# 图像超分辨率重建技术介绍

## 问题简介

图像分辨率是一组用于评估图像中蕴含细节信息丰富程度的性能参数，包括时间分辨率、空间分辨率及色阶分辨率等，体现了成像系统实际所能反映物体细节信息的能力。相较于低分辨率图像，高分辨率图像通常包含更大的像素密度、更丰富的纹理细节及更高的可信赖度。但在实际上情况中，受采集设备与环境、网络传输介质与带宽、图像退化模型本身等诸多因素的约束，我们通常并不能直接得到具有边缘锐化、无成块模糊的理想高分辨率图像。提升图像分辨率的最直接的做法是对采集系统中的光学硬件进行改进，但是由于制造工艺难以大幅改进并且制造成本十分高昂，因此物理上解决图像低分辨率问题往往代价太大。由此，从软件和算法的角度着手，实现图像超分辨率重建的技术成为了图像处理和计算机视觉等多个领域的热点研究课题。

## 1.2 应用领域

1955年，Toraldo di Francia在光学成像领域首次明确定义了超分辨率这一概念，主要是指利用光学相关的知识，恢复出衍射极限以外的数据信息的过程。1964年左右，Harris和Goodman则首次提出了图像超分辨率这一概念，主要是指利用外推频谱的方法合成出细节信息更丰富的单帧图像的过程。1984 年，在前人的基础上，Tsai和 Huang 等首次提出使用多帧低分辨率图像重建出高分辨率图像的方法后, 超分辨率重建技术开始受到了学术界和工业界广泛的关注和研究。

图像超分辨率重建技术在多个领域都有着广泛的应用范围和研究意义。主要包括：

（1） 图像压缩领域

在视频会议等实时性要求较高的场合，可以在传输前预先对图片进行压缩，等待传输完毕，再由接收端解码后通过超分辨率重建技术复原出原始图像序列，极大减少存储所需的空间及传输所需的带宽。

（2） 医学成像领域

对医学图像进行超分辨率重建，可以在不增加高分辨率成像技术成本的基础上，降低对成像环境的要求，通过复原出的清晰医学影像，实现对病变细胞的精准探测，有助于医生对患者病情做出更好的诊断。

（3） 遥感成像领域

高分辨率遥感卫星的研制具有耗时长、价格高、流程复杂等特点，由此研究者将图像超分辨率重建技术引入了该领域，试图解决高分辨率的遥感成像难以获取这一挑战，使得能够在不改变探测系统本身的前提下提高观测图像的分辨率。

（4） 公共安防领域

公共场合的监控设备采集到的视频往往受到天气、距离等因素的影响，存在图像模糊、分辨率低等问题。通过对采集到的视频进行超分辨率重建，可以为办案人员恢复出车牌号码、清晰人脸等重要信息，为案件侦破提供必要线索。

（5） 视频感知领域

通过图像超分辨率重建技术，可以起到增强视频画质、改善视频的质量，提升用户的视觉体验的作用。

## 1.3 现有深度学习方法及其不足

在深度学习技术被引入图像超分辨率领域之前，研究者们主要依赖于插值方法（如最近邻和双三次插值）和基于特征提取的方法（如稀疏编码、邻居嵌入和贝叶斯森林）。这些传统方法无法有效地恢复图像的高频细节，导致输出图像质量不佳。

自21世纪中叶以来，深度学习技术在图像分类等视觉任务中取得了显著进展，尤其是在ImageNet挑战赛之后。这些技术随后被应用于超分辨率任务，显著提高了图像质量。例如，Dong等人提出的SRCNN模型通过卷积层增强了上采样图像。

为了训练深度学习模型，研究者定义了多种损失函数，主要是基于像素级的误差度量，如均方误差（MSE）和平均绝对误差（MAE）。这些损失函数主要关注图像的定量质量，通过PSNR、RMSE和SSIM等指标进行评估。

尽管基于像素差异最小化的现有方法在定量评估方面表现出色，但它们往往无法保证输出图像的自然性和感知质量。例如，树木和房屋的细微细节可能无法被充分恢复（如下图1.c）。



图1 PIRM数据集的图像获得的示例结果

为了改善感知质量，研究者提出了使用生成对抗网络（GANs）和图像分类网络的中间特征的方法。这些方法通过训练模型来最小化VGG19等分类网络中间层的特征差异。尽管这些方法显著提高了感知性能，但它们仍有局限性：

间接考虑感知质量：感知质量的提高可能受限，因为用于提取中间特征的网络模型原本是为图像分类任务设计的，而不是直接针对感知质量优化。

鉴别器的局限性：鉴别器网络可能学习到的准则可能与感知质量无关，例如，它可能仅依赖于高频分量来区分图像，导致超分辨率模型在低频区域添加不期望的纹理。

定量质量的牺牲：为了提高感知质量，基于GAN的方法可能会牺牲图像的定量质量，例如，SRGAN模型虽然在感知性能上优于其他模型，但在定量质量指标（如RMSE）上表现更差。

# 实现细节

本次练习运用了Four-pass perceptual super-resolution with enhanced upscaling (4PP-EUSR)方法实现了图像超分辨率重建，这个模型建立在先前提出的增强型上采样超分辨率（EUSR）模型基础之上，旨在提升放大图像的感知质量，同时维持图像的传统定量性能。

## 2.1 创新

4PP-EUSR模型在训练阶段采用了多通道上采样策略，通过模型的多个上采样路径生成多个上采样图像。这使得模型能够学习到不同的上采样特征，并考虑到上采样过程中可能出现的各种情况。该模型使用了两个定性分数预测器来直接评估上采样图像的美学和主观质量分数。这些预测器基于人类评价者的数据进行训练，有助于提升图像的感知质量。同时，4PP-EUSR引入了鉴别器网络，用于区分生成的图像与真实图像。这促使超分辨率模型生成更加自然的图像，类似于GANs的使用，但专注于超分辨率任务。以往的方法可能在提高感知质量的同时牺牲了定量性能，而4PP-EUSR模型通过其设计实现了在不降低定量质量的前提下，提升图像的感知质量。

在结构设计方面，4PP-EUSR模型基于EUSR模型，该模型本身就是一种先进的多尺度超分辨率方法。4PP-EUSR进一步通过多通道上采样和定性分数预测器增强了其性能。

总的来说，4PP-EUSR模型的创新之处在于其综合考虑了定量和感知质量，并通过多通道上采样和定性评分预测器的使用，提高了超分辨率图像的自然性和视觉吸引力，同时保持了图像的细节和结构。这些创新使得4PP-EUSR在图像超分辨率领域中相比以往的深度学习方法有了显著的提升。

## 2.2 实现流程

所提出方法的架构可以分解为四个组成部分（见图2）：一个多尺度上采样模型、以多通道方式使用模型、一个鉴别器和定性分数预测器。

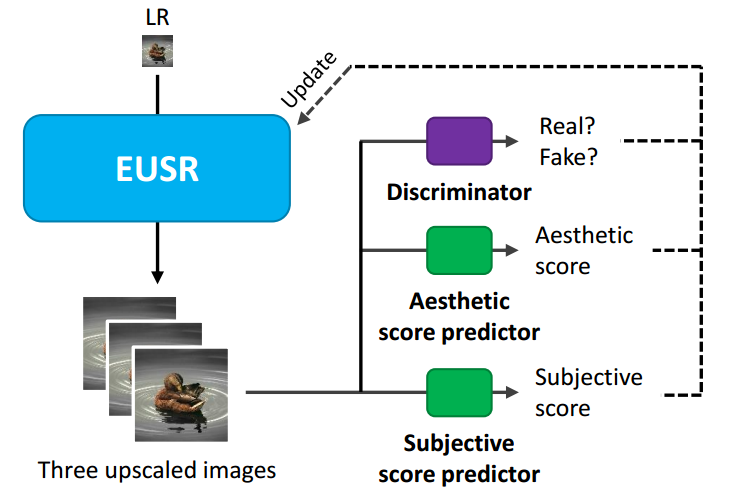


图2 4PP-EUSR方法概述

·模型的第一部分是一个多尺度上采样模型，它负责生成三个上采样图像；在训练阶段，模型使用多个上采样路径来产生三个输出图像。第一个图像直接来自×4上采样路径，第二个图像是通过×2路径两次上采样得到的，第三个图像是通过×8路径上采样然后通过双三次插值下采样得到的，描述细节在2.3多通道上采样中；

·鉴别器网络在训练阶段使用，目的是区分生成的上采样图像和真实的高分辨率图像。鉴别器网络试图尽可能准确地识别图像，而超分辨率模型则被训练为使鉴别器难以区分它们，使用鉴别器网络和定性分数预测器的输出来更新超分辨率模型。这意味着模型在训练过程中不仅考虑定量质量指标，还考虑感知质量指标，描述细节在2.4鉴别器网络中；

·模型还包括两个定性分数预测器，它们分别评估上采样图像的美学分数和主观质量分数。这些预测器基于人类评价者的数据进行训练，描述细节在2.5定性评分预测因子中；

在训练过程中，使用了多种损失函数，包括重建损失、对抗性损失、美学分数损失和主观分数损失，描述细节在2.6训练细节中。

## 2.3 多通道上采样

原始的EUSR模型支持2倍、4倍和8倍的多尺度超分辨率。在训练阶段，模型利用所有这些放大路径生成三个输出图像，其中我们使得给定图像的输出图像具有相同的4倍放大因子，具体如下（图3）。第一个图像直接从4倍路径生成。第二个图像通过将给定图像通过2倍路径两次传递生成。第三个图像通过将从8倍路径得到的图像进行2倍双三次降采样生成。

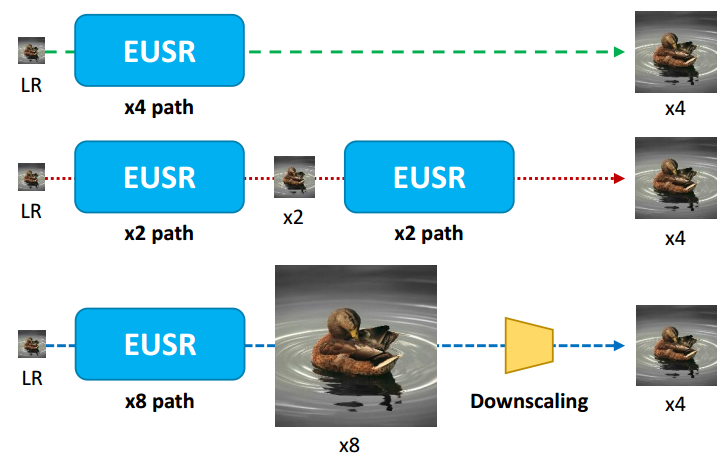


图3 多通道升级过程，从共享的预训练EUSR模型中产生3个升级图像，放大系数为4

因此，对于每个输入图像，该模型被使用了四次。多尺度模型的原始目的是支持单个模型上的可变缩放因子。另一方面，模型多次通过放大扩展了它，并具有不同的目标，即改善固定缩放因子下的放大图像质量。由于训练时使用了从不同放大路径获得的三个图像，模型必须学习减少可能在通过4倍路径直接放大、通过2倍路径进行两次传递放大以及通过8倍路径放大和降采样时出现的伪影。这样可以防止模型过度拟合特定模式，从而使模型能够处理各种放大场景。

## 2.4鉴别器网络

在训练阶段，4PP-EUSR模型利用了一个鉴别器网络，该网络旨在区分生成的图像和真实图像。虽然鉴别器试图尽力识别放大的图像，但超分辨率模型的训练目标是让鉴别器难以区分它们与真实图像的差异。这有助于放大模型生成更自然的图像。受到SRGAN 的启发，4PP-EUSR模型的鉴别器网络由多个卷积层组成，后面跟着带有α = 0.2的LeakyReLU激活函数和两个全连接层，如图4所示。

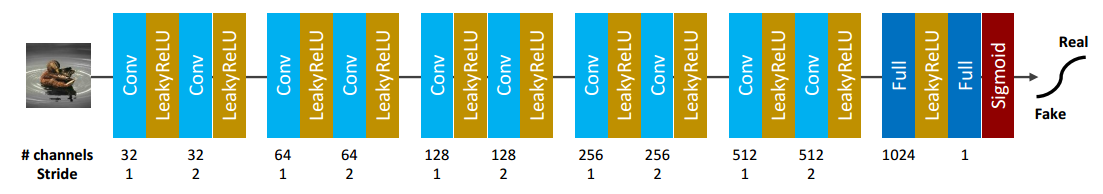


图4 鉴别器网络的结构

最后的sigmoid激活函数确定输入图像是真实图像还是伪造图像的概率。由于输入图像块的大小不同，它比原始的SRGAN模型多了两个卷积层。

## 2.5定性评分预测因子

在提升感知上的超分辨率中的一个主要思想是利用深度学习模型对图像的感知质量进行分类，而不是一般的图像分类器。为此，使用两个深度网络分别预测图像的审美和主观质量分数。该方法将著名的图像分类器的最后一层替换为具有softmax激活函数的全连接层，该层生成10个分数类别的概率，使用MobileNetV2 作为基础图像分类器，因为它比其他图像分类器更快，并支持各种尺寸的输入图像。

## 2.6训练细节

超分辨率模型的训练过程，包括三个阶段：预训练多尺度超分辨率模型、训练定性分数预测器和以感知方式训练超分辨率模型。

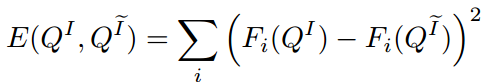
### 2.6.1 预训练多尺度超分辨率模型

使用EUSR模型，包含32个和1个局部残差块（local residual blocks）在残差模块（residual module）和上采样部分（upscaling part）。预训练使用DIV2K数据集的训练集（800张图片），采用L1重建损失函数。训练步骤包括随机裁剪训练图像以获取16个48×48像素的图像块，然后随机选择一个上采样路径（×2、×4或×8）进行训练。使用Adam优化方法更新参数，初始学习率为10^-4，每200,000步减半。

### 2.6.2 训练定性分数预测器

使用MobileNetV2作为基础图像分类器，替换最后一层为全连接层，以预测图像的质量分数。

作为训练定性分数预测器的损失函数，采用了平方地球移动距离：

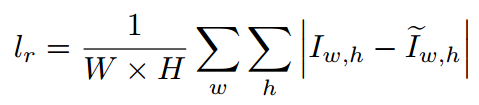


### 2.6.3 训练超分辨率模型

微调预训练的EUSR模型，并与判别器网络和两个训练有素的定性分数预测器一起训练。4PP-EUSR模型输出三个4倍上采样的图像，判别器被训练以区分真实图像和上采样图像。

采用六种损失函数，包括：

重建损失 (lr): 像素级L1损失，公式为：



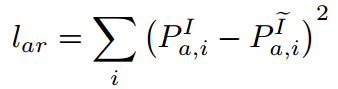
对抗损失 (lg): 使用sigmoid交叉熵损失，公式为：



审美分数损失 (las)：



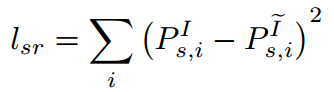
审美表示损失 (lar)：



主观分数损失 (lss) ：



主观表示损失 (lsr)：



所有损失函数计算后取平均，使用Adam优化方法训练4PP-EUSR和判别器，学习率为10^-5和2×10^-5。整个训练过程是在TensorFlow框架上实现的，通过多阶段训练和多种损失函数的结合，旨在提升超分辨率图像的定量和感知质量。

# 方法评估

## 3.1 优点

4PP-EUSR模型在提升上采样图像的感知质量的同时，保持了传统的定量性能。实验结果显示，该模型在定量质量（如PSNR和SSIM）和感知质量（如NIQE、SR评分和PI）之间取得了良好的平衡。



表1 原论文结果

图5为本地测试结果。

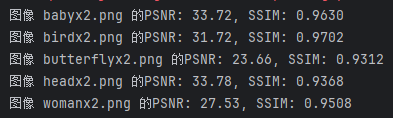


图5 本地测试结果

复现4PP-EUSR模型的结果时，得到的PSNR（峰值信噪比）和SSIM（结构相似性）值比原文中报告的值要高，可能的原因如下：

训练数据的差异：复现时使用的训练数据集为BSD100数据集，与原文中使用的DIV2K数据集不同，即使是微小的变化也可能影响最终的评估结果。

**模型超参数**：如果模型的超参数（如学习率、优化器参数、批量大小等）设置不同，可能会导致性能上的显著差异。

**随机性**：深度学习训练过程中存在随机性，如参数初始化、数据洗牌等，这可能导致每次训练的结果都略有不同。

除此之外，模型在很多方面还具有优势：

通过在训练阶段使用多通道上采样，模型能够学习减少直接上采样、两次通过×2通道上采样以及通过×8通道上采样后再下采样可能产生的伪影。这种方法有助于模型适应不同的上采样场景；模型采用了两个定性分数预测器，直接评估上采样图像的美学和主观质量分数。这些预测器基于人类评价者的数据集训练，有助于生成感知质量更高的图像；通过使用判别器网络，模型在训练过程中学习生成更自然的图像。判别器试图区分真实图像和上采样图像，而超分辨率模型则训练为使判别器难以进行区分；4PP-EUSR模型可以根据定量和感知质量的不同优先级，通过调整损失函数的权重来训练，提供了在定量和感知质量之间进行权衡的灵活性；与其他一些模型相比，4PP-EUSR模型拥有较少的参数数量，这可能有助于减少计算资源的消耗，并可能加快模型的运行速度；在训练过程中，模型学会了减少在上采样过程中可能出现的特定伪影，如格网状纹理和椒盐噪声，从而生成更自然、更高质量的图像。

## 3.2 缺点

本次模型训练300个epoch，训练了大约三个小时，有较高的计算成本和较长的训练时间。虽然4PP-EUSR模型的参数数量相对较少，但相比于更简单的模型，它可能仍然需要更多的存储空间和计算资源。此外，模型使用了多种损失函数和多个训练阶段，这可能导致调参过程变得复杂，需要大量的实验来找到最优的参数组合，而且尽管模型在特定数据集上表现良好，但可能在未见过的新数据上泛化能力不足，特别是如果训练数据和实际应用中的数据分布不一致时。

## 3.3 改进方向

**计算效率**：优化模型以减少计算资源的需求，使其更适合实时或资源受限的应用场景。

**轻量化模型**：通过模型剪枝、量化或知识蒸馏等技术，减少模型的大小和复杂性，同时保持性能。

**泛化能力**：通过在更多样化的数据集上训练，或使用数据增强技术，提高模型对不同来源图像的泛化能力。

**感知质量的优化**：进一步研究和改进感知质量的度量方法，使模型能更好地捕捉人类视觉感知的特性。

**对抗性训练的稳定性**：研究更稳定的对抗性训练方法，以提高模型在感知质量方面的性能。

**损失函数的改进**：设计新的损失函数或改进现有损失函数，以更准确地反映图像的视觉效果和质量。

**注意力机制**：引入或改进注意力机制，使模型能更集中于图像的关键部分，提高上采样的质量。

**跨模态迁移**：探索将模型从一个数据集迁移到另一个不同但相关的数据集的能力。

# 实验结果

## 4.1 图像超分辨率重构

图6为对set5数据集进行下采样2倍的图像和重构图像对比。

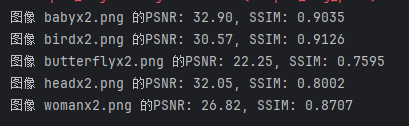


图6 重构图像对比

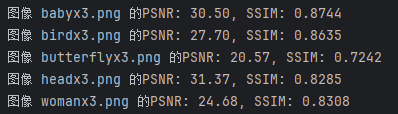
**将输入输出图像存储在results文件夹中。**

## 4.2 PSNR、SSIM指标

下采样X2：



下采样X3:



下采样X4:

