

# SeD

状态：CVPR 2024

单位：中国科学技术大学

文章链接：[arxiv.org/pdf/2402.19387](https://arxiv.org/pdf/2402.19387)

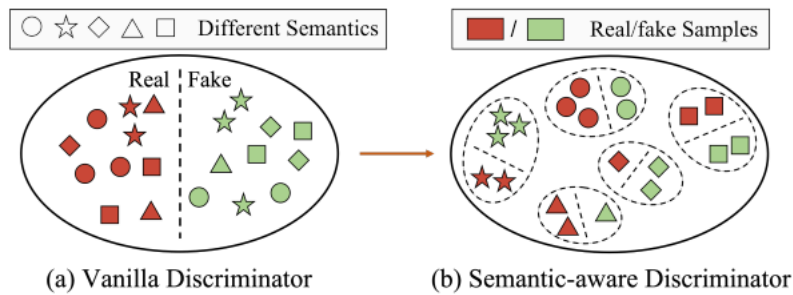
Github 链接：<https://github.com/lbc12345/SeD>

## 目录

|          |   |
|----------|---|
| 摘要 ..... | 1 |
| 背景 ..... | 1 |
| 贡献 ..... | 2 |
| 实现 ..... | 2 |
| 实验 ..... | 3 |

## 摘要

GAN 模型在 SR 领域有着较广泛的应用，但其效果仍旧较为粗糙。本文提出了一种根据语义进行划分的训练方法，即如下图所示，原版的 GAN 是在真实数据与生成数据的整体上进行比较，而本文提出在判别器运作时额外引入语义信息，以让判别器在更小的范围内做出更为精确的比较，从而取得更好的效果，同时提高生成器的能力。



## 背景

深度学习在 SISR 任务上得到了广泛的应用，早期的 CNN，Transformer 都达到了较好的效果。但其在像素层级上设置的损失函数使得模型难以输出具有足够的想象力，即输出图像的质感较差。

GAN 模型则通过判别器 **D** 和生成器 **G** 的对抗训练取得了较好的效果。最初使用的是图像层级的判别器，但同样会因为图像分布范围大而得到模糊的输出。因此，块/像素层级的判别器也得到了利用，但其会造成更大的计算量。

然而，这些工作都没有考虑到不同语义范围的图像有着不同的特征，其分布自然也不可一概而论。要解决这一问题，自然的想法就是向生成器 **G** 中加入额外的语义信息，但问题在于：1.低分辨率的图像可能会造成语义信息的错误。2.额外的语义信息提取会造成预测时的开销。因此，本工作决定将语义信息放在判别器 **D** 中，从而间接提高生成器 **G** 的效果。

相关工作包括三部分：

- 1.SISR 任务
- 2.基于 GAN 的 SR 方案
- 3.提取语义信息的 PVM 方案

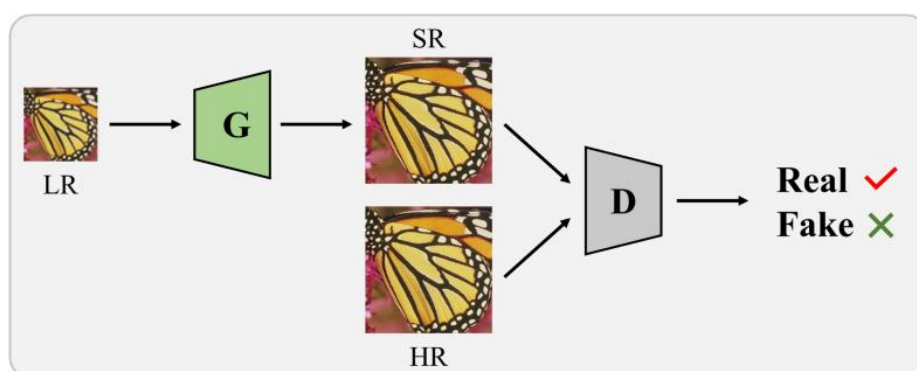
## 贡献

1.本文指出了语义信息在高质量超分任务中的重要性，并构建了 SeD 架构来进行超分。

2.为了更好的向 **D** 中注入语义信息，本文推出了 SeFB 模块，使用 cross-attention 方式加入控制。

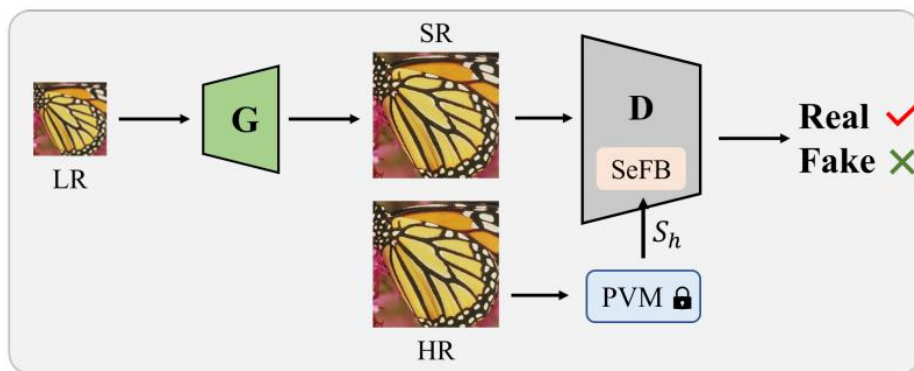
3.通过实验论证了这一思路的可行性。

## 实现



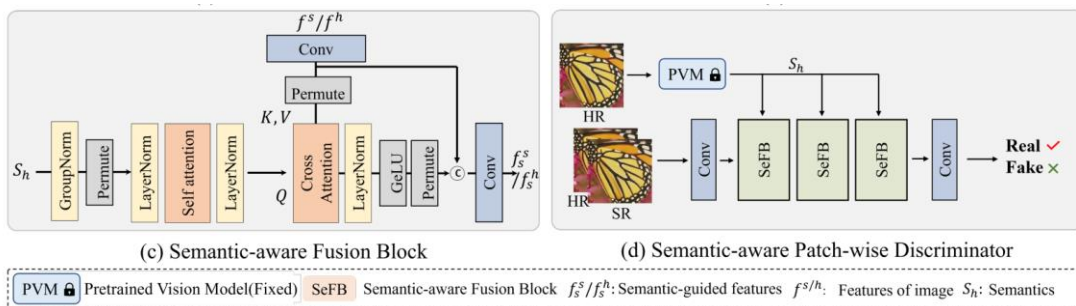
(a) Vanilla GAN

最为基础的 GAN 结构如图，**D** 用于对比 **HR/SR**，并判断二者哪一个才是真实图像。但由于 **D** 是被所有数据共同训练的，所以其需要分辨的是所有数据构成的大分布，由此使得 **D** 的判别不够准确，从而进一步影响到 **G** 的生成不够清晰。



(b) SeD GAN

而本文的方案则是给  $D$  额外加入了一个  $S_h$  (语义信息) 作为条件, 可以帮助  $D$  进一步缩小分布的范围, 从而取得效果更佳的分辨能力。虽然图中只有  $SR$  进入  $D$  中进行判断, 但由后续具体结构图可知, 依旧是  $SR/HR$  共同输入  $D$  中。

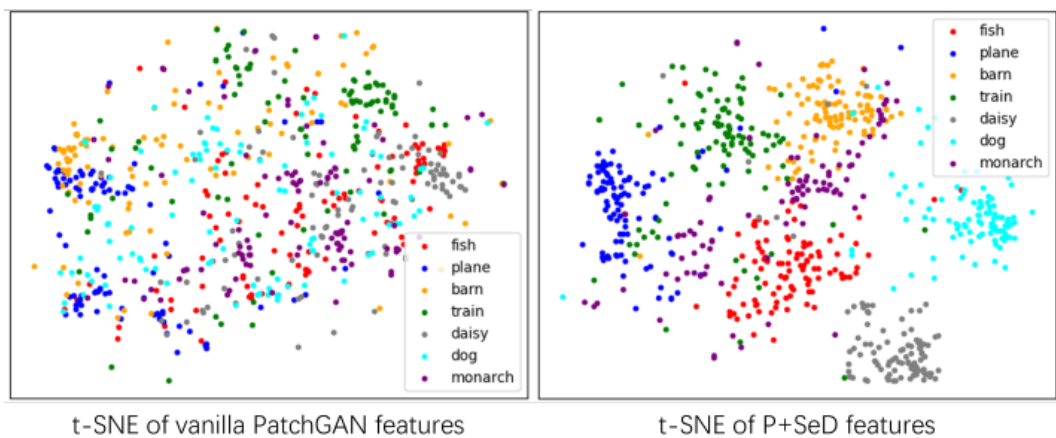


具体结构图如上, 大致就是使用 corss attention 引入的额外控制条件。

## 实验

| Method          | Set5                      | Set14                     | DIV2K                     | Urban100                  | Manga109                  |
|-----------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| SFTGAN          | 0.080/30.06/0.848         | -/-                       | 0.133/28.08/0.771         | 0.134/24.34/0.723         | 0.072/28.17/0.856         |
| ESRGAN          | 0.076/30.44/0.852         | 0.133/26.28/0.699         | 0.115/28.20/0.777         | 0.123/24.37/0.734         | 0.065/28.41/0.859         |
| USRGAN          | 0.079/30.91/0.866         | 0.143/27.15/0.736         | 0.132/28.77/0.793         | 0.133/24.89/0.750         | 0.063/28.75/0.872         |
| RRDB+LDL        | 0.069/31.03/0.861         | 0.121/26.94/0.721         | 0.101/28.95/0.795         | 0.110/25.50/0.767         | 0.055/29.41/0.875         |
| RRDB+DualFormer | 0.068/31.40/0.872         | 0.121/27.53/0.741         | 0.103/29.30/0.802         | 0.115/25.73/0.774         | 0.053/29.90/0.886         |
| RRDB+P          | 0.070/30.67/0.860         | 0.130/26.92/0.724         | 0.111/28.71/0.792         | 0.120/24.86/0.752         | 0.058/28.60/0.872         |
| RRDB+P+SeD      | <b>0.064</b> /31.22/0.867 | <b>0.117</b> /27.37/0.736 | <b>0.094</b> /29.27/0.802 | <b>0.106</b> /25.93/0.779 | <b>0.048</b> /29.99/0.888 |
| RRDB+U          | 0.072/31.13/0.869         | 0.127/27.52/0.739         | 0.110/29.28/0.802         | 0.125/25.61/0.768         | 0.056/29.49/0.882         |
| RRDB+U+SeD      | <b>0.069</b> /31.73/0.880 | <b>0.123</b> /27.94/0.757 | <b>0.102</b> /29.85/0.818 | <b>0.112</b> /26.20/0.788 | <b>0.047</b> /30.46/0.897 |
| SwinIR+LDL      | 0.065/31.03/0.861         | 0.118/27.22/0.732         | 0.094/29.12/0.801         | 0.102/26.23/0.792         | 0.047/30.14/0.888         |
| SwinIR+P        | 0.070/31.49/0.876         | 0.127/27.66/0.747         | 0.103/29.66/0.815         | 0.107/26.22/0.790         | 0.048/30.18/0.895         |
| SwinIR+P+SeD    | <b>0.061</b> /31.44/0.870 | <b>0.115</b> /27.53/0.742 | <b>0.090</b> /29.53/0.810 | <b>0.097</b> /26.45/0.794 | <b>0.044</b> /30.48/0.896 |
| SwinIR+U        | <b>0.064</b> /31.38/0.869 | 0.120/27.64/0.744         | <b>0.095</b> /29.56/0.810 | 0.103/26.09/0.786         | 0.049/29.99/0.889         |
| SwinIR+U+SeD    | 0.067/31.64/0.874         | <b>0.117</b> /27.84/0.750 | 0.096/29.79/0.816         | <b>0.102</b> /26.46/0.796 | <b>0.045</b> /30.58/0.898 |

由于本文的方案只是对  $D$  加入了一个额外的条件, 所以可以兼容大部分 GAN based 方案。上图可得出, 加入 SeD 训练方案后, 数据指标 LPIPS/PSNR/SSIM 大都得到了一定程度的提高。



此外，具体绘制出多种语义的图像分布情况，可以看出，在加入 **SeD** 后，同种语义信息的图像分布更为接近，这也十分符合现实的实际情况。