



WORLD ARTIFICIAL INTELLIGENCE
CONFERENCE
2023世界人工智能大会



北京大學
PEKING UNIVERSITY

专题论坛: 智能媒体计算

刘家瑛

北京大学 王选计算机研究所

liujaying@pku.edu.cn

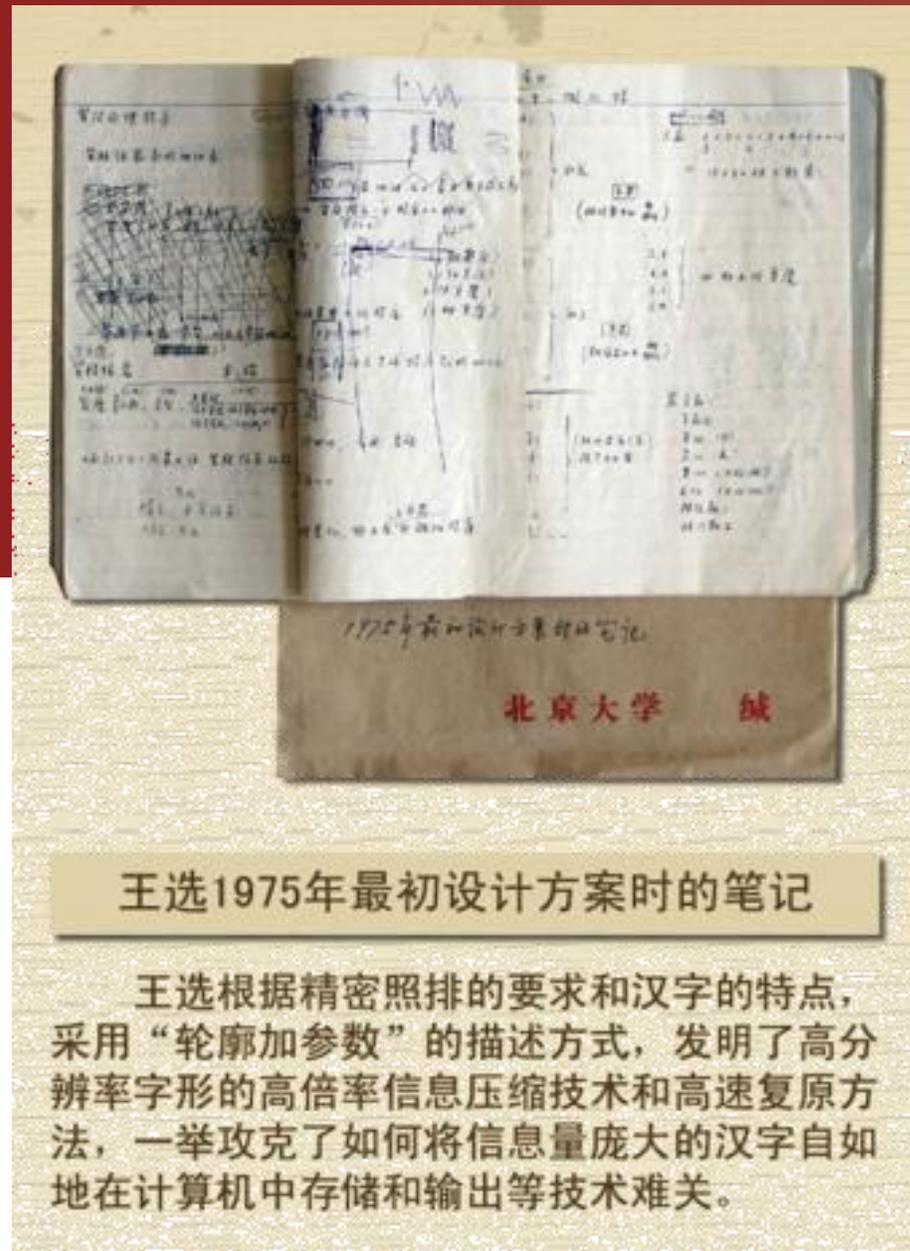
北京大学 王选计算机研究所



- 北京大学的二级教学科研机构
- 研究方向包括图形图像处理技术与数字出版应用、认知计算与知识服务技术、跨媒体智能处理与分析技术、数字文档处理技术等
- 建有硕士、博士培养点及博士后流动站，电子出版新技术国家工程研究中心，中国文字字体设计与研究中心，新闻出版智能媒体技术重点实验室等科研基地

北京大学 王选计算机研究所

- **1974年8月**，国家计委批准立项重点科技攻关项目“汉字信息处理系统工程”，简称为“**748工程**”
- **1975年**，王选老师提出了汉字字形信息压缩及快速复原的技术方案



王选计算机研究所 发展历程

- **1976年**, 北京大学成立“748工程会战组”
- **1977年**, 成立了“北京大学汉字信息处理技术研究室”



旧图书馆一楼



王选的原理性样机主要设计手稿

1976年, 王选决策跳过当时世界流行的第二代、第三代照排机, 直接研制当时尚无商品的第四代激光照排系统, 大胆选择了技术上的跨越发展。

科学研究有时可以采取迂回策略, 用创新的设计, 绕过按常规方式发展会遇到的巨大困难, 实现技术发展的跨越, 往往能够走一条高效益的、事半功倍的捷径。

——王选

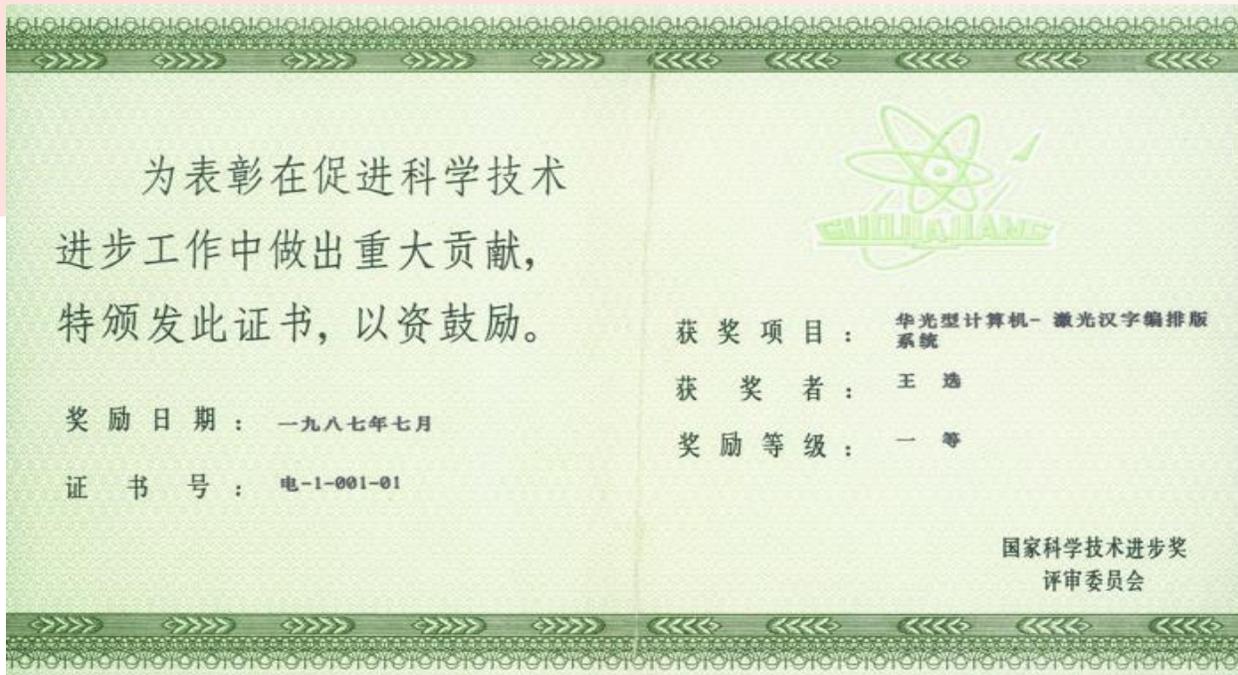
王选计算机研究所 发展历程

- **1983年**，研制成功华光Ⅱ型计算机—激光汉字编辑排版系统
- **北京大学计算机科学技术研究所正式成立**

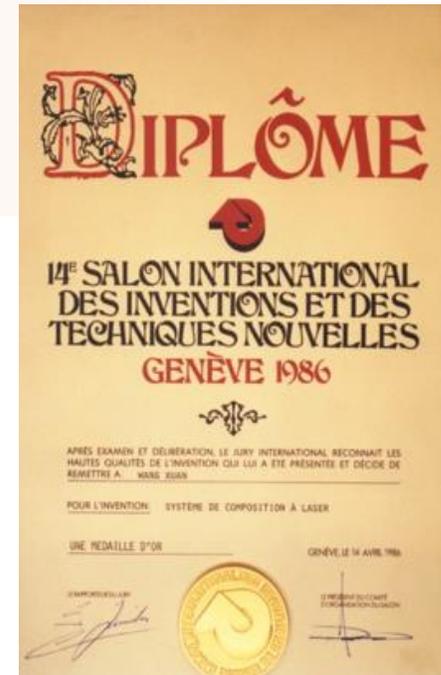
图为新华社印刷厂的试验车间里，周培源（左二）、卢嘉锡（右三）、黄辛白（左一）等领导 and 专家认真听取王选（右一）介绍华光Ⅱ型系统的运行情况。



王选计算机研究所 发展历程

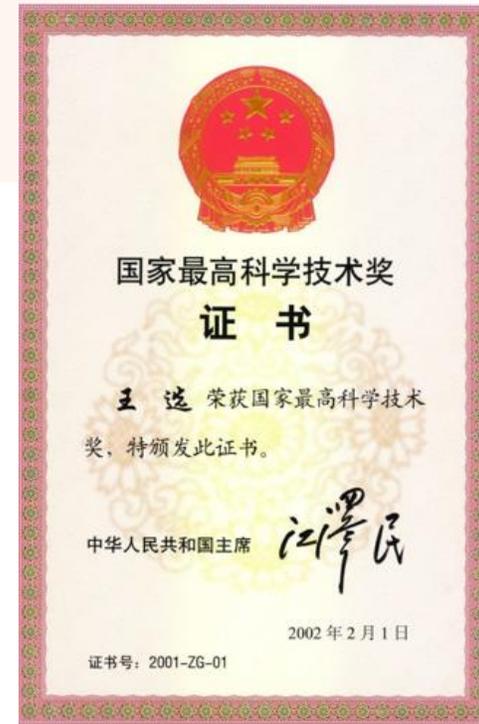


华光型计算机 — “激光汉字编排版系统”
获1986年度国家科技进步一等奖



华光III型系统 获 **第14届**
日内瓦国际发明展览金牌奖

王选计算机研究所 发展历程



2002年, 王选老师荣获 2001年度国家最高科学技术奖

王选计算机研究所 发展历程



2019年，北京大学计算机科学技术研究所 更名为 **北京大学 王选计算机研究所**

王选计算机研究所 发展历程



坚持“顶天立地”产学研道路

巩固现有媒体技术成果

发展智能媒体技术

STRUCT Group

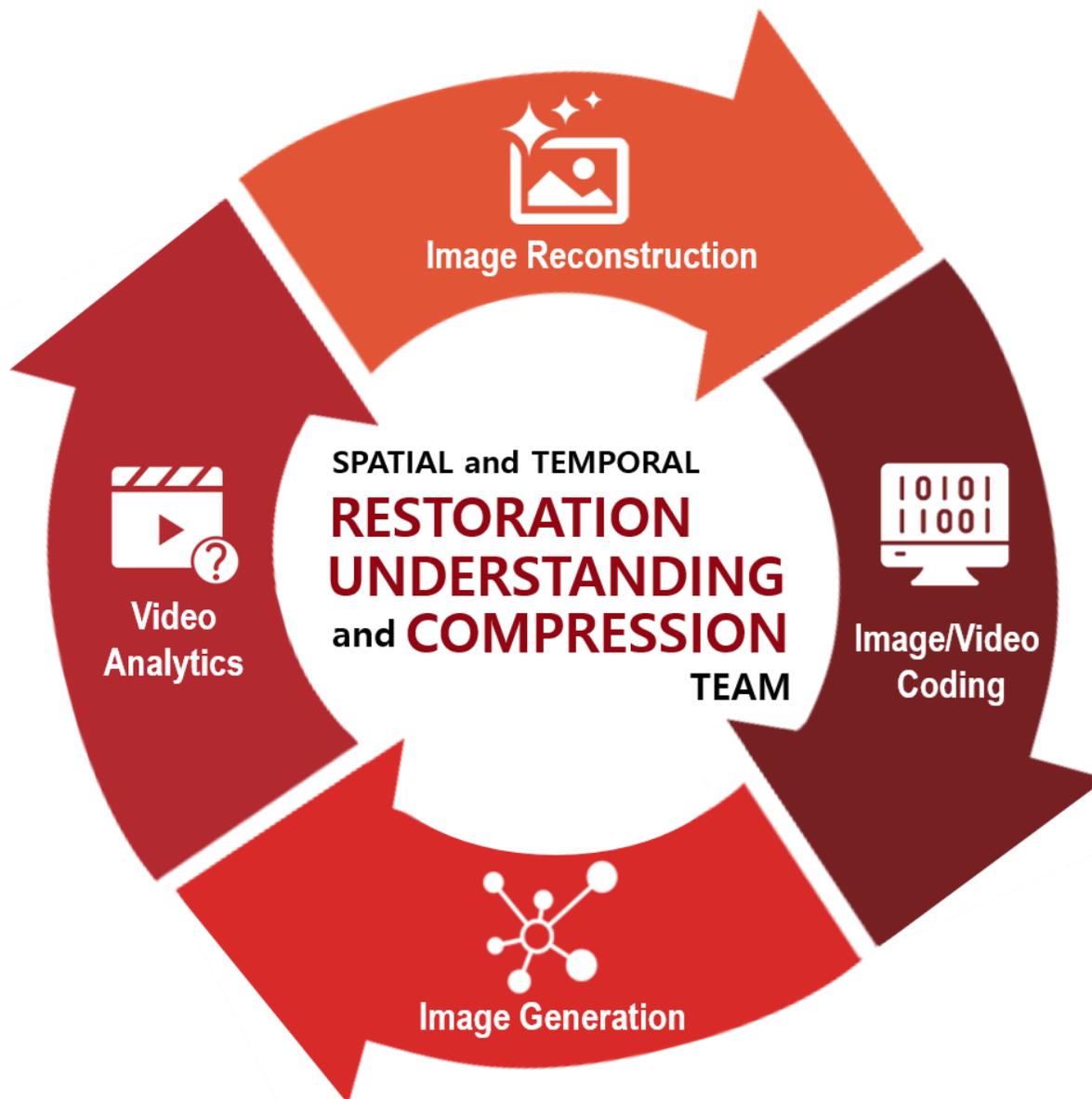
智能影像计算

北京大学 王选计算机研究所

数字视频研究室

Spatial and Temporal **Restoration**,
Understanding and **Compression** Team

- PI: **刘家瑛**
- 邮箱: liujiaying@pku.edu.cn
- 网页: <http://www.wict.pku.edu.cn/struct/>



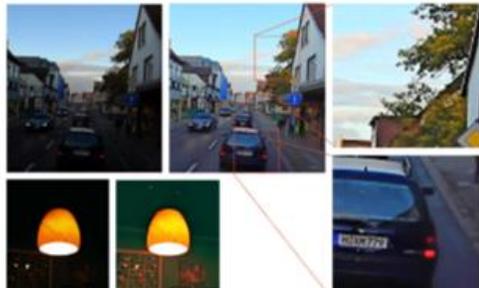


- 100+ Research Paper, including IJCV, IEEE TPAMI, TIP, TMM, TCSVT, CVPR, AAAI, ICCV, IJCAI
- 70+ Patents applied, 50+ Issued



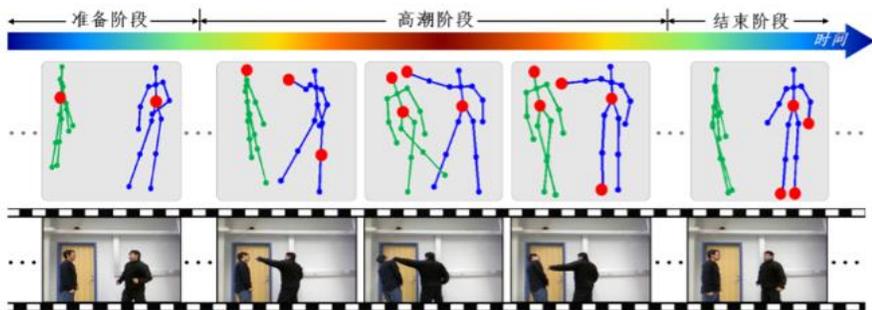
Intelligent Image/Video Editing

- Image / Typography Stylization
- Image / Video Restoration
- Image / Video Enhancement, *i.e.* derain, dehaze, contrast enhance



动作识别

● 基于注意力机制的动作识别

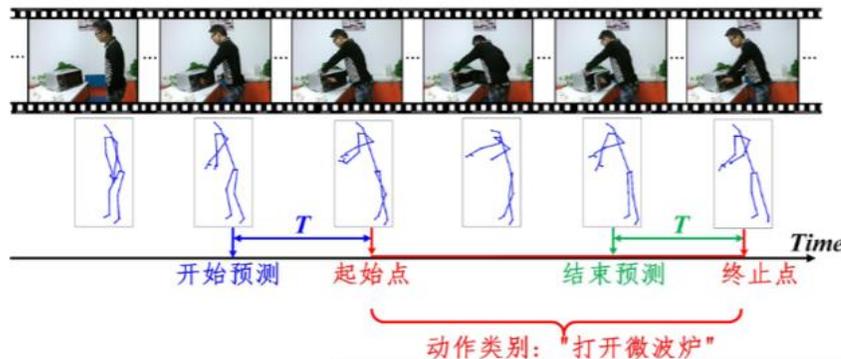


PKU-MMD Action Analysis Dataset

- 1 多模态**
 4种模态
- 2 多角度**
 3个拍摄角度
- 3 类内多样性**
 66位不同的拍摄主体
- 4 类间多样性**
 41个日常动作 + 10个交互动作

动作检测

● 基于循环神经网络的动作检测与预测



Video Action Analytics



代表性工作: 艺术字智能化设计

Awesome Typography: Statistics-Based Text Effects Transfer

Shuai Yang, Jiaying Liu, Zhouhui Lian, Zongming Guo, CVPR 2017



代表性工作: 艺术字智能化设计

Awesome Typography: Statistics-Based Text Effects Transfer

Shuai Yang, Jiaying Liu, Zhouhui Lian, Zongming Guo, CVPR 2017









和

Q

元

享

陈

R

爰

氣





和

Q

元

亨

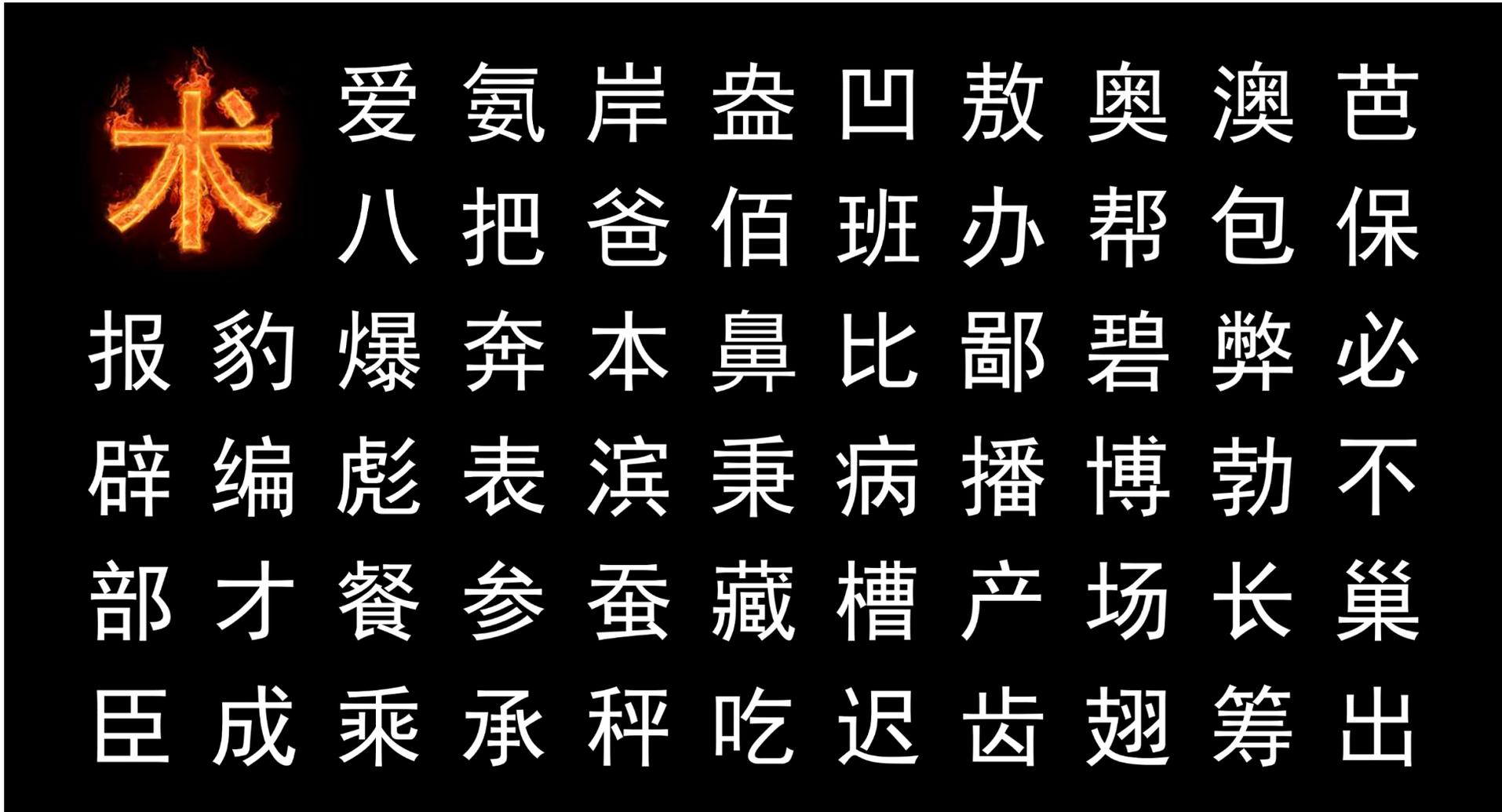
陈

R

爰

氣

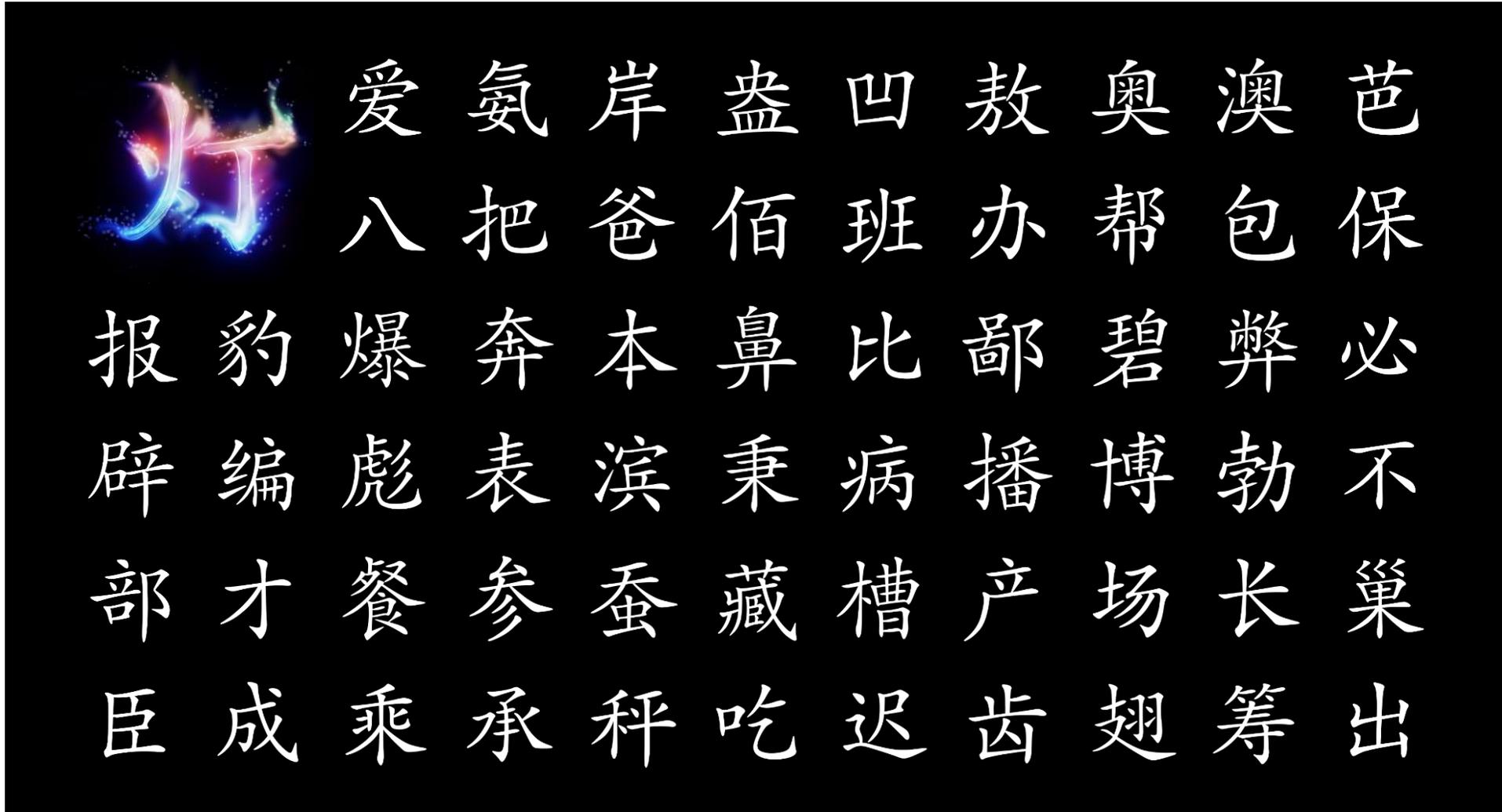
艺术字体库生成



艺术字体库生成



艺术字体库生成

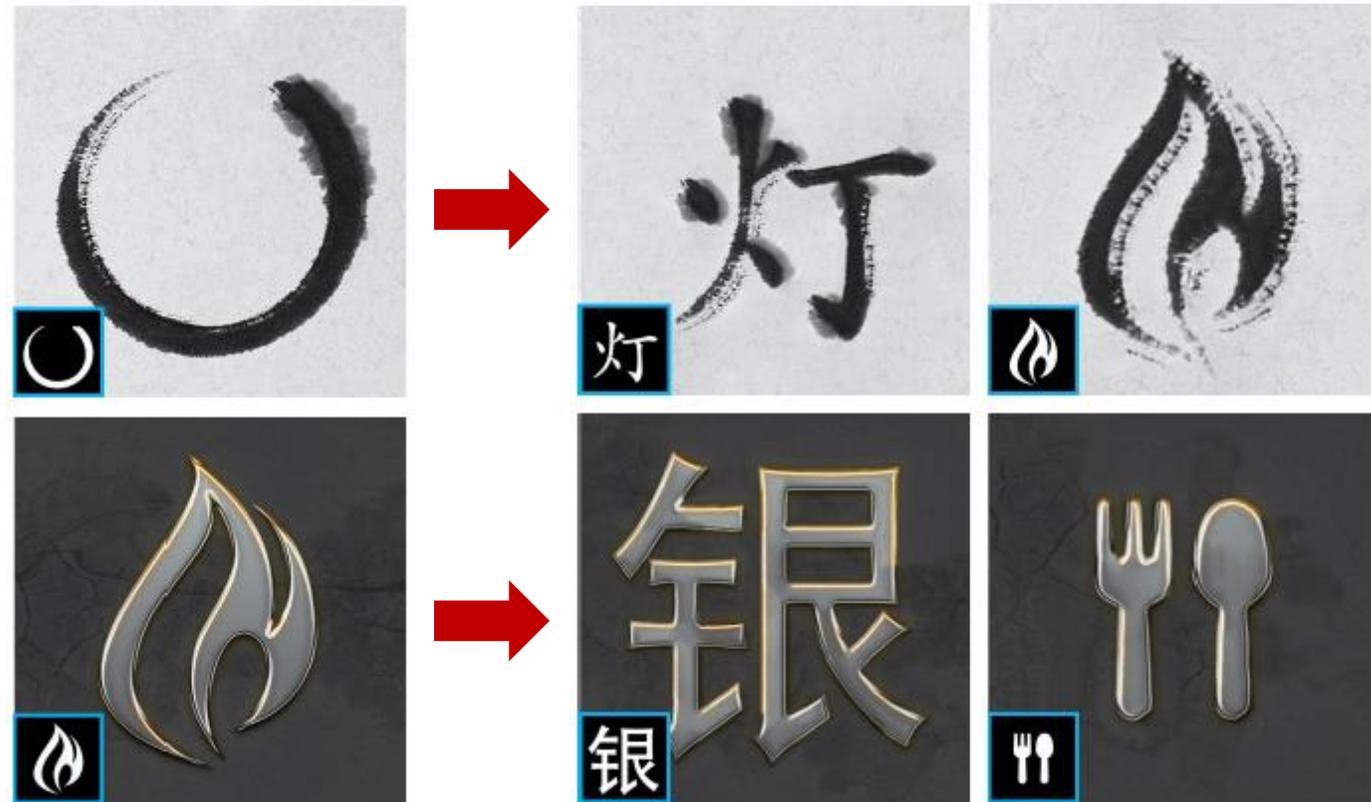


艺术字体库生成



应用场景扩展

- 基于笔画的图标/符号渲染
- 图标 → 图标/文字



无监督艺术字生成与图文设计

Context-Aware Text-Based Binary Image Stylization and Synthesis

Shuai Yang, Jiaying Liu, Wenhan Yang, and Zongming Guo, ACM MM 2018/IEEE TIP 2019



实验结果: 不同语言/字体间转换



不同语言的艺术字生成



不同字体的艺术字生成

实验结果: 图文设计

- S' 来自背景图像



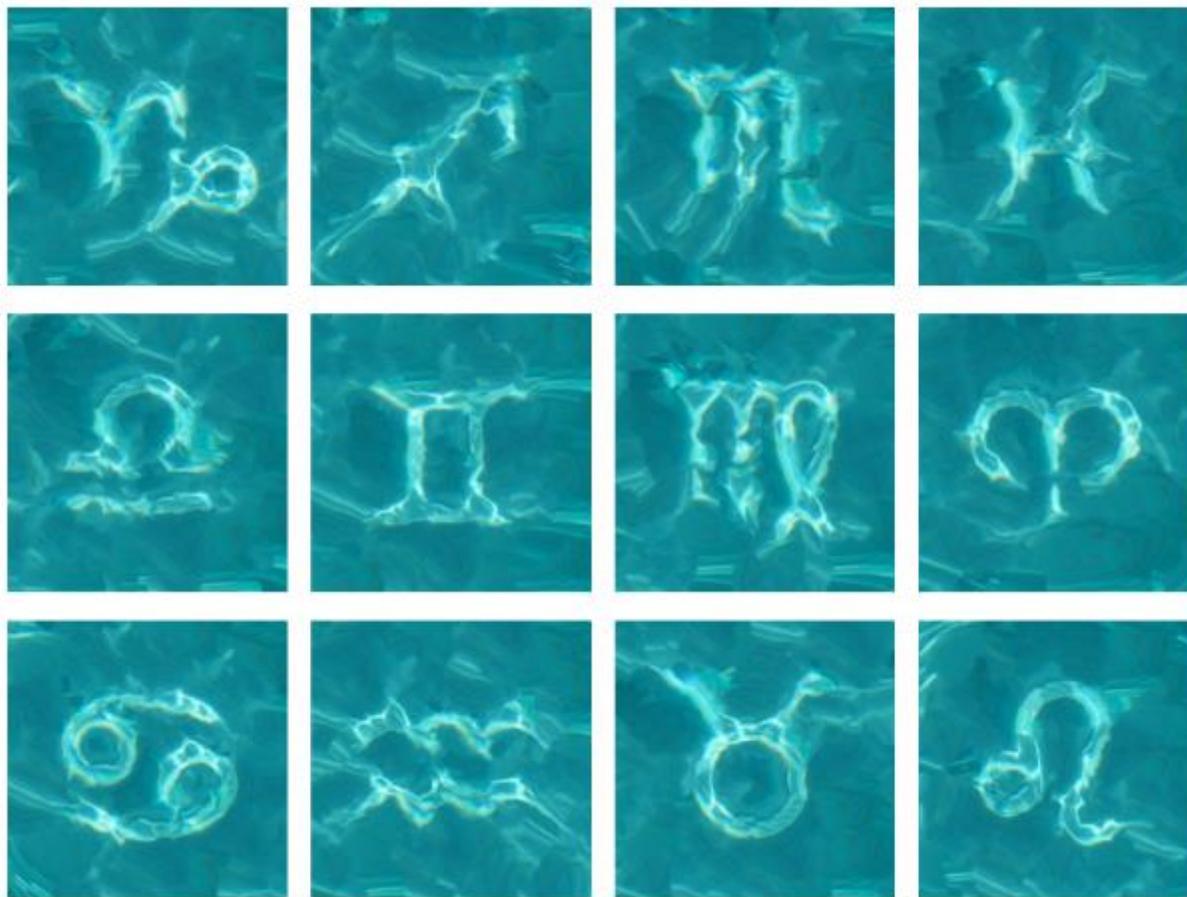
实验结果: 图文设计

- S' 来自其他风格图像



其他艺术化表达

- Symbol Design



其他艺术化表达

- Icon Rendering

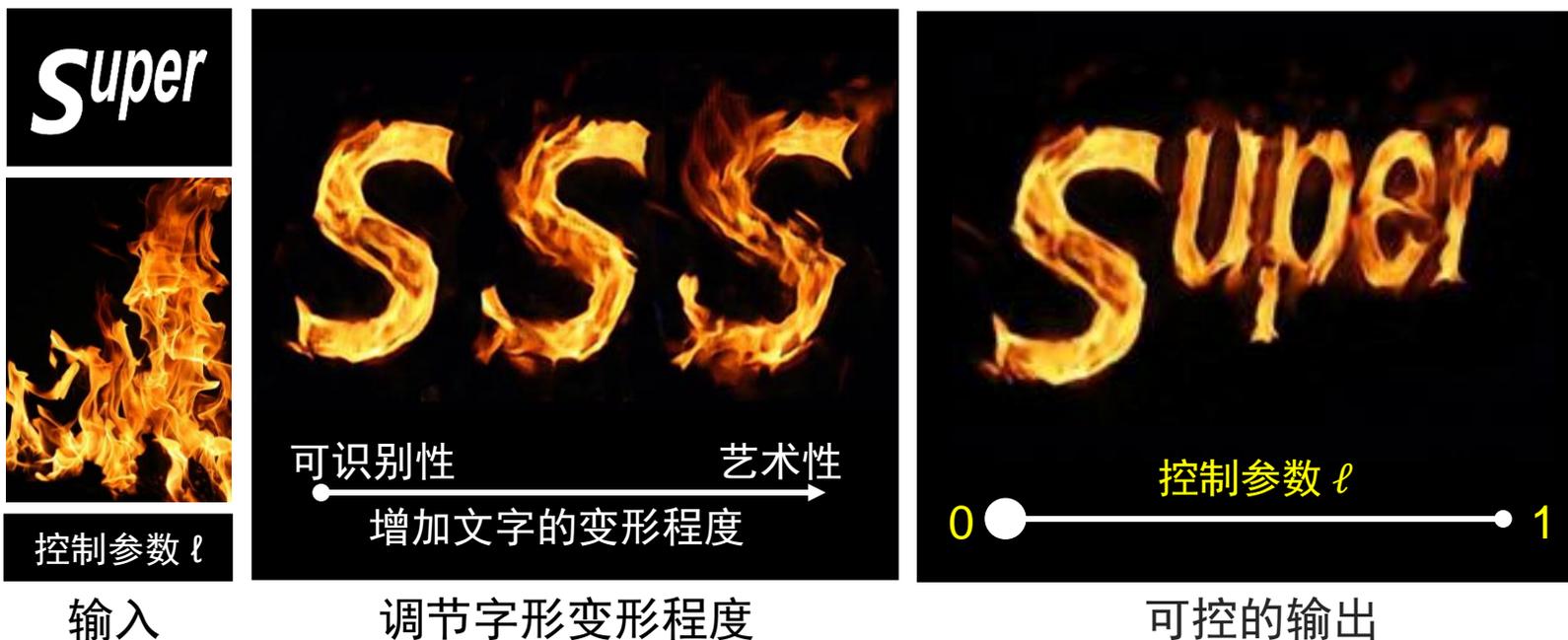


Rendering emoji icons with the painting style of Van Gogh using “The Starry Night”

风格程度可控的艺术字生成

Controllable Artistic Text Style Transfer via Shape-Matching GAN

Shuai Yang, Zhangyang Wang, Zhaowen Wang, Ning Xu, Jiaying Liu and Zongming Guo, ICCV 2019



实验结果: 多尺度结构迁移结果



参考风格图

MAPLE

目标文字



调节字形变形程度

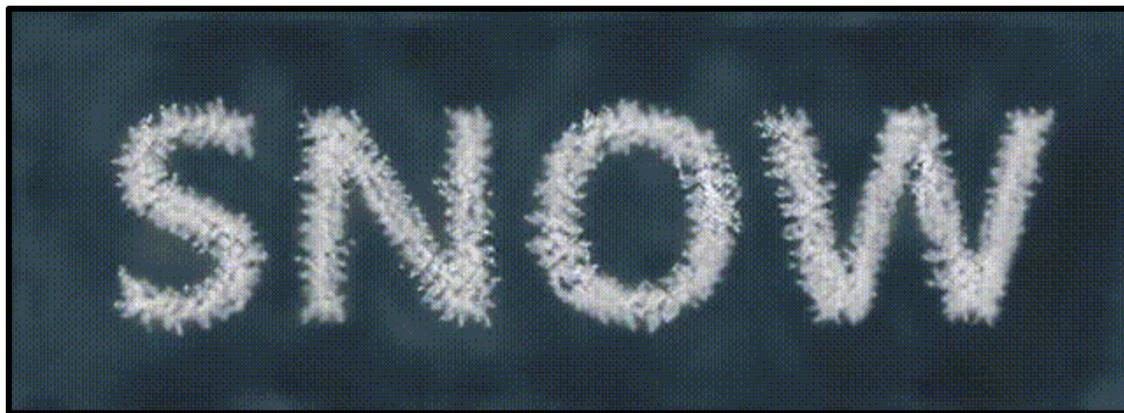
实验结果: 多尺度结构迁移结果



参考风格图

SNOW

目标文字

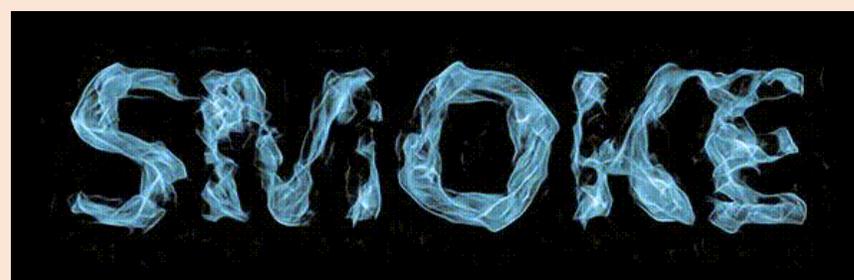
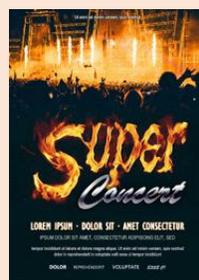


调节字形变形程度

实验结果: 应用场景

- 不同的应用场景

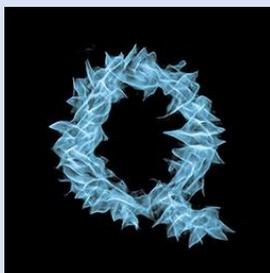
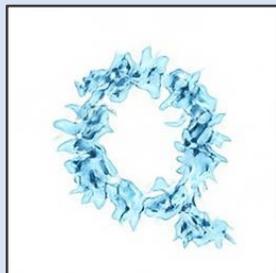
动态文字生成



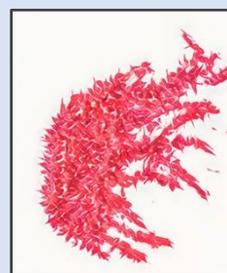
对输入添加随机噪声生成随机结果

对输入添加连续变化的噪声

多样化的结构/纹理混搭



图标/符号风格化





WORLD ARTIFICIAL INTELLIGENCE
CONFERENCE
2023世界人工智能大会

智能感知与媒体计算

Intelligent Visual Computing

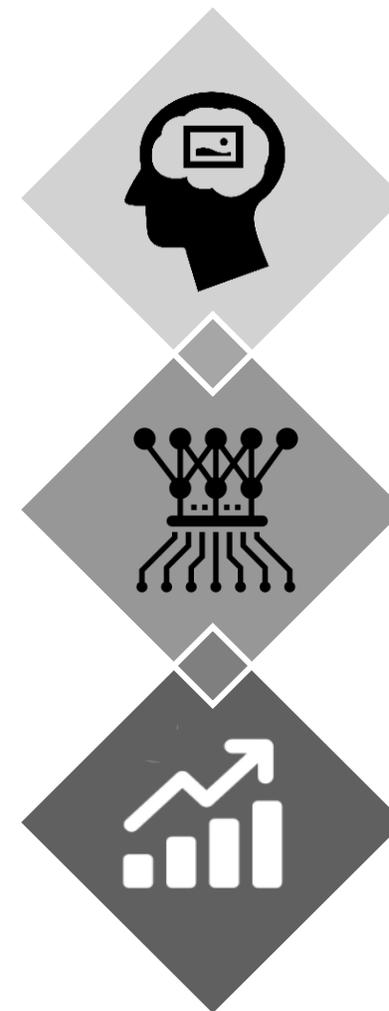
刘家瑛 北京大学



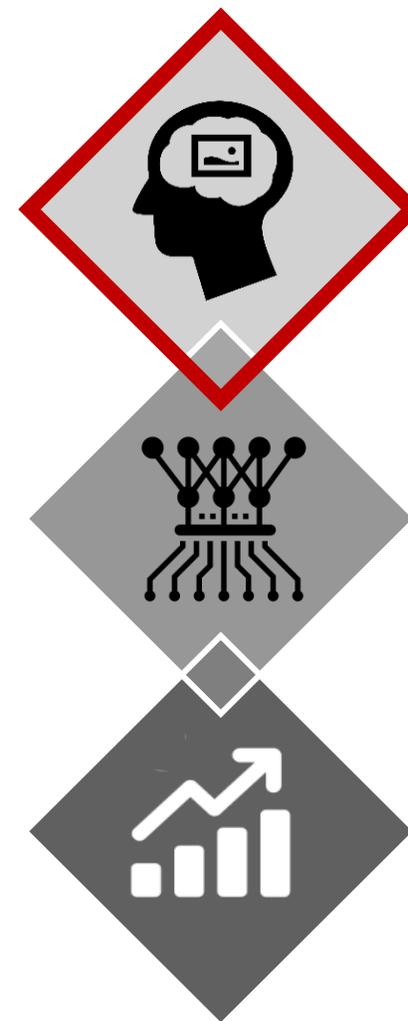
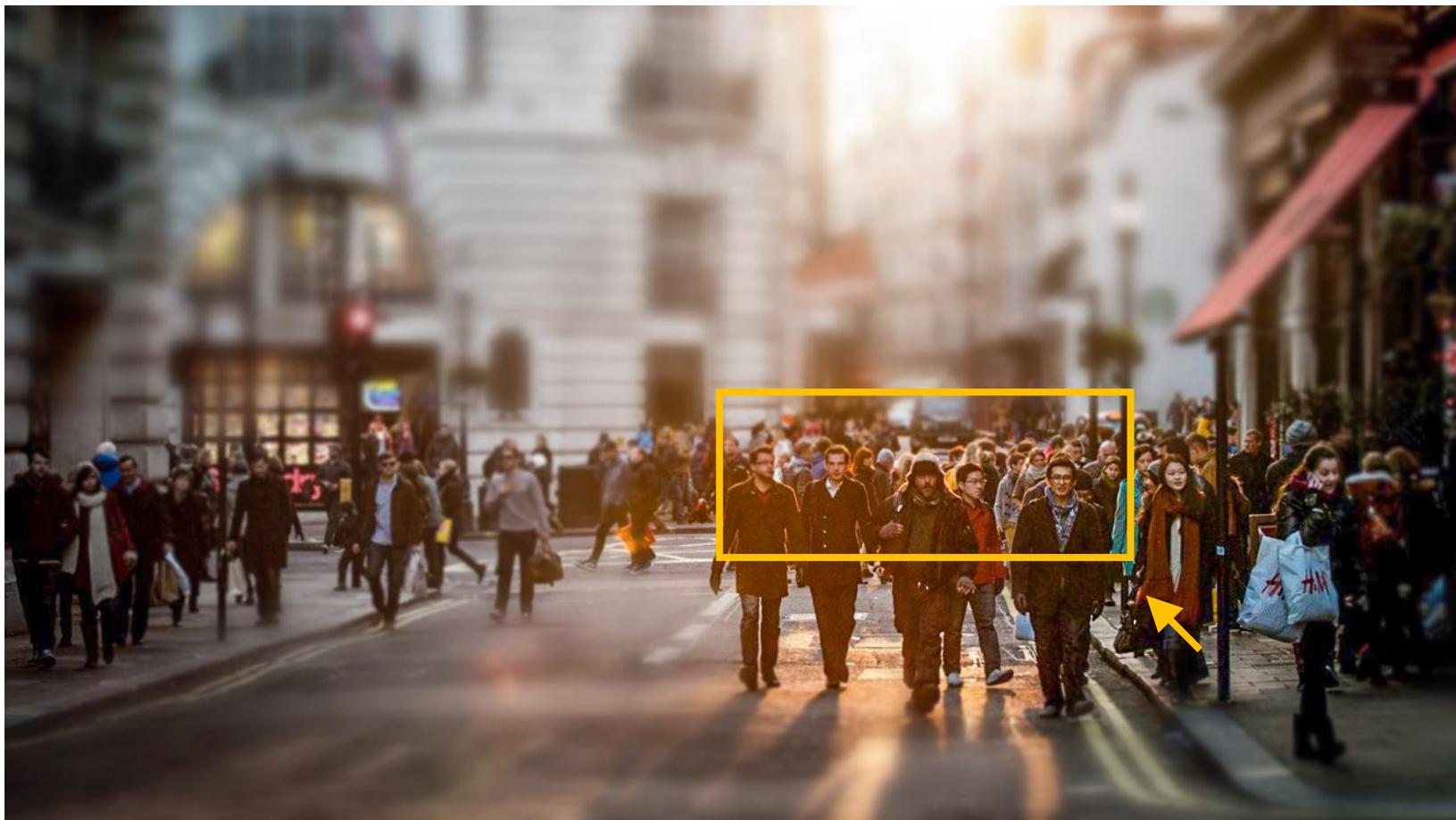
北京大学
PEKING UNIVERSITY



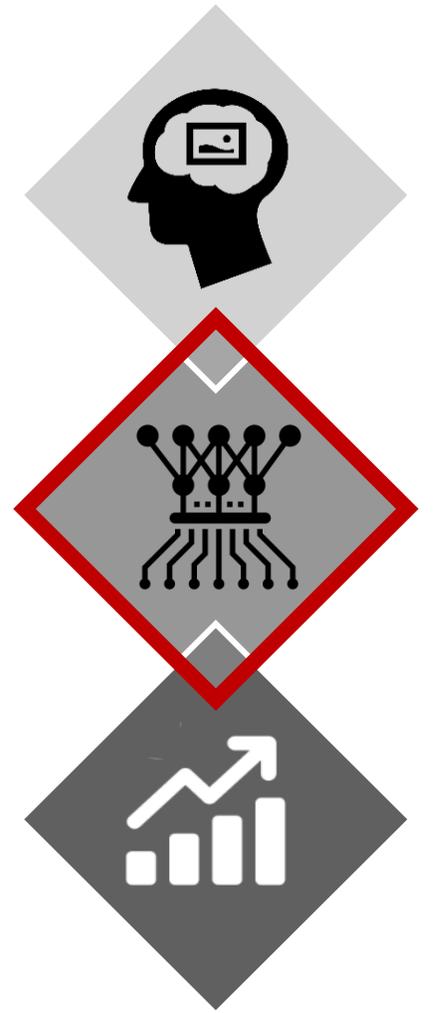
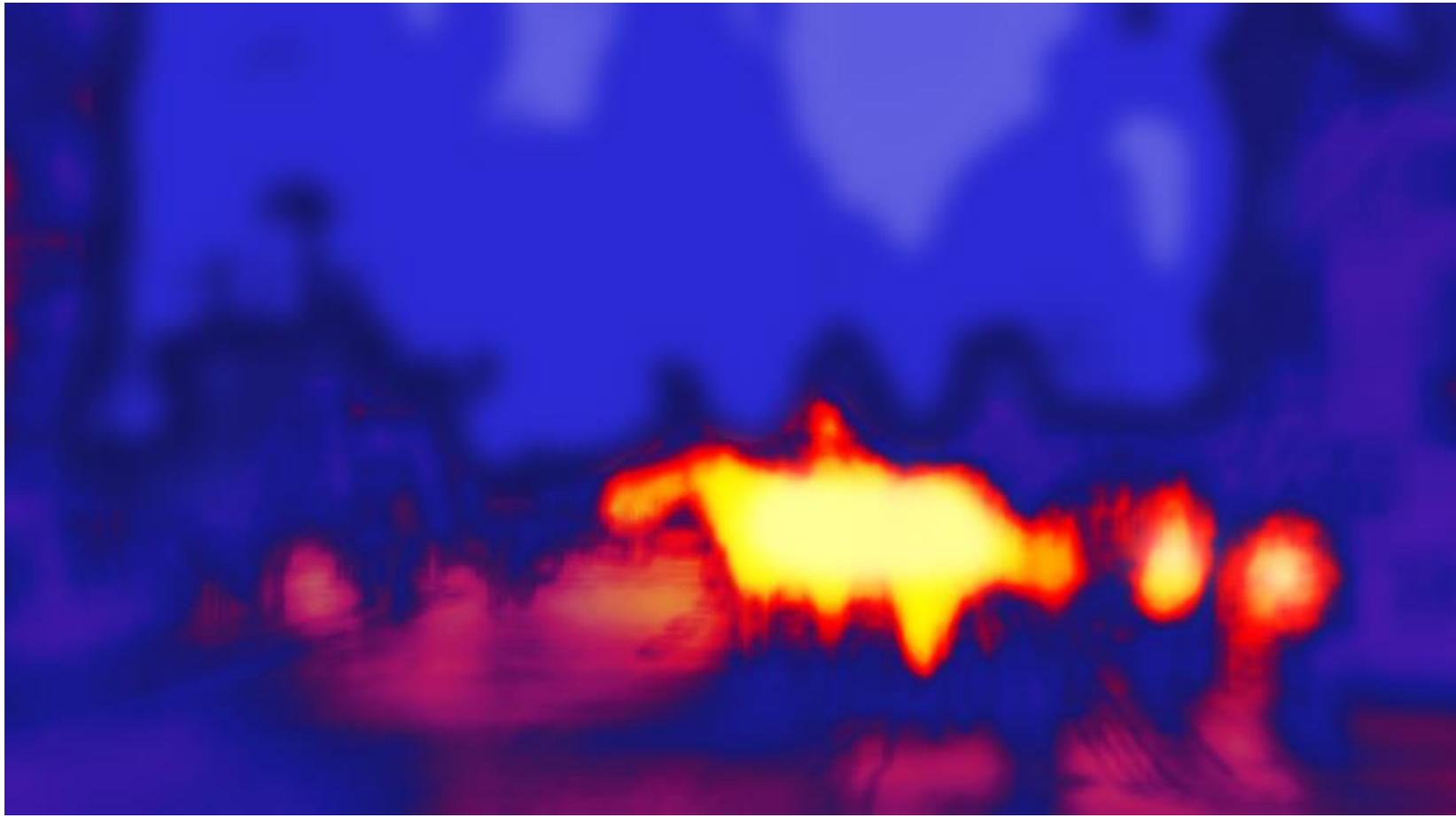
SCENE



WHAT HUMANS SEE



MACHINE FEATURES



MACHINE ANALYTICS

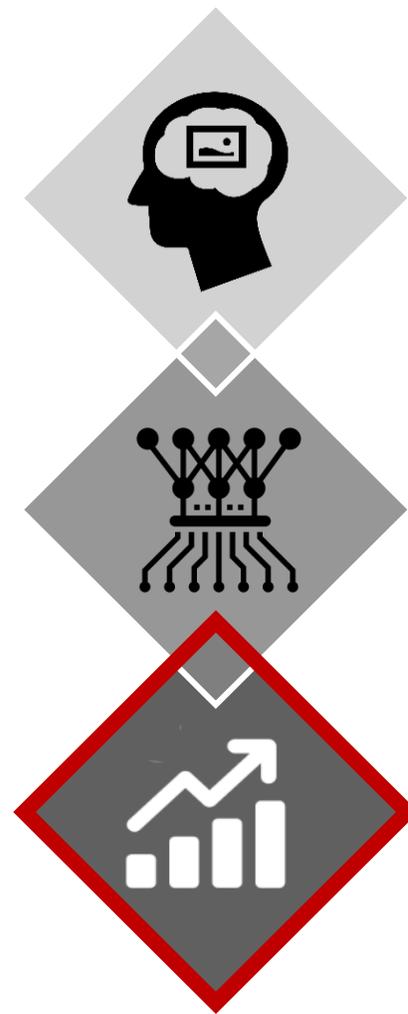
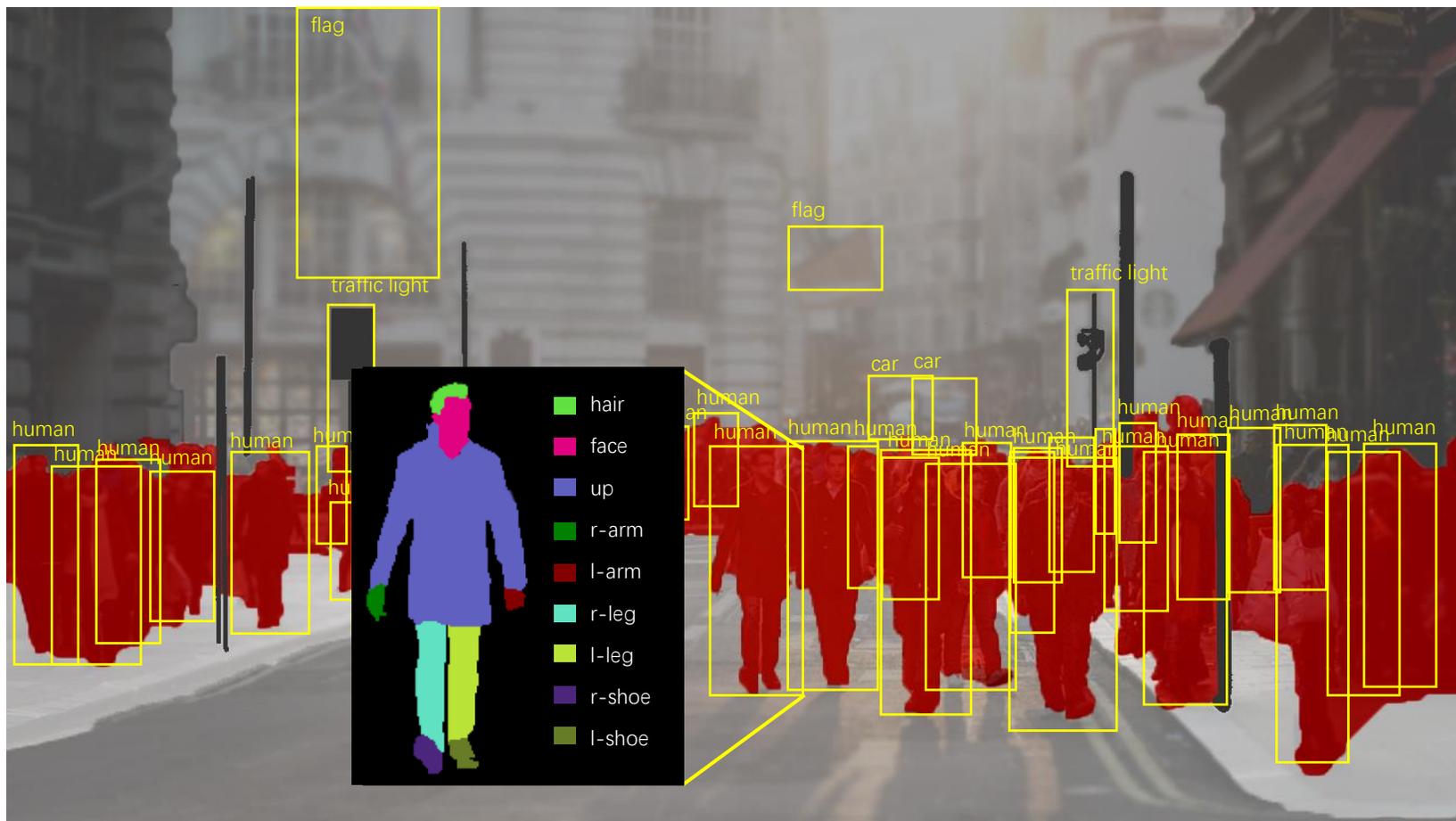


IMAGE CODING FOR WHOM?

▷ HIGH-LEVEL FEATURES
FOR MACHINE VISION

▷ RECONSTRUCTED IMAGES
FOR HUMAN VISION



MACHINE VISION

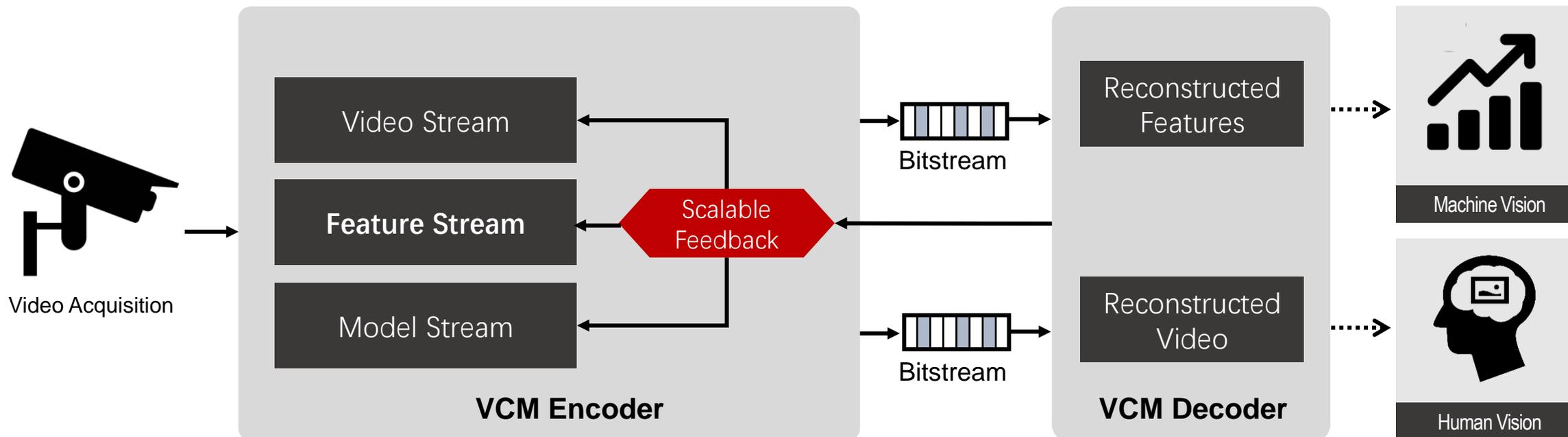
- 1) Feature Extraction
 - Redundant → Compact
 - **Compressed** Information
- 2) **Regression**
 - Feature → Label
 - Further Compressed

HUMAN VISION

- 1) Feature Extraction
 - Degraded → High-Quality
 - **Enhanced** Information
- 2) Guided **Reconstruction**
 - Feature → Image
 - Information Generation

IMAGE CODING NEXTGEN

- Scalable (according to utilizations)
- Efficient compression for joint human and machine vision



Ling-Yu Duan, Jiaying Liu, Wenhan Yang, Tiejun Huang, Wen Gao. Video Coding for Machines: A Paradigm of Collaborative Compression and Intelligent Analytics. arXiv:2001.03569, 2020

INFORMATION DENSITY SPECTRUM

- **Descriptor coding** for efficient machine vision analytics (low bit-rate)
- Sophisticated **video codecs** for improved human vision (high bit-rate)

▷ HIGH-LEVEL FEATURES FOR MACHINE VISION

▷ RECONSTRUCTED IMAGES FOR HUMAN VISION

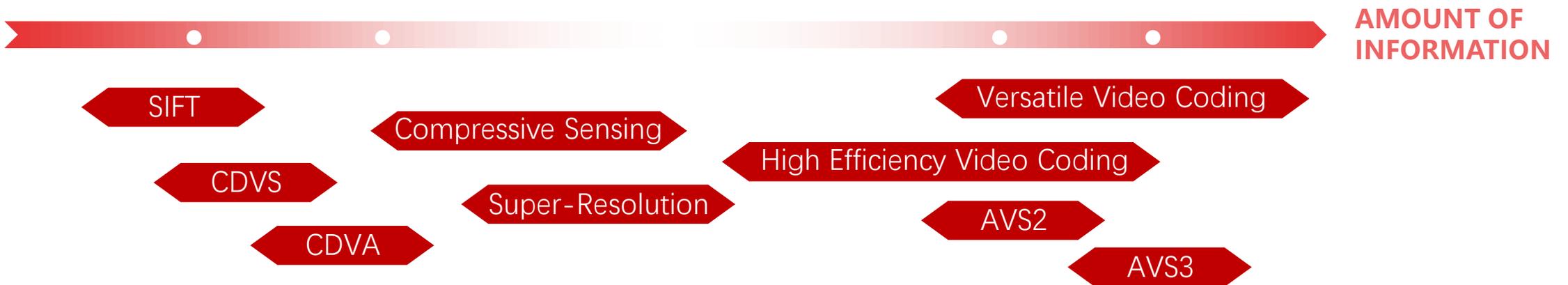
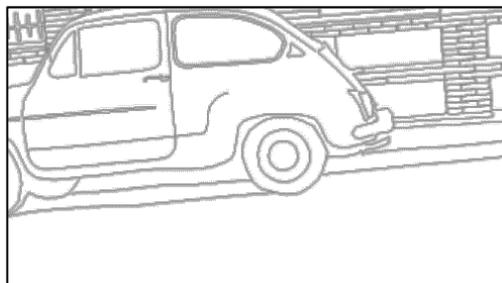
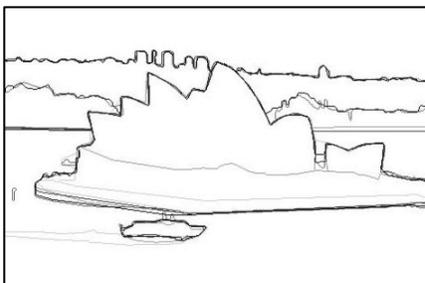
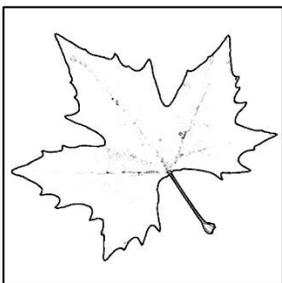


IMAGE REPRESENTATIONS



EDGES



△ PROS

- Efficient for structural information
- Maintain scalability
- Sparse and light-weight
- Supports smooth scaling

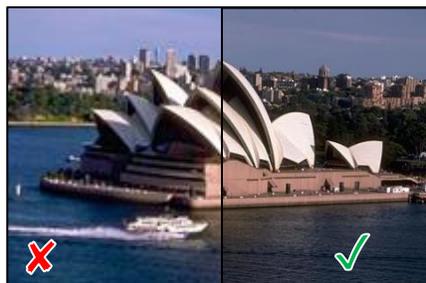
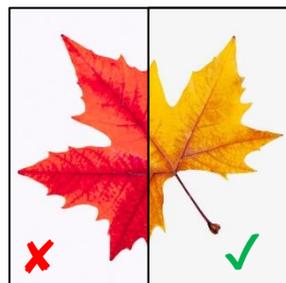
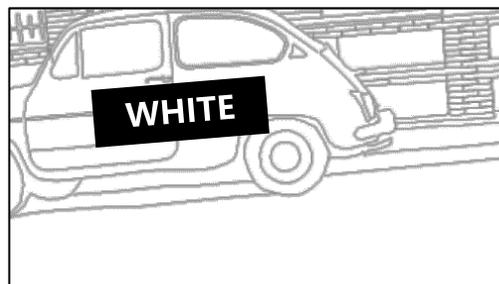
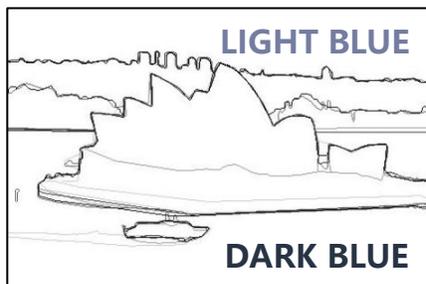
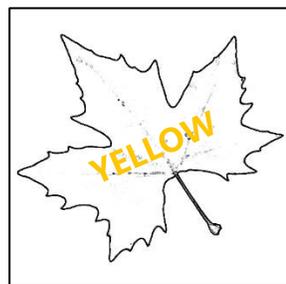
▽ CONS

- Inefficient for details in images
- Ambiguous in color

IMAGE REPRESENTATIONS



COLOR



△ PROS

- Avoid color ambiguity
- Sparse and compact
- Related to visual fidelity

▽ CONS

- Usually randomly distributed
- Inefficient for further compress

HUMAN FACES



Analytics of Faces

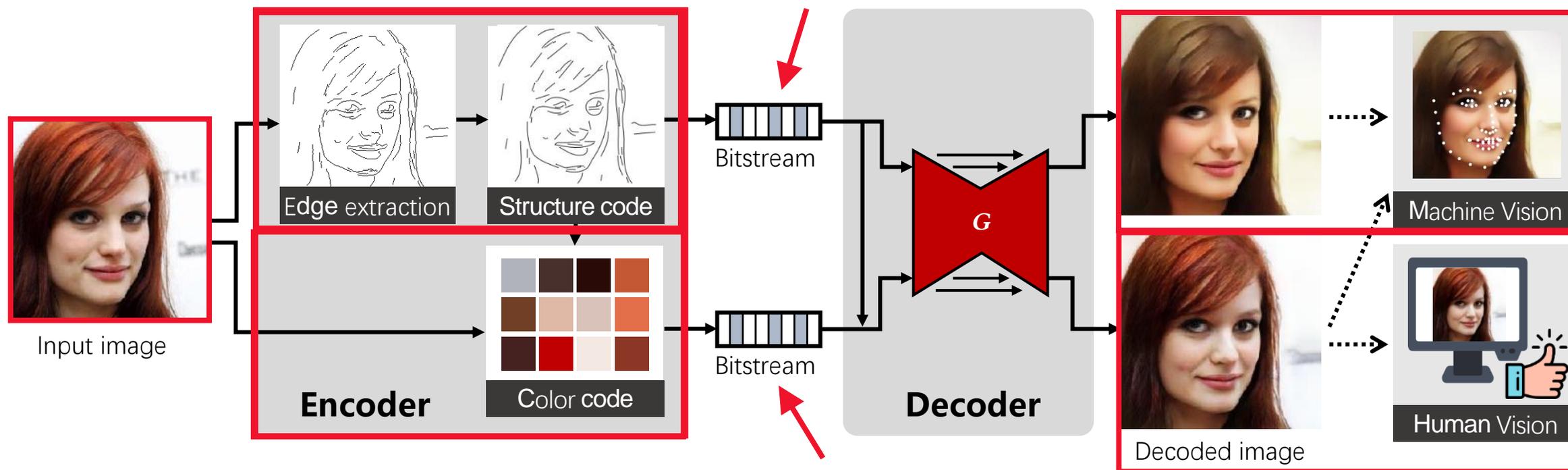
Faces are naturally salient area in images we are looking at.

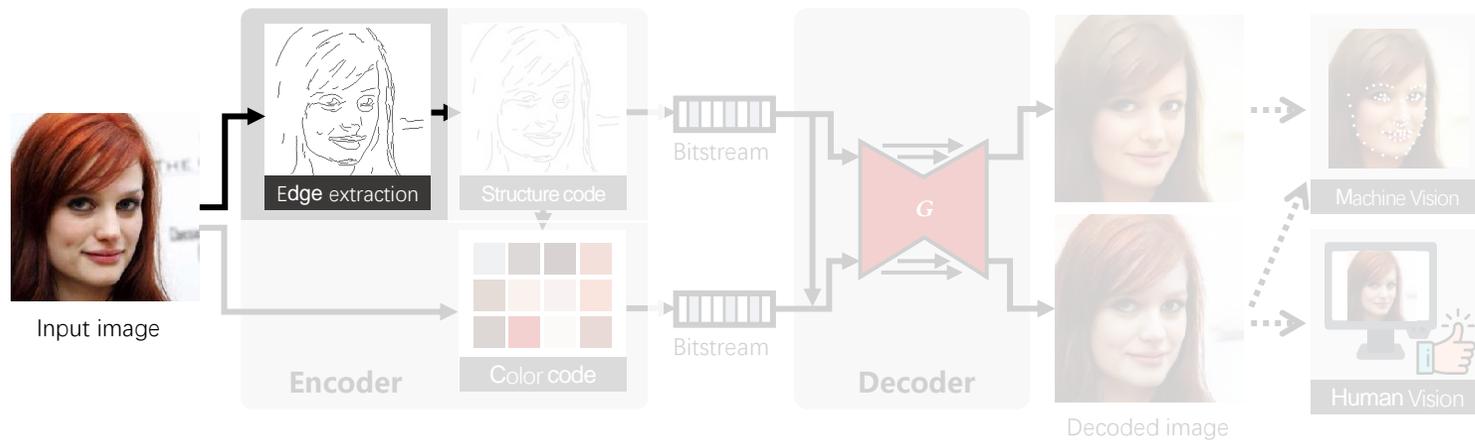
Machine vision systems to analysis faces have been widely developed.

It is the reflection of humanity in technology.

SCALABLE FRAMEWORK

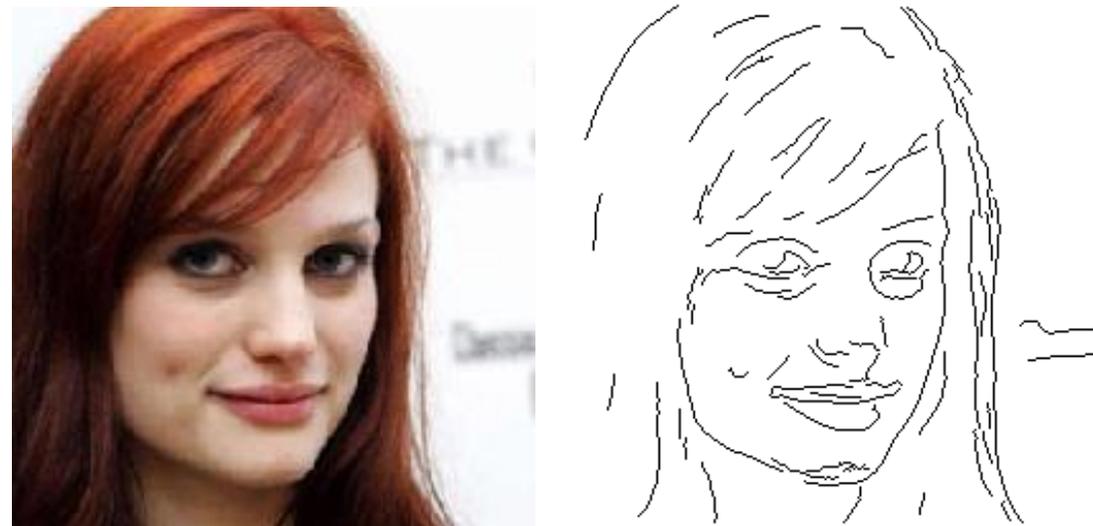
- Conceptual compression to achieve high quality with compact features
- Scalable bit-stream for different tasks
- Vectorized Edges + Sparse Pixels



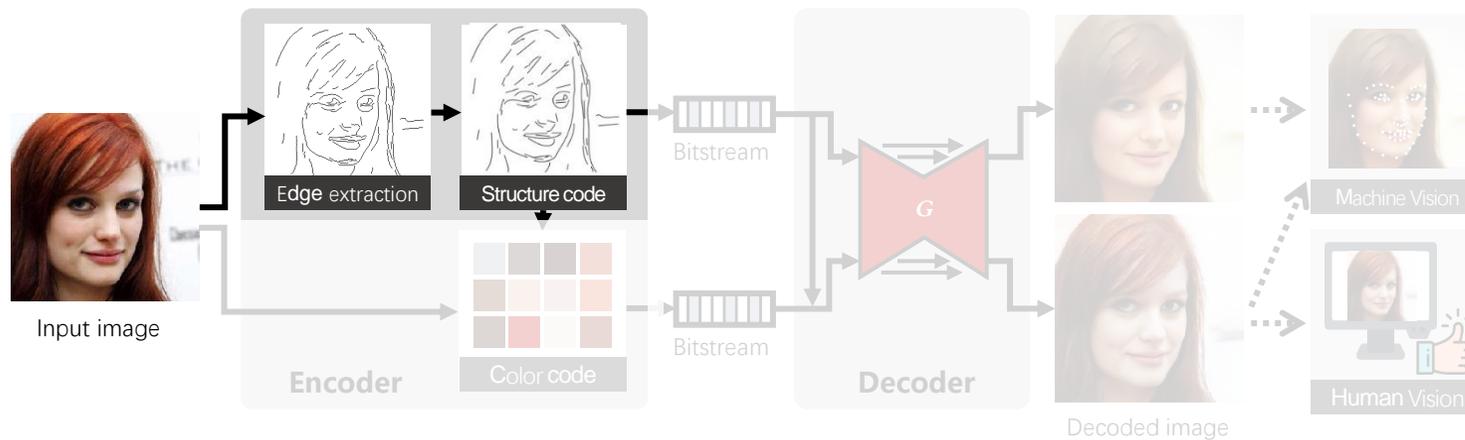


ENCODER • EDGE

- Edge detection via structured forests

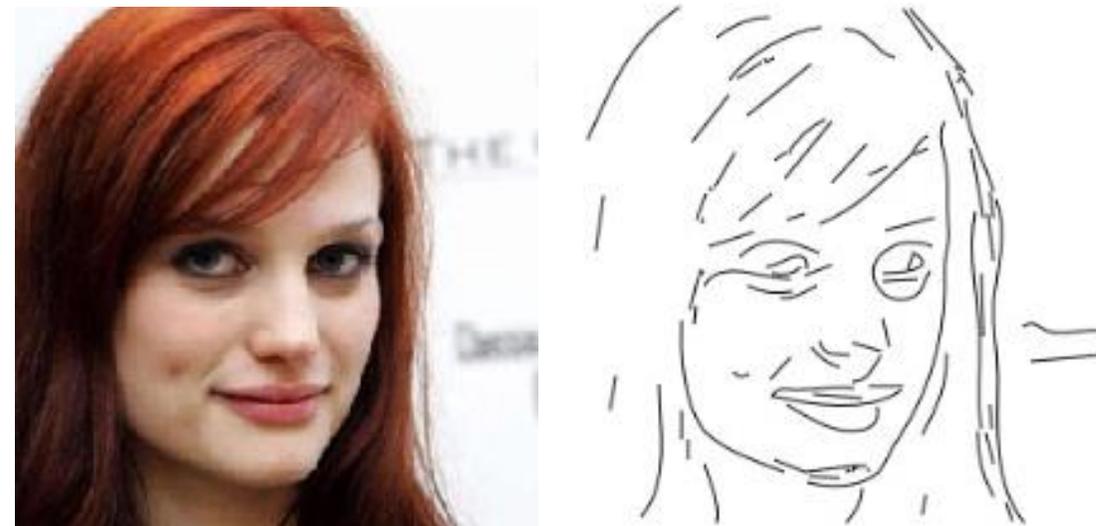


P. Dollar and C. L. Zitnick. Structured forests for fast edge detection. ICCV, 2013.

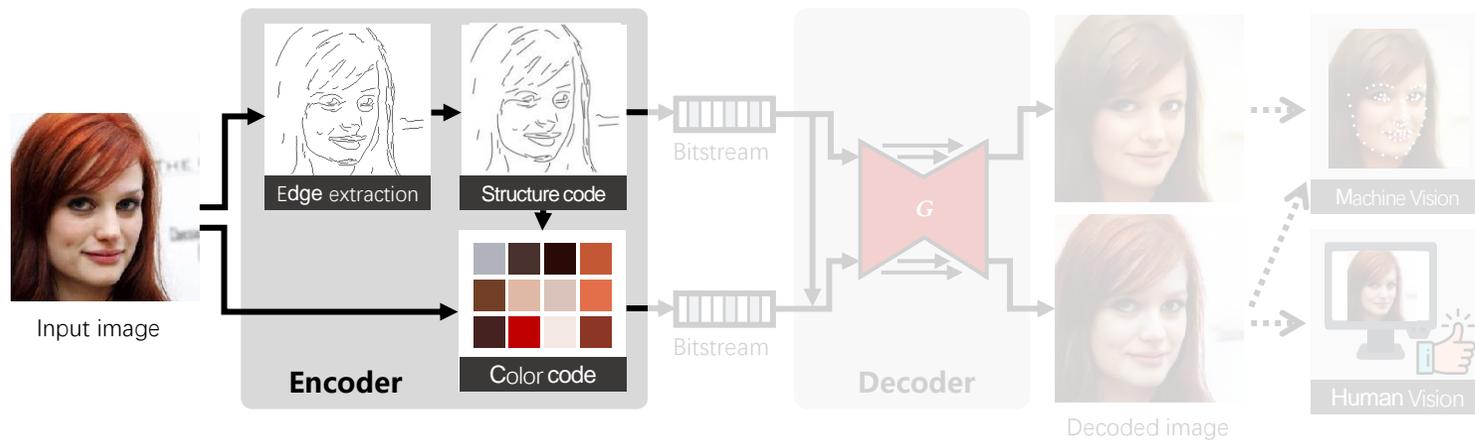


ENCODER • EDGE

- Edge detection via structured forests
- *AutoTrace* to convert edge pixels to vectorized representations
 - Represented by lines and curves
 - Short and trivial edges are screened
- Prediction for Partial Matching (PPM) to losslessly compress vectors



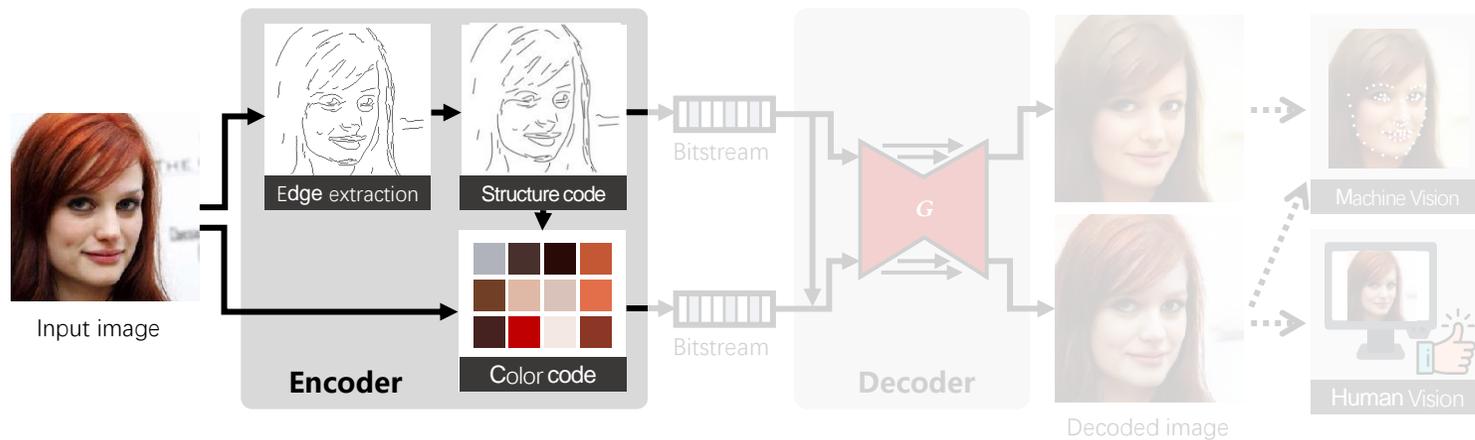
M.Weber. AutoTrace: a program for converting bitmap to vector graphic. 1998. <http://autotrace.sourceforge.net/>



ENCODER • COLOR

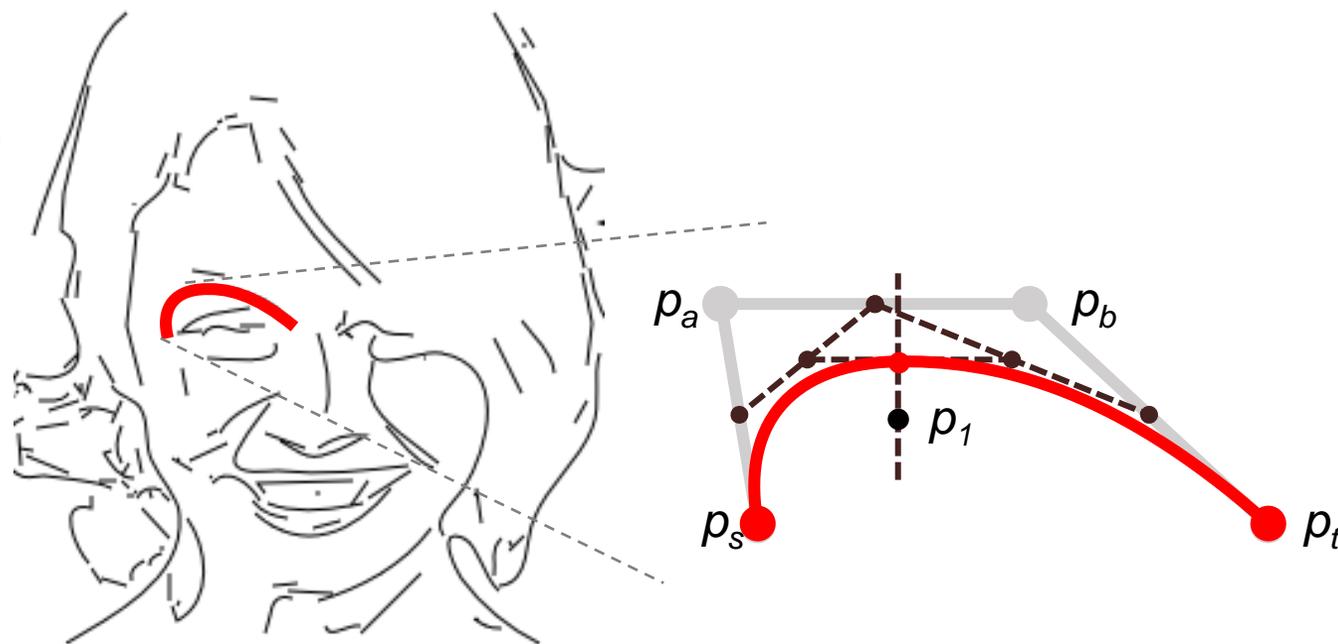
- Sparse pixels sampled according to edges
 - Segments: sample on both sides

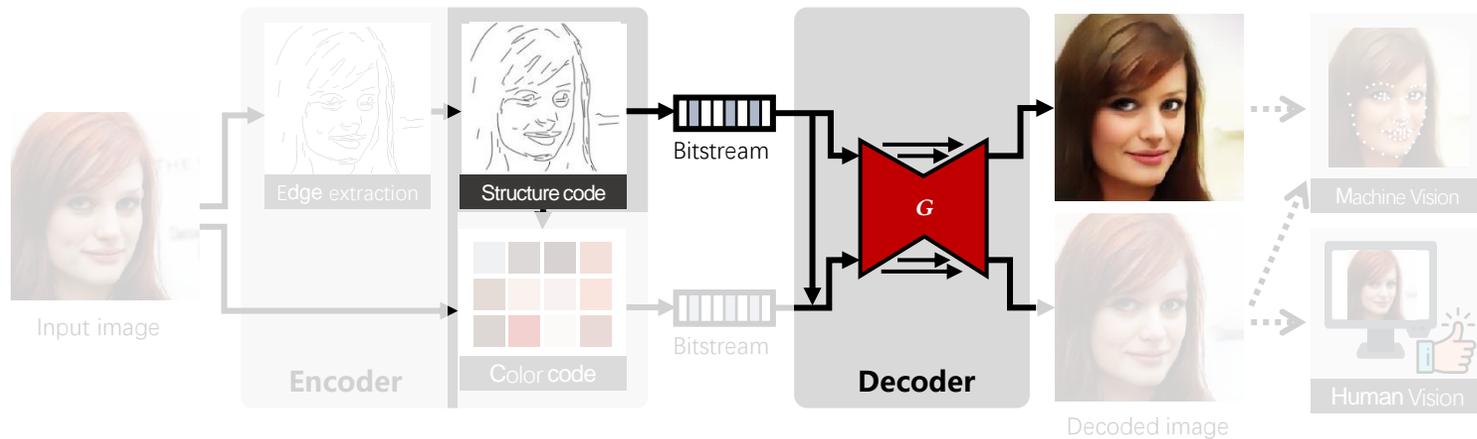




ENCODER • COLOR

- Sparse pixels sampled according to edges
 - Segments: sample on both sizes
 - Curves: sample on areas with steepest gradients

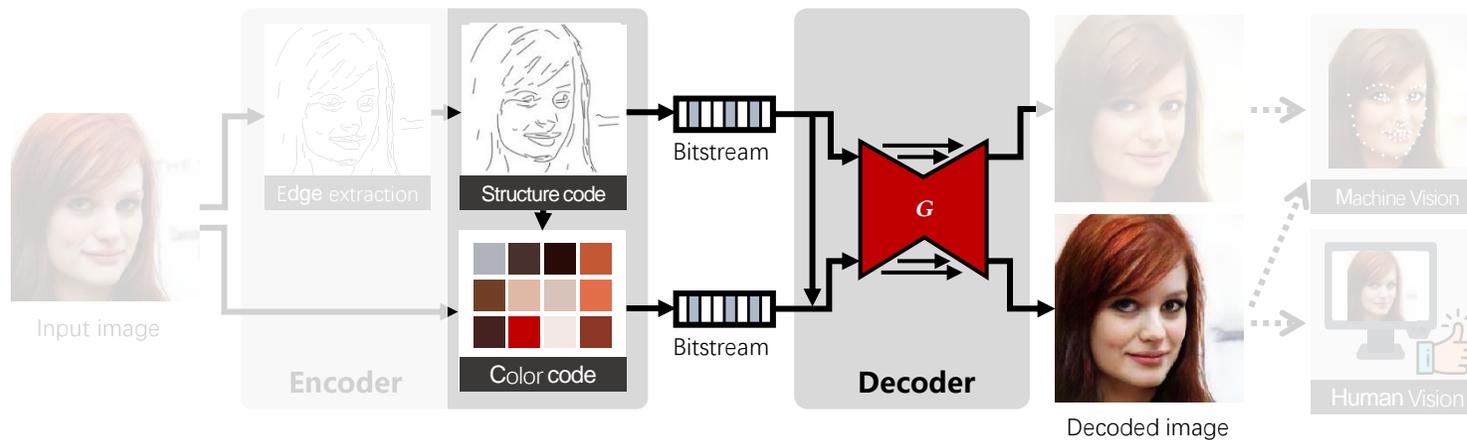




DECODER • MACHINE VISION

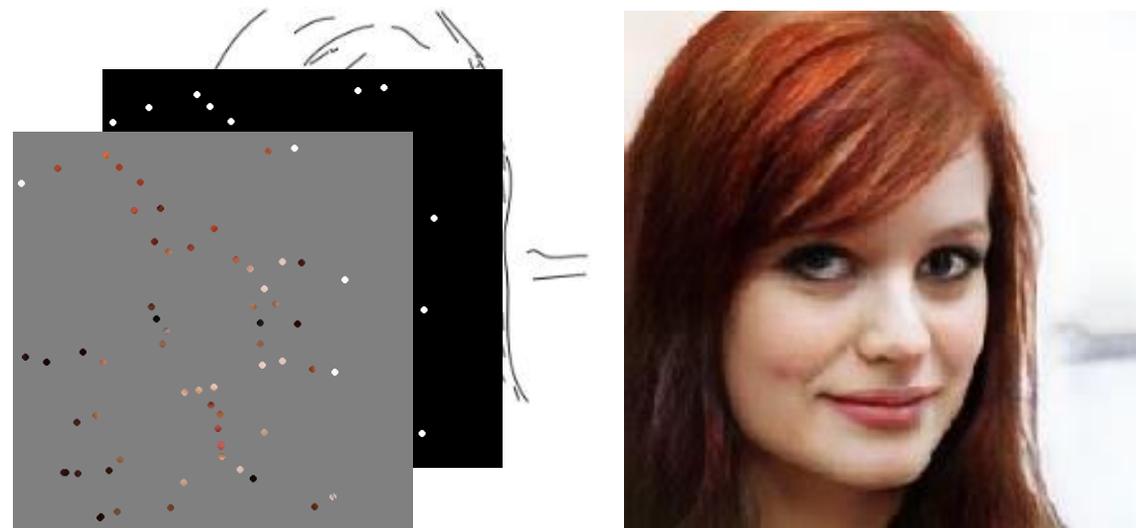
- Image-to-image translation
 - Render pixels with vectorized representations
 - Edge-to-RGB translation

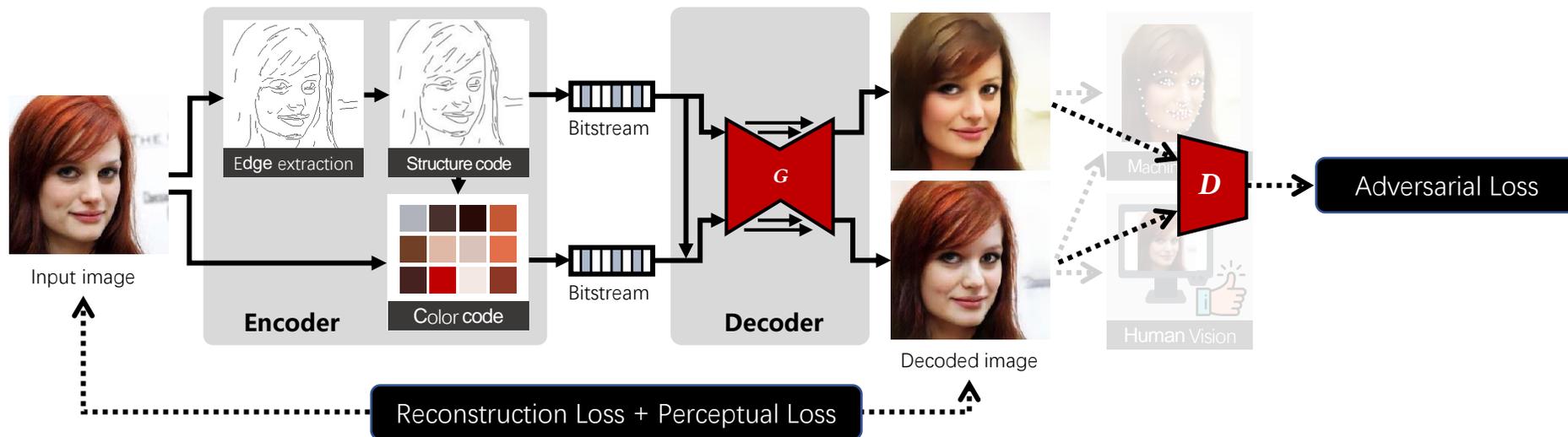




DECODER • HUMAN VISION

- Image-to-image translation
 - Render pixels with vectorized representations
 - Generate masks for completion synthesis
 - Image inpainting





LOSS FUNCTIONS

- Reconstruction Loss

$$\mathcal{L}_r = \mathbb{E}[\lambda_1 \| I_G - I \| + \lambda_2 \text{SSIM}(I_G, I)]$$

- Perceptual Loss

$$\mathcal{L}_p = \mathbb{E}[\lambda_3 \text{PERC}(I_G, I)]$$

- Adversarial Objective

$$\mathcal{L}_G = -\mathbb{E}[D(I_G, E, M)]$$

$$\mathcal{L}_D = \mathbb{E}[\text{ReLU}(\tau + D(I_G, E, M))] + \mathbb{E}[\text{ReLU}(\tau - D(I, E, M))]$$



EXPERIMENTAL RESULTS

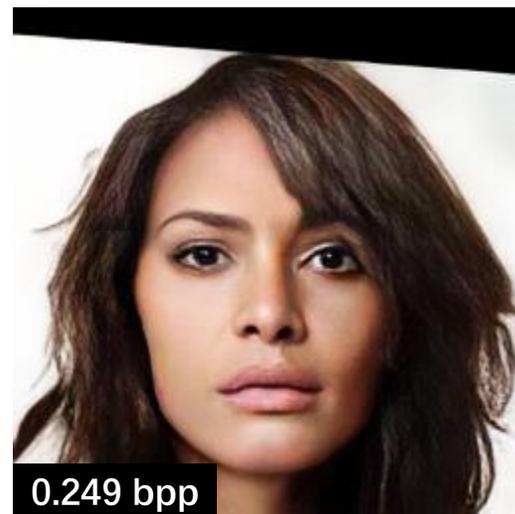
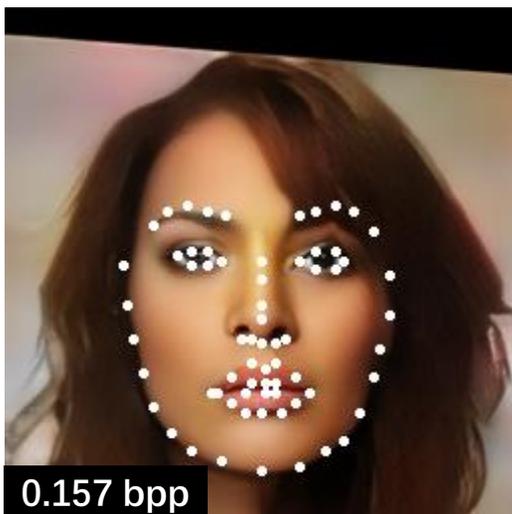
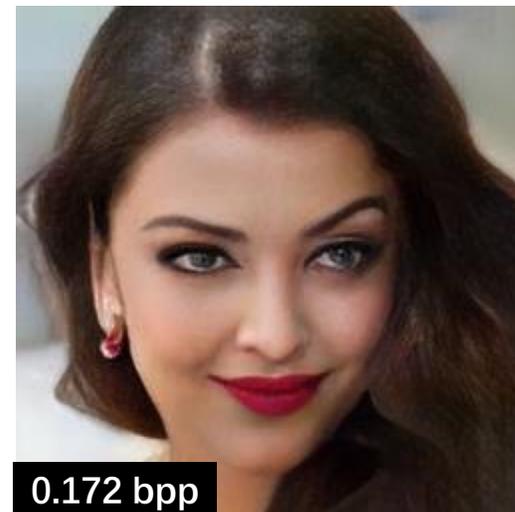
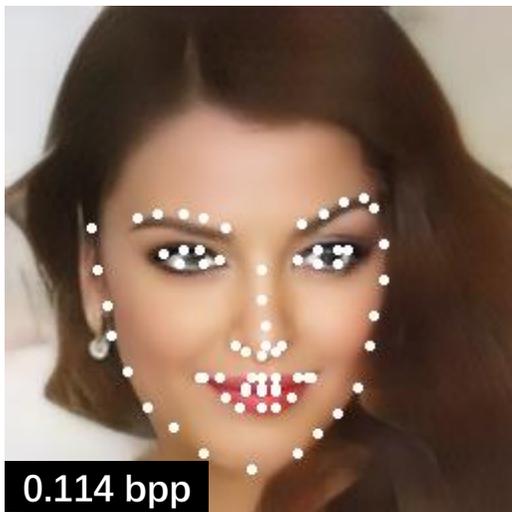
HUMAN VISION

Subjective preference survey.
Measuring fidelity and Aesthetics.

MACHINE VISION

Evaluate facial landmark detection.
Measuring information preservation.

SCALABLE OUTPUT



MACHINE VISION

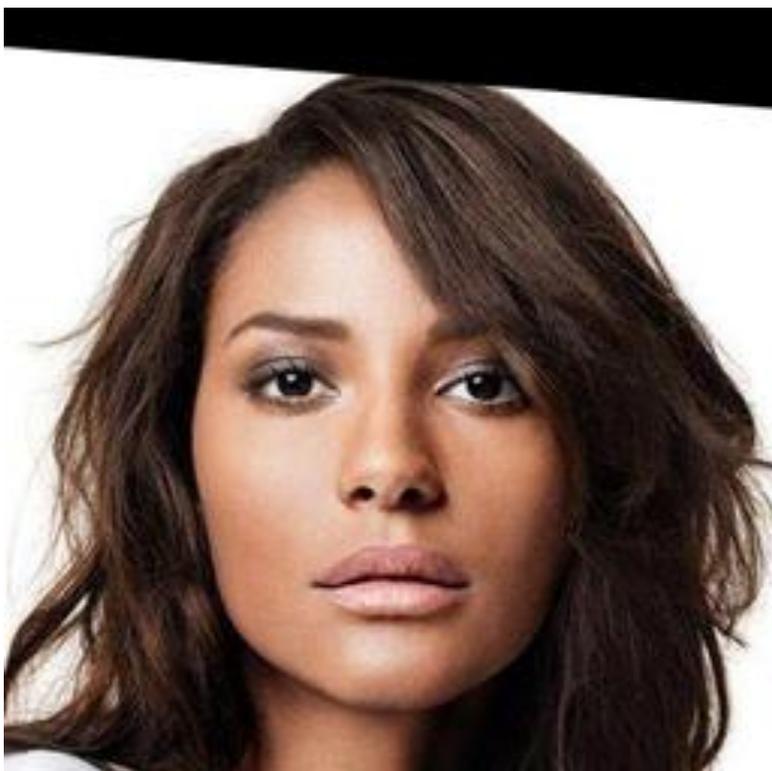


INPUT IMAGE

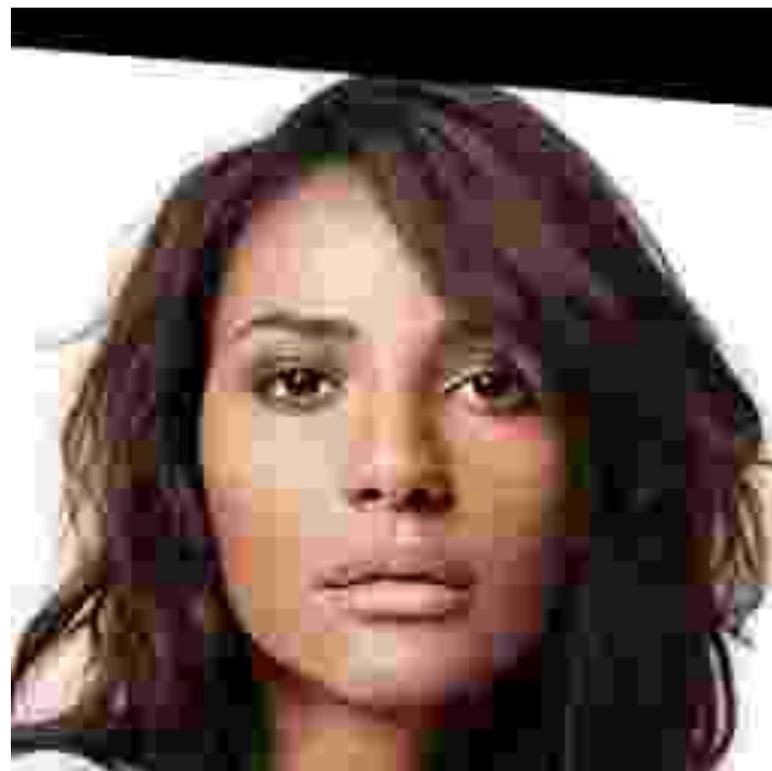


HUMAN VISION

HUMAN PERCEPTION

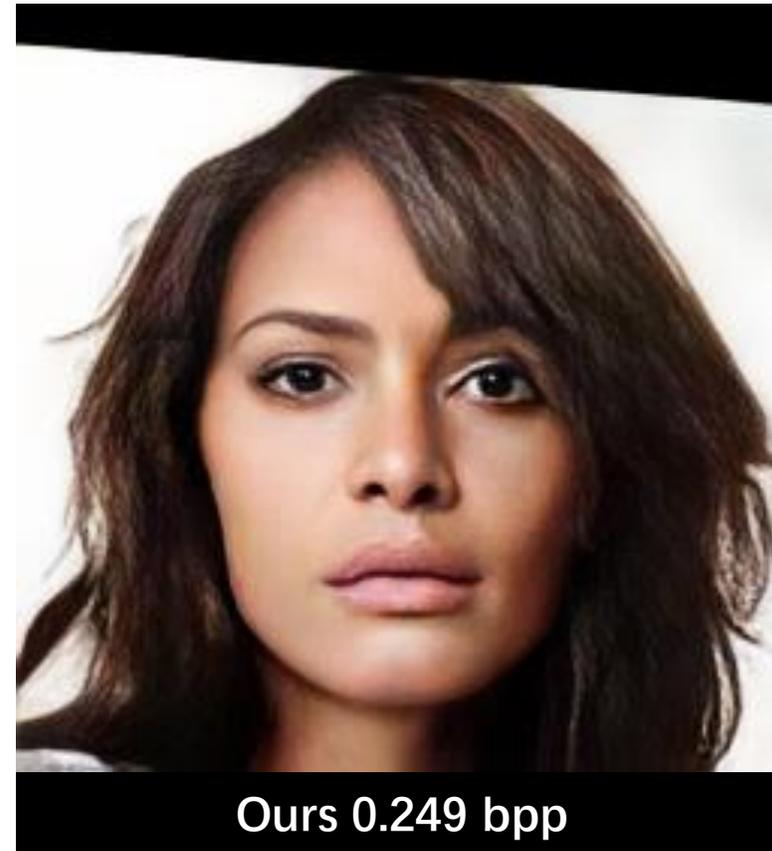


Input image



JPEG 0.266 bpp

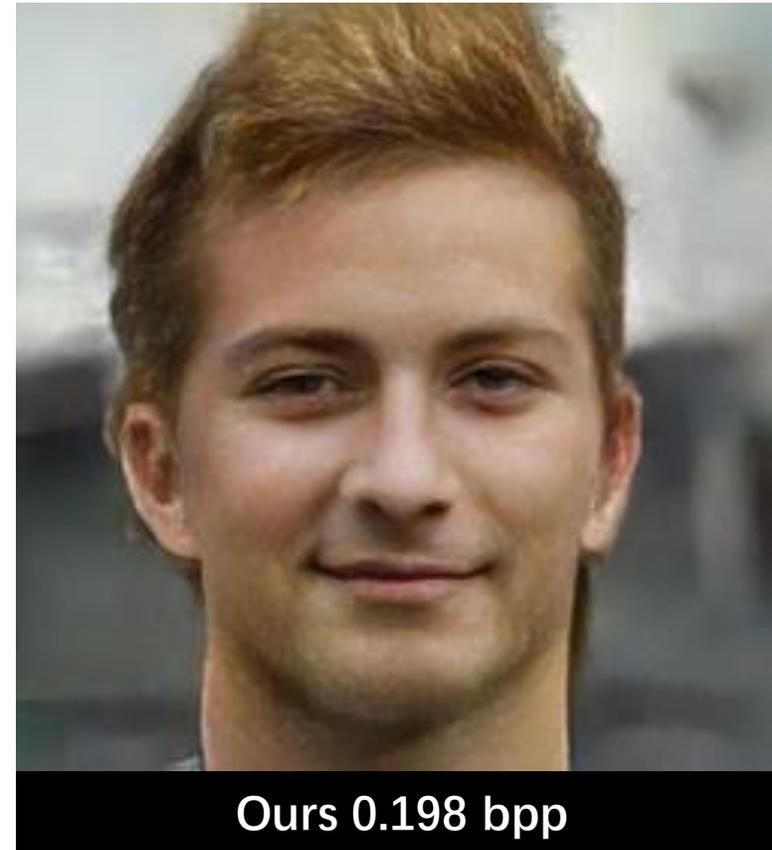
HUMAN PERCEPTION



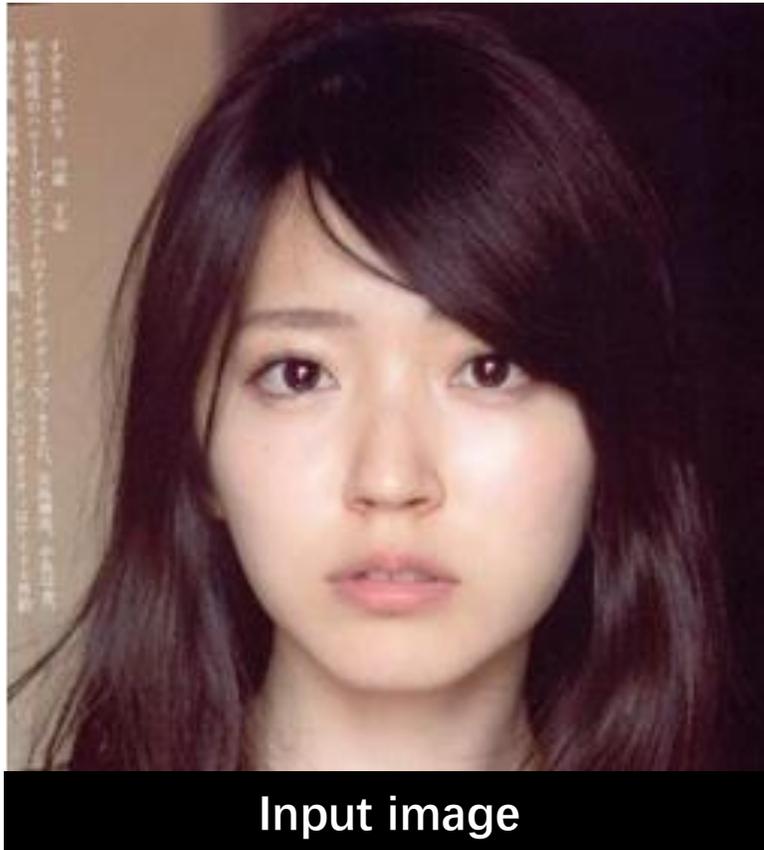
HUMAN PERCEPTION



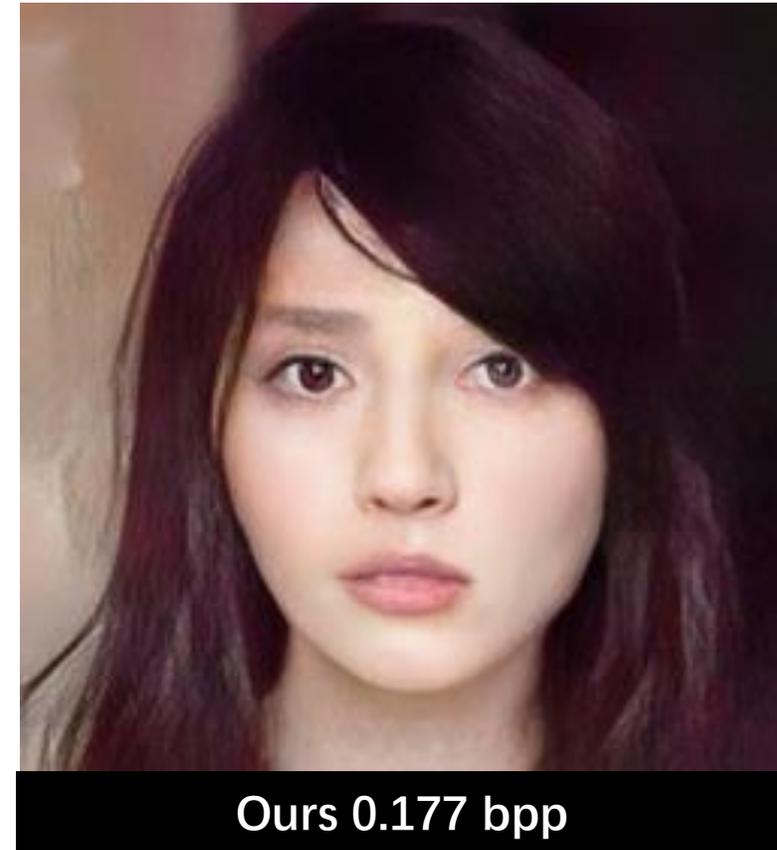
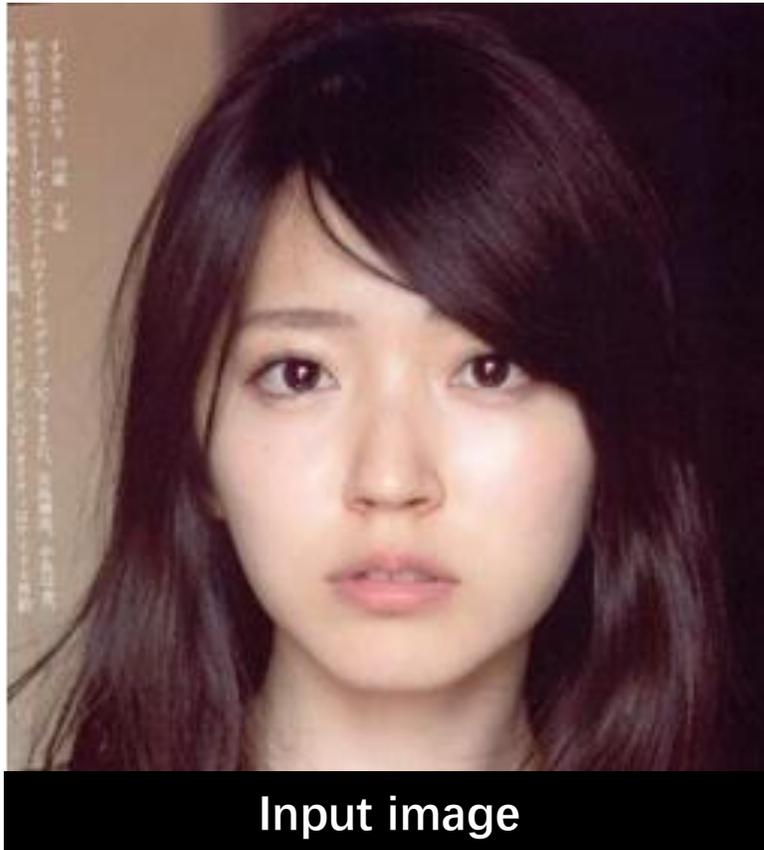
HUMAN PERCEPTION



HUMAN PERCEPTION

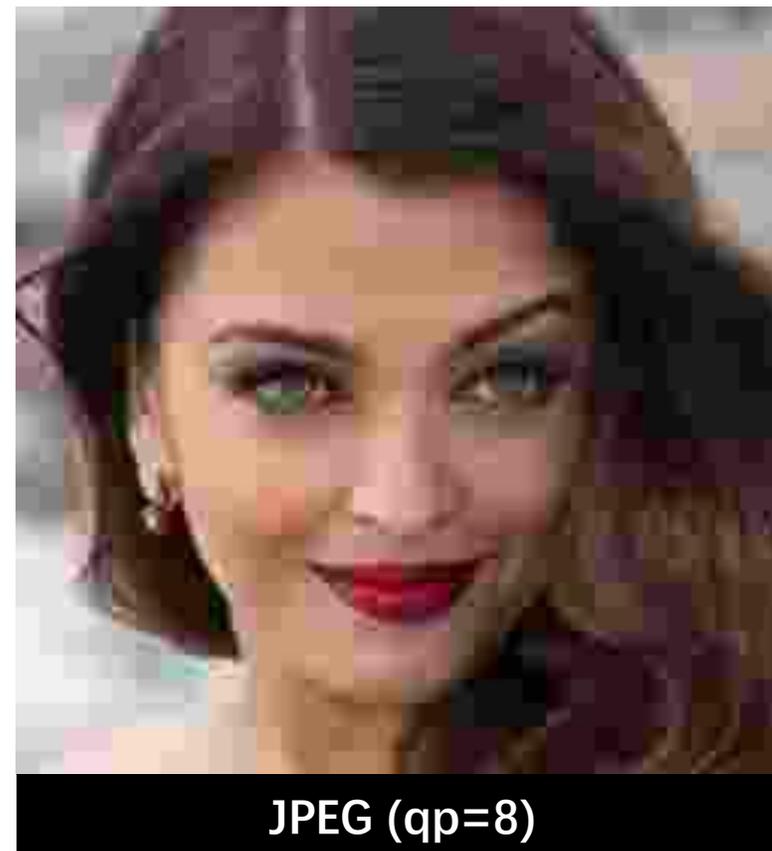
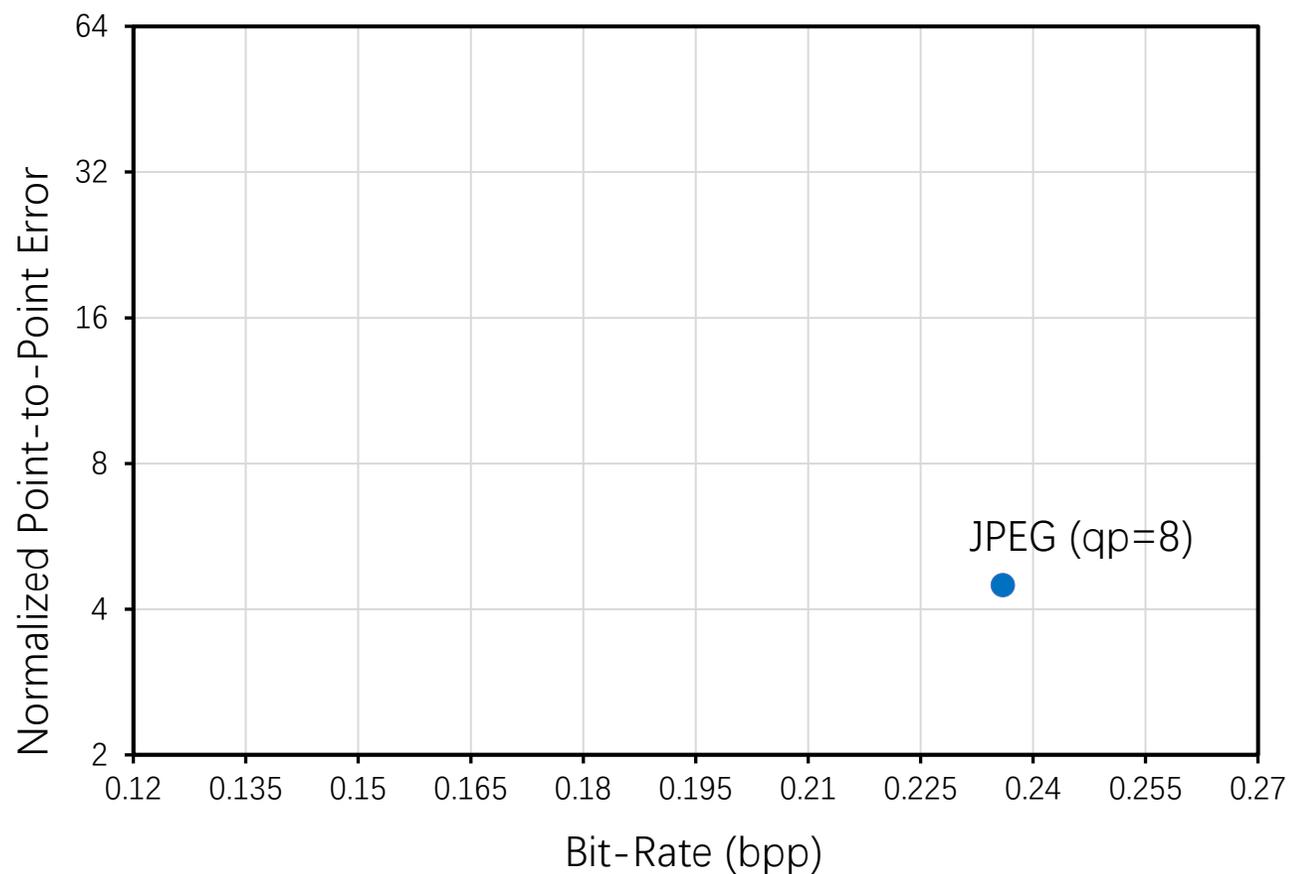


HUMAN PERCEPTION



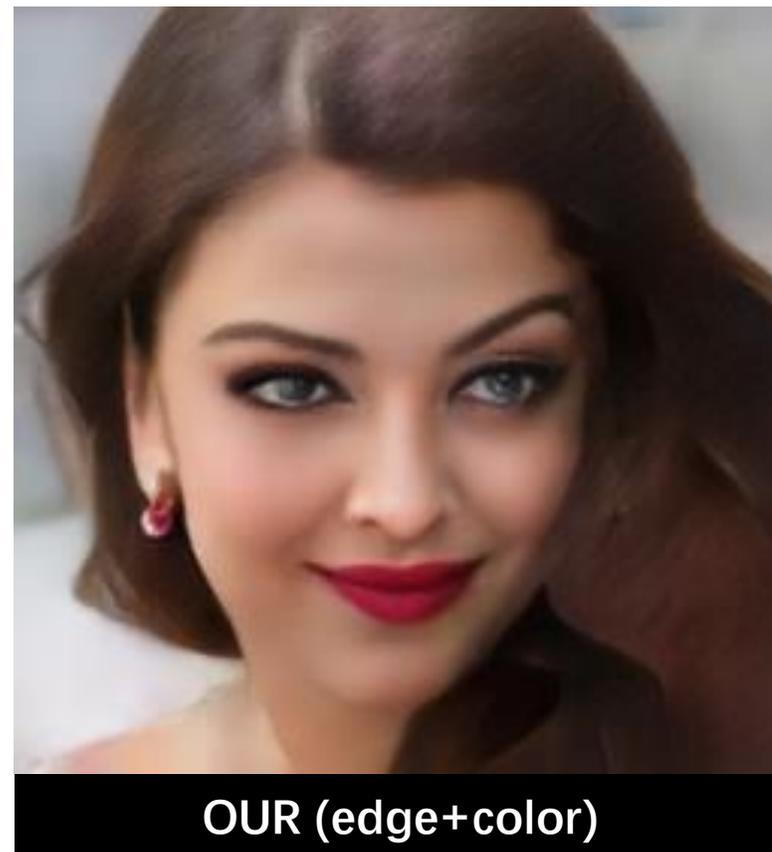
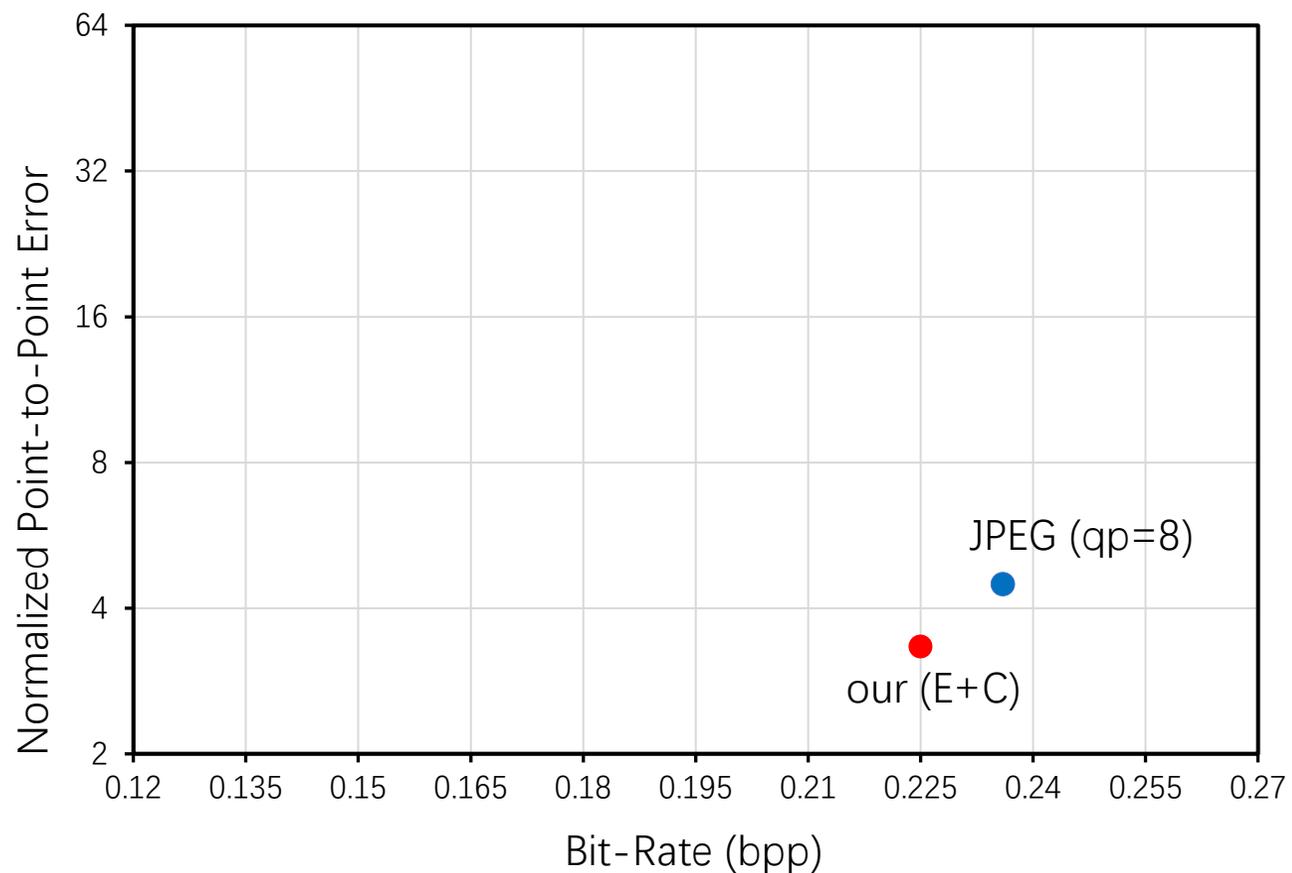
MACHINE VISION

LANDMARK DETECTION ACCURACY



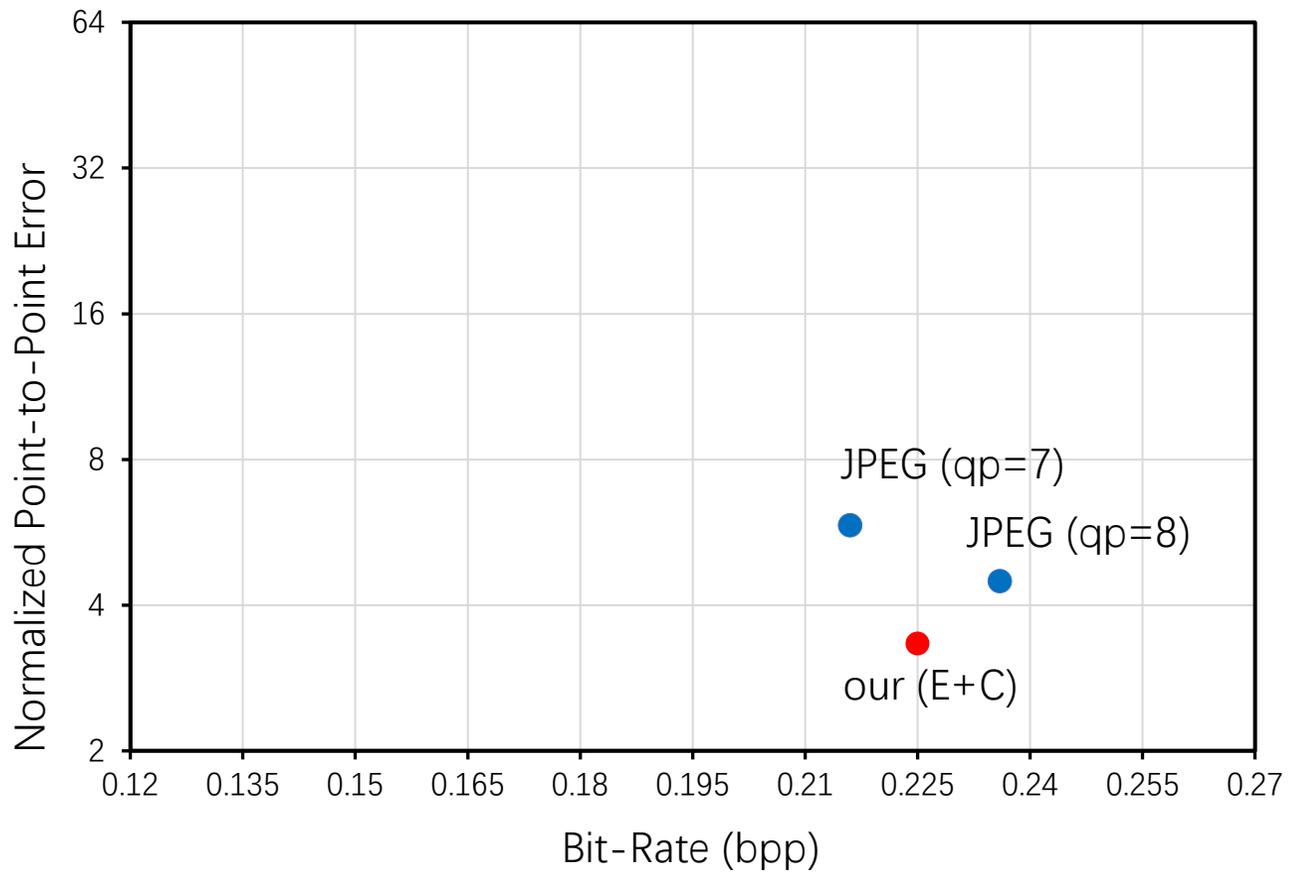
MACHINE VISION

LANDMARK DETECTION ACCURACY



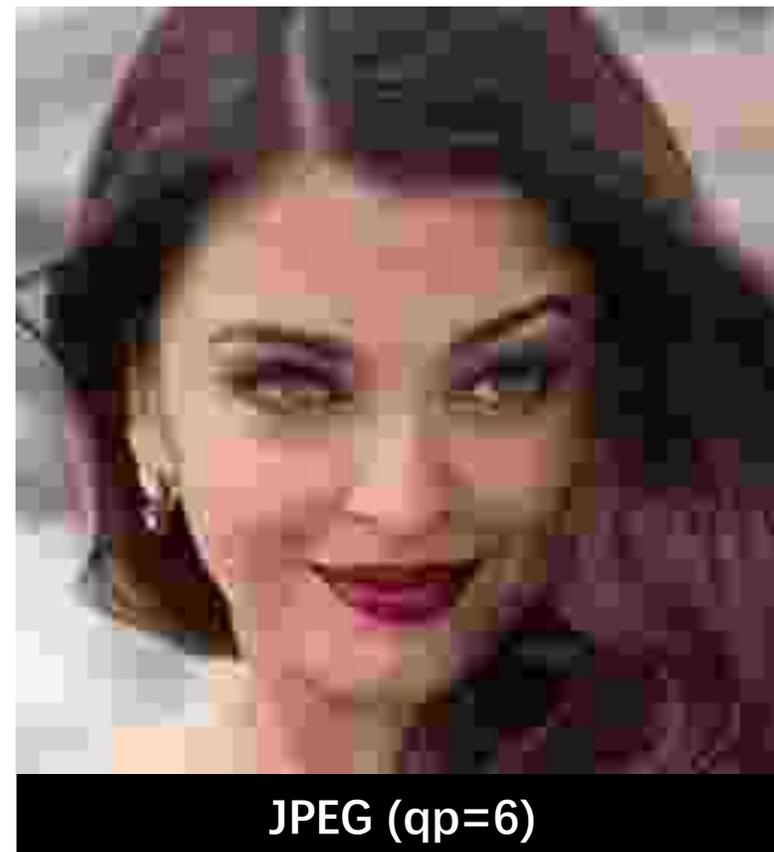
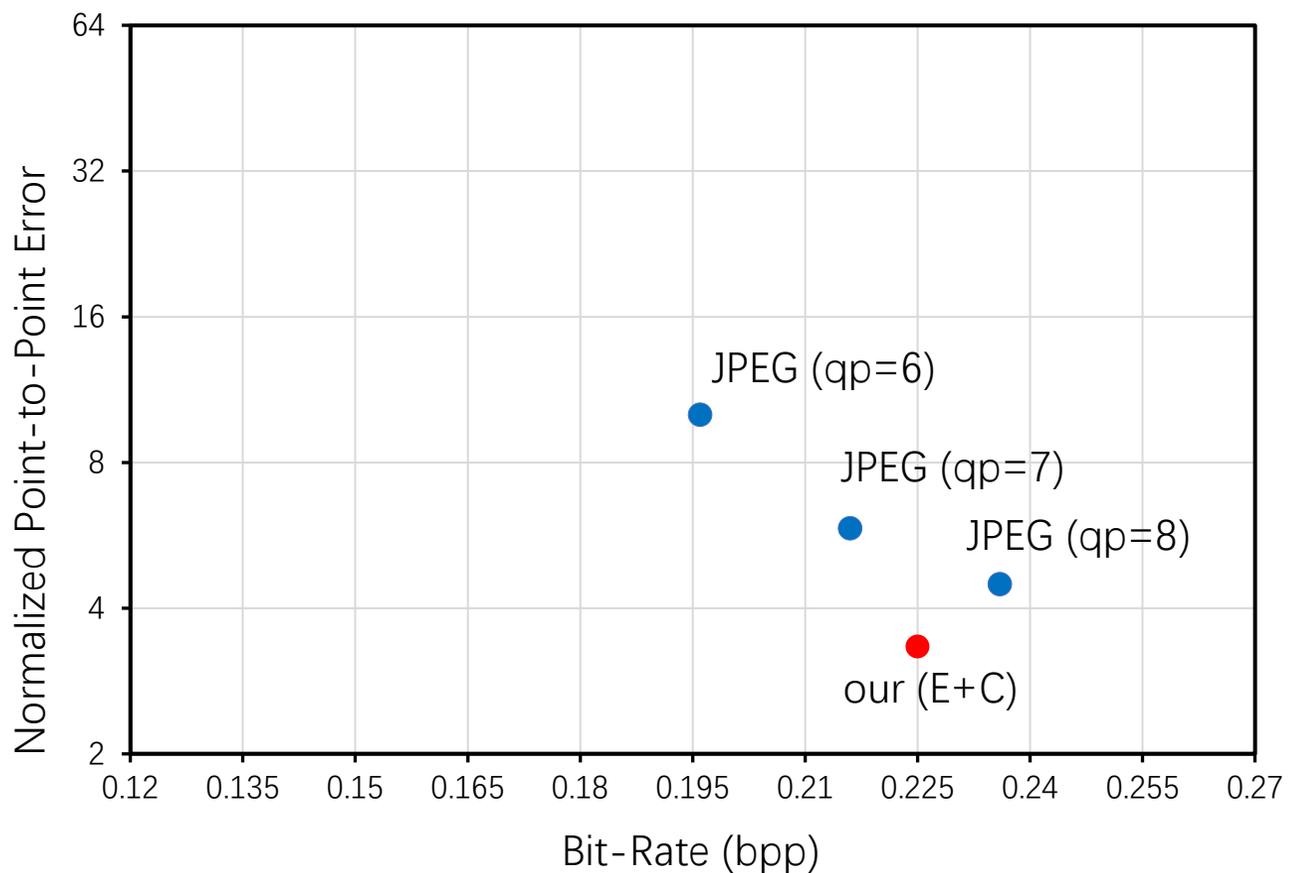
MACHINE VISION

LANDMARK DETECTION ACCURACY



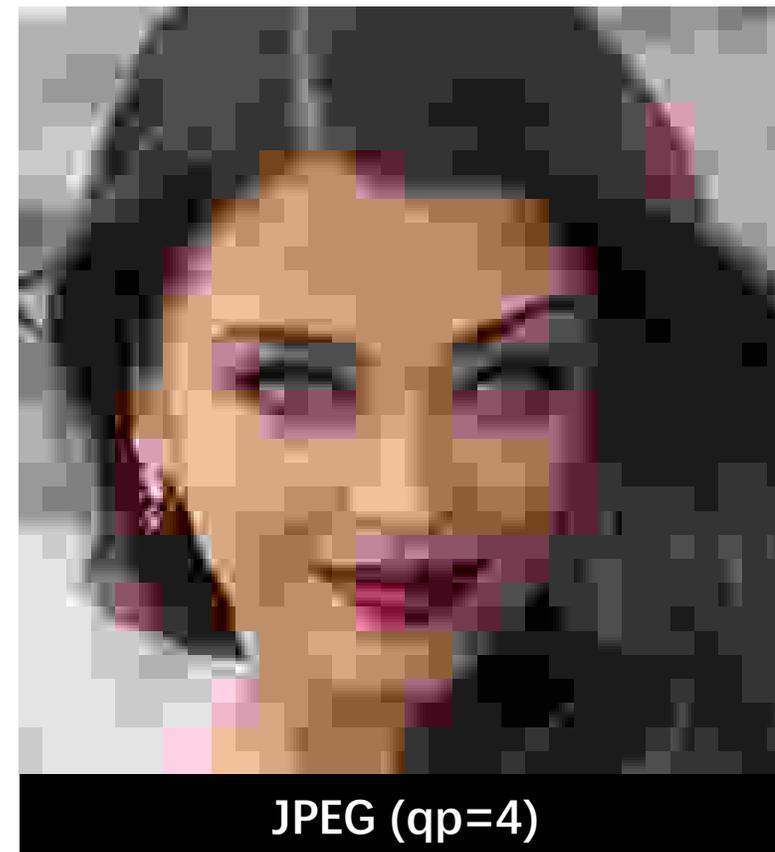
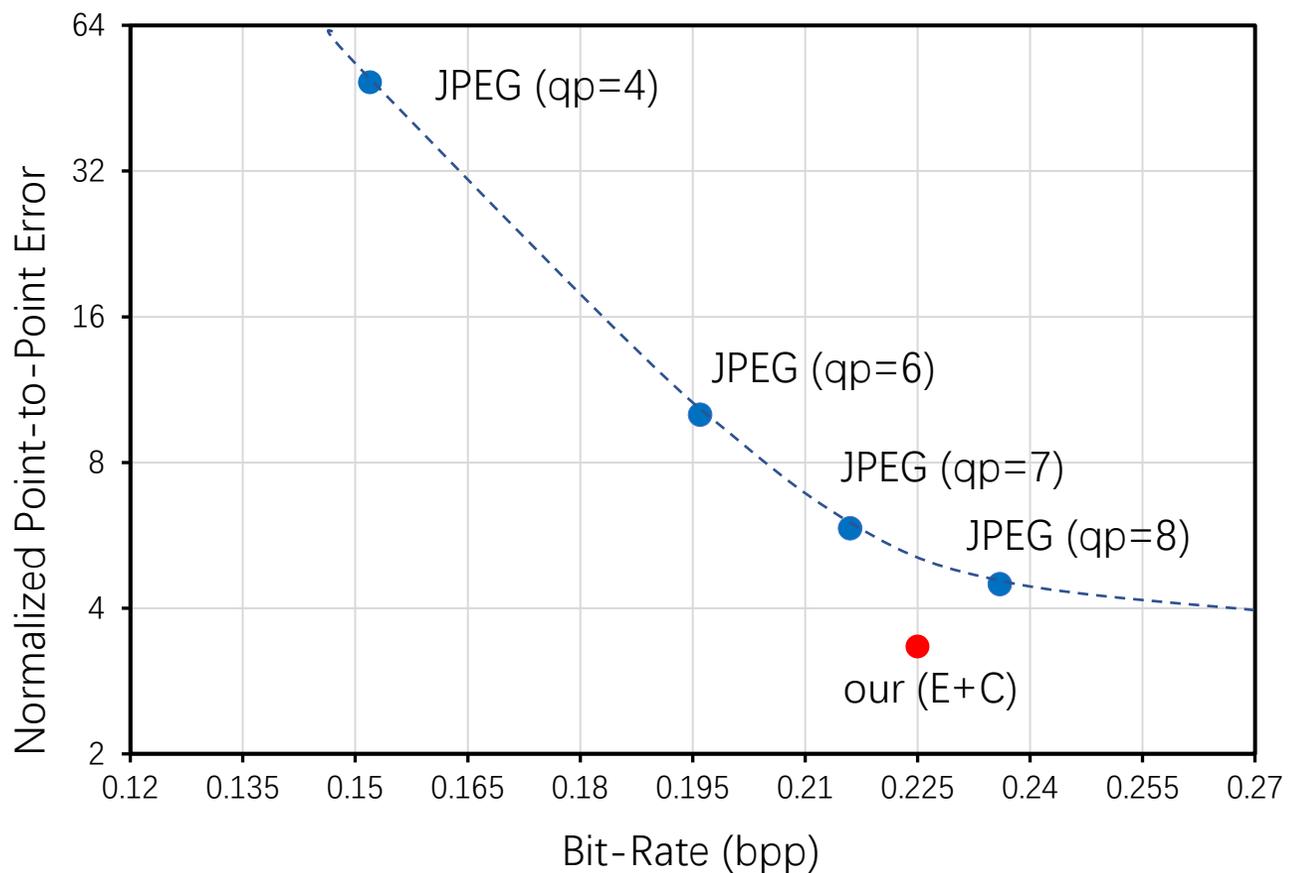
MACHINE VISION

LANDMARK DETECTION ACCURACY



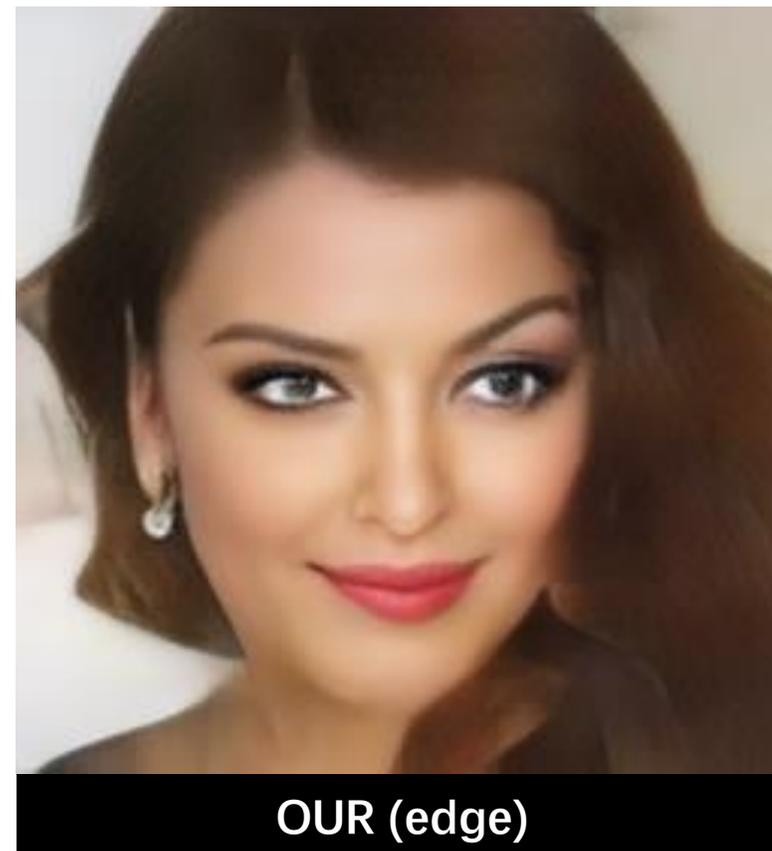
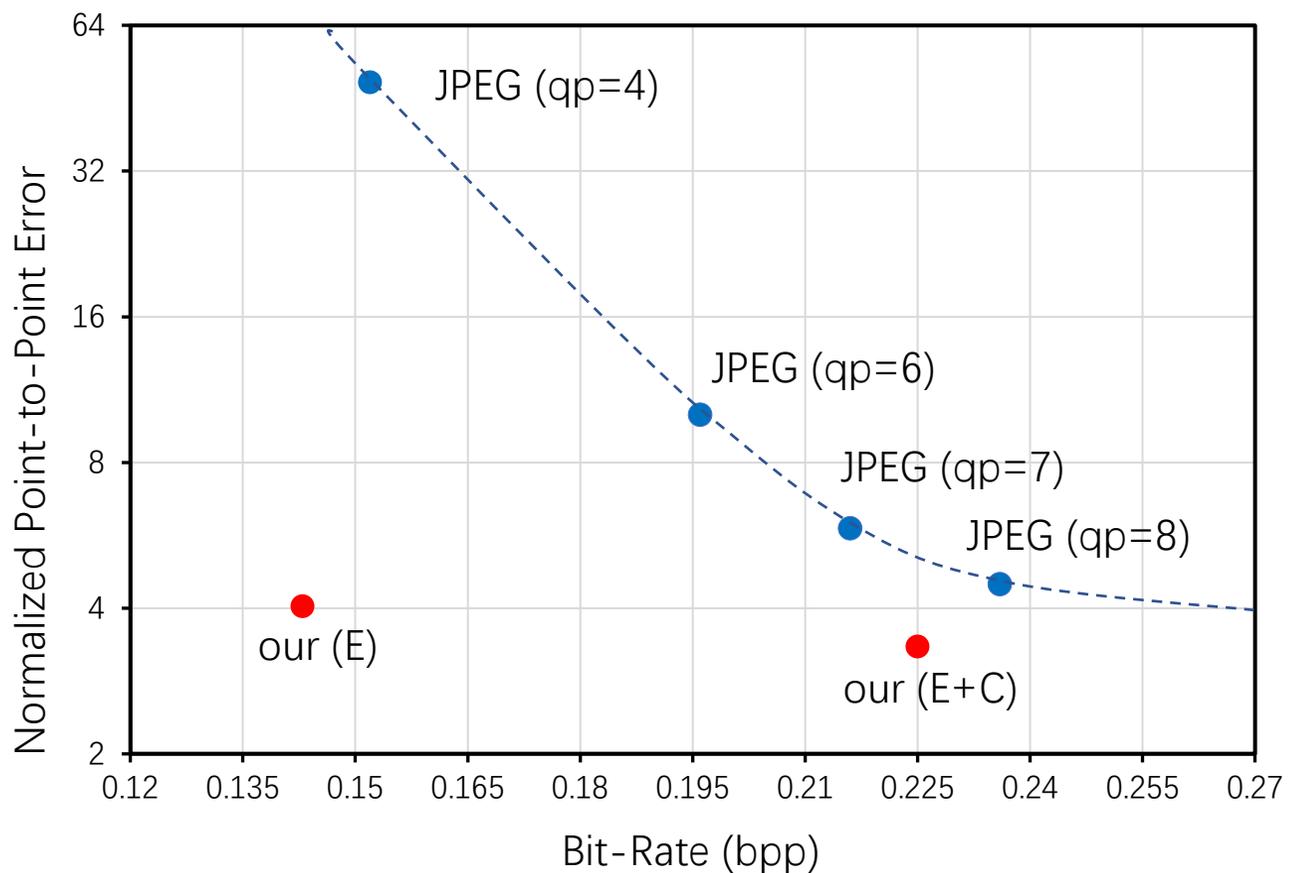
MACHINE VISION

LANDMARK DETECTION ACCURACY



MACHINE VISION

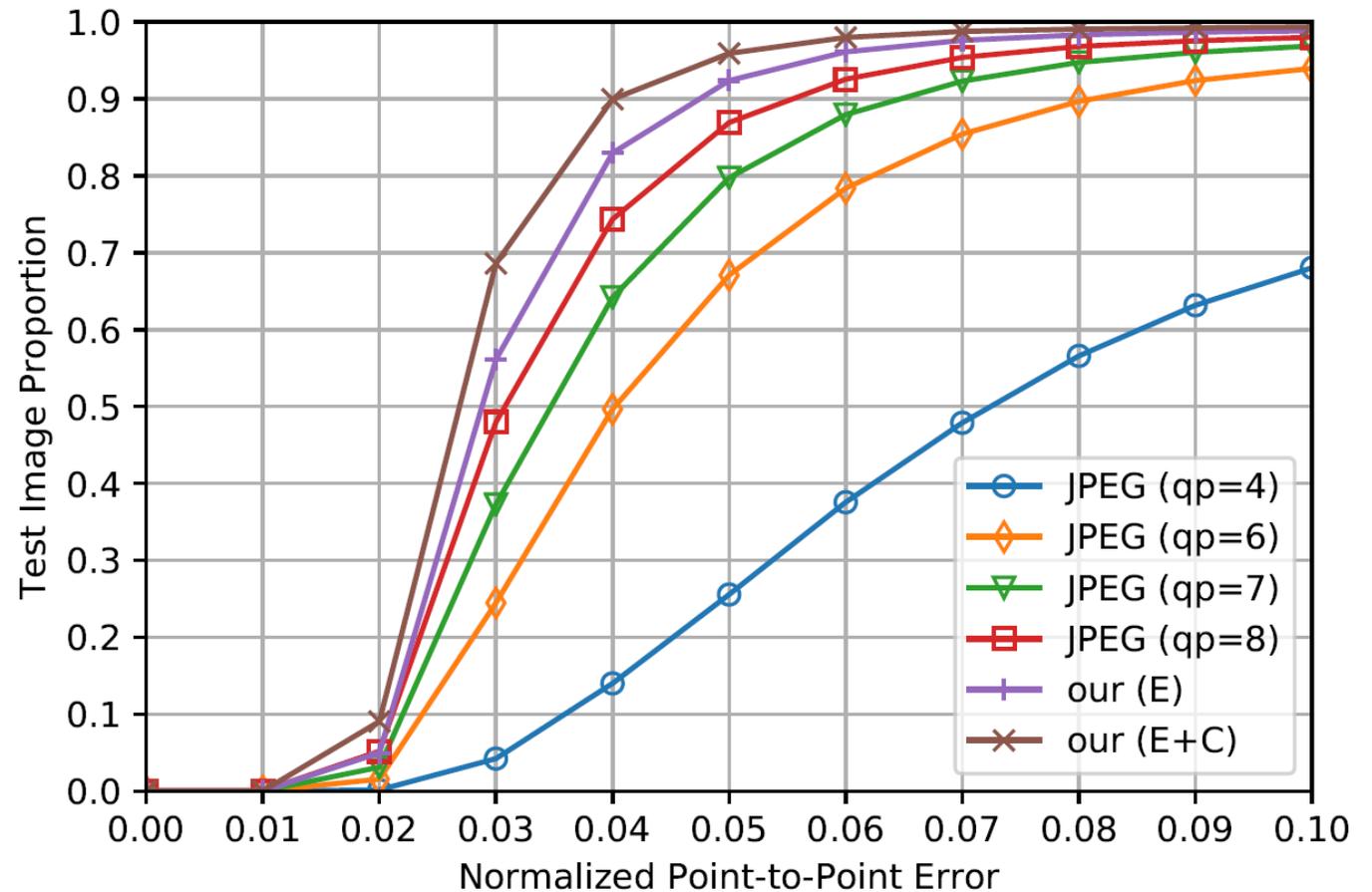
LANDMARK DETECTION ACCURACY



MACHINE VISION

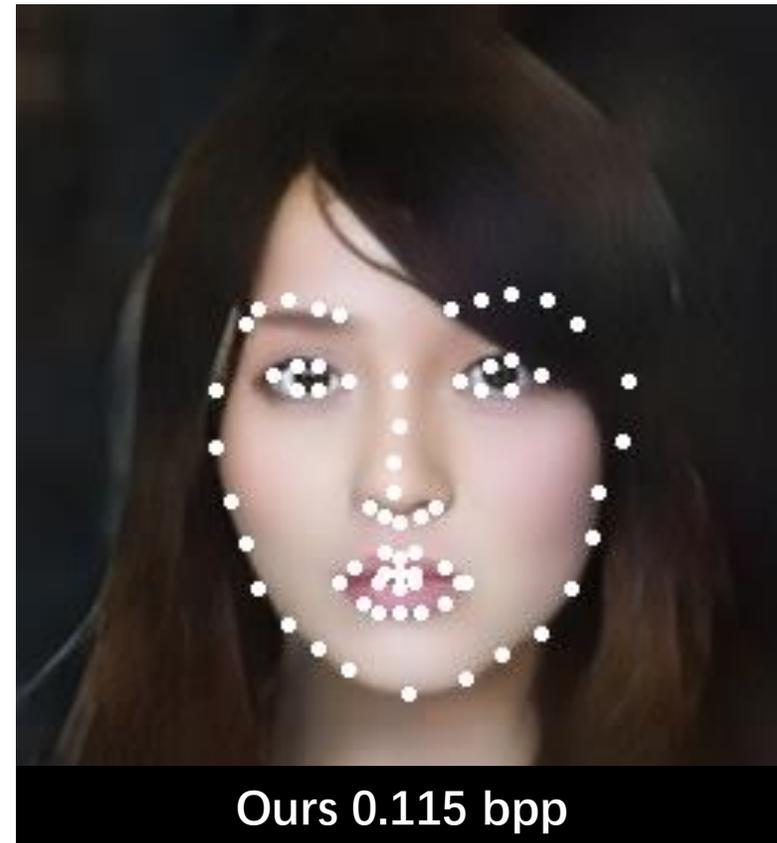
LANDMARK DETECTION ACCURACY

- Quantitatively evaluate the accuracy of facial landmark detection on the reconstructed images.
- Results show statistically improved accuracy at a lower bit-rate.
- While the basic layer maintain a high accuracy, the enhancing layer provide more fidelity.



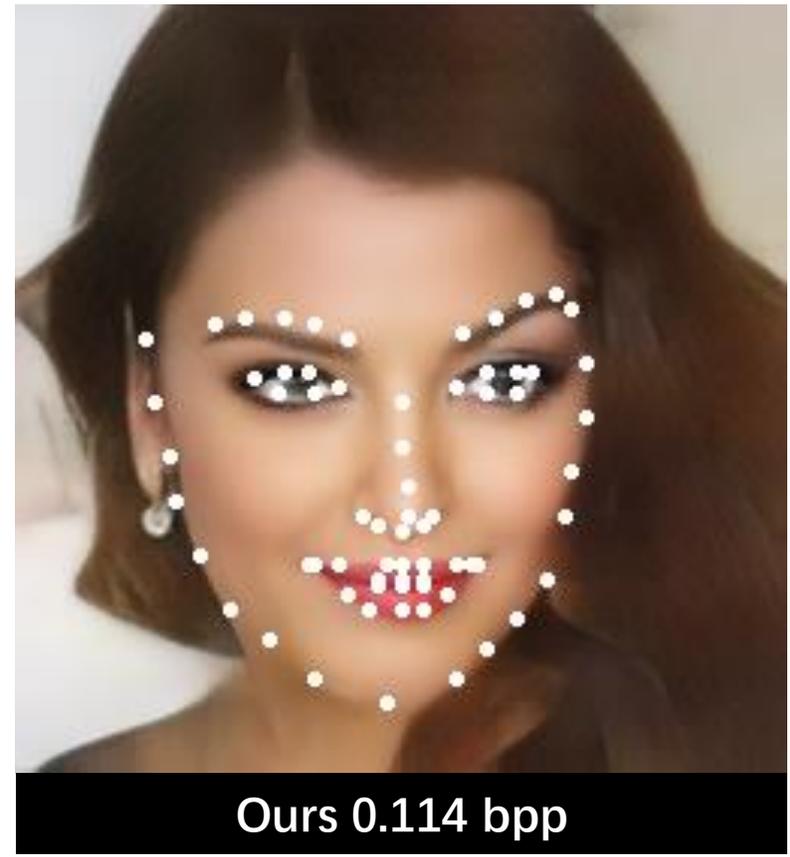
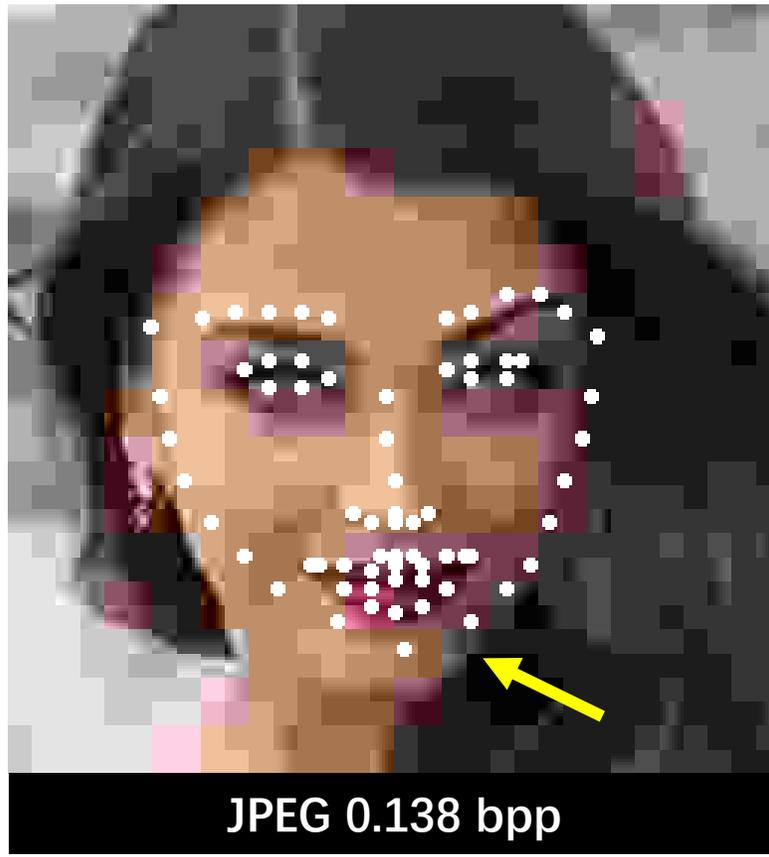
MACHINE VISION

LANDMARK DETECTION RESULTS



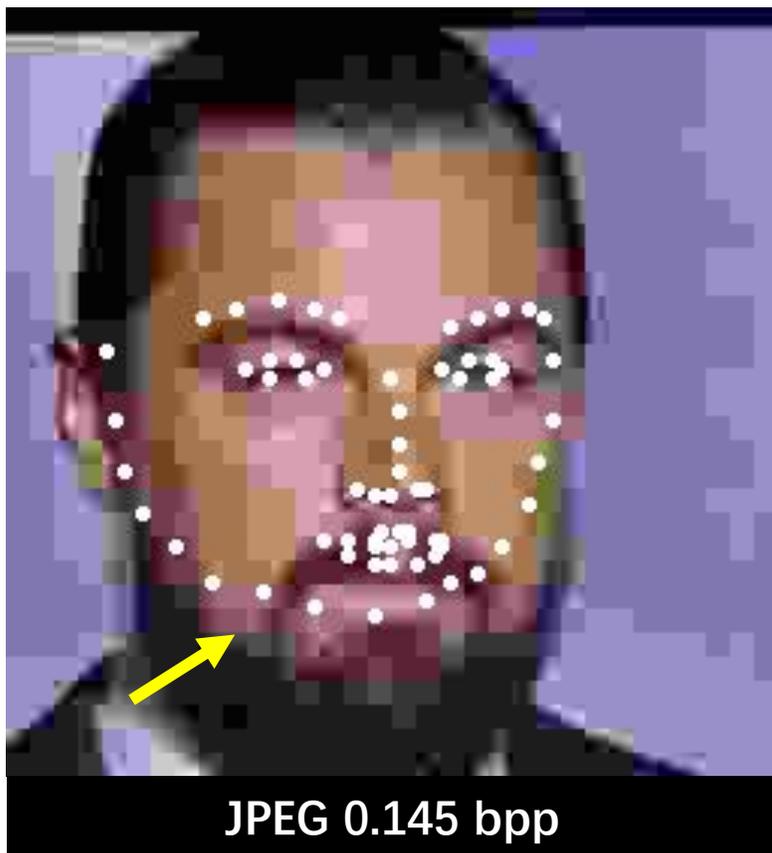
MACHINE VISION

LANDMARK DETECTION RESULTS



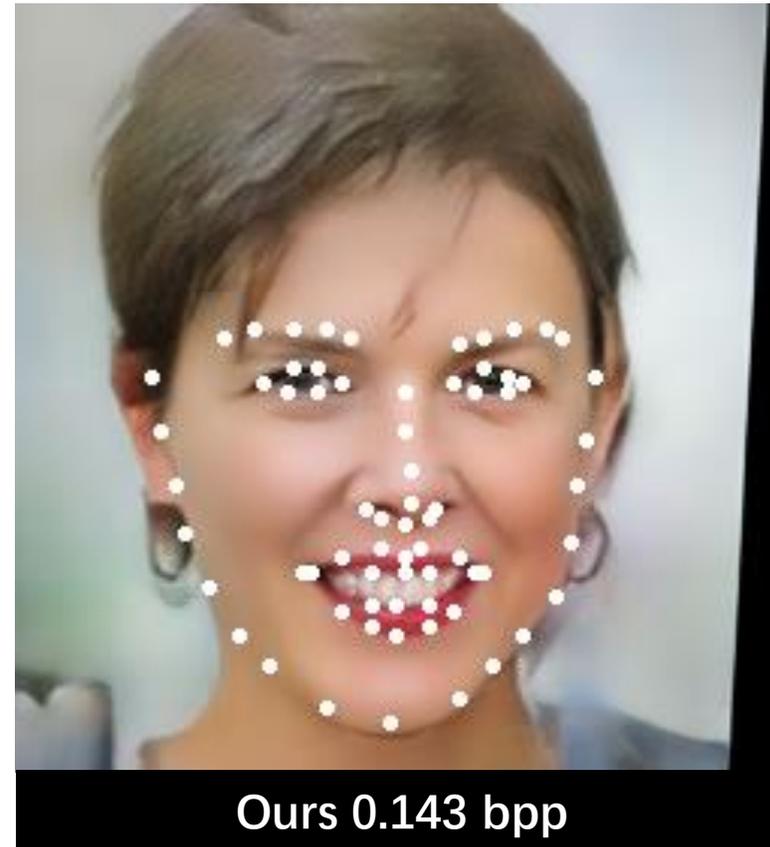
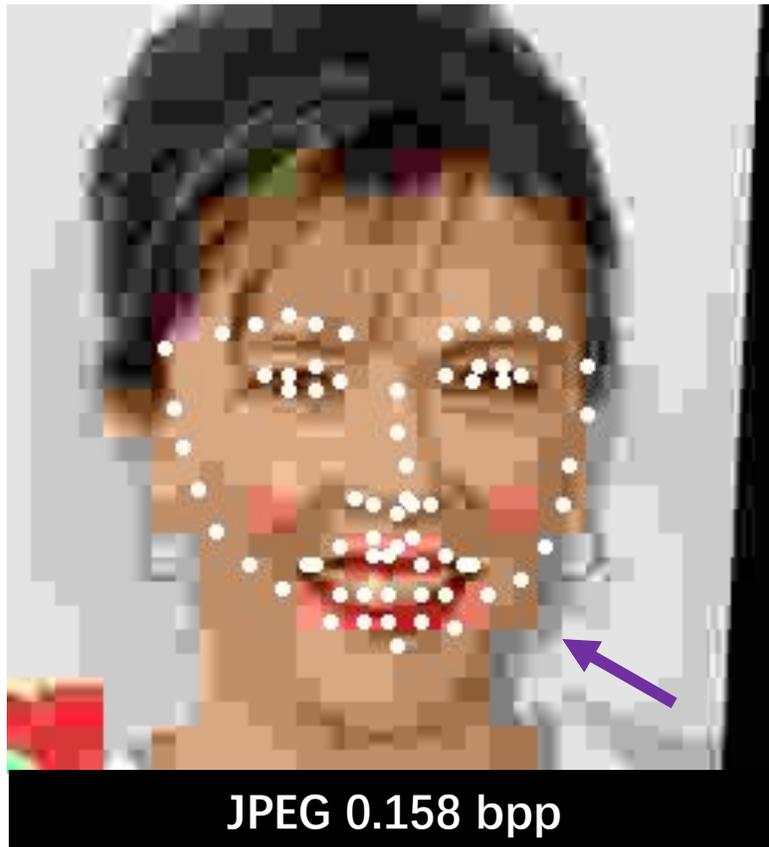
MACHINE VISION

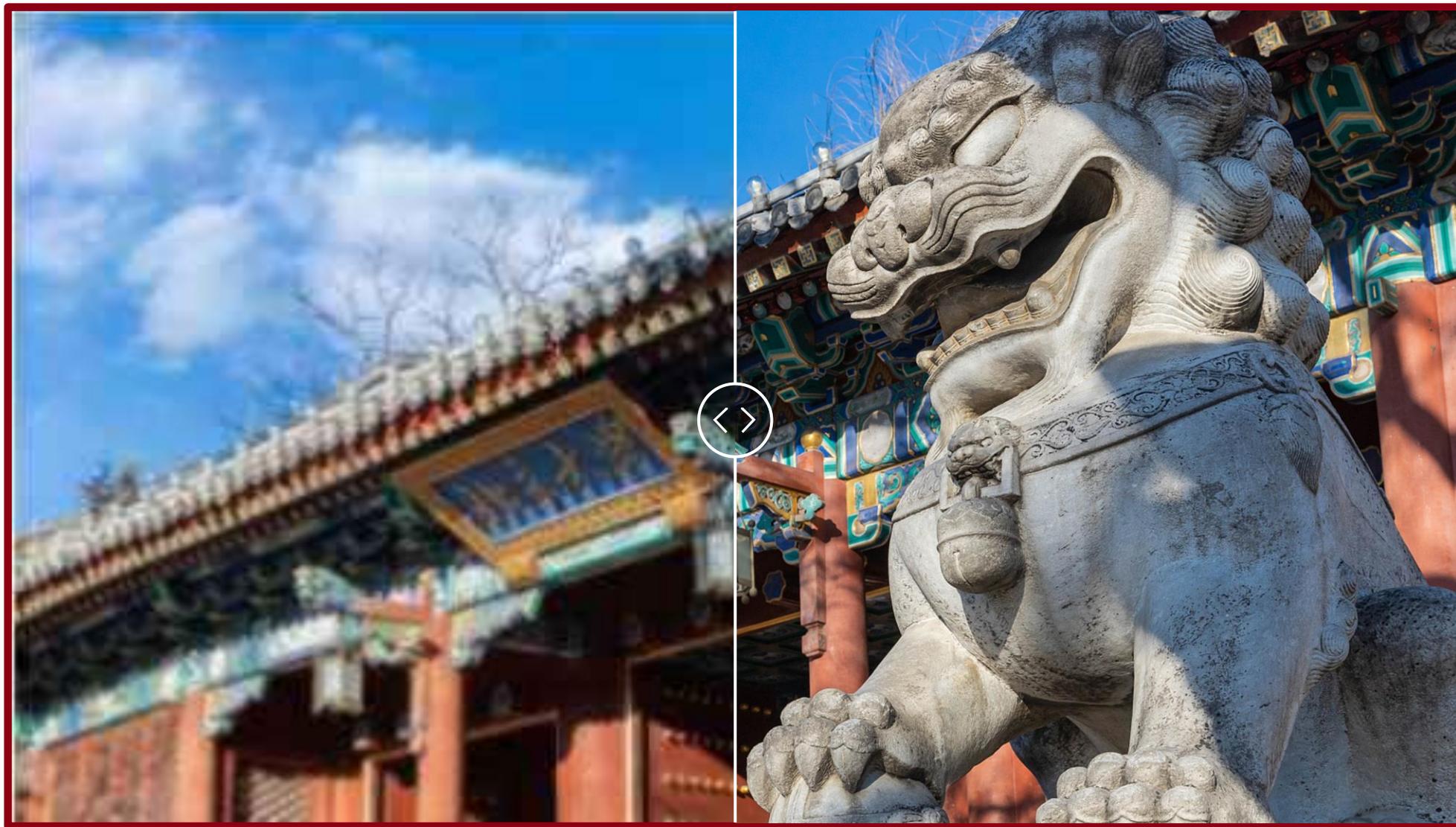
LANDMARK DETECTION RESULTS



MACHINE VISION

LANDMARK DETECTION RESULTS







真实场景下的自学习超分辨率

Self-Learning Super-Resolution in Real-World Scenes

樊泽嘉 北京大学



超高清视界



超分辨率技术



超分辨率技术

- 老照片的修复



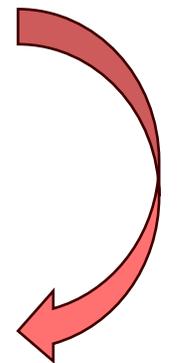
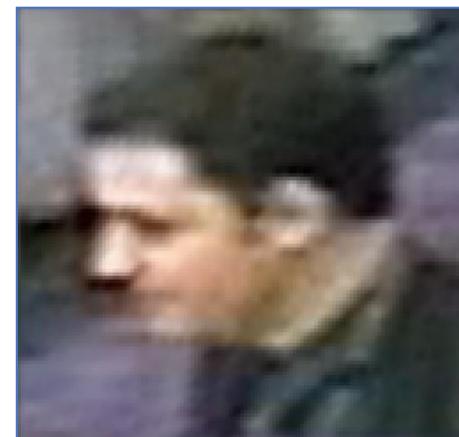
超分辨率技术

- 老照片的修复



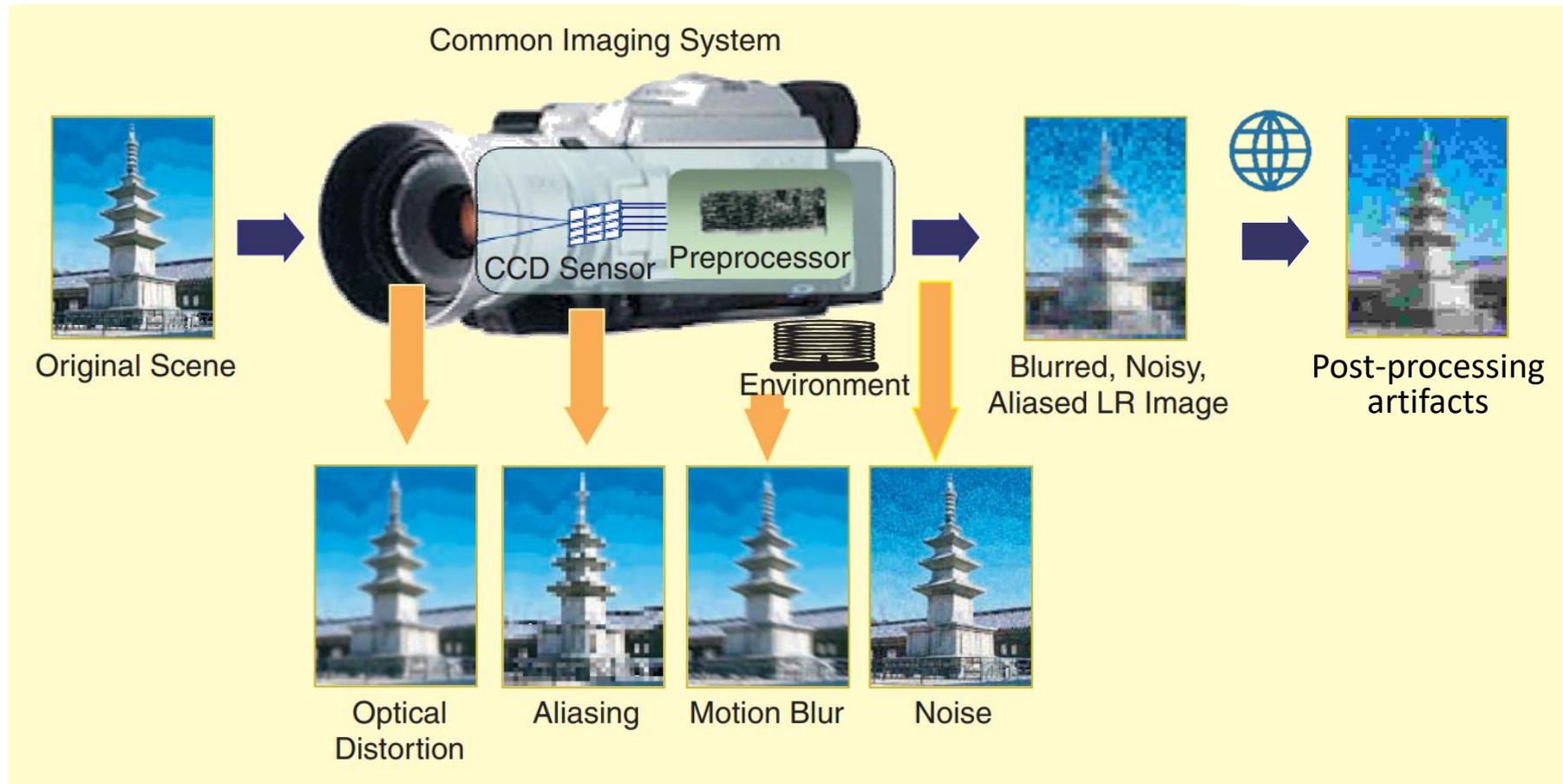
超分辨率技术

- 监控视频的识别分析



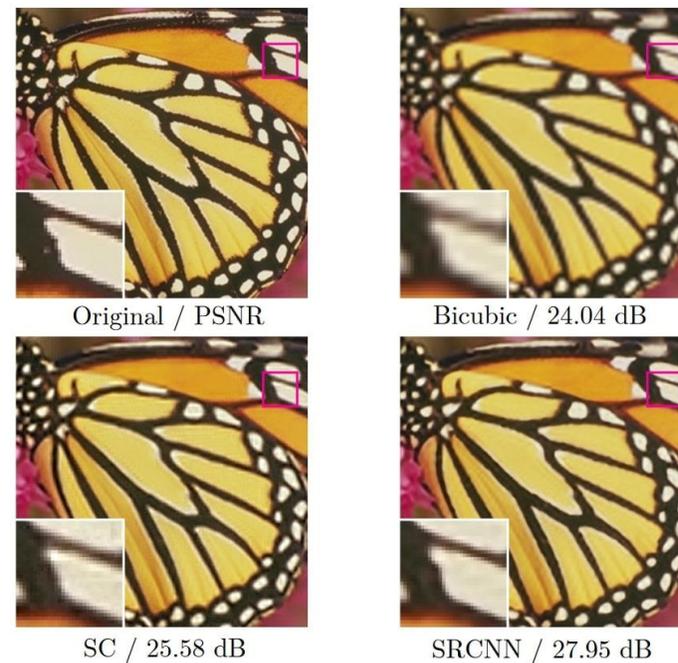
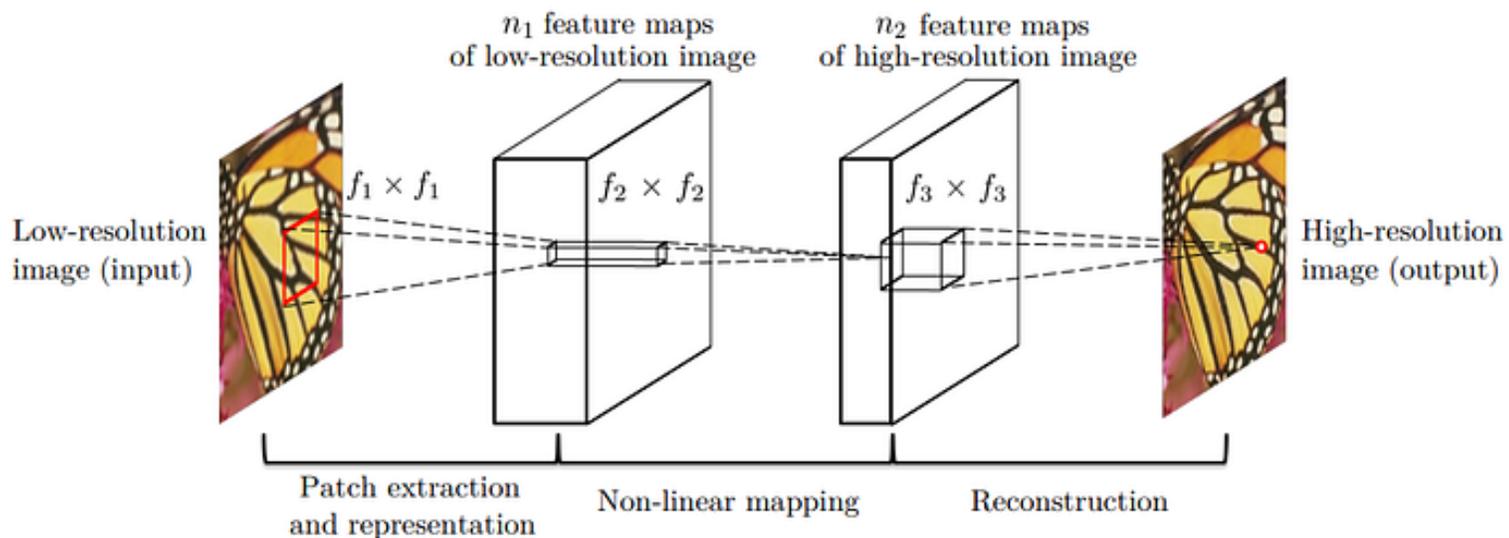
真实世界超分辨率

- 复杂多变的降质情况



相关工作

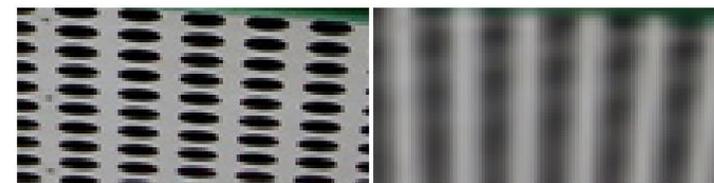
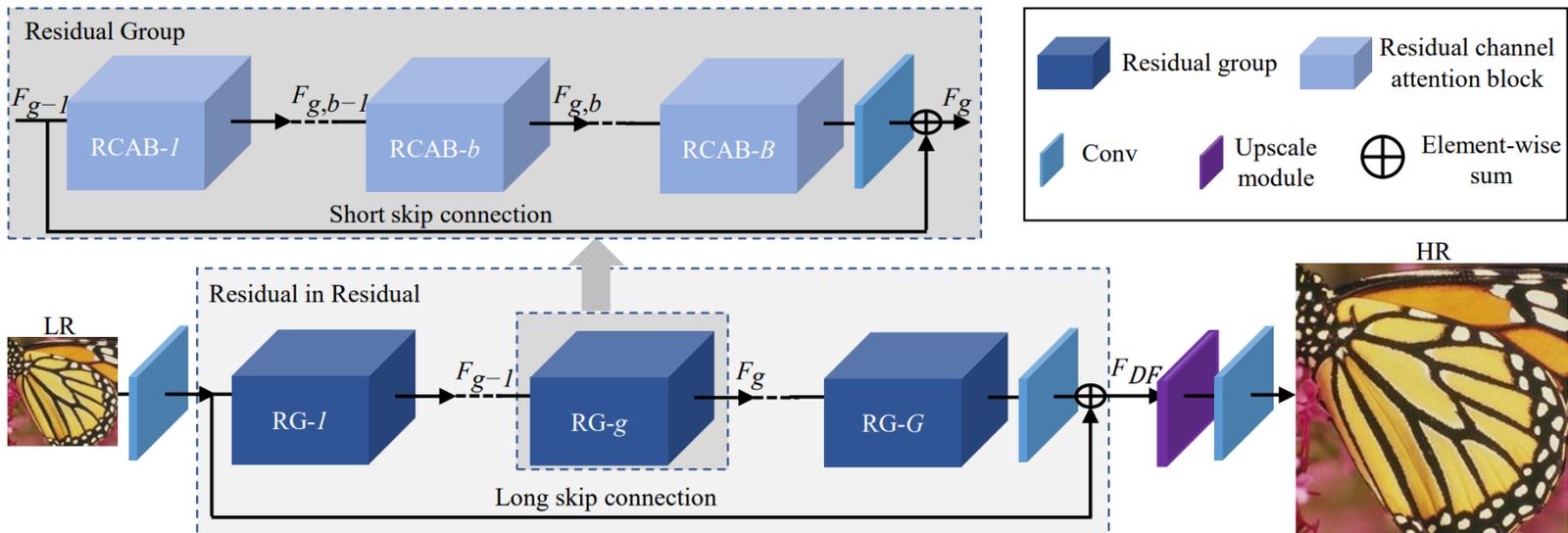
2014
SRCNN



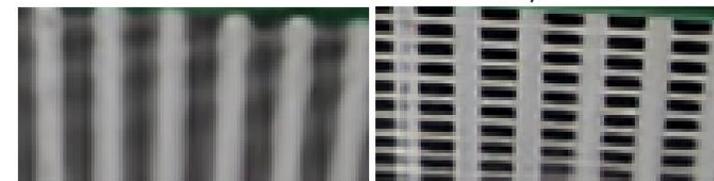
相关工作

2014
SRCNN

2018
RCAN



HR
PSNR/SSIM 21.08/0.6788



SRCNN [1]
22.13/0.7635

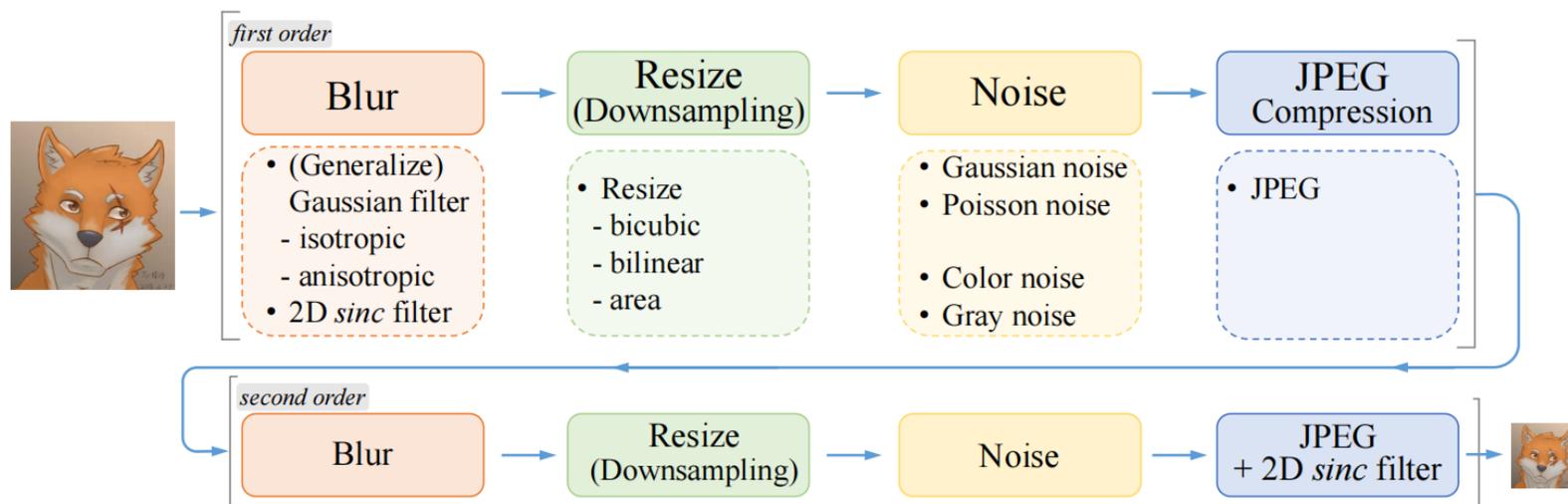
RCAN
25.64/0.8830

相关工作

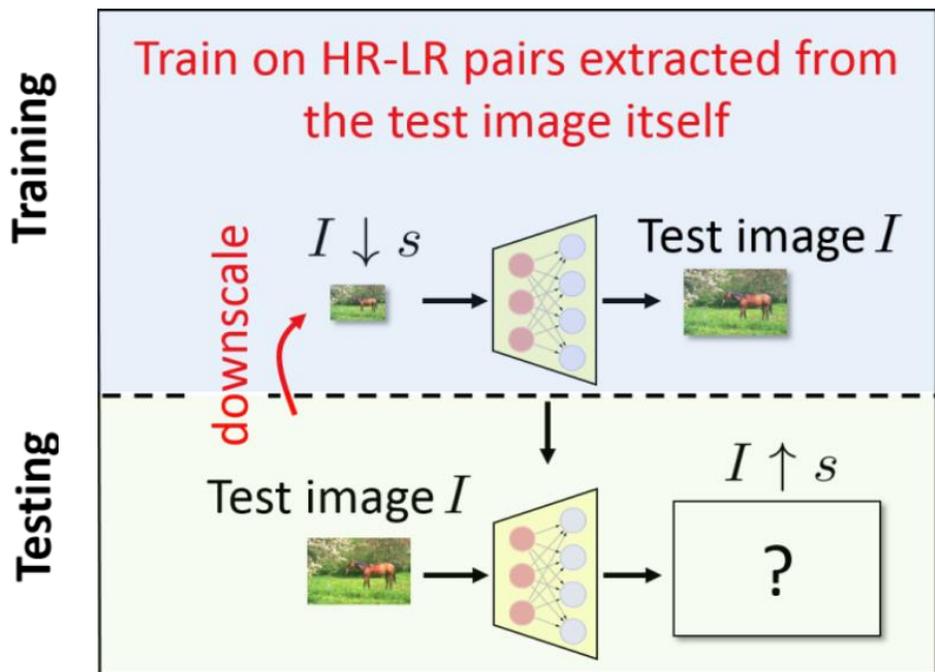
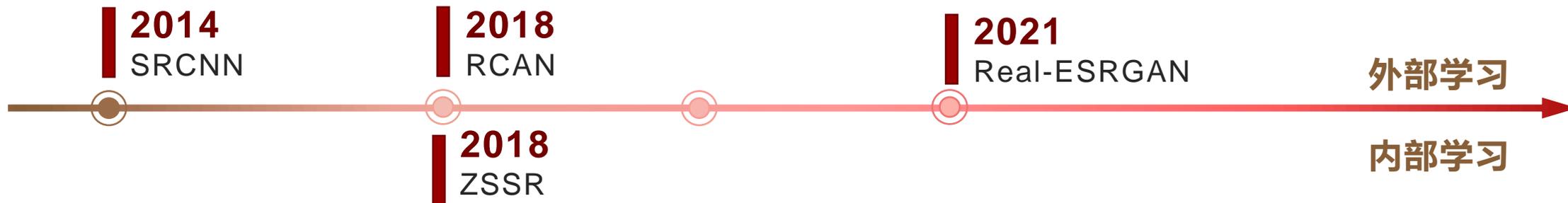
2014
SRCNN

2018
RCAN

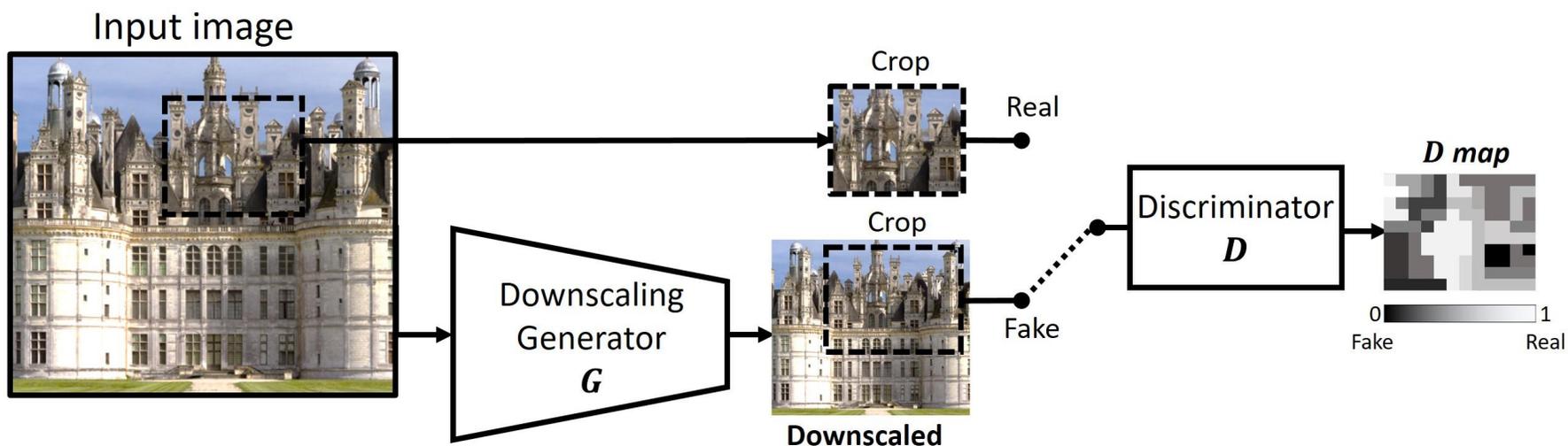
2021
Real-ESRGAN



相关工作

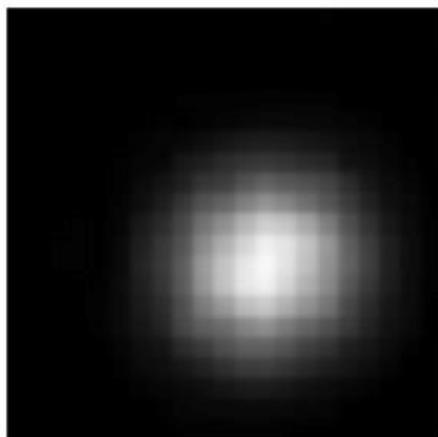


相关工作



研究问题

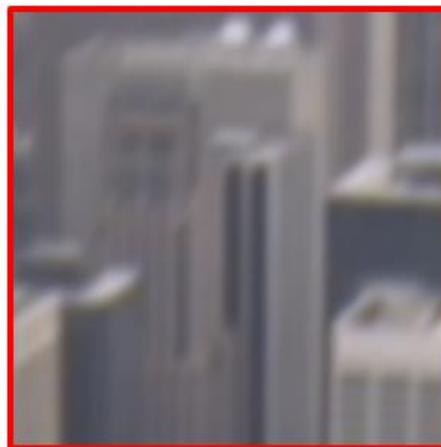
- 视频超分辨率方法通常依赖于成对的合成数据
- 训练数据与测试场景不匹配



Unknown
Blind Kernel



Input
(PSNR/SSIM)



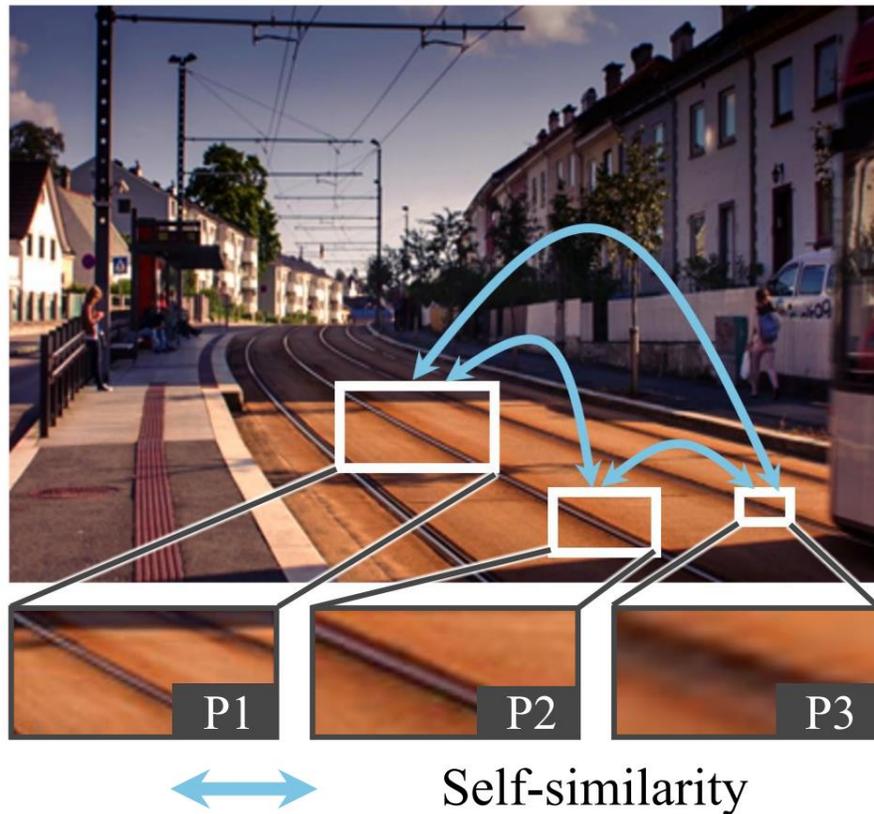
Supervised VSR
23.97/0.5910



SLVSR
26.41/0.7147

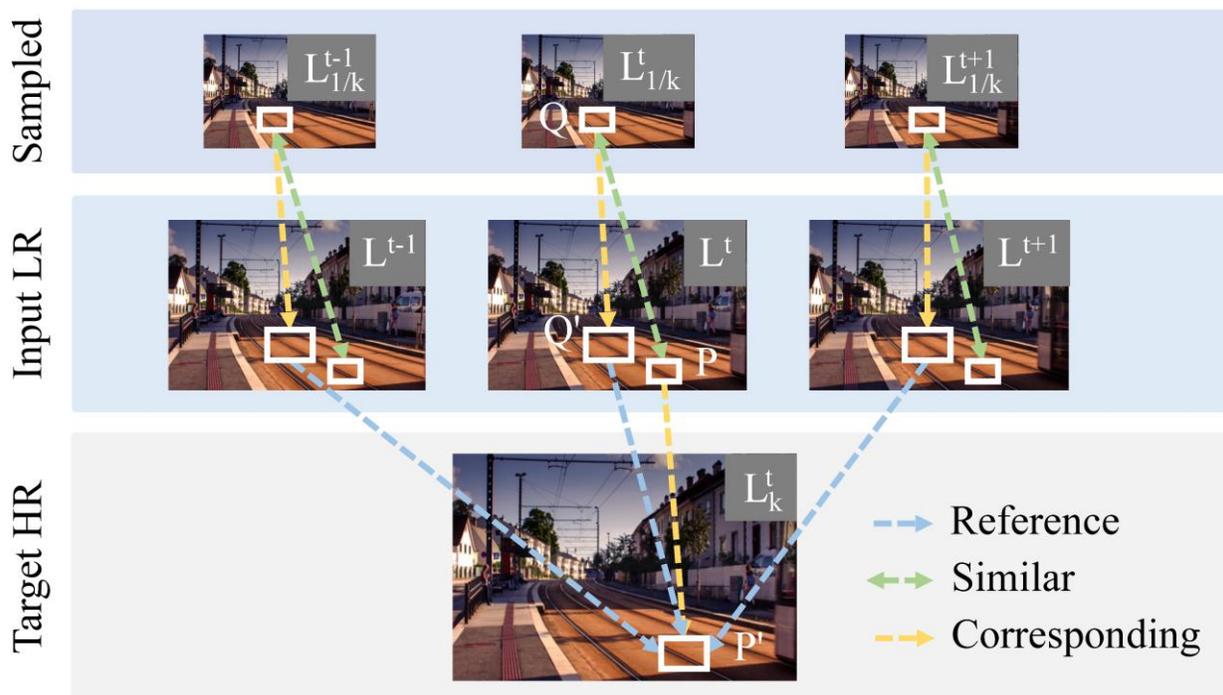
研究目标

- 旨在实现自学习的零样本视频超分辨率框架
- 不依赖外部数据

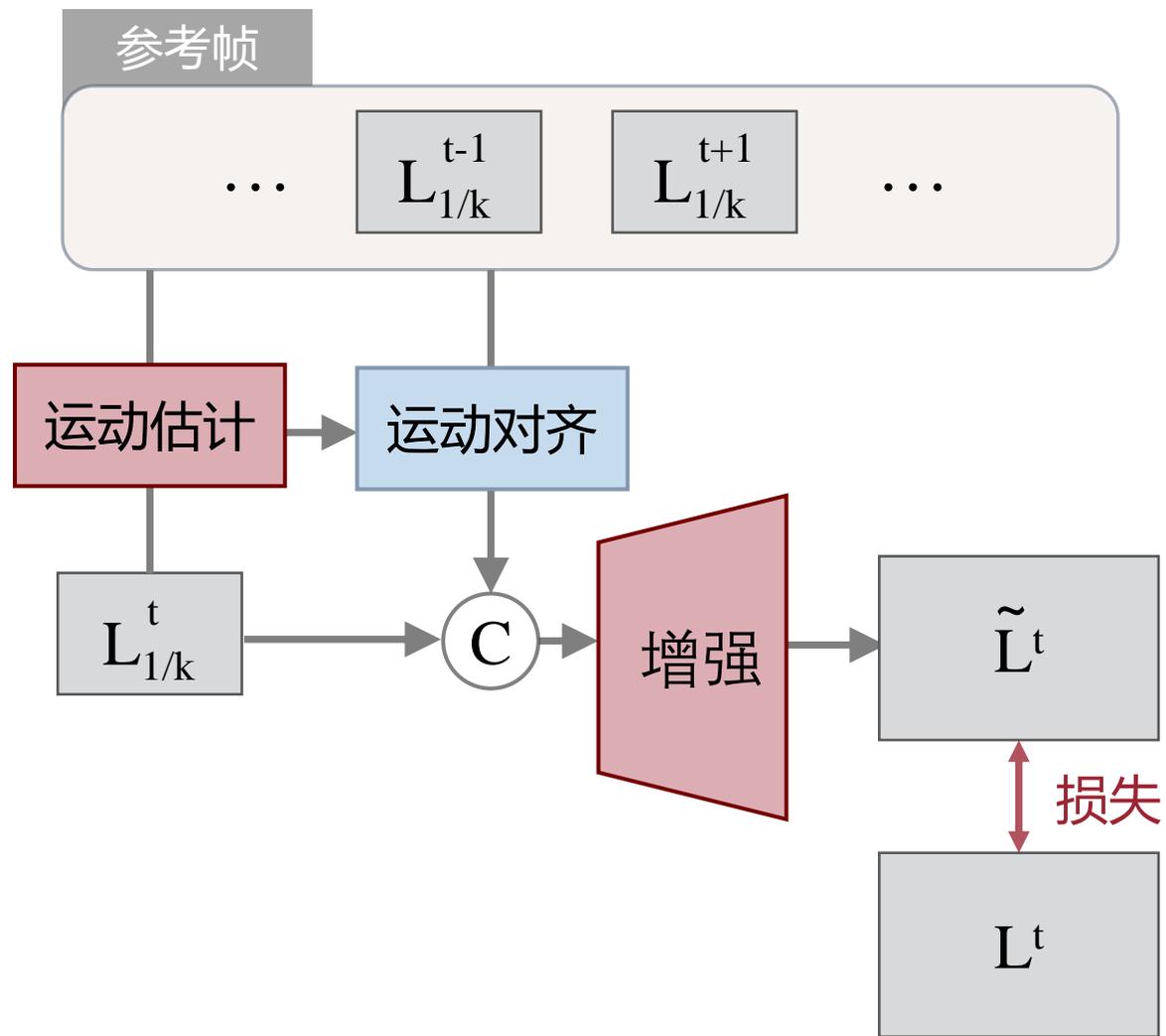
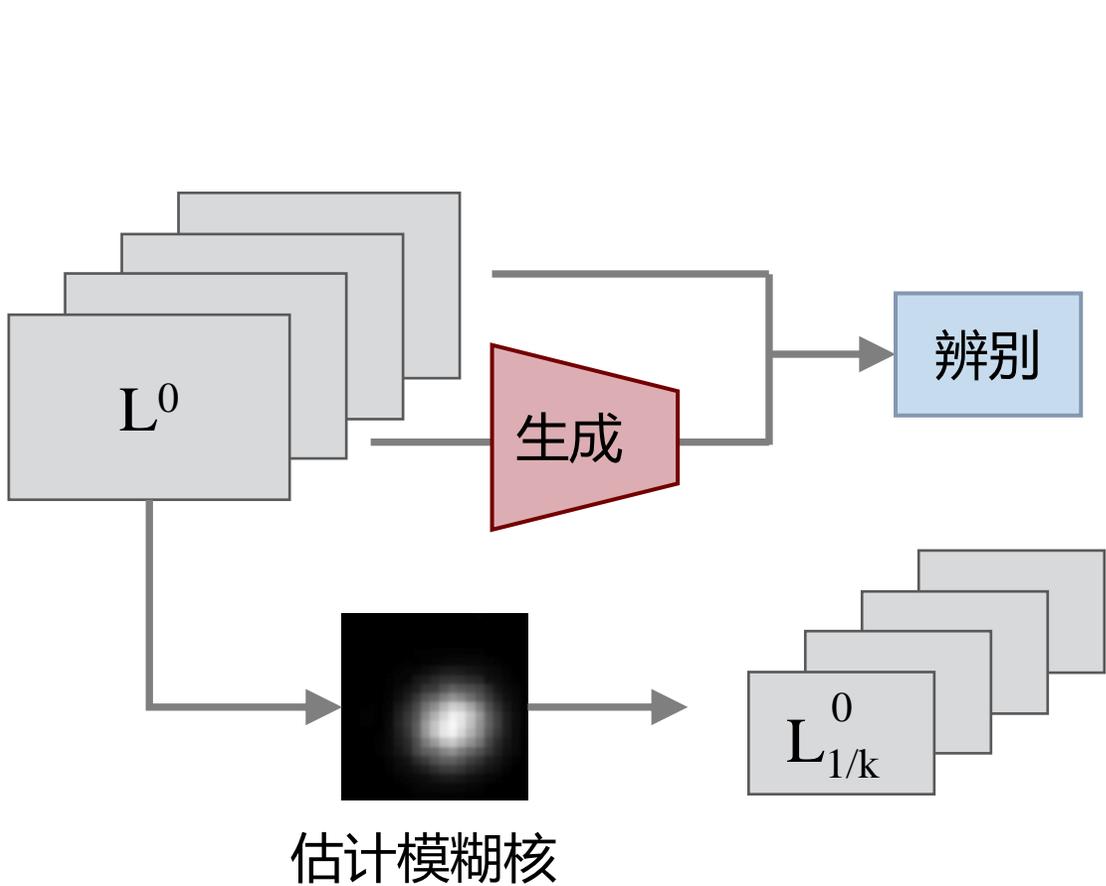


基于自学习的视频超分辨率

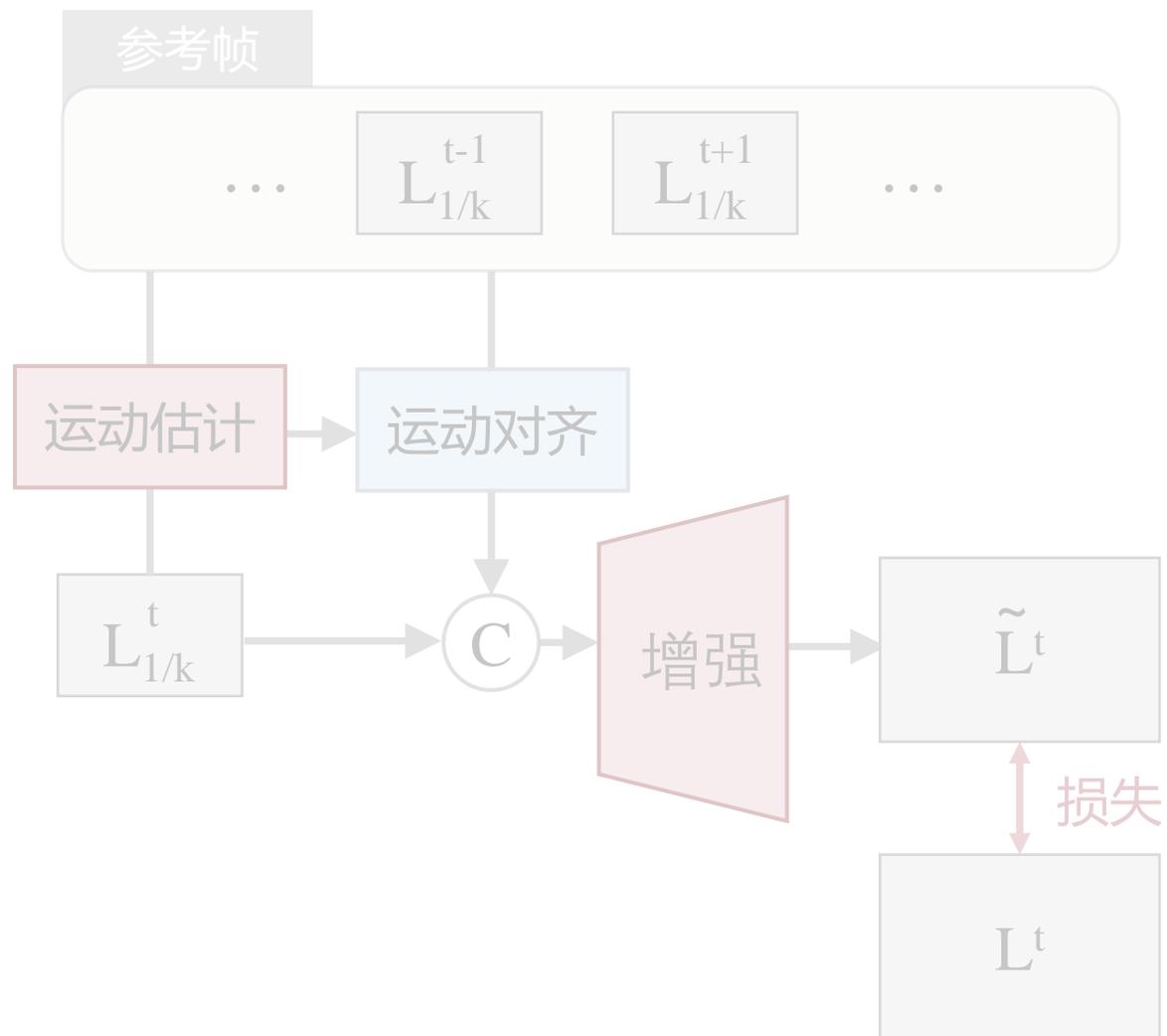
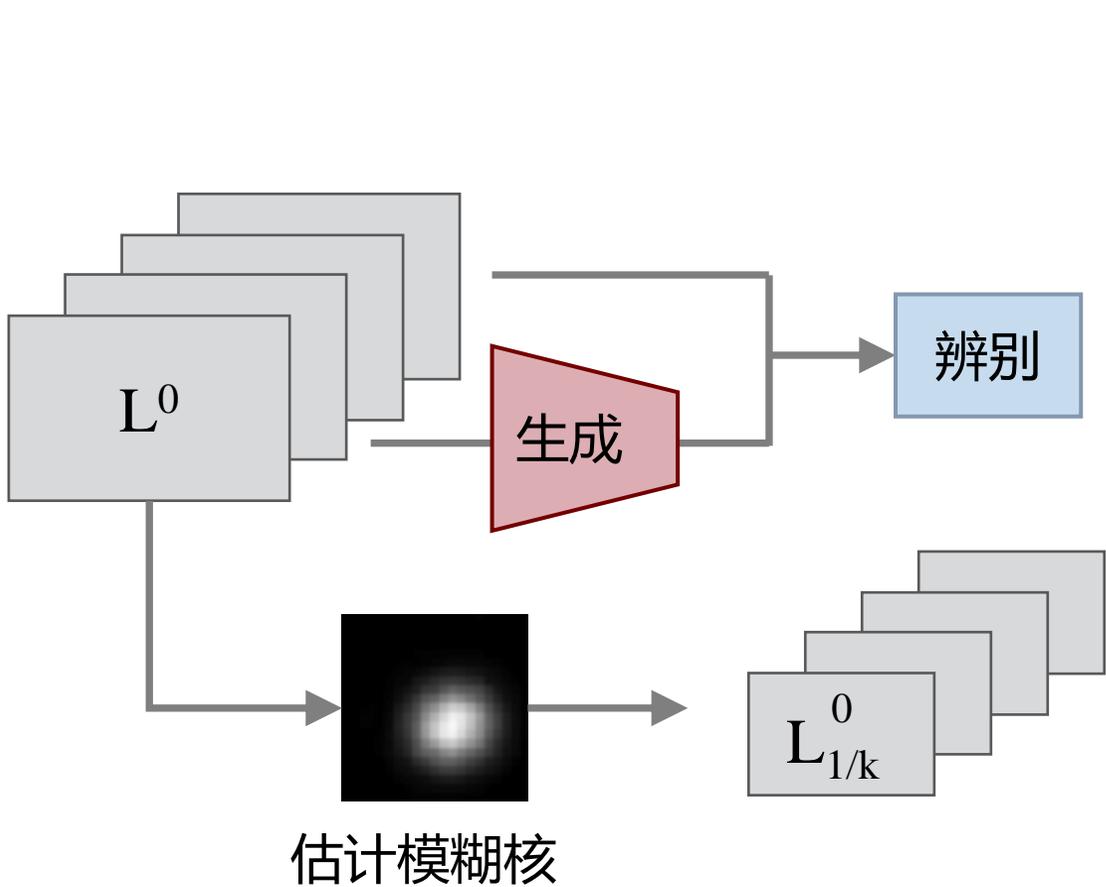
- 跨尺度空间自相似性
- 时域自相似性



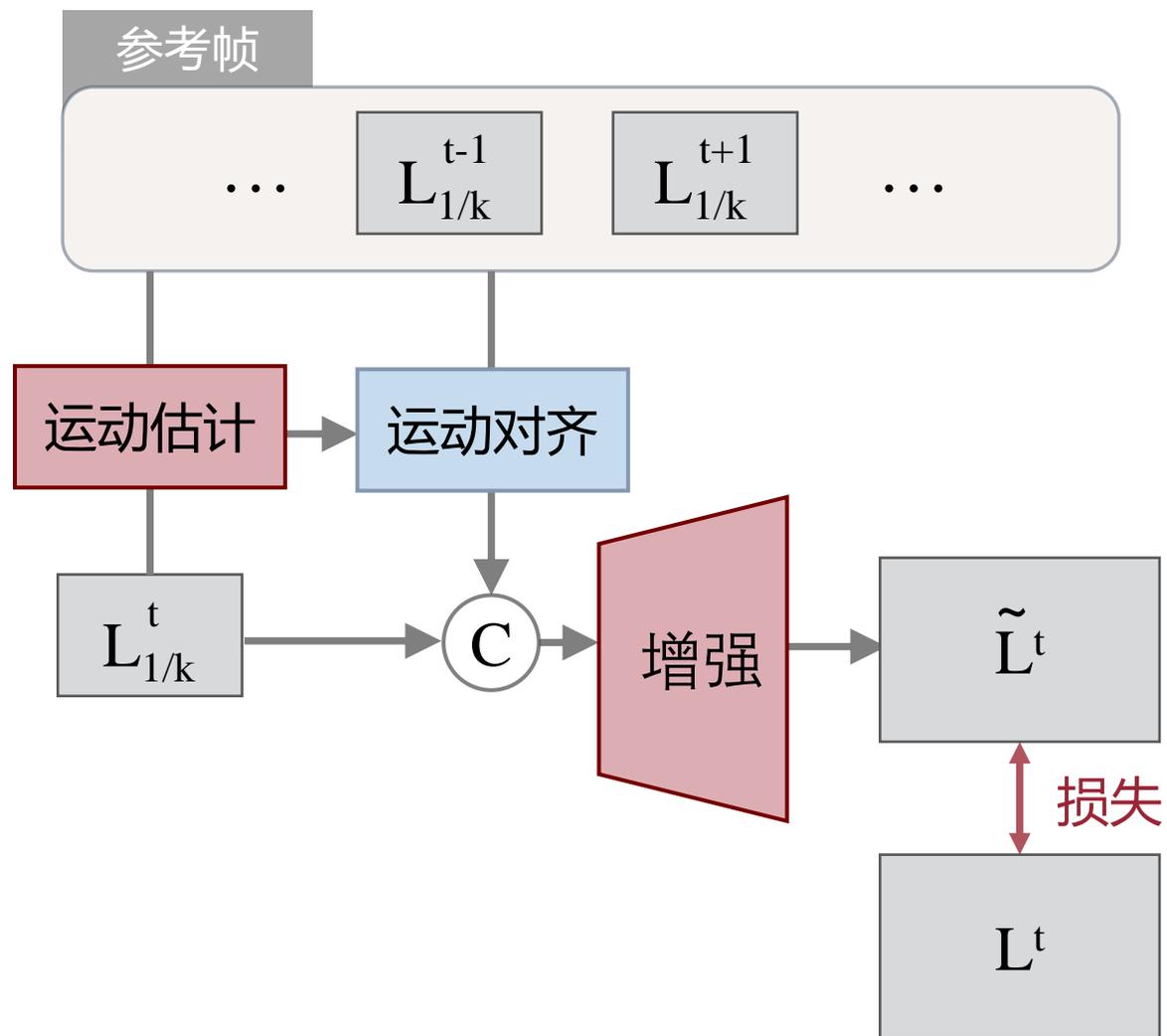
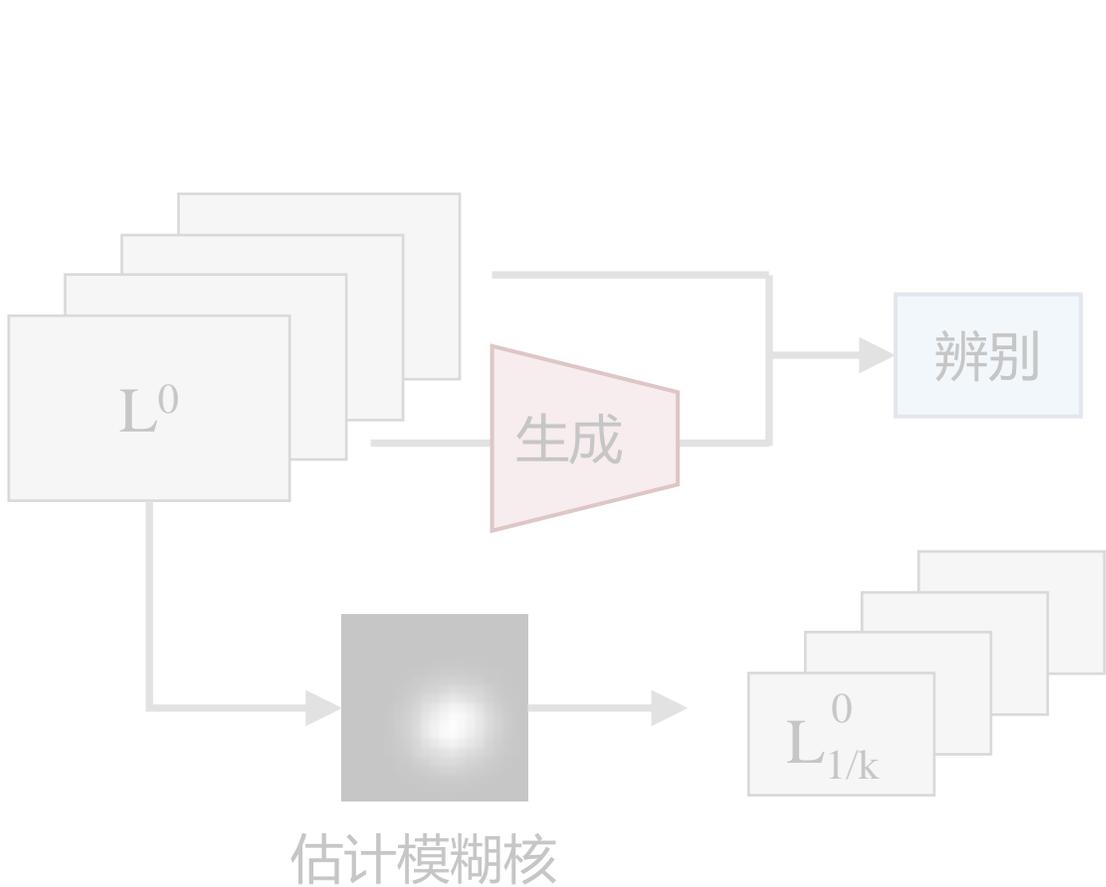
基于自学习的视频超分辨率



基于自学习的视频超分辨率

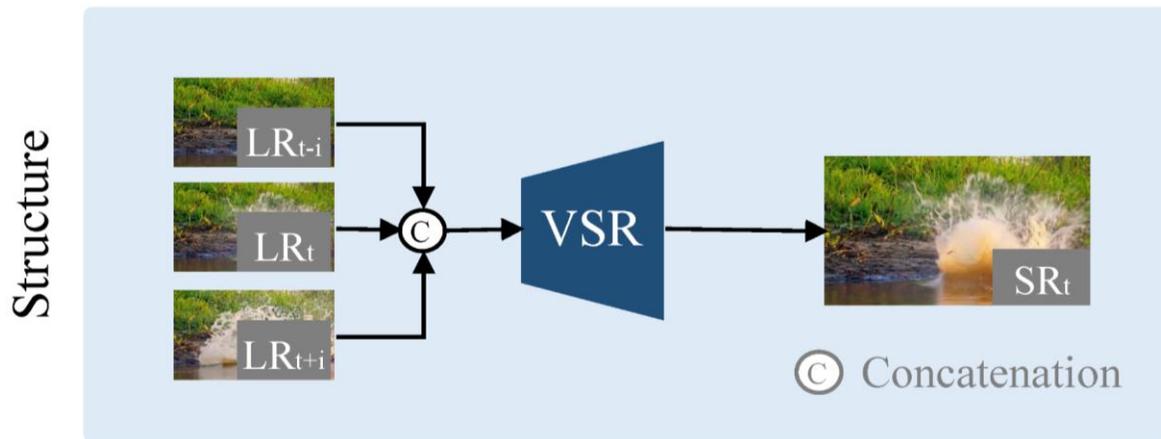


基于自学习的视频超分辨率



基于自学习的视频超分辨率

- 针对时间上下文，使用间隔扩展的采样或逆序采样
- 对于空间上下文，进行旋转、镜像和仿射变换

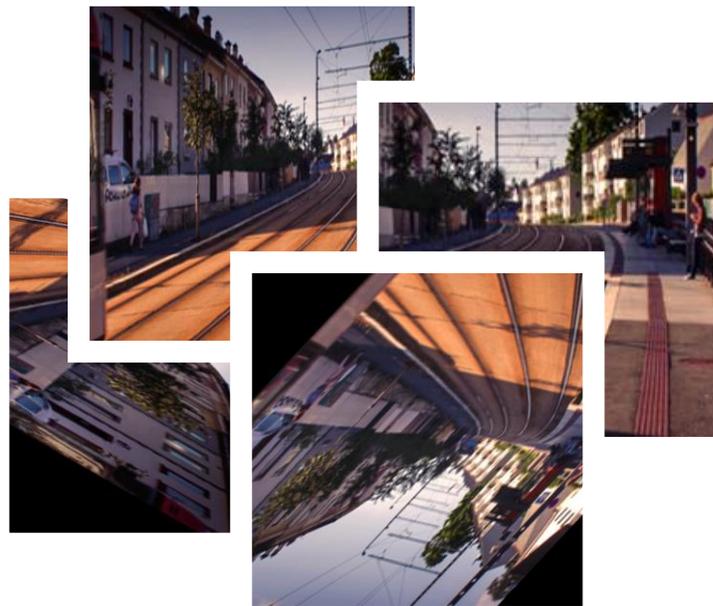


基于自学习的视频超分辨率

- 针对时间上下文，使用间隔扩展的采样或逆序采样
- 对于空间上下文，进行旋转、镜像和仿射变换



generate



实验结果

- 盲模糊核情况下

Method	RCAN (1 Frame)	DUF (7 Frames)	ZSSR (1 Frame)
Calendar	21.46/0.6787	20.08/0.6139	21.54/0.6808
City	25.41/0.6293	23.84/0.5784	25.24/0.6261
Foliage	24.13/0.6453	22.54/0.5697	24.11/0.6431
Walk	25.32/0.7794	23.65/0.7527	25.31/0.7855
Average	24.08/0.6832	22.53/0.6287	24.05/0.6839
Method	KernelGAN (1 Frame)	SinGAN (1 Frame)	SLVSR (3 Frames)
Calendar	22.17/0.7705	20.71/0.6118	25.10/0.8431
City	26.07/0.7026	24.74/0.5744	26.93/0.7459
Foliage	25.76/0.7835	24.15/0.6021	28.66/0.8651
Walk	26.96/0.8433	24.43/0.7465	27.86/0.8574
Average	25.24/0.7750	23.45/0.6337	27.14/0.8279

实验结果

- 盲模糊核情形下



LR input



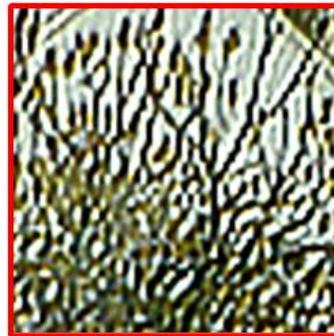
RCAN



DUF



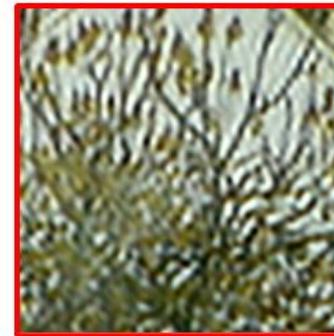
ZSSR



KernelGAN



SinGAN



SLVSR



Ground Truth

实验结果

- 盲模糊核情形下



LR input



RCAN



DUF



ZSSR



KernelGAN



SinGAN



SLVSR



Ground Truth

实验结果

- 盲模糊核情形下



LR input



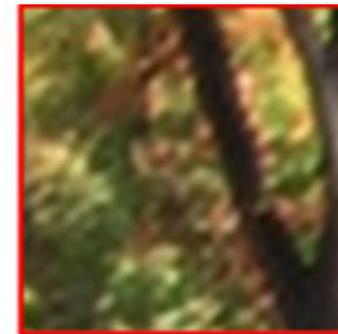
RCAN



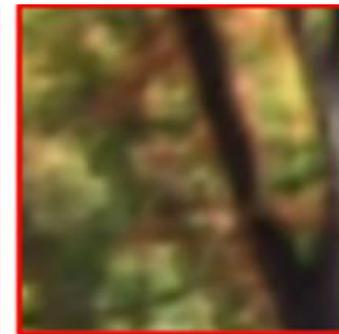
DUF



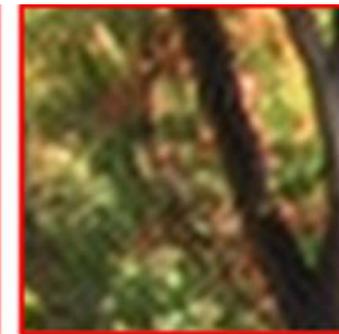
ZSSR



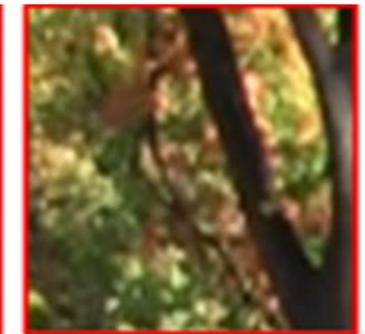
KernelGAN



SinGAN



SLVSR



Ground Truth

实验结果

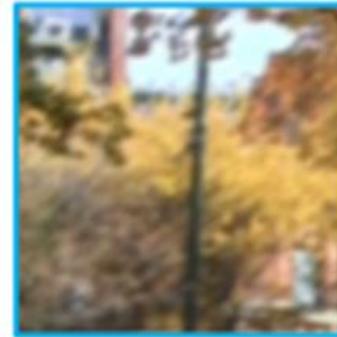
- 盲模糊核情形下



LR input



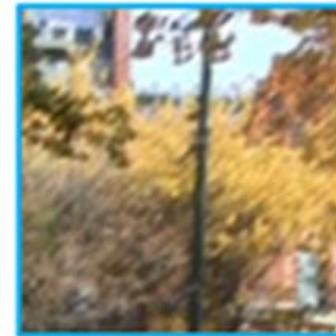
RCAN



DUF



ZSSR



KernelGAN



SinGAN



SLVSR



Ground Truth

实验结果

- 真实拍摄情形下



LR input



RCAN



DUF



ZSSR



KernelGAN



SinGAN



SLVSR

实验结果

- 真实拍摄情形下



LR input



RCAN



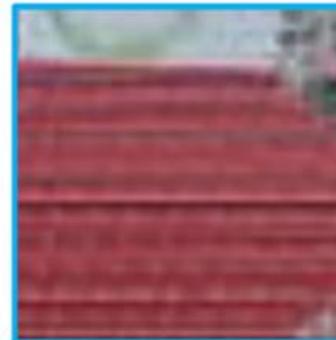
DUF



ZSSR



KernelGAN

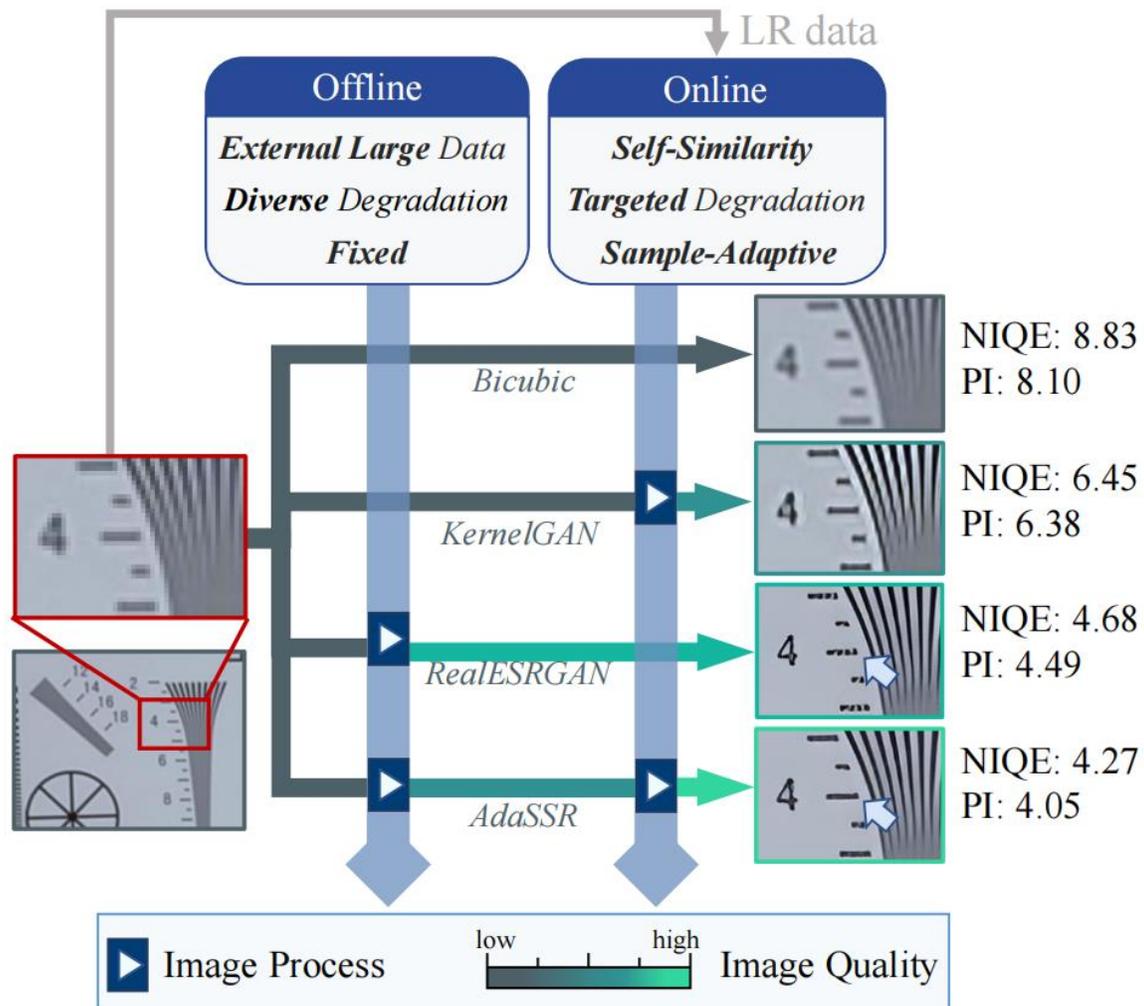


SinGAN

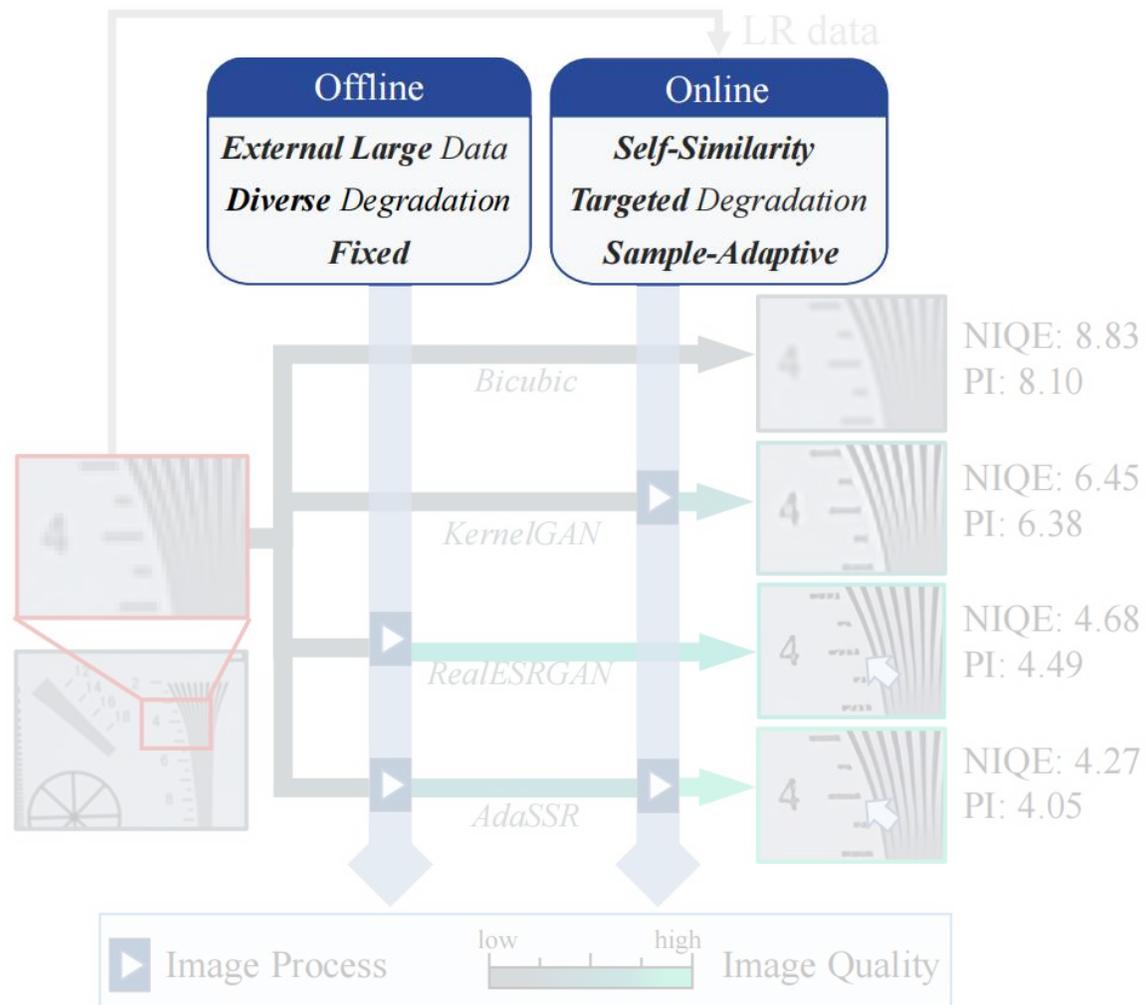


SLVSR

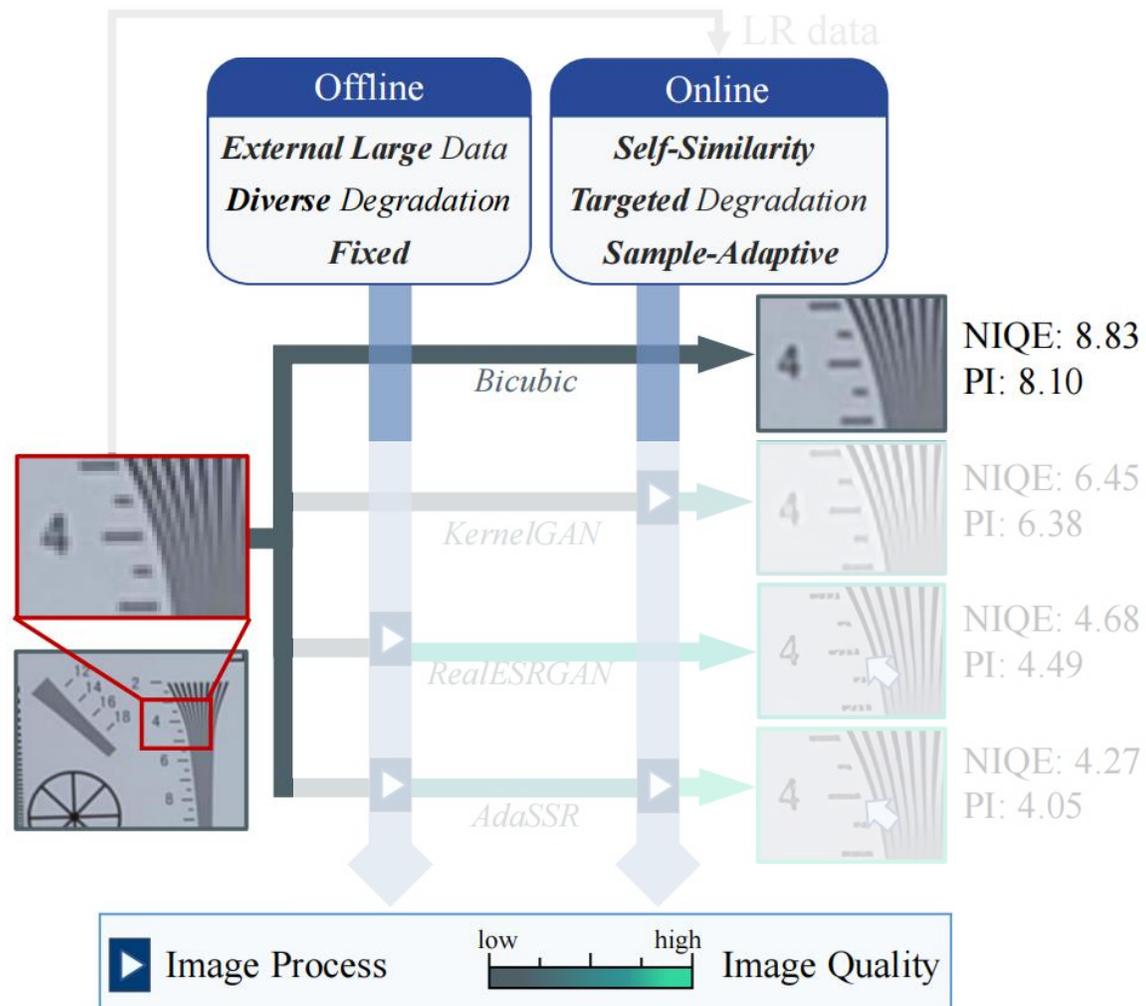
半监督真实场景超分辨率



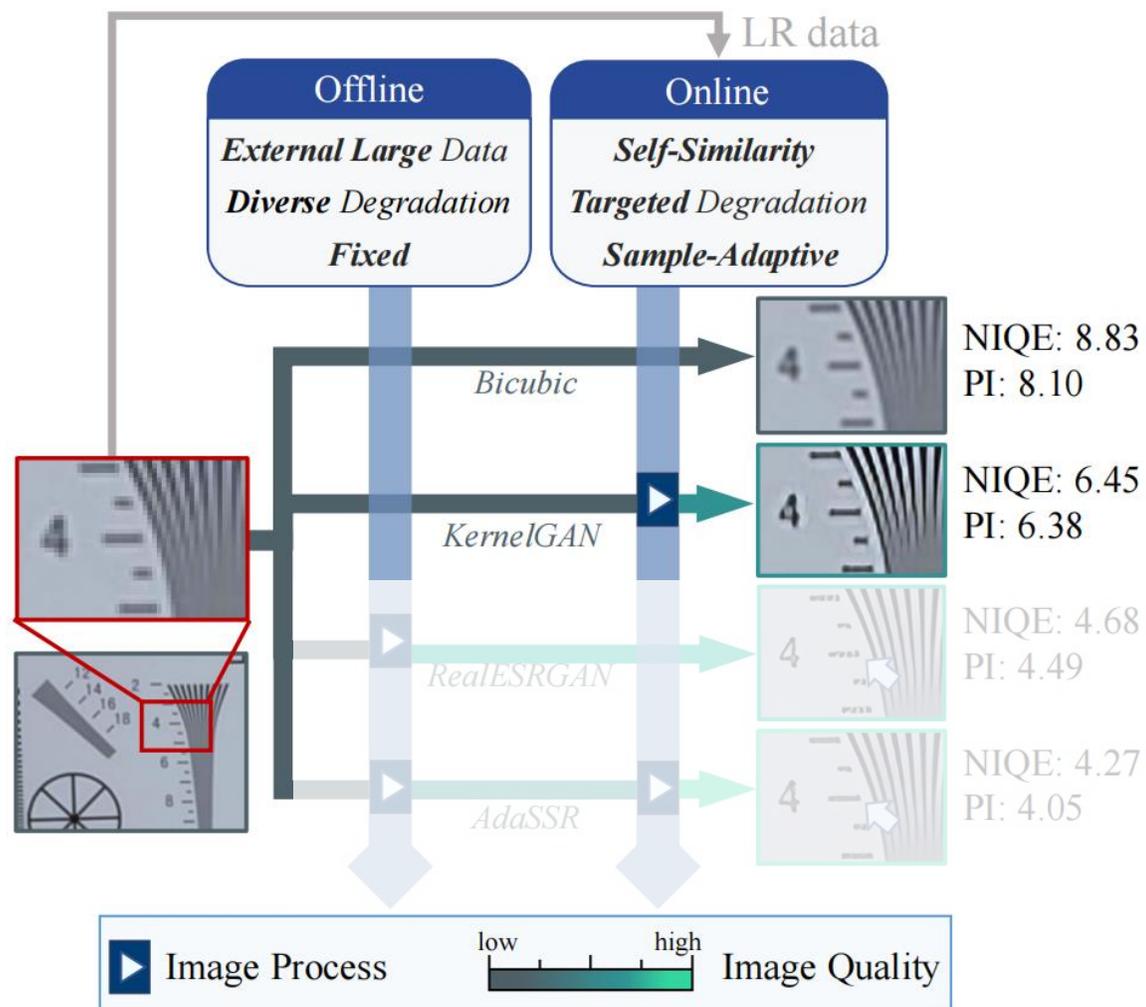
半监督真实场景超分辨率



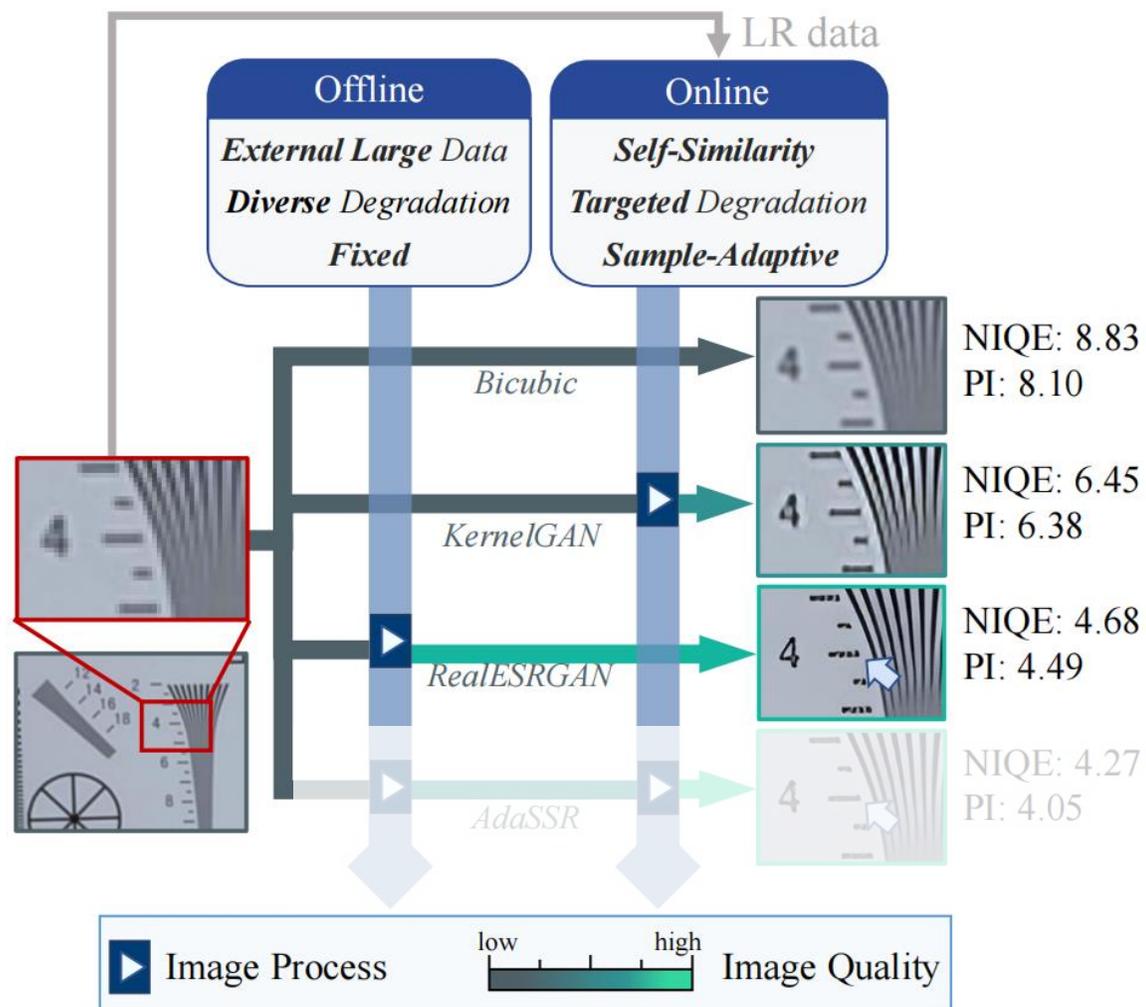
半监督真实场景超分辨率



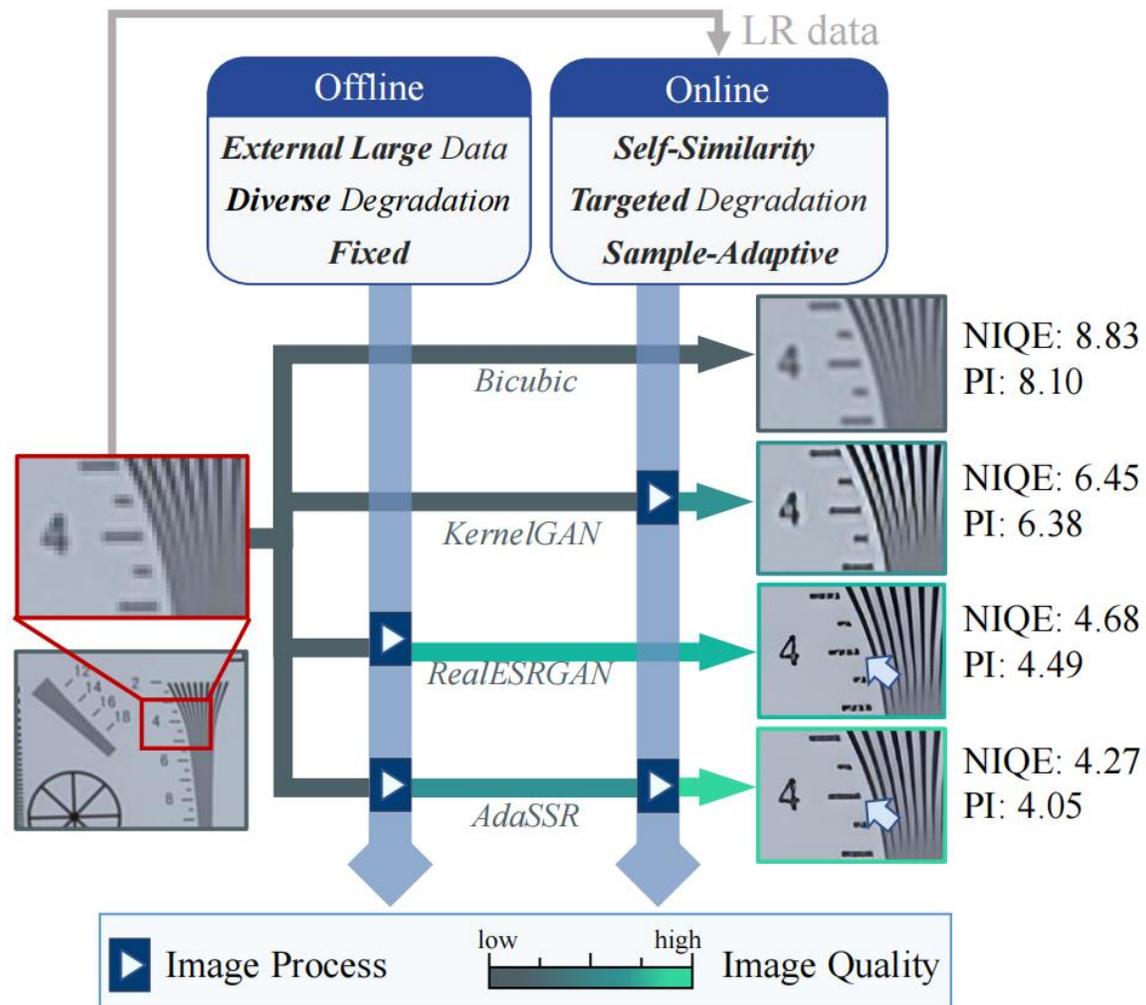
半监督真实场景超分辨率



半监督真实场景超分辨率

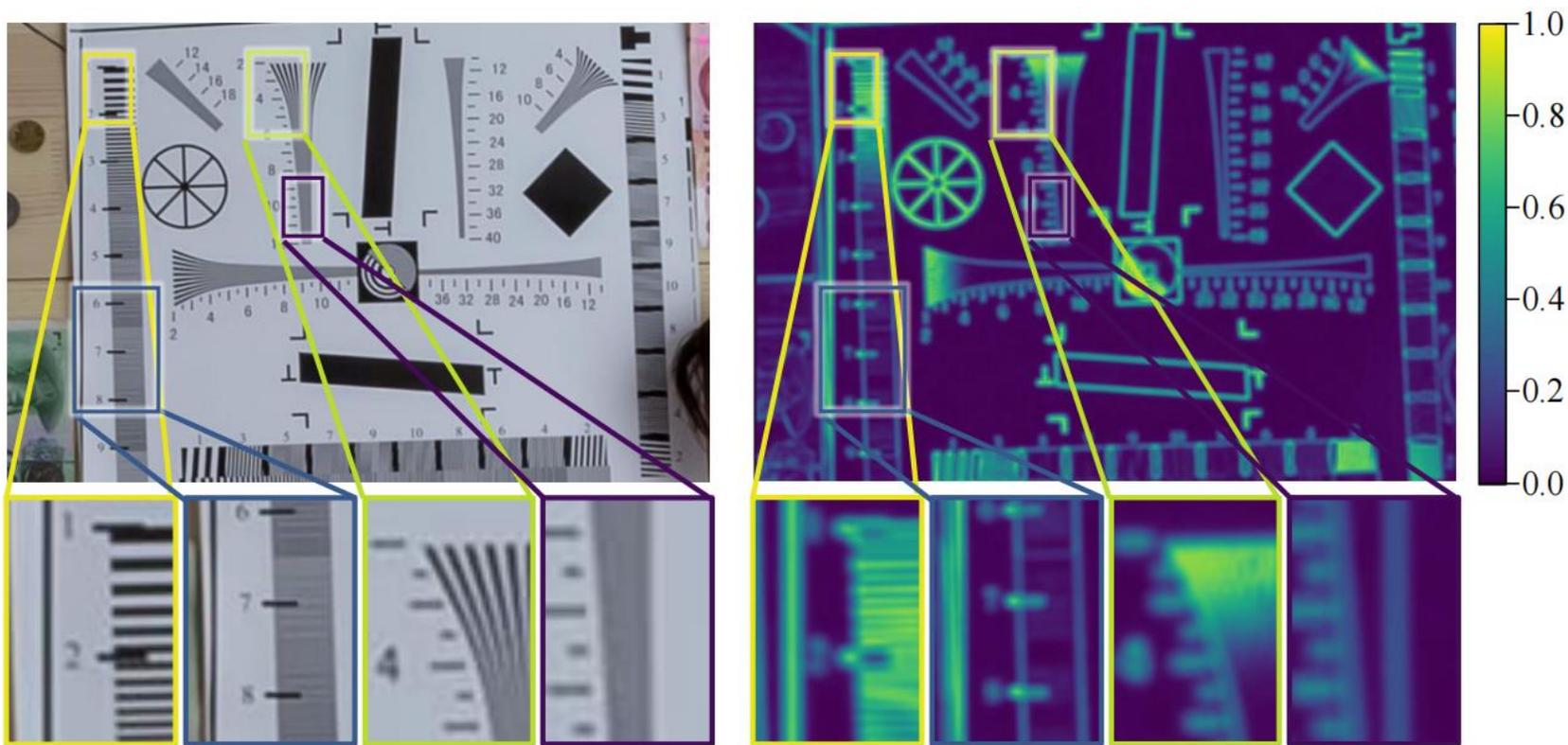


半监督真实场景超分辨率



半监督真实场景超分辨率

- 在外部学习的先验基础上，进行权重图引导的内部学习

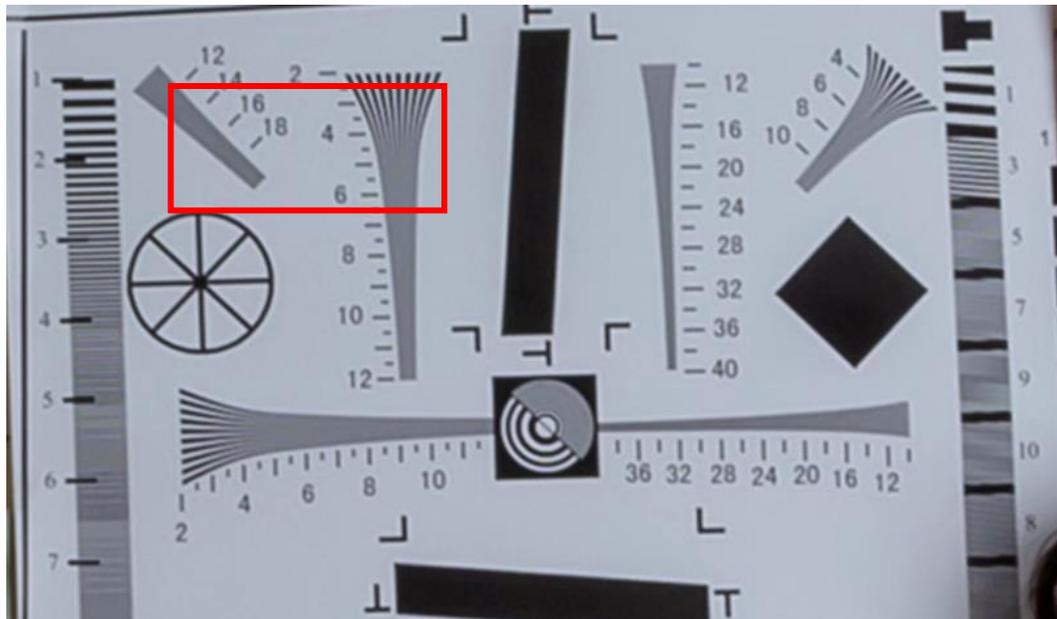


实验结果

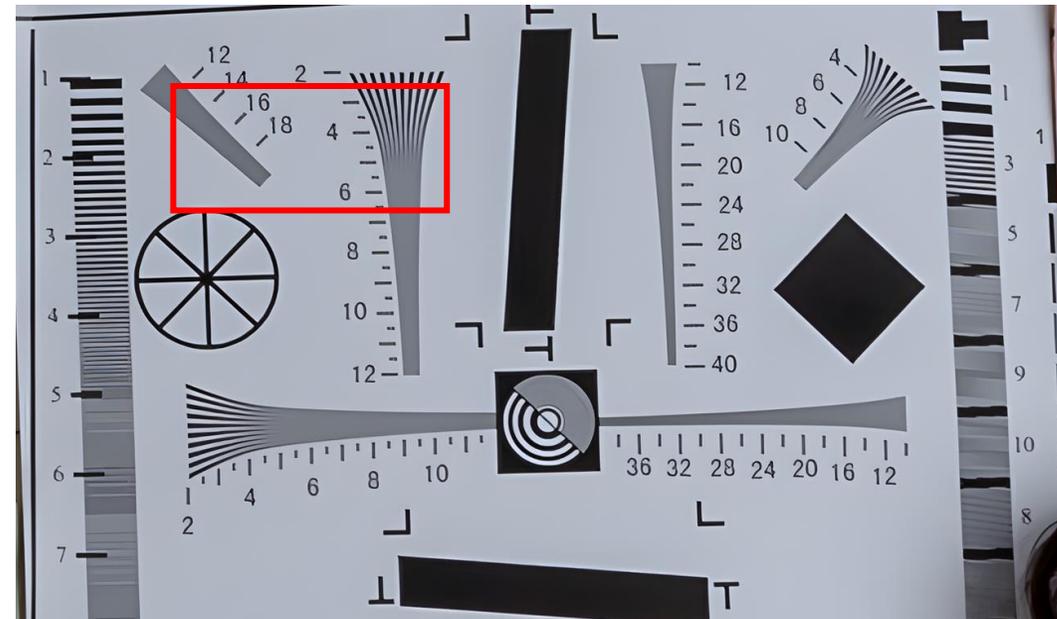
- 在真实场景采集的低质图像上

	NIQE ↓	BRISQUE ↓	PI ↓	LPIPS ↓	SSEQ ↓	NRQM ↑
Bicubic	8.83	67.15	8.10	0.477	57.74	2.72
<i>Self-learning method</i>						
ZSSR	7.91	62.65	7.54	0.406	52.84	3.03
KernelGAN	6.45	52.53	6.38	0.317	48.07	3.80
<i>Supervised method</i>						
RCAN	8.55	66.68	7.84	0.442	56.62	3.01
HAT	5.26	32.23	5.08	0.247	25.58	5.38
SWINIR	4.68	33.79	4.71	0.252	28.66	5.50
BSRGAN	4.65	25.35	4.46	0.269	23.04	5.98
RealESRGAN	4.68	29.13	4.49	0.273	24.51	5.87
<i>Semi-supervised method</i>						
AdaSSR+SWINIR	4.56	32.84	4.56	0.244	28.92	5.67
AdaSSR+BSRGAN	4.54	25.15	4.34	0.263	23.69	6.09
AdaSSR+RealESRGAN	4.25	24.47	4.02	0.265	24.21	6.34

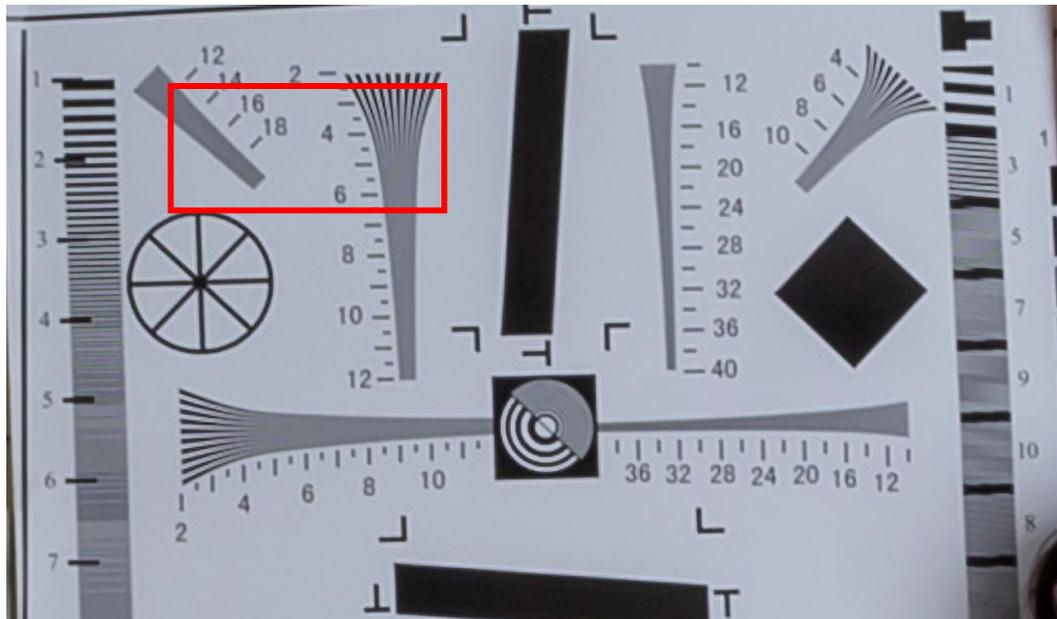
Input



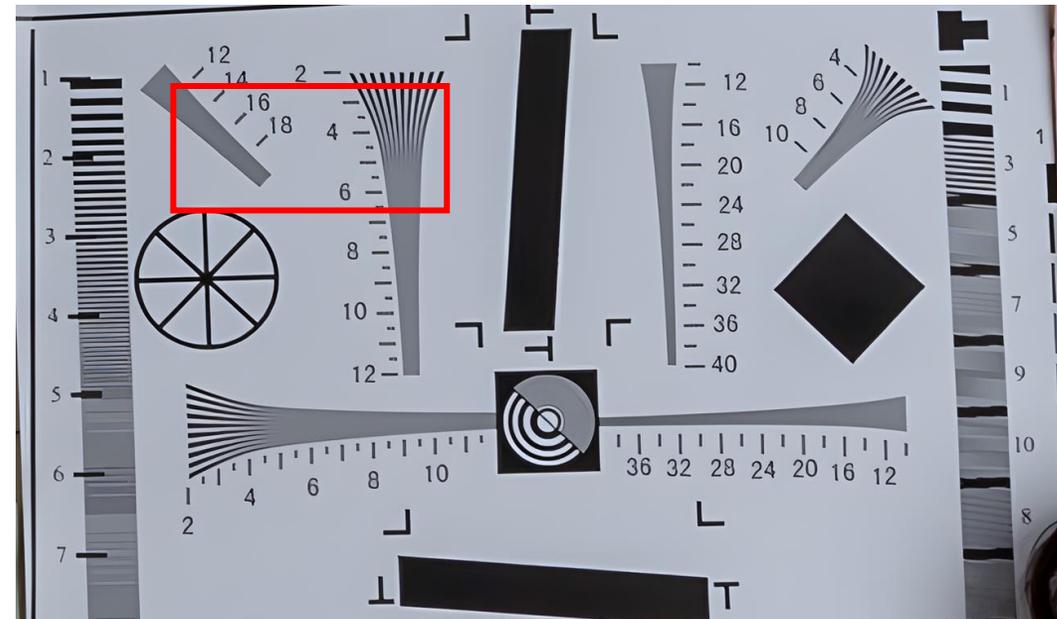
AdaSSR



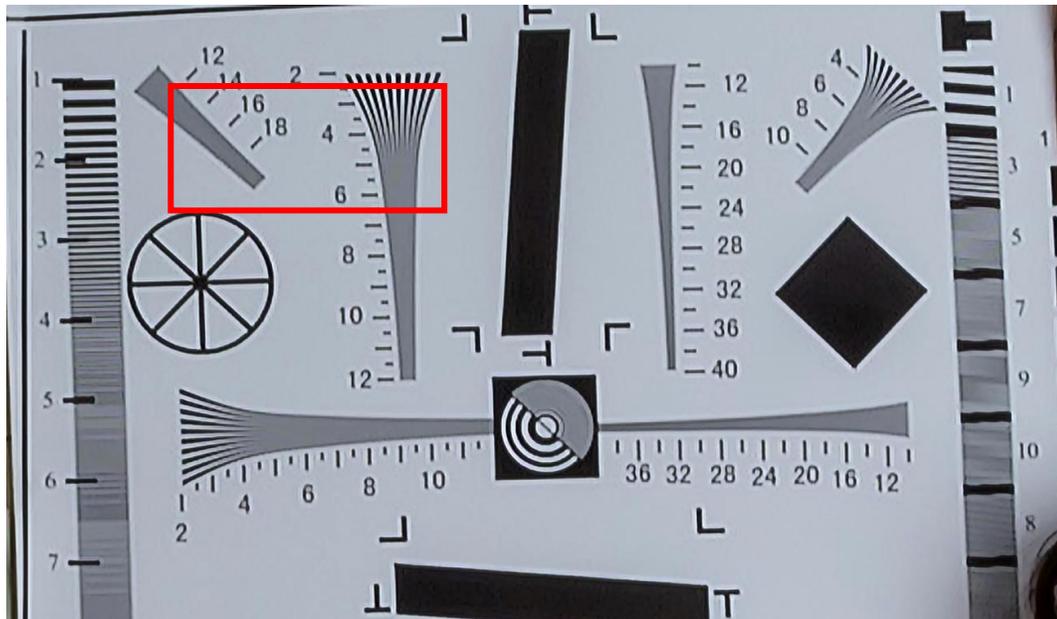
ZSSR



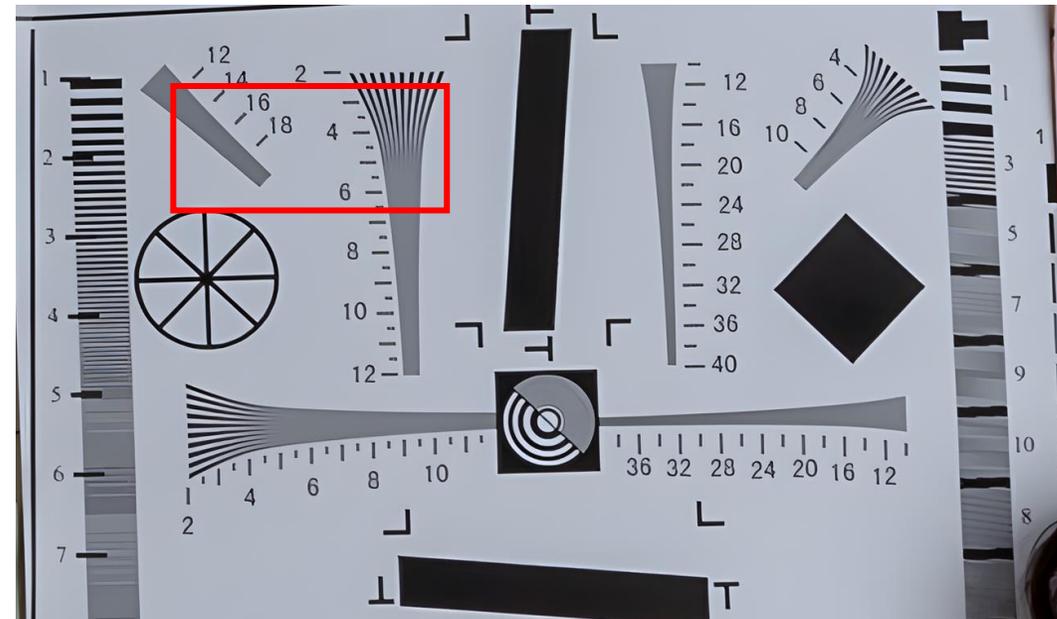
AdaSSR



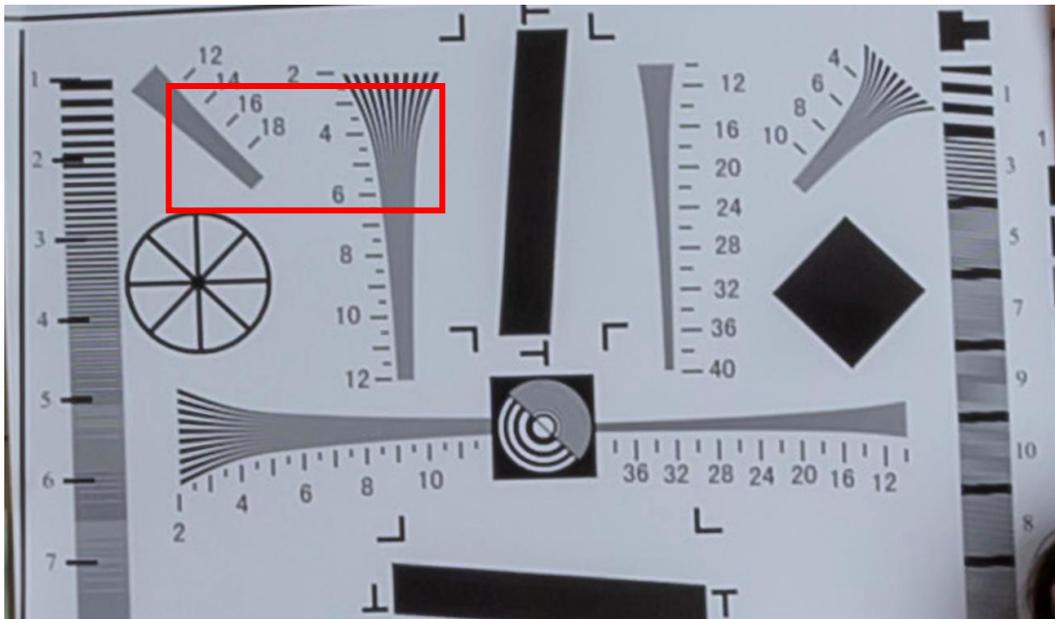
KernelGAN



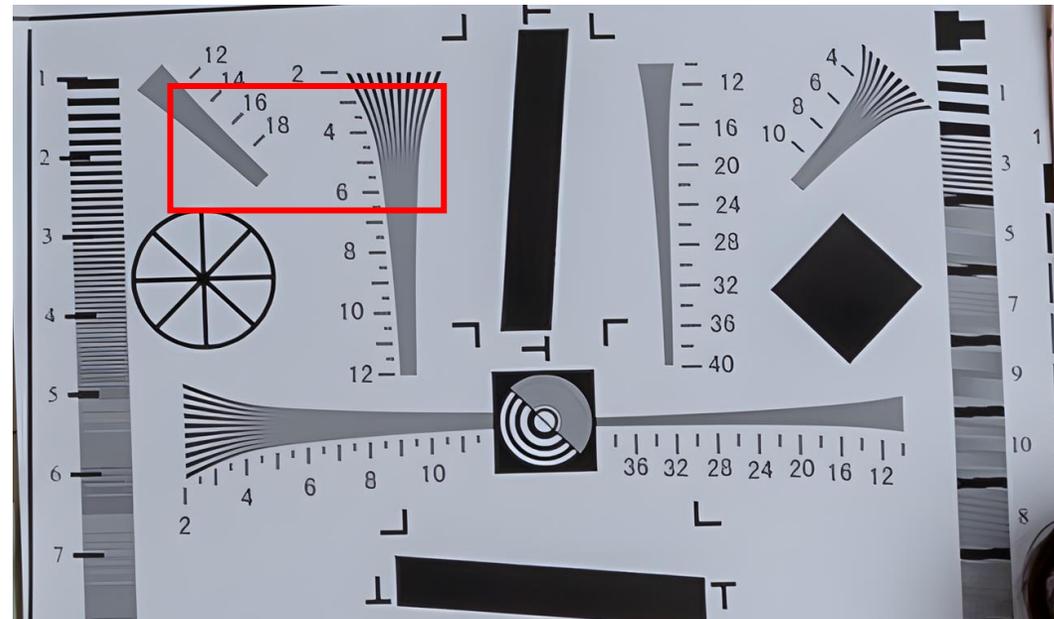
AdaSSR



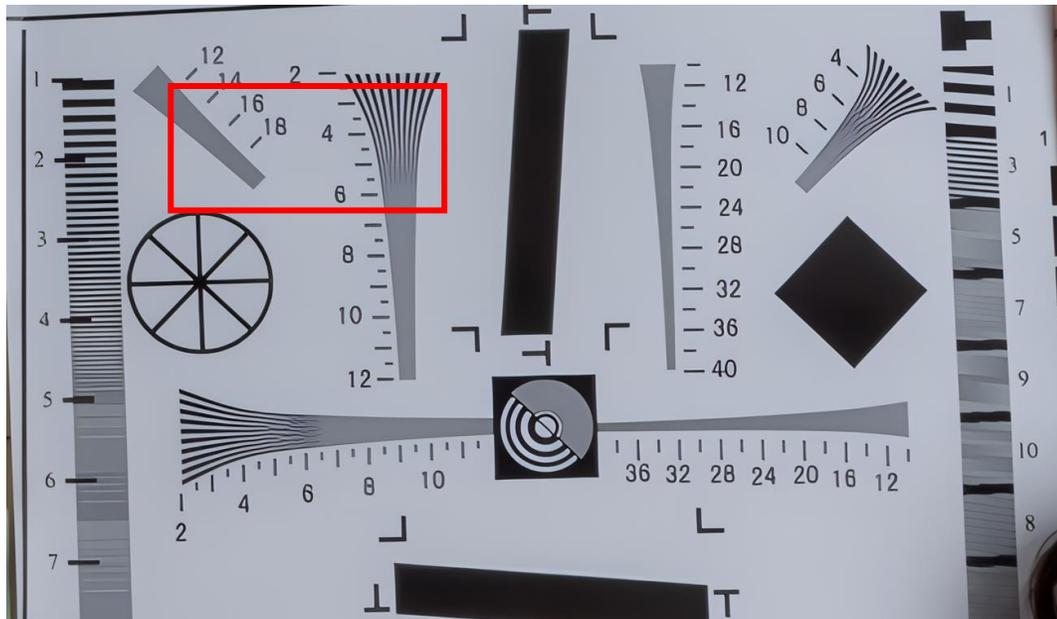
RCAN



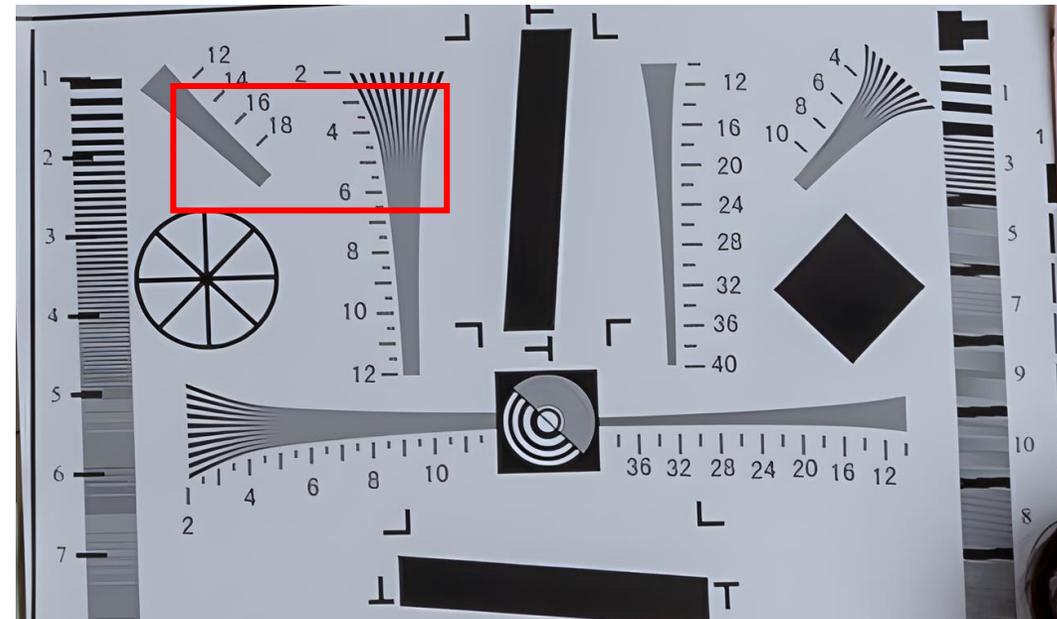
AdaSSR



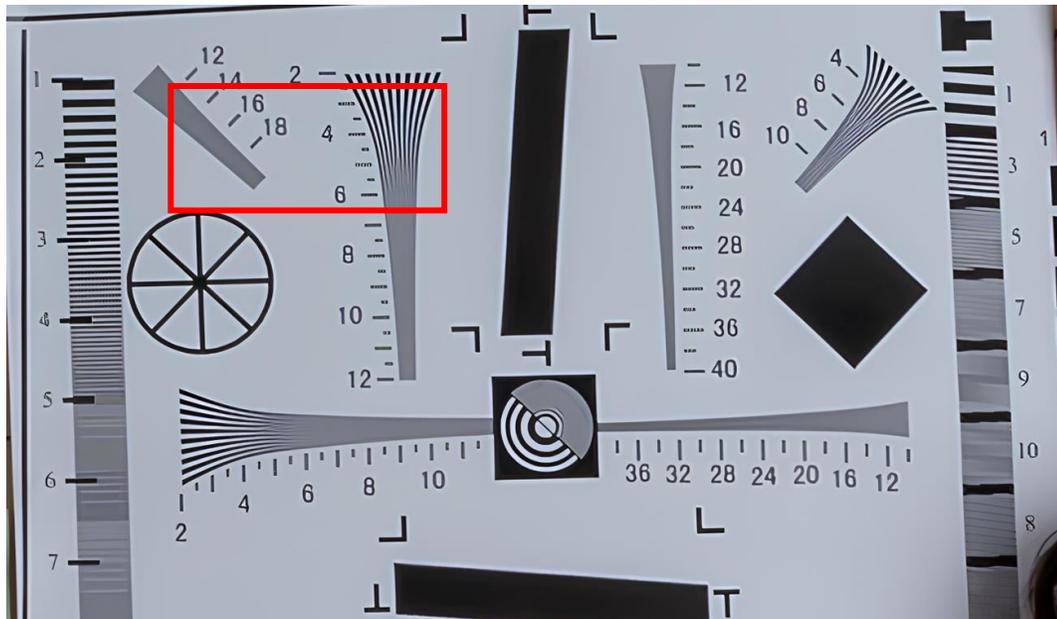
HAT



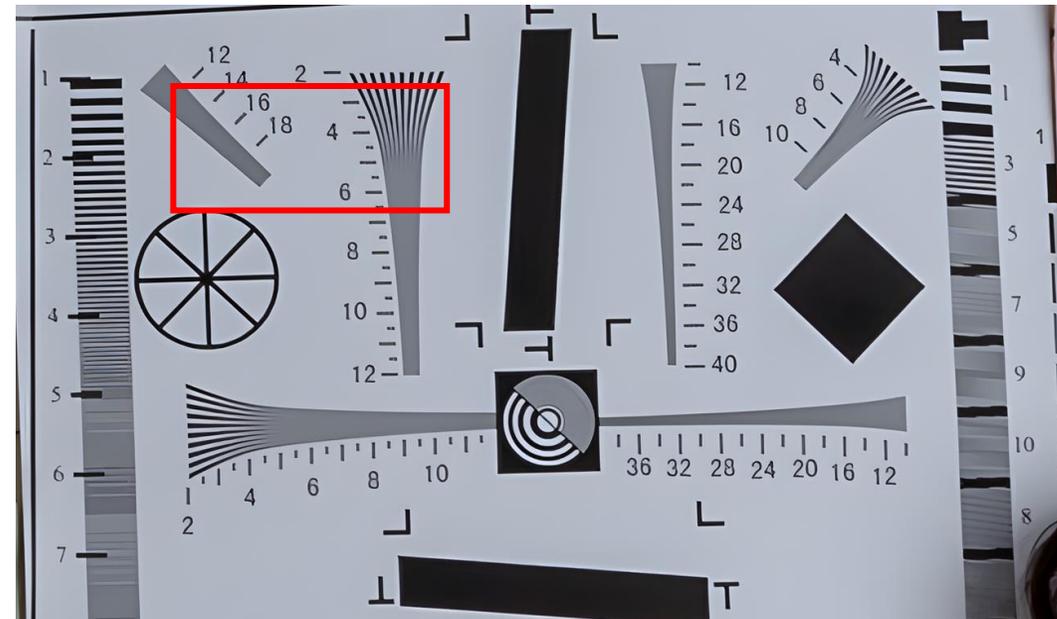
AdaSSR



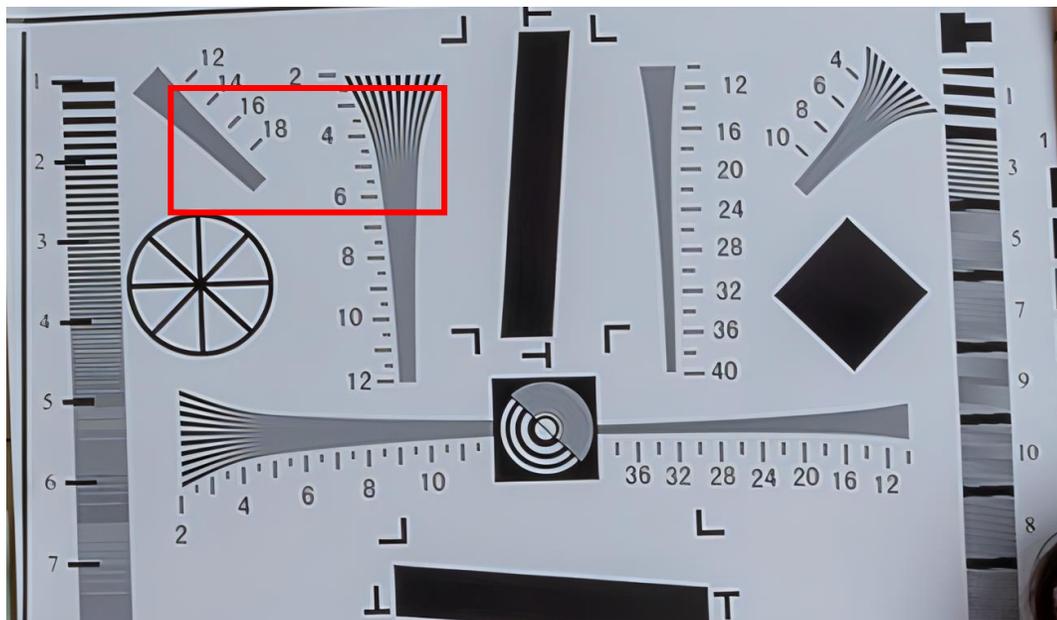
SWINIR



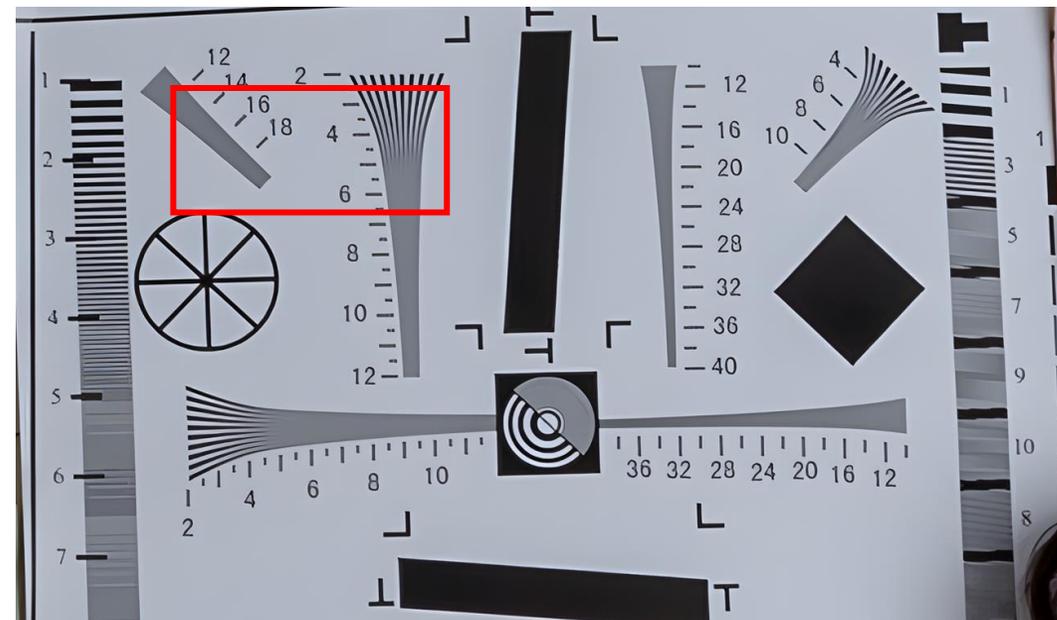
AdaSSR



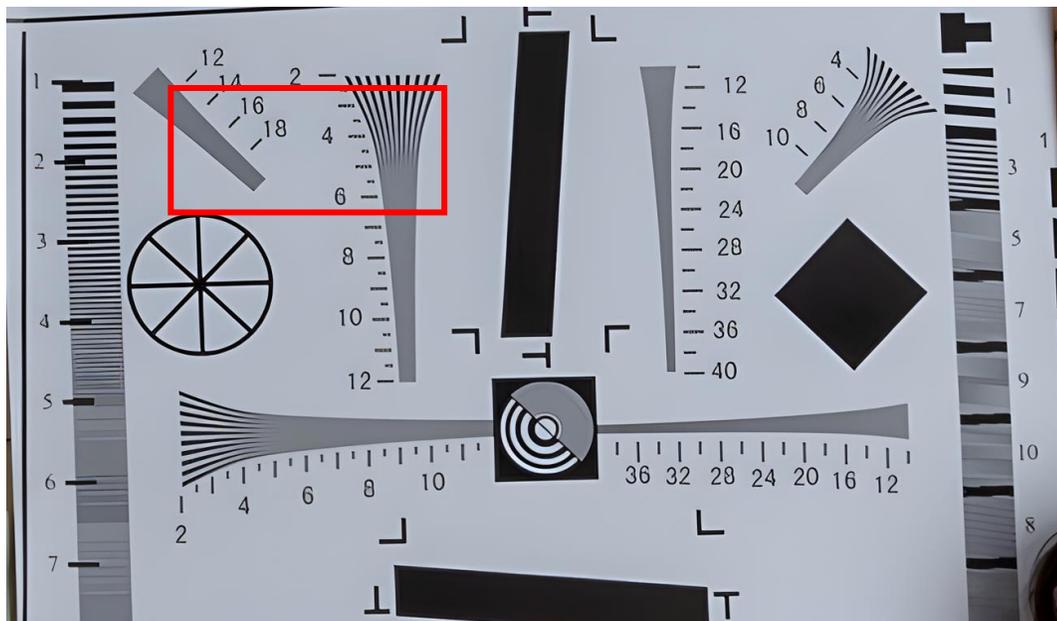
BSRGAN



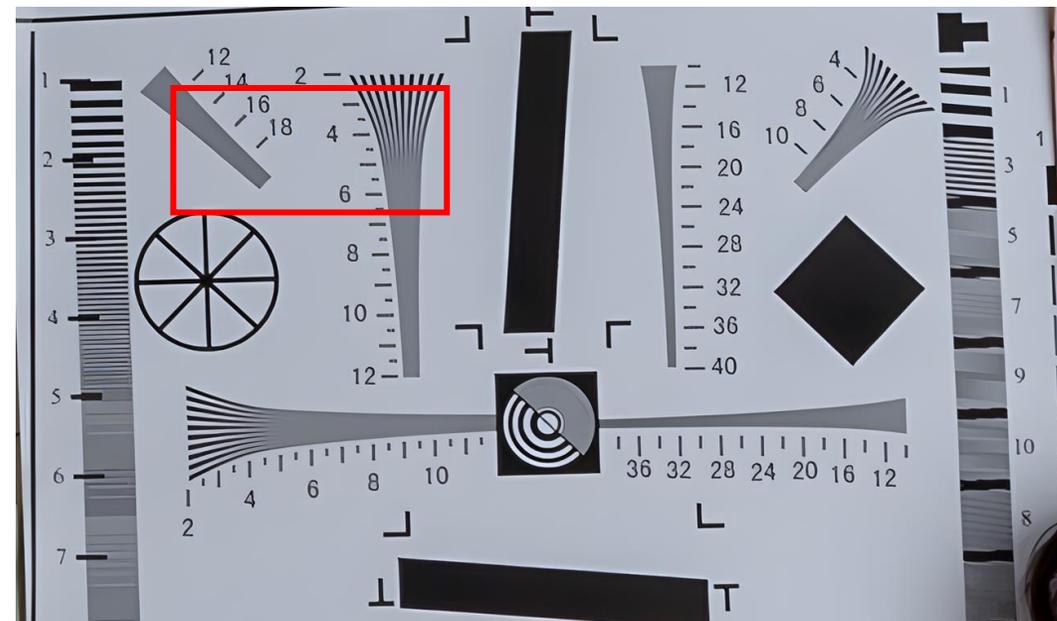
AdaSSR



RealESRGAN



AdaSSR



拍摄结果



超分辨率结果



期望结果



期望结果



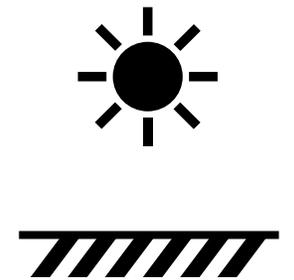


暗光信号解耦与智能成像

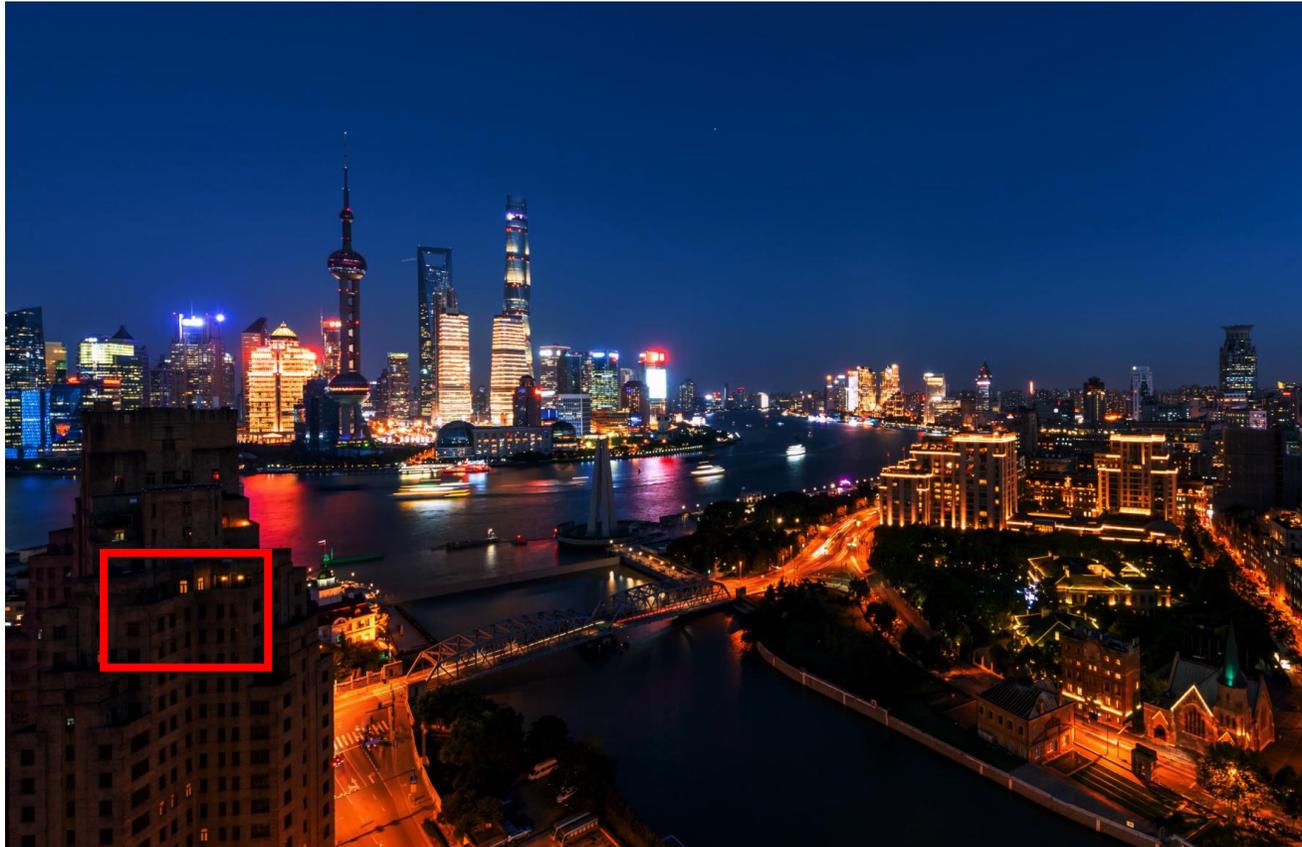
Decoupling of Low-Light Signals and Intelligent Imaging

黄浩峰 北京大学









任务场景



手机拍照



数码摄影



监控视频

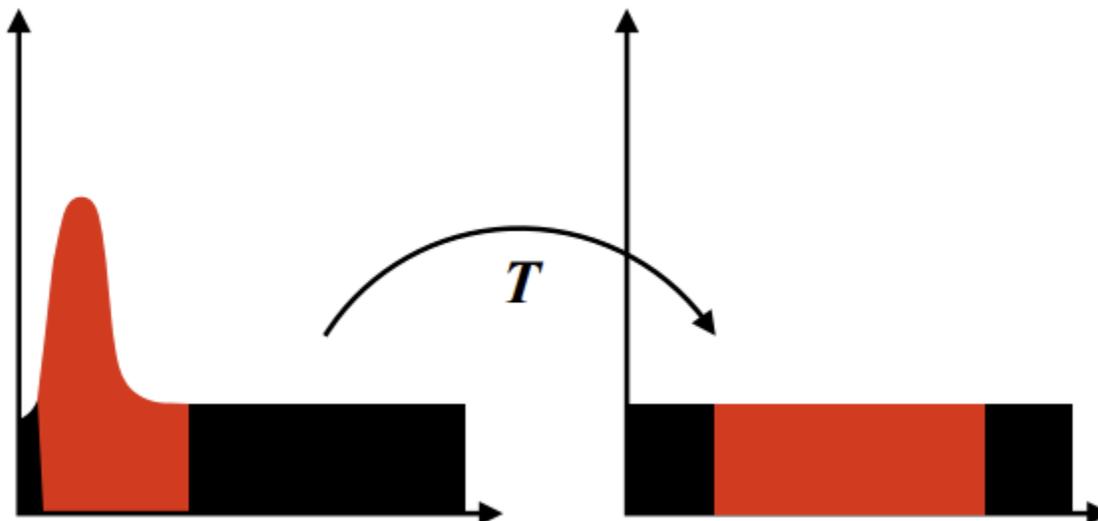
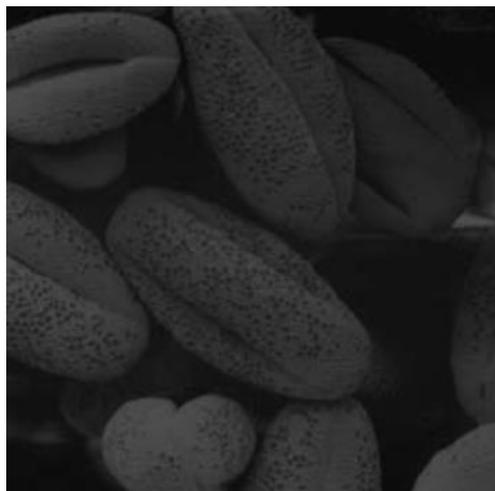
物体不可视、低对比度、强噪声



相关工作

- 传统方法
- 监督学习
- 半监督学习
- 无监督学习

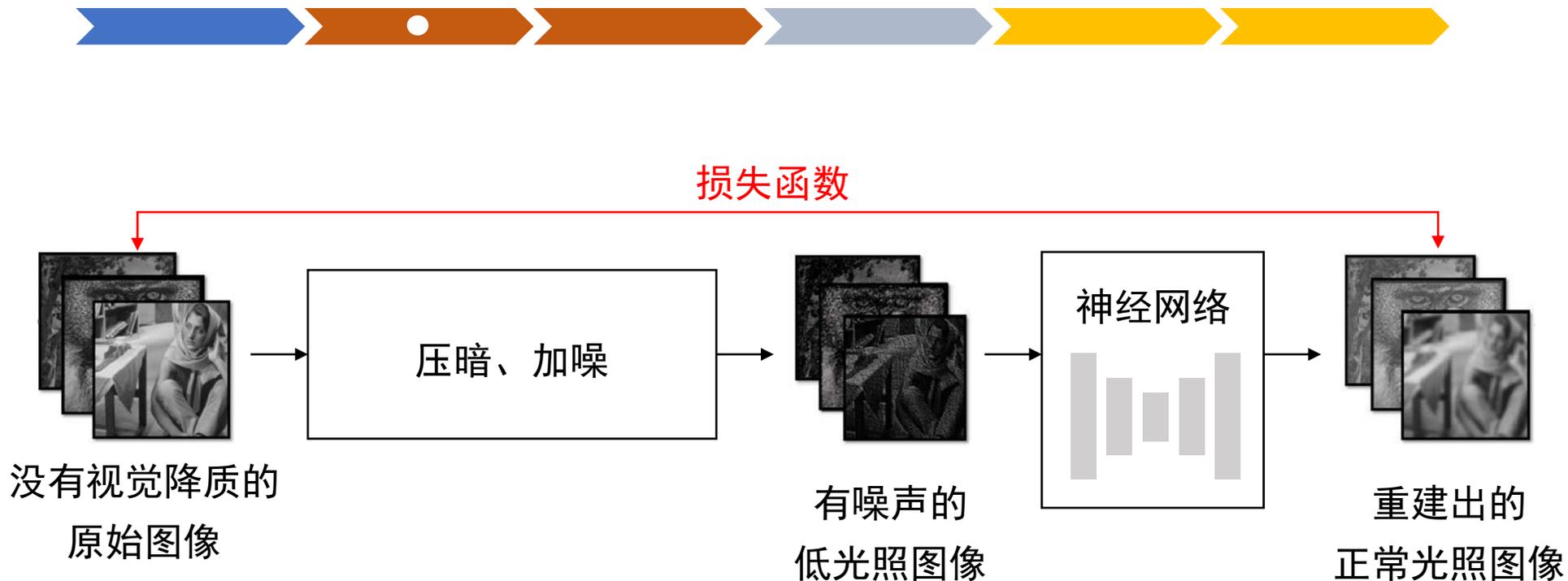
Histogram
Equalization



相关工作

- 传统方法
- 监督学习
- 半监督学习
- 无监督学习

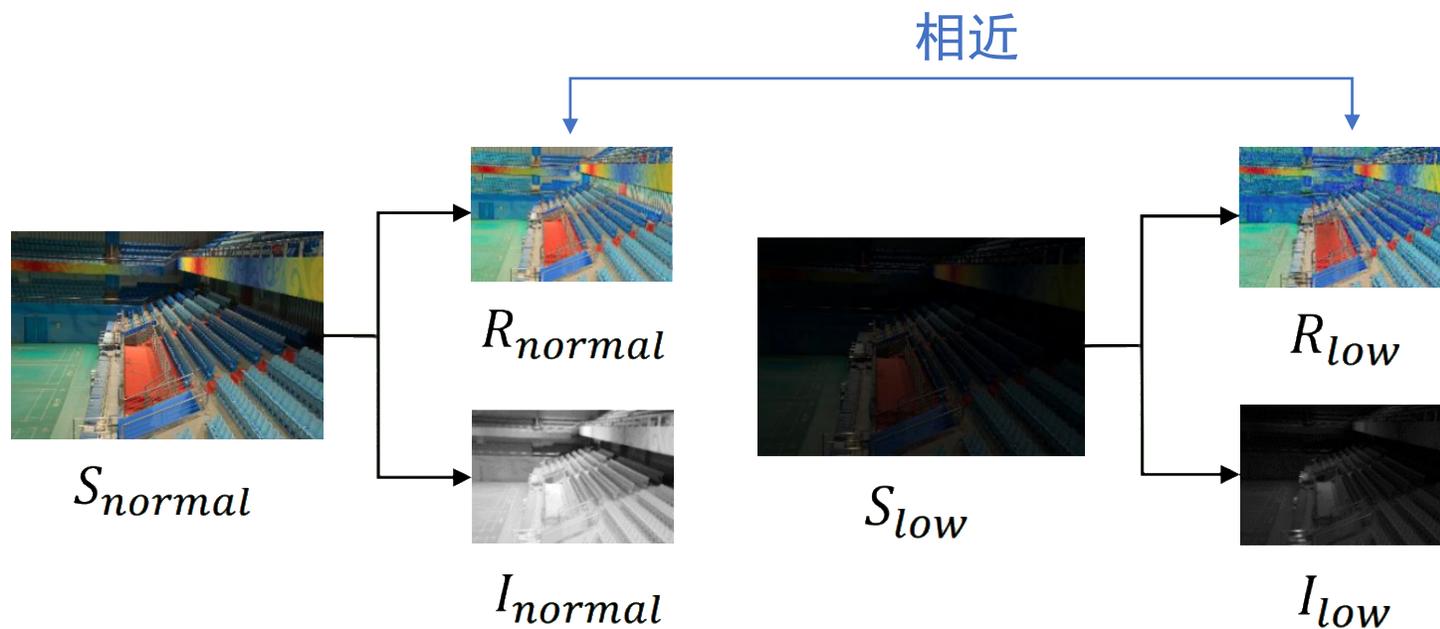
2017 PR
LLNet



相关工作

- 传统方法
- 监督学习
- 半监督学习
- 无监督学习

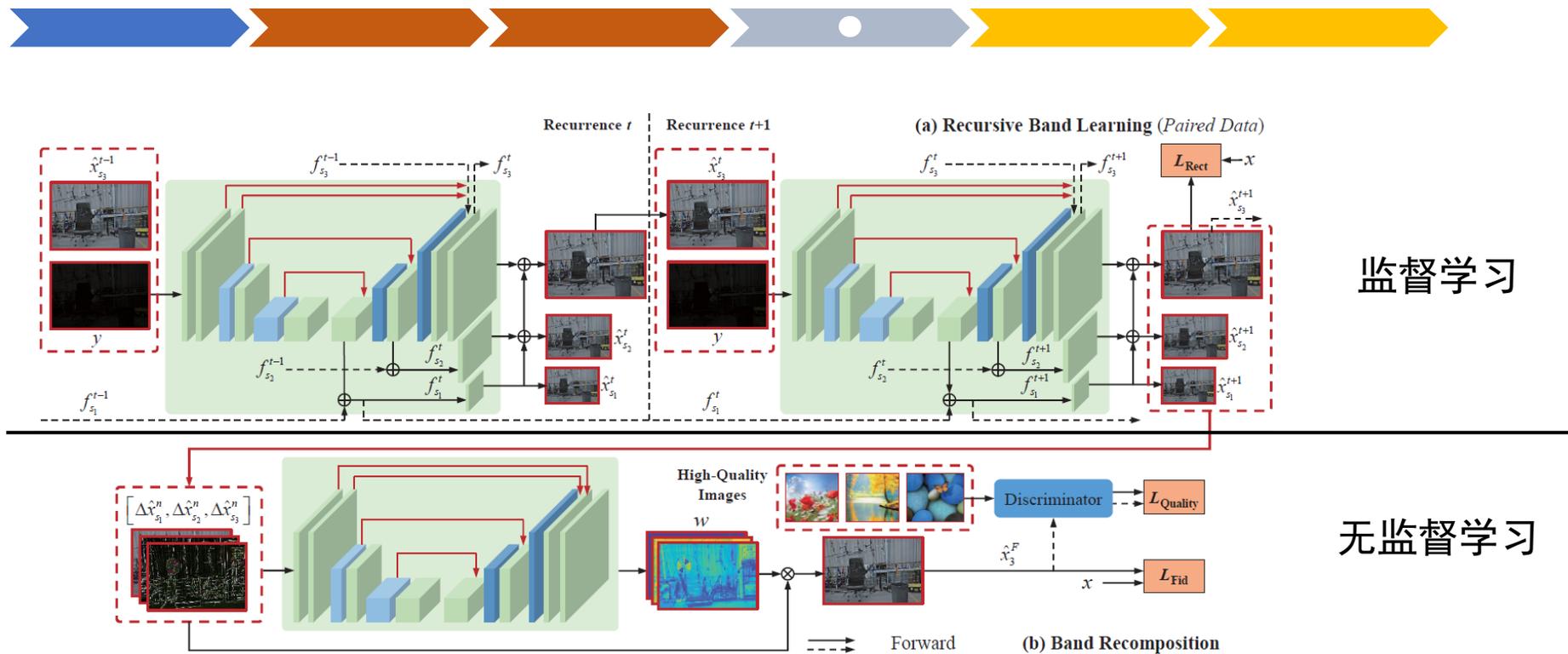
2018 BMVC
RetinexNet



相关工作

- 传统方法
- 监督学习
- 半监督学习
- 无监督学习

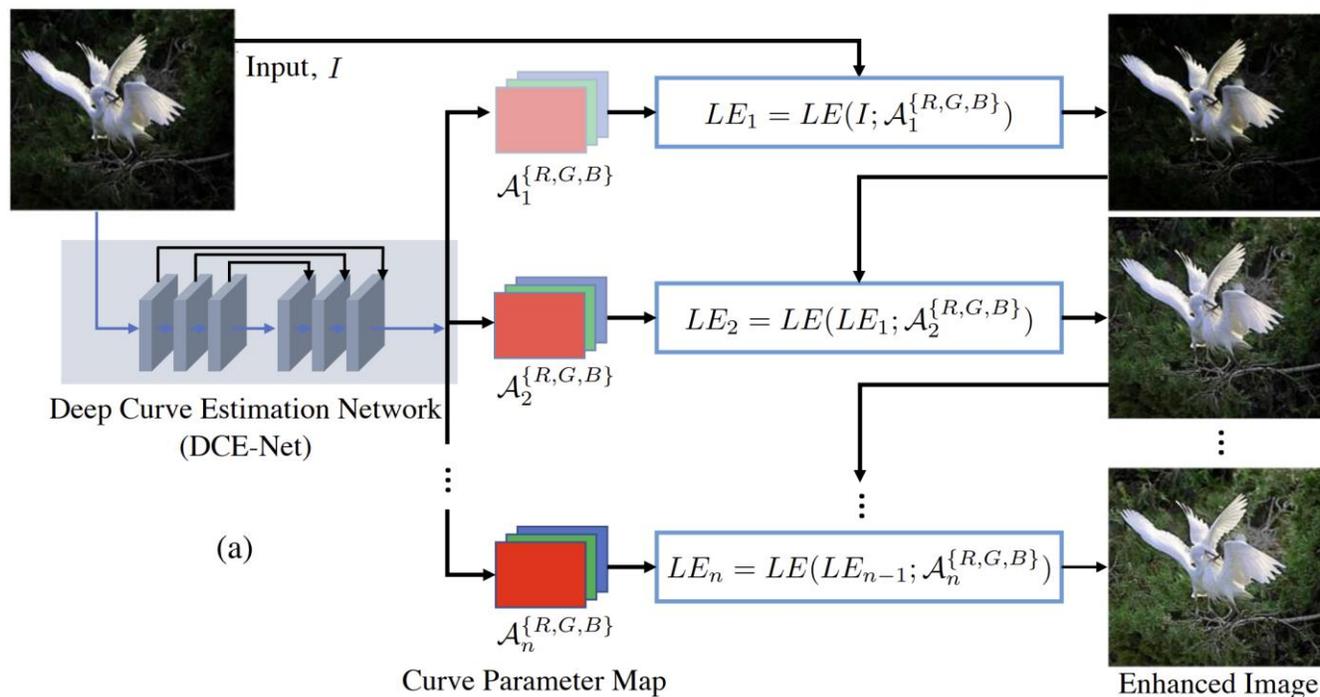
2020 CVPR
DRBN



相关工作

- 传统方法
- 监督学习
- 半监督学习
- 无监督学习

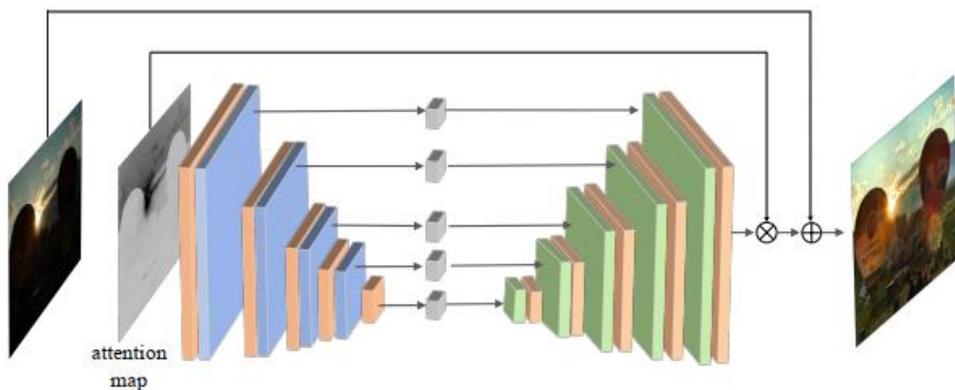
2020 CVPR
Zero-DCE



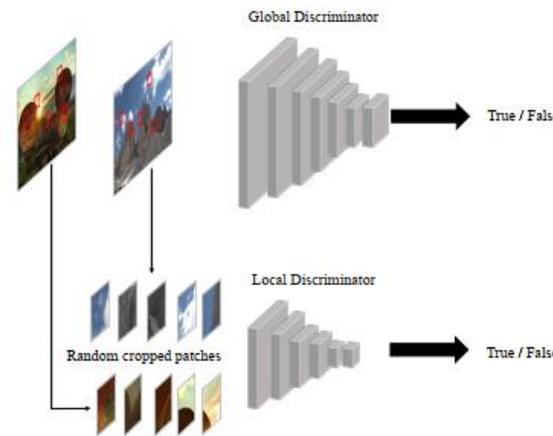
相关工作

- 传统方法
- 监督学习
- 半监督学习
- 无监督学习

2021 IEEE TIP
EnlightenGAN



自注意力机制 +
编码器-解码器的生成器网络结构



全局-局部的判别器设计

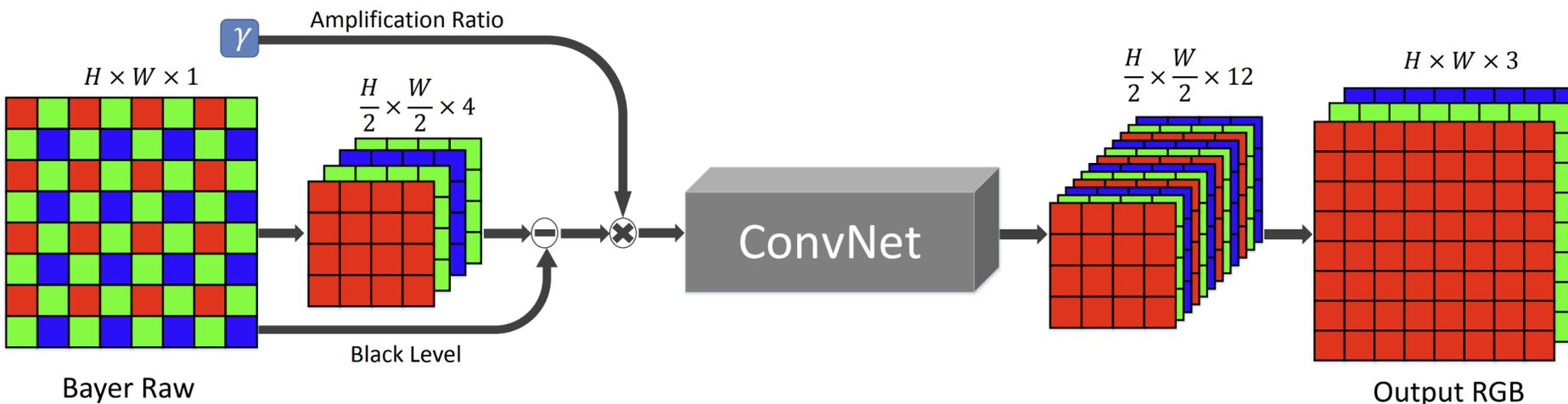
极暗光场景的挑战



相关工作

- 数据集构建
- 模型优化
- 噪声建模

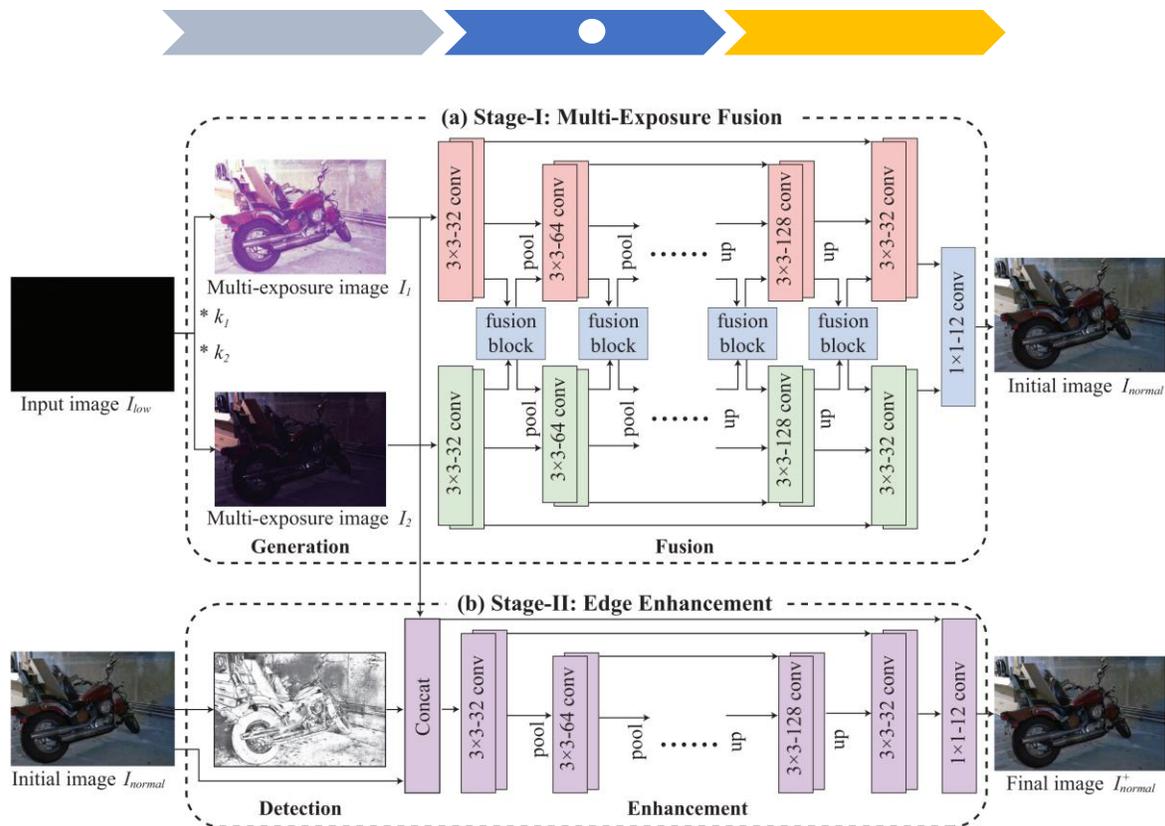
2018 CVPR
SID



相关工作

- 数据集构建
- 模型优化
- 噪声建模

2019 AAAI
EEMEFN

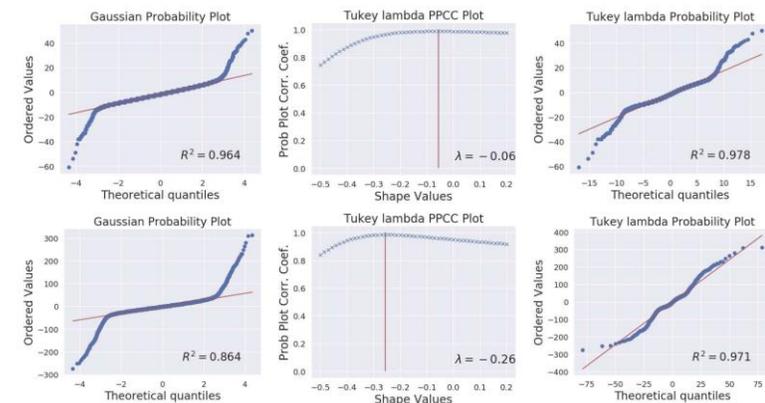
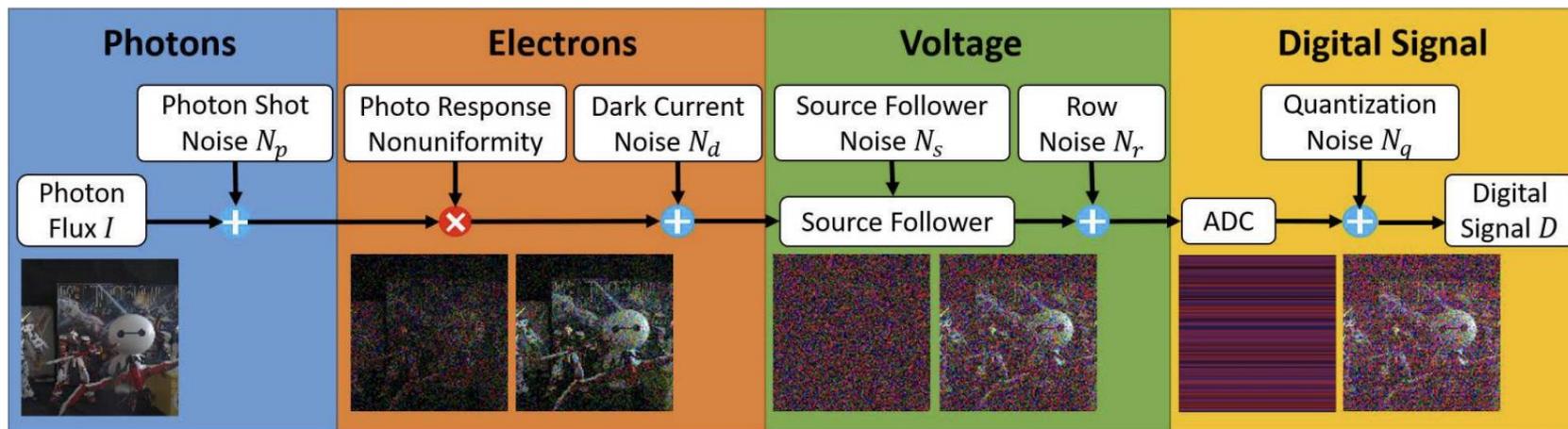


相关工作

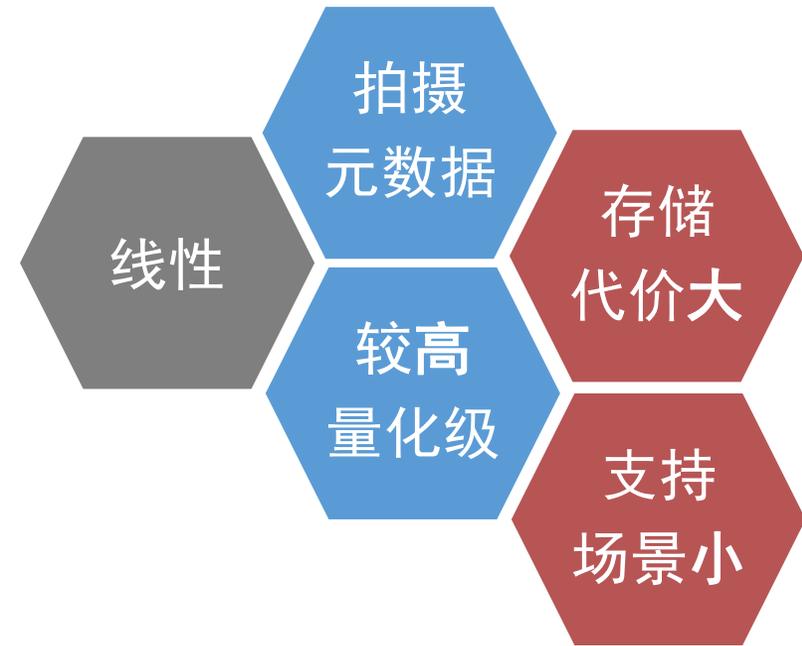
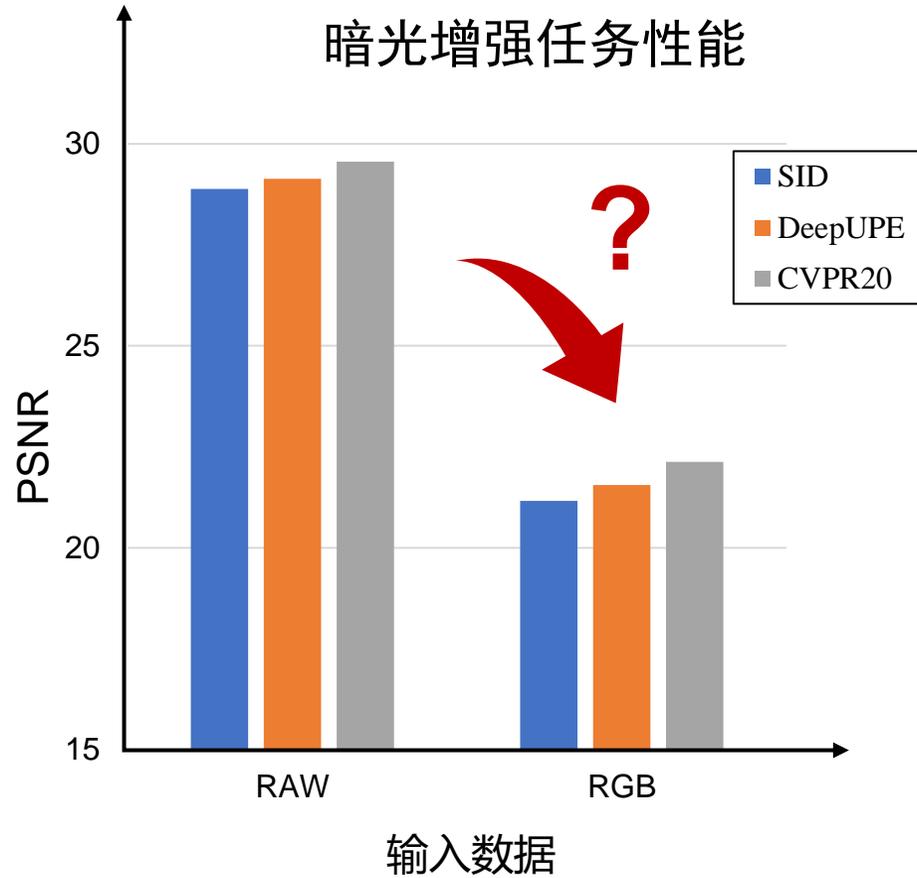
2020 CVPR
Wei et al.



- 数据集构建
- 模型优化
- 噪声建模

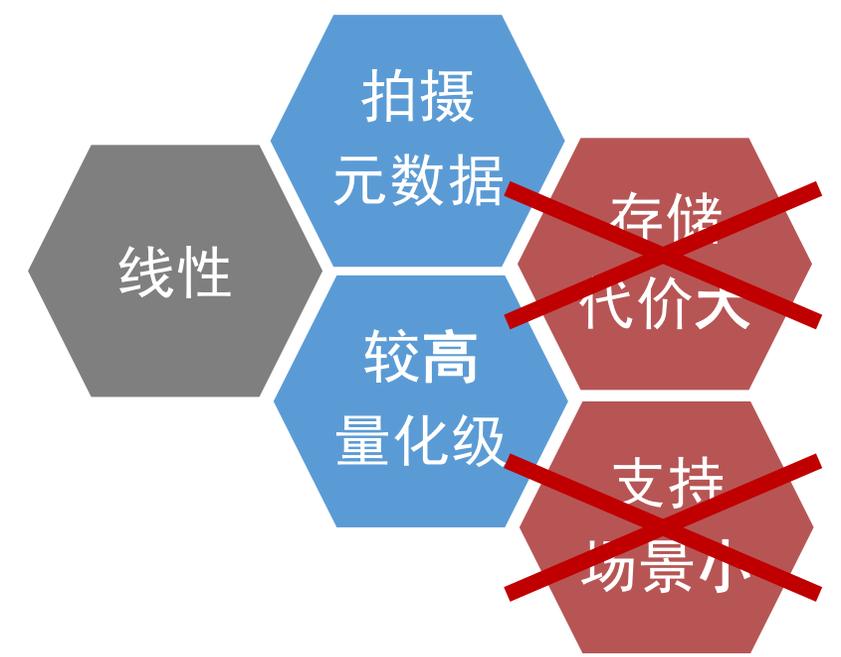
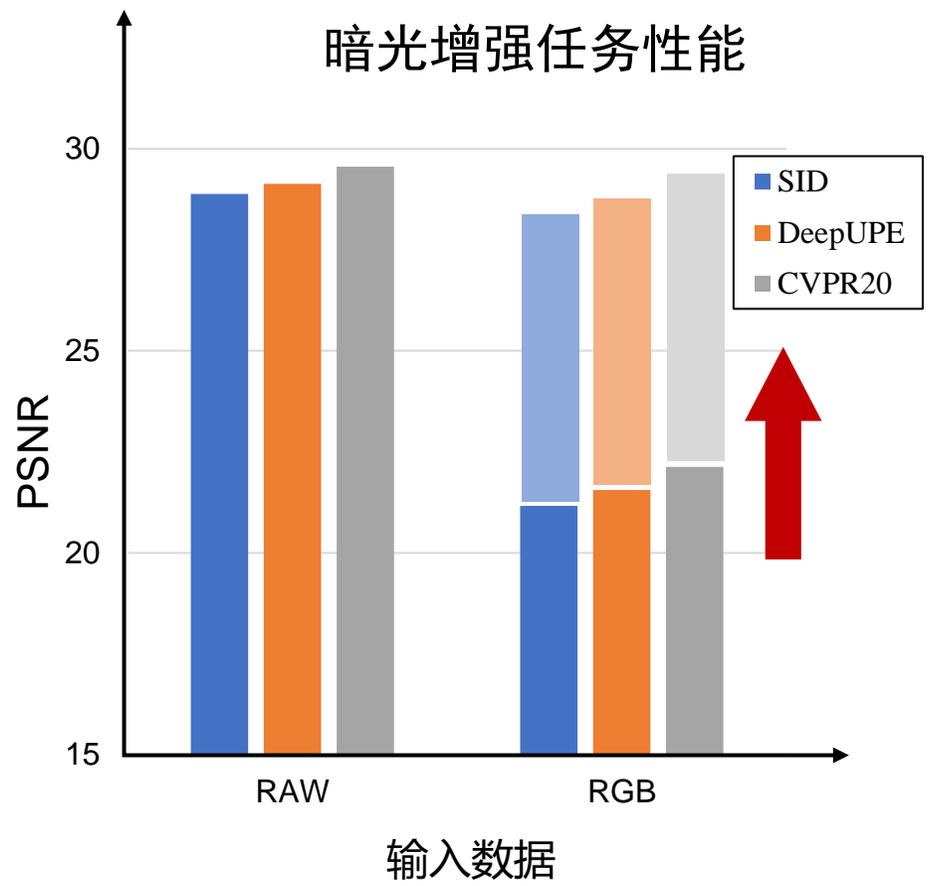


研究问题



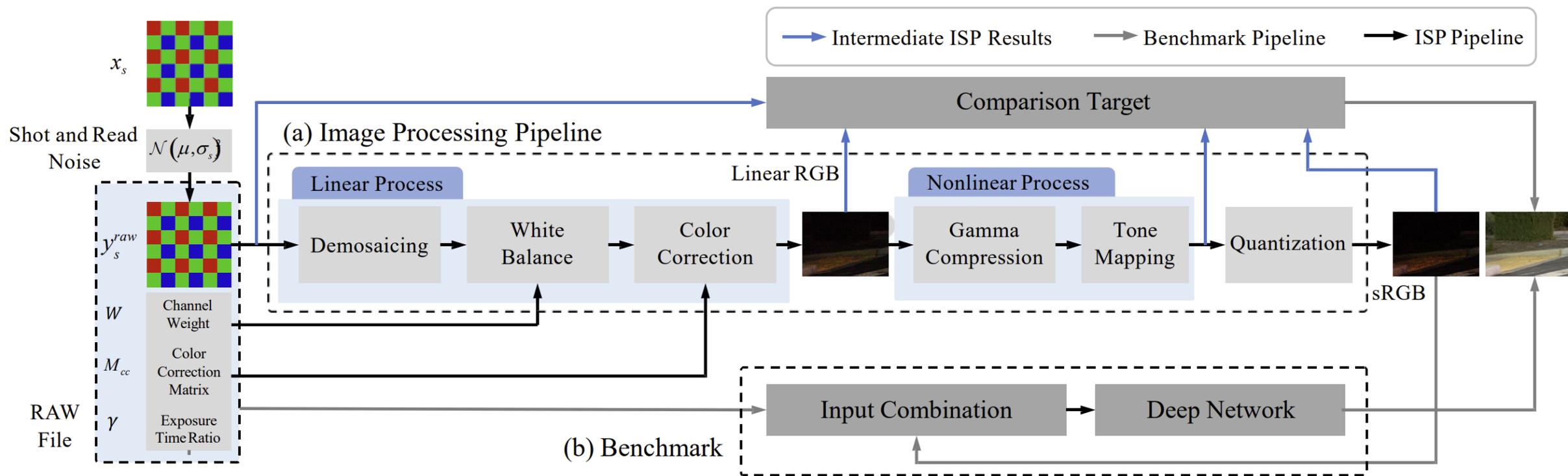
RAW图像特性

研究目标

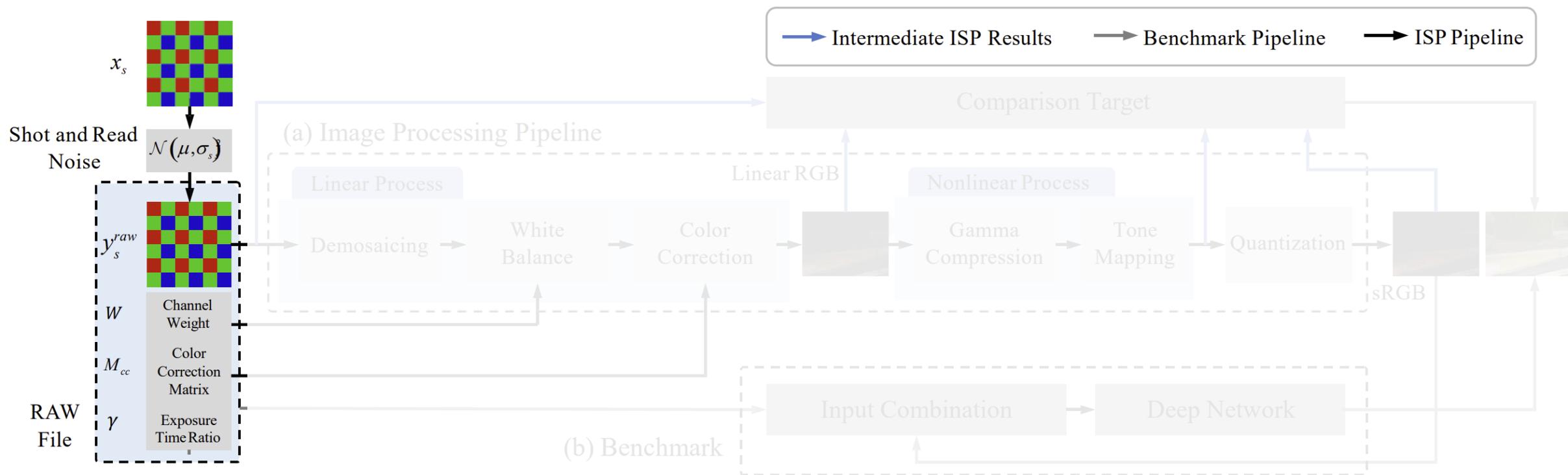


RAW图像特性

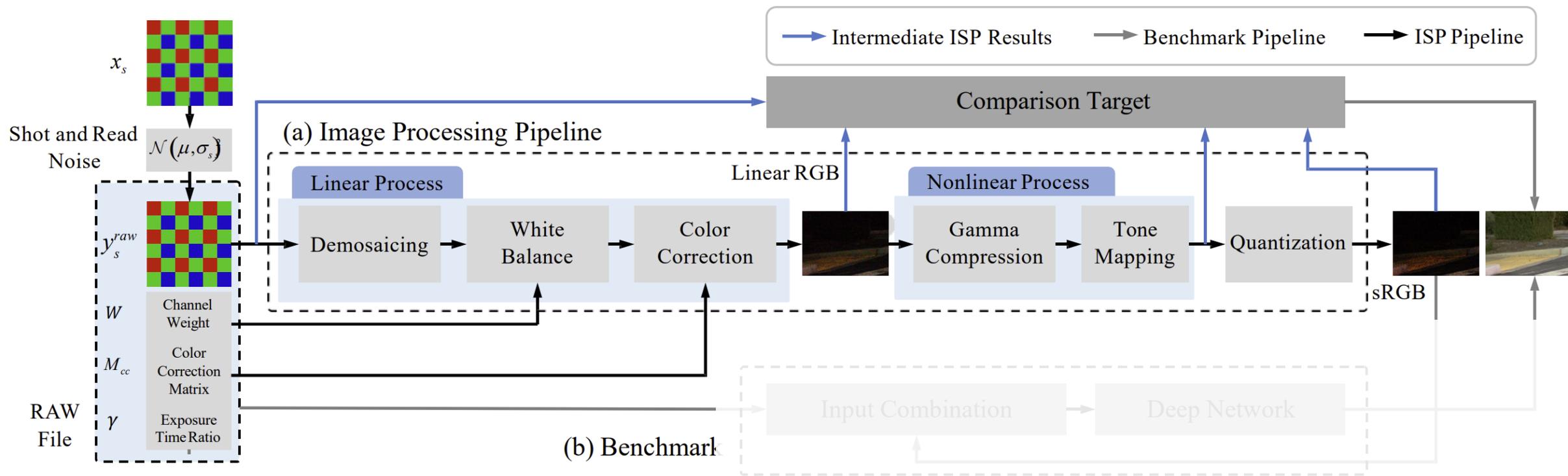
信号特性解耦及量化



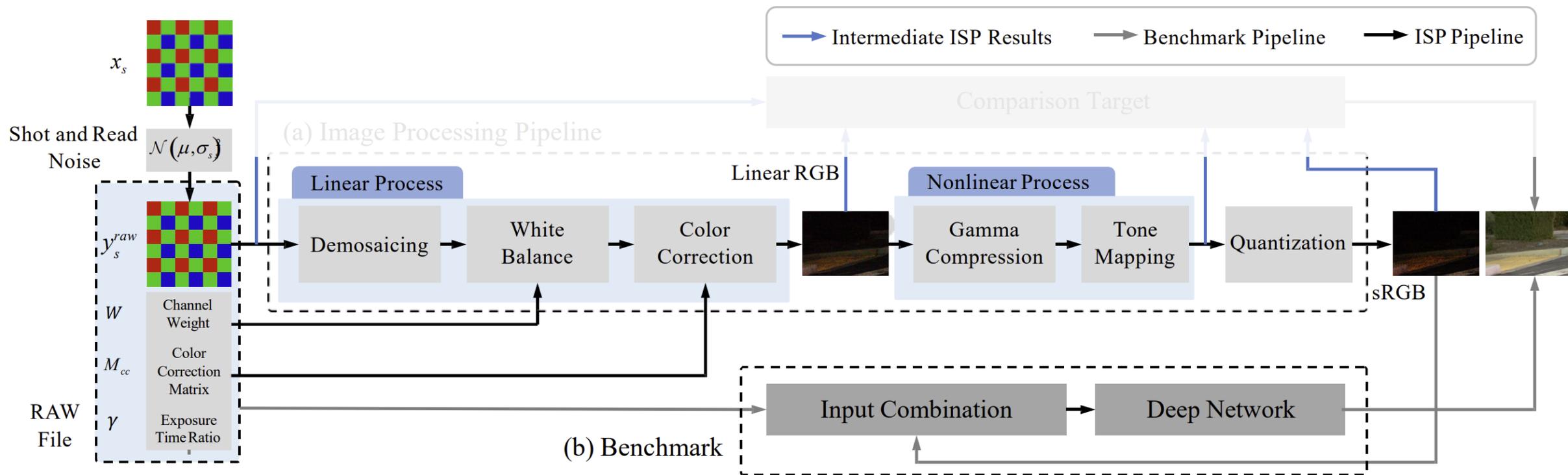
信号特性解耦及量化



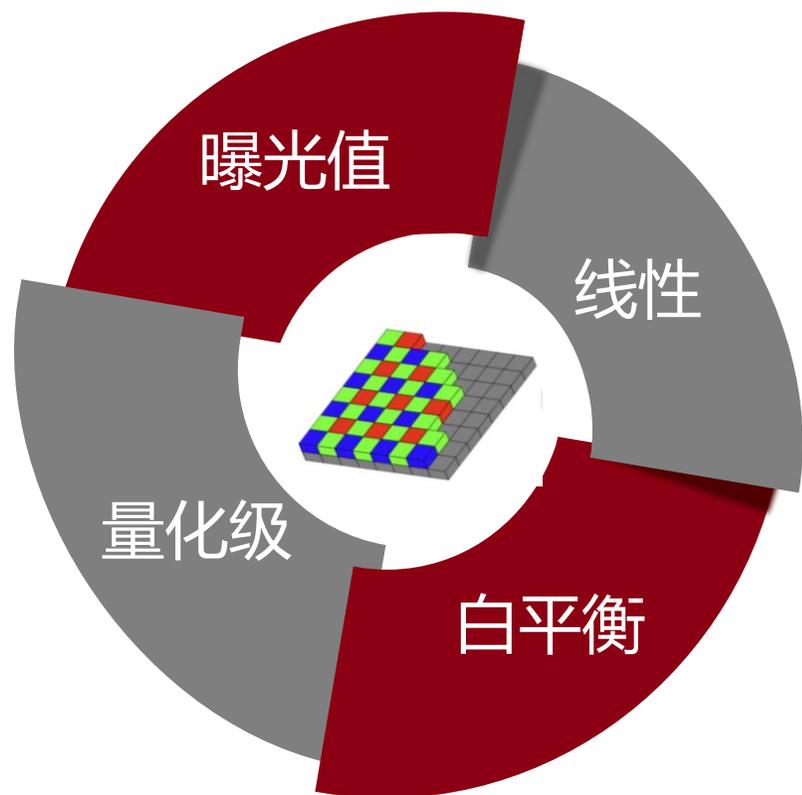
信号特性解耦及量化



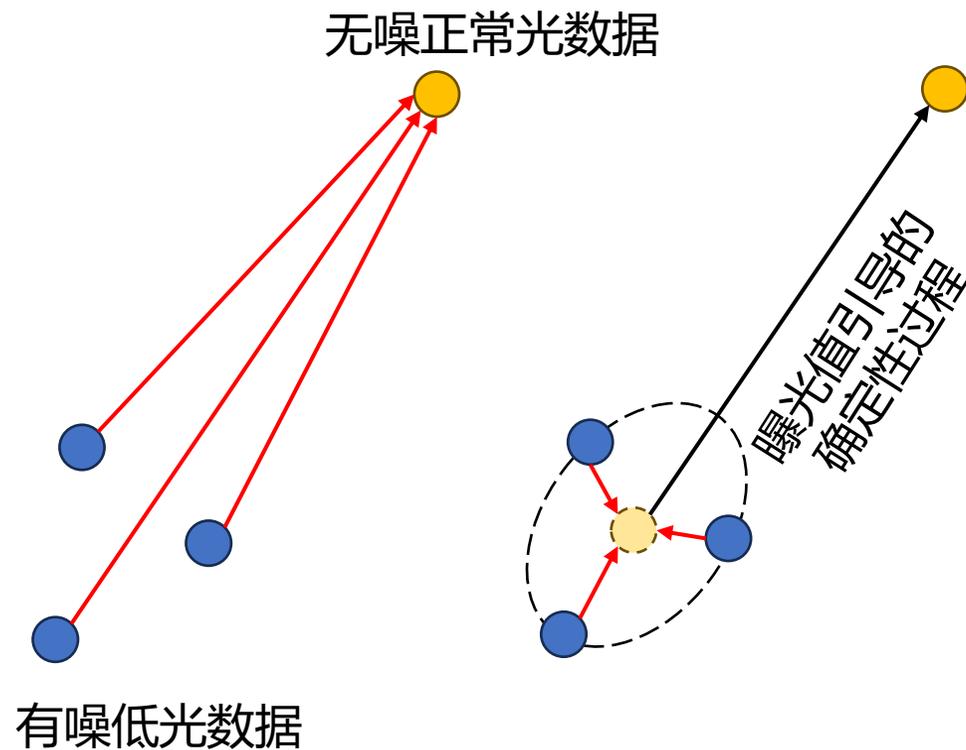
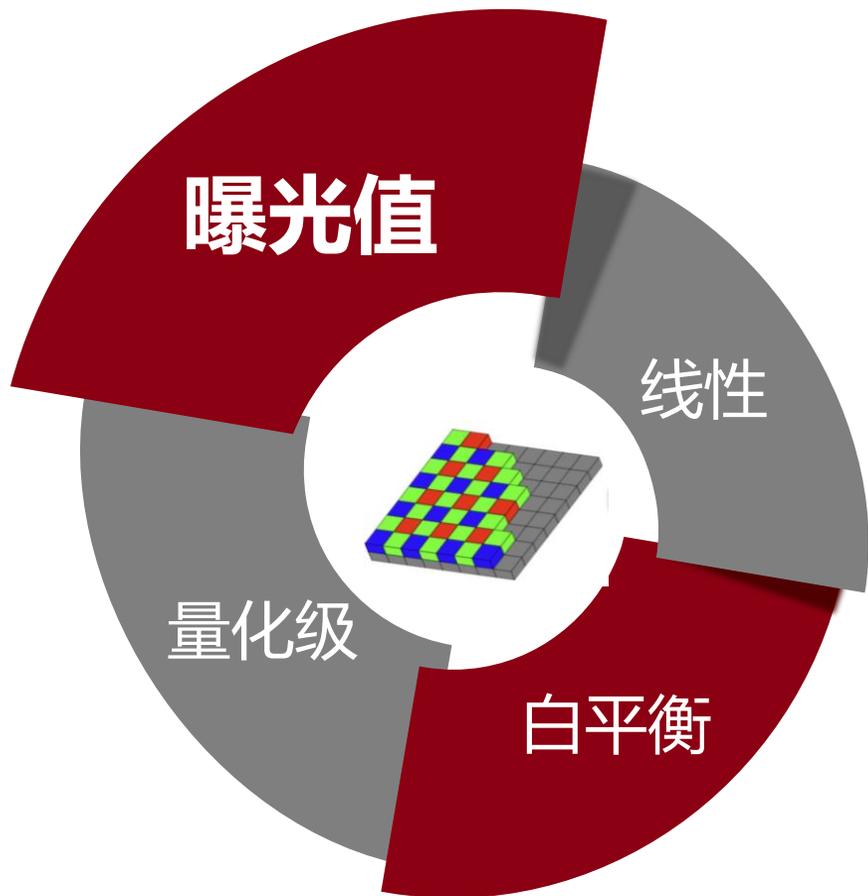
信号特性解耦及量化



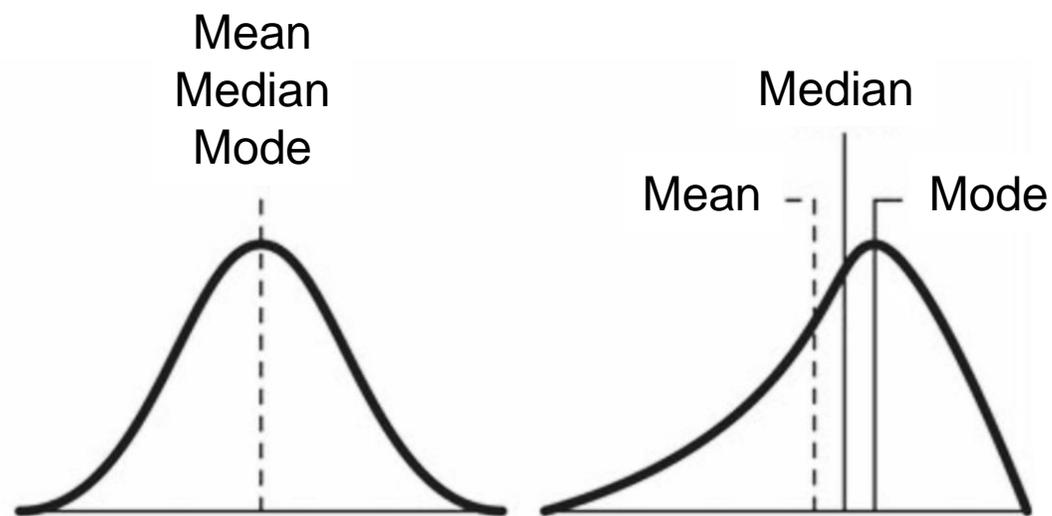
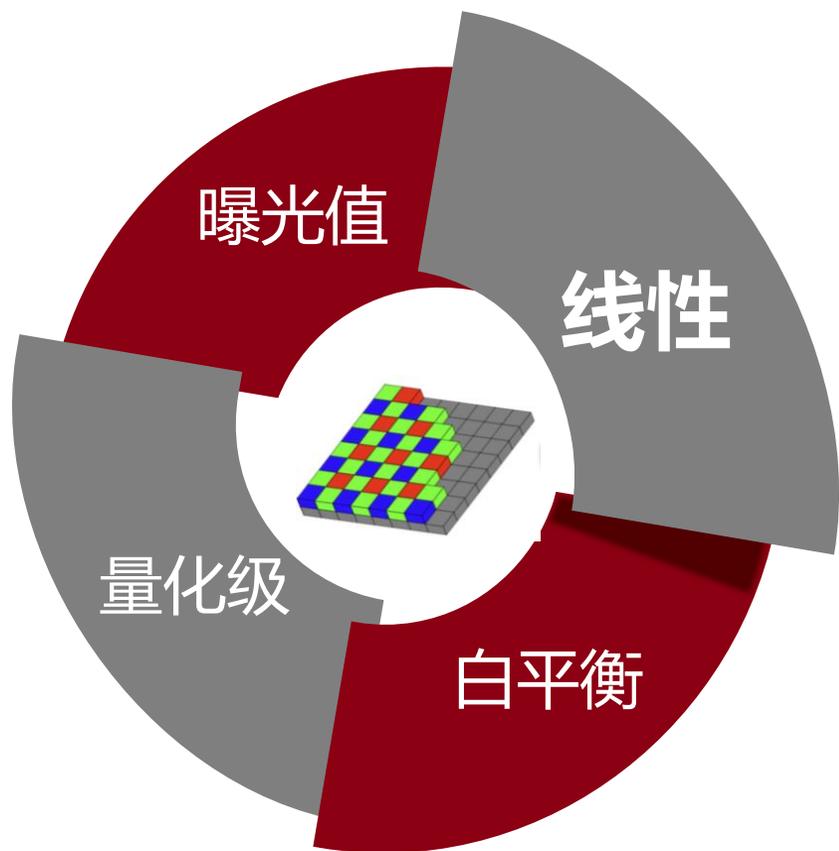
信号特性解耦及量化



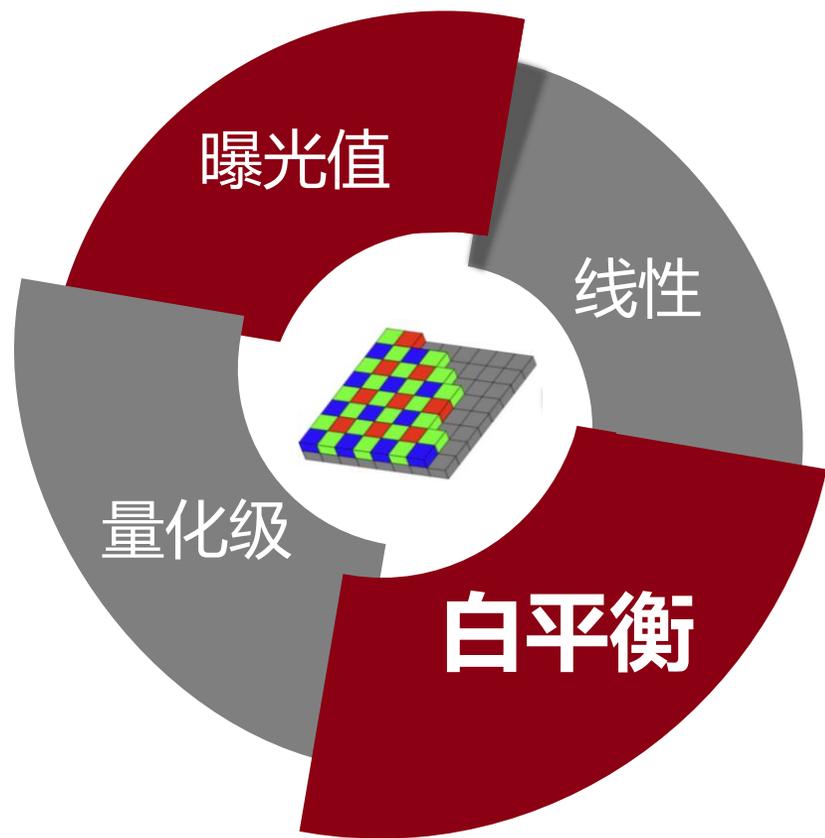
信号特性解耦及量化



信号特性解耦及量化



信号特性解耦及量化

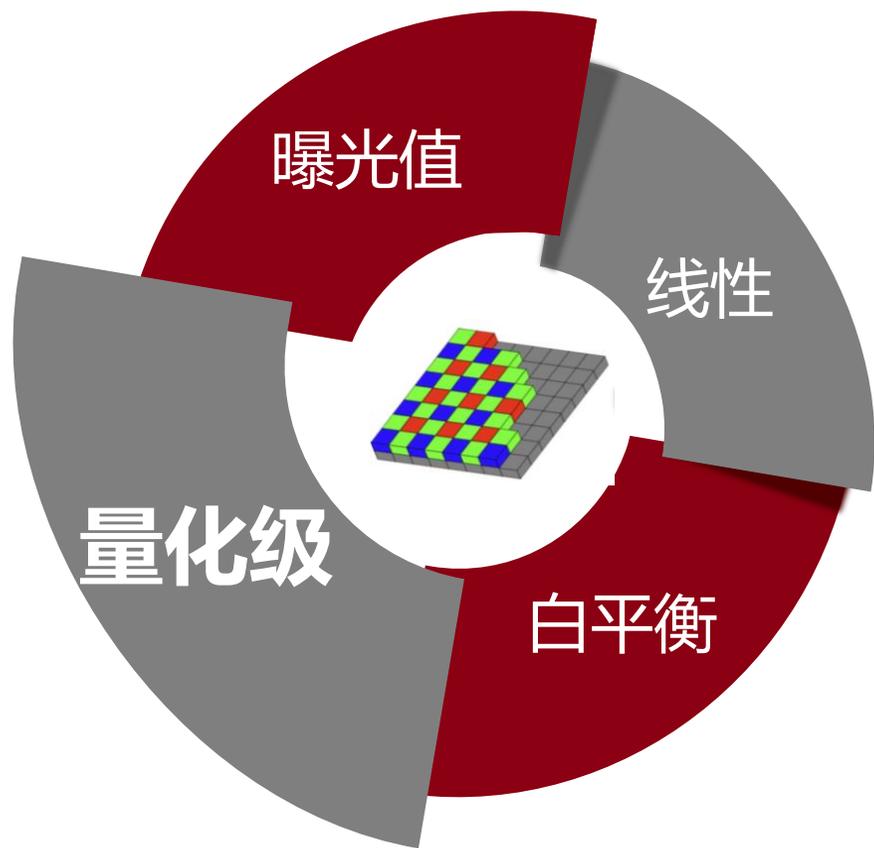


不准确的
白平衡参数

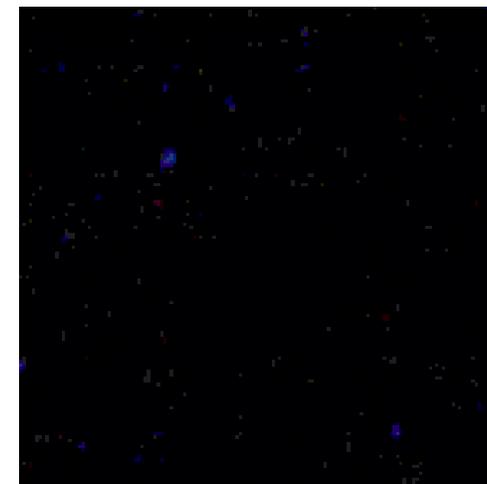


准确的
白平衡参数

信号特性解耦及量化

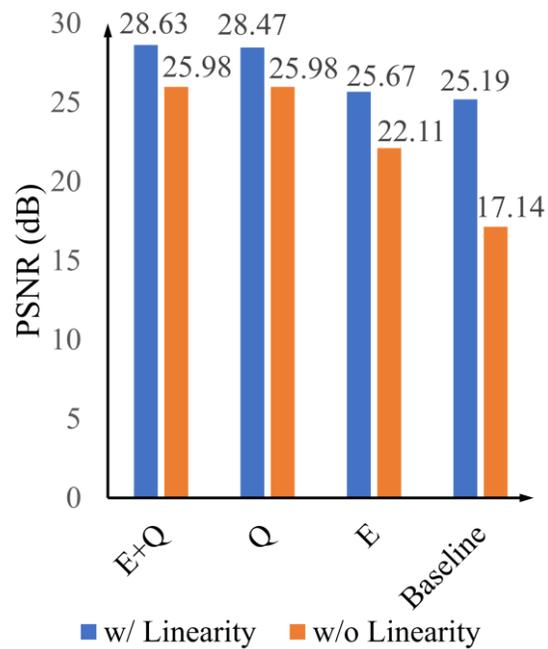


提亮后的16bit图像

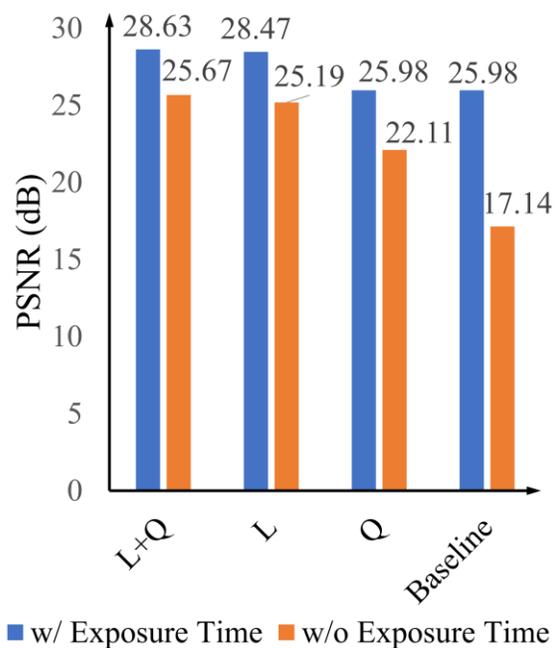


提亮后的8bit图像

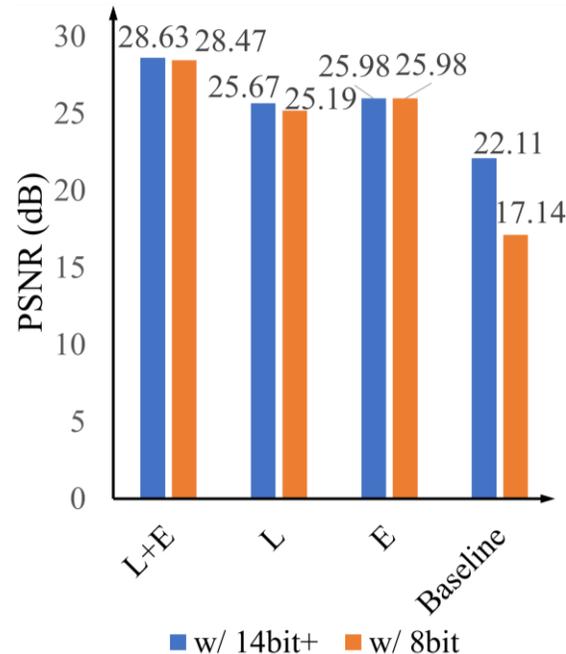
信号特性解耦及量化



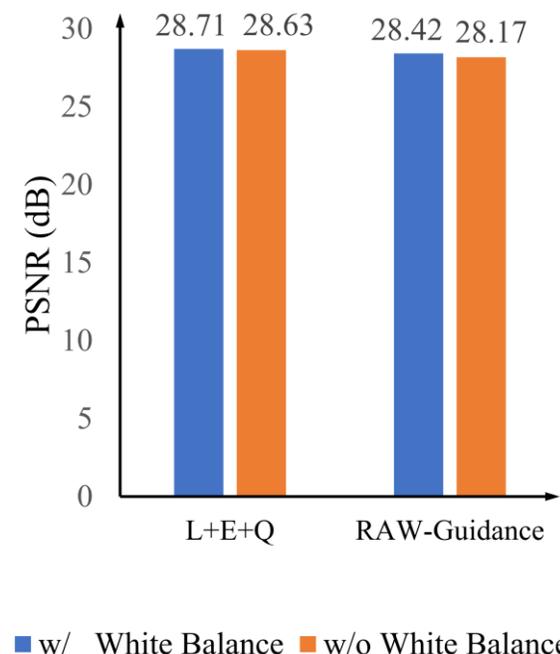
线性



曝光值

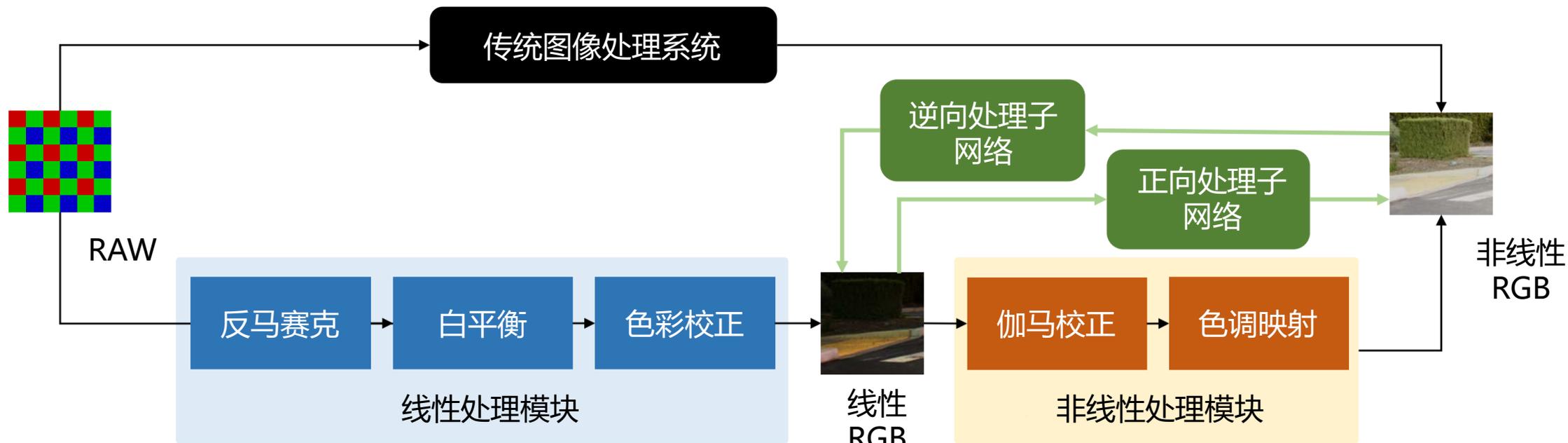


量化级

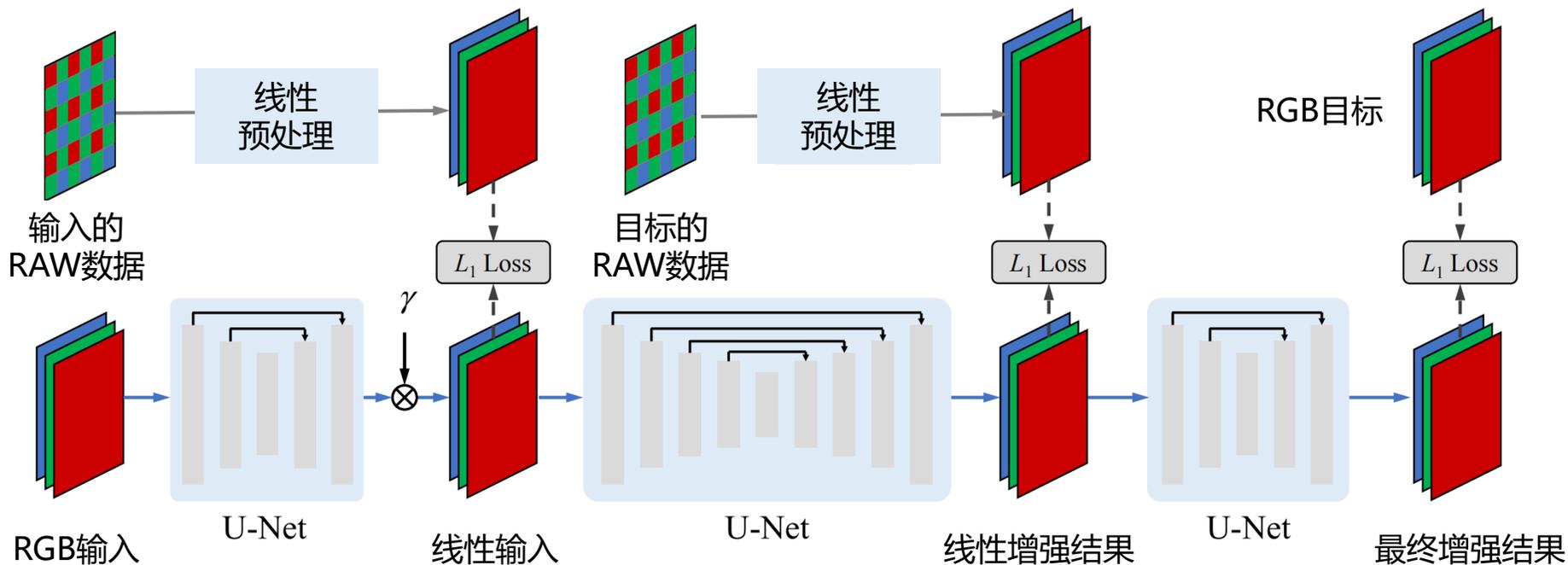


白平衡

RAW领域引导的曝光增强



RAW领域引导的曝光增强



实验结果

方法	PSNR \uparrow	SSIM \uparrow	VIF \uparrow	NIQE \downarrow
HE [1]	5.90	0.028	0.095	28.27
BPDHE [38]	10.67	0.072	0.051	18.64
Dehazing [5]	12.81	0.103	0.077	25.79
MSR [9]	10.04	0.070	0.116	32.24
MF [40]	13.87	0.111	0.108	33.13
LIME [28]	12.59	0.102	0.118	32.33
BIMEF [29]	13.06	0.110	0.086	26.27
LLNet [11]	14.21	0.221	0.047	18.02
SICE [12]	14.26	0.366	0.011	6.89
KinD [13]	13.50	0.109	0.048	17.73
DeepUPE [30]	12.10	0.070	0.028	18.30
本工作方法	28.42	0.880	0.139	15.23

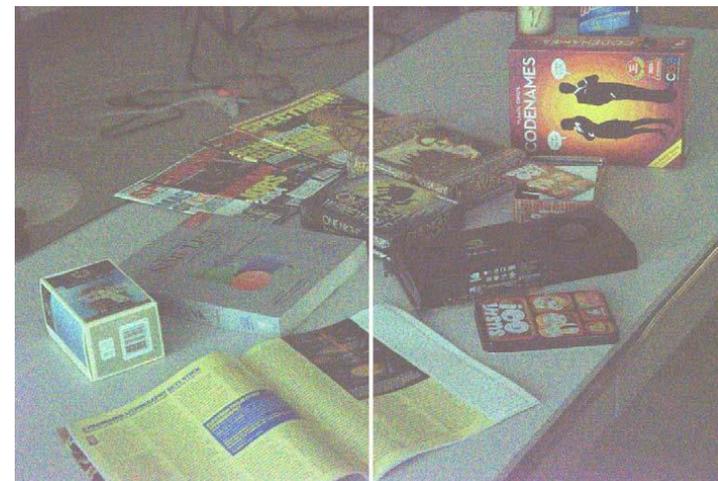
实验结果



输入图像 / 目标图像



本工作



直方图均衡化

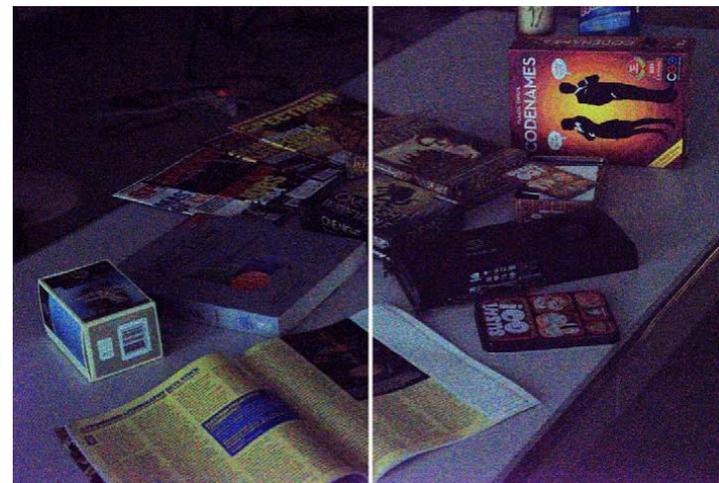
实验结果



输入图像 / 目标图像



本工作



BPDHE

实验结果



输入图像 / 目标图像



本工作



Dehazing

实验结果



输入图像 / 目标图像



本工作



MSR

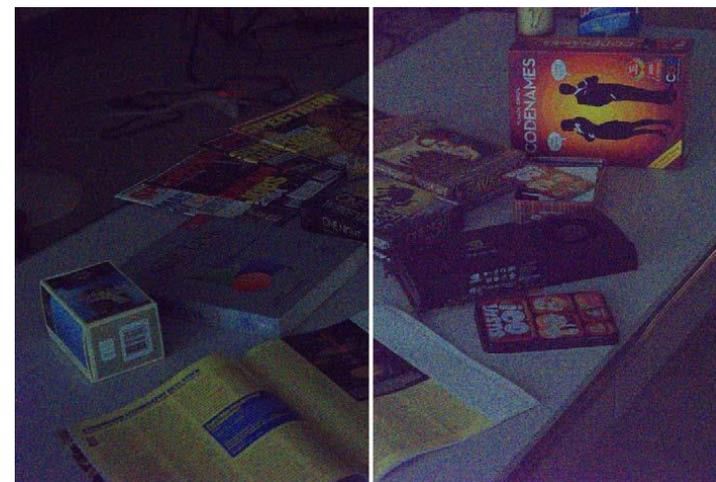
实验结果



输入图像 / 目标图像



本工作



MF

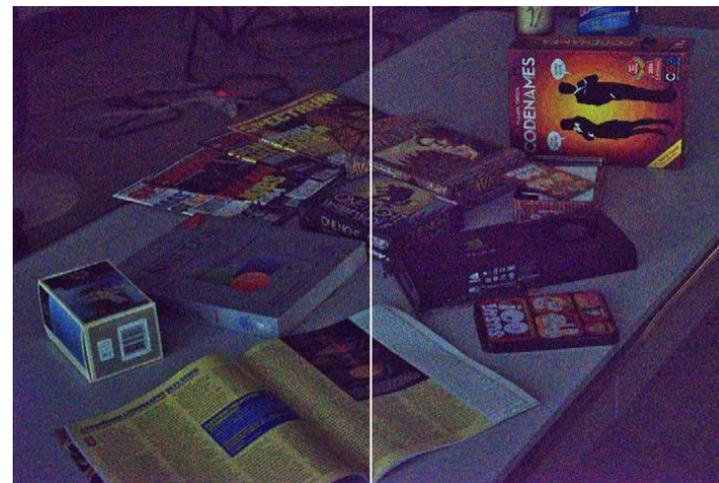
实验结果



输入图像 / 目标图像



本工作



LIME

实验结果



输入图像 / 目标图像



本工作



BIMEF

实验结果



输入图像 / 目标图像



本工作



LLNet

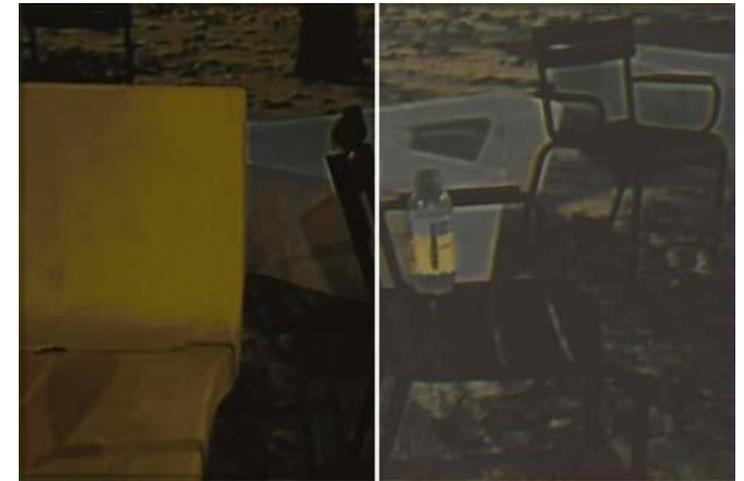
实验结果



输入图像 / 目标图像



本工作



SICE

实验结果



输入图像 / 目标图像



本工作



KinD

实验结果



输入图像 / 目标图像



本工作



DeepUPE

人眼视觉质量提升



面向机器视觉



面向视觉



从低质视觉增强到跨域适配理解

From Low-Quality Visual Enhancement to Cross-Domain Adaptation Understanding

汪文靖 北京大学



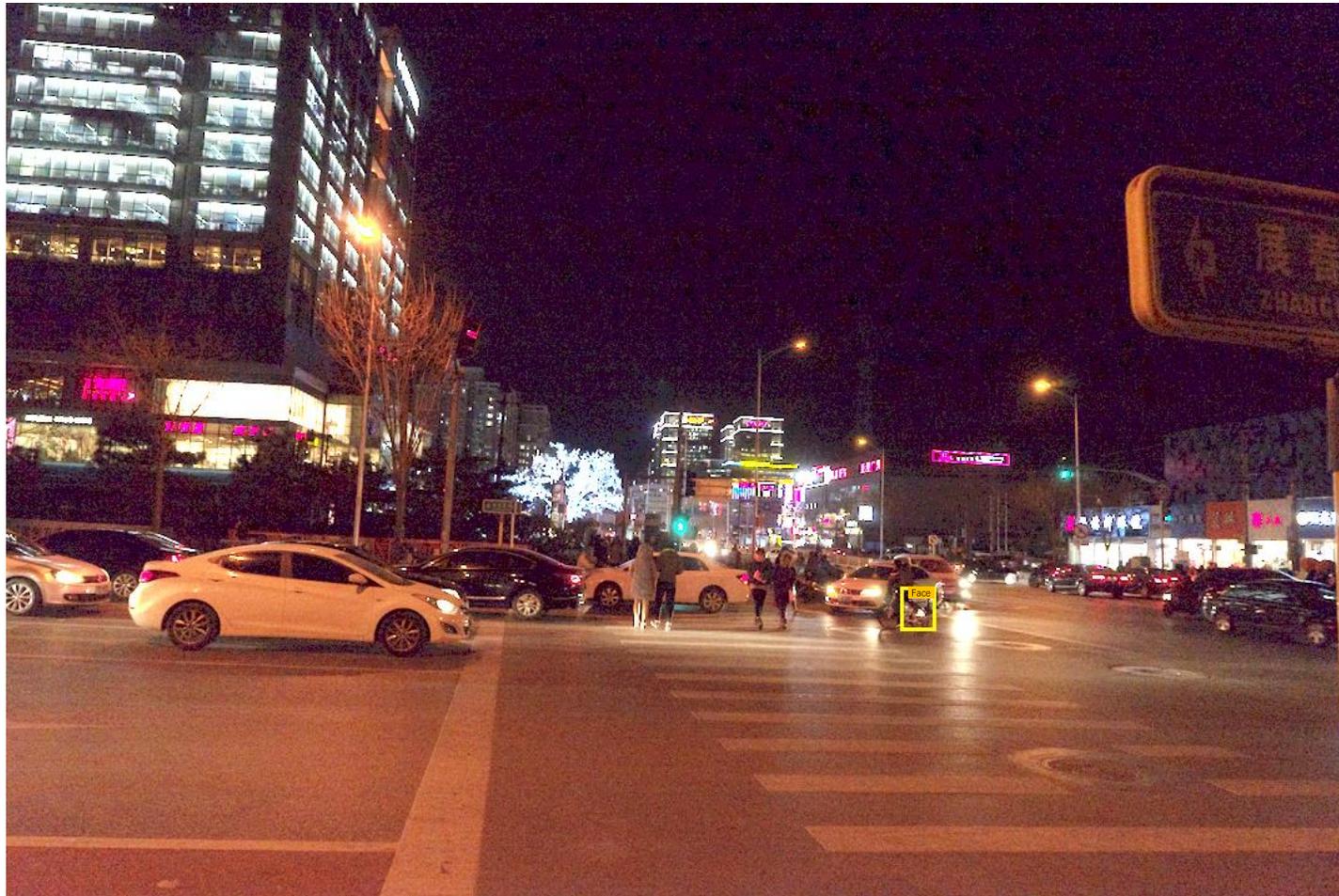
人脸检测



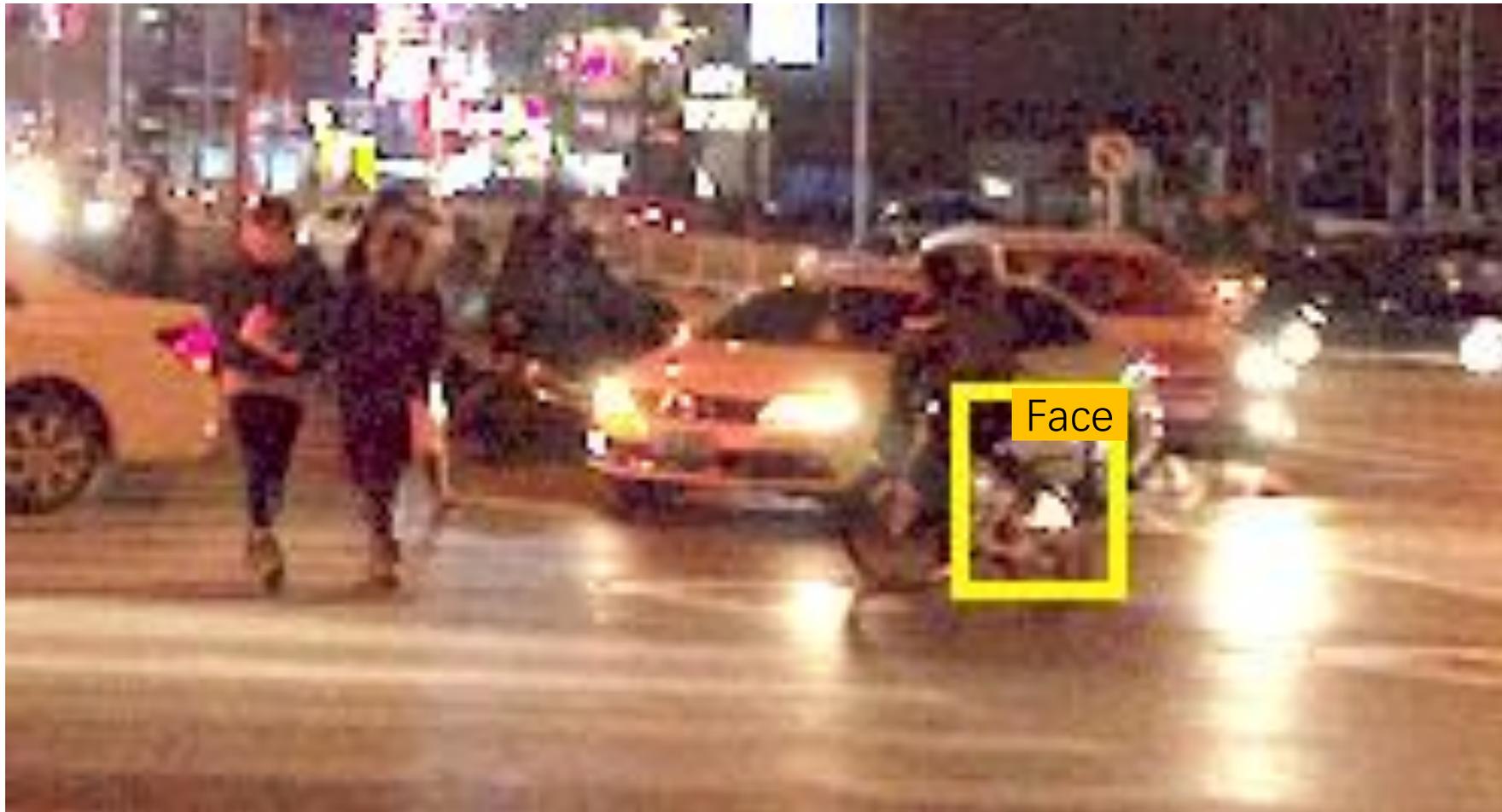
低光照人脸检测



低光照人脸检测



低光照人脸检测



低光照降低下游机器分析任务性能

- 夜间自动驾驶



低光照降低下游机器分析任务性能

- 监控视频分析



低光照降低下游机器分析任务性能

- 图像检索



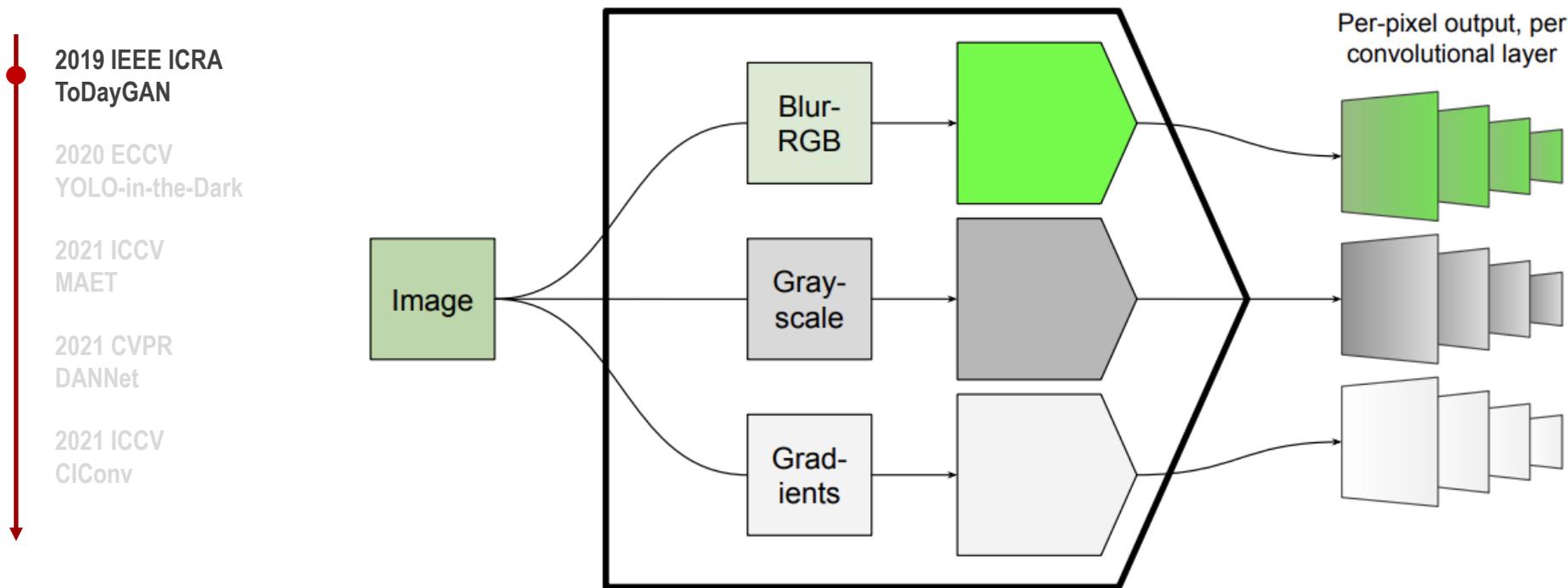
面向低光照的无监督域迁移技术

- ✗ 重新标注一套低光照数据费时费力
- ✓ “正常光照 → 低光照” 的无监督域迁移



相关工作

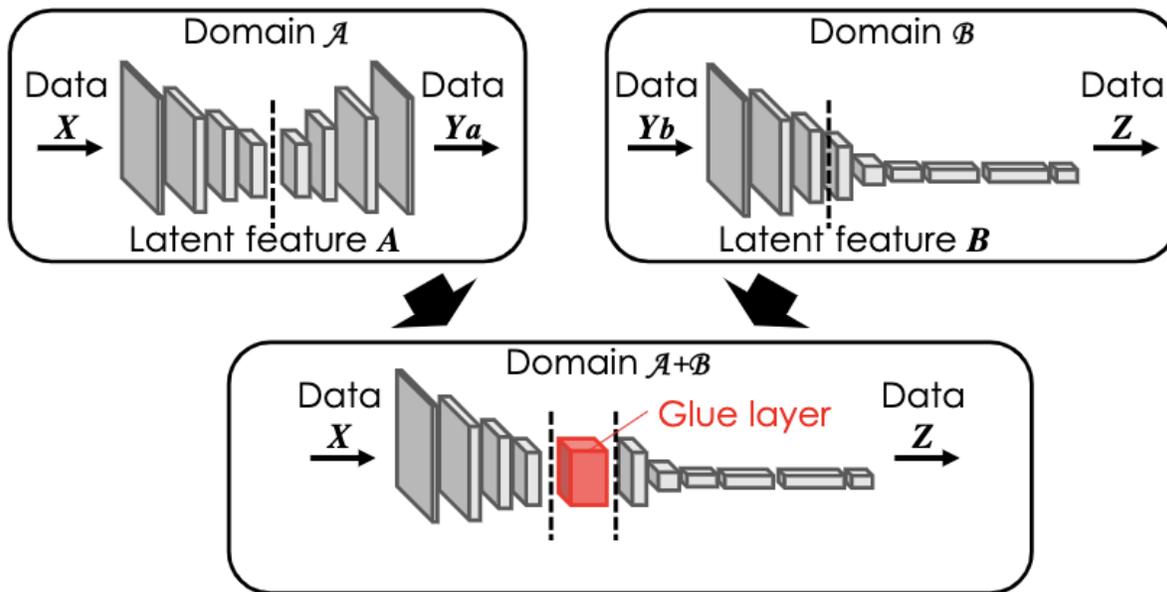
- “夜晚 → 白天” 图像转换



相关工作

- 组合正常光照与低光照域模型

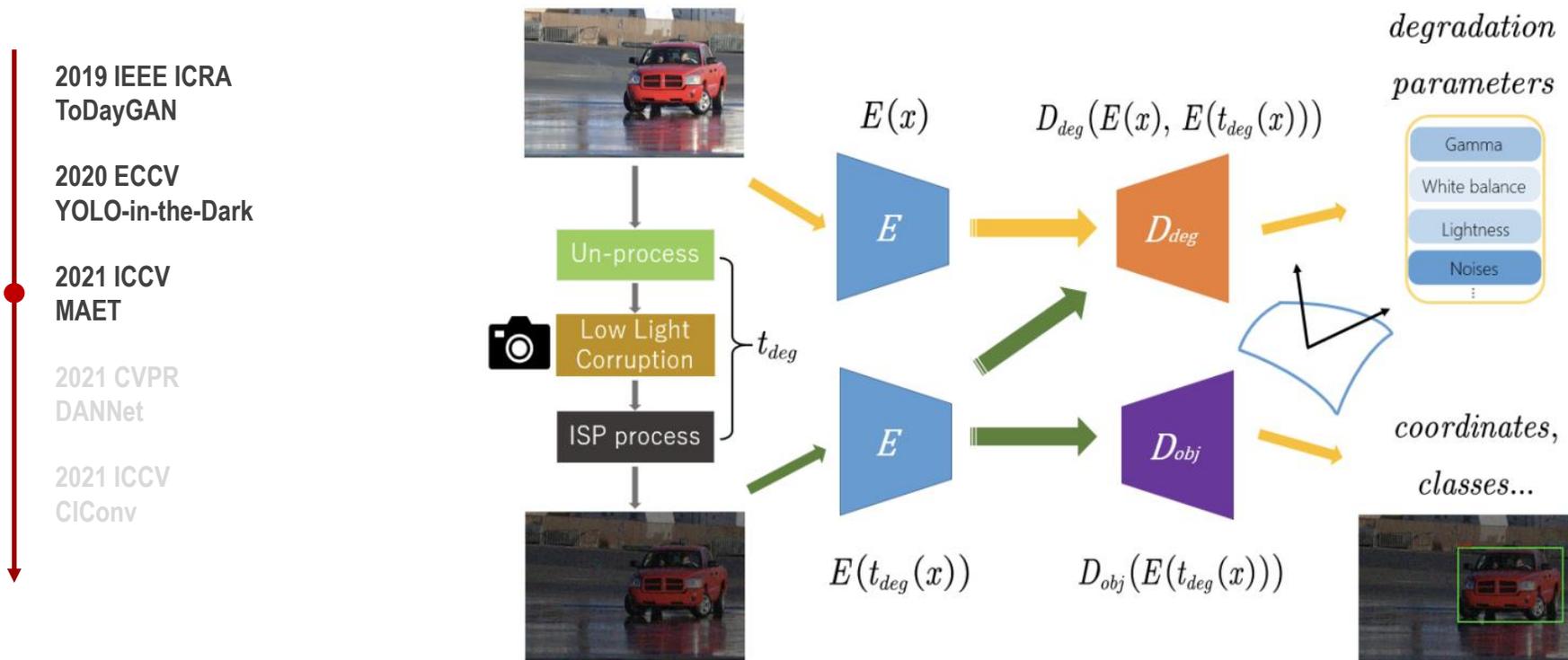
Domain \mathcal{A} : train by RAW image with annotation based on RGB image
 Domain \mathcal{B} : train by RGB image with annotation (object label, bounding boxes)



- 2019 IEEE ICRA
ToDayGAN
- 2020 ECCV
YOLO-in-the-Dark
- 2021 ICCV
MAET
- 2021 CVPR
DANNet
- 2021 ICCV
CIConv

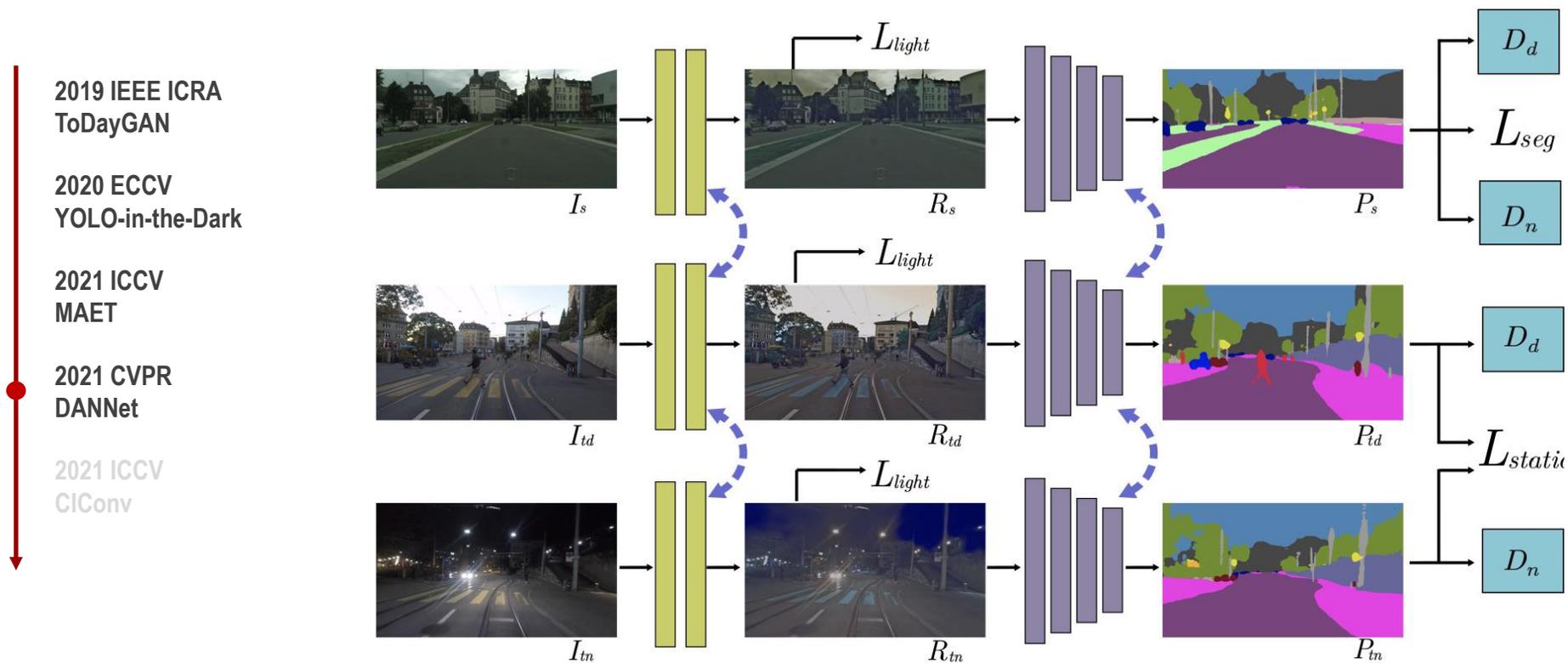
相关工作

- “增强-检测” 多任务自编码变换



相关工作

- “白天→微光→夜晚” 逐步地将模型迁移至目标域



相关工作

- 基于光照不变的边缘提取算子

$$\text{CIConv}(x, y) = \frac{\log(\text{CI}^2(x, y, \sigma = 2^s) + \epsilon) - \mu_S}{\sigma_S}$$

2019 IEEE ICRA
ToDayGAN

2020 ECCV
YOLO-in-the-Dark

2021 ICCV
MAET

2021 CVPR
DANNet

2021 ICCV
CIConv



Input

光照不变信息

问题难点

- 正常光照与低光照的域间隙分为

- 像素层级外观差距



- 物体层级语义差异



问题难点

- 正常光照与低光照的域间隙分为

- 像素层级外观差距



底层视觉域间隙

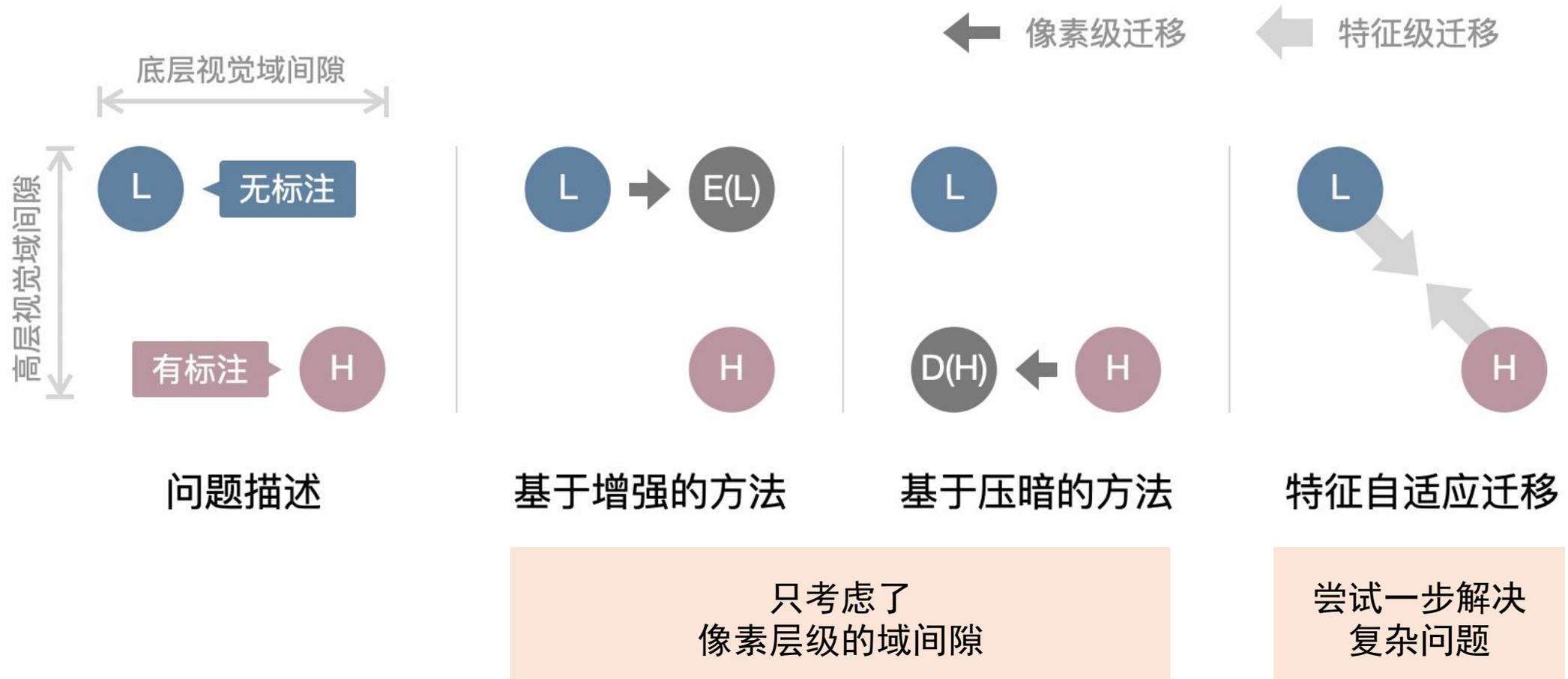
- 物体层级语义差异



高层视觉域间隙

相关工作

- 总结与建模



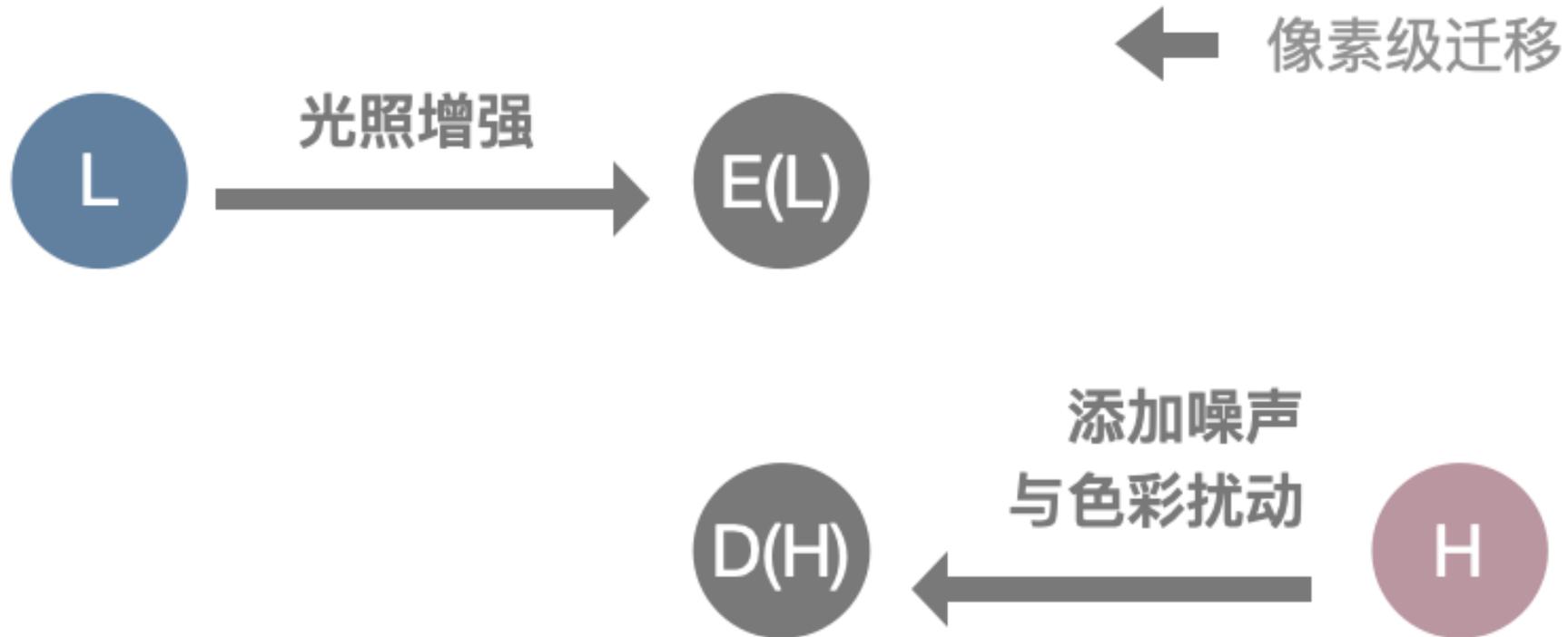
高层-底层视觉联合迁移框架

- 方法范式



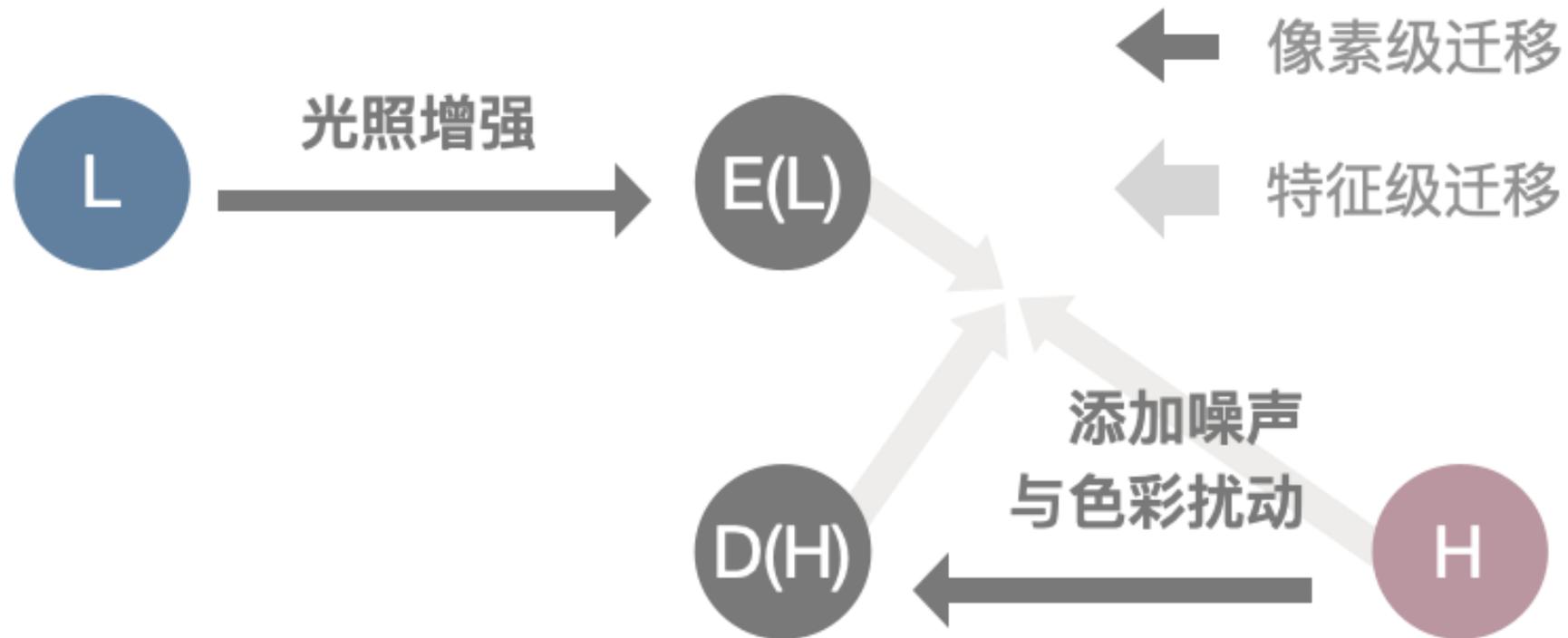
高层-底层视觉联合迁移框架

- 方法范式



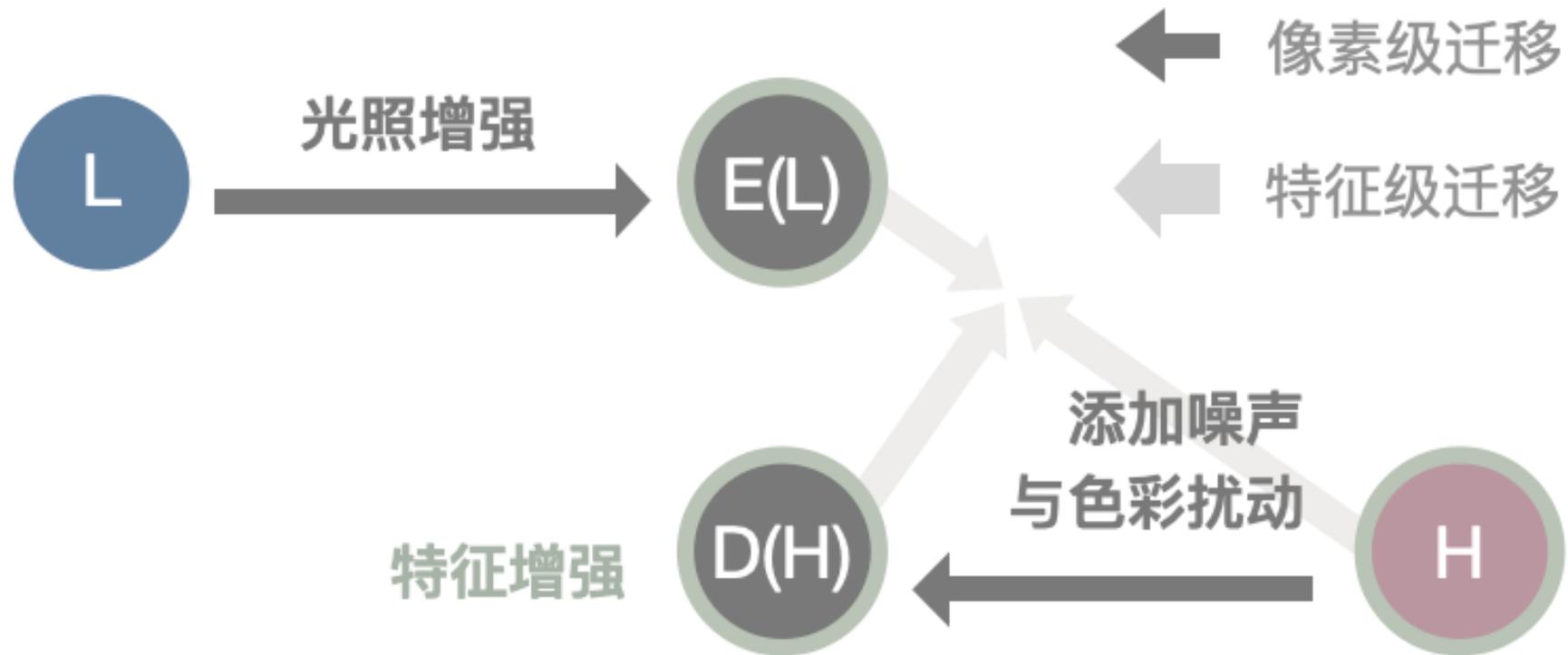
高层-底层视觉联合迁移框架

- 方法范式



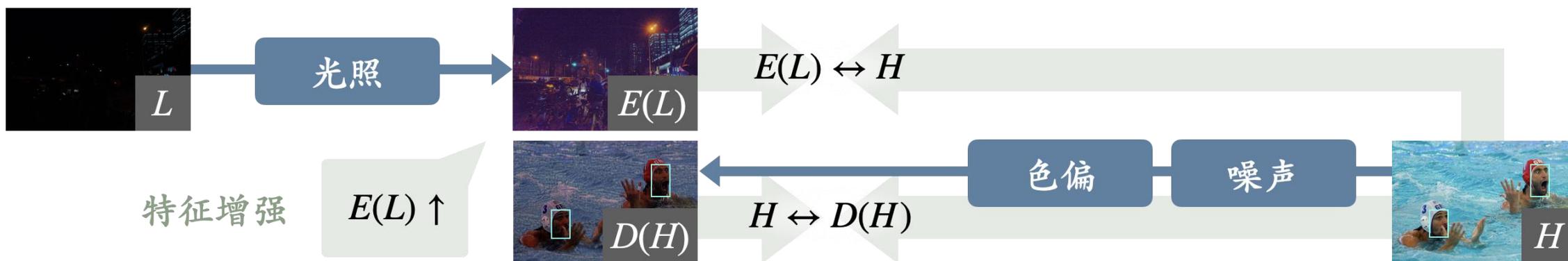
高层-底层视觉联合迁移框架

- 方法范式



高层-底层视觉联合迁移框架

- 总迁移框架



高层-底层视觉联合迁移框架

- 面向高层视觉的提亮曲线族

条件

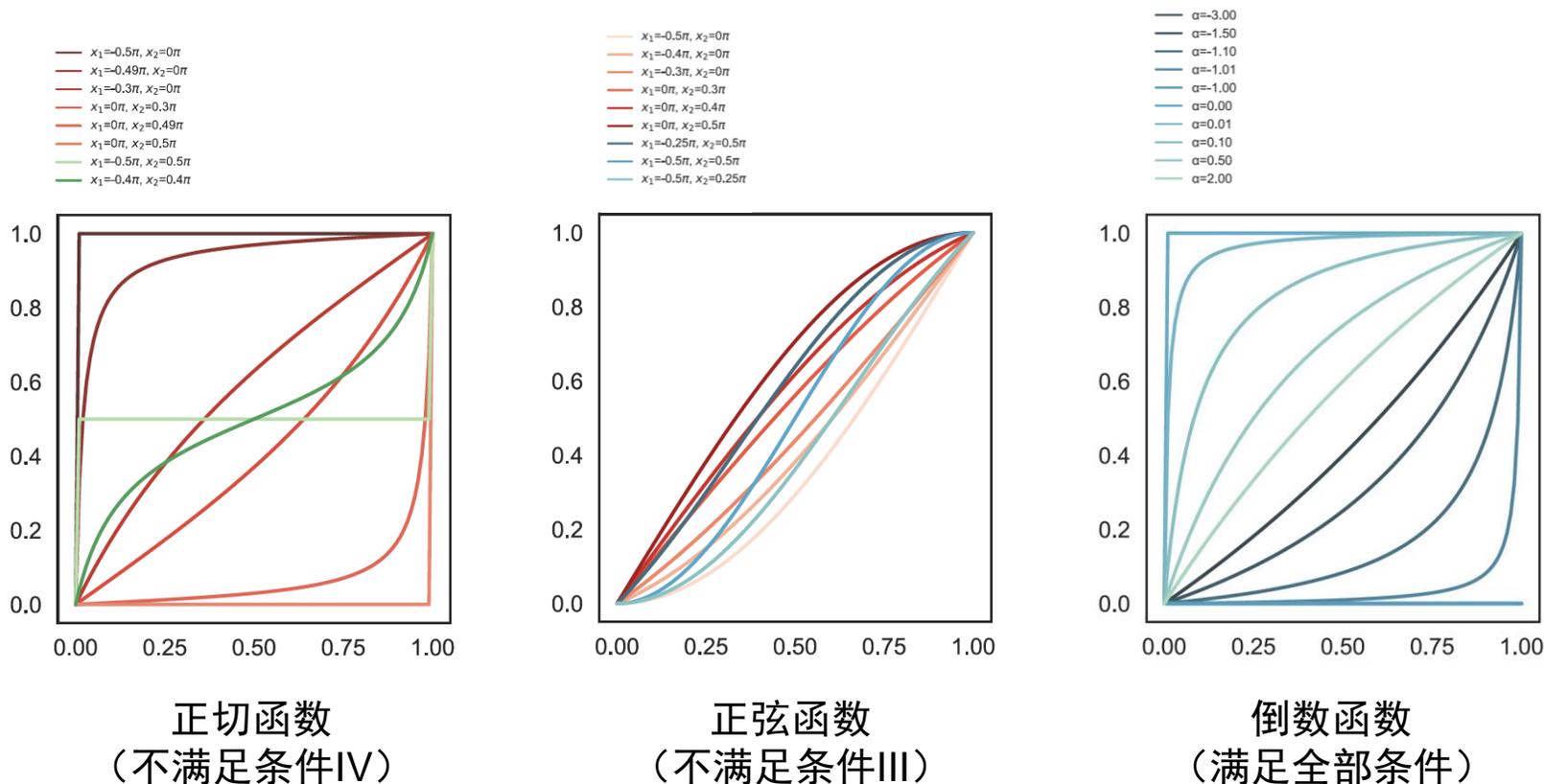
- I. 经过 $(0,0)$ 和 $(1,1)$: $f(0, \alpha) = 0, f(1, \alpha) = 1$
- II. 单调可微: $f(\cdot, \alpha): [0,1] \rightarrow [0,1]$ 单调可微
- III. 保证表示范围: $\forall x_0 \in [0,1]$ 和 $\forall y_0 \in [0,1], \exists \alpha_0$ 使 $f(x_0, \alpha_0) = y_0$
- IV. 防止退化为水平线: $\forall \alpha, \forall x \in (0,1), \partial f(x, \alpha) / \partial x \neq 0$



在初等函数中
求解满足**条件**的
曲线

高层-底层视觉联合迁移框架

- 面向高层视觉的提亮曲线族



高层-底层视觉联合迁移框架



输入图像



正切函数 (不满足条件IV)



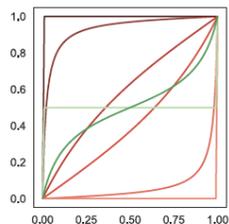
正弦函数 (不满足条件III)



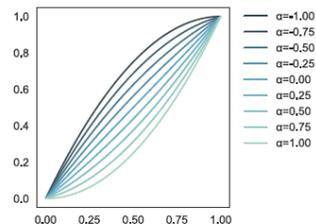
倒数函数 (满足全部条件)

高层-底层视觉联合迁移框架

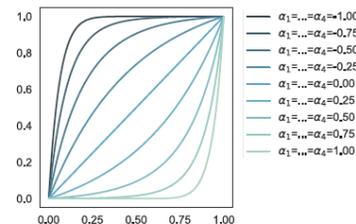
- 面向高层视觉的提亮曲线族



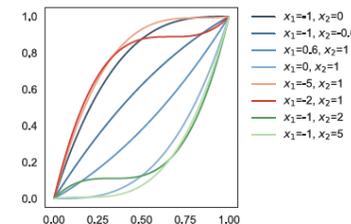
(a) Tangent



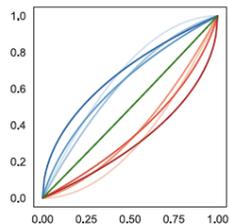
(b) Quadratic



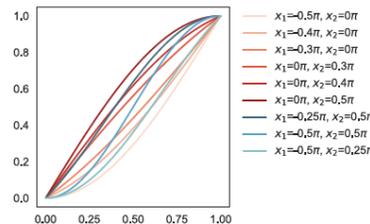
(c) Iterative Quadratic (Zero-DCE n=4)



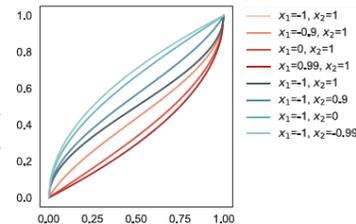
(d) Cubic



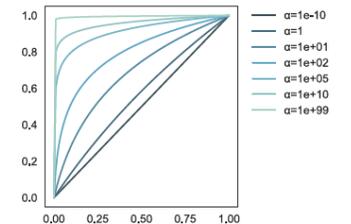
(e) Circle



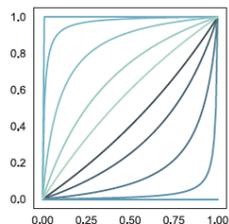
(f) Sine



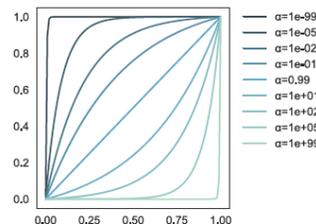
(g) Arcsine



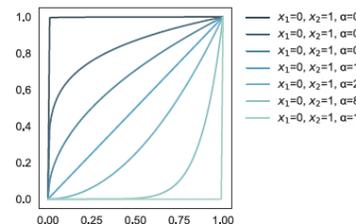
(h) Logarithmic



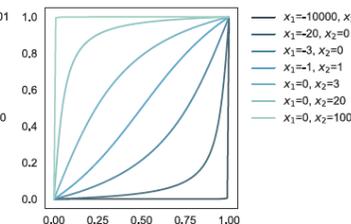
(i) Reciprocal



(j) Exponential



(k) Power



(l) Arctangent

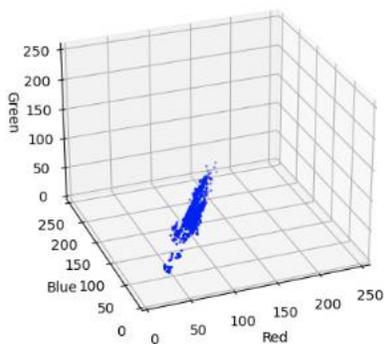
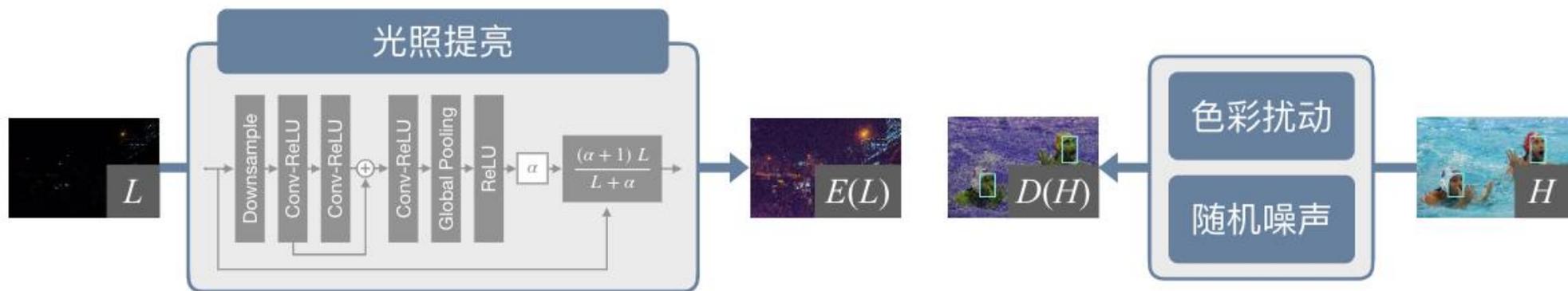
高层-底层视觉联合迁移框架

- 面向高层视觉的提亮曲线族

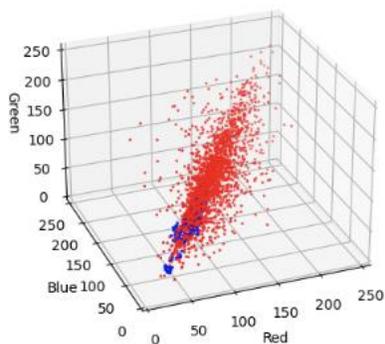
函数名	解析式	数值	条件 III	条件 IV
对数	$\frac{\log(\alpha x + 1)}{\log(\alpha + 1)}$	$\alpha > 0$	1/2	✓
倒数	$\frac{(\alpha + 1)x}{x + \alpha}$	$\alpha \in (-\infty, -1) \cup (0, +\infty)$	✓	✓
指数	$\frac{\alpha^x - 1}{\alpha - 1}$	$\alpha \in [0, 1) \cup (1, +\infty)$	✓	✓
幂	$AT(x^\alpha, x_1, x_2)$	$0 < x_1 < x_2, \alpha > 0$	✓	✓
反正切/反余切	$AT(\arctan(x), x_1, x_2)$	$x_1 < x_2$	✓	✓

高层-底层视觉联合迁移框架

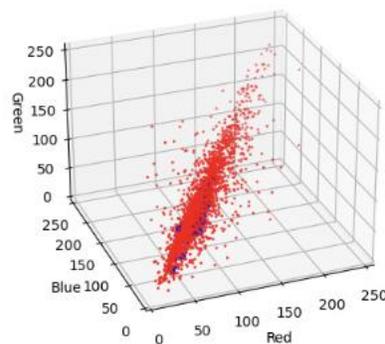
- 双向底层视觉迁移



$E(L)$



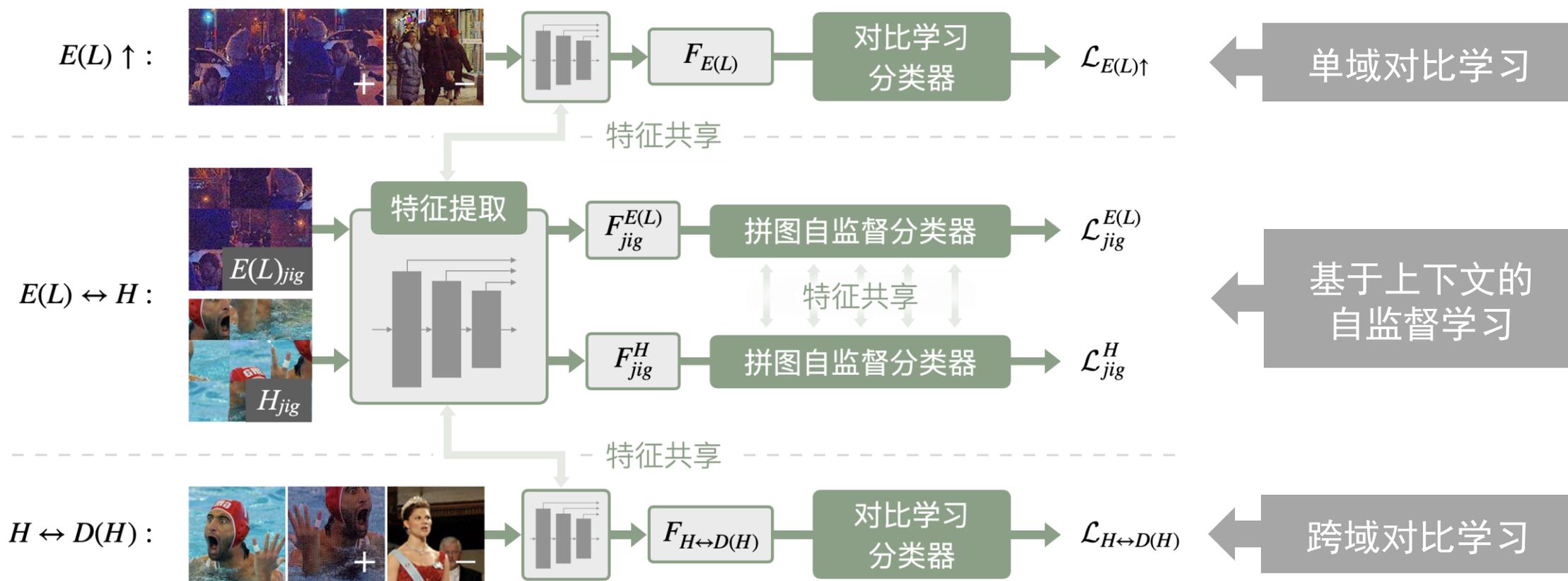
H



$D(H)$

高层-底层视觉联合迁移框架

- 多任务高层视觉迁移



实验结果

• 低光照人脸检测性能通测

方法类别	方法	mAP (%)
人脸检测	Faster-RCNN	1.7
	SSH	6.9
	RetinaFace	8.6
	SRN	9.0
	SFA	9.3
	PyramidBox	12.5
	Small Hard Face	16.1
	DSFD	16.1
基于压暗 骨架: DSFD	MUNIT	29.7
	CycleGAN	31.9
	CUT	32.7

方法类别	方法	mAP (%)
基于提亮 骨架: DSFD	SICE	4.7
	RetinexNet	12.0
	KinD	15.8
	EnlightenGAN	31.3
	LIME	40.7
	Zero-DCE++	40.9
	Zero-DCE	41.3
	MF	41.4
特征迁移 骨架: DSFD	OSHOT	25.4
	Progressive DA	28.5
	Bidirectional DA	33.7
	Pseudo Labeling	35.1
	本方法	45.9

实验结果



(a) DSFD

(b) SICE

(c) RetinexNet

(d) KinD

(e) LIME

(f) MF



(g) EnlightenGAN 重训

(h) EnlightenGAN

(i) Zero-DCE 重训

(j) Zero-DCE

(k) 本方法

(l) 真值

实验结果



(a) DSFD

(b) SICE

(c) RetinexNet

(d) KinD

(e) LIME

(f) MF



(g) EnlightenGAN 重训

(h) EnlightenGAN

(i) Zero-DCE 重训

(j) Zero-DCE

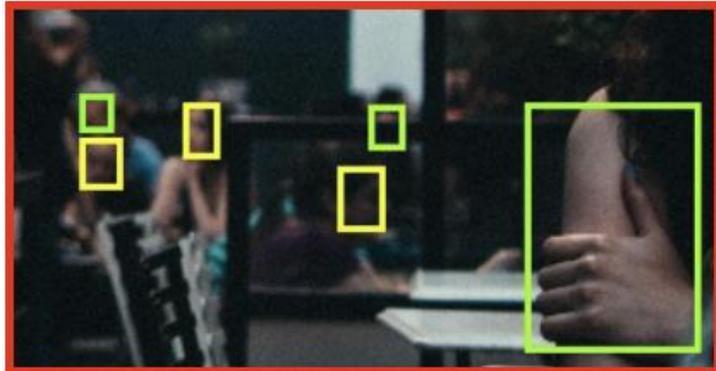
(k) 本方法

(l) 真值

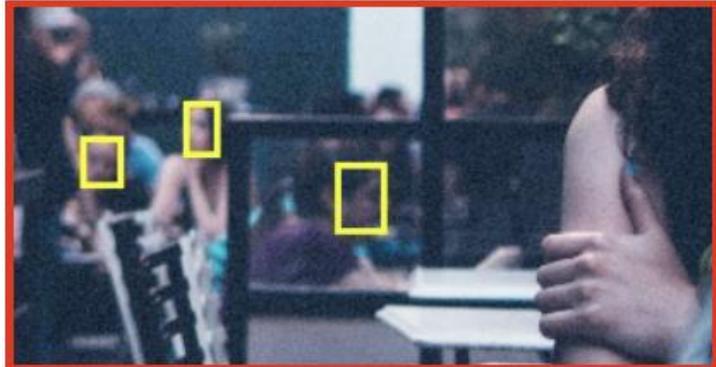
实验结果

- 非训练数据集上的困难样本

基线方法



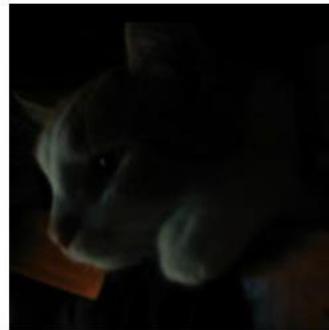
本方法



实验结果

- 分类任务, Top-1 准确率 46.9% → 60.7%

基线方法



Dog, 82.6%



Cup, 100.0%



Motorbike, 53.0%



Car, 60.0%

本方法



Cat, 51.6%



Boat, 90.0%



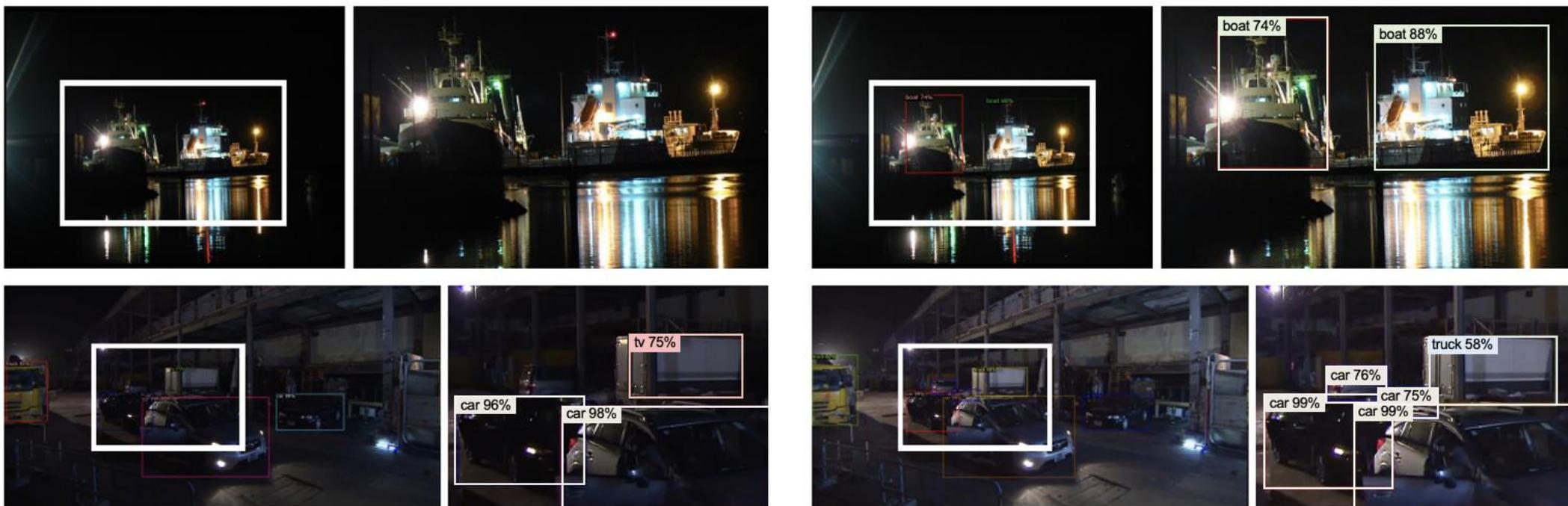
Motorbike, 99.7%



Bicycle, 99.7%

实验结果

- 通用物体检测任务, mAP 29.3% → 30.4%



基线方法

本方法

实验结果

- 街景分割任务, mIoU 17.1% \rightarrow 23.1%



基线方法

本方法

从低质视觉增强到跨域适配理解



从低质视觉增强到跨域适配理解



从低质视觉增强到跨域适配理解



更细致地分析人的行为



走路 × 跑步 × 跳舞 ✓



行为分析



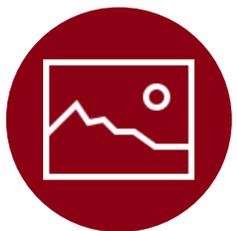
噪声鲁棒的视频序列行为分析

Noise-Robust Video Sequence Behavior Analysis

林里浪 北京大学



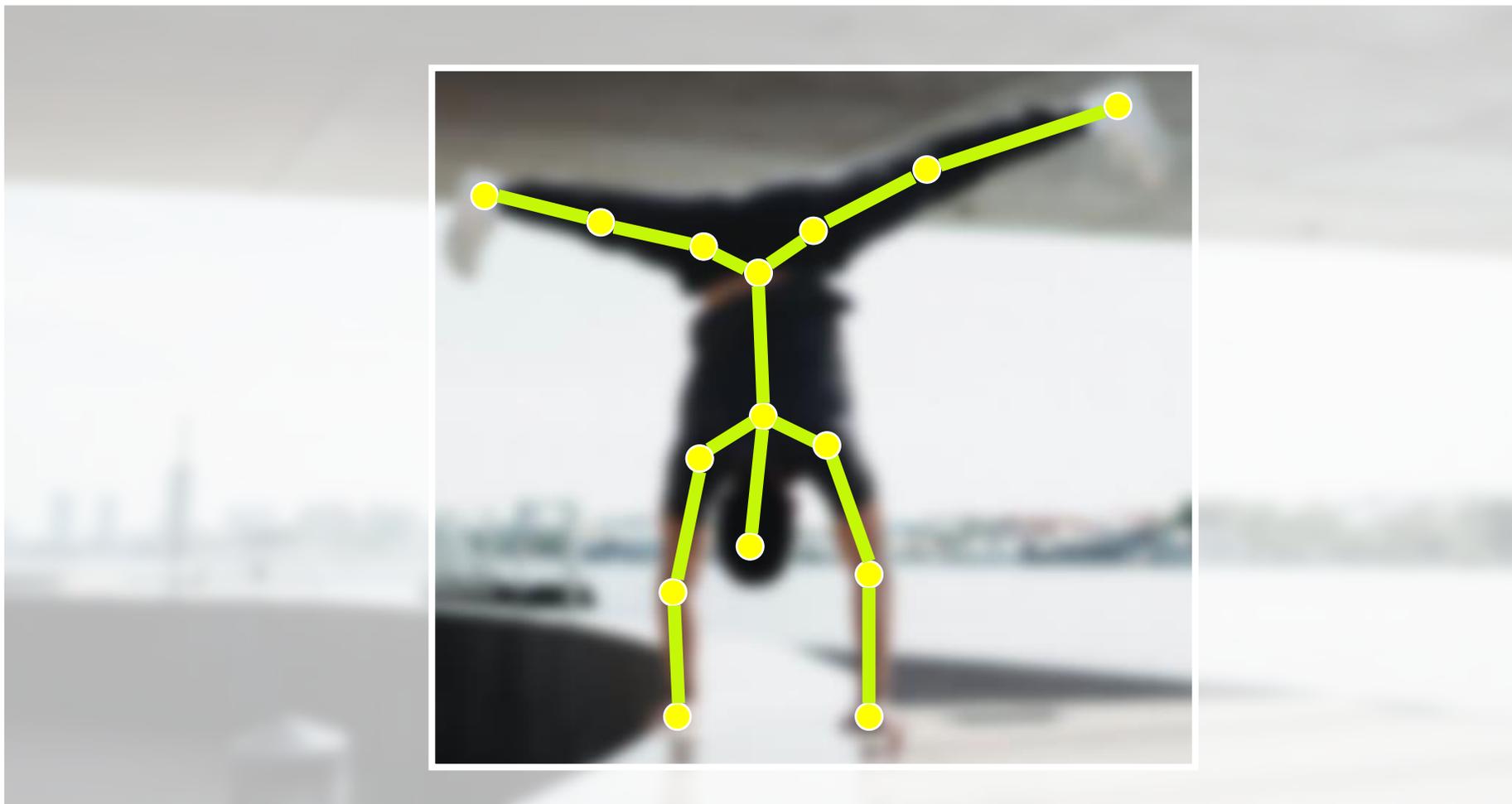
图像数据



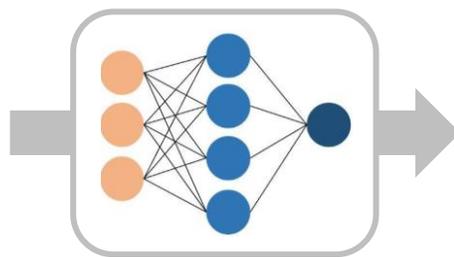
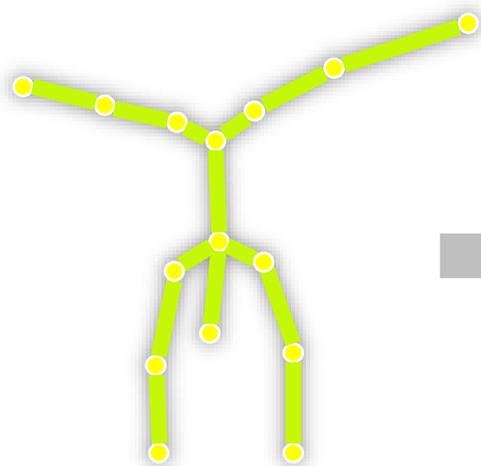
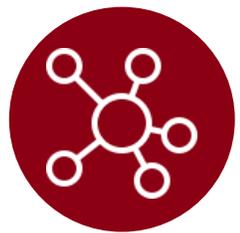
人体定位



骨架提取



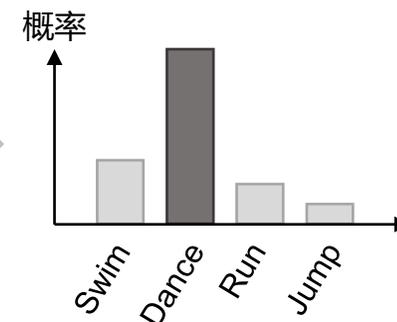
动作分析



神经网络

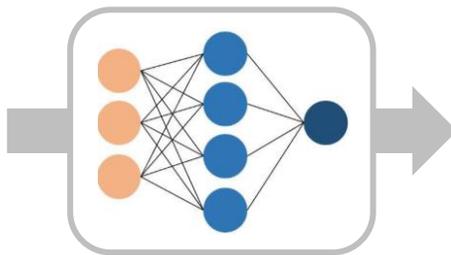
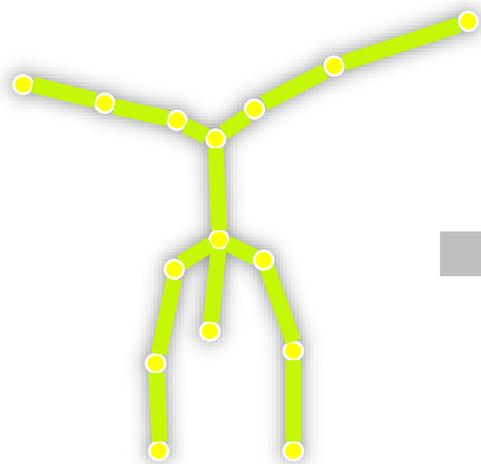
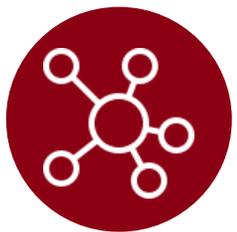


特征



概率分布

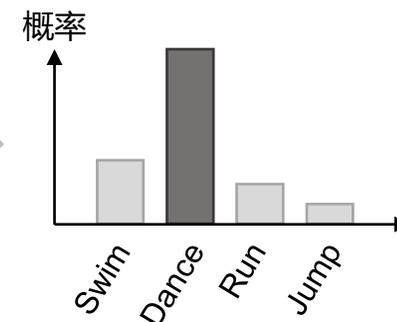
动作分析



神经网络



特征



概率分布

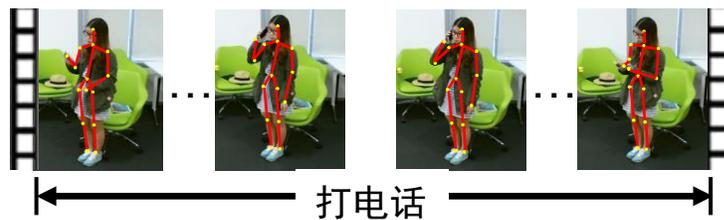


一个男人
正在跳街舞

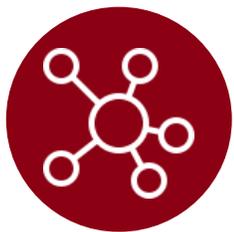
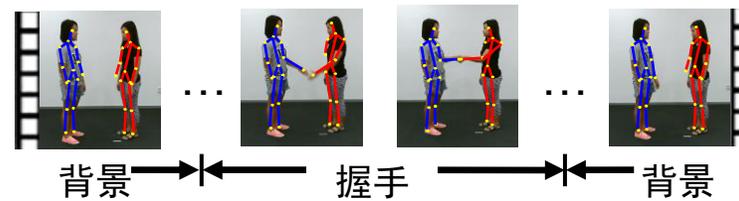
动作分析



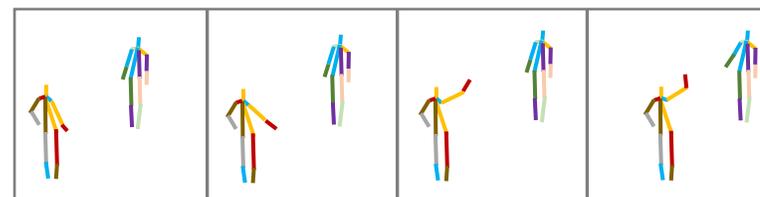
动作
识别



动作
检测



动作
生成

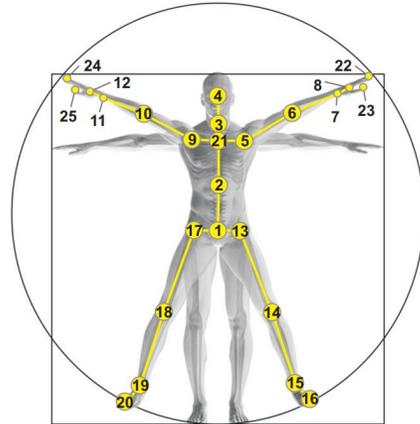


输入文本：A朝B挥手

人体骨架

优点

- 高级特征
- 鲁棒表示
- 深度信息
- 在线处理



缺点

- 视觉丢失
- 噪声问题



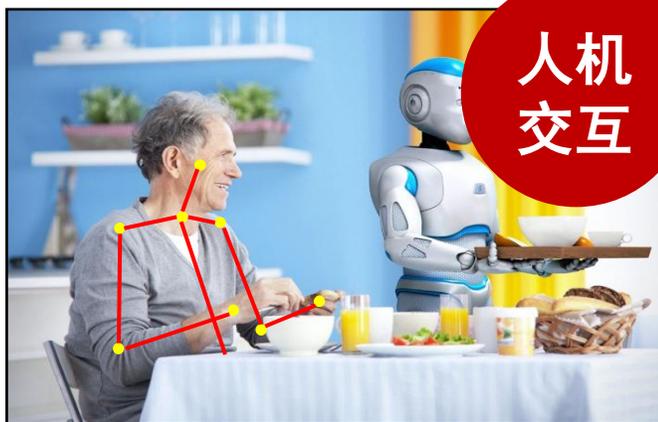
动作分析应用领域



视频
监控



健康
监护



人机
交互



运动
分析

相关工作

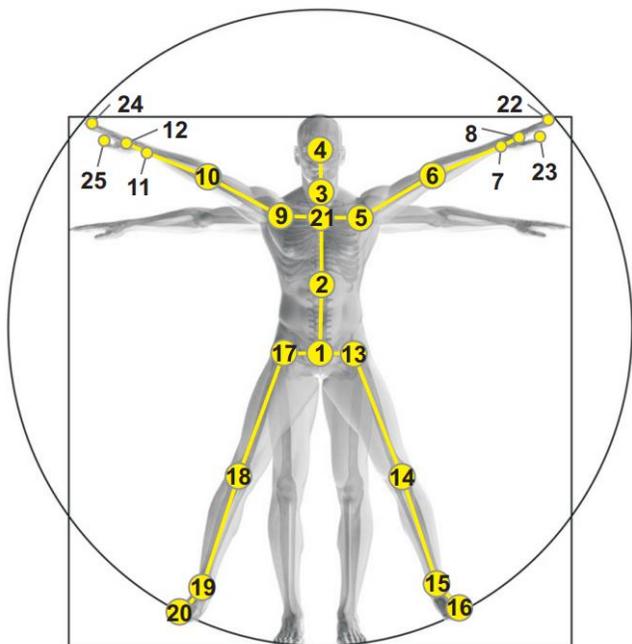
2014 CVPR
NW-UCLA



相关工作

2014 CVPR
NW-UCLA

2016 CVPR
NTU RGB+D

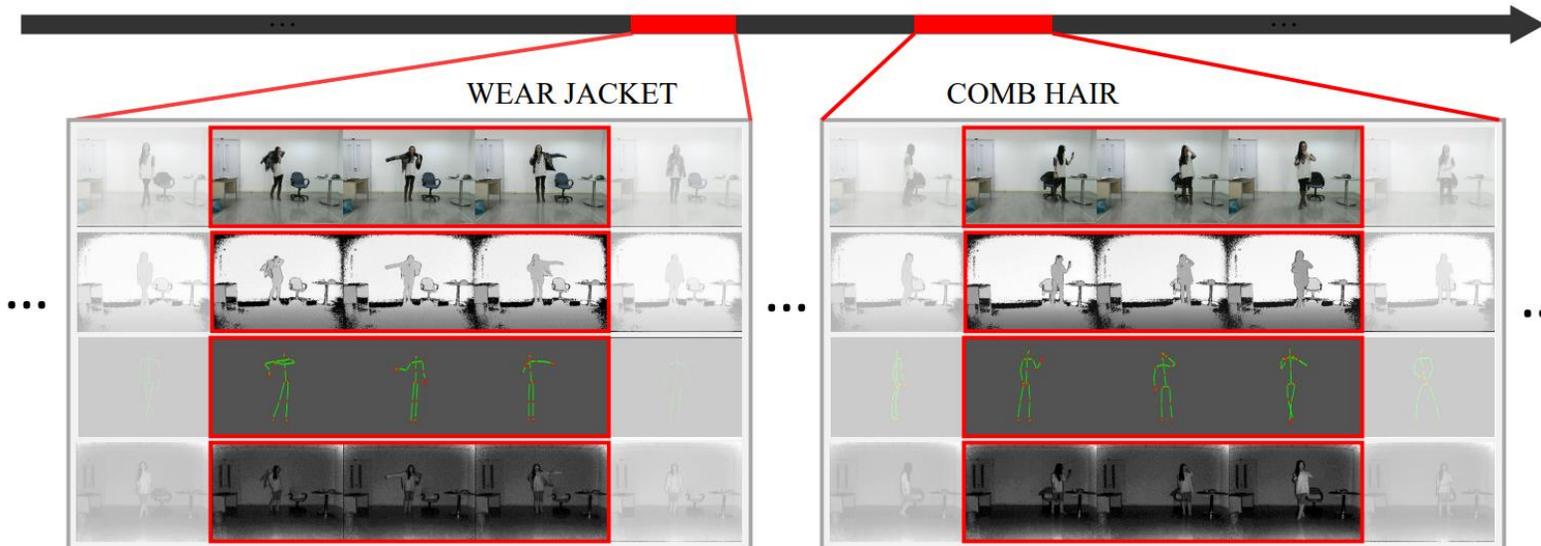
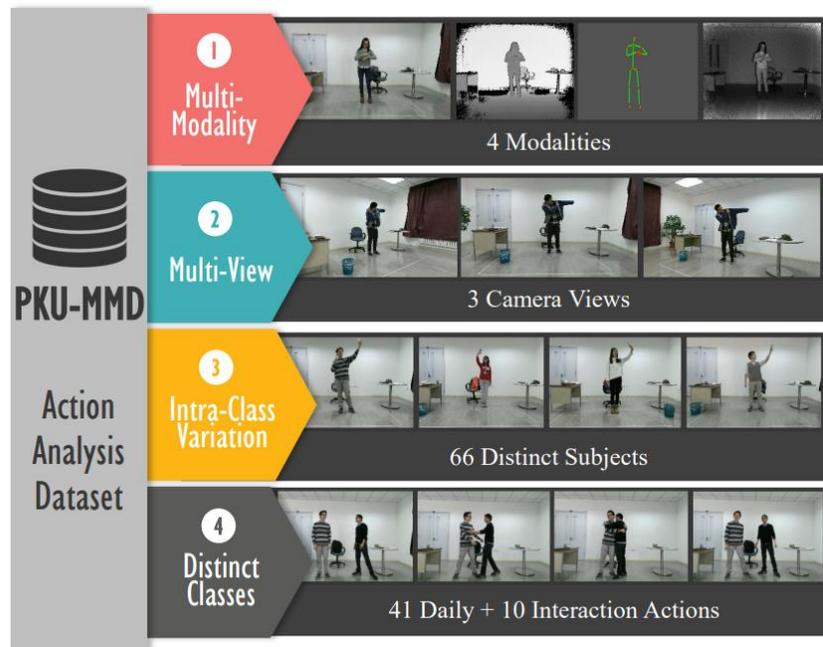


相关工作

2014 CVPR
NW-UCLA

2017 ACM MM
PKU-MMD

2016 CVPR
NTU RGB+D



相关工作

2014 CVPR
NW-UCLA

2017 ACM MM
PKU-MMD

2016 CVPR
NTU RGB+D

2019 TPAMI
NTU RGB+D 120



相关工作

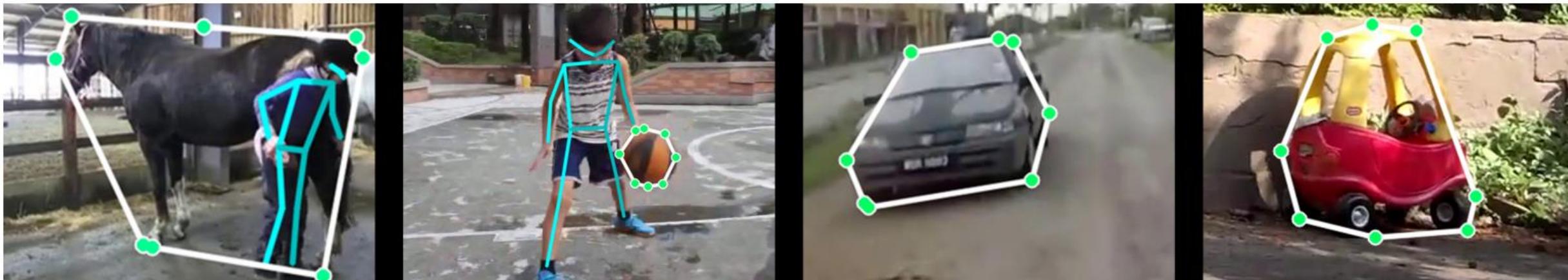
2014 CVPR
NW-UCLA

2017 ACM MM
PKU-MMD

2023 CVPR
SKP

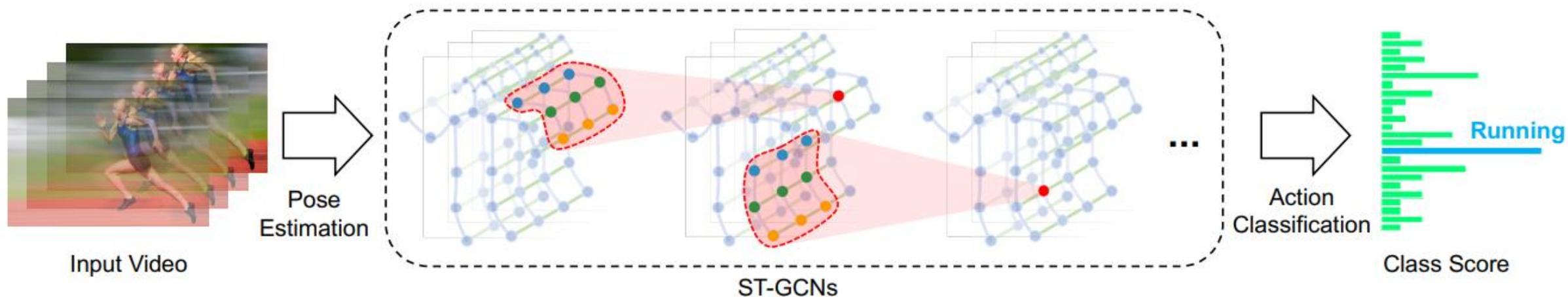
2016 CVPR
NTU RGB+D

2019 TPAMI
NTU RGB+D 120



相关工作

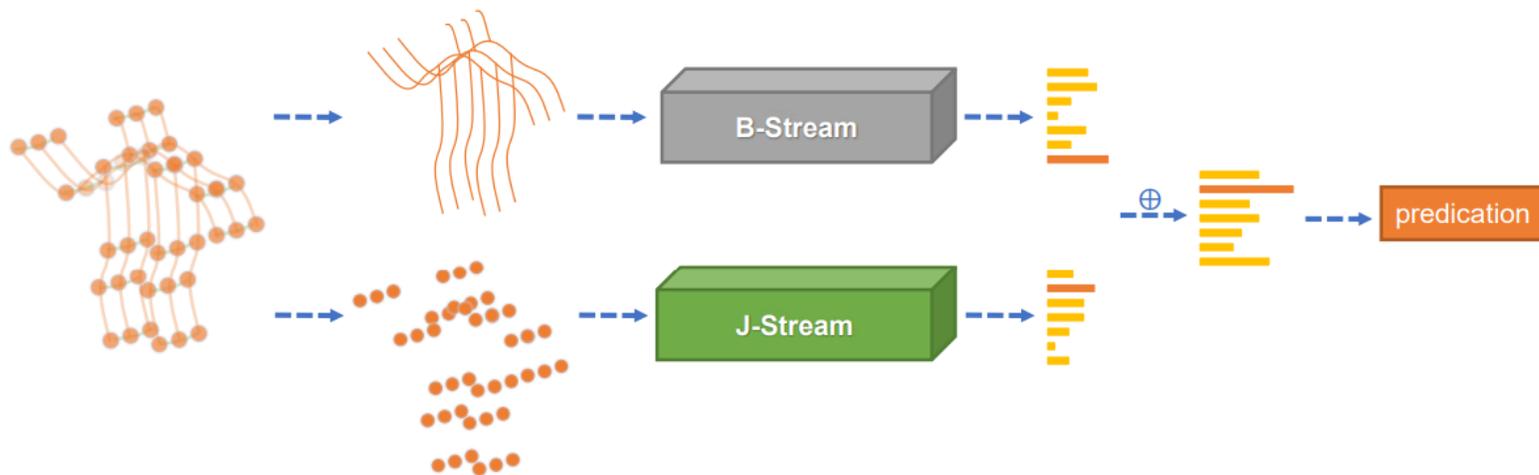
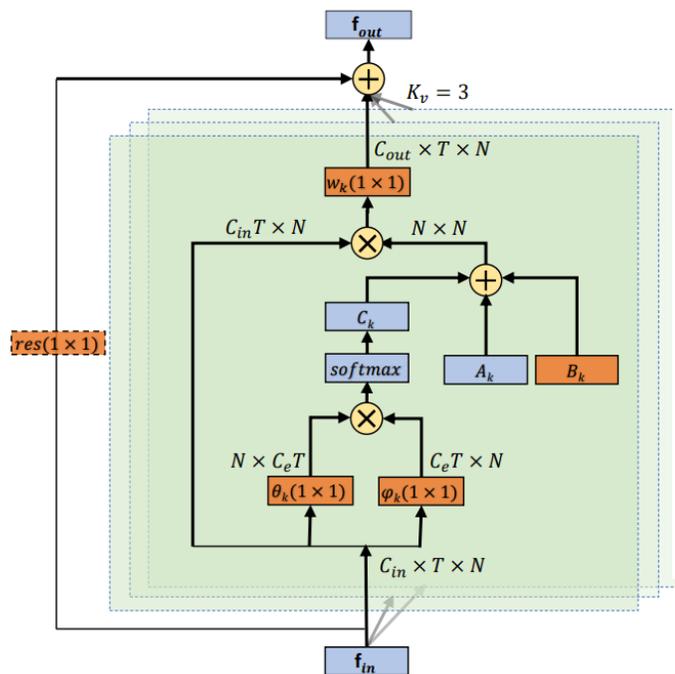
2018 AAAI
ST-GCN



相关工作

2018 AAAI
ST-GCN

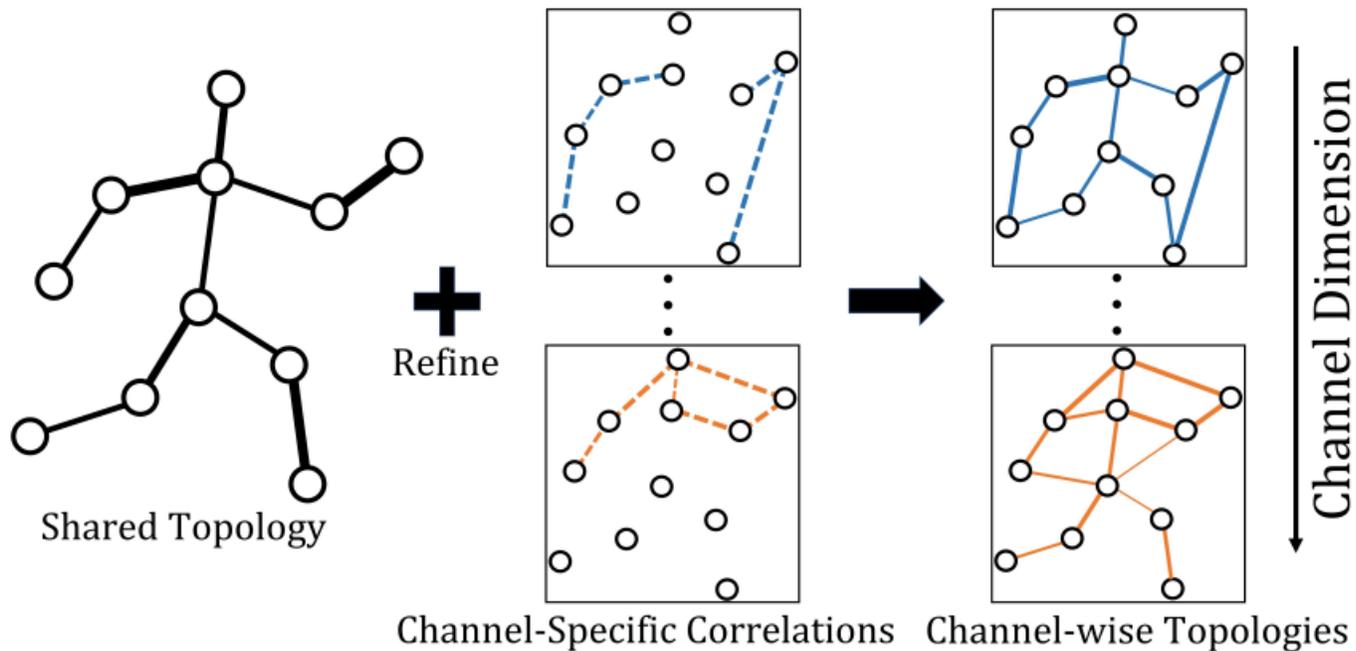
2019 CVPR
2s-AGCN



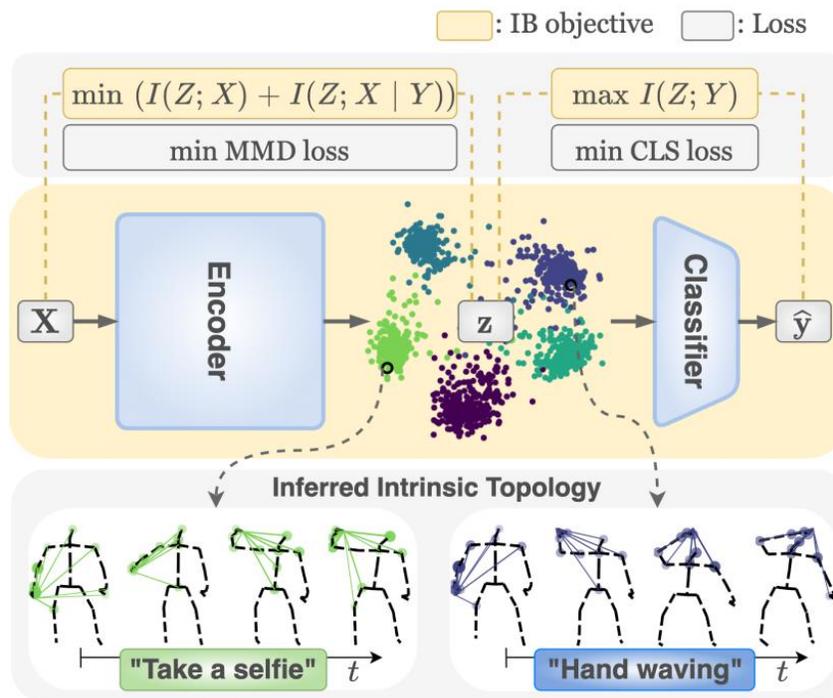
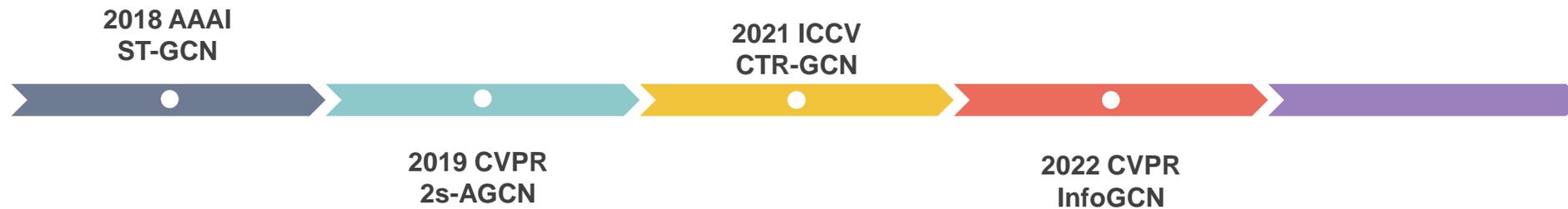
相关工作



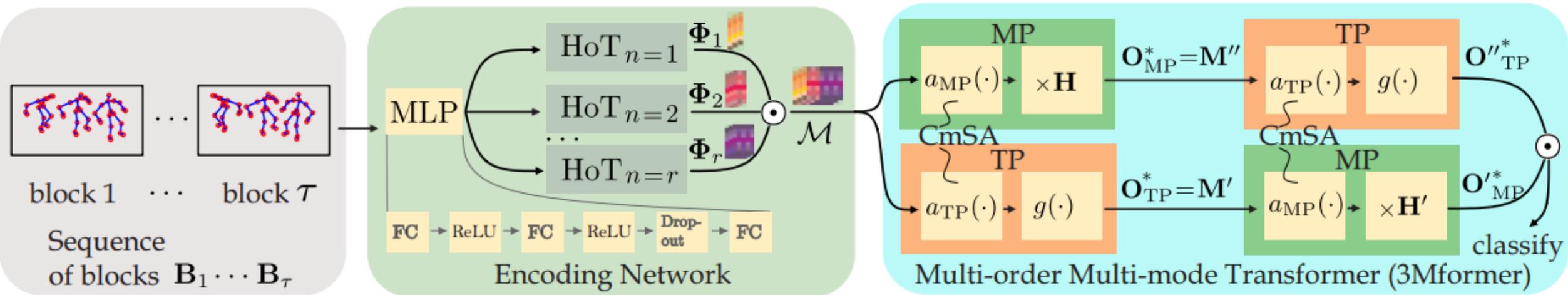
2019 CVPR
2s-AGCN



相关工作



相关工作



基于噪声适应的人体骨架动作识别

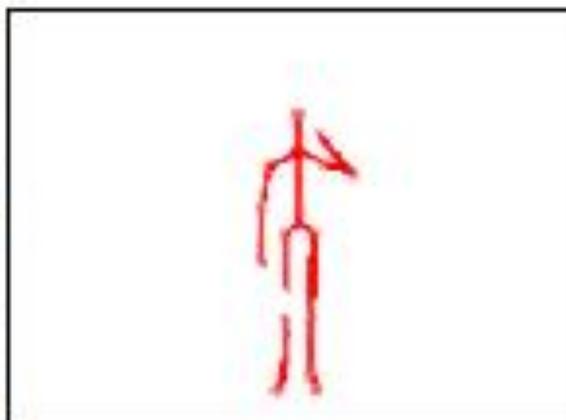
- 研究问题

实验室环境
(低噪声)

真实场景
(高噪声)



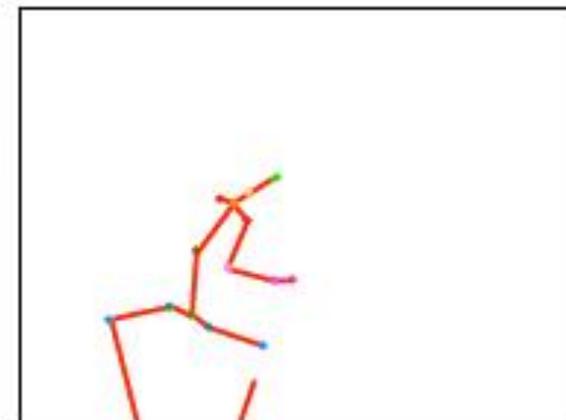
彩色图像



人体骨架



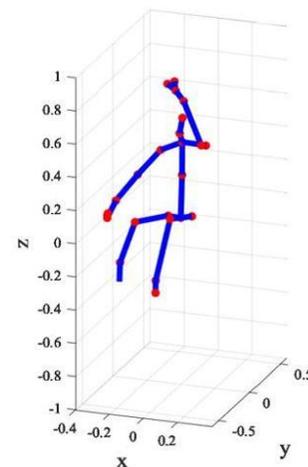
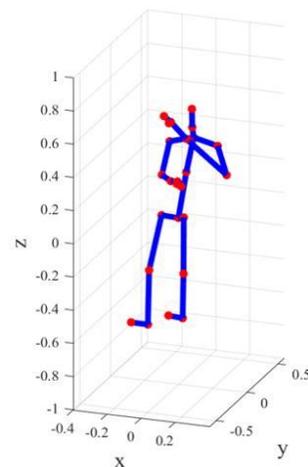
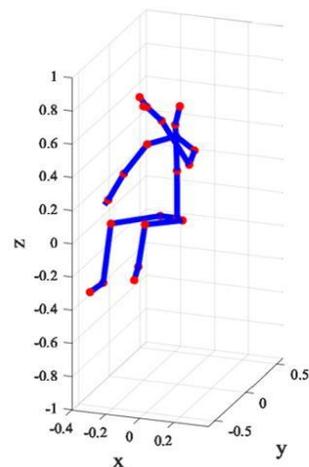
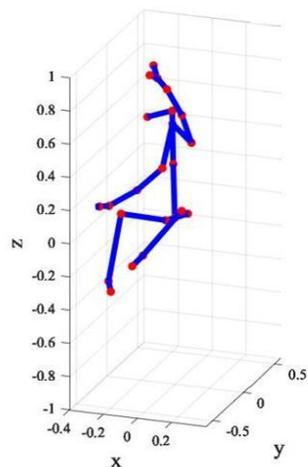
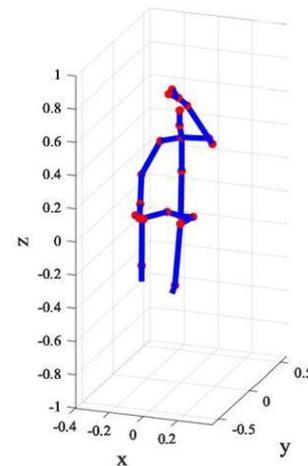
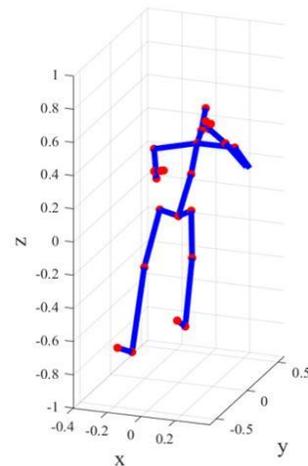
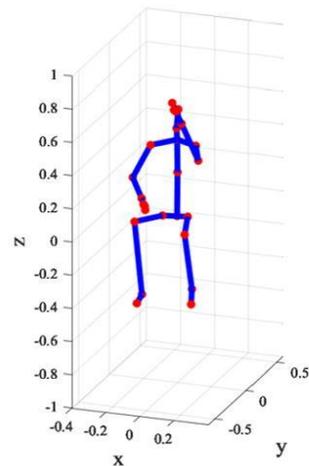
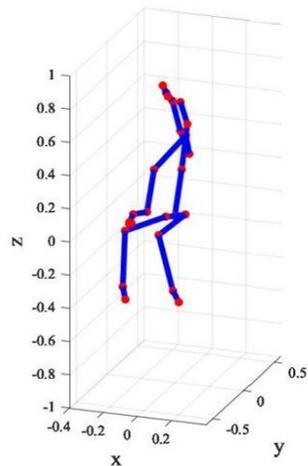
彩色图像



人体骨架

基于噪声适应的人体骨架动作识别

- 骨架去噪



原始噪声骨架

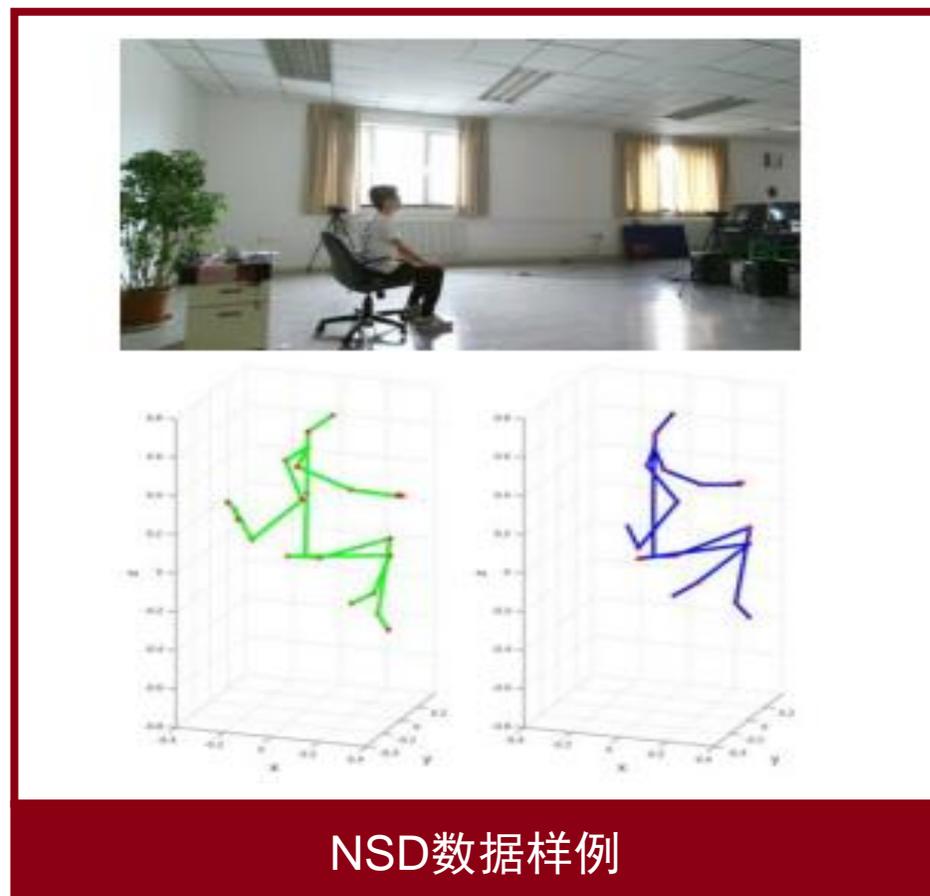
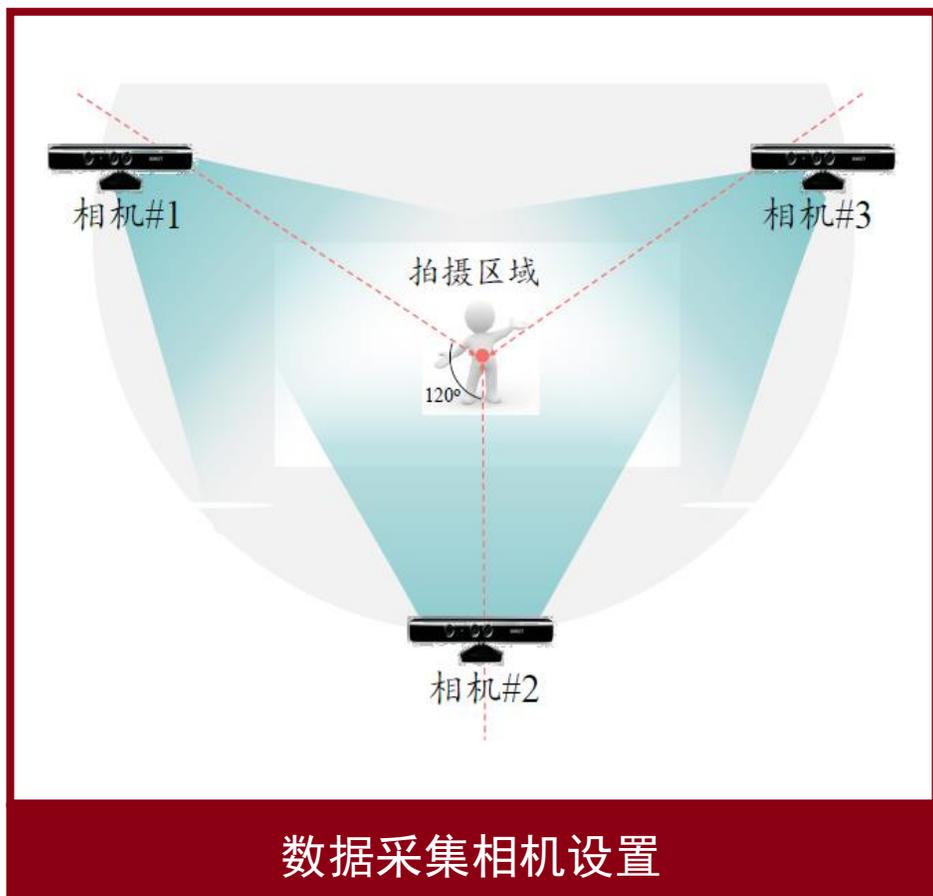
R-SD

PE-AE

G-SD

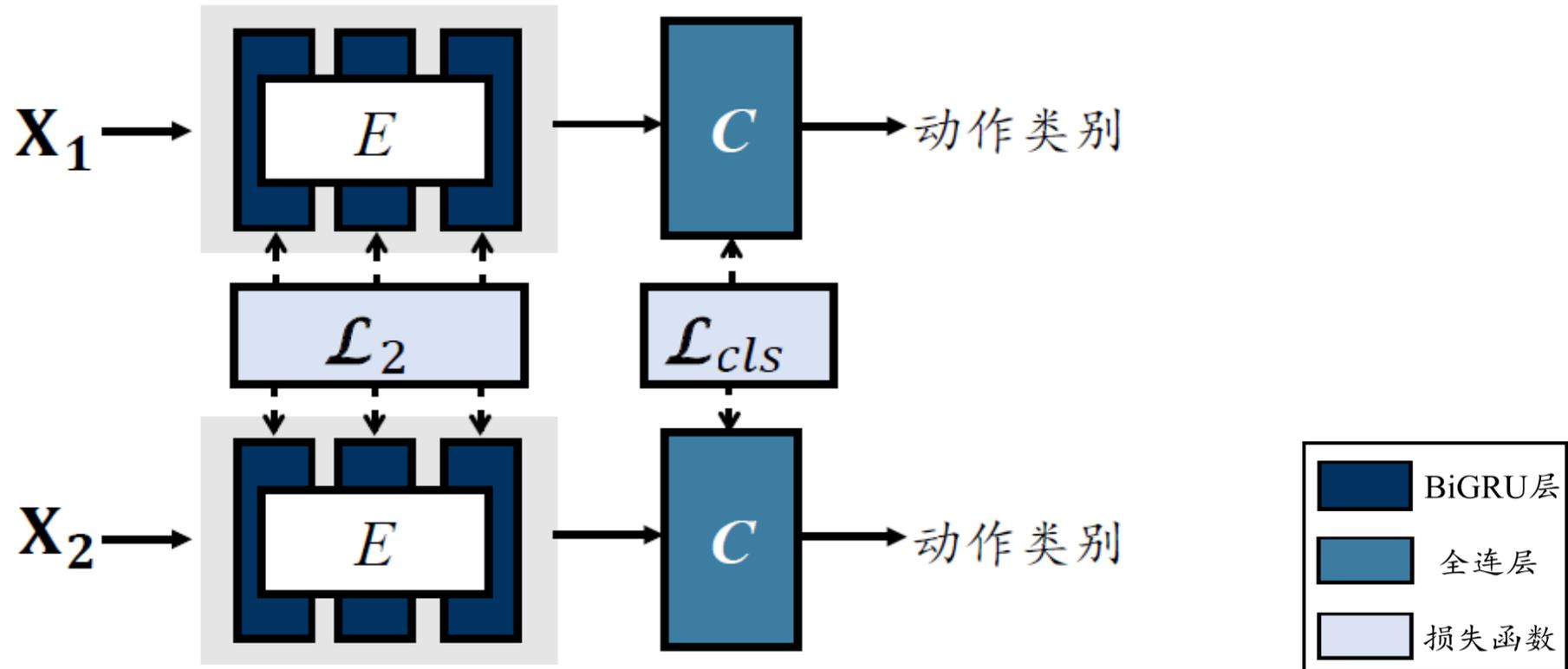
噪声数据集建立

- NSD (Noisy Skeleton Dataset) 数据库概览



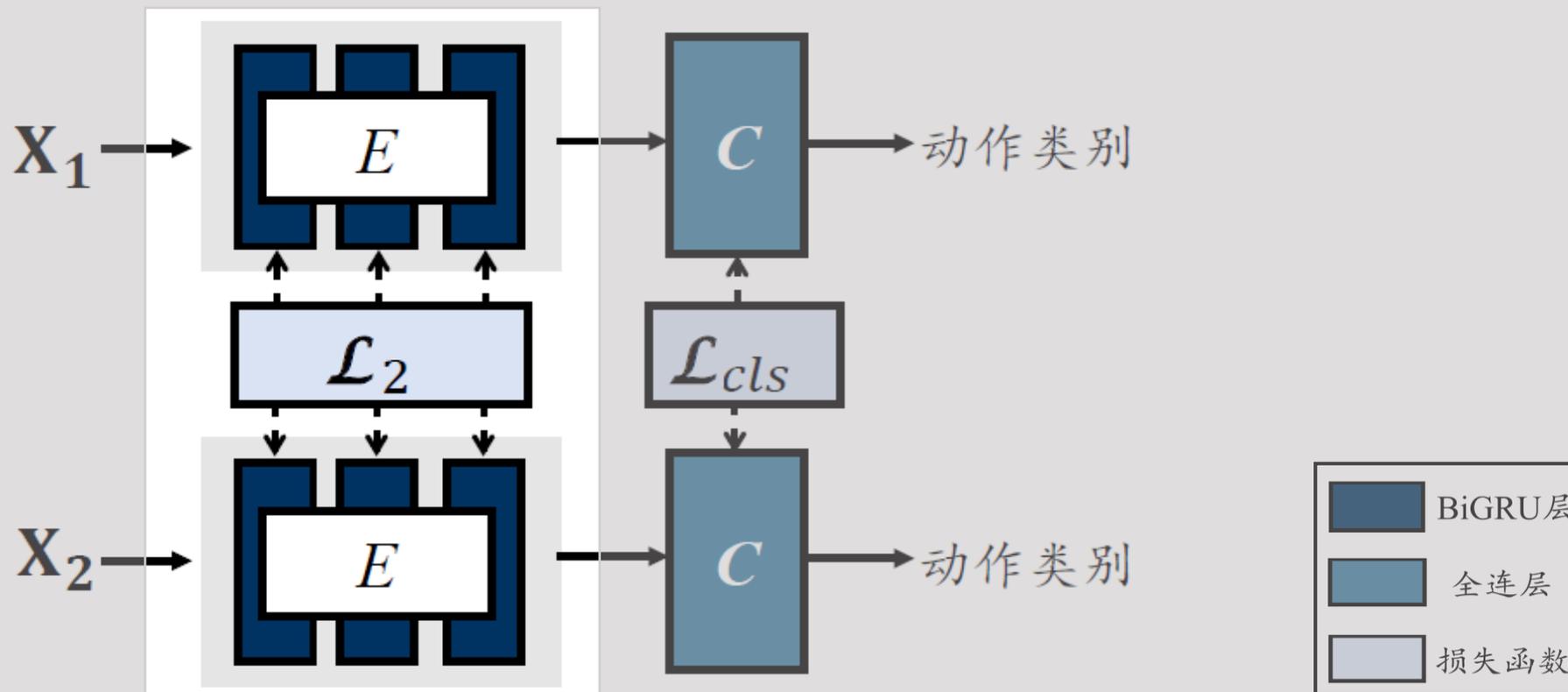
基于噪声适应的鲁棒特征学习

- 基于回归的噪声适应模型



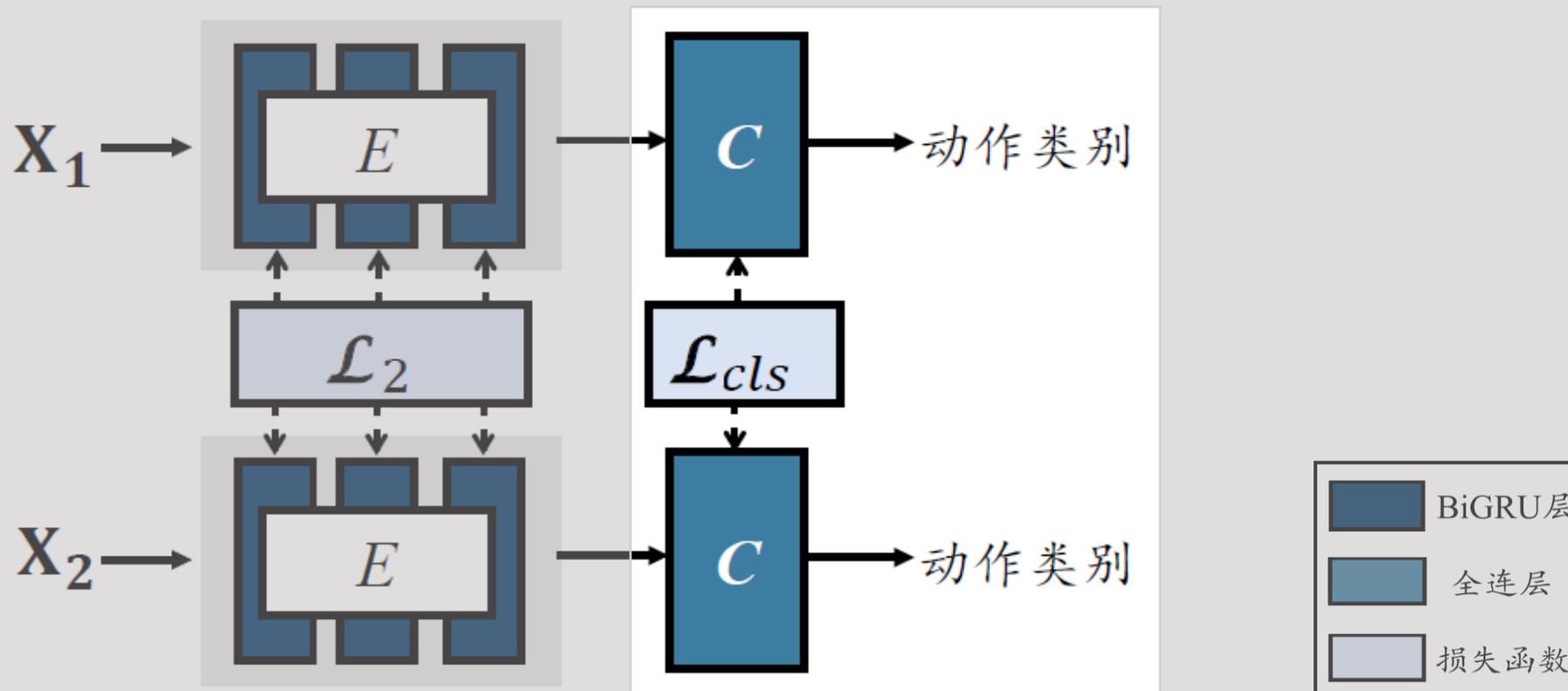
基于噪声适应的鲁棒特征学习

- 基于回归的噪声适应模型



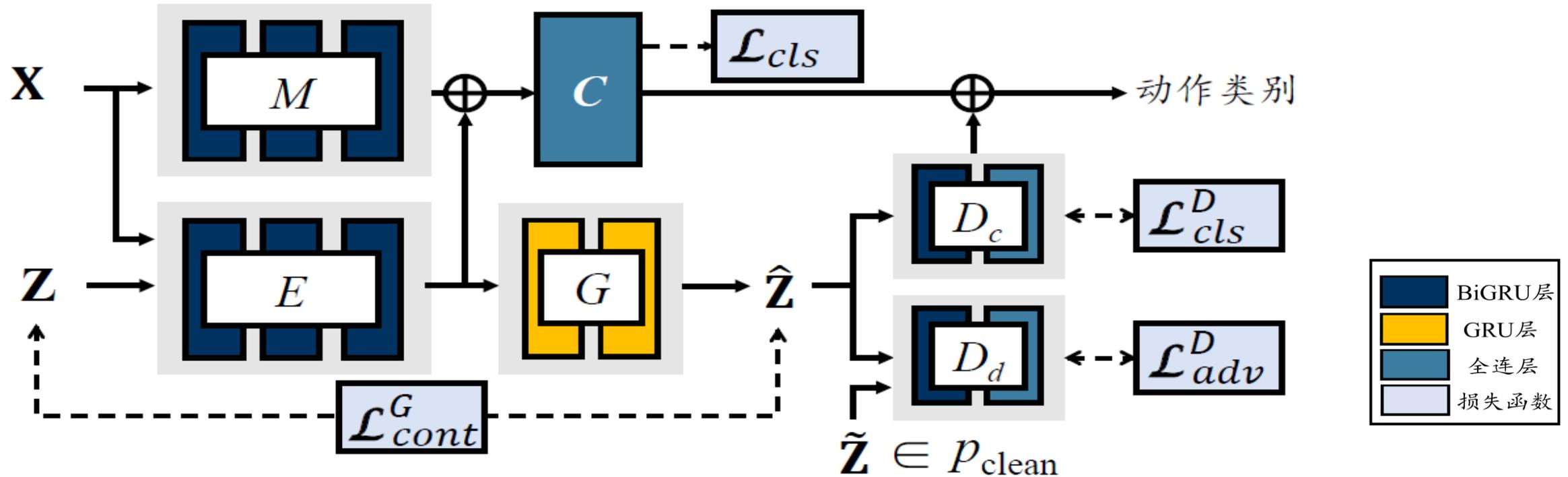
基于噪声适应的鲁棒特征学习

- 基于回归的噪声适应模型



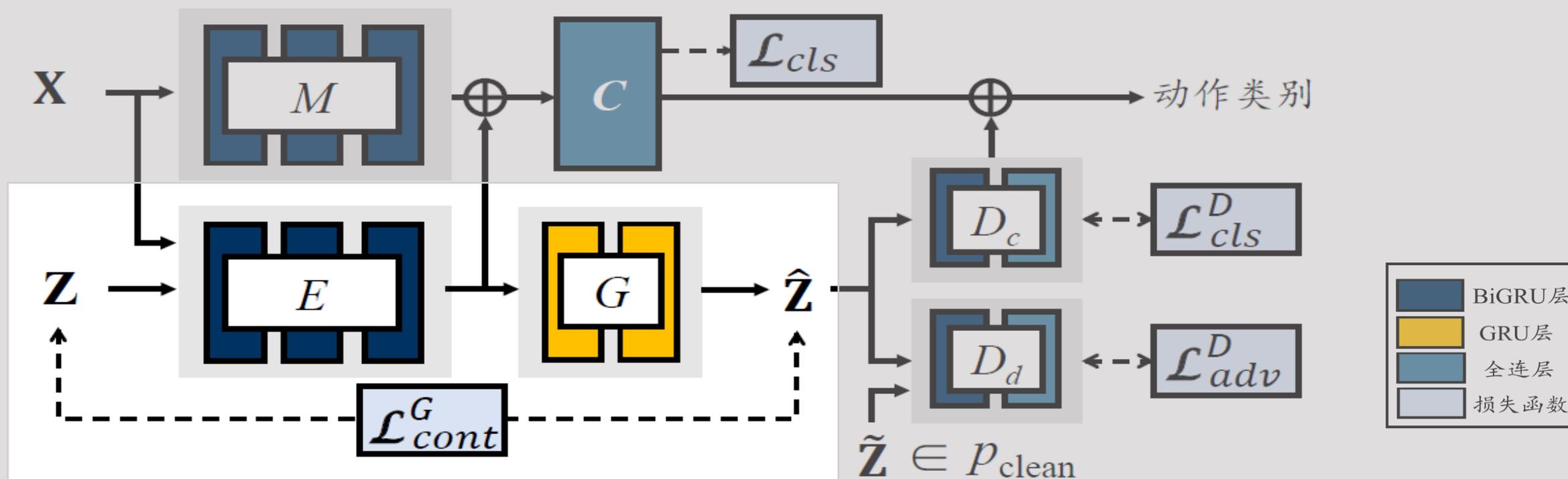
基于噪声适应的鲁棒特征学习

- 基于生成的噪声适应模型



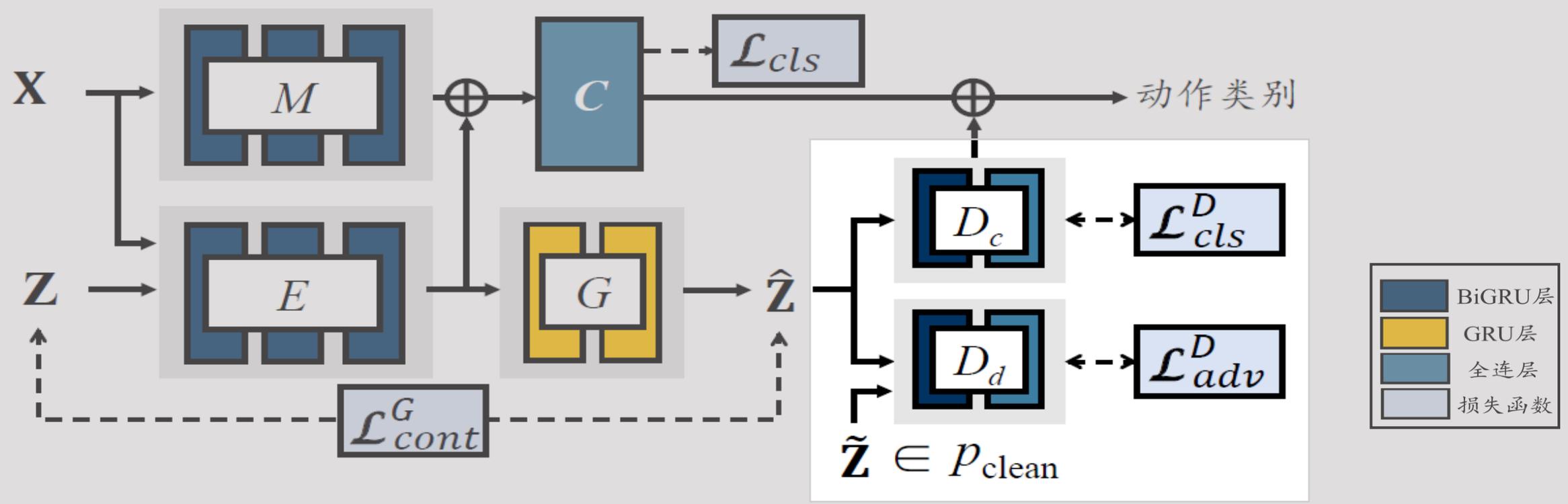
基于噪声适应的鲁棒特征学习

- 基于生成的噪声适应模型



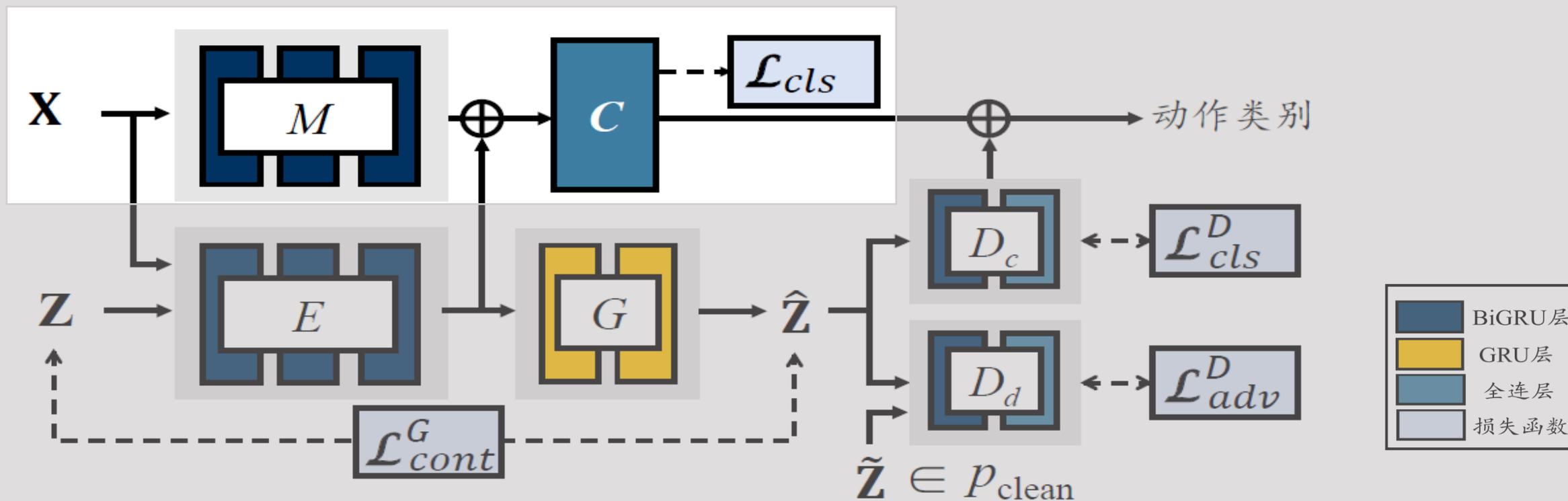
基于噪声适应的鲁棒特征学习

- 基于生成的噪声适应模型



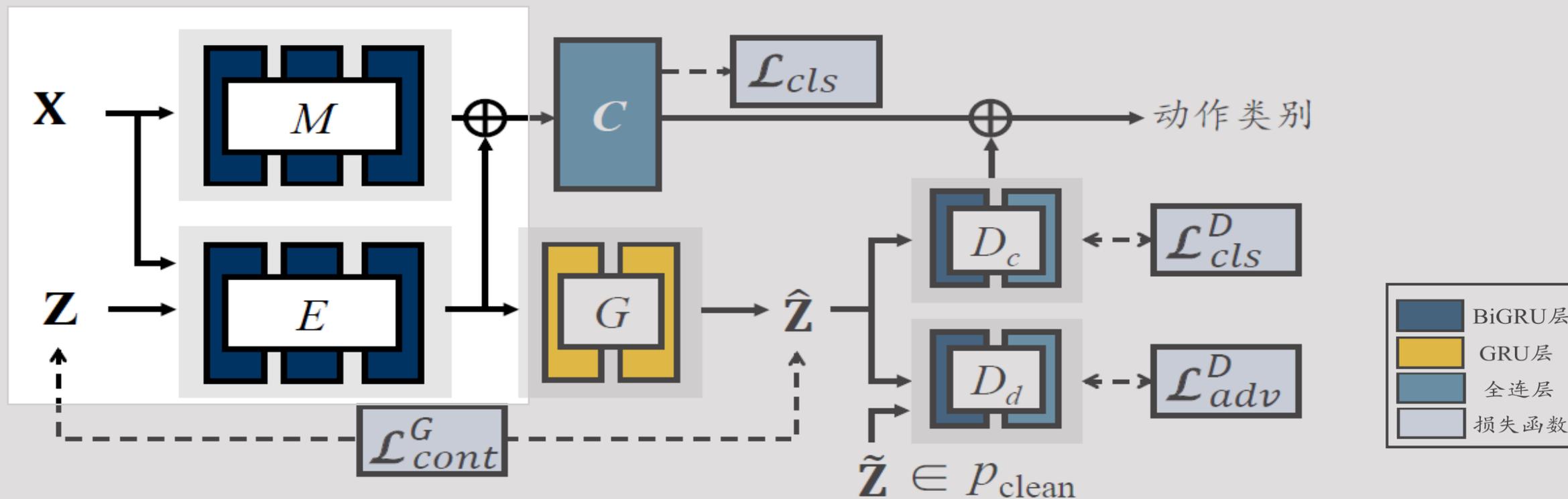
基于噪声适应的鲁棒特征学习

- 基于生成的噪声适应模型



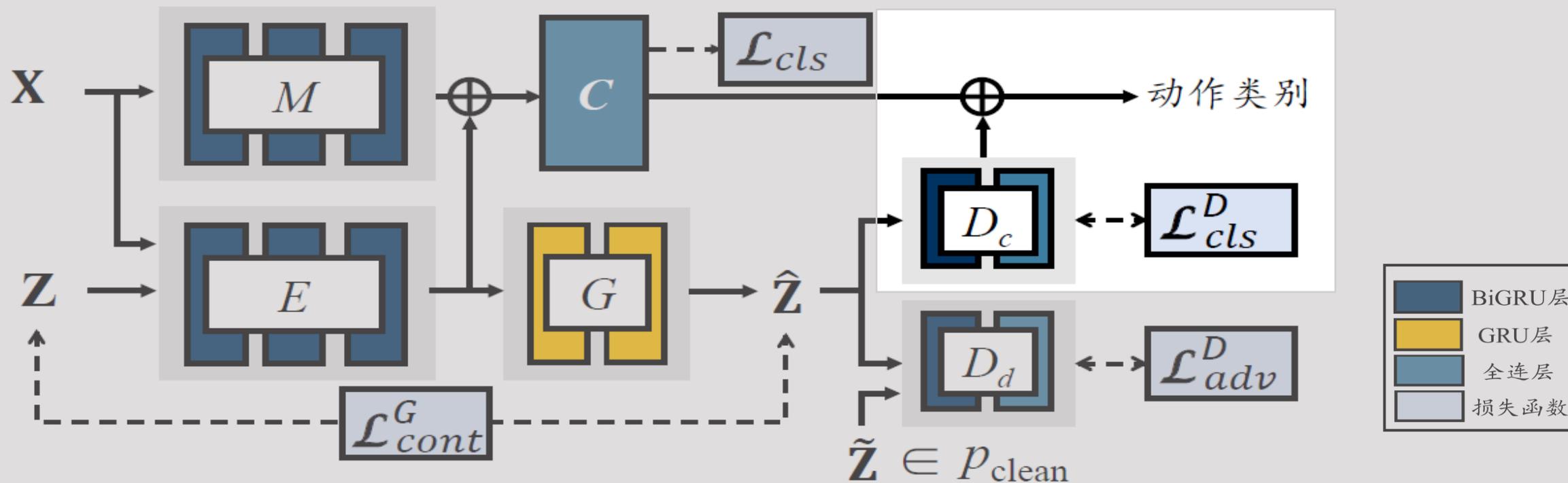
基于噪声适应的鲁棒特征学习

- 基于生成的噪声适应模型



基于噪声适应的鲁棒特征学习

- 基于生成的噪声适应模型



实验结果

- NSD数据集

测试方法 (Acc. %)	Cross-Subject	Cross-View
STA-LSTM	44.3	28.6
TPN ¹	46.9	29.7
VA-LSTM ²	50.0	34.5
ST-GCN ³	48.2	35.8
Denoised-LSTM ⁴	38.1	26.1
Baseline	50.7	34.6
R-NAN	<u>55.3</u>	40.5
G-NAN	55.5	<u>36.3</u>

¹Hu *et al.* Temporal Perceptive Network for Skeleton-Based Action Recognition. *BMVC 2017*.

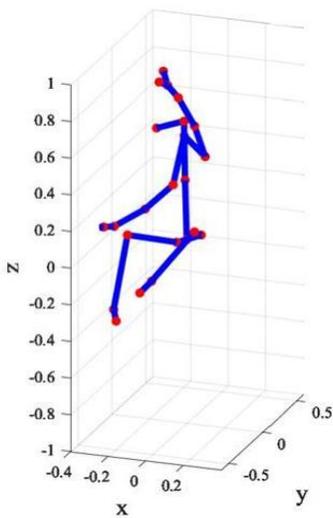
²Zhang *et al.* View adaptive recurrent neural networks for high performance human action recognition from skeleton data. *ICCV 2017*.

³Yan, *et al.* Spatial temporal graph convolutional networks for skeleton-based action recognition. *AAAI 2018*.

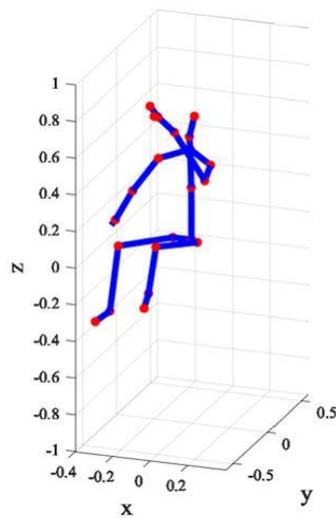
⁴Demisse *et al.* Pose encoding for robust skeleton-based action recognition. *CVPRW 2018*.

实验结果

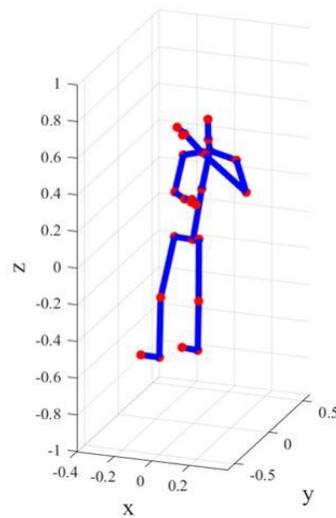
- 和骨架去噪方法的比较 - 主观分析



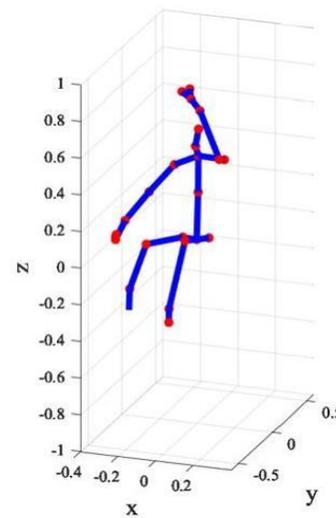
原始噪声骨架



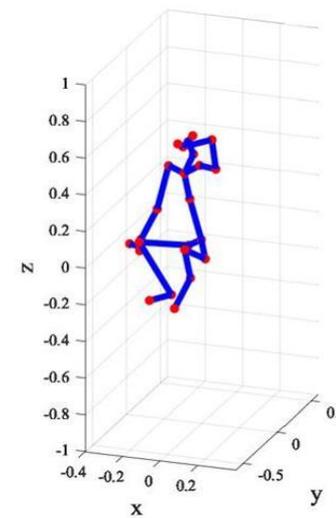
R-SD



PE-AE



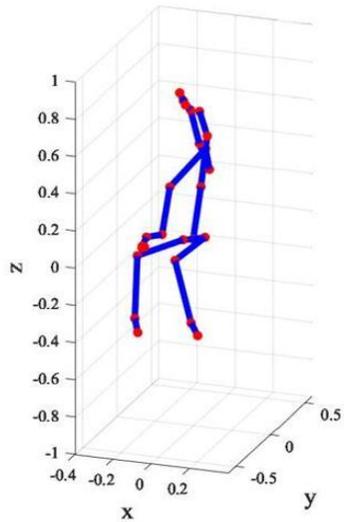
G-SD



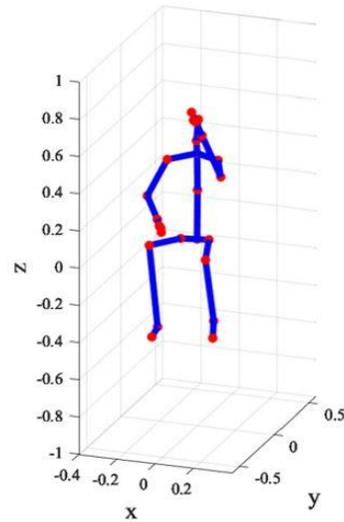
G-NAN

实验结果

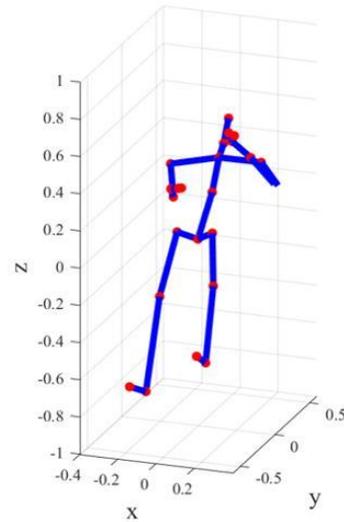
- 和骨架去噪方法的比较 - 主观分析



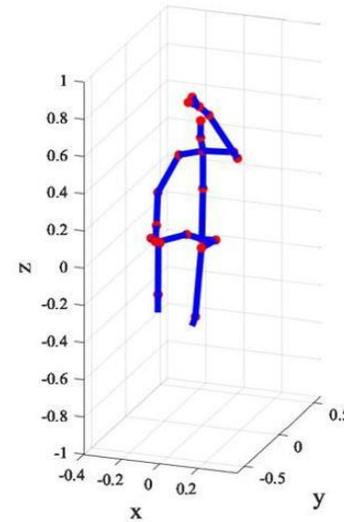
原始噪声骨架



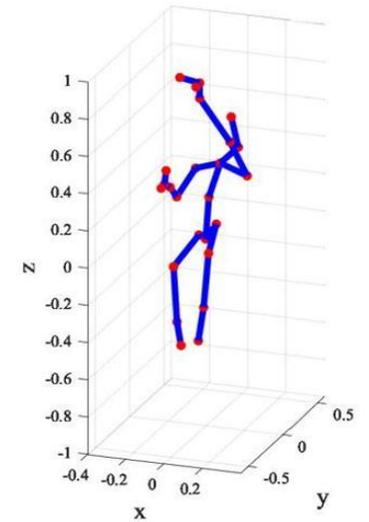
R-SD



PE-AE



G-SD



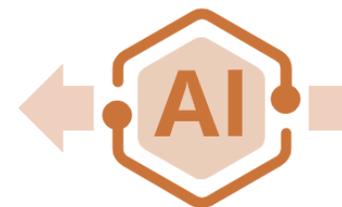
G-NAN

语义引导的生成



“一个男人
正在跳街舞”

语义引导的生成



“一个男人
正在跳街舞”



跨越语义鸿沟的开放世界图像生成

Open-World Image Generation across Semantic Gap

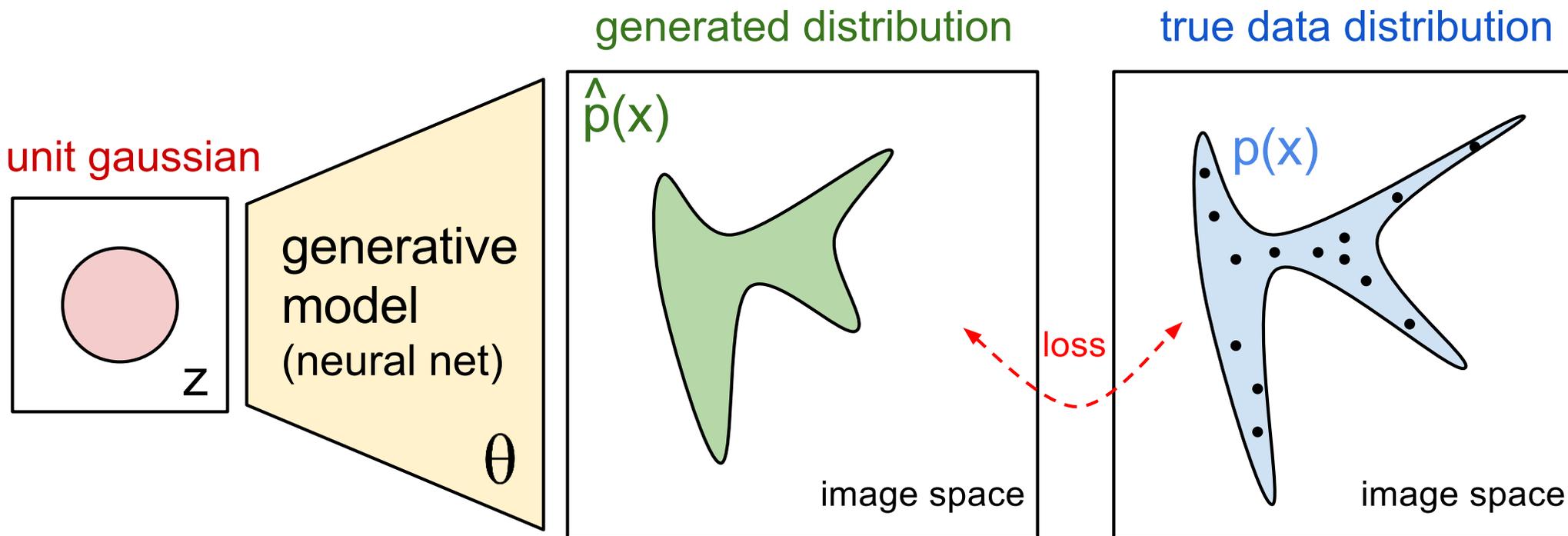
马逸扬 北京大学



“在上海跳舞的女孩”

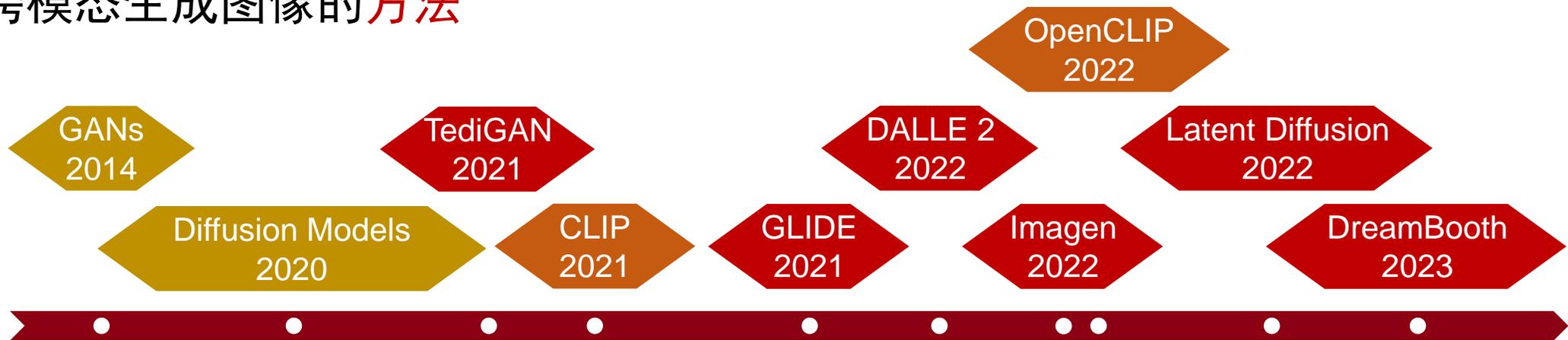
相关工作

- 生成模型：合成符合真实数据分布的“伪数据”

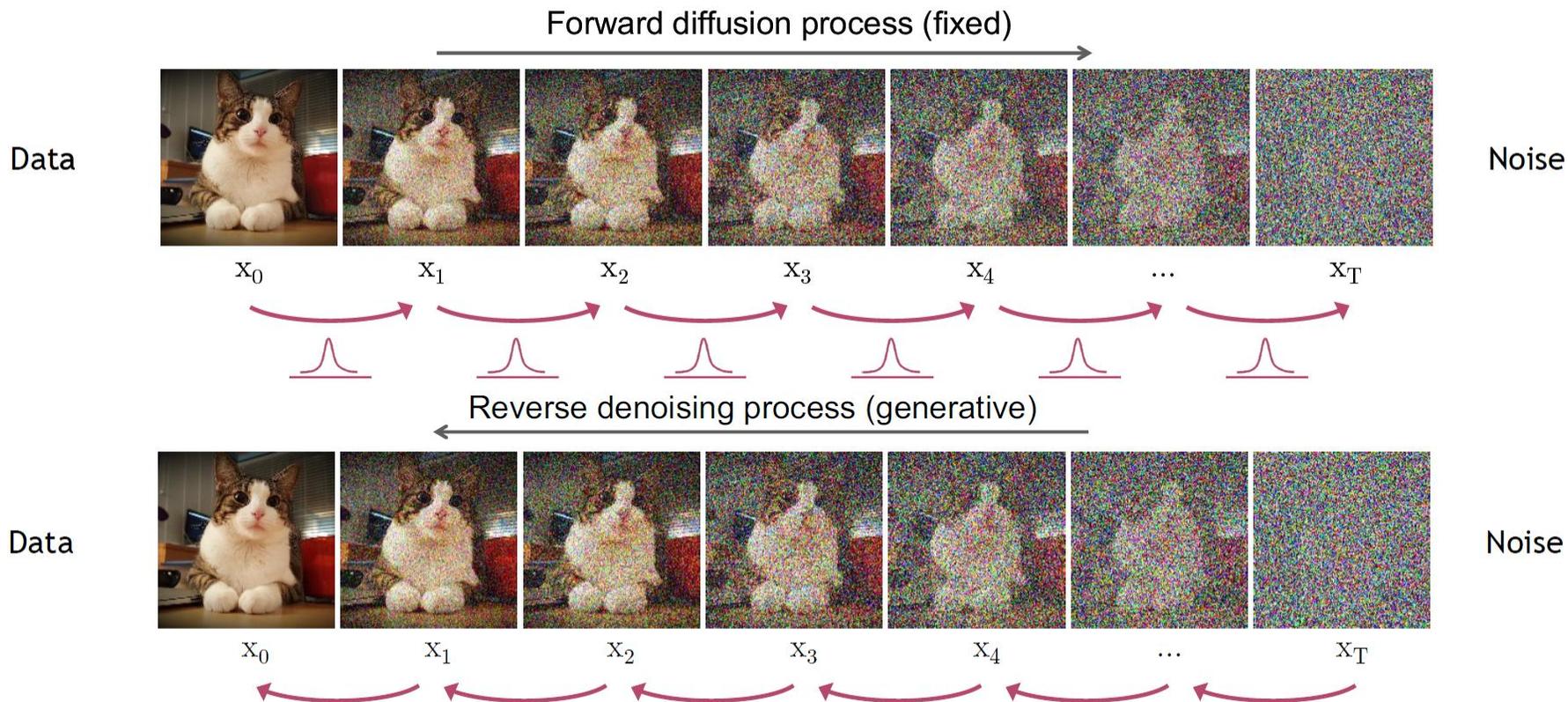


相关工作

- 高质量的生成模型
- 衡量跨模态语义对齐的标准
- 跨模态生成图像的方法



相关工作



相关工作

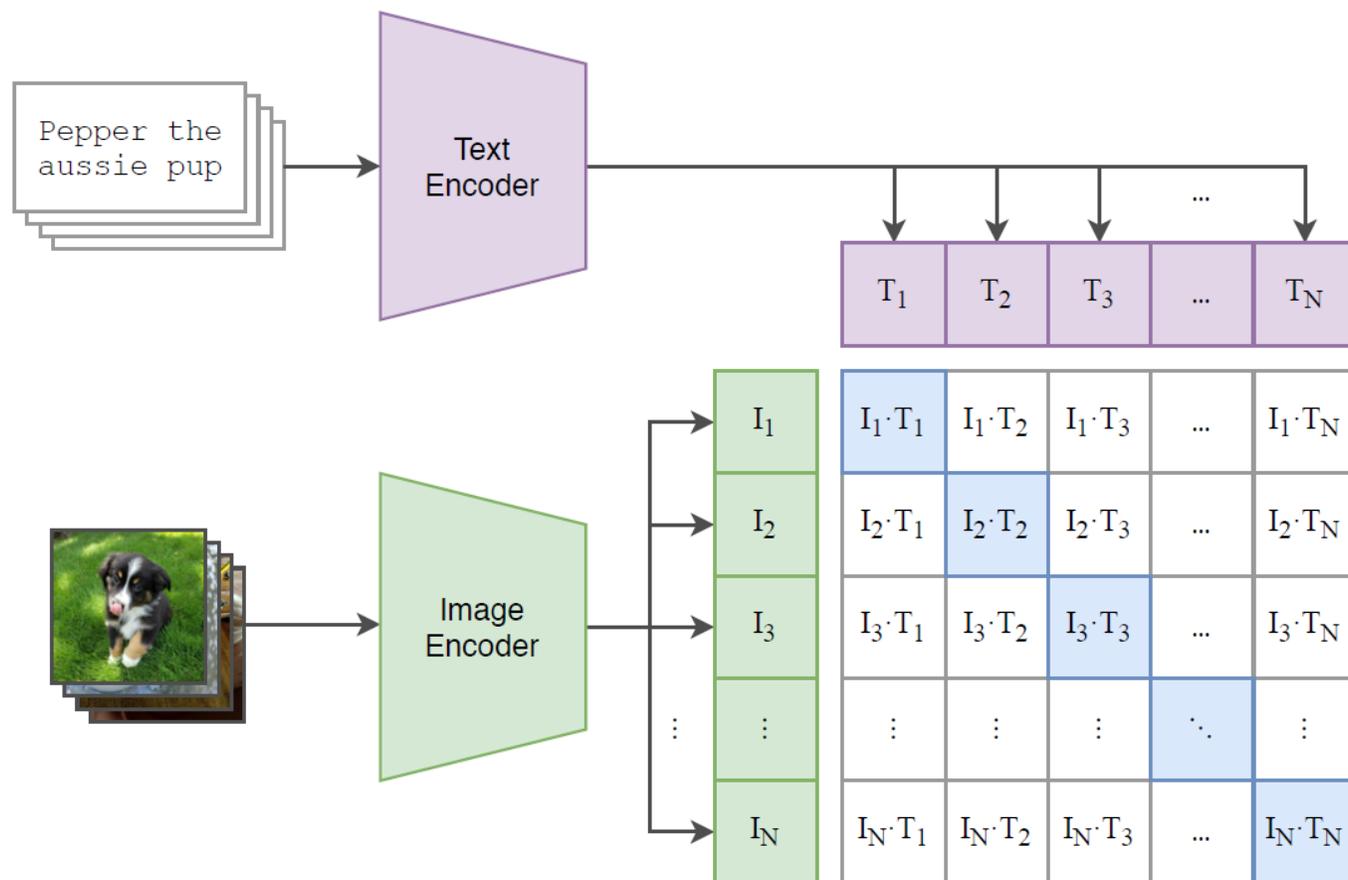
DDPM
2020

CLIP
2021

DALLE 2
2022

Imagen
2022

LDM
2022



相关工作

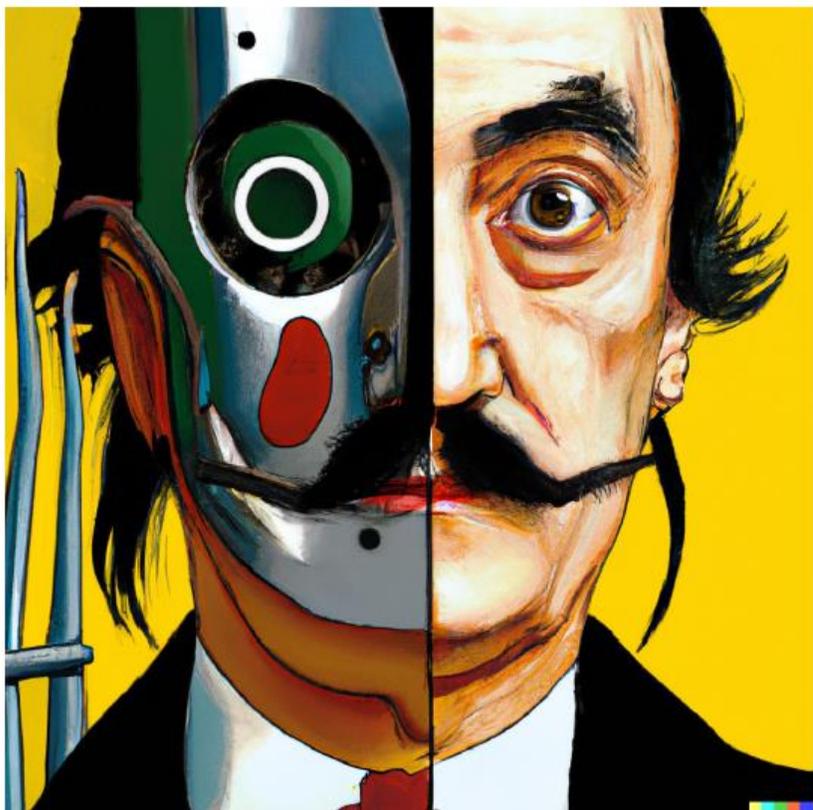
DDPM
2020

CLIP
2021

DALLE 2
2022

Imagen
2022

LDM
2022



vibrant portrait painting of Salvador Dalí with a robotic half face



a shiba inu wearing a beret and black turtleneck

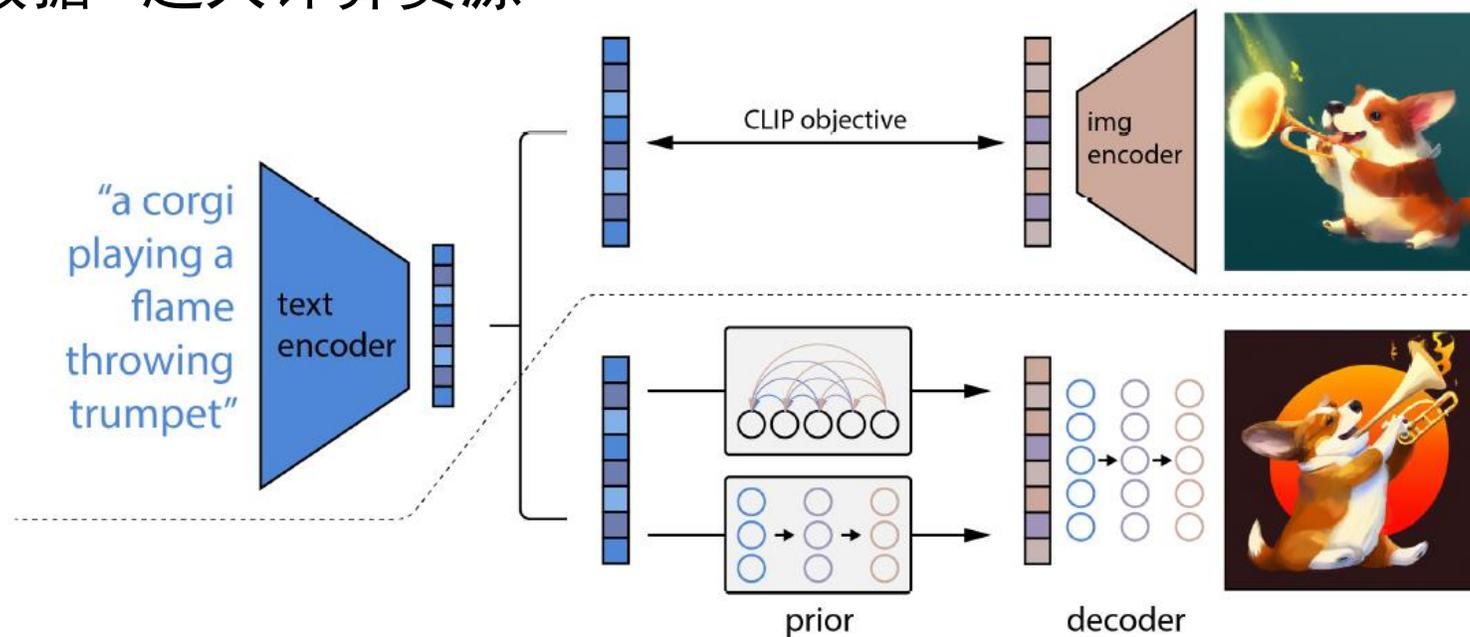


a close up of a handpalm with leaves growing from it

相关工作



- 串联两个跨模态扩散模型+一个上采样扩散模型
- 超大规模数据+超大计算资源



相关工作

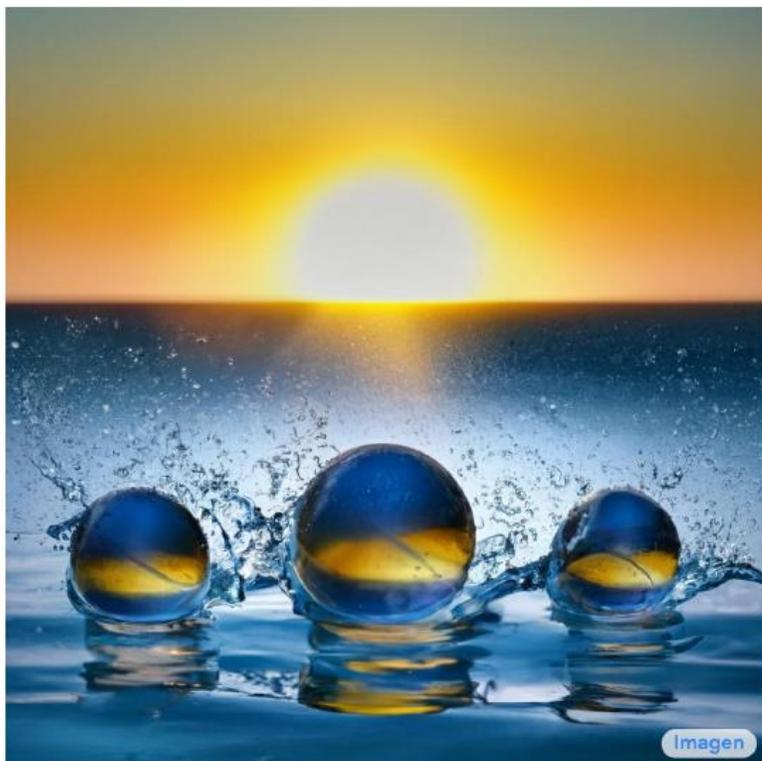
DDPM
2020

CLIP
2021

DALLE 2
2022

Imagen
2022

LDM
2022



Three spheres made of glass falling into ocean. Water is splashing. Sun is setting.



Vines in the shape of text 'Imagen' with flowers and butterflies bursting out of an old TV.

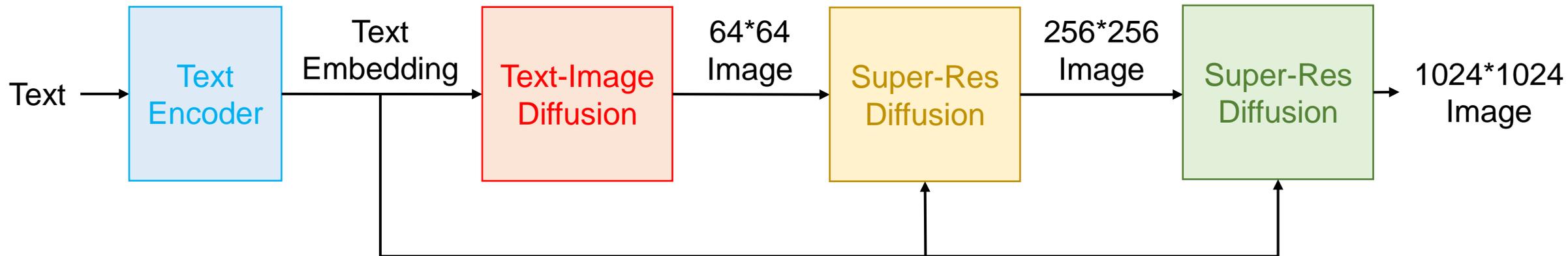


A strawberry splashing in the coffee in a mug under the starry sky.

相关工作



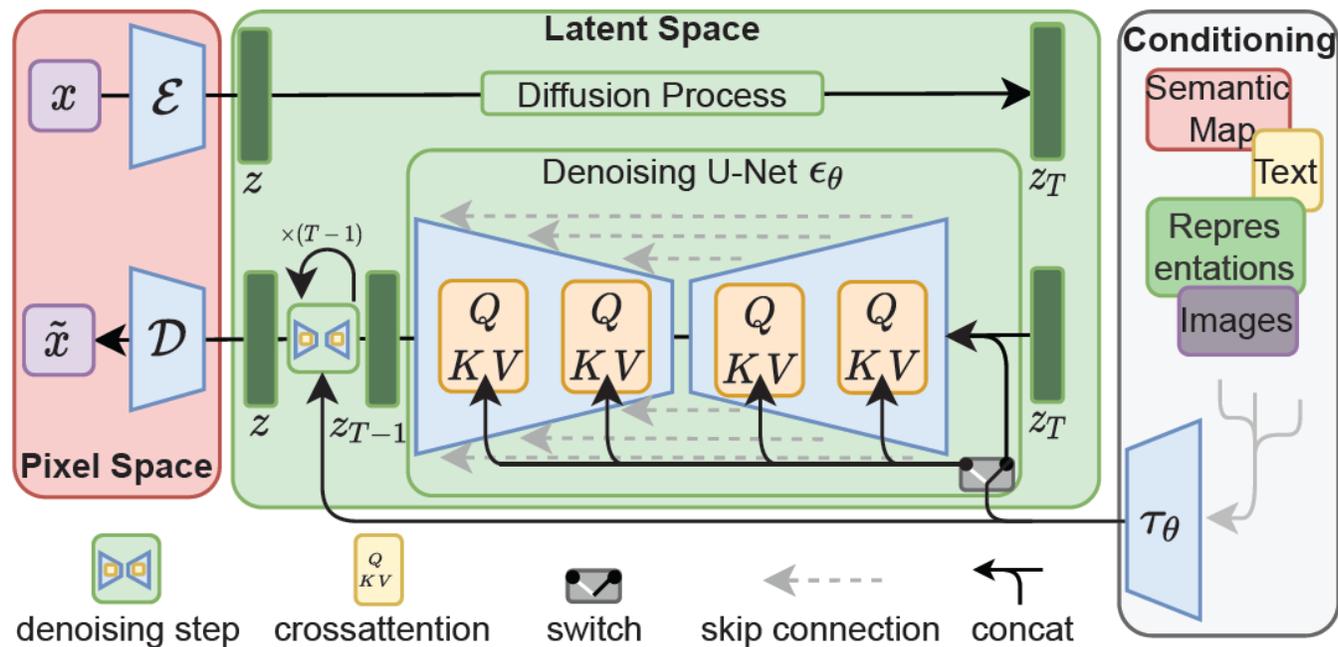
- 一个跨模态扩散模型+上采样扩散模型



相关工作



- 降低扩散模型的训练&采样代价



相关工作

DDPM
2020

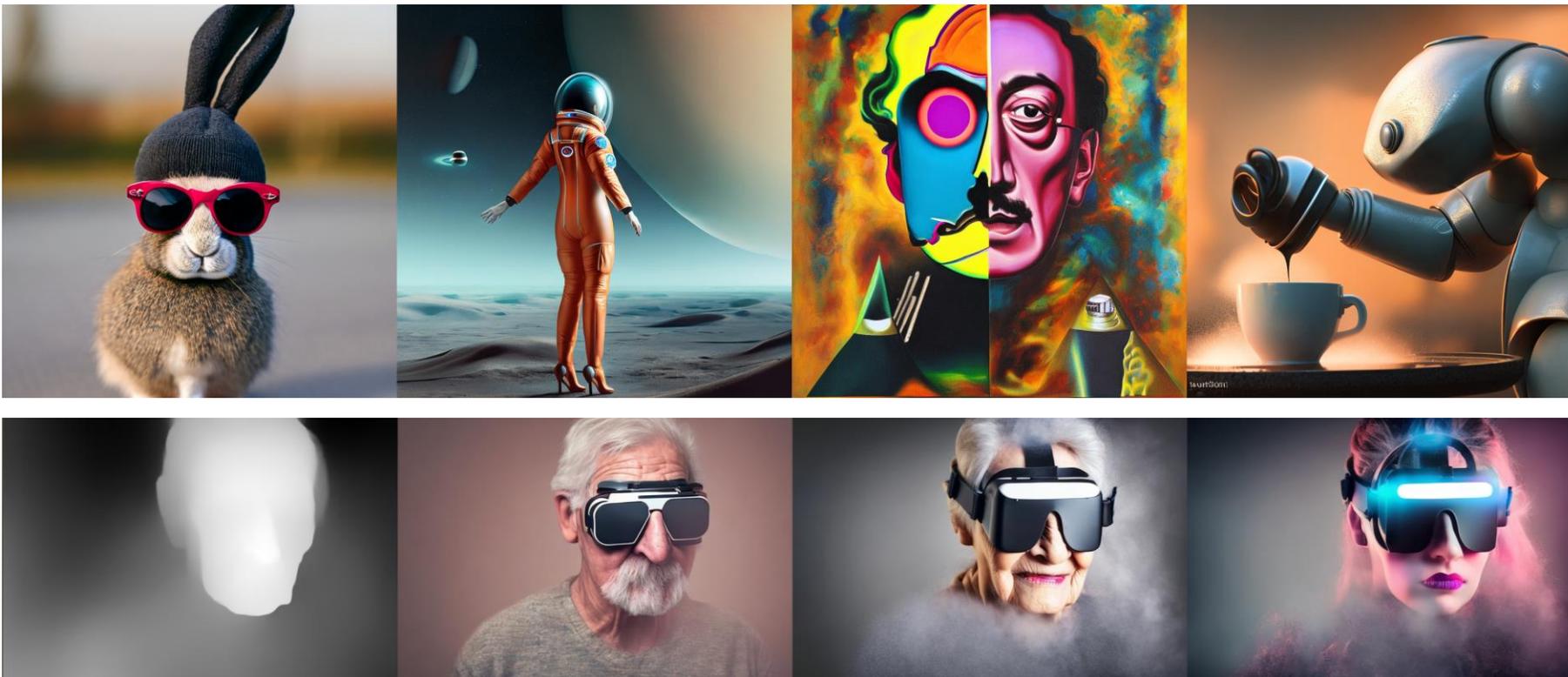
CLIP
2021

DALLE 2
2022

Imagen
2022

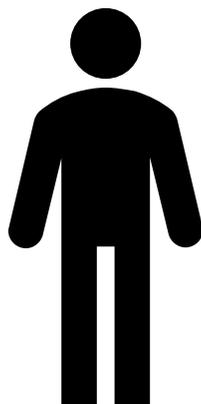
LDM
2022

- 基于其技术的大规模项目：Stable Diffusion



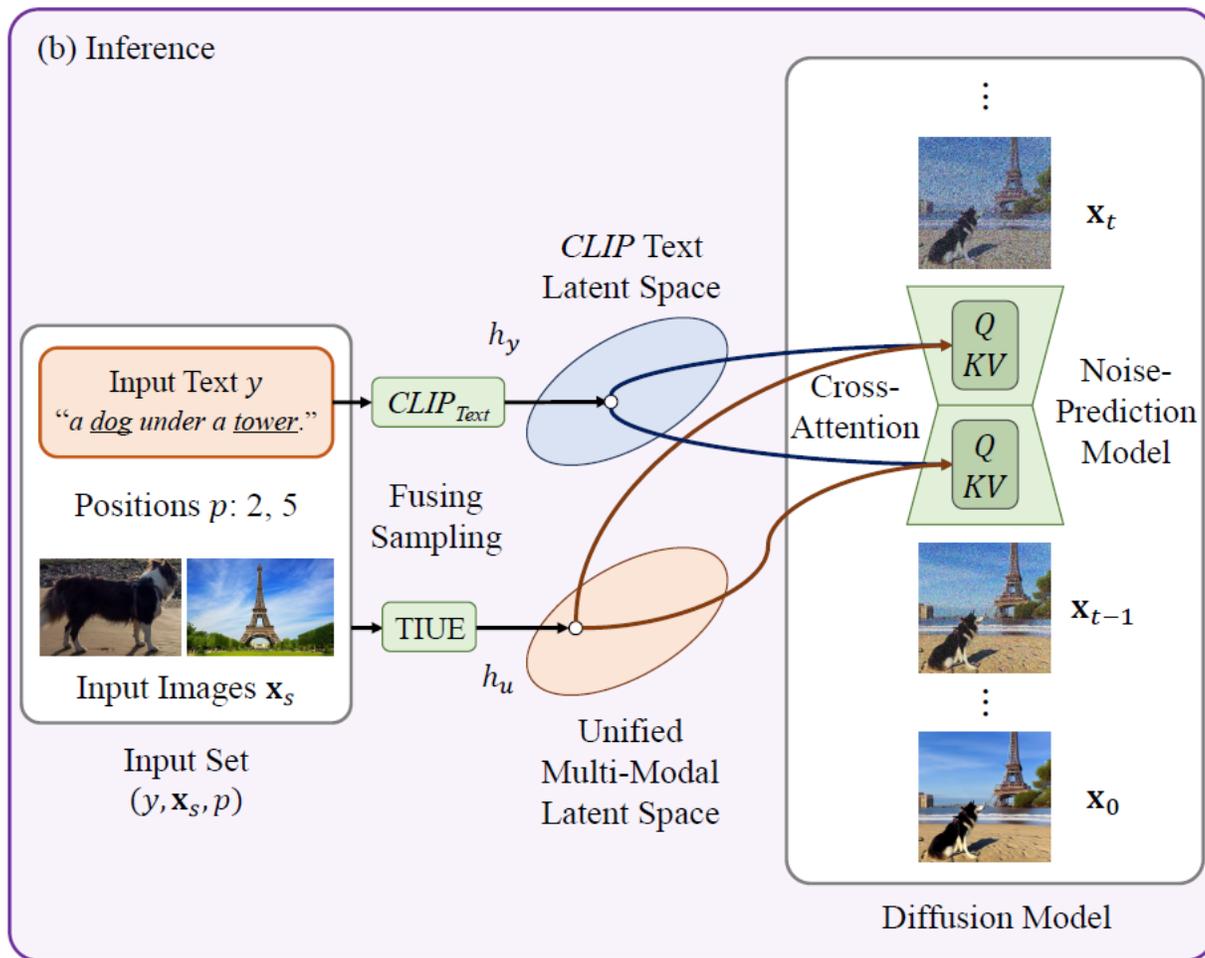
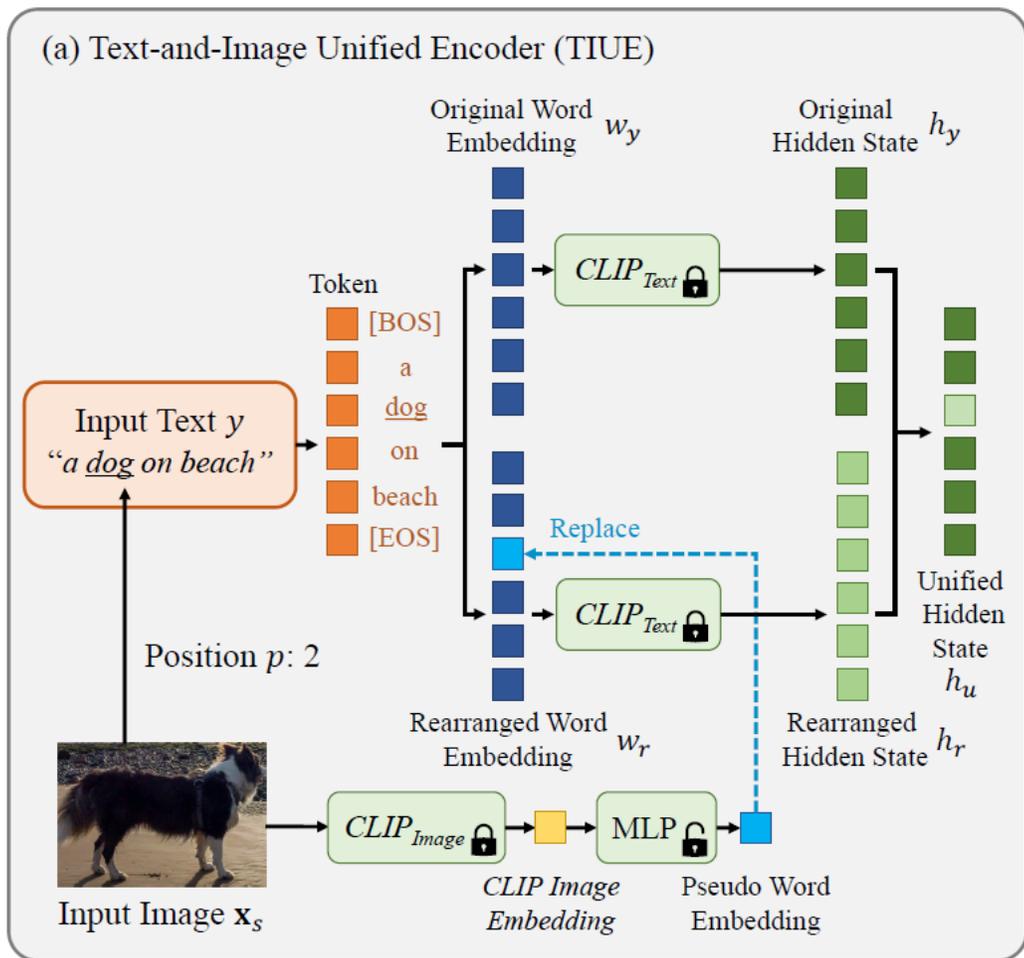
研究目标

- 文到图模型：定制化程度有限

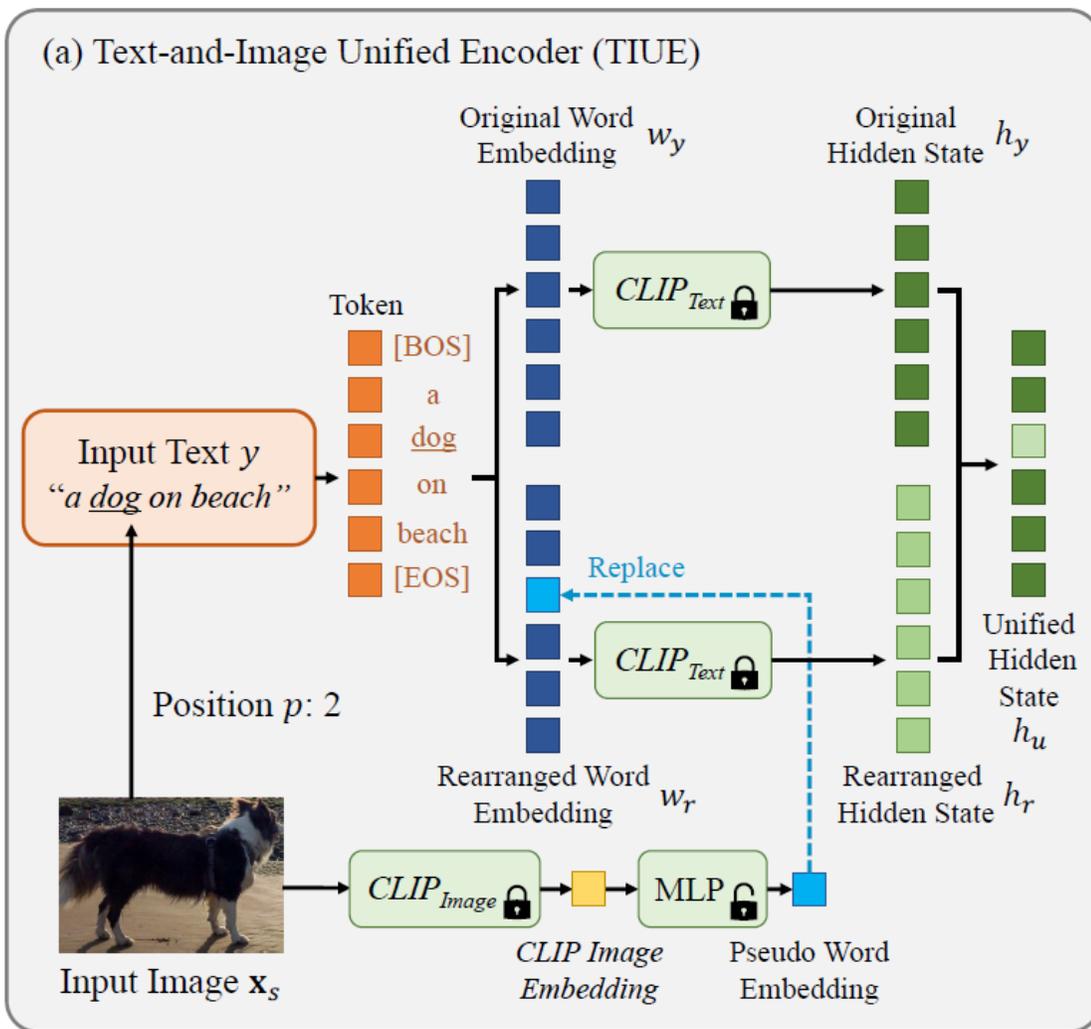


小狗、棕白相间、
白色脖子、棕色尖耳、
.....

定制主体的多模态引导生成



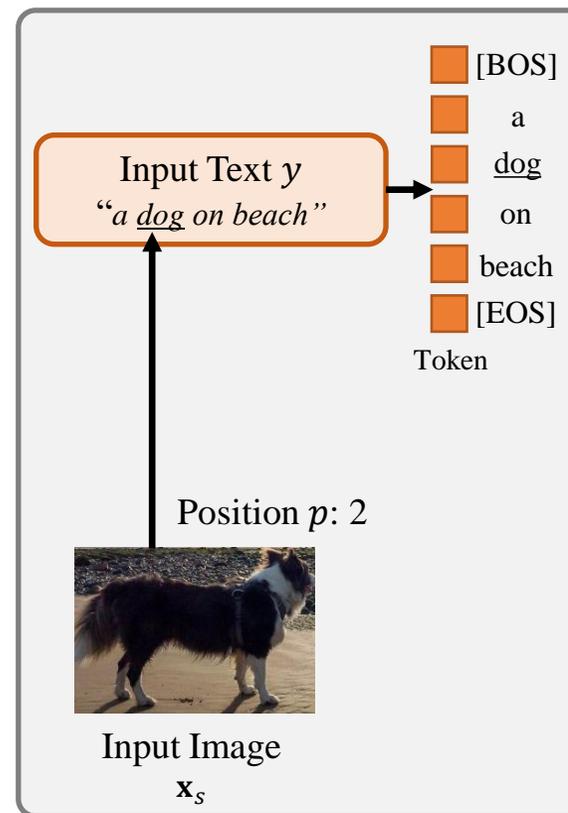
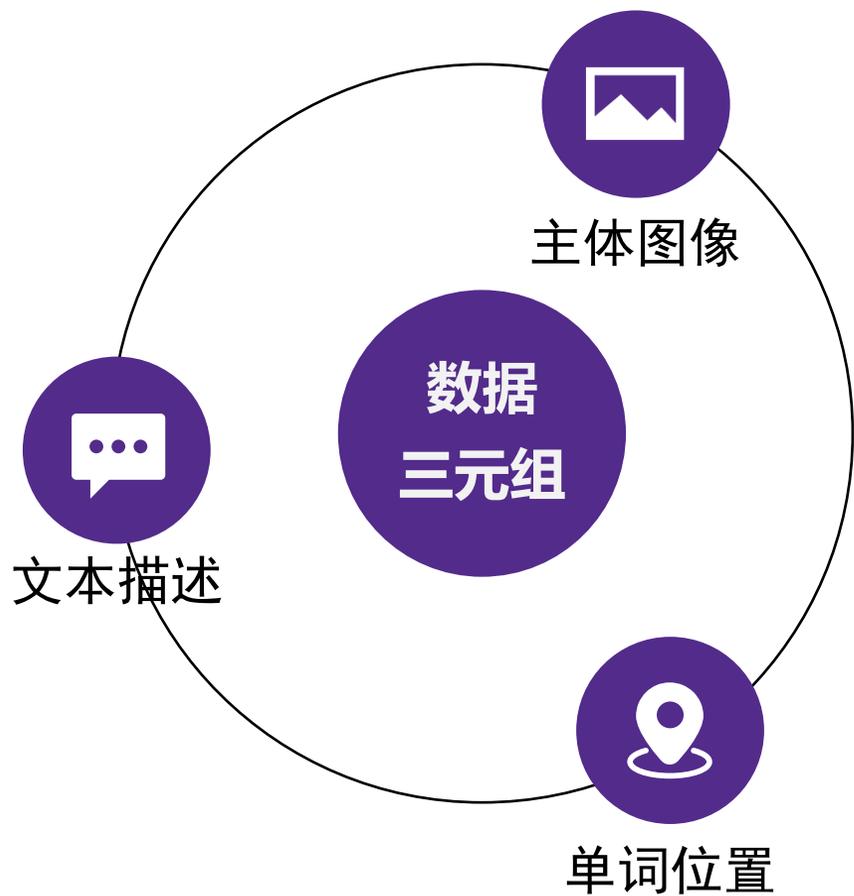
多模态联合编码器



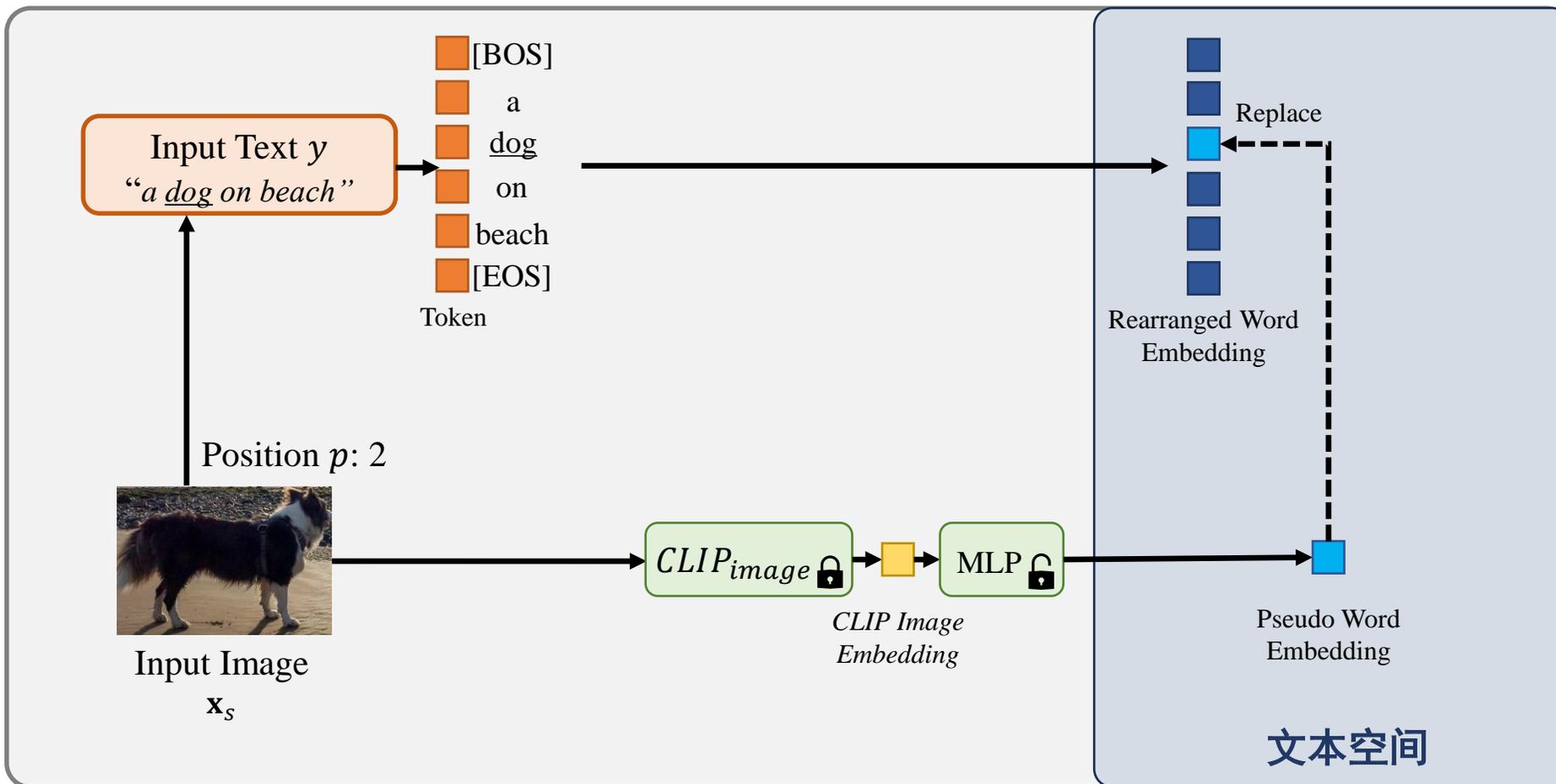
统一文本+主体编码

统一语义隐空间

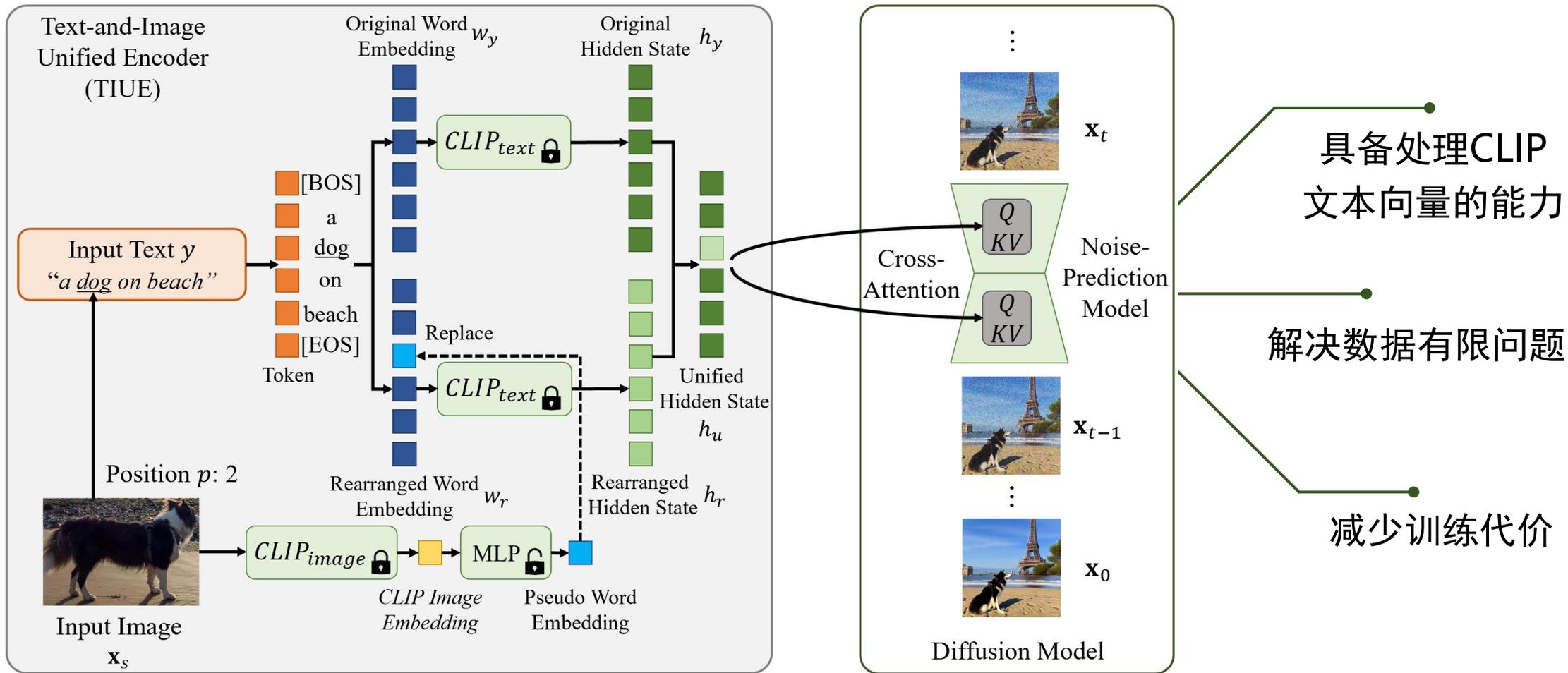
训练数据



统一编码至文本空间



使用预训练文到图模型



定制主体的多模态引导生成

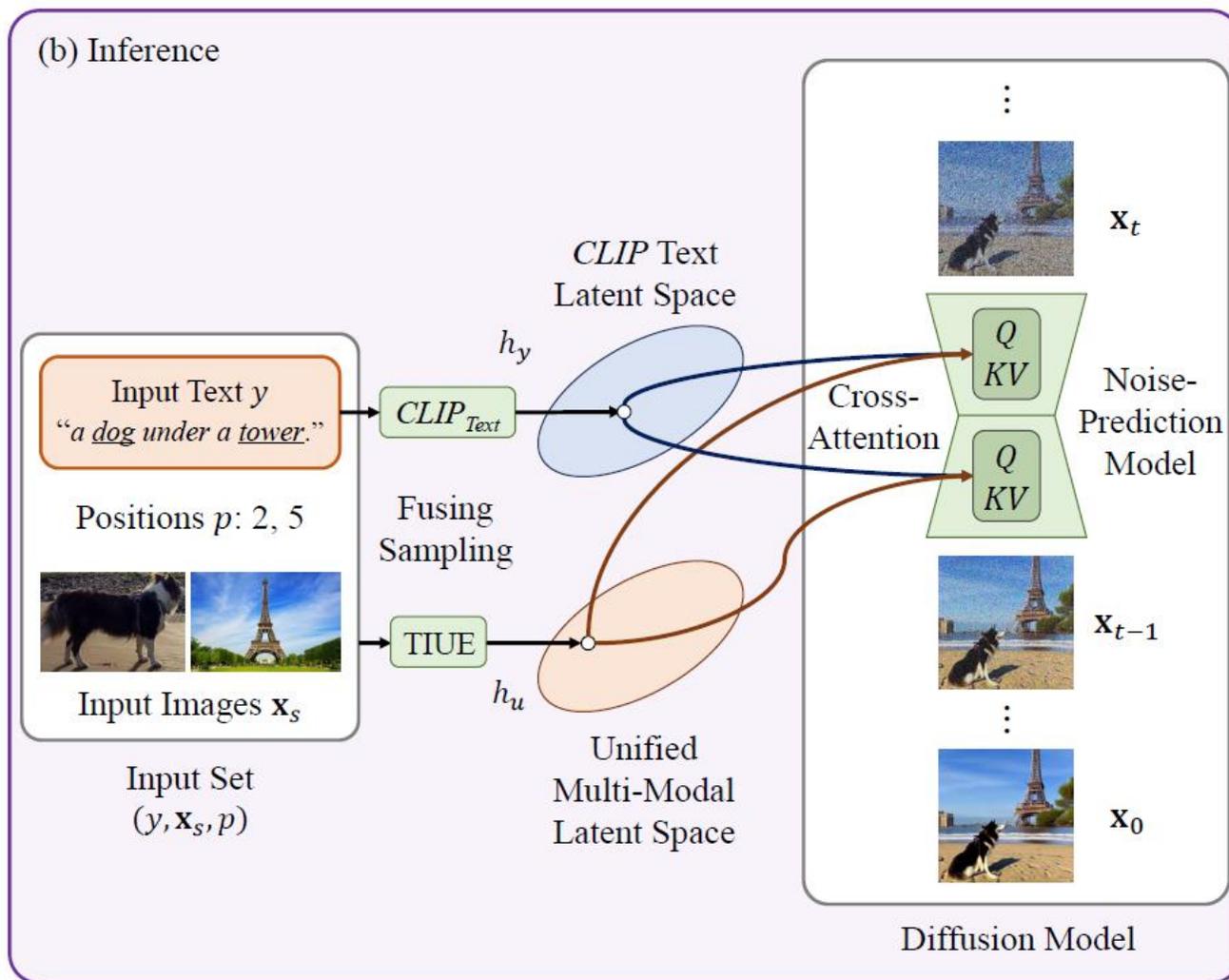
• 挑战:

主体图像的背景影响生成

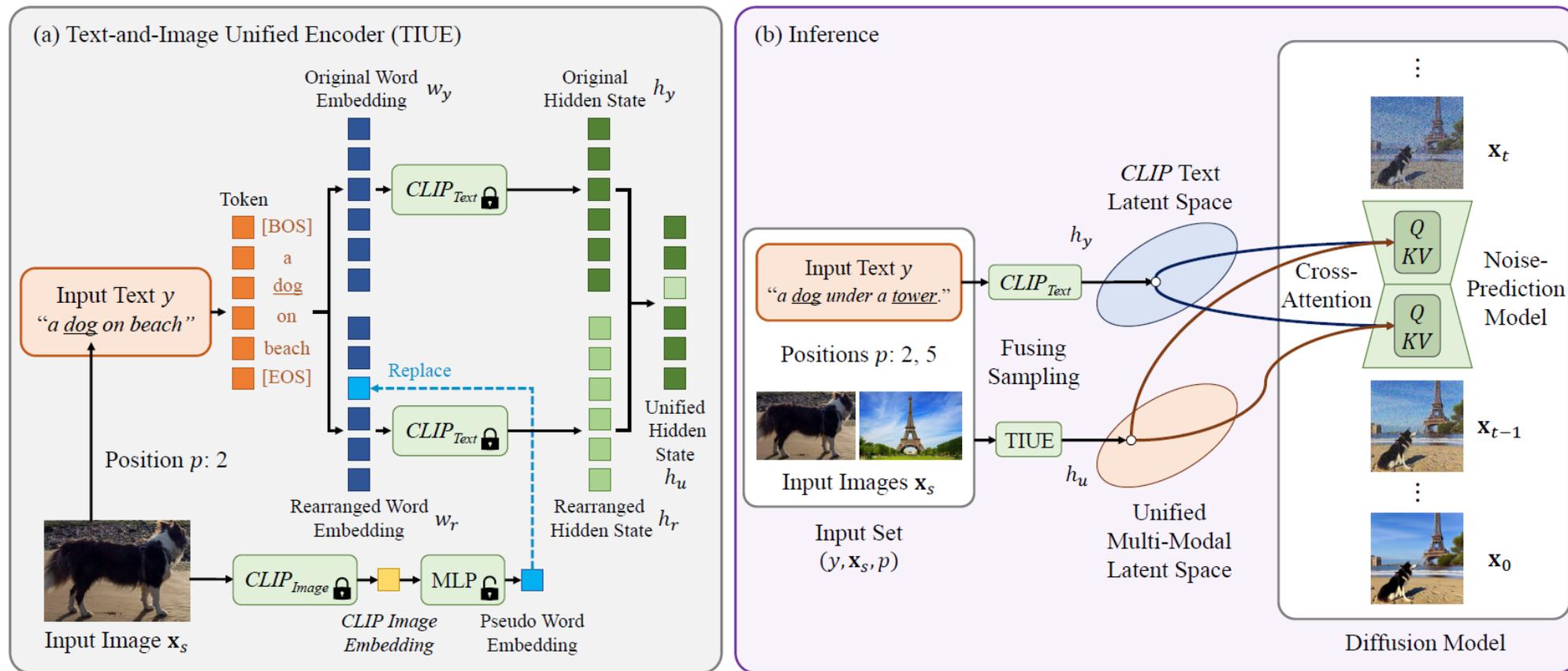
• 方案:

采样时混合引导

$$\hat{\epsilon}_f = \alpha \hat{\epsilon}_u + (1 - \alpha) \hat{\epsilon}_y$$



定制主体的多模态引导生成



无需任何采样时的参数模型精调

实用、高效的主体定制化生成

实验结果



input image



a car running in front of the Eiffel Tower in the winter.



a car running on the road with forest in the background.



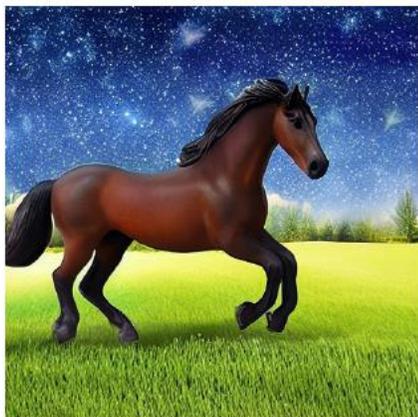
a car on the snow.



a car with a mountain in the background under sunny sky.



input image



a horse running on grass in front of trees under a starry sky with galaxies.



a toy horse on a woody table.



a shirt with a horse on it.



an astronaut is riding a horse, plastic material.

实验结果



input image



a car running in the countryside at sunset.



a car on a bridge above the ocean in the afternoon.



a green car in front of the Big Ben.



a cat is playing with a toy car.



input image



a dog on top of snow with fuji mountain in the background.



a dog is next to a container on the beach.



a dog on grass with the ocean at sunrise.



a dog on the glacier wearing a hat.

实验结果



input image



a bin on the floor with flowers in it.



a bin robot made of the material of the bin.



a bin tower on grass.



a sculpture of a bin made of plaster.



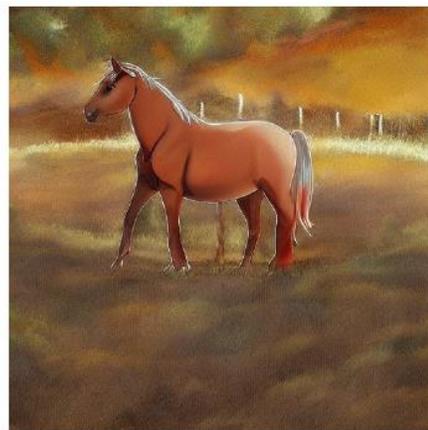
input image



a horse eating grass with snow mountain in the background.



Egyptian pyramids with a horse in saddle in front of them.

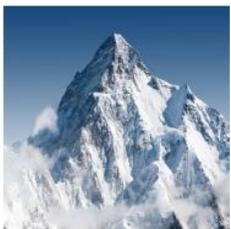


an anime painting of a cute horse.



a blanket with a horse on it.

实验结果



input image



a dog in front of a mountain.



a house in the mountain
with a chimney.



forest in the mountain
at night with illumination.



a painting of the mountain
hanging on the wall.



input image



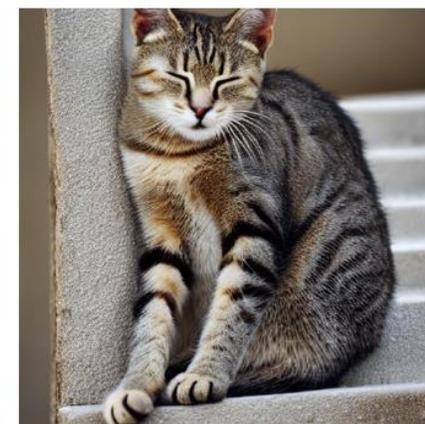
a cat playing on the beach at
sunset.



a cat standing on the cliff
with a volcano in the
background.



a cat on the wood floor
indoor with dim sunshine.



a sleeping cat on stairs.

实验结果



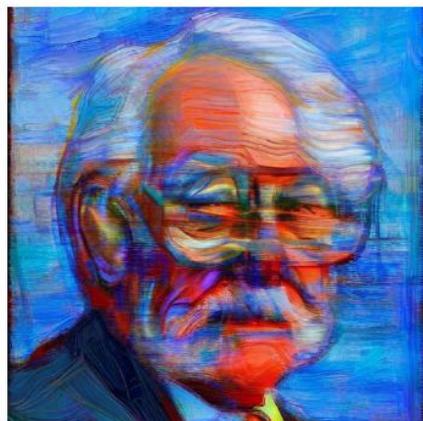
input image



a painting of an open country, in the style of the painting.



a painting of a city at night, in the style of the painting.



a portrait of an old man, in the style of the painting.



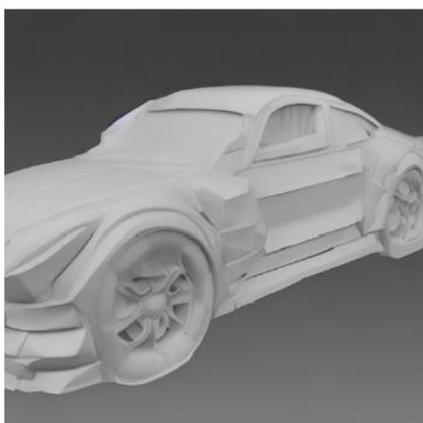
a painting of a car, in the style of the painting.



input image



a sculpture of a dog made of the material of the sculpture.



a sculpture of a car made of the material of the sculpture.



a sculpture of a city made of the material of the sculpture.



a sculpture of a man made of the material of the sculpture.

实验结果



input image



a car is next to another car on a poster.



input image



a dog is under a tower.



实验结果

input image	a <u>cat</u> under a tree in the summer at dusk.	a <u>cat</u> in front of acropolis.	a <u>cat</u> on the grass under a starry sky with galaxy in the early morning.	a <u>cat</u> with a grand canyon in the background in a cloudy day.
	<p>Ours</p> 			
	<p>DreamBooth</p> 			
	<p>SD Origin Caption</p> 			
	<p>SD Detailed Caption</p> 			



STRUCT @ PKU

Spatial and Temporal Restoration, Understanding and Compression Team



智能媒体计算专题论坛

主 办：北京大学王选计算机研究所数字视频研究室
承 办：AI TIME
协 办：东浩兰生(集团)有限公司

Panel讨论

科研生活的“预训练”
“任务驱动”与“多模态”

Panel 嘉宾



李泽超

南京理工大学
计算机科学与工程学院 教授



高盛华

上海科技大学 研究员



陈智能

复旦大学 青年研究员



严骏驰

上海科技大学
信息学院 副教授



杨文瀚

鹏城实验室 副研究员

话题一

上午报告中您印象深刻的工作，
以及您对该方向的一些建议

话题二

你的小组科研经历？
你的 STRUCT “科研预训练” 的
特性与故事？

话题三

导师如何引导科研的
“任务驱动”？ From scratch,
finetuning or prompt?

话题四

谈谈你觉得最有收获或印象最深
的一次“任务驱动”经历

话题五

从 CNN 到 Transformer 到大模型，
浪潮袭来，科研人员应该站在浪潮之
巅还是坚持坐稳自己的冷板凳？

话题六

大模型变革下的科研压力 与就业焦虑

话题七

大模型，知识爆炸下导师焦虑吗？
如何在任务驱动下寻求发展，
分享更多成长经验？

话题八

如何制定合适的“任务”？当你和合作者对“任务”的目标函数认知不一致时，你是如何解决和应对的？

话题九

你在北大的“多模态”
学习 / 生活？

话题十

当“预训练”，“任务驱动”与
“多模态”发生冲突的时候，
该如何平衡与抉择



Demo 展示

北京大学 STRUCT 团队

Cross-Modality Generation

Lowlight Understanding

Action Recognition

Lowlight Enhancement

Super Resolution

Derain

Configuration

Input Source

 File Camera

Input Image



Text description

Begin Generate

Output

Generated Image



Example Image



Paper Information

Yiyang Ma, Huan Yang, Bei Liu, Jianlong Fu, Jiaying Liu. "AI Illustrator: Translating Raw Descriptions into Images by Prompt-based Cross-Modal Generation". ACM Multimedia, Lisbon, Portugal, October 2022.



北京大學
PEKING UNIVERSITY

Q&A