

多模态人脸识别与属性学习

韩 琥

中科院计算所 VIPL研究组

2018/4/20

hanhu@ict.ac.cn

- 1 背景与研究现状
- 2 RGB-D多模态人脸识别 [FG'18, ICB'18]
- 3 表观属性分析 [CVPR'18, ICPR'18, TPAMI'17, TPAMI'15]
- 4 由表及里属性分析 [ICPR'18, IJCB'17]
- 5 总结



FaceID中的TrueDepth技术

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences



关键词：结构光、深度感知、3万顶点、百万分之一重合率(FAR)



常见的多模态人脸识别

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

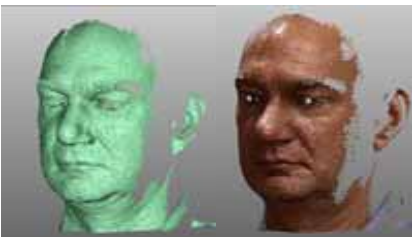
静态
图像



视频
序列



三维
图像
(S+T)



RGB-D
(T+D)



近红外
热像
...



多模态识别范式

静态图像+视频序列 (广义)

静态图像+模拟画像

3D人脸识别：纹理+形状

RGB-D人脸识别：形状+深度

彩色 + 近红外，彩色 + 热像

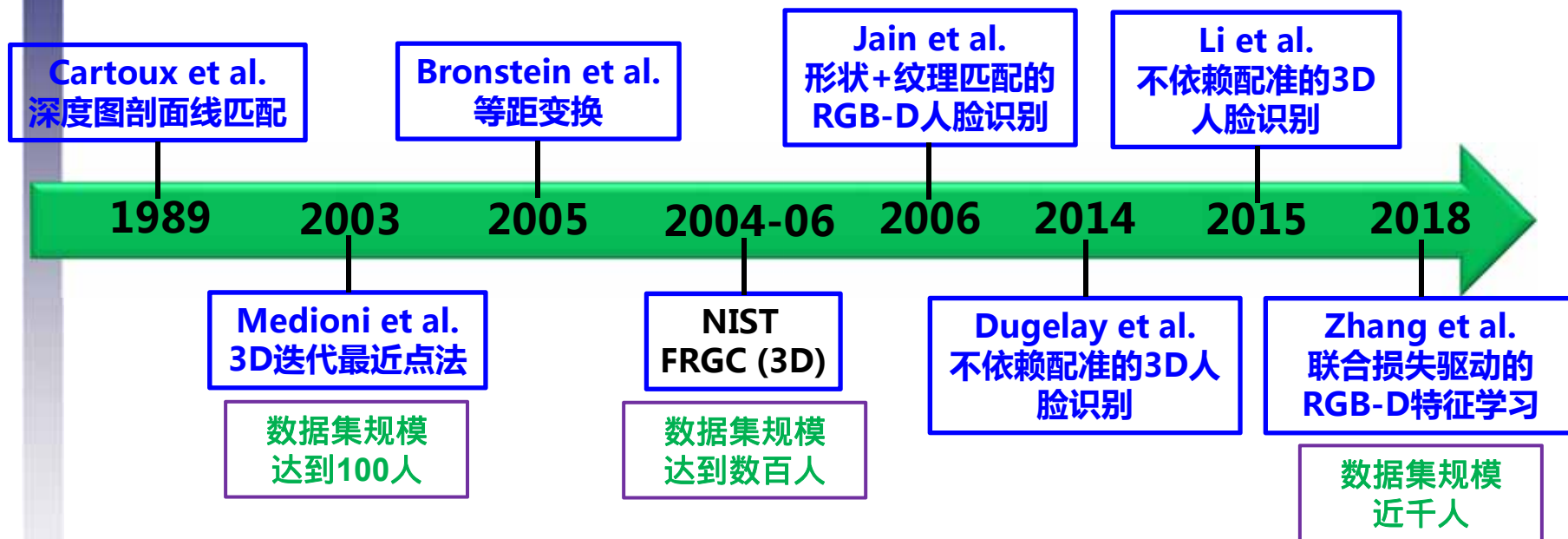


多模态人脸识别

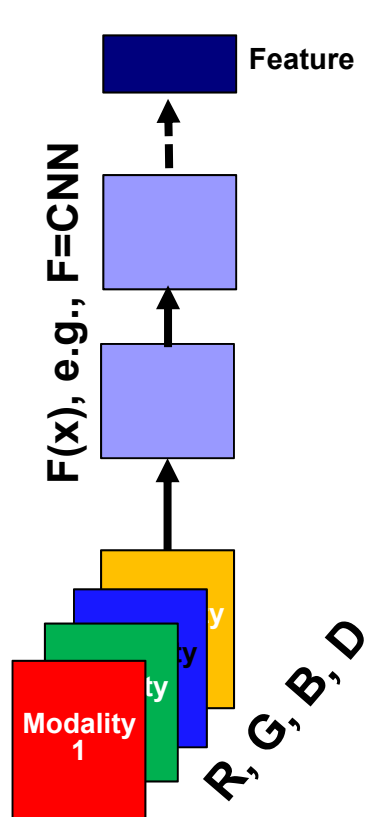
中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

■ 基于RGB-D/3D的人脸识别存在已久

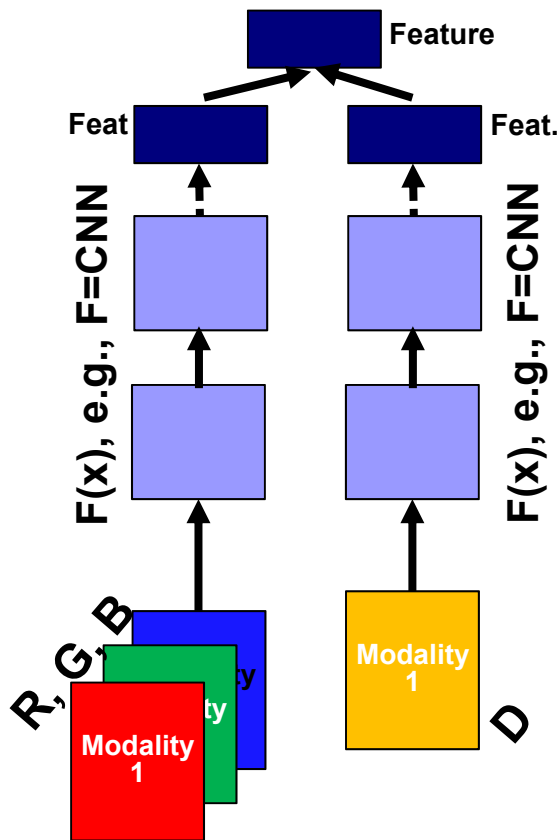


主流框架



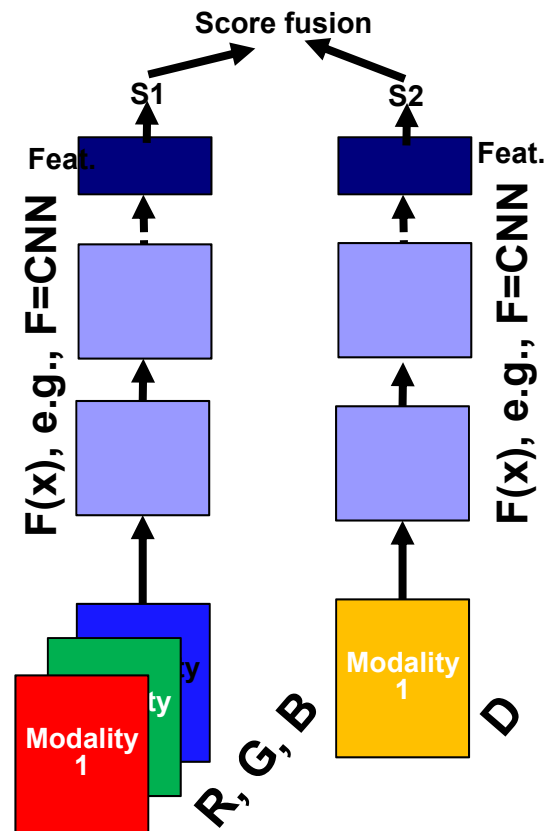
信号层融合

输入信号的多通道扩展较容易



特征层融合

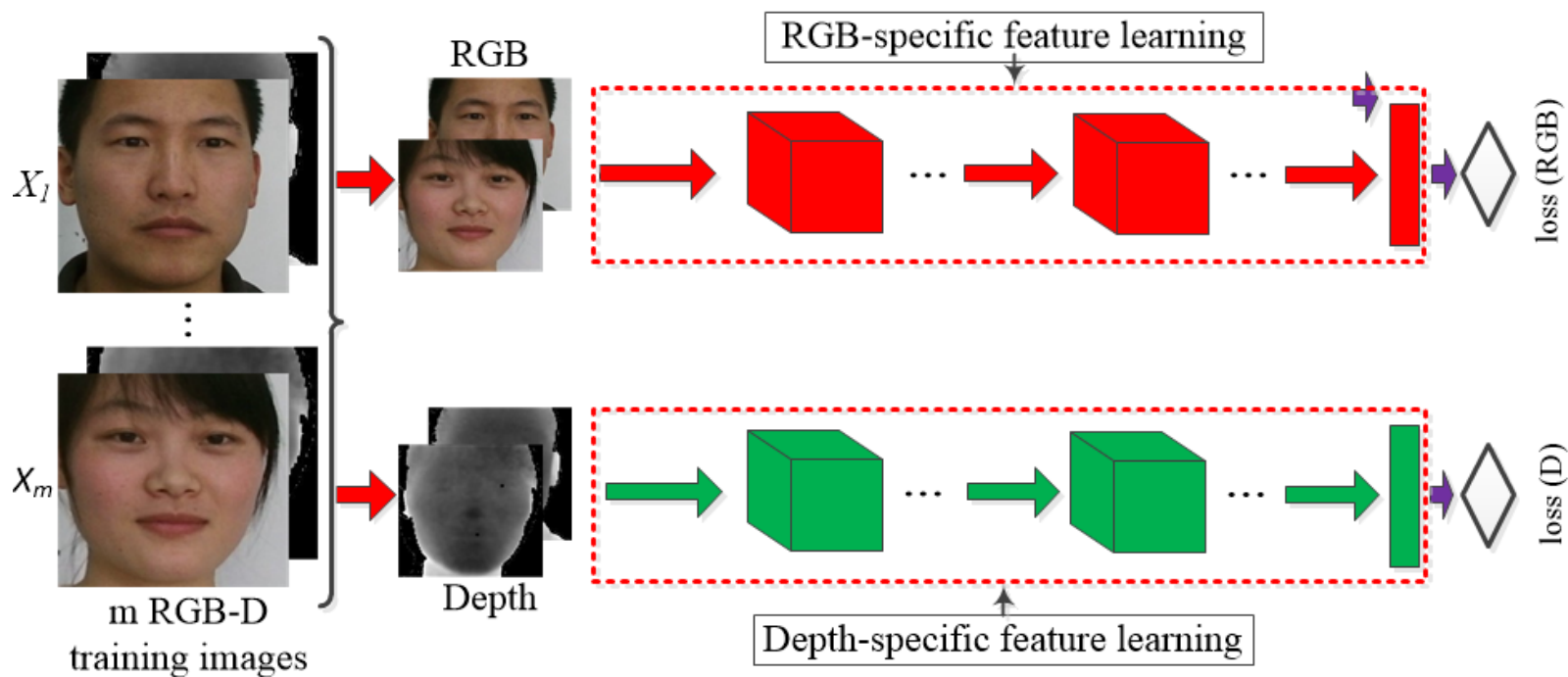
需要仔细设计



得分层融合

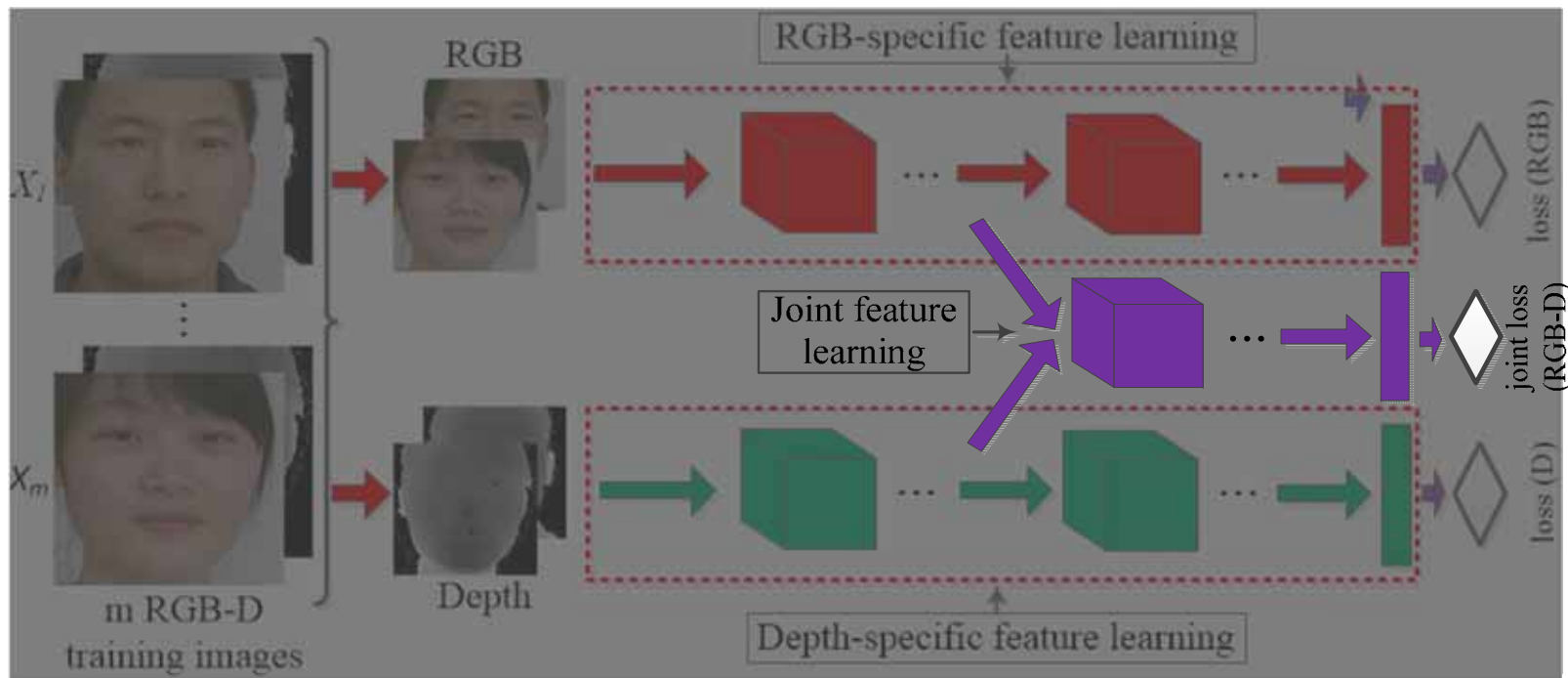
通道独立建模再融合较容易

- 独立与联合损失共同驱动的互补特征学习[FG18]



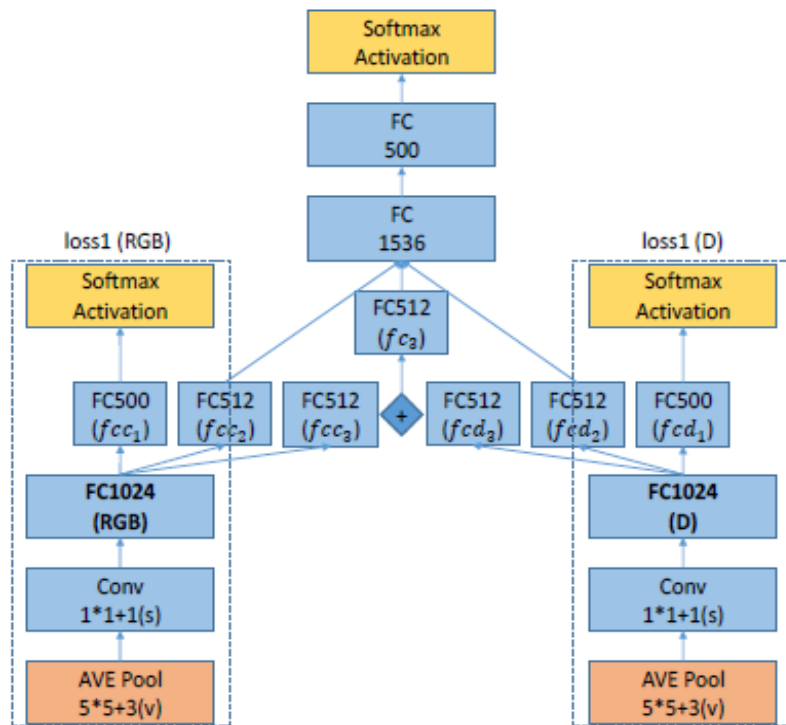
不同模态特征学习过程中完全没有交互

- 独立与联合损失共同驱动的互补特征学习[FG18]

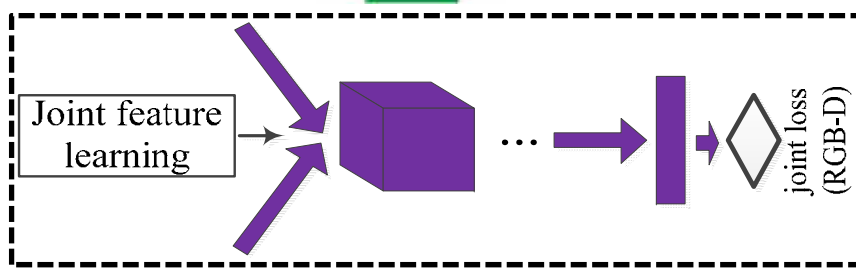


通过引入模态联合损失，增强模态间的交互

- 独立与联合损失共同驱动的互补特征学习[FG18]



基于Inception V2网络
构造联合损失的例子



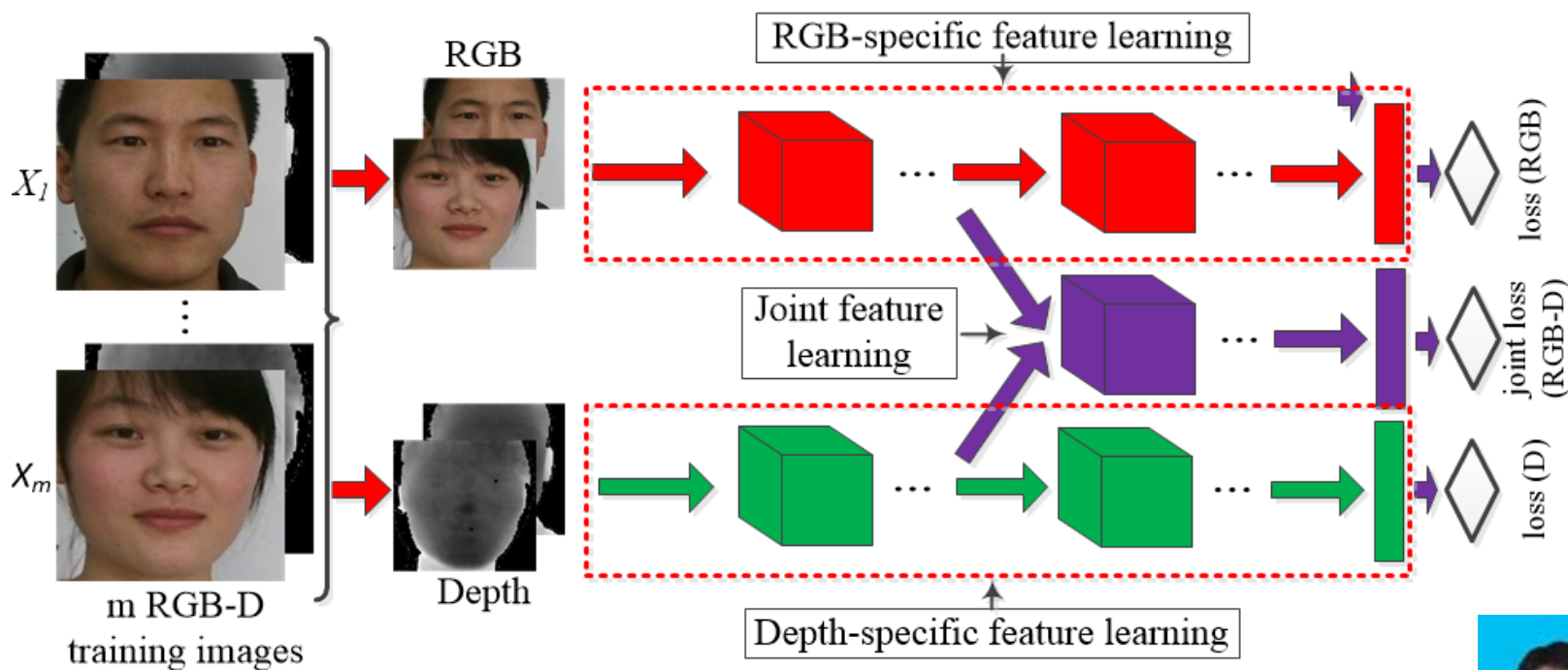


RGB-D多模态人脸识别

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

- 独立与联合损失共同驱动的互补特征学习[FG18]



同时保持模态个性化特征与模态共性特征



张浩 (硕)



RGB-D多模态人脸识别

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

■ RGB-D评测数据集

(a)									Kinect I 52人
(b)									Kinect I 106人
(c)									Kinect I 31人
(d)									Kinect II 509人
(e)									RealSense II 750人



RGB-D多模态人脸识别

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

- 信号层融合 vs. 分数层融合 vs. 互补特征学习
 - 基于Inception v2网络实现的上述3种RGB-D识别方法

融合方法	Our DB	BUAA	IIIT-D	Eurecom	VAP
信号层融合	91.5	97.3	97.6	96.0	82.2
相似度层融合	96.8	90.1	97.8	96.3	88.6
互补特征学习	97.3	90.8	98.6	96.3	90.8

基于AlexNet, VGG-16, ResNet-87等网络模型的实验得到相似的趋势



RGB-D多模态人脸识别

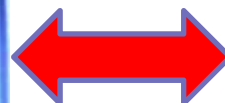
中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

- 基于ResNet-87网络设计的互补特征学习模型，在复杂场景条件下目前取得了99.1%的首选识别准确，比单独使用RGB模态提升2%



测试图像示例，~25万张
(姿态变化大，光照复杂)



注册图像示例
每人1张**准正面**注册图像

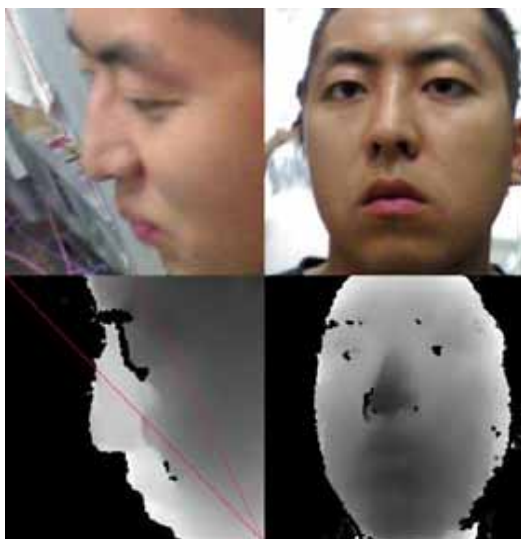


RGB-D多模态人脸识别

中科院计算所

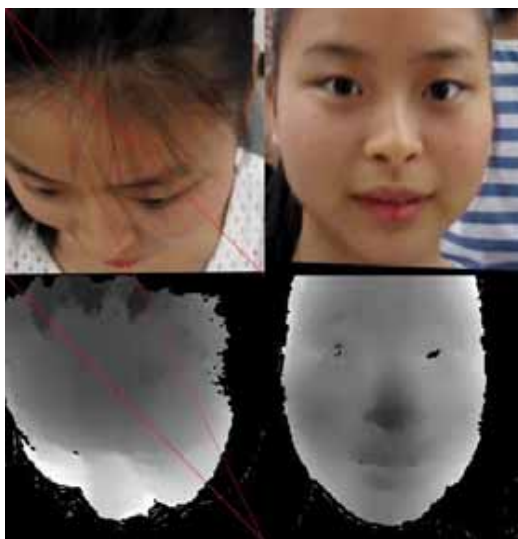
Institute of Computing Technology, Chinese Academy of Sciences

- 测试集中的首选识别错的0.9%图像示例



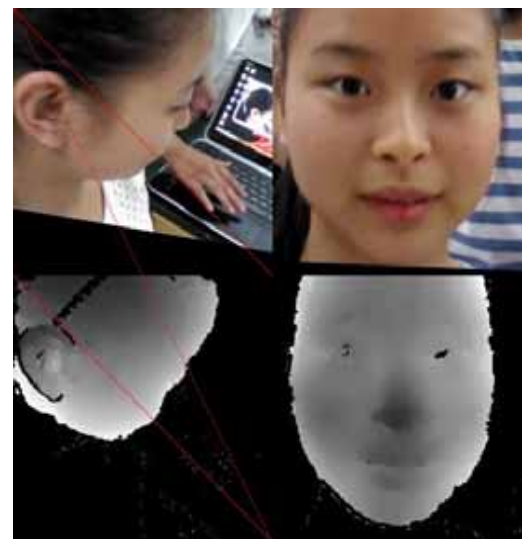
测试图像

注册图像



测试图像

注册图像



测试图像

注册图像



RGB-D多模态人脸识别

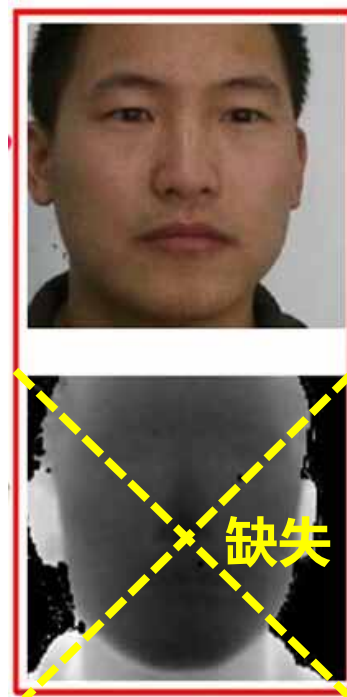
中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

- 模态出现缺失时如何处理？



测试图像

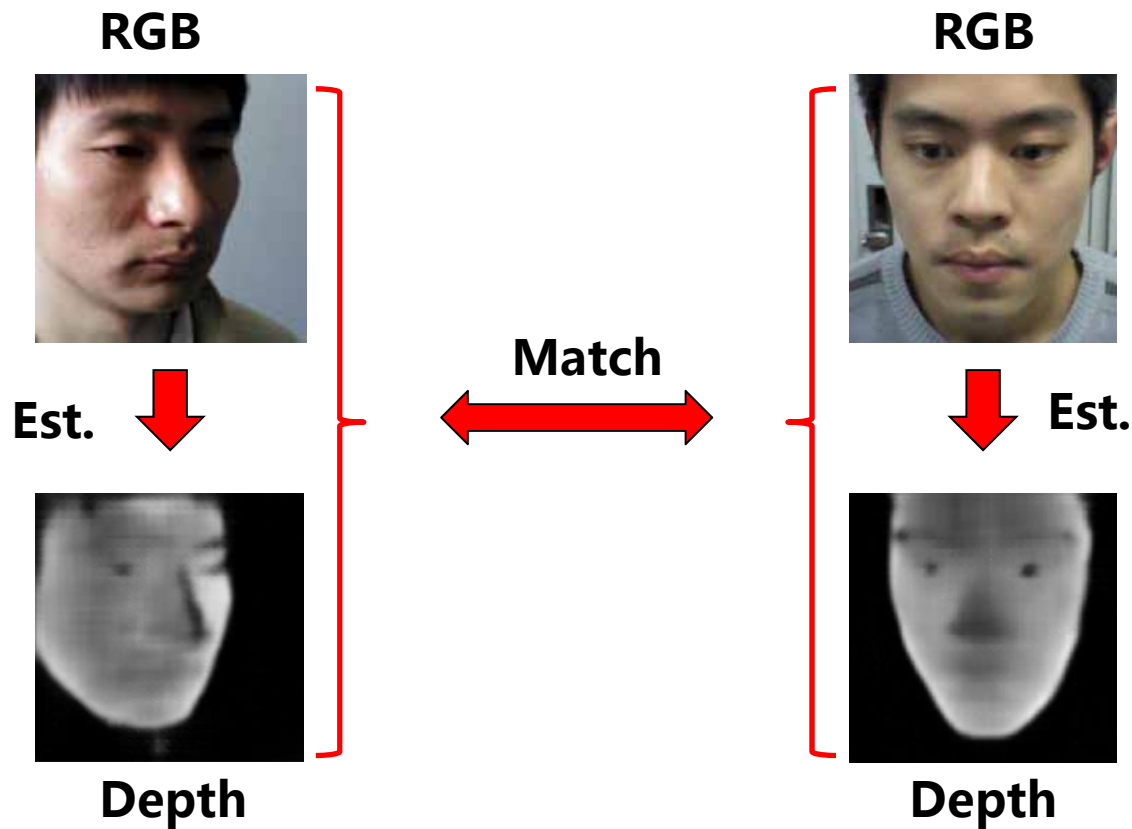


注册图像

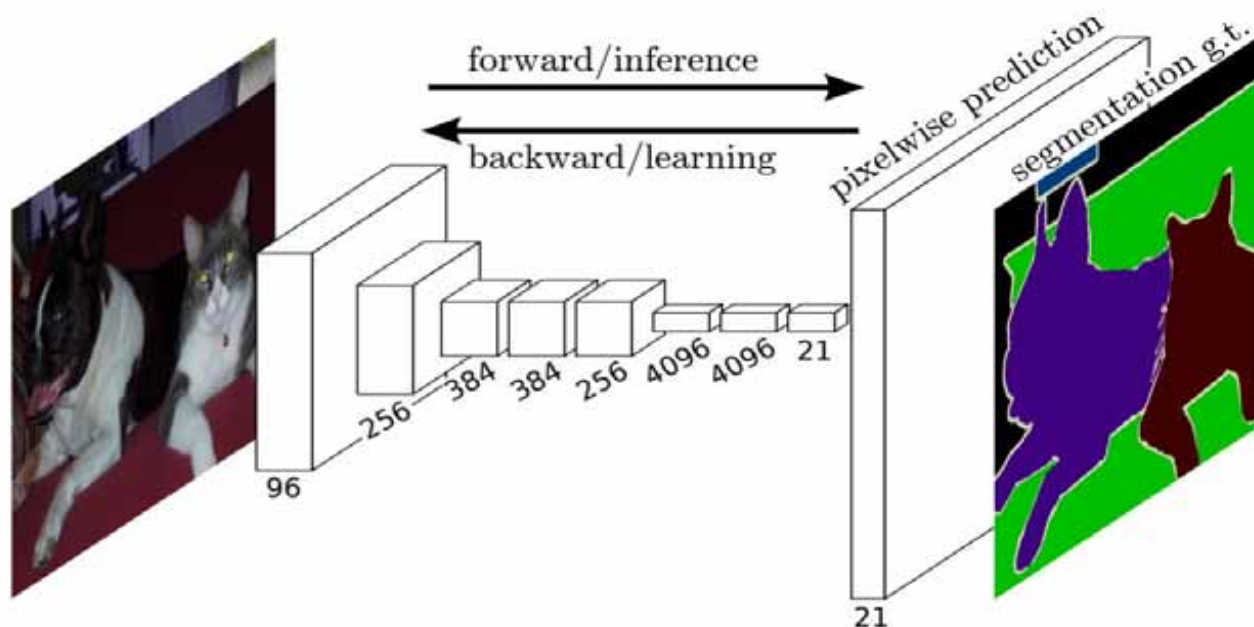


RGB-D多模态人脸识别无法回避的问题

■ 人脸深度生成



- 人脸深度生成-可能的方法
 - Fully convolutional network (FCN)



如何使生成的深度保持人与人之间的区分力？



RGB-D多模态人脸识别

- 保持区分力的人脸深度生成 [ICB18]

$$\min_{W_F, W_C} \underbrace{\|d - Z\|_2}_{\text{L2 loss}} + \log \underbrace{\sum_j e^{z_j} - z_y}_{\text{Softmax loss}}$$

$$Z = F_{FCN}(x, W_{FCN})$$

- 所生成的深度Z，既要和真实深度相比具有较小的误差，又要保持不同人之间的区分度



崔继运 (硕)

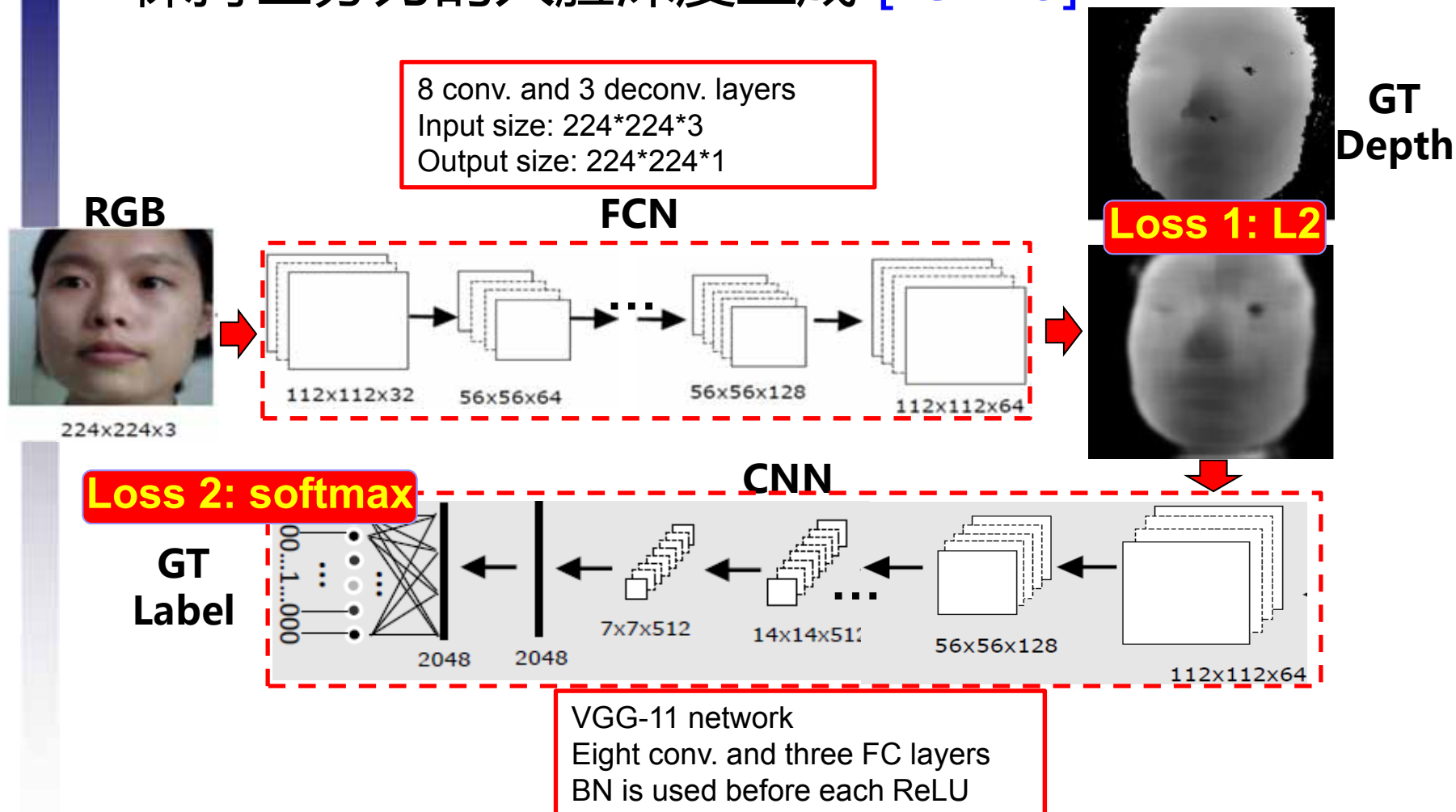


RGB-D多模态人脸识别

中科院计算所

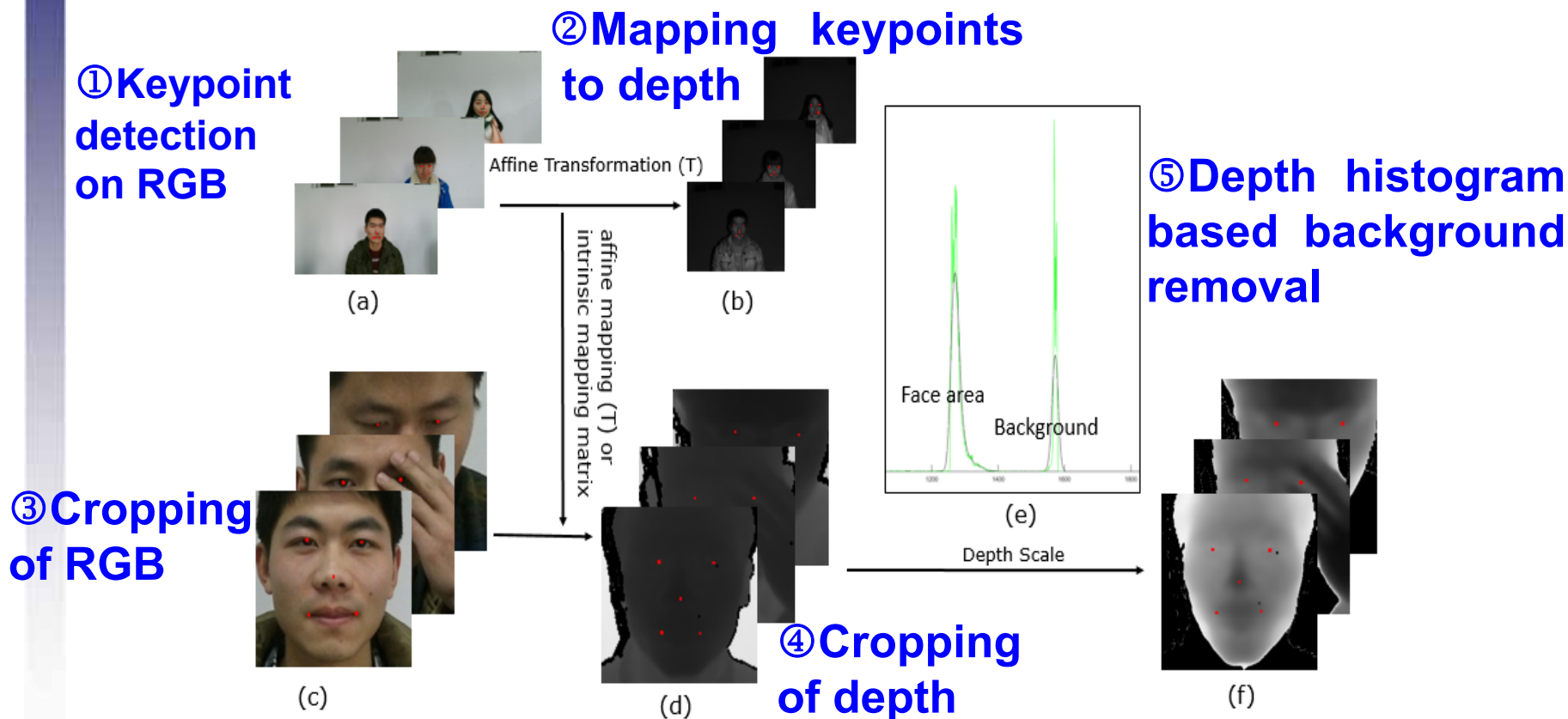
Institute of Computing Technology, Chinese Academy of Sciences

■ 保持区分力的人脸深度生成 [ICB18]



■ 保持区分力的人脸深度生成 [ICB18]

□ 数据预处理很重要！





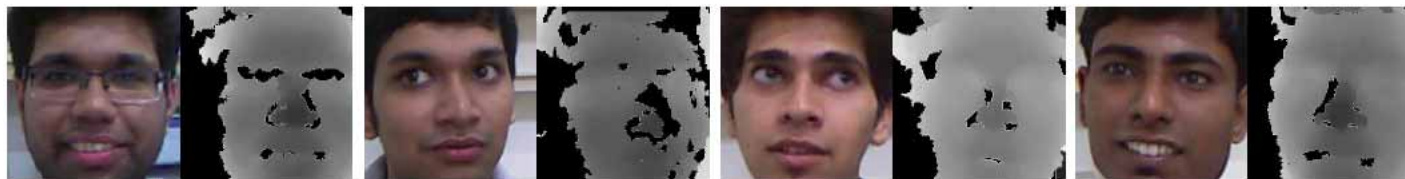
RGB-D多模态人脸识别

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

- 保持区分力的人脸深度生成 [ICB18]
 - 评测数据集

IIIT-D



BUAA



Our
internal



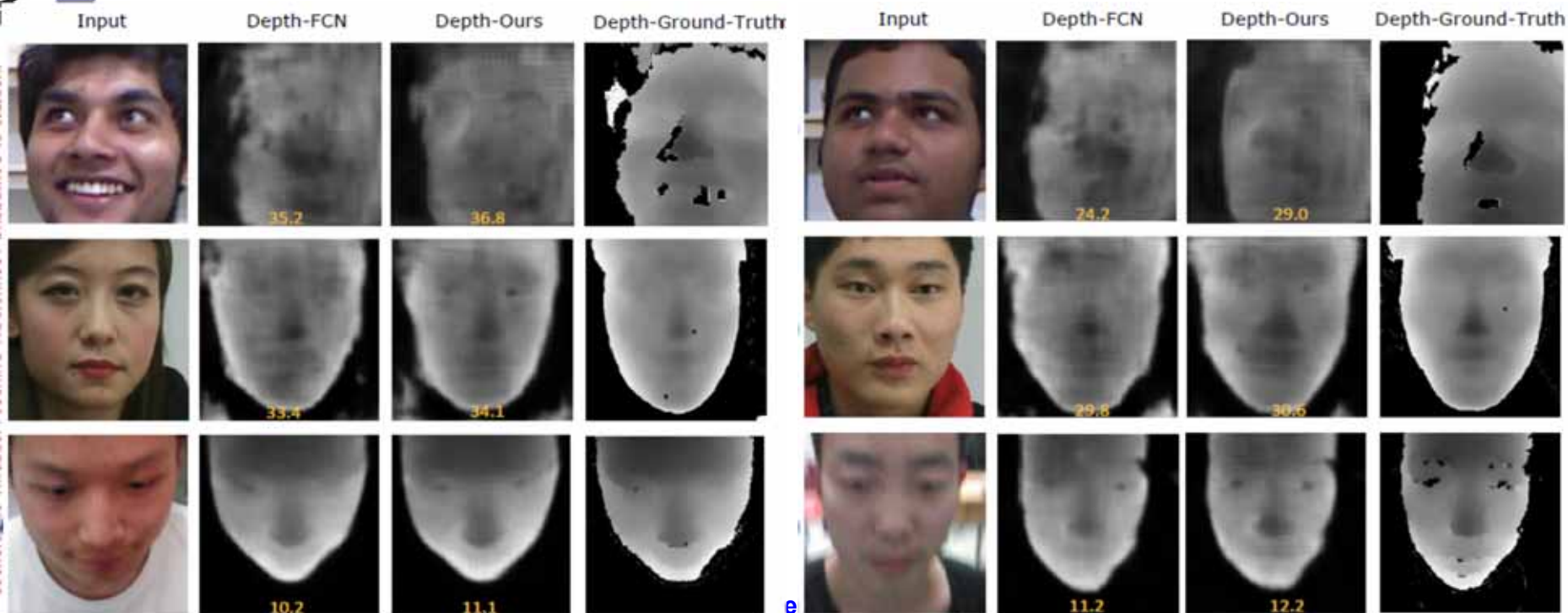


RGB-D多模态人脸识别

中科院计算所

深度生成结果的可视化与MAE误差

Method	Pixel-wise MAE		
	Lock3DFace	IIIT-D	Our Dataset
FCN [24]	22.3	40.2	12.9
Proposed approach	22.4	40.1	13.4





RGB-D多模态人脸识别

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

- 将所生成的深度用于RGB-D人脸识别

Method	BUAA	IIIT-D	Our dataset
RGB	94.5%	95.9%	93.3%
GT Depth	79.8%	78.8%	94.2%
FCN Depth	50.5%	69.6%	52.7%
Our Depth	80.1%	82.1%	80.8%
RGB+GT Depth	95.6%	96.5%	98.0%
RGB+FCN Depth	94.2%	95.7%	94.8%
RGB+Our Depth	94.9%	96.5%	96.1%

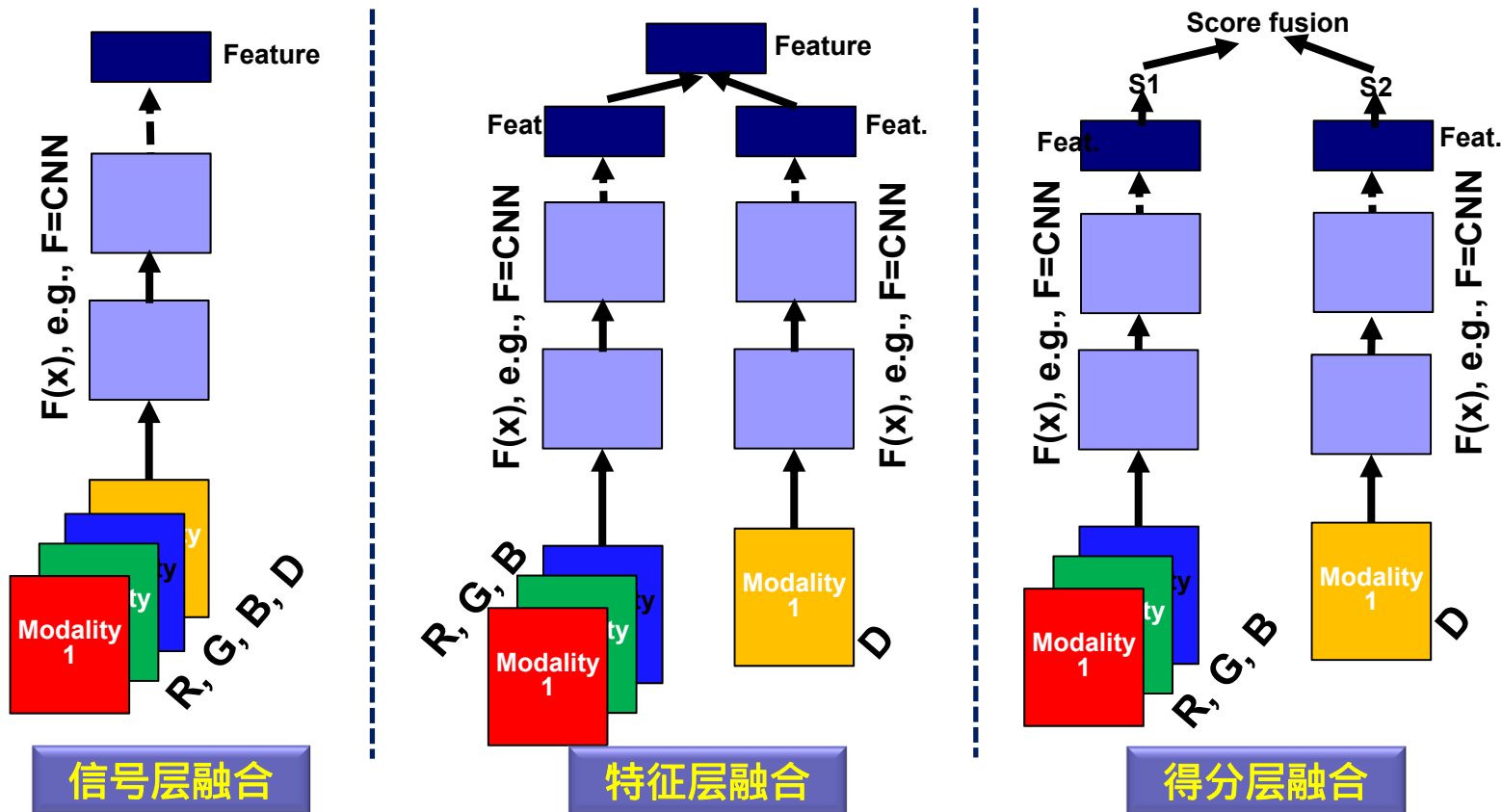
生成的深度可以提高人脸识别准确率！



RGB-D多模态人脸识别-小结

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences



1) 理论上，融合越靠近前端，信息损失越少 ≠ 在具体实现中前端融合更容易取得好的效果；

2) 从信息论的角度，生成深度并不能增加信息，但可以视作一种差异化的特征抽取路径，与直接从RGB图像抽取特征形成互补；



- 1 背景与研究现状
- 2 RGB-D多模态人脸识别 [FG'18, ICB'18]
- 3 表观属性分析 [CVPR'18, ICPR'18, TPAMI'17, TPAMI'15]
- 4 由表及里属性分析 [ICPR'18, IJCB'17]
- 5 总结

- 人脸图像除了表征身份之外，还蕴含着丰富的属性特征

视觉可见属性 (表观属性)

年龄: 40岁左右

性别: 男

种族: 白人

发型: 棕色短发

小胡子: 有

络腮胡: 有

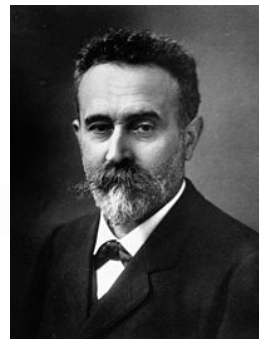
斑痣: 有

伤疤: 有



- 人脸表观属性分析的重要性/意义
 - 维基百科：Attribute is a characteristic of an object (person, thing, etc.)
 - 属性对于定义或立即不同的目标对象类别非常重要
 - 人脸为什么是人脸？
 - 眼睛、眉毛、鼻子、嘴巴、耳朵
 - 一张人脸图像为什么是张三的人脸图像？
 - 浓眉大眼、高鼻梁、鹅蛋脸、斑痣、纹身...





■ 人脸表观属性分析方法发展历程

- 1888年，法国生物特征识别专家贝蒂荣(Alphonse Bertillon)创建了一套人体测量学的生物特征识别方法用于犯罪嫌疑人认定
- 包含：年龄、眼睛颜色、颧骨宽度

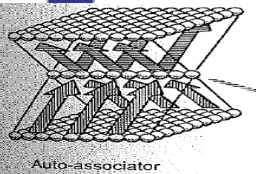


依赖于人工测量或抽取这些属性!



表观属性分析

人脸表观属性分析研究发展历程



MIT: Cottrell & Metcalfe
把基于Auto-Endoder的特征降维用于性别和表情识别

PCA特征

北卡: Ricanek & Tesafaye
构建了首个大规模年龄、性别、种族数据库MORPH (1.3万人, 5.5万图像)

生物启发特征+SVM

MSU: Han & Jain
首次研究了人与机器在属性识别上的性能差异(可控), 并发现机器在年龄、性别和种族的识别上已经可以超过人类

多任务、多标记

计算所: Han et al.; UMD: Rama;
基于深度多任务学习的属性学习; 构建并发布了LFW+数据集(表观年龄)

1990

1999

2006

2008

2013-15

2014-15

2017

塞浦路斯学院: Lanitis构建了FGNET年龄估计数据库 (82人, 1002张图像)

PCA特征

哥大: Kumar等人构建了包含10个属性的大规模名人数据库PubFig (6万图像, 200人) 仅部分公开

手工设计特征+SVM

NIST组织了人脸的年龄和性别分析评测

港中文: Tang et al. 构建了大规模互联网名人的40属性数据集CelebA (20万图像)

深度特征+SVM



表观属性分析

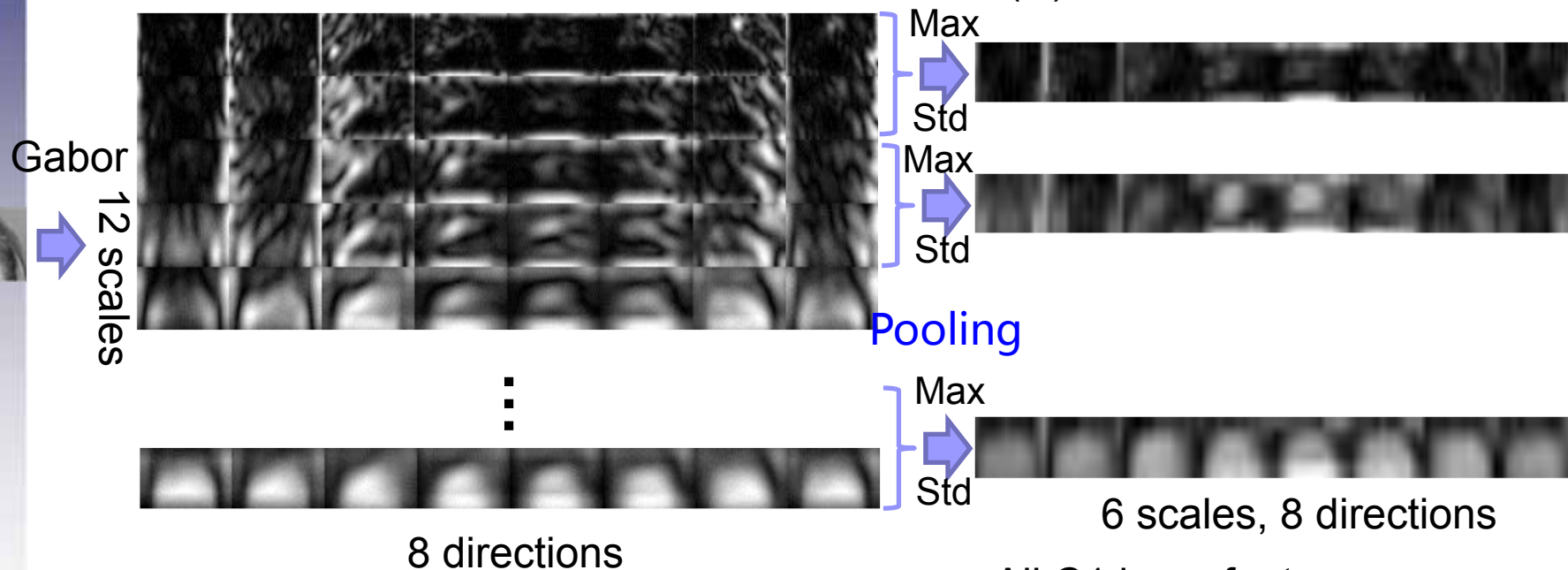
- 人脸表观属性分析研究发展历程
 - 手工设计特征→深度特征
 - 组合式方法→端到端学习
 - 单任务学习→多任务/多标记学习
- 驱动力：需求与数据的变化
 - 简单场景→复杂场景
 - 小规模有限个体的数据集→大规模数据集
 - 年龄、性别等单个或少数属性→数十种属性

■ 人脸属性判别特征学习 [ICB13, TPAMI 15]

□ 生物启发的属性特征表示-BIF [Guo CVPR09]

S1 层: Simulate the simple (S) cell units

C1 层: Simulate the complex (C) cell units



BIF能够捕捉大量人脸图像细节特征，但可能包含大量的冗余信息

All C1 layer features are concatenated into a 4280D feature vector



■ 人脸属性判别特征学习 [ICB13, TPAMI 15]

□ 形式化

$$\mathbf{D} = \{(\mathbf{x}_i, y_i) : \mathbf{x}_i \in \mathbb{R}^d, y_i \in \mathbb{N}, i \in [1, m]\}$$

D: 原始的细节特征丰富
但可能存在冗余的特征



$$\mathbf{S} = \{\mathbf{x}'_i : \mathbf{x}'_i \in \mathbb{R}^{d'}, \mathbf{x}'_i \subset \mathbf{x}_i, i \in [1, m]\}$$

$$d' \ll d$$

S: 对多个人脸属性判别
有针对性的特征

□ 基于Boosting方法的特征选择



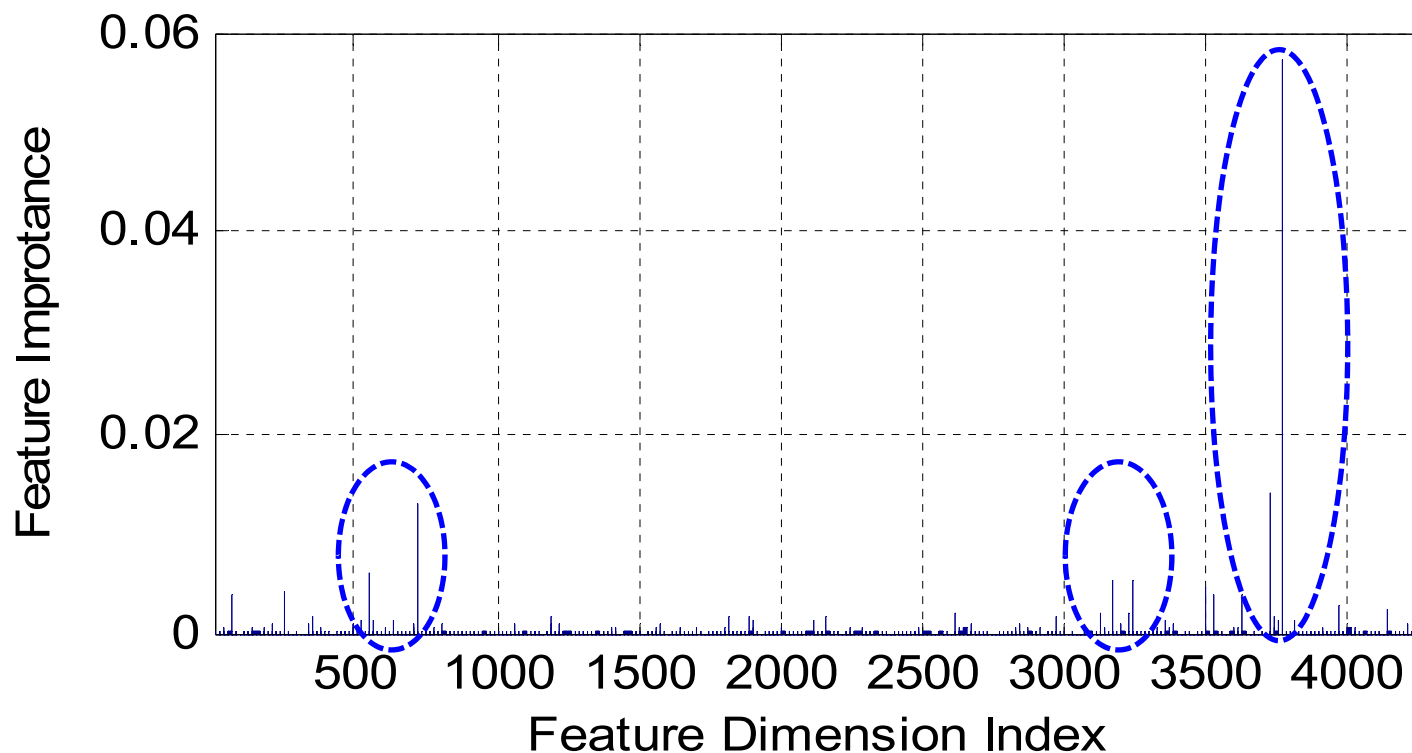
表观属性分析

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

■ 人脸属性判别特征学习 [ICB13, TPAMI 15]

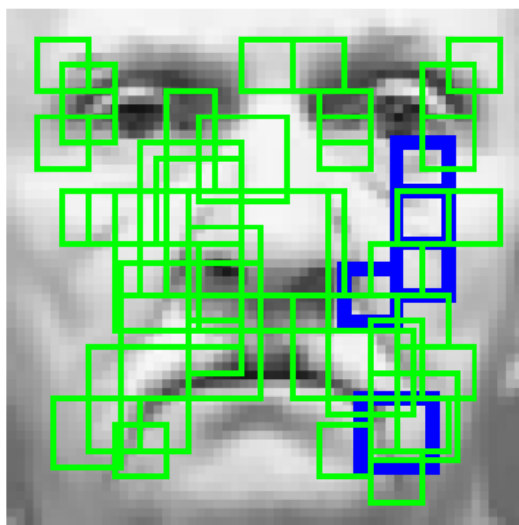
□ 可视化



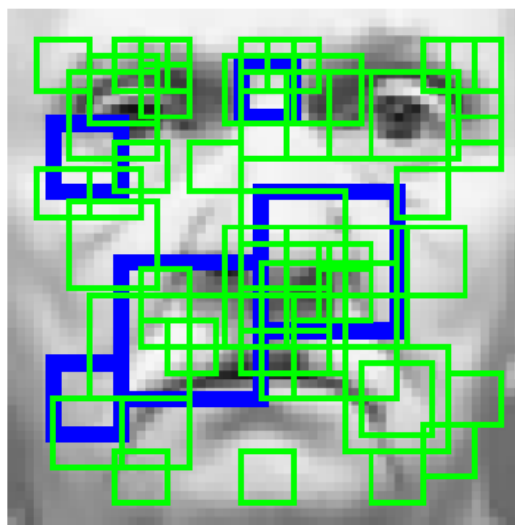
Selected 800 out of 4280 dimensions

■ 人脸属性判别特征学习 [ICB13, TPAMI 15]

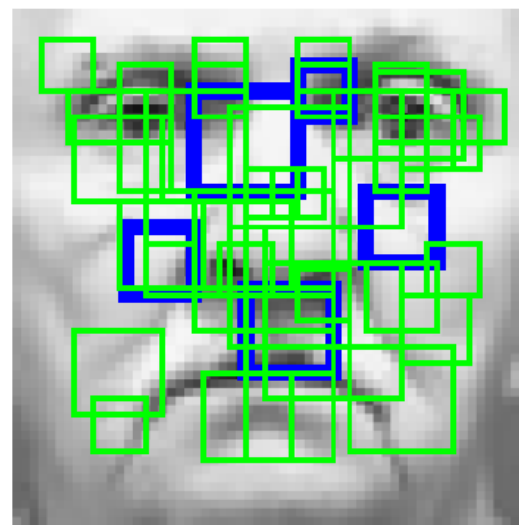
□ 可视化



年龄估计任务



性别估计任务

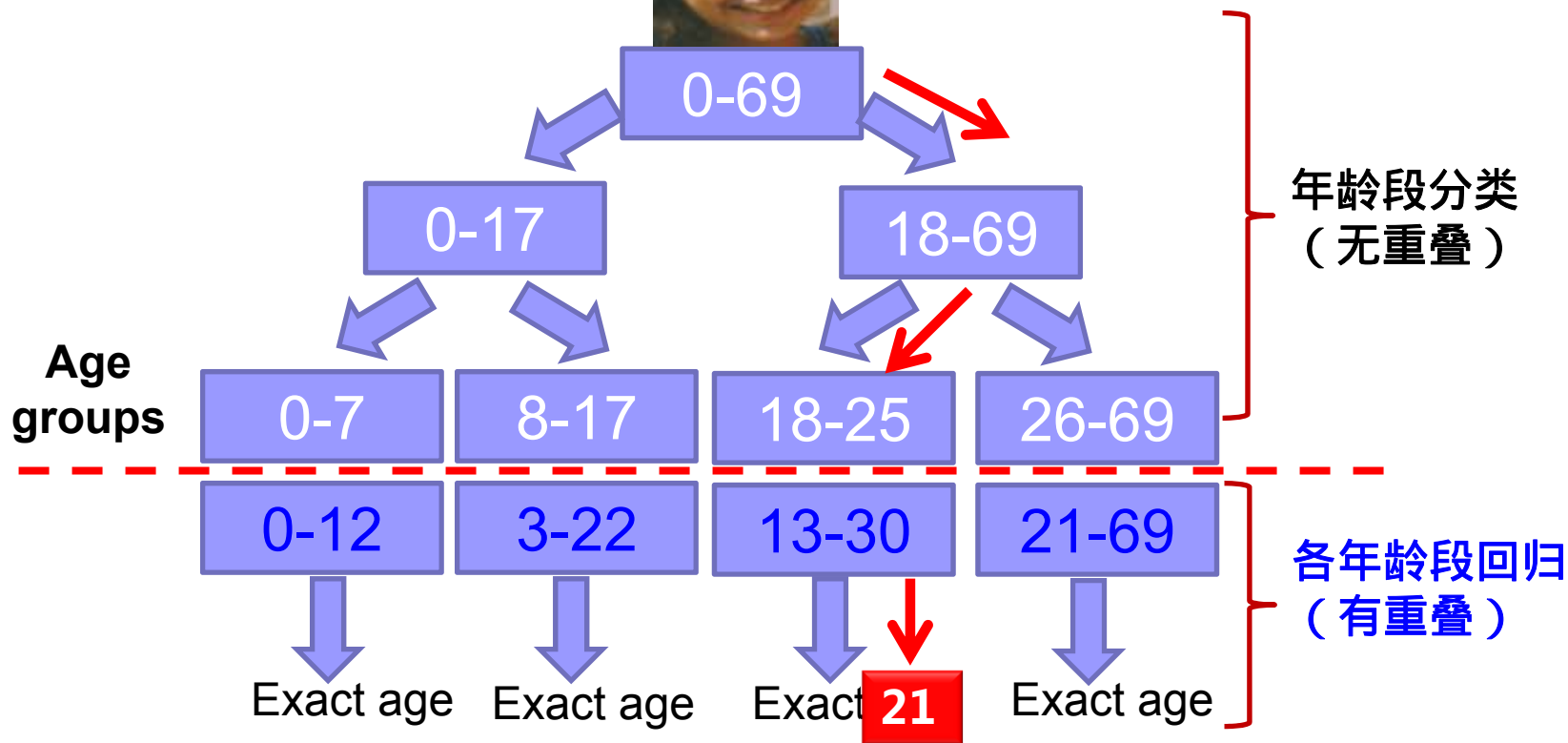


种族估计任务

蓝色框: top 5 important features;
绿色框: top 6-50 important features;

■ 人脸属性判别特征学习 [ICB13, TPAMI 15]

□ 层级式年龄分类器





表观属性分析

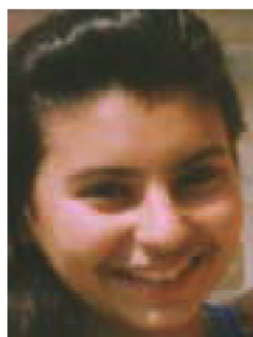
■ 人脸属性判别特征学习 [ICB13, TPAMI 15]

□ 算法的性能 vs. 人的能力的对比

- 基于Amazon Mechanical Turk获得人对年龄、性别、种族的估计结果
- 每张图像至少3个人标注，然后进行投票或取均值



How many years old is the person in the image? Enter your answer using digits 0-9 only.



What is the gender of the person in the image? Give your answer by checking one gender category.

☐ Male ☐ Female



What is the race of the person in the image? Give your answer by checking one race category.

☐ White ☐ Black



表观属性分析

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

- 人脸属性判别特征学习 [ICB13, TPAMI 15]
 - NIST 2014年的FRVT报告直接引用了我们总结的年龄估计结果

Publication	MAE (years)	CS(5)
Luu et al. [18]	4.1	73%
Chao et al. [7]	4.4	NA
Chang et al. [6]	4.5	74.7%
Han et al. [15]	4.6	74.8%
Choi et al. [8]	4.7	73%
Guo et al. [14]	4.8	47%
Wu et al. [23]	5.9	62%
Suo et al. [19]	6.0	55%
Thukral et al. [20]	6.2	NA
Geng et al. [9]	6.8	65%

(a) Published methods [15],
using LOPO testing protocol



表观属性分析

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

- 人脸属性判别特征学习 [ICB13, TPAMI 15]
 - 人与算法在人脸年龄、性别、种族估计中的对比

任务	数据库	算法	人类
Age estimation	FGNET	3.8 yr.	4.7 yr.
	MORPH	3.6 yr.	6.3 yr.
Gender classification	FERET	96.8%	n/a
	MORPH	97.6%	96.9%
Race classification	MORPH	99.1%	97.8%
	PCSO	98.7%	96.5%

在大规模数据集上取得了超越人的性能！

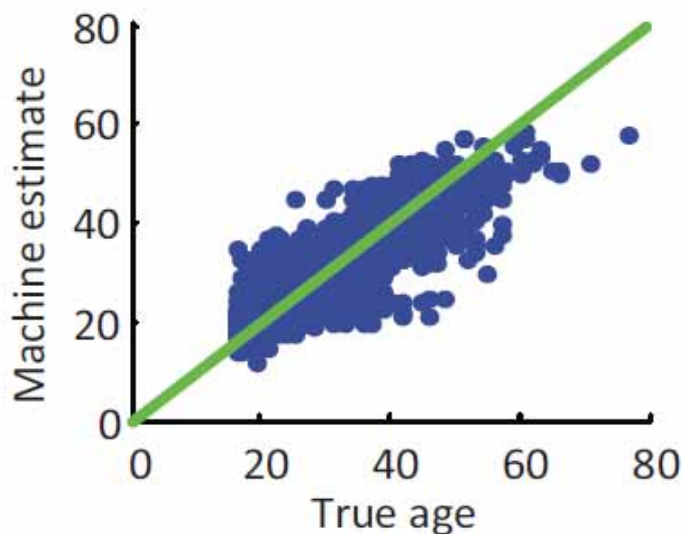


表观属性分析

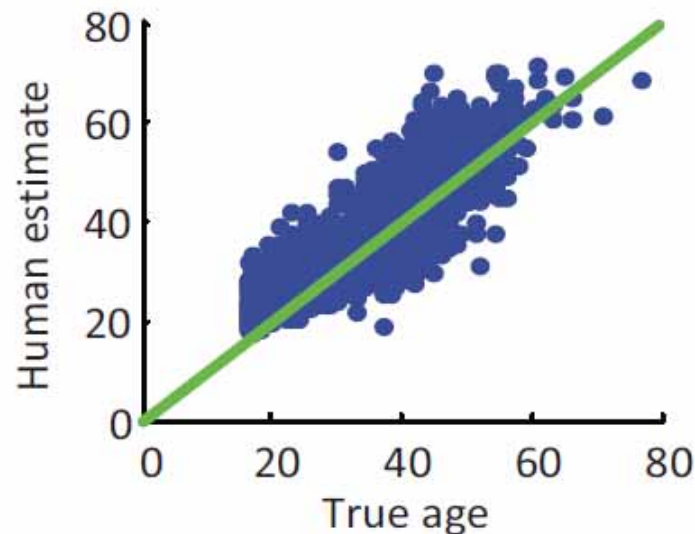
中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

- 人脸属性判别特征学习 [ICB13, TPAMI 15]
 - 人对年龄估计的偏差



算法的年龄估计



人类的年龄估计

- 大量人脸属性的估计



模型



28-year

male

white

eye glasses

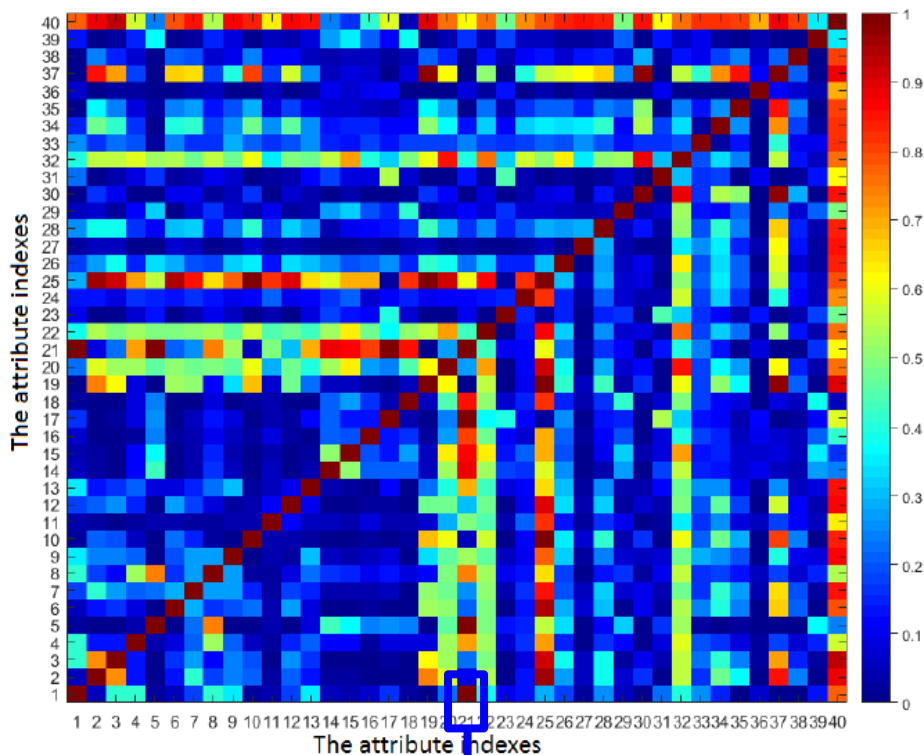
short hair

...

- 每个属性构建一个模型？

- 数据缺乏、分布不均衡、标注不一致

大量人脸属性的估计



CelebA 中 40 个属性的共生矩阵

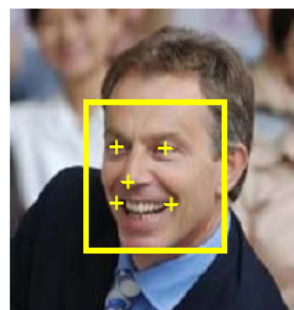
(5 O'ClockShadow, Male)



表观属性分析

中科院

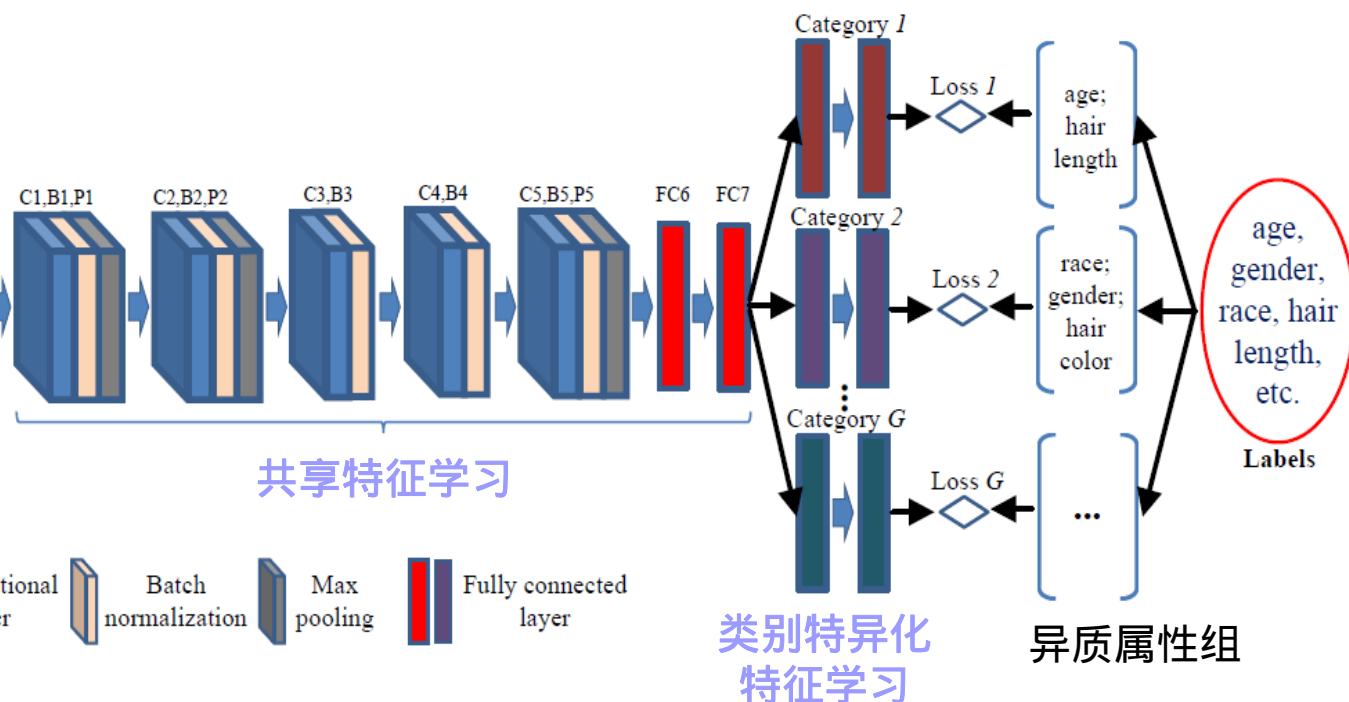
深度多任务属性学习 [FG17, TPAMI 17]



Detection



Alignment

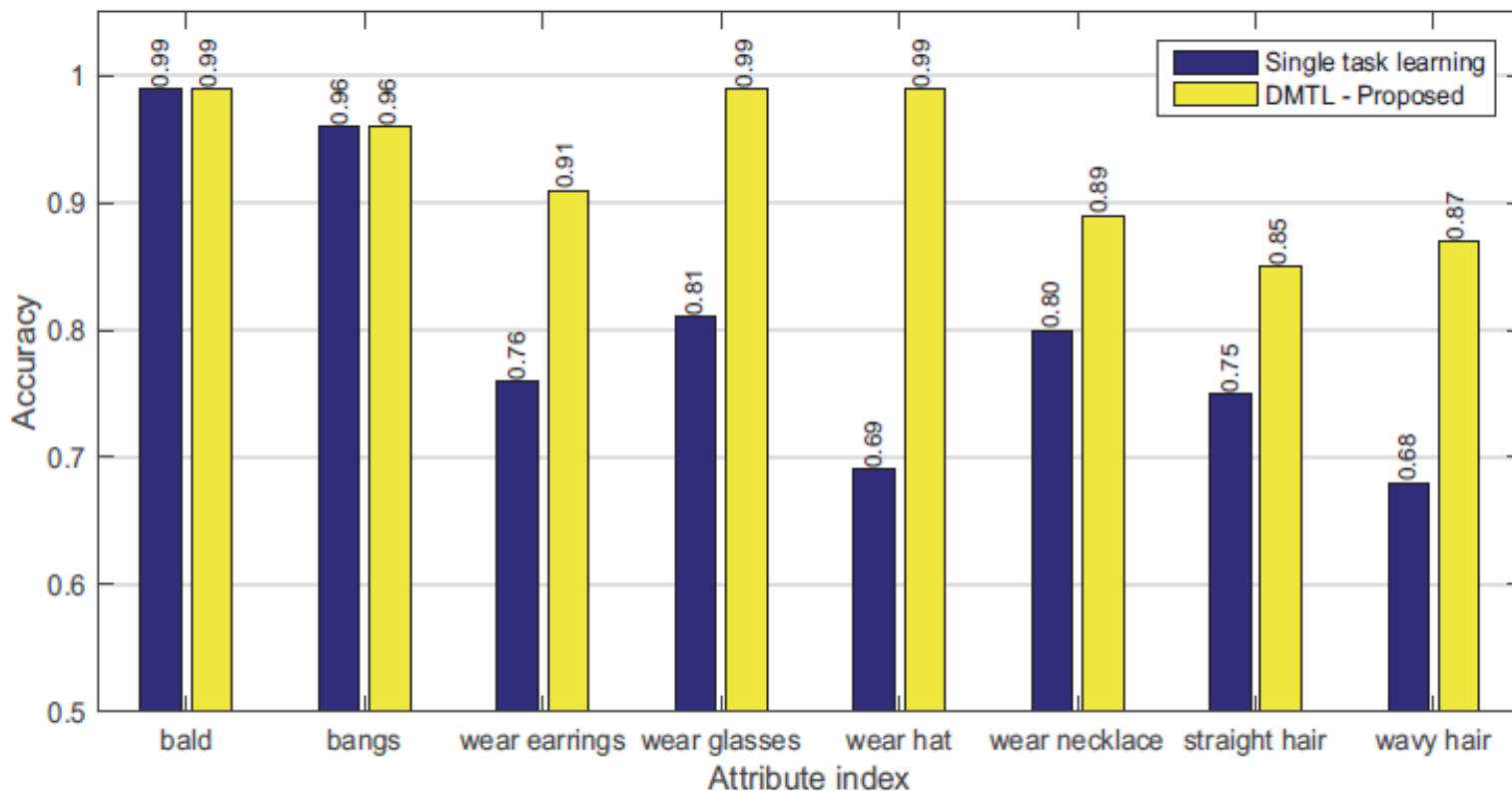


- 深度多任务属性学习 [FG17, TPAMI 17]
 - 主要结果—年龄、性别、种族估计

Approach	MORPH II			LFW+		
	Age ²	Gender	Race	Age ²	Gender	Race
Guo and Mu [19]	3.92/70.0	98.5	99.0	NA	NA	NA
Yi <i>et al.</i> [20]	3.63/NA	98.0	99.1	NA	NA	NA
DIF [16]	3.8/75.0	97.6	99.1 ³	7.8/42.5 ⁴	94 ⁴	90 ^{3,4}
DEX [17]	3.25/NA	NA	NA	NA	NA	NA
DEX [17] ¹	2.68/NA	NA	NA	NA	NA	NA
Proposed	3.0/85.3	98.0	98.6	4.5/75.0	96.7	94.9

1 The IMDB-WIKI database was used for network pre-training.

- 深度多任务属性学习 [FG17, TPAMI 17]
 - 主要结果—多任务学习vs.单任务学习



■ 深度多任务属性学习





表观属性分析

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

- 实现准确的年龄估计对人和算法都具有挑战性



35岁左右



45岁左右

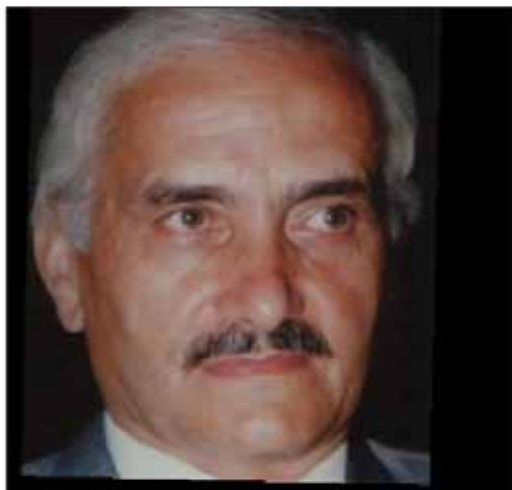
They are identical twins separated at birth, and grow up in different homes/environments.



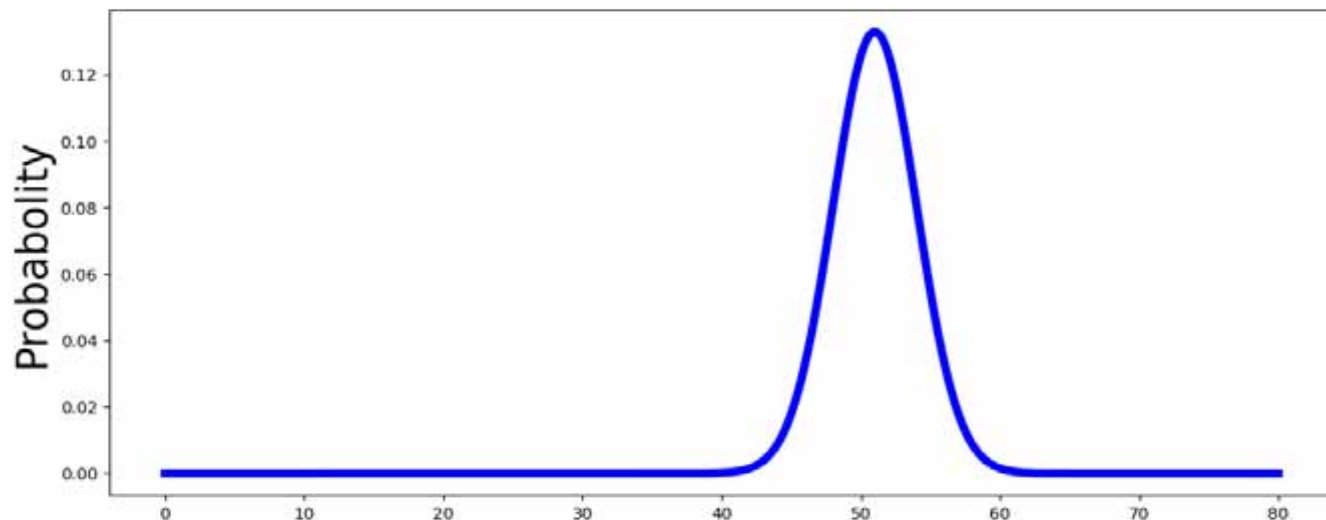
表观属性分析

中科院

- 基于分布学习的年龄估计
 - 用一个年龄分布(如高斯分布)代替单一年龄值



Age: 51



目前基于分布学习的年龄估计方法通常假设年龄分布的均值和方差是已知的。这种假设对于通过众包收集的表观年龄数据是合理的，但对生理年龄而言，只能人为设定某种方差。



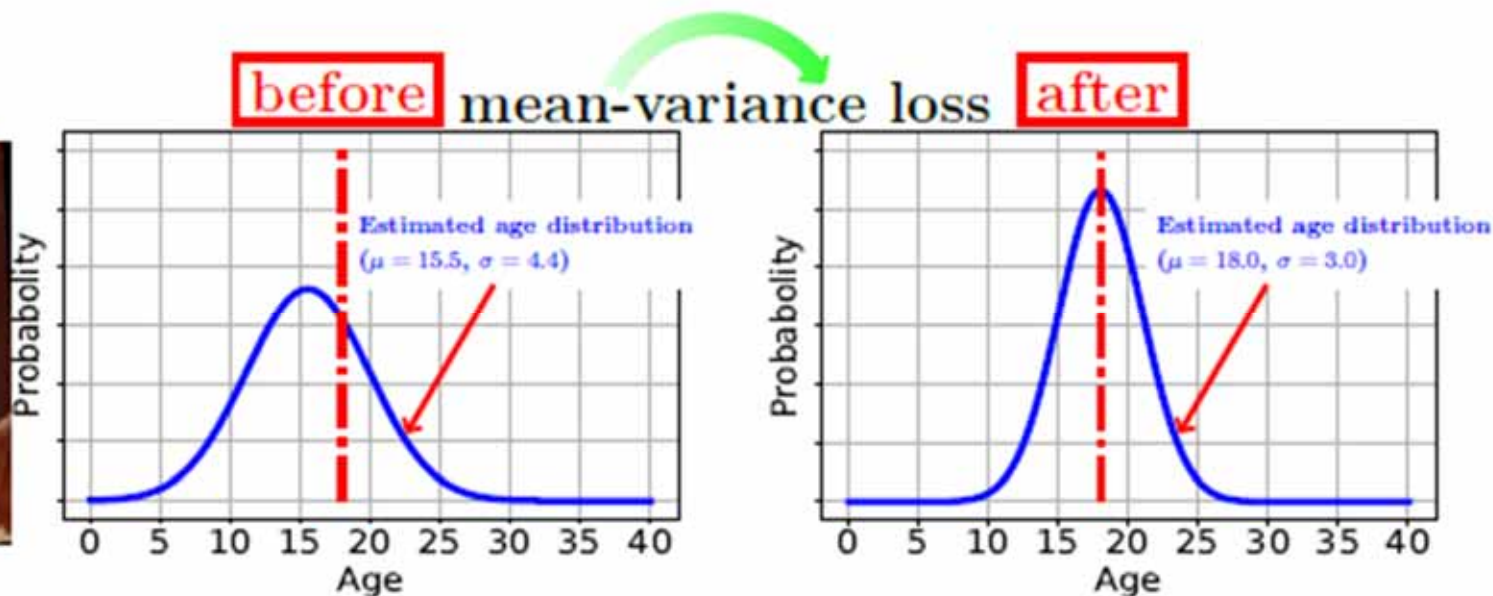
表观属性分析

中科院计算所

- 基于均值-方差损失的年龄分布学习 [CVPR18]
 - 形式化：同时最小化年龄分布的均值与真实年龄值之间的误差，以及年龄分布的方差；
 - 不需要给定/假设年龄的标准差



age: 18



■ 基于均值-方差损失的年龄分布学习 [CVPR18]

□ 形式化

✓ 均值损失部分

$$\mathcal{L}_m = \frac{1}{2N} \sum_{i=1}^N (m_i - y_i)^2 = \frac{1}{2N} \sum_{i=1}^N \left(\sum_{j=1}^N j * p_{i,j} - y_i \right)^2$$

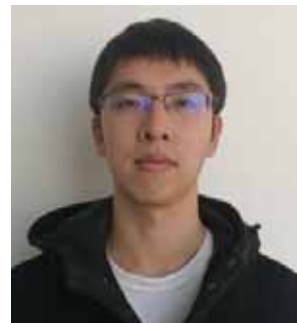
✓ 方差损失部分

$$\mathcal{L}_v = \frac{1}{N} \sum_{i=1}^N v_i = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^K p_{i,j} * \left(j - \sum_{k=1}^K k * p_{i,k} \right)^2$$

✓ 梯度求导

$$\frac{\partial \mathcal{L}_m}{\partial z_{i,j}} = \frac{(m_i - y_i)}{N} p_{i,j} (j - m_i)$$

$$\frac{\partial \mathcal{L}_v}{\partial z_{i,j}} = \frac{p_{i,j}}{N} ((j - m_i)^2 - \alpha)$$



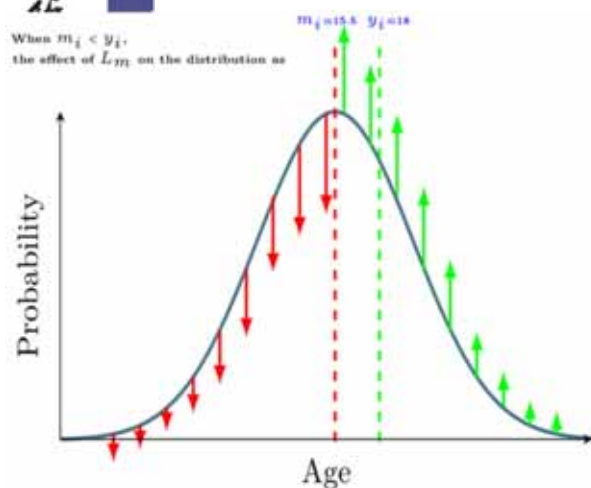
潘虹宇 (硕)



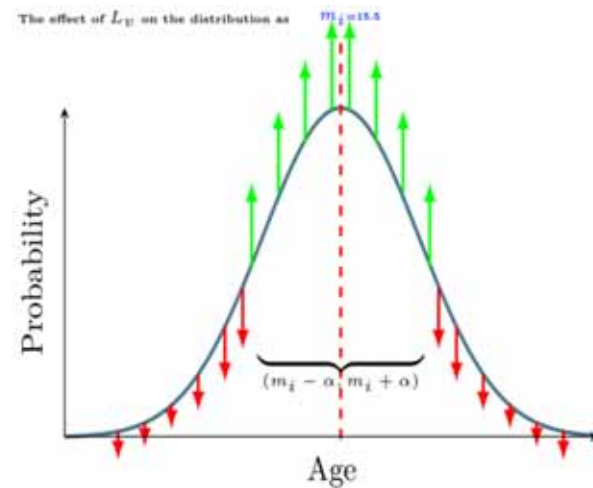
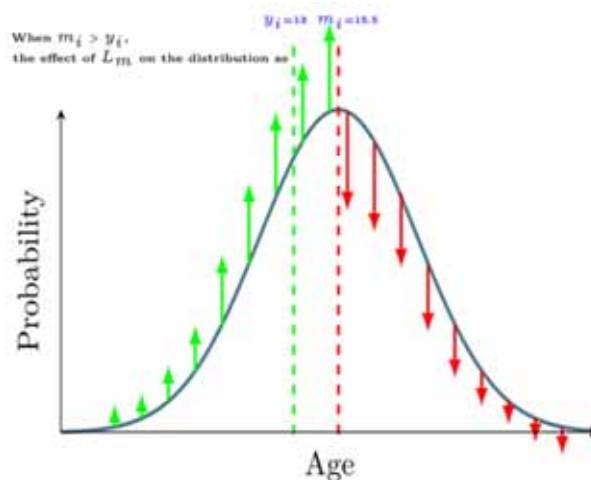
表观属性分析

中科院

- 基于均值-方差损失的年龄分布学习 [CVPR18]
 - 均值-方差损失的对分布学习的三个作用



驱动年龄分布向正确年龄平移



促使分布更趋集中

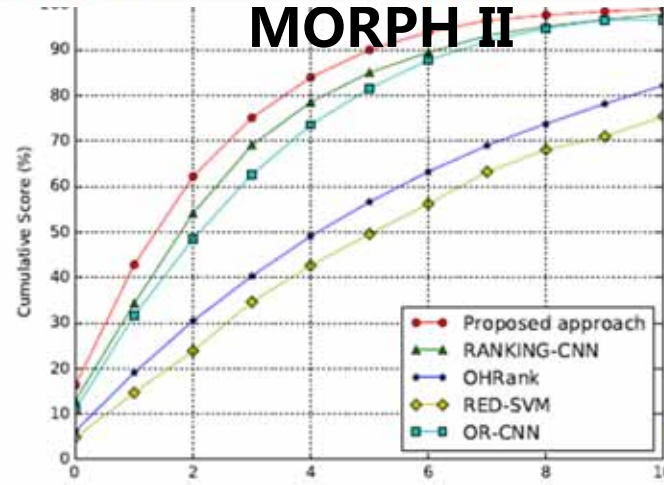
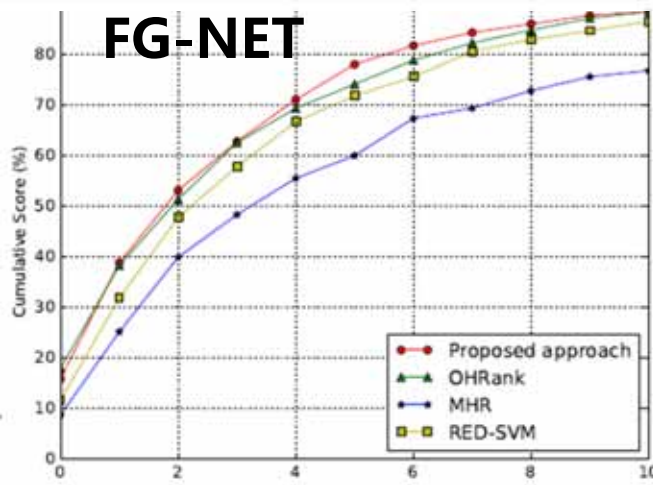
不同的年龄值对损失计算的贡献程度不同

■ 基于均值-方差损失的年龄分布学习 [CVPR18]

□ 年龄估计结果

Method	MAE	Protocol
RED-SVM [3]	5.24	LOPO
OHRank [4]	4.48	LOPO
DEX [38]	4.63	LOPO
DEX*	3.09	LOPO
Proposed	4.10	LOPO
Proposed*	2.68	LOPO

Method	MAE	Protocol
RED-SVM [3]	6.49	RS
OHRank [4]	6.07	RS
OR-CNN [33]	3.27	RS
DEX [38]	3.25	RS
DEX*	2.68	RS
DIF [16]	3.00	SE
Ranking-CNN [5]	2.96	RS
Proposed	2.41/2.80	RS/SE
Proposed*	2.16/2.79	RS/SE





表观属性分析

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

- 基于均值-方差损失的年龄分布学习 [CVPR18]
 - 方法的推广能力→图像审美
 - AADB数据库

Method	Cls	Reg	[27]	EMD [22]	Ours
ρ	0.5923	0.6239	0.6782	0.6682	0.6647

欢迎来我们的
海报前交流！

- 人脸表观属性的研究近年来发展迅速，深度学习与多任务学习相结合的方法取得了好的效果



- 准确年龄估计仍然有进一步提升的空间，特别是复杂场景下的准确年龄估计
- 学术界亟需大规模的复杂场景年龄数据集

- 1 背景与研究现状
- 2 RGB-D多模态人脸识别 [FG'18, ICB'18]
- 3 表观属性分析 [CVPR'18, ICPR'18
TPAMI'17, TPAMI'15]
- 4 由表及里属性分析 [ICPR'18, IJCB'17]
- 5 总结

- 人脸图像除了表征身份之外，还蕴含着丰富的属性特征

视觉可见属性 (表观属性)

年龄: 40岁左右
性别: 男
种族: 白人
发型: 棕色短发
小胡子: 有
络腮胡: 有
斑痣: 有
伤疤: 有



视觉不可见属性 (内在/生理属性)

心率: 72bpm
血压: 120/80 mmHg



由表及里属性分析

中科院计算所

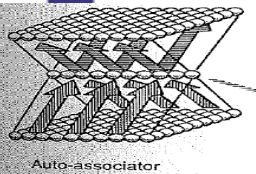
Institute of Computing Technology, Chinese Academy of Sciences





由表及里属性分析

中国科学院



人脸属性分析研究发展历程

PCA特征

ICA + FFT

生物启发特征+SVM

MIT: Cottrell & Metcalfe
把基于Auto-Encoder的特征降维用于性别和表情识别

北卡: Ricanek & Tesafaye
构建了首个大规模年龄、性别、种族数据库MORPH (1.3万人, 5.5万图像)

MIT: Pho等人
首次研究了基于普通摄像头的**非接触式**心率估计

MSU: Han & Jain
首次研究了人与机器在属性识别上的性能差异(可控), 并发现机器在年龄、性别和种族的识别上已经可以超过人类

1990

1999

2006

2008

2010

2013-15

2014-15

“由表及里”

塞浦路斯学院: Lanitis构建了FGNET年龄估计数据库 (82人, 1002张图像)

哥大: Kumar等人
构建了包含10个属性的大规模名人数据库PubFig (6万图像, 200人) 仅部分公开

NIST组织了人脸的年龄和性别分析评测

港中文: Tang et al.
构建了大规模互联网名人的40属性数据集CelebA (20万图像)

PCA特征

手工设计特征+SVM

深度特征+SVM



由表及里属性分析

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

- 心率：心脏每分钟跳动的次数

健康
水平

人的静息心率
在每分钟50-
100次之间

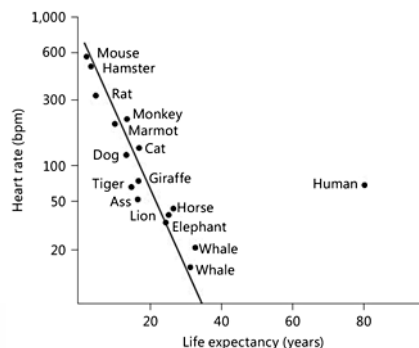


PresenterMedia

情报、审讯、
心理战（测谎
仪收集的数据
中包含心率、
血压等）

情绪
状态

寿命



哺乳动物寿
命与静息心
率成反比¹

1. <https://www.karger.com/Article/FullText/435947>; <http://www.chinanews.com/jk/2012/09-29/4222475.shtml>;



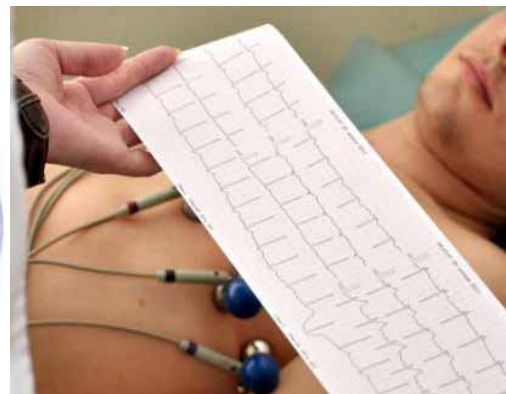
由表及里属性分析

中科院计算所



人工
计数

心电图



心率
带



智能
手表



- 1) 都需接触皮肤
(卫生/过敏)
- 2) 带有侵扰性
(需要用户配合)

从脸上看出你的心跳数？

- 基于光电容积描记法 (photoplethysmography, rPPG) 的心率测量

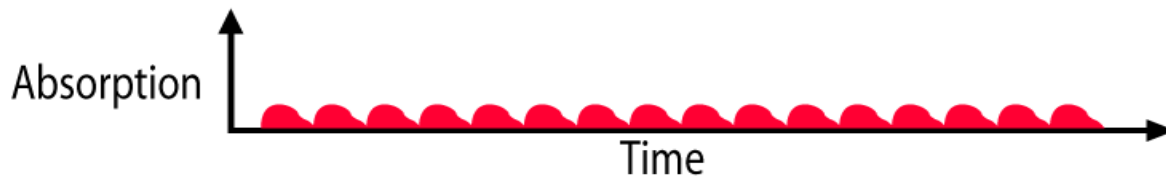


心跳带动动脉血
周期性涌动



放大后的
肤色变化

周期性动脉血涌
动引起微弱的周
期性肤色变化



肤色变化：每秒1次
普通摄像头：每秒
25-30 FPS

通过分析单位时间内人脸颜色变化的周期数目，可以得出心率数



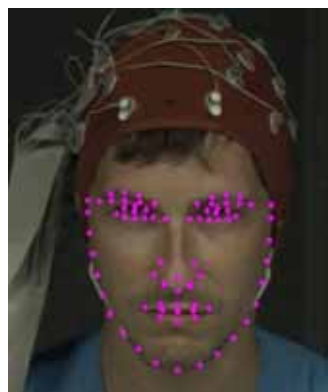
基于rPPG的无感式心率测量

■ 基于rPPG的无感式心率测量 [IJCB17]

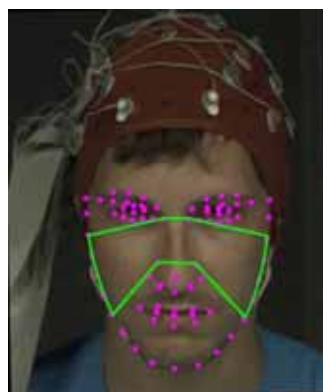
- **基本方法**：找人脸、定五官、选区域、调颜色、算周期



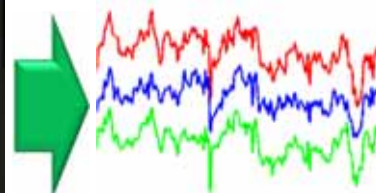
找人脸



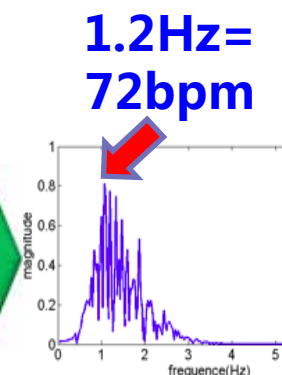
定五官



选区域



调颜色



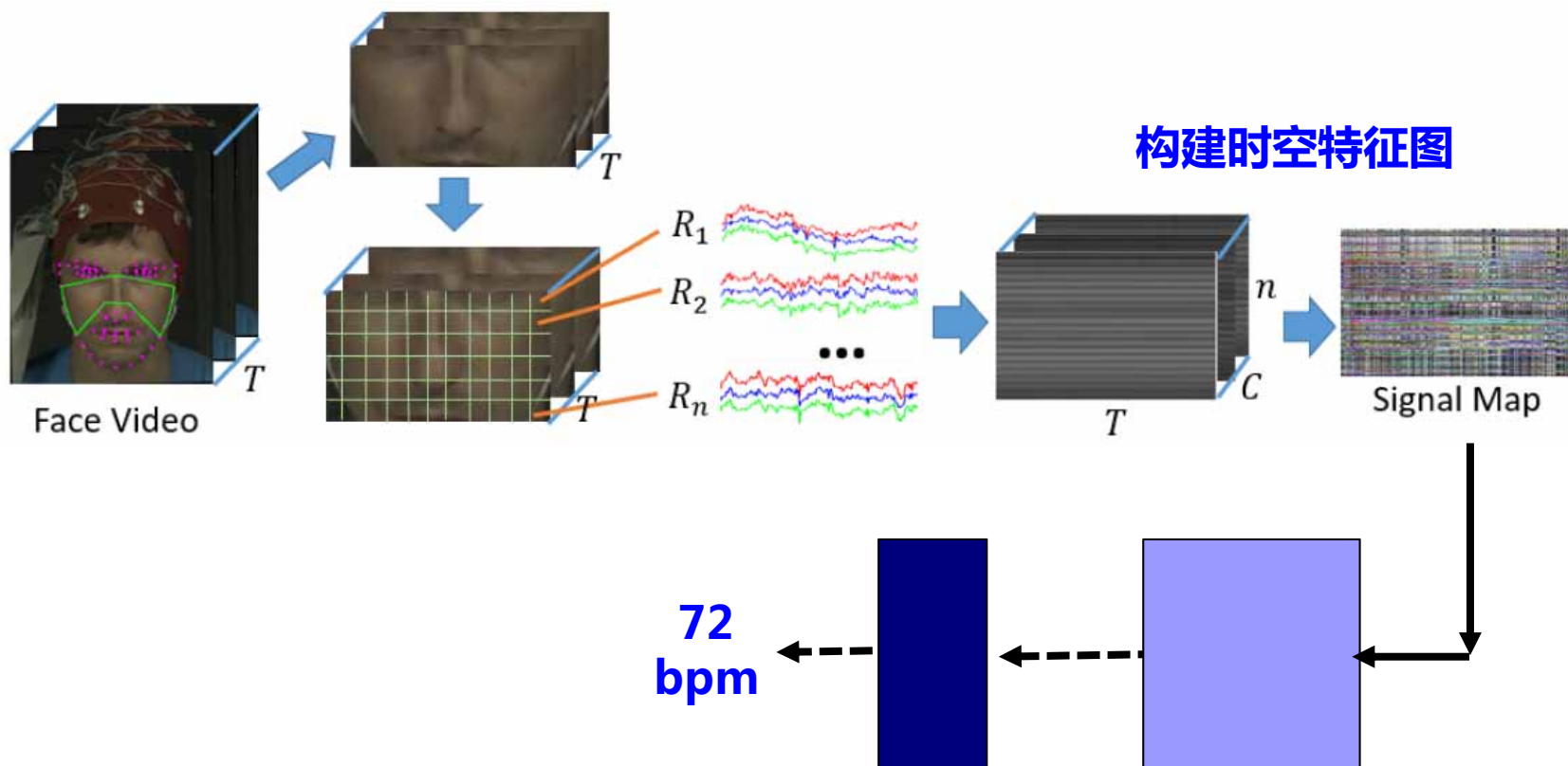
算周期

完全基于信号分析技术，不需要学习



牛雪松 (直博)

■ 基于端到端学习的心率估计 [ICPR18]





真能做到吗？

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences



医用指夹式心率仪





总结

- 小型化和低成本的新型RGB-D传感器技术发展，推动了RGB-D多模态人脸识别的研究，基于RGB-D的人脸识别的安全度达到了相当于6位数字口令的安全度水平（1:100万）
- 多模态人脸识别仍然缺乏**简洁高效**的多模态信息抽取方法，特别是单网络结构的多模态特征学习网络设计
- 学术界仍然缺少公开的大规模、复杂场景多模态数据集，在一定程度上限制了新方法的探索



总结

- 近年来针对人脸/人体属性分析研究的关注度较高；对于一些数据收集容易、分布相对均衡的属性，属性分析的准确度达到了实用的基本要求（性别、戴眼镜、戴帽子）
- 大规模人脸属性数据集的标注仍然是一个难题，标注过程中易产生歧义和不一致性，部分属性的数据严重不均衡

■ 创建并公开的一些数据集

□ LFW+

- ✓ 除了LFW13000+图像的基础上，增加了2000张左右青少年人脸图像
- ✓ 通过亚马逊众包标注了每张人脸图像的年龄、性别、种族
- ✓ 用途：表观年龄估计、多属性分析

□ FG-NET Apparent Age

- ✓ 通过亚马逊众包标注了每张人脸图像的年龄
- ✓ 用途：表观年龄估计与真实年龄估计差异对比

□ 人脸心率估计数据库

- ✓ 超过100人，每人超过20段视频，包含连续GT心率值
- ✓ 包含了光照、姿态变化
- ✓ 预计几个月之后发布

□ <https://sites.google.com/site/huhanhomepage/download>



中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

感谢关注！ Questions?