

恢复真实细节的放大任意图像超分辨率

王嵩

摘要

超分辨率任务旨在保证图像的质量的同时、提升图像的分辨率以得到高清的图像。基于深度学习的超分技术这几年得到了迅速的发展，研究的方向从最初单尺度超分变成现在的任意尺度的超分。之前所提出的 LIIF 只能对模糊图像进行超分辨率处理，并且缺乏生成令人愉悦的感知细节的能力。而这篇论文提出隐式像素流（IPF）来建模模糊 INR 分布和尖锐真实世界分布之间的坐标相关性。对于模糊边缘附近的每个像素，IPF 为像素的坐标指定偏移，以便原始 RGB 值可以被更适合形成更清晰边缘的相邻像素的 RGB 值替换。

法线图在定义和提取形状分量和细节分量后，进行归一化至 0-255 的范围后能作为数字图像进行处理，故也可进行超分的任务。本次课程的论文复现任务拟对该模型进行微调，将输入数据由 RGB 图像改为法线图，通过对法线图进行处理，完成法线图的超分。

关键词：法线图；图像超分辨率；任意尺度

1 引言

在科学技术飞速发展的今天，摄像机的技术也不断发展，可以得到各种的超高清的图像，给人们的生活和娱乐带来了许多的便利。这时，人们对于低分辨率的老照片也有了高清的需求，希望模糊的老照片也能变得高清。这篇论文提出通过建模清晰图像的像素分布与其模糊对应点之间的内在关系来生成具有锐利边缘的照片逼真图像，他们发现模糊的边缘可以通过压缩周围的像素来很容易地转换为清晰的边缘，受此启发，在恢复锐利边缘时，用邻近位置更合适的值替换每个像素的原始 RGB 值。

超分辨率任务旨在保证图像的质量的同时、提升图像的分辨率以得到高清的图像。基于深度学习的超分技术这几年得到了迅速的发展，研究的方向从最初单尺度超分变成现在的任意尺度的超分。这篇论文对 2021 的 LIIF 模型进行改进，可以在完成任意分辨率的超分的同时，丰富图像的细节纹理。法线图在定义和提取形状分量和细节分量后，进行归一化至 0-255 的范围后能作为数字图像进行处理，故也可进行超分的任务。本次课程的论文复现任务拟通过对法线图进行处理，完成法线图的超分的基础上，对细节纹理进行更好的处理。

2 相关工作

2.1 单一图像超分辨率

单一图像超分辨率旨在从低分辨率 (LR) 图像中恢复高分辨率 (HR) 图像。最近的 SISR 方法主要有两个分支，面向 psnr 的方法和感知驱动的方法。第一类中的方法通常通过精心设计的生成器输出 SR 图像。这样的 SR 结果通常具有较高的 PSNR 值，因为生成器是根据像素级损失的目标函数优化的，包括 L1 损失和 MSE 损失。这一类中的大多数方法，包括 ESPCN^[1]，EDSR^[2]，RDN^[3]等，都是利用亚像素卷积来生成 SR 图像。这些模型只能处理具有单个整数上升因子的 SR。

近年来，任意放大倍数的超分辨率引起了社会的广泛关注。Hu 等通过动态预测上尺度滤波器的权重提出了 Meta-SR^[4]。这是第一个实现任意上标因子超分辨率的工作。Wang 等^[5]设计了一个插件模

块，将特定比例的模型转换为任意比例的模型，成功地解决了不对称 SR 问题。这些方法大多受到像素级损失的监督，这可能会给大因子 SR 恢复带来模糊。

2.2 隐式神经表征

内隐神经表征 (INR) 在三维视觉任务中得到了广泛的应用。由于它具有高效连续表示的特性，相对于点云、体素和网格等其他形式的 3D 数据具有明显的优势。最近的一些超分辨率工作也成功地用隐式函数表示二维视觉数据。Sitzmann 等^[6]提出了正弦表示网络，通过周期性激活来建模高质量的细节。Dupont 等^[7]提出了一个生成模型来表示二维和三维连续信号。Skorokhodov 等^[8]应用 INR 进行高分辨率图像生成。虽然上述两种方法可以从随机采样的潜在编码中生成感知愉快的图像，但它们无法编码真实世界的图像和超分辨率的任意给定图像。

LIIF^[9]提出了连续图像表示，可以以任意分辨率表示。然而，在 LIIF 中只考虑 L1 损失，这导致图像模糊。自然而真实的细节对于任意放大的超分辨率也是必要的。不同的是，所复现论文^[10]的目标是学习隐式像素流，通过修改空间对应关系，将原始模糊的 INR 分布转换为清晰的 INR 分布。

3 本文方法

3.1 本文方法概述

论文提出隐式像素流 (IPF) 来建模模糊 INR 分布和尖锐真实世界分布之间的坐标相关性。对于模糊边缘附近的每个像素，IPF 为像素的坐标指定偏移，以便原始 RGB 值可以被更适合形成更清晰边缘的相邻像素的 RGB 值替换。通过 IPF 修改 INR 域坐标和图像域像素之间的关系，将原始模糊的 INR 分布转换为清晰的 INR。

图像的连续表示的定义是通过拟合一个连续函数，对图像进行预处理得到潜在码，最后实现由图像像素坐标得到 RGB 值的预测，从而实现图像的任意尺度上采样。具体来说，就是通过训练一个 Encoder 来提取图像特征，形成潜在码，再由像素坐标作为 Decoder 的输入得到 RGB 值。具体公式如下，

$$M = E(I) \quad (1)$$

$$v_{x,y} = f(m^*, [x, y, c_h, c_w]) \quad (2)$$

其中 E 表示具有标准 SR 架构的编码器，用于提取输入图像的局部连续表示。

论文方法基于 e 得到的连续图像表示，通过 F 提取像素流表示，然后通过隐式像素流函数 g 输出像素流值。我们的方法还使用了另一个隐式函数 C 来计算隐式置信度 (IC)。得到的像素流经过隐式置信度 c 调制后，加入到高分辨率坐标中，使 LIIF 的模糊分布变得更加清晰。灰色箭头表示稳定优化过程的双重约束模块。用彩色编码图对像素流场进行可视化。中心白色表示不需要偏移，其他颜色表示不同方向和幅度的像素流向量，具体的模型如图 1 所示：

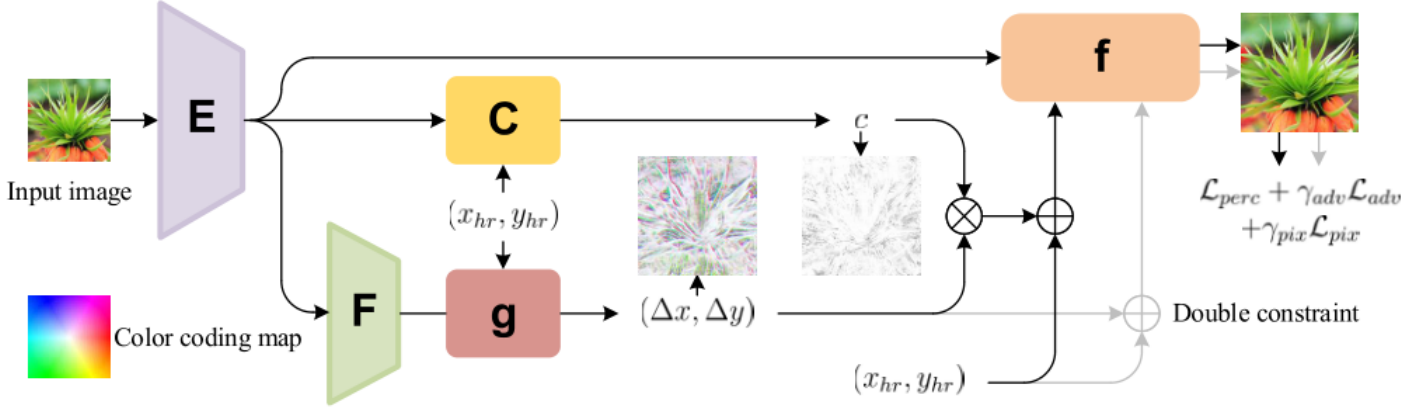


图 1: 模型示意图

3.2 隐式像素流模块

通过观察模糊图像和清晰图像之间的差异发现，通过将模糊图像附近的像素点靠近边缘，可以很容易地获得清晰的边缘，从而使颜色变化更快，边缘变得更清晰。根据隐式图像函数的性质，我们可以将上述观察表述如下：

$$v_{x,y} = f(m'*, [x', y', c_h, c_w]) \quad (3)$$

其中 $m'*$ 是最接近 (x, y) 的特征。这个方程意味着 (x, y) 的 RGB 值由 (x', y') 而不是 (x, y) 决定，而 $(x', y') = (x + \Delta x, y + \Delta y)$, $(\Delta x, \Delta y)$ 表示偏移距离，在原论文中， $(\Delta x, \Delta y)$ 被定义为像素流。则上式可以表示为：

$$v_{x,y} = f(m'*, [x + \Delta x, y + \Delta y, c_h, c_w]) \quad (4)$$

$$(\Delta x, \Delta y) = g(p, x, y, c_h, c_w) \quad (5)$$

式中， $g(\cdot)$ 表示隐式像素流函数， p 为像素流 p 在坐标 (x, y) 附近的特征图得到的潜码。具体地说，我们以连续表示 M 为先验提取像素流表示 P 。我们使用另一个卷积神经网络 F 提取 P 给定 M 作为输入。即 $P = F(M)$ ， P 与 M 具有相同的尺度。 F 用于将连续图像表示修改为适合计算坐标偏移值的版本，而不是 RGB 值。因此，原始连续图像表示 M 和我们提出的像素流表示 P 构成了感知驱动隐式图像函数的新表示。

3.3 损失函数定义

论文的损失函数分为 I^f 和 I^d 两部分， I^f 由感知损失，生成对抗损失和像素损失三个部分组成，具体公式如下：

$$\zeta^f = \zeta_{perc} + \gamma_{adv} \zeta_{adv} + \gamma_{pix} \zeta_{pix} \quad (6)$$

γ_{adv} 和 γ_{pix} 是平衡不同损失的权重。除了对 I^f 的监督之外，我们还对由偏移量 $(\Delta x, \Delta y)$ 计算的预测 RGB 值 I^d 施加了另一个双重约束监督，而没有对隐式置信度 $c_{x,y}$ 进行调制：

$$I_{x,y}^d = \sum \frac{S_t}{S} f(m'*, [x + \Delta x, y + \Delta y, c_h, c_w]) \quad (7)$$

则 I^d 的损失函数可同样定义为 ζ^d ，因此，最终损失函数定义为：

$$\zeta = \zeta^f + \zeta^d \quad (8)$$

在训练和测试过程中，将采样的 HR 坐标 (x_{hr}, y_{hr}) 赋给 (x, y) ，用于计算 I^f 和 I^d 。

4 复现细节

4.1 与已有开源代码对比

该论文没有开源代码，但是工作是基于 LIIF^[9]的工作进行的，所以复现该论文时参考了 LIIF 的代码。而本次课程的论文复现的数据并不是自然图像，而是对法线图，旨在通过对法线图进行处理，完成法线图的超分的基础上，恢复出更清晰的三维表面纹理，从而实现对三维表面的超分。

与 LIIF 相比，复现的代码做了如下的改进：

1. loss 模块增加了感知损失和生成对抗损失，使其变成一个 GAN 网络模型。
2. 因为是对法线图进行超分，而法线图记录的是三维表面上对应位置上的法向量 (x,y,z) ，并不能将其直接超分，我们的做法是将法线图分离成细节分量和形状分量两部分，其中形状分量是用低通滤波器实现平滑操作后得到的，存储的信息是邻域内的平均法线，只表现原来的法线图的形状信息，细节分量是输入法线图和其形状分量之间的相对关系，我们得到的细节分量实际上和形状分量的相对法线，表现为原来法线图的细节信息。因此，在图片载入的代码中加入了法线图分离细节分量和形状分量的模块。
3. 我们的数据集的均值和 LIIF 中的图像均值是不一样的，因此在图像归一化时，修改了均值，改为 $[0.422369, 0.426868, 0.836872]$ 。

4.2 实验环境搭建

所有的模型都由 PyTorch 实现，并在 Nvidia 3090 GPU 上进行训练，每 200 个 epoch 将学习率 10⁻⁴ 降低到一半，直到微调停止在 1000 个 epoch。对于超分辨率图像的质量，只用了 PSNR 作为指标。

为了训练隐式像素流函数，我们首先确定预训练 LIIF^[9]编解码的参数，使用 8 个 EDSR^[2]块来构建流量编码器 F。隐式像素流函数和隐式置信度函数分别为 5 层 MLP 和 4 层 MLP，分别包含 256 个隐藏神经元和 64 个隐藏神经元。

4.3 创新点

- 将法线图当作一种数字图像的数据进行超分，并用超分后的法线图三维表面重建，从而实现三维表面的超分。
- 将图像超分模型迁移到法线图上，结合三维重建的方法，使用二维数据的超分方法来实现三维超分，并且有一定的效果。

5 实验结果分析

5.1 实验结果

实验结果分析包括法线图对比和三维重建结果对比两个部分，其中法线图对比是超分结果和 Groundtruth 的对比，三维重建结果对比则是将两者进行三维重建后的对比。

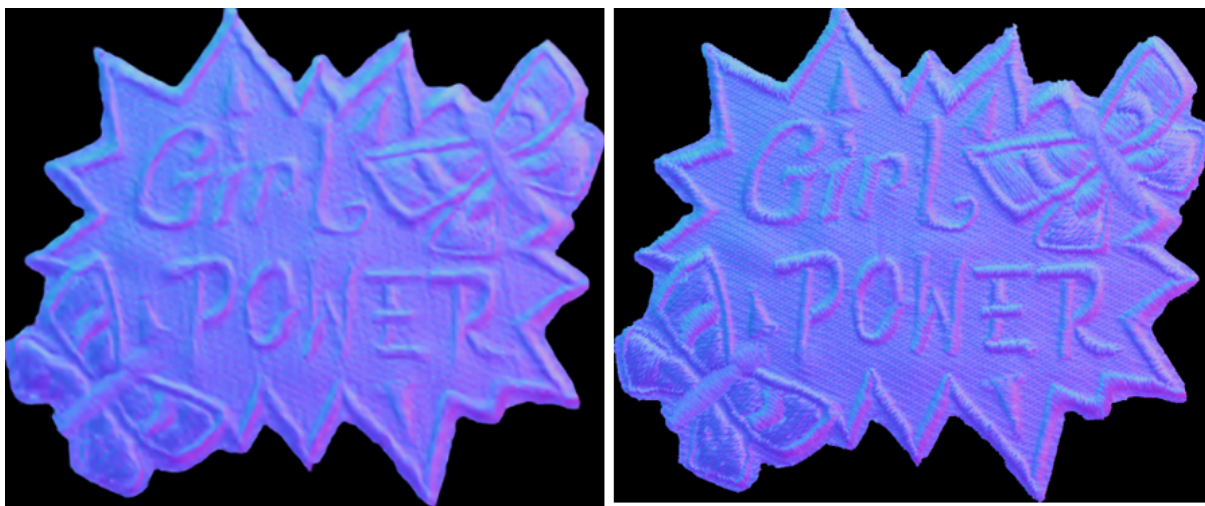


图 2: x8 超分结果对比

左边是训练出的模型对下采样 x8 的法线图进行超分后的结果，右边是数据集的 GroundTruth 法线图。

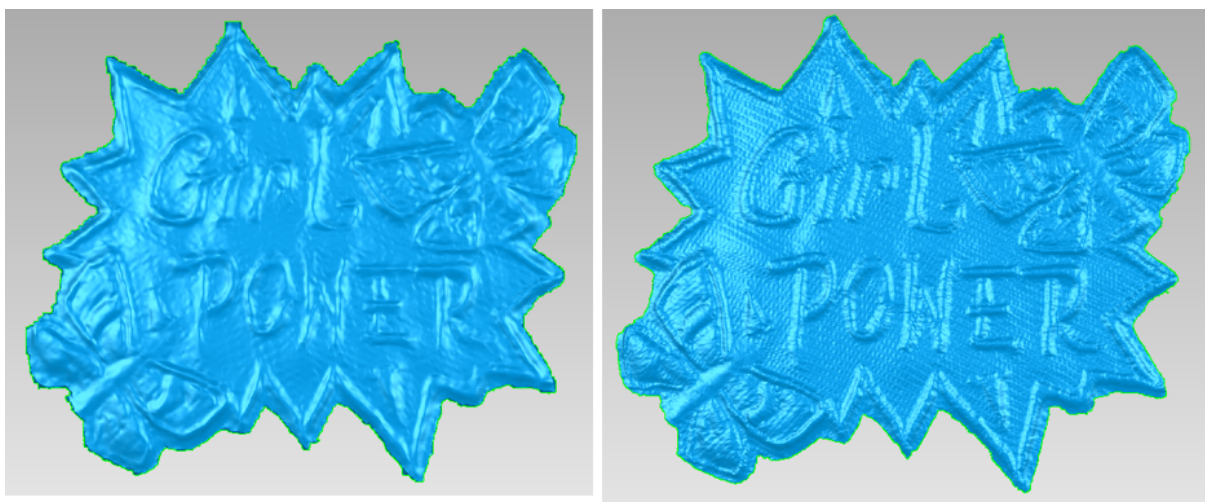


图 3: x8 三维结果对比

左边是训练出的模型对下采样 x8 的法线图进行三维表面重建后的结果，右边是对法线图数据集中的 GroundTruth 法线图进行三维表面重建的结果。

5.2 分析

从法线图的超分结果上看，对于法线图上较大的细节，模型能够很好的恢复出来，而对于表面细致的纹理，恢复的效果较差，基本上没有恢复出细节。可能的原因是，x8 下采样的数据集中的法线图分辨率过于低，平均分辨率只有 40 作用，这就使得 x8 下采样后的法线图无法提供很好的细节信息，细节恢复效果固然也会较差。

从三维表面重建的结果上看，超分后的法线图恢复出的三维表面形状比较好，但由于法线图本身超分效果较差，使得三维表面重建后的细节也较差，恢复的三维表面过于光滑。

6 总结与展望

这篇论文提出隐式像素流（IPF）来建模模糊 INR 分布和尖锐真实世界分布之间的坐标相关性。对于模糊边缘附近的每个像素，IPF 为像素的坐标指定偏移，以便原始 RGB 值可以被更适合形成更清晰边缘的相邻像素的 RGB 值替换。我根据论文提出的方法，参考 LIIF 的代码进行论文方法的复现，

并将原论文的数据改为使用法线图超分，从而实现基于法线图的三维表面超分。代码复现有一定成果，基本能够实现法线图的超分，也能用超分后的法线图进行三维表面重建，但重建的效果较差，还有很大的提升空间。

在复现代码的过程中，我学习到了如何去搭建一个深度学习网络，包括数据集如何作为输入来训练，不同训练集之间如何对模型进行微调。同时，测试过程中也学习到了如何调试模型去达到最优的效果。这一复现的过程，大大提高了我的编程能力，以及模型搭建的能力，为我后面做自己的课题奠定了很好的基础。

参考文献

- [1] SHI W. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[J]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 1874-1883.
- [2] B.Lim. Enhanced deep residual networks for single image super-resolution[J]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 136-144.
- [3] Y.Zhang, Y.Tian. Residual dense network for image super-resolution[J]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 2472-2481.
- [4] HU X, MU H. A magnification-arbitrary network for super-resolution[J]. IEEE Conference on Computer Vision and Pattern Recognition, 2019: 1575-1584.
- [5] WANG L, WANG Y . Learning a single network for scale-arbitrary super-resolution[J]. IEEE International Conference on Computer Vision, 2021: 4801-4810.
- [6] SITZMANN V, MARTEL J N P. Learning a single network for scale-arbitrary super-resolution[J]. arXiv, 2020: 2006.09661.
- [7] DUPONT E, TEH Y W, DOUCET A. Generative models as distributions of functions[J]. arXiv, 2021: 2102.04776.
- [8] SKOROKHODOV I, IGNATYEV S, ELHOSEINY M. Adversarial generation of continuous images[J]. arXiv, 2020: 2011.12026.
- [9] CHEN Y, LIU S, WANG X. Learning continuous image representation with local implicit image function [J]. IEEE Conference on Computer Vision and Pattern Recognition, 2021: 8628-8638.
- [10] MA C, YU P, LU J. Recovering Realistic Details for Magnification-Arbitrary Image Super-Resolution [J]. IEEE TRANSACTIONS ON IMAGE PROCESSING, 2021: 3669-3683.