

深度人脸感知研究

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



VALUE及其在线Webinar活动

中科院计算所

- VALUE: **V**ision **A**nd **L**earning **S**eminar(视觉与学习研讨会)
- VALUE年度大会
 - **不投稿, 免注册费**, 每年四月末举行, 今年第5届, 共**700**人报名参会
 - 邀请大会报告, 顶会牛刊论文Poster展示, Tutorial, Panel讨论...
- VALUE QQ群(共**4800+**人的社区)
 - VALUE A群: 将满(2000人)——不再接受入群申请
 - VALUE B群: 已满(2000人)——不再接受入群申请
 - VALUE C群: 目前900+人, 还可以申请加入, 群号**480822477**
- VALUE每周在线Webinar活动(依托QQ群视频功能)
 - **每周三晚上有VALUE在线Webinar报告会**(有网络上Win QQ即可)
 - 活动形式: 佳作讲解, Panel讨论, 研究组巡讲, 热点技术...
 - 高峰在线参加人数**640+**人, 迄今举办**50余场**, 报告/讨论嘉宾**100+**人
 - 颜水成, 王晓刚, 屠卓文, 沈春华, 吴建鑫, 凌海滨, 杨铭, 王井东, 王乃岩, 张弛原, 华刚, 鲁继文等...(1/3来自海外或港澳地区)
- 关于VALUE的更多详情: <http://valser.org>



中科院计算所

精彩直播推荐 群视频-VALSE(364188996) 反馈 管理

主席模式 当前状态: 仅管理员可以上台

仅台上的人和管理员可以说话

群视频成员(250)

- 山世光T计算所
- 贾伟T中科院
- 朱军-清华
- 白翔T华科
- 张磊-I-微软
- 吕乐
- 张兆翔T自动化所
- 孟德宇T西交
- 董乐T电子科大
- 禹之鼎D-CMU
- 程明明T南开
- 邓亚峰I-百度
- 王振D吉林大学

Thinking as a Researcher
and Practicing as an Engineer

Lei Zhang

Microsoft Research
November 26, 2015

库存: x0 购买

送给: 张磊-I-微软(30803... 数量: 9 赠送

自由说话 播放伴奏 上台

99% 21:14



VALUE及其在线Webinar活动

■ VALUE主要组织者

□ 联合发起人

- 山世光, 潘纲, 刘青山, 颜水成, 李学龙

□ 指导委员会 (拼音序)

- 白翔, 高新波, 何晓飞, 黄华, 李学龙, 刘青山, 马毅, 潘纲, 山世光, 徐东, 颜水成, 杨健, 余凯, 周昆, 周志华

■ VALUE Online主要组织者

□ 主要发起人

- 山世光, 白翔, 程明明, 孟德宇, 彭玺, 禹之鼎, 贾伟

□ 首届VALUE Online理事会 (拼音序)

- 白翔, 程明明, 邓成, 董乐, 董伟生, 纪荣嵘, 贾伟, 林惊, 鲁继文, 孟德宇, 彭玺, 山世光, 王乃岩, 王瑞平, 禹之鼎, 章国锋, 张利军, 张兆翔, 郑海永, 郑伟诗, 左旺孟



提纲

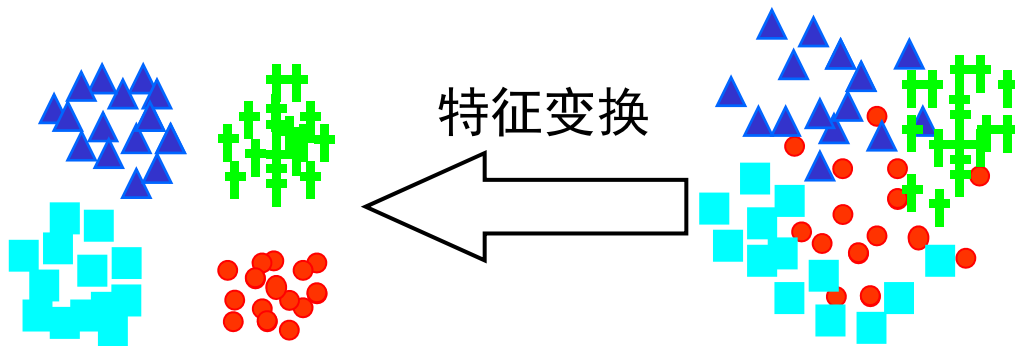
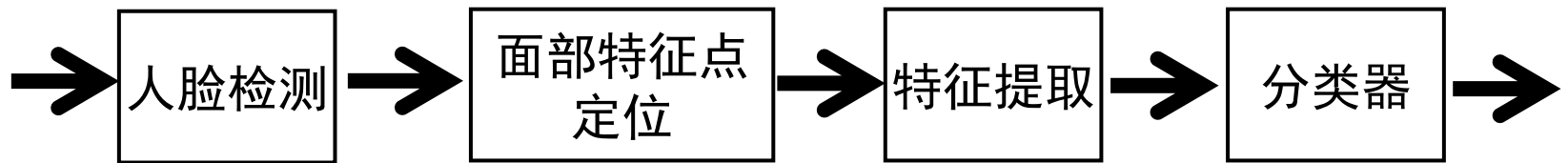
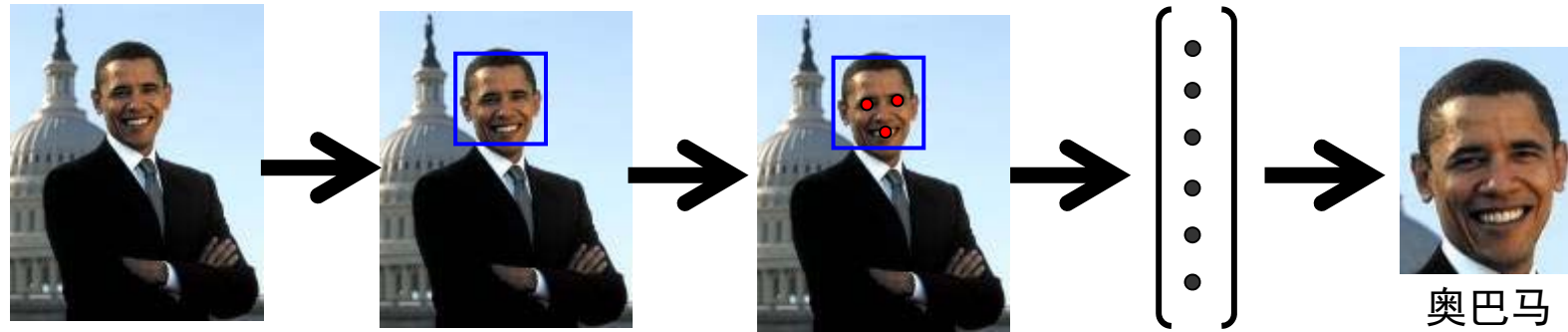
中科院计算所

- 人脸识别历史与近期进展
- 我们近期的部分工作



人脸识别的一般技术流程(DL前)

中科院计算所

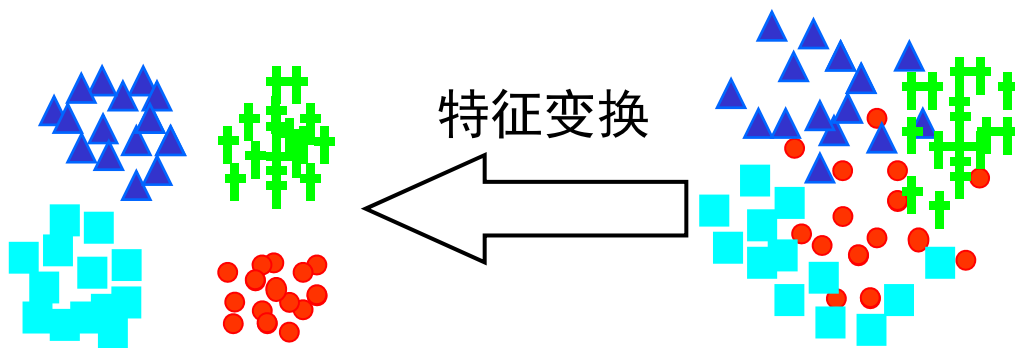
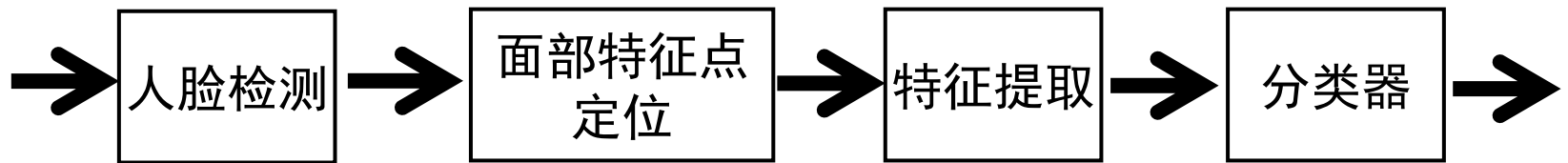
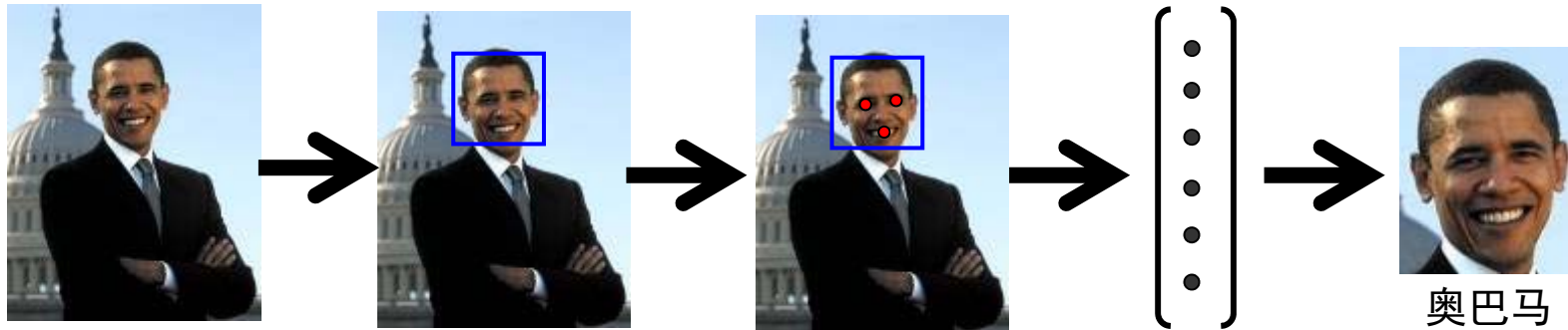


两步法:
1. 人工局部(卷积)特征
Gabor, LBP, SIFT, ...
2. 特征变换
PCA, LDA, KDA, ML...



人脸识别的一般技术流程(DL时代)

中科院计算所



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



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]

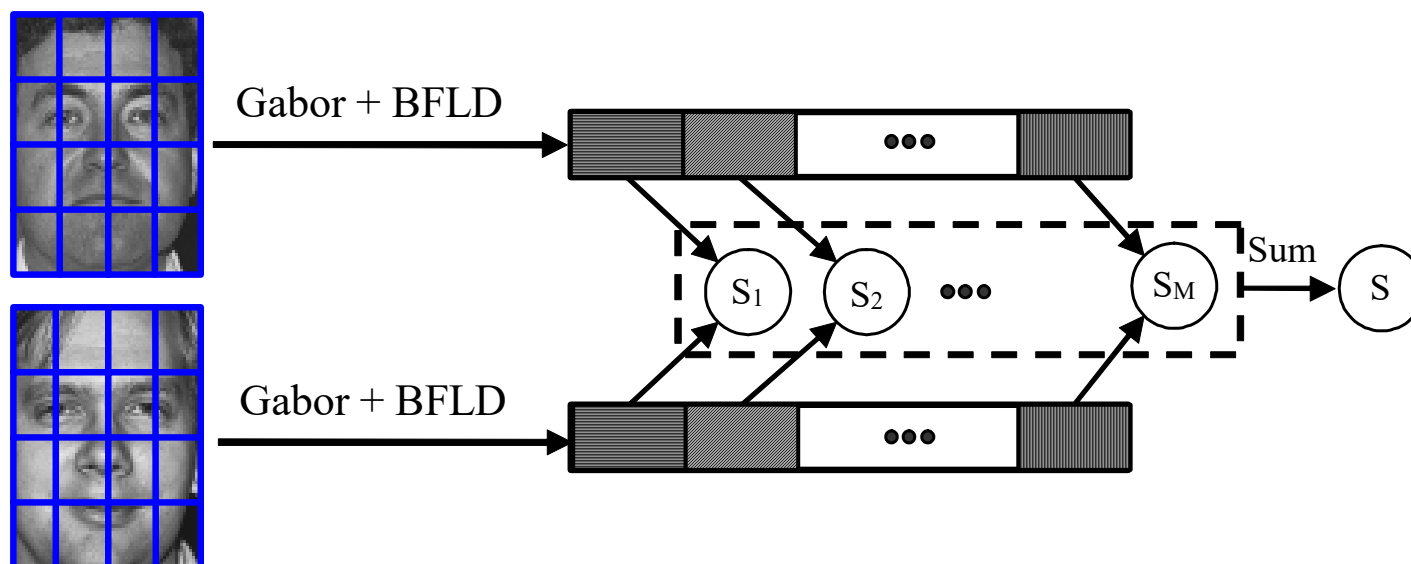




从测试库变迁看人脸识别

■ 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%

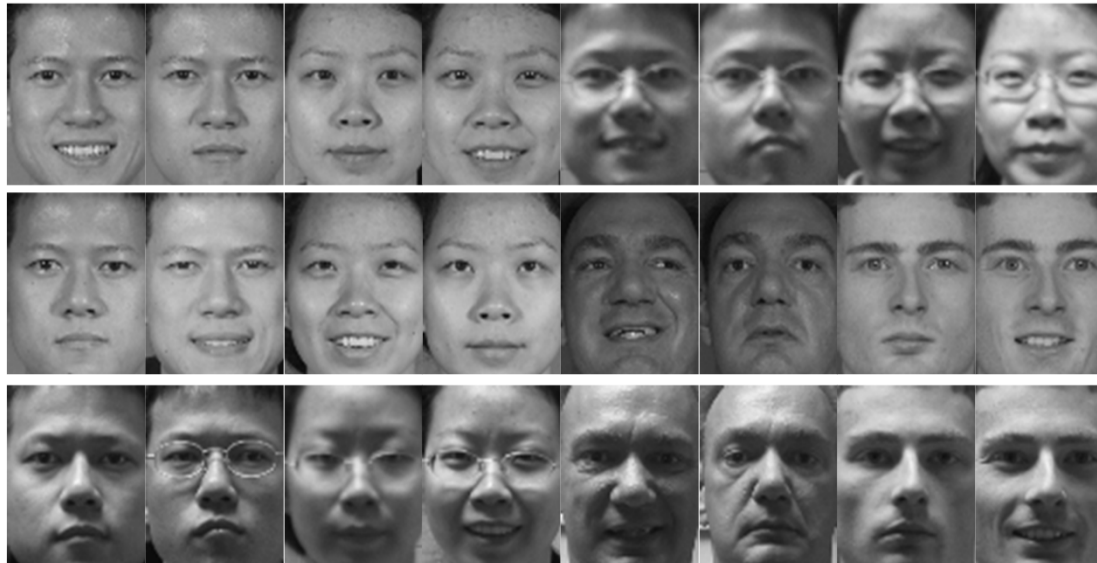
1. Claudio A.Perez , LeonardoA.Cament, LuisE.Castillo. Methodological improvement on local Gabor face recognition based on feature selection and enhanced Borda count. Pattern Recognition 44 (PR2011), 951–963
2. Georgios Tzimiropoulos, Stefanos Zafeiriou, Maja Pantic. Subspace Learning from Image Gradient 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

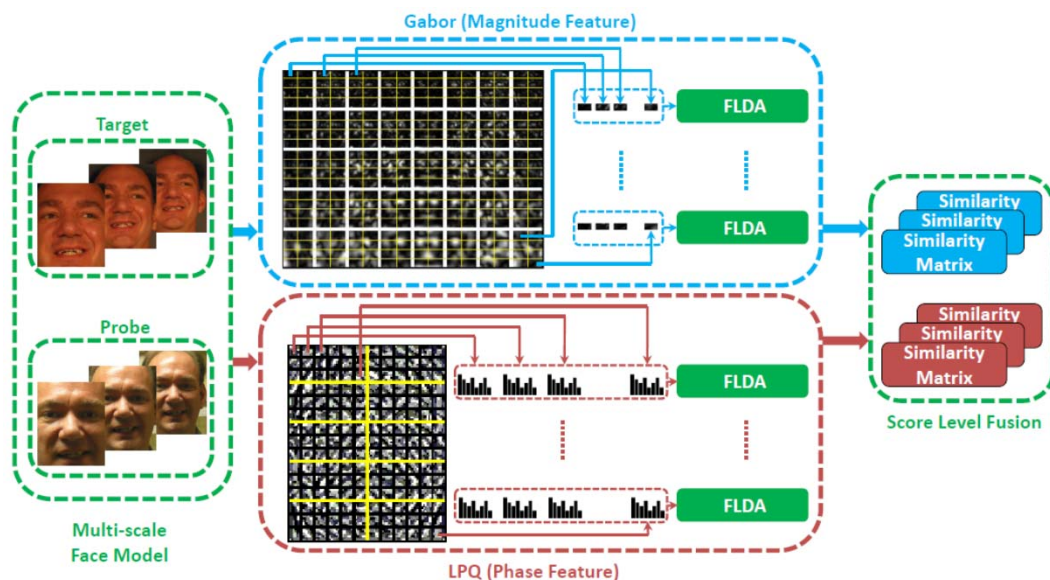
中科院计算所

- 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

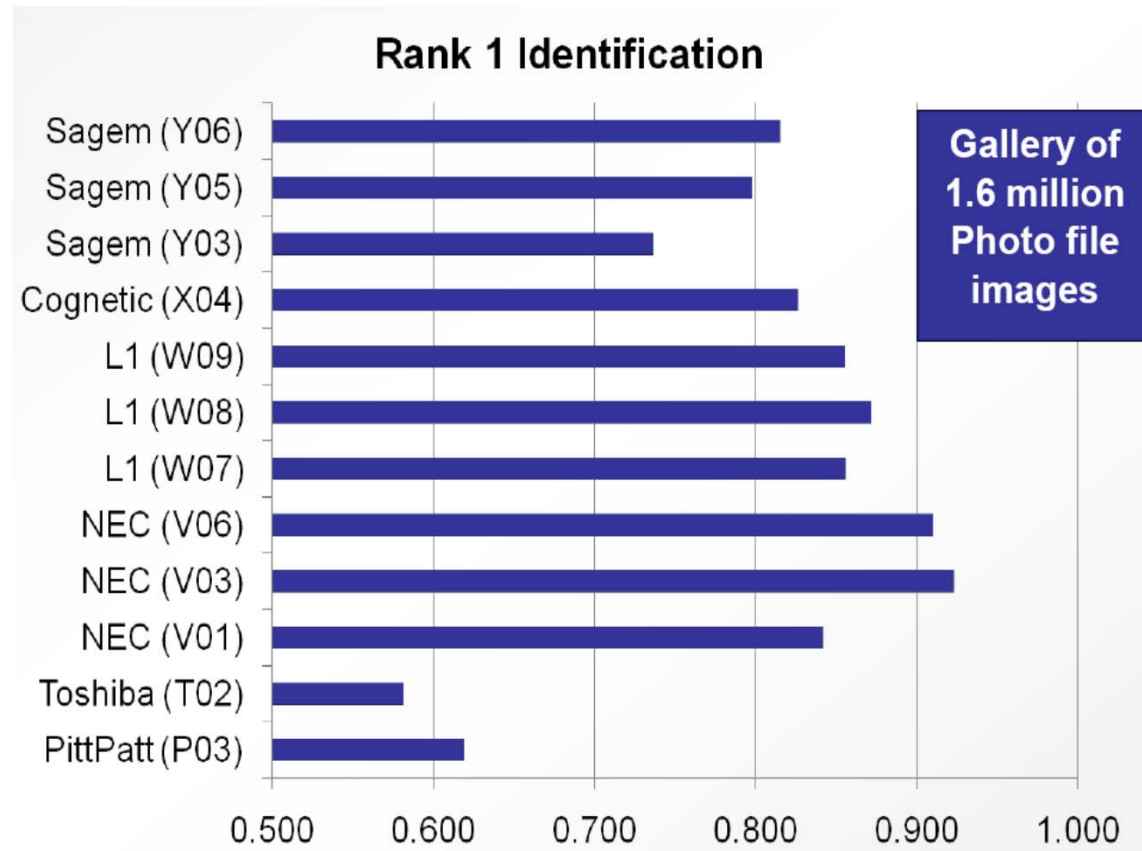
[Li 12] Y.Li, S. Shan, H. Zhang, S. Lao, X. Chen. Fusing Magnitude and Phase Features for Robust Face Recognition, ACCV2012



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%



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1%
(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



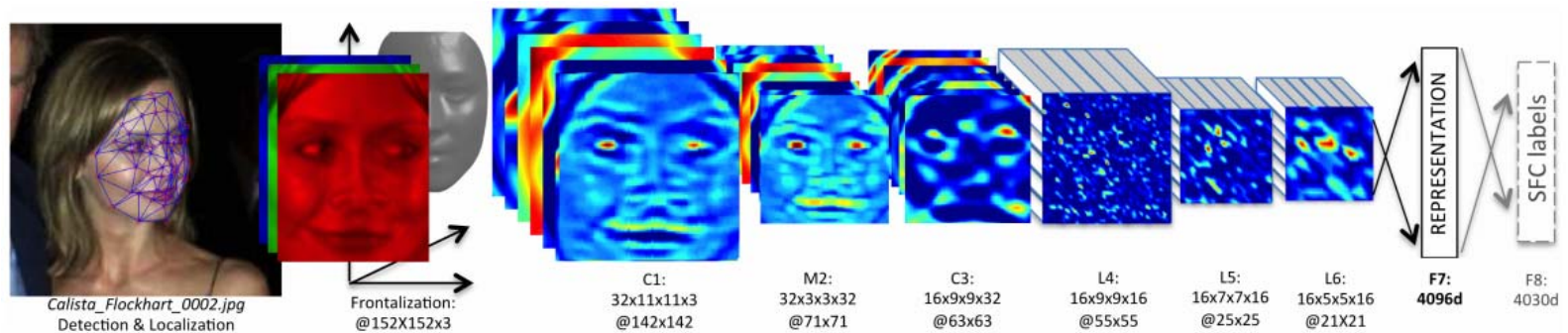
Huang G B, Ramesh M, Berg T, et al. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report, University of Massachusetts, Amherst, 2007.



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

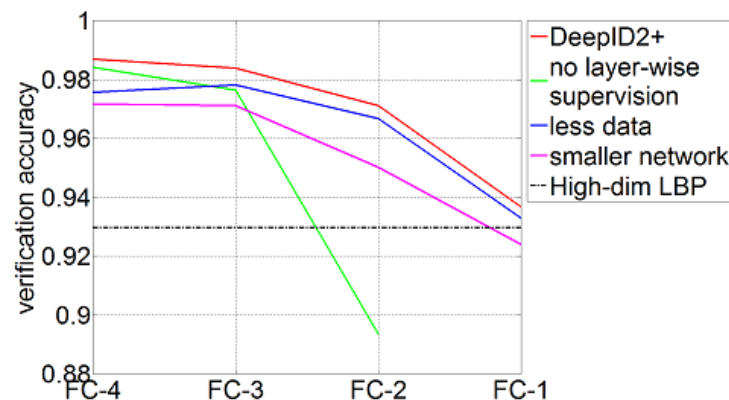
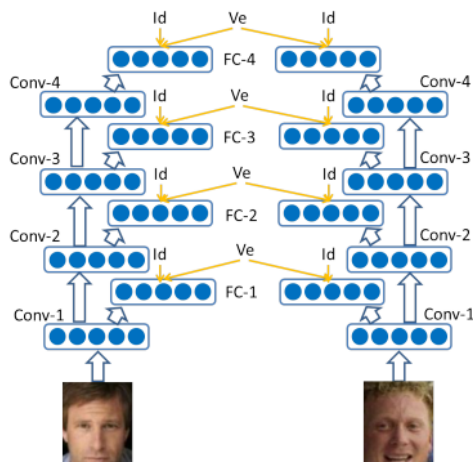
[1] Taigman Y, Yang M, Ranzato M A, et al. Deepface: Closing the gap to human-level performance in face verification. CVPR, 2014.



DeepID2+

中科院计算所

- 在25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



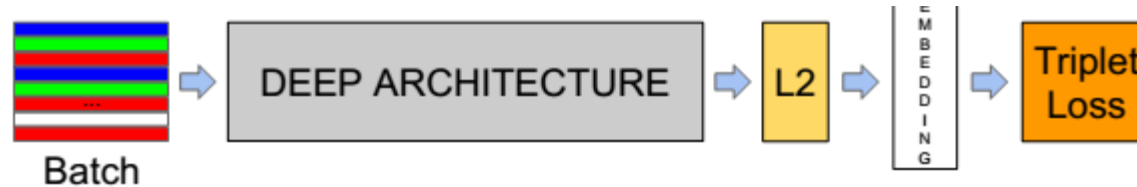
Sun Y, Wang X, Tang X. Deeply learned face representations are sparse, selective, and robust. arXiv preprint, 2014.



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



问题与挑战

■ 学术界

- 没有真正有效的、接近实战的公共评测库
 - 对实战应用缺乏了解，数据显著少于工业界
- 挑战
 - 超越工业界的理论进步；与工业界分工合作

■ 工业界

- 雨后春笋，热火朝天，江湖沸腾
 - 有误导大众和媒体的现象
 - 对学术界缺乏尊重和回馈
 - 抢人自己做
- 挑战：客户信任，理性竞争，



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



关于DL的一些讨论

- DL带来观念的变革
- DL是类脑信息处理方法吗？
- DL有理论吗？
- DL不能做什么？
- 数据驱动的学习不再需要领域知识？
- 工业界抢了学术界的饭碗？
- CV研究者沦为ML研究者的实验员？
- DL未来工作？



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有理论吗？

- DL理论匮乏
 - 收敛性, bound
 - 局部极值, 初值很重要
 - 复杂度理论
- 但不完全是black box
 - 与传统“分步”做法的关系
 - 比Kernel更“显式”
 - 层级可视化提供了很多线索
 - 逐层抽象 or 分层“非线性”？



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?



工业界抢了学术界的饭碗？

- 工业界看不起学术界？
 - 工业界重视大数据收集和并行实现
 - 学术界重视理论和新的模型
- CV学术界应该更smart更前瞻
 - 新的网络设计
 - 网络结构学习
 - 新的优化方法
 - 算法层面的加速学习
 - 更Smart的数据收集
 - 大而脏乱差标注数据的高效利用
 - 大规模无监督数据的利用



CV研究者沦为ML研究者的实验员

- 有这个危险
 - CV本身缺少理论体系
 - “分步”法主宰CV太多年
- CV仍有机会
 - 实际上ML也从CV获益良多
 - 基于学习的CV理论?
 - CV研究者应该与ML有更多互动
 - “几何”和“结构”的可学习性?



DL领域的未来工作

- DL的理论
- 网络结构本身的学习
- 小数据条件下的DL
- 领域知识的嵌入
- 带反馈的深度网络
- 大而脏乱差数据条件下的DL
- 无监督条件下的DL
- 深度模型的迁移与适应
- 面向视频分析的DL模型
- “非线性”的更多来源
- 新的优化方法和学习策略



What's missing from DL?

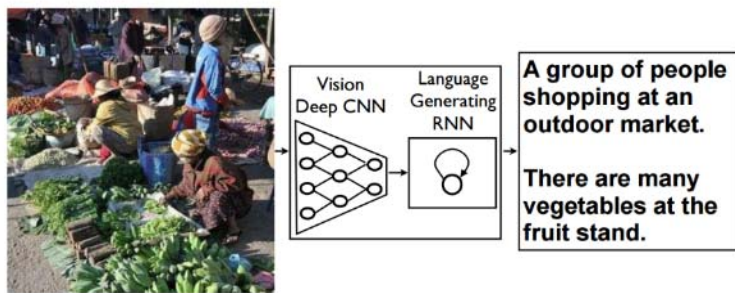
- Yann LeCun's Points
 - Theory
 - Reasoning, structured prediction
 - Memory, short-term/working/episodic memory
 - Unsupervised learning that actually works



热点与趋势

中科院计算所

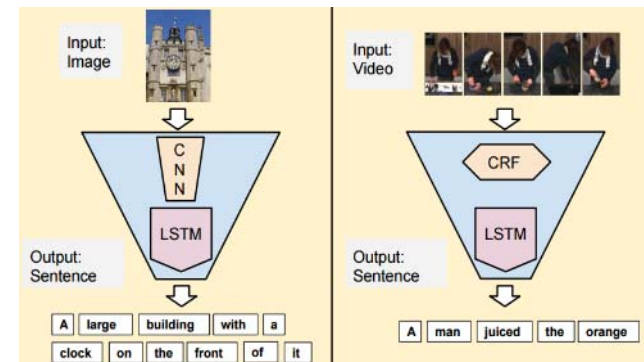
- Deep Learning for Vision in CVPR2015
 - 仅标题中出现Deep的就有87篇，出现Convolution的47篇，出现Neural的40篇，出现Network的51篇，Recurrent 7篇
 - Going deeper, 优化...
 - Fully Convolutional Network (for segmentation等)
 - Vision and Language(for看图说话, Google, Fei-fei, Microsoft, UCB)
 - RNN with LSTM (for 时序处理)
 - Deep Learning for XXX



Show and Tell: A Neural Image Caption Generator (a work from Google)



From Captions to Visual Concepts and Back (a work from Microsoft)



Long-term Recurrent Convolutional Networks for Visual Recognition and Description(a work from UTA/UML/USB)



热点与趋势

- Deep Learning for Vision in CVPR2015
 - 仅标题中出现Deep的就有87篇，出现Convolution的47篇，出现Neural的40篇，出现Network的51篇，Recurrent 7篇
 - Going deeper, 优化...
 - Fully Convolutional Network (for segmentation等)
 - Vision and Language(for看图说话, Google, Fei-fei, Microsoft, UCB)
 - RNN with LSTM (for 时序处理)
 - Deep Learning for XXX
 - Image segmentation
 - Face recognition; Human Re-ID;
 - Object detection; Saliency detection
 - Metric learning; Transfer learning; Attribute learning; Hashing;
 - 3D ShapeNet; Depth estimation; stereo matching;
 - Crowd counting; weather prediction;
 - Action recognition, Even recognition;



我们近期的部分工作

中科院计算所

- 四个竞赛情况
 - ACM ICMI EmotioW 2014 challenge
 - IEEE FG2015 PaSC video FR challenge
 - ICCV15 ChaLearn Apparent Age Estimation
 - ICCV15 ChaLearn Culture Event Recognition
- VIPL SDK情况
 - 人脸检测v4.x, 特征定位v4.x, 人脸识别v4.x
- 三个基于小数据的DL学习方法
 - DAE for face alignment [ECCV2014]
 - DAE for pose normalization [CVPR2014]
 - Bi-shift AE for Domain Adaptation [ICCV2015]



我们在DL上的工作

■ 四个竞赛情况

- ACM ICMI EmotioW 2014 challenge
- IEEE FG2015 PaSC video FR challenge
- ICCV15 ChaLearn Apparent Age Estimation
- ICCV15 ChaLearn Culture Event Recognition

■ VIPL SDK情况

- 人脸检测v4.x, 特征定位v4.x, 人脸识别v4.x

■ 三个基于小数据的DL学习方法

- DAE for face alignment [ECCV2014]
- DAE for pose normalization [CVPR2014]
- Bi-shift AE for Domain Adaptation [ICCV2015]

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



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.





EmotioW 2014: Data

中科院计算所

- Challenging data
 - AFEW* 4.0 database
 - audio-video clips collected from movies showing close-to-real-world conditions

Attribute of AFEW 4.0	Description
Length of sequences	300-5400ms
Number of annotators	3
Emotion categories	Anger, disgust, fear, happiness, neutral, sadness, and surprise
Audio/Video format	Audio: WAV; Video: AVI
# of samples	1368
# of subjects	428
# of movies	111

*Acted Facial Expression in Wild



EmotioW 2014: Protocols

中科院计算所

- Evaluation protocols
 - Dataset division: training, validation, and testing
 - The test labels were unknown.
 - Either audio/video modality or both can be used.

Set	# of subjects	Min. Age	Max. Age	Avg. Age	# of Males	# of Females
Train	177	5	76	34	102	75
Val	136	10	70	35	78	58
Test	115	5	88	34	64	51

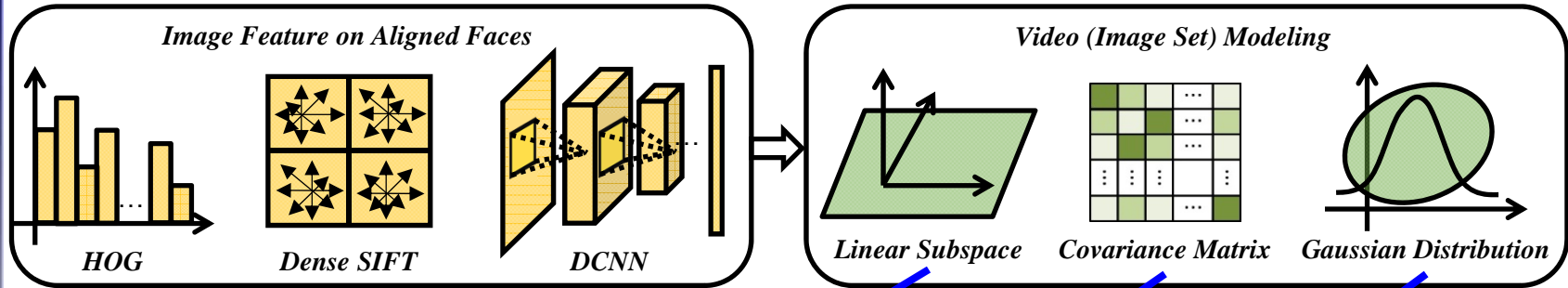
	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Train	92	66	66	105	102	82	54
Val	59	39	44	63	61	59	46
Test	58	26	46	81	117	53	26



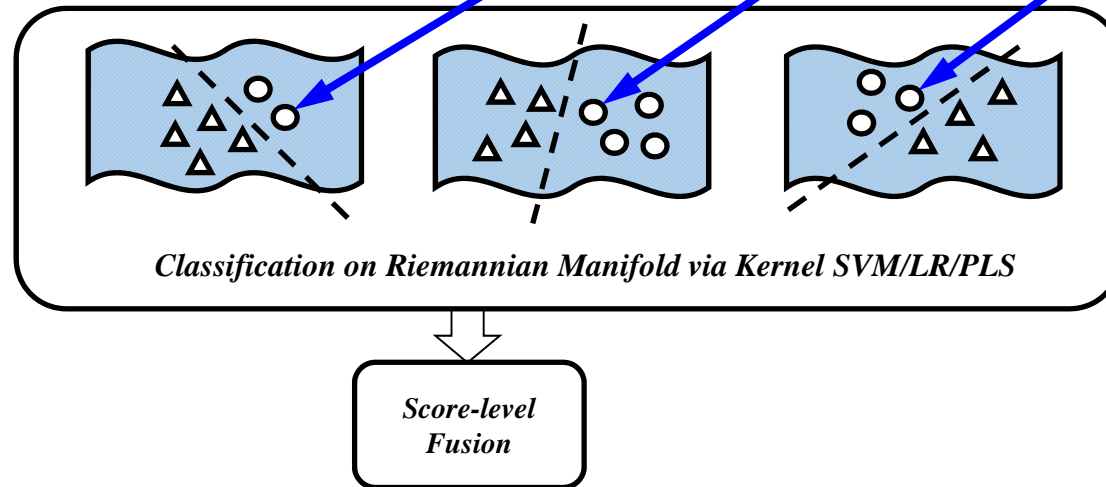
Our method

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

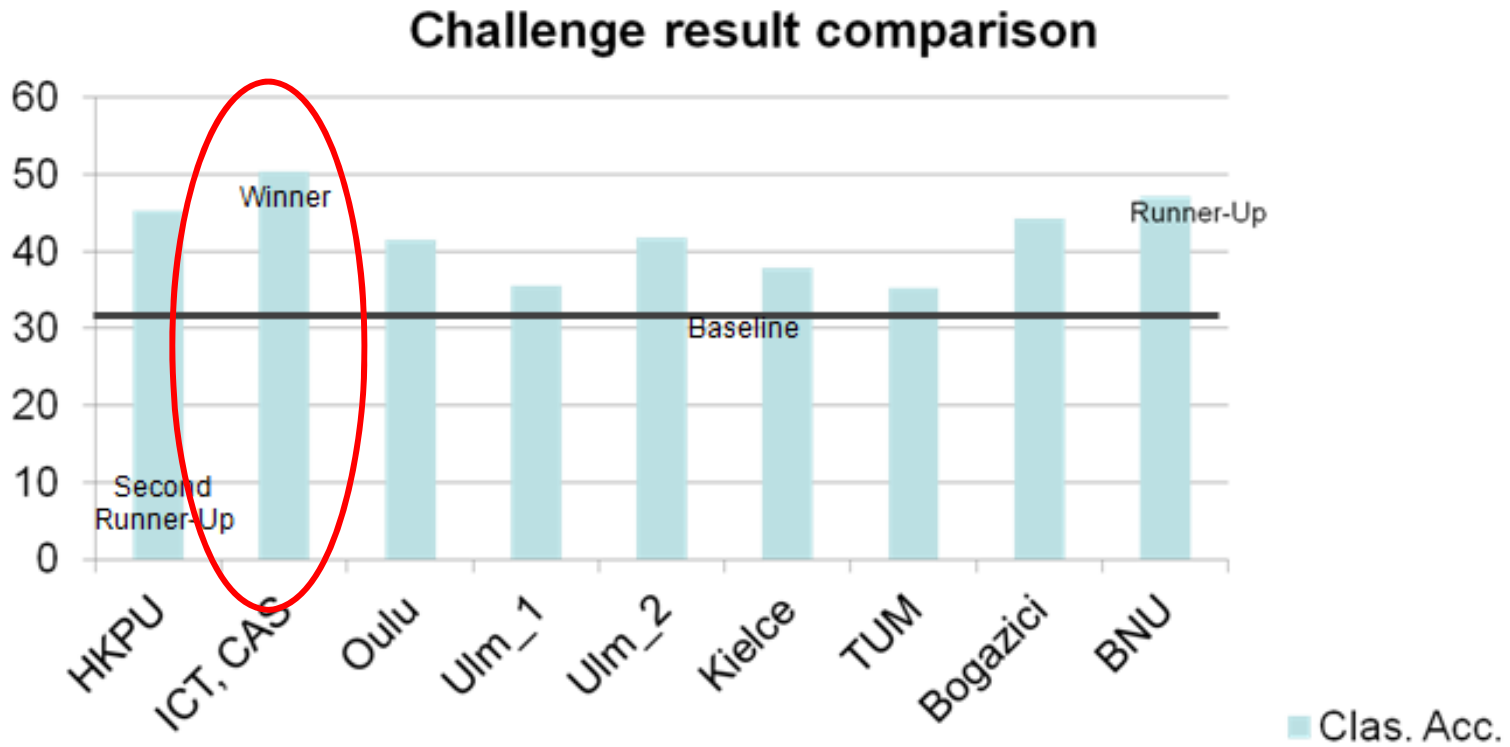
■ Combine multiple features

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





我们在DL上的工作

■ 四个竞赛情况

- ACM ICMI EmotioW 2014 challenge
- **IEEE FG2015 PaSC video FR challenge**
- ICCV15 ChaLearn Apparent Age Estimation
- ICCV15 ChaLearn Culture Event Recognition

■ VIPL SDK情况

- 人脸检测v4.x, 特征定位v4.x, 人脸识别v4.x

■ 三个基于小数据的DL学习方法

- DAE for face alignment [ECCV2014]
- DAE for pose normalization [CVPR2014]
- Bi-shift AE for Domain Adaptation [ICCV2015]



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





FG 2015 Video FR Challenge

中科院计算所

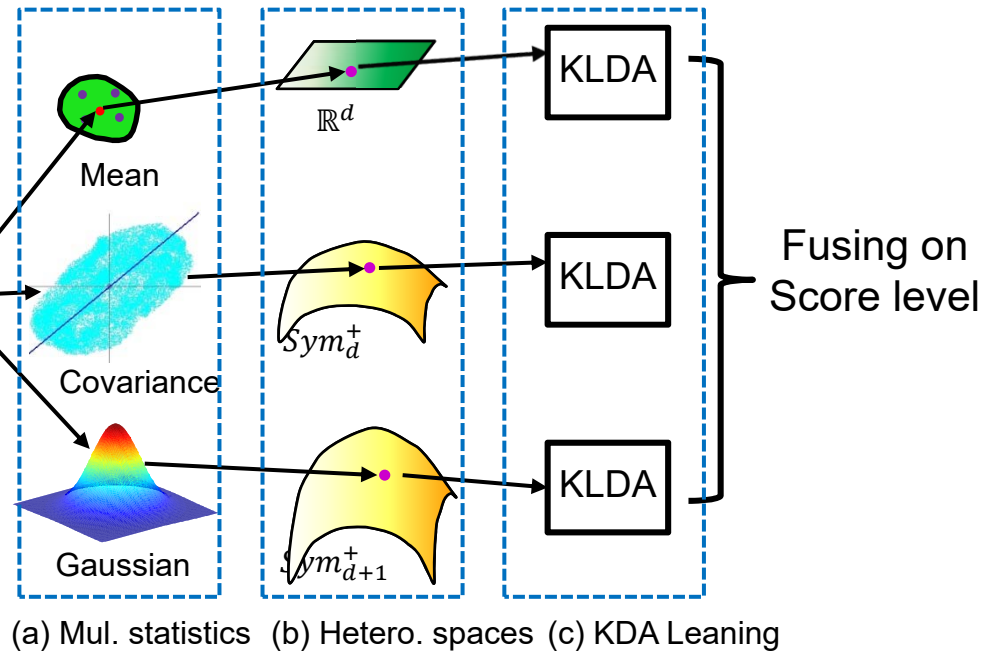
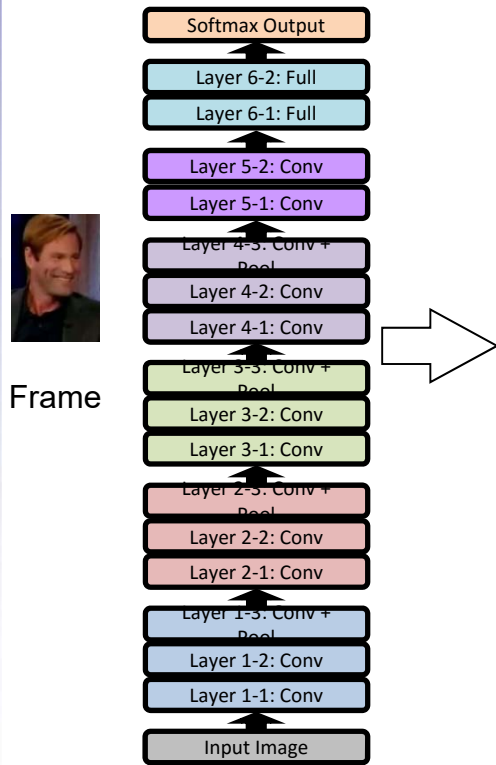
- Videos for testing in the PaSC datasets

Number of Subjects	265
Total Videos	2,802
Total Control Videos	1,401
Total Handheld Videos	1,401
Control Videos per Subject	4 to 7
Handheld Videos per Subject	4 to 7
Number of Locations	6

[Beveridge, BTAS'13]

Our Method

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



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



Training Models

■ Training DCNN

- Caffe, Jia'13 → **14 Cov. Layers (from 5)**
- Pre-train: **CFW**
 - Start learning rate: 0.01
 - 153,461 images from 1520 persons
- Fine-tune: **PaSC training set + COX**
 - Start learning rate: 0.001
 - PaSC training set
 - 170 persons, 38113 images
 - COX training set (our own, surveillance-like videos)
 - 1000 persons, 147,737 video frames

■ Features exploited finally

- **2,048** dimensional features of fc 6-2 layer for each frame



Training Models

- Training HERML
 - 1,165 videos from 470 person, from two heterogeneous datasets
 - PaSC training set
 - 170 persons, 265 videos
 - COX training set
 - 300 persons, 900 videos (3 videos/person)
- Final feature dimensions (per video)
 - 1320 (440*3)-dimensional (KLDA features)

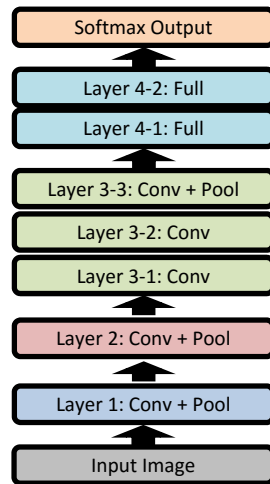


Evaluation Results

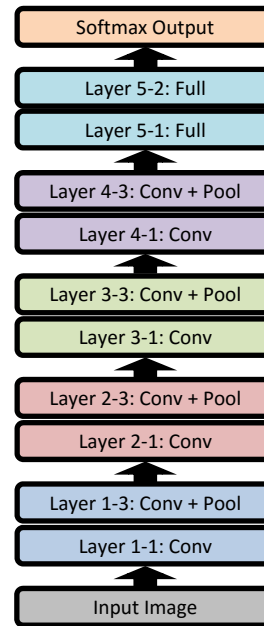
中科院计算所

■ The deeper the better

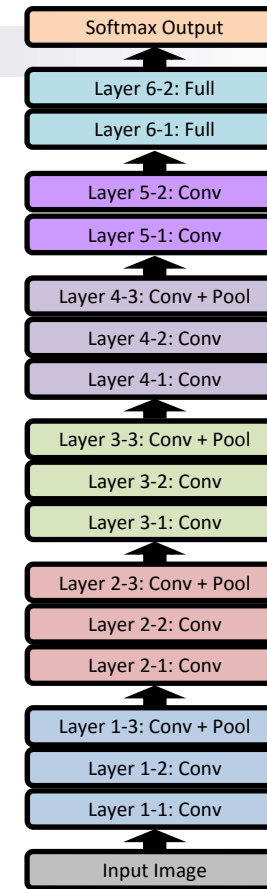
DCNN for single frame



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%



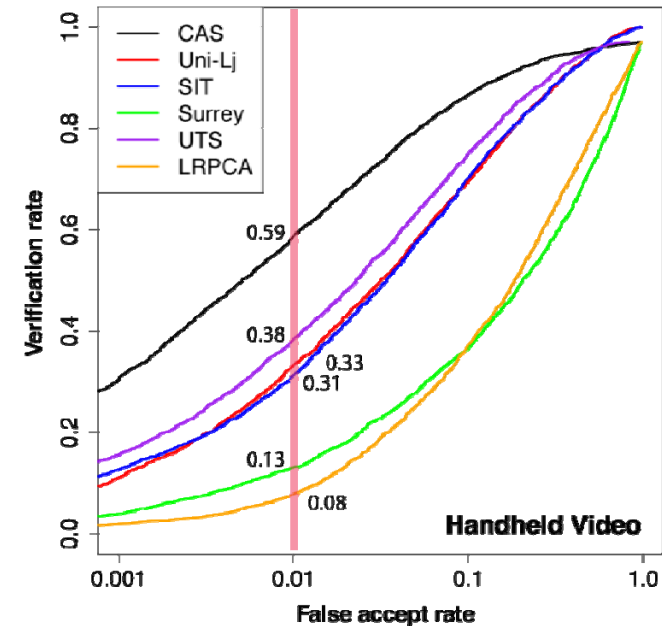
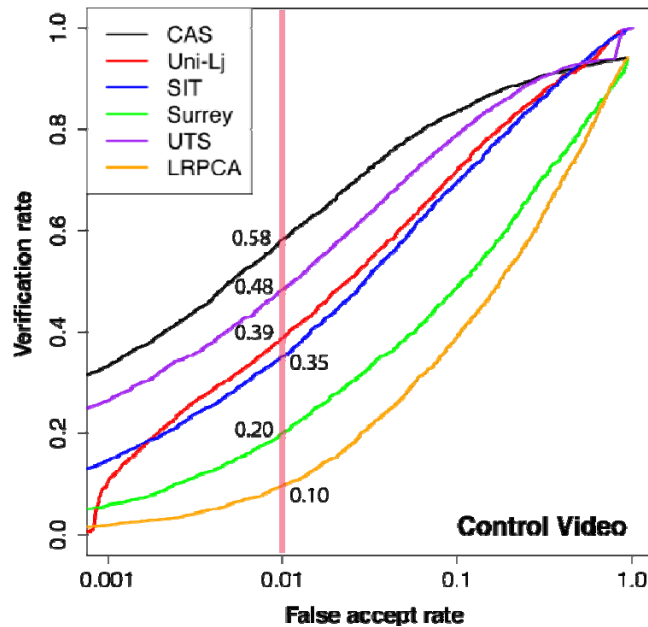
Primary Results

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





我们在DL上的工作

■ 四个竞赛情况

- ACM ICMI EmotioW 2014 challenge
- IEEE FG2015 PaSC video FR challenge
- ICCV15 ChaLearn Apparent Age Estimation
- ICCV15 ChaLearn Culture Event Recognition

■ VIPL SDK情况

- 人脸检测v4.x, 特征定位v4.x, 人脸识别v4.x

■ 三个基于小数据的DL学习方法

- DAE for face alignment [ECCV2014]
- DAE for pose normalization [CVPR2014]
- Bi-shift AE for Domain Adaptation [ICCV2015]

ICCV2015 ChaLearn Looking at People 感知年龄估计竞赛

刘昕等



中国科学院计算技术研究所

Institute of Computing Technology, Chinese Academy of Sciences



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

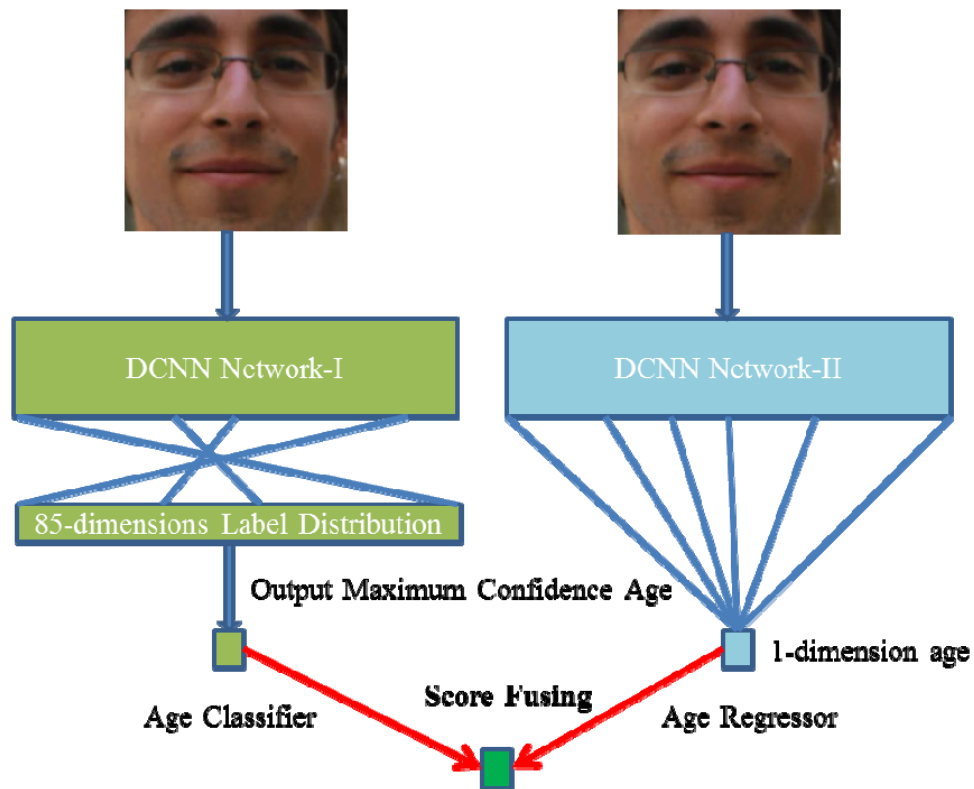
Apparent Age图像示例：每张图像由10名志愿者标注，取均值作为apparent age，并保存标注的标准差



我们的参赛方法：AgeNet

■ 方法介绍

- 端到端学习年龄回归器与年龄分类器并融合

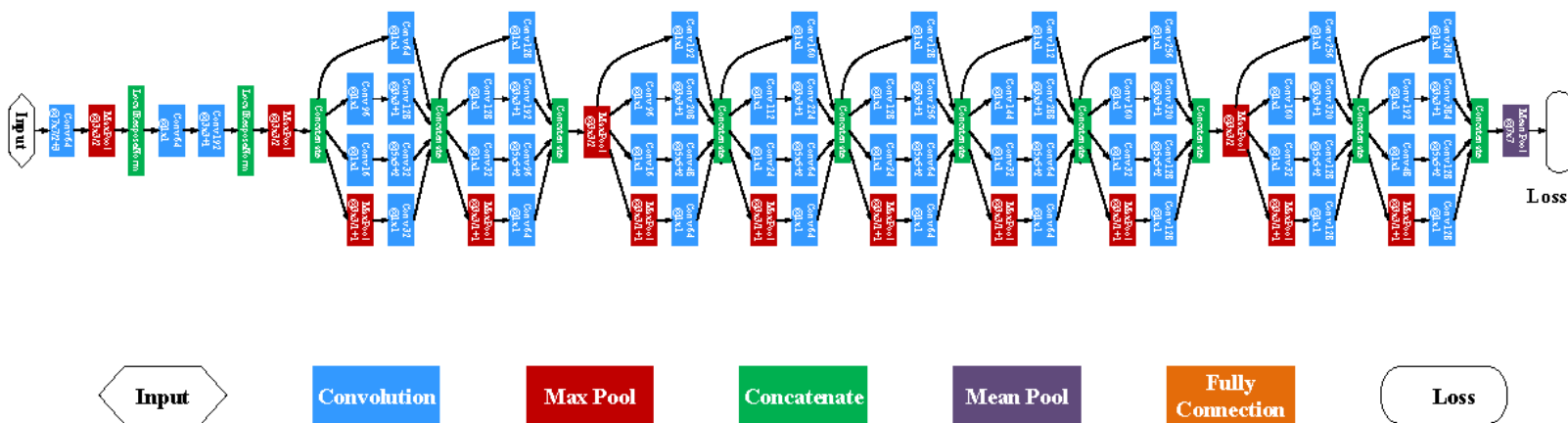




我们的参赛方法：AgeNet

中科院计算所

- 大规模深度卷积网络
 - 对GoogLeNet（22个卷积层）进行改造
 - 去掉两个辅助损失层
 - 加速收敛的策略



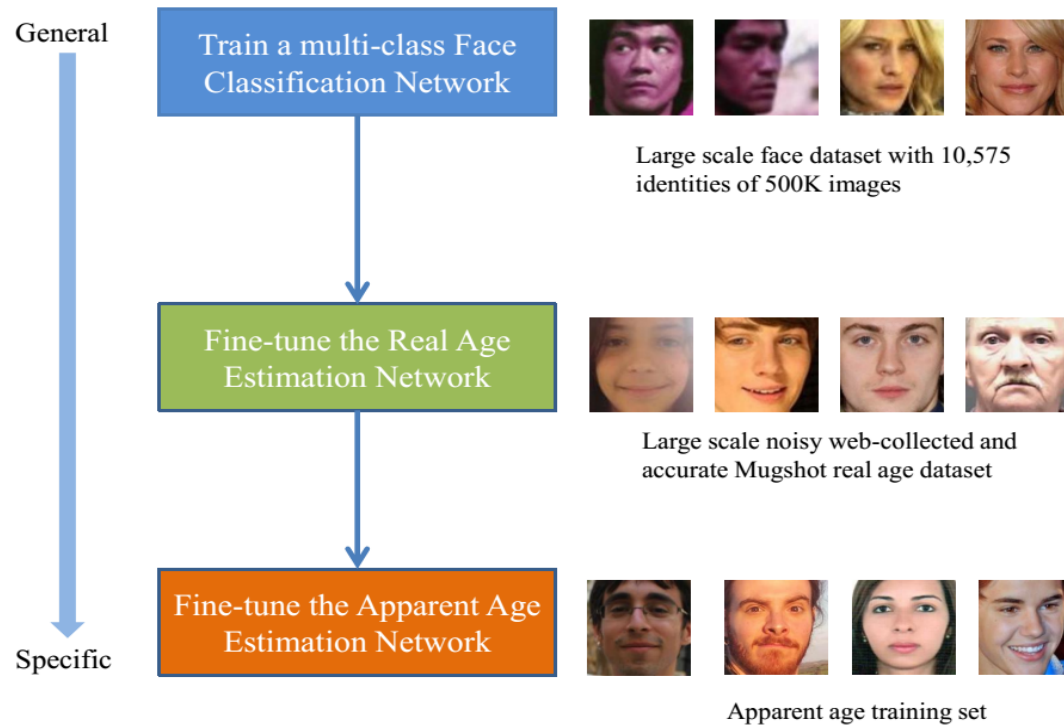


我们的参赛方法：AgeNet

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





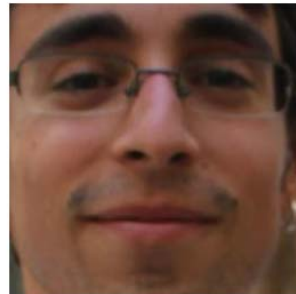
我们的参赛方法：AgeNet

中科院计算所

- 人脸归一化与模型集成
 - 两种人脸归一化模板和两种Crop Size
 - 分别学习回归和分类CNN，形成8个子模型



(1) Exterior



(2) Inner

我们使用的两种人脸对齐模板，人脸均被归一化为256x256

归一化模板	Crop Size	年龄估计器
Exterior	248	Regressor
Exterior	248	Classifier
Exterior	227	Regressor
Exterior	227	Classifier
Inner	248	Regressor
Inner	248	Classifier
Inner	227	Regressor
Inner	227	Classifier



实验结果

■ 性能评价指标

- 主办方采用平均归一误差 (Mean Normalized Error):

$$\bar{\epsilon} = \frac{1}{N} \sum_{i=1}^N \left[1 - \exp\left(\frac{-(y_i - \hat{y}_i)^2}{2\sigma_i^2}\right) \right]$$

- 文献中普遍采用的Mean Absolute Error (MAE)

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$



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



我们在DL上的工作

■ 四个竞赛情况

- ACM ICMI EmotioW 2014 challenge
- IEEE FG2015 PaSC video FR challenge
- ICCV15 ChaLearn Apparent Age Estimation
- **ICCV15 ChaLearn Culture Event Recognition**

■ VIPL SDK情况

- 人脸检测v4.x, 特征定位v4.x, 人脸识别v4.x

■ 三个基于小数据的DL学习方法

- DAE for face alignment [ECCV2014]
- DAE for pose normalization [CVPR2014]
- Bi-shift AE for Domain Adaptation [ICCV2015]

ICCV2015 ChaLearn Looking at People 社会事件识别竞赛

刘梦怡，刘昕等



中国科学院计算技术研究所

Institute of Computing Technology, Chinese Academy of Sciences



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





问题概要

- 基于图像的文化事件识别任务
 - 挑战：复杂图像内容的语义理解
 - 事件发生场景(scene)
 - 事件涉及典型道具(object: 如花灯/国旗/自行车)
 - 人物形态(pose / garment)
 - 人-人/人-物/物-物之间的交互(interaction)
 - 解决思路
 - 图像内容(特征)的层次挖掘
 - 全局背景(context)与局部线索(cues)信息融合



实验结果(1/6)

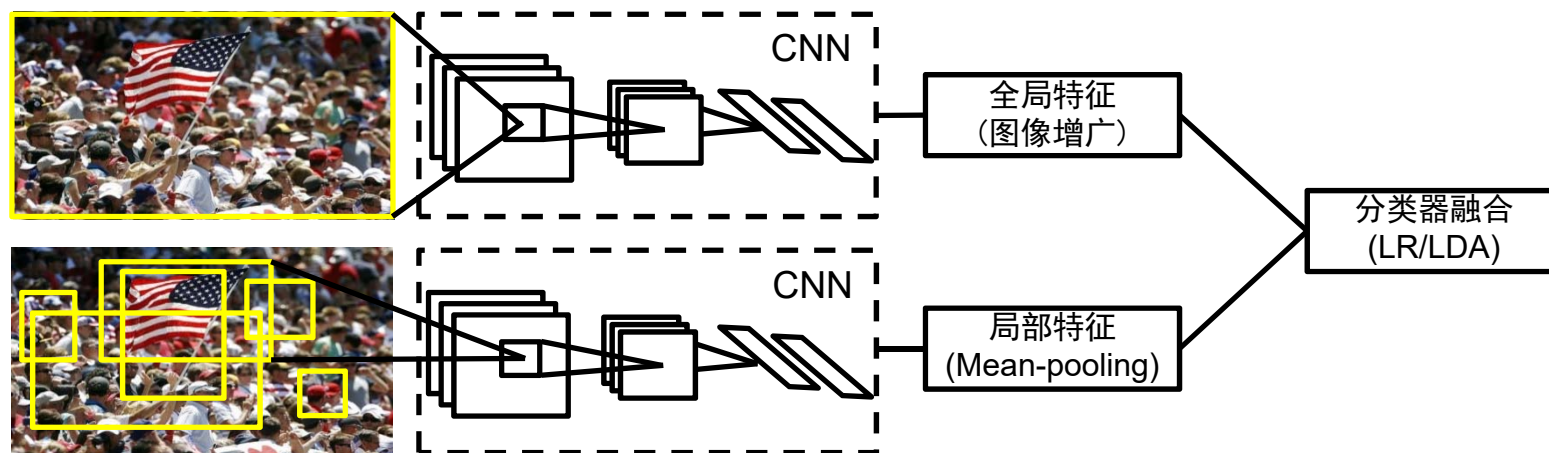
■ 实验数据与协议

- 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

- 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



实验结果

中科院计算所

■ 多模型多特征融合结果

Models	LR	LDA	Fusion LR+LDA
GoogLeNet, 预训练加fine-tuning, 局部特征融合 (loss1)	0.805	0.804	0.820
+GoogLeNet, 预训练加fine-tuning, 局部特征融合 (loss2)	--	--	--
+GoogLeNet, 预训练加fine-tuning, 局部特征融合 (loss3)	0.813	0.804	0.824
+VGG16, 预训练加fine-tuning, 局部特征融合	--	--	--
+VGG19, 预训练加fine-tuning, 局部特征融合	0.830	0.826	0.839
+VGG16, 预训练加fine-tuning, 全局特征	--	--	--
+VGG19, 预训练加fine-tuning, 全局特征	0.841	0.831	0.846
+VGG16, 预训练, 局部特征融合	--	--	--
+VGG19, 预训练, 局部特征融合	0.845	0.829	0.850



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



我们在DL上的工作

■ 四个竞赛情况

- ACM ICMI EmotioW 2014 challenge
- IEEE FG2015 PaSC video FR challenge
- ICCV15 ChaLearn Apparent Age Estimation
- ICCV15 ChaLearn Culture Event Recognition

■ 三个基于小数据的DL学习方法

- DAE for face alignment [ECCV2014]
- DAE for pose normalization [CVPR2014]
- Bi-shift AE for Domain Adaptation [ICCV2015]



我们在DL上的工作

- 四个竞赛情况
 - ACM ICMI EmotioW 2014 challenge
 - IEEE FG2015 PaSC video FR challenge
 - ICCV15 ChaLearn Apparent Age Estimation
 - ICCV15 ChaLearn Culture Event Recognition
- 三个基于小数据的DL学习方法
 - **DAE for face alignment [ECCV2014]**
 - DAE for pose normalization [CVPR2014]
 - Bi-shift AE for Domain Adaptation [ICCV2015]

J. Zhang, S. Shan, M. Kan, X. Chen. Coarse-to-Fine Auto-Encoder Networks (CFAN) for Real-Time Face Alignment. ECCV2014 (oral)



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Human Knowledge
- 知识驱动
- → 数据驱动
- → 混合驱动

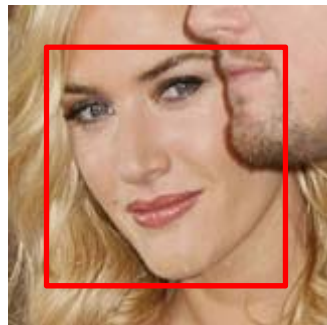


Problem

中科院计算所

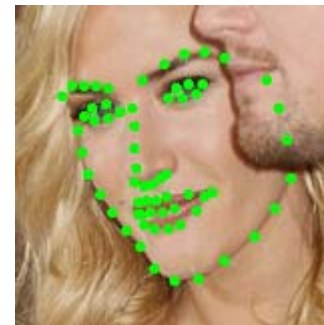
■ Face Alignment

- Predict facial landmarks from detected face



Detected face region $I(u,v)$

Goal



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



Problem

中科院计算所

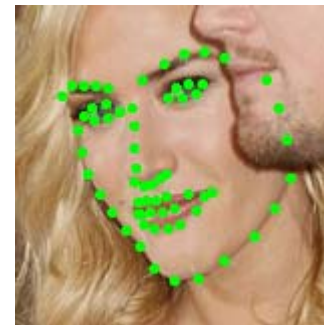
■ Face Alignment

- Predict facial landmarks from detected face



Detected face region $I(u,v)$

Goal



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

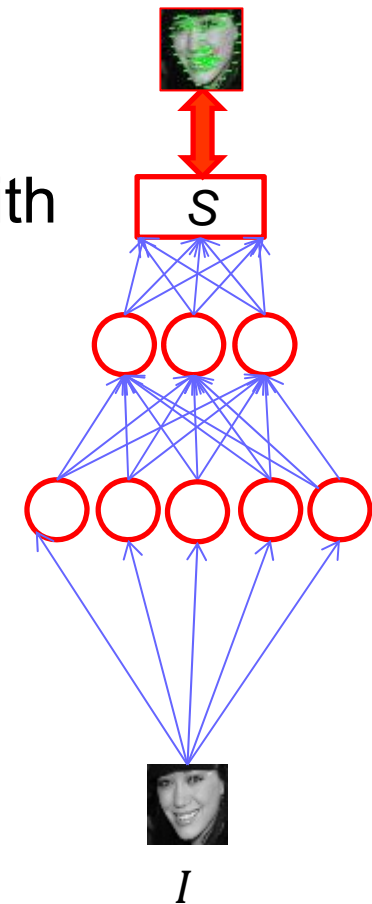
$$S = H(I), I \in R^{w \times h}, S \in R^{2L},$$



Motivation

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

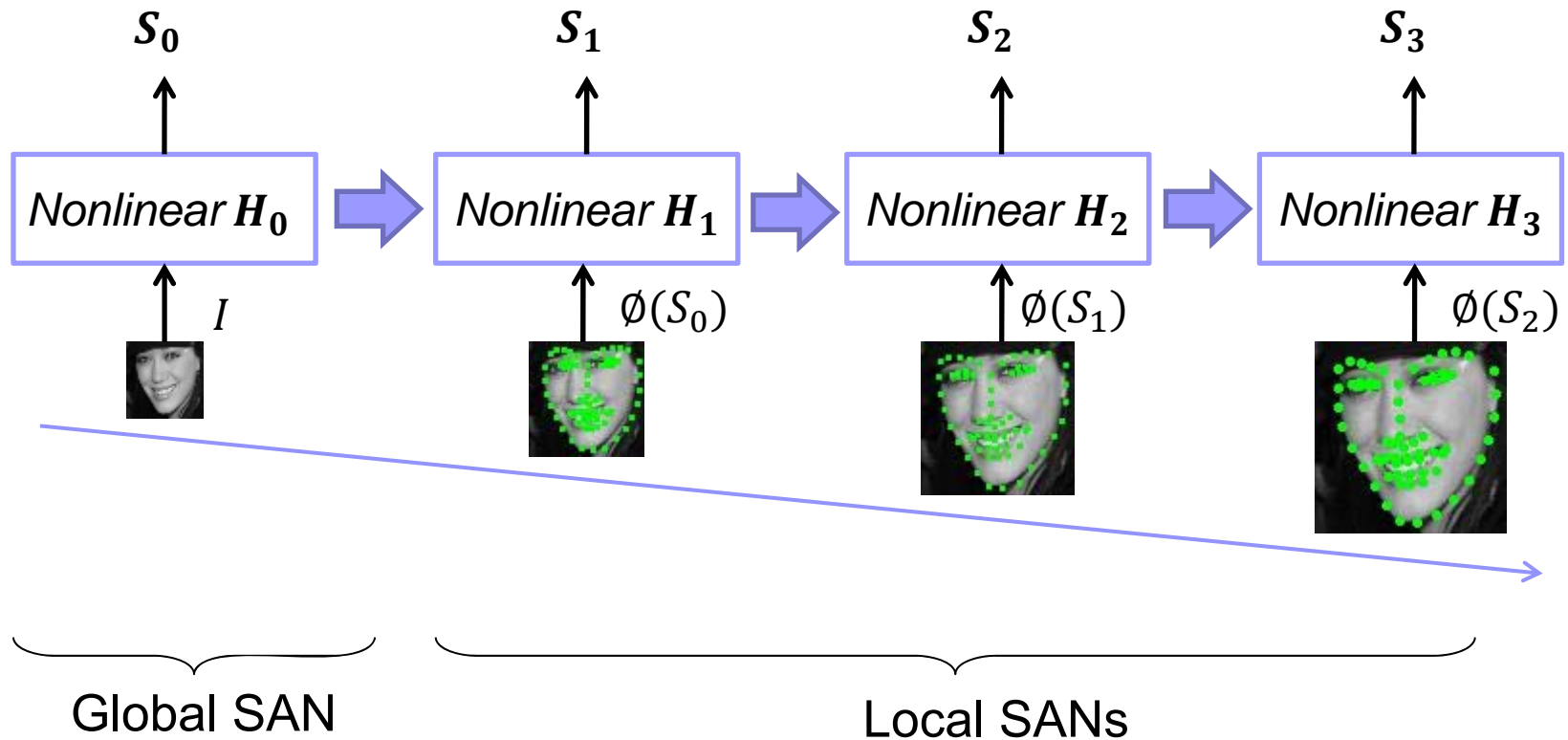




Our Method

中科院计算所

■ Schema of Coarse-to-Fine AE Networks



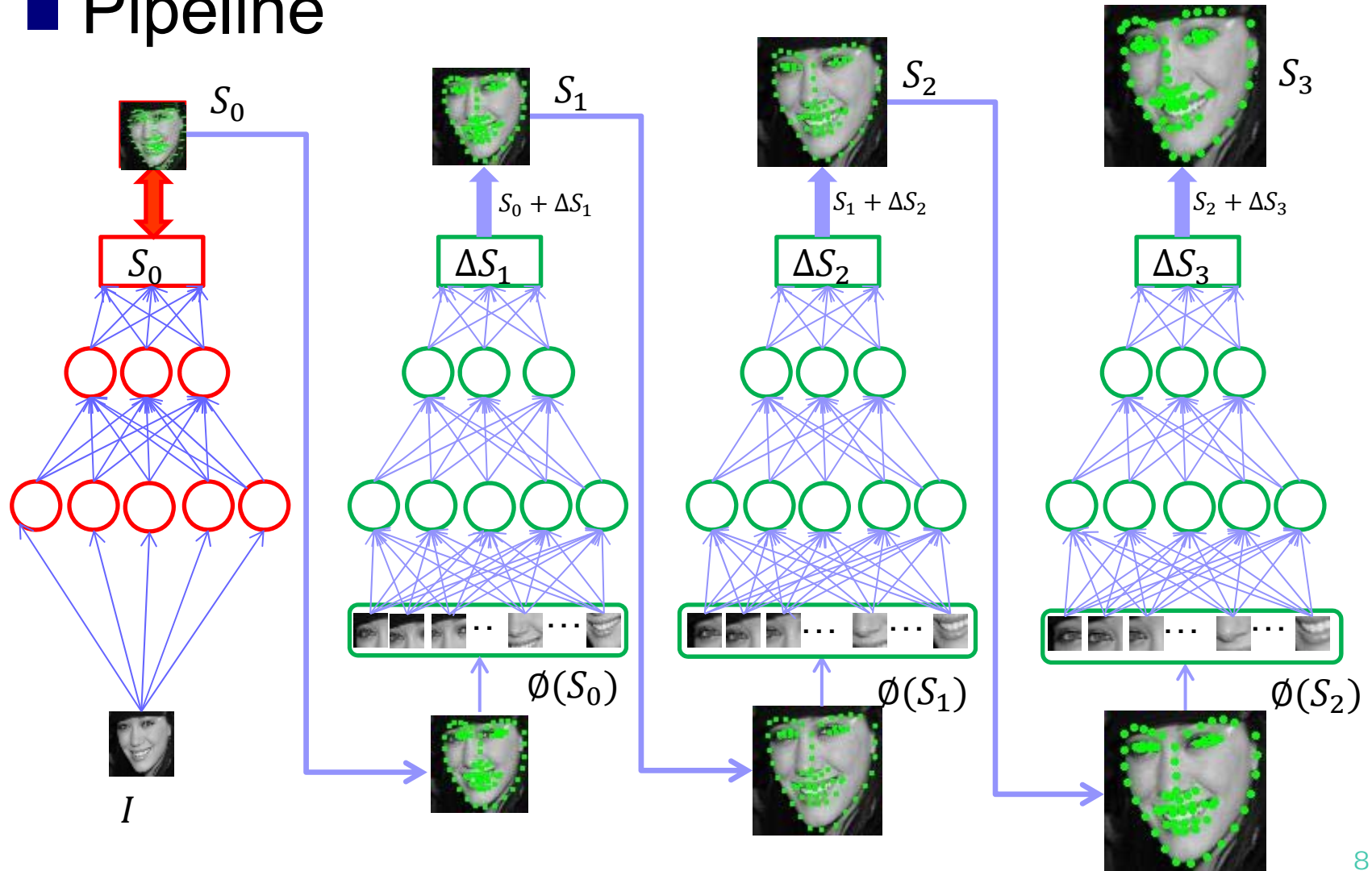
SAN: Stacked Auto-encoder Network



Our Method

中科院计算所

■ Pipeline





Experiments(1/8)

■ Datasets

□ XM2VTS [Messer'99]

- 2360 face images collected over 4 sessions under the controlled settings

□ LFPW [Belhumeur'11]

- 1132 training images and 300 test images collected from wild condition

□ HELEN [Le'12]

- 2330 high-resolution face images collected from the wild, 2000 images for training and 330 images for test

□ AFW [Zhu'12]

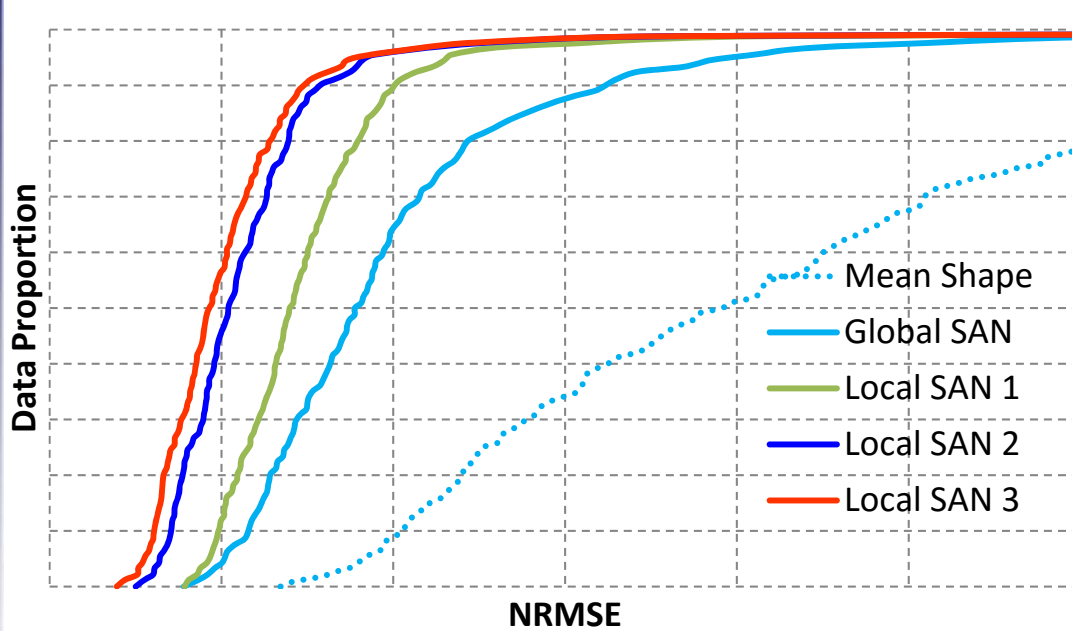
- 205 images with 468 faces collected from the wild



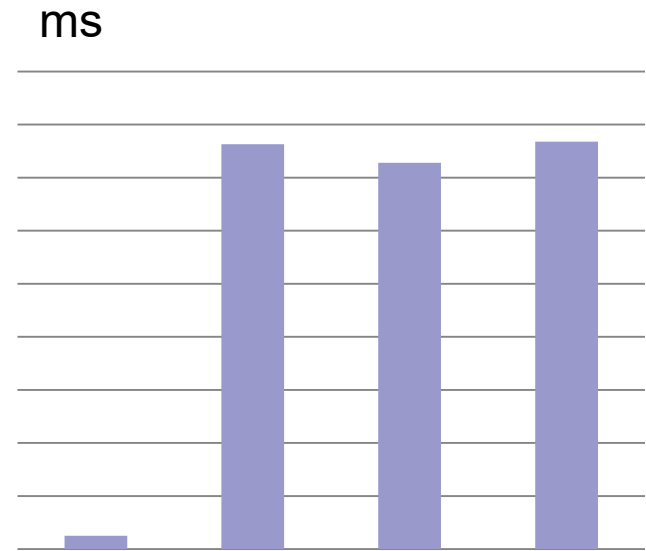
Experiments(2/8)

中科院计算所

■ Evaluation of Successive SANs



Performance gain of each SAN
(Conduct on LFPW)



Run Time (ms)



Experiments(3/8)

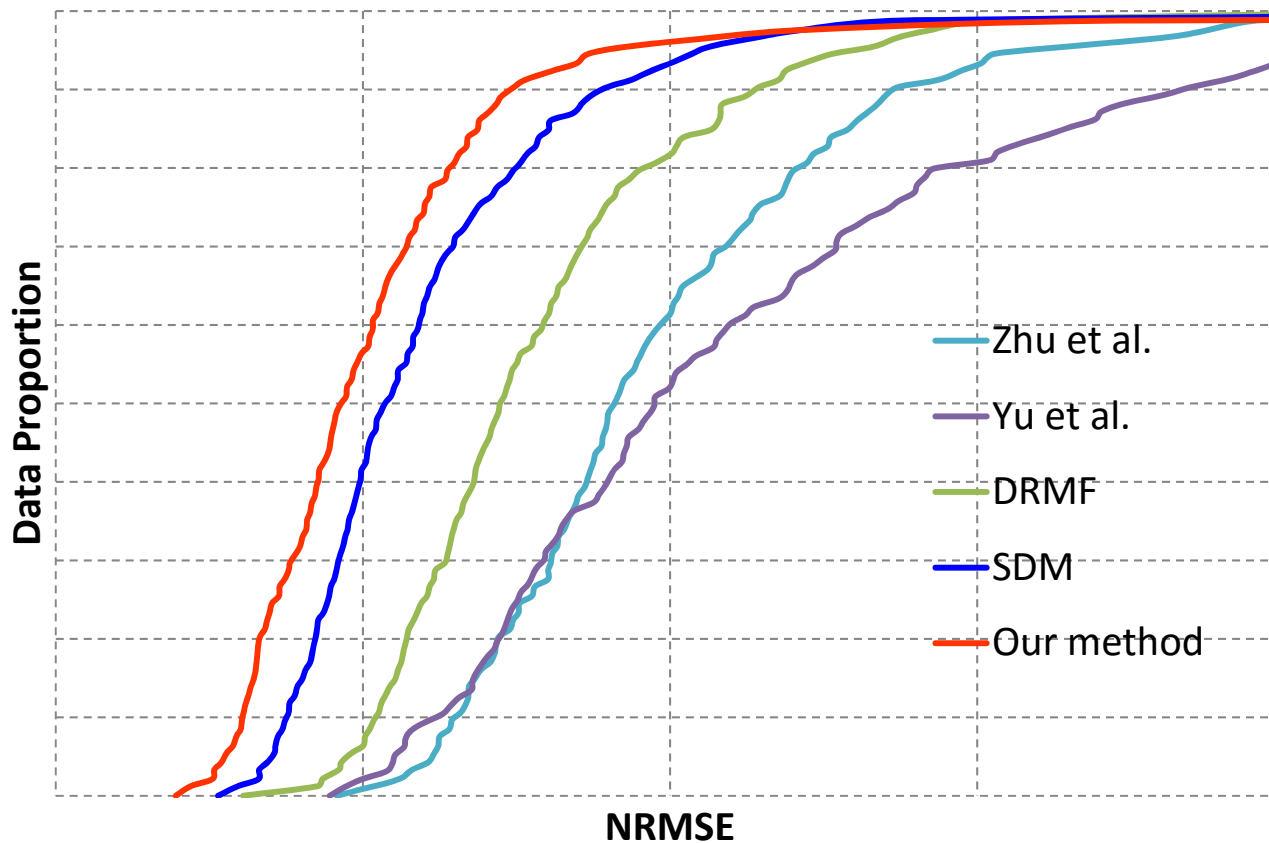
- Comparative Methods
 - Local Models with Regression Fitting
 - SDM [Xiong'13]
 - DRMF [Asthana'13]
 - Tree-structured Models
 - Zhu et al. [Zhu'12]
 - Yu et al. [Yu'13]
 - Deep Model
 - DCNN [Sun'13]



Experimental Result(4/8)

中科院计算所

■ Performance comparisons on HELEN

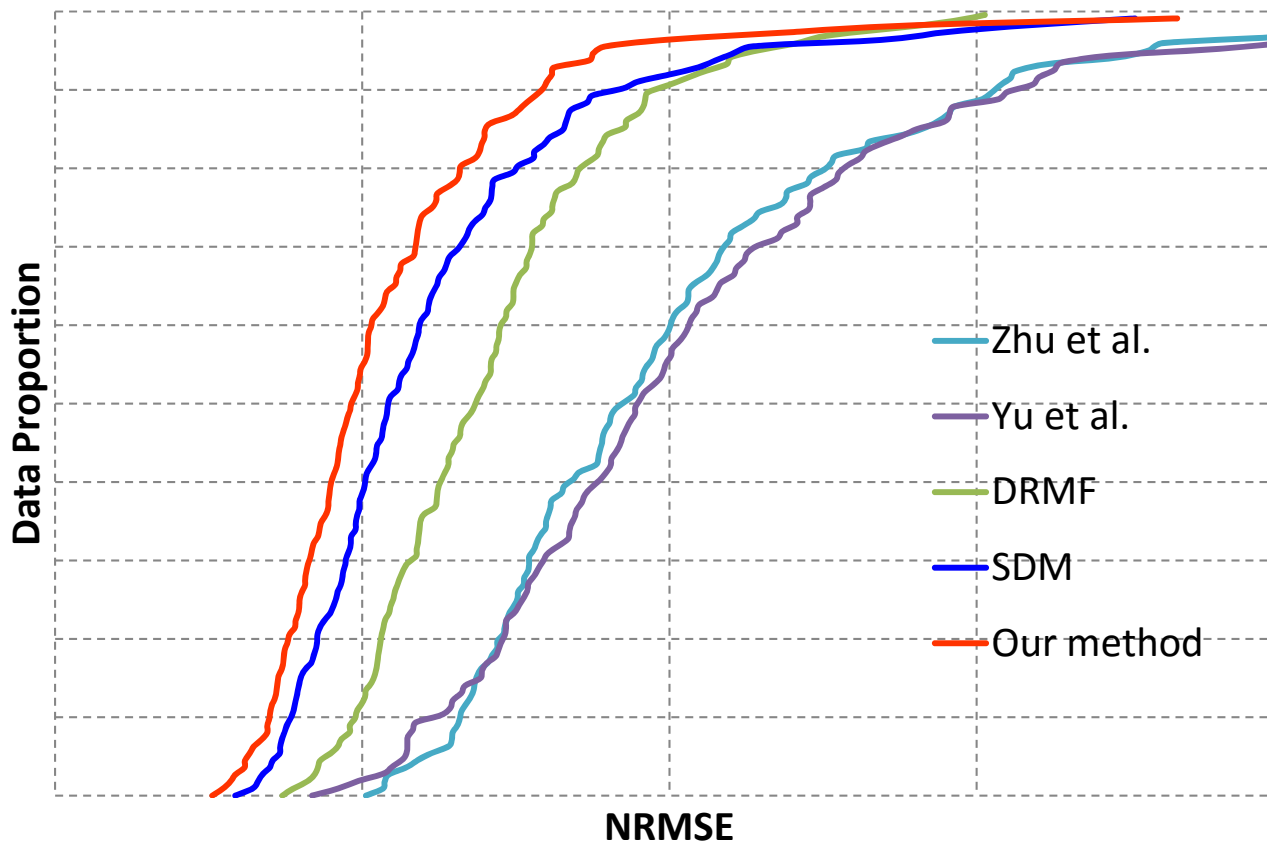




Experimental Result(5/8)

中科院计算所

■ Performance comparisons on LFPW

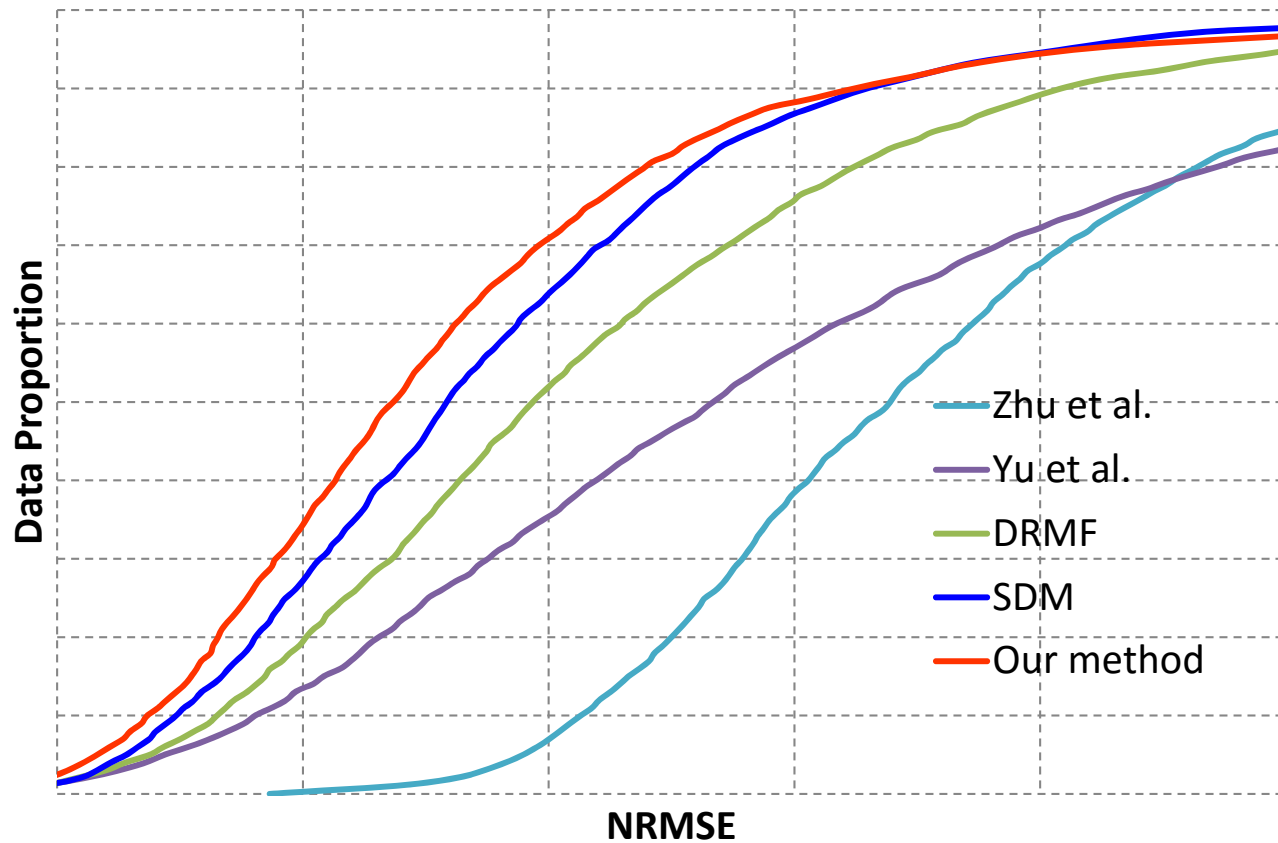




Experimental Result(6/8)

中科院计算所

■ Performance comparisons on XM2VTS

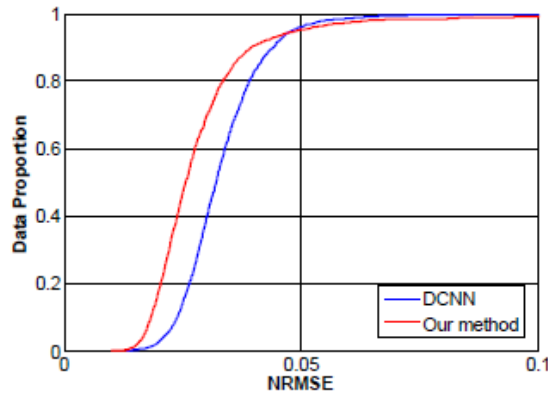




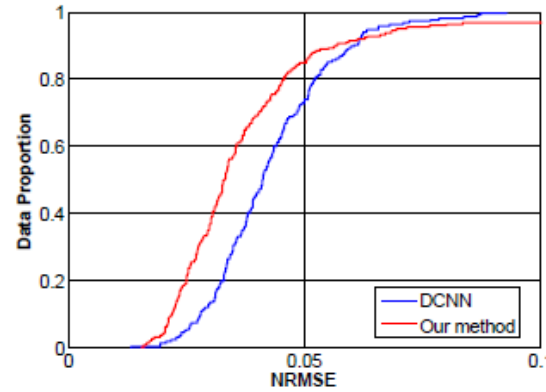
Experimental Result(7/8)

中科院计算所

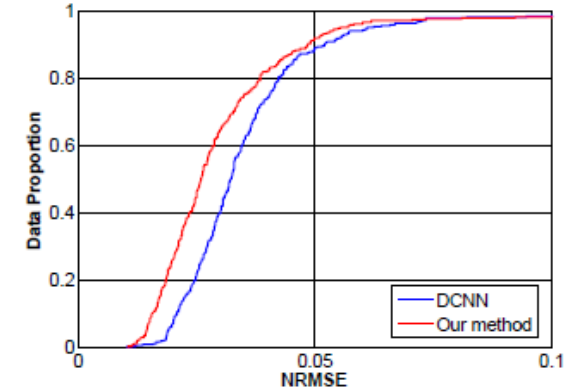
■ Comparisons with DCNN* [Sun et al., CVPR'13]



XM2VTS



LFPW



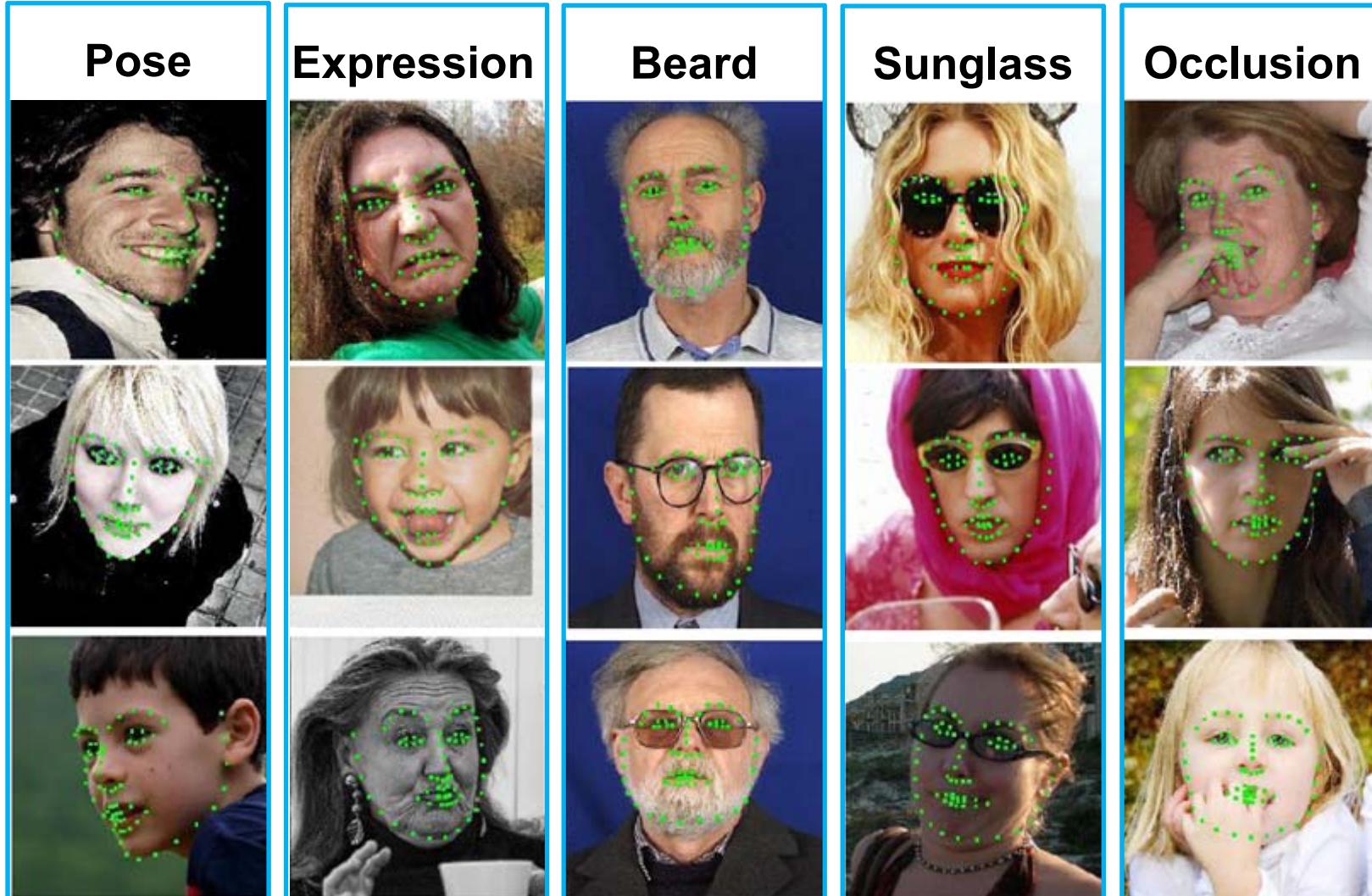
HELEN

Note: The performance is evaluated in terms of five common landmarks



Experimental Result(8/8)

中科院计算所





我们在DL上的工作

■ 四个竞赛情况

- ACM ICMI EmotioW 2014 challenge
- IEEE FG2015 PaSC video FR challenge
- ICCV15 ChaLearn Apparent Age Estimation
- ICCV15 ChaLearn Culture Event Recognition

■ VIPL SDK情况

- 人脸检测v4.x, 特征定位v4.x, 人脸识别v4.x

■ 三个基于小数据的DL学习方法

- DAE for face alignment [ECCV2014]
- **DAE for pose normalization [CVPR2014]**
- Bi-shift AE for Domain Adaptation [ICCV2015]

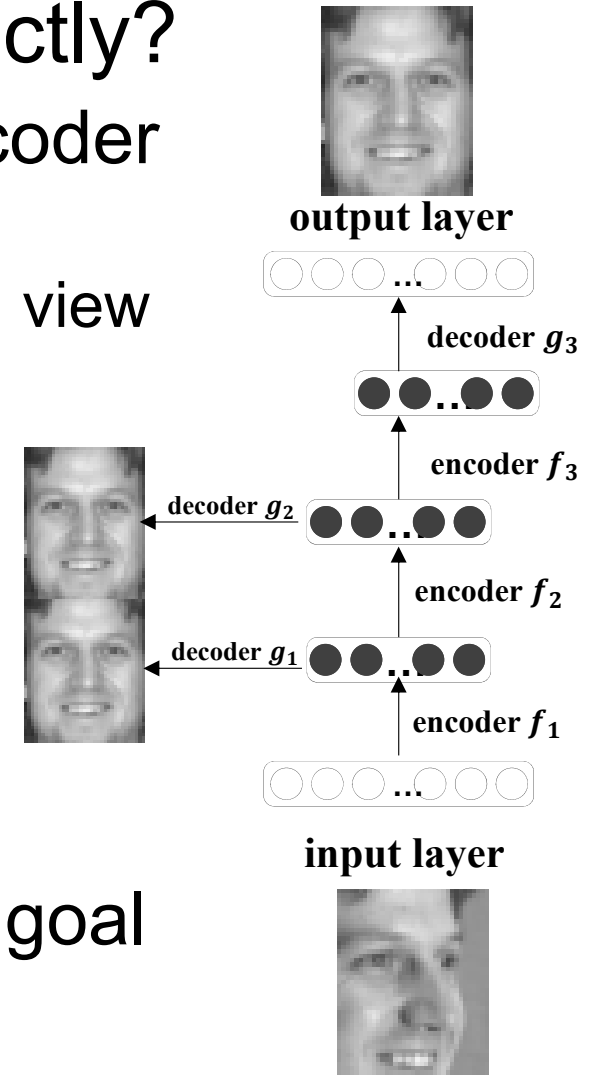
M. Kan, S. Shan, H. Chang, X. Chen. Stacked Progressive Auto-Encoder (SPAЕ) for Face Recognition Across Poses. CVPR2014



Motivation

中科院计算所

- How about deep model directly?
 - Stacked de-noising Auto-Encoder
 - Regard non-frontal view as contaminated version of frontal view
- Unfortunately, fail again
 - Complex non-linear model
 - Easily overfit to “Small” data
- Our idea -- priors
 - Pose changes smoothly
 - Progressively reach the final goal



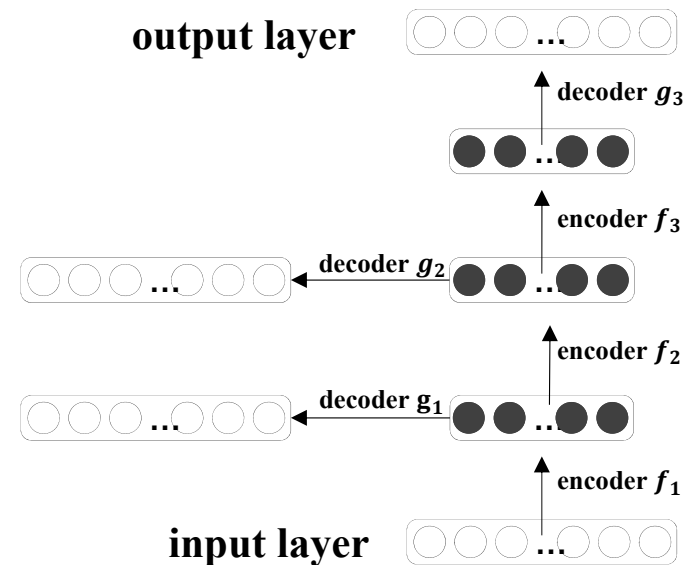
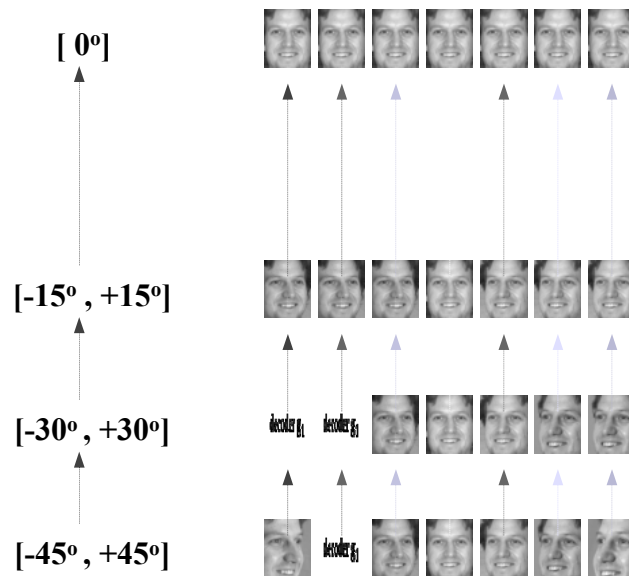


Our Method

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





Experimental Results

中科院计算所

0°
↑
15°
↑
30°
↑
45°

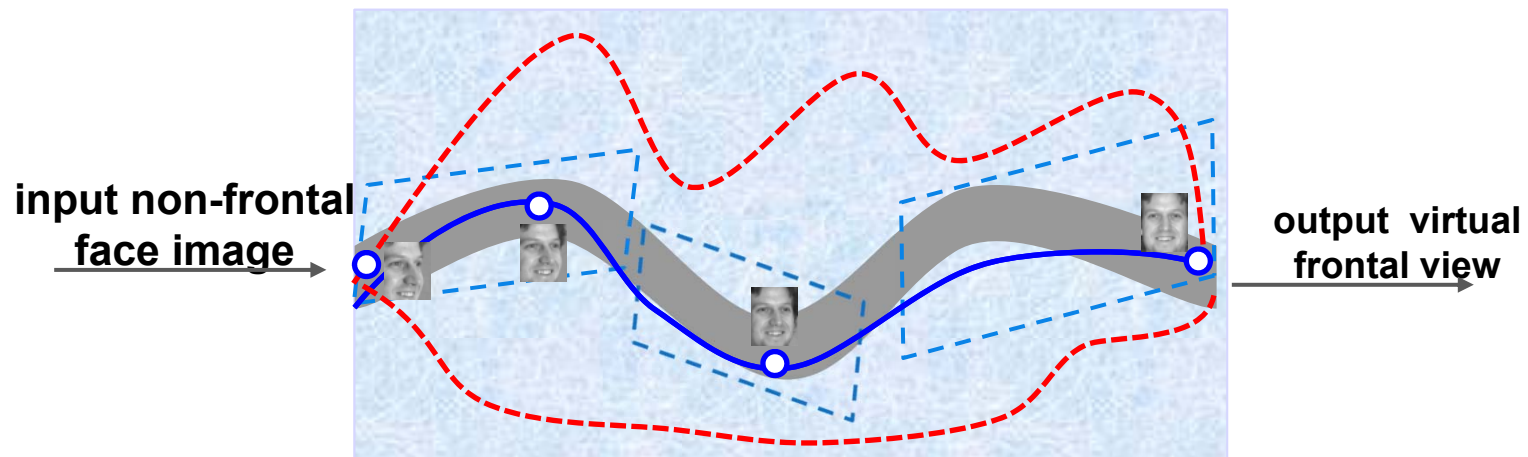




Our Method

■ Discussion

- Medium goals restrict the model, thus alleviate overfitting
 - Multi-view database provides the medium goals
- Otherwise, too many feasible solutions





Experimental Results

中科院计算所

■ Comparison on Multi-PIE

Methods		Probe Pose						Average	Pose Estimation
		-45°	-30°	-15°	+15°	+30°	+45°		
3D	Asthana11	74.1%	91.0%	95.7%	95.7%	89.5%	74.8%	86.8%	Automatic
	MDF	78.7%	94.0%	99.0%	98.7%	92.2%	81.8%	90.7%	
2D	PLS	51.1%	76.9%	88.3%	88.3%	78.5%	56.5%	73.3%	Manually
	CCA	53.3%	74.2%	90.0%	90.0%	85.5%	48.2%	73.5%	
	GMA	75.0%	74.5%	82.7%	92.6%	87.5%	65.2%	79.6%	
	DAE	69.9%	81.2%	91.0%	91.9%	86.5%	74.3%	82.5%	N/A
	our SPAE	84.9%	92.6%	96.3%	95.7%	94.3%	84.4%	91.4%	

■ Comparison on FERET

Methods		Probe Pose								Average	Pose Estimation
		bb -60°	bc -40°	bd -25°	be -15°	bf +15°	bg +25°	bh +40°	bi +60°		
3D	Blanz03	95%	95%	97%	100%	97%	96%	95%	91%	95.8%	Automatic
	MDF	87%	97%	99%	99%	100%	99%	98%	92%	96.4%	
2D	PLS	39%	59%	76%	76%	77%	72%	53%	37%	60.0%	Manually
	CCA	40%	66%	83%	85%	84%	88%	70%	39%	69.4%	
	StackFlow	48%	70%	89%	96%	94%	82%	62%	42%	72.9%	
	DAE	62%	91%	93%	96%	96%	94%	83%	61%	84.5%	N/A
	Our SPAE	77%	96%	98%	99%	99%	99%	95%	77%	92.5%	



我们在DL上的工作

■ 四个竞赛情况

- ACM ICMI EmotioW 2014 challenge
- IEEE FG2015 PaSC video FR challenge
- ICCV15 ChaLearn Apparent Age Estimation
- ICCV15 ChaLearn Culture Event Recognition

■ VIPL SDK情况

- 人脸检测v4.x, 特征定位v4.x, 人脸识别v4.x

■ 三个基于小数据的DL学习方法

- DAE for face alignment [ECCV2014]
- DAE for pose normalization [CVPR2014]
- **Bi-shifting AE for Domain Adaptation [ICCV2015]**

M. Kan, S. Shan, X. Chen. Bi-shifting Auto-Encoder for Unsupervised Domain Adaptation. ICCV2015



Related Works

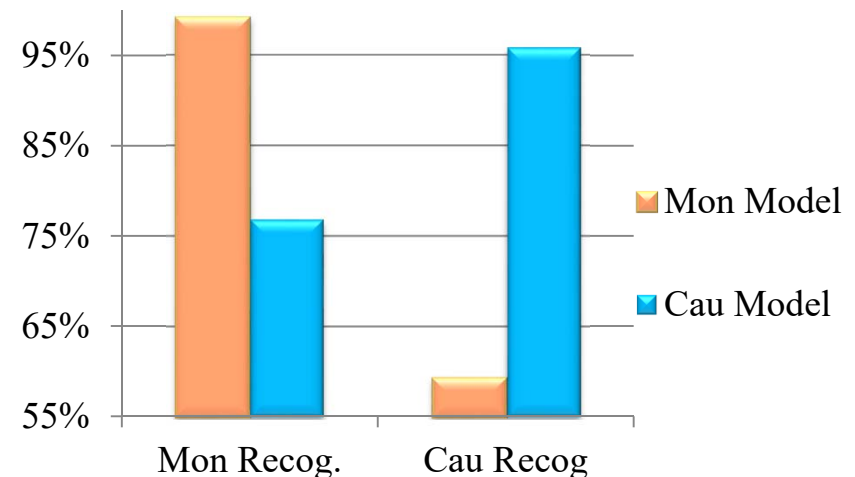
中科院计算所

- Problem
 - Training and testing data from different domains
- Domain Adaptation
 - Distribution Discrepancy between domains
 - Labeled source domain and Unlabeled target domain
 - Discrepancy → performance degeneration

- $\mathcal{Y}_S = \mathcal{Y}_T,$

- $P_S(y|x) = P_T(y|x).$

- But, $\mathcal{X}_S \neq \mathcal{X}_T$ or $P_S(x) \neq P_T(x).$

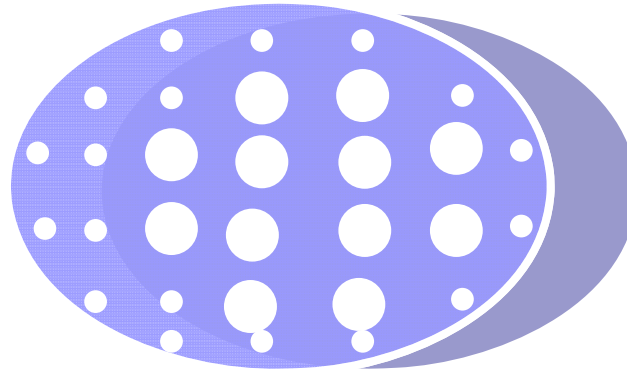




Related Works

中科院计算所

- Instance-based Approaches
 - Resampling/reweighting source domain
 - Sample Selection Bias/ Covariant Shift¹
 - Sample selection bias: $p_1(x) \neq p_2(x), p_1(y|x) \neq p_2(y|x)$
 - Covariant Shift: $p_1(x) \neq p_2(x), p_1(y|x) = p_2(y|x)$
 - Criterion: equalize the distribution of each sample



1. Quionero-Candela, et al, Data Shift in Machine Learning, MIT Press 2009



Related Works

- Feature-based Approaches
 - Feature representation that minimize the discrepancy
 - structural correspondence learning [Blitzer et al., EMNLP 2006]
 - Transfer Component Analysis [Pan et al., 2009]
 - Sampling Geodesic Flow [Gopalan et al., ICCV 2011]
 - Geodesic Flow Kernel [Gong et al., CVPR 2011]
 - Criterion
 - maximum mean discrepancy
 - K-L divergence
 - Bregman divergence
 - Low-rank



Related Works

中科院计算所

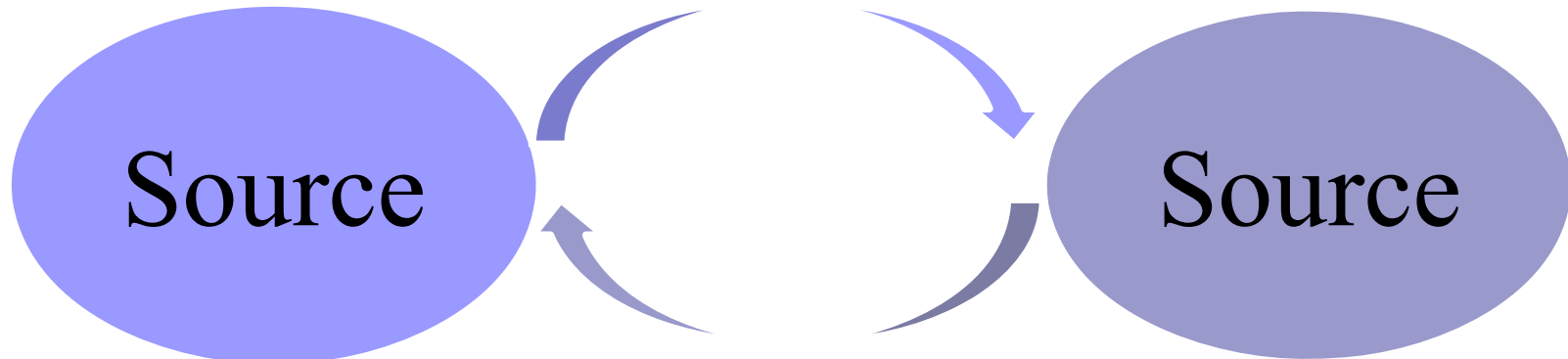
- Model-based Approaches
 - Transductive SVM [Chen et al. PRL 2003]
 - Domain Adaption SVM [Bruzzone et al. PAMI 2010]
 - Addaptive AML [Duan, et al. TPMAI 2012]
- Criterion
 - Distribution distance
 - Iteratively include more target domain samples



Bi-shifting Auto-Encoder

中科院计算所

- Main Idea
 - Shifting samples between domain
- Approach
 - Non-linear model for shifting
 - Sparse-reconstruction for closeness measurement
 - Shifted domain and desirable domain

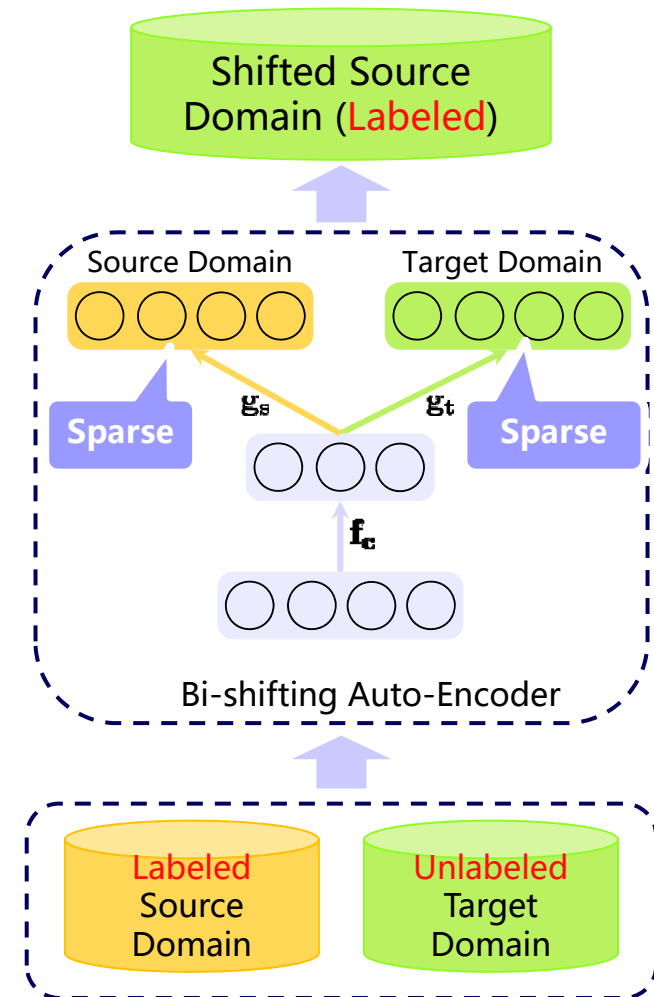




Bi-shifting Auto-Encoder

中科院计算所

- Overview
 - Shifting model
 - Auto-Encoder
 - sparse representation
 - Shifted source domain
 - with class label
 - Learn classification model
 - On shifted source samples

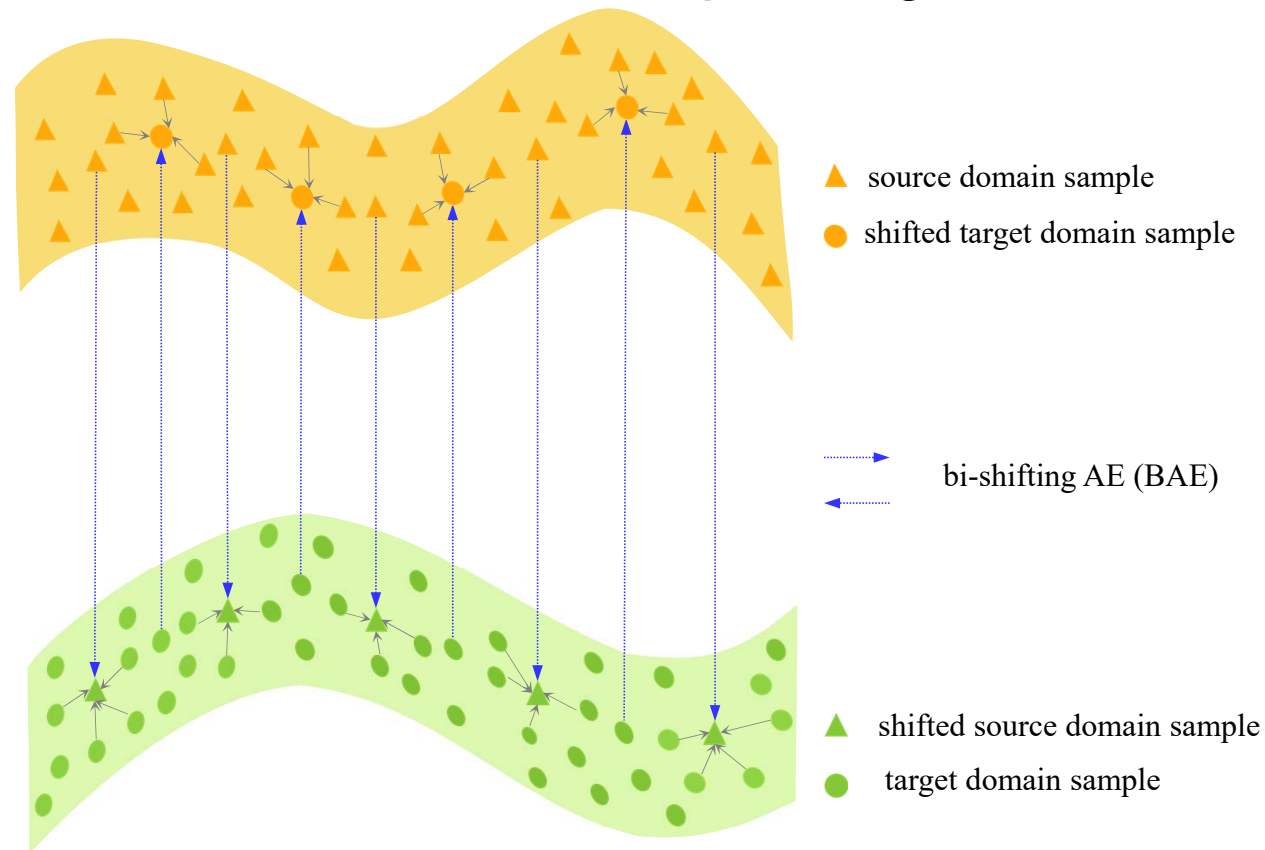




Bi-shifting Auto-Encoder

中科院计算所

- Why sparse reconstruction for closeness measurement?
 - Distance between corresponding local structure





Bi-shifting Auto-Encoder

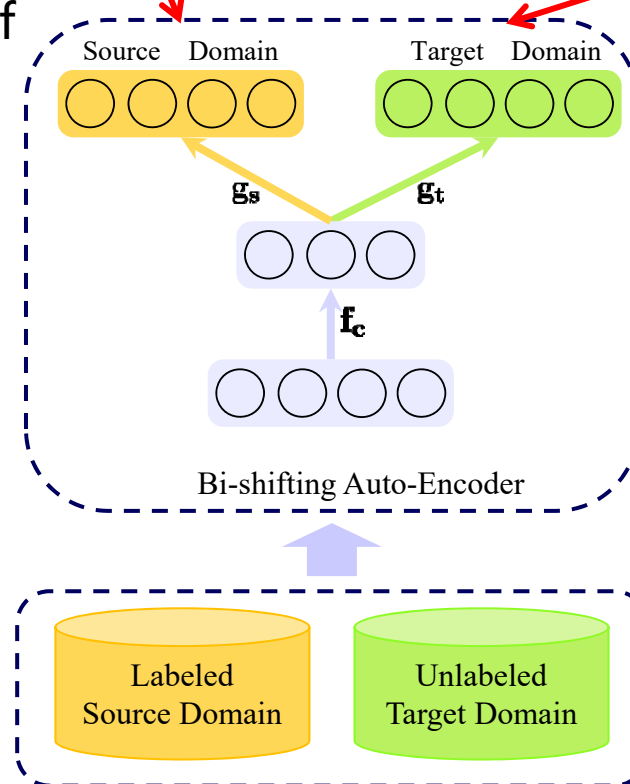
中科院计算所

Formulation $\min \Omega_1 + \Omega_2$

$$\min \Omega_1 = \min_{\mathbf{f}_c, \mathbf{g}_s, \mathbf{g}_t, \mathbf{B}_s} \underbrace{\|\mathbf{X}_s - \mathbf{g}_s(\mathbf{f}_c(\mathbf{X}_s))\|_2^2}_{\text{Reconstruct Itself}} + \underbrace{\|\mathbf{X}_t \mathbf{B}_t - \mathbf{g}_t(\mathbf{f}_c(\mathbf{X}_s))\|_2^2}_{\text{Closeness to target domain constrained by sparse representation}} + \gamma \sum_{i=1}^{n_s} |\beta_i^t|_1$$

Reconstruct Itself

Closeness to target domain constrained by sparse representation





Bi-shifting Auto-Encoder

中科院计算所

■ Formulation

$$\begin{aligned} \min_{\mathbf{f}_c, \mathbf{g}_s, \mathbf{g}_t, \mathbf{B}_s, \mathbf{B}_t} & \|\mathbf{X}_s - \mathbf{g}_s(\mathbf{f}_c(\mathbf{X}_s))\|_2^2 + \|\mathbf{X}_t \mathbf{B}_t - \mathbf{g}_t(\mathbf{f}_c(\mathbf{X}_s))\|_2^2 \\ & + \|\mathbf{X}_s \mathbf{B}_s - \mathbf{g}_s(\mathbf{f}_c(\mathbf{X}_t))\|_2^2 + \|\mathbf{X}_t - \mathbf{g}_t(\mathbf{f}_c(\mathbf{X}_t))\|_2^2 \\ & + \gamma \left(\sum_{i=1}^{n_s} |\beta_i^t|_1 + \sum_{i=1}^{n_t} |\beta_i^s|_1 \right). \end{aligned}$$

■ Shifted Source Domain

$$(\mathbf{X}_s, \mathbf{y}_s) \rightarrow (\mathbf{G}_t, \mathbf{y}_s)$$

$$\mathbf{G}_t \triangleq \mathbf{g}_t(\mathbf{f}_c(\mathbf{X}_s))$$

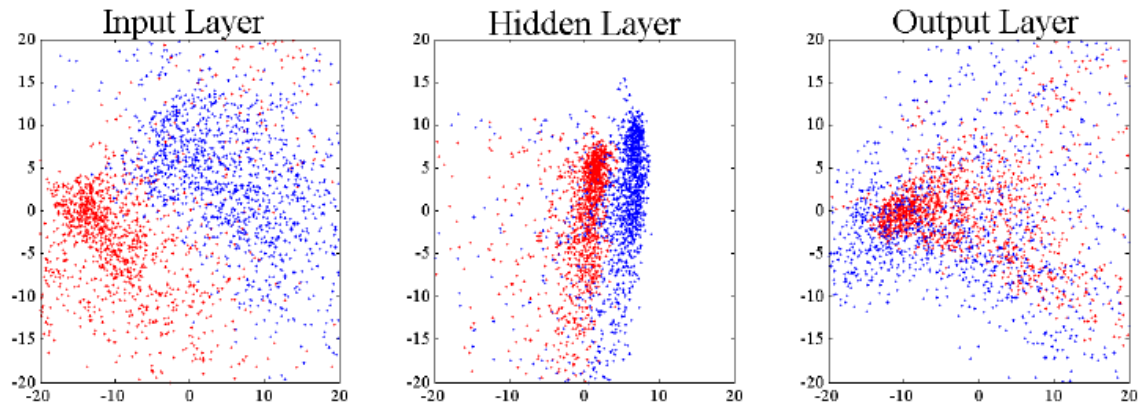


Experiments

中科院计算所

■ Shifting across poses on MultiPIE

	$-45^\circ \rightarrow 0^\circ$	$-30^\circ \rightarrow 15^\circ$	$-15^\circ \rightarrow 30^\circ$	$0^\circ \rightarrow 45^\circ$	$-45^\circ \rightarrow -15^\circ$	$-15^\circ \rightarrow 15^\circ$	$15^\circ \rightarrow 45^\circ$	Average
Source LDA	0.665	0.693	0.669	0.617	0.703	0.719	0.637	0.672
Target PCA	0.615	0.632	0.583	0.541	0.651	0.632	0.541	0.599
ITL	0.745	0.727	0.653	0.597	0.721	0.714	0.634	0.684
SGF+LDA	0.716	0.714	0.669	0.629	0.735	0.748	0.629	0.691
GFK (PCA,LDA)	0.751	0.754	0.699	0.615	0.767	0.761	0.624	0.710
Landmarks (PCA,LDA)	0.747	0.759	0.701	0.640	0.763	0.763	0.642	0.716
LTSL (LDA)	0.736	0.735	0.698	0.642	0.752	0.767	0.640	0.710
AE+LDA	0.735	0.708	0.702	0.656	0.746	0.739	0.649	0.705
BAE+LDA(Ours)	0.795	0.794	0.763	0.698	0.803	0.796	0.693	0.763





Experiments

中科院计算所

■ Shifting across ethnicity

	Cau→Mon	Mon→Cau	Average
Source LDA	0.679	0.676	0.678
ITL	0.801	0.775	0.788
SGF+LDA	0.790	0.751	0.771
GFK (PCA,LDA)	0.738	0.721	0.730
Landmarks (PCA,LDA)	0.718	0.763	0.741
LTSL (LDA)	0.791	0.793	0.792
AE+LDA	0.784	0.786	0.785
BAE+LDA(Ours)	0.892	0.826	0.859



Experiments

中科院计算所

■ Shifting across lighting modality

	VIS→NIR	NIR→VIS	Average
Source LDA	0.816	0.779	0.798
ITL	0.858	0.877	0.868
SGF+LDA	0.841	0.832	0.837
GFK (PCA,LDA)	0.850	0.867	0.859
Landmarks (PCA,LDA)	0.859	0.871	0.865
LTSL (LDA)	0.868	0.878	0.873
AE+LDA	0.827	0.846	0.837
BAE+LDA(Ours)	0.904	0.920	0.912



Summary and discussion

- DL (esp. CNN) wins with “big” data
 - So, collect big data...
 - The deeper, the better
- No ability to collect big data? Or, big data is impossible?
 - Past experiences help to build model
 - Data structure help to design network
 - Priors help to design the objective functions



中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

Thanks, Q & A

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

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