**Accurate Image Super-Resolution Using Very Deep Convolutional Networks**

**（**准确的图像超分辨率使用很深的卷积网络**）**

**摘要**

我们提出一种高精度的单图像超分辨率（SR）方法。 我们的方法使用了一个非常深的卷积网络，灵感来自用于ImageNet分类的VGG-net [19]。 我们发现，增加我们的网络深度表明精度的显着提高。 我们的最终模型使用20个重量层。 通过在深层网络结构中多次级联小的滤波器，在大的图像区域上的上下文信息被有效地利用。 然而，在非常深的网络中，收敛速度在训练期间成为一个关键问题。 我们提出一个简单而有效的培训程序。 我们只学习残差，使用极高的学习率（比SRCNN [104]高104倍），通过可调节的梯度限幅。 我们提出的方法执行比现有方法更准确，视觉改进在我们的结果很容易注意到。

1. **介绍**

我们解决了给定低分辨率（LR）图像生成高分辨率（HR）图像的问题，通常称为单图像超分辨率（SISR）[12]，[8]，[9]。 SISR广泛应用于计算机视觉应用，从安全和监控成像到医疗成像，需要更多的图像细节。在计算机视觉领域已经研究了许多SISR方法。早期的方法包括插值，如双三次插值和Lanczos重采样[7]更强大的方法利用统计图像先验[20,13]或内部补丁复现[9]。目前，学习方法广泛用于模拟从LR到HR补丁的映射。邻居嵌入[4,15]方法内插补丁子空间。稀疏编码[25,26,21,22]方法使用基于稀疏信号表示的学习的紧凑词典。最近，随机森林[18]和卷积神经网络（CNN）[6]也被用于精度的大幅提高。其中，Dong et al。 [6]已经证明，CNN可以用于以端到端的方式学习从LR到HR的映射。他们的方法，称为SRCNN，不需要任何工程特征，在其他方法中通常是必要的[25,26,21,22]，并显示了最先进的性能。虽然SRCNN成功地将深度学习技术引入超分辨率（SR）问题，我们在三个方面找到其局限性：第一，它依赖于小图像区域的上下文;第二，训练收敛太慢;第三，网络只适用于单一规模。在这项工作中，我们提出一种新的方法来实际解决问题。上下文我们利用分布在非常大的图像区域上的上下文信息。对于大比例因子，通常情况下，包含在小补丁中的信息对于细节恢复（不适当）是不够的。我们使用大的接受域的非常深的网络考虑了大的图像上下文。收敛我们建议一种加速培训的方法：剩余学习CNN和极高的学习率。因为LR图像和HR图像在很大程度上共享相同的信息，所以明确地建模残差图像，这是HR和LR图像之间的差异是有利的。当输入和输出高度相关时，我们提出了一种有效学习的网络结构。此外，我们的初始学习率是SRCNN的104倍高[6]。这通过残差学习和梯度限幅来实现。比例因子我们建议单一模型SR方法。刻度通常是用户指定的，可以是任意的，包括分数。例如，可能需要在图像查看器中平滑放大或调整到特定维度。在准备所有可能的情况下训练和存储许多尺度相关模型是不切实际的。我们发现单个卷积网络对于多尺度因子超分辨率是足够的。贡献总之，在这项工作中，我们提出了一种基于非常深的卷积网络的高度精确的SR方法。如果使用小的学习速率，非常深的网络收敛太慢。提高收敛速率与高学习率导致爆炸梯度，我们解决问题与残差学习和梯度限幅。此外，我们扩展我们的工作以在单个网络中应对多尺度SR问题。与如图1所示的现有技术方法相比，我们的方法相对准确和快速。

2.相关工作

SRCNN是基于深度学习的SR方法的代表性的现有技术方法。因此，让我们分析和比较它与我们提出的方法。

2.1。图像超分辨率卷积网络

模型SRCNN由三个层组成：补丁提取/表示，非线性映射和重建。分别使用空间尺寸为9×9,1×1和5×5的滤波器。在[6]，Dong et al。尝试准备更深的模型，但没有在一周的培训后观察到优秀的表现。在某些情况下，更深的模型性能较差。他们的结论是，更深层的网络不会带来更好的性能（图9）。然而，我们认为，增加深度显着提高了性能。我们成功使用20个重量层（每层3×3）。我们的网络非常深（20比3 [6]），用于重建（接受字段）的信息要大得多（41×41比13×13）。训练对于训练，SRCNN直接建模高分辨率图像。高分辨率图像可以被分解为低频信息（对应于低分辨率图像）和高频信息（残留图像或图像细节）。输入和输出图像共享相同的低频信息。这表明SRCNN用于两个目的：将输入携带到结束层并重建残差。将输入结束到概念上类似于自动编码器。训练时间可能花在学习这个自动编码器上，使得学习其他部分（图像细节）的收敛速度显着降低。相比之下，由于我们的网络直接模拟残差图像，我们可以有更快的收敛，甚至更好的准确性。缩放与大多数现有的SR方法一样，SRCNN是针对单个缩放因子进行训练的，并且应该仅使用指定的缩放。因此，如果需要新的规模，则必须训练新的模型。为了应对多尺度SR（可能包括分数因子），我们需要为每个所关注的尺度构建单个单尺度SR系统。然而，为所有可能的情况准备多个单独的机器以处理多个尺度是无效的和不切实际的。在这项工作中，我们设计和训练单个网络以有效地处理多尺度SR问题。这证明工作非常好。我们的单机与给定子任务的单一规模专家相比较。对于三个比例因子（×2,3,4），我们可以将参数数量减少三倍。除了上述问题，还有一些细微的差别。我们的输出图像具有与输入图像相同的大小，通过在训练期间对每一层填充零点，而来自SRCNN的输出小于输入。最后，我们简单地对所有层使用相同的学习速率，而SRCNN对不同层使用不同的学习速率，以实现稳定收敛。

3.建议方法

3.1。建议网络

对于SR图像重建，我们使用一个非常深的卷积网络，灵感来自Simonyan和Zisserman [19]。配置如图2所示。我们使用d层，除了第一层和最后一层之外的层是相同类型的：64×3×3×64的滤波器，其中滤波器在64个通道上的3×3空间区域上操作（特征图）。第一层对输入图像进行操作。用于图像重建的最后一层由大小为3×3×64的单个滤波器组成。网络采用内插的低分辨率图像（到期望的大小）作为输入并预测图像细节。建模图像细节通常用于超分辨率方法[21,22,35,3]，我们发现基于CNN的方法可以从这个领域特定的知识中获益。在这项工作，我们证明明确建模图像细节（残差）有几个优点。这些将在后面的第4.2节中进一步讨论。使用非常深的网络来预测密集输出的一个问题是每次应用卷积运算时特征映射的大小减小。例如，当大小为（n + 1）×（n + 1）的输入应用于接收场大小为n×n的网络时，输出图像为1×1。这与其他超分辨率方法因为许多需要周围像素来正确地推断中心像素。这种中心 - 环绕关系是有用的，因为周围区域对这种错误问题（SR）提供了更多的约束。对于靠近图像边界的像素，这种关系不能被充分利用，并且许多SR方法裁剪结果图像。然而，如果所需的环绕区域非常大，则该方法是无效的。裁剪后，最终图像太小，无法在视觉上令人满意。为了解决这个问题，我们在卷积前填零，以保持所有特征映射（包括输出图像）的大小相同。事实证明，零填充工作出奇的好。为此，我们的方法不同于大多数其他方法，在图像边界附近的像素也被正确地预测的意义上。一旦图像细节被预测，它们被添加回输入ILR图像以给出最终图像（HR）。我们使用这种结构用于我们的工作中的所有实验。

3.2。训练

我们现在描述最小化的目标，以找到我们的模型的最优参数。令x表示内插的低分辨率图像，y表示高分辨率图像。给定训练数据集{x（i），y（i）} N i = 1，我们的目标是学习预测值y = f（x）的模型f，其中y是目标HR图像的估计。我们最小化在训练集上平均的均方误差1 2 || y - f（x）|| 2最小化。残留学习在SRCNN中，网络必须保留所有输入细节，因为图像被丢弃，输出是从学习的特性单独生成的。对于许多重量层，这成为需要非常长期记忆的端对端关系。为此，消失/爆炸梯度问题[2]可能是至关重要的。我们可以简单地用残差学习来解决这个问题。由于输入和输出图像很大程度上相似，我们定义残差图像r = y - x，其中大多数值可能为零或小。我们要预测这个残差图像。损失函数现在变为1 2 || r - f（x）|| 2，其中f（x）是网络预测。在网络中，这在损耗层中反映如下。我们的损失层有三个输入：残差估计，网络输入（ILR图像）和地面真实HR图像。损失计算为重建图像（网络输入和输出之和）和地面实况之间的欧几里得距离。通过使用基于反向传播的微型批量梯度下降优化回归目标来进行训练（LeCun等人[14]）。我们将momentum参数设置为0.9。训练通过重量衰减（L2罚分乘以0.0001）正规化。非常深的网络的高学习率训练深层模型可能无法在实际的时间限制内收敛。 SRCNN [6]没有显示超过三个重量层的优越的性能。虽然可能有各种原因，一种可能性是他们在网络收敛之前停止了他们的训练过程。他们的学习率10-5太小，网络不能在一个星期内在一个共同的GPU上收敛。看图。 9的[6]，不容易说，他们更深的网络已经收敛，他们的表现饱和。虽然更多的培训将最终解决这个问题，但增加深度到20不似乎实际与SRCNN。这是一个基本的经验法则，使学习率高，提高培训。但是简单地设置学习率高也可以导致消失/爆炸梯度[2]。因为这个原因，我们建议一个可调节的梯度限幅，以最大限度地提高速度，同时抑制爆炸梯度。可调节梯度剪辑梯度剪辑是一种经常用于训练循环神经网络的技术[17]。但是，据我们所知，其使用在培训CNN方面有限。虽然存在许多方法来限制梯度，但是通常的策略之一是将单个梯度限幅到预定范围[-θ，θ]。使用剪辑，渐变在一定范围内。随着通常用于训练的随机梯度下降，学习率被乘以调整步长。如果使用高学习速率，则很可能θ被调整为小以避免在高学习速率方案中的爆炸梯度。但是随着学习速率退火以变小，有效梯度（梯度乘以学习速率）趋近于零，并且如果学习速率在几何学上减小，则训练可以采用指数级数的迭代来收敛。对于最大收敛速度，我们将梯度剪切为[-θγ，θγ]，其中γ表示当前学习速率。我们发现可调梯度限幅使我们的收敛过程非常快。我们的20层网络训练在4小时内完成，而3层SRCNN需要几天的训练。多尺度虽然非常深的模型可以提高性能，但现在需要更多的参数来定义网络。通常，为每个缩放因子创建一个网络。考虑到经常使用小数因子，我们需要一种经济的方式来存储和检索网络。为此，我们还训练一个多尺度模型。使用这种方法，参数在所有预定比例因子之间共享。训练多尺度模型是直接的。几个指定尺度的训练数据集合成一个大数据集。数据准备与SRCNN [5]类似，但有一些区别。输入色块大小现在等于接收场的大小，图像被分成没有重叠的子图像。迷你批次由64个子图像组成，其中来自不同尺度的子图像可以在同一批次中。我们使用MatConvNet1包实现我们的模型[23]。

4.了解属性

在本节中，我们研究我们提出的方法的三个属性。首先，我们显示大深度是SR的任务所必需的。非常深的网络利用图像中的更多上下文信息，并用许多非线性层来模拟复杂函数。我们通过实验验证更深层的网络比浅层网络提供更好的性能。第二，我们显示我们的剩余学习网络收敛比标准CNN快得多。此外，我们的网络提供了显着的性能提升。第三，我们表明，我们的方法与单个网络执行以及使用多个网络训练每个尺度的方法。我们可以有效地降低多网络方法的模型容量（参数数量）。

4.1。更深，更好

卷积神经网络通过在相邻层的神经元之间实施局部连接模式来利用空间 - 局部相关[1]。换句话说，层m中的隐藏单元将层m-1中的单元的子集作为输入。它们形成空间连续的接受场。每个隐藏单元对于接收字段之外的关于输入的变化不响应。因此，该架构确保学习的滤波器对空间局部输入模式产生最强的响应。然而，堆叠许多这样的层导致变得越来越全面（即响应于更大的像素空间区域）的滤波器。换句话说，非常大支持的过滤器可以有效地分解成一系列小过滤器。在这项工作中，我们使用相同大小，3×3，所有层的过滤器。对于第一层，接收场的大小为3×3。对于下一层，接收场的大小在高度和宽度上都增加2。对于深度D网络，接收场具有大小（2D + 1）×（2D + 1）。其大小与深度成正比。在SR的任务中，这对应于可以被利用来推断高频分量的上下文信息的量。大的接受域意味着网络可以使用更多的上下文来预测图像细节。由于SR是一个不良的反问题，收集和分析更多的邻近像素给出更多的线索。例如，如果存在完全包含在接受场中的一些图像图案，则认为该图案被识别并用于超分辨图像是合理的。此外，非常深的网络可以利用高非线性。我们使用19个整流线性单位，我们的网络可以模拟具有中等数量通道（神经元）的非常复杂的函数。制作薄深网络的优点在Simonyan和Zisserman中得到了很好的解释。我们现在实验表明，非常深的网络显着提高了SR性能。我们训练和测试深度范围从5到20的网络（只计算重量层，不包括非线性层）。在图3中，我们显示结果。在大多数情况下，性能随着深度的增加而增加。随着深度的增加，性能迅速提高。

4.2。残留学习

由于我们已经有一个低分辨率的图像作为输入，预测高频分量足以达到SR的目的。虽然在以前的方法中已经使用了预测残差的概念[21,22,26]，但是没有在深度学习的SR框架的背景下进行研究。在这项工作中，我们提出了一种学习残留图像的网络结构。我们现在详细研究这种修改对标准CNN结构的影响。首先，我们发现这个剩余网络收敛得更快。通过实验比较两个网络：剩余网络和标准非剩余网络。我们使用深度10（权重层）和比例因子2.各种学习速率的性能曲线在图4中示出。所有使用上面已经提到的相同的学习速率调度机制。第二，在收敛时，剩余网络显示出优越的性能。在图4中，当训练完成时，剩余网络给出较高的PSNR。另一点是，如果使用小的学习速率，网络不会收敛在给定的时期数。如果使用初始学习速率0.1，剩余学习网络的PSNR在10个时期内达到36.90。但是如果使用0.001代替，网络从未达到相同的性能水平（其性能在80个时期后为36.52）。以类似的方式，残余和非残留网络在10个时期之后显示出显着的性能差距（对于速率0.1，36.90对27.42）。简而言之，这种对标准非残留网络结构的简单修改非常强大，可以在输入和输出图像高度相关的其他图像恢复问题中探索该想法的有效性。

4.3。单个模型的多个尺度

训练期间的缩放是一种关键技术，为网络配备多分辨率的超分辨率机器

胸部鳞。我们的多尺度机器可以执行不同规模的许多SR过程，其容量远小于单尺度机器组合的容量。我们从一个有趣的实验开始如下：我们用一个比例因子应变训练我们的网络，并在另一个比例因子下测试。这里，考虑广泛用于SR比较的因子2,3和4。对数据集“Set5”[15]尝试可能的对（应变，stest）。实验结果总结在表2中。如果应变6 = stest，性能降低。对于比例因子2，用因子2训练的模型给出PSNR为37.10（以dB为单位），而用因子3和4训练的模型分别给出30.05和28.13。在单尺度数据上训练的网络不能处理其他尺度。在许多测试中，它甚至比双三次内插，用于生成输入图像的方法更差。我们现在测试用尺度增加训练的模型是否能够在多个比例因子下执行SR。上面使用的相同网络用多个比例因子strain = {2,3,4}训练。此外，我们实验案例应变= {2,3}，{2,4}，{3,4}更多的比较。我们观察到网络处理在训练期间使用的任何规模。当应变= {2,3,4}（表2中的2,3,4）时，其每个尺度的PSNR与单个尺度网的相应结果相当

工作：37.06对37.10（×2），33.27对32.89（×3），30.95对30.86（×4）。另一种模式是对于大规模（×3,4），我们的多尺度网络优于单尺度网络：我们的模型（×2,3），（×3,4）和（×2,3,4）给出PSNR 33.22 ，33.24和33.27，而（×3）给出了32.89。类似地，（×2,4），（×3,4）和（×2,3,4）分别给出30.86,30.94和30.95（相对于30.84×4模型）。由此，我们观察到训练多个尺度提高了大尺度的性能。

5.实验结果

在本节中，我们评估我们的方法对几个数据集的性能。我们首先描述用于训练和测试我们的方法的数据集。接下来，给出训练所需的参数。在概述了我们的实验设置之后，我们将我们的方法与几个最先进的SISR方法进行比较。

5.1。培训和测试的数据集

训练数据集不同的基于学习的方法使用不同的训练图像。例如，RFL [18]有两种方法，其中第一种使用91张来自Yang等人的图像。 [25]，第二个使用291图像与添加200图像从伯克利分割数据集[16]。 SRCNN [6]使用非常大的ImageNet数据集。我们使用291图像作为本节其他方法的基准与[18]。此外，使用数据增加（旋转或旋转）。对于上一节中的结果，我们使用91个图像来快速训练网络，因此性能可能略有不同。测试数据集对于基准测试，我们使用四个数据集。数据集'Set5'[15]和'Set14'[26]经常用于其他作品的基准[22,21,5]。数据集“Urban100”，最近由Huang等人提供的城市图像的数据集。 [11]，是非常有趣，因为它包含许多有挑战性的图像失败了许多现有的方法。最后，数据集“B100”，在Timofte等人使用的Berkeley分割数据集中的自然图像。 [22]和Yang和Yang [24]作为基准，也被采用。

5.2。培训参数

我们提供用于训练我们的最终模型的参数。我们使用深度网络20.训练使用大小为64的批次。动量和重量衰减参数分别设置为0.9和0.0001。对于重量初始化，我们使用He et al。 [10]。这是使用整流线性单元（ReLu）的网络的理论上可靠的过程。我们训练所有实验超过80个时代（9960次迭代，批次大小64）。学习率最初设置为0.1，然后每20个时期减少10的因子。总的来说，学习率降低了3倍，并且学习在80个时期之后停止。训练大约4小时在GPU Titan Z.

5.3。基准

对于基准，我们遵循Huang等人公开可用的框架。 [21]。它能够将许多最先进的结果与相同的评估程序进行比较。框架对于图像的颜色分量应用双三次插值，对于亮度分量使用复杂的模型，如在其他方法[4]，[9]，[26]中。这是因为人类视觉对强度的细节比对颜色更敏感。此框架在图像边界附近合成像素。对于我们的方法，这个过程是不必要的，因为我们的网络输出全尺寸的图像。然而，为了公平比较，我们也裁剪像素相同的量。

5.4。与最先进的方法的比较

我们提供定量和定性比较。比较方法是A + [22]，RFL [18]，SelfEx [11]和SRCNN [5]。在表3中，我们提供了对几个数据集的定量评估的总结。我们的方法胜过这些数据集中的所有以前的方法。此外，我们的方法相对较快。基于CPU实现的SRCNN的公共代码比Dong等使用的代码慢。 al [6]在他们的论文中基于GPU实现。在图6和图7中，我们比较我们的方法和topperforming方法。在图6中，只有我们的方法完美地重建中间的线。类似地，在图7中，轮廓在我们的方法中是干净和生动的，而在其他方法中它们严重模糊或扭曲。

6.结论

在这项工作，我们提出了一个超分辨率方法使用非常深的网络。由于收敛速度慢，训练非常深的网络是困难的。我们使用剩余学习和极高的学习率来优化非常深的网络快速。收敛速度最大化，我们使用梯度限幅，以确保训练的稳定性。我们已经证明，我们的方法在基准图像上优于现有方法。我们相信我们的方法很容易应用于其他图像恢复问题，如去噪和压缩伪像删除。