分类号：TP391　　单位代码：10363

密　级：公 开　　学　号:2210320114

题　目基于深度特征交互的图像超分辨率重建算法研究

题目　Research on Image Super-Resolution Algorithm Based on Depth Feature Interaction Reconstruction

学　生　姓　名：储曦

校　内　导　师：王凤随（教授）

专　　业：控制科学与工程

研　究　方　向：数字图像处理

论文答辩日期：2024 年 5 月 23 日

安徽工程大学学位论文原创性声明

本人郑重声明：我恪守学术道德，崇尚严谨学风。所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已明确注明和引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品及成果的内容。论文为本人亲自撰写，我对所写的内容负责，并完全意识到本声明的法律后果由本人承担。

学位论文作者签名：

日期：　　年　　月　　日

安徽工程大学学位论文版权使用授权书

学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅或借阅。本人授权安徽工程大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密 □，在 ____ 年解密后适用本版权书。

本学位论文属于

不保密 □。

学位论文作者签名：　　指导教师签名：

日期：　　年　　月　　日　　日期：　　年　　月　　日

基于深度特征交互的图像超分辨率重建算法研究

摘 要

图像作为视觉信息的重要载体，其分辨率直接决定着视觉信息的丰富程度和颗粒度。由于设备能耗、传输容量和拍摄环境限制等多方面因素的限制，低分辨率图像常常缺失许多关键细节，这不仅损害了图像的观感质量，也削弱了后续图像处理流程的效能。为了解决这一问题，图像超分辨率重建技术得以发展，其核心目标是将低分辨率图像重建成高分辨率图像，从而尽可能地还原图像中所丢失的细节。图像超分辨率重建作为计算机视觉中经典的底层视觉问题，是一个典型的逆问题。低分辨率图像丢失大量的图像纹理细节，而且一张低分辨率图像往往对应多张适配的高分辨率图像，从而导致无法正确地恢复高分辨率图像的图像纹理细节。因此，图像超分辨率重建具有重要研究价值和应用价值，且富有挑战性。本文针对轻量化图像超分辨率重建算法中如何实现从低分辨率图像中学习丢失的图像纹理细节的问题，以及如何正确地恢复高分辨率图像的图像纹理细节的问题，从网络结构、算法设计上展开相关研究，主要工作如下：

（1）轻量化图像超分辨率重建算法主要由卷积神经网络构建，很少使用 Transform 构建。因为 Transform 中多头自注意力机制的高计算复杂度，使得其难以应用在轻量化图像超分辨率重建算法中。为降低 Transform 的计算复杂度，提出一种基于高阶门控交互的单图像超分辨率重建算法。由于低分辨率图像丢失的图像纹理细节为图像的

高频信息。因此，该算法主要挖掘低分辨图像中的高频信息以恢复高分辨率图像。首先使用递归门控卷积代替多头自注意力，降低 Transform 的计算复杂度。递归门控卷积采用递归的形式划分特征的维度，实现特征的高阶空间交互。其次，针对特征的高频分量在空域中难以提取的问题，提出使用残差全局自适应滤波块提取特征中的高频分量，使得特征中的高频分量获得更高的权重，网络的注意力集中于特征的高频部分，提升算法的超分辨重建性能。最后，结合残差全局自适应滤波块与递归门控卷积，重新构建 Transform 网络。最终实验表明，该算法在多个测试数据集上与现有的先进算法相比实现了网络性能与网络计算复杂度之间的最佳平衡。

（2）进一步考虑卷积神经网络的归纳偏置中的局部性，从特征的全局交互角度分析，提出一种基于迭代高阶门控交互的单图像超分辨率重建算法，该算法针对递归门控卷积在特征交互阶数较低时，特征维度过小无法充分交互的问题，提出迭代广域门控卷积，进一步实现特征的全局交互，降低卷积神经网络的归纳偏置中的局部性。该算法通过利用增强混合通道注意力代替 Transform 中的前馈神经网络块，进一步降低 Transform 网络参数量，并提高网络计算效率。该算法为解决残差全局自适应滤波块对特征输入大小过于敏感的问题，在频域中提出利用大核频域卷积重新映射特征的高频区域，大核频域卷积对输入更具有自适应性，而且更轻量化。该算法通过残差控制单元分析残差特征，对网络的残差特征进行约束与优化。通过实验表明，残差控制单元能有效地加速算法收敛，且能提升算法的超分辨率重建性能。

该算法与现有的先进算法相比，具有最小的计算复杂度且实现了最优的结果。

（3）通过将 Transform 架构中有效的组件引入 CNN，优化 CNN 的架构，基于此提出特征蒸馏与结构损失的单图像超分辨率重建算法。通过分析 Transform 特征交互的过程将其拆解为像素的全局交互与通道混洗，首先采用蒸馏链接提升网络通道混洗的能力，为进一步降低特征蒸馏链接的计算复杂度，提高网络的计算效率，采用自蒸馏链接代替传统的蒸馏链接。针对像素的全局交互提出大核卷积块，扩大网络的像素感受野，进一步提升算法的超分辨率重建性能。其次，为解决如何正确地恢复的高分辨率图像的图像纹理细节的问题，设计一种新的损失函数，即结构损失函数，用于约束算法优化网络中的权重使得特征能够更好的交互，提升算法的超分辨率重建性能。最终实验表明，该算法在多个测试数据集上与现有的先进算法相比，以最低的参数量和计算复杂度实现了最优的性能，而且在超分辨率重建的图像有着更好的视觉效果。

关键词：图像超分辨率重建，深度学习，特征交互，卷积神经网络，Transform 网络，轻量化网络，频率域学习

RESEARCH ON IMAGE SUPER-RESOLUTION RECONSTRUCTION ALGORITHM BASED ON DEPTH FEATURE INTERACTION

ABSTRACT

Images serve as crucial carriers of visual information, and their resolution directly determines the richness and granularity of such information. However, due to limitations in equipment energy consumption, transmission capacity, and shooting environments, low-resolution images often lack essential details. This not only compromises the visual quality but also hampers subsequent image processing efficiency. To address this issue, image super-resolution reconstruction technology is developed with the core objective of reconstructing low-resolution images into high-resolution ones to restore lost details as much as possible. Image super-resolution reconstruction is a classic low-level visual problem in computer vision that represents a typical inverse problem. Low-resolution images lose numerous texture details while corresponding to multiple adaptive high-resolution counterparts, thereby hindering accurate restoration of texture details in high-resolution images. In summary, image super-resolution reconstruction holds significant research and application

value while presenting various challenges. This paper focuses on addressing how to learn lost image texture details from low-resolution images within lightweight image super-resolution reconstruction algorithms and how to accurately restore the texture details in high-resolution images through relevant network structure design and algorithmic advancements. The main work is as follows:

(1) The construction of lightweight image super-resolution reconstruction algorithms primarily relies on convolutional neural networks, with limited utilization of Transform. This is due to the high computational complexity associated with the multi-head self-attention mechanism in Transform, making its application challenging in lightweight image super-resolution reconstruction algorithms. To address this issue and reduce the computational complexity of Transform, A single image super-resolution reconstruction algorithm based on high-order gating interaction is proposed. Because the lost information of the low-resolution image is often high-frequency information, the algorithm mainly digs the high-frequency information in the low-resolution image to restore the high-resolution image. Firstly, recursive gated convolution is used to replace the multi-head self-attention to reduce the computational complexity of Transform. Recursive gated convolution adopts the recursive form to divide the dimensions of the features, and realizes the high-order spatial interaction of the features. The second contribution of

this study is the proposal of a residual global adaptive filtering block to address the challenge of extracting high-frequency components in the spatial domain. By assigning higher weights to these high-frequency components, our approach effectively directs the network's attention towards them, thereby enhancing the super-resolution reconstruction performance. Finally, the Transform network is rebuilt by combining the residual global adaptive filtering block and recursive gated convolution. Experiments show that the algorithm achieves the best balance between network performance and network computing complexity compared with the existing advanced algorithms on multiple benchmark datasets.

(2) Further considering the locality of convolutional neural network in inductive paranoia, from the perspective of global interaction of features, a single image super-resolution reconstruction algorithm based on iterative high order gated convolution is proposed. The algorithm aims at the problem that the recursive gated convolution cannot fully interact with the feature dimension too small when the feature interaction order is low, and proposes iterative wide-area gated convolution to further realize the global interaction of features and reduce the locality of convolutional neural network in inductive paranoia. The algorithm further reduces the number of parameters of the Transform network and improves the computing efficiency of the network by replacing the feedforward neural network block in the Transform with enhanced contrast-aware channel attention. To

solve the problem that the residual global adaptive filtering block is sensitive to the size of the feature input, in the frequency, the large kernel frequency convolution is proposed to remap the high frequency region of the feature. The large kernel frequency convolution is more adaptable to the input and more lightweight. The algorithm analyzes the residual features through the residual control unit to constrain and optimize the residual features of the network. Through experiments, the residual control unit can effectively accelerate the convergence of the algorithm and improve the super-resolution reconstruction performance of the algorithm. Compared with the existing advanced algorithms, the algorithm has the minimum computational complexity and achieves the best results.

(3) By analyzing the effective components in the Transform architecture, the CNN architecture is further optimized, and based on this, a single image super-resolution reconstruction algorithm based on feature distillation and structural loss is proposed. By analyzing the process of feature interaction in the Transform, it is disassembled into the global interaction of pixels and channel mixing. Firstly, the distillation link is used to improve the ability of network channel mixing. In order to further reduce the computational complexity of feature distillation link and improve the computational efficiency of the network, the self-distillation link is used instead of the traditional distillation link. For the global interaction of pixels, a large kernel convolution block is proposed to expand the receptive

field of pixels in the network, and further improve the super-resolution reconstruction performance of the algorithm. To solve the problem of errors in the image texture details of the high-resolution image restored by the network, a new loss function is designed, namely the structural loss function is used to constrain the algorithm to optimize the weights in the network so that the features can interact better and improve the super-resolution reconstruction performance of the algorithm. Finally, the experiments show that the proposed algorithm achieves the lowest number of parameters and computational complexity compared with the existing advanced algorithms on multiple test datasets, and achieves the optimal performance, and the super-resolution reconstruction of the image has better visual effect.

KEY WORDS: image super-resolution reconstruction, deep learning, feature interaction, convolution neural network, Transform network, lightweight network, frequency domain learning

目录

第1章 绪论

1.1 研究背景与意义

视觉是人类获取信息的重要方式，对于我们的日常生活和工作来说具有不可替代的重要性。在计算机视觉领域[1, 2, 3]，图像的分辨率对于获取图像信息和还原真实场景具有重要意义。低分辨率的图像往往包含的信息量较少，图像质量差，部分内容丢失，难以满足我们对图像的需求。例如，在图 1-1 中，低分辨率图像中难以清晰获取到优质的内容和图像细节，这不仅影响了人们的日常生活体验，也会影响计算机算法的部署和最终效果。因此，通过算法提高图像分辨率以获取更多的信息，是计算机视觉领域的重要研究方向。这将有助于改善人们的生活质量和提升计算机算法的应用效果。

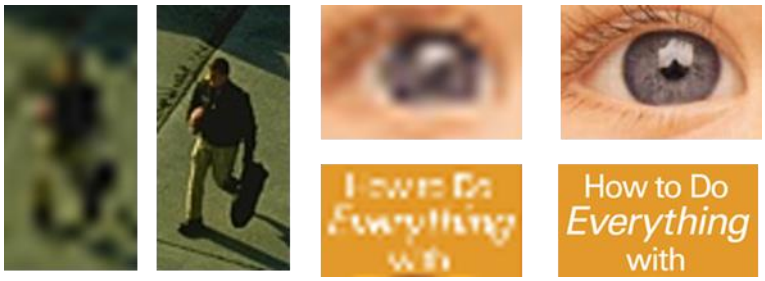


图 1-1 低分辨率与高分辨图像对比图

图像超分辨率(super-resolution，SR)重建是一种通过算法从低分辨率图像生成高分辨率图像的技术。随着数字图像处理技术的发展，越来越多的高分辨率图像被生成，但由于采集设备的限制，这些高分辨率图像的数量仍然有限。因此，图像超分辨率重建技术的发展成为了一项重要的研究方向。通过图像超分辨率重建技术，我们可以从低分辨率图像中提取出更多的信息，从而更好地还原原始场景的细节。这种技术的研究和应用对许多领域具有重要意义。例如，医学影像[4, 5, 6]中的高分辨率图像可以提供更详细的诊断信息，卫星遥感图像[7, 8]的高分辨率版本可以更精确地识别地表特征，工业检测[9, 10, 11]中的高分辨率图像可以提高产品质量控制的准确性。在日常生活中，我们常常需要高质量的图像来获取信息，比如在社交媒体上分享照片或观看电影。低分辨率的图像往往难以提供足够的细节，使得我们难以获得满意的视觉体验。通过图像超分辨率重建技术，我们可以生成更清晰、更逼真的高分辨率图像，提升人们的视觉享受。在许多计算机视觉

应用中，高分辨率图像对于算法的性能和准确性至关重要。例如，图像识别[12, 13]、目标跟踪[14, 15]和人脸识别[16, 17]等任务需要对图像进行详细分析和处理。通过提高图像分辨率，我们可以提供更多的信息和细节，从而提升这些算法的性能和准确性。

综上所述，图像超分辨率重建的研究背景和意义主要在于提升图像质量、拓展应用领域以及促进算法发展。通过不断的探索和创新，图像超分辨率重建技术有望在各个领域发挥更大的作用，为人类的生活和工作带来更多的便利和效益。

1.2 研究现状

图像超分辨率重建根据所做的任务不同，可以进一步分为固定倍数图像超分辨率重建、盲超分辨率重建、细分领域内的超分辨率重建及其轻量化超分辨重建。其中固定倍数超分辨重建、盲超分辨率重建、轻量化超分辨率重建主要面向自然图像，根据所做任务的不同得以区分。对于图像来说自然图像仅占图像中的一部分，根据图像我们可以划分为：自然图像、高光谱图像、双目图像等等。根据采样倍数又可以划分为固定倍数上采样重建任务和任意倍数上采样重建任务。本节首先回顾与总结上述几种超分辨率重建技术，之后对轻量化超分辨率重建技术的研究现状进行进一步的分析与总结。

1.2.1 固定倍数超分辨率重建

固定倍数超分辨率重建主要面向单一退化下的重建任务，例如下采样、去雨、去噪、去模糊等等。随着深度学习技术的快速进步，卷积神经网络(convolutional neural network, CNN)在诸如物体识别、人脸检测等高级视觉任务中展现了卓越的性能。在此背景下，越来越多的研究人员开始尝试探索将深度学习应用到超分辨率重建任务中。例如 Dong 等人[18]第一次提出利用深度学习进行超分辨率重建工作，即 SRCNN(super-resolution convolutional neural network)，该工作首先使用双三次上采样提升低分辨图像的分辨率，随后输入到一个三层的卷积神经网络中学习低分辨率和高分辨率图像间的复杂非线性映射关系，从而显著提高了超分辨率重建的效果。这一开创性的工作促使更多的研究员使用深度学习进行超分辨率重

建任务，Kim 等人[19]在 SRCNN 的基础之上提出一个更深的 SR 网络，即 VDSR(very deep super-resolution network)，网络的性能远优于仅有三层的 SRCNN。上述两个网络使用双三次上采样之后的低分辨率图像输入网络用于训练，这样极大的增加的网络的复杂度，基于此 ESPCN(efficient sub-pixel convolutional neural network)[20]使用低分辨率图像直接输入网络，并使用亚像素卷积层(sub-pixel convolution layer)重建高分辨率图像，基于此往后的多数工作使用亚像素卷积层重建高分辨图像。

伴随高层视觉的发展，残差网络因为其出色的性能表现被广泛应用于各类任务之中。在超分辨率重建领域，研究者们也开始采用残差网络以提高重建效果。其中 Lim 等人[21]首先对残差网络的基础模块进行了创新性改进，去除了残差网络中的批归一化(batch normalization, BN)层，基于这个新的基本残差块构建了一个新的残差网络用于图像超分辨重建，即 EDSR(enhanced deep super-resolution network)。为进一步增加网络的层数，进一步扩大网络的参数量，以提升超分辨率重建网络的性能，Zhang 等人[22]提出 RDN (residual dense network)网络。该网络不仅使用残差链接来提升网络的性能，同时使用密集链接实现多层级之间的特征交互。

在深度学习算法的研究不断深入的过程中，涌现了大量优秀的网络结构。循环神经网络(recurrent neural network, RNN)结构作为一种重要的网络结构，已被广泛应用于高层视觉、自然语义处理等任务中。针对图像超分辨率重建任务，Li 等人[23]使用 RNN 结构构建 SRFBN(super-resolution feedback network)，该网络巧妙地运用了反馈机制处理网络中的特征，即使用上一次循环的输入通过反馈机制进入下一次循环实现特征的高效利用，以提升网络的超分辨率重建性能。此外，随着注意力机制的兴起，Zhang 等人[24]利用通道注意力机制，提出一个新的网络，即 RCAN(residual channel attention network)，进一步挖掘图像中的特征实现更多的特征交互，并让网络的注意力集中在重要特征的通道之上，提升网络的超分辨率重建性能。紧接着，Dai 等人[25]进一步发展了这一思路，SAN (second-order attention network)，该网络通过引入二阶注意力机制，实现了更为精细的像素级注意力控制，显著增强了网络的特征表达能力和特征间的相关性学习，从而推动了超分辨率重建技术的发展。

1.2.2 盲超分辨率重建

固定倍数超分辨重建在面对图像的单一退化取得了良好的效果，但是在真实世界中图像的退化过程不是单一的，往往更加的复杂是多种退化过程叠加所形成低分辨率图像。固定倍数超分辨率重建网络在面对多级复杂退化的低分辨图像时，其超分辨率重建效果大打折扣。由此研究人员为应对真实世界下的复杂退化从而提出盲超分辨率重建任务。

首先研究人员对特定多级退化进行研究，Zhang 等人[26]使用特定的退化模糊核对高分辨率图像进行模糊处理，并嵌入随机噪声。随后将退化过程进行编码与低分辨率图像一起输入网络，以此实现多级退化下的图像超分辨率重建任务。该工作提出 SRMD(super-resolution multiple degradations)，该网络首次实现了对于多级退化下的低分率图像超分辨率重建。Zhang 等人[27]通过半二次分裂算法(half-quadratic splitting algorithm, HQS)构建端到端的网络，即 USRNet (unfolding super-resolution network)，该网络交替的从数据中学习先验，推理最大后验概率来优化网络，从而在不同程度的图像退化情况下都能显著提高超分辨率的效果。这种方法为特定类型的退化图像的超分辨率重建提供了一个创新的路径。

上述工作中对于图像退化过程中的模糊核都是固定的，这显然不服从真实世界图像的退化过程。当图像的退化过程与上述网络所设计的过程差异较大时，其超分辨率重建性能明显受到影响。基于此 IKC(iterative kernel correction)由 Gu 等人[28]提出，该网络显式的估计模糊核，利用超分辨率重建之后的结果更正对模糊核的估计，采用不断迭代的方式优化网络。该网络使用可以更准确地估计图像退化过程中的模糊核，从而提升网络的超分辨率重建性能。Ji 等人[29]设计了一个新的退化模型，旨在更准确地模拟真实世界图像的模糊和噪声特性。通过这个模型，能够生成更接近真实情况的低分辨率图像。其次，构建了一个新的网络 RealSR(real-world super-resolution)，该网络通过对抗生成网络显式的估计退化过程并实现超分辨率重建。

研究人员通过对数据的分析发现，图像的退化过程不仅能显式的估计，也可以通过数据建模进行隐式估计。基于此 Yuan 等人[30]提出 CinCGAN(cycle-in-cycle adversarial networks)，该网络首先将噪声和模糊的输入被映射到无噪声的低分辨率空间。然后使用预训练的网络对降噪后的图像进行上采样。最后，以端到端的

方式微调这两个模块以获得高分辨率输出。Wang 等人[31]通过对比学习将不同退化映射到高维空间进行隐式的建模，并且提出一个由退化感知指导的网络，即 DASR(degradation-aware super-resolution)网络。Wang 等人[32]设计 CRDNet (distortion-specific super-resolution network with contrastive regularization)，网络首先通过全局特征编码对退化进行隐式的建模，以指导网络进行超分辨重建，并采用 U 型网络降低网络的计算复杂度，最后使用对比正则化优化网络输出。

1.2.3 细分领域的超分辨率重建

1.2.3.1 高光谱图像超分辨率重建

高光谱图像通常具有多个波段，每个波段包含丰富的光谱信息，但分辨率通常较低，难以直接用于许多应用。故此，高光谱图像超分辨率重建的主要目的是从低分辨率的高光谱图像中恢复出高质量的高分辨率图像。该技术通常采用深度学习模型，利用大量数据进行训练，以便从低分辨率图像中学习到更高层次的特征表示，进而生成高分辨率图像。高光谱图像超分辨率重建技术在许多领域有着广泛的应用，如农业、环境监测、地质勘探等。

第一个基于深度学习的高光谱图像超分辨率重建算法由 Xie 等人[33]提出，他们基于卷积神经网络提出 MHNet(multispectral / hyperspectral fusion net)，并使用受近端梯度下降(proximal gradient)算法训练该网络。大部分网络利用已知的先验点扩散函数(point spread function)和光谱响应函数(spectral response function)解决高光谱图像超分辨率重建任务，但是这些先验知识极为有限，甚至是不可用的。为解决这一问题 Zheng 等人[34]提出了一种基于无监督深度学习的融合网络，即 HyCoNet(hyperspectral coupled network)。为进一步使得空间和光谱信息得到更有效地交互，Hou 等人[35]提出 PDE-Net(posterior distribution-based embedding network)，该网络从概率的角度描述如何将高维光谱信息嵌入像素空间中，并提出了一种隐式的高光谱特征嵌入方案，从而提升网络超分辨率重建性能。Zhang 等人[36]提出使用对偶常微分方程(dual ordinary differential equations, Dual ODEs)设计高光谱图像超分辨率网络，该网络设计了多分支空间-光谱特征提取模块，使得网络中的深度特征实现了更高效率的交互，更充分地挖掘低分辨率特征中的

有效信息重建高分辨率高光谱图像。随着研究者对高光谱图像超分辨率重建技术的深入探索，网络的性能和运行效率均得到显著提升，因而在实际应用中得到了日益广泛地推广。

1.2.3.2 双目图像超分辨率重建

双目图像具有两个视角，每个视角都包含了物体的不同角度和深度信息。通过将这两个视角的图像结合起来，可以利用视差（即同一物体在两个视角下的位置差异）来恢复出更高质量的高分辨率图像。根据实际需求和应用场景，将双目超分辨率重建技术应用于各个领域，如医疗、安防、自动驾驶等。同时持续改进和优化算法性能以提高应用效果和效率。

Jeon 等人[37]开创性地开发了一种基于深度学习的双目图像超分辨率重建方法，该方法通过隐式学习机制掌握了双目低分辨率图像与对应高分辨率图像间的复杂映射规律。具体操作上，他们首先对输入的双目图像进行平移操作以实现初步的对准，然后将这些图像送入卷积神经网络中，由网络自动学习并建立起低分辨率到高分辨率图像的映射模型。从网络的角度来说，使用双网络分别学图像的亮度与图像的色彩，达到更好的重建效果。Dai 等人[38]通过在网络的输出对视图的视差进行估计，代替在输入图像平移以对齐重建后的高分辨率图像进一步提升重建性能。在网络上使用递归与迭代同时完成视差估计以及双目图像超分辨率重建。Zhang 等人[39]提出 ACLRNet(Attention-guided correspondence learning restoration network)，该网络使用注意力引导的对应学习方法，在视差和全向注意力的指导下学习自视图和跨视图的特征，并在学习到的对应关系的基础上，通过关联两个视图的特征来恢复立体图像。

1.2.3.3 任意倍率图像超分辨率重建

已有的图像超分辨率重建技术面对不同下采样尺度都要单独训练网络，不能用一个模型同时处理多个上采样倍数任务。故此，任意倍率图像超分辨率重建任务被提出，用于处理需要不同超分倍率的应用场景。首先 Lim 等人[21]提出一个新的网络，即 MDSR (multi-scale deep super-resolution)，该网络将上采样倍数为 2、3、4 的网络整合为一个网络可以直接进行推理。但是面对实际情况，大多数

时候 2、3、4 倍数无法应对复杂多变的实际情况，故此 Lu 等人[40]提出 Meta-SR (meta-learning for super-resolution)，该网络在末端使用以超分倍数为指导的上采样滤波器实现任意倍率超分辨重建。Wang 等人[41]进一步设计多尺度感知特征适应块和尺度感知上采样层组成一个新的任意尺度超分辨率重建网络，即 ArbSR(scale-arbitrary super-resolution)，其中尺度感知上采样模块依靠动态卷积实现，并设计多尺度感知特征，而且该网络实现非对称倍数超分辨率重建（长宽分别进行 2.5 倍和 4 倍超分辨）。Meta-SR 通过显式建模实现超分倍数的编码与上采样滤波器对齐，而这种对齐方式需要额外的输入而且很难实现较好的超分效果。Chen 等人[42]构建一种隐式映射对齐采样倍数与上采样滤波器，即 LIIF(local implicit image function)。该网络以图像坐标和坐标周围的特征作为输入，预测给定坐标处的 RGB 值作为输出。Yu 等人[43]提出 SAFANet(scale aware frequency attention network)，该网络首先引入频域注意力学习来优化尺度自适应特征，使其明确关注高频成分，从而提高网络超分辨率重建性能。同时设计尺度引导的连续重建模块利用隐式重建函数重建任意分辨率的图像，从而自适应地实现图像的任意尺度超分辨率重建。

1.2.4 轻量化超分辨率重建

上述超分辨重建任务中主要针对网络的性能，但是在实际运用当中这些网络是难以部署在边缘设备例如：手机、车载设备等，或是需要实时应用的设备上的例如：VR 眼镜、直播等。由于部署端的设备或是计算资源受限，大部分网络模型难以部署。故此研究人员提出轻量化超分辨率重建以此应对现实生活中设备或是计算资源受限的情况。

基于网络架构的设计，Dong 等人[44]首先提出该任务通过改进 SRCNN 提出 FSRCNN(fast super-resolution convolutional neural network)，该网络直接采用低分辨率图像直接输入网络，最后使用转置卷积生成高分辨率图像，这样使得网络的计算复杂度大大下降。并使用更多的卷积层处理特征以提高网络的超分辨率重建性能。Shi 等人[20]通过分析上采样层提出一种新的上采样方式即亚像素卷积层(sub-pixel convolution layer)，并使用该上采样层构建了一个新网络 ESPCN(efficient sub-pixel convolutional neural network)更高效的实现了超分辨率

重建。Hui 等人[45]提出利用特征蒸馏的方式细化不同层级之间的特征，并以此提出网络 IDN((information distillation network)。特征蒸馏进一步消除网络中的冗余通道加快网络的推理速度以及内存消耗。在 IDN 的基础上，Hui 等人[46]进一步提出 IMDN(information multi-distillation network)，该网络首先进一步改进了特征蒸馏的方式，更优的实现了网络的计算复杂度与网络性能之间的平衡。Liu 等人[47]提出 RFDN(residual feature distillation network)，该网络利用点卷积进行特征蒸馏进一步简化了网络。实现了更低的计算复杂度，且取得了更优的超分辨率重建效果。Luo 等人[48]利用蝶形结构设计网络实现 LatticeNet(lattice network)，该网络实现了分层提取上下文信息，使得特征得到充分地交互，提升网络超分辨率重建的性能。研究人员通过对网络架构的不断改进，轻量化超分辨率重建算法得到不断的发展。

基于轻量化技术的设计，权重共享，Kim 等人[49]使用递归学习的方式构建网络 DRCN(deeply-recursive convolutional network)，通过卷积之间的权重共享，极大的降低的网络的参数量，并且实现了优异的性能。Tai 等人[50]进一步改进了 DRCN 中的残差结构并以此提出 DRRN(deep recursive residual network)，该网络提出局部残差链接进一步增强超分性能。神经网络架构搜索，Chu 等[51]首先应用这一技术在轻量化超分辨率重建上提出网络 FALSR(fast, accurate and lightweight super-resolution)进一步实现了超分性能与计算复杂度之间的平衡。在此基础上，Zhan 等[52]为实现实时移动端超分，进一步对网络进行剪枝处理，该网络在推理阶段甚至只需要 40ms 实现了在移动端的实时超分。知识蒸馏，知识蒸馏技术在高层视觉中有着广泛的应用，但是在超分辨重建里的应用较少。Lee 等[53]首次将该技术应用到轻量化超分辨率重建中，具体的来说通过构建变分自编码器架构，利用教师模型训练学生模型中的解码器以生成高分辨率图像。在此基础上，Wang 等人[54]通过自蒸馏的方式训练教师与学生网络，其能更进一步压缩模型大小，降低模型复杂度，更好的实现模型性能与计算复杂度之间的平衡。结构重参化，结构重参化技术将多个卷积的权重在网络推理阶段重参化为一个卷积，使得网络的运行速度大大提升。Zhang 等人[55]提出 ECBSR (edge-oriented convolution block for super resolution)网络引入了拉普拉斯算子提取特征的细节纹理从而提升网络超分性能，而且在推理时多个卷积会结构重参化为一个卷积加速网络推理。在此

基础上，Wang 等人[56]提出 RepSR(re-parameterization and batch normalization for super resolution)网络，该网络首先去除 ECBSR 中使用的拉普拉斯算子提高网络的训练速度。同时对网络中的批归一化层进行分析，在网络推理时冻结批归一化层，以杜绝批归一化层对网络带来的负面影响。RepSR 进一步提升了网络的超分性能。随着各钟新技术的出现，轻量化超分辨率重建的方法也随之增多，相信会有更多有效的方法被提出。

基于注意力机制的设计，Hui 等人[47]进一步提出的 IMDN，该网络中使用了对比度通道注意力，简单的来说通过计算特征的方差利用多层感知机混洗特征最后对输入特征做乘积实现对比度通道注意力机制，提升网络的超分性能。进一步的，Li 等人[57]通过在对比度通道注意力上加入特征均值信息进一步修改对比度通道注意力机制，同时引入增强的空间注意力机制[58]进一步提高模型的性能。Park 等人[59]通过自注意力机制设计一种针对残差的注意力网络，即 DRSAN(dynamic residual self-attention network)，有效的改进了残差结构，同时设计了不同大小的网络以应对不同的部署环境。Gao 等人[60]通过多路径卷积设计了一种高频注意力机制，有效的提取特征中的高频信息，并由此构建网络，即 VLESR(lightweight and efficient super-resolution)。Zhao 等人[61]通过对空间注意力与通道注意力的分析，进一步提出像素注意力机制。并由此构建网络，即 PAN(Pixel attention network)。此后基于像素注意力机制的改进[62, 63, 64]层出不穷。Liu 等人[64]针对图像的纹理细节提出 CTE-Net(Contextual Texture Enhancement Network)并利用自注意力机制进一步修改像素注意力机制和通道注意力机制，并设计模块融合注意力子图，提升了网络的超分辨重建性能，恢复更优的图像纹理细节。

基于 Transform 的设计，随着高层视觉的发展 Transform 架构[65, 66]展现出比 CNN 架构更优秀的性能，自然众多底层视觉的研究员也将 Transform 架构引入底层视觉[67]在超分辨率重建领域取得了巨大的进展。但是 Transform 架构往往伴随着模型高复杂度和高的内存消耗，难以应用于轻量化超分辨率重建中。由此，Zou 等人[62]利用 CNN 与 Transform 构建了一个混合网络即 SCET(self-calibrated efficient transformer)，该网络首先进一步改进了像素注意力并网络的最后使用 Transform 块混合深层特征，进一步提升网络的超分辨率重建性能。Gao 等人[68]

在网络进一步增加 Transform 块，综合考虑通道注意力与空间注意力结构充分利用 CNN 与 Transform 构建网络 LBNet(lightweight bimodal network)。Lu 等人[69]利用权重共享设计 CNN 网络，并对 transform 中的自注意力机制进一步改进提出 ESRT(efficient super-resolution transformer)网络，进一步提升网络超分辨率重建的性能。上述网络通过组合 CNN 与 Transform 的组合构建网络，但如何构建一个完全基于 Transform 架构的轻量化超分辨率重建网络依然是一个挑战。Liu 等人[70]根据 Swin-Transform[66]提出 SwinIR(swin-transform image restoration)网络首次实现了一个完全由 Transform 架构实现的轻量化超分辨率重建网络。随后，Choi 等人[71]对 Transformer 中的基于滑动窗口的自注意力机制进一步改进，提出一种自适应的窗口更多的获取特征信息，实现网络性能的提升，但是在网络的计算复杂度上也略微有提升。Wang 等人[72]通过分析网络中的感受野，提出一种性能自注意力机制即，全局自注意力(omni self-attention, OSA)机制，并由此构建 Omni-SR(omni super-resolution)网络，由于 OSA 更全局的关注特征，网络超分辨率重建性能取得了巨大的提升。Zheng 等人[73]提出 EMT(Efficient Mixed Transformer)网络，该网络设计了两个模块，首先设计了条纹窗口注意力模块，该模块使用条纹窗口注意力(striped window self-attention, SWSA)机制代替 Transform 中的滑动窗口注意力机制，以获得更大的感受野提升网络的超分辨率重建性能，其次设计像素混洗模块，该模块利用像素混洗机制代替 Transform 中的多头自注意力机制，从而降低网络复杂度，最后实现网络性能与计算复杂度之间的平衡。

轻量化超分辨率重建的总结，首先对于网络架构来说轻量化超分辨率重建已有重大的进展，但是上述网络仅从像素域去考虑处理特征，没有从频域的角度考虑处理特征。从轻量化技术的设计角度来说，虽然众多的技术被应用于轻量化超分辨重建任务中，这降低了算法的计算复杂度，但是构建出的网络性能有所损耗，如何更合理的应该轻量化技术还有待思考。从注意力的角度来说，研究人员已经设计了很多优越的注意力机制，但是这些注意力机制只是简单的添加在网络上并没有代替网络的某一部分。故此，如何添加在网络中使用注意力机制，仍然需要思考。从 Transform 的角度来说，现有算法虽然对 Transform 架构有所改进，但是多头自注意力中的矩阵相乘带来的巨大计算量还是难以解决。故此，如何设计一个具有较低计算复杂度的 Transform 网络依然是一个挑战。

1.3 论文的主要研究内容

在实际生活中超分辨率重建往往是要部署与边缘设备上例如：手机、相机、车载摄像头等，或是部署于需要实时的应用上例如：VR 眼镜、游戏画面优化等。故此本文针对轻量化超分辨重建任务展开研究。通过上述章节对轻量超分辨率重建任务的分析，本文主要解决网络中计算复杂度高，内存消耗大等问题。其次，为了进一步提升网络在图像超分辨率重建方面的性能，本文从特征提取的视角出发，对网络结构进行了细致的分析与优化，旨在研究如何更有效地构建轻量化且高效的超分辨率重建网络。通过这种方式，能够更好地捕捉和利用图像特征，从而达到提升网络超分辨率重建性能的目的。主要研究内容如下：

（1）基于高阶门控交互的单图像超分辨率重建算法

首先对 CNN 与 Transform 架构进行分析，提出在 Transform 中的多头自注意力机制不一定是 Transform 架构展现良好效果的主要原因，但是多头自注意力机制中的矩阵运算带来了大量的计算复杂度，所以有必要使用另外更高效的方法代替多头自注意力机制降低 Transform 模型的高复杂度并保持一定的超分辨率重建性能，实现模型性能与计算复杂度之间的平衡。具体地说，首先利用高阶门控交互模块代替 Transform 架构中的多头自注意力机制，重新构建一个没有多头自注意力的 Transform 架构。其次，从频域的角度提出一种自适应的滤波器，提升高频细节在特征中的权重，从而帮助网络恢复图像的纹理细节。

（2）基于迭代高阶门控交互的单图像超分辨率重建算法

对高阶门控交互模块进一步分析，发现一种新卷积单元，即广域门控卷积。他与普通卷积有类似的感受野但是参数量远远低于普通卷积。随后在使用迭代的方式使特征进行交互，由此设计了一个新的模块迭代高阶门控交互模块。其次，通过分析轻量化超分辨重建网络中的常用注意力机制，提出一种性能像素注意力机制即大核频域卷积注意力，此注意力类似于像素注意力但是能更全局的混合像素特征。通过广域门控卷积进一步修改增强的空间注意力提升网络性能。进一步的分析 Transform 架构，发现其中的多层感知机主要作用为对于特征在通道上的混洗与重排，于是采用一种增强的均值对比注意力机制以更小的运算量实现对通道的混洗与重排，进一步筛选对网络性能有增益的特征提高网络性能。最后为分

残差链接在网络中的作用于影响提出残差控制单元，通过可学习的方式控制残差特征加速网络收敛，并提高网络性能。

（3）基于特征蒸馏与结构损失的单图像超分辨率重建算法

特征蒸馏是轻量化超分辨率重建构建网络的常用方式，通过分析特征蒸馏的过程，提出一种新的蒸馏方式，即特征自蒸馏。其有效的降低了网络的参数量与计算复杂度。其次，在设计网络的过程中引入大核卷积自注意力模块，扩大网络对于特征的感受野进一步提升网络性能。通过对 Transform 架构的分析选择层归一化的位置并进行实验。最后，设计了一个崭新的损失函数即结构损失，其有效地提取图像中的结构信息约束网络权重，并使得网络恢复出更优质的图像纹理细节。

1.4 论文的组织结构

论文内容分为六章，其各章内容如下所示：

第一章为绪论，首先重点阐述超分辨率重建的研究背景与意义，并介绍了超分辨率重建在实际生活中的应用。随后详细阐述了超分辨重建的研究现状，其通过不同的任务分别进行介绍其发展历程与重要算法。最后介绍本文的三方面研究内容，以及论文结构安排。

第二章为图像超分辨重建的基础理论，首先给出了图像超分辨重建问题的定义。随后，进一步介绍从图像降质模型和图像重建模型，其中图像重建模型将从网络的角度划分两个部分进行叙述：基于卷积神经网络、Transform 的超分辨率重建算法。最后，介绍图像超分辨重建的评价指标。

第三章为基于高阶门控交互的单图像超分辨率重建算法，该章节主要由三节组成，首先主要探讨了 Transform 架构中的多头自注意力机制，对网络的影响。介绍设计该算法的动机与整体算法流程。随后，重点介绍网络的主要模块设计思路与构成。最后，介绍算法的消融实验与最后对比结果。

第四章为基于迭代高阶门控交互的单图像超分辨率重建算法，该章节也由三节组成，首先进一步分析 Transform 架构中各个模块的作用，进一步简化 Transform 架构构建网络。随后重点介绍网络中迭代广域门控卷积、改进的注意力机制和残差控制单元。最后介绍算法的消融实验与最后对比结果，并综合分析

算法。

第五章为基于特征蒸馏与结构损失的单图像超分辨率重建算法，该章节由四部分组成，首先总结上述两章的工作，并重新思考 CNN 架构尝试设计一种崭新的架构进一步实现算法的性能与计算复杂度之间的平衡。随后，介绍网络的设计与新损失函数的构建。最后，介绍算法的消融实验与最后结果对比。

第六章总结与展望，首先总结本文所做的所有工作，然后对总结算法不足以及未来的工作。

第2章 图像超分辨率重建的基础理论

2.1 引言

本章首先叙述超分辨率重建问题的定义，其中包括图像降质模型与图像重建模型。随后，介绍超分辨率重建中常用的数据集与评价指标。最后，对本章进行总结。

2.2 图像超分辨率重建问题的定义

图像超分辨率(super-resolution, SR)重建是一种利用图像处理手段提升图像分辨的技术，旨在从包含有限信息的低分辨率(low-resolution, LR)图像中恢复重建出细节更丰富、更加清晰的高分辨率(high-resolution, HR)图像。如图 2-1 所示，图像超分辨率重建相当于图像退化过程的逆向操作，是一个数学上典型的逆问题，具有显著的欠定性和病态性。图像超分辨率重建作为计算机视觉中经典的低层视觉问题，不仅在学术研究中具有重要的科学价值，在实际中也有着广阔的应用价值。

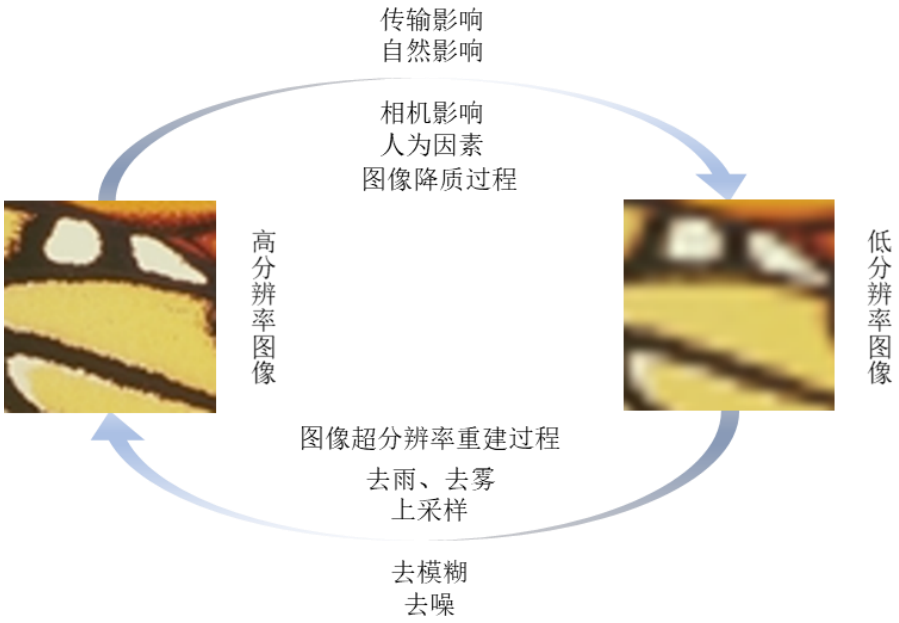


图 2-1 图像降质与超分辨率重建

2.2.1 图像降质模型

图像降质是一个多级复杂的过程，对于实际生活中图像降质可能是因为图像传输过程中压缩降质例如：JEPG 压缩伪影等，拍摄过程中自然环境影响例如：雨雾天气等，相机内部电路影响例如：电流噪声等，人为因素影响例如：运动模糊等。由于实际生活中图像降质过程过于复杂，不便于研究。故此，研究员提出了多种图像降质退化假设，便于对图像降质进行建模。图像的质量降低可以通过多种不同的退化模型来模拟，具体地说，图像降质退化模型可分为：双三次下采样(bicubic-down, BI)降质、高斯模糊(blur-down, BD)降质、双三次插值下采样和高斯白噪声(bicubic-down and noise, BN)降质、高斯模糊下采样和高斯白噪声(blur-down and noise, BN)降质模型。

首先介绍最简单的双三次下采样降质模型，这种降质模型仅对高分辨率图像进行双三次下采样生成低分辨率图像。这种降质过程难以模拟实际生活的降质过程，但是从超分辨重建的角度来说，在降质的过程中高分辨率图像的纹理细节大量丢失，且难以恢复，故大部分图像超分辨重建的低分辨率图像均由该降质模型生成。双三次下采样降质模型可以用公式(2-1)表示如下：

$$I_{\mathrm{LR}}=I_{\mathrm{HR}}\downarrow_{s} \tag{2-1}$$

式中，I_{HR} 代表高分辨率图像，I_{LR} 低分辨率图像。$\downarrow_{s}$ 代表 s 倍的双三次下采样。

高斯模糊降质模型对应于现实生活中的运动模糊、相机模糊等。这种降质过程相较于 BI 过程更复杂，具体的说就是高分辨率图像与高斯模糊核进行卷积得到降质后的图像，其可以概括为如公式(2-2)。

$$I_{\mathrm{LR}}=\left(I_{\mathrm{HR}}*k\right)\downarrow_{s} \tag{2-2}$$

式中，k 为模糊核一般选取不同方向相同性质的高斯模糊核，$*$ 表示卷积操作。

双三次插值下采样和高斯白噪声，即在图像 BI 降质过程后对图像施加噪声。如公式(2-3)表示：

$$I_{\mathrm{LR}}=I_{\mathrm{HR}}\downarrow_{s}+n \tag{2-3}$$

式中，n 表示白噪声一般选取高斯白噪声。

高斯模糊下采样和高斯白噪声降质模型，是上三种降质模型的合集。这种降质过程对应于实际生活中的降质模型，但是这种降质往往难以恢复，而且当降质

的方式不匹配时恢复出的高分辨率图像往往没有很好的效果。这种降质模型生成的低分辨率图像主要用于盲超分的训练。其可以用公式(2-4)表示：

$$I_{\mathrm{LR}}=\left(I_{\mathrm{HR}} * k\right)\downarrow_{s}+n \tag{2-4}$$

进一步的，研究人员为应对 JEPG 压缩带来的伪影，往往会在 BN 降质之后再进行 JEPG 压缩。如公式(2-5)所示。

$$I_{\mathrm{LR}}=\left(\left(I_{\mathrm{HR}} * k\right)\downarrow_{s}+n\right)_{\mathrm{JEPG}} \tag{2-5}$$

式中，$(\bullet)_{\mathrm{JEPG}}$ 代表 JIEPG 压缩降质。

2.2.2 图像重建模型

上一节本文介绍了对于图像退化问题的建模，但构建一个图像超分辨重建模型是复杂而又困难的，已有的一些不基于深度学习的方法例如：最近邻插值[76, 77]、双三次插值[78]、最大后验概率法[79, 80]、迭代反投影法[81]等等。但是随着深度学习的发展，基于深度学习的方法实现的效果要远远优于这些传统方法。本文将基于深度学习的图像超分辨率重建模型，定义为如公式(2-6)：

$$I_{\mathrm{SR}}=f\left(I_{\mathrm{LR}}, \theta\right) \tag{2-6}$$

式中，I_{SR} 为超分辨率重建之后的图像，I_{LR} 为输入的低分辨率图像。f 为一个深度神经网络表示 I_{LR} 与 I_{SR} 之间的映射关系。θ 为 f 中的参数。

众所周知深度学习模型中最重要的部分是网络的构成，例如给定数据集$\mathcal{D}$中训练数据集$\mathcal{D}_{\mathrm{Tr}}$与测试数据$\mathcal{D}_{\mathrm{Te}}$，使用两个不同的神经网络定义为 f^{1} 和 f^{2}，其中 f^{1} 拟合数据能力远强于 f^{2}。使用相同的损失函数进行训练，所得到的结果往往是大相径庭的。具体的说，利用训练完的神经网络 f^{1} 和 f^{2} 分别在测试数据集$\mathcal{D}_{\mathrm{Te}}$上分别计算超分后的图像与高分辨率图像的误差均值。如公式(2-7)与公式(2-8)所示。

$$loss_{1}=\mathbb{E}_{\mathcal{D}_{\mathrm{Te}} \subset \mathcal{D}}\left(I_{\mathrm{HR}}-f_{\mathcal{D}_{\mathrm{Tr}}}^{1}\left(I_{\mathrm{LR}}, \theta^{1}\right)\right) \tag{2-7}$$

$$loss_{2}=\mathbb{E}_{\mathcal{D}_{\mathrm{Te}} \subset \mathcal{D}}\left(I_{\mathrm{HR}}-f_{\mathcal{D}_{\mathrm{Tr}}}^{2}\left(I_{\mathrm{LR}}, \theta^{2}\right)\right) \tag{2-8}$$

式中，I_{HR} 为测试数据集中的高分辨率样本。$f_{\mathcal{D}_{\mathrm{Tr}}}^{i}(\bullet)$ 为第 i 个在训练数据上训练的神经网络。θ^{i} 为对应神经网络的参数。$\mathbb{E}_{\mathcal{D}_{\mathrm{Te}} \subset \mathcal{D}}(\bullet)$ 代表在测试数据集$\mathcal{D}_{\mathrm{Te}}$的均值。

最后我们会必然得出一个结论 $\mathrm{loss}_1 \ll \mathrm{loss}_2$。这是因为 f^{1} 拟合数据的能力远

强于 f^2。故，众多研究员开始聚焦于如何设计一个拟合数据能力强的神经网络。接下来本文将介绍两种在深度学习中最常用的网络：卷积神经网络和 Transform 网络。

2.2.2.1 卷积神经网络

CNN 主要使用卷积与激活函数构建网络用于学习低分辨率图像与高分辨率图像之间的关系。因为本文主要对图像（二维信号）进行处理。故，首先介绍二维卷积，卷积操纵通过对卷积核进行翻转，随后在输入图像对卷积区域求和得到一个新的像素，随后进行平移对后续像素进行相同操作。二维卷积公式如公式(2-9)所示。

$$y=\sum_{i=0}^{k}\sum_{j=0}^{k}g(i,j)\times h(k-i,k-j) \tag{2-9}$$

式中，y 为卷积之后的输出。k 为卷积核的大小。g 为输入图像。

为更清晰地展示二维卷积操作，本文在像素层面上对卷积操作进行可视化，选择一个 3 × 3 大小已经翻转过的卷积核对图像进行卷积，如下图 2-2。图中仅为卷积操作中的第一步，卷积核在图像的对应区域进行加权求和，随后伴随着平移对图像的所有像素进行遍历得到最后输出的图像。

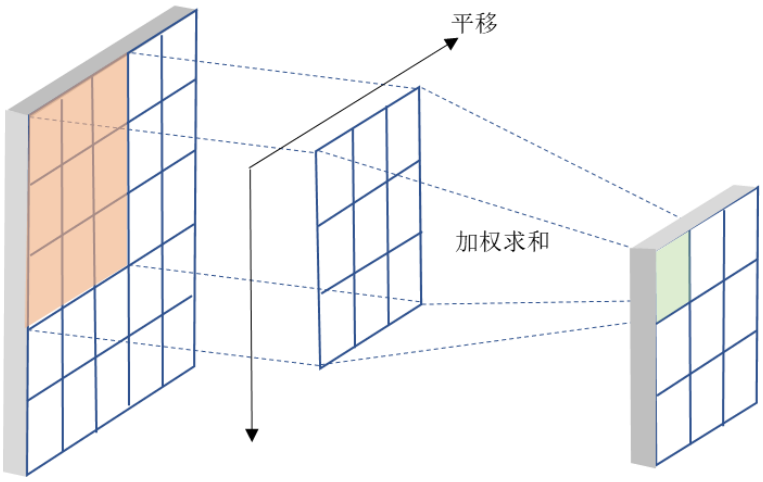


图 2-2 二维卷积示意图

在 CNN 中不仅有传统卷积，也有一些新的卷积例如：深度可分离卷积[82]，蓝图卷积[83]，深度重参化卷积[84]，可变性卷积[85]，蛇形卷积[86]等。接下来本文将重点介绍深度可分离卷积与蓝图卷积中的两个组成部分：深度卷积与点卷积。

因为本文不使用深度重参化卷积、可变形卷积、蛇形卷积，所以在此不过多叙述。

深度卷积其实就是按照输出维度进行分组的分组卷积，为便于叙述本文设输入与输出的特征维度均为 3，其运算过程如下图 2-3 所示。

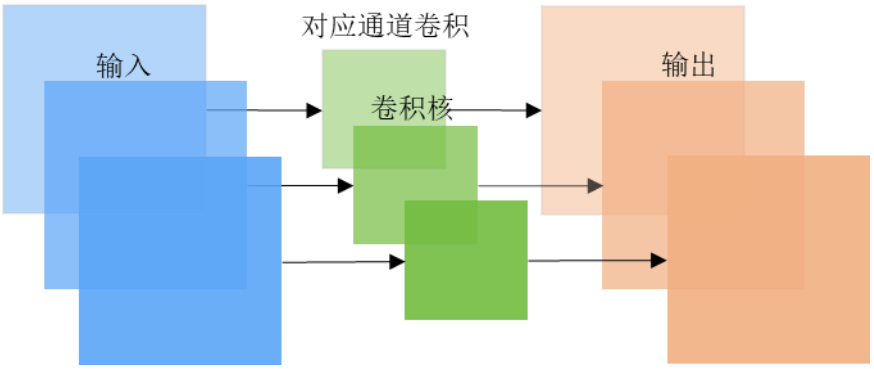


图 2-3 深度卷积示意图

如图 2-3 输入维度为 3 每个特征仅对应一个卷积核分别对特征进行卷积得到输出。接下来介绍点卷积，为便于介绍输入输出维度均设置为 3。点卷积与普通卷积相同，但是其卷积核大小为 1，其运算过程如图 2-4 所示。其中卷积核分为了三组，每组三个卷积核分别卷积三个输入特征，得到输出。

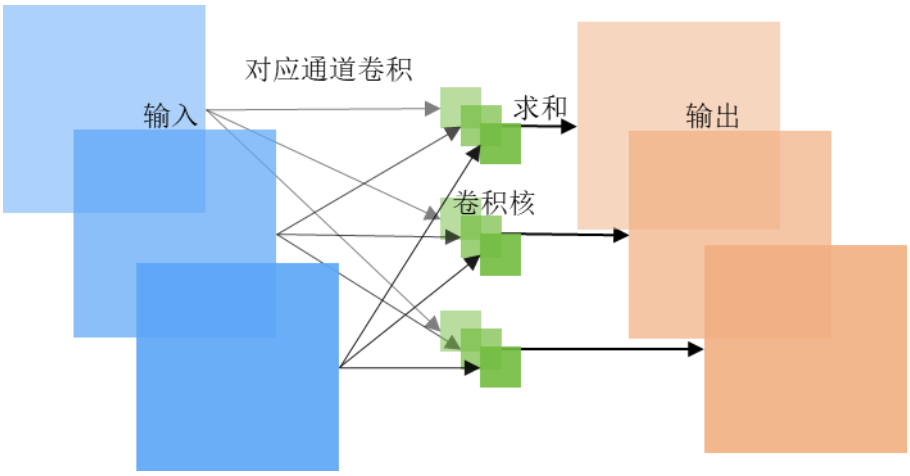


图 2-4 点卷积示意图

在卷积神经网络中，激活函数的作用至关重要。由于网络每一层的基本计算是线性的，仅仅涉及输入信号的加权求和，若无激活函数介入，即便增加网络层数，最终输出也仅为输入数据经过一系列线性变换的结果。然而，现实世界中许多问题的解决需要捕捉和表达非线性关系，而线性变换本身无法实现这一点。因此，引入激活函数能够为网络引入非线性因素，使得网络具备学习和表示复杂函数映射的能力，从而有效处理各种复杂的模式识别和预测任务。所以在深度学习

模型中，激活函数的选择对于模型的性能至关重要。不同的激活函数具有不同的特性，适用于不同的任务。以下是几种常见的激活函数：

ReLU 函数[82](Rectified Linear Unit)：ReLU 函数是目前最常用的激活函数之一，其公式如公式(2-10)所示。

$$f(x)=\max(0,x) \tag{2-10}$$

ReLU 函数的优点是计算速度快，而且可以有效地避免梯度消失问题。然而，当输入小于 0 时，ReLU 函数的梯度为 0，这意味着对应的神经元将不会对后续的学习过程产生影响，导致模型无法学习到这些神经元所代表的特征。

PReLU(Parametric ReLU)函数[88]：PReLU 函数是 ReLU 函数的改进版，如公式(2-11)所示。

$$f(x)=\begin{cases}\alpha\cdot x & x<0\\ x & x\geq 0\end{cases} \tag{2-11}$$

PReLU 函数的优点是可以缓解 ReLU 函数的当输入小于 0 时，ReLU 函数的梯度为 0 的问题，同时也可以通过调整 α 的值来控制函数的非线性程度。

Sigmoid 函数[89]：Sigmoid 函数的表达式如公式(2-12)。

$$f(x)=\frac{1}{1+e^{-x}} \tag{2-12}$$

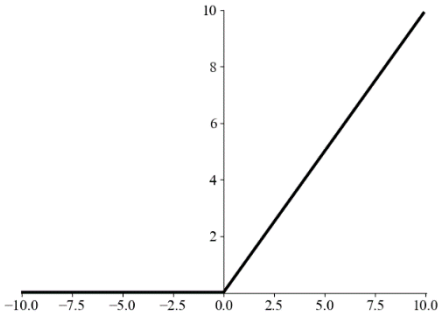
Sigmoid 函数可以将任意范围的输入值映射到(0, 1)之间，常用于将神经元的输出值转化为概率分布。然而，Sigmoid 函数在输入值较大或较小时，其梯度接近于 0，可能会导致梯度消失问题。

Tanh 函数[90]：Tanh 函数的表达式为公式(2-13)。

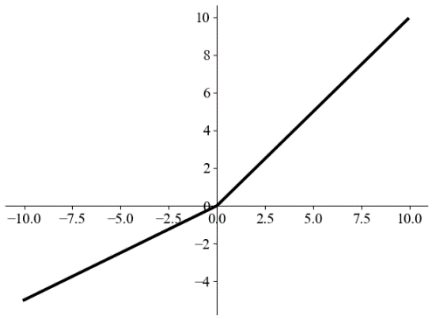
$$f(x)=\tanh(x) \tag{2-13}$$

Tanh 函数可以将任意范围的输入值映射到(-1, 1)之间，与 Sigmoid 函数类似，常用于将神经元的输出值转化为概率分布。与 Sigmoid 函数相比，Tanh 函数在输入值较大或较小时的梯度更大，可以缓解梯度消失问题。

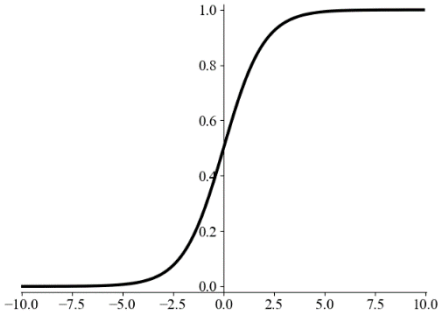
除了上述激活函数，还有一些其他的激活函数如 Softmax 函数、Leaky ReLU 函数等，在此处不一一列举。



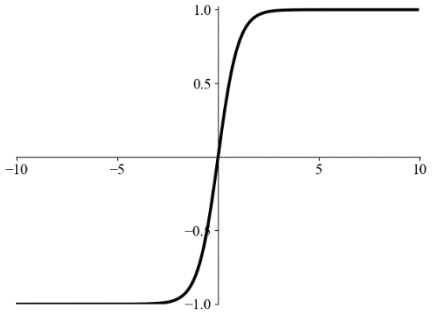
(a) ReLU 函数示意图



(b) PRelu 函数示意图



(c) Sigmoid 函数示意图



(d) Tanh 函数示意图

图 2-5 激活函数示意图

研究人员设置了不同的激活函数放置于卷积层之后，使得神经网络能够拟合复杂的非线性函数。故此，在超分辨率重建任务中 CNN 占据着重要的地位。

2.2.2.2 Transformer 网络

Vision Transformer (ViT)[65]是一个基于 Transformer[91]的自注意力机制的深度学习模型，专门为图像识别任务而设计。自从它在 2020 年被提出以来，ViT 已经在计算机视觉领域产生了广泛的影响，被认为是计算机视觉领域的一个重大突破。ViT 的主要结构是多个 Transformer 编码器层，每个编码器层都包含一个多头自注意力子层和一个前馈神经网络层。多头自注意力子层负责计算每个图像块之间的相关性，而前馈神经网络层则负责将图像块向量映射到更高的维度空间。通过这种方式，ViT 能够捕获图像中的全局依赖关系并有效地提取图像的特征。

Transform 架构可以归结为如下图 2-6 所示，其由层归一化，多头自注意力机制以及一个残差链接组成第一个块，第二个块由层归一化前馈神经网络以及一个残差链接组成。

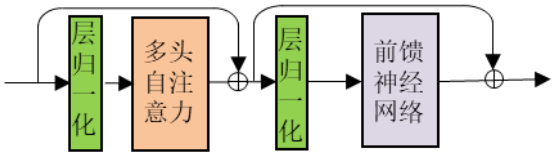


图 2-6 Transform 主要结构

在 Transform 结构中多头自注意力机制，使得特征实现全局特征交互。自注意力机制与多头自注意力机制如下图 2-7 所示。

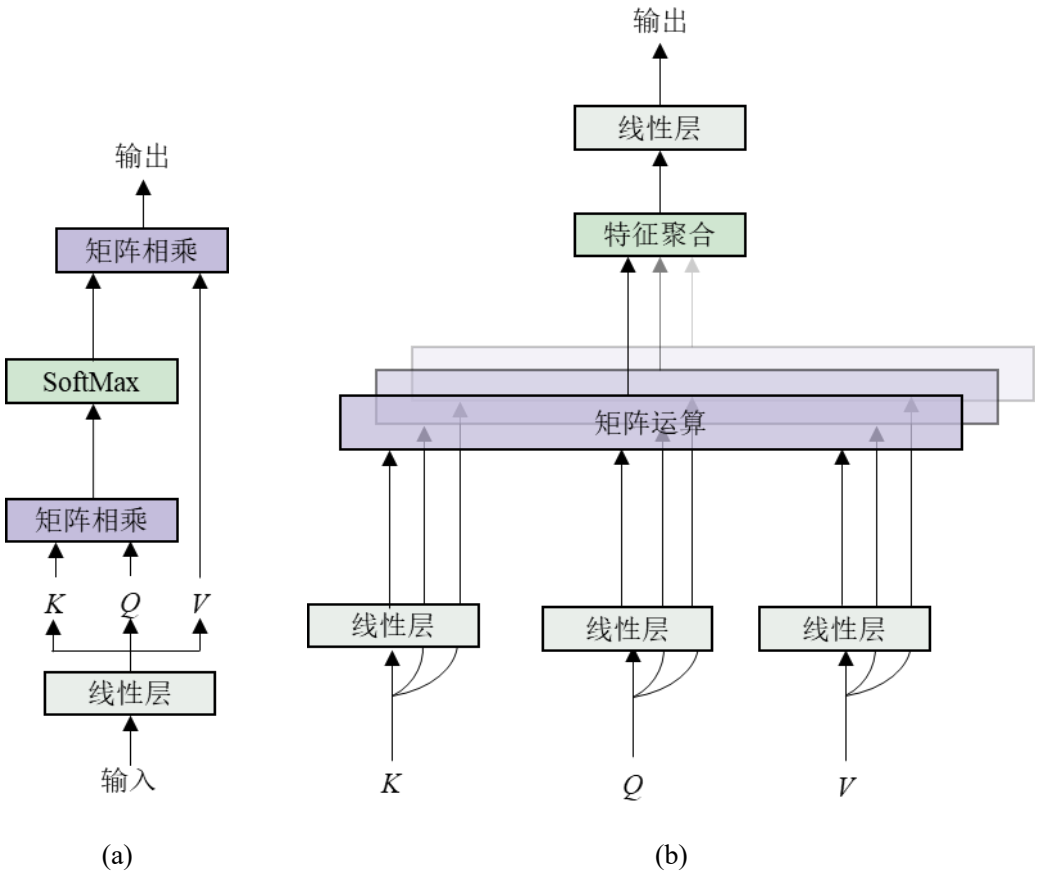


图 2-7 自注意力与多头自注意力结构图

如图 2-7(a)所示自注意力机制，首先输入通过线性层，生成三组向量 K、Q、V。K 和 Q 首先做矩阵相乘，随后通过 SoftMax 激活函数，得到注意力图。最后与 V 做矩阵相乘。图 2-7(b)展示了多头自注意力机制，其与自注意力机制的区别是在三组向量 K、Q、V 将会通过线性层并且被拆分为多组，随后再进行与自注意力相同的矩阵运算得到多组输出，最后合并这多组输出通过一个线性层混洗特征得到输出。

接下来，介绍 Transform 第二块中的前馈神经网络结构，如图 2-8。其实，前馈神经网络也就是多层感知机，主要依靠线性层对特征的通道进行混洗。

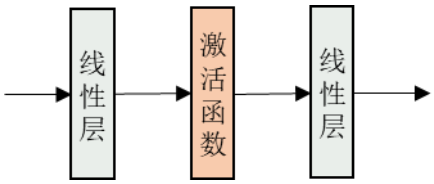


图 2-8 前馈神经网络结构图

如图 2-8 输入特征经过线性层、激活函数和线性层之后得到输出。

2.3 图像超分辨率重建的常用数据集

公开可用的数据集对于评估不同算法的性能起着核心作用，并为基于深度学习的技术提供了标准化的训练材料。在构建训练集时，应尽量增加图像样本的数量，并确保它们能够代表多样的场景类别，以此提高模型的泛化能力。而测试集则需要精心挑选，以便全面检验算法在各种环境下的适应性和准确性，确保覆盖广泛的场景变化。针对图像超分辨率重建这一特定任务，数据集中的图像应当具有较高的原始质量，并且保留丰富的细节信息，这对于提升重建后图像的质量尤为关键。随着超分辨率重建领域的不断发展，研究人员收集了越来越多的图像数据集用于图像超分辨重建的研究，这些数据集包含不同数量、不同分辨率的图像，涵盖不同种类的场景。

常用的图像超分辨重建测试集主要包括 Set5[92]、Set14[93]、B100[94]、Urban100[95]、Manga109[96]。训练集主要包括 DIV2K[97]、Flickr2K[98]、DIV8K[99]、LSDIR[100]以及 ImageNet[101]各数据的特点如表 2-1 所示。

表 2-1 数据集特点表

数据集	图像数量	平均图像尺寸	场景种类
Set5	5	31 3× 336	人、鸟、蝴蝶
Set14	14	492 × 446	人、动物、昆虫等
B100	100	320 × 480	人、动物、职务等
Urban100	100	984 × 797	城市、建筑、道路等
Manga109	109	435 × 381	漫画
DIV2K	1000	1972 × 1437	人、动物、建筑等
Flickr2K	2650	1782 × 1698	人、动物、建筑等
DIV8K	1500	7680 × 4320	人、动物、建筑等
LSDIR	84991	1782 × 1698	人、动物、建筑等
ImageNet	1281167	256 × 256	人、动物、建筑等

图像处理早期的两个数据集 Set5 和 Set14 的平均图像分辨率分别为 313 × 336 和 492 × 446。研究人员考虑到 Set5 和 Set14 数据集图像数量较小，且覆盖场景不足，不能很好地评测算法在复杂场景下的性能。于是研究人员收集了数据量更大的 B100 数据集，用于图像超分辨率重建的评测。其中 B100 数据分别包含 100 张图像，涵盖的场景也更加丰富。虽然 B100 数据集提供了更多的测试数据，但这个数据集中的图像分辨率较小，图像清晰度以及图像的纹理细节质量仍较低。为了更好地评估图像超分辨率重建算法的性能，研究人员又收集了包含 100 张城市场景高分辨率图像的 Urban100 数据集。该数据集图像尺寸较大且细节十分丰富，是图像超分辨率重建领域难度较大的一个数据集。由于 Urban100 数据集只包含城市场景，为了评测算法在不同场景下的性能，研究人员还收集了漫画风格的 Manga109 数据集。随着深度学习在图像超分辨率重建领域的应用，数据集除了提供测试数据外，还需要为神经网络提供训练数据。图像超分辨率重建主要使用的训练数据集为 DIV2K、Flickr2K、DIV8K 以及 LSDIR。DIV2K 数据集是首个转为图像超分辨率重建收集的数据集，其中共包含 1000 张图像，其中 800 张作为训练集，100 张作为验证集，100 张作为测试集。为进一步提升神经网络的性能，研究人员制作了一个更大的数据集，即 Flickr2K 数据集包含 2650 张高分辨率图像，涵盖多种多样的场景。在 DIV2K 与 Flickr2K 的基础上制作了

分辨率更高、图像数量更多的数据集分别为 DIV8K 和 LSDIR。ImageNet 是一个一个目标分类的数据集，其图像数量多，但是分辨率较低。部分算法用该数据集作为预训练数据集。

2.4 图像超分辨率重建的评价指标

为了评估重建图像的质量，需要采用一系列的评价指标。客观评价则是通过具体的数学模型和公式对重建图像的某些特性进行量化评估。在图像超分辨率重建领域中，最常用的客观评价指标是峰值信噪比（peak signal-to-noise ratio, PSNR）和结构相似性（structure similarity index measure, SSIM）。

峰值信噪比表示信号最大可能功率和影响它表示精度的破坏性噪声功率的比值，常用对数分贝(dB)单位来表示。在图像超分辨率重建领域，通常将真值图像的最大量化值作为信号最大可能功率，重建结果与参考图像间的差异作为噪声，计算峰值信噪比结果。峰值信噪比越高，则表示重建结果与参考图像越接近。一般来说，对于灰度图像，峰值信噪比在该通道内计算；对于 RGB 彩色图像，通常将图像投影到 YCbCr 空间并只在 Y 通道计算峰值信噪比；对于高光谱图像，通常在所有通道内一起计算峰值信噪比。PSNR 通过测量重建图像 $\hat{I}$ 与高分辨率图像 I 之间的均方误差(mean-square error, MSE)来评估图像质量。如公式(2-14)和公式(2-15)所示。

$$\mathrm{MSE}=\frac{1}{N}\sum_{i\in N}[I(i)-\hat{I}(i)]^2 \tag{2-14}$$

$$\mathrm{PSNR}=10\cdot\log_{10}(\frac{\mathrm{MAX}_I^2}{\mathrm{MSE}}) \tag{2-15}$$

式中，图像中第 i 个像素由 $I(i)$来表示，N 为图像 I 的总像素数。MAX 值图像的最大量化值，比如对于 8 比特量化的 RGB 图像来说，MAX = 255。

结构相似性通过比较两幅图像的结构相似性来评估重建效果。其公式如公式(2-16)所示。

$$SSIM(I,\hat{I})=\left[C_l(I,\hat{I})\right]^{\alpha}\left[C_c(I,\hat{I})\right]^{\beta}\left[C_s(I,\hat{I})\right]^{\gamma} \tag{2-16}$$

式中，$C_l(I,\hat{I})$、$C_c(I,\hat{I})$、$C_s(I,\hat{I})$ 分别表示亮度、对比度以及结构上的相似性。

α 、 β 、 γ 为超参数。$C_l(I,\hat{I})$、$C_c(I,\hat{I})$、$C_s(I,\hat{I})$的公式如公式(2-17)~公式(2-19)所示。

$$C_l(I,\hat{I})=\frac{2\mu_I\mu_{\hat{I}}+C_1}{\mu_I^2+\mu_{\hat{I}}^2+C_1} \tag{2-17}$$

$$C_c(I,\hat{I})=\frac{2\sigma_I\sigma_{\hat{I}}+C_2}{\sigma_I^2+\sigma_{\hat{I}}^2+C_2} \tag{2-18}$$

$$C_s(I,\hat{I})=\frac{\sigma_{I\hat{I}}+C_3}{\sigma_I\sigma_{\hat{I}}+C_3} \tag{2-19}$$

μ_I 和 $\mu_{\hat{I}}$ 分别表示真值图像和超分辨结果的均值，σ_I 和 $\sigma_{\hat{I}}$ 分别 表示真值图像和超分辨结果的标准差，$\sigma_{I\hat{I}}$ 表示真值图像和超分辨结果的协方差，C_1、C_2 和 C_3 为三个较小的常数参数以避免分母为零。μ，σ，$\sigma_{I\hat{I}}$ 公式由公式(2-20)~公式(2-22)表示。

$$\mu=\frac{1}{N}\sum_{i\in N}I(i) \tag{2-20}$$

$$\sigma=\sqrt{\frac{1}{N}\sum_{i\in N}\left(I(i)-\mu_I\right)^2} \tag{2-21}$$

$$\sigma_{I\hat{I}}=\frac{1}{N}\sum_{i\in N}\left(I(i)-\mu_I\right)\left(\hat{I}(i)-\mu_{\hat{I}}\right) \tag{2-22}$$

通常情况下，将参数设置为 $\alpha=\beta=\gamma=1$，$C_3=C_2/2$，此时上式可以进一步简化为公式(2-23)。

$$SSIM(I,\hat{I})=\frac{2\mu_I\mu_{\hat{I}}+C_1}{\mu_I^2+\mu_{\hat{I}}^2+C_1}\cdot\frac{2\sigma_{I\hat{I}}+C_2}{\sigma_I^2+\sigma_{\hat{I}}^2+C_2} \tag{2-23}$$

除了 PSNR 和 SSIM，还有一些其他的客观评价指标，如边缘信息保真度（edge information fidelity，EIF）、梯度保真度（gradient similarity measure，GSM）等。这些指标从不同的角度对重建图像的质量进行评估，本文图像评估中不使用这些指标故在此不顾多叙述。

2.5 本章小节

本章首先介绍了超分辨率重建问题的定义，介绍了图像降质模型，与图像重建模型。在图像重建模型中分析网络对于算法的影响，由此证明网络设计的重要性。并具体的介绍了图像重建模型中常用的两种网络：CNN 与 Transform 架构。随后，介绍超分辨重建算法中针对自然图像常用的训练集与测试集。最后，具体介绍图像超分辨率重建中最常用的两个评价指标 PSNR 与 SSIM。

第3章 基于高阶门控交互的单图像超分辨率重建算法

3.1 引言

本章提出一种新的基于 Transform 架构的轻量化图像超分辨率重建的网络模型。针对 Transform 中的多头自注意力机制(multi-head slef-attention, MSA)或基于滑动窗口的多头自注意力机制(window-based multi-head slef-attention, W-MSA)运算复杂度高，花费时间长，内存消耗大等问题做出改进。首先 MSA 中的运算复杂度主要是因为 MSA 中复杂的矩阵运算。本章主要针对 MSA 进行改进尽量降低 Transform 架构中的参数量与计算复杂度，实现轻量化超分辨率重建方法。

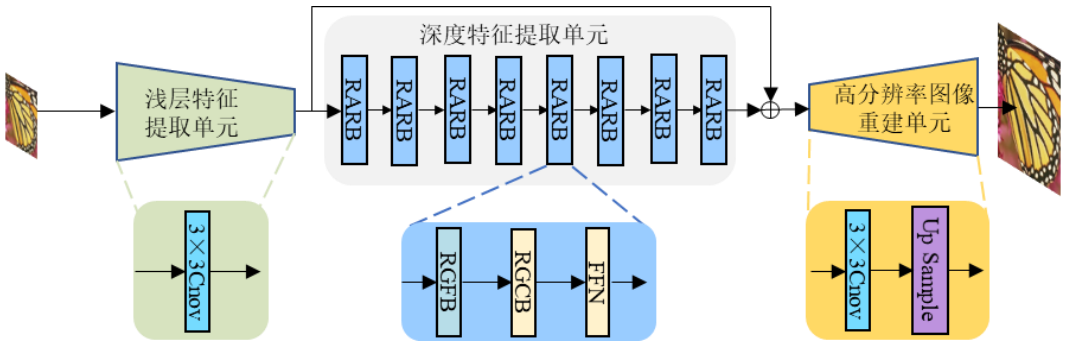
3.2 高阶门控网络

3.2.1 网络结构

高阶门控卷积网络(high-order gated convolutional networks, HGCN)由浅层特征提取单元(shallow feature extraction modules, SFM)、深度特征提取单元(deep feature extraction modules, DFM)、高分辨率图像重建单元组成(HR image reconstruction module, IRM)，其结构图如图 3-1 所示。浅层特征提取单元的结构图如图 3-1(a)所示。浅层特征提取单元通过卷积使输入映射到高维度空间生成深度特征，其过程可以由公式(3-1)表示。

$$F_{\mathrm{S}} = H_{\mathrm{SF}}(x) \tag{3-1}$$

式中，输入 $x \in \mathbb{R}^{H\times W\times 3}$，其中 H、W 分别为输入的宽和高，输入的通道数为 3。深度特征使用 $F_{\mathrm{S}} \in \mathbb{R}^{H\times W\times C}$ 表示，其中 C 为 F_{S} 的通道数。H_{SF} 是一个 3×3 卷积其输入维度为 3 输出维度为 C。



(a)SFM 结构图　　(b)RARB 结构图　　(c)IRM 结构图

图 3-1 高阶门控卷积网络结构图

随后深度特征将会在深度特征提取单元中通过残差全局自适应滤波块(residual global adaptive filtering block, RGFB)、递归门控卷积块(recursive gated convolution block, RGCB)和前馈神经网络(feedforward neural network, FNN)所组成的全局自适应滤波和高阶门控卷积块(residual global adaptive filtering and recursive gated convolution block, RARB)，进行一系列的特征交互，RARB 的结构图如图 3-1(b)。深度特征交互的过程可以描述为如公式(3-2)和公式(3-3)所示。

$$F_{\mathrm{D}} = H_{\mathrm{DF}}(x) \tag{3-2}$$

$$H_{\mathrm{DF}} = H_{\mathrm{FFN}}(H_{\mathrm{RGCB}}(H_{\mathrm{RGFB}}(\bullet))) \tag{3-3}$$

式中，H_{DF} 代表 DFM。F_{D} 是 DFM 的输出。H_{RGFB}、H_{RGCB}、H_{FFN} 分别代表 RGFB、RGCB、FFN。

最后通过一个残差链接聚合浅层特征与深层特征得到聚合特征，F_{UP} 输入 IRM 得到最后的高分辨图像即输出。重建过程能被描述为如公式(3-4)~公式(3-6)所示。IRM 的结构图如图 3-1(c)所示。

$$F_{\mathrm{UP}} = F_{\mathrm{S}} + F_{\mathrm{D}} \tag{3-4}$$

$$I_{\mathrm{SR}} = H_{\mathrm{IR}}(F_{\mathrm{UP}}) \tag{3-5}$$

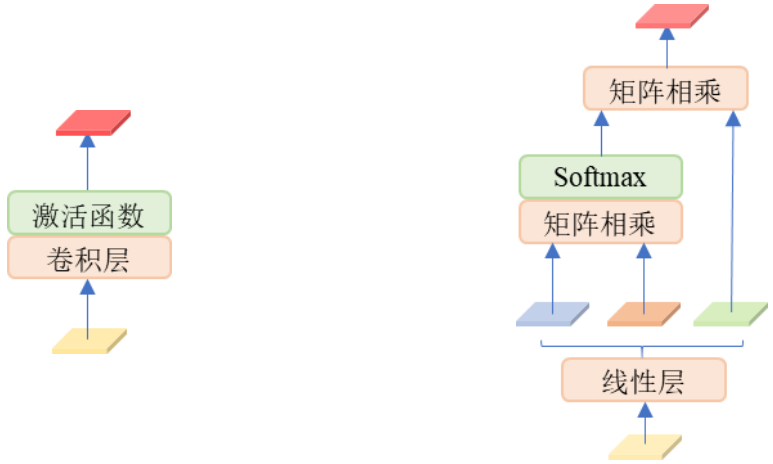
式中，F_{UP} 代表聚合特征。输出使用 $I_{\mathrm{SR}} \in \mathbb{R}^{s\cdot H\times s\cdot W\times 3}$ 代表，其中 s 是上采样倍数。H_{IR} 代表 IRM 由一个 3×3 的卷积(convolution)与像素混洗上采样(pixel-shuffle up-sampling)组成，具体可由公式(3-6)所示。

$$H_{\mathrm{IR}} = H_{\mathrm{UP}}(H_{\mathrm{Conv}}(\bullet)) \tag{3-6}$$

式中，3×3 的卷积与像素混洗上采样分别用 H_{Conv} 和 H_{UP} 代表。

3.2.2 递归门控卷积模块

在章节 2.2.3 中本文介绍了 Transform 网络中的多头自注意力机制(muti-head self-attention, MSA)。MSA 的计算复杂度主要是因为其中多次的矩阵乘法运算，为简化这个过程本文提出利用哈达玛积实现自注意力机制，哈达玛积的计算复杂度为$\mathcal{O}(n)$，而 MSA 中矩阵运算的计算复杂度为$\mathcal{O}(n^2)$。所以利用哈达玛积代替 MSA 中的矩阵运算能有极大的减少计算复杂度。由此提出递归门控卷积模块。递归门控卷积模块是极其轻量，灵活，高效的。首先本文将更进一步地分析 Transform 中的 MSA，首先 MSA 与普通的卷积进行比较如图 3-6。



(a) 卷积示意图　　(b) MSA 示意图

图 3-2 卷积与 MSA 对比图

图 3-6 中输入特征（黄色）输出特征（红色），显然卷积没有更全面的处理特征。MSA 采用两次矩阵相乘进行显式且全面的特征空间交互。进一步地说，MSA 具有对特征的全局建模能力，这导致 Transform 架构实现的效果远远优于 CNN，但是 MSA 的高复杂度使得网络难以部署在边缘设备上，而且难以构建轻量化网络。所以需要一个更轻量高效的方式实现显式且全面的特征空间交互，故此提出递归门控卷积。递归门控卷积设计示意图如下图 3-7。

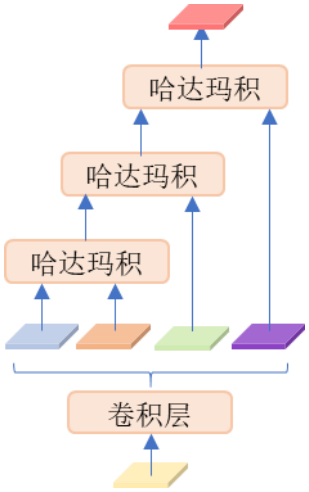


图 3-3 递归门控卷积设计示意图

在图 3-7 中首先利用卷积层生成不同层面的特征，随后这些特征利用哈达玛积进行空间交互。较于 MSA，递归门控卷积能更轻松的实现显示全面的特征交互，而且特征交互的阶数是可以拓展到任意阶。接下来，本文会详细叙述递归门控卷积。首先，卷积层使用蓝图卷积，其由点卷积（1×1 卷积）与深度可分离卷积(depth-wise convolution, DWConv)组成，其结构图如图 3-8 所示。

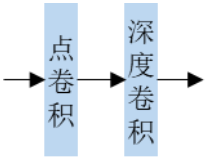


图 3-4 蓝图卷积结构图

在章节 2.2.2 中已有介绍深度可分离卷积，相较于普通卷积其参数量远低于普通卷积，所以能使递归门控卷积做到轻量。在递归门控卷积中一部分特征为门控特征与交互特征进行交互，划分这种特征的方式有很多种，例如：在 MSA 中直接将输入维度放大三倍分别生成 K、Q、V 实现高阶特征交互。在递归门控卷积中采用递归的方式划分门控特征与交互特征。递归门控卷积结构图如图 3-9 所示。

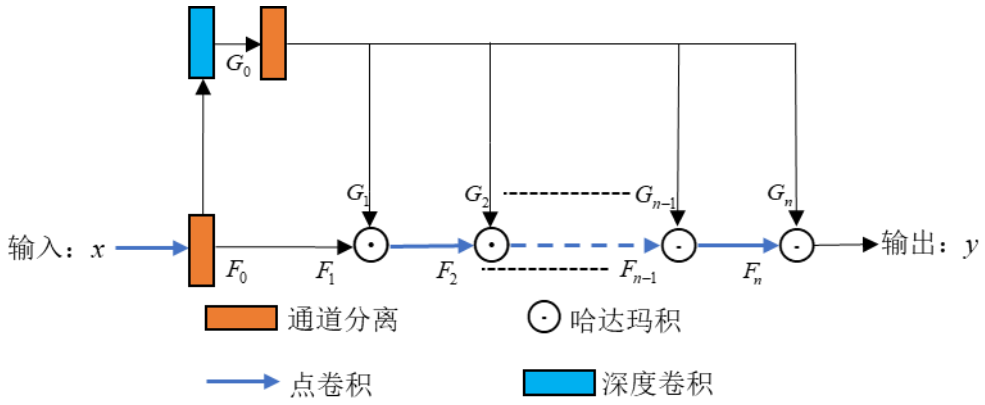


图 3-5 递归门控卷积结构图

为便于叙述，将输入设为 $x \in \mathbb{R}^{H\times W\times C}$。输入首先会通过一个点卷积，随后通过通道分离操作生成门控特征与交互特征，如公式(3-7)。

$$F_0, G_0 = H_S(Conv_1(x)) \tag{3-7}$$

式中，$F_0 \in \mathbb{R}^{H\times W\times C_F^0}$ 与 $G_0 \in \mathbb{R}^{H\times W\times C_G^0}$ 分别为交互特征与门控特征，其中 $C_F^0 = C/2^{n-1}$、$C_G^0 = C[(2^{n-1}-1)/2^{n-1}]$。$H_S$ 为通道分离操作，$Conv_1$ 代表点卷积，使输入的维度 C 上升到 $2C$。

接下来，门控特征利用深度卷积学习特征中的信息，随后通过一个超参数 α 缩放特征避免反向传播的过程中梯度爆炸，如公式(3-8)所示。

$$G_g = \{G_1, G_2, \cdots\cdots G_{n-1}, G_n\} = H_S(\alpha \odot H_{DW}(G_0)) \tag{3-8}$$

式中，H_{DW} 代表深度可分离卷积，G_g 是一组特征 $\{G_1, G_2, \cdots\cdots G_{n-1}, G_n\}$ 通过通道分离操作生成，其中第 q 组特征 $G_q \in \mathbb{R}^{H\times W\times C_G^q}$，$C_G^q = C/2^{n-q}$，$1 \le q \le n-1$。

为对齐门控特征与交互特征的编号，设 $F_0 = F_1$。门控特征与交互特征通过哈达玛积交互的过程如公式(3-9)所示。

$$F_k = Conv_2(F_{k-1} \odot G_{k-1}) \tag{3-9}$$

式中，G_{k-1} 是 G_g 中第 $k-1$ 组特征。F_{k-1} 是第 $k-1$ 次特征交互的输出。G_{k-1} 和 F_{k-1} 进行哈达玛积之后通过一个点卷积 $Conv_2$ 混洗特征得到输出 F_k，即第 k 次空间交互后的输出，特别的 $F_n = y$，y 为递归门控卷积的输出。

最后，深度卷积的卷积核大小也是需要考虑的问题，在大部分 CNN 中的卷积核大小为 3 × 3。在 Transform 中的全局建模能力是由于 K、Q、V 之间的全局交互，可以理解为其拥有全局感受野。所以，对于 CNN 传统 3 × 3 的卷积，其感受野要小很多，于是大核卷积被用于许多高级视觉任务中，并取得了令人印象深

刻的性能，值得注意的是，对于在高层视觉中 ConvNext[102]和 Swin Transformer[103]中，默认的窗口或内核大小是 7 × 7，在底层视觉中，例如 ShuffleMixer[105]等网络也表现出优秀的效果。因此，因此本文在设置深度卷积的卷积核大小也使用大卷积核，并将其设置为 7 × 7。

3.2.3 残差全局自适应滤波块

在图像降质的过程中主要丢失的是图像的高频信息，所以在超分重建过程中恢复高频信息卓为重要。从空域来说，图像的高频信息是分散的、稀疏的，很难做到针对性处理。从网络的角度来说，对于 CNN 由于卷积的归纳偏置（局部性和平移不变性）使得高频语义信息，例如：图像边缘信息，图像细节部分难以得到重建。对于 Transform 虽然拥有 MSA 得以处理图像的全局特征，但是图像的高频信息在图像中是分散的、稀疏的难以得到较高的权重。综上设计一种针对高频信息的处理方式解决空域中高频信息的分散，解决 CNN 中的归纳偏置，解决 Transform 中的高频权重不适定问题尤为重要。而且这种算法需要做到简单且轻量，方便部署在网络中。

大部分的网络主要从像素域处理深度特征，而在像素域很难提取特征的高频信息。反观频域，在没有进行频谱搬移操作时，高频信息主要居中与频谱的中心位置如图 3-2。

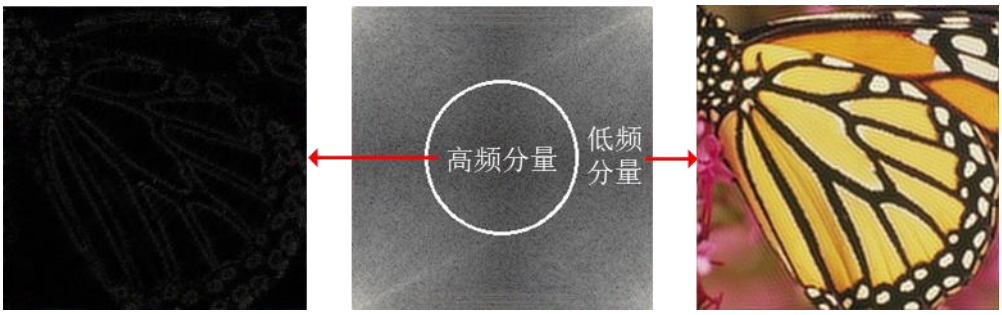


图 3-6 频谱示意图

由此在频域中提取特征的高频信息是极其容易的，在白色圆圈中的区域为高频分量，圆圈之外的为低频分量。明显的，在频域中提取图像的高频细节直接且高效。分别提取这些分量，通过逆傅里叶变换转为空域后，高频分量明显分布不均匀，较难以提取，低频分量明显较为模糊因为高频细节的丢失。

为更好的介绍残差全局自适应滤波块，我们首先介绍离散傅里叶变换

(discrete Fourier transform, DFT)在一维情况下其公式如公式(3-10)所示。

$$X[k]=\sum_{n=0}^{N-1}x[n]e^{-j(2\pi/N)kn}=\sum_{n=0}^{N-1}x[n]W_N^{kn} \tag{3-10}$$

式中，j 是虚单位，$W_N=\mathrm{e}^{-j(2\pi/N)}$。公式(3-10)中的 DFT 公式可以通过时域和频域中的采样从连续信号的傅里叶变换导出。由于 $X[k]$在长度为 N 的区间上重复，因此在 N 个连续点 $k=0, 1, \ldots\ldots, N-1$ 处取 $X[k]$的值就足够了。具体来说，$X[k]$表示频率 $\omega_k=2\pi k/N$ 处的序列 $x[n]$的频谱。

逆离散傅里叶变换(inverse discrete Fourier transform, IDFT)的定义由公式(3-11)给出：

$$x[n]=\frac{1}{N}\sum_{k=0}^{N-1}X[k]e^{j(2\pi/N)kn} \tag{3-11}$$

对于离散实信号 $x[n]$，它的 DFT 是共轭对称的，即 $X[N-k]=X^*[k]$。反之亦然，如果我们对共轭对称的 $X[k]$执行 IDFT，则可以恢复离散实信号 $x[n]$。根据上述的 DFT 的共轭对称性，显然可以得到在执行 DFT 之后的离散实信号 $x[n]$的频谱 $X[k]$仅需要一半的频谱信息即可完整的保存 $x[n]$所含有的信息，即{X[k] : $0\le k\le\lceil N/2\rceil$}，其中$\lceil\cdot\rceil$为上取整。

DFT 在现代信号处理算法中得到了广泛的应用其主要原因有：首先 DFT 的输入和输出都是离散的，因此可以很容易地由计算机处理，其次存在计算 DFT 的有效算法，即快速傅里叶变换(fast Fourier transform, FFT)对应的也存在逆快速傅性特性，并将计算 DFT 的复杂度从$\mathcal{O}(N^2)$降低到$\mathcal{O}(N\log N)$，极大的提高了 DFT 运算的速度与效率。

接下来将 DFT 扩展到二维，其公式如公式(3-12)所示。

$$X[u,v]=\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}x[m,n]e^{-j2\pi\left(\frac{um}{M}+\frac{vn}{N}\right)} \tag{3-12}$$

式中，输入信号为 $x[m, n], 0 \leqslant m \leqslant M-1, 0 \leqslant n \leqslant N-1$。输入信号经过二维 DFT 后得到的频谱为 $X[u, v]$。M, N 为图像的长度和宽度。在二维 DFT 同样有其二维 IDFT，同时也具有上述的共轭对称性，及其二维 FFT 和二维 IFFT 在此不多赘述。

输入信号通过 FFT 映射到频域得到一组频域特征，在频域中通过一个由外

部输入的自适应算子赋予高频即频谱中心区域较高的权重，随后通过 IFFT 得到最后的输出。我们设输入实离散信号为$x\in\mathbb{R}^{H\times W\times C}$。RGFB 的公式可以简写为公式(3-13)如下所示。

$$y=\mathcal{F}^{-1}\left(\mathcal{H}\left(\mathcal{F}\left(x\right)\right)\right)+x \tag{3-13}$$

式中，$\mathcal{F}$和$\mathcal{F}^{-1}$分别表示 FFT 与 IFFT. $\mathcal{H}$代表一个由外部输入的自适应频域滤波算子。y为最后的输出信号。

在构建 RGFB 的过程中，自适应算子$\mathcal{H}$是要最重要的，接下来叙述自适应算子$\mathcal{H}$的构建方式。首先借助先验知识在未搬移的频谱中，其中心位置代表图像的高频信息所以在网络权重初始化的过程中其中心位置的权重应该较高，周围区域的权重应该较低。其次算子的大小也需要被确定，根据频域的共轭对称性，将算子的大小设置为$H\times\lceil W/2\rceil$。H, W为训练设置中输入的特征尺寸。综上，本文对自适应算子在权重初始化之后进行可视化如图 3-3 所示。

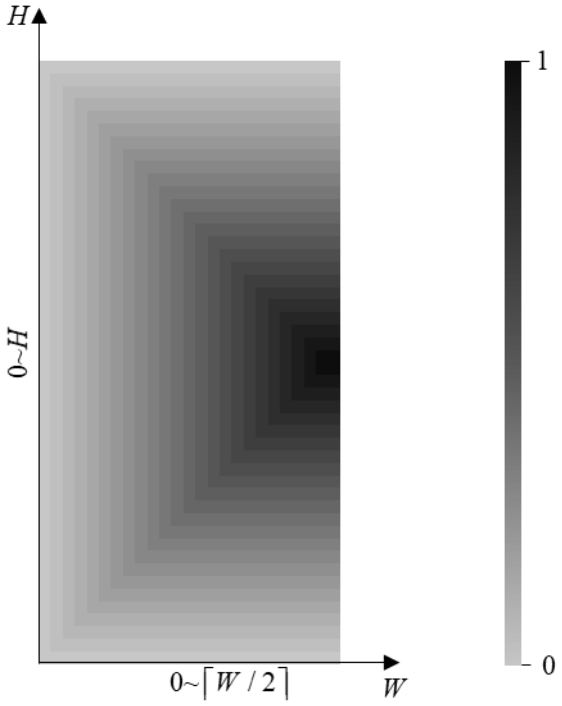


图 3-7 自适应算子初始化权重可视化图

在频域中如何让$\mathcal{H}$与特征图进行交互，从注意力的角度来说，自适应算子相当于高频注意力机制，使得网络的着重处理特征高频的部分。大部分的注意力机制采用哈达玛积，即对应元素相乘，例如空间注意力。但是空间注意力对于所有维度仅用一张注意力图与特征进行交互。对于高频部分来说，本文从小波变换的

角度分析，小波变换中仅将高频部分分为三种：水平、竖直、对角线分量。本文对比与小波变换将对角线分量分解为左上到右下与右上到左下的高频分量，综上，将高频分量分解为高频分为水平高频分量，竖直高频分量，左上到右下的对角线高频分量和右上到坐下的对角线高频分量，共四组滤波器，即四张注意力图与特征进行交互，充分的提取特征中的高频部分。本文将这种注意力称为分组的频域空间注意力机制。这种新的注意力机制，较于普通空间注意力机制能更好的实现对注意力图与特征之间的交互，较于像素注意力机制参数更少，但是将特征分为四组容易带来固定的归纳偏置。这种归纳偏置与 CNN 中卷积的平移不变性类似，本文将其称为滤波变性。具体地来说，再分为四组的滤波器中，按照先验知识提取固定的高频分量，但是特征中这种高频分量是否与高分辨率图像对应，甚至是否存在是不可知的。所以需要对特征进一步的处理，使得所有特征中均有这些高频分量再由自适应滤波提取对应的高频分量。故，本文加入层归一化(layer norm, LN)在滤波之前对特征进行处理。

对于训练来说可以将训练图像大小裁剪成一致的大小，但是测试时对于网络的输入大小将会是任意的，所以如何让算子拥有对任意输入大小有自适应性是接下来要考虑的问题。为了简单快速的与输入图像的尺寸进行对齐，本文采用双三次采样的方式，对自适应权重径向上或下采样。双三次采样具有一定的采样精度。、且速度较快符合本文所需求的目标。

残差全局自适应滤波块由 LN 与全局自适应滤波器(global adaptive filter, GF)以及一个残差链接组成。其中加入 LN 层是为了起到对特征之间的交互，已消除后续分为四组对应滤波带来的归纳偏置，而且能更加稳定地训练网络。残差全局自适应滤波块由图 3-4 所示。

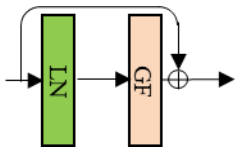


图 3-8 残差全局自适应滤波块结构示意图

残差全局自适应滤波块中 GF 的结构图如下图 3-5 所示。

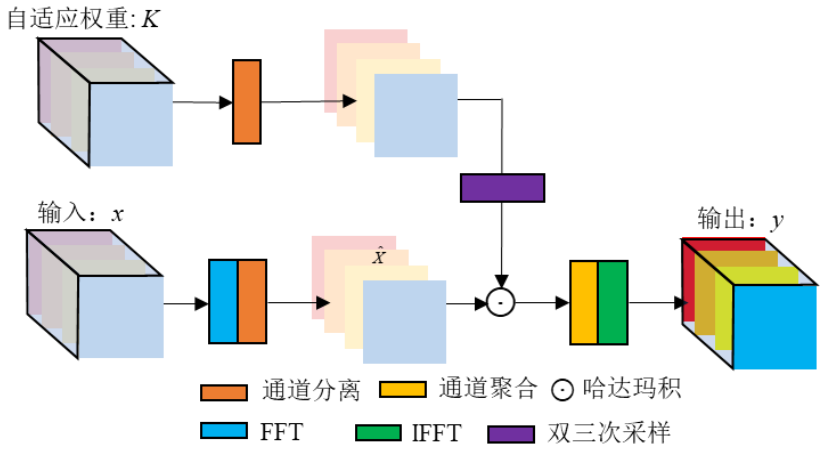


图 3-9 GF 结构示意图

首先，我们设输入 $x \in \mathbb{R}^{H\times W\times C}$。对齐进行 FFT 后得到 $\hat{X} \in \mathbb{C}^{H\times (W/2)\times C}$，这个过程如公式(3-14)所示。

$$\hat{X} = \mathcal{F}(x) \tag{3-14}$$

随后引入外部输入自适应滤波器的权重 $K \in \mathbb{C}^{H\times (W/2)\times n}$，其中 $n \in \mathbb{Z}$，$1 \le n \le C$ 且能整除 C。在上文中经过分析 $n = 4$。本文会在实验章节对 n 进行消融实验并对其进行分析。K 首先会进行双三次采样与特征尺寸进行对齐，随后通过通道分离与特征进行交互，如公式(3-15)所示

$$\tilde{X} = H_{\mathrm{S}}(H_{\mathrm{B}}(K)) \odot H_{\mathrm{S}}(\hat{X}) \tag{3-15}$$

式中，$\tilde{X}$ 为自适应滤波器权重与特征交互之后的输出。H_{B} 代表双三次采样。H_{S} 代表通道分离操作。$\odot$ 代表哈达玛积。

最后交互后的特征通过通道聚合与 IFFT 得到输出 y，如公式(3-16)所示。

$$y = \mathcal{F}^{-1}(H_{\mathrm{C}}(\tilde{X})) \tag{3-16}$$

式中，H_{C} 代表通道聚合操作。

3.3 实验结果与分析

3.3.1 实验设置

（1）数据集和指标

参照文献[46]，训练图像由来自 Flickr2K 的 2650 张图像和来自 DIV2K 的 800 张图像组成。为了评估不同 SR 网络的性能，本文使用 Set5、Set14、B100、Urban100

和 Manga109 这 5 个标准基准数据集。平均峰值信噪比(PSNR)和 Y 通道(即亮度)上的结构相似性(SSIM)被用作评估指标。在计算复杂度与内存消耗上选取，参数量与 FLOPs 与其他先进网络进行对比。

（2）降质模型

使用双三次下采样的方式对图像进行下采样倍数为 2、3、4 的降质。如章节 2.2.1 中，图像降质过程可以由如公式(3-17)表示。

$$I_{\mathrm{LR}}=I_{\mathrm{HR}}\downarrow_{s} \tag{3-17}$$

式中，I_{HR} 代表高分辨率图像，I_{LR} 低分辨率图像。$\downarrow_{s}$ 代表 s 倍的双三次下采样。

（3）本章算法的实现细节

本章算法由 8 个全局自适应滤波和高阶门控卷积块组成，通道数设置为 64。所有深度卷积的核大小设置为 7。采用随机旋转 90°、180°、270°和水平翻转的数据增强方法。批量大小设置为 32，每个 LR 输入的 patch 大小设置为 48×48。模型由 Adam[106]优化器进行训练。初始学习率设置为 1×10^{-3}，迭代 2.5×10^{5} 后使用余弦退火学习率衰减方法衰减到 1×10^{-7}。L1 损失用于优化总 1×10^{6} 迭代的模型。L1 损失函数如公式(3-18)所示。

$$\mathcal{L}^{\ell_1}=\frac{1}{N}\sum_{i\in N}\left|I_{\mathrm{SR}}(i)-I_{\mathrm{HR}}(i)\right| \tag{3-18}$$

式中，I_{SR} 为超分辨率重建后的图像，I_{HR} 为高分辨图像。$I_{*}(i)$表示图像 I_{*}的第 i 个像素，其中*代表 HR 或 SR。N 为图像的像素个数。

（4）实验环境

GPU 是 GeForce RTX 3060 GPU 上，使用 Pytorch 实现本章算法。CPU 是第 12 代的英特尔酷睿 i5-12490F。代码和数据集保存在固态磁盘 DS42HKVS-78BFKYO 上。操作系统为 Windows10。

3.3.2 实验结果

在本节将本章算法与已有的轻量化超分辨重建算法(DRCN[44], DRRN[50], IDN[45], CARN[104], IMDN[46], RFDN[47], DRSAN[59], LBNet[68], ShuffleMixer[105])进行对比。其中 RFDN 是使用特征蒸馏实现特征的长远距离交互，是具有代表性的 CNN 算法。DRSAN 使用多头自注意力增强残差链接，并使用 CNN 构建主体网

络。LBNet 利用 CNN 与 Transform 相结合构建网络，是具有代表性的 CNN-Transform 算法。ShuffleMixer 使用更大的卷积核实现网络，是具有代表性的大核卷积算法。由于 LBNet 仅提供在下采样因子为 3 和 4 的测试结果，没有在上采样倍数为 2 的尺度上进行测试。所以本章算法在上采样倍数为 3 和 4 的尺度上与 LBNet 进行对比。具体的对比结果如表 3-1 所示。其中 FLOPs 在高分辨率图像为 1280 × 720 测得，即上采样倍数为 2、3、4 所对应的低分辨率图像输入大小为 640 × 360、430 × 240、320 × 180。表中最优结果用红色标注、次优结果用蓝色标出。

表 3-1 本章算法与先进算法结果对比表

方法	上采样倍数	参数量	FLOPs	set5	Set14	B100	Urban100	Manga109
				PSNR/SSIMM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Bicubic	×2	-		33.66/0.9299	30.24/0.8688	29.56/0.8431	26.88/0.8403	30.80/0.9339
DRCN[44]		1,774K	17,974G	37.63/0.9588	33.04/0.9118	31.85/0.8942	30.75/0.9133	37.55/0.9732
DRRN[50]		298K	6,796G	37.74/0.9591	33.23/0.9136	32.05/0.8973	31.23/0.9188	37.88/0.9749
IDN[45]		553K	124.6G	37.83/0.9600	33.30/0.9148	32.08/0.8985	31.27/0.9196	38.01/0.9749
CARN[104]		1,592K	222.8 G	37.76/0.9590	33.52/0.9166	32.09/0.8978	31.92/0.9256	38.36/0.9765
IMDN[46]		694K	161G	38.00/0.9605	33.63/0.9177	32.19/0.8996	32.17/0.9283	38.88/0.9774
RFDN[47]		534K	95G	38.05/0.9606	33.68/0.9184	32.16/0.8994	32.12/0.9278	38.88/0.9773
DRSAN[59]		419K	85.5G	37.99/0.9606	33.57/0.9177	32.16/0.8999	32.11/0.9279	-/-
ShuffleMixer[105]		394K	91G	38.01/0.9606	33.63/0.9180	32.17/0.8995	31.89/0.9257	38.83/0.9774
本章算法		454K	91.2G	37.98/0.9605	33.66/0.9181	32.19/0.9000	32.17/0.9285	38.92/0.9776
Bicubic	×3	-		30.39/0.8682	27.55/0.7742	27.21/0.7385	24.46/0.7349	26.95/0.8556
DRCN[44]		1,774K	17,974G	33.82/0.9226	29.76/0.8311	28.80/0.7963	27.15/0.8276	32.24/0.9343
DRRN[50]		298K	6,796G	34.03/0.9244	29.96/0.8349	28.95/0.8004	27.53/0.8378	32.71/0.9379
IDN[45]		553K	56.3G	34.11/0.9253	29.99/0.8354	28.95/0.8013	27.42/0.8359	32.71/0.9381
CARN[104]		1,592K	118.8G	34.29/0.9255	30.29/0.8407	29.06/0.8034	28.06/0.8493	33.50/0.9440
IMDN[46]		703K	71.5G	34.36/0.9270	30.32/0.8417	29.09/0.8046	28.17/0.8519	33.61/0.9445
RFDN[47]		541K	42.2G	34.41/0.9273	30.34/0.8420	29.09/0.8050	28.21/0.8525	33.67/0.9449
DRSAN[59]		419K	43.2G	34.41/0.9272	30.27/0.8413	29.08/0.8056	28.19/0.8529	-/-
LBNet[68]		736K	68.4G	34.47/0.9277	30.38/0.8417	29.13/0.8061	28.42/0.8559	33.82/0.9460
ShuffleMixer[105]		415K	43G	34.40/0.9272	30.37/0.8423	29.12/0.8051	28.08/0.8498	33.69/0.9448
本章算法		462K	47.1G	34.43/0.9271	30.40/0.8430	29.13/0.8059	28.19/0.8527	33.80/0.9456
Bicubic	×4	-		28.42/0.8104	26.00/0.7027	25.96/0.6675	23.14/0.6577	24.89/0.7866
DRCN[44]		1,774K	17,974G	31.53/0.8854	28.02/0.7670	27.23/0.7233	25.14/0.7510	28.93/0.8854
DRRN[50]		298K	6,796G	31.68/0.8888	28.21/0.7720	27.38/0.7284	25.44/0.7638	29.45/0.8946
IDN[45]		553K	32.3G	31.82/0.8903	28.25/0.7730	27.41/0.7297	25.41/0.7632	29.41/0.8942
CARN[104]		1.592K	91G	32.13/0.8937	28.60/0.7806	27.58/0.7349	26.07/0.7837	30.47/0.9084
IMDN[46]		715K	41G	32.21/0.8948	28.58/0.7811	27.56/0.7353	26.04/0.7838	30.45/0.9075
RFDN[47]		550K	23.9G	32.24/0.8952	28.61/0.7819	27.57/0.7360	26.11/0.7858	30.58/0.9089
DRSAN[59]		419K	30.5G	32.15/0.8935	28.54/0.7813	27.54/0.7364	26.06/0.7858	-/-
LBNet[68]		742K	38G	32.29/0.8960	28.68/0.7832	27.62/0.7382	26.27/0.7906	30.76/0.9111
ShuffleMixer[105]		411K	28G	32.21/0.8953	28.66/0.7827	27.61/0.7366	26.08/0.7835	30.65/0.9093
本章算法		474K	26.5G	32.36/0.8969	28.75/0.7847	27.65/0.7389	26.31/0.7928	30.83/0.9125

如表 3-1 所示，本文提出的算法与先进算法相比较在上采样倍数为 4 时，取得了最佳效果，但是在上采样倍数为 2 和 3 时，仅有部分取得最佳效果。接下来，本文将从参数量与客观评价上进行分析。

（1）参数量与计算量分析

本章算法的参数量与 FLOPs 分别为 474K 与 26.5G。对比于 CNN-Transform 类的算法，本章算法较于 LBNet 参数量下降了 268K，计算量 FLOPs 下降了 11.5G。这主要是因为递归门控卷积的高效与轻量化。对于 CNN 类算法，本章算法较于 RFDN 参数量下降了 76K，但是计算量 FLOPs 上升 2.6G。计算量的上升主要是因为引入了哈达玛积作为全局特征交互，较于一般卷积这种特征交互方式计算复杂度较高，但是远远小于 MSA 中的矩阵运算。综上，本章算法实现了 Transform 架构的轻量化，与大部分优秀算法相比均有优势。

（2）客观指标 PSNR 与 SSIM 分析

本章算法在上采样倍数为 4 时，较于 LBNet 算法在数据集 Urban100 的 PSNR 与 SSIM 至少提升 0.04dB、0.0022。Urban100 数据集主要是城市景观，富含大量的高频部分，残差全局自适应滤波模块充分的提取特征中的高频部分，并使得网络充分利用这些高频特征去重建图像中的高频信息，所以得到了较大的提升。较于 RFND 算法在 Manga109 数据集上 PSNR 与 SSIM 分别有 0.25dB、0.0036 的提升，这主要归功于 Transform 架构的优越性以及递归门控卷积带来的全局建模能力，其能更全面的关注特征实现更优的超分辨率重建效果。在上采样倍数为 2 和 3 时，本章算法仅取得了部分最佳结果，这主要是因为训练网络时选取的低分辨率图像块大小固定为 48 × 48，但是在测试时低分辨率图像输入大小随着上采样倍数的缩小而上升，在 3.2.2 章节中自适应滤波器的权重是由上三次上采样得到与输入特征相匹配的尺寸。当输入特征的尺寸远大于自适应滤波器权重的尺寸这会导致部分的低频信息丢失，使得超分辨率重建后的结果下降。综上，本章算法较于其他先进算法在客观指标结果上取得了不错的效果。

为进一步可视化本章算法与先进算法的比较，选取上采样倍数为 4 的 Urban100 数据集计算 PSNR，同 FLOPs 与参数量汇总在图中比较，如图 3-10 所示。

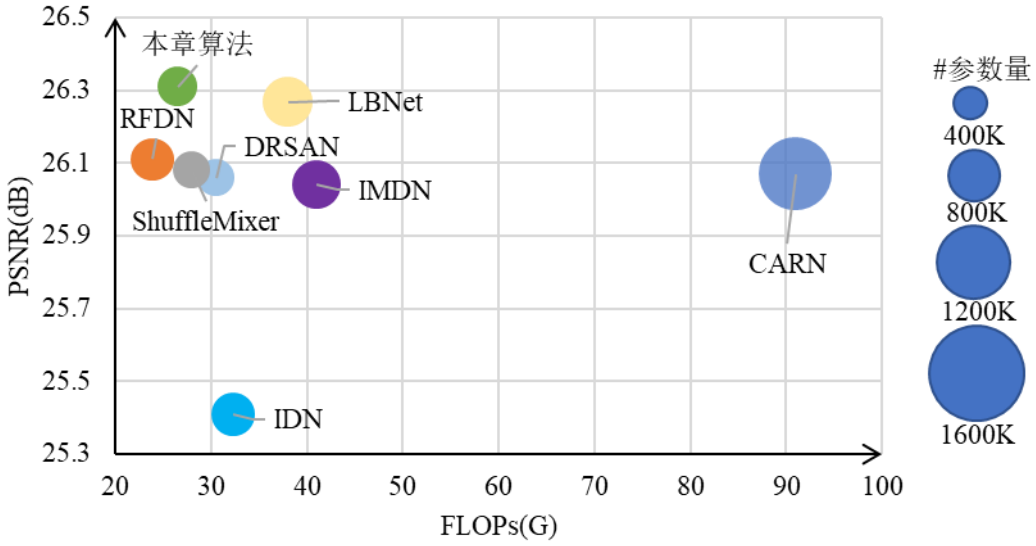


图 3-10 本章算法与先进算法比较可视化图

从图 3-10 中可以看出本章算法在参数量、计算量与客观结果实现了更优的平衡。无论是对比 CNN-Transform 算法例如 LBNet 或是 CNN 算法例如 RFDN。

为进一步的比较本章算法与其他先进算法，选取 Urban100 数据集中 img012、img024 与 img062 进行本章算法重建后的图像与其他先进算法重建后的图像的视觉效果进行对比如图 3-11。

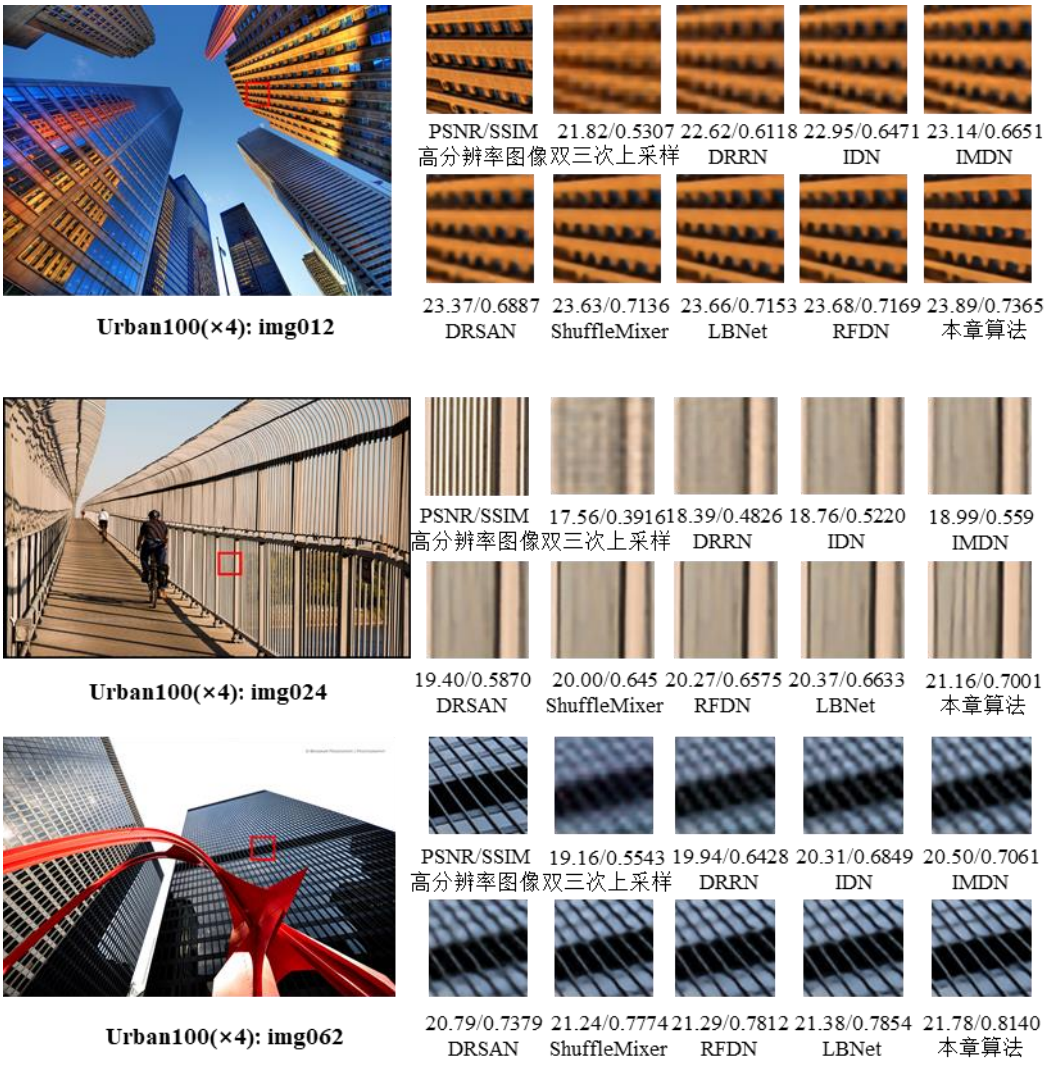


图 3-11 本章算法与其他先进算法视觉效果对比图

从图 3-11 中明显可以看出，本章算法在图像细节上恢复的质量更贴近与高分辨率图像，实现了最优的视觉效果。但是相较于高分辨率图像其图像质量还有略微差距，在网络的设计上要更多的考虑更有效的特征交互方式，以恢复更高的图像质量。

3.3.3 算法分析

在上节中本文针对结果简要的分析了本章算法，在本节中本文进一步的讨论算法中各个模块的作用，更详细的分析本章算法。首先，针对浅层特征提取模块进行实验，选取四种卷积通过不同的方式设计浅层提取模块分别为 1 × 1 卷积、3 × 3 卷积、7 × 7 蓝图卷积、两个 7 × 7 蓝图卷。在上采样倍数为 4 的数据集上重新训练网络，采用 Set5 数据集计算 PSNR 作为评价指标。具体如下表 3-2。

表 3-2 浅层特征提取模块消融结果表

方法	PSNR(dB)	参数量(K)
1 × 1 卷积	32.20	256
3 × 3 卷积	32.36	3392
7 × 7 蓝图卷积	32.26	1792
两个 7 × 7 蓝图卷	32.32	6784

从表中可以看出选择 3×3 卷积在 PSNR 与参数量能获得最优的效果。3×3 卷积是一种简单高效的方式提取图像中的特征。蓝图卷积虽然提升了像素感受野但是对于特征的通道交互较少很难提取更优的特征。

在 CNN 中一般卷积之后需要一个激活函数，但是在浅层特征提取模块中并没有使用激活函数，故此本文选用层归一化(LN)、RELU、LRELU、GELU 作为放置于卷积层之后，在上采样倍数为 4 的数据集上重新训练网络，采用 Set5 数据集计算 PSNR 作为评价指标。具体结果如下表 3-3。

表 3-3 浅层特征提取激活函数消融结果

方法	NO	LN	RELU	LRELU	GELU
PSNR(dB)	32.36	32.20	32.25	32.27	32.30

通过表 3-3 发现不使用激活函数达到最优效果，其原因主要是非线性对图像特征影响较大，在高层视觉中对图像进行映射或是编码的过程也不使用激活函数。故此，在构建网络时浅层特征提取主要使用 3×3 卷积且不加入激活函数等非线性操作。

残差全局自适应滤波块是本章算法重要的组成部分，其提取特征中的高频分量，增强超分辨率重建效果。模型 1 不使用残差全局自适应滤波块，模型 2 残差全局自适应滤波块不使用残差链接，模型 3 使用残差全局自适应滤波块并带有残差链接。在上采样倍数为 4 的数据集上重新训练网络，采用 Set5、Set14、B100、Urban100、Manga109 数据集计算 PSNR 作为评价指标。其结果如表 3-4 所示。

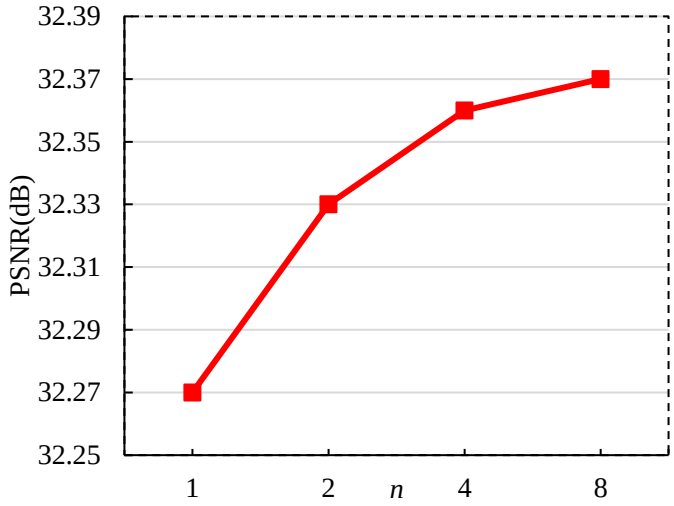
表 3-4 残差全局自适应滤波块消融结果表

数据集 / 方法	Set5	Ste14	B100	Urban100	Manga109
模型 1	32.27	28.70	27.63	26.20	30.75
模型 2	29.18	25.22	23.40	22.20	28.13
模型 3	32.36	28.75	27.65	26.31	30.83

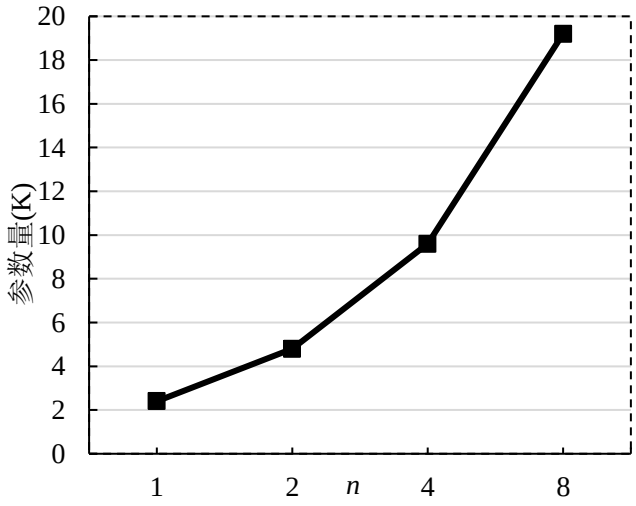
从表 3-4 中可以发现使用残差全局自适应滤波块对算法的超分辨率重建性能提升极高。例如在 Set5 与 Urban100 上分别提升 0.09dB 与 0.11dB。但是在使

用残差全局自适应滤波块不使用残差链接时，算法超分辨率重建性能大幅度下降。这主要是因为没有残差链接，在滤波之后的征低频分量大量丢失，导致最后重建的高分辨率图像质量大幅下降。故此，残差链接在残差全局自适应滤波块中起到了重要作用。

接下来，本文将分析在残差全局自适应滤波块中的滤波器分组的个数 n，本文将其分为 1、2、4、8 组。在上采样倍数为 4 的数据集上重新训练网络，采用 Set5 数据集计算 PSNR 作为评价指标。其消融结果如图 3-12 所示。



(a) PSNR 对比图



(b) 参数量对比图

图 3-12 残差全局自适应滤波块消融结果图

从图 3-12 中可以看出，在 $n=1$ 时残差全局自适应滤波块等同于空间注意力

机制。显然，这样无法提取较好的高频特征，所以继续提高分组数量。当 n = 4 时 PSNR 有显著提升参数量仅少量提升。当 n = 8 时 PSNR 仅有略微提升，但是参数量有较高的提升。所以最后残差全局自适应滤波块的滤波器组数，选择为 n = 4。

为进一步证明残差全局自适应滤波块的作用，本文对残差全局滤波块的输出特征进行可视化。首先提取出残差全局滤波块的输出特征，对齐进行归一化处理，对特征进行取均值操作，最后进行 Uint8 编码可视化特征图，其可视化结果如图 3-13 所示。

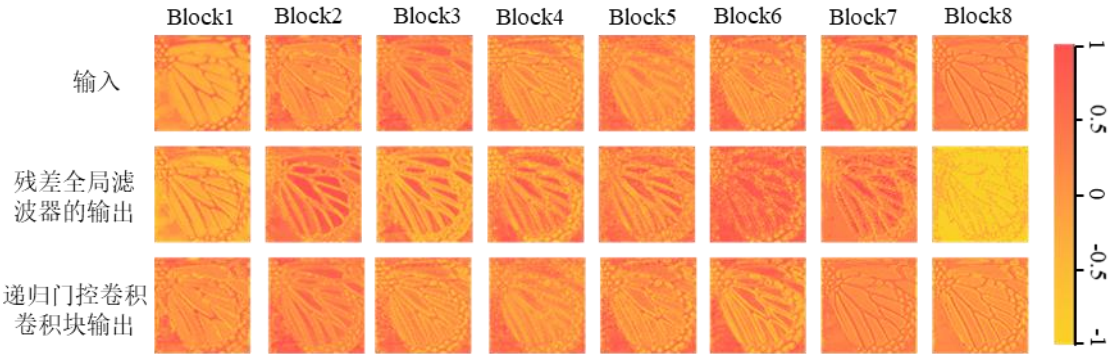


图 3-13 残差全局自适应滤波块输出特征可视化

图 3-13 中，Block i 代表第 i 个 RARB 块。从图 3-13 可以看出残差全局自适应滤波块充分的提取出了高频分量，提高高频分量的权重。使得网络更加关注特征中的高频部分，以提升网络的超分辨率重建效果。

递归门控卷积块也是算法的重要组成部分，首先对递归门控卷积块中的超参数 α 进行消融实验。选取 α 的值为 1、2、3、4，在上采样倍数为 4 的数据集上重新训练网络，采用 Set5 数据集计算 PSNR 作为评价指标。其消融结果如图 3-14 所示。

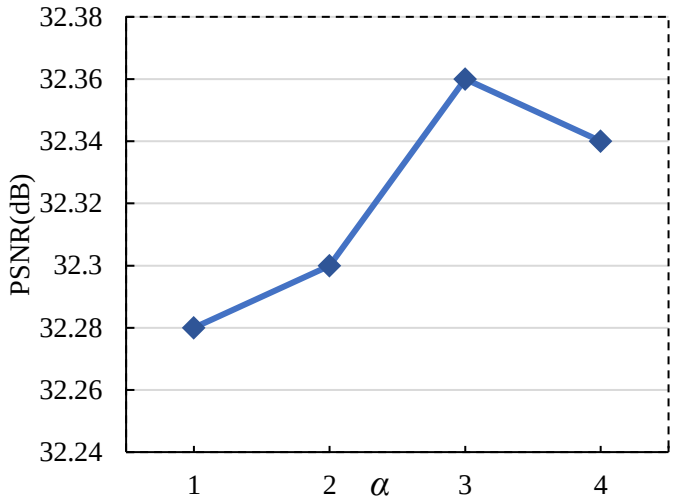


图 3-14 递归门控卷积块超参数 α 消融结果

从图 3-14 中可以看出当 $\alpha=3$ 时，取得最优效果。

在递归门控卷积可以扩展到任意阶数，所以递归门控卷积的阶数也需要进行消融实验。本文选取阶数为 2、3、5 以及混合阶数分别构建模型 1、模型 2、模型 3 与模型 4，在上采样倍数为 4 时，DIV2K 数据集上重新训练网络，采用 Set5、Set14、B100、Urban100、Manga109 数据集计算 PSNR 与 SSIM 作为评价指标。其结果如表 3-5 所示。

表 3-5 递归门控卷积块高阶交互阶数消融结果

方法	参数量(K)	FLOPs(G)	Set5 PSNR/SSIMM	Set14 PSNR/SSIM	B100 PSNR/SSIM	Urban100 PSNR/SSIM	Manga109 PSNR/SSIM
模型 1	422	24.1	32.12/0.8940	28.58/0.7811	27.56/0.7357	25.91/0.7809	30.40/0.9076
模型 2	436	26.5	32.16/0.8946	28.60/0.7813	27.56/0.7358	25.98/0.7826	30.41/0.9078
模型 3	465	26.6	32.16/0.8941	28.60/0.7816	27.58/0.7362	26.04/0.7839	30.42/0.9082
模型 4	460	26.5	32.20/0.8970	28.64/0.7830	27.60/0.7365	26.14/0.7850	30.47/0.9095

从表 3-5 中可以看出在选择混合阶数时，参数量计算量最小而且实现最佳超分辨率重建效果。

3.4 本章小结

本章提出了基于高阶门控交互的单图像超分辨率重建算法。首先，使用残差全局自适应滤波块挖掘深度特征中的高频信息，使得网络的注意力集中于深度特征的高频区域，提升算法的超分辨率重建性能。并对残差全局自适应滤波块处理的深度特征进行可视化，证实该模块能有效地提取深度特征中的高频信息。其次，提出使用递归门控卷积代替 Transform 中 MSA 结构，以解决 Transform 计算复

杂度高难以在轻量化超分辨率重建算法中使用的问题。并对递归门控卷积中的阶数做了充分地分析后，选择了最合适的混合阶数构建网络实现算法。在上采样倍数为 4 时，该算法与先进的 CNN 算法，例如 RFDN 相比数量下降了 76K，在五个基准数据集上的 PSNR 和 SSIM 至少分别提升了 0.08dB 和 0.0029。与先进的 CNN-Transform 算法相比，例如 LBNet 相比参数量下降了 268K，计算量 FLOPs 下降了 11.5G，在五个基准数据集上的 PSNR 和 SSIM 至少分别提升了 0.03dB 和 0.0007。在视觉效果上，本章提出的算法较于先进算法也有着更优的视觉效果。本章所提出的算法与现有的先进算法相比实现了计算复杂度和参数量以及超分辨率重建性能的最佳平衡。

第4章 基于迭代高阶门控交互的单图像超分辨率重建算法

4.1 引言

在上一章节中，本文探讨了基于 Transform 架构构建的基于高阶门控交互的单图像超分辨率重建算法。这个算法利用递归特征进行高阶门控交互。虽然高阶门控交互阶数较于 MSA 更高，但是在初次交互时特征的维度较低，难以提取到优质特征。本文以此为出发点重新思考高阶门控卷积，并分析普通卷积。深度可分离卷积等卷积的像素与通道上的感受野，由此提出一种新的卷积，即迭代广域门控卷积。同时对轻量化超分辨率重建算法中常用的轻量化注意力机制进行增强以提高超分辨率重建效果。在上一章节中，残差全局自适应滤波块由于在输入图像较大时自适应权重难以与特征尺寸相匹配。故此，需要一种更有效的方式处理频域内的特征，由此提出频域大核卷积。最后，进一步的分析 Transform 架构中 MLP 的作用并选取增强的通道混合注意力模块代替 MLP 进一步轻量化 Transform 架构。同时，为分析残差连接的作用，提出残差控制单元分析残差链接。通过实验发现残差控制单元能加速模型收敛，并提高模型超分辨率重建性能。

4.2 迭代高阶门控网络

4.2.1 网络结构

与上一章节类似，迭代高阶门控卷积网络(iterative high order gated convolutional networks, IGCN)由浅层特征提取单元(shallow feature extraction modules, SFM)、深度特征提取单元(deep feature extraction modules, DFM)、高分辨率图像重建单元组成(HR image reconstruction module, IRM)。其中深度特征提取单元由迭代门控卷积和增强注意力块(iterative gated convolution and enhanced attention block, ICEAB)组成。

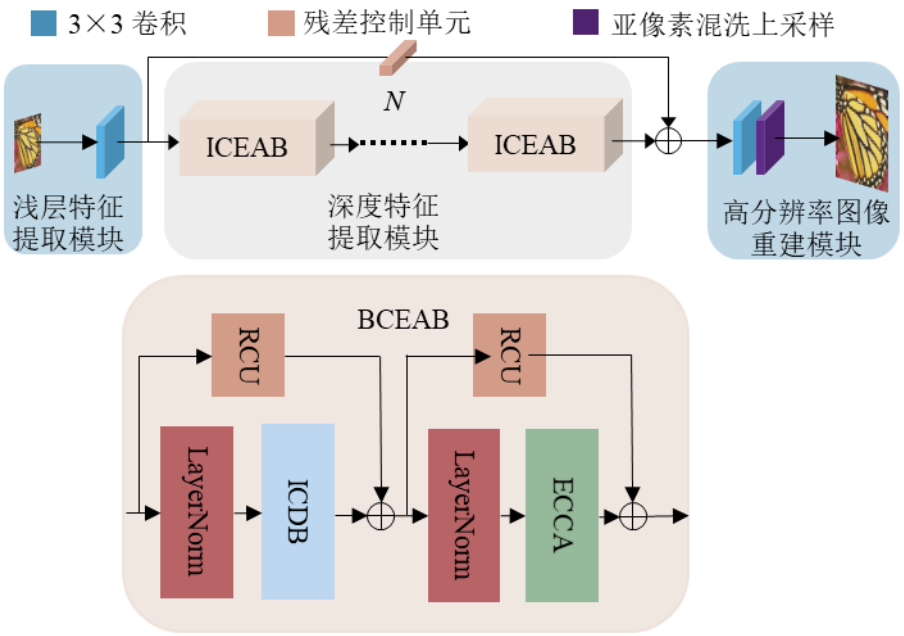


图 4-1 迭代高阶门控网络结构图

首先浅层特征提取模块依旧是由一个 3 × 3 卷积组成，使输入图像 $I_{\mathrm{LR}} \in \mathbb{R}^{H\times W\times 3}$ 的维度从 3 升高到 C 得到深度特征 $F_{\mathrm{S}} \in \mathbb{R}^{H\times W\times C}$。其过程如公式(4-1)所示。

$$F_{\mathrm{S}}=H_{\mathrm{Conv}}(I_{\mathrm{LR}}) \tag{4-1}$$

式中 H_{Conv} 代表 3 × 3 卷积。

随后深度特征将会在深度特征提取单元中通过迭代门控卷积和增强注意力块。迭代门控卷积和增强注意力块由层归一化(layer norm, LN)、迭代门控卷积蒸馏块(iterative gated convolution distillation block, ICDB)、增强混合通道注意力模块(enhanced contrast-aware channel attention, ECCA)以及带由残差控制单元的残差链接组成，其结构如图 4-1 所示。深度特征交互过程可以由公式(4-2)所示。

$$F_{i+1}=H^{i}_{\mathrm{BCEAB}}(F_i), i=0,\ldots,n \tag{4-2}$$

式中，H^{i}_{BCEAB} 代表第 i 个 ICEAB 块，F_i 与 F_{i+1} 分别代表 ICEAB 块的输入特征与输出特征，特别的 $F_0=F_{\mathrm{S}}$，$F_{\mathrm{D}}=F_n$。

最后通过一个残差链接聚合浅层特征与深层特征得到聚合特征，F_{UP} 输入 IRM 得到最后的高分辨图像即输出，如公式(4-3)所示。

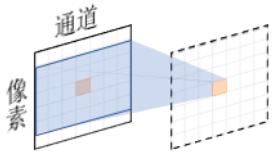
$$I_{\mathrm{SR}}=H_{\mathrm{UP}}(H_{\mathrm{Conv}}(F_{\mathrm{D}}+H_{\mathrm{RCU}}(F_{\mathrm{S}}))) \tag{4-3}$$

式中，H_{Conv} 代表 3×3 卷积。H_{UP} 代表亚像素混洗上采样。F_D 为深度特征提取模块的输出。

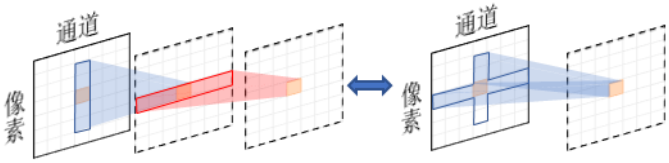
4.2.2 迭代高阶门控模块

4.2.2.1 卷积感受野分析

首先对普通卷积、深度可分离卷积的感受野，将感受野拆解为像素与通道。普通卷积如图 4-2 所示。



(a) 普通卷积感受野



(b) 深度可分离卷积感受野

图 4-2 感受野示意图

在图 4-2 中普通卷积在像素域感受野有限，在通道域有全局的感受野。对于深度可分离卷积，它将普通卷积分解为深度卷积和点卷积。深度可分离卷积在像素域上的感受野与普通卷积类似，但是在通道域中仅对关注一个通道。点卷积在通道域中有全局感受野，但是在像素域中仅关注单独像素。通过先做深度卷积随后在做点卷积再做深度卷积，合并后的感受野也是有限的，其大小远低于普通卷积的感受野，所以使用深度可分离卷积处理特征的效果要低于普通卷积。需要设计一种新的卷积，与普通卷积拥有类似的感受野且在卷积的设计上与深度可分离卷积类似，以此提高特征交互效果同时降低参数量。

本文受到门控卷积启发，重新分析深度卷积与点卷积，重新规划位置并引入哈达玛积，设计了一种新的卷积，即广域门控卷积。其拥有深度可分离卷类似的设置，而且拥有与普通卷积的感受野。其感受野如图 4-3 所示。

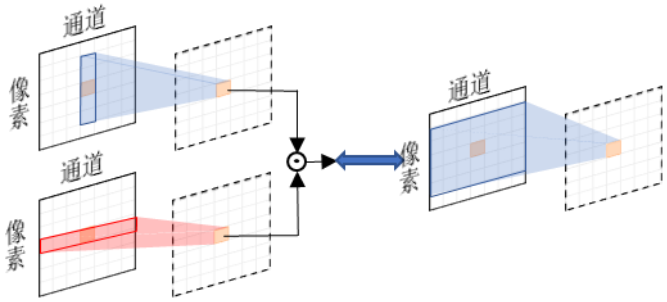


图 4-3 广域门控卷积示意图

从图 4-3 可以看出利用哈达玛积可以使得特征更良好的交互，达到更大的感受野。而且广域门控卷积在设置上与深度可分离卷积类似，仅采用点卷积与深度卷积。这说明广域门控卷积的参数量要远远低于普通卷积，而且在感受野上普通卷积相同。

通过分析卷积的感受野，广域门控卷积更适合于轻量化网络模型的搭建。本文再将其迭代化能够实现高阶特征交互，使得其更加灵活与高效。接下来本文叙述迭代高阶门控卷积的构建。

4.2.2.2 高阶门控卷积的构建

为构建高阶门控卷积，首先分析普通门控卷积，普通门控卷积结构图如下图 4-4 所示，其过程由公式(4-4)与公式(4-5)表示。

$$F_{n-1}, G_{n-1} = H_{\mathrm{S}}(H_{\mathrm{Conv}}(F_{n-2})) \tag{4-4}$$

$$F_n = \phi(F_{n-1}) \odot \sigma(G_{n-1}) \tag{4-5}$$

式中，H_{Conv} 与 H_{S} 表示卷积与通道分离操作。$F_{n\text{-}2}$ 通过卷积与通道分离生成交互特征 $F_{n\text{-}1}$ 与门控特征 $G_{n\text{-}1}$。ϕ 与 σ 代表激活函数，ϕ 拥有多种选择，例如 RELU、LeakyReLU、GELU 等激活函数。σ 主要代表 Sigmoid 激活函数。$\odot$ 代表逐元素乘积，即哈达玛积。

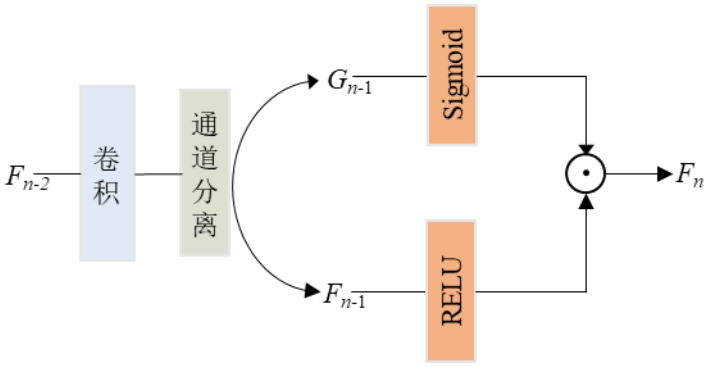


图 4-4 普通门控卷积结构图

普通门控卷积主要利用普通卷积与哈达玛积对输入特征进行处理。而且中间有两次额外的非线性操作，这极大的增加了算法的计算复杂度，所以本文首先对普通门控卷积进行分解，将普通卷积分解为点卷积与深度卷积。如图 4-5(a)所示，其过程可以由公式(4-6)与公式(4-7)表示。

$$F_{n-1}, G_{n-1} = H_{\mathrm{S}}(H_{\mathrm{PWConv}}(F_{n-2})) \tag{4-6}$$

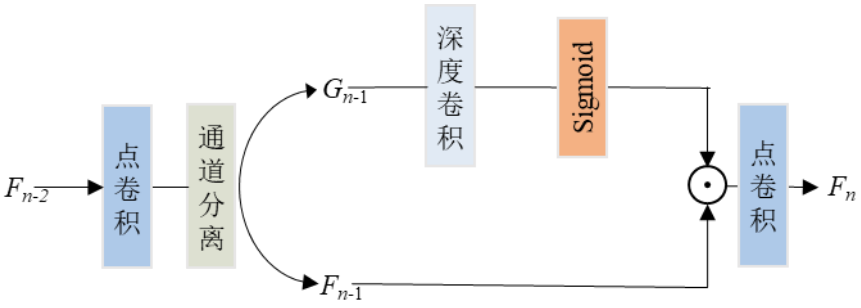
$$F_n = H_{\mathrm{PWConv}}(F_{n-1} \odot \sigma(H_{\mathrm{DWConv}}(G_{n-1}))) \tag{4-7}$$

式中，H_{PWConv} 代表点卷积。H_{DWConv} 代表深度卷积。

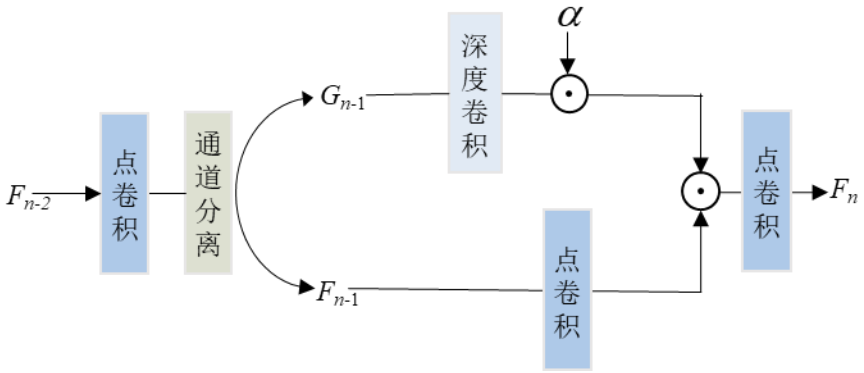
随后，利用在上一节提出的广域门控卷积进一步在交互特征分支中加入一个点卷积，进一步促进特征交互，扩大感受野，其结构如图 4-5(b)所示，其规程由公式(4-8)与公式(4-9)表示。

$$F_{n-1}, G_{n-1} = H_{\mathrm{S}}(H_{\mathrm{PWConv}}(F_{n-2})) \tag{4-8}$$

$$F_n = H_{\mathrm{PWConv}}(H_{\mathrm{PWConv}}(F_{n-1}) \odot H_{\mathrm{DWConv}}(G_{n-1})) \tag{4-9}$$



(a) 分解的门控卷积



(b) 不带有激活函数的广域门控卷积

图 4-5 广域门控卷积结构图

首先，在图 4-5(a)中利用点卷积与深度卷积构建了广域门控卷积，而且删去了一个激活函数，降低计算复杂度。最后再哈达玛积之后加入点卷积，混洗特征。在图 4-5(b)中进一步利用超参数 α 控制门控特征，并在交互特征分路加入点卷积形成广域门控卷积。接下来要在广域门控卷积中实现特征的高阶空间交互，首先将门控特征进一步的拆分，随后在每次特征交互之后加入点卷积，混洗特征并，并扩大深度卷积感受野，迭代广域门控卷积可以扩展到任意阶，本文选取三阶广域门控卷积做详细说明，结构图如图 4-6 所示，其过程由公式(4-10)~公式(4-13)表示。

首先，设输入特征为 $F_{n-2}\in\mathbb{R}^{H\times W\times C}$，同样使用点卷积扩张输入特征维度随后进行通道分离生成交互特征 $F_{n-1}^{0}\in\mathbb{R}^{H\times W\times C_{base}}$ 与门控特征 $G_{n-1}^{0}\in\mathbb{R}^{H\times W\times 6C_{base}}$，其中 H、W、C 为特征的高、宽、以及输入维度，为方便控制迭代广域门控卷积的参数量，引入超参数 C_{base} 控制交互特征与门控特征的维度。

$$F_{n-1}^{0},G_{n-1}^{0}=H_{\text{S}}(H_{\text{PWConv}}(F_{n-2})) \tag{4-10}$$

G_{n-1}^{0} 通过深度卷积与通道分离得到三个门控特征分别为 $G_{n-1}^{1}\in\mathbb{R}^{H\times W\times C_{base}}$、$G_{n-1}^{2}\in\mathbb{R}^{H\times W\times 2C_{base}}$、$G_{n-1}^{3}\in\mathbb{R}^{H\times W\times 3C_{base}}$。其过程由公式(4-11)表示。

$$G_{n-1}^{1},G_{n-1}^{2},G_{n-1}^{3}=H_{\text{S}}(\alpha\odot H_{\text{DWConv}}(G_{n-1}^{0})) \tag{4-11}$$

交互卷积在与门控卷积进行哈达玛积之前会通过点卷积进行维度交互，三个门控特征分别与交互特征进行哈达玛积，其过程如公式(4-12)所示。交互特征与第三个门控特征交互后会利用点卷积进行一次维度交互，其过程如公式(4-13)所示。

$$F_{n-1}^{i}=H_{\text{PWConv}_i}(F_{n-1}^{i-1})\odot G_{n-1}^{i},i=1,2,3 \tag{4-12}$$

$$F_n = H_{\mathrm{PWConv}_4}(F_{n-1}^3) \tag{4-13}$$

式中，F_{n-1}^i 为第 i 次空间交互之后的结果。H_{PWConv_i} 为第 i 次哈达玛积之前对交互特征进行的点卷积。H_{PWConv_4} 为最后一次点卷积。

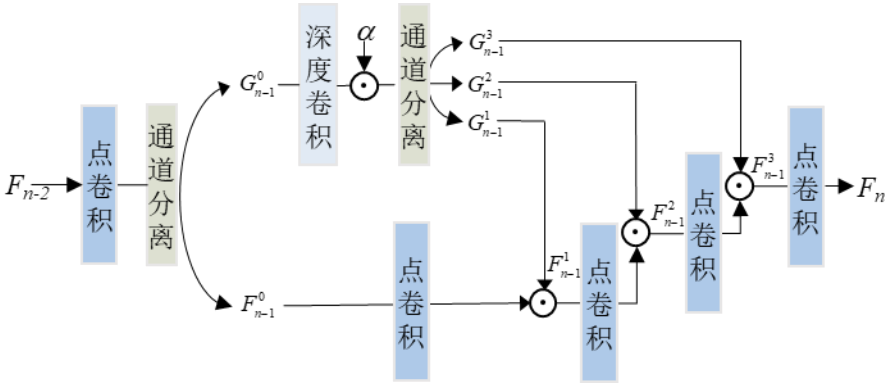


图 4-6 迭代广域门控卷积结构图

本节介绍了本章算法最重要的一个结构迭代广域门控卷积，相较于上一章节的递归门控卷积，首先解决了在低阶空间交互时维度过少导致低阶交互没有意义的问题。而且通过分析感受野进一步修改门控卷积用较少的参数有效地扩大了卷积的感受野。接下来本文将叙述对在轻量化超分辨率重建算法中常用的轻量化注意力模块，并对齐进行改进。频域大核卷积也将在轻量化注意力模块叙述，并探讨频域卷积与注意力机制的关系。

4.2.3 轻量化注意力模块

本节首先介绍在轻量化超分辨重建算法中常用的两个轻量化注意力模块分别为增强空间注意力与混合通道注意机制，他们的原始结构如图 4-7 和图 4-8 所示。

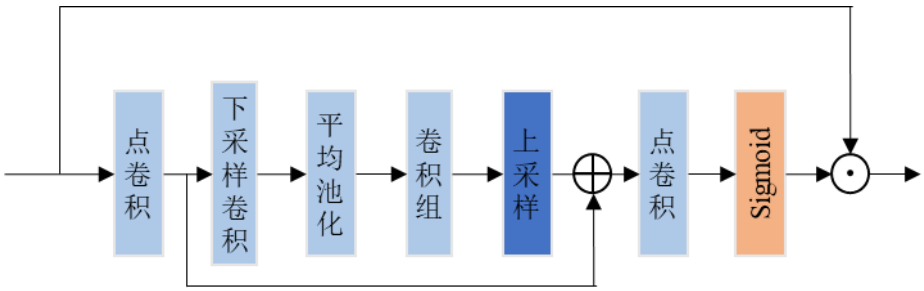


图 4-7 增强空间注意力结构图

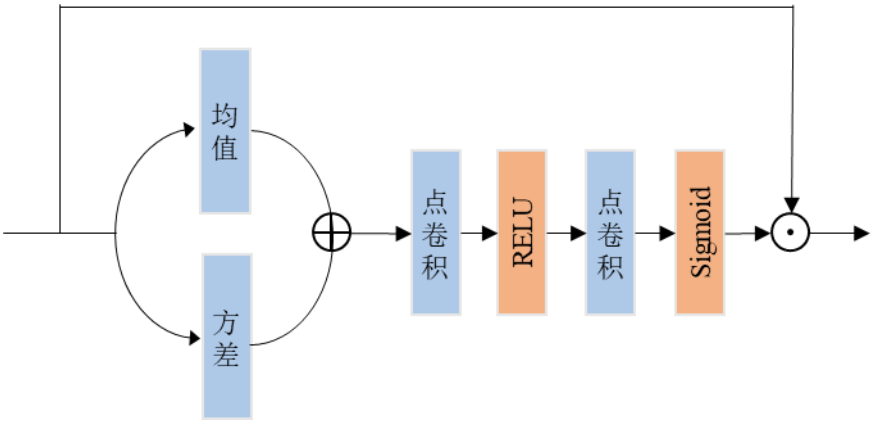


图 4-8 混合通道注意结构图

增强空间注意力通过一个点卷积收缩特征维度，以降低注意力模块的参数量随后使用用一个步长为 2 卷积核大小为 3 的卷积进行下采样，在图 4-7 标注为下采样卷积，再通过平均池操作进一步下采样特征，利用一组卷积学习特征，最后上采样并进行一次残差链接，这样做的原因是在下采样过程中学习低分辨率特征最后通过残差链接滤去低分辨率信息，获得高分辨信息的注意力图。最后高分辨率信息注意力图经过 Sigmoid 的激活函数与原始特征做哈达玛积得到输出特征。本文利用广域门控卷积对增强空间注意力机制进行优化与修改，修改后的注意力，即广域门控卷积空间注意力，其结构如图 4-9 所示。

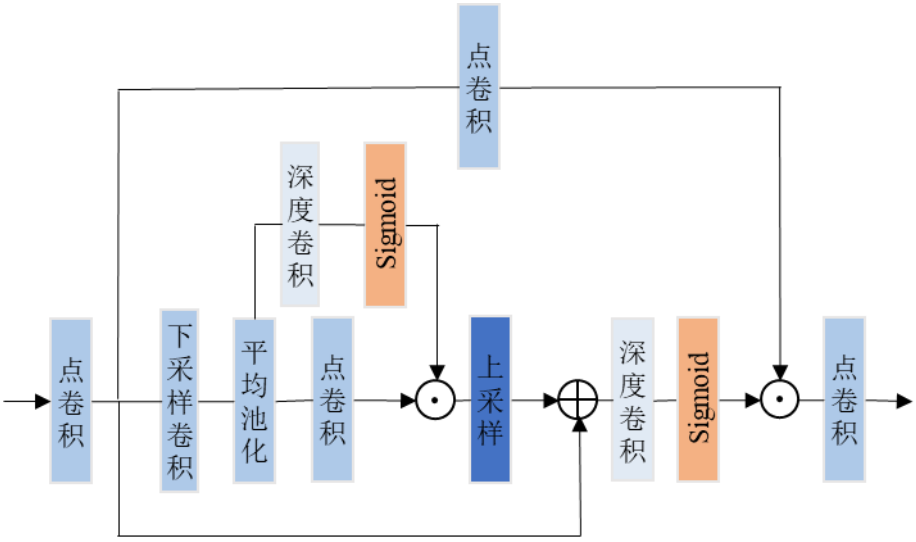


图 4-9 广域门控卷积空间注意力结构图

通过广域门控卷积代替增强空间注意力中的卷机组学习特征中的低频信息，从参数量上来说原始的增强空间注意力主要使用普通卷积这样参数量往往较高，无法很好地应用到算法中。后续文献[57]将普通卷积换为蓝图卷积降低了参数量，

但是从感受野上来说要小于普通卷积。所以广域门控卷积空间注意力能更好的平衡参数量与感受野以实现更优的超分辨率重建效果。

具体地说，广域门控卷积注意力的第一步与增强空间注意力一样，都是通过一个点卷积收缩维度。为便于解释，设输入为 $x \in \mathbb{R}^{H \times W \times C}$ 引入超参数 C_{base} 以控制收缩后的维度，其过程可由公式(4-14)表示。

$$F = H_{\text{PWConv}}(x) \tag{4-14}$$

式中，H_{PWConv} 为点卷积，其输出为 $F \in \mathbb{R}^{H \times W \times C_{base}}$ 。

随后下采样过程中也与增强空间注意力相同，其过程可表示为公式(4-15)。

$$F_{\text{down}} = H_{\text{P}}(H_{\text{DConv}}(F)) \tag{4-15}$$

式中，H_{DConv} 为步长为 2 卷积核大小为 3 的深度卷积，H_{P} 为平均池化操作。其输出用 F_{down} 表示。

下采样后的卷积会经过第一次广域卷积，其过程由公式(4-16)表示。

$$F_{\text{G}} = H_{\text{PWConv}}(F_{\text{down}}) \odot \sigma(H_{\text{DWConv}}(F_{\text{down}})) \tag{4-16}$$

式中，H_{DWConv} 与 H_{PWConv} 分别为深度卷积与点卷积，σ 为 Sigmoid 激活函数。其输出为 F_{G}。值得注意的是，为节省参数，在这次广域门控卷积中门控特征与交互特征均为 F_{down}。

第一次广域门控卷积后的特征会进行上采样和一次残差链接，如公式(4-17)所示。

$$F_{\text{up}} = H_{\text{UP}}(F_{\text{G}}) + F \tag{4-17}$$

式中，H_{UP} 为上采样操作，输出的上采样特征记为 F_{up}。

最后上采样的特征与原始特征会进行一次广域控卷积，即上采样特征为门控特征，原始特征为交互特征。这个过程用公式(4-18)表示。

$$y = H_{\text{PWConv}}(F) \odot \sigma(H_{\text{DWConv}}(F_{\text{up}})) \tag{4-18}$$

式中，$y \in \mathbb{R}^{H \times W \times C}$ 为广域门控卷积注意力的输出。

混合通道注意其中的混合是混合特征的均值与方差，随后通过点卷积收缩与扩张维度最后通过 Sigmoid 激活函数与原始特征相乘得到最后的输出。本文对混合通道注意进行增强提出增强混合通道注意，其结构图如图 4-10 所示。

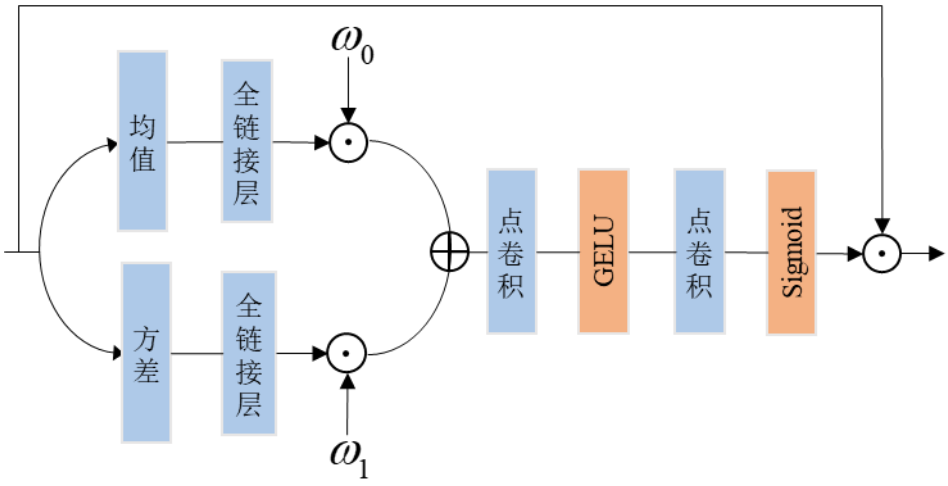


图 4-10 增强混合通道注意结构图

与原始的混合通道注意相比本文在均值与方差之后加入全链接层混洗均值与方差，并引入外部可学习参数 ω_0 和 ω_1 控制均值与方差加权求和。随后的过程与原始的混合通道注意相似，并且把 RELU 激活函数更换为 GELU 激活函数。

首先，设输出特征为 $x \in \mathbb{R}^{H\times W\times C}$ 获取特征的方差与均值该过程可以由公式(4-19)和公式(4-20)表示。

$$F_{\text{avg}} = H_{\text{AVG}}(x) \tag{4-19}$$

$$F_{\text{var}} = H_{\text{VAR}}(x) \tag{4-20}$$

式中 H_{AVG} 表示取特征的均值，H_{VAR} 表示取特征的方差。输出的均值与方差分别由 $F_{\text{avg}} \in \mathbb{R}^{1\times 1\times C}$ 与 $F_{\text{var}} \in \mathbb{R}^{1\times 1\times C}$ 代表。

随后特征的均值与方差分别通过两个全链接层，随后与外部输入的可学习参数加权求和，该过程可表示为公式(4-21)。

$$F_{\text{sum}} = \omega_0 \odot H_{FC}(F_{\text{avg}}) + \omega_1 \odot H_{FC}(F_{\text{var}}) \tag{4-21}$$

式中，H_{FC} 代表全链接层。$F_{\text{sum}} \in \mathbb{R}^{1\times 1\times C}$ 代表输出。$\odot$ 代表哈达玛积。

最后通过由两个点卷积组成的多层感知机收放维度后，通过 Sigmoid 激活函数与原始特征做哈达玛积之后得到最后的输出 $y \in \mathbb{R}^{H\times W\times C}$。这个过程如公式(4-22)所示。

$$y = F \odot \sigma(H_{\text{MLP}}(F_{\text{sum}})) \tag{4-22}$$

式中，H_{MLP} 为多层感知机。

通过对两个在轻量化超分辨率重建算法中常用的轻量化注意力机制的改进，

可以进一步增强算法超分辨率重建的性能。并且本文在改进轻量化注意力机制时，仅引入了较少的参数对齐修改，但是达到了优越的性能。该部分会在章节 4.3.3 中详细讨论。

在上一章节，本文提出残差全局自适应滤波器，通过分析它有一个明显的缺点也就是当输入图像过大时，会损失部分低频特征导致算法超分辨率重建性能下降。为解决这个问题，本章节提出频域大核卷积块。频域大核卷积块从频域中来说是在做卷积但是在空域中实质上是在做像素注意力，随后本文会详细讨论。首先，频域大核卷积块的结构图如图 4-11 所示。

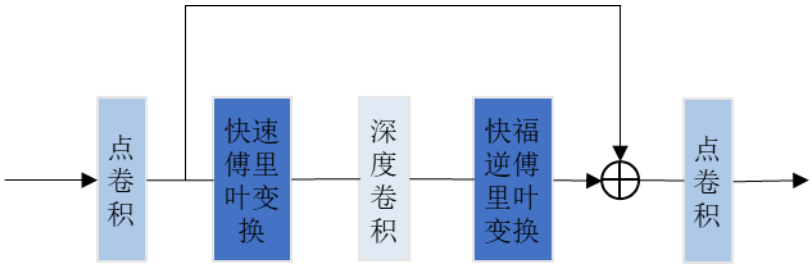


图 4-11 频域大核卷积块结构示意图

频域大核卷积块为减少参数，也利用点卷积收缩输入特征为维度。为便于叙述，设输入特征为 $x \in \mathbb{R}^{H \times W \times C}$，并引入超参数 C_{base} 控制输入特征收缩后的维度。如公式(4-23)所示。

$$F = H_{\text{PWConv}}(x) \tag{4-23}$$

式中，H_{PWCovn} 代表点卷积，其输出为 $F \in \mathbb{R}^{H \times W \times C_{base}}$ 。

随后 F 进行快速傅里叶变换，随后进行深度卷积、快速傅里叶逆变换与残差链接。最后通过点卷积扩张维度得到输出该过程可以由公式(4-24)所示。

$$y = H_{\text{PWConv}}(\mathcal{F}^{-1}(H_{\text{DWConv}}(\mathcal{F}(F))) + x) \tag{4-24}$$

式中，H_{DWConv} 为深度卷积。输出为 $y \in \mathbb{R}^{H \times W \times C}$ 。$\mathcal{F}^{-1}$ 与 $\mathcal{F}$ 分别代表逆快速傅里叶变换和傅里叶变换。傅里叶变换已经在章节 3.2.2 中叙述过，在此不再赘述。

接下来，本文将叙述如何选择深度卷积核的大小与频域卷积与注意力机制的关系。在章节 3.2.2 中，本文叙述过在频域中频谱具有共轭对称性，所以在选择卷积核时，可以选择一个非对称的卷积核，例如：17×9 等非对称卷积核。其他卷积核大小会在章节 4.3.3 进一步讨论与叙述。而选择的核的大小本文认为要尽可能的大一些这样可以充分的筛选低频区域与高频区域，从而映射后的特征为高

频图像特征。根据卷积定理其公式如公式(4-25)所示。

$$F(u,v)*H(u,v)\overset{\mathcal{F}^{-1}}{\Leftrightarrow} f(x,y)\odot h(x,y) \tag{4-25}$$

式中，其中 $F(u,v)$和 $H(u,v)$分别是二维离散图像 $f(x,y)$和 $h(x,y)$的傅里叶变换。卷积用$*$表示，$\odot$表示 Hadamard 积。逆傅里叶变换用$\mathcal{F}^{-1}$代表。

从公式(4-25)中可知频域的卷积实质上空域上的哈达玛积，对卷积核进行一定的补 0 可以使得卷积核与特征尺寸相匹配，实现特征进行空域中的哈达玛积。在上述两个注意模块中最后均使用哈达玛积与特征交互，故在频域中的卷积也是一种注意力机制，但是与空间注意力不同的是频域大核卷积块从频域中提取注意力图，作用于特征。这样做与空域卷积不同，在频域中卷积重新映射高频与低频分量，即在空域中的低频域高频区域，这样在空域中的感受野进一步扩大进一步地说，能得到一个在空域中全局的注意力图。

本节介绍了迭代门控卷积和增强注意力块主要结构，接下来的一节会介绍进一步增强超分辨率重建性能的两个单元，特征蒸馏与残差控制单元。

4.2.4 残差控制单元与特征蒸馏

首先我们介绍残差控制单元，其结构图如图 4-12 所示。残差门控单元的设想是利用一些权重控制残差特征以筛选残差特征。接下来，本文将利用一种简单的方式实现残差控制单元。具体地说，利用深度点卷积即可实现该过程，在章节 2.2.2.1 中介绍了深度卷积，深度点卷积即为卷积核大小为 1 的深度卷积。

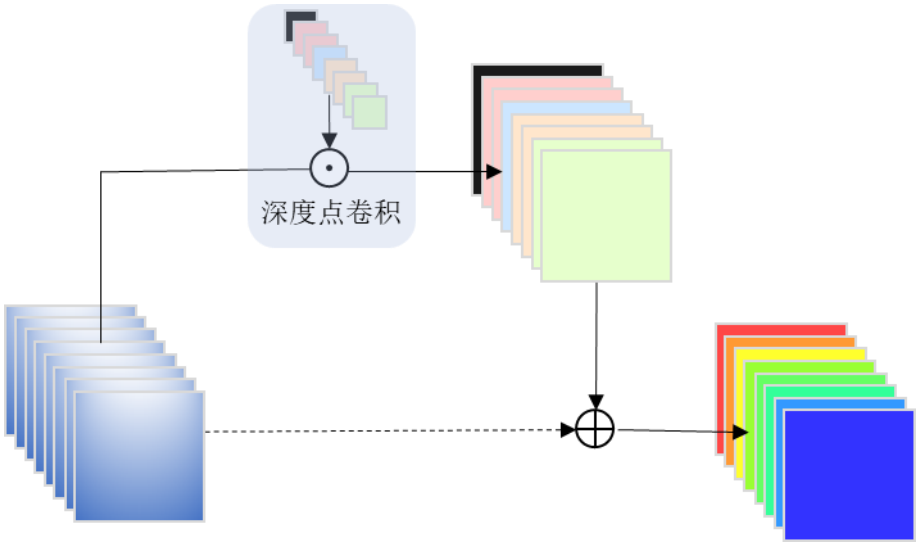


图 4-12 残差门控单元结构图

在图 4-12 虚箭头线代表特征通过任意的特征交互块，深度点卷积可以分解为一个外部引入的可学习参数与残差特征进行哈达玛积交互。随后交互后的残差特征与原始特征相加。残差控制单元可以加快算法的收敛速度，而且实现更好的超分辨率重建性能，具体细节将在章节 4.3.3 中进一步分析。

特征蒸馏是在轻量化超分辨率重建算法中常用的加强长距离特征交互的方法。本章算法也是用这种方式实现特征的长距离交互，具体地说，特征蒸馏链接将在输入，与每次高阶特征交互之后进行。具体的说，利用点卷积下放到特征维度的一半，并采用激活函数最后再输出之前进行通道聚合其过程可以由如公式(4-26)~公式(4-29)所示。

$$F_{\mathrm{dl\_1}} = DL_1(F_{\mathrm{in}}) \tag{4-26}$$

$$F_{\mathrm{dl\_2}} = DL_2(F_0 \odot G_0) \tag{4-27}$$

$$F_{\mathrm{dl\_3}} = DL_3(F_1 \odot G_1) \tag{4-28}$$

$$F_{\mathrm{dl\_4}} = DL_4(F_2 \odot G_2) \tag{4-29}$$

式中，$F_{\mathrm{in}} \in \mathbb{R}^{H\times W\times C}$ 为输入特征，$F_{\mathrm{dl\_}i}(i=1,2,3,4)$ 代表第 i 次特征蒸馏后的特征。为便于叙述引入超参数 C_{base}，$F_k \in \mathbb{R}^{H\times W\times (k+1)\times C_{base}}$，$G_k \in \mathbb{R}^{H\times W\times (k+1)\times C_{base}}$ (k=0, 1, 2)代表第 k+1 次进行哈达玛积的交互特征与门控特征。DL_i (i=1, 2, 3, 4)代表第 i 次蒸馏链接，其由点卷积与激活函数组成，输入维度与输出维度如表 4-1 所示。

表 4-1 特征蒸馏链接输入输出维度表

蒸馏链接	输入维度	输出维度
DL_1	C	C//2
DL_2	C_{base}	C_{base}//2
DL_3	$2C_{base}$	$2C_{base}$//2
DL_4	$3C_{base}$	$3C_{base}$//2

特征蒸馏策略完成之后，迭代门控卷积和增强注意力块的所有机制与结构全部介绍完毕。本文更详细地展现迭代门控卷积和增强注意力块的结构，如图 4-13。

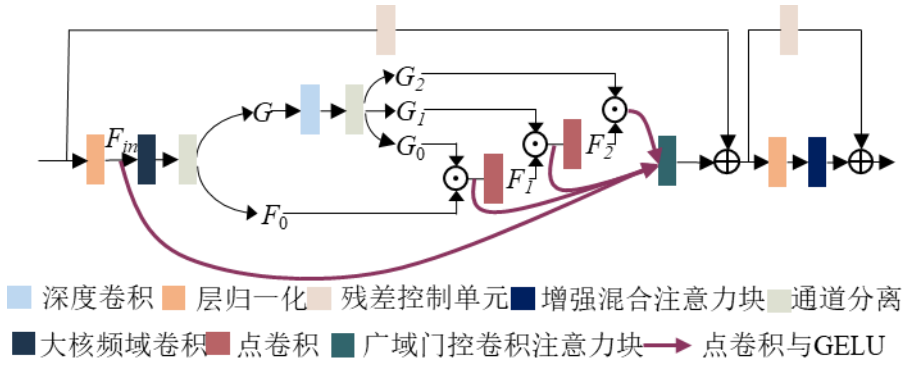


图 4-13 迭代门控卷积和增强注意力块详细结构图

4.3 实验结果与分析

4.3.1 实验设置

（1）数据集和指标

参照文献[46]，训练本章算法-T 的图像由来自 Flickr2K 的 2650 张图像和来自 DIV2K 的 800 张图像组成。训练本章算法-L 的图像为 DIV2K 的 800 张图像。为了评估不同 SR 网络的性能，本文使用 Set5、Set14、B100、Urban100 和 Manga109 这 5 个标准基准数据集。平均峰值信噪比(PSNR)和 Y 通道(即亮度)上的结构相似性(SSIM)被用作评估指标。在计算复杂度与内存消耗上选取，参数量与 FLOPs 与其他先进网络进行对比。

（2）降质模型

使用双三次下采样的方式对图像进行下采样倍数为 2、3、4 的降质。如章节 2.2.1 中，图像降质过程可以由如公式(4-30)表示。

$$I_{\mathrm{LR}}=I_{\mathrm{HR}}\downarrow_{s} \tag{4-30}$$

式中，I_{HR} 代表高分辨率图像，I_{LR} 低分辨率图像。$\downarrow_{s}$ 代表 s 倍的双三次下采样。

（3）本章算法的实现细节

本章算法-T 由 8 个迭代门控卷积和增强注意力块组成，通道数设置为 64，超参数 C_{base} 设置为 24。采用随机旋转 90°、180°、270°和水平翻转的数据增强方法。批量大小设置为 32，每个 LR 输入的 patch 大小设置为 48×48。模型由 AdamW 优化器进行训练。初始学习率设置为 1×10^{-3}，迭代 1×10^{6} 后使用余弦退火学习率衰减方法衰减到 1×10^{-7}。L1 损失用于优化总 1×10^{6} 迭代的模型。L1

损失函数。

本章算法-L 由 14 个迭代门控卷积和增强注意力块组成，通道数设置为 64，超参数 C_{base} 设置为 32。采用随机旋转 90°、180°、270°和水平翻转的数据增强方法。批量大小设置为 64，每个 LR 输入的 patch 大小设置为 48×48。模型由 AdamW 优化器进行训练。初始学习率设置为 1×10^{-4}，迭代 1×10^{6} 后使用余弦退火学习率衰减方法衰减到 1×10^{-7}。L1 损失用于优化总 1×10^{6} 迭代的模型。L1 损失函数如公式(4-31)所示。

$$\mathcal{L}^{\ell_1}=\frac{1}{N}\sum_{i\in N}\left|I_{\text{SR}}(i)-I_{\text{HR}}(i)\right| \tag{4-31}$$

式中，I_{SR} 为超分辨率重建后的图像，I_{HR} 为高分辨图像。$I_{*}(i)$表示图像 I_{*}的第 i 个像素，其中*代表 HR 或 SR。N 为图像的像素个数。

（4）实验环境

本章算法-T 在 GPU 是 GeForce RTX 3060 GPU 上，使用 Pytorch 实现本章算法。CPU 是第 12 代的英特尔酷睿 i5-12490F。代码和数据集保存在固态磁盘 DS42HKVS-78BFKYO 上。操作系统为 windows10。

本章算法-L 在 GPU 为 GeForce RTX 3090 GPU 上，使用 Pytorch 实现本章算法。CPU 是 CPU 为 Intel(R) Xeon(R) Platinum 8358P。操作系统为 ubuntu 20.04。

4.3.2 实验结果

在本节将本章算法-T 与已有的轻量化超分辨重建算法(CARN[104], IMDN[46], PAN[61], RFDN[47], DRSAN[59], ShuffleMixer[105])进行对比。从注意力的角度来说，PAN 网络利用像素注意机制，IMDN 使用混合通道注意力机制，RFDN 使用增强空间注意力机制，这些网络是使用注意力机制提升网络性能的典型网络。从残差控制的角度来说 DRSAN 使用 MSA 控制残差特征。从特征蒸馏链接的角度来说，IMDN 与 RFDN 使用蒸馏链接进一步提升超分辨率重建性能。其对比结果如表 4-2 所示，其中 FLOPs 在高分辨率图像为 1280 × 720 测得，即上采样倍数为 2、3、4 所对应的低分辨率图像输入大小为 640 × 360、430 × 240、320 × 180。表中最优结果用红色标注、次优结果用蓝色标出。

表 4-2 本章算法-T 与先进算法比较表

方法	上采样倍数	参数量[K]	Multi-Adds[G]	Set5 PSNR/SSIM	Set14 PSNR/SSIM	BSD100 PSNR/SSIM	Urban100 PSNR/SSIM	Manga109 PSNR/SSIM
Bicubic	×2	-	-	33.66/0.9299	30.24/0.8688	29.56/0.8431	26.88/0.8403	30.80/0.9339
CARN[104]		1592	222.8	37.76/0.9590	33.52/0.9166	32.09/0.8978	31.92/0.9256	38.36/0.9765
IMDN[48]		694	158.8	38.00/0.9605	33.63/0.9177	32.19/0.8996	32.17/0.9283	38.88/0.9774
PAN[67]		261	70.5	38.00/0.9605	33.59/0.9181	32.18/0.8997	32.01/0.9273	38.70/0.9773
RFDN[47]		534	95	38.05/0.9606	33.68/0.9184	32.16/0.8994	32.12/0.9278	38.88/0.9773
DRSAN[59]		419	85.5	37.99/0.9606	33.57/0.9177	32.16/0.8999	32.10/0.9279	-/-
ShuffleMixer[105]		394	91	38.01/0.9606	33.63/0.9180	32.17/0.8995	31.89/0.9257	38.83/0.9774
本章算法-T		306	62.7	38.09/0.9608	33.76/0.9186	32.17/0.8999	32.27/0.9288	39.03/0.9779
Bicubic	×3	-	-	30.39/0.8682	27.55/0.7742	27.21/0.7385	24.46/0.7349	26.95/0.8556
CARN[104]		1592	118.8	34.29/0.9255	30.29/0.8407	29.06/0.8034	28.06/0.8493	33.50/0.9440
IMDN[48]		703	71.5	34.36/0.9270	30.32/0.8417	29.09/0.8046	28.17/0.8519	33.61/0.9445
PAN[67]		261	39	34.40/0.9271	30.36/0.8423	29.11/0.8050	28.11/0.8511	33.61/0.9448
RFDN[47]		541	42.2	34.41/0.9273	30.34/0.8420	29.09/0.8050	28.21/0.8525	33.67/0.9449
DRSAN[59]		419	43.2	34.41/0.9272	30.27/0.8413	29.08/0.8056	28.19/0.8529	-/-
ShuffleMixer[105]		415	43	34.40/0.9272	30.37/0.8423	29.12/0.8051	28.08/0.8498	33.69/0.9448
本章算法-T		314	28.9	34.45/0.9275	30.40/0.8427	29.13/0.8055	28.27/0.8538	33.91/0.9461
Bicubic	×4	-	-	28.42/0.8104	26.00/0.7027	25.96/0.6675	23.14/0.6577	24.89/0.7866
CARN[104]		1592	90.9	32.13/0.8937	28.60/0.7806	27.58/0.7349	26.07/0.7837	30.47/0.9084
IMDN[48]		715	40.9	32.21/0.8948	28.58/0.7811	27.56/0.7353	26.04/0.7838	30.45/0.9075
PAN[67]		272	28.2	32.13/0.8948	28.61/0.7822	27.59/0.7363	26.11/0.7854	30.51/0.9095
RFDN[47]		550	23.9	32.24/0.8952	28.61/0.7819	27.57/0.7360	26.11/0.7858	30.58/0.9089
DRSAN[59]		419	30.5	32.15/0.8935	28.54/0.7813	27.54/0.7364	26.06/0.7858	-/-
ShuffleMixer[105]		411	28	32.21/0.8953	28.66/0.7827	27.61/0.7366	26.08/0.7835	30.65/0.9093
本章算法-T		327	17	32.37/0.8969	28.67/0.7370	27.61/0.7370	26.20/0.7883	30.69/0.9105

从表 4-2 可以看出本章算法-T 在所有上采样倍数中均取得最优效果，而且计算量 FLOPs 最小，参数量仅比 PAN 网络略高。

（1）参数量与计算量分析

本章算法进一步简化了 Transform 架构，所以在参数量取得了进一步的下降。例如对比使用相同特征蒸馏链接的 RFND 网络，参数量与计算量 Multi-Adds 分别下降 223K 与 6.9G。虽然本章算法对比 PAN 网络参数量略高，但是计算量 Multi-Adds 下降 11.2G，这主要是因为 PAN 网络采用了权重共享方法，虽然该方法可以减少参数量但是会增加计算量。总的来说，参数量与计算量的下降主要归因于本章算法提出的迭代广域门控卷积的高效与轻量，并对 Transform 架构进一步的简化。具体细节将在章节 4.3.3 探讨。

（2）客观指标 PSNR 与 SSIM 分析

对比于主要使用注意力机制的 PAN 网络在上采样倍数为 4 时，在测试数据集 Set5 的 PSNR 与 SSIM 上提升了 0.24 与 0.0021。这证明本章算法增强的注意

力机制即为即为极为有效的。利用特征蒸馏链接与空间注意力的 RFDN 网络在上采样倍数为 4 时，本章算法-T 在测试数据集 Set5 的 PSNR 与 SSIM 提升了 0.13 与 0.0017。这进一步证明了 Transform 架构的优越性，在本章算法中同样也使用了频域中处理特征，不同于上一章节，本章节采用的频域卷积有效的规避了在输入过大时，低频信息丢失对超分辨率重建的影响。进一步的探究了如何在频域出处理特征。总的，来说本章算法-T 在所有上采样倍数中取得了最优的效果，实现了计算量与超分辨率重建性能的最佳平衡。

为进一步的与具有代表性的 CNN-Transform 的轻量化超分辨率重建算法(LBNet[68], SCET[62])以及 Transform 的轻量化超分辨率重建算法(ESRT[69], SwinIR[70], NGswin[71])对比，本文将使用本章算法-L 在上采样倍数为 4 时，同等参数量下与先进算法对比，其对比结果如表 4-3，其中 D2K 代表 DIV2K，DF2K 代表 DIV2K 数据集和 Flickr2K 数据集。

表 4-3 本章算法-L 与先进算法比较表

方法	上采样倍数	训练集	参数量[K]	Sct5 PSNR/SSIM	Sct14 PSNR/SSIM	BSD100 PSNR/SSIM	Urban100 PSNR/SSIM	Manga109 PSNR/SSIM
LBNet[62]	×3	DF2K	736	34.47/0.9277	30.38/0.8417	29.13/0.8061	28.42/0.8559	33.82/0.9460
SCET[57]		DF2K	683	34.53/0.9278	30.43/0.8441	29.17/0.8075	28.38/0.8559	34.29/0.9503
ESRT[63]		DF2K	770	34.42/0.9268	30.43/0.8433	29.15/0.8063	28.46/0.8574	33.95/0.9455
SwinIR[64]		DF2K	886	34.62/0.9289	30.54/0.8463	29.20/0.8082	28.66/0.8624	33.98/0.9478
NGswin[71]		D2K	1007	34.52/0.9282	30.53/0.8456	29.19/0.8078	28.52/0.8603	33.89/0.9470
本章算法-L		D2K	793	34.63/0.9290	30.53/0.8456	29.20/0.8076	28.59/0.8607	34.06/0.9476
LBNet[62]	×4	DF2K	742	32.29/0.8960	28.68/0.7832	27.62/0.7382	26.27/0.7906	30.76/0.9111
SCET[57]		DF2K	683	32.27/0.8963	28.72/0.7847	27.67/0.7390	26.33/0.7915	31.10/0.9155
ESRT[63]		DF2K	751	32.19/0.8947	28.69/0.7833	27.69/0.7379	26.39/0.7962	30.75/0.9100
SwinIR[64]		DF2K	897	32.44/0.8976	28.77/0.7858	27.69/0.7406	26.47/0.7980	30.92/0.9151
NGswin[71]		D2K	1019	32.33/0.8963	28.78/0.7859	27.66/0.7396	26.45/0.7963	30.80/0.9128
本章算法-L		D2K	805	32.53/0.8987	28.79/0.7860	27.68/0.7403	26.48/0.7974	30.95/0.9142

从表 4-3 中可以看出本章算法-L 仅在 DIV2K 上进行训练但与在 DF2K 上训练的 SwinIR，在测试数据集 Set5 上 PSNR 与 SSIM 分别提升了 0.09 与 0.0009。与一样在 DIV2K 上训练的 NGswin 在测试数据集 Set5 上 PSNR 与 SSIM 分别提升了 0.2 与 0.0024。这证明本文简化后的 Transform 网络更优地平衡了计算复杂度与超分辨率重建性能。

为进一步可视化本章算法与先进算法的比较，选取上采样倍数为 4 的 Set5 数据集计算 PSNR，同 FLOPs 与参数量汇总在图中比较，如图 4-14 所示。

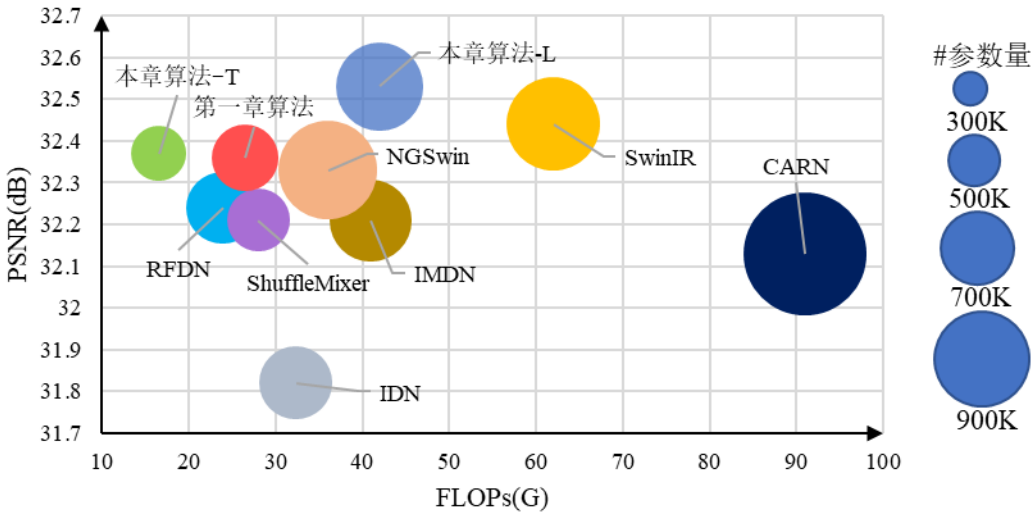


图 4-14 本章算法与先进算法比较可视化图

从图 4-14 可以明显看出，本章算法-T 与第一章算法有着更低的参数量与计算复杂度但是实现了相近的超分辨率重建性能。本章算法-L 虽然参数量与计算复杂度高于第一章算法，但是超分辨率重建性能明显优于第一章算法。这说明本章算法所设计的模块较于第一章算法更轻量化且能保持较优的性能，同时模块更加灵活可以设计不同大小的网络与其他先进算法进行对比。

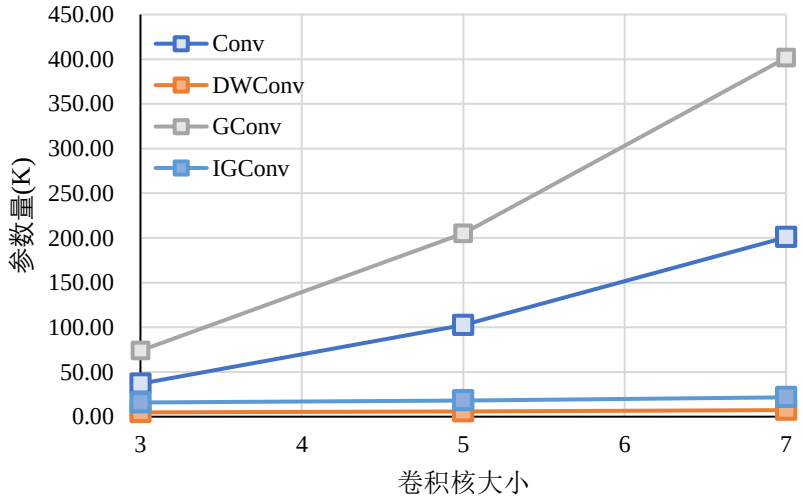
为进一步的比较本章算法-T 与其他先进算法的超分辨率重建后图像的视觉效果，选取 Manga109 数据集中图像纹理较为简单的 ShimatteIkouze_vol01 和图像问题较为复杂的 ARMS 进行本章算法重建后的图像与其他先进算法重建后的图像的视觉效果进行对比如图 4-15

图 4-15 本章算法-T 与其他先进算法视觉效果对比图

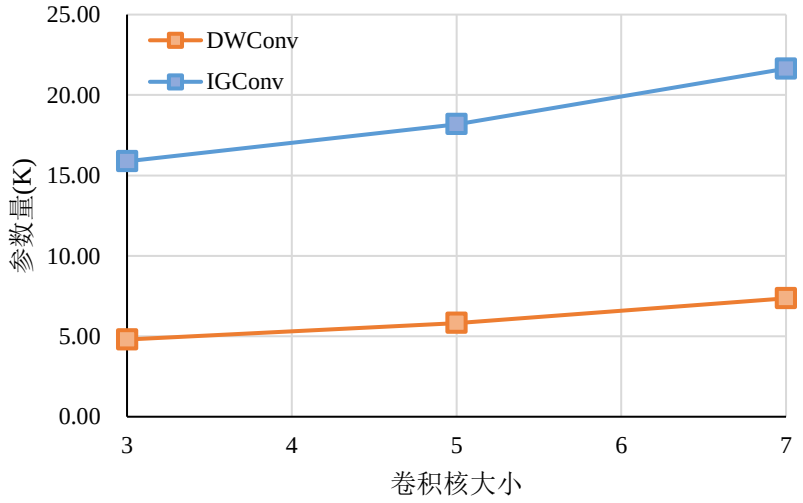
从图像细节上来说本章算法-T 较于其他先进算法更加接近高分辨率图像。这主要归因于迭代广域门控卷积对于图像全局的把控，以及频域大核卷积充分地分解低频区域与高频区域，挖掘特征的高频信息，使得图像的高频部分得以恢复。

4.3.3 算法分析

在上节中本文针对结果简要地分析了本章算法，在本节中本文进一步地讨论算法中各个模块的作用，更详细地分析本章算法。对本文提出的迭代广域门控卷积与其他卷积在同等卷积大小下进行参数量对比，如图 4-16 所示，其中普通卷积、深度卷积、门控卷积、迭代广域门控卷积分别由 Conv、DWconv、GConv、IGConv 所代表。在图 4-16(a)中所有卷积将同时进行对比。图 4-16(b)将进一步清晰地对比深度卷积与迭代广域门控卷积。



(a) 普通卷积、深度卷积、门控卷积、与迭代广域门控卷积对比图



(b) 深度卷积与迭代广域门控卷积对比图

图 4-16 卷积对比图

在图 4-16 中，显然普通卷积与门控卷积的参数量要远远高于深度卷积与迭代广域门控卷积。特别是随着卷积核的扩大普通卷积与门控卷积参数量大大增加，不适合用在轻量化超分辨率重建算法中。迭代广域门控卷积较于深度卷积虽然在参数量上略高于深度卷积，但是在章节 4.2.2.1 中广域门控卷积的感受野要大于深度卷积，而且迭代广域门控卷积能实现任意阶的高阶空间交互，得到比深度卷积更优质的特征。

为验证本文设计的注意力机制的有效性针对注意力机制进行消融实验，实验结果如下表 4-4 所示，其中 FLCB 表示频域大核卷积注意力块，WGSA 代表广域

门控卷积空间注意力，ECCA 表示增强混合通道注意力，ESA 表示增强空间注意，CCA 表示混合通道注意力。模型 1 不使用任何注意机制（注：不使用增强混合通道注意时，Transform 架构中的后归一化层于残差链接也会舍去），模型 2 仅使用增强空间注意力，模型 3 在模型 2 的基础上加入大核频域卷积块，模型 4 在模型 3 的基础上用混合通道注意力代替增强混合通道注意力并加入增强空间注意力，模型 5 使用本文提出的所有注意力。上述所有模型在原训练集上重新训练，并采用测试数据集 Set5 上计算 PSNR 作为评价指标。

表 4-4 注意力机制消融结果表

方法	参数量[K]	FLCB	WGSA	ECCA	ESA	CCA	PSNR[dB]
模型 1	260	✘	✘	✘	✘	✘	32.22
模型 2	290	✘	✘	✔	✘	✘	32.28
模型 3	311	✔	✘	✔	✘	✘	32.32
模型 4	324	✔	✘	✘	✔	✔	32.33
模型 5	327	✔	✔	✔	✘	✘	32.37

从表 4-4 中明显在使用本文所有注意的模型 5 对比于不使用注意力机制的模型 1 的 PSNR 上升 0.15。当与使用原始注意力，即增强空间注意与混合通道注意力的模型 4 作比较 PSNR 上升 0.04。由此证明本文设计的注意力机制能够增强算法的超分辨率重建性能。

接下来对频域大核卷积中的卷积核进行消融实验，其结果如表 4-5 所示。分别设计了四种卷积核应用于频域卷积，他们分别是模型 1 使用点卷积，模型 2 使用卷积核大小为 5×5 的深度卷积，模型 3 使用点卷积与卷积核大小为 7×3 的深度卷积，模型 4 使用卷积核大小为 17×9 的深度卷积，上述所有模型在原训练集上重新训练，并采用测试数据集 Set5、Set14、BSD100、Urban100、Manga109 上计算 PSNR 作为评价指标。

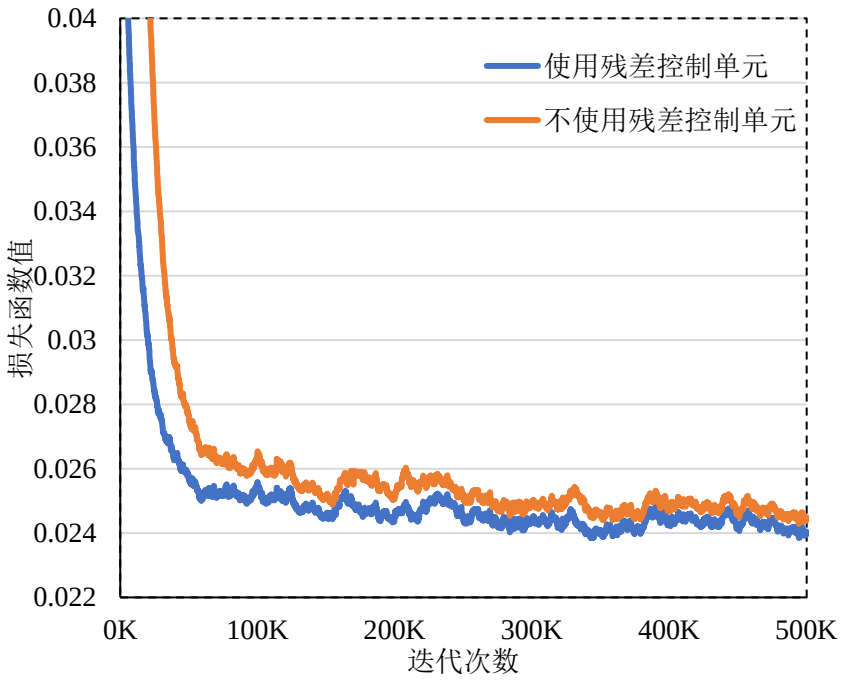
表 4-5 频域大核卷积卷积核大小消融结果表

方法	Set5	Set14	BSD100	Urban100	Manga109
模型 1	32.33	28.63	27.58	26.10	30.53
模型 2	32.34	28.64	27.60	26.14	30.55
模型 3	32.34	28.65	27.60	26.16	30.60
模型 4	32.37	28.67	27.61	26.20	30.69

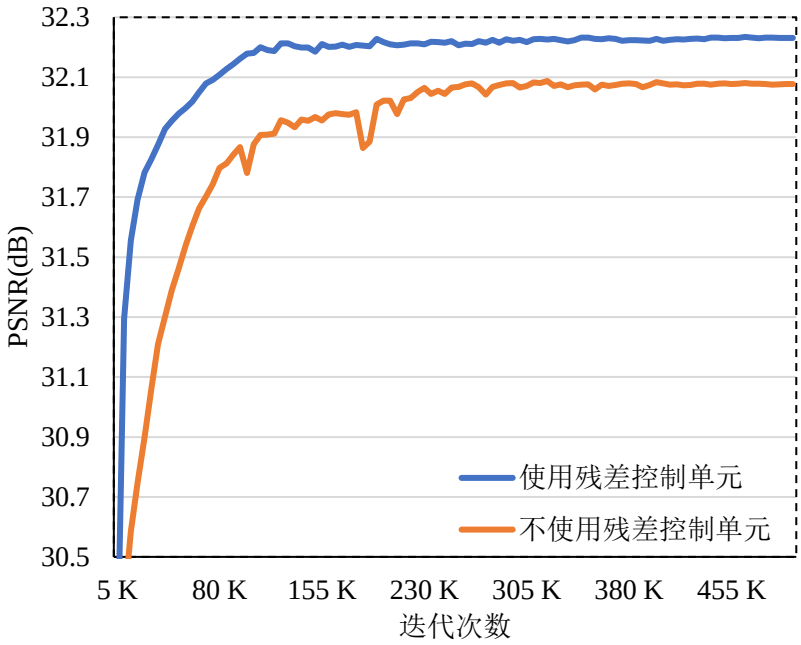
从表 4-5 中，当频域感受野较小时比如使用点卷积或者卷积核大小为 7×7 的

深度卷积实现的效果都是较差的。所以在频域中进行卷积还是需要一个较大的卷积核。

残差控制单元通过控制残差特征对网络产生影响，本文在 DIV2K 上重新训练使用残差控制单元与不使用残差控制单元的网络，具体结果如图 4-17 所示。其中损失函数对比图中的损失函数值进行指数平滑，平滑指数选取 0.99。同时对残差控制单元中的权重进行可视化如图 4-18，其中残差链接编号从 1 ~ 9，1 ~ 8 分别对应本章算法-T 的 8 个迭代门控卷积和增强注意力块中的两个残差控制单元。编号 9 代表长残差链接中的残差控制单元。随后，残差控制单元中的参数取均值，进一步分析残差链接。



(a) 损失函数对比图



(b) PSNR 对比图

图 4-17 残差控制单元分析图

从图 4-17(a)可知，在使用残差控制单元时，损失函数曲线下降速度要远快于不使用残差控制单元的损失函数曲线，且在训练 455K 轮次之后，在带有残差控制单元损失函数值要低于不带有残差控制单元损失函数值。这说明残差控制单元有着加速算法收敛的效果且能提升算法的性能。从图 4-17(b)可知，使用残差控制单元的 PSNR 值明显从开始训练一直到训练结束都高于不带有残差控制单元的 PSNR 值，这进一步证明残差控制单元对残差特征的控制是有效的，提升算法的性能。

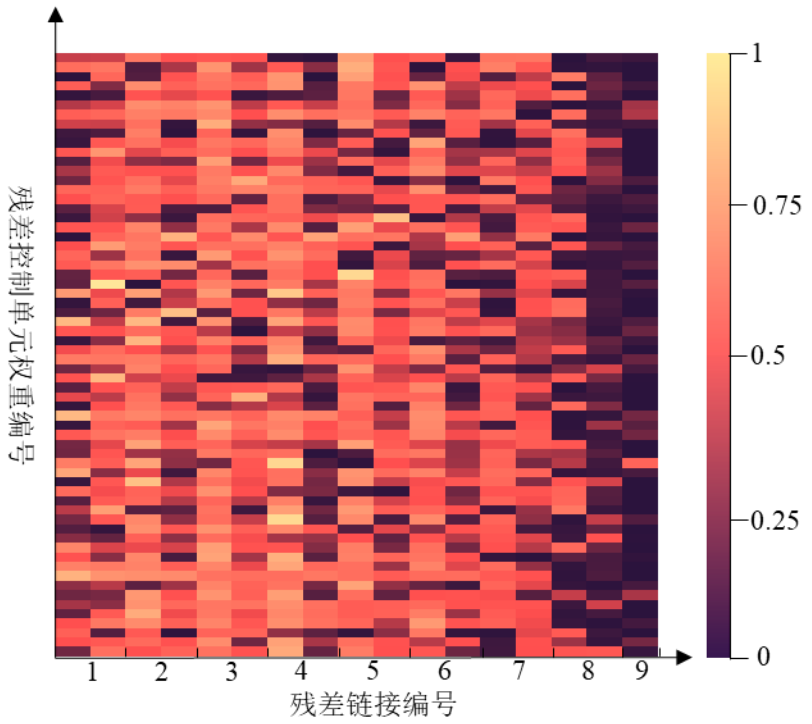


图 4-18 残差控制单元权重可视化图

从图 4-18 中在前 1~8 个残差链接编号中仅有部分权重趋于 0，这说明残差控制单元筛选有效特征给予输出，控制残差特征对后续深度交互块的影响。编号为 9 的长残差链接基本全部趋于 0，长残差链接是浅层特征提取模块的输出与深度特征交互模块的输出的聚合。本文分析浅层特征提出模块的输出为低分辨率特征，毕竟仅用一个卷积提取特征难以获得高分辨率信息，故在重建高分辨率图像时这些特征往往是不重要的，但是在训练初期这些特征可以有效的阻止网络训练崩溃，所以后续网络可以设置长残差链接权重衰减代替可学习权重，进一步加速网络训练。

表 4-6 残差控制单元权重均值表

残差链接编号	RES_1^1	RES_1^2	RES_2^1	RES_2^2	RES_3^1	RES_3^2	RES_4^1	RES_4^2	RES_5^1
均值	0.45	0.60	0.96	0.43	1.02	0.48	1.09	0.32	0.91
残差链接编号	RES_5^2	RES_6^1	RES_6^2	RES_7^1	RES_7^2	RES_8^1	RES_8^2	RES_9^1	RES
均值	0.37	0.74	0.28	0.43	0.31	0.31	0.08	0.04	0.52

从表 4-6 中 RES_i^j (i=1, 2, 3, 4, 5, 6, 7, 8 j = 1, 2)代表第 i 个迭代门控卷积和增强注意力块的第 j 个残差控制单元。RES_9^1 代表长残差链接。最后的 RES 是 RES_1^1 到 RES_8^2 的均值。对表格进行分析，RES_1^1 到 RES_8^2 残差控制单元的权重主要集中于 0.3 ~ 1 之间最后整体的均值为 0.52，这进一步说明不是所有的残差特征对深

度交互后的特征起到积极作用。在构建网络时可以考虑给残差链接设置收放因子。

4.4 本章小结

本章提出了基于迭代高阶门控交互的单图像超分辨率重建算法。首先，设计了迭代广域门控卷积，该卷积较于深度可分离卷积有着更大的感受野，较于普通卷积有着更小的参数量与计算复杂度。同时，该卷积解决了递归门控卷积在特征交互阶数较低时，特征维度过小无法实现有效地实现特征交互的问题。其次，设计了大核频域卷积块利用频域产生注意力图作用于像素域，与在像素域上的注意力机制相比该模块拥有更大的像素感受野，能充分地提取特征的高频信息。随后，针对轻量化常用的注意力机制进行改进，更优地实现深度特征的交互，提升算法的超分辨率重建性能。最后，针对残差特征设计残差控制单元为每个残差特征设置单独的缩放因子控制残差特征，该单元能有效地提升算法的收敛速度与超分辨率重建性能。本章所提出的算法分别设计了两个不同参数量的网络分别为本章算法-T 与本章算法-L。在上采样倍数为 4 的情况下，本章算法-L 与 Transform 算法相比，例如 NGSwin 在超分辨率重建性能相近的情况下，参数量下降了 214K。本章算法-T 与 CNN 算法相比，例如 RFDN，参数量与计算复杂度分别下降 223K 与 6.9G，在五个基准数据集上的 PSNR 和 SSIM 至少分别提升了 0.04dB 和 0.001。本章所提出的算法与现有的先进算法相比具有最小的计算复杂度且实现了最优客观指标。

第5章 基于特征蒸馏与结构损失的单图像超分辨率重建算法

5.1 引言

在上两个章节本文通过分析 Transform 架构，并替换其中的多头自注意力机制与前馈神经网络。由此得出一个结论 Transform 架构中主要是因为其层归一化、全局特征的交互和通道维度的交互，提升算法性能。以此为指导本章使用 CNN 的架构加上从 Transform 中分析的架构优越性，重新规划 CNN 架构的设计。首先重新分析全局特征交互与前馈神经网络，选择利用大核卷积进行全局特征交互，使用特征自蒸馏块代替前馈神经网络，进一步调整网络结构。更重要的是，通过分析频谱，提出一种行的损失函数约束网络，即结构损失。

5.2 特征自蒸馏网络

5.2.1 模型及网络结构

超分辨率重建模型流程框图以及特征自蒸馏网络(feature self-distillation network, FSDN)的结构，如图 5-1 所示。

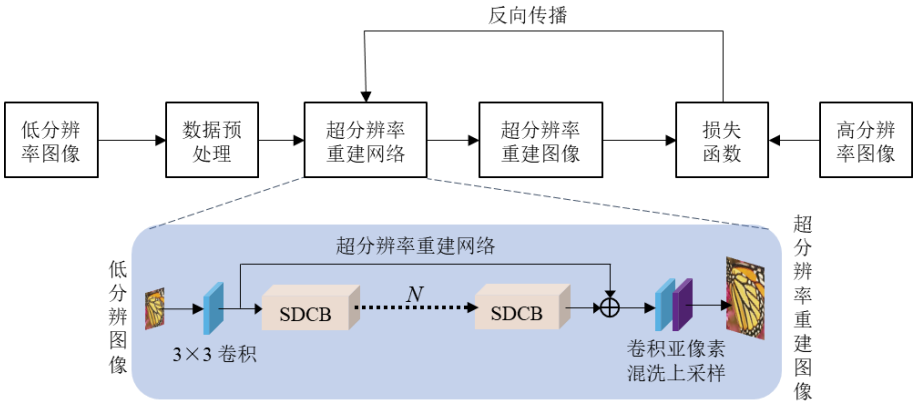


图 5-1 模型及网络结构图

特征首先通过一个卷积核大小为 3 的卷积（图中缩写为 3×3 卷积），卷积的输入维度为 3 输出维度为 C。设输入 $I_{\mathrm{LR}} \in \mathbb{R}^{H\times W\times 3}$，该过程如公式(5-1)所示。

$$F = H_{\mathrm{Conv}}(I_{\mathrm{LR}}) \tag{5-1}$$

式中，H_{Conv} 代表卷积核大小为 3 的卷积，浅层特征 $F_{\text{S}} \in \mathbb{R}^{H\times W\times C}$ 为卷积的输出。

通过卷积核大小为 3 的卷积后的特征 F_{S} 会进入自蒸馏大核卷积块(self-distillation and large kernel convolution block, SDCB)如公式(5-2)所示。自蒸馏大核卷积块由自蒸馏块与大核卷积块组成在章节 5.2.2 与 5.2.3 详细介绍。

$$F_k = H_{\text{SDCB}}^{k}(F_{k-1}), k = 1,2,3\cdots N \tag{5-2}$$

式中，H_{SDCB}^{k} 为第 i 个自蒸馏大核卷积块。F_{k-1} 和 F_k 分别为自蒸馏卷积块的输入与输出。特别的 $F_0 = F_{\text{S}}, F_N = F_{\text{D}}$。

经过 N 个自蒸馏卷积块之后，通过长残差链接聚合浅层特征与深层特征之后，利用卷积亚像素混洗上采样重建高分辨率图像。卷积亚像素混洗上采样由一个核大小为 3 的卷积与亚像素混洗上采样组成，高分辨率图像重建过程如公式(5-3)如下。

$$I_{\text{SR}} = H_{\text{up}}(H_{\text{Conv}}(F_{\text{D}} + F_{\text{S}})) \tag{5-3}$$

式中，H_{Conv} 代表卷积核大小为 3 的卷积，H_{up} 代表亚像素混洗上采样。$F_{\text{D}}, F_{\text{S}}$ 为深度特征与浅层特征。输出使用 $I_{\text{SR}} \in \mathbb{R}^{s\cdot H\times s\cdot W\times 3}$ 代表，其中 s 是上采样倍数。

5.2.2 特征自蒸馏链接

自蒸馏卷积块，结构如图 5-2(a)所示，输入特征通过通道分离 $F_{\text{in}} \in \mathbb{R}^{H\times W\times C}$ 生成原特征 $F_{\text{I}} \in \mathbb{R}^{H\times W\times p\times C}$ 与卷积特征 $F_{\text{C}} \in \mathbb{R}^{H\times W\times (1-p)\times C}$，其中 p 为超参数控制 F_{I} 与 F_{C} 的维度，该过程如公式(5-4)所示。

$$F_{\text{I}}, F_{\text{C}} = H_{\text{S}}(F_{\text{in}}) \tag{5-4}$$

式中，H_{S} 代表通道分离操作。

随后卷积特征会进入蓝图残差块进行卷积。蓝图卷积结构如图 5-2(d)所示，蓝图残差块结构如图 5-2(c)所示。卷积过程如公式(5-5)所示。

$$F_{\text{B}} = \phi(H_{\text{PWConv}}(H_{\text{DWConv}}(F_{\text{C}})) + F_{\text{C}}) \tag{5-5}$$

式中，H_{PWConv}，H_{DWConv} 分别为点卷积与深度卷积。ϕ 代表激活函数，算法中选择 GELU 函数作为激活函数。

卷积特征通过蓝图残差块之后与原特征进行通道聚合并通过点卷积混洗特

征，该过程如公式(5-6)所示。

$$F_{\text{out}}=\phi(H_{\text{PWConv}}H_{\text{C}}(F_{\text{I}},F_{\text{B}})) \tag{5-6}$$

式中 H_{C} 代表通道聚合。

自蒸馏链接卷积块结构如图 5-2(b)所示，自蒸馏链接卷积块由三个自蒸馏卷积块，蓝图卷积的作用是缩放最后输出维度减少参数。经过特征聚合后点卷积混洗自蒸馏特征得到输出。该过程由公式(5-7)~公式(5-10)所示。

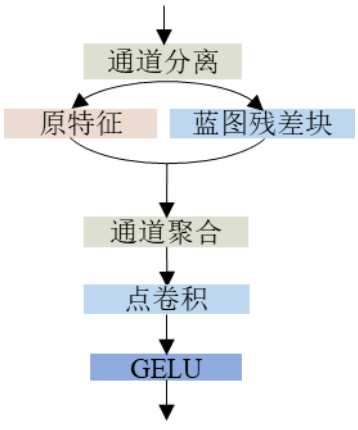
$$F_i=H_{\text{SDB}}^{i}(F_{i-1}),i=1,2,3 \tag{5-7}$$

$$F_4=H_{\text{BC}}(F_3) \tag{5-8}$$

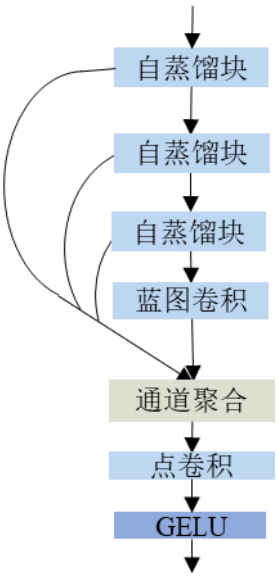
$$F_{\text{cat}}=H_{\text{C}}(F_{\text{I1}},F_{\text{I2}},F_{\text{I3}},F_4) \tag{5-9}$$

$$F_{\text{out}}=\phi(H_{\text{PWConv}}(F_{\text{cat}})) \tag{5-10}$$

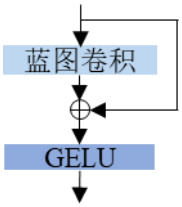
式中，H_{SDB}^{i} 代表第 i 个自蒸馏块。$F_{i\text{-}1}$，F_i 分别代表第 i 个 H_{SDB}^{i} 的输出，特别的 $F_0=F_{\text{in}}$。$F_4\in\mathbb{R}^{H\times W\times(1-p)\times C}$ 为蓝图卷积的输出。聚合特征 $F_{\text{Ii}}\in\mathbb{R}^{H\times W\times(1-p)\times C}$ 为在 H_{SDB}^{i} 中的原特征，其输出为 $F_{\text{cat}}\in\mathbb{R}^{H\times W\times 4(1-p)\times C}$。最后自蒸馏链接卷积块的输出为 $F_{\text{out}}\in\mathbb{R}^{H\times W\times C}$。



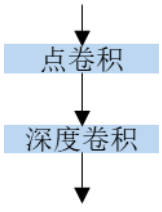
(a)自蒸馏卷积块结构



(b) 自蒸馏链接卷积块结构



(c)蓝图残差块



(d)蓝图卷积结构

图 5-2 自蒸馏块结构图

特征自蒸馏与特征蒸馏不同的是，不需要点卷积收缩特征的维度，而且能通过超参数 p 控制蒸馏特征，使其有自适应性。而且只有部分特征进行卷积进一步

的节省参数。具体细节会在章节 5.4.3 中叙述。

5.2.3 大核卷积块

大核卷积注意力模块依靠普通卷积与空洞卷积组合而成，利用普通卷积与空洞卷积可以获得更大的感受野。普通卷积与空洞卷积的结构图 5-3 所示。

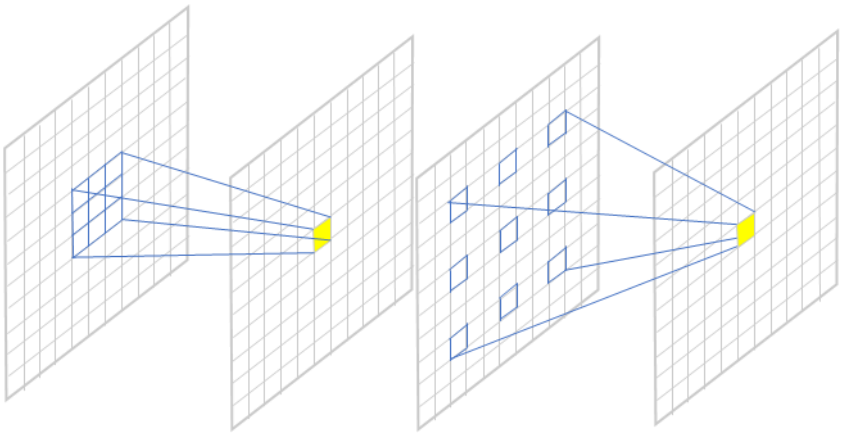


图 5-3 普通卷积与空洞卷积示意图

从图 5-3 中可以看出普通深度卷积（左）的权重集中感受野较小，而空洞深度卷积（右）权重稀疏感受野较大，但是不关注蓝色权重外的像素。故，组合这两种卷积可以获得更大的感受野而且能关注到具体像素，如图 5-4 所示。

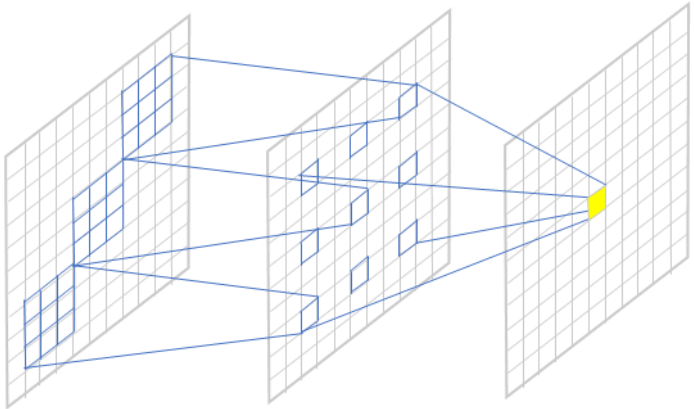


图 5-4 普通卷积与空洞卷积组合示意图

如图 5-4 所示普通卷积与空洞卷积相结合之后感受野较于普通卷积的感受野明显变大。通过文献[107]可以得知只要卷积的感受野足够的大其足以模拟 Transform 架构中的多头自注意力机制实现全局的特征交互。故，本文引入大核卷积的设计，构建大核卷积块。在上两个章节对门控卷积中的哈达玛积进行细致的分析后可以得知其能够实现特征的全面交互。故此，本文利用普通卷积与深度

卷积扩大卷积感受野，同时利用哈达玛积进一步实现全局特征交互。大核卷积块的结构如图 5-5 所示。

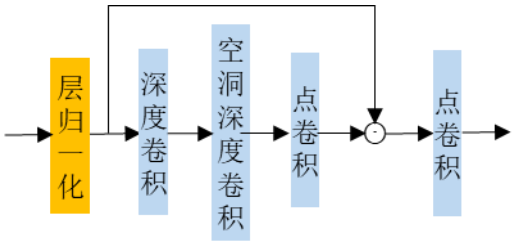


图 5-5 大核卷积块结构图

如图 5-5 所示，进一步利用深度卷积代替普通卷积降低大核卷积块的参数，同时利用点卷积混洗通道维度，该过程由公式(5-11)~公式(5-13)表示。

$$F_{\text{nrom}} = H_{\text{LN}}(F_{\text{in}}) \tag{5-11}$$

$$F_{\text{h}} = H_{\text{PWConv}}(H_{\text{DWConv}}(H_{\text{DWConv-d}}(F_{\text{norm}}))) \odot F_{\text{norm}} \tag{5-12}$$

$$F_{\text{out}} = H_{\text{PWConv}}(F_{\text{h}}) \tag{5-13}$$

式中，H_{LN} 代表层归一化，$H_{\text{DWConv-d}}$ 表示空洞深度卷积。F_{in} 为输入特征，F_{norm} 为通过层归一化之后的特征，F_{h} 为进行哈达玛积之后的特征，F_{out} 为输出特征。

至此，本章算法的网络结构介绍完毕，接下来汇总自蒸馏大核卷积块的结构如图 5-6 所示。

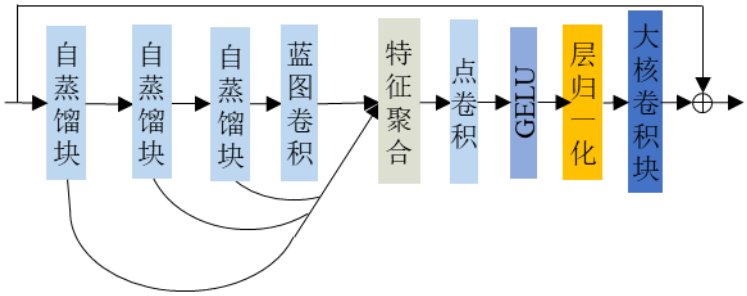


图 5-6 自蒸馏大核卷积块的结构图

从图 5-6 中可以发现与 Transform 架构不同，本章算法仅采用一个层归一化与一个残差链接，而且归一化层的位置处于结构的中间位置，一般 Transform 架构归一化层一般置于最前。本文对层归一化的位置进行消融实验后发现放置于结构中间能达到最优效果且能稳定训练网络，详细细节会在章节 5.4.3 叙述。

本章算法将 Transform 架构中的全局特征交互块与前馈神经网络块进行了调换。这样做的原因是，因为在自蒸馏卷积块中有卷积核大小为 3 的卷积学习特征的像素信息，而 Transform 架构中前馈神经网络块仅对特征的通道层面进行混洗，

不学习特征的像素信息。为此，自蒸馏卷积块中可以学习特征的浅层像素信息为后续的全局特征交互块，即大核卷积块提供更优质的特征。故，本文将自蒸馏卷积块放置于大核卷积块之前。

5.3 结构损失

5.3.1 图像结构分析

在上两章节都使用了在频域中处理特征均取得优越的效果。上章节都是从网络结构的角度分析特征，本章从另外一个角度思考，利用损失函数约束网络的输出，优化网络权重从而使得特征更好的交互。

本文利用一张图像与一张旋转后的图像，分别对它们进行傅里叶变换，并拆解为幅度谱与相位谱。随后交换两张图像的相位谱进行逆傅里叶变换后可视化结果，如图 5-7 所示。

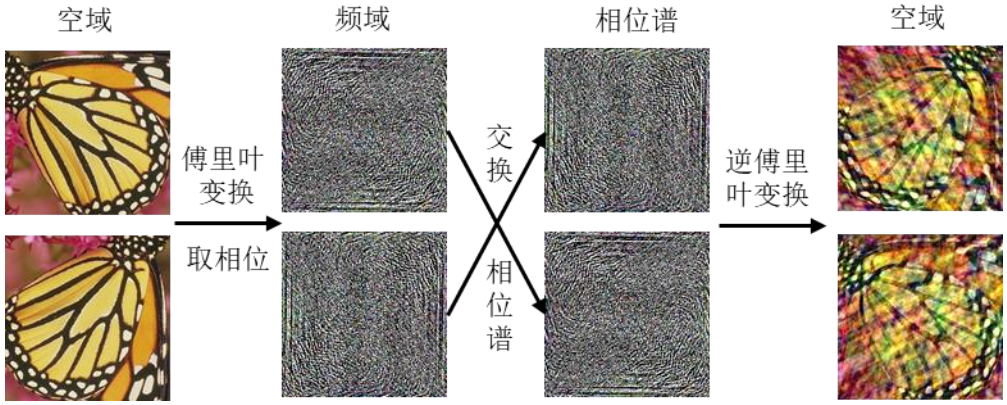


图 5-7 相位谱分析图

从图 5-7 可知在交换相位谱之后，显然没有旋转的图像进行了旋转，而应该旋转后的图像没有进行旋转。所以，相位谱决定这图像的结构信息，这对于超分辨重建来说是及其重要的。

进一步对不同分辨率下的图像相位谱进行分析，如下图 5-8。选取一张低分辨率图像与一个高分辨率图像分别对它们的相位谱进行可视化之后进行对比。

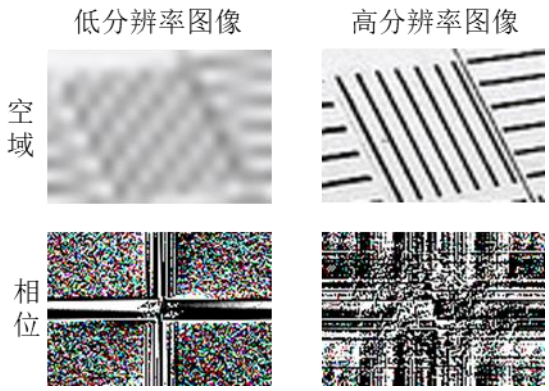


图 5-8 高低分辨率相位谱对比图

从图 5-8 中可知低分辨率图像的相位谱含有大量的噪声，分布于相位谱的四周。而高分辨率图像的相位谱排列规整。在超分辨率重建算法恢复图像时，如果不能恢复正确的相位往往恢复出的图像也是变质的，如下

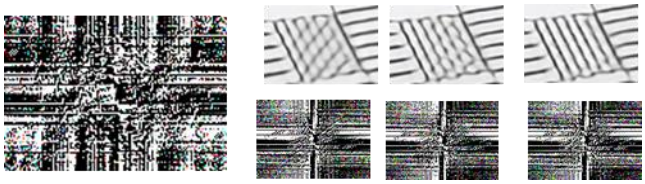


图 5-9 图像恢复相位谱对比图

显然在图 5-9 中因为相位没有正确的匹配，所以应该是直线的区域产生了变质。故此，利用相位谱约束算法的输出，优化算法中的权重使得算法中的特征更好的产生交互，提升超分辨率重建的性能是有必要的。

5.3.2 损失函数构建

傅里叶变换已在前章节介绍过，故此在本节不赘述，首先给出一对傅里叶变换如 $f(x,y)\overset{\mathcal{F}}{\Leftrightarrow}F(u,v)$，其中法$f(x,y)$为离散实信号，$F(u,v)$为$f(x,y)$的傅里叶变换。$F(u,v)$可以分解为幅度谱与相位谱，如公式(5-14)。

$$F(u,v)=\left|F(u,v)\right|e^{j\phi(u,v)} \tag{5-14}$$

式中，$\left|F(u,v)\right|$为幅度谱，$\phi(u,v)$为相位谱。其公式为公式(5-15)和公式(5-16)。

$$\left|F(u,v)\right|=\sqrt{R^2(u,v)+I^2(u,v)} \tag{5-15}$$

$$\phi(u,v)=\arctan(I(u,v)/R(u,v)) \tag{5-16}$$

式中，$R(u,v), I(u,v)$为 $F(u,v)$的实部与虚部。

给出 ℓ_1 损失函数的定义如公式(5-17)。ℓ_1 损失函数是图像超分辨重建中最常用的损失函数，是针对像素的损失函数。超分辨率重建后的图像与高分辨率图像对齐每个像素之间做差得出差值约束算法。

$$\mathcal{L}^{\ell_1} = \frac{1}{N}\sum_{i\in N}\left|I_{\mathrm{SR}}(i) - I_{\mathrm{HR}}(i)\right| \tag{5-17}$$

式中，I_{SR} 代表超分辨率重建后的图像。I_{HR} 代表高分辨率图像。$I_*(i)$是图像 I_*的第 i 个像素，*代表 SR 或 HR。N 是图像 I_*像素个数

随后利用 ℓ_1 损失函数构建相位谱损失函数，即结构损失，如公式(5-18)。超分辨率重建后的图像的相位谱与高分辨率图像的相位谱对齐每个点之间做差得出差值约束算法。

$$\mathcal{L}^{\ell_\phi} = \frac{1}{N}\sum_{i\in N}\left|\phi_{\mathrm{SR}}(i) - \phi_{\mathrm{HR}}(i)\right| \tag{5-18}$$

式中，超分辨率重建后的图像的相位谱由 ϕ_{SR} 表示。ϕ_{HR} 表示高分辨率图像的相位谱。$\phi_*(i)$ 是相位谱 ϕ_* 的第 i 个像素，*代表 SR 或 HR。N 是相位谱 ϕ_* 像素个数。

最后，整合两个损失函数如公式(5-19)所示。

$$\mathcal{L}^{\ell_t} = \mathcal{L}^{\ell_1} + \varepsilon\mathcal{L}^{\ell_\phi} \tag{5-19}$$

式中，ε 是一个超参数，用于约束结构损失。在章节 5.4.3 中会对其进行详细的消融实验。

在网络初始化时，一切参数都是随机的所以算法的输出基本上是随机噪声这样导致相位谱差异过大导致网路崩溃。为防止这一情况本文采用从训练开始让 ε 逐渐线性增大。具体方法在章节 5.4.2 中叙述。

5.4 实验结果与分析

5.4.1 实验设置

参照文献[57]，训练图像由来自 Flickr2K 的 2650 张图像和来自 DIV2K 的 800 张图像组成。为了评估不同 SR 网络的性能，我们使用了 Set5、Set14、B100、Urban100 和 Manga109 这 5 个标准基准数据集。平均峰值信噪比(PSNR)和 Y 通道(即亮度)上的结构相似性(SSIM)被用作评估指标。在计算复杂度与内存消耗上

选取，参数量与 FLOPs 与其他先进网络进行对比。

（1）降质模型

使用双三次下采样的方式对图像进行下采样倍数为 2、3、4 的降质。如章节 2.2.1 中，图像降质过程可以由如公式(5-20)表示。

$$I_{\mathrm{LR}}=I_{\mathrm{HR}}\downarrow_{s} \tag{5-20}$$

式中，I_{HR} 代表高分辨率图像，I_{LR} 低分辨率图像。$\downarrow_{s}$ 代表 s 倍的双三次下采样。

（2）本章算法的实现细节

本章算法由 8 个自蒸馏大核卷积块组成，通道数设置为 64。控制蒸馏特征维度的超参数 p 设定为 0.5。采用随机旋转 90°、180°、270°和水平翻转的数据增强方法。批量大小设置为 64，每个 LR 输入的 patch 大小设置为 48×48。模型由 Adan 优化器进行训练。初始学习率设置为 1×10^{-3}，训练 1×10^{6} 迭代次数。

（3）损失函数设置

本章提出的结构损失用于优化总 1×10^{6} 迭代的模型。结构损失函数如公式(5-21)所示。

$$\mathcal{L}^{\ell_t}=\frac{1}{N}\sum_{i\in N}\left(\left|I_{\mathrm{SR}}(i)-I_{\mathrm{HR}}(i)\right|+\varepsilon\left|\phi_{\mathrm{SR}}(i)-\phi_{\mathrm{HR}}(i)\right|\right) \tag{5-21}$$

式中，I_{SR} 为超分辨率重建后的图像，I_{HR} 为高分辨图像。$I_{*}(i)$表示图像 I_{*}的第 i 个像素，其中*代表 HR 或 SR。N 为图像的像素个数。超分辨率重建后的图像的相位谱由 ϕ_{SR} 表示。ϕ_{HR} 表示高分辨率图像的相位谱。$\phi_{*}(i)$ 是相位谱 ϕ_{*} 的第 i 个像素，*代表 SR 或 HR。N 是图像 I_{*}和相位谱 ϕ_{*} 像素个数。

本章针对 ε 设置为从第 0 个迭代次数到第 1×10^{3} 个迭代次数，ε 从 0 线性增长至 0.02。

（4）实验环境

GPU 是 GeForce RTX 3060 GPU 上，使用 Pytorch 实现本章算法。CPU 是第 12 代的英特尔酷睿 i5-12490F。代码和数据集保存在固态磁盘 DS42HKVS-78BFKYO 上。操作系统为 windows10。

5.4.2 实验结果

在本节将本章算法与已有的轻量化超分辨重建算法(IMDN[46], RFDN[47],

ShuffleMixer[105], BSRN[57], VapSR[108])进行对比。BSRN 使用特征蒸馏链接与注意力机制构建网络，VapSR 使用大核卷积注意力构建网络。通过与他们进行对比证明了本章算法的有效性。上述先进算法均使用 ℓ_1 损失函数训练网络，为进一步证明本章提出的结构损失函数的有效性将重新训练 BSRN 与 VapSR 其结果将在章节 5.4.3 进一步分析与讨论。其对比结果如表 5-1 所示，其中 FLOPs 在高分辨率图像为 1280 × 720 测得，即上采样倍数为 2、3、4 所对应的低分辨率图像输入大小为 640 × 360、430 × 240、320 × 180。

表 5-1 本章算法与先进算法比较表

方法	上采样倍数	参数量(k)	FLOPs(G)	set5 PSNR/SSIMM	Set14 PSNR/SSIM	B100 PSNR/SSIM	Urban100 PSNR/SSIM	Manga109 PSNR/SSIM
Bicubic		-	-	33.66/0.9299	30.24/0.8688	29.56/0.8431	26.88/0.8403	30.80/0.9339
IMDN[46]		694	158.8	38.00/0.9605	33.63/0.9177	32.19/0.8996	32.17/0.9283	38.88/0.9774
RFDN[47]		534	95	38.05/0.9606	33.68/0.9184	32.16/0.8994	32.12/0.9278	38.88/0.9773
ShuffleMixer[105]	×2	394	91	38.01/0.9606	33.63/0.9180	32.17/0.8995	31.890/0.9257	38.83/0.9774
BSRN[57]		332	73	38.10/0.9610	33.74/0.9193	32.24/0.9006	32.34/0.9303	39.14/0.9782
VapSR[108]		329	74	38.08/0.9612	33.77/0.9195	32.27/0.9011	32.45/0.9316	-/-
本章算法		288	66.7	38.11/0.9609	33.86/0.9198	32.26/0.9008	32.49/0.9315	39.18/0.9782
Bicubic		-	-	30.39/0.8682	27.55/0.7742	27.21/0.7385	24.46/0.7349	26.95/0.8556
IMDN[46]		703	71.5	34.36/0.9270	30.32/0.8417	29.09/0.8046	28.17/0.8519	33.61/0.9445
RFDN[47]		541	42.2	34.41/0.9273	30.34/0.8420	29.09/0.8050	28.21/0.8525	33.67/0.9449
ShuffleMixer[105]	×3	415	43	34.40/0.9272	30.37/0.8423	29.12/0.8051	28.08/0.8498	33.69/0.9448
BSRN[57]		340	33.3	34.46/0.9277	30.47/0.8449	29.18/0.8068	28.39/0.8567	34.05/0.9471
VapSR[108]		337	33.6	34.52/0.9284	30.53/0.8452	29.19/0.8077	28.43/0.8583	-/-
本章算法		296	30.5	34.57/0.9284	30.49/0.8444	29.19/0.8071	28.47/0.8586	34.08/0.9474
Bicubic		-	-	28.42/0.8104	26.00/0.7027	25.96/0.6675	23.14/0.6577	24.89/0.7866
IMDN[46]		715	41	32.21/0.8948	28.58/0.7811	27.56/0.7353	26.04/0.7838	30.45/0.9075
RFDN[47]		550	23.9	32.24/0.8952	28.61/0.7819	27.57/0.7360	26.11/0.7858	30.58/0.9089
ShuffleMixer[105]	×4	411	28	32.21/0.8953	28.66/0.7827	27.61/0.7366	26.08/0.7835	30.65/0.9093
BSRN[57]		352	19.4	32.35/0.8966	28.73/0.7847	27.65/0.7387	26.27/0.7908	30.84/0.9123
VapSR[108]		342	19.5	32.38/0.8978	28.77/0.7852	27.68/0.7398	26.35/0.7941	-/-
本章算法		308	17.9	32.41/0.8976	28.78/0.7853	27.68/0.7392	26.37/0.7946	30.98/0.9139

从表 5-1 可以看出本章算法在参数量计算量上均取得了最优，在所有上采样倍数上，大部分数据集都取得了最佳结果。

（1）参数量分析

在上采样倍数为 4 时，与蒸馏链接的 BSRN 网络进行对比本章算法参数量下降了 44K 计算量下降了 1.5G。这主要是因为本章算法所采用的自蒸馏链接大大降低了参数量，所使用的大核卷积的计算复杂度要低于 BSRN 中使用的注意力机制。与使用大核卷积的 VapSR 网络进行对比本章算法参数量下降了 34K 计算量下降了 1.6G，这主要是因为本章算法所使用的大核卷积块较于 VapSR 的大核卷积注意力机制的计算量小，而且使用自蒸馏链接进一步降低参数量。

（2）客观指标 PSNR 与 SSIM 分析

本章算法较于 BSRN 在上采样倍数为 4 使，在数据集 Urban100 上 PSNR 与 SSIM 分别提升了 0.1dB 与 0.0038。大核卷积块提供的全局特征交互极大地提升算法的超分辨率重建性能，结构损失进一步约束网络优化网络的参数使得特征更好的交互恢复更正确的高分辨率图像，从而使得算法的超分辨率重建性能进一步提高。

为进一步可视化本章算法与先进算法的比较，并且与第一章算法、第二章算法-T 与第三章算法-L 进行比较，选取上采样倍数为 4 的 Set5 数据集计算 PSNR，同 FLOPs 与参数量汇总在图中比较，如图 5-10 所示。

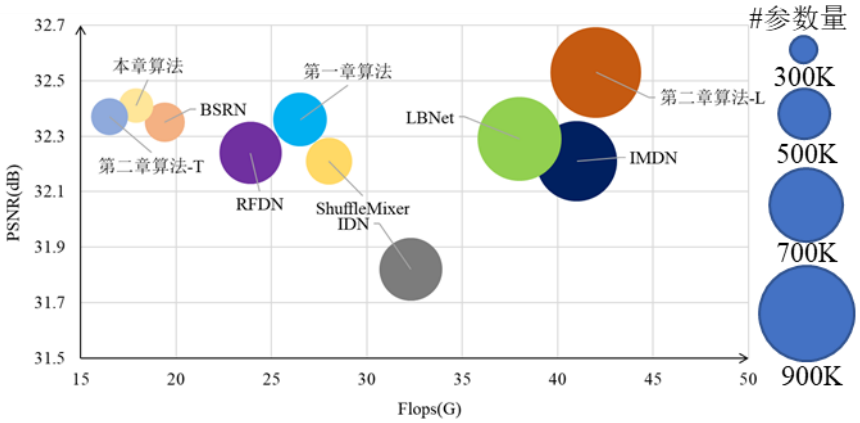


图 5-10 本章算法与先进算法比较可视化图

由图 5-10 所示，本章算法较于第一章算法在参数量与 FLOPs 上有着明显的下降，且在超分辨率重建性能上也有相应的提升。本章算法与第二章算法-T 相比虽然 FLOPs 略高，但是本章算法的超分辨率重建性能要优于第二章算法-T。本章算法与第二章算法-L 相比虽然超分辨率重建性能略差，但是 FLOPs 和参数量仅有第二章算法-L 的三分之一。综上所述，本章算法相较于前两章算法，实现了更优的超分辨率重建性能与计算复杂度之间的平衡。

本章算法与其他先进算法的超分辨率重建的视觉效果，本文选取 Urban100 中的 img006，img011 和 img096 与其他先进算法进行重建效果对比，如图 5-11 所示。

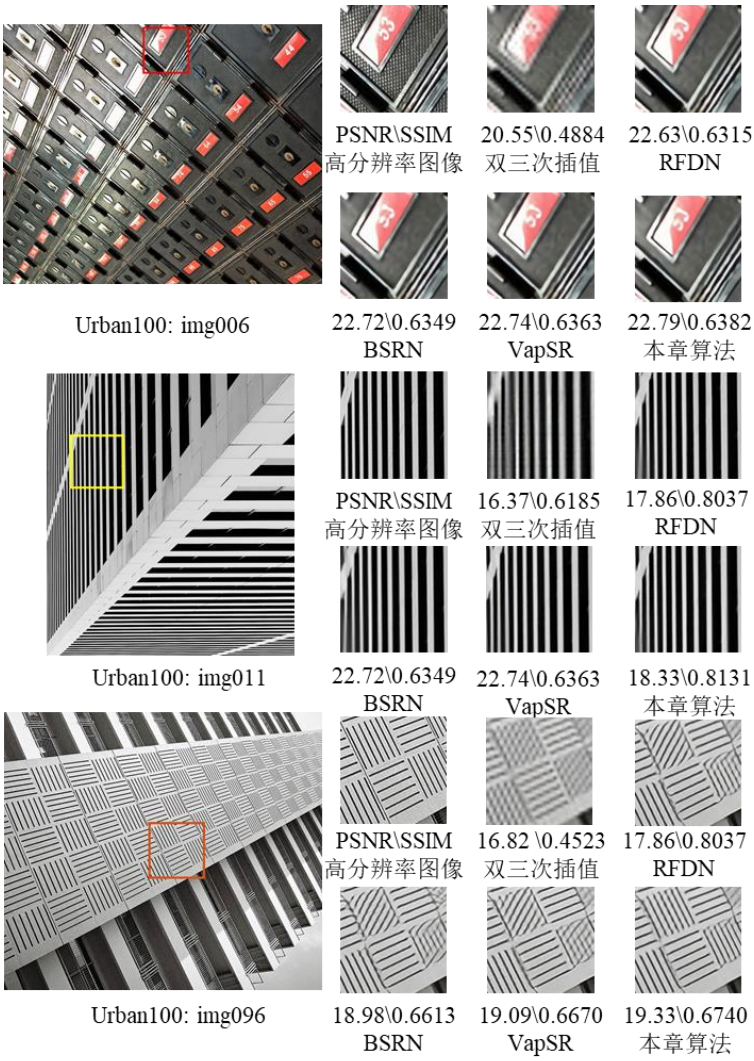


图 5-11 本章算法与其他先进算法视觉效果对比图

从图 5-11 可以看出本章算法在图像细节上恢复的质量更贴近与高分辨率图像，实现了最优的视觉效果。而且因为结构损失函数，本章算法超分辨率重建的图像与高分辨率图像有着更类似的结构，更加贴近高分辨率图像。

5.4.3 算法分析

章节首先对算法的网络结构进行分析，首相对网络结构中层归一化的位置进行消融实验，其实验结果如表 5-2 与图 5-12 所示，其中模型 1 把归一化层放置于自蒸馏大核卷积块最前端，模型 2 把归一化层放置于自蒸馏大核卷积块最后端，模型 3 把归一化层放置于大核卷积块前。上述模型在原训练集上重新训练，表 5-2 采用 Set5、Set14、B100、Urban100、Manga109 计算 PSNR 作为评价指标。图 5-12 采用 Set5 作为验证集绘制 PSNR 曲线。

表 5-2 层归一化消融结果表

方法	Set5	Set14	B100	Urban100	Manga109
模型 2	32.35	28.70	27.64	26.29	30.93
模型 3	32.41	28.78	27.68	26.37	30.98

如表 5-2 所示，在采用归一化层放置于大核卷积块前，即模型 3 时，算法性能最佳。因为模型 1 在训练过程中不稳定具体如图 5-12 所示，所以并没有完整训练。故，表中仅给出模型 2 与模型 3 训练完成后测得的 PSNR 值。

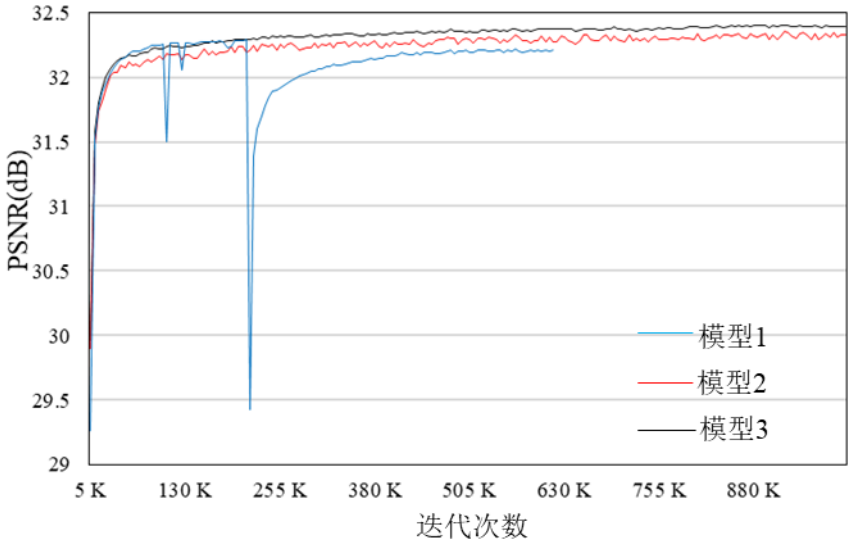
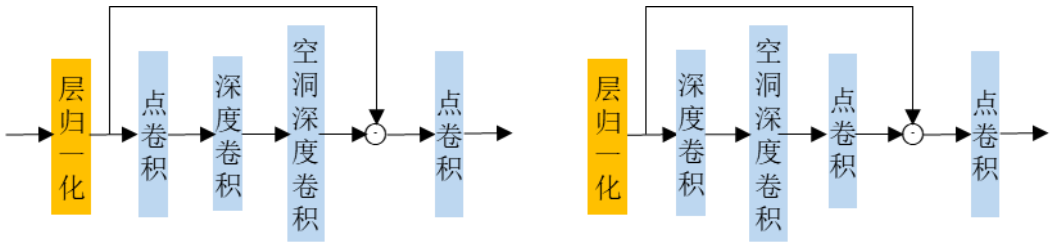


图 5-12 层归一化消融结果 PSNR 曲线图

从图 5-12 可以看出，模型 1，即归一化层放置于自蒸馏大核卷积块最前端时，网络的训练不稳定，所以不能采用这种网络结构训练网络。模型 2，即归一化层放置于自蒸馏大核卷积块最后端虽然完整训练 PSNR 曲线有些许抖动不能一直稳定提升。而模型 3，即归一化层放置于大核卷积块前，PSNR 曲线平滑且能稳定提升。所以归一化层放置于大核卷积块前可以有效的稳定网络训练且实现最佳结果。

接下来对大核卷积块进行分析，设立三个模型，模型 1 不使用大核卷积块，模型 2 使用点卷积在前的大核卷积块如图 5-13(a)所示，模型三采用点卷积在后的大核卷积块如图 5-13(b)所示。上述模型在原训练集上重新训练，表 5-3 采用 Set5、Set14、B100、Urban100、Manga109 上计算 PSNR 与 SSIM 为指标。图 5-14 采用 Set5 为验证集绘制 PSNR 曲线。



(a) (b)

图 5-13 大核卷积块消融结构图

如图 5-13 所示，本文对大核卷积块中点卷积的位置进行改变观察算法超分辨重建性能。首先从理论上进行分析点卷积位置对网络带来的影响。首先如图 5-13(a)点卷积置于大核卷积块前。根据图 5-6 可知大核卷积块之前已经有一个点卷积混洗自蒸馏特征，随后通过层归一化在进入点卷积，该过程的大核卷积块由公式(5-22)和公式(5-23)所示。

$$F = H_{\mathrm{PWConv}}(H_{\mathrm{LN}}(H_{\mathrm{PWConv}}(F_{\mathrm{in}}))) \tag{5-22}$$

$$F_{\mathrm{out}} = H_{\mathrm{PWConv}}(H_{\mathrm{DDWConv}}(F) \odot F_{\mathrm{ln}}) \tag{5-23}$$

式中，H_{PWConv} 为点卷积，H_{LN} 为归一化层。为便于叙述将两个深度卷积简记为 H_{DDWConv}。F_{ln} 为输入特征进行层归一化之后的特征。F_{out} 为输出特征。

从公式(5-22)中点卷积、归一化层均为在通道维度上的线性操作。故，上述公式可以归结为公式(5-24)。

$$F = H_{\mathrm{PLP}}(F_{\mathrm{in}}) \tag{5-24}$$

式中 H_{PLP} 为线性化之后的点卷积与层归一化。

综上，在归一化层之前的点卷积完全可以代替层归一化后的点卷积。而且公式(5-23)中特征 F 通过两个深度卷积进行像素层面交互后紧接着与 F_{ln} 进行哈达玛积，哈达玛积也为像素层面交互，最后进行一次通道上的交互。这样的设计使得特征交互的方式非常紧凑不利于发掘特征中的信息。反观图 5-13(b)中的大核卷积块结构，其可以归结为公式(5-25)和公式(5-26)

$$F = H_{\mathrm{DDWConv}}(H_{\mathrm{LN}}(H_{\mathrm{PWConv}}(F_{\mathrm{in}}))) \tag{5-25}$$

$$F_{\mathrm{out}} = H_{\mathrm{PWConv}}(H_{\mathrm{PWConv}}(F) \odot F_{\mathrm{ln}}) \tag{5-26}$$

从上述公式中明显可以看出，输入特征首先径向维度通道交互，随后经过两个深度卷积像素层面交互，最后点卷积混洗深度卷积的输出进行通道上的交互后

再与 $F_{\ln}$ 进行哈达玛积，即像素层面交互。哈达玛积之后的点卷积再次混洗特征完成通道上的交互。这样的设计特征的交互方式是穿插进行更有利于发掘特征中的信息。进一步通过实验，可以证明上述理论。

表 5-3 大核卷积块消融结果表

方法	参数量(k)	FLOPs(G)	set5 PSNR/SSIMM	Set14 PSNR/SSIM	B100 PSNR/SSIM	Urban100 PSNR/SSIM	Manga109 PSNR/SSIM
模型 1	214	12.5	32.13/0.8937	28.57/0.7811	27.56/0.7355	25.98/0.7823	30.40/0.9339
模型 2	308	17.9	32.35/0.8964	28.76/0.7849	27.65/0.7384	26.34/0.7942	30.91/0.9130
模型 3	308	17.9	32.41/0.8976	28.78/0.7853	27.68/0.7392	26.37/0.7946	30.98/0.9139

从表 5-3 中可以看出没有大核卷积块时，因为算法缺少全局特征交互所以性能有极大的下降。模型 2 与模型 3 虽然有着相同的参数量与计算量，但是在性能上例如 Set5 数据集模型 3 比模型 2 的 PSNR 要高出 0.06。这说明大核卷积块应该按照图 5-13(b)的方式所设计。

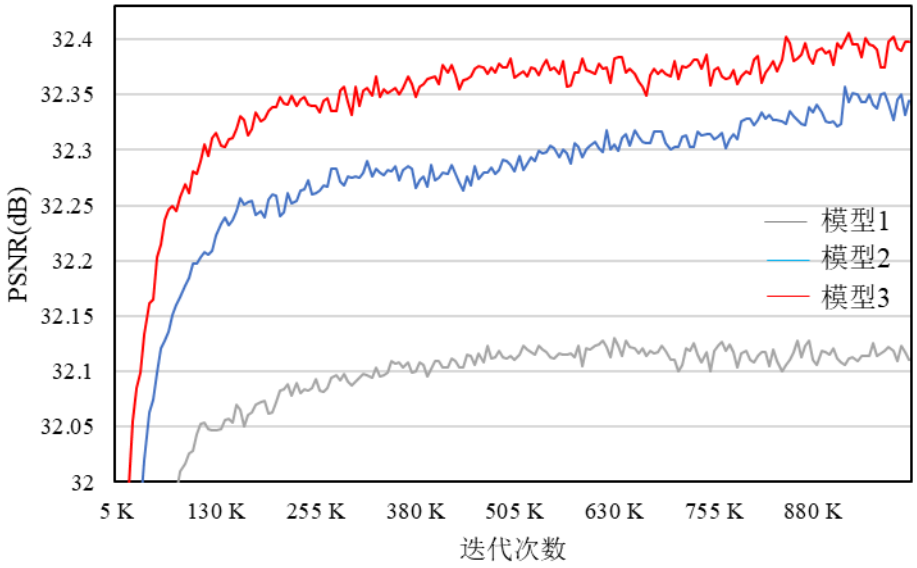


图 5-14 大核卷积块消融结果 PSNR 曲线图

从图 5-14 中可以看出模型 2 的 PSNR 曲线呈跳跃式上升这说明，大核卷积块发掘特征中的信息速度较慢，而模型 3 的 PSNR 曲线呈缓步上升，这说明大核卷积块发掘特征中的信息速度较快，特征得到良好的交互。所以在设计网络结构时，通道维度与像素维度的特征交互应该穿插设计。

对于网络设计的消融实验探讨完毕，接下来将探讨损失函数对算法带来的影响。首先对损失函数中超参数 ε 进行消融实验，分别设置 5×10^{-4}、1×10^{-3}、2×10^{-3} 和 5×10^{-3}。其消融结果如表 5-4 与图 5-15 所示。表 5-4 中重新设定的 ε 的值，均在原始数据集训练相同轮次，特别的 ε ＝0 时，结构损失等同于 ℓ_1 损失函数。

采用 Set5 数据集计算 PSNR 评价指标，并计算损失最后 100 个迭代次数的平均值记录于表中。

表 5-4 ε 取值消融结果表

ε 取值	PSNR(dB)	损失均值
0	32.31	0.02433
5×10^{-4}	32.32	0.02456
1×10^{-3}	32.35	0.02519
2×10^{-3}	32.41	0.02539
5×10^{-3}	32.38	0.02891

从表 5-4 可以看出当 $\varepsilon = 2\times10^{-3}$ 时，算法的超分辨率重建性能最优。根据损失的均值可以看出，当 $\varepsilon = 5\times10^{-3}$ 时，损失的均值较高网络难以收敛到最优解。其他情况下损失的均值相近差距不大。进一步对结构损失的值进行可视化如图 5-15 所示。

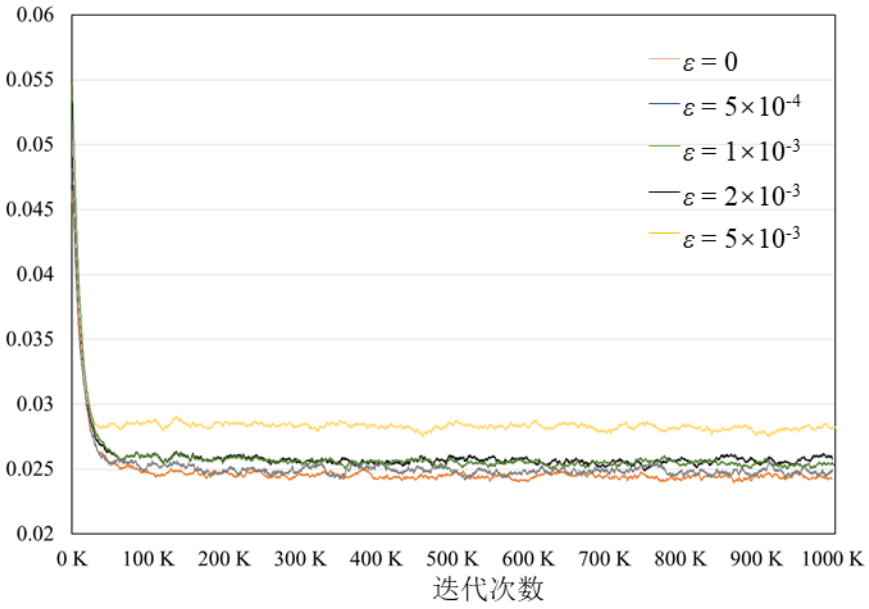


图 5-15 ε 取值结构损失曲线图

从图 5-15 中可以看出，当 $\varepsilon = 5\times10^{-3}$ 时算法的结构损失的曲线在 100K 迭代后明显高于其他取值，算法难以收敛。其他情况下，结构损失曲线相近，算法可以收敛，例如 $\varepsilon = 2\times10^{-3}$ 时虽然曲线略微在 $\varepsilon = 0$ 的上方，但是算法的超分辨率重建效果要更好。

为进一步验证结构损失的有效性，本在 BSRN 与 VapSR 采用结构损失对其在原训练数据集上重新训练，并采用 Set5、Set14、B100、Urban100、Manga109 计算 PSNR 作为评价指标。其训练结果如表 5-5 和视觉效果如图 5-16。

表 5-5 结构损失定量比较表

Method	上采样倍数	set5	Set14	B100	Urban100	Manga109
		PSNR/SSIMM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
BSRN	×4	32.35/0.8966	28.73/0.7847	27.65/0.7387	26.27/0.7908	30.84/0.9123
VapSR		32.38/0.8978	28.77/0.7852	27.68/0.7398	26.35/0.7941	-/-
BSRN-N		32.40/0.8972	28.73/0.7845	27.66/0.7386	26.35/0.7928	30.87/0.9126
VapSR-N		32.43/0.8978	28.80/0.7858	27.69/0.7400	26.40/0.7956	30.96/0.9139

表 5-5 中 BSRN-N 与 VapSR-N 分别代表用结构损失训练的算法。与原始算法相比，采用结构损失的算法在全体数据集上的 PSNR 与 SSIM 均有提升，例如在 Urban100 数据集上 BSRN 采用结构损失训练后 PSNR 提升 0.08dB，VapSR 的 PSNR 提升 0.05dB。进一步选取 Urban100 数据集中的 img096 对比采用结构损失后训练的算法与原始算法超分辨率重建图像的视觉效果图，如图 5-16。

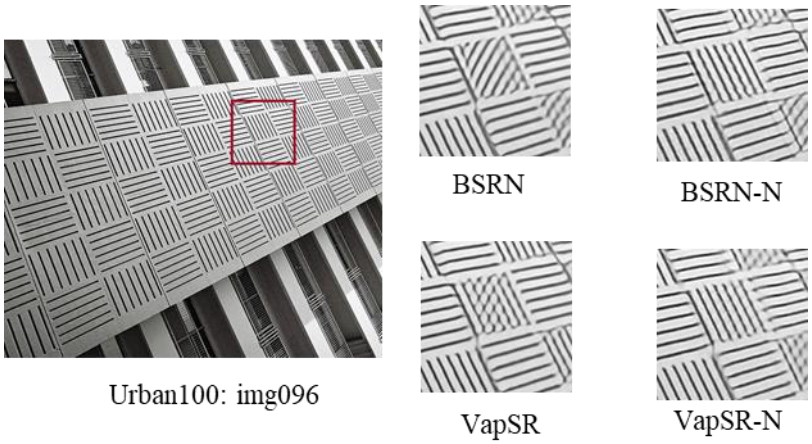


图 5-16 结构损失视觉效果对比图

从图 5-16 可以明显看出在采用结构损失训练之后，图像的视觉效果有明显的提升。这进一步证明本章节提出的结构损失的优越性，能够更良好的约束算法，使网络获得更好的参数，实现更优质的特征交互。

5.5 本章小结

本章提出了基于特征蒸馏与结构损失的单图像超分辨率重建算法。首先，通过拆解 Transform 架构将其分解为全局特征的交互和通道维度的交互，并以此为指导优化 CNN 架构。通过引入 Transform 架构中的层归一化，进一步稳定 CNN 架构的训练，同时设计了不同的层归一化的位置构建网络，找到最合适的位置放置层归一化实现算法。其次在网络中，设计自蒸馏链接实现长远距离的特征通道交互，与传统的蒸馏链接相比，自蒸馏链接具有更低的参数量与计算复杂度。提

出大核卷积块实现全局的特征交互，并设计了不同的大核卷积块对比超分辨率重建性能，选择最优的大核卷积块的设计方式构建网络实现算法。最后，利用结构损失函数约束算法的输出，优化算法中的权重使得算法中的特征更好的产生交互，提升算法的超分辨率重建性能且能更准确的恢复高分辨率图像。使用结构损失重新训练先进算法 VapSR 之后得到的 VapSR-N 在 Urban100 数据集上 PSNR 与 SSIM 提升 0.05dB 和 0.0015。在上采样倍数为 4 时，本章算法较于先进算法 BSRN 相比，参数量和计算复杂度分别下降 32K 与 1.5G，在五个基准测试数据集上 PSNR 与 SSIM 提升的均值分别为 0.076dB 和 0.0015。本章所提出的算法与现有的先进算法相比以最小的参数量与计算复杂度实现了最优的超分辨率重建性能，而且本章算法能更准确地恢复高分辨率图像，超分辨重建后的图像的视觉效果要优于现有的先进算法。

第6章 总结与展望

6.1 工作总结

本文围绕图像超分辨率重建这一科学问题，提出了针对轻量化图像超分辨率重建提出了不同的算法，详细的分析现有 CNN 与 Transform 网络中轻量化网络的不足并加以修改实现更优的算法性能与算法计算复杂度之间的平衡。本文的研究工作总结如下：

（1）详细探讨了图像超分辨率重建领域的成果，并介绍了相关理论基础，分析并总结了轻量化超分辨率重建算法的研究现状。

（2）提出了基于高阶门控交互的单图像超分辨率重建算法，设计了残差全局自适应滤波块，其能给予特征中的高频分量更高的权重，进一步挖掘特征的高频信息，解决低分辨率图像纹理细节丢失较多的问题。同时，使用递归门控卷积代替多头自注意力，降低 Transform 的计算复杂度，使得 Transform 网络应用于轻量化超分辨率重建算法中。

（3）提出了基于迭代高阶门控交互的单图像超分辨率重建算法，设计迭代广域门控卷积，提高网络的感受野。同时解决递归门控卷积在阶数较低的特征交互过程中，特征维度过窄的问题。提出频域大核卷积改善了网络对输入的敏感，使得网络对输入特征的大小更具有自适应性，使得深度特征能够充分地实现高低频交互。提出广域门控卷积空间注意力和增强混合通道注意力，进一步增强网络的超分辨率重建性能，同时使用增强混合通道注意力代替 Transform 中的前馈神经网络层减少网络的计算复杂度。最后，使用残差控制单元，调整残差特征系数，使得网络能够更快速的收敛且提升网络的超分辨率重建性能。

（4）提出基于特征蒸馏与结构损失的单图像超分辨率重建算法，提出自蒸馏链接降低特征蒸馏链接的计算复杂度，提高网络的计算效率，提升网络对长远距离的特征通道混洗能力。针对像素的全局交互提出大核卷积块，进一步提升算法的超分辨率重建性能，同时引入 Transform 中的层归一化稳定网络训练。结构损失函数用于约束算法优化网络，解决网络恢复的高分辨率图像的图像纹理细节出现错误的问题，进一步提升算法的超分辨率重建视觉效果。

6.2 今后工作展望

虽然本文取得了一定的研究成果，针对轻量化超分辨率重建算法设计了不同的网络结构，但是在轻量化超分辨率重建中还有诸多问题有待解决，主要包括以下几个方面：

（1）针对真实退化下的轻量化超分辨率重建，本文设计的轻量化超分辨率重建算法都是针对双三次下采样降质的情况下进行超分辨率重建。但是对于真实世界的图像降质来说，仅仅是双三次下采样降质是不足以应对的。但是针对真实世界退化设计的超分辨率重建算法，无论是在模型大小，还是算法的计算复杂度都过于庞大，所以针对真实世界退化设计的超分辨率重建算法进行轻量化，使得其更高效是十分有必要的。

（2）边缘设备下的实时超分辨率重建，虽然轻量化超分辨率重建算法已取得重大进展，但是在边缘设备上例如：车载设备，监控摄像，AR，VR 等边缘设备上，难以做到实时且高效。所以进一步研究轻量化超分辨率重建如何在边缘设备上的部署方式，以及如何提高网络结构运行效率是十分有必要的。

参考文献

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks[J], Communications of the ACM, 2017(60): 84-90.

[2] Guo M, Lu C, Hou, Q, et al. SegNeXt: Rethinking convolutional attention design for semantic segmentation[C]. In Proceedings of Conference and Workshop on Neural Information Processing Systems, 2022: 1140-1156.

[3] Shermeyer J, Van Etten A. The effects of super-resolution on object detection performance in satellite imagery[C], In Proceedings of Conference on Computer Vision and Pattern Recognition Workshop, 2019: 0-10.

[4] Shi F, Cheng J, Wang L, et al. LRTV: MR image super-resolution with lowrank and total variation regularizations[J]. IEEE Transactions on Medical Imaging, 2015, 34(12): 2459-2466.

[5] 徐军, 刘慧, 郭强, 等. 结合反卷积的 CT 图像超分辨重建网络[J]. 计算机辅助设计与图形学学报, 2018, 30(11): 2084-2092.

[6] Zhang Y, Li K, Li K, et al. MR image super-resolution with squeeze and excitation reasoning attention network[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2021: 13425-13434.

[7] Lei S, Shi Z, Zou Z. Super-resolution for remote sensing images via local-global combined network[J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(8): 1243-1247.

[8] Lei S, Shi Z, Zou Z. Coupled adversarial training for remote sensing image super-resolution[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 58(5): 3633-3643.

[9] Zhou K, Xiao Y, Yang J, et al. Encoding structure-texture relation with P-Net for anomaly detection in retinal images[C]. In Proceedings of European Conference on Computer Vision, 2020: 360-377.

[10] Zavrtanik V, Kristan M, Skoˇcaj D. Draem-a discriminatively trained reconstruction embedding for surface anomaly detection[C]. In Proceedings of International Conference on Computer Vision, 2021: 8330- 8339.

[11] Schl¨uter H, Tan J, Hou B, et al. Natural synthetic anomalies for self-supervised anomaly detection and localization[C]. In Proceedings of European Conference on Computer Vision, 2022: 474-489.

[12] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2016: 1132-1140.

[13] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2017: 230-242.

[14] Gupta D, Arya D, and Gavves E. Rotation equivariant siamese networks for tracking[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2021: 360-377.

[15] Yan B, Peng H, Wu K, et al. LightTrack: Finding lightweight neural networks for object tracking via one-shot architecture search[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2021: 454-469.

[16] Marriott R, Romdhani S, Chen L. A 3D GAN for improved large-pose facial recognition[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2021: 13445-13455.

[17] Fu C, Wu X, Hu Y, et al. DVG-Face: Dual variational generation for heterogeneous face recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 21 (4): 2938-2952.

[18] Dong C, Loy C C, He K, et al. Learning a deep convolutional network for image super-resolution[C]. In Proceedings of European Conference on Computer Vision, 2014: 184-199.

[19] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2016: 1646-1654.

[20] Shi W, Jose C, Ferenc H, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2016: 1874-1883.

[21] Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition Workshop, 2017: 1132-1140.

[22] Zhang Y, Tian Y, Kong Y, et al. Residual dense network for image super-resolution[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2018: 2472-2481.

[23] Li Z, Yang J, Liu Z, et al. Feedback network for image super-resolution[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2018: 3867-3876.

[24] Zhang Y, Li K, Li K, et al. Image super-resolution using very deep residual channel attention networks[C]. In Proceedings of European Conference on Computer Vision, 2018: 1646-1654.

[25] Dai T, Cai J, Zhang Y, et al. Second-order attention network for single image super-resolution[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2019: 1132-1142.

[26] Zhang K, Zuo W, Zhang L. Learning a single convolutional super-resolution network for multiple degradations[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2017: 3262-3271.

[27] Zhang K, Gool L V, Timofte R. Deep unfolding network for image super-resolution[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2020: 3214-3223.

[28] Gu J, Lu H, Zuo W, et al. Blind super-resolution with iterative kernel correction[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2019: 1604-1613.

[29] Ji X, Yun C, Ying T, et al. Real-world super-resolution via kernel estimation and noise injection[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition Workshop, 2020: 1914-1923.

[30] Yuan Y, Liu S, Zhang J, et al. Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition Workshop 2018: 701-710.

[31] Wang, L, Wang Y, Dong Xi, et al. Unsupervised degradation representation learning for blind super-resolution[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2021: 10581-10590.

[32] Wang X, Ma J and Jiang J. Contrastive learning for blind super-resolution via a distortion-specific network[J]. In IEEE/CAA Journal of Automatica Sinica 2024, 10(1): 78-89

[33] Xie Q, Zhou M, Zhao Q, et al. Multispectral and hyperspectral image fusion by MS/HS fusion net[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2019: 1585-1594.

[34] Zheng K, Gao L, Liao W, et al. Coupled convolutional neural network with adaptive response function learning for unsupervised hyperspectral super-resolution[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59 (3): 2487-2502.

[35] Hou J, Zhu Z, Zeng H, et al. Deep posterior distribution-based embedding for hyperspectral image super-resolution[J]. In IEEE Transactions on Image Processing. 2022, 31, pp. 5720-5732

[36] Zhang X, Song C, You T, et al. Dual ODE: Spatial–spectral neural ordinary differential equations for hyperspectral image super-resolution[J]. In IEEE Transactions on Geoscience and Remote Sensing. 2024, 62: 1-15,

[37] Jeon D S, Baek S-H, Choi I, et al. Enhancing the spatial resolution of stereo images using a parallax prior[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2018: 1721-1730.

[38] Dai Q, Li J, Yi Q, et al. Feedback network for mutually boosted stereo image super-resolution and disparity estimation[C]. In Proceedings of the 29th ACM International Conference on Multimedia, 2021: 1323-1345.

[39] Zhang S, Wei Y, Feng J, et al. Stereo image restoration via attention-guided correspondence learning[J]. In IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 1: 1-17.

[40] Fu Y, Chen J, Zhang T, et al. Residual scale attention network for arbitrary scale image super-resolution[J]. Neurocomputing, 2021, 427: 201-211.

[41] Wang, L, Wang Y, Lin Z, et al. Learning a single network for scale-arbitrary super-

resolution[C]. In Proceedings of International Conference on Computer Vision, 2021: 4781-4790.

[42] Chen Y, Liu S, Wang X. Learning continuous image representation with local implicit image function[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2021: 8624-8634.

[43] Yu W, Li Z, Liu Q, et al. Scale-aware frequency attention network for super-resolution[J]. Neurocomputing, 2023, 554: 126584-126601

[44] Dong C, Chen C, Tang X. Accelerating the super-resolution convolutional neural network[C]. In Proceedings of European Conference on Computer Vision, 2016: 391-407.

[45] Hui Z, Wang X, Gao X. Fast and accurate single image super-resolution via information distillation network[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2018: 723-731.

[46] Hui Z, Gao X, Yang Y, et al. Lightweight image super-resolution with information multi-distillation network[C]. In Proceedings of the 27th ACM International Conference on Multimedia, 2019: 2024-2032.

[47] Liu J, Tang J, Wu G. Residual feature distillation network for lightweight image super-resolution[C]. In Proceedings of European Conference on Computer Vision, 2020: 41-55.

[48] Luo X, Xie Y, Zhang Y, et al. LatticeNet: Towards lightweight image super-resolution with lattice block[C]. In Proceedings of European Conference on Computer Vision, 2020: 272-289.

[49] Kim J, Lee J K, Lee K M. Deeply-recursive convolutional network for image super-resolution[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2016: 1637-1645.

[50] Tai Y, Yang J, Liu X. Image super-resolution via deep recursive residual network[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2017: 2790-2798.

[51] Chu X, Zhang B, Ma H, et al. Fast, accurate and lightweight super-resolution with neural architecture search[C]. In Proceedings of International Conference on Pattern

Recognition, 2020: 59-64.

[52] Zhan Z, Gong Y, Zhao P, et al. Achieving on-mobile real-time super-resolution with neural architecture and pruning search[C]. In Proceedings of International Conference on Computer Vision, 2021: 4821-4831.

[53] Lee W, Lee J, Kim D, et al. Learning with privileged information for efficient image super-resolution[C]. In Proceedings of European Conference on Computer Vision, 2020: 465-482.

[54] Wang Y, Lin S, Qu Y, et al. Towards compact single image super-resolution via contrastive self-distillation[C]. In Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, 2021: 1122-1128.

[55] Zhang, X, Hui Z, Li Z. Edge-oriented convolution block for real-time super resolution on mobile devices[C]. In Proceedings of the 29th ACM International Conference on Multimedia, 2021: 4034-4043.

[56] Wang X, Dong C, Shan Y. RepSR: Training efficient VGG-style super-resolution networks with structural re-parameterization and batch normalization[C]. In Proceedings of the 30th ACM International Conference on Multimedia, 2022: 2556-2564.

[57] Li Z, Liu Y, Chen X, et al. Blueprint separable residual network for efficient image super-resolution[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition Workshop, 2022: 833-843.

[58] Liu J, Zhang W, Tang Y, et al. Residual feature aggregation network for image super-resolution[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2020: 2356-2365.

[59] Park K, Soh J W, Cho N I. A dynamic residual self-attention network for lightweight single image super-resolution[J]. In IEEE Transactions on Multimedia, 2023, 25: 907-918.

[60] Gao D and Zhou D. A very lightweight and efficient image super-resolution network[J]. In Expert Systems with Applications Appl, 2022, 213: 298-309.

[61] Zhao H, Kong X, He J, et al. Efficient image super-resolution using pixel attention[C]. In Proceedings of the European Conference on Computer Vision Workshop, 2020:

56-72.

[62] Zou W, Ye T, Zheng W, et al. Self-calibrated efficient transformer for lightweight super-resolution[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition Workshop, 2022: 929-938.

[63] Du Z, Liu D, Liu J, et al. Fast and memory-efficient network towards efficient image super-resolution[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition Workshop, 2022: 852-861.

[64] Liu D, Wang X, Han R, et al. CTE-Net: Contextual texture enhancement network for image super-resolution[J]. In IEEE Transactions on Multimedia, 2024, 1: 1-14.

[65] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: transformers for image recognition at scale[C]. In Proceedings of International Conference on Learning Representations, 2021: 210-231.

[66] Liu Z, Lin Y, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]. In Proceedings of International Conference on Computer Vision, 2021: 9992-10002.

[67] Chen H, Yunhe Wang Y, Tianyu Guo T, et al. Pre-trained image processing transformer[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2021: 12294-12305.

[68] Gao G, Wang Z, Li J, et al. Lightweight bimodal network for single-image super-resolution via symmetric CNN and recursive Transformer[C]. In Proceedings of International Joint Conferences on Artificial Intelligence, 2022: 913-919.

[69] Lu Z, Li J, Liu H, et al. Transformer for single image super-resolution[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition Workshop, 2022: 456-465.

[70] Liang J, Cao J, Sun G, et al. SwinIR: image restoration using swin Transformer[C]. In Proceedings of International Conference on Computer Vision Workshop, 2021: 1833-1844.

[71] Choi H, Lee J, Yang J. N-Gram in swin Transformers for efficient lightweight image super-resolution[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2023: 2071-2081.

[72] Wang H, Chen X, Ni B, et al. Omni aggregation networks for lightweight image super-resolution[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2023: 22378-22387.

[73] Zheng L, Zhu J, Shi J, et al. Efficient mixed Transformer for single image super-resolution[J]. Engineering Applications of Artificial Intelligence, 2024, 133: 108035-108045.

[74] Schoenberg I. Cardinal interpolation and spline functions[J]. Journal of Approximation Theory, 1969, 2(2): 167-206.

[75] Olivier R, Hanqiang C. Nearest neighbor value interpolation[J]. International Journal of Advanced Computer Science & Application, 2012, 3(4):25-30.

[76] Hou H, Andrews H. Cubic splines for image interpolation and digital filtering[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1978, 26(6): 508-517.

[77] 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.

[78] Nguyen N, Milanfar P, Golub G. A computationally efficient super-resolution image reconstruction algorithm[J]. IEEE Transactions on Image Processing, 2001, 10(4): 573-583

[79] Liu C, Sun D. On bayesian adaptive video super-resolution[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(2): 346-360.

[80] 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.

[81] Zhang L, Wu X. An edge-guided image interpolation algorithm via directional filtering and data fusion[J]. IEEE Transactions on Image Processing, 2006, 15(8): 2226- 2238.

[82] Chollet F. Xception: Deep learning with depth-wise separable convolutions[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2017: 1800-1807.

[83] Haase D, Amthor M. Rethinking depth-wise separable convolutions: how intra-

kernel correlations lead to improved MobileNets [C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2020: 14600-14609.

[84] Cao J, Li Y, Sun M. DO-Conv: Depth-wise over-parameterized convolutional layer[J]. IEEE Transactions on Image Processing, 2022, 31: 3726-3736.

[85] Dai J, Qi H, Xiong Y, et al. Deformable convolutional networks[C]. In Proceedings of International Conference on Computer Vision, 2017: 764-773.

[86] Qi Y, He Y, Qi X, et al. Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation[C]. In Proceedings of International Conference on Computer Vision, 2023: 6047-6056.

[87] Nair V, Geoffrey E. Rectified linear units improve restricted Boltzmann machines[C]. In Proceedings of International Conference on Machine Learning, 2010: 643-669.

[88] He K, Zhang X, Ren S et al. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification[C]. In Proceedings of International Conference on Computer Vision, 2015: 1026-1034

[89] Han J, Moraga C. The influence of the sigmoid function parameters on the speed of backpropagation learning[J]. From Natural to Artificial Neural Computation, 1995, 930: 195-201.

[90] Marra S, Iachino M, Morabito F. Tanh-like activation function implementation for high-performance digital neural systems[C]. In 2006 Ph.D. Research in Microelectronics and Electronics, 2006: 237-240.

[91] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]. In Proceedings of Conference and Workshop on Neural Information Processing Systems, 2017: 6000-6010.

[92] Bevilacqua M, Roumy A, Guillemot C, et al. Low-complexity single image super-resolution based on nonnegative neighbor embedding[C]. In Proceedings of British Machine Vision Conference, 2012: 1-10.

[93] Zeyde R, Elad M, Protter M. On single image scale-up using sparse-representations[C]. In International Conference on Curves and Surfaces, 2010: 711-730

[94] Martin D, Fowlkes C, Tal D, et al. A database of human segmented natural images

and its application to evaluating segmentation algorithms and measuring ecological statistics[C]. In Proceedings of International Conference on Computer Vision, 2001, 2: 416-423.

[95] Huang J, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2015: 5197-5206.

[96] Matsui Y, Ito K, Aramaki Y, et al. Sketch-based manga retrieval using manga109 dataset[J]. Multimedia Tools and Applications, 2017, 76 (20): 21811-21838.

[97] Agustsson E, Timofte R. NTIRE 2017 challenge on single image super-resolution: dataset and study[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition Workshop, 2017: 1122-113

[98] Timofte R, Agustsson E, Gool L V, et al. NTIRE 2017 challenge on single image super-resolution: methods and results[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition Workshop, 2017: 1110-1121.

[99] Gu S, Lugmayr A, Danelljan M et al. DIV8K: DIVerse 8K resolution image dataset[C]. In Proceedings of International Conference on Computer Vision Workshop, 2019: 3512-3516.

[100] Li Y, Zhang K, Liang J, et al. LSDIR: a large scale dataset for image restoration[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition Workshop 2023: 1775-1787.

[101] Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2009: 248-255.

[102] Liu Z, Mao H, Wu C, et al. A ConvNet for the 2020s[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2022: 11966-11976.

[103] Liu Z, Lin Y, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]. In Proceedings of International Conference on Computer Vision, 2021: 9992-10002.

[104] Ahn N, Kang B, Soh K A. Fast, Accurate, and lightweight super-resolution with cascading residual network[C]. In Proceedings of European Conference on

Computer Vision, 2018: 256-272.

[105] Sun L, Pan J, and Tang J. Shufflemixer: an efficient convnet for image super-resolution[C]. In Proceedings of Conference and Workshop on Neural Information Processing Systems, 2022: 17314-17326.

[106] Kingma D P, Ba J. Adam: a method for stochastic optimization[C]. In Proceedings of International Conference on Learning Representations, 2015: 14-28.

[107] Ding X, Zhang X, Han J, et al. Scaling up your kernels to 31×31: revisiting large kernel design in CNNs[C]. In Proceedings of Conference on Computer Vision and Pattern Recognition, 2022: 11953-11965.

[108] Zhou L, Cai H, Gu J, et al. Efficient image super-resolution using vast-receptive-field attention[C]. In Proceedings of the European Conference on Computer Vision Workshop, 2022: 256–272.

致谢

时光荏苒，三年的研究生生活即将画上句号。站在毕业的门槛上，我感慨万千，心中充满了对过去的回忆和对未来的憧憬。在这段宝贵的时光里，我得到了许多人的关心、支持和帮助，让我在学术道路上不断前行。

首先，我要衷心感谢我的导师王凤随教授。他不仅在学术上给予了我精心的指导，更在人生道路上给予了我宝贵的建议。他严谨的治学态度、深厚的学术造诣和无私的奉献精神，让我深受感染。他的悉心教诲和耐心指导，让我在学术研究中少走了许多弯路，取得了今天的成绩。

我还要感谢实验室的同门。我们一起度过了无数难忘的时光，共同探讨学术问题、分享生活点滴。他们的热情帮助和无私支持，让我感受到了集体的力量和温暖。我们在一起成长、进步，建立了深厚的友谊，这段经历将永远留在我的记忆中。

此外，我还要感谢家人的支持和关爱。他们一直是我坚强的后盾，在我遇到困难和挫折时给予我无尽的鼓励和支持。他们的理解和包容，让我能够专心致志地追求自己的梦想。家人的爱和关怀是我前进的动力，也是我永远的依靠。

最后，我要感谢学校和学院提供的良好学习环境和资源。学校优美的校园环境、丰富的学术资源和先进的实验设施，为我的学习和研究提供了有力保障。学院的培养和教育，让我在学术上不断成长和进步。

回顾这三年的研究生生活，我收获了很多宝贵的经验和知识。感谢所有帮助过我的人，你们的支持和鼓励是我不断前进的动力。在未来的日子里，我将继续努力，不断追求卓越，为实现自己的梦想而奋斗。

作者：

2024 年 5 月 26 日

攻读学位期间取得的学术成果情况

1 发表学术论文情况

[1] Fengsui Wang, Xi Chu, HorSR: High-Order Spatial Interactions and Residual Global Filter for Efficient Image Super-Resolution[J]. Signal Processing: Image Communication, 2024, https://doi.org/10.1016/j.image.2024.117148

[2] 储曦,王凤随.基于部分特征卷积聚合的图像超分辨率重建方法 [J/OL]. 重庆工商大学学报(自然科学版),1-11[2024–05-22].http://kns.cnki.net/kcms/detail/50.1155N.20240411.0851.002.html.

2 参与科研项目情况

（1）安徽省自然科学基金(2108085MF197)

（2）安徽高校省级自然科学研究重点项目(KJ2019A0162)

（3）安徽工程大学国家自然科学基金预研项目(Xjky2022040)