深度学习之能与不能

山世光
中科院计算所

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

基于机器学习的视觉信息处理

中科院计算所

- 任务

$x \longrightarrow F(x) \longrightarrow y$

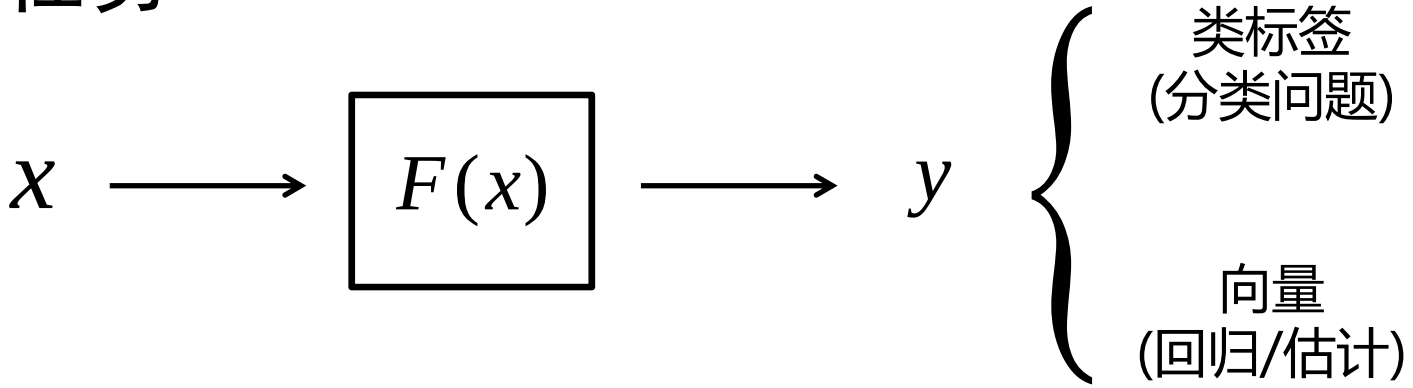
y：类标签（分类问题）；向量（回归/估计）

通常是高度非线性的！

基于机器学习的视觉信息处理

中科院计算所

- 任务

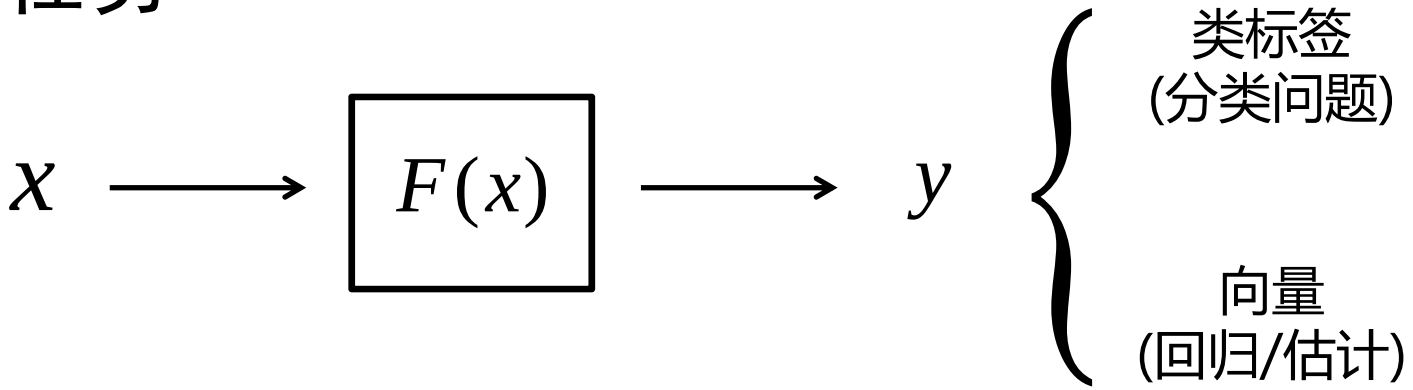


- 过去：人工设计+部分学习F
  - 早期神经网络(一棵老树)
  - Kernel方法：隐式非线性映射，可学性不佳
  - 流形学习：试图学习显式映射，可扩展性差
  - 分治思想指导下的“分段线性逼近”

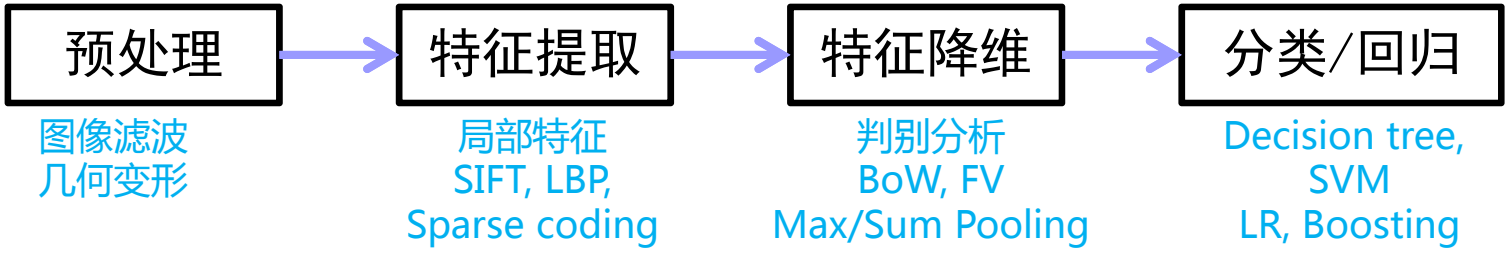
基于机器学习的视觉信息处理

中科院计算所

- 任务



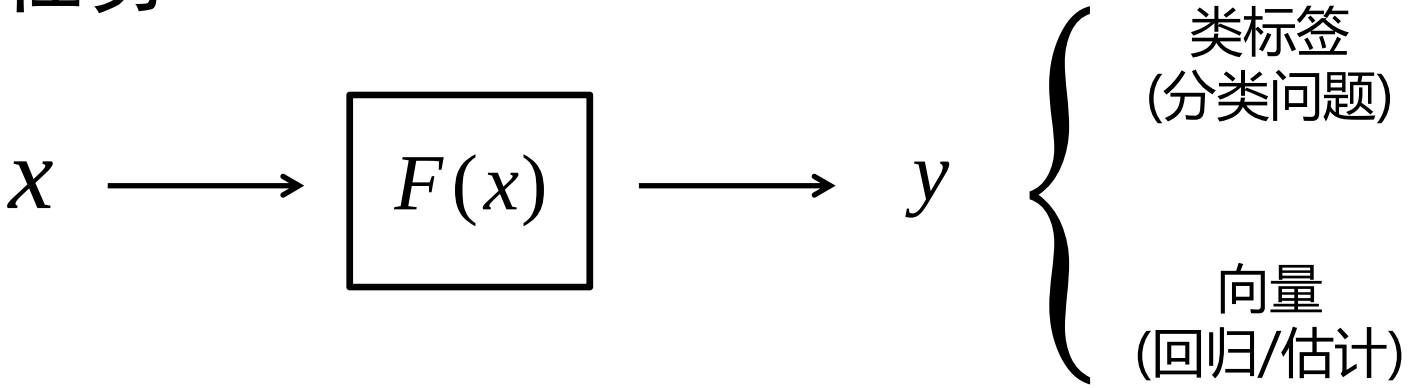
- 过去：人工设计+部分学习F
  - 分段处理(分段线性逼近)
  - 领域知识指导，部分进行学习



基于机器学习的视觉信息处理

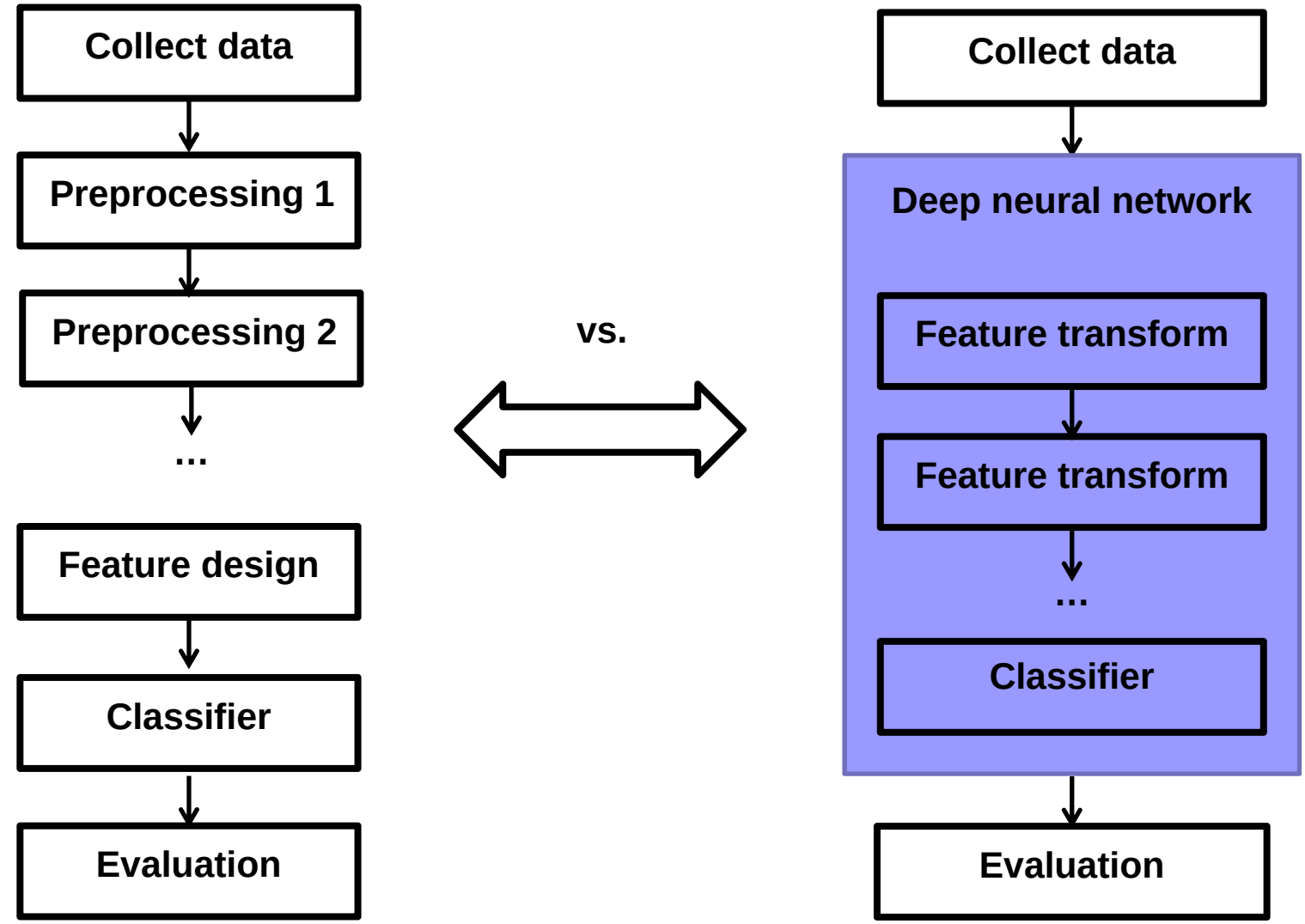
中科院计算所

- 任务



- 深度学习：学习显式的非线性映射
  - 分层非线性→逐层语义抽象
  - End-to-End(E2E) Learning

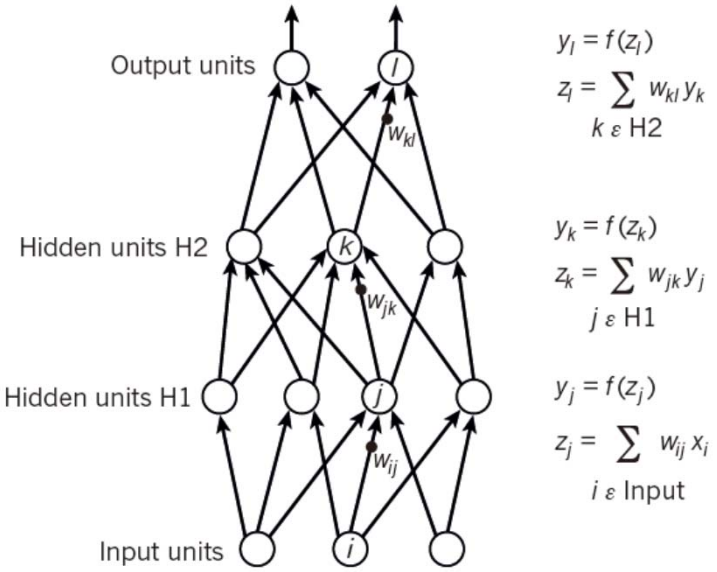
E2E Learning System

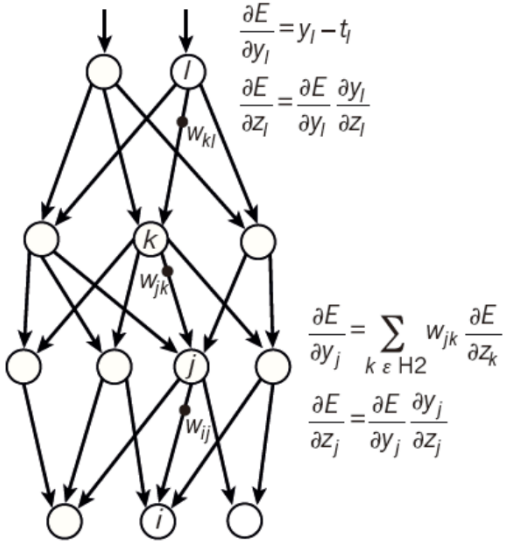


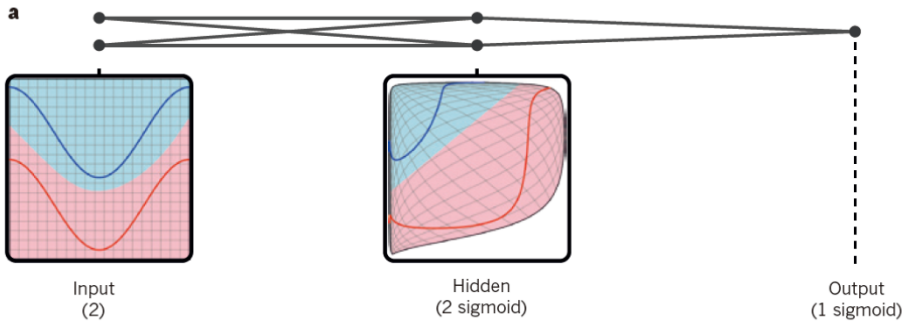
Credit to Dr. Xiaogang Wang

神经网络的发展

- 1980s
  - 多层网络结构
  - Error梯度反向传播
  - 非线性激活函数



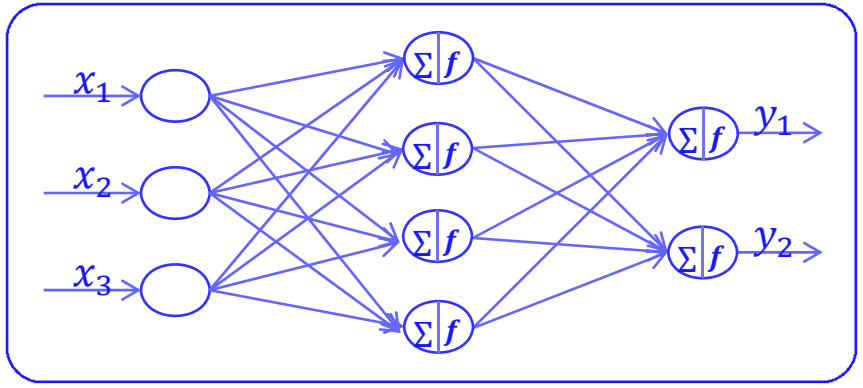




Rumelhart D E, **Hinton** G E, Williams R J. Learning representations by back-propagating errors. Cognitive modeling, **1988**.

神经网络的发展

- 1980s
  - 多层网络结构
  - Error梯度反向传播
  - 非线性激活函数
- 不成功
  - 优化困难：梯度消失
  - 训练数据少
  - 计算资源不足



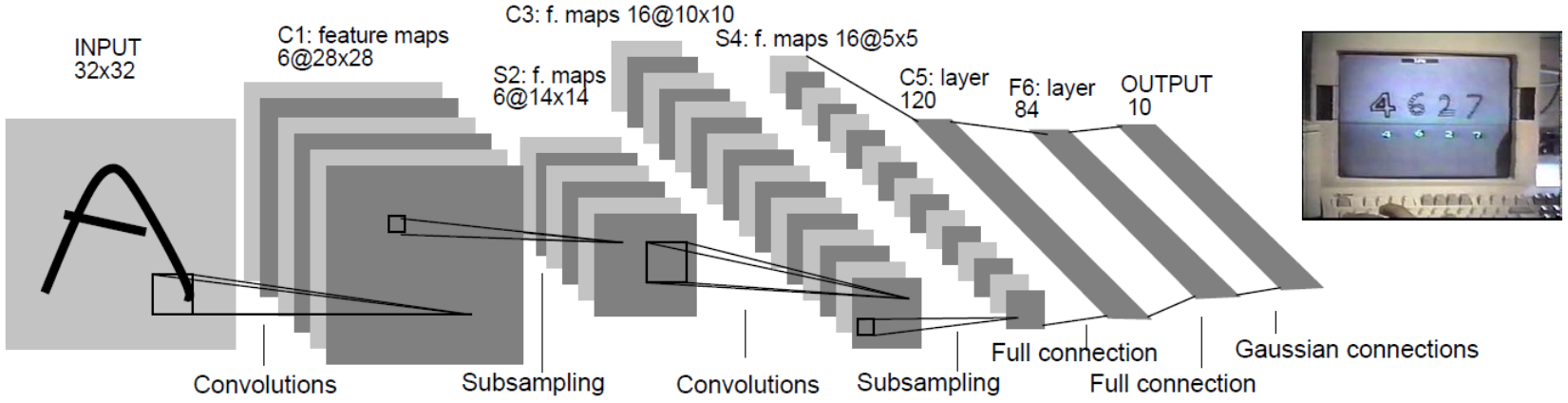
- Decision tree
- SVM
- Boosting
- Sparse coding
- Graph models

Rumelhart D E, **Hinton** G E, Williams R J. Learning representations by back-propagating errors. Cognitive modeling, **1988**.

Credit to Dr. Xiaogang Wang (modified)

神经网络的发展

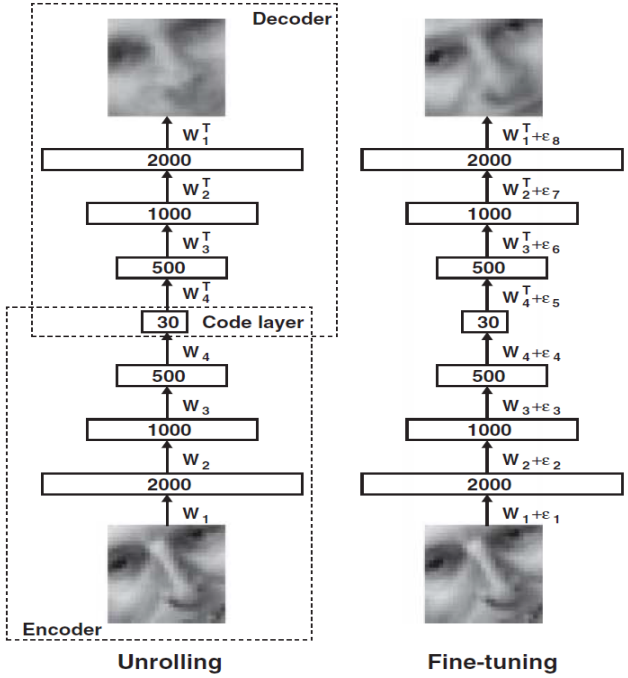
- 1990s：卷积神经网络CNN
  - 局部卷积操作
  - Pooling操作
  - 非线性激活函数



LeCun Y, Bottou L, **Bengio** Y, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998.

深度学习的崛起

- 2000s
  - 无监督学习
  - 分层预训练
  - 各种网络结构
    - RBM, DBN, DAE
  - 得名“深度”学习



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

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

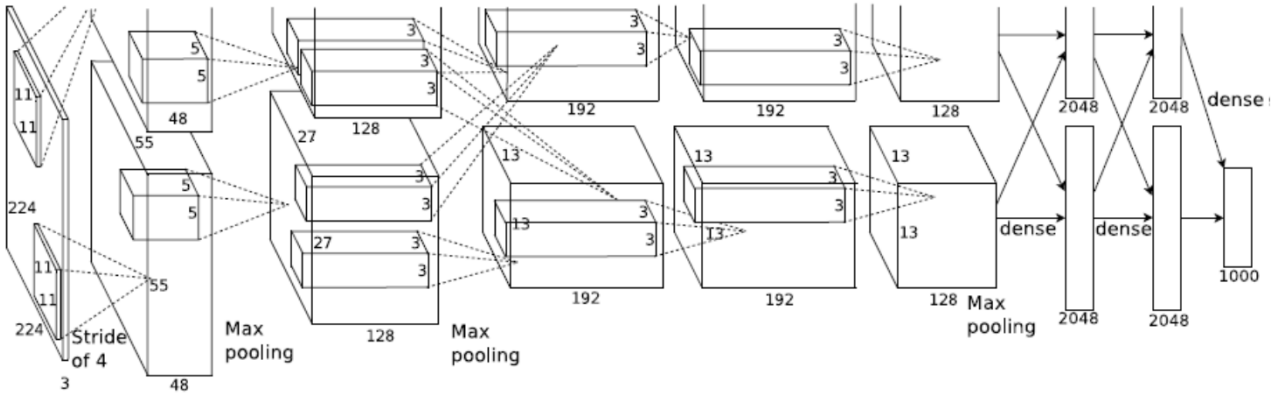
Decoding: $\tilde{\mathbf{h}}_1 = \sigma(\mathbf{W'}_2\mathbf{h}_2+b_3)$

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

Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks. Science, **2006**.

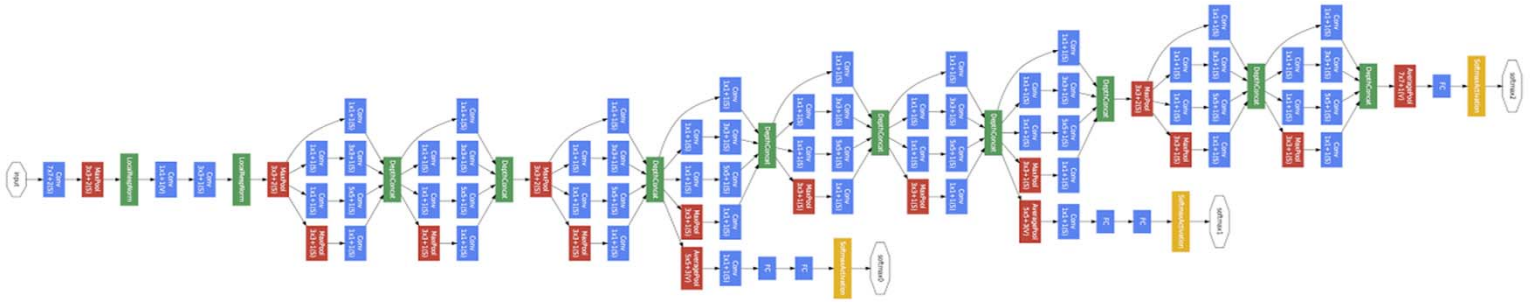
深度学习的崛起

- 2012
  - Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton. ImageNet classification with deep convolutional neural. NIPS12
  - CNN：老树发新芽
    - 大数据训练：ImageNet
    - 非线性部分：ReLU
    - 防止过拟合：数据增广，DropOut
    - 其他：双GPU实现，LRN



深度学习的崛起

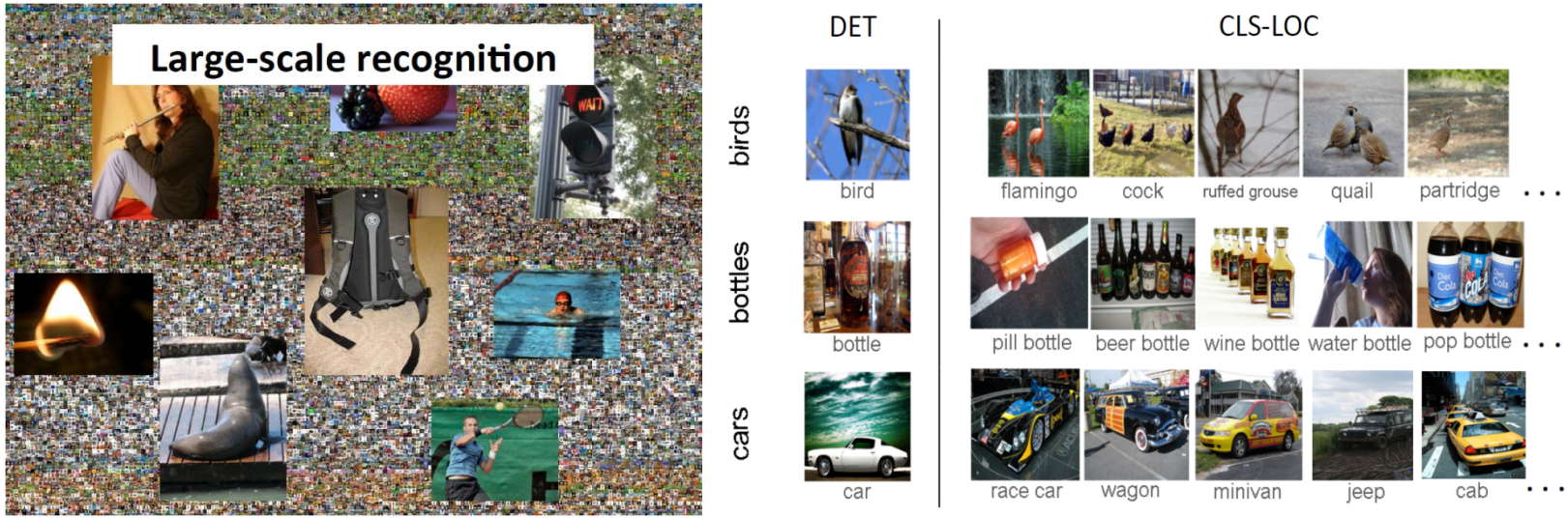
- 2012—至今
  - 非线性函数：ReLU
  - 防止过拟合的策略：Dropout
  - 更深的网络结构[2]
  - 优化策略的持续改进



[2] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. CVPR 2015.

人工神经网络发展概况

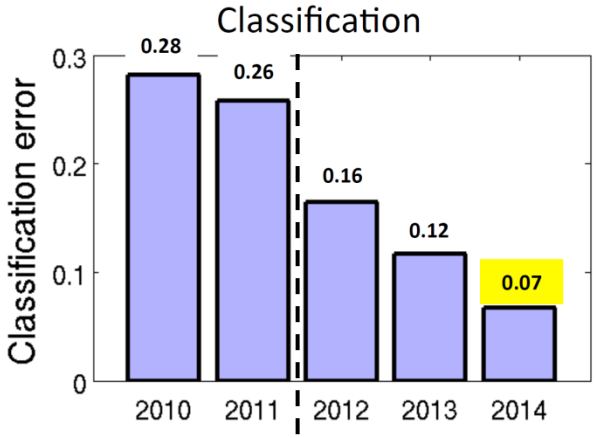
- ImageNet竞赛 (2014)
  - 物体分类定位任务：1000类，1,431,167幅图像



Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge. arXiv preprint, 2014.

人工神经网络发展概况

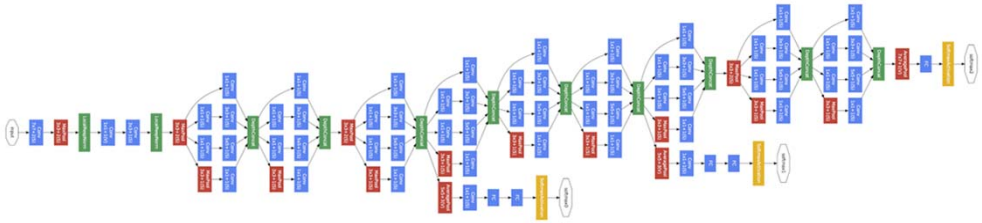
- 分类任务：1000类，1,431,167幅图像



传统方法
Fisher vector

深度方法
AlexNet
ZF-Net
GoogLeNet

2012&2013年，8层深度卷积网络



2014年，24层深度卷积网络

深度学习促进人脸识别

中科院计算所

- 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新闻图片
    - ~5749明星,
    - 1680人多于2张图
  - 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.

深度学习促进人脸识别

- 过去两年正确率的提升
  - 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]

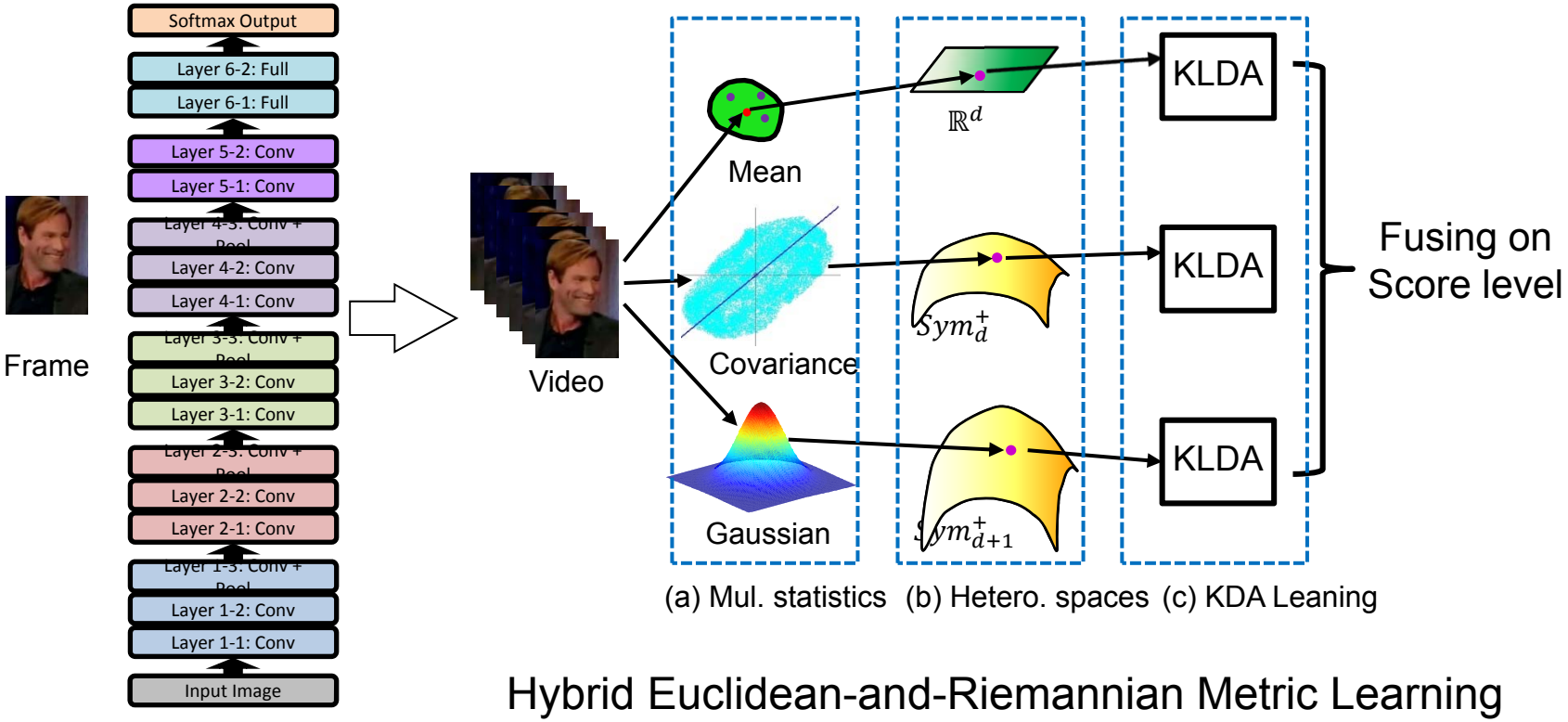
过去2年错误率从5%下降到1%
(错300对→错60对)

FG 2015 Video FR Challenge

- Task: video-to-video face verification
  - 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

Our Method

- DCNN (single frame feature)
- HERML(set model and classification)



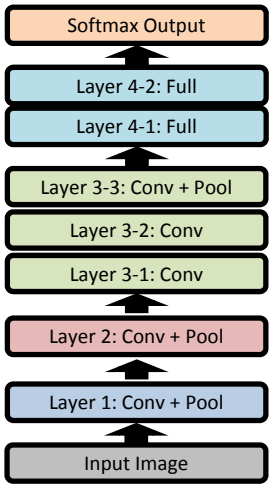
DCNN [Jia'13]

Hybrid Euclidean-and-Riemannian Metric Learning (HERML) [Huang, Wang, Shan, Chen, ACCV'14]

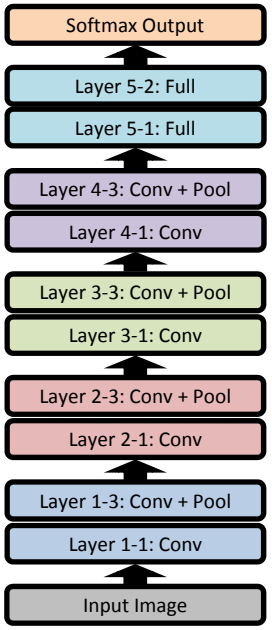
Evaluation Results

- The deeper the better

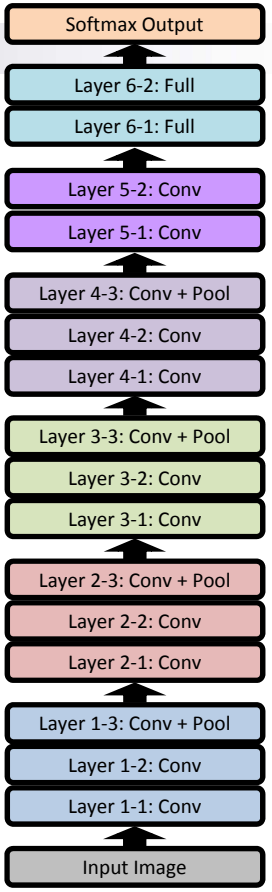
DCNN for single frame



control：41.40%，
handheld：41.62%



control: 47.41%
handheld: 48.02%



control: 54.76%
handheld: 56.20%

DCNN + HERML (set models)

control：46.61%，
handheld：46.23%

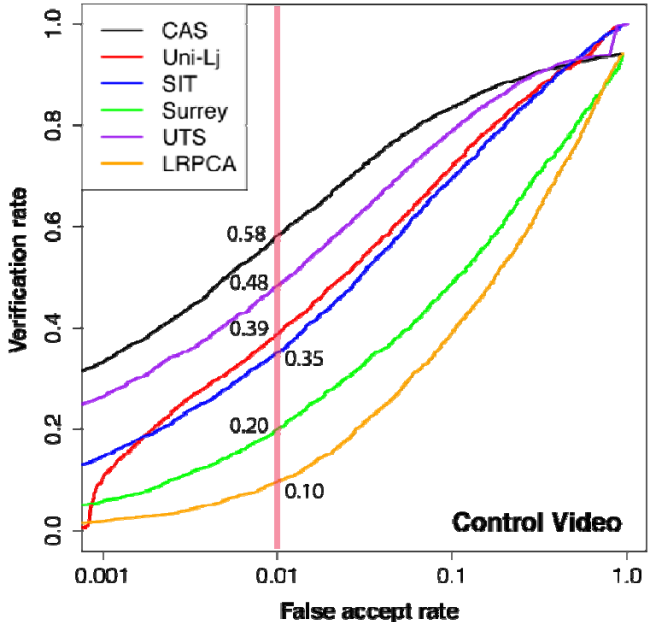
control： 56.20%，
handheld：54.41%

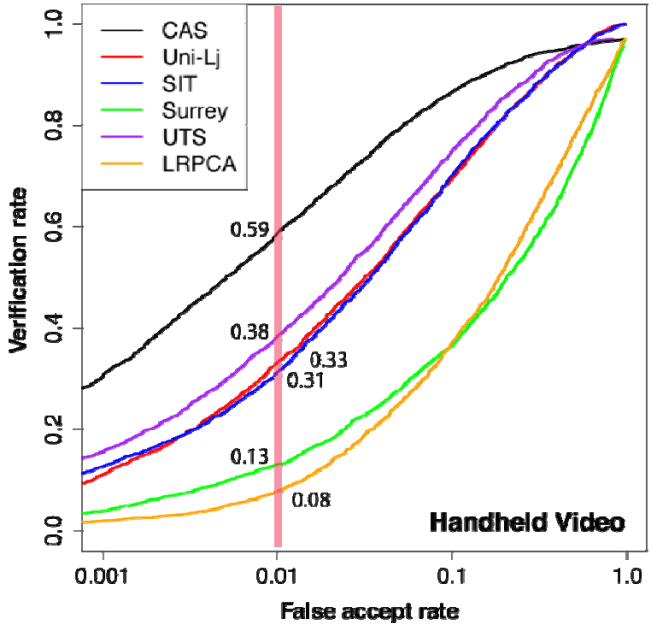
control：58.63%，
handheld：59.14%

Primary Results

- Image features
  - HOG < Dense SIFT << DCNN

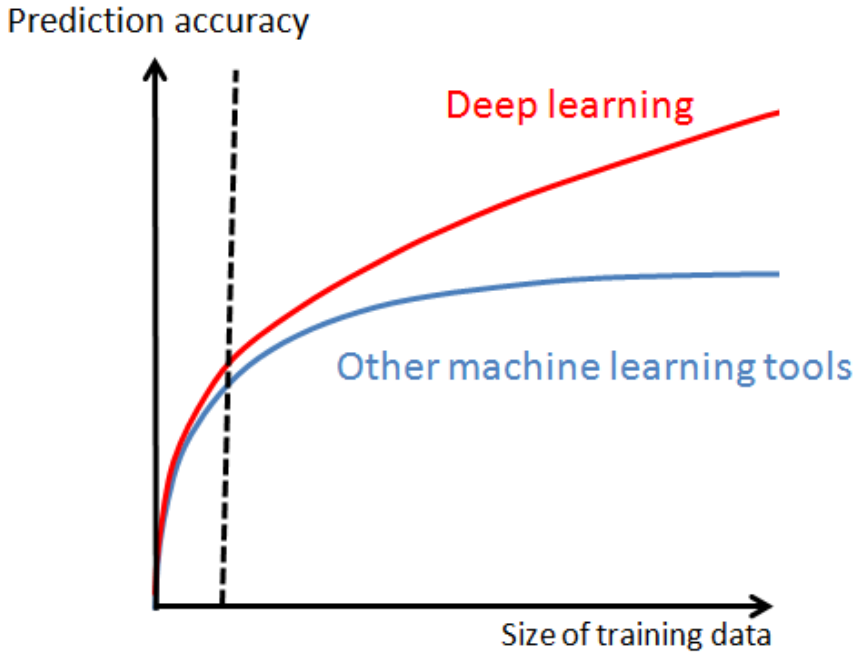
Method	HOG		Dense SIFT		DCNN	
	Control	Handheld	Control	Handheld	Control	Handheld
HERML	25.26	19.28	**33.82**	**28.93**	**58.63**	**59.14**





深度学习成功的条件

- “大”数据支撑
  - 容易Under-fitting，需要提升模型复杂度
- “大”容量模型
  - “深度”层级连接网络
  - 更高的模型复杂度
- “大”机器支持
  - 多核CPU，Cluster
  - GPU！
- 算法进步
  - ReLU，数据增广，DropOut，SGD，
- “大”社区协同
  - 开放源码，开放数据，开放模型，开放文章



DL是类脑信息处理方法吗？

- DL受到脑信息处理方式启发
  - 神经元连接方式
  - 分层逐级抽象
    - 初级视觉神经元的“类小波编码”
  - 分布式的记忆
- 并不“类脑”
  - 脑的“计算机理”尚不清晰
  - 脑皮层连接更多样、更复杂
  - 学习过程未必需要大量数据

DL有理论吗？

- DL理论匮乏
  - 收敛性，错误率的bound
  - 模型复杂度理论缺失
  - 不怕局部极值?
- 但不完全是black box
  - 显式的、分层非线性
  - 层级可视化提供了很多线索
    - 逐层抽象
  - 优于传统“分段”方法
  - 比Kernel更“显式”

DL不能做什么？

- 用做Feature Learning最成功
  - 学到的特征具有良好的通用性
  - 传统分类器或回归似乎还可用
- 非常倚重大数据
  - 小数据深度学习不可靠
- 在简单问题上未必需要深度学习
  - 人脸识别的例子
- 目前的DL不学习“自身结构”
  - 调试经验很重要

DL带来观念的变革

- 小数据→控制模型复杂度，避免过拟合
- 大数据→提高模型复杂度，担心欠拟合
- 人工领域知识驱动→数据驱动的学习思想
  - “大数据+简单模型”是错误的！
  - 人工特征→特征学习
  - 维数灾难(降维)→高维有益(升维)
- 分步、分治→ E2E的全过程学习
  - 分步无法“全过程”最优
  - 协同学习(joint learning)思想
- 软硬件更优的协同（优化和加速）

数据驱动的学习不需要领域知识？

- 大数据驱动确实减少了对领域知识的依赖
- 但是，CNN在CV领域的成功，本身就说明了领域知识的重要性
  - 卷积操作
  - MaxPooling操作
  - 权值共享减少参数
  - 非线性的来源
  - 各种超参数的调试
- 小数据条件下，领域知识尤其重要

DL可能的未来工作

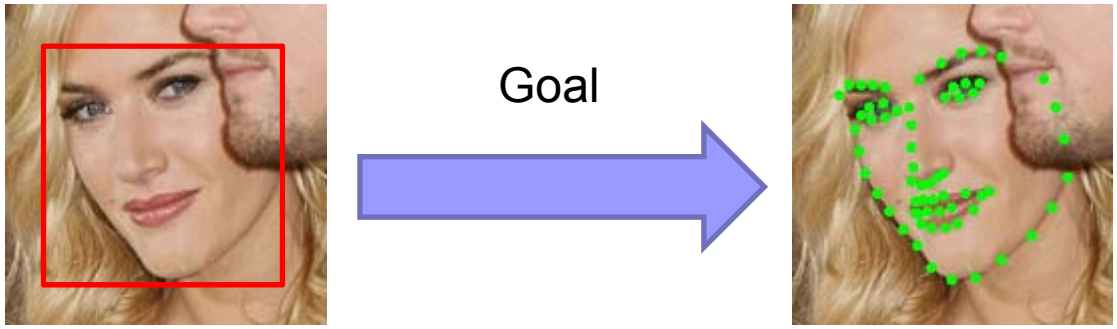
- 建立DL学习理论
- 网络结构设计、学习和优化方法
  - “非线性”的更多来源
  - 新的优化和训练算法
  - 网络结构本身的学习，借鉴人的学习机理
- 领域问题适应
  - 领域知识的嵌入
  - 深度模型的迁移与适应
  - 面向时序信号分析的DL模型
- 数据
  - “小数据”条件下的DL

小数据如何训练？

- 目前DL的胜利更多是大数据的胜利！
- 小数据没有机会了吗？
  - 迁移（预训练）
  - 利用Domain Knowledge

小数据条件下应用DL的例子

- 面部特征定位
  - Predict facial landmarks from detected face



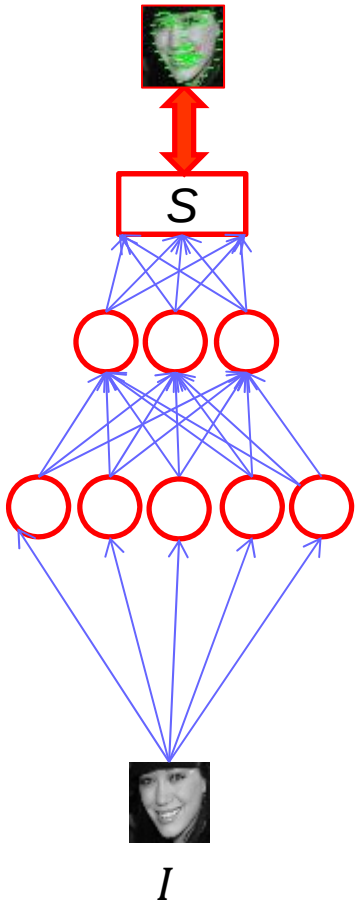
Detected face region $I(u,v)$

Facial landmarks $S=(x_1,y_1, x_2, y_2, \ldots, x_L, y_L)$

$$S = H(I),\ I \in R^{w*h}, S \in R^{2L},$$

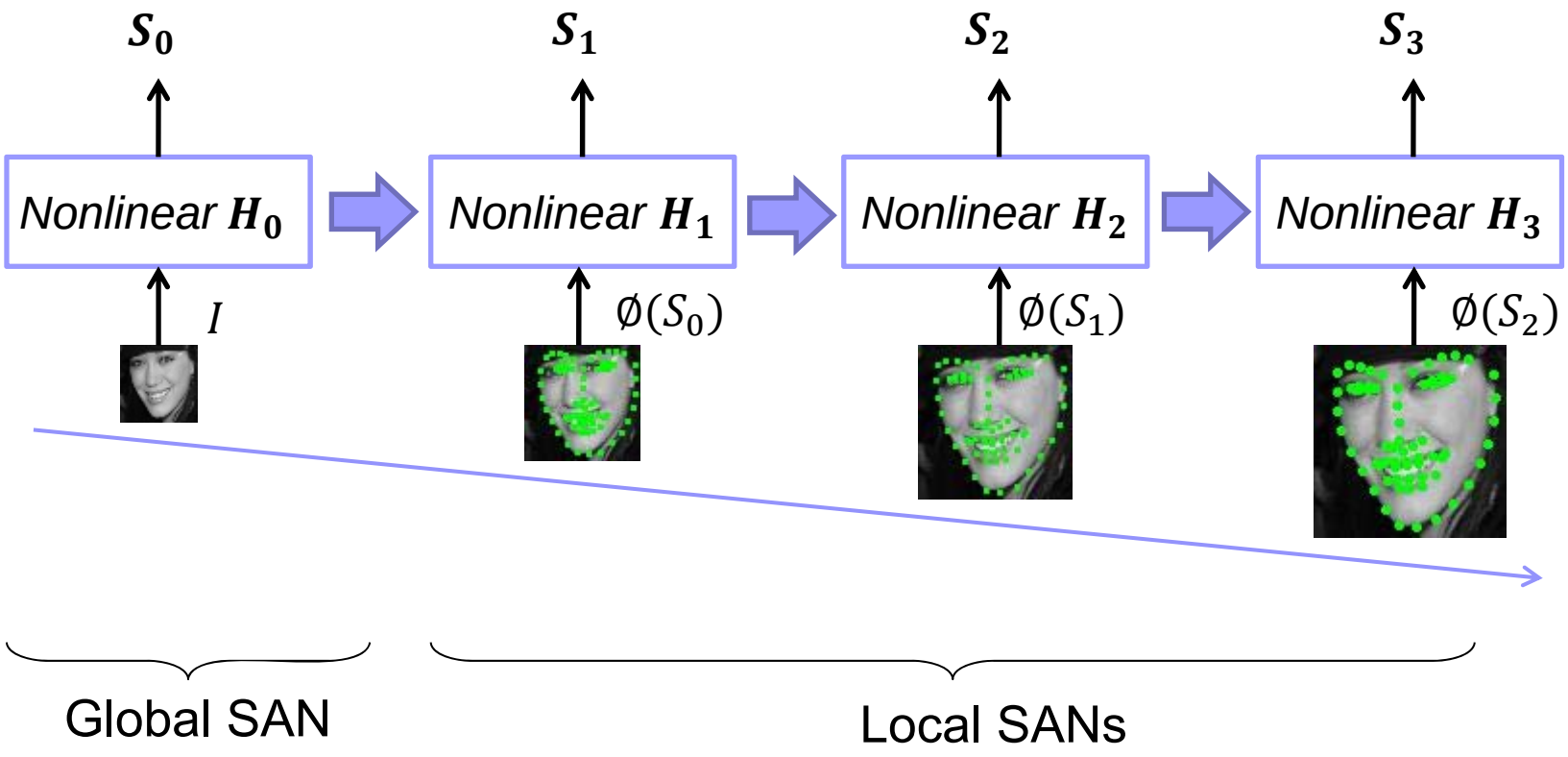
小数据条件下应用DL的例子

- 直接用DL？
  - OK，但不理想
    - 小数据--少于3000幅训练图像
    - 非常容易过拟合
- 我们的思路 – exploiting priors
  - 分段非线性
  - 配合由粗到细的策略
  - 避免卷积部分的学习
    - 采用人工设计的特征SIFT



小数据条件下应用DL的例子

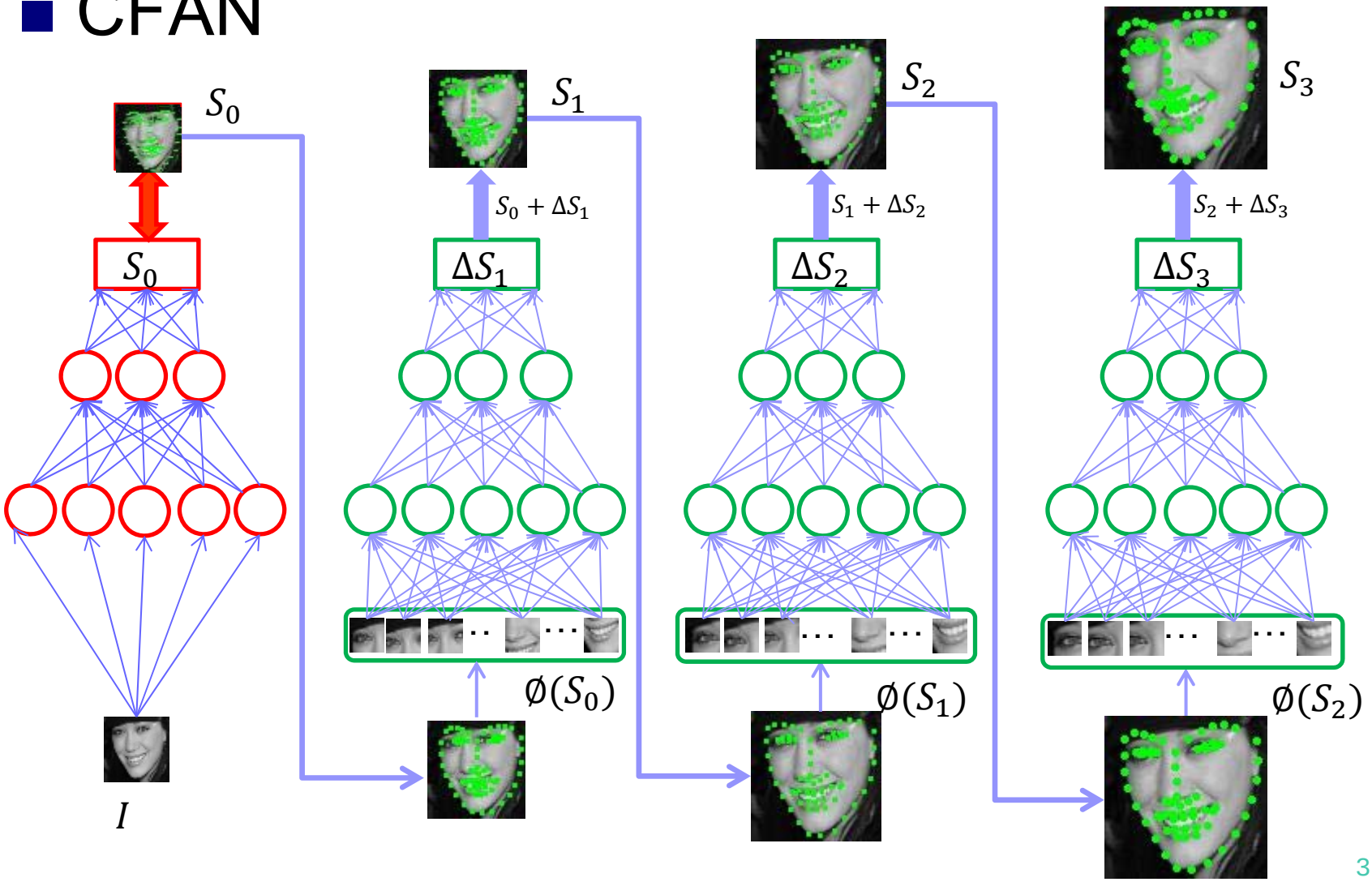
- CFAN: Coarse-to-Fine AE Networks



SAN: Stacked Auto-encoder Network

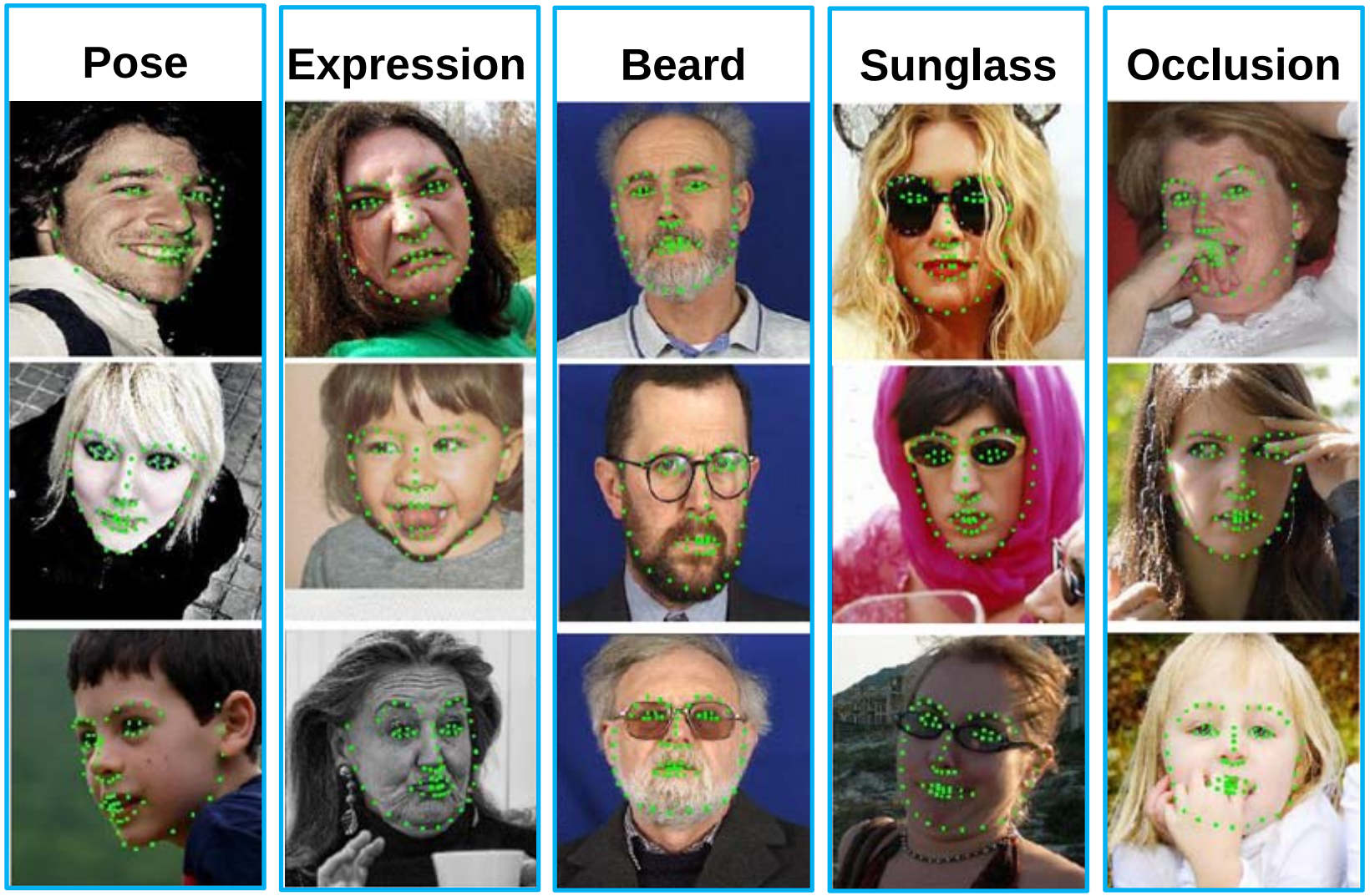
小数据条件下应用DL的例子

- CFAN



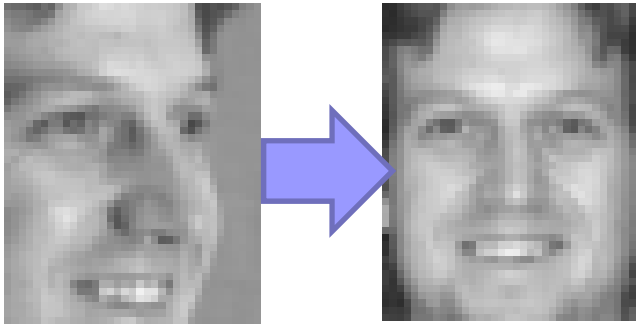
小数据条件下应用DL的例子

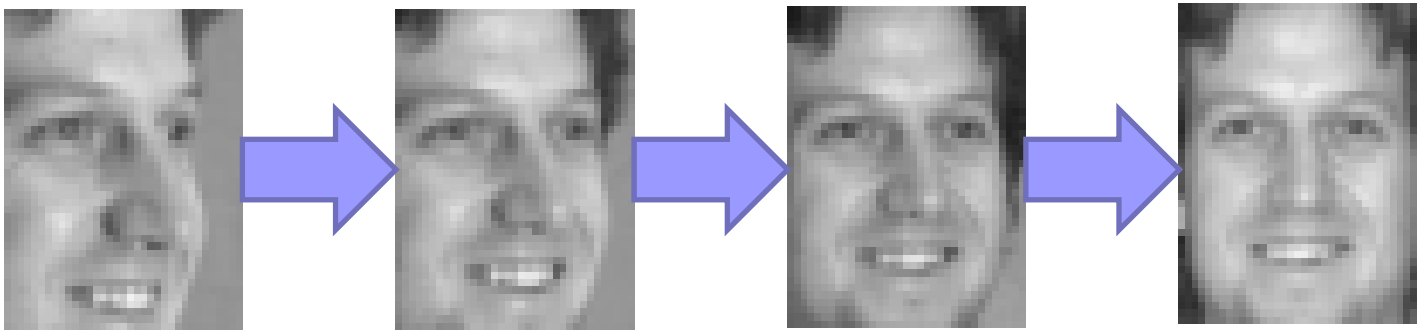
中科院计算所



另一个例子

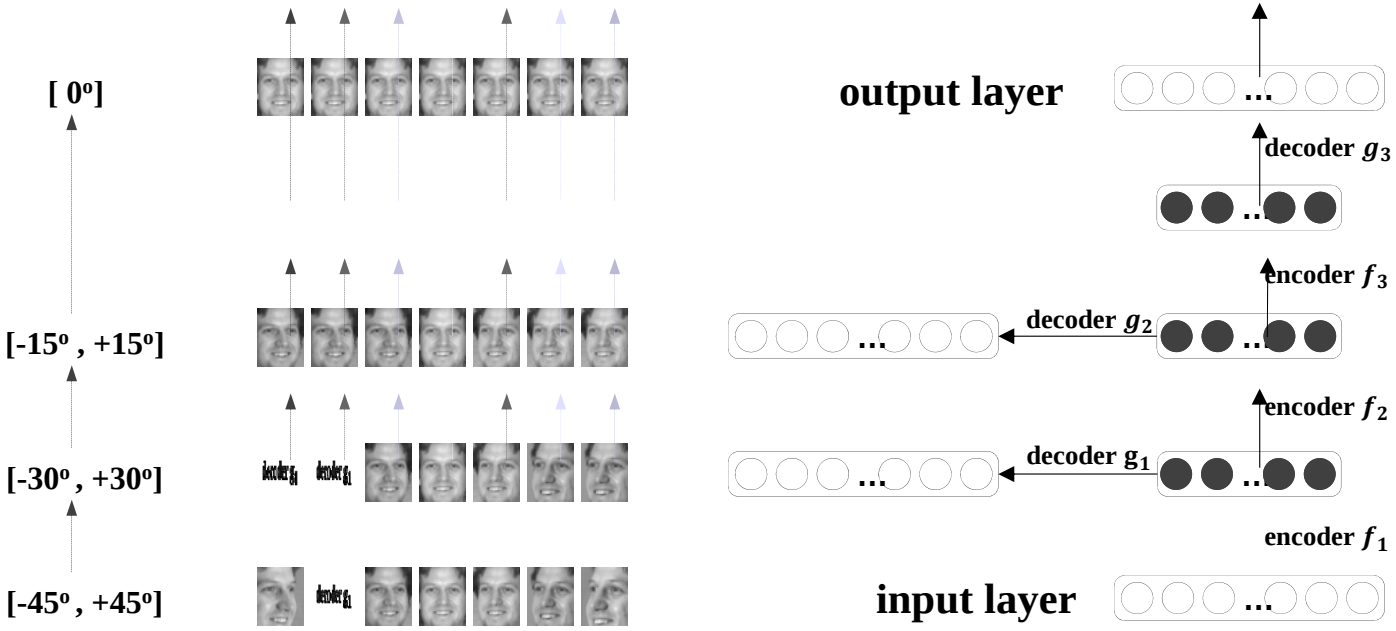
- 直接用DL
  - 失败！小数据障碍
- 我们的思路
  - 领域先验知识：姿态变化平滑变化
  - 渐进的达到目标





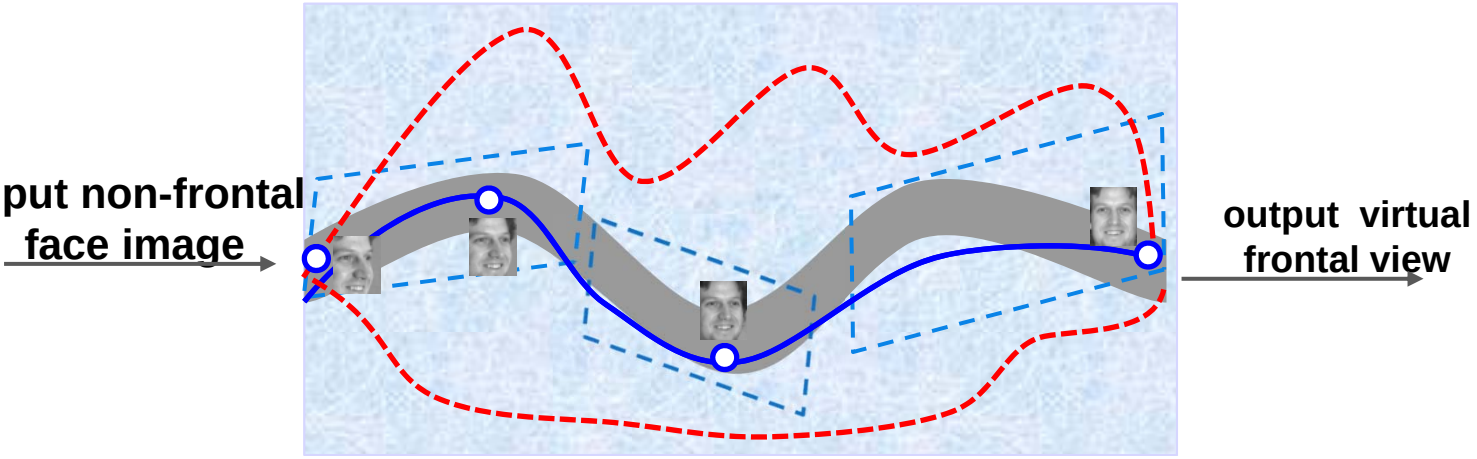
另一个例子

- Stacking multiple Progressive single-layer Auto-Encoders (SPAE)
  - 每一层实现较小的姿态变化



另一个例子

- Stacking multiple Progressive single-layer Auto-Encoders (SPAE)
  - 每一层实现较小的姿态变化
  - 靠中间目标约束每层的非线性变换



深度学习还能走多远？

- 类比SVM，Boosting…
  - DL将成为一个标准模块（尤其是特征学习）
- 人类知识全部来自数据吗？
  - 个人认为：NO！
  - 统计学习(深度学习）是归纳法
  - 我们同样需要演绎推理
    - 举一反三
    - 触类旁通：迁移
    - 无师自通：悟道

现在是数据为王，但要不要搞终身制？

Thank you!

Q&A