

CCF走进高校-南京邮电大学

图像跨域重建与识别

汇报人：王楠楠

西安电子科技大学

2020.10.23



汇报提纲

一. 研究背景与意义

二. 图像跨域重建

- ✓ 异质人脸图像合成
- ✓ 图像超分辨重建
- ✓ 其他相关应用

三. 图像身份识别

- ✓ 异质人脸图像识别
- ✓ 跨模态行人重识别
- ✓ 行为定位与识别

汇报提纲

一. 研究背景与意义

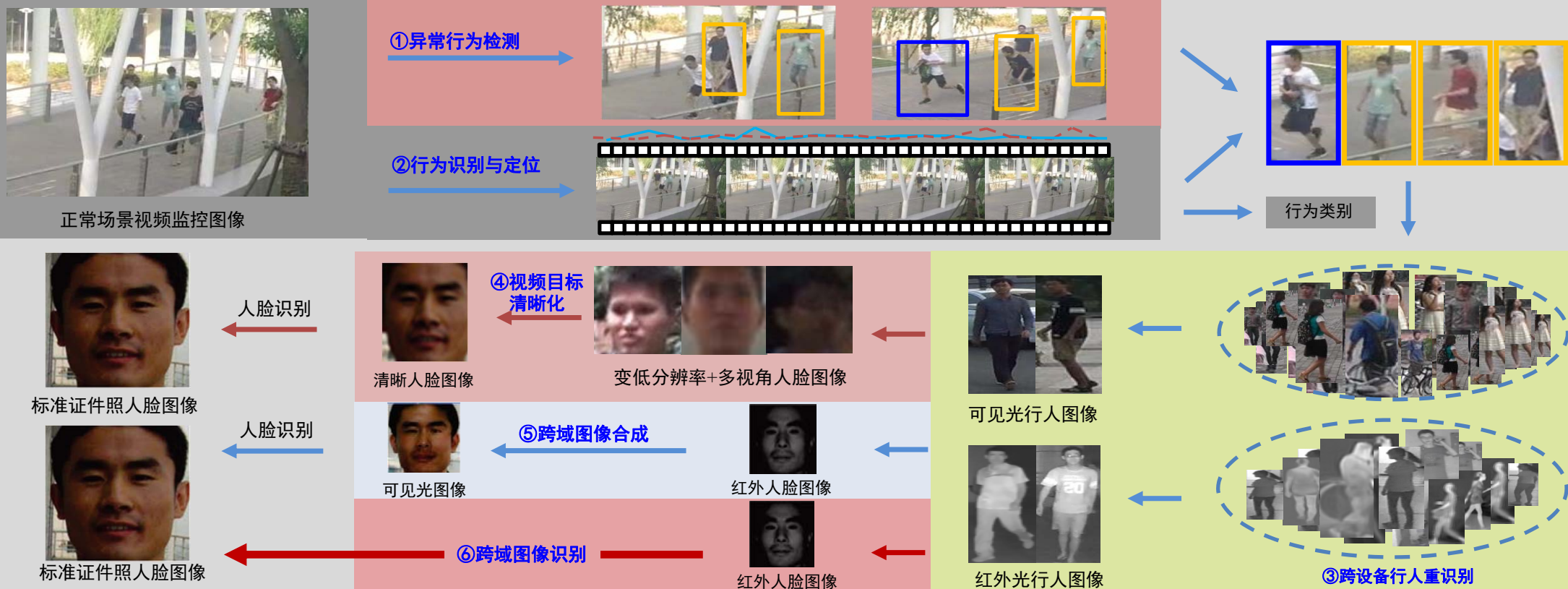
二. 图像跨域重建

- ✓ 异质人脸图像合成
- ✓ 图像超分辨重建
- ✓ 其他相关应用

三. 图像身份识别

- ✓ 异质人脸图像识别
- ✓ 跨模态行人重识别
- ✓ 行为定位与识别

研究背景



面向跨设备多场景视频理解的图像跨域重建与识别

研究背景

实现城市级视频监控系统“找得到、看得清、看得懂”
需借助面向跨设备多场景视频理解的可视分析技术



找得到

研究背景

实现城市级视频监控系统“找得到、看得清、看得懂”
需借助面向跨设备多场景视频理解的可视分析技术

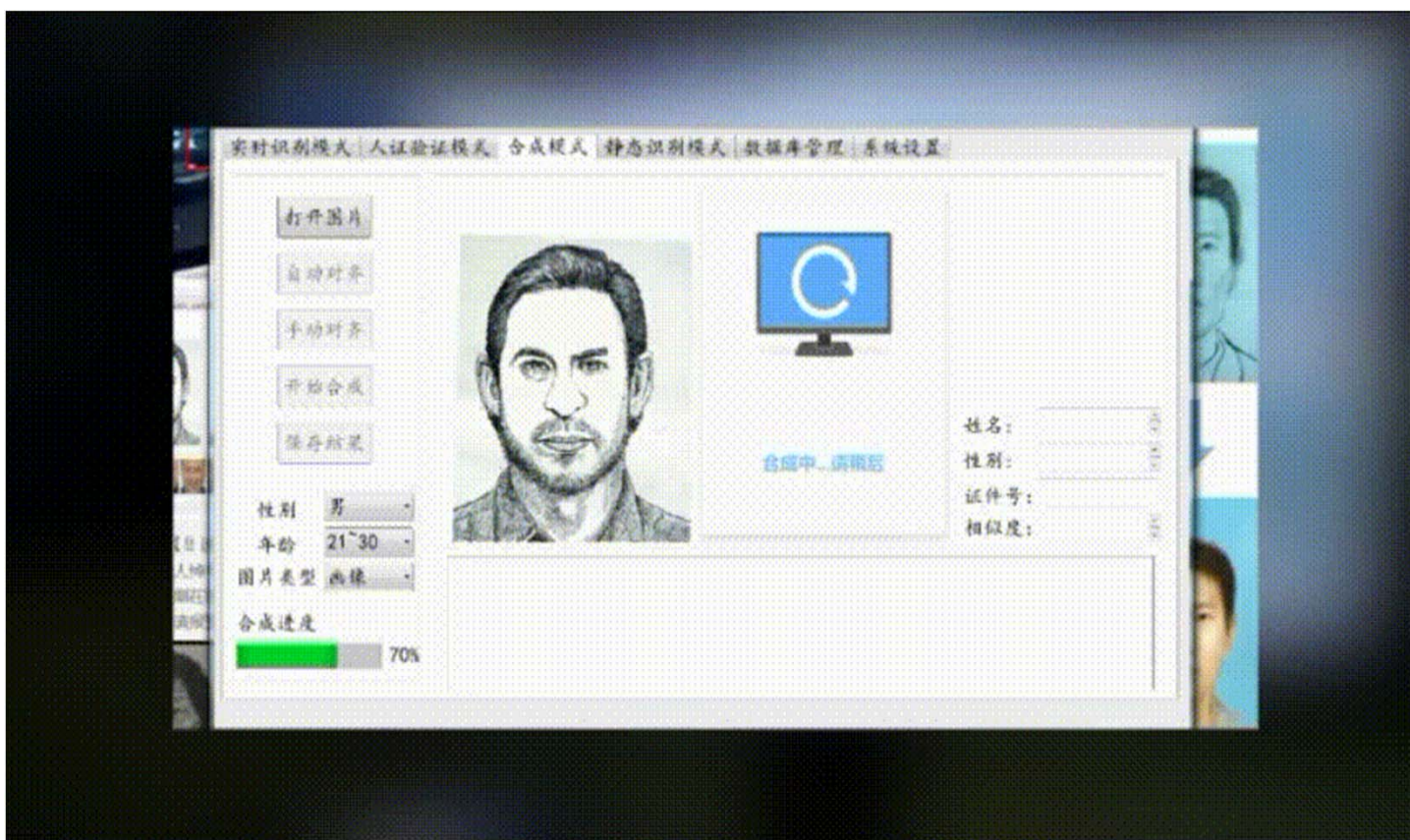
上海市公安局801案件



看得清

研究背景

实现城市级视频监控系统“找得到、看得清、看得懂”
需借助面向跨设备多场景视频理解的可视分析技术



看得懂

汇报提纲

一. 研究背景与意义

二. 图像跨域重建

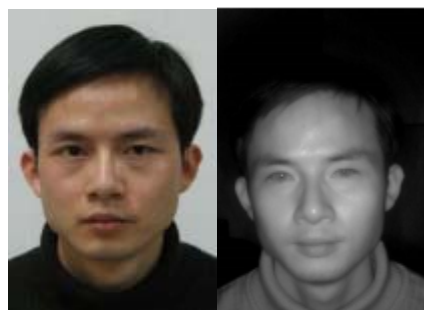
- ✓ 异质人脸图像合成
- ✓ 图像超分辨率重建
- ✓ 其他相关应用

三. 图像身份识别

- ✓ 异质人脸图像识别
- ✓ 跨模态行人重识别
- ✓ 行为定位与识别

研究背景

跨光谱



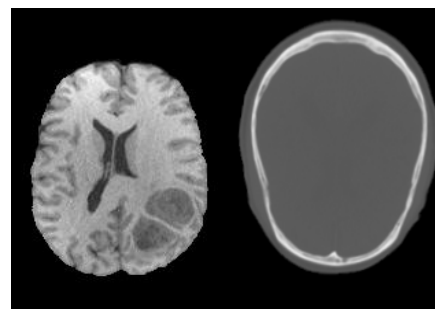
可见光 近红外

跨模式



可见光 画像

跨介质



脑部MRI 脑部CT

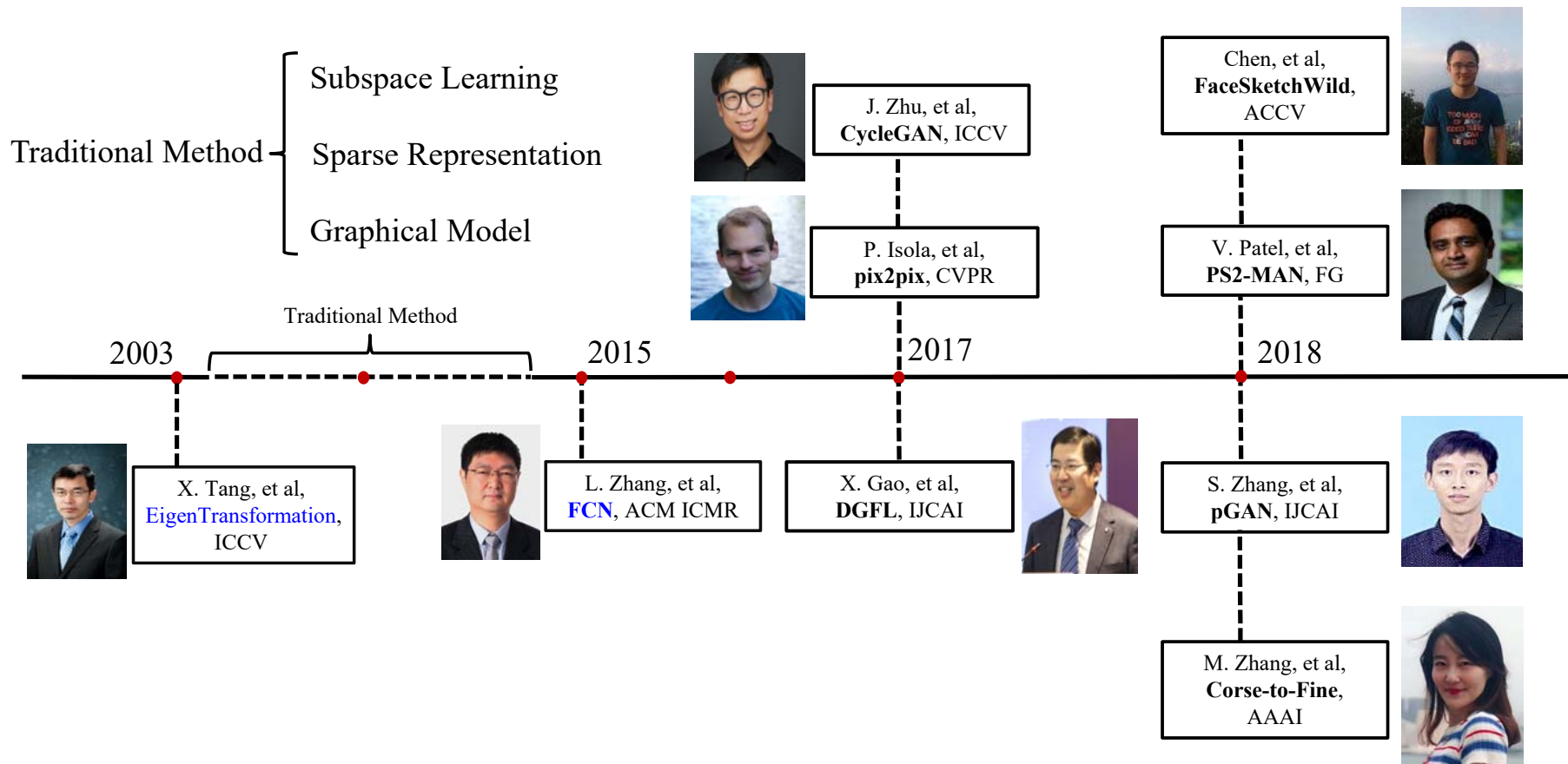
跨模态



可见光 红外

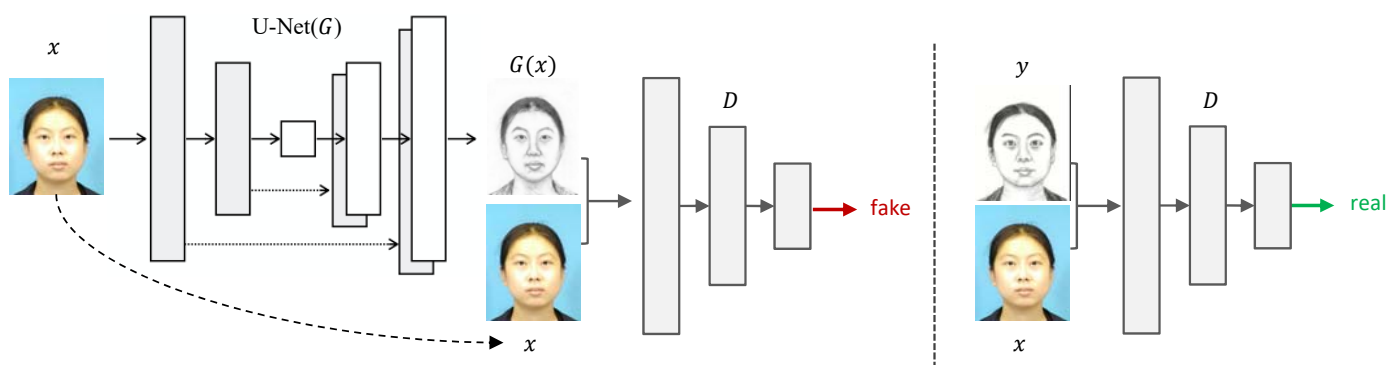
- **模态差异**：同一目标通过**不同传感器**所形成的不同类型的数据，由于传感器对不同物理信息的敏感度不同产生的差异。
- **目标差异**：同一目标由于采集设备的**角度**，**遮挡物**以及目标本身**形变**（如人脸表情、人体姿态等）产生的差异。

发展历史



相关工作

pix2pix(conditional GAN) 图像到图像翻译模型



$$L_{L1}(G) = E_{x,y,z}[\|y - G(x, z)\|_1]$$

$$\underline{L_{cGAN}(G, D) = E_{x,y}[\log D(x, y)] + E_{x,z}[\log(1 - D(x, G(x, z)))]}$$

思想：将条件生成对抗网络用做图像到图像的翻译模型，通过判别器引入对抗损失 L_{cGAN}

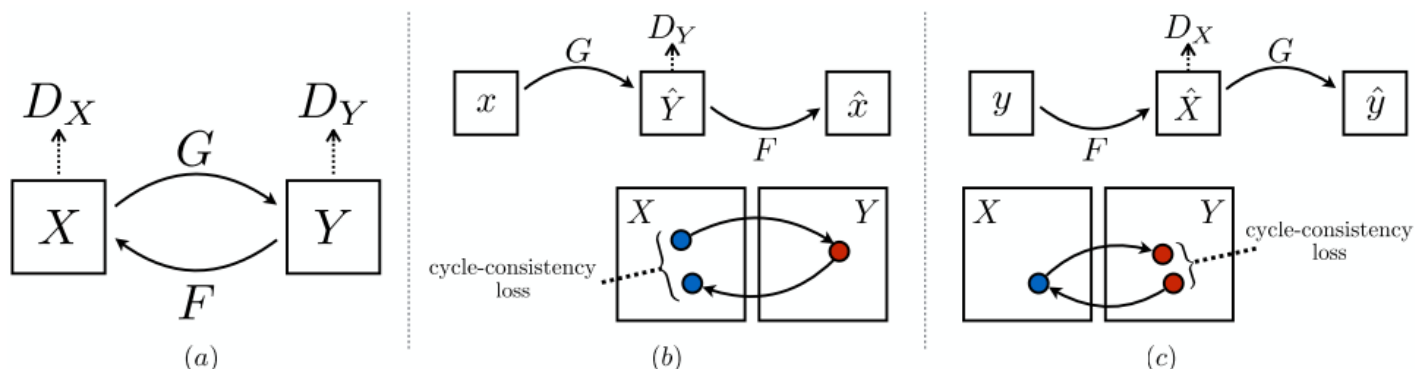
优势：相比于L1损失，对抗损失能使生成的图像更加逼真，与真实图像更加一致，分布更切合真是图像分布

局限：会引入形变与不合理的纹理，人工效应较为明显

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou and Alexei A. Efros. Image-to-Image Translation with Conditional Adversarial Networks. CVPR 2017.

相关工作

CycleGAN 图像到图像翻译模型



$$L_{GAN}(G, D_Y, X, Y) = E_{y \sim p_{data}(y)}[\log D_Y(y)] + E_{x \sim p_{data}(x)}[\log(1 - D_Y(G(x)))]$$

$$L_{GAN}(F, D_X, X, Y) = E_{x \sim p_{data}(x)}[\log D_X(x)] + E_{y \sim p_{data}(y)}[\log(1 - D_X(G(y)))]$$

$$L_{cyc}(G, F) = E_{x \sim p_{data}(x)}[\|F(G(x)) - x\|_1] + E_{y \sim p_{data}(y)}[\|G(F(y)) - y\|_1]$$

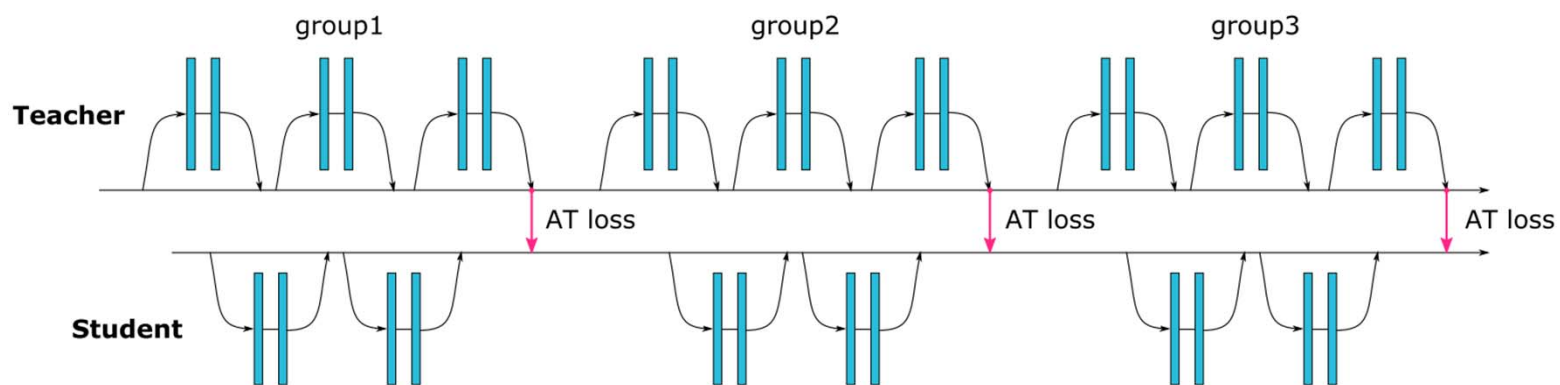
思想：移除L1损失，仅使用对抗损失约束两个域的生成图像，同时引入环一致损失，使网络只注重更加容易的变换（如颜色，纹理），而较少改变内容（形状，物体）

优势：不需要成对的训练数据，实现两个域之间的图像翻译，效果明显

局限：移除了L1损失，仍有内容身份信息丢失

相关工作

Attention Transfer 知识迁移



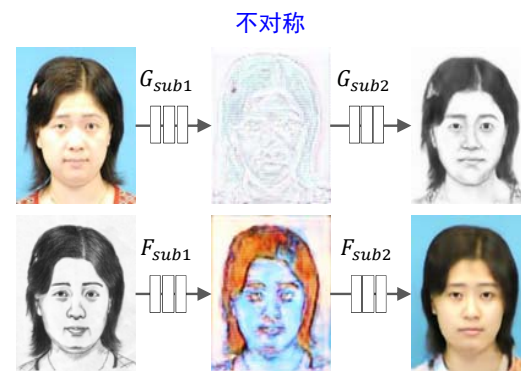
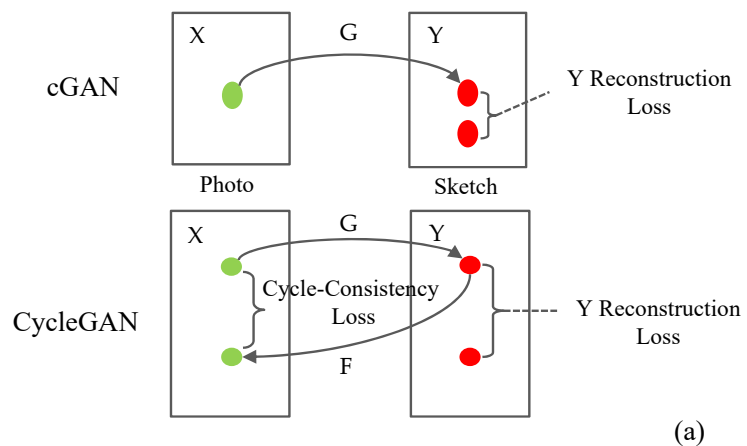
思想：知识迁移，也称作知识蒸馏，是模型压缩的一种手段。这种方法利用大型网络中的知识，来监督训练小型网络，使得小型网络在具有与大型网络一致性能的同时，减少网络的参数量。

提供思路：目前知识迁移大多用于分类判别网络，能否用于图像合成？怎么使用？

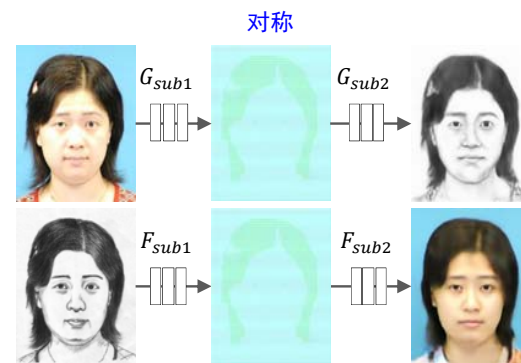
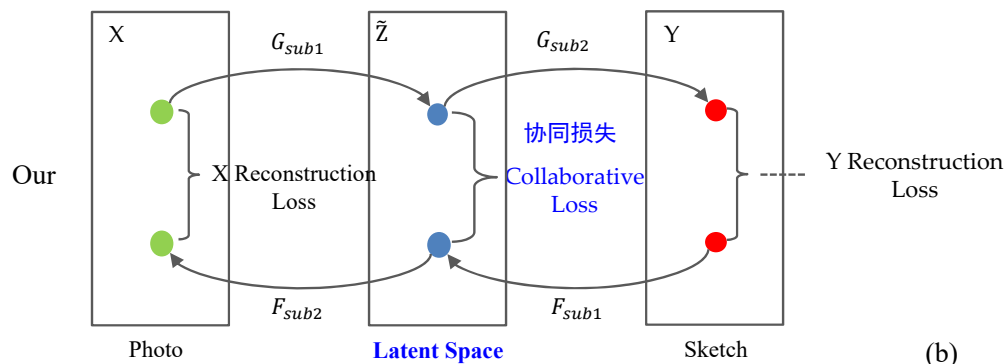
Sergey Zagoruyko and Nikos Komodakis. Paying More Attention to Attention: Improving The Performance of Convolutional Neural Networks via Attention Transfer. ICLR 2017.

基于深度协同网络的人脸画像合成

Motivation



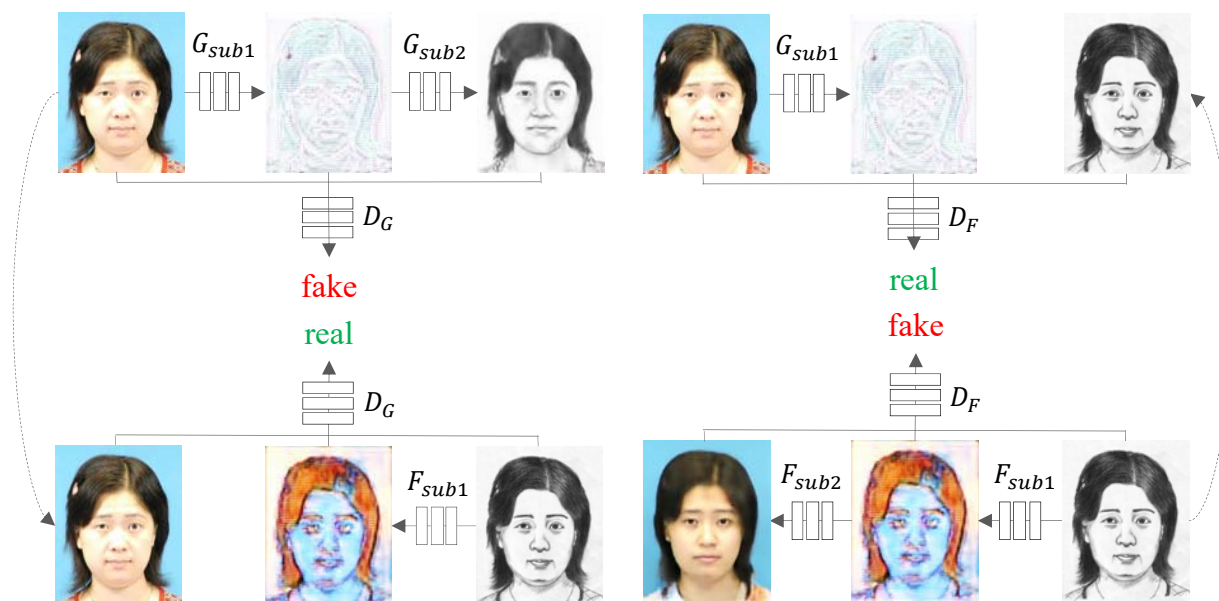
(a)



(b)

基于深度协同网络的人脸画像合成

Method



$$\begin{aligned}
 \mathcal{L}(G, F, D_G) = & E_{x,y,z}[\log D_G(x, F_{sub1}(y, z), y)] + \\
 & E_{x,z}[\log(1 - D_G(x, G_{sub1}(x, z), G(x, z)))] + \\
 & E_{x,y,z}[\|G_{sub1}(x, z) - F_{sub1}(y, z)\|_1] \\
 & E_{x,y,z}[\|G(x, z) - y\|_1] +
 \end{aligned}$$

← 协同对抗损失
← 重构损失

$$\begin{aligned}
 \mathcal{L}(G, F, D_F) = & E_{x,y,z}[\log D_F(x, G_{sub1}(x, z), y)] + \\
 & E_{y,z}[\log(1 - D_F(F(y, z), F_{sub1}(y, z), y))] + \\
 & E_{x,y,z}[\|F_{sub1}(y, z) - G_{sub1}(x, z)\|_1] \\
 & E_{x,y,z}[\|G(y, z) - x\|_1] +
 \end{aligned}$$

基于深度协同网络的人脸画像合成

Comparisons

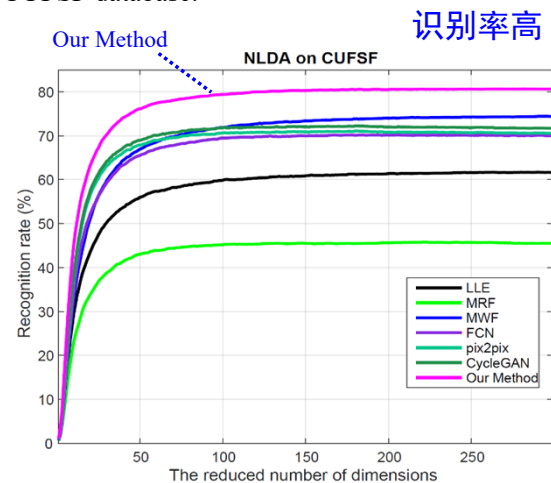
视觉效果好



Average SSIM score (%) on the CUFS database and the CUFSF database
图像质量评价高

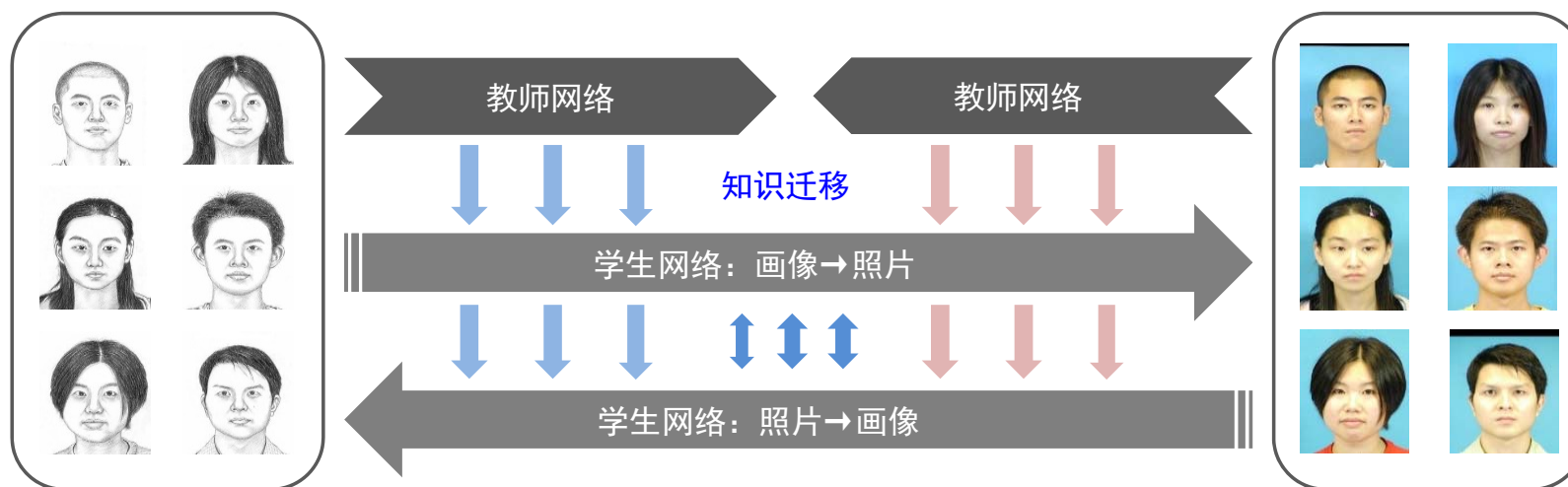
Methods	CUFS(%)	CUFSF(%)
pix2pix	49.39	39.65
CycleGAN	49.61	34.56
Our Method	52.44	42.24

Face recognition accuracy against variations of the number of dimensions reduced by NLDA on the CUFSF database.



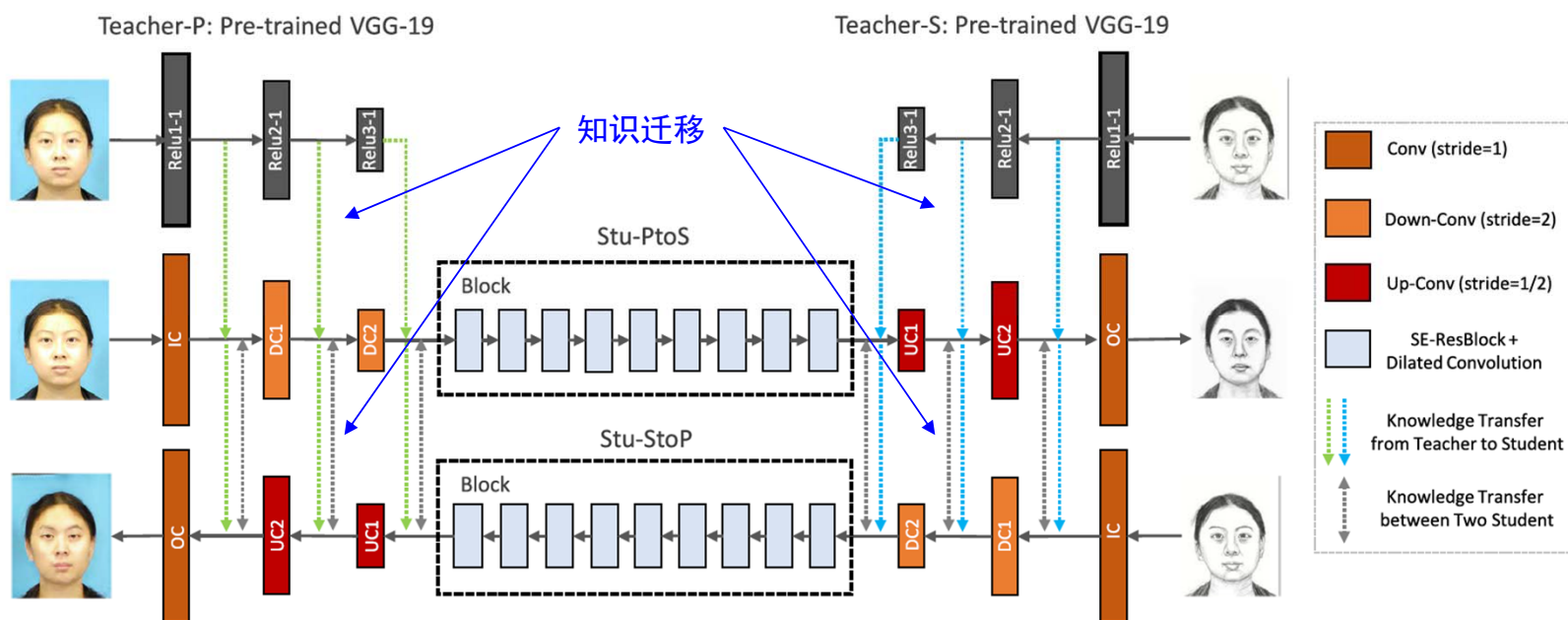
基于知识迁移的人脸画像合成

Motivation



基于知识迁移的人脸画像合成

Method

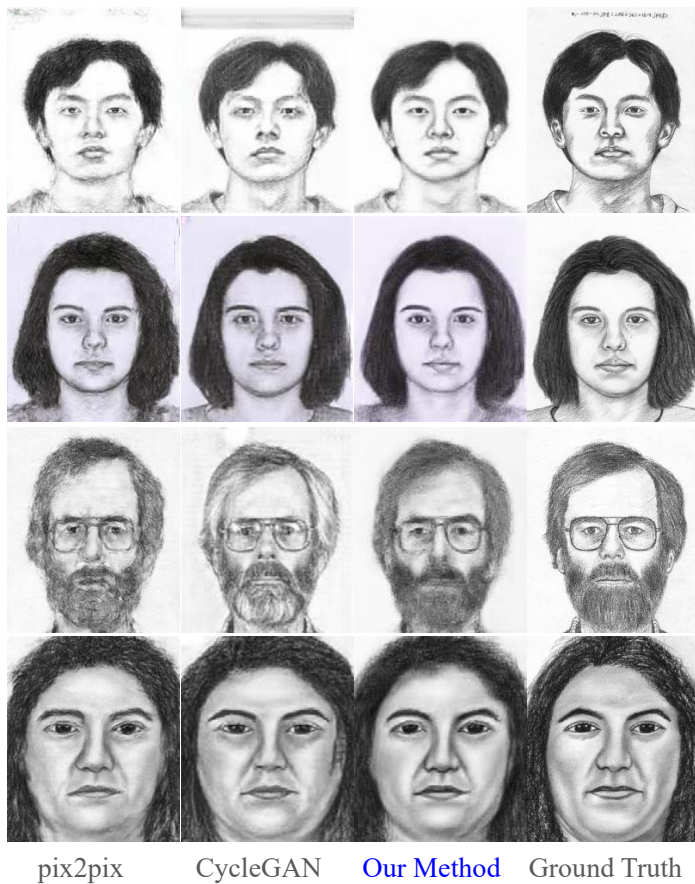


Mingrui Zhu, Nannan Wang, Jie Li and Xinbo Gao. Face Photo-Sketch Synthesis via Knowledge Transfer. IJCAI 2019.

基于知识迁移的人脸画像合成

Comparisons

视觉效果好



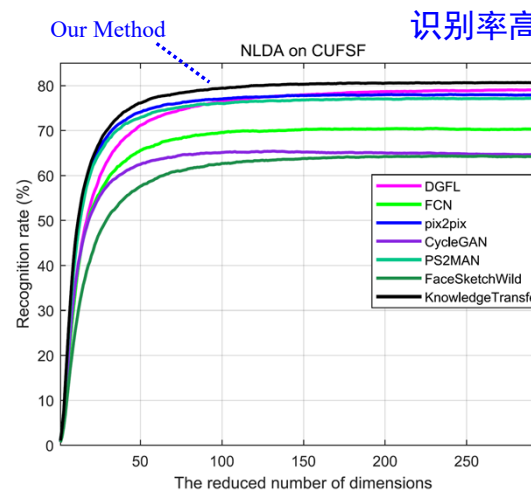
pix2pix CycleGAN Our Method Ground Truth

Average **FSIM** score (%) on the CUFS database and the CUFSF database

图像质量评价高

Methods	CUFS(%)	CUFSF(%)
pix2pix	0.7363	0.7283
CycleGAN	0.7219	0.7088
Our Method	0.7373	0.7311

Face recognition accuracy against variations of the number of dimensions reduced by **NLDA** on the CUFSF database.



识别率高

汇报提纲

一. 研究背景与意义

二. 图像跨域重建

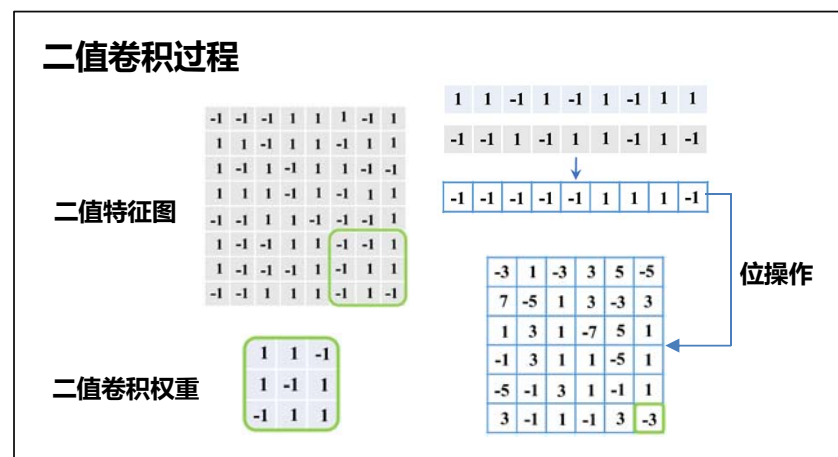
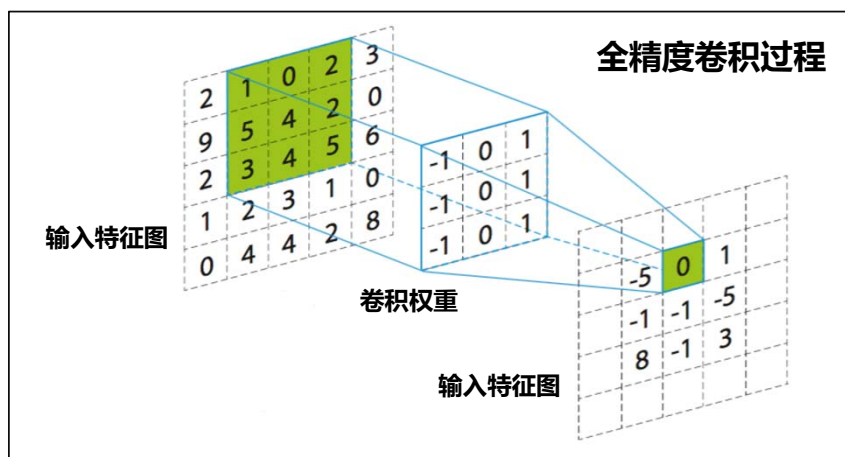
- ✓ 异质人脸图像合成
- ✓ 图像超分辨重建
- ✓ 其他相关应用

三. 图像身份识别

- ✓ 异质人脸图像识别
- ✓ 跨模态行人重识别
- ✓ 行为定位与识别

背景

- 基本思想：将网络中的特征图及卷积权重由全精度数值转换为二值数，网络的卷积过程即可由位操作而实现，可大幅降低模型的推理时间及存储资源的占用。



缺陷：

- 全精度数到二值数的量化过程存在严重的信息丢失，会导致模型性能剧烈下降。
- 现有的模型二值化方法仅针对于图像分类、识别等任务，直接部署于图像超分辨率重建任务难以取得满意的效果。

相关工作

- 基本思想：考虑到量化过程造成的**信息损失**，研究者们提出了多种方法来**降低量化误差**，有效**提升模型性能**。

Bengio等人提出了BNN，网络的权重和激活均为二值数。

BNN

```
for  $k = 1$  to  $L$  do
   $W_k^b \leftarrow \text{Sign}(W_k)$ 
   $s_k \leftarrow a_{k-1}^b W_k^b$ 
   $a_k \leftarrow \text{BatchNorm}(s_k, \theta_k)$ 
  if  $k < L$  then
     $a_k^b \leftarrow \text{Sign}(a_k)$ 
  end if
end for
```

Zhou等人在BNN的基础上增加了可学习的增益项。

DoReFa

Weights:

$$r_o = f_{\omega}^k(r_i) = 2 \text{quantize}_k \left(\frac{\tanh(r_i)}{2 \max(|\tanh(r_i)|)} + \frac{1}{2} \right) - 1.$$

Activations:

$$f_{\alpha}^k(r) = \text{quantize}_k(r).$$

Quantize_K : $r_o = \frac{1}{2^k - 1} \text{round}((2^k - 1)r_i)$

Lin等人通过将多个量化卷积进行加权来逼近全精度卷积的性能。

ABC-Net

Weights:

$$B_i = F_{u_i}(\mathbf{W}) := \text{sign}(\bar{\mathbf{W}} + u_i \text{std}(\mathbf{W})),$$

Activations:

$$\mathbf{A} = H_v(\mathbf{R}).$$

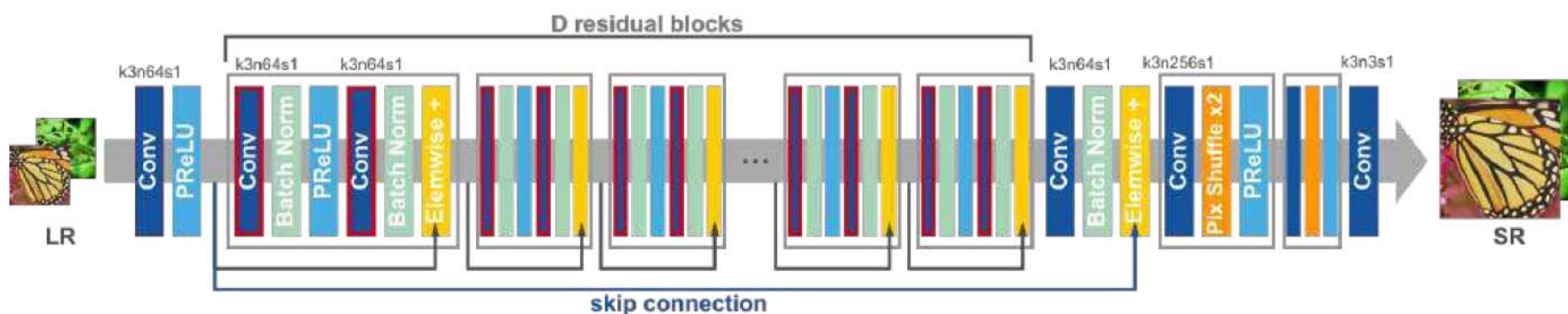
$$\text{Conv}(\mathbf{W}, \mathbf{R}) \approx \text{Conv} \left(\sum_{m=1}^M \alpha_m \mathbf{B}_m, \sum_{n=1}^N \beta_n \mathbf{A}_n \right)$$

缺陷:

- 全精度增益项**无法提升图像高频区域特征量化的精度**，而该精度对于图像超分辨任务中高频信息的预测过程具有严重影响。
- 多个量化卷积加权求和会**显著增加模型的参数数量和计算复杂度**。

相关工作

- 基本思想：将模型量化应用于图像超分辨率重建任务，采用了一种基于二值化权重和全精度特征图的卷积方式。



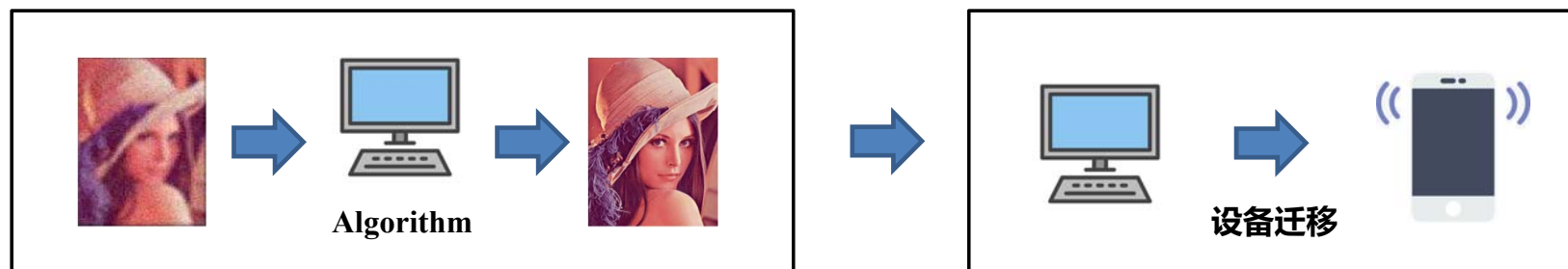
图中带有红框的卷积层的卷积权重为二值数。

缺陷：

- 仅对网络的权重进行了量化，而模型中的特征图依然为全精度数值。网络并不是标准的二值化网络，其卷积计算的速度未能充分展现。
- 该量化卷积过程对于图像信息保留程度较低，模型性能受限。

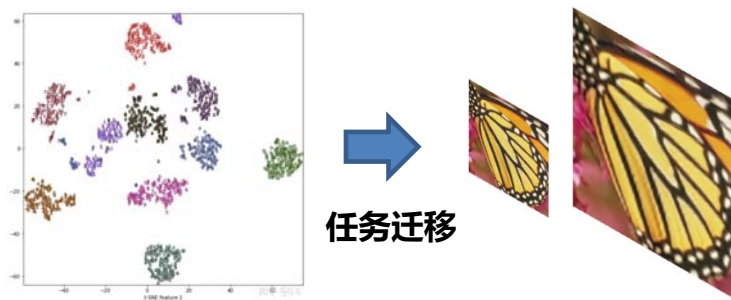
Motivation

1. 受制于存储和计算资源的限制，目前图像超分辨率网络大多**难以直接应用于低计算性能的设备中**。作为一种能够显著降低模型大小和推理时间的方法，面向图像超分辨任务的模型二值化方法研究具有重大的应用价值。



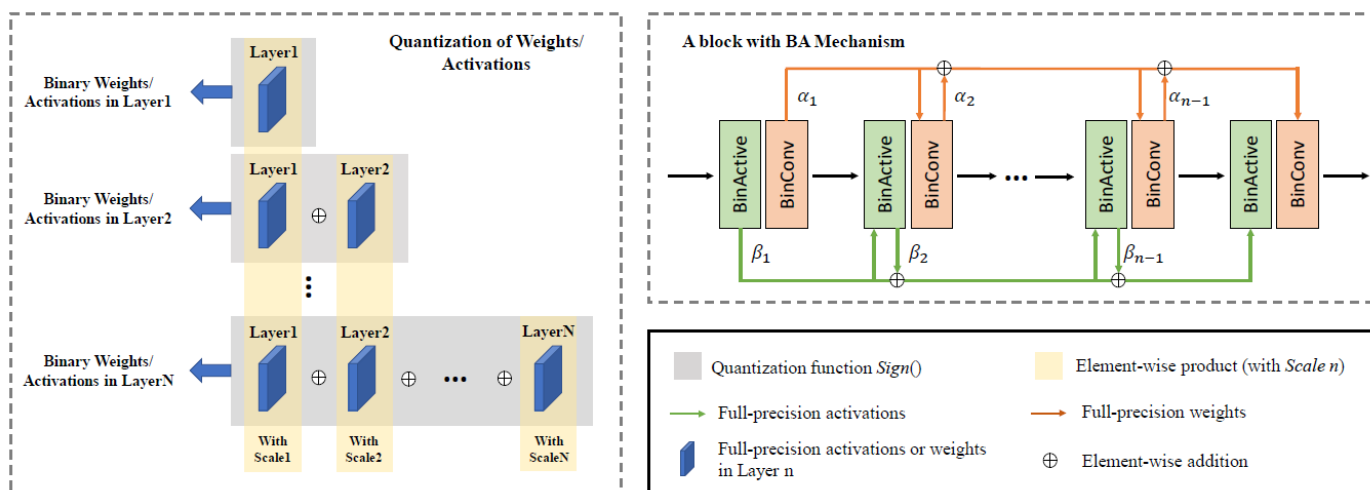
2. 现有二值化算法直接部署于图像超分辨率重建任务上**难以达到满意的效果**：

- 图像超分辨任务对于数值量化精度的依赖性；
- 网络模型的计算成本。



方法过程

- 基本思想：提出一种面向图像超分辨任务的模型量化方法，能够沿着网络推理的方向逐步提升模型量化的精度，显著提升图像重建模型的性能。



- 第一个基于二值神经网络(权值和特征图均为二值数)的图像超分辨率重建工作。
- 提出了一种比特累积机制，该机制能够通过迭代的主题实现二值卷积到全精度卷积的近似，在不引入额外推理参数的情况下显著提升了模型性能。

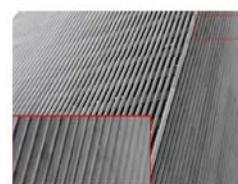
实验结果

客观评价:

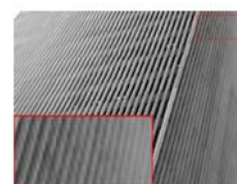
Methods	Scale	Set5		Set14		Urban100	
		PSRN	SSIM	PSRN	SSIM	PSRN	SSIM
SRResNet_Ma et.al.	×2	35.66	0.946	31.56	0.897	28.76	0.882
SRResNet_BAM	×2	37.51	0.956	33.03	0.912	30.79	0.915
SRResNet_Ma et.al.	×4	30.34	0.864	27.16	0.756	24.48	0.728
SRResNet_BAM	×4	31.57	0.883	28.16	0.769	24.30	0.755

Methods	Scale	Set5		Set14		B100		Urban100	
		PSRN	SSIM	PSRN	SSIM	PSRN	SSIM	PSRN	SSIM
SRResNet	×2	37.76	0.958	33.27	0.914	31.95	0.895	31.28	0.919
Bicubic	×2	33.66	0.930	30.24	0.869	29.56	0.843	26.88	0.840
SRResNet_BNN	×2	35.21	0.942	31.55	0.896	30.64	0.876	28.01	0.869
SRResNet_DoReFa	×2	36.09	0.950	32.09	0.902	31.02	0.882	28.87	0.880
SRResNet_ABC	×2	36.34	0.952	32.28	0.903	31.16	0.884	29.29	0.891
SRResNet_BAM	×2	37.21	0.956	32.74	0.910	31.60	0.891	30.20	0.906
SRResNet	×3	34.07	0.922	30.04	0.835	28.91	0.798	27.50	0.837
Bicubic	×3	30.39	0.868	27.55	0.774	27.21	0.739	24.46	0.735
SRResNet_BNN	×3	31.18	0.877	28.29	0.799	27.73	0.765	25.03	0.758
SRResNet_DoReFa	×3	32.44	0.903	28.99	0.811	28.21	0.778	25.84	0.783
SRResNet_ABC	×3	32.69	0.908	29.24	0.820	28.35	0.782	26.12	0.797
SRResNet_BAM	×3	33.33	0.915	29.63	0.827	28.61	0.790	26.69	0.816
SRResNet	×4	31.76	0.888	28.25	0.773	27.38	0.727	25.54	0.767
Bicubic	×4	28.42	0.810	26.00	0.703	25.96	0.668	23.14	0.658
SRResNet_BNN	×4	29.33	0.826	26.72	0.728	26.45	0.692	23.68	0.683
SRResNet_DoReFa	×4	30.38	0.862	27.48	0.754	26.87	0.708	24.45	0.720
SRResNet_ABC	×4	30.78	0.868	27.71	0.756	27.00	0.713	24.54	0.729
SRResNet_BAM	×4	31.24	0.878	27.97	0.765	27.15	0.719	24.95	0.745

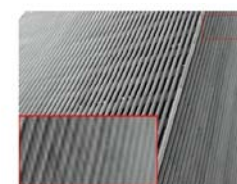
主观评价:



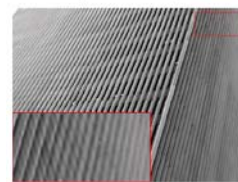
SRResNet
(PSNR/SSIM)



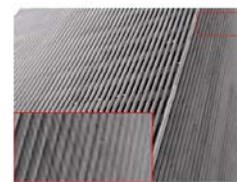
Bicubic
(20.96 / 0.522)



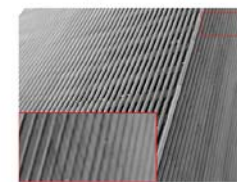
SRResNet-BNN
(21.97 / 0.590)



SRResNet-DoReFa
(22.59 / 0.615)



SRResNet-ABC
(22.71 / 0.612)



SRResNet-BAM
(22.91 / 0.639)



SRResNet
(W-32 A-32)



SRResNet-BNN
(W-1 A-1)

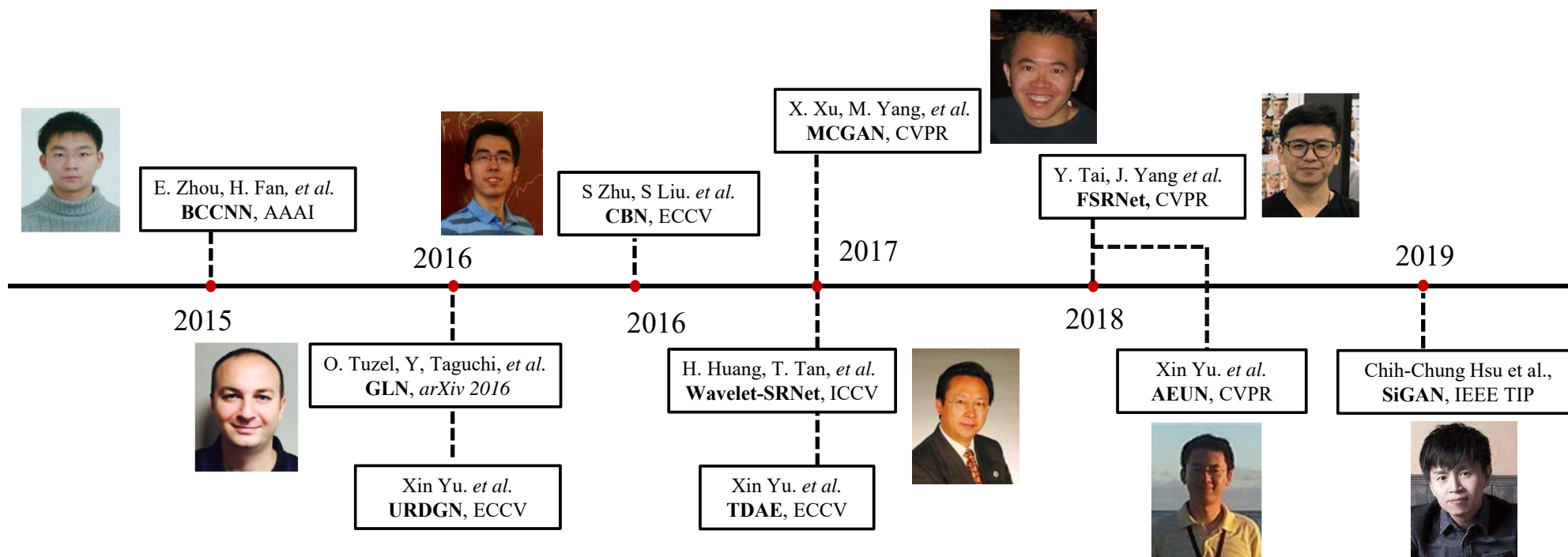


SRResNet-ABC-Net
(W-1 A-1)



SRResNet-BAM(Ours)
(W-1 A-1)

发展历史

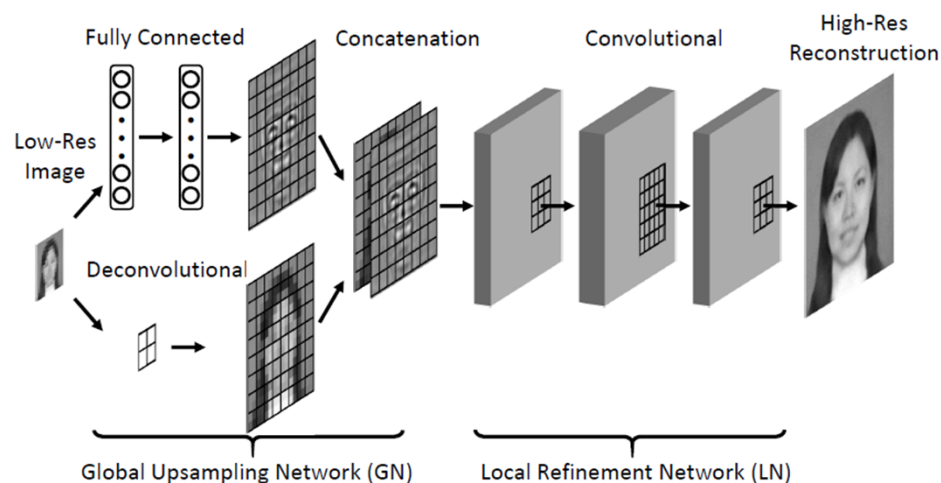


BCCNN: Bi-channel Convolutional Neural Network
GLN: Global Local Network
URDGN: Ultra-Resolving by Discriminative Generative Networks
CBN: Cascaded Bi-Network
MCCGN: multi-class GAN

TDAE: Transformative Discriminative Autoencoder
FSRNet: Face Super-Resolution Network
AEUN: Attribute Embedded Upsampling Network
SiGAN: Siamese GAN

相关工作

- 基本思想：构建**二分支网络**，利用**全局特征**和**局部特征**的递进式学习提升生成图像质量。



4× GN		8× GN	
deconv4	fc-512 fc-256 fc-512 fc-(128 × 128)	deconv8	fc-256 fc-256 fc-256 fc-(128 × 128)
concatenation		concatenation	

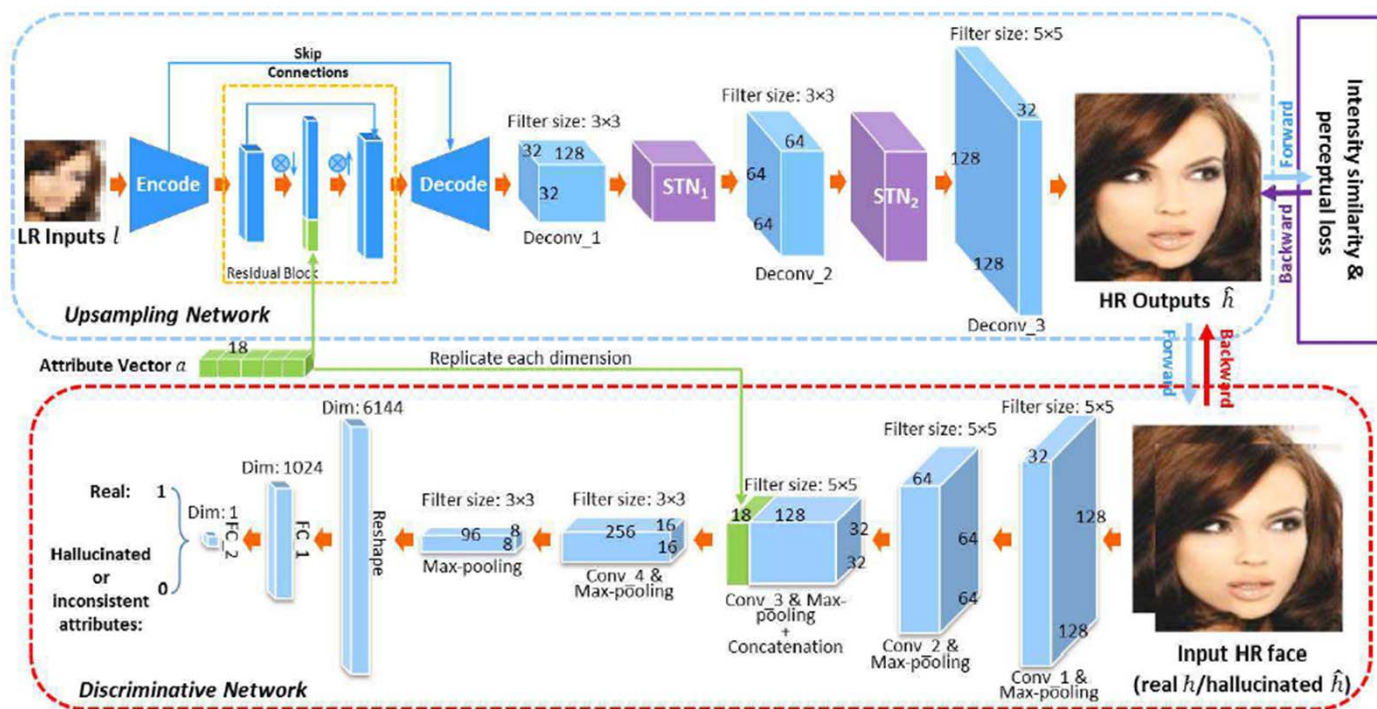
4 Layer LN (LN4)	6 Layer LN (LN6)	8 Layer LN (LN8)
conv5-16 conv7-64 conv5-16	conv5-16 conv7-32 conv7-64 conv7-32 conv5-16	conv5-16 conv7-32 conv7-64 conv7-64 conv7-32 conv5-16
conv5-1	conv5-1	conv5-1

缺陷：

- 全局上采样网络中的全连接层的**参数数量十分庞大**，对硬件成本依赖较高。
- 人脸图像的重建过程为考虑**人脸的先验信息**，人脸的高频区域重建质量较低。

相关工作

- 基本思想：通过属性嵌入的方式，更改人脸的表示向量，使重建图像具有属性一致特性

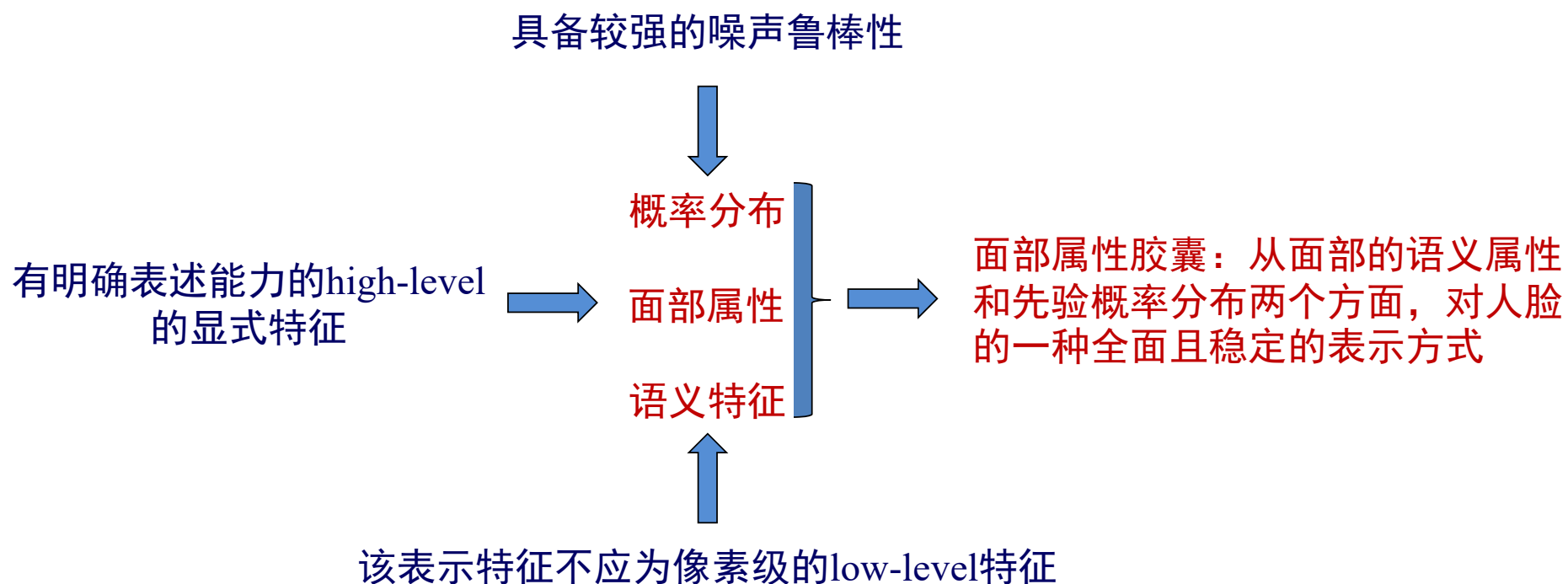


缺陷:

- 属性嵌入需要人为操作，在属性无法预知的场合难以取得较好的重建效果。
- 图像的编码方式对于噪声的鲁棒性较低。在现实场景中，图像中的噪声对于图像重建的质量会产生极大影响。

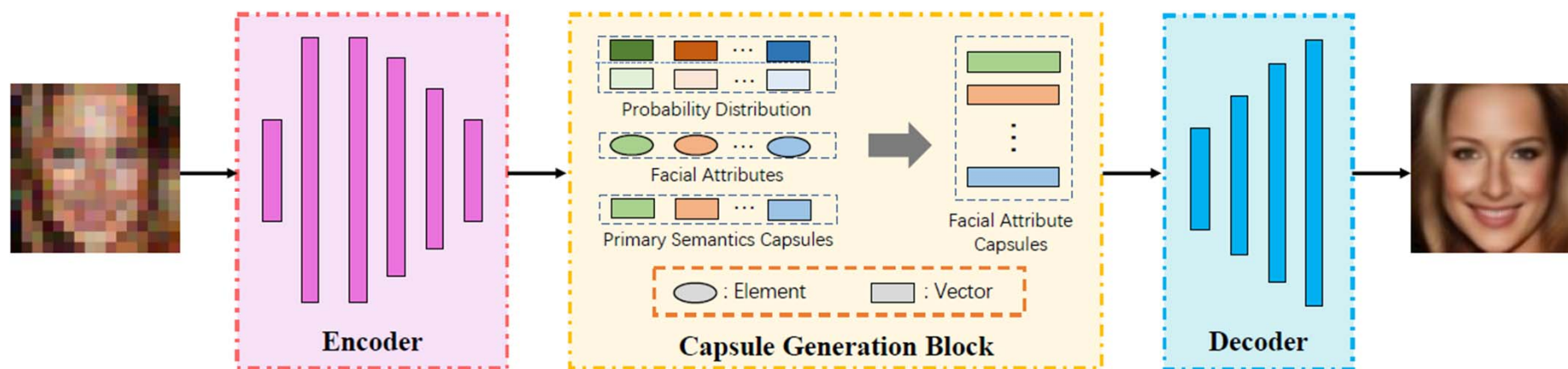
Motivation

面对噪声人脸超分辨问题，如何学习人脸图像在不同噪声环境下的不变表示特征？



方法过程

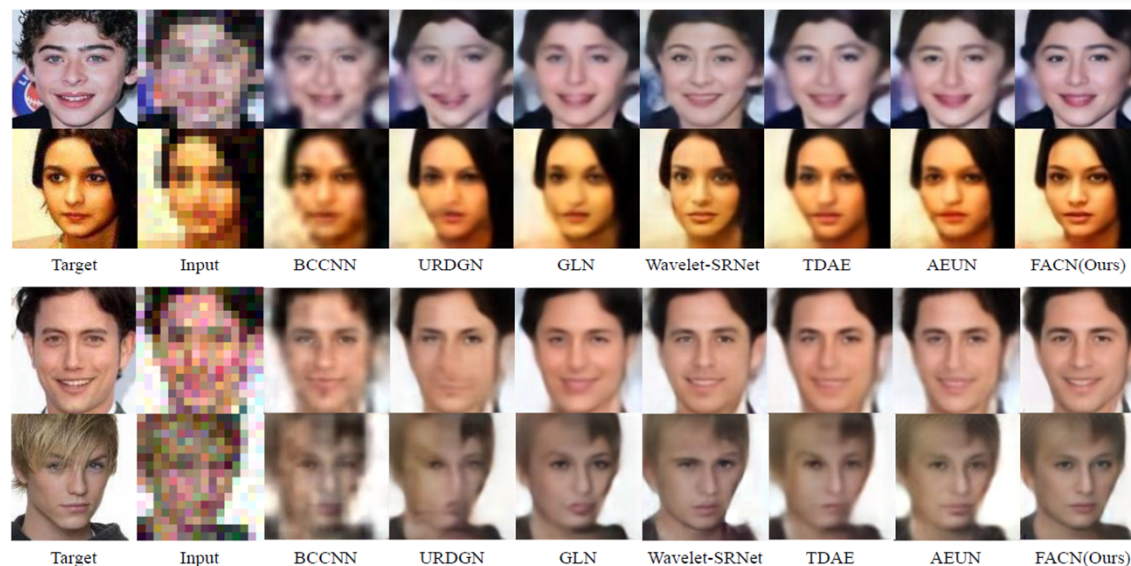
- 基本思想：引入了**胶囊网络**的概念进行人脸超分辨，以**人脸属性**和**先验概率分布**相结合的方式构建出一种新型的胶囊结构，对于**噪声**和**模糊**具有较强**鲁棒性**



- 与现有的基于向量的表示方法相比，基于胶囊的表示模型能够有效地减少由于任务本身的模糊和噪声所引入的歧义。
- 所提出的人脸属性胶囊通过语义属性和概率分布两个方面对于人脸进行了更加全面的描述，对于噪声和模糊具有强鲁棒性。

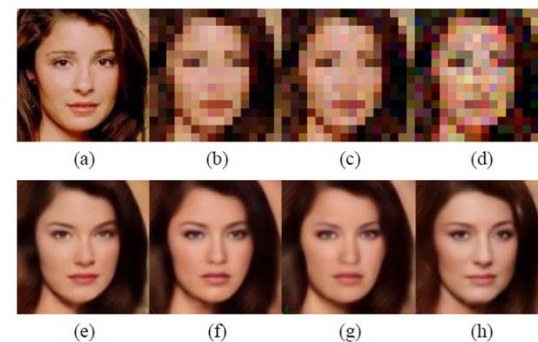
实验结果

Methods	<i>Bic</i>			<i>BicN</i>			<i>BBicN</i>		
	PSRN	SSIM	IFC	PSRN	SSIM	IFC	PSRN	SSIM	IFC
Bicubic	23.98	0.6505	0.6588	23.13	0.6088	0.4497	21.09	0.5329	0.3457
BCCNN	25.29	0.7135	0.9524	23.94	0.6615	0.6677	22.21	0.6154	0.5453
GLN	26.04	0.7427	1.0783	24.01	0.6718	0.7179	22.45	0.6365	0.5758
URDGN	24.54	0.6785	0.6981	23.80	0.6444	0.5502	21.01	0.5482	0.3650
Wavelet-SRNet	24.43	0.6891	0.7835	23.95	0.6768	0.7270	22.48	0.6428	0.6035
TDAE	26.29	0.7411	1.1523	24.16	0.6778	0.7321	22.81	0.6511	0.6211
AEUN	26.37	0.7477	1.1605	24.24	0.6801	0.7535	22.83	0.6514	0.6254
FACN (Ours)	26.79	0.7684	1.2515	24.61	0.7009	0.8060	23.14	0.6714	0.6775



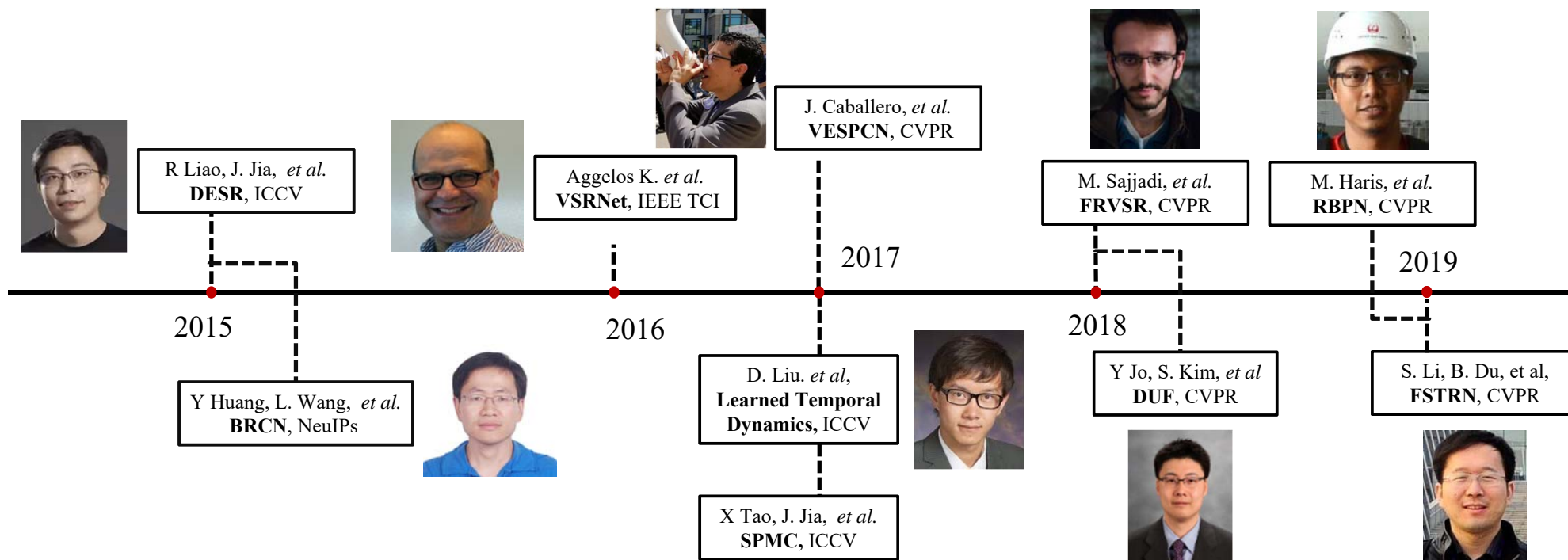
Methods	Performance	Methods	Performance
Bicubic	0.8058	BCCNN	0.8570
URDGN	0.8212	GLN	0.8580
Wavelet-SRNet	0.8820	TDAE	0.8680
AEUN	0.8694	FACN(Ours)	0.8922

以人脸识别率为评价标准，与现有的方法相比，我们的方法具有**较高的识别准确率**。



随着噪声等级的增加（从左至右），重建人脸的质量并没有明显降低。

发展历史

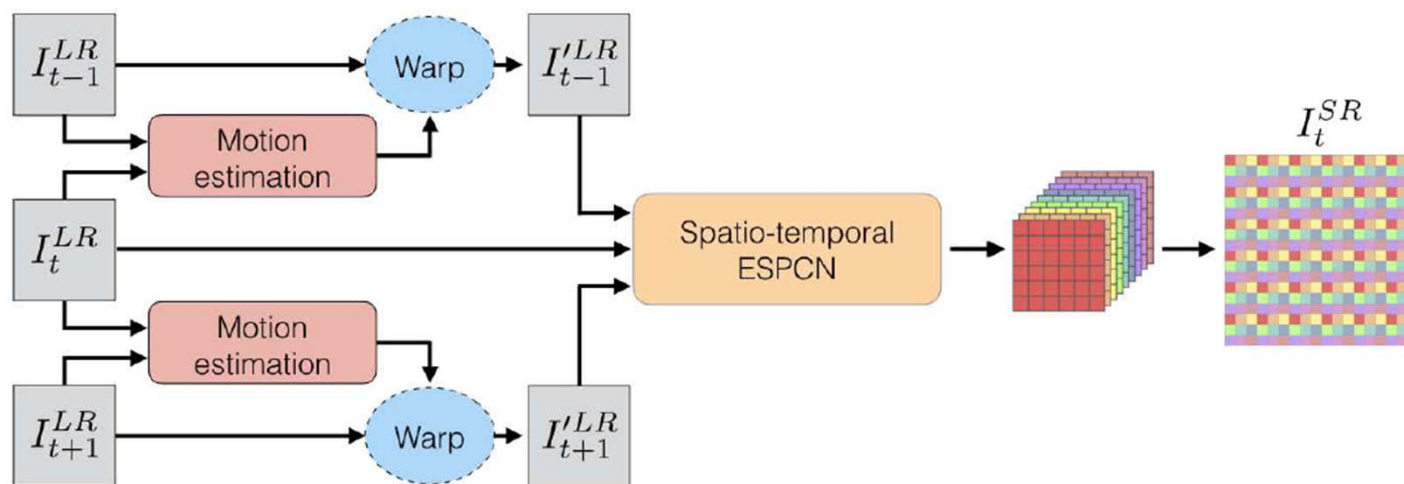


DESR: Draft Ensemble Super Resolution
BRCN: Bidirectional Recurrent Convolutional Networks
LTD: Learned Temporal Dynamics
SPMC: Sub-Pixel Motion Compensation
VESPCN: Video-ESPCN(Efficient Sub-Pixel Convolutional Network)

FRVSR: Frame-Recurrent Video Super-Resolution
DUF: Dynamic Upsampling Filters
RBPN: Recurrent Back-Projection Network
FSTRN: Fast Spatio-Temporal Residual Network

相关工作

- 基本思想：构建**运动补偿模块**，将**运动补偿**后的各帧图像输入网络进行高分辨率图像重建。



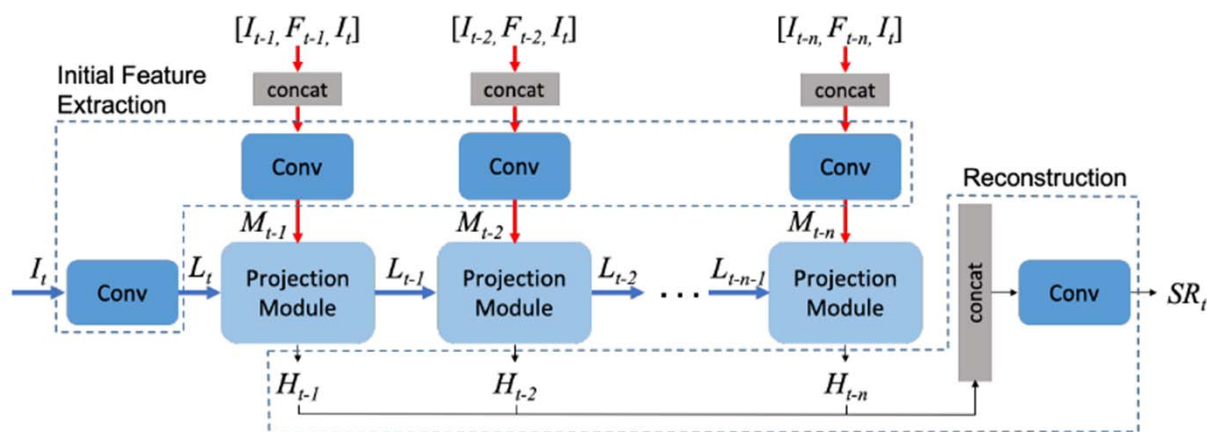
缺陷：

- 图像重建质量对于运动补偿的精度依赖性高，容易**因为补偿的误差而造成性能显著下降**。
- 模型**受限的运动建模能力**导致其性能并不能够随着网络的输入帧数量的增多而提高。

Jose Caballero, Christian Ledig, Andrew Aitken, Alejandro Acosta, Johannes Totz, Zehan Wang, Wenzhe Shi. Real-time video super-resolution with spatio-temporal networks and motion compensation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 4778-4787.

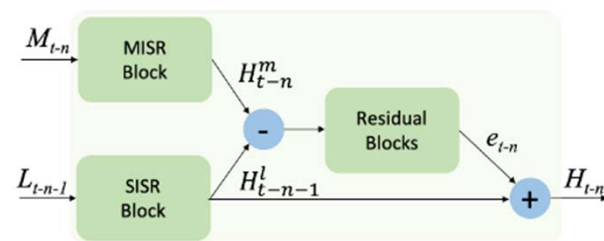
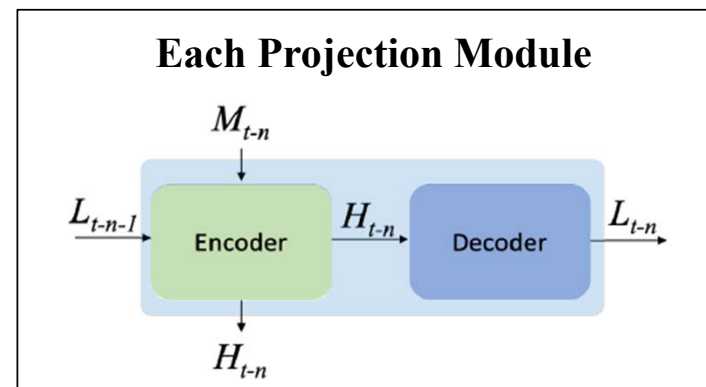
相关工作

□ 基本思想：通过**迭代反向投影网络**，逐步的将相邻帧组合在一个**迭代细化框架**中。

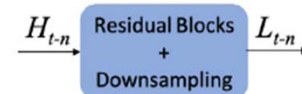


缺陷：

- 迭代的反向投影网络**计算量庞大**，生成结果时间过长。
- 相邻帧逐步结合的方式会使得**各帧信息的利用率受限**，从而消耗了极大的计算资源也未能够取得满意的表现。



(a) Encoder (the back-projection)

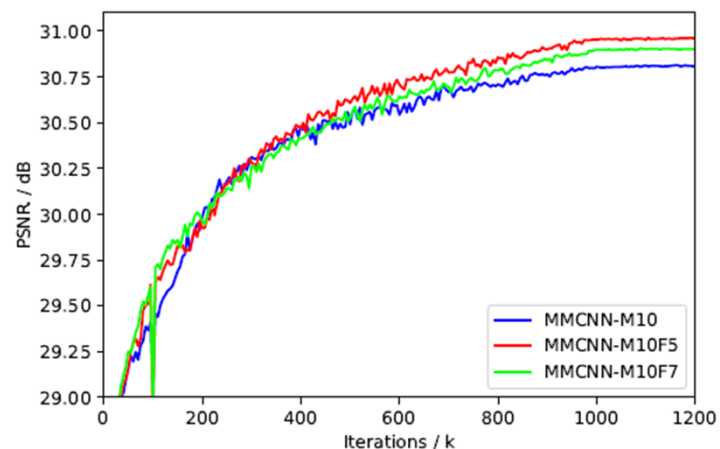
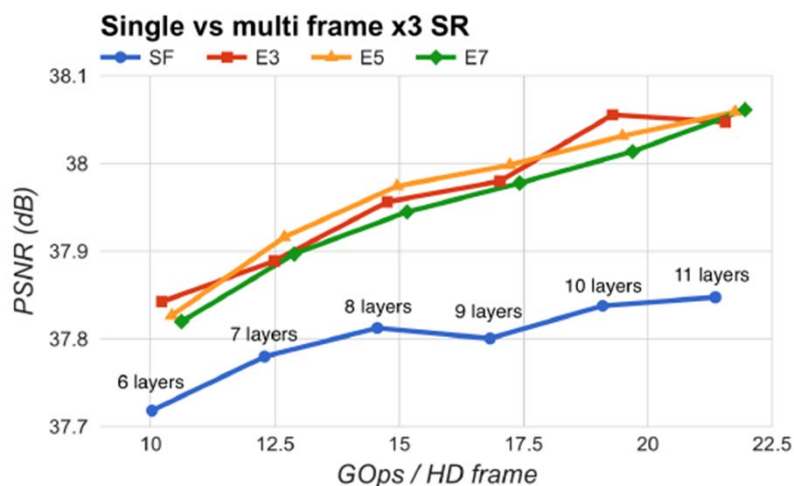


(b) Decoder

Jing Liao, Yuan Yao, Lu Yuan, Gang Hua, and Sing Bing Kang, "Visual attribute transfer through deep image analogy," ACM Transactions on Graphics, TOG, vol. 36, no. 4, 2017.

Motivation

面对视频中复杂的时序依赖关系，如何有效的提升模型对于帧间运动的建模能力



输入帧数量过多使得网络难以对视频中复杂的时序依赖关系进行建模。

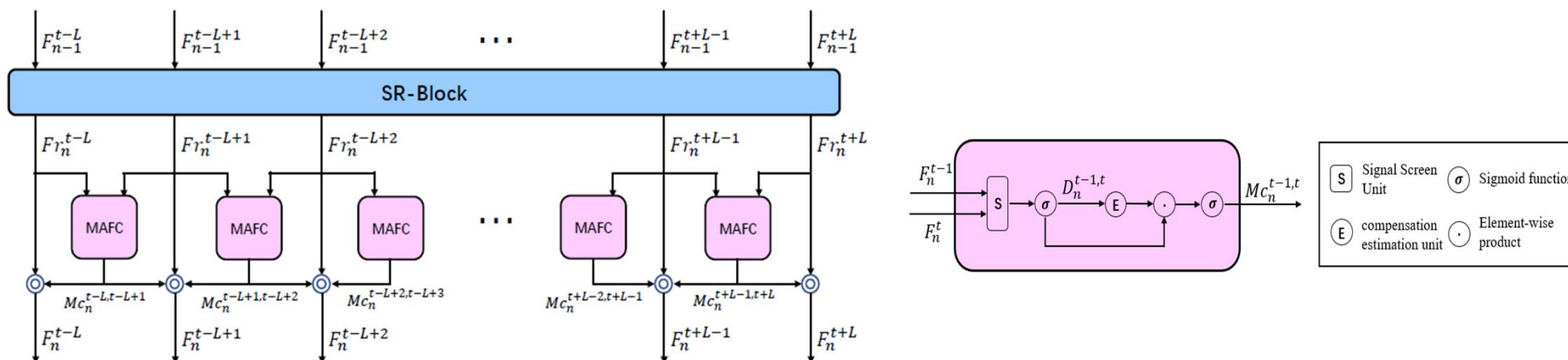
Jose Caballero, Christian Ledig, Andrew Aitken, Alejandro Acosta, Johannes Totz, Zehan Wang, and Wenzhe Shi. Real-time video super-resolution with spatio-temporal networks and motion compensation. CVPR, pages 4778–4787, 2017.

网络的学习能力是有限的，因此，并不意味着输入帧数量越多效果就越好。

Wang Z, Yi P, Jiang K, et al. Multi-memory convolutional neural network for video super-resolution[J]. IEEE Transactions on Image Processing, 2018, 28(5): 2530-2544.

方法过程

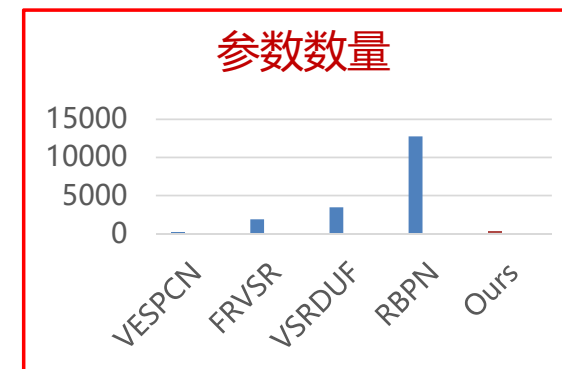
- 基本思想：将视频的帧间运动补偿操作转移到图像特征级进行，以遗忘，保留和增强的运动补偿方式，提升补偿的精度以及重建图像的质量



- 特征级的运动补偿操作能够更加充分的捕获到帧间的运动关系，同时也能够避免在运动补偿不准确时存在的重建质量急剧下降问题。
- 运动补偿的自适应反馈机制能够显著提升网络对于复杂的视频时序依赖关系建模的能力。

实验结果

Methods	$T = 3$		$T = 5$		$T = 7$	
	PSRN	SSIM	PSRN	SSIM	PSRN	SSIM
Bicubic	29.95	0.8416	29.95	0.8416	29.95	0.8416
DESR	32.19	0.8929	32.30	0.8953	32.09	0.8929
VESPCN	33.07	0.9097	33.14	0.9112	32.79	0.9055
LIU et.al	32.70	0.9033	32.82	0.9063	32.66	0.9033
SPMC	33.03	0.9066	33.23	0.9099	33.44	0.9132
FRVSR	33.26	0.9105	33.42	0.9129	33.53	0.9147
VSR_DUF	34.38	0.9290	34.14	0.9245	33.82	0.9214
FSTRN	32.96	0.9059	33.11	0.9089	33.07	0.9085
RBPN	33.16	0.9084	33.67	0.9158	33.91	0.9232
MAFN	34.15	0.9237	34.59	0.9279	34.81	0.9318



与其他方法相比，我们的模型具有更加轻量级的结构以及更高的图像重建质量。



汇报提纲

一. 研究背景与意义

二. 图像跨域重建

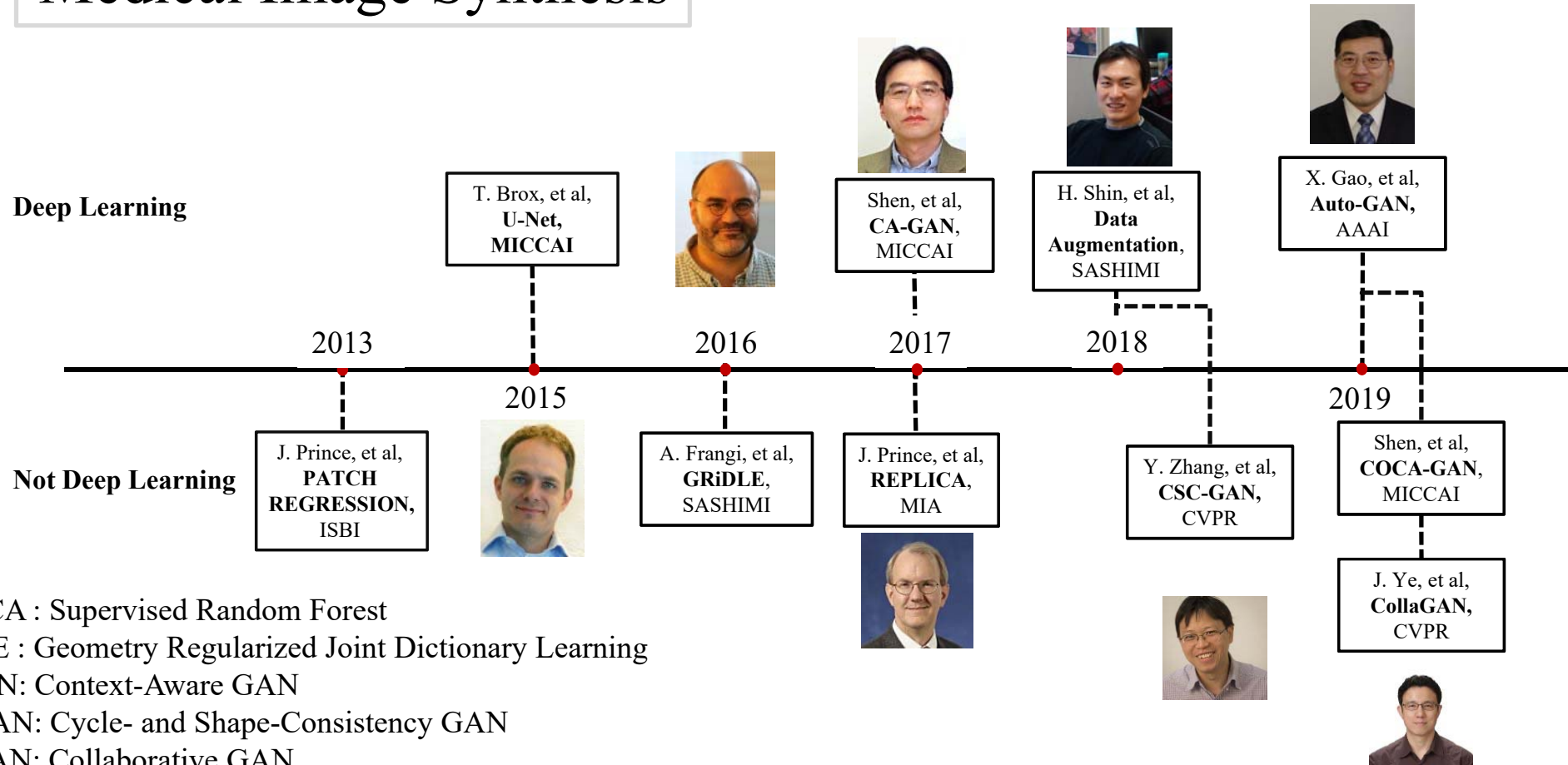
- ✓ 异质人脸图像合成
- ✓ 图像超分辨重建
- ✓ 其他相关应用

三. 图像身份识别

- ✓ 异质人脸图像识别
- ✓ 跨模态行人重识别
- ✓ 行为定位与识别

03. Related Work

Medical Image Synthesis



REPLICa : Supervised Random Forest

GRiDLE : Geometry Regularized Joint Dictionary Learning

CA-GAN: Context-Aware GAN

CSC-GAN: Cycle- and Shape-Consistency GAN

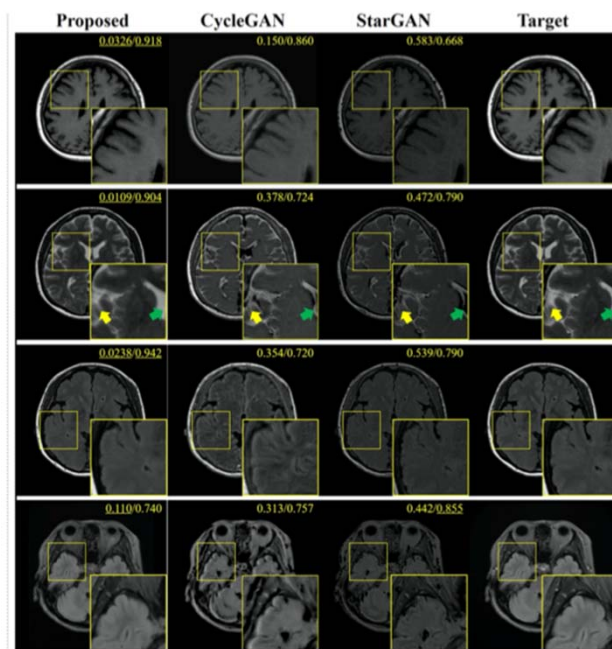
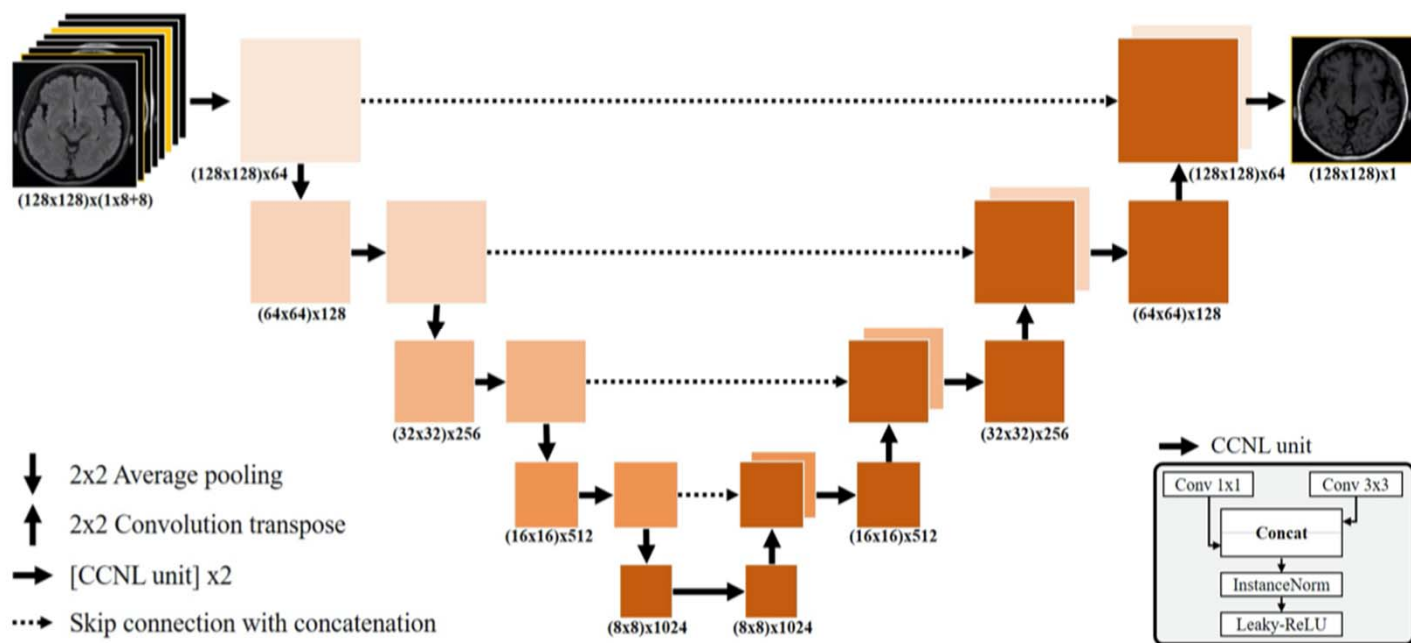
CollaGAN: Collaborative GAN

COCA-GAN: Common-Feature-Learning-Based Context-Aware GAN

Auto-GAN: Self-Supervised Collaborative Learning

Related Work: CollaGAN

基于协同生成对抗网络的医学图像数据补齐

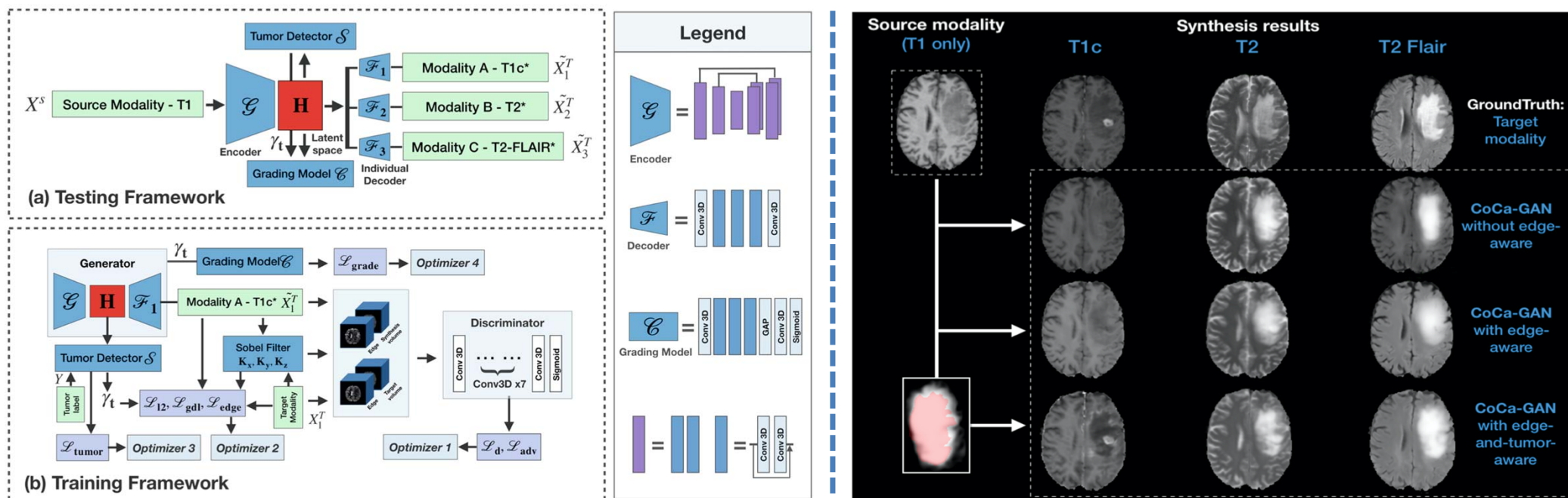


- 通过级联多种模态的医学图像作为输入，从尽量多的不同医学图像模态中，学习输入模态与目标模态的关系，利用多种医学图像生成缺失的模态。

Lee, D., Kim, J., Moon, W. J., & Ye, J. C. "CollaGAN: Collaborative GAN for missing image data imputation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR19)*, pp. 2487-2496, June 16-20, 2019, Long Beach, CA, USA

Related Work: Coca-GAN

基于共同空间学习的上下文相关生成对抗网络



- 将输入模态映射到一个**隐空间**，并引入**病灶区域检测**对肿瘤实质的边缘进行监督，进而提高模型对**肿瘤细节**的合成效果
- 通过**多个解码器**，构建隐空间与多个目标模态的**映射关系**，将隐空间的模态表示解码到不同的目标模态。

Pu Huang, Dinggang Shen, et. al "CoCa-GAN: Common-Feature-Learning-Based Context-Aware Generative Adversarial Network for Glioma Grading," *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, October 7-12, 2019, Shenzhen, China.

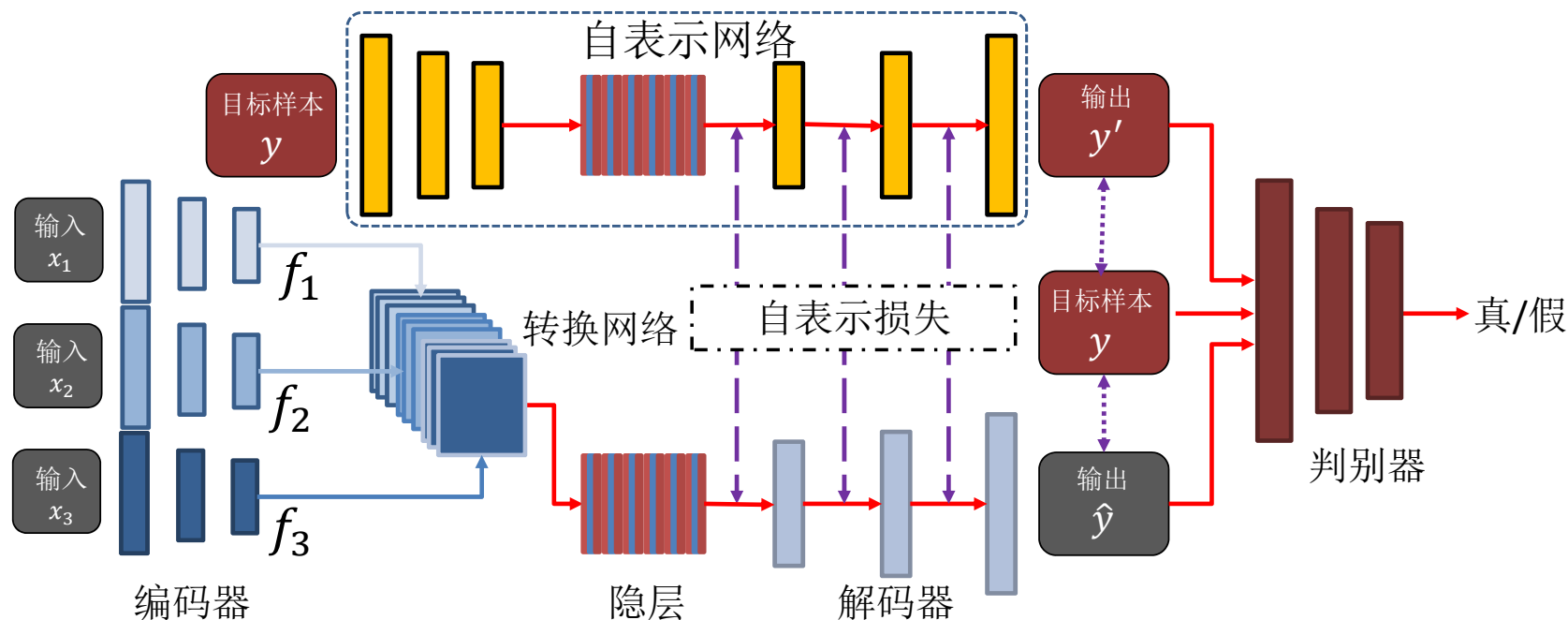
Motivation

医学图像现状：

- 不同模态的医学图像是对同一器官的不同表示，可以为诊断提供不同的信息；
—————> 如何有效利用不同模态的有效信息？
- 医学图像数据的获取极其昂贵，很难获取大量的样本；
—————> 如何在小样本数据下进行医学图像合成？
- 由于各地的医疗水平和病人的经济承受能力不同，不同病人的医学图像模态数量不一；
—————> 如何合成任一缺失模态？

Our Method: Auto-GAN

基于自监督协同学习的医学图像合成

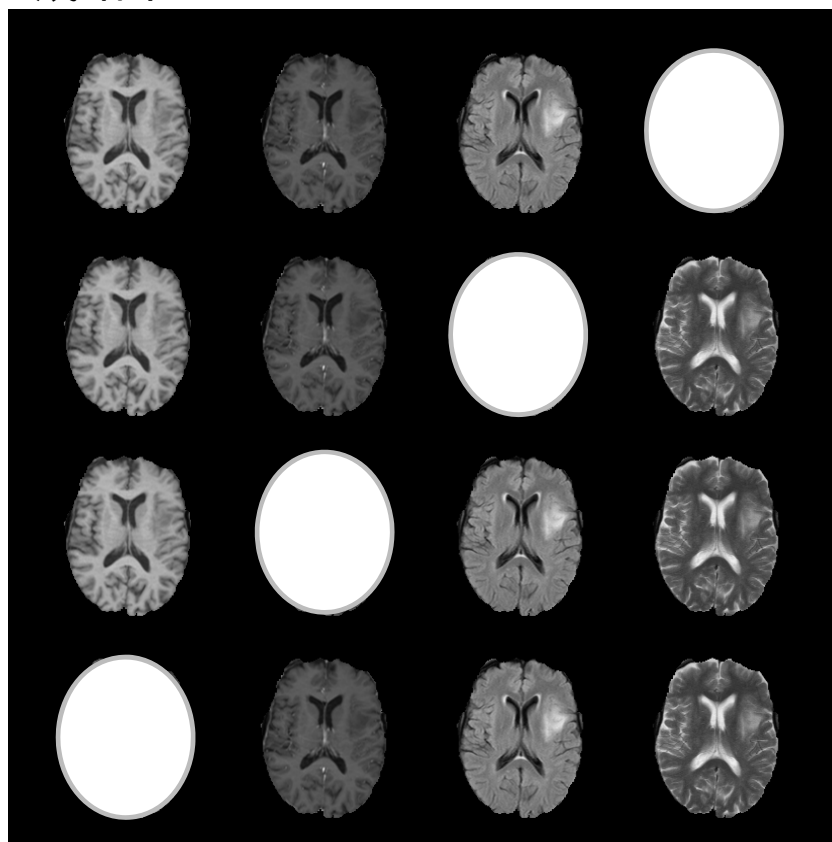


- 通过引入“多编码器-单解码器”的生成网络结构，构建多模态图像向单模态图像的映射，对输入的多模态图像的有效信息进行融合
- 将自编码器作为自表示网络，为解码器提供任意目标模态的优化方向，提高了模型对目标模态的重构能力。

Bing Cao, Han Zhang, Nannan Wang, Xinbo Gao, Dinggang Shen, “Auto-GAN: Self-Supervised Collaborative Learning for Medical Image Synthesis,” *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI-20)*, February 7-12, 2020, New York, New York, USA.

Our Method: Auto-GAN

生成结果

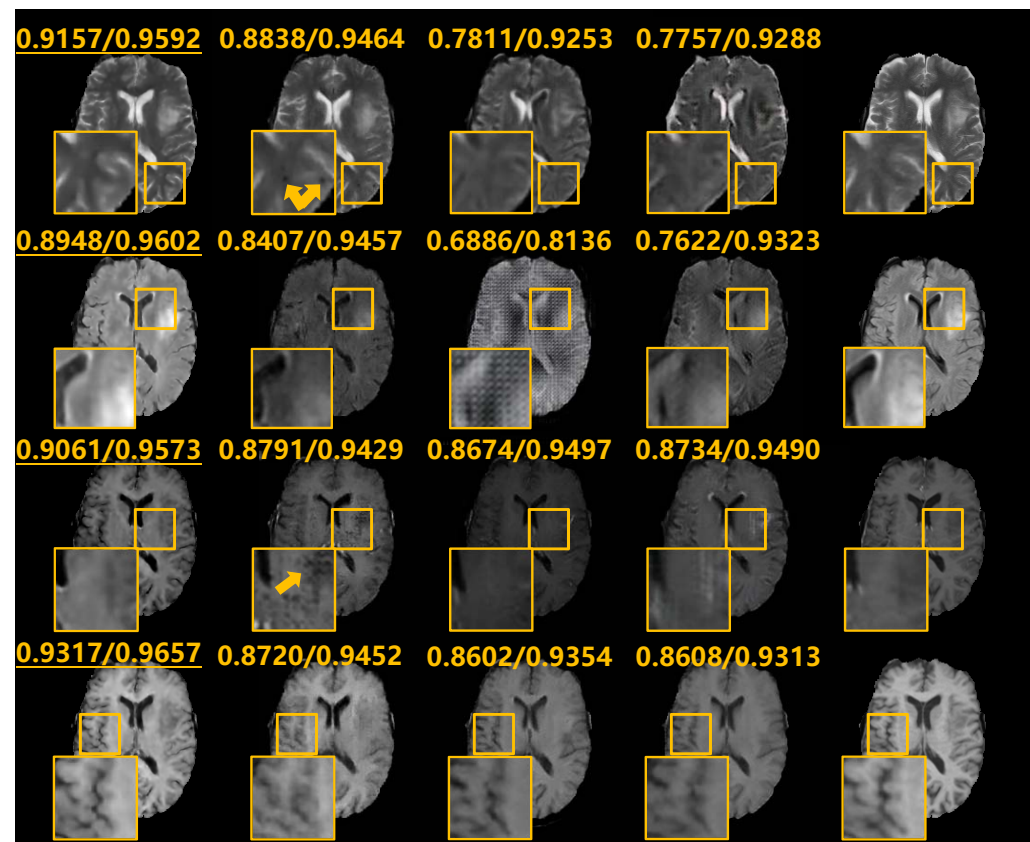


T1

T1-C

T2-FLAIR

T2



0.9157/0.9592 **0.8838/0.9464** **0.7811/0.9253** **0.7757/0.9288**

0.8948/0.9602 **0.8407/0.9457** **0.6886/0.8136** **0.7622/0.9323**

0.9061/0.9573 **0.8791/0.9429** **0.8674/0.9497** **0.8734/0.9490**

0.9317/0.9657 **0.8720/0.9452** **0.8602/0.9354** **0.8608/0.9313**

Proposed

Pix2Pix

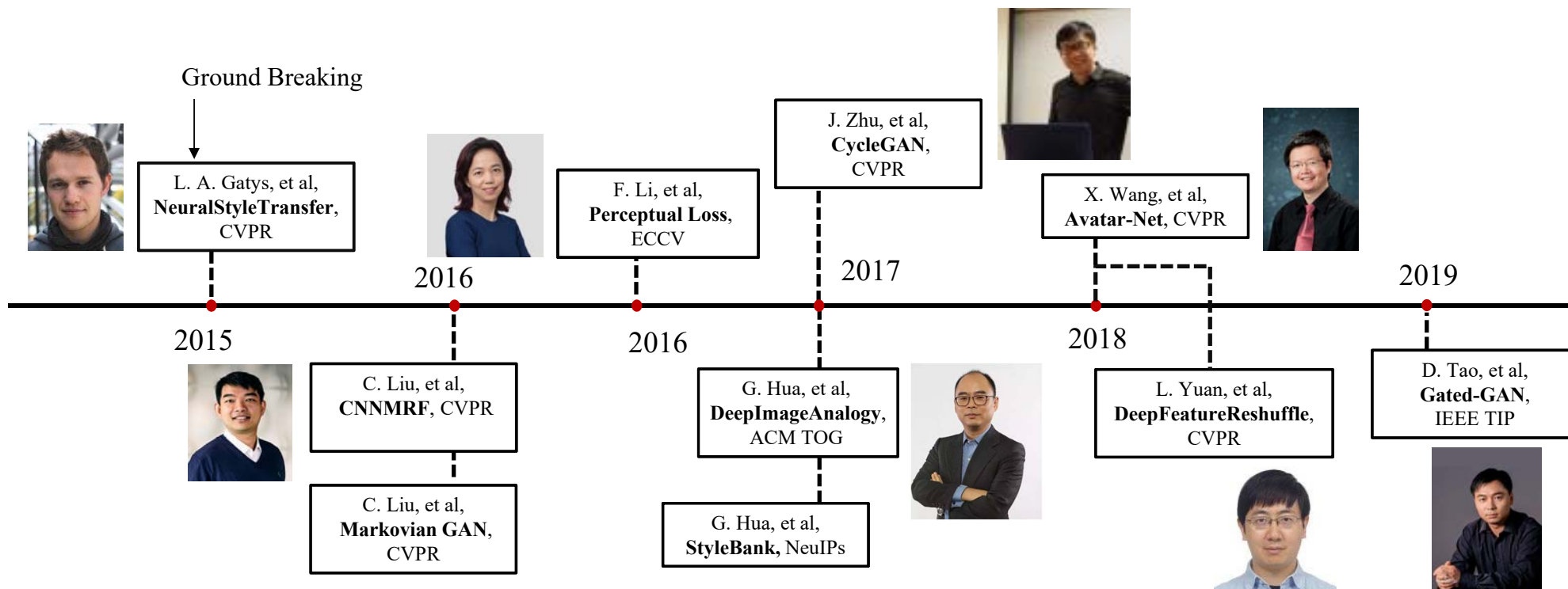
CycleGAN

StarGAN

Ground Truth

Bing Cao, Han Zhang, Nannan Wang, Xinbo Gao, Dinggang Shen, "Auto-GAN: Self-Supervised Collaborative Learning for Medical Image Synthesis," *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI-20)*, February 7-12, 2020, New York, New York, USA.

发展历史



基于**优化**的方法：NeuralStyleTransfer, CNNMRF, DeepImageAnalogy

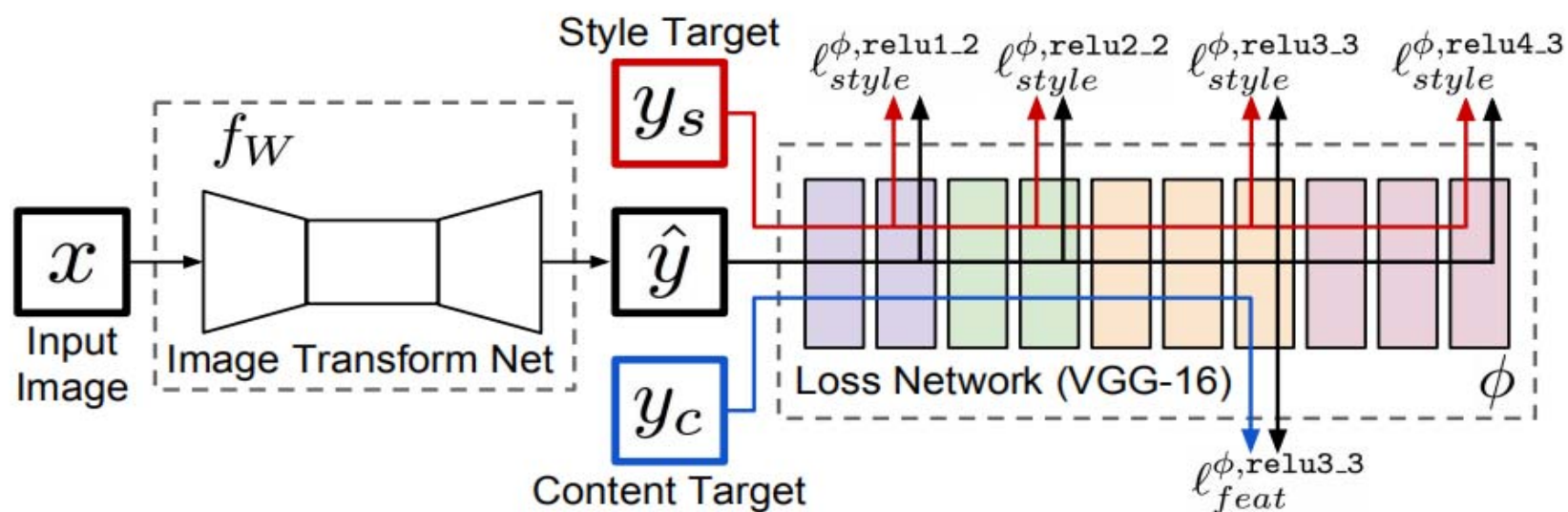
基于**前馈神经网络**的方法：Perceptual Loss, StyleBank,

基于**零样本学习**的方法：Avatar-Net, DeepFeatureReshuffle

基于**GAN**的方法：Markovian GAN, CycleGAN, Gated-GAN

相关工作

- 基本思想：构建**图像转换网络**，利用**感知损失**约束生成图像的内容及风格信息。



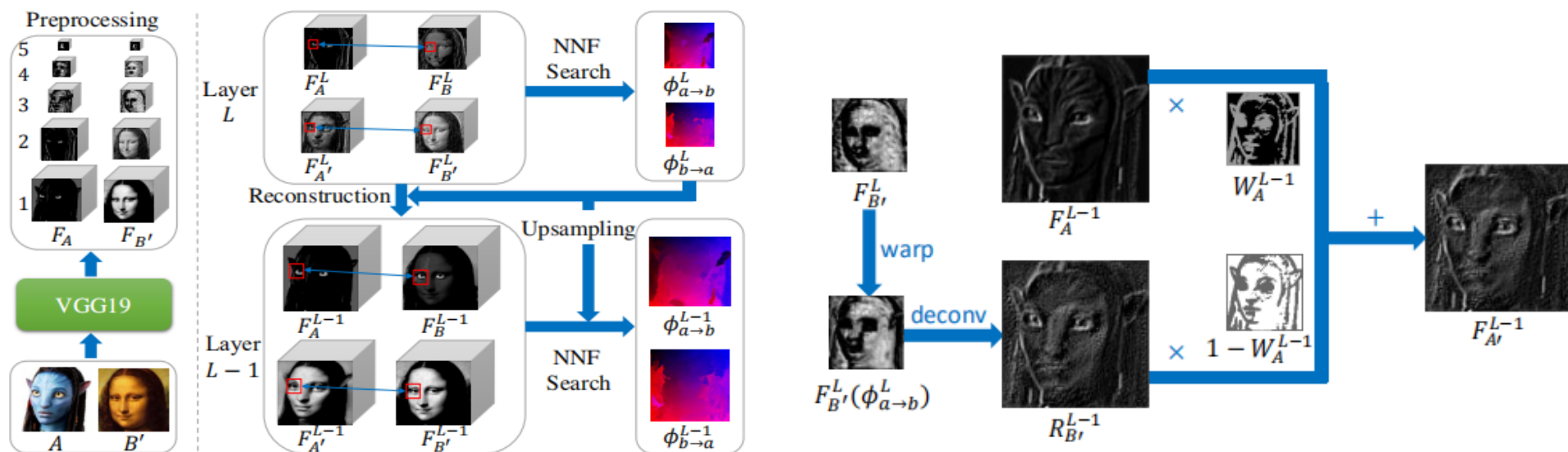
缺陷：

- 不同的艺术图像需要训练新的网络，单张艺术图像不能充分体现流派风格。
- 风格损失对整张图像的纹理信息进行约束，破坏了图像的语义结构。

J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in European Conference on Computer Vision. Springer, 2016, pp. 694–711.

相关工作

- 基本思想：通过匹配跨域图像深度特征，逐层重构特征，迁移跨域图像的视觉属性



缺陷：

- 逐层进行特征匹配及重构计算量庞大，生成结果时间过长。
- 仅适用于语义及结构高度一致的跨域图像。采用欧几里得距离匹配图像深度特征，使跨域图像之间的相似性对跨域特征匹配结果有较大影响。

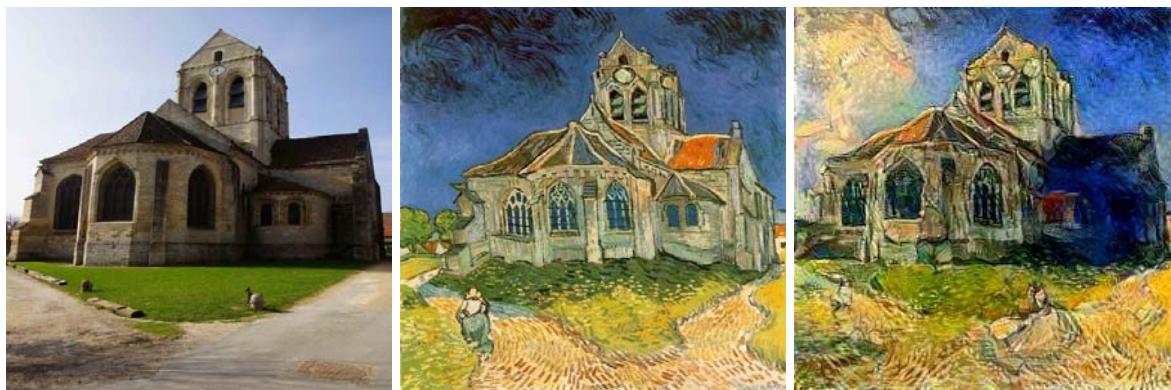
Jing Liao, Yuan Yao, Lu Yuan, Gang Hua, and Sing Bing Kang, "Visual attribute transfer through deep image analogy," ACM Transactions on Graphics, TOG, vol. 36, no. 4, 2017.

Motivation

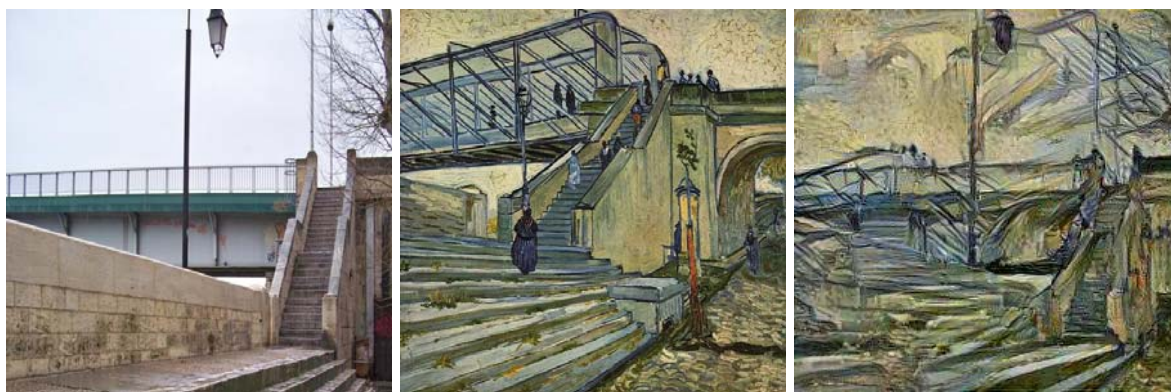
“艺术风格是一个人或一个群体的艺术作品中不变的元素或表达方式。”

——美国艺术史学家Meyer Schapiro

奥维尔小镇教堂



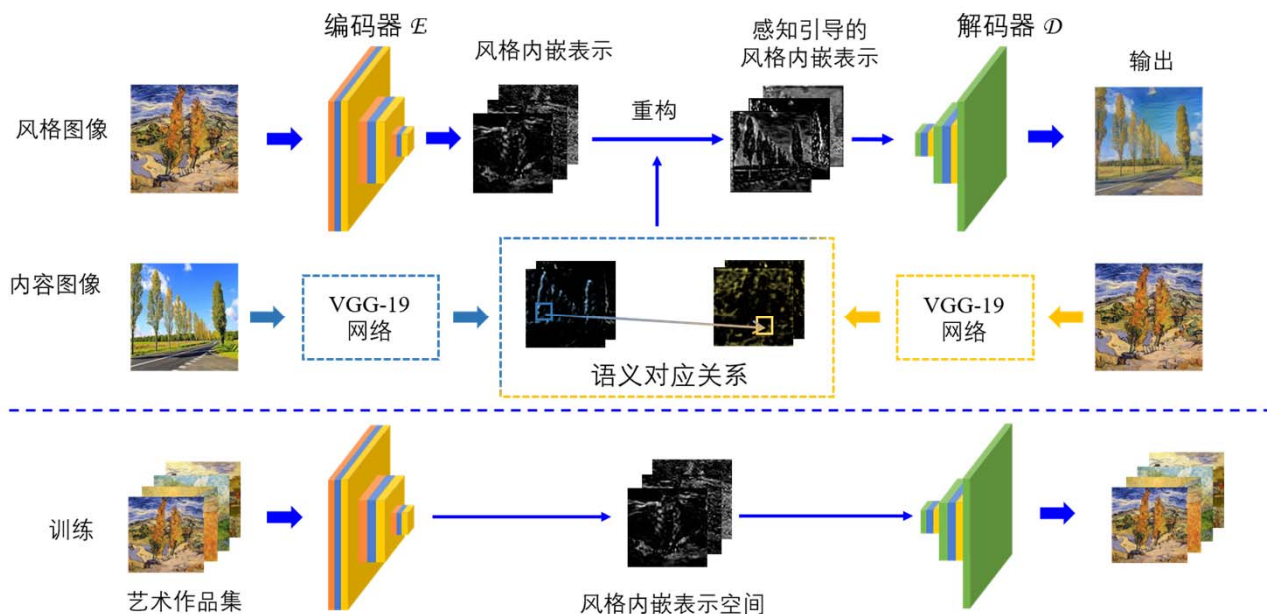
阿尔勒天桥



- **色彩模式和纹理**迁移破坏图像语义一致性
 - 艺术家在艺术作品中表达他们对真实世界的感受。
 - 模仿艺术作品的创作机制，采用**感知引导**，增强跨域图像的**语义一致性**。
 - **流派整体风格**而非单张图像的色彩和纹理模式。
 - **跨域特征匹配**构建语义对应关系。

方法过程

基本思想：构建**编码-特征重构-解码**模型，学习**流派风格特征**，保持**跨域图像语义一致性**



➤ 流派风格编码器损失

$$L_{identity} = \|x_s - D(E(x_s))\|_1$$

➤ 构建语义对应关系

跨域PatchMatch

$$f_{c \rightarrow s}^L(p) = \arg \min_q \sum_{x \in M(p), y \in M(q)} -corr(D_c^L(x), D_s^L(y)),$$

NNF块搜索

$$NN(i) = \arg \min_j -corr(D_c^L(x_i), D_s^L(x_j))$$

- 训练艺术流派编码解码网络，构建编码表示空间，学习**流派风格表示**。
- 设计**跨域图像特征相似性衡量标准**，利用图像的**深度特征**作为感知引导，在编码表示空间**重构具有结构一致性的风格特征表示**。

实验结果

输入



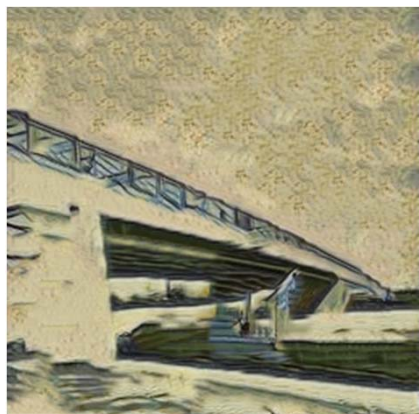
NeuralStyleTransfer



DeepImageAnalogy



我们的方法 (ERPG)

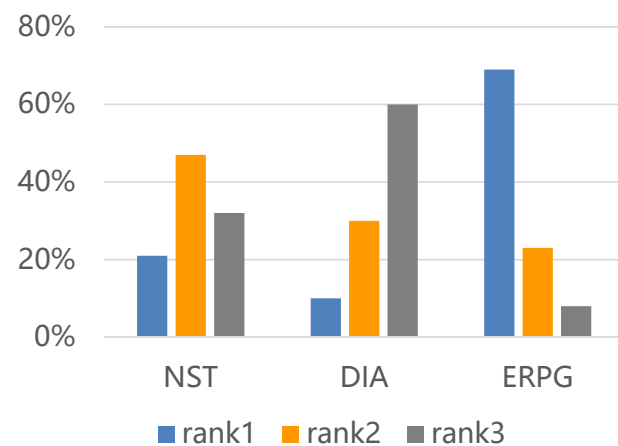


- 感知引导保持了跨域图像的语义结构。
- 迁移整体风格特征，增强“手绘感”。

客观评价

方法	SSIM	HOSA
NeuralStyleTransfer	0.3524	20.3579
DeepImageAnalogy	0.4737	24.4083
ERPG	0.6433	37.1470

主观评价



汇报提纲

一. 研究背景与意义

二. 图像跨域重建

- ✓ 异质人脸图像合成
- ✓ 图像超分辨率重建
- ✓ 其他相关应用

三. 图像身份识别

- ✓ 异质人脸图像识别
- ✓ 跨模态行人重识别
- ✓ 行为定位与识别

汇报提纲

一. 研究背景与意义

二. 图像跨域重建

- ✓ 异质人脸图像合成
- ✓ 图像超分辨重建
- ✓ 其他相关应用

三. 图像身份识别

- ✓ 异质人脸图像识别
- ✓ 跨模态行人重识别
- ✓ 行为定位与识别

异质人脸图像识别相关工作

■ 基于共同空间映射的异质人脸识别工作发展时间线



IIT-Delhi
Forensic &
Composite



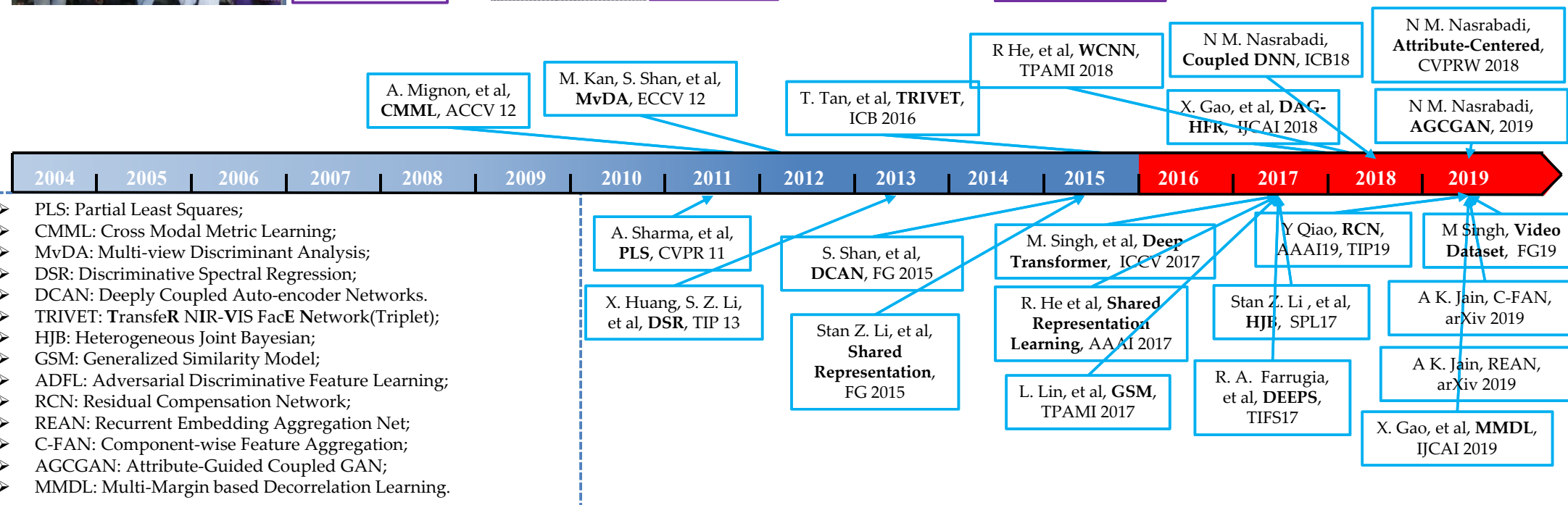
Ran. He
NIR



N. M. Nasrabadi
Composite & TIR



A. K. Jain
Video face



异质人脸图像识别相关工作

■ 基于特征表示的异质人脸识别工作发展时间线



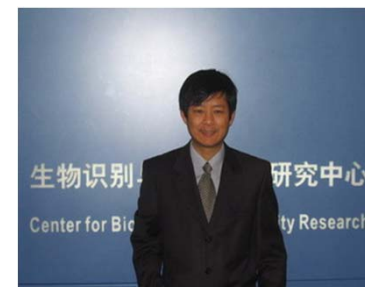
A. K. Jain
Forensic & Composite



IIT-Delhi
Forensic & Composite



Jiwen Lu
NIR & Sketch



S Z. Li
ID Face

B. Klare and A. K. Jain, **A feature-based approach**, SPIE 10

H. Bhatt, et al, **MCWLD**, TIFS 12

H. Galoogahi and T. Sim, **LRBP**, ICIP 12;
Gabor Shape, ACM MM 12;
HAoG, ICME 12;

S. Ouyang, et al, **Facial attributes model**, ACCV 14

S. Saxen, et al, **CNNs**, ECCVW 2016

J. Lu, et al, **SLBFLE**, TPAMI 2018

J. Lu, et al, **CA-LBFL**, TPAMI 2018



A. K. Jain, et al, **LFDA**, TPAMI 11

W. Zhang, X. Wang, and X. Tang, **CITE**, CVPR 11

A. K. Jain, **Kernel Prototype Similarity**, TPAMI 13

P. Mittal, et al, **Transfer learning**, ICB 2015

X. Gao, et al, **Re-ranking**, TIP 2019

X. Gao, et al, **DLFace**, PR19

S Z. Li, et al, **LBL**, IJCV 2019

- CITE: Coupled Information-Theoretic Encoding;
- LFDA: Local Feature-based Discriminant Analysis;
- LRBP: Local Radon Binary Pattern;
- HAoG: Histogram of Averaged Oriented Gradients;
- MCWLD: Multiscale Circular Weber's Local Descriptor;
- CA-LBFL: Context-Aware Local Binary Feature Learning;
- SLBFLE: Simultaneous Local Binary Feature Learning and Encoding.
- LBL: Large-scale bisample learning

异质人脸图像识别相关工作

■ 基于人脸合成的异质人脸识别工作发展时间线



X. Tang
Viewed Sketch



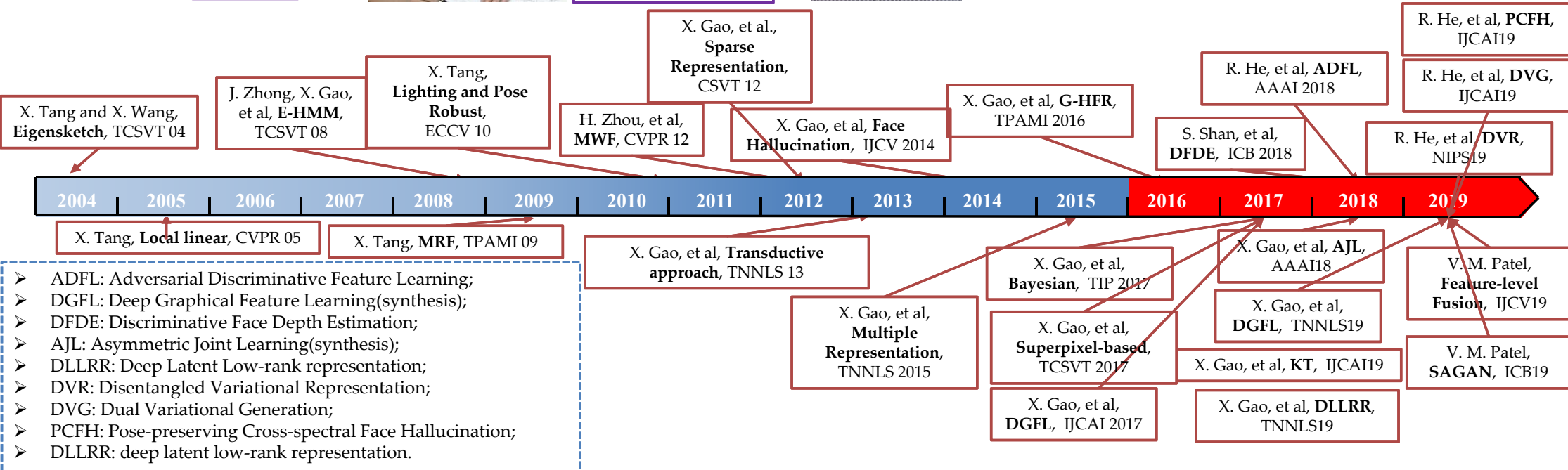
X. Gao
Forensic& Composite Sketch& NIR



Ran. He
NIR

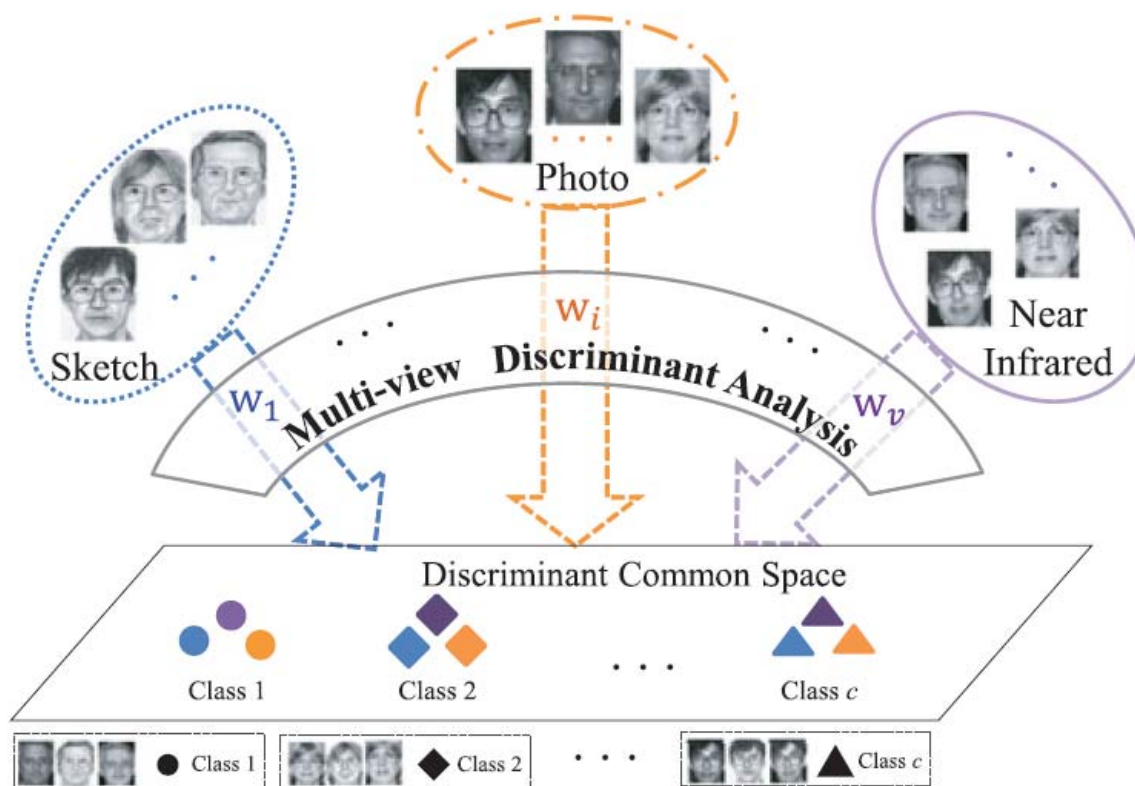


Vishal M. Patel
TIR



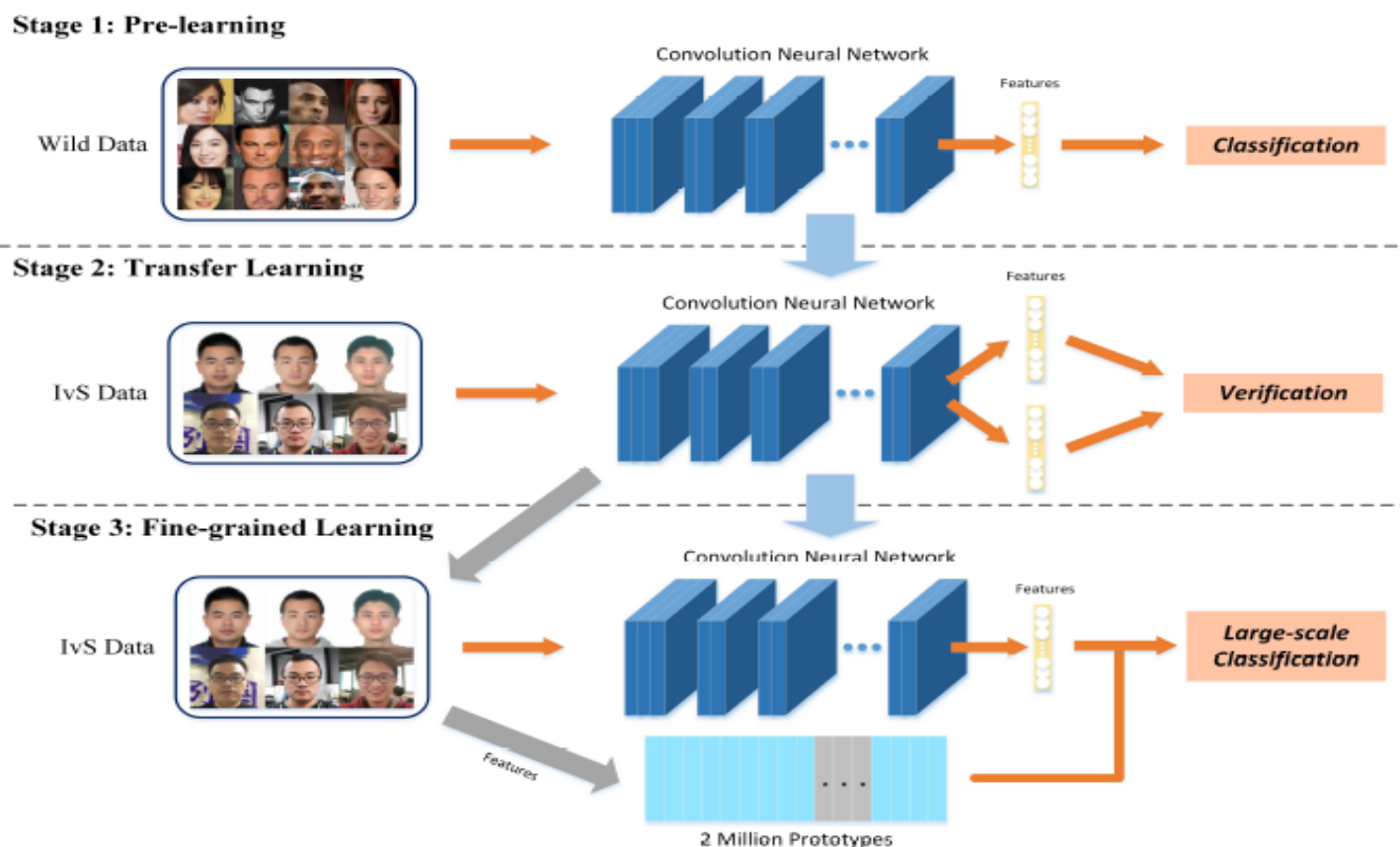
- ADFL: Adversarial Discriminative Feature Learning;
- DGFL: Deep Graphical Feature Learning(synthesis);
- DFDE: Discriminative Face Depth Estimation;
- AJL: Asymmetric Joint Learning(synthesis);
- DLLRR: Deep Latent Low-rank representation;
- DVR: Disentangled Variational Representation;
- DVG: Dual Variational Generation;
- PCFH: Pose-preserving Cross-spectral Face Hallucination;
- DLLRR: deep latent low-rank representation.

基于多视角判别分析的异质人脸识别



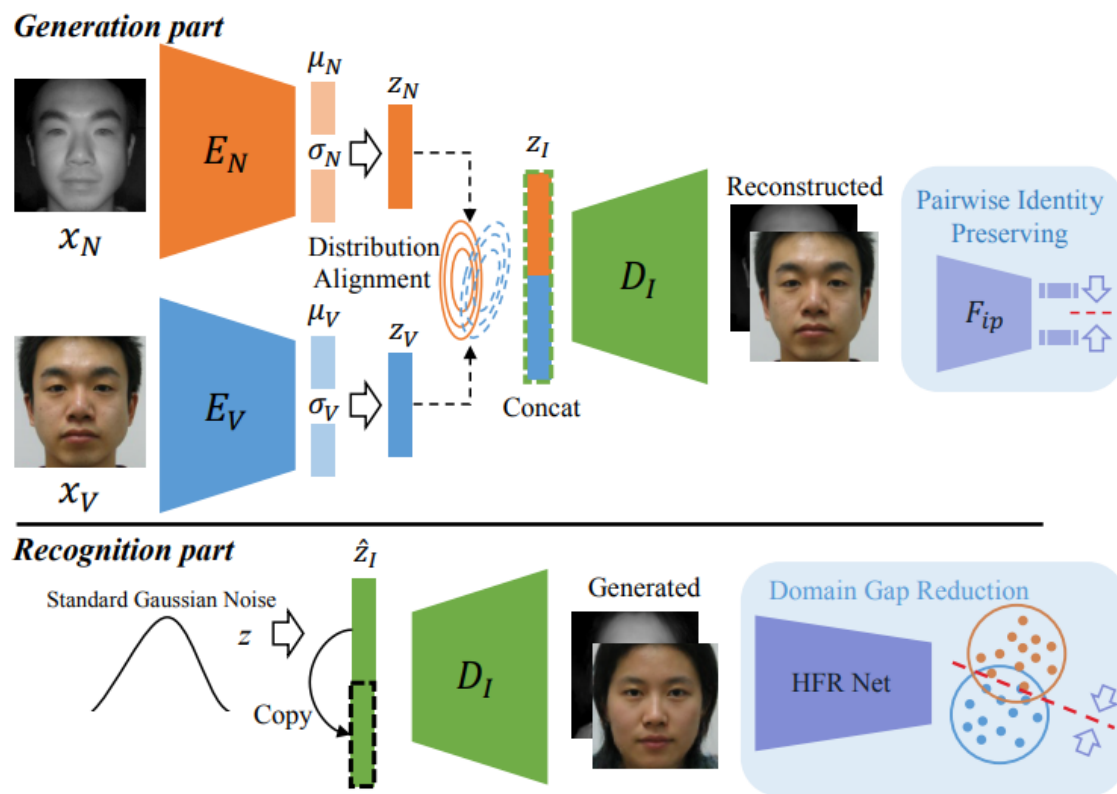
- 作者通过联合学习多个线性映射将不同模态人脸映射到一个共同可分辨空间下，以减少异质图像之间的模态差异。

基于大规模双样本学习的异质人脸识别



- 作者提出基于深度学习的大规模双样本学习算法，解决训练数据有限的问题；并提出一种大规模身份证人脸数据库，约两百万张。

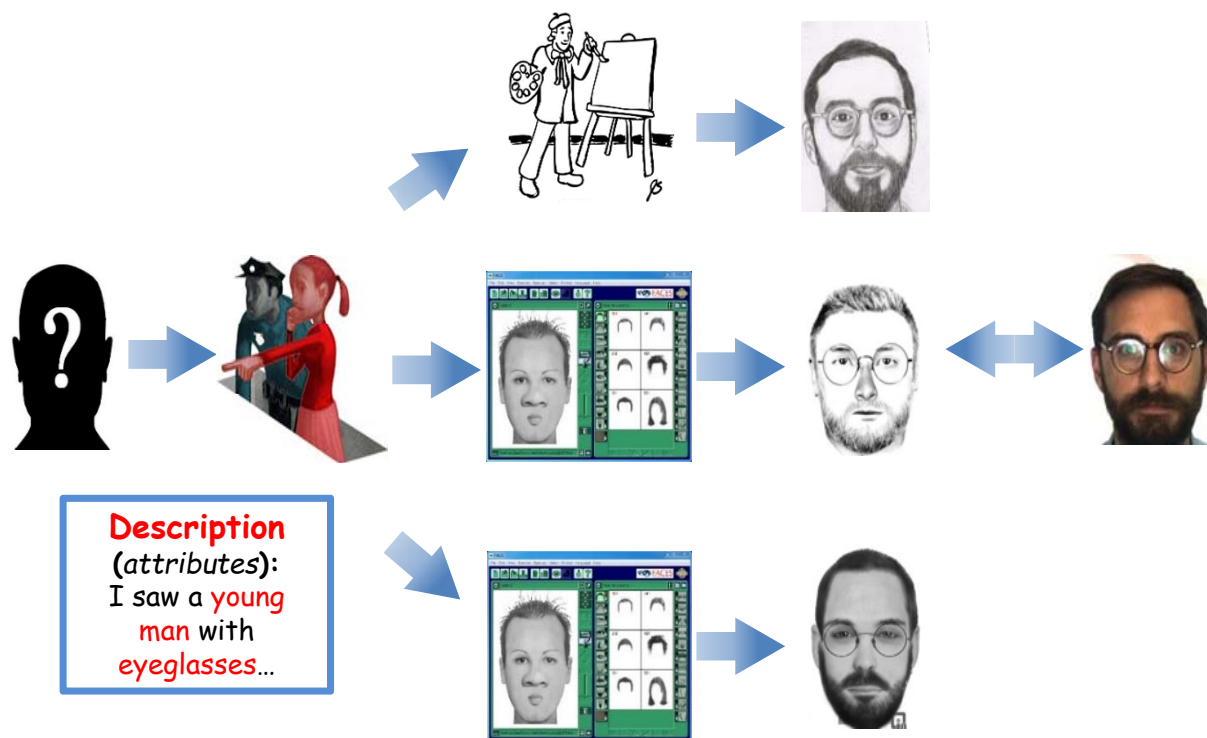
基于对偶变分生成学习的异质人脸识别



- 作者提出一种基于对偶变分学习的人脸数据生成方法，通过数据增广的方式来缓解异质人脸图像数据规模较少的问题。

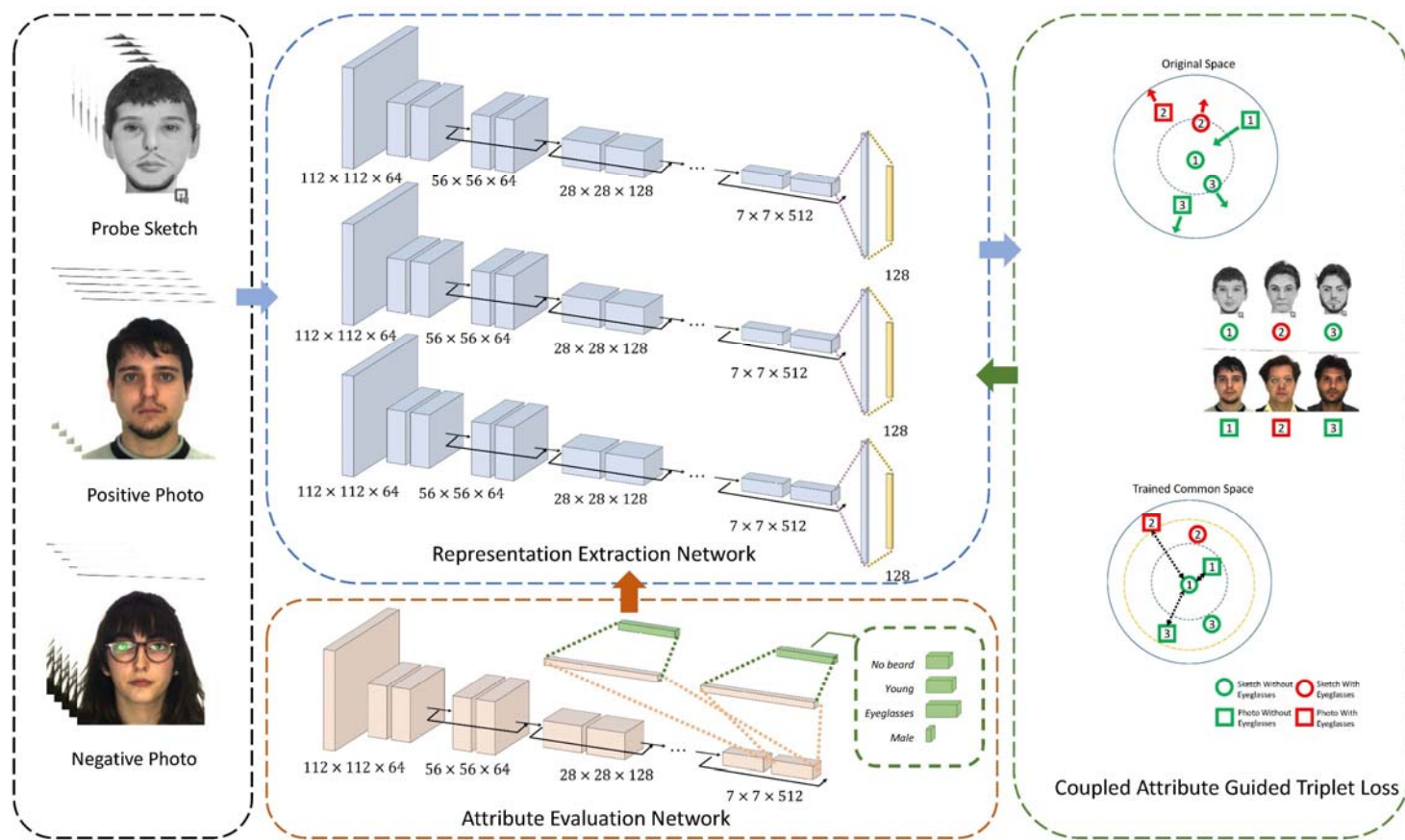
基于双属性表征学习的异质人脸识别

- 基本思想：从异质图像特殊的生成机制分析出发，我们发现人脸语义属性信息在不同模态下均具有较好鲁棒性，如何设计模型来有效融合语义属性信息和图像信息是拟解决的关键问题。



基于双属性表征学习的异质人脸识别

- 主要贡献：1) 提出一种端对端的基于双属性表征学习方法，可以提取语义属性指导的可判别性嵌入式表征；2) 提出一种新的双属性指导的损失函数，可以有效减少由属性检测网络误差带来的不利影响。



基于双属性表征学习的异质人脸识别

- 实验结果：在多种不同场景的异质人脸数据库上超越了当前领先的识别算法，证明了该算法良好的性能。另外，我们也增大测试数据集的数量来验证算法性能，使其更符合显示真实场景。

TABLE I
RECOGNITION ACCURACIES OF THE STATE-OF-THE-ART METHODS AND
THE PROPOSED CAL-HFR METHOD ON THE UoM-SGFS SETA
DATABASE.

Algorithm	Rank-1	Rank-10	Rank-50
VGGFace [51]	9.33	31.07	59.73
PCA[52]	2.80	8.40	17.73
CBR [53]	5.73	18.80	43.33
P-RS [29]	22.13	49.33	69.87
DEEPS [20]	31.60	66.13	86.00
DAG-HFR[18]	80.93	92.40	96.13
CAL-HFR	82.40	93.87	98.40

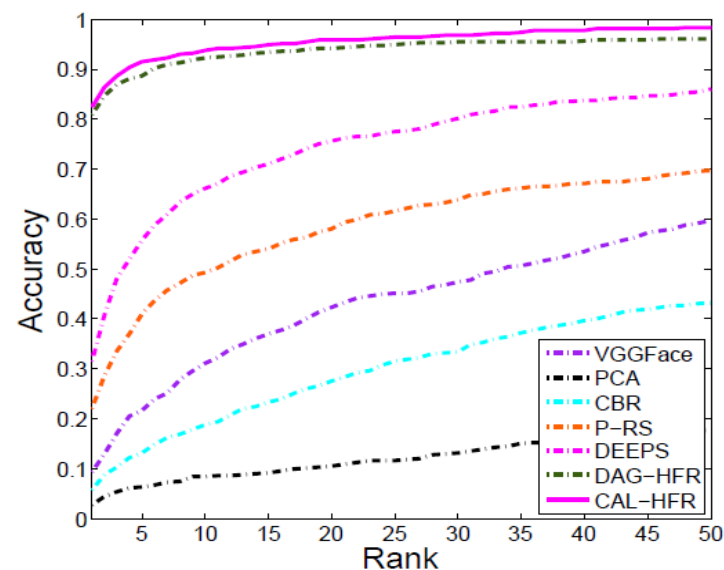
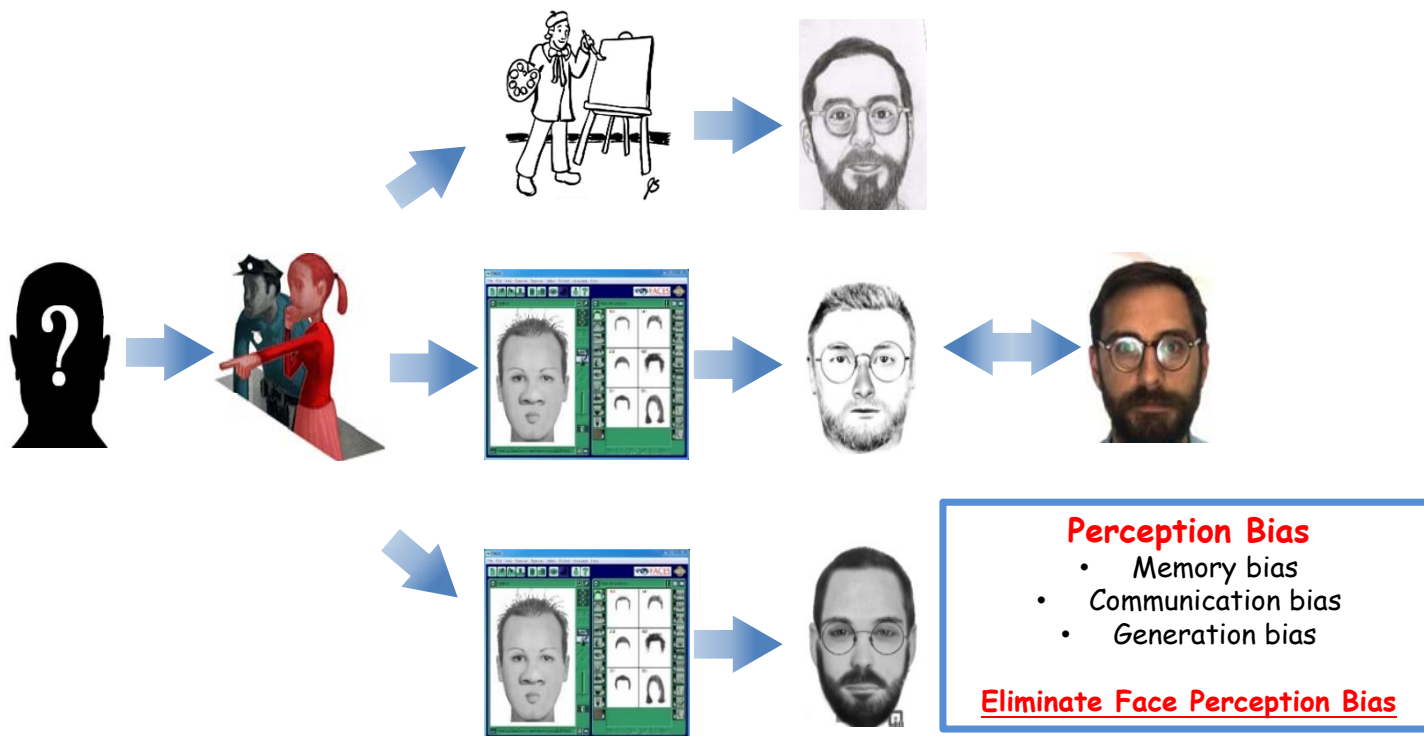


Fig. 8. Cumulative match score comparison of the VGGFace methods, the PCA method, the CBR method, the P-RS method, the DEEPS method, the DAG-HFR method and our proposed CAL-HFR method on the UoM-SGFS SetA database.

Liu D, Gao X, Wang N, et al. Coupled Attribute Learning for Heterogeneous Face Recognition. IEEE Transactions on Neural Networks and Learning Systems, 2020.

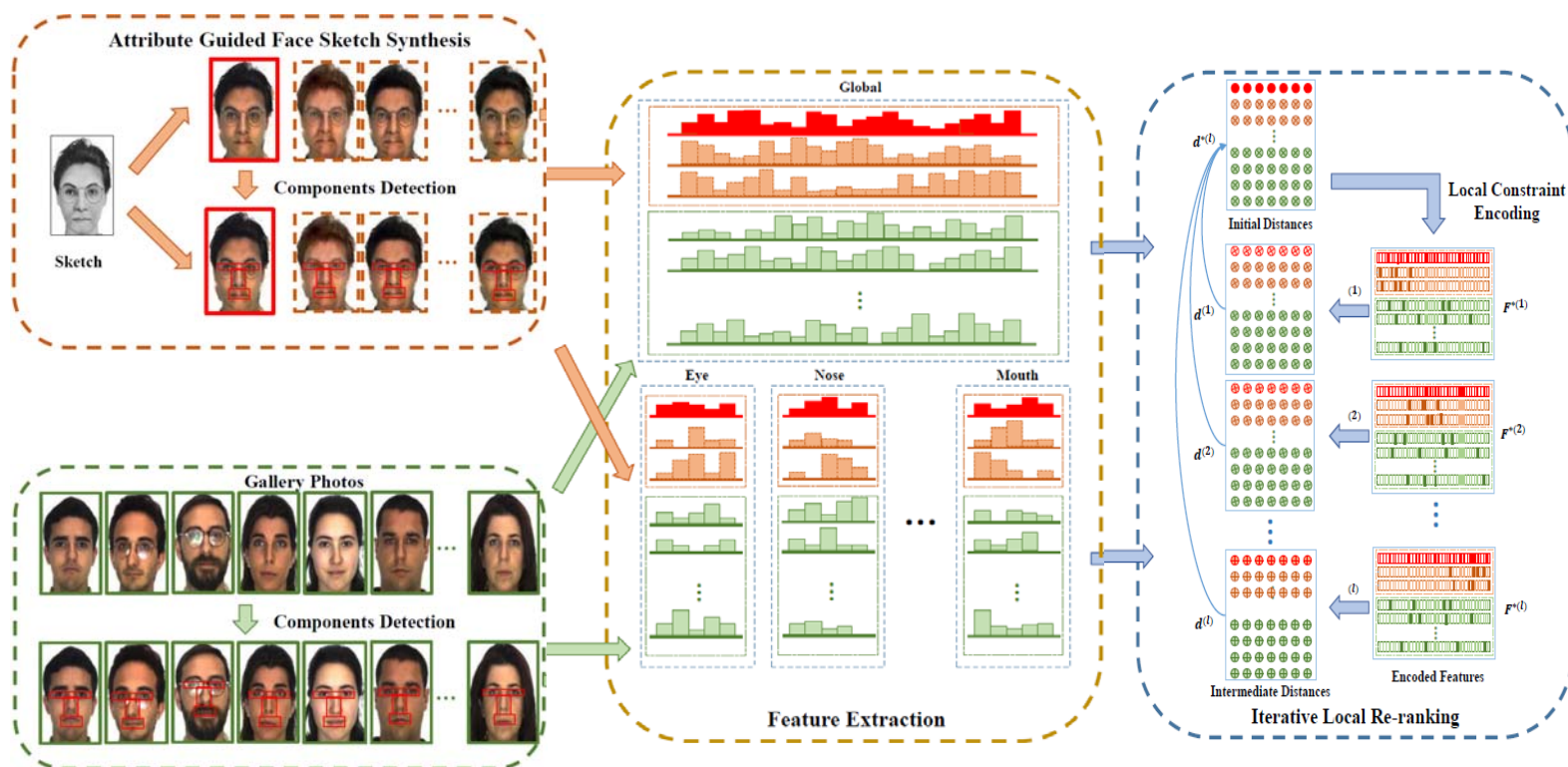
基于属性合成和局部重检索的异质人脸识别

- 基本思想：异质人脸图像复杂的制作过程会不可避免地引入认知偏差，并且由于制作成本高导致训练图像数量较少。如何利用多语义属性指导合成人脸图像，并设计重检索算法提高识别效果是拟解决的难点。



基于属性合成和局部重检索的异质人脸识别

- 主要贡献：1) 提出基于语义指导的跨模态人脸合成算法来减少感知偏差带来的干扰；2) 提取具有模态不变性的局部特征来完成重检索任务。



基于属性合成和局部重检索的异质人脸识别

- 实验结果：在几种主流的异质人脸数据库上超越了当前领先的相关算法，证明了该算法良好的识别性能和泛化能力。消融实验进一步证明了该算法的有效性。

Algorithms	Accuracy (E-PRIP)	Accuracy (PRIP-VSGC)
Fisherface ([46])	35.30%	21.87%
MCWLD ([47])	24.00 %	15.40 %
SSD-based ([37])	53.30%	45.30%
Transfer Learning ([3])	60.20%	52.00%
CNNs ([43])	65.60%	51.50%
DEEPS ([39])	80.80%	54.90%
Baseline	79.73%	64.53%
SCA ([32])	78.13%	66.40%
k-RF([33])	80.93%	68.27%
ILR (Proposed)	83.47%	68.67%

Algorithms	Rank-1 Accuracy	Rank-10 Accuracy
PCA ([48])	2.80%	8.40%
CBR ([30])	5.73%	18.80%
VGGFace ([49])	9.33%	31.07%
P-RS ([16])	22.13%	49.33%
DEEPS ([39])	31.60%	66.13%
Baseline	80.27%	97.07%
SCA ([32])	82.67%	94.27%
k-RF([33])	83.73%	97.20%
ILR (Proposed)	84.53%	98.13%

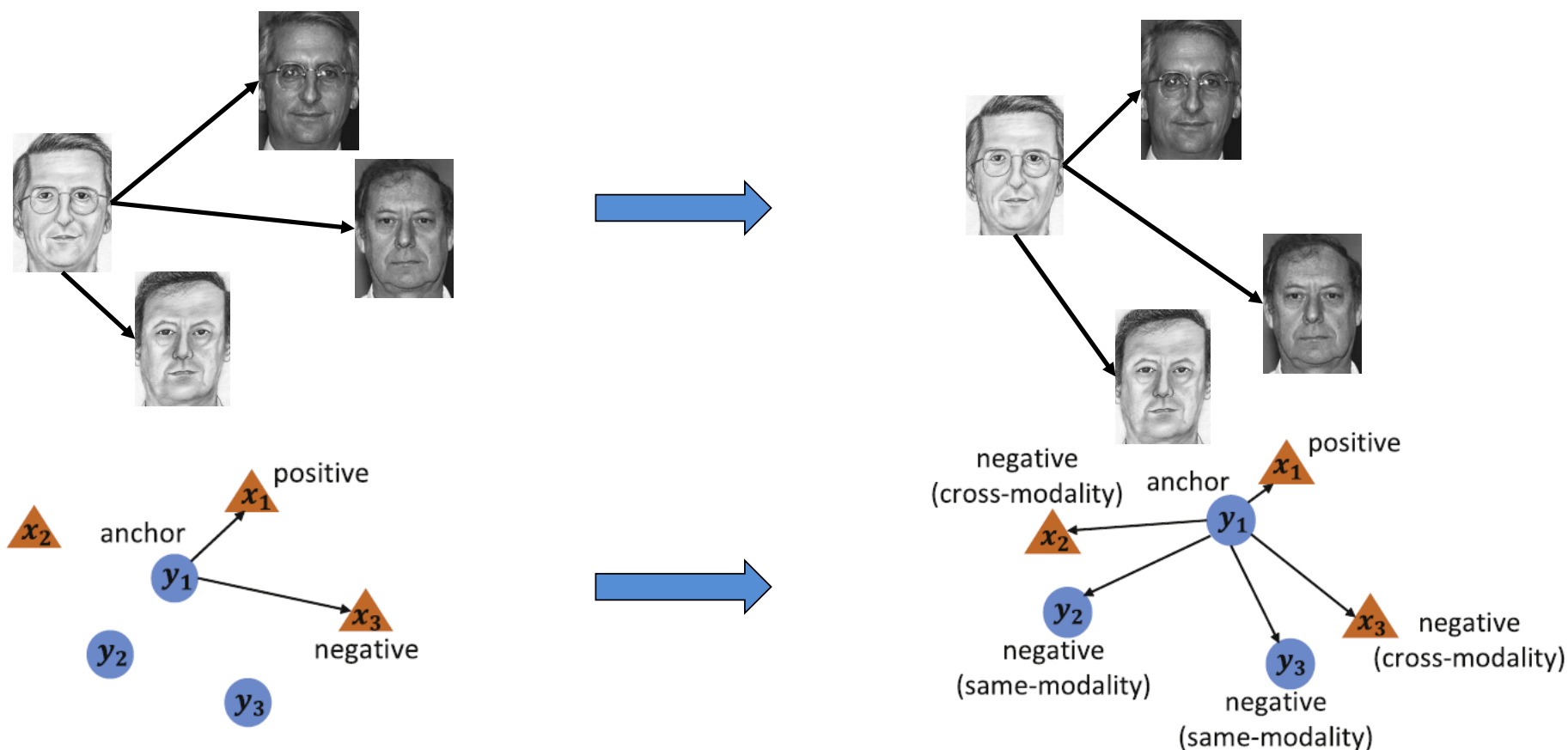
Table 1: Rank-10 recognition accuracies of the state-of-the-art approaches and our method on the E-PRIP and PRIP-VSGC databases.

Table 2: Recognition accuracies of the state-of-the-art methods and the proposed iterative local re-ranking method on the UoM-SGFS SetA Database.

Liu D, Gao X, Wang N, et al. Iterative local re-ranking with attribute guided synthesis for face sketch recognition. Pattern Recognition, 2020.

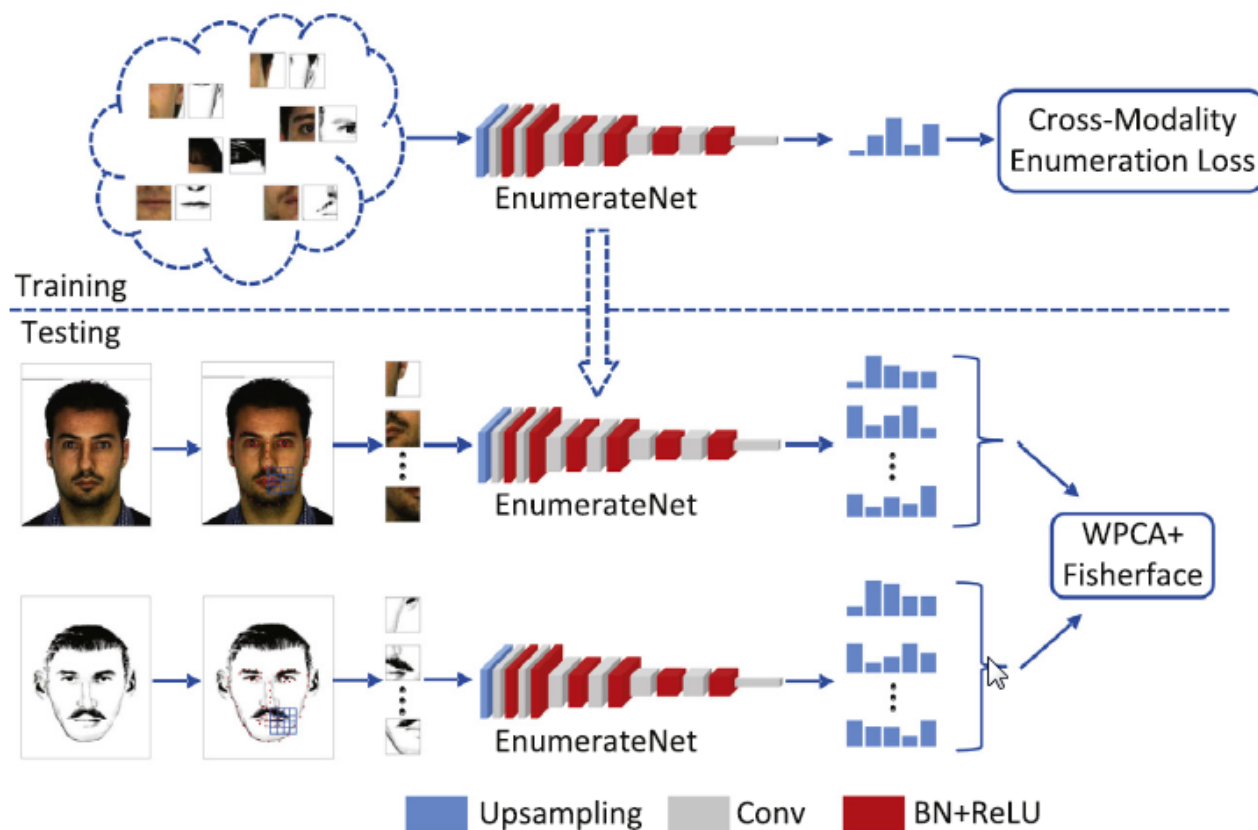
基于深度局部描述子的异质人脸识别

- 基本思想：该方法拟结合深度卷积网络与局部区域表征，来减少异质人脸图像的类内差异，增加其类间差异以提高识别效果。



基于深度局部描述子的异质人脸识别

- 主要贡献：1) 提出一种深度局部描述子学习框架，可以从输入图像块中学习局部可判别信息；2) 利用枚举损失函数来减少局部图像块的模态差异，以提高识别效果。



基于深度局部描述子的异质人脸识别

- 实验结果：在跨模态人脸数据库上达到了领先水平，证明了该算法良好的识别性能和泛化能力。

Comparison with state-of-the-art methods on CASIA NIR-VIS 2.0 dataset (rank-1, %).

Method	Accuracy	Method	Accuracy
CSR [22]	33.80	D-RS [31]	28.20
KDSR [72]	37.50	PCA+HCA [7]	23.70
LCFS [73]	35.40	H2(LBP3) [74]	43.80
C-DFD [75]	65.80	CDFL [75]	71.50
Gabor+RBM [38]	86.20	Recon.+UDP [76]	78.50
CEFD [77]	85.60	TRIVET [41]	95.70
VGG-Face [69]	62.09	CNN [65]	85.90
IDNet [78]	87.10	SeetaFace [79]	68.03
CenterLoss [80]	87.69	Lightened CNN [81]	91.88
IDR-128 [51]	97.33	DLFace	98.68

Comparison with state-of-the-art methods on NJU-ID dataset (rank-1, %).

Method	Accuracy	Method	Accuracy
PCA [60]	23.00	CSR [22]	29.30
CCA [70]	20.30	CDFE [20]	24.60
MvDA [24]	16.50	CMML [25]	20.30
HMLCR [71]	27.70	ESPAC [6]	20.80
CML [27]	30.90	DLFace	43.46

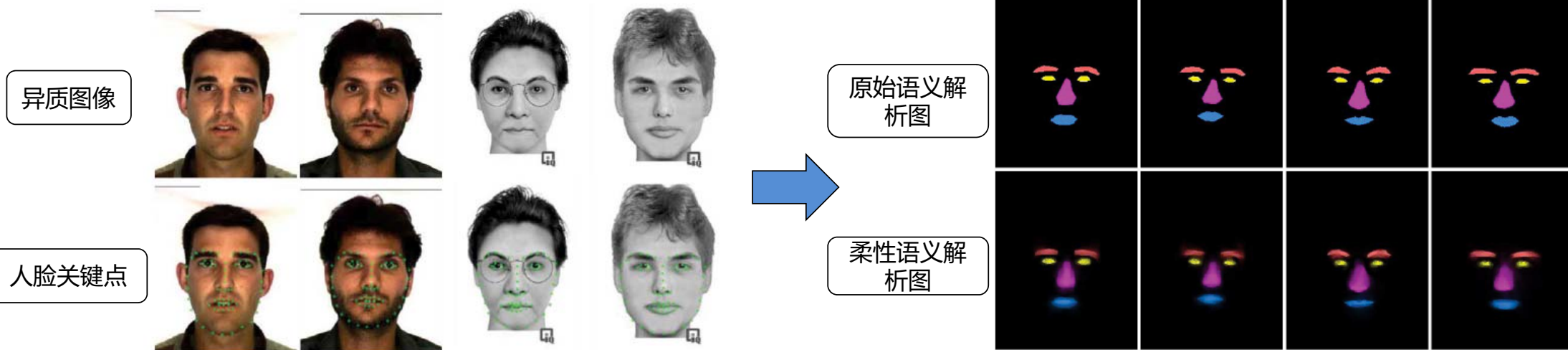
Comparison with state-of-the-art methods on forensic sketch dataset (%).

Method	Rank-10	Rank-50
PCA [60]	6.07	12.68
Fisherface [60]	10.71	17.14
D-RS [31]	11.21	20.80
G-HFR [5]	18.21	31.96
DLFace	40.73	57.64

Peng C, Wang N, Li J, et al. DLFace: Deep local descriptor for cross-modality face recognition. Pattern Recognition, 2019.

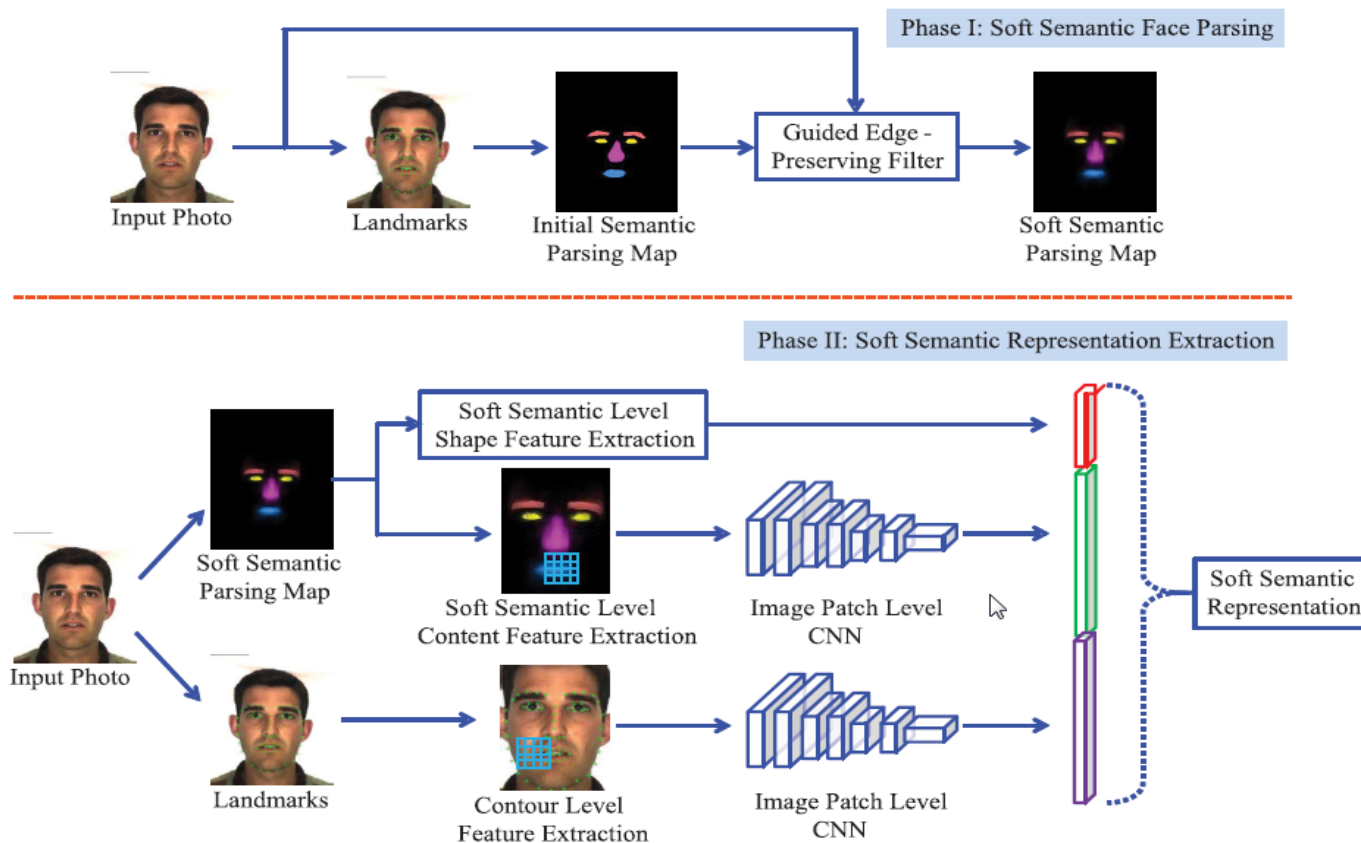
基于柔性语义表征的异质人脸识别

- 基本思想：人脸语义信息对于减少异质图像模态差异有着重要作用。如何设计更加准确的语义解析图来提高识别效果是拟解决的问题。



基于柔性语义表征的异质人脸识别

- 主要贡献：1) 提出一种柔性语义解析算法来提取更加准确的异质人脸语义信息；2) 利用分层级语义表征框架并融合深度网络来提高识别效果。



基于柔性语义表征的异质人脸识别

- 实验结果：我们在目前主流的跨模态异质人脸数据库上验证算法的性能，结果证明该算法超过目前最先进的识别算法。

TABLE I

RECOGNITION PERFORMANCE (RANK-10, %) BETWEEN PHOTO DOMAIN AND COMPOSITE SKETCH DOMAIN ON THE E-PRIP DATABASE AND PRIP-VSGC DATABASE

Method	Accuracy on e-PRIP	Accuracy on PRIP-VSGC
Fisherface [61]	35.30	21.90
MCWLD [35]	24.00	15.40
SSD [73]	53.30	45.30
Composite-Components [74]	70.10	61.60
TLDNN [41]	60.20	52.00
CNN [75]	65.60	51.50
DEEPS [1]	80.80	54.90
DLFace [59]	82.80	74.60
AADCNN [72]	76.40	72.30
Proposed	84.60	77.12

TABLE II

RECOGNITION PERFORMANCE (%) BETWEEN PHOTO DOMAIN AND COMPOSITE SKETCH DOMAIN ON THE UoM-SGFS DATABASE

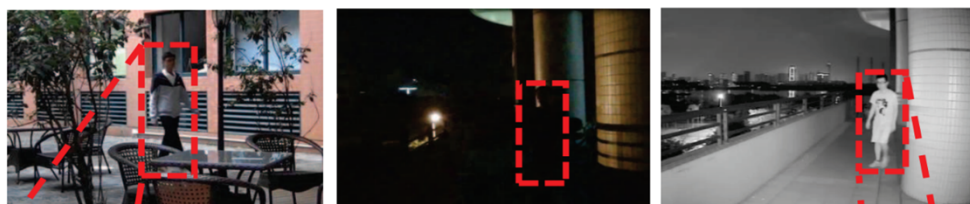
Method	Accuracy on subset A		Accuracy on subset B	
	Rank-1	Rank-10	Rank-1	Rank-10
PCA [61]	2.80	8.40	5.33	9.87
HAOG [76]	13.60	37.33	21.60	42.27
CBR [22]	5.73	18.80	7.60	25.47
LGMS [77]	21.87	51.20	43.47	73.60
VGG-Face [78]	9.33	31.07	16.13	48.00
D-RS [37]	22.13	49.33	40.80	70.80
D-RS+CBR [62]	25.87	56.00	42.93	75.87
DEEPS [1]	31.60	66.13	52.17	82.67
DLFace [59]	64.80	92.13	72.53	94.80
Proposed	70.16	94.60	73.83	95.10

Peng C, Wang N, Li J, et al. Soft Semantic Representation for Cross-Domain Face Recognition. IEEE Transactions on Information Forensics and Security, 2020.

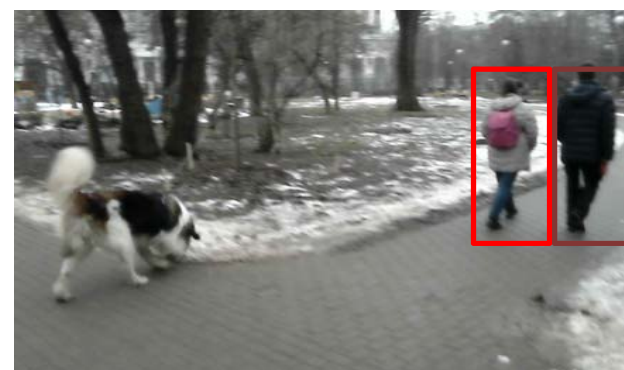
研究背景

■ 夜间监控场景

可见光图像（日间） 可见光图像（夜间） 红外图像（夜间）



跨模态行人图像



可见光图像

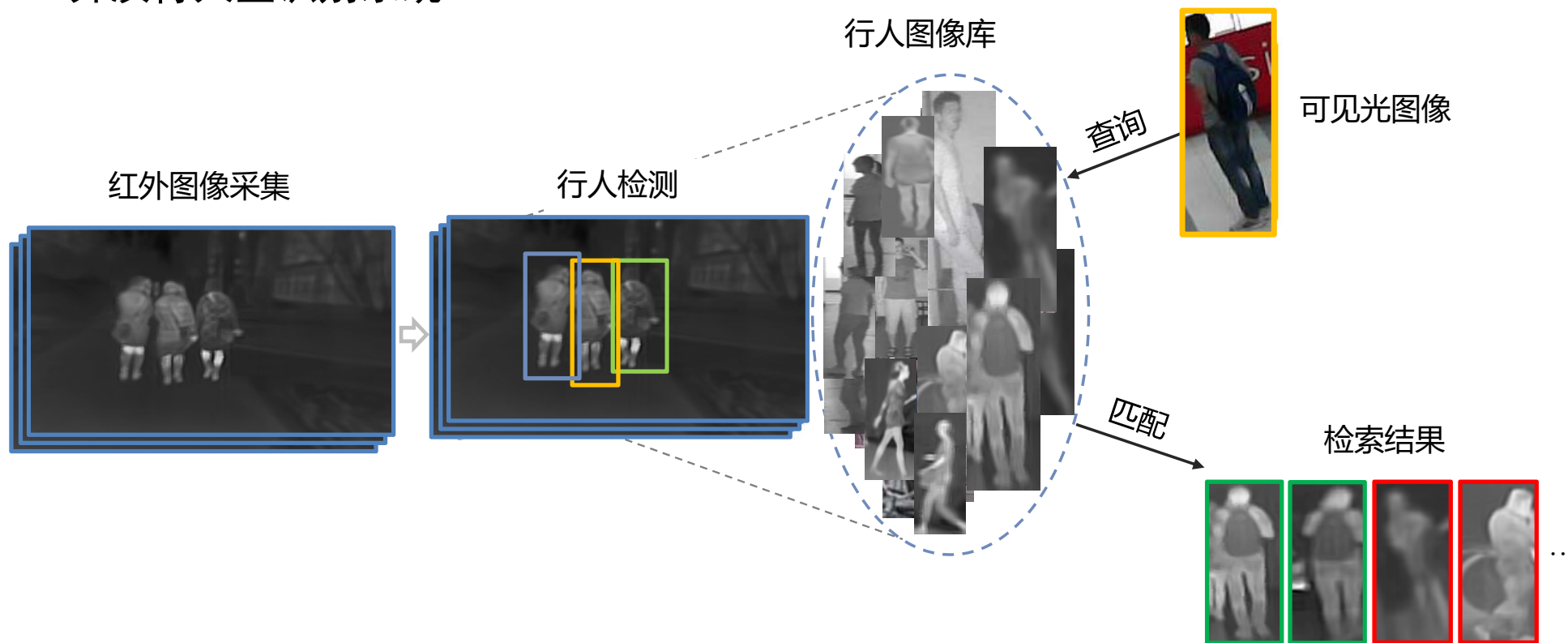


热红外图像

极端光照条件下，可见光到红外的行人重识别问题可弥补普通行人重识别系统的不足

研究背景

■ 异质行人重识别系统

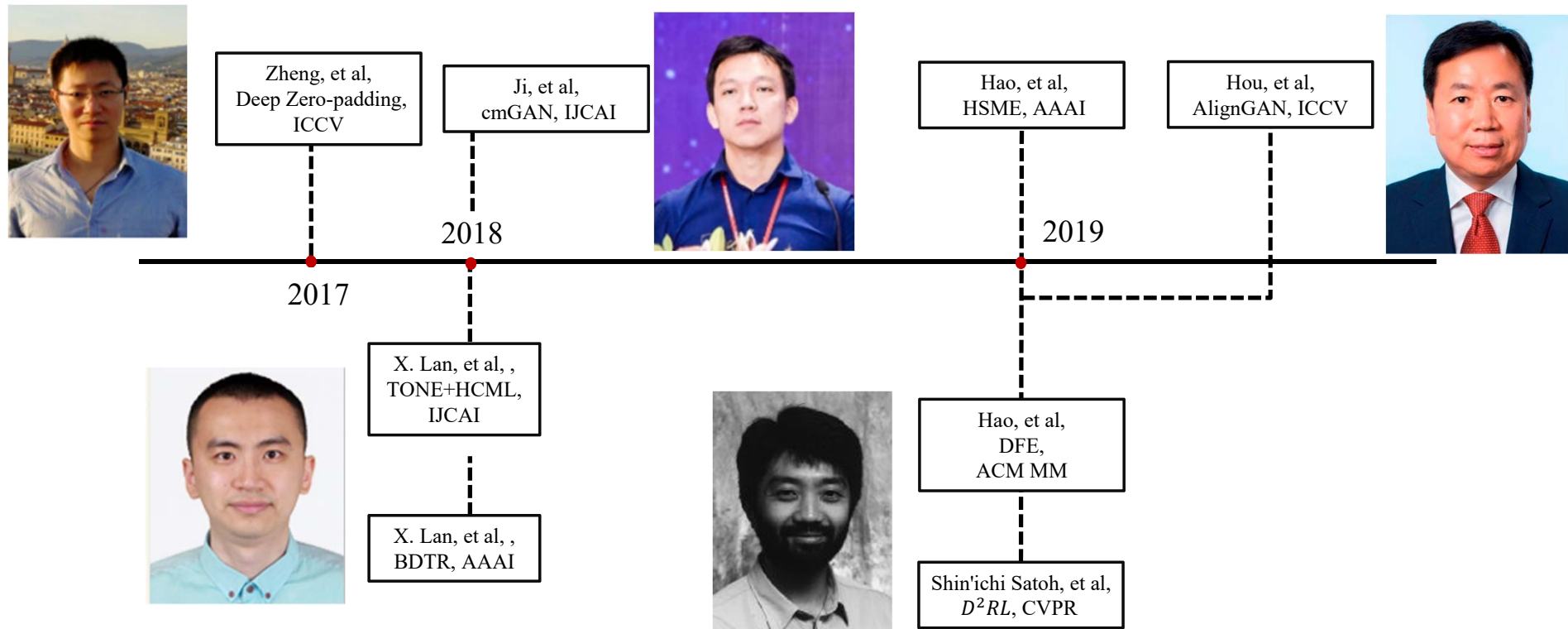


第一阶段：行人检测

第二阶段：行人重识别

跨模态行人重识别相关工作

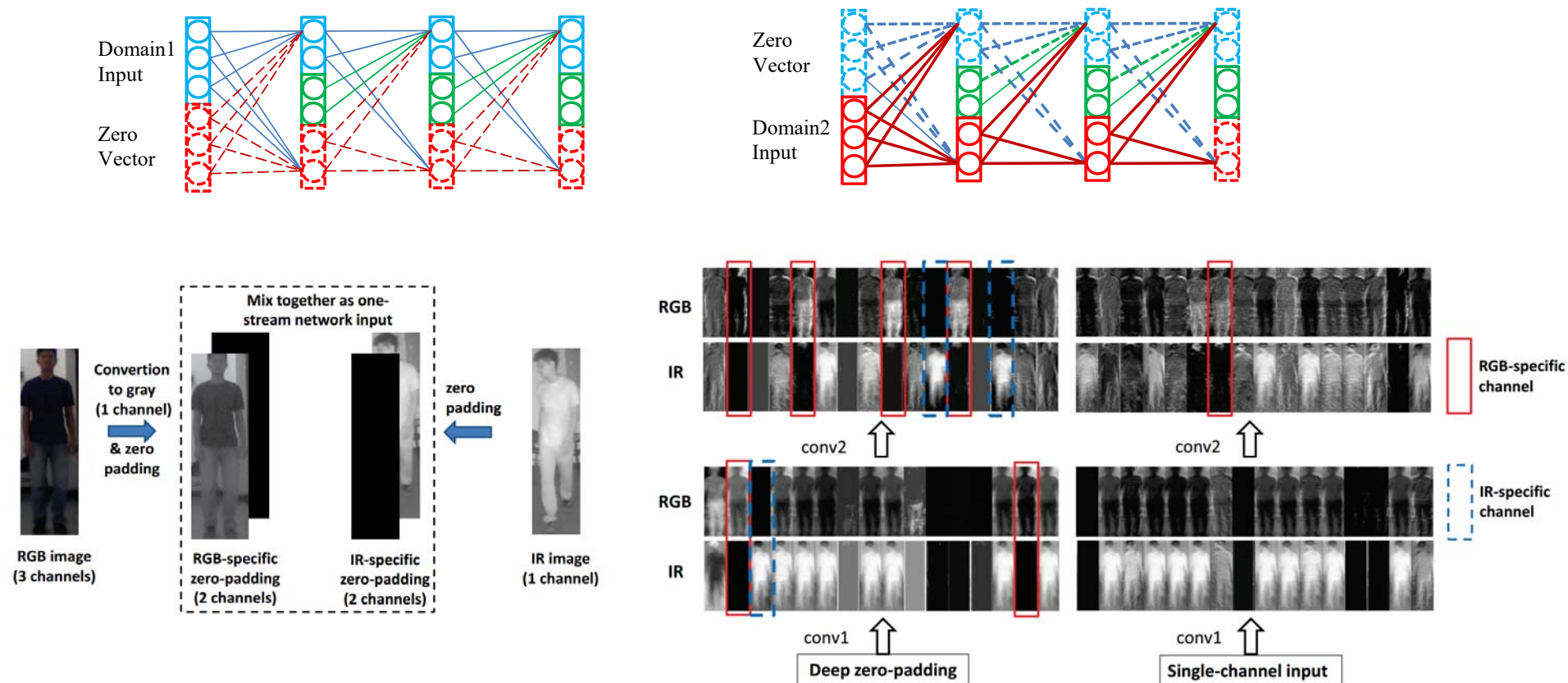
■ 现有工作发展时间线



TONE: **T**w**O**-stream CNN **N**etwork
HCML: **H**ierarchical **C**ross-modality **M**etric **L**earning
BDTR: **B**i-directional **D**ual-constrained **T**op-**R**anking loss
HSME: **H**yper**S**phere **M**anifold **E**mbedding

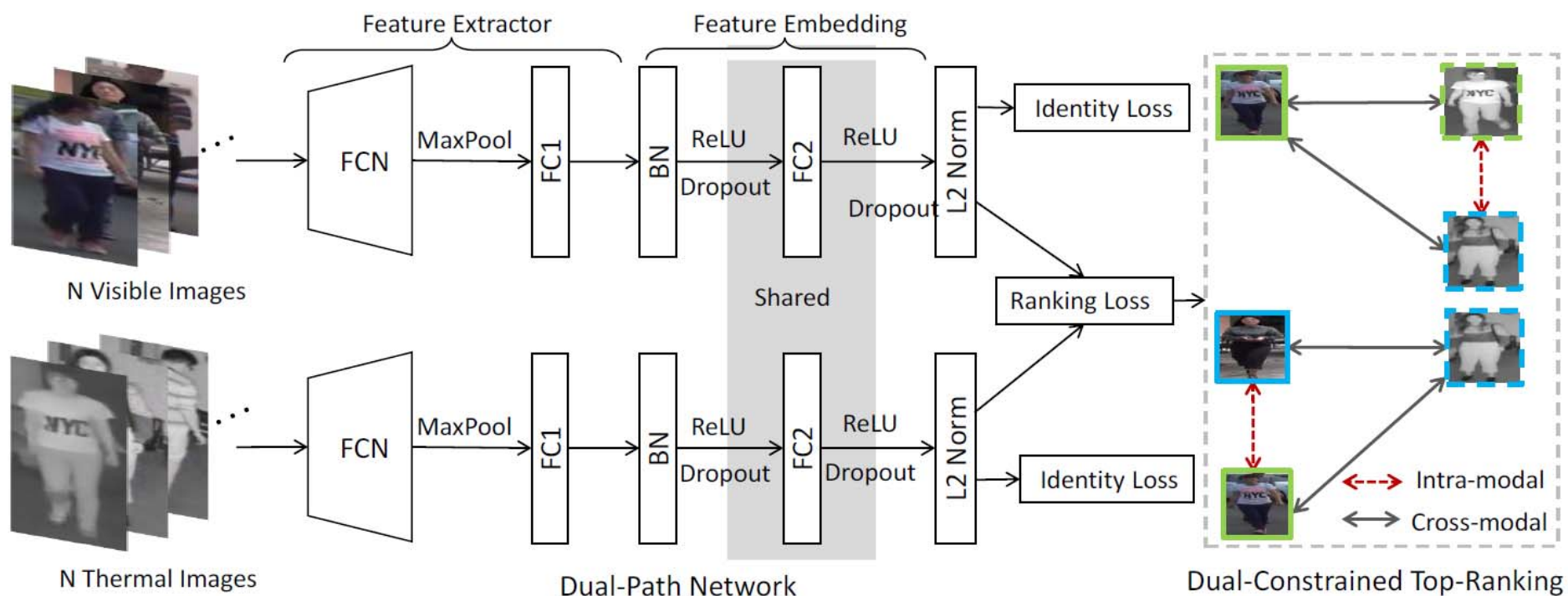
cmGAN: cross-modality GAN
DFE: **D**ual-alignment **F**eature **E**mbedding
 D^2RL : **D**ual-level **D**iscrepancy **R**eduction **L**earning

基于深度零填充的跨模态行人重识别



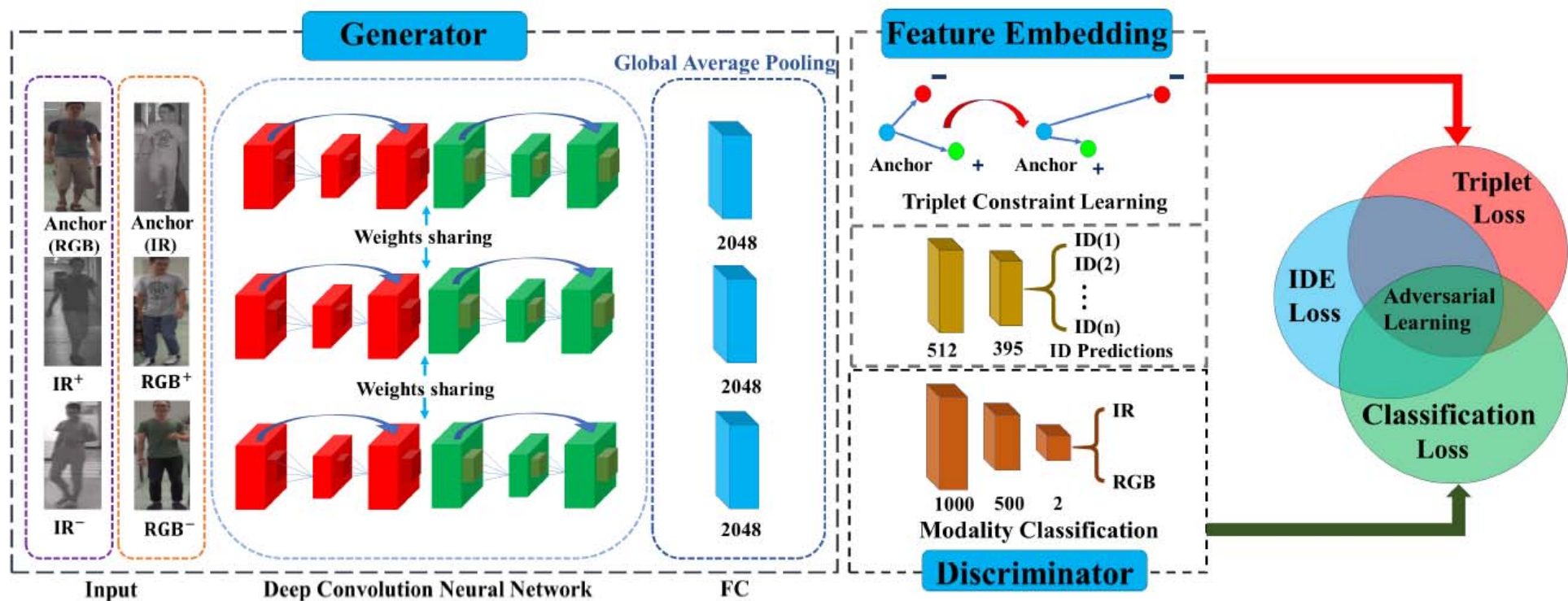
- 作者使用深度零填充的方法让网络自适应地提升共享神经元的作用；同时该工作还提出了一个大型的异质行人重识别数据库SYSU-MM01。

基于双向双约束排序损失的跨模态行人重识别



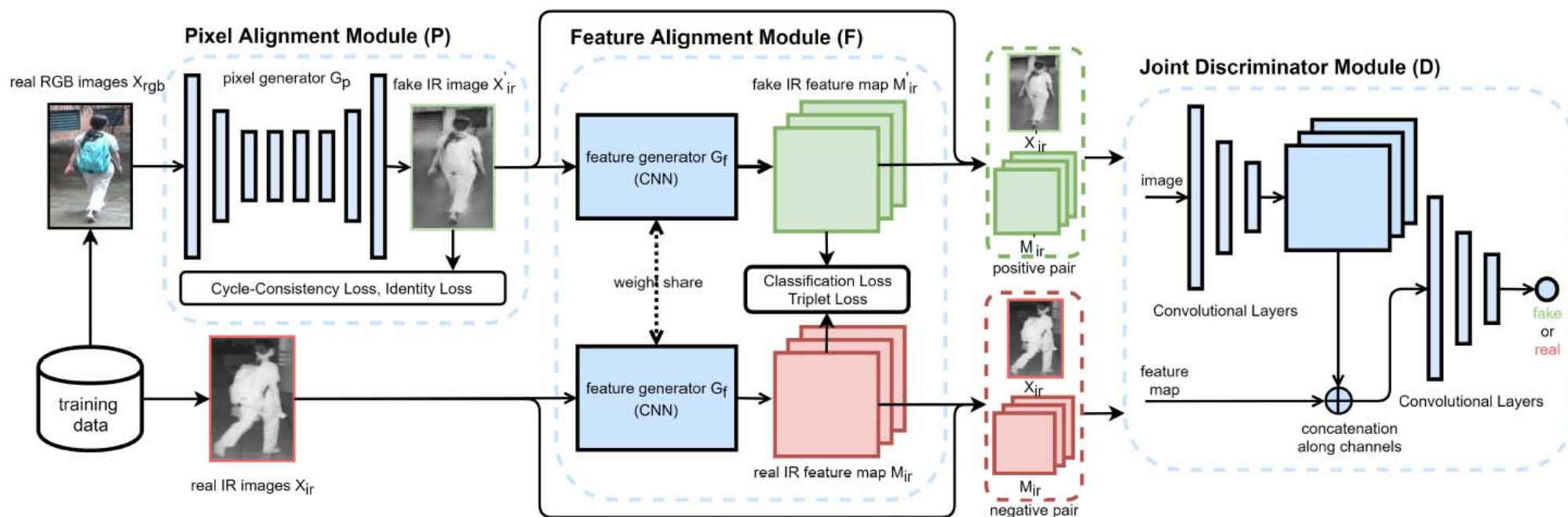
- 作者使用了权值共享的双流网络进行特征提取于嵌入表示，使用双向双约束排序损失对模型进行优化.

基于生成对抗训练的跨模态行人重识别



- 作者设计了一个模态判别器，使用对抗训练的思想获得具有模态不变性的深度特征表示；使用三元组损失和交叉熵损失的融合损失作为目标函数训练生成器。

像素-特征联合双对齐生成对抗网络

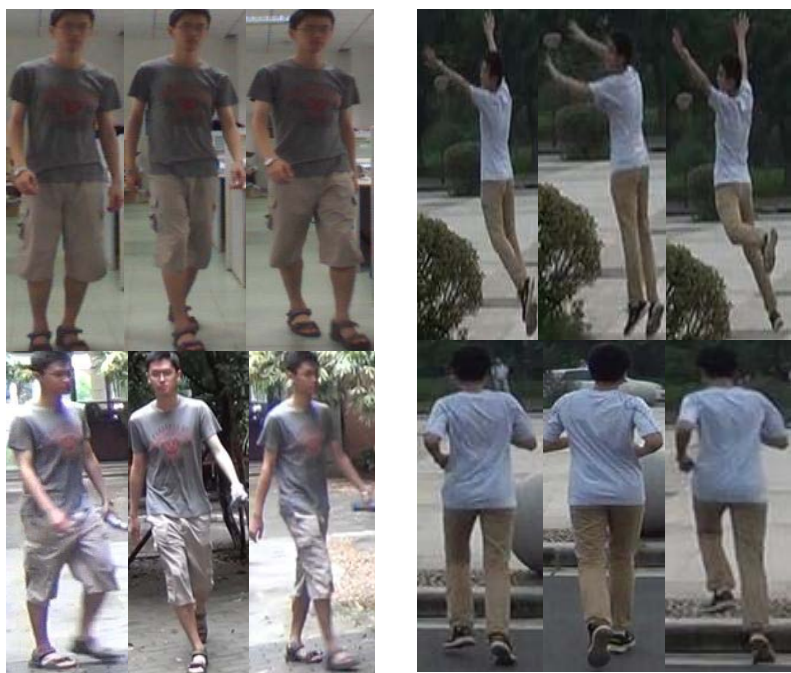


- 作者从生成的图像的像素级别和特征级别两个层面对生成图像联合进行优化，从而使生成图像在特征和像素层面都更接近真实的红外图像。

主要工作与创新

跨模态行人重识别面临的挑战

可见光行人图像



红外行人图像

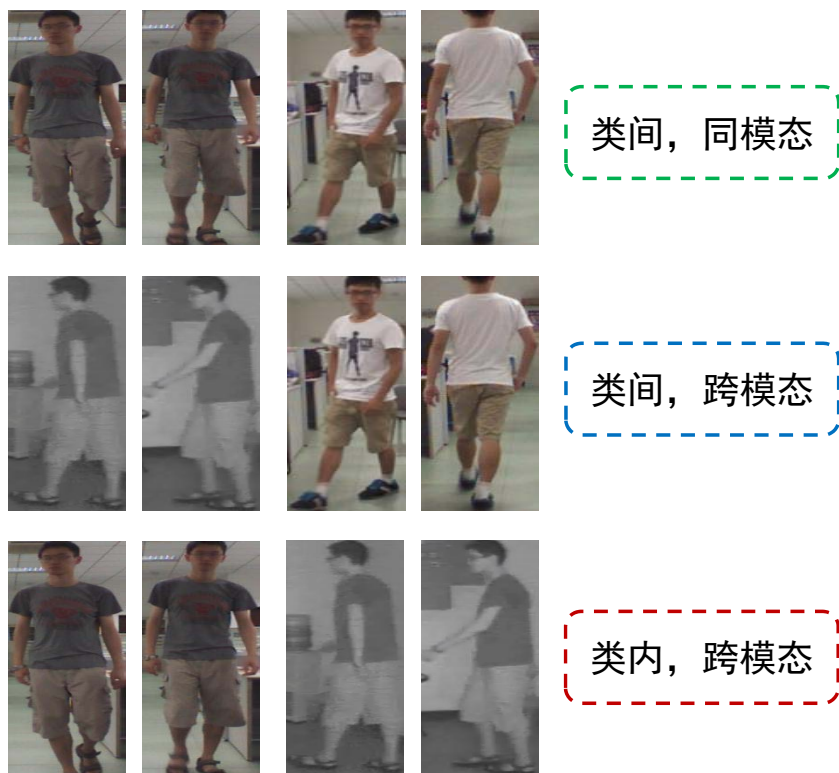


- 在同模态内和不同模态间，由于姿态和摄像机视角，行人图像存在**空间不对齐**的情况。
- 在不同模态间，由于不同传感器对不同物理信息敏感度的差异存在**模态不对齐**的情况。

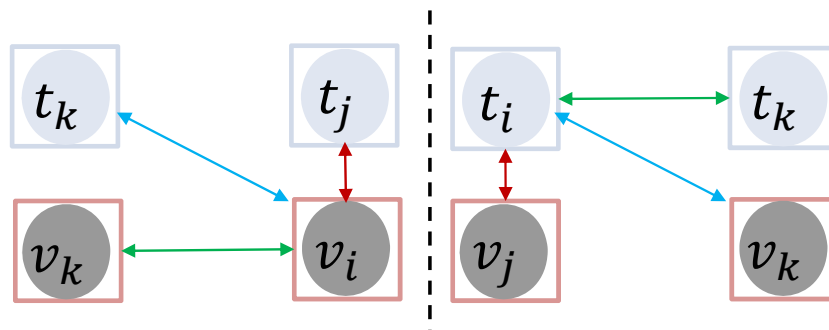
基于超球面流形嵌入的跨模态行人重识别

基本思想：考虑到跨模态行人重识别的复杂类内-类间关系，设计更全面双向排序损失，更细致地学习跨模态行人图像特征的相似性度量。

一个批次内所有图像可能存在的关系



双向排序损失



同模态排序损失

$$L_{intra} = \sum_{i=1}^N \max[(\rho + d(v_i, t_j) - \min_{\forall y_k \neq y_i} d(v_i, v_k)), 0]$$

$$+ \sum_{i=1}^N \max[(\rho + d(t_i, v_j) - \min_{\forall y_k \neq y_i} d(t_i, t_k)), 0]$$

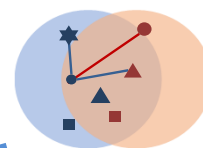
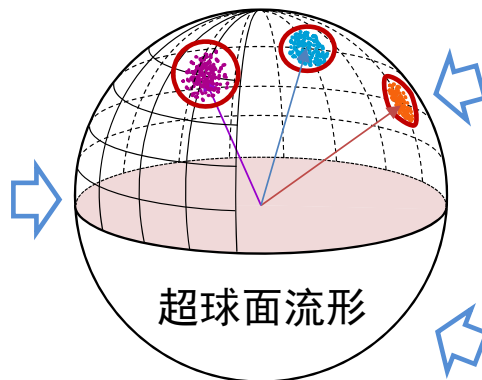
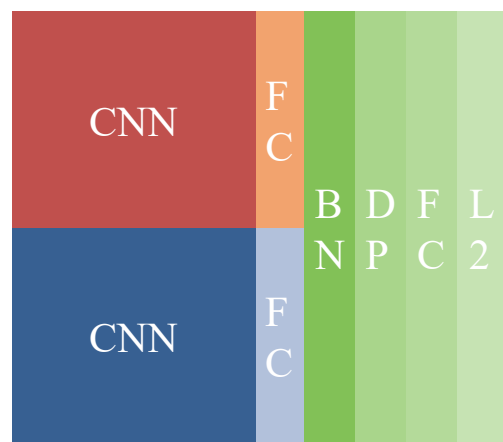
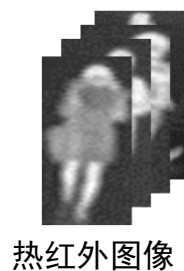
跨模态排序损失

$$L_{cross} = \sum_{i=1}^N \max[(\rho + d(v_i, t_j) - \min_{\forall y_k \neq y_i} d(v_i, t_k)), 0]$$

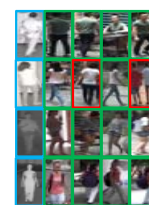
$$+ \sum_{i=1}^N \max[(\rho + d(t_i, v_j) - \min_{\forall y_k \neq y_i} d(t_i, v_k)), 0]$$

基于超球面流形嵌入的跨模态行人重识别

- 主要贡献：1) 设计双流神经网络提取具有模态不变性的特征表示，将跨域图像投影到共同超球面流形，使用超球面流形约束不同模态特征，扩大跨模态数据中的类间-类内距离比。2) 使用SVD分解，对超球面流形上的中心向量解相关，进一步扩大类间距离。

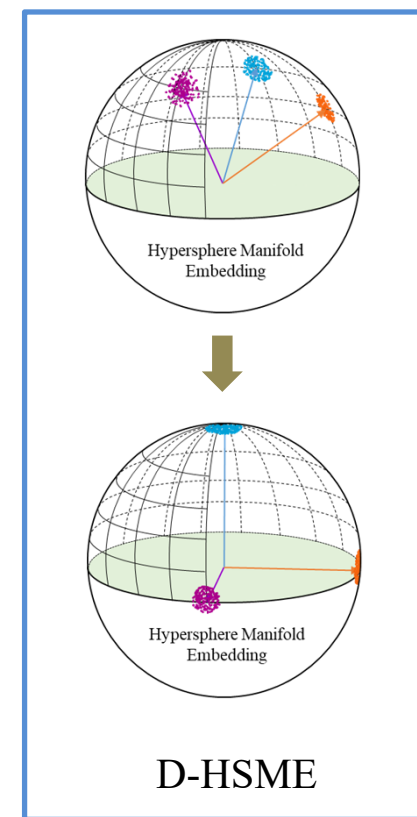


Reciprocal Ranking Loss



Sphere Loss

$$L_{sphere} = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s \cos \theta_{y_i}}}{\sum_{j=1}^C e^{s \cos \theta_j}}$$



基于超球面流形嵌入的跨模态行人重识别

- 实验结果：在两个主流的跨模态数据库SYSU-MM01和RegDB上超越了几个当时领先的算法；双向排序损失的对称性保证了不管是从可见光往红外识别，还是从红外往可见光识别，我们的模型都表现出良好的性能。

Comparison with the state-of-the-art methods.
CMC(%) and mAP(%)

Datasets	RegDB				SYSU-MM01			
	r=1	r=10	r=20	mAP	r=1	r=10	r=20	mAP
HOG	13.49	33.22	43.66	10.31	2.76	18.25	31.91	4.24
MLBP	2.20	7.33	10.90	6.77	2.12	16.23	28.32	3.86
LOMO	0.85	2.47	4.10	2.28	1.75	14.14	26.63	3.48
One-stream	13.11	32.98	42.51	14.02	12.04	49.68	66.74	13.67
Two-stream	12.43	30.36	40.96	13.42	11.65	47.99	65.50	12.85
Zero-Padding	17.75	34.21	44.35	18.90	14.80	54.12	71.33	15.59
TONE	16.87	34.03	44.10	14.92	12.52	50.72	68.69	14.42
TONE+XQDA	21.94	45.05	55.73	21.80	14.01	52.78	68.60	14.42
TONE+HMCL	24.44	47.53	56.78	20.80	14.32	53.16	69.17	16.16
BCTR	32.67	57.64	66.58	30.99	16.12	54.90	71.47	19.15
BDTR	33.47	58.42	67.52	31.83	17.01	55.43	71.96	19.66
Ours(HSME)	41.34	65.21	75.13	38.82	18.03	58.31	74.43	19.98
Ours(D-HSME)	50.85	73.36	81.66	47.00	20.68	62.74	77.95	23.12

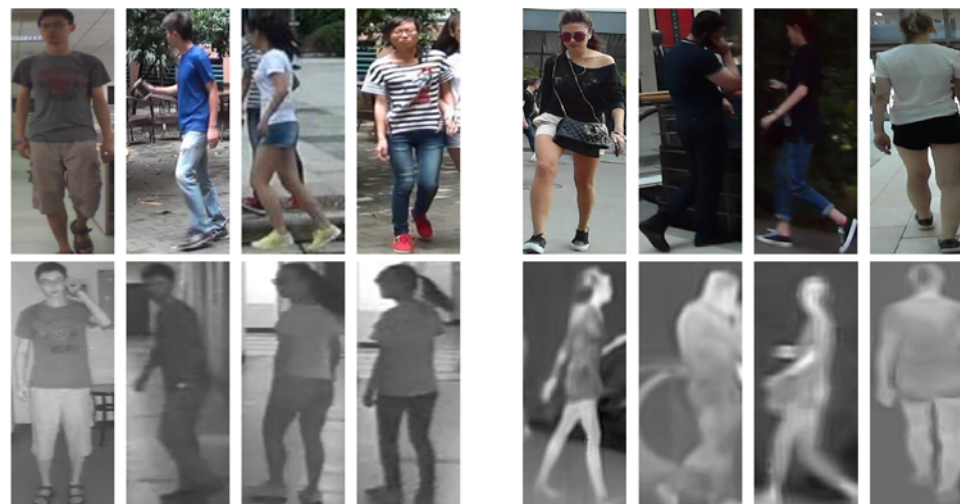
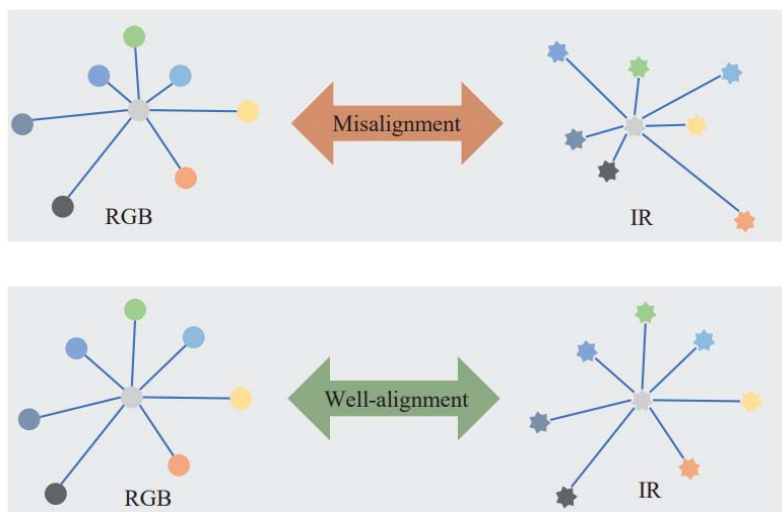
Different query settings

Visible to Thermal				
Methods	r=1	r=10	r=20	mAP
TONE+HCML	24.44	47.53	56.78	20.08
Zero-Padding	17.75	56.42	67.52	31.83
BDTR	33.47	58.42	67.52	31.83
Ours(HSME)	41.34	65.21	75.13	38.82
Ours(D-HSME)	50.85	73.36	81.66	47.00
Thermal to Visible				
Methods	r=1	r=10	r=20	mAP
TONE+HCML	21.70	45.02	55.58	22.24
Zero-Padding	16.63	34.68	44.25	17.82
BDTR	32.72	57.96	68.86	31.10
Ours(HSME)	40.67	65.35	75.27	37.50
Ours(D-HSME)	50.15	72.40	81.07	46.16

Hao Y, Wang N, Li J, Gao X. HSME: Hypersphere Manifold Embedding for Visible Thermal Person Re-Identification. In Proceedings of the AAAI Conference on Artificial Intelligence 2019 Jul 17 (Vol. 33, pp. 8385-8392).

基于双对齐特征嵌入的跨模态行人重识别

- 基本思想：跨模态行人重识别的挑战主要来源于空间不对齐和模态不对齐两方面，设计了双对齐特征嵌入网络同时解决这两个问题。

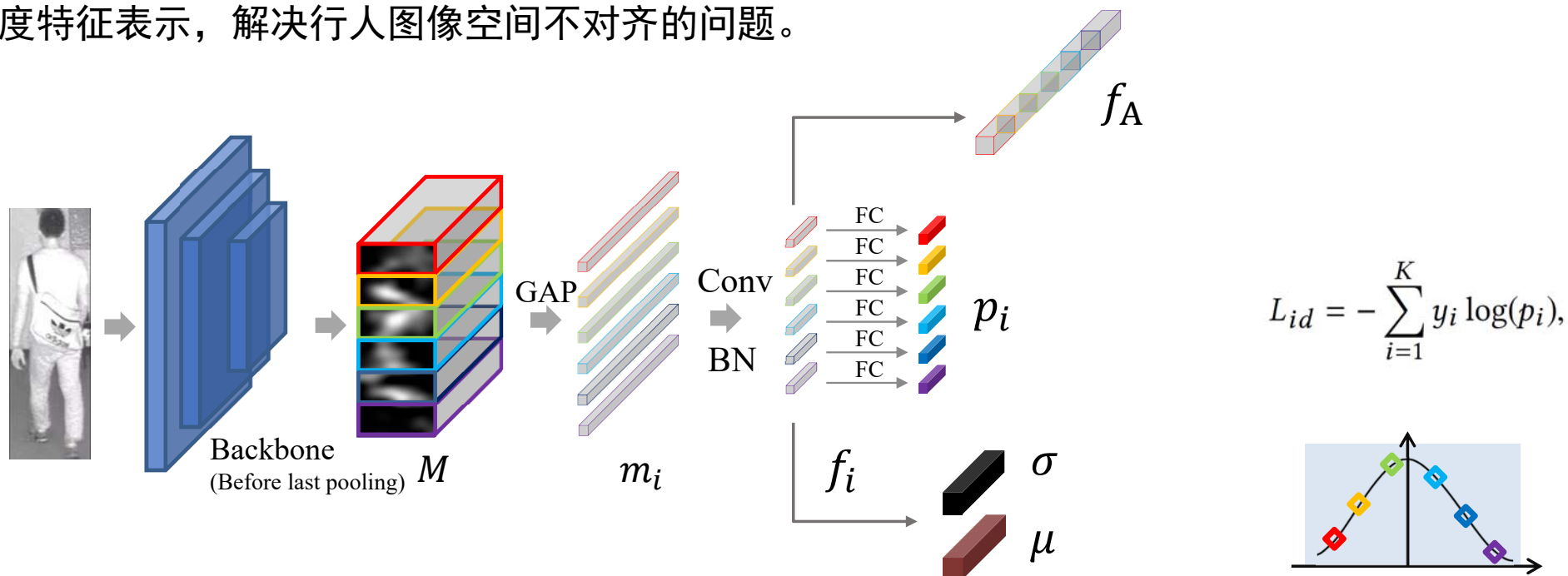


- 跨模态行人重识别的模态不对齐（上半部分为模态不对齐时，在不同的域内，特征分布不一致。下半部分为对齐后，特征分布表现出一致性。）

- 跨模态行人重识别的空间不对齐（上半部分为可见光图像，下半部分为不同波段的红外图像，上下均为同一人。）

基于双对齐特征嵌入的跨模态行人重识别

- 创新点1: 在PCB框架的基础上进行调整, 使用基于局部分块的骨干网络提取行人图像的细粒度特征表示, 解决行人图像空间不对齐的问题。

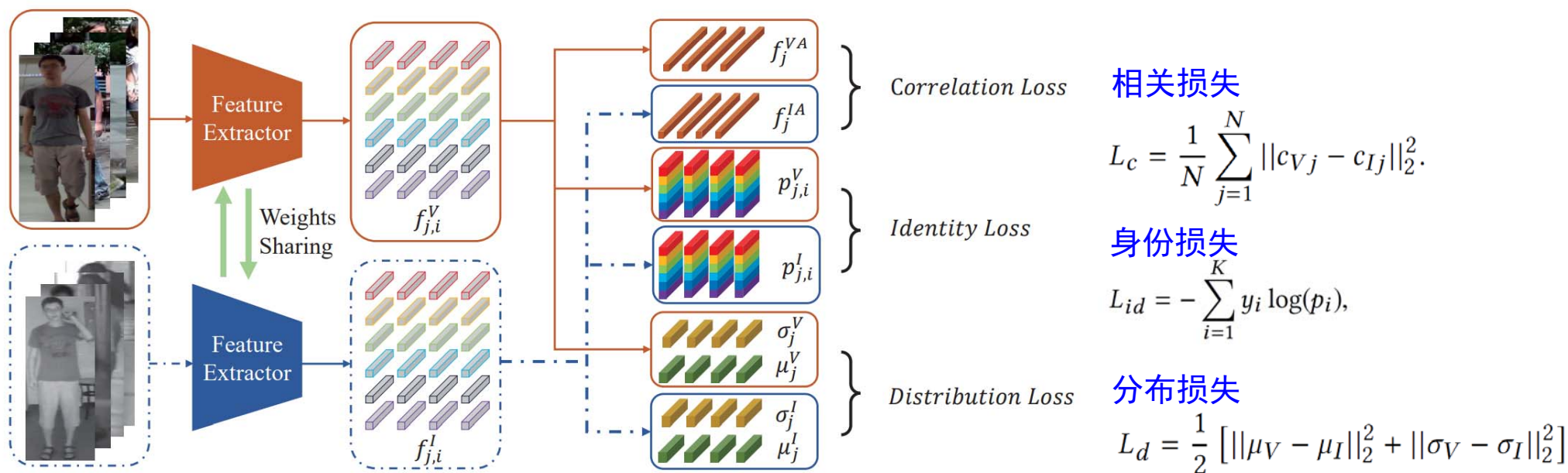


$$\mu_I = \frac{1}{K} \sum_{i=1}^K f_i^I, \sigma_I = \sqrt{\frac{1}{K} \sum_{i=1}^K [(f_i^I - \mu_I) \cdot (f_i^I - \mu_I)]}, \mu_V = \frac{1}{K} \sum_{i=1}^K f_i^V, \sigma_V = \sqrt{\frac{1}{K} \sum_{i=1}^K [(f_i^V - \mu_V) \cdot (f_i^V - \mu_V)]}.$$

计算特征的均值和方差用于分布损失, 角标及上标处的 \$I, V\$ 分别表示红外数据和可见光数据

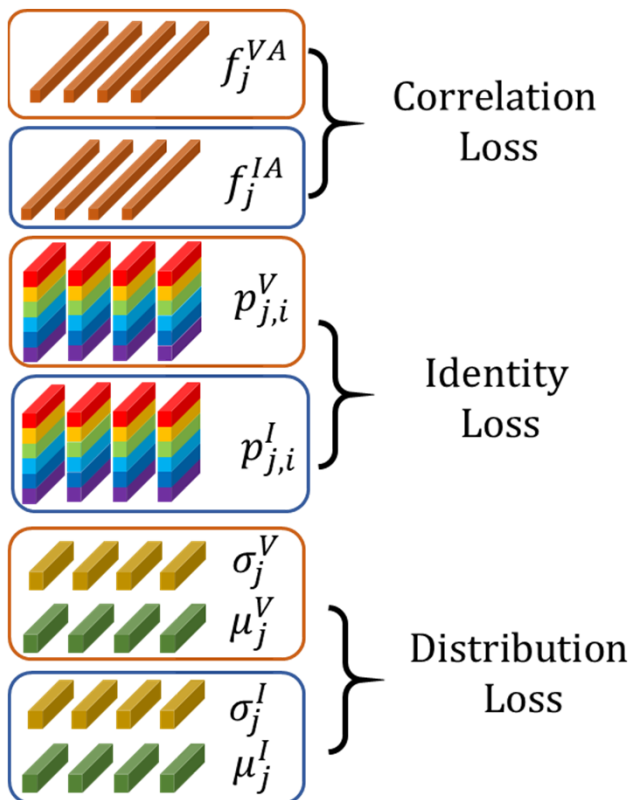
基于双对齐特征嵌入的跨模态行人重识别

- 创新点2: 使用双流网络结构同时提取可见光和红外图像的深度特征表示, 设计了两个基于分布度量的**相关损失**和**分布损失**, 解决跨模态行人重识别的模态不对齐问题。



基于双对齐特征嵌入的跨模态行人重识别

相关损失



$$f_j^{VA} (j = 1, 2, \dots, N)$$

$$f_j^{IA} (j = 1, 2, \dots, N)$$

$$x = [x_1, x_2, \dots, x_n]^T$$

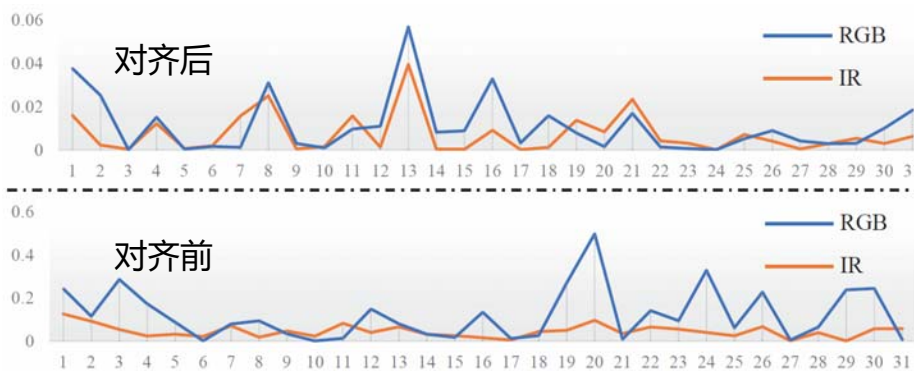
$$y = [y_1, y_2, \dots, y_n]^T$$

$$c_{xy} = \frac{\sum_{j=1}^n (x_j - \hat{x})(y_j - \hat{y})}{\sqrt{\sum_{j=1}^n (x_j - \hat{x})^2} \sqrt{\sum_{j=1}^n (y_j - \hat{y})^2}}$$

$$C_I = [c_{I1}, c_{I2}, \dots, c_{IN}]^T$$

$$C_V = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1N} \\ c_{21} & c_{22} & \dots & c_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ c_{N1} & c_{N2} & \dots & c_{NN} \end{bmatrix}$$

$$L_c = \frac{1}{N} \sum_{j=1}^N \|c_{Vj} - c_{Ij}\|_2^2$$



使用相关损失对齐后，比较32组跨模态图像的相关系数，发现对齐后的更具有有一致性。

基于双对齐特征嵌入的跨模态行人重识别

- 实验结果：在两个主流的跨模态数据库SYSU-MM01和RegDB上超越取得了领先的效果。消融学习的结果证明了相关损失和分布损失对模态对齐具有重要作用。基线模型的效果说明了消除空间不对齐可以有效提高跨模态行人重识别的准确度。

在RegDB数据库上与SOTA方法的对比，以及消融学习的结果。CMC(%) and mAP(%)

Method	Rank 1	Rank 10	Rank 20	mAP
HOG	13.49	33.22	43.66	10.31
GSM	17.28	34.47	45.26	15.06
LOMO	0.85	2.47	4.10	2.28
One-stream	13.11	32.98	42.51	14.02
Two-stream	12.43	30.36	40.96	13.42
Zero-Padding	17.75	34.21	44.35	18.90
TONE+XDQA	21.94	45.05	55.73	21.80
TONE+MLAPG	17.82	40.29	49.73	18.03
TONE+HMCL	24.44	47.53	56.78	20.80
BCTR	32.67	57.64	66.58	30.99
BDTR	33.47	58.42	67.52	31.83
HSME	41.34	65.21	75.13	38.82
D-HSME	50.85	73.36	81.66	47.00
Ours(baseline)	57.50	75.63	83.47	57.30
Ours(DFE w/o L_c)	64.18	81.48	87.84	63.60
Ours(DFE w/o L_d)	62.48	81.29	88.16	61.10
Ours(DFE)	70.13	86.32	91.96	69.14

在SYSU-MM01上与SOTA方法的对比，CMC(%) and mAP(%)

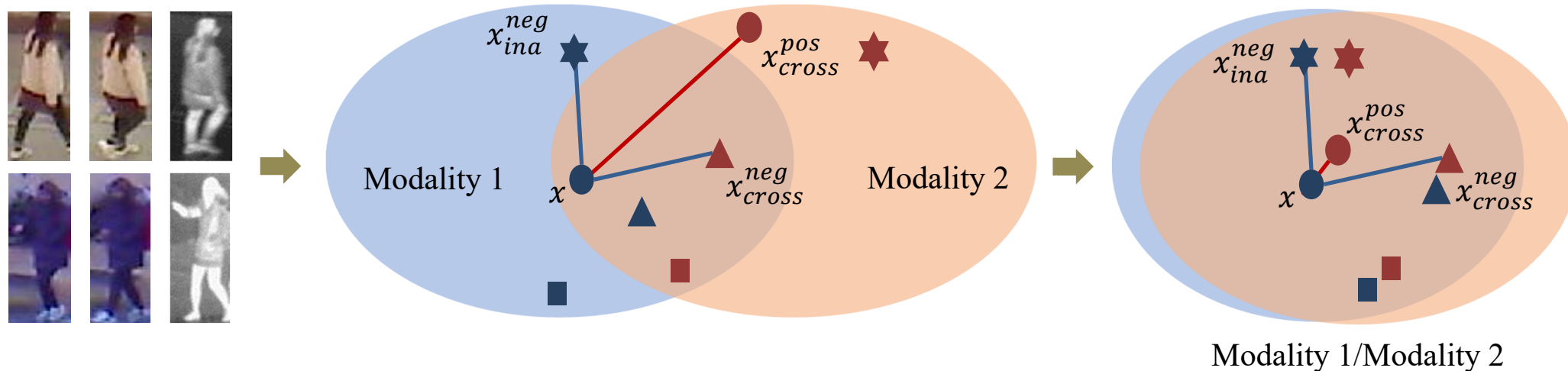
Method	all-search								indoor-search							
	single-shot				multi-shot				single-shot				multi-shot			
	r1	r10	r20	mAP	r1	r10	r20	mAP	r1	r10	r20	mAP	r1	r10	r20	mAP
HOG+Euclidean	2.76	18.25	31.91	4.24	3.82	22.77	37.63	2.16	3.22	24.68	44.52	7.25	4.75	29.06	49.38	3.51
HOG+CRAFT	2.59	17.93	31.50	4.24	3.58	22.90	38.59	2.06	3.03	24.07	42.89	7.07	4.16	27.75	47.16	3.17
HOG+CCA	2.74	18.91	32.51	4.28	3.25	21.82	36.51	2.04	4.38	29.96	50.43	8.70	4.62	34.22	56.28	3.87
HOG+LFDA	2.33	18.58	33.38	4.35	3.82	20.48	35.84	2.20	2.44	24.13	45.50	6.87	3.42	25.27	45.11	3.19
LOMO+CCA	2.42	18.22	32.45	4.19	2.63	19.68	34.82	2.15	4.11	30.60	52.54	8.83	4.86	34.40	57.30	4.47
LOMO+CRAFT	2.34	18.70	32.93	4.22	3.03	21.70	37.05	2.13	3.89	27.55	48.16	8.37	2.45	20.20	38.15	2.69
LOMO+CDFE	3.64	23.18	37.28	4.53	4.70	28.23	43.05	2.28	5.75	34.35	54.90	10.19	7.36	40.38	60.44	5.64
LOMO+LFDA	2.98	21.11	35.36	4.81	3.86	24.01	40.54	2.61	4.81	32.16	52.50	9.56	6.27	36.29	58.11	5.15
Asymmetric FC	9.30	43.26	60.38	10.82	13.06	52.11	69.52	6.68	14.59	57.94	78.68	20.33	20.09	69.37	85.08	13.04
Two-stream	11.65	47.99	65.50	12.85	16.33	58.35	74.46	8.03	15.60	61.18	81.02	21.49	22.49	72.22	88.61	13.92
One-stream	12.04	49.68	66.74	13.67	16.26	58.14	75.05	8.59	16.94	63.55	82.10	22.95	22.62	71.74	87.82	15.04
Zero-padding	14.80	54.12	71.33	15.95	19.13	61.40	78.41	10.89	20.58	68.38	85.79	26.92	24.43	75.86	91.32	18.64
cmGAN	26.97	67.51	80.56	27.80	31.49	72.74	85.01	22.27	31.63	77.23	89.18	42.19	37.00	80.94	92.11	32.76
Ours(baseline)	43.13	85.37	93.67	43.56	48.69	88.94	95.63	36.92	45.24	86.40	94.75	53.75	50.38	92.02	96.98	44.48
Ours(DFE w/o L_c)	45.81	86.81	94.59	46.37	52.26	90.58	96.14	39.76	51.32	89.17	96.46	58.73	57.97	92.42	97.6	49.66
Ours(DFE)	48.71	88.86	95.27	48.59	54.63	91.62	96.83	42.14	52.25	89.86	95.85	59.68	59.62	94.45	98.07	50.60

部分检索结果排序列表，绿色为正确的结果，红色为错误的检索结果。



基于模态对抗网络的可见光-热红外行人重识别

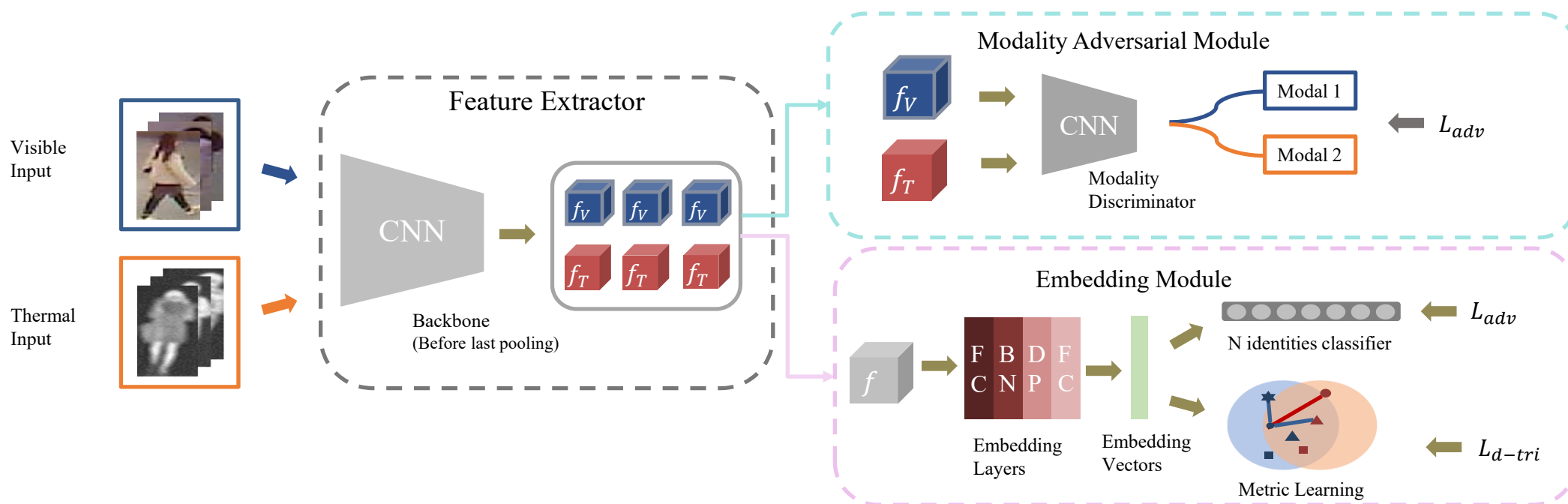
- 基本思想：使用对抗学习策略，缩小可见光图像与热红外图像在特征空间的差异性，即提高图像特征的**模态不变性**。



- 使用对抗学习提高特征的模态不变性，从而使不同模态图像的特征空域重合度更高。
- 使用双约束三元组损失与身份鉴别损失提高深度特征的判别性以保证图像特征的身份辨识度。

基于模态对抗网络的可见光-热红外行人重识别

- 主要贡献：设计**模态对抗网络**，将跨域图像投影到特征空间，使用生成-判别模型优化特征提取器以使模型学习提取**模态不变特征**；双约束三元组损失从可见光-热红外与热红外-可见光**两个匹配方向**使用度量学习的方法约束特征提取器的学习过程，从而提高特征的判别性。



模态对抗损失

$$L_{adv}(X^V, X^T; E_f, D_m) = -\mathbb{E}_{x^V \sim X^V} \log(D_m(f^V)) \\ + \mathbb{E}_{x^T \sim X^T} (1 - \log(D_m(f^T)))$$

双约束三元组损失

$$L_{d-tri}(X^V, X^T; E_f, L_e) = \mathbb{E}_{(x^V, y^V) \sim (X^V, Y^V)} \max[(m + d_{cross}^{pos} - d_{cross}^{neg}), 0] \\ + \mathbb{E}_{(x^T, y^T) \sim (X^T, Y^T)} \max[(m + d_{cross}^{pos} - d_{cross}^{neg}), 0]$$

基于模态对抗网络的可见光-热红外行人重识别

- 实验结果：在两个主流的跨模态数据库SYSU-MM01和RegDB上超越了几个当时领先的算法；从消融实验可以看出对抗学习思想对跨可见光-热红外行人重识别模型的性能有明显提高。同时该方法在可见光-红外数据库上也取得了当时领先的效果。

在RegDB数据库上与SOTA方法的对比，以及消融学习的结果。CMC(%) and mAP(%)

Method	Rank-1	Rank-10	Rank-20	mAP
HOG [32]	13.49	33.22	43.66	10.31
MLBP [5]	2.20	7.33	10.90	6.77
LOMO [33]	0.85	2.47	4.10	2.28
One-stream [4]	13.11	32.98	42.51	14.02
Two-stream [4]	12.43	30.36	40.96	13.42
Zero-Padding [4]	17.75	34.21	44.35	18.90
TONE+XDQA [34]	21.94	45.05	55.73	21.80
TONE+MLAPG [33]	17.82	40.29	49.73	18.03
TONE+HMCL [5]	24.44	47.53	56.78	20.80
BCTR [6]	32.67	57.64	66.58	30.99
BDTR [6]	33.47	58.42	67.52	31.83
HSME [23]	41.34	65.21	75.13	38.82
MANN w/o L_{adv}	31.38	56.25	66.90	29.86
MANN(Ours)	48.67	71.55	80.06	41.11

在SYSU-MM01数据库上与SOTA方法的对比，以及消融学习的结果。CMC(%) and mAP(%)

Method	Rank-1	Rank-10	Rank-20	mAP
HOG [32]	2.76	18.25	31.91	4.24
One-stream [4]	12.04	49.68	66.74	13.67
Two-stream [4]	11.65	47.99	65.50	12.85
Zero-Padding [4]	14.80	54.12	71.33	15.95
TONE [5]	12.52	50.72	68.69	14.42
TONE+XDQA [34]	14.01	52.78	68.60	14.42
TONE+HMCL [5]	14.32	53.16	69.17	16.16
BCTR [6]	16.12	54.90	71.47	19.15
BDTR [6]	17.01	55.43	71.96	19.66
HSME [23]	18.03	58.31	74.43	19.98
MANN w/o L_{adv}	25.34	68.65	82.86	27.99
MANN(Ours)	30.04	74.34	86.98	32.18

Hao Y, Li J, Wang N, Gao X. Modality Adversarial Neural Network for Visible-Thermal Person Re-identification. Pattern Recognition, 107, 107533.

谢谢！

