

密 级\_\_\_\_\_



**桂林电子科技大学**  
GUILIN UNIVERSITY OF ELECTRONIC TECHNOLOGY

# 硕 士 学 位 论 文

(全日制专业学位硕士)

题目 基于卷积神经网络的图像超分辨率重建研究

(英文) Research on Image Super-resolution Reconstruction

Based on Convolutional Neural Network

研 究 生 学 号: 1803303056

研 究 生 姓 名: 朱朝辉

指导教师姓名、职称: 刘建明 教授

申 请 学 位 类 别: 工程硕士

领 域: 计算机技术

论 文 答 辩 日 期: 2021 年 6 月 1 日



## 独创性（或创新性）声明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢中所罗列的内容以外，论文中不包含其他人已经发表或撰写过的研究成果；也不包含为获得桂林电子科技大学或其它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做了明确的说明并表示了谢意。

申请学位论文与资料若有不实之处，本人承担一切相关责任。

本人签名：朱朝辉 日期：2021.6.05

## 关于论文使用授权的说明

本人完全了解桂林电子科技大学有关保留和使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属桂林电子科技大学。本人保证毕业离校后，发表论文或使用论文工作成果时署名单位仍然为桂林电子科技大学。学校有权保留送交论文的复印件，允许查阅和借阅论文；学校可以公布论文的全部或部分内容，可以允许采用影印、缩印或其它复制手段保存论文。（保密的论文在解密后遵守此规定）

本学位论文属于保密在\_\_\_\_年解密后适用本授权书。

本人签名：朱朝辉 日期：2021.6.05

导师签名：刘事 日期：2021.6.05



## 摘要

图像超分辨率一直以来都是图像处理领域的一个非常热门的研究课题。在现实生活中,受到现实环境和拍照设备的影响,人们往往无法获得包含细节信息的高分辨率图像。图像超分辨率技术可以将因环境、设备、人为因素得到的低分辨率图像重建出清晰的、包含更多细节信息的高分辨率图像。随着卷积神经网络在图像超分辨率的重建工作中的蓬勃发展,图像超分辨率的重建效果越来越好。本文主要对现有的卷积神经网络图像超分辨率算法的网络模型进行改进,研究的内容主要为如下两个部分:

(1) 考虑到大多数卷积神经网络模型不能充分利用卷积层所提取的特征,而且随着网络模型深度的不断加深,网络模型的训练难度也相应增大。针对这些问题,结合密集连接和残差网络的优点,提出了一种基于多尺度残差密集块卷积神经网络,该网络模型主要由特征提取、多尺度残差密集块和多尺度亚像素卷积组成。首先,通过特征提取获取图像的浅层信息;其次通过多尺度残差密集块对提取到的浅层特征进行更深层次的提取,残差密集块采用局部密集特征融合和局部残差融合相结合的方式增加对以往卷积层的利用,并将多个残差密集块进一步进行全局特征融合,增加网络特征信息的复用性,同时在残差密集块内采用多尺度卷积来增加网络的宽度,获取图像的区域信息;然后将浅层特征信息和深层特征信息进行全局残差融合,以保证对低频信息的有效利用;最后通过多尺度亚像素卷积对图像进行重建,最终得到高质量的重建图像

(2) 考虑到大多数卷积神经网络的图像超分辨率重建方法中并没有充分利用图像的固有非局部相似特性,并且目前大多数基于卷积神经网络的图像超分辨率算法主要在加深网络深度和构建更复杂的网络模型来学习更多的图像特征,却很少利用高阶的图像特征。针对这些问题,结合图像的非局部特征和二阶统计特征的优点,提出了一种基于注意力机制与多尺度特征融合的网络结构。在特征提取过程中,将图像的非局部特征和二阶统计特征相结合组成一个注意力机制模块,以增强图像的结构信息的提取,同时采用多尺度特征模块提取不同尺度下的图像特征信息,获取更多的图像区域信息,并将两种方式提取到的图像信息进行融合,最后通过亚像素卷积对图像进行重建,得到最终的高分辨率图像。

**关键词:** 超分辨率重建; 卷积神经网络; 多尺度残差密集块; 密集连接; 残差网络; 注意力机制

## Abstract

Image super-resolution has always been a very popular research topic in the field of image processing. In real life, affected by the real environment and camera equipment, people often cannot obtain high-resolution images containing detailed information. Image super-resolution technology can reconstruct low-resolution images obtained due to environmental, equipment, and human factors into clear, high-resolution images that contain more detailed information. With the vigorous development of convolutional neural networks in the reconstruction of image super-resolution, the reconstruction effect of image super-resolution is getting better and better. This article mainly improves the network model of the existing convolutional neural network image super-resolution algorithm. The main content of the research is mainly the following two parts.

(1) Considering that most convolutional neural network models cannot make full use of the features extracted by the convolutional layer, and as the depth of the network model continues to deepen, the difficulty of training the network model increases accordingly. In response to these problems, combining the advantages of dense connections and residual networks, a multi-scale residual dense block convolutional neural network is proposed. The network model is mainly composed of feature extraction, multi-scale residual dense blocks and multi-scale sub-pixel convolution. First, the shallow information of the image is obtained through feature extraction; secondly, the extracted shallow features are extracted deeper through the multi-scale residual dense block. The residual dense block uses a combination of local dense feature fusion and local residual fusion. Ways to increase the use of the previous convolutional layer, and further perform global feature fusion of multiple residual dense blocks to increase the reusability of network feature information. At the same time, multi-scale convolution is used in the residual dense blocks to increase the network's Width, obtain the regional information of the image; then perform global residual fusion of the shallow feature information and the deep feature information to ensure the effective use of low-frequency information; finally, the image is reconstructed through multi-scale sub-pixel convolution, and finally high quality is obtained Reconstructed image

(2) Considering that most image super-resolution reconstruction methods based on convolutional neural networks do not make full use of the inherent non-local similarity characteristics of images, and most of the current image super-resolution algorithms based on convolutional neural networks are mainly deepening Network depth and building more complex network models to learn more image features, but seldom use high-level image

features. In response to these problems, combining the advantages of non-local features and second-order statistical features of images, a network structure based on the fusion of attention mechanism and multi-scale features is proposed. In the feature extraction process, the non-local features of the image and the second-order statistical features are combined to form an attention mechanism module to enhance the extraction of the structural information of the image. At the same time, the multi-scale feature module is used to extract image feature information at different scales. Get more image area information, and merge the image information extracted by the two methods, and finally reconstruct the image through sub-pixel convolution to obtain the final high-resolution image.

**Keywords:** Super-resolution reconstruction; Convolutional neural network; Multi-scale dense residual block; Dense connection; Residual network; Attention mechanism

# 目录

摘要 .....	I
Abstract.....	II
第一章 绪论 .....	1
§1.1 图像超分辨率重建的研究背景与意义 .....	1
§1.2 图像超分辨率重建研究现状 .....	2
§1.2.1 基于插值的方法 .....	2
§1.2.2 基于重建的方法 .....	2
§1.2.3 基于学习的方法 .....	3
§1.3 本文的主要工作内容和章节安排 .....	5
第二章 卷积神经相关理论 .....	6
§2.1 卷积神经网络 .....	6
§2.1.1 卷积层 .....	6
§2.1.2 非线性激活层 .....	7
§2.1.3 反向传播算法 .....	10
§2.2 残差网络和密集连接网络 .....	10
§2.2.1 残差网络 .....	10
§2.2.2 密集连接网络 .....	11
§2.3 非局部特征和注意力机制 .....	13
§2.3.1 非局部特征 .....	13
§2.3.2 注意力机制 .....	13
§2.4 亚像素卷积 .....	13
§2.5 神经网络优化算法 .....	14
§2.6 图像质量评价标准 .....	16
§2.7 本章小结 .....	18
第三章 基于多尺度残差密集块的图像超分辨率重建算法 .....	19
§3.1 引言 .....	19
§3.2 多尺度残差密集融合网络 .....	19
§3.2.1 浅层特征提取 .....	20
§3.2.2 多尺度残差密集融合 .....	20
§3.2.3 全局特征融合 .....	22



§3.2.4 多尺度亚像素卷积 .....	22
§3.3 实验结果与分析 .....	23
§3.3.1 参数设置 .....	23
§3.3.2 损失函数选择 .....	23
§3.3.3 消融实验分析 .....	24
§3.3.4 多尺度亚像素卷积分析 .....	25
§3.3.5 重建结果对比与分析 .....	26
§3.4 本章小结 .....	28
第四章 基于注意力机制与多尺度特征融合的图像超分辨率重建算法 .....	29
§4.1 引言 .....	29
§4.2 注意力机制与多尺度特征融合网络 .....	30
§4.3 浅层特征提取模块 .....	30
§4.4 注意力机制与多尺度信息融合模块 .....	30
§4.4.1 多尺度特征融合模块 .....	31
§4.4.2 注意力机制融合模块 .....	32
§4.5 图像重建模块 .....	35
§4.6 实验结果与分析 .....	35
§4.6.1 亚像素卷积对网络性能的提升 .....	35
§4.6.2 消融实验分析 .....	36
§4.6.3 重建结构与对比分析 .....	36
§4.7 本章小结 .....	39
第五章 总结与展望 .....	40
§5.1 总结 .....	40
§5.2 展望 .....	40
参考文献 .....	42
致谢 .....	46
作者在攻读硕士期间的主要研究成果 .....	47



## 第一章 绪论

本章首先简单介绍一些关于图像超分辨率重建技术的基础知识，并对其在人们的生活、生产中的应用以及研究的重要意义进行了说明。其次对图像超分辨率的研究现状尤其是基于卷积神经网络的重建算法的研究进展做一个简单的介绍。

### § 1.1 图像超分辨率重建的研究背景与意义

图像是人们获取外界信息的主要来源之一，与人们的生活紧密相关。图像分辨率则是图像包含信息量的重要衡量指标，图像分辨率越高则说明图像存储更丰富的细节信息，具有更强的信息表达能力。图像的获取通常是通过相机、摄像头等成像电子设备，然而现实生活中往往因为这些成像设备的条件限制和成像环境因素的影响，人们很难得到高分辨率的图像，从而导致获取的这些图像不能满足我们的需求。为了解决上述问题，我们就需要从软件方面着手来获取高分辨率的图像。图像分辨率重建技术已在人们生活中的各行各业得到广泛的应用。下面介绍几个典型的应用。

#### （1）医学诊断领域

在医院中，医生通过医疗器材的影像设备拍摄来获取病人的诊断图像进而来辅助医生对病人的诊断和治疗，但是由于医学器材的设备成本及其现有技术的制约，仅通过医疗器材的影像设备很难获取高分辨率的医学图像。因此运用图像超分辨率技术，将现实中的低分辨率图像经过处理得到高分辨率图像可以使医生获得包含更丰富医学信息的医学图像，从而帮助医生对病人做出更加准确的诊断和治疗。

#### （2）视频监控领域

随着人工智能的发展，人脸识别逐渐成为实际应用中常见的技术，特别是在监控领域，但是现实生活中由于受到外界天气因素以及监控硬件设备的制约，视频监控采集到图像信息往往分辨率较低，图像信息模糊，无法辨识图形的细节信息。因此运用图像超分辨率技术，将得到的低分辨率模糊图像进行处理得到清晰的高分辨率图像，警务人员就可以获取到犯罪嫌疑人的面部特征的细节信息，这极大的方便了警务人员对犯罪案件的侦破。

#### （3）遥感成像领域

宇宙中的卫星拍摄图像的过程容易受到大气层中气流等环境因素的影响，会使得拍摄到的图形出现模糊、不清晰等情况，这样会使得探测的目标的难度增大。引入超分辨率重建技术之后，可以将模糊、不清晰图像重建出具有丰富细节信息的高分辨率

图像，这样需要检测的目标就会更加容易检测和识别。

由以上可知，图像超分辨率技术在实际生活中应用非常广泛，因此对这项技术进行更深入的研究就显得非常重要。

## § 1.2 图像超分辨率重建研究现状

本章主要从算法角度来对目前应用广泛的重建技术进行介绍。在算法方面基本可以分为三类：基于插值的方法<sup>[1]</sup>、基于重建的方法<sup>[2]</sup>和基于学习的方法<sup>[3]</sup>。

### § 1.2.1 基于插值的方法

基于插值的方法主要是利用现有低分辨率图像像素点周围的信息，在中心点像素的四周通过插入特定函数计算得来像素进行重建。由于这种方式是利用数学公式来计算得到插入点像素的值，所以这种方法的效率是比较高的，但也存在不少问题，比如计算得来的像素与中心点的像素值过于平滑，得到的图像的边缘细节以及高频纹理细节部分的效果较差。插值法常见的算法主要有最近邻插值算法<sup>[4]</sup>、双线性插值算法<sup>[5]</sup>以及双三次插值算法<sup>[6-8]</sup>

最近邻插值算法并不像插值法那样对图像的像素进行数值计算，而是将距离待求像素最近的像素值赋值给所求像素，这种方式虽然简单计算量小，但是这样做会导致图像的灰度值不连续等问题，重建出来的图像的边缘部分会出现明显的锯齿痕迹。

双线性插值算法是对线性插值算法的改进得来的，线性插值是利用图像平面两个像素点来确定需要插入像素点的位置，而双线性插值是对两个方向上进行扩展，利用平面图像临近的四个像素点来确定插入像素点的位置，并且这临近的四个像素点不在同一直线上。这种方法的计算比最近邻插值算法复杂，也避免了其导致的图像灰度值不连续等问题，不过这种方法会造成重建图像高频信息的丢失，得到的重建图形边缘轮廓会较为模糊。

双三次插值算法可以称得上是双线性插值算法的升级改造版本，它是利用待插入像素点四周的像素值进行三次插值计算的来的。这种方法不仅考虑到了临近像素灰度值的影响，并且将附近像素值之间的变化率计算在内，因而这方法的效果比前两种更好，但是因为参与计算的像素点数量更多，计算量也更大更复杂。

### § 1.2.2 基于重建的方法

基于重建的算法主要是利用数学方法和逆向思维对低分辨率图像进行转化、融合

进而得到更易求解的高分辨率图像。其中比较代表性的算法有迭代反投影法<sup>[9]</sup>凸集投影法<sup>[10]</sup>和最大后验概率估计法<sup>[11]</sup>等。

迭代反投影法首先通过运用插值的方法将低分辨率图像放大至高分辨率图像大小，并将得到的图像作为模拟低分辨率图像，再将模拟的低分辨率图像和原来的低分辨率图像作差，将差值映射到空间域不断的进行迭代，直到其达到设置的迭代次数最大值为止。迭代反射投影算法的思路比较简单直观，求解速度也相对较快，但是由于没有充分图像先验知识，得到的高分辨率图像也不一定是唯一最优的。

凸集投影法是一种充分利用了集合论的超分辨率重建算法，它将图像的先验知识引入进去，并将高分辨率图像所拥有的一系列约束条件用凸集的形式表现出来，将先验知识化成集合的约束。然后利用集合相交的理论获得各个凸集的交集，得到的集合就是高分辨率图像所属的集合。最后在这个集合内用迭代反投影法对图像进行高分辨率重建。这个算法是对迭代反投影法的改进，充分利用了图像的先验知识，但其同样无法避免重建结果不唯一的情况。

最大后验概率估计算法是一种概率算法，它将高分辨率图形和低分辨率图像看作两个不同的随机过程并将超分辨率重建问题看作是数学求最优解的问题。这种方法不仅充分利用了图像的先验知识，并且通过数学问题求得的最优解保证了得到的重建结果是唯一的，同时在重建图像的高频部分也能取得较好的效果。但是这种方法计算量较大。

### § 1.2.3 基于学习的方法

基于学习的方法一般是利用卷积神经网络学习低分辨率图像和高分辨率图像之间的特征变换和非线性映射，建立起低分辨率图像和高分辨率图像之间的映射关系。随着近年来卷积神经网络的蓬勃发展，目前这种方法已经成为图像超分辨率算法的主要方法，取得了比其他传统方法都要好的重建效果。通常，卷积神经网络在图像超分辨率重建算法中主要有在网络体系结构、损失函数、学习原则和策略等方面有较大的区别和不同。

目前，网络体系结构设计逐渐成为深度学习的重要方向。Dong 等人提出的 SRCNN<sup>[12]</sup>是卷积神经网络在图像超分辨率重建方面的首次应用，在当时取得了比所有传统方法都要好的重建效果，从此奠定了卷积神经网络在图像超分辨率领域新的统治地位。后来 Dong 等人又在此基础上改进提出了 FSRCNN<sup>[13]</sup>，在这个网络模型中，首次运用了反卷积<sup>[14]</sup>，因此不需要在网络外部用插值等方法对图像进行尺寸放大，并且在网络非线性映射部分用小尺寸的卷积层，这样就大大减少了计算量，模型的训练速度也提升了不少。但是反卷积在采进行相邻插值计算时，上采样会将特征图像的像

素点重复计算多次,会影响图像的重建效果。随后 Shi 等人提出了 ESPCN<sup>[46]</sup>网络,引入了亚像素卷积,它是通过在最后扩展特征图的通道数来提高图像的分辨率而不是像反卷积一样通过放大特征图来达到提高图像分辨率的目的。因此仅仅需要较小的卷积核就能得到较好的重建效果。而后何凯明等人提出 ResNet<sup>[15]</sup>网络结构,开启了残差学习在图像超分辨率领域的广泛应用。残差学习仅仅通过是学习低分辨率图像和高分辨率图像之间的残差部分信息,省去了其相似部分的学习过程,避免了大量的重复性工作,减少了模型的训练难度并提高了学习能力。Kim 等提出的 VDSR<sup>[16]</sup>就是全局残差学习的代表,它由 20 层的 VGG<sup>[17]</sup>网路组成,VDSR 在最后一个卷积层使用残差学习,大大提升了网络的收敛速度。Kim 等人后来又提出了和 VDSR 类似的网络模型 DRCN<sup>[18]</sup>,这两个模型在处理非线性映射部分非常类似,不同的是 DRCN 在非线形映射部分将同一个卷积层重复应用多次,以达到减少参数的目的,并且为采用了多监督的策略来加速网络模型的训练。类似的模型如 Zhang 等提出的 RCAN<sup>[19]</sup>也采用多监督策略,利用多路径来提取图像不同方面的特征,这些路径可以相互交叉,这样大大增强了网络模型的学习能力。相似的还有 DSRN<sup>[20]</sup>利用两条路径分别在低维和高维空间提取图形的不同方面的特征,并进行信息的交换来提升学习能力。Huang 等人提出了一种基于一种跳跃连接的 DenseNet<sup>[21]</sup>的网络结构,这种网路模型可以让每个卷积层都与前面的所有卷积层连接,这样可以充分利用每层卷积提取到的特征信息,支持特征信息的重复使用,从而达到信息的最大利用。密集连接不仅能有助于减轻梯度消失等问题,还能增强信息传播并促进图像特征重用,最终取得的效果也较以前的模型有较大提升。Tong 等人在 DenseNet 网络模型的基础上进行改进,提出了 SRDenseNet<sup>[22]</sup>网络模型,它将密集块中每一层的特征图都输入后面的所有层,使所有的特征图都串联起来,取得了较好的实验效果。Tai 等人提出的 MemNet<sup>[23]</sup>网络使用残差单元递归的替换密集块中的卷积核,并在不同的块之间添加密集连接,这样可以将前端的信息更好的传递到后端,减少图像信息的丢失。Zhang 等人提出的 RDN<sup>[24]</sup>网络也是用了类似的网路结构,基本单元通过密集块紧密连接在一起,在密集块内进行残差学习,通过密集连接和残差学习融合了前面所有块中的特征信息。

在卷积神经网络中,损失函数有着举足轻重的作用,它用于重建误差指导修正模型。在图像超分辨重建过程中,损失函数可以测量两个图像之间的像素差异。主要包括 L1 损失函数<sup>[25]</sup>(平均绝对误差)和 L2 损失函数<sup>[26-27]</sup>,实验效果中运用 L1 损失函数的模型比运用 L2 损失函数的模型具有更高的收敛性和性能。后来,人们将内容损失<sup>[28]</sup>引入进来对图像的感知质量进行评估,它通过预先训练的分类网络模型来对图像之间的语义之间的差异进行评测。由于重建的图像与目标图像具有相同的纹理等样式,为了加于区分引进了纹理损失<sup>[29]</sup>,纹理损失的统计量是通过计算在纹理变化的平均值得来的,然而确定图像块的大小进行匹配纹理完全是凭借经验,图像的大小合理

的与否则影响图形纹理区域伪影的出现。

在学习卷积神经网络模型研究中，为了加速深层网络的训练，Sergey<sup>[30]</sup>等人提出了 Batch Normalization(BN)，它可以降低网络上一层参数变化之后导致后续网络结点不停的去调整数据分布，即降低内部自身数据的协变量偏移。由于 BN 层校准了数据的中间分布特征，减轻了消失的梯度，可以使模型采用更高的学习率，加快了网络的训练和收敛速度，防止模型出现梯度消失和梯度爆炸。但 Lim<sup>[31]</sup>等人认为 BN 层会造成图像丢失其缩放比例信息，导致网路模型的灵活性降低，提出了删除 BN 层来节约内存成本来开发更大的网络模型。

### § 1.3 本文的主要工作内容和章节安排

通过对图像超分辨率重建技术现状的分析，发现目前基于卷积神经网络的图像超分辨率重建算法比先前的传统重建算法具有很大的优势和效果。因此本文利用卷积神经网络的强大学习能力构建了两类图像超分辨率重建算法。主要内容如下：

(1) 绪论：简单介绍了本文课题的在实际生活中的应用价值，通过分析现有的关于图像超分辨率重建领域的诸多算法，大致介绍了这项技术的研究现状。

(2) 卷积神经理论的简单介绍：这一章主要是对本文所用到的卷积神经相关知识进行介绍，为后续的内容介绍奠定理论知识基础。

(3) 针对多数基于卷积神经网络的图像超分辨率重建方法不能充分利用所有卷积层提取特征的问题，构建了一种基于多尺度残差密集块的图像超分辨率重建算法。该算法通过残差网络与密集连接构建的多尺度残差密集块来提升图像特征复用能力，并提出多尺度的亚像素卷积的重建方法，提升最终重建效果。

(4) 针对多数基于卷积神经网络的图像超分辨率重建方法中并没有充分利用图像的固有非局部相似特性和高阶的图像特征构建了基于注意力机制与多尺度特征融合的图像超分辨率图像重建的算法。该算法通过注意力机制对图像的非局部特征和二阶统计特征进行融合，通过多尺度特征提取模块提取不同尺度下的图像特征信息，最后用亚像素卷积实现图像的重建。

(5) 总结与展望：对本文的主要研究内容和研究成果进行概括和分析，对文章存在的不足进行总结，并对未来的研究重点进行了展望。

## 第二章 卷积神经相关理论

本章对课题中运用到的深度学习理论进行了阐述,为后面的基于深度学习的图像超分辨重建实验打下理论知识基础。

### § 2.1 卷积神经网络

卷积神经网络<sup>[32]</sup> (ConvNet/CNN) 是一种常用的深度学习算法。它是一种根据生物神经行为衍生出来的一种的算法模型。ConvNet 的架构类似于人脑神经元的连接模式。卷积神经网络可以将输入的图像的不同部分分配各自的重要性(可学习的权重和偏差),并且彼此之间能够相互区分开来。在原始的神经网络中,过滤器是手工设计的,但在卷积神经网络中,经过足够多的训练 ConvNet 能够通过学习得到这些过滤器,这比原来的预处理要简单许多。

#### § 2.1.1 卷积层

卷积层是 ConvNet 的最关键的组成部分。卷积层一般包含一个或者多个用来学习参数的过滤器(卷积核),每个过滤器的高度和宽度都小于输入层的图像的高度和宽度,通过将每个过滤器与输入层进行卷积,来计算由神经元组成的激活图。使用卷积层是为了提取输入的图像的特征,按照惯例,一般前面的卷积层负责提取图像的低级特征,如边缘、颜色、渐变方向等;后面的卷积用于获取高级特征,从而使我们通过卷积神经网络得到图像的全面特征,来完成我们的计算机视觉任务。卷积层通常运用的是互相关的运算方式,这种运算方式是从数据的左上方开始计算,按照从左往右,从上往下的顺序,对输入的数据进行全方位的计算。当卷积的滑动窗口滑动到某一位置时,窗口中的图像子矩阵和过滤器中的权重矩阵进行乘积运算,并将得到的结果与每一个偏置相加,最终得到数据即为输出矩阵中数值。由于卷积操作会改变原有图像的尺寸大小,因此每次卷积前都会对图像的边缘进行填充以此来使得到的卷积后的图像尺寸不会变化。图像的卷积操作如图 2-1 所示:



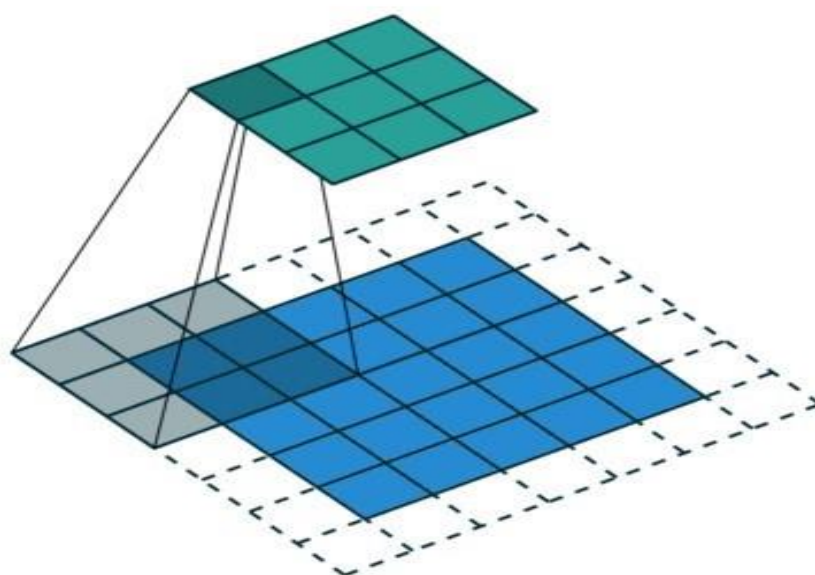


图 2-1 卷积操作

在进行卷积操作时，整个图层对应的是同一个卷积核，并且卷积核的尺寸通常都比对应图像的尺寸要小的多，因此在更新卷积核的权重时，只需要更新很小一部分的权值，这样即使随着网络结构的加深，整个网络的参数也不会有很大程度的提升，这也是卷积神经网络的一个很大的优点，可以实现权值共享。

### § 2.1.2 非线性激活层

ConvNet 中的非线性激活函数层是由激活函数构成的，它将非神经网络中节点之间的输入信号和输出信号通过非线性映射来完成转换。在一个神经网络中，将输入的  $X$  和对应的卷积核参数  $W$  进行乘积求和，然后将激活函数应用到求和的结果，将激活函数作用之后的值作为下一层的输入。激活函数可以使网络模型具有更强的表现形式，使其能够学习更加复杂数据形式，可以获取任意输入和输出之间的函数映射关系。现阶段常用的激活函数重要有三种，分别是 Sigmoid 函数<sup>[33]</sup>、双曲正切函数（tanh）<sup>[34]</sup>以及修正线性单元<sup>[35]</sup>（Rectified Linear Linear Unit Relu）。

#### （1）Sigmoid 函数

Sigmoid 函数可用  $f(x) = 1 / (1 + \exp(-x))$  来表示，他的取值范围是 0 到 1 之间，它在坐标轴上的形状类似 S 型曲线。Sigmoid 函数虽然简单易理解，但是也存在诸多缺点：首先是 Sigmoid 函数会存在梯度消失的问题，当输入非常大或者非常小的时候神经元的梯度是接近 0 的，这将会导致反向传播更新权重时权重系数过小网络模型训练困难。其次是它的输出不是以原点为中心，这将会导致梯度下降以及权重更新出现 Z 字型下降等问题，并且会在反向传播进行参数更新优化时变得困难，网络收敛速度

也会相对缓慢。Sigmoid 函数图像如下图 2-2 所示：

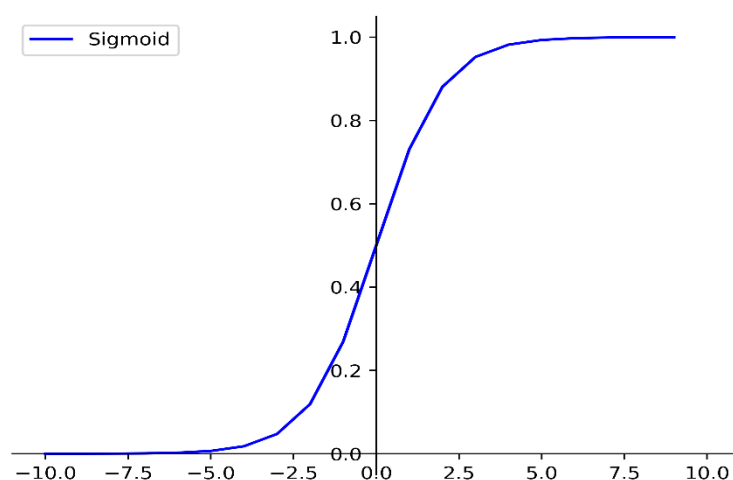


图 2-2 Sigmoid 激活函数

## (2) 双曲正切函数 (Tanh)

Tanh 函数可用  $f(x) = (1 - \exp(-2x)) / (1 + \exp(-2x))$  公式表示。这个函数是由 Sigmoid 函数推导变化得来的，它的取值范围在 -1 到 1 之间。它修正了 Sigmoid 函数不以原点为中心的问题，相比 Sigmoid 函数来说更容易进行优化，收敛速度也提升了许多，但是仍然没有解决梯度消失的问题。Tanh 函数的函数图像如下图 2-3 所示：

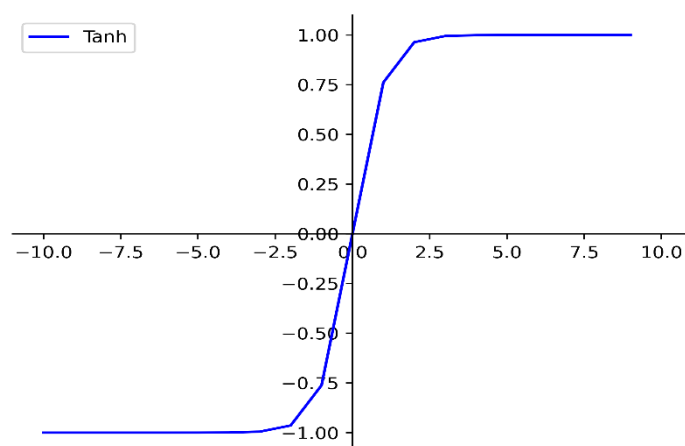


图 2-3 Tanh 激活函数

## (3) 修正线性单元

ReLU 修正线性单元是线性分段函数，函数表达式为  $f(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases}$ 。Alex

Krizhevsky 等人证明 ReLU 的收敛速度比 Tanh 函数提升 6 倍左右。ReLU 函数在区间为正值时，函数的线性表达式为  $y=x$ ，因此网络模型的局部梯度在  $x>0$  时永远为 1，在  $x<0$  时永远为 0，这样就解决了上述两个函数梯度消失的缺点。但是 ReLU 函数也存在局限性，一方面，它只能在隐藏层发挥作用，不能同时作用在输入层和输出层；另一方面函数的输出值为 0，会导致权重不会在被激活函数所更新，这样会导致部分神经元死亡，无法进入整个网络的训练，这部分无法训练的区域称为“死亡区”。因此在实验中设置学习率的时候要大小适当，避免出现这种情况的发生。ReLU 相关激活函数如下图 2-4 所示：

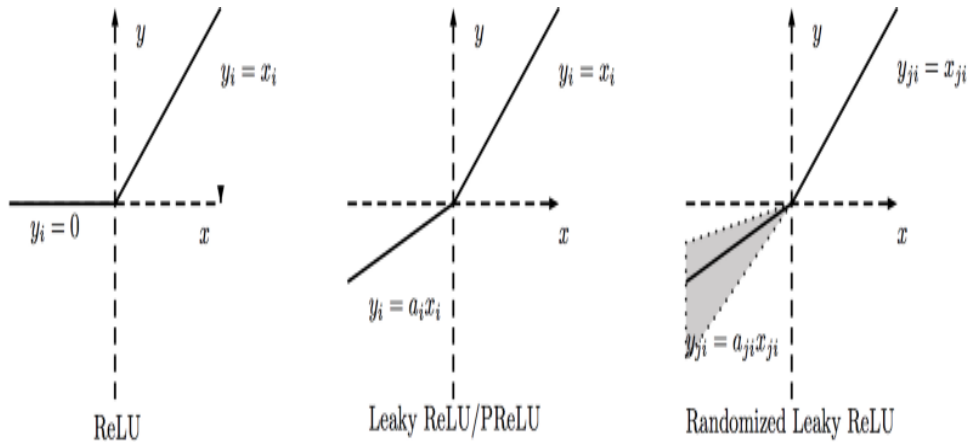


图 2-4 ReLU 相关激活函数

图 2-4 中间的是 Leaky ReLU<sup>[36]</sup>和 PreLU<sup>[37]</sup>，Leaky ReLU 函数表达式为

$$f(x) = \begin{cases} ax & x < 0 \\ x & x \geq 0 \end{cases} \quad (2-1)$$

它主要是在神经元死亡方面进行改善，解决了神经元死亡的问题，此处的  $a$  是一个较小的常数，这样就可以使得负半轴的数值不是零，保留了其对应的数值信息。PreLU 和 Leaky ReLU 很类似，它把负区间上的斜率当做神经元中的参数进行训练，通过反向传播可以得到这个参数，得到效果要比 Leaky ReLU 好。图 2-4 最右边的是 Randomized Leaky ReLU<sup>[38]</sup>，函数表达式为

$$y_{j,i} = \begin{cases} x_{j,i} & x_{j,i} \geq 0 \\ \alpha_{j,i} x_{j,i} & x_{j,i} < 0 \end{cases} \quad (2-2)$$

它可以称的上是 Leaky ReLU 的 random 版本( $\alpha$  是随机的)，其中  $\alpha$  是一个服从高斯分布的  $U(1,u)$  随机数值，并且在网络训练过程中不断的进行修正。

### § 2.1.3 反向传播算法

反向传播算法<sup>[39]</sup>是用来训练神经网络模型最常用且最有效的方法，它解决了神经网络模型以前诸多无法解决的问题。卷积神经网络使深度学习发掘出了巨大的潜力，更重要的是为深度网络量身定做的很多优秀算法，使深度学习产生了强大的功能。反向传播算法是一种在深度模型结构中相对各个参数求取损失函数的算法。它是一种利用链式法则，从输出层开始，反方向求取各个卷积层的反传误差，从后往前使用相应的优化算法更新每层神经元的参数，使得在整个模型迭代训练期间不断缩小误差值，最终使损失函数完成收敛，得到接近真实值的输出。

## § 2.2 残差网络和密集连接网络

### § 2.2.1 残差网络

通常来讲，网络模型越复杂，层次越深图像的重建效果就会越好，但随着网络模型的加大也会产生例如梯度爆炸、梯度消失等问题，难以训练出好的网络模型。神经网络中的一层通常可以看作是  $y=H(x)$ ，残差网络的残差块可以表示为  $H(x)=F(x)+x$  即为  $F(x)=H(x)-x$ ，在单位映射中  $y=x$  是观测值， $H(x)$  是预测值，因此  $F(x)$  便对应残差叫做残差网络。这样在网络训练过程中，输入值保持不变，恒等映射的值与网络学习值相加，这样网络训练过程中学习的参数就大大减少，同时保证了深层网络的有效性。残差单元可用如下公式表示：

$$Y = F(x, \{W_i\}) + W_s x \quad (2-3)$$

其中  $F(x, \{W_i\})$  表示输入  $x$  经过一个网络层  $i$  的输出结果， $W_s x$  表示前一个的输入  $x$  经过参数变换后的到和  $F(x, \{W_i\})$  和前面部分具有相同的维度，因此才能够进行相加。残差单元如图 2-5 所示：

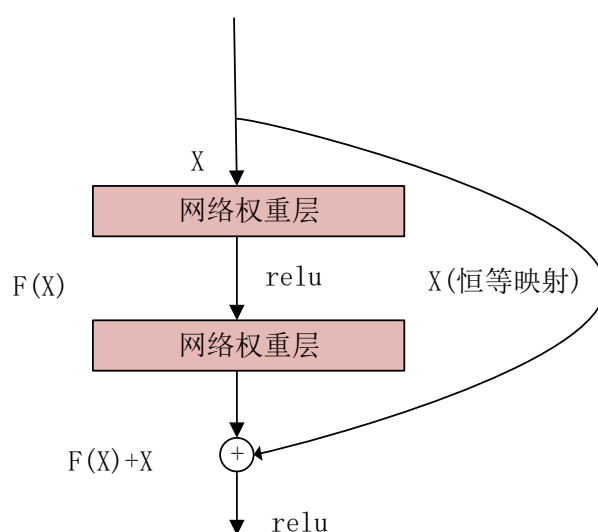


图 2-5 残差单元

残差网络一般遵循 VGG 网络中  $3 \times 3$  的卷积层设计。一个残差块由两个  $3 \times 3$  卷积组成，每个卷积都有相同数量的输出通道。残差网络的计算要跳过这两个卷积之后的 BN 层和 ReLU 激活函数，并将输入添加到 ReLU 激活函数之前。这种设计需要进入残差单元的特征图和经过两个  $3 \times 3$  卷积之后的特征图保持一致，如果前后两个特征图维度不一致，就需要引入额外的  $1 \times 1$  卷积来保证两个特征图维度的统一。

### § 2.2.2 密集连接网络

在卷积神经网络中，输入的图像通过多次卷积来获取高级特征。在 ResNet 中利用恒等映射来降低网络的计算量促进梯度的传播，并通过逐通道相加的方式来获取图像信息，而 DenseNet 却是通过特征图通道堆叠的方式来获取图像的信息，这种结构的设计目的是考虑到只是单纯的串联卷积核会导致已获取的信息出现丢失的情况，因此便借鉴了 ResNet 进行卷积层之间的跳跃连接，以此来使得信息的有效传递。但 DenseNet 并不是在连接点处对特征图进行求和，而是将特征图并联在一起，这样可以保证网络中的每一层卷积都能获取先前所有卷积层提取到的信息，将自身的特征图间接的传递给了所有的后续层，减少了信息的丢失，从而获取更加丰富的特征表示。由于每层都能接收到所有先前层的特征图，这样一来网络结构就会更薄、更紧凑，网络中的通道数也更少。与 ResNet 类似，DenseNet 的跳跃连接并联操作并不会使网络中的梯度随着网络的加深而消失，因为卷积层的跳跃连接并联操作可以使得每个卷积层的信息得到充分利用，使用相对较浅的网络就能获得大量的特征图信息，这在一定程度上减轻了模型的深度参数的数量，并且提高了网络模型的训练效率。图 2-6 展示的

是 DenseNet 的网络结构如下所示

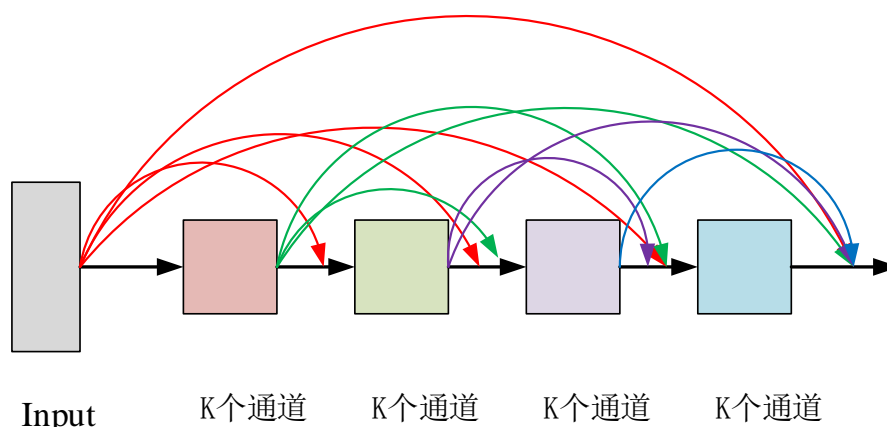


图 2-6 DenseNet 中的密集块

左边是输入层，剩下的是 4 个隐藏层，数据从输入层输入数据，通过隐藏层输出特征图。隐藏层产生的特征图数量定义为生长率(Growth Rate)，用字母  $K$  表示。上图中  $K=4$  则每个隐藏层产生 4 个特征图，在特征图达到过渡层的时候就会达到 16 个。DenseNet 的特征图就是通过其中的隐藏层来进行增长的。这几个隐藏层共同构成了一个密集块 (Dense Block)，若干个密集块则构成了一个简单的密集连接神经网络，然而通过密集连接后，特征图的数量会逐渐增多，因此每个密集块的过渡层中会对特征图进行适当的降维处理。一般会设置一个压缩系数  $\rho$  ( $\rho$  的取值范围为  $(0,1]$ )，假设到达过度层的特征图数量为  $W$ ，那么进行过压缩之后就会有  $\rho W$  个通道数的特征图。由于要对不同层的特征图进行合并，所以要同一个密集块中的特征图尺寸大小相同，但是不同密集块之间，特征图的大小可以不同，为了保证最终的不同密集块之间的特征图能够并联在一起，需要在密集块的过渡区设置池化层，池化层负责协调不同密集块之间特征图的融合，保证后续特征图的并联。完整的 DenseNet 网络结构如下图 2-7 所示：

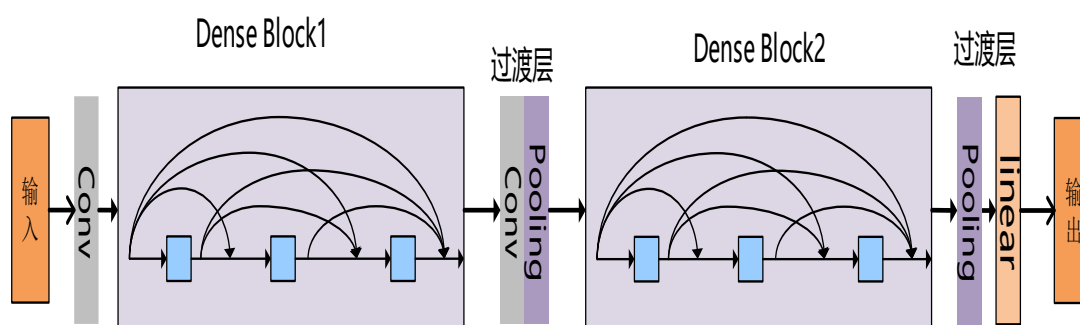


图 2-7 DenseNet 网络结构

## § 2.3 非局部特征和注意力机制

### § 2.3.1 非局部特征

非局部特征在传统的图像重建领域取得了很好的重建效果，它的基本思想是通过提取图像中相似的图像块，并利用这些相似的图像块之间的结构信息来对图像进行重建。Buades 等人<sup>[40]</sup>提出了一种基于图像所有像素的非局部平均操作来提升图像去噪能力的方法，该方法主要是在整个图像内提取相似的图像块，并图像块之间结构的相似性，赋予图像块不同大小的权值，通过加权平均得到去噪图像。在图像处理领域，捕捉图像的长距离依赖是至关重要的一个问题，一般想要捕捉长距离依赖，通常的做法是堆叠卷积层，随着卷积层的加深，感受野就越大，就可以把原来非相邻的图像像素点纳入到一个整体考虑，获取图像信息分布的广度也就越高，但这种堆叠的卷积层意味着更多的参数，更复杂的数据关系，使得模型训练更加的困难，得到的效果并不理想。Wang<sup>[41]</sup>等人提出了一种非局部网络，将非局部操作作为一种简洁高效的组件，用于捕捉神经网络中较远像素点之间的依赖关系。它与递归操作和卷积层的叠加不同，它可以通过计算任意两个位置之间的交互直接捕捉远程依赖关系，不仅仅局限于相邻点，摒弃了距离的概念。因此，非局部思想是一种经典的计算机视觉方法，目前在神经网络学习中有着广泛的应用。

### § 2.3.2 注意力机制

近年来，注意力机制在神经网络中被广泛应用，Hu 等人的提出的 SENet<sup>[42]</sup>通过学习特征通道之间相关性，通过模型的训练来自动获取每个特征通道的权重参数，并通过网络的全局损失函数来自适应的矫正特征通道间的相应强度。该网络模型在图像分类工作中取得了良好的效果。Woo 等人随后提出的 CBAM<sup>[43]</sup>在 SENet 的基础上增加了空间注意力机制，将空间和通道注意力机制结合起来，取得了比 SENet 更好的效果。Zhang<sup>[44]</sup>等人将注意力机制运用到图像超分辨率重建任务中，通过注意力机制来区分对待不同的特征通道，以此来提升网络的特征表达能力。然而这些网络模型仅仅注意到了图像的一阶特征，却忽略了图像的高阶特征。因此 Dai<sup>[45]</sup>等人设计了一个二阶注意力网络，通过二阶统计特征量来捕捉图像特征，取得了较好的实验效果。

## § 2.4 亚像素卷积

在传统的定义中图像是由许许多多的像素组成，像素是图形的基本组成单位。但



其实在微观上，实际的物理像素之间还存在着看不见的像素，即亚像素。在超分辨率领域，无法被传感器监测出来的亚像素可以通过算法近似的计算出来，相当于推理出图像纹理细节等高频信息。Shi 等人在 ESPCN<sup>[46]</sup>网络模型中首次应用了亚像素卷积的概念，在文章中作者将亚像素卷积放在了网络的最末端，直接从 LR 图像重建出 HR 图像。这样就意味着每幅低分辨率图像不需要进行前期的处理就可以直接作为输入送入到的网络模型中去，由于输入处理的图像的分辨率降低，因此我们在开始使用尺寸较小的滤波器来提取图像的初始信息，这样一来就大大降低了计算的复杂性和内存使用量。在图像重建部分用可用亚像素卷积来完成图像的特征图像到高分辨率图像的构建，其构建的过程如图 2-8 所示，假设目标倍数为  $r$ ，输入的低分辨率特征图像大小为  $H \times W$ ，将其与通道数为  $r^2$  的  $H \times W$  亚像素卷积进行卷积，得到  $H \times W \times r^2$  个像素值，在将其重新排列组合成大小为  $rH \times rW$  的目标图像。

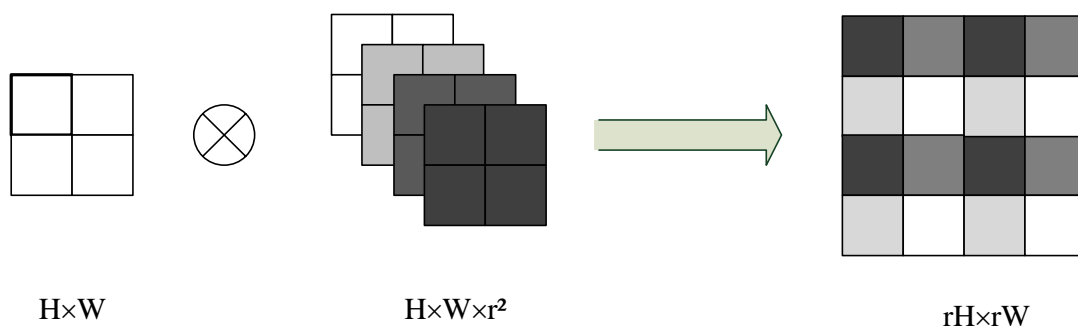


图 2-8 亚像素卷积上采样

## § 2.5 神经网络优化算法

神经网络优化算法是利用某种方案使得网络模型的收敛速度加快或者对损失函数进行优化。以便求得最优的参数，使得损失函数的值最小。在神经网络中常用的优化算法有：随机梯度下降法<sup>[47]</sup>、动量梯度下降法<sup>[48]</sup>和 Adam 算法<sup>[49]</sup>这三种。

### (1) 随机梯度下降法

随机梯度下降法是一种简单且有效的方法，它是从所有的样本中随机抽取其中的一组，训练后按照梯度更新一次，然后在抽取一组，在更新一次，在样本量很大的情况下，可能不用训练完所有的样本就可以得到一个可接受范围内的网络模型了，由于每次只进行一组样本的更新，网络模型的更新所需的参数就会大大减少，加快了网络的训练速度，但这样也会造成优化方向不确定的问题，会最终导致获得的结果可能只是局部最优而不是全局最优，从而降低了网络模型的精度。

### (2) 动量梯度下降法



动量梯度下降法借鉴了物理学中的思想，想像一下将球滚动到无摩擦的碗内，累计的动量并没有停止在碗的底部，而是将其向前推动，这就产生了动量。上面的随机梯度下降法的下降过程并不是直接走向最优点的，而是来回往复的下降，这样就会浪费许多网络训练的时间；还有就是这个随机梯度下降算法需要训练之前自己设置一个学习率，设置过大的学习率会导致参数过量的更新，梯度的摆动过大误差也会较大，设置较小的学习率会使训练的迭代次数增加，这样会增加模型的训练时间，造成资源和时间的浪费。而动量梯度下降法就可以设置较大的学习率，并能减少训练过程中的训练次数，使网络模型的收敛速度更快更加稳定，不会出现来回往复的下降问题。该算法主要是在随机梯度下降的过程中设置一个冲量，用来达到对梯度下降加速的目的。动量梯度下降算法可以用下面的公式来表示：

$$V_{dw} = \theta V_{dw} + (1 - \theta) dw \quad (2-4)$$

$$V_{db} = \theta V_{db} + (1 - \theta) db \quad (2-5)$$

$$W = W - \phi V_{dw} \quad (2-6)$$

$$b = b - \phi V_{db} \quad (2-7)$$

其中  $dw$  和  $db$  分别代表本次迭代所产生的梯度值， $V_{dw}$  和  $V_{db}$  代表前一轮迭代过程中积累的梯度动量。这两个指数的加权平均梯度可以减缓梯度下降的幅度。 $\phi$ 、 $\phi$  是引入的超参，经过大量实验表明， $\phi=0.9$  时效果较好。

### （3）Adam 算法

Adam 算法（Adaptive moment estimation）是 RMSprop 和随机梯度下降法结合两者相结合的一种方法。该算法运用 RMSprop 的平方梯度来缩放网络模型的学习率，并使用梯度移动的平均值而非梯度本身来进行计算，这样可以使得它针对不同的参数设置单独不同的学习率，是一种自适应的学习率方法。Adam 算法因其使用自适应矩阵而得名，它通过对第一阶矩阵和第二阶矩阵的估计来针对性的来调节网络模型中的学习率。并对之前算法存在的波动问题进行了改进，使得网络模型的收敛速度和稳定性都得到了明显的提升。Adam 算法可以用下面的公式来表示：

$$V_{dw} = \phi_1 V_{dw} + (1 - \phi_1) dw \quad (2-8)$$

$$V_{db} = \phi_1 V_{db} + (1 - \phi_1) db \quad (2-9)$$

$$S_{dw} = \phi_2 S_{dw} + (1 - \phi_2) dw^2 \quad (2-10)$$

$$S_{db} = \phi_2 S_{db} + (1 - \phi_2) db^2 \quad (2-11)$$

上述公式中的  $S_{dw}$  和  $S_{db}$  分别表示权重和偏置的二阶指数加权的平均梯度。由于移动的平均梯度值和迭代后的值差距较大，因此需要对到的值进行适当的修正，修正公式为： $V_{dw}^c = V_{dw} / (1 - \phi_1^t)$ ， $V_{db}^c = V_{db} / (1 - \phi_1^t)$ ， $S_{dw}^c = S_{dw} / (1 - \phi_2^t)$ ， $S_{db}^c = S_{db} / (1 - \phi_2^t)$  把修正后的权重和偏置结合更新方程可以到的新的公式：

$$W = W - \phi \frac{V_{dw}^c}{\sqrt{S_{dw}^c + \omega}} \quad (2-12)$$

$$b = b - \phi \frac{V_{db}^c}{\sqrt{S_{db}^c + \omega}} \quad (2-13)$$

上述公式中的  $\phi$  对应动量梯度下降算法中的  $\phi$  值，同样取值为 0.9，参数  $\phi_2$  对应的是类似  $\phi$  的指定值一般取值 0.999，参数  $\omega$  则是一个平滑项，一般取值  $10^{-8}$ ，其中的学习率  $\phi$  则需要在网络模型训练中进行微调。

## § 2.6 图像质量评价标准

图像质量一般是指人对图像的视觉感知评估。图像质量评估<sup>[50]</sup>（IQA）包括基于人类感知的主观评估（即图像的视觉真实感）和客观评估。虽然人类的主观评估符合我们的现实需求，但是这种评估方式耗费大量的人力和时间，并且受到的人为因素过多，无法得到稳定评估结果。因此基于客观评估的计算方法才是主流常用的方法。本文主要介绍两常用的客观评价方法：峰值信噪比<sup>[51]</sup>（PSNR）和结构相似性<sup>[52]</sup>（SSIM）。

### （1）峰值信噪比

峰值信噪比（PSNR）目前最常用的图像质量客观评测方法，通常在图像压缩、图像修复之后会将其与原来的图像进行对比，它是通过比较原图像和被处理图像之间像素值的差异来对图像质量进行评价的。它的公式如下：

$$PSNR = 10 \cdot \log_{10} \left( \frac{k^2}{\frac{1}{N} \sum_{i=1}^N (I(i) - \hat{I}(i))^2} \right) \quad (2-14)$$

通常情况下  $k$  采用 8 位表示法，及  $k$  为 255。由上述公式可知 PSNR 仅与像素间的误差 MSE 有关。而 MSE 它是计算具有  $N$  个像素的真实图像  $I$  和重建图像  $\hat{I}$  之间的均方误差。PSNR 因为只关注图像像素之间的差异，而人往往注重视觉的感官体验，但是为了和文献著作中的实验结果去理性比较，其仍然具有很大的优越性，因此 PSNR 目前仍然是图像超分辨率领域最广泛的评价标准。对于彩色图像一般是计算

RBG 三个通道的 PSNR 值以后在取其平均值，也可以计算 RGB 三个通道的均方误差后取其平均值。

### (2) 结构相似性

通常来说，人类的视觉提取图像的结构更加容易，因此便根据图像的亮度、对比度、结构等方面对比，提出了结构相似性指数（SSIM）来测量图像之间的结构相似性。对于具有  $N$  个像素点的图像  $I$ 。假设图像的亮度为  $\gamma_i$ ，图像的对比度为  $\lambda_i$ ，分别对应图像特征强度的平均值和偏差。其中

$$\gamma_i = 1/N \sum_{i=1}^N I(i) \quad (2-15)$$

$$\lambda_i = \left( \frac{1}{N-1} \sum_{i=1}^N (I(i) - \gamma_i)^2 \right)^{1/2} \quad (2-16)$$

公式中的  $I(i)$  代表的是图像  $I$  的第  $i$  个像素点的特征强度，图像的像素点亮度和对比度的比较  $C_l(I, \hat{I})$  和  $C_c(I, \hat{I})$  可用如下公式表示：

$$C_l(I, \hat{I}) = \frac{2\gamma_I \gamma_{\hat{I}} + c_1}{\gamma_I^2 + \gamma_{\hat{I}}^2 + c_1} \quad (2-17)$$

$$C_c(I, \hat{I}) = \frac{2\lambda_I \lambda_{\hat{I}} + c_1}{\lambda_I^2 + \lambda_{\hat{I}}^2 + c_2} \quad (2-18)$$

上面公式中的  $c_1 = (k_1 m)^2$ ， $c_2 = (k_2 m)^2$  是为了防止分母为 0 的常量，其中  $k_1$ ， $k_2$  是远小于 1 的两个常数，通常取值  $k_1 = 0.01$ ， $k_2 = 0.03$ 。图像的标准化后的像素值可以表示为  $((I - \gamma_I) - \lambda_I)$ ，用原图像的像素值和图像亮度及对比度的相关性来衡量结构相似性，等同于图像  $I$  和图像  $\hat{I}$  之间的相关系数。因此它们之间的比较函数  $C_s(I, \hat{I})$  可用如下公式表示：

$$\lambda_{I\hat{I}} = \frac{1}{N-1} \sum_{i=1}^N (I(i) - \gamma_I)(\hat{I}(i) - \gamma_{\hat{I}}) \quad (2-19)$$

$$C_s(I, \hat{I}) = \frac{(\gamma_{I\hat{I}} + c_3)}{\gamma_I \gamma_{\hat{I}} + c_3} \quad (2-20)$$

公式中的  $\gamma_{I\hat{I}}$  是图像  $I$  和图像  $\hat{I}$  之间的协方差， $c_3$  也是为了防止分母为 0 的常量，最终由上述公式可的 SSIM 的表达式：

$$SSIM(I, \hat{I}) = [C_l(I, \hat{I})]^a [C_c(I, \hat{I})]^\beta [C_s(I, \hat{I})]^\delta \quad (2-19)$$

其中  $\alpha$ ,  $\beta$ ,  $\delta$  是调整图像要素的控制参数, 如果把这些参数设置成 1 则可得到下面的公式:

$$SSIM(I\hat{I}) = \frac{(2\gamma_I\gamma_{\hat{I}} + c_1)(\lambda_I + c_2)}{(\gamma_I^2 + \gamma_{\hat{I}}^2 + c_1)(\lambda_I^2 + \lambda_{\hat{I}}^2 + c_2)} \quad (2-21)$$

公式计算的时候会从图像上取一个  $m \times m$  大小的窗口, 然后根据窗口的移动来进行计算, 最后取其结果的平均值作为全局的 SSIM。这种评价方式是将传统的亮度、对比度通过结构信息来进行表达, 用均值来作为亮度的估值, 协方差来作为结构相似度的度量, 标准差来作为对比度的估值。SSIM 从人眼观察图像的角度去评估图像的重建质量, 与 PSNR 相比更能满足人们在直观视觉方面的需求, 因此这种图像评价方式也得到了广泛的应用。

## § 2.7 本章小结

本章主要介绍了卷积神经网络的基本概念并对其组成部分的内容作了简单介绍。描述了卷积层的计算方式、卷积核对特征信息提取的影响; 对卷积层节点之间的常用的激活函数进行了分析和介绍, 并介绍了神经网络中对权重和参数进行更新的重要算法—反向传播算法, 解释了网络模型训练中参数和权重更新的原理和方法。随后对神经网络模型常用到的残差网络、密集连接网络和亚像素卷积的概念和原理进行了简单的介绍, 对注意力机制和非局部特征的原理和其在超分辨领域的应用进行了分析与介绍, 为后续的介绍打下理论基础。其次对图像重建中用到的优化方法进行了讲解, 不同的优化方法对模型的收敛和重建效果都会有显著的影响。最后对图像重建的常用指标峰值信噪比和结构相似性进行了介绍, 这些指标是后文对提出的网络模型的实验结果分析的重要评价标准, 对衡量模型重建效果和算法优劣有着重要意义。

## 第三章 基于多尺度残差密集块的图像超分辨率重建算法

### § 3.1 引言

卷积神经网络对图像超分辨率重建算法的发展产生了重大影响,随后越来越多的学者将图像超分辨率的研究放在了卷积神经网络上面来。Dong 等人基于卷积神经网络的强大学习能力,首次将卷积神经网络运用到图像超分辨率任务上,并构建出了 SRCNN<sup>[12]</sup>网络模型,虽然 SRCNN 仅有三层卷积神经网络,但却获得了明显优于传统算法的实验效果。随后的 FSRCNN<sup>[13]</sup>为了提升网络的运算速度,将原本需要进行插值放大的输入图像取消,换成原始的 LR 图像,将更多的计算花销转移到网络模型上面去,最后利用反卷积来达到高分辨率图像重建的目的。后来 Kim 等人提出了包含 20 层残差块的 VDSR<sup>[16]</sup>模型,使得网络模型的深度进一步加深,其对网络的设计进行了很大的改进,图像的重建质量也有很大的提升。但是上述基于卷积神经网络的算法都没有充分的利用网络的层次特征学习能力,即网络模型的先前层次的特征信息对后续的输出影响较小,导致后续层的性能相对较差。

为了解决上述问题,本章提出了一种可以充分利用全局和中间特征的多尺度残差密集块卷积神经网络。网络中主要包含多个多尺度残差密集块,每个模块都可以将提取到的图像特征进行密集特征融合,并在每个块内添加局部残差融合,保证低频信息的有效利用,然后将所有的多尺度残差密集块进行全局特征融合,从而使得网络模型学习到整个特征空间相关性的内容,形成了图像特征信息机制的全局记忆。在公共数据集上的大量实验表明,所提的方法展现出了较好的实验效果。

### § 3.2 多尺度残差密集融合网络

为了高效的学习到深层网络产生的分层特征,并能进行多尺度重建,提出了多尺度残差密集块卷积神经网络,图 3-1 展示了网络模型的体系结构。该网络模型主要包括浅层特征提取、多尺度残差密集融合、全局残差融合、多尺度亚像素卷积重建四个部分。其中浅层特征提取主要由两个卷积构成,它的主要功能是将输入进来的低分辨率图像进行浅层特征提取,并将其作为后续多尺度残差密集块的输入;多尺度残差密集融合主要是利用密集连接和残差学习的思想,在多尺度残差密集块内的各个卷积层之间进行密集连接和残差融合,并将各个多尺度残差密集块之间也利用密集连接进行全局信息融合,达到图像信息最大程度的利用;全局残差融合是将浅层特征和后面提

取得到深层次特征进行残差融合；最后利用多个尺度的亚像素卷积进行最终重建得到高分辨率图像。

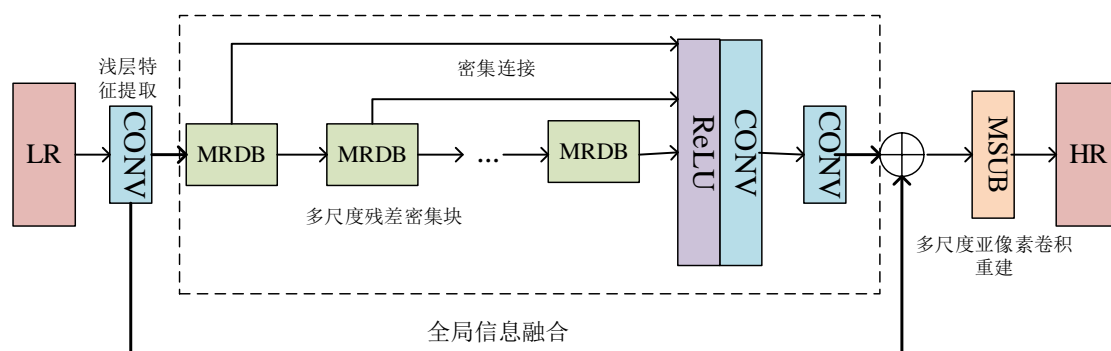


图 3-1 全局网络结构示意图

### § 3.2.1 浅层特征提取

浅层特征的提取主要由两个  $3 \times 3$  卷积层来组成，其过程可用下面的公式表示：

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

其中， $I_{LR}$  表示输入的低分辨率图像， $H_{conv}$  代表卷积操作对图像特征的提取， $F_0$  代表浅层特征提取后的到的特征图，它作为后续深层特征的输入，因为低分辨率图像和高分辨率图像有着相似的图形轮廓，因此低分辨率图像中有大量的浅层信息可以用来进行图像的超分辨率重建，后面将提取到的浅层特征与深层特征进行残差融合，用来提升图像的重建质量。

### § 3.2.2 多尺度残差密集融合

随着卷积神经网络模型的不加深，深层次的高频特征信息得以提取，但这些网络都面临前面卷积层提取到的特征信息不能随着网络卷积层的叠加而持续传递的问题，这就会导致许多前面卷积层已获取的信息的流失，同时也浪费了计算资源，为了最大限度的利用每个卷积层提取的特征信息，提出了多尺度残差密集块（如图 3-2），每个密集残差块包含三种不同尺度的卷积核，卷积核的大小分别为 3、5 和 7。每一个多尺度残差密集块都同时包含特征密集连接和残差融合，使得前一层的信息能够传递到后面的每一层，以此保证网络中提取到的特征信息的能够充分有效的利用。

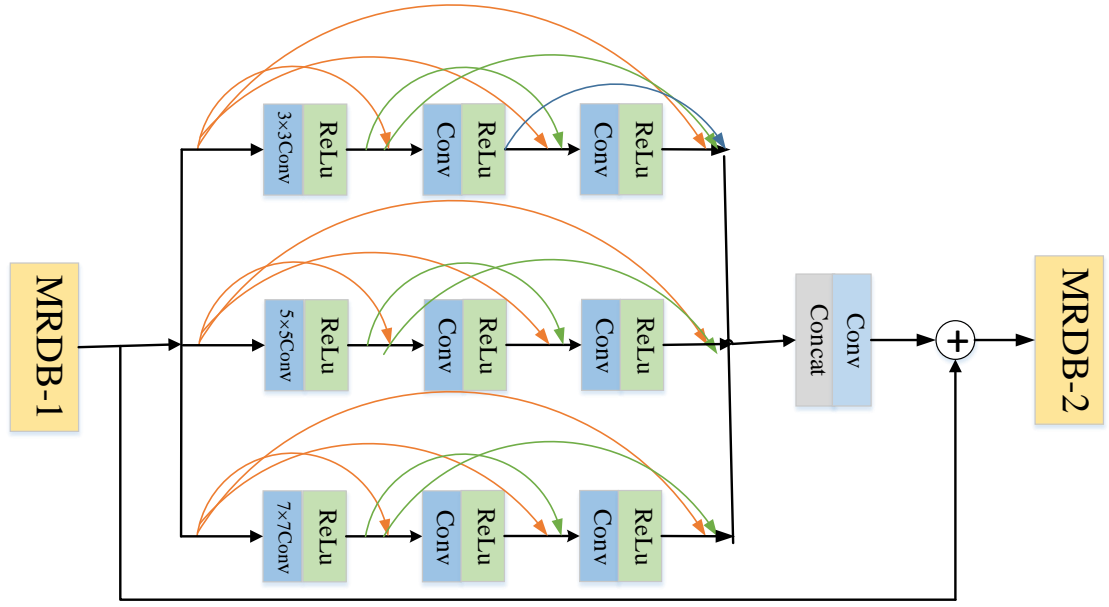


图 3-2 多尺度残差密集块结构示意图

将前面提取的浅层特征  $F_0$  作为  $MRDB$  的输入，用  $MRDB$  学习图像的深层特征，分别用  $MRDB_{d-1}$  和  $MRDB_d$  和分别表示第  $d$  个多尺度残差密集块的输入和输出，则第  $d$  个  $MRDB$  可以通过式 (3-2) 获得：

$$F_d = O_{MRDB,d}(F_{d-1}) = O_{MRDB,d}(O_{MRDB,d-1}(\dots(O_{MRDB,1}(F_0)))) \quad (3-2)$$

其中， $O_{MRDB,d}$  是对输入到  $MRDB$  中的特征图进行密集连接和残差连接操作。 $F_0$  是前面提取到的浅层特征信息，作为第一个  $MRDB$  的输入， $F_d$  是第  $d$  个  $MRDB$ ，则其输入是第  $d-1$  个  $MRDB$  的输出  $F_{d-1}$ ，这样便能使得每个  $MRDB$  的都能与前后的  $MRDB$  直接相连，从而实现特征图像信息的连续传输。每个卷积层的具体操作如下公式所示：

$$F_{d,i} = \max(0, w_{d,i} * [F_{d,1}, F_{d,2}, \dots, F_{d,i-1}] + b_i) \quad (3-3)$$

其中， $w_{d,i}$  是第  $i$  个卷积层的权重， $b_i$  代表卷积层的偏执； $[F_{d,1}, F_{d,2}, \dots, F_{d,i-1}]$  表示由  $d-1$  个  $MRDB$  通过卷积操作在第  $d$  个  $MRDB$  中生成的特征映射。例如，假设上一个  $MRDB$  有  $G_0$  个特征映射，每个  $MRDB$  模块中的每个卷积层输出  $g$  个特征映射，则每个  $MRDB$  块最终产生  $G_0 + (i-1)g$  个特征映射。这样就可以将每个  $MRDB$  的以及每个卷积层的输出都连接到后续所有的  $MRDB$  和卷积层中，保证了网络信息的充分利用。

由于每个  $MRDB$  包含提取到的前面所有卷积层的信息，这就会造成提取的特征过多，网络训练困难等问题，为了减少特征图的数量，使用了  $1 \times 1$  卷积层来降低特征图的数量，使得在下一个  $MRDB$  的输入中将特征图降低到所需要的数量需求，在保

证尽量不损失特征图的基础上来达到特征图信息的传递。

### § 3.2.3 全局特征融合

通过 *MRDB* 模块的进行深层次的特征提取可以得到图像的深层次的特征，将多个 *MRDB* 获取到的信息进行密集连接融合，最大程度的利用提取到的特征信息。该过程如下图公式所示

$$F_{DF} = H_{DF}(MRDB_1, MRDB_2, \dots, MRDB_d) \quad (3-4)$$

其中， $F_{DF}$  是指多个 *MRDB* 模块进行密集连接得到的深层次特征， $H_{DF}$  是指密集连接操作，将多个密集残差块 *MDRB* 产生的特征图进行合并连接，并通过  $1 \times 1$  卷积层来进行降维，降低得到的特征图的数量。最后将全局融合的深层特征  $F_{DF}$  和前面提取到的浅层特征进行残差融合后进行上采样操作，最终完成图像的重建工作。该过程如下图公式所示：

$$HR = f_{\uparrow}(F_{DF} \oplus F_0) \quad (3-5)$$

其中， $f_{\uparrow}$  表示上采样重建， $F_{DF}$  和  $F_0$  分别表示获取到的深层特征和浅层特征， $\oplus$  表示残差融合操作， $HR$  指最终的到重建图像。

### § 3.2.4 多尺度亚像素卷积

先前的超分辨率重建算法在将低分辨率图像输入网络之前通常采用插值算法将图像放大至目标倍数的方式，但是这种方式会网络参数增多，增加网络模型的训练难度。目前大多数图像超分辨率算法都取消了这种插值算法，取而代之的是在图像的特征提取之后，在网络末端采用反卷积和亚像素卷积操作来将图像放大至目标倍数。

考虑到图像具有区域相关性，因此使用不同尺寸的卷积核来获取不同区域大小的相关信息<sup>[52]</sup>，以提高图像最终的重建效果。提出的多尺度亚像素卷积便是在融合特征输出后进行不同尺度的卷积操作，在分别对不同尺度的卷积核输出的特征图进行亚像素卷积操作，随后进行特征融合，使得不同尺度的信息融合在一起，最后进行特征融合，重建出最终的高分辨率图像。多尺度亚像素卷积如图 3-3 包含三种不同比例的亚像素卷积分支，卷积核的大小分别为 3、5 和 7。可用如下公式进行表示：

$$F_{MSUB} = \text{Conv}(\text{Concat}[f_{\text{sub-conv}3}(F_{DF}), f_{\text{sub-conv}5}(F_{DF}), f_{\text{sub-conv}7}(F_{DF})]) \quad (3-6)$$



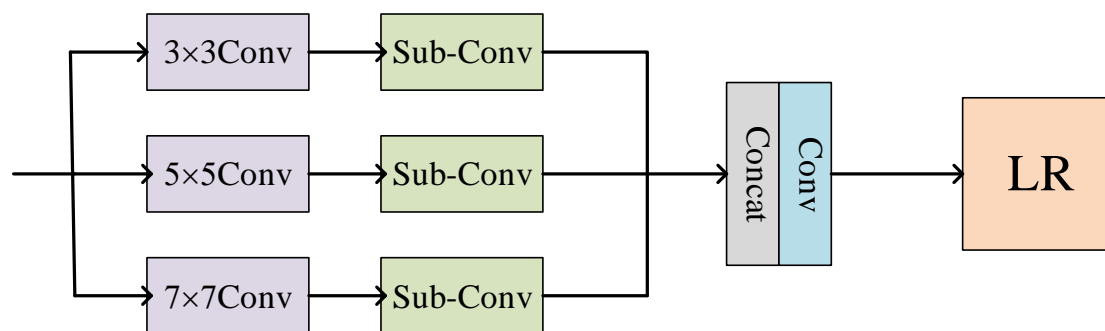


图 3-3 多尺度亚像素卷积结构示意图

### § 3.3 实验结果与分析

#### § 3.3.1 参数设置

进行本次实验使用的系统是 Ubuntu16.04，使用的框架是 Pytorch，使用的 IDE 为 PyCharm，电脑的硬件配置为 Intel® Core™ i5-8500 CPU @3.00GHz × 6，NVIDIA GTX1066/PCIe/SSE2。实验使用的是 Timofte 等发布的数据集 DIV2K<sup>[53]</sup>来训练网络模型，对 DIV2K 数据进行数据增强：将图像块进行翻转，以及多角度的旋转。对测试数据集则采用了 Set5<sup>[54]</sup>，Set14<sup>[55]</sup>和 BSD100<sup>[56]</sup>这三个基准数据集对算法的效果进行验证。本次算法实验的学习率设置为  $1e-4$ ，每经过 200 个小批次更新便将学习率减半；每个小批次输入样本的数目为 8；初始卷积通道数为 32；卷积核步长为 1；激活函数为 ReLU；在训练数据的迭代时，使用 Adam 对网络中各个层的权重进行更新。

#### § 3.3.2 损失函数选择

在相同实验环境条件下，网络参数相同，训练批次相同（均为 200 次），相比于常用的损失函数均方误差(MSE)，以放大比例为×2 的重建任务中，在 Set5 测试集下，L1 损失函数和 MSE 损失函数的 PSNR 值如图 3-4 所示：

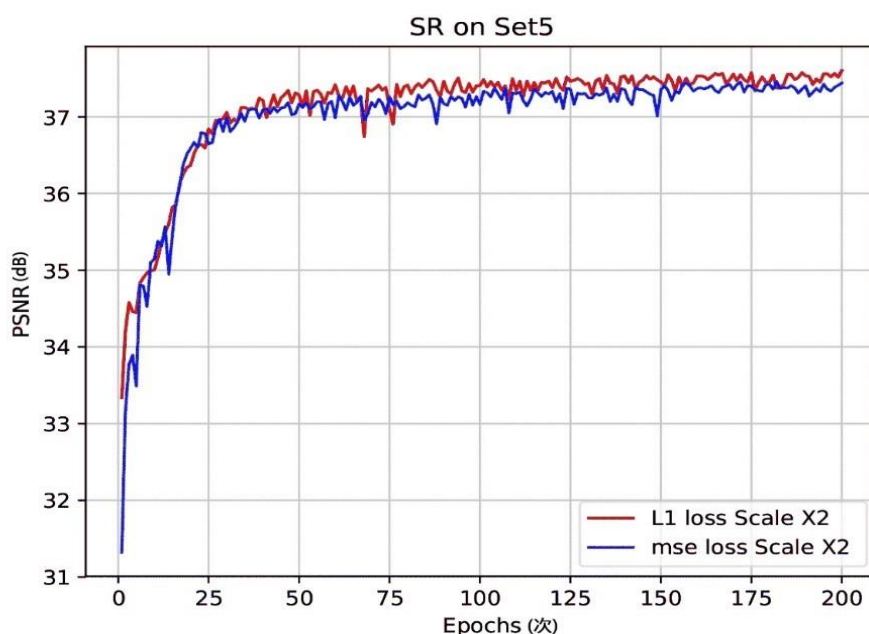


图 3-4 L1 损失和 MSE 损失对比效果图

### § 3.3.3 消融实验分析

为了证明网络结构的更改对重建图像效果的影响，我们对网络结构进行了消融实验分析，我们以放大尺寸为 2、针对测试集 Set5 为例，比较了不同模块对网络结构的影响。由于该网络使用了多尺度特征提取、局部密集连接和全局密集连接，因此准备了一个基准实验组和局部添加组，在添加组分别将这几个模块分别添加。

表格 3-1 不同模块添加在测试集 Set5、比例为 2 条件下的 PSNR 平均值

多尺度模块	局部密集连接	全局密集连接	PSNR
√	√	√	38.00
√	√	×	37.68
√	×	√	37.63
×	√	√	37.82

其中，符号“×”和“√”分别表示不存在某个模块和存在某个模块。对于多尺度特征融合模块，“×”并不是不存在该模块，而是将多尺度换成单一尺度 $3\times 3$ 的卷积核，其他的完全一样。从上表可以看到，网络中添加局部密集连接和全局密集连接分别都可以使得实验的结果有不同程度的提升，主要是因为这两个模块能使网路中的特征图像信息能够充分的得到利用，减少前端网络中特征信息的丢失。添加多尺度模块则可以增加网络的宽度，提取到更加丰富的图像信息。

### § 3.3.4 多尺度亚像素卷积分析

为了验证多尺度亚像素卷积的重建图像性能，简单设计了一个模型对比试验，该模型剔除了模型的其他模块仅包含特征提取层和重建层。该模型中重建模块分别使用亚像素卷积和多尺度亚像素卷积进行图像重建。模型按照 $\times 2$  缩放比例进行模型训练，并且在模型训练中实验参数全部相同。实验都使用 Set5 测试集对模型进行实验结果对比。亚像素卷积和多尺度亚像素卷积在重建结果的 PSNR 值在 Set5 测试集下的变化图如下图 3-5 所示：

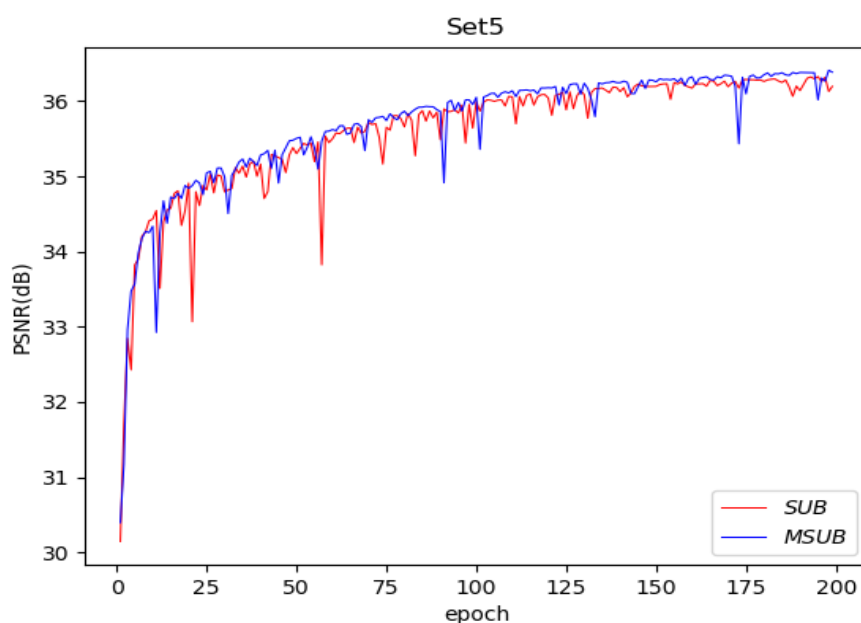


图 3-5 亚像素卷积和多尺度亚像素卷积重建效果对比

从上图可以看出采用多尺度亚像素卷积进行重建的图像的 PSNR 值明显高于亚像素卷积的图像的 PSNR 值，并且这两种方式都有相同的收敛趋势，因此本章提出的网络模型使用多尺度亚像素卷积进行最终的图像重建，能够获得更好的结果。两者在测试集 Set5、Set14 并且放大比例为  $\times 2$  下的 PSNR 平均值如表 3-2 所示：

表格 3-2 两种上采样方式在测试集 Set5、Set14 且比例为 2 条件下的 PSNR 平均值

模型	Set5	Set14
亚像素卷积	36.326/0.22s	32.232/0.45s
多尺度亚像素卷积	36.410/0.27s	32.278/0.58s

我们可以看到，在网络参数以及训练数据等其他实验环境相同的情况下，多尺度亚像素卷积的 PSNR 比亚像素卷积的更高，而且两者运行的时间是相差无几的。

### § 3.3.5 重建结果对比与分析

为了更大程度的保证网络模型中各个卷积层信息的利用和融合，在图像信息提取的时候采用了多尺度密集残差块，更好的利用了图像所有的分层特征，利于高频信息的恢复。在图像重建的时候采用了多尺度亚像素卷积，有效的解决图像不同区域具有相关性的问题。因此本文提出的算法的 PSNR 平均值与传统的 Bicubic 算法相比有明显的提高，相比于其他基于 CNN 的算法也均有提高，也实现了更好的重建效果。不同算法在不同测试集中以×2、×3、×4 的比例放大时的 PSNR 和 SSIM 分别如表 3-3、表 3-4、表 3-5 所示：

表 3-3 不同算法在测试集 Set5 的 PSNR 和 SSIM 平均值

算法	x2	x3	x4
Bicubic	33.66 / 0.9299	30.39 / 0.8682	28.42./0.8104
SRCNN	36.66 / 0.9542	32.75 / 0.9090	30.48./0.8628
FTSR	36.88 / 0.9550	32.97 / 0.9112	30.55./0.8645
DRSR	37.41 / 0.9521	33.60 / 0.9168	31.18./0.8836
VDSR	37.53 / 0.9587	33.66 / 0.9213	31.35./0.8838
本文方法	<b>37.94 / 0.9598</b>	<b>34.22 / 0.9255</b>	<b>32.11./0.8943</b>

表 3-4 不同算法在测试集 Set14 的 PSNR 和 SSIM 平均值

算法	x2	x3	x4
Bicubic	30.24 / 0.8688	27.55 / 0.7742	26.00 / 0.7027
SRCNN	32.42 / 0.9063	29.28 / 0.8208	27.49 / 0.7503
FTSR	32.43 / 0.9060	29.31 / 0.8212	27.61 / 0.7549
DRSR	32.95 / 0.9063	29.69 / 0.8314	27.83 / 0.7624
VDSR	33.03 / 0.9124	29.77 / 0.8314	28.01 / 0.7674
本文方法	<b>33.55 / 0.9192</b>	<b>30.22 / 0.8435</b>	<b>28.47 / 0.7885</b>

表 3-5 不同算法在测试集 BSD100 的 PSNR 和 SSIM 平均值

算法	x2	x3	x4
Bicubic	29.56 / 0.8431	27.21 / 0.7382	25.96 / 0.6675
SRCNN	31.36 / 0.8879	28.41 / 0.7863	26.90 / 0.7101
FTSR	31.24 / 0.8862	28.42 / 0.7875	26.78 / 0.7084
DRSR	31.55 / 0.8924	28.54 / 0.7917	27.00 / 0.7092
VDSR	31.90 / 0.8960	28.82 / 0.7976	27.29 / 0.7251

本文方法	<b>32.14 / 0.9085</b>	<b>29.04 / 0.8222</b>	<b>27.52 / 0.7592</b>
------	-----------------------	-----------------------	-----------------------

为了更清晰直观的对比不同算法的重建效果，本章选用测试集 Set5 下的几张图像以 $\times 2$  放大比例下的重建结果作为对比，如下图 3-6 和图 3-7 所示，图中从左到右依次为原始输入的低分辨率图像、真实高分分辨率图像、Bicubic、SRCNN、FTSR、DRSR 和 VDSR 以及本章所提算法重建后的图像。

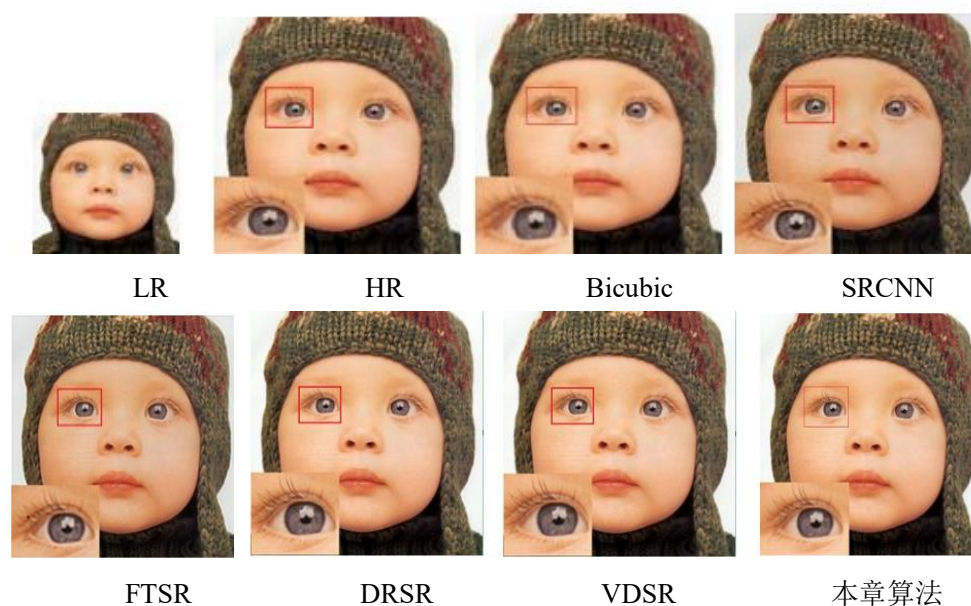


图 3-6 放大比例为 2，Set5 数据集中“Baby”的对比结果

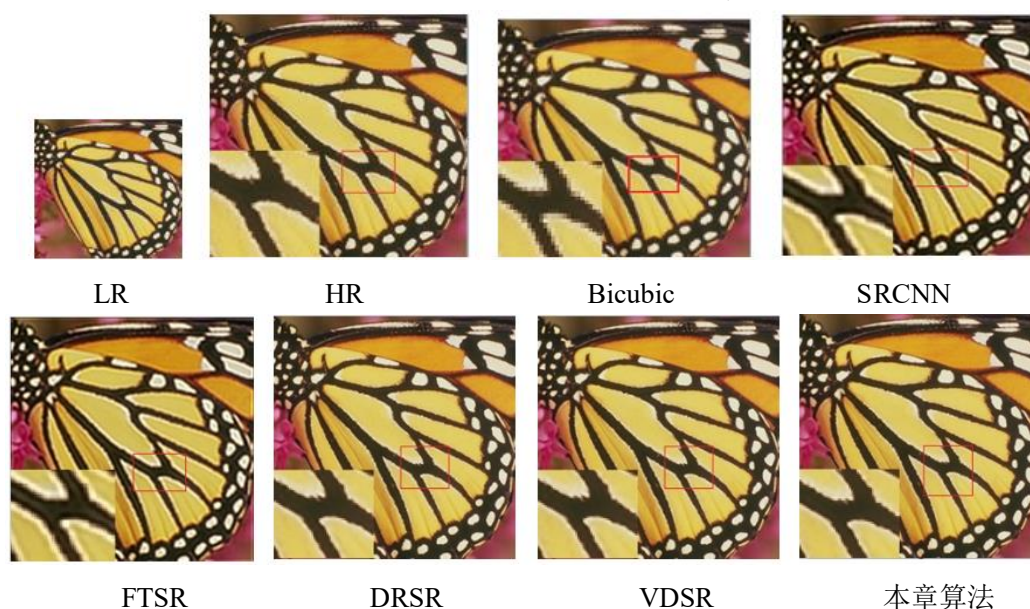


图 3-7 放大比例为 2，Set5 数据集中“Butterfly”的对比结果

通过对上述两组图片的观察可以发现，Bicubic 算法重建得到的图像效果是最差的，它的图像边缘细节模糊并周围还存在着许多伪影；SRCNN 和 FTSR 这两组算法重建的效果于 Bicubic 算法相比有着相对清晰的轮廓，虽然伪影较少，但是依然存在

边缘细节模糊问题；而 VDSR 和 DRSR 这两种算法与前面的集中算法相比较，实现了更多边缘细节上的重建，即这两种算法获得了图像更多的高频信息；本章算法与前几种算法相比，细节信息上则更加丰富，从“Baby”图中可以看到本章算法重建得到的图像眼睫毛和眼睑的显示效果更清晰，从“Butterfly”图中可以看到蝴蝶纹路的边缘与真实 HR 图像更为接近，都有更多的边缘细节。

### § 3.4 本章小结

本章针对多数图像超分辨率重建算法存在的不能充分利用网络层次特征的学习能力，提出了一种基于多尺度残差密块的卷积神经网络图像超分辨率算法。通过构建多个多尺度密集残差块对图像特征进行密集残差融合，实现各层特征的存储复用，使得每层卷积提取到的特征信息都能得到最大程度的利用；同时采用多尺度的卷积核来提取不同类型的图像特征信息，增加网路的宽度；其次通过对浅层特征和深层次特征进行全局残差融合，保证低频信息的有效利用；最后采用多尺度亚像素卷积进行图像重建。实验结果表明，本文提出的方法重建效果优于其他算法，能够重建出更多边缘细节，得到质量较好的高分辨率图片，在客观评价指标和主观视觉上都展现了更好的重建效果。



## 第四章 基于注意力机制与多尺度特征融合的图像超分辨率重建算法

### 建算法

#### § 4.1 引言

图像超分辨率重建在图像领域的广泛应用引发了诸多学者探究图像重建方法的热潮，而卷积神经网络在处理这方面的问题展现了无与伦比的优异性能，伴随着计算机硬件运算能力的提升，构建的网络模型也越来越复杂，这些都促进了卷积神经网络重建算法的大量应用。Dong 等人首次将卷积神经网络应用到图像超分辨率领域，仅仅使用了三层卷积神经网络 SRCNN<sup>[12]</sup>就取得了明显优于传统方法的效果，展现了卷积神经网络在特征提取和表示方面的优越性。随后的 FSRCNN<sup>[13]</sup>为了提升网络的运算速度，将原本需要进行插值放大的输入图像取消，换成原始的低分辨率图像，将更多的计算花销转移到网络模型上面去。最后利用反卷积来达到高分辨率图像重建的目的。然而，尽管基于卷积神经网络的图像超分辨率重建方法取得的比较好的效果，但是仍然存在以下问题：（1）图像的而非局部特性在传统方法中得到了广泛应用<sup>[57]</sup>，但是在大多数基于 CNN 的图像超分辨率重建方法中并没有充分利用图像的固有非局部相似特性，（2）目前大多数基于 CNN 的图像超分辨率算法主要其中在加深网络深度和更复杂的网络模型来学习更多的图像特征，却很少利用高阶的图像特征，文献<sup>[58]</sup>表明图像的高阶特征比传统的一阶特征包含更丰富的特征表现能力。（3）大多数基于 CNN 的图像超分辨率算法只在单一尺度下提取图像特征信息，忽略了不同尺度下图像的细节信息。

为了解决上述问题，本章提出了一种基于注意力机制与多尺度特征融合的网络结构。该网络主要由多个注意力机制模块组成，而每个注意力机制模块主要包含非局部特征和二阶统计特征，并将二者提取的特征信息融合，便于获取图像结构信息和高阶图像信息，而后将多个注意力机制模块进行残差信息融合，同时在多个注意力机制模块的前后添加多尺度特征模块用来提取不同尺度下的图像信息，获取更多的图像局部信息，最后通过亚像素卷积对图像进行重建，得到最终的高分辨率图像。在公共数据集上的大量实验表明，所提的方法展现出了较好的实验效果。

## § 4.2 注意力机制与多尺度特征融合网络

为了充分的利用图形固有的非局部特征和图像的高阶特征信息，提出了注意力机制与多尺度特征融合的卷积神经网络如下图4-1所示，该网路模型主要由三部分组成，分别是浅层特征提取模块、注意力机制与多尺度信息融合模块和重建模块。其中浅层特征提取主要是有两个卷积构成，它的主要功能是将输入的低分辨率图像进行浅层特征提取，并将其作为后续模块的输入；注意力机制与多尺度信息融合模块和重建模块主要是利用了非局部特征和二阶统计特征的思想，每个注意力机制模块都有这两个模块组成，并将二者提取的特征信息进行融合，并将多个注意力机制模块进行残差融合，同时在注意力机制模块前后加入多尺度特征提取模块提取不同尺度下的图像特征信息，并将其获取的信息与注意力机制残差融合模块进行信息融合，最后将得到的特征信息通过亚像素卷积进行重建，得到最终的高分辨率图像。

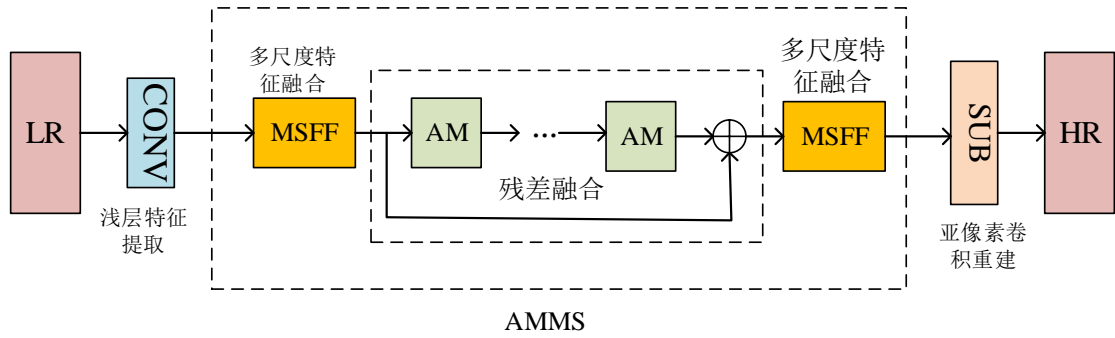


图 4-1 全局网络结构

## § 4.3 浅层特征提取模块

浅层特征的提取主要由两个 $3 \times 3$ 卷积层来组成，其过程可用下面的公式表示：

$$F_{SF} = H_{SF}(I_{LR}) \quad (4-1)$$

其中， $I_{LR}$ 表示输入的低分辨率图像， $H_{SF}$ 代表卷积操作对特征的提取， $H_{SF}$ 代表提取得到浅层特征信息，它将作为后续深层特征提取的输入。

## § 4.4 注意力机制与多尺度信息融合模块

将上面提取到的浅层特征 $F_{SF}$ 经过基于注意力机制与多尺度特征融合(AMMS)模块进行深层次的特征提取公式如下所示：



$$F_{DF} = H_{AMMS}(F_{SF}) \quad (4-2)$$

其中  $H_{AMMS}$  表示深层特征提取，对于 AMMS 模块，主要由两个多尺度特征融合模块(DF-MS)模块和一个注意力机制融合模块(DF-AM)组成。

#### § 4.4.1 多尺度特征融合模块

一般来说图像会包多种类型的特征，特征的不同，其对应的特征尺寸也不尽相同，比如，图像的平滑区域和边缘区域对应的特征尺度会较大，而图像的纹理区域对应的特征尺度相对较小。CNN 网络通过卷积层来提取图像的特征，但是不同大小的卷积核能提取不同类型的特征，当卷积尺寸较大时，CNN 可以提取图像的较大尺度的特征，当卷积核尺寸较小时，CNN 则可以提取到图像的小尺度的细节特征。GoogLeNet 通过不同尺度的卷积核提取图像的特征，取得了较好的实验效果。受 GoogLeNet 的启发，我们利用一个多个不同尺度大小的卷积核来组成多尺度特征融合模块来提取图像不同类型的特征信息，以便更好的获取图像的结构信息。如下图 4-2 所示，该模块主要由多尺度特征提取层、特征融合层、降维层。

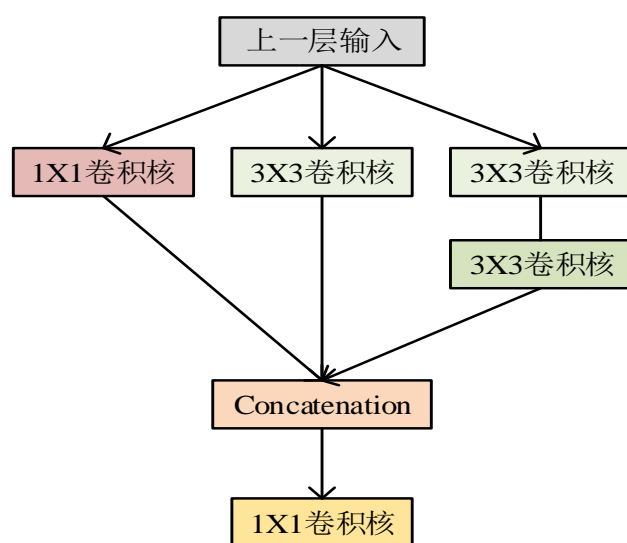


图 4-2 多尺度特征融合模块

对多尺度特征提取层，采用三个尺度的卷积核，分别为  $1 \times 1$ 、 $3 \times 3$ 、 $3 \times 3$  和  $3 \times 3$ ，其中两个  $3 \times 3$  卷积相当于一个  $5 \times 5$ 。其中  $1 \times 1$  的卷积核能够将上一层的特征保留并与其他尺度的特征进行融合，使得网络包含更加丰富的特征信息。在每一个卷积核后面采用一个非线性层，提高网络模型的非线性映射能力。在提取到图像的多尺度特征之后，我们利用 concatenation 层将不同尺度下卷积核提取到的特征图进行叠加融合，然后用  $1 \times 1$  卷积核用来降低融合之后特征图的维度，因为  $1 \times 1$  既可以保持 concatenation

层的所有特征信息,而且由于其通道数较少,卷积之后得到的多尺度特征图也会减少,这样就可以方便网络模型的训练。

#### § 4.4.2 注意力机制融合模块

对于注意力机制融合模块,结构如图 4-3 所示,由  $M$  个注意力机制模块残差融合而成。用  $F_i, F_o$  分别来表示注意力机制融合模块的输入和输出特征。

$$F_o = H_{RF-AM}(F_i) = W_{RF} F_i + H_{AM}(F_i) \quad (4-3)$$

其中  $H_{RF-AM}$  表示  $M$  个 AM 模块残差融合,  $H_{AM}$  表示 AM 模块,  $W_{RF}$  表示残差融合操作。通过残差融合可以将前面的低层信息传递到深层丰富图像的特征信息,增加网络的模型训练的稳定性 and 效率。

注意力机制模块如图 4-3 所示,它主要通过非局部特征和二阶特征这两种方式来进行图像进行特征提取,然后将提取到的特征进行融合作为整个注意力模块的输出。

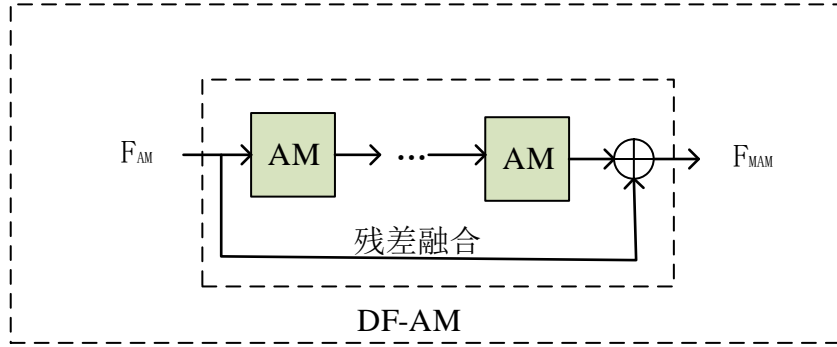


图 4-3 注意力机制融合模块

用  $F_{g-1}, F_g$  分别表示  $M$  个注意力机制模块的输入和输出特征图,对于非局部的特征提取公式如下:

$$F_{NL} = H_{Con-NL}(F_{g-1}) \quad (4-4)$$

其中  $H_{Con-NL}$  表示的是非局部特征提取操作,  $F_{NL}$  是经过操作之后得到的特征图,对于二阶特征提取公式如下:

$$F_{SO} = H_{Con-SO}(F_{g-1}) \quad (4-5)$$

其中  $H_{Con-SO}$  表示二阶特征提取操作,  $F_{SO}$  是经过操作之后得到的特征图。二者特征提取之后在进行特征融合,其特征融合过程可用下面公式表示:

$$F_g = H_C(F_{NL}, F_{SO}) \quad (4-6)$$

其中  $H_c$  表示特征融合操作，然后在用一个  $1 \times 1$  的卷积进行降维，来降低特征图的数量，减少模型的计算损失。注意力模块对图像特征的处理可用下面公式表示：

$$F_g = H_{AM}(F_{g-1}) \quad (4-7)$$

其中  $H_{AM}$  表示基于非局部和二阶特征融合的注意力模块，通过特征融合，可以有效的利用图像的固有属性和图像的高阶图像信息，显著的提升网络模型的性能。

(1) 非局部操作：非局部操作的提出就是为了获取图像间的长距离依赖关系，将图像中非相邻的像素点纳入到一个整体去考虑，这样就可以充分利用图像中的结构信息和图像的自相似性。将非局部操作融合到 CNN 计算中，可以使模型得到整个图像的特征相关性，并且非局部操作对网络中特征图大小并不会产生影响，这样就使得非局部操作可以很方便的嵌入到任意网络架构中。

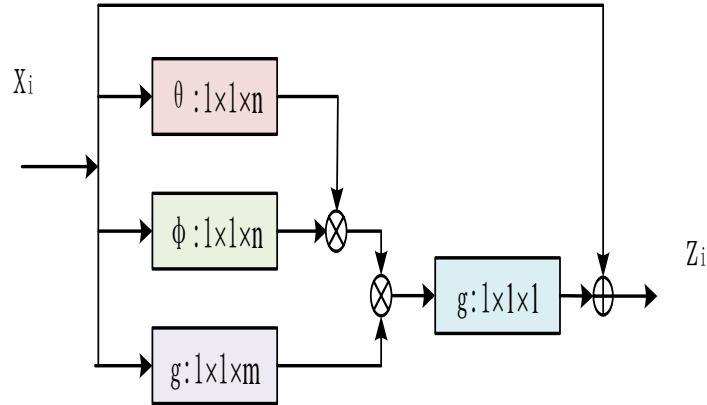


图 4-4 非局部操作模块

卷积神经网络关于非局部操作的表达式为：

$$y_i = \frac{1}{\sigma(x)} \sum_{\forall j} f(x_i, x_j) g(x_j) \quad (4-8)$$

其中  $x$  是表示的是输入图像， $y$  表示的是输出图像， $i$  代表图像位置索引， $j$  代表的是图像中所有的位置索引， $f$  是计算函数，用来计算  $x_i$  和  $x_j$  的相关性， $g$  是一个线性函数，用来进行信息的线性变化， $\sigma(x)$  是归一化函数，用来保证经过变换后整体信息不变。

$$g(x_j) = W_g x_j \quad (4-9)$$

对于  $f$  我们通常采用 embedded Gaussian，在嵌入式空间中计算高斯距离。表达式为：

$$f(x_i, x_j) = e^{\theta(x_i)^T \phi(x_j)} = e^{(W_\theta x_i)^T W_\phi x_j} \quad (4-10)$$

其中  $W_\theta, W_\phi$  分别是可以学习的权重。将  $\sigma(x)$  考虑进去其实就是 softmax 形式，表达式为

$$y = \text{softmax}(x^T W_\theta^T W_\phi x) g(x) \quad (4-11)$$

以上都是基本的非局部操作算子，利用这些算子可以构造模块。表达式为

$$z_i = W_z y_i + x_i \quad (4-12)$$

可以看出上面构造出了残差形式， $W_z$  表示可以学到的参数矩阵， $+x_i$  表示残差连接， $y_i$  表示非局部算子。该模块可以看作一个残差模块，可以在不影响原始网络结构模型的基础上将该非局部模块嵌入到网路框架中。

(2) 二阶特征提取：传统的卷积和全局平均池化只能获取图像的一阶特征，受文献<sup>[59]</sup>启发我们采用全局协方差池化（global covariance pooling, GCP）操作来获取图像的二阶统计特征。如图 4-5 所示：

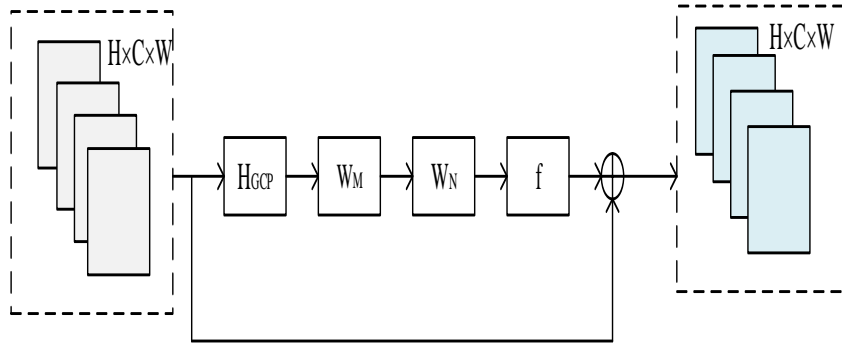


图 4-5 二阶特征提取模块

首先通过 GCP 来提取图像的二阶统计特征，表达式为：

$$Z_C = H_{GCP}(y_c) = \frac{1}{C} \sum_i^C y_c(i) \quad (4-13)$$

其中  $H_{GCP}$  表示全局协方差池化操作， $y_c$  表示输入的特征， $y(i)$  表示在  $i$  处的输入特征。

$$s = f(W_M \phi(W_N Z_C)) \quad (4-14)$$

其中  $W_M, W_N$  是学习到的权重矩阵， $f$  和  $\phi$  分别代表 sigmoid 和 ReLu 函数，经过

sigmoid 函数，将其值缩放到 0 到 1 之间，对特征图进行重新的校准，然后将特征图与原来每个通道中的特征图的值进行对应相乘。公式如下所示：

$$x_c = s_c \cdot x_c \quad (4-15)$$

其中  $s_c$  是指尺度因子， $x_c$  代表信道特征。

## § 4.5 图像重建模块

为了降低 LR 图像进行上采样后进行网络训练带来的参数数量提升的问题，大多数图像超分辨率算法都取消了将 LR 输入网络前的使用插值算法将图像放大至目标 HR 图像的大小，取而代之的是在图像特征提取之后，运用反卷积或者亚像素卷积操作将图像方法至指定分辨率。考虑到反卷积会存在大量的补 0 的区域，这会影响到图像的重建效果，因此本文采用效果更好的亚像素卷积来进行图像的重建。

$$I_{HR} = f_{sub}(F_{DF}) \quad (4-16)$$

其中  $f_{sub}$  表示亚像素卷积操作， $F_{DF}$  是经过特征提取得到的深层特征图， $I_{HR}$  是最终得到的重建高分辨率图像。

## § 4.6 实验结果与分析

本章实验使用的系统是 Ubuntu16.04，整个实验所使用的框架是在系统上搭建的 Pytorch，进行实验所使用的 IDE 为 PyCharm。电脑的硬件配置为 Intel® Core™ i5-8500 CPU @ 3.00GHz × 6，NVIDIA GTX1066/PCIe/SSE2。实验使用的是 Timofte 等发布的数据集 DIV2K<sup>[53]</sup>来训练网络模型，测试数据集则采用了 Set5<sup>[54]</sup>，Set14<sup>[55]</sup>和 BSD100<sup>[56]</sup>这三个基准数据集，这些数据集常用于图像的 SR 重建。为了与其他算法进行比较，我们采用了 PSNR 和 SSIM 这两个指标来对实验的重建效果评价。实验的初始学习率为 1e-4，然后每次迭代  $2 \times 10^5$ ，学习率减少一半，使用 Adam 对网络中各个层的权重进行更新。

### § 4.6.1 亚像素卷积对网络性能的提升

为测量亚像素卷积对模型重建效果的影响，设计两个模型进行实验比较：把低分辨率图像输入前进行插值放大的模型记做  $L'$ ，把原始低分辨率图像直接输入并采用亚像素卷积进行重建的模型  $L$ 。两个模型的其余参数设定完全相同。具体的实验结果

如下表 3-2 所示：

表 4-1 两个模型对应的参数数量以及 Set5 下的运行时间

模型	参数数量/运行时间
$L'$	3.27M/0.23s
$L$	<b>3.26M/0.19s</b>

从上表中的结果可以发现，使用亚像素卷积进行图像重建，使得网络模型在参数数量方面有着较为明显的减少，网络的运算速度也有相应提升。

### § 4.6.2 消融实验分析

为了研究注意力机制和多尺度特征融合模块对该网路模型的影响，我们用消融学习的思想来对网络模型进行对比分析，我们以放大尺寸为 4 为例，比较了不同模块对网络结构的影响，由于该网络使用了多尺度特征融合模块、非局部操作和二阶特征提取操作，因此准备了一个基准实验组和局部添加组，在添加组分别将这几个模块分别添加。

表格 4-2 不同模块添加在测试集 Set5、比例为 4 条件下的 PSNR 平均值

多尺度特征融合模块	非局部操作	二阶特征提取操作	PSNR
√	√	√	31.76
√	√	×	31.62
√	×	√	31.50
×	√	√	31.72

其中，符号“×”和“√”分别表示不存在某个模块和存在某个模块。从上表可以看到，网络中添加多尺度特征融合模块、非局部操作和二阶特征提取操作都可以使得实验的结果有所提升，全部模块都采用的实验效果是最好的。

### § 4.6.3 重建结构与对比分析

由于采用了注意力机制将图像的非局部特征和二阶统计特征进行了融合，获取了图像的固有结构信息和高阶图形信息特征，并利用多尺度卷积提取不同尺度下的图像信息，因此本章所提出的算法的 PSNR 平均值与传统的 Bicubic 算法相比有明显的提高，相比于其他基于 CNN 的算法也均有提高，实现了更好的重建效果。不同算法在不同测试集中以 x2、x3、x4 的比例放大时的 PSNR 和 SSIM 值分别如下述表 4-3、表 4-4、表 4-5 所示：

表 4-3 不同算法在测试集 Set5 的 PSNR 和 SSIM 平均值

算法	x2	x3	x4
Bicubic	33.66 / 0.9299	30.39 / 0.8682	28.42 / 0.8104
SRCNN	36.66 / 0.9542	32.75 / 0.9090	30.48 / 0.8628
DRSR	37.41 / 0.9521	33.60 / 0.9168	31.18 / 0.8836
VDSR	37.53 / 0.9587	33.66 / 0.9213	31.35 / 0.8838
本文方法	<b>37.93 / 0.9598</b>	<b>33.98 / 0.9243</b>	<b>31.74 / 0.8872</b>

表 4-4 不同算法在测试集 Set14 的 PSNR 和 SSIM 平均值

算法	x2	x3	x4
Bicubic	30.24 / 0.8688	27.55 / 0.7742	26.00 / 0.7027
SRCNN	32.42 / 0.9063	29.28 / 0.8208	27.49 / 0.7503
DRSR	32.95 / 0.9063	29.69 / 0.8314	27.83 / 0.7624
VDSR	33.03 / 0.9124	29.77 / 0.8314	28.01 / 0.7674
本文方法	<b>33.34 / 0.9136</b>	<b>29.95 / 0.8365</b>	<b>28.28 / 0.7736</b>

表 4-5 不同算法在测试集 BSD100 的 PSNR 和 SSIM 平均值

算法	x2	x3	x4
Bicubic	29.56 / 0.8431	27.21 / 0.7382	25.96 / 0.6675
SRCNN	31.36 / 0.8879	28.41 / 0.7863	26.90 / 0.7101
DRSR	31.55 / 0.8924	28.54 / 0.7917	27.00 / 0.7092
VDSR	31.90 / 0.8960	28.82 / 0.7976	27.29 / 0.7251
本文方法	<b>32.07 / 0.8985</b>	<b>28.98 / 0.7993</b>	<b>27.37 / 0.7316</b>

为了更清晰直观的对比不同算法的重建效果，本章选用测试集 Set14 和 BSD100 下的几张图像以 x2 放大比例下的重建作为对比，如下图 4-6 和图 4-7 所示，图中最左侧的为真实 LR 图像，其余从左到右依次为真实 HR 图像、Bicubic、SRCNN、DRSR、VDSR 以及本章所提方法重建后的图像。



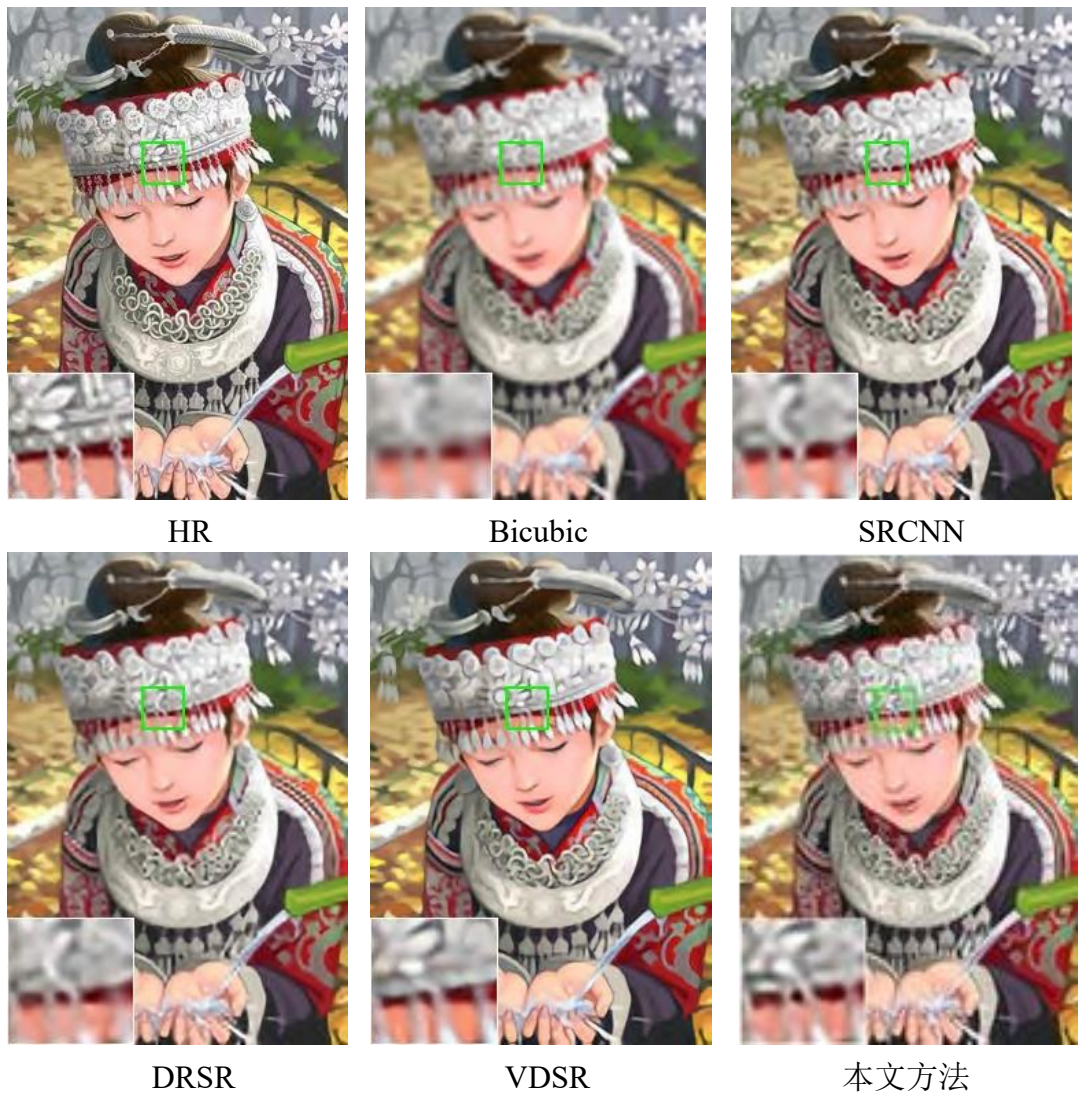


图 4-6 放大比例为 2，Set14 数据集中“girl”的对比结果

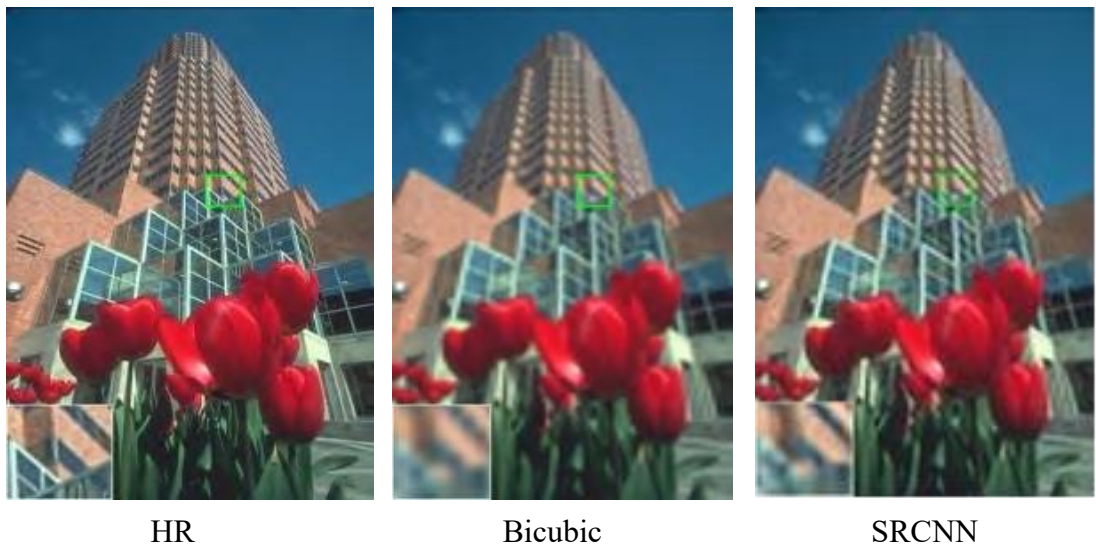






图 4-7 放大比例为 2，BSD100 数据集中“city”的对比结果

由上述两组图片不难发现，Bicubic 算法重建后的得到图像效果是最差的，图像轮廓模糊并且图形的细节信息也存在伪影；SRCNN 算法相较于 Bicubic 算法来说，能够大致看到图像的轮廓和边缘信息，但是与原来的高分辨率图像来说依然存在大量的边缘细节信息没有重建成功；而 DRSR 和 VDSR 相较于本章所提算法，虽然图像质量上差距不大，但是由于本章方法引入了注意力机制和多尺度信息融合，且使用亚像素卷积法进行最终的重建任务，使得本文算法能够获取图像原有的结构信息和高阶图像特征，在结构轮廓方面更加具有优势。因此本章方法在客观指标上取得了更好的结果。

## § 4.7 本章小结

本文提出了一种基于注意力机制与多尺度特征融合的卷积神经网络模型。结合非局部特征和二阶统计特征的优点，构建一个注意力机制模块，将图像的非局部信息和二阶特征信息融合起来，使得我们的网络结构能够学习到图像的固有结构特征信息和高阶特征，同时，利用了不同尺度的卷积核来提取图像的特征，获取更多的图像区域信息并将二者提取的信息进行融合，最后通过亚像素卷积进行图像的重建工作。在公共数据集上的大量实验表明，所提的方法展现出了较好的实验。

## 第五章 总结与展望

### § 5.1 总结

图像分辨率是图像包含信息量多少的重要衡量指标，图像分辨率越高，意味着图像包含的信息量越多。但由于现实环境的复杂和成像硬件条件的限制以及诸多不确定因素的存在，我们通常不能得到满足需求的高分辨率图像，因此在软件方面对图像进行超分辨率重建就显得尤为重要。

基于卷积神经网络的图像超分辨率重建通过卷积神经网络学习低分辨率图像和高分辨率图像之间的对应关系，通过学习训练来获得网络模型。随着近年来卷积神经网络的蓬勃发展，基于卷积神经网络模型的图像超分辨率取得了显著的成效，目前这种方法已经成为图像超分辨率重建领域主流的方法。本文内容主要是围绕基于卷积神经网络的图形超分辨率算法进行研究改进，主要工作如下：

(1) 提出了一种基于多尺度残差密集块的卷积神经网络的图像超分辨率重建算法，该算法主要包含多个多尺度残差密集块，每个多个多尺度残差密集块内都对提取到的图像特征进行密集特征融合和局部残差融合，来增加对以往卷积层的利用，同时在残差密集块内采用多尺度卷积来增加网络的宽度，获取图像的区域信息；然后将浅层特征信息和深层特征信息进行全局残差融合，以保证对低频信息的有效利用，从而使得网络模型学习到整个特征空间相关性的内容，形成了图像特征信息机制的全局记忆。

(2) 提出了一种基于注意力机制与多尺度与多尺度特征融合的图像超分辨率算法，该算法主要由多个注意力机制模块组成，而每个注意力机制模块主要包含非局部特征和二阶统计特征，并将二者提取的特征信息融合，便于获取图像结构信息和高阶图像信息，而后将多个注意力机制模块进行残差信息融合，保证对低频信息的有效利用，同时在多个注意力机制模块的前后添加多尺度特征模块用来提取不同尺度下的图像信息，获取更多的图像局域信息，最后通过亚像素卷积对图像进行重建，得到最终的高分辨率图像。

### § 5.2 展望

图像超分辨率重建任务不仅要按一定比例因子放大图像比例，同时还要保证图像的丰富的高频信息，即从原图像上获取更多有用的信息。本文主要研究的是基于卷积

神经网络的图像超分辨率重建算法，与目前的诸多算法比较，本文提出的算法实验结果虽有不同程度的提升，但是目前的工作内容仍然存在不足与改进的空间。

（1）在基于多尺度残差密集块的卷积神经网络的图像超分辨率重建算法中，为了保证更多的特征信息融合，将大量的特征图进行并联，但该算法只是粗略的将特征信息进行融合，但却并未进行图形纹理细节的分割，因此未来将会在进行更加有效的信息融合方面进行探索。

（2）基于注意力机制与多尺度特征融合的图像超分辨率算法中通过注意力机制模块，可将图像的非局部特征和二阶特征有效的融合起来，并利用了不同尺度的卷积来提取图像的特征，但是如何将非局部特征和二阶特征提取到的特征信息进行更加有效的融合是未来值得探讨的问题。

（3）本文提出的两种重建方法虽然获得了不错的实验效果，但是在面临不同尺度的缩放因子时需要重新训练不同的网路模型，这种做法并不能满足现实生活中人们对此的需求，因此未来工作中将进一步探索能够进行不同比例因子的网络模型，使训练出来的模型能够更加符合现实的需求。

## 参考文献

- [1] 肖进胜,饶天宇,贾茜等.改进的自适应冲击滤波图像超分辨率插值算法[J].计算机学报, 2015, 38(6): 1131-1139.
- [2] Ren C, He X, Teng Q, et al. Single image super-resolution using local geometric duality and non-local similarity[J]. IEEE Transactions on Image Processing, 2016, 25(5): 2168-2183.
- [3] 王世平,毕笃彦,刘坤等.一种多映射卷积神经网络的超分辨率重建算法[J].西安电子科技大学学报 (自然科学版),2018 (04): 155-160.
- [4] Bätz M, Eichenseer A, Seiler J, et al. Hybrid super-resolution combining example-based single-image and interpolation-based multi-image reconstruction approaches[C]. IEEE International Conference on Image Processing. IEEE, 2015: 58-62.
- [5] Kim K I, Kwon Y. Single-image super-resolution using sparse regression and natural image prior[J]. IEEE transactions on pattern analysis and machine intelligence, 2010, 32(6): 1127-1133.
- [6] Xiao J, Liu E, Zhu L. Improved image super-resolution algorithm based on convolutional neural network[J]. Acta Optica Sinica, 2017, 37(3): 0318011.
- [7] Song X, Liu W, Liu J, et al. Deep CNN jointing low-high level feature for image super-resolution[C]. Tenth International Conference on Graphics and Image Processing. International Society for Optics and Photonics, 2019, 11069: 110693O.
- [8] Wang Y, Wang L, Wang H, et al. End-to-end image super-resolution via deep and shallow convolutional networks[J]. IEEE Access, 2019, 7: 31959-31970.
- [9] Irani M, Peleg S. Improving resolution by image registration[J]. GVGIP: Graphical Models and Image Processing, 1991, 53(3): 231-239
- [10] Stark H, Oskoui P. High-resolution image recovery from image-plane arrays, using convex projections[J]. Journal of the Optical Society of America A, 1989, 6(11): 1715-1726.
- [11] Schultz R R, Stevenson R L. Extraction of high-resolution frames from video sequences[J]. IEEE transactions on image processing, 1996, 5(6): 996-1011
- [12] Dong C, Loy C C, He K, et al. Learning a deep convolutional network for image super-resolution[C]. European conference on computer vision. Springer, Cham, 2014: 184-199.
- [13] 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, 2017.
- [14] Fergus R, Taylor G W, MD Zeiler. Adaptive deconvolutional networks for mid and high level feature learning[C]// International Conference on Computer Vision. IEEE Computer Society, 2011.

- 
- [15] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [16] Kim J, Kwon Lee J, Mu Lee K. Accurate image super-resolution using very deep convolutional networks[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1646-1654.
- [17] Simonyan K, Zisserman. A Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.
- [18] Kim J, Lee J K, Lee K M. Deeply-Recursive Convolutional Network for Image Super-Resolution[J]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [19] Zhang Y, Li K, Li K, et al. Image super-resolution using very deep residual channel attention networks[C]. Proceedings of the European Conference on Computer Vision. 2018: 286-301.
- [20] Guo D, Niu Y, Xie P. Speedy and accurate image super-resolution via deeply recursive CNN with skip connection and network in network[J]. IET Image Processing, 2019, 13(7): 1201-1209.
- [21] Huang G, Liu Z, Laurens V, et al. Densely Connected Convolutional Networks[J]. IEEE Computer Society, 2016.
- [22] Tong T, Li G, Liu X, et al. Image Super-Resolution Using Dense Skip Connections[C]// IEEE International Conference on Computer Vision. IEEE Computer Society, 2017.
- [23] Tai Y, Yang J, Liu X, et al. MemNet: A Persistent Memory Network for Image Restoration[C]// IEEE International Conference on Computer Vision. IEEE Computer Society, 2017.
- [24] Zhang Y, Tian Y, Kong Y, et al. Residual Dense Network for Image Super-Resolution[J]. IEEE, 2018.
- [25] Lai W S, Huang J B, Ahuja N, et al. Deep laplacian pyramid networks for fast and accurate super-resolution[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 624-632.
- [26] Ren C X, Dai D Q, Yan H. Robust classification using  $\ell_2$ , 1-norm based regression model[J]. Pattern Recognition, 2012, 45(7): 2708-2718.
- [27] Ahn N, Kang B, Sohn K A. Fast, accurate, and lightweight super-resolution with cascading residual network[C]. Proceedings of the European Conference on Computer Vision. 2018: 252-268.
- [28] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution[C]. European conference on computer vision. Springer, Cham, 2016: 694-711.
- [29] Gatys L, Ecker A S, Bethge M. Texture synthesis using convolutional neural networks[C]. Advances in neural information processing systems. 2015: 262-270.
- [30] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[J]. arXiv preprint arXiv:1502.03167, 2015.
- [31] Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution[C]. Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2017:

- [32] 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.
- [33] Rafferty J, Shellito P, Hyman N H, et al. Practice parameters for sigmoid diverticulitis[J]. Diseases of the colon & rectum, 2006, 49(7): 939-944.
- [34] Malfliet W, Hereman W. The tanh method: I. Exact solutions of nonlinear evolution and wave equations[J]. Physica Scripta, 1996, 54(6): 563.
- [35] Hara K, Saito D, Shouno H. Analysis of function of rectified linear unit used in deep learning[C]. International Joint Conference on Neural Networks. IEEE, 2015: 1-8.
- [36] Zhang X, Zou Y, Shi W. Dilated convolution neural network with LeakyReLU for environmental sound classification[C]. 22nd International Conference on Digital Signal Processing. IEEE, 2017: 1-5
- [37] QingJie W, WenBin W. Research on image retrieval using deep convolutional neural network combining L1 regularization and PRelu activation function[C]. IOP Conference Series: Earth and Environmental Science. IOP Publishing, 2017, 69(1): 012156.
- [38] Xu B, Wang N, Chen T, et al. Empirical evaluation of rectified activations in convolutional network[J]. arXiv preprint arXiv: 1505.00853, 2015.
- [39] 杨鹤标, 龚文彦. 基于卷积神经网络的反向传播算法改进[J]. 计算机工程与设计, 2019, 40(01): 126-130.
- [40] Buades A, Coll B, Morel J M. A non-local algorithm for image denoising[C]// Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE, 2005.
- [41] Wang X, Girshick R, Gupta A, et al. Non-local Neural Networks[J]. 2017.
- [42] Jie H, Li S, Gang S. Squeeze-and-Excitation Networks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018.
- [43] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional Block Attention Module[J]. Springer, Cham, 2018.
- [44] Zhang Y, Li K, Li K, et al. Image Super-Resolution Using Very Deep Residual Channel Attention Networks. 2018.
- [45] Dai T, Cai J, Zhang Y, et al. Second-order Attention Network for Single Image Super-Resolution[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2019.
- [46] Shi W, Caballero J, F Huszár, et al. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network[J]. IEEE, 2016.
- [47] Bottou L. Stochastic Gradient Descent Tricks[M]. Neural Networks Tricks of the Trade. Springer Berlin Heidelberg, 2012: 421-436.
- [48] Lu H, Jesmanowicz A, Li S J. Momentum-weighted conjugated gradient descent algorithm for

- gradient coil optimization[J]. Magnetic Resonance in Medicine, 2004, 51(1): 158-164.
- [49] Kingma D P, Ba J. Adam: A method for Stochastic Optimization[J]. Computer Science, 2014, 11(22): 87-102.
- [50] H.R.Sheikh, M.F.Sabir, A.C.Bovik, et al. "A statistical evaluation of recent full reference image quality assessment algorithms," IEEE Transactions on Image Processing, vol. 15, 2006.
- [51] 肖祥元, 景文博, 赵海丽. 基于峰值信噪比改进的图像增强算法[J]. 长春理工大学学报(自然科学版), 2017, 40(04): 83-86+92.
- [52] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE transactions on image processing, 2004, 13(4): 600-612.
- [53] Agustsson E, Timofte R. Ntire 2017 challenge on single image super-resolution: Dataset and study[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2017: 126-135.
- [54] Bevilacqua M, Roumy A, Guillemot C, et al. Single-image super-resolution via linear mapping of interpolated self-examples[J]. IEEE Transactions on image processing, 2014, 23(12): 5334-5347.
- [55] Timofte R, De Smet V, Van Gool L. A+: Adjusted anchored neighborhood regression for fast super-resolution[C]//Asian conference on computer vision. Springer, Cham, 2014: 111-126.
- [56] Martin D R, Fowlkes C C, Malik J. Learning to detect natural image boundaries using local brightness, color, and texture cues[J]. IEEE transactions on pattern analysis and machine intelligence, 2004, 26(5): 530-549.
- [57] Dong W, Lei Z, Shi G. Nonlocally Centralized Sparse Representation for Image Restoration[C]//IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society. IEEE, 2013.
- [58] Li P, Xie J, Wang Q, et al. Is Second-Order Information Helpful for Large-Scale Visual Recognition[C]//2017 IEEE International Conference on Computer Vision (ICCV). IEEE, 2017.
- [59] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional Block Attention Module[J]. Springer, Cham, 2018.

## 致谢

时间真是个神奇的东西，身处其中浑然不觉，然而一回首却往事如烟。时光匆匆，三年的研究生生涯转眼间到了说离别的时候，随之而去的还有青春，许多的美好回忆。三年的时光我收获良多、感触良多，在此我把我最真诚的感谢送给所有帮助关心我的人。

首先要感谢的是我的导师刘建明老师，在短短的三年研究生涯中，刘老师在生活学习中多个方面给与了我尽心的指导和帮助。虽然刘老师公务繁忙，但在我们遇到困难问题的时候，他总是尽量抽出时间帮我们分析问题，提出建议。刘老师的宽容、大格局、渊博的知识、为人处世的生活态度、严谨的科研态度都是我终身学习的榜样，刘老师的教诲我都铭记在心，终生受用。

其次要感谢实验室团队的张华成老师、林科老师、崔更申老师。他们在实验室工程项目上认真负责，一丝不苟的工作态度深深的影响了我，非常感谢三年的研究生涯中你们的无私指导和帮助。

在生活和学习中，还要感谢实验中的师兄、师姐、同门、师弟和师妹们，谢谢你们给了我一个温馨祥和的实验室学习环境，使得三年求学生涯并不乏味孤单。此外还要感谢我的室友们，三年的生活互相关照，互相帮助，谢谢你们了。

最后，要感谢我的父母对我学业的支持与鼓励，他们一直在身后的默默付出与支持一直是我学习与进步的不懈动力。



## 作者在攻读硕士期间的主要研究成果

### 学术论文

- [1] 刘建明, 朱朝辉. 基于多尺度残差密集块的图像超分辨率重建. (中文核心在投)

### 软件著作

- [1] 刘建明, 朱朝辉. 基于卷积神经网络的图像超分辨率重建系统 V1.0. 登记号: 2020SR0240646. 2020-6-11.