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

2022

深度学习

Lecture 1：简 介

王 亮

智能感知与计算研究中心(CRIPAC)
模式识别国家重点实验室(NLPR)
中科院自动化研究所(CASIA)

目录

课程目标

- 在现实世界的不同任务和数据集合的背景下，初步了解深度学习这样一种代表性的多变量数据分析方法
- 理解深度学习算法的原理
- 对于不同类型的问题，如何选择合适的深度学习方法
- 如何将现实任务抽象为学习的问题
- 怎样利用现有工具来实现设计的模型

前序课程基础

- 微积分与线性代数
- 概率论与数理统计
- 机器学习
- 程序设计语言(例如：Python, C, C++等)

课程大纲

1. 课程简介
2. 数学基础
3. 前馈网络
4. 卷积神经网络
5. 循环神经网络
6. 正则化方法及优化算法
7. 深度生成式网络
8. 强化学习
9. 注意及记忆机制
10. 图神经网络
11. 深度学习相关应用

课程要求

- 课堂讲授(3小时/周) + 课程设计作业 + 文献阅读
- 最终成绩 = 50% 期末考试 + 25% 课程设计作业 + 25% 文献阅读
- 课程设计作业需要分组进行，每组原则上要求5-10人，最终需要提交按照规范格式完成的技术报告
- 课程设计包含presentation环节，鼓励大家主动申请进行报告（有加分），如最终申请人数较少，将采取随机抽签的方式进行

课程要求

- 课程设计作业鼓励大家提前进行分组，充分讨论交流，互相协作完成；
- 小组中的每位同学都应有明确的任务划分，独立思考并完成所承担的部分，最终需要在课程报告中明确指出不同成员的分工情况；
- 严禁剽窃，抄袭；
- 注意学术道德问题及引用规范；

课程要求

课程设计作业评价规则:

1. 研究意义及动机是否明确?
2. 方案是否可行有效?
3. 整个解决方案是否完整，是否存在缺陷?
4. 是否具备创新性?
5. 技术报告是否逻辑清晰，叙述准确?

课程要求

课程设计作业技术报告要求4-8页，提交电子版；原则上要求按照如下结构完成，英文中文均可，鼓励使用英文。

- 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

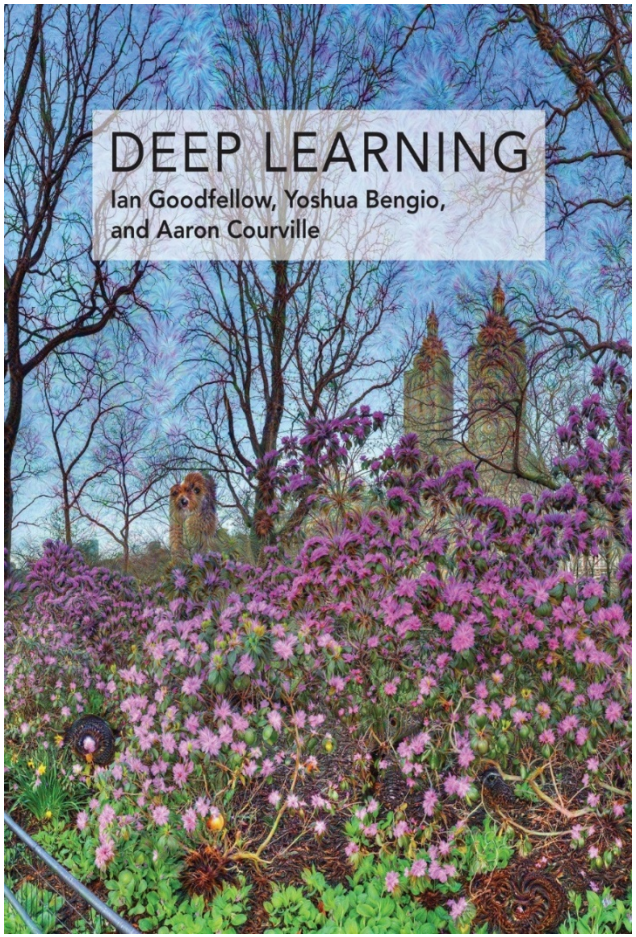
课程要求

- 技术报告推荐格式：CVPR的论文格式，地址如下:

https://cvpr2022.thecvf.com/author-guidelines

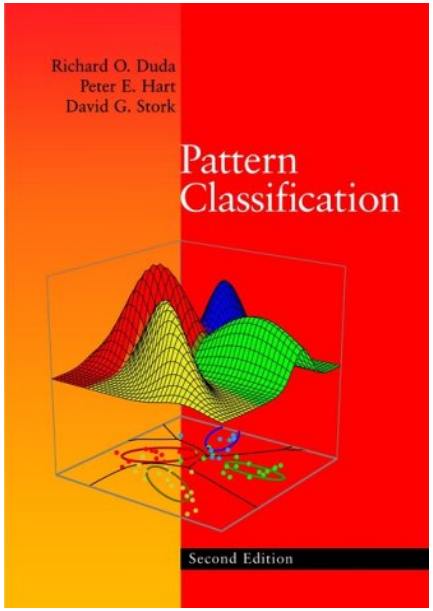
- 鼓励提交课程设计代码，要求给出详细的readme.txt代码说明文档，给出其参数配置及运行方法；

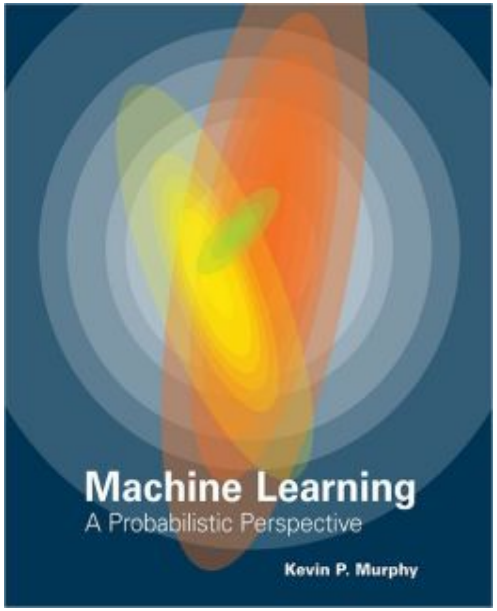
参考教材

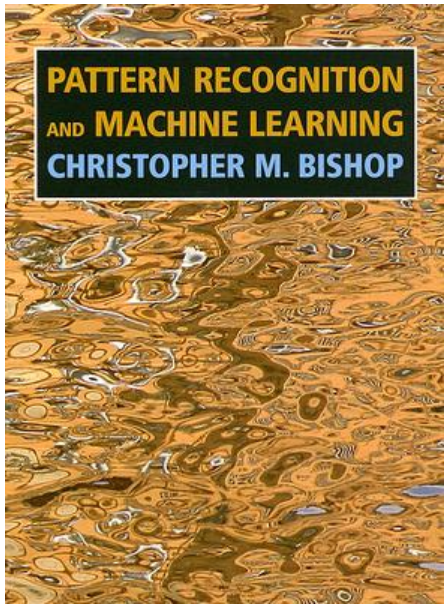


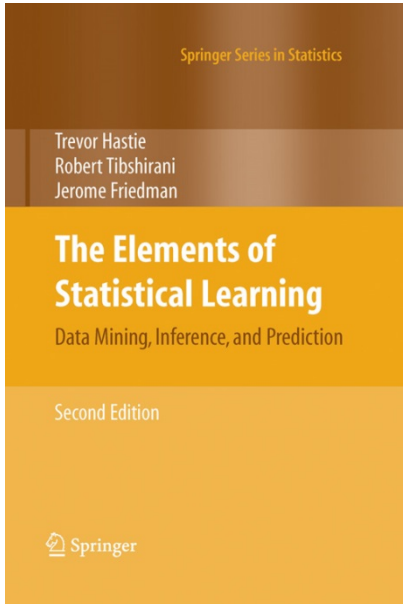
http://www.deeplearningbook.org/lecture_slides.html

经典文献


Richard O. Duda
Peter E. Hart
David G. Stork
Pattern Classification
Second Edition


Machine Learning
A Probabilistic Perspective
Kevin P. Murphy


PATTERN RECOGNITION
AND MACHINE LEARNING
CHRISTOPHER M. BISHOP


Springer Series in Statistics
Trevor Hastie
Robert Tibshirani
Jerome Friedman
The Elements of Statistical Learning
Data Mining, Inference, and Prediction
Second Edition
Springer

相关资源

- https://pytorch.org/tutorials/ (Learning Pytorch)
- https://www.tensorflow.org/tutorials (Learning Tensorflow)
- http://scikit-learn.org/stable/ (Scikit Learn: a Python library for machine learning)

相关资源

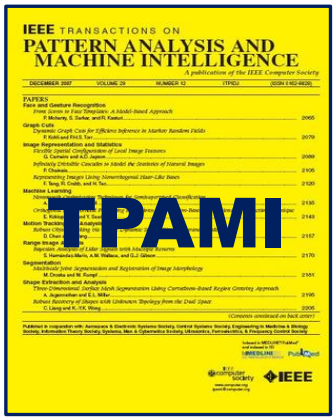
Name	Language	Link	Note
TensorFlow	Python/C++/Java	http://www.tensorflow.org	A deep learning based python library
Pytorch	Python/C++/Java	https://pytorch.org/	A python deep learning library
Caffe	C++	http://caffe.berkeleyvision.org/	A deep learning framework by Berkeley
Overfeat	Lua	http://cilvr.nyu.edu/doku.php?id=code:start	A convolutional network image processor
Deeplearning4j	Java	http://deeplearning4j.org/	A commercial grade deep learning library
Word2vec	**C**	https://code.google.com/p/word2vec/	Word embedding framework
GloVe	C	http://nlp.stanford.edu/projects/glove/	Word embedding framework
Doc2vec	C	https://radimrehurek.com/gensim/models/doc2vec.html	Language model for paragraphs and documents
StanfordNLP	Java	http://nlp.stanford.edu/	A deep learning-based NLP package
DGL	Python	https://www.dgl.ai/	Deep Graph Library
Pyg	Python	https://pytorch-geometric.readthedocs.io/	Geometric deep learning extension library

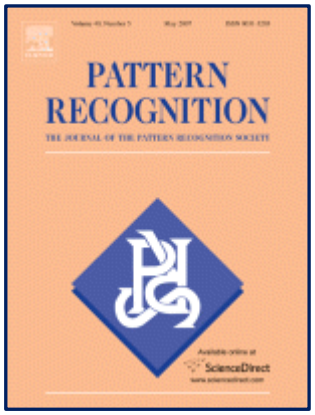
相关资源

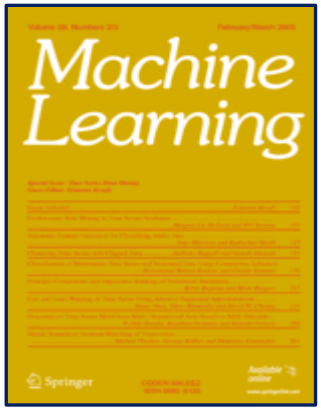
推荐阅读最新的机器学习和人工智能方向的研究论文，领域内顶级学术会议包括：

- **NeurIPS**：Advances in Neural Information Processing Systems
- **ICML**：International Conference on Machine Learning
- **ICLR**：International Conference on Learning Representations
- **CVPR**：IEEE Conference on Computer Vision and Pattern Recognition
- **ICCV**：IEEE International Conference on Computer Vision
- **AAAI**：AAAI Conference on Artificial Intelligence
- **KDD**：SIGKDD Conference on Knowledge Discovery and Data Mining
- ……

相关资源


IEEE TRANSACTIONS ON
PATTERN ANALYSIS AND MACHINE INTELLIGENCE
TPAMI


PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY


Machine Learning


KNOWLEDGE AND DATA ENGINEERING
TKDE


2018
CVPR
SALT LAKE CITY • JUNE 18-22


ICCV17


NIPS 2018
Monday December 03 – Saturday December 08, 2018
Palais des Congrès de Montréal, Montréal CANADA
BY-NC-SA by Paco CT


ECCV 2018
European Conference on Computer Vision
8 – 14 September 2018 | Munich, Germany


ICML | 2018
Thirty-fifth International Conference on Machine Learning

目录

深度学习

Artificial Intelligence

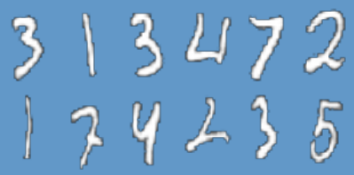
Any technique that enables computers to mimic human behavior

Machine Learning

Ability to learn without explicitly being programmed

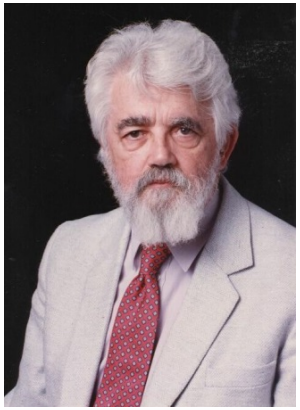
Deep Learning

Learn underlying features in data using neural networks



人工智能(AI)学科起源

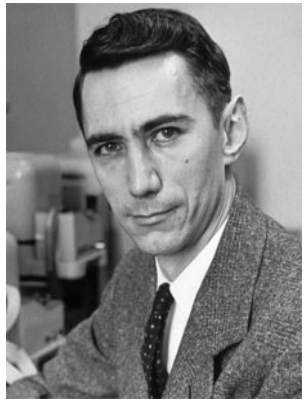
1956年夏天约翰·麦卡锡等人在美国达特茅斯学院开会研讨“如何用机器模拟人的智能”，会上提出“人工智能”这一概念，标志着人工智能学科的诞生



约翰·麦卡锡
John McCarthy
达特茅斯学院

马文·明斯基
Marvin Minsky
哈佛大学

纳撒尼尔·罗彻斯特
Nathaniel Rochester
IBM公司

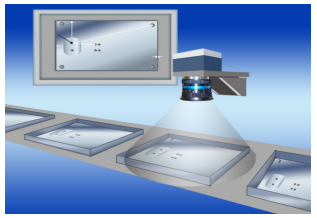


克劳德·香农
Claude Shannon
贝尔电话实验室

AI的基本概念

研究目的：探寻智能本质，研制出具有类人智能的智能机器

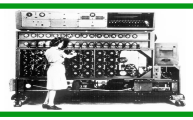
研究内容：能够模拟、延伸和扩展人类智能的理论、方法、技术及应用系统



表现形式：

- **会看**：图像识别、文字识别、车牌识别
- **会听**：语音识别、说话人识别、机器翻译
- **会说**：语音合成、人机对话
- **会行动**：机器人、自动驾驶汽车、无人机
- **会思考**：人机对弈、定理证明、医疗诊断
- **会学习**：机器学习、知识表示

AI发展历史

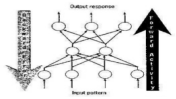


1936:**Turing Machine from Alan Turing**

1946: **Birth of ENIAC**

1956: Birth of AI

1966: **Turing Award**

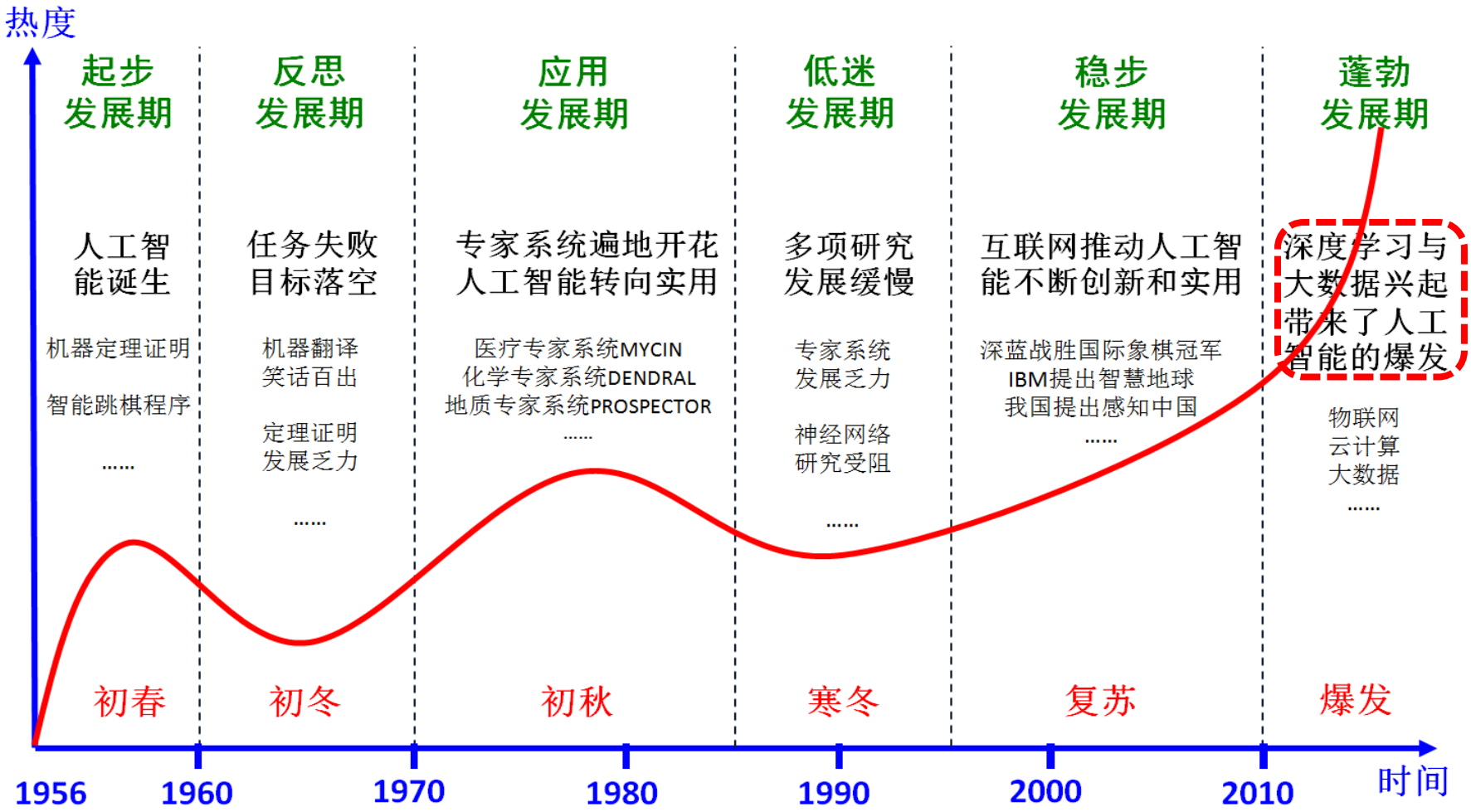


1986: **The BP Algorithm**

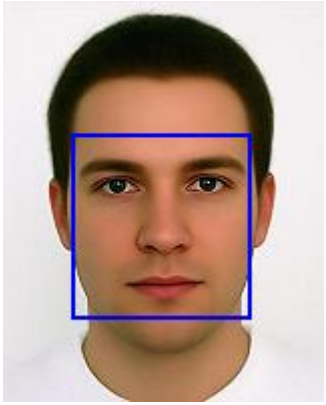
2006: **DNN, Deep Learning**

AlphaGo **2016: AlphaGo**

AI发展历史


热度
起步发展期
人工智能诞生
机器定理证明
智能跳棋程序
.....
初春
1956
反思发展期
任务失败目标落空
机器翻译笑话百出
定理证明发展乏力
.....
初冬
1960
1970
应用发展期
专家系统遍地开花
人工智能转向实用
医疗专家系统MYCIN
化学专家系统DENDRAL
地质专家系统PROSPECTOR
.....
初秋
1980
低迷发展期
多项研究发展缓慢
专家系统发展乏力
神经网络研究受阻
.....
寒冬
1990
稳步发展期
互联网推动人工智能不断创新和实用
深蓝战胜国际象棋冠军
IBM提出智慧地球
我国提出感知中国
.....
复苏
2000
蓬勃发展期
深度学习与大数据兴起带来了人工智能的爆发
物联网
云计算
大数据
.....
爆发
2010
时间

人类感知



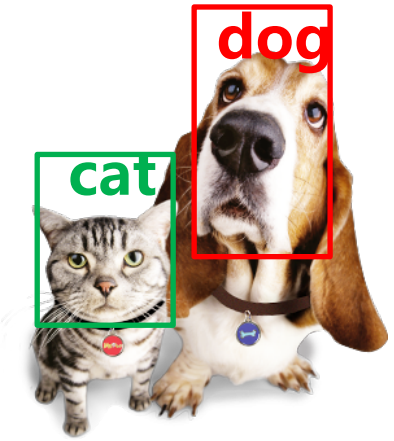
人脸识别

语音识别

人类已经发展出高度复杂的技能来感知环境，并根据他们所观察到的内容采取行动

我们希望赋予机器类似的功能

手写字体识别



物体识别

机器学习初识

A pattern is a defined entity that could be given a name：模式是可以指定名称的已定义实体

- 指纹,
- 手写字符,
- 人脸,
- 语音,
- DNA序列,
- …

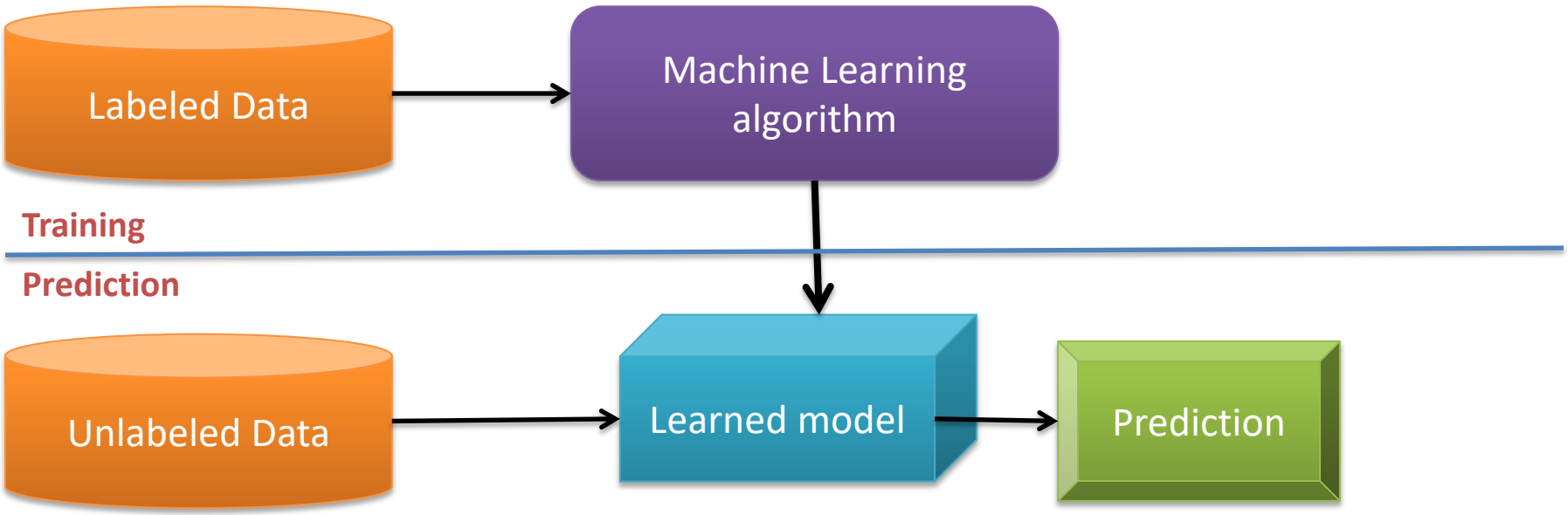
机器学习是研究机器如何能够：

- 感知环境,
- 学会区分特定模式,
- 对模式的类别做出合理的决策

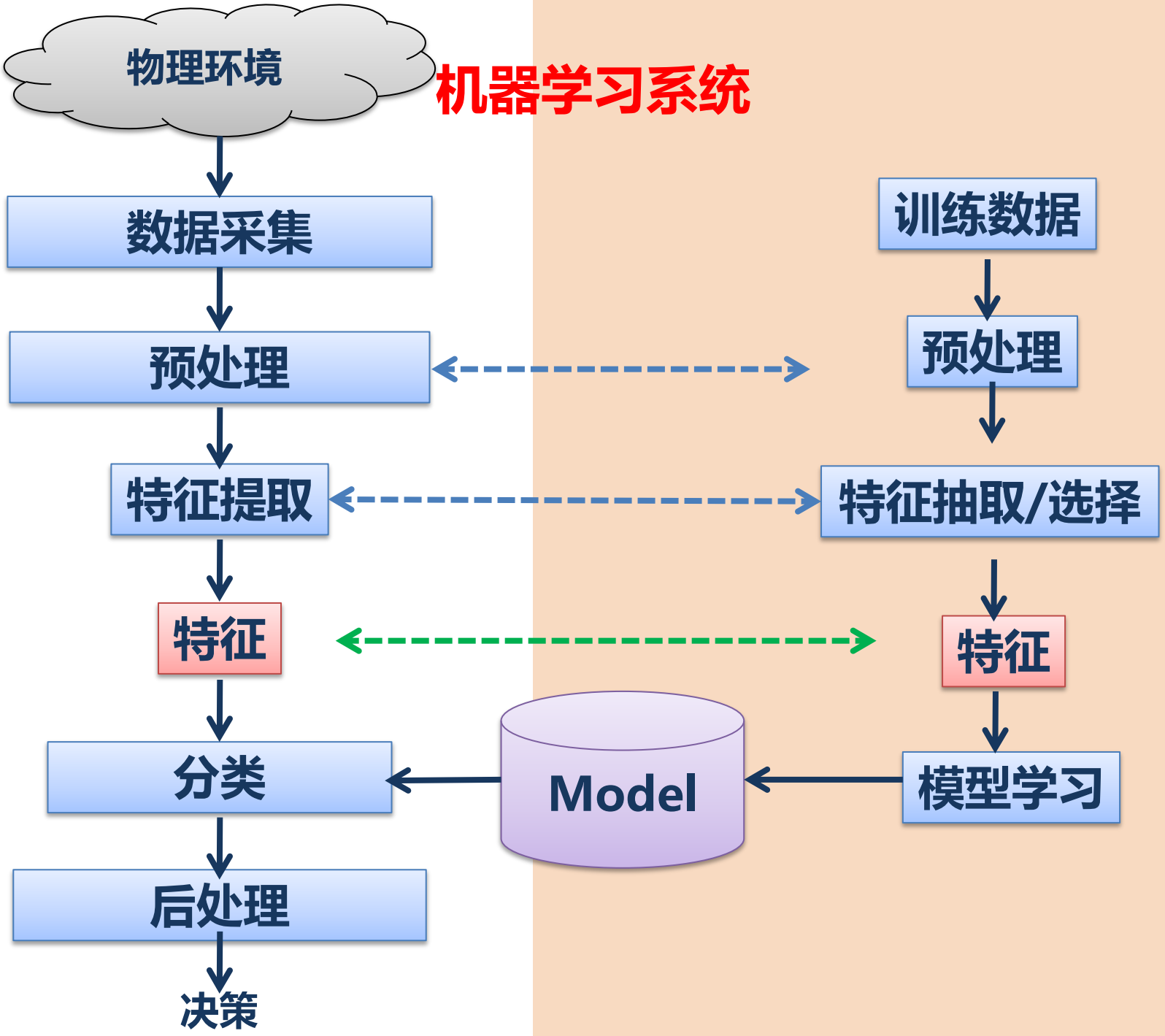
机器学习基础

Arthur Samuel (1956): 在不直接针对问题进行编程的情况下，赋予计算机学习能力的一个研究领域。

Tom Mitchell (1997): 对于任务T和性能度量P，如果计算机程序在T上以P衡量的性能随着经验E而自我完善，那么就称这个计算机程序从经验E学习。



能够从数据中进行学习，并且根据数据进行预测的一类方法。


物理环境
机器学习系统
数据采集
预处理
特征提取
特征
分类
后处理
决策
训练数据
预处理
特征抽取/选择
特征
模型学习
Model

机器学习系统

数据获取：Data acquisition and sensing:

- 测量物理变量：Measurements of physical variables
- 需要关注: 带宽，分辨率，敏感度，失真，延迟...

预处理:

- 去除数据噪声
- 将感兴趣的模式与背景分离

特征提取:

- 找到合适的特征

机器学习系统

模型学习:

- 学习特征与模式类别之间的映射

分类:

- 使用提取到的特征和学习到的模型将模式划分为类别

后处理:

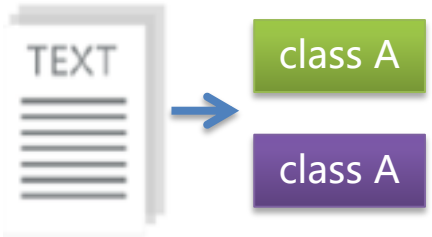
- 结果的可信度
- 利用整体信息提高性能
- 结合专家指导知识

机器学习算法分类

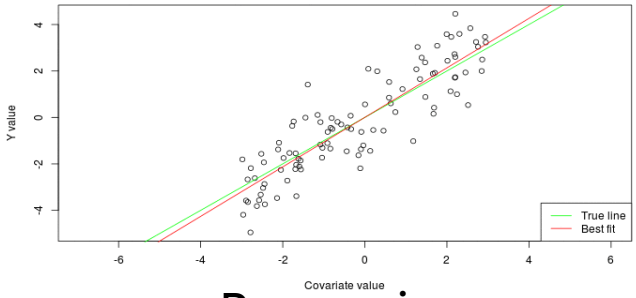
监督学习: Learning with a **labeled** training set, such as Classification, Regression.
Example: email *classification* with already labeled emails

无监督学习: Discover **patterns** in **unlabeled** data, such as Clustering, Dimensionality Reduction.
Example: *cluster* similar documents based on text

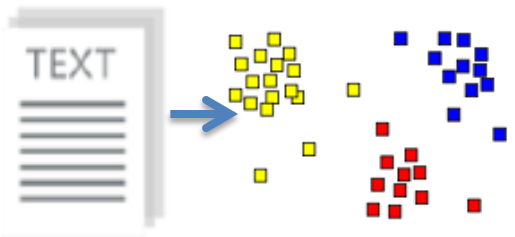
强化学习: learn to **act** based on **feedback/reward**
Example: learn to play Go, reward: *win or lose*



Classification



Regression



Clustering

机器学习方法分类

Supervised Learning

- Neural Networks
- Naïve Bayes
- SVM
- Decision tree algorithm
- Ensemble (Boosting)
- Markov random field (MRF)

……

Unsupervised Learning

- K-Means
- Gaussian mixture model
- PCA
- LDA
- PageRank

…….

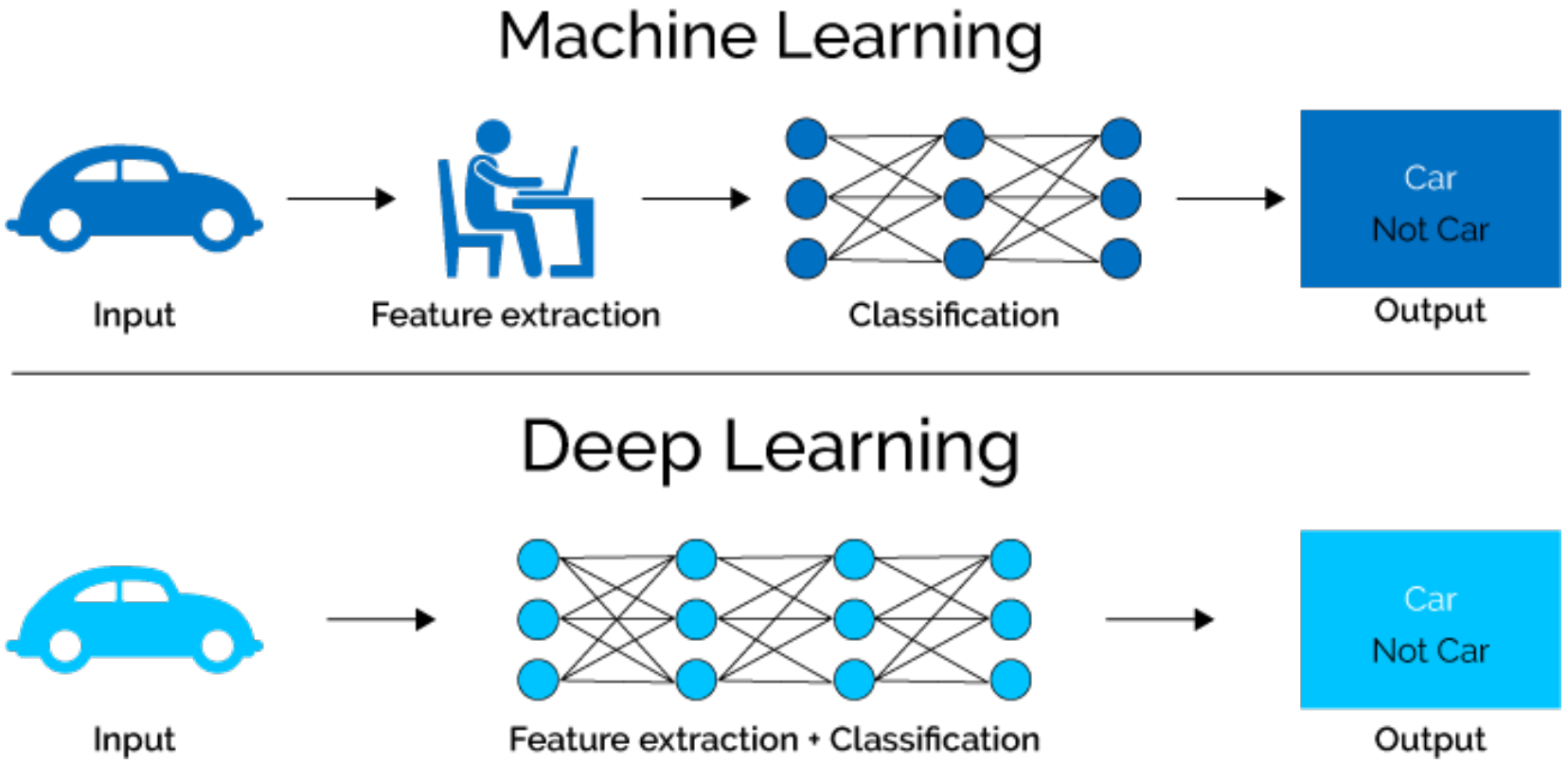
Reinforcement Learning

- Model-free RL
- Model-based RL

……

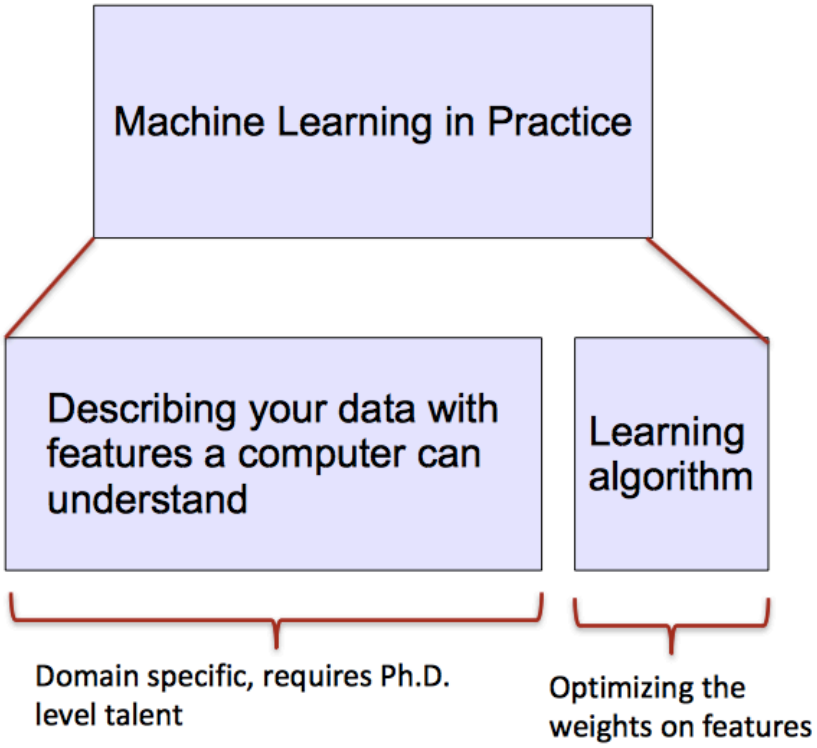
机器学习 vs 深度学习

深度学习是机器学习的一个分支，可以理解为包含多个隐含层的神经网络结构；其思想受人脑启发，通过使用多层的神经网络来自动从数据中学习特征；



机器学习 vs 深度学习

目前多数机器学习方法能产生较好效果的原因主要是**human-designed representations and input features**；
而机器学习算法在其中的作用主要是**optimize weights**，
以便改善最终预测的结果。



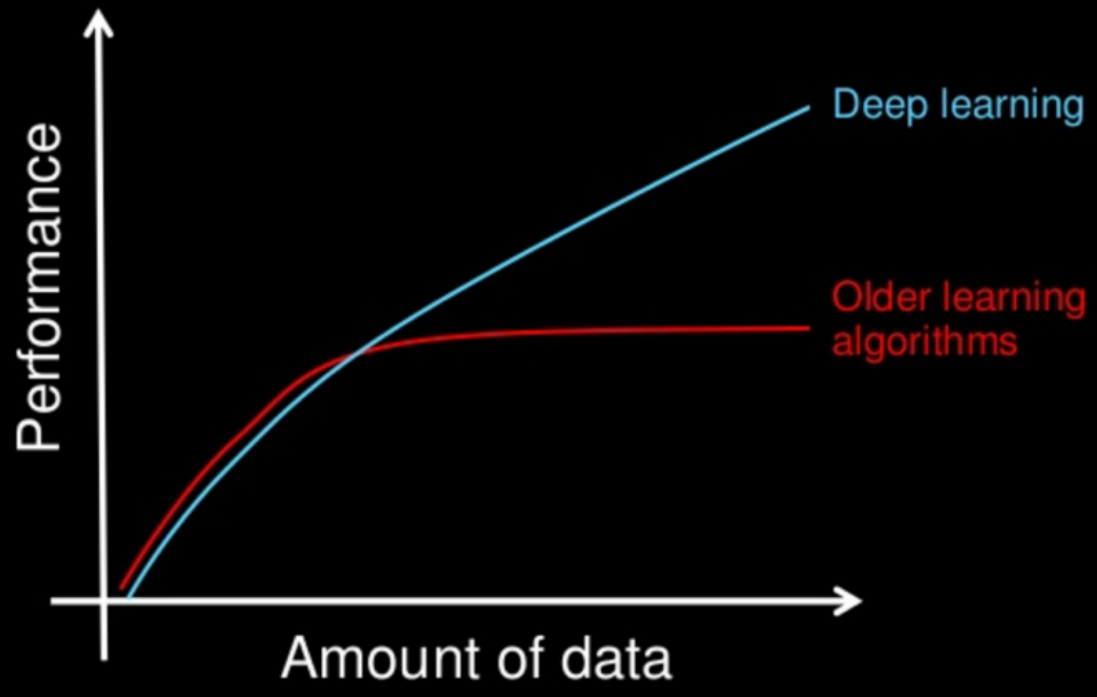
Feature	NER
Current Word	✓
Previous Word	✓
Next Word	✓
Current Word Character n-gram	all
Current POS Tag	✓
Surrounding POS Tag Sequence	✓
Current Word Shape	✓
Surrounding Word Shape Sequence	✓
Presence of Word in Left Window	size 4
Presence of Word in Right Window	size 4

机器学习 vs 深度学习

Difference between ML and DL

Machine Learning	Deep Learning
Achieves good results using small data	Needs large scale data
Training is faster	High-intensity computing
Needs feature engineering	Features and classifier be learned automatically
Accuracy reaches the bottleneck	Accuracy is infinite theoretically
Religious mathematics theory	Exists plentiful lack of theories

Why deep learning



How do data science techniques scale with amount of data?

发展现状

- 在**语音识别**方面的进展：
  - 利用深度学习类方法打破了许多长期维持的性能记录；
  - Microsoft 和 Google 都在其产品中部署了基于深度学习算法的语音识别系统；
- 在**计算机视觉**方面的进展：
  - 卷积神经网络在视觉任务中成为主流的基础网络结构；
  - 基于Transformer的网络逐渐在各项视觉任务中刷新最好效果；
  - 自监督、多模态数据融合等在计算机视觉中愈发重要；
- 在**自然语言处理**方面的进展：
  - 语言模型，词性标注，句法分析；
  - 细粒度情感分析，机器翻译，问答系统；
  - 大规模预训练模型；

发展现状


nature
THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE
At last — a computer program that can beat a champion Go player PAGE 484
ALL SYSTEMS GO
CONSERVATION
SONGBIRDS À LA CARTE
Illegal harvest of millions of Mediterranean birds
PAGE 452
RESEARCH ETHICS
SAFEGUARD TRANSPARENCY
Don't let openness backfire on individuals
PAGE 459
POPULAR SCIENCE
WHEN GENES GOT 'SELFISH'
Dawkins's calling card forty years on
PAGE 462
NATURE.COM/NATURE
28 January 2016 £10
Vol. 529, No. 7587

深度神经网络

- **起源:**
  - 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*
- **缺陷:**

参数量巨大->高计算复杂度

数据集规模有限->过拟合问题

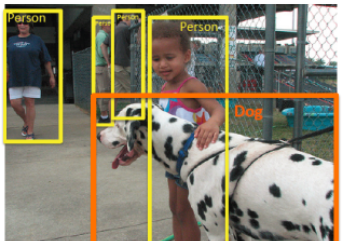
重要推动因素

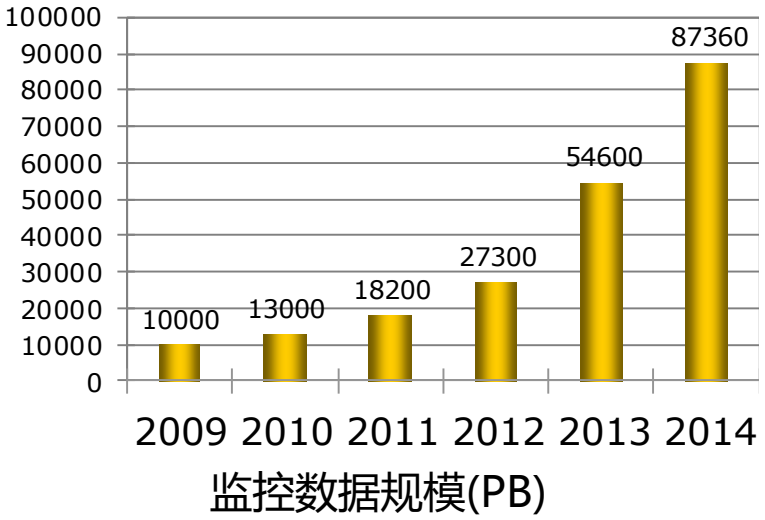
大数据

IMAGENET Large Scale Visual Recognition Challenge (ILSVRC) 2010-2013

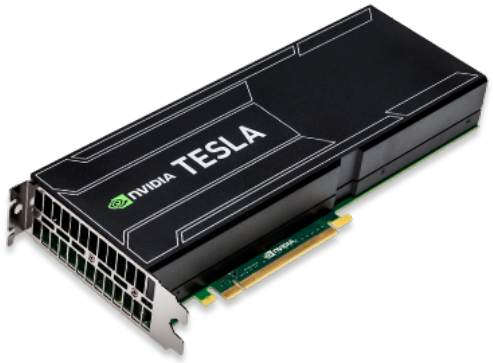
~~20 object classes~~ ~~22,591 images~~

200 object classes	456,191 images	DET NEW
1000 object classes	1,431,167 images	CLS-LOC





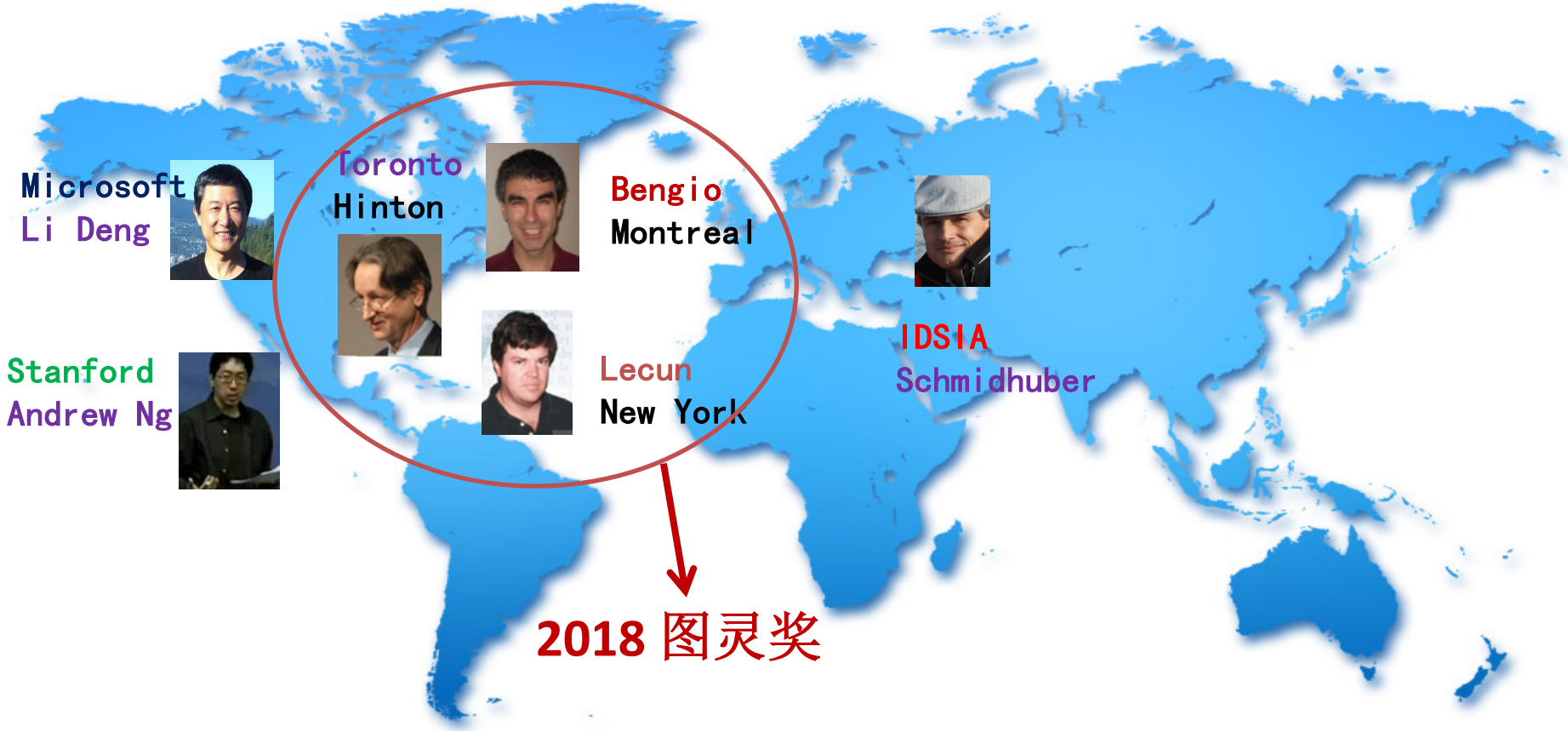
算力提升



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

深度神经网络可以有效的拟合

领域开拓者



引领深度学习经历3个主要阶段

深度学习：深度神经网络复苏

Breakthrough in 2006

Reducing the Dimensionality of Data with Neural Networks

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

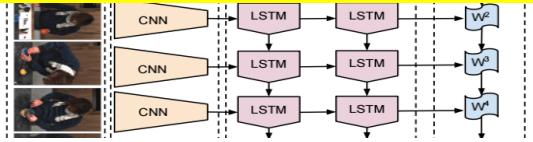
High-dimensional data can be converted to low-dimensional codes by training a multilayer neural network with a small central layer to reconstruct high-dimensional input vectors. Gradient descent can be used for fine-tuning the weights in such "autoencoder" networks, but this works well only if the initial 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

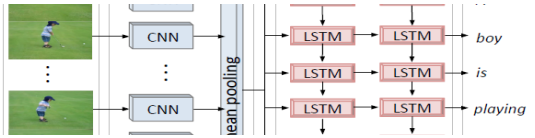
2006 VOL 313 SCIENCE www.sciencemag.org

ImageNet: 74% vs. 85%

RNN for sequence analysis



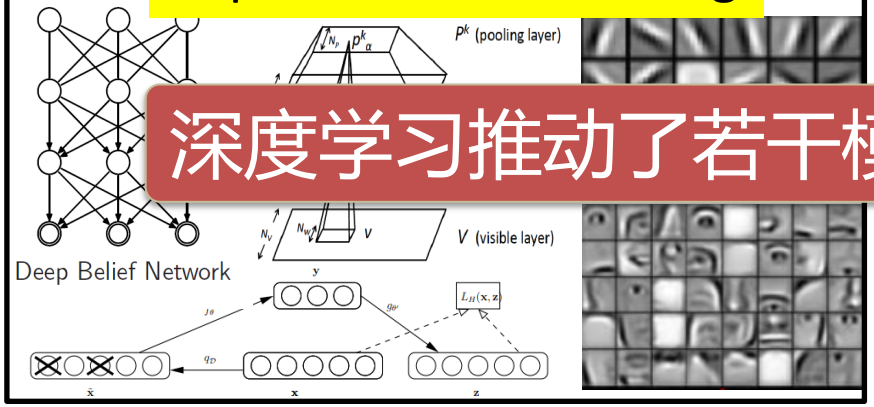
Activity recognition, CVPR2015



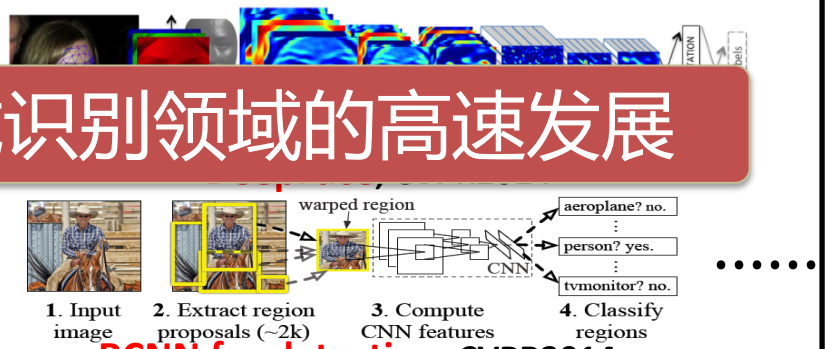
Video caption, CVPR2015

2006 2012 2015

Representation learning



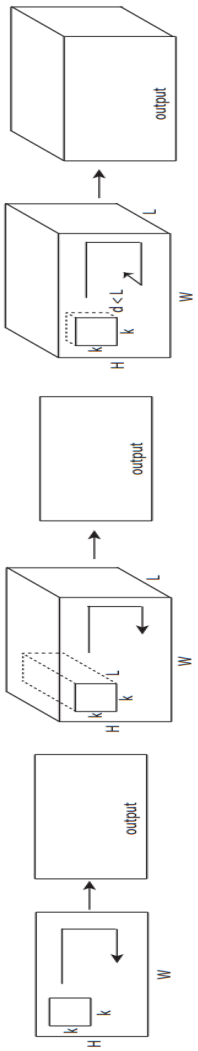
CNN for visual tasks



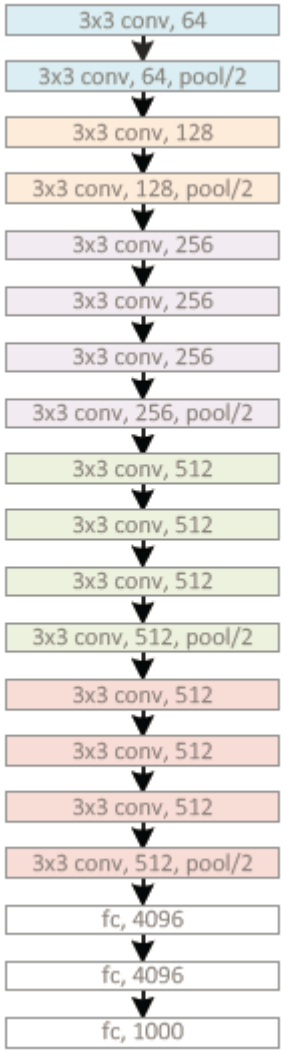
RCNN for detection, CVPR2014

深度学习推动了若干模式识别领域的高速发展

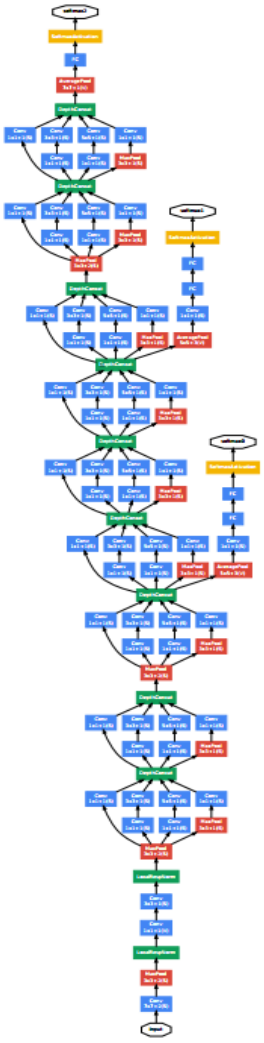
深度神经网络



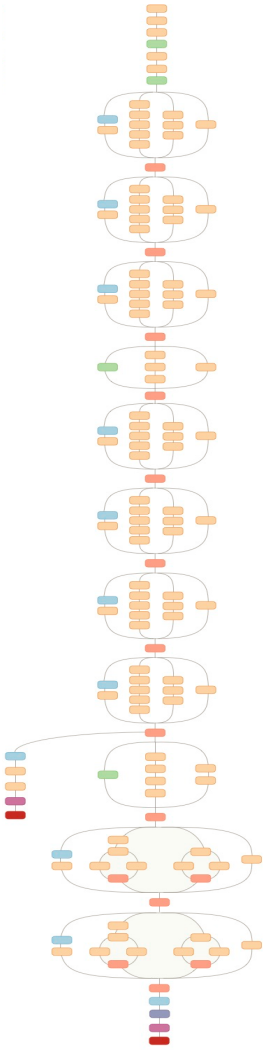
C3D



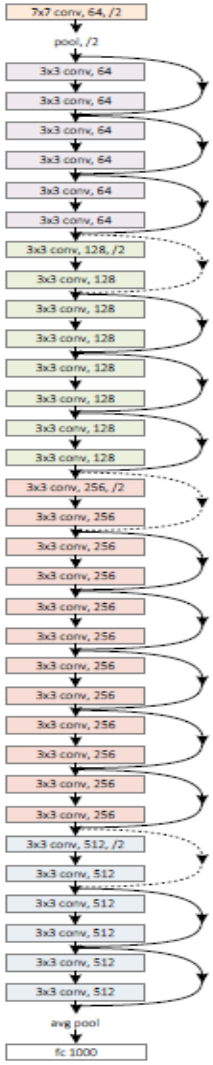
VGG-19



GoogleNet



Inception-3

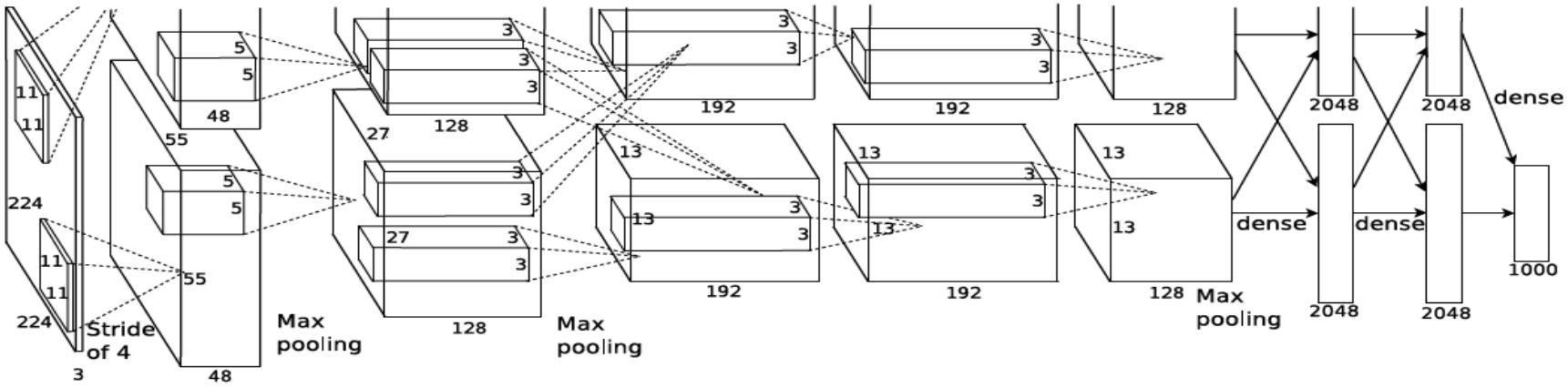


Residual

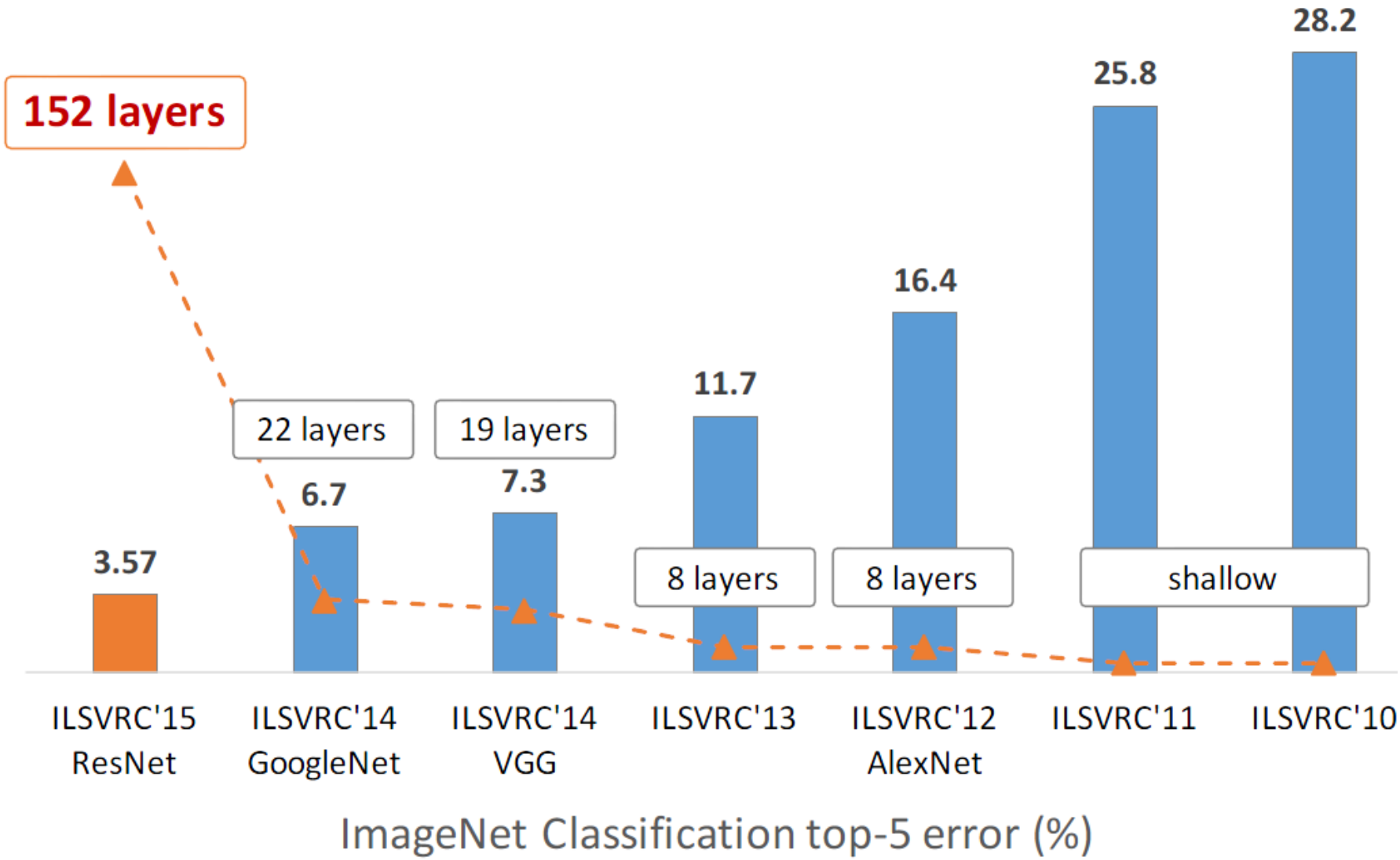
...

ImageNet竞赛

- 传统方法
  74% 2011
- 深度模型
  85% 2012
  89% 2013
  92% 2014



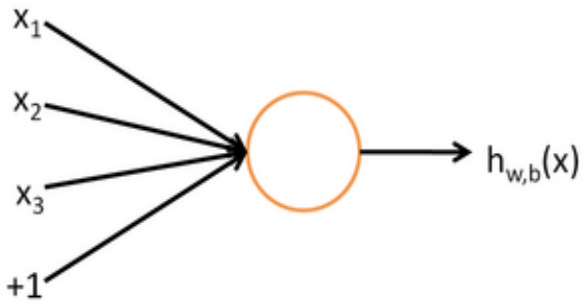
ImageNet竞赛



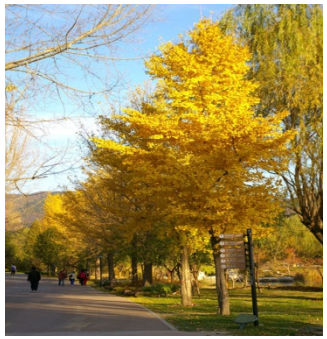
Slide Courtesy: Kaiming He

深度网络: 全连接

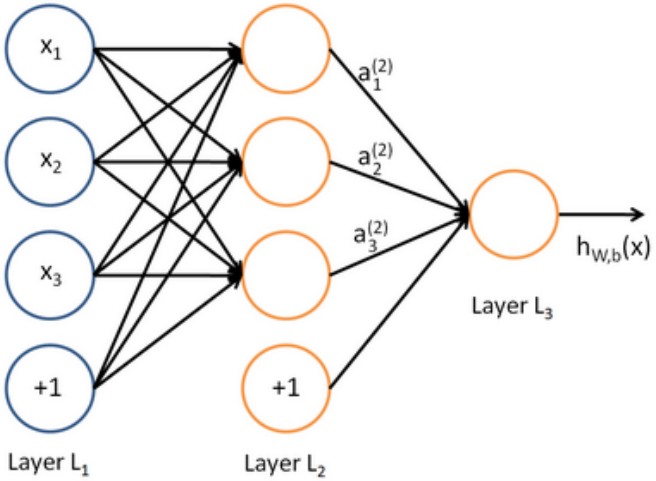
- 每一个单元与之前所有单元均存在连接



$$h_{W,b}(x) = f(W^T x) = f(\sum_{i=1}^{3} W_i x_i + b)$$



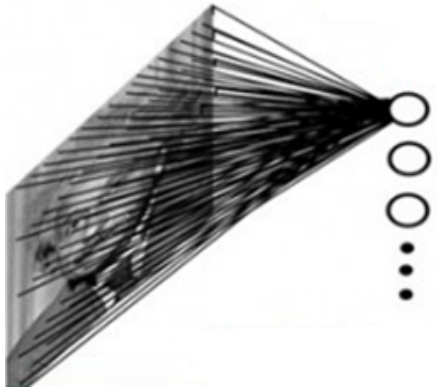
高维输入



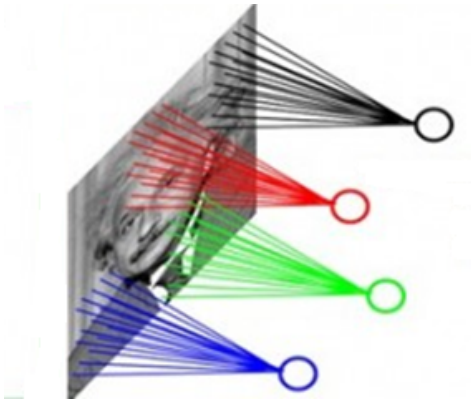
网络参数规模很大

卷积网络: 局部连接+权值共享

- 局部感受野

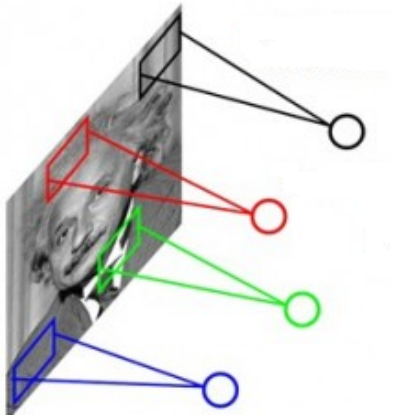


全连接

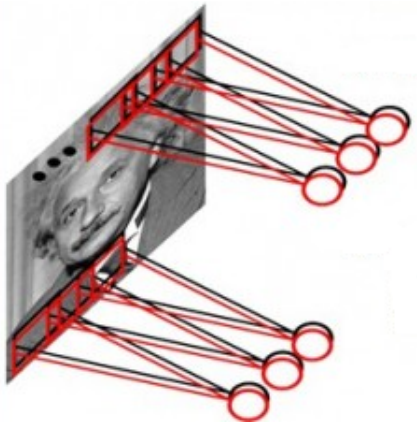


局部连接

- 所有局部区域共享相同的filter weights



局部连接

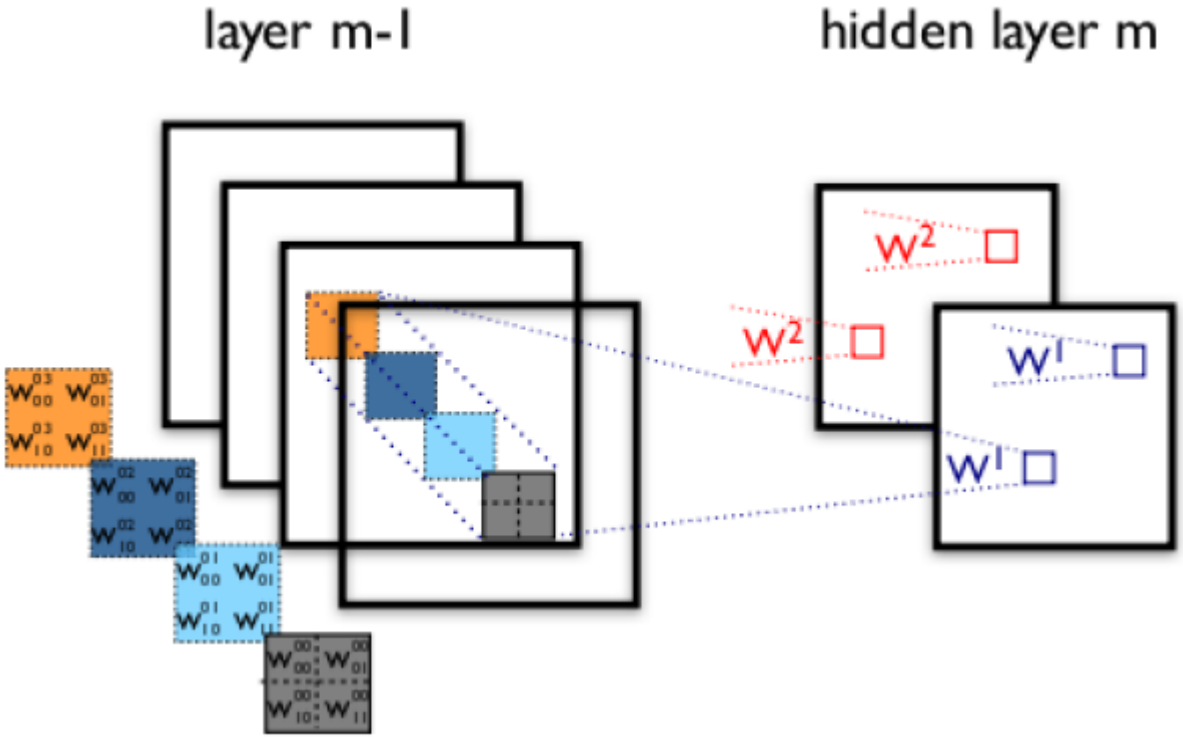


权值共享

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

卷积网络: 多个滤波器

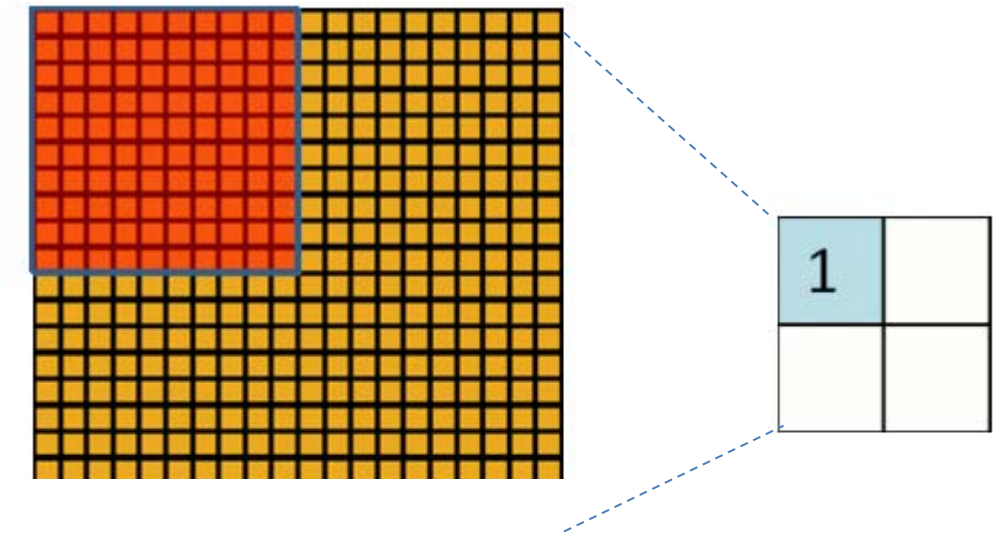
- 使用N个不同的filters获得N个feature maps



w_1: weights of the 1^{st} filter, w_2: weights of the 2^{nd} filter

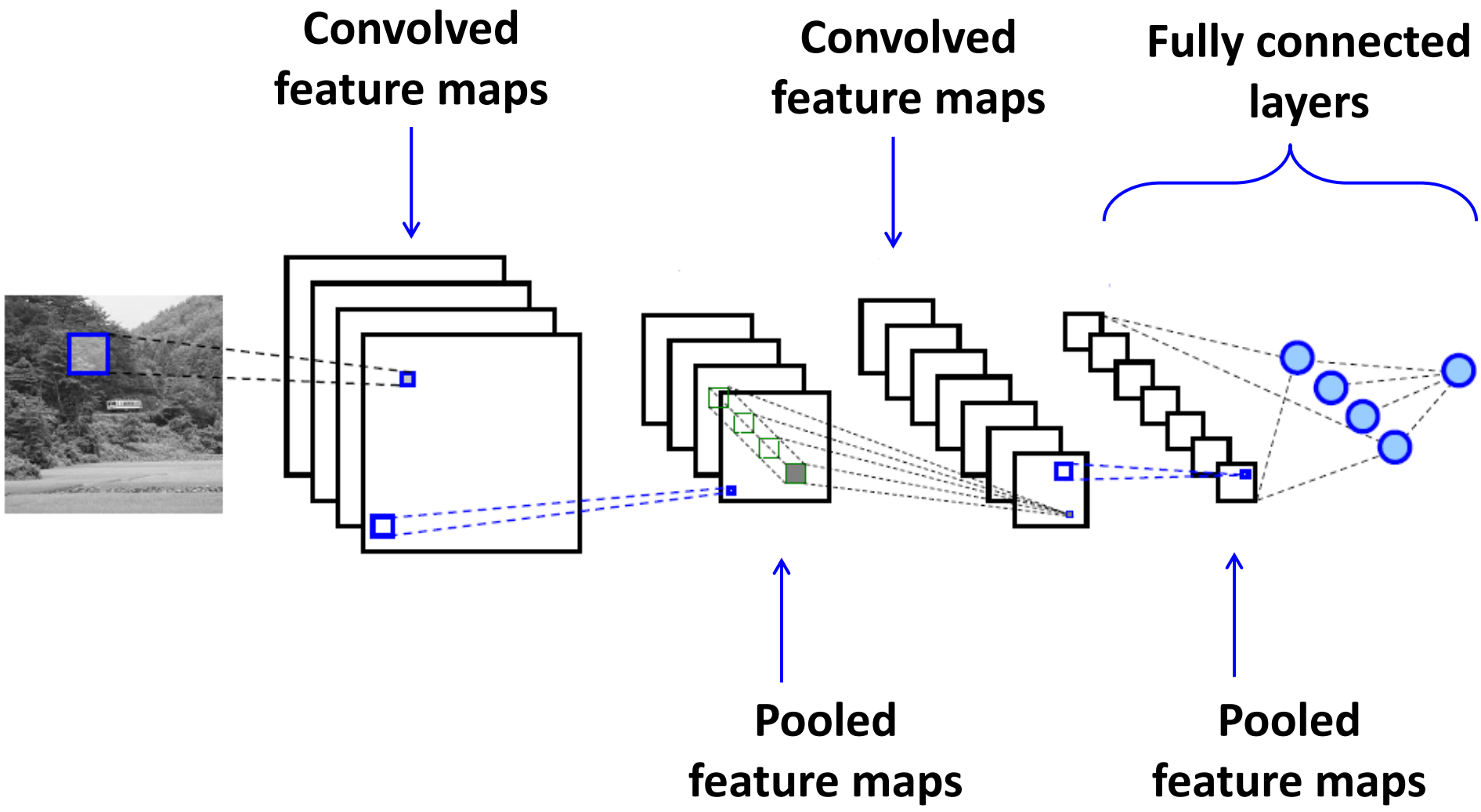
卷积网络: 池化

- 采用最大/均值池化来保持平移不变性

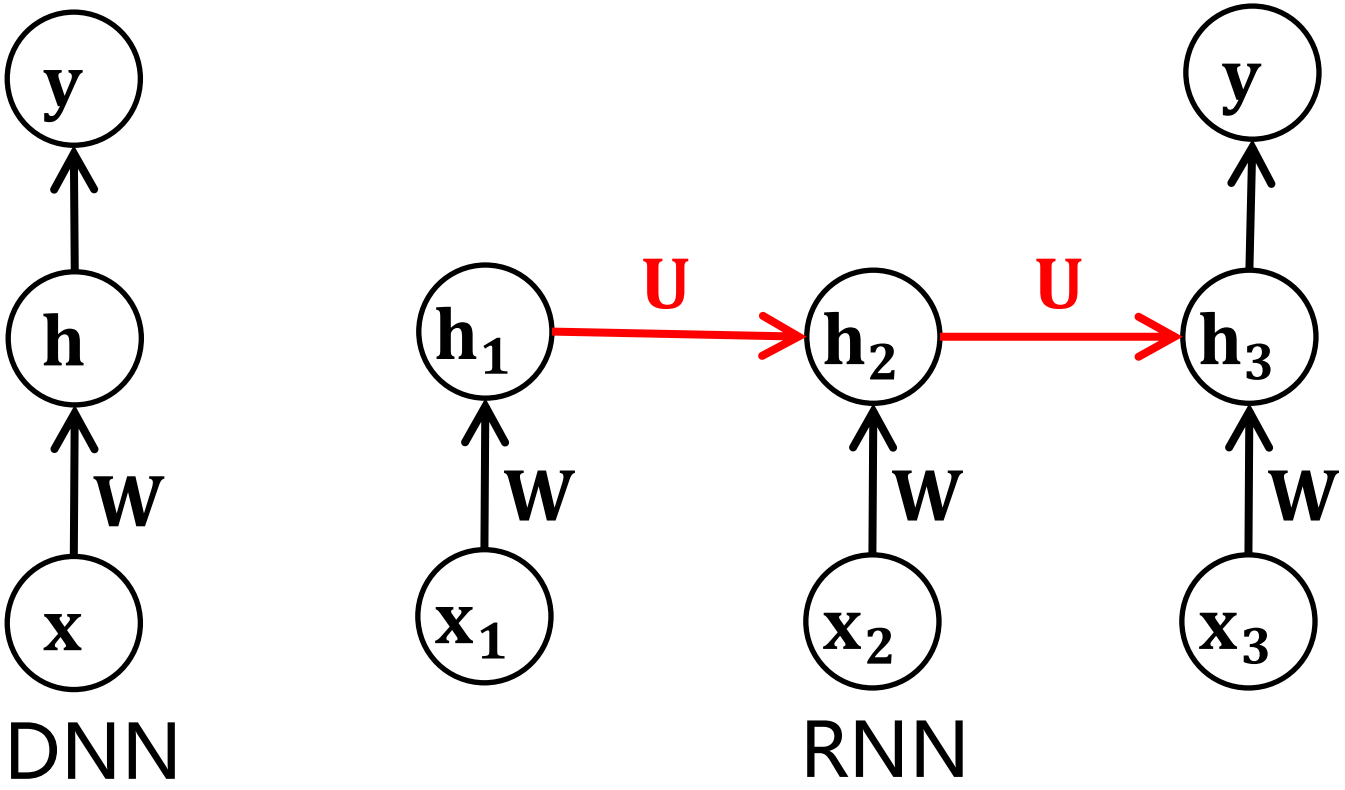


Convolved feature　　Pooled feature

整体结构


Convolved
feature maps
Convolved
feature maps
Fully connected
layers
Pooled
feature maps
Pooled
feature maps

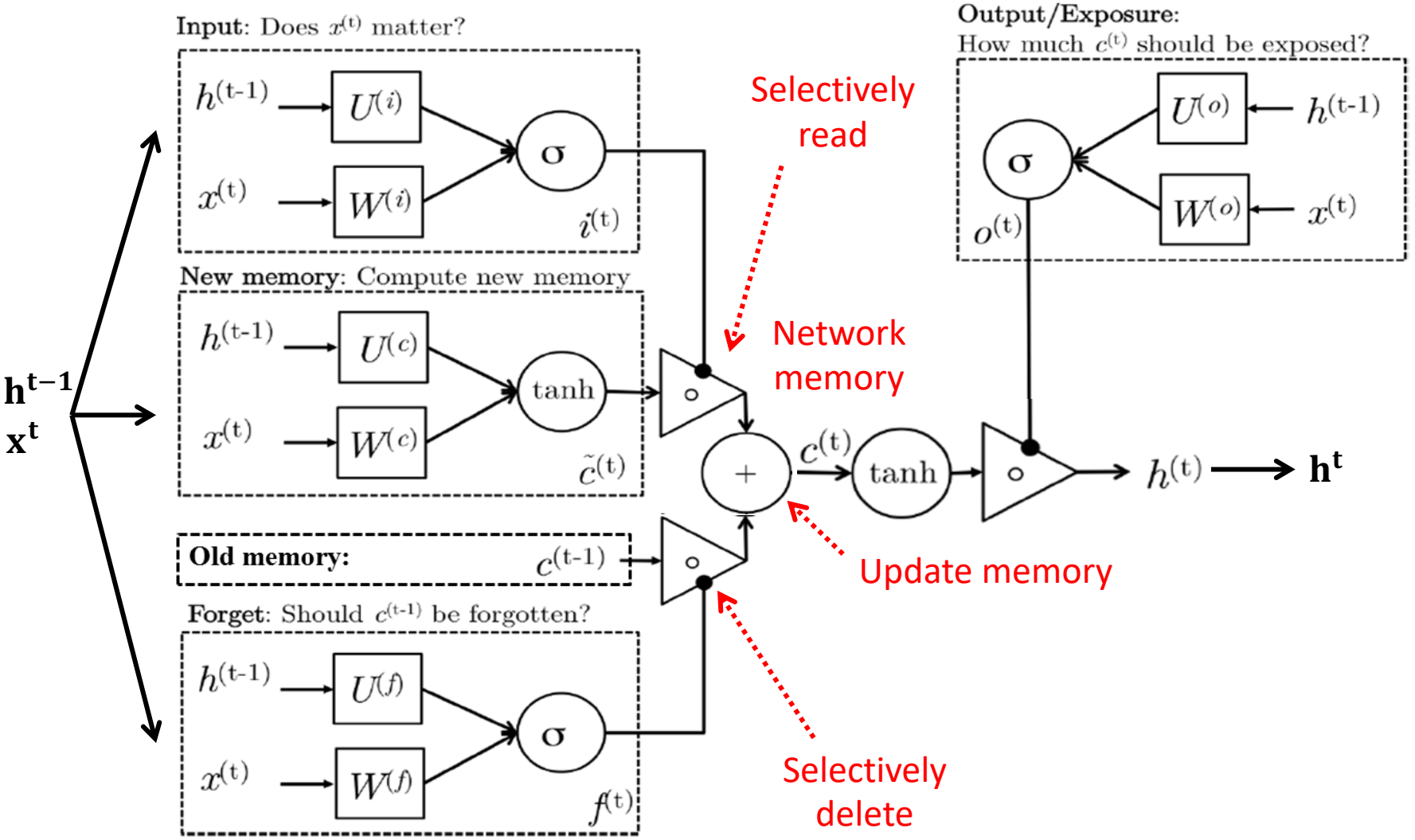
循环神经网络



- $\mathbf{x_t} \in \mathbb{R}^d, \mathbf{h_t} \in \mathbb{R}^n, \mathbf{W} \in \mathbb{R}^{d \times n}, \mathbf{U} \in \mathbb{R}^{n \times n}$
- $\mathbf{h}_t = \sigma(\mathbf{x}_t\mathbf{W} + \mathbf{h}_{t-1}\mathbf{U})$

循环神经网络用来建模前向长距离上下文关系

长短时记忆机制 Long Short-Term Memory

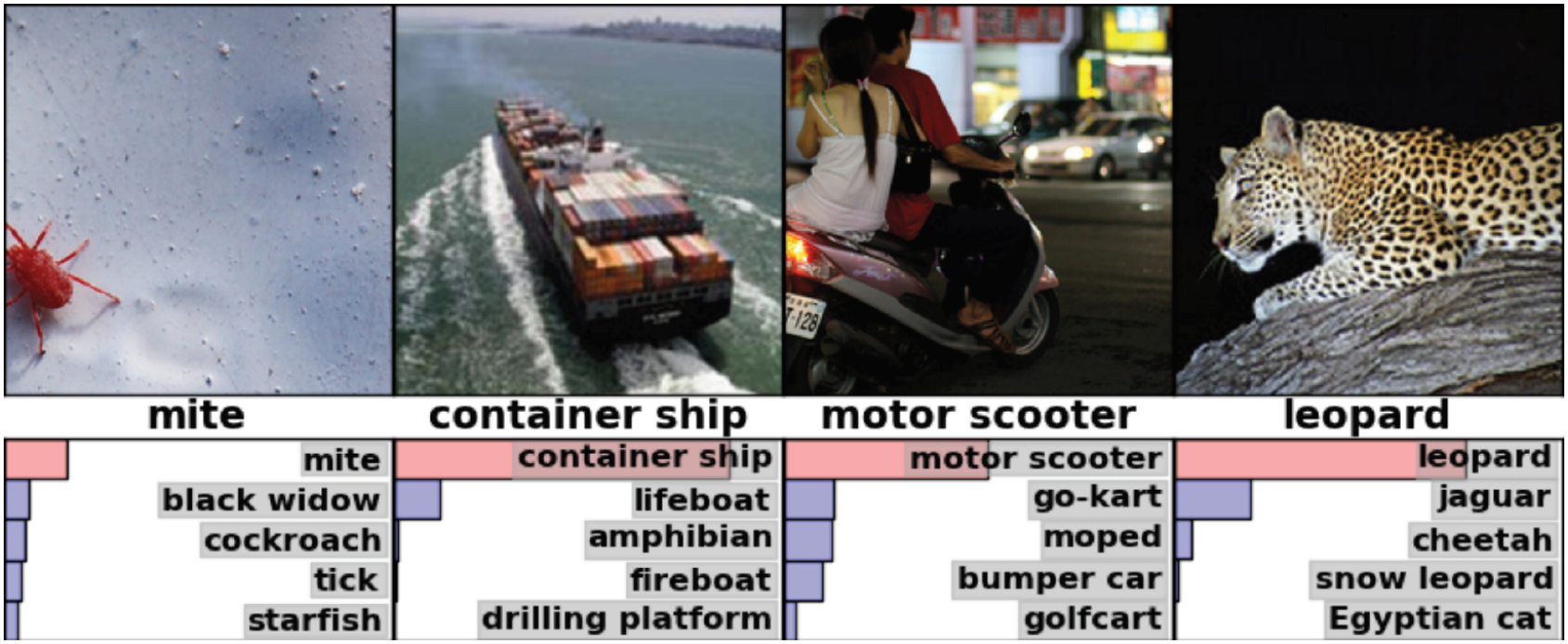


[Hochreiter and Schmidhuber, Neural computation1997]

目录

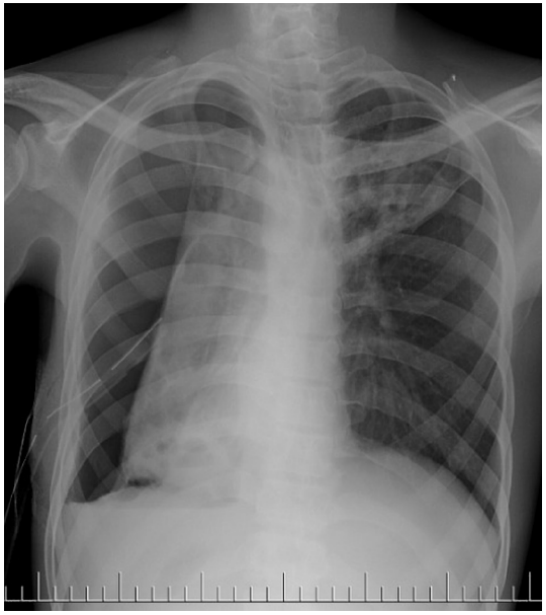
在视觉领域获得成功

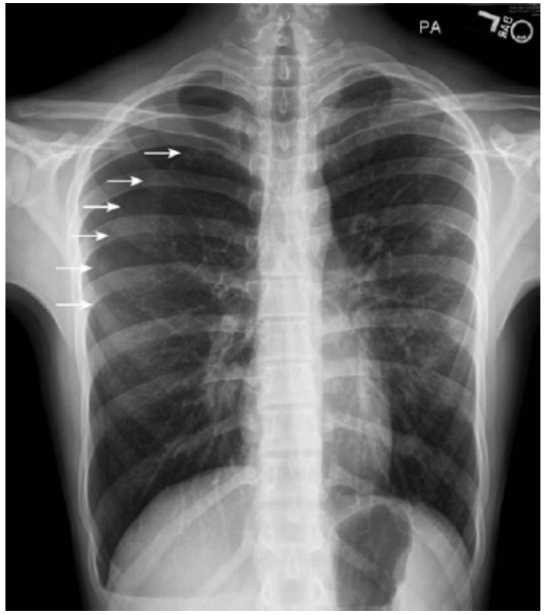
图像识别

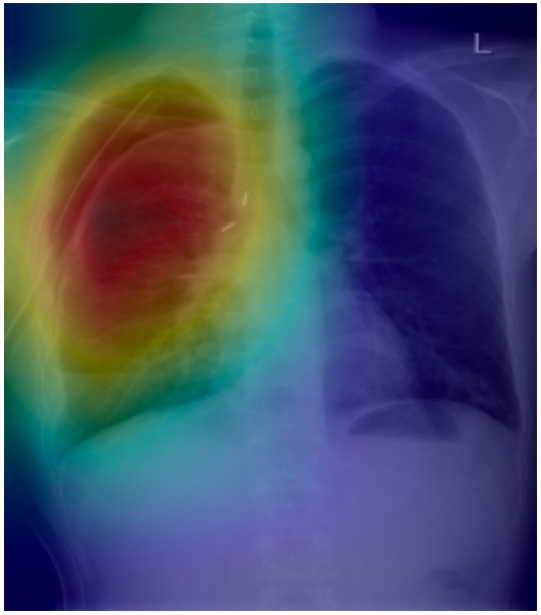


在视觉领域获得成功

辅助X光扫描进行疾病诊断

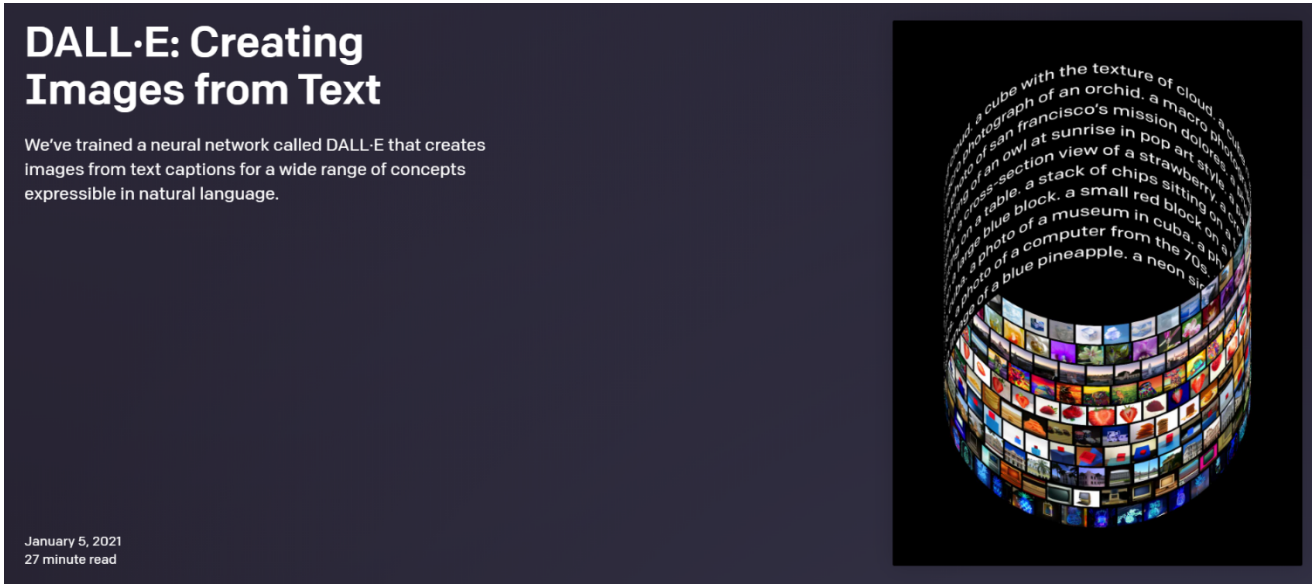






在自然语言处理获得成功

按照文字描述生成图片



TEXT PROMPT

an illustration of a baby daikon radish in a tutu walking a dog

AI-GENERATED IMAGES

https://openai.com/blog/dall-e/

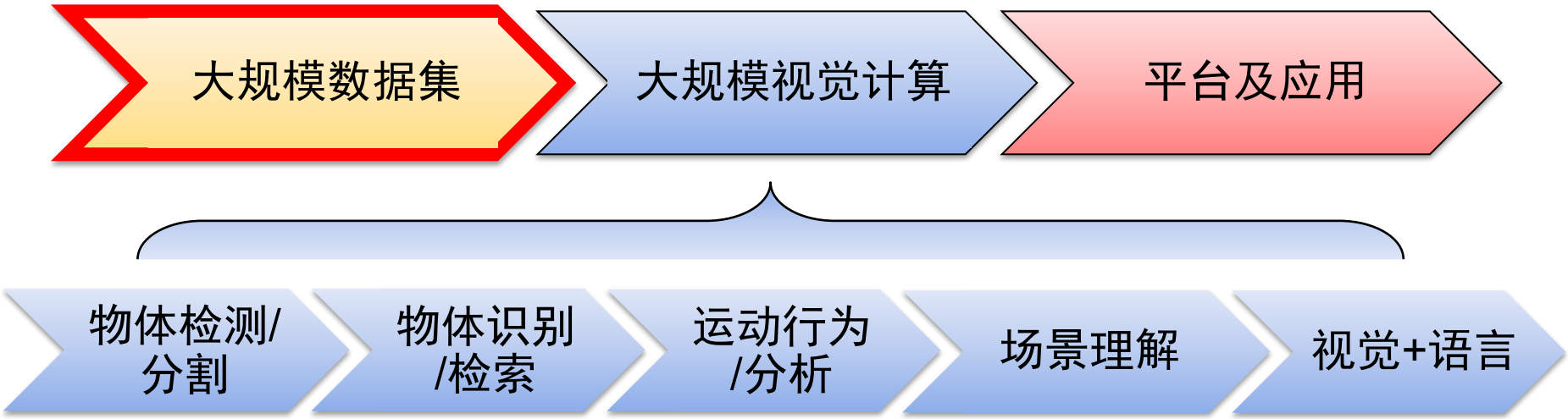
在语音领域获得成功

音乐生成

其他领域

更多.....

机器视觉



ImageNet

IMAGENET

14,197,122 images, 21,841 synsets indexed

Object detection (DET)[top]

Task 1a: Object detection with provided training data

Ordered by number of categories won

Team name	Entry description	Number of object categories won	mean AP
CUImage	Ensemble of 6 models using provided data	109	0.662751
Hikvision	Ensemble A of 3 RPN and 6 FRCN models, mAP is 67 on val2	30	0.652704
Hikvision	Ensemble B of 3 RPN and 5 FRCN models, mean AP is 66.9, median AP is 69.3 on val2	18	0.652003
NUIST	submission_1	15	0.608752
NUIST	submission_2	9	0.607124

Object localization (LOC)[top]

Task 2a: Classification+localization with provided training data

Ordered by localization error

Team name	Entry description	Localization error	Classification error
Trimps-Soushen	Ensemble 3	0.077087	0.02991
Trimps-Soushen	Ensemble 4	0.077429	0.02991
Trimps-Soushen	Ensemble 2	0.077668	0.02991
Trimps-Soushen	Ensemble 1	0.079068	0.03144
Hikvision	Ensemble of 3 Faster R-CNN models for localization	0.087377	0.03711

Microsoft COCO

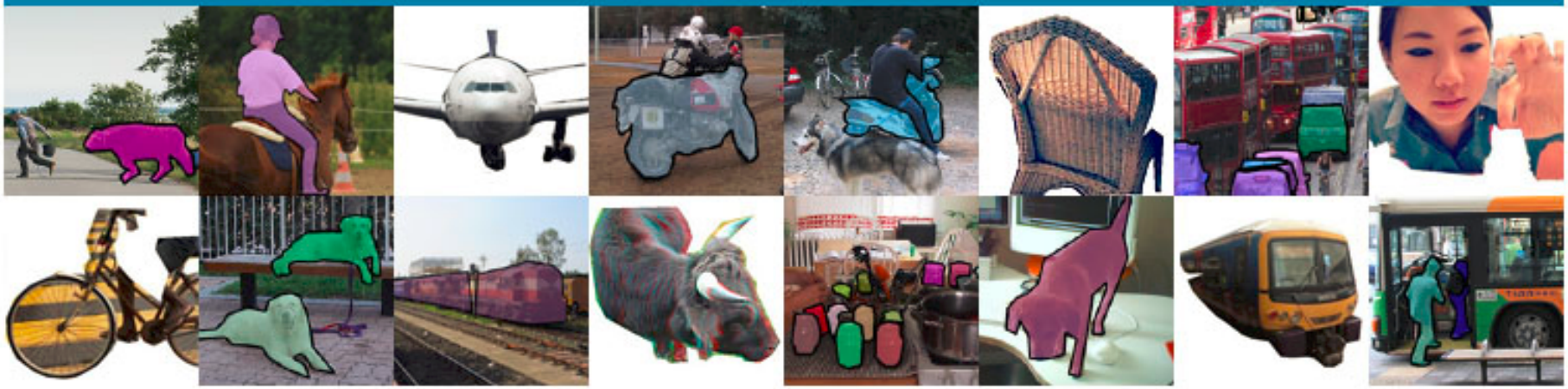
Object Detection/Semantic Segmentation/Pose Estimation/Image Captioning

- Object segmentation
- Recognition in Context
- Multiple objects per image
- More than 300,000 images
- More than 2 Million instances
- 80 object categories
- 5 captions per image
- Keypoints on 100,000 people

Challenges: Detections | Captions | Keypoints

	M1	M2	M3	M4	M5	date
Human[5]	0.638	0.675	4.836	3.428	0.352	2015-03-23
Google[4]	0.273	0.317	4.107	2.742	0.233	2015-05-29
MSR[11]	0.268	0.322	4.137	2.662	0.234	2015-04-08
Montreal/Toronto[10]	0.262	0.272	3.932	2.832	0.197	2015-05-14
MSR Captivator[12]	0.25	0.301	4.149	2.565	0.233	2015-05-28

Dataset examples

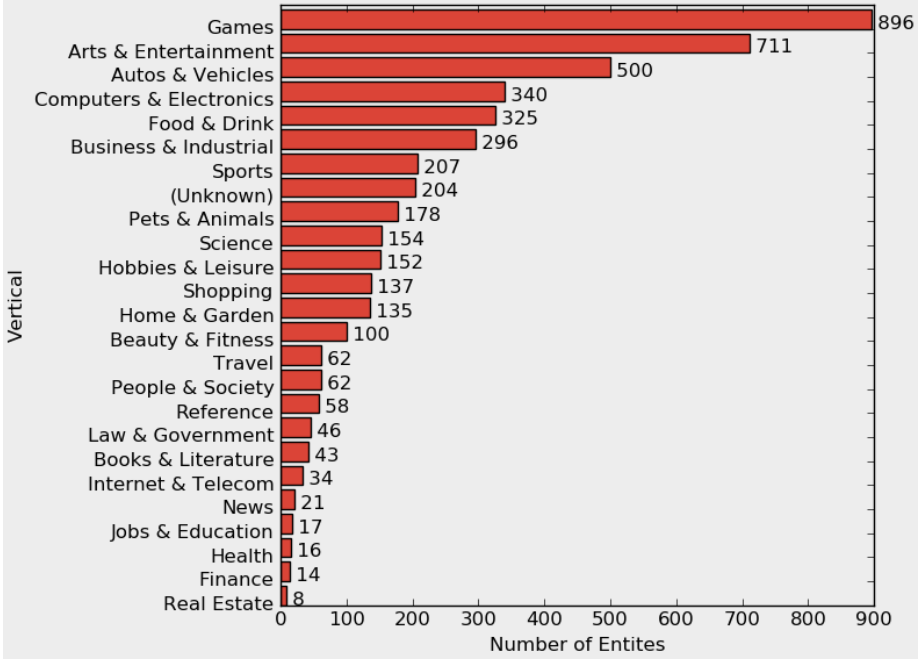


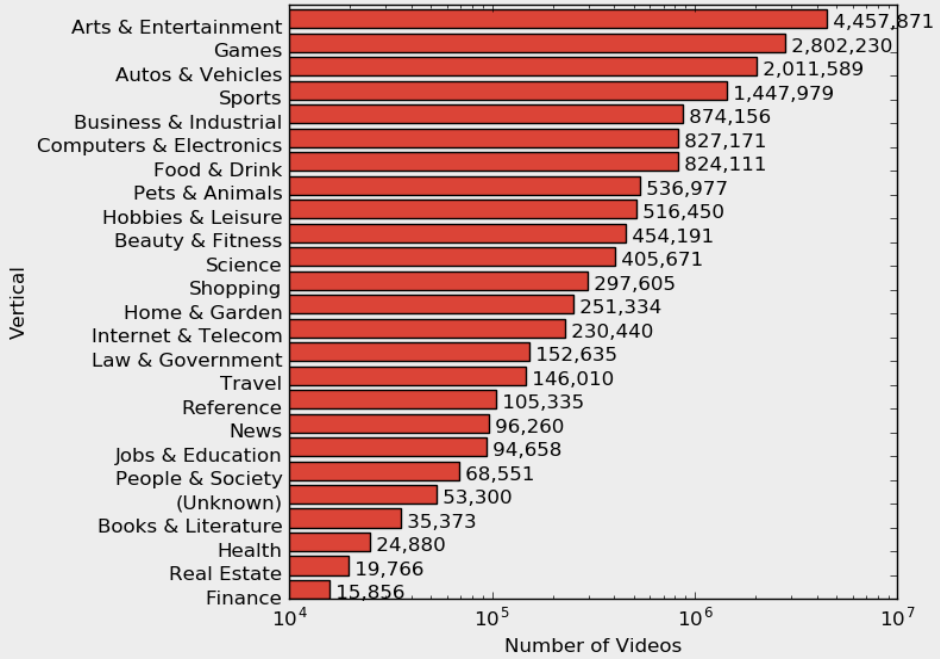
Google YouTube-8M

7 Million	450,000	3.2 Billion	4716	3.4
Video URLs	Hours of Video	Audio/Visual Features	Classes	Avg. Labels / Video

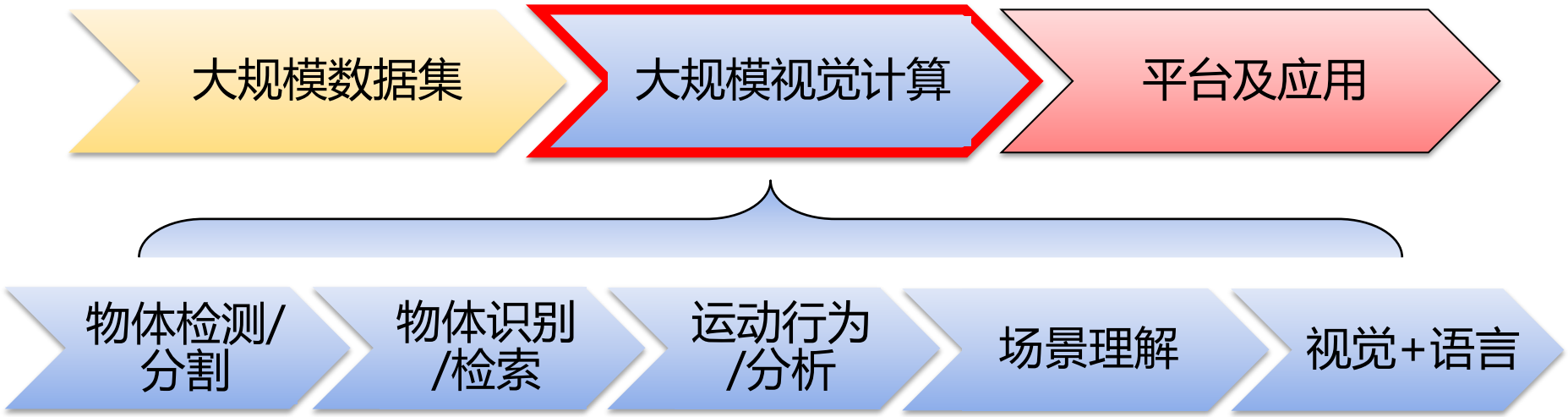
The videos are sampled uniformly to preserve the diverse distribution of popular content on YouTube, subject to a few constraints selected to ensure dataset quality and stability:

- Each video must be public and have at least 1000 views
- Each video must be between 120 and 500 seconds long
- Each video must be associated with at least one entity from our target vocabulary
- Adult & sensitive content is removed (as determined by automated classifiers)

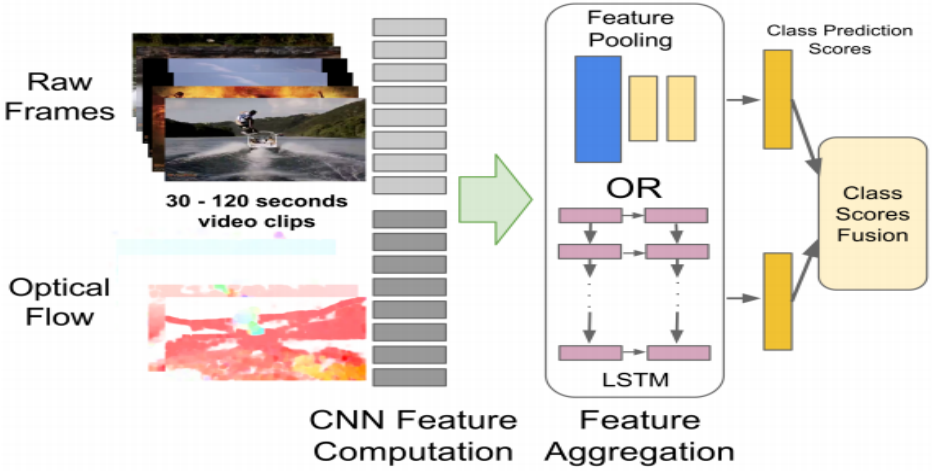




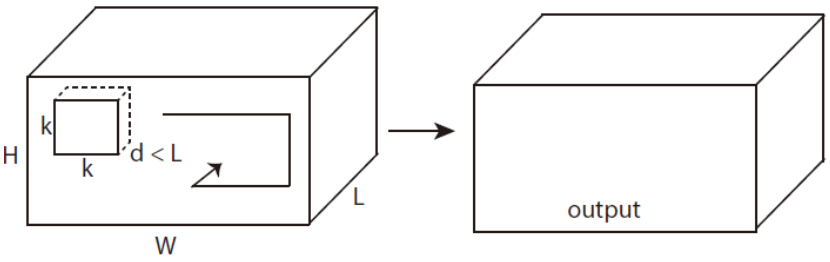
机器视觉



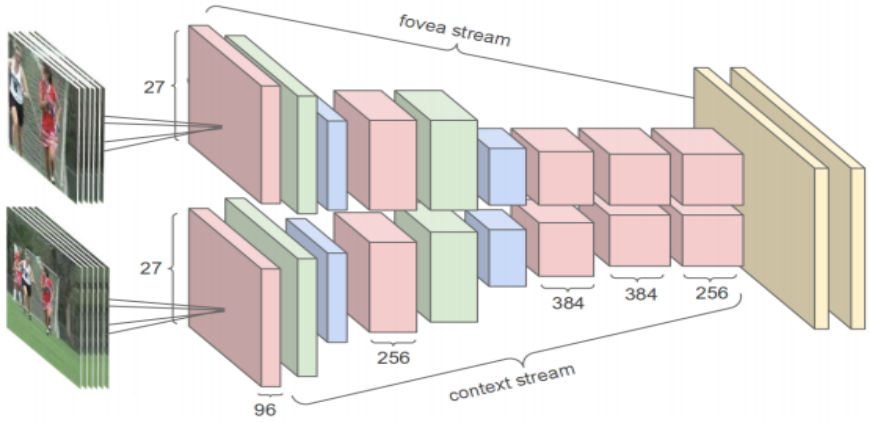
视频分类



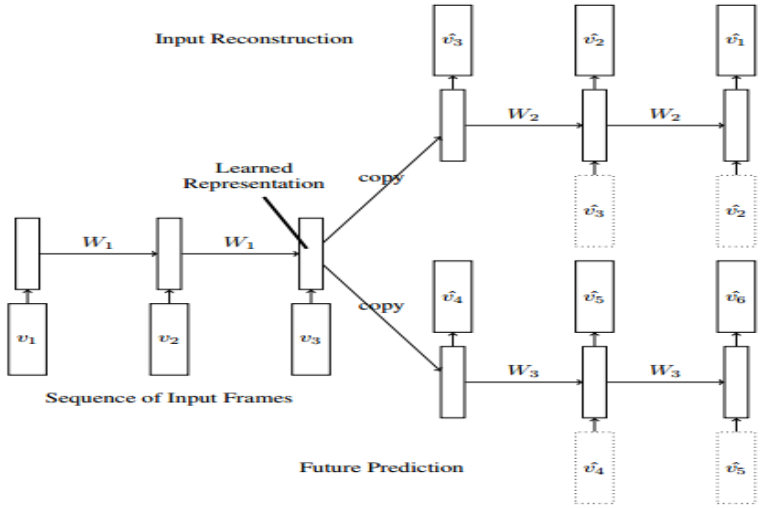
Long-range Videos Modelling, CVPR2015



3D Convolution, ICCV2015

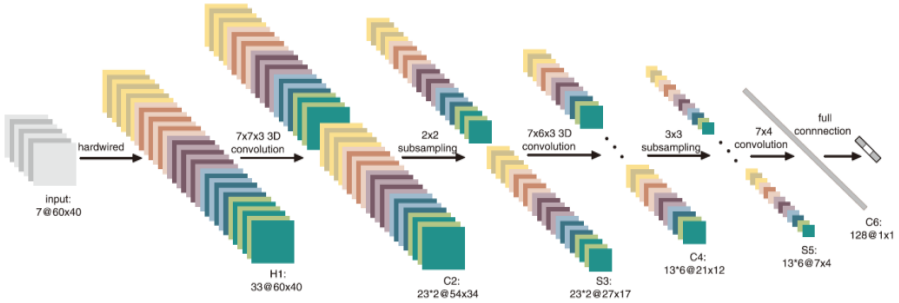


Multi-resolution CNN, CVPR2014

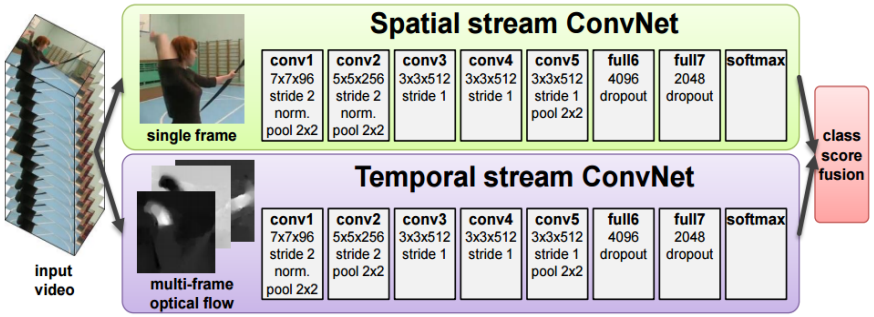


Autoencoder-RNN, ICML2015

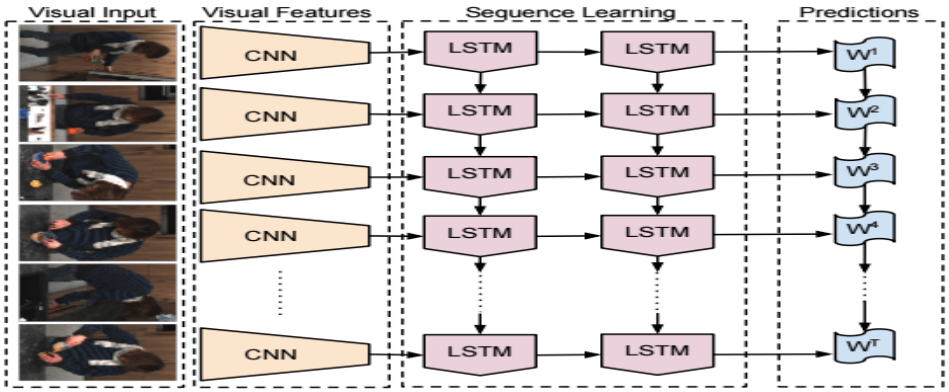
行为识别



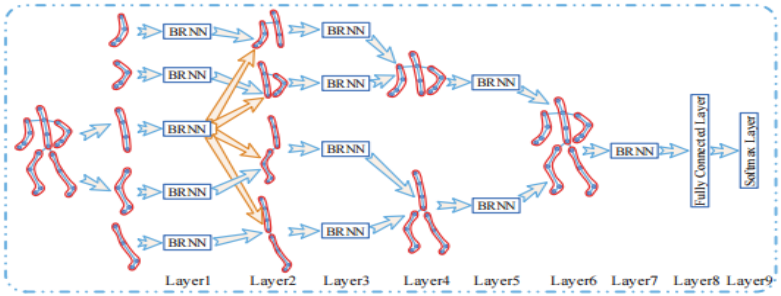
3D CNN, ICML2010



Two-stream CNN, NeurIPS2014

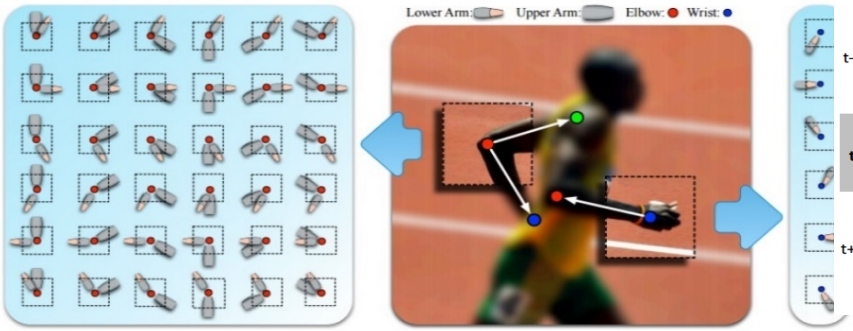


CNN + LSTM-RNN, CVPR2015

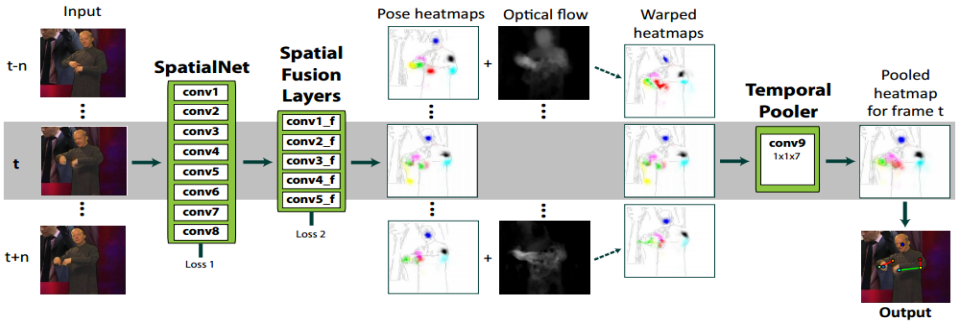


Skeleton + LSTM-RNN, CVPR2015

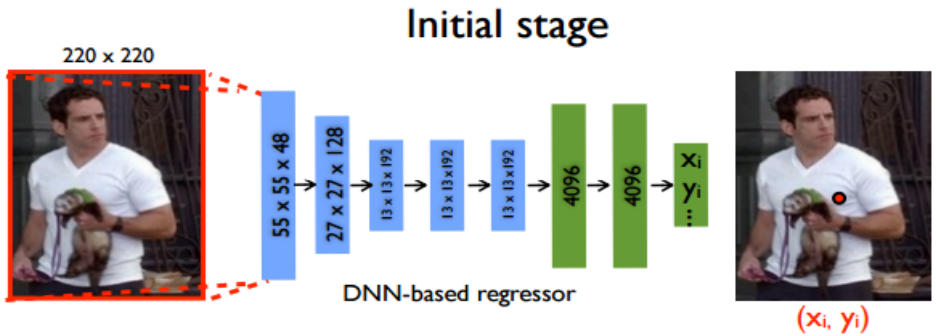
姿态估计

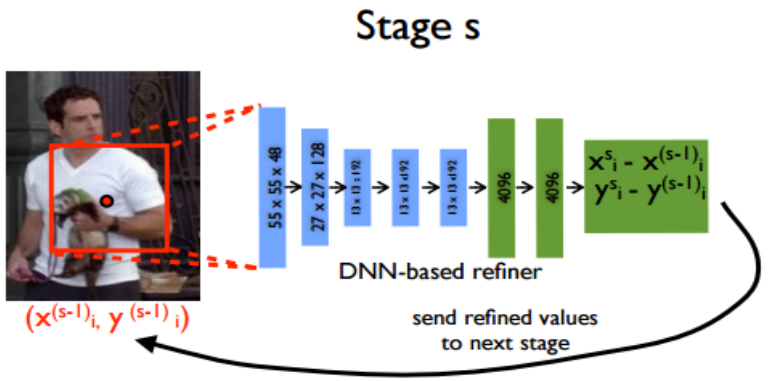


CNN + Graphical Model, NeurIPS2014



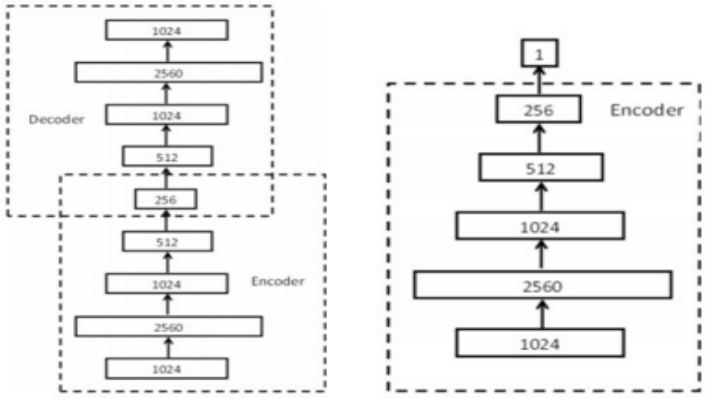
Flowing ConvNets, ICCV2015

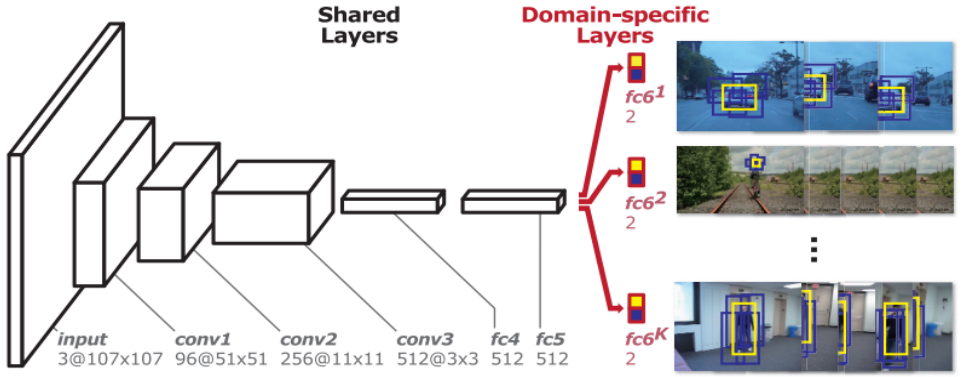




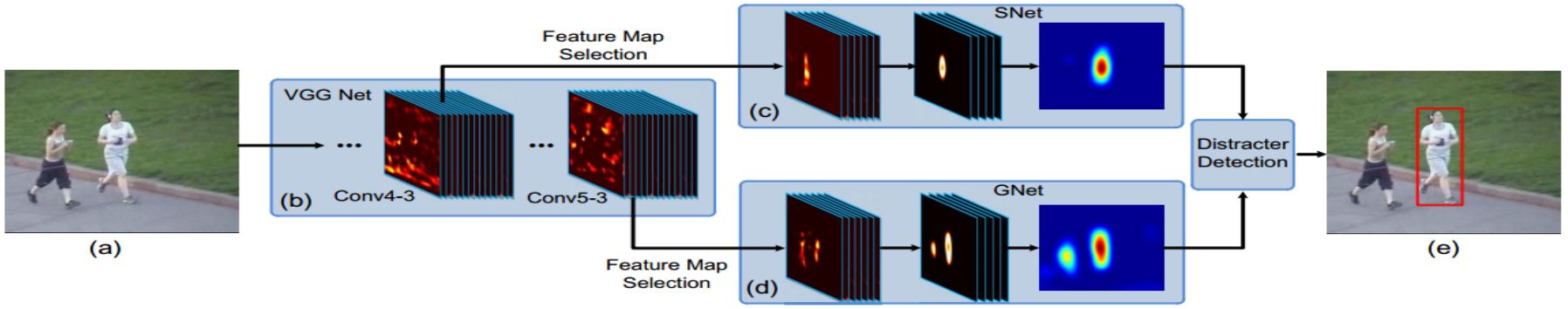
DNN-based Regression, CVPR2014

视觉跟踪



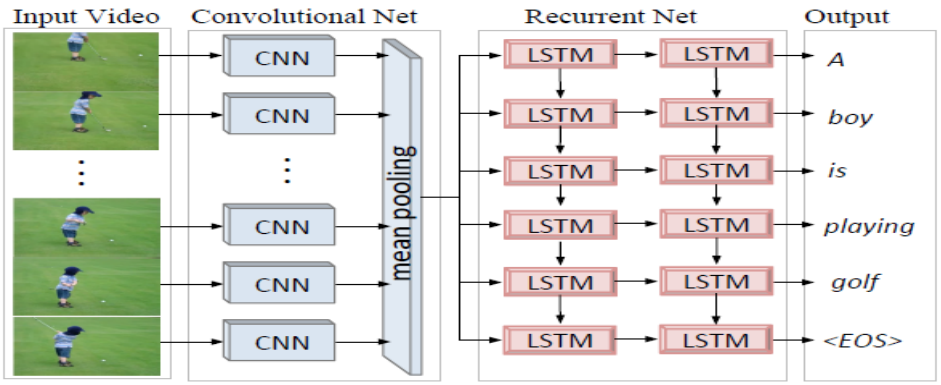


Stacked Denoising Auto-encoder, NeurIPS2013 **Multi-domain Networks**, CVPR2016

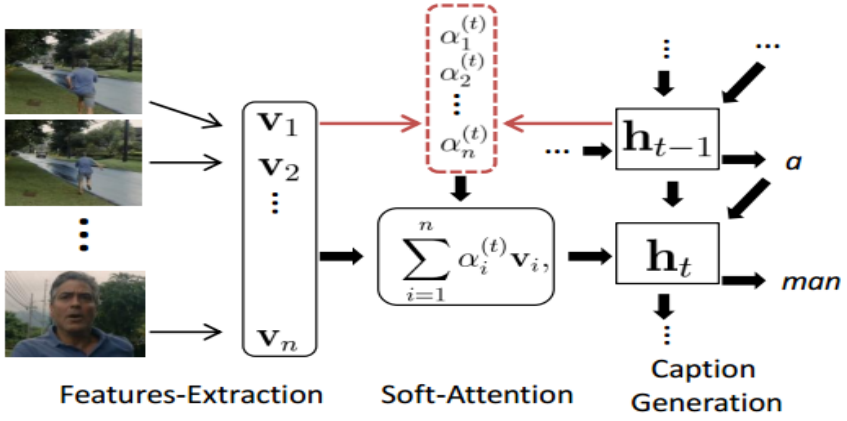


Fully Convolutional Networks, ICCV2015

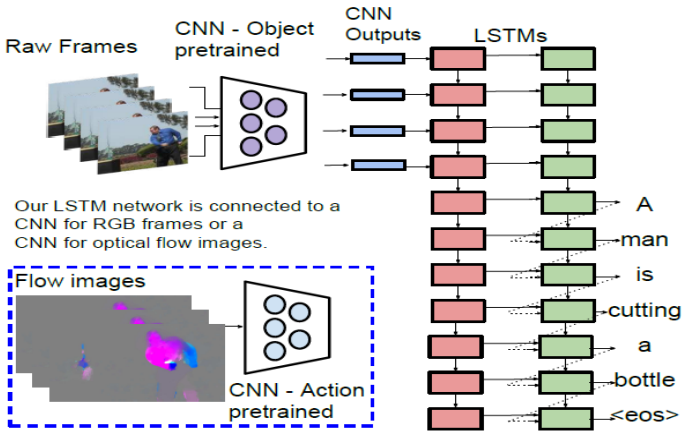
视频注释生成



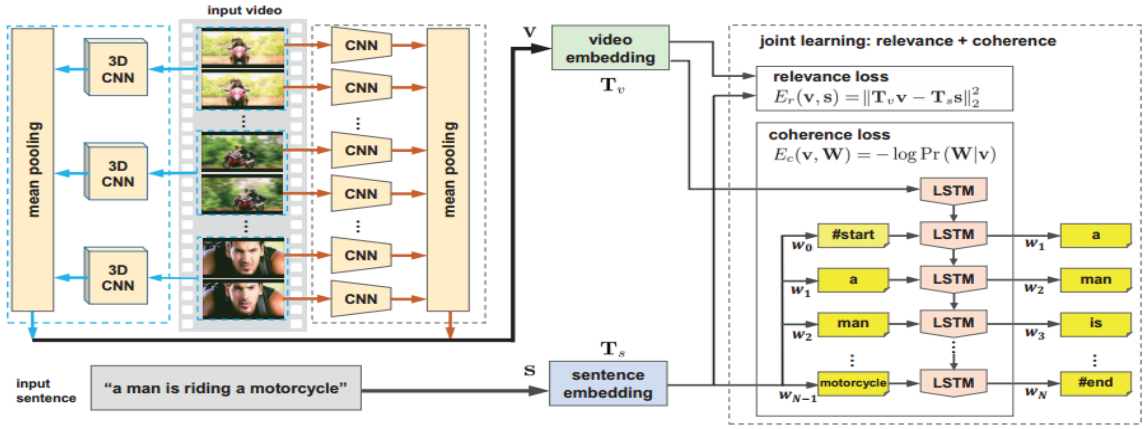
CNN + LSTM-RNN, CVPR2015



Attention-based LSTM-RNN, ICCV2015

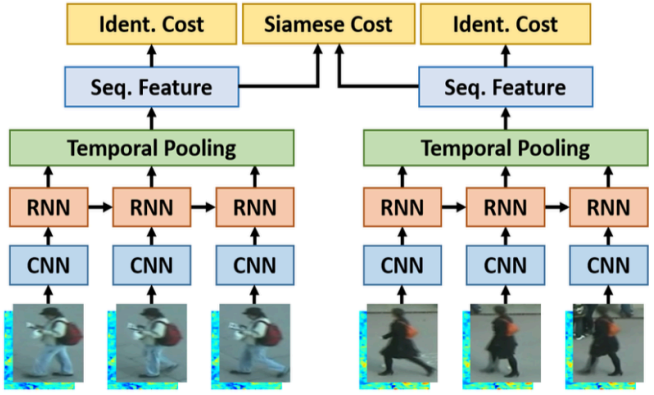


Video2Text, ICCV2015

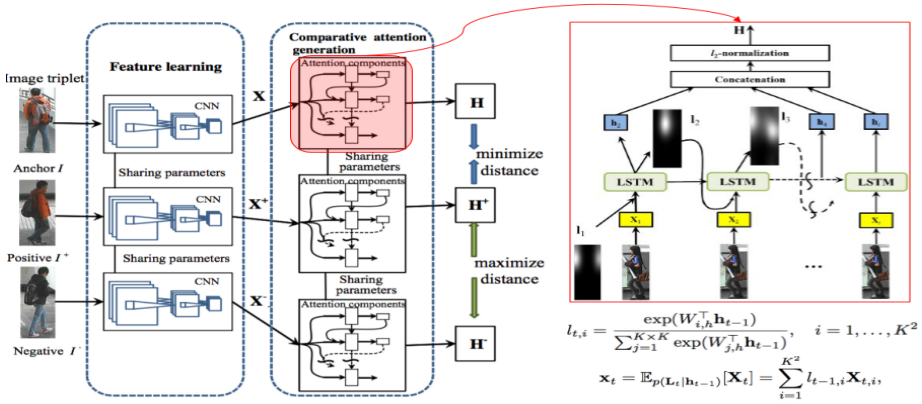


Visual-semantic Embedding + LSTM-RNN, CVPR2016

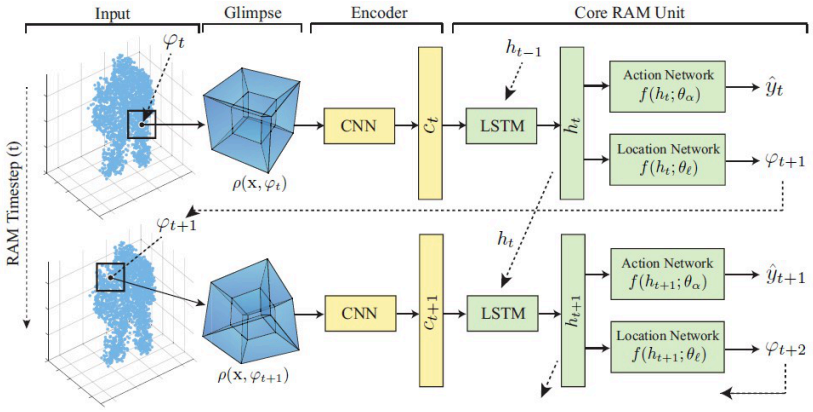
行人再识别



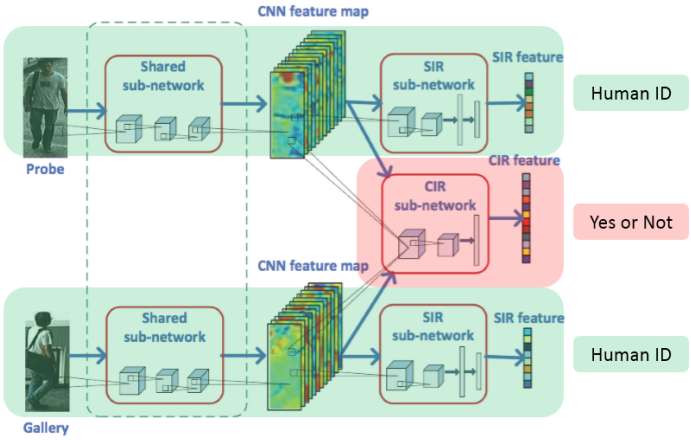
Recurrent CNN, CVPR2016



End-to-end Attention, TIP2016

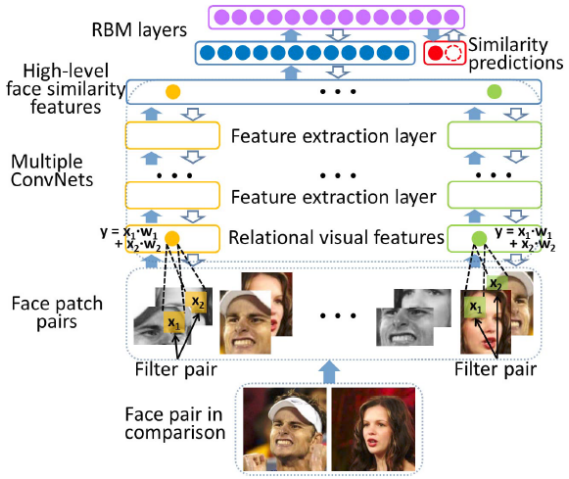


Sequential Local Attention, CVPR2016

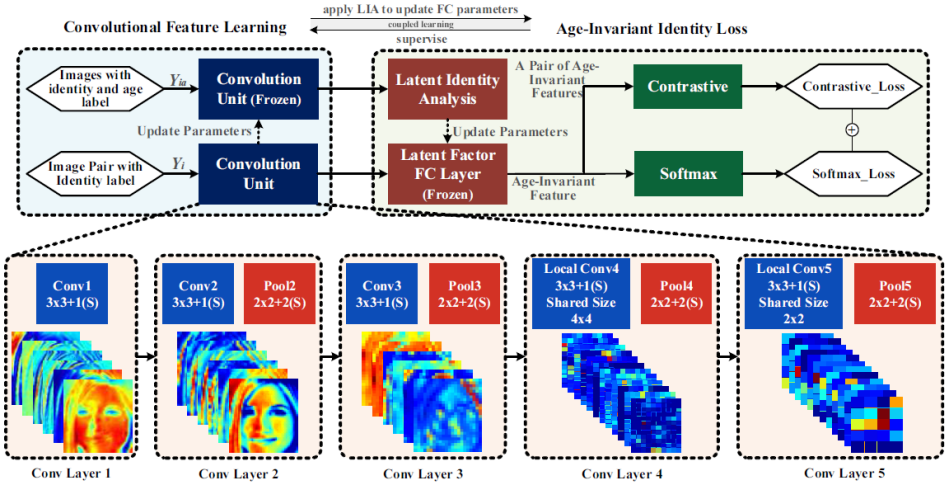


Single-/Cross-image Representations, CVPR2016

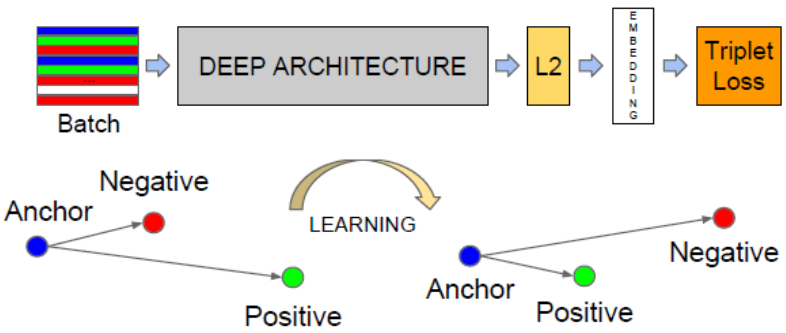
人脸识别



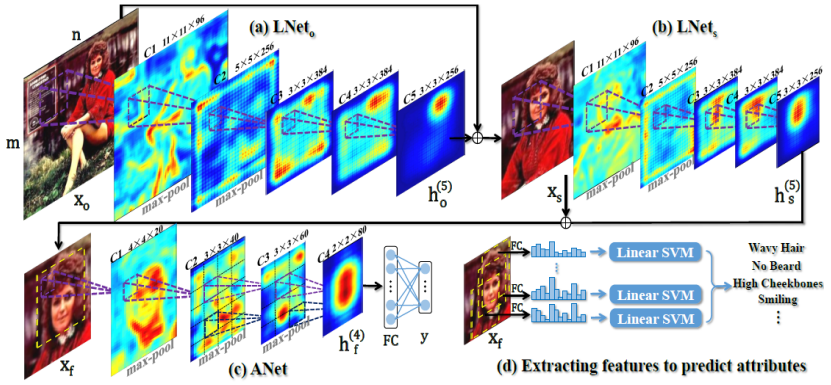
Hybrid ConvNet-RBM, ICCV13



LF-CNNs Model, CVPR16

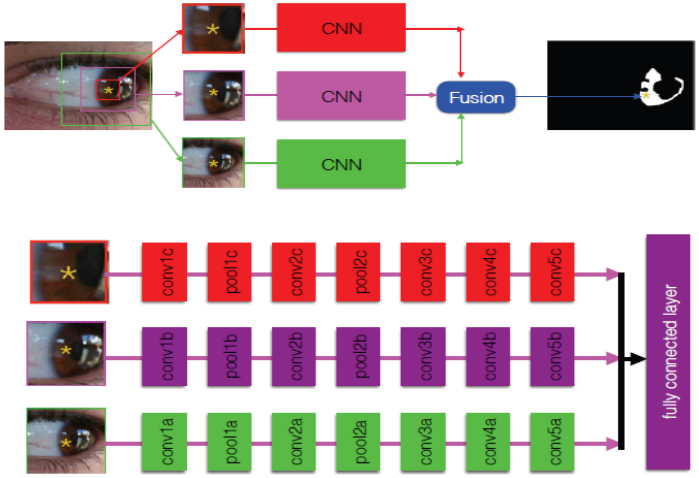


FaceNet, CVPR15

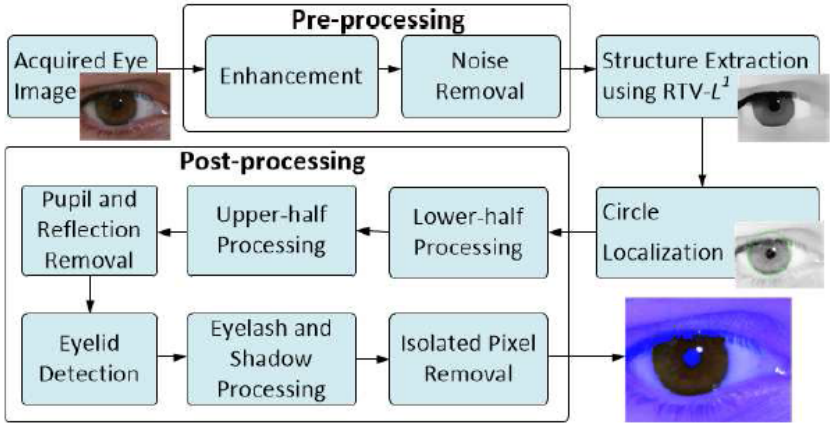


Face LNet and ANet, ICCV15

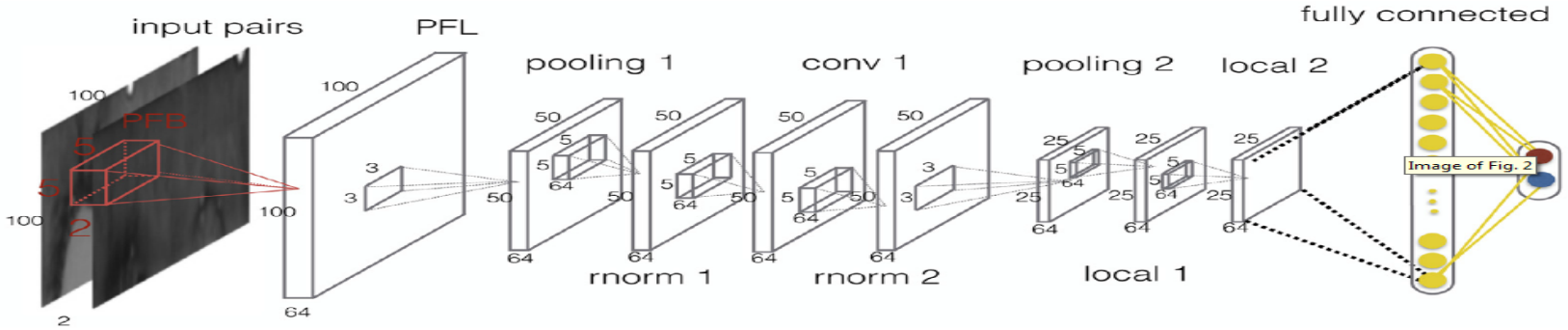
虹膜识别



HCNNs, ICB16

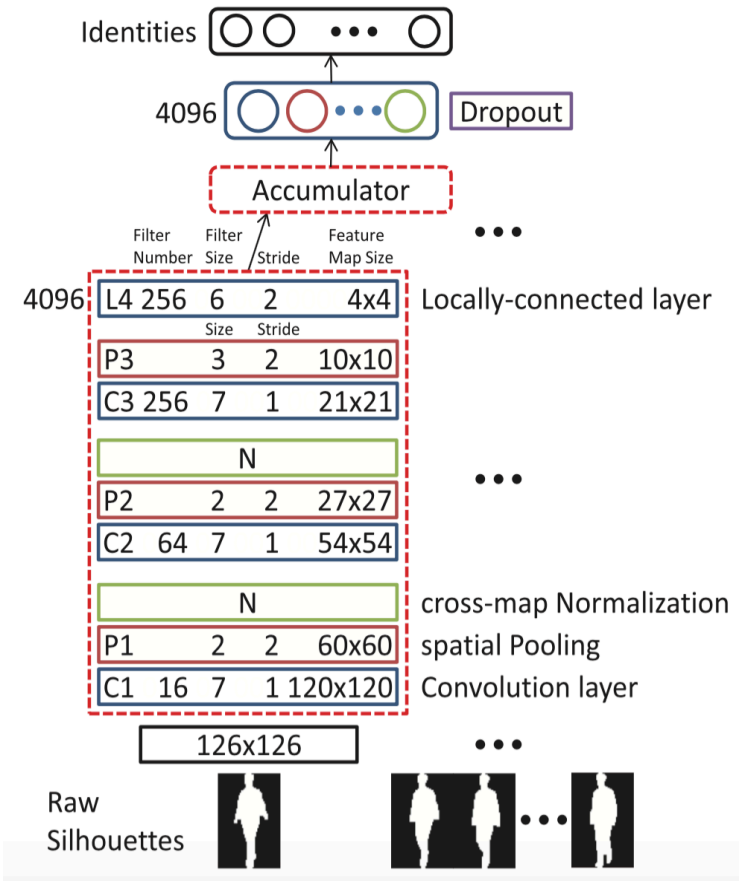


Iris Segmentation, ICCV15

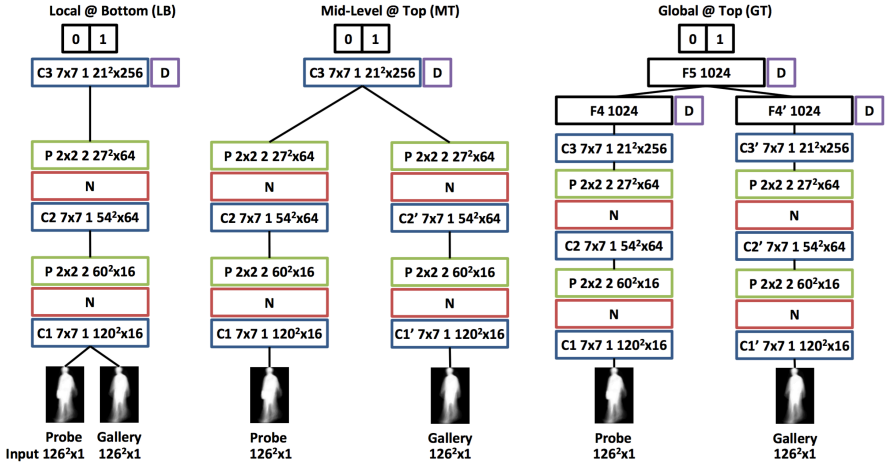


DeepIris, PRL15

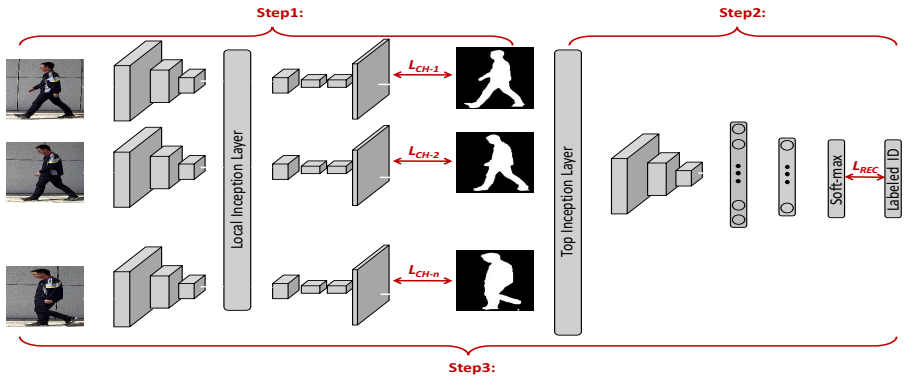
步态识别



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

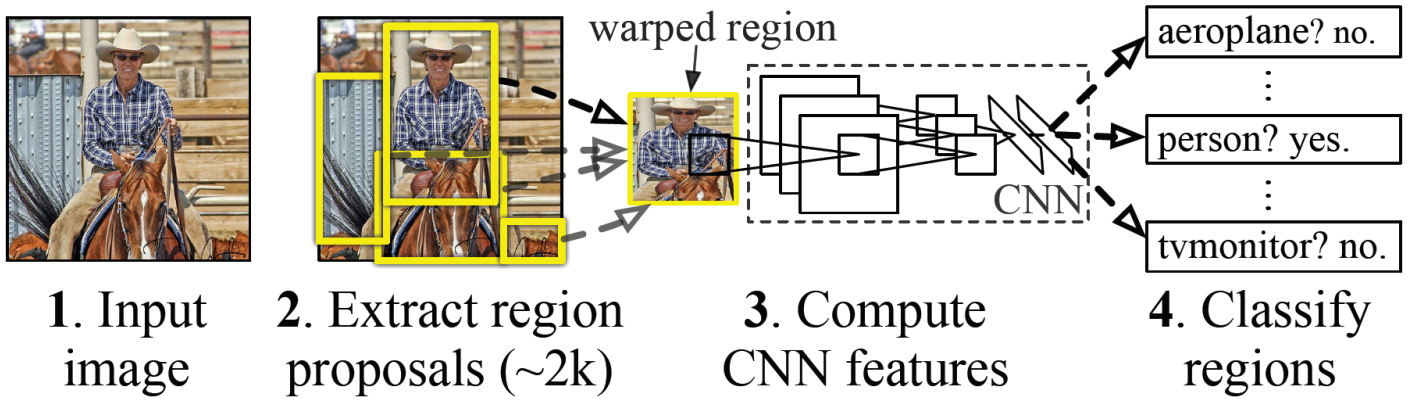


Cross-view Gait Identification, TPAMI 2016

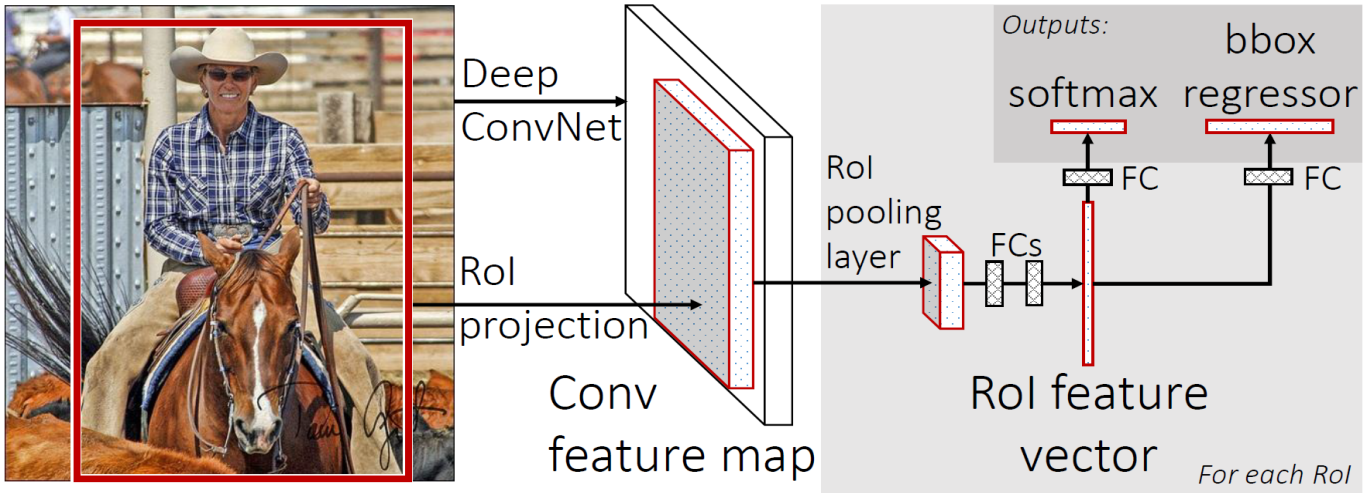


Joint Gait Segmentation and Recognition, submitted to CVPR 2017

物体检测

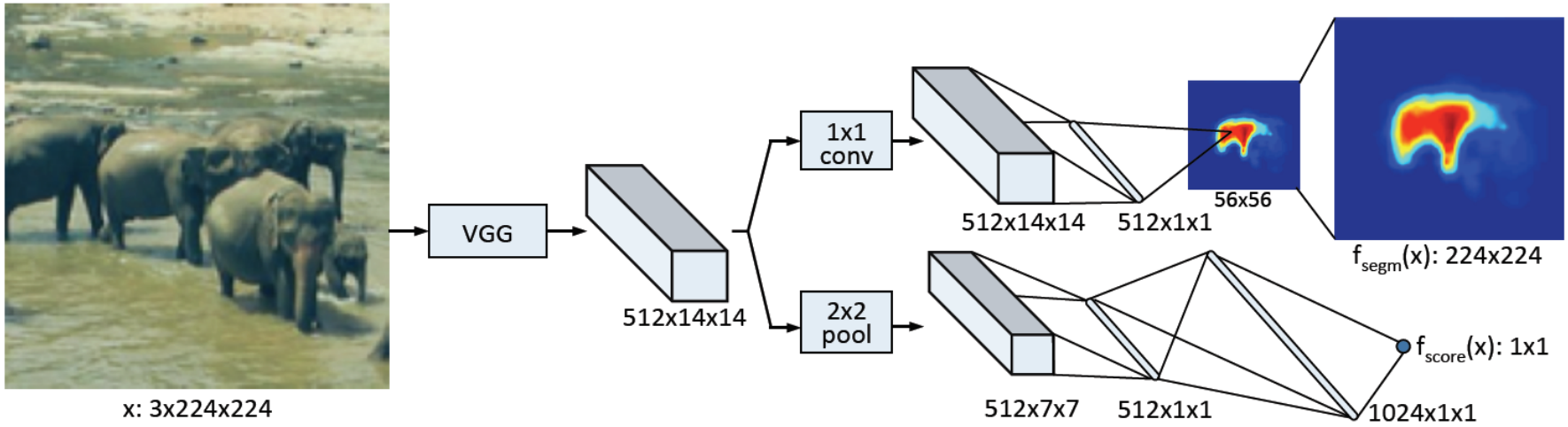


RCNN, CVPR14

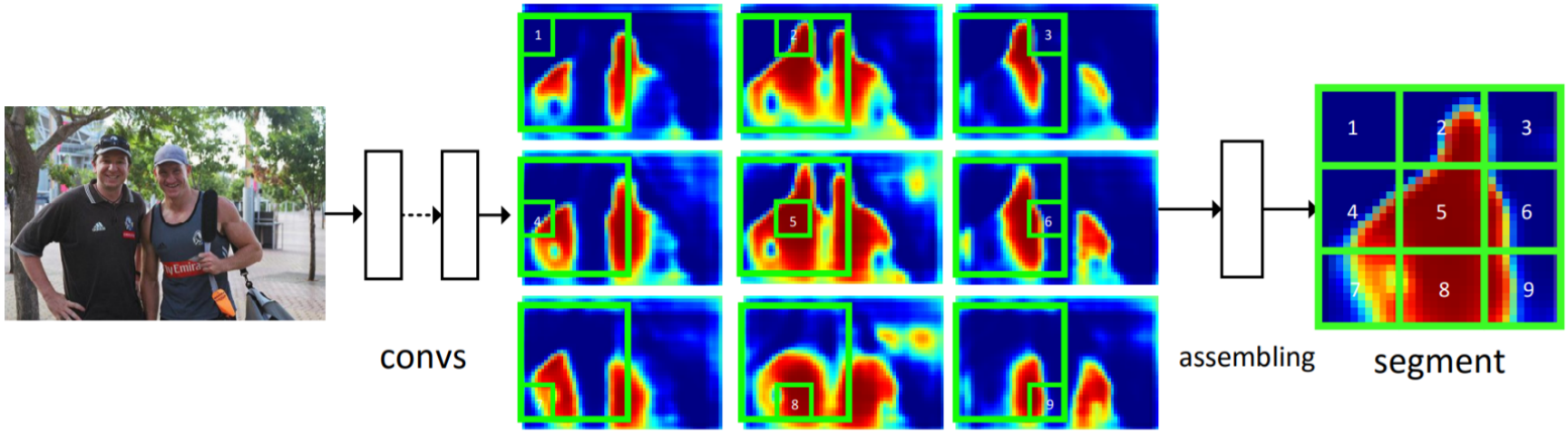


Fast R-CNN, ICCV15

语义分割

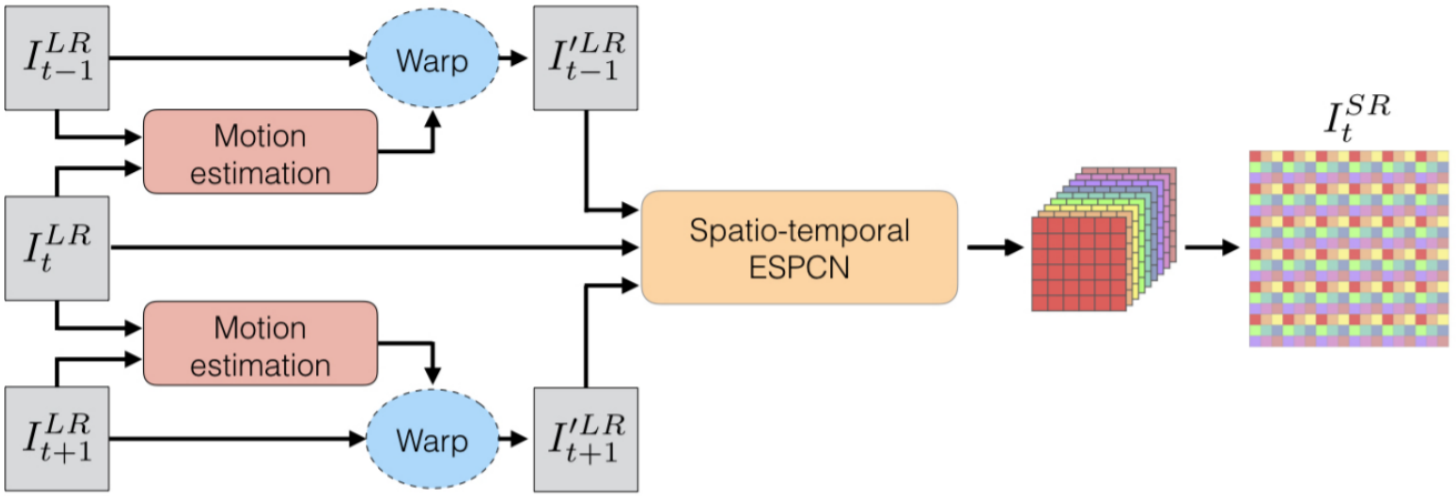


Instance-aware semantic segmentation, **CVPR16**

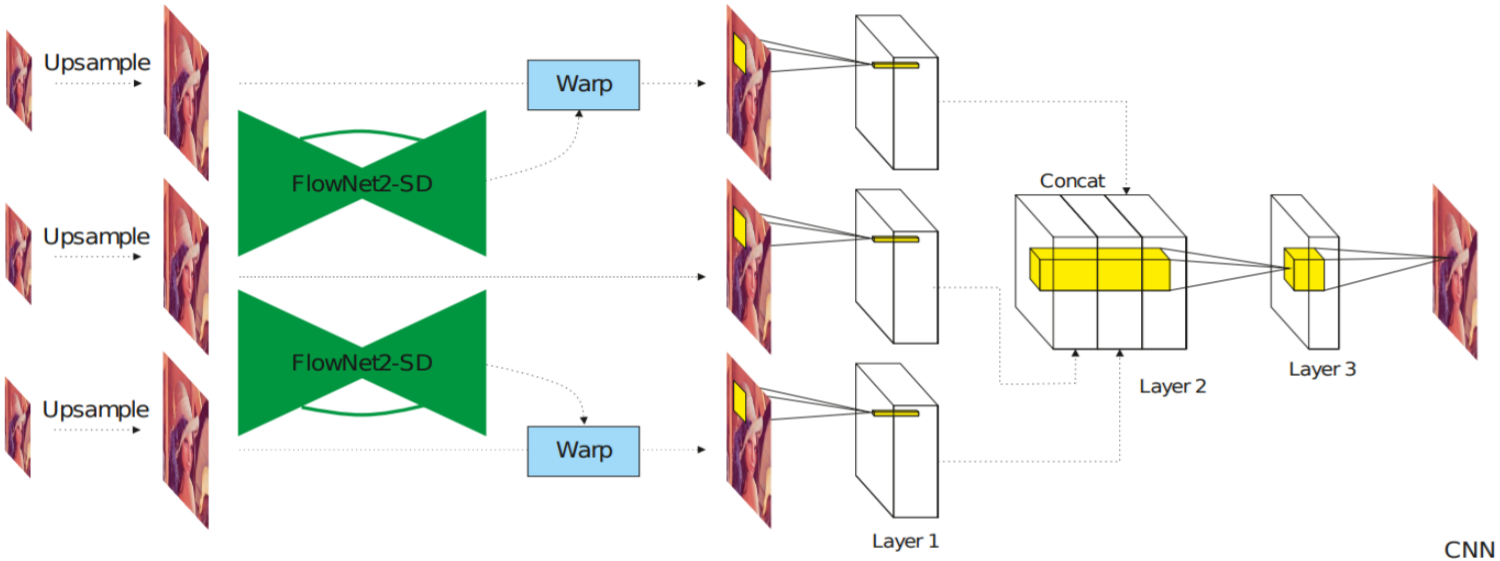


Learning to segment object candidates, **NeurIPS15**

视频超分辨率

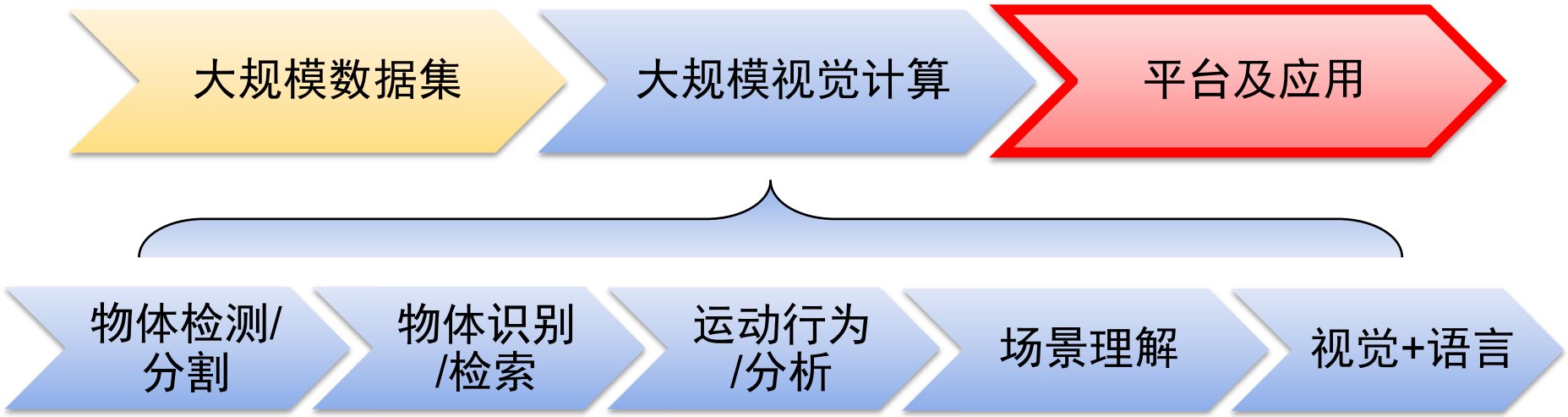


Real-Time Video Super-Resolution, CVPR17

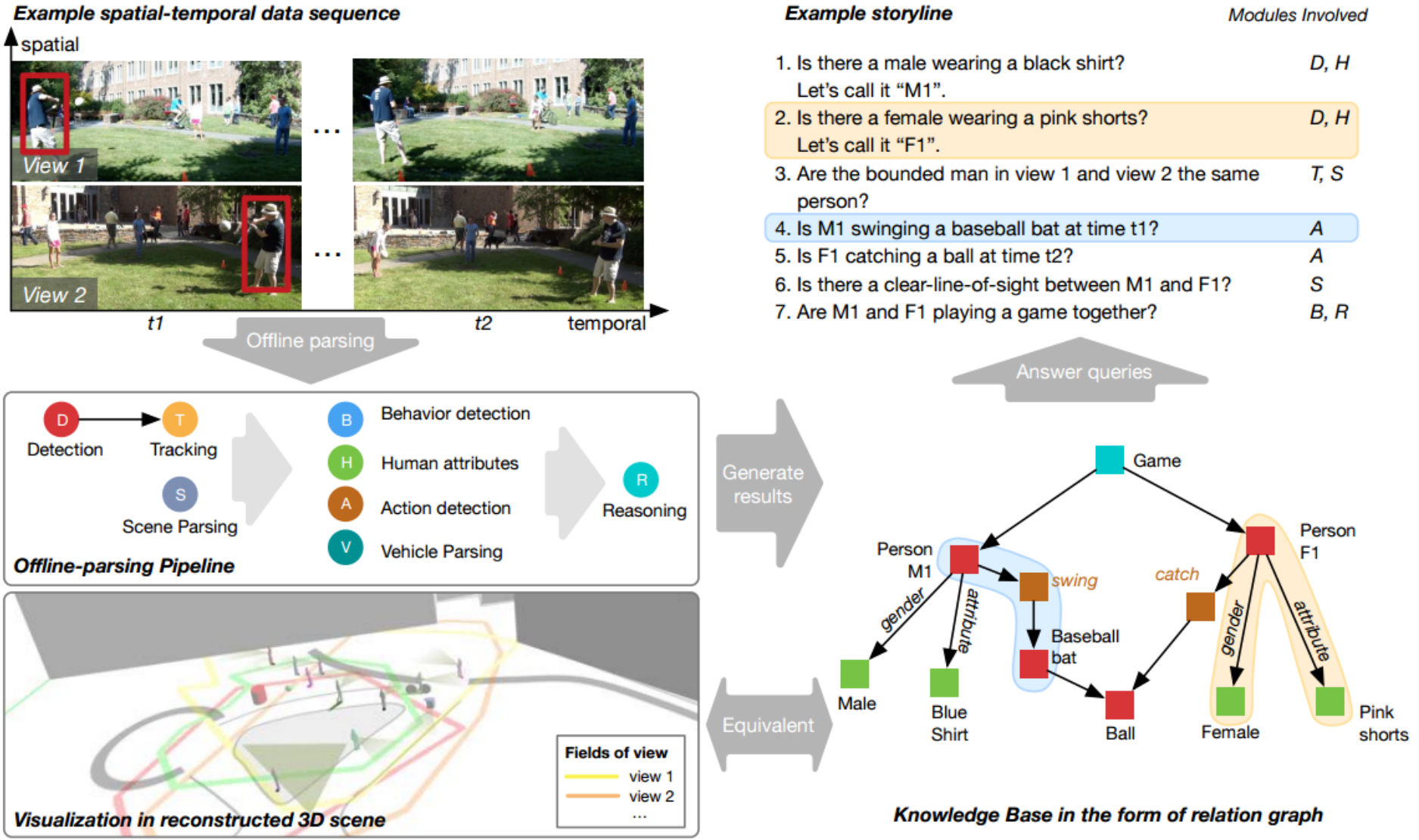


Video Super-Resolution with Motion Compensation, GCPR2017

机器视觉

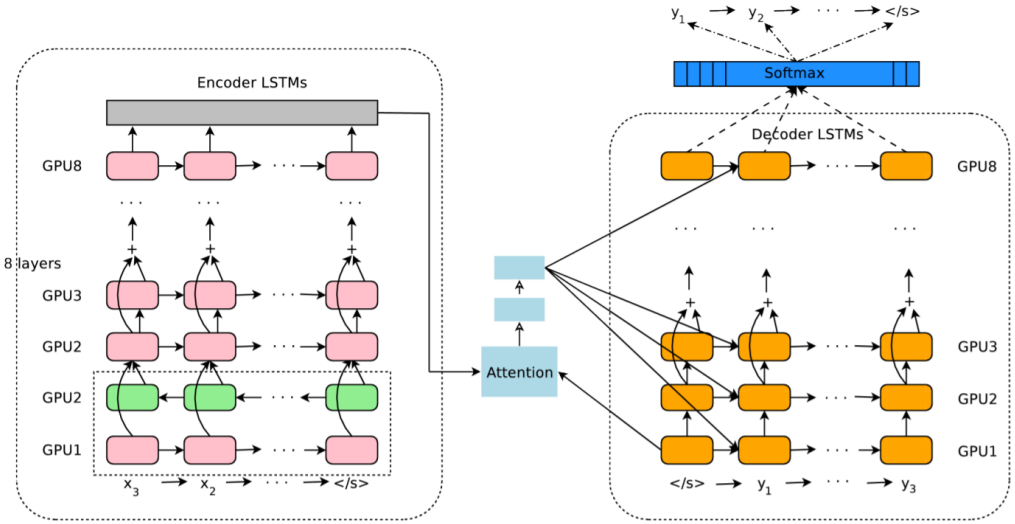


视觉图灵测试

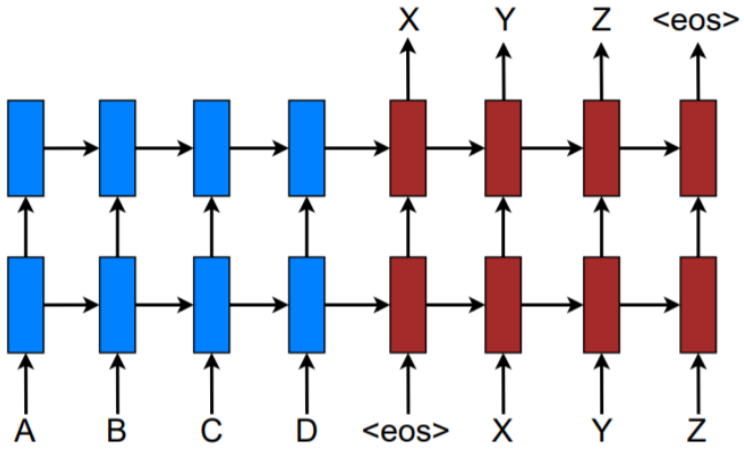


Q. Hang et al., A Restricted Visual Turing Test for Deep Scene and Event Understanding, arXiv 1512.01715.

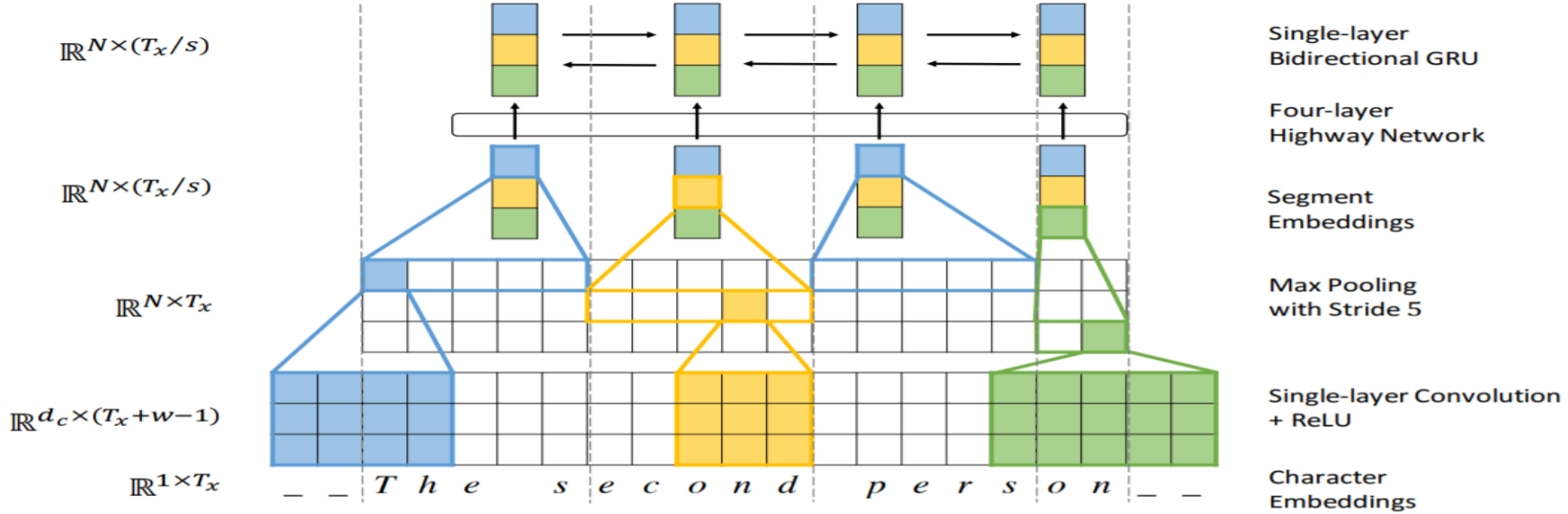
机器翻译



GNMT, Arxiv16

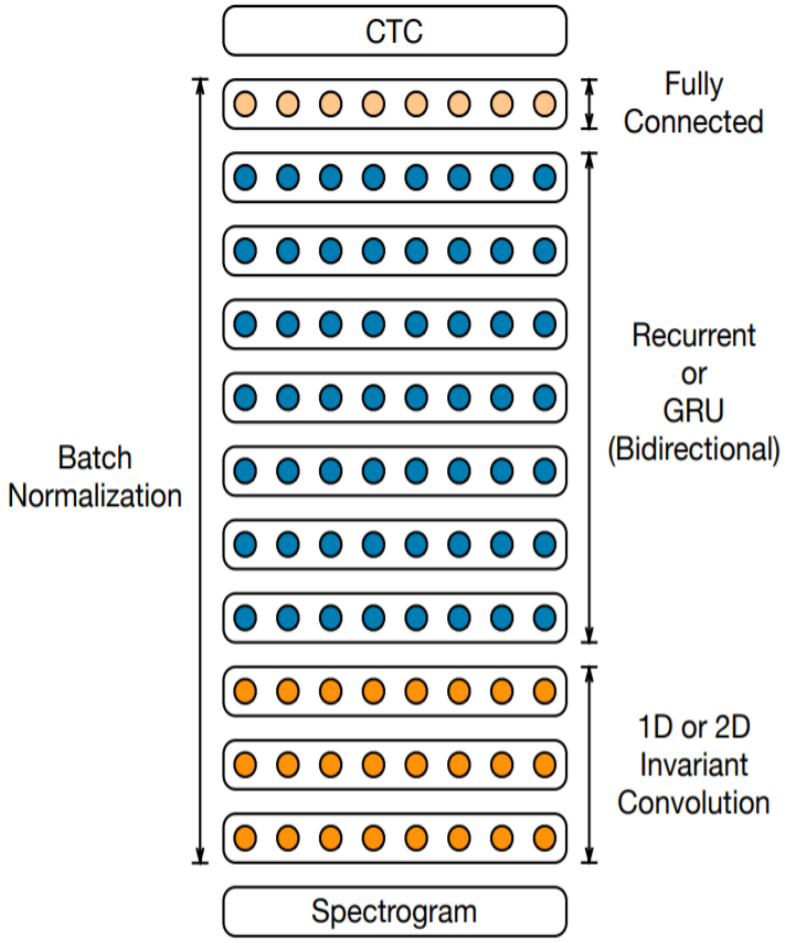


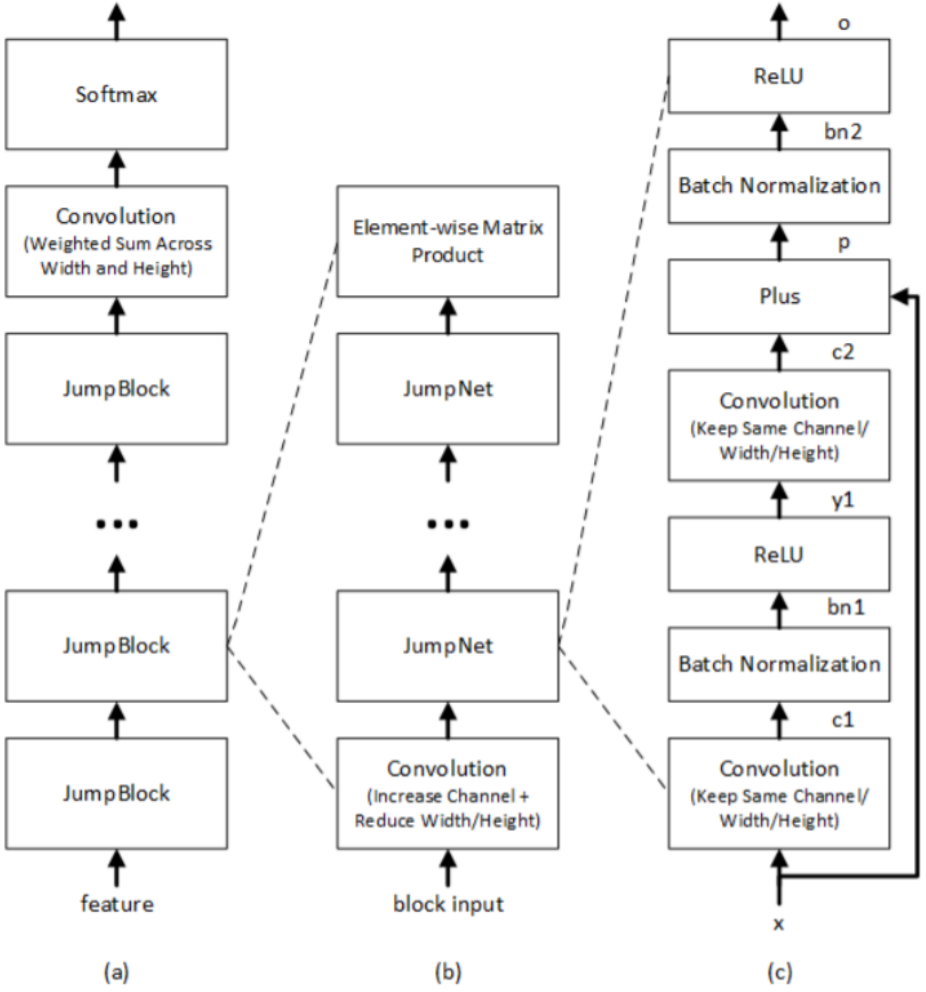
Attention-based NMT, EMNLP 2015



Fully Character-Level NMT, TACL17

语音识别

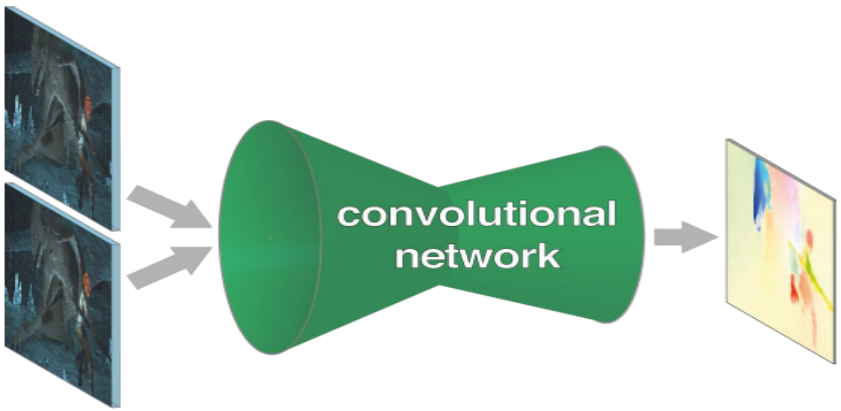




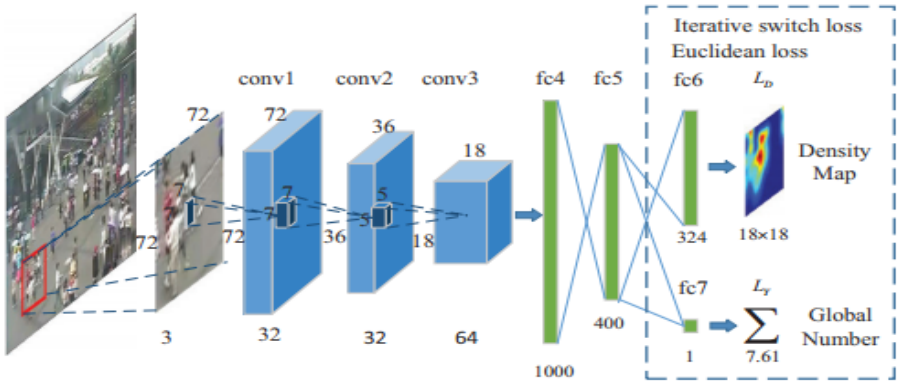
Speech Recognition in English and Mandarin, TASLP16

Conversational Speech Recognition, TASLP17

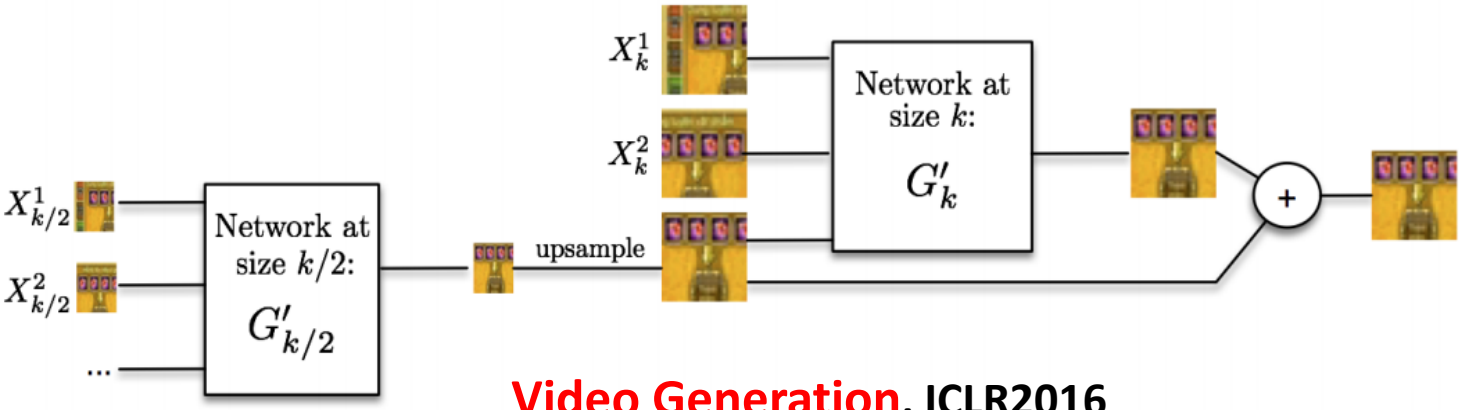
其他...



Optical Flow Prediction, ICCV2015



Cross-scene Crowd Counting, CVPR2015



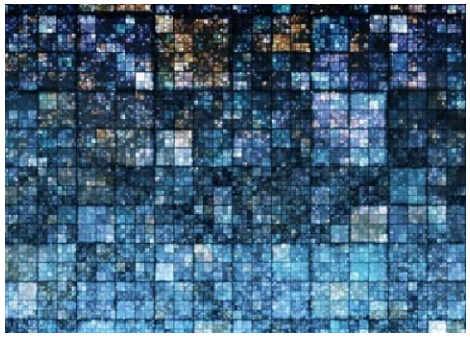
Video Generation, ICLR2016

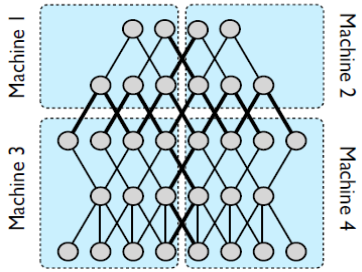
在许多应用中均实现了最先进的性能

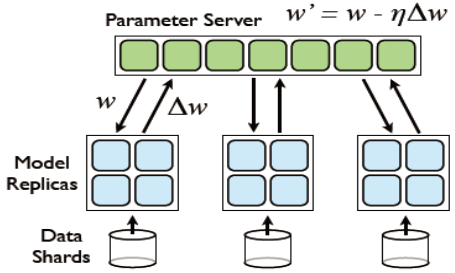
目录

未来研究方向

- 大规模深度学习
  - 多GPU学习
  - 分布式系统

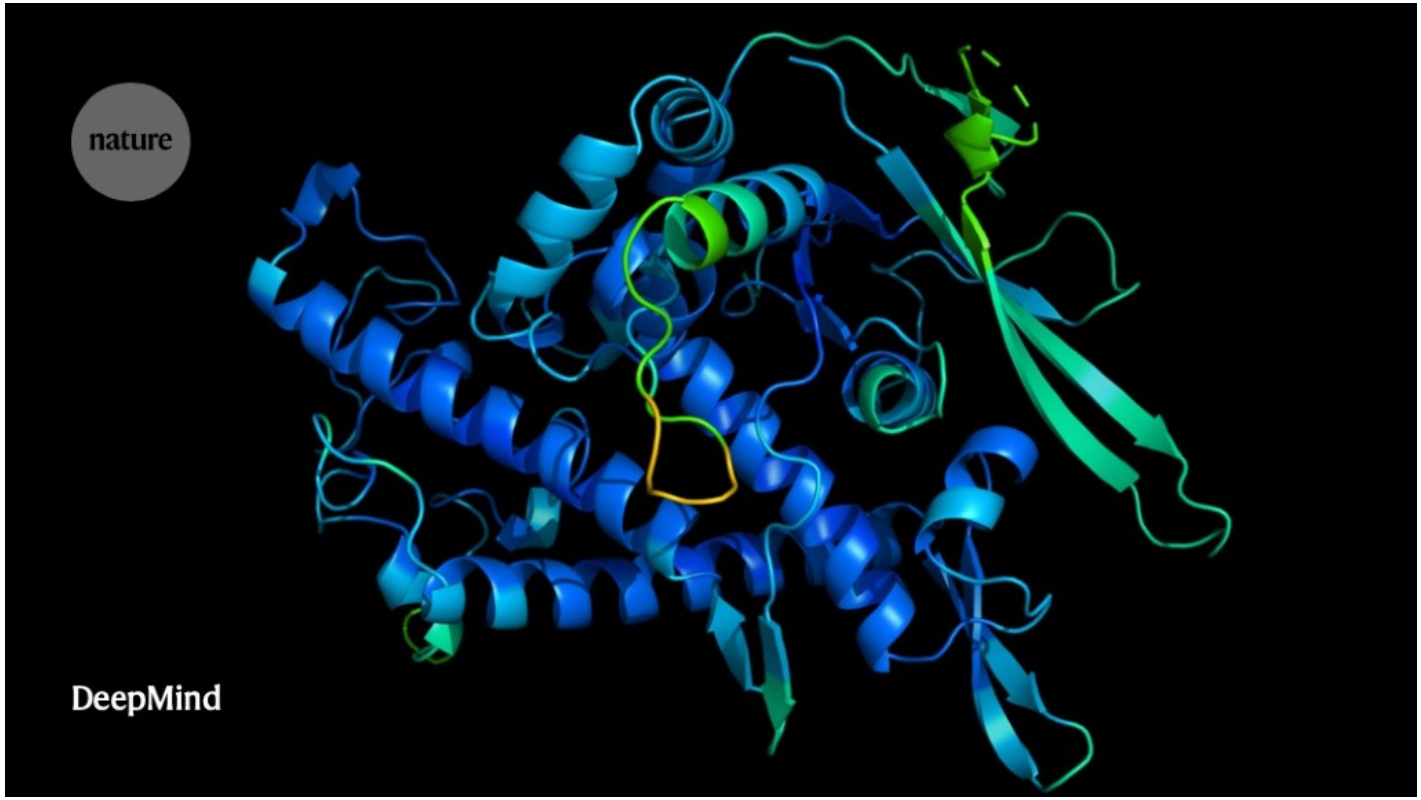






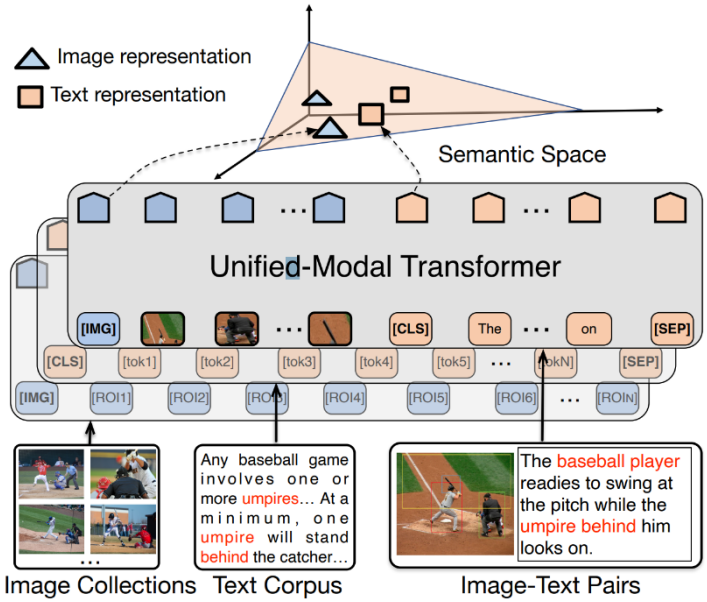
未来研究方向

- AI for science
  - 人工智能（深度学习）成为科学家的新生产工具，催生新的科研范式。

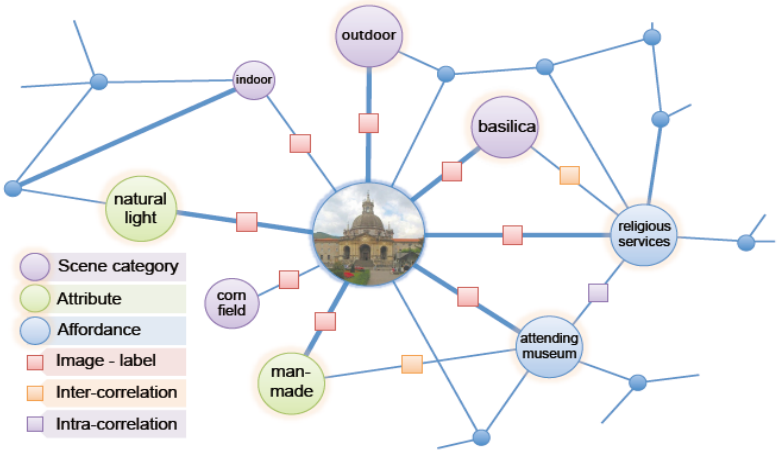


未来研究方向

- 多模态学习
  - 视频与其他数据模态密切相关，例如文本、音频
  - 多模态预训练模型



Unimo [Li et al., ACL2021]



Multimodal Knowledge Base [Zhu et al., arXiv16]

未来研究方向

- 脑启发的深度模型
  - 传统的神经网络受到人类认知机制的启发
  - 未来可参考神经科学和脑科学领域的先进成果

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

参考文献

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

参考文献

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

参考文献

14. **(GNN)**: Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks, ICLR, 2017

15. **(GNN)**: Veličković P, Cucurull G, Casanova A, et al. Graph attention networks,ICLR , 2018

16. **(Transformer):**Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need, NIPS, 2018

17. **(Transformer in CV):** Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale, ICLR, 2021

课程设计选题示例

人体行为识别

可能的思路:
1. 给定视频片段，能否确定其中是否包含行人?
2. 如果包含行人，能否跟踪其轨迹，并且检测其身体部分及姿态?
3. 根据上述信息，能否最终确定该行人完成了什么行为或者动作?

建议:
该任务涉及到的内容十分广泛，建议选择其中的一个具体的部分进行课程设计。

可用数据集:
[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/)

课程设计选题示例

人体行为识别

相关论文：
[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://ai.stanford.edu/~bangpeng/ppmi.html)
[3] Poselet (http://www.cs.berkeley.edu/~lbourdev/poselets/)
[4] 2D articulated human pose estimation

可以采用网上提供的代码来获得时空特征 (spatial-temporal features)。
一个可用的示例：
http://vision.ucsd.edu/~pdollar/research/research.html. By Piotr Dollar.

课程设计选题示例

图像分类

建议:
图像分类/目标识别一直是计算机视觉领域的重要研究课题之一。研究人员已经开发了各种不同的局部描述符、特征编码方案和分类方法。例如SIFT, HoG等等。
SIFT features: http://www.vlfeat.org/~vedaldi/code/sift.html.

可用数据集:
[1]Caltech101/256:
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/

课程设计选题示例

多模态学习

可能的思路:

主要尝试挖掘和理解不同模态信息之间的潜在关系，例如：

1. 给定电影海报的照片(图像)，我们是否可以检索得到电影的相关预告片(视频)或最新的新闻报道(文本)?
2. 给定一个啤酒品牌图片，我们是否可以在一个给定的视频片段中检索哪些帧包含该品牌的啤酒?
3. 通过考虑更实际的应用场景，能否在现有的方法上做出改进?

课程设计选题示例

多模态学习

相关论文及工具：

[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/)

课程设计选题示例

- 图像分割 (Image Segmentation)
- 人脸识别 (Face Recognition)
- 文本分类 (Text Classification)
- 智能问答 (Question Answering)
- 图像去噪/超分辨率 (Image Denosing/Super-resolution)
- 图像检索 (Image Retrieval)
- 跟踪 (Tracking)
- 数据挖掘（Data Mining）
- ……

注意：上述题目难度有所差异。我们在评估每个小组的报告时会考虑该因素，尽量保证课程设计成绩的公平性。

致谢

本课件中部分材料借鉴以下课程：

- 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

Thank You !