

通用视觉框架OpenMMLab
第6讲 底层视觉与MMEdition (上)

吕健勤 教授
2021年5月



本节内容：

- 图像超分辨率 Super Resolution
 - 基于卷积网络的模型 SRCNN 与 FSRCNN
 - 损失函数
 - 对抗生成网络 GAN 简介
 - 基于 GAN 的模型 SRGAN 与 ESRGAN
 - 视频超分辨率介绍
- 实践 MMEdition 1



下节预告：

- 图像补全 Inpainting
- 图像转译 Translation
- 抠图 Matting
- 实践 MMEdition 2

什么是图像超分辨率

OpenMMLab



图像超分辨率：根据从低分辨率图像重构高分辨率图像



- 提高图像的分辨率
 - 高分图像符合低分图像的内容
 - 恢复图像的细节、产生真实的内容
- > 双线性或双立方插值不能恢复图像的高频细节



超分辨率相关的研究在1984年首次发表 [1]，而超分辨率这个术语在1990年被首次提出 [2]。

1. R. Y. Tsai and T. S. Huang, "Multiframe image restoration and registration," in Advances in Computer Vision and Image Processing, vol. 1, chapter 7, pp. 317-339, JAI Press, Greenwich, Conn, USA, 1984.
2. M. Irani and S. Peleg. 1991, "Super Resolution From Image Sequences" ICPR, 2:115--120, June 1990.

应用：经典游戏高清重制



nearest neighbor

应用：动画高清重制

OpenMM Lab

来源: <https://animerefiner.com/>



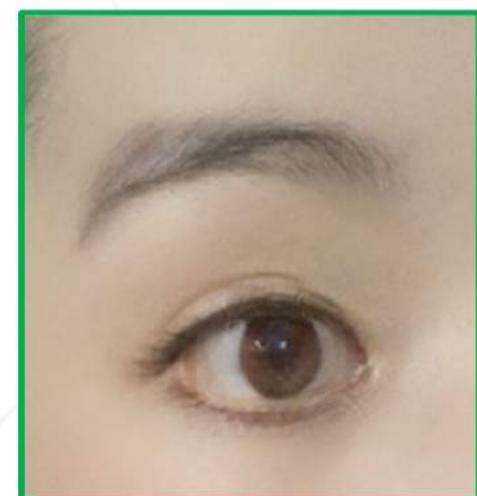
Before (480x270px)



After (1920 x 1080 px, FullHD)

©TOMASON
upconverted by RADIUS5 Inc.

应用：照片修复



应用：节约传输高清图像的带宽

在网络上传输高清视频

最终分辨率： 1500×1000

网络流量：100kb



在网络上传输原图1/4的像素值

在网络上传输低清视频、再在本地恢复

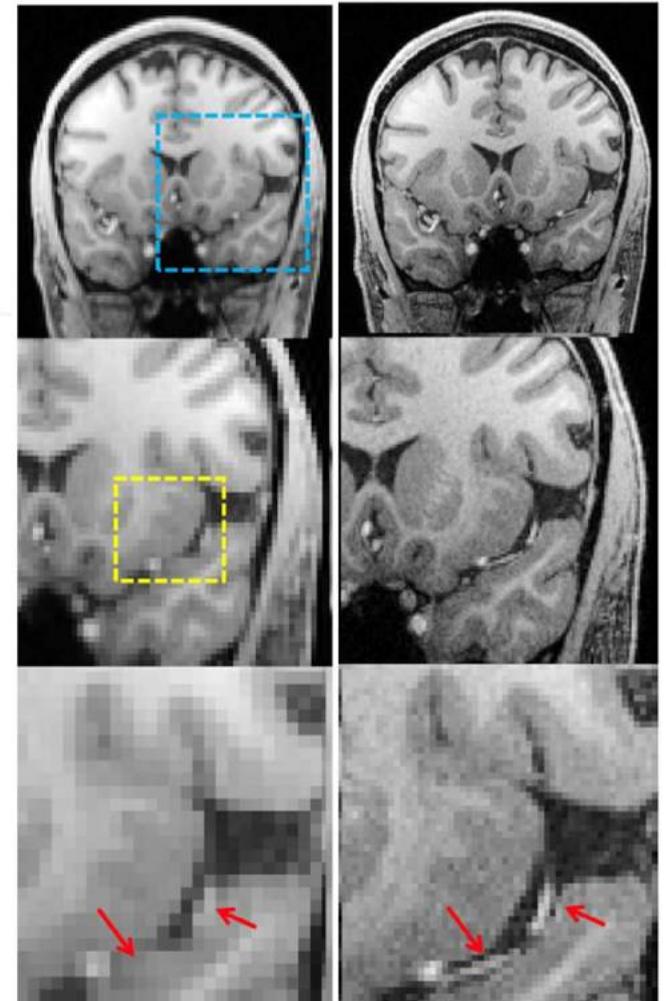
最终分辨率： 1500×1000

网络流量：**25kb**



在用户设备上恢复高清图像

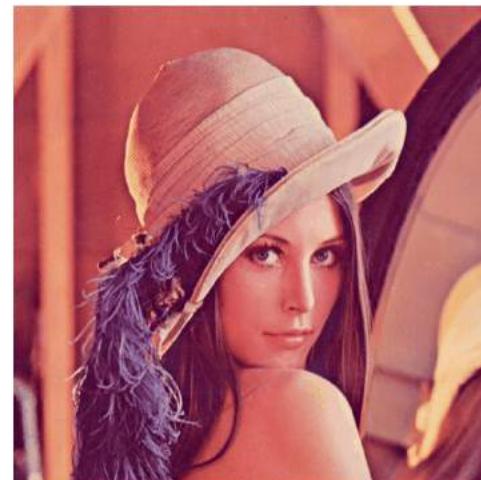
- 医疗影像
- 卫星影像
- 监控系统（车牌或人脸）
- 空中监察
- ...



多图超分



2 倍放大

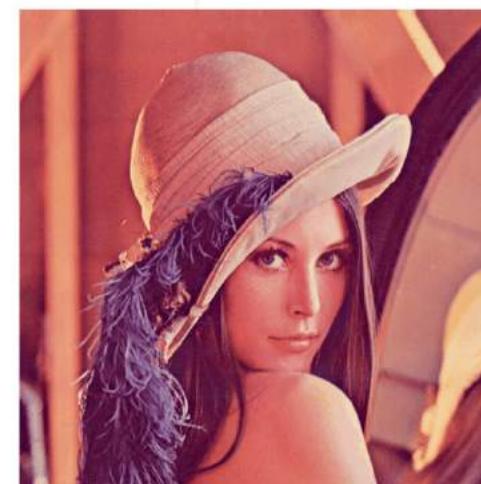


常见于遥感影像

单图超分



2 倍放大



单图超分是**不适定问题** ill-posed problem;
对于给定低分辨率图像，高分辨率图像不能唯一确定。

1. 基于已知数据学习高低分辨率图像之间的关系：先验知识
2. 在符合先验知识的条件下恢复高清图像



目标：内容更真实、细节更丰富



低分辨率图像



2 倍放大

先验知识

图像数据集
高分-低分图像对

高分辨率



...



低分辨率

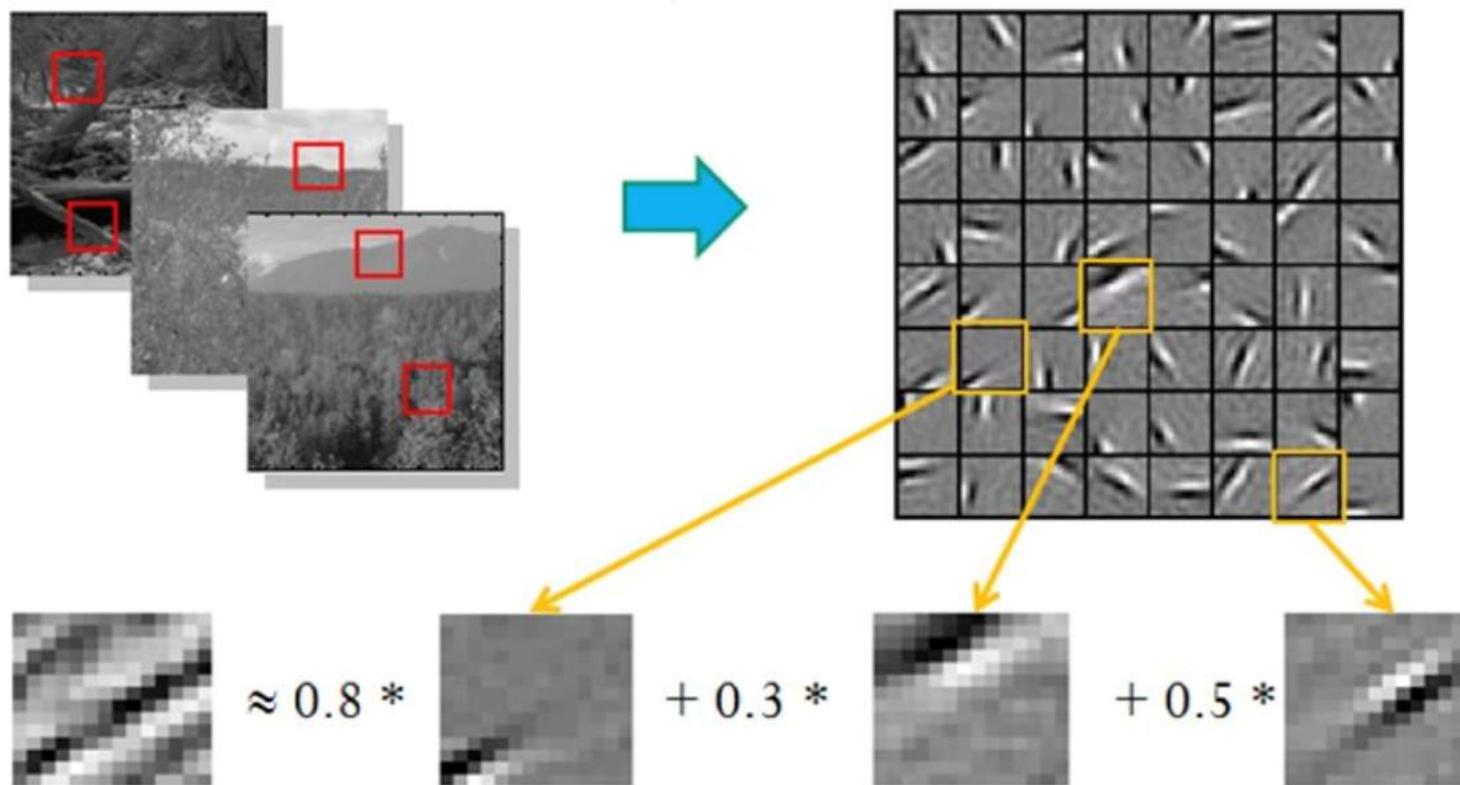


...

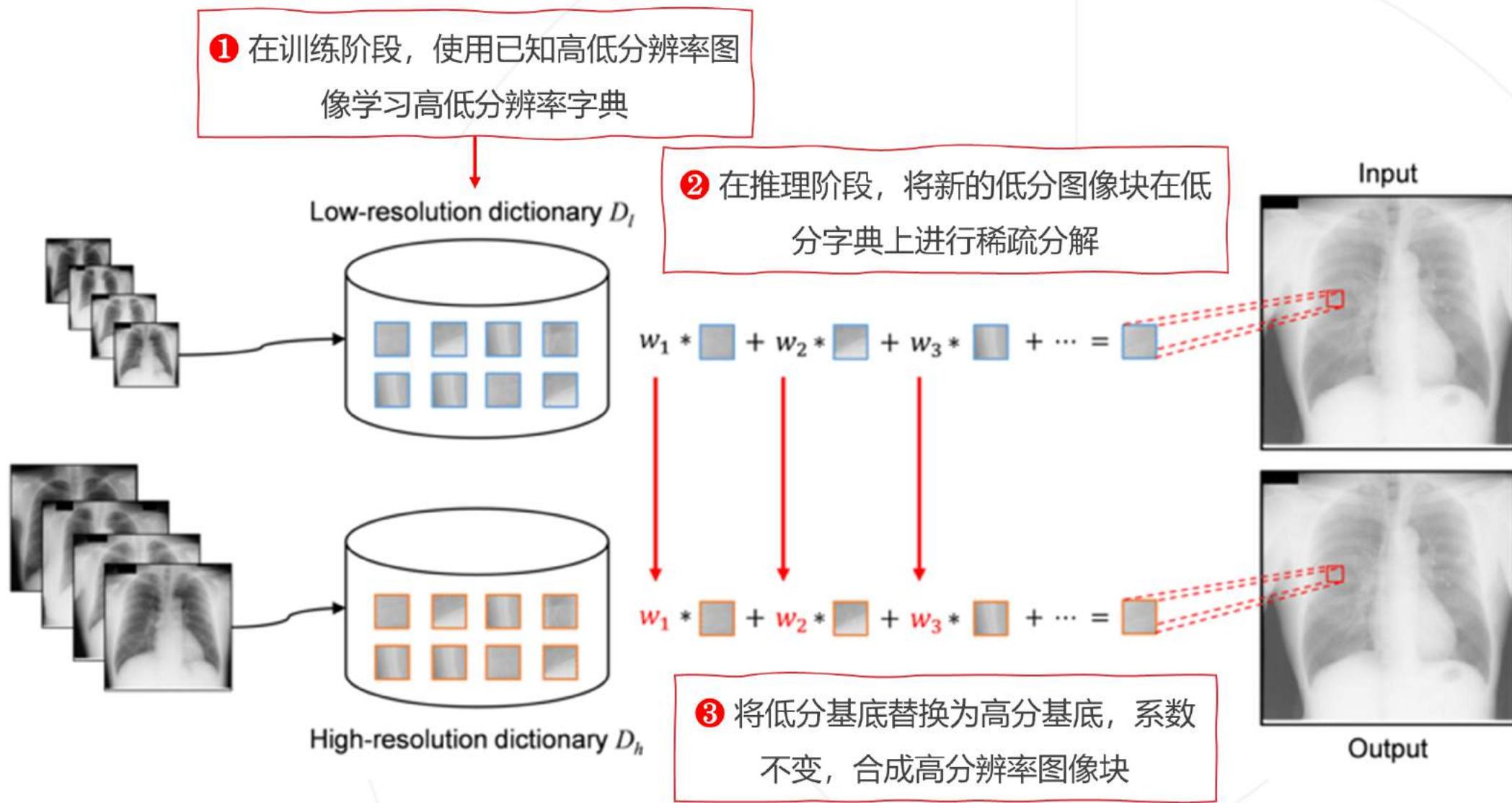


高分辨率图像

通过无监督学习方法，构建一组“基图像块”，使得图像中的任意一块区域可以由**少数**基图像块线性组合而成。



$$[a_1, \dots, a_{64}] = [0, 0, \dots, 0, \mathbf{0.8}, 0, \dots, 0, \mathbf{0.3}, 0, \dots, 0, \mathbf{0.5}, 0]$$



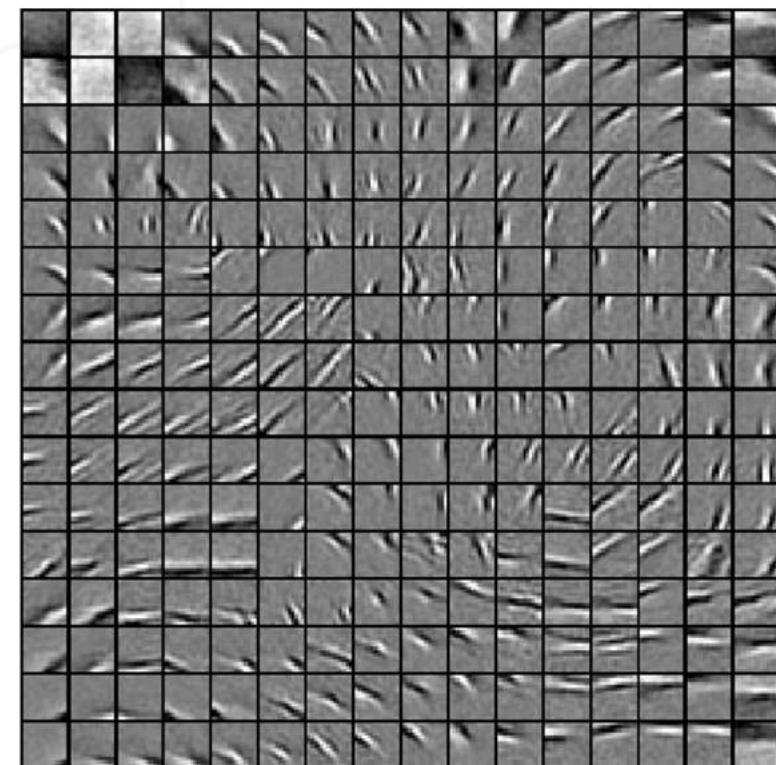
即便已经学习出字典，对低分辨率图像块进行系数分解、得到系数仍然是一个相对复杂的优化问题。

训练和推理阶段都很耗时

在推理阶段，基于字典求解系数仍然
涉及复杂的计算

$$\text{Image} \approx 0.8 * \text{Dictionary}_1 + 0.3 * \text{Dictionary}_2 + 0.5 * \text{Dictionary}_3$$

Represent x_i as: $a_i = [0, 0, \dots, 0, \mathbf{0.8}, 0, \dots, 0, \mathbf{0.3}, 0, \dots, 0, \mathbf{0.5}, \dots]$



基于卷积网络和普通损失函数

使用卷积神经网络，端到端从低分辨率图像恢复高分辨率图像

代表算法：SRCNN 与 FSRCNN

使用生成对抗网络

采用生成对抗网络的策略，鼓励产生细节更为真实的高分辨率图像

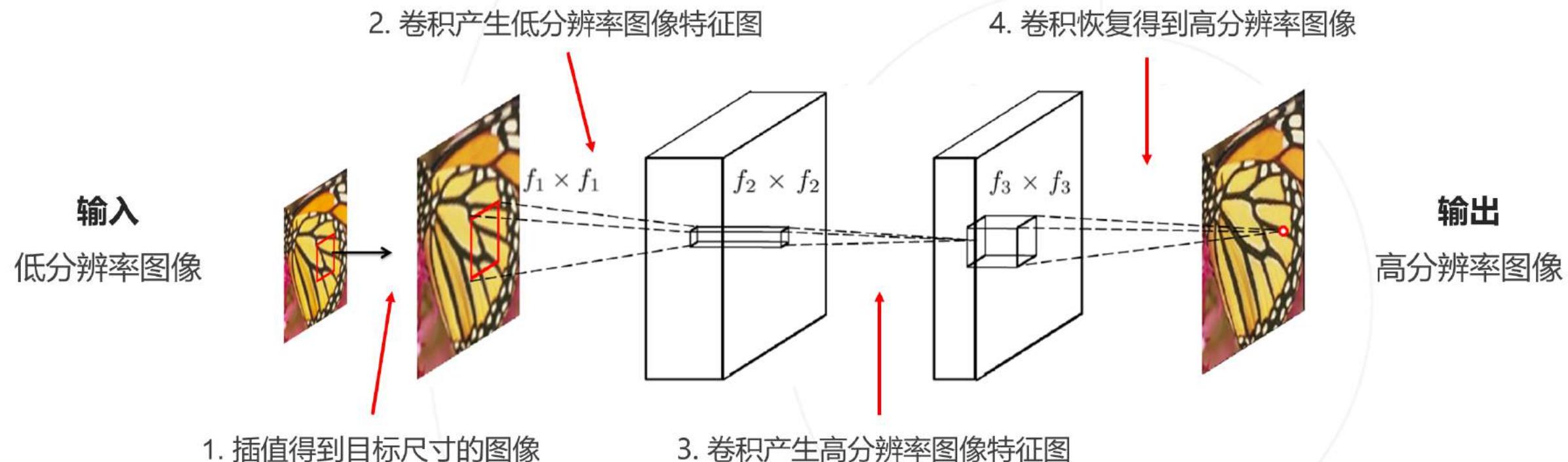
代表算法：SRGAN 与 ESRGAN

卷积网络模型

SRCNN、FSRCNN、SRResNet

SRCNN 是首个基于深度学习的超分辨率算法，证明了深度学习在底层视觉的可行性。

模型仅由三层卷积层构成，可以端到端学习，不需要额外的前后处理步骤。

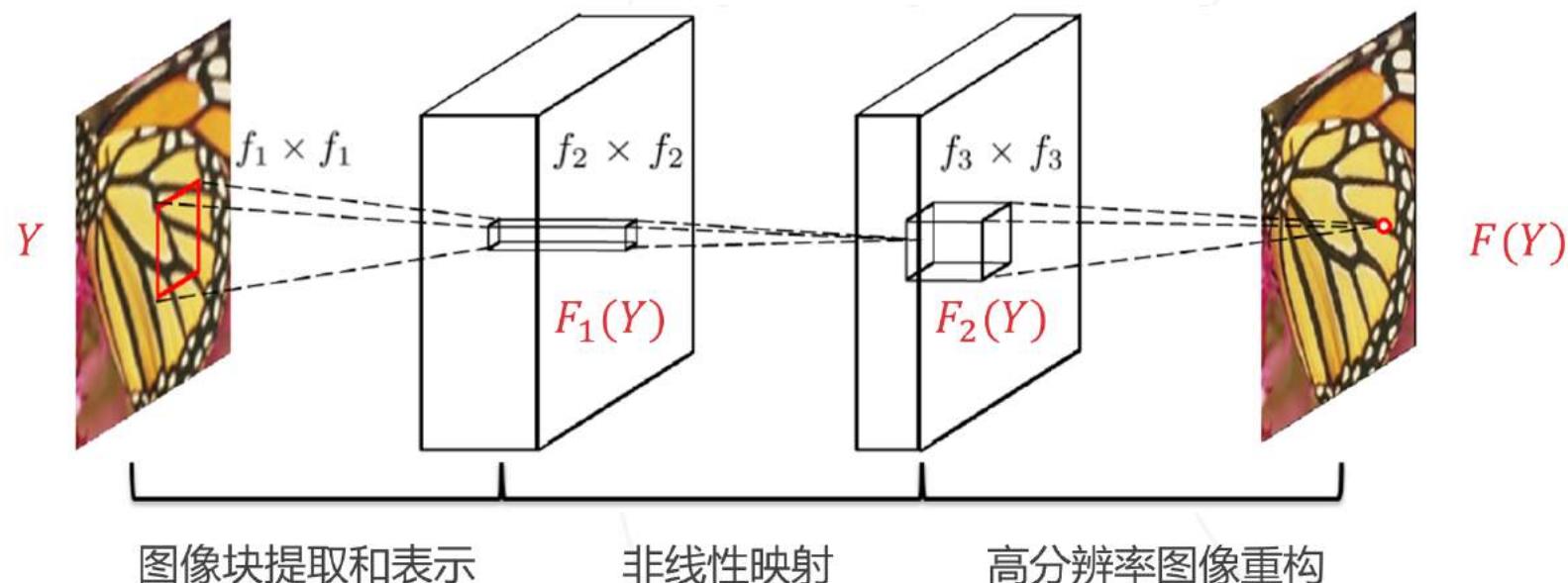


SRCNN 的单个卷积层有明确的物理意义：

第一层：提取图像块的低层次局部特征；

第二层：对低层次局部特征进行非线性变换，得到高层次特征；

第三层：组合邻域内的高层次特征，恢复高清图像。



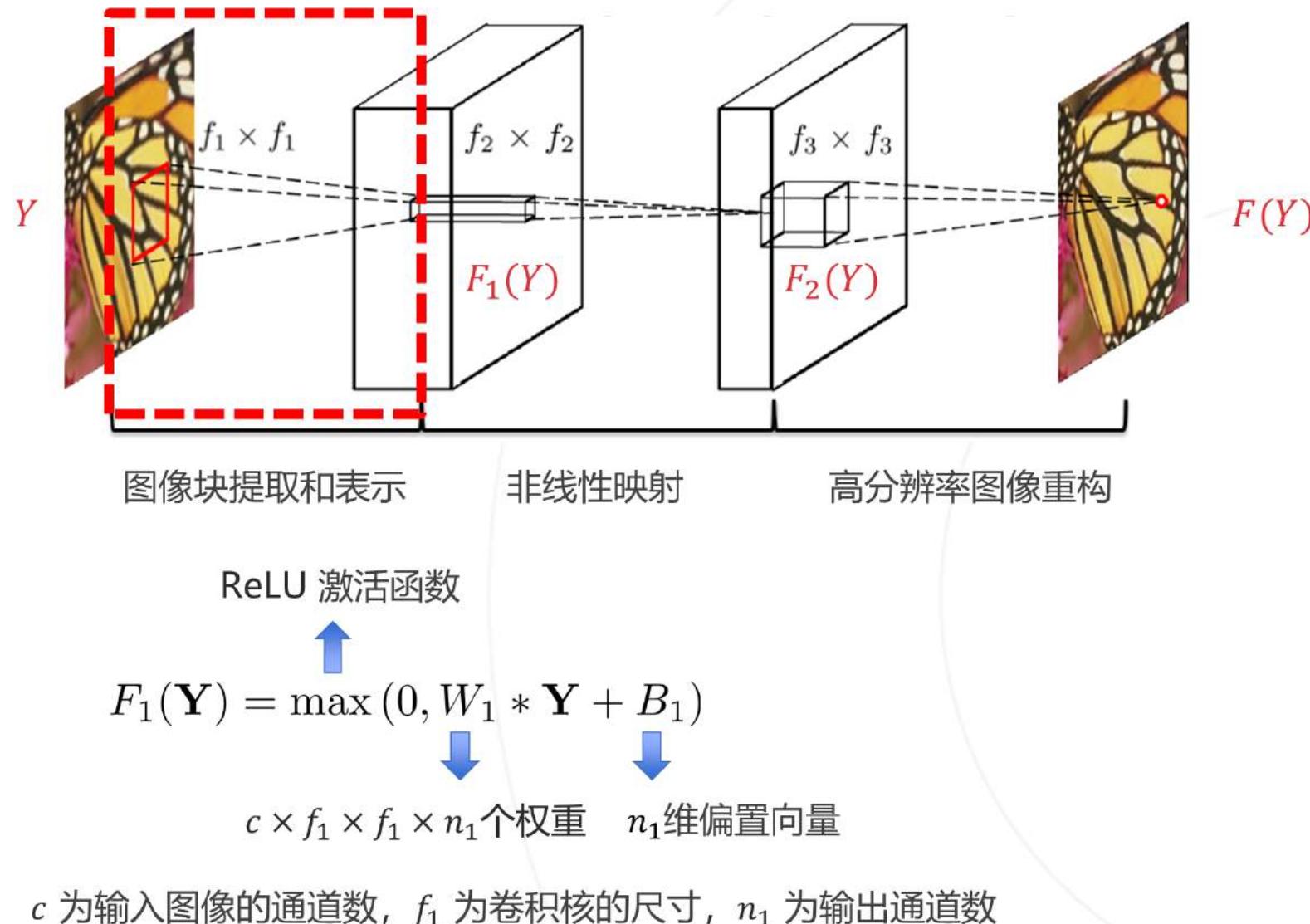
$$F_1(\mathbf{Y}) = \max(0, W_1 * \mathbf{Y} + B_1)$$

$$F_2(\mathbf{Y}) = \max(0, W_2 * F_1(\mathbf{Y}) + B_2)$$

$$F(\mathbf{Y}) = W_3 * F_2(\mathbf{Y}) + B_3$$

三层卷积核大小可以设置为
9-1-5 或 9-5-5

第一层：特征提取

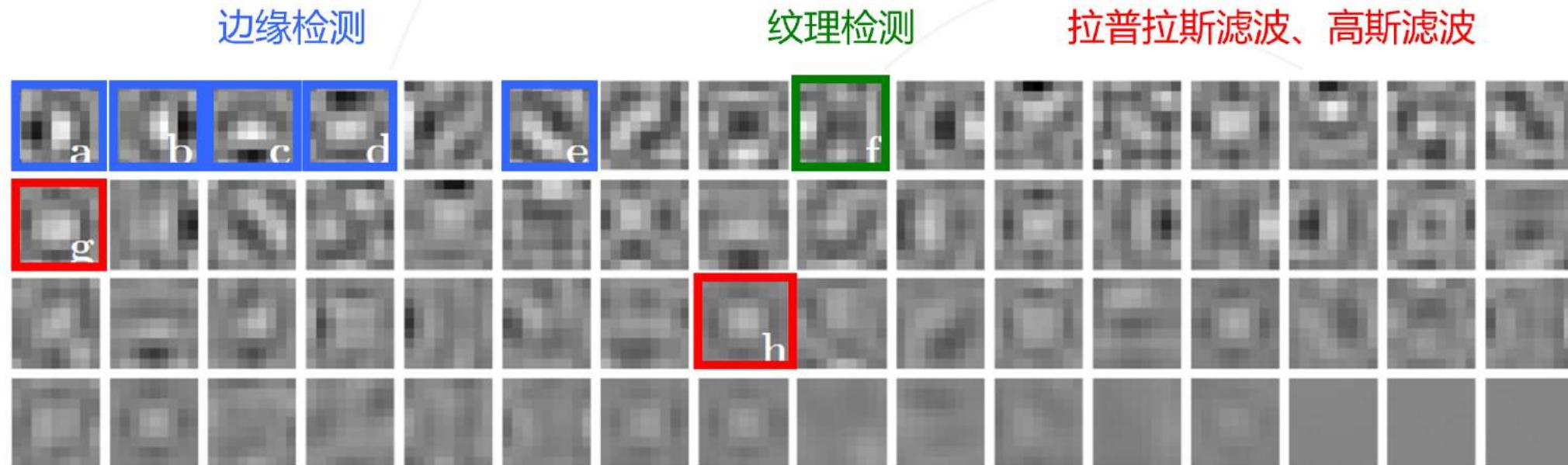


经典方法通常将图像切分成小块，并基于一系列基底对图像块进行分解（常用算法有PCA、DCT、Haar小波等），分解系数向量即为图像块在基底上的表示。

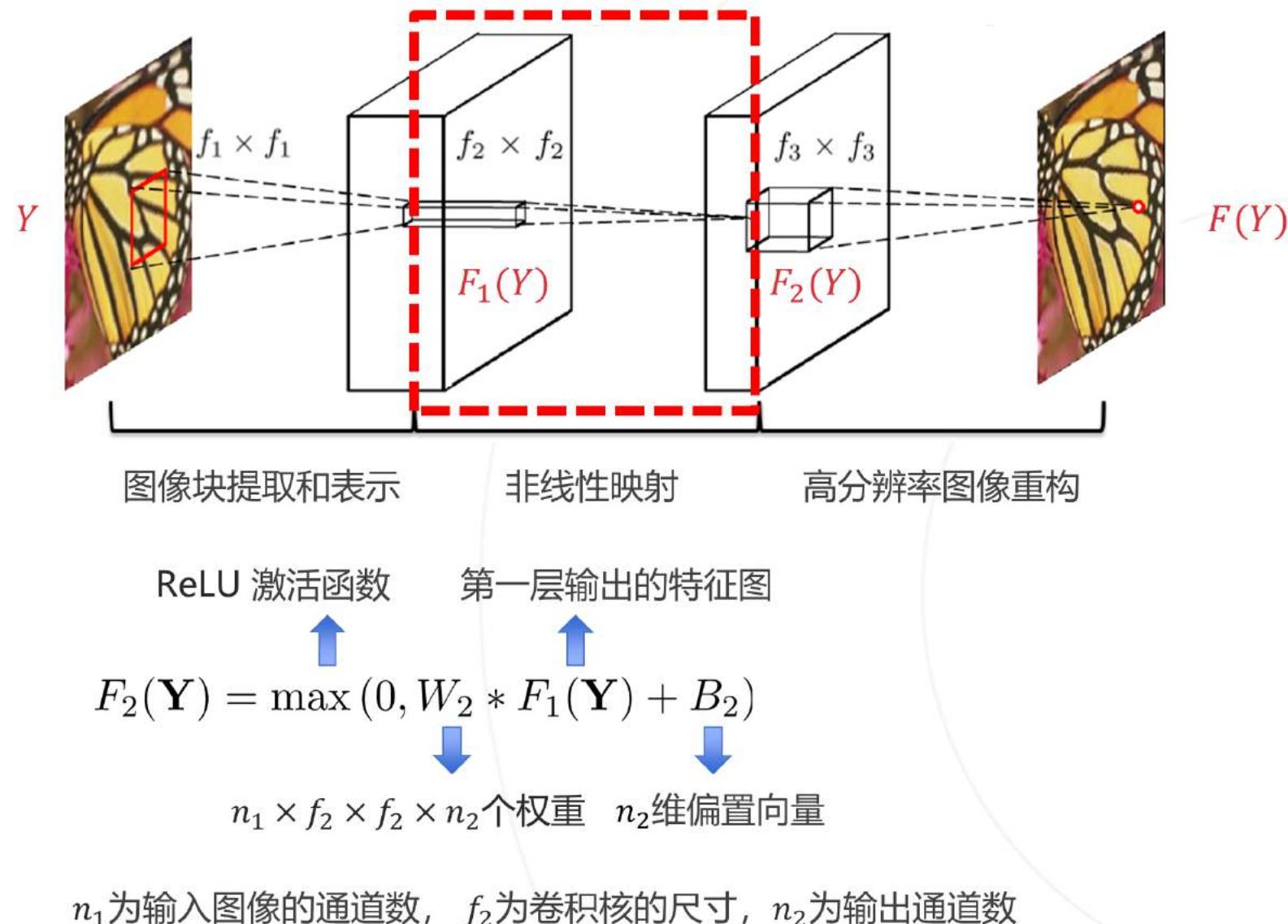
这个操作等价于用一系列卷积核（对应经典方法中的基底）对原图像进行卷积。 $F_1(Y)$ 中每个像素位置上的 n_1 维度的向量即为对应图像块在基底上的表示。

使用神经网络，基底可以从数据中学习出来。

在 ImageNet 数据集上训练的 SRCNN 可以学习到不同的低层次特征所对应的卷积核。



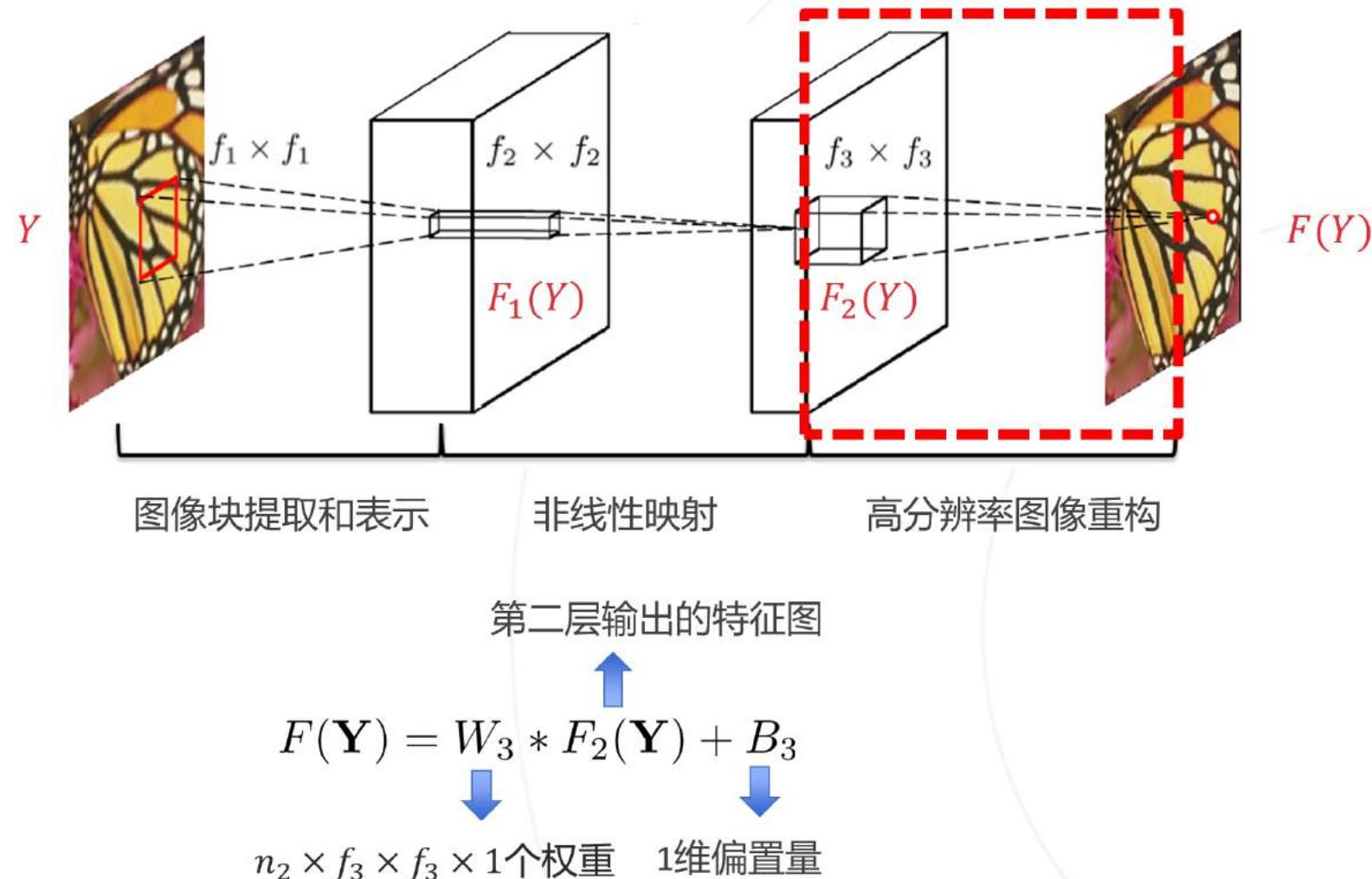
第二层：非线性映射



$f_2 = 1$ 时, 第二层卷积将 $F_1(Y)$ 每个位置上 n_1 维度的特征非线性映射为一个 n_2 维的特征。

该特征可以看作是图像块在高分基底上的表示, 在后一层中用于重构。

非线性映射可以有很多层, 但实验表明只应用单层卷积层就可以达到较好的效果。



n_2 为输入图像的通道数, f_3 为卷积核的尺寸, 输出通道数为 1

第三层的卷积核对应高分辨率基底, 用 $F_2(Y)$ 中的系数对高分基底加权求和即可得到高分图像块。第三层卷积完成这个过程。

三个步骤与稀疏编码方法中的步骤一一对应。

- 准备数据：

将 ImageNet 数据集中的图像作为高分图像，降采样再插值升采样得到的图像作为低分图像

- 需要学习的参数：

$$\Theta = \{W_1, W_2, W_3, B_1, B_2, B_3\}$$

- 损失函数：逐像素计算恢复图像和原高分图像的平方误差 (Mean Squared Error, MSE)

$$L(\Theta) = \frac{1}{n} \sum_{i=1}^n \|F(\mathbf{Y}_i; \Theta) - \mathbf{X}_i\|^2,$$

最小化损失函数即可鼓励网络完美恢复高分辨率图像

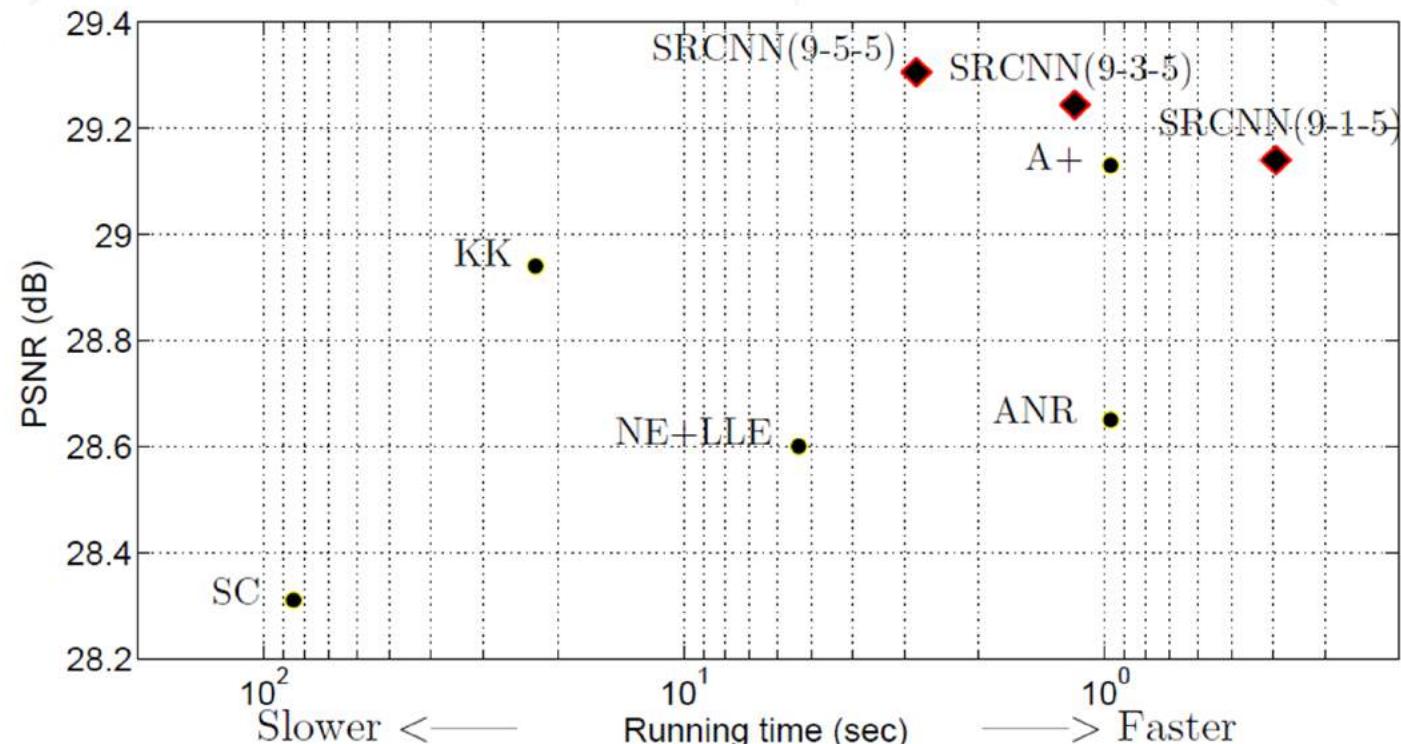
- 通过标准的 SGD 训练模型

$$\Delta_{i+1} = 0.9 \cdot \Delta_i - \eta \cdot \frac{\partial L}{\partial W_i^\ell}, \quad W_{i+1}^\ell = W_i^\ell + \Delta_{i+1},$$

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right)$$

峰值信噪比 (Peak signal-to-noise ratio, **PSNR**) 为最大信号能量与平均噪声能量的比值，值越大恢复效果越好。

SRCNN 在性能和速度上全面超越深度学习前的算法



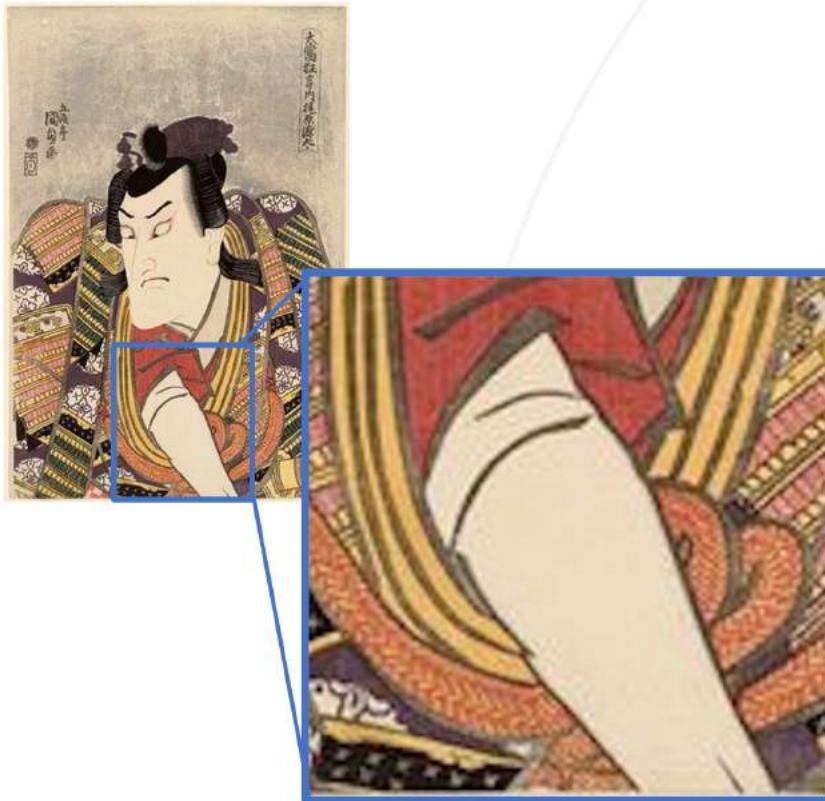
SC – Sparse coding by Yang et al. TIP 2010

NE+LLE – Neighbour embedding + non-negative least squares by Chang CVPR 2004

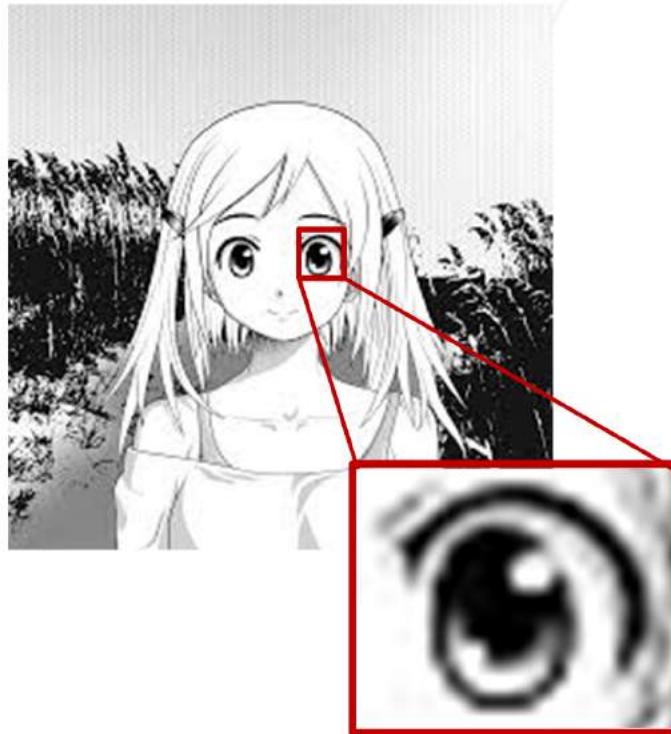
KK – Sparse regression and natural image prior by Kim and Kwon TPAMI 2010

ANR – Anchored Neighbourhood Regression by Timofte et al. ICCV 2013

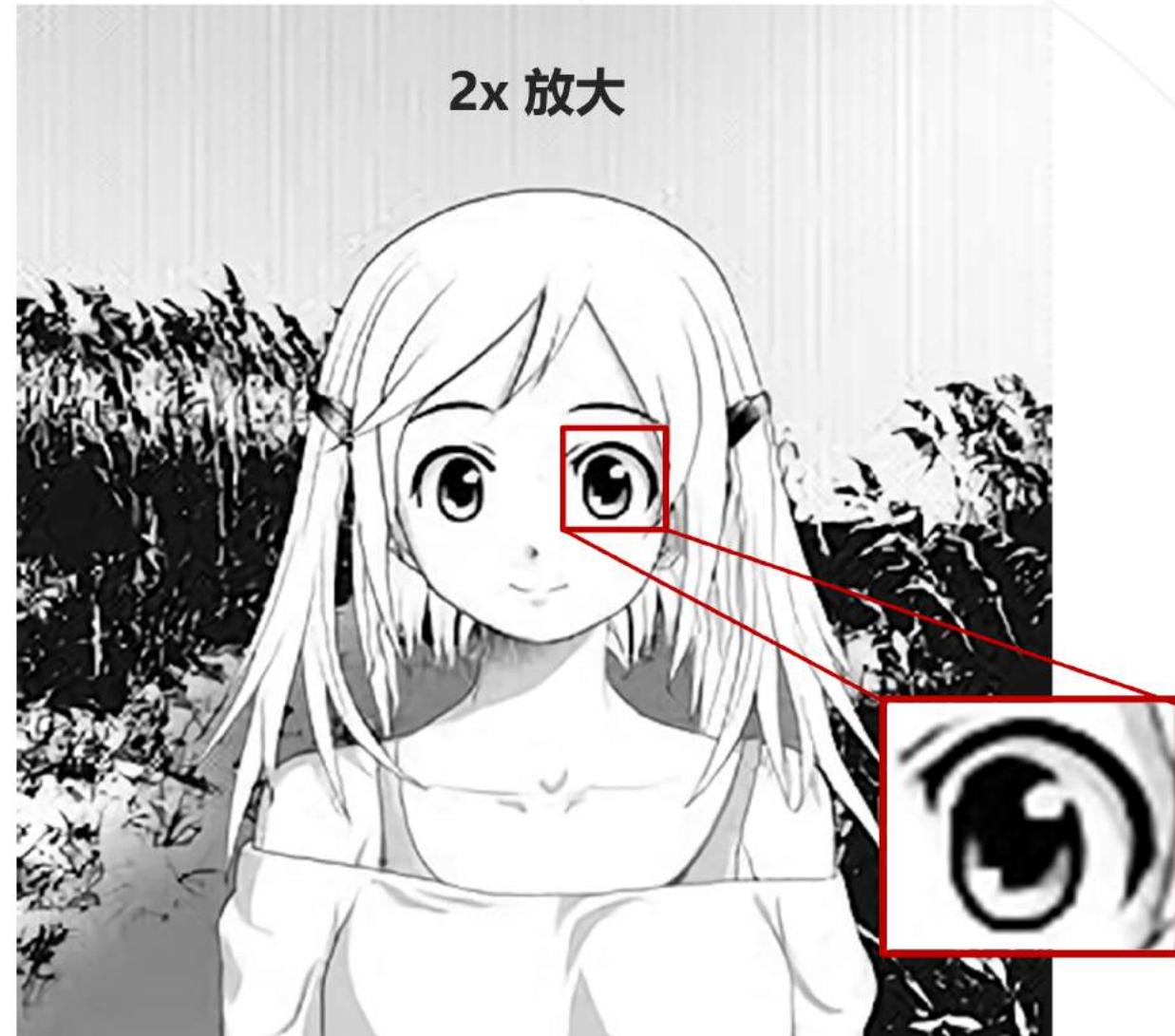
A+ – Adjusted Anchored Neighbourhood Regression by Timofte et al. ACCV 2014



原图

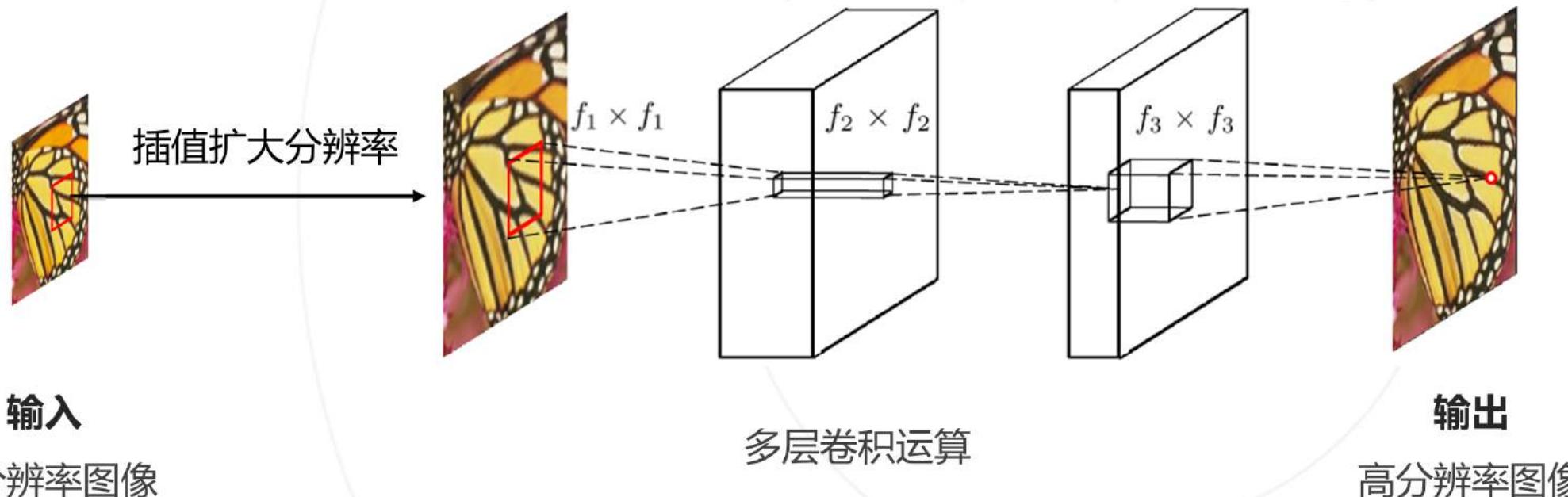


2x 放大



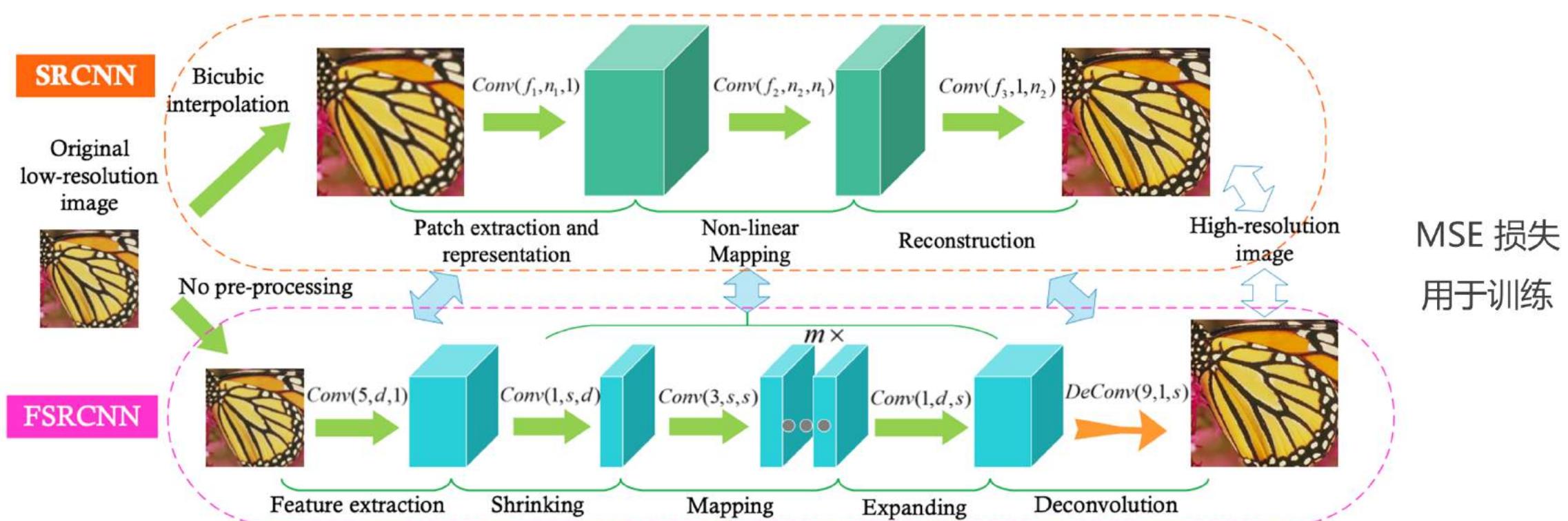
SRCNN 先对低分图像进行插值，再在高分辨率下进行卷积运算；
然而插值不产生额外信息，因而产生一定的冗余计算；

在学术数据集上，SRCNN 的速度在 1~10 FPS，达不到实时的标准。

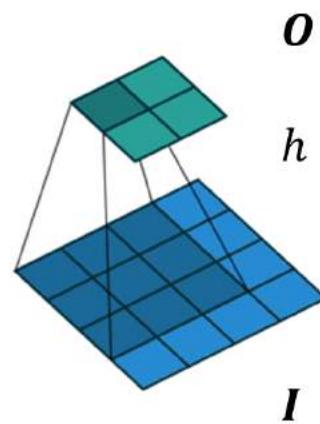


FSRCNN 在 SRCNN 的基础上针对速度进行了改进：

1. 不使用插值，直接在低分辨率图像上完成卷积运算，降低运算量
2. 使用 1×1 的卷积层对特征图通道进行压缩，进一步降低卷积的运算量
3. 若干卷积层后再通过转置卷积层提高图像分辨率



卷积



卷积核

$$h = \begin{pmatrix} 1 & 4 & 1 \\ 1 & 4 & 3 \\ 3 & 3 & 1 \end{pmatrix}$$

$$C = \begin{pmatrix} 1 & 4 & 1 & 0 & 1 & 4 & 3 & 0 & 3 & 3 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 4 & 1 & 0 & 1 & 4 & 3 & 0 & 3 & 3 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 4 & 1 & 0 & 1 & 4 & 3 & 0 & 3 & 3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 4 & 1 & 0 & 1 & 4 & 3 & 0 & 3 & 3 & 1 \end{pmatrix}$$

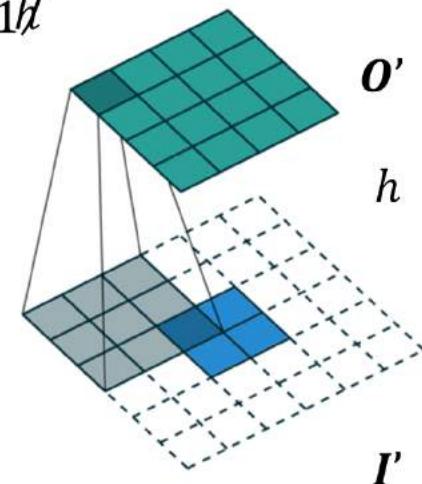
$$O = I * h$$

将卷积运算表示成
矩阵乘法

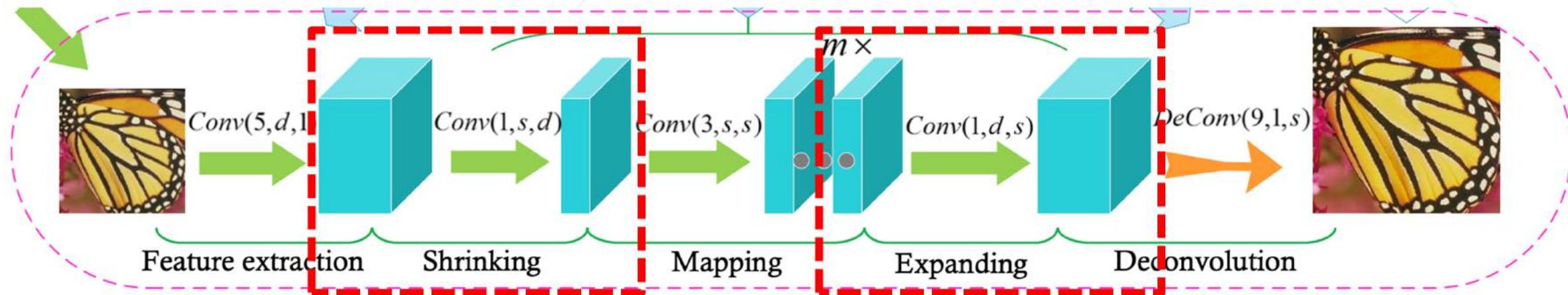
$$O = I * h = C I$$

$$O' = I' \star h = C^T I'$$

转置卷积



转置卷积通过矩阵转置，调换输入输出之间的运算关系，可以对特征图进行升采样。

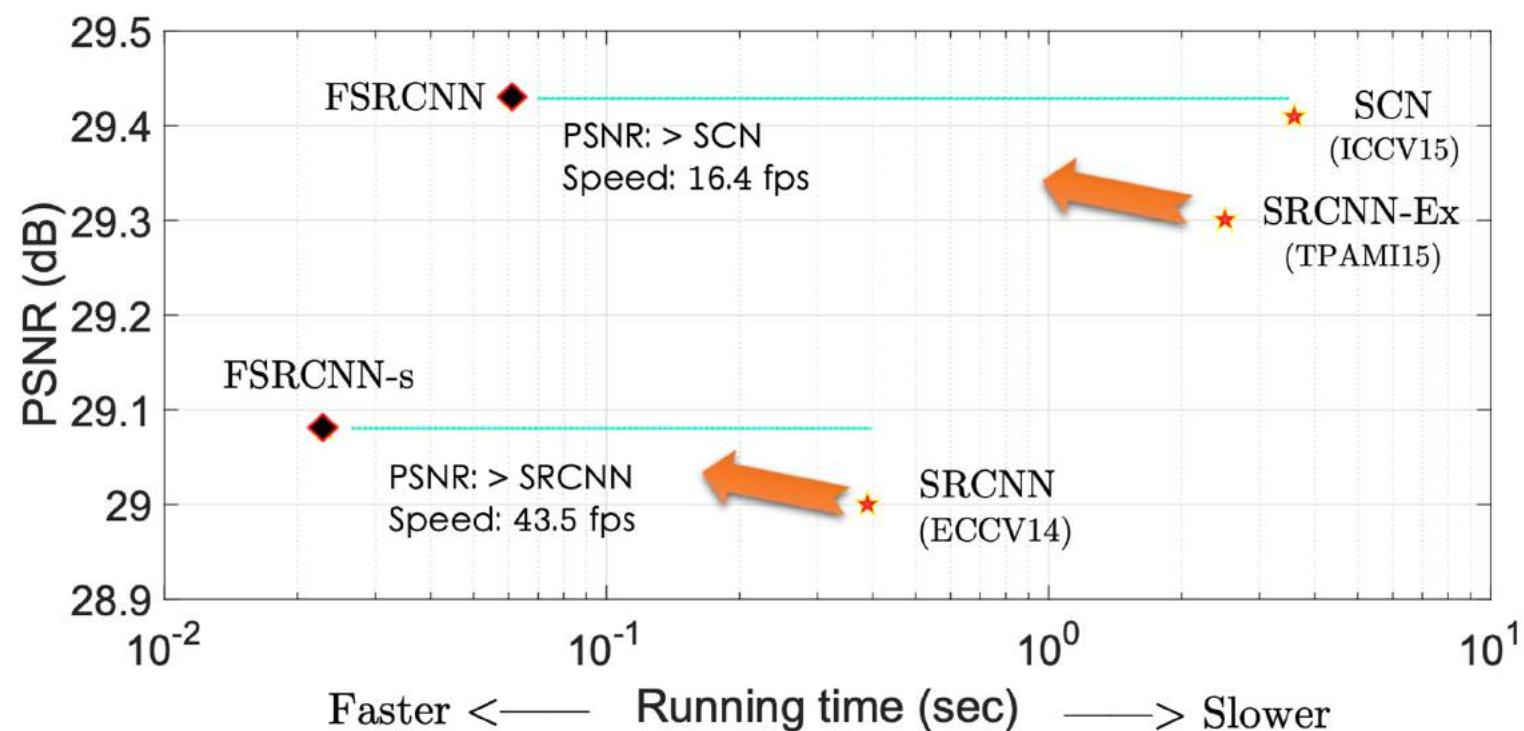


SRCNN 使用一个卷积层完成特征映射，因为特征维度 d 通常较大，卷积运算量较大。

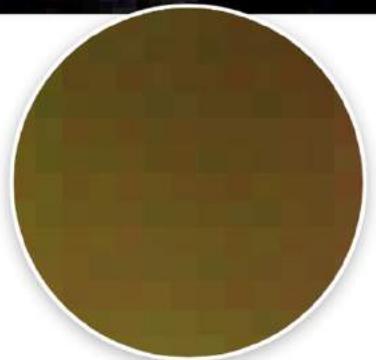
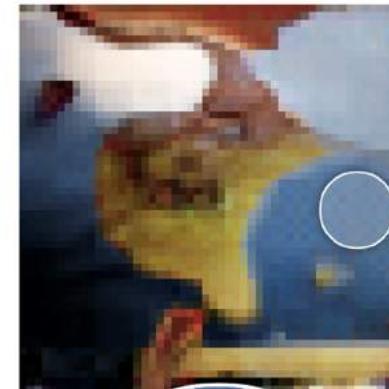
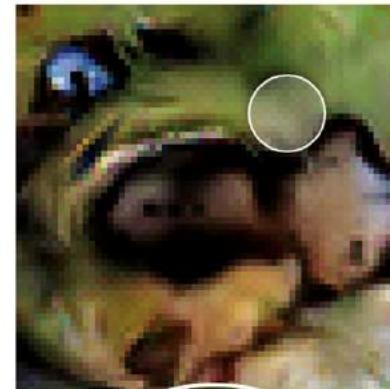
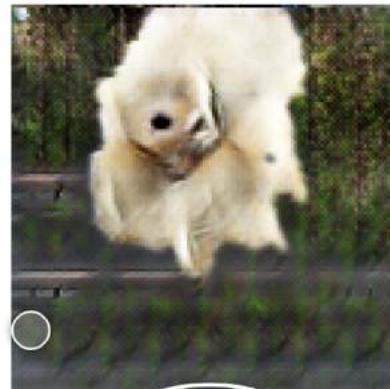
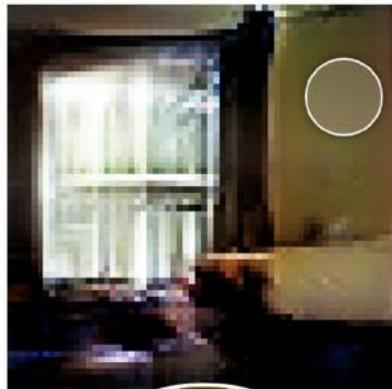
FSRCNN 中

1. 增加缩小层，其中 $s << d$ ，将维度减少到 s ，从而减少特征映射过程的计算量；
2. 增加放大层，将特征维度还原到 d ，防止结果精度受到影响。

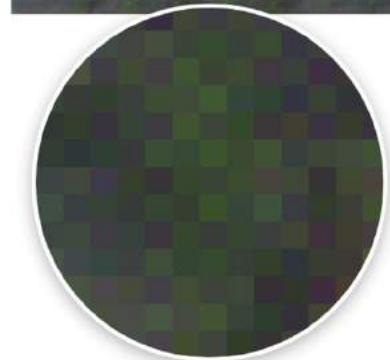
1. 基于 CPU 进行推理，速度可以达到实时；
2. 在处理不同上采样倍数时，只需要微调反卷积的权重，特征映射层的参数额可以保持不变，大幅加快训练速度。



转置卷积会在图像上产生棋盘格效应，降低生成高分辨率图像的质量。



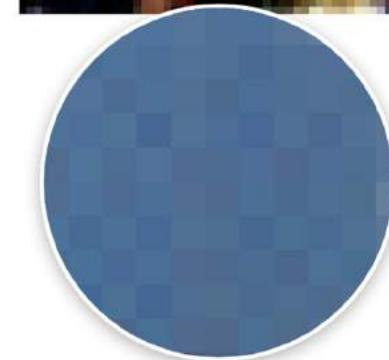
Radford, et al., 2015 [1]



Salimans et al., 2016 [2]



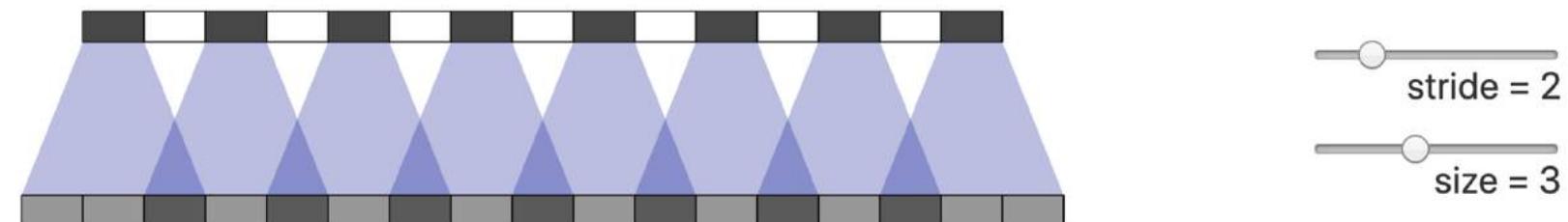
Donahue, et al., 2016 [3]



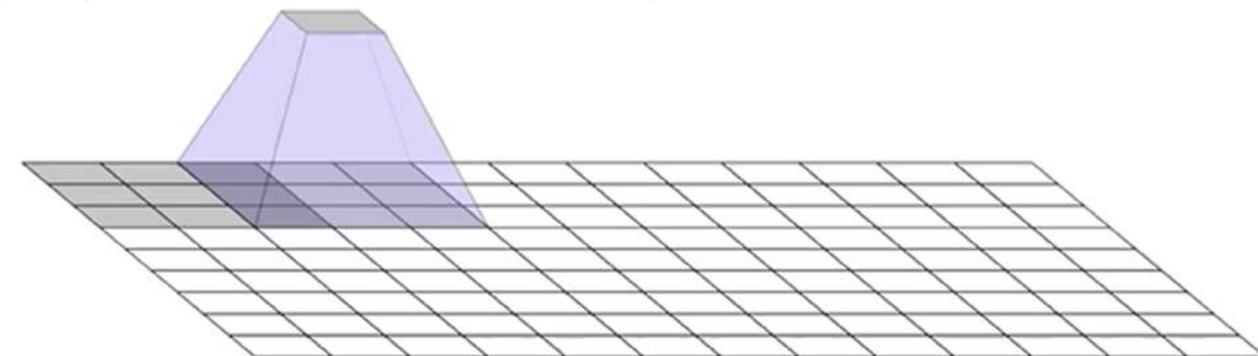
Dumoulin, et al., 2016 [4]

当转置卷积的步长与卷积核不同时，不同位置的输出卷积核叠加不同次数合成，造成棋盘缺陷。

一维示意图：



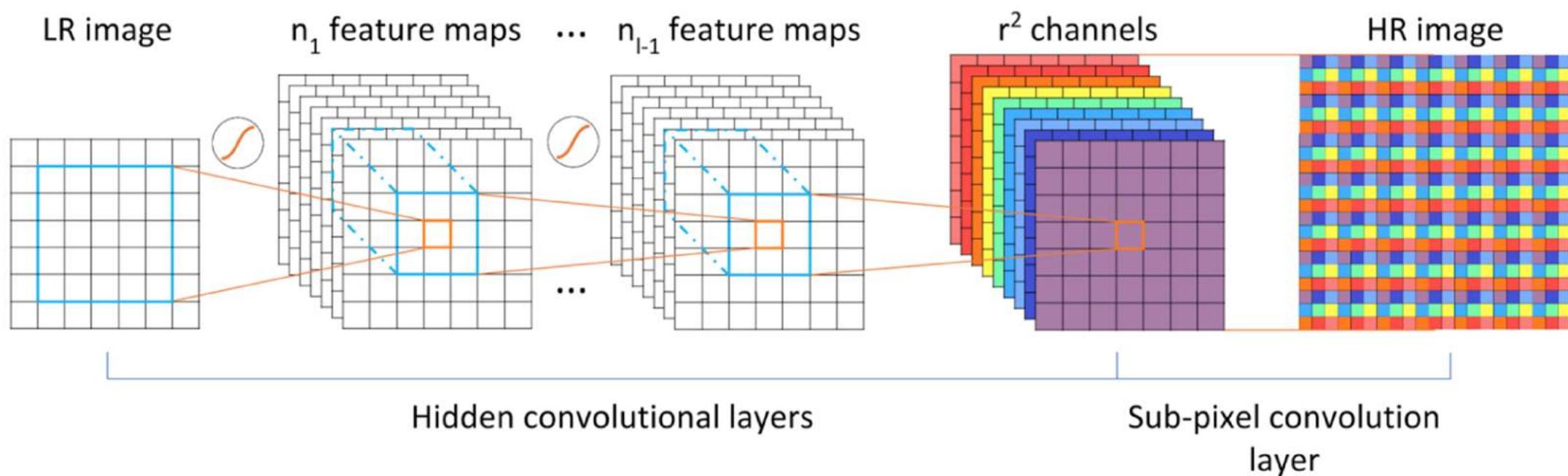
二维示意图：



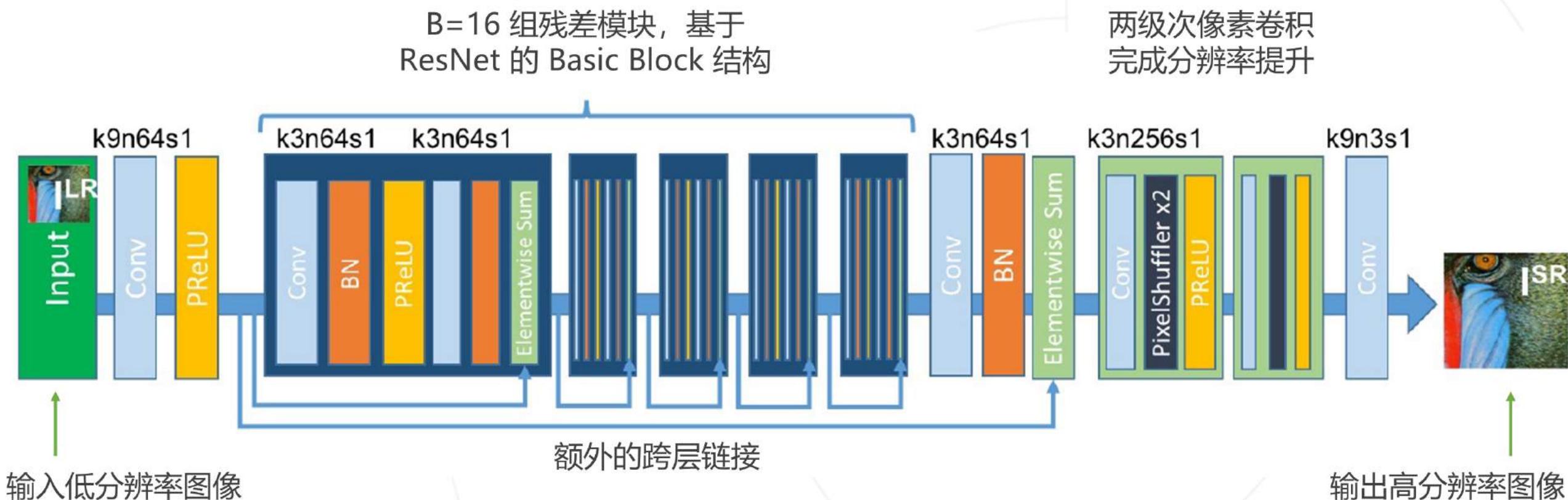
次像素卷积 Subpixel convolution

次像素卷积提供了另一种升采样方法：

为得到 r 倍分辨率的特征图，先通过卷积得到 r^2 通道的特征图，
再将每个位置的 r^2 个响应排列成 $r \times r$ 面积的图像



Twitter 于 2016 年提出的 SRResNet 模型使用类似 ResNet 的网络结构从低分图像生成高分图像。



图像超分辨率中的损失函数

MSE Loss, Perceptual Loss

逐像素计算的损失函数

比较恢复图像与原始高分图像的**每个像素值**，并计算均方误差。

例如：SRCNN 和 FSRCNN 中用到的均方误差损失 (MSE Loss)

感知损失函数

比较恢复图像与原始高分图像的**语义特征**，并计算损失。

语义特征的计算由预先训练的神经网络模型给出。例如：使用在 ImageNet 数据集上预训练好的神经网络计算语义特征。

均方误差损失函数是常用的逐像素计算的损失函数，定义如下：

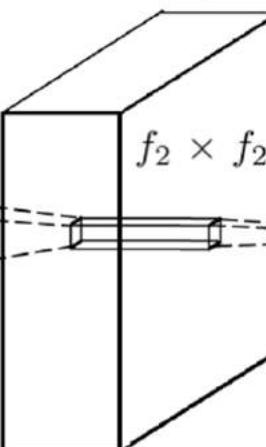
$$L(\Theta) = \frac{1}{n} \sum_{i=1}^n \|F(\mathbf{Y}_i; \Theta) - \mathbf{X}_i\|^2$$

反传损失函数优化网络参数

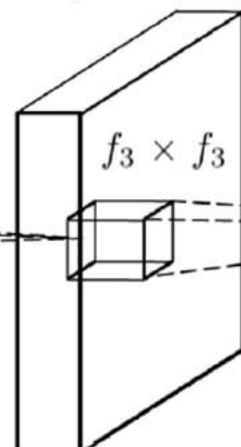


输入

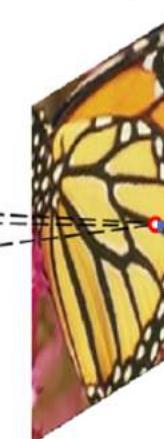
低分辨率图像



多层卷积运算



原始高分辨率图像



输出

恢复高分辨率图像

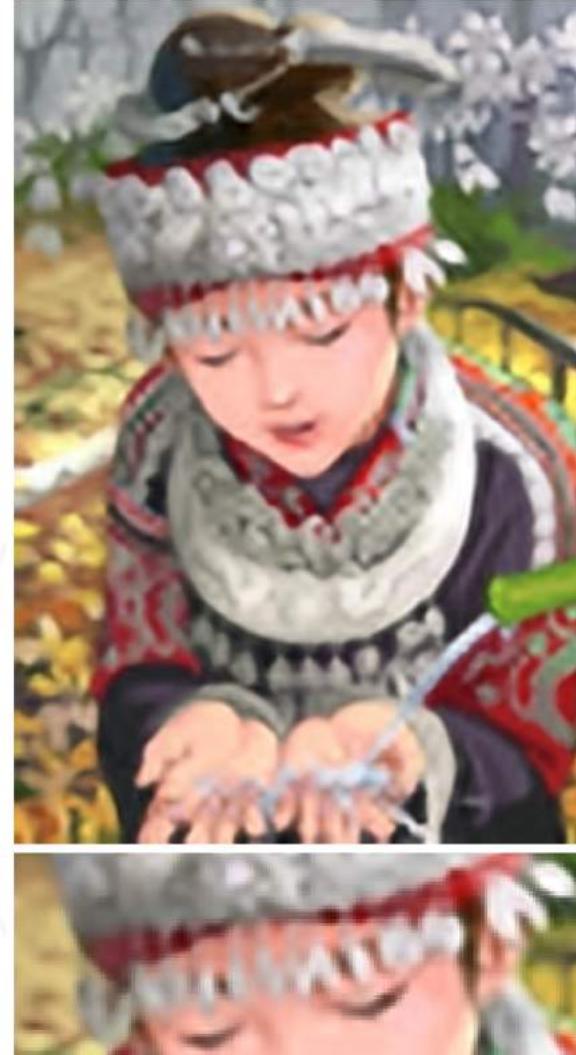
$$L(\Theta)$$



原始高分辨率图像



插值扩大分辨率



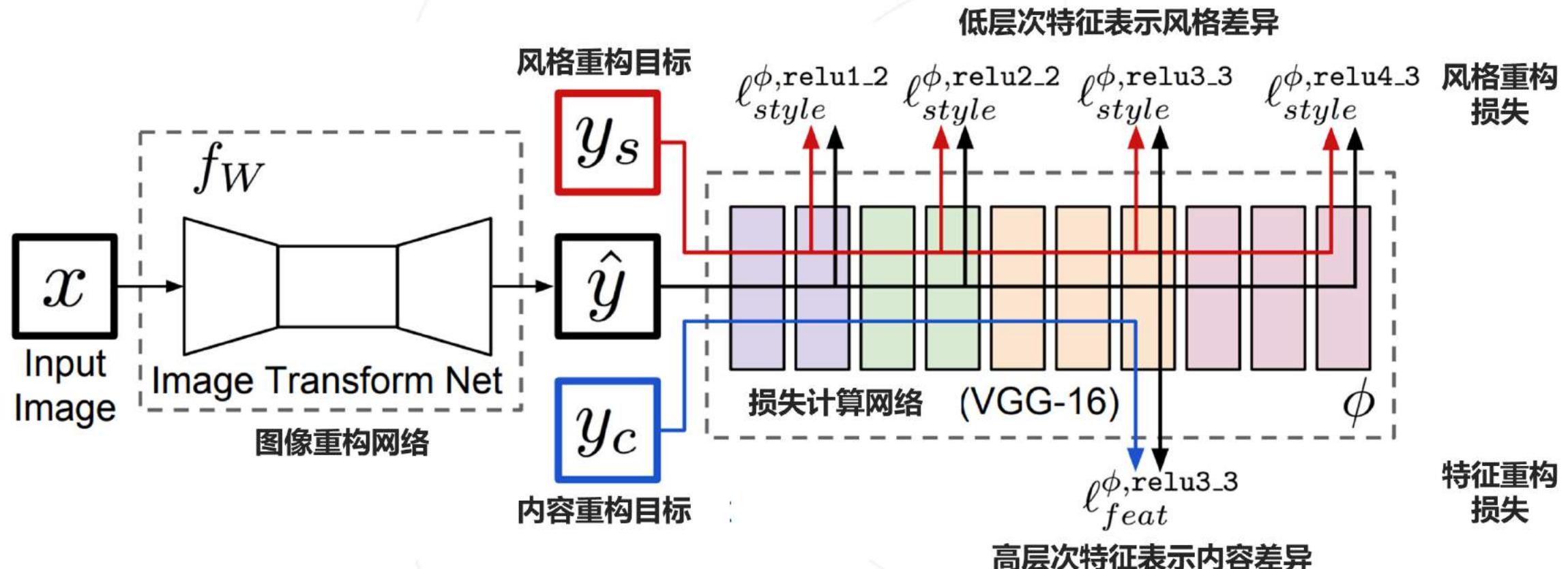
恢复高分辨率图像
用了均方误差

在一些细节，特征上缺失，造成局部的模糊

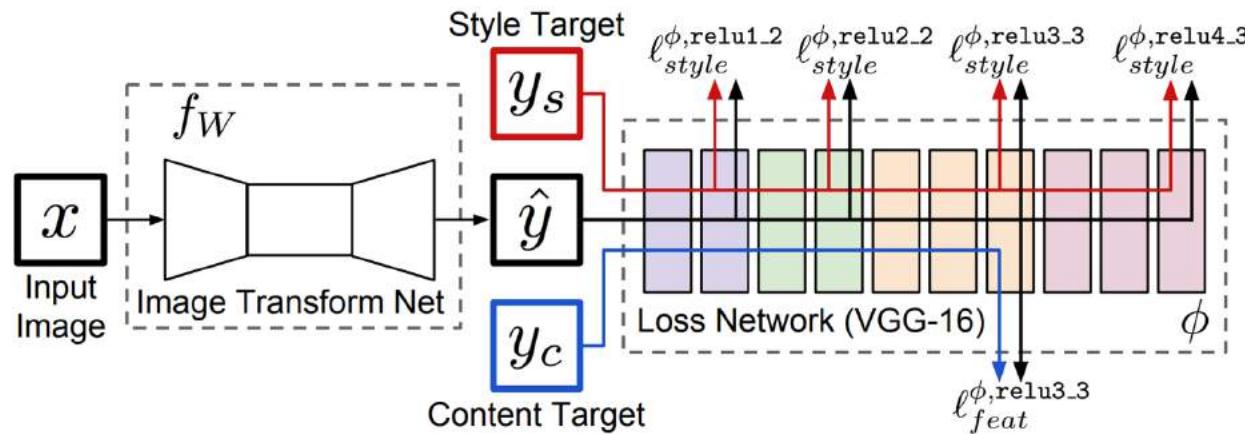
比较恢复图像与原始高分图像的语义特征，并计算损失

损失网络一般是训练图像分类任务得到的模型构成，例如 VGG 网络

损失网络不参与学习，在训练过程中参数保持不变



特征重构损失直接计算特征图之间的欧氏距离



$$\ell_{feat}^{\phi,j}(\hat{y}, y) = \frac{1}{C_j H_j W_j} \|\phi_j(\hat{y}) - \phi_j(y)\|_2^2$$

特征图尺寸大小 $C_j \times H_j \times W_j$

重构图像在第 j 层得到的特征图

目标图像在第 j 层得到的特征图

减少特征重构损失

输出图像在感知理解上越接近目标图像

不要求两者完全匹配



原始高分辨率图像



插值扩大分辨率



恢复高分辨率图像
用了均方误差



恢复高分辨率图像
用了感知损失

对抗生成网络介绍

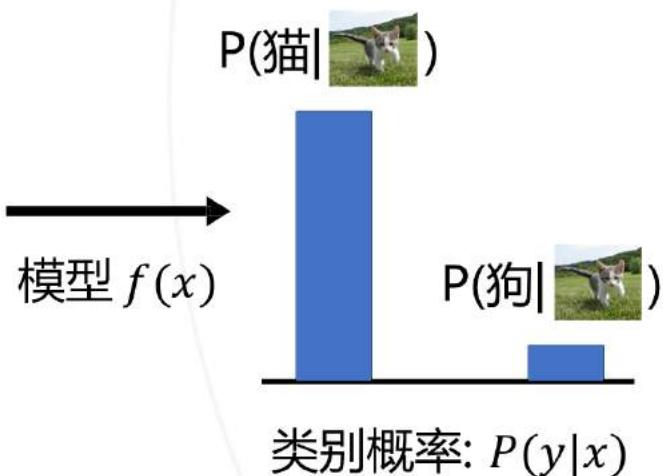
Generative Adversarial Network

对抗生成网络是一种基于神经网络的**无监督学习模型**，可以建模数据的分布，并通过采样生成新数据。

监督学习

学习数据对之间的映射关系

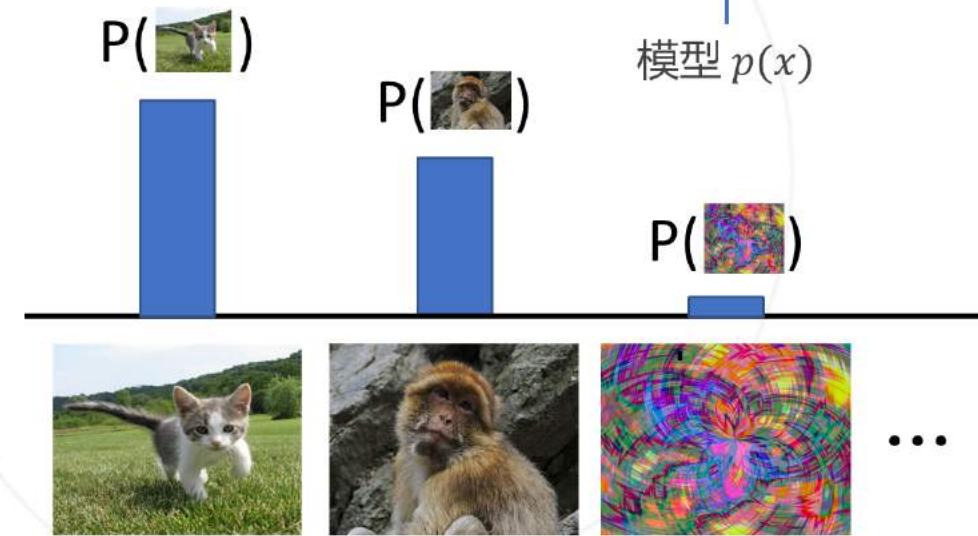
分类: $P(y|x) = f(x)$



无监督学习

学习数据的内在结构

概率密度估计: $p(x)$



使用 GAN 生成图像

估计人脸图像的分布 p_{face} , 再从 p_{face} 中采样即得到人脸图像。



输入视频中的马



输出视频中变为斑马



使用普通损失函数训练的模型 使用对抗训练方法训练的模型

细节还有些模糊

细节恢复得更好

bicubic
(21.59dB/0.6423)



SRResNet
(23.53dB/0.7832)



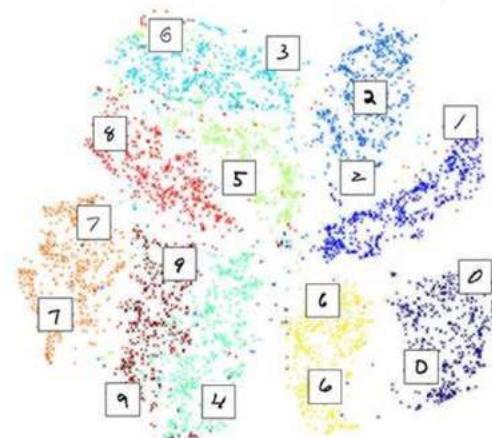
SRGAN
(21.15dB/0.6868)



original



- 图像是高维空间中的数据点，但由于真实图像中的像素值不能取任意值，这些数据点不会充满整个高维空间，仅分布在一个相对低维的结构上。
- 可以通过数学函数，将数据映射到低维空间上，也可以从低维空间逆映射回高维空间。
- 映射函数 G 可以由神经网络构成，其中的参数可以基于数据学习出来。



低维空间 R^2



0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9

高维空间 $R^{28 \times 28}$

前提：数据在高维空间形成一个概率分布 p_{data} ，通常不可解析表示

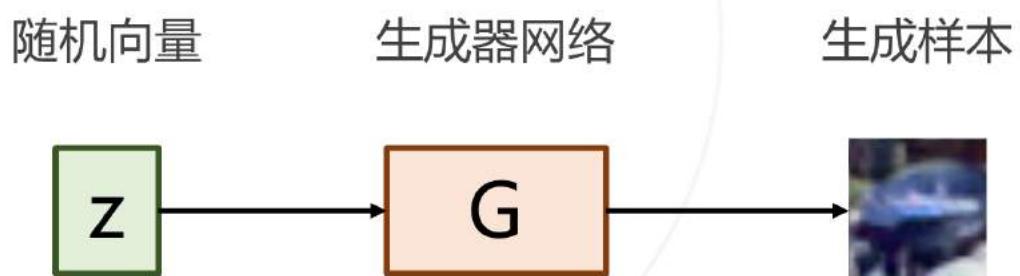
问题：如何表示 p_{data} ，并从中采样以生成新的数据？

解决思路：

1. 基于简单的分布产生随机向量 $z \sim p_z$ ，例如低维的均匀或高斯分布
2. 使用神经网络 G 对随机变量进行变换 $x = G(z)$ ，产生复杂的概率分布 p_x ， G 称为生成器网络
3. 调整 G 网络的参数，使 p_x 拟合 p_{data}

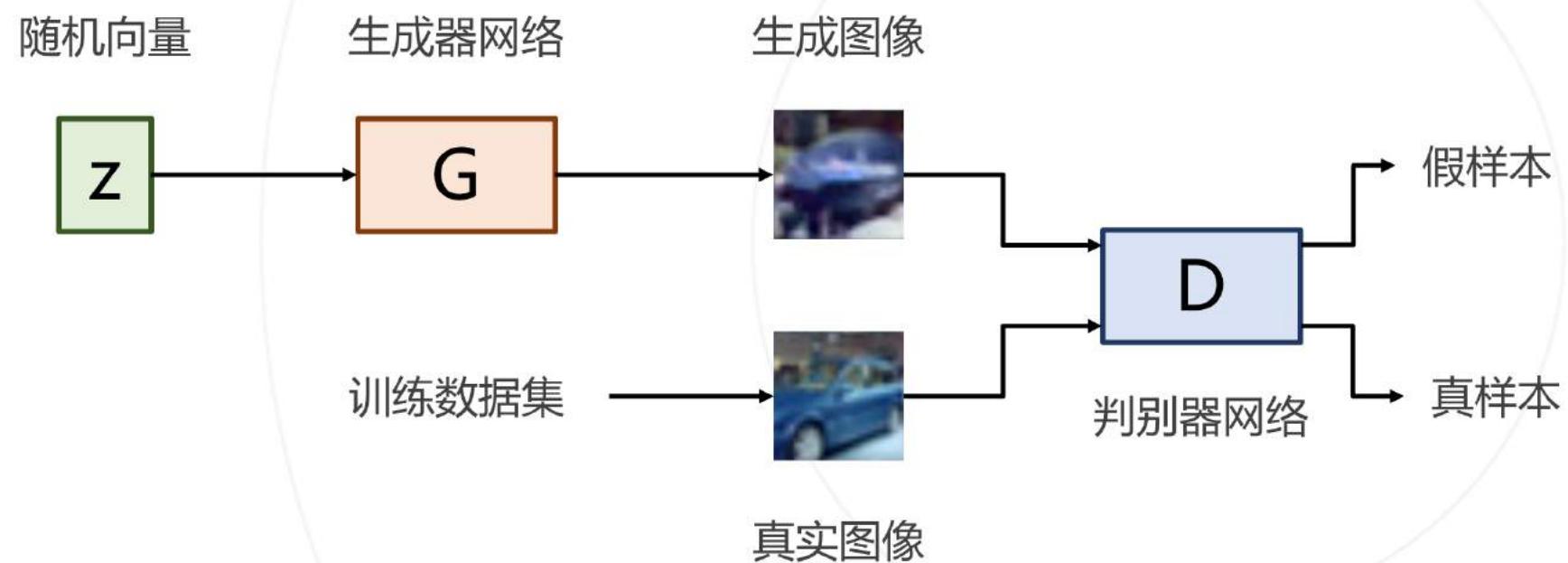
达到目的：

1. p_z 和 G 隐式表示 p_{data}
2. 从 p_z 采样，再经过 G 网络变换，等效从 p_{data} 采样



问题：我们希望 p_x 与 p_{data} 近似，但二者没有闭式表达，无法直接计算“差距”或损失函数。

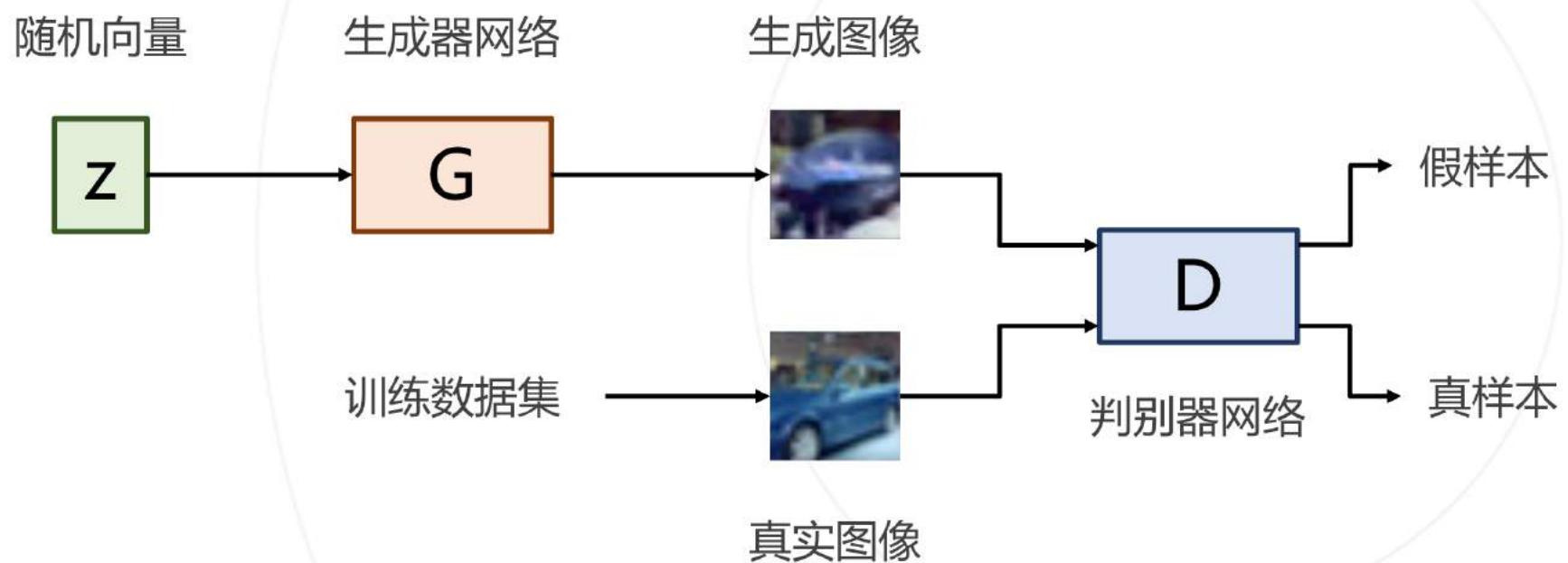
思路：如果 p_x 与 p_{data} 有差别，那么它们的样本就可以区分 → 使用一个分类网络区分两类样本，将分类正确率作为两个概率分布的“差距”。二者越接近，分类正确率应该越低。



判别器网络 D 和生成器网络 G 采用对抗的方式进行训练：

- 训练 D 网络时降低分类损失，尽力分辨 G 网络产生的假样本
- 训练 G 网络时提高分类损失，尽力迷惑 D 网络，使之无法区分真假样本

二者相互对抗相互进步，最优状态下 G 网络可以生成以假乱真的样本



$$\min_{\mathbf{G}} \max_{\mathbf{D}} \left(E_{x \sim p_{\text{data}}} [\log \mathbf{D}(x)] + E_{z \sim p(z)} [\log (1 - \mathbf{D}(\mathbf{G}(z)))] \right)$$

判别器希望对真实样本输出 $D(x) = 1$

判别器希望对生成样本输出 $D(x) = 0$

生成器希望对生成样本输出 $D(x) = 1$

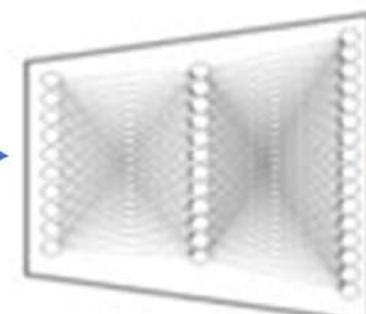
- 对于给定的 G 网络，训练出最佳判别器网络，记录对应的分类损失（的负值）
- 在所有可能的 G 网络中，找到使得上述损失最大（对应负值最小）的 G 网络。
- 可以证明，最优 G 网络满足 $p_G = p_{\text{data}}$



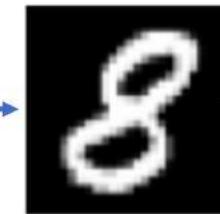
手写数字图像数据集



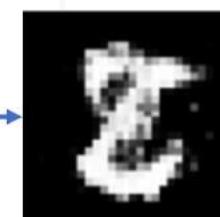
随机向量



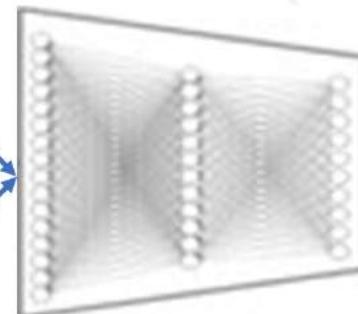
生成器网络



真实数据



生成数据



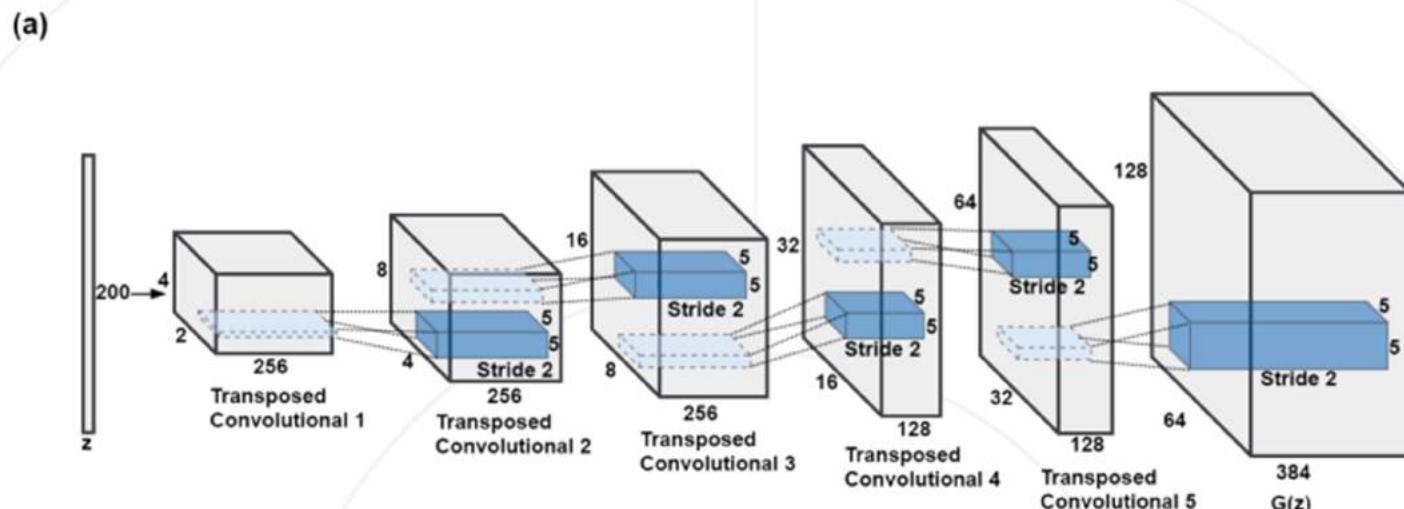
判别器网络



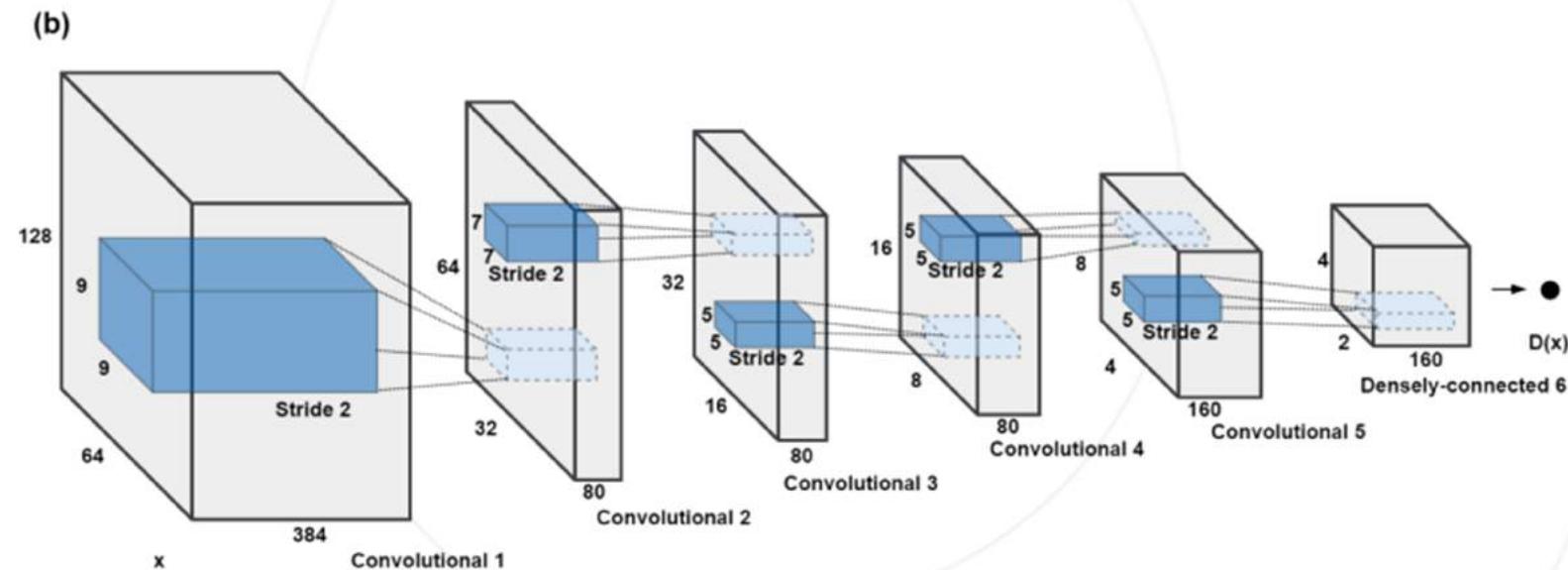
假
真

基于多层感知器的 GAN 模型生成手写数字图像

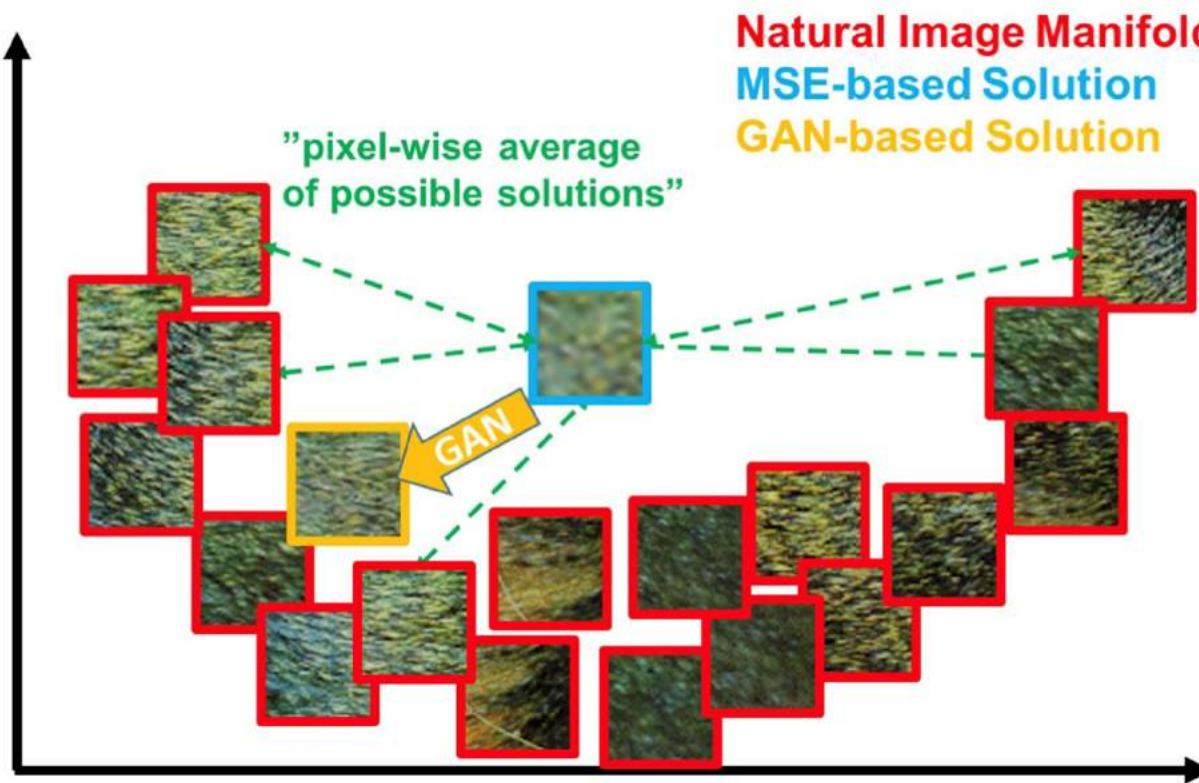
DCGAN 的生成器网络由转置卷积网络构成，可以从特征向量开始，逐层升采样，得到一定分辨率的图像



DCGAN 的判别器网络由普通的卷积网络构成，完成图像的真假分类



逐像素的 MSE 损失会产生过于平滑（模糊）的图像，因为 MSE 倾向给出所有可能高分图像的平均值作为最优解。

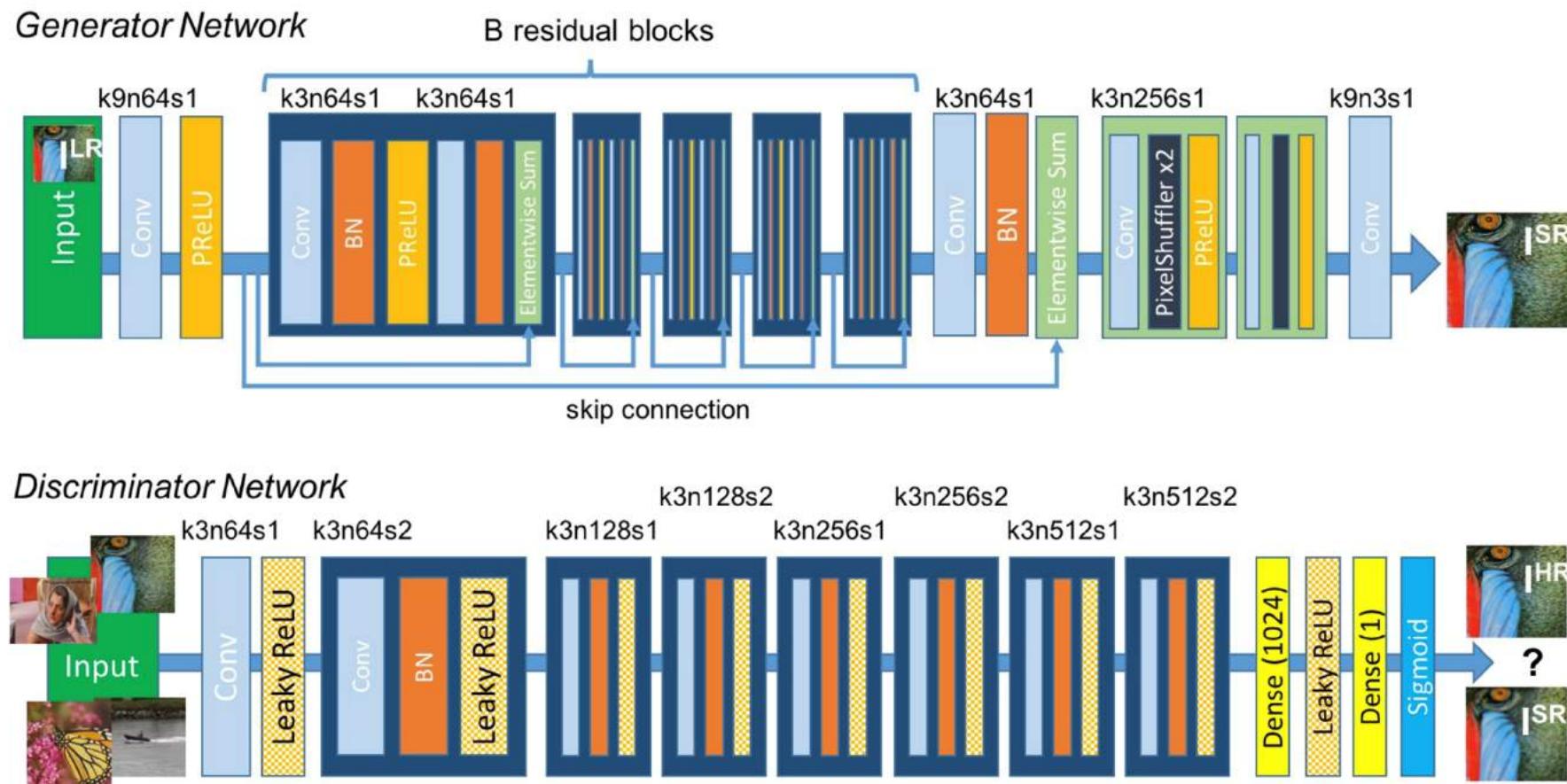


使用 GAN 可以鼓励网络产生更为真实，
细节更为丰富的高分辨率图像。

超分辨率模型中的 GAN

SRGAN、ESRGAN

SRGAN 在 SRResNet 的基础上额外增加了判别器网络，用于区分训练集中的高分图像（真实图像）以及 SRResNet 恢复的高分图像（虚假图像）



SRGAN 在训练生成器时同时优化三个损失

$$l^{SR} = \underbrace{l_X^{SR}}_{\text{content loss}} + \underbrace{10^{-3} l_{Gen}^{SR}}_{\text{adversarial loss}}$$

同时优化判别器网络，与生成器产生对抗

$$\begin{aligned} & \min_{\theta_G} \max_{\theta_D} \mathbb{E}_{I^{HR} \sim p_{\text{train}}(I^{HR})} [\log D_{\theta_D}(I^{HR})] + \\ & \mathbb{E}_{I^{LR} \sim p_G(I^{LR})} [\log(1 - D_{\theta_D}(G_{\theta_G}(I^{LR})))] \end{aligned}$$

MSE Loss 鼓励恢复图像的像素值与高分图像相近

$$l_{MSE}^{SR} = \frac{1}{r^2 WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} (I_{x,y}^{HR} - G_{\theta_G}(I^{LR})_{x,y})^2$$

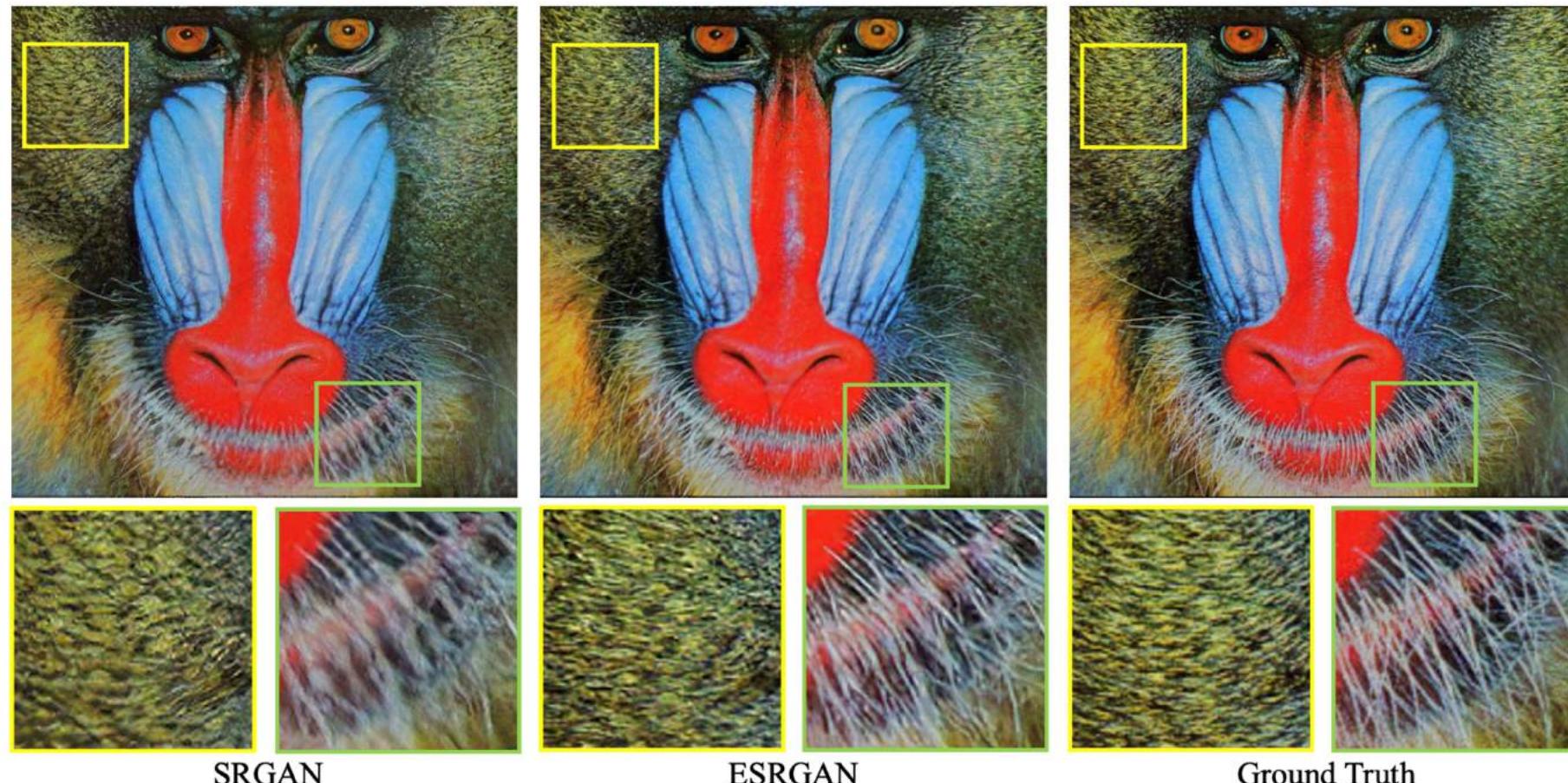
Perceptual Loss 鼓励恢复图像的高层特征与高分图像相近

$$l_{VGG/i.j}^{SR} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(G_{\theta_G}(I^{LR}))_{x,y})^2$$

Adversarial Loss 鼓励恢复图像更真实、细节更丰富

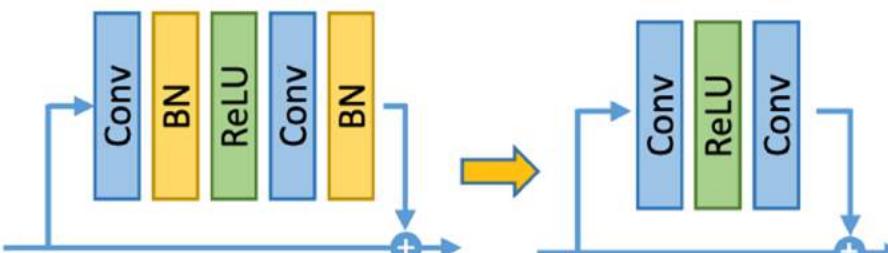
$$l_{Gen}^{SR} = \sum_{n=1}^N -\log D_{\theta_D}(G_{\theta_G}(I^{LR}))$$

Enhanced SRGAN (ESRGAN) 从网络结构、感知损失、对抗损失三个角度对 SRGAN 进行了全面改进，在超分辨率效果上取得了很大的提升，同时获得了 PIRM2018 超分辨率挑战赛冠军。



更强的能力，更容易训练

Residual Block (RB)

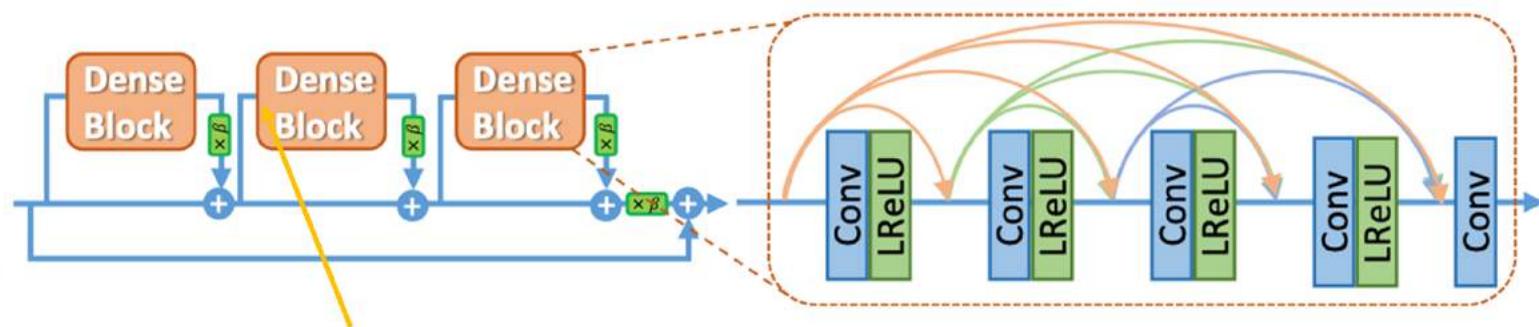


SRGAN

移除 BN 层

BN 在超分辨率网络（尤其是深层网络以及使用 GAN 的时候）会产生一些负面效果

Residual in Residual Dense Block (RRDB)



增加更多跨层链接

引入衰减因子

残差分支和主分支求和之前乘以衰减系数 β ，提高训练稳定性

形成 Dense Block 结构

RaGAN 鼓励生成网络产生变化分明的图像边界与细节纹理。

$$D(x_r) = \sigma(C(\text{Real})) \rightarrow 1 \quad \text{Real?}$$

$$D(x_f) = \sigma(C(\text{Fake})) \rightarrow 0 \quad \text{Fake?}$$

a) Standard GAN



$$D_{Ra}(x_r, x_f) = \sigma(C(\text{Real}) - \mathbb{E}[C(\text{Fake})]) \rightarrow 1$$

$$D_{Ra}(x_f, x_r) = \sigma(C(\text{Fake}) - \mathbb{E}[C(\text{Real})]) \rightarrow 0$$

b) Relativistic GAN

计算 mini batch 中所有真实样本和伪造样本的均值

More realistic than fake data?

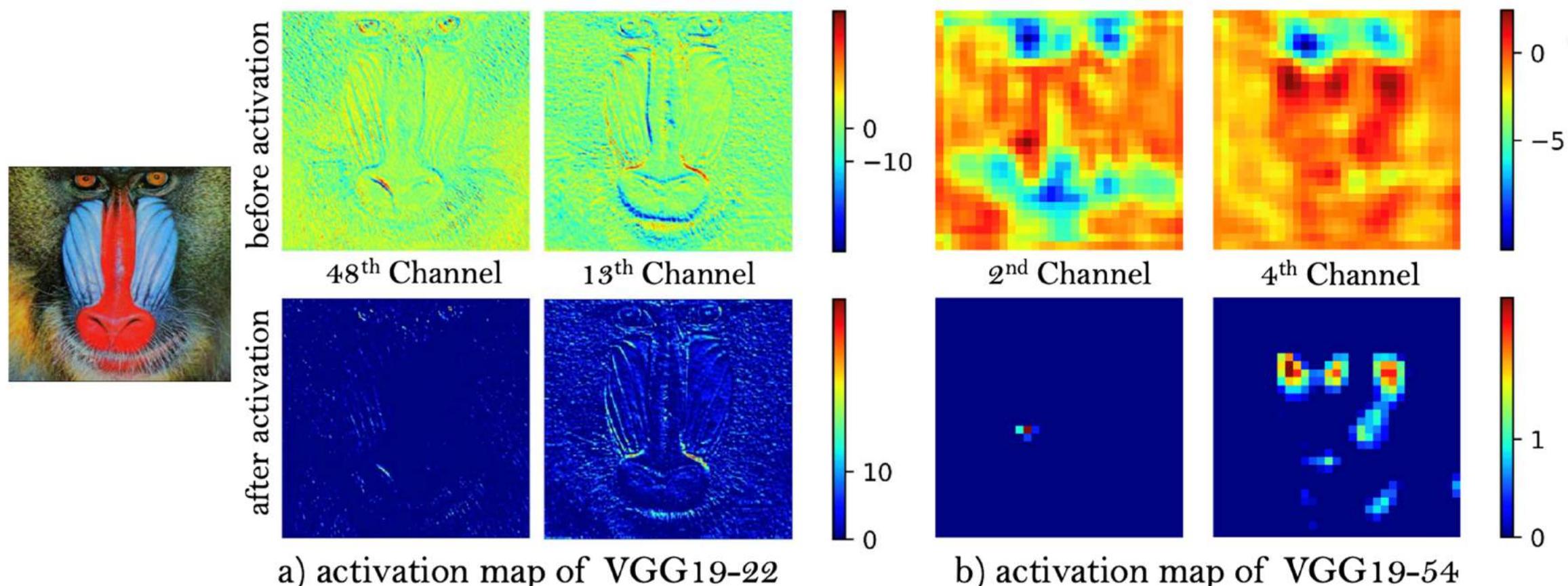
Less realistic than real data?

GAN 中的判别器尝试区别真实图像
与伪造图像

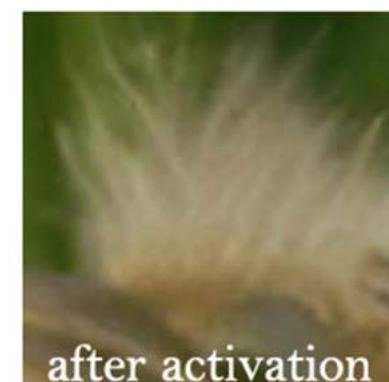
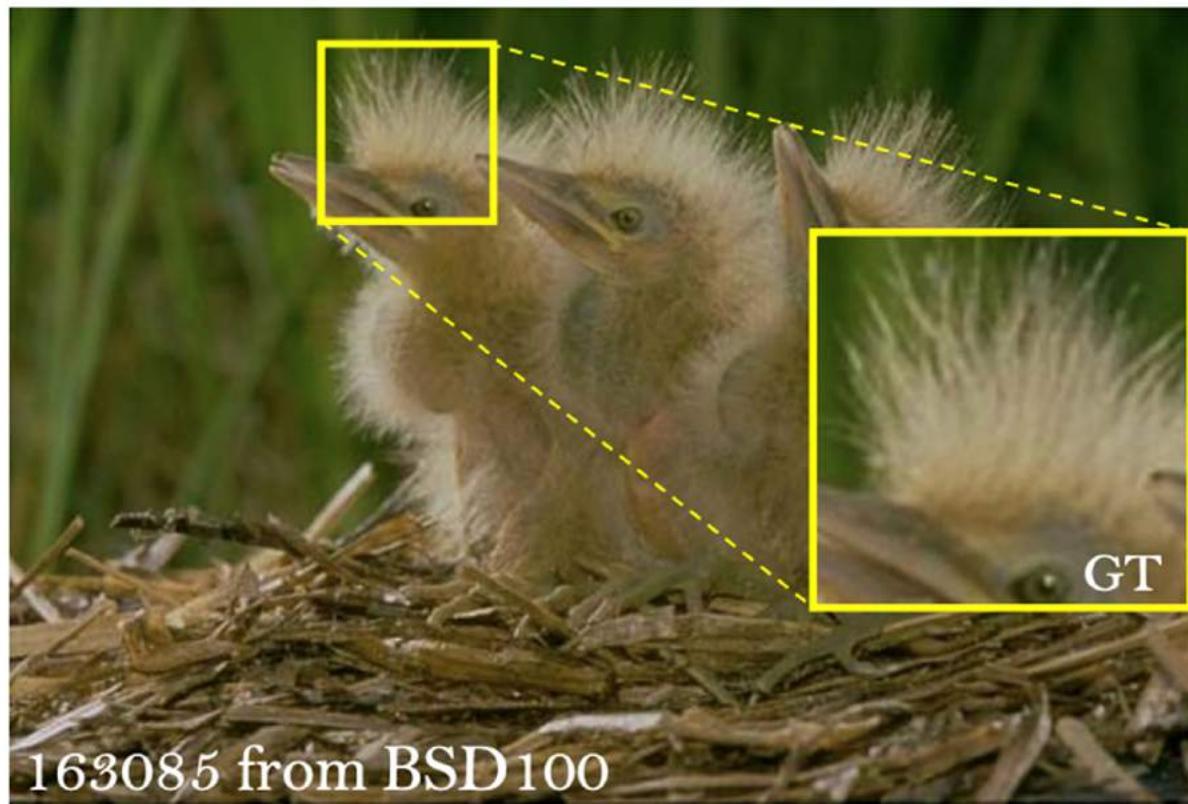
RaGAN 中的判别器尝试判断图像之间的相对真实性

- 真实图像是否比伪造图像更真实?
- 伪造图像是否比真实图像更虚假?

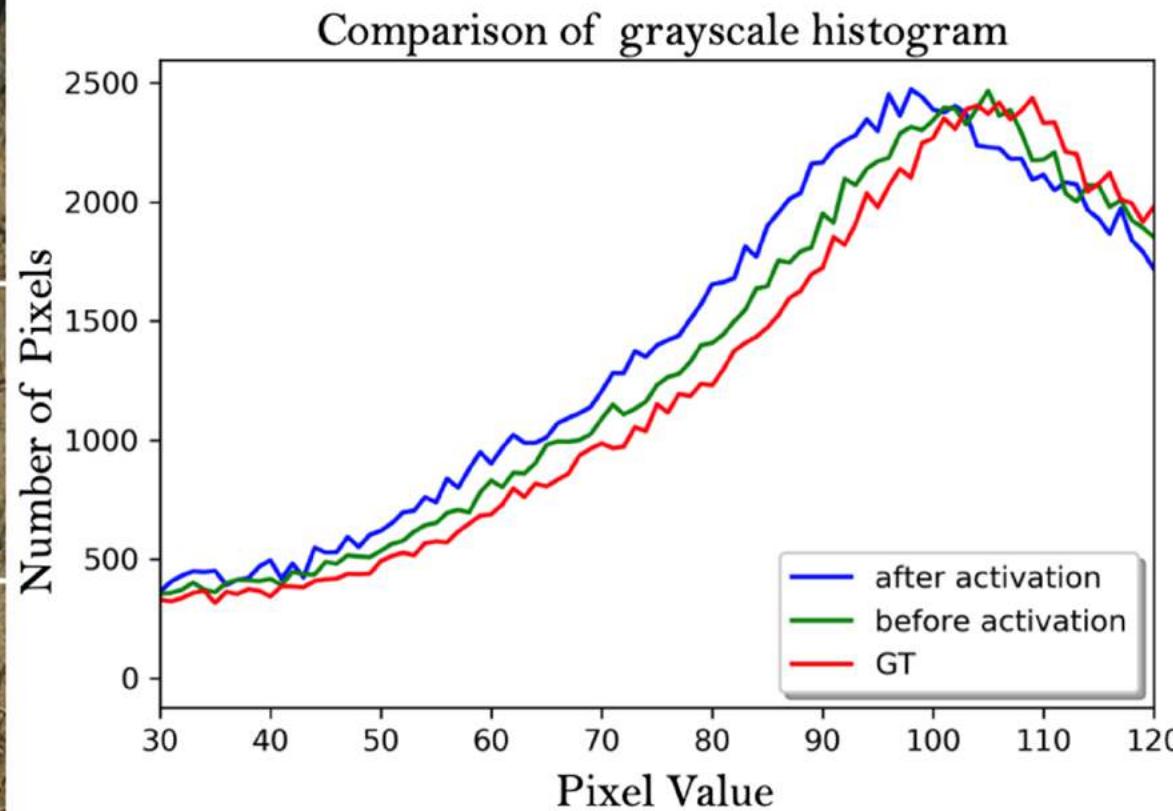
原因1：激活后的响应值更稀疏，削弱监督信号的强度。



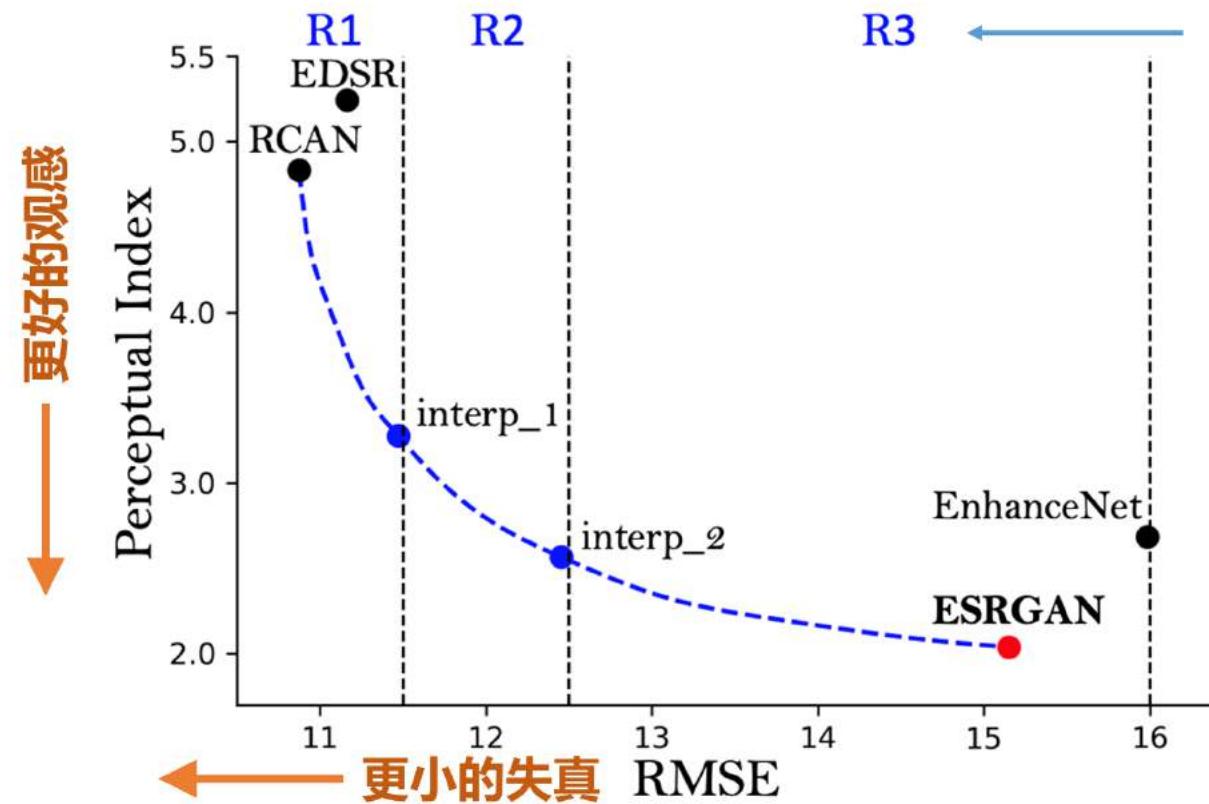
使用更为密集的、激活前的响应计算感知损失有助于恢复图像中的细节信息。



原因2：使用激活后的响应容易使恢复图像产生亮度上的偏移。



观感-失真平面 (Perception-Distortion Plane)



每个失真区间内部比较，
观感指数越低越好

Results on PIRM self val dataset

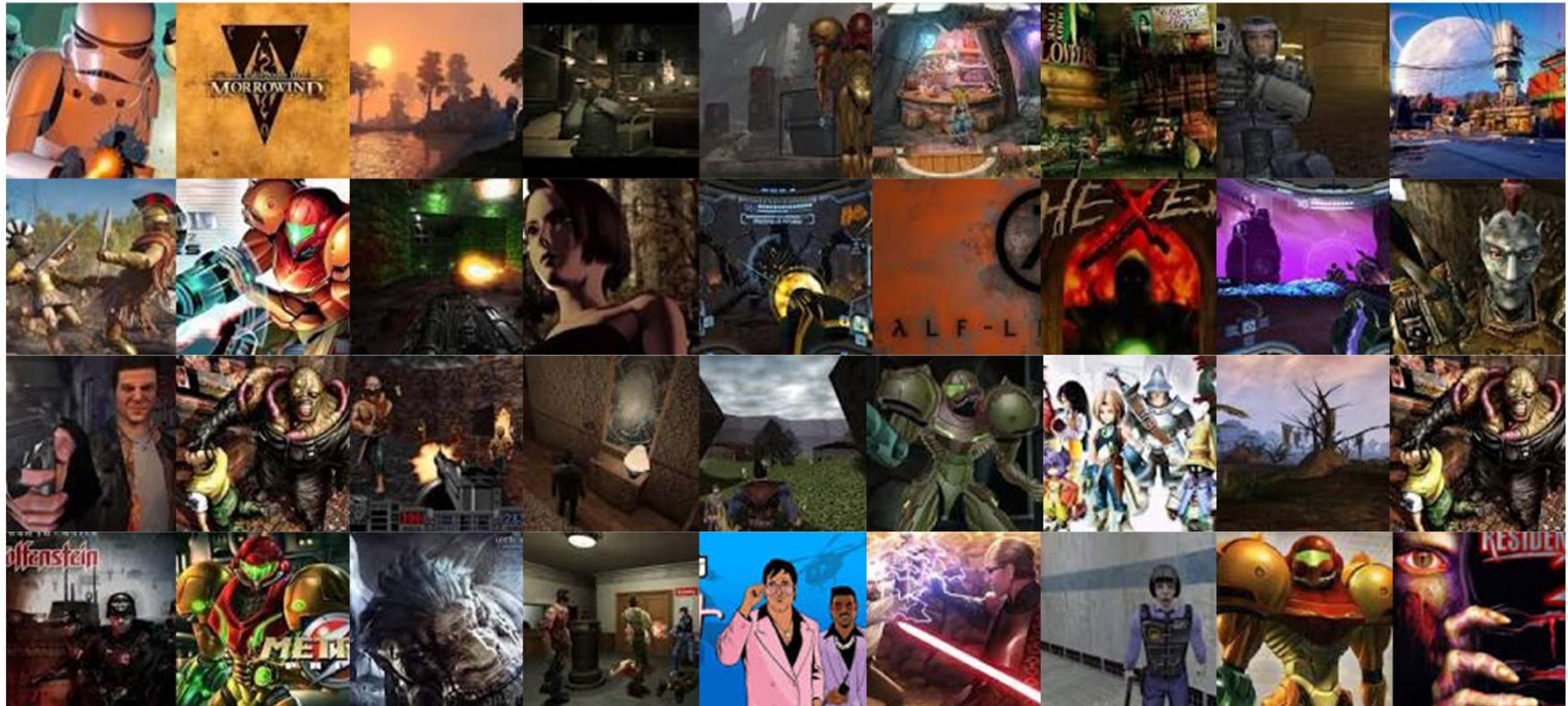
Method	PI	RMSE
ESRGAN	2.040	15.15
interp_2	2.567	12.45
EnhanceNet	2.688	15.99
interp_1	3.279	11.47
RCAN	4.831	10.87
EDSR	5.243	11.16

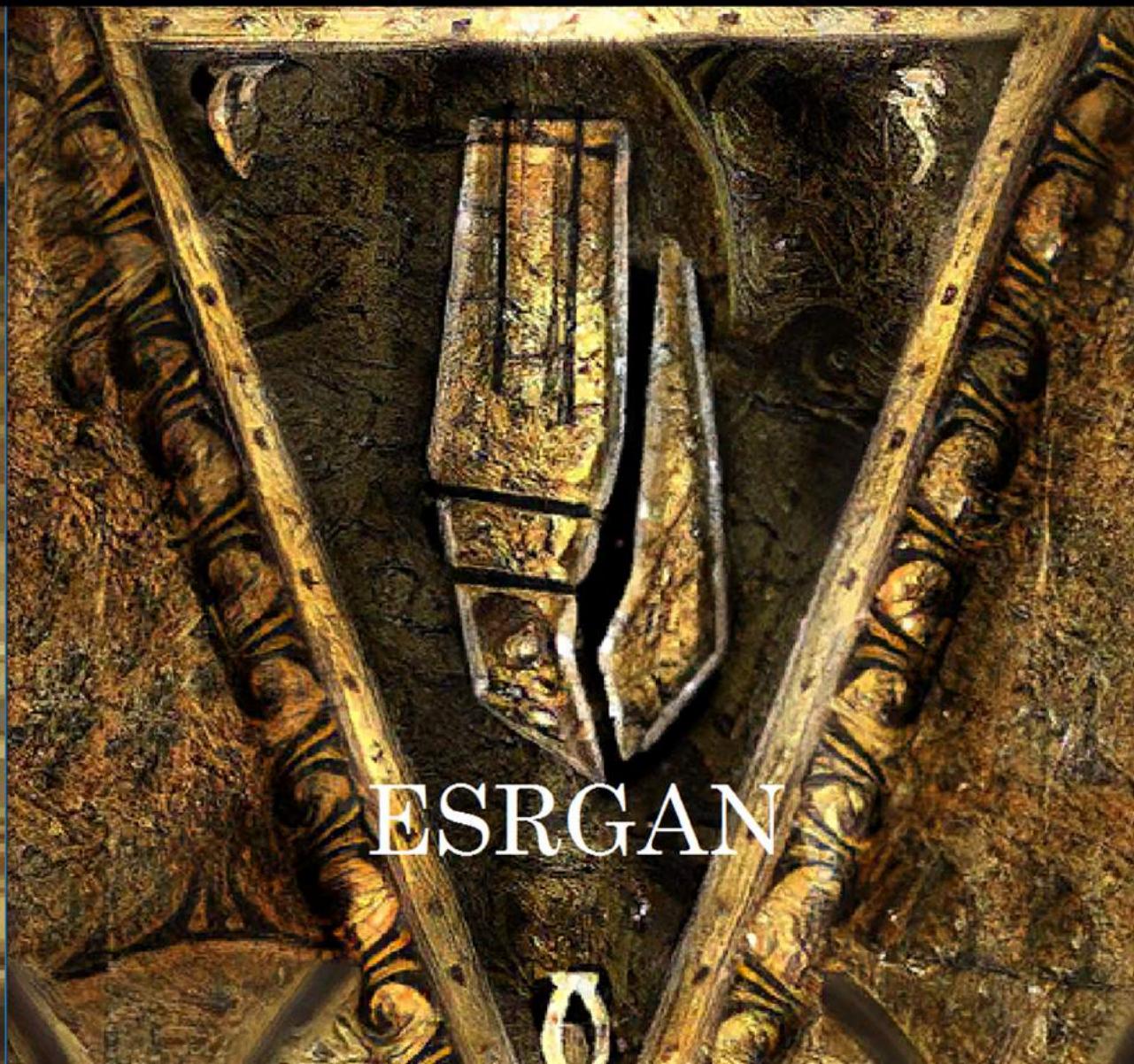
$$\text{perceptual index} = \frac{1}{2}((10 - \text{Ma}) + \text{NIQE})$$

Ma = C. Ma et al., Learning a no-reference quality metric for single-image super-resolution, CVIU 2017

NIQE = A. Mittal et al., Making a “completely blind” image quality analyzer. IEEE Signal Process. Lett 2017

ESRGAN 修复经典游戏





视频超分辨率

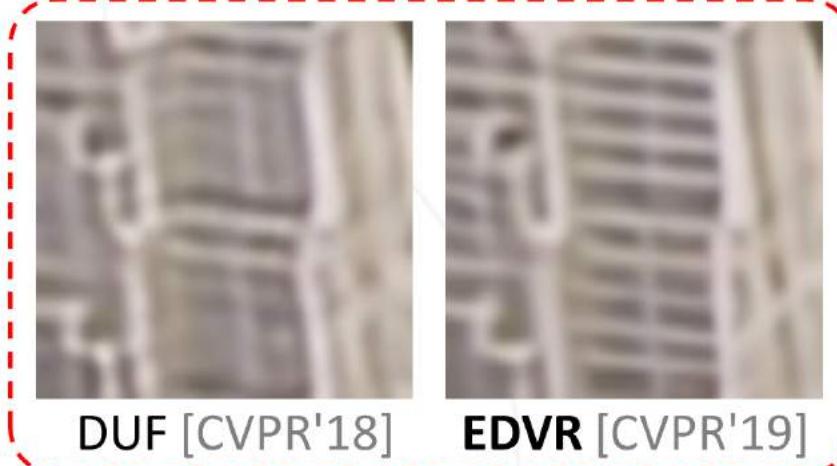
Video Super Resolution



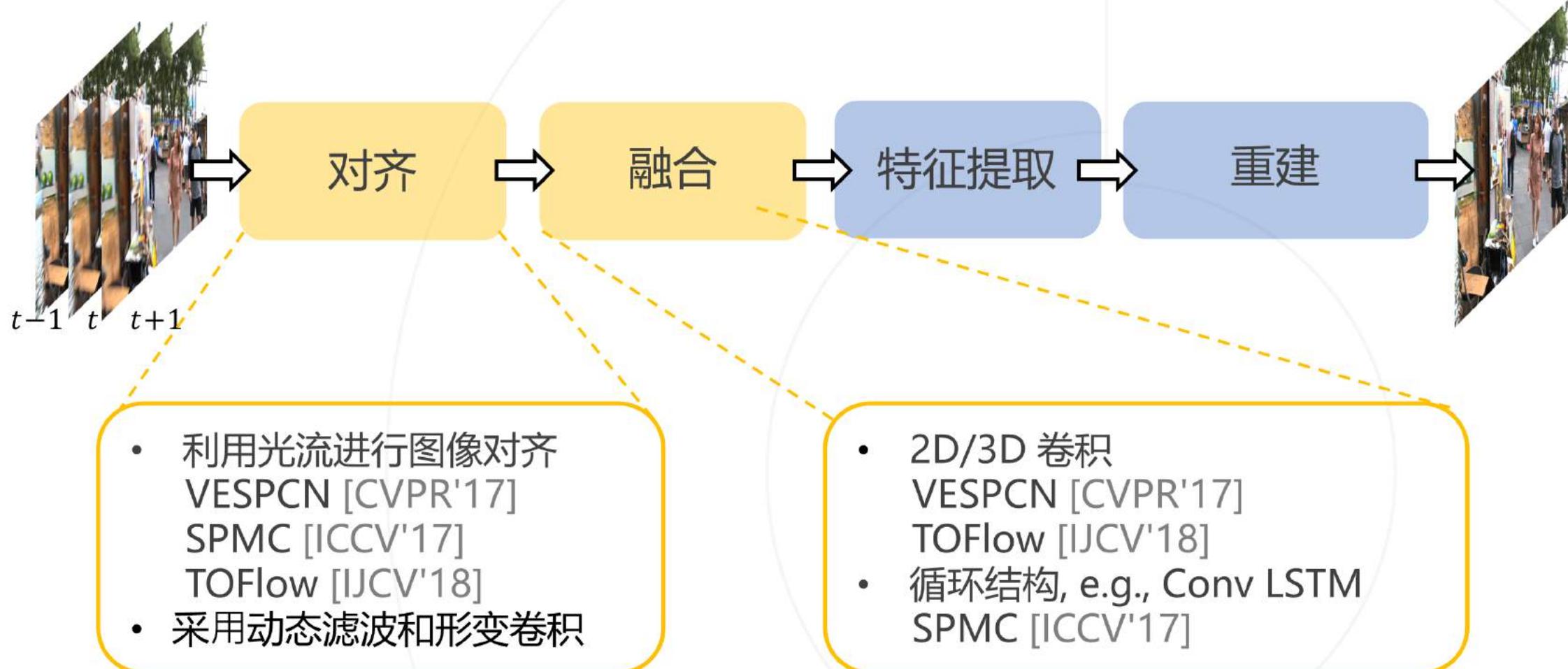
图像超分

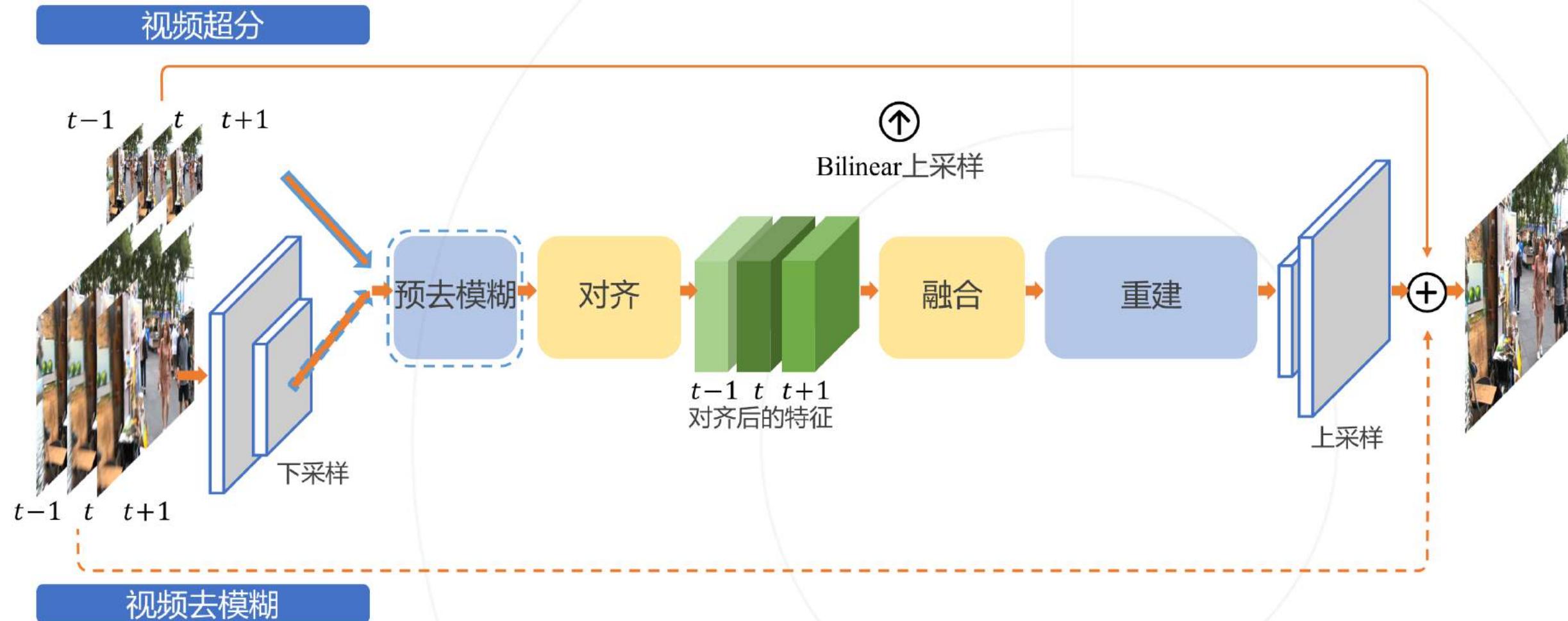


视频超分

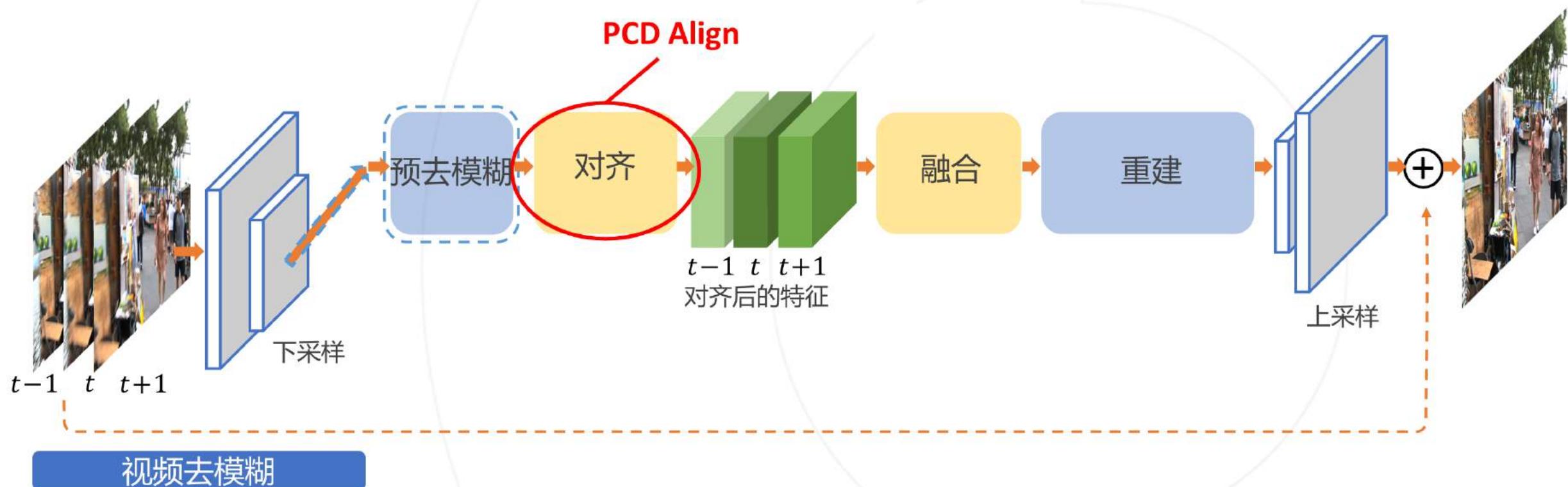


Vid4数据集中的测试图像

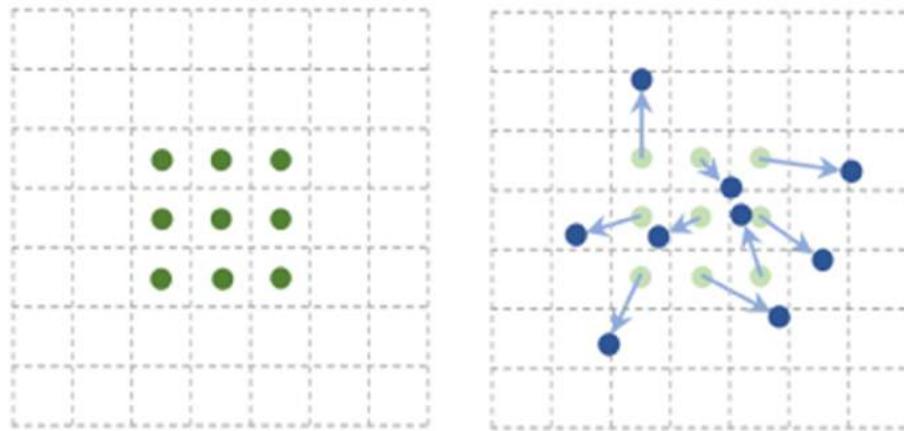
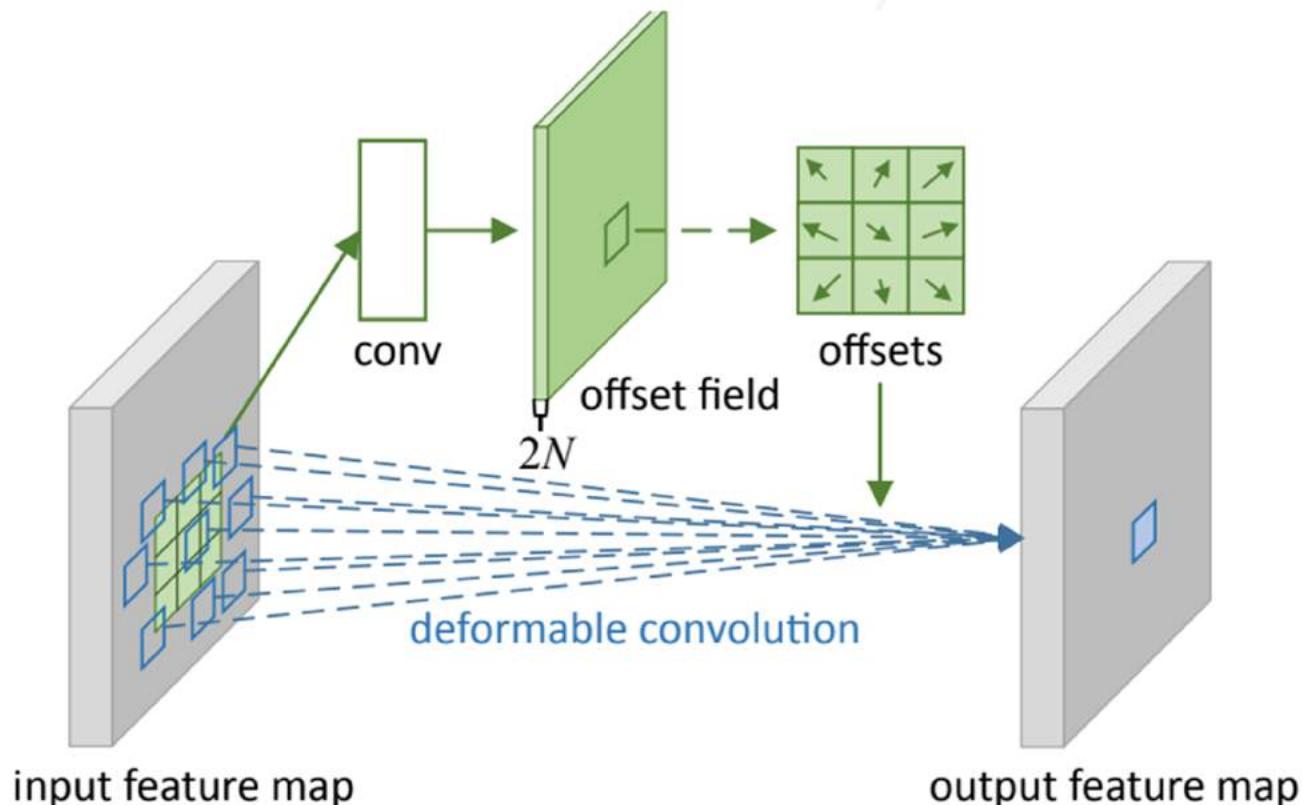




- 适用于不同视频复原任务的通用框架
- PCD: 通过金字塔级联变形对齐处理大的运动，使用形变卷积以由粗到细的方式在特征级别进行帧对齐

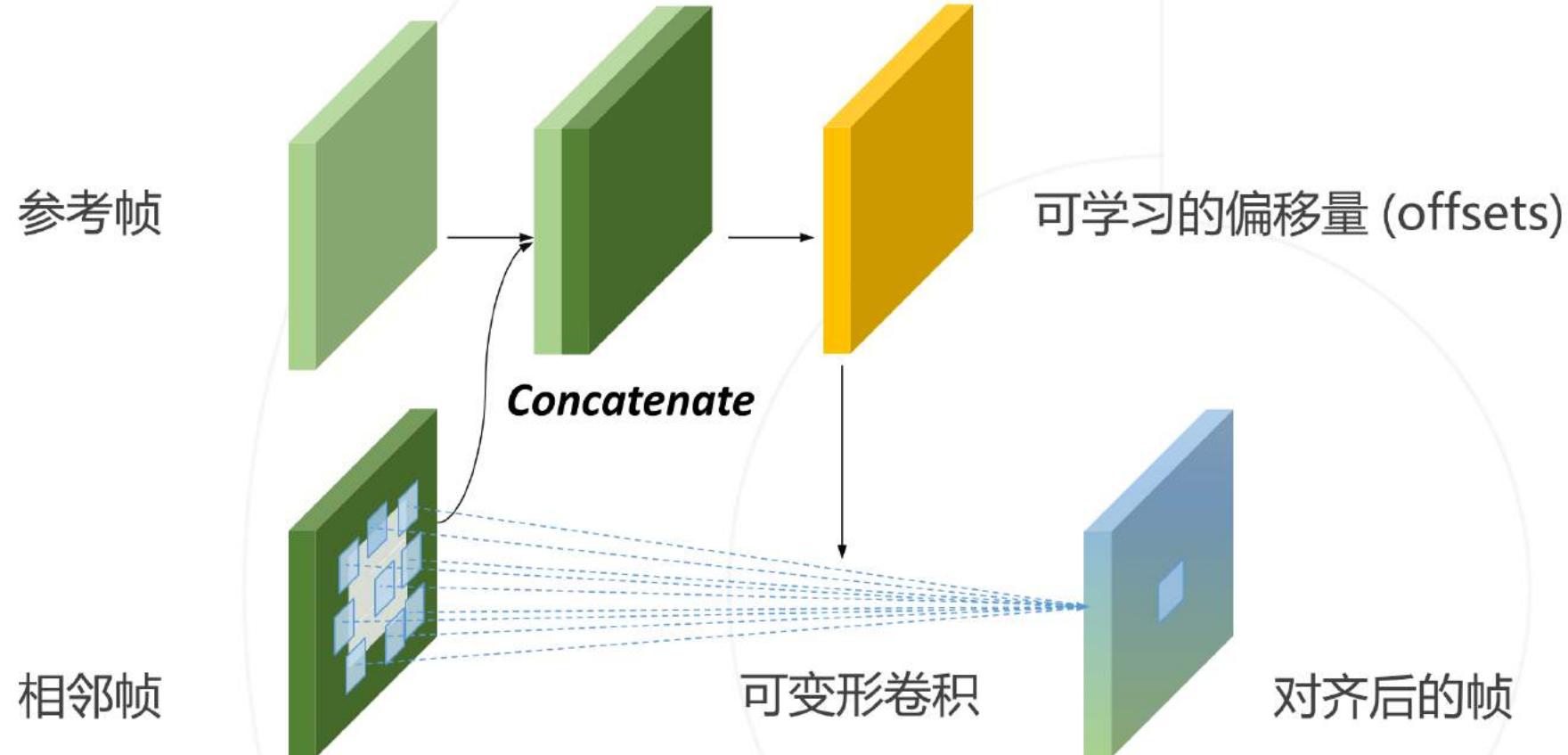


可变形卷积 Deformable Convolution



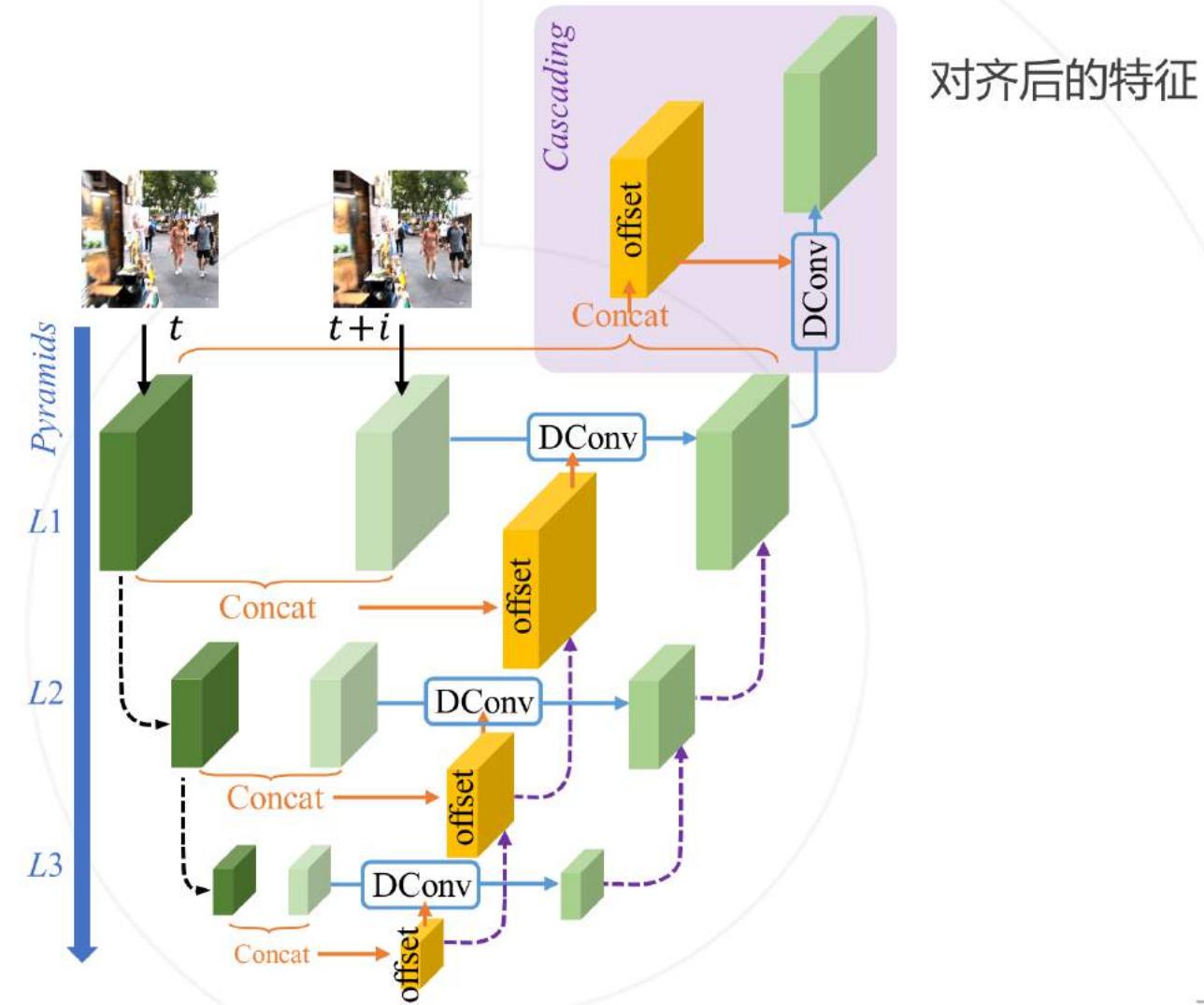
标准卷积的常规采样
网格 (绿点)

在可变形卷积中具有
增加的偏移量 (浅蓝
色箭头) 使采样位置
能够灵活变化 (深蓝
色点)。

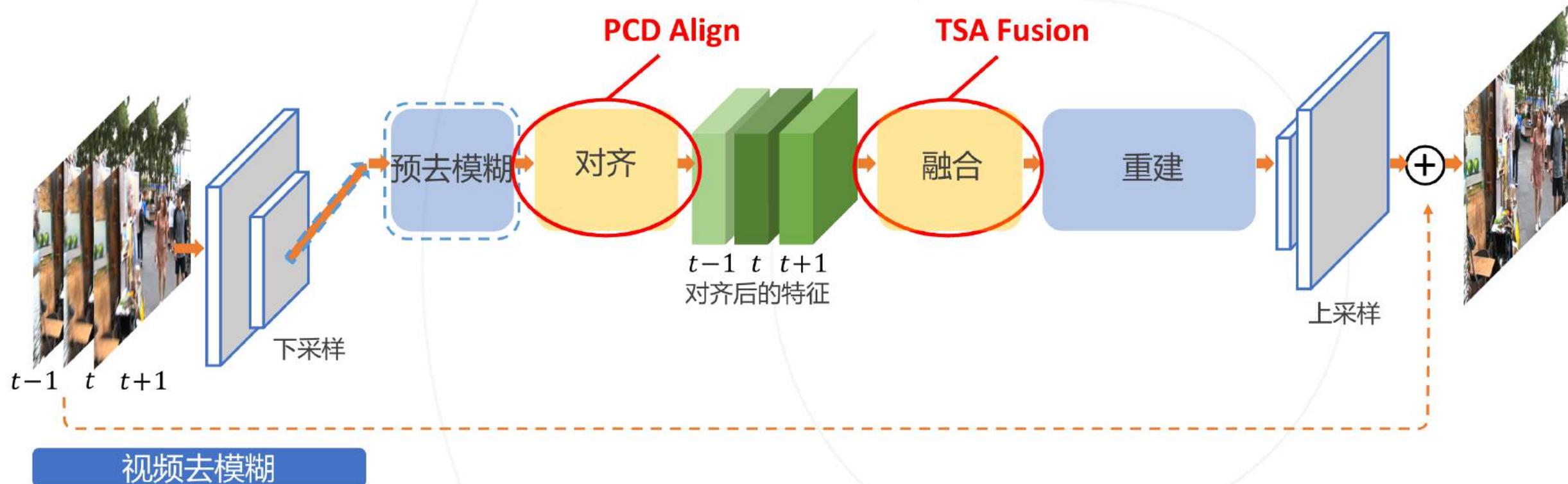


PCD: Alignment with Pyramid, Cascading and Deformable Convolution

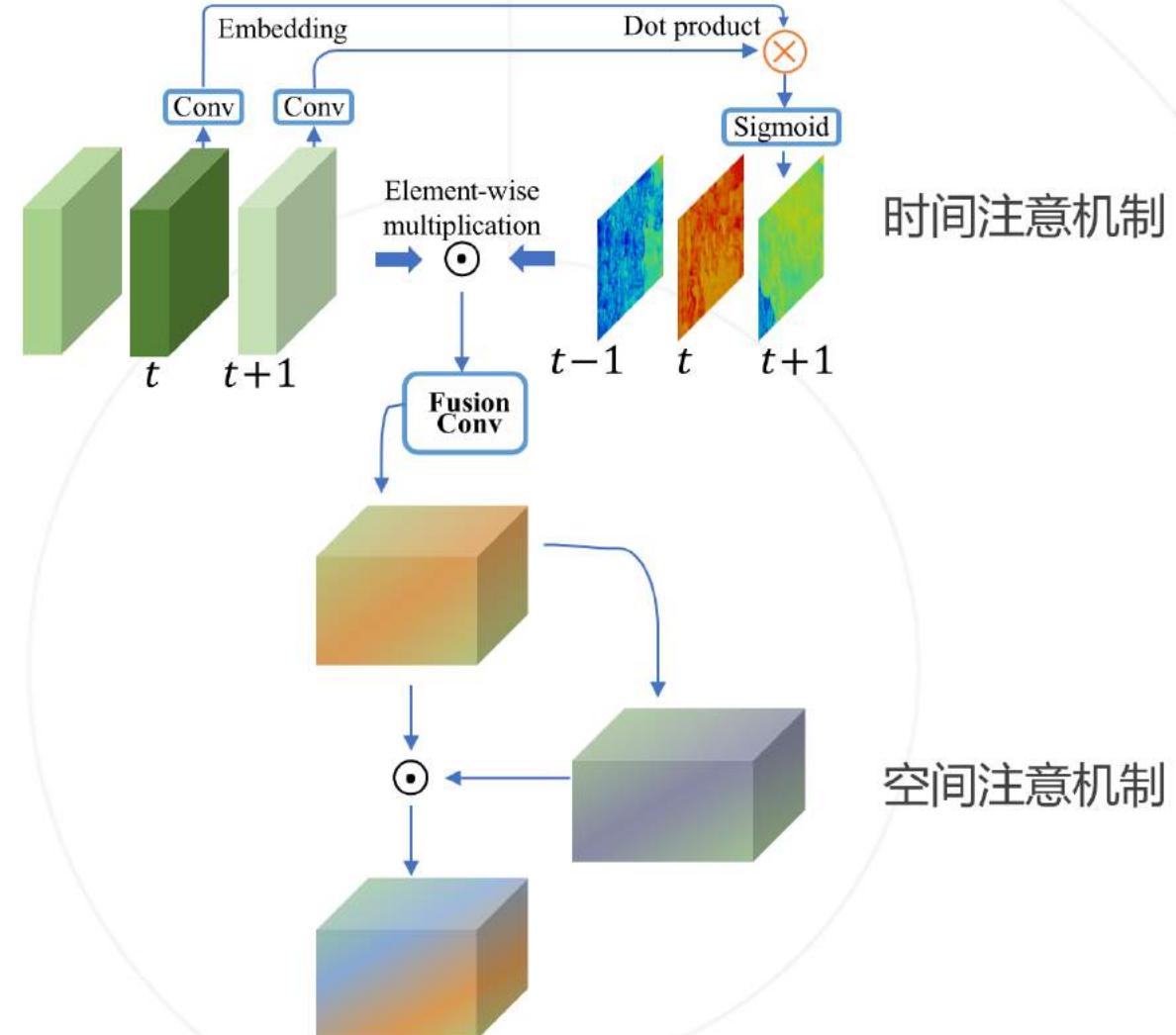
- 为了处理快速而复杂的运动
 - ✓ pyramidal processing
 - ✓ cascading refinement



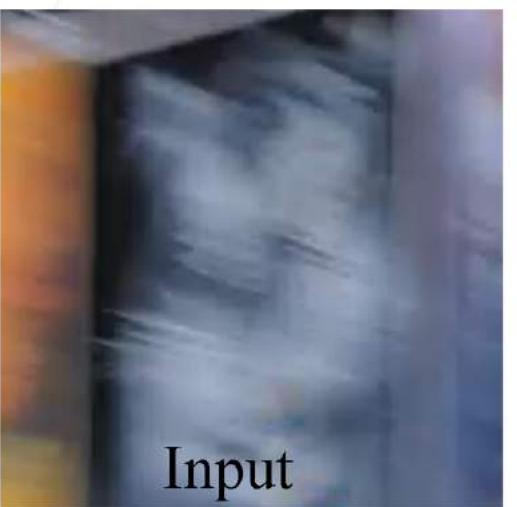
- 适用于不同视频复原任务的通用框架
- PCD**: 通过金字塔级联变形对齐处理大的运动，使用形变卷积以由粗到细的方式在特征级别进行帧对齐
- TSA**: 时空注意力机制

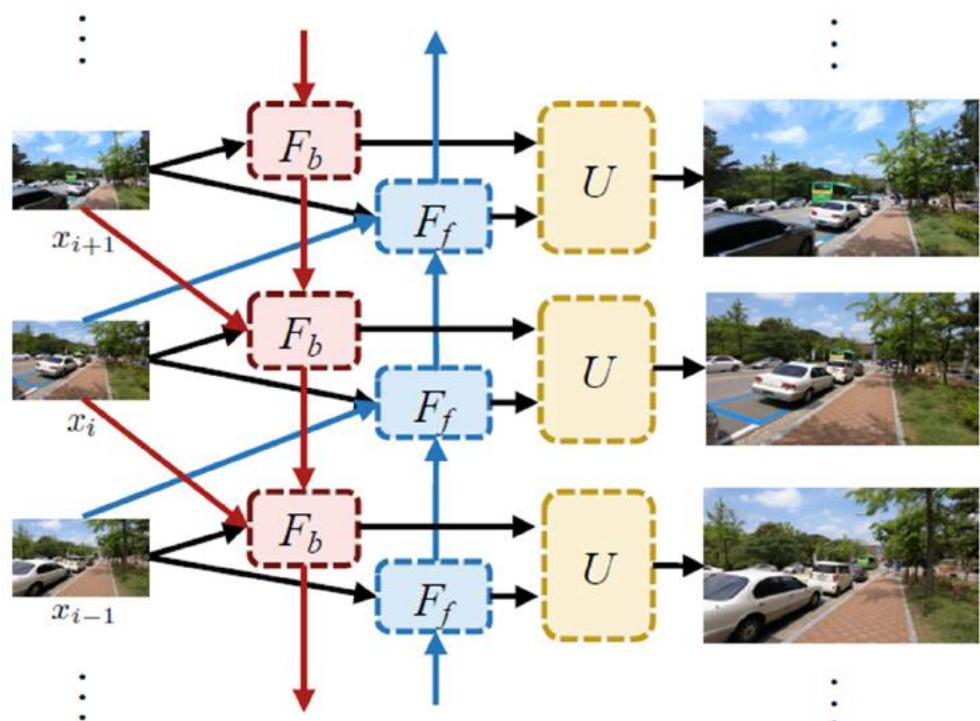


TSA: Fusion with Temporal and Spatial Attention

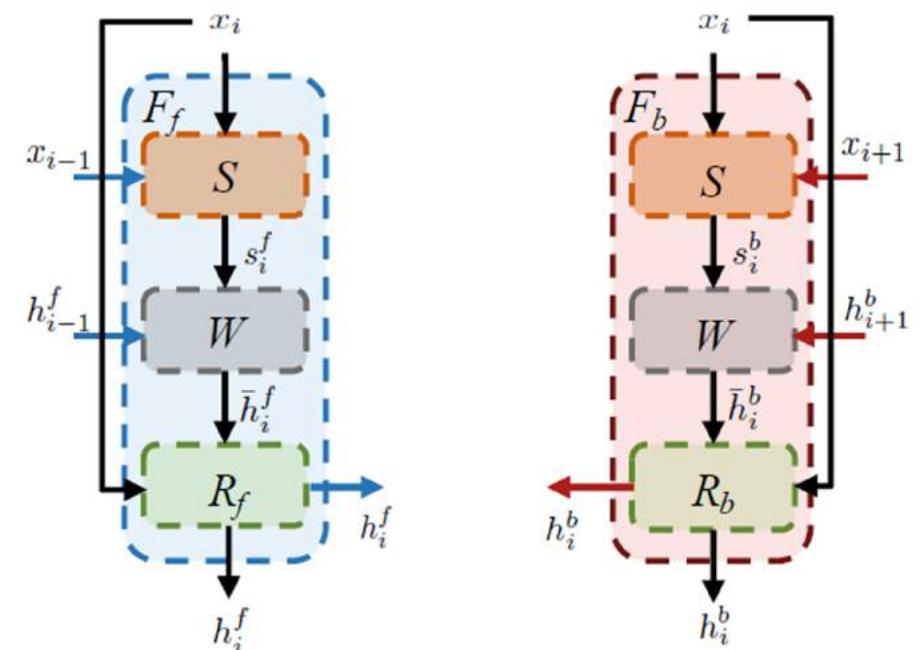


- 由于遮挡, 模糊和未对齐等等问题, 相邻帧的信息不足, 不同的相邻帧应该有不同的权重
- 我们通过以下方式在每帧上分配像素级聚合权重:
 - ✓ temporal attention 时间注意机制
 - ✓ spatial attention 空间注意机制





(a) BasicVSR architecture



(b) Forward and backward propagation branches

BasicVSR在结构上更简单，效果比EDVR更好

谢谢大家