

轻量化超分辨率卷积神经网络研究

**Research on Lightweight
Super-resolution Convolutional Neural
Networks**

学科专业： 信息与通信工程

作者姓名： 杨炳旺

指导教师： 杨爱萍

天津大学电气自动化与信息工程学院

二零一九年十一月

独创性声明


本人声明所呈交的学位论文是本人在导师指导下进行的研究工作和取得的研究成果，除了文中特别加以标注和致谢之处外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得 **天津大学** 或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名： 签字日期：2019年11月28日

学位论文版权使用授权书

本学位论文作者完全了解 **天津大学** 有关保留、使用学位论文的规定。特授权 **天津大学** 可以将学位论文的全部或部分内容编入有关数据库进行检索，并采用影印、缩印或扫描等复制手段保存、汇编以供查阅和借阅。同意学校向国家有关部门或机构送交论文的复印件和磁盘。

(保密的学位论文在解密后适用本授权说明)

学位论文作者签名：

导师签名：

签字日期：2019年11月29日

签字日期：2019年11月29日

摘要

超分辨率技术能够在硬件设备性能受限情况下，利用算法提高图像分辨率，恢复图像细节，获取高质量的图像。基于卷积神经网络的深度学习方法能有效提取图像内部特征，学习低分辨率图像与高分辨率图像之间的映射关系，较好地实现超分辨率重建。本文基于卷积神经网络，针对现阶段超分辨网络的效率问题进行了一系列研究，主要工作如下：

针对目前基于深度学习的超分辨网络模型较深，参数量、运算量较大，无法适应于实际场景等问题。本文在卷积神经网络基础上，深入研究轻量化卷积神经网络关键技术，设计出一种轻量、快速、有效的超分辨率网络。该网络首先采用分组卷积有效降低模型参数，并引入 1×1 卷积层来解决信息流通问题；同时通过密集连接加强特征复用，提高网络效率；为进一步提高超分辨率重建质量，引入注意力机制提升网络的特征选择能力。实验结果表明，所提方法不仅能有效地恢复图像细节，同时具有较快的运行速度和较少的模型参数。

目前超分辨率网络对图像的多尺度信息利用不足，且往往采用线性方式对不同感受野的信息进行融合，无法实现感受野尺寸自适应改变。为充分提取图像不同尺度特征，论文提出一种基于动态卷积核选择的多尺度残差超分辨率网络。首先，采用动态卷积核选择卷积层，可自适应利用不同尺度的特征图，实现对图像多尺度信息的提取；在此基础上设计残差嵌套结构，可充分利用局部特征信息，加强反向传播时信息流通；另外，引入权重归一化层对每一层网络的权重进行归一化处理以加快网络收敛速度，并提升超分性能。实验结果表明，与现有主流方法相比，所提算法在客观和主观两方面均可取得良好的超分辨率重建效果，同时仅包含少量参数。

关键词： 超分辨率，卷积神经网络，分组卷积，注意力机制，多尺度

ABSTRACT

In the case of limitations of hardware devices, super-resolution algorithm is used to improve the resolution of the image, recover the details of image and reconstruct high-quality image. Deep learning method based on convolutional neural network can effectively extract the internal features of the image, learn the mapping relationship between the low-resolution image and the high-resolution image to reconstruct high-resolution image. Based on the convolutional neural network, this paper focuses on the efficiency of the super-resolution network. The main works are summarized as follows:

At present, the super-resolution network model is relatively deep, and demands large amount of parameters and computational resources, so it is hard to be applied to resolve the real-world problems. Based on the convolutional neural network, this paper concentrates on technologies of lightweight convolutional neural networks, designs a lightweight, fast and effective super-resolution network. In this network, the group convolution is firstly used to reduce model parameters effectively, and 1×1 convolution layer is employed to create a linear combination of the output of the group convolution layer. Meanwhile, feature reuse is encouraged through dense connection to improve network efficiency. In order to further improve the quality of reconstruction image, the attention mechanism is developed to boost the feature discriminative ability of the network. Experimental results show that the proposed method not only recovers image details effectively, but also has faster running speed and fewer parameters.

The multi-scale information of image is insufficiently utilized by super-resolution network, and the information of different receptive fields is linearly aggregated, so the size of receptive fields can not be adaptively changed. In order to fully extract the different scale features of images, a multi-scale residual super-resolution network based on dynamic selection mechanism of convolutional kernel is proposed. Firstly, the selective kernel convolution is presented to extract multi-scale information from feature maps of different scales. On this basis, the residual nested structure is designed to make full use of the local feature information and strengthen the information flow

during back propagation. In addition, the weight normalization layer is introduced to normalize the weight of each layer to accelerate network convergence and improve the performance of the network. Experimental results show the proposed method achieves better results on both objective and subjective evaluation indicators compared with the state-of-the-art super-resolution methods, and only contains a few parameters.

KEY WORDS: Image super resolution, Convolution neural network, Group convolution, Attention mechanism, Multi-scale

目 录

摘 要	I
ABSTRACT	III
第 1 章 绪论	1
1.1 课题研究背景与意义	1
1.2 图像超分辨率研究现状	2
1.2.1 基于插值的方法	2
1.2.2 基于重建的方法	3
1.2.3 基于学习的方法	4
1.3 论文主要工作和章节安排	7
1.3.1 论文主要工作	7
1.3.2 论文章节安排	8
第 2 章 基于卷积神经网络的图像超分辨率	9
2.1 卷积神经网络基础理论	9
2.1.1 神经元模型	10
2.1.2 卷积层与池化层	10
2.1.3 激活函数	12
2.1.4 反向传播与优化方法	14
2.2 基于卷积神经网络的图像超分辨率	15
2.2.1 SRCNN 与纯卷积结构	15
2.2.2 反卷积层与亚像素卷积层	16
2.2.3 残差学习与递归结构	17
2.2.4 批归一化层移除和金字塔结构	18
2.3 图像超分辨算法质量评价	20
2.3.1 主观评价	20
2.3.2 客观评价	20
2.4 损失函数	21
2.5 超分辨率研究常用数据集	22
2.6 本章小结	24
第 3 章 基于分组卷积和注意力机制的超分辨率网络	25
3.1 轻量化网络设计	26
3.2 基于分组卷积和注意力机制的超分辨率网络设计	27
3.2.1 网络整体结构	27

3.2.2 级联的记忆分组卷积网络.....	28
3.3 实验与结果分析	32
3.3.1 实验设置	32
3.3.2 实验结果	33
3.3.3 消融实验	37
3.4 本章小结	40
第4章 基于动态卷积核选择的多尺度超分辨率网络.....	41
4.1 多尺度网络设计	42
4.2 基于动态卷积核选择的多尺度超分辨率网络设计.....	43
4.2.1 网络整体结构	43
4.2.2 多重级联的多尺度残差网络.....	44
4.2.3 权重归一化层	47
4.3 实验与结果分析	47
4.3.1 实验设置	47
4.3.2 实验结果	48
4.3.3 消融实验	53
4.4 本章小结	54
第5章 总结与展望	55
5.1 全文总结	55
5.2 研究展望	56
参考文献	57
发表论文和参加科研情况说明	63
致 谢	65

第 1 章 绪论

1.1 课题研究背景与意义

二十一世纪是信息技术发展迅猛的时代,随着网络技术和终端设备的不断发展,智能终端设备已经走进了千家万户,许多数字成像设备也成为人们日常生活的一部分。现如今,除了从事摄像的专业人士,拍摄图像不再需要沉重的单反相机或者其他专业摄像设备,移动智能终端如智能手机、智能平板已经成为人们拍摄图像的重要途径。图像作为重要的信息传递载体,是一种最直观,最生动的表达信息的方式。数字图像的清晰程度与分辨率大小有密切的联系。在数字图像处理领域,图像所包含的信息量在相当程度上取决于其分辨率的大小。高分辨率图像包含清晰的纹理和细节,以及更加丰富的语义信息,不仅能够展现更好的视觉效果,还可以提取图像附加的细节信息用于更高层次的计算机视觉问题^[1],例如图像分类问题^[2]、人脸识别问题^[3]以及语义分割^[4]、目标检测问题^[5]等。计算机视觉领域的热点问题都是以高质量清晰图像为基础,算法的精度在一定程度上取决于所用数据集图像的质量。

获取高分辨率图像最直接方法即是采用高精度的硬件设备,如高清摄像机。通常情况下,硬件设备的性能决定着图像信号的传递和接收,优良的高清设备对图像的信号的衰减较小,呈现出的图像画面更为清晰,但在实际应用中这种方案在很多场景或大规模部署时非常困难^[6],一方面原因在于高精度传感器和光学器件通常需要复杂的制造工艺和昂贵的成本,另一方面,在某些特殊场景中,如航空遥感、医疗图像等,硬件设备的制造和部署都存在难以攻克的难题。由于硬件设备的诸多限制,关于以软件方式来获取高分辨率图像成为热门的研究方向,图像超分辨率重建技术就是利用低质量、低分辨率图像(或图像序列)来产生高质量、高分辨率图像的技术。图像超分辨率技术成本较低,不需要高精度的硬件设备就能得到较高清晰的图像,因此具备较高的实用价值。

由于图像超分辨率技术的高效低成本优势,该技术有着极其广泛的应用,近些年来,其在图像传输^[7]、视频监控^[8]、航空遥感^[9]、医疗图像^[10]、工业生产^[11]领域都发挥着重要作用。在视频监控方面,对监控图像进行超分辨率重建从而获取高分辨率图像,其能有效提高图像视频的清晰度,从而能更准确地识别目标对象或感兴趣的人,维护社会安全稳定;在航空遥感领域,超分辨率技术突破成像

系统的固有限制，提高图像分辨率，有益于实现多光谱遥感图像中对区域更好的分类；在医疗图像中，超分辨率技术能有效改善核磁共振图像的质量，提高图像分辨率，辅助医生做出准确的诊断；在工业生产中，超声波检测图像较为模糊，常采用超分辨率技术获取清晰的检测图像；提高检测准确率。在其他方面例如图像修复、手机拍照变焦、旧电影重构等，超分辨率技术也有着重要的应用，在未来，图像超分辨率技术能带来更好视觉体验并服务于高精度的视觉任务，如自动驾驶，智慧交通等方面。因此，图像超分辨率技术的研究具有重要的研究意义和现实价值。

1.2 图像超分辨率研究现状

图像超分辨率重建是一个病态的逆问题，低分辨率图像和高分辨率图像之间存在多种映射关系，因此通过一张低分辨率图像求解高分辨率图像并没有唯一解。目前，根据对高分辨率图不同求解方式，图像超分辨率算法研究一般可分为三个种类，即基于插值的方法、基于重建的方法、基于学习的方法。

1.2.1 基于插值的方法

基于插值的方法是三种方法中最易实现，算法复杂度最小的一种算法，其利用周围相邻像素点间的关联性和连续性来求解未知像素点的灰度值，通过这种方法来增加图像单位面积内的像素点的数目，以得到高分辨率图像。经典的插值算法包括最近邻插值算法^[12]、双线性插值算法^[13]、双三次插值算法^[14]等。最近邻插值，又称零阶插值，其变换后图像的每一个插入像素点的灰度值等于原图像中与它最邻近的像素点的灰度值，是一种最简单的灰度值插值方法。其优点是计算量较小，算法实现简单，缺点是插值后图像质量不够理想，容易在边缘处出现锯齿效应和块效应；双线性插值则是基于水平和垂直方向上的信息进行线性插值，即对周围四个像素点的灰度值进行内插。其优点是能产生平滑的效果，避免出现锯齿效应和块效应，缺点是图像的部分高频信息会产生丢失，致使图像的边缘出现模型，影响视觉体验；双三次插值，又称为双立方插值，其通过增加插值邻域的大小来提升插值效果，插值邻域扩充为 4×4 ，即利用目标像素点周围的16个像素点的灰度值进行内插。其优点是能产生比双线性插值更加平滑的效果，同样能有效去除锯齿效应和块效应，能得到比前两种插值方法更好的插值效果，缺点是其计算复杂度比前两种方法要大。

由以上阐述可知，基于插值的超分辨率方法有着较好的时间复杂度，但因为算法较为简单，因此效果略显不足。

1.2.2 基于重建的方法

基于重建的方法，通过数学方法对形成低分辨率图像的过程进行建模，利用已知的图像先验知识由低分辨率图像估计高分辨率图像，对图像缺失的高频信息进行重建。常见的方法有：迭代反投影法^[15]，凸集投影法^[16]和最大后验概率法^[17]等。迭代反投影法通过反复迭代反投影过程以最小化模拟的和原始的低分辨率图像之间的差异，当差异小于设定的阈值时则停止迭代，此时能得到较好的高分辨率图像。其优点是算法易于理解，但因为超分辨率问题是一个病态的逆问题，因此问题的解是不确定、不唯一的，对反投影算子的选取较为困难，此类方法也没有充分利用图像的先验知识。超分辨率问题估计高分辨率解的空间受到一组约束（闭凸集）的限制，进而产生相较于原解空间更小的解集合空间，这些约束表征了期望的属性，例如对数据的保真度、平滑度、清晰度等，与最终求得的解的属性保持一致。凸集投影法将凸集投影用于超分辨率重建，凸集投影法在给定高分辨图像空间中的一个点的情况下，将问题简化为迭代寻找与所有给定的凸约束相交的最近似的解，并最终获得高分辨率图像的一个估计。其优点是简单易行，能够很好的利用先验信息来保持高分辨率图像的边缘信息和细节信息，缺点是计算成本比较高，收敛速度比较慢，收敛的效果在很大程度上依赖初始值的选取，并且最终得到的解不唯一。最大后验概率法基于贝叶斯理论，将高分辨率图像的估计看作一个统计估计问题。这种包含了以先验知识为约束的方法为解决病态的超分辨率逆问题提供了一种理论框架。经典的超分辨率模型可以写成：

$$b = AX + e \quad (1-1)$$

b 代表观察到的低分辨率图像； A 代表一些造成图像降质的线性处理； X 为高分辨率图像； e 代表噪声。这些因素都被假定为随机的，对高分辨率图像的最大后验概率估计就是最大化后验概率 $P\{X|b\}$ ，假定 \hat{X}_{MAP} 表示 X 的最大后验方法，则

$$\hat{X}_{MAP} = \arg \max_X P\{X|b\} = \arg \max_X [\log P\{b|X\} + \log P\{X\}] \quad (1-2)$$

在式 1-2 中， $\log P\{b|X\}$ 表示对数似然函数； $\log P\{X\}$ 表示对数先验模型。高分辨率图像的先验模型可由先验知识得到，似然函数可由噪声的统计量确定。其优点是存在唯一解，在保证合理有效的先验信息的情况下，能保持良好的边缘，缺点是需要较大的计算量并且有较慢的收敛速度。

由以上阐述可知，基于重建的超分辨率方法能够较好地保持图像的边缘信息，但当超分方法尺度较大的时候，先验知识较难有效恢复图像的高频信息，使得超分后图像的纹理区域过于平滑。此外，基于重建的方法的重建效果和收敛速度十分依赖于超分辨率图像的初值的估计，选择合适的初值是实现超分辨率重建的重要一环。

1.2.3 基于学习的方法

基于学习的方法在近几年成为研究的热点。相较于传统方法，基于学习的方法能取得较好的超分效果，其将低分辨率图和与之对应的高分辨率图进行一一配对，并以此作为训练集，从中推测出测试的样本与训练图像集间的邻域关系，并通过构造的最优权值约束来求得图像的先验知识以重建与测试样本的高分辨率图像接近的高分辨率图像。基于学习的方法能得到更多的关于高分辨率图像的高层信息以弥补低分辨率图像自身对高分辨率图像的高频信息缺失，可得到较好的超分辨率重建效果。

常见的浅层学习方法包括基于样例学习的方法^[18]，基于邻域嵌入的方法^[19]，基于稀疏表示的方法^[20]等。基于样例学习的方法先将图像切块，例如 5×5 、 7×7 块，之后通过对构建的高低分辨率的图像数据集进行学习，获取与测试的低分辨率图像块最接近的低高分辨率图像对，从中提取高频信息补充到低分辨图像中，从而完成对低分辨率图像的超分辨率重建^[21]。其优点是能较好的重建低分辨率图像的高频信息，实现较高质量的超分辨率重建；其缺点是需要高度切合的训练样本，算法的好坏在一定程度上取决于训练样本的选择，另外其也对训练集噪声比较敏感；邻域嵌入法基于流形学习，首先假设高分辨率图像块和与之对应的低分辨率图像块在局部包含相近的几何流形。先通过图像数据集上的训练，求得高分辨率图像块和低分辨率图像块的几何流形结构，然后为测试的低分辨图像找出其在低分辨率图像块训练集中的 K 近邻表示，并推导出 K 近邻表示系数，并对高分辨率图像块训练集中对应的 K 个近邻表示利用这些系数进行加权组合求得高分辨率图像块。其优点是能充分训练集，降低训练集数目，但重建的效果会出现过拟合或者欠拟合现象。基于稀疏表示的方法，通过构建一个过完备的字典，该字典由对高分辨率图像块进行大量随机抽样得到，接着对每个输入的低分辨图像块求出其在该字典中的稀疏表示，并通过对稀疏系数的加权组合来获得高分辨率图像。其优点是避免了邻域嵌入方法中邻域大小的选择问题，缺点则是只适用于某些领域的超分辨率，对于通用图像的超分辨率效果不够理想，原因在于过完备字典的选择有很大的局限性。

近年来基于深度学习的方法成为研究的热点，卷积神经网络能够有效地提取图像特征，并进行特征变换和非线性映射，从大量数据中学习到低分辨率图像与高分辨率图像之间繁杂的映射关系^[22]。目前方法主要通过加深网络模型、利用残差学习、加强特征复用等技术来改善超分辨率效果。Dong 等^[23]率先提出了基于卷积神经网络的图像超分辨率算法 SRCNN(Image Super-Resolution Using Deep Convolutional Networks)，首次将卷积神经网络(CNN)应用到超分辨率重建领域，

其网络结构设计非常简洁，由三个卷积层构成。网络首先将低分辨率通过三次插值放大，然后学习插值之后的低分辨率图与高分辨率图之间的映射关系，实现了一个端到端的学习框架实现超分辨率重建，但是其包含的卷积层个数过少，不能充分地提取图像的特征；为提升网络的特征提取能力，Dong 等提出 SRCNN^[23]的改进版 FSRCNN^[24]，增加了卷积层数，同时把网络的输入由插值后图像替换成原始低分辨率图像，在网络的最后引入反卷积层来进行上采样以重建高分辨率图像，该算法不仅能获得更好的超分辨率效果还加快了网络运行的速度，但是反卷积层可能会给图像带来棋盘效应；为克服反卷积层带来的棋盘效应，同时加快网络上采样速度，ESPCN^[25]引入了亚像素卷积层来取代反卷积操作对图像上采样，其采用低分辨率图像作为网络输入，亚像素卷积层较反卷积层拥有更少的运算量，因而能大幅提高了网络速度。

上述三种网络均只包含了较少的卷积层，不能较好地学习低分辨率图像与高分辨图像之间的映射关系；为提升网络的表达能力，VDSR^[26]借鉴了 VGG16 的结构，其移除了 VGG16 的池化层，采用包含 20 个卷积层的神经网络来实现图像超分辨率，同时为克服深层网络训练困难的问题，引入了残差学习思想，网络学习高分辨率图像与低分辨率图像之间的残差，加快了网络收敛速度，提升了学习效率，但是也带来了更多的网络参数量；为减少网络的参数量，同时实现更深的网络结构，DRCN^[27]将递归结构引入超分辨率网络中，递归结构能在不引入更多网络参数的情况下，堆叠更多的卷积层来增加网络的感受野，为避免梯度消失问题，网络采用循环监督和跳跃连接保证网络收敛，网络效果较 SRCNN^[23]有较大提升；DRRN(Image Super-Resolution via Deep Recursive Residual Network)^[28]受 ResNet^[29]、DRCN^[27]的启发，采用递归结构，与 DRCN 不同的是，递归块由包含两个卷积层的残差结构构成以加强信息的流通，同时采用全局残差学习，利用更深的网络层数获取性能的提升。

VDSR、DRCN、DRRN 均采用上采样后的图像作为网络输入，这样会增加额外的计算开销。为减少网络的计算开销，并克服当放大倍数较大时，网络采用一次上采样操作难以恢复高质量图片的缺陷，LapSRN^[30]参考了图像金字塔结构，设计出一种拉普拉斯金字塔结构，构造了一个逐级放大的超分辨率网络，同时采用了更加鲁棒的 Charbonnier 损失函数对每一级的结果进行监督，实现了单网络多尺度超分辨率，获得了较好地超分辨率效果，但是网络还是采用了传统的卷积层与批归一化层的组合；批归一化层对特征进行了正则化，限制了网络的灵活性，同时带来大量的内存占用，为克服批归一化层的缺陷，EDSR(Enhanced Deep Residual Networks for Single Image Super-Resolution)^[31]移除了批归一化层，从而可以构建更深的模型来提升超分辨率网络的效果，并取得较高的峰值信噪比指标，

但是其采用简单的残差结构,没有能充分利用网络中各个卷积层的特征;为充分利用网络中提取的各个层次的特征,RDN(Residual Dense Network for Image Super-Resolution)^[32]同时借鉴了 ResNet 和 DenseNet^[33]的思想,构建了残差密集块,残差密集块结合局部残差学习和局部密集连接的思想以充分提取局部特征,在整体网络结构上采用了超过 100 个卷积层的网络得到较好的超分效果;DBPN(Deep Back-Projection Networks for Super-Resolution)^[34]针对超分辨率网络对高分辨率图像和低分辨率图像之间的相互依赖关系的挖掘不够充分的问题,基于传统的迭代反投影^[15]思想,构建了一种迭代上下采样的端到端的可学习的超分网络,为每个迭代子模块设置误差反馈机制,在迭代过程总进行误差矫正,在放大倍数较大时取得了更好的效果,但是因为需要在高分辨率特征图上进行卷积,使得网络运算量较大;为实现轻量化设计,CARN(Fast, Accurate, and Lightweight Super-Resolution with Cascading Residual Network)^[35]采用级联的结构,通过将卷积核尺寸为 1×1 卷积层和残差结构进行级联,实现一个轻量的超分辨率网络结构。上述基于卷积神经网络的图像超分辨率方法均采用 $L1$ 损失函数或者均方误差损失函数,这种损失函数虽然能带来较高的峰值信噪比,但会使得超分辨率重建图像比较平滑,缺乏细节信息,为获得更好的主观效果,恢复更多的细节信息,SRGAN^[36]借鉴了生成对抗网络(Generative Adversarial Network)的思想,利用生成网络和判别网络的相互对抗,采用由感知损失函数和对抗损失函数组合的复合损失函数学习出视觉效果更好的高分辨率图像。

可见,目前的大部分超分辨率网络虽然能取得相对较好的效果,但受模型参数多,计算量大等问题的影响,在计算能力有限,内存容量不足的情况下,基于深度学习方法的超分辨率重建很难应用到实际场景中。针对上述问题,目前出现了多种轻量化方法,前文所述的 DRCN^[27]和 DRRN^[28]方法采用递归结构,能有效减少网络参数,但这两种方法采用上采样后的图像作为网络输入,并且使用递归构造了较深的网络,这种方式增加了网络的运行速度和计算量;而 IDN(Fast and Accurate Single Image Super-Resolution via Information Distillation Network)^[37]采用信息蒸馏结构来实现快速的图像超分辨率,但其没有充分考虑到各个特征通道之间的依赖性。此外,现有的图像超分辨率网络不能充分利用图像的多尺度信息,Shi^[38]基于 Inception 结构构建了一个多尺度超分辨率网络;Li^[39]提出了一种多尺度残差结构(Multi-scale Residual Network for Image Super-Resolution),通过多路分支的融合来提取多尺度信息。但是,上述多尺度方法都是通过线性方式融合多尺度信息,没有考虑感受野的动态调整的问题。本文针对上述问题,首先利用分组卷积和注意力机制设计一种轻量、快速的超分辨率网络,实验结果表明其能有效实现精度与模型参数、运行时间之间的平衡;此外,论文还提出一种动态卷积

核选择的超分辨率网络，其能动态地调整感受野的大小，充分利用图像的多尺度信息，实验表明该网络能在保持较少参数数量的同时，有效提升超分辨率效果。

1.3 论文主要工作和章节安排

1.3.1 论文主要工作

本文研究基于卷积神经网络的图像超分辨率算法，主要针对现有的图像超分辨率网络的参数量大，运算时间较长的问题，研究如何在不明显降低精度的情况下，减少网络的参数量以提升超分网络的效率。首先，本文提出了基于分组卷积和注意力机制的图像超分辨率网络，利用分组卷积有效降低网络的参数量，并采用注意力机制在引入极少参数数量的情况下，有效地提升网络的性能，实现轻量的图像超分辨率网络。此外，为提升超分辨率网络的精度，实现更好的参数量与网络性能之间的均衡，针对超分辨率网络对于图像多尺度信息关注不足的缺陷，本文还提出另一种轻量的超分辨率网络结构，其采用动态卷积核选择卷积层，通过在卷积神经网络中对卷积核的动态选择以实现自适应地调整其感受野的大小，以此来实现包含多尺度信息的图像超分辨率网络。具体工作如下：

针对基于卷积神经网络的超分辨率的参数量较大，计算复杂度高的问题。本文提出了一种轻量、快速、有效的超分辨率网络结构。首先，通过引入分组卷积层，实现对模型参数数量的有效降低；然而，分组卷积的弊端在于各个分组之间的通道信息不流通，因此，在分组卷积层后添加卷积核大小为 1×1 卷积以融合分组卷积层的通道信息；另一方面，为加强特征复用，避免网络出现梯度消失的现象，本文引入密集连接结构，并采用以分组卷积层后的卷积核大小为 1×1 卷积层为密集连接结构主要连接对象的方式来有效降低原始密集连接的参数量，同时保证了网络的性能没有明显下降。另外，为进一步提高超分辨率网络重建效果，在引入少量参数的情况下增加注意力机制，有效提升了网络的性能。经实验证明，所提算法不仅能大幅提高网络的运行效率，还有较好地恢复图像细节。

为提升超分辨率网络的重建效果，在网络参数与超分辨率重建质量之间实现更好的均衡，本文设计了另一种有效的图像超分辨率网络。针对目前超分辨率网络无法完全利用图像多尺度信息的问题，提出一种基于动态卷积核选择超分辨率网络以充分获取网络中不同尺度的信息。其主要通过动态卷积核选择卷积层来实现感受野的自适应选择，并以此为基础构建残差嵌套结构实现对局部特征的充分利用；在整体结构上则采用多重级联机制实现特征特征复用并加强特征信息的流

通；此外，网络还引入权值归一化层以稳定网络的训练过程，加快网络的收敛速度，提升超分辨率网络的重建效果。

1.3.2 论文章节安排

本文主要对轻量化超分辨率卷积神经网络进行研究，共分五个章节，具体的章节安排如下：

第1章为绪论，主要介绍了课题的研究背景和意义，并详细地综述了本领域的研究现状，并给出本篇文章的主要工作以及结构安排。

第2章介绍了卷积神经网络相关知识，在此基础上研究了基于卷积神经网络的超分辨率发展过程中所提出的关键技术，最后介绍了图像超分辨率研究常用评价指标、损失函数以及数据集等。

第3章提出了基于分组卷积和注意力机制的轻量级超分辨率网络。针对超分辨率网络参数量大，计算复杂度大的问题，设计了一种轻量、快速的超分辨率网络，采用分组卷积有效降低参数，利用改进的密集连接和注意力机制提高网络的性能，通过实验证明该网络能取得较好的超分辨率效果，同时有着较快的运行速度。

第4章提出了基于动态卷积核选择的多尺度轻量级超分辨率网络。为提升超分辨率网络的效率与性能，实现更好的超分辨率重建效果与网络参数量的均衡，针对超分辨率网络对多尺度信息利用不足的缺陷，提出动态卷积核选择的多尺度网络，其能实现动态选择感受野提取图像的多尺度特征，同时结合了权重归一化层以提高超分辨率网络的重建效果。实验证明该网络能有效恢复图像细节，同时保持较少参数。

第5章是总结和展望，总结的全文主要工作，并对本领域未来研究重点和方向进行展望。

第 2 章 基于卷积神经网络的图像超分辨率

目前,卷积神经网络已经在图像超分辨率领域取得了较好的效果,其能从大量数据中学习高低分辨率图像之间的映射关系,进而从低分辨率图中恢复出高分辨率图像。本章首先介绍卷积神经网络基础理论,包括卷积层、池化层的等原理及实现过程,反向传播、优化器、激活函数等相关知识;然后,在此基础上介绍目前主流的图像超分辨率神经网络,总结并阐述超分辨率网络发展中的关键技术;最后,给出图像超分辨率任务采用的评价指标、损失函数以及网络训练中常用的数据集。

2.1 卷积神经网络基础理论

人工神经网络的研究在上世纪早期就已经出现,其目的是利用抽象的数学模型或者计算模型来模拟或近似生物神经网络,通常可定义为由大量神经元相互联结形成的具有自适应性的非线性网络系统^[40]。神经网络的发展在近年来十分迅速,目前已经众多领域取得应用。卷积神经网络作为神经网络的重要发展方向,在计算机视觉领域已经实现了良好的性能。

在上世纪 60 年代,Hubei 和 Wiese^[41]等的生物学家通过对猫的视觉系统的研究提出了感受野的概念;之后,第一个基于感受野的认知机模型由 Fukushima 提出,对卷积神经网络的研究就此开启。

现代的标准卷积神经网络结构主要由输入层,卷积层,池化层,全连接层以及输出层构成。卷积层和池化层的作用是提取图像的多种特征以完成特定的任务。一个典型的用于图像分类任务的卷积神经网络结构如图 2-1 所示:

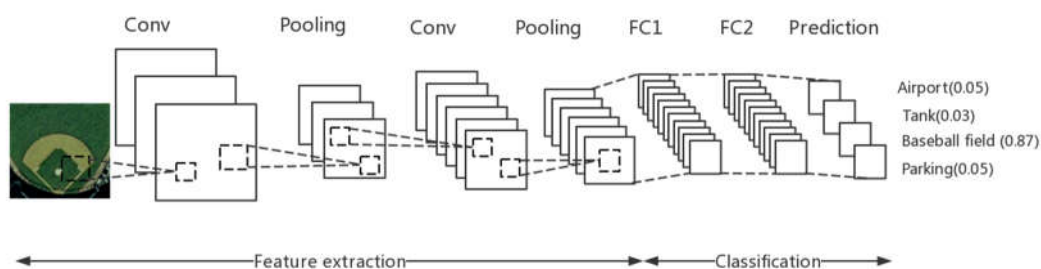


图 2-1 卷积神经网络结构

与用于其他计算机视觉任务的卷积神经网络不同，用于超分辨率的卷积神经网络主体架构中通常不包含池化层和全连接层。本节将详细介绍有关卷积神经网络的相关知识，如神经元模型、卷积层、池化层以及激活函数等。

2.1.1 神经元模型

第一个关于神经元的抽象数学模型由 McCulloch 和 W.Pitts^[42]提出，被称为 MP 神经元模型，如图 2-2 所示。其中 x_i 表示神经元同时接受的第 i 个输入信号； w_k 表示该神经元与输入信号之间的连接的权重；神经元的偏置项由 b_k 表示，则神经元的输出 y_k 可表示为：

$$y_k = f(b_k + \sum_i^m (x_i \times w_{ki})) \quad (2-1)$$

式 2-2 中， $f(\cdot)$ 表示激活函数(Activation function)。

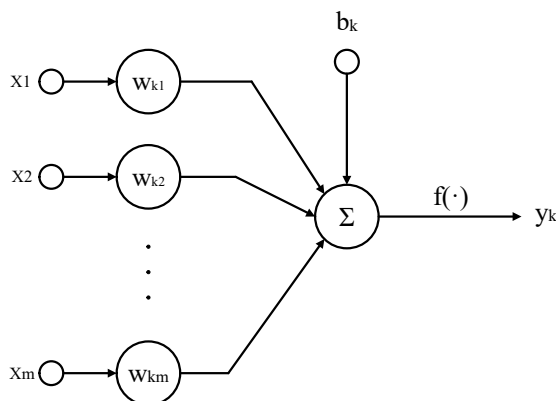


图 2-2 神经元模型

2.1.2 卷积层与池化层

卷积神经网络中的卷积层对输入特征图进行卷积操作提取特征，卷积操作如图 2-3 所示，其使用固定尺寸的卷积核在输入特征图上进行滑动，在滑动过程中，卷积核参数与输入特征图对应位置元素相乘后对所有值求和得到输出特征图对应位置的值。

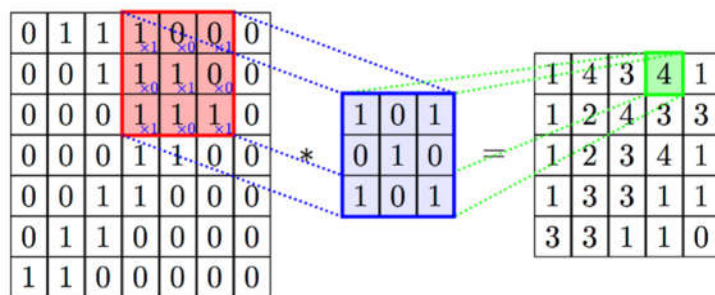


图 2-3 卷积过程

图 2-3 给出了一张 7×7 的特征图与一个 3×3 卷积核进行卷积的过程，卷积核移动的步长为 1，通过在输入图像中取出与卷积核大小相同的块，再与卷积核进行对应位置元素相乘后求和得到输出特征图上对应点的值，使用这个块在输入图像上进行滑动得到最终的卷积结果，经过卷积后的特征图大小为 5×5 。卷积层可以使用一个或多个卷积核提取特征，其输入也有可能是一张或多张特征图，经过卷积层得到的特征图数量与卷积核数量保持一致。卷积层包含两个重要的机制，一是局部连接机制^[43]，二是权重共享机制。卷积层的各个节点只与上一层特征图的局部区域进行连接，以学习图像的局部特征，局部连接的过程如图 2-4 所示。

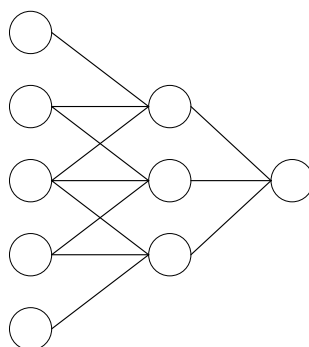


图 2-4 局部连接

通常将连接的局部区域称为“局部感受野”，对于二维图像，局部区域像素的空间关联性较强。随着卷积层的增多，网络越来越深，对应其在浅层网络特征图上的感受野也越来越大。更深层次卷积层通过融合浅层局部信息，可获取更全局的特征。卷积层的权重共享是指利用局部窗口在输入图像进行滑动时，与之作卷积操作的卷积核采用同一组权重矩阵和偏置参数，使得在提取图像特征时不需要考虑图像特征的位置。由于局部连接和权重共享机制，有效地减少了卷积神经网络的计算量，降低了网络规模，推动了卷积神经网络的发展。

池化层又称为下采样层，池化层的每个神经元对局部区域进行池化操作，其过程是通过选取局部区域中某个特征统计量来替换局部区域的单个点的属性^[44]。常用的两种池化方法有最大池化和均值池化^[45]，如图 2-5 所示。

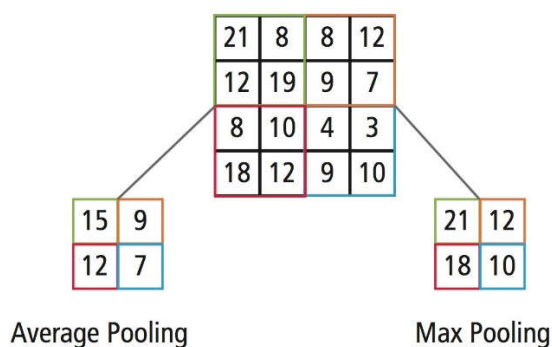


图 2-5 池化操作

其中，均值池化(Average pooling)采用局部区域(如上图中 2×2)中所有元素的均值作为池化结果，而最大池化(Max pooling)则采用区域中所有元素的最大值作为池化结果。与卷积层一样，池化层通过滑动局部区域得到最终的池化结果。图 2-5 给出了利用 2×2 的窗口对一张 4×4 的特征图进行池化的过程，最终得到大小为 2×2 的特征图。池化层作用是能实现对特征的选择和对信息的过滤，提取区域内的主要特征。通过池化操作，保持特征的尺度不变性^[46]，使得网络更加关注特征的相对位置而不是精确位置；另外，池化操作能有效降低特征图的维度，减小输入到下一层的特征图尺寸，从而减少了网络的参数个数和计算量，提高网络的运行效率，同时也在一定程度上控制了网络训练中的过拟合。

2.1.3 激活函数

激活函数^[43]的主要作用是神经元的输出引入非线性因素，由神经网络万能近似定理^[47]可知，一个前馈神经网络如果含有非线性隐藏层，则在由足够多隐藏单元情况下，能够以任意精度逼近任何从一个有限维空间到另一个有限维空间的 Borel 可测函数。目前流行的激活函数主要包括 Sigmoid 激活函数，Tanh 激活函数，ReLU 激活函数，LeakyReLU 激活函数等。

Sigmoid 的函数曲线如图 2-6 所示，其数学表达式为：

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (2-2)$$

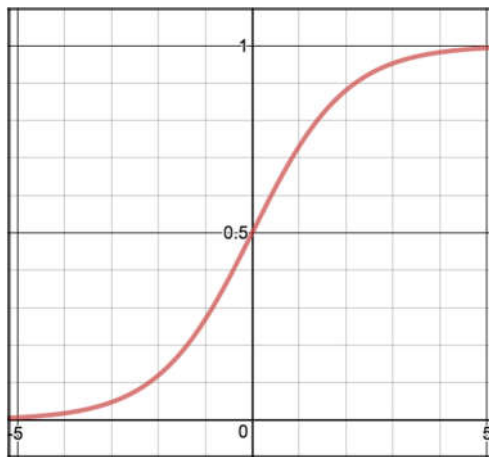


图 2-6 Sigmoid 激活函数

Sigmoid 函数以任意值作为输入并输出 0 到 1 之间的另一个值，当输入值很小趋向负无穷时，其输出值无限接近于 0，当输入值很大趋向于正无穷时，其输出值则无限接近于 1。当输入值在 Sigmoid 函数的左右两端时，其梯度接近于 0，此时会造成网络的较多神经元接近饱和，使得参数更新缓慢，影响网络的训练结果，最终收敛于局部最优。

Tanh 的函数曲线如图 2-7 所示，其数学表达式为：

$$\text{Tanh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2-3)$$

Tanh 函数是 Sigmoid 函数的修改或扩展版本。Sigmoid 函数的输出值在 0 和 1 之间；但是，在 Tanh 的情况下，值的范围在 -1 和 1 之间。Tanh 能实现比 Sigmoid 函数更快的收敛速度，但同时也存在与 Sigmoid 函数一样的问题，其在 -2 到 +2 范围外的梯度变化非常小。

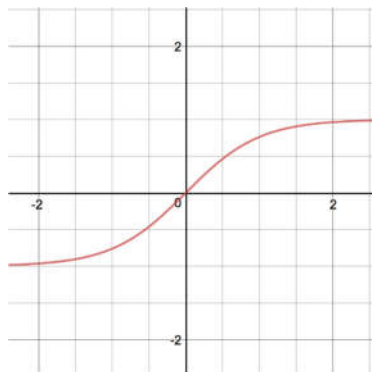


图 2-7 Tanh 激活函数

ReLU 的函数曲线如图 2-8 所示，其数学表达式为：

$$\text{ReLU}(x) = \max(0, x) \quad (2-4)$$

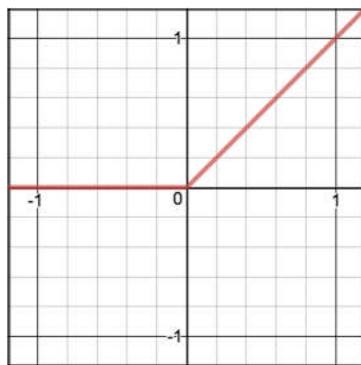


图 2-8 ReLU 激活函数

ReLU 激活函数称为线性整流单元，是深度学习中最常用的激活函数。当函数的输入为负数值的时候，输出为 0，但对于正数值，其和线性函数一样输出输入的值。研究表明，在训练时采用 ReLU 激活函数收敛速度快于 Sigmoid 函数和 Tanh 函数，并能有效地避免梯度消失问题^[2]。而针对 ReLU 函数在负数区域会造成神经元死亡问题，另一种激活函数 LeakyReLU 通过给定一个很小的，恒定的梯度值来解决。LeakyReLU 的函数曲线如图 2-9 所示，其数学表达式为：

$$\text{LeakyReLU}(x) = \max(\alpha x, x) \quad (2-5)$$

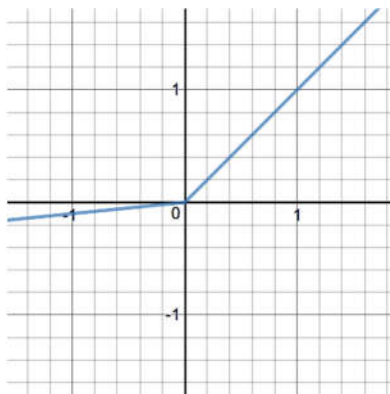


图 2-9 LeakyReLU 激活函数

2.1.4 反向传播与优化方法

卷积神经网络在训练时主要有两个过程，分别是前向传播过程和反向传播过程^[48]。前向传播过程是模型推断过程，是指数据输入后，逐层经过卷积运算和一些其他运算得到最终输出的过程；而反向传播过程则是对模型参数进行更新的过程，基于损失函数计算网络输出与真实值之间的误差，并根据链式求导法则，对每一层进行求导，以此来更新每一层的参数，使最终误差降至最低，即模型收敛，以此时近似最优的参数来对网络进行推导，得到较为理想的结果。反向传播理论真正解决了神经网络的训练问题，使得神经网络的研究进入新时代。

针对反向传播中的参数更新策略，传统的思路是采用梯度下降算法^[49]，通过链式求导法则对每一层的参数进行求导后得到参数的梯度。梯度下降法通过固定一个更新步长，通常称之为学习率 η ，参数在每次反向传播的过程都沿着梯度的方向更新相同的步长，直到网络的损失函数不再下降即模型收敛时则停止更新。更新过程如下：

$$\theta_{t+1} = \theta_t - \eta \Delta \theta_t \quad (2-6)$$

θ_t 表示当前的参数值， θ_{t+1} 表示更新后的参数值， $\Delta \theta_t$ 表示梯度量。由于传统的梯度下降法更新参数时需要计算所有样本的梯度，在数据集样本数量较大时训练速度比较缓慢，且容易收敛于局部最优。随机梯度下降优化(SGD)算法^[50]通过随机挑选一个样本计算梯度来更新全部权值，这种方法能有效提高网路的训练速度，但也会造成网络收敛过程中目标函数波动比较大的问题。因此，其对学习率的选择要求比较高，学习率偏大时会导致网络在最优点附近徘徊，不能收敛到最小值，而学习率偏小时则会使网络收敛速度变慢，且容易收敛到局部极小值点。

为避免训练过程陷入局部最优，并加快网络的收敛速度，基于动量的梯度下降算法^[51]不依赖当前的梯度来更新权重，而是使用动量代表当前梯度，即当前与过去梯度的移动平均值，算法在梯度下降较快的区域通过动量的累计加快梯度下

降，从而也一定程度上避免了陷入局部极小值区域的概率。而在全局最优点时，之前累计的动量会与当前梯度成反方向，从而减小更新的步长，避免了在最优点徘徊。另外一种基于动量的梯度下降法为 NAG^[52]，通过预测梯度下降的方向来加快网络收敛速度。除了基于动量的改进，还有基于自适应学习率优化方法。Duchi^[53]提出了将学习率除以历史梯度平方的累积和的平方根，以动态调整训练过程中学习率的 AdaGrad 优化方法，这种方法能很好地实现自适应学习率调整，但是由于历史梯度的累加，会导致学习率越来越小。为克服学习率不断减小的问题，RMSProp^[43]方法采用历史梯度的指数衰减平均，而并非 AdaGrad 那样采用累积的平方梯度和，从而能有效舍去遥远的梯度。现在最常见的优化算法 Adam^[54]，是动量方法与 RMSProp 的组合，综合了动量法和自适应法的优势，在实践中取得良好的效果。

2.2 基于卷积神经网络的图像超分辨率

近年来，基于卷积神经网络的图像超分辨率算法实现了较大的发展，除了依赖于卷积神经网络技术的整体进步，例如在残差结构、密集结构等方面。超分辨率自身的技术创新如亚像素卷积层、移除批归一化层等技术也起到了至关重要的作用。为了更深刻地理解超分辨率网络的设计思想，本节将介绍超分辨率网络发展中的相关关键技术。

2.2.1 SRCNN 与纯卷积结构

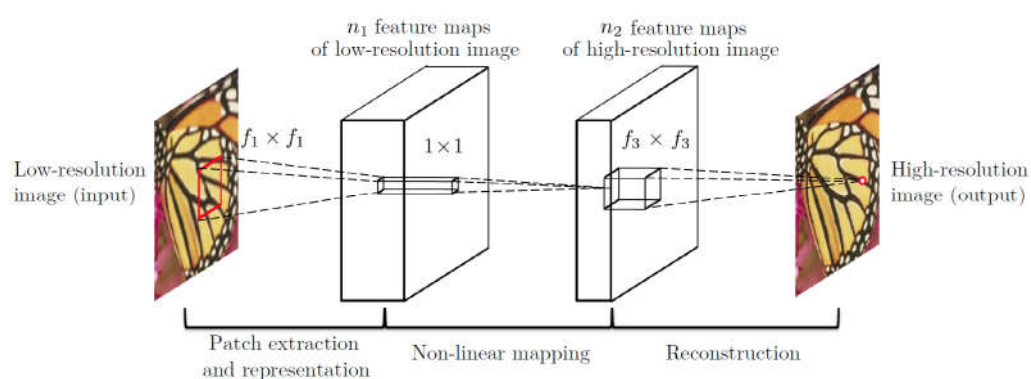


图 2-10 SRCNN 结构

SRCNN(Image Super-Resolution Using Deep Convolutional Networks)[23]首次将卷积神经网络引入单幅图像超分辨率领域，奠定了纯卷积结构是深度学习超分辨率主要结构的基础，之后几乎所有用于图像超分辨率的网络都采用纯卷积结构。纯卷积结构，指的是整个网络结构只包含卷积层，而不包含其他层。SRCNN 的

网络结构如图 2-10 所示，其首次将超分辨率网络分成三个步骤，分别是图像的特征提取与特征表示、特征的非线性映射以及重建过程。这三个步骤也被之后的很多单幅图像超分辨率的网络所沿用，成为网络设计的关键技术。

2.2.2 反卷积层与亚像素卷积层

SRCNN 通过将低分辨率图像利用双三次线性插值放大到目标尺寸后再输入到网络进行训练，在网络运行中不会改变特征图的尺寸。但这样的方法存在两个弊端。首先，这种方法会带来大量的运算量。卷积操作过程中，运算量与特征图的大小成正比关系，如果在一个放大两倍的特征图上进行卷积运算，其运算量是在原始尺寸特征图上进行卷积运算的四倍。因此，输入放大的图像会降低网络运行的速度；其次，通过双三次线性插值会给图像带来无法消除的人工痕迹，导致网络输出图像不够理想。FSRCNN^[24]通过在网络最后引入反卷积层^[55]放大图像，可加快网络运行速度，也奠定了超分辨率网络可先在低分辨率图上做卷积，在网络后期再进行放大的基础。

反卷积是卷积过程的逆操作，也称为转置卷积，或者小数步长卷积。卷积层的反向传播过程就是对卷积层输入梯度的反卷积过程。如果将卷积运算视为输入与卷积矩阵 C 相乘，则对误差的反向传播过程则相当于将输入梯度乘以卷积矩阵的转置，这也是反卷积被称为转置卷积的由来。反卷积过程与卷积操作完全相反，在前向运算时输入与卷积核矩阵的转置相乘，而反向传播时相乘的是卷积矩阵。反卷积有时被称为小数步长卷积，对于步长 $s > 1$ 的直接卷积，则其对应的反卷积的步长就小于 1。图 2-11 给出了输入特征图尺寸为 5×5 ，卷积核大小为 3×3 ，步长为 $s = 2$ 的卷积所对应的反卷积操作，通过在两个元素之间插入 $s - 1$ 个 0，这使得卷积核移动速度比步长为 1 时要慢，因此被称为小数步长。另外，通过在输入特征图元素之间插入 0 像素，反卷积层则可用来进行上采样，反卷积层中的权重是可以学习的，不需要事先设定好插值方法。

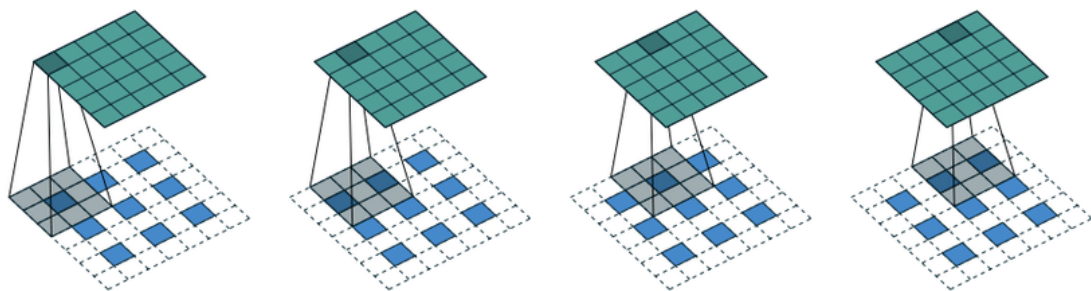


图 2-11 反卷积过程

通过在网络最后采用反卷积层进行上采样,使得网络的特征提取与非线性映射过程都在低分辨率特征图上进行,极大地减少了参数量。

反卷积层需要在输入特征图的元素之间填充零像素再进行卷积操作,这会造成较大的计算成本。为了实现更好的实时性,ESPCN^[25]提出了一种新的上采样算法,通常称为亚像素卷积层,具体操作由图 2-12 所示。

假设放大倍率是 r ,亚像素卷积操作的过程是,通过卷积层得到通道数是 r^2 ,尺寸与低分辨率特征图相同的一系列特征图,然后将特征图每个像素的 r^2 个通道重新排列成一个 $r \times r$ 的单通道特征图,其能够将 $H \times W \times r^2$ 的特征图重新

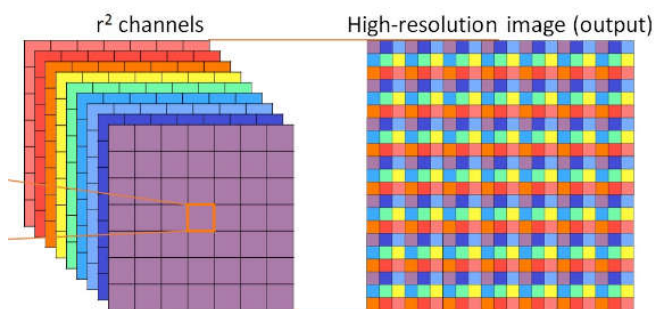


图 2-12 亚像素卷积过程

排列成 $rH \times rW \times 1$ 的单通道图像以实现图像上采样。亚像素卷积只是简单地进行像素重排,此操作没有参数,不用进行额外的计算,因此能有效提高运行速度。

2.2.3 残差学习与递归结构

根据 VDSR^[26]中对高低分辨率图像之间关系的研究,网络输入的低分辨率图像与网络输出高分辨率图像在很大程度上具有相似性。具体来说,指的是低分辨率与高分辨率图像之间的差异主要在于高频信息的不同,其低频信息很相近。因此,在网络学习中,可只学习高分辨率图像与低分辨率图像之间的高频残差,可通过在网络中增加一个长程连接,使网络输出与输入的低分辨率图像相加使得网络中主要对残差进行学习。另外,为保持所有特征图与最终的输出图像的尺寸一致,VDSR 在每次卷积前都会对图像进行补零操作,这种操作也能加强对边缘像素的预测。区别于残差网络 ResNet(Deep residual learning for image recognition)^[29]结构,通常将 VDSR 采用残差学习称为全局残差学习。而残差结构^[29]主要用于图像分类等任务,其目的是为了解决网络层数加深时带来的梯度消失和梯度爆炸^{[56][57]}进而导致网络训练不收敛等问题,残差结构也可称为局部残差学习,其通过一个短程连接,也可称为跳跃连接机制跳过一层或多层而进行连接。利用卷积神经网络对残差映射进行拟合,假设网络输入为 x ,则其可表达为:

$$H(x)=F(x)+x \quad (2-7)$$

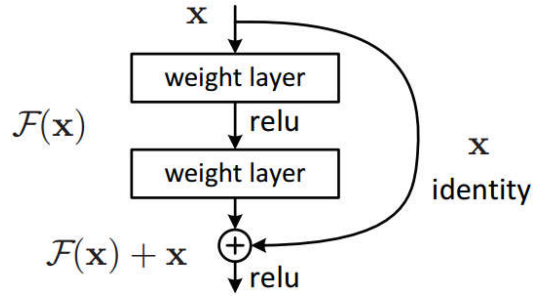


图 2-13 残差网络结构

通过残差结构能有效解决反向传播时梯度在回传过程中出现消失的问题，加强了网络前后层之间的联系。残差网络结构如图 2-13 所示，该结构被用于很多超分辨率网络。

递归神经网络在自然语言处理研究中发挥了非常重要的作用，DRCN^[27]首次将递归结构引入到超分辨率领域，通过不停的递归调用相同的模块，使得特征图循环多次通过该模块。通过递归结构，可有效增加网络的感受野，同时保持较少的参数。另外，由于递归结构加深了网络层数，可能会造成网络训练上的困难。DRRN(Image Super-Resolution via Deep Recursive Residual Network)^[28]基于递归残差网络结构，通过局部残差学习与全局残差学习相结合有效地解决了训练困难问题，大大提升了网络性能。VDSR、DRCN、DRRN 网络结构如图 2-14 所示。

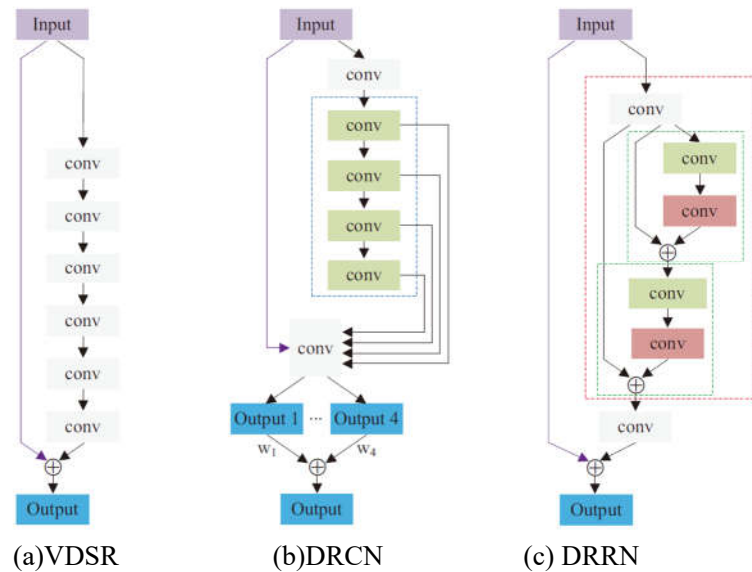


图 2-14 残差学习结构与递归结构

2.2.4 批归一化层移除和金字塔结构

基于卷积神经网络的图像超分辨率网络的研究取得较大进展的一个重要因素在于 EDSR^[31]网络对批归一化层的移除，这对之后的超分辨率网络起到了指导作用。其原因在于批归一化主要用来解决高层计算机视觉问题，如图像分类或者目标检测等，作用在于稳定网络训练，加快网络收敛，解决中间层数据分布发生

变化等情况，而直接将其应用于超分辨这样低层的计算机视觉问题显然是不合适的。通过移除批归一化层，在相同的计算资源下能设计出更深的网络结构，取得更好的性能。

多数超分辨率网络在网络最后采用单一的上采样模块来实现对低分辨率图像进行放大，但这种操作在放大倍率较大时产生的超分辨率图像质量较差。针对此问题，LapSRN^[30]参考了图像金字塔结构提出了拉普拉斯金字塔网络结构，通过逐层上采样，分级预测残差的方式实现超分放大。通过逐层上采样机制，较好地优化了当放大倍率较大时，在网络最后采用单一的大倍率上采样层而造成的超分辨率重建效果不理想问题；同时，该网络能在训练大倍率超分网络时得到中间倍率的结果，可省去为特定倍率放大训练特定模型的麻烦。LapSRN 网络结构如图 2-15 所示。

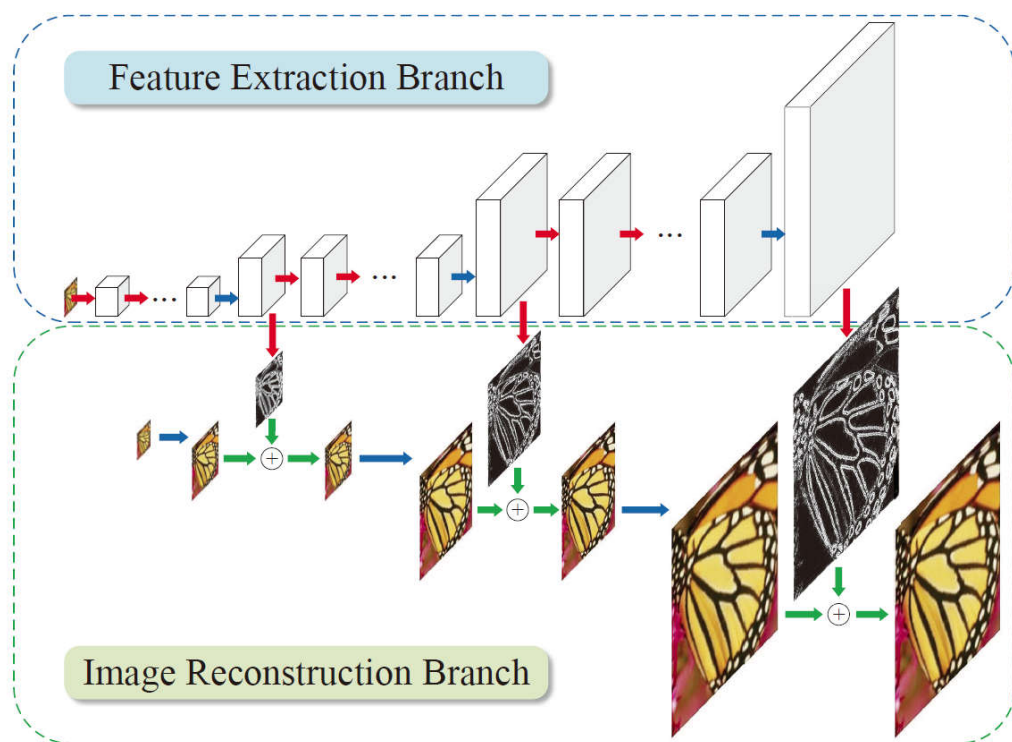


图 2-15 金字塔网络结构

在图像超分辨率研究过程中，还提出很多超分辨率卷积网络相关技术，如密集连接结构^[58]，其通过引用密集连接加强特征复用，提升超分辨率网络的性能。基于对抗生成网络(GAN)的图像超分辨率^[36]利用对抗生成网络来重建具有较好主观视觉效果图像。这些关键技术与深度学习领域的先进思想相结合，可设计出一系列高效的超分辨率算法。

2.3 图像超分辨率算法质量评价

图像超分辨率重建性能需要合理有效的评价标准，不同算法需要采用相同的质量评价标准进行评估才能体现公平性。图像质量评价通常可分为两个种类：主观评价和客观评价。主观评价指标主要依据人眼的主观视觉去评价图像的质量，定性地分析超分辨率图像重建质量；而客观评价指标主要是通过科学的方法计算出超分辨率重建图像与给出的参考图像之间某种属性上的差异，定量分析图像的优劣。

2.3.1 主观评价

主观评价主要由观察者对一幅图像进行直接观察，依据观察者自身的主观感觉对图像进行评价，最终的结果则是观察者通过给出评分的形式呈现。惯用的主观评分标准如表 2-1 所示。

表 2-1 图像质量主观评价

评价分数	相对评价	绝对评价
1	组内最差的	很差
2	差于组内平均水平的	差
3	组内平均水平	一般
4	好于组内平均水平的	好
5	组内最好的	很好

主观评价又可分为对图像质量的绝对评价和组内进行对比的相对评价两种。主观评价简单易行，但受观察者的主观意识影响较大，不同的人对同一幅图像评价的结果可能存在一定的不同。

2.3.2 客观评价

客观评价指标通过建立数学模型来计算待评价图像与参考图像之间的误差，以此来得到两幅图像的误差的具体数值。常见的客观评价指标有：均方误差(MSE)，峰值信噪比(PSNR)^[59]，结构相似度(SSIM)^[60]。

均方误差计算两幅图像各个像素点之间误差的均方值，当待评价图像与参考图像越接近时，均分误差越小。而峰值信噪比作为主流的图像评价指标，能有效评价超分辨率重建图像与参考图像的整体误差，值越大表示重建质量越好，整体失真度越低。均方误差与峰值信噪比的数学表达式可表示如下：

$$MSE = \frac{\sum_{i=1}^M \sum_{j=1}^N [S(i, j) - H(i, j)]^2}{M \times N} \quad (2-8)$$

$$PSNR = 10 \log_{10} \left(\frac{255^2}{MSE} \right) = 10 \log_{10} \left[\frac{255^2 \times M \times N}{\sum_{i=1}^M \sum_{j=1}^N [S(i, j) - H(i, j)]^2} \right] \quad (2-9)$$

式 2-8、2-9 中, M 和 N 表示图像的长度和宽度, S 和 H 分别表示超分重建后图像和真实图像, (i, j) 则表示像素点位置。

由式 2-8、2-9 中可以看出, 均方误差和峰值信噪比能对图像总体误差进行衡量, 但没有考虑图像的内部结构, 其可能对两幅完全不同的图像计算出相同的数值, 因为两幅图像的整体像素差异相同。而图像的结构相似度指标 SSIM^[60] 结合图像的结构信息来对图像质量进行评价, 其能更好的接近人眼对图像相似度的评判。若比较的两个区域的结构信息越相近, 则 SSIM 值越大。其数学表达式为:

$$\left\{ \begin{array}{l} SSIM(X, Y) = L(X, Y)^\alpha C(X, Y)^\beta S(X, Y)^\lambda \\ L(X, Y) = \frac{2\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \\ C(X, Y) = \frac{2\sigma_x \sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \\ S(X, Y) = \frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3} \end{array} \right. \quad (2-10)$$

式 2-10 中, α 、 β 、 λ 为大于 0 的调整参数, 通常取 1; μ_x 和 μ_y 表示像素均值, σ_x 和 σ_y 表示像素的方差, σ_{xy} 表示协方差, C 是为避免除 0 而加入的常数, 通过取值较小。令 $C_3 = C_2 / 2$, 则最终的计算公式是:

$$SSIM(X, Y) = \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (2-11)$$

本文中主要采用 PSNR 和 SSIM 作为超分辨率算法评价标准。

2.4 损失函数

基于卷积神经网络的图像超分辨率是一种有监督的学习任务, 因此需要成对的输入和标签, 一般使用低分辨率图像作为输入, 将高分辨率图像作为标签。在网络训练过程中需要通过损失函数计算网络生成的图像与真实图像的误差, 并通过反向传播来指导参数的更新。因此, 损失函数设计是网络训练的重要环节, 对

网络的性能有较大影响。假设训练集的第 i 个样本输入和标签分别为 x^i 和 y^i ，则其损失函数可表示为：

$$L(\theta) = \frac{1}{n} \sum_{i=1}^n L(y^i, f(x^i, \theta)) \quad (2-12)$$

式 2-12 中， θ 代表网络的参数。图像超分辨率常用的损失函数有 L_1 、 MSE 损失函数。其数学表达式分别为：

$$Loss_{L1} = \frac{1}{n} \sum_{i=1}^n |y^i - \hat{y}^i| \quad (2-13)$$

$$Loss_{MSE} = \frac{1}{n} \sum_{i=1}^n (y^i - \hat{y}^i)^2 \quad (2-14)$$

根据 Zhao^[61]的研究，由于平方的关系，MSE 损失对大的误差有较强的惩罚，对小的误差惩罚较弱，因此其对异常值比较敏感，最终会使图像趋于平滑，而 L_1 损失函数对异常值的敏感度不高，主流的超分辨率模型基本上使用 L_1 损失函数。

2.5 超分辨率研究常用数据集

数据集对网络模型的训练结果有着较大的影响，通常可分为训练集和测试集。在超分辨率领域，目前主流的训练数据集有 Yang91^[62]，General200^[63]，DIV2K^[64] 等。Yang91^[62]包含 91 张尺寸从 78×78 像素到 450×450 像素不等的彩色图像，其场景包含多种场景和多种物品种类，图 2-16 给出了该训练集中的部分图片。



图 2-16 Yang91 部分图片

General200^[63]数据集有 200 张 320×480 的彩色图像, 包含人像、动物、建筑、风景等多种风格类型, 提供了更加丰富的超分辨率数据样本。图 2-17 给出了 General200 数据集中的一些图片。

DIV2K^[64]数据集则是 2017 年在 NTIRE 超分辨率竞赛上提出的新数据集, 其包含更丰富更高清的图像, 数据集包含 800 张训练图片和 100 张验证图片, 图片分辨率为 2040×1356 。无论是图片数量还是图片分辨率都远超前两种数据集, 因此能让网络学习到更加丰富的特征信息。DIV2K 数据集的部分图片如图 2-18 所示。

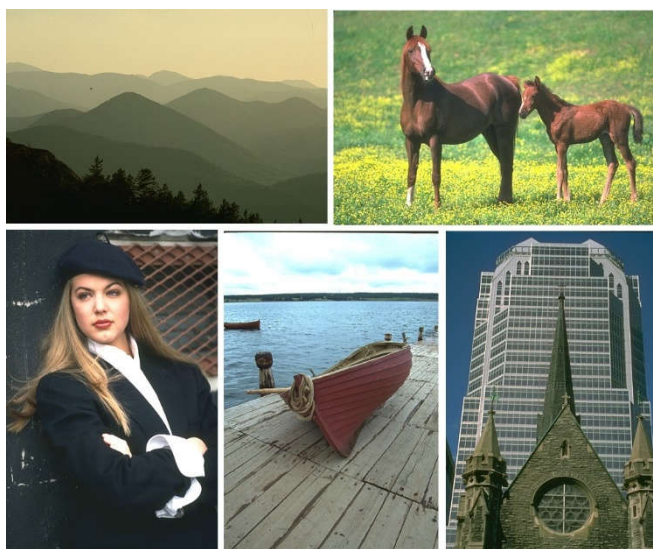


图 2-17 General200 部分图片

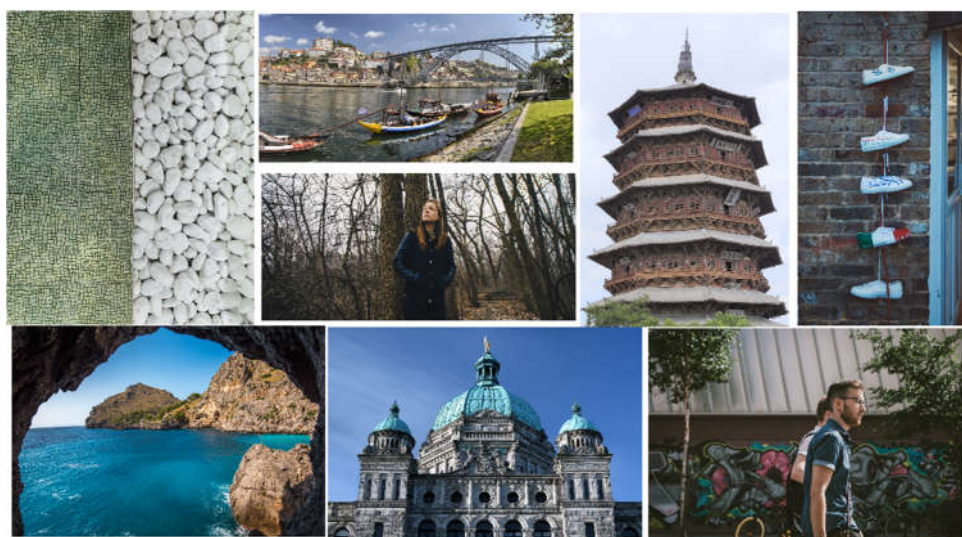


图 2-18 DIV2K 部分图片

除训练集外, 超分辨率测试集主要包含四个基准数据集, 分别是包含 5 张图片的 Set5^[65]数据集, 包含 14 张图片的 Set14^[66]数据集, 包含丰富场景的 BSD100^[67]

数据集以及只包含城市场景的 Urban100^[68]数据集，通常在以上四个基准数据集上评估超分辨率算法的优劣。

2.6 本章小结

本章主要对基于卷积神经网络的超分辨率的基础理论知识进行概述。首先介绍关于卷积神经网络的基础知识，包括神经元模型的原理，卷积层、池化层的具体实现过程，常用的激活函数，反向传播算法和优化方法等；接着介绍了关于超分辨率网络发展过程中所应用的一些关键技术，这些关键技术推动了超分辨率网络的发展；最后，介绍了图像超分辨率算法质量评价指标，以及常用的损失函数、数据集等。

第 3 章 基于分组卷积和注意力机制的超分辨率网络

基于深度学习的超分辨率方法在近年来取得较好的效果，但随着网络层数加深和结构加宽，网络的整体参数量也随之大幅增加，因而需要更多的计算资源和内存消耗，难以用于实际场景。因此，轻量化超分辨率网络成为超分辨率领域的重要研究方向之一。轻量化网络的一个重要目标即减少网络参数，如前述提到的 DRCN^[27]采用递归结构能有效减少网络的参数量，DRRN(Image Super-Resolution via Deep Recursive Residual Network)^[28]通过增加残差结构来改进 DRCN^[27]以提升超分辨率的重建效果，与标准的基于卷积神经网络的超分辨率方法相比，这两种模型能有效减少网络参数，获得较好的超分辨率效果。然而，上述两种方法采用插值方法将低分辨率图像放大到指定倍数后再输入到网络，这会增加网络的计算量并消耗大量内存；另外，通过递归结构来增加网络的深度虽然能有效地提升网络的效果，但是会增加网络运行的时间。为提升网络运行速度，IDN(Fast and Accurate Single Image Super-Resolution via Information Distillation Network)^[37]提出信息蒸馏结构来构建快速图像超分辨率网络，通过对特征图通道进行压缩来提高网络效率，但此网络没有充分考虑各特征图通道之间的相互依赖性。

为提升网络效率，实现在显著降低网络参数并提升网络运行时间的同时，尽可能保持良好的超分辨率重建性能。本章提出一种轻量级分组卷积超分辨率网络(Lightweight Group Convolutional Network, LGCN)，以下简称 LGCN。该网络基于分组卷积和注意力机制，主要由三部分组成：特征提取网络，记忆分组卷积网络以及重建网络。网络的主体部分为记忆分组卷积网络，其由数个子网络级联构成，而子网络中主要包含记忆单元和注意力机制单元，所设计的记忆单元由分组卷积层、卷积核为 1×1 的卷积层组成，并通过密集连接融合多层次的信息以实现记忆机制，密集连接能有效利用模块中的各个卷积层的特征信息，不仅能实现特征复用，同时能学习的特征的多层表示。而其中的分组卷积层因能在各个分组内各自卷积而能有效减少网络参数，卷积核为 1×1 的卷积层不仅能解决分组卷积层带来的通道信息不流通的问题，还能作为密集连接的中间层实现特征复用，并且降低密集连接带来的参数量，基于此，记忆单元能以较少的参数量完成对特征的充分利用；而注意力机制单元则融合了通道注意力机制和空间注意力机制同时对特征图的通道和特征的空间进行针对性的加权来实现对特征图进行更好的特征选择，提升网络效果。

3.1 轻量化网络设计

近年来,在图像分类、目标检测领域,出现了一些轻量且高效的神经网络^{[69][70][71]}以满足在移动端或者嵌入式设备中应用。例如, SqueezeNet(SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size)^[69]在 AlexNet^[2]的基础上,将卷积核大小为 3×3 的卷积层替换成卷积核大小为 1×1 的卷积层来实现效率的提升,能够保持与 AlexNet 相似的精度的同时,参数量可减少 50 倍。而分组卷积的出现使得轻量化网络设计进入一个新时代,分组卷积的概念首次出现是在 AlexNet 中,其主要目的是为了实现在多个 GPU 上进行训练。ResNext(Aggregated Residual Transformations for Deep Neural Networks)^[72]通过将 ResNet 网络中的常规的卷积核尺寸为 3×3 的卷积层替换成分组卷积层,证明了分组卷积在减少参数量方面的有效性。MobileNet(MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications)^[70]提出了分组卷积更加极致的版本,深度可分离卷积,通过将标准的卷积层替换成深度卷积层和逐点卷积层组合来构建一个轻量深层卷积神经网络,其中深度卷积是分组卷积的一种极端形式,其分的组数为整个输入通道数,文中证明了这种结构在各种计算机视觉任务中都是有效的。ShuffleNet(ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices)^[71]应用了深度卷积层,但是其发现在以往的深度可分离卷积层网络结构中,逐点卷积层占有很大一部分参数量,消耗很多的运算资源,因此提出了通道混合操作代替逐点卷积层来实现特征通道之间的信息流通,并验证了可在保持网络精度的同时有效地降低运算量。在另一方面, Hu^[73]提出了 SENet(Squeeze-and-Excitation Networks),其对特征通道之间相互依赖关系进行建模并对各个特征通道进行加权,通过学习的方式来获取每个特征通道的权重,加强了对当前任务有用的特征通道,并抑制无用的特征通道。

受 ResNext 网络和 SENet 网络启发,本章基于分组卷积和注意力机制设计超分辨率网络。在所设计的网络中,记忆分组卷积子网络应用了分组卷积层和通道加权机制,但与 ResNext 中的残差块设计不同的是,记忆分组卷积子网络包含了几个交替放置的分组卷积层和卷积核尺寸为 1×1 的卷积层,并通过密集连接来高效提取特征信息;另一方面,除通道注意力机制外,子网络还引入空间注意力机制,空间注意力机制能够对特征图进行空间上的加权,通过融合注意力机制基础上实现对特征图的进一步特征选择,过滤无用的特征,使得网络对有用的特征更加敏感。实验证明,本章提出的超分辨率网络能有效提升运行速度,并能获得较好的超分辨率效果。

3.2 基于分组卷积和注意力机制的超分辨率网络设计

3.2.1 网络整体结构

如图 3-1 所示,所设计的轻量级分组卷积网络(LGCN)主要包含三个子网络,分别是特征提取网络(F-Net)、记忆分组卷积网络(C-Net)和重建网络(R-Net)。在网络的整体结构上,为了得到更好的超分辨率效果,采用了分层特征融合和长程连接来有效融合网络浅层特征与高层特征,以使网络能学习到更丰富的特征信息。与 SRCNN^[23]和一些先将低分辨率图上采样再作为网络输入等方法不同,该网络直接输入低分辨率图像,并在网络最后采用反卷积层进行上采样得到特定尺寸的图像。

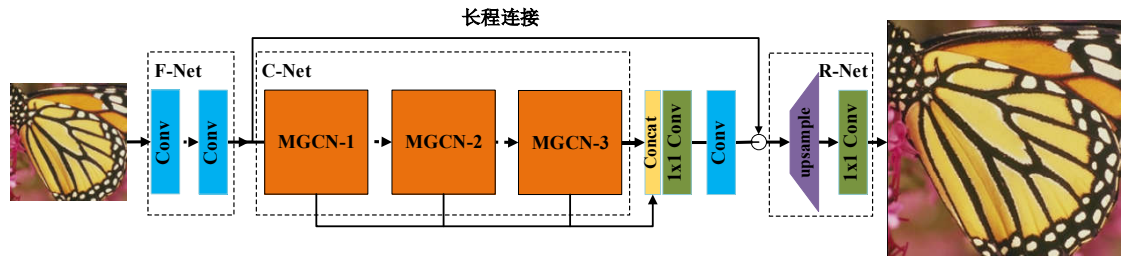


图 3-1 网络整体结构

假定 I_{LR} 和 I_{SR} 分别代表超分辨率网络的输入和输出,即输入的低分辨率图像和超分辨率重建后的图像。对于特征提取网络,其包含两个卷积层,主要作用是从低分辨率图像提取基本的图像特征,该过程可表示为:

$$F_0 = H_F(I_{LR}) \quad (3-1)$$

式 3-1 中, $H_F(\cdot)$ 代表特征提取网络中的非线性映射, F_0 代表提取到的浅层视觉特征,它将被用作下一个网络,即记忆分组卷积网络(C-Net)的输入。

记忆分组卷积网络级联了数个子网络。假设网络包含 K 个记忆分组卷积子网络(MGCN),则记忆分组卷积网络的过程可以表示为:

$$F_k = H_{MGCN,k}(H_{MGCN,k-1}(\cdots(H_{MGCN,1}(F_0)))) \quad k = 1, 2, \dots, K \quad (3-2)$$

式 3-2 中, F_k 表示第 k 个子网络的输出,而 $H_{MGCN,1}(\cdot)$ 表示第 k 个子网络所包含的运算,子网络中主要包含了记忆单元和注意力单元。所设计的子网络可高效地利用网络自身内部所有卷积层提取到的特征,其细节会在下一节进行介绍。

然后,为充分利用记忆分组卷积网络提取的特征,在网络中加入分层特征融合和长程连接,如图 3-1 中所示,可以表示为:

$$F_G = H_{HF}([F_1, \cdots F_k \cdots F_K]) + F_0 \quad (3-3)$$

式 3-3 中, F_G 表示经过分层特征融合和长程连接的操作后的输出特征图, $[F_1, \cdots F_k \cdots F_K]$ 表示由记忆分组卷积子网络 $1, \dots, k \dots K$ 所提取的特征图的堆叠。关于网络所需要的记忆分组卷积子网络的数量,在本章第四节会有讨论。 $H_{HF}(\cdot)$ 表

示分层特征融合的操作,其由一个卷积核大小为 1×1 的卷积层和一个卷积核大小为 3×3 的卷积层组成,前者用来融合之前的记忆分组卷积子网络提取到的特征,后者则用来进一步提取图像特征。

在网络最后是重建网络。重建网络由一个反卷积层和一个卷积核大小为 1×1 的卷积层组成。为得到特定尺寸的图像,首先采用反卷积层生成高分辨率图像,可表示为:

$$F_{UP} = H_{UP}(F_G) \quad (3-4)$$

式 3-4 中, F_{UP} 和 $H_{UP}(\cdot)$ 分别表示上采样后特征图和反卷积层的运算。然后经过一个卷积核大小为 1×1 的卷积层得到整个网络的输出。

$$I_{SR} = H_{REC}(F_{UP}) = H_{LGCN}(I_{LR}) \quad (3-5)$$

式 3-5 中, $H_{REC}(\cdot)$ 和 $H_{LGCN}(\cdot)$ 分别表示最终的重建卷积层和整个网络的映射。

3.2.2 级联的记忆分组卷积网络

本节将详细介绍超分辨网络中的记忆分组卷积网络(C-Net),其级联了数个子网络(MGCN),如 3-1 所示。其中,子网络结构如图 3-2 所示。子网络包含了一个记忆单元和一个注意力机制单元。由先前卷积层所提取特征首先通过记忆单元进一步提取特征,然后通过注意力机制单元对特征图的在通道和空间上进行加权以提升模型的特征表示能力。

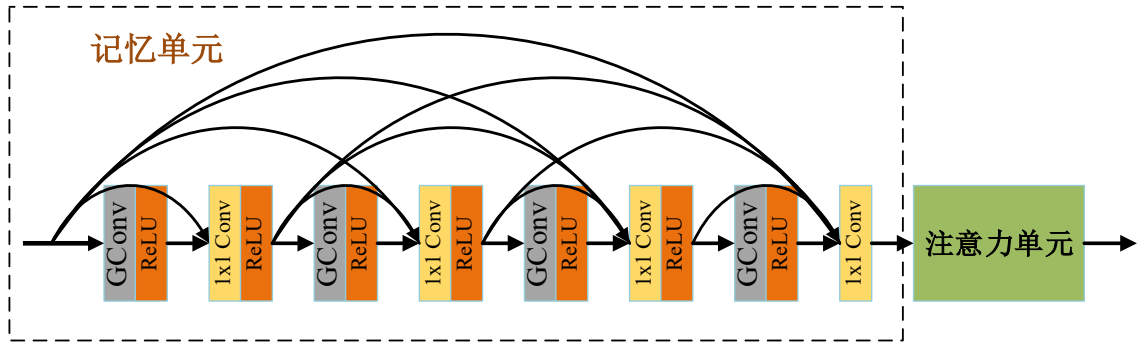


图 3-2 记忆分组卷积子网络(MGCN)

● 记忆单元

如图 3-2 所示,记忆单元包含数个交替放置的分组卷积层和卷积层尺寸为 1×1 的卷积层,并通过密集连接的方式将当前层之前的所有层特征图传递到当前层完成特征融合,这种类似于让当前层加强对网络前层特征记忆的方式称为记忆机制,也是记忆单元名字的由来。

a)记忆机制。网络中间卷积层的输出通过直接连接的方式传递到网络的后端卷积层来构成记忆机制。如图 3-2 所示。与 DenseNet(Densely Connected Convolutional Networks)^[33]的结构不同,记忆单元采用卷积核大小为 1×1 的卷积

层作为中间层来收集前面卷积层的特征信息。假设 F_d 是记忆单元的输出, F_{k-1} 是记忆分组卷积的输入。则在第 k 个记忆分组卷积子网络中的记忆单元中第 j 个分组卷积层的输出可表示为:

$$\begin{aligned} F_g^{k,1} &= \tau(W_g^{k,1}[F_{k-1}]) \\ F_g^{k,j} &= \tau(W_g^{k,j}[F_1^{k,j-1}]) \quad j = 2, \dots, J \end{aligned} \quad (3-6)$$

式 3-6 中, τ 表示激活函数, J 表示分组卷积层的个数, $W_g^{k,j}$ 表示第 j 个分组卷积层的权重, $F_1^{k,j-1}$ 表示第 $j-1$ 个卷积核为 1×1 的卷积层的输出。然后, 在第 k 个记忆分组卷积子网络中的记忆单元中第 i 个卷积核为 1×1 的卷积层的输出可表示为:

$$\begin{aligned} F_1^{k,1} &= \tau(W_1^{k,1}[F_{k-1}, F_g^{k,1}]) \\ F_1^{k,i} &= \tau(W_1^{k,i}[F_{k-1}, F_1^{k,1}, F_1^{k,2}, \dots, F_1^{k,i-1}, F_g^{k,i}]) \quad i = 2, \dots, I-1 \end{aligned} \quad (3-7)$$

其中 $W_1^{k,i}$ 表示第 i 个卷积核为 1×1 的卷积层的权重, τ 表示激活函数, $[F_{k-1}, F_g^{k,1}]$ 、 $[F_{k-1}, F_1^{k,1}, F_1^{k,2}, \dots, F_1^{k,i-1}, F_g^{k,i}]$ 表示特征图堆叠操作。 I 表示卷积核为 1×1 的卷积层的数量, 最终的记忆单元的输出可表示为:

$$F_d = F_1^{k,I} = W_1^{k,I}[F_{k-1}, F_1^{k,1}, F_1^{k,2}, \dots, F_1^{k,I-1}, F_g^{k,I}] \quad (3-8)$$

卷积核大小为 1×1 的卷积层通过密集连接能够使得记忆单元融合网络前层的特征, 使得网络可以学习到丰富的多层次特征。而且通过这种连接能加强信息流通从而加快网络信息的反向传播。另一方面, 因为通过这种连接实现了对前层特征的复用, 也能降低网络的参数量使得网络更加高效。

b) 分组卷积。如图 3-3(b)所示, 相较于标准卷积, 分组卷积能有效减少模型参数。假设 D_k 是分组卷积层的卷积核尺寸, m 和 n 分别代表分组卷积层的输入和输出通道数。通过在特征图周围的零填充操作, 分组卷积层的输入和输出特征图的尺寸是一样的, 并设为 D_F 。下面将对比使用标准卷积的网络和使用分组卷积的网络之间的运算量。当使用标准卷积时, 网络中记忆单元的运算量为:

$$4 \times (D_k \times D_k \times M \times N \times D_F \times D_F) + (2 \times N + 3 \times N + 4 \times N + 5 \times N) \times M \times D_F \times D_F \quad (3-9)$$

式 3-9 包含了 4 个卷积核尺寸为 $D_k \times D_k$ 的标准卷积和 4 个卷积核尺寸为 1×1 的卷积的运算量总和。当采用分组卷积层时, 记忆单元的运算量为:

$$4 \times (D_k \times D_k \times M \times \frac{N}{G} \times D_F \times D_F) + (2 \times N + 3 \times N + 4 \times N + 5 \times N) \times M \times D_F \times D_F \quad (3-10)$$

式 3-10 中的 G 代表着分组卷积分的组数, 因此运算量减少的比例为:

$$\begin{aligned} & \frac{4 \times (D_k \times D_k \times M \times \frac{N}{G} \times D_F \times D_F) + (2 \times N + 3 \times N + 4 \times N + 5 \times N) \times M \times D_F \times D_F}{4 \times (D_k \times D_k \times M \times N \times D_F \times D_F) + (2 \times N + 3 \times N + 4 \times N + 5 \times N) \times M \times D_F \times D_F} \quad (3-11) \\ &= \frac{4D_k^2 + 14G}{4GD_k^2 + 14G} \end{aligned}$$

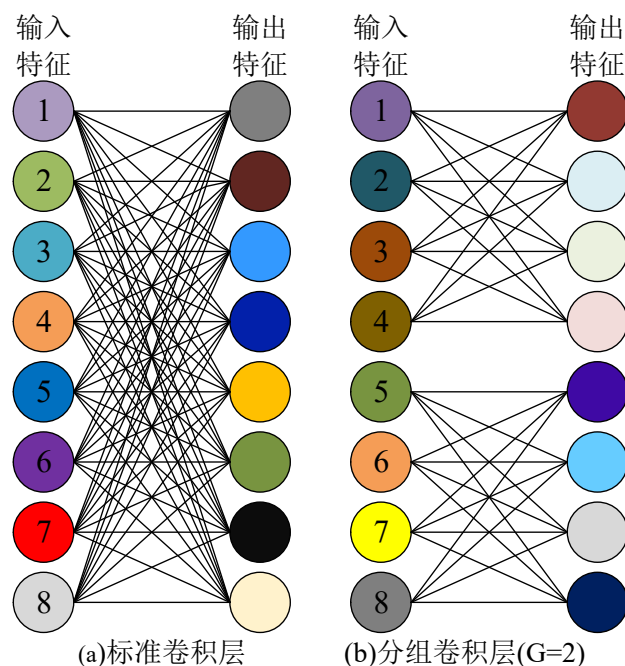


图 3-3 标准卷积结构与分组卷积结构比较

在本章网络中，分组卷积层的卷积核尺寸为 3×3 ，也就是 $D_k = 3$ 。分的组数 G 可以设成 2 到 64 之间，也就意味着相应的计算量可以减少 1.6 到 3.4 倍，这取决于分组的数量。

● 注意力机制单元

为实现更好的超分辨重建效果，网络引入注意力机制，其能在只增加微小的参数量的情况下，较有效地提升超分辨率网络的重建效果。在提出的注意力机制单元中，融合了通道注意力机制和空间注意力机制，并将注意力机制单元放置在记忆单元之后，实现对特征的加权效果。

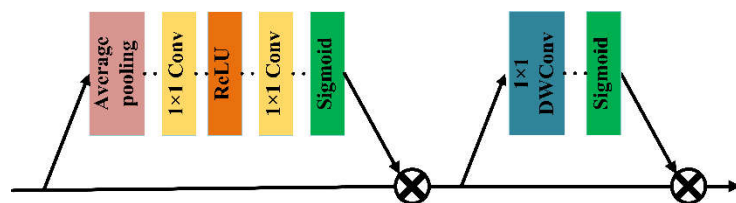


图 3-4 注意力机制

a) 通道注意力机制。通道注意力机制首次在 SENet^[73]被提出并广泛用于图像分类和目标识别问题。在此之前，基于卷积神经网络的超分辨率方法通常认为特征图每个通道的信息对网络性能的影响是一样的，但在实际中特征图各个通道的重要性不尽相同，由此，可以通过对特征图的通道进行加权来使得网络更关注有用的特征通道，而减少对无用通道的关注，并以此来提升网络的表达能力。因

此, 受 SENet 启发, 记忆分组卷积网络中引入通道注意力机制, 与之不同的是, 本文提出的通道注意力机制采用卷积核为 1×1 的卷积层而不是全连接层来学习权重参数, 使其能获得灵活的输入输出, 并获取较好的加权效果, 如图 3-4 所示。

首先, 在每个通道的特征图上提取全局空间信息。假设输入为 $U=[u_1, \dots, u_c, \dots, u_C]$, 其包含 C 个通道的特征图, 并且特征图尺寸为 $H \times W$ 。应用全局平均池化层生成各个通道的统计量 z , 则 z 的第 c 项可表示为:

$$z_c = S(u_c) = \frac{1}{H \times W} \sum_i^H \sum_j^W u_c(i, j) \quad (3-12)$$

式 3-12 中, $S(\cdot)$ 表示全局平均池化操作, $u_c(i, j)$ 表示在第 c 个特征图 u_c 中 (i, j) 位置的像素值。

为从全局平均池化收集的信息中充分提取通道之间的依赖性, 通过两个卷积层来学习各通道的权重映射, 并采用 Sigmoid 函数作为门机制以获得 0 到 1 之间的权重输出。

$$\alpha = \sigma(W_e \delta(W_s z)) \quad (3-13)$$

式 3-13 中 σ 和 δ 分别表示 Sigmoid 函数和 ReLU 激活函数。 W_s 和 W_e 分别表示两个卷积层的权重。最终得到各通道的权重统计量 α 。

b) 空间注意力机制。卷积过程中每一个通道代表一种滤波器, 因此各个通道代表着不同的意义, 一些通道主要提取竖直方向的边缘信息, 而另一些通道提取水平方向的信息。对于超分辨率领域, 每个通道的特征图应有自身关注的区域, 例如依靠复杂滤波器提取边缘或者边缘信息的特征图, 应当更加注重细节信息; 而对于包含天空等没有高频信息区域的特征图, 应当更加关注低频信息。因此, 可利用空间注意力机制对每个通道的特征图进行特征图空间上的加权, 如图 3-4 所示。

利用前述深度卷积层, 为每个通道生成不同的空间加权特征图。假设输入为 x , 则各通道特征图的权重可表示为:

$$\beta = \sigma(W_d(x)) \quad (3-14)$$

式 3-14 中, W_d 表示深度卷积层的参数。 σ 表示 Sigmoid 函数。

因此, 如图 3-4 所示, 所设计的记忆分组卷积子网络的输出 F_k 可表示为:

$$F_k = F_{sp}(F_{ch}(F_d, \alpha), \beta) \quad (3-15)$$

式 3-15 中, F_d 表示记忆单元的输出, 其首先通过通道注意力机制进行通道上的加权, 进而通过空间注意力机制对每个单独的特征图进行空间上的加权。通过两种注意力机制, 首先能使得网络更加关注有用的特征通道, 其次能使得网络能关注更为有用的特征图区域, 完成更加高效的特征提取, 因此, 能够有效地抑制噪声的影响, 并使得网络更加关注有用的特征信息。

3.3 实验与结果分析

为更好理解所设计网络的超分辨率性能,本节首先介绍实验时一些参数的设置,包括训练和测试中数据集、优化方法和训练平台等方面的设置,以及网络实现过程中的参数设置;之后,通过在超分辨率领域的基准数据集上与一些经典方法关于超分辨率重建效果和算法运行的速度等方面进行对比,验证轻量级分组卷积超分辨率网络(LGCN)的性能;最后,通过消融实验证明网络各模块的有效性。

3.3.1 实验设置

(a) 训练与测试设置

为得到较好的超分辨率重建效果,本章所有实验的训练数据集均采用来自于 Yang^[62]的 91 张图片和来自于 Berkeley Segmentation Dataset(BSD)^[63]数据集的 200 张图片,一共 291 张图片。数据在训练前通过三种方式进行增强:(1)对图像进行 90 度,180 度,270 度的旋转;(2)取图像的水平镜像;(3)对图像分别进行 0.9、0.8、0.7、0.6 倍率的下采样,通过以上方法可有效扩充数据集,使网络学习到更加丰富的特征。在测试时,采用 Set5^[65]、Set14^[66]、BSD100^[67]和 Urban100^[68]四个基准数据集作为测试集对网络性能进行评估。四个基准数据集中的 RGB 通道的图像都先转换为 Ycbcr 颜色通道,在训练和测试时,网络只处理 Y 通道,对于另外两个通道采用双三次线性插值进行上采样。对于低分辨率数据集,通常采用双三次线性插值的方法将高分辨率图像下采样得到低分辨率数据集。

训练过程中,随机抽取 64 个图像块作为一个训练批次。对于不同的上采样倍数,为实现 2 倍、3 倍和 4 倍放大,首先,用图像块尺寸为 $29^2/58^2$, $15^2/45^2$, $11^2/44^2$ 的 LR/HR 图像对进行训练,然后采用图像块尺寸为 $35^2/70^2$, $25^2/75^2$, $19^2/76^2$ 的 LR/HR 图像对对模型进行微调以充分地对数据集进行训练;最后,又采用 $64^2/128^2$, $43^2/129^2$, $32^2/128^2$ LR/HR 图像对再次进行微调。网络训练的优化器采用 ADAM^[54]优化器,参数设置为 $\beta_1 = 0.9$, $\beta_2 = 0.999$ 以及 $\varepsilon = 10^{-8}$ 。学习率初始化为 10^{-4} ,并在每次训练 2.5×10^5 个迭代次数后减小 10 倍,训练的总迭代次数为 5×10^5 。另外,在两次微调过程中,学习率设置为 10^{-5} ,训练轮次为 2.5×10^5 次迭代。网络参数通过 He^[74]方法进行初始化,其中偏置项被设置为 0。本章实验的环境为 Python3.5,Pytorch,Ubuntu 16.04,GPU 为 NVIDIA TITAN XP,显存为 12GB。

(b) 网络参数设置

在记忆分组卷积子网络(MGCN)中,包含 4 个分组卷积层和 4 个卷积核尺寸为 1×1 的卷积层,即 $I=J=4$,并且将特征提取网络所有卷积层的卷积核尺寸设置

为 3×3 。对于所有的卷积核尺寸为 3×3 的卷积层以及反卷积层，采用周围填充零像素的方式保持特征图尺寸不变。除了卷积核尺寸为 1×1 的卷积层，所有卷积层的输出都经过 ReLU 激活函数。对于不同的放大倍数，网络中反卷积层采用不同的卷积核尺寸，例如，对于两倍放大，采用卷积核尺寸为 4×4 ，步长为 2；对于三倍放大，采用卷积核尺寸为 5×5 ，步长为 3。为避免放大倍数较大时参数量过大，在四倍放大时，采用两个卷积核尺寸为 4×4 ，步长为 2 的反卷积层进行上采样。第一个卷积层的通道数为 $C_0=128$ ，其余层的通道数设置为 $C=64$ ，不包括通道注意力机制中的两个卷积层，通道注意力集中的第一个卷积层的通道数为 $C/8=8$ 。因网络只处理 Y 通道，所以网络最后一层只有一个通道输出。

3.3.2 实验结果

为验证本章提出的轻量级分组卷积超分辨率网络(LGCN)的有效性，本节从客观指标和主观效果两方面与目前主流的超分辨率方法进行对比，对比方法包括：Bicubic、SRCNN^[23]、VDSR^[26]、LapSRN^[30]、MS-LapSRN^[75]、DRRN^[28]和 IDN^[37]。

(a) 客观评价指标对比

对于每种方法，在四个基准数据集上计算超分辨率重建后的图像与真实图像之间的峰值信噪比(PSNR)以及结构相似性(SSIM)值。如表 3-1 所示，LGCN 在多数放大尺寸以及多数数据集上取得了最好的效果。在 Set5 数据集上，当放大 2 倍或 4 倍时，LGCN 相较于 IDN 的 PSNR 值有着微小提升。但 Set5 数据集只包含 5 张图片，是最小的数据集。在放大 4 倍时，LGCN 在 Set14 数据集上比 IDN 算法的 PSNR 上高 0.05dB，而在数据量更大的数据集 Urban100 上则有 0.08dB 的提升。另外，LGCN 在所有数据集上，对于所有放大倍数都取得最好的 SSIM 值，也意味着所提轻量级分组卷积超分辨率网络(LGCN)所重建图像与原始图像有最好的相似度。

(b) 主观效果对比

为直观比较各算法性能，从 Set14^[66]、BSD100^[67]和 Urban100^[68]中选取三张放大 4 倍的图像进行对比，如图 3-5 所示。从图中可以看出，所对比的多数方法对于红色框内的花儿恢复比较模糊，因为这些方法对高频细节丢失较严重，而 LGCN 对花儿细节的恢复则较为清晰。图 3-6 中，LGCN 不仅能清晰恢复港口木板的水平缝隙，还能在船体边缘恢复出更加清晰的凸起，如红色框所示。对于图 3-7，LGCN 能获取比其他方法更加清晰的曲线。由此，可以看出本章方法能有效恢复图像细节，例如花上的点，水平线，以及曲线段等。

表 3-1 客观评价指标对比

算法	尺度	Set5		Set14		BSD 100		Urban 100	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	2	33.66	0.930	30.24	0.869	29.56	0.843	26.88	0.841
SRCNN ^[23]	2	36.65	0.954	32.29	0.903	31.36	0.888	29.52	0.895
VDSR ^[26]	2	37.53	0.958	32.97	0.913	31.90	0.896	30.77	0.914
LapSRN ^[30]	2	37.52	0.959	33.08	0.913	31.80	0.895	30.41	0.910
DRRN ^[28]	2	37.74	0.959	33.23	0.913	<u>32.05</u>	0.897	<u>31.23</u>	0.919
MS-LapSRN ^[75]	2	37.72	<u>0.959</u>	33.24	0.914	32.00	0.898	31.01	0.917
IDN ^[37]	2	37.83	0.960	<u>33.30</u>	<u>0.914</u>	32.08	<u>0.899</u>	31.27	<u>0.920</u>
LGCN	2	<u>37.75</u>	0.960	33.31	0.918	<u>32.05</u>	0.901	31.27	0.924
Bicubic	3	30.39	0.868	27.55	0.774	27.21	0.739	24.46	0.735
SRCNN ^[23]	3	32.75	0.909	29.30	0.822	28.41	0.786	26.24	0.799
VDSR ^[26]	3	33.67	0.921	29.78	0.832	28.83	0.799	27.14	0.830
LapSRN ^[30]	3	33.82	0.923	29.87	0.832	28.82	0.798	27.07	0.828
DRRN ^[28]	3	34.03	0.924	29.96	0.835	<u>28.95</u>	0.800	27.56	<u>0.838</u>
MS-LapSRN ^[75]	3	34.01	0.924	29.96	<u>0.836</u>	28.92	<u>0.801</u>	27.39	0.835
IDN ^[37]	3	<u>34.11</u>	<u>0.925</u>	<u>29.99</u>	0.835	<u>28.95</u>	<u>0.801</u>	27.42	0.836
LGCN	3	34.12	0.926	30.01	0.837	28.96	0.803	<u>27.47</u>	0.840
Bicubic	4	28.42	0.810	26.00	0.703	25.96	0.668	23.14	0.658
SRCNN ^[23]	4	30.48	0.863	27.61	0.754	26.91	0.712	24.53	0.724
VDSR ^[26]	4	31.35	0.882	28.03	0.770	27.29	0.726	25.18	0.753
LapSRN ^[30]	4	31.54	0.885	28.19	0.772	27.32	0.728	25.21	0.756
DRRN ^[28]	4	31.68	0.888	28.21	0.772	27.38	0.728	25.44	0.764
MS-LapSRN ^[75]	4	31.74	0.888	28.25	0.773	27.42	<u>0.731</u>	<u>25.45</u>	<u>0.765</u>
IDN ^[37]	4	31.82	<u>0.890</u>	<u>28.25</u>	<u>0.773</u>	<u>27.41</u>	0.730	25.41	0.763
LGCN	4	<u>31.79</u>	0.891	28.30	0.780	27.42	0.735	25.49	0.766



清晰图像



(a) 细节清晰图像



(b) LGCN(本章算法)

(c) IDN^[37]

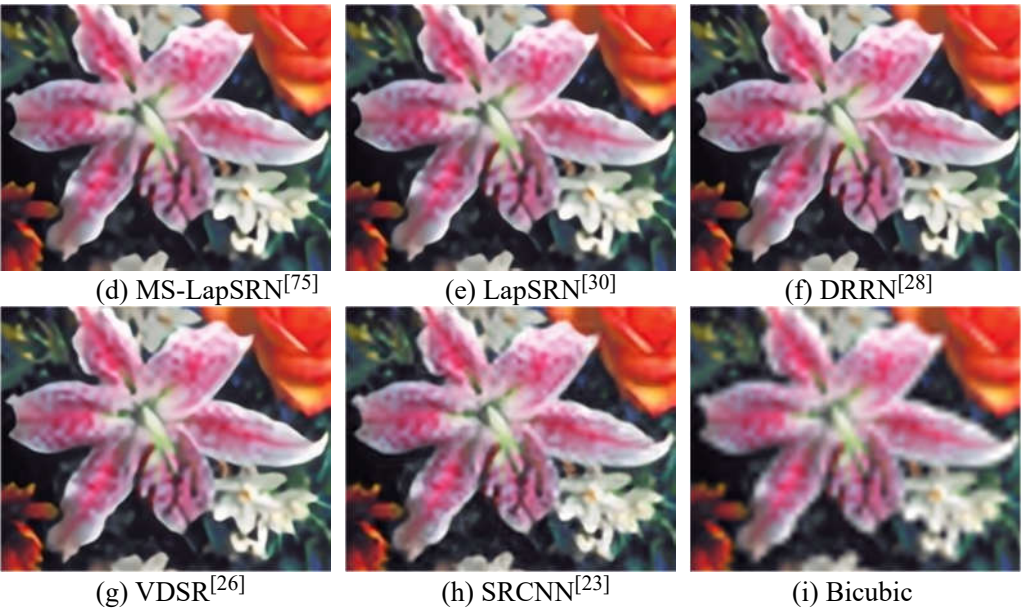
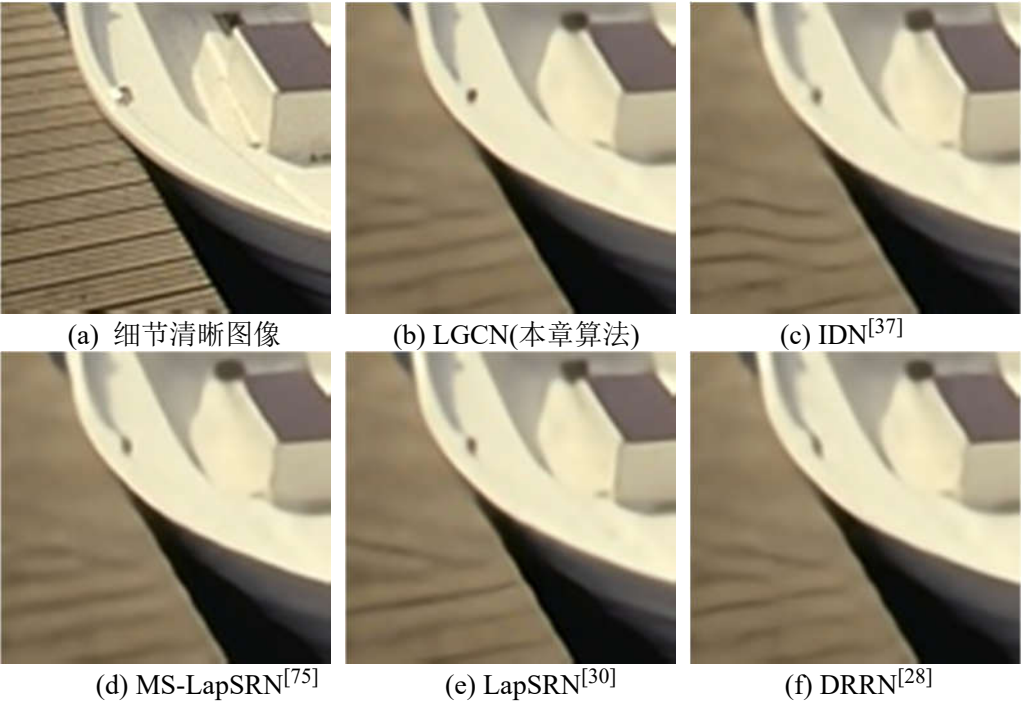


图 3-5 “花” 图片细节对比



清晰图像



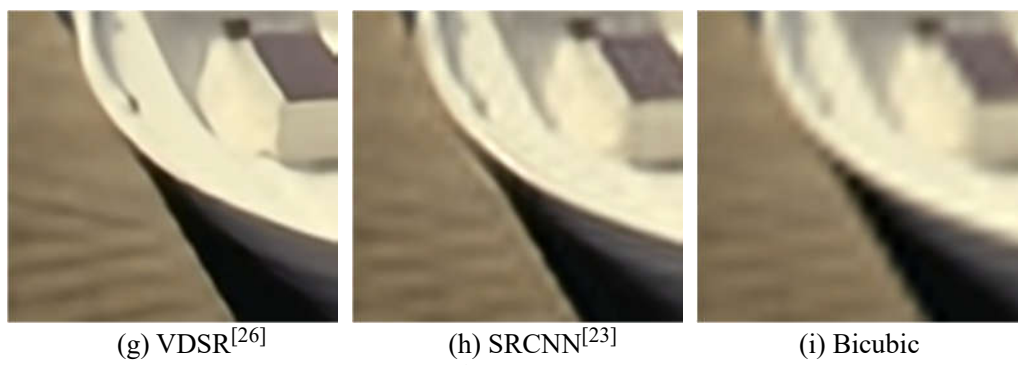
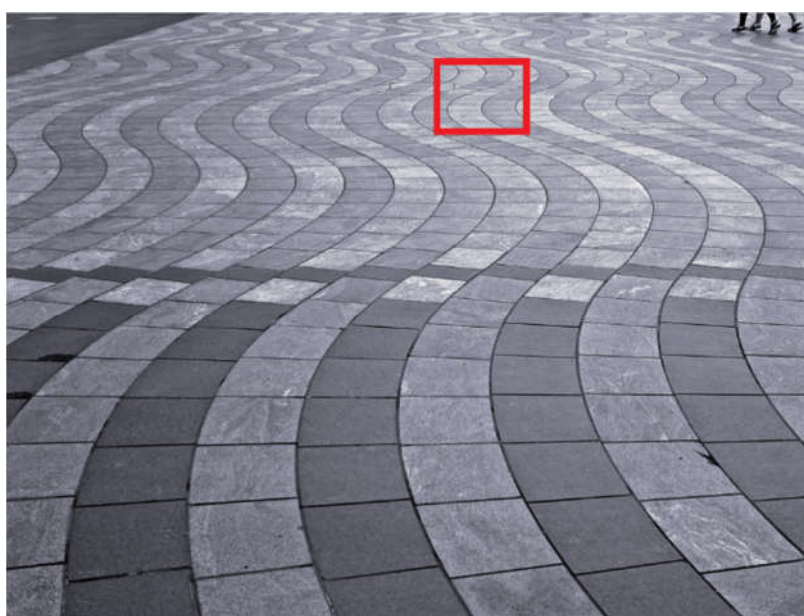
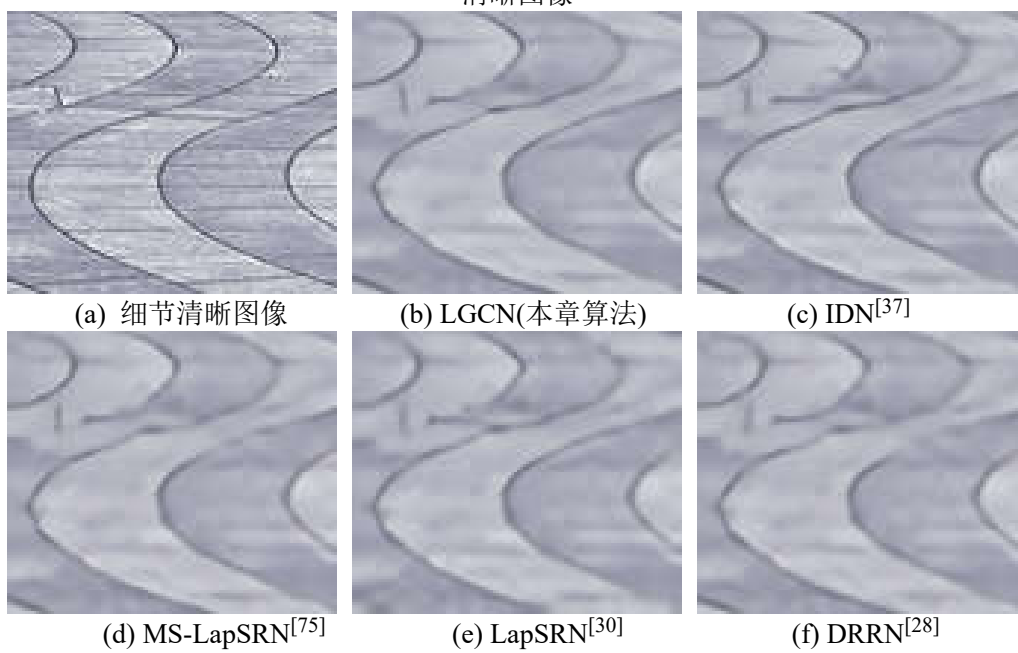


图 3-6 “船”图片细节对比



清晰图像



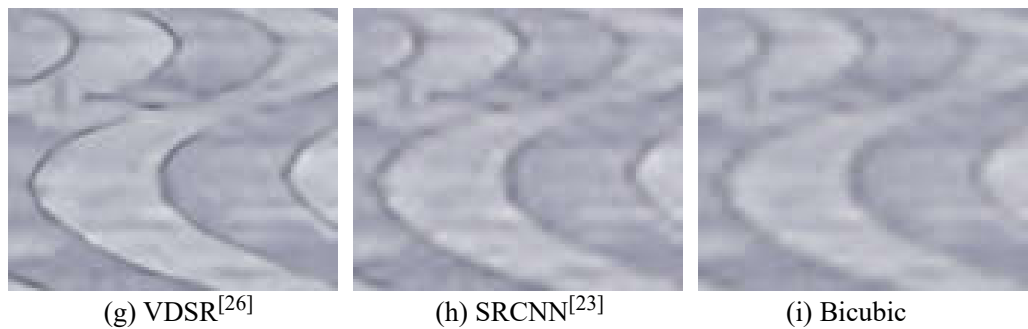


图 3-7 “路面” 图片细节对比

(c) 运行时间对比

在相同的硬件设备上执行算法的开源代码，比较算法的运行时间。设备的配置为主频为 3.2GHz 的 Inter Core i7 CPU，内存为 16GB，显卡为 NVIDIA TITAN XP，GPU 显存 12GB。由于 SRCNN 的代码是在 CPU 上实现并测试的，因此使用 PyTorch 重新构建代码，然后在 GPU 上测试其运行时间。测试时放大倍数为 4，数据集为 Set14。如图 3-8 所示，LGCN 在 Set14 数据集上处理一张图片的平均时间约为 0.01 秒，和 IDN 算法的运行时间相当，而与其他方法相比，LGCN 具备更快的运行速度。具体的，LGCN 比 LapSRN 约快 4 倍，比 MS-LapSRN 快 5 倍。另外，相较于先将低分辨率特征图上采样再输入到网络进行运算的网络，LGCN 比 SRCNN 快 6 倍，比 VDSR 和 DRRN 快 10 倍以上。LGCN 取得较快运行速度的原因之一，是 LGCN 在低分辨率图像上进行卷积运算，而并非和 SRCNN、VDSR 或 DRRN 一样先进行上采样再输入到网络；另一个原因在于 LGCN 没有采用 DRRN 和 MS-LapSRN 所包含的递归结构，递归结构虽然能有效降低模型参数，但也会带来更长的运行时间。此外，对于 LapSRN，其在放大 4 倍的过程中会在放大 2 倍的高分辨率特征图上进行运算，这种操作也会增加网络运算时间。

3.3.3 消融实验

本节通过消融实验分析网络中各重要模块的有效性。为了简化操作，所有的消融实验结果都在 Set5 数据集上进行 4 倍放大测试。

表 3-2 记忆单元各部件分析

记忆单元	1×1 卷积	✓	✓
	密集连接		✓
PSNR (dB)		31.52	31.57 31.79

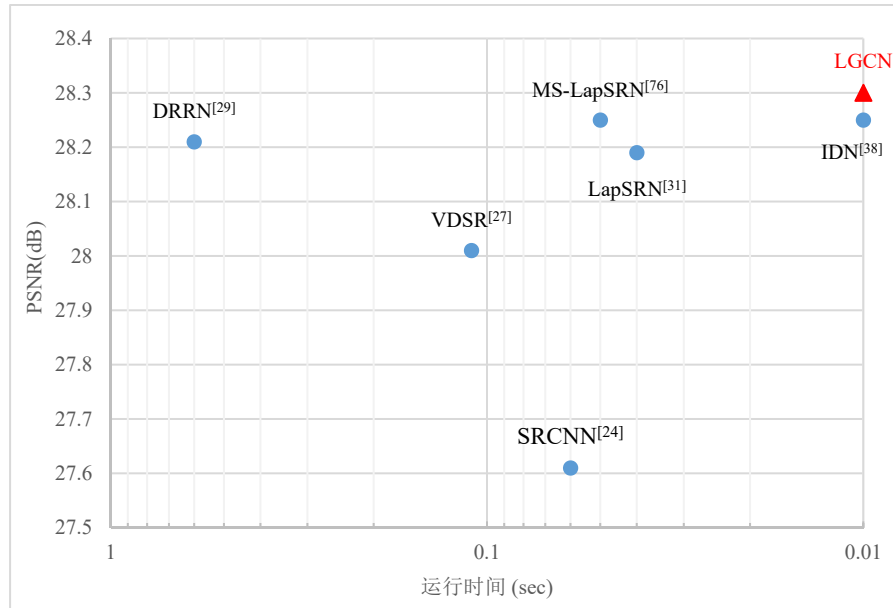


图 3-8 运行时间比较

(1) **记忆单元** 表 3-2 给出了关于卷积核尺寸为 1×1 的卷积层和密集连接对超分辨率重建结果产生的影响。这三个模型都包含 3 个记忆分组卷积子网络，表 3-2 中第一列是初始模型，只包含四个串联的分组卷积，不包括卷积核尺寸为 1×1 的卷积层和密集连接。由于所提网络中的密集连接采用卷积核尺寸为 1×1 的卷积层为中间层来聚集浅层的信息，因此对于第二行中的“密集连接”，意味着同时包含卷积核为 1×1 的卷积层和密集连接操作。在表 3-2 中，初始模型的 PSNR 值为 31.52dB，当仅仅添加卷积核为 1×1 的卷积层到记忆分组卷积子网络时，PSNR 相较于初始模型提升了 0.05dB。该实验证明卷积核为 1×1 的卷积层能对分组卷积层的输出进行线性融合以提升网络的表达能力。然而，如同 ResNet 中讨论过的那样，网络会由于简单地添加了卷积核为 1×1 的卷积层变得更深，这样可能会造成梯度消失现象，使得网络发生退化。通过密集连接加强前后层的信息流通，可有效地避免梯度消失现象，PSNR 值提升至 31.79dB。因此，通过实验，记忆模块被证明对网络重建效果有较大的提升。

表 3-3 注意力机制分析

	无注意力	通道注意力	空间注意力	通道+空间注意力
参数量	650K	653K	651K	654K
PSNR (dB)	31.66	31.73	31.71	31.79

(2) **注意力机制** 表 3-3 展示了通道注意力机制与空间注意力机制的不同组合方式的效果比较。其中无注意力表示记忆分组卷积网络不包含注意力单元。实验可知，只包含通道注意力机制的注意力单元较无注意力单元提升了 0.07dB，参数量仅增加 3K；而只包含空间注意力机制的注意力单元较无注意力单元提升了 0.05dB，参数仅增加 1K；同时采用两种注意力机制时，网络取得了最高的 PSNR

值 31.79。由以上比较可知，混合的注意力机制能有效提升网络性能，同时只带来微小的参数量增加。

表 3-4 分组卷积数分析

分组数量	1	2	4
参数量	875K	653K	543K
PSNR (dB)	31.90	31.79	31.63

(3) 分组卷积 对于不同的分组卷积的分组数，例如 $G = 1$ 、2 或者 4，训练不同的模型，其参数量和对应的 PSNR 值如表 3-4 所示。对于 $G = 1$ 的情况，即卷积层为标准卷积层而非分组卷积层。当组数为 2 时，PSNR 值下降了 0.11dB，参数量也下降了 222K。当组数变为 4 时，PSNR 值相比于 $G=2$ 时下降了 0.16dB，同时参数量也相对应的下降了 110K。显而易见的是，当组数增加时，参数量也明显降低，但网络性能也相应变差。为了实现一种平衡的效果，在本章网络(LGCN)中选择 $G = 2$ ，即分组数量为 2。如图 3-9 所示，当组数为 2 时，网络既实现了较为轻量的模型，同时也展现出良好的重建效果。虽然 DRRN，MS-LapSRN 比 LGCN 的参数量更少，但这两个网络都采用了递归结构构建更深的网络，会导致更高的运算开销，网络运行时间是所提章网络的 5 到 10 倍。而对于 SRCNN，其只拥有 3 个卷积层，只能取得很有效重建效果。而所提 LGCN 相较于 IDN，在 PSNR 值上也有着 0.05dB 的提升。总体来说，本章所提的轻量级分组卷积超分辨率网络(LGCN)在只含有相对较少的参数的情况下，取得了更好的超分辨率重建效果。

表 3-5 分组卷积子网络个数分析

K	2	3	4
参数量	515K	653K	792K
PSNR(dB)	31.66	31.79	31.89

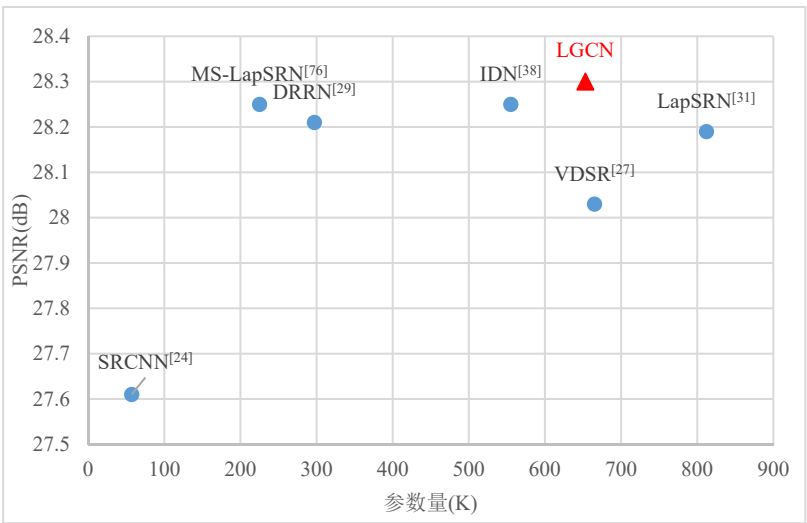


图 3-9 参数量比较

(4) 分组卷积子网络数量分析。对于不同数量 $K=2、3、4$ 的分组卷积子网络训练不同的模型，以考量分组卷积子网络的数量对网络性能的影响。表 3-5 给出了当不同的分组卷积子网络数量所对应的参数量以及重建图像的 PSNR 值。从表中可以看出，堆叠更多的分组卷积子网络能获取更好的超分辨率重建性能，但是也会带来更多网络参数。为了更好的实现两者之间的平衡，LGCN 使用 3 个分组卷积子网络，即 $K=3$ 。

3.4 本章小结

本章提出一种轻量、快速的超分辨率算法，可以高效地对低分辨率图像进行超分辨率重建。针对现有的大多数超分辨率算法存在的参数量大，运行速度慢等问题，本章首先采用分组卷积降低参数量，进而为解决分组卷积带来的通道信息不流通的问题，引入卷积核尺寸为 1×1 的卷积层融合通道信息，并以卷积核为 1×1 的卷积层作为中间层构建高效的密集连接结构以实现特征的复用；另外，提出结合通道注意力机制和空间注意力机制的注意力机制使得网络更加关注有用的特征通道和特征图区域。实验结果表明，无论从客观评价指标还是主观视觉效果本算法都取得较好的结果，并保持较快的运行速度以及较少的模型参数。

第 4 章 基于动态卷积核选择的多尺度超分辨率网络

在利用卷积层提取图像特征时，不同的卷积核尺寸对应着不同感受野，可以提取特征图中不同尺度的结构信息。图像的超分辨率重建中对低分辨率图像的特征提取通常需要不同的感受野以适应不同尺度的特征信息，而现有的基于卷积神经网络的超分辨率方法基本都只采用一个的较大或是较小的感受野，单一尺度上特征提取往往不能完整地获取图像的特征信息，不足以完全重建图像的高频纹理信息。现有的多尺度超分辨率方法中，Shi 提出 MSSRNet (Single image super-resolution with dilated convolution based multi-scale information learning inception module)^[38]在同一层利用不同尺寸的卷积核提取特征，并将提取到的多个尺度的特征直接相加进行融合，Li^[39]提出了多尺度残差超分辨率网络 (Multi-scale Residual Network for Image Super-Resolution)，利用卷积核尺寸不同的两路分支实现多尺度特征提取，利用堆叠的方式融合多尺度特征。但这两种网络均采用线性组合的方式融合多尺度特征信息，无法完成感受野的自适应调整，有着较大的局限性。

基于上述问题，为提取图像不同尺度的特征，实现感受野的动态调整，更加完整地提取图像的特征信息，本章提出一种动态卷积核选择多尺度残差超分辨率网络 (Selective Kernel based Multi-Scale Residual Network, SKMRN)。网络的整体结构如图 4-1 所示。网络分成特征提取网络、多重级联的多尺度残差网络，以及重建网络。网络采用了动态卷积核选择卷积层实现了对多种尺寸卷积核的动态选择，以完成对特征图的多尺度信息的提取，通过动态卷积核选择卷积层构建动态卷积核选择子残差块，并以子残差块为基础构建多尺度残差块，利用这种残差嵌套结构实现了对低分辨率特征图信息的充分利用，克服了现有多尺度网络缺乏对感受野的动态调节的缺陷；在整体结构上，网络采用多重级联机制加强网络中信息流通，有效避免梯度消失问题；此外，还引入权重归一化层加快网络训练并提升超分辨率网络的重建效果。

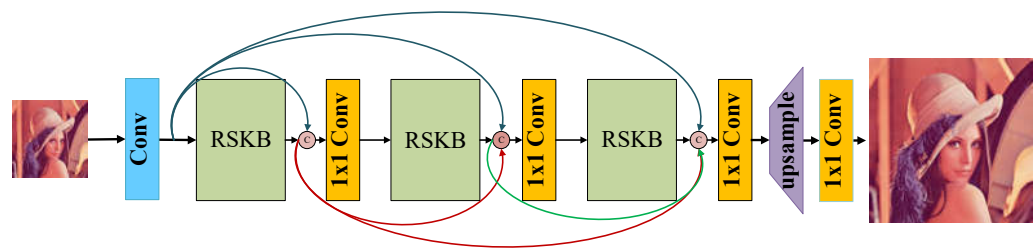


图 4-1 网络整体结构(SKMRN)

4.1 多尺度网络设计

卷积神经网络起源于人类对视觉皮层的工作原理的研究,并由此提出感受野的概念。在视觉皮层中,同一区域包含神经元的感受野大小是不同的,使得神经元在一个阶段处理不同尺度的信息。基于此,关于多尺度卷积神经网络设计的研究在近年来取得较大的进展。**GoogleNet^[76]提出的 Inception 模块在网络中使用不同大小的卷积核提取同一层的不同特征,通过对不同卷积核的并行融合增加了提取特征的丰富性。**Shi^[38]利用 Inception 结构设计出一种多尺度超分辨率网络,提取低分辨率图像的特征信息,在基准数据集上取得了较好的超分辨率效果。Li^[39]提出多尺度残差结构,采用两个不同尺寸的卷积核同时提取特征,然后将两路输出特征图堆叠起来再次进行相同操作完成多尺度特征的提取。通过这种局部多路径学习,模型能够很好地从多个尺度提取图像特征,有效提高性能。

然而,上述多尺度网络没有利用到皮层神经元的另外一个特性,即感受野大小的自适应改变。大量实验表明神经元的感受野大小不是固定的,而是通过刺激来调节的^[77],而上述这些具有多个分支的网络存在一定的内部机制来实现在下个卷积层根据输入的内容调整神经元感受野大小,具体实现方式为在下个卷积层对不同分支的多尺度特征信息通过线性组合的方式进行融合,但这种线性组合的融合方式可能满足不了神经元强大的调整能力的需要。为实现以非线性的方式来融合多个卷积核的信息, Li^[78]提出一种可选择卷积核的单元模块用于分类任务,其包含多个由不同卷积核尺寸构成的分支,最终通过由这些分支的特征信息决定的权重值对各个分支进行融合,由不同的权重值决定最终得到的融合层中采用的感受野大小。而在超分辨率领域,目前还没有对于自适应卷积核尺寸的神经网络的研究。因此本章对可选择卷积核模块进行改进,提出一种适应超分辨率任务的动态卷积核选择模块。

因此, **本章将动态卷积核选择卷积层引入图像超分辨率任务**,同时改进了原始可选择卷积核模块,去除掉模块中的批归一化层,改权重归一化层以适应超分辨率任务,权重归一化层能有效抑制噪声,同时完成参数重写,避免训练时出现梯度爆炸。本文在以上基础上设计出具有残差嵌套机制的多尺度残差块,每个多尺度残差块包含三个动态卷积核选择子残差块,通过这种残差嵌套的设计能充分利用局部信息,使得大量的低频信息通过跳跃连接直接传输到网络后层,提高训练速度,加强反向传播时的信息流通。而在整体结构上,所设计网络采用多重级联机制,利用卷积核尺寸为 1×1 的卷积层作为瓶颈层减少网络的参数量。通过这种设计,本章算法取得较好的超分辨率重建效果,并且相较于其他算法有着较少的参数量。

4.2 基于动态卷积核选择的多尺度超分辨率网络设计

4.2.1 网络整体结构

如图 4-1 所示,本章所提出的动态卷积核选择多尺度残差网络,简称 SKMRN,主要分为三个部分,分别是浅层特征提取网络,多重级联的多尺度残差网络以及最后的重建网络。网络结构在整体上采用多重级联机制,与传统的 DenseNet(Densely Connected Convolutional Networks)^[33]中密集连接不同,SKMRN 采用卷积核为 1×1 的卷积层作为瓶颈层以减少网络参数;多重级联机制能够实现特征的重复利用,并加强网络中前后层的信息流通以有效融合网络浅层特征和高层特征进而学习更丰富的特征信息,瓶颈层能有效降低网络后层的参数量,使网络更高效。SKMRN 直接采用低分辨率图像作为输入,而非上采样后的图像,在网络后层采用亚像素卷积层进行上采样以获得指定尺寸的高分辨率图像。

设 SKMRN 的输入和输出分别为 I_{LR} 和 I_{SR} ,代表着低分辨率图像和超分重建完成后的图像。对于浅层特征提取网络,如图 4-1 中蓝色框所示,其包含一个卷积层,主要用于从低分辨率图像中提取基本的图像特征。可表示为:

$$F_0 = H_C(I_{LR}) \quad (4-1)$$

在式 4-1 中, $H_C(\cdot)$ 表示浅层特征提取网络的非线性映射关系, F_0 则表示卷积层提取到的特征图,被用作网络后层网络的输入。

在浅层特征提取网络之后是多重级联的多尺度残差网络,其包含数个多尺度残差块(RSKB),并通过多重级联进行连接,图 4-1 中带有字母 C 的圆圈表示特征图堆叠操作,假设网络包含 N 个多尺度残差块,则多重级联的多尺度残差网络中多尺度残差块的输出可表示为:

$$\begin{aligned} F_R^1 &= H_{RSKB,1}(F_0) \\ F_R^n &= H_{RSKB,n}(F_1^{n-1}) \quad n=2 \cdots N \end{aligned} \quad (4-2)$$

式 4-2 中, F_R^1 表示第一个多尺度残差块的输出,其输入为浅层特征提取网络的输出 F_0 , F_R^n 表示第 n 个多尺度残差块的输出,其输入为前面一个卷积核为 1×1 的卷积层的输出。 $H_{RSKB,n}$ 表示第 n 个多尺度残差块所包含的映射关系,其主要由残差嵌套机制构成。而卷积核为 1×1 的卷积层的输出是:

$$F_1^n = \tau(W_1^n[F_0, F_R^1 \cdots, F_R^n]) \quad n=1 \cdots N \quad (4-3)$$

式 4-3 中, τ 表示激活函数, N 表示卷积核为 1×1 的卷积层的数量, W_1^n 表示第 n 个卷积核为 1×1 的卷积层的权重参数, $[F_0, F_R^1 \cdots, F_R^n]$ 表示对特征图的堆叠操作,即图 4-1 中的带 C 的圆圈的操作, F_1^n 表示第 n 个卷积核为 1×1 的卷积层

的输出。多重级联的多尺度残差网络的最终输出表示为式 4-4，即最后一个卷积核为 1×1 的卷积层的输出。

$$F_M = F_1^N \quad (4-4)$$

经多重级联的多尺度残差网络提取的特征图最后经过重建网络进行放大，其包括一个亚像素卷积层和一个卷积核大小为 1×1 的卷积层。为实现特定尺寸的超分辨率放大，首先采用亚像素卷积层得到特定尺寸高分辨率图像，其中亚像素卷积层包含一个标准卷积层以得到特定通道数的特征图，之后再行像素重排操作，整个过程表示为：

$$F_{UP} = H_{UP}(F_M) \quad (4-5)$$

式 4-5 中， F_{UP} 和 $H_{UP}(\cdot)$ 分别表示高分辨率特征图和亚像素卷积层的操作。网络的最后经过一个卷积核为 1×1 的卷积层以得到整个网络的输出。

$$I_{SR} = H_{REC}(F_{UP}) = H_{SKMRN}(I_{LR}) \quad (4-6)$$

式 4-6 中， $H_{REC}(\cdot)$ 和 $H_{SKMRN}(\cdot)$ 分别表示最后的重建卷积层以及整个网络的映射关系。

动态卷积核选择多尺度残差网络(SKMRN)主要由三个子网络组成，首先是由一个卷积层组成的浅层特征提取网络完成特征的粗略提取；接着，特征图输入到多重级联的多尺度残差网络中学习进一步的非线性映射；最后，经过由亚像素卷积层和重建卷积层构成的重建网络得到指定尺寸的高分辨率图像。多重级联的多尺度残差网络作为网络的主要部分，将在 4.2.2 节详细介绍。

4.2.2 多重级联的多尺度残差网络

如图 4-1 所示，多重级联的多尺度残差网络通过多重级联机制连接了数个多尺度残差块(RSKB)。多尺度残差块(RSKB)的结构如图 4-2 所示，其采用残差嵌套机制，每个多尺度残差块包含三个动态卷积核选择的子残差块(SKB)和一个卷积层并采用跳跃连接构成残差结构。

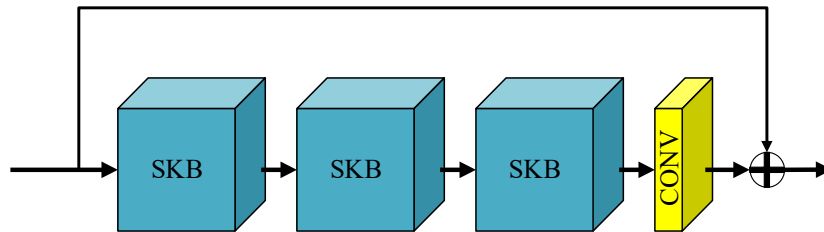


图 4-2 多尺度残差块(RSKB)

(a)多尺度残差块(RSKB)。低分辨率的原始图像和特征图蕴含着丰富的特征信息，超分辨率网络的目标是恢复更加有用的信息，对于特征图中丰富的低频信息可以通过恒等跳跃连接直接传递到网络后层。为了构建更加有效的残差学习机

制，每个多尺度残差块采用多个动态卷积核选择子残差块。假设每个多尺度残差块中使用了 D 个动态卷积核选择子残差块，第 m 个多尺度残差块中的第 d 个动态卷积核选择子残差块的输出可表示为：

$$F_m^d = H_m^d(F_m^{d-1}) = H_m^d(H_m^{d-1}(\cdots H_m^1(F_{m-1})\cdots)) \quad (4-7)$$

其实第 m 个多尺度残差块中的第 d 个动态卷积核选择子残差块的输入为 F_m^{d-1} ，输出为 F_m^d ，动态卷积核选择子残差块的运算表示为 $H_m^d(\cdot)$ 。为了使主干网络更加关注更加有用的特征信息，多尺度残差块采用跳跃连接获得残差块的输出，可表示为：

$$F_m = F_{m-1} + W_m F_m^D = F_{m-1} + W_m H_m^D(H_m^{D-1}(\cdots H_m^1(F_{m-1})\cdots)) \quad (4-8)$$

第 m 个多尺度残差块最后的一个卷积层的参数为 W_m 。跳跃连接使得多尺度残差块学习残差信息，在训练时，大量的低频信息通过跳跃连接快速直接传递到网络后层，加强了网络的信息流通，加快了网络收敛速度。

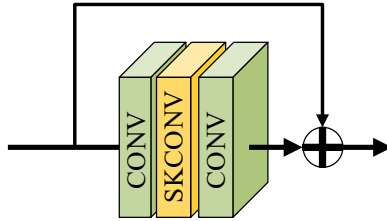


图 4-3 动态卷积核选择子残差块

(b)动态卷积核选择子残差块。动态卷积核选择子残差块的结构如图 4-3 所示，其包含三个卷积层以及一个跳跃连接。为了构建更加高效的残差模块，子残差块被设计为两个卷积核尺寸为 1×1 的卷积层以及一个动态卷积核选择卷积层 (SKCONV) 的组合，首先利用前面的卷积核尺寸为 1×1 的卷积层对特征图进行降维，之后通过动态卷积核选择卷积层进行多尺度特征提取，最后通过另一个卷积核尺寸为 1×1 的卷积层恢复到原始维度。可以表示为：

$$F_m^d = \tau(W_1^2(W_{sk}(\tau(W_1^1(F_m^{d-1})))))) + F_m^{d-1} \quad (4-9)$$

式 4-9 中， F_m^{d-1} 和 F_m^d 为第 m 个多尺度残差块中的第 d 个动态卷积核选择残差块的输入和输出， W_1^1 和 W_1^2 分别表示两个卷积核尺寸为 1×1 的卷积层的参数， τ 表示激活函数， W_{sk} 表示动态卷积核选择卷积层的参数。

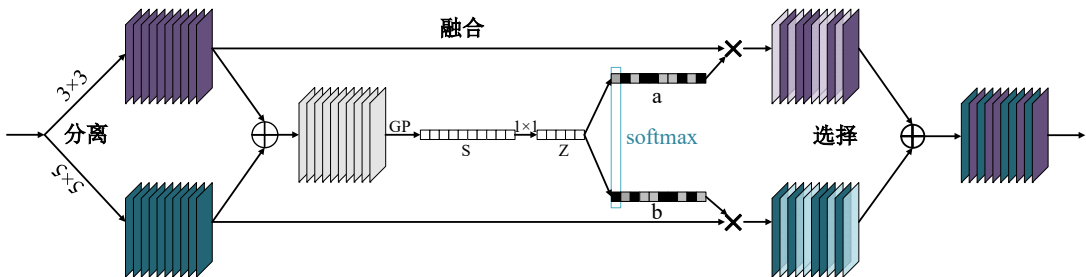


图 4-4 动态卷积核选择卷积层

对于动态卷积核选择卷积层，其结构如图 4-4 所示。其主要包含三个操作，分离，融合，以及选择。下面分析包含两个分支的情况，即含有两个不同的卷积核。

分离：假设给定的输入是 X ，其包含 C' 个通道，首先在两个分支进行不同的卷积运算得到两组包含 C 个通道的尺寸为 $H \times W$ 的特征图 \tilde{U} 和 \hat{U} ，表示为：

$$\begin{aligned}\tilde{U} &= \tilde{F}(X) \\ \hat{U} &= \hat{F}(X)\end{aligned}\quad (4-10)$$

其中两个卷积核大小分别是 3×3 和 5×5 。与原始 SKNet^[78]不同的是，本章网络移除了批归一化层，而采用权重归一化层。因此式 4-10 中 $\tilde{F}(\cdot)$ 和 $\hat{F}(\cdot)$ 操作，由上章提到的组卷积层、权重归一化层以及 ReLU 激活函数组成。

融合：对于动态卷积核选择卷积层，其目标是使得神经元能够根据输入的内容动态地调整其感受野，其基本思想是使用门机制来控制输入到下一个神经元的承载着多个分支的不同尺度的信息的信息流。为实现这个目标，首先将两个分支的结果采用逐元素相加进行合并：

$$U = \tilde{U} + \hat{U} \quad (4-11)$$

然后，为得到每个通道特征图的全局空间信息，采用全局平均池化(GP)生成各个通道特征图的全局统计量 s ，则 s 的第 c 项可以写成：

$$s_c = F_{gp}(U_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W U_c(i, j) \quad (4-12)$$

其中 H 和 W 为特征图的尺寸， U_c 表示第 c 个通道的特征图， $U_c(i, j)$ 表示在点 (i, j) 的像素值。为实现对特征图信息流的精确和自适应控制，采用一个卷积核尺寸为 1×1 的卷积层来学习其映射关系，并通过降维来提升效率：

$$z = F_1(s) = \delta(w_n(W_1(s))) \quad (4-13)$$

其中 δ 表示激活函数， w_n 表示权重归一化^[79]， W_1 表示卷积核为 1×1 的卷积层的参数。维数减少的倍数设置为 r 。

选择：采用跨通道的注意力机制来动态地选择来自不同空间尺度的信息，选择的依据则是之前学习到的特征向量 z ，具体做法为对不同分支的同一维度的特征向量做 softmax 操作，以此决定卷积层最终输出中该维度采用哪个分支的特征图。可表示如下：

$$a_c = \frac{e^{A_c z}}{e^{A_c z} + e^{B_c z}}, b_c = \frac{e^{B_c z}}{e^{A_c z} + e^{B_c z}} \quad (4-14)$$

其中 A 和 B 代表两个全连接层，用于将特征向量的通道数恢复到 C ，因为本节以两路分支为例，所以只包含 A 、 B 两个全连接层，全连接层能进一步学习映射关系的得到两组特征向量， a 和 b 代表对于两个分支的权重向量，其由两组特征向量经过 softmax 操作得到。全连接层参数可看成一个矩阵， A_c 代表全连接

层 A 的第 c 行, a_c 代表 a 的第 c 项, B_c 和 b_c 同理。最后, 通过对各个不同卷积核的分支的特征图赋予权重以得到最终的输出特征图 V 。

$$V_c = a_c \cdot \tilde{U}_c + b_c \cdot \hat{U}_c \quad a_c + b_c = 1 \quad (4-15)$$

最终卷积层输出为 $V = [V_1, V_2, \dots, V_C]$, 即经过动态卷积核选择后的输出特征图。

4.2.3 权重归一化层

对于批归一化层, 其已经在 EDSR^[31]中被证实其不适合用于超分辨率任务, 主要原因在于: 第一, 批归一化层基于一个小批量数据计算均值和方差, 而不是基于全部的训练集, 而对于图像超分辨率的网络中, 通常采用小图像块和小批量数, 使得均值和方差在各个小批量中差异很大, 使得统计数据不稳定, 容易在梯度计算时引入噪声, 因此不适用于对噪声敏感的模型; 第二, 训练时和测试时采用不同的公式可能使得逐点像素值预测的准确性下降。针对这个问题, 本文所设计网络采用权重归一化层^[79]。

权重归一化^[79]是神经网络中权重向量的参数重写方法, 其将权重向量的长度与它们的方向分离。权重归一化不会在训练时引入批量数据之间的依赖关系, 并且在训练和测试中具有相同的公式。假设输出 y 可写成:

$$y = w \cdot x + b \quad (4-16)$$

式 4-16 中, w 是一个 k 维的权重向量, b 是一个标量偏置项, x 是输入的 k 维特征向量。权重归一化层根据新的参数 g 和 v 采用下式来重写了权重向量。

$$w = \frac{g}{\|v\|} v \quad (4-17)$$

式 4-17 中, v 是一个 k 维的向量, g 是一个标量, $\|v\|$ 代表 v 的欧式范数, 通常标量 g 的值为 $\|w\|$, 即权重 w 的大小, 与参数 v 独立, 在训练过程中优化 g 与 v 。权重归一化层能加快网络的收敛速度, 并在测试集上获得更高的准确率。

4.3 实验与结果分析

本节首先介绍网络训练与测试时的超参数设置、优化方法以及训练集、测试集; 然后对网络实现中的参数做进一步的说明; 之后, 在测试集上将动态卷积核选择多尺度残差网络(SKMRN)与其他优秀的超分辨率方法进行对比以证明 SKMRN 的有效性; 最后, 通过消融实验来分析网络中各设计的有效性。

4.3.1 实验设置

(a) 训练与测试设置

本章所有实验采用的训练数据集均为 DIV2K^[64]数据集，如第二章所述，DIV2K^[64]数据集是一个新提出的包含高质量图片的数据集，其包括 800 张训练图片，能够使得网络学习到丰富的特征。对于测试集，本章采用与上一章相同的 Set5^[65]、Set14^[66]、BSD100^[67]以及 Ubran100^[68]四个基准数据集来评估网络的超分辨率重建性能。采用双三次线性插值的方式将高分辨率图像下采样到指定尺寸以制作低分辨率数据集。

训练过程采用尺寸为 64×64 的 RGB 通道的低分辨率图像块作为网络输入，对数据进行水平翻转和 90 度旋转以实现数据增强，并随机抽取 64 个图像块作为一个训练批次。网络采用第二章介绍的 ADAM^[54]优化器更新参数，基础参数设置为： $\beta_1 = 0.9$ ， $\beta_2 = 0.999$ 以及 $\epsilon = 10^{-8}$ 。训练时总的迭代次数设置为 6×10^5 ，学习率初始化为 10^{-4} ，每迭代 4×10^5 次，学习率减半。网络的权重和偏置初始化为服从均匀分布 $U(-w, w)$ 的随机值，其中 $w = 1/\sqrt{c_{in}}$ ， c_{in} 输入特征图的通道数。本章实验的环境为 Python3.5, Pytorch, Ubuntu 16.04, 显卡为 NVIDIA TITAN XP, 显存大小为 12GB。

(b) 网络参数设置

网络采用 3 个多尺度残差块以及 3 个卷积核为 1×1 的卷积层，即 $M=3$ 。在多尺度残差块中，包含 3 个动态卷积核选择子残差块，以及一个卷积核尺寸为 1×1 的卷积层，即 $D=3$ 。浅层特征提取网络中卷积层的卷积核尺寸设置为 3×3 。对于除卷积核尺寸为 1×1 的卷积层外的所有卷积层均采用零填充的方式保持卷积前后特征图尺寸不变，并都经过 ReLU 激活函数。对于不同的放大倍数，网络采用不同上采样结构，对于两倍和三倍放大，网络采用一个亚像素卷积层上采样；对于四倍放大，网络采取先上采样两倍，再上采样两倍的策略，即采用两个亚像素卷积层。网络的动态卷积核选择残差块中的前一个卷积核尺寸为 1×1 的卷积层的输出通道数设置为 32，对于动态卷积核选择卷积层中的融合部分中，卷积层维数降低倍数 r 设置为 2，即输出通道数为 16，亚像素卷积层中的卷积层输出通道数设置为该层输入通道数与放大倍数平方的乘积，即 $64 \times s^2$ ，最后一个卷积层的输出通道数为 3，即输出三通道 RGB 图像，其余卷积层的输出通道数设置为 64。

4.3.2 实验结果

为了对动态卷积核选择多尺度残差网络(SKMRN)的有效性进行验证，本节首先从客观评价指标和主观视觉效果两个方面与近年来先进的超分辨率网络进行对比，在此之后对各个网络的参数量进行分析。对比的算法主要包括 Bicubic、SRCNN^[23]、VDSR^[26]、DRCN^[27]、LapSRN^[30]、MS-LapSRN^[75]以及 CARN^[35]。

(a) 客观评价指标对比

如第二章所述，客观评价指标包括峰值信噪比(PSNR)^[59]以及结构相似性(SSIM)^[60]值。在四个基准数据集 Set5^[65]、Set14^[66]、BSD100^[67]以及 Urban100^[68]上分别计算各个方法的平均的 PSNR 值与 SSIM 值，结果图表 4-1 所示。SKMRN 在所有的尺度上对于所有的数据集都取得最高的 PSNR 值和 SSIM 值，对比近年来先进的超分辨率网络都取得了客观指标上的较大提升。具体的，SKMRN 在放大 2 倍时，在 Set5 数据集上比 CARN 算法高 0.1dB，在图片尺寸较大，数据量也较多的 Urban100^[68]数据集上较 CARN 提升了 0.22dB 由此说明 SKMRN 能取得与原始图像最相近的像素相似度和结构相似度。

表 4-1 客观评价指标

算法	尺度	Set5		Set14		BSD100		Urban 100	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	2	33.66	0.930	30.24	0.869	29.56	0.843	26.88	0.841
SRCNN ^[23]	2	36.65	0.954	32.29	0.903	31.36	0.888	29.52	0.895
VDSR ^[26]	2	37.53	0.958	32.97	0.913	31.90	0.896	30.77	0.914
LapSRN ^[30]	2	37.52	0.959	33.08	0.913	31.80	0.895	30.41	0.910
DRCN ^[27]	2	37.63	0.959	33.04	0.912	31.85	0.894	30.75	0.913
MS-LapSRN ^[75]	2	37.72	0.959	33.24	0.914	32.00	<u>0.898</u>	31.01	0.917
CARN ^[35]	2	<u>37.76</u>	<u>0.959</u>	<u>33.52</u>	<u>0.916</u>	<u>32.09</u>	0.897	<u>31.92</u>	<u>0.926</u>
SKMRN	2	37.86	0.963	33.59	0.919	32.15	0.901	32.15	0.936
Bicubic	3	30.39	0.868	27.55	0.774	27.21	0.739	24.46	0.735
SRCNN ^[23]	3	32.75	0.909	29.30	0.822	28.41	0.786	26.24	0.799
VDSR ^[26]	3	33.67	0.921	29.78	0.832	28.83	0.799	27.14	0.830
LapSRN ^[30]	3	33.82	0.923	29.87	0.832	28.82	0.798	27.07	0.828
DRCN ^[27]	3	33.82	0.923	29.76	0.831	28.80	0.796	27.15	0.828
MS-LapSRN ^[75]	3	34.01	0.924	29.96	0.836	28.92	0.801	27.39	0.835
CARN ^[35]	3	<u>34.29</u>	<u>0.926</u>	<u>30.29</u>	<u>0.840</u>	<u>29.06</u>	<u>0.803</u>	<u>28.06</u>	<u>0.846</u>
SKMRN	3	34.38	0.928	30.33	0.841	29.10	0.805	28.28	0.854
Bicubic	4	28.42	0.810	26.00	0.703	25.96	0.668	23.14	0.658
SRCNN ^[23]	4	30.48	0.863	27.61	0.754	26.91	0.712	24.53	0.724
VDSR ^[26]	4	31.35	0.882	28.03	0.770	27.29	0.726	25.18	0.753
LapSRN ^[30]	4	31.54	0.885	28.19	0.772	27.32	0.728	25.21	0.756
DRCN ^[27]	4	31.53	0.885	28.02	0.767	27.23	0.723	25.14	0.751
MS-LapSRN ^[75]	4	31.74	0.888	28.25	0.773	27.42	0.731	25.45	0.765
CARN ^[35]	4	<u>32.13</u>	<u>0.893</u>	<u>28.60</u>	<u>0.780</u>	<u>27.58</u>	<u>0.734</u>	<u>26.07</u>	<u>0.783</u>
SKMRN	4	32.20	0.895	28.63	0.781	27.60	0.735	26.21	0.789

(b) 主观评价指标对比

为更直观的比较各个算法的超分辨率算法的超分辨率重建效果，在 Set14^[66]、BSD100^[67]以及 Urban100^[68]中选取较难恢复的放大 4 倍的图像进行主观视觉比较。如图 4-5 中所示，SKMRN 对图片红色框内手的细节恢复得比其他先进的网络要更加清晰，其他网络对于手指、指甲部分的高频细节恢复比较模糊，SKMRN 对这一部分要恢复更加完整。对于图 4-6，SKMRN 能够正确地恢复地面的圆形的

划痕，而有些网络则出现横向的错误纹理，SKMRN 恢复的纹理比大部分对比算法更加清晰。而对于图片 4-7，明显地观察到 SKMRN 能恢复出更多的横向条纹线，具体来说，较其他 要多恢复两条横条纹。综上所述，本章所提的动态卷积核选择多尺度残差超分辨率网络(SKMRN)能较好地复原图像的纹理细节信息，恢复出来的图片整体质量更高。

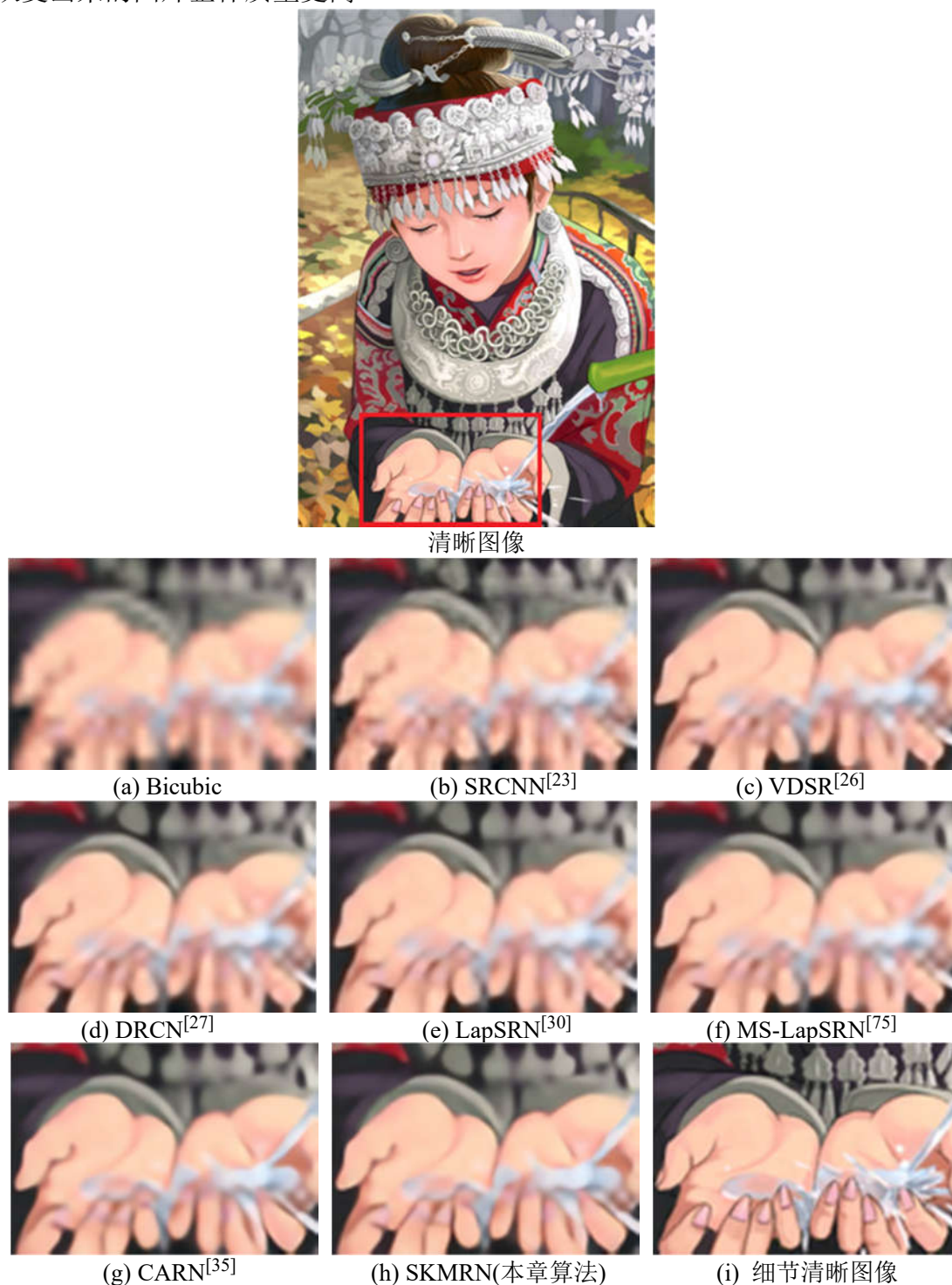


图 4-5 “Comic” 图像细节对比

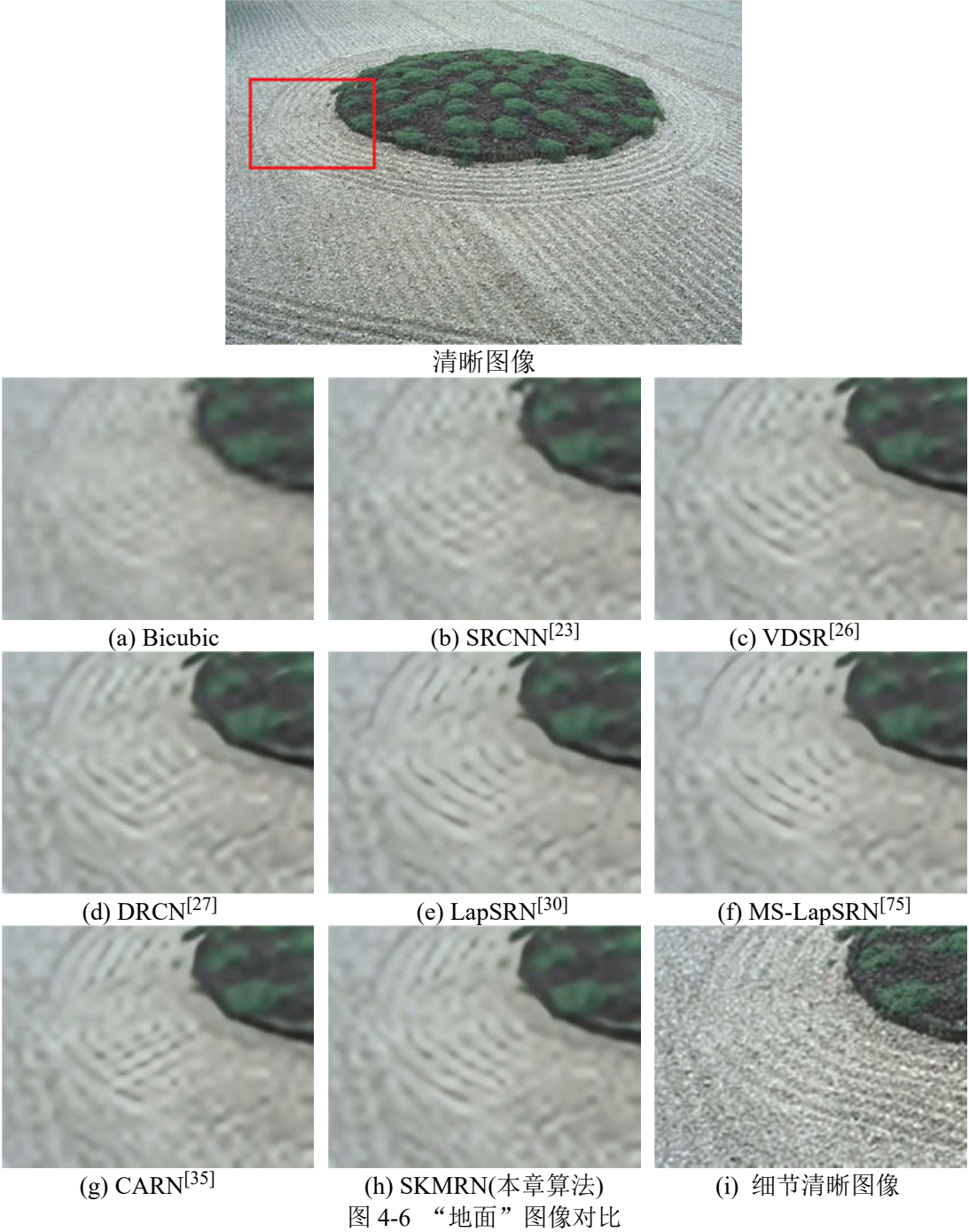


图 4-6 “地面” 图像对比



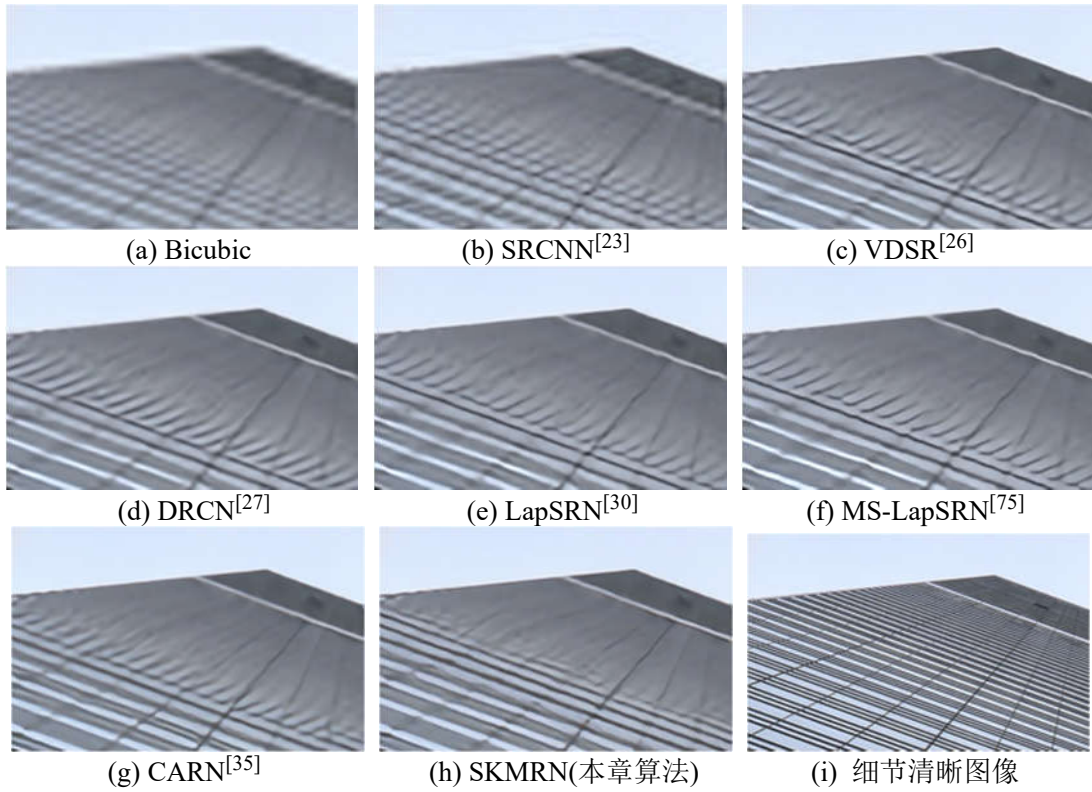


图 4-7 “建筑”图像对比

(c) 参数量分析

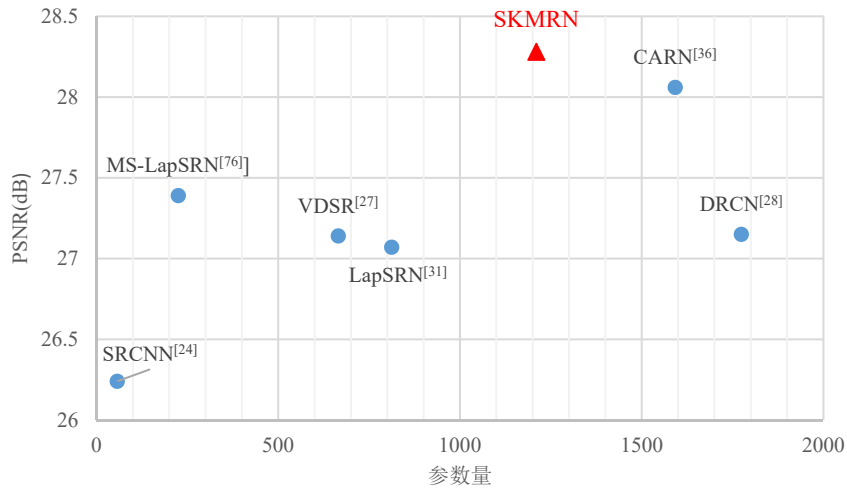


图 4-8 参数量对比

图 4-8 给出了在 Ubran100^[68]数据集上，尺度为 3 的模型参数量与模型效果对比图。从图中可以看出，SKMRN 能取得较好的超分辨率效果，并且有着较少的参数量，其参数量相较于 CARN^[35]减少了 300k，相较于 DRCN 减少了 500k，并且获得了更好的超分辨率性能，相较于 CARN^[35]，PSNR 值提升了 0.22dB。VDSR^[26]以及 LapSRN^[30]虽然有着较少的参数量，但恢复的超分辨率图像质量不够理想，而且 VDSR^[26]采用插值后图像作为输入，LapSRN^[30]采用两段式放大，

这种做法都会在高分辨率图像上进行运算，无疑会增加网络的运算量，使得网络运行时间增加。另外，MS-LapSRN^[75]和 DRCN^[27]一样都采取递归结构来构建网络，MS-LapSRN^[75]虽然有着较少的参数量，但有着很深的网络深度，在超分辨率性能上也与 SKMRN 有着明显差距。

4.3.3 消融实验

为验证本章算法的各组件的合理性与有效性，本节将介绍两个消融实验，分别是关于动态卷积核选择卷积层的多尺度分支数量选择的消融实验，以及关于权重归一化层的消融实验。

表 4-2 不同分支数量分析

	3×3	3×3+5×5	3×3+5×5+7×7
参数量	894K	1218K	2133K
PSNR (dB)	28.09	28.28	28.35

(1) 多尺度分支数量。表 4-2 给出了不同分支数量的网络的参数量和和在 Urban100^[68]数据集放大 3 倍的效果。其中第一列是只包含一路分支即卷积核尺寸为 3×3 卷积层，第二列为包含两路分支即卷积核尺寸为 3×3 卷积层和卷积核尺寸为 5×5 卷积层，第三列则是包含了三路分支的网络即卷积核尺寸为 3×3、5×5 以及 7×7 的卷积层。表 4-2 中，可以看出包含两路分支的网络相较于一路分支以微小的参数量的增加获取了效果的有效提升，PSNR 值提升了 0.19dB。而三路分支网络的参数量增加了 915K，PSNR 值却只提升了 0.07dB。因此，本章网络(SKMRN)中为获得参数量与效果的平衡，选取两路分支的卷积层。

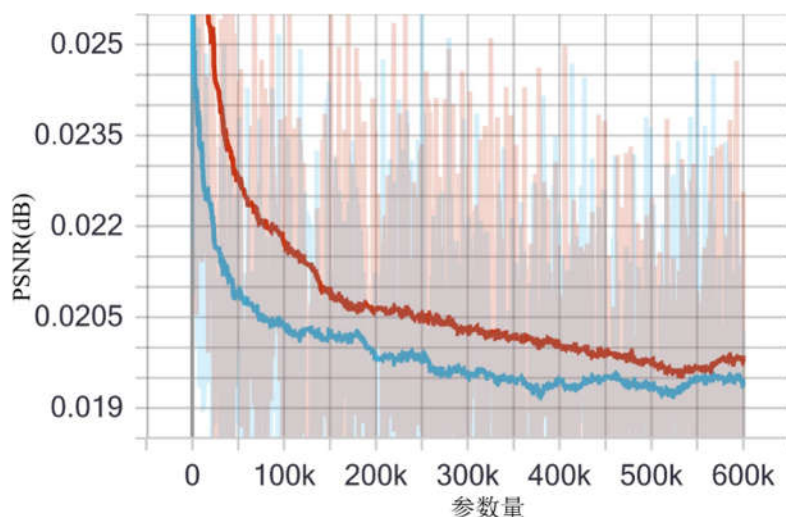


图 4-9 权重归一化层(蓝色)和无归一化层(红色)损失函数收敛曲线

(2) 权重归一化层。图 4-9 给出了采用权重归一化层与不采用权重归一化层训练时 $L1$ 损失函数的下降曲线。由图 4-9 可以看出，采用权重归一化层的网络

收敛速度比不采用权重归一化层的网络更快。除此之外，权重归一化层也会提升网络的重建效果，在 Urban100^[68]数据集上放大 3 倍时，PSNR 值相较于不采用权重归一化层的网络提升 0.22dB。因此，权重归一化层的使用不仅能加快网络收敛的速度，也能有效的提升超分辨率网络的性能。

4.4 本章小结

现有的超分辨率网络通常采用单一的卷积核尺寸，很难提取到图像的多个尺度特征。针对此问题，本章提出一种基于动态卷积核选择的多尺度超分辨率网络，网络引入动态卷积核选择卷积层实现对不同尺寸卷积核的动态选择，实现以非线性方式融合不同尺度的特征信息，对卷积层的感受野进行自适应调整；此外，基于动态卷积核选择卷积层设计了残差嵌套结构加强局部信息流通，充分利用局部的特征信息；并引入权重归一化层加速网络收敛，并实现更高的超分辨率重建质量。通过在各种放大倍数的情况下对四个基准数据集进行测试，结果表明本章网络能获得细节清晰的超分辨率图像，在主观和客观评价上都取得较大提升，并相较于其他超分辨率方法保持较少的参数量。最后，通过消融实验验证了本章中各种设计的有效性。

第5章 总结与展望

5.1 全文总结

随着信息时代的来临，数字图像成为生活中最重要的信息载体，人们对于图像质量的追求越来越高。但在现实中，常常受摄像设备与拍摄环境的限制，无法获取高分辨率图像。因此，通过研究图像超分辨率技术，实现从低分辨率图像重建出高分辨率图像成为一个重要的技术方向。图像超分辨率技术是图像处理领域的研究热点与难点，近年来，在图像传输、视频监控、医疗、遥感图像等领域中都发挥着重要的作用。本文通过调研图像超分辨率领域的研究现状发现卷积神经网络能够有效提取复杂的图像特征，完成特征映射，基于卷积神经网络的图像超分辨率方法在重建质量上明显优于传统算法，拥有广阔的应用前景。基于此，本文着重研究了基于卷积神经网络的图像超分辨率方法，首先介绍了卷积神经网络的基本原理与方法，包括基于卷积神经网络的超分辨率方法的关键技术等；然后针对超分辨率网络速度慢，参数多的问题，提出一种轻量，快速的超分辨率网络；最后针对超分辨网络对多尺度信息不够关注的缺点，提出基于动态卷积核选择的多尺度超分辨率网络。本文的主要研究工作如下：

现有的超分辨率网络虽然能获得较好的超分辨率效果，但大多数网络都存在参数量大，运行速度慢的问题，无法应用到实际的应用中。本文针对以上问题，提出一种轻量级分组卷积超分辨率网络，其首先利用分组卷积有效降低模型参数量；并以卷积核尺寸为 1×1 的卷积层为中间层构建密集连接结构，以实现特征的充分利用；最后通过注意力机制以微小的参数量增加来获取较大的性能提升。实验结果表明，该网络能有效降低网络参数量，并提高网络运行速度，同时保持着较好的超分效果。

目前的超分辨率方法缺乏对图像的多尺度信息的关注，大多数网络只采用一个大感受野或者一个小感受野来提出特征，但是单一尺度的特征提取往往不能充分利用图像特征。基于此，本文提出一种基于动态卷积核选择残差网络的多尺度图像超分辨率方法，采用动态卷积核选择卷积层，其包含多种卷积核的不同分支，并在网络中实现对不同分支进行动态选择以构成最终的卷积层输出；以动态卷积核选择卷积层为基础，构建子残差块，并通过残差嵌套机制构建多尺度残差块，最终在整体上通过多重级联的结构实现图像超分辨率重建。除此之外，网络还采

用权重归一化层加速网络训练过程并提升网络效果。实验证明,该网络能取得较好的超分辨率效果,并保持较少的参数量。

5.2 研究展望

本文分别提出两种有效的图像超分辨率算法。为解决网络参数量大,运行速度慢的问题,提出了基于分组卷积和注意力机制的轻量级超分辨率网络;针对现有超分辨率网络对图像的多尺度信息的关注不足的问题,提出基于动态卷积核选择的多尺度超分辨率网络。两种超分辨率算法分别取得较好的超分辨效果,但同时也存在一定的问题与缺陷,需要进一步研究解决,在未来,本课题的主要研究内容包括:

论文中所提轻量级分组卷积网络(LGCN)采用分组卷积和卷积核尺寸为 1×1 的卷积层来减少网络参数量,但研究表明网络中卷积核为 1×1 的卷积层参数在整体参数中占比较大,除此之外,网络的精度还有待提高。在未来,通过采用更为合适的网络结构取代卷积核为 1×1 的卷积层和分组卷积层以实现更好的精度与速度的平衡。

论文提出的多尺度图像超分辨率算法虽然能取得较好的超分辨率效果,但采用大卷积核的分支时会带来较多的参数量,针对这个问题,在未来可以采用空洞卷积层或者其他更为高效的卷积层代替卷积核尺寸大的卷积层以减少网络分支过多带来的参数量过多问题。

本文采用的超分辨率算法以 PSNR 和 SSIM 值为评价标准,研究表明这两种指标往往会使图像变得平滑,缺乏较多的高频细节。近年来,对抗生成网络取得了较快的发展,其能生成具有更好主观视觉效果图像,在未来的研究中需要更加注重对抗生成网络的应用。另外,对于本文采用的损失函数 $L1$ 函数,其功能也比较单一,生成的图像平滑程度也比较高。在未来的研究中,一种新的图像评价指标,以及更能指导网络学习的损失函数也是重点内容。

参考文献

- [1] Tian J, Jing A, and Kai Kuang Ma. A Survey on super-resolution imaging [J]. Signal, Image and Video Processing, 2011: 329–342.
- [2] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C]. International Conference on Neural Information Processing Systems, 2012:1097-1105.
- [3] Yi S, Wang X, Tang X. Deep learning face representation from predicting 10,000 classes [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2014: 1891-1898.
- [4] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation [C]. International Conference on Medical Image Computing and Computer-assisted Intervention, 2015:234-241.
- [5] Ren S, Girshick R, Girshick R, et al. Faster R-CNN: Towards real-Time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.
- [6] Park S C, Park M K, Kang M G. Super-resolution image reconstruction: a technical overview [J]. IEEE Signal Processing Magazine, 2003, 20(3):21-36.
- [7] Romano Y, Isidoro J, Milanfar P. RAISR: Rapid and accurate image super resolution [J]. IEEE Transactions on Computational Imaging, 2016, 3(1):110-125.
- [8] 朱海, 王国中, 范涛. 基于深度超分辨率重建的监控图像人脸识别[J]. 电子测量技术, 2018(16): 61-64.
- [9] 何林阳. 航空图像超分辨率重建关键技术研究[D]. 中国科学院大学, 2016.
- [10] 邢晓羊, 魏敏, 符颖. 基于特征损失的医学图像超分辨率重建[J]. 计算机工程与应用, 2018, 54(20): 207-212.
- [11] 金阳. 木材空气耦合超声图像的超分辨率重构方法[D]. 浙江农林大学, 2018.
- [12] Thévenaz P, Blu T, Unser M. Image interpolation and resampling [J]. Handbook of Medical Imaging, 2000, 87(3):393-420.
- [13] Blu T, Thévenaz P, Unser M. Linear interpolation revitalized [J]. IEEE Transactions on Image Processing, 2004, 13(5):710.
- [14] Keys R G. Cubic convolution interpolation for digital image processing [J]. IEEE Transactions on Acoust. Speech and Signal Processing, 1981: 37.
- [15] Irani M, Peleg S. Super resolution from image Sequences [C]. International

- Conference on Pattern Recognition, 1990, 90:115-120.
- [16] Stark H, Oskoui P. High-resolution image recovery from image-plane arrays, using convex projections [J]. Journal of the Optical, 1989, 6(11):1715-1726.
- [17] Schultz R, Stevenson R. Improved definition video frame enhancement [C]. Proceedings of the IEEE international Conference on Acoustics, Speech and Signal Processing, 1995:2169-2172.
- [18] Freeman W T, Jones T R, Pasztor E C. Example-based super-resolution [J]. IEEE Computer Graphics and Applications, 2002, 22(2):56-65.
- [19] Chang H, Yeung, DitYan, et al. Super-resolution through neighbor embedding [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2004: 275-282.
- [20] Yang J, Wright J, Huang T, et al. Image super-resolution as sparse representation of raw image patches [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2008: 1-8.
- [21] 王泽思. 基于插值和样例的超分辨率图像处理算法的研究[D]. 浙江师范大学, 2014.
- [22] Hayat, K. Multimedia super-resolution via deep learning: A survey [J]. Digital Signal Processing, 2018, 81:198-217.
- [23] Dong C, Loy C C, He K, et al. Image super-resolution using deep convolutional networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(2):295-307.
- [24] Dong C, Chen C L, Tang X. Accelerating the super-resolution convolutional neural network [C]. European Conference on Computer Vision. 2016:391-407.
- [25] Shi W, Caballero J, Huszár F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network [C]. IEEE Conference on Computer Vision and Pattern Recognition. 2016:1874-1883.
- [26] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016:1646-1654.
- [27] Kim J, Lee J K, Lee K M. Deeply-recursive convolutional network for image super-resolution [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016:1637-1645.
- [28] Tai Y, Yang J, Liu X. Image super-resolution via deep recursive residual network [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017:2790-2798.
- [29] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [30] Lai W S, Huang J B, Ahuja N, et al. Deep laplacian pyramid networks for fast

- and accurate super-resolution [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017:5835-5843.
- [31] Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution [C]. IEEE Computer Vision and Pattern Recognition Workshops. IEEE, 2017:1132-1140.
- [32] Zhang Y, Tian Y, Kong Y, et al. Residual dense network for image super-resolution [C]. IEEE Computer Vision and Pattern Recognition Workshops, 2018:2472-2481.
- [33] Huang G, Liu Z, Maaten L V D, et al. Densely connected convolutional networks [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017:2261-2269.
- [34] Haris M, Shakhnarovich G, Ukita N. Deep back-projection networks for super-resolution [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018:1664-1673.
- [35] Ahn N, Kang B, Sohn K A. Fast, accurate, and lightweight super-resolution with cascading residual network [C]. European Conference on Computer Vision, 2018: 252-268.
- [36] Ledig C, Theis L, Huszar F, et al. Photo-realistic single image super-resolution using a generative adversarial network [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 4681-4690.
- [37] Hui Z, Wang X, Gao X. Fast and accurate single image super-resolution via information distillation network [C]. IEEE Computer Vision and Pattern Recognition, 2018:723-731.
- [38] Shi W, Jiang F, Zhao D. Single image super-resolution with dilated convolution based multi-scale information learning inception module [C]. IEEE International Conference on Image Processing, 2017:977-981.
- [39] Li J, Fang F, Mei K, et al. Multi-scale residual network for image super-resolution [C]. European Conference on Computer Vision, 2018: 517-532.
- [40] Kohonen T. Self-organised formation of topologically correct feature map [J]. Biological Cybernetics, 1982, 43(1):59-69.
- [41] Hubel D H, Wiesel T N. Receptive fields of single neurones in the cat's striate cortex [J]. Journal of Physiology, 1959, 148(3):574.
- [42] McCulloch W S, Pitts W. A logical calculus of the idea immanent in nervous activity [J]. Bulletin of Mathematical Biology, 1943, 5(4):115-133.
- [43] Lecun Y, Bengio Y, Hinton G. Deep learning [J]. Nature, 2015, 521(7553):436.
- [44] 李彦冬, 郝宗波, 雷航. 卷积神经网络研究综述[J]. 计算机应用, 2016, 36(9): 2508-2515.
- [45] Boureau Y L, Roux N L, Bach F, et al. Ask the locals: Multi-way local pooling

- p>for image recognition [C]. IEEE International Conference on Computer Vision, 2011: 2651-2658.
- [46] Gu, J, Wang, Z, Kuen, J, Ma, L, Shahroudy, A, et al. Recent advances in convolutional neural networks [J]. Pattern Recognition, 2018, 77:354-377.
- [47] Hornik K. Multilayer feedforward networks are universal approximators [J]. Neural Networks, 1989, 2(5):359-366.
- [48] Hecht-Nielsen. Theory of the backpropagation neural network [C]. IEEE International Joint Conference on Neural Networks, 1989: 593-605.
- [49] Cauchy, A. Méthode générale pour la résolution des systemes d'équations simultanées [J]. Comptes Rendus de l'Academie des Science, 1847, 25:536-538.
- [50] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.
- [51] Polyak B T. Some methods of speeding up the convergence of iteration methods [J]. USSR Computational Mathematics and Mathematical Physics, 1964, 4(5):1-17.
- [52] Sutskever I, Martens J, Dahl G, et al. On the importance of initialization and momentum in deep learning [C]. International Conference on International Conference on Machine Learning, 2013: 2176–2184.
- [53] Duchi J, Hazan E, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization [J]. Journal of Machine Learning Research, 2011, 12(7):257-269.
- [54] Kingma, D.P. and Ba, J. Adam: A method for stochastic optimization [C]. International Conference on Learning Representations, 2015.
- [55] Zeiler M D, Krishnan D, Taylor G W, et al. Deconvolutional networks [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2010:2528-2535.
- [56] Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult [J]. IEEE Transactions on Neural Networks, 1994, 5(2): 157-166.
- [57] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks [C]. International Conference on Artificial Intelligence and Statistics, 2010: 249-256.
- [58] Tong T, Li G, Liu X, et al. Image super-resolution using dense skip connections [C]. IEEE International Conference on Computer Vision, 2017:4809-4817.
- [59] Dodgson, N. A. Quadratic interpolation for image resampling [J]. IEEE Transactions on Image Processing, 1997, 6(9):1322-1326.
- [60] Wang, Z, Bovik, A.C, Sheikh H R. Structural similarity based image quality assessment [J]. Digital Video Image Quality and Perceptual Coding, 2017: 225-242.

- [61] Zhao H, Gallo O, Frosio I, et al. Loss functions for image restoration with neural networks [J]. IEEE Transactions on Computational Imaging, 2017, 3(1):47-57.
- [62] 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.
- [63] Arbelaez P, Maire M, Fowlkes C, et al. Contour detection and hierarchical image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(5):898-916.
- [64] Timofte R, Lee K M, Wang X, et al. NTIRE 2017 challenge on single image super-resolution: methods and results [C]. IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017:1122-1131.
- [65] Bevilacqua, M, Roumy, A, Guillemot, C, et al. Low-complexity single-image super-resolution based on nonnegative neighbor embedding [C]. The British Machine Vision Conference, 2012:135.1-135.10.
- [66] Roman Zeyde, Michael Elad, Matan Protter. On single image scale-up using sparse-representations [C]. International Conference on Curves and Surfaces. Springer-Verlag, 2012:711-730.
- [67] Martin D, Fowlkes C, Tal D, et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics [C]. IEEE International Conference on Computer Vision, 2001, 2: 416-423.
- [68] Huang J B, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2015:5197-5206.
- [69] Iandola F N, Han S, Moskewicz M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. arXiv:1602.07360,2016.
- [70] Howard A G, Zhu M, Chen B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861,2017.
- [71] Zhang X, Zhou X, Lin M, et al. ShuffleNet: An extremely efficient convolutional neural network for mobile devices [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018:6848-6856.
- [72] Xie S, Girshick R, Dollar P, et al. Aggregated residual transformations for deep neural networks [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017:5987-5995.
- [73] Jie H, Li S, Albanie S, et al. Squeeze-and-Excitation networks [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018:7132-7141.
- [74] He K, Zhang X, Ren S, et al. Delving deep into rectifiers: surpassing human-level performance on imagenet classification [C]. IEEE International

- Conference on Computer Vision, 2015:1026-1034.
- [75] Lai W S, Huang J B, Ahuja N, et al. Fast and accurate image super-resolution with deep laplacian pyramid networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018:2259-2263.
- [76] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2015:1-9.
- [77] Spillmann L, Dresch-Langley B, Tseng C H. Beyond the classical receptive field: The effect of contextual stimuli [J]. Journal of Vision, 2015, 15(9): 7-7.
- [78] Li X, Wang W, Hu X, et al. Selective kernel networks [C]. IEEE Computer Vision and Pattern Recognition, 2019:510-519.
- [79] Salimans T, Kingma D P. Weight normalization: a simple reparameterization to accelerate training of deep neural networks [C]. Advances in Neural Information Processing Systems, 2016: 901-909.

发表论文和参加科研情况说明

发表的论文:

- [1] 杨爱萍, 杨炳旺, 曲畅, 等. 基于透射率融合与优化的水下图像复原[J]. 天津大学学报(自然科学与工程技术版), 2019, 52(10): 1033-1044.

发明专利:

- [1] 杨爱萍, 杨炳旺, 王金斌, 等. 一种基于深度反投影网络的高效超分辨率方法(专利申请号: CN201811637583.3)

参与的科研项目:

- [1] 面向无人驾驶汽车的恶劣天气环境下视觉计算技术, 国家自然科学基金重点基金项目, No.61632018.
- [2] 基于稀疏表示的单幅图像联合盲复原及识别, 国家自然科学基金面上项目, No. 61472274.

致 谢

时间总是转瞬即逝，转眼间两年半的硕士研究生生涯即将结束，离别之际，思绪万千。回顾两年半的时间，研一时，学习专业课程为科研打下基础，研二则是在专业领域钻研文献，努力创新，取得科研成果，研三则撰写毕业论文，总结研究生生涯，每个阶段都有着很多不同的想法，也感谢一路陪我走过的老师、同学、朋友、亲人。

首先要感谢我的导师杨爱萍副教授。自入学以来，杨爱萍老师在科研和生活上都对我有很大的帮助，在科研工作中，杨老师严谨认真、精益求精的态度深深的影响着我，当我遇到问题时，总是能给予我很多的帮助和支持，在科研中也为我提供优质的资源，在论文撰写期间，杨老师总是能提出建设性的意见，帮助我不断地修改完善。生活中，杨老师平易近人，给我们很多关爱，科研的时间带给我温暖的感觉，感谢杨老师为我们营造和谐融洽的实验室氛围。再次由衷感谢杨老师近三年来对我们的教导和无微不至的关怀，您的教诲与训诫我会一直铭记于心！另外，还要感谢冀中老师的指导，您的指导意见对我有很大的帮助。

其次，感谢实验室的同学们，感谢王南、张越、曲畅、杨苏辉、王金斌、田鑫、赵美琪师兄师姐们在学术及人生方向的帮助和指导，感谢鲁立宇、王海新对我的鼎力相助，无论生活还是科研都给我提供了很多建设性意见。感谢张兵、王前、邢金娜、刘瑾、邵明瑄师弟师妹们在生活和学习上的各种帮助和陪伴。

最重要的是要特别感谢我的父母。父母给予我物质和精神上的无私奉献，你们的关心是我最温暖的依靠，这让我心生感激但又有无限愧疚，在此由衷感谢多年来的陪伴。

