

# 零样本超分辨率重建——计算机视觉课程报告

张伟锋

21821163

计算机科学与技术学院

21821163@zju.edu.cn

## 摘要 (Abstract)

在本课程作业中,本人介绍的是Assaf Shocher等发表<sup>1</sup>在CVPR2018上的文章“Zero-Shot” Super-Resolution using Deep Internal Learning[1]。在这篇文章中首次提出基于卷积神经网络的无监督图像超分辨率算法。本文主要介绍了这篇文章的研究背景,创新点和特色,以及算法的流程,并通过实验验证了论文中提到的主要实验结论。

### 1. 研究背景

图像超分辨率 (Super-Resolution, SR) 问题是指利用单张或多张低分辨率的图像复原出原始的高分辨率的图像,本文我们关注的是基于单张图片的超分辨率问题。图像超分辨率有相当广泛的应用,如在计算机图像处理软件中对图像的尺寸进行调整,获取更高质量的医学和卫星图像,对老旧照片或绘画进行修复等。

根据综述[2]的分类,单图片超分辨率方法可以大致分为基于重构 (reconstruction-based) 的方法和基于学习 (learning-based) 的方法。

基于重构的方法有基于插值的方法,如其中双三次插值 (Bicubic Interpolation) [3]是应用相当广泛的方法,常作为其他算法的预处理方法。此外,还有基于图像内部特征的方法,如Sun等[4]引入梯度场先验 (Gradient Profile Prior, GPP) 作为从低分辨率图像复原高分辨率图像的约束。梯度场先验可以用来描述一幅图片梯度的形状和尖锐程度,并且其数值在图像分辨率发生改变时仍然相当稳定。这一方法可以有效地保持图像中的边缘等细节。Glasner等[5]则发现一些图像块 (patches) 会以相同尺度或不同尺度在同一幅自然图像中重复出现,基于这一观察,作者设计出一种在高斯金字塔中寻找相邻尺度层中相似图片块来复原超分辨率图像的方法。

基于学习的方法通常从包含大量高低分辨率对 (HR-LR pairs) 的数据集中学习高低分辨率图像的内在联系。早期的方法有基于流行学习的方法和基于稀疏编码的方法。Chang等[6]受到流形学习中的局部线性嵌入

(Locally Linear Embedding) 的启发,认为高分辨率和低分辨率图像中相对应的两个图像块在各自的特征空间中形成具有相似局部几何结构 (local geometry) 的流形。Yang等[7]则借助压缩感知领域的理论,认为高分辨率和低分辨率图像中相对应的两个图像块在各自的字典下具有相同的稀疏表示。因此,根据低分辨率图像块得到其在低分辨率字典下的稀疏表示,结合高分辨率字典就可以重构出高分辨率的图像块。随着近年来深度神经网络在计算机视觉领域的快速发展,深度学习方法逐渐成为图像超分辨率算法领域的主流。其中,最为经典的方法是SRCNN[8],这一方法先使用传统方法 (如双三次插值) 将低分辨率图像扩展到高分辨率图像的大小,然后利用深度神经网络学习这一输入图像与高分辨率图像的残差。随后的许多方法都是在这一方法的基础上改进网络结构、损失函数或学习策略,具体见综述[9]。

以上提到的基于学习的方法通常都需要大量的高低分辨率图像对作为训练数据,而我们将要介绍的这篇论文“Zero-Shot” Super-Resolution using Deep Internal Learning (ZSSR) [1]不需要任何训练集,而是利用单一图像内部结构的重现性来训练一个与该图像相关的卷积神经网络,是第一个基于卷积神经网络的无监督图像超分辨率算法。

### 2. 论文的创新点

该论文最大的创新点在于指出了基于学习的方法普遍存在的一个问题,即难以处理非理想条件下的超分辨率问题。基于学习的方法都依赖于大量的理想条件下的高低分辨率对作为训练数据,即在训练数据中,从所有的高分辨图像得到对应低分辨率图像的过程通常是固定的、一致的,例如具有相同的下采样核 (downsampling kernel)。但是在现实中,这一过程在不同的图像之间通常是不同的,例如相机和传感器不同,或是成像条件 (抖动、能见度等) 的不同,这导致了不同的高低分辨图像对具有不同的下采样核和不同的噪声特性。在这样的非理想的条件下,基于监督学习的单一的CNN结果无法泛化到能处理所有不同的条件,因此基于学习的方法很难在非理想的条件下取得很好的结果。

为了解决这一问题,作者提出不依赖于任何其他的

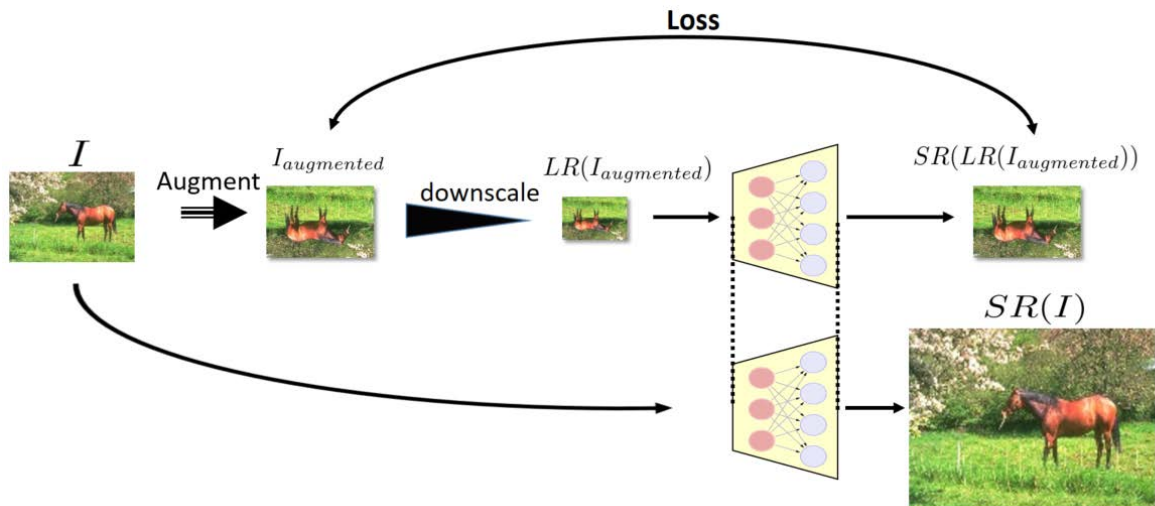


图 1 ZSSR模型示意图

训练数据，而是显示地利用图像内部信息的重现性为每一张特定的图片单独训练一个轻量级的CNN。这样，训练好的CNN可以适应该特定图片的特性，使得我们可以在处理该图片特定条件下的超分辨率问题。在理想条件下，ZSSR的效果可以匹敌利用大量训练数据得到的模型；而在非理想条件下，ZSSR的效果显著优于当前最好的模型。具体来说，论文中提出的ZSSR模型具有以下特性：

（1）由于其为每幅图像训练一个特定的CNN网络，模型可以处理非理想条件下的超分辨率问题。

（2）模型不需要预训练，并且由于其很轻量，对计算资源要求很小。

（3）在产生高分辨图像时（测试时），如果给定从高分辨率到低分辨率图像产生过程的真实信息，如下采样核、噪声特性等，ZSSR模型可以利用这些信息得到更好的效果。而这些信息对于基于监督学习的模型通常用处不大，因为他们需要根据这些信息生成新的高低分辨率图像对，重新进行模型的训练，而这一过程将相当耗时。

其实，本文使用的方法并非完全的创新，因为使用自然图像中图像块的重现性进行超分辨率计算已经在[5]中提出，本文的作者之一也是该文章的作者，ZSSR中训练CNN的方法也和SRCNN相似，但文章最为创新的地方就在于指出了基于学习的方法的不足，并提出了合适的解决方法。此外，实验的论证也相当充分。

### 3. ZSSR模型

#### 3.1. 理论基础

ZSSR模型的理论基础是自然图像的一些统计学特

征。在自然图像中，小的图像块会以相同尺度或不同尺度在同一幅图像中重复出现，文献[5]和[10]通过几百张自然图像经验性地证实了这一规律，并证明对于几乎所有的自然图像上几乎所有小图像块，这一规律都成立。

此外，[10]经验性地表明一幅图片中图像块的内部熵（internal entropy）远小于一组图片中图像块的外部熵（external entropy），这进一步表明了一幅图像内部的统计特征和一组图像的外部统计特征相比，具有更强的预测能力。

因此，ZSSR模型利用了图片自身的内在信息，结合深度学习，提出了对每个图片训练特定CNN的方法来解决图像超分辨率问题。

#### 3.2. 算法流程

ZSSR模型如图 1 所示，其主要思想是：在训练阶段，把低分辨率图像X当做高分辨率图片，用它得到更低分辨率的图片，用这些训练CNN，使其学习从低分辨率图像预测高分辨率图像。在预测阶段，把低分辨率图像作为CNN的输入，预测对应的高分辨率图像。

算法流程如下所示

输入：低分辨率图像X，放大尺度因子  $s$ ，最大迭代次数  $\max\_iter$ ，学习率阈值  $\alpha_0$

输出：高分辨率图像Y

##### 1. 利用单幅图片构建训练集。

1.1 将低分辨率图像X进行数据扩张，包括是不同尺度的缩放，4 种角度的旋转（ $0^\circ$ ， $90^\circ$ ， $180^\circ$ ， $270^\circ$ ）和水平、竖直 2 个方向的镜面对称，得到的图像称为高分辨率父类（HR fathers）

1.2 将HR fathers按给定的下采样核（若未指定，则使用双三次插值）和缩小因子  $1/s$ 进行下采样，得到低分辨率子

类 (LR sons)

## 2. 利用训练集训练CNN。

while (迭代次数 < max\_iter) or (学习率  $\alpha > \alpha_0$ ), do

2.1 从HR fathers中选取固定大小 (128×128) 的图像块  $P_H$ ，将LR sons中对应的图像块，根据给定的噪声类型添加噪声 (默认为无噪声)，在此基础上使用双三次插值，得到与  $P_H$  大小相等的图像块  $P_L$

2.2 训练一个CNN使其能预测  $P_H$  和  $P_L$  之间的残差  $P_R$ ，优化目标为最小化 ( $P_L + P_R - P_H$ ) 的L1范数

2.3 根据重构误差的斜率和方差，动态调整学习率

## 3. 使用训练好的CNN计算超分辨率图像

3.1 将X经过旋转和镜面对称得到的 8 幅图像分别作为CNN的输入，得到HR fathers的初始超分辨率图像

3.2 将 8 幅初始超分辨率图像进行旋转和镜面对称的逆变换，取得到的 8 幅图像中对应像素点的中位数构成最终的超分辨率图像Y

下面，我们对算法中的一些具体细节进行说明。

在第 1 步中，由于训练集只包含一张图片，作者对其进行数据扩张，以得到更多的高低分辨率图像对。此外，当图片很小而放大尺度因子  $s$  很大时，为了提升算法的鲁棒性，作者提出了一种渐进式 (gradually) 的方法。即在 1 和  $s$  之间，选取一系列中间值作为放大尺度因子 ( $s_1 < s_2 < \dots < s_m = s$ )。从  $s_1$  开始，依次在这些放大尺度因子下进行训练和预测，而最终预测的结果及其数据扩张后的结果会加入到HR fathers，用于更大的放大尺度因子下的训练。实验表明使用该方法在PSNR指标上有约 0.2dB的提升。

在第 2 步中，算法中使用固定大小的图像块而可以加快训练速度，并使得运行时间与图片X的大小无关。此外，模型中使用的CNN的结构是：8 层全卷积网络，每层 64 个通道，使用ReLU作为激活函数，优化器为ADAM，初始学习率为 0.001。

而在第 3 步中，在 3.1 和 3.2 后均可以加入反向投影 (back-projection) 模块，进一步微调以得到更好的结果。

## 4. 实验与分析

### 4.1. 实验准备

为了验证ZSSR模型的效果，我们在Set14数据集上对比Bicubic插值，SRCNN和ZSSR模型的效果<sup>1</sup>。其中Bicubic插值使用的是Matlab上的标准函数imresize，SRCNN使用的是原论文作者提供的测试代码<sup>2</sup>，ZSSR

使用的也是论文作者提供的基于Python 2 的代码<sup>3</sup>，但做了部分修改，使其能在Python 3 环境下运行。我们原计划比较论文中的主要基线模型EDSR+，但由于该模型较为庞大，训练时间非常长，故没有比较。评价指标为超分辨率领域常用的PSNR和SSIM。实验环境为CPU：Intel Core i5-4570，GPU：GeForce GTX 1060。

### 4.2. 实验结果

我们选择了 3 幅图片的结果作为展示，以下均为长宽方向各放大 2 倍的结果。图片对应的说明为“所用方法: PSNR值/SSIM值”，其中ZSSR\_one和ZSSR\_gradually分别表示不使用和使用渐进式方法。Set14 上的其他实验结果见项目中的results文件夹。



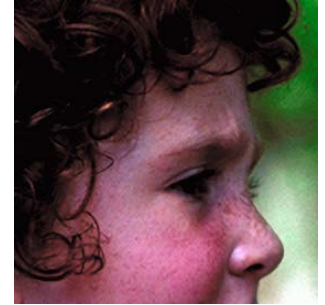
Bicubic: 31.44/0.8182



SRCNN: 31.96/0.8405



ZSSR\_one: 32.14/0.8486



ZSSR\_gradual: 32.20/0.8499



Bicubic: 28.55/0.8920



SRCNN: 30.59/0.9118



ZSSR\_one: 31.65/0.9332



ZSSR\_gradual: 32.25/0.9384

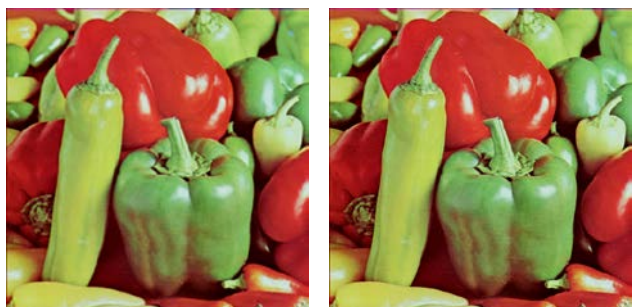
<sup>1</sup> 项目的Github地址为

<https://github.com/Weifeng73/Zero-Shot-Super-resolution>

<sup>2</sup> <http://mmlab.ie.cuhk.edu.hk/projects/SRCNN.html>

<sup>3</sup> <https://github.com/assafshocher/ZSSR>





Bicubic: 30.49/0.8417

SRCNN: 31.05/0.8533



ZSSR\_one: 32.34/0.8646

ZSSR\_gradual: 32.29/0.8624

在以上三组超分辨率图像中,Bicubic插值得到的图像都出现了模糊的现象,而SRCNN和ZSSR视觉效果接近。定量来看,ZSSR在PSNR和SSIM指标上明显优于SRCNN,结果与论文基本一致,可以认为,ZSSR模型确实也从单张图片中学习到图片的内部特征,而与外部的大量训练数据相比,这些特征对于构建超分辨率图像更有帮助。此外,我们也发现,渐进式的方法总体来说由于非渐进的方法,但也不一定总能得到更好的效果。

为了验证文章中提到的非理想条件情况,我们重复了论文中展示的在随机生成的下采样核和真实图片下的实验。结果如下所示:



Bicubic

SRCNN



ZSSR

Bicubic



SRCNN

ZSSR

在以上两种情况均没有高分辨率图像的真实值,但从视觉效果来看,Bicubic和SRCNN都出现了明显的模糊,而ZSSR生成的图片更为清晰,这表明对比插值方法和基于大量训练数据学习的方法,ZSSR确实可以在非理想条件下重构出质量更高的超分辨率图片。

#### 4.3. 结论

通过实验,我们验证了ZSSR模型在理想和非理想条件下均能得到较好的超分辨率图像,并且在非理想条件下,明显优于所对比的基线模型,主要原因是ZSSR能有效地利用真实的下采样核等信息。ZSSR模型较为轻量,在本人的电脑上训练和测试一幅图片的时间在1分钟左右,尽管已经相当短,但由于要针对每张图片进行训练,还不能满足一些应用的实时性要求。

#### Reference

- [1] Shocher A , Cohen N , Irani M . "Zero-Shot" Super-Resolution using Deep Internal Learning[C]. in Proc. of CVPR, 2018.
- [2] Nasrollahi K , Moeslund T B . Super-resolution: a comprehensive survey[J]. Machine Vision and Applications, 2014, 25(6):1423-1468.
- [3] Keys R G . Cubic convolution interpolation for digital image processing[J]. IEEE Trans. on Acoust. Speech. & Signal Processing, 1981, 37..
- [4] J. Sun, J. Sun, Z. Xu, and H.-Y. Shum, Image super-resolution using gradient profile prior. in Proc. of CVPR, 2008.
- [5] D. Glasner, S. Bagon, and M. Irani, Super-resolution from a single image[C], In Proc. of ICCV, 2009
- [6] Chang H, Yeung D Y, Xiong Y. Super-resolution through neighbor embedding[C]. in Proc. of CVPR, 2004. 1: 1-1
- [7] Yang J, Wright J, Huang T S, et al. Image super-resolution via sparse representation[J]. IEEE transactions on image processing, 2010, 19(11): 2861-2873.
- [8] Chao Dong, Chen Change Loy, Kaiming He, Xiaoou Tang. Learning a Deep Convolutional Network for Image Super-Resolution, in Proc of ECCV, 2014
- [9] Wang Z, Chen J, Hoi S C H. Deep learning for image super-resolution: A survey[J]. arXiv preprint arXiv:1902.06068, 2019.
- [10] M. Zontak and M. Irani. Internal statistics of a single natural image. in Proc. of CVPR, 2011.