



南京大学

本科生毕业论文 (申请学士学位)

论 文 题 目 基于深度学习的图像超分辨率算法研究

作 者 姓 名 薛雨婷

学科、专业方向 计算机科学与技术系

研 究 方 向 计算机视觉

指 导 教 师 唐杰 副教授

2019 年 5 月 10 日

学 号：**151220138**

论文答辩日期：**2019** 年 **5** 月 **xx** 日

指 导 教 师： (签字)

Research on Image Super-Resolution based on Deep Learning

by

XUE Yu-Ting

Supervised by

Associate Professor TANG Jie

A dissertation submitted to
the graduate school of Nanjing University
in partial fulfilment of the requirements for the degree of

BACHELOR

in

Computer Application



Department of Computer Science and Technology
Nanjing University

May 10, 2019

南京大学本科生毕业论文中文摘要首页用纸

毕业论文题目： 基于深度学习的图像超分辨算法研究

计算机科学与技术系 专业 2015 级学士姓名： 薛雨婷
指导教师（姓名、职称）： 唐杰 副教授

摘 要

图像超分辨是目前计算机视觉领域中一个非常热门的问题,其任务是从一张低分辨率的图像重建出一张高分辨率的图像。近年来随着深度学习的迅速发展和卷积神经网络的广泛运用,许多图像超分辨的研究都着眼于设计基于卷积神经网络的图像超分辨算法。为了使卷积神经网络模型取得更好的性能,加深或者加宽网络结构成为了一个设计趋势,然而同时也带来了参数量和计算量的增加以及大量的内存消耗。本文主要探究注意力模块的设计及其与深度超分辨模型的联合机制。通过将注意力机制加入卷积网络中,我们能够得到更加高效的网络设计,在提升网络性能的同时减少了计算量和参数量。参考 EDSR^[1] 算法的基础结构,通过改进其中的残差模块,可以得到不同的基于注意力机制的超分辨模型:

- (1) 仅将通道注意力加入了卷积网络中,通过将 EDSR 结构中的残差模块全部替换为 CAB 模块得到 CAN 结构;
- (2) 仅将空间注意力加入了卷积网络中,通过将 EDSR 结构中的残差模块全部替换为 SAB 模块得到 SAN 结构;
- (3) 将通道注意力与空间注意力机制结合起来加入卷积网络中,通过将 EDSR 结构中的残差模块全部替换为 CASAB 模块得到 CASAN 结构;
- (4) 加入不同尺度上的通道与空间注意力相结合的注意力机制到卷积网络中,通过将 EDSR 结构中的残差模块全部替换为 OAB 模块得到 OAN 结构。

本文将从这几个网络结构入手,通过大量的实验对比分析他们的优劣,最后运用不同尺度上的注意力机制与卷积神经网络相结合所得模型结构来解决图像超分辨的问题。

关键词：深度学习; 注意力机制; 图像超分辨

南京大学本科生毕业论文英文摘要首页用纸

THESIS: Research on Image Super-Resolution
based on Deep Learning
SPECIALIZATION: Computer Application
POSTGRADUATE: XUE Yu-Ting
MENTOR: Associate Professor TANG Jie

Abstract

Image super-resolution is a very popular problem in the field of computer vision. Its task is to reconstruct a high-resolution image from a low-resolution image. In recent years, with the rapid development of deep learning and the widespread use of convolutional neural networks, many image super-resolution studies have focused on designing image super-resolution algorithms based on convolutional neural networks. In order to achieve better performance of the convolutional neural network model, deepening or widening the network structure has become a design trend, but it also brings about an increase in the amount of parameters and calculations and a large amount of memory consumption. This paper mainly explores the design of attention block and its joint mechanism with deep super-resolution model. By incorporating attention mechanisms into the convolutional network, we are able to get more efficient network designs that improve network performance while reducing the amount of computation and parameters. Referring to the basic structure of the EDSR^[1] algorithm, different attention mechanism models are obtained by improving the residual block:

- (1) Only the channel-wise attention is added to the convolutional network, and the CAN structure is obtained by replacing all the residual blocks in the EDSR structure with the CA blocks.
- (2) Only the spatial attention is added to the convolutional network, and the SAN structure is obtained by replacing all the residual blocks in the EDSR structure with the SA blocks.
- (3) The channel-wise attention and spatial attention mechanism are combined into the convolution network, and the CASAN structure is obtained by replacing all the residual blocks in the EDSR structure with the CASA blocks.

- (4) The attention mechanism on different scales which combined with the channel-wise attention and spatial attention is added to the convolutional network, and the OAN structure is obtained by replacing all the residual blocks in the EDSR structure with the OA blocks.

This paper will start with these several network structures, analyze their advantages and disadvantages through a large number of experiments, and finally use the attentional mechanism of different scales and the convolutional neural network to solve the problem of image super-resolution.

keywords: Deep Learning, Attention Mechanism, Single Image Super-resolution

目 次

目 次	9
插图清单	11
附表清单	13
1 绪论	1
1.1 研究背景及意义	1
1.2 研究现状及发展趋势	2
1.3 本文主要研究内容及章节安排	4
2 相关工作	7
2.1 现有算法简介	7
2.2 EDSR 算法结构	10
2.2.1 EDSR 算法简介	10
2.2.2 残差模块	11
2.2.3 EDSR 算法模型结构	12
2.3 注意力机制	12
2.4 Squeeze-and-Excitation 模块	14
2.5 不足与可改进之处	15
2.6 本章小结	15
3 改进的模型结构	17
3.1 CASAB 模块	17
3.1.1 Channel-wise Attention (CA) 单元	18
3.1.2 Spatial Attention (SA) 单元	19
3.2 CAN 模型	20
3.3 SAN 模型	21
3.4 CASAN 模型结构	22

3.5	OAN 模型	24
3.5.1	OAB 模块	24
3.5.2	OAN 模型结构	25
3.6	本章小结	26
4	实验与分析	27
4.1	数据集与评估标准	27
4.2	实验设置及实现细节	27
4.3	参数选择与分析	28
4.4	不同模型效果分析	29
4.5	与现有算法的比较与分析	32
4.6	本章小结	33
5	总结与展望	35
	致 谢	37
	参考文献	39

插图清单

1-1	低分辨率图像与高分辨率图像的联系 ^[4]	1
1-2	图像超分辨问题中的深度神经网络 ^[5]	3
2-1	ESPCN 模型网络 ^[6]	8
2-2	SRCNN、FSRCNN、VDSR、DRCN 和 LapSRN 的网络结构对比。红色箭头表示卷积层；蓝色箭头表示反卷积（上采样）；绿色箭头表示异或运算；橙色箭头表示递归层。 ^[7]	10
2-3	原始 ResNet、SRResNet 和改进后的残差模块对比 ^[1]	11
2-4	EDSR 算法模型结构 ^[1]	12
2-5	注意力机制 ^[8]	13
2-6	Squeeze-and-excitation 网络结构 ^[2]	15
3-1	CSAR 模块结构示意图 ^[3]	17
3-2	通道间注意力操作（CA 单元） ^[3]	18
3-3	空间注意力操作（SA 单元） ^[3]	20
3-4	CAN 模型结构示意图	20
3-5	CAB 模块结构示意图	21
3-6	SAN 模型结构示意图	21
3-7	SAB 模块结构示意图	22
3-8	CASAN 模型结构示意图	23
3-9	OAN 模型结构示意图	24
3-10	OAB 模块结构示意图	25
4-1	原始 HR 图片与 EDSR 算法、CAN 算法、SAN 算法、CASAN 算法及 OAN 算法恢复出的图片局部效果展示	30

附表清单

- 4-1 基于 CASAN 模型在 2 倍规模上的不同参数实验结果。在配置 (Set) 一列中, RB 表示残差块的个数 (n_resblocks), F 表示特征纬度 (n_feats), B 表示批处理参数 (batch_size), P 表示输入图像块大小 (patch_size), RE 表示下降率 (reduction)。四个测试集上的 PSNR 分数一栏中, 得分最高的用 **highlighted** 表示, 得分第二高的用 underlined 表示。..... 28
- 4-2 基础模型 EDSR 与改进后模型 CAN、SAN、CASAN 和 OAN 的实验表现结果。四个测试集上的 PSNR 分数一栏中, 得分最高的用 **highlighted** 表示, 得分第二高的用 underlined 表示。..... 31
- 4-3 OAN 模型与当下现有其它超分辨率算法模型效果对比。四个测试集上的 PSNR 分数一栏中, 得分最高的用 **highlighted** 表示, 得分第二高的用 underlined 表示。..... 32

第一章 绪论

1.1 研究背景及意义

图像超分辨是目前计算机视觉领域中一个非常热门的问题，其任务是从一张低分辨率的图像重建出一张高分辨率的图像。图像超分辨在图像处理方面的应用十分广泛，在大量的电子图像应用中，高分辨率的图像的需求量很大且需求很频繁。图像超分辨技术可用于安防监控、医疗图像、HDTV 和卫星图像等需要图像中高频细节的场景。

许多研究中都假设低分辨率图像是高分辨率图像的双三次降采样版本，但其他的退化因子如模糊、信息提取或是噪声等在实际应用中也同样应该被考虑到。广义的超分辨问题将低分辨率数据视为经过了低通滤波与降采样并加入噪声的高分辨率数据（如图 1-1）。由于任何的低分辨率输入都存在着多种超分辨方案，即从低分辨率到高分辨空间是一个一对多的映射关系，它的正确的解集是非平凡的，所以图像超分辨实际上是一个不确定问题。当尺度上推的因子很大的时候，从低分辨率图像中重建出高频细节将变得非常有难度。

由于许多低分辨率技术都假设大部分高频数据都是冗余的，并且可以从部分低频数据中准确地预测得来，因此从低分辨率图像大片区域中收集到的有效有用的上下文信息便格外重要，解决该问题需要从上下文信息中捕捉到充分的知识用于重构高分辨图像中的高频细节。为了解决这样的不确定问题，大量的学习算法被运用来研究低分辨率与高分辨图像对之间的映射关系。

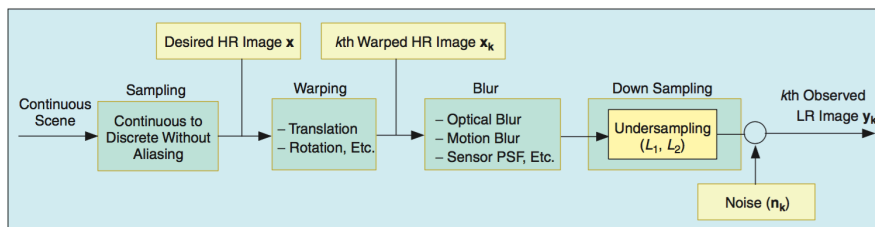


图 1-1: 低分辨率图像与高分辨率图像的联系^[4]

1.2 研究现状及发展趋势

由于有无穷多数量的分辨率图像可以通过下采样得到相同的低分辨率图像，因此图像超分辨问题是一个不确定问题，并没有唯一解。目前已经有大量的图像超分辨方法被提出，包括基于插值的方法（interpolation-based methods），基于重构的方法（reconstruction-based methods）和基于实例的方法（example-based methods）。

基于插值的方法通过使用图像相临近位置的像素点的数值作为计算对象，通过插入算子估计出来像素数值，从而重构出高分辨率图像。基于插值的方法关键即在于所插入的算子函数的选择，目前现有的插值算子的函数有最近邻插值函数、线性插值函数和样条插值函数等。一个好的算子函数有助于快速重构出高分辨率图像。然而，由于插值算法仅仅只能够从相邻的像素值之间的到有用信息来重构高分辨率图像，而无法利用图像中本身的纹理、区域信息、上下文等细节信息，故在实际的应用中一般不会采用该类基于插值的方法。

鉴于基于插值的方法的局限性，H. Demirel^[9] 在 2011 年提出了基于小波变换的小波域方法来改进插值算法。通过小波变换来实现空间和频率的局部变换，从而提取出有效信息来重构高分辨率图像。基于重构的超分辨方法主要针对图像中空间信息的增强和重构，在一些特定的情况下能够取得不错的效果。但该方法仍然没能充分利用输入低分辨率图像的先验知识，同时经常在大尺度因子上的表现很差，所以在实际的应用当中也并不理想。

鉴于前两种方法的实际运用中的局限性，以及近年来深度学习算法的迅速发展和卷积神经网络的优越性能，许多图像超分辨方法研究都着眼于基于实例学习的方法。很多基于卷积神经网络的图像超分辨算法都取得了不俗的表现，同时许多人在尝试沿着这个思路用更深的网络来学习出更好的超分辨模型。自从 SRCNN^[10] 算法被提出来预测双三次上采样得到的图像与高分辨率图像之间的非线性关系之后，大量的卷积神经网络结构被用于研究超分辨问题。由于更深的卷积神经网络结构拥有更大的感受野来捕捉到更多的上下文信息，VDSR^[11] 与 DRCN^[12] 这两种分别使用了全局残差学习和递归层来提升超分辨准确率的深度网络结构被提出。为了控制模型的参数量并且维持持久记忆，DRRN^[13] 模型使用了带有全局与局部残差学习的递归模块，并且在 MemNet 中设计了含有密集连接的记忆模块。在这些算法中，低分辨率输入图像都需要在进入网络结构前先进行双三次插值处理来达到所需要的尺寸，然而双三次插值

的预处理将不可避免地提升计算复杂度并且可能会产生新的噪声。为了使卷积神经网络实现更好的性能，加深或者加宽网络结构成为了一个设计趋势，然而随之带来的参数量和计算量的增加，以及大量的计算和内存消耗，都使得这些方法难以被直接运用到实际应用中。

为了缓解巨大的计算负担，克服预处理上采样结构的弊端，Dong 等人^[14]提出了一种利用反卷积操作在网络结构末端放大图像尺寸的算法。之后 Shi 等人^[6]提出了一种更加有效的次像素卷积层来代替反卷积层，这一算法后来通过在模块中采用残差学习和多路径级联被进一步扩充为 EUM 算法^[14]。越来越多的深度网络（例如 SRResnet^[15]，EDSR^[1] 和 SRDenseNet^[16]）通过采用这种在网络结构末端再放大图像尺寸的策略，都在较小的计算负担下取得了较好的表现。RDN 算法^[17]中使用了紧密连接卷积层来提取大量本地特征，并且采用局部和全局特征聚变操作来自适应地在低分辨率空间融合层次结构特征。

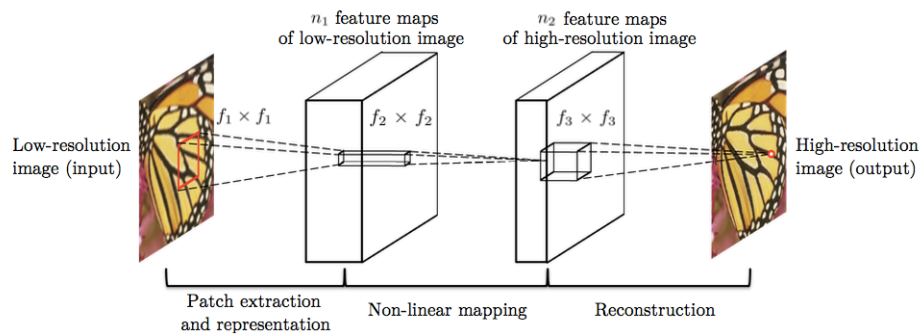


图 1-2: 图像超分辨率问题中的深度神经网络^[15]

在 VGGNets^[18] 算法中显示了增加网络的深度可以显著提升其学习能力的表现效果。通过调节每一层中输入的分配，批归一化操作^[19]增加了深度网络中学习过程的稳定性。在这些工作的基础上，ResNets 表现出了学习适当更深层次的可能性，并尝试通过使用跳跃连接来加强网络模型^{[20][21]}。高速路神经网络结构^[22]很好地介绍了一种门机制来规范近路连接中的信息流。沿着这些工作的思路，还会有更多不同形式的网络层之间的连接方式，给其代表的深度网络结构的学习效果带来显著的效果提升。

另一种密切相关的研究方向集中在提升一个网络结构中包含的可计算元素的功能形式的方法上。在先前的工作中，无论是在独立的空间结构上，还是在由 1×1 卷积核构成的标准卷积滤波器连接而成的结构上，跨通道连接都被认为是典型的特征值之间的新的组合方式。这类研究大多都侧重于简化模型或是降低计算复杂度，这反映了通道关系可以被制定为由局部感受野中与实例无关的

功能的组合这一猜想。

除此之外，传统的卷积神经网络方法通常采用拓扑级联网络，例如 VDSR^[11] 和 DRCN^[12]，在这些方法中，每一层的特征图都被无区别的送入网络，却忽视了特征图本身之间的关系和所包含信息量的差异。针对这一问题，Hu 等人^[3] 尝试将基于通道（channel-wise）的注意力机制和基于空间（spatial-wise）的注意力机制结合起来加入卷积神经网络中，自适应地再校准特征响应来提升网络的表示能力。Hu 等人在文章^[2] 中则认为，提供一个通过使用全局信息、带有能够简化模型动态性机制和各通道间非线性依赖性的单元结构，更加能够简化学习过程，并且显著增强网络的表现能力。

1.3 本文主要研究内容及章节安排

沿着将通道注意力机制和空间注意力机制相结合来加入卷积神经网络以提升网络性能的思路，本文从将注意力机制与深度学习结合起来解决图像超分辨问题的角度进行研究。参考 EDSR^[1] 算法网络结构，以 EDSR 网络结构框架为基础，改进其中的残差模块，得到不同的注意力机制模型。

通过将 EDSR 网络结构中的残差模块全部替换为 CAB 模块得到 CAN 结构，在 CAN 结构中仅将通道注意力加入了卷积网络中；通过将 EDSR 网络结构中的残差模块全部替换为 SAB 模块得到 SAN 结构，在 SAN 结构中仅将空间注意力加入了卷积网络中；而通过将 EDSR 网络结构中的残差模块全部替换为 CASAB 模块得到 CASAN 结构，在 CASAN 结构中则将通道注意力与空间注意力机制结合起来，一并加入卷积网络中；最终在 CASAN 结构上进一步优化，加入了不同尺度上的注意力机制，得到 OAN 结构。本文将从这几个网络结构入手，对比分析用注意力机制与卷积网络相结合所得模型结构来解决图像超分辨问题的优劣。

本文一共分为 5 个章节，主要章节安排如下：

第一章 绪论。描述了图像超分辨问题的研究背景及意义，阐述了图像超分辨的研究现状和发展趋势，初步介绍本文研究内容和算法，给出本文主要结构。

第二章 相关工作。介绍了现有较主流的基于深度学习来解决图像超分辨问题的算法，着重介绍了本文主要参考的 EDSR 算法框架；介绍了注意力机制和 Squeeze-and-Excitation 模块；分析了现有算法的部分局限性。

第三章 改进的模型结构。介绍了本文在 EDSR 模型结构基础上改进提出的 CAN 模型、SAN 模型、CASAN 模型和 OAN 模型。

第四章 实验与分析。在 CASAN 模型上进行不同参数的实验对比，得到最佳参数；将 EDSR 模型和提出的 4 种改进模型进行实验对比，得到最佳模型；将最佳参数下的最佳模型与现有算法模型进行对比分析。

第五章 总结与展望。总结文章整体的内容，分析展望文章未来可以继续深入研究优化的方向。

第二章 相关工作

2.1 现有算法简介

SRCNN 算法^[5] 近年来,深度学习(DL)尤其是卷积神经网络(CNN)凭借其强大的学习能力,被广泛用于解决超分辨这一不适定问题。SRCNN 算法作为深度学习在超分辨率重建问题上的开山之作,它凭借比较简单的网络结构便取得了比当时最新技术更高的准确率,同时由于其使用的滤波器和卷积层的数量较少,它在实际应用实践中的速度也较快。

SRCNN 算法通过对输入的低分辨图像进行双三次插值,从而将低分辨率图像放大为目标尺寸作为图像输入,然后通过对图像块特征的提取和表示、特征非线性映射和图像重构这三次卷积操作,最终使用均方误差(MSE)作为损失函数。在 SRCNN 网络结构中,卷积核的数量越多、大小越大,所取得的实验效果就越好;然而网络层次则并非越深越好,由于没有使用池化层和全连接层,模型对初始参数和学习率十分敏感。

FSRCNN 算法^[14] FSRCNN 算法主要为对 SRCNN 的改进,它首先将 SRCNN 中对初始数据进行的双三次插值预处理替换成了在网络最后插入一个反卷积层。即将原始的低分辨图像直接输入到网络中,最终再通过反卷积层放大尺寸,从而将计算复杂度由高分辨图像的尺寸降低为低分辨图像的尺寸。同时 FSRCNN 算法改变了特征维数,通过使用数量更多、规格更小的映射层来替代单个大规模的映射层,从而节省计算时间并提升计算结果。FSRCNN 算法还采用了更小的滤波尺寸和更少的滤波,从而将冗余的参数去掉,来共享其中的映射层,当需要训练不同上采样倍率的模型时,该算法只需要对最后的反卷积层进行微调即可,这样节省了计算时间。

ESPCN 算法^[6] ESPCN 算法也是省去了 SRCNN 算法中一开始对数据进行的双三次插值预处理。双三次插值处理放大了目标尺寸,从而带来的额外计算量,同时插值算法处理的好坏也会对实验结果产生直接的影响,故 ESPCN

算法选择了单张图像超分辨（SISR）技术作为替代，该算法尝试从原始单个低分辨率数据中学习隐式的冗余信息，从而重建丢失的高分辨率信息（如图 2-1）。ESPCN 算法的核心是亚像素卷积层，与传统算法的反卷积计算相比，它能够在训练时间内通过非常小的附加运算开销便将低分辨率数据超分辨率化。

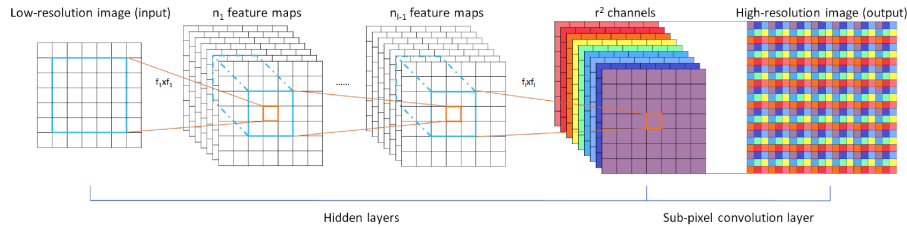


图 2-1: ESPCN 模型网络^[6]

VDSR 算法^[11] VDSR 算法针对 SRCNN 算法的三个方面的不足进行了改进，即 SRCNN 算法过于依赖小块图像区域上的上下文信息、训练收敛的非常慢以及网络仅能用于单一尺寸。VDSR 算法通过加深网络结构和使用更大的感受野，从而获得更多、更大图像区域的上下文信息；该算法通过使用残差学习神经网络和梯度下降来获得极高的学习率，从而加速网络收敛；同时由于为所有可能需要的场景训练并存储大量不同尺度的模型是不切实际的，所以该算法找到一个能够进行多尺度的图像超分辨重构的网络。

总之，VDSR 算法期望获得一个基于非常深的卷积网络的高准确率高分辨模型，并通过残差学习和梯度下降来解决高学习率下高效收敛速率带来的梯度爆炸问题，并将算法延伸为具有可以在一个单独的网络中解决多种尺度的超分辨问题的能力。

DRCN 算法^[12] DRCN 算法充分多地重复使用相同的卷积层，同时避免在多次递归后参数量的上涨。该算法通过使用递归监督，并且从输入数据到卷积层之间使用跳跃连接，从而来避免梯度爆炸/消失问题。跳跃连接方法在输入和输出数据高度相关时尤为有效。

DRRN 算法^[13] 由于非常深的网络结构通常会带来大量的参数，与简单的模型相比，大模型会需要更多的存储空间并更难以适用于移动设备上。DRRN 算法正是为了解决这一问题而诞生的，它提出了一种建立在一个非常深

的高效的网络结构上的方法。该方法相比于 VDSR 算法和 DRCN 算法，在需要更少的参数的同时又能取得更好的结果。

DRRN 算法的创新之处主要在两个方面。一是采用了全局与局部残差学习相结合的方式，在一个非常深的网络结构中，由于大量的图像细节可能在多层运算后便被丢失了，这将会引发梯度问题。而引入强化残差单元结构，又称为多重路径模型的局部残差学习，使得特征分支将大量丰富的图像细节带去下层结构，同时也有利于梯度流；二是残差单元中的多权重递归学习使得模型更加简洁，DRRN 模型有一个包含了许多残差单元的递归模块，其中残差单元分享不同的权重集合。同时为了解决梯度爆炸/消失问题，DRRN 算法设计了一个多路径结构的递归模块，从而节约了 DRCN 算法中对每个递归进行的监督的开销。

LapSRN 算法^[7] LapSRN 算法根据作者总结出的对先前算法的三个主要问题进行了改进，即先前算法中网络结构相对较小，由于其有限的网络能力，其并不能很好的完全学习 LR 到 HR 之间的映射关系；先前算法多使用 L2 型损失函数，但这会使得重构过于平滑，不太符合人类的视觉从而不可避免地产生模糊的预测；最后，大多数现有算法都是在一步上采样操作中重构出高分辨图像，这会提升在获得大采样因子时候的训练复杂度。

LapSRN 算法采用多级的结构，每一级首先采用一串卷积层来提取特征映射，然后用一个反卷积层来使图形尺寸上采样 2 倍，最后使用一个卷积层来预测次级残差。在每一级上预测的残差值都将被用于重建高分辨图像的上采样及其他操作（如图 2-2）。通过这些改进，LapSRN 算法获得比先前现有算法更高的准确率、更快的运算速度以及更优的重构方式。

SRDenseNet 算法^[16] SRDenseNet 算法使用了稠密串联卷积网络，它加强了信息在网络中的流通，缓解了梯度消失问题。同时，它允许复用上一层的特征，避免重复学习繁冗的特征。与先前算法不同的是，SRDenseNet 使用了跳跃连接，将低层次卷积层的输出特征和高层次卷积层的输出特征结合起来，为低分辨的重建提供丰富的信息。重构层也被整合用于重建图像的细节和加速重构进程。由于不同深度层之间的信息是互补的，所以低层卷积的输出特征和高层卷积的输出特征的结合能够使超分辨重建算法的性能得到提升。

SRResNet 算法成功地解决了时空开销过大的问题，但它仅仅使用了未加修

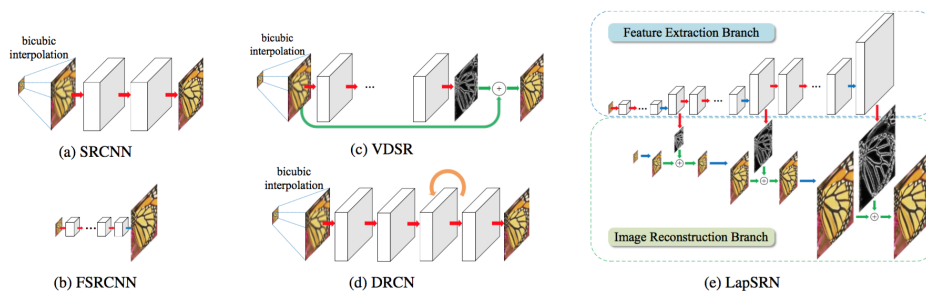


图 2-2: SRCNN、FSRCNN、VDSR、DRCN 和 LapSRN 的网络结构对比。红色箭头表示卷积层；蓝色箭头表示反卷积（上采样）；绿色箭头表示异或运算；橙色箭头表示递归层。^[7]

改的 ResNet 结构。由于 ResNet 最初是被提出来解决更高层次的计算机视觉问题，例如图像分类和检测问题的，所以直接将 ResNet 复用于低层次的计算机视觉问题如超分辨问题的效果可能不太理想。

2.2 EDSR 算法结构

2.2.1 EDSR 算法简介

深度神经网络被广泛用于超分辨问题领域的同时，也展现出其在结构最优性问题上的局限性。首先，神经网络中重构操作的表现对细小结构上的变化都十分敏感。其次，同一模型上的不同的初始化方式和训练技术都会导致不同的表现结果。所以模型结构的设计和最优方案方式的选择在神经网络的训练中都十分重要。

同时，大多数现存的超分辨算法都将不同尺度因子上的超分辨问题看作各自独立的不同的问题，而忽视了思考超分辨问题中不同尺度中的内在联系关系。例如，很多算法中都需要单独训练许多特定尺度的网络来解决不同尺度的问题。尽管 VDSR 算法提供了一个可以解决多种尺度问题的单一网络，但它的计算时间开销和空间开销却也大大高于其他特定尺度的算法。

EDSR^[1] 算法在 SRResNet 结构^[16] 的基础上，首先移除了一些不必要的模块来简化网络结构。当一个模型过于复杂的时候，训练该网络的过程将是不平凡的。不仅如此，EDSR 算法还研究了一个能够将在不同尺度模型上训练的信息转移的模型训练方法。为了在训练过程中使用尺度独立的信息，EDSR 算法

通过预训练的低尺度模型来训练高尺度模型。同时 EDSR 算法还提出一个新的多尺度模型，从而在不同尺度下共享绝大多数的参数。该多尺度模型相较于先前现有的单尺度模型显著降低了参数量，同时也取得了相当优秀的表现成绩。

2.2.2 残差模块

残差模块在解决计算机视觉领域问题方面表现不凡，SRResNet 算法中首次将残差模块应用于解决图像超分辨问题，而在 EDSR 算法中则更进一步改进了残差网络结构来提升其表现结果。

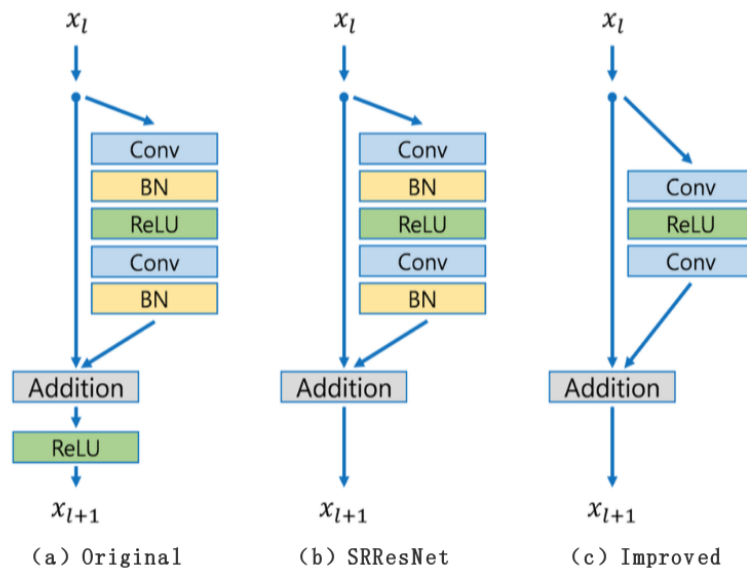


图 2-3: 原始 ResNet、SRResNet 和改进后的残差模块对比^[1]

如图 2-3 所显示，EDSR 算法中的残差模块移除了原始 ResNet 和 SRResNet 结构中的批标准化层，避免批标准化操作将特征值标准化后会降低网络的灵活度。同时由于批标准化层需要占用与它之前的卷积层相当的存储空间开销，因此移除批标准化层后算法使用的 GPU 内存也大大降低。实验^[1]表明，改进后的残差模块模型比原始的 SRResNet 模型节省了将近 40% 的内存，因此在有限的计算资源下，改进后的残差模块算法可以建立一个更大的模型来提升计算结果。

2.2.3 EDSR 算法模型结构

提升一个网络模型性能最简单的方法便是增加其参数数量，而在卷积神经网络中，可以通过堆叠很多层或者增加滤波器的数量来提升模型性能。一个拥有深度（层数） B 和宽度（特征通道数） F 的 CNN 结构大约需要 $O(BF)$ 的存储空间和 $O(BF^2)$ 的参数数量。因此，在有限的计算资源下增大宽度 F 而不是深度 B 可以最大限度地提升模型的性能。

然而，在增加特征图的数量超过了某个特定阈值后，将会造成训练过程中数值上的不稳定。为了解决这一问题，EDSR 算法采用因子为 0.1 的残差尺度，在每个残差模块中，在每个卷积层后都有一个固定尺度的层。当采用大量的滤波器时，这些模块很大程度上保证了训练过程的稳定性。

EDSR 模型结构大体上与 SRResNet 很相似，其采用了改进过后的残差模块，并且在模型的最后没有在残差模块之外再设定一个 ReLU 激活层。同时 EDSR 模型没有使用残差尺度层，因为其在每个卷积层上都只使用 64 个特征图。在最终的 EDSR 模型中，基础模型的深度被扩大为 32，宽度被扩大为 256，并使用 0.1 的尺度因子。EDSR 模型结构如图 2-4 所示。

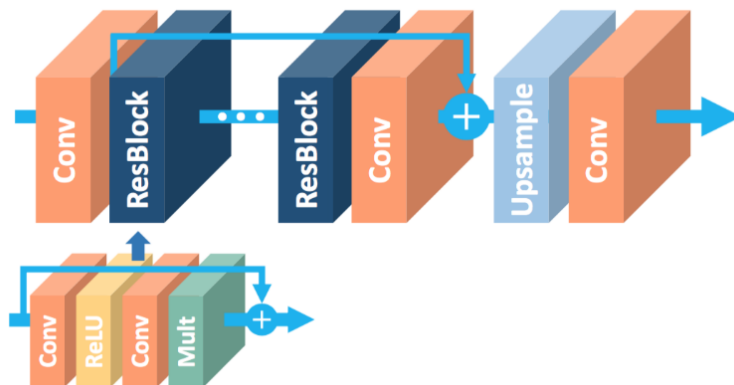


图 2-4: EDSR 算法模型结构^[1]

2.3 注意力机制

注意力可以被理解为一种对于可得的计算资源的有偏向的分配给最具有信息价值的特征部分的策略。注意力机制在许多工作中表现出很大的实用价值，如序列学习、图像定位与解释、图像描述等。在这些应用中，注意力机制可以被合并于一层或多层运算后紧接着的一个计算操作组，它代表了用于不同模块

之间适配的高层次上的抽象化。在近期的研究工作中曾有人尝试将通道和空间注意力^[23]与卷积神经网络结合起来。在关于图像分类的残差注意力网络研究^[24]中则介绍了一种基于沙漏模块^[25]的强有力的注意力机制，并应用于深度残差网络作用的不同领域中。

在卷积神经网络中，空间注意力机制的目标是重新校验输入信息中最有价值、最重要部分的特征。一些研究致力于将注意力模块整合在深度神经网络结构上，例如图像生成、图像描述、图像分类和图像恢复等。在 Xu 等人的文章^[26]中提出了一种用于图像描述生成的视觉注意力模型，其中使用了硬池化来挑选最有可能被注意到的区域或者是软池化根据注意力权重来平均空间特征。Xu 等人在另一篇文章^[27]中进一步优化了空间注意力模型，通过两个堆叠的空间注意力模型来解决视觉问题。同时通过考查一个网络结构中卷积特征通道之间不同的相互依赖关系，Hu 等人在文章^[2]中介绍了一种基于通道的注意力机制，提出一种 Squeeze-and-Excitation 模块来对图像分类问题自适应地调整不同通道上特征的权重。而受 SE 模块的启发，Chen 等人在文章^[28]中通过将多个空间和通道注意力模块堆叠在不同层上来处理图像描述问题。除了空间和通道注意力，Wang 等人在文章^[29]中运用了语义分割概率图作为先导知识，并介绍了语义注意力机制来将空间特征运用于现实文本生成。

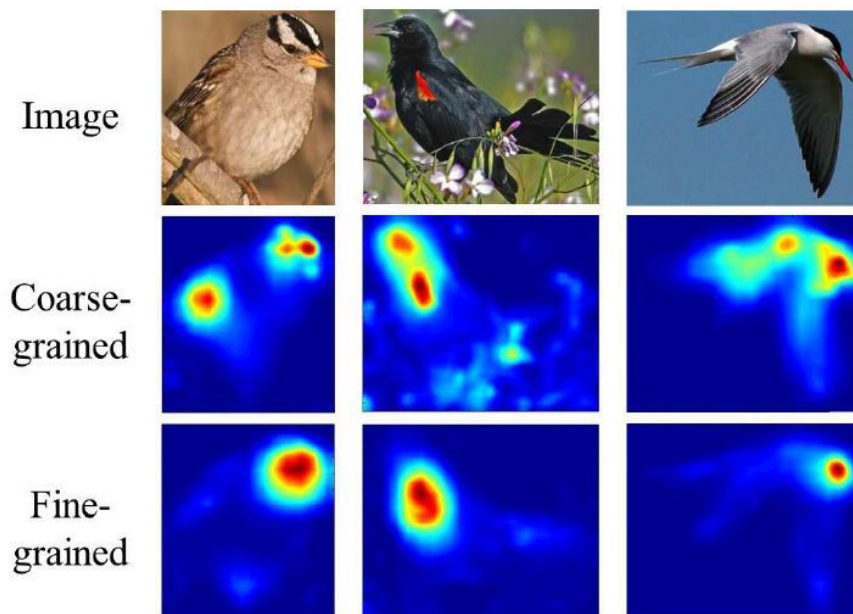


图 2-5: 注意力机制^[8]

受注意力机制的启发，考虑到在不同特征图之间和单个特征图内部不同种

类信息之间对图像超分辨有着不同程度的贡献，通过将通道注意力和空间注意力结合起来加入残差模块来自适应局部或全局的模块特征表示，从而捕捉到更加重要的信息。同时适当地忽略信息量较少和重要程度较低的特征，简化输入信息，略微降低了深度网络的计算量和复杂度，降低噪声输入对最终输出结果的影响。

2.4 Squeeze-and-Excitation 模块

Squeeze-and-Excitation^[2] 网络最早被提出用于解决图像分类问题，与其他神经网络不同的是，SE 模块更加关注图像各通道之间的联系与关系。它提出了一种使网络可以重新校准特征的机制，从而可以学习使用全局信息来有选择地强调有用的信息并忽略掉不怎么有用的信息。

Squeeze-and-Excitation 模块是一个计算单元，由 squeeze 和 excitation 这两步操作组成。因为每个学习到的滤波器都只能在局部的感受野上进行操作，从而导致了变换输出 \mathbf{U} 中的每个单元都无法利用该区域以外的上下文信息。为了解决这个问题，在 squeeze 操作中通过使用全局平均池化操作生成通道统计，从而将全局空间信息压缩成一个通道描述符。形式上，数据 $\mathbf{z} \in \mathbf{R}^C$ 是由在空间维度 $\mathbf{H} \times \mathbf{W}$ 上压缩 \mathbf{U} 得来的，则 \mathbf{z} 中第 c 个元素计算方式为：

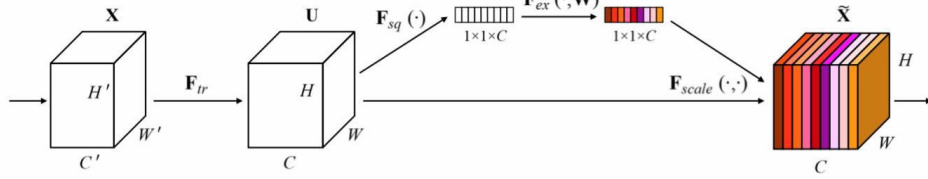
$$\mathbf{z} = \mathbf{F}_{sq}(\mathbf{u}_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (2-1)$$

为了将 squeeze 操作中得到的信息汇聚起来，紧接着进行 excitation 操作来充分捕捉通道依赖性。excitation 操作选用了一种用 sigmoid 激活的简单门机制：

$$\mathbf{s} = \mathbf{F}_{ex}(\mathbf{z}, \mathbf{W}) = \sigma(g(\mathbf{z}, \mathbf{W})) = \sigma(\mathbf{W}_2 \delta(\mathbf{W}_1 \mathbf{z})) \quad (2-2)$$

其中 δ 是指 ReLU 函数， $\mathbf{W}_1 \in \mathbf{R}^{\frac{C}{r} \times C}$ 且 $\mathbf{W}_2 \in \mathbf{R}^{C \times \frac{C}{r}}$ 。为了限制模型的复杂度并增加普适度，通过非线性结构周围添加两个全连接 (FC) 层瓶颈来参数化门机制，例如降维层参数为 \mathbf{W}_1 且降维比例为 τ 加上一个 ReLU 函数和一个参数为 \mathbf{W}_2 的升维层。数据块的最终输出通过激活操作重构转换输出 \mathbf{U} 得到：

$$\tilde{\mathbf{x}}_c = \mathbf{F}_{scale}(\mathbf{u}_c, \mathbf{s}_c) = \mathbf{s}_c \cdot \mathbf{u}_c \quad (2-3)$$

图 2-6: Squeeze-and-excitation 网络结构^[2]

其中 $\tilde{X} = [\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_C]$ 并且 $\mathbf{F}_{scale}(\mathbf{u}_c, s_c)$ 指的是标量 s_c 和特征映射 $\mathbf{u}_c \in \mathbf{R}^{H \times W}$ 之间对应通道的乘积。激活操作将特定输入描述符 z 和通道特征集合相互映射起来，从本质上来看，SE 模块引入了基于输入信息的动态特性，有助于提高特征分辨性。SE 模块的整体结构如图 2-6 所示。

2.5 不足与可改进之处

目前现有的几个比较具有代表性的超分辨算法中，仍存在一些不足。例如在 SRCNN 算法中，由于其输入图像是在一个高维度的低分辨率特征空间上被处理的，故其映射到的高维度高分辨率特征空间十分复杂，所需的运行时间开销也会相应的过大。虽然 VDSR 风格的结构解决了这一问题，但因为其需要双三次插值预处理过的图像作为输入，故这类结构的算法的计算时间开销和空间开销大大高于其他特定尺度的算法。同时许多算法中存在对于图像层面的上下文信息没能充分利用的问题，以及增加了大量的参数后更容易产生过拟合的问题。

故本文针对这些不足之处，将注意力机制与卷积神经网络结合起来。引入通道注意力机制，提升模型对于特征值各通道之间的联系与关系，同时引入空间注意力机制，加入空间注意力单元，使得模型对于每个特征值通道之间的关系信息也能够被表示、保存下来。通过对现有算法的改良优化，使得网络模型能够更好地利用全局的上下文信息，增强网络模型对不同信息的分辨能力，使其具有有选择地强调有用信息并忽略无用信息的能力。

2.6 本章小结

在本章节中，首先介绍了现有的一些比较主流的基于卷积神经网络的图像超分辨问题算法，并着重介绍了 EDSR 算法的整体框架和其内部优化后的残差

模块。同时介绍了注意力机制，沿着先前研究中将通道注意力和空间注意力与卷积神经网络结合起来运用于计算机视觉问题的思路，介绍了基于通道注意力机制的 **Squeeze-and-Excitation** 模块。最终分析了当前现有的一些算法的局限性与不足，指出通过将注意力机制与卷积神经网络结合起来来解决图像超分辨率问题这一改进思路。

第三章 改进的模型结构

3.1 CASAB 模块

深度网络生成的特征中包含了通道和空间区域里不同种类的信息，其中不同的特征对于高频细节重构所起到的作用与贡献度也不相同。故可以通过提高网络对高贡献率的信息的敏感度，使其更加关注并学习更有价值的信息，从而来提升网络结构的性能和其计算结果表现。基于该想法，本文参考引入了 Hu 等人文章^[3]中的 CSAR 模块，其结构如图 3-1 所示。CASAB 模块结构与 CSAR 模块相似，都是由通道注意力 CA 单元和空间注意力 SA 单元所组合构成。通过计算研究特征值在不同通道之间和不同空间区域之间的相互依赖关系，并将两种注意力机制相结合来优化残差模块，从而自适应地调整各特征的代表值。

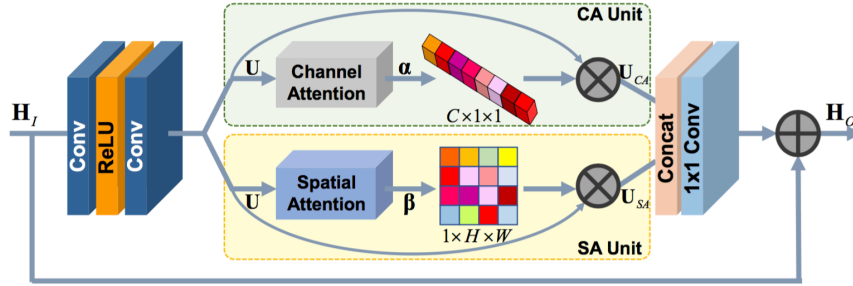


图 3-1: CSAR 模块结构示意图^[3]

通过将 CA 单元和 SA 单元结合起来加入残差模块从而提出 CASAB 模块。令 H_1 和 H_0 分别作为 CASAB 模块的输入与输出，并且令 $\Phi(\cdot)$ 作为 CA 单元和 SA 单元结合的模块操作，则 CASA 模块可以被表示为：

$$H_0 = \mathbf{Q}(H_1) = H_1 + \Phi(U) = H_1 + \Phi(\mathbf{R}(H_1)) \quad (3-1)$$

此处 $\mathbf{Q}(\cdot)$ 和 $\mathbf{R}(\cdot)$ 分别表示 CASAB 模块函数和残差分支。残差分支包含两个堆叠的带有 ReLU 激活函数的卷积层，

$$U = \mathbf{R}(H_1) = W_{\mathbf{R}}^2 * \delta(W_{\mathbf{R}}^1 * H_1 + b_{\mathbf{R}}^1) + b_{\mathbf{R}}^2 \quad (3-2)$$

此处 $\{W_{\mathbf{R}}^i\}_{i=1}^2$ 和 $\{b_{\mathbf{R}}^i\}_{i=1}^2$ 分别为残差分支的权重和偏差值， \mathbf{U} 是残差特征的集合。

为了获取更多的信息，CASAB 模块采用注意力结合模块 $\Phi(\cdot)$ 来重构分支特征 \mathbf{U} 。首先在残差特征 \mathbf{U} 上分别计算 CA 单元 $\Phi_{CA}(\cdot)$ 和 SA 单元 $\Phi_{SA}(\cdot)$ ，从而来获得通道权重特征图 U_{CA} 和空间权重特征图 U_{SA} ，然后两个重构过的特征图集合连接起来形成 1×1 卷积层的输入，用来融合两种学习调整过权重的注意力重构的特征值，所有操作总结如下：

$$\begin{aligned} U_{CA} &= \Phi_{CA}(U), \\ U_{SA} &= \Phi_{SA}(U), \\ \Phi(U) &= W_{\Phi} * [U_{CA}, U_{SA}] + b_{\Phi} \end{aligned} \quad (3-3)$$

其中 $[\cdot]$ 代表特征连接操作。

3.1.1 Channel-wise Attention (CA) 单元

CA^[3] 单元的目标是在全局范围进行特征再校准，计算每个通道的统计数据并将它们用于有选择地强调信息丰富的特征图部分，相应地压缩一些冗余的、无用的特征图部分。CA 单元的结构如图 3-2 所示，

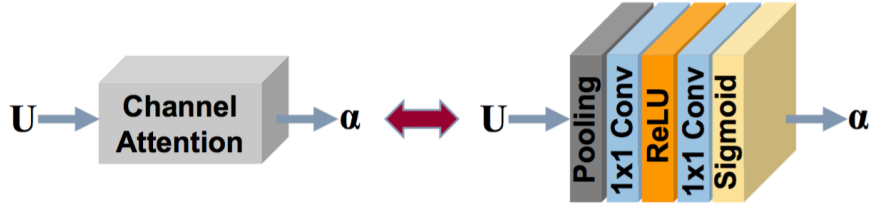


图 3-2: 通道间注意力操作 (CA 单元) ^[3]

令 $\mathbf{U} = [u_1, u_2, \dots, u_c]$ 作为 CA 单元的输入，其中有 C 个 $H \times W$ 尺寸的特征映射。为了生成通道统计数据 $z \in \mathbf{R}^{1 \times H \times W}$ ，通过在空间维度 $H \times W$ 的每个单独的特征通道上进行全局平均池化操作，如前文中提到的 Squeeze-and-Excitation 模块^[2]中的那样， z 中的第 c 的元素计算为：

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (3-4)$$

这里 $u_c(i, j)$ 表示第 c 个通道 u_c 在位置 (i, j) 处的值。为了将特征值映射图

中不同的注意力分配为不同的类别，采用了一种带 **sigmoid** 激活的门机制来统计数据 z ，计算过程如下：

$$\alpha = \sigma(W_{CA}^2 * \delta(W_{CA}^1 * U + b_{CA}^1) + b_{CA}^2) \quad (3-5)$$

其中 $\sigma(\cdot)$ 和 $\delta(\cdot)$ 分别表示 **sigmoid** 函数和 **ReLU** 函数， $*$ 表示卷积操作。 $W_{CA}^1 \in \mathbf{R}^{C \times C \times 1 \times 1}$ 和 $b_{CA}^1 \in \mathbf{R}^C$ 为第一个卷积层的权重和偏差值，其紧接着 **ReLU** 激活层并被用来以下降比率 r 来减少通道 z 的数目。接下来，通道 z 的数目通过另一个参数为 $W_{CA}^2 \in \mathbf{R}^{C \times C \times 1 \times 1}$ 和 $b_{CA}^2 \in \mathbf{R}^C$ 的卷积层提升回到原来的初始数目。通过 **sigmoid** 函数 $\sigma(\cdot)$ 将通道间注意力权重 $\alpha \in \mathbf{R}^{C \times 1 \times 1}$ 调整到在 $[0,1]$ 之间的值，然后用于重构输入的特征值，操作如下：

$$U_{CA} = \Phi_{CA}(U) = f_{CA}(U, \alpha) \quad (3-6)$$

其中 $f_{CA}(\cdot)$ 是通道特征值上特征通道和相对应的特征权重值的乘积， U_{CA} 是通道重新校准后的输出结果， $\Phi_{CA}(\cdot)$ 代表以输入值 U 为条件的 **CA** 单元。

通过以上操作，**CA** 单元可以通过输入的通道数据值来自适应地调整通道特征值，通过将更加重要的、包含更多有效信息的通道特征值放大，而将不那么重要的、较多冗余或是噪声数据的通道特征值缩小，来帮助网络结构显著提升通道之间特征值的差异性。

3.1.2 Spatial Attention (SA) 单元

使用深度卷积神经网络之后图像超分辨的表现得到了显著的提升，然而大多数基于卷积神经网络的模型并不具备对不同类型信息的分辨能力，对不同类型的信息同等对待处理会导致模型的表现能力受限。从另一个方面说，随着神经网络深度的增加，从较早的层中得来的长期信息更容易在后期的层中被弱化甚至消失，这对图像超分辨很不利。通过引入空间注意力机制，来获得更多的信息特征并将长期特征保留下来，赋予网络结构识别不同本地区域的能力并且更加关注包含更多信息的区域。

如图 3-3所示，令 $\mathbf{U} = [u_1, u_2, \dots, u_C]$ 作为空间注意力模块的输入，其中有 C 个 $H \times W$ 尺寸的特征映射。模块使用两层神经网络，紧接着使用 **sigmoid** 函

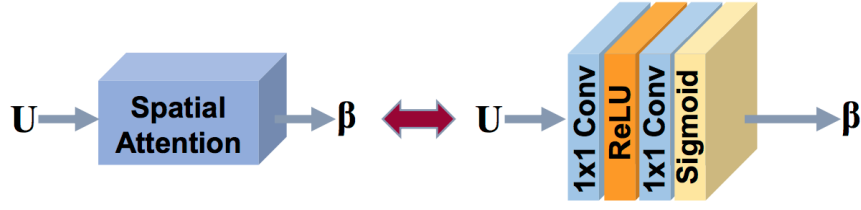


图 3-3: 空间注意力操作 (SA 单元) [3]

数来生成空间注意力掩码 $\beta \in \mathbf{R}^{1 \times H \times W}$ 。空间注意力模块的定义如下:

$$\beta = \sigma(W_{SA}^2 * \delta(W_{SA}^1 * U + b_{SA}^1) + b_{SA}^2) \quad (3-7)$$

其中 $\sigma(\cdot)$ 和 $\delta(\cdot)$ 分别表示 sigmoid 函数和 ReLU 函数, $*$ 表示卷积操作。第一个参数为 $W_{SA}^1 \in \mathbf{R}^{C \times C \times 1 \times 1}$ 和 $b_{SA}^1 \in \mathbf{R}^C$ 的卷积层用来产生每通道注意力图, 之后由第二个 1×1 且参数为 W_{SA}^2 和 b_{SA}^2 的卷积层融合为一个注意力映射关系图。不仅如此, sigmoid 函数 $\sigma(\cdot)$ 还将注意力映射关系范围归一化在 $[0,1]$ 之间来获得空间注意力平缓掩码 β 。输入特征在空间上被 β 调整的过程可以表示为:

$$U_{SA} = \Phi_{SA}(U) = f_{SA}(U, \beta) \quad (3-8)$$

其中 $f_{SA}(\cdot)$ 是每个特征图的空间位置及其对应的空间注意力权重的元素乘法, $\Phi_{SA}(\cdot)$ 表示 SA 模型。

通过 SA 单元, 可以在本地自适应地调制特征值, 使其可以与全局通道调制相互作用来增强网络结构的表示能力。

3.2 CAN 模型

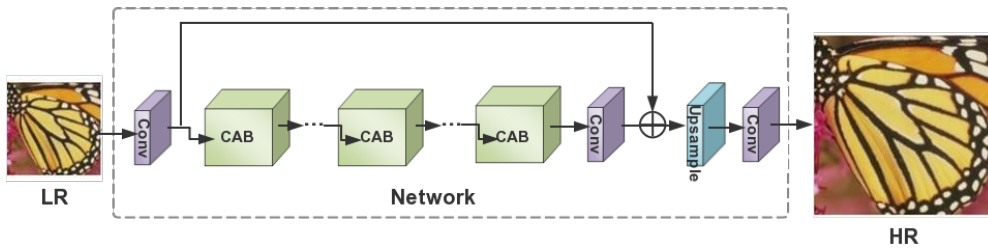


图 3-4: CAN 模型结构示意图

CAN 模型以 EDSR 模型结构为基础，将通道注意力机制与卷积神经网络结合起来，通过将 EDSR 算法中的残差模块全部替换为 CAB 模块，得到改进后的 CAN 模型，其结构如图 3-4所示。

CAB 模块 CAB 模块在先前 EDSR 算法中改进的残差模块的基础上进行了进一步的优化。原始的残差模块并没有对输入数据同一通道之间进行任何计算操作，故可能忽略了在同一通道中，特征信息相互之间可能存在的关系信息。而 CAB 模块则是在其基础上加入了 CA 单元，通过 CA 单元来计算输入信息各通道之间的相互关系，并更新每个通道的权重值。CAB 模块由一次卷积操作，一次 ReLU 线性整流操作，第二次卷积操作和一次 CA 操作构成，其内部结构如图 3-5所示。

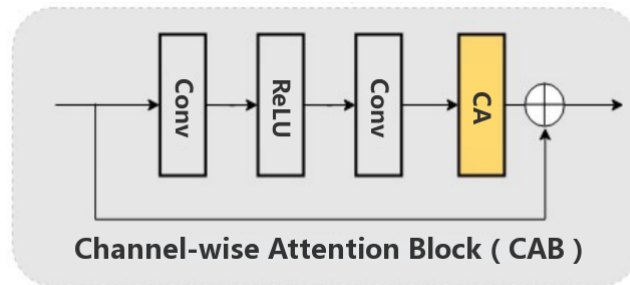


图 3-5: CAB 模块结构示意图

3.3 SAN 模型

SAN 模型以 EDSR 模型结构为基础，将空间注意力机制与卷积神经网络结合起来，通过将 EDSR 算法中的残差模块全部替换为 SAB 模块，得到改进后的 SAN 模型，其结构如图 3-6所示。

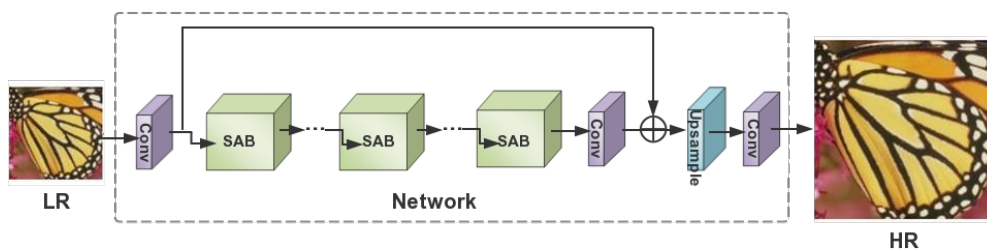


图 3-6: SAN 模型结构示意图

SAB 模块 SAB 模块在先前 EDSR 算法中改进的残差模块的基础上进行了进一步的优化。由于原始的残差模块并没有对输入数据不同空间区域之间进行任何计算操作，忽略了特征信息在不同的空间区域中相互之间可能存在的关系信息。而 SAB 模块则是在其基础上加入了 SA 单元，通过 SA 单元引入空间注意力机制，从而计算输入信息不同空间区域之间的相互关系。SAB 模块由一次卷积操作，一次 ReLU 线性整流操作，第二次卷积操作和一次 SA 操作构成，其内部结构如图 3-7 所示。

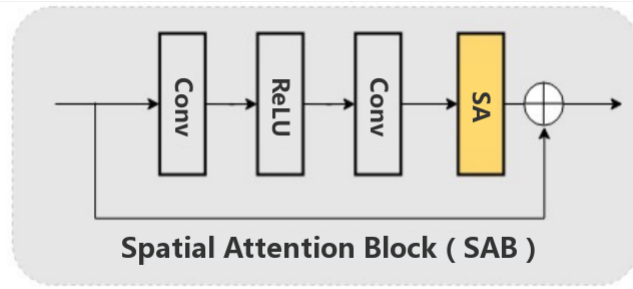


图 3-7: SAB 模块结构示意图

3.4 CASAN 模型结构

CASAN 模型同样以 EDSR 模型为基础，但与 CAN 模型和 SAN 模型不同的是，CASAN 模型同时将通道注意力机制和空间注意力机制结合起来，一并加入了卷积神经网络当中。通过将 EDSR 算法中的残差模块全部替换为上文介绍的 CASAB 模块，得到改进后的 CASAN 模型。CASAN 模型首先对低分辨率的输入信息进行一次卷积处理，然后对处理过后的输入信息进行多次 CASAB 模块操作，最后对输出的信息进行上采样重构到目标图像大小，再进行一次卷积处理后得到最终超分辨率图像输出，其结构如图 3-8 所示。

在网络结构的一开始，使用一个卷积层来从低分辨率输入图像中提取出原始特征。该原始特征作为后续 CASAB 模块的输入特征，同时也被用于全局残差特征学习，

$$F_{InputF} = f_{conv}(X) \quad (3-9)$$

其中 $f_{conv}(\cdot)$ 代表对初始输入图像信息的卷积操作， F_{InputF} 是提取后得到的初始特征值。接着将 F_{InputF} 传入第一个 CASAB 模块进行运算，假设整个网络

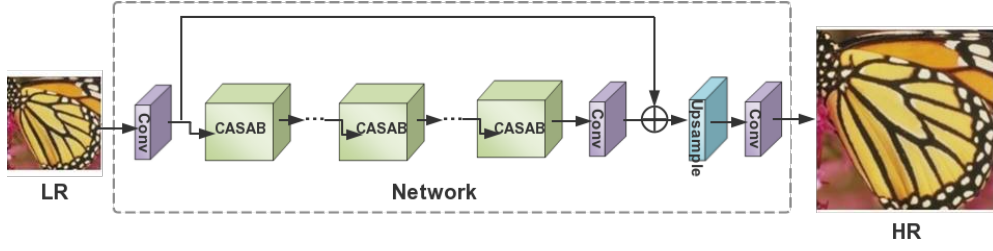


图 3-8: CASAN 模型结构示意图

结构中有 M 个 CASAB 模块来进行特征值转换，则经过所有 CASAB 模块运算过后的输出为：

$$\begin{aligned}
 F_{casa} &= f_{conv}(\psi^{(M)}(F_{InputF})) + F_{InputF} \\
 &= f_{conv}(\psi_M(\psi_{M-1}(\cdots(\psi_1(F_{InputF}))\cdots))) + F_{InputF}
 \end{aligned}
 \tag{3-10}$$

其中 $f_{conv}(\cdot)$ 代表卷积操作， $\psi_m(\cdot) (m = 1, 2, \dots, M)$ 代表第 m 个 CASAB 模块函数， F_{casa} 代表经过 M 个 CASAB 模块运算后的输出值。最后再对该经过转换的输出特征值进行上采样重新扩大到目标超分辨图片规模大小，得到最终的输出结果。

CASAN 模型通过将可结合的注意力模块加入深度网络，从两个方面优化了算法结构，一是由于可结合的注意力模块只会重构残差特征值，故不会破坏残差模块中特征映射图中的良好特性，同时信息流仍然会得到提升；二是由于两个注意力单元被结合进一个残差模块，这样就可以很便利地将通道间注意力和空间注意力机制通过堆叠的多个 CASAB 模块运用到多层特征中去，从而使更多的多层重要信息被捕捉。CASAN 模型分别从通道间注意力机制与空间注意力机制两个方面入手，分辨出不同输入信息之间的价值差异。通过在 CASAB 模块内部对通道与通道之间、空间区域之间的关系进行计算衡量，得到不同通道、空间上特征值所对应的不同权重值。对包含有更多信息量、更加有价值的输入权重提高，使其在后续的计算中产生更多的作用；而将信息量较少的冗余输入信息或是噪声杂音的权重降低，减少其在后续计算中的参与度，并减轻其对后续计算的干扰。注意力机制的加入使得 CASAN 模型结构的效果得到了提升，同时也降低了计算过程中的参数量和计算量，节约了网络中的计算和存储开销。

3.5 OAN 模型

OAN 模型在 CASAN 模型的基础上进一步创新优化，在 CASAN 模型中通道注意力和空间注意力机制相结合的基础上，再在注意力机制的计算上加入不同的尺度。通过将 EDSR 算法中的残差模块全部替换为加入了在不同尺度上计算的通道与空间注意力相结合的 OAB 模块，得到改进后的 OAN 模型，其结构如图 3-9 所示。

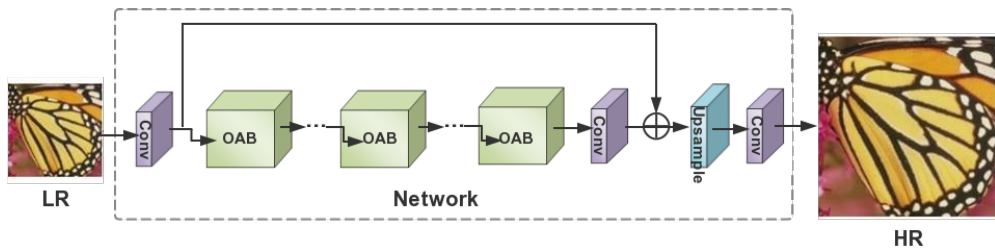


图 3-9: OAN 模型结构示意图

3.5.1 OAB 模块

受到卷积神经网络与注意力机制相结合来解决图像超分辨率问题这一思路的启发，并且在 CASAB 模块的基础上优化创新出了 OAB 模块。由于 CASAB 模块对于权重值的计算和更新都仅仅基于单一尺度，例如 CASAB 模块中的 CA 单元，在计算通道特征值相互之间的关系前，先对各通道上特征值进行了平均值池化操作，取得每个通道上的一个平均特征值后，再对各通道之间的平均特征值进行运算获得新的权重值；而 CASAB 模块中的 SA 单元则是去除了池化操作，保留原始空间大小，对一个空间区域上的所有特征值之间的相互关系计算新的权重值。

而 OAB 模块则在注意力机制中加入了不同尺度上的计算，通过先对特征值进行不同尺寸的池化操作后，再分别通过 CA 单元和 SA 单元计算通道之间和空间区域之间的关系并更新插值出各自的新权重值，最终将不同尺寸下的新权重值相加后进行一次卷积操作还原成目标尺寸，得到最终的更新后权重值。

在 OAB 模块当中，首先对输入的待处理的特征值集合分别取 $\frac{1}{2}$ 、 $\frac{1}{4}$ 、 $\frac{1}{8}$ 的空间大小，在不同大小的区域上进行最大值池化操作，接着进行各自区域上的注意力机制操作，得到新的权重值；最后，分别将 $\frac{1}{2}$ 、 $\frac{1}{4}$ 、 $\frac{1}{8}$ 空间大小上的新权

重值进行插值，统一到同一个尺寸后相加并输入一次卷积操作，接着再用一个 1×1 卷积来将输出扩大到目标规模尺寸，并通过 sigmoid 函数激活，得到最终输出结果，其结构如图 3-10所示。

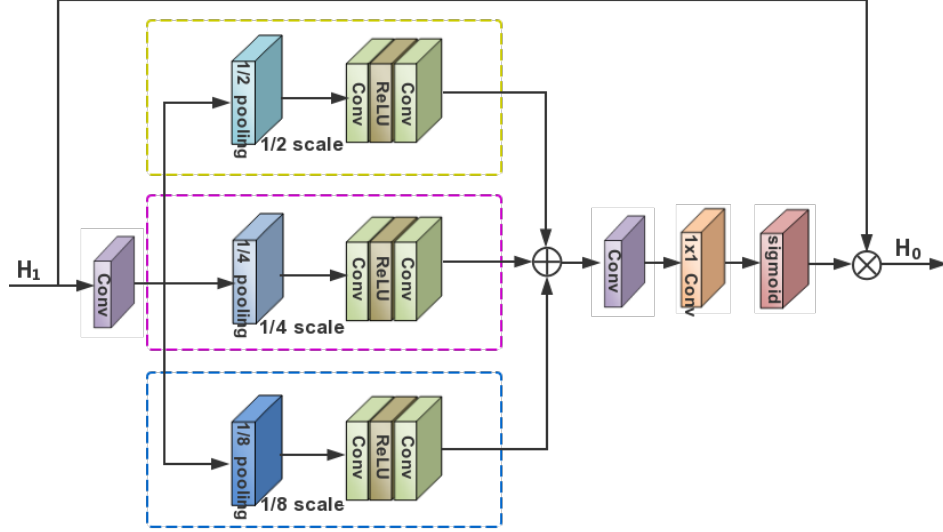


图 3-10: OAB 模块结构示意图

3.5.2 OAN 模型结构

OAN 模型同样也是以 EDSR 模型为基础，首先对低分辨率的输入信息进行一次卷积处理。区别在于将 EDSR 模型当中的残差模块替换成为了 OAB 模块，而后对处理过后的输入信息进行多次 OAB 模块计算，最后对输出的信息进行上采样重构到目标图像大小，再进行一次卷积处理后得到最终超分辨率图像输出。

若以 F_{InputF} 作为提取后得到的特征值，将 F_{InputF} 传入第一个 OAB 模块进行运算，假设整个网络结构中有 M 个 OAB 模块来进行特征值转换，则经过所有 OAB 模块运算过后的输出为：

$$\begin{aligned} F_{oa} &= f_{conv}(\psi^{(M)}(F_{InputF})) + F_{InputF} \\ &= f_{conv}(\psi_M(\psi_{M-1}(\cdots(\psi_1(F_{InputF}))\cdots))) + F_{InputF} \end{aligned} \quad (3-11)$$

其中 $f_{conv}(\cdot)$ 代表卷积操作， $\psi_m(\cdot) (m = 1, 2, \dots, M)$ 代表第 m 个 OAB 模块函数， F_{oa} 代表经过 M 个 OAB 模块运算后的输出值。最后再对该经过转换的

输出特征值进行上采样重新扩大到目标超分辨图片规模大小，得到最终的输出结果。

3.6 本章小结

本章节中分别介绍了在 EDSR 模型基础上改进优化出的 4 种模型结构，分别为加入了通道注意力机制的 CAN 模型；加入了空间注意力机制的 SAN 模型；加入了通道注意力与空间注意力机制相结合的 CASAN 模型；以及加入了不同尺度上通道与注意力机制相结合的 OAN 模型。

第四章 实验与分析

4.1 数据集与评估标准

训练集数据 实验中采用的训练集为 DIV2K，其中包含 800 张高质量训练图像来训练模型，100 张验证图像来验证模型，100 张测试图像来测试模型。同时在这些训练图像上进行数据扩大，通过随机水平翻转、随机旋转 90° 、 180° 或 270° 来实现数据增强。

测试集数据 测试集数据选取现在普遍使用的 Set5, Set14, B100 和 Urban100 作为基准，它们分别包含 5, 14, 100 和 100 张图片。Set5、Set14 和 B100 包含了自然景观类的图片，Urban100 则较为具有挑战性，它包含了许多城市景观的图片，其中包括许多不同频段中的细节，很多现有算法都不能很好的处理 Urban100 数据集。

评估标准 以峰值信噪比 (PSNR) 作为实验结果评估标准，其计算公式为：

$$PSNR = 10 \times \log_{10} \left(\frac{(2^n - 1)^2}{MSE} \right) \quad (4-1)$$

其中 MSE 是初始输入图像与经过计算后的输出图像之间的均方误差。

4.2 实验设置及实现细节

整体实验设置如下：

- 工具：PyTorch 1.0.0, Python3
- 设备：GTX 1080Ti GPU
- 输入图片大小： 48×48
- 优化器：Adam 优化器
- 初始学习率： 10^{-4} ，每隔 200 轮降低一半
- 损失函数：L1 型损失函数

表 4-1: 基于 CASAN 模型在 2 倍规模上的不同参数实验结果。在配置 (Set) 一列中, RB 表示残差块的个数 (n_resblocks), F 表示特征纬度 (n_feats), B 表示批处理参数 (batch_size), P 表示输入图像块大小 (patch_size), RE 表示下降率 (reduction)。四个测试集上的 PSNR 分数一栏中, 得分最高的用 **highlighted** 表示, 得分第二高的用 underlined 表示。

Scale	Set	MultAdds	Params	Set5 PSNR	Set14 PSNR	B100 PSNR	Urban100 PSNR
2×	RB16 F48 B16 P64 RE16	197.5G	857K	37.89	<u>33.48</u>	32.09	31.83
	RB8 F64 B16 P64 RE16	197.2G	854K	37.84	33.40	32.07	31.71
	RB16 F48 B8 P64 RE16	197.5G	857K	37.83	33.43	32.08	31.78
	RB16 F48 B32 P64 RE16	197.5G	857K	37.96	33.55	32.12	31.92
	RB16 F48 B16 P32 RE16	197.5G	857K	37.85	33.39	32.05	31.69
	RB16 F48 B16 P48 RE16	197.5G	857K	37.90	33.42	32.08	31.79
	RB16 F48 B16 P64 RE4	197.5G	857K	37.89	33.46	<u>32.10</u>	<u>31.87</u>
	RB16 F48 B16 P64 RE8	197.5G	857K	<u>37.92</u>	33.43	<u>32.10</u>	31.81

首先针对提出的 CASAN 模型设置了 8 组不同参数 (参与验证的参数包括 CSAR 模块的个数、特征的纬度、下降率、批处理参数及输入块大小) 在规模 2 上的对照试验。通过对比 8 组对照实验的计算量、参数量及 PSNR 值评分结果来比较不同参数下模型效果的优劣, 从而选出实验的最佳参数。

接着针对基础网络结构 EDSR 模型, 和在其基础上改进优化的来的 CAN 模型、SAN 模型、CASAN 模型和 OAN 模型分别在规模 2、规模 3 (除 OAN 模型) 和规模 4 上设置对照实验。对比 5 种模型的计算量、参数量及 PSNR 值评分结果, 分析评估出其中最优的模型, 用来在后文中与当下现有算法结构比较。

最终在相同的参数设置下, 分别在输出规模 2 和规模 4 上将 OAN 模型与当下现有的一些超分辨算法进行比较, 通过对比不同算法之间的计算量、参数量及 PSNR 值评分结果, 来判断本文提出的改进后的 OAN 模型的性能效果是否得到了提升。

4.3 参数选择与分析

针对提出的 CASAN 模型, 从残差模块个数、特征纬度、下降率、批处理参数和输入图像块大小这五个参数角度出发, 以 16 个残差块、48 个特征纬度、16 为下降率、16 为批处理参数以及 64 为图像块大小作为基础参数, 在 2 倍规模上共设置了 7 组对照实验来进行不同参数的对比, 实验结果见表 4-1。

残差块的个数和特征纬度 当残差块个数减少为 8 而特征纬度增加为 64 时, 计算量和参数量有些微的降低, 但同时在各个测试集上的 PSNR 评分结果也有一些降低。总体来说, 残差块个数为 16, 特征纬度为 48 时的基础参数设置的效果要更好一些。

批处理参数 当批处理参数降低为 8 时, 但同时在各个测试集上的 PSNR 评分结果均有一些降低; 而当批处理参数增加为 32 时, 在各个测试集上的 PSNR 评分结果都有很大提升, 并且均为 8 组实验中的最高得分。批处理参数的改变并不影响参数量和计算量, 故可见当批处理参数增加为 32 时, 模型的效果最好。

输入图像块大小 输入图像块大小也不影响参数量和计算量。当输入图像块降低为 32 和 48 时, 其在各测试集上的 PSNR 评分也不如基础参数设置下的模型理想, 故还是初始参数中的 64 大小的输入图像块效果较好。

下降率 下降率也同样不影响参数量和计算量。当下降率降低为 4 和 8 时, 其在各个测试集上的 PSNR 评分表现相较于初始参数设置下的模型都有少许的提升, 但整体的提升效果不如批处理参数增加为 32 时的模型。

4.4 不同模型效果分析

使用上一节的最有参数, 即 16 个残差块、48 个特征纬度、16 为下降率、32 为批处理参数以及 64 为图像块大小, 针对基础网络结构 EDSR 模型, 和在其基础上改进优化的来的 CAN 模型、SAN 模型、CASAN 模型和 OAN 模型分别在规模 2、规模 3 (除 OAN 模型) 和规模 4 上设置对照实验。实验结果见表 4-2。同时 EDSR 模型、CAN 模型、SAN 模型、CASAN 模型与 OAN 模型恢复出的同一图片中的局部效果对比见图 4-1。

CAN 模型 CAN 模型与基础 EDSR 模型对比, 计算量和参数量都有些微的增加, 但在各测试集上的 PSNR 得分表现并没有明显的提升, 在个别规模的测试集上的 PSNR 得分反而还有少许降低。总体来说, 加入了通道注意力机制的 CAN 模型优化并不明显, 表现并不理想。

SAN 模型 SAN 模型与基础的 EDSR 模型对比, 计算量和参数量有较多的增加, 但在各个测试集上的 PSNR 得分表现的提升并不明显, 但基本都略高于 EDSR 模型得到的 PSNR 评分, 相较于 CAN 模型没有出现评分降低的情况。总体来说, 加入了空间注意力机制的 SAN 模型在计算量、参数量和最终效果之

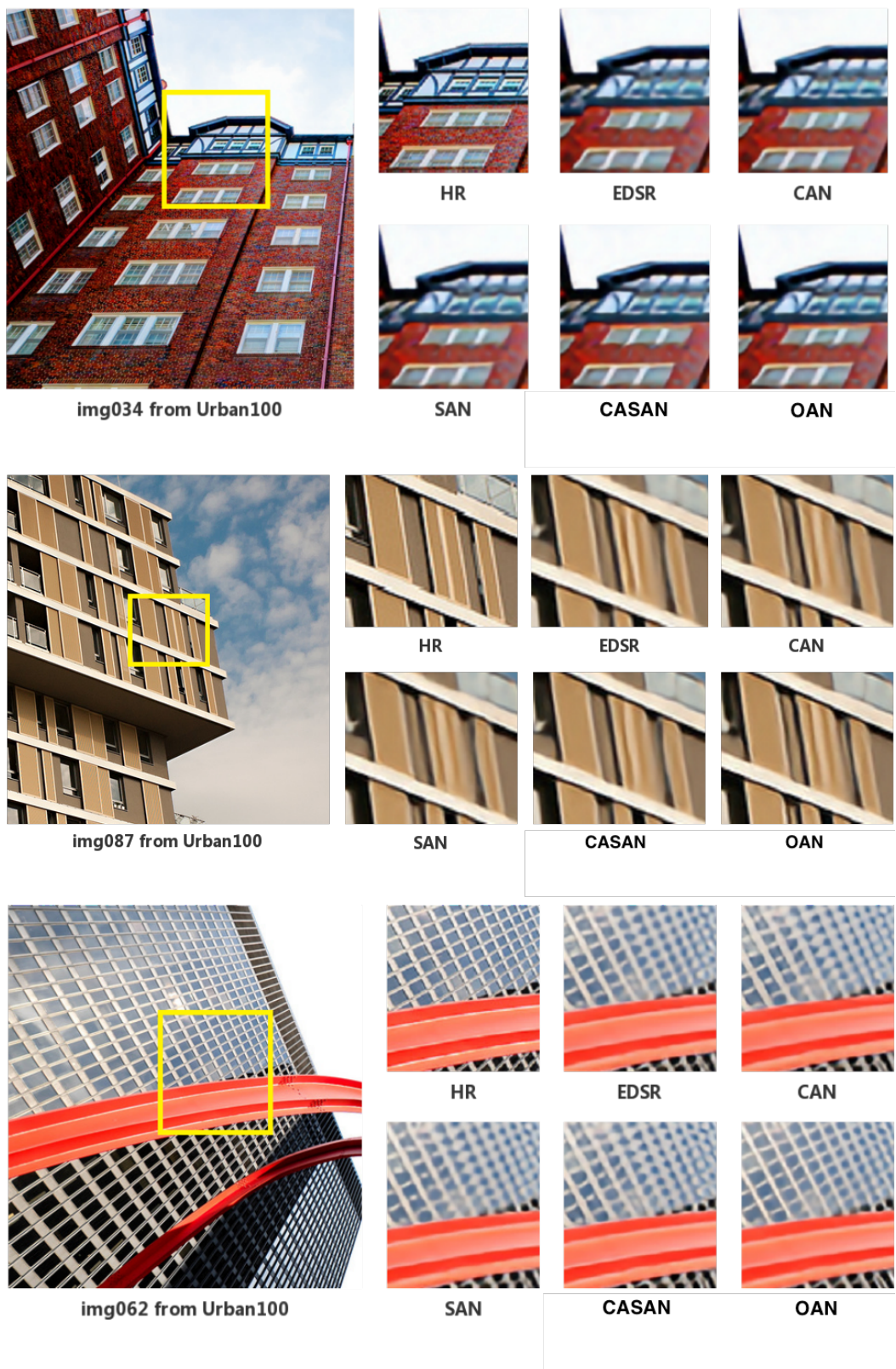


图 4-1: 原始 HR 图片与 EDSR 算法、CAN 算法、SAN 算法、CASAN 算法及 OAN 算法恢复出的图片局部效果展示

表 4-2: 基础模型 EDSR 与改进后模型 CAN、SAN、CASAN 和 OAN 的实验表现结果。四个测试集上的 PSNR 分数一栏中，得分最高的用 **highlighted** 表示，得分第二高的用 underlined 表示。

Scale	Method	MultAdds	Params	Set5 PSNR	Set14 PSNR	B100 PSNR	Urban100 PSNR
2×	EDSR ^[1]	178.9G	772K	37.86	33.43	32.11	31.78
	CAN	179.1G	776K	37.91	33.44	32.08	31.72
	SAN	188.8G	820K	37.90	33.42	32.10	31.77
	CASAN	197.5G	857K	<u>37.96</u>	<u>33.55</u>	<u>32.12</u>	<u>31.92</u>
	OAN	190.1G	882K	38.06	33.59	32.21	32.27
3×	EDSR ^[1]	90.7G	876K	34.23	30.20	<u>29.02</u>	27.88
	CAN	90.8G	880K	<u>34.28</u>	<u>30.21</u>	29.01	<u>27.90</u>
	SAN	95.1G	924K	34.23	<u>30.21</u>	29.01	27.87
	CASAN	98.9G	961K	34.30	30.23	29.04	27.98
4×	EDSR ^[1]	64.8G	855K	31.92	28.45	27.50	25.81
	CAN	64.2G	859K	31.09	28.46	27.50	25.82
	SAN	67.3G	903K	32.02	28.47	27.51	25.83
	CASAN	69.4G	940K	<u>32.07</u>	<u>28.48</u>	<u>27.52</u>	<u>25.88</u>
	OAN	67.6G	965K	32.16	28.59	27.58	26.08

间的性价比并不高。

CASAN 模型 CASAN 模型与基础的 EDSR 模型对比，计算量和参数量有显著的增加，同时各个测试集上的 PSNR 得分表现也有较大的提升。CASAN 模型在各个规模的各测试集上的 PSNR 得分表现都要优于 CAN 模型和 SAN 模型，说明通道注意力和空间注意力相结合的算法取得了较好的优化效果。

OAN 模型 OAN 模型与基础的 EDSR 模型相比，计算量和参数量也有较大的增加，同时各测试集上的 PSNR 得分表现也有显著的提升。而作为在 CASAN 模型结构上改进优化得到的 OAN 模型，其计算量略小于 CASAN 模型，而参数量则比 CASAN 模型的增加较多，同时各个测试集上的 PSNR 评分结果也都高于 CASAN 模型。说明 OAN 模型在 CASAN 模型的基础上加入了多尺度上的注意力机制计算，对模型性能的优化提升效果明显。

表 4-3: OAN 模型与当下现有其它超分辨率算法模型效果对比。四个测试集上的 PSNR 分数一栏中, 得分最高的用 **highlighted** 表示, 得分第二高的用 underlined 表示。

Scale	Method	MultAdds	Params	Set5 PSNR	Set14 PSNR	B100 PSNR	Urban100 PSNR
2×	SRCNN ^[5]	52.7G	57K	36.66	32.42	31.36	29.50
	FSRCNN ^[14]	6.0G	12K	37.00	32.63	31.53	29.88
	VDSR ^[11]	612.6G	665K	37.53	33.03	31.90	30.76
	DRCN ^[12]	17974.3G	1774K	37.63	33.04	31.85	30.75
	LapSRN ^[7]	29.9G	813K	37.52	33.08	31.80	30.41
	DRRN ^[13]	6796.9G	297K	37.74	33.23	32.05	31.23
	MemNet ^[30]	623.9G	677K	37.78	33.28	32.08	31.31
	SelNet ^[31]	225.7G	974K	<u>37.89</u>	33.61	32.08	-
	IDN ^[32]	123.5G	551K	37.83	33.30	32.08	31.27
	CARN ^[33]	222.8G	964K	37.76	33.52	<u>32.09</u>	<u>31.92</u>
	OAN(ours)	190.1G	882K	38.06	<u>33.59</u>	32.21	32.27
4×	SRCNN ^[5]	52.7G	57K	30.48	27.49	26.90	24.52
	FSRCNN ^[14]	6.0G	12K	30.71	27.59	26.98	24.62
	VDSR ^[11]	612.6G	665K	31.35	28.01	27.29	25.18
	DRCN ^[12]	17974.3G	1774K	31.53	28.02	27.23	25.14
	LapSRN ^[7]	6796.9G	297K	31.54	28.19	27.32	25.21
	DRRN ^[13]	623.9G	677K	31.68	28.21	27.38	25.44
	MemNet ^[30]	83.1G	1417K	31.74	28.26	27.40	25.50
	SelNet ^[31]	149.4G	872K	32.00	28.49	<u>27.44</u>	-
	IDN ^[32]	30.9G	555K	31.82	28.25	27.41	25.41
	CARN ^[33]	90.9G	1112K	<u>32.13</u>	28.60	27.58	<u>26.07</u>
	OAN(ours)	67.6G	965K	32.16	<u>28.59</u>	27.58	26.08

4.5 与现有算法的比较与分析

分别在输出规模 2 和规模 4 上将 OAN 模型与当下现有的一些超分辨率算法进行比较, 通过对比不同算法之间的计算量、参数量及 PSNR 值评分结果, 来判断本文提出的改进后的 OAN 模型的性能效果是否得到了提升。其实验结果总结在表 4-3 中。

从表中结果可以看出, 一些较为早期的算法如 SRCNN 和 FSRCNN 由于结构较为简单, 其计算量和参数量也较小, 但同时它们在各个测试集上的 PSNR 得分结果也都较差。在与现有算法的比较当中, OAN 算法在规模 2 和规模 4 上的 Set5、B100 和 Urban100 测试集上 PSNR 得分都为最高分。仅在规模 2 时的

测试集 Set14 上 PSNR 得分略微低于 SelNet 算法；在规模 4 时的测试集 Set14 上的 PSNR 得分略低于 CARN 算法，测试集 B100 上的 PSNR 得分与 CARN 算法持平。但 OAN 算法在规模 2 上的计算量和参数量都要小于 SelNet 算法，同时在规模 4 上的计算量和参数量也同样明显小于 CARN 算法。故 OAN 算法总体在平衡计算量、参数量与最终效果上表现较好，整体性价比要优于 SelNet 算法和 CARN 算法。

而与其它算法相比，OAN 算法基本均在最终结果上更是有了较为明显的提升。其计算量与参数量比 DRCN 算法少了几个数量级，同时计算结果也远远优于一些提出时间较早的基础算法，而与近几年提出的算法相比，其计算结果也要略优，说明 OAN 算法与现有许多算法相比已经有了一定的提升和优化，其通过将卷积神经网络与注意力机制相结合，并在不同的尺度上计算通道和空间注意力并更新特征值集合的权重，在解决图像超分辨问题上表现优异。

4.6 本章小结

本章节中首先介绍了所采用的训练数据集和测试数据集，以及所遵循的评分标准；然后介绍了所进行试验的配置；接着在 CASAN 模型上设置不同的参数进行实验对比和分析；同时分别对基础 EDSR 模型和改进后的 CAN 模型、SAN 模型、CASAN 模型和 OAN 模型实验效果进行了对比和分析；最后将 OAN 模型表现结果与当前现有的许多超分辨算法进行比较和分析。

第五章 总结与展望

本文针对当下较为热门的基于卷积神经网络的图像超分辨算法，沿着将通道注意力机制和空间注意力机制相结合来加入卷积神经网络以提升网络性能的思路，从将注意力机制与深度学习结合起来解决图像超分辨问题的角度进行研究。

本文参考了 EDSR^[1] 算法网络结构，以 EDSR 网络结构框架为基础，通过将 EDSR 网络结构中的残差模块全部替换为 CAB 模块得到 CAN 结构，从而仅将通道注意力加入了卷积神经网络中；通过将 EDSR 网络结构中的残差模块全部替换为 SAB 模块得到 SAN 结构，从而仅将空间注意力加入了卷积神经网络中；而通过将 EDSR 网络结构中的残差模块全部替换为 CASAB 模块得到 CASAN 结构，在 CASAN 结构中则将通道注意力与空间注意力机制结合起来，一并加入卷积网络中；最终在 CASAN 结构上进一步优化，加入了不同尺度上的注意力机制，得到 OAN 结构。

同时本文设置对照实验验证了 CASAN 模型相较于 EDSR 模型、CAN 模型和 SAN 模型的表现更优，同时 OAN 模型相较于 CASAN 模型的表现也得到了提升。实验证明了通过通道和空间注意力机制相结合的加入得到的 CASAN 模型，确实改进提升了算法模型在超分辨问题上的表现，而在 CASAN 模型基础上再加入不同尺度上的注意力机制得到的 OAN 模型，则更加进一步提升了其在超分辨问题上的表现。在 OAN 模型与当下现有的许多基于卷积神经网络的图像超分辨算法的比较当中，OAN 模型也得到了比较优异的实验结果，实验证明 OAN 模型实验结果确实比当前现有的许多算法模型有了优化和提升。

基于上述工作，在今后的实验中，可以进一步尝试更多尺度下的注意力机制相结合的效果，同时在注意力机制与深度学习的结合与运用方面进一步优化、探索。

致 谢

大学四年如白驹过隙，不知不觉中已经临近毕业。在我的论文即将完稿之际，我首先想感谢我的指导老师唐杰副教授对我的精心指导和热心帮助。唐老师为人和蔼可亲，关心爱护学生，对于毕业设计的指导方面亲力亲为，在自身繁重的教学任务之中仍然能够抽出宝贵的时间来及时指正我毕业设计中的不足之处，令我十分感动。同时也感谢我的师兄刘杰，在我的整个毕业设计过程中对我的指导和启发，每每当我遇到自身难以解决的问题时，师兄刘杰总不辞繁琐，给我一些指导和帮助。最后感谢多媒体组实验室为我毕业设计实验所提供的硬件条件，也感谢全体答辩组委会的老师能够在百忙之中抽空对我的毕业设计进行指正。

参考文献

- [1] 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 : 136 – 144.
- [2] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2018 : 7132 – 7141.
- [3] HU Y, LI J, HUANG Y, et al. Channel-wise and Spatial Feature Modulation Network for Single Image Super-Resolution[J]. arXiv preprint arXiv:1809.11130, 2018.
- [4] 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.
- [5] DONG C, LOY C C, HE K, et al. Learning a deep convolutional network for image super-resolution[C] // European conference on computer vision. 2014 : 184 – 199.
- [6] 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] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2016 : 1874 – 1883.
- [7] 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.
- [8] WANG D, SHEN Z, SHAO J, et al. Multiple granularity descriptors for fine-grained categorization[C] // Proceedings of the IEEE international conference on computer vision. 2015 : 2399 – 2406.

- [9] DEMIREL H, ANBARJAFARI G. Image resolution enhancement by using discrete and stationary wavelet decomposition[J]. IEEE transactions on image processing, 2011, 20(5): 1458 – 1460.
- [10] SCHULTER S, LEISTNER C, BISCHOF H. Fast and accurate image upscaling with super-resolution forests[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 3791 – 3799.
- [11] 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.
- [12] KIM J, KWON LEE J, MU LEE K. Deeply-recursive convolutional network for image super-resolution[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1637 – 1645.
- [13] TAI Y, YANG J, LIU X. Image super-resolution via deep recursive residual network[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 3147 – 3155.
- [14] DONG C, LOY C C, TANG X. Accelerating the super-resolution convolutional neural network[C] // European conference on computer vision. 2016: 391 – 407.
- [15] LEDIG C, THEIS L, HUSZÁR F, et al. Photo-realistic single image super-resolution using a generative adversarial network[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4681 – 4690.
- [16] TONG T, LI G, LIU X, et al. Image super-resolution using dense skip connections[C] // Proceedings of the IEEE International Conference on Computer Vision. 2017: 4799 – 4807.
- [17] ZHANG Y, TIAN Y, KONG Y, et al. Residual dense network for image super-resolution[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 2472 – 2481.
- [18] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

- [19] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[J]. arXiv preprint arXiv:1502.03167, 2015.
- [20] 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.
- [21] HE K, ZHANG X, REN S, et al. Identity mappings in deep residual networks[C] // European conference on computer vision. 2016: 630–645.
- [22] SRIVASTAVA R K, GREFF K, SCHMIDHUBER J. Training very deep networks[C] // Advances in neural information processing systems. 2015: 2377–2385.
- [23] WOO S, PARK J, LEE J-Y, et al. Cbam: Convolutional block attention module[C] // Proceedings of the European Conference on Computer Vision (ECCV). 2018: 3–19.
- [24] PARK J, WOO S, LEE J-Y, et al. Bam: Bottleneck attention module[J]. arXiv preprint arXiv:1807.06514, 2018.
- [25] NEWELL A, YANG K, DENG J. Stacked hourglass networks for human pose estimation[C] // European Conference on Computer Vision. 2016: 483–499.
- [26] XU K, BA J, KIROS R, et al. Show, attend and tell: Neural image caption generation with visual attention[C] // International conference on machine learning. 2015: 2048–2057.
- [27] XU H, SAENKO K. Ask, attend and answer: Exploring question-guided spatial attention for visual question answering[C] // European Conference on Computer Vision. 2016: 451–466.
- [28] CHEN L, ZHANG H, XIAO J, et al. Sca-cnn: Spatial and channel-wise attention in convolutional networks for image captioning[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 5659–5667.

- [29] WANG X, YU K, DONG C, et al. Recovering realistic texture in image super-resolution by deep spatial feature transform[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018 : 606 – 615.
- [30] TAI Y, YANG J, LIU X, et al. Memnet: A persistent memory network for image restoration[C] // Proceedings of the IEEE international conference on computer vision. 2017 : 4539 – 4547.
- [31] CHOI J-S, KIM M. A deep convolutional neural network with selection units for super-resolution[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2017 : 154 – 160.
- [32] HUI Z, WANG X, GAO X. Fast and accurate single image super-resolution via information distillation network[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018 : 723 – 731.
- [33] 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 (ECCV). 2018 : 252 – 268.