Programming

---

- <http://scikit-learn.org/stable/> (Scikit Learn: a Python library for machine learning)
- <http://ubcmatlabguide.github.io/> (YAGTOM: Yet Another Guide TO Matlab)
- <https://github.com/saitjr/C-Toturials> (Learning C)

# Some Resources: Deep Learning Tools

Name	Language	Link	Note
Pylearn2	Python	<a href="http://deeplearning.net/software/pylearn2/">http://deeplearning.net/software/pylearn2/</a>	A machine learning library built on Theano
Theano	Python	<a href="http://deeplearning.net/software/theano/">http://deeplearning.net/software/theano/</a>	A python deep learning library
Caffe	C++	<a href="http://caffe.berkeleyvision.org/">http://caffe.berkeleyvision.org/</a>	A deep learning framework by Berkeley
Torch	Lua	<a href="http://torch.ch/">http://torch.ch/</a>	An open source machine learning framework
Overfeat	Lua	<a href="http://cilvr.nyu.edu/doku.php?id=code:start">http://cilvr.nyu.edu/doku.php?id=code:start</a>	A convolutional network image processor
Deeplearning4j	Java	<a href="http://deeplearning4j.org/">http://deeplearning4j.org/</a>	A commercial grade deep learning library
Word2vec	C	<a href="https://code.google.com/p/word2vec/">https://code.google.com/p/word2vec/</a>	Word embedding framework
GloVe	C	<a href="http://nlp.stanford.edu/projects/glove/">http://nlp.stanford.edu/projects/glove/</a>	Word embedding framework
Doc2vec	C	<a href="https://radimrehurek.com/gensim/models/doc2vec.html">https://radimrehurek.com/gensim/models/doc2vec.html</a>	Language model for paragraphs and documents
StanfordNLP	Java	<a href="http://nlp.stanford.edu/">http://nlp.stanford.edu/</a>	A deep learning-based NLP package
TensorFlow	Python	<a href="http://www.tensorflow.org">http://www.tensorflow.org</a>	A deep learning based python library

# Some Resources: Conferences

---

A very good project will comprise a **publishable or nearly-publishable** piece of work. So, for inspiration, you might also look at some recent machine learning and artificial intelligence research papers.

Main machine learning and artificial intelligence conferences are ICML, NIPS, CVPR, AAAI and IJCAI. You can find papers from the following websites:

<https://icml.cc/Conferences/2017/Schedule> (ICML17)

<https://nips.cc/Conferences/2017/Schedule> (NIPS17)

<http://openaccess.thecvf.com/CVPR2017.py> (CVPR17)

<https://www.aaai.org/Library/AAAI/aaai17contents.php> (AAAI17)

<https://ijcai-17.org/accepted-papers.html> (IJCAI17)

# Outline

---

**1** Course Information

**2** What is Deep Learning

**3** Applications of Deep Learning

**4** Future Directions

# Deep Learning

## ARTIFICIAL INTELLIGENCE

Any technique that enables computers to mimic human behavior



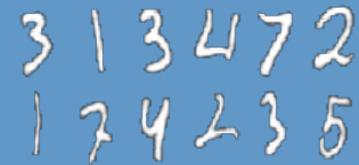
## MACHINE LEARNING

Ability to learn without explicitly being programmed



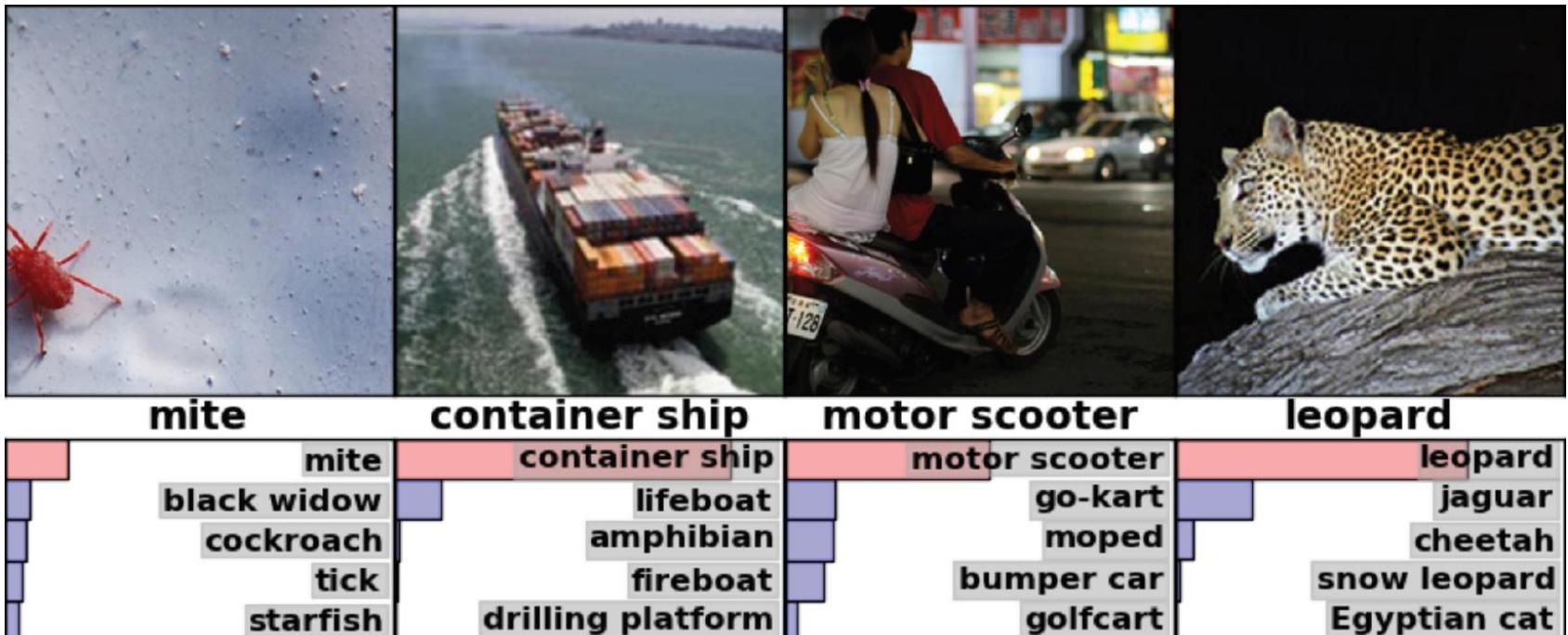
## DEEP LEARNING

Learn underlying features in data using neural networks



# Deep Learning Success for Vision

## Image Recognition



# Deep Learning Success for Vision

---

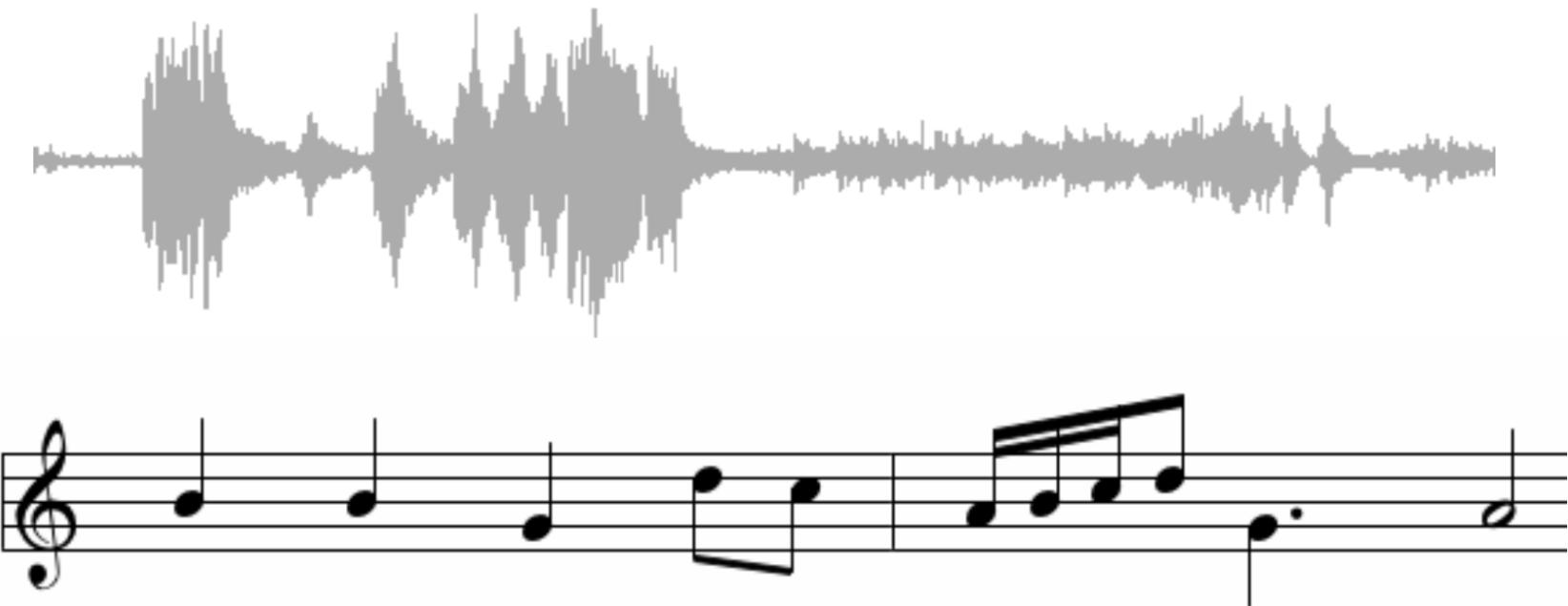
Detect pneumothorax in real X-Ray scans



# Deep Learning Success for Audio

---

## Music Generation

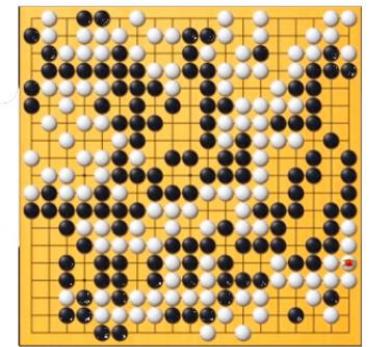


# Deep Learning Success

So many more ...

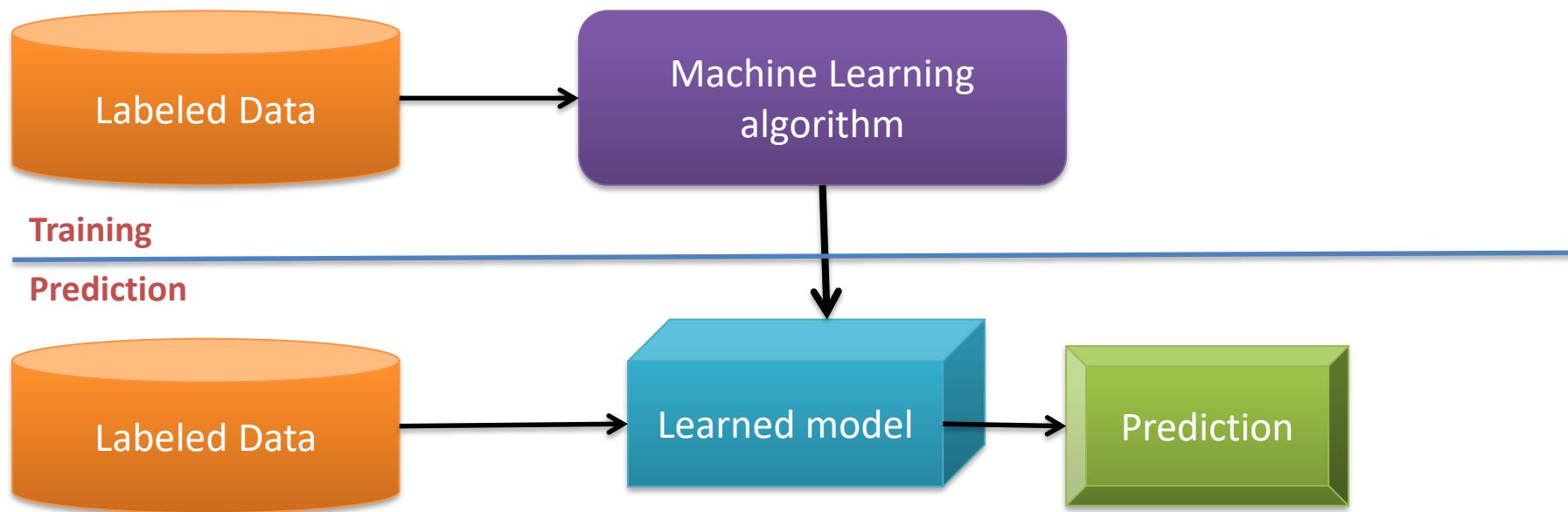


AlphaGo  
ZERO



# Machine Learning Basics

Machine learning is a field of computer science that gives computers the ability to **learn without being explicitly programmed**



Methods that can learn from and make predictions on data

# Types of Learning

有监督

**Supervised:** Learning with a **labeled training** set

Example: email *classification* with already labeled emails

无监督

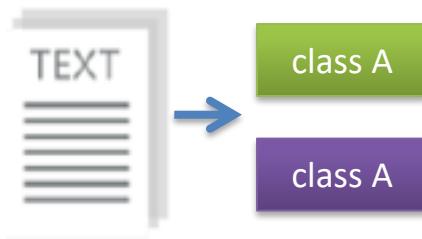
**Unsupervised:** Discover **patterns** in **unlabeled** data

Example: *cluster* similar documents based on text

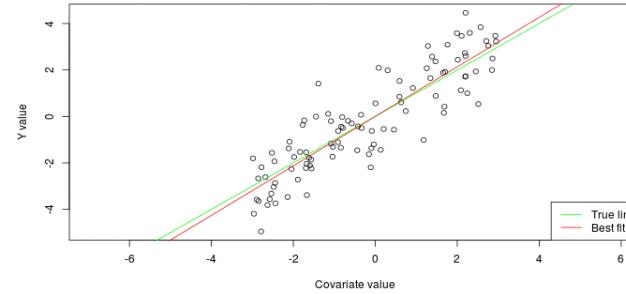
强化学习

**Reinforcement learning:** learn to **act** based on **feedback/reward**

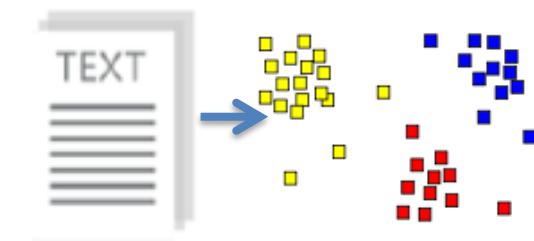
Example: learn to play Go, reward: *win or lose*



Classification



Regression



Clustering

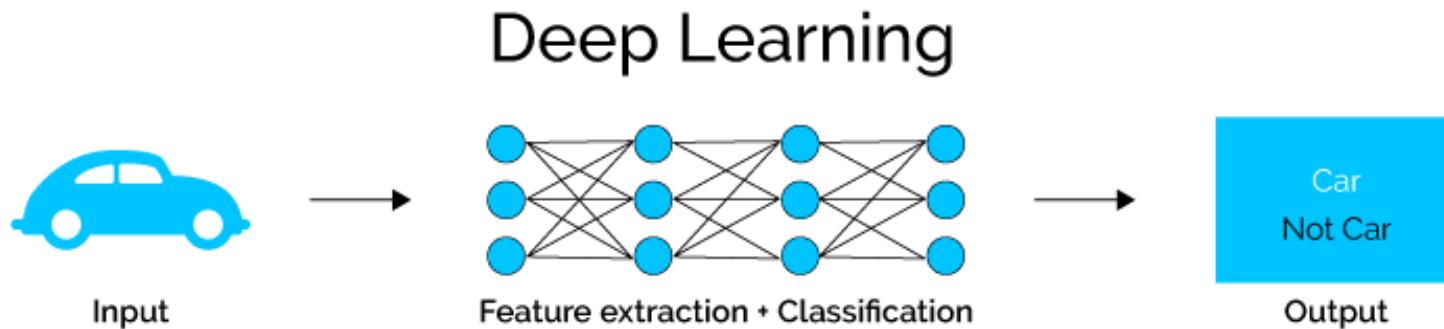
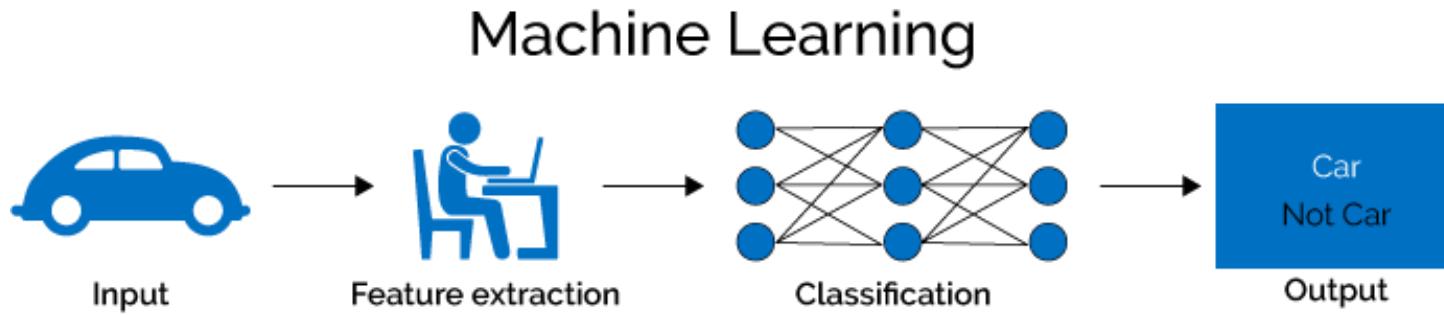
Anomaly Detection  
Sequence labeling

...

# ML vs. Deep Learning

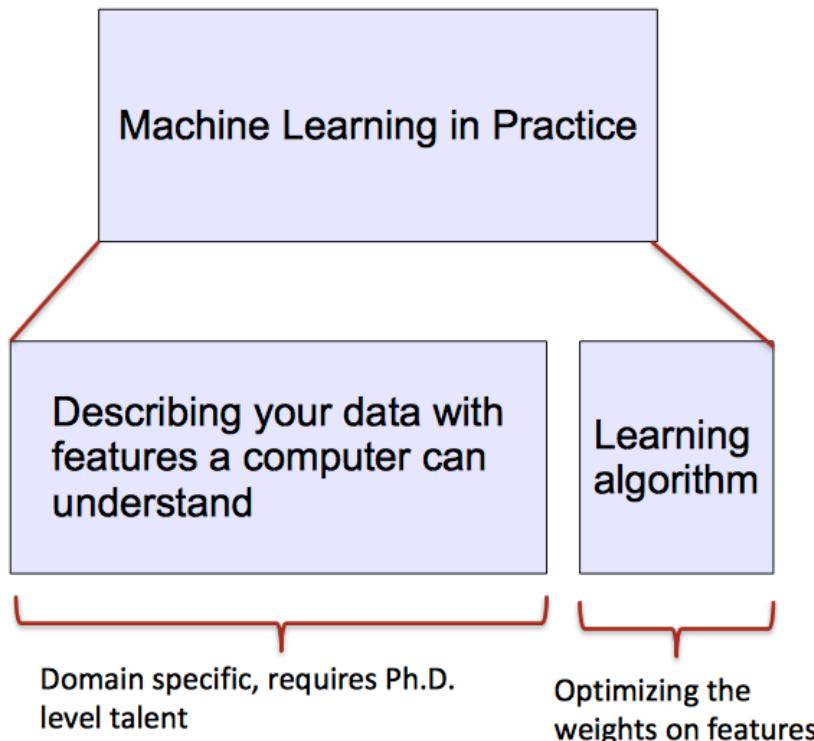
A machine learning subfield of learning **representations** of data.  
Exceptional effective at **learning patterns**.

Deep learning algorithms attempt to learn (multiple levels of) representation by using a **hierarchy of multiple layers**  
If you provide the system **tons of information**, it begins to understand it and respond in useful ways.



# ML vs. Deep Learning

Most machine learning methods work well because of **human-designed representations** and **input features**  
ML becomes just **optimizing weights** to best make a final prediction



Feature	NER
Current Word	✓
Previous Word	✓
Next Word	✓
Current Word Character n-gram	all
Current POS Tag	✓
Surrounding POS Tag Sequence	✓
Current Word Shape	✓
Surrounding Word Shape Sequence	✓
Presence of Word in Left Window	size 4
Presence of Word in Right Window	size 4

# Deep Learning Today

---

- Advancement in **speech recognition** in the last 2 years
  - A few long-standing performance records were broken with deep learning methods
  - Microsoft and Google have both deployed DL-based speech recognition systems in their products
- Advancement in **Computer Vision**
  - Feature engineering is the bread-and-butter of a large portion of the CV community, which creates some resistance to feature learning
  - But the record holders on ImageNet and Semantic Segmentation are convolutional nets
- Advancement in **Natural Language Processing**
  - Fine-grained sentiment analysis, syntactic parsing
  - Language model, machine translation, question answering

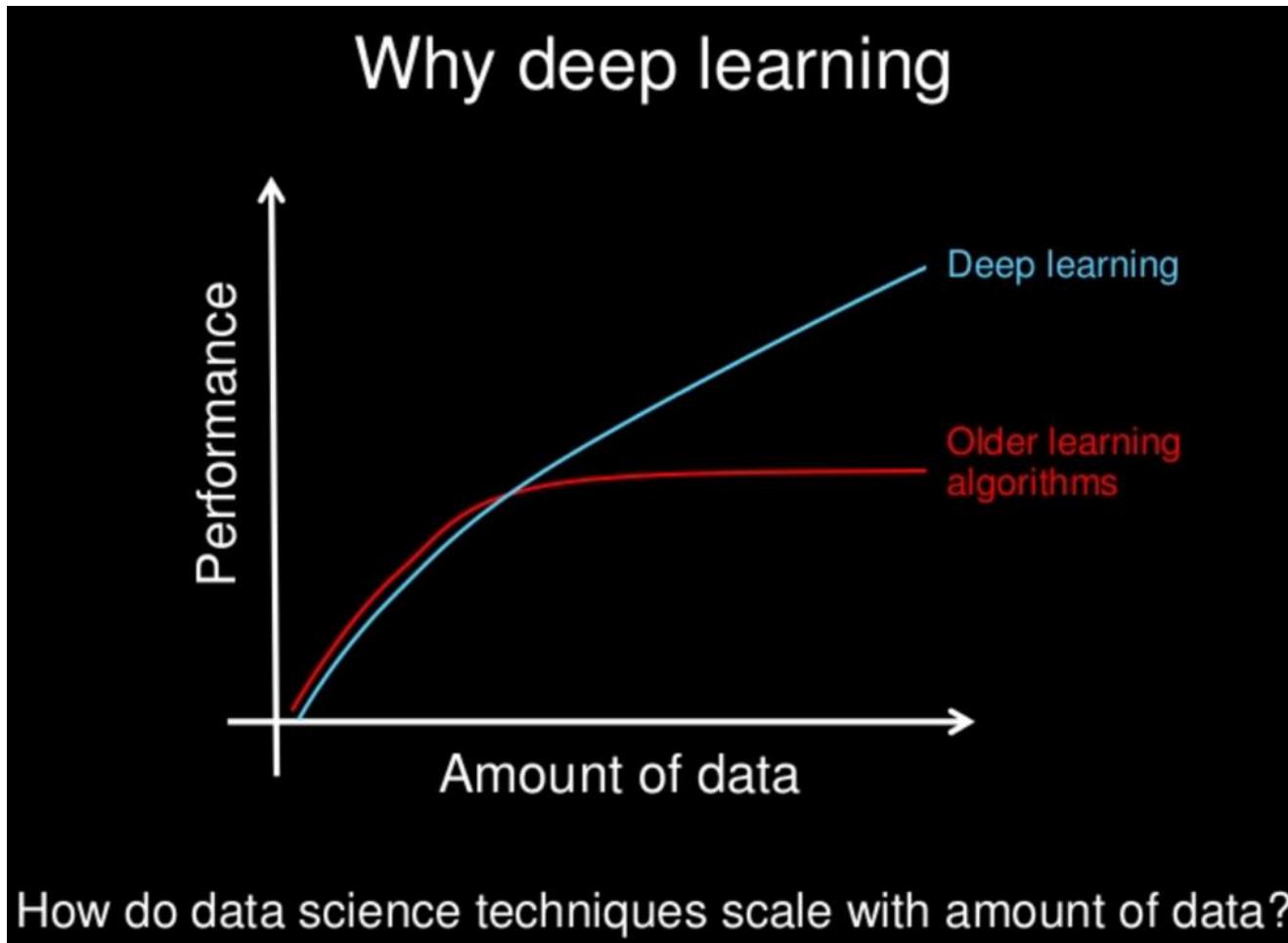
# Deep Learning Today



In “Nature” 27 January 2016:

- “DeepMind’s program AlphaGo beat Fan Hui, the European Go champion, five times out of five in tournament conditions...”
- “AlphaGo was not preprogrammed to play Go: rather, it learned using a general-purpose algorithm that allowed it to interpret the game’s patterns.”
- “...AlphaGo program applied **deep learning** in neural networks (convolutional NN) — brain-inspired programs in which connections between layers of simulated neurons are strengthened through examples and experience.”

# Why is Deep Learning Important



Deep Learning is finally enabling us to cross that line in places we weren't able to before.

# Big Year for AI



1936: Turing Machine from Alan Turing



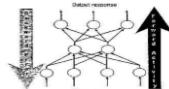
1946: Birth of ENIAC



1956: Birth of AI



1966: Turing Award



1986: The BP Algorithm



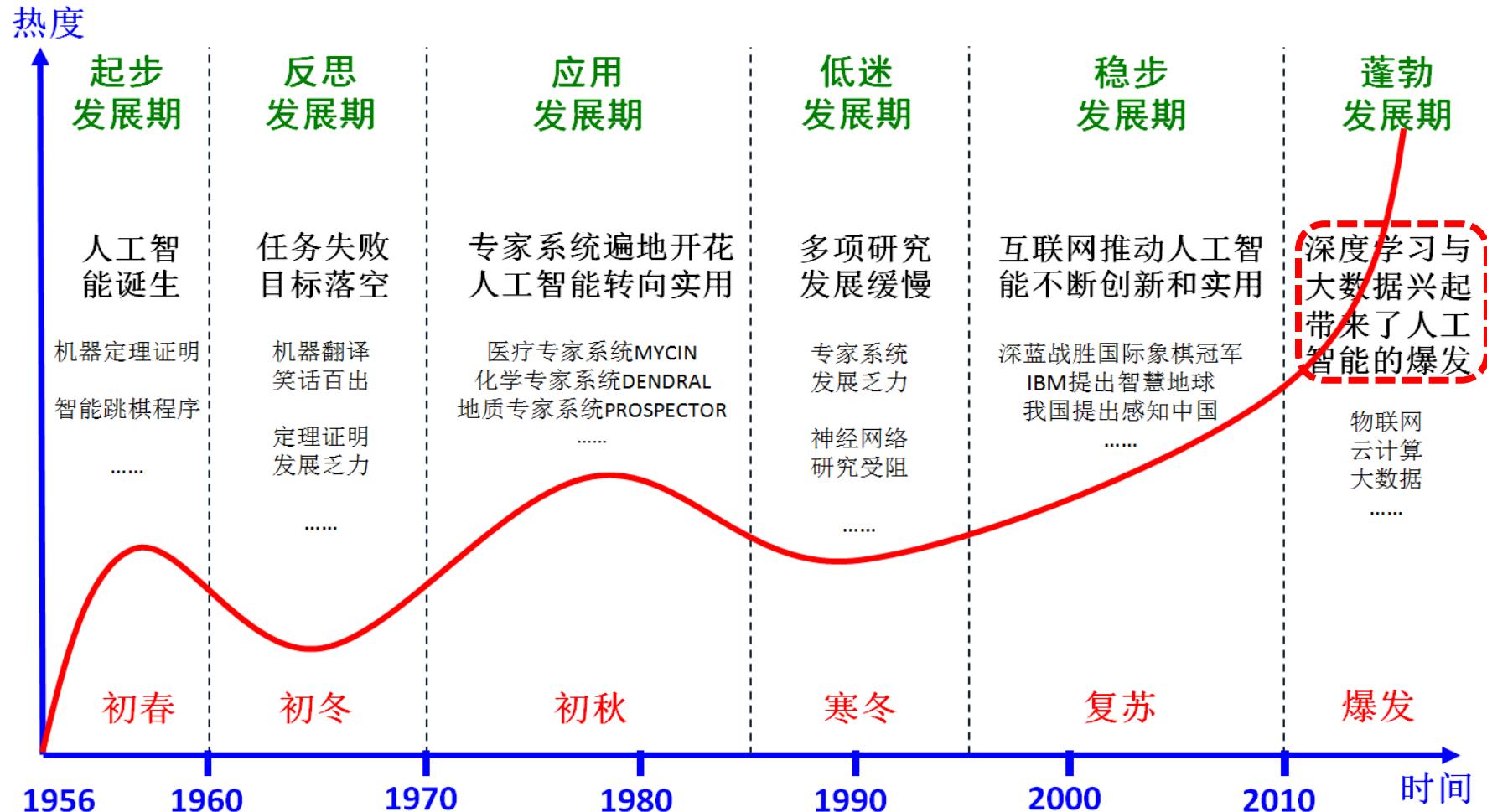
2006: DNN, Deep Learning



AlphaGo

2016: AlphaGo

# The History of AI



# Deep Neural Networks (DNN)

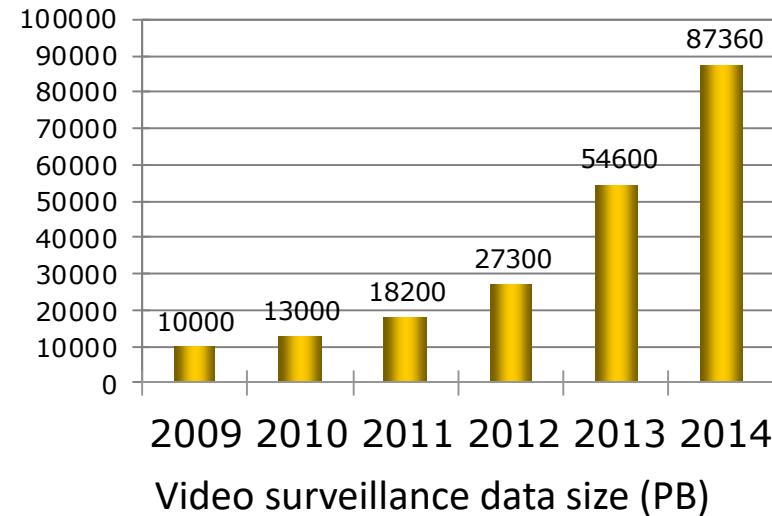
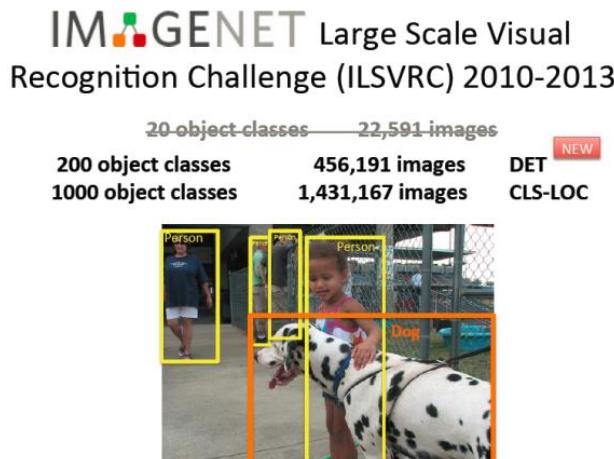
- **Originate from:**
  - 1962 – simple/complex cell, *Hubel and Wiesel*
  - 1970 – efficient error **backpropagation**, *Linnainmaa*
  - 1979 – deep **neocognitron**, **convolution**, *Fukushima*
  - 1987 – **autoencoder**, *Ballard*
  - 1989 – **convolutional neural networks (CNN)**, *Lecun*
  - 1991 – deep **recurrent neural networks (RNN)**, *Schmidhuber*
  - 1997 – **long short-term memory (LSTM)**, *Schmidhuber*
- **Two drawbacks:**

Large numbers of parameters → High computational cost

Small training set → Over-fitting problem

# Two Recent Developments

## Big Data



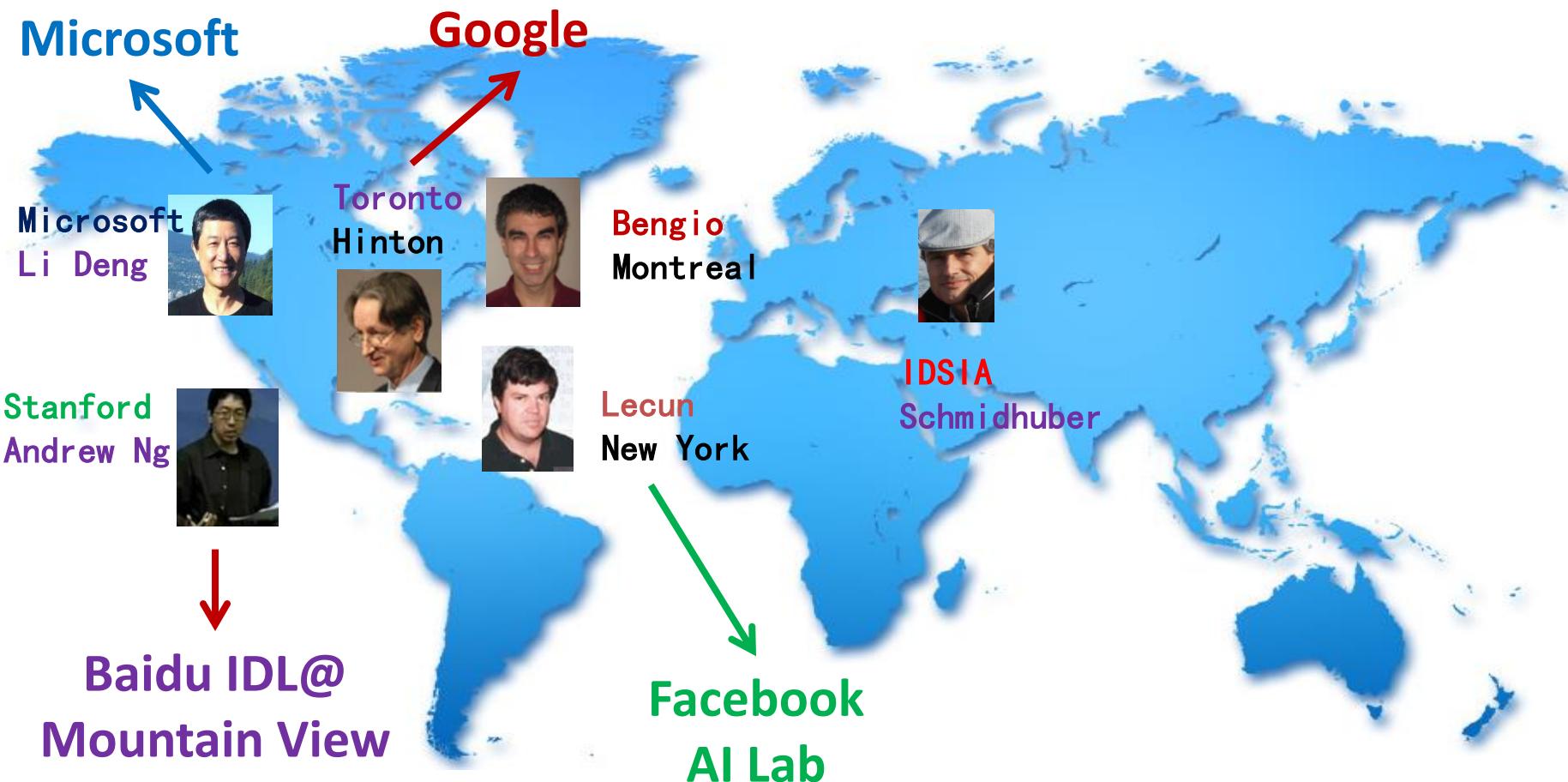
## Cheap Computation



Features	Tesla K40	Tesla K20X	Tesla K20
Number and Type of GPU	1 Kepler GK110B		1 Kepler GK110
Peak double precision floating point performance	1.43 Tflops	1.31 Tflops	1.17 Tflops
Peak single precision floating point performance	4.29 Tflops	3.95 Tflops	3.52 Tflops
Memory bandwidth (ECC off)	288 GB/sec	250 GB/sec	208 GB/sec
Memory size (GDDR5)	12 GB	6 GB	5 GB
CUDA cores	2880	2688	2496

DNN can thus be fitted efficiently

# Deep Learning Pioneers (2006-)



**They lead deep learning through three major stages!**

# Deep Learning: The Resurgence of DNN

Breakthrough in 2006

## Reducing the Dimensionality of Data with Neural Networks

G. E. Hinton\* and R. R. Salakhutdinov

High-dimensional data can be converted to low-dimensional codes by training a multilayer neural network with a small central layer to reconstruct high-dimensional input vectors. Gradient descent can be used for fine-tuning the weights in such "autoencoder" networks, but this works well only if the initial weights are close to a good solution. We describe an effective way of initializing the weights that allows deep autoencoder networks to learn low-dimensional codes that work much better than principal components analysis as a tool to reduce the dimensionality of data.

Dimensionality reduction facilitates the classification, visualization, communication, and storage of high-dimensional data. A simple and widely used method is principal components analysis (PCA), which finds the directions of greatest variance in the data set and represents each data point by its coordinates along each of these directions. We describe a nonlinear generalization of PCA that uses an adaptive, multilayer "encoder" network

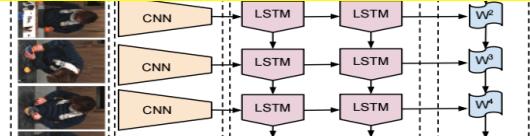
2006 VOL 313 SCIENCE www.sciencemag.org

2006

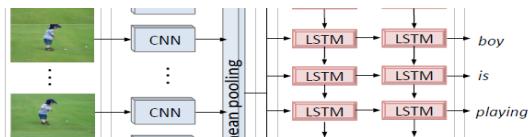
ImageNet: 74% vs. 85%



RNN for sequence analysis



Activity recognition, CVPR2015

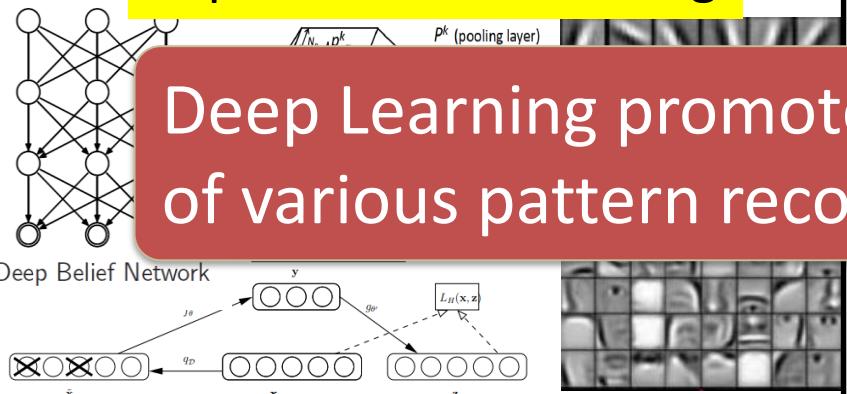


Video caption, CVPR2015

2012

2015

Representation learning



Deep Learning promotes the fast development of various pattern recognition areas

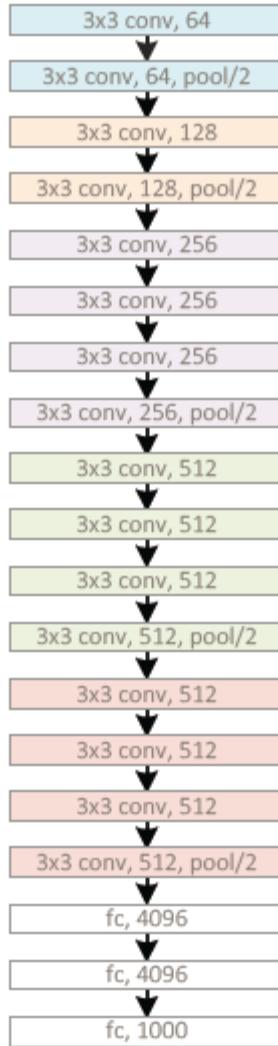
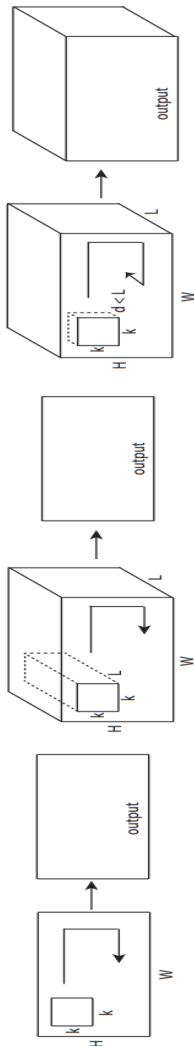
CNN for visual tasks



1. Input image  
2. Extract region proposals (~2k)  
3. Compute CNN features  
4. Classify regions

RCNN for detection, CVPR2014

# Deep Neural Networks



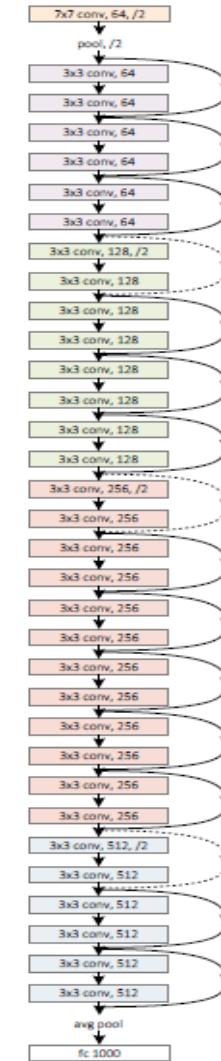
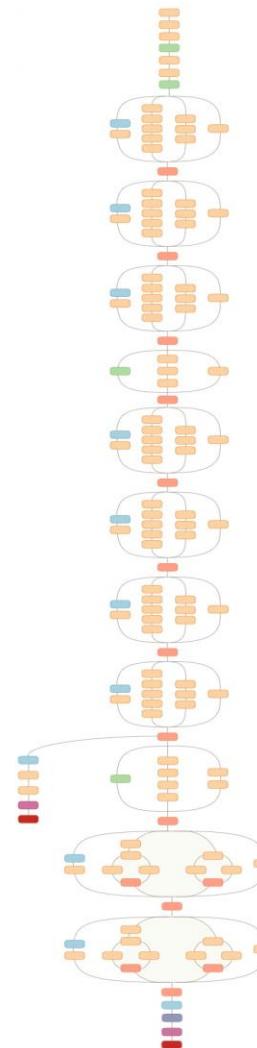
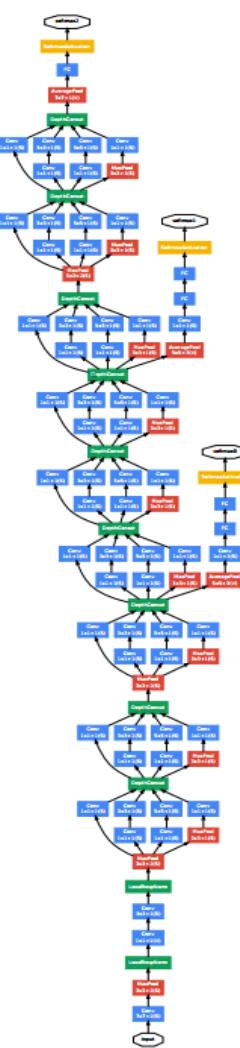
C3D

VGG-19

GoogleNet

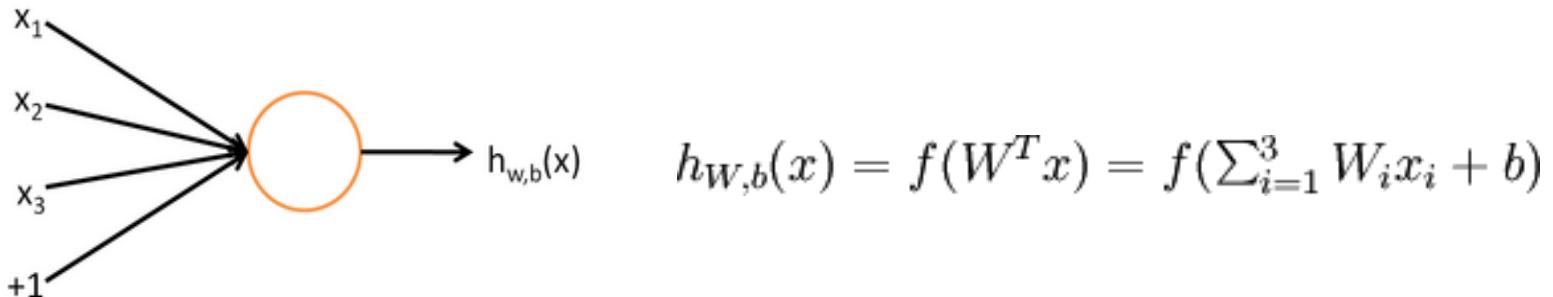
Inception-3

Residual

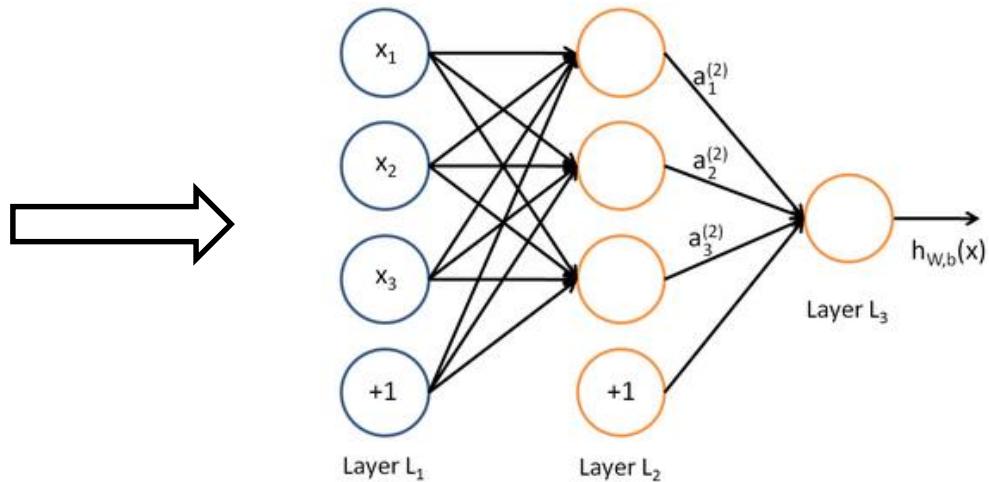


# DNN: Full Connection

- Each unit fully connects to all previous units



High-dimensional input

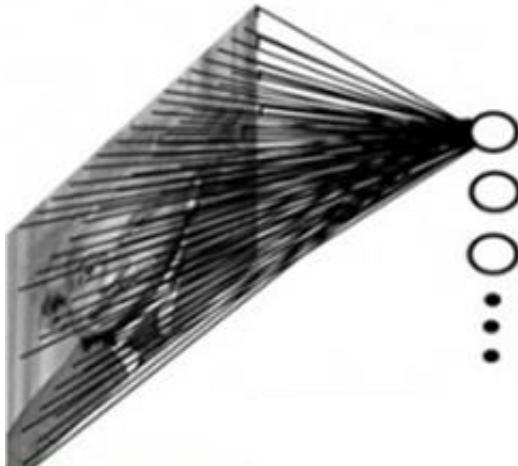


Large numbers of network parameters

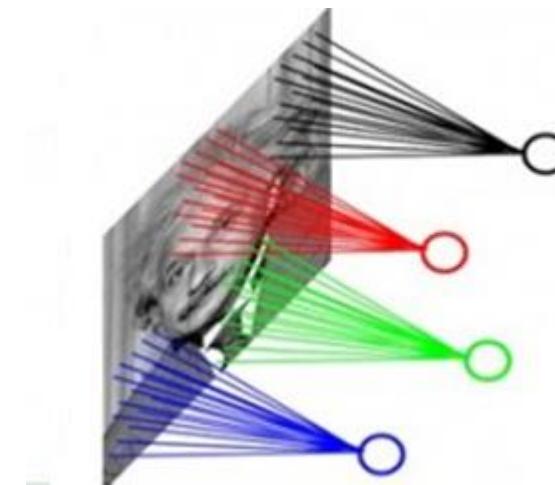
# CNN: Local Connection

<http://blog.csdn.net/stdcoutzyx/article/details/41596663>

- Local receptive field, biologically-plausible



Full connection

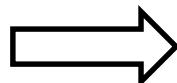


Local connection

- Image filtering



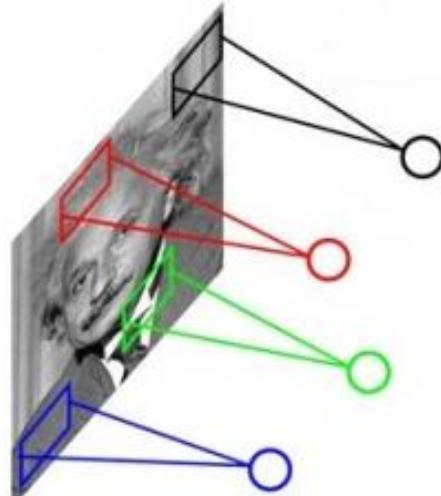
$$\ast \quad \begin{matrix} \textcolor{blue}{\bullet} & & \\ & \textcolor{blue}{\bullet} & \\ & & \textcolor{blue}{\bullet} \end{matrix} \quad w_9$$



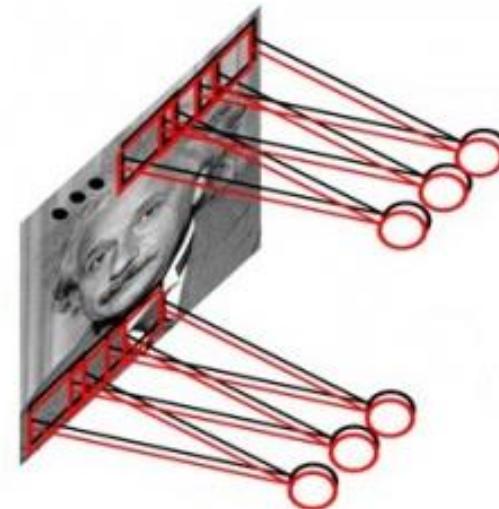
$$y = p_1 w_1 + \cdots + p_9 w_9$$

# CNN: Weight Sharing

- All local regions share the same filter weights



Local connection



Weight sharing

- Image convolution

1 x1	1 x0	1 x1	0	0
0 x0	1 x1	1 x0	1	0
0 x1	0 x0	1 x1	1	1
0	0	1	1	0
0	1	1	0	0

Image

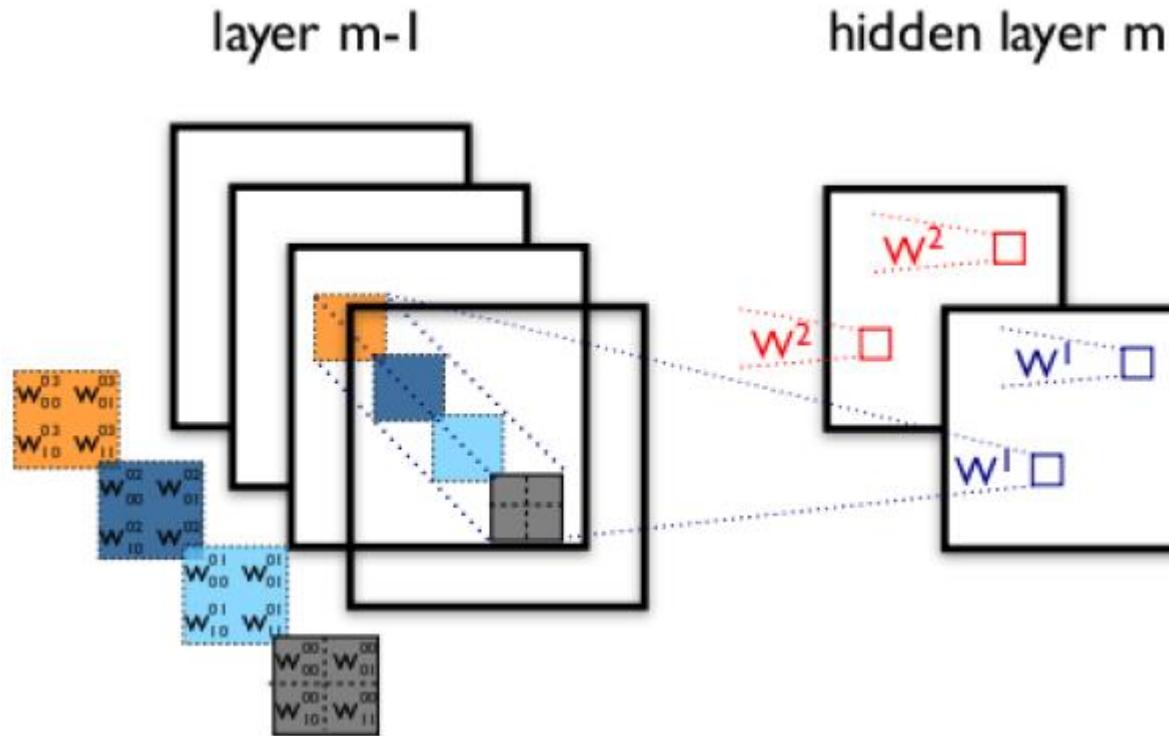
4		

Convolved Feature

$$h_{ij}^k = \tanh((W^k * x)_{ij} + b_k)$$

# CNN: Multiple Filters

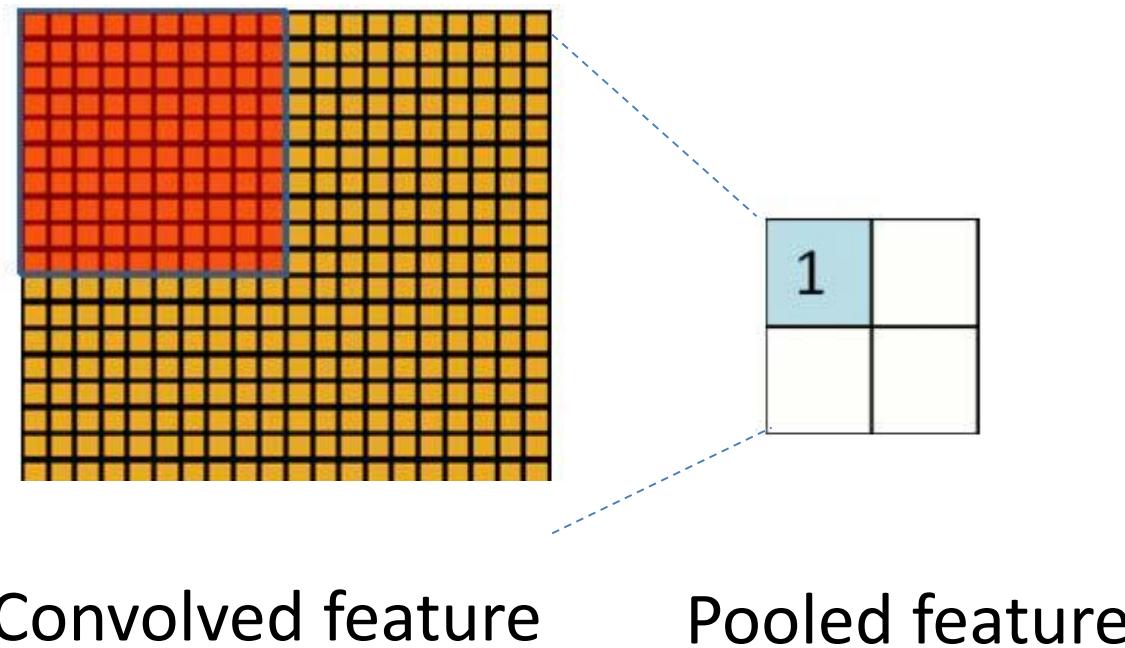
- Use  $N$  different filters obtain  $N$  feature maps



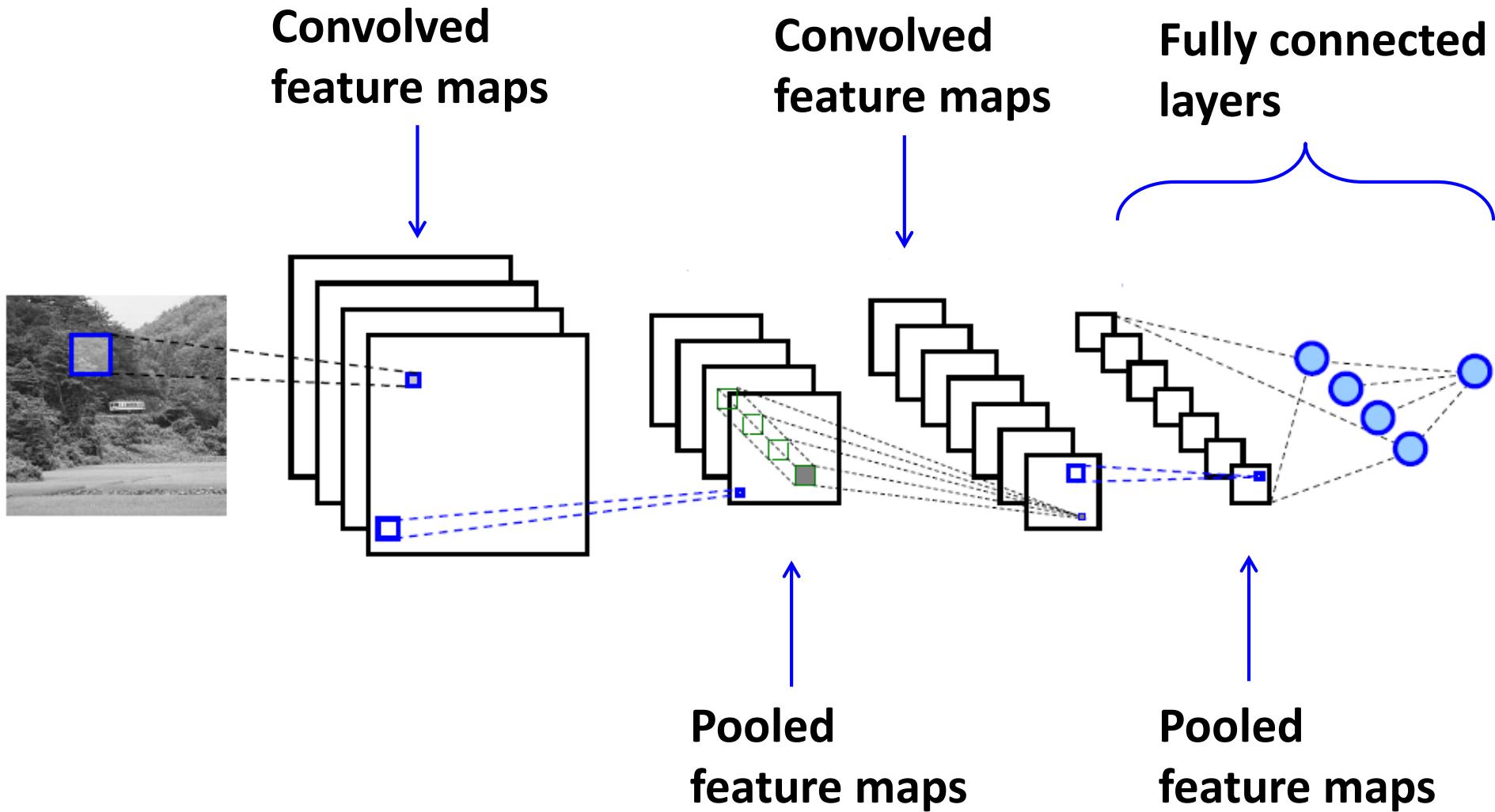
$w_1$ : weights of the 1<sup>st</sup> filter,  $w_2$ : weights of the 2<sup>nd</sup> filter

# CNN: Pooling

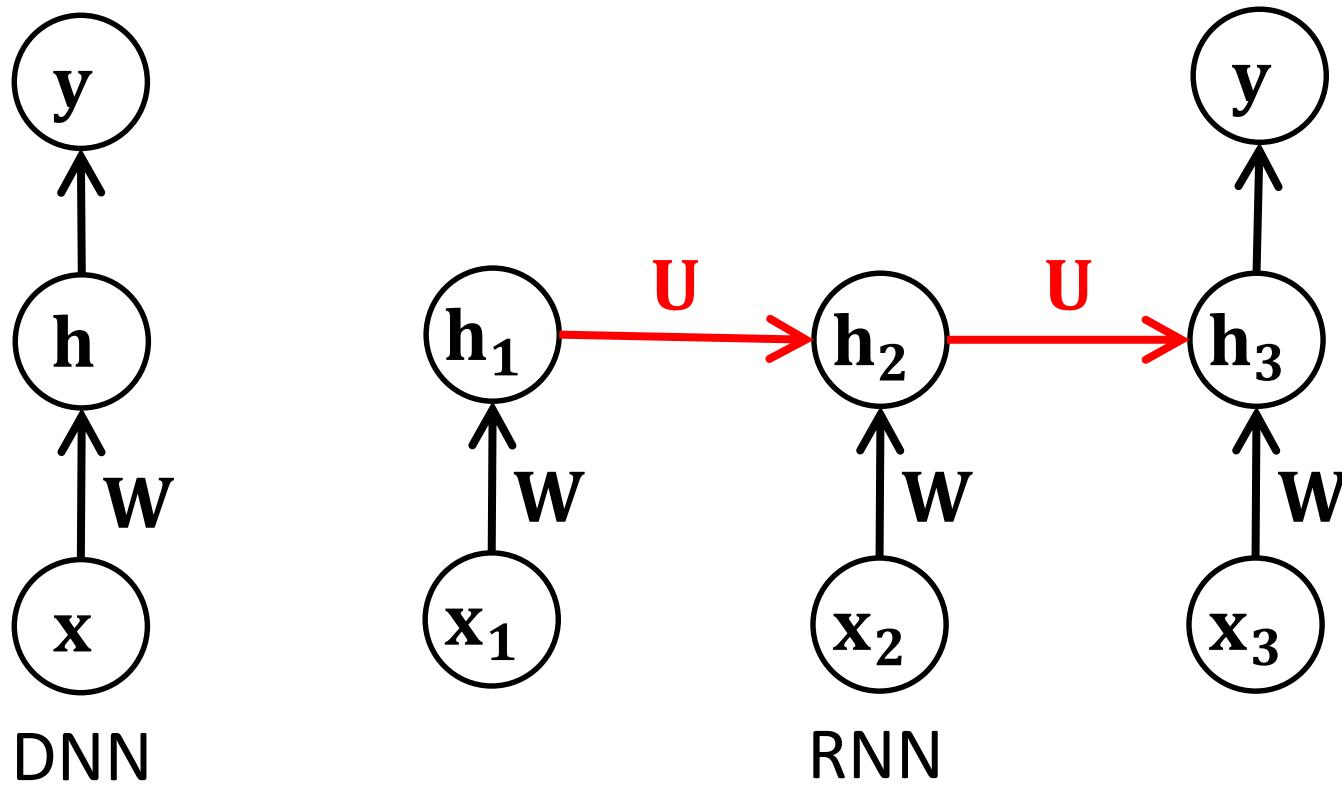
- Be invariant to image shifting using max (or average) pooling



# The Whole Architecture

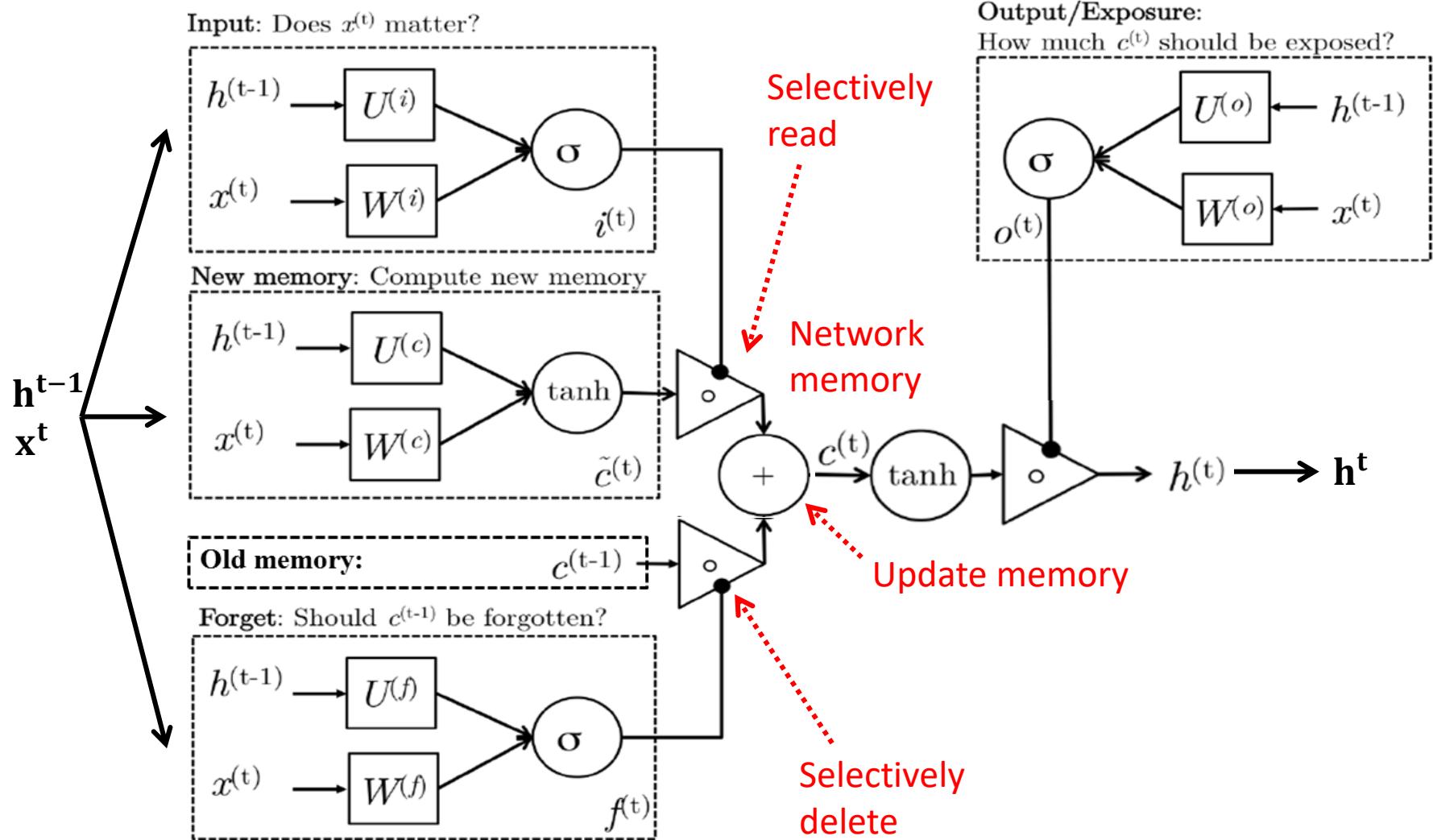


# RNN



- $\mathbf{x}_t \in \mathbb{R}^d, \mathbf{h}_t \in \mathbb{R}^n, \mathbf{W} \in \mathbb{R}^{d \times n}, \mathbf{U} \in \mathbb{R}^{n \times n}$
- $\mathbf{h}_t$  RNN aim to model long-range contextual information in forward direction

# LSTM



[Hochreiter and Schmidhuber, Neural computation 1997]

# Outline

---

**1** Course Information

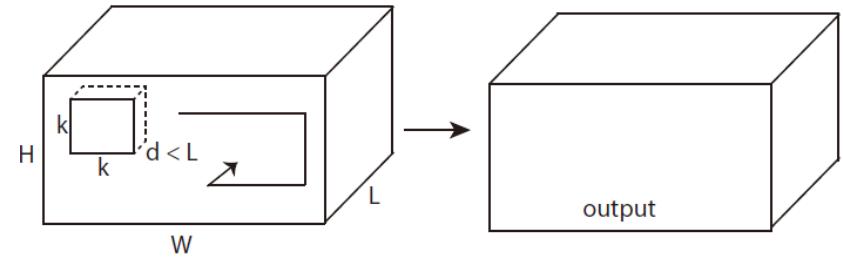
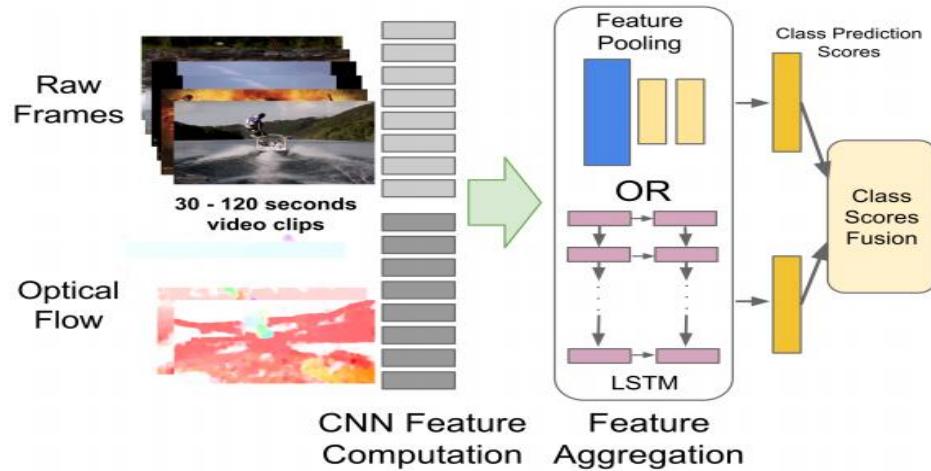
**2** What is Deep Learning

**3** Applications of Deep Learning

**4** Future Directions

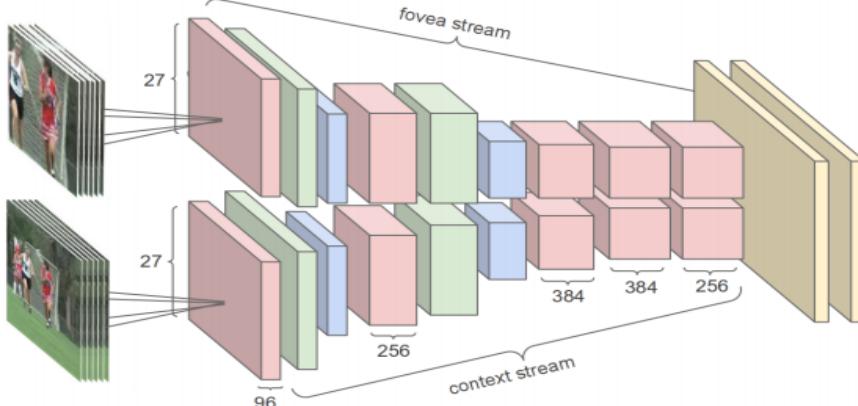
# Application for Video Classification

CNN: convolutional neural networks; RNN: recurrent neural networks

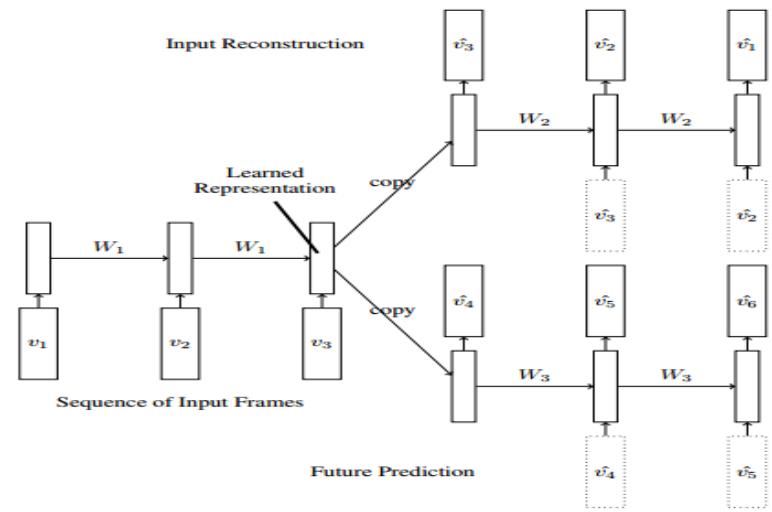


3D Convolution, ICCV2015

Long-range Videos Modelling, CVPR2015



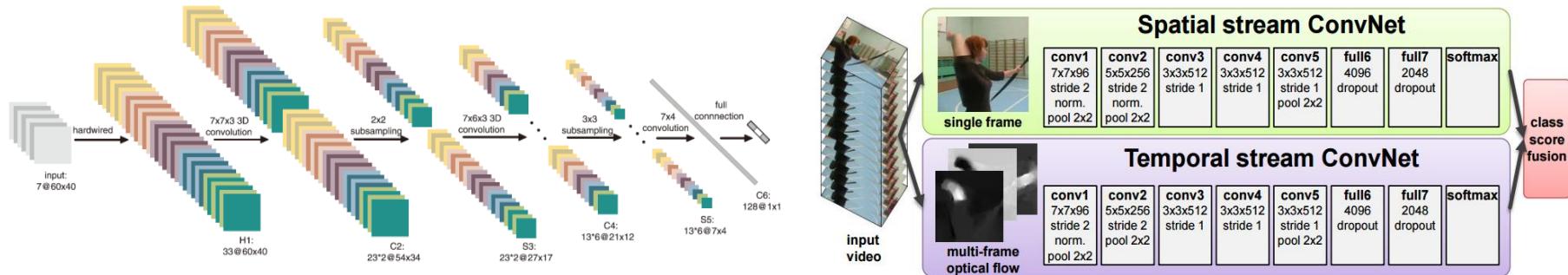
Multi-resolution CNN, CVPR2014



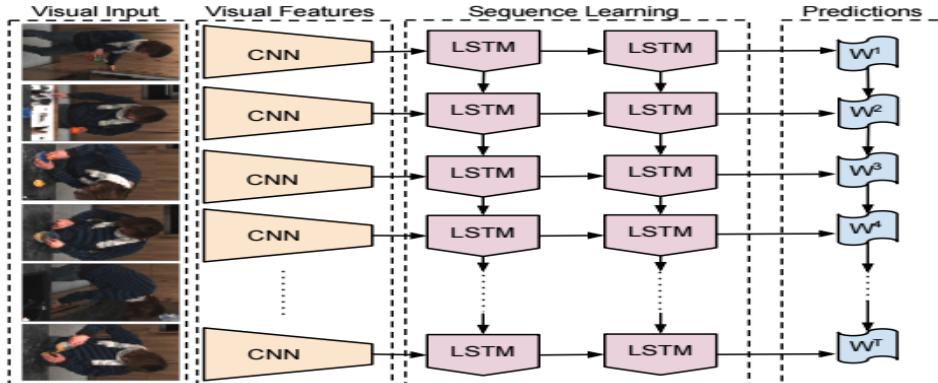
Autoencoder-RNN, ICML2015

# Application for Action Recognition

CNN: convolutional neural networks; LSTM: long short term memory; RNN: recurrent neural networks

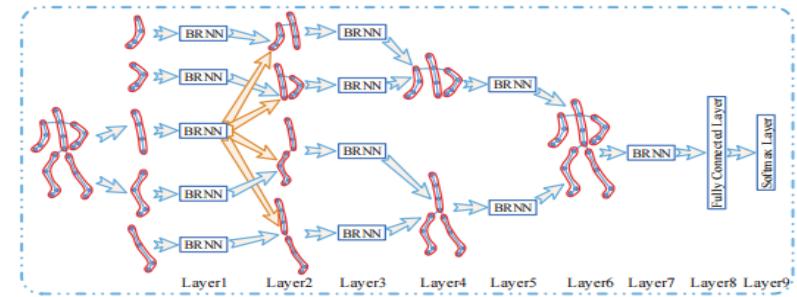


3D CNN, ICML2010



CNN + LSTM-RNN, CVPR2015

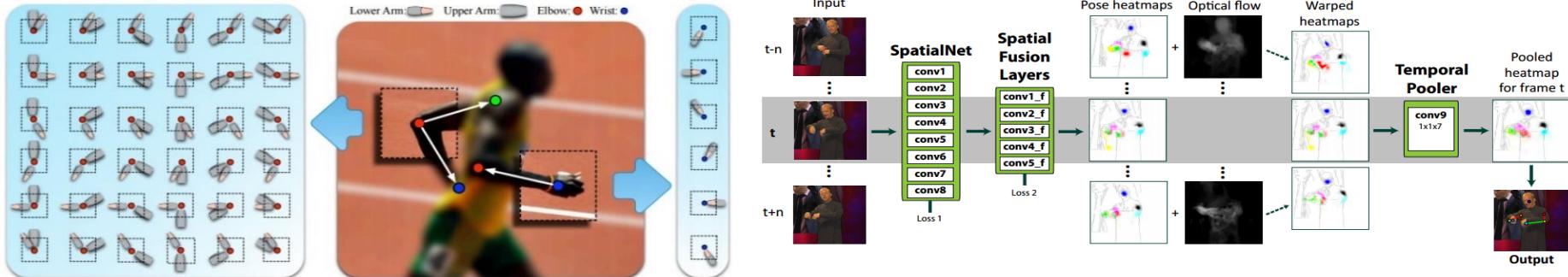
Two-stream CNN, NIPS2014



Skeleton + LSTM-RNN, CVPR2015

# Application for Pose Estimation

CNN: convolutional neural networks; DNN: deep neural networks



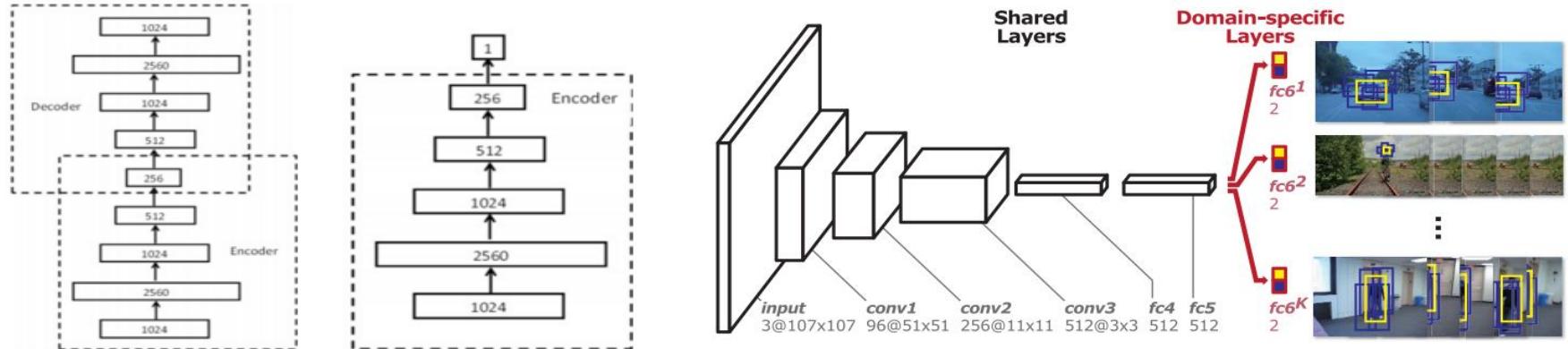
**CNN + Graphical Model, NIPS2014**

**Flowing ConvNets, ICCV2015**



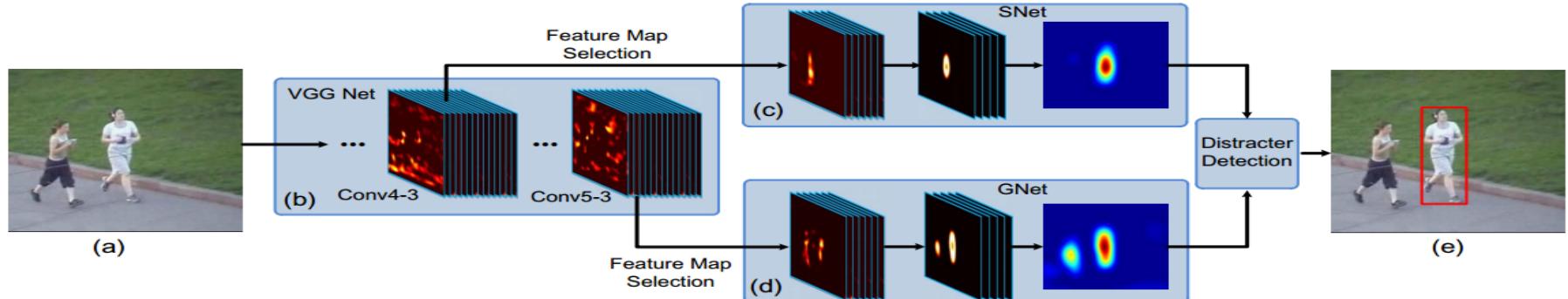
**DNN-based Regression, CVPR2014**

# Application for Video Tracking



Stacked Denoising Auto-encoder, NIPS2013

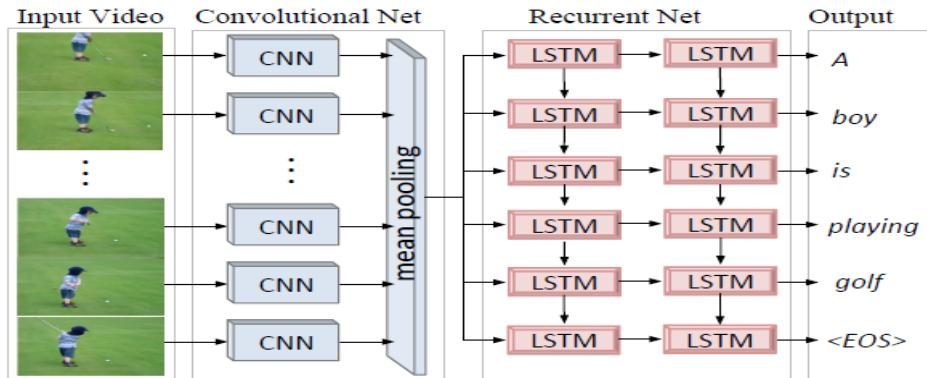
Multi-domain Networks, CVPR2016



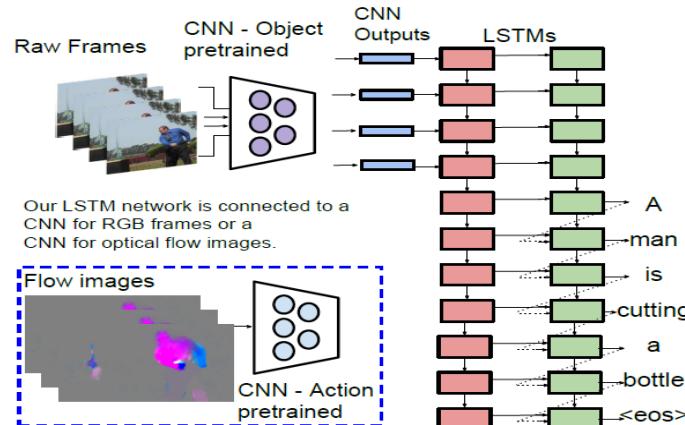
Fully Convolutional Networks, ICCV2015

# Application for Video Caption

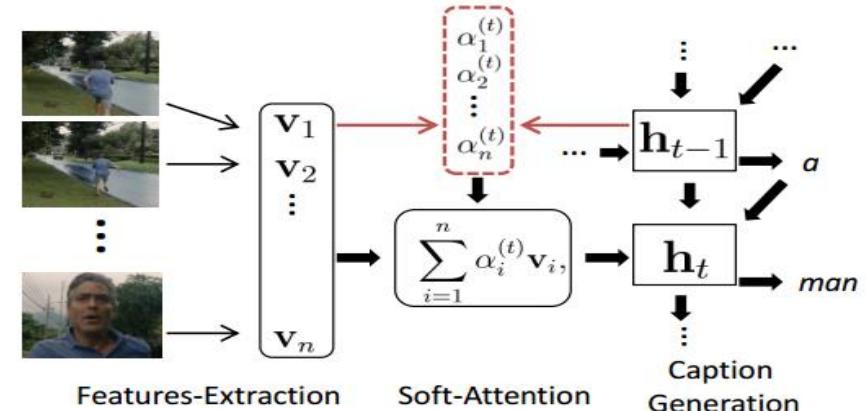
CNN: convolutional neural networks; LSTM: long short term memory; RNN: recurrent neural networks



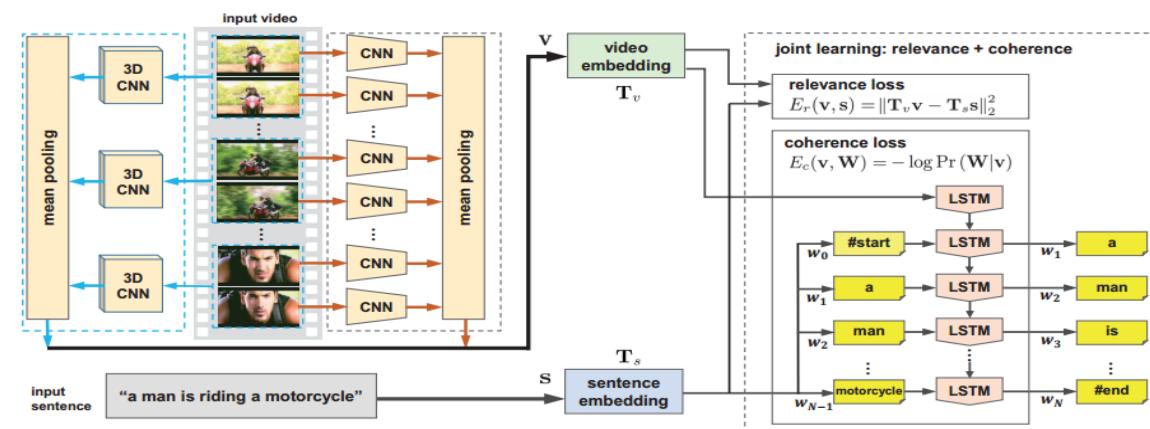
**CNN + LSTM-RNN, CVPR2015**



**Video2Text, ICCV2015**

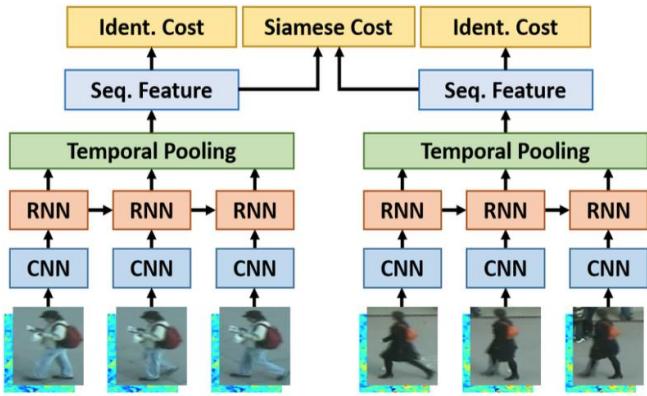


**Attention-based LSTM-RNN, ICCV2015**

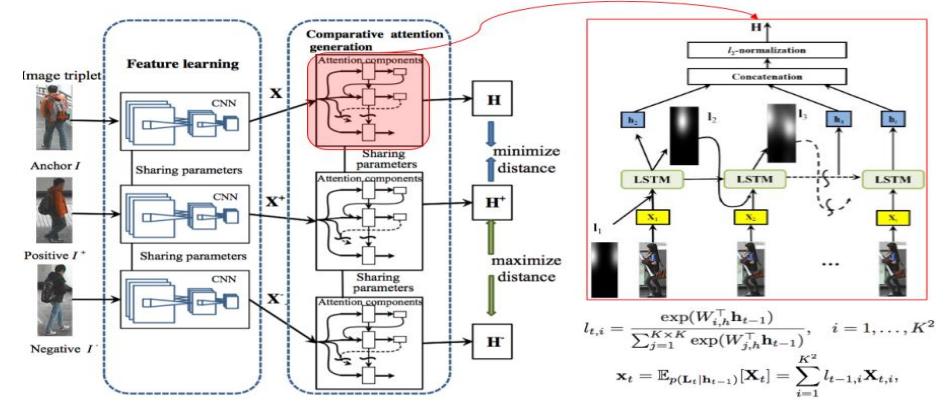


**Visual-semantic Embedding + LSTM-RNN, CVPR2016**

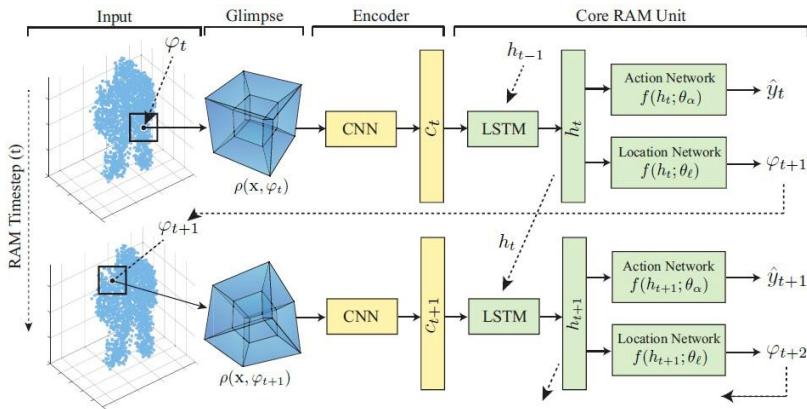
# Application for Person Re-identification



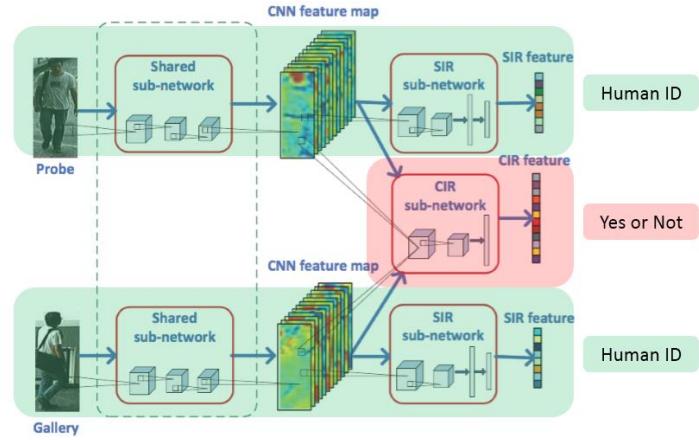
Recurrent CNN, CVPR2016



End-to-end Attention, TIP2016

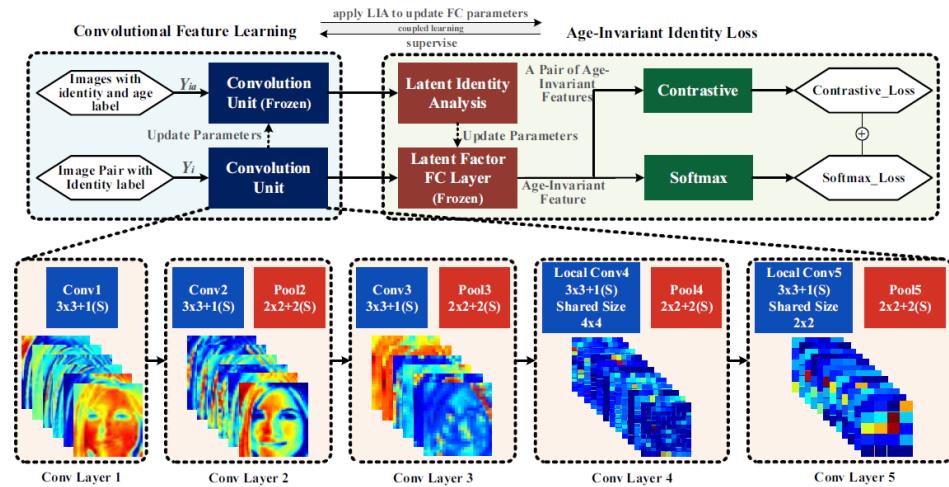
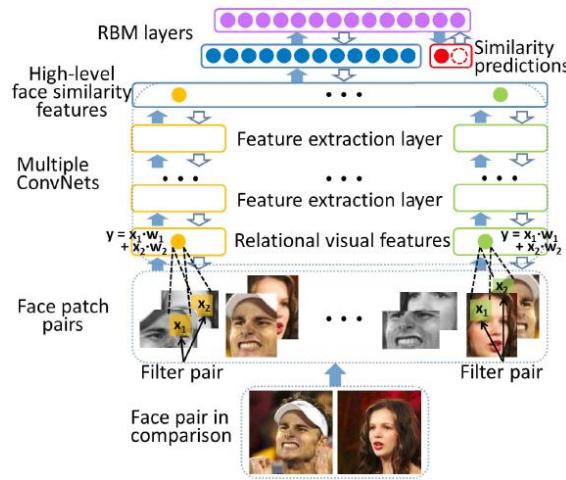


Sequential Local Attention, CVPR2016

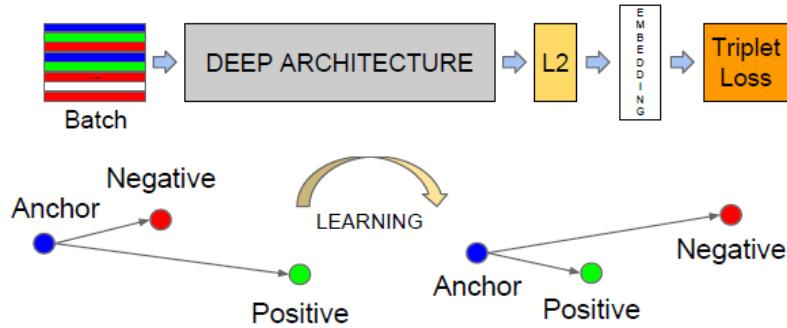


Single-/Cross-image Representations, CVPR2016

# Application for Face Recognition

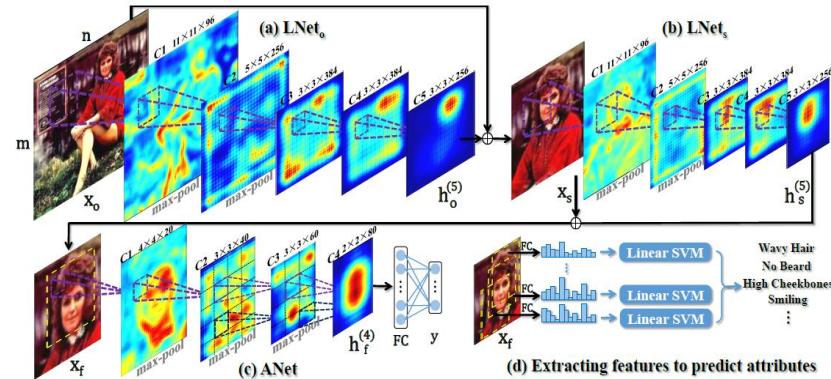


Hybrid ConvNet-RBM, ICCV13



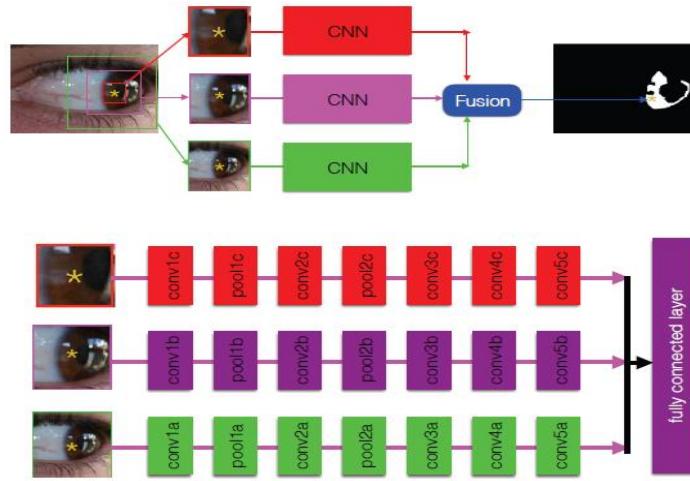
FaceNet, CVPR15

LF-CNNs Model, CVPR16

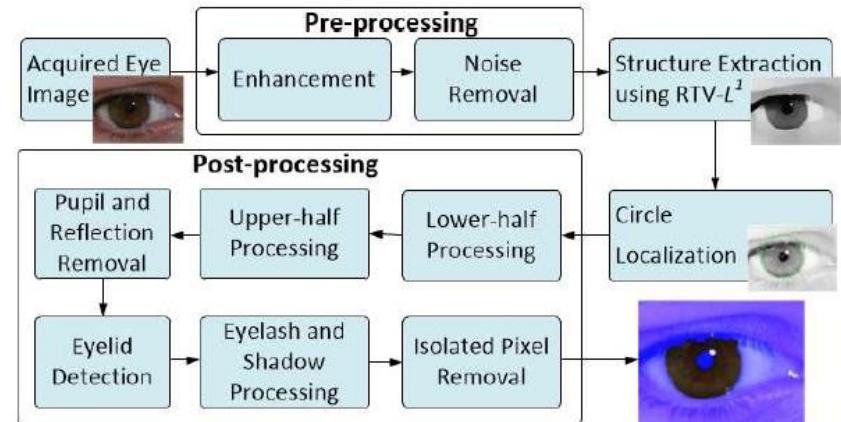


Face LNet and ANet, ICCV15

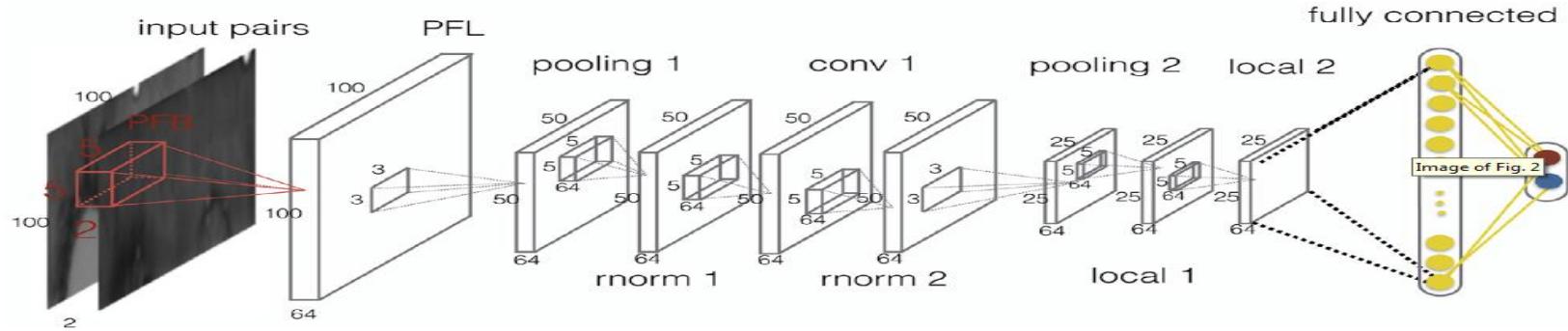
# Application for Iris Recognition



HCNNs, ICB16

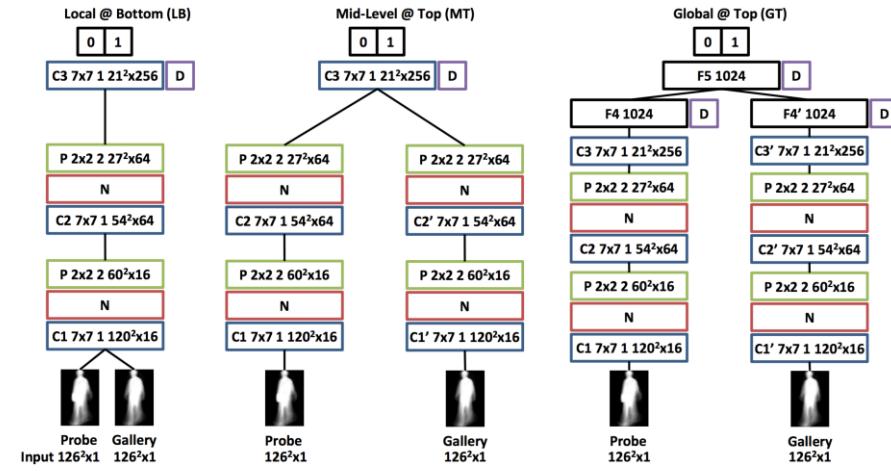
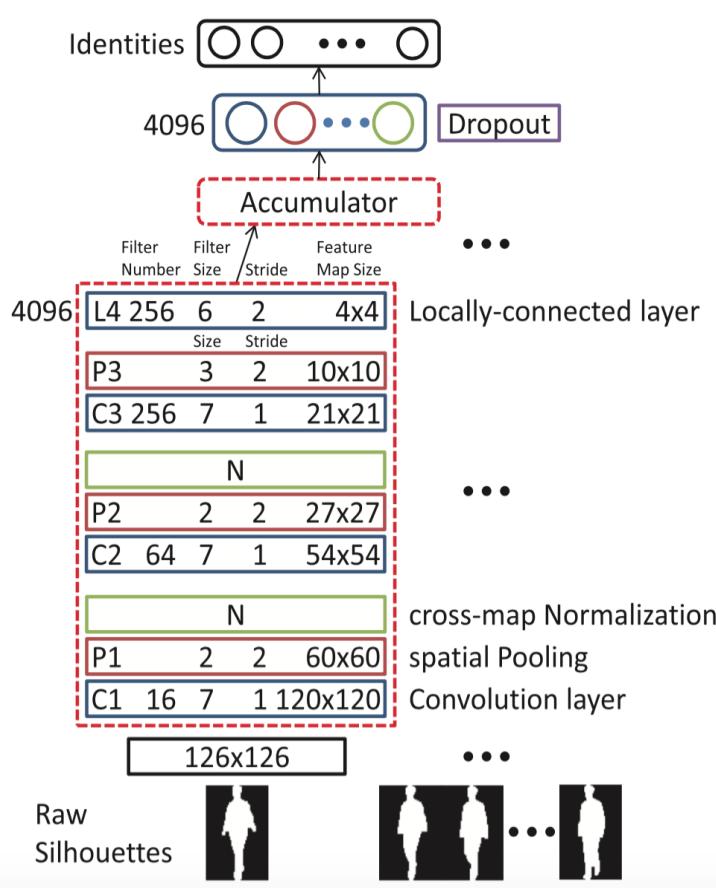


Iris Segmentation, ICCV15

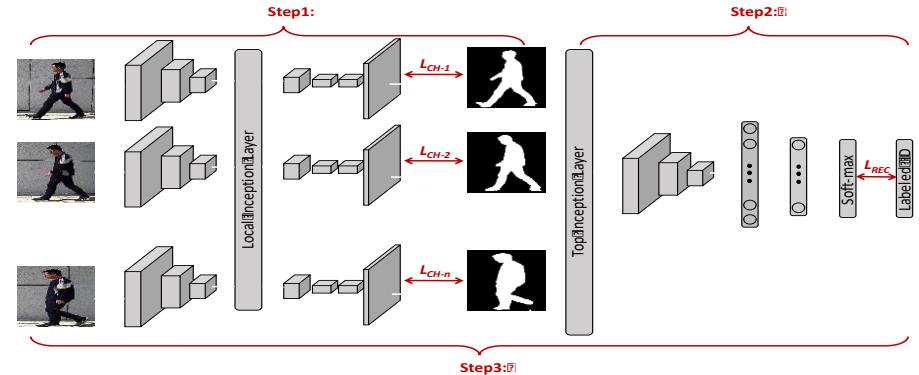


DeepIris, PRL15

# Application for Gait Recognition

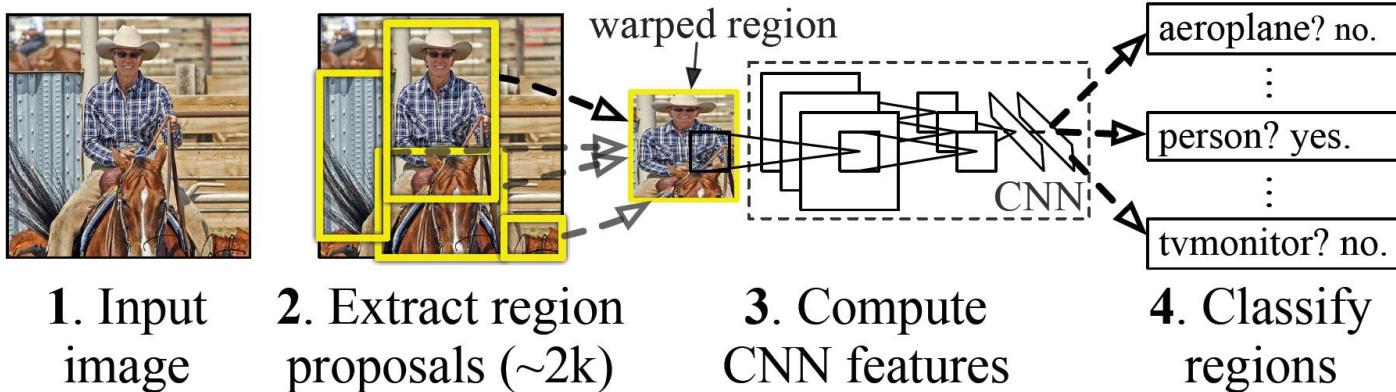


Cross-view Gait Identification, TPAMI 2016

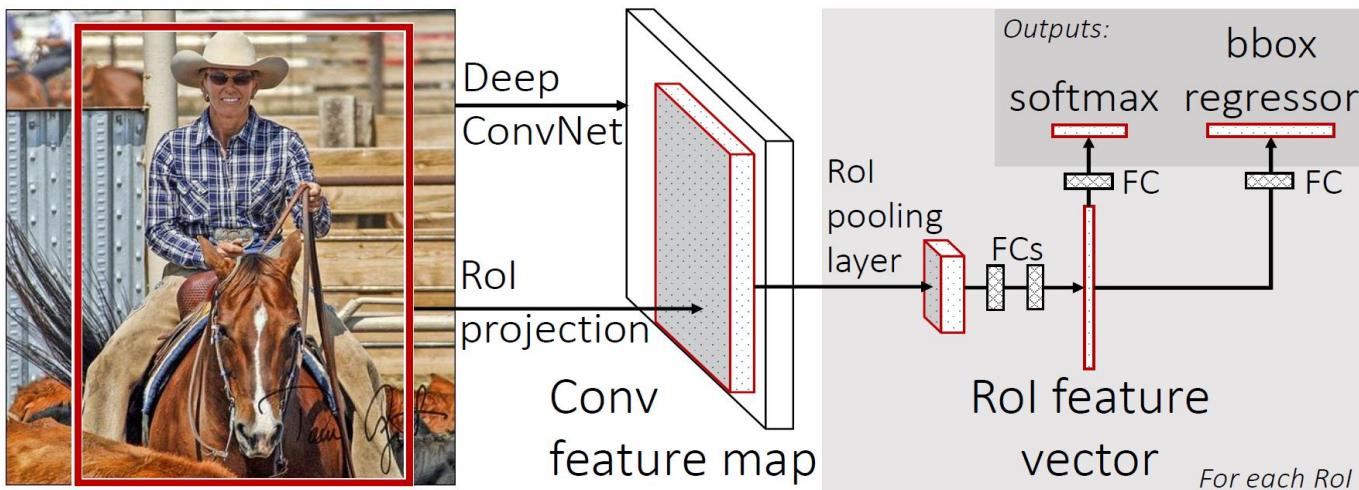


Deep Gait Feature Learning by Image Set, submitted to TIP 2017

# Application for Object Detection

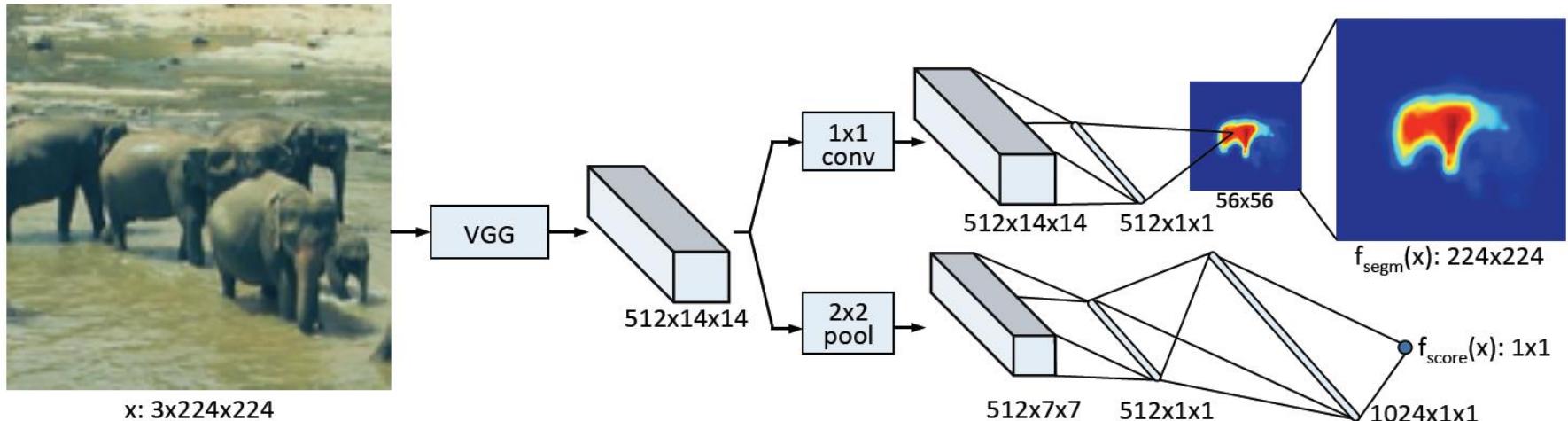


RCNN, CVPR14

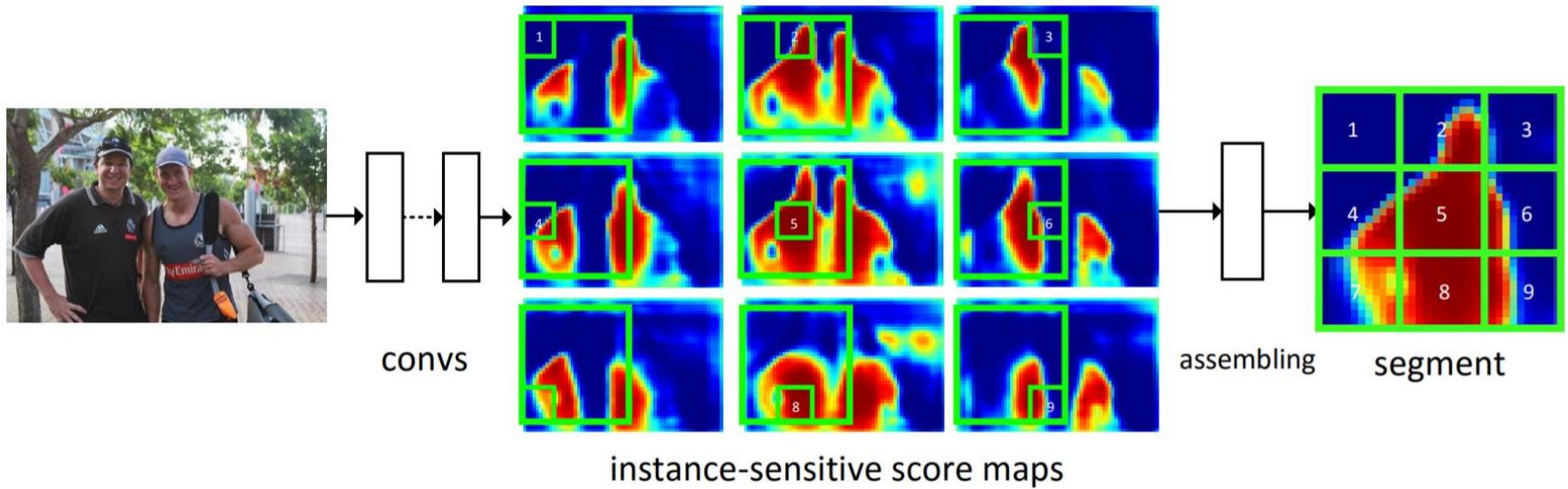


Fast R-CNN, ICCV15

# Application for Semantic Segmentation

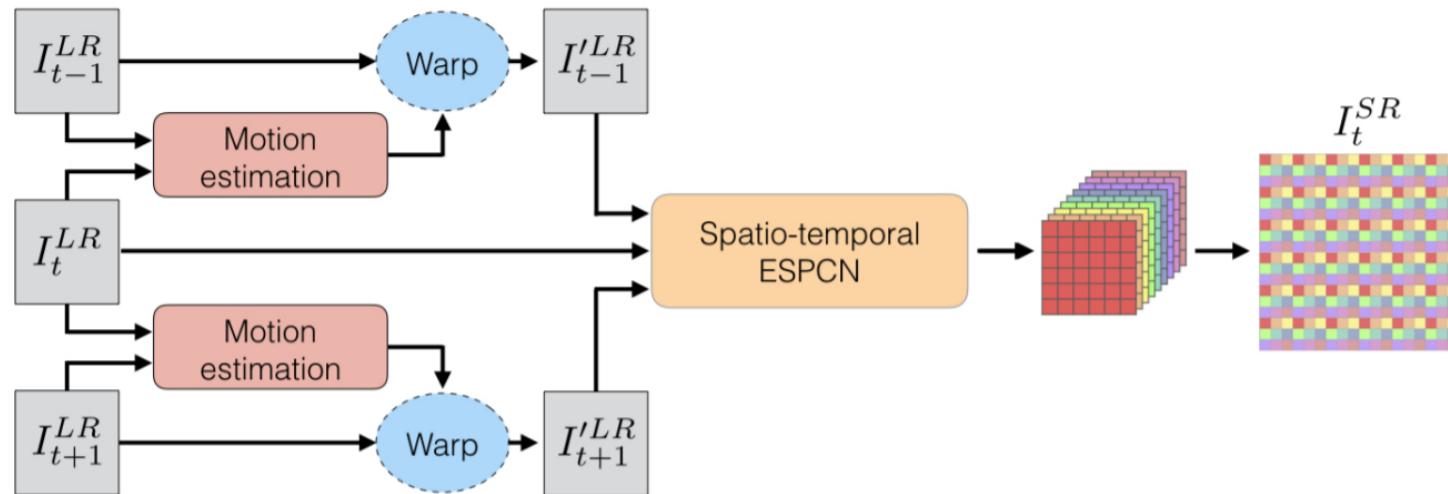


Instance-aware semantic segmentation, CVPR16

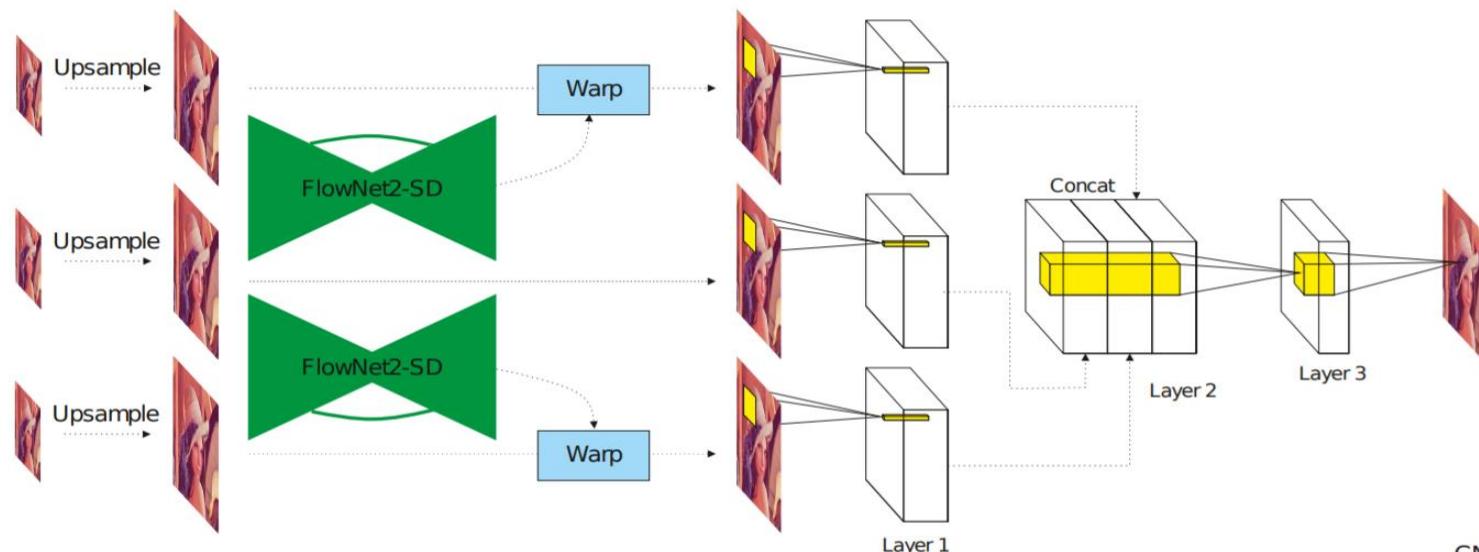


Learning to segment object candidates, NIPS15

# Application for Video Super-resolution



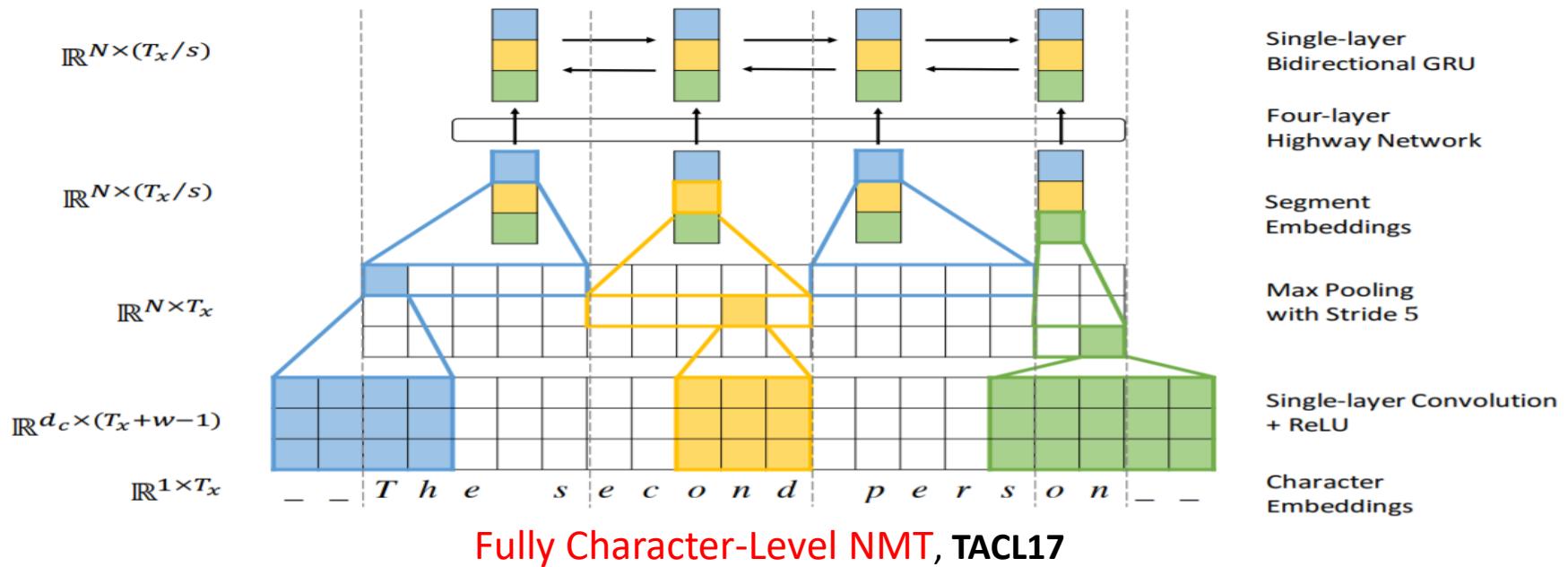
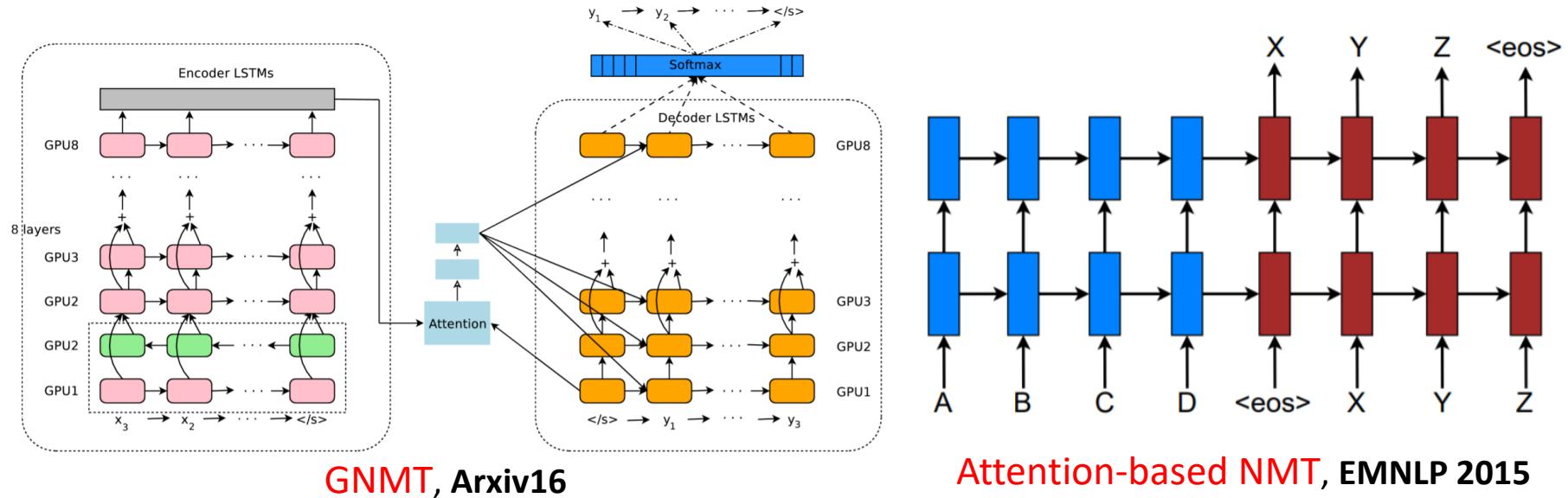
Real-Time Video Super-Resolution, CVPR17



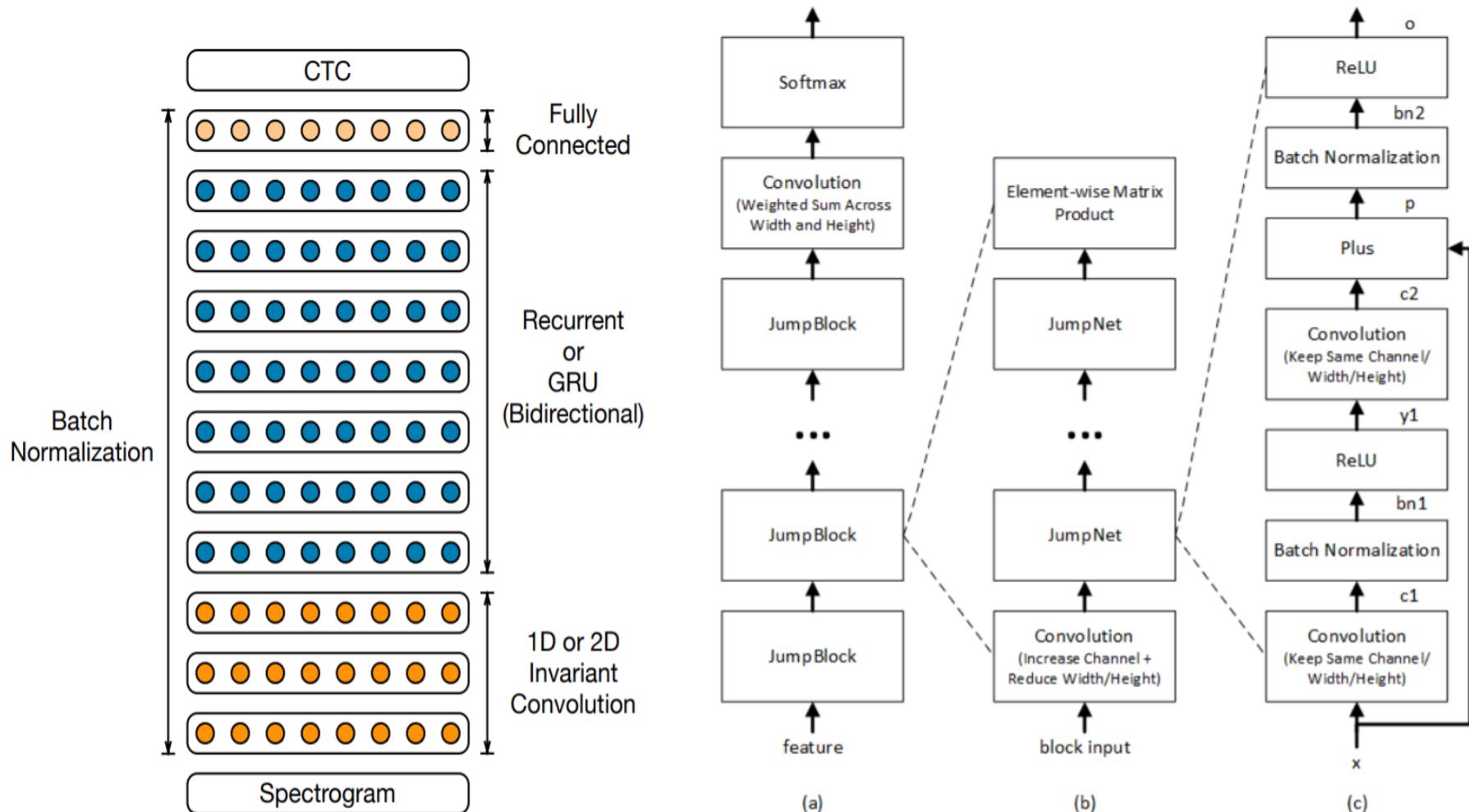
Video Super-Resolution with Motion Compensation, GCPR2017

CNN

# Application for Machine Translation



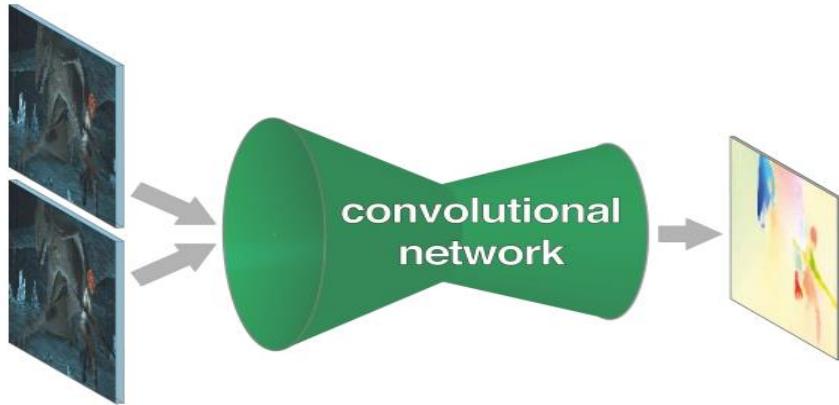
# Application for Speech Recognition



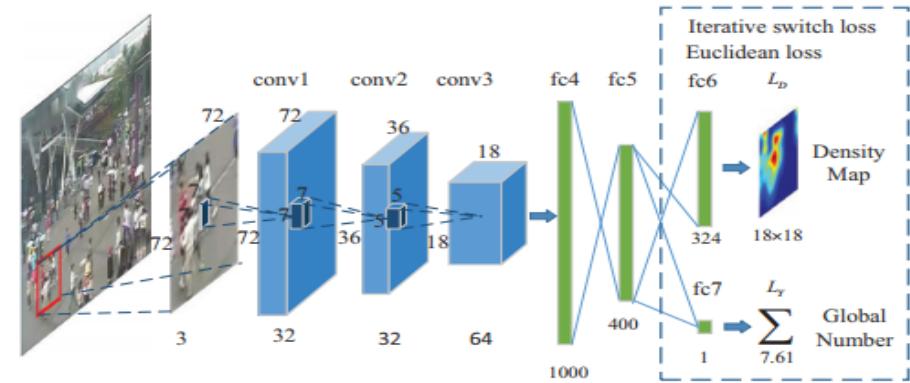
Speech Recognition in English  
and Mandarin, TASLP16

Conversational Speech Recognition,  
TASLP17

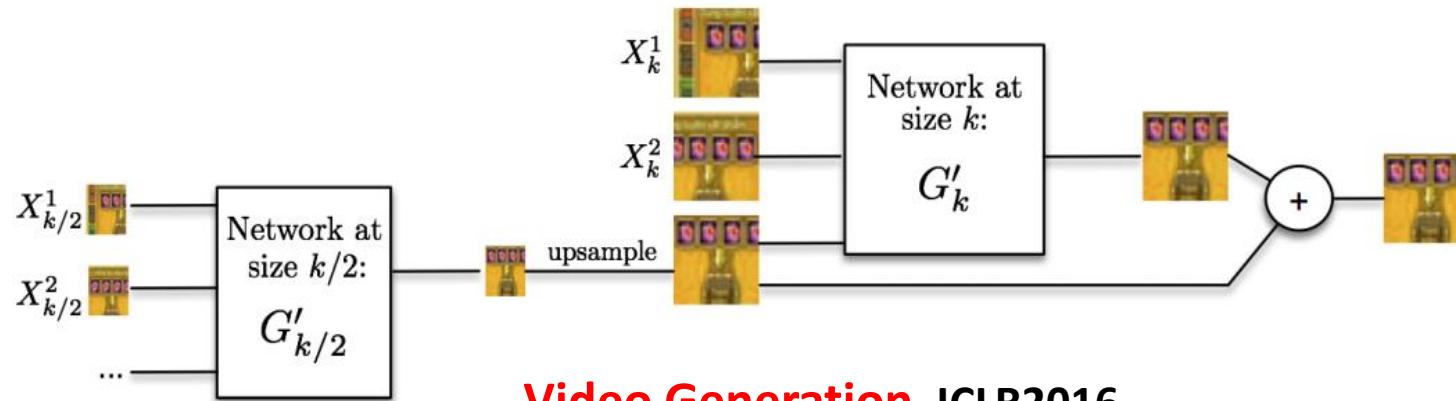
# Applications for Others



Optical Flow Prediction, ICCV2015



Cross-scene Crowd Counting, CVPR2015



Video Generation, ICLR2016

Achieve state-of-the-art results in various applications

# Outline

---

**1** Course Information

**2** What is Deep Learning

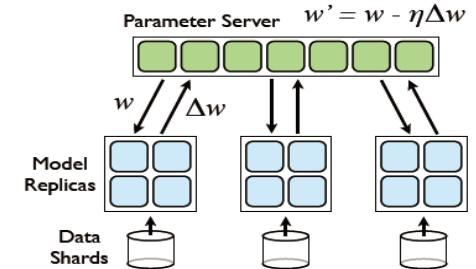
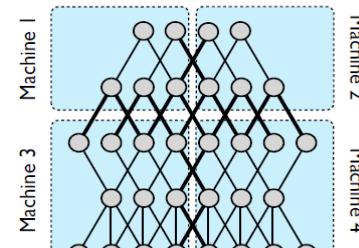
**3** Applications of Deep Learning

**4** Future Directions

# Direction 1

## ■ Large scale deep learning

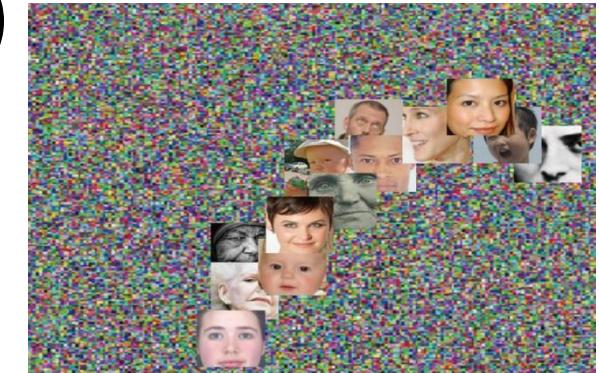
- multiple GPUs
- distributed system



# Direction 2

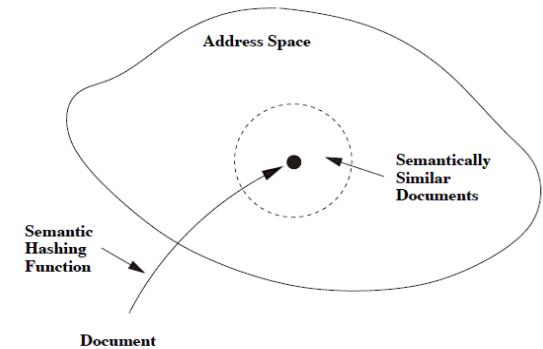
## ■ Unsupervised (semi-supervised) video feature Learning

- lack of supervised information in real applications



Unsupervised learned features

Manifold assumption



# Direction 3

## Multimodal learning

— videos are closely related to other data modalities, e.g., text, audio



### Image classification

man    woman    street    building  
people    kiss    walk

### Image captioning

Man kissing woman on a street

### Visual QA

Q: What is the sailor doing?

A: He is kissing a nurse in a white dress.

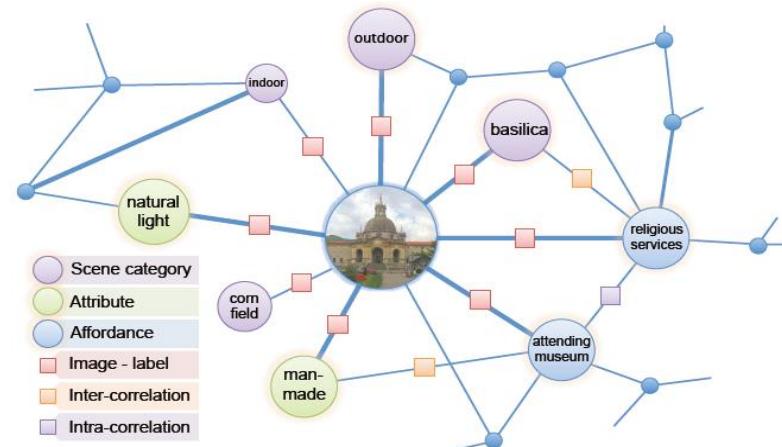
Q: Why are they kissing?

A: To celebrate V-J Day in Times Square.

## Question answering



	<b>Class</b> auditorium	<b>Affordances</b> community and social work, taking class for personal interest, religious practices, waiting, attending the performing arts
	<b>Class</b> landing deck	<b>Affordances</b> transportation and material moving work, in transit / traveling, military work
	<b>Class</b> candy store	<b>Affordances</b> transporting things or people, asphalt, natural light, far-away horizon, man-made
		<b>Attributes</b> congregating, indoor lighting, spectating, enclosed area, glossy

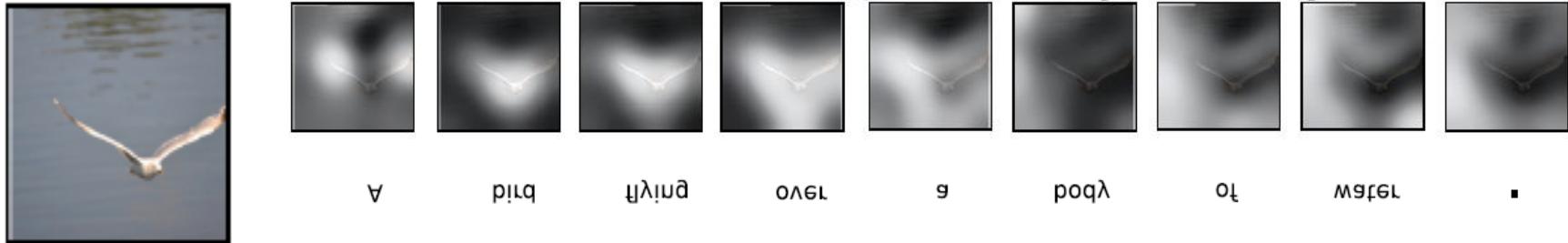


Multimodal Knowledge Base  
[Zhu et al., arXiv16]

# Direction 4

---

- Brain-inspired deep models
  - conventional neural networks are inspired by ***human cognitive mechanism***
  - refer to advanced achievements in the fields of neuroscience and brain science



Attention-based image caption [Xu et al., ICML15]

# References

---

1. **(Image Recognition) AlexNet** : Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, NIPS, 2012
2. **(Action Recognition)**: Shuiwang Ji, Wei Xu, Ming Yang, Kai Yu, 3D Convolutional Neural Networks for Human Action Recognition, ICML, 2010
3. **(Detection) R-CNN**: Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, CVPR, 2014
4. **(Detection) Faster R-CNN**: Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, arXiv:1506.01497
5. **(Segmentation) F-CNN**: Jonathan Long, Evan Shelhamer, Trevor Darrell, Fully Convolutional Networks for Semantic Segmentation, CVPR, 2015
6. **(Tracking)** Chao Ma, Jia-Bin Huang, Xiaokang Yang and Ming-Hsuan Yang, Hierarchical Convolutional Features for Visual Tracking, ICCV, 2015
7. **(Super-resolution)** Chao Dong, Chen Change Loy, Kaiming He, Xiaoou Tang, Learning a Deep Convolutional Network for Image Super-Resolution, ECCV, 2014

# References

---

8. **(Edge Detection)**: Gedas Bertasius, Jianbo Shi, Lorenzo Torresani, DeepEdge: A Multi-Scale Bifurcated Deep Network for Top-Down Contour Detection, CVPR, 2015
9. **(Face Recognition)**: Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, Lior Wolf, DeepFace: Closing the Gap to Human-Level Performance in Face Verification, CVPR, 2014
10. **(Question Answering)**: Hyeonwoo Noh, Paul Hongsuck Seo, and Bohyung Han, Image Question Answering using Convolutional Neural Network with Dynamic Parameter Prediction, arXiv:1511.05765
11. **(Video Caption)**: Jeff Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, Trevor Darrell, Long-term Recurrent Convolutional Networks for Visual Recognition and Description, CVPR, 2015
12. **(Text Classification)**: Xiang Zhang, Junbo Zhao, Yann LeCun, Character-level Convolutional Networks for Text Classification, NIPS, 2015
13. **(Retrieval)**: Fang Zhao, Yongzhen Huang, Liang Wang, Tieniu Tan, Deep Semantic Ranking Based Hashing for Multi-Label Image Retrieval, CVPR, 2015

# 1.Human Action Recognition

---

**Project idea:** Suppose that we are given a few video clips. Can we classify whether these videos contain humans ? If yes, can we track the humans in the videos, detect the pose of human body parts, and eventually infer what activity the human make? In this project, you are encouraged **to implement your own human action recognition system.** Since this area of research is so vast, we do not recommend you to try a general problem. More specified is better. Here are some references as a starting point.

Following is a list of data sets you could use.

- [1] KTH:<http://www.nada.kth.se/cvap/actions/>
- [2] Weizmann:<http://www.wisdom.weizmann.ac.il/~vision/SpaceTimeActions.html>
- [3] Hollywood Human Actions  
dataset:<http://www.irisa.fr/vista/Equipe/People/Laptev/download.html>
- [4] VIRAT Video Dataset:<http://www.viratdata.org/>  
([http://groups.inf.ed.ac.uk/calvin/articulated\\_human\\_pose\\_estimation\\_code/](http://groups.inf.ed.ac.uk/calvin/articulated_human_pose_estimation_code/))

# 1.Human Action Recognition

---

Here are some references as a starting point.

- [1] CVPR 2011 Tutorial on Human Activity Recognition  
(<http://cvrc.ece.utexas.edu/mryoo/cvpr2011tutorial/>)
- [2] Human Activity Recognition Summer course  
(<http://www.cs.sfu.ca/~mori/courses/cmpt888/summer10/>)
- [2] Stanford Vision lab ([http://vision.stanford.edu/discrim\\_rf/](http://vision.stanford.edu/discrim_rf/))  
(<http://ai.stanford.edu/~bangpeng/ppmi.html>)
- [3] Poselet (<http://www.cs.berkeley.edu/~lbourdev/poselets/>)
- [4] 2D articulated human pose estimation

You could use any code from the web for computing spatial-temporal features. One good example is the spatial-temporal interest point proposed by Piotr Dollar. Source code available at <http://vision.ucsd.edu/~pdollar/research/research.html>.

---

## 2. Image Categorization

---

**Project idea:** Image categorization/object recognition has been one of the most important research problems in the computer vision community. Researchers have developed a wide spectrum of different local descriptors, feature coding schemes, and classification methods.

In this project, **you will implement your own object recognition system**. You could use any code from the web for computing image features, such as SIFT, HoG, etc.

For computing SIFT features, you could use <http://www.vlfeat.org/~vedaldi/code/sift.html>.

**You're also encouraged to try the state-of-the-art deep learning methods:**

<http://deeplearning.net/>

Following is a list of data sets you could use.

[1] Caltech101/256:

[http://www.vision.caltech.edu/Image\\_Datasets/Caltech101/Caltech101.html](http://www.vision.caltech.edu/Image_Datasets/Caltech101/Caltech101.html)

[2] The PASCAL Object Recognition Database Collection:

<http://pascallin.ecs.soton.ac.uk/challenges/VOC/databases.html>

[3] LabelMe:<http://labelme.csail.mit.edu/>

[4] Face in the wild:<http://vis-www.cs.umass.edu/lfw/>

[5] ImageNet:<http://www.image-net.org/index>

[6] TinyImage:<http://groups.csail.mit.edu/vision/TinyImages/>

# 3. Multi-modal Social Media

---

**Project idea:** In this project, we encourage students to infer the underlying relations between different modalities of information on the Web. Here are some examples.

(1) Given a photo of movie poster (image), can we retrieve related trailers (video) or latest news articles (documents) of the movie?

To make project simpler, we recommend focusing on less than five movies (eg. 'Rise of the planet of the apes' and 'The smurfs'). You first download posters and trailers from some well-organized sites such as [imdb.com](http://imdb.com) or [itunes.com](http://itunes.com). They will be used as training data to learn your classifiers. Now your job is to gather raw data from [youtube.com](http://youtube.com) or Flickr, and classify them. In this project, we encourage you to explore the possibility to build classifiers to be learned from one information modality (eg. images), and to be applicable to other modalities (eg. trailer videos).

---

# 3. Multi-modal Social Media

---

(2) Given a beer label (image), can we search for which frames of a given video clip the logo or bottle appears?

Suppose that you are a big fan of Guinness beer. You can easily download the clean Guinness logo or cup images by Google image search. These images can be used to learn your detector, which can discover the frames that the logo appears in the video clips. For testing, you can download some video clips from youtube.com.

The above examples are just two possible candidates, and any new ideas or problem definitions are welcome.

For this purpose, one may take advantage of some source codes available on the Web as unit modules (e.g., near-duplicated image detection, object recognition, action recognition in video).

Another interesting direction is to improve the current state-of-the-arts methods by considering more practical scenarios.

---

# 3. Multi-modal Social Media

---

## Related Papers and Software::

- A good example of how a machine learning technique is successfully applied to real systems (ex. Google news recommendation).
- [1] Das, Datar, Garg, Rajaram. Google news personalization: scalable online collaborative filtering. WWW 2007.
- One of most popular approaches to near duplicated image detection is LSH families.
- [2] <http://www.mit.edu/~andoni/LSH/> (This webpage links several introductory articles and source codes).
- Various hashing techniques in computer vision (papers and source codes).
- [3] Spectral Hashing (<http://www.cs.huji.ac.il/~yweiss/SpectralHashing/>)
- [4] Kernelized LSH (<http://www.eecs.berkeley.edu/~kulis/klsh/klsh.htm>)
- Recognition in video
  - [5] Naming of Characters in Video (<http://www.robots.ox.ac.uk/~vgg/data/nface/index.html>)
  - [6] Action recognition in Video  
(<http://www.robots.ox.ac.uk/~vgg/data/stickmen/index.html>)
- Recognition in images
  - [7] Human pose detection (Poselet) (<http://www.eecs.berkeley.edu/~lbourdev/poselets/>)
  - [8] General object detection (<http://people.cs.uchicago.edu/~pff/latent/>)

# Other Projects

---

4. Image Segmentation
5. Face Recognition
6. Text Classification
7. Question Answering
8. Image Denosing/Super-resolution
9. Image Retrieval
10. Tracking
11. .....

**Note that** not all topics are equally difficult. But we'll take this into consideration when evaluating each team's performance and make the assessment as fair as possible.

# Acknowledgement

---

Some of the materials in these slides are drawn inspiration from:

- Alexander Amini, MIT University, Introduction to Deep Learning course
- Hung-yi Lee, National Taiwan University, Machine Learning and having it Deep and Structured course
- Chenglin Liu, CASIA, Pattern Recognition course
- Fei-Fei Li, Standord University, CS231n Convolutional Neural Networks for Visual Recognition course

# Questions?

---

Thank You !



WeChat Group for Deep Learning