

大数据驱动的深度学习及其人脸识别应用

山世光

中科院计算所

sgshan@ict.ac.cn



中国科学院计算技术研究所

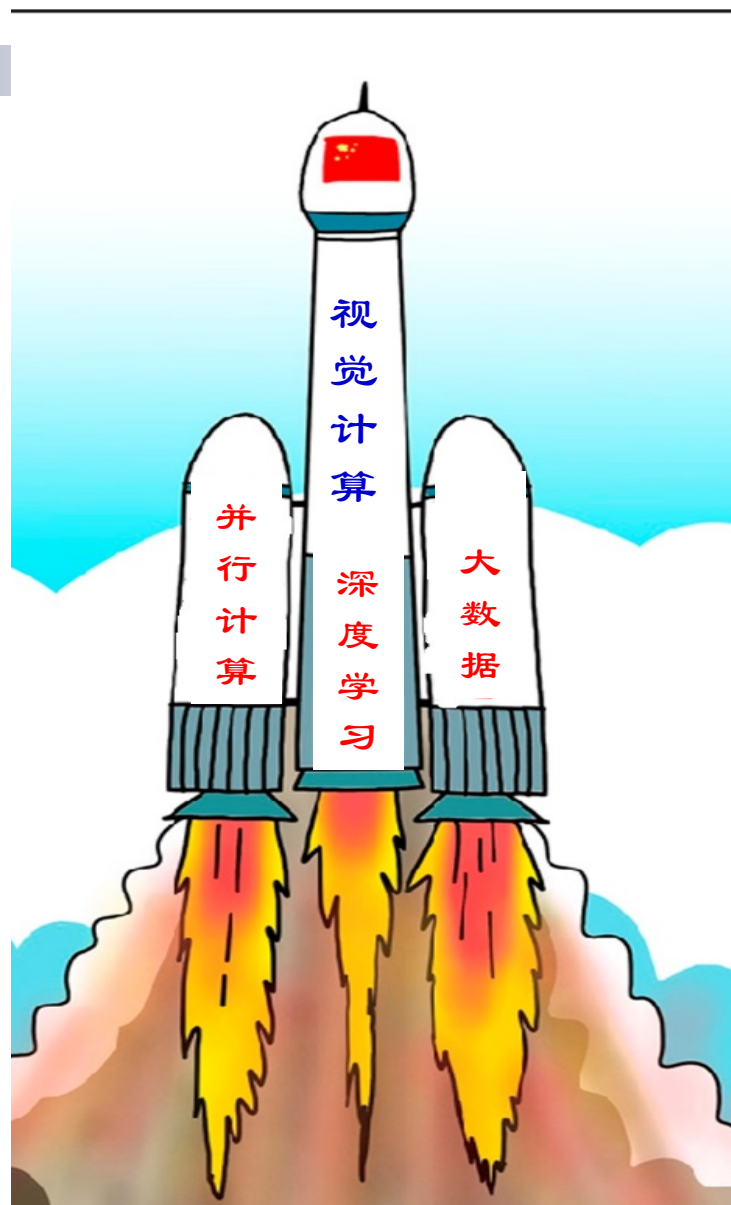
Institute of Computing Technology, Chinese Academy of Sciences



提纲

中科院计算所

- 人脸识别进展概况
- 关于深度学习的讨论
- 我们近期的部分工作





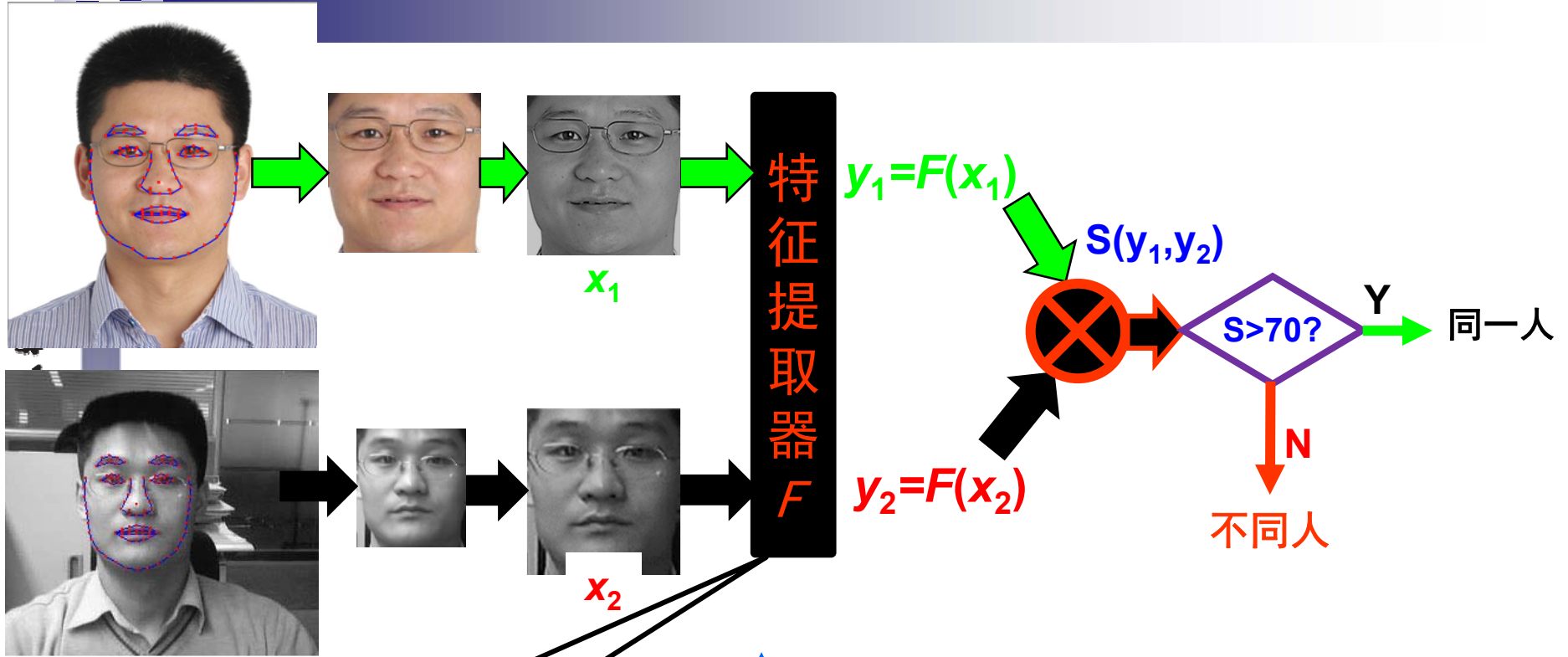
人脸识别：科幻 or 现实？

中科院计算所



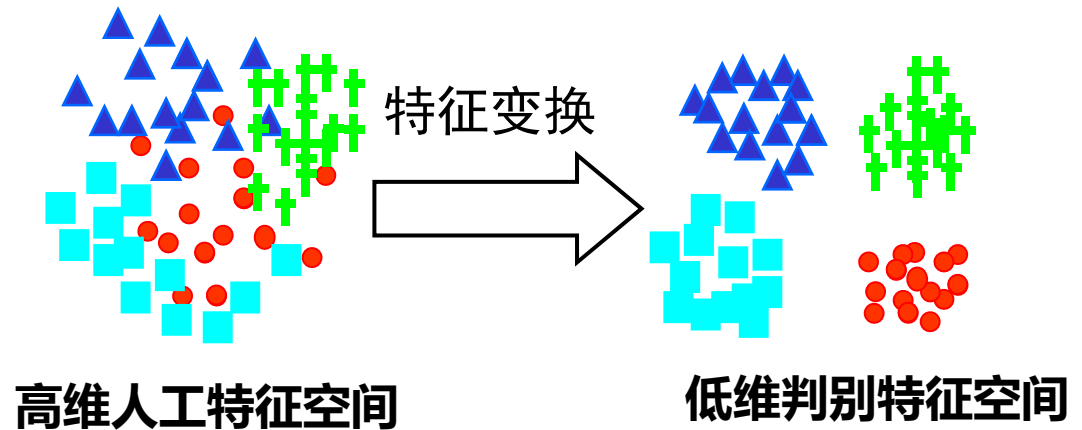


人脸识别基本原理



两步法:

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





人脸识别技术史本质上是 评测人脸库更替史！



从测试库变迁看人脸识别

中科院计算所

■ 学术界的里程碑（数据库）

□ ORL, E Yale B, AR: 1990~ (人数少于130人)

- 首选识别率: 95%~99% [J.Wright et al, 2008]
- 方法: 稀疏表示

$$(\ell^1): \quad \hat{x}_1 = \arg \min \|x\|_1 \quad \text{subject to} \quad Ax = y.$$

$$\min_i r_i(y) \doteq \|y - A \delta_i(\hat{x}_1)\|_2$$



ORL(40人, 每人10幅)

AR(126人, 每人26幅)



从测试库变迁看人脸识别

中科院计算所

■ 学术界的里程碑（数据库）

- ORL, E Yale B, AR: 1990~ (人数少于130人)
- 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 TIP10]

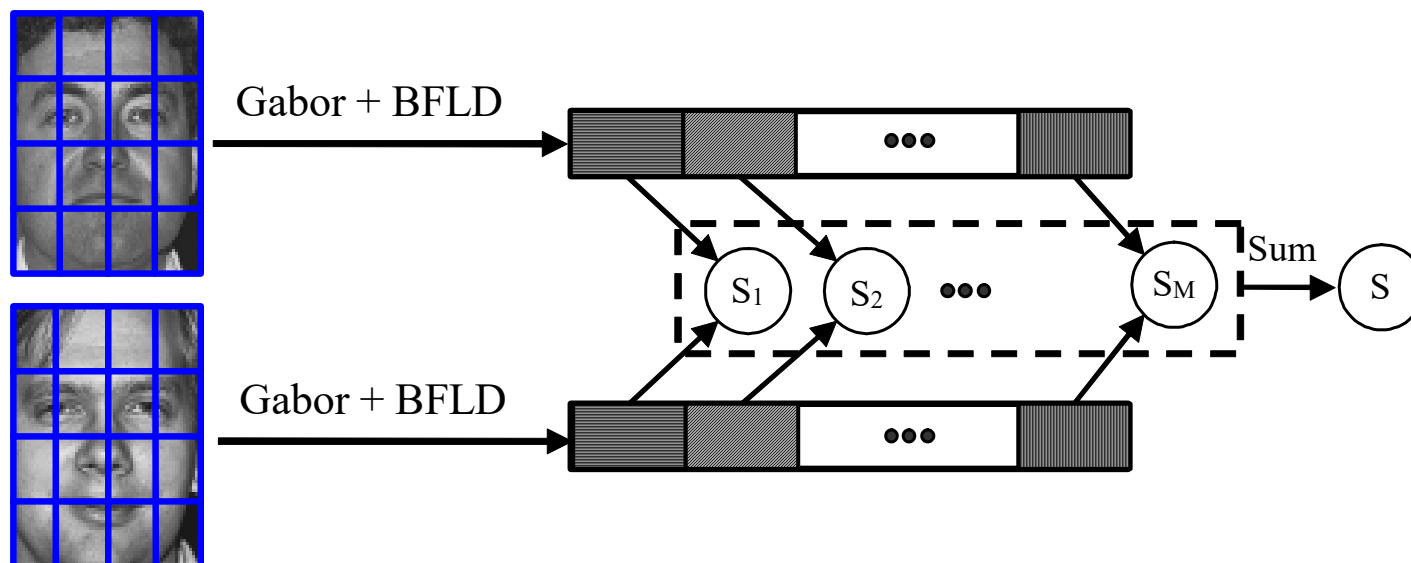




从测试库变迁看人脸识别

■ 学术界的里程碑（数据库）

- ORL, E Yale B, AR: 1990~ (人数少于130人)
- 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 TIP10]
 - 方法: Gabor幅值局部编码+Gabor相位局部编码+分块LDA





人脸识别性能对比——FERET

中科院计算所

对比方法	美国NIST发布的FERET测试集			
	FB	FC	Dup.I	Dup.II
我们的方法[T IP10]	99%	100%	94%	93%
[1]LGP + Borda Count (PR11)	99.8%	99.5%	89.2%	86.8%
[2]Image Gradient Orientations(T PAMI12)	-	-	88.9%	85.4%
[3]LGBP+Whitened PCA (ICB09)	98.1%	98.9%	83.8%	81.6%
[4]Oriented Edge Magnitudes (ECCV10)	98.1%	99%	79%	79.1%
[5]Improved SRC (T PAMI12)	96.6%	58.8%	71.6%	61.5%
[6]LBP (ECCV04)	97%	79%	66%	64%

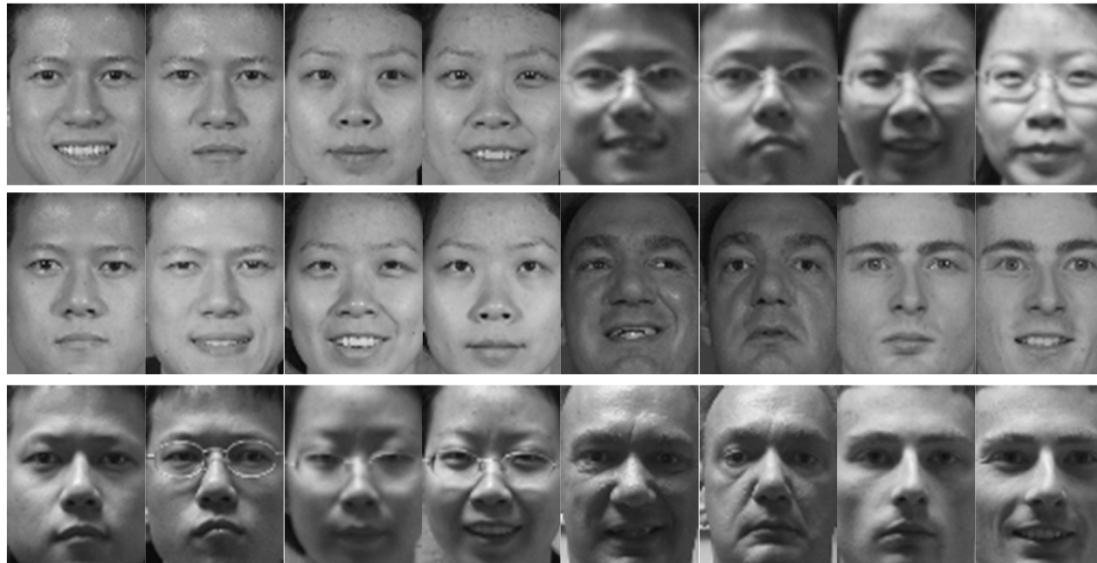
1. Claudio A.Perez , LeonardoA.Cament, LuisE.Castillo. Methodological improvement on local Gabor face recognition based on feature selection and enhanced Borda count. Pattern Recognition 44 (PR2011), 951–963
2. Georgios Tzimiropoulos, Stefanos Zafeiriou, Maja Pantic. Subspace Learning from Image Gradient Orientations. IEEE Transactions on Pattern Analysis And Machine Intelligence, IEEE T PAMI2012
3. Hieu V. Nguyen, Li Bai, and Linlin Shen, Local Gabor Binary Pattern Whitened PCA: A Novel Approach for Face Recognition from Single Image Per Person. ICB 2009, LNCS 5558, pp. 269–278, 2009
4. Andrew Wagner, John Wright, Arvind Ganesh, Zihan Zhou, Hossein Mobahi, and Yi Ma. Toward a Practical Face Recognition System: Robust Alignment and is And Machine Intelligence, IEEE T PAMI2012
5. Ngoc-Son Vu, Alice Caplier. Face Recognition with Patterns of Oriented Edge Magnitudes. ECCV2010
6. A. Timo, H. Abdenour, and P. Matti. Face recognition with Local Binary Patterns. ECCV 2004



从测试库变迁看人脸识别

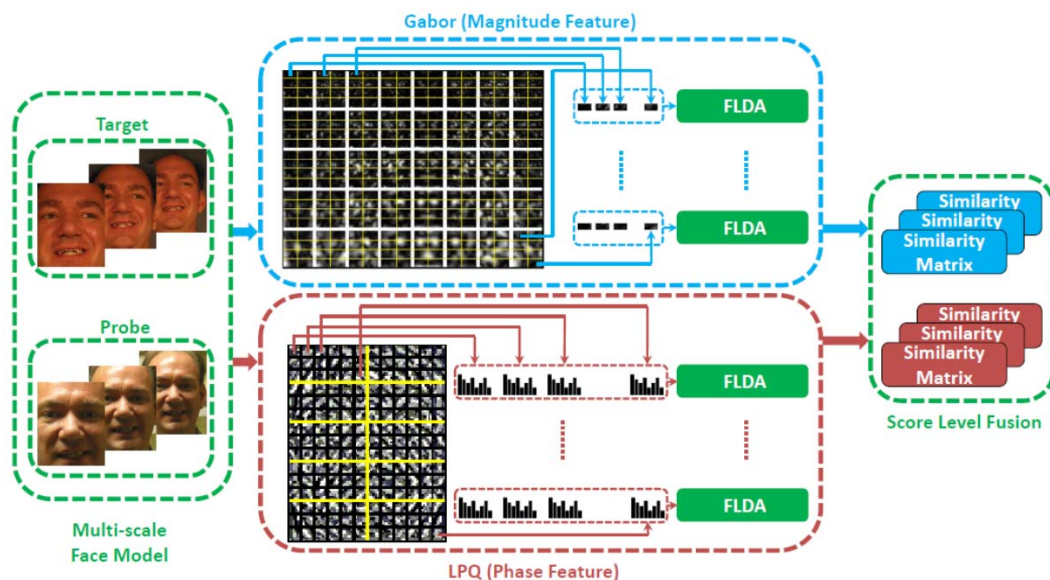
中科院计算所

- 学术界的里程碑（数据库）
 - ORL, E Yale B, AR: 1990~ (人数少于150人)
 - FERET: 1994~2010 (1196人, 每人2~5张图)
 - 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]



从测试库变迁看人脸识别

- 学术界的里程碑（数据库）
 - ORL, E Yale B, AR: 1990~ (人数少于150人)
 - FERET: 1994~2010 (1196人, 每人2~5张图)
 - 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]
 - 方法: Gabor幅值局部编码+LPQ+分块LDA





人脸识别性能对比——FRGC

中科院计算所

Methods	FRGC test set		Verification Rate (when FAR=0.1%)	
	Exp.1	Exp.4	Exp.1	Exp.4
FRGC Baseline (Eigenfaces)	66%	12%		
Hybrid Fourier [Hwang 2006]	91%	74%		
KFA [Liu 2006]	92%	76%		
DCT_EFM [Liu 2008]	n/a	84%		
Gabor+LDA [Han 2010]	97%	78%		
LBP & Gabor + KLDA+SN [Tan 2010]	N/A	88%		
Our methods [Su 2009]	98%	89%		
RTF + RCF [Deng 2010]	99%	93.5%		
Our Methods [Li 2012]	99%	96%		

[Hwang 06] W. Hwang, et. al, Multiple Face Model of Hybrid Fourier Feature for Large Face Image Set, In CVPR'06.

[Liu 06] C. Liu, Capitalize on dimensionality increasing techniques for improving face recognition performance, In PAMI 2006.

[Liu 08] Z. Liu and C. Liu, Fusion of the complementary Discrete Cosine Features in the YIQ color space for face recognition, in CVIU 2008.

[Han 10] Z. Han, C. Fang, X. Ding, A Discriminated Correlation Classifier for Face Recognition, Proc. of 2010 ACM Sym. on Applied Computing, 2010

[Tan 10] X.Tan, B.Triggs. Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions. IEEE T IP 19(6), 2010.6

[Deng 13] Deng, W., Hu, J., Guo, J., Cai, W., Feng, D.: Emulating biological strategies for uncontrolled face recognition., PR, 2013

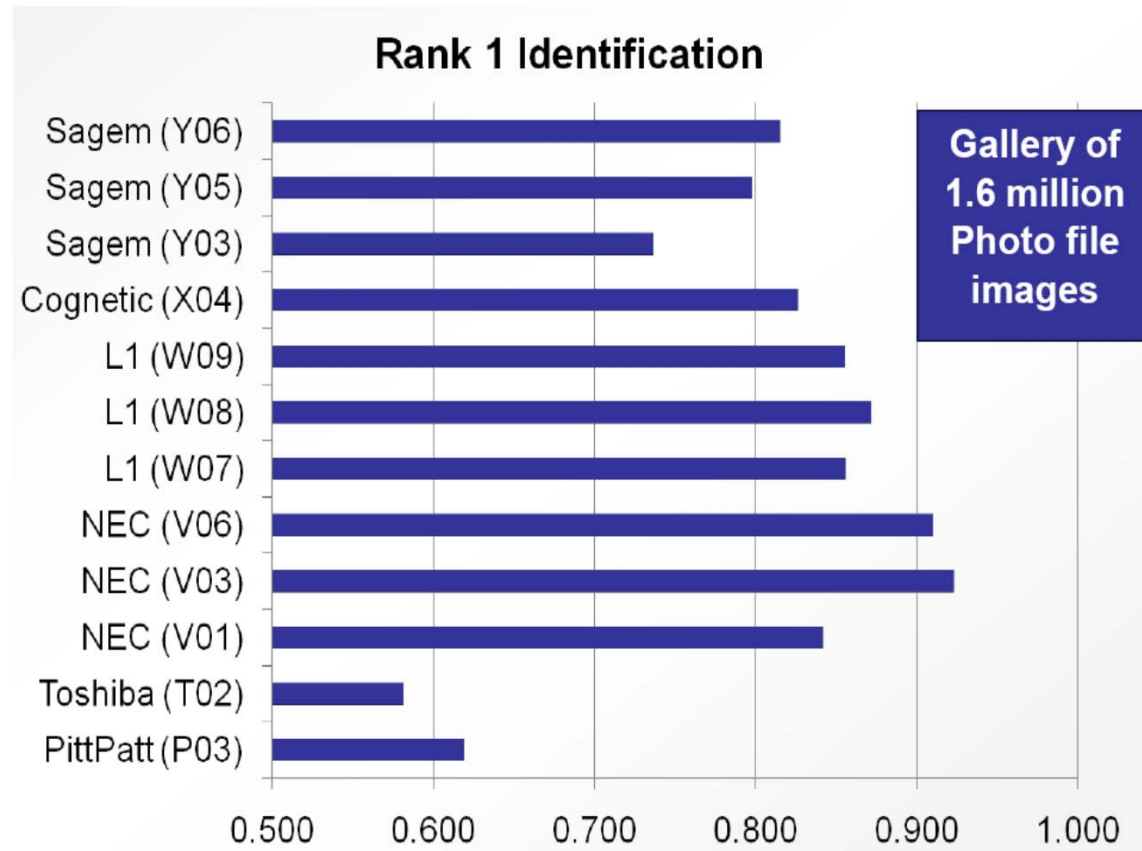
[Li 12] Y.Li, S. Shan, H. Zhang, S. Lao, X. Chen. Fusing Magnitude and Phase Features for Robust Face Recognition, ACCV2012



从测试库变迁看人脸识别

中科院计算所

- 美国NIST MBE 2010结果
 - 证件照 vs. 证件照





从测试库变迁看人脸识别

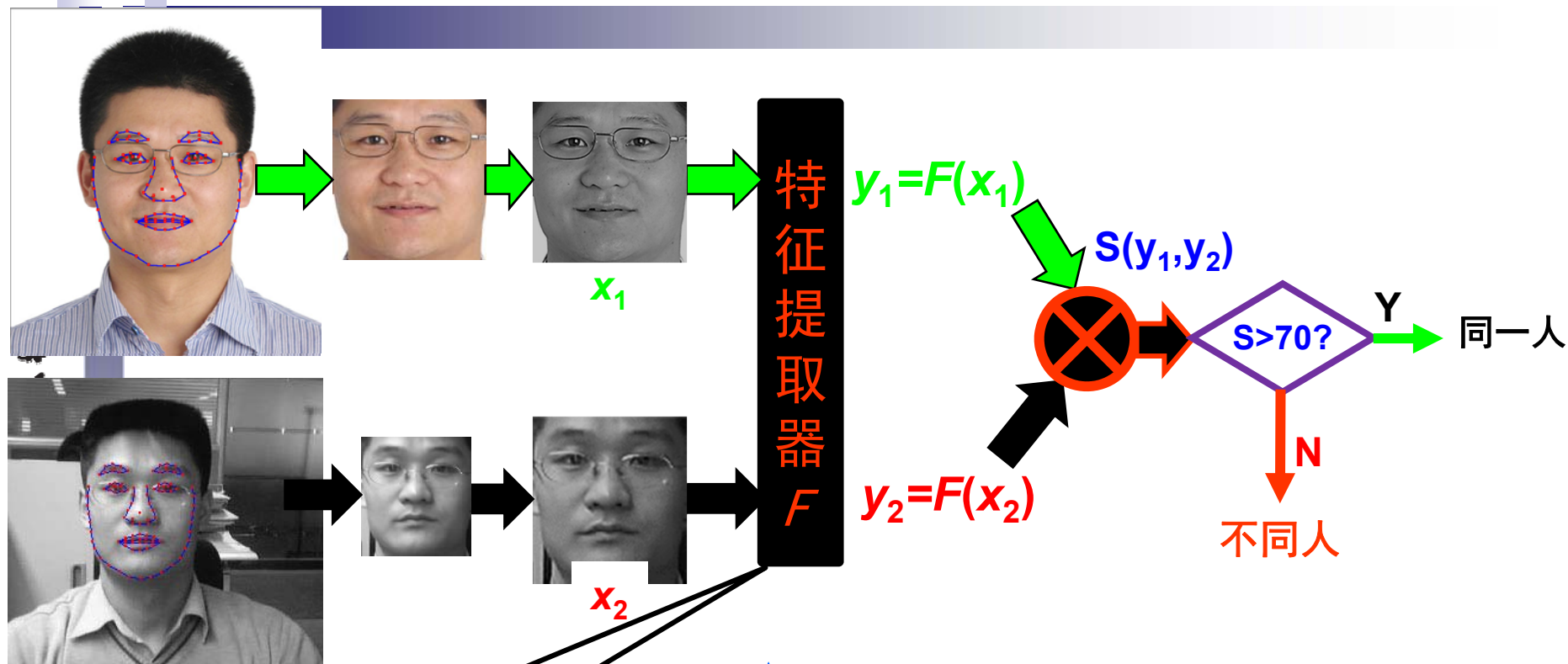
中科院计算所

- 中科院计算所的人脸识别系统
 - 在270万人的证件照数据库上进行比对
 - 测试样本和注册样本均为证件照
 - 10万测试样本，2700亿次比对

对比系统	识别率
我们的系统	91.57%
某对比系统	~60%

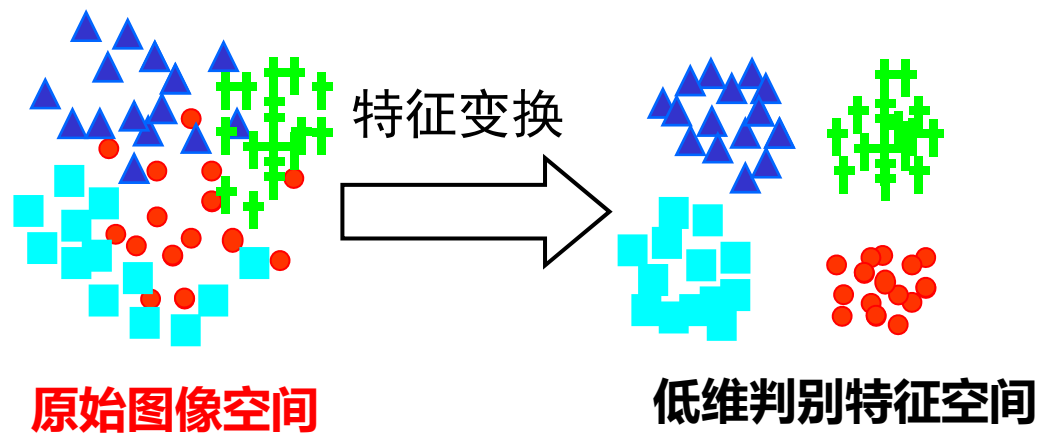


人脸识别基本原理



端到端的特征学习

1. 多层卷积+多变换
2. 非线性特征提取
3. 损失函数优化





过去2年的新进展——深度学习

中科院计算所

- **人脸识别技术本质上是评测人脸库更替史！**
- 学术界的里程碑（数据库）

过去2年错误率从5%下降到0.5%
(错300对→错30对)

- 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]



从测试库变迁看人脸识别

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
 - 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



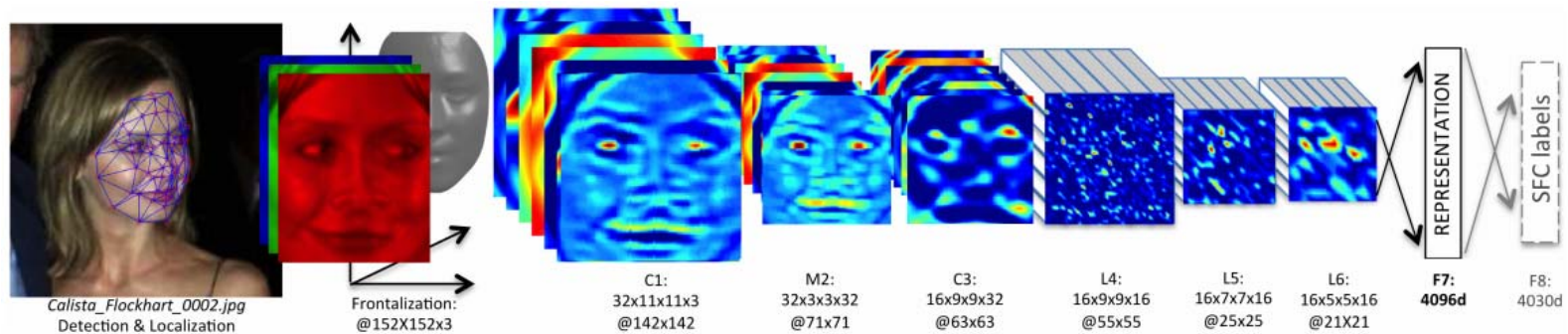
Huang G B, Ramesh M, Berg T, et al. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report, University of Massachusetts, Amherst, 2007.



过去2年的新进展——深度学习

中科院计算所

- 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

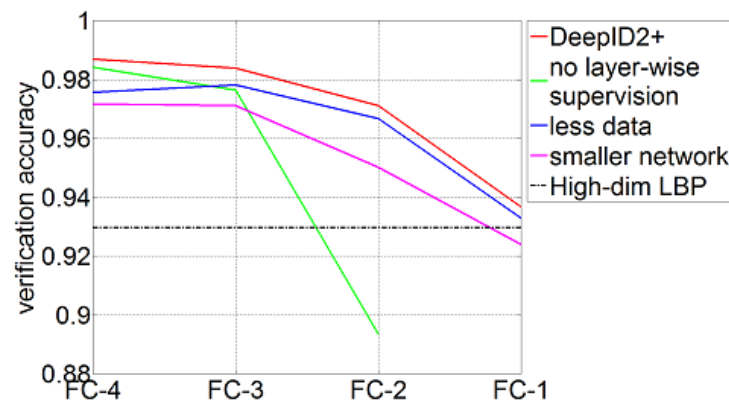
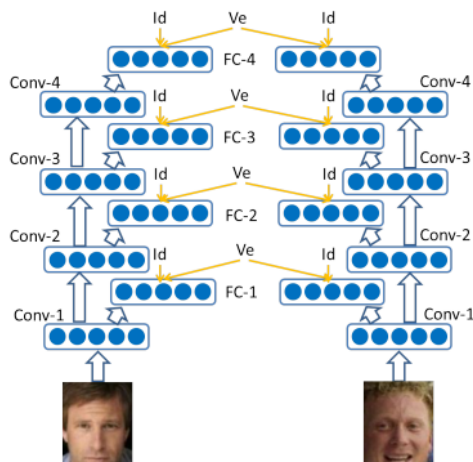
[1] Taigman Y, Yang M, Ranzato M A, et al. Deepface: Closing the gap to human-level performance in face verification. CVPR, 2014.



DeepID2+

中科院计算所

- 在50个人脸Patch上分别训练CNN（4个卷积层，4个全连接层，4个verification损失信号和1个identification损失信号）
- 训练数据：10K人，202K名人图像



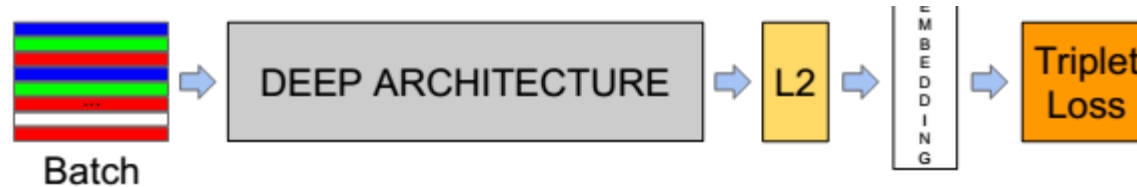
Sun Y, Wang X, Tang X. Deeply learned face representations are sparse, selective, and robust. arXiv preprint, 2014.



FaceNet

中科院计算所

- 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



从测试库变迁看人脸识别

- 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&I) [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]

□ 然后呢? !



研究与应用现状概况

- 研究方面
 - 全面沦陷于深度学习(CNN)的应用
 - 在新理论和新方法方面落伍于国际
 - 优化方法, 新的网络结构等方面
- 应用方面
 - 似乎“超前”于国际: 户口/护照查重系统
 - 各类“刷脸应用”雨后春笋
 - 可见光的活体检测技术不过关
 - 身份证上的应用政策局限性: 加水纹
 - 安防动态布控人脸识别初见成效
 - 但距离全面应用尚有非常大的距离



提纲

中科院计算所

- 人脸识别历史与近期进展
- 关于深度学习的讨论
- 我们近期的部分工作

2006



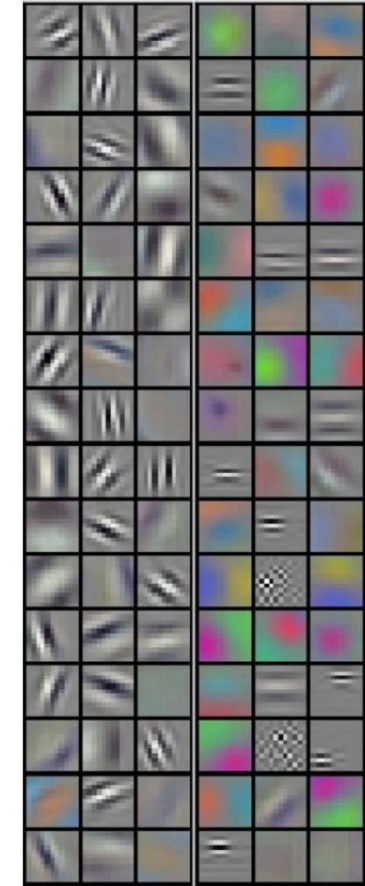
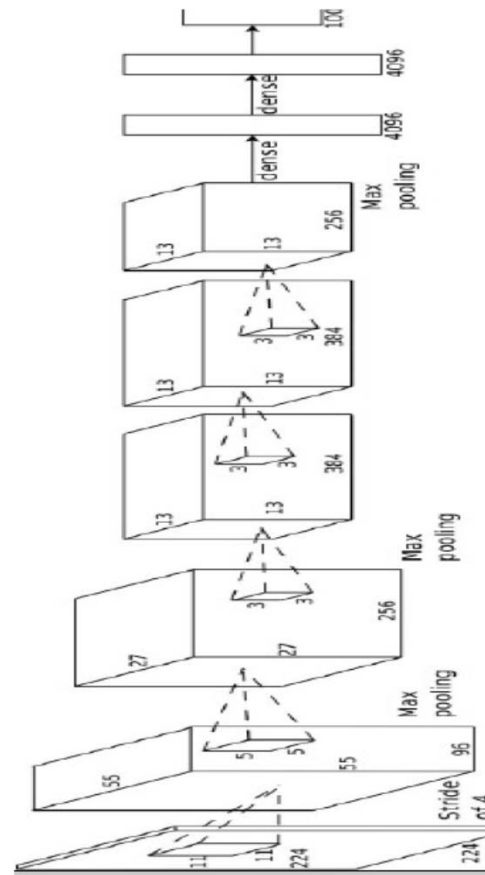


ImageNet with Deep CNN

中科院计算所

■ 方法：大规模CNN网络【650K神经元, 60M参数】

4M	FULL CONNECT
16M	FULL 4096/ReLU
37M	FULL 4096/ReLU
	MAX POOLING
442K	CONV 3x3/ReLU 256fm
1.3M	CONV 3x3ReLU 384fm
884K	CONV 3x3/ReLU 384fm
	MAX POOLING 2x2sub
	LOCAL CONTRAST NORM
307K	CONV 11x11/ReLU 256fm
	MAX POOL 2x2sub
	LOCAL CONTRAST NORM
35K	CONV 11x11/ReLU 96fm





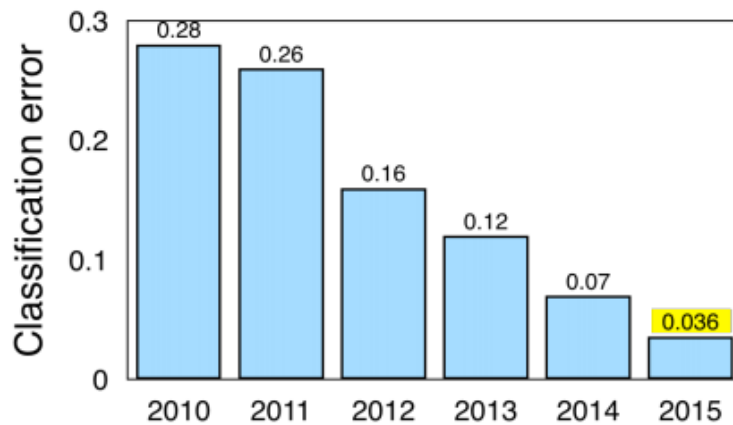
对智能领域的局势判断

中科院计算所

智能方向疑似跨越式发展

□ 诸多超出预期的进步

■ 视觉识别, 语音识别, DeepMind, Alphago...



人脸识别: LFW上错误率5%下降到**0.5%**

VOC图像分割: **50%→75%**(交并比)

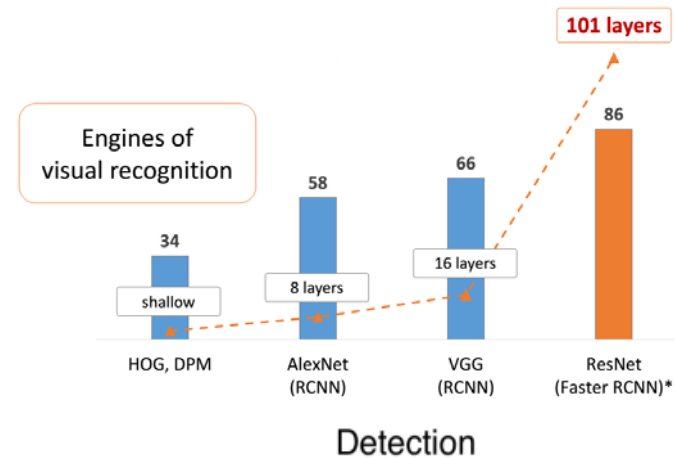
看图说话(图题生成):

Bleu:1-Gram: 66.3 (Flickr 30k测试)

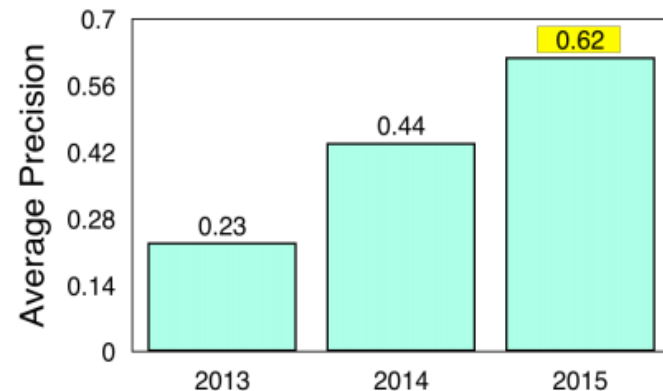
Bleu:2-Gram: 42.3 (Flickr 30k测试)

MSCOCO2015上 (Google系统)

27.3%优于人, 31.7%通过图灵测试



Detection



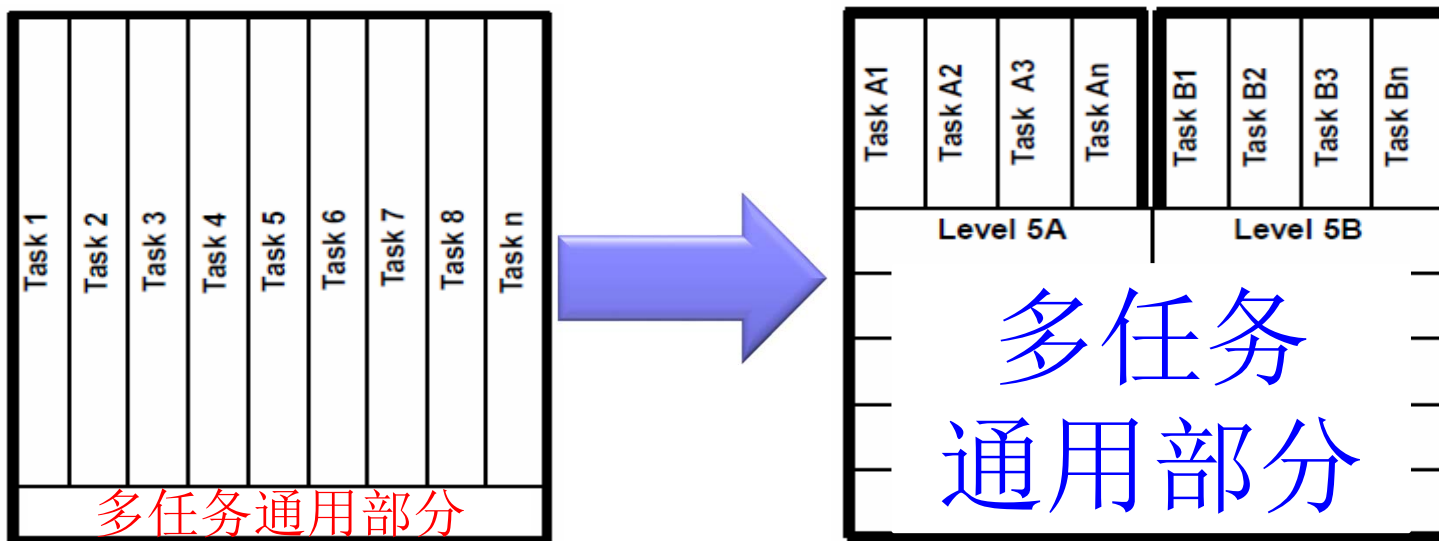


对智能领域的局势判断

中科院计算所

■ 智能方向疑似跨越式发展

- 诸多超出预期的进步
- 割据式的研究格局正被打破
 - 小：人脸识别与物体识别
 - 大：视觉计算与语言计算:ImageNet→Visual Genome
 - AI历史
 - 通用智能→专家系统→构建于通用智能之上的、协同的专家系统

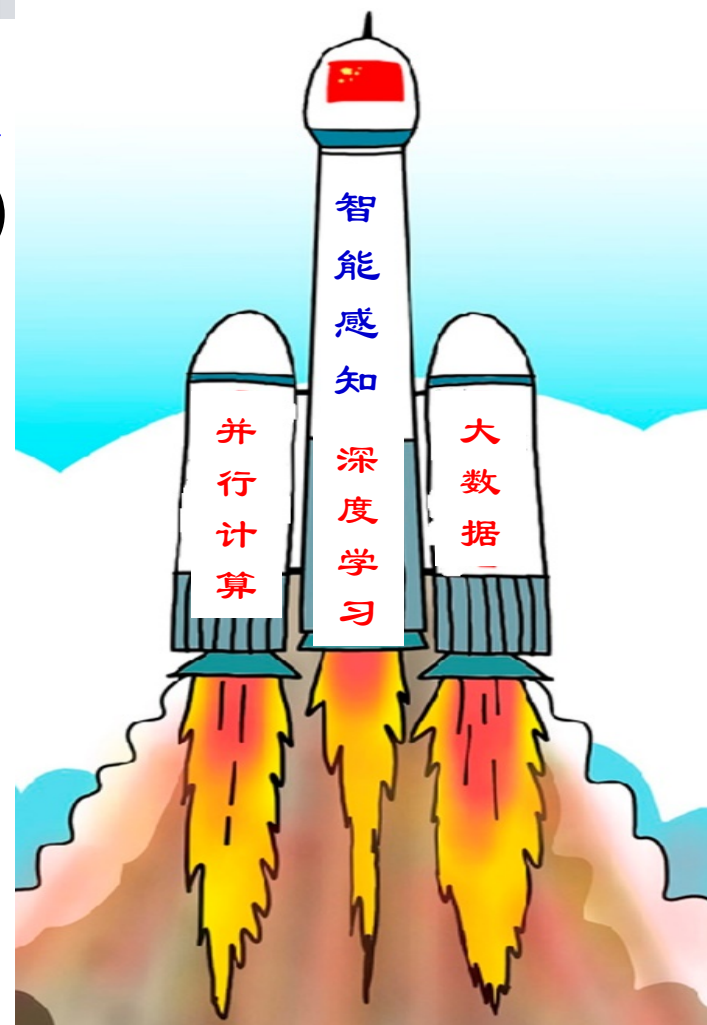


■ 智能方向疑似跨越式发展

- 三大引擎助推(since 2011)
- 深度学习：箭体
- 大数据：燃料
- 并行计算：燃烧效率

■ 为什么深度学习复兴？

- “容量(复杂度)” 足够大，因而可以“装”下大数据
 - 例: 650K神经元, 60M参数
- 大数据使得其学习可能
- 并行计算使得学习时间可控





DL带来观念的变革

- 人工领域知识驱动→数据驱动的学习思想
- 小数据时代→控制模型复杂度避免过拟合
- 大数据时代→提高模型复杂度避免欠拟合
 - “大数据+简单模型”是错误的!
 - 维数灾难(降维)→高维有益(升维)
- 分步、分治思想
 - →协同学习(joint learning)思想
 - →End-to-end的全过程学习
- 软硬件协同 (GPU训练, 算法加速)
- 全社区协同 (开源, 开放模型, 开放数据)



DL是类脑智能方法吗？

- DL的确受到脑信息处理方式启发
 - 初级视觉神经元的“类Gabor小波编码”
 - 感受野的概念，分层逐级抽象
- 不“类脑”
 - 本质上，脑的计算“机制”尚不清晰
 - 大量化学过程，机理不清楚
 - 脑的连接更多样、更复杂、规模更大
 - Top-down，反馈回路，巨量链接
 - 学习过程未必需要大量数据
 - 先天——生物进化的结果（大数据长期训练）
 - 后天学习——更多演绎推理，迁移学习



DL不能做什么？

- 用做“特征学习”或“非线性变换”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
 - 提供了实现复杂非线性变换的可行手段
- 严重倚重大数据，小数据深度学习不可靠
 - 需要引入领域知识，或者进行迁移学习
- 难以演绎推理
 - DL是归纳学习，难以举一反三，更难无师自通
- 在一些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



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
- CNN在CV领域的成功，已经说明了领域知识的重要性
 - 卷积操作：过去Gabor, Haar, SIFT...
 - Pooling操作：Max, Sum...
- 小数据条件下，领域知识尤其重要

Data is king, and DL is queen?



关于深度学习——开放问题

- 深度学习理论
- 数据角度
 - 基于无监督大数据的深度学习
 - 大而脏乱差数据条件下的DL
 - 小数据条件下的DL
- 模型设计与学习角度
 - 新的模型结构及部件
 - 网络结构本身的学习
 - 经验和领域知识的便捷嵌入
 - 感知阶段可反馈的深度网络
 - 深度模型的知识迁移与适应
 - 新的优化方法和学习策略



关于深度学习——开放问题

中科院计算所

- 双引擎带来超出预期的进步
 - 机器可以做的比人好!
 - 数据多大碗，就有多大产!
 - 大数据归纳(和记忆)的进步
- 但，需要更多能力引擎
 - 演绎和推理能力：原理和常识
 - 触类畅通：迁移;他山之石
 - 举一反三：小数据学习能力
 - 无师自通：悟；自学能力
 - 勤学好问：交互





警告!!!

- 历史经常重演
 - 神经网络的兴衰史
- 相比CV领域的狂热ML领域远未一边倒
- 建议
 - 要会DL
 - 但不要只会DL



提纲

中科院计算所

- 人脸识别历史与近期进展
- 关于深度学习的讨论
- 我们近期的部分工作



我们近期的相关工作

■ 与DL相关的学术文章

- Stacked Progressive Auto-Encoder (SPAEC) 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](#)]



在特征或测度学习方面

中科院计算所

问题：较小规模数据条件下的视觉表示与测度学习

思路：在数据驱动模型中嵌入专家经验或领域知识

分段渐进的学习策略

由粗到细的、多阶段深度自编码网络
【ECCV'14】

人脸姿态分层渐进正面化深度自编码网络
【CVPR'14】

图像分辨率分层渐进的深度自编码网络
【ECCV'14】

模型中嵌入领域知识或规律

基于人脸部件区块的多测度协同学习
【CVPR'13, CVPR'15】

表情运动单元友好的深度学习模型
【IEEE FG'13】

充分利用多类型的数据

利用无监督、弱监督数据的中层特征学习
【IEEE T IP'13】

利用其它任务标注数据集的特征学习
【ACM ICML'14表情识别竞赛**冠军**】
【ICCV'15事件分类竞赛**冠军**】
【ICCV'15年龄估计竞赛**亚军**】

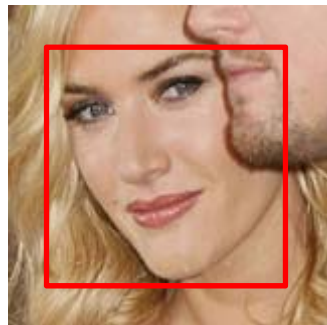


由粗到细的、多阶段深度自编码网络

中科院计算所

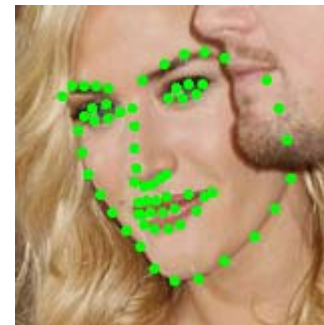
■ Face Alignment

- Predict facial landmarks from detected face



Detected face
region $I(u,v)$

Goal



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

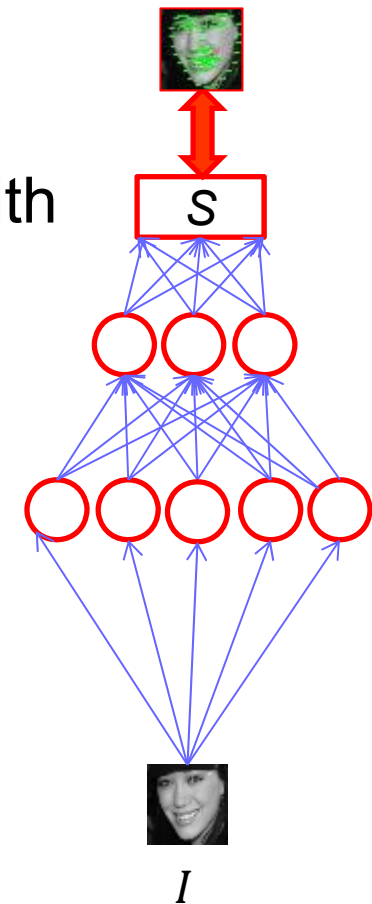
$$S = H(I), I \in R^{w \times h}, S \in R^{2L},$$



由粗到细的、多阶段深度自编码网络

中科院计算所

- Directly apply Stacked Auto-Encoder (SAE)? OK, but not good. Why?
 - Easily overfit to **small data**
 - Typically only thousands of images with landmark annotations
- Our ideas – **exploiting priors**
 - Features are partially handcrafted
 - SIFT, shape-indexed
 - Better initialization
 - Coarse to fine

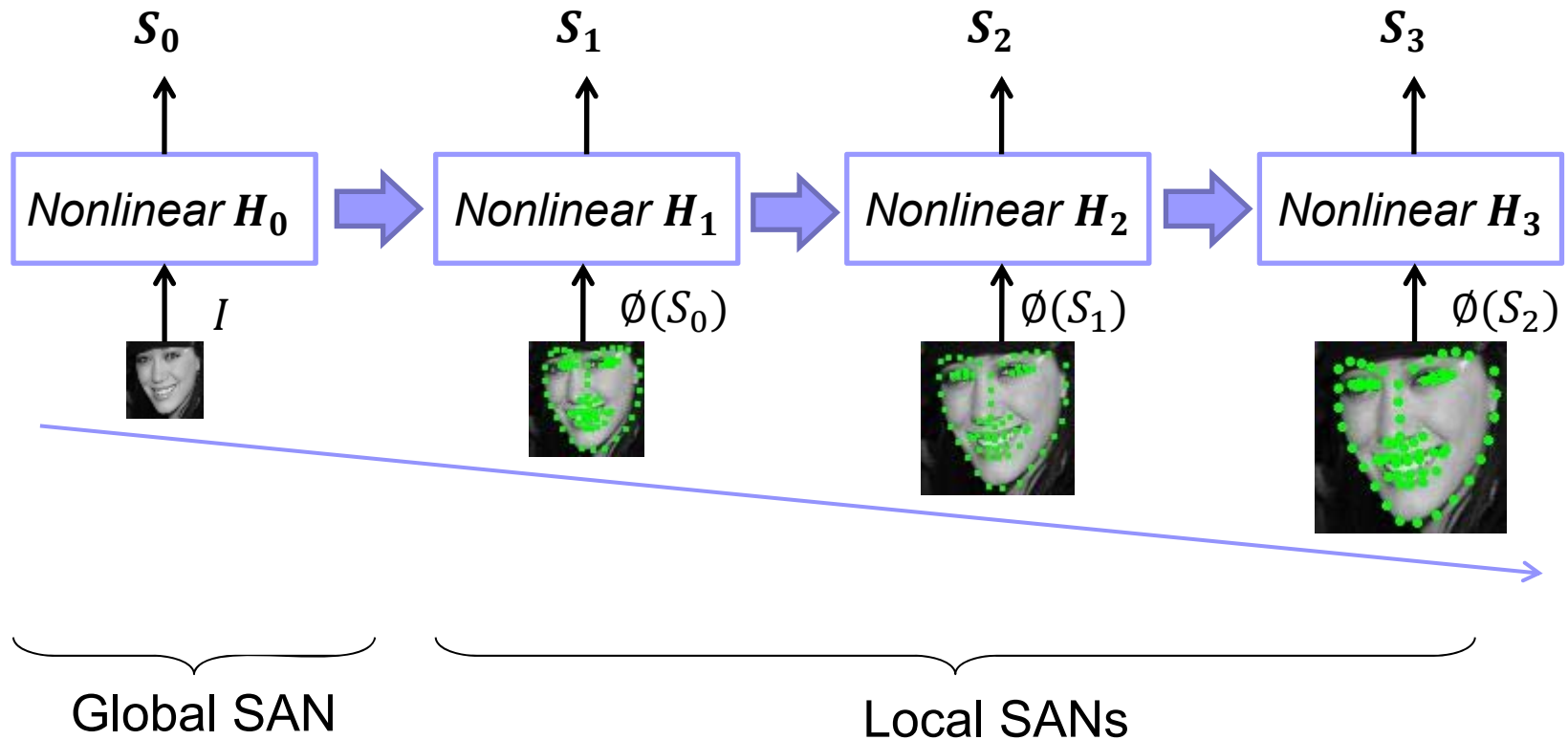




由粗到细的、多阶段深度自编码网络

中科院计算所

■ Schema of Coarse-to-Fine AE Networks



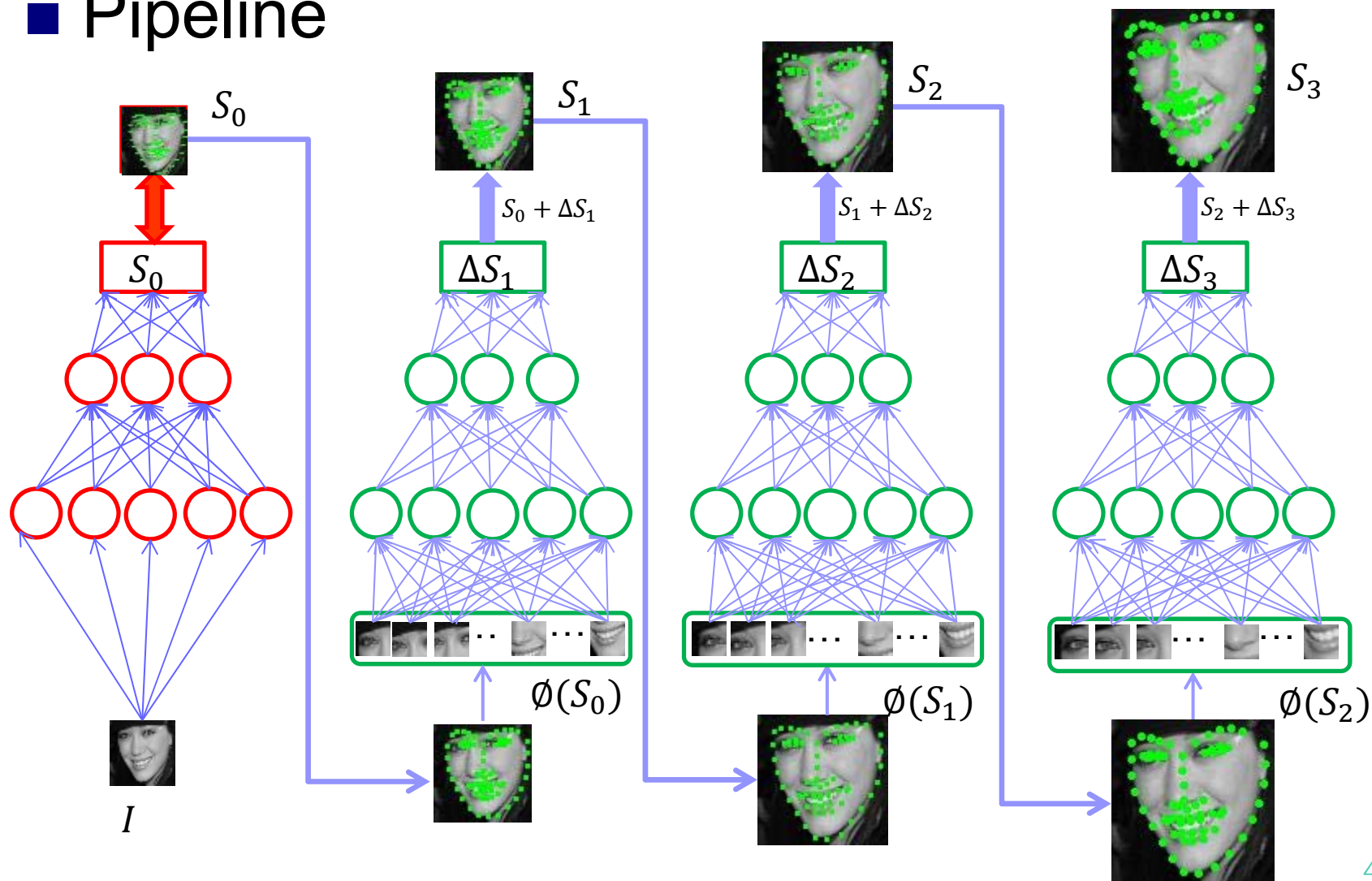
SAN: Stacked Auto-encoder Network



由粗到细的、多阶段深度自编码网络

中科院计算所

■ Pipeline



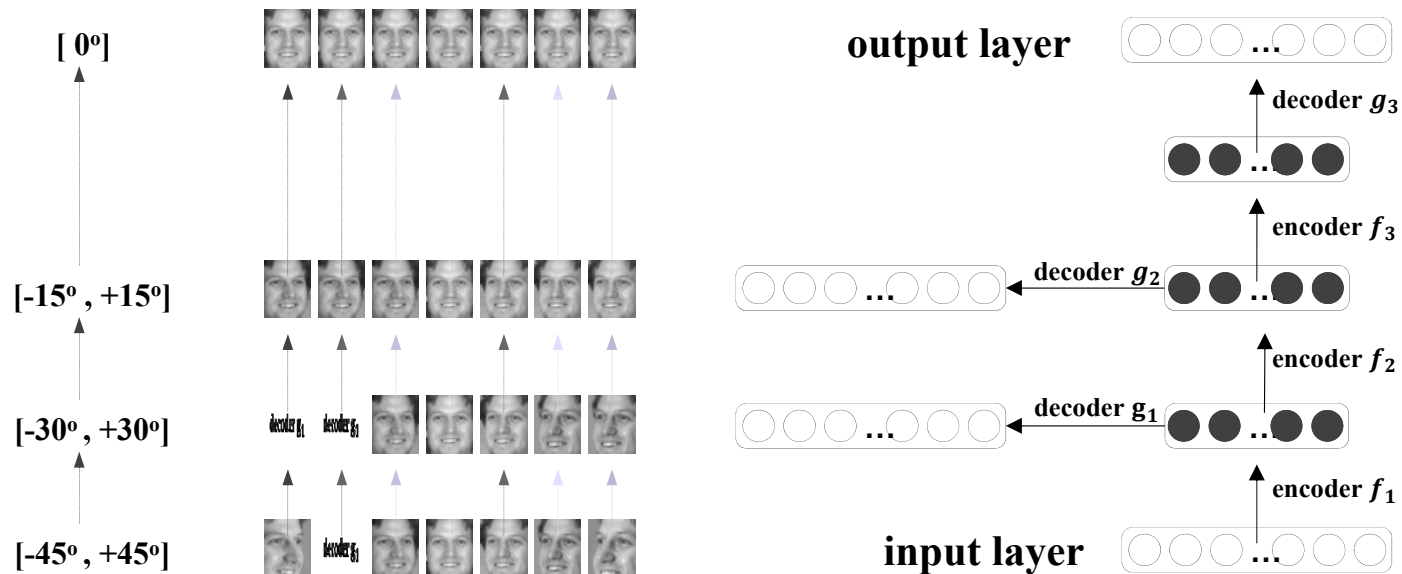


人脸姿态分层渐进正面化深度自编码网络

中科院计算所

Basic idea

- Stacking multiple Progressive single-layer Auto-Encoders
- Each PAE maps non-frontal faces to another with smaller pose





人脸姿态分层渐进正面化深度自编码网络

中科院计算所

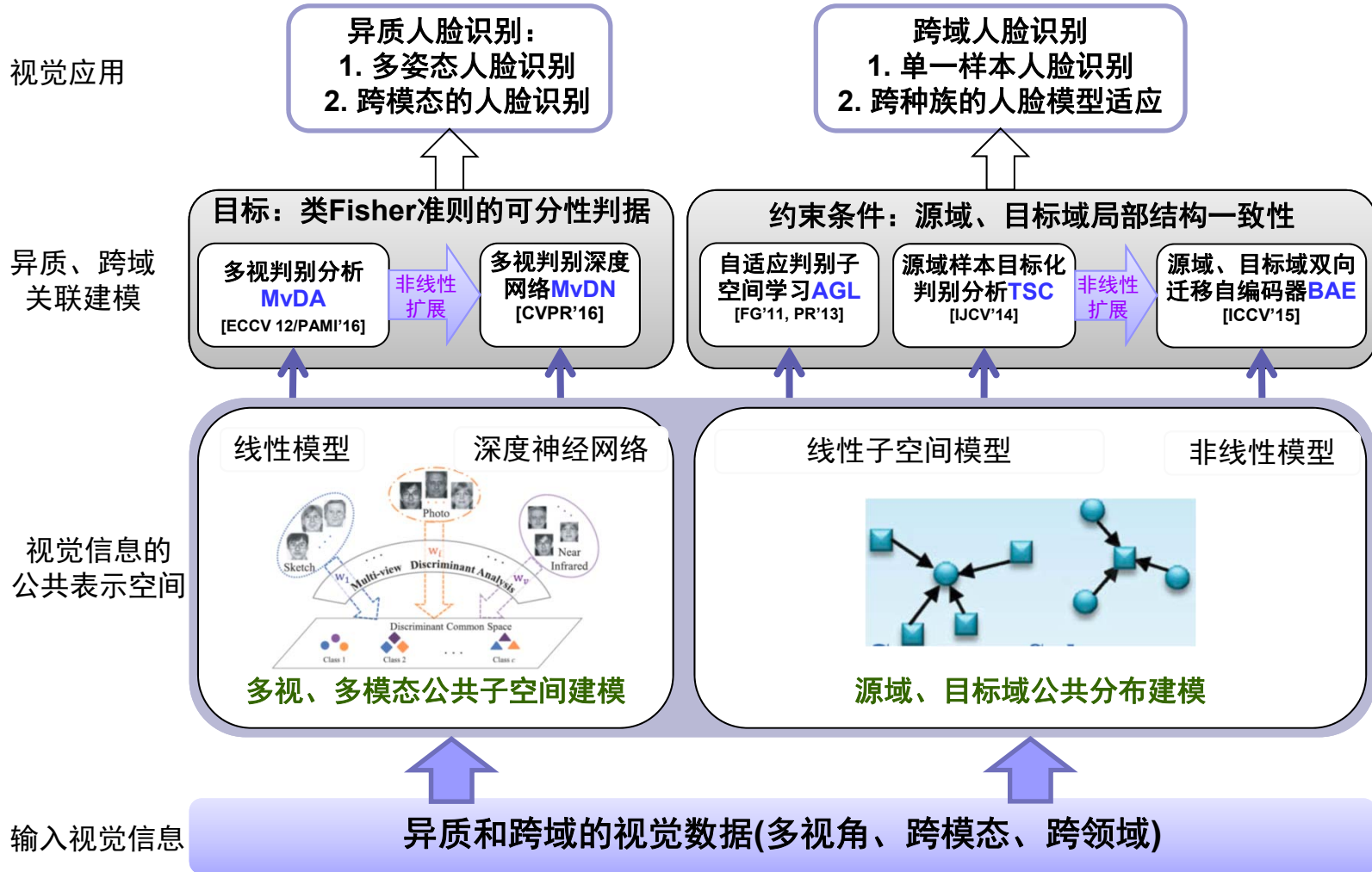
0°
↑
15°
↑
30°
↑
45°





在异质跨域视觉模式识别方面

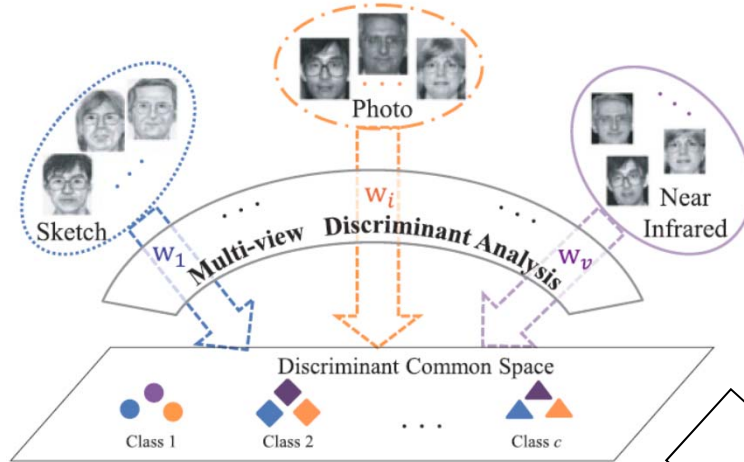
中科院计算所





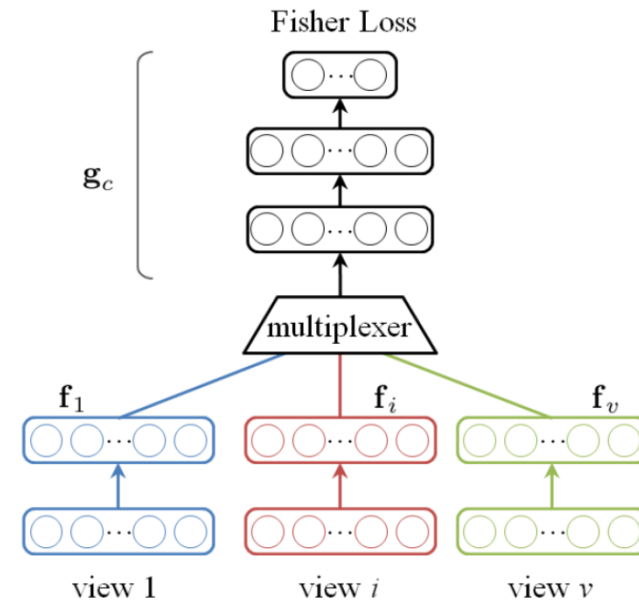
Multi-view Discriminant Analysis

中科院计算所



ECCV'12, TPAMI'16
Multi-view Discriminant
Analysis (MvDA)

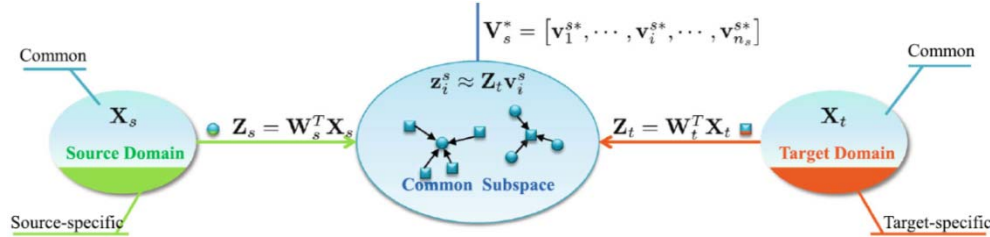
CVPR'16
Multi-view Discriminant Network
(MvDN)



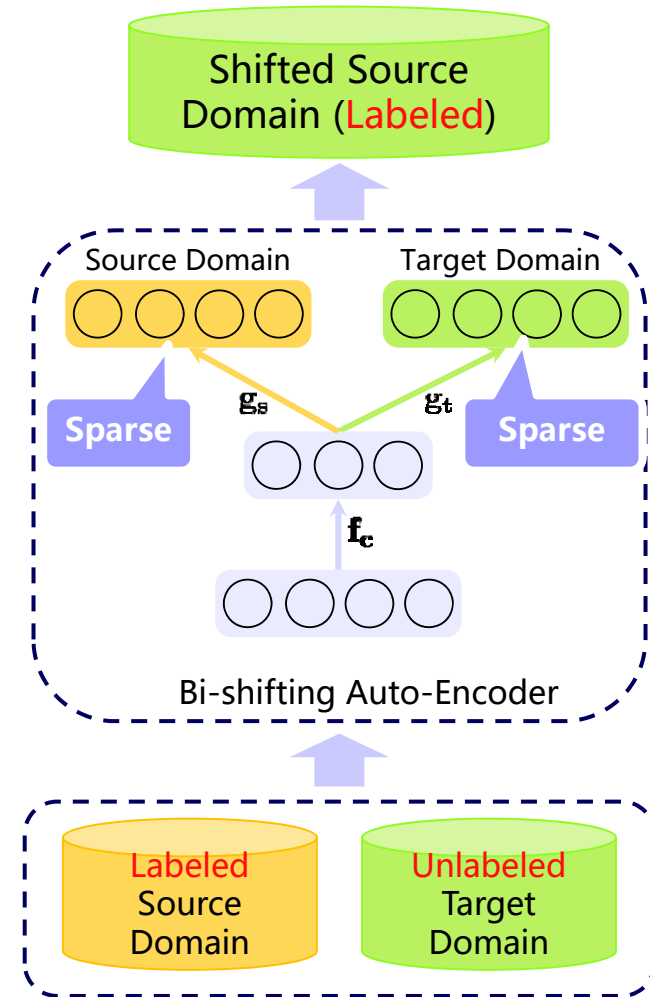
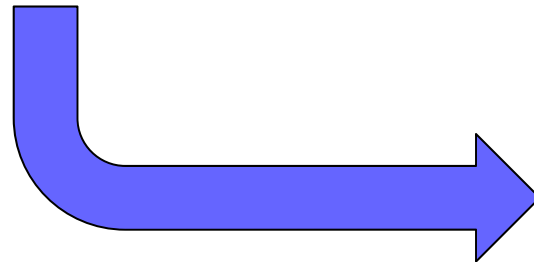


Bi-shifting Auto-Encoder [ICCV15]

中科院计算所



IJCV2014



ICCV2015



我们近期的相关工作

- 五个竞赛情况
 - ACM ICMI EmotiW 2014 challenge(冠军)
 - IEEE FG2015 PaSC video FR challenge (冠军)
 - ICCV15 Apparent Age Estimation (亚军)
 - ICCV15 Culture Event Recognition (冠军)
 - 阿里大规模图像搜索大赛 (第一名)
- 深度化的VIPL SDK情况
 - 人脸检测v4.x
 - 特征定位v4.x
 - 人脸识别v4.x



EmotioW 2014: Task

中科院计算所

■ Task

- Classify a sample audio-video clip into one of the seven categories
 - Neutral, anger, disgust, fear, happy, sad, surprise

■ Challenge

- Close-to-real-world conditions
 - Large variations e.g. head pose, illumination, partial occlusion, etc.

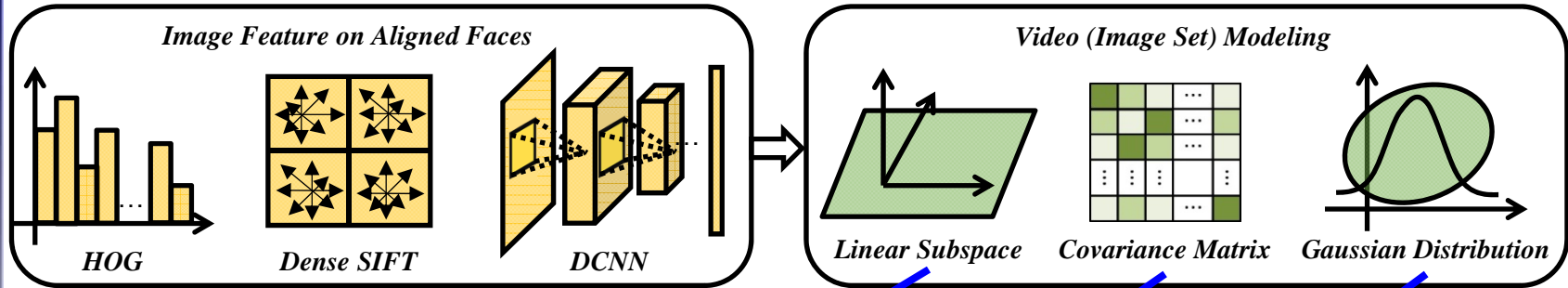




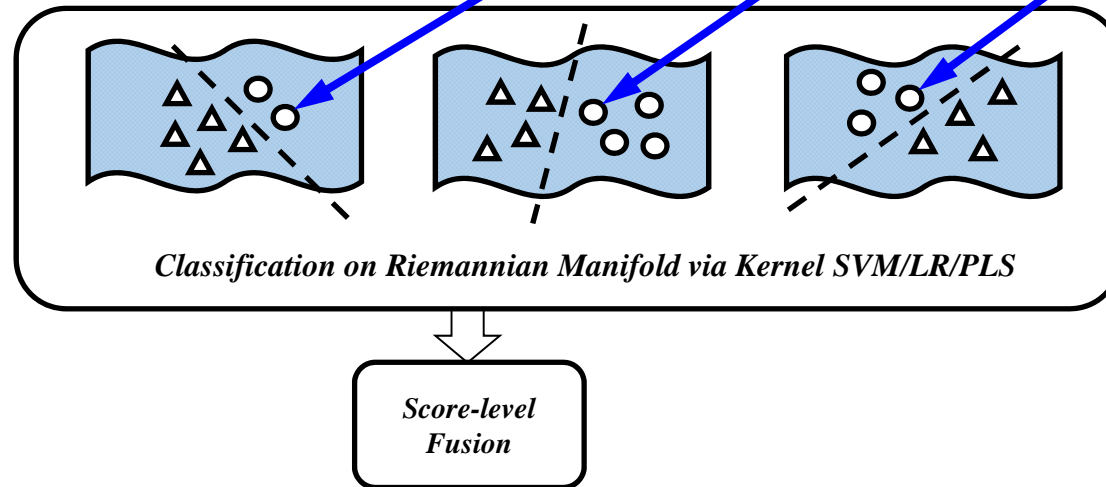
Our method

中科院计算所

Stage 1: Emotion Video Representation



Stage 2: Emotion Video Recognition



M. Liu, R. Wang, S. Li, Z.Huang, S.Shan, X. Chen. Combining Multiple Kernel Methods on Riemannian Manifold for Emotion Recognition in the Wild. ACM ICMI 2014



Our method

- Image features
 - Aligned face images: 64x64; Features: HOG, dense SIFT, DCNN.
- DCNN
 - CaffeNet trained on CFW database
 - Trained over 150,000 face images from 1520 subjects
 - **Identities are served as supervised label in the deep networks**
 - Architecture 3@237x237 > 96@57x57 > 96@28x28 > 256@28x28 > 384@14x14 > 256@14x14 > **256@7x7** > 4096 > 1520
 - Output of the last convolutional layer as final image features: 256x7x7=12, 544
- HOG
 - Block size: 16x16; stride: 8; # of blocks: 7x7=49
 - # of cells per block: 2x2; # of bins: 9; # of total dims: 2x2x9x49=1764
- Dense SIFT
 - Block size: 16x16; stride: 8; # of points: 7x7=49
 - # of dims per point: 4x4x8=128; # of total dims: 128x49=6272



Our Results

中科院计算所

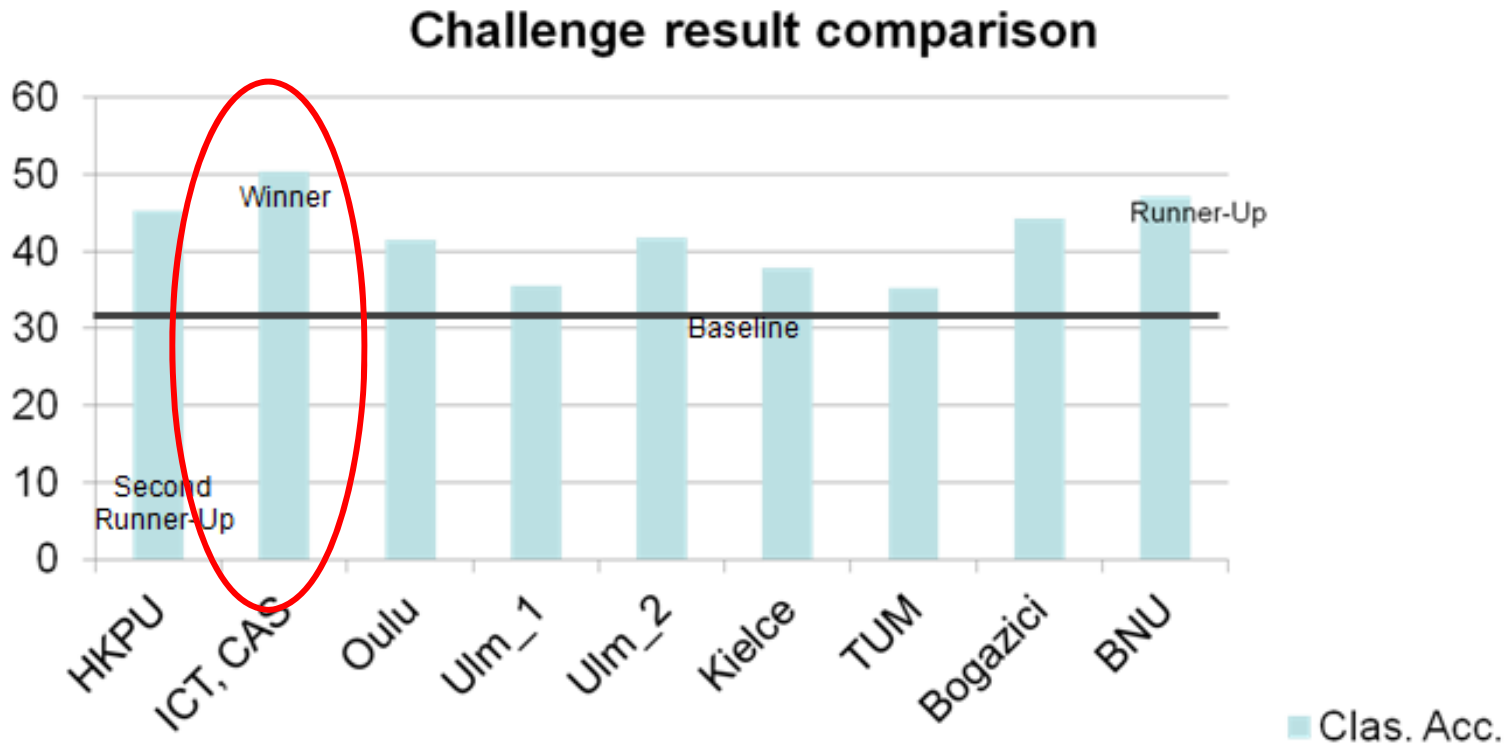
■ Combine multiple features

Methods		Accuracy (%)	
		Validation set	Test set
Baseline (<i>provided by EmotiW organizers</i>)		34.40	33.70
Audio (<i>OpenSMILE Toolkit</i>)		30.73	--
Video	HOG	38.01	--
	Dense SIFT	43.94	--
	DCNN (<i>Caffe-CFW</i>)	43.40	--
	HOG + Dense SIFT	44.47	--
	HOG + Dense SIFT + DCNN (<i>Caffe-CFW</i>)	45.28	--
Audio + Video (<i>HOG+Dense SIFT</i>)		46.36	46.68
Audio+Video (<i>HOG + Dense SIFT + DCNN (Caffe-CFW)</i>)		48.52	50.37



Final Results of Competition

中科院计算所





我们近期的相关工作

- 五个竞赛情况
 - ACM ICMI EmotiW 2014 challenge(冠军)
 - IEEE FG2015 PaSC video FR challenge (冠军)
 - ICCV15 Apparent Age Estimation (亚军)
 - ICCV15 Culture Event Recognition (冠军)
 - 阿里大规模图像搜索大赛 (第一名)
- 深度化的VIPL SDK情况
 - 人脸检测v4.x
 - 特征定位v4.x
 - 人脸识别v4.x



FG 2015 Video FR Challenge

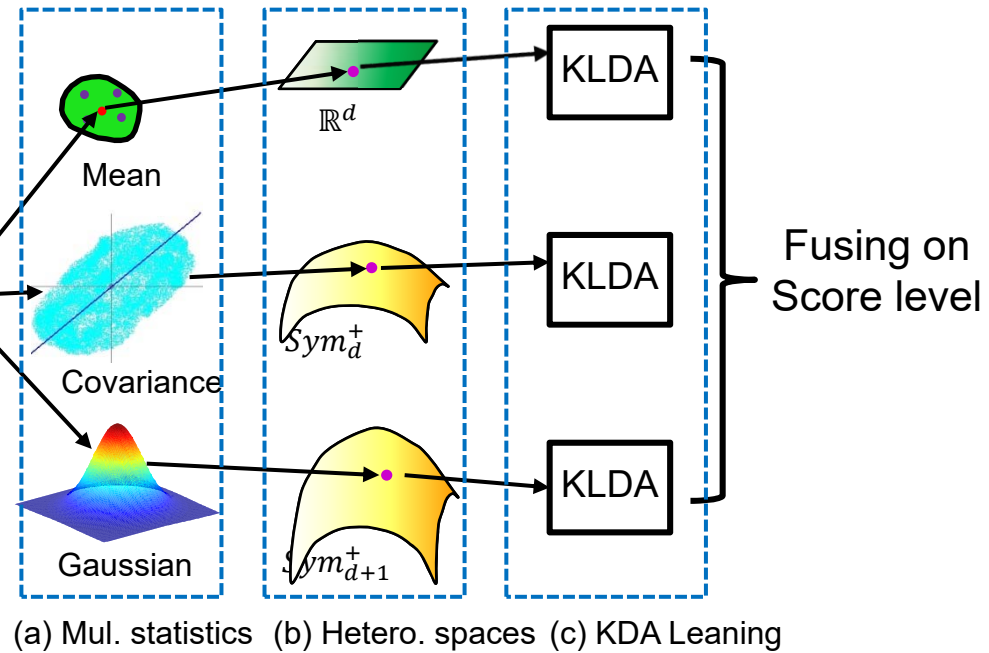
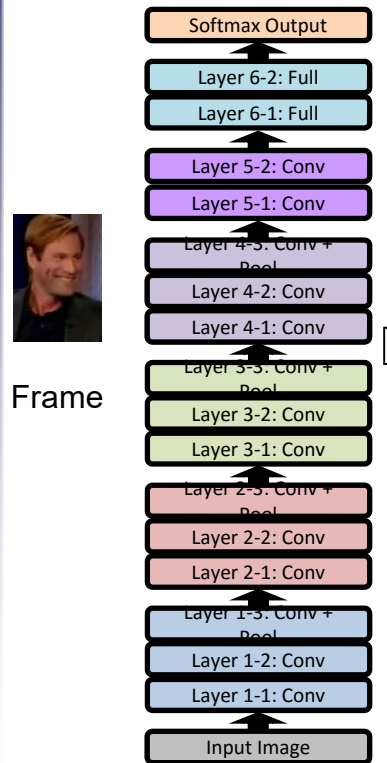
中科院计算所

- 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)



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



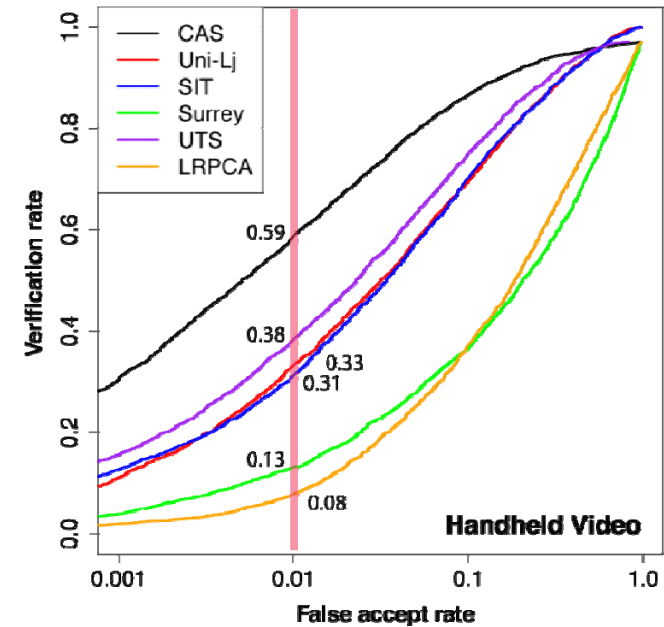
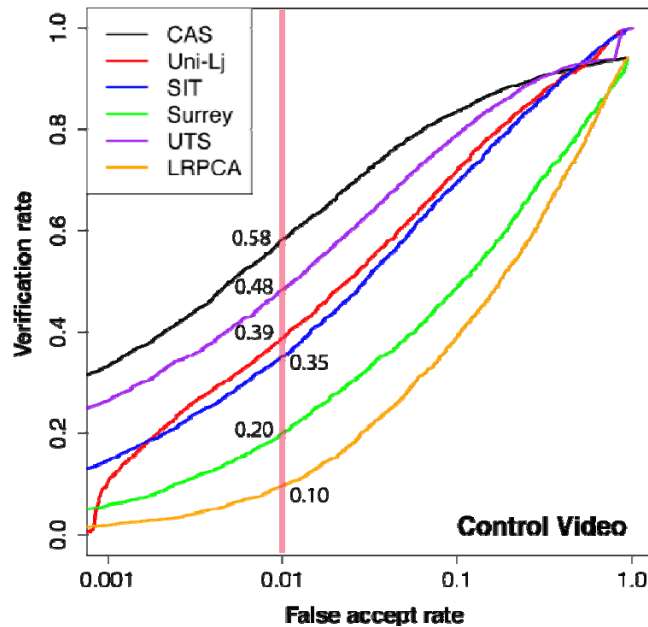
Primary Results

中科院计算所

■ 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





我们近期的相关工作

- 五个竞赛情况
 - ACM ICMI EmotiW 2014 challenge(冠军)
 - IEEE FG2015 PaSC video FR challenge (冠军)
 - **ICCV15 Apparent Age Estimation (亚军)**
 - ICCV15 Culture Event Recognition (冠军)
 - 阿里大规模图像搜索大赛 (第一名)
- 深度化的VIPL SDK情况
 - 人脸检测v4.x
 - 特征定位v4.x
 - 人脸识别v4.x



Apparent Age Estimation

中科院计算所

- ICCV2015 Apparent Age Estimation Challenge介绍
 - 共4699张图像，分为train, validation和test三个子集
 - 开发阶段：train训练，在validation上汇报性能(500次提交机会)
 - 测试阶段：train+val训练，在test上汇报性能(5次提交机会)



Apparent age: 20 Std: 3.99 Apparent age: 22 Std: 2.59 Apparent age: 49 Std:4.39 Apparent age: 20 Std: 4.31



Apparent age: 12 Std: 3.34 Apparent age: 63 Std: 5.10 Apparent age: 29 Std: 5.09 Apparent age: 27 Std: 2.90

数据子集	图像数量
train	2,476
validation	1,136
test	1,087

Apparent Age图像示例：每张图像由10名志愿者标注，取均值作为apparent age，并保存标注的标准差

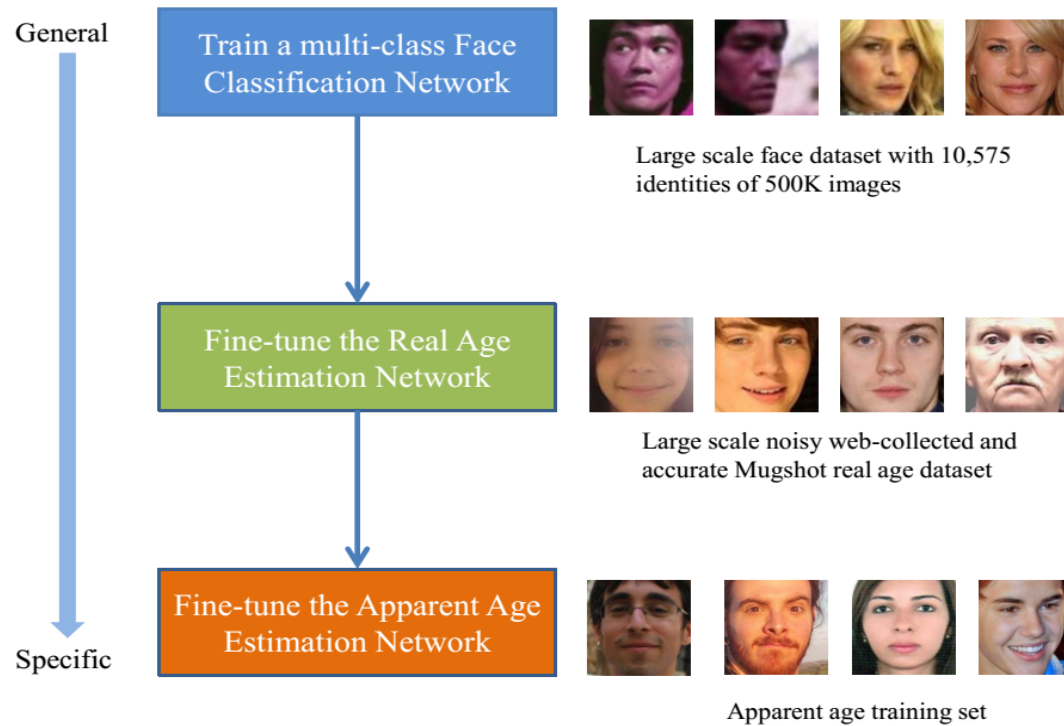


我们的参赛方法：AgeNet

中科院计算所

■ AgeNet中的迁移学习策略

- 人脸识别网络 → 真实年龄估计网络 → 社会年龄估计网络





实验结果

中科院计算所

■ 最终竞赛结果（我们是亚军）

Rank	Team	方法简介	开发阶段	最终测试
1	CVL_ETH	20 VGGNet, 101 reunions softmax normalized output of the last layer	0.295116	0.264975
2	ICT-VIPL	8 GooletLet, AgeNet Age Regressor and Classifier	0.292297	0.270685
3	Ageseer	VGGNet, Prediction of age codes, fusion of regressors, such as lasso, global and local quadratic regressor, and random forest	0.327321	0.287266
3	WVU_CVL (郭国栋组)	GoogleLet + 10 age groups, RF, SVR, and fusion	0.316289	0.294835
4	SEU_NJU (耿新组)	VGGNet, Fusion of different network setups, softmax loss and KL-divergence for training.	0.380615	0.305763



我们近期的相关工作

- 五个竞赛情况
 - ACM ICMI EmotiW 2014 challenge(冠军)
 - IEEE FG2015 PaSC video FR challenge (冠军)
 - ICCV15 Apparent Age Estimation (亚军)
 - **ICCV15 Culture Event Recognition (冠军)**
 - 阿里大规模图像搜索大赛 (第一名)
- 深度化的VIPL SDK情况
 - 人脸检测v4.x
 - 特征定位v4.x
 - 人脸识别v4.x



问题概要

中科院计算所

■ 社会文化事件图像示例

- 从上至下依次为：美国独立日，环法自行车赛，热气球嘉年华，赶牛艺术节，西班牙西红柿节





实验结果(1/6)

■ 实验数据与协议

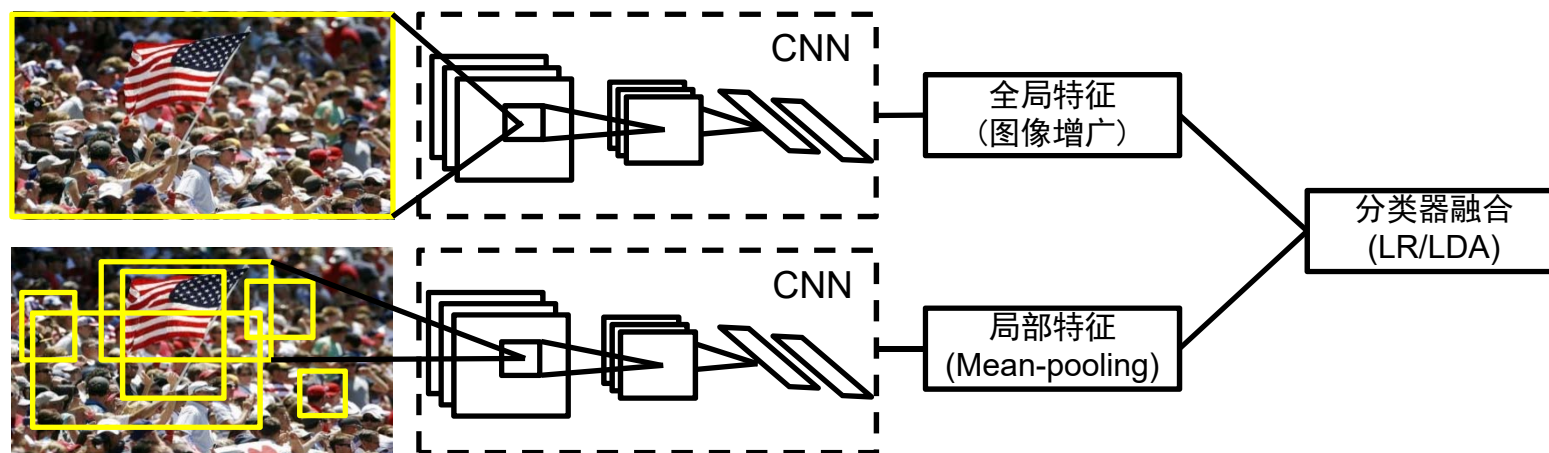
- 数据集规模：共计28,705图像，100类事件
 - 开发阶段：训练集14,332图，校验集5,704图
 - 测试阶段：测试集8,669图（只能提交5次结果）
 - 数据分布基本一致
- 测试指标：mAP(mean Average Precision)
 - 对于每类事件根据每个样本得到的预测分值绘制准确率-召回率(precision-recall)曲线
 - 计算曲线下区域面积即AP(Average precision)
 - 对所有事件类别计算AP的均值，即mAP



我们的参赛算法

中科院计算所

- 全局图像特征提取
 - 缩放图像以保持短边长为256像素
 - 按上中下或左中右采样三张256x256图像
- 局部目标特征提取
 - 基于Selective Search每张图像采样若干局部区域
 - 滤除噪声区域
 - 长或宽小于原图20%的区域
 - (2)长宽比>2的区域
 - 缩放所有局部区域至256x256





我们的参赛算法

■ 模型训练

□ VGGNet

- 16/19层深度网络结构，单一softmax loss层(4096维)
- 基于ImageNet数据预训练，Event数据fine-tuning
 - 全局策略：42,996样本，迭代30K次；
 - 局部策略：1,794,988样本，迭代120K次

□ GoogLeNet

- 22层深度网络结构，三个softmax loss层(1024维)
- 基于ImageNet数据预训练
- Event数据fine-tuning
 - 全局策略：42,996样本，迭代40K次；
 - 局部策略：1,794,988样本，迭代100K次



竞赛成绩

中科院计算所

■ 最终测试集结果比较（我们是冠军）

排名	参赛队伍	方法简介	开发阶段	最终测试
1	VIPL-ICT-CAS	Region fine-tuned VGGNet +GoogLeNet; LR+LDA	0.783	0.854
2	FV	VGGNet+Places-CNN; LR	0.770	0.851
3	MMLAB	VGGNet+GoogLeNet; FV; SVM	0.717	0.847
4	NU&C	CaffeNet; RAW+LDA; iterNN	--	0.824
5	CVL_ETHZ	Pre-trained VGGNet on Places and ImageNet	0.662	0.798
6	SSTK	13 pre-trained VGGNet	0.740	0.770
7	MIPAL_SNU	Person/face/region GoogLeNet;	0.801	0.763



我们近期的相关工作

- 五个竞赛情况
 - ACM ICMI EmotiW 2014 challenge(冠军)
 - IEEE FG2015 PaSC video FR challenge (冠军)
 - ICCV15 Apparent Age Estimation (亚军)
 - ICCV15 Culture Event Recognition (冠军)
 - 阿里大规模图像搜索大赛 (第一名)
- 深度化的VIPL SDK情况
 - 人脸检测v4.x
 - 特征定位v4.x
 - 人脸识别v4.x



问题描述(3/3)

中科院计算所

query



1st ✓



2nd ✓



3rd ✓



4th ✓



5th ✓



6th ✓



7th ✓



8th ✓



9th ✗



10th ✓



...

19th ✓



20th ✓



...

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

- 我们的特色
 - 基于自步学习的自动同款挖掘
 - 有噪声情况下的DCNN学习
 - 快速LDA作为后端算法



我们的方法——自动同款挖掘

中科院计算所

■ Motivation

- Step2: Discovering more using the “seeds”
- Kicking out “noise” and “outliers”

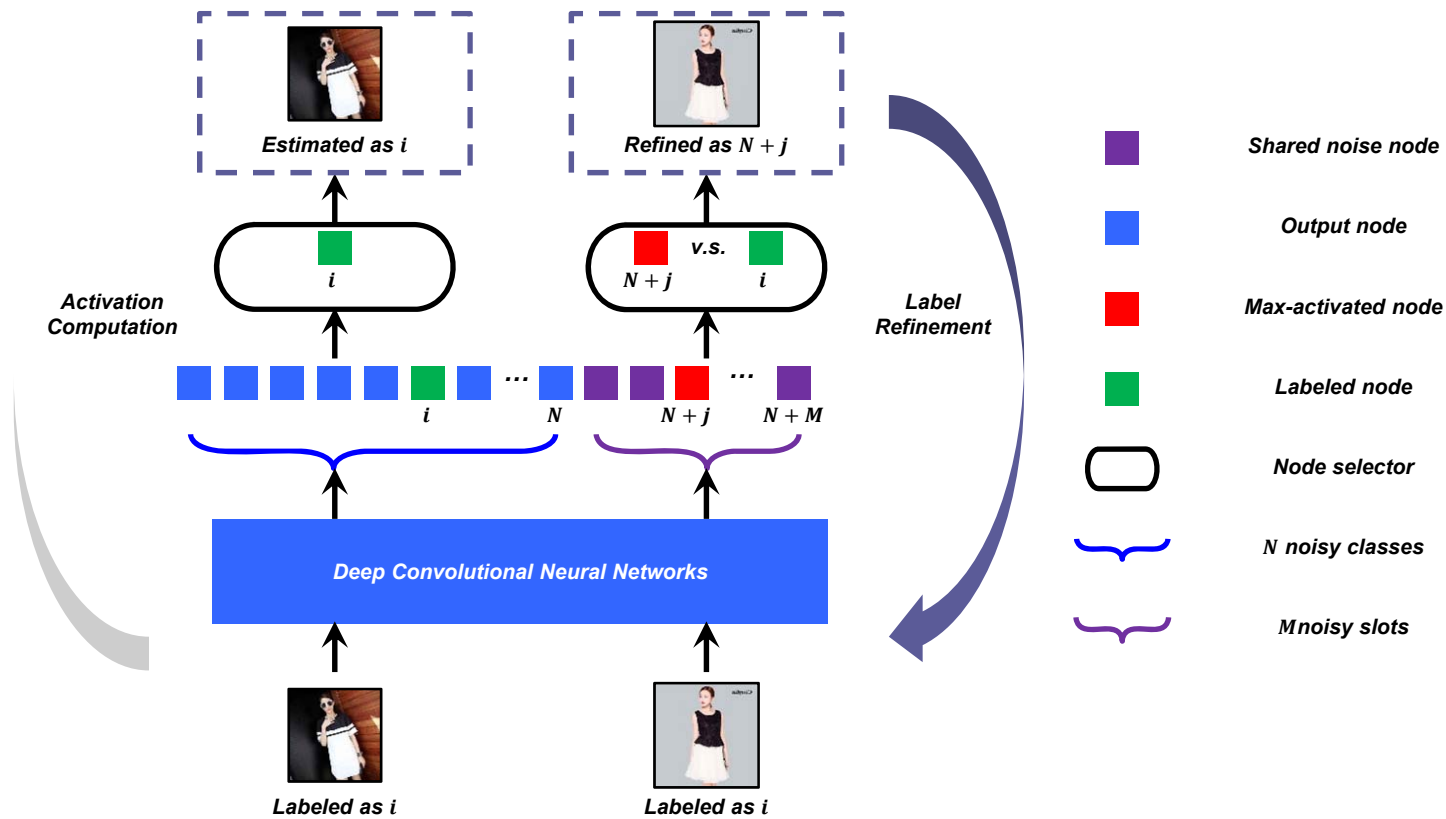




我们的方法——自纠错DCNN

中科院计算所

- Simulate human “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%
Benchmark: MNIST					
LeNet	99.3	96.8	84.2	56.0	20.6
LeNet + Proposed Method	99.3	99.2	98.9	98.6	98.4
Benchmark: CIFAR-10					
LeNet	64.5	47.9	34.5	23.0	15.5
LeNet + Proposed Method	63.6	59.9	52.9	49.9	34.6



Experimental results

中科院计算所

Performance on validation and test data.

	Method						mAP	
	Same style mining		DCNN fine-tuning		LDA training		Val	Test
	Similarity	SPL-eSVM	Normal	Noise-robust	Standard	Improved		
1	✓	✗	✓	✗	✗	✗	0.26	--
2	✓	✗	✗	✓	✗	✗	0.32	--
3	✗	✓	✓	✗	✗	✗	0.31	0.38
4	✗	✓	✗	✓	✗	✗	0.38	0.44
5	✗	✓	✗	✓	✓	✗	<u>0.42</u>	0.48
6	✗	✓	✗	✓	✗	✓	0.40	0.4929



我们近期的相关工作

■ 五个竞赛情况

- ACM ICMI EmotiW 2014 challenge(冠军)
- IEEE FG2015 PaSC video FR challenge (冠军)
- ICCV15 Apparent Age Estimation (亚军)
- ICCV15 Culture Event Recognition (冠军)
- 阿里大规模图像搜索大赛 (第一名)

■ 深度化的VIPL SDK情况

- 人脸检测v4.x; 特征定位v4.x; 人脸识别v4.x
- Demos



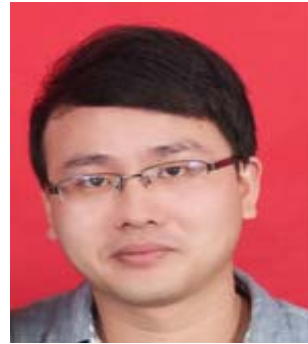
Thanks, Q & A

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences



Jie Zhang



Zhiwu Huang



Yan Li



Mengyi Liu



Xin Liu



Shaoxin Li



Meina Kan



Ruiping Wang



Shiguang Shan



Xilin Chen

Lab of Visual Information Processing and Learning (VIPL) @ICT@CAS

Academic codes available at: <http://vipl.ict.ac.cn/resources/codes>