

# 不同模型在超分辨率领域效果的复现和比较

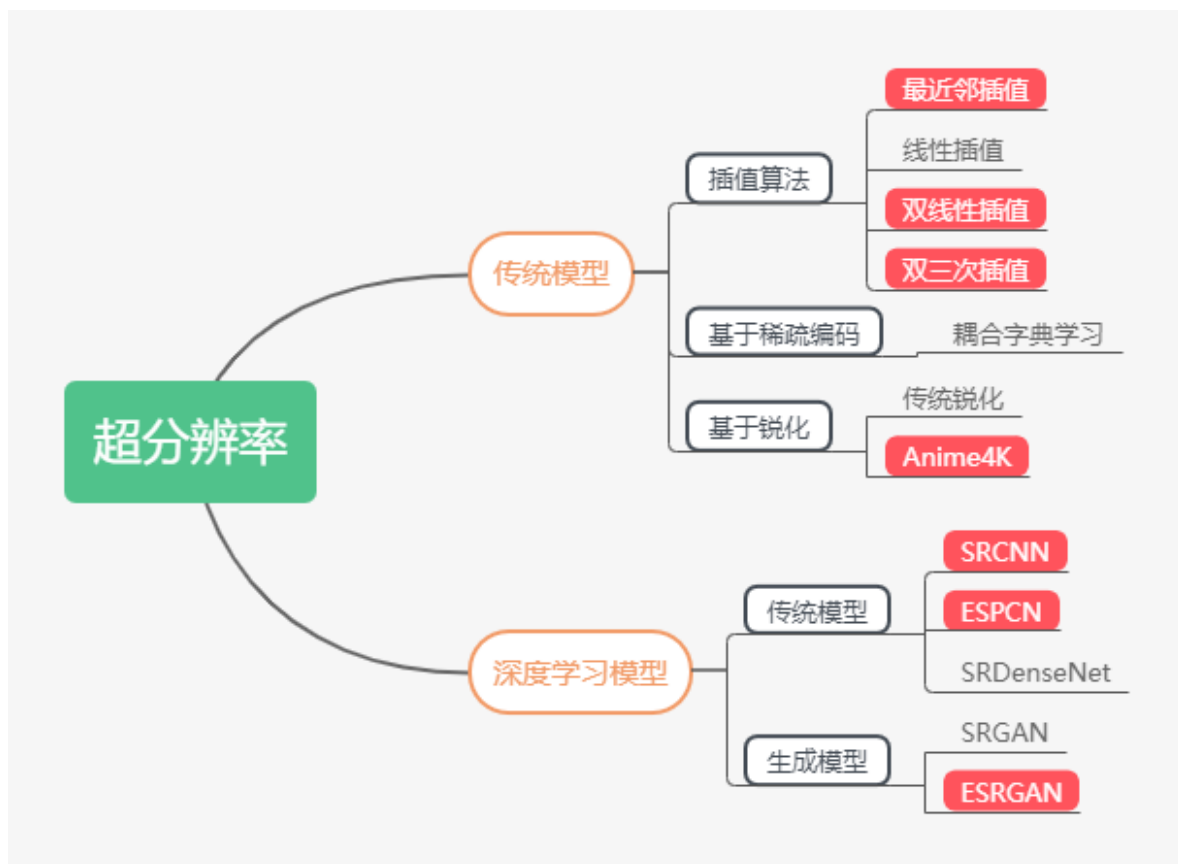
## 一. 课题背景

在用小数据集训练 GAN 的数据增强方案 (NeurIPS-2020-training-generative-adversarial-networks-with-limited-data-Paper [1]) 的来进行基于单视频的数据集过程生成对应人物头像的实验中, 上述增强方法在传统GAN及其改进上 (WGAN, LSGAN等) 有着较好的效果, 但结果图片分辨率从 128 到 256 扩展时, 网络结构不再稳定, 加上视频截取的数据集, 大部分大小分布在 128 左右的区间内, 所以说无法进行更高分辨率的训练, 所以选择尝试使用超分辨率算法进行最终模型输出的提升。



64 x 64 和 128 x 128 模型隐空间插值结果

超分辨率是通过硬件或软件的方法提高原有图像的分辨率, 通过一系列低分辨率的图像来得到一幅高分辨率的图像过程就是超分辨率重建, 总的来说图像超分辨率大体分为两种方式: 一是利用计算机视觉等传统算法进行图像信息的增强, 二是利用深度学习模型对未知的图像信息进行预测, 而本实验将采用不同领域的超分辨率算法进行复现和效果的比较。



## 二. 拟解决问题

1. 比较不同超分辨率算法的优劣，并分析其合理性
2. 探究超分辨率算法能否在本实验背景下对 GAN 的生成结果有很好的提升
3. 除了超分辨率以外有没有更好的方法解决该问题

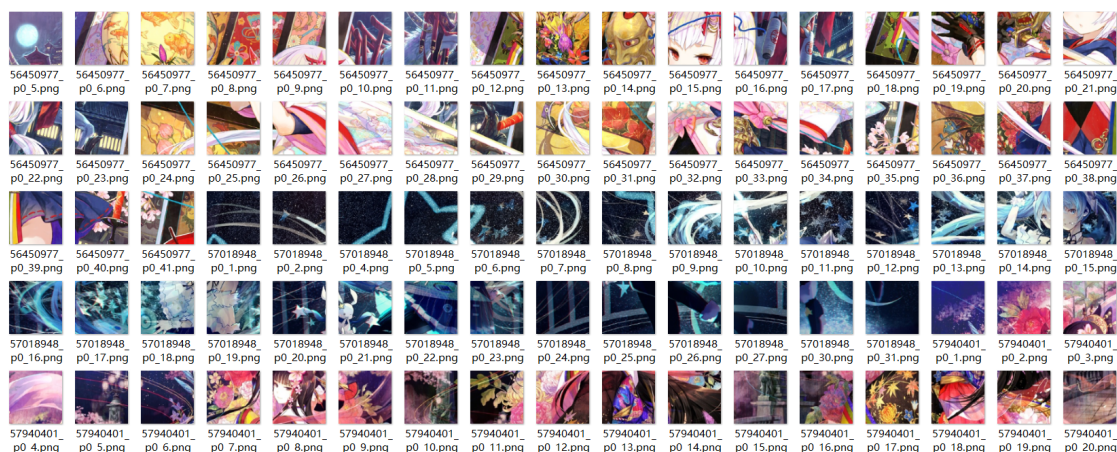
### 三. 项目步骤

1. 阅读 Anime4K, SRCNN, ESPCN, SRGAN, ESRGAN 的论文，并提取核心思想
2. 采用合适的实验数据集，并经过处理将其分为 train 集，val 集和 test 集
3. 复现这几个模型的效果，并用 PSNR, SSIM 等评价指标衡量生成图像的质量
4. 对这几个模型进行结果比对和分析

### 四. 实验实现

#### 0. 数据集处理

数据集来源：为了保证任务能更好的契合GAN数据集的风格，选取了2016年-2018年 pixiv 月榜图片，切割成大小为 128 x 128 的小块左右作为数据集，经过图片指纹去重，低密度信息切块的去除，得到约1.5w 张不同细节的图片块



图片指纹去重：将图片按照哈希指纹变换为13x13的指纹特征，在特征空间内判断重复度，如果重复度过高则剔除

低密度信息切块去除：将每个图片映射成像素直方图，用聚类检测颜色聚集程度，如果聚集程度过高，说明当前图片可能是纯色的图，或者细节很少的图，将其去除

#### 1. 建立评价指标模型

使用 **PSNR** 指标和 **SSIM** 指标来评测超分辨率的生成结果

**PSNR** 指标：峰值信噪比，是一种全参考的图像质量评价指标，数值越大表示失真越小

$$MSE = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W (X(i, j) - Y(i, j))^2$$
$$PSNR = 10 \log_{10} \left( \frac{(2^n - 1)^2}{MSE} \right)$$

**SSIM** 指标：结构相似性，也是一种全参考的图像质量评价指标，它分别从亮度、对比度、结构三方面度量图像相似性，值越大，表示图像失真越小

$$\mu_X = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X(i, j) \quad \sigma_X^2 = \frac{1}{H \times W - 1} \sum_{i=1}^H \sum_{j=1}^W (X(i, j) - \mu_X)^2$$

$$\sigma_{XY} = \frac{1}{H \times W - 1} \sum_{i=1}^H \sum_{j=1}^W ((X(i, j) - \mu_X)(Y(i, j) - \mu_Y))$$

$$l(X, Y) = \frac{2\mu_X\mu_Y + C_1}{\mu_X^2 + \mu_Y^2 + C_1} \quad c(X, Y) = \frac{2\sigma_X\sigma_Y + C_2}{\sigma_X^2 + \sigma_Y^2 + C_2} \quad s(X, Y) = \frac{\sigma_{XY} + C_3}{\sigma_X\sigma_Y + C_3}$$

$$SSIM(X, Y) = l(X, Y) \cdot c(X, Y) \cdot s(X, Y)$$

## 2. 算法模型

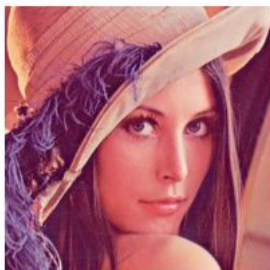
### Anime4K:

Anime4K 是针对快速放大视频而设计的，但在单张图片上也有还不错的效果，原理是尝试最小化残差（r）的线宽，将 LR\_U 及其残差作为输入，挤压残差像素，使残差线变细，以此产生锐化细节的效果

## Image Decomposition

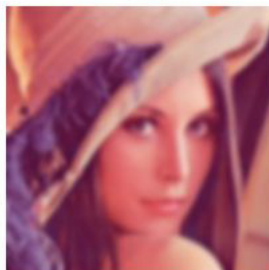
High Resolution  
Image

HR



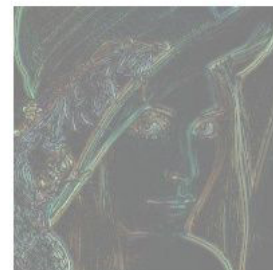
Upscaled Low Resolution  
Image

LR\_U



High-Frequency  
Residual

r



= +

Equivalently, the following is true :

$$r = (HR - LR\_U)$$

### SRCNN:

SRCNN 是非常经典的将深度学习应用到神经网络的设计，也是将深度学习引入到超分辨率领域的开山之作

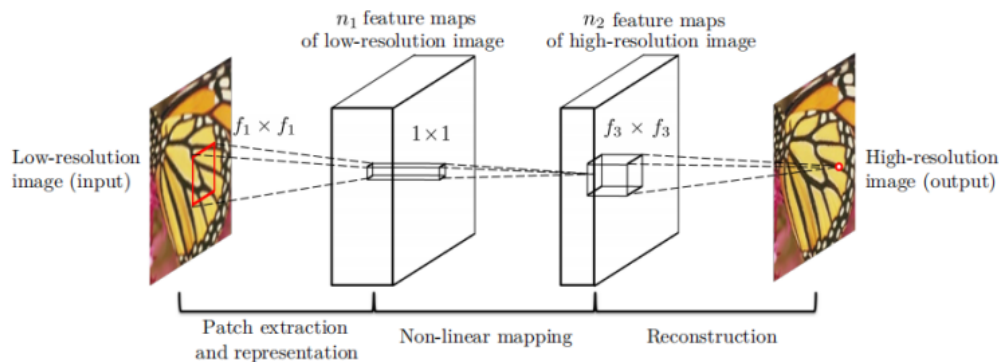
SRCNN 的结构非常简单，整个卷积网络仅包括三个卷积层

第一层卷积对低分辨率的图进行卷积操作 feature maps A

第二层卷积对 feature maps A 再进行卷积操作生成细化的 feature maps B

第三层卷积对 feature maps B 进行卷积生成超分辨率的图像





**Fig. 2.** Given a low-resolution image  $\mathbf{Y}$ , the first convolutional layer of the SRCNN extracts a set of feature maps. The second layer maps these feature maps nonlinearly to high-resolution patch representations. The last layer combines the predictions within a spatial neighbourhood to produce the final high-resolution image  $F(\mathbf{Y})$ .

**ESPCN:**

ESPCN 网络利用亚像素的思想, 将 SRCNN 中的常规上采样进行了改进

ESPCN 模型只在模型的末端进行上采样，可以使得在低分辨率空间保留更多的纹理区域，在视频超分中也可以做到实时，而且直接使用亚像素卷积的方式进行上采样，相比于将 LR 插值到 HR，这种上采样方式可以学习到更好、更为复杂的方式，可以获得更好的重建效果，这种重排像素的上采样方法后来被称作 PixelShuffle，也是常用的采样方法

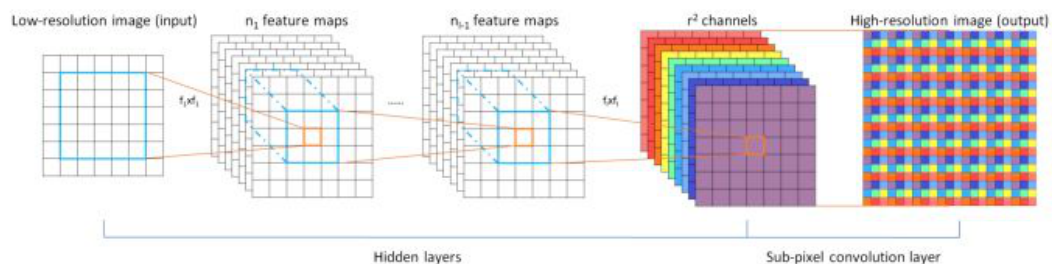
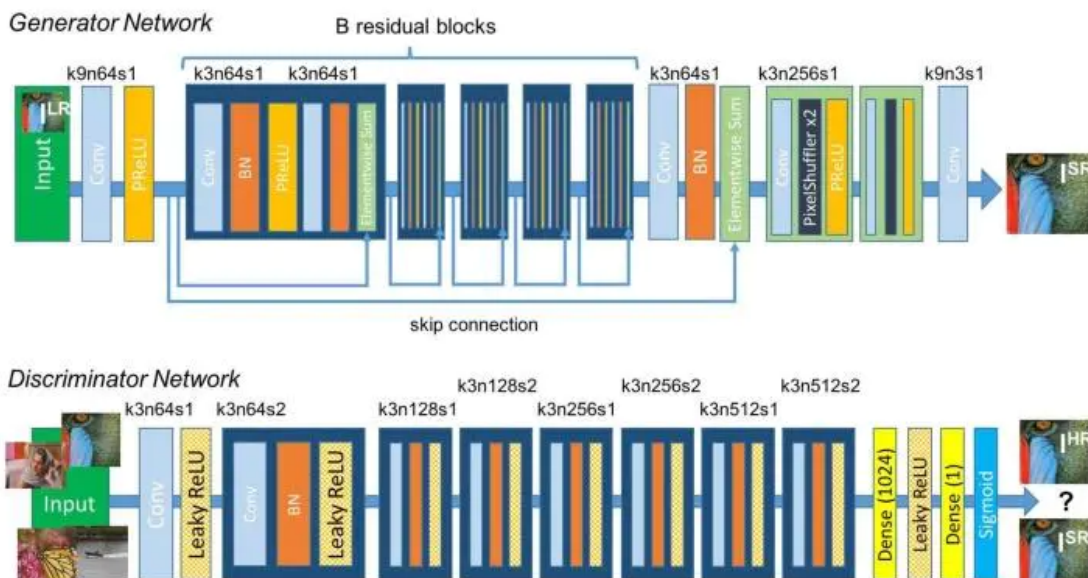


Figure 1. The proposed efficient sub-pixel convolutional neural network (ESPCN), with two convolution layers for feature maps extraction, and a sub-pixel convolution layer that aggregates the feature maps from LR space and builds the SR image in a single step.

**SRGAN:**

SRGAN 和 SRCNN 最大的区别就是引入了生成对抗网络，与 SRCNN 的均方误差不同，SRGAN 在考虑生成图像和原图的同时，也在生成图片的特征图与真实图片的特征图差距，并用一个内容误差来表示，利用 GAN 双网络的性质，对这两个方面进行优化



## ESRGAN:

ESRGAN 对前者 SRGAN 进行了三方面优化

1. 改进了网络结构,对抗损失,感知损失
2. 引入Residual-in-Residu Dense Block
3. 使用激活前的VGG特征来改善感知损失

在 SRGAN 中，用了传统的残差网络，使用了 BN 层，但实验发现这样会导致残影，所以 ESRGAN 中改为了将残差块用密集连接的方式连在一起方法，去掉了 BN 层，同时在网络的总损失中加入了 L1 误差，来避免出现过于平滑的现象，同时还引入了网络插值的概念，在论文中作者为了平衡感知质量和 PSNR 等评价值，首先基于 PSNR 方法训练的得到的网络 G\_PSNR，然后再进行整合

3. 不同算法比较

Anime4K:

指标名称	指标数值
PSNR	18.21
SSIM	0.639

SRCNN:

指标名称	指标数值
PSNR	25.73
SSIM	0.749

ESPCN:

指标名称	指标数值
PSNR	26.44
SSIM	0.735

ESRGAN:

指标名称	指标数值
PSNR	30.51
SSIM	0.7611

---

## 五. 实验结果分析

可以明显看到，SRCNN 和 ESPCN 由于神经网络层数过小，无法达到很好的预测效果，本质上不会比传统的插值方法好太多，也不能像生成模型那样生成不同细节，而 SRGAN 和 ESRGAN 的生成模型则能更好的实现超分辨率任务，无论是在细节丰富程度上还是图像锐化度上都要优于传统的算法和 SRCNN 和 ESPCN，但缺点是生成模型由于网络较大，也没有加入优化，他的单张图片的处理速度要慢很多，很难应用到实时的视频放大，而针对速度优化的 Anime4K 算法则将速度加快到了极致，能保证在还一般的效果下，达到深度学习模型的 10 倍~20 倍的处理速度。

处理速度	处理效果
传统插值	SRGAN 和 ESRGAN
Anime4K	SRCNN 和 ESPCN
SRCNN 和 ESPCN	Anime4K
SRGAN 和 ESRGAN	传统插值

这几个模型的效果都没有想象中的好，因为本质上超分辨率是一个低维到高维数据的映射，映射的方式很多，对应细节也不同，所以说如果数据集很小而且没有针对性，虽然也能训练出比较好的效果，但是这样的话网络只能训练出对常规细节的填补，并不能训练出对针对性细节的填补效果。

所以其实超分辨率在本实验中并不是合适的方案，以 ESRGAN 为例，用作超分辨率的模型本是一个生成模型，那其实他是可以跟原本的GAN整合在一起，合并成一个组合网络共同训练，得到的效果会更好，实际上现在针对高像素图片生成的生成模型就是采取这种方法，它会以一种渐进式的方法组合堆叠网络，在保证稳定的前提下达到最好的效果。