

# 低质量视觉物体学习理解



中国科学院 信息工程研究所  
INSTITUTE OF INFORMATION ENGINEERING, CAS

葛仕明

CCF-CV走进高校@济南大学，2020年11月28日

# 引子

## IBM沃森首次登陆美国医疗界 医生饭碗堪忧

2017.2.17

李敬敬 (编译) 健康界 2017-02-17

### 摘要

认知计算平台Watson for Oncology将于3月初进军美国医疗界，该消息引发不少议论与猜测。这一智能平台究竟有何“神技”？未来是否会抢医生的“饭碗”呢？

来源：健康界 作者：李敬敬（编译）

近日，IBM Watson Health宣布其认知计算平台Watson for Oncology将于今年3月初在美国佛罗里达州（Florida）的木星医疗中心（Jupiter Medical Center）正式上线，**该项诊疗支持技术虽早已入驻印、韩、泰等国，但却是首次登陆美国公立医院。**事实上，不少医生对此技术颇有微词，担心其未来会抢了自己的“饭碗”。

认知计算平台究竟是什么？其融入医疗领域面临何种挑战？美权威医疗信息网站Becker's Hospital Review就此对IBM Watson Health 副总裁Anil Jain博士进行了深度采访。

今日头条 首页 / 科技 / 正文

2018.5.31

3

## IBM大裁员，AI落地太难

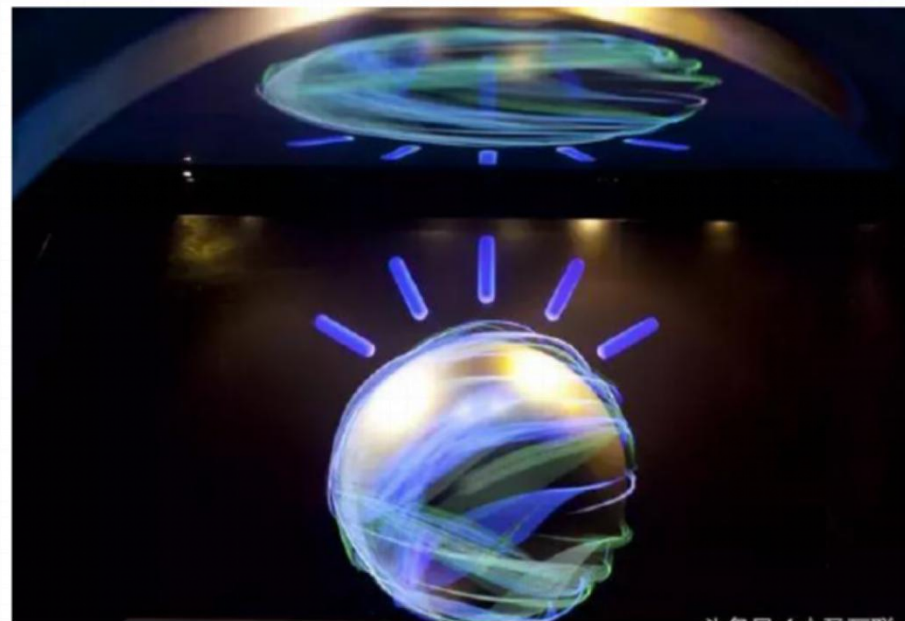
中盈互联 2018-05-31 17:49:22

微博

Qzone

微信

源 | 商业技术评论 文 | rebootman



# 引子

---



“Mission: Impossible”

# 目录

---

深度模型  
知识蒸馏

蒙面人脸  
检测识别

低分辨率  
人脸识别

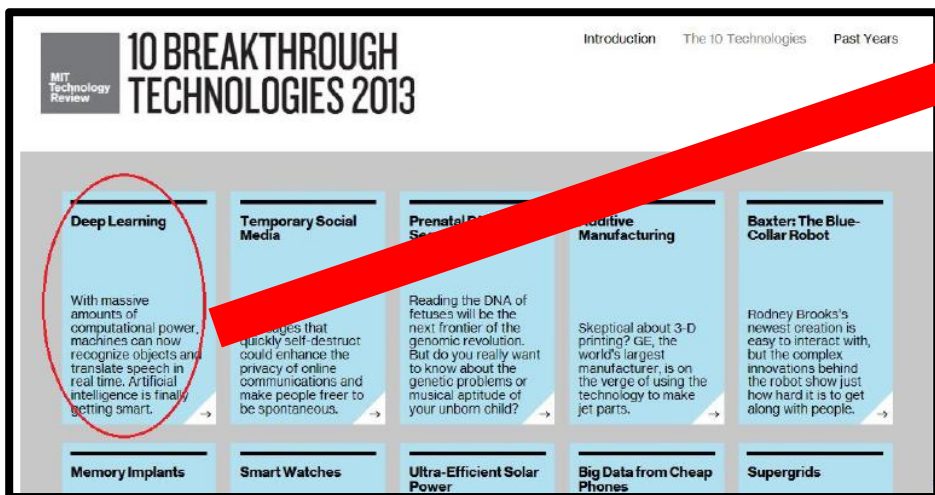
弱小目标  
深度跟踪

# 深度学习在很多AI任务上取得了突破

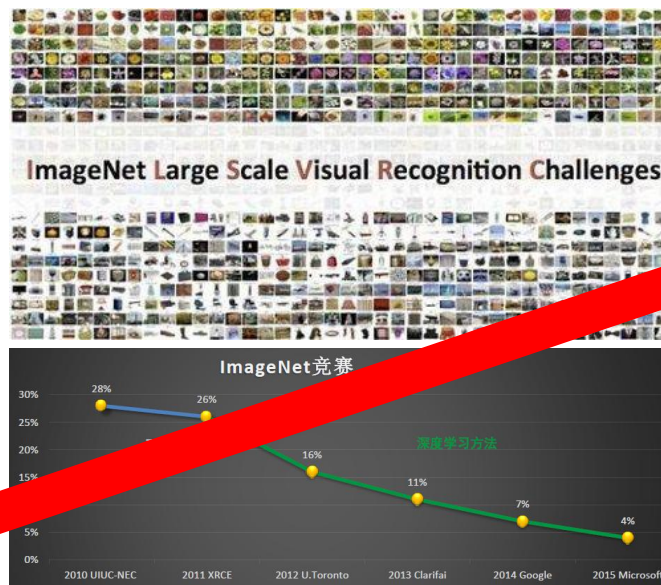


“没有可与深度学习竞争的人工智能技术”

——Yoshua Bengio



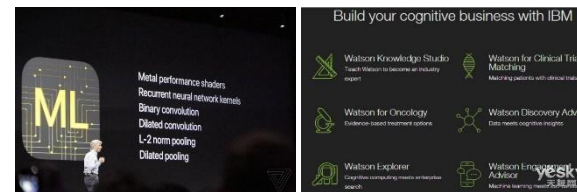
2013: 深度学习被评为10大科技突破之首



2014-2015: 在通用任务和特定任务都取得极大进展（超过人眼水平）



2016: AI在围棋游戏击败人类



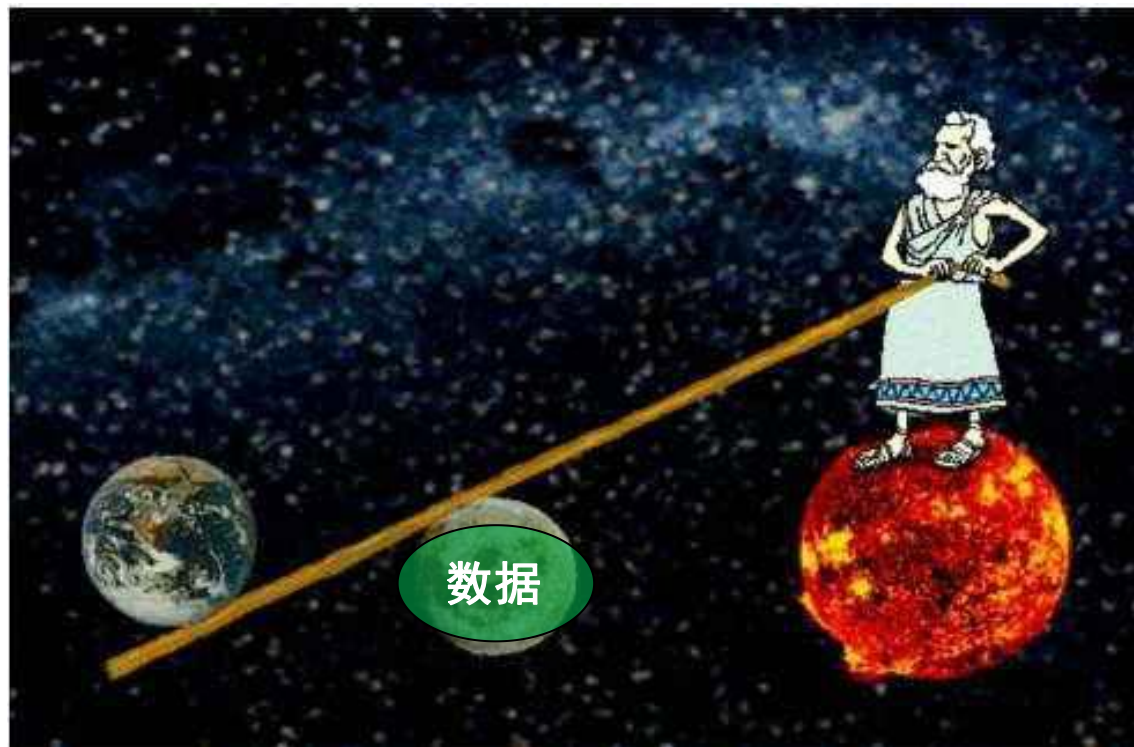
2017, 谷歌、百度、苹果、IBM全面推进AI战略

AI First  
All in AI  
AI in AI  
.....

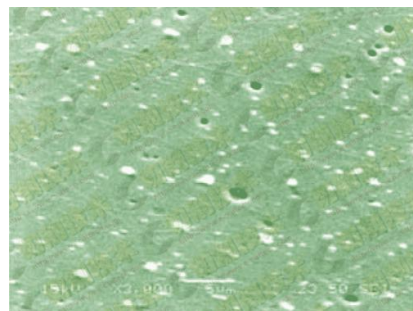
世界充满“爱”(AI)

海量高质的数据是当前深度学习成功的关键

# 实际场景中难以获得海量高质的数据



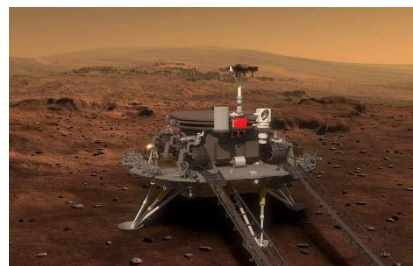
“给我一个支点（数据），我就能撬动地球”



工业表面缺陷



城市视频监控



火星探测

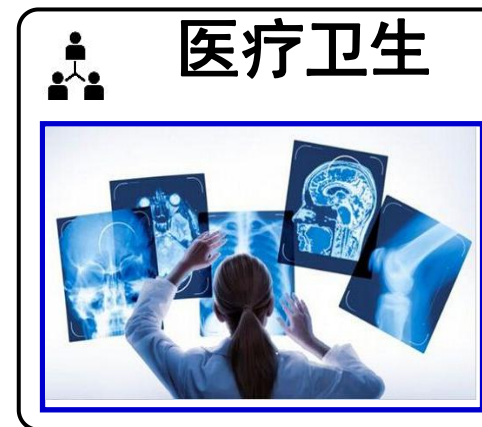


医学数据



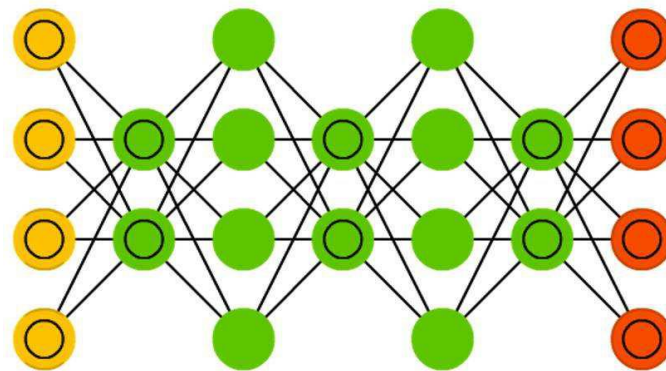
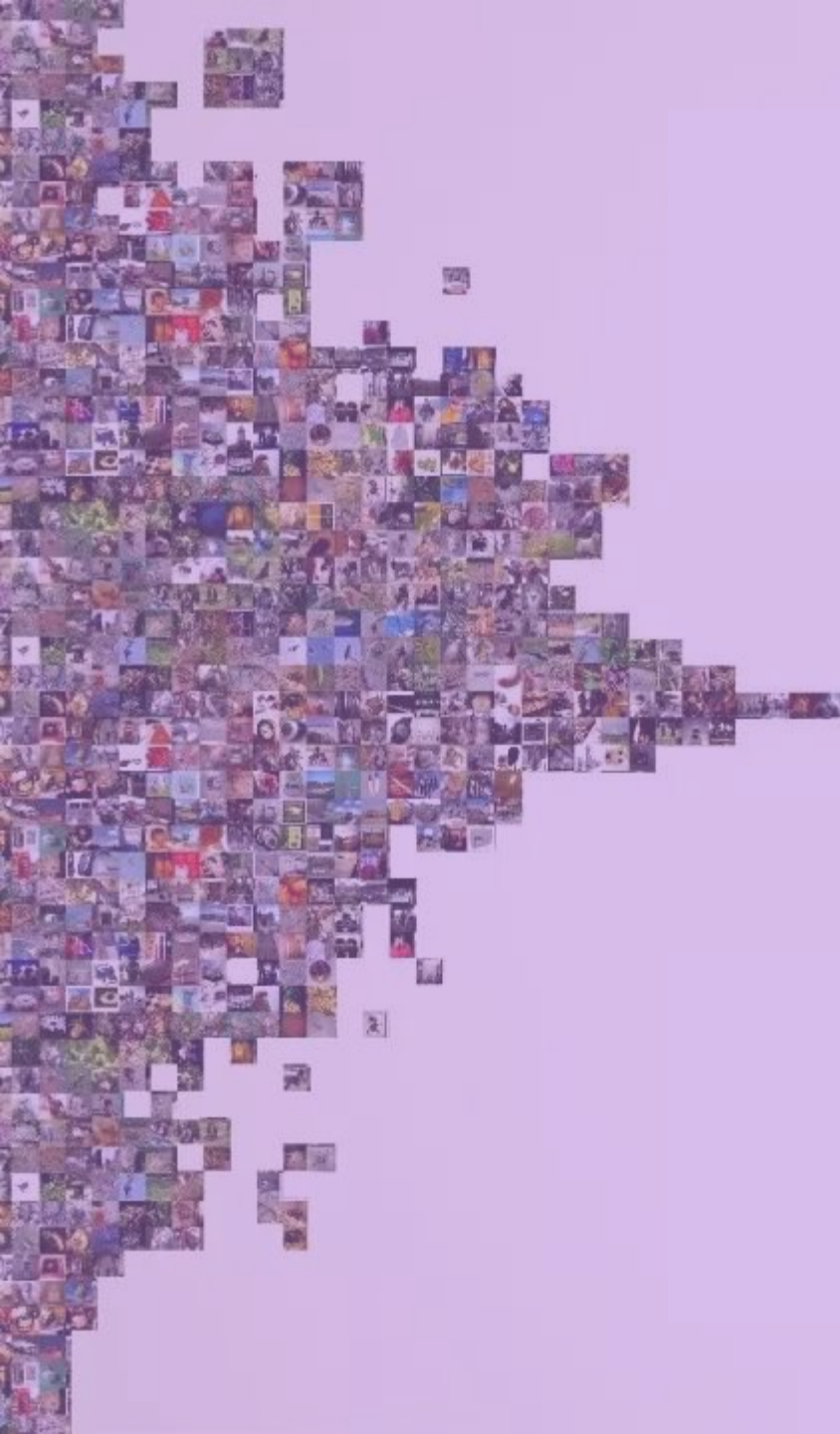
没有数据，或者非常有限

# 视觉物体理解的机遇及挑战



**关键问题**      源自信号、语义、模型、分布上的多层次混杂  
求解知识表达、蒸馏、修复、迁移和共享难题

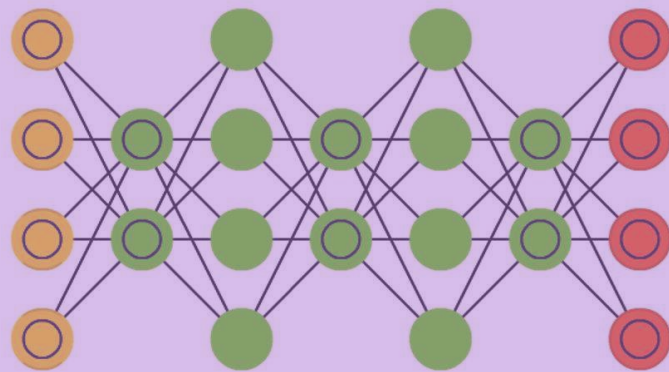




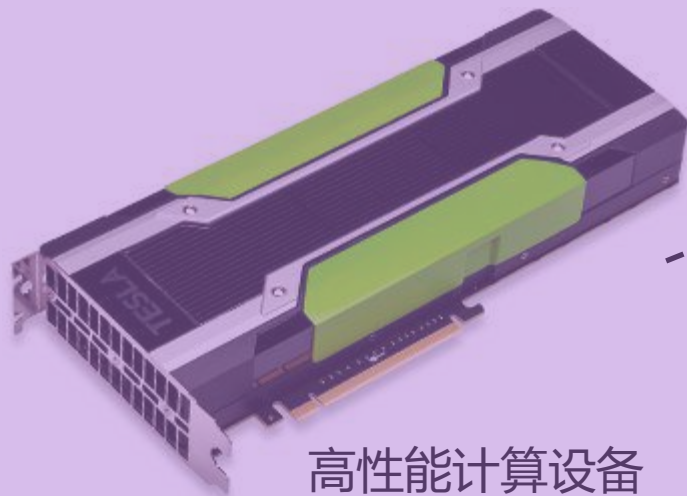
训练的深度模型从海量数据中汲取了丰富的知识(修炼了内功), 通过模型复用提供知识的有效利用



高性能计算设备



深度神经网络

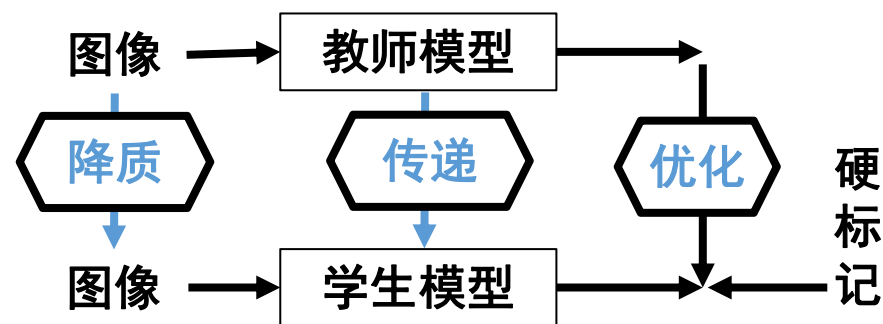
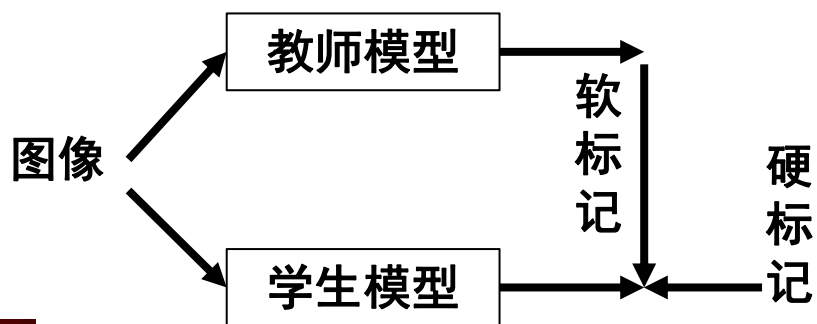
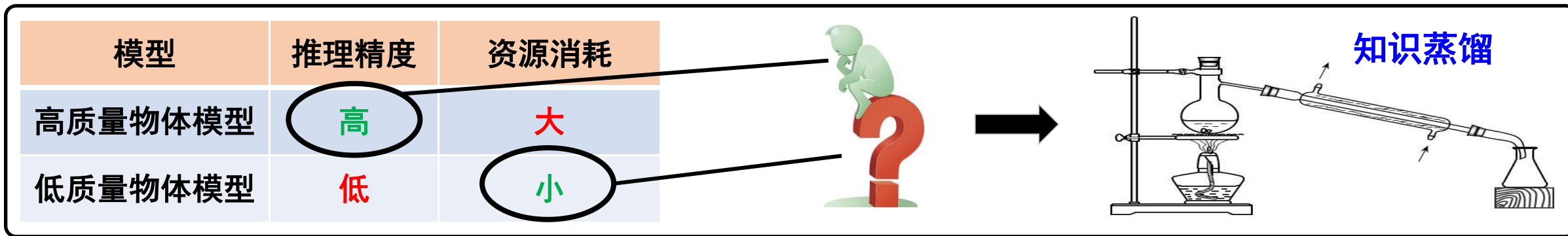


高性能计算设备

耗费大量的时间成本、精力成本和能耗成本训练得到很多高性能的模型，如果不能加以利用则是一种浪费；模型复用提供一种经济的手段来利用已训练的深度模型“红利”。

# 知识蒸馏

## ■ 复用深度模型，求解知识表达、蒸馏、修复和迁移问题

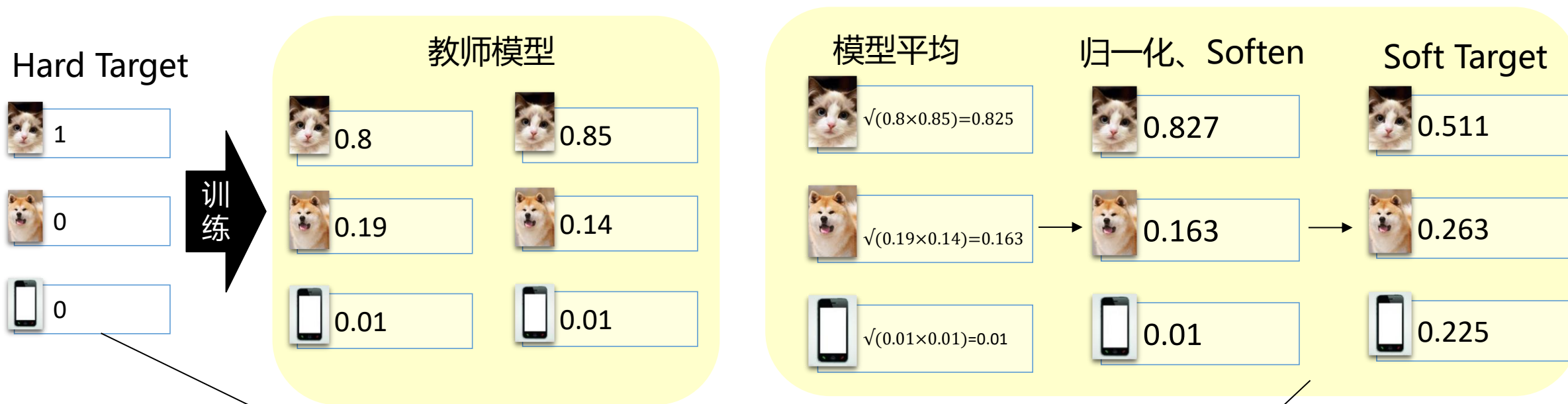


知识蒸馏(Hinton,2014)

Geoffrey Hinton, Oriol Vinyals, and Jeffrey Dean. Distilling Knowledge in a Neural Network, NIPS 2014 Workshop

# 知识蒸馏

- 利用大模型或集成(教师)的预测去训练更小、更便宜的模型(学生)
- 学生模型能够捕获教师模型的大部分优点, 实现轻量级、更快的推理



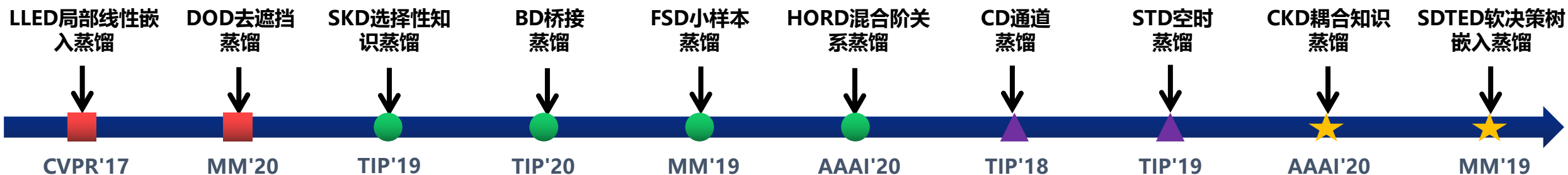
$$L = \alpha L^{(hard)} + (1 - \alpha) L^{(soft)}$$

Geoffrey Hinton, Oriol Vinyals, and Jeffrey Dean. Distilling Knowledge in a Neural Network, NIPS 2014 Workshop

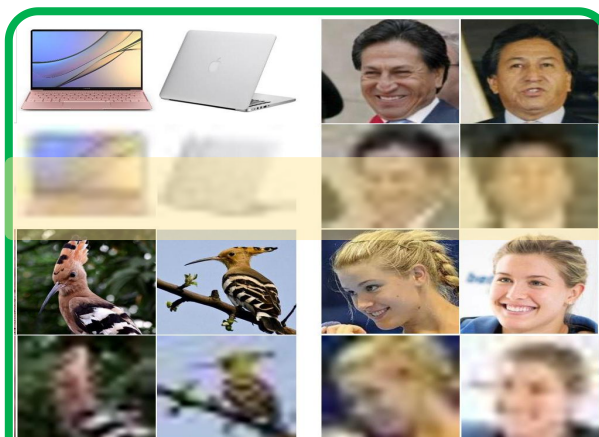


# 跨质量知识蒸馏

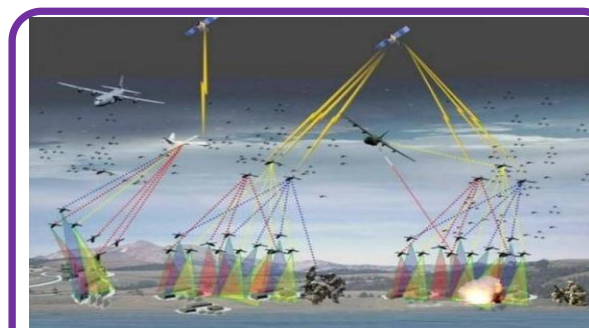
## ■ 提出跨质量知识蒸馏框架和系列算法，应用于多种场景



蒙面人脸分析



低分物体识别



VisDrone-SOT@ICCV2019冠军  
Anti-UAV@CVPR2020冠军

物体检测跟踪



模型理解防御

# 三个例子

## ■ 蒙面人脸检测识别

- S Ge, J Li, Q Ye, and Z Luo. Detecting Masked Faces in the Wild with LLE-CNNs. In: CVPR 2017
- C Li, S Ge, D Zhang, and J Li. Look Through Masks: Towards Masked Face Recognition with De-Occlusion Distillation. In: ACM MM 2020

## ■ 低分辨率人脸识别

- S Ge, S Zhao, C Li, and J Li. Low-Resolution Face Recognition in the Wild via Selective Knowledge Distillation. TIP, 2019
- S Ge, S Zhao, X Gao, and J Li. Fewer-Shots and Lower-Resolutions: Towards Ultrafast Face Recognition in the Wild. In: ACM MM, 2019
- S Ge, S Zhao, C Li, Y Zhang, and J Li. Efficient Low-Resolution Face Recognition via Bridge Distillation. TIP, 2020

## ■ 弱小目标深度跟踪

- S Ge, Z Luo, C Zhang, Y Hua, and D Tao. Distilling Channels for Efficient Deep Tracking. TIP, 2020
- C Zhang, S Ge, K Zhang, and D Zeng. Accurate UAV Tracking with Distance-Injected Overlap Maximization. In: ACM MM 2020

# 目录

---

深度模型  
知识蒸馏

蒙面人脸  
检测识别

低分辨率  
人脸识别

视频弱小  
目标跟踪

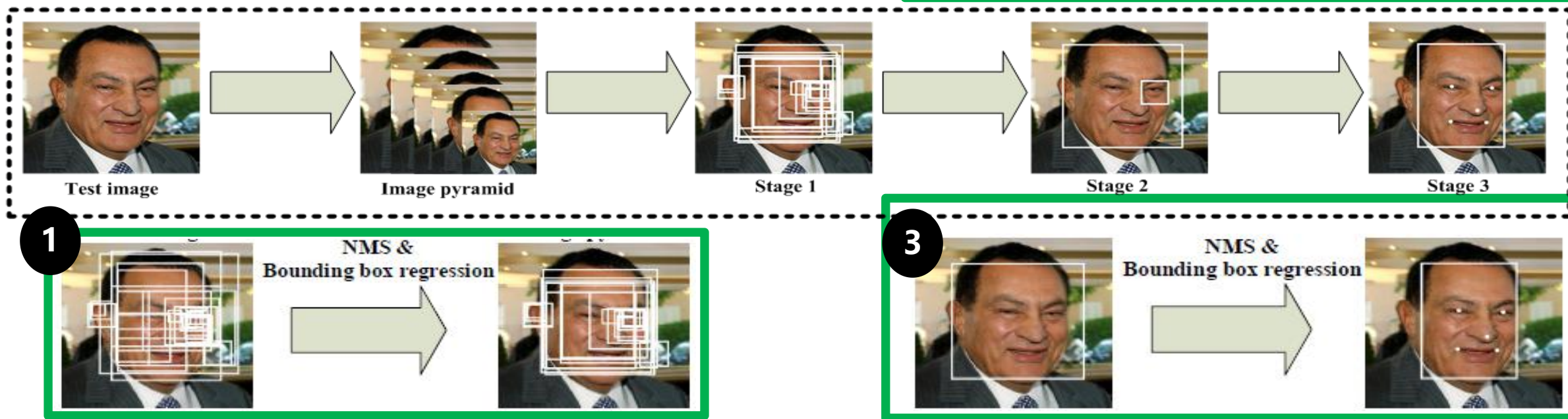
# 蒙面人脸检测与识别

- 城市治理、国家安全、公共卫生等领域具有十分重要意义及应用价值



# 检测：深度学习方法

- 单步检测 (SSD、YOLO)、两步检测 (RCNN)
- 级联了3个多任务卷积神经网络
- 分类和修正器联合分类和回归任务实现

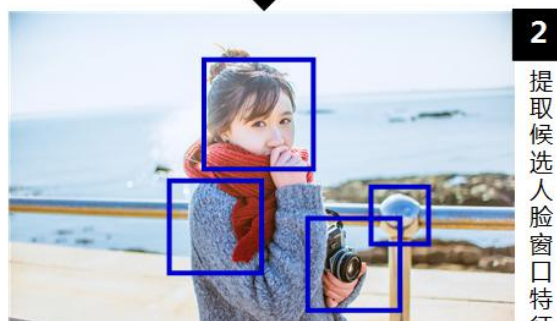


Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters. 2016

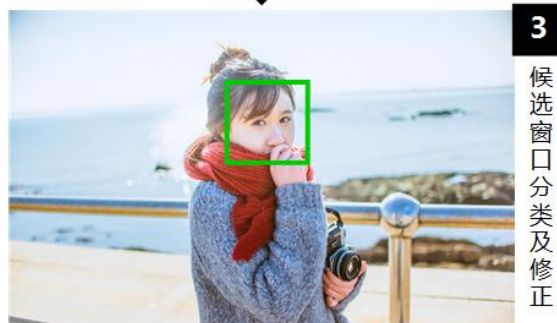
# 检测：两步人脸检测



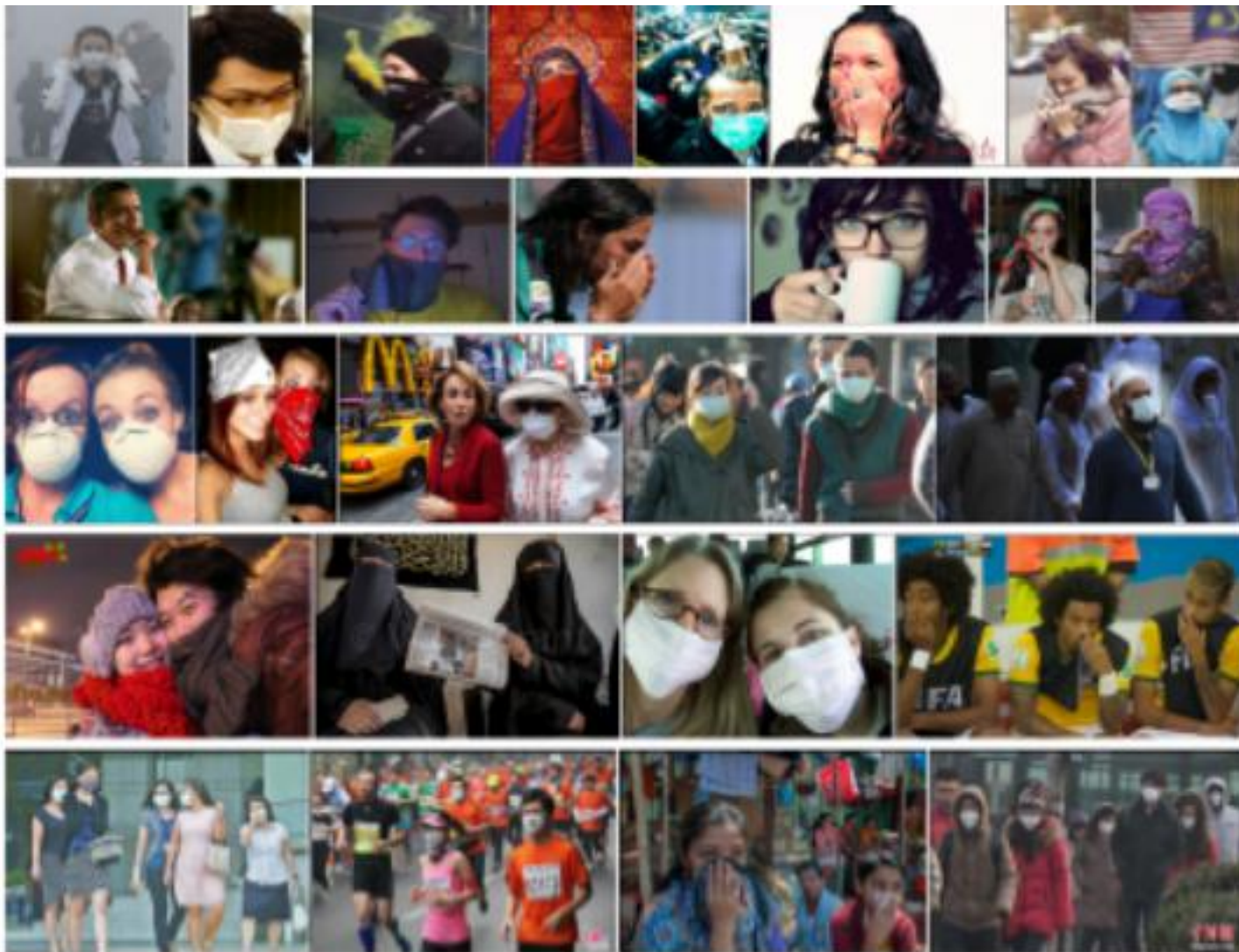
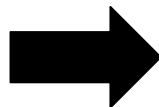
1 产生候选人脸窗口



2 提取候选人脸窗口特征



3 候选窗口分类及修正

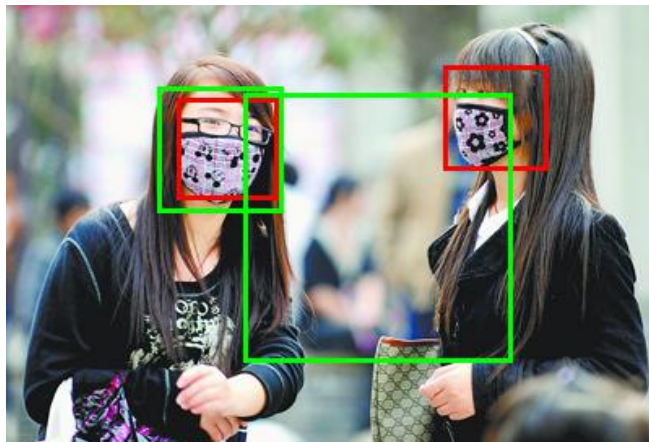


# 检测：蒙面人脸上的问题

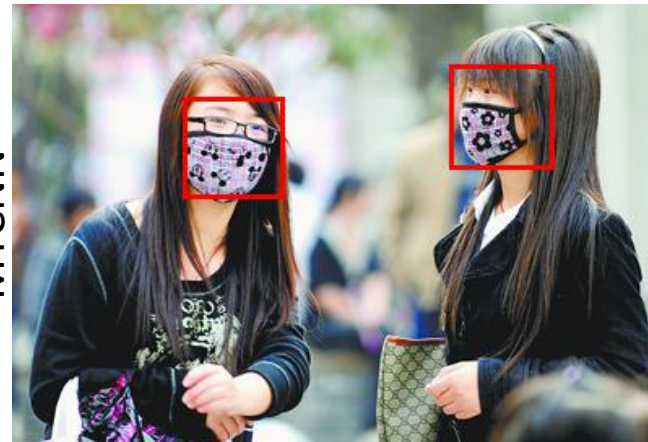
SURF



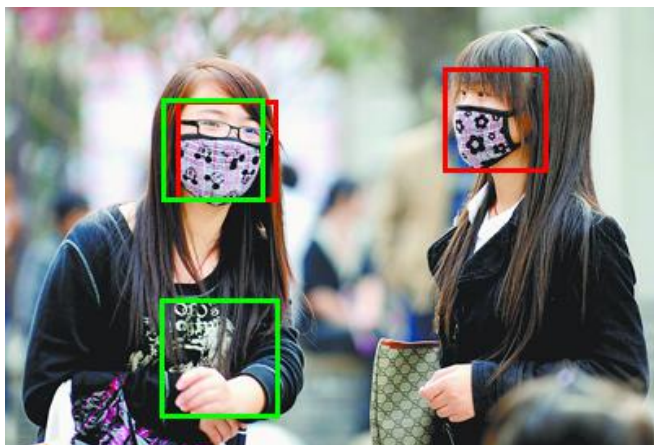
HH



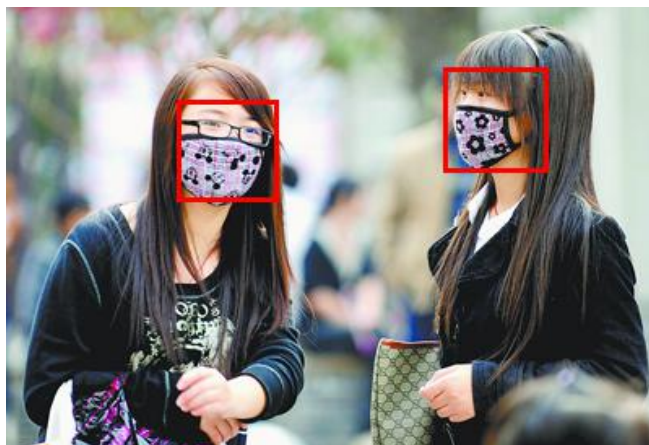
MTCNN



NPD



HPM



- 遮挡造成纹理/结构/部件等缺失，导致手工表征错误
- 遮挡造成的人脸结构缺失导致DPM方法失败
- 由于训练中未含蒙面人脸，导致深度模型检测失败

从数据和模型两方面进行优化，利用蒙面人脸数据辅助进行模型优化

# 检测：数据集MAFA (MAsked FAcEs)

**Face**  
(Location & Orientation?)

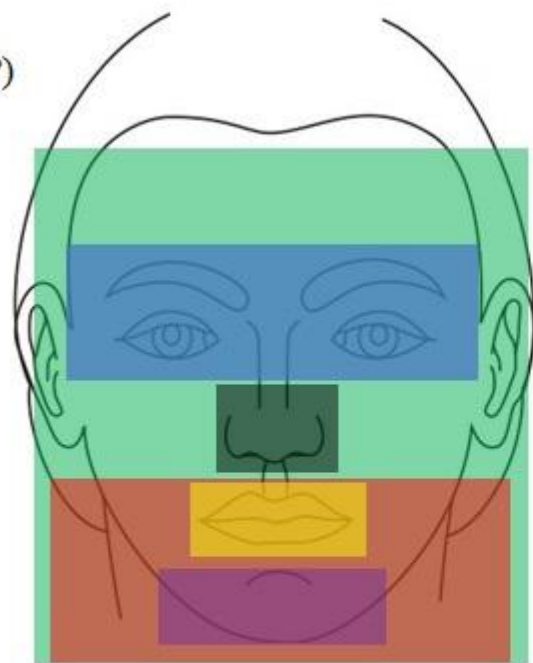
**Mask**  
(Location & Type?)

**Eyes**  
(Location & Glasses?)

**Nose**  
(Occlusion?)

**Mouth**  
(Occlusion?)

**Chin**  
(Occlusion?)

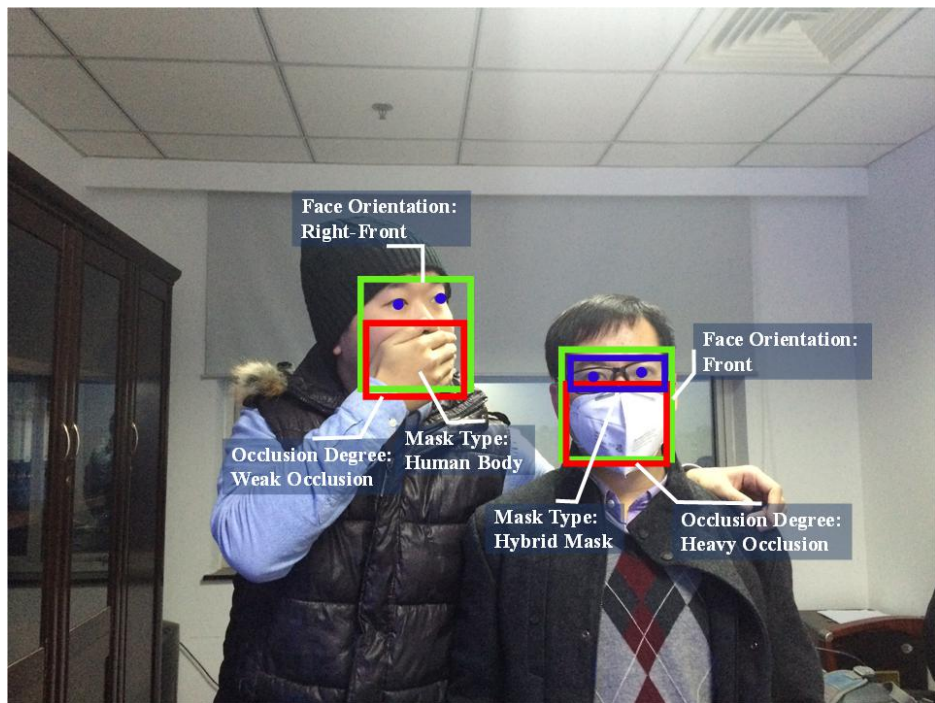


(a) Definition of facial attributes for annotation

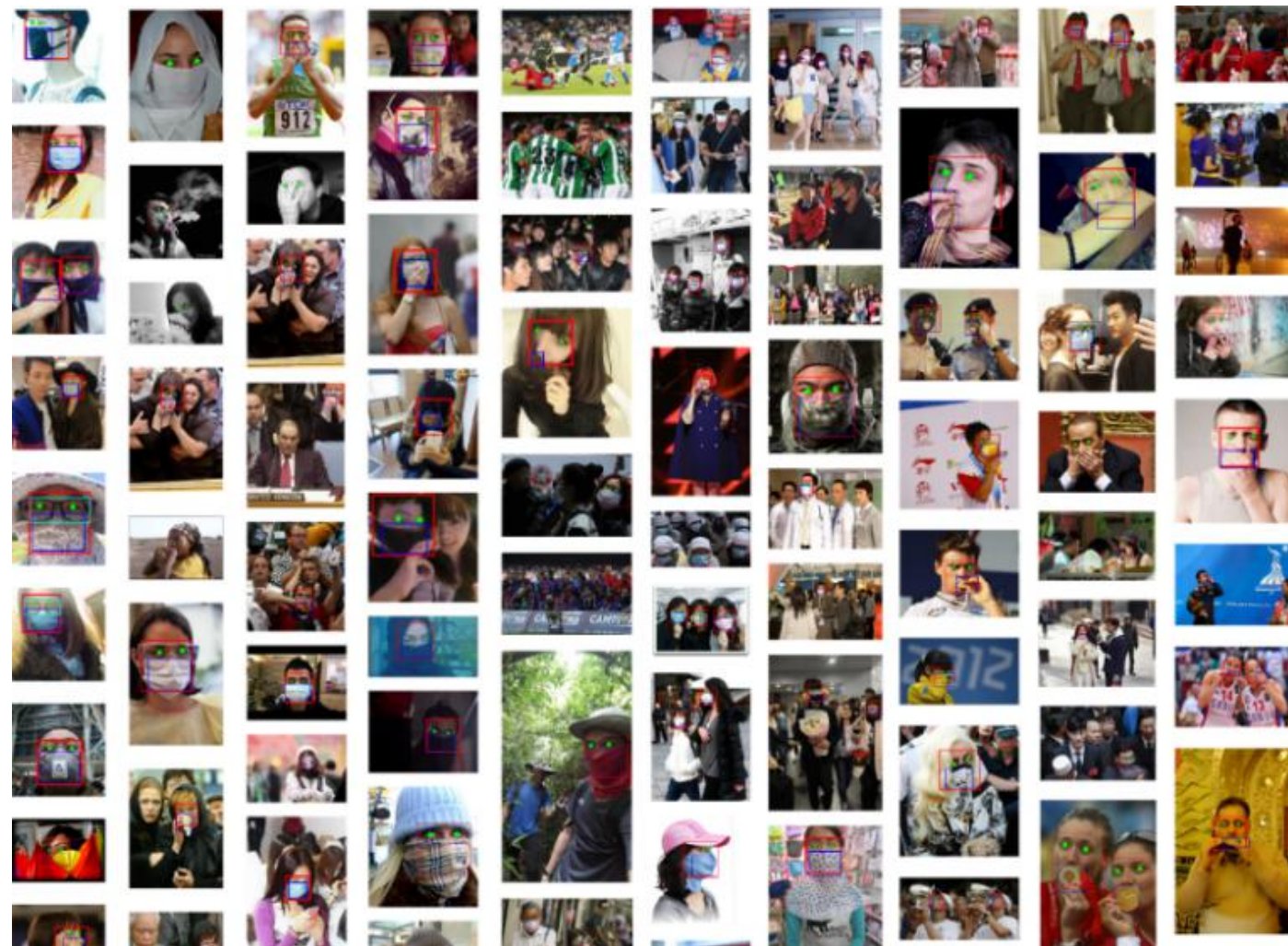


(b) Representative facial images with different attributes

# 检测：数据集MAFA (MAsked FAcEs)



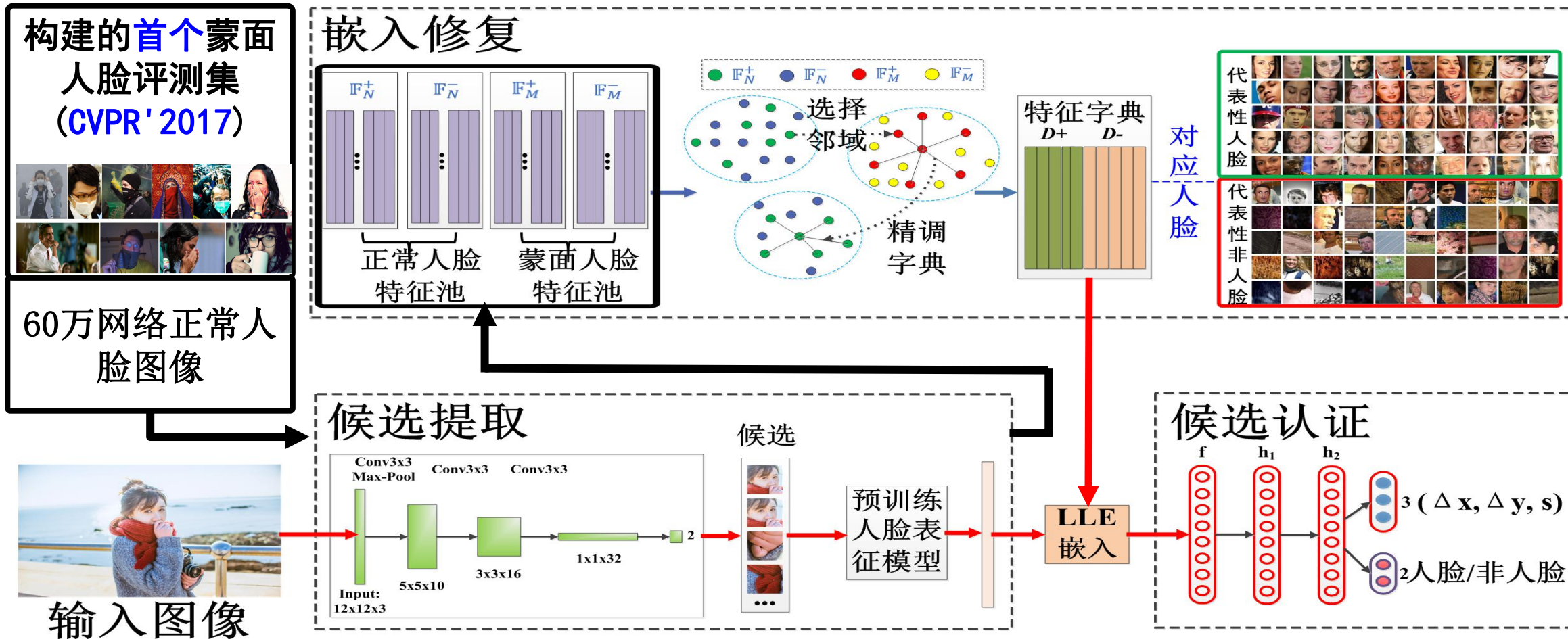
- 训练集：25876图，44746蒙面脸
- 测试集：4935图，6354蒙面脸
- 标注人脸/人眼/遮挡物位置、姿态、遮挡程度、遮挡物类型



Detecting Masked Faces in the Wild with LLE-CNNs. In CVPR 2017

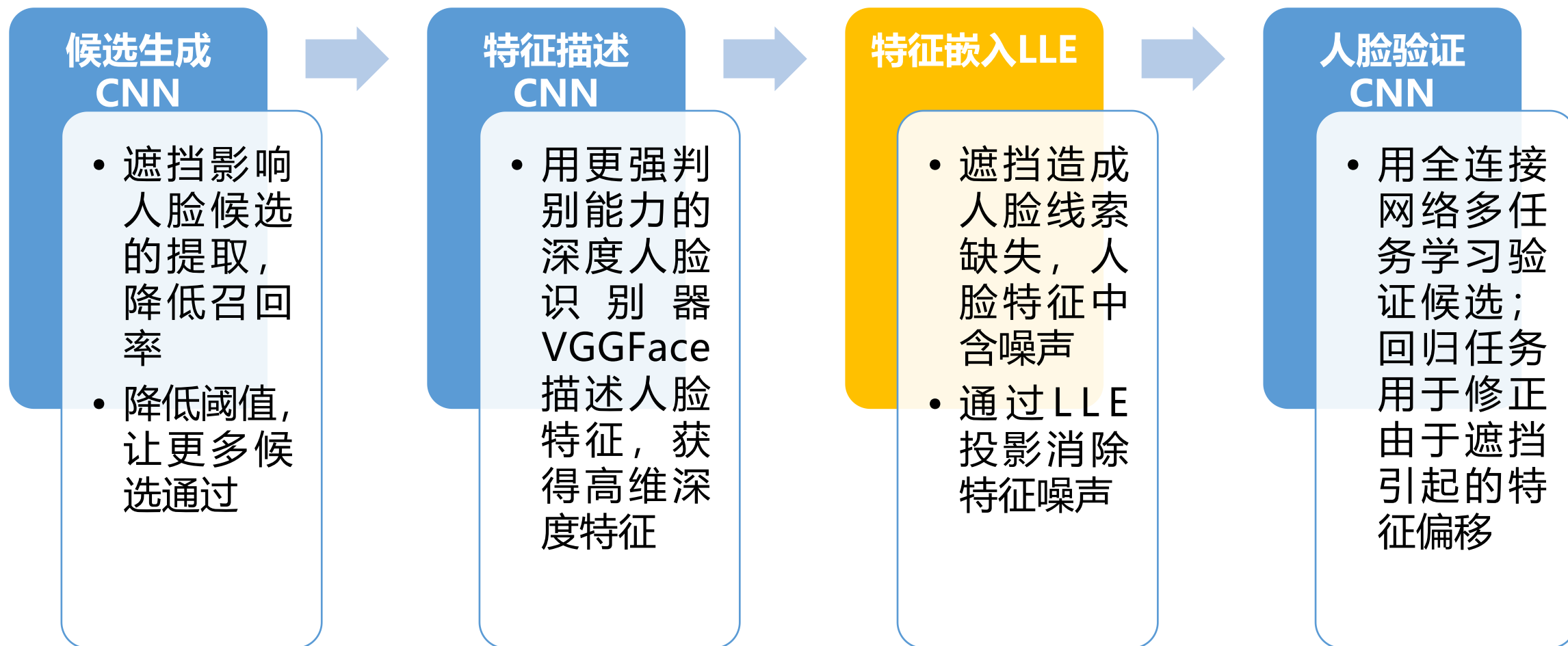
# 检测：模型LLE-CNNs

## ■ LLE (Locally Linear Embedding) 嵌入，实现蒙面人脸检测判别表征修复



Detecting Masked Faces in the Wild with LLE-CNNs. In CVPR 2017

# 检测：模型LLE-CNNs



Detecting Masked Faces in the Wild with LLE-CNNs. In CVPR 2017

# 检测：模型LLE-CNNs

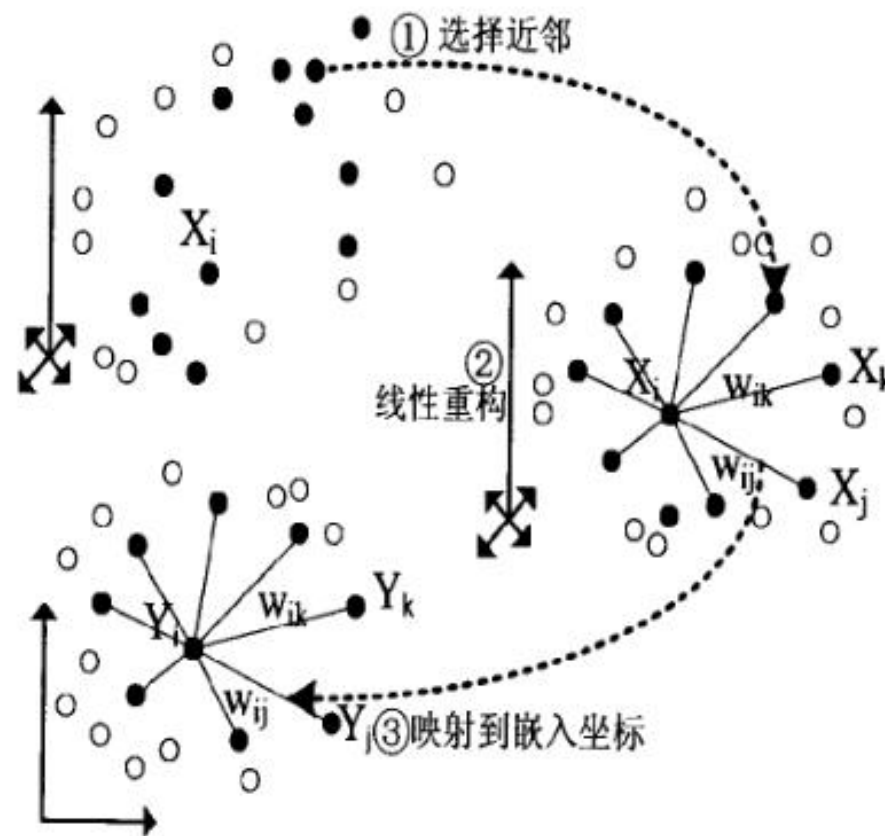
- 原始的LLE：利用自身数据内部规律进行降维，每次都要进行紧邻求解W然后降维

$$J(w) = \sum_{i=1}^m \left\| x_i - \sum_{j=1}^k w_{ij} x_j \right\|_2^2 \Rightarrow W \Rightarrow J(y) = \sum_{i=1}^m \left\| y_i - \sum_{j=1}^k w_{ij} y_j \right\|_2^2$$

- 从非蒙面人脸中寻找最能代表人脸和非人脸组成特征字典，对蒙面人脸特征进行嵌入

$$\hat{v}_i = \arg \min_v \left\| x_i - Dv \right\|_2^2, \quad \text{s.t. } v \succeq 0.$$

- $D = [D^+ D^-]$  是字典， $x_i$  是高维深度特征， $v$  是投影后特征



Sam Roweis, Lawrence Saul. Nonlinear dimensionality reduction by locally linear embedding. Science, 2000

## 检测：模型LLE-CNNs

- 通过遮挡人脸和正常人脸的比较，利用正常人脸数据集构造D
  - 通过寻找与正常人脸特征最相似的人脸或非人脸来修正 $x_i$

$$\min_{D^+} \frac{1}{|\mathbb{F}_M^+|} \sum_{\mathcal{F}_1 \in \mathbb{F}_M^+} \|\mathbf{x}_1 - \mathbf{D}^+ \alpha_1\|_2^2$$

$$- \frac{1}{|\mathbb{F}_M^-|} \sum_{\mathcal{F}_2 \in \mathbb{F}_M^-} \|\mathbf{x}_2 - \mathbf{D}^+ \alpha_2\|_2^2$$

$$\text{s.t. } \|\alpha_1\|_0 = \|\alpha_2\|_0 = 1, \alpha_1 \succeq 0, \alpha_2 \succeq 0,$$

$$\mathbf{D}_i^+ \in \{\mathbf{x}_0 | \mathcal{F}_0 \in \mathbb{F}_N^+\}$$

## 检测：实验结果

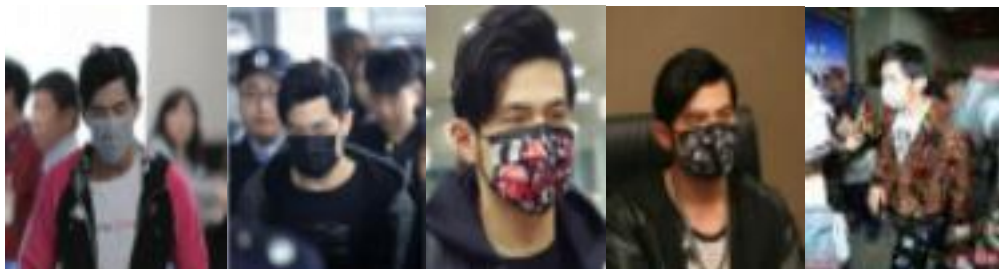
Table 1. Average Precision (%) on the Testing Set of MAFA

Attributes	SURF [18]	NPD [20]	ZR [37]	HH [22]	HPM [7]	MT [35]	OUR	Min ↑
Left	0.02	1.01	5.02	7.91	1.29	6.89	<b>17.2</b>	9.29
Left-Fr.	2.17	4.37	29.3	28.5	26.6	31.9	<b>61.7</b>	29.8
Front	19.7	16.9	45.5	51.6	64.4	62.2	<b>79.6</b>	15.2
Right-Fr.	1.93	2.34	13.8	20.4	18.9	20.2	<b>54.5</b>	34.1
Right	0.02	0.23	1.34	5.43	0.93	1.94	<b>14.3</b>	8.87
Weak	18.1	5.87	37.1	47.7	58.5	56.2	<b>75.8</b>	17.3
Medium	12.7	17.0	13.9	46.4	34.8	45.6	<b>67.9</b>	22.3
Heavy	0.05	0.52	7.12	5.59	5.31	5.24	<b>22.5</b>	15.4
Simple	10.7	12.8	39.3	45.3	54.7	51.6	<b>74.3</b>	19.6
Complex	11.8	8.52	33.3	42.1	46.1	48.2	<b>71.6</b>	23.4
Body	12.3	4.12	21.4	34.7	23.4	30.4	<b>62.0</b>	27.3
Hybrid	0.17	0.63	7.64	7.58	6.00	6.48	<b>24.2</b>	16.6
<b>All</b>	<b>16.1</b>	<b>19.6</b>	<b>41.6</b>	<b>50.9</b>	<b>60.0</b>	<b>60.8</b>	<b>76.4</b>	<b>15.6</b>

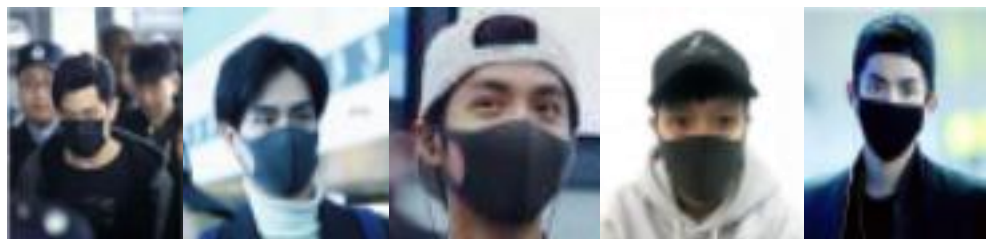
# 识别：问题与思路

## ■ 遮挡造成数据难建模，由以偏概全到主动修复

遮挡挑战



同一人，不同遮挡，外观差异大



不同人，相似遮挡，情况数量多



有效部分建模，信息不足

现有方法  
“以偏概全”

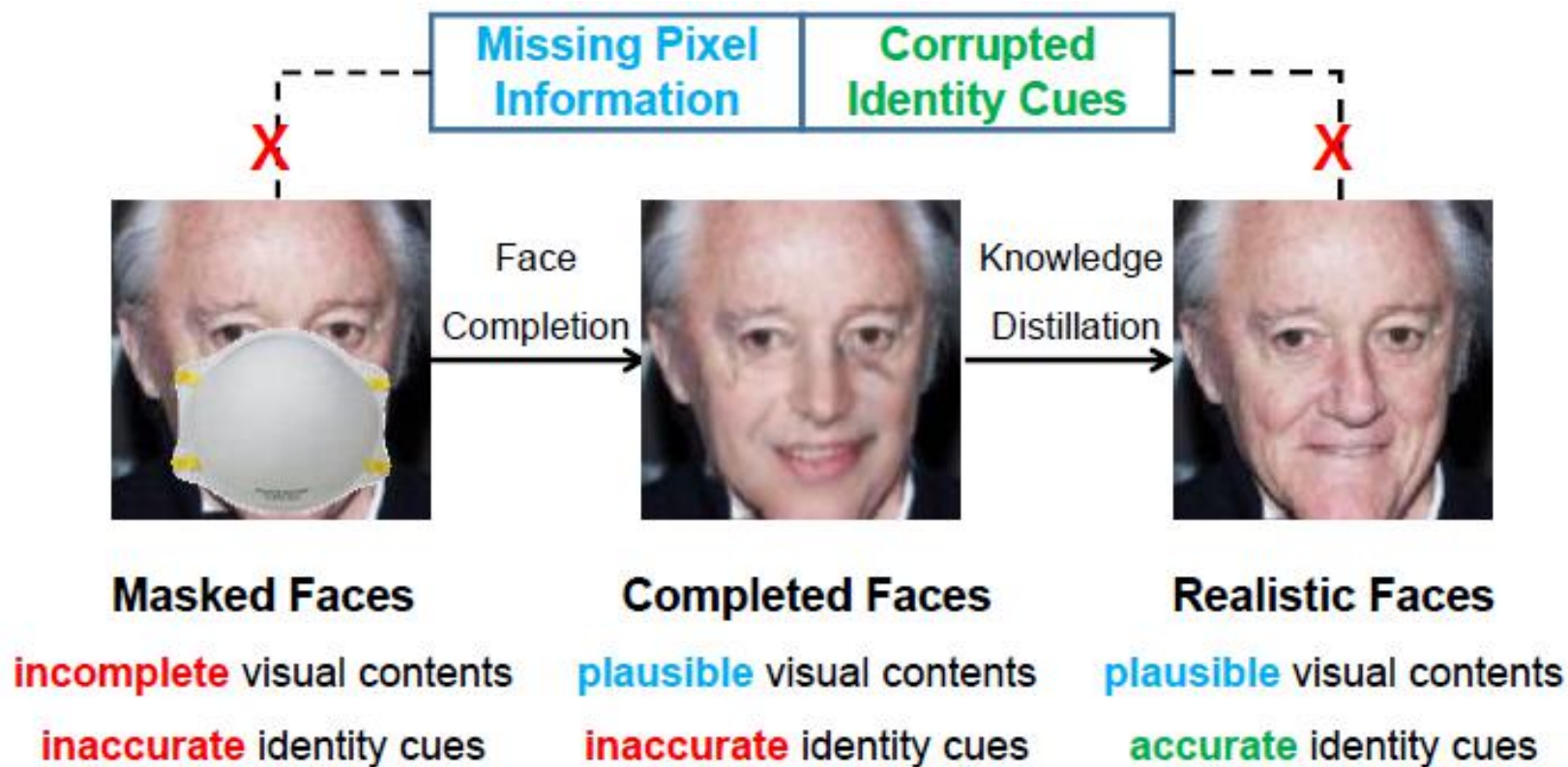


遮挡修复建模，信息完整

我们方法  
“主动修复”

# 识别：去遮挡蒸馏

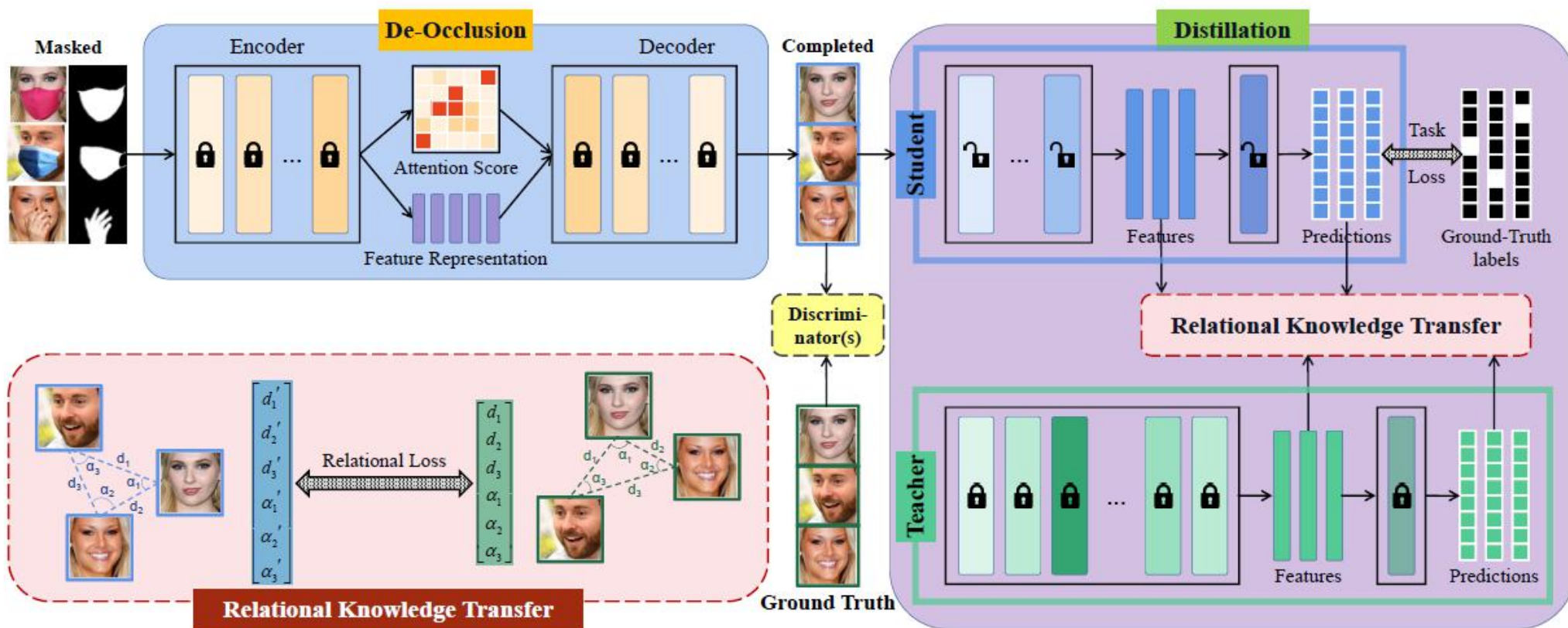
- 基于判别（粗粒度）：人脸修复显式地约束信息复原
- 基于生成（不对称）：知识蒸馏填补领域鸿沟



Look Through Masks: Towards Masked Face Recognition with De-Occlusion Distillation. In: ACM MM 2020

# 识别：去遮挡蒸馏

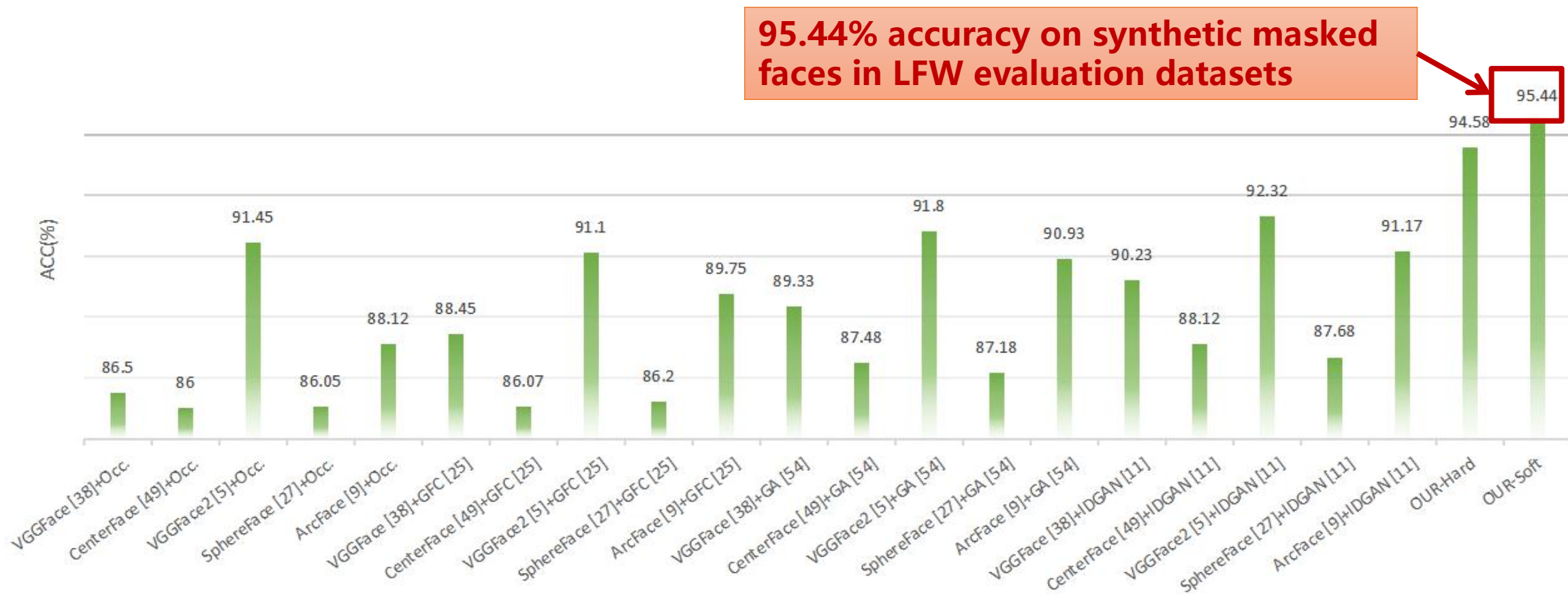
- 去遮挡模块：恢复人脸外观信息
- 蒸馏模块：通过软蒸馏重构特征



Look Through Masks: Towards Masked Face Recognition with De-Occlusion Distillation. In: ACM MM 2020

# 识别：实验结果

- 在合成和真实数据集上进行了验证



Look Through Masks: Towards Masked Face Recognition with De-Occlusion Distillation. In: ACM MM 2020

# 识别：实验结果

- 在合成和真实数据集上进行了验证

state-of-the-art performance on realistic masked faces in AR Database

Table 1: Rank-1 recognition accuracy on AR Database by several existing recognizers.

	AR1:Recognition of faces with glasses (%)					AR2:Recognition of faces with scarfs (%)				
	Occ.	RPCA [52]	GL [10]	GFC [26]	GA [56]	Occ.	RPCA [52]	GL [10]	GFC [26]	GA [56]
PCA [50]	61.4	64.2	70.0	82.9	88.6	37.5	32.2	40.8	72.6	78.3
GPCA [24]	73.3	71.6	76.6	88.4	92.7	56.2	54.0	60.9	80.3	87.2
LPP [15]	45.7	61.4	59.0	83.5	90.0	43.0	38.3	47.1	75.9	80.1
SR [53]	59.2	57.3	60.6	85.7	90.3	51.8	47.7	56.7	79.8	86.4
VGGFace [40]	85.4	84.5	87.9	91.7	95.9	75.9	79.6	83.5	89.9	92.2
VGGFace2 [5]	<u>88.3</u>	86.7	<u>89.0</u>	<u>94.2</u>	97.0	<u>78.2</u>	<u>81.4</u>	<u>85.7</u>	<u>93.1</u>	<u>93.3</u>
SphereFace [29]	<u>87.5</u>	<u>87.2</u>	<u>89.0</u>	93.7	<u>97.5</u>	<u>78.0</u>	79.8	83.9	92.8	93.1
ArcFace [9]	85.5	85.2	87.6	92.3	95.5	76.4	79.2	82.6	90.2	91.9
OUR	-	<b>92.1</b>	<b>93.3</b>	<b>97.2</b>	<b>98.0</b>	-	<b>84.4</b>	<b>86.8</b>	<b>93.3</b>	<b>94.1</b>

# 目录

---

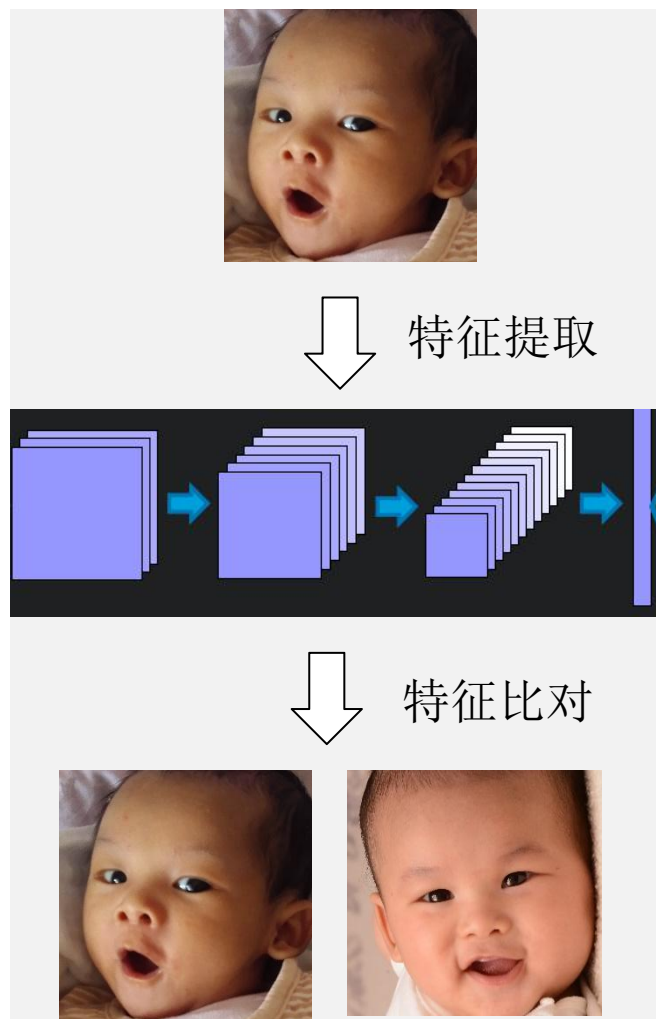
深度模型  
知识蒸馏

蒙面人脸  
检测识别

低分辨率  
人脸识别

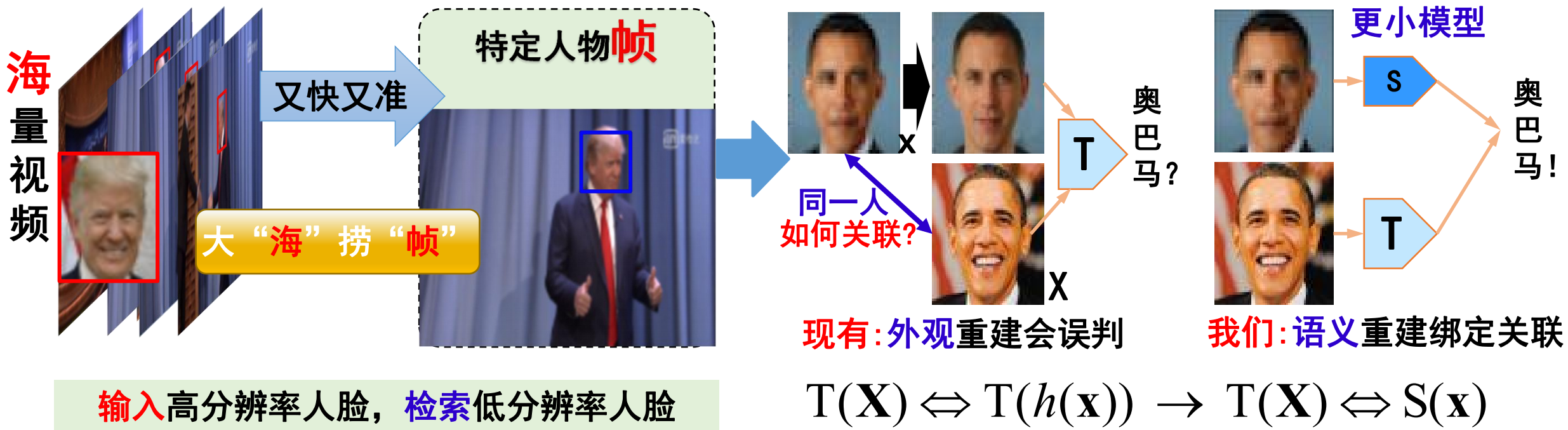
视频弱小  
目标跟踪

# 人脸识别介绍



# 低分辨率人脸识别

■ **问题与思路**: 细节缺失使语义难关联, 由外观重建到语义重建

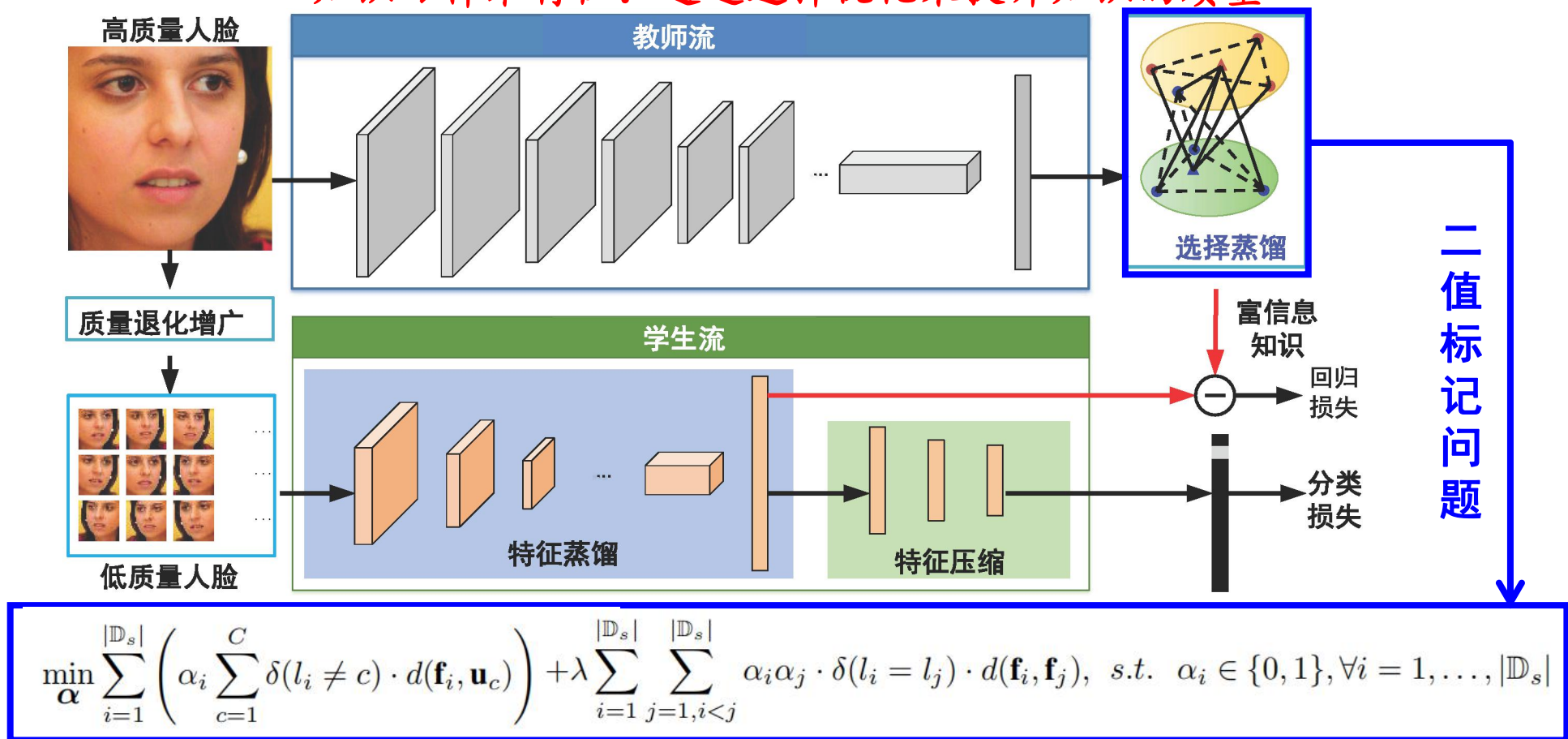


如何跨越场景、质量、模型的鸿沟?

# 选择性知识蒸馏

## 选择性知识蒸馏算法 (SKD), 优化迁移知识

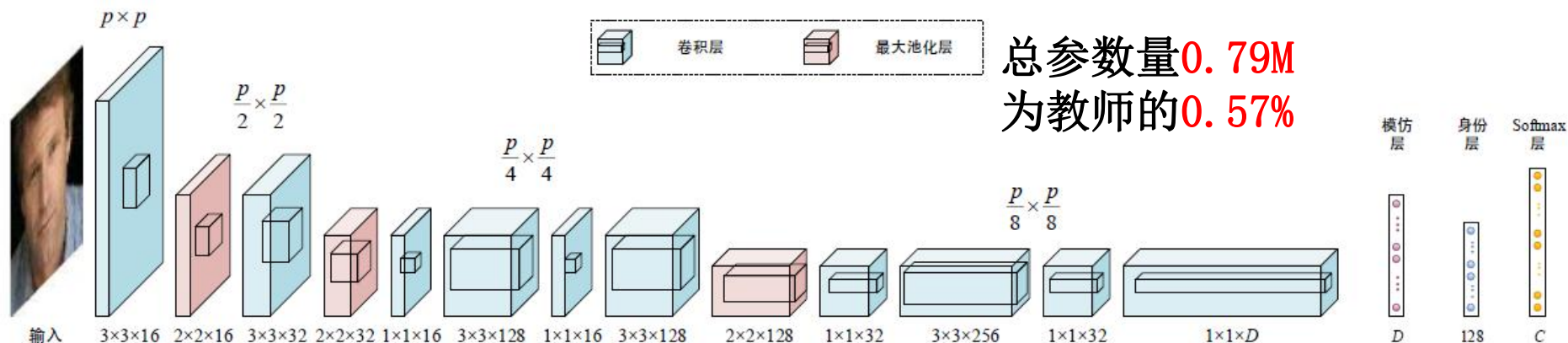
知识=>样本特征: 通过选择优化来提升知识的质量



Low-Resolution Face Recognition in the Wild via Selective Knowledge Distillation. TIP, 2019

# 选择性知识蒸馏

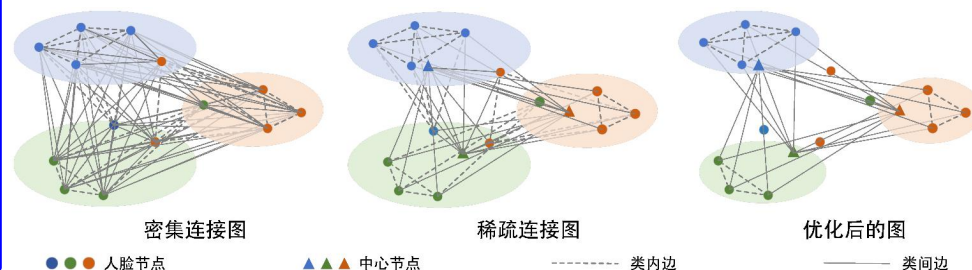
## ■ 学生网络：10卷积层、3最大池化层和3全连接层



□ 初始化：通过分类损失训练学生网络

□ **选择：基于图优化算法进行样本选择**

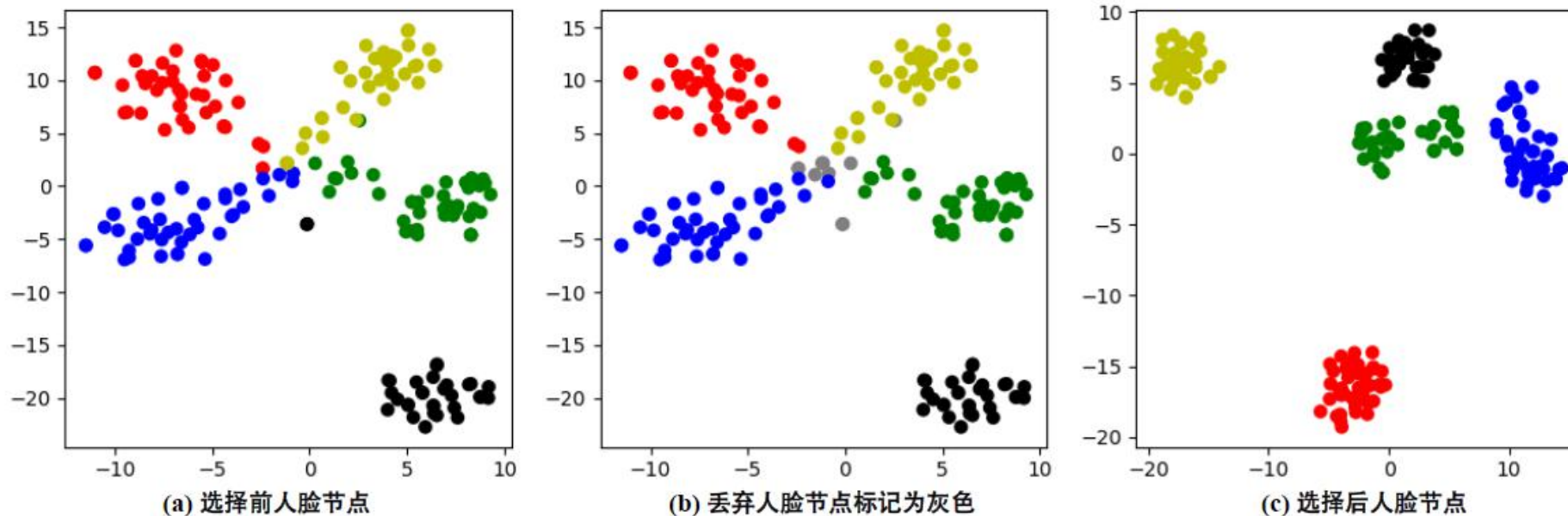
出发：每类的离群点极可能是出错样本  
构图：点表示人脸，边表示两人脸相似度  
优化：引入类中心节点，减少边数量  
求解：使用图割算法



□ 迁移：通过分类和蒸馏损失训练学生

# 选择性知识蒸馏

## 知识选择的t-SNE可视化结果



(d) 丢弃人脸节点对应图像 (边框颜色为人脸节点对应的类别)

被丢弃的人脸常具有侧姿、遮挡、光照不一致或外观模糊等特点，导致教师模型提取的特征知识有问题

# 选择性知识蒸馏

## ■ 实验结果

- UCCS数据集: <https://vast.uccs.edu/Opensetface/>

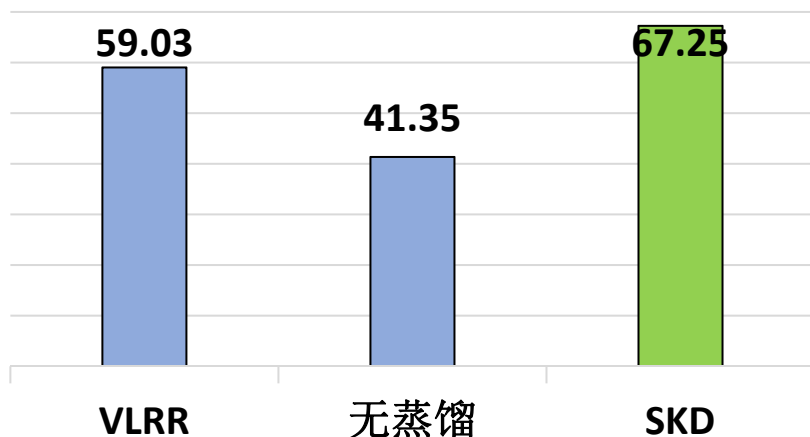
The UnConstrained College Students (UCCS) Dataset was collected over several months using Canon 7D camera fitted with Sigma 800mm F5.6 EX APO DG HSM lens, taking images at one frame per second, during times when many students were walking on the sidewalk. We have labeled in total **1732 identities**. Each labeled sequence contains around **10 images**.



# 选择性知识蒸馏

## ■ 实验结果

UCCS集 (16x16)



选择性知识蒸馏SKD提升**25.9%**,  
比VLRR (CVPR2016) 提升**8.22%**

【VLRR】Zhangyang Wang, Shiyu Chang, Yingzhen Yang, Ding Liu, and Thomas S. Huang. Studying very low resolution recognition using deep networks, In IEEE CVPR 2016

效能分析

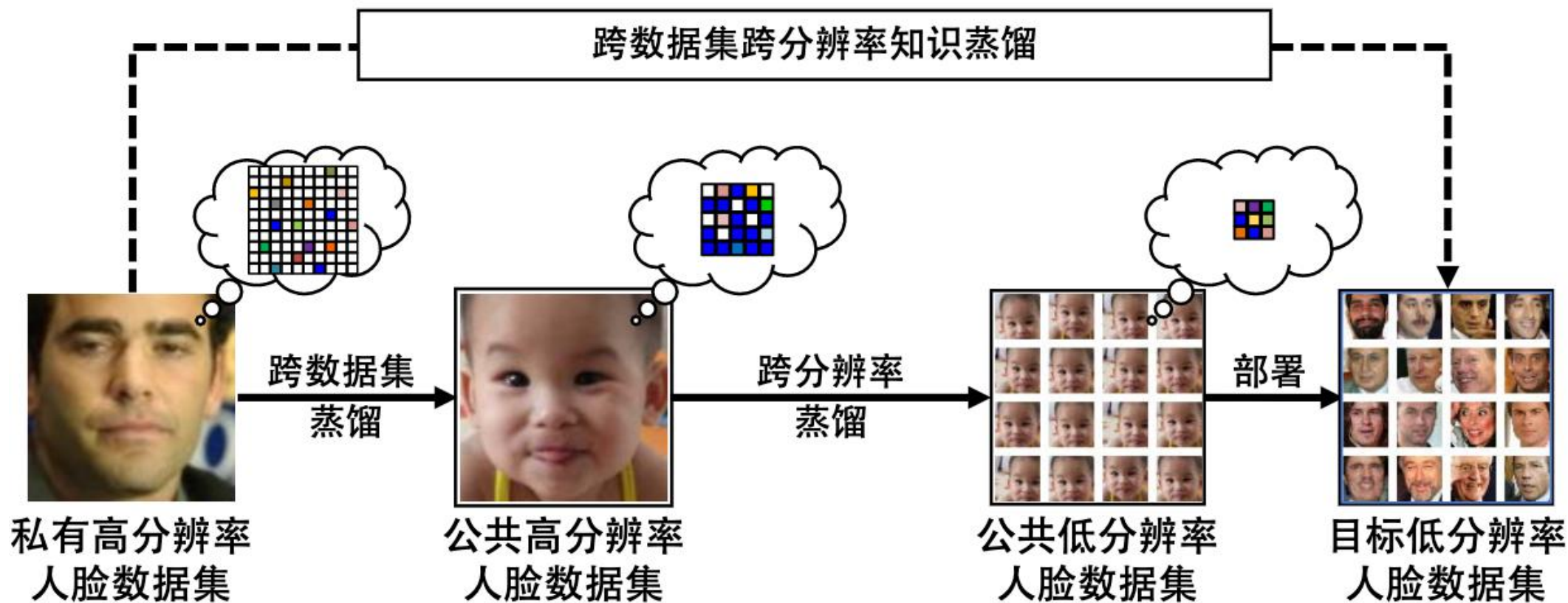
模型	推理内存 (MB)	速度@CPU (张/秒)
教师	93	1.02
SKD	0.15	418

推理内存占用比教师**减少620倍**,  
CPU上处理速度比教师**快410倍**



# 桥接蒸馏

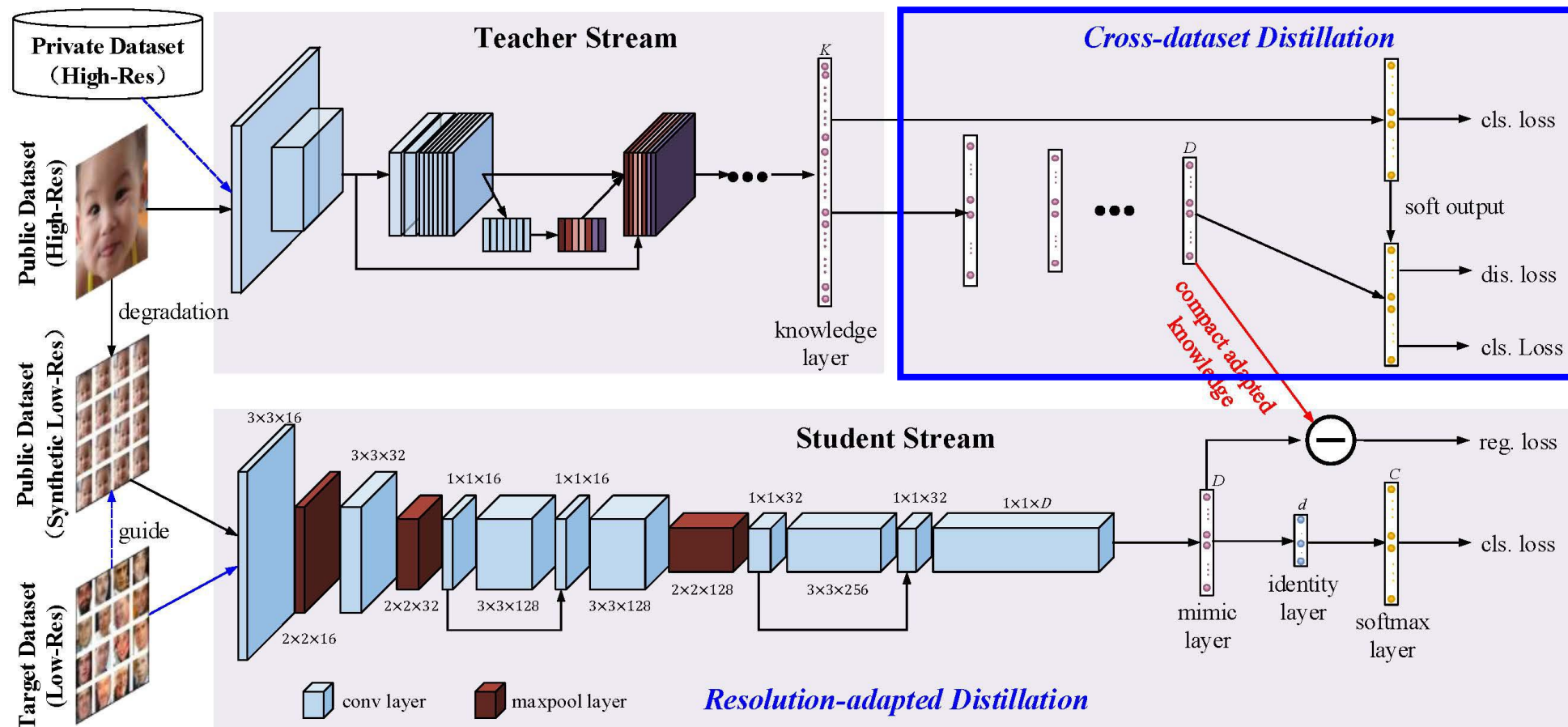
- **问题：** 教师模型和学生模型的训练数据存在**数据集分布**和**分辨率**的差异
- **思路：** 引入公共人脸数据集作为连接两个模型的桥梁



# 桥接蒸馏

## 方法：桥接蒸馏，提升跨质量知识的有效迁移

知识=>样本特征：通过适应来提升知识的质量

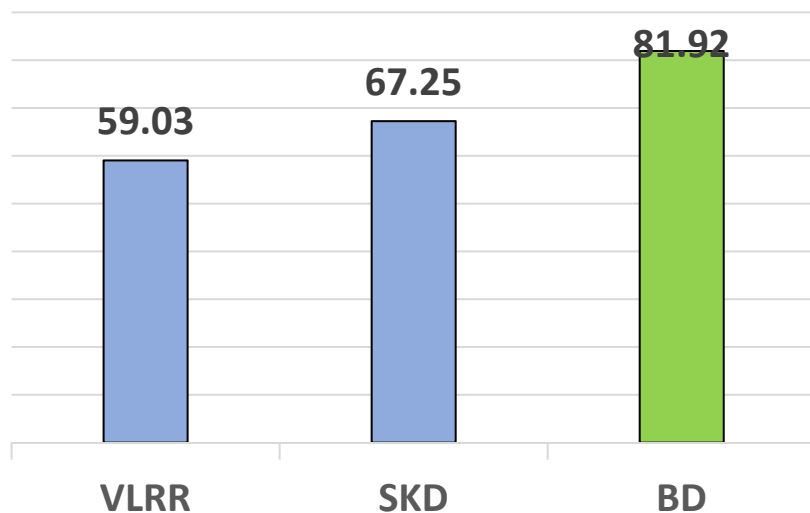


Efficient Low-Resolution Face Recognition via Bridge Distillation. TIP, 2020

# 桥接蒸馏

## ■ 实验结果

### UCCS集(16x16)



渡桥蒸馏BD相比选择性知识蒸馏SKD提升**14.67%**

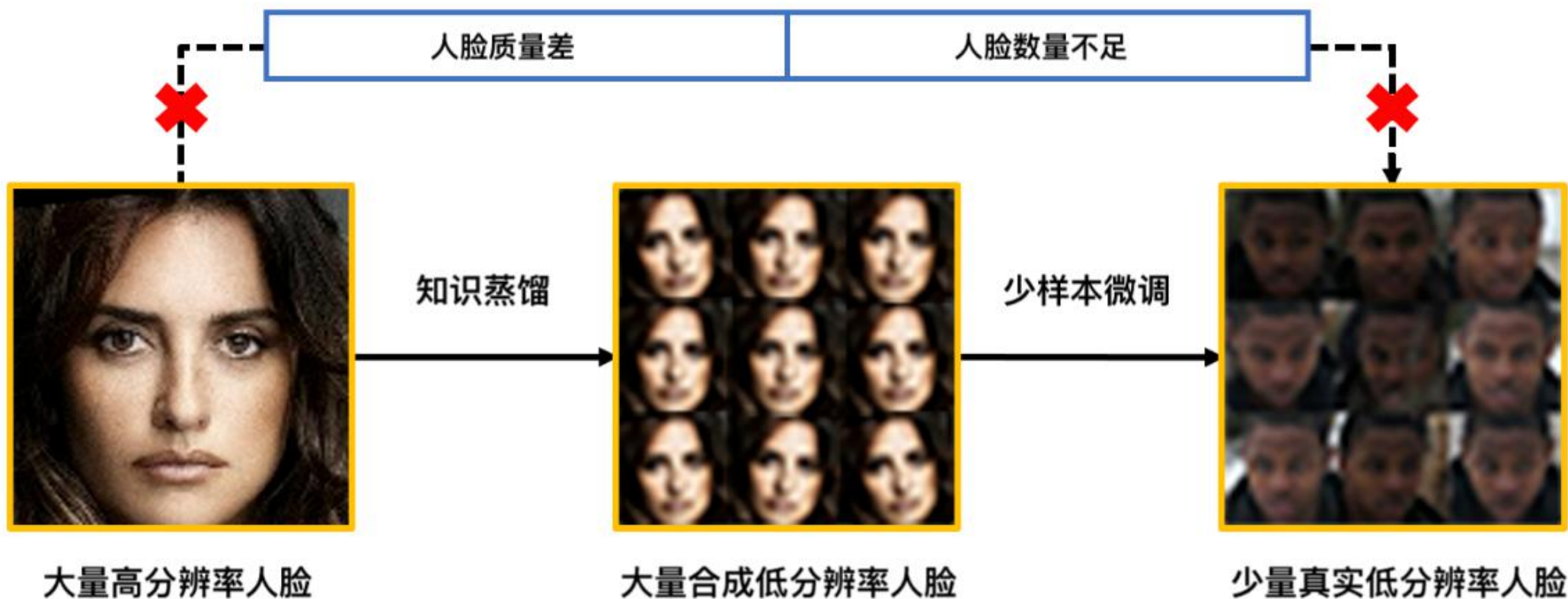
### 效能分析

模型	推理内存(MB)	速度@CPU(张/秒)
教师	35.41	0.01
SKD	0.15	418
BD	<b>0.057</b>	<b>934</b>

BD模型推理内存占用比SKD模型**减少了2.6倍**，CPU上**处理速度比快2.2倍**

# 小样本蒸馏

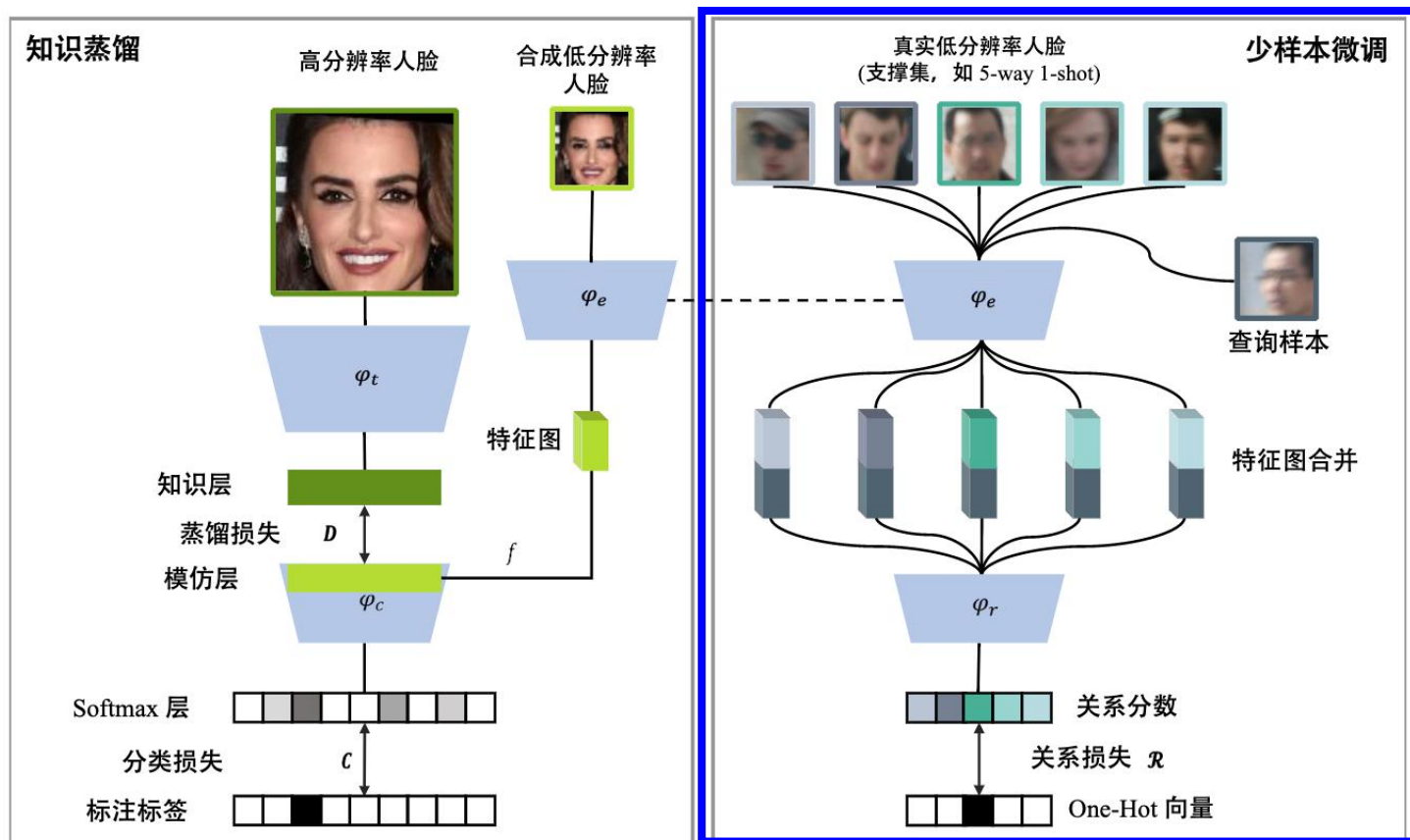
- **问题：** 真实场景下的低分辨率人脸样本获取和标注难
- **思路：** 结合知识蒸馏和小样本学习



# 小样本蒸馏

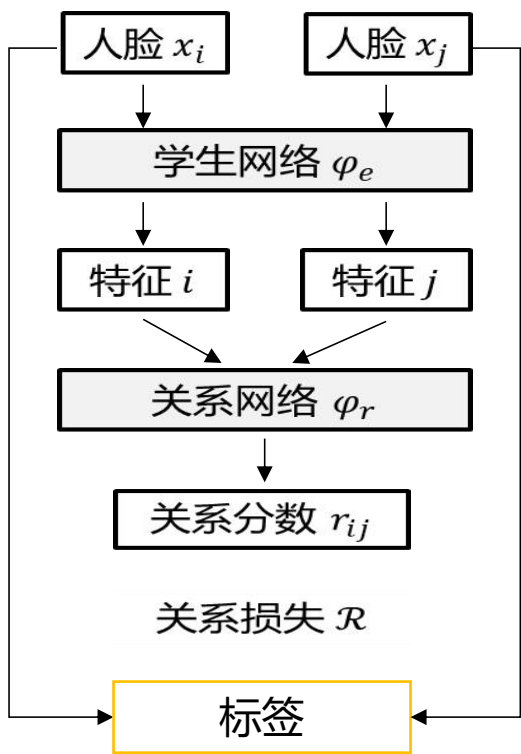
## ■方法：小样本蒸馏(FSD)，提升了跨质量知识的有效迁移

知识=>样本特征：通过小样本微调来提升知识的迁移



Fewer-Shots and Lower-Resolutions: Towards Ultrafast Face Recognition in the Wild. In ACM MM, 2019

# 小样本蒸馏



$$r_{i,j} = \varphi_r([\varphi_e(x_i; \mathbf{w}_e), \varphi_e(x_j; \mathbf{w}_e)]; \mathbf{w}_r)$$

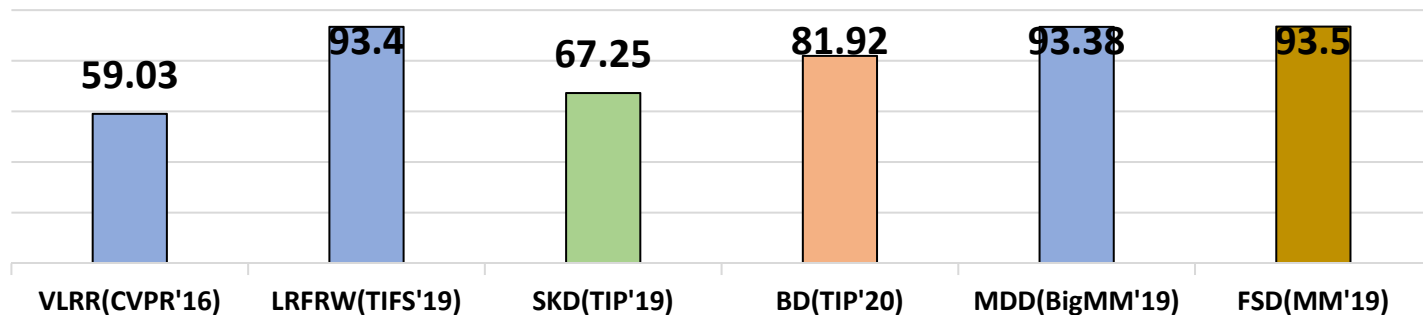
$$\mathcal{R}(\{\mathbf{w}_e, \mathbf{w}_r\}; \mathbb{D}_T) = \sum_{i=1}^c \sum_{j=1}^m (r_{i,j} - 1(y_i == y_j))^2.$$

## 小样本微调

- 在训练网络的每轮迭代中:
  - ✓ 构造支撑集和查询集。随机从训练集中选择 $c$ 类，其中每类选取 $l$ 个样本组成支撑集，剩余的样本组成查询集
  - ✓ 通过关系模块计算查询集中每个样本和支撑集中每个类别的关系分数 $r_{ij}$
  - ✓ 使用平方误差计算关系分数和标签的关系损失 $\mathcal{R}$
  - ✓ 计算梯度并反向传播
  - ✓ 更新网络参数
- 直到收敛

# 小样本蒸馏

## ■ 实验结果



小模型下比混合域蒸馏MDD和LRFRW均有提升，比桥接蒸馏提升11.58%

【LRFRW】 On low-resolution face recognition in the wild: Comparisons and new techniques, IEEE TIFS 2019.

模型	精度 (%)	内存 (MB)	张/秒@CPU
教师(ArcFace)	90.56	35.41	0.01
SKD(TIP'19)	67.25	0.15	418
BD(TIP'20)	81.92	0.057	934
FSD(MM'19)	93.50	0.03	1460

# 目录

---

深度模型  
知识蒸馏

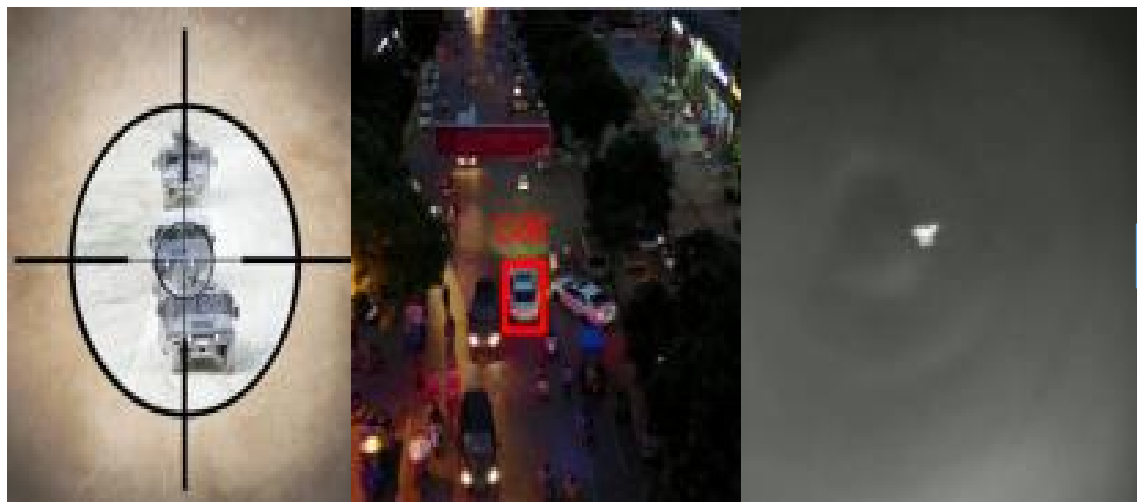
蒙面人脸  
检测识别

低分辨率  
人脸识别

视频弱小  
目标跟踪

# 视频弱小目标跟踪

■ **问题与思路**: 目标弱小使模型难适应, 由一成不变到随机应变



战场环境(地面、空对地、地对空)目标

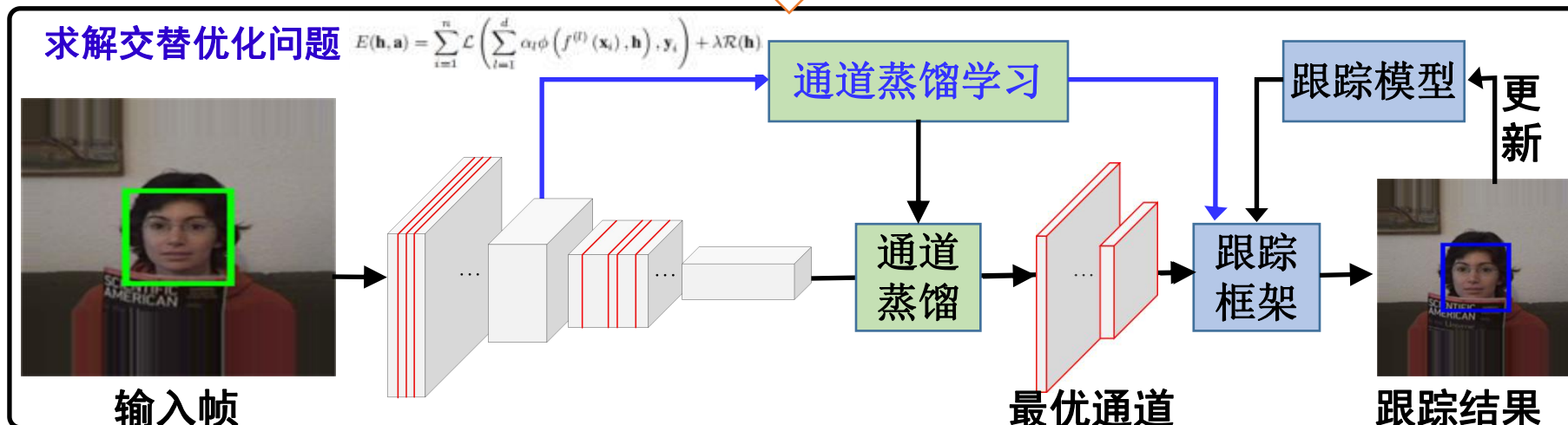
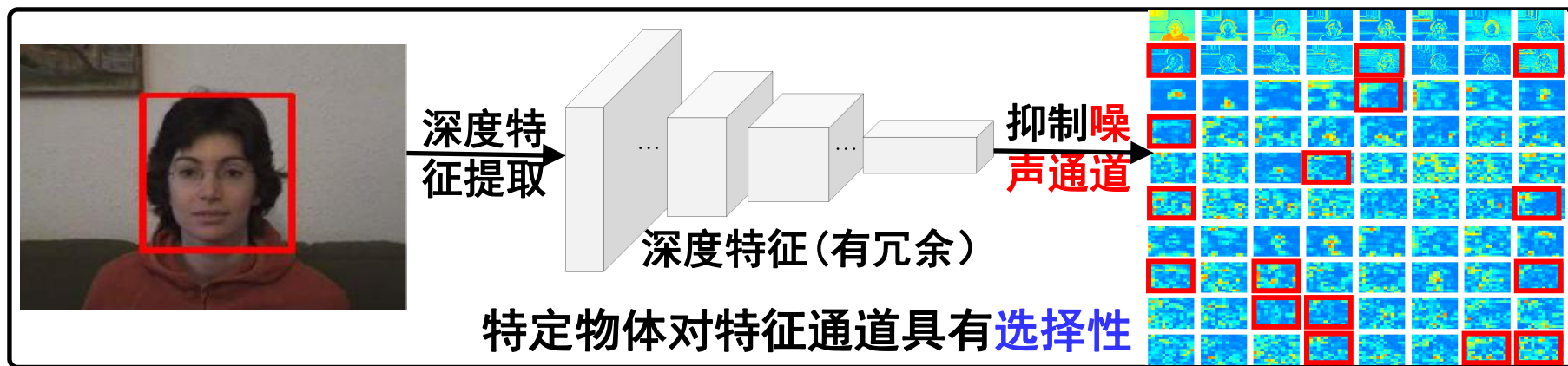


现有: 固定模型难适应 我们: 模型适应目标

如何利用通用的视觉模型自适应表征特定的目标?

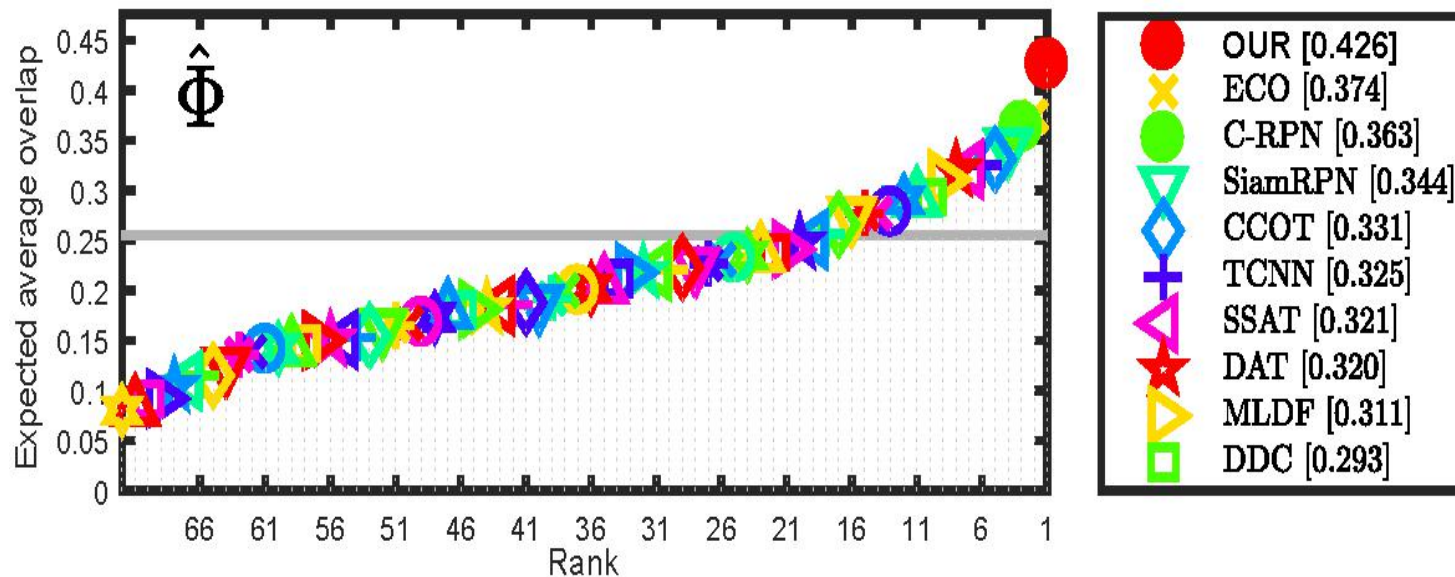
# 通用目标跟踪

- 利用深度特征的通道选择性，提出了**通道蒸馏**深度跟踪方法



# 实验结果

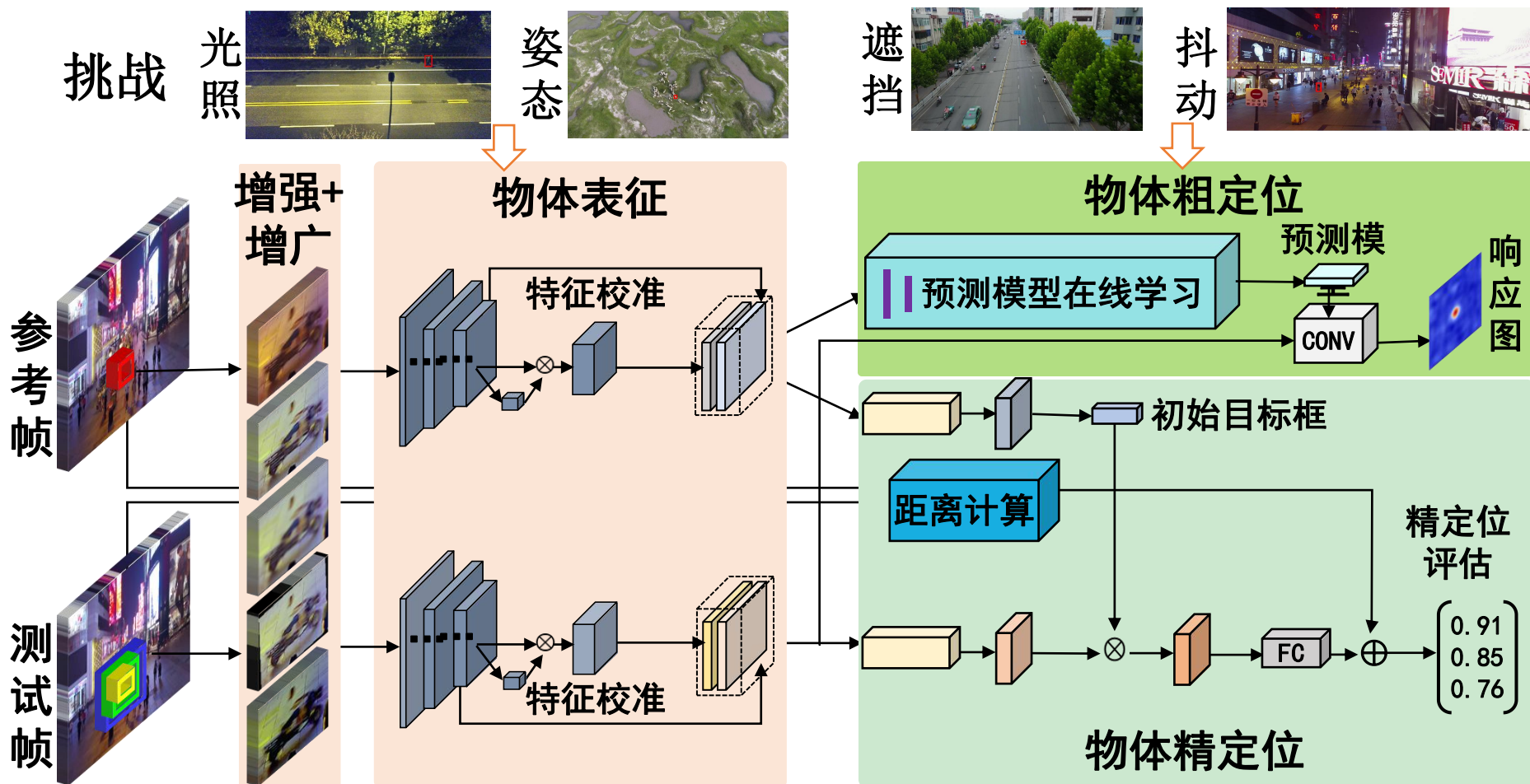
■ 评测集VOT2016上，性能比最好方法(CVPR'17)提升13.9%



地面场景VOT2016数据集上的目标跟踪性能

# 无人机目标跟踪

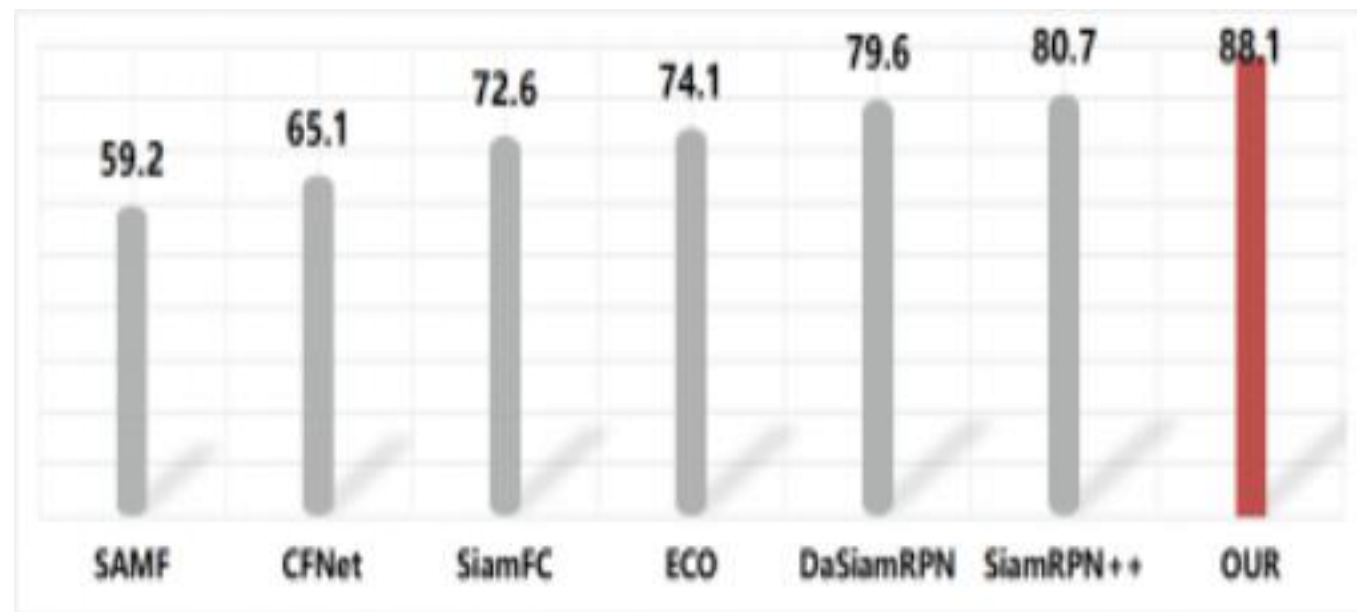
## 结合数据增强与增广，提升表征，抑制模型漂移



Accurate UAV Tracking with Distance-Injected Overlap Maximization. In: ACM MM 2020

# 实验结果

- 评测集UAV123上，无人机目标跟踪精度比最好方法(CVPR'19)提升7.4%

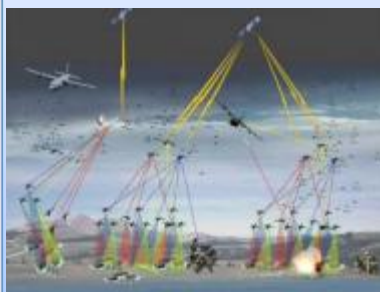


空对地场景UAV123数据集上的无人机目标跟踪性能

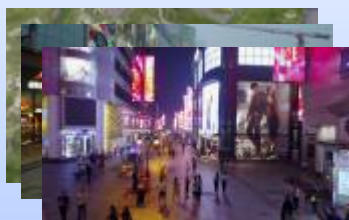
# 实验结果

## ■ 视觉领域国际评测中，获得第一

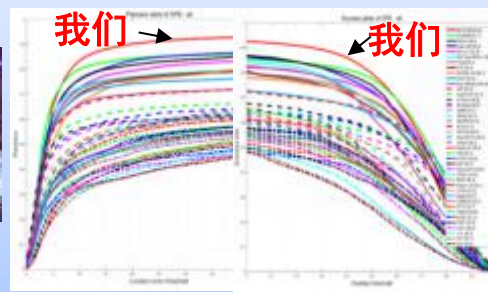
### ICCV 2019, 无人机目标跟踪国际评测冠军



空对地



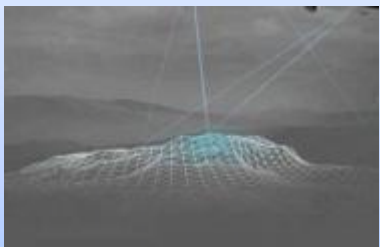
272,117视频帧, 260万个标注框  
长时遮挡、黑夜、目标弱小、相机抖动



结果



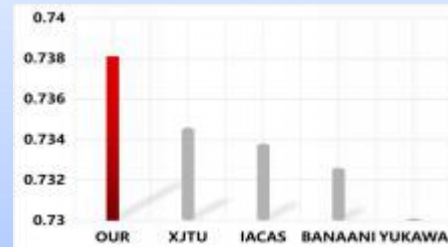
### CVPR 2020, 反无人机国际评测冠军



地对空



302,148视频帧, 24万个标注框  
多机型、多模态、目标弱小、遮挡运动



结果



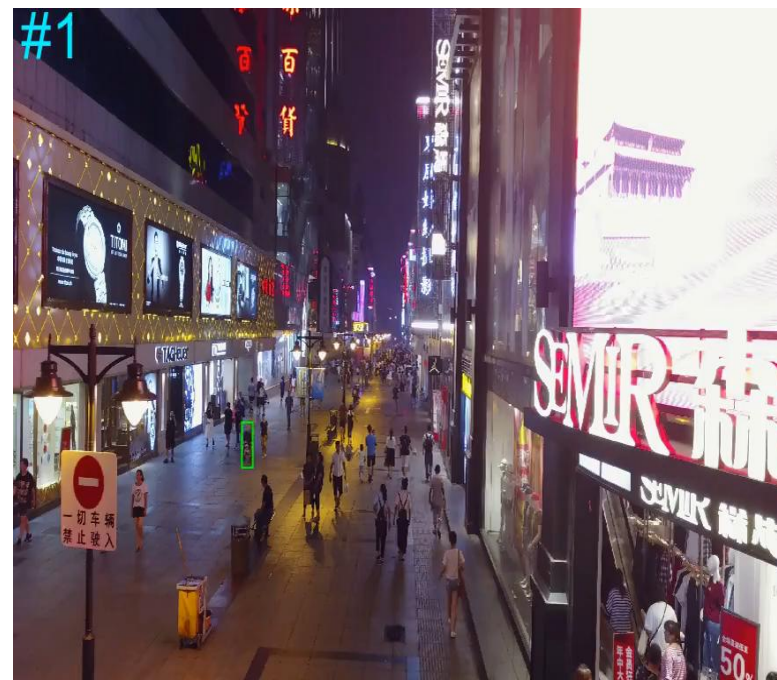
# 实验结果



低光照



长时遮挡



相机和物体运动

黄:其他 绿:OURS

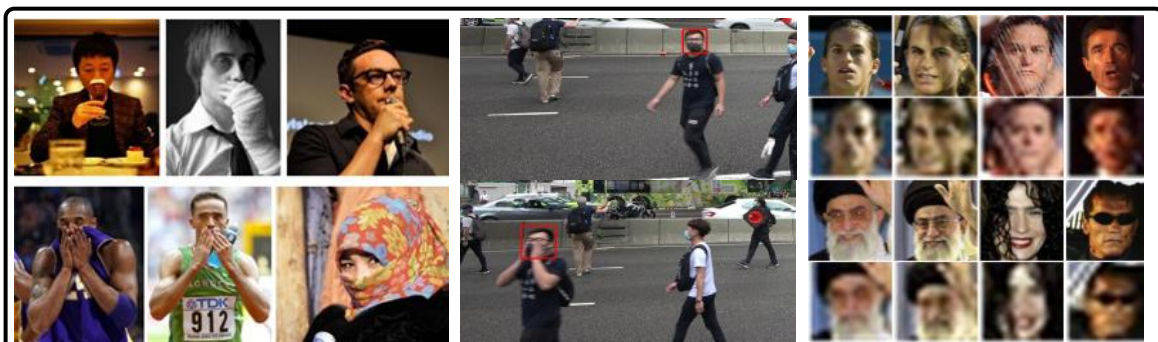
# 总结与展望

## ■ 低质量视觉物体学习理解

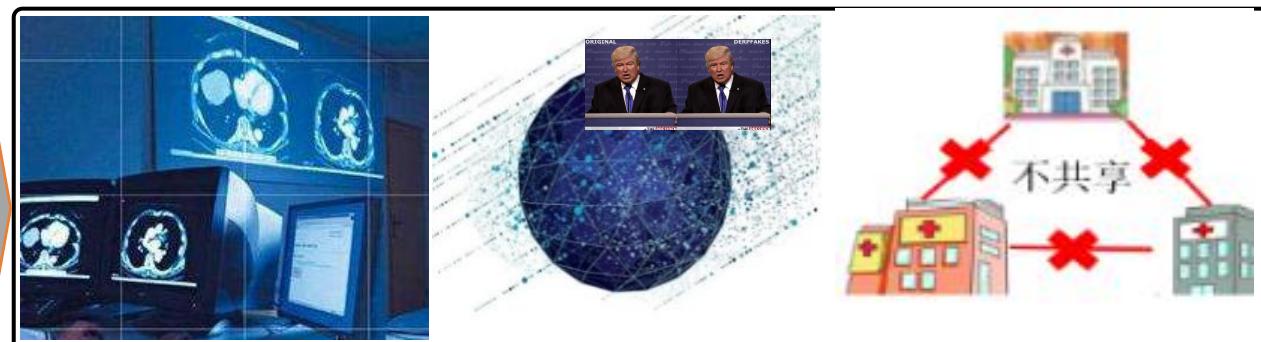
- 蒙面人脸检测识别
- 低分辨率人脸识别
- 视频弱小目标跟踪

## ■ 低质量数据解析与共享学习

- 人工智能+安全



“个体”低质量：遮挡、运动、低分辨率



“集体”低质量：弱标注、强对抗、泛孤立

# 谢谢，请批评指正！



中国科学院 信息工程研究所  
INSTITUTE OF INFORMATION ENGINEERING, CAS

