

# 数据驱动的深度人脸识别 研发进展及挑战

山世光

中科院计算所

[sgshan@ict.ac.cn](mailto:sgshan@ict.ac.cn)



中国科学院计算技术研究所  
Institute of Computing Technology, Chinese Academy of Sciences



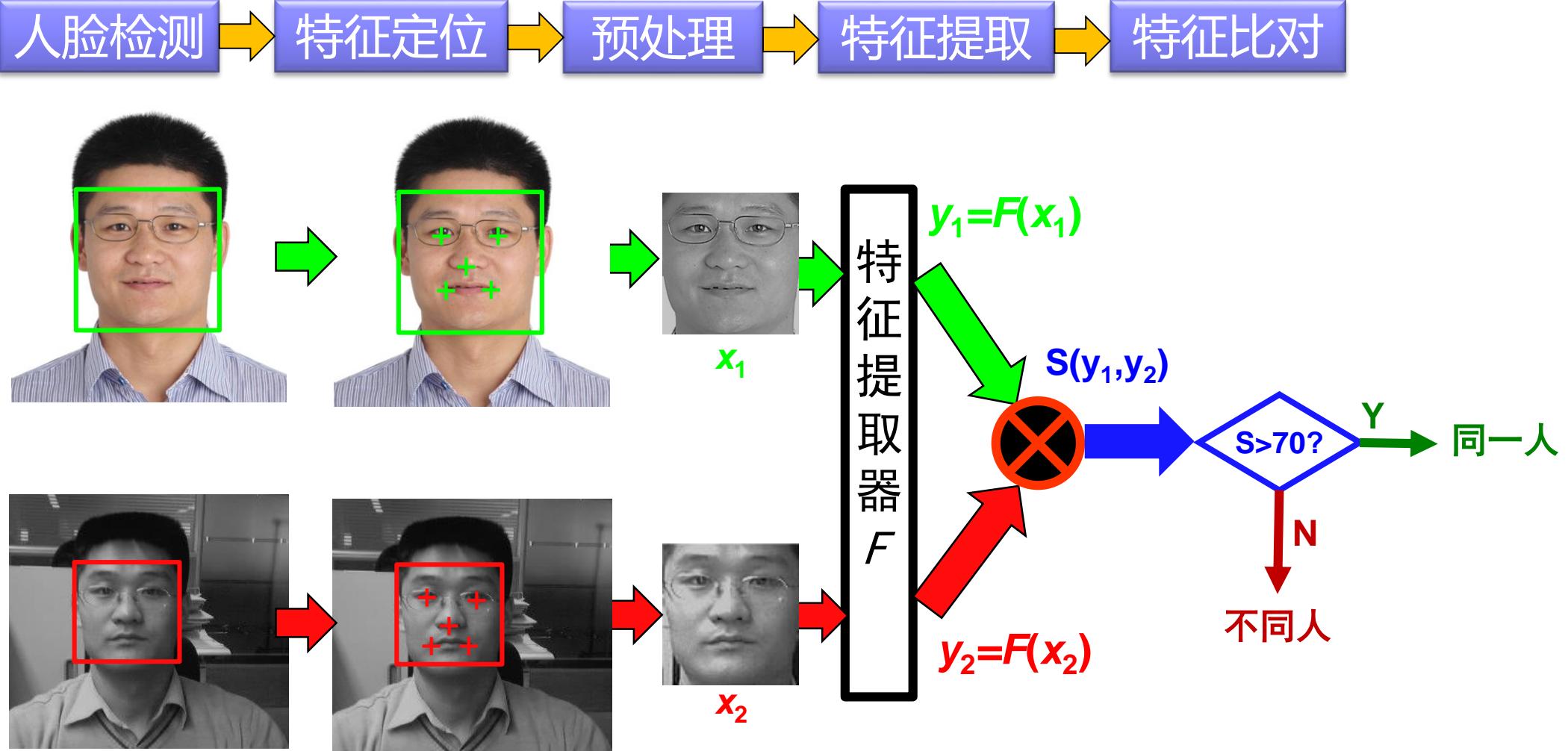
# 提纲

- 背景与概况
- 人脸识别技术的深度化
  - 人脸检测技术
  - 面部特征点定位技术
  - 判别特征学习与匹配方法
  - SeetaFace人脸识别引擎
- 总结与讨论



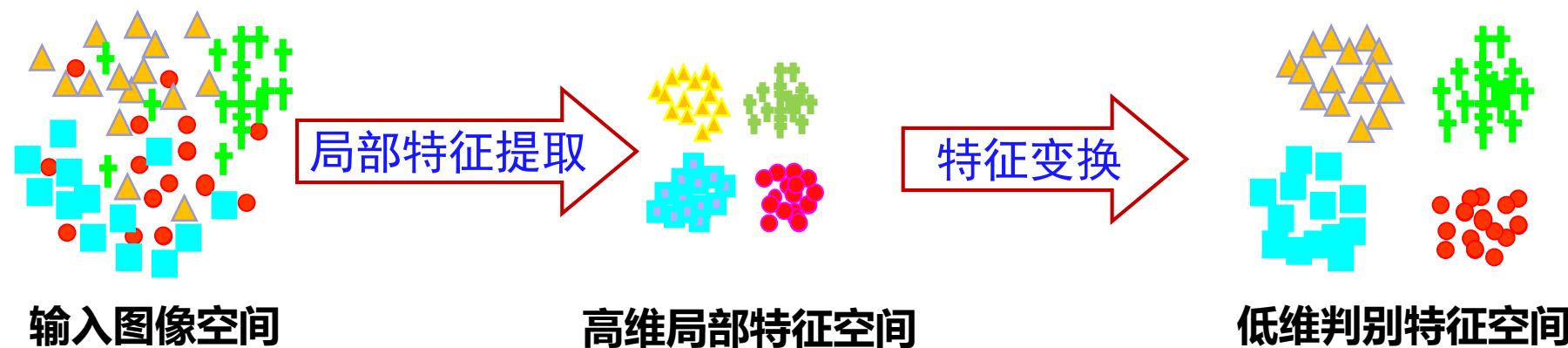
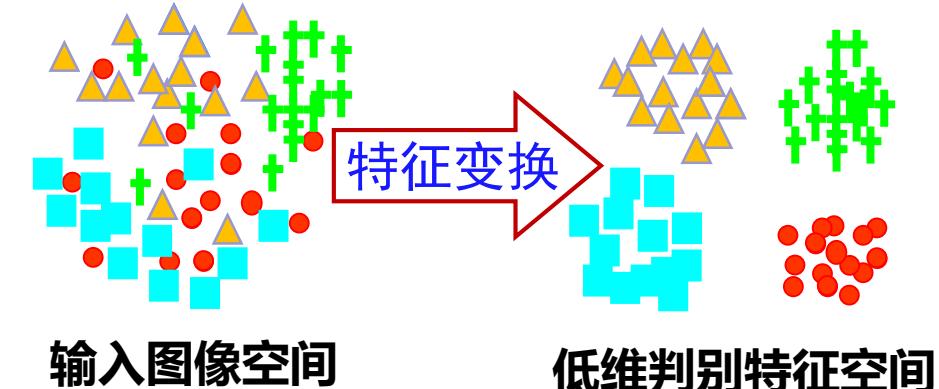
# 全自动人脸识别系统流程

中科院计算所



# 特征提取器F（传统方法总结）

- 第一代：完全人工设计特征
  - 几何特征，图像模板，Fourier谱
- 第二代：（子空间）变换特征
  - PCA, LDA, LPP, SR...
- 第三代：人工设计局部特征+变换特征
  - Gabor滤波器，LBP + PCA, LDA等





# 特征提取器F（传统方法总结）

- 分步法：预处理+人造特征+分类器学习
  - 预处理
    - 光照归一、姿态矫正、去遮挡、去表情...
  - 人造特征（Knowledge-driven）
    - 滤波器（Haar-like, SURF, Gabor）
    - 梯度统计（SIFT, HOG, LBP）
  - 分类器学习（Data-driven）
    - 线性变换(PCA, LDA, MDS, Linear SVM)
    - 非线性方法
      - 流形学习, Kernel, 非线性升维, 分段线性（AdaBoost, DT）



# 特征提取器F（传统方法总结）

- 分步法的缺陷 → 端到端/Joint Learning
  - 通用特征不适应特定问题
  - 问题特定特征设计困难
  - 各个步骤最优未必全局最优
- 非线性能力的缺失 → 直接的非线性映射
  - 流形学习
    - 非显式变换；Novel样本不适用；分段线性
  - Kernel方法
    - 黑盒子；有限种类核函数；核函数不可学



# 人脸识别近期发展的最大特点

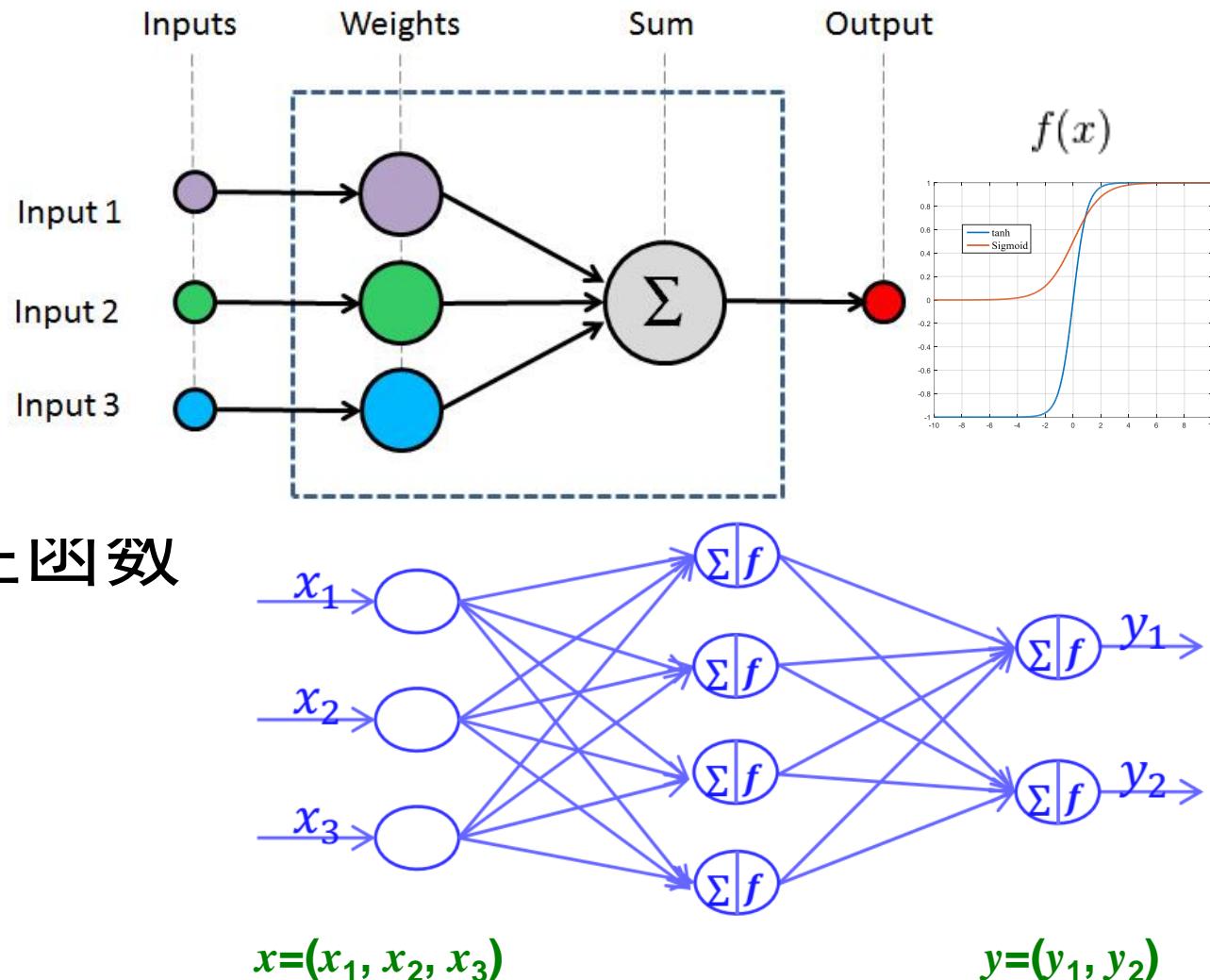
- 通用的检测与识别技术几乎完全适用于人脸处理！

**最深刻的变化：人脸识别  
不再搞特殊化了！**

# 卷积神经网络(CNN)变迁

- 前馈神经网络初步
  - 加权求和(卷积)
  - 非线性操作
    - 二值化, Sigmoid
    - ReLU
  - 层级前馈连接
- 本质上要学习非线性函数

$$y=F(x)$$

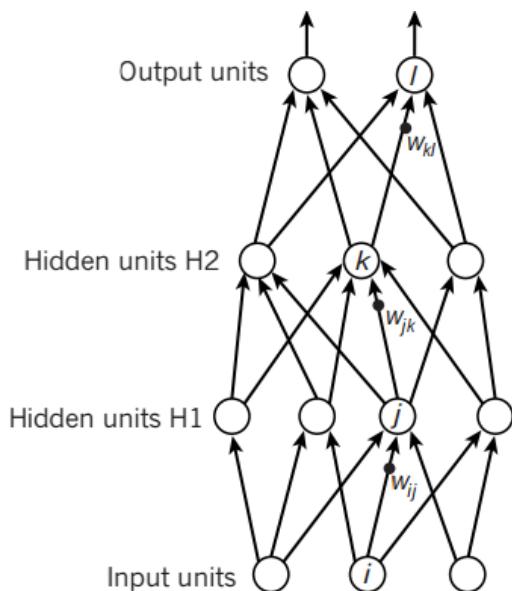


# 卷积神经网络(CNN)变迁

## ■ 优化方法（学习权重和偏置）

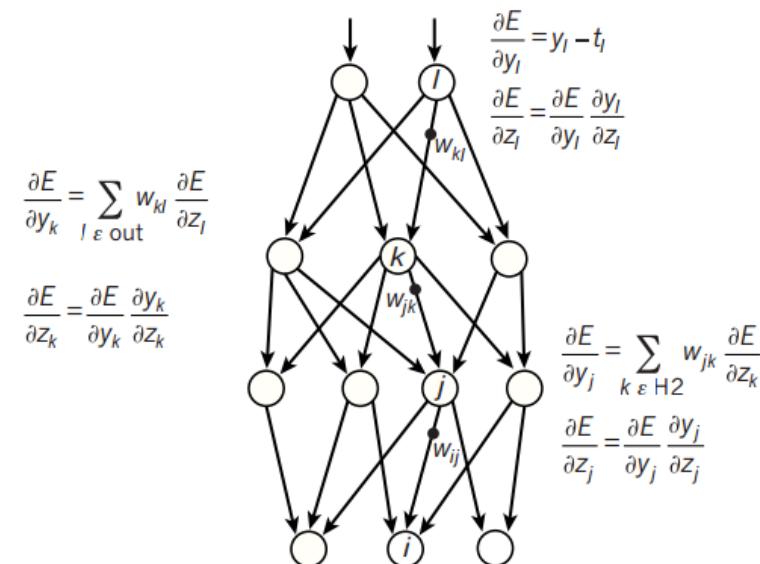
- 1974年Webos在博士论文中首次提出BP算法，但未引发关注
- 目前广泛使用的BP算法诞生于1986年
- 以全连接层为例：链式求导，梯度反向传播

### 网络前向计算



$$\begin{aligned}y_l &= f(z_l) \\z_l &= \sum_{k \in H2} w_{kl} y_k \\y_k &= f(z_k) \\z_k &= \sum_{j \in H1} w_{jk} y_j \\y_j &= f(z_j) \\z_j &= \sum_{i \in \text{Input}} w_{ij} x_i\end{aligned}$$

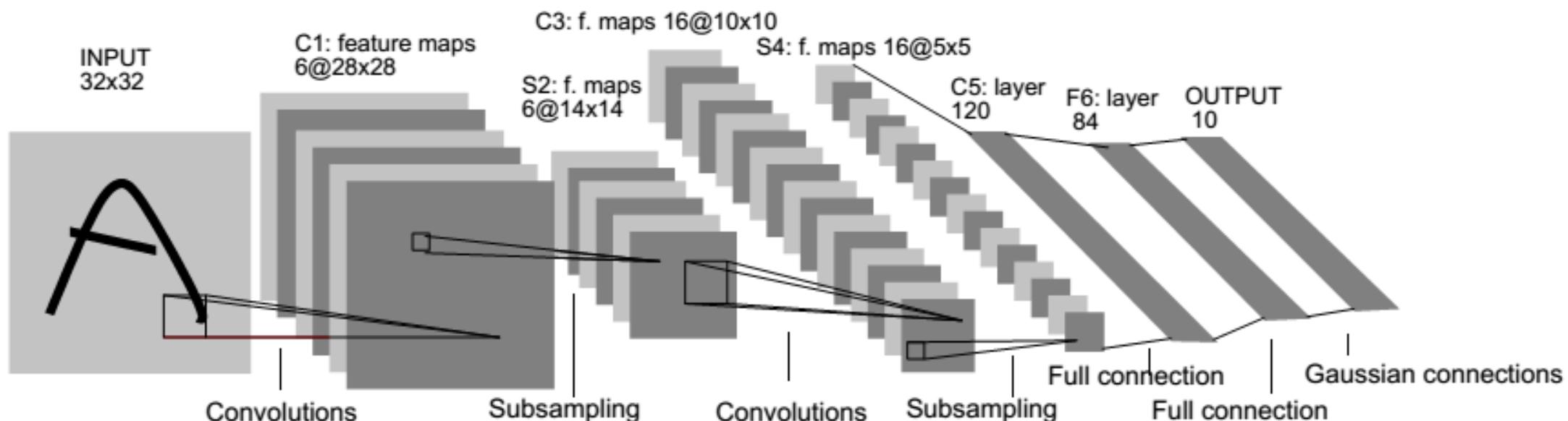
### 误差反向传播



Werbos, P.. Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. PhD thesis, Harvard University, 1974  
 Rumelhart, David E.; Hinton, Geoffrey E.; Williams, Ronald J. (8 October 1986). "Learning representations by back-propagating errors". *Nature* 323 (6088): 533–536.

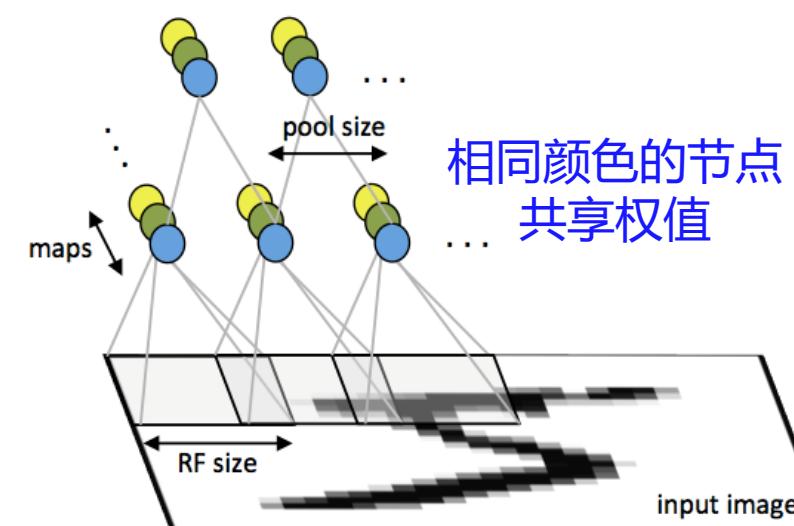
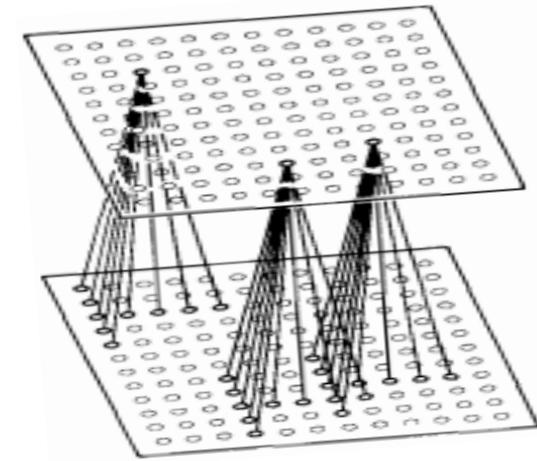
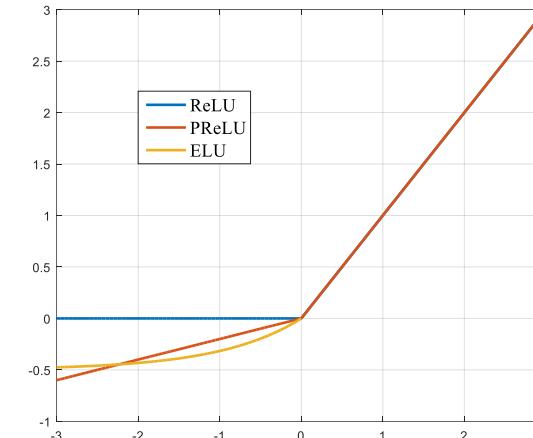
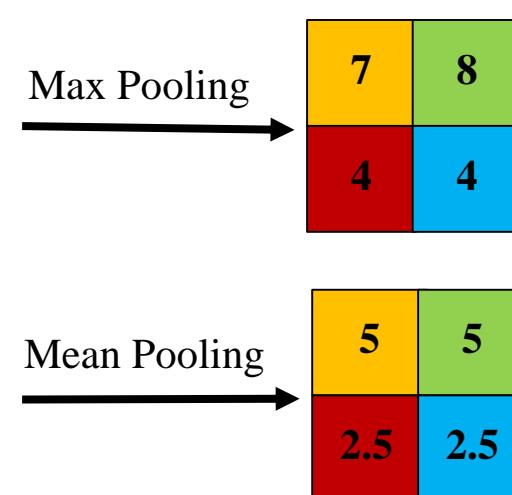
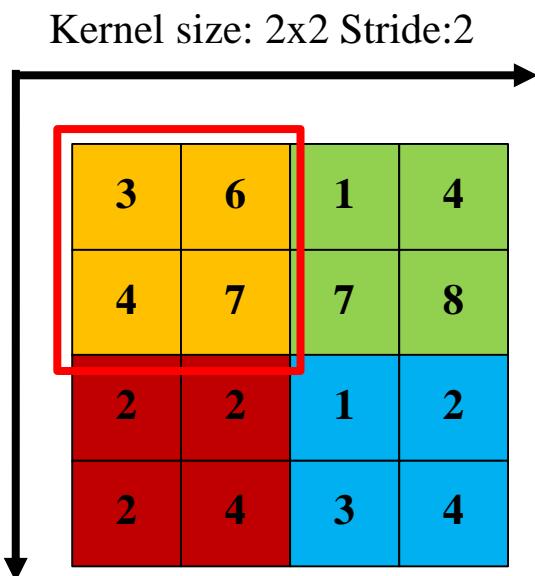
# 卷积神经网络(CNN)

- CNN层叠以下三级操作+全连接层
  - 卷积层：局部连接（卷积，滤波器...）
  - Pooling层：下采样，降维，获取不变性
  - 非线性激活

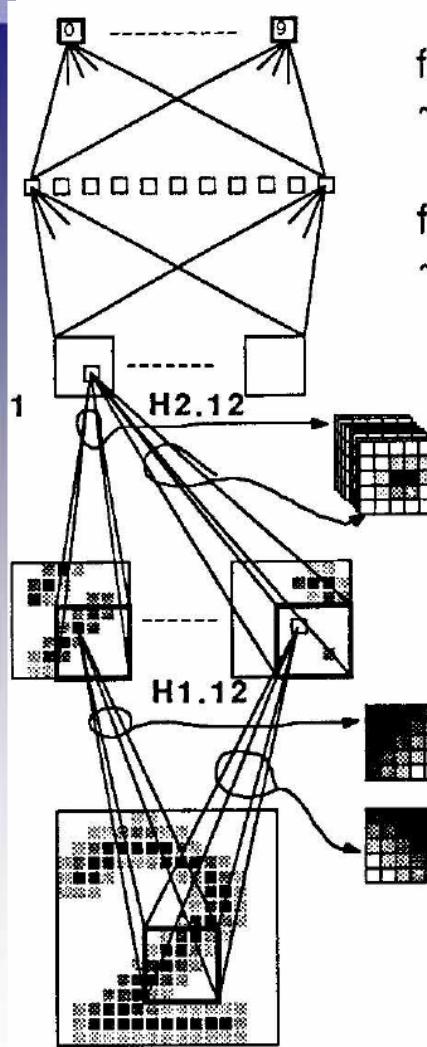


# 卷积神经网络(CNN)

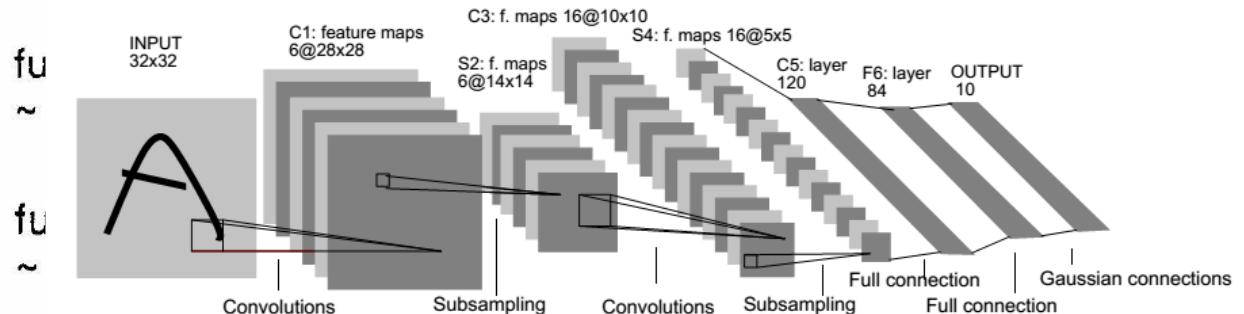
- 卷积层：局部连接（卷积，滤波器...）
  - 共享权重（若干种滤波器）
    - 大大减少需要学习的权重数
- Pooling层：降维，不变性
- 非线性激活



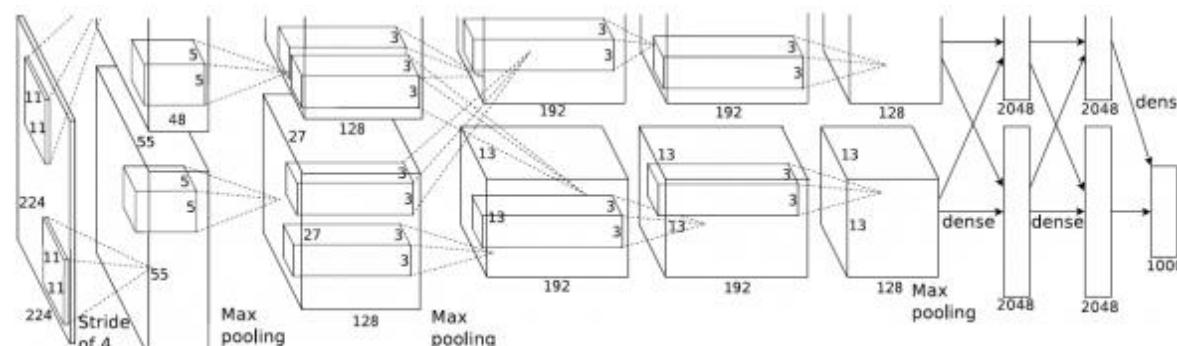
# 卷积神经网络(CNN)变迁



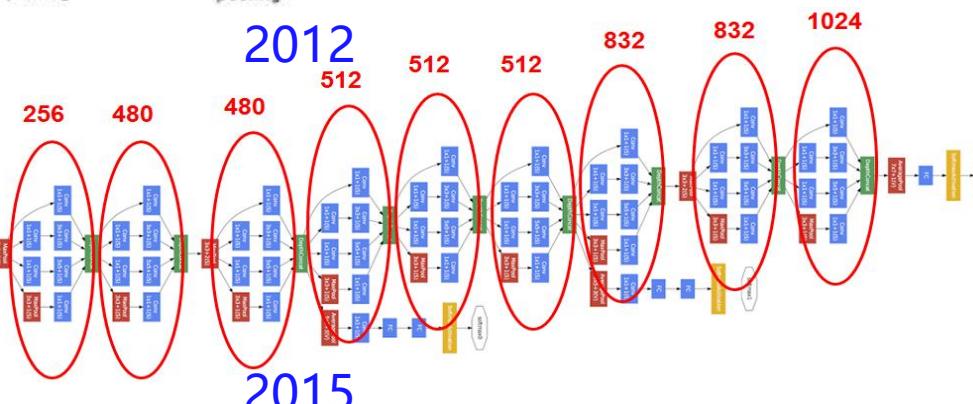
1989



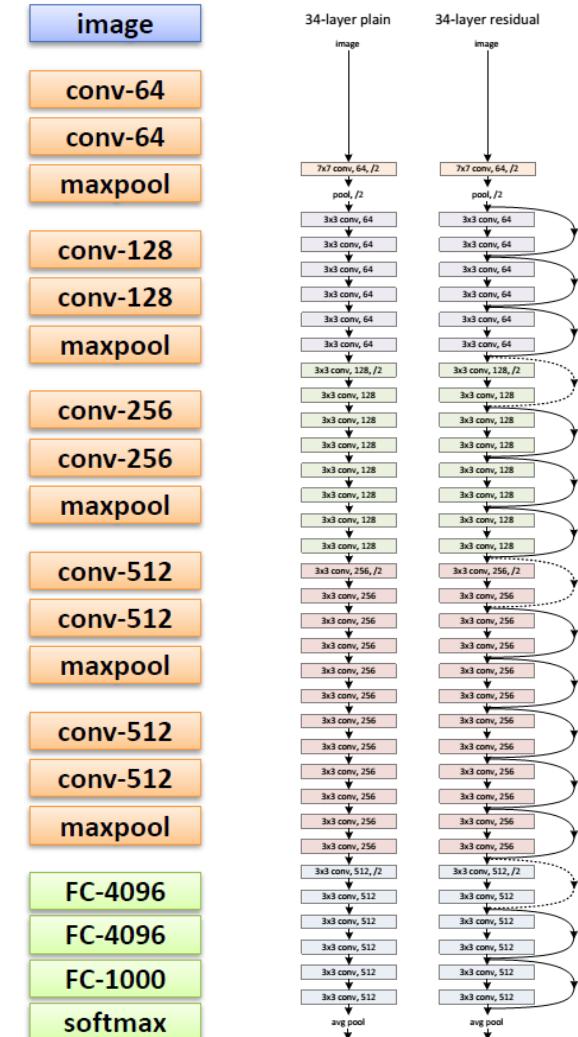
1998



2012



2015



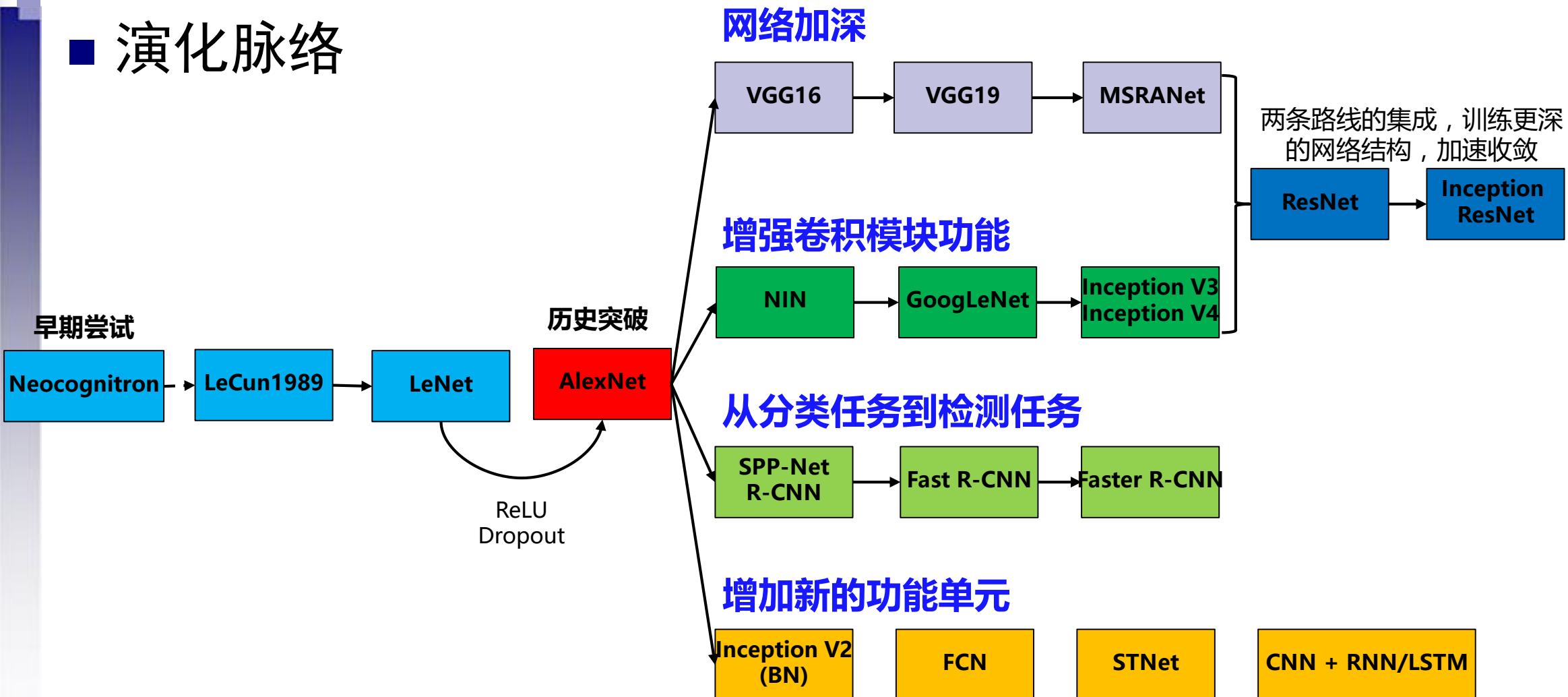
2014

2015

12

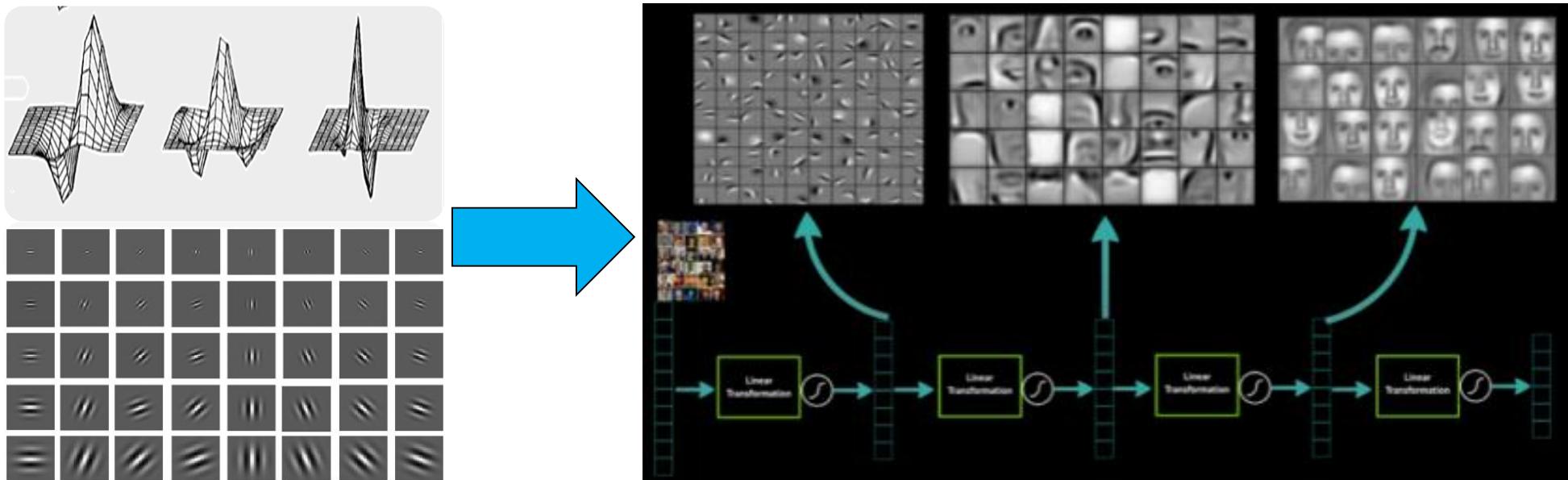
# 卷积神经网络(CNN)变迁

## ■ 演化脉络



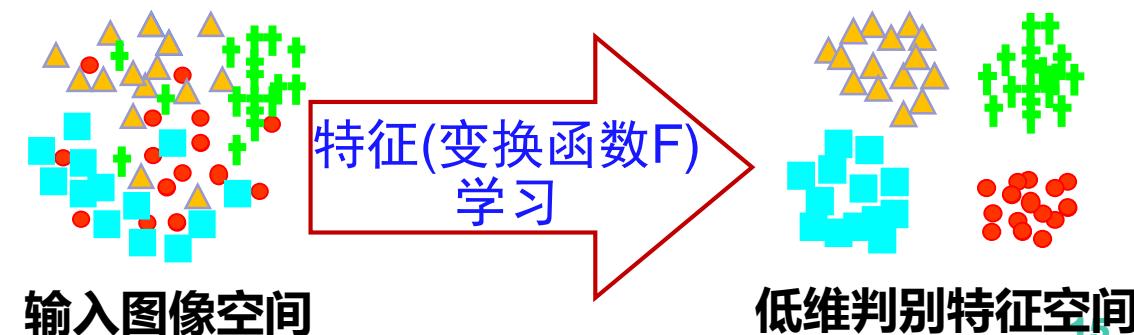
# 卷积神经网络(CNN)

- 全连接层本质上是非线性变换： $y=f(Wx)$
- 卷积层本质上是层级的、滤波型局部特征
  - 与之前局部特征的不同
    - Gabor：权值固定、人为设定（加窗傅里叶型函数）
    - CNN：权值通过数据驱动的方式学习而来（最优目标函数）



# 特征提取器F（深度学习时代）

- 第一代：完全人工设计特征 —— 知识驱动
  - 几何特征，图像模板，Fourier谱
- 第二代：（子空间）变换特征 —— 数据驱动
  - PCA, LDA, LPP, SR...
- 第三代：人工设计局部特征 + 特征变换 —— 知识 + 数据驱动
  - Gabor滤波器，LBP + PCA, LDA等
- 第四代：特征学习 —— 完全数据驱动
  - 局部特征参数可学习
  - 变换可学习
  - 非线性





# 人脸识别测试集和性能的变迁

- 人脸识别技术史本质上是评测人脸库更替史！
- 学术界的里程碑（数据库）
  - ORL, Extended Yale B: 1990~2012 (人数少于40人)
    - 首选识别率: 95%~99% [J.Wright et al, 2008]
  - FERET: 1994~2010 (1196人, 每人2~5张图)
    - 首选识别率: 99%~94% (Dup.I&II) [S.Xie, S.Shan, X.Chen, IEEE T IP10]
  - FRGC v2.0: 2004~2012 (约500人, 每人50+张图)
    - 拒识率4% @ FAR=0.1% [Y.Li, S.Shan, H.Zhang, S.Lao, X.Chen, ACCV12]
  - LFW: 2007~至今 (~5749明星, 1680人多于2张图)
    - 正确率95.17% [D.Chen, X. Cao, F. Wen, J. Sun, CVPR13]
    - 正确率97.35% [Y.Taigman, M. Yang, M.Ranzato, L. Wolf, CVPR14]
    - 正确率99.47% [Y. Sun, X. Wang, and X. Tang, CVPR14]
    - 正确率99.63% [F. Schroff, D. Kalenichenko, and J. Philbin, CVPR15]



# LFW人脸识别评测

- 人脸识别技术史本质上是评测人脸库更替史！
- 学术界的里程碑（数据库）
  - ORL, Extended Yale B: 1990~2012 (人数少于40人)
    - 首选识别率: 95%~99% [J. Wright et al., 2008]

过去3年错误率从5%下降到0.4%  
(错300对→错24对)

- LFW: 2007~至今 (~5749明星, 1680人多于2张图)
  - 正确率95.17% [D.Chen, X. Cao, F. Wen, J. Sun, CVPR13]
  - 正确率97.35% [Y.Taigman, M. Yang, M.Ranzato, L. Wolf, CVPR14]
  - 正确率99.47% [Y. Sun, X. Wang, and X. Tang, CVPR14]
  - 正确率99.63% [F. Schroff, D. Kalenichenko, and J. Philbin, CVPR15]

# LFW人脸识别评测

## ■ Labeled Face in the Wild (LFW)

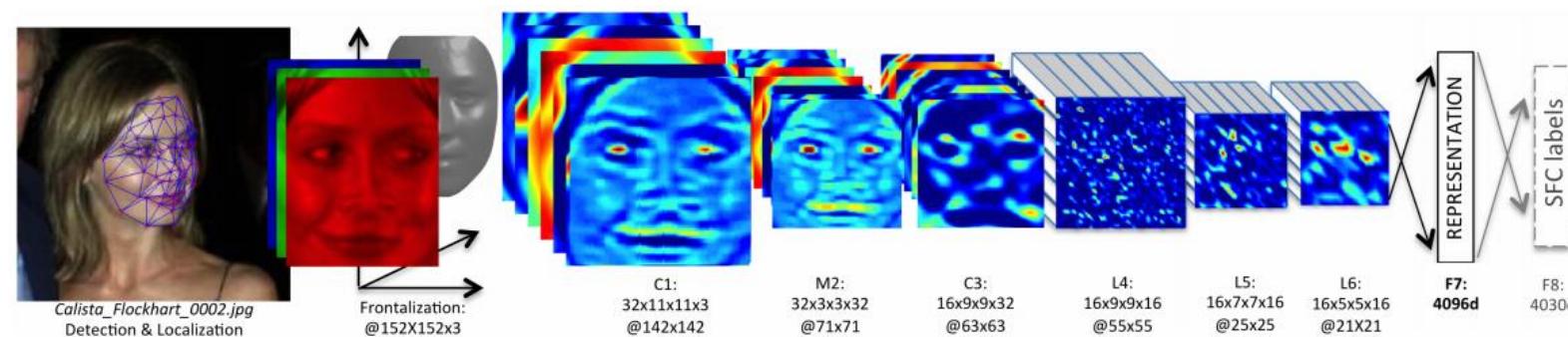
- 非限定条件下的人脸识别
- 数据来源于因特网
- 国外名人，Yahoo新闻
- 广为人知的测试模式
  - 训练集：**无限制**
  - 验证任务测试集
    - 共**6000图像对**



Huang G B, Ramesh M, Berg T, et al. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report, University of Massachusetts, Amherst, 2007.

# 过去2年的新进展——深度学习

- DeepFace(Facebook)
  - 8层网络，人脸3D正面化预处理
  - 训练数据：4K人，4.4M图像



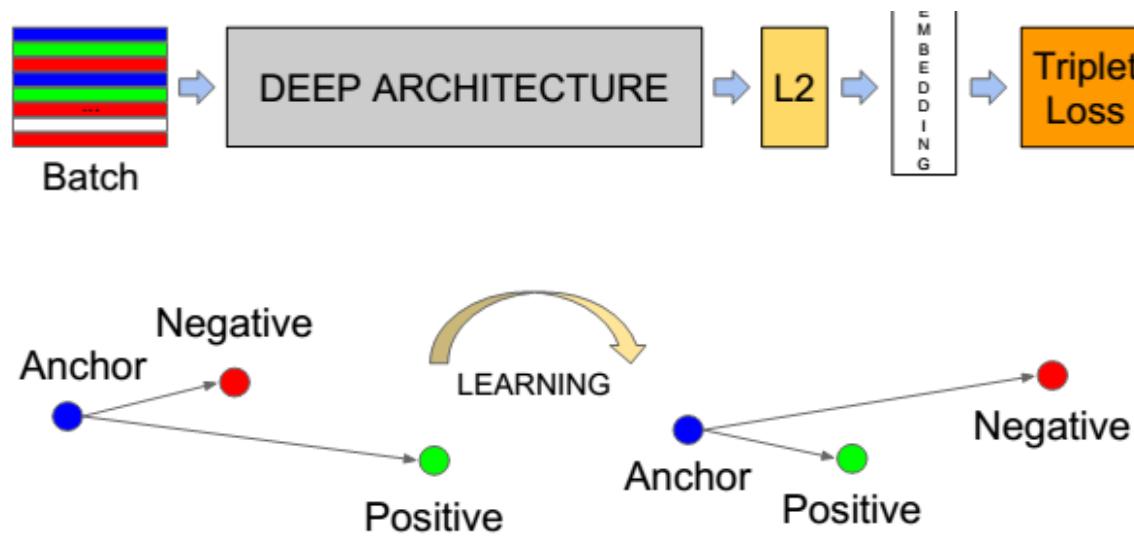
## DeepFace

[1] Taigman Y, Yang M, Ranzato M A, et al. Deepface: Closing the gap to human-level performance in face verification. CVPR, 2014.



# FaceNet

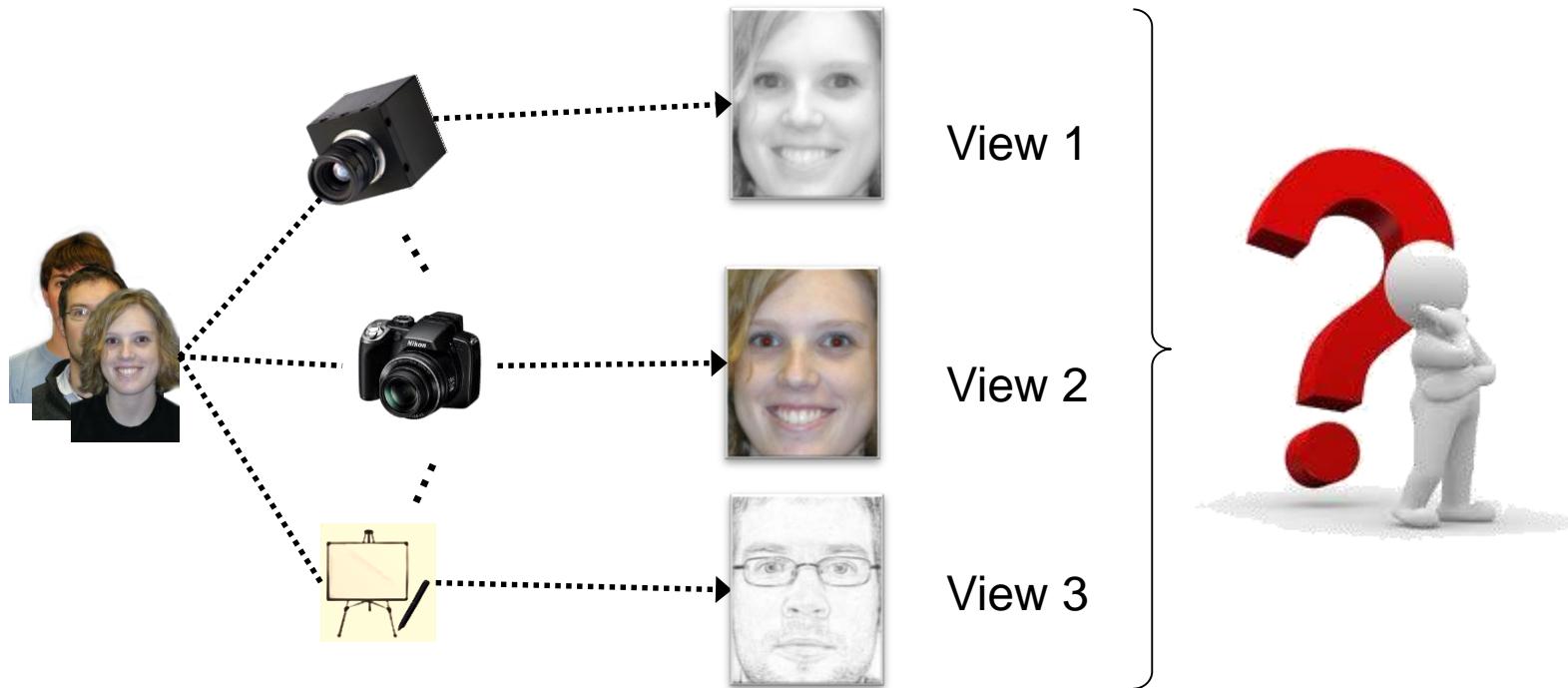
- 深层网络 ( 22 层 ) + 海量数据(800万人，2亿张图像) + Triplet Loss [F. Schroff, D. Kalenichenko, and J. Philbin, CVPR15]



$$\sum_i^N \left[ \|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right]_+$$

FaceNet中使用的Triplet Loss, 训练集中类别数为800万, 如果使用softmax loss, 输出层节点将达到800万个, 需要至少32GB显存 (假设上一个隐层节点1024个), 而triplet loss则不需要额外占用显存

# 很多问题需要解决，例1：多模态人脸比对



Meina Kan, Shiguang Shan, Xilin Chen. Multi-view Deep Network for Cross-view Classification. **CVPR 2016**.

# 很多问题需要解决，例1：多模态人脸比对

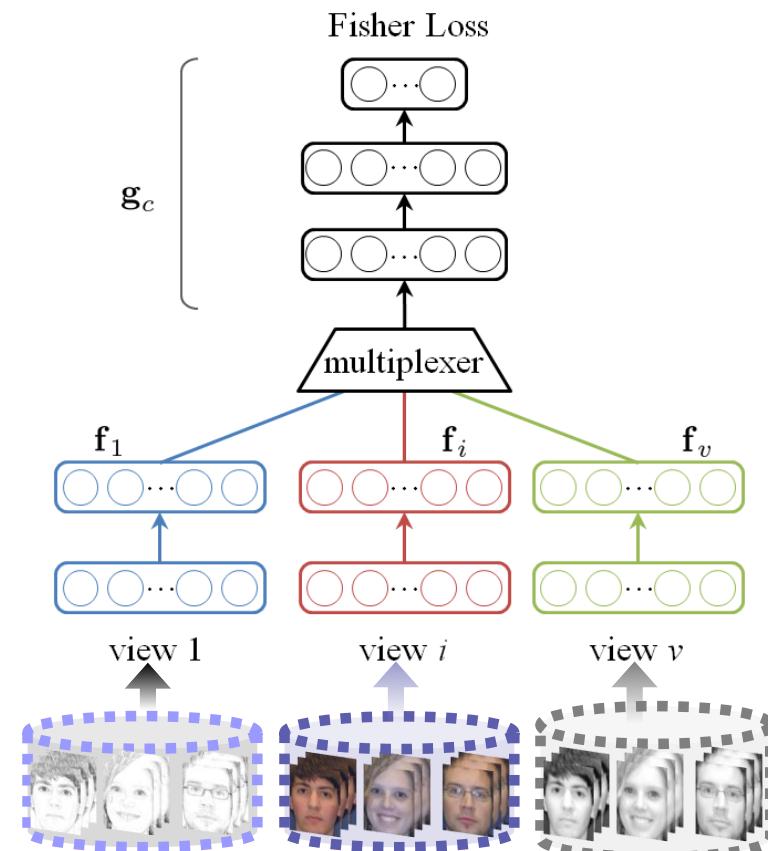
- 多视图判别分析深度网络(MvDN)
  - 基本框架：包含两个子网络

$$\mathbf{y}_i^j = \mathbf{g}_c \left( \mathbf{f}_i \left( \mathbf{x}_i^j \right) \right)$$

公共子网络      视图特定子网络

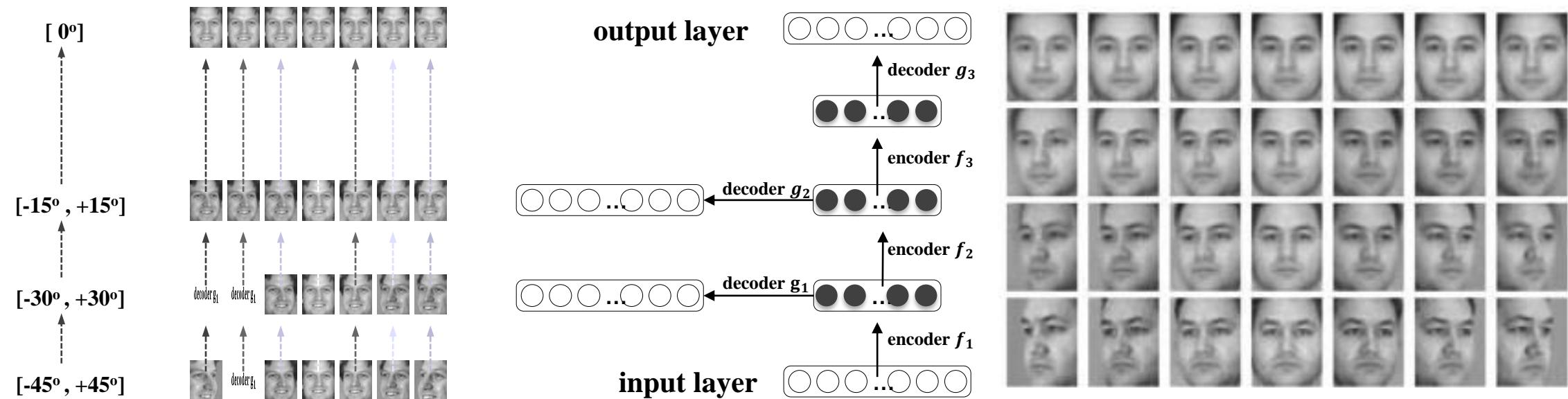
$$[\mathbf{g}_c^*, \mathbf{f}_i^*]_{i=1}^c = \arg \min_{\mathbf{g}_c, \mathbf{f}_i | i=1}^{c} Tr \left( \frac{\mathbf{S}_W^y}{\mathbf{S}_B^y} \right)$$

Meina Kan, Shiguang Shan, Xilin Chen. Multi-view Deep Network for Cross-view Classification. **CVPR 2016**.



# 很多问题需要解决，例2：大姿态问题

- 问题：非正面人脸→虚拟正面人脸
- 思路：每一层转正较小的偏转角度
- 文章：Meina Kan, et al. Stacked Progressive Auto-Encoder (SPAЕ) for Face Recognition Across Poses, CVPR2014





# 人脸识别测试集和性能的变迁

- 人脸识别技术史本质上是评测人脸库更替史！
- 学术界的里程碑（数据库）
  - ORL, Extended Yale B: 1990~2012 (人数少于40人)
    - 首选识别率: 95%~99% [J.Wright et al, 2008]
  - FERET: 1994~2010 (1196人, 每人2~5张图)
    - 首选识别率: 99%~94% (Dup.I&II) [S.Xie, S.Shan, X.Chen, IEEE T IP10]
  - FRGC v2.0: 2004~2012 (约500人, 每人50+张图)
    - 拒识率4% @ FAR=0.1% [Y.Li, S.Shan, H.Zhang, S.Lao, X.Chen, ACCV12]
  - LFW: 2007~至今 (~5749明星, 1680人多于2张图)
    - 正确率95.17% [D.Chen, X. Cao, F. Wen, J. Sun, CVPR13]
    - 正确率97.35% [Y.Taigman, M. Yang, M.Ranzato, L. Wolf, CVPR14]
    - 正确率99.47% [Y. Sun, X. Wang, and X. Tang, CVPR14]
    - 正确率99.63% [F. Schroff, D. Kalenichenko, and J. Philbin, CVPR15]

■ 应用成熟了吗？

# IEEE FG 2015视频人脸验证竞赛

## ■ 任务：视频vs视频1:1验证

### □ Exp. 1: Controlled case

- Video-to-video verification
- 1920\*1080 video captured by mounted camera

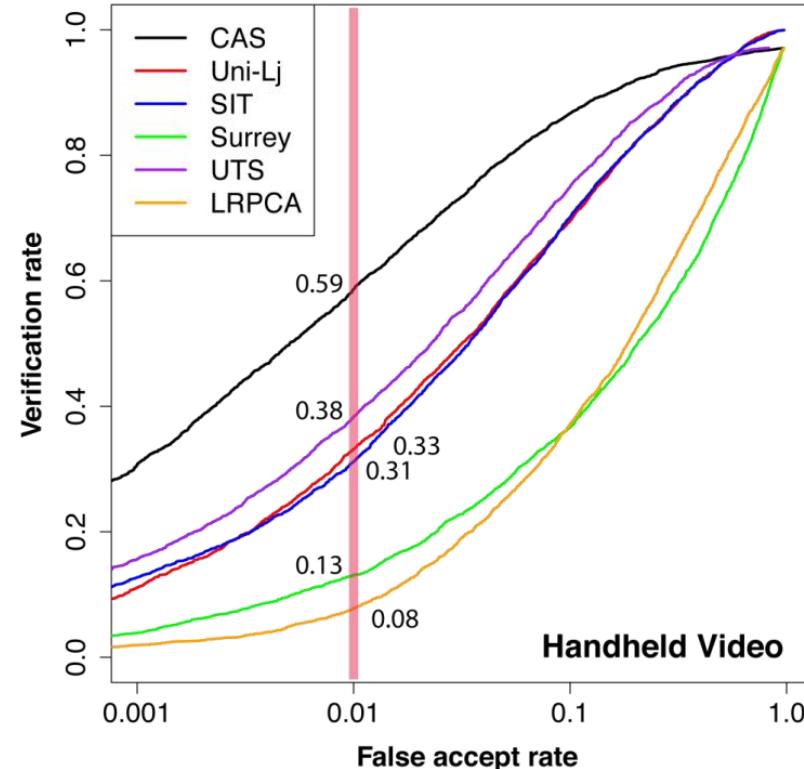
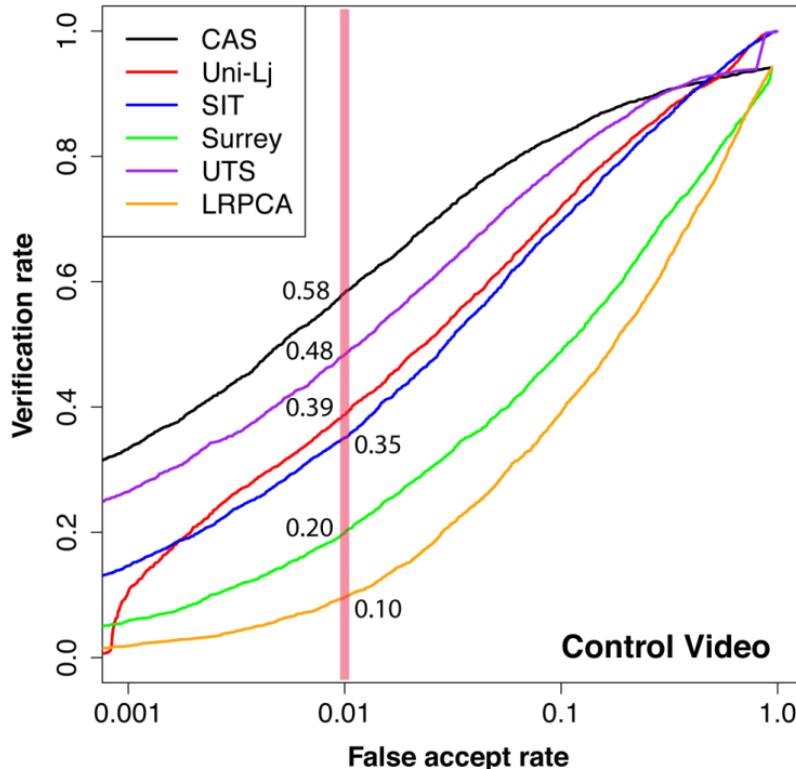
### □ Exp. 2: Handheld case

- Video-to-video verification
- Varying resolution from 640\*480~1280\*720
- Videos from a mix of different handheld point-and-shoot video cameras



# IEEE FG 2015视频人脸验证竞赛结果

Method	HOG		Dense SIFT		DCNN	
	Control	Handheld	Control	Handheld	Control	Handheld
HERML	25.26	19.28	<b>33.82</b>	<b>28.93</b>	<b>58.63</b>	<b>59.14</b>





# 人脸识别的应用现状

- 1:N (配合照检索应用, 基本成熟)
  - 亿级照片库: ~90% (首选正确率)
- 1:1验证 (配合用户, 可应用)
  - 门禁考勤 (99%@FAR=0.01%)
  - 身份证vs现场照片
    - 场景1: 卡内小图 (~90%@FAR=0.01%)
    - 场景2: GA原大图 (~95%@FAR=0.01%)
    - 场景3: 中图水印 (~85%@FAR=0.01%)
- 1:N+1 (万级以上非配合用户, 任重道远)
  - LFW Open Set (596人) : 90+%@FAR=1%
  - 监控库中1000人: 50%@FAR= $10^{-6}$  (猜测) ——百万分之一的虚警率
  - 监控库中1万人: 10%@FAR= $10^{-7}$  (猜测) ——千万分之一的虚警率



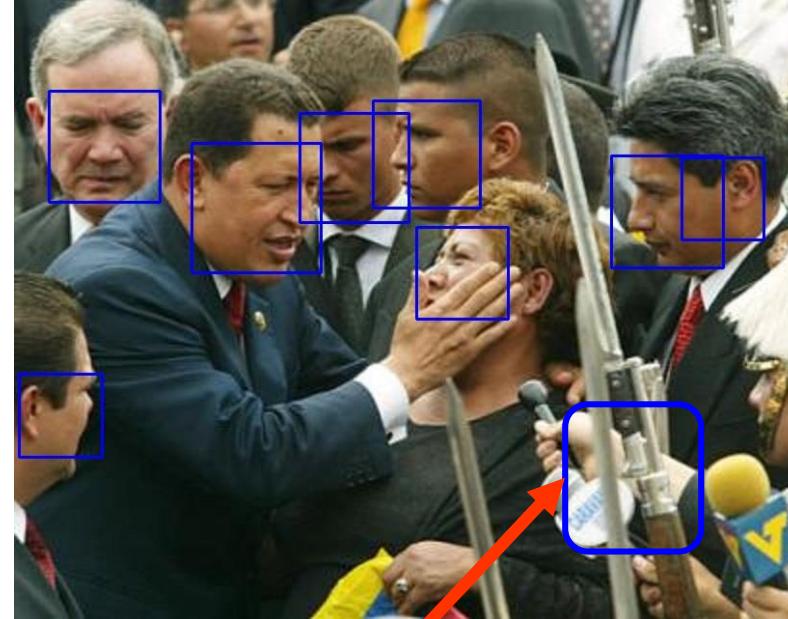
# 提纲

- 背景与概况
- 人脸识别技术的深度化
  - 人脸检测技术
  - 面部特征点定位技术
  - 判别特征学习与匹配方法
  - SeetaFace人脸识别引擎
- 总结与讨论

# 人脸检测技术



评价指标： 检测率 (1-漏检率) +

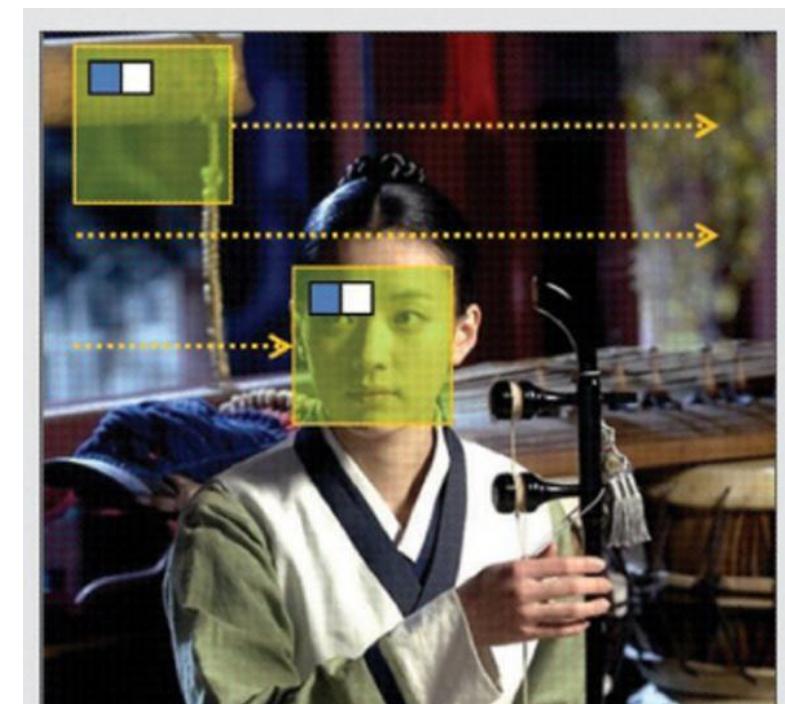


误检率 (假脸)

# 2000年，人脸检测的第一次突破

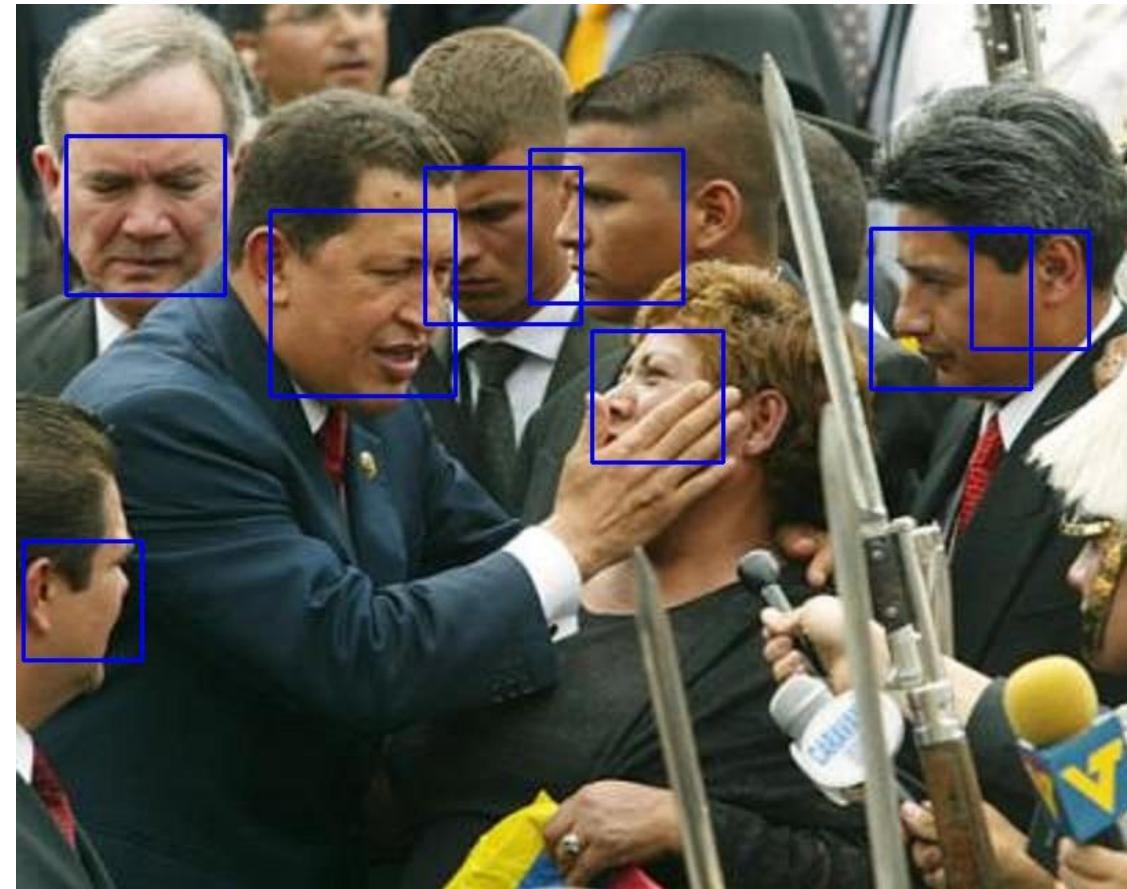
## ■ Viola and Jones (2000)

- 滑动窗口式:所有可能的位置、尺度和长宽比
- Haar-like feature + AdaBoost + Cascade



# 物体检测：内涵与总体思路

- 有没有？
- 在哪里？





# 2014年以来的变迁

## ■ 有没有？

- 从人造特征+分类器学习 “两步法”
- 特征和分类器 End-to-end 学习
- 二类分类 → 多类分类 (人脸,人体,马,牛,羊...)

## ■ 在哪里&有多大？

- Check 所有可能位置和大小
- 只 check 可能位置和大小
- 回归位置和大小(更准确)



# 2014年以来的变迁

■ 有没有?

从人造特征+分类器学习“两步法”

两个关键词：

深度(特征)学习 & BB回归

→ 只check可能位直和大小

→ 回归位置和大小(更准确)



# 基于深度学习的物体检测

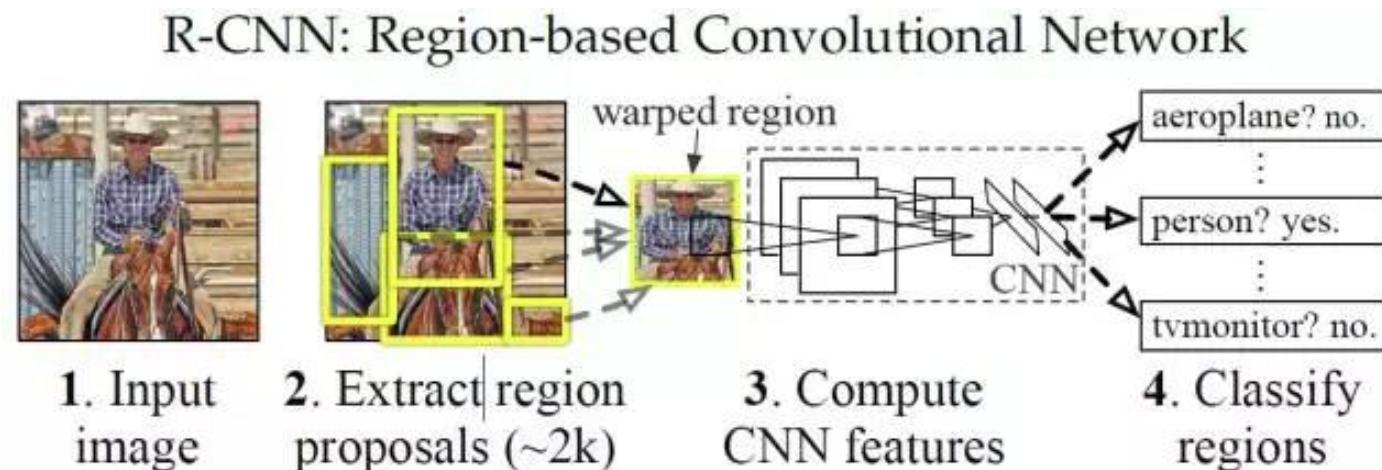
- Region CNN系列
  - R-CNN → Fast R-CNN → Faster R-CNN
- 回归位置
  - YOLO
- 传统方法与深度方法的结合
  - FacenessNet
  - Cascade CNN (coarse-to-fine)

# R-CNN (R. Girshick, CVPR14, TPAMI15)

## 关键点

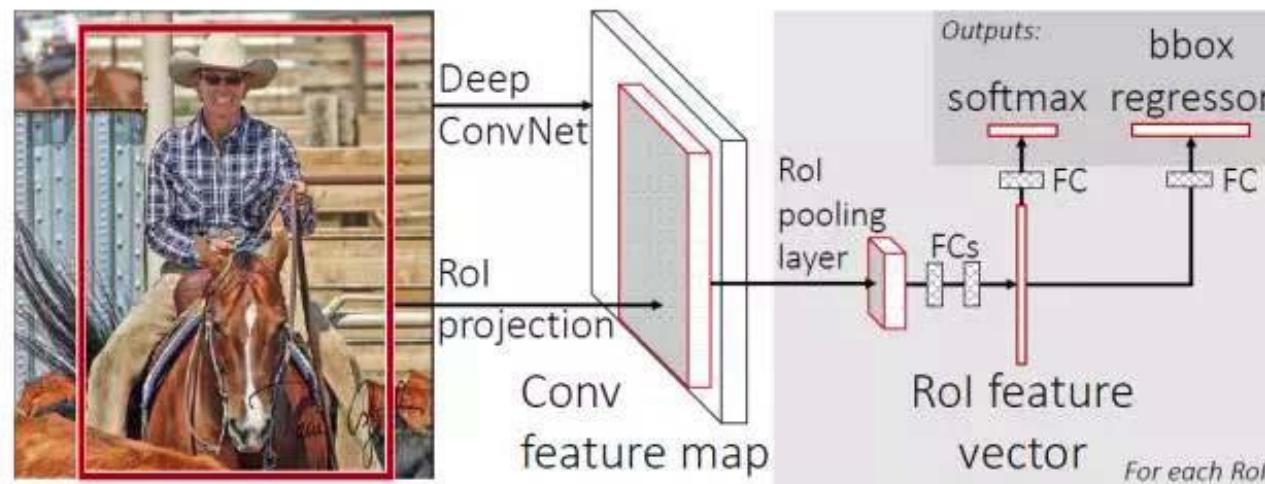
- Selective Search产生约2K候选窗口
- 每个候选窗缩放到227\*227用CNN提取特征
- SVM分类 (+Bounding box regression)

## Too Slow!! GPU还要40+s



# Fast R-CNN (R. Girshick, ICCV2015)

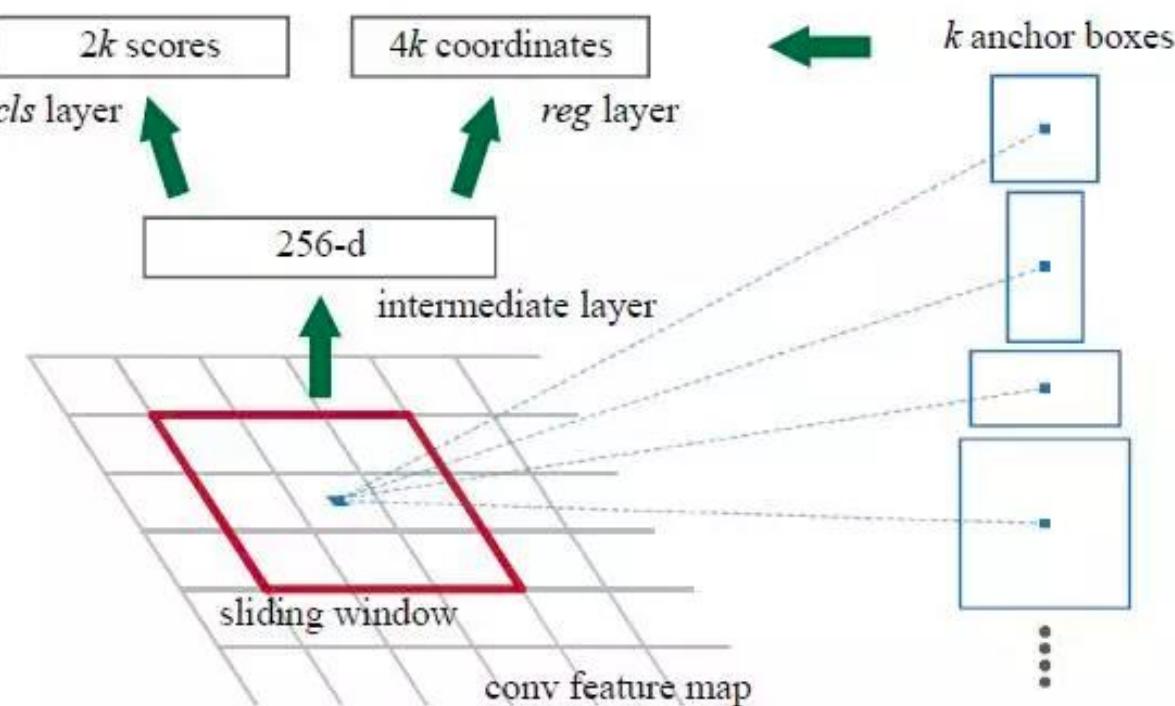
- 采用简单的pooling策略解决RoI窗口大小不一问题
  - 每个候选RoI窗 → 下采样为  $7 \times 7$  (特征:  $7 \times 7 \times 512$ )
- 多任务损失函数
  - Softmax分类 (替换了SVM)
  - **Bounding Box Regression**



缺点：还是不够快！

# Faster R-CNN (S.Ren et al NIPS2015)

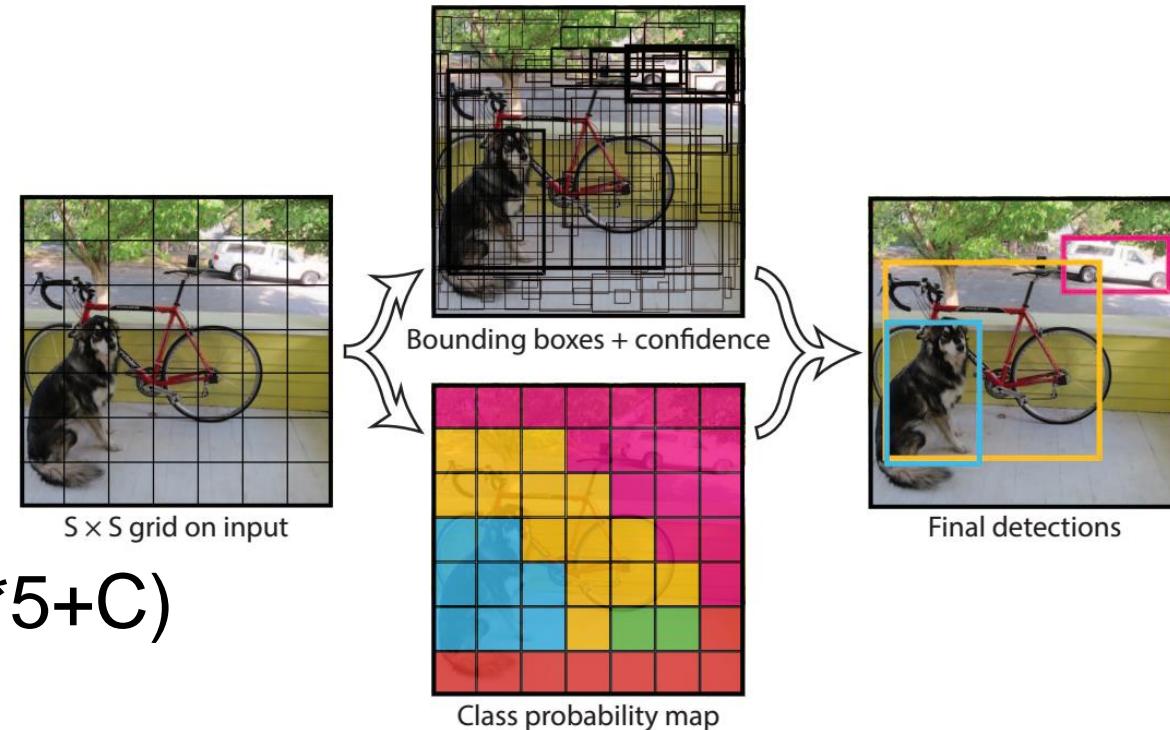
- 用 **Region Proposal Networks** 替换 Selective Search(2-3s)
  - 采用 **anchor boxes** 机制，在最后层卷积图上滑窗，用  $K$ (特征图数) 维特征进行分类和 BB 回归
  - 共 9 种 Anchor Boxes
    - 3 种尺度，3 种长宽比
  - 卷积图大小  $40 \times 60$
  - 最终  $40 \times 60 \times 9$  个候选
  - 取前面若干(300)即可
- 精度高，速度快
  - 5fps (GPU 上)



# YOLO(CVPR2016)

作者: Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi

- 将输入图像划分为 $S \times S$  grid
- 对grid cell预测
  - B bounding boxes: x,y,w,h
  - 每个BB的信度(conf.)
  - C class probabilities
  - 最终预测张量
    - $S \times S \times (B \times 5 + C)$
- 网络全连接层→张量( $7 \times 7 \times 2 \times 5 + C$ )
- 用NMS去除冗余
  - Non-max suppression
- **速度快: 45fps (GPU上)**





# CNN Cascade

H. Li , Z. Lin , X. Shen, J. Brandt and G. Hua, CVPR 2015

- 混合传统方法和深度学习方法
  - 简单CNN→复杂CNN→更复杂CNN
    - #1 CNN: 1卷积层+1全连接层, 12x12, 步长4像素
    - #2 CNN: 1卷积层+1全连接层, 24x24, 更多卷积核
    - #3 CNN: 2卷积层+1全连接层, 更多的卷积核
  - 效果
    - 精度: 在FDDB人脸库上, **85.67%@fp=100**
    - 速度: 640x480图14fps (CPU上, 最小人脸80x80)
    - 速度: 640x480图100 fps (Titan Black GPU)



# 物体(人脸)检测：总结与问题

## ■ 总结

- 深度学习方法带来性能大提升(**工业界90%以上，我们:92%**)
- Bounding Box Regression (BBR)兴起
  - 遍历式滑窗受到挑战；有利于检测精度
- 样本组织方式的变革
  - 传统：归一化大小，预处理，虚拟样本生成
  - 现在：无需这么麻烦 (hard negative mining很重要)

## ■ 问题

- 高精度、全姿态时的速度问题
  - CPU, 手机, 嵌入式设备/低配置终端
- 小尺度，大遮挡、复杂光照情况时的漏检



# 提纲

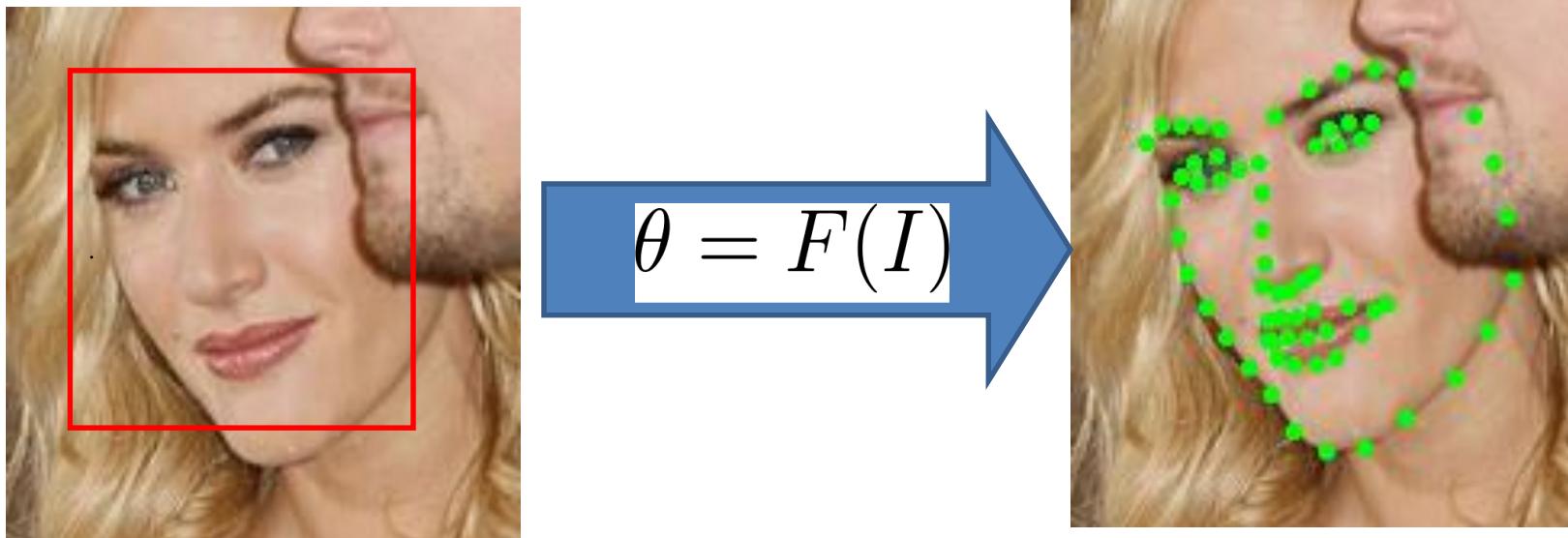
- 背景与概况
- 人脸识别技术的深度化
  - 人脸检测技术
  - 面部特征点定位技术
  - 判别特征学习与匹配方法
  - SeetaFace人脸识别引擎
- 总结与讨论

# 面部关键特征定位技术

- ASM/AAM → 级联回归模型 → 深度级联回归

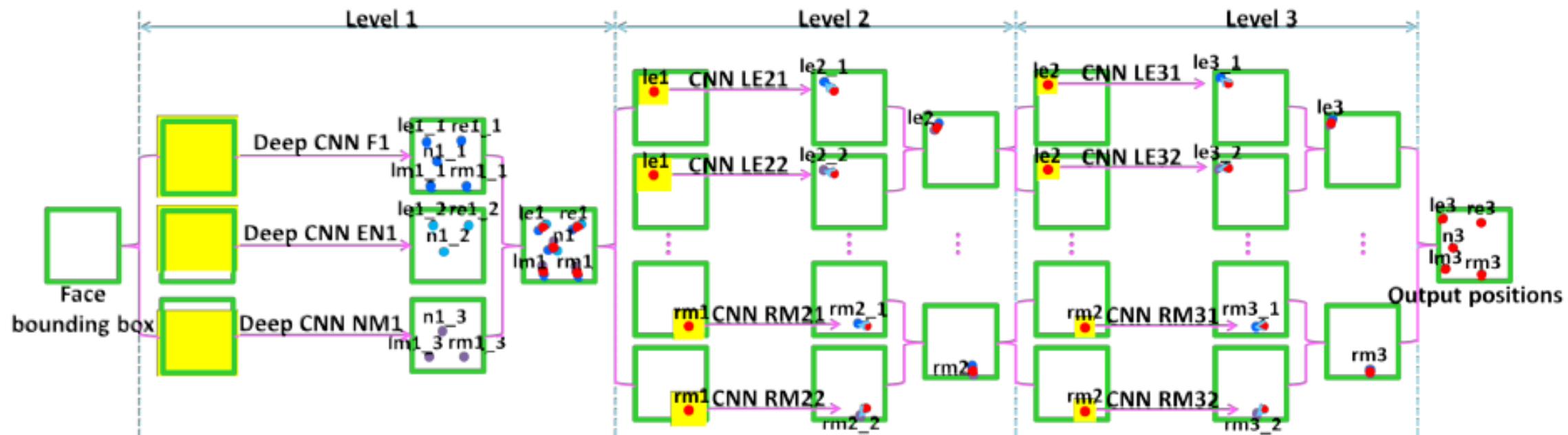
$$\theta = F(I) = f_n(f_{n-1}(\dots f_1(\theta_0, I), I), I)$$

$$\theta_i = f_i(\theta_{i-1}, I), \quad i = 1, \dots, n$$



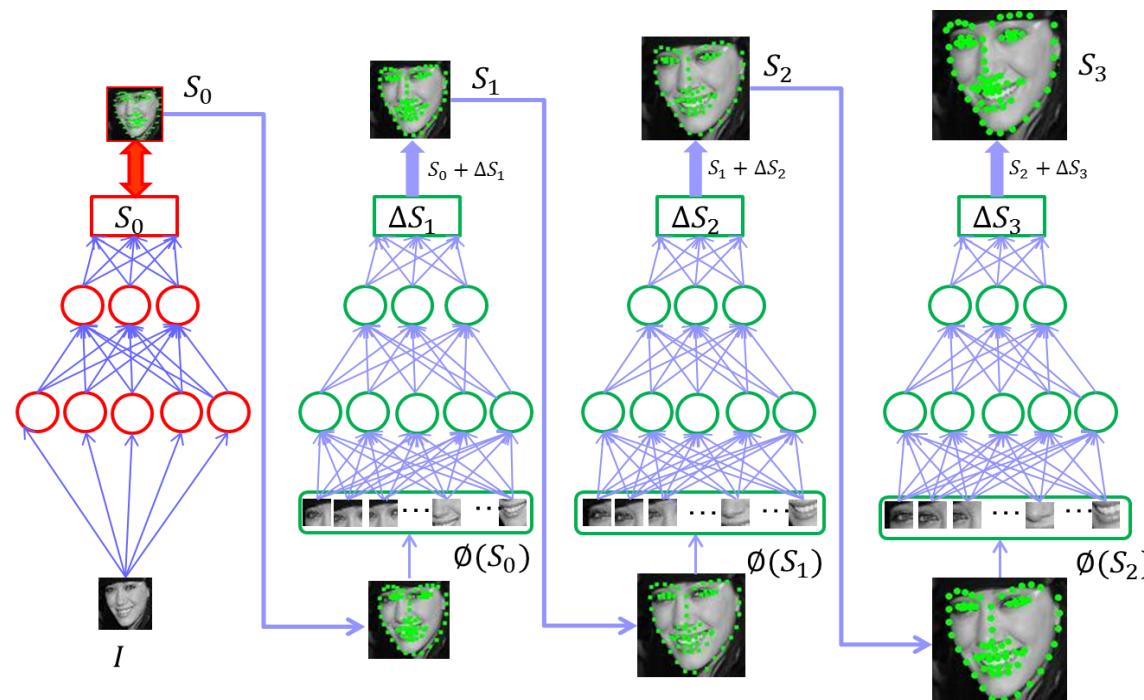
# 深度级联回归

- 基于多级DCNN实现f函数
  - Yi Sun, Xiaogang Wang, Xiaoou Tang, CVPR2013

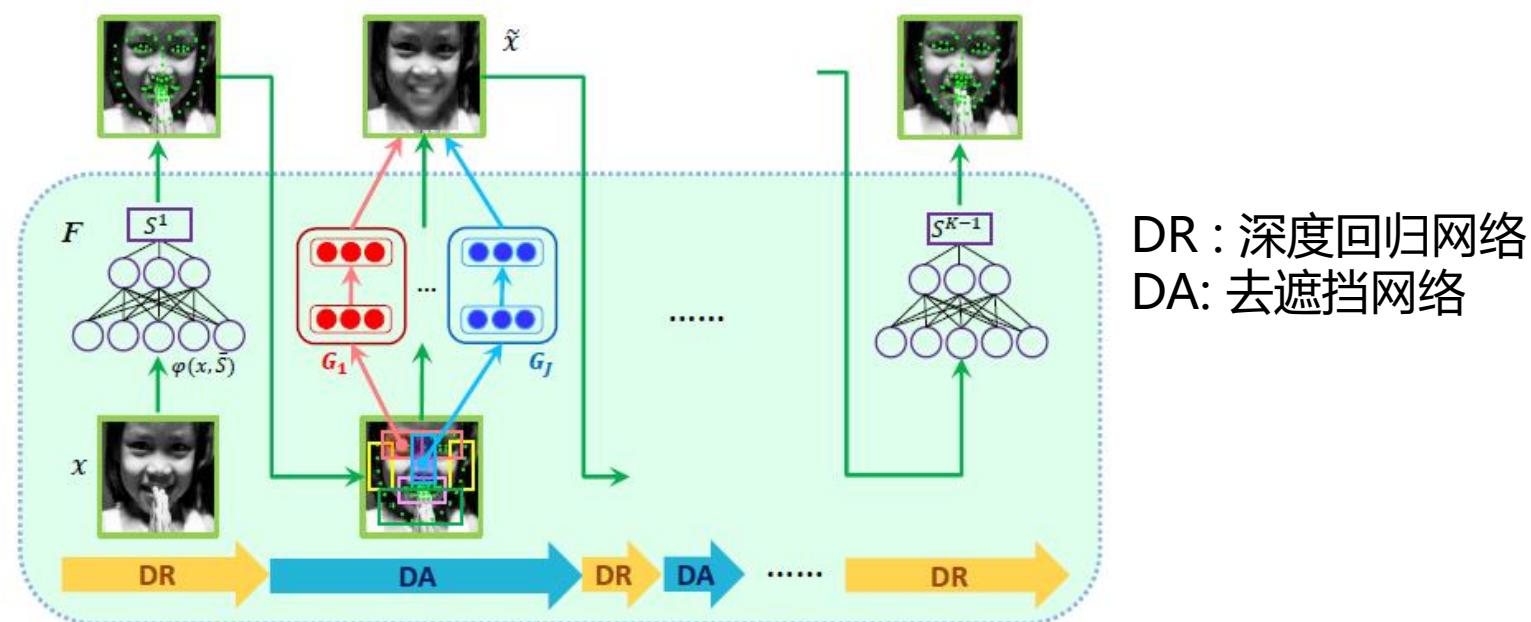


# 深度级联回归

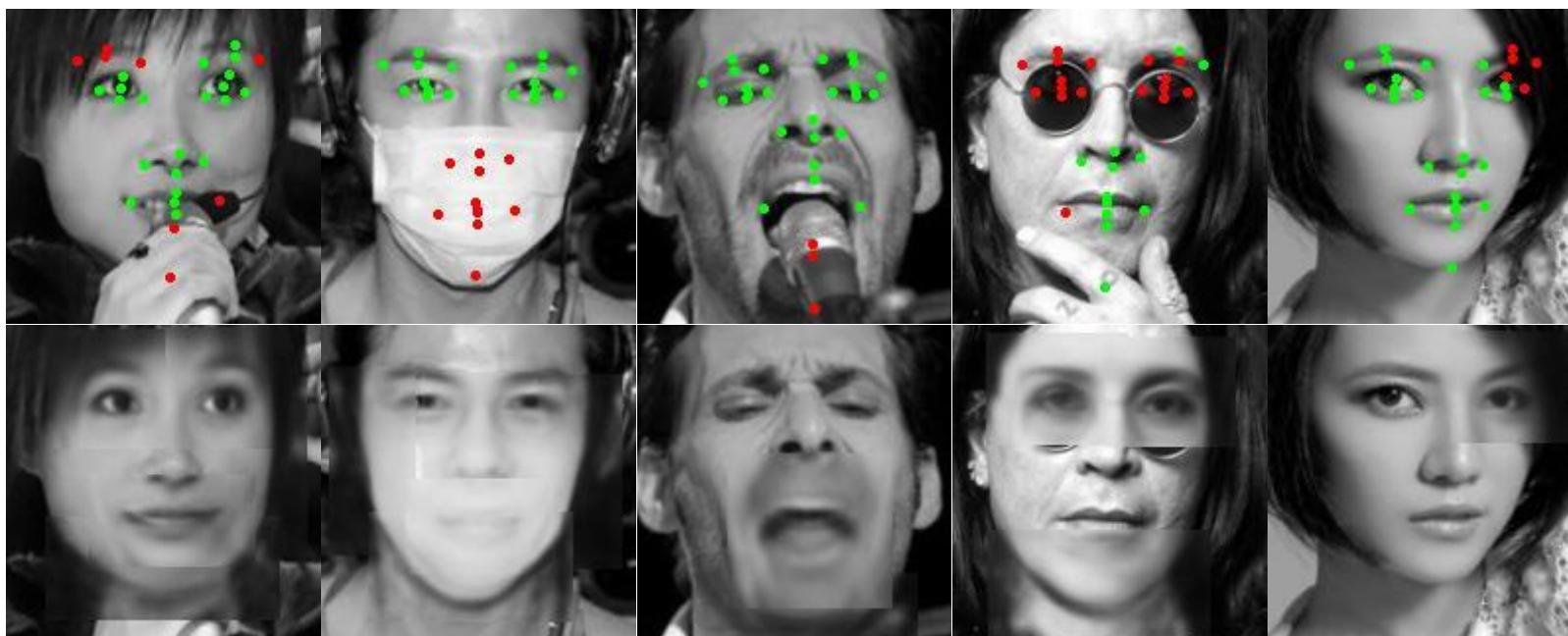
- 基于多级DCNN实现f函数
- 基于多级Auto-encoder实现f函数
  - Jie Zhang, Shiguang Shan, Meina Kan, Xilin Chen, [ECCV2014](#)



- 基于多级DCNN实现f函数
- 基于多级Auto-encoder实现f函数
- 对遮挡鲁棒的特征点定位方法
  - Jie Zhang, Meina Kan, Shigang Shan, Xilin Chen, **CVPR2016**



- 基于多级DCNN实现f函数
- 基于多级Auto-encoder实现f函数
- 对遮挡鲁棒的特征点定位方法
  - Jie Zhang, Meina Kan, Shigang Shan, Xilin Chen, **CVPR2016**





# 提纲

- 背景与概况
- 人脸识别技术的深度化
  - 人脸检测技术
  - 面部特征点定位技术
  - 判别特征学习与匹配方法
  - **SeetaFace人脸识别引擎**
- 总结与讨论

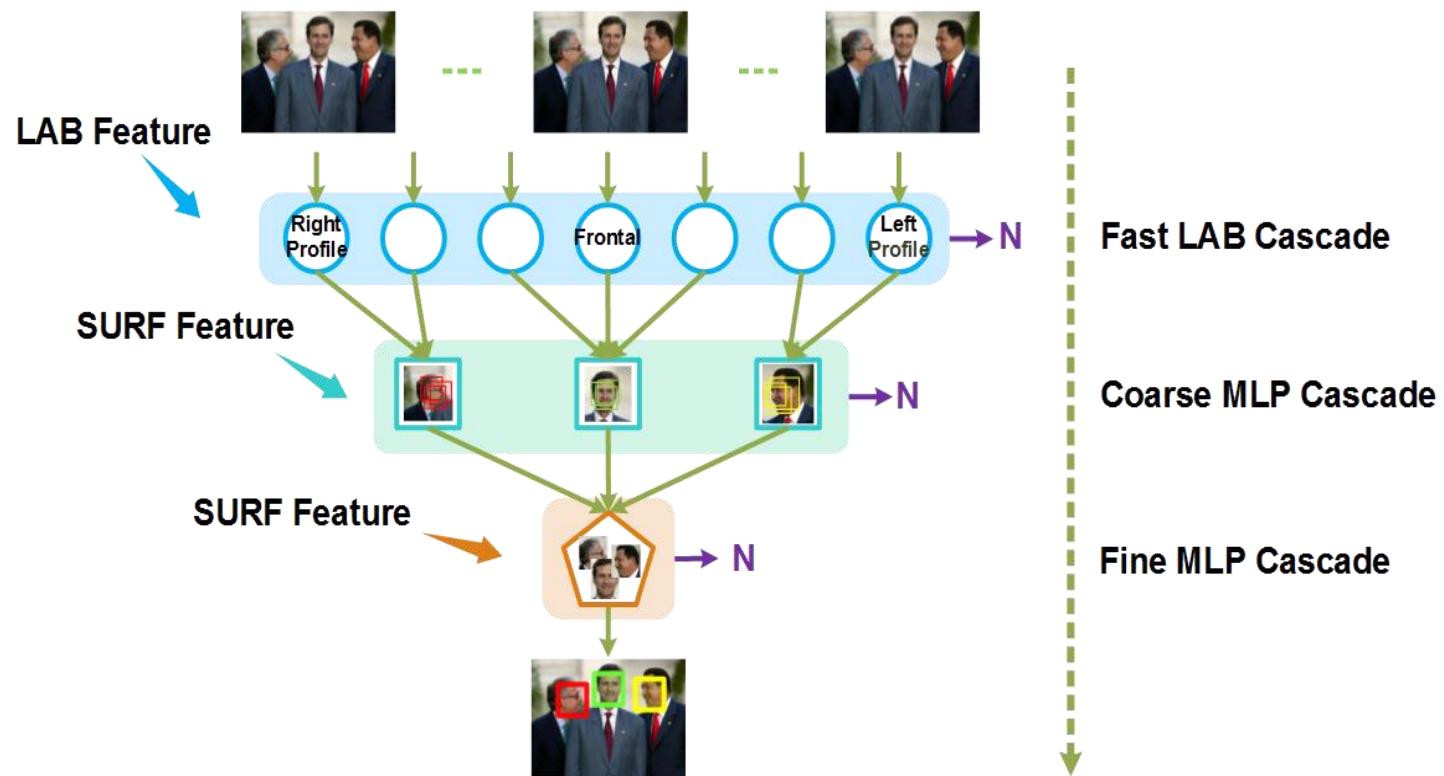


# SeetaFace引擎的组成

- SeetaFace Detector
  - Funnel-Structured cascade
  - 传统特征+级联结构+MLP
- SeetaFace Alignment
  - Coarse-to-Fine Auto-encoder Networks
  - 深度级联形状回归模型
- SeetaFace Identification
  - A CNN with 7 conv + 3 FCs
  - 深度卷积神经网络

# 1、SeetaFace检测器

- 漏斗结构的级联人脸检测模型
  - 区别于已有的parallel, tree, or pyramid structure



Shuzhe Wu, Meina Kan, Zhenliang He, Shiguang Shan, Xilin Chen. *Funnel-structured cascade for multi-view face detection with alignment awareness*. Neurocomputing (to appear).



# 1、SeetaFace检测器

- 特点：由粗到精，平衡了速度和精度
  - #1. Adaboost with LAB features
    - Rapidly reject most non-face windows while keeping high face recall
  - #2. Coarse MLP cascade with SURF feature
    - Roughly refine candidate windows at low time cost
  - #3. Fine MLP cascade
    - With shape-indexed SURF feature
    - Accurately determine face windows

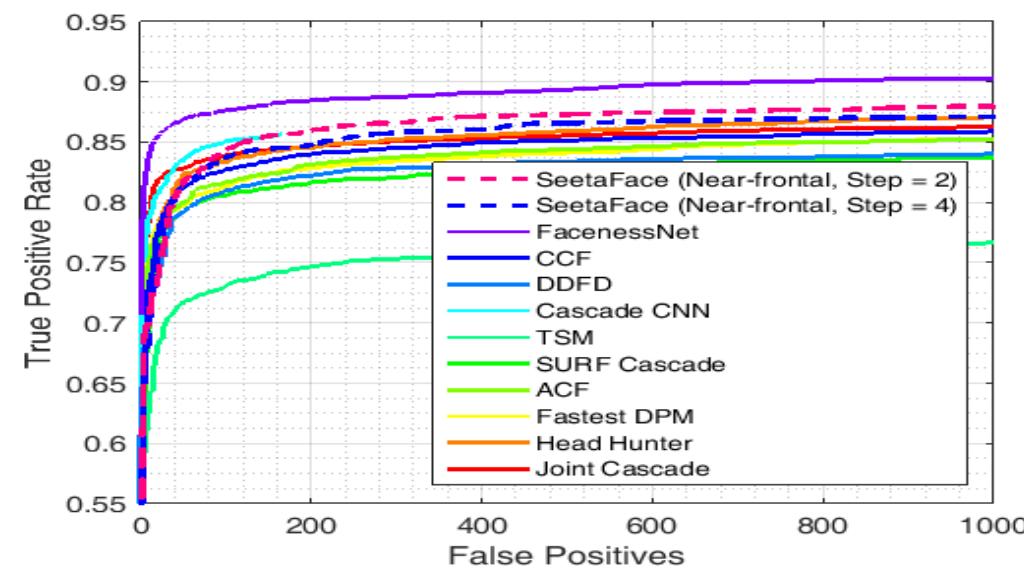
# 1、SeetaFace检测器

- 开源版本的基本情况
  - 准正面人脸检测器
    - 可以检测45度以内人脸
  - 额外的增强模块
    - 非极大值抑制（NMS）
    - BBR (Bounding box regression)



# 1、SeetaFace检测器：精度

- Accuracy on FDDB (2,845 images)
  - Standard: IoU > 0.5
- TPR=**84.4%**@100FPs
  - FacenessNet: 87%@100 FPs
  - Cascade CNN: 85%@100 FPs





# 1、SeetaFace检测器：速度

- 在VGA size image (640×480)上的速度
  - 单颗3.4GHz i7 CPU上

Detector	Min Face Size	Speed on CPU	Speed on GPU
SeetaFace Detector	40×40	16 FPS	-
	80×80	55 FPS	-
Cascade CNN <sup>†</sup>	80×80	9.1 FPS	100 FPS
FacenessNet <sup>‡</sup>	80×80	-	20 FPS

- Model size
  - 1.6 MB (cf. FacenessNet > 1GB)

<sup>†</sup> Haoxiang Li, Zhe Lin, Xiaohui Shen, Jonathan Brandt, and Gang Hua. *A convolutional neural network cascade for face detection*. CVPR 2015.

<sup>‡</sup> Shuo Yang, Ping Luo, Chen Change Loy, and Xiaoou Tang. *From Facial Parts Responses to Face Detection: A Deep Learning Approach*. ICCV 2015.



# 1、SeetaFace检测器：其他

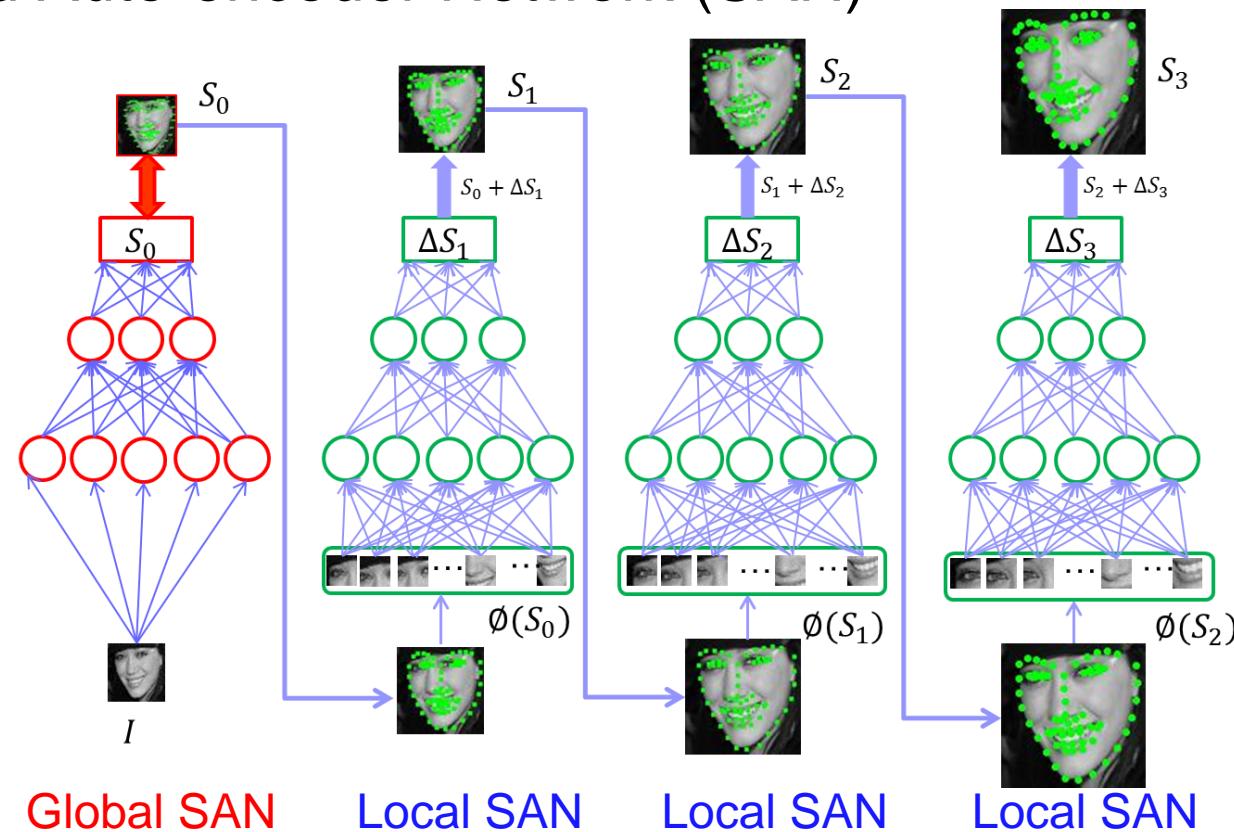
- C++, 不依赖于任何第三方库函数
- 界面简单易用
  - Define
    - `vipl::fd::FaceDetector my_detector(model_path)`
  - Configure (optional)
    - `my_detector.SetMinFaceSize(size)`
    - `my_detector.SetWindowStep(step_x, step_y)`
    - `my_detector.SetScoreThresh(thresh)`
  - Detect
    - `Std::vector<vipl::FaceInfo> faces =`  
`my_detector.Detect(img)`
  - Play with the detected faces ☺

## 2、SeetaFace Alignment



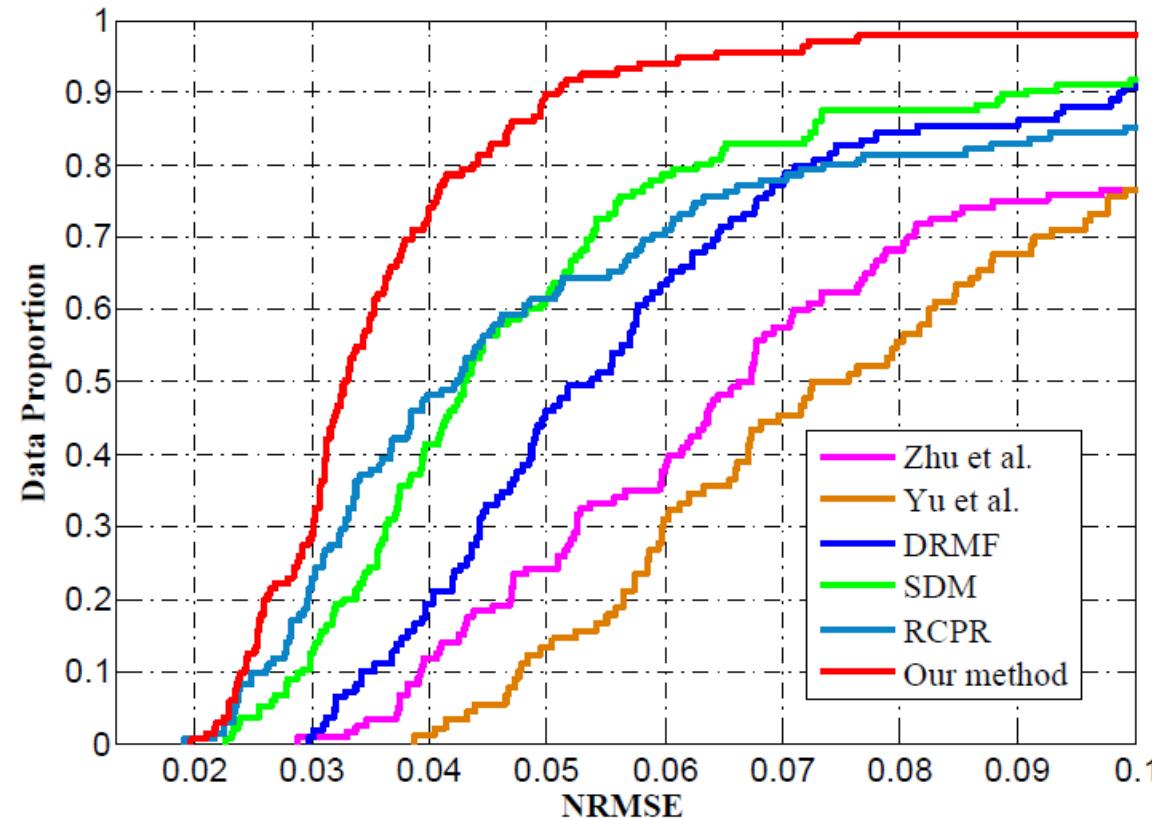
## 2、SeetaFace Alignment

- CFAN: Coarse-to-Fine Auto-encoder Network
  - Stacked Auto-encoder Network (SAN)



## 2、SeetaFace Alignment

- CFAN在IBUG数据集上的效果



Evaluated in terms of 68 landmarks



## 2、SeetaFace Alignment

### ■ 开源版本

- C++，没有第三方函数依赖
- 为加速，只用了2个阶段
- 训练集包括23,000余幅标注了5点的人脸图像

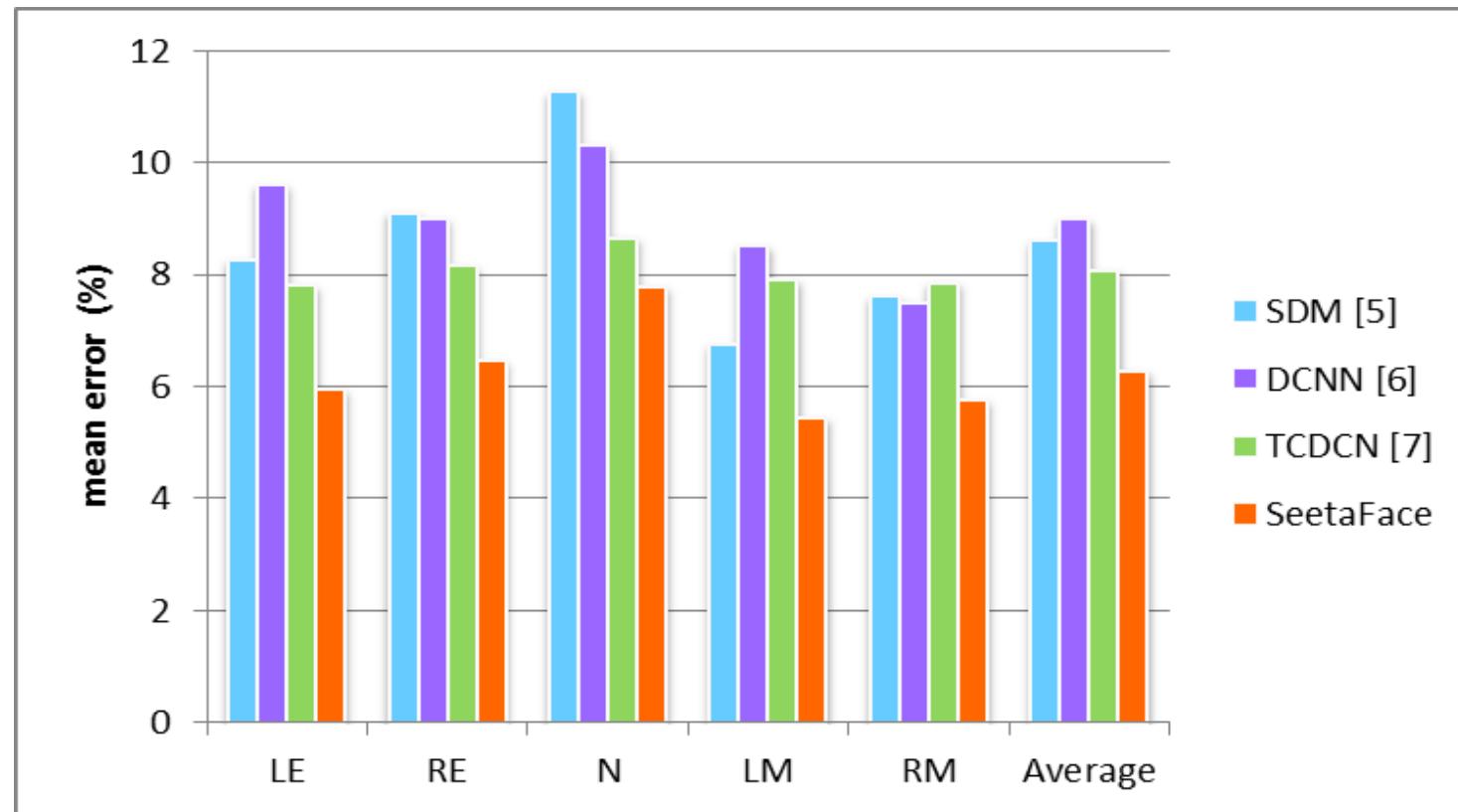
Time for detecting five facial landmarks

Method	Speed (ms)	Model Size (M)
Four Stages	13.8	3.97
Four Stages+ Faster SIFT	7.6	3.97
Two Stages+ Faster SIFT	3.9	1.98

\*Test on single 3.4GHz i7 CPU

## 2、SeetaFace Alignment

- SeetaFace Alignment在AFLW数据集上的定位误差及对比情况





## 2、SeetaFace Alignment

### ■ 如何使用：三个接口

- 初始化模型

```
PointDetectInitialize(char * model_path)
```

- 定位关键点

```
DetectLandmarks(VIPLImageData gray_im,  
VIPLFaceInfo face_info, VIPLPoint *points)
```

Face bbox

Image

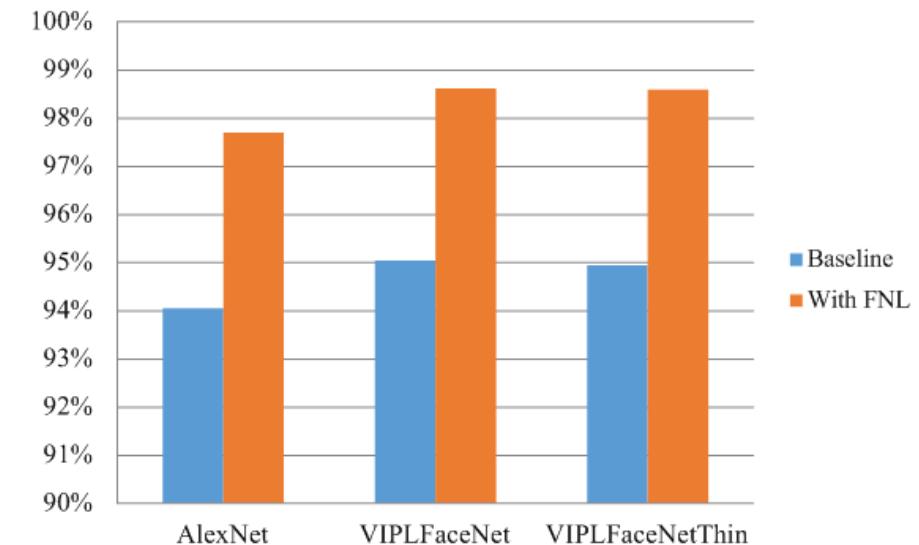
Landmarks

- 释放模型

```
PointDetectClose ()
```

### 3、SeetaFace Identification

- 对AlexNet的修改(→faster, →deeper)
  - kernel size of conv1: from 11x11 → 9x9
  - 5x5 conv kernel → two 3x3 conv kernels
    - Leading to a deeper network (5 conv → 7 conv)
  - #Nodes of the 2nd FC layer: 4096 → 2048
  - Exploit Fast Normalization Layer(**FNL**)
    - Faster and improved accuracy
  - Remove LRN layers
  - Remove group operation
    - Training on 1 GPU





### 3、SeetaFace Identification

#### ■ Structure comparison of AlexNet and SeetaFace

AlexNet	VIPLFaceNet
Conv1: 96x11x11, S:4, Pad:0	Conv1: 48x9x9, S:4, Pad:0
LRN	—
Pool1: 3x3,S:2	Pool1: 3x3,S:2
Conv2: 256x5x5, G:2, S:1, Pad:2	Conv2: 128x3x3,S:1, Pad:1
LRN	—
—	Conv3: 128x3x3,S:1, Pad:1
Pool2: 3x3,S:2	Pool2: 3x3,S:2
Conv3: 384x3x3, S:1, Pad:1	Conv4: 256x3x3,S:1, Pad:1
Conv4: 384x3x3, G:2, S:1, Pad:1	Conv5: 192x3x3,S:1, Pad:1
—	Conv6: 192x3x3,S:1, Pad:1
Conv5: 256x3x3, G:2, S:1, Pad:1	Conv7: 128x3x3,S:1, Pad:1
Pool3: 3x3,S:2	Pool3: 3x3,S:2
FC1, 4,096	FC1, 4,096
Dropout1: dropout_ratio:0.5	Dropout1: dropout_ratio:0.5
FC2, 4,096	FC2, 2,048
Dropout2: dropout_ratio:0.5	Dropout2: dropout_ratio:0.5
FC3, Number of training classes	FC3, Number of training classes



### 3、SeetaFace Identification

- Results on LFW test set/protocol of View 2
  - Trained on WebFace (500,000 images of 10,000 subjects)

Method	Accuracy	# of Network	# of Training Images
High-dim LBP[27]	95.17%	–	–
Fisher Vector Face[118]	93.10%	–	–
DeepFace[133]	97.35%	3	4M
DeepID[126]	97.45%	25	200K
DeepID2[126]	99.15%	25	200K
Gaussian Face[97]	98.52%	–	–
DeepID2+[127]	99.47%	25	290K
DeepID2+(Single)[127]	98.70%	1	290K
DeepID3 [123]	99.53%	1	290K
WSTFusion[134]	98.37%	–	10M
VGGFace[106]	98.95%	1	2.6M
FaceNet[115]	99.63%	1	200M
AlexNet + FNL [73]	97.70%	1	500K
<b>VIPLFaceNetFull + FNL</b>	<b>98.62%</b>	1	500K
<b>VIPLFaceNet + FNL</b>	<b>98.60%</b>	1	500K



### 3、SeetaFace Identification

- 开源版本训练集
  - 1.6万人
  - 140万幅图像
- 特征提取
  - FC2的2048个节点输出作为特征
- 特征比对
  - 向量夹角余弦Cosine ( $\Theta$ )



### 3、SeetaFace Identification

- 精度
  - 全自动系统(采用SeetaFace检测和对齐)在LFW上97.1%
- 时间性能
  - 训练时间为AlexNet的20%
  - 测试时间为AlexNet的60%
    - **120ms** (单颗3.4GHz i7 CPU上)



# 3、SeetaFace Identification

## ■ 接口简单易用

- Initialize face recognition model

```
VIPLFaceRecognizer FR;  
FR.LoadModel(const char* model_path)
```

- Extract face feature from an input image

```
FR.ExtractFeatureWithCrop(const VIPLImageData  
gray_im,VIPLPoint * points, FaceFeatures feats)
```

Image

Landmarks

Identity feature

- Compute similarity of two face features

```
FR.CalcSimilarity(FaceFeatures const fc1,  
FaceFeatures const fc2);
```



# SeetaFace引擎总结

- 下载源码：<https://github.com/seetaface/>
- 全开源的、C++版、全自动人脸识别引擎
  - BSD-2 license（全免费使用）
  - 无第三方函数库依赖
- 精度和速度可满足很多应用需求 (**单颗3.4GHz i7 CPU**)
  - SeetaFace Detector: **20ms**/VGA图像（最小人脸80x80）
  - SeetaFace Alignment: **5ms**/face
  - SeetaFace Identification: **120ms**/face
  - **Totally ~200ms/image (on single i7 CPU core )**
- Performance
  - SeetaFace Detector: DR=**84.4%**@100FPs (VGA图像实时)
  - LFW上**~98.6%** (半自动系统)
  - LFW上**~97.1%** (全自动系统)



# 提纲

- 背景与概况
- 人脸识别技术的深度化
  - 人脸检测技术
  - 面部特征点定位技术
  - 判别特征学习与匹配方法
  - SeetaFace人脸识别引擎
- 总结与讨论



# 挑战：人脸识别应用层面

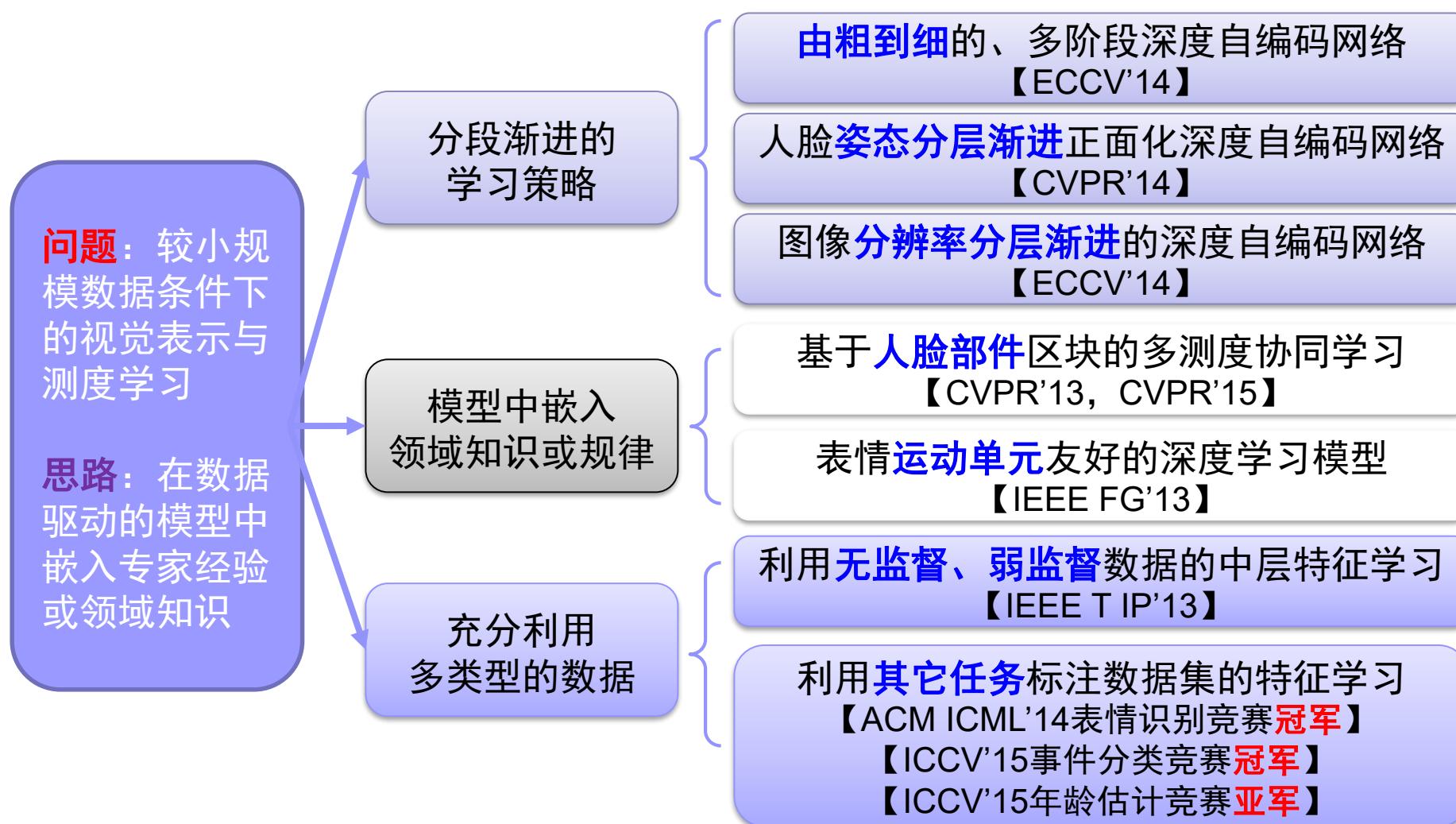
- 姿态：超过45度角度情况下（俯视、仰视和侧面）
- 光照
  - 白天可见光，晚上(近)红外
  - 室外严重偏光或背光条件
- 分辨率(远距离): 40x40像素以下
- 遮挡（墨镜、口罩）
- 大尺度整容
- 开放监控场景下的人脸布控
  - 千万分之一的虚警率



# 挑战：深度学习方面

- 理论层面：优化方法，收敛性，可解释性，可学习性
- 应用层面：大数据可得性，异构计算，速度与功耗
- 模型相关
  - 从参数学习到结构学习？
  - 新结构：层内连接，跨层连接，多任务子网络互联
  - 如何在预测阶段进行反馈与网络调制？
  - 归纳学习vs.演绎推理：如何赋予机器演绎推理能力？
- 数据相关
  - 大数据是解决CV问题的终极道路吗？
  - 举一反三：大数据是否学习之必需？
  - 无师自通：如何获取无监督学习能力？
  - 基于脏、乱、差、小数据的深度学习

# 小规模数据条件下的深度学习





# 近期的相关工作

中科院计算所

- Stacked Progressive Auto-Encoder (SPAЕ) for Face Recognition Across Poses [[CVPR2014](#)]
- Coarse-to-Fine Auto-encoder Networks (CFAN) for Real-time Face Alignment [[ECCV2014](#)]
- Deep Network Cascade for Image Super-resolution [[ECCV2014](#)]
- Shape Driven Kernel Adaptation in Convolutional Neural Network for Robust Facial Traits Recognition [[CVPR2015](#)]
- Bi-shift Auto-Encoder for Domain Adaptation [[ICCV2015](#)]
- Leveraging Datasets with Varying Annotations for Face Alignment via Deep Regression Network [[ICCV2015](#)]
- A Unified Multiplicative Framework for Attribute Learning [[ICCV2015](#)]
- Multi-view Discriminant Deep Network [[CVPR2016](#)]
- Occlusion-free Face Alignment: Deep Regression Networks Coupled with De-corrupt AutoEncoders [[CVPR2016](#)]
- Deep Supervised Hashing for Fast Image Retrieval [[CVPR2016](#)]

# 阿里巴巴大规模图像搜索竞赛



## ***Problem Measurement***

- *mean Average Precision (mAP) @ 20*

## ***Goal***

- *Recalling as more same styles as possible in top 20 results;*
  - *Ranking relevant samples as front as possible;*
- *Retrieving as soon as possible (compact data representation).*

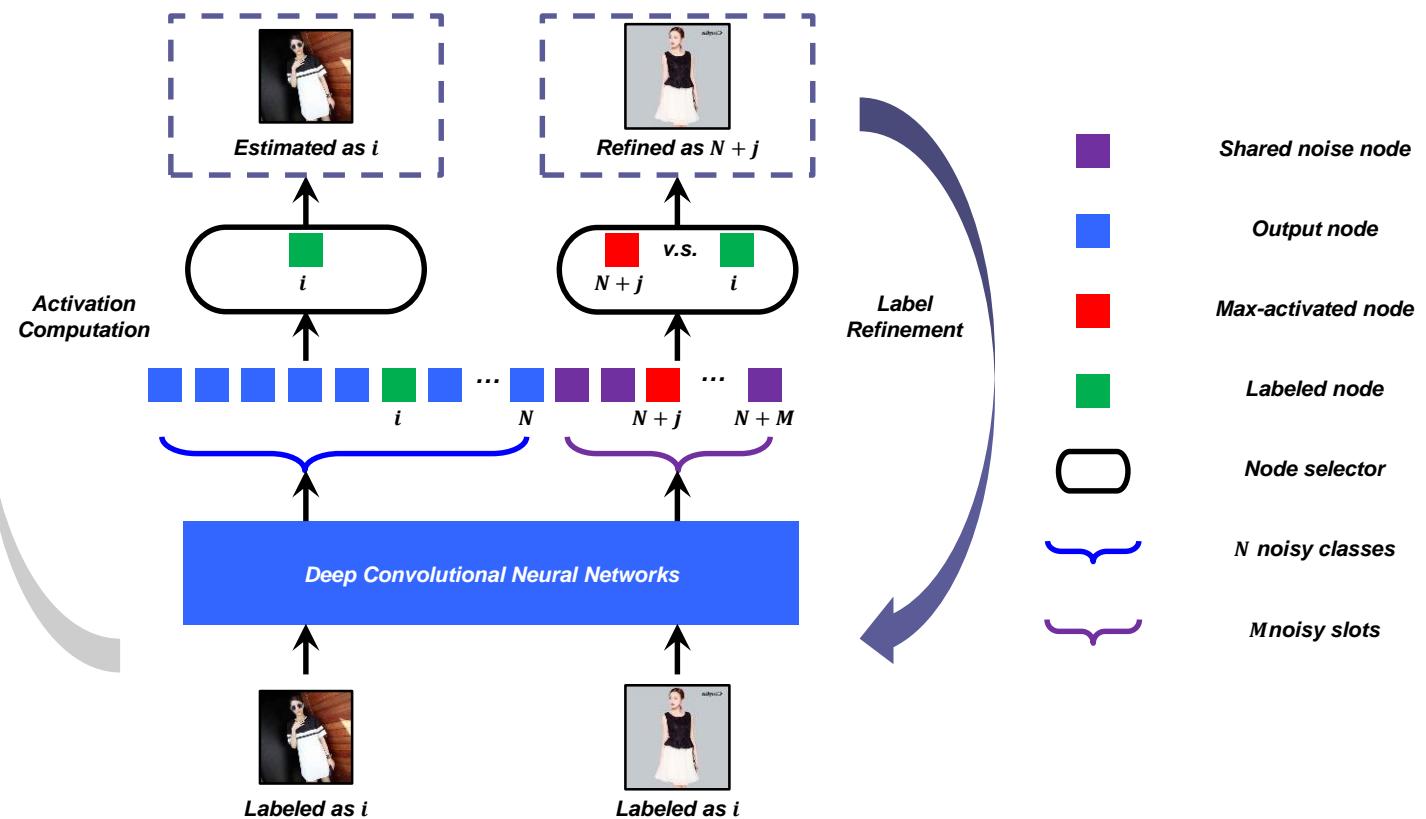
# 我们的方法——自动同款挖掘

- 问题：大规模无监督图像数据库
- 解决方法
  - 自动聚类+基于自步学习的“同款商品垂钓”



# 我们的方法——自纠错DCNN

- 自动聚类的结果会有大量错误，怎么办？
  - 模拟人的自纠错学习 “error-correcting” learning





# 我们的方法——自纠错DCNN

中科院计算所

## ■ *Evaluation of proposed method*

Method	Accuracy with different noise ratio (%)				
	0%	20%	40%	60%	80%
<i>Benchmark: MNIST</i>					
LeNet	99.3	96.8	84.2	56.0	20.6
LeNet + Proposed Method	99.3	99.2	<b>98.9</b>	<b>98.6</b>	<b>98.4</b>
<i>Benchmark: CIFAR-10</i>					
LeNet	64.5	47.9	34.5	23.0	15.5
LeNet + Proposed Method	63.6	<b>59.9</b>	<b>52.9</b>	<b>49.9</b>	<b>34.6</b>





# 总结 (Take Home Message)

- 人脸的检测与识别不再具有特殊性！
- 深度学习极大推进了人脸识别能力！
- 甚至在多数场景上超过人眼能力！
- **单从性能来看，工业界走在了学术界前面！**
- 万人以上规模的黑名单视频监控场景尚不成熟！
- SeetaFace为大家提供了一个不错的baseline😊
- 深度学习在人脸识别领域dominate并不表明其他问题上也会dominate！

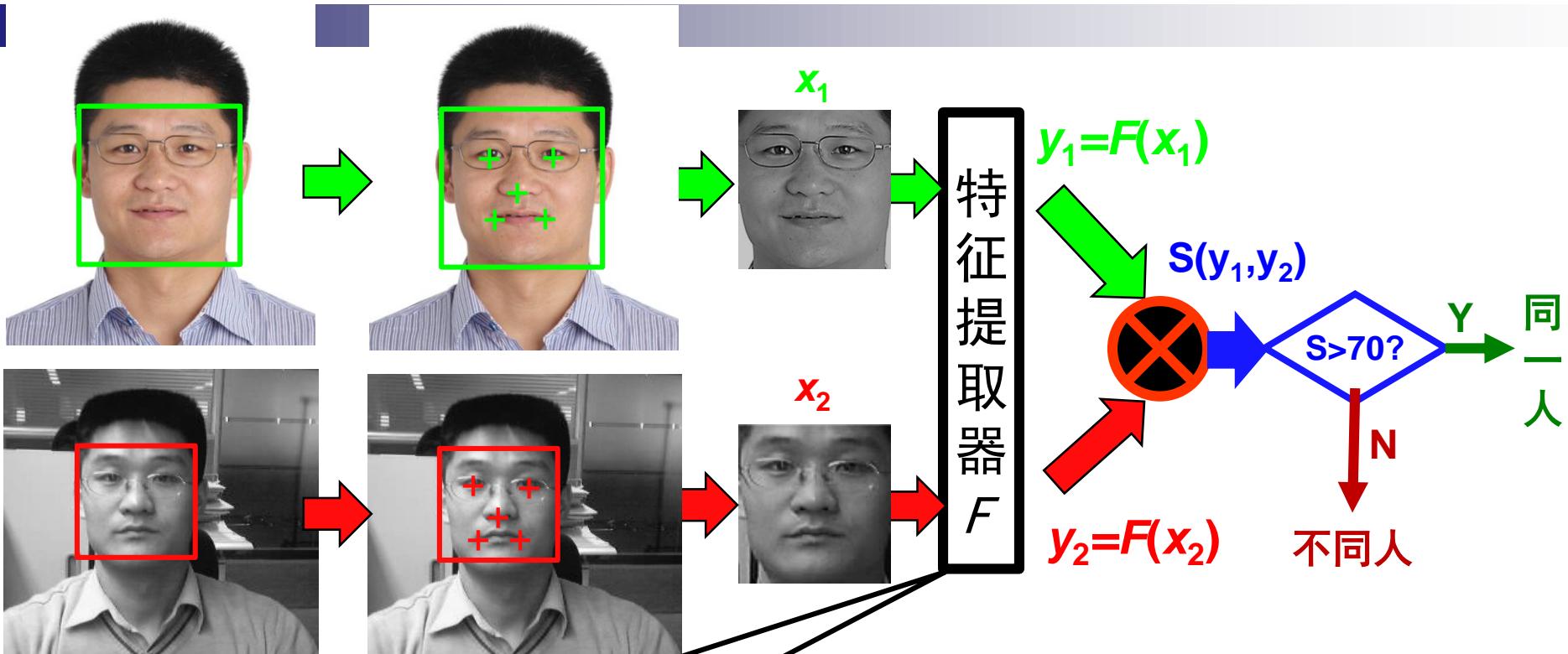
谢谢！



中国科学院计算技术研究所  
Institute of Computing Technology, Chinese Academy of Sciences

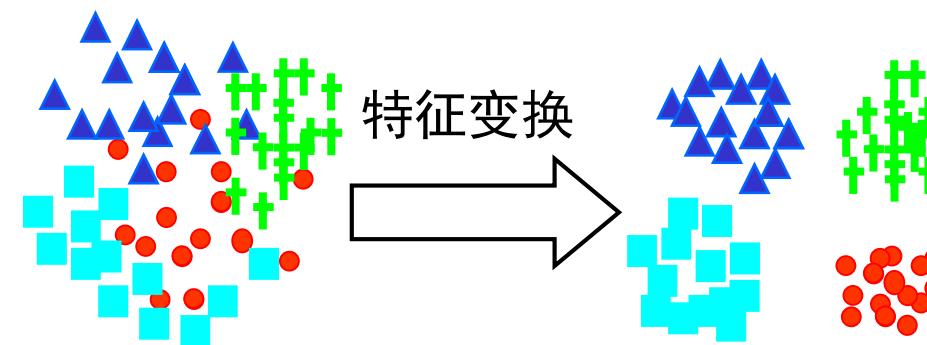
# 全自动人脸识别系统流程(DL之前)

中科院计算所



两步法：

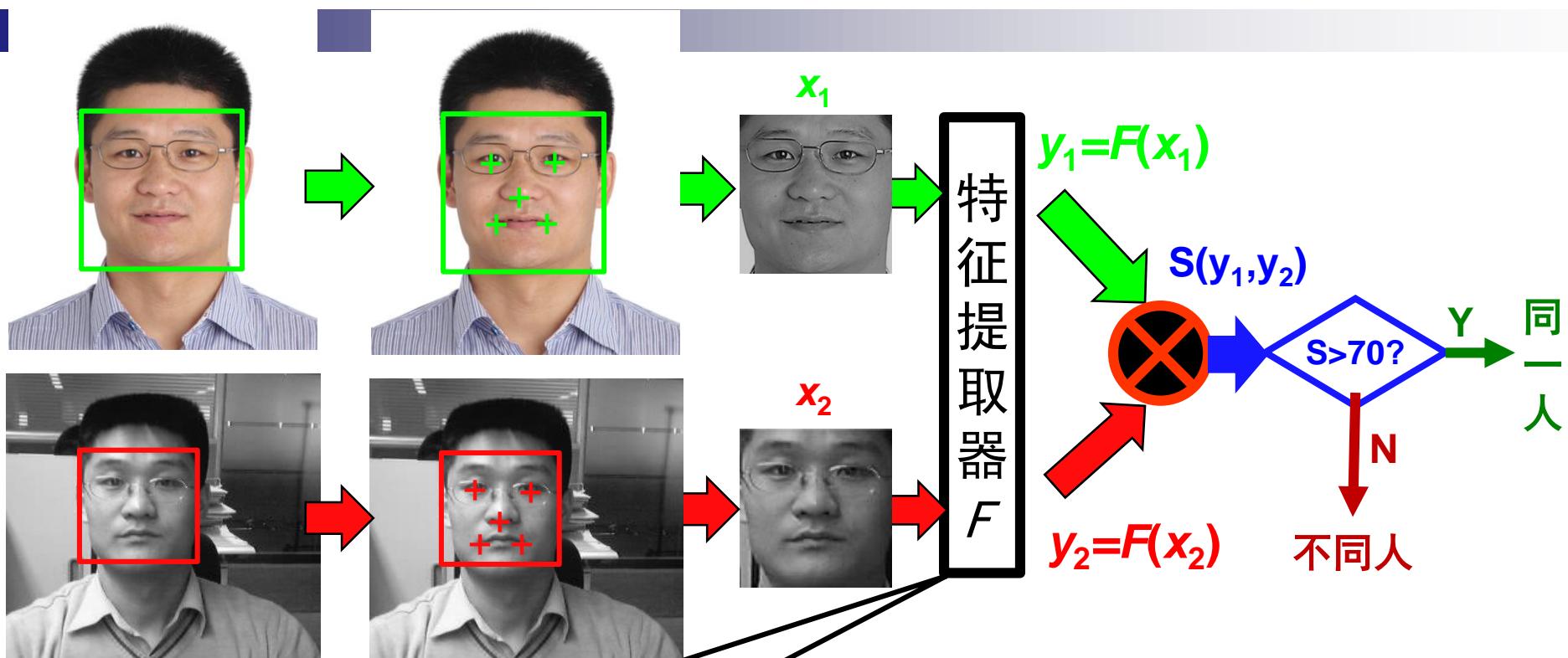
1. 人工(卷积)特征： Gabor, LBP, SIFT, ...
2. 特征变换： PCA, LDA, KDA, ML...



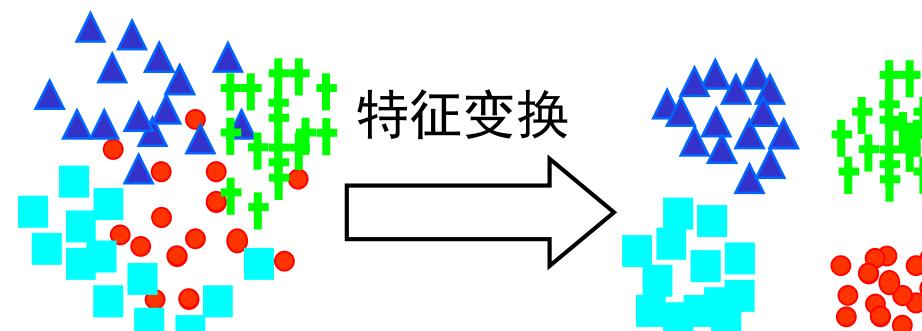


# 全自动人脸识别系统流程(DL)

中科院计算所



从图像开始直接学习  
特征!



输入图像

低维判别特征空间