

图像跨域重建理论与方法

Cross-domain Image Reconstruction

报告人：高新波

重庆邮电大学 / 西安电子科技大学

汇报提纲

- 一. **前言** (Preface) : 什么是图像跨域重建?
- 二. **画像-照片合成** (Sketch-Photo Synthesis)
- 三. **图像超分辨重建** (Image Super-Resolution Reconstruction)
- 四. **其他相关应用** (Other Related Applications)
- 五. **结束语** (Conclusions)



前言：Preface

新一代人工智能发展规划 国发〔2017〕35号

AI 1.0 → AI 2.0



2020

2025

2030

- **第一步**，到2020年人工智能（AI）总体技术和应用与世界先进水平同步，人工智能产业成为新的重要经济增长点
 - 新一代人工智能**理论和技术**取得重要进展，**大数据智能**、**跨媒体智能**、**群体智能**、**混合增强智能**、**自主智能系统**等基础理论和核心技术实现重要进展
 - 人工智能产业竞争力进入国际第一方阵。人工智能核心产业规模超过1500亿元，带动相关产业规模超过1万亿元

前言：Preface



中华人民共和国中央人民政府

www.gov.cn

简 | 繁 | EN | 注册 | 登录



国务院

总理

新闻

政策

互动

服务

数据

国情

国家政务服务平台

首页 > 新闻 > 直播 > 国务院政策吹风会 > 解读政策

新一代的人工智能主要是大数据基础上的人工智能

2017-07-21 13:31 来源： 中国政府网

【字体：大 中 小】 打印 分享 微信 微博 +

国务院新闻办公室于2017年7月21日（星期五）上午10时举行国务院政策例行吹风会，请科学技术部副部长李萌介绍《新一代人工智能发展规划》有关情况，并答记者问。

【央视新闻记者】规划的名称叫《新一代人工智能发展规划》，那么如何解读新一代和以往的人工智能有什么区别？另外，人工智能是否可以简单理解为机器人？

【李萌】关于人工智能的划代，国际社会和国内专家都有从不同角度进行过研究，但共识是非常高的，1956年提出人工智能概念以后，经过60多年的演进，特别是在移动互联网、大数据、超级计算、传感网、脑科学等新理论新技术的引领，再加上经济社会发展强烈需求的驱动，人工智能出现了一些新特点，包括《规划》当中讲到“它呈现出深度学习、跨界融合、人机协同、群智开放和自主智能的新特点”。新一代的人工智能主要是大数据基础上的人工智能。

它有几个特点：一是从人工知识表达到大数据驱动的知识学习技术；二是从分类型处理的多媒体数据转向跨媒体的认知、学习、推理，这里讲的“媒体”不是新闻媒体，而是界面或者环境。三是从追求智能机器到高水平的人机、脑机相互协同和融合。四是从聚焦个体智能到基于互联网和大数据的群体智能，它可以把很多人的智能集聚融合起来变成群体智能。五是从拟人化的机器人转向更加广阔的智能自主系统，不是一个单纯的机器人人才叫人工智能，比如说智能工厂、智能无人机系统，这都是人工智能。

科技部副部长李萌介绍AI2.0时，指出：新一代人工智能主要是**大数据**基础上的人工智能

[AI 1.0 vs. AI 2.0](#) [AI 3.0?](#)

AI 1.0 基于单一媒体的分析与应用

AI 2.0 基于**跨媒体**的分析与应用，强调跨媒体源的采集、分析、检索和推理，即**跨媒体智能**

非结构化数据：如图像、视频，占到 90% 以上的比例。这些数据虽然**形式上是异构的**，但在**语义上相互关联**

跨媒体智能：多元异构大数据的分析、识别、检索和推理

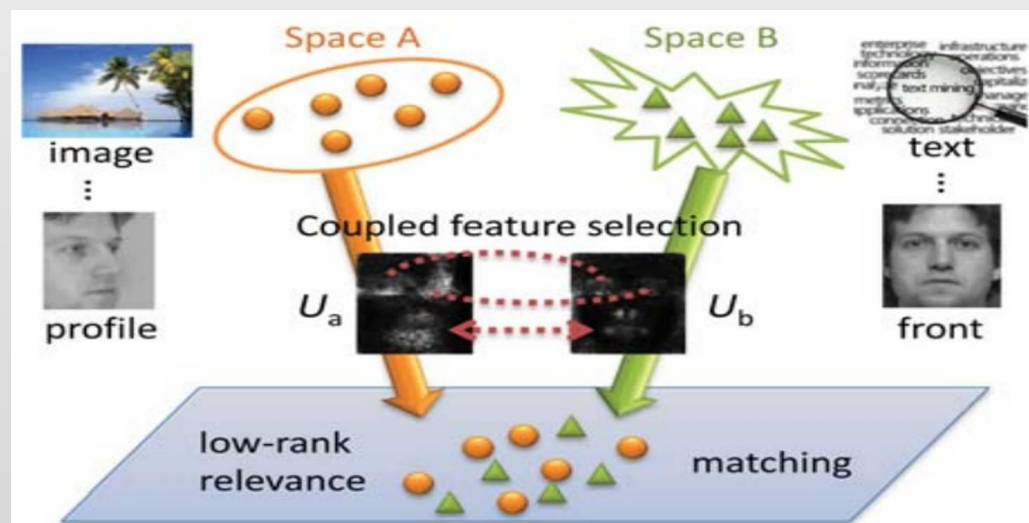
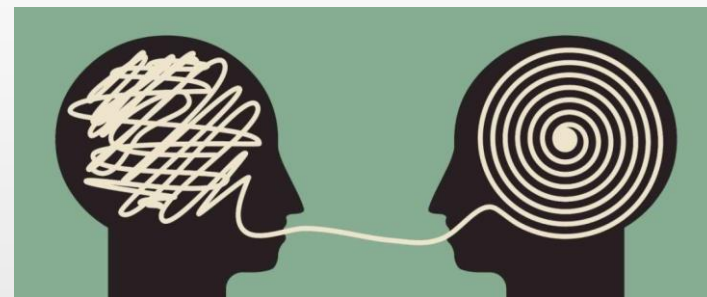
前言：Preface

跨媒体智能面临两大科学问题： **语义鸿沟**（Semantic Gap）和**异构鸿沟**（Heterogeneity Gap）

- **语义鸿沟**：是指计算机**表示**与人类所**理解**的**语义概念**不一致（自下而上 vs. 自上而下）
- **异构鸿沟**：是指不同模态的数据的**特征表示**不一致

AI 2.0中跨媒体智能的研究内容：

1. 跨媒体统一表征的理论和模型
2. 跨媒体的关联、理解与挖掘
3. 跨媒体知识图谱的构建与学习
4. 跨媒体的知识演化和推理
5. **跨媒体的描述与生成**
6. 跨媒体的智能搜索引擎
7. 跨媒体的智能应用

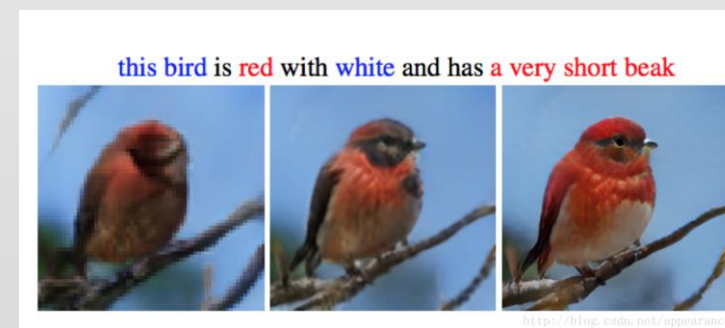
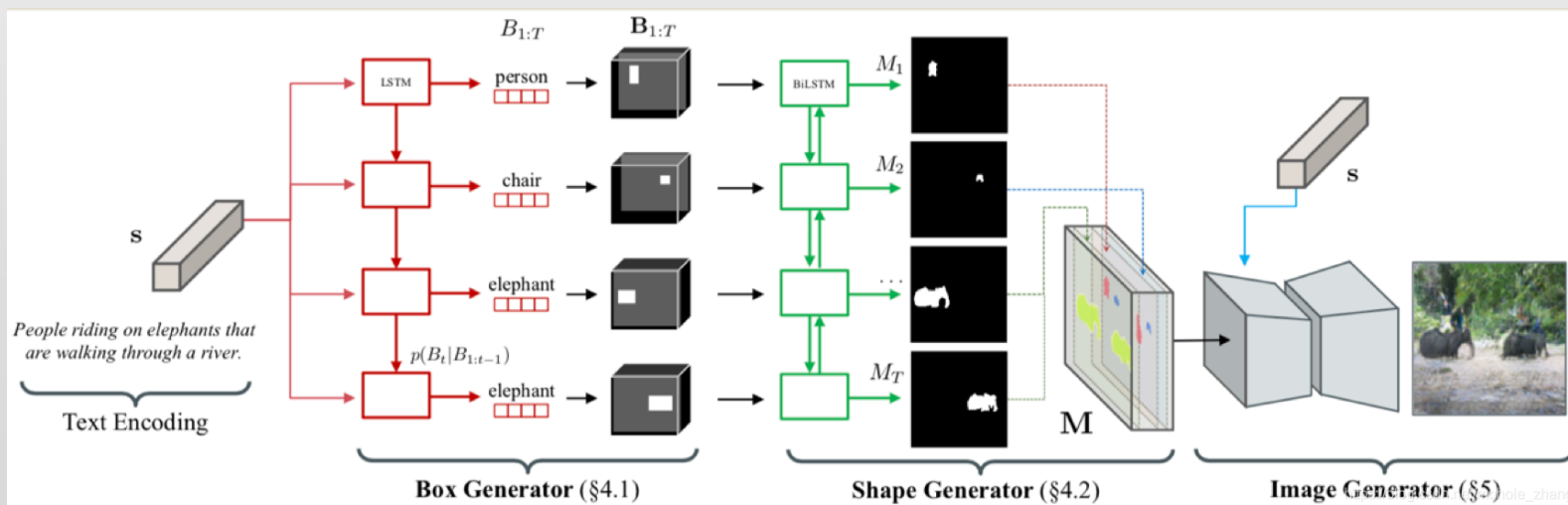
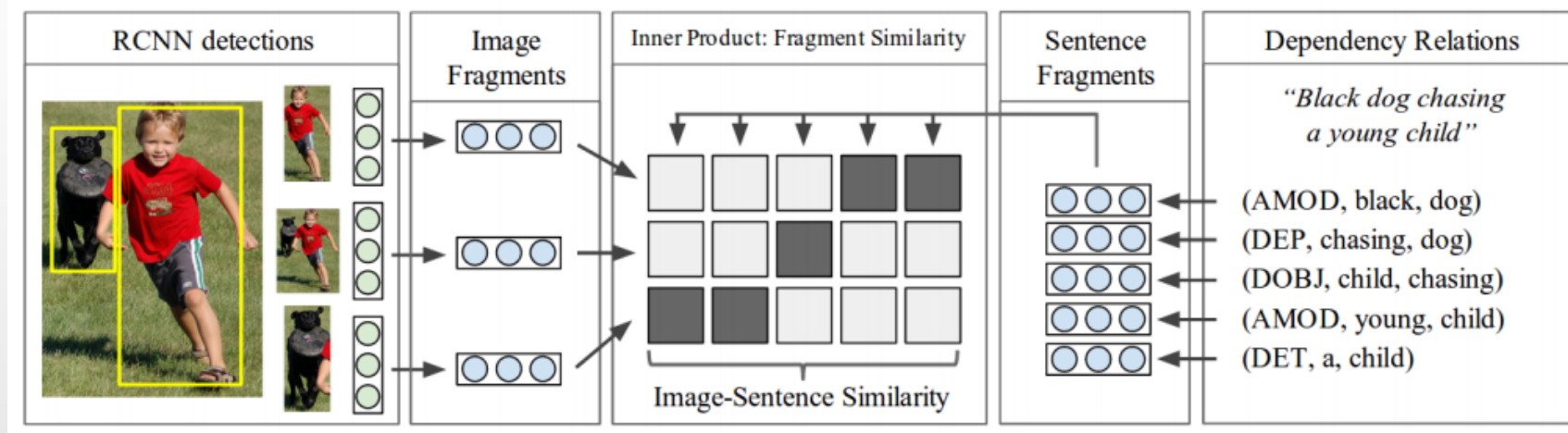


前言: Preface

跨媒体的描述与生成

► 图像翻译

- 图像→图像
- 图像→文字
- 文字→图像
- 文字→视频



前言：Preface

图像翻译（Image Translation）：图像→图像

跨域图像 & 图像跨域重建



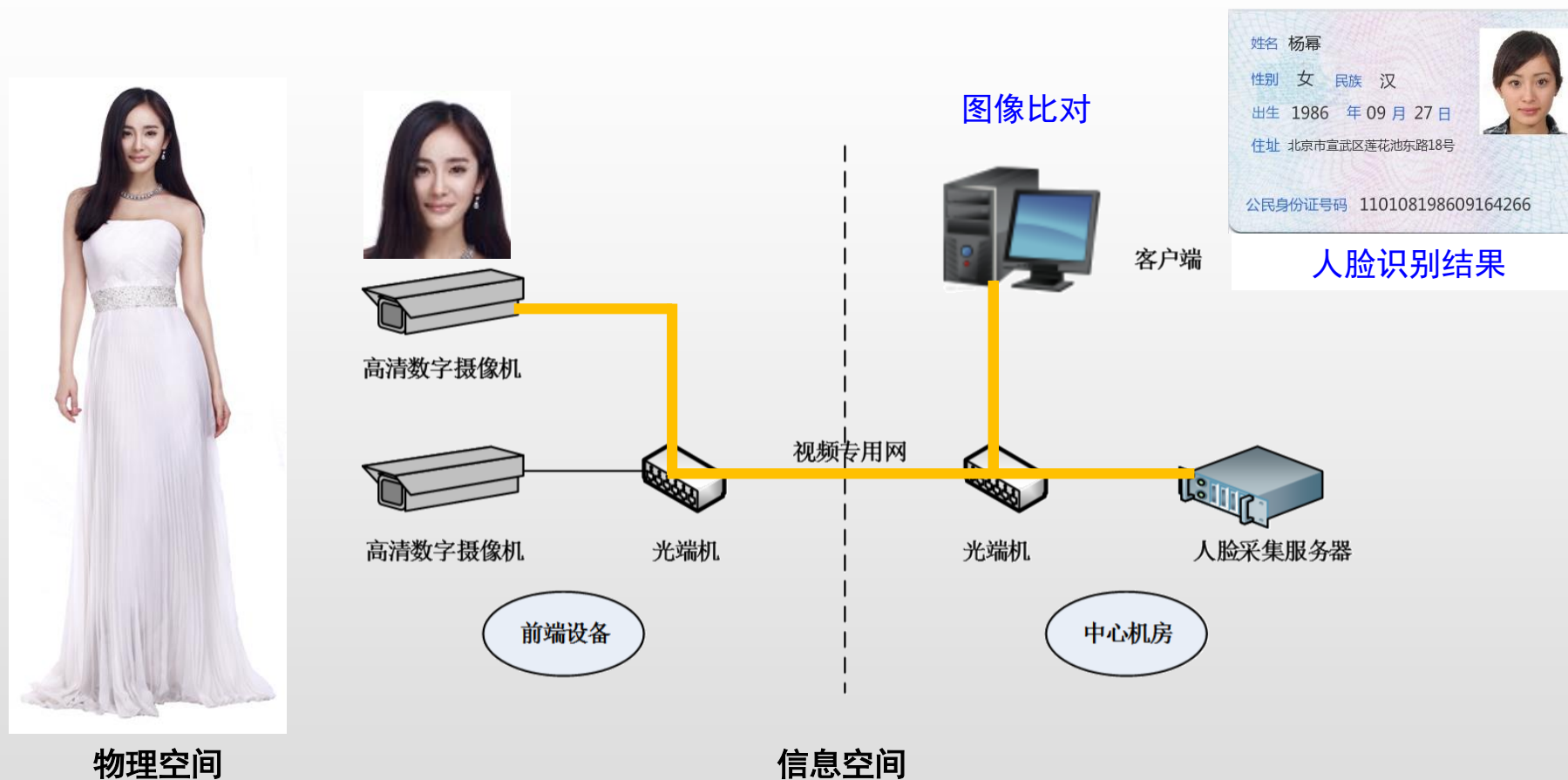
- **跨域图像**：通过**不同传感器**所感知到的**不同模态**的图像
- **图像跨域重建**：利用跨域图像之间的**内容关联性**和**表达互补性**，由一个域的图像生成另外一个域的图像的过程

汇报提纲

- 一. 前言 (Preface) : 什么是跨域图像重建?
- 二. 画像-照片合成 (Sketch-Photo Synthesis)
- 三. 图像超分辨重建 (Image Super-Resolution Reconstruction)
- 四. 其他相关应用 (Other Related Applications)
- 五. 结束语 (Conclusions)

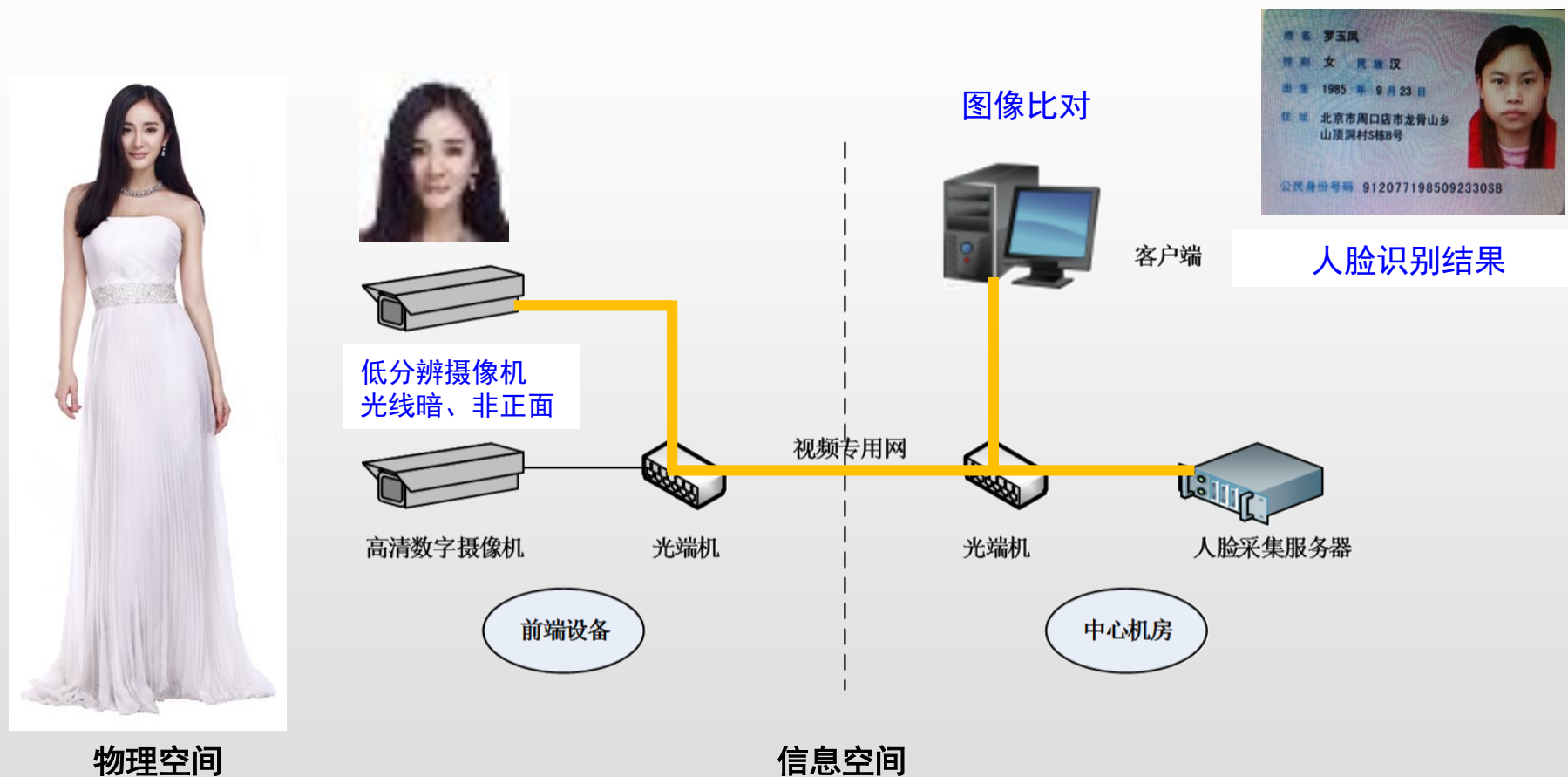


画像-照片合成：应用背景



传统的人脸识别系统（二元空间：物理-信息系统）

画像-照片合成：应用背景



传统的人脸识别系统（二元空间：物理-信息系统）

画像-照片合成：应用背景

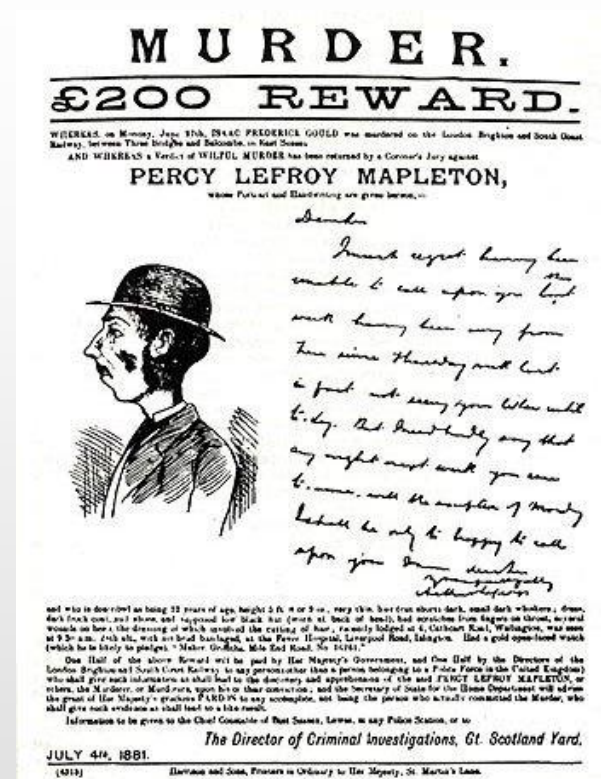


真实案件：2002年XX市公安局801案件

画像-照片合成：应用背景



Spring and Autumn and Warring States Period
(Historical Records, The biography of Wu Zixu)



Mapleton in 1881
The first *Police composite picture* to appear
on a 'wanted' poster and in a newspaper.

基于画像追凶的历史（国内 公元前484 vs. 欧洲 公元1881）

画像-照片合成：应用背景



Lois Jibson
Huston Police Dept., Texas

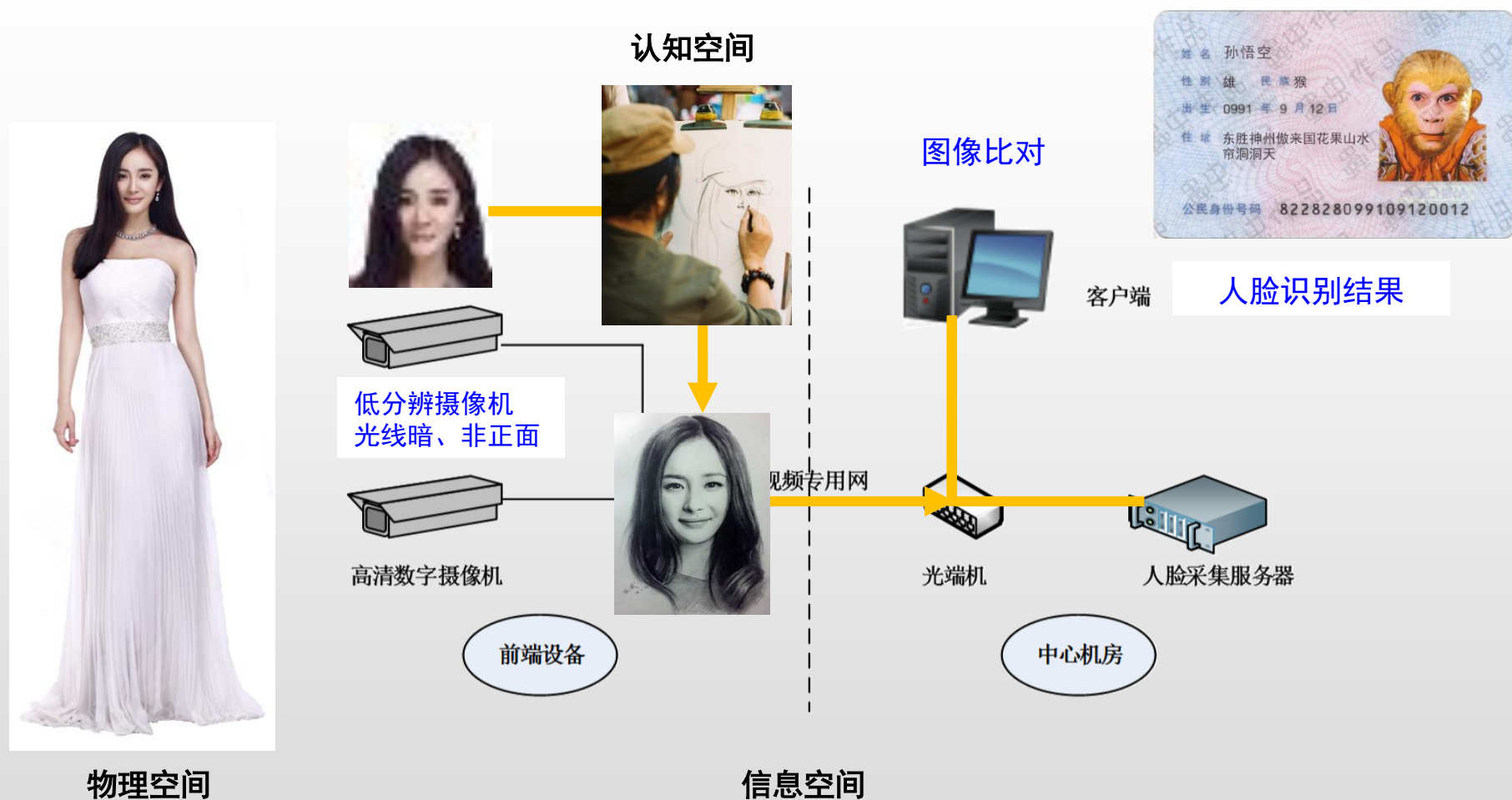


Jianhua Li
Shenzhen Police Dept.,
China



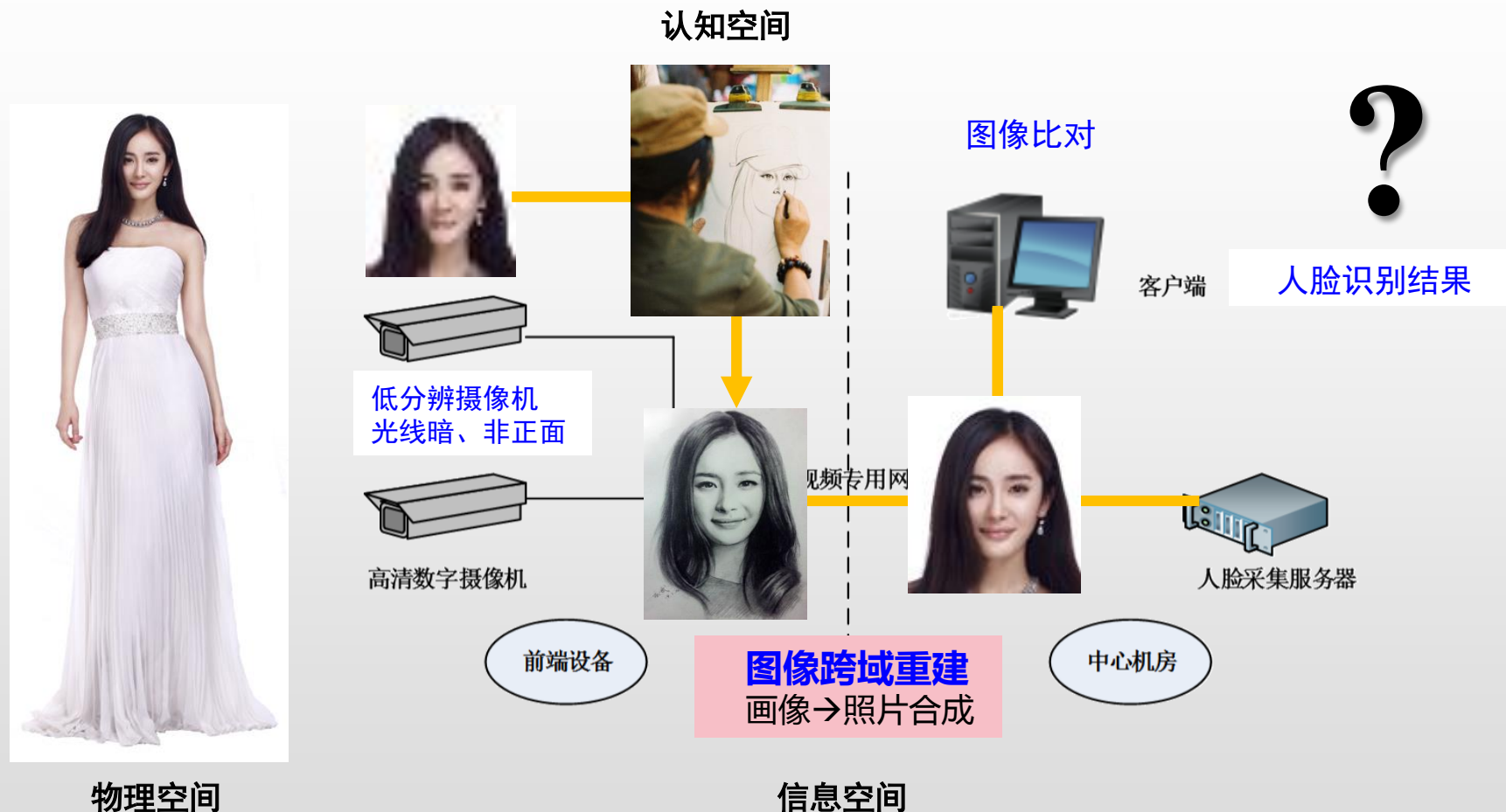
基于画像追凶的国内外成功案例

画像-照片合成：应用背景



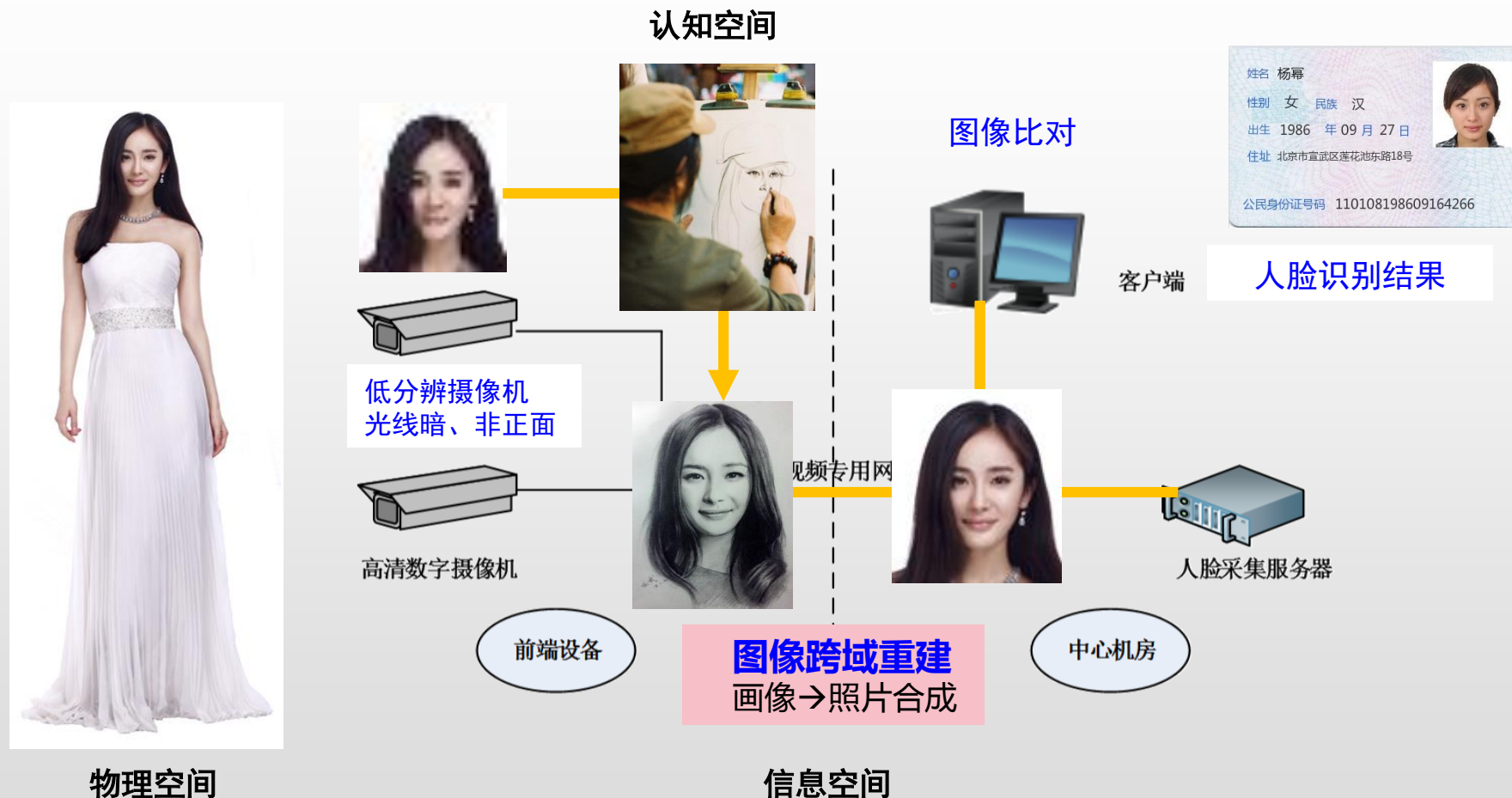
基于三元空间融合的人脸识别系统 (物理-信息-认知系统)

画像-照片合成：应用背景



基于三元空间融合的人脸识别系统的关键：**人机接口/交互**

画像-照片合成：应用背景



基于三元空间融合的人脸识别系统的关键：人机接口/交互

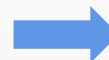
画像-照片合成：应用背景



北大硕士章莹颖
(UIUC)



低质量的监控视频画面



林宇辉模拟画像



协助确定嫌疑人身份



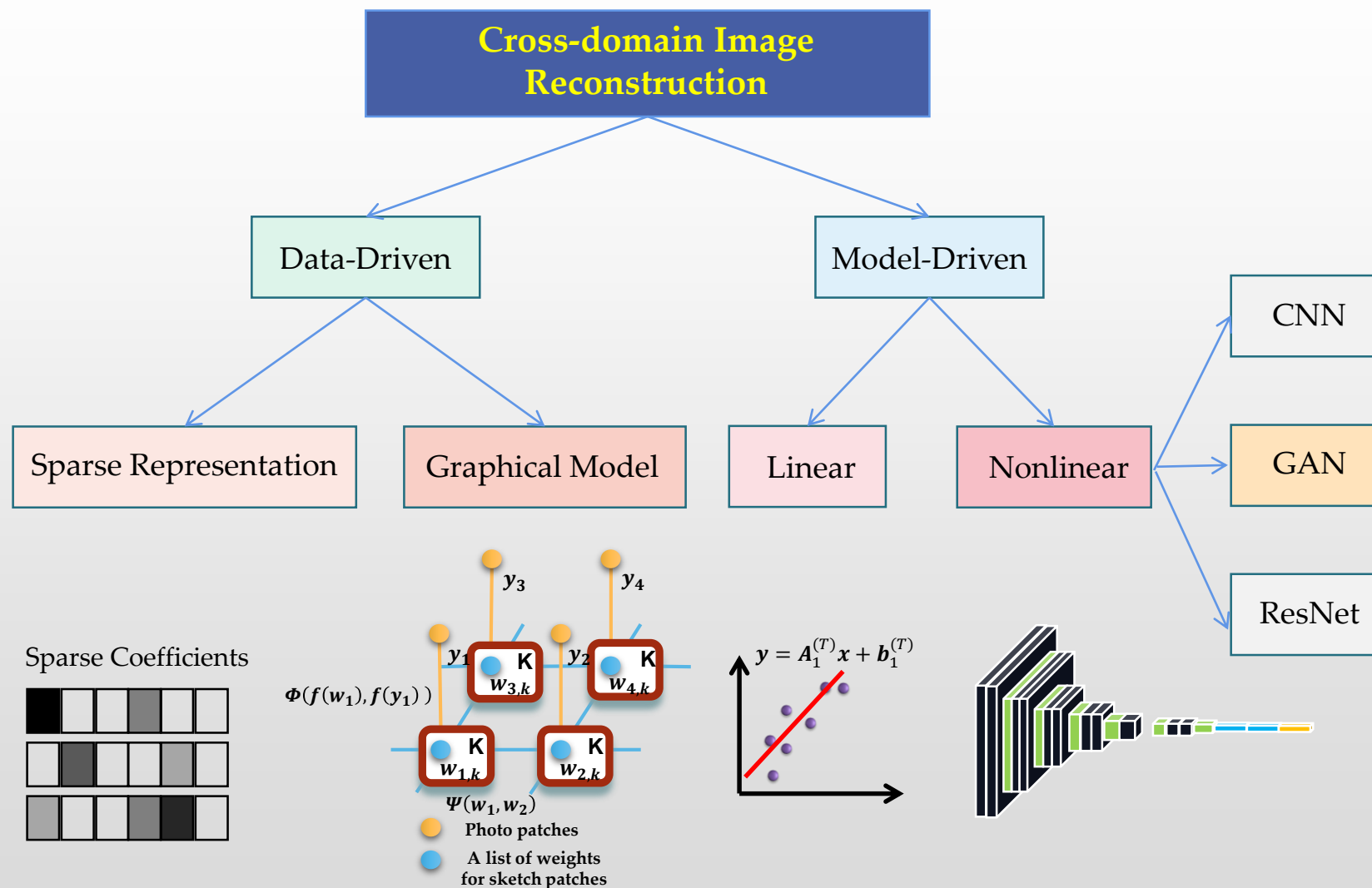
人脸识别



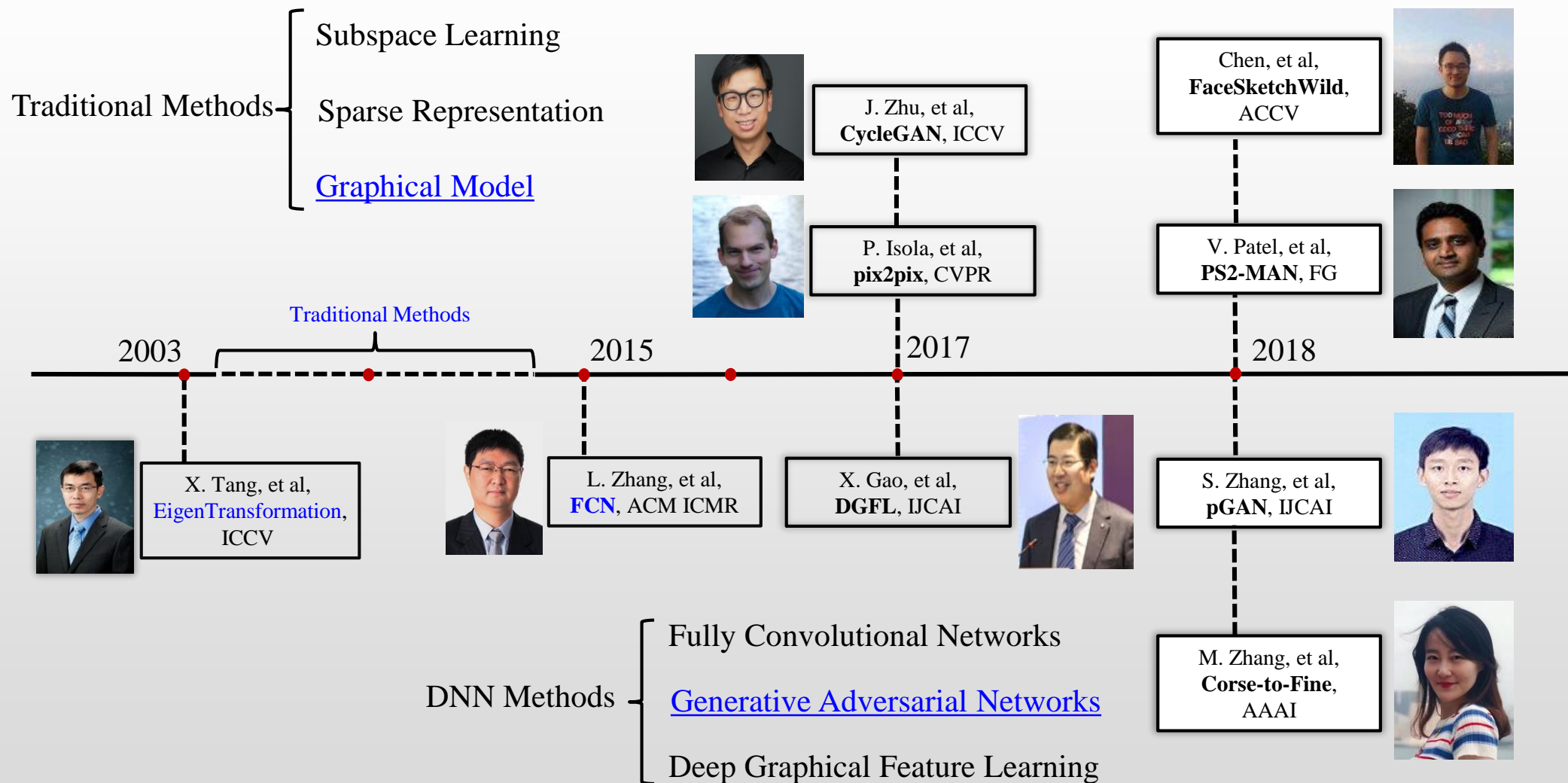
图像跨域重建：异质图像合成

基于三元空间融合的人脸识别系统的应用案例 (2017.6.9)

画像-照片合成：基本方法

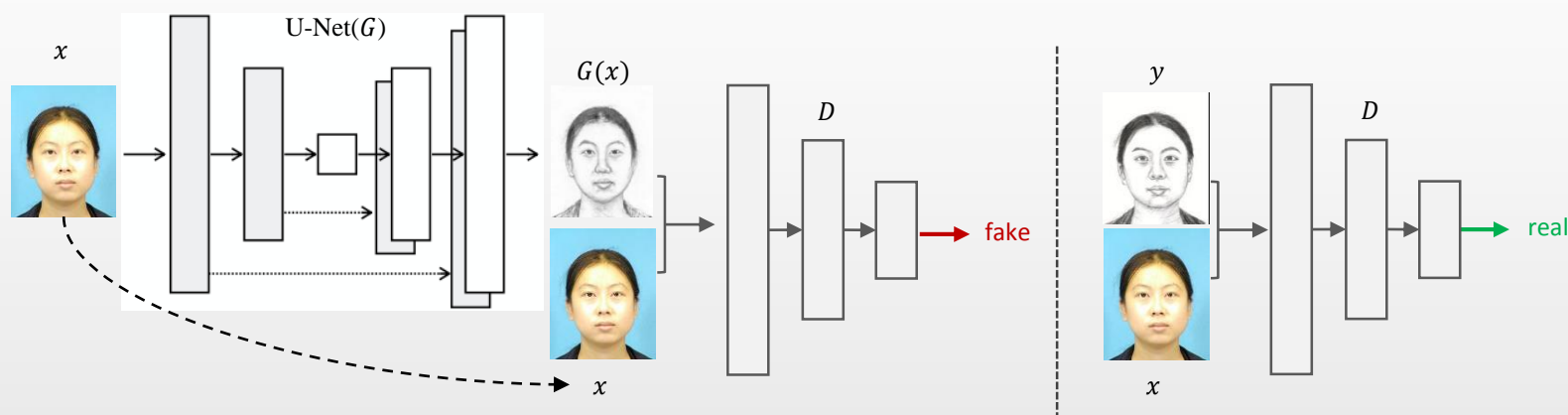


画像-照片合成：发展历程



画像-照片合成: pix2pix, conditional GAN

Pix2pix (conditional GAN) 图像到图像翻译模型: 对抗损失函数设计



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

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

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}

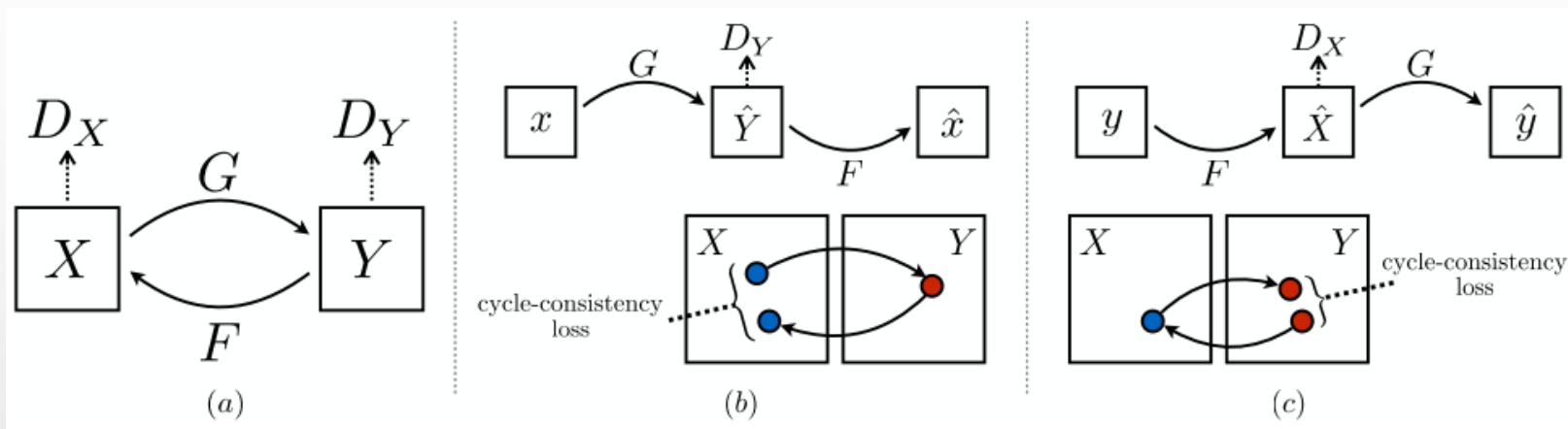
优势：相比于L1损失，对抗损失能使生成的图像更加逼真，与真实图像更加一致，分布更切合真是图像分布

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.**】

画像-照片合成: cycleGAN

CycleGAN 图像到图像翻译模型: 对抗损失+环一致损失



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

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

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

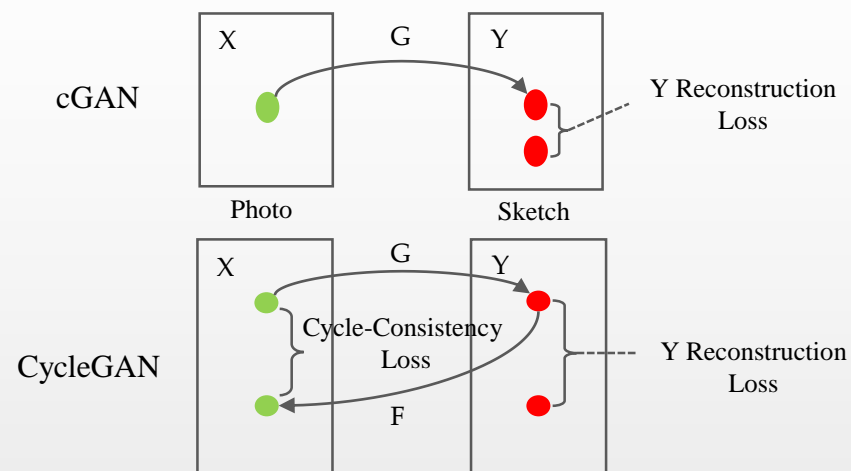
思想: 移除L1损失, 仅使用对抗损失约束两个域的生成图像, 同时引入环一致损失, 使网络只注重更加容易的变换 (如颜色, 纹理), 而较少改变内容 (形状, 物体)

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

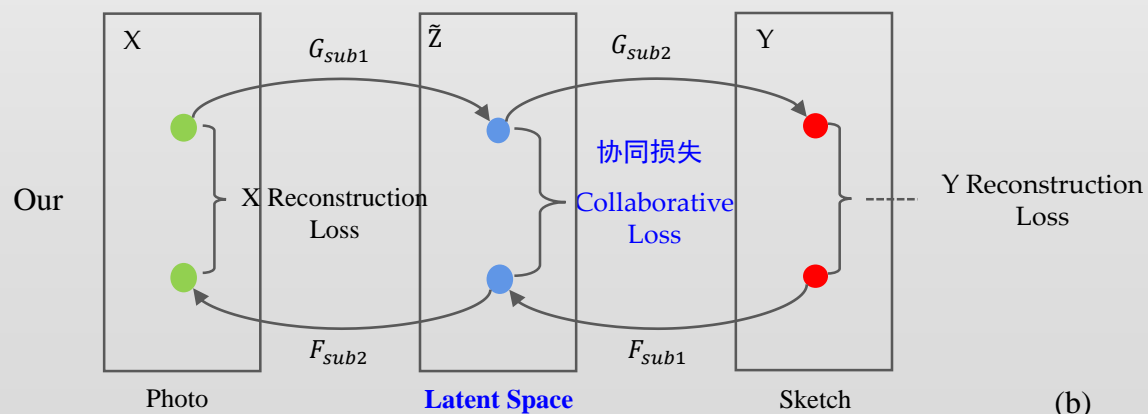
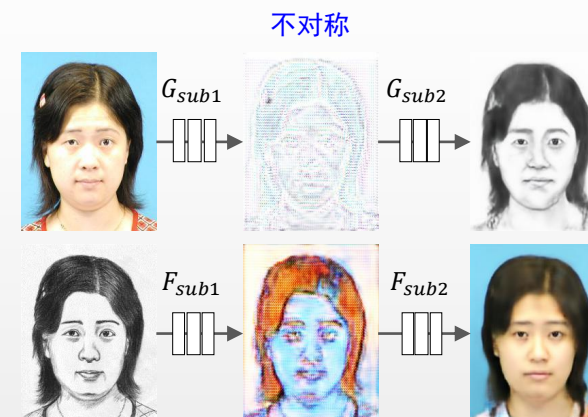
局限: 移除了L1损失, 仍有内容身份信息的丢失

①基于深度协同网络的人脸画像合成

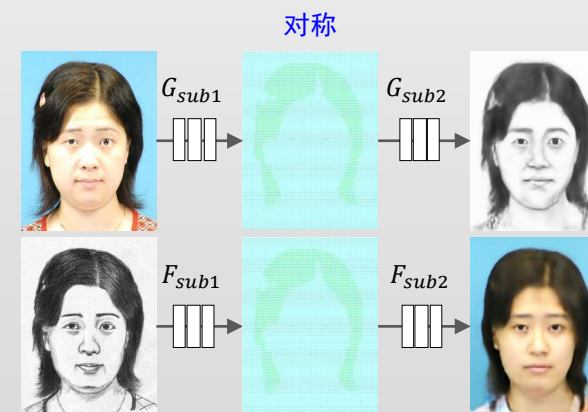
Motivation



(a)

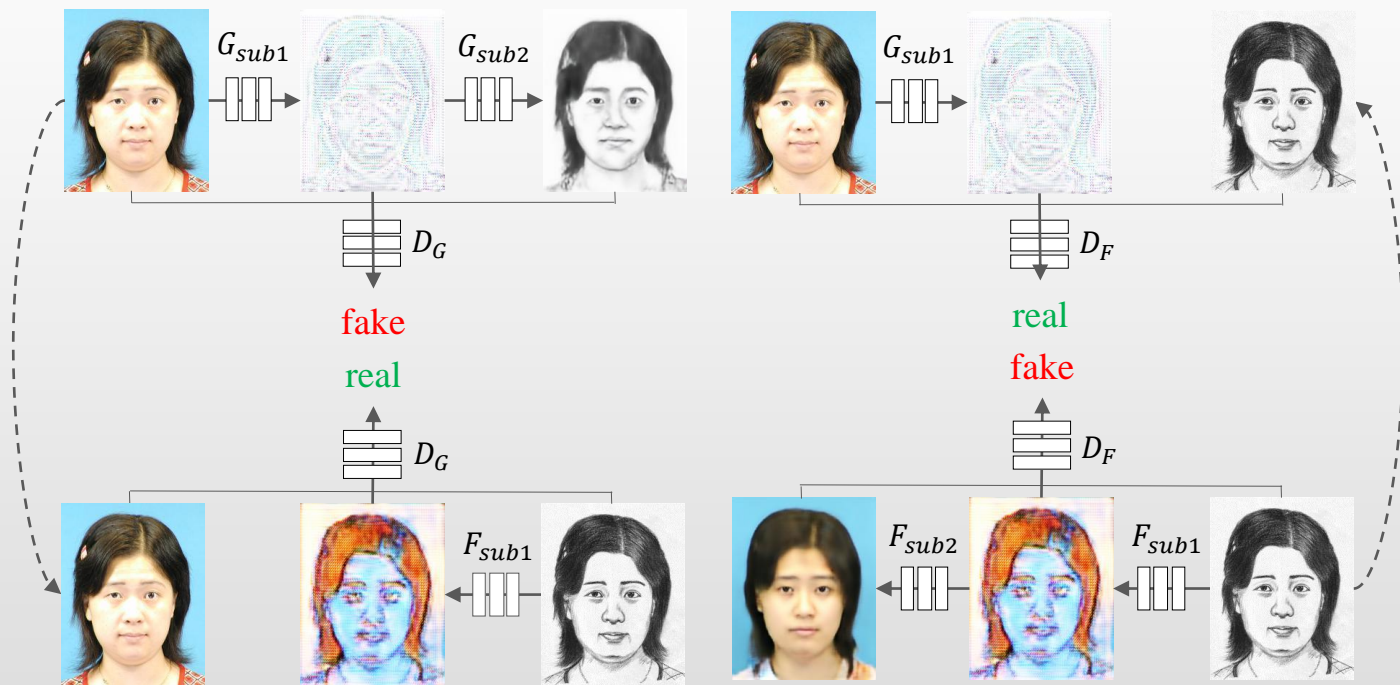


(b)



①基于深度协同网络的人脸画像合成

Deep Collaborative Framework: 协同对抗损失



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

协同对抗损失
重构损失

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

【Mingrui Zhu, Jie Li, Nannan Wang and Xinbo Gao. A Deep Collaborative Framework for Face Photo-Sketch Synthesis. **IEEE TNNLS 2019**】.

①基于深度协同网络的人脸画像合成

Comparisons

视觉效果好



pix2pix

CycleGAN

Our Method

Ground Truth

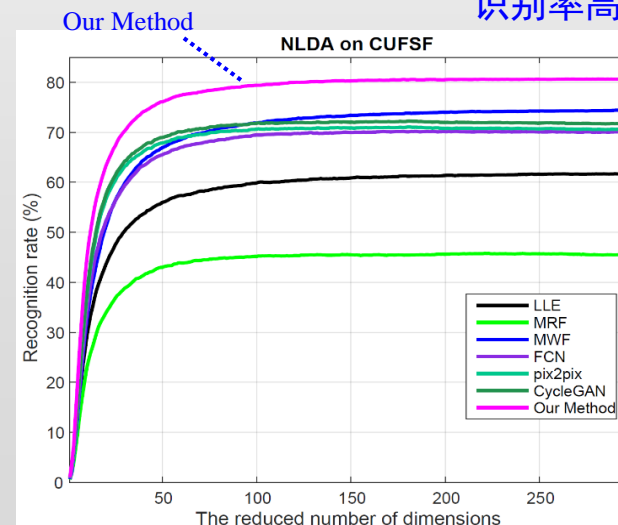
Average SSIM score (%) on the CUFS database and the CUFSF database

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

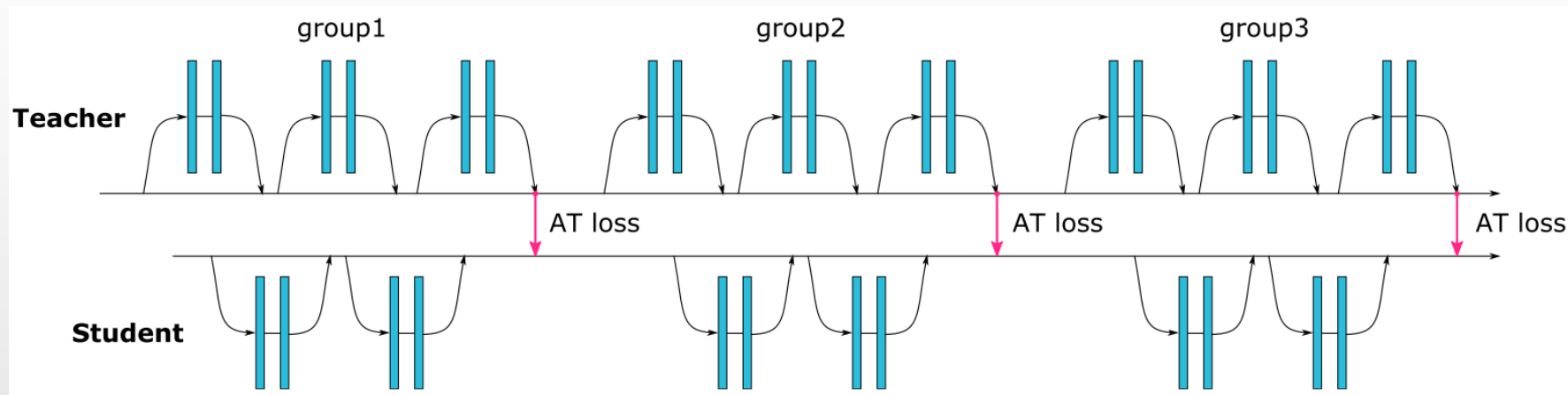
Face recognition accuracy against variations of the number of dimensions reduced by NLDA on the CUFSF database.

识别率高



画像-照片合成: Attention Transfer

Attention Transfer 知识迁移、知识蒸馏: 模型压缩、网络轻量化



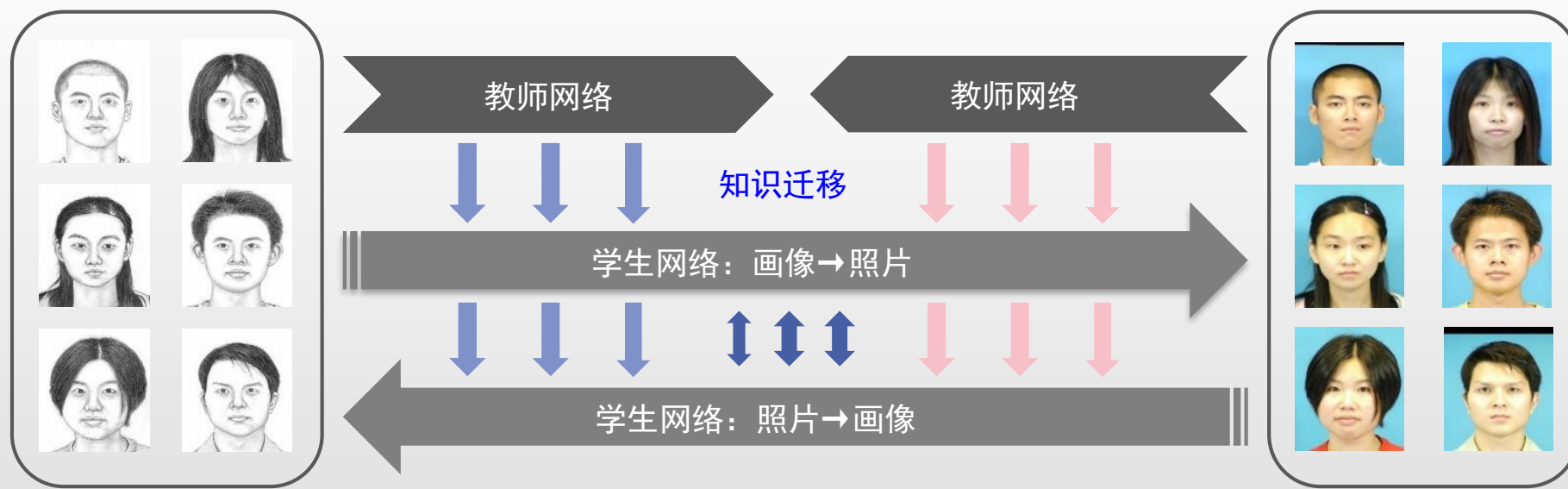
思想: 知识迁移, 也称作**知识蒸馏**, 是模型压缩的一种手段。这种方法利用大型网络中的知识, 来监督训练小型网络, 使得小型网络在具有与大型网络一致性能的同时, 减少网络的参数量。

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.**】

②基于知识迁移的人脸画像合成

Motivation

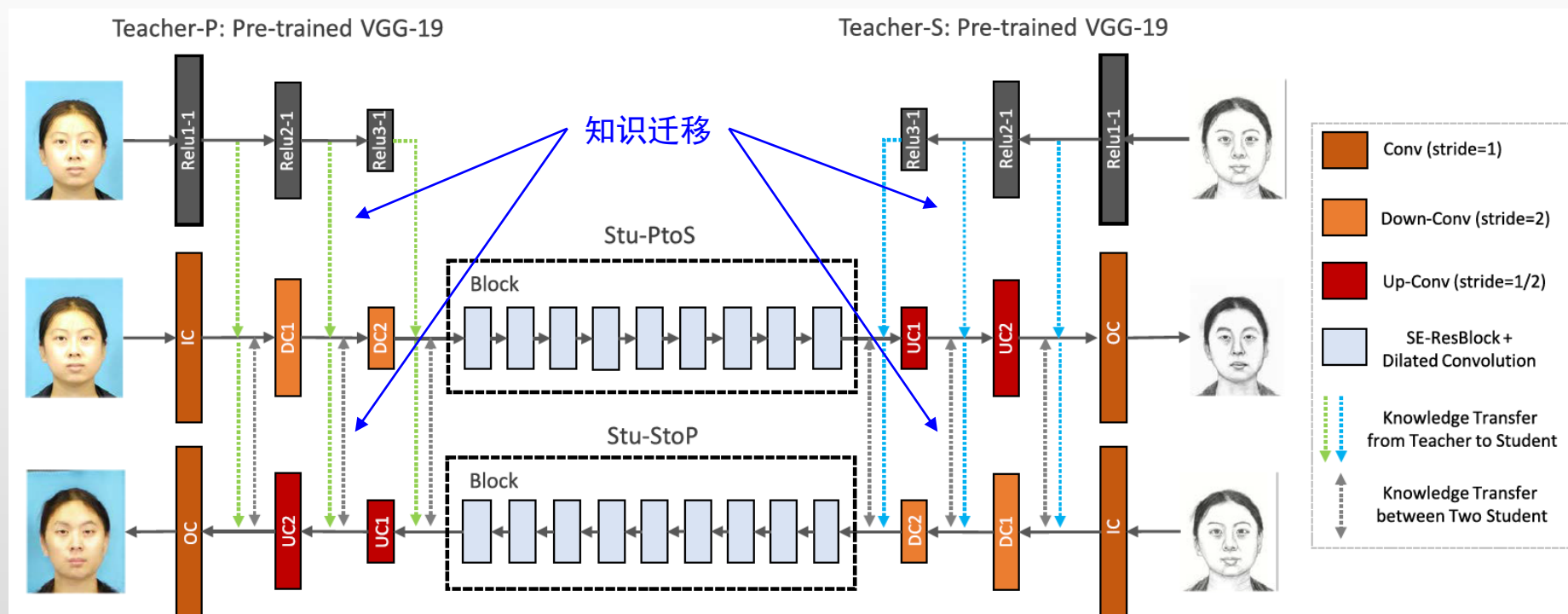


思想：解决训练样本不足造成的深层网络过拟合与浅层网络表征能力不足的矛盾

思路：提取与人脸照片画像合成相似，在充足训练数据上训练的任务中的知识，并迁移到人脸照片-画像合成任务，指导异质人脸图像的合成，学习更好的异质人脸图像间的映射关系，提升人脸照片-画像合成的效果。

②基于知识迁移的人脸画像合成

Knowledge Transfer Based Photo-Sketch Synthesis Method:



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

②基于知识迁移的人脸画像合成

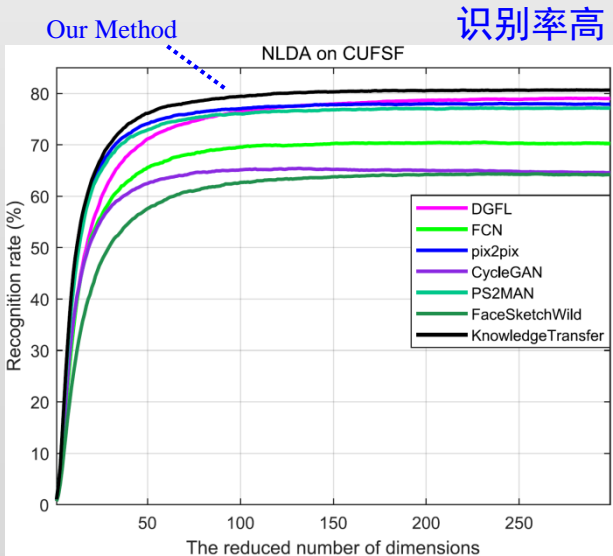
Comparisons



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

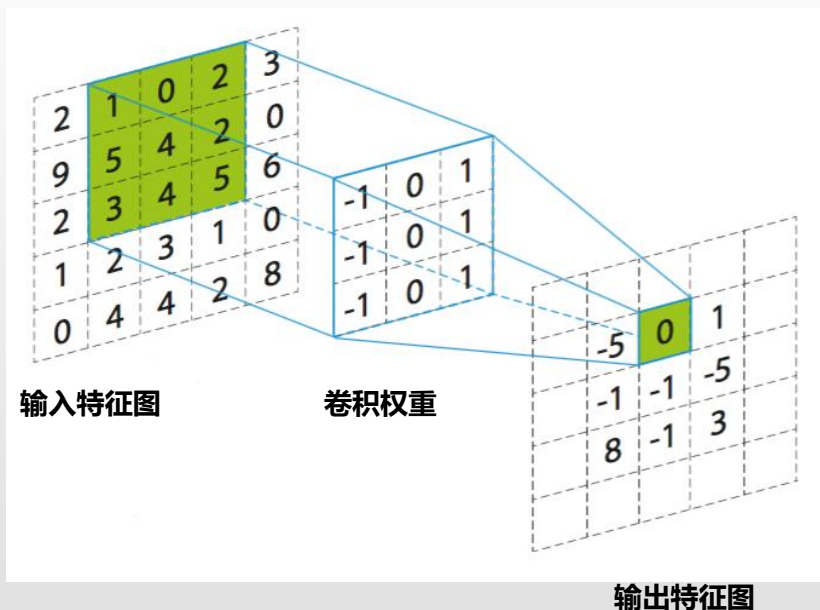
Face recognition accuracy against variations of the number of dimensions reduced by **NLDA** on the CUFSF database.



③基于Transformer的人脸画像合成

Motivation

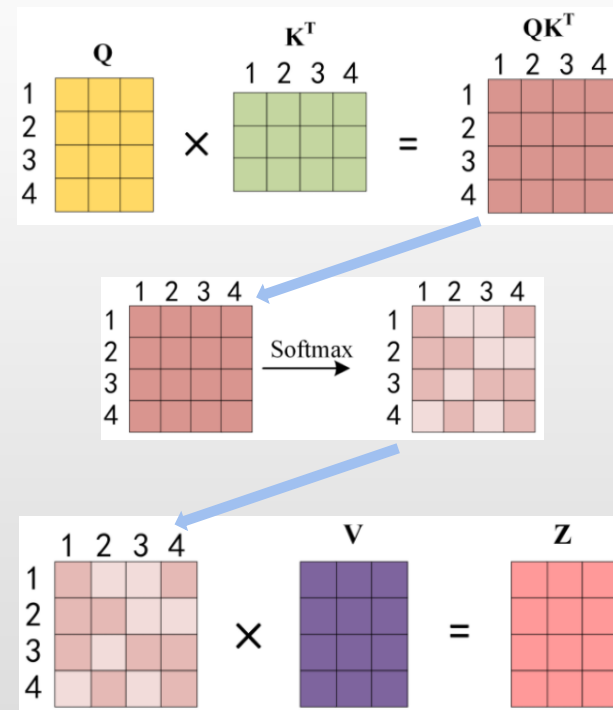
Convolution



特点：平移不变性、局部特性

缺陷：无法捕捉长范围空间相关性

Transformer: Self-attention

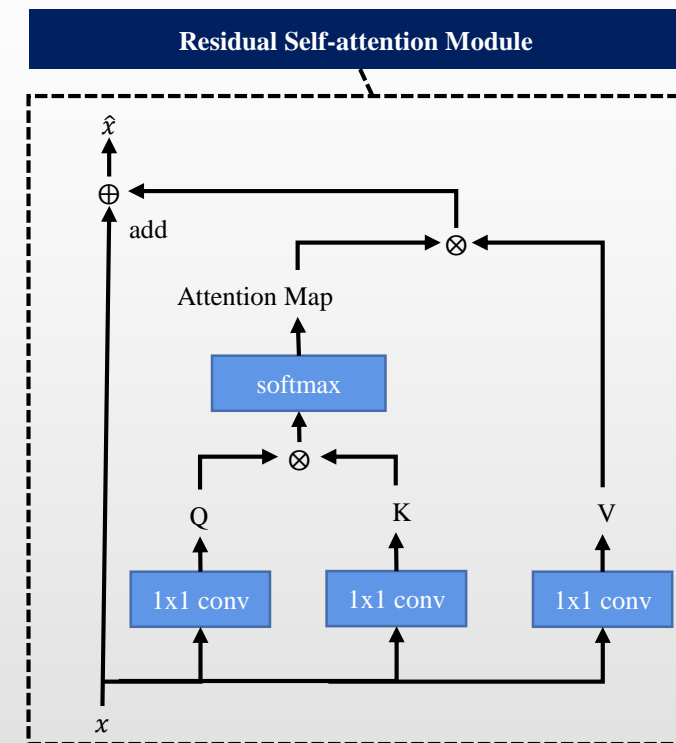
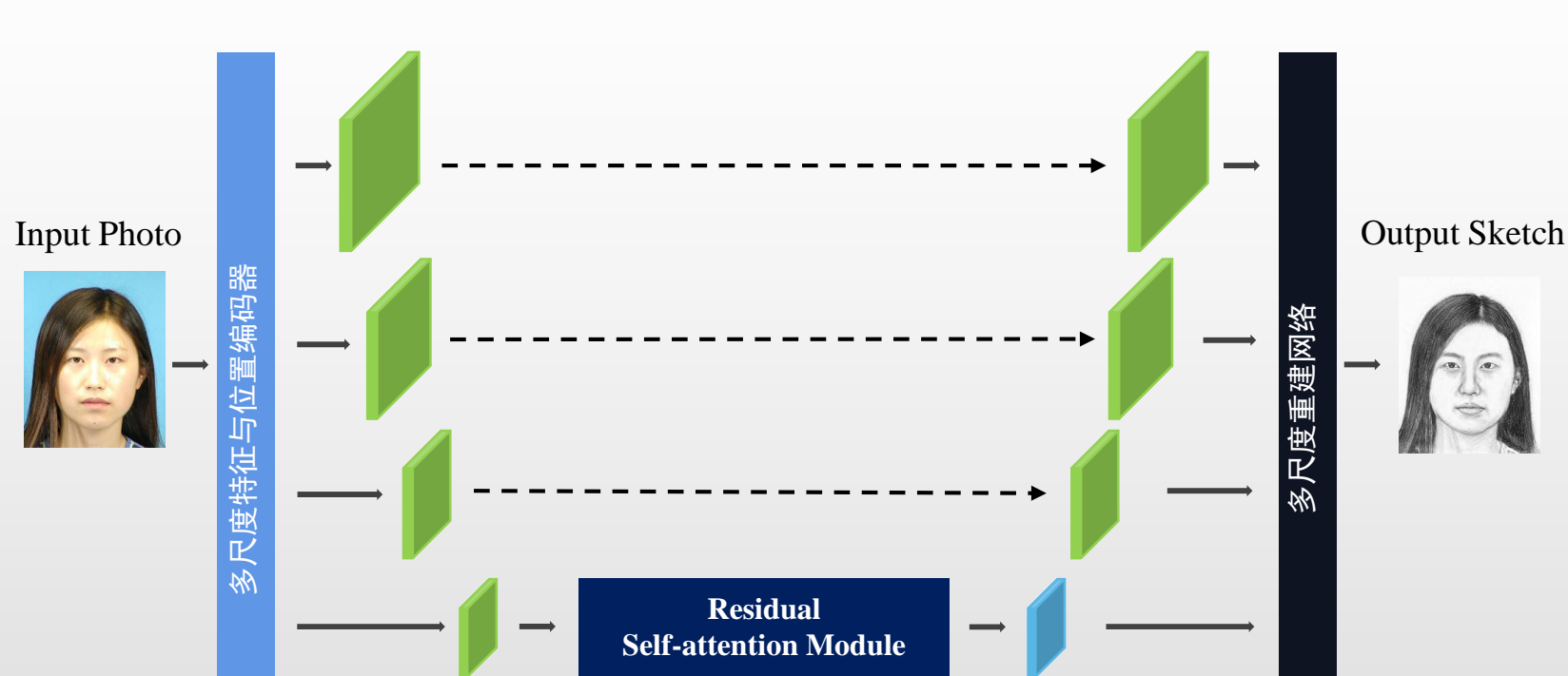


特点：捕捉长范围空间相关性

缺陷：计算量较大

③基于Transformer的人脸画像合成

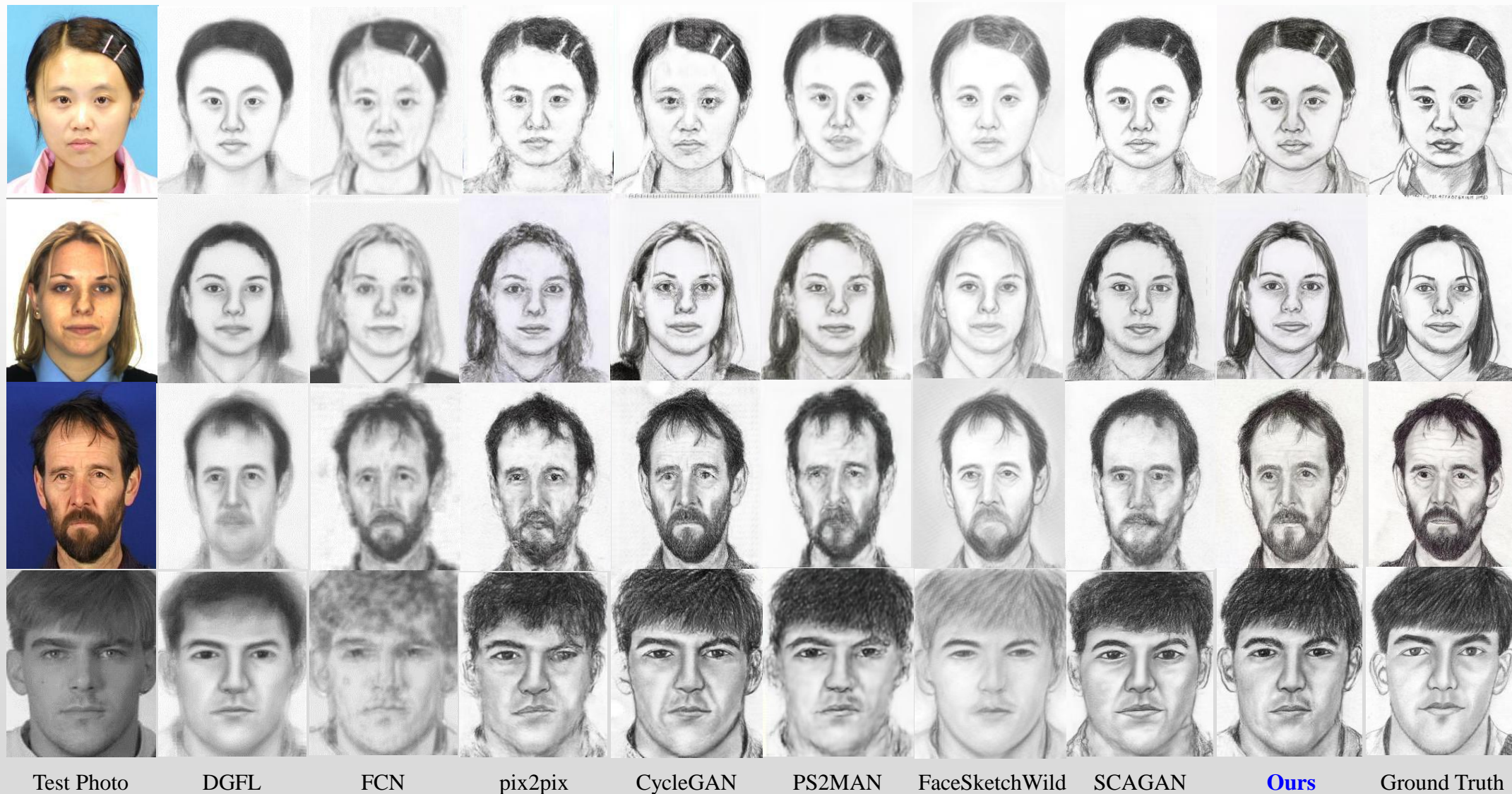
A Sketch-Transformer Network



- 利用**自注意力机制**捕捉不同位置特征向量的**长空间依赖关系**的能力，保持重建画像的空间结构
- 将人脸语义标签信息作为自注意力编码模块的**位置编码**，增强自注意力模块的**语义感知能力**

③基于Transformer的人脸画像合成

Qualitative Comparisons:

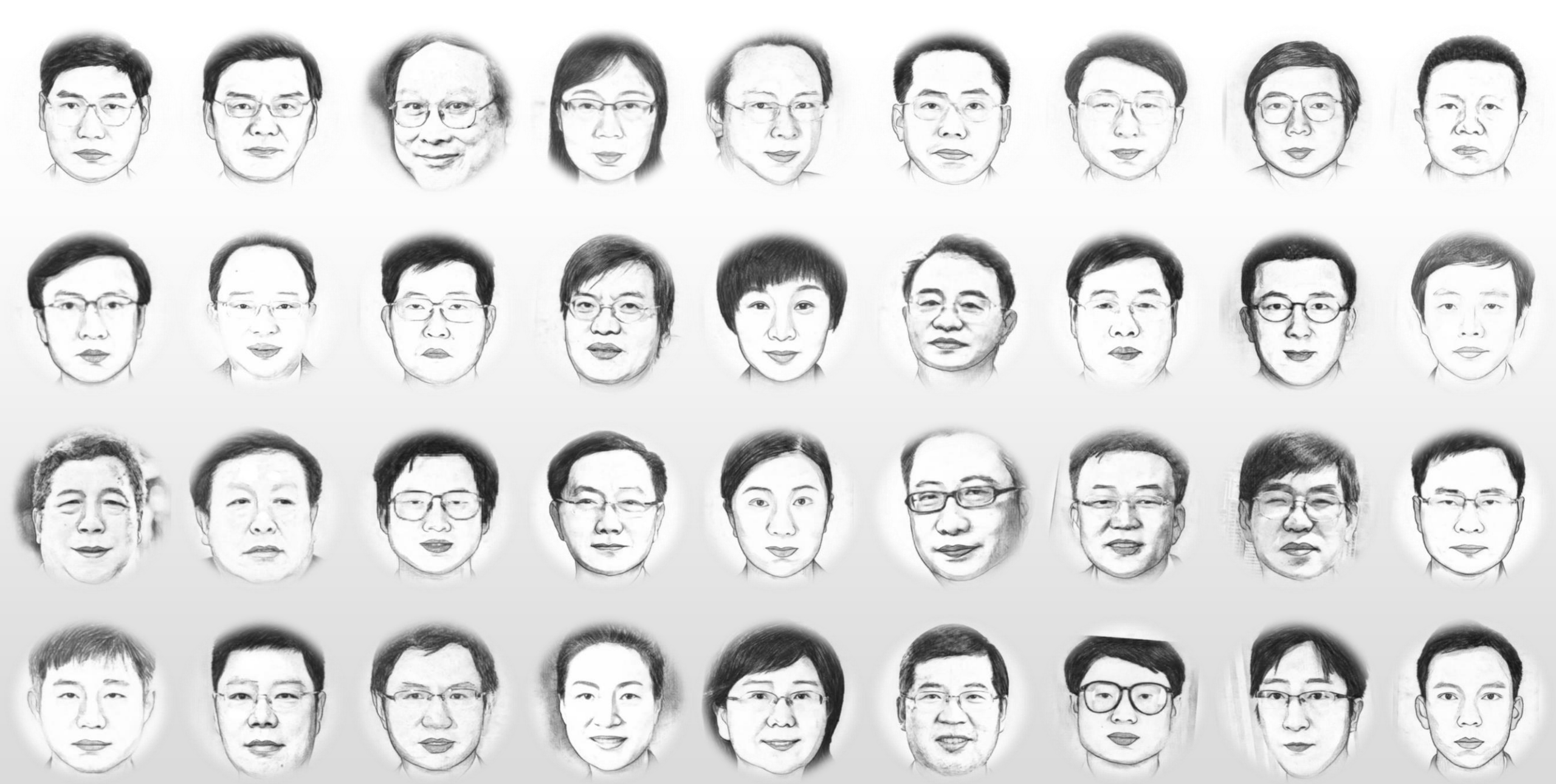


③基于Transformer的人脸画像合成

Quantitative Comparisons:

			DGFL	FCN	pix2pix	CycleGAN	PS2MAN	FaceSketchWild	SCAGAN	Sketch-Transformer
CUFS	Photo	LPIPS(alex) ↓	-	-	0.1993	0.2096	0.2464	-	0.1727	0.1538
		LPIPS(squeeze) ↓	-	-	0.1830	0.2094	0.2158	-	0.1643	0.1310
		LPIPS(vgg) ↓	-	-	0.3525	0.3882	0.3254	-	0.3053	0.2738
		FSIM ↑	-	-	0.7726	0.7450	0.7819	-	0.7937	0.7851
		FID ↓	-	-	73.56	80.44	65.04	-	80.53	27.88
	Sketch	LPIPS(alex) ↓	0.3316	0.4517	0.2263	0.2139	0.2961	0.2807	0.2408	0.1807
		LPIPS(squeeze) ↓	0.2635	0.3596	0.1552	0.1529	0.2265	0.2210	0.1722	0.1233
		LPIPS(vgg) ↓	0.3654	0.4350	0.3734	0.3598	0.3707	0.3639	0.3627	0.3019
		FSIM ↑	0.7079	0.6936	0.7363	0.7219	0.7230	0.7114	0.7086	0.7350
		FID ↓	70.81	69.93	44.91	23.76	48.95	59.26	38.61	20.92
CUFSF	Photo	LPIPS(alex) ↓	-	-	0.2463	0.2557	0.3145	-	0.1735	0.2199
		LPIPS(squeeze) ↓	-	-	0.2005	0.2002	0.2853	-	0.1469	0.1714
		LPIPS(vgg) ↓	-	-	0.4019	0.3791	0.4237	-	0.3128	0.3474
		FSIM ↑	-	-	0.7777	0.7645	0.7812	-	0.8395	0.7861
		FID ↓	-	-	39.82	14.46	78.03	-	18.84	15.22
	Sketch	LPIPS(alex) ↓	0.3524	0.4793	0.2408	0.2371	0.3288	0.3288	0.2188	0.1971
		LPIPS(squeeze) ↓	0.2794	0.3895	0.1628	0.1589	0.2397	0.2473	0.1500	0.1349
		LPIPS(vgg) ↓	0.3972	0.5305	0.3824	0.3744	0.4170	0.4053	0.3536	0.3400
		FSIM ↑	0.6957	0.6624	0.7283	0.7088	0.7233	0.6821	0.7270	0.7259
		FID ↓	57.33	124.40	35.52	14.62	64.42	59.76	18.32	9.39





部分人脸照片-画像合成结果展示, Part of Photo-Sketch Synthesis Results



部分人脸照片-画像合成结果展示, Part of Photo-Sketch Synthesis Results



肖像

合成剪纸

真实剪纸

肖像

合成剪纸

真实剪纸

[部分人脸照片-剪纸合成结果展示](#) , [Part of Photo-Paper-Cutting Synthesis Results](#)

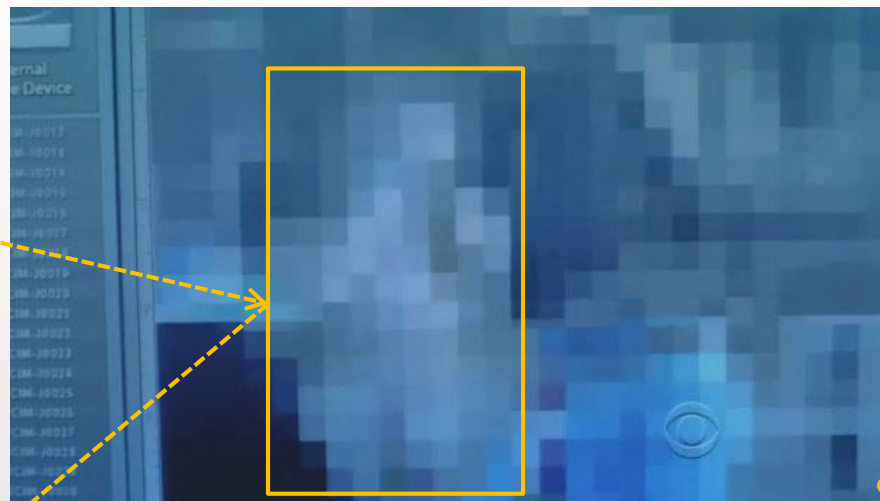
汇报提纲

- 一. 前言 (Preface) : 什么是跨域图像重建?
- 二. 画像-照片合成 (Sketch-Photo Synthesis)
- 三. **图像超分辨重建 (Image Super-Resolution Reconstruction)**
- 四. 其他相关应用 (Other Related Applications)
- 五. 结束语 (Conclusions)

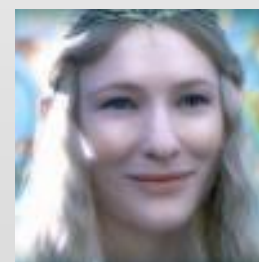
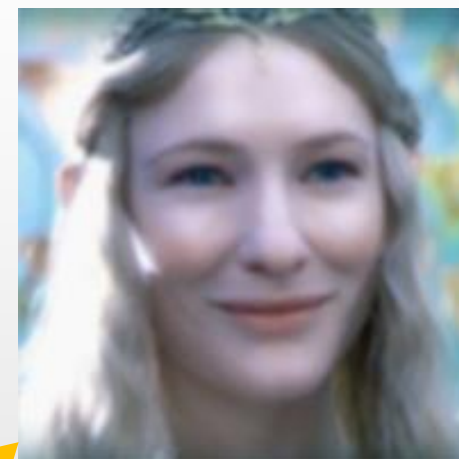


图像超分辨重建：应用背景

为什么要进行图像超分辨重建？



Interpolation– blurred!



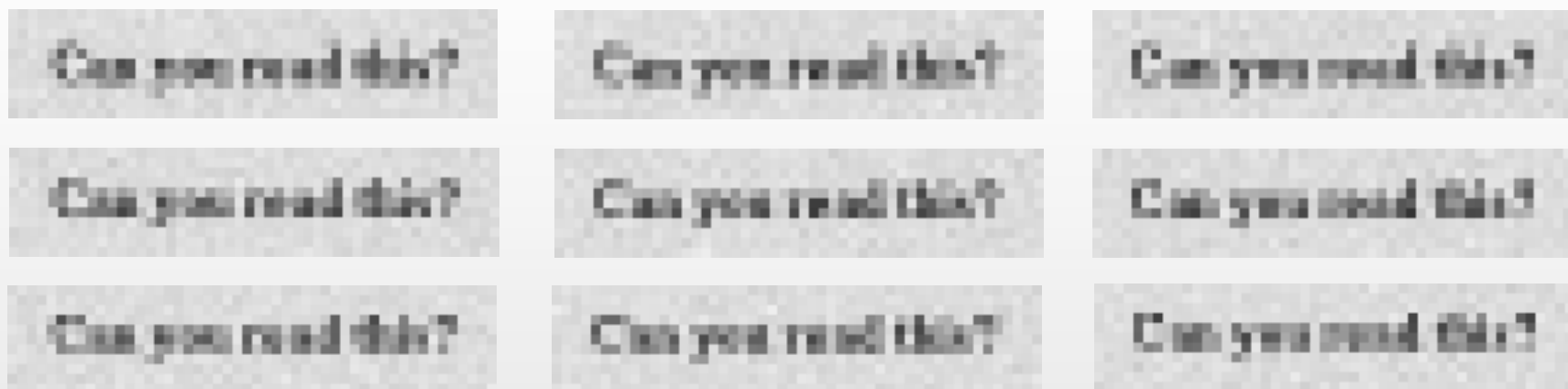
LR image



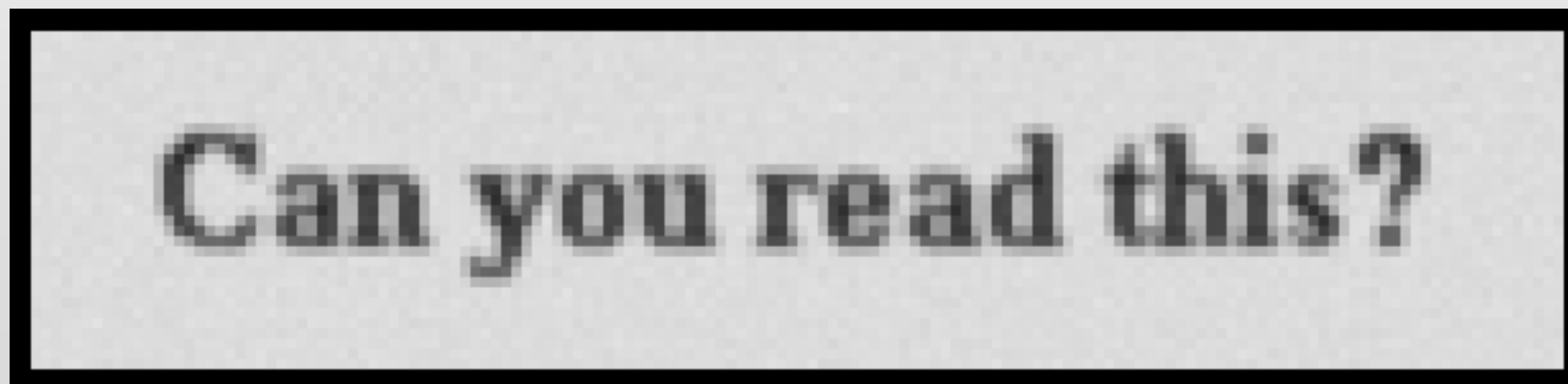
True HR image

图像超分辨重建：应用背景

Multiple Frame Image Super-Resolution Reconstruction



3:1 scale-up in each axis using 9 images, with pure global translation between them



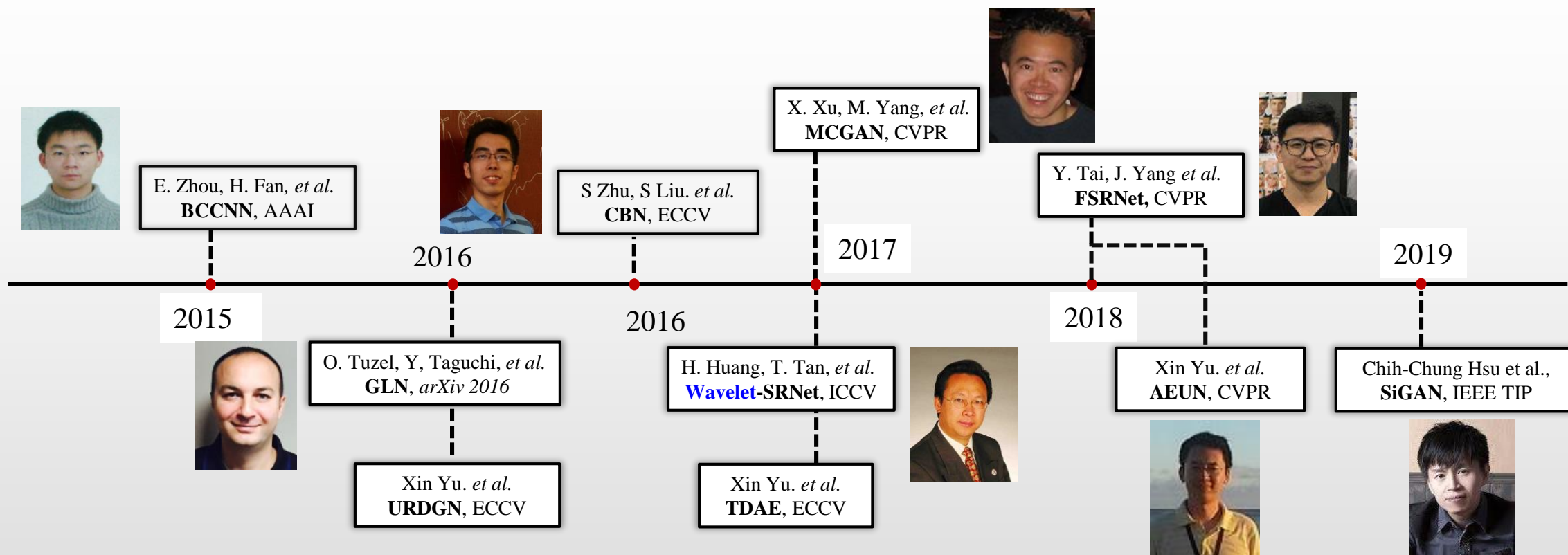
图像超分辨重建：应用背景

Ex. Surveillance: License plate recognition



LPR, Reconnaissance Crimes...

图像超分辨重建：发展历程



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

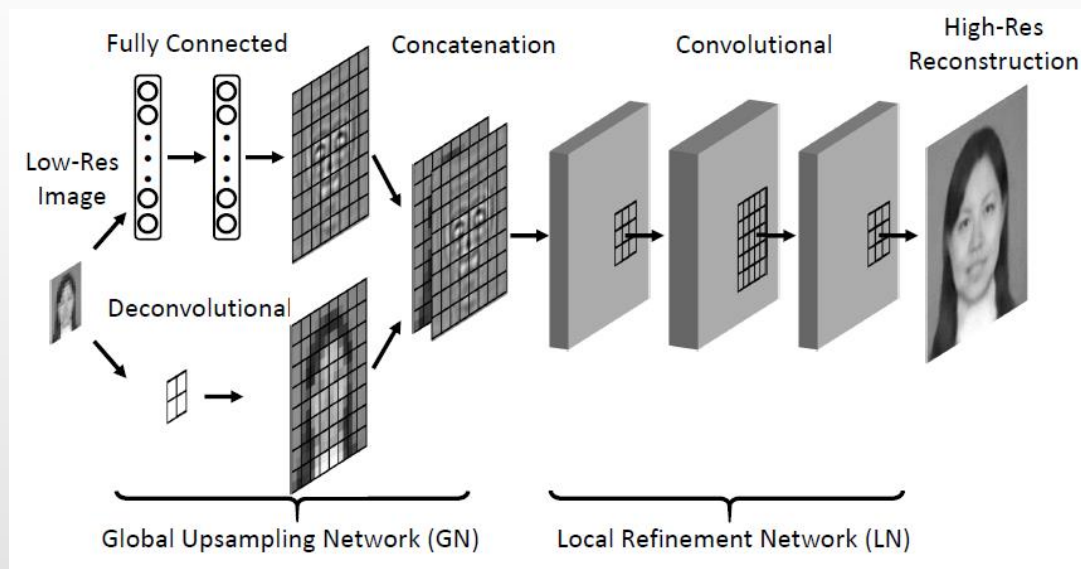
FSRNet: Face Super-Resolution Network

AEUN: Attribute Embedded Upsampling Network

SiGAN: Siamese GAN

图像超分辨重建：相关工作-GLU Network

- 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



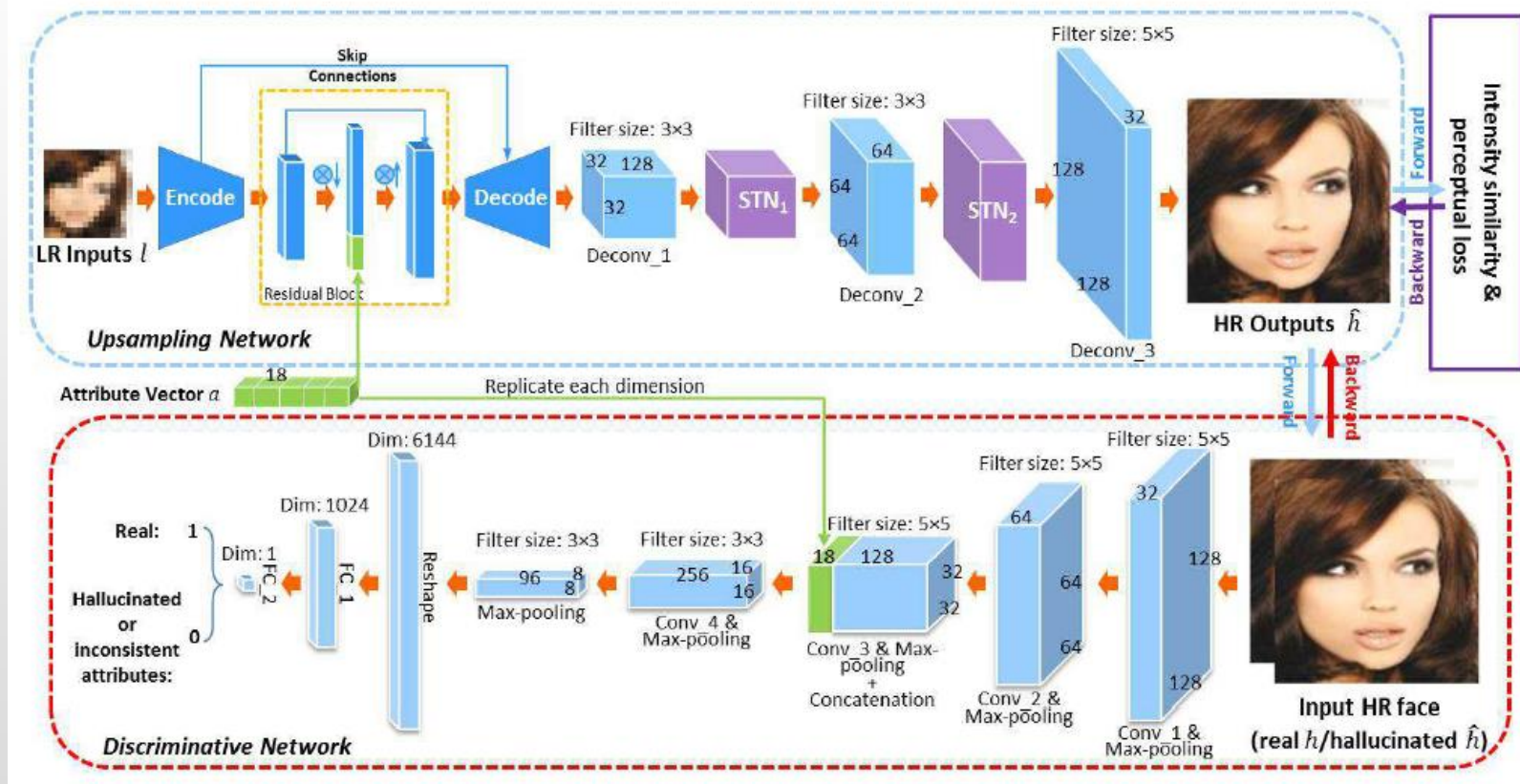
4 × GN		8 × GN	
deconv4	fc-512 fc-256 fc-512 fc-(128 × 128)	deconv8	fc-256 fc-256 fc-256 fc-(128 × 128)
concatenation		concatenation	
4 Layer LN (LN4)		6 Layer LN (LN6)	8 Layer LN (LN8)
conv5-16 conv7-64 conv5-16		conv5-16 conv7-32 conv7-64 conv7-32 conv5-16	conv5-16 conv7-32 conv7-64 conv7-64 conv7-64 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未考虑人脸的先验信息，人脸的高频区域重建质量较低

图像超分辨重建：相关工作-属性嵌入

- 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



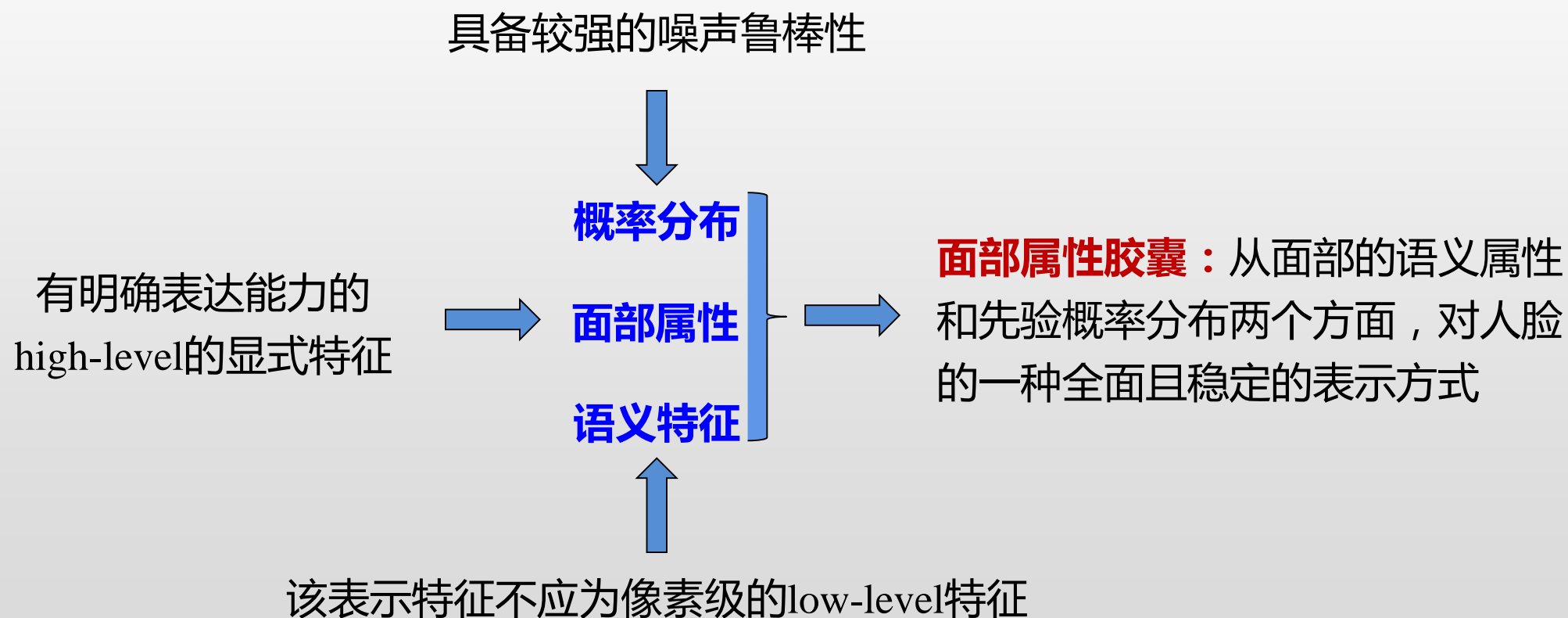
缺陷：

- 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。

Yu X, Fernando B, Hartley R, et al. Super-resolving *very low-resolution* face images with supplementary attributes [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR2018). 2018: 908-917.

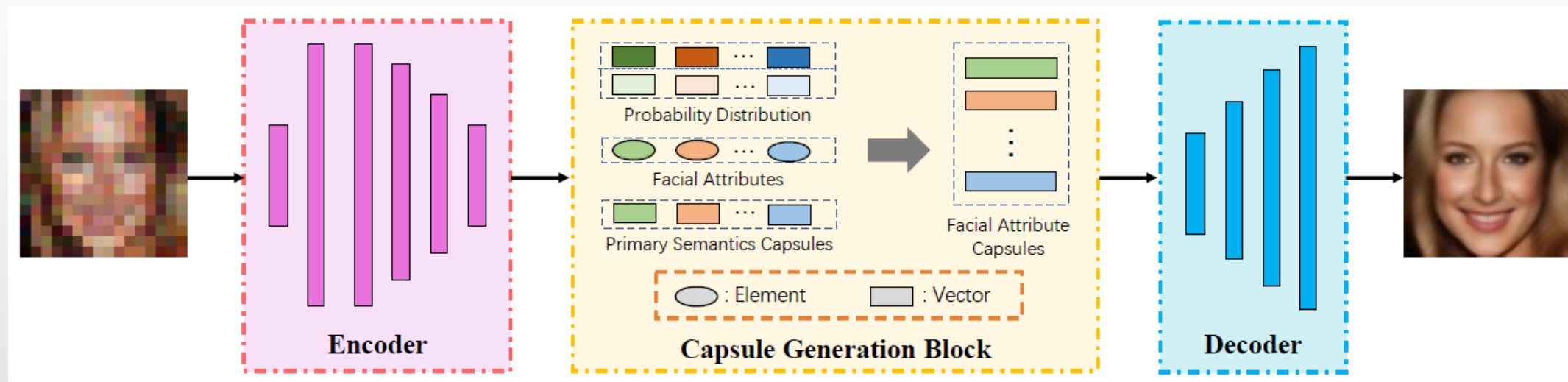
噪声人脸图像超分辨重建：Motivation

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**？



噪声人脸图像超分辨重建：方法过程

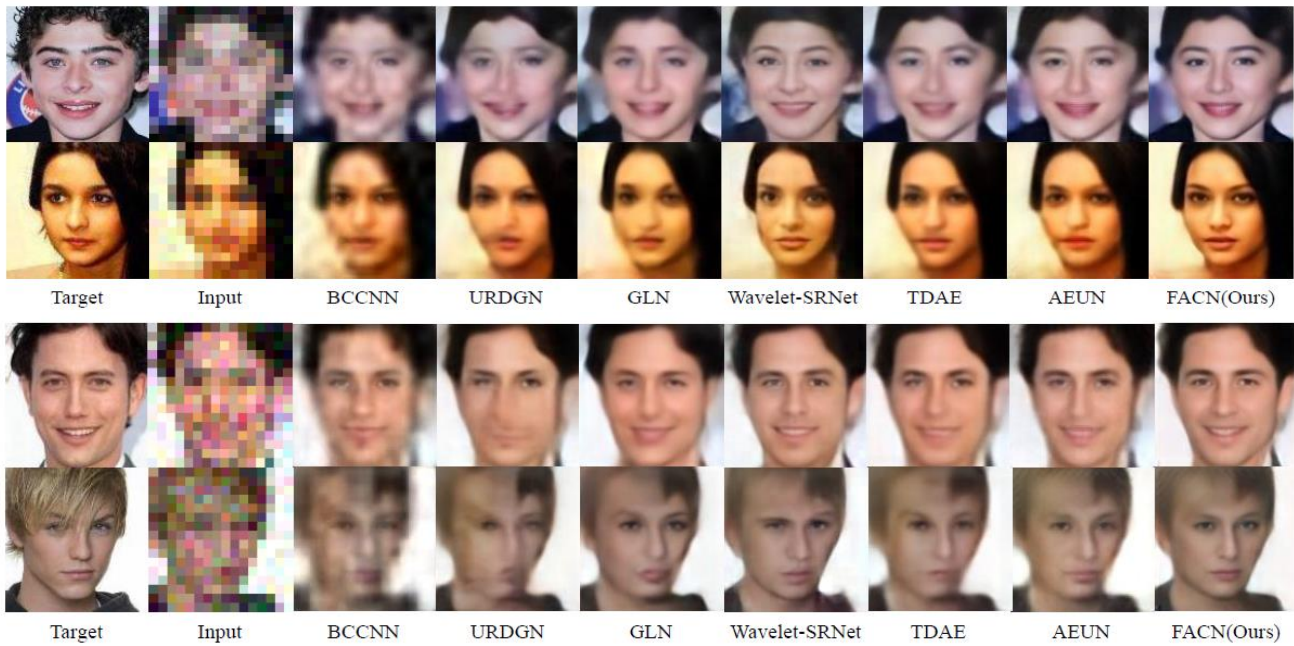
- **基本思想**：引入了胶囊网络的概念进行人脸超分辨，以人脸属性和先验概率分布相结合的方式构建出一种新型的胶囊结构，对于噪声和模糊具有较强鲁棒性



- 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

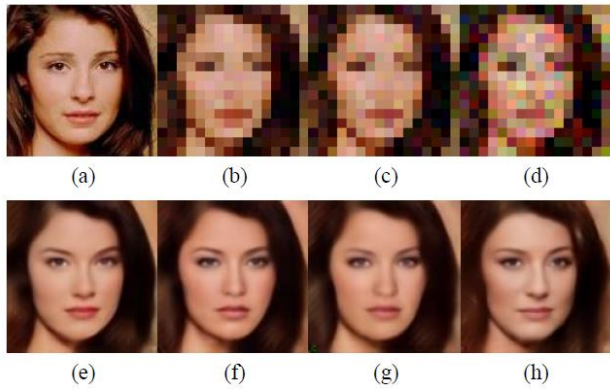
噪声人脸图像超分辨重建：实验结果

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



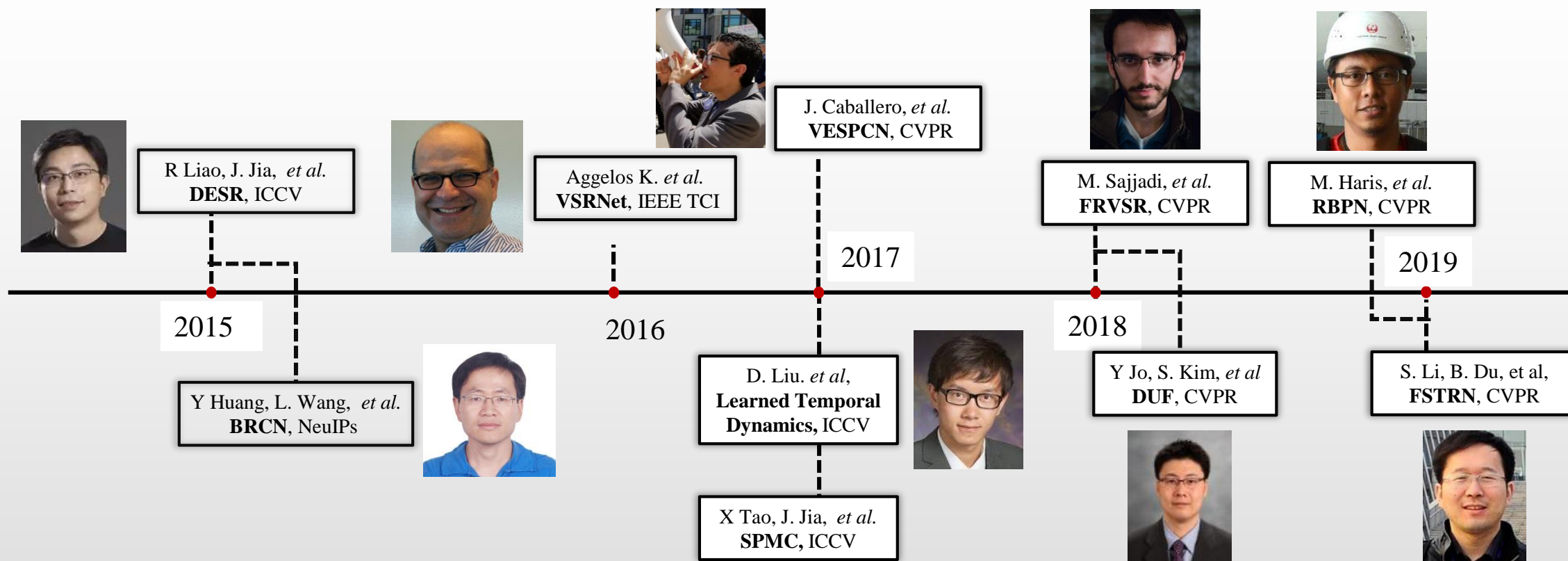
Methods	Performance	Methods	Performance
Bicubic	0.8058	BCCNN	0.8570
URDGN	0.8212	GLN	0.8580
Wavelet-SRNet	0.8820	TDAE	0.8680
AEUN	0.8694	FACN(Ours)	0.8922

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**。

视频超分辨重建：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

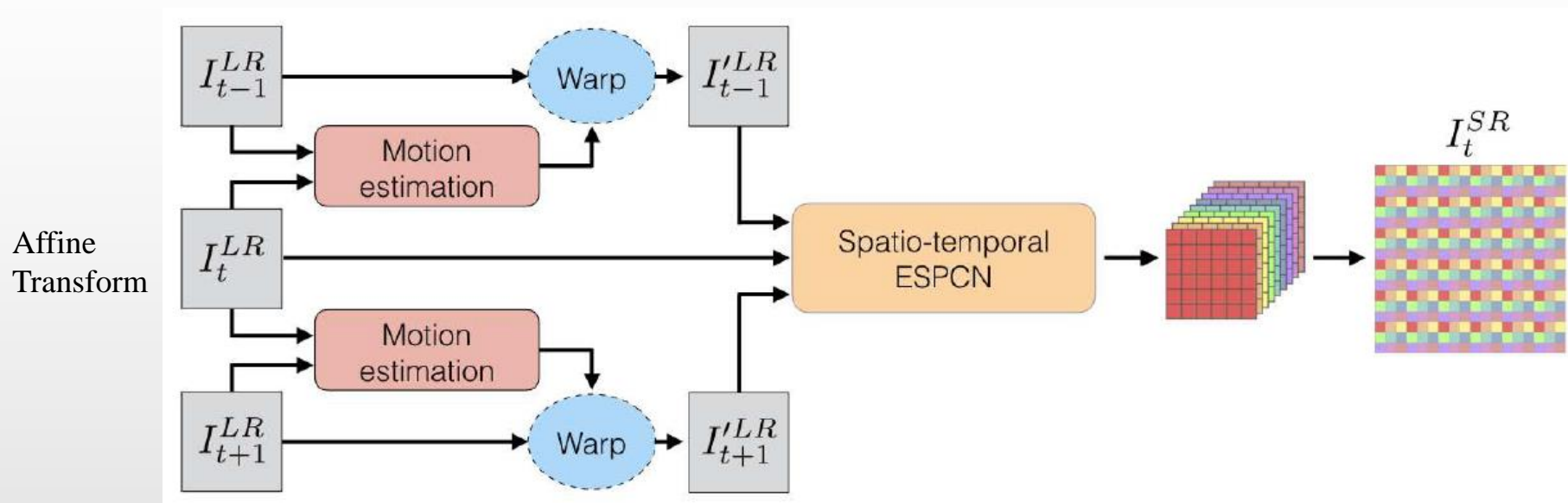
FSTRN: Fast Spatio-Temporal Residual Network

DUF: Dynamic Upsampling Filters

RBPN: Recurrent Back-Projection Network

视频超分辨重建：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



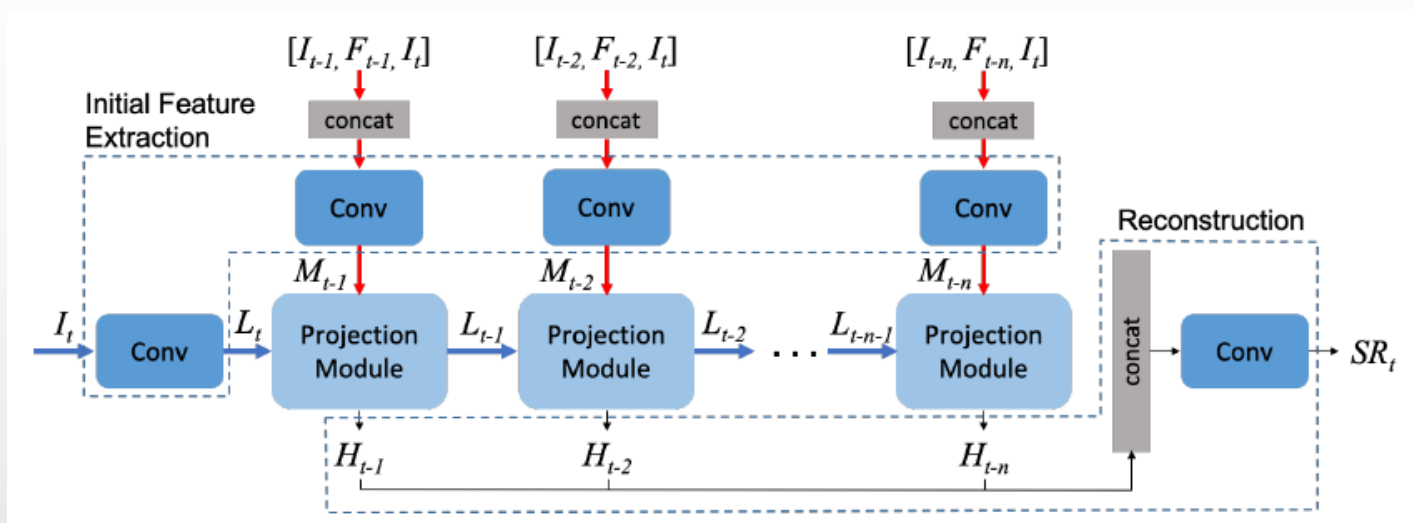
缺陷：

- 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. CVPR 2017: 4778-4787.】

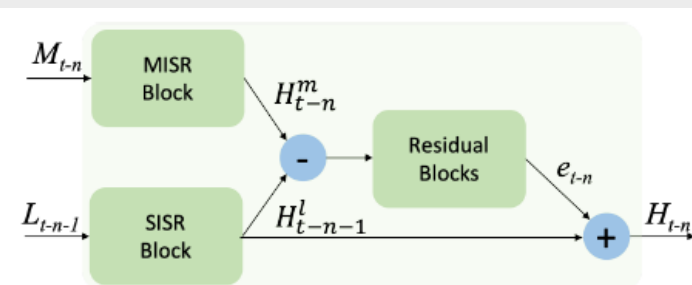
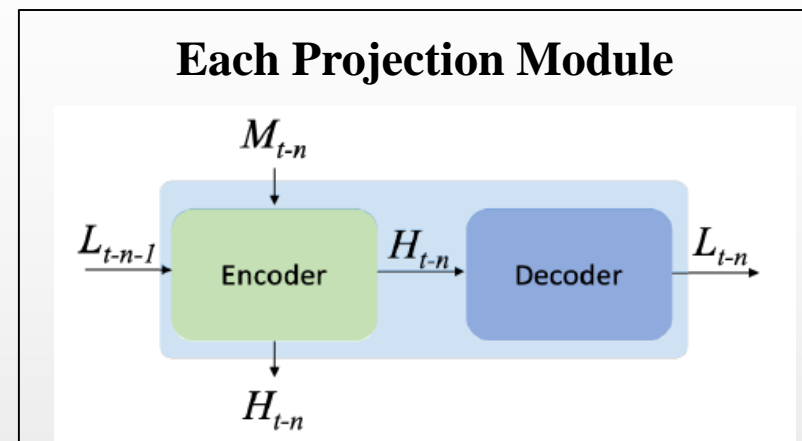
视频超分辨重建：相关工作

□ **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。

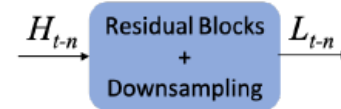


缺陷：

- 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



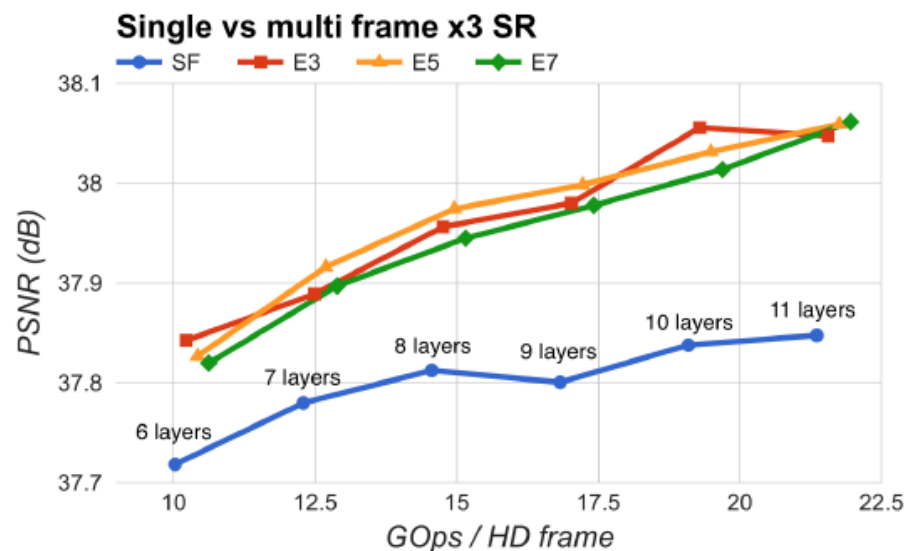
(a) Encoder (the back-projection)



(b) Decoder

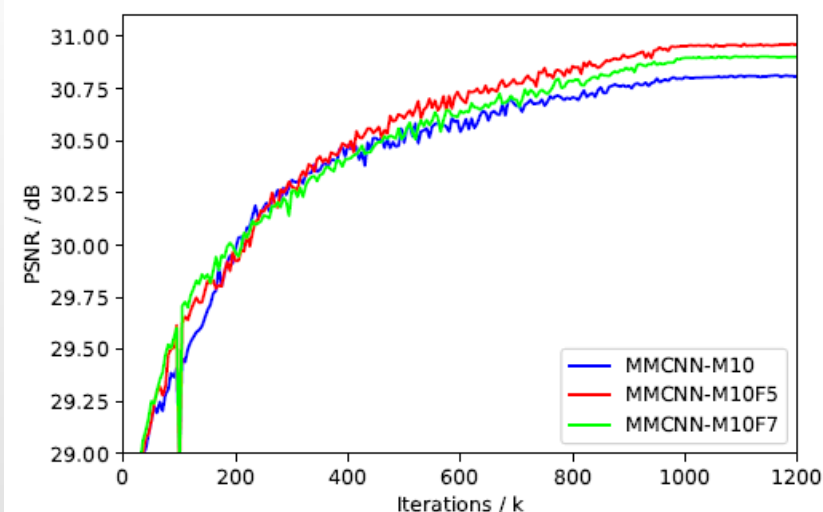
人脸视频超分辨重建: Motivation

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.

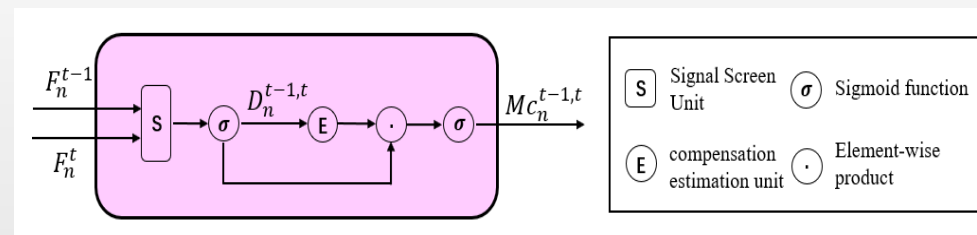
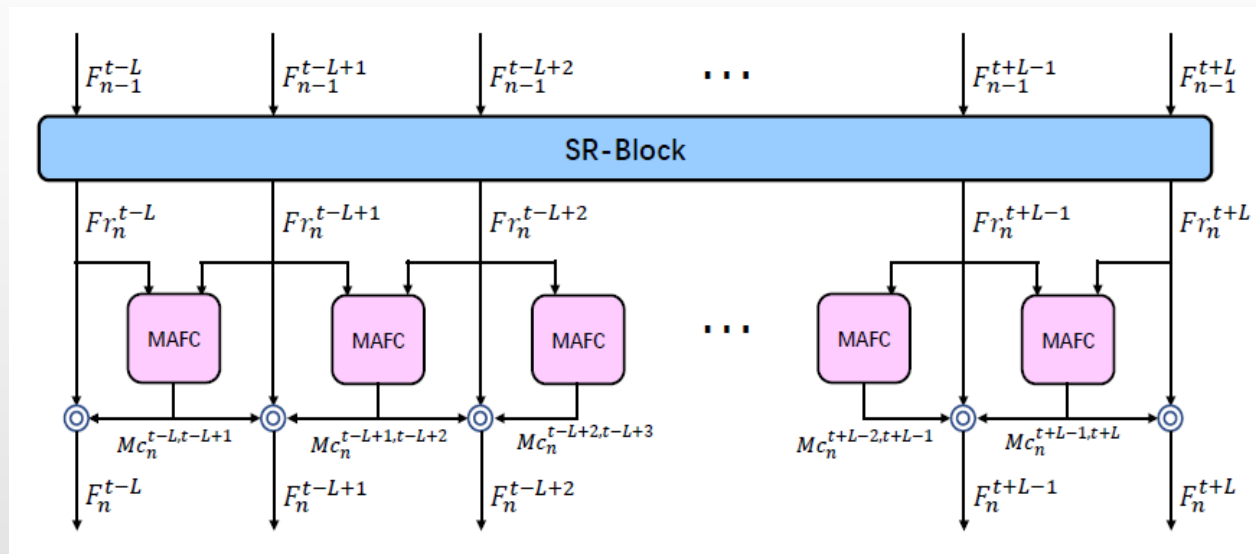


网络的学习能力是有限的，因此并不意味着输入帧数量越多效果就越好

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.

人脸视频超分辨重建：方法过程

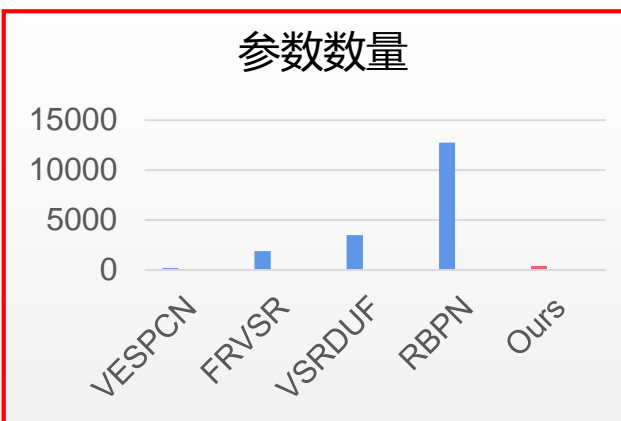
- 基本思想：将视频的帧间运动补偿操作转移到图像特征级进行，以遗忘、保留和增强的运动补偿方式，提升补偿的精度以及重建图像的质量



- 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

人脸视频超分辨重建：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

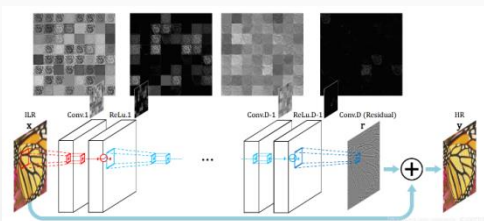


与其他方法相比，我们的模型具有更加轻量级的结构以及更高的图像重建质量

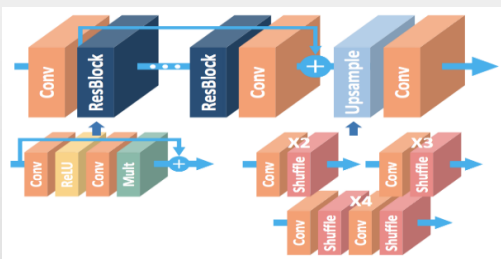


图像超分辨率重建的二值网络：研究背景

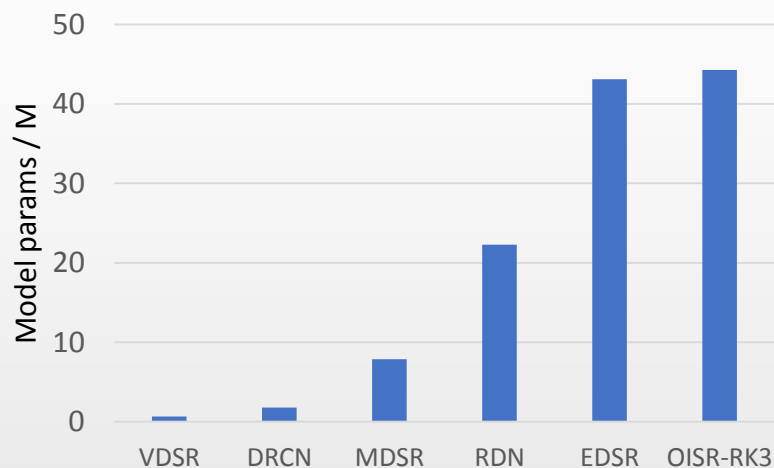
VDSR,
CVPR2016



EDSR, NTIRE2017
Champion

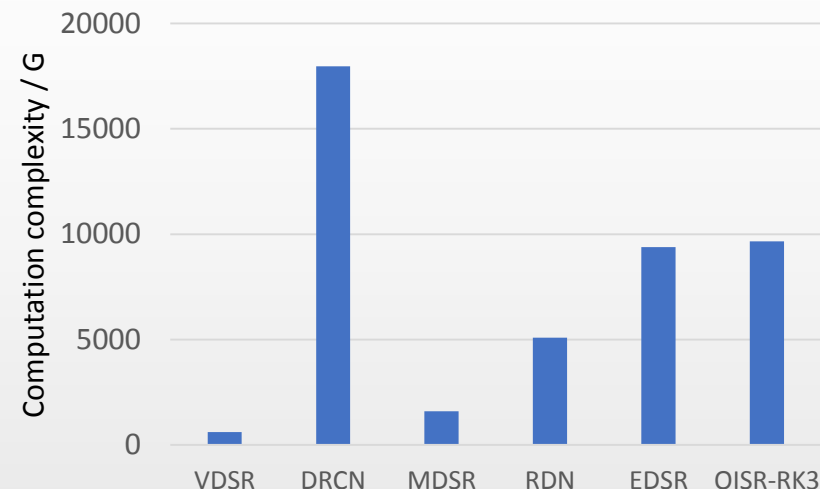


Model parameters



42 million 32-bit floating-point parameters

Computation complexity



9×10^9 floating-point multiplication and accumulation operations

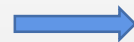
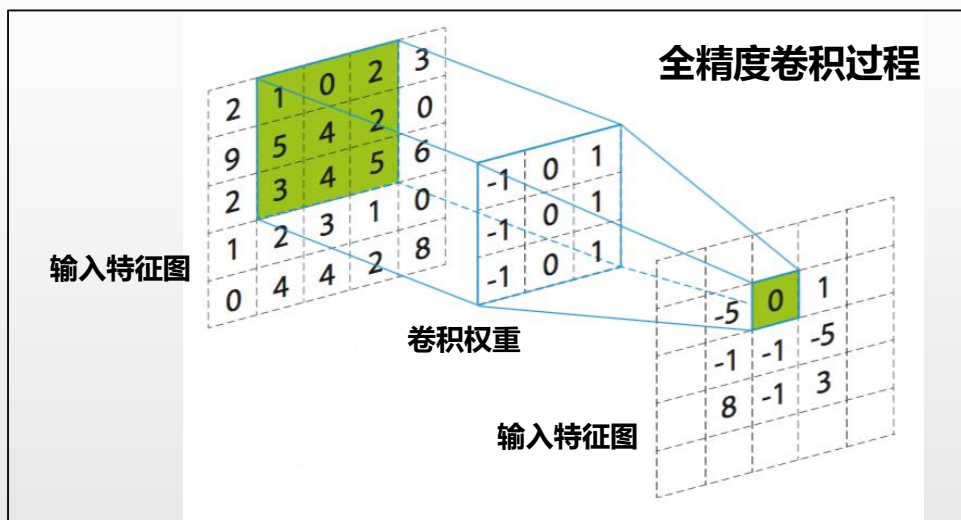
缺陷：

- 现有的深度图像超分辨率模型需要大量的模型存储空间及计算资源消耗，无法直接部署于手机等资源有限的移动端设备。

图像超分辨重建的二值网络：研究背景

- 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。

32bit



高性能计算设备

二值卷积过程

二值特征图

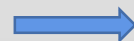
1	1	-1	1	-1	1	-1	1
-1	-1	1	-1	1	1	-1	-1

↓

-1	-1	-1	-1	-1	1	1	-1
----	----	----	----	----	---	---	----

1	1	-1
1	-1	1
-1	1	1

-3	1	-3	3	5	-5
7	-5	1	3	-3	3
1	3	1	-7	5	1
-1	3	1	1	-5	1
-5	-1	3	1	-1	1
3	-1	1	-1	3	-3



低性能计算设备

1bit

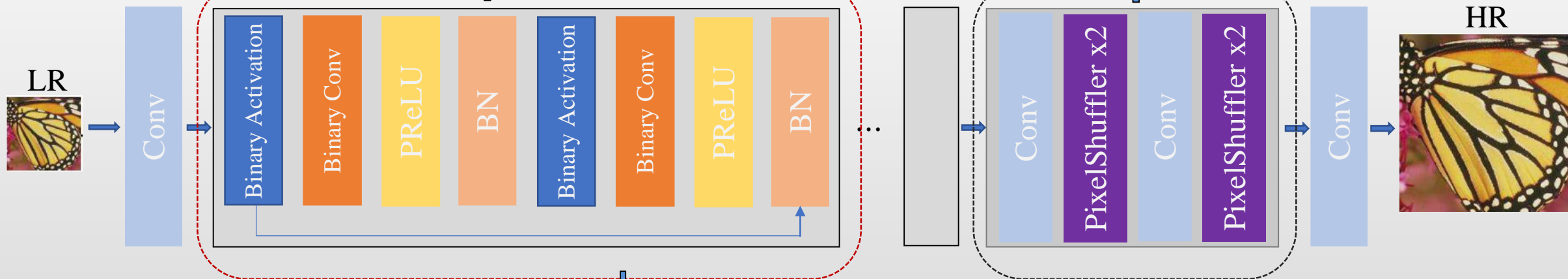
图像超分辨重建的二值网络：研究动机

二值超分辨率重建网络面临以下问题：

二值网络的计算速度未能充分展现，
模型效率受限。

二值卷积块中的BN层为全精度操作

上采样块中依然采用全精度权值和激活

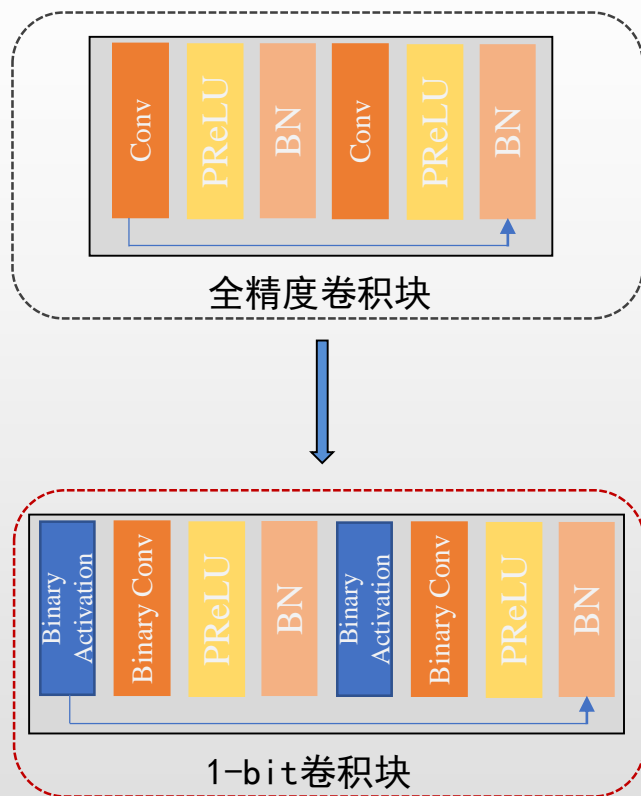


图中红框内的卷积操作为二值卷积操作

量化卷积过程依然存在较大的前向量化误差和反向梯度近似误差

二值网络对于图像信息保留
程度较低，模型性能受限

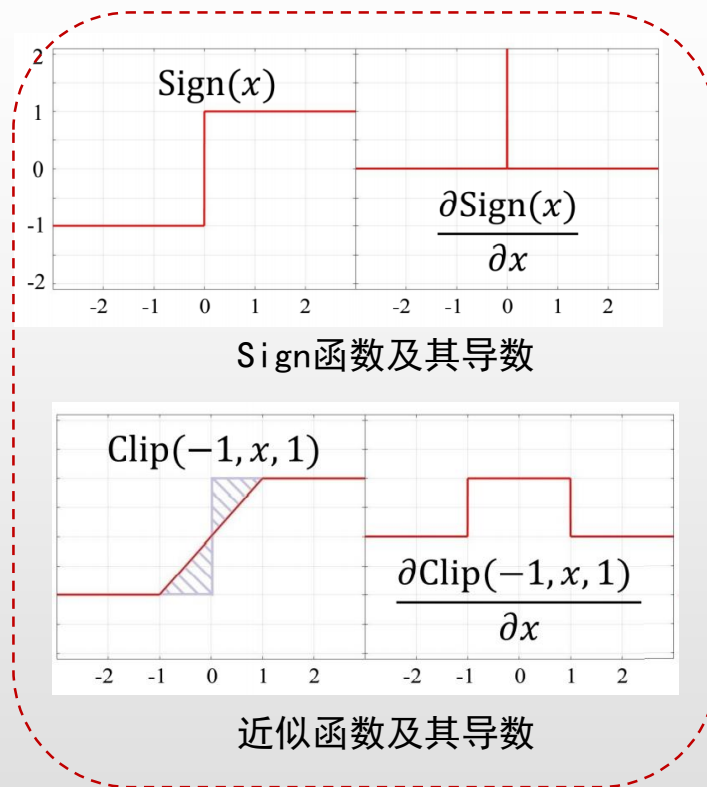
图像超分辨重建的二值网络：研究动机



全精度特征图



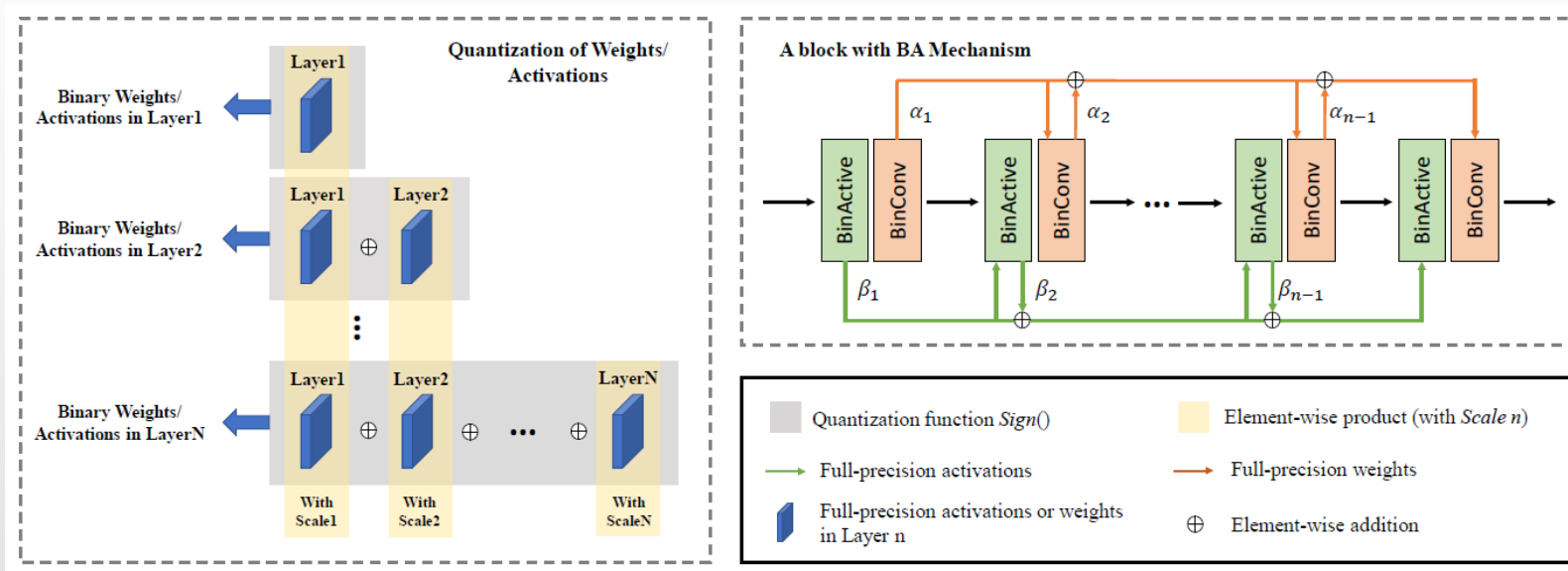
1-bit特征图



量化卷积过程存在较大的前向量化误差和反向梯度近似误差，图像超分辨任务对于数值量化精度有高依赖性。如何提高二值卷积的精度，实现优化的准确性？

图像超分辨重建的二值网络：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

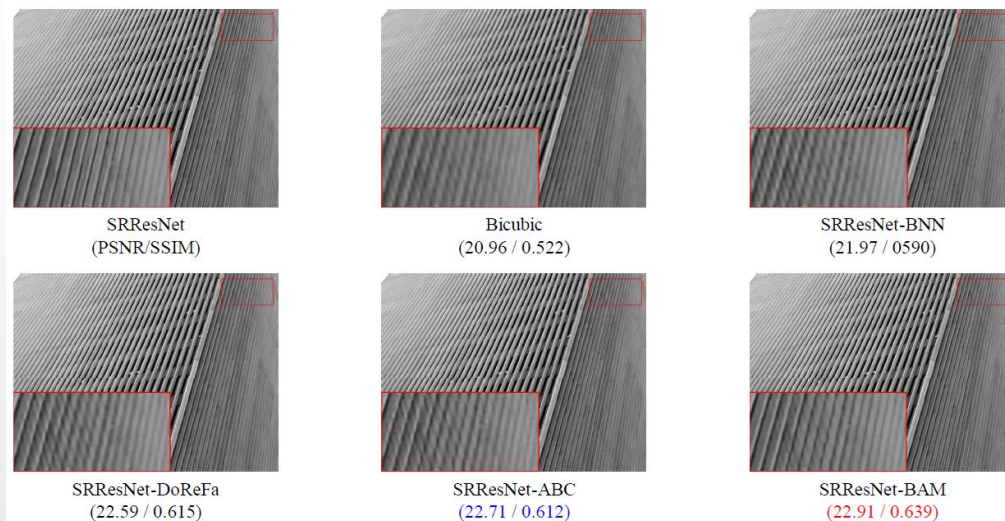
图像超分辨重建的二值网络：实验结果

客观评价：

Methods	Scale	Set5		Set14		B100		Urban100	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
VDSR	×2	37.53	0.959	33.05	0.913	31.90	0.896	30.77	0.914
Bicubic	×2	33.66	0.930	30.24	0.869	29.56	0.843	26.88	0.840
VDSR_BNN	×2	34.43	0.936	30.94	0.882	30.05	0.856	27.54	0.860
VDSR_DoReFa	×2	34.70	0.933	31.22	0.876	30.25	0.849	28.25	0.865
VDSR_ABC	×2	35.35	0.939	31.71	0.886	30.68	0.861	28.77	0.878
VDSR_BAM	×2	36.60	0.953	32.41	0.905	31.32	0.886	29.43	0.895
VDSR	×3	33.66	0.921	29.77	0.831	28.82	0.798	27.14	0.828
Bicubic	×3	30.39	0.868	27.55	0.774	27.21	0.739	24.46	0.735
VDSR_BNN	×3	31.01	0.874	28.15	0.791	27.57	0.755	25.01	0.758
VDSR_DoReFa	×3	31.79	0.895	28.68	0.806	27.98	0.766	25.53	0.782
VDSR_ABC	×3	32.01	0.898	28.86	0.808	28.08	0.770	25.80	0.787
VDSR_BAM	×3	32.52	0.907	29.17	0.819	28.29	0.782	26.07	0.799
VDSR	×4	31.35	0.884	28.01	0.767	27.29	0.725	25.18	0.752
Bicubic	×4	28.42	0.810	26.00	0.703	25.96	0.668	23.14	0.658
VDSR_BNN	×4	29.02	0.827	26.55	0.724	26.29	0.685	23.55	0.685
VDSR_DoReFa	×4	29.39	0.837	26.79	0.728	26.45	0.689	23.81	0.696
VDSR_ABC	×4	29.59	0.841	29.63	0.730	26.51	0.687	23.96	0.699
VDSR_BAM	×4	30.31	0.860	27.46	0.749	26.83	0.706	24.38	0.720

将本方法部署于VDSR模型时，领先SOTA结果**1.25dB** (x2 Set5), 与全精度网络相差**0.58dB** (x2 B100)。

主观评价：



图像超分辨重建的推广应用：超高清电视

- 跨域图像重建可以解决超高清视频源稀缺的问题：由高清视频到超高清视频的转换



720p高清图片



4K/8K超高清图片



4K/8K高清电视



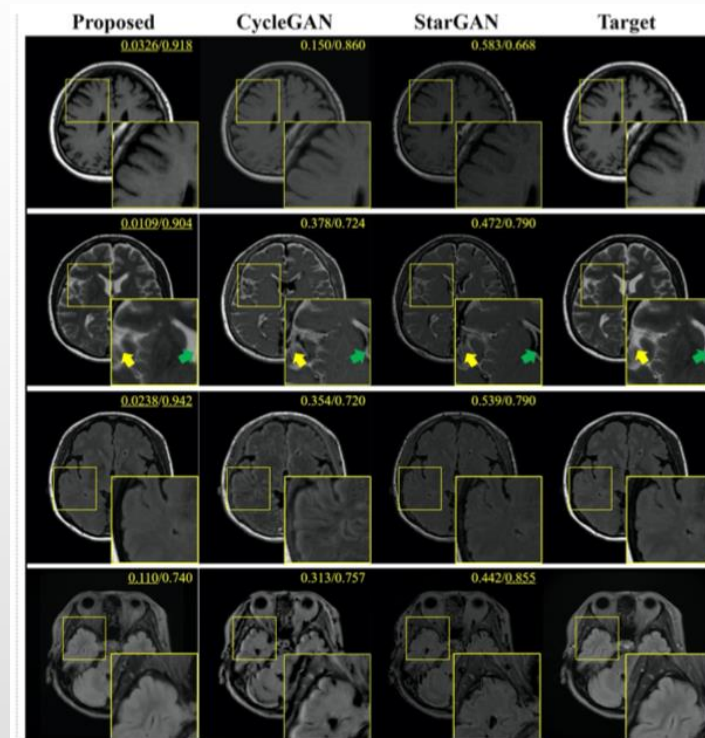
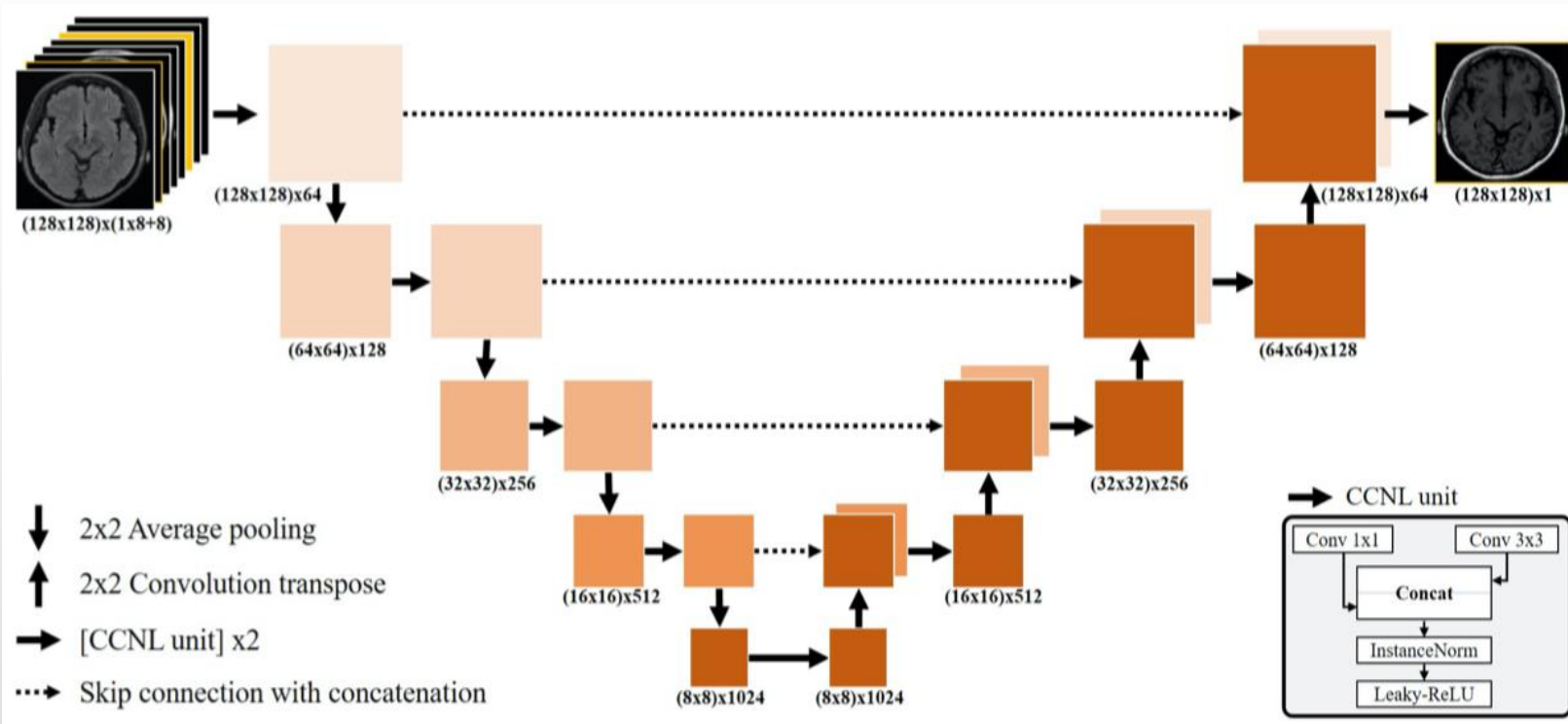
汇报提纲

- 一. 前言 (Preface) : 什么是跨域图像重建?
- 二. 画像-照片合成 (Sketch-Photo Synthesis)
- 三. 图像超分辨重建 (Image Super-resolution Reconstruction)
- 四. 其他相关应用 (Other Related Applications)
- 五. 结束语 (Conclusions)



医学图像合成：相关工作-CollaGAN

基于协同生成对抗网络的医学图像数据补齐

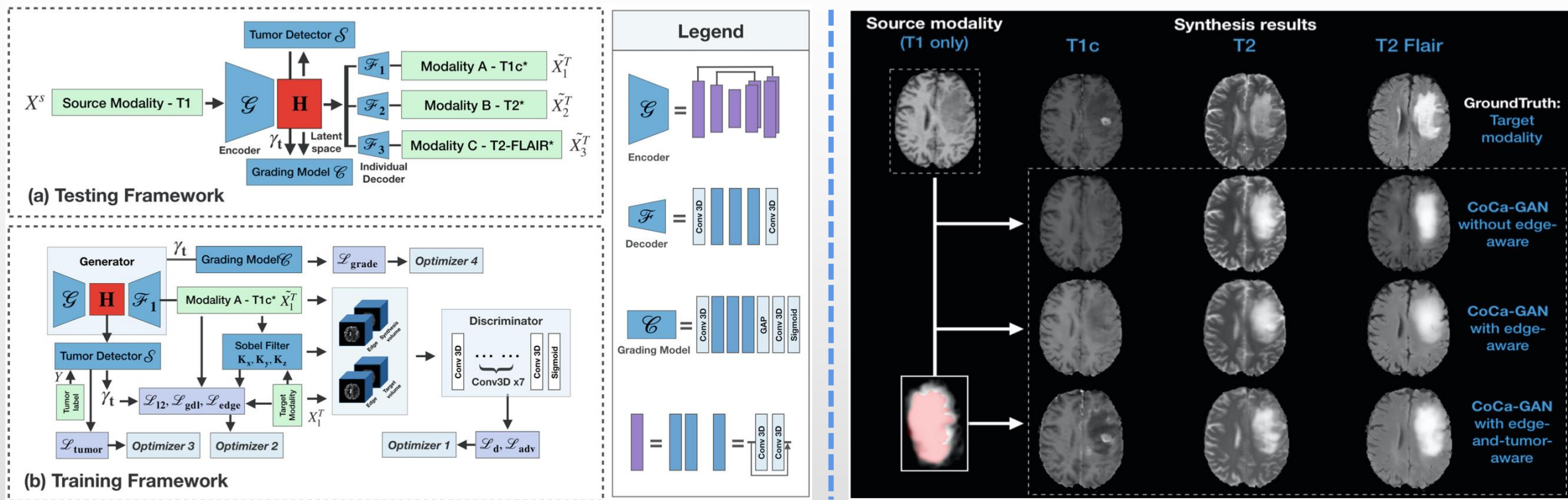


- 通过级联多种模态的医学图像作为输入，从尽量多的不同医学图像模态中，学习输入模态与目标模态的关系，利用多种医学图像生成缺失的模态。

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

医学图像合成：相关工作-CoCaGAN

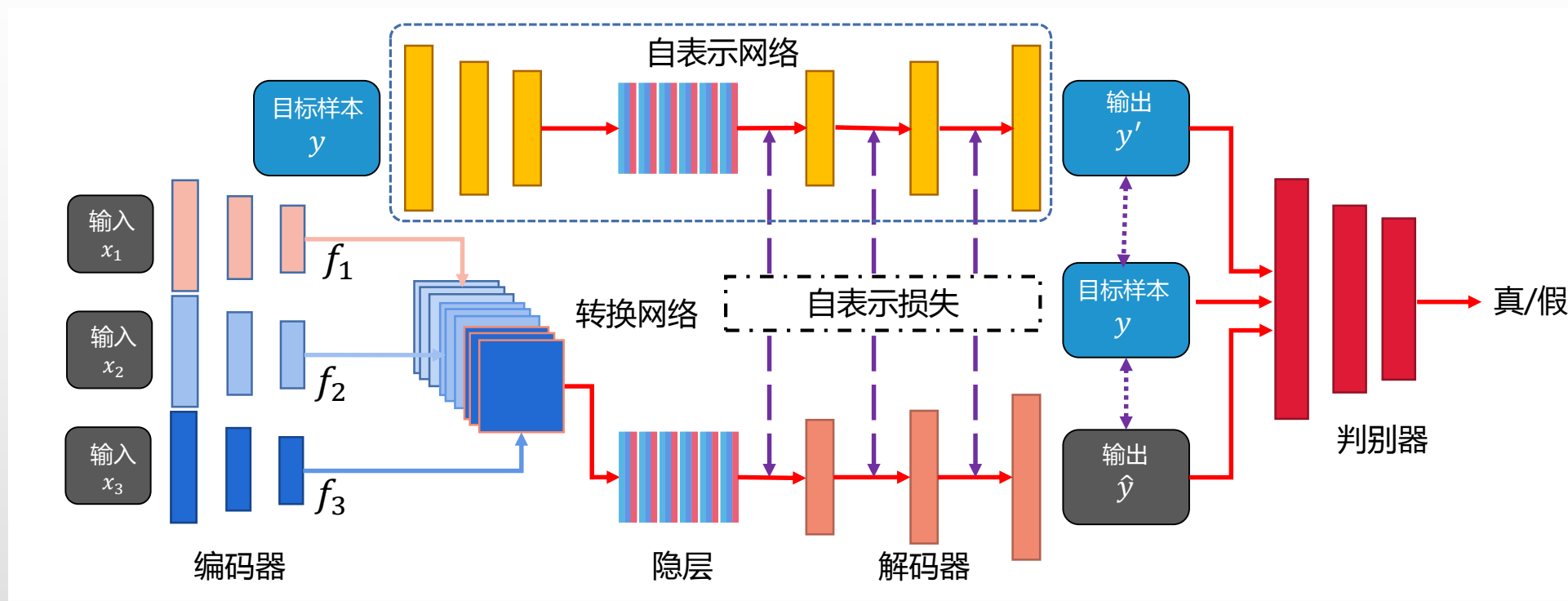
基于共同空间学习的上下文相关生成对抗网络



- 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

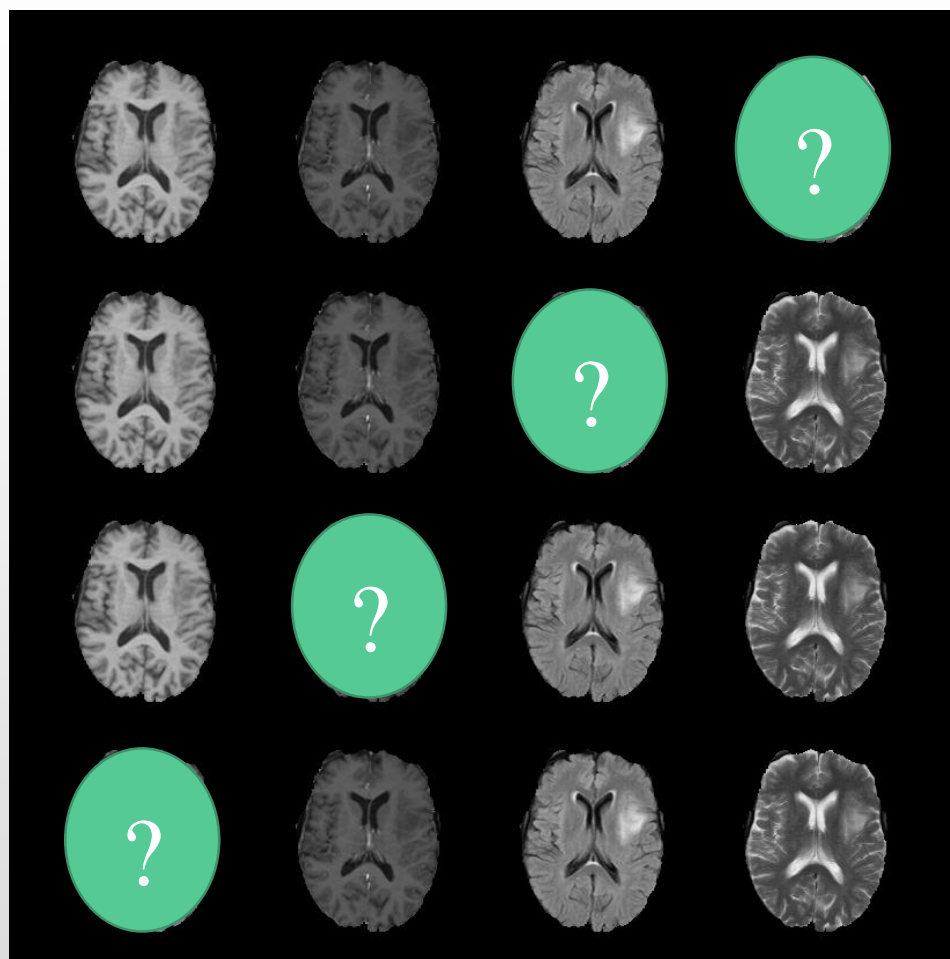
医学图像合成：AutoGAN

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

医学图像合成: AutoGAN

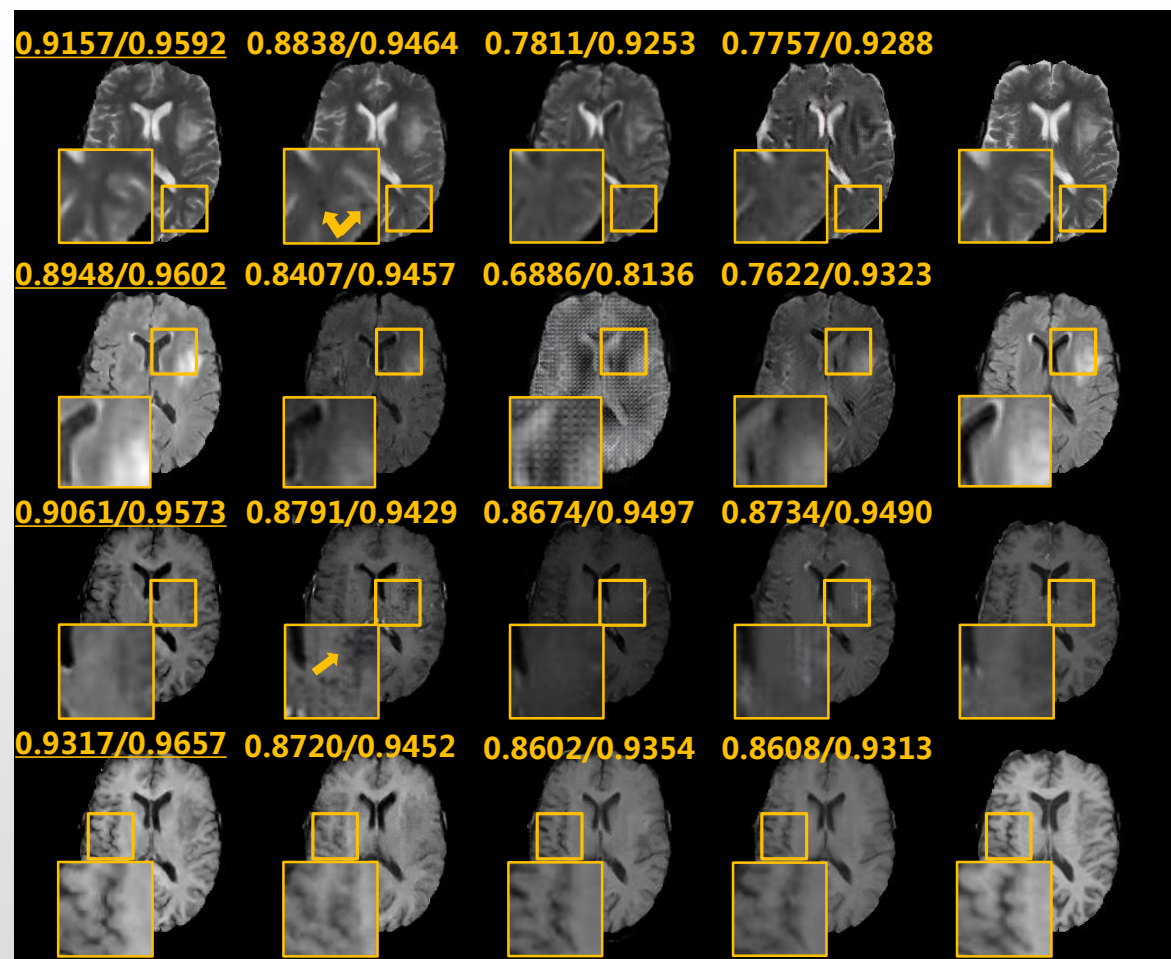


T1

T1-C

T2-FLAIR

T2



Proposed

Pix2Pix

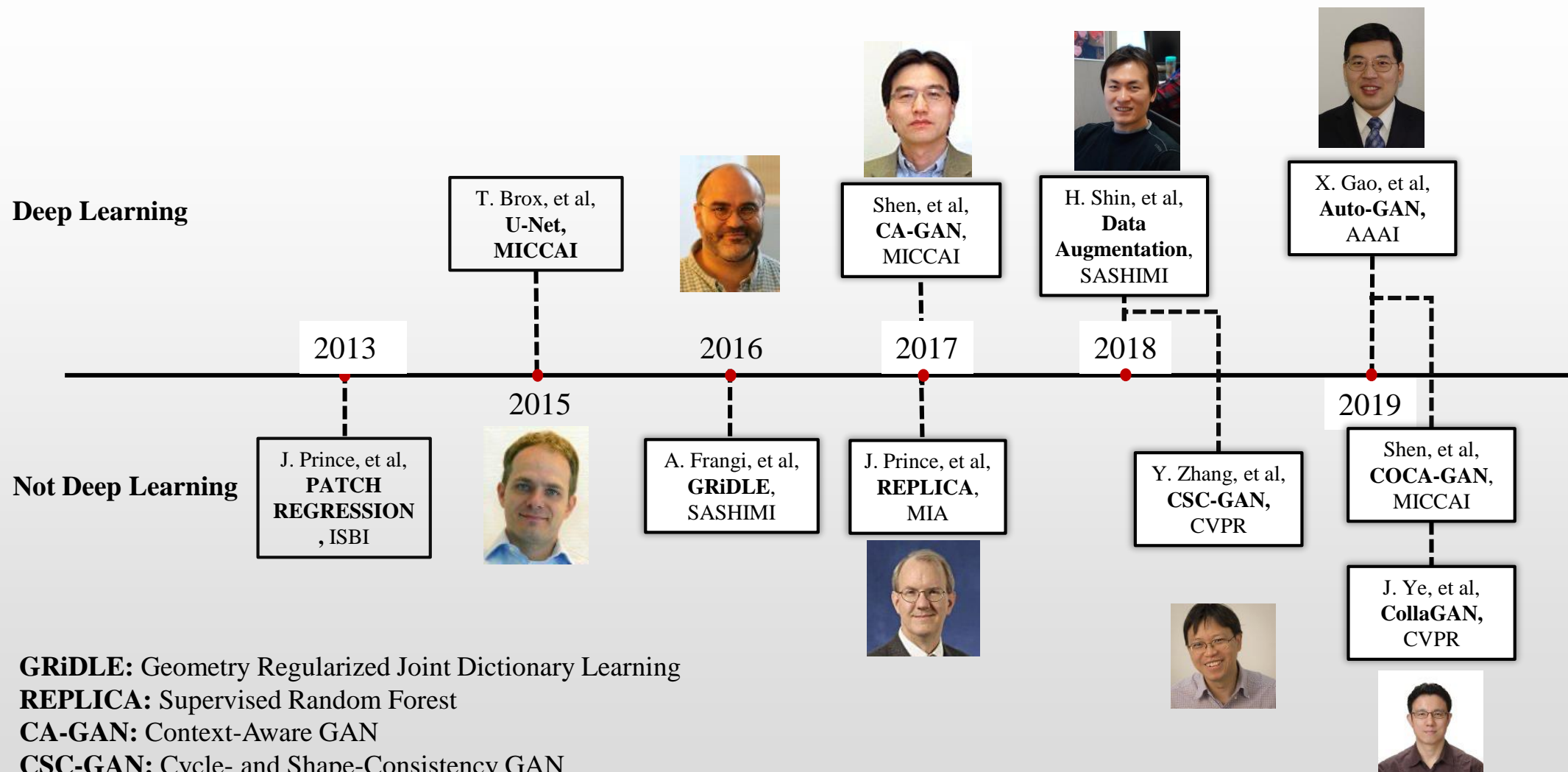
CycleGAN

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.

医学图像合成：发展历程



GRiDLE: Geometry Regularized Joint Dictionary Learning

REPLICA: Supervised Random Forest

CA-GAN: Context-Aware GAN

CSC-GAN: Cycle- and Shape-Consistency GAN

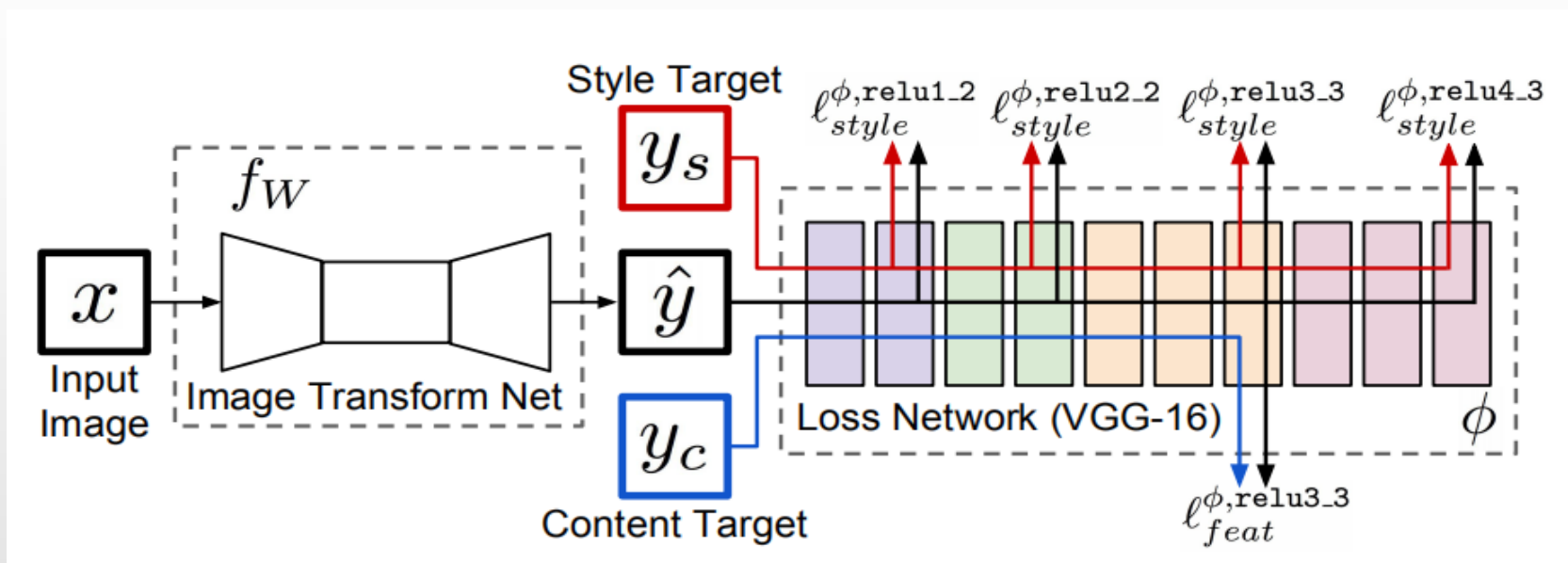
CollaGAN: Collaborative GAN

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

Auto-GAN: Self-Supervised Collaborative Learning

绘画风格转换：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

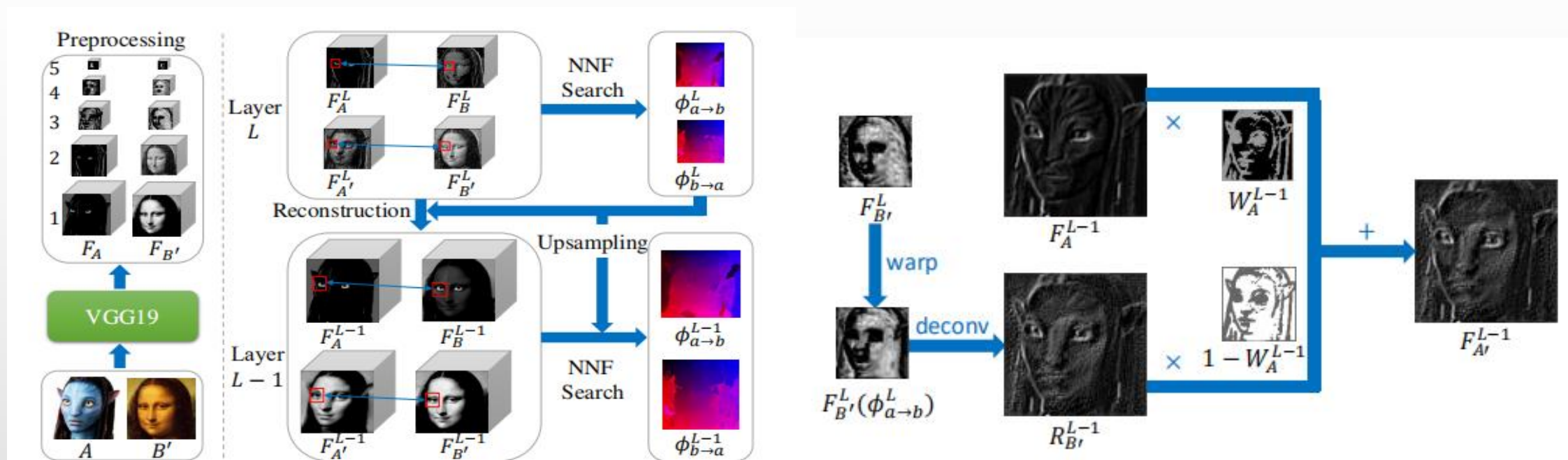


缺陷：

- 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

绘画风格转换：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：

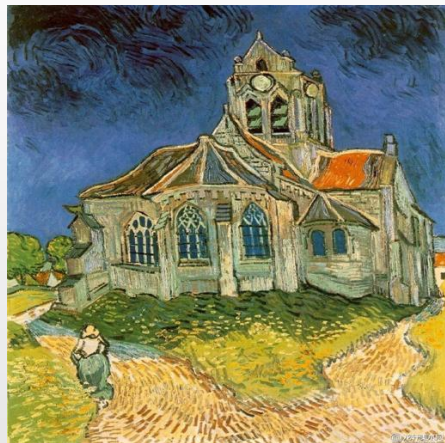
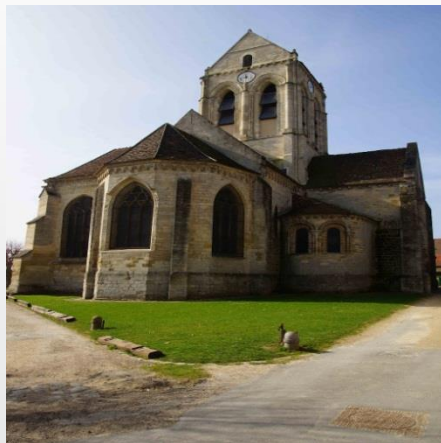
- 逐层进行特征匹配及重构计算量庞大，生成结果时间过长。
- 仅适用于语义及结构高度一致的跨域图像。采用欧几里得距离匹配图像深度特征，使跨域图像之间的相似性对跨域特征匹配结果有较大影响。

绘画风格转换：Motivation

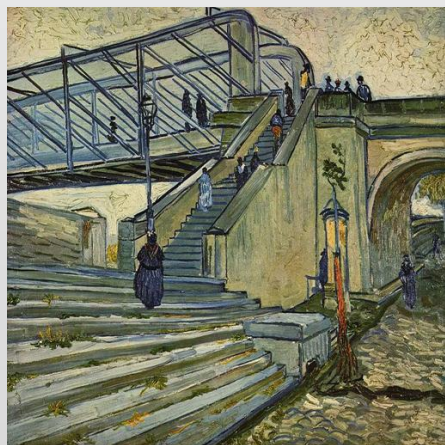
“艺术风格是一个人或一个群体的艺术作品中不变的元素或表达方式。”

——美国艺术史学家Meyer Schapiro

奥维尔小镇教堂



阿尔勒天桥

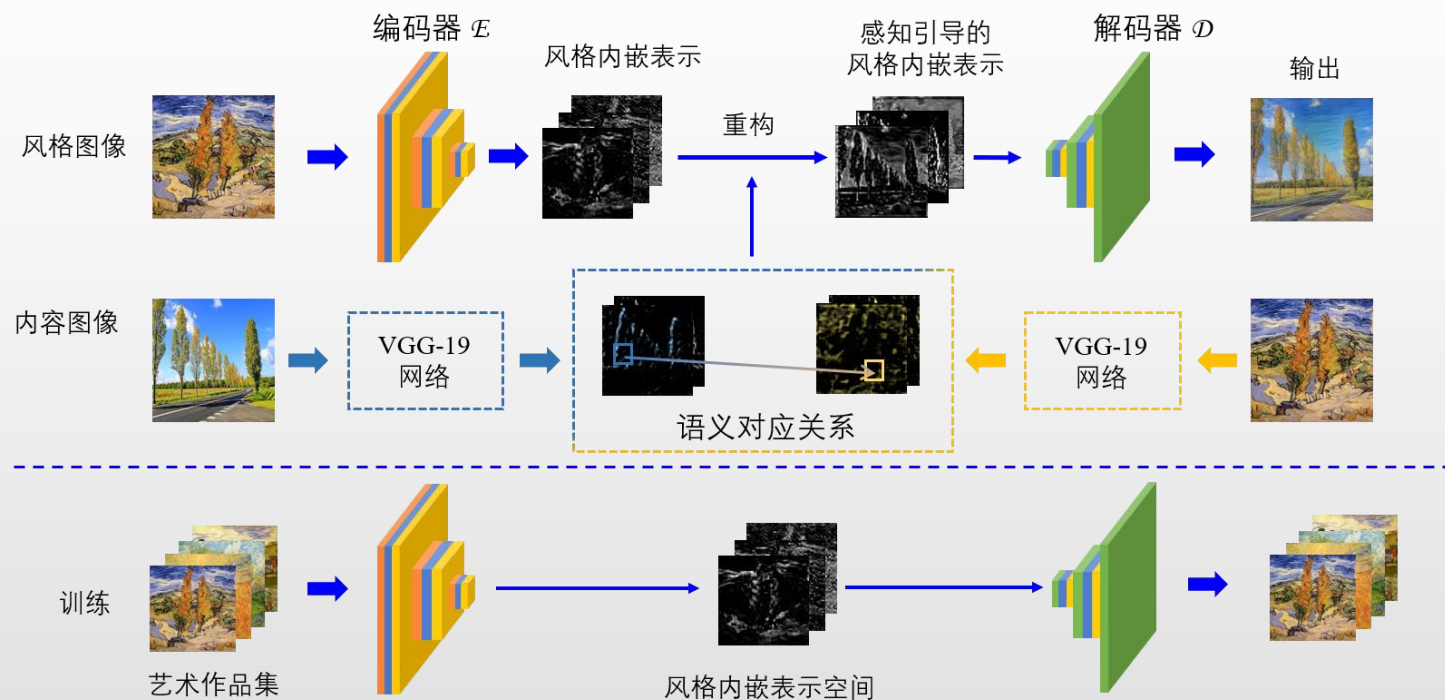


□ 色彩模式和纹理迁移破坏图像语义一致性

- 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

绘画风格转换：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

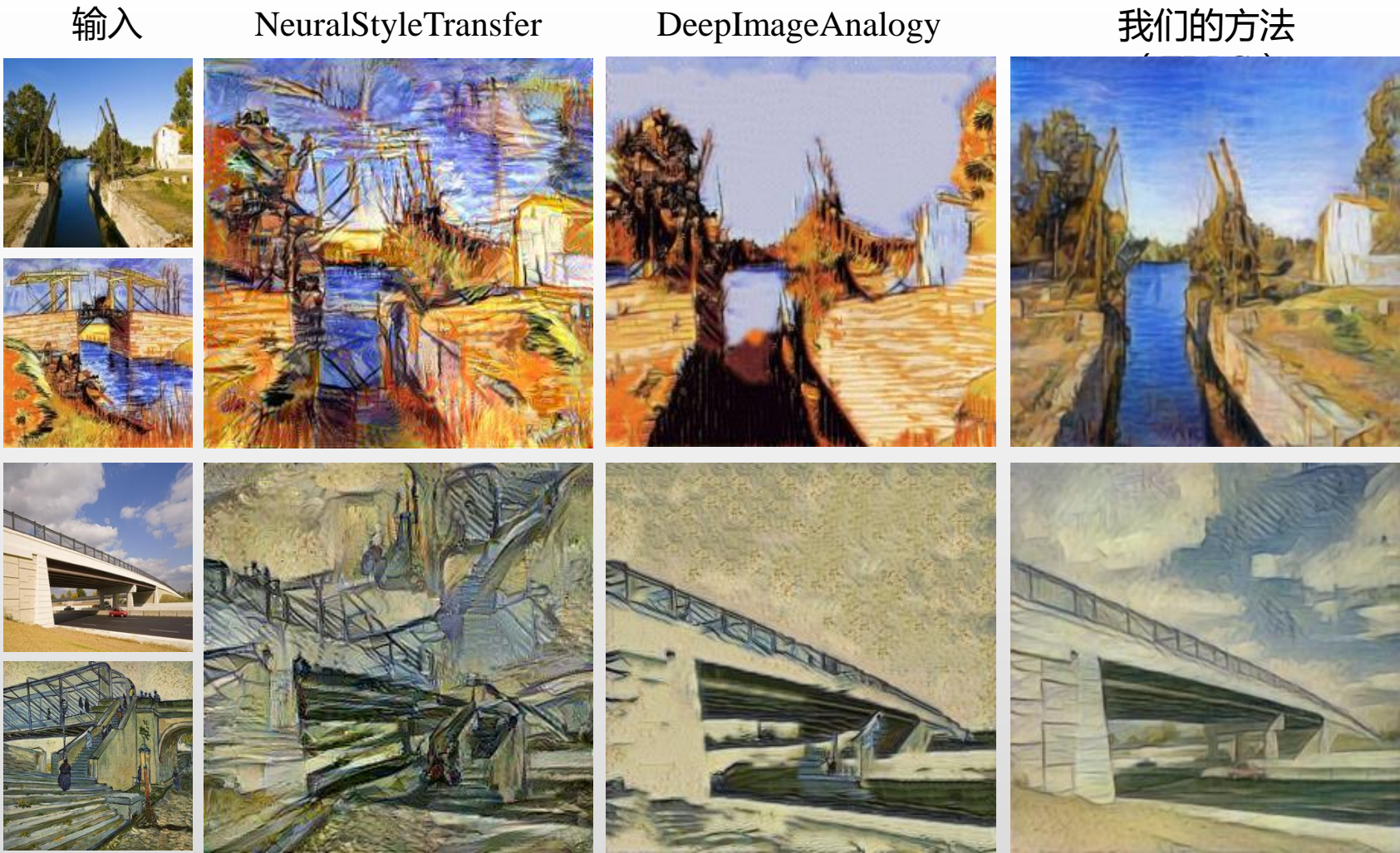
$$f_{c \rightarrow s}^L(p) = \arg \min_q \sum_{x \in M(p), y \in M(q)} -corr(D_c^L(x), D_s^L(y)),$$

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

绘画风格转换：实验结果

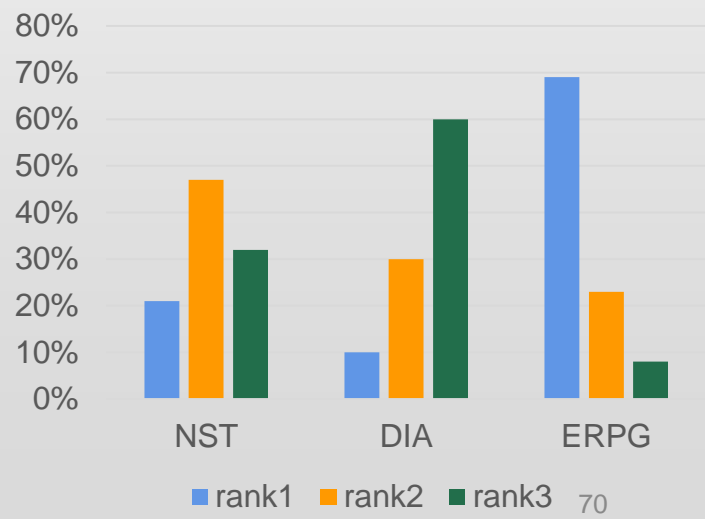


感知引导保持了跨域图像的语义结构；迁移整体风格特征，增强“手绘感”

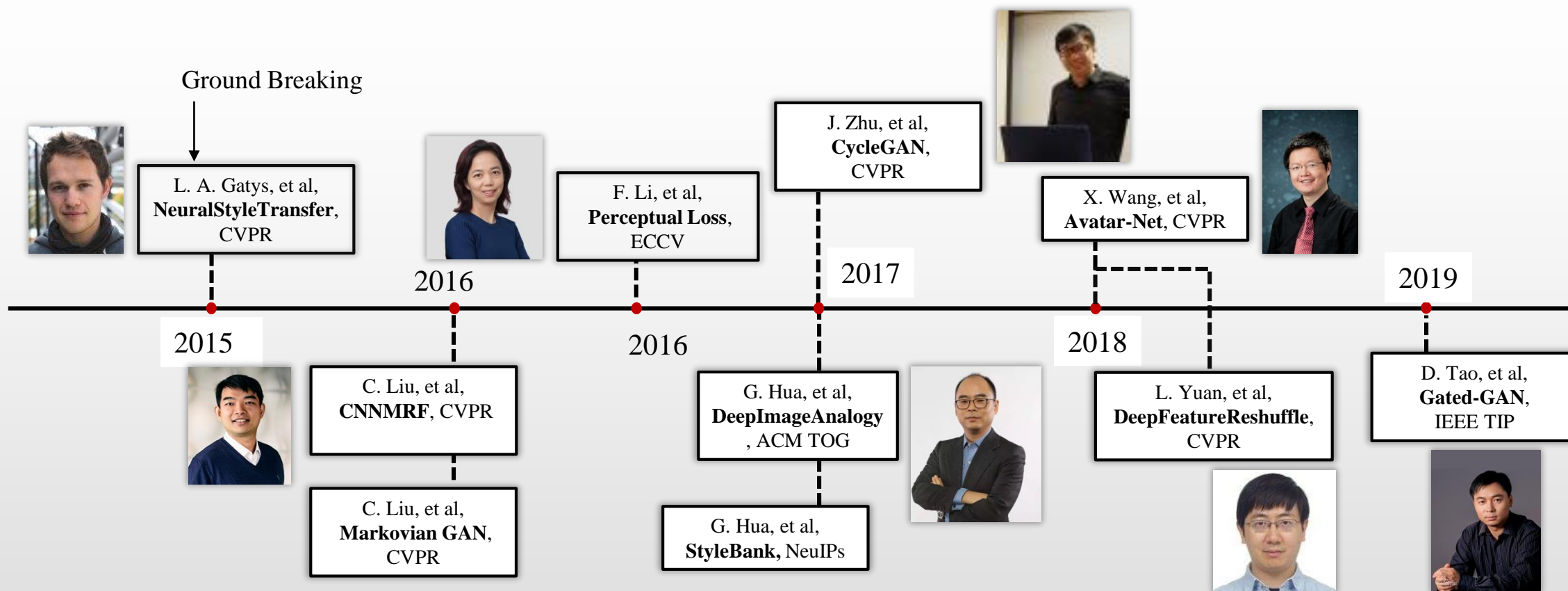
客观评价

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

主观评价



绘画风格转换：发展历程



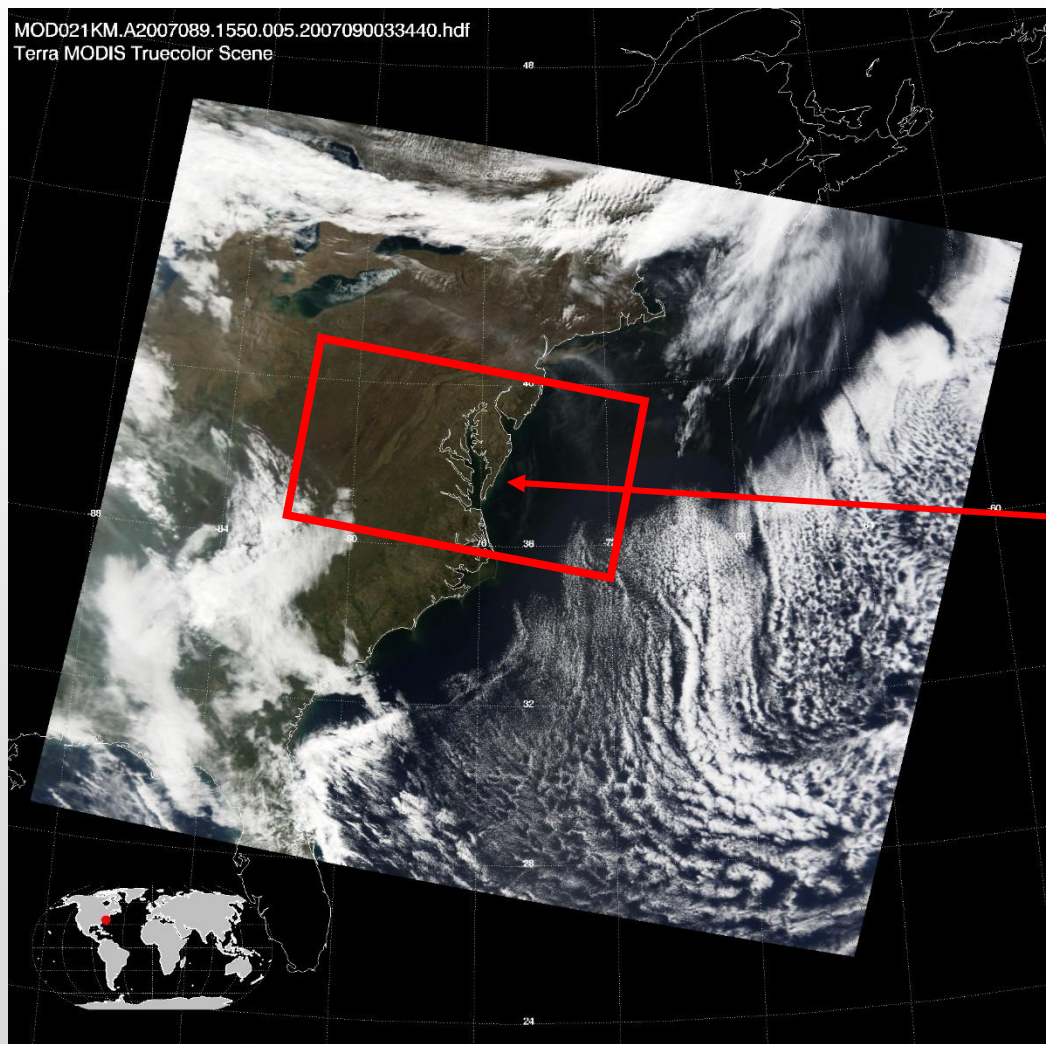
基于优化的方法：NeuralStyleTransfer, CNNMRF, Deep Image Analogy

基于零样本学习的方法：Avatar-Net, Deep Feature Reshuffle

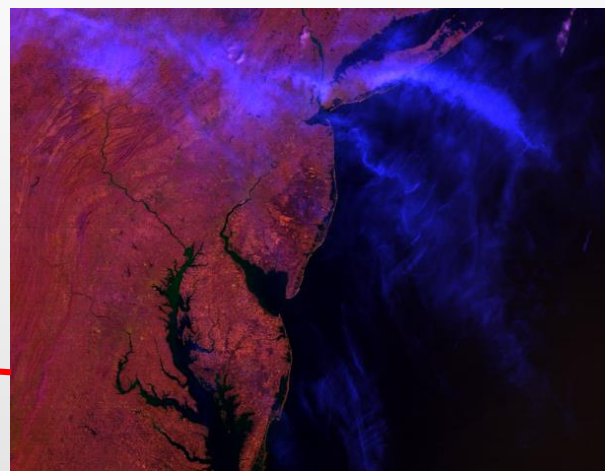
基于前馈神经网络的方法: Perceptual Loss, StyleBank,

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

遥感图像缺失信息补全：研究背景

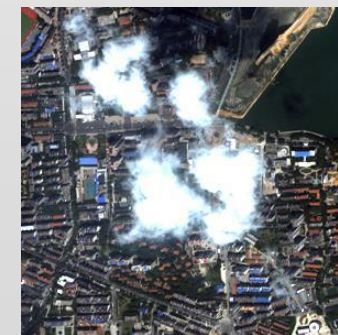
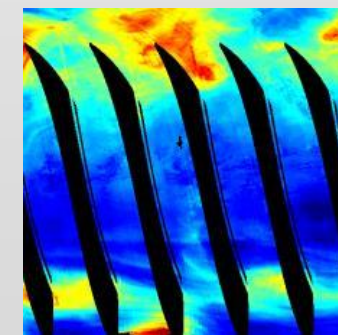
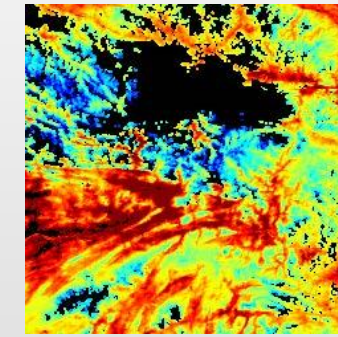
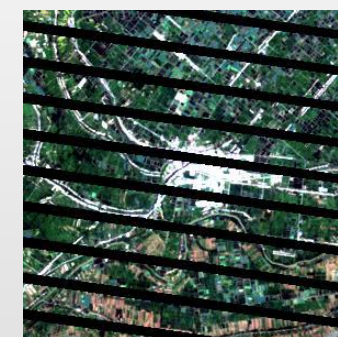
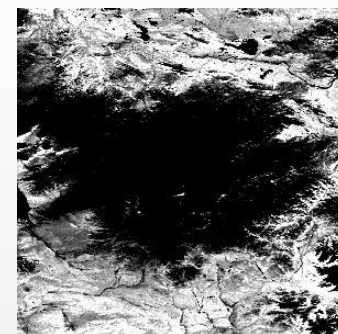
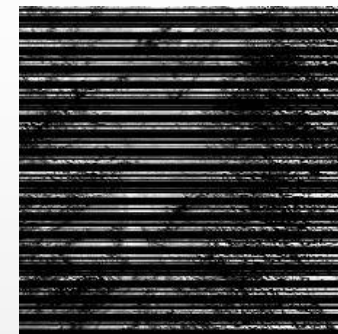


伪彩图像(R 2.1um, G 0.65um, B 1.38um)

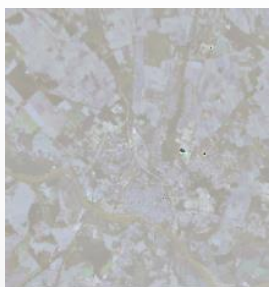


遥感图像获取过程中，会受到天气因素(云、雾霾、积雪)的影像，产生缺失信息。

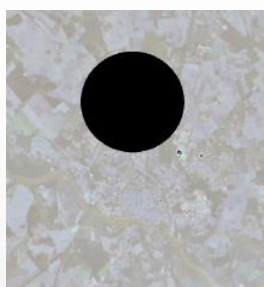
硬件故障引起的缺失信息 成像条件引起的缺失信息



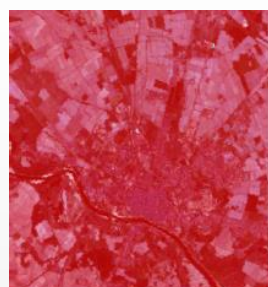
基于联合邻域回归的多时相遥感数据缺失信息复原



(a) 原始图像



(b) 仿真有云图像



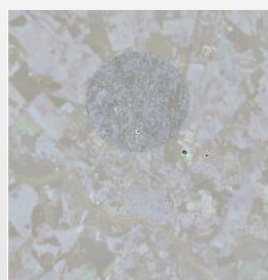
(c) 参考图像



(d) 直接替代法



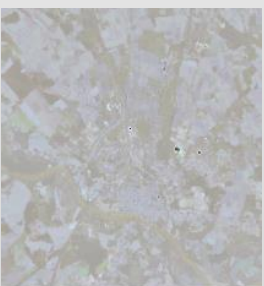
(e) K-SVD



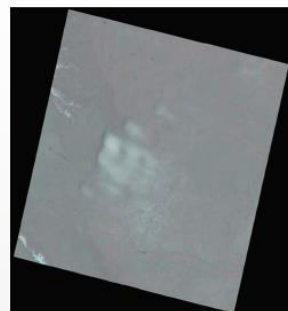
(f) ANR



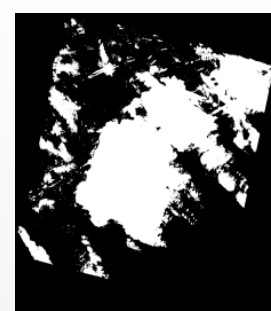
(g) A+



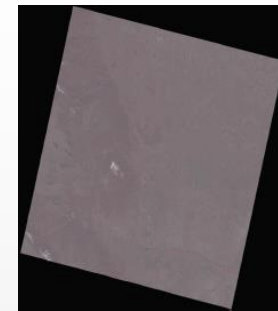
(h) 我们的方法



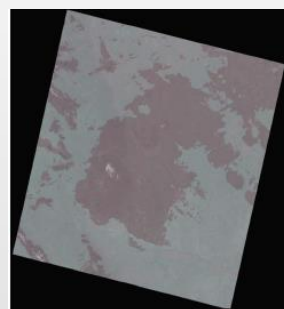
(a) 原始图像



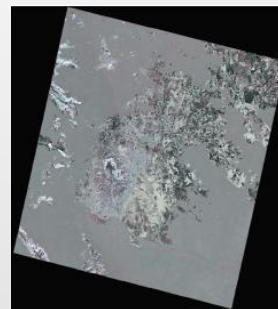
(b) Fmask云图像



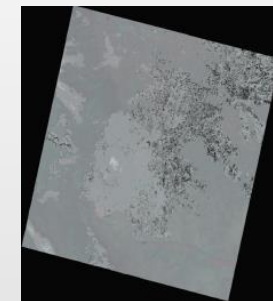
(c) 参考图像



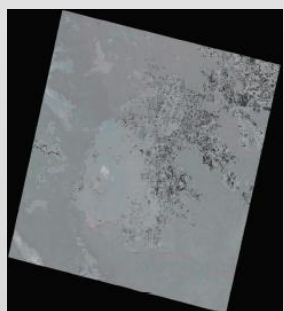
(d) 直接替代法



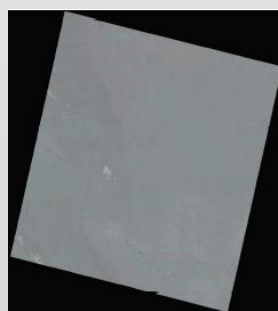
(e) K-SVD



(f) ANR



(g) A+



(h) 我们的方法

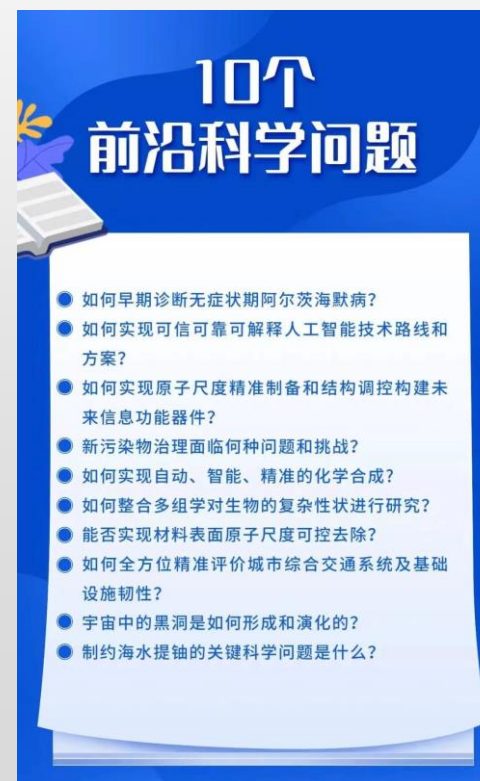
汇报提纲

- 一. 前言 (Preface) : 什么是跨域图像重建?
- 二. 画像-照片合成 (Sketch-Photo Synthesis)
- 三. 图像超分辨重建 (Image Super-resolution Reconstruction)
- 四. 其他相关应用 (Other Related Applications)
- 五. 结束语 (Conclusions)



结束语

- **图像跨域重建**：属于AI2.0中的**跨媒体智能**研究的范畴，主要研究跨媒体的生成，又称作**图像翻译**（Image Translation）
- **典型应用包括**：画像-照片合成，可见光-近红外图像转化，图像超分辨重建，医学图像跨模态生成，图像风格转化，多光谱图像补全
- **未来研究方向**：
 - 训练样本规模：大样本 → 小样本 → 零样本学习
 - 有监督 → 半监督 → 部分监督 → 弱监督 → 无监督学习
 - 基于可信、可靠、可解释的图像跨域重建模型
 - 基于数据、知识（模型）双驱动图像跨域重建算法



致谢



王楠楠 教授



彭春蕾 副教授



朱明瑞 准聘副教授



马卓奇 准聘副教授



曹兵 助理教授



刘德成



辛经纬



郝毅



姜馨蕊



丁鑫鹏



李国璋



李柯宇

致谢



An aerial view of a high-speed train, likely a Shinkansen, traveling through a lush forest. The train is white with a blue stripe and is moving along a curved track. The surrounding trees are in various stages of bloom, with many showing pink and white blossoms, suggesting a spring setting. The train's front is aerodynamic, and it has multiple windows. The tracks are surrounded by a green fence, and the overall scene is vibrant and scenic.

谢谢!