Abstract

近年来，深度学习以其强大的表达能力在图像超分辨率重建领域为人们所关注。然而，尽管这些CNN-based方法能够在很大程度上提升超分效果，但是，针对遥感图像，我们要求能产生更真实的纹理细节，这是他们不能满足的。Moreover, most previous researchers 只是针对一个尺度的超分任务训练一个特定尺度的模型，这显然不能满足我们现实的需求。在这篇文章中，针对遥感图像，我们提出了a novel approach，主要对以前的方法进行了三个方面的改进：1）以progressive reconstruction的方式生成多个尺度的超分结果；2）使用残差密集网络，并加入细节关注机制来充分利用LR中的特征信息；3）使用对抗学习来生成LR中损失的信息，进一步提升视觉质量。实验结果show that我们的模型在数量和质量上都有很好的效果。

1. Introduction

总结

单张图像的超分辨率重建aims at 从低质量的低分辨率图像生成其对应的视觉效果理想的高分辨率图像。图像的超分辨率重建作为一种切实可行的有效的图像处理方法，被广泛应用于遥感、安防、医学等领域，这篇文章主要针对的就是遥感图像。实际上，一张低分图像可以由无数张高分图像通过降采样得到，也就是说图像超分本质上是一个ill-posed问题and no unique solution。In order to tackle this problem, 大量的SR算法被提出，主要包括interpolation based，sparse coding based and deep learning based.

相关工作总结及其缺点

Recently, 深度卷积网络以其强大的学习能力在图像超分辨率领域取得了impressive 成果。Dong et al. 首先提出了将CNN引入图像超分，证明了CNN-based方式相比传统方法有很大优势。Kim et al. 在前者的基础上，通过增加网络深度、加入递归监督机制，提出了VDSR和DRCN，效果相比前者有了相当大的提升。这也充分证明了通过增加网络的深度能够有效提高图像超分的performance，基于此Lim et al.提出了更深的利用残差学习策略的网络模型EDSR，其不再以粗略上采样后的LR图像作为输入，在很大程度上降低了计算量和复杂度，并以残差学习的方式减轻了由于网络深度带来的训练困难。随着网络深度的增加，底层的特征由于没有直接与subsequent layer相连，会逐渐被遗忘。Zhang et al.提出了一种残差密集网络RDN，使用残差密集块来充分利用所有卷积层得到的分级特征。Li et al.提出了MSRN，引入多个size的卷积核来检测图像特征，并通过多个尺度卷积核得到的特征之间的相互联系来获得最优的图片信息，另外进行全局特征融合以达到充分利用底层特征的目的。尽管他们achieve great improvements on performance，但最后得到的超分图像都趋于平滑，我们认为这是由于这些方法在处理图像特征时完全没有侧重，对高低频信息的处理缺乏灵活性。Furthermore, 这些方法只能针对一个scale factor训练特定的模型，这显然限制了图像超分在real-world scenarios的应用。

为了解决这些缺点我们做了哪些改进

To address these drawbacks, 我们提出了一种新颖的基于双关注机制的多尺度对抗网络结构。首先，我们通过双注意力机制从特征细节与特征通道两个方面来适应性地学习更有用的特征，让网络根据feature detail和feature channel两个不同维度的语义依赖性来改善超分结果。Besides, 我们设计的网络结构是以progressive reconstruction的方式逐步生成多个尺度的超分图像，都是由previous level进行x2的上采样得到。由于LR是通过HR降采样得到，这一过程中必然会损失很多细节信息，为了生成与原图更接近的纹理细节，我们employ adversarial loss 来调整用MSE训练得到较粗略的超分图像，得到更sharper edges、更真实的纹理以及视觉效果更理想的结果。

Contributions

The main contributions of our work can be summarized in three-fold:

1. 提出了一种基于双关注机制的残差密集网络结构，不仅从内容细节也从特征通道两个维度来捕捉更有用的特征信息。
2. 利用尺度之间的相关性，以层层递进的方式对中间层特征逐步上采样得到多个尺度的超分图像。
3. 通过对抗学习调整超分的粗略结果，生成LR中损失的细节信息，让最终得到的超分结果能有更让人信服的纹理细节且与真实图像更相像。

接下来的文章结构概述

本文的组织结构安排如下：我们将会在sec2详细介绍网络结构以及相关模块的实现细节，训练细节说明以及对比实验会安排在sec3，sec4是这篇文章的总结。

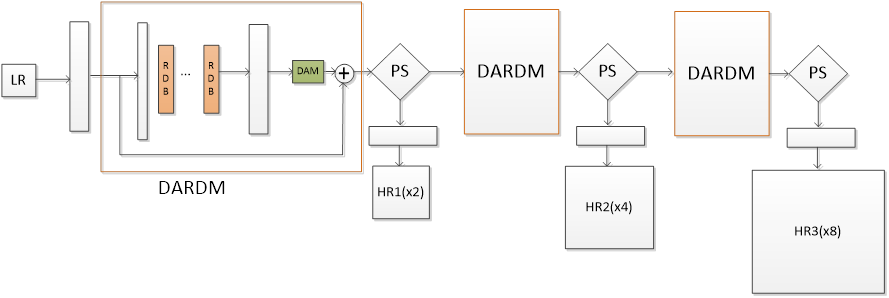
1. Methods

Given LR-HR training pairs，这里表示为（xi, yi），我们的目标是希望通过提出的算法方程G（参数为θ）得到一个ypi能够与ground truth yi尽可能地相似。为了得到最优的ypi，我们的任务是通过训练不断地优化算法参数，It can be formulated as



其中L表示损失函数，N指的是图片的数量。接下来，我们将从网络结构，残差密集块，双关注机制以及adversarial loss四个方面对算法模型Gθ进行详细阐述。

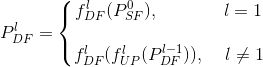
* 1. Network Architecture



As show in Fig2，我们的DAMSR主要由两个部分组成：浅层特征提取部分，多尺度重建部分。其中多尺度重建部分有三个级别分别对应三个不同尺度，每个级别包含密集特征提取模块、上采样模块以及图像重建模块。这里，我们用I(SR)表示输入，I(HR)表示其对应的gt，Ii(SR)表示我们的算法输出结果。首先对于浅层特征提取部分，我们可以表示

D:\firefoxdown\CodeCogsEqn(6).gif

其中fSF表示卷积核为3x3的卷积操作。P0会作为多尺度重建部分的输入，我们将多尺度重建部分每个级别中密集特征提取模块的输出表示为



其中l表示尺度的level（l=1, 2, 3），fDF(l)表示在第l级别的操作，这里主要包括一个用于改变浅层特征通道数的1x1卷积，B个残差密集块，用于收集该级别所有残差密集块特征以进行一个密集特征融合的1x1卷积和双关注模块。fUP(l)表示在第l级别的上采样操作，这里我们引用了Shi et al.的思想，通过sub-pixel convolution操作得到高分辨率特征图。

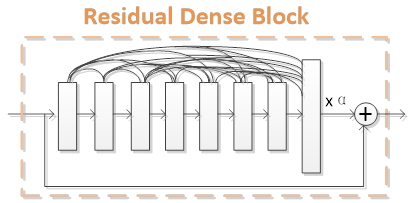
然后是每个级别的上采样模块，可以表示为

D:\firefoxdown\CodeCogsEqn(3).gif

PUP(l)表示每个级别的上采样后的特征，将作为图像重建模块的输入，即

https://latex.codecogs.com/gif.latex?I_%7BSR%7D%5E%7Bl%7D%20%3D%20f_%7BRec%7D%5E%7Bl%7D%28P_%7BUP%7D%5E%7Bl%7D%29

这里fRec(l)表示卷积核为3x3的卷积操作，ISR(l)即为最终每个级别的输出。

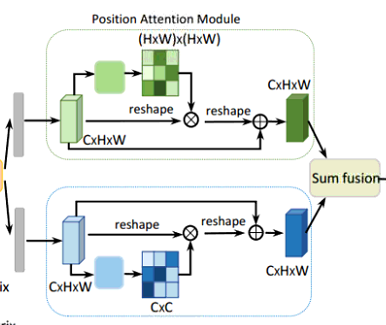


* 1. Residual Dense Block

残差密集块的结构如Fig3，该结构主要包括密集连接，residual scaling，以及residual fusion三个部分。

密集连接包含C个3x3卷积，能够让底层特征直接参与到上层特征构造，使得底层特征经过多次卷积后不被遗忘，从而实现持续的记忆。Residual scaling 是通过将学习到的残差特征与一个0到1之间的固定因子相乘来保持训练的稳定性。Residual fusion是通过将输入与输出做pixel-wise sum来进一步提高网络的表达能力，获得更丰富的语义信息。

* 1. Dual Attention Mechanism



以往基于CNN的超分方法大都在提取特征信息时完全没有侧重，虽然也有方法提出基于Channel上的关注机制（zhang et al.），但是他们却忽略了空间细节信息的重要性。基于此，我们提出了分别从特征细节与特征通道两个维度来适应性地学习更有用信息的双关注机制。

如Fig4所示，双关注机制主要包括两个平行的注意模块，一个是Detail Attention Module，一个是Channel Attention Module。对于DA模块，我们通过自注意力机制来获取特征图中任意两个位置间的空间依赖性。即通过一系列操作来学习由两个位置间的特征相似性决定的特征权重，并以此来更新特征的空间细节信息。具体我们可以表示为



其中。。。

对于CA模块，我们用相似的自注意力机制来捕获两个特征通道之间的通道依赖性，以此来更新特征的通道信息。具体我们可以表示为



其中。。。

最后，这两个平行关注模块学习到的有侧重性的特征以pixel-wise sum的方式进行融合，以达到进一步增强特征的目的。

* 1. Adversarial Learning

尽管通过双关注机制让我们的超分效果有很大的提升，但是其虽然能够让网络更关注图像的高频细节信息从而有更理想的视觉效果，但是对于HR降采样到LR中损失的信息，我们很难从LR中学习到。受到SRGAN的启发，我们希望通过adversarial learning的方式，通过生成器和判别器的博弈来生成LR中没有的而且与真实图像尽可能相似的特征。

考虑到计算成本以及Gan结构收敛的困难性，我们只采用adversarial loss来对粗略结果进行精细化调整。Adversarial loss的公式表达为

https://latex.codecogs.com/gif.latex?L_%7Badv%7D%20%3D%20%5Csum_%7Bi%3D1%7D%5E%7BN%7D-logD%28G%28I_%7Bi%7D%5E%7BLR%7D%29%29

这里G表示生成器，D表示判别器。对于判别器，我们采用的是VGG-style network来逐步缩小特征的空间维度，由于篇幅限制，这里不做细节描述。

1. Experiments
   1. Experimental Setting

**Loss function** 考虑到直接使用Gan结构训练很难收敛，我们两步走的策略，首先用most widely used MSE loss进行初步训练 as defined below：

https://latex.codecogs.com/gif.latex?L_%7Bmse%7D%20%3D%20%5Cfrac%7B1%7D%7BN%7D%5Csum_%7Bi%3D1%7D%5E%7BN%7D%5Cleft%20%5C%7C%20I_%7Bi%7D%5E%7BSR%7D%20-I_%7Bi%7D%5E%7BHR%7D%5Cright%20%5C%7C_%7B2%7D%5E%7B2%7D

接下来再加入adversarial loss对上一步的初步结果进行调整：

https://latex.codecogs.com/gif.latex?L%20%3D%20L_%7Bmse%7D%20&plus;%5Cgamma%20L_%7Badv%7D

**Datasets** 训练数据集我们采用的是从Deepglobe中选取的2500张尺寸为1024x1024的遥感图像，validation images and test images 分别为100张和20张。另外测试集还包括从西北工业大学的NWPU-RESISC45遥感数据集中选取的100张图像。

**Parameter setting**参考【】，我们随机选取16个尺寸为32x32的LR patches作为输入，使用的优化器为Adam optimizer by setting β1=0.9，β2=0.999，and learning rate设置为0.0001，迭代次数为200个epochs。We implement the DAMSR with Tensorflow framework and train them using two NVIDIA 1080Ti GPUs with no special initialization。

* 1. Ablation Experiments

Ablation investigation主要研究双关注机制以及adversarial learning对图像超分在评价指标表现以及视觉质量的影响。整体结构去掉DA以及AL部分，并且设置RDB的数量B为6，每个RDB中卷积层的数量C为6，密集连接的growth rate为32，作为我们的baseline。

|  |  |  |  |
| --- | --- | --- | --- |
|  | X2 | X4 | X8 |
| Baseline |  |  |  |
| DA |  |  |  |
| AL |  |  |  |

接下来逐步加入DA和adversarial learning，通过观察评价指标以及图像视觉质量的变化来评估这两个组成部分的影响力。PSNR与SSIM在DeepGlobe测试集上的对比结果如Table1，It can be seen that DA模块对每个尺度的提升效果都很明显，特别是X2的情况提升了近。。个点，我们猜测是因为相对来说其包含的特征信息量更多，DA模块能够学习到的特征权重也就更加准确更具有贡献性。而AL对评价指标的提升效果就不那么理想了，我们认为是由于这个模块的主要侧重于生成LR图像中损失的纹理细节以从视觉上提升图像质量。

Fig5

视觉质量的对比如Fig5所示，这里我们只展示X4倍数，可以观察到baseline的图像过于平滑，加入DA后有所改善，而AD则让整个图像拥有了更清晰的纹理和shaper edge看起来更加真实。

* 1. Comparison Experiments

在这部分实验中，我们主要与Bicubic，VDSR，DRCN，SRGAN，EDSR，MSRN，RDN，RCAN这几种主流的超分方法进行了quantitative和qualitative两个方面的对比。

**Quantitative** Table2 shows the multi-scale evaluation results of the average peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) on Geepglobe测试集。从表中可以看出，the proposed method outperforms all the previous work on x2 and x4两个尺度by a considerable magin，除了 x8的情况与RCAN差不多。我们猜测这是因为在大倍数的情况下，双关注机制对特征的增强效果已经没有那么显著。

|  |  |  |  |
| --- | --- | --- | --- |
|  | X2 | X4 | X8 |
| Bicubic |  |  |  |
| VDSR |  |  |  |
| DRCN |  |  |  |
| SRGAN |  |  |  |
| EDSR |  |  |  |
| MSRN |  |  |  |
| RDN |  |  |  |
| RCAN |  |  |  |
| Proposed |  |  |  |

Fig 6

**Qualitative** Fig6是图像在x4倍数下的质量对比结果，从图中可以观察到most of previous methods suffer from不同程度的由于图像平滑带来的边界和纹理模糊。即便是SRGAN也存在因为特征利用不充分而产生的noticeable artifacts。相反，用我们的方法重建的图像则拥有更清晰的纹理and shaper edges。同样，在x2和x8的情况下我们的超分质量也领先于其他的方法，由于篇幅原因，这里我们仅show visual comparisons on scale x4。

1. Conclusion

In this paper, we proposed a dual attention multi-scale framework for remote sensing image super resolution. The dual attention mechanism is proposed to adaptively learning more useful features by considering interdependency among feature positions and feature channels. Furthermore, we introduce adversarial learning to guiding the model to recover more detailed textures lost in LR. Extensive experimental evaluations and empirical analysis well demonstrate the proposed model achieves superiority over the state-of-the-arts.

1. Reference