OpenMMLab

# 通用视觉框架OpenMMLab

# 第7讲 底层视觉与MMEditing (下)

吕健勤 教授

2021年5月

# 内容概要

## 本节内容：

- 图像转译 Translation
  - Pix2pix 模型
  - CycleGAN 模型
- 图像补全 Inpainting
  - Context Encoders 模型
  - Global & Local 模型
- 抠图 Matting
- 实践 MMEditing 2

## 上节回顾：

- 图像超分辨率
- 视频超分辨率
- 使用 MMEditing 完成超分辨率任务

# 图像转译

Image-to-Image Translation

# 什么是图像转译

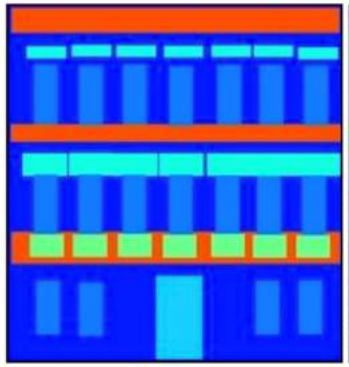
保持图像内容不变，将图像从一种形态转换为另一种形态

Labels to Street Scene

input

output

Labels to Facade



input

output

BW to Color

input

output

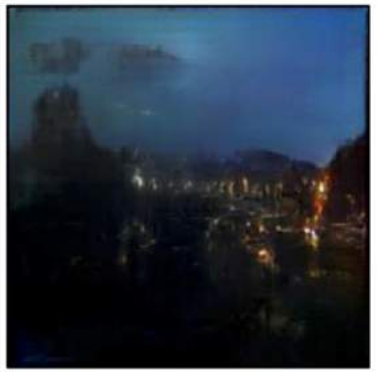
Aerial to Map

input

output

Day to Night

input



output

Edges to Photo

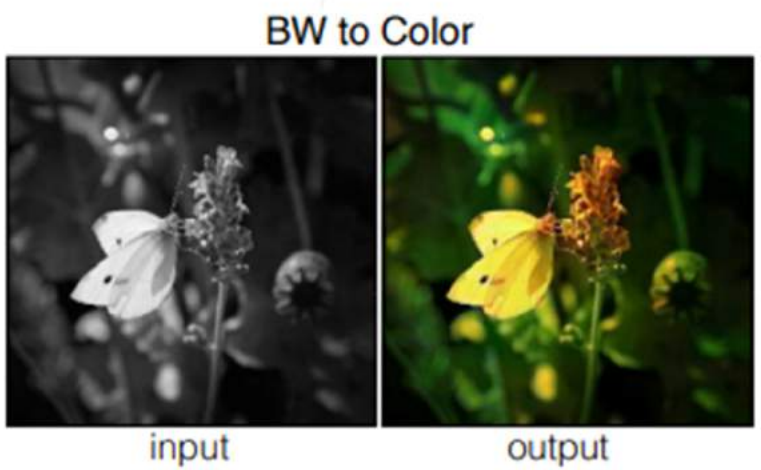
input

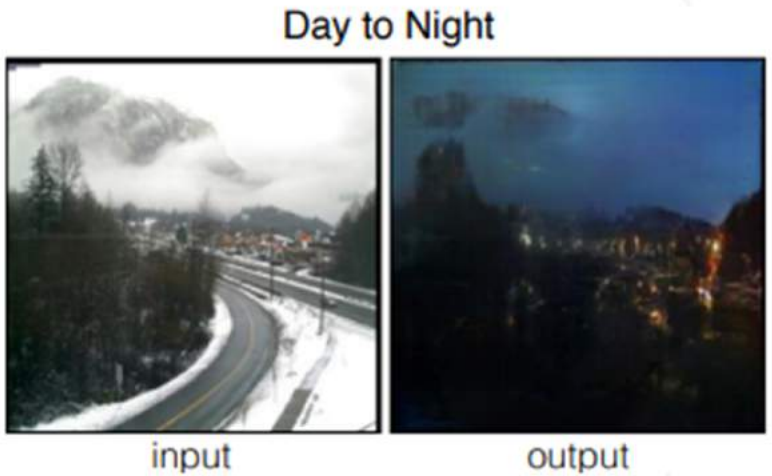
output



Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." Proceedings of the IEEE international conference on computer vision. 2017.

# 应用：相片特效




Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." Proceedings of the IEEE international conference on computer vision. 2017.

# 应用：卡通人像

Landmark Assisted CycleGAN for Cartoon Face Generation

# 应用：航拍图像生成地图

航拍图像

自动转译

地图

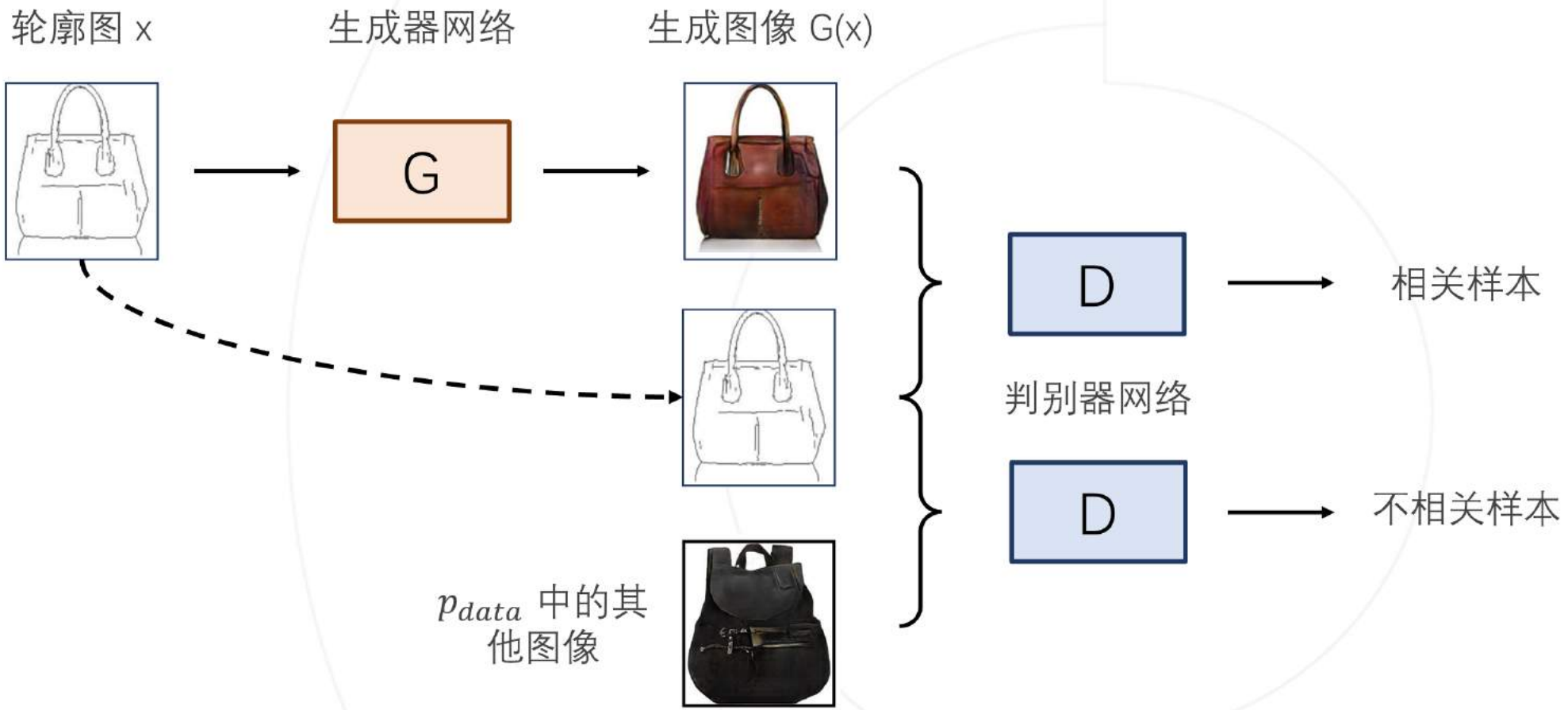
# 基本思路

**思路**：使用 GAN 模型，生成器完成图像转译，判别器判断生成图像为真实或合成

# 基本思路

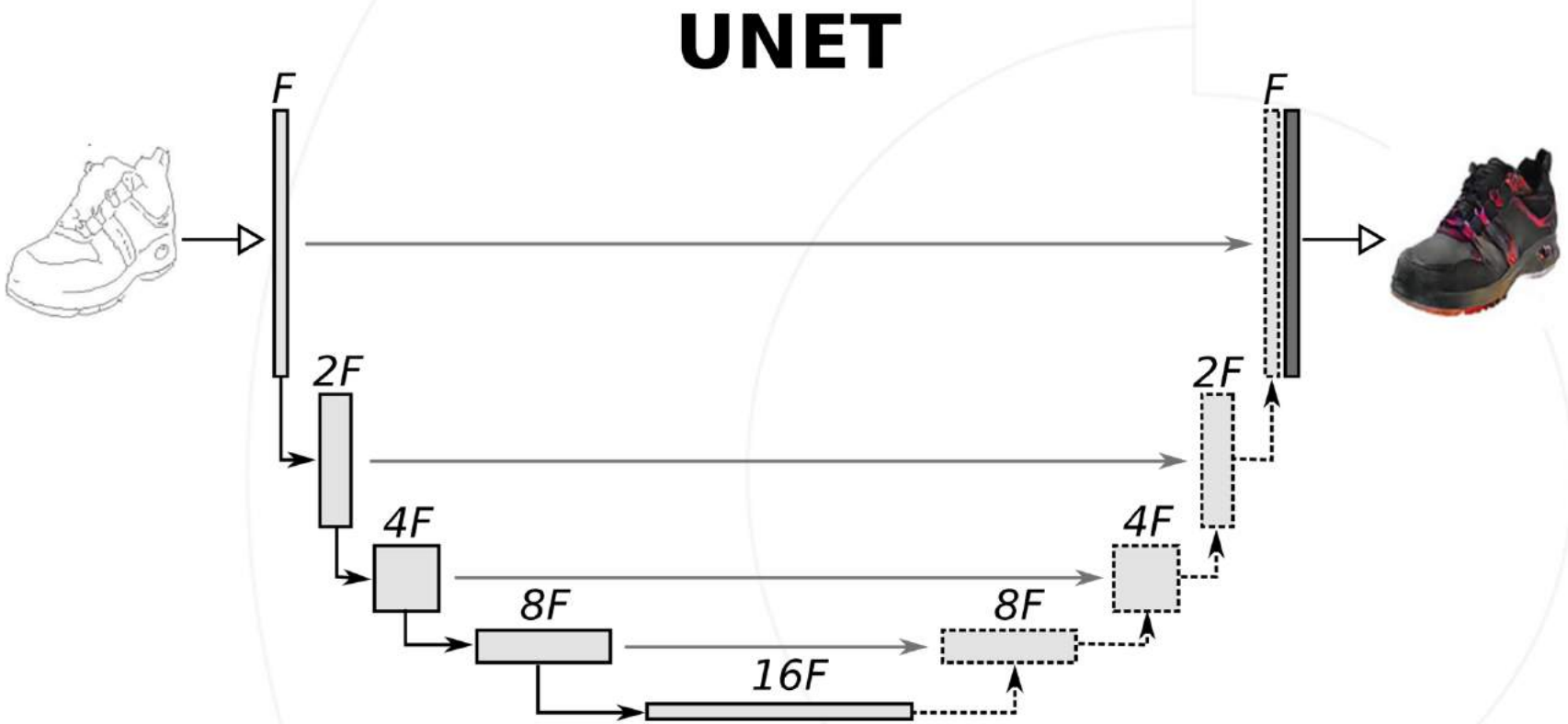
**问题**：普通 GAN 模型中的生成器，只能判断生成图像是否符合数据分布 $p_{data}$，不能判断生成图像是否与输入图像相关联。

**解决思路**：改造判别器网络，使之接受两个图像为输入，判断二者是否相关联



# Pix2Pix 的生成器网络

Pix2Pix 使用 UNet 作为生成器网络，可以让输入和输出共享一些底层特征





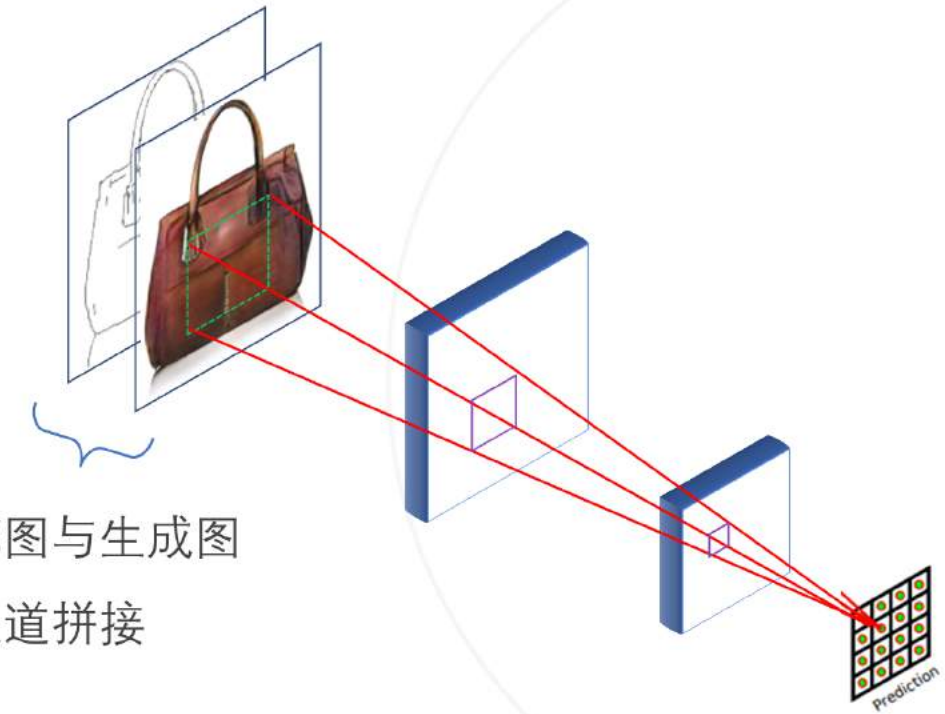

图片来源：https://nchlis.github.io/2019_10_30/page.html

# Pix2Pix 的判别器网络 PatchGAN

Pix2Pix 使用 PatchGAN 作为判别器网络；

PatchGAN 是一个全卷积结构，每个分类输出对应输入图像的某个**局部区域；**

基于Patch而不是全局，可以鼓励产生更好的局部效果，如纹理等。




轮廓图与生成图

按通道拼接

# Pix2Pix 的训练数据

Pix2Pix 使用配对的训练数据

例：一个训练样本包含 { 轮廓图，真实图像 }

对于一些任务，我们可以以很低的成本获取图像对

例：收集常规图像，使用边缘检测算法生成轮廓图

# Pix2Pix 的损失函数

1. GAN 损失函数

$$\mathbb{E}_x[\log D(x, y)] + \mathbb{E}_{x,y}\left[\log\left(1 - D\left(x, G(x)\right)\right)\right]$$

{ 轮廓图，真实图像 }　　{ 轮廓图，生成图像 }

2. L1 或 L2 损失函数，使生成图像尽量与已知的配对图像相近

$$\mathbb{E}_{x,y}\|G(x) - y\|_{1 \text{ or } 2}$$

交替训练 G、D 网络，训练完成后丢弃 D 网络，使用 G 网络完成图像转译任务

# Pix2Pix 的例子

轮廓图 → 图像


手绘图 → 图像


地图 → 卫星图

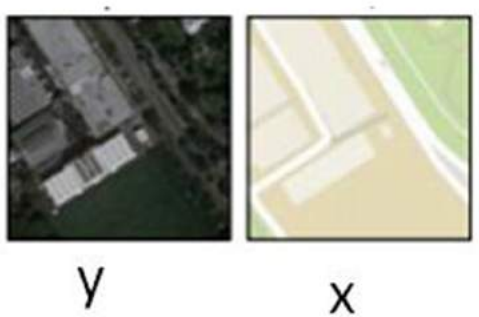
# 配对数据集的问题

Pix2Pix 要求训练数据成对出现，称为**配对数据集**(Paired Dataset)

- 一些转译场景中，采集配对数据的成本很高
  - 例：地图与卫星数据


容易获取

使用算法生成轮廓




采集成本高

需要人工绘制

- 更多转译场景中，几乎不可能采集到配对数据
  - 例：同样形状姿态的马和斑马
  - 例：普通照片与相同内容的油画

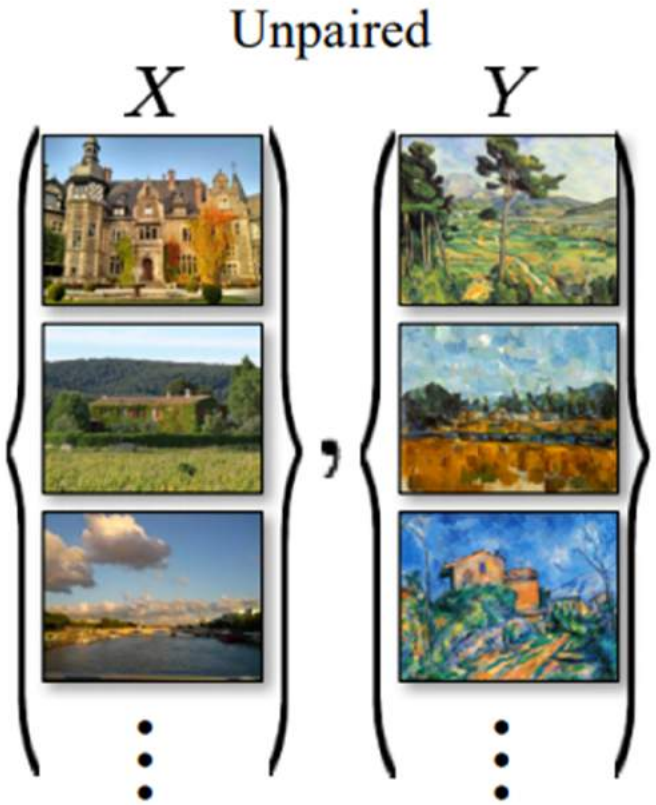
实际生活中

几乎不可能出现

# 非配对数据集

实际场景中，更容易获取的数据是**非配对数据集**(Unpaired Dataset)。

我们可以在两个图像领域分别采集数据，但二者在内容上并没有配对关系




问题：如何使用**非配对数据**训练转译模型?

# 使用非配对数据集 CycleGAN 2017

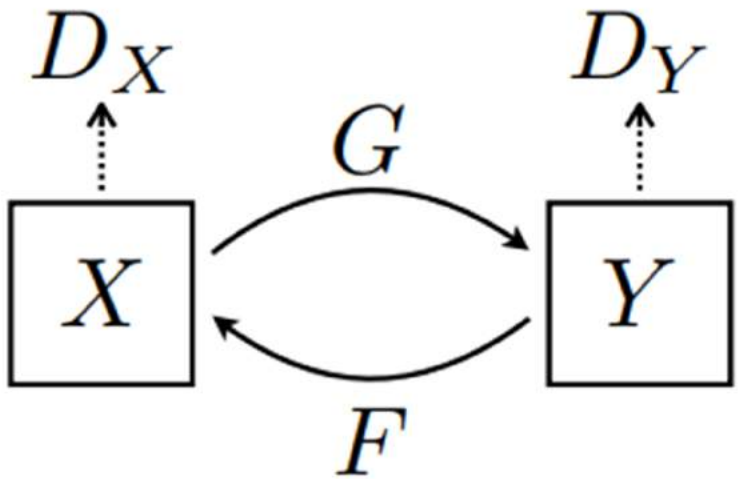
已知图像域 X 和 Y

生成网络 G : X 到 Y 的映射

生成网络 F : Y 到 X 的映射

判别网络 Dx：分辨图像是否符合 X 的分布 Px

判别网络 Dy：分辨图像是否符合 Y 的分布 Py

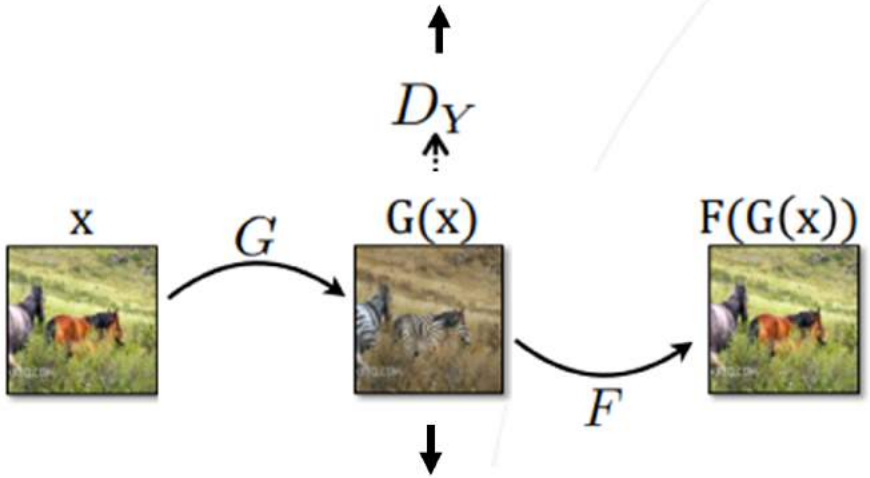





Zhu et al., Unpaired Image-to-Image Translation using CycleConsistent Adversarial Networks, ICCV 2017

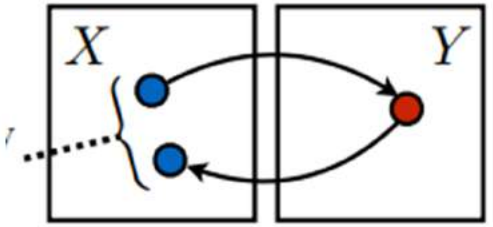
# 循环重构

是否符合 Y 域的数据分布



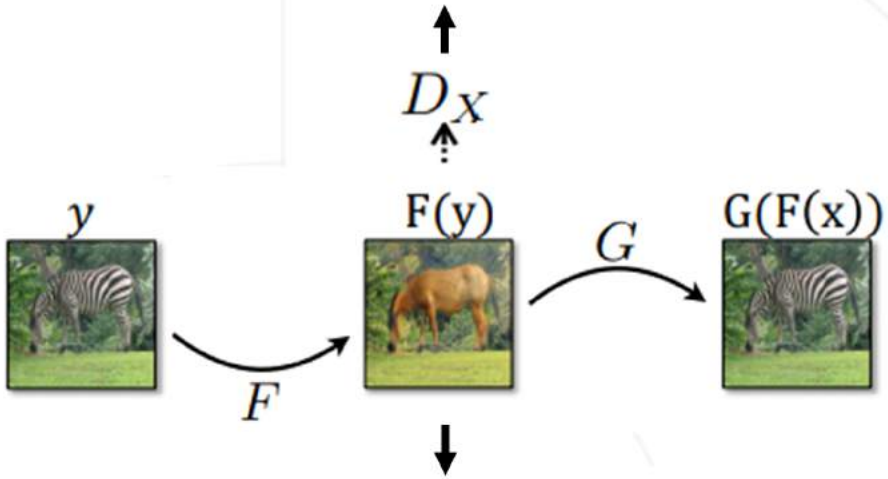

在 Y 域没有对应样本

在 X 域比较
计算重构损失

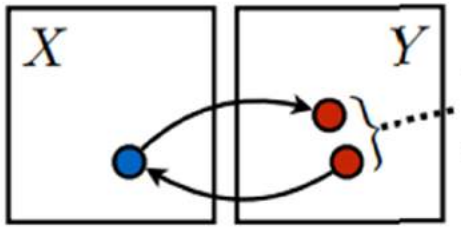



$$\|F(G(x)) - x\|_1$$

是否符合 X 域的数据分布




在 X 域没有对应样本




在 Y 域比较
计算重构损失

$$\|G(F(y)) - y\|_1$$

# CycleGAN 的损失函数

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) + \mathcal{L}_{\text{GAN}}(F, D_X, Y, X) + \lambda \mathcal{L}_{\text{cyc}}(G, F),$$

$$\mathcal{L}_{\text{cyc}}(G, F) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1].$$

对抗训练损失 GAN Loss

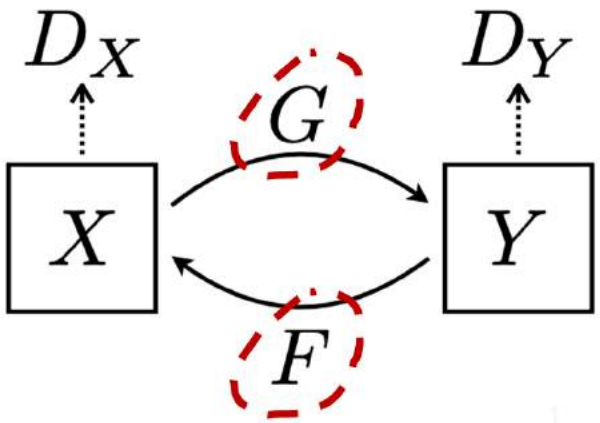
鼓励 G 和 F 生成的图像符合目标域的图像风格

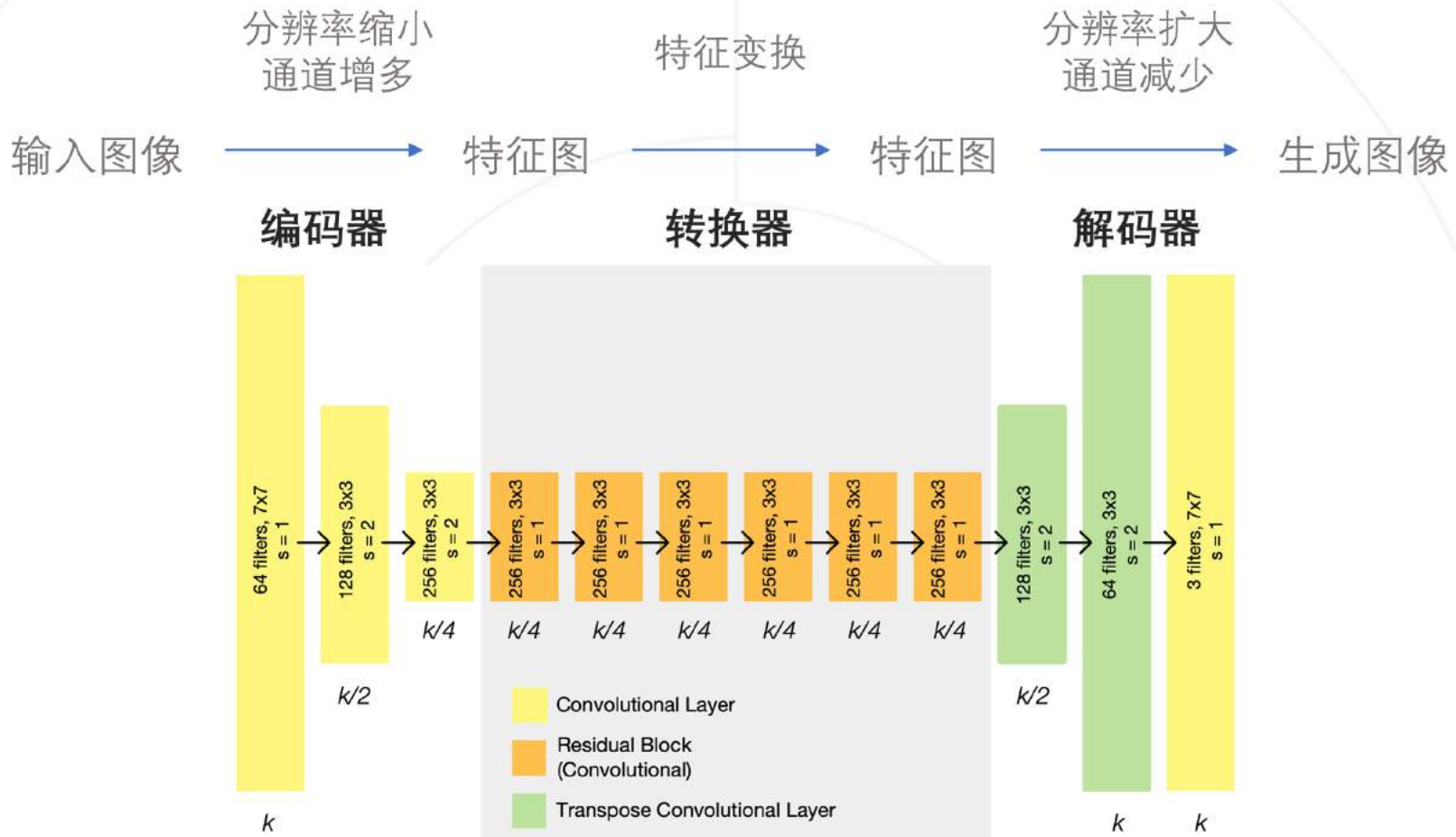
循环重构损失 Cycle Consistency Loss

鼓励 G 和 F 生成与输入图像**相对应**的样本

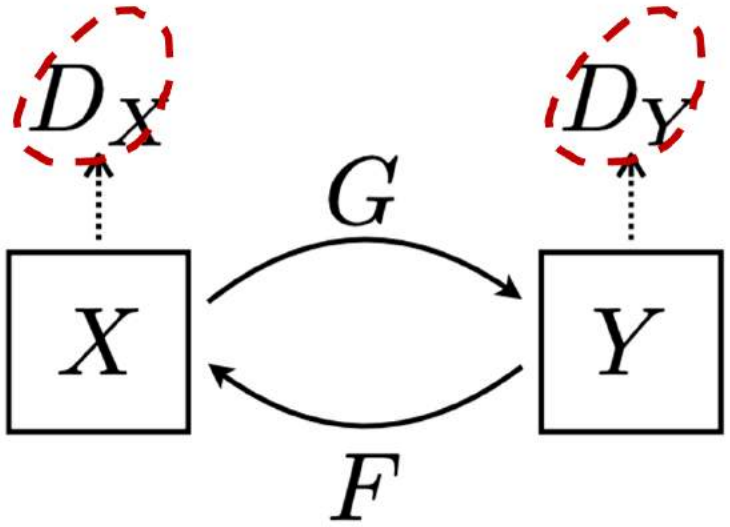
# CycleGAN 生成网络的结构

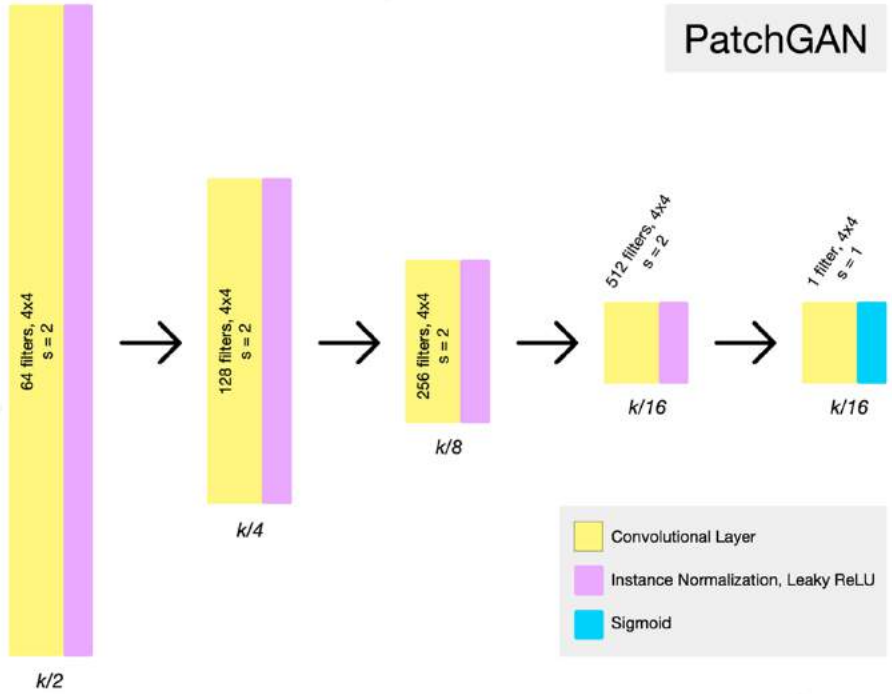
生成器采用 bottleneck 结构



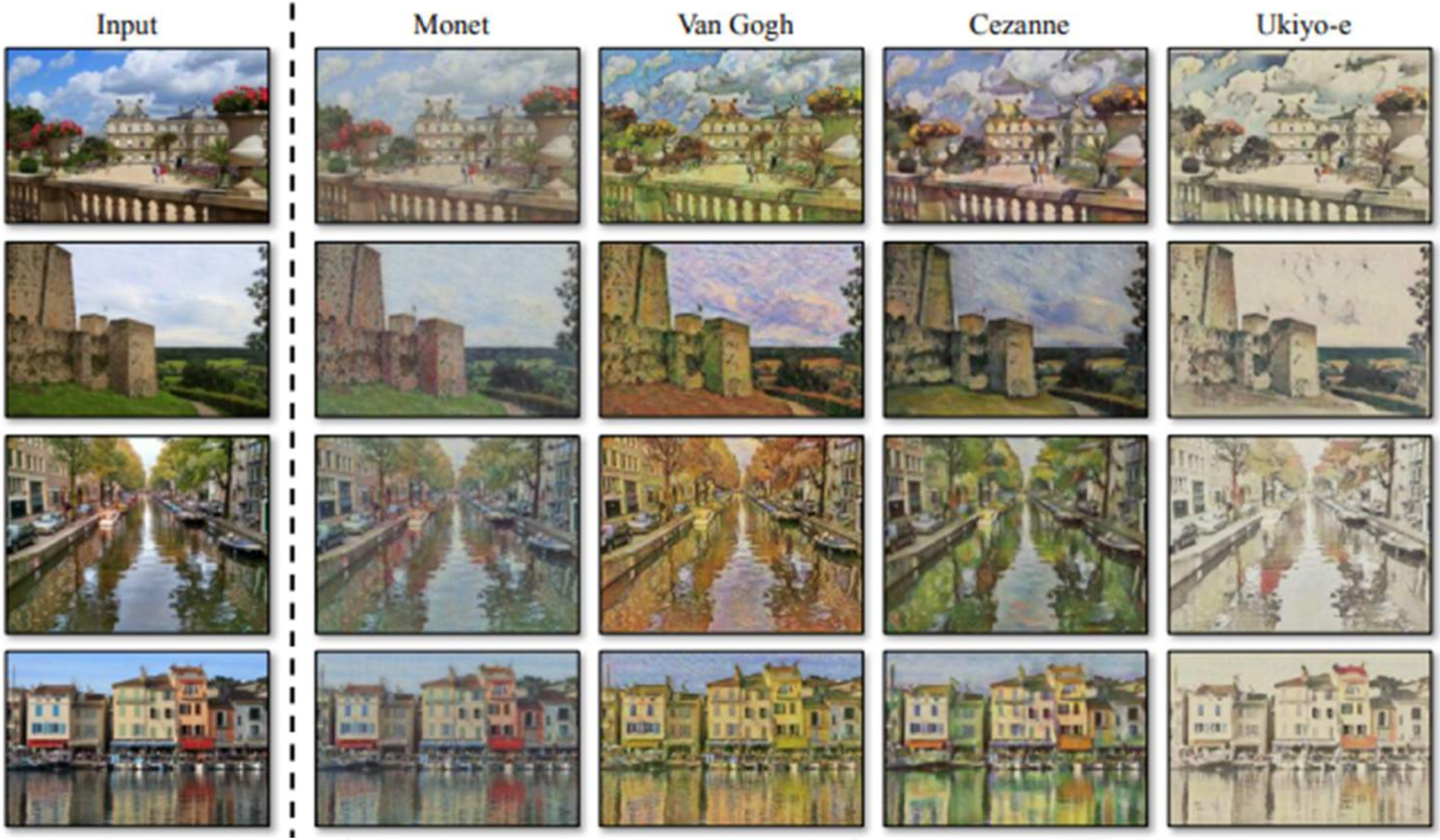
# CycleGAN 判别网络的结构




与 Pix2Pix 相同，使用 PatchGAN 结构，基于局部图像判断真实/生成

# CycleGAN 的效果：风格变化



# CycleGAN 的效果：季节变换

winter Yosemite → summer Yosemite

summer Yosemite → winter Yosemite

# 谢谢大家

OpenMMLab

# 通用视觉框架OpenMMLab

# 第7讲 底层视觉与MMEditing (下)

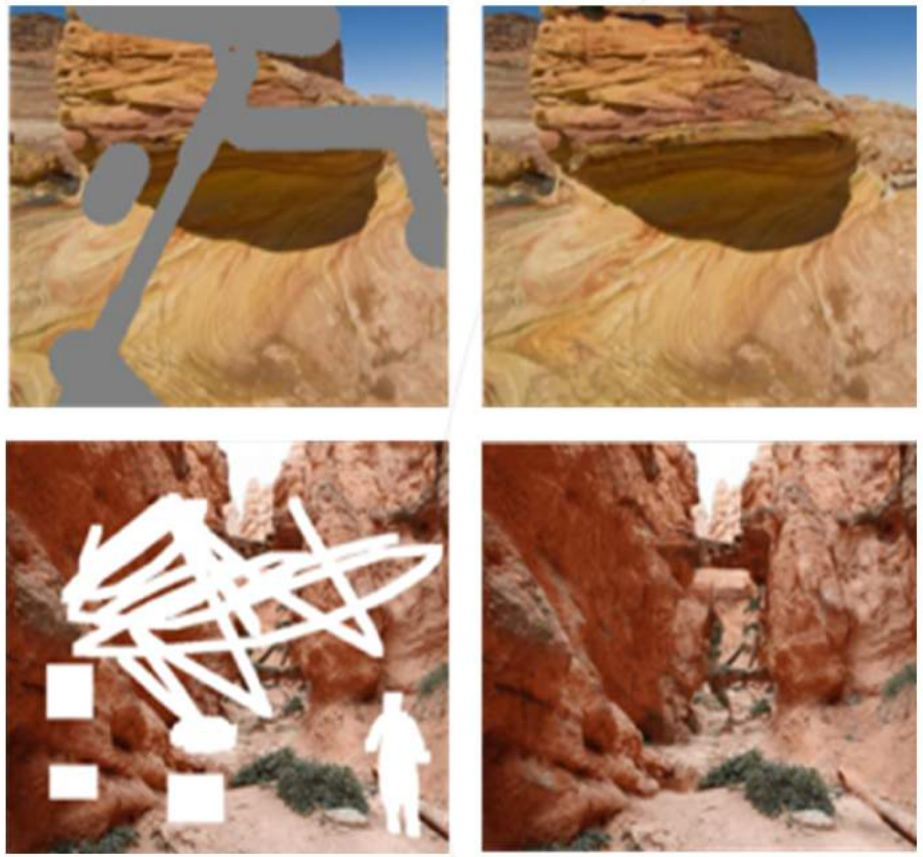
吕健勤 教授

2021年5月

# 图像修复

Inpainting

# 任务描述



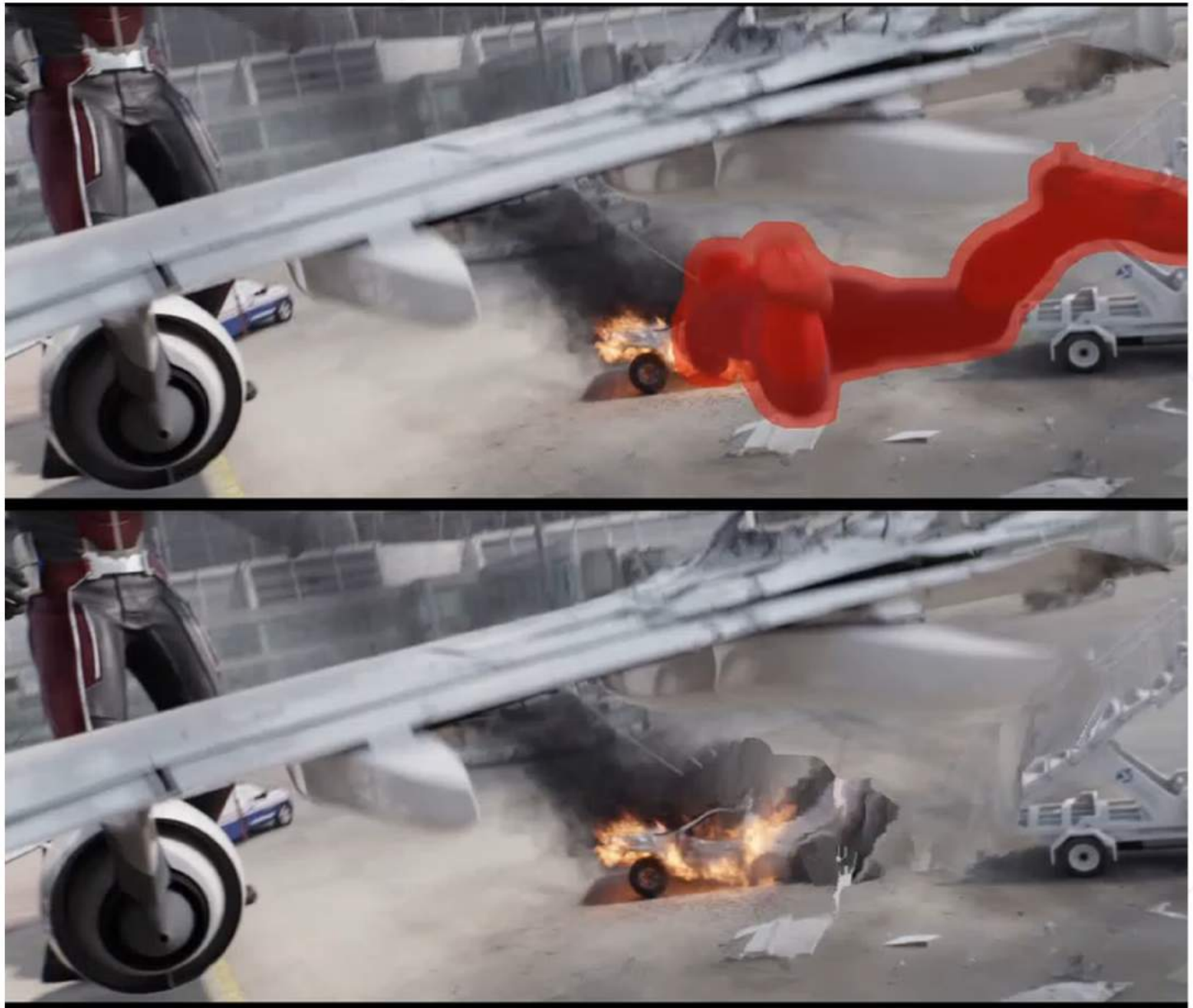
## 目标

修复图像中的受损区域

## 应用

去除水印、消除人像、视频修复

# 应用：视频编辑



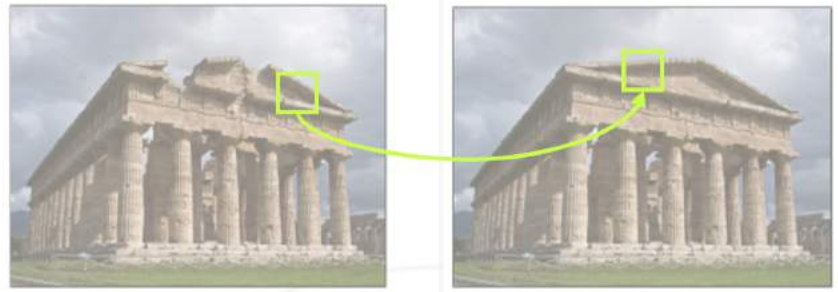
# 图像修复的技术演进

OpenMMLab

**传统视觉**

PatchMatch (2009)

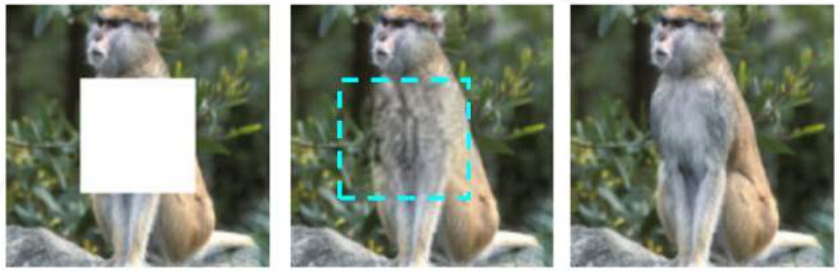
基于区块匹配，从原图上匹配相似的区域进行补全



**深度学习早期探索**

Context Encoder (2016)

使用**编码解码器**结构和**对抗训练**机制，对抗训练只考虑全局，局部恢复效果不佳



(a) Input (b) CE (c) GT

**更多样的深度模型更好的恢复效果**

Global & Local (2017)

在 CE 的基础上加入局部的对抗训练，获得较好效果

DeepFill (2018) v2 (2019)

Pconv (2018)

加入 Attention 机制

单阶段➔双阶段

# 解决思路

**思路**：图像中的内容通常是结构化的，有规律可循的，可以根据已有的部分推测缺失的内容

**难点**：不是定问题 (ill-posed problem)，有无穷多解

缺失的原图

推测缺失的内容

人类艺术家



Pathak, Deepak, et al. "Context encoders: Feature learning by inpainting." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

# 如何利用已有图像部分

传统方法

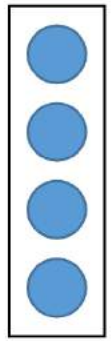
PatchMatch

在原图上寻找相似图块进行修补

借助深度学习

编解码器

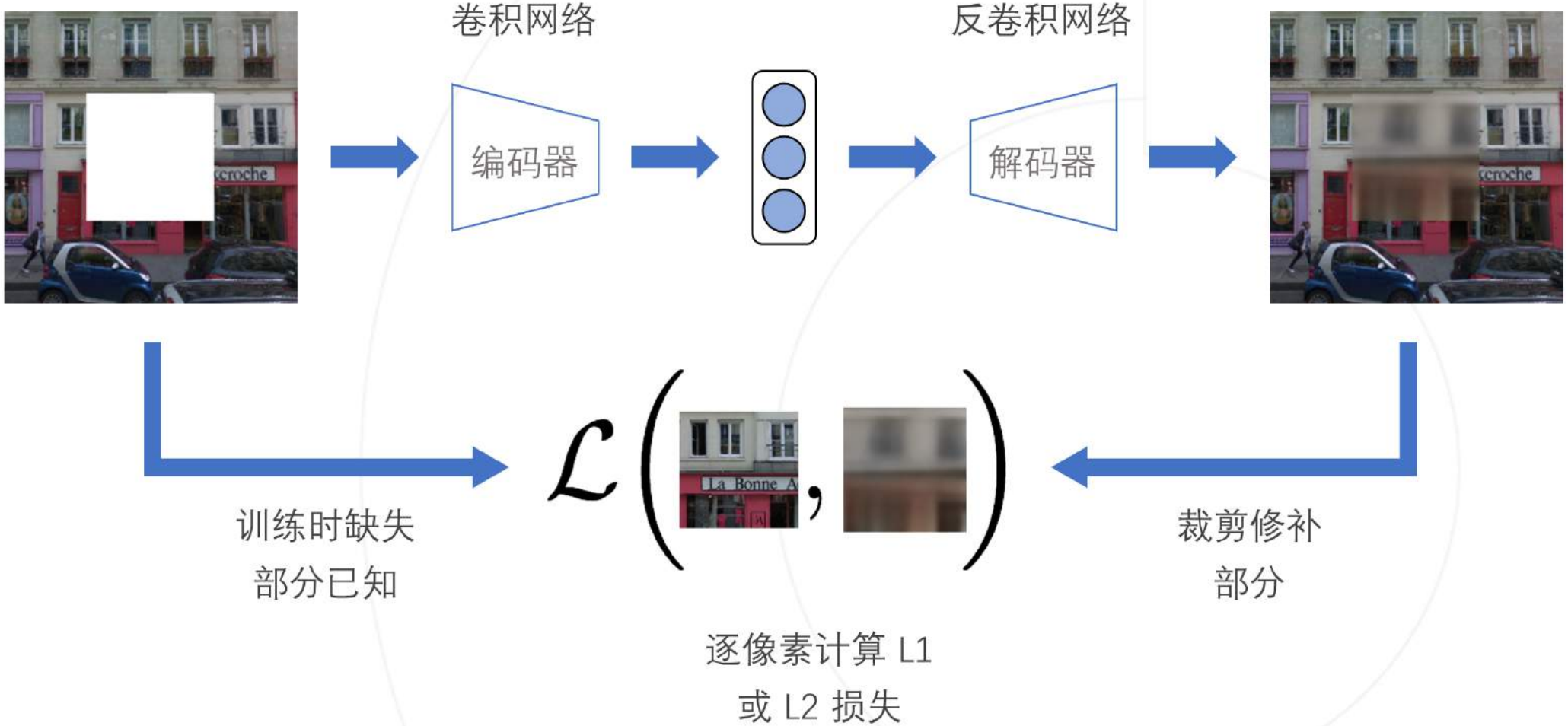
将图像编码成语义特征



从特征恢复完整图像

1. 特征对缺失有鲁棒性
2. 模型包含一定的先验知识

# 编码器与解码器恢复






Pathak, Deepak, et al. "Context encoders: Feature learning by inpainting." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

# 生成更真实的图像

**问题**：L2 损失倾向给出所有合理解的均值，恢复部分相对模糊，高频信息丢失

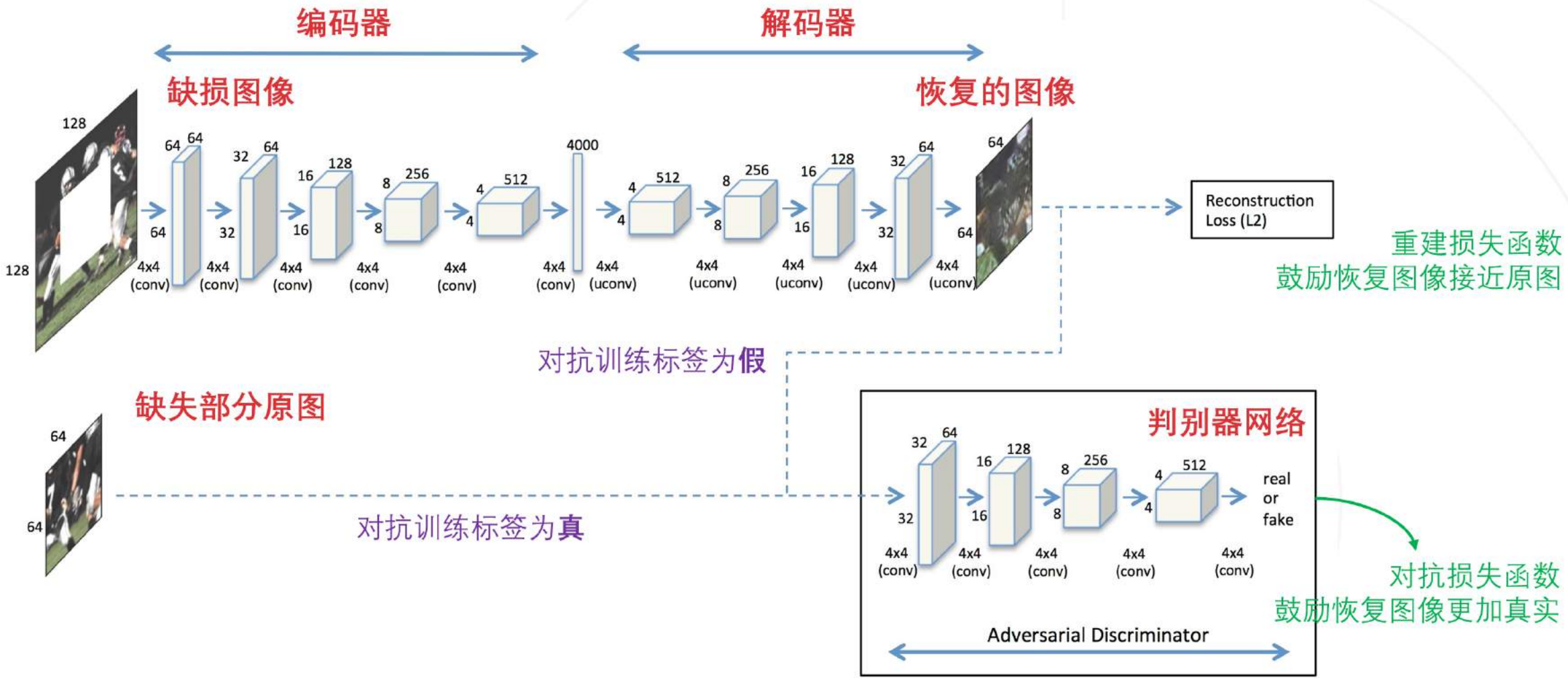



**解决方案**：加入对抗训练，用判别器分辨恢复图像，鼓励编解码器生成更真实的图像



Pathak, Deepak, et al. "Context encoders: Feature learning by inpainting." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

# 完整网络结构






Pathak, Deepak, et al. "Context encoders: Feature learning by inpainting." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

# 对抗训练模型的恢复效果

缺损图像

使用 L2 损失

加入 GAN



Pathak, Deepak, et al. "Context encoders: Feature learning by inpainting." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

# 对抗训练模型的恢复效果



Pathak, Deepak, et al. "Context encoders: Feature learning by inpainting." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

# Context Encoders 的问题

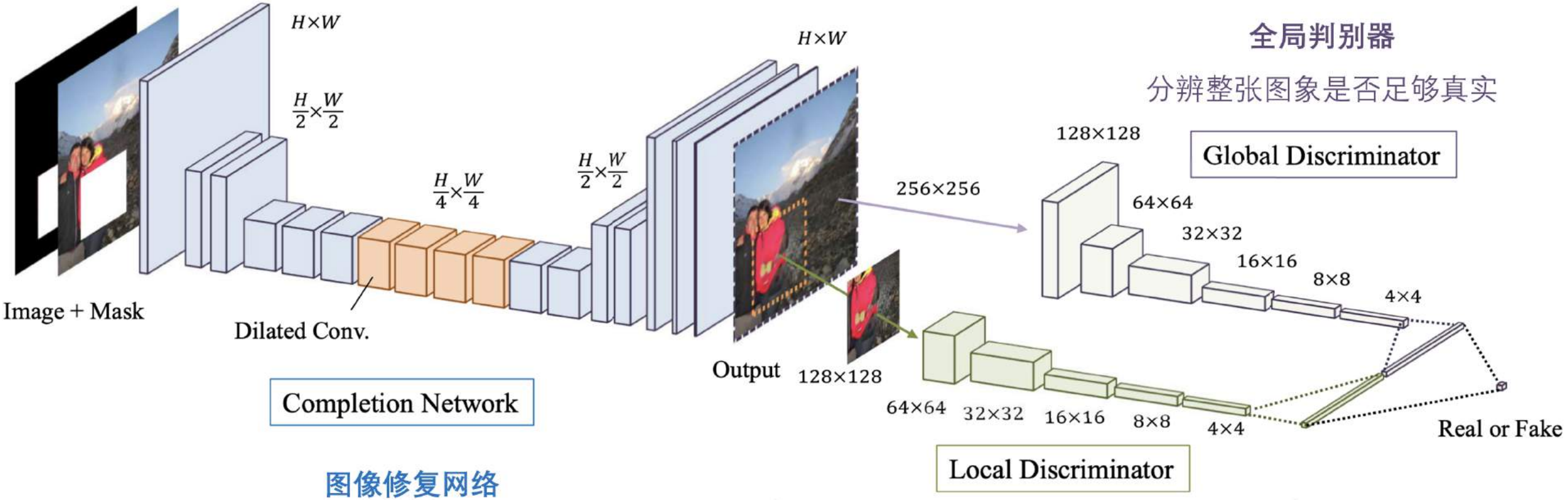
**问题**：Context Encoders 的恢复效果仍然不是很好

**分析**：判别器以全图为输入，局部区域的监督信号不强

**解决方案**：

针对局部区域加入判别器，强化局部区域的恢复效果

# Global & Local 2017




**全局判别器**

分辨整张图象是否足够真实

**图像修复网络**

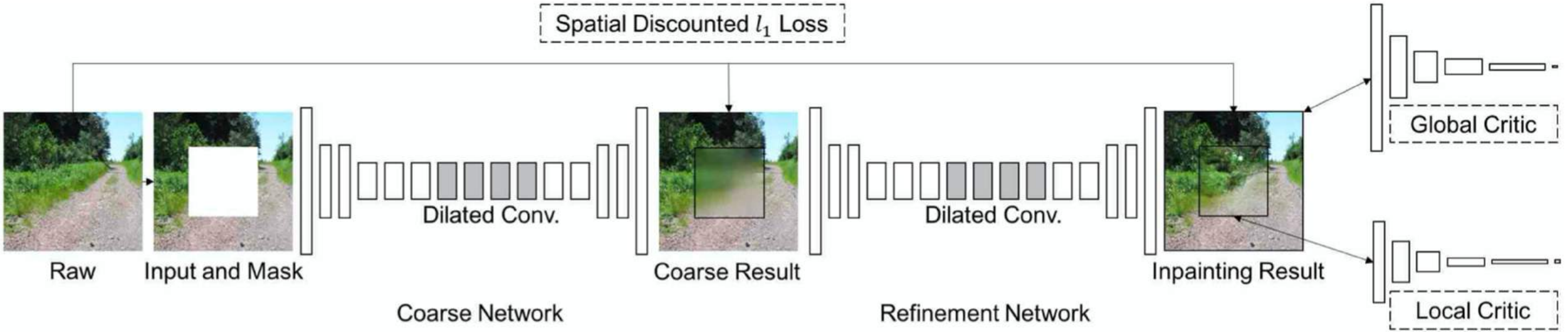
1. 叠加图像和掩膜作为输入
2. 加入空洞卷积捕捉上下文信息

**局部判别器**

分辨恢复区域是否足够真实

# Deepfill 2018

包含粗糙、精炼两个阶段，第二阶段加入 contextual attention，尾部分别使用 Global 和 Local 的判别器




**粗糙网络**

1. 叠加图像和掩膜作为输入
2. 输出初步的修复结果
3. 优化目标是图像重建的pixel维度L1 Loss

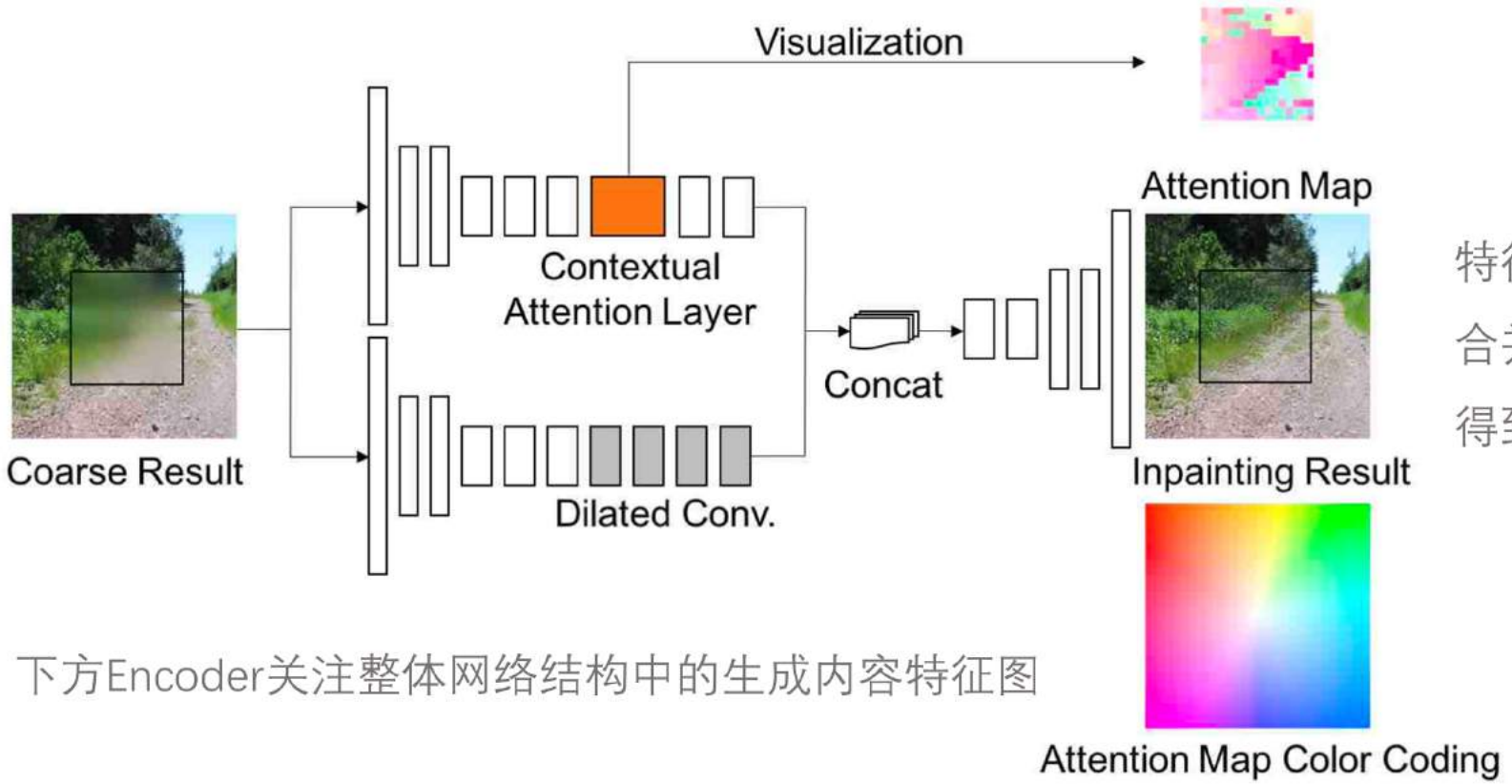
**精炼网络**

1. 进一步扩大感受野，优化修复结果
2. 优化目标包括L1 Loss以及Global、Local GAN 判别器损失

# Deepfill 2018

上方Encoder关注学习已知背景的特征信息，得到Attention Map

精炼网络结构平行分为两个Encoder



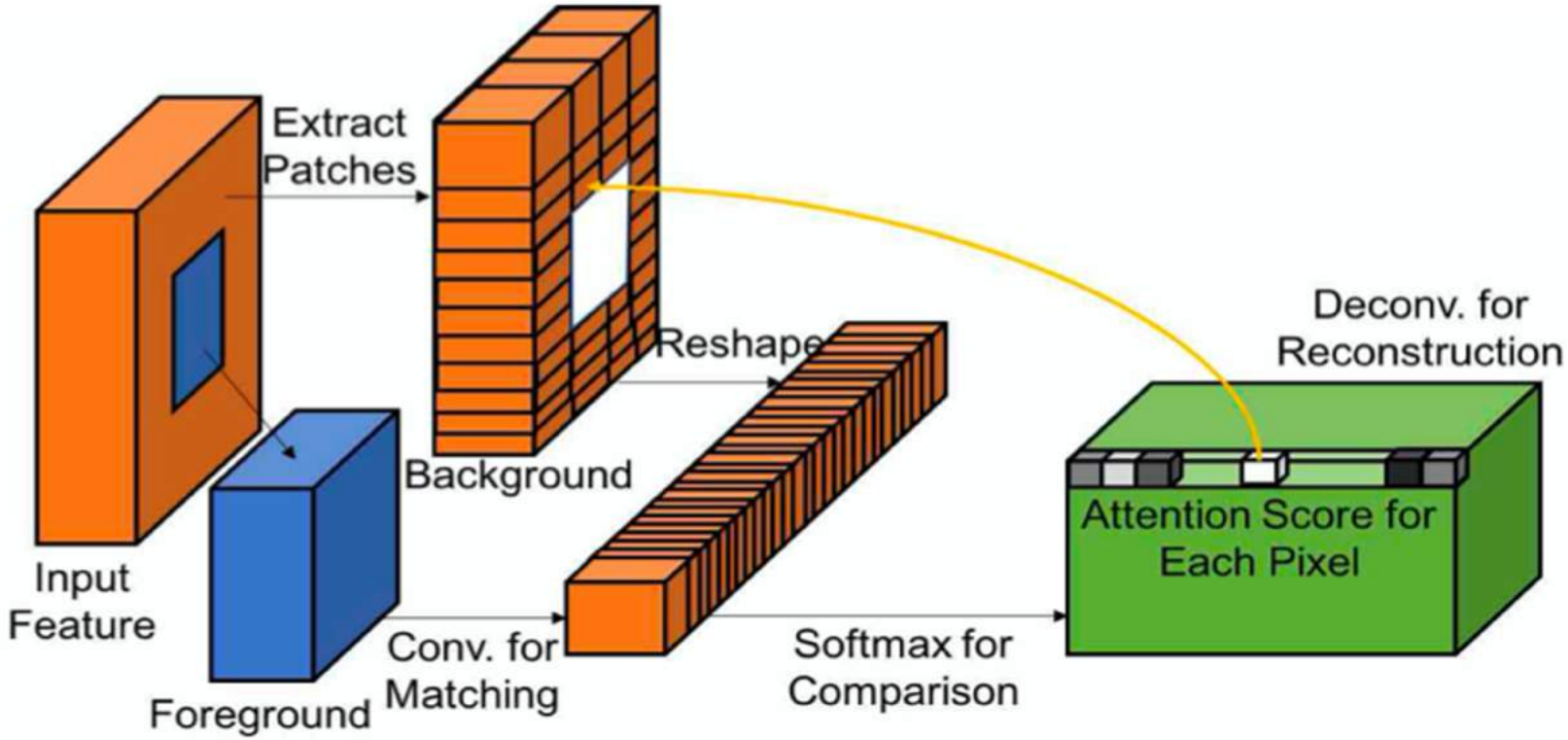

特征图和Attention Map合并后通过简单Decoder得到修复结果

下方Encoder关注整体网络结构中的生成内容特征图

# Contextual Attention Layer

在已知背景区域提取3*3的图像块，并reshape为卷积核

输入特征图



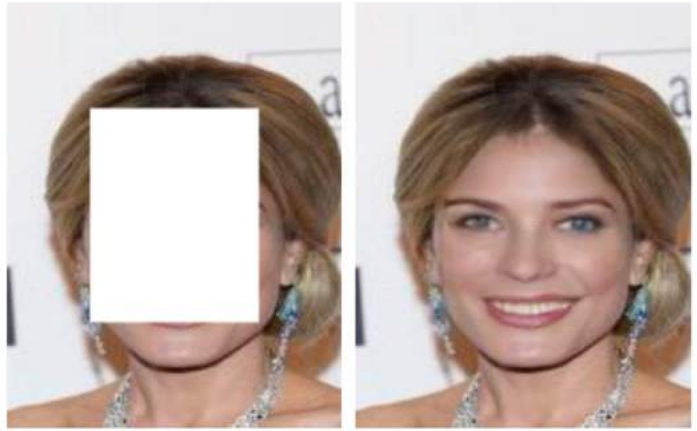

选取分数最高的图像块作为卷积核，反卷积得到前景区域。

未知前景区域和背景卷积核计算余弦相似度，通过softmax得到attention score

# 结果

Deepfill可以取得较好的结果

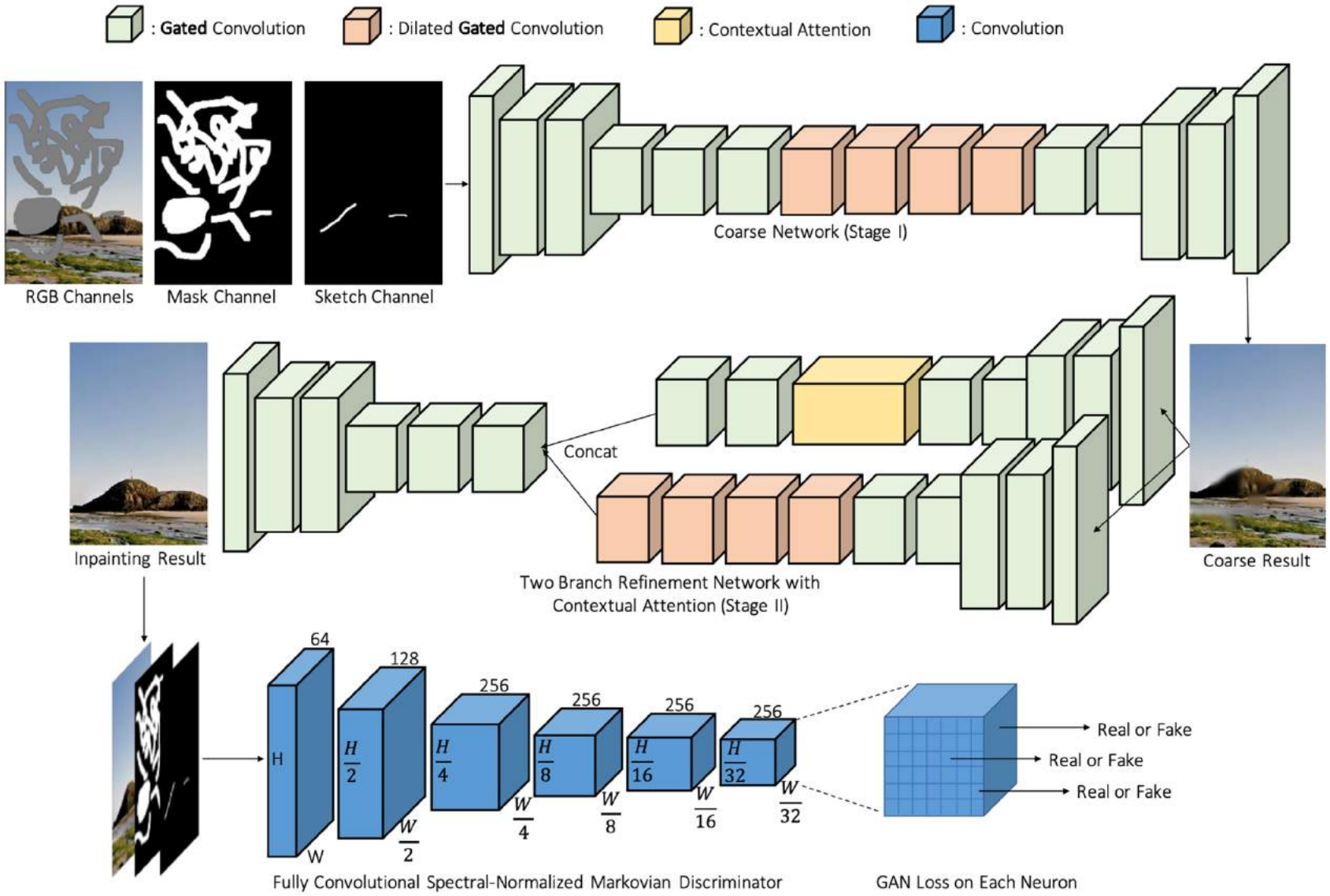
- 自然图像
- 纹理
- 人脸



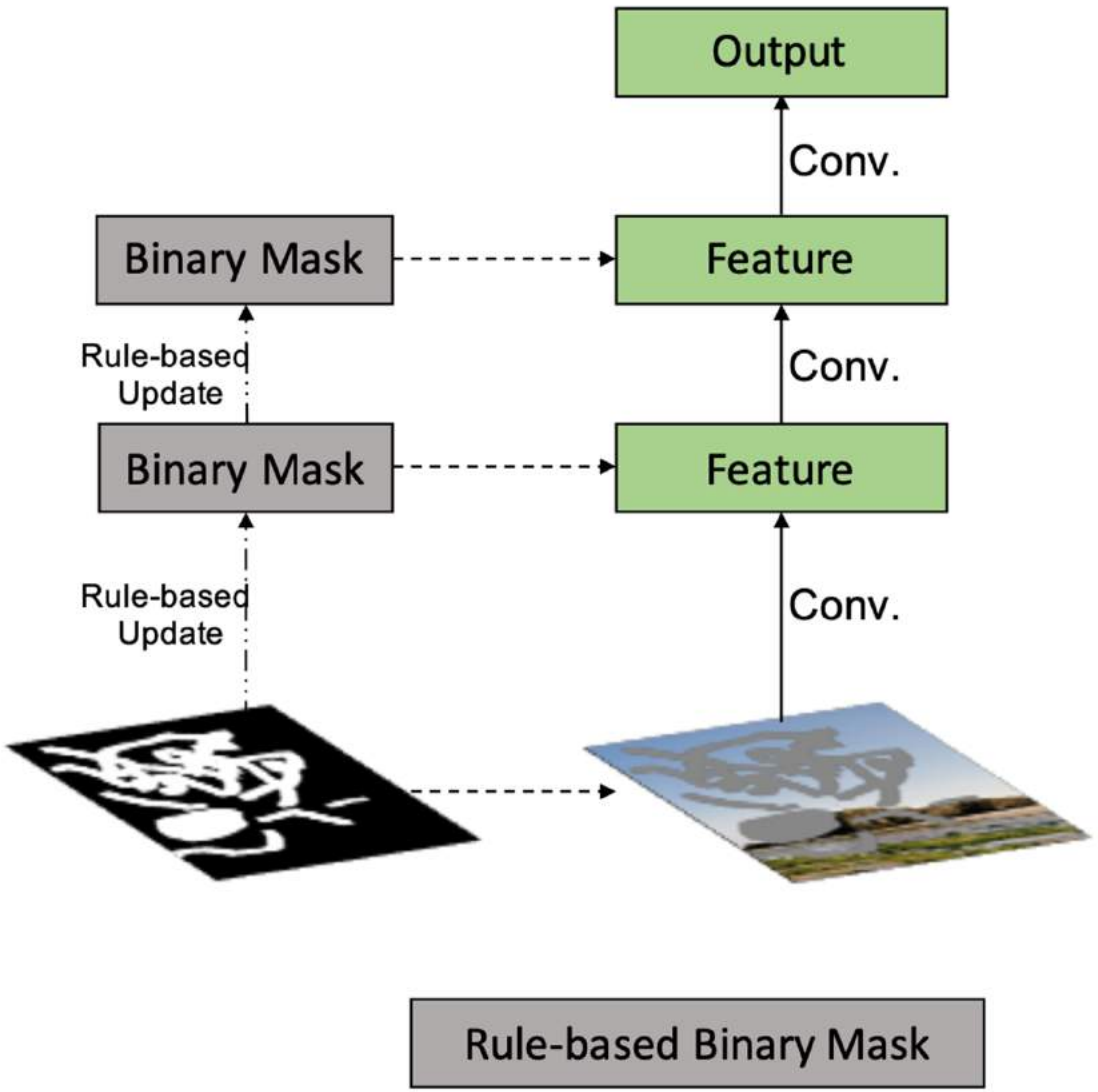
# Deepfill v2 2019

沿用了v1的两阶段结构，并有如下改进：

- 输入增加了用户草图
- 为了进一步处理不规则掩码的部分卷积（Partial Conv），提出了门控卷积（Gated Conv）
- 判别器改为PatchGAN，并增加了谱归一化（SN）



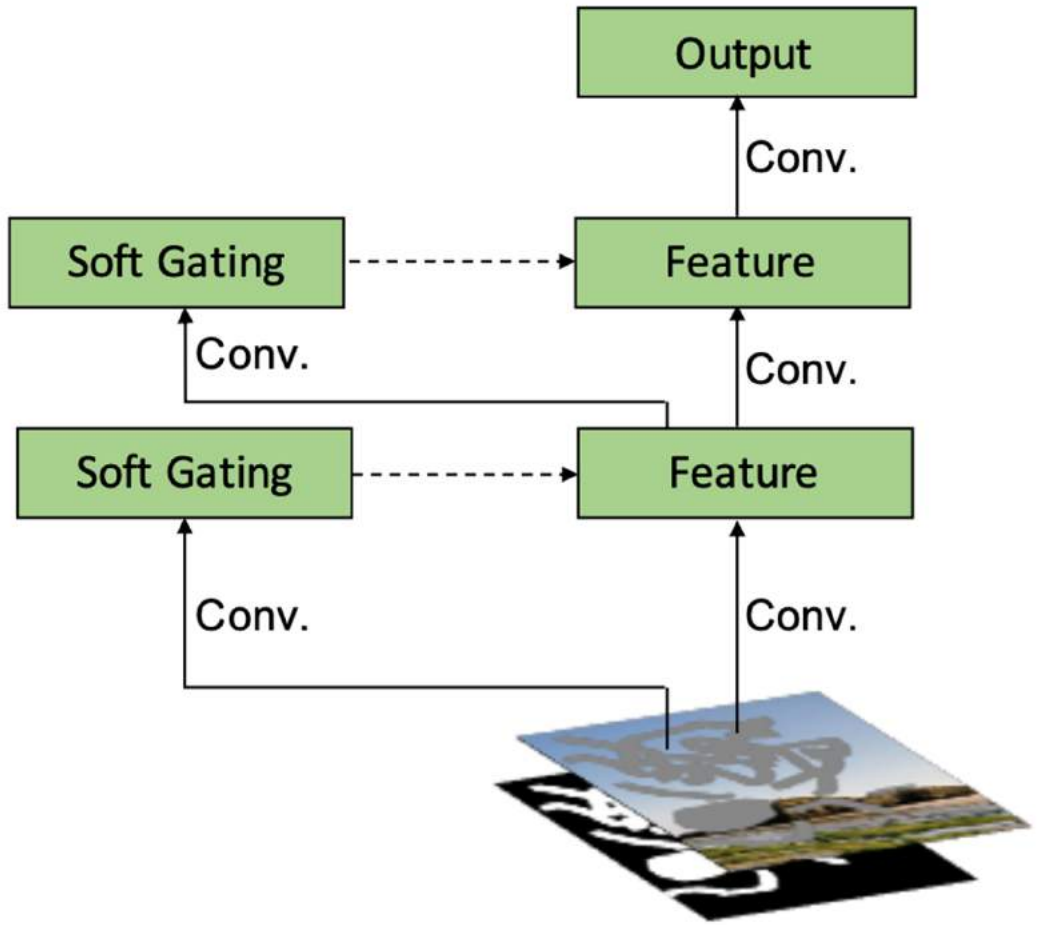
# 部分卷积，Partial Convolution




**部分卷积存在的问题：**

- 粗暴地将所有像素位置分为合法或非法，没有考虑不同层合法像素的个数的连贯性
- 规则上无法给出输入用户草图的指导
- 随着网络的深入，所有的掩膜最后都会逐渐变为1，丧失部分卷积的意义
- 同一层的每个通道都延用同一个掩膜，缺少灵活性

# 门控卷积，Gated Convolution




**门控卷积：**

- 对每一个通道都进行掩膜更新
- 通过标准卷积和S型激活函数更新掩膜，使得值落在0-1之间
- 将得到的soft gating掩膜再应用到特征图上
- 使得整个掩膜更新过程可学习

# 结果


在Deepfill的基础上，v2还可以增加用户草图输入，在填充图像的基础上对原图进行部分编辑，达到更为复杂的效果。

# 谢谢大家

OpenMMLab

# 通用视觉框架OpenMMLab

# 第7讲 底层视觉与MMEditing (下)

吕健勤 教授

2021年5月

# 抠图

Image Matting

# 什么是抠图

## 目标

对图片中已知区域做精细化分割

## 建模

$\boldsymbol{C}_i = \alpha_i \boldsymbol{F}_i + (1 - \alpha_i)\boldsymbol{B}_i$

已知每点像素值 $\boldsymbol{C}_i$ ，预测透明度 $\alpha_i$

## 应用

人物抠像，背景替换，影视制作

# 证件照抠图



| 原照片 | 预览证件照 |
| --- | --- |
|  |   |