

河北地質大學

畢業設計（論文）

題目：基於深度學習的 OCT 圖像重建與分析

姓 名 郭澤釗

學 號 421109030207

專 業 計算機科學與技術

指導教師 趙占芳 教授

學 院 信息工程學院

2025 年 06 月 01 日

毕业设计任务书

学 院 信息工程学院 专 业 计算机科学与技术

班 级 4211090307 姓 名 郭泽钊

毕业设计起止日期 2025 年 1 月 06 日-2025 年 5 月 16 日

实习地点 河北省石家庄市藁城区河北大道 601 号河北地质大学

毕业设计题目 基于深度学习的 OCT 图像重建与分析

毕业设计内容与要求：

1. 毕业设计内容：

毕业论文要求针对欠采样 OCT 图像数据集设计并实现基于深度学习模型的重建和分析。

2. 毕业论文要求

(1) 阅读相关文献，梳理汇总关于欠采样 OCT 图像重建的任务及其方法，分析当前方法的优缺点。

(2) 深度调研现有一系列深度学习模型（CNN、GAN、Transformer）的优劣，设计一种新的深度学习模型来重建 OCT 图像。

(3) 定量分析本文提出的方法与主流方法的差异，采用 PSNR、SSIM 等常规的图像质量指标。定性分析本文提出的方法与主流方法重建出的图像与原始高分辨率图像的重建情况。采用消融和对比实验评估本文提出方法中模块的有效性。

(4) 提交成果列表

- ①一篇 1000 字以上的文献综述；
- ②一篇 2000 字以上的外文翻译；
- ③毕业论文打印版和电子版各一份；
- ④设计的程序清单一份，含平台开发环境及使用说明书。

指导教师签字 赵洪

2025 年 3 月 14 日

指导教师评语

该生选题具有较好的理论与实际应用价值,工作量饱满,工作量较大,很好完成了毕业设计任务书内容。论文结构合理,行文流畅,图表规范。该生具有较好的独立思考能力,能独立运用所学知识解决实际问题。建议成绩为优秀。

建议成绩:

优 良 中 及格 不及格



指导教师签字 赵江华

2025 年 5 月 21 日

评议教师评语

该生选题符合专业方向,有一定的实用价值。针对OCT图像重建问题给出了可行的方案,设计过程完整,实验数据完备,论证过程严谨,工作量饱满。论文结构完整,章节安排得当。论文格式规范,图表清晰,文献引用全面。语言表述流畅。

评议教师签字 刘建东

2025 年 5 月 23 日

答辩小组评语

答辩过程中表现优秀，能对问题作出回答
及时反应，对问题理解深入。
论文书写工整，符合学术相关规范。

建议成绩：

优/良 中 及格 不及格

答辩小组组长签字 王培孝

2025年 5 月 29 日

答辩委员会评语

终审成绩：

优 良 中 及格 不及格

答辩委员会主任签字_____

年 月 日

原创性声明

本人郑重声明：所呈交的毕业设计（论文）是本人在导师指导下，进行研究工作所取得的成果。除文中已经注明引用的内容外，本设计（论文）不包含任何他人或集体已经发表的作品内容，也不包含本人为获得其他学位而使用过的材料。对本设计（论文）所涉及的研究工作做出贡献的其他个人或集体，均已在文中以明确方式标明。本设计（论文）原创性声明的法律责任由本人承担。

设计（论文）作者签名：

日期：2025 年 5 月 16 日

关于毕业设计（论文）版权使用授权的说明

本人完全了解河北地质大学关于收集、保存、使用设计（论文）的以下规定：学校有权采用影印、缩印、扫描、数字化或其它手段保存设计（论文）；学校有权提供本设计（论文）全文或者部分内容的阅览服务；学校有权将设计（论文）的全部或部分内容编入有关数据库进行检索、交流；学校有权向国家有关部门或者机构送交设计（论文）的复印件和电子版。

（保密的设计（论文）在解密后适用本授权说明）

设计（论文）作者签名：

日期：2025 年 5 月 16 日

导师签名：



日期：2025 年 5 月 16 日

摘 要

光学相干断层扫描 (Optical Coherence Tomography, OCT) 是一种具备高空间分辨率的非侵入式成像技术, 它已经被应用到许多领域, 包括眼科等临床诊断场景。尽管密集采集可以获得高质量图像, 但其所带来的扫描时间增加及患者不适问题, 限制了 OCT 在临床中的高效应用。为提升采集效率, 欠采样策略被广泛采用, 但由此带来的图像分辨率下降亟需通过有效的重建方法加以弥补。此外, 卷积神经网络 (Convolutional Neural Networks, CNNs) 在建模过程中普遍对低频信息更为敏感, 导致其在高频细节恢复方面存在不足。

针对上述问题, 本文提出了一种混合注意力结构保持网络 (Hybrid Attention Structure Preserving Network, HASPN), 对欠采样 OCT 图像进行超分辨率重建。该方法基于非锐化掩码 (Unsharp Masking, USM) 中的高频成分分解设计双分支网络架构, 分别学习图像的低频结构和高频细节。同时采用自适应膨胀卷积通道注意力机制 (Adaptive Dilated Convolution-based Channel Attention, ADCCA) 和增强空间注意力机制 (Enhanced Spatial Attention, ESA), 从而进一步增强网络在通道以及空间维度上的建模能力。实验在公开数据集 OCT2017 上进行, 在 4 倍与 8 倍的放大倍数下, HASPN 在峰值信噪比 (PSNR) 和结构相似性 (SSIM) 指标上均优于现有主流方法。而且模型在多个未见领域的数据上也具有良好的跨域泛化能力, 证明了其在临床 OCT 图像智能重建方面中的应用潜力。

关键词: 光学相干断层扫描; 超分辨率重建; 非锐化掩码; 注意力机制

ABSTRACT

Optical Coherence Tomography (OCT) is a non-invasive imaging technique with high spatial resolution, widely used in clinical diagnostics such as ophthalmology. Although dense sampling enables the acquisition of high-resolution images, it significantly increases scanning time and causes patient discomfort, thereby limiting the clinical efficiency of OCT. To address this issue, under-sampling strategies are commonly adopted to improve acquisition speed. However, under-sampling inevitably leads to a decline in image resolution, which necessitates the development of effective reconstruction methods. In addition, Convolutional Neural Networks (CNNs) inherently show higher sensitivity to low-frequency information, making them less effective in recovering high-frequency image details.

To overcome these challenges, this paper proposes a Hybrid Attention Structure Preserving Network (HASPN) for super-resolution reconstruction of under-sampled OCT images. Inspired by the high-frequency component decomposition in Unsharp Masking (USM), HASPN adopts a dual-branch architecture to learn the low-frequency structures and high-frequency details of images separately. Furthermore, Adaptive Dilated Convolution-based Channel Attention (ADCCA) and Enhanced Spatial Attention (ESA) are integrated to enhance the network's modeling capability on channel and spatial dimensions. Experiments conducted on the publicly available OCT2017 dataset demonstrate that HASPN outperforms existing state-of-the-art methods at $4\times$ and $8\times$ upscaling factors in terms of Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). Moreover, HASPN exhibits strong cross-domain generalization on out-of-distribution datasets, validating its potential for practical application in intelligent clinical OCT image reconstruction.

Keywords: Optical Coherence Tomography; Super-resolution; Unsharp Masking; Attention Mechanism

目 录

摘要	I
ABSTRACT	II
第一章 绪论	1
1.1 研究背景及意义	1
1.2 国内外研究现状	2
1.3 研究的目标和内容	3
1.4 章节安排	4
第二章 相关技术介绍	5
2.1 卷积神经网络（CNN）	5
2.2 非锐化掩码（USM）	6
2.3 注意力机制	7
2.4 变换器（Transformer）	7
2.5 生成对抗网络（GAN）	8
第三章 数据预处理与实验设计	10
3.1 数据集说明	10
3.1.1 数据来源	10
3.1.2 样本构成	10
3.1.3 标注标准	10
3.1.4 伦理声明	11
3.2 预处理流程	11
3.3 评估指标	11
3.3.1 峰值信噪比	11
3.3.2 结构相似性指数	12
3.4 实验细节及环境	12
第四章 模型设计与优化	14
4.1 网络整体架构图	14
4.2 混合注意力机制	15

4.3 高频分解.....	17
4.4 损失函数设计.....	17
第五章 实验与结果分析.....	19
5.1 定性对比.....	18
5.2 定量对比.....	21
5.3 补充实验.....	23
5.3.1 双分支有效性验证	23
5.3.2 模型超参数实验	23
5.3.3 模型消融实验	24
5.3.4 模型泛化实验	25
第六章 模型不足与展望.....	26
6.1 模型不足.....	26
6.2 未来展望.....	26
第七章 总结.....	27
参考文献.....	28
致谢	32

第一章 绪论

本章主要对光学相干层析成像技术的发展过程以及在临床方面的应用现状进行系统梳理,分析数据采集环节存在的主要问题及传统方法的不足之处,从而提出基于深度学习的研究框架,并阐述研究目的及全文结构布局。

1.1 研究背景及意义

光学相干层析成像 (Optical Coherent Tomography, OCT) 是一种基于低相干干涉及非接触、非侵入成像的,分辨率为亚微米级的成像技术,可以提供组织结构的细胞级横向截面图像^[1]。从 20 世纪 90 年代开始应用,就已经在眼科、心血管、皮肤、牙科等多个临床领域,尤其是在视网膜疾病诊断和治疗领域起到了不可替代的作用^[2-5]。与传统的医学影像学检查方法,例如计算机断层扫描 (Computed Tomography, CT) 和磁共振成像 (Magnetic Resonance Imaging, MRI) 等,比较起来, OCT 的优点包括:成像速度快、穿透力强及分辨率高等特点,从而受到医疗图像处理方面的极大关注^[6]。

但是 OCT 图像的采集过程常常会因为硬件的限制以及扫描速度与噪声等因素,很难实现对整个视野内的高密度图像扫描^[7]。特别是在临床上需要对大量的患者进行扫描的情况下,为了提升扫描的速度以及减轻患者的不适感,人们往往选择用欠采样的方式对 OCT 图像进行快速采集。虽然这种方法可以有效提升采集速度,但是却会导致采集到的图像质量变差,出现如边缘模糊,结构丢失以及空间分辨率低等问题从而影响了后面的诊断准确性^[8]。所以如何从这些经过欠采样之后得到的数据中还原高质量的 OCT 图像也就成为了 OCT 图像处理研究中的一个重点问题。

传统的图像重建方法如插值算法 (双线性插值、样条插值等)^[9]、稀疏表示^[10]、压缩感知^[11]等方法可以在一定程度上提高图像的清晰度,但是存在以下问题:需要人为设定先验模型,适应性差;对复杂结构或纹理的处理效果差,容易产生伪影,且在低采样率下重建效果较差;计算复杂度高,不能满足实时重建的需求。随着人工智能技术的快速发展,尤其是深度学习在图像超分辨率、图像去噪、医学图像分割等方面的应用越来越广泛,越来越多的研究者开始尝试使用深度神经网络来实现 OCT 图像重建^[12-14]。

近年来,卷积神经网络 (Convolutional Neural Network, CNN)^[15]、生成对抗网络 (Generative Adversarial Network, GAN)^[16]、变换器 (Transformer)^[17]等深度学习方

法在医学图像领域取得了丰硕的成果。这些方法具有强大的非线性建模能力，可以从大量的样本中学习从欠采样图像到全采样图像的复杂的映射关系。例如基于 U-Net 结构的模型可以很好地保留图像边缘以及局部细节，GAN 类模型可以生成更加逼真的图像纹理。此外，Transformer 结构由于其强大的全局建模能力，近年来也被引入到 OCT 图像重建任务中来捕捉长距离依赖以及空间结构信息。

OCT 图像质量对之后的任务（组织分割、病灶检测等）也非常重要。质量低或失真的重建图像会影响之后任务的准确度，甚至导致误诊^[18]。所以基于深度学习的 OCT 图像重建不仅是图像增强任务，也是医疗辅助诊断系统的重要基础任务。

基于深度学习的欠采样 OCT 图像重建和分析研究具有重要的理论研究意义以及广阔的现实应用前景。基于构建的高效神经网络模型可以极大提高 OCT 图像重建质量，为智能医疗提供更可靠的依据。该研究方向在未来可以实现对临床成像设备的实时部署，使医学影像向高效、智能的方向发展^[19]。

1.2 国内外研究现状

为了提高 OCT 图像的空间分辨率和质量，国内外的学者对欠采样的 OCT 图像进行了各种重建方法的研究。早期大部分是用传统的图像处理方法和稀疏表示理论，近几年来转向了深度学习，取得了很大的进步。

在传统方法上，Fang 等人^[20]提出了一种基于图像稀疏性的重建框架，该框架可以同时执行图像插值和图像去噪，从而实现高效的 OCT 图像重建。Abbasi 等人^[21]提出了一种非局部加权稀疏表示（Non-local Weighted Sparse Representation, NWSR）方法，该方法通过融合多个具有冗余信息的噪声图像补丁的稀疏表示来提高图像的视觉质量。Wang 等人^[22]将压缩感知（Compressed Sensing, CS）技术与数字滤波器相结合，在只保留 70% B-扫描采样点的情况下重建出结构清晰的 OCT 血管造影图像，这表明 CS 技术在加快采集过程方面具有巨大的优势。

随着深度学习的发展，研究者们也开始尝试使用深度学习的方法来对 OCT 图像进行重建，实现更加复杂的非线性映射，获得更好的图像重建效果。国外方面，Huang 等人^[23]首次将 GAN 技术应用于 OCT 图像超分辨率重建中，不仅提高了图像的空间分辨率，而且降低了图像的噪声，开启了深度学习在 OCT 图像重建中的应用趋势。Lee 等人^[24]为了克服扫频源 OCT（Swept-Source OCT）数据进行密集采样时所带来的问题，提出利用深度神经网络进行图像重建的方法，极大地减少了对完整光谱数据的需求，减轻了数据采集的压力。Gao 等人^[25]设计了一种称为 HARNet 的深度学习网络，可以增强 6×6mm 的 OCT 血管造影图像，结果显示该方法在噪声抑制和血管结构连通性恢

复方面具有明显的优势。

在国内，相关研究也在持续进行，Ke 等人^[26]把金字塔增强通道注意力机制（Pyramid Enhanced Channel Attention, PECA）融入到 GAN 结构当中去，这样就使得模型对 OCT 图像细节的恢复能力又得到提升。而 Lu 等人^[27]则探讨了 Transformer 模型在 OCT 图像重建方面的应用情况，他们指出 Transformer 这类模型具备的全局建模能力，在处理长距离依赖信息的时候具有天然的优势，并且也证明了其在临床应用中的可行性和前景。

1.3 研究的目标和内容

本研究希望可以提出一种深度学习驱动的 OCT 图像重建与分析方法，以此来改善欠采样 OCT 图像的空间分辨率以及诊断质量以及改进 OCT 图像重建时高频细节的恢复程度，本文参考了神经网络的频率原则（Frequency Principle, F-Principle），即神经网络在训练之初会更倾向于先学习低频分量，对于高频细节的建模能力相对较弱^[28]。这种特性在 OCT 图像重建任务当中体现得更为明显，高频部分往往包含着很多边缘结构和病理特征，若是不能正确地重建高频部分，那么整个图像的可诊断性就会大打折扣。

为了解决这个问题，本文采用非锐化掩码（Unsharp Masking, USM）的思想对图像进行高频分解，提高神经网络对高频成分的建模能力^[29]。即首先对低分辨率图像以及与之对应的高分辨率标签图像进行平滑滤波（如高斯模糊），得到它们的低频近似图像。再将原始图像与平滑后的图像相减，得到近似的高频成分，分别作为高频分解后的低分辨率图像和高分辨率图像。

网络设计方面，本文采用双分支结构进行重建。主分支输入原始低分辨率图像，主要负责建模图像主要结构信息和整体轮廓，辅助分支输入高频分解的低分辨率图像，主要负责图像边缘细节和局部纹理的重建。两个分支的特征在各自分支编码和上采样后，在融合模块中进行通道拼接和特征融合，最后经过残差结构输出高质量重建图像。

这种结构既可以让网络尽可能利用低频结构信息以保证重建图像的大致稳定，又可以大幅提高网络对于高频细节的感知和建模能力。通过 USM-inspired 的频率解耦和双分支重建机制，我们有效解决了传统网络重建建模高频信息能力弱的问题，为 OCT 图像的超分辨率重建开辟了新的方向。

1.4 章节安排

本论文共分为七章，章节内容安排如下：

第一章，绪论介绍光学相干断层扫描（OCT）技术的临床应用背景和研究意义，对比传统重建方法存在的不足及深度学习技术的优势，概括国内外研究现状，明确本文地研究目标和内容，并概述全文结构。

第二章，系统阐述本文所用到的深度学习模型与关键技术，包括卷积神经网络（CNN）、非锐化掩码（USM）、注意力机制、变换器（Transformer）架构和生成对抗网络（GAN）及其原理与适用性。

第三章，详细说明实验数据集的构成、来源与预处理过程，定义图像质量评估指标（PSNR、SSIM），描述实验环境配置与训练参数设置。

第四章，提出混合注意力结构保持网络（HASPNN），阐述其双分支结构设计原理、混合注意力机制（ADCCA 和 ESA）、高频分解策略以及混合损失函数，分析各个模块的优化思路。

第五章，通过视觉效果做定性对比和用 PSNR、SSIM 等指标进行定量对比验证模型性能，开展消融实验、超参数实验及跨域泛化实验，验证关键模块的有效性与鲁棒性。

第六章，总结 HASPNN 在计算复杂度、实时性及极端样本适应性方面的不足，提出未来的改进方向，如模型轻量化、自监督学习与多任务联合训练等。

第七章，归纳本文的研究成果与创新点，总结 HASPNN 在 OCT 图像重建任务中的优势，并展望其在医学影像领域的扩展应用。

第二章 相关技术介绍

本章简要回顾卷积神经网络（CNN）、非锐化掩码（USM）、注意力机制、变换器（Transformer）和生成对抗网络（GAN）等关键技术为模型设计提供重要支撑。

2.1 卷积神经网络（CNN）

卷积神经网络（CNN）是深度学习的一种模型，被广泛应用在图像分类、目标检测、语音识别等场景中。相比于传统的全连接神经网络，它使用了卷积层和池化层可以有效地从输入数据中提取出空间上的特征，同时还可以大大减少模型的参数量。其主要优点是自动提取特征、局部感受野、参数共享、平移不变性^[15]。

卷积神经网络一般有多个层级组成，卷积层（Convolutional Layer）、池化层（Pooling Layer）、全连接层（Fully Connected Layer）等。在这样的结构中，输入的图像经过卷积层做局部特征提取，再经过池化层做降维，最后通过全连接层来做分类或者回归等。卷积神经网络结构如图 2.1 所示。

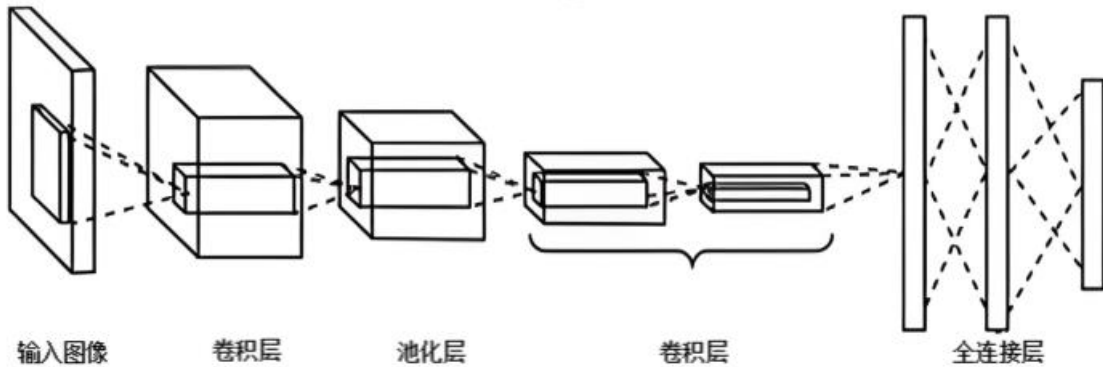


图 2.1 卷积神经网络结构图

图中所示的 CNN 模型包含了多个功能特定的基本模块：

a) 输入图像：输入的图像是希望 CNN 进行处理和分类的原始数据，一般为多通道图像（如 RGB 图像）。

b) 卷积层：卷积层是 CNN 的核心，用于提取输入图像中的局部特征。假设输入为 $X \in \mathbb{R}^{H \times W}$ （高度 H 和宽度 W 的二维图像），卷积操作为：

$$S(i, j) = (X * K)(i, j) = \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} X(i+m, j+n) \cdot K(m, n) \quad (2.1)$$

其中， K 是卷积核， (i,j) 是卷积结果的坐标。

c) 池化层：池化层用于下采样操作，减小特征图的尺寸，降低计算量，同时保留图像的关键信息。常用的池化操作包括最大池化和平均池化。假设输入为特征图 $X \in \mathbb{R}^{H \times W}$ ，池化操作输出尺寸为 $S \in \mathbb{R}^{H' \times W'}$ ，具体为：

$$S(i,j) = \max_{(m,n) \in \text{pooling region}} X(i+m, j+n) \quad (2.2)$$

这里的最大池化操作提取每个局部区域内的最大值。

d) 全连接层：经过卷积和池化操作后，得到的特征图一般会被展平（flatten）后传给全连接层。全连接层的计算形式为：

$$Y = f(WX + b) \quad (2.3)$$

其中 W 是权重矩阵， b 是偏置项， f 是激活函数（如 ReLU）。全连接层一般用于分类任务，输出的是类别的概率分布。

2.2 非锐化掩码（Unsharp Masking）

非锐化掩模（Unsharp Masking，简称 USM），USM 最早来源于传统的摄影过程，通过在图像中增强图像的边缘来产生一个锐利的图像^[29]。虽然名字叫非锐化，但是它并不是对图像进行模糊处理，而是通过对图像中的锐利细节，特别是边缘和细小特征的选择性增强。在现在的数字图像处理中，USM 经常用来增强图像中的细节和图像中的边缘。USM 原理图如图 2.2 所示。

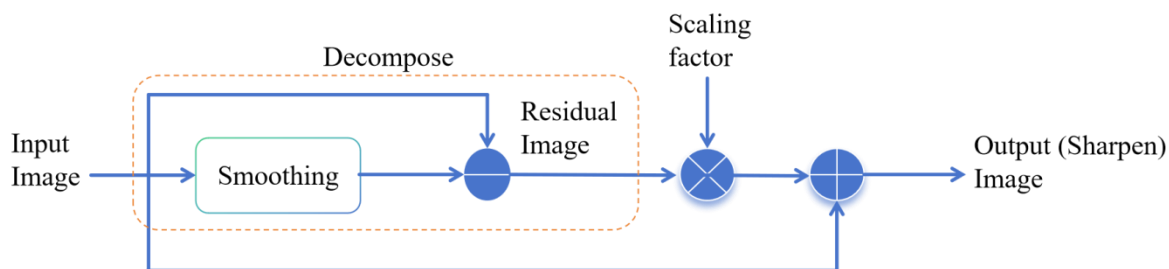


图 2.2 非锐化掩码（Unsharp Masking）原理图

由图 2.2 所示，USM 过程通常包含以下几个步骤：

- 1) 对原图进行模糊，得到低通图像。
- 2) 减去模糊后的图像，从而得到原图中的高频部分（如边缘和细节）。
- 3) 放大高频部分，再和原始图像相加进行图像锐化。

具体来说，非锐化掩膜操作的具体公式表述如下：

$$I_{\text{sharp}} = I_{\text{original}} + k \cdot (I_{\text{original}} - I_{\text{blurred}}) \quad (2.4)$$

其中, I_{sharp} 是锐化后的图像; I_{original} 是原始图像; I_{blurred} 是模糊后的图像, 一般采用高斯滤波器来实现。 k 是一个缩放因子, 控制锐化效果的程度。

2.3 注意力机制

图像超分辨率重建 (Super-Resolution, SR) 是把从低分辨率 (Low-Resolution, LR) 图像变成高分辨率 (High-Resolution, HR) 图像的过程, 其在医学成像、卫星遥感、视频增强等许多领域均有应用^[30]。传统的卷积神经网络中会把每个区域都平等地对待, 这样就无法有效地关注图像中的关键信息, 从而影响了模型性能表现。

注意力机制 (Attention Mechanism) 最早是在自然语言处理 (Natural Language Processing, NLP) 领域提出, 用于模拟人类视觉的注意力分配方式^[31]。在图像处理任务中, 注意力机制可以让网络关注图像中更有价值的部分 (如边缘、纹理等), 从而提高图像重建质量。因此, 越来越多的超分模型使用注意力机制来提高模型对细节信息的表达能力。

图像超分中注意力机制增强特征表示能力有以下常见类型:

(1) 通道注意力 (Channel Attention): 找到对重建任务更有贡献的特征通道, 比如 SE-Net^[32]。

(2) 空间注意力 (Spatial Attention): 关注图像中的重要部分, 例如边缘, 纹理丰富的区域, 比如 Non-LocalNet^[33]。

(3) 联合注意力机制 (Joint Channel-Spatial Attention): 结合通道和空间信息, 改进整体性能, 例如 CBAM^[34]。

采用注意力机制之后, 网络能够更为灵活地对特征的重要程度进行调整, 从而提升超分图像的重建精度与细节表现情况。

2.4 变换器 (Transformer)

Transformer 最早是由 Vaswani 等人在 2017 年首次提出于, 其特征在于不依赖传统的循环或者卷积结构, 而采用“自注意力机制” (Self-Attention) 来建模全局依赖关系^[17]。在此基础上, Transformer 强大的建模能力开始逐渐被应用于计算机视觉任务中, 并已发展为一种新的模型如 ViT (Vision Transformer) 等模型^[35]。

图像超分辨率旨在将低分辨率图像中还原成对应的高分辨率图像。传统的图像超分辨率算法多是基于卷积神经网络 (CNN) 进行局部特征提取, 但是 CNN 由于固定

的感受野，在处理长距离依赖或者结构信息时效果较差。而 Transformer 的出现，解决了这一问题，极大地提高了超分辨率任务的重建性能。其数学公式如下：

设输入为查询 Q 、键 K 和值 V ，维度均为 d ，注意力输出为：

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (2.5)$$

其中， Q 、 K 、 V 一般是通过线性映射从输入获得。 $softmax$ 用于获得注意力权重矩阵。

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^O \quad (2.6)$$

其中每一注意力头计算为：

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (2.7)$$

同时，利用多个注意力头可以有效地提取不同类型地语义和结构性信息。

其次，Transformer 每个 Block 中还有残差连接和前馈网络：

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2 \quad (2.8)$$

整个模块如下所示：

$$Output = LayerNorm\left(x + FFN\left(LayerNorm\left(x + Attention(x)\right)\right)\right) \quad (2.9)$$

2.5 生成对抗网络（GAN）

生成对抗网络（Generative Adversarial Networks, GAN）是由 Ian Goodfellow 等人于 2014 年提出的一种利用博弈论中的“对抗”思想生成更加真实的样本的方法^[16]。

GAN 由两部分构成：生成器（Generator）和判别器（Discriminator）。生成器试图生成以假乱真的样本，判别器试图判断输入的样本是真是假。

图像超分辨率任务中，传统方法会最小化 MSE 等像素级损失，这样得到的峰值信噪比较高，但是重建出的图像会比较平滑、缺乏真实感。SRGAN 出现之后，它引导生成器去学习数据的高频纹理以及感知一致性，从而让生成的图像显得更真实，更具视觉吸引力^[36]。

由图 2.3 可知，SRGAN 把低分辨率图当作输入条件，送到生成器里面去做重建，判别器要分辨输入的是生成的超分辨率图还是原始的高分辨率图。

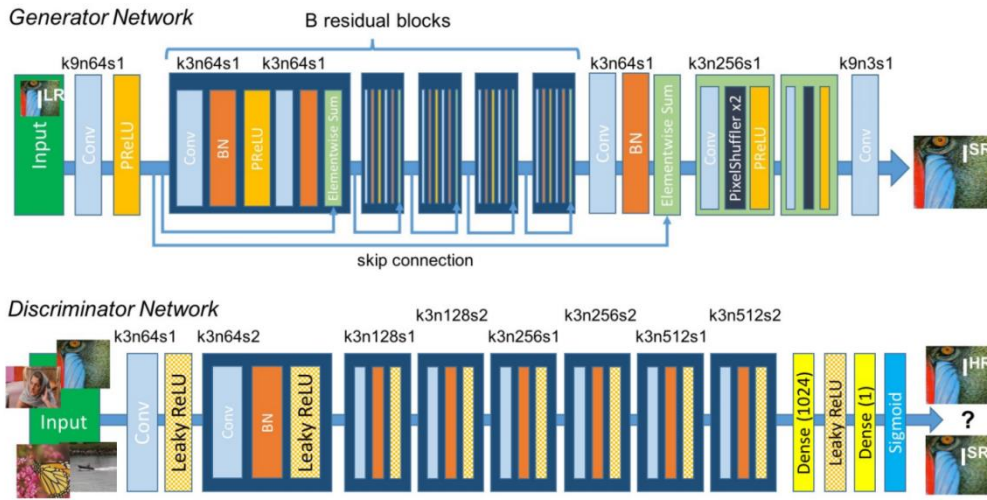


图 2.3 超分辨率生成对抗网络（SRGAN）原理图

SRGAN 的损失函数可表述如下：

$$L_G = L_{content} + L_{per} + \lambda L_{adv} \quad (2.10)$$

其中， $L_{content}$ 是内容损失，用来保证图像的语义一致性。 L_{per} 是基于感知的特征损失。 L_{adv} 是对抗损失，用于改善图像的真实感。 λ 是权重超参数（一般为 $1e-3$ ），用来调整对抗损失的重要性。

$$L_{content} = \frac{1}{N} \sum_{i=1}^N \|I_i^{SR} - I_i^{HR}\|_2^2 \quad (2.11)$$

其中， I_i^{SR} 是第 i 张生成器输出的超分图像。 I_i^{HR} 是第 i 张真实高分辨率图像。

$$L_{per} = \frac{1}{N} \sum_{i=1}^N \|\psi^l(I_i^{SR}) - \psi^l(I_i^{HR})\|_2^2 \quad (2.12)$$

其中， ψ^l 是预训练 VGG-19 网络第 l 层的激活特征图。通过感知损失使得生成的图像具有更好的感知质量，使生成的图像在语义层面上更接近真实的图像，而不仅仅是像素上的匹配。

$$L_{adv} = -\log D(G(I_{LR})) \quad (2.13)$$

其中， $D(\cdot)$ 是判别器输出，表示输入图像被判断为“真实”的概率； $G(I_{LR})$ 是生成图像。通过最小化该损失，促使生成器产生高频细节更多、视觉上更加真实的图像。

第三章 数据预处理与实验设计

本章主要介绍所用数据集、预处理方法与评估指标，并阐述实验配置及训练细节。

3.1 数据集说明

OCT2017 是一个被广泛用作视网膜疾病自动分类研究的公开光学相干断层扫描（Optical Coherence Tomography, OCT）图像数据集。OCT2017 数据集由美国加州大学圣地亚哥分校（University of California, San Diego）医学团队收集整理，并在 Kermany 等人发表于 2018 年的研究中首次公开^[37]。

3.1.1 数据来源

OCT2017 数据集包含了 84,495 张视网膜 OCT 图像，分辨率为 512×496 或者 768×496，其中这 84495 张图像分别来自 4686 名不同的患者，这些图像都是通过频域 OCT（Spectral-Domain OCT, SD-OCT）设备获取的。

3.1.2 样本构成

本研究所采用的 OCT2017 数据集按疾病类型共分为四类：

- (1) 脉络膜新生血管（Choroidal Neovascularization, CNV）：共计 37,455 张图像，主要用于表示湿性年龄相关性黄斑变性（wet Age-related Macular Degeneration, AMD）；
- (2) 糖尿病性黄斑水肿（Diabetic Macular Edema, DME）：共计 11,598 张图像，对应糖尿病引发的黄斑水肿病变；
- (3) 视网膜黄斑硬化症（Drusen）：共计 8,866 张图像，代表干性 AMD 的典型临床表现；
- (4) 正常视网膜（Normal）：共计 26,565 张图像，表示健康视网膜结构。

上述图像均以 JPEG 格式存储，并且严格按照患者独立性的原则划分出训练集、验证集与测试集，保证模型评估的客观性和科学性。

3.1.3 标注标准

每张图像的标签都是由专业的眼科医生根据临床诊断的结果进行标注，保证了高

质量的地面真值（Ground Truth）。图像的文件名遵循统一格式，以疾病的种类，随机化后的患者 ID 以及该患者的图像编号组成，方便对数据进行管理和追踪。

3.1.4 伦理声明

OCT2017 数据集的收集和使用遵守严格的伦理标准^[37]。具体来说，所有的图像都去除了可识别的个人信息，从而保证患者隐私。此外，数据集的发布和使用也得到了相关机构的伦理批准。

3.2 预处理流程

OCT2017 数据集原始数据一共包含了 84,495 幅图像，包括了正常和异常的视网膜图像。本研究首先从中选取了正常视网膜图像子集，共计 1300 张图像作为训练集，200 张图像作为验证集，100 张图像作为测试集。

由于 GPU 资源有限，本文对图像进行了随机裁剪，将每幅图像裁剪成 256x256 的高分辨率（HR）图像作为地面真值。为了获得低分辨率（LR）图像，本研究对高分辨率图像进行了下采样处理，分别得到了 2 倍、4 倍和 8 倍下采样的图像。通过这种方式，本研究得到了相应的低分辨率和高分辨率图像对（LR-HR 图像对）。

此外，为了检验模型的泛化能力，本文从 OCT2017 数据集中随机选取了 100 幅视网膜病变（包括 DME、CNV 以及 Drusen）的 OCT 图像，用于创建对应的子数据集。这些子数据集将会用来验证所构建的网络模型在不同病变图像上的表现。

3.3 评估指标

在本研究中，采用了峰值信噪比（PSNR）和结构相似性指数（SSIM）作为主要的图像质量评估指标，用以评估超分辨率重建图像的质量。PSNR 和 SSIM 都是在图像处理领域广泛使用的标准指标，可以量化图像重建的效果，而且在超分辨率任务当中提可以给出直观的质量评价。

3.3.1 峰值信噪比

PSNR（Peak Signal-to-Noise Ratio）是图像质量评价中最常用的评价指标之一，它是通过比较图像间的差异来评价重建图像的质量^[38]。PSNR 的单位是分贝（dB），它的值越大，说明图像质量越好。

PSNR 的计算公式如下：

$$PSNR = 10 \cdot \log_{10} \left(\frac{I_{max}^2}{MSE} \right) \quad (3.1)$$

其中， I_{max} 是图像中最大的像素值，一般对 8 位图像而言， $I_{max} = 255$ 。MSE 为均方误差（Mean Squared Error），其计算公式为：

$$MSE = \frac{1}{N} \sum_{i=1}^N \|I_i^{SR} - I_i^{HR}\|_2^2 \quad (3.2)$$

PSNR 值越大，说明重建图像与原始图像的差异越小，图像质量越好。在超分辨率任务中，PSNR 通常用来衡量图像整体亮度和对比度是否一致。

3.3.2 结构相似性指数

SSIM（Structural Similarity Index）是一种衡量图像质量的指标，它会考量图像的亮度、对比度和结构信息，以此来判断图像间的相似之处^[39]。SSIM 不同于 PSNR，它不但关注像素级别的误差，而且还会顾及图像的结构信息，因此它能够提供更加符合人眼感知的质量评估。SSIM 的计算公式为：

$$SSIM = \frac{(2\mu_{I_{SR}}\mu_{I_{HR}} + C_1)(2\sigma_{I_{SR}I_{HR}} + C_2)}{(\mu_{I_{SR}}^2 + \mu_{I_{HR}}^2 + C_1)(\sigma_{I_{SR}}^2 + \sigma_{I_{HR}}^2 + C_2)} \quad (3.3)$$

其中， $\mu_{I_{SR}}$ 和 $\mu_{I_{HR}}$ 分别代表图像 I_{SR} 和 I_{HR} 的均值； $\sigma_{I_{SR}}$ 和 $\sigma_{I_{HR}}$ 分别代表图像 I_{SR} 和 I_{HR} 的标准差； $\sigma_{I_{SR}I_{HR}}$ 是图像 I_{SR} 和 I_{HR} 间的协方差； C_1 和 C_2 是用于稳定计算的常数。SSIM 的取值范围为 0 到 1 之间，值越接近 1，表示图像之间的相似度越高，图像质量越好。在超分辨率任务中，SSIM 能够更好地反映图像的结构信息，是衡量图像视觉质量的重要指标。

3.4 实验细节及环境

在训练过程中，HASPN 的俩个主要超参数分别为混合注意力残差块（HARB）的数量 G 和空间注意力残差块（SARB）的数量 M 。其中， G 和 M 分别被设置为 20、5。所有网络均使用 Adam 优化器更新网络参数，其超参数设置为 $\beta_1 = 0.9$ ， $\beta_2 = 0.999$ ，初始学习率为 1×10^{-4} 。为了提升模型的收敛效果，所有网络的学习率均每经过 20 个周期衰减 50%。每个批次的样本数量设置为 2，并对各个模型训练 200 个周期的完整，保证其在训练过程中收敛稳定。本实验均在 Pytorch 2.1.0 框架（支持 Python3.10）

中完成，并在配备 16GB 显存的 NVIDIA RTX 4090 GPU 上完成训练和测试工作。

第四章 模型设计与优化

本文提出混合注意力结构保持网络（HASPN）模型，采用双分支架构与混合注意力机制来提升高频细节恢复能力。

4.1 网络整体架构图

针对神经网络对高频信息建模能力差的问题，本文提出了一种新颖的网络结构即混合注意力结构保持网络（Hybrid Attention Structure Preserving Network, HASPN），其整体框架如图 4.1 所示。HASPN 包含两个分支，即原始低频重建分支和分解高频重建分支。并且在模型中还采用双分支权重共享的方法来大大减少模型的参数量。

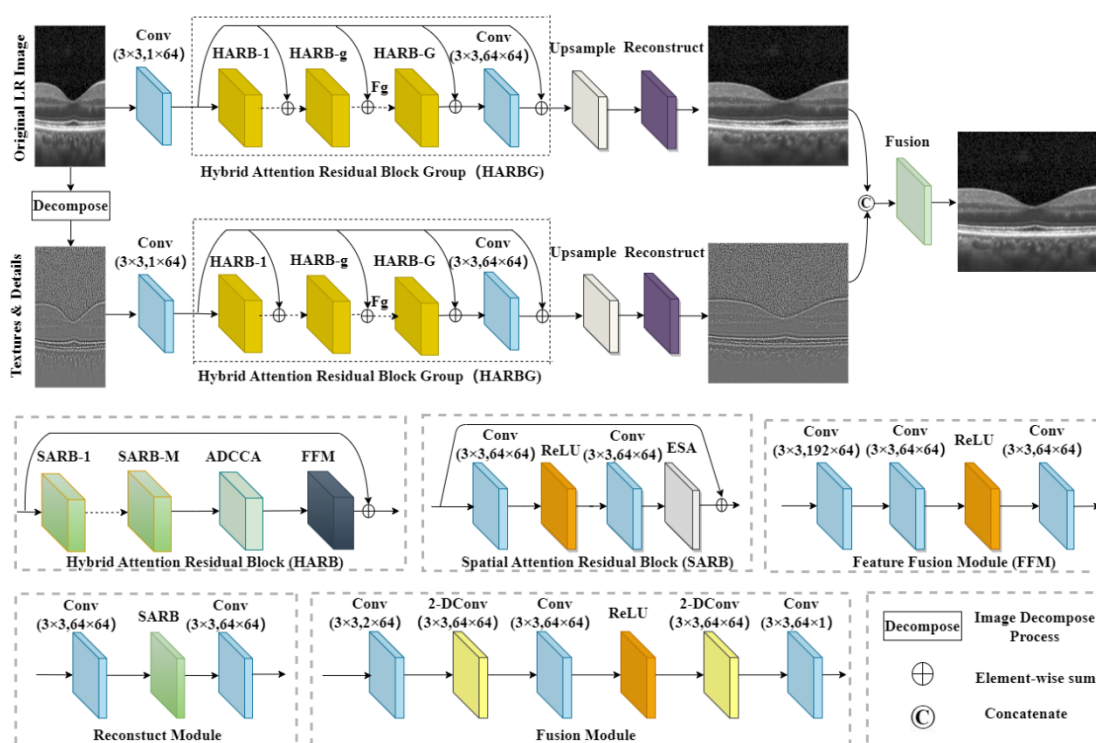


图 4.1 混合注意力结构保持网络（HASPN）架构图

具体来讲，每个重建分支都由三部分组成，分别是浅层特征提取、深层特征提取、上采样重建。首先，浅层特征提取阶段由一个卷积核大小为 3×3 的卷积来提取模型的浅层特征。其次，深层特征提取由 G 个混合注意力残差块（Hybrid Attention Residual Block, HARB）和一个卷积核大小为 3×3 的卷积组成。混合注意力残差块由 M 个空间注意力残差块（Spatial Attention Residual Block, SARB）、一个基于自适应空洞卷积的

通道注意力 (Adaptive Dilated Convolution-based Channel Attention, ADCCA) 和一个特征融合模块 (Feature Fusion Module, FFM) 组成。

其中空间注意力残差块是经典“卷积-激活-卷积”的残差结构与增强空间注意力 (Enhanced Spatial Attention, ESA) 结合, 以提高空间信息的表达能力^[40]。而特征融合模块是普通的卷积层与经典的“卷积-激活-卷积”残差结构结合, 以实现多尺度特征的有效融合。然后进入到上采样重建阶段, 上采样重建阶段包含上采样模块 (Upsample Module, UM) 和重建模块 (Reconstruct Module, RM)。本研究在上采样过程中采用双线性插值的方法来提高图像的分辨率, 在重建模块中将经典残差块中“激活”的结构替换为空间注意力残差块来重新实现重建。最后两个重建分支的输出通道拼接后送入融合模块 (Fusion Module, FM) 来得到最终的重建图像。这个融合模块与特征融合模块的区别就是在每俩个普通卷积层之间增加一个膨胀因子为 2, 卷积核大小为 3×3 的空洞卷积来增大感受野, 更好地融合俩个分支间的上下文信息。

4.2 混合注意力机制

针对超分辨重建任务, 已有研究充分证明了注意力机制在提升重建质量的有效性^[30]。注意力机制可以引导网络关注特征图的关键信息区域。基于此, 本文设计了一种新颖的混合注意力机制——联合通道-空间注意力机制, 其结构如图 4.2 所示。

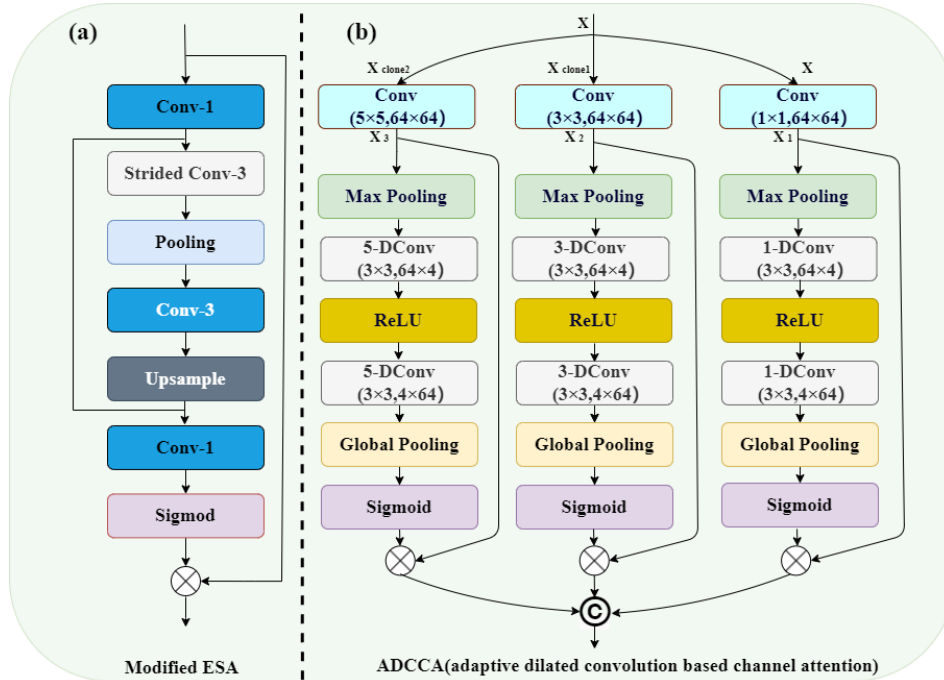


图 4.2 混合注意力机制架构图 (a) 改良增强空间注意力 (b) 基于自适应膨胀卷积的通道注意力

图 4.2.(a) 是其中的空间注意力 ESA 的具体组成。首先, 该模块采用一个卷积核大

小为 1×1 的卷积以及一个卷积核大小为 3×3 、步长为 2 的卷积分别用来减小网络的通道、空间维度，从而大大减少了该注意力模块的参数。紧接着，利用一个池化窗口为 7×7 、步长为 3 的最大池化被用来进一步降低特征图的尺寸并扩大网络的感受野。随后，接着使用一个 3×3 卷积来对下采样后的特征图进行特征提取。为了恢复特征图的原始尺寸，双线性插值以及一个卷积核大小为 1×1 的卷积被用来分别恢复其下采样的空间、通道维度。最终，经过 Sigmoid 函数得到空间注意力图，将其与输入特征图相乘实现对空间维度上的注意力加权。值得注意的是，与原始 ESA 采用多组卷积模块不同，本研究采用一个卷积核大小为 3×3 的卷积来进行特征提取，实验结果表明，该改进证明了本文提出的改良增强空间注意力 (Modified ESA) 相于比原始 ESA，PSNR 值提高了 1.87dB，SSIM 值提高了 0.017。这意味着采取一个卷积大小为 3×3 的卷积不仅可以减少模型的参数而且可以提高模型的重建效果。

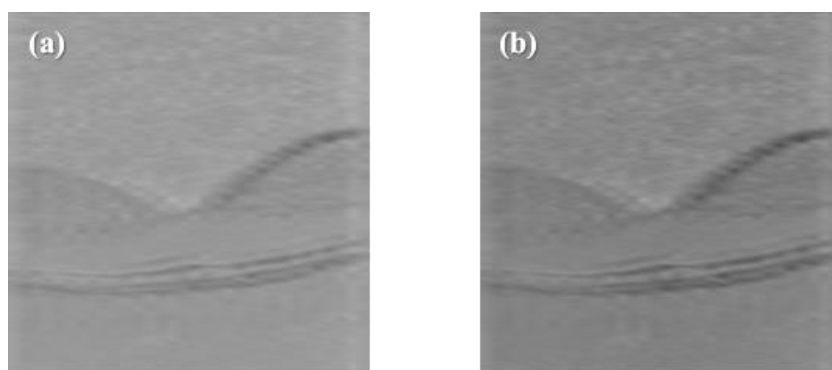


图 4.3 ESA 模块特征图可视化对比 (a) ESA 模块的输入特征图 (b) ESA 模块的输出特征图

为了验证本文提出的 ESA 模块的有效性，通过可视化 ESA 模块的输入特征图与输出特征图，如图 4.3 所示。图 4.3.(a)为 ESA 模块的输入特征图，图 4.3.(b)则为 ESA 模块的输出特征图。从图中可以明显观察出，ESA 模块的输出特征图相比于输入特征图更加清晰地展现了视网膜层。这表明 ESA 模块可以有效的加强与重建任务相关联的重要空间信息，并且抑制不相关的或者冗余的背景信息，从而为后面的图像重建提供更加准确的特征信息。

为了使模型能更好的分辨不同通道特征的重要程度，本文提出了一个新颖的通道注意力模块 ADCCA，其结构图如图 4.2(b)所示。ADCCA 模块中加入多尺度信息融合，利用不同的膨胀因子和卷积核大小来提取多尺度的特征。ADCCA 模块的结构设计借鉴了 SENet^[32]，但结构与功能有所不同，首先 ADCCA 模块使用最大池化操作对输入特征图进行一次“压缩” (squeeze)，即空间压缩。然后模块用不同的膨胀因子的“膨胀卷积-激活函数-膨胀卷积”结构实现对特征的非线性变换以及通道关系建模，完成“激发” (excitation) 的过程，最后通过全局平均池化将特征图压缩成 1×1 的通道描述

向量，通过 Sigmoid 激活函数得到各个通道的重要性权重，从而完成通道注意力的自适应调整。

4.3 高频分解

Xu 等人^[41]得出在 CNN 处理图像的时候，对低频信息比高频信息更敏感，但高频信息对超分辨重建很重要。为了应对这种情况，本文采用 USM 的一种启发式方法，将图像分解成包含高频内容的残差图像，以此提升高频信息重建性能。本文先使用了高斯核大小为 5×5 且 X 方向上标准差为 1.5 的高斯模糊器来模糊图像，接着用原始图像减去模糊图像得到残差图像，从而得到高频信息凸显的残差图像。与传统的 USM 方法不同，本文并未采用直接将残差图像和原始图像相加的方式增强图像细节，而是设计了一个双分支的重建策略，主要分支用于重建低分辨率的低频原始图像，辅助分支用于重建对应的高频残差图像，最后两个分支的输出在通道上进行拼接，再通过一个专门设计的融合模块对原始图像和残差图像信息进行融合，从而提高图像的重建效果。

4.4 损失函数设计

为了在保持更清晰的视网膜结构同时追求更高的 PSNR 值，本文提出了一种新颖的混合损失函数，其结构如下：

$$L = L_{\alpha} + L_{\beta} + L_{\gamma} \quad (4.1)$$

其中， L_{α} 、 L_{β} 和 L_{γ} 分别代表重构的原始图像和原始图像的地面真值间的损失，重构的残差图像和残差图像的地面真值间的损失和最终的重构图像与原始图像的地面真值间的损失。

L_{α} 、 L_{β} 和 L_{γ} 均采用相同的度量函数，定义如下：

$$L_{\alpha, \beta, \gamma} = L_{pix} + L_{per} + L_{gra} \quad (4.2)$$

Lim 等人^[42]发现，尽管最小化 L2 范数可以最大化 PSNR 值，但采用 L1 范数往往能导致更好的网络收敛。因此，本文采取 L1 范数来衡量模型输出与地面真值的像素级误差。像素损失的定义如下：

$$L_{pix} = \frac{1}{N} \sum_{i=1}^N \|I_i^{SR} - I_i^{HR}\|_1 \quad (4.3)$$

其中 I_i^{SR} 是第 i 张模型输出的超分图像。 I_i^{HR} 是第 i 张真实高分辨率图像。

然而仅使用像素损失难以确保模型在视觉效果上的最优表现。因此，本文还引入了一个感知损失来提升模型输出的超分图像与 HR 图像的视觉相似性^[43]。具体来说，其利用了一个预训练的 VGG19 网络^[44]来提取第 L 层的高级语义信息，并采用 L2 范数来衡量超分图像与真实图像之间特征空间的差异。感知损失的定义如下：

$$L_{per} = \frac{1}{N} \sum_{i=1}^N \|\varphi^L(I_i^{SR}) - \varphi^L(I_i^{HR})\|_2^2 \quad (4.4)$$

其中 $\varphi^L(I_i^{SR})$ 和 $\varphi^L(I_i^{HR})$ 分别表示第 L 层提取的第 i 张超分图像的特征和第 L 层提取的第 i 张真实高分辨率图像的特征。

最后，为避免模型在最小化像素损失时引入过度平滑的结果，本文额外采取了一个梯度损失来显式惩罚图像的梯度。梯度损失定义如下：

$$L_{gra} = \frac{1}{N} \sum_{i=1}^N \|\nabla I_i^{SR} - \nabla I_i^{HR}\|_1 \quad (4.5)$$

其中， ∇I_i^{SR} 和 ∇I_i^{HR} 分别表示第 i 张超分辨图像和经过梯度算子运算后的第 i 张真实高分辨率图像施加梯度算子后的结果。梯度算子 ∇I_i^{SR} 定义如下，*表示卷积运算：

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad (4.6)$$

$$\nabla I_i^{SR} = \frac{\partial I_i^{SR}}{\partial x} + \frac{\partial I_i^{SR}}{\partial y} \quad (4.7)$$

$$\frac{\partial I_i^{SR}}{\partial x} = G_x * I_i^{SR}, \frac{\partial I_i^{SR}}{\partial y} = G_y * I_i^{SR} \quad (4.8)$$

第五章 实验与结果分析

本章通过多组对比和消融实验，验证本文提出方法的有效性与优越性能，凸显其泛化能力。

5.1 定性对比

为了证明本文提出的模型 HASPN 的优越性，通过对比现有多个主流方法：Bicubic, SRCNN^[45], FSRCNN^[46], EDSR^[42], RDN^[47], RCAN^[48], SRGAN^[36], ESRGAN^[49], RFANet^[40], RVSRNet^[50], TDPN^[51], SwinIR^[52], ESRT^[53]。

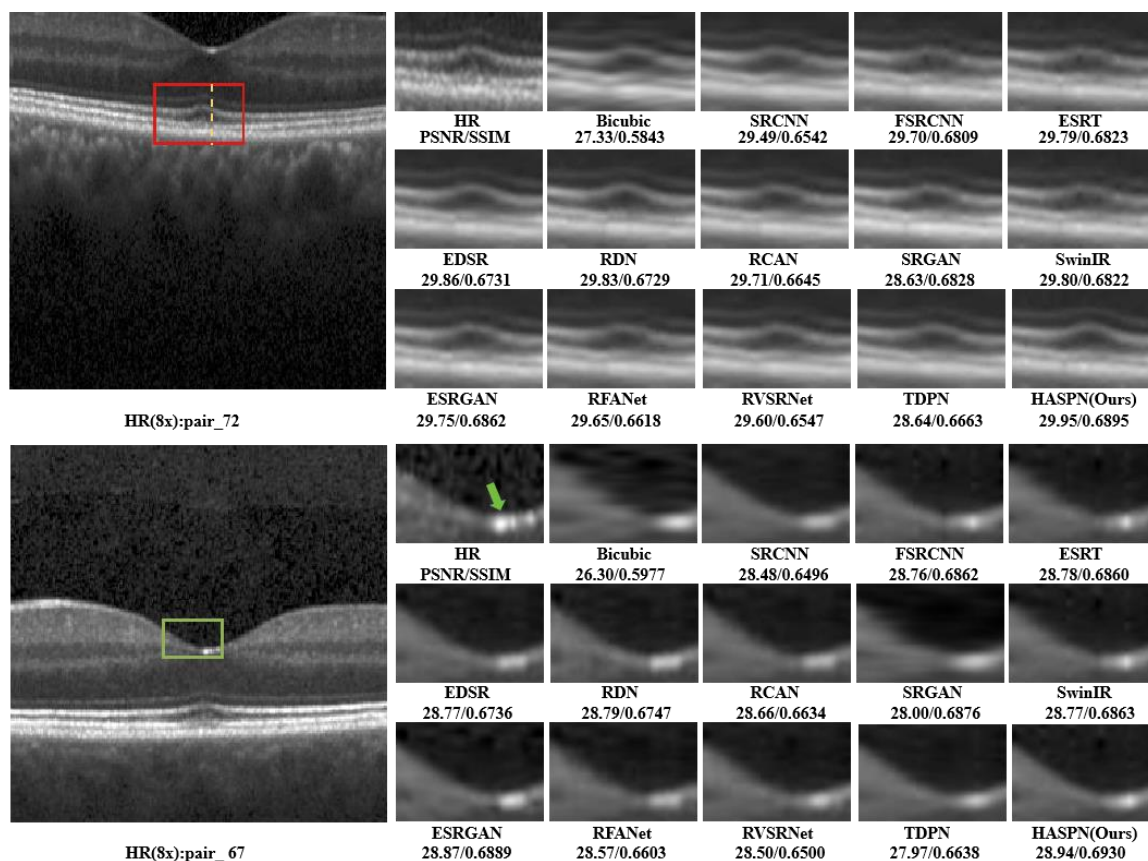


图 5.1 不同方法在 8 倍超分辨率下的可视化对比

如图 5.1 中的 pair_72 所示，Bicubic 重建图像中的外节（Outer Segment, OS）区域存在明显的不连续性。由于插值方法本身的局限性，图像在外界膜（External Limiting Membrane, ELM）和视网膜色素上皮层（Retinal Pigment Epithelium, RPE）边缘区域出现了大量振铃伪影。采用反卷积作为上采样方式的 FSRCNN，其重建图像中的 ELM

区域过于模糊，且受到伪影的干扰，严重影响了图像的最终质量^[54]。让人感到意外的是，SRCNN 在重建的 ELM 区域里，对比度和清晰度要比 Bicubic 和 FSRCNN 更高，不过这些方法在 RPE 层的重建效果还是不太好，有很严重的结构扭曲现象。

EDSR、RDN 和 RCAN 通过使用宽通道残差块、密集残差连接和通道注意力机制等设计来增强网络的特征提取能力，在视觉效果上比 Bicubic、SRCNN 和 FSRCNN 好。而且重建的结果也比 Bicubic、SRCNN 和 FSRCNN 接近于 GAN 类方法（如 SRGAN、ESRGAN）和 RFANet。TDPN 在重建视网膜结构上比以上方法更加清晰，但在细节上不如本文方法。同时 TDPN 与 RVSRNet 都没有很好的重建 RPE 层中细小的颗粒状结构。虽然 SwinIR 和 ESRT 在 RPE 层的细粒度结构重建方面表现良好，但在 ELM 重建方面对比度较低，并且在 OS 区域出现了伪影。相比之下，本文提出的 HASPN 模型在细节结构的还原方面更具优势，所重建的 ELM 对比度明显优于其他方法，且 OS 区域更加连续，展现出更好的结构还原能力。

此外，在图 5.1 中的 pair_67 中，Bicubic、SRGAN 与 RFANet 所重建的内界膜（Inner Limiting Membrane, ILM）呈现出阶梯状结构。对于中央凹（由绿色箭头指示），除了 FSRCNN、SwinIR、ESRT 和本文提出的 HASPN 方法之外，其余方法与 HR 图像相比存在较大的差异。RDN 和 ESRGAN 没有正确地重建 INL，TDPN 在 INL 的重建上与 HR 图像的差异最小，但是重建的 ILM 与中央凹仍有些模糊，SwinIR 和 ESRT 在 INL 的重建上也不好，导致图像边缘模糊。

而本文提出的 HASPN 模型在这几个结构的重建上表现较好，不仅重建出了完整的 INL 结构，而且在 ILM 区域的视觉效果与 HR 图像最为接近，这些结果也充分说明了 HASPN 在视网膜图像细节结构重建上的优势。

为了进一步体现本方法在重建视网膜层级结构的性能，本文选取图 5.1 中 HR(8x): pair_72 所示的橙色虚线位置对应的 A-scan，对 HASPN、TDPN、SRGAN（PSNR 排名第二）以及 ESRT（SSIM 排名第二）四种方法进行了对比，并将结果绘制在图 5.2 中。从图 5.2 中可以观察到，在第 65 至 70 行的峰值区域对应于图 5.1 所示的视网膜内/外节连接（IS/OS junction）。结果显示 HASPN 与 TDPN 重建出来的结构和真实图像非常接近，而 ESRT 与真实图像之间有较大的差别。

特别是第 70-75 行对应的外节（OS），HASPN 重建出的 OS 区域几乎和真实图像重合，而 TDPN 和 ESRT 重建出的 OS 区域存在一定的偏差，说明 HASPN 在保持视网膜层结构完整性的优势上更为突出。在第 80-85 行对应的视网膜色素上皮层（RPE）中，TDPN、ESRT 和 HASPN 重建出的结果也较为接近真实图像，但 SRGAN 重建出的结果与真实图像有较大的偏差，可能是因为 SRGAN 在重建过程中容易导致图像边

缘过度平滑。

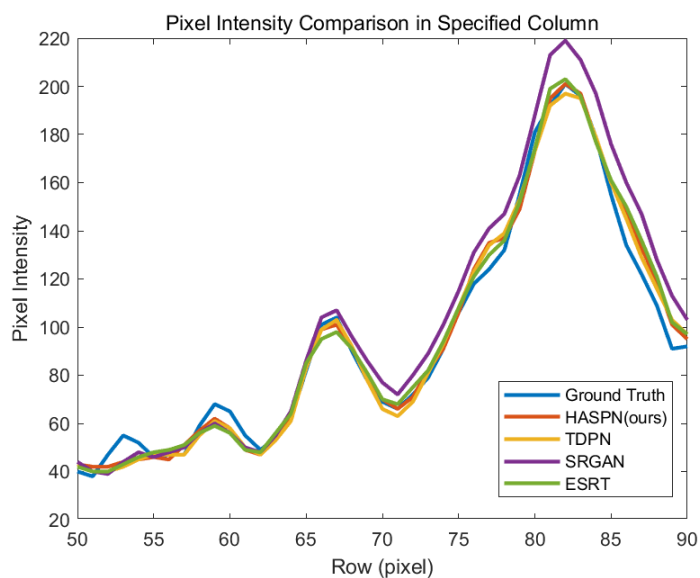


图 5.2 不同方法在图 5.1 中 pair_72 的橙色虚线处进行 8 倍超分时的单个 A-scan 对比

5.2 定量对比

在表 5.1 中，本文从模型复杂度指标即每秒浮点运算次数（FLOPs）、参数量（Parameter）、平均延迟时间（Average Latency）和图像质量指标即 PSNR、SSIM 两个方面对 HASPN 与其他主流方法在 2 倍、4 倍和 8 倍放大倍率进行了定量对比。

表 5.1 各方法在 2×、4×、8×超分重建下的定量对比。↑/↓ 表示数值越高/低性能越好，最佳与次优结果分别用粗体和下划线标注

Model	upscale	FLOPs[G]↓	Param[M]↓	Average Latency[S]↓	PSNR↑	SSIM↑
Bicubic	2x	-	-	0.052014	31.73	0.8410
SRCNN	2x	15.0158	0.057281	0.000371	33.62	0.8877
FSCRNN	2x	0.662979	0.010289	0.000487	33.66	0.8884
EDSR	2x	2591.4	39.5402	0.087134	33.59	0.8871
RDN	2x	1444.9	22.0472	0.089945	33.67	0.8897
RCAN	2x	999.595	15.3685	0.060812	32.55	0.8667
SRGAN	2x	32.5143	0.491911	0.002044	32.42	0.8620
ERSGAN	2x	1098	16.6385	0.066602	33.73	0.8913
RFANet	2x	642.309	6.4015	0.061185	<u>33.69</u>	<u>0.8898</u>
RVSRNet	2x	422.278	6.4421	0.022905	32.18	0.8486

续表 5.1

Model	upscale	FLOPs[G]↓	Param[M]↓	Average Latency[S]↓	PSNR↑	SSIM↑
TDPN	2x	1900	0.564972	0.096445	31.36	0.8761
SwinIR	2x	48.0272	0.762889	0.040426	33.54	0.8861
ESRT	2x	43.1854	0.639061	0.061485	33.49	0.8862
HASPN	2x	1940.5	0.579052	0.108374	32.65	0.8881
Bicubic	4x	-	-	0.042997	27.68	0.6633
SRCNN	4x	7.5079	0.057281	0.000172	29.44	0.7102
FSCRNN	4x	0.397546	0.012305	0.000474	<u>30.11</u>	<u>0.7637</u>
EDSR	4x	1373.2	40.7204	0.044873	29.95	0.7593
RDN	4x	727.313	22.1211	0.042283	29.92	0.7589
RCAN	4x	504.671	15.4423	0.052618	30.01	0.7507
SRGAN	4x	21.433	0.565767	0.001517	29.68	0.7402
ERSGAN	4x	565.921	16.6551	0.036978	29.74	0.7557
RFANet	4x	325.959	6.4753	0.063954	29.97	0.7604
RVSRNet	4x	211.139	6.4421	0.010950	29.11	0.6939
TDPN	4x	1037.6	0.564972	0.058766	30.06	0.7629
SwinIR	4x	28.8874	0.836745	0.088075	29.96	0.7607
ESRT	4x	24.0506	0.676053	0.036188	29.95	0.7611
HASPN	4x	1060.1	0.579052	0.107809	30.14	0.7650
Bicubic	8x	-	-	0.041999	25.74	0.5686
SRCNN	8x	3.7539	0.057281	0.000168	27.79	0.6276
FSCRNN	8x	0.264832	0.016337	0.000450	28.25	0.6733
EDSR	8x	764.072	41.9005	0.026542	28.29	0.6538
RDN	8x	368.531	22.1949	0.021923	28.27	0.6541
RCAN	8x	257.21	15.5162	0.059235	28.14	0.6421
SRGAN	8x	15.8923	0.639623	0.001417	<u>28.35</u>	0.6624
ERSGAN	8x	452.247	16.6884	0.052569	28.34	0.6645
RFANet	8x	167.715	6.5492	0.062491	28.04	0.6392
RVSRNet	8x	105.57	6.4421	0.005082	27.94	0.6232
TDPN	8x	693.935	0.564972	0.070226	27.64	0.6482
SwinIR	8x	19.3175	0.910601	0.049893	28.30	0.6739
ESRT	8x	14.4833	0.713045	0.018722	28.28	<u>0.6742</u>
HASPN	8x	709.793	0.579052	0.105395	28.55	0.6786

由上表可知，在 2 倍超分任务中，ESRGAN 和 RFANet 分别达到了最高和次高，但是 HASPN 虽然在 2 倍超分下没有达到最高的 PSNR，但是其 SSIM 只比 ESRGAN 稍低一些，表现依旧很优秀。FSRCNN 在保持最低模型复杂度的同时，其性能仅次于 HASPN，但是其重建图像的视觉效果还是不如部分主流方法。

在 8 倍超分任务中，SRGAN 和 ESRT 分别获得了次高的 PSNR 和 SSIM，而本文所提出的 HASPN 在 4 倍和 8 倍超分任务中分别获得了最好的整体性能。但是需要指

出的是，HASPN 的 FLOPs 较大导致其延迟较高，可能对本研究在临床环境中的实用化产生一定的影响。后续本研究将致力于剪枝、量化、蒸馏等降低模型计算复杂度的技术以提高模型实用性。

5.3 补充实验

5.3.1 双分支有效性验证

为了验证双分支分别增强低频和高频特征的效果，本文对最终超分辨率图像与原始分支重建图像之间谱幅值差异图，以及最终超分辨率图像与高频分解分支重建图像之间谱幅值差异图进行了可视化分析。如图 5.3 所示，最终的超分辨率图像相比于原始分支重建图像含有更多的高频成分，相比于高频分解分支重建图像含有更多的低频成分。说明高频分解分支很好的补充了原始分支在高频特征上的不足，降低了神经网络的频谱偏移，提高了整个网络的性能。

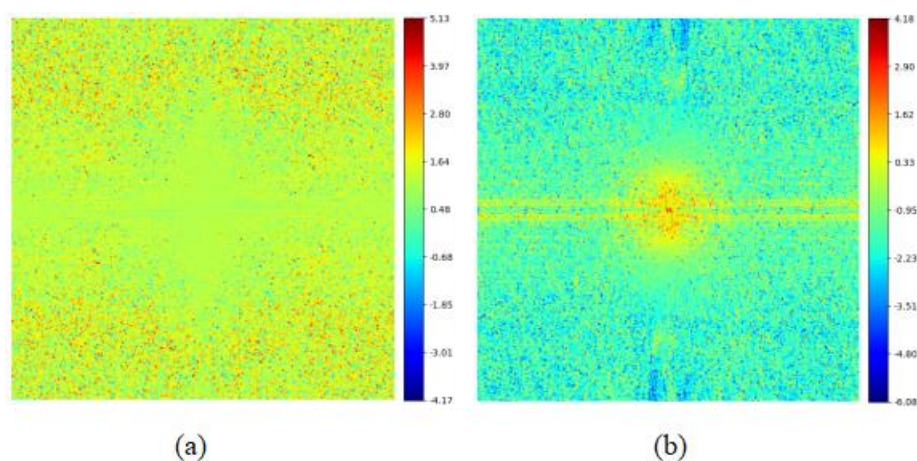


图 5.3 频谱幅值差异图 (a) 最终超分辨率图像与原始分支重建图像的差异图 (b) 最终超分辨率图像与高频分解分支重建图像的差异图

5.3.2 模型超参数实验

之后本文对模型的超参数进行了系统的调整，减小 HARB (G)、SARB (M) 和通道数 (C)，探究网络深度和宽度对模型的影响。由表 5.2 可知，随着 G、M 和 C 的增加，模型的性能得到了明显的提高，这说明扩大网络的深度和宽度可以提高模型的特征提取和表示能力。其中，G、M 和 C 分别设置为 20、5、64 时，模型在 4 倍超

分任务中得到最好的 PSNR 和 SSIM 值。同时 G 对于 PSNR 和 SSIM 的提升较大，尤其是在取值 8 和 16 时，M 对于 SSIM 的提升较小，尤其是在取值 2 和 4 时，C 对于 PSNR 的提升较小，尤其是在取值 16 和 32 时。

表 5.2 不同深度与宽度配置的 HASPN 架构在 4×超分上的定量对比，G、M、C 分别表示每层 HARB、SARB、通道数

G	M	C	PSNR↑	SSIM↑
20	5	64	30.14	0.7650
16	5	64	29.57	0.7564
8	5	64	28.66	0.7474
20	4	64	29.83	0.7546
20	2	64	29.17	0.7536
20	5	32	29.12	0.7352
20	5	16	29.08	0.7287

5.3.3 模型消融实验

为验证所提出关键模块的有效性，本研究分别对 ESA、ADCCA 以及高频重建分支 (Tex Branch) 进行了消融实验。如表 5.3 所示，由 Model 1 重建的图像在 PSNR 和 SSIM 方面表现一般，但在引入 ADCCA 和 ESA 后，模型性能显著提升。当所有关键模块均被集成时，模型达到最佳性能。

表 5.3 不同关键模块消融在 8 倍超分任务中的对比

Model	ESA	ADCCA	Tex Branch	PSNR↑	SSIM↑
1	×	×	×	28.01	0.6359
2	√	×	×	28.11	0.6404
3	×	√	×	28.23	0.6492
4	√	√	×	28.26	0.6540
5	√	√	√	28.55	0.6786

表 5.4 不同损失函数组成部分消融在 8 倍超分任务中的对比

Loss	Pix	Per	Gra	PSNR↑	SSIM↑
1	√	×	×	28.18	0.6645
2	×	√	×	25.68	0.4913
3	×	×	√	28.17	0.6756
4	√	√	√	28.55	0.6786

此外，本研究还对混合损失函数进行了消融实验，以验证其各组成部分的作用。如表 5.4 所示，当同时使用像素损失 (Pix)、感知损失 (Per) 和梯度损失 (Gra) 时，

模型表现最优，进一步证明了混合损失在提升模型性能方面的重要作用。

5.3.4 模型泛化实验

最后，本研究使用第 3.2 节中提到的 DME、CNV 和 DRUSEN 子数据集，对基于正常视网膜数据训练的模型进行了泛化能力测试。

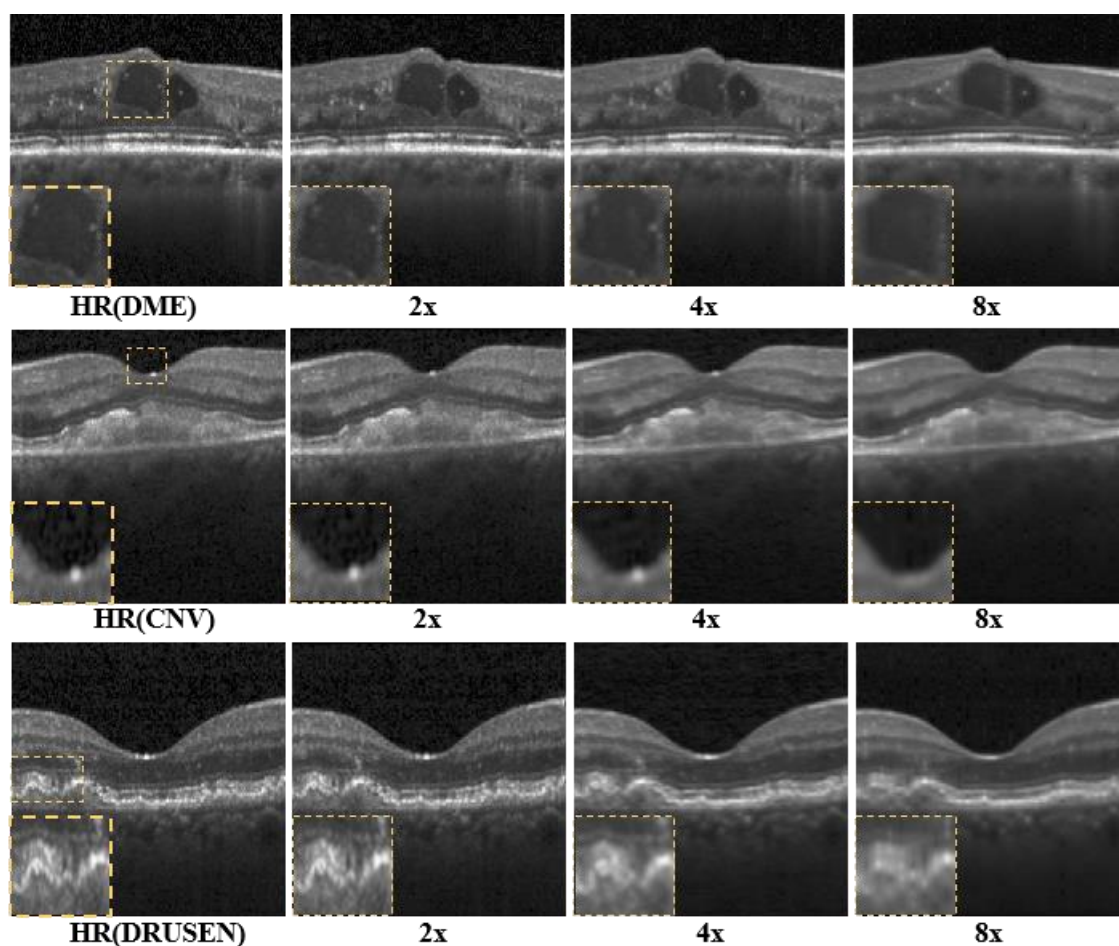


图 5.4 HASPN 在 DME、CNV 及 DRUSEN 子数据集上的泛化性能

如图 5.4 可见, HASPN 能够在 2 倍、4 倍和 8 倍超分辨率下重建出视网膜层结构。2 倍放大下重建的图像与高分辨率图像非常接近,但在 8 倍超分辨率下,一些细节结构(图中橙色虚线框所示)没有完全重建出来。除了这些复杂细节之外,该方法可以很好地重建大部分视网膜层结构。由此可以看出 HASPN 具有很好的泛化能力,有着在临床上应用的潜力。

第六章 模型不足与展望

本章分析模型在效率与鲁棒性上的不足，展望自监督学习、轻量化及多任务优化等发展方向。

6.1 模型不足

尽管本文的 HASPN 由于采用了双分支的权重共享机制有效的控制了模型的参数大小，使 Parameter 始终处于正常水平，但由于整个网络比较复杂，存在多层混合注意力残差块以及空洞卷积等计算，导致整个模型的 FLOPs 和 Average Latency 较大，如表 5.1 所示。因此，其在一定程度上限制了该模型的临床实时应用，特别是对计算资源有硬性要求的设备上推理速度较慢。

模型虽然在标准测试集以及多个病变子集中有较好的泛化能力，但是对于一些极端病变样本以及噪声较大的图像，模型的鲁棒性有待进一步提高。训练时对于高质量标注数据的依赖度较大，且获取成本高，限制了模型的广泛应用和迁移能力。

6.2 未来展望

针对以上不足之处，今后可以关注几个方面，第一，持续开展网络架构的轻量化研究工作，同模型剪枝，量化以及知识蒸馏这样的技术手段相结合来减轻计算压力并缩减推理延迟时间；第二，探究是否可用自监督学习或者无监督方式下的重建办法，从而削减对大量已标注资料所存在依靠性问题；第三，考虑把联合重建及后续的临床任务纳入多任务学习框架内进行探索，进而达成端到端的图像重新构建目标，增强诊断准确度的同时也能加快诊断进程；第四点是，针对不同的 OCT 装置还有采集协议去设计各自专属化的适配模块，从而改善整个模型对不同机器的适应范围。

综上所述，HASPN 在欠采样 OCT 图像重建上具有较好的性能基础和应用前景，但模型效率和实际临床应用上仍存在很大的改进空间，未来将围绕实用化展开研究。

第七章 总结

本研究提出了一种新颖的混合注意力结构保持网络（HASPNet），旨在提高 OCT 图像的采样效率，同时实现接近密集采样质量的高空间分辨率重建。利用公开发表的视网膜 OCT 数据集 OCT2017，对 HASPNet 进行了横向超分辨率重建任务的研究，实验结果显示，HASPNet 在超分辨率重建任务上与当前各种超分方法相比较在 4 倍和 8 倍放大倍数下均取得了最优的性能。通过对 HASPNet 模型进行分析，从定性、定量两方面表明 HASPNet 能够以更优的方式恢复出空间分辨率高、结构信息丰富的图像。同时，系统研究了网络深度、宽度的变化对超分辨效果的影响。并做了大量的消融实验来证明模型中关键模块和混合损失的有效性。实验结果表明 HASPNet 具有良好的泛化性，可以用于跨域的 OCT 图像重建。接下来的工作将会研究基于自监督学习方法来改善欠采样 OCT 图像的重建效果，并且将 HASPNet 推广到其他的医学成像领域，如 MRI、CT 等，以拓宽 HASPNet 在医学研究和临床中的应用。

参考文献

- [1] HUANG D, SWANSON E A, LIN C P, et al. Optical coherence tomography [J]. science, 1991, 254(5035): 1178-81.
- [2] VAN VELTHOVEN M E, FABER D J, VERBRAAK F D, et al. Recent developments in optical coherence tomography for imaging the retina [J]. Progress in retinal and eye research, 2007, 26(1): 57-77.
- [3] SHIN E-S, ANN S H, SINGH G B, et al. OCT-defined morphological characteristics of coronary artery spasm sites in vasospastic angina [J]. JACC: Cardiovascular Imaging, 2015, 8(9): 1059-67.
- [4] BARAN U, CHOI W J, WANG R K. Potential use of OCT-based microangiography in clinical dermatology [J]. Skin research and technology, 2016, 22(2): 238-46.
- [5] OTIS L L, EVERETT M J, SATHYAM U S, et al. Optical coherence tomography: a new imaging: technology for dentistry [J]. The Journal of the American Dental Association, 2000, 131(4): 511-4.
- [6] 袁贤琳, 李百灵. 光学相干断层成像技术在医学应用中的研究进展 [J]. 中国处方药, 2021, 19(11): 71-3.
- [7] YANAGIHARA R T, LEE C S, TING D S W, et al. Methodological challenges of deep learning in optical coherence tomography for retinal diseases: a review [J]. Translational Vision Science & Technology, 2020, 9(2): 11-.
- [8] XU J, ISHIKAWA H, WOLLSTEIN G, et al. 3D OCT eye movement correction based on particle filtering; proceedings of the 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, F, 2010 [C]. IEEE.
- [9] 朱杨兴. 医学断层图像插值重建算法的研究 [D]; 东南大学, 2006.
- [10] 王彩艳. 基于稀疏表示的图像超分辨率重建算法研究 [D]; 武汉科技大学, 2023.
- [11] LIU X, KANG J U. Compressive SD-OCT: the application of compressed sensing in spectral domain optical coherence tomography [J]. Optics express, 2010, 18(21): 22010-9.
- [12] HE Z, XIAO Z, XU Z, et al. MFGAN: OCT image super-resolution and enhancement with blind degradation and multi-frame fusion; proceedings of the International Workshop on Advanced Imaging Technology (IWAIT) 2025, F, 2025 [C]. SPIE.
- [13] LIAO J, ZHANG T, SHEPHERD S, et al. Semi-supervised assisted multi-task learning for oral optical coherence tomography image segmentation and denoising [J]. Biomedical Optics Express, 2025, 16(3): 1197-215.

- [14] QUAN X, HOU G, YIN W, et al. A multi-modal and multi-stage fusion enhancement network for segmentation based on OCT and OCTA images [J]. Information Fusion, 2025, 113: 102594.
- [15] ALZUBAIDI L, ZHANG J, HUMAIDI A J, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions [J]. Journal of big Data, 2021, 8: 1-74.
- [16] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks [J]. Communications of the ACM, 2020, 63(11): 139-44.
- [17] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [J]. Advances in neural information processing systems, 2017, 30.
- [18] 何锦涛, 陈明惠, 贾文字, et al. 眼底 OCT 图像中糖尿病性黄斑水肿的分割 [J]. 光电工程, 2018, 45(7): 170605-1--8.
- [19] YANG D, RAN A R, NGUYEN T X, et al. Deep learning in optical coherence tomography angiography: Current progress, challenges, and future directions [J]. Diagnostics, 2023, 13(2): 326.
- [20] FANG L, LI S, MCNABB R P, et al. Fast acquisition and reconstruction of optical coherence tomography images via sparse representation [J]. IEEE transactions on medical imaging, 2013, 32(11): 2034-49.
- [21] ABBASI A, MONADJEMI A, FANG L, et al. Optical coherence tomography retinal image reconstruction via nonlocal weighted sparse representation [J]. Journal of biomedical optics, 2018, 23(3): 036011-.
- [22] WANG L, CHEN Z, ZHU Z, et al. Compressive-sensing swept-source optical coherence tomography angiography with reduced noise [J]. Journal of Biophotonics, 2022, 15(8): e202200087.
- [23] HUANG Y, LU Z, SHAO Z, et al. Simultaneous denoising and super-resolution of optical coherence tomography images based on generative adversarial network [J]. Optics express, 2019, 27(9): 12289-307.
- [24] LEE W, NAM H S, SEOK J Y, et al. Deep learning-based image enhancement in optical coherence tomography by exploiting interference fringe [J]. Communications Biology, 2023, 6(1): 464.
- [25] GAO M, GUO Y, HORMEL T T, et al. Reconstruction of high-resolution 6× 6-mm OCT angiograms using deep learning [J]. Biomedical Optics Express, 2020, 11(7): 3585-600.
- [26] 柯舒婷, 陈明惠, 郑泽希, et al. 生成对抗网络对 OCT 视网膜图像的超分辨率重建 [J]. Chinese Journal of Lasers, 2022, 49(15): 1507203--9.
- [27] 芦焱琦, 陈明惠, 秦楷博, et al. 基于金字塔长程 Transformer 的 OCT 图像超分辨率重建 [J]. Chinese Journal of Lasers, 2023, 50(15): 1507107--12.

- [28] XU Z-Q J, ZHANG Y, LUO T, et al. Frequency principle: Fourier analysis sheds light on deep neural networks [J]. arXiv preprint arXiv:190106523, 2019.
- [29] MALIN D F. Unsharp masking [J]. AAS Photo-Bull, Issue 16 (1977/3), p 10-13, 1977, 16: 10-3.
- [30] ZHU H, XIE C, FEI Y, et al. Attention mechanisms in CNN-based single image super-resolution: A brief review and a new perspective [J]. Electronics, 2021, 10(10): 1187.
- [31] HU D. An introductory survey on attention mechanisms in NLP problems; proceedings of the Intelligent Systems and Applications: Proceedings of the 2019 Intelligent Systems Conference (IntelliSys) Volume 2, F, 2020 [C]. Springer.
- [32] HU J, SHEN L, SUN G. Squeeze-and-excitation networks; proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, F, 2018 [C].
- [33] WANG X, GIRSHICK R, GUPTA A, et al. Non-local neural networks; proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, F, 2018 [C].
- [34] WOO S, PARK J, LEE J-Y, et al. Cbam: Convolutional block attention module; proceedings of the Proceedings of the European conference on computer vision (ECCV), F, 2018 [C].
- [35] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale [J]. arXiv preprint arXiv:201011929, 2020.
- [36] LEDIG C, THEIS L, HUSZÁR F, et al. Photo-realistic single image super-resolution using a generative adversarial network; proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, F, 2017 [C].
- [37] KERMANY D S, GOLDBAUM M, CAI W, et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning [J]. Cell, 2018, 172(5): 1122-31.e9.
- [38] KORHONEN J, YOU J. Peak signal-to-noise ratio revisited: Is simple beautiful?; proceedings of the 2012 Fourth international workshop on quality of multimedia experience, F, 2012 [C]. IEEE.
- [39] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity [J]. IEEE transactions on image processing, 2004, 13(4): 600-12.
- [40] LIU J, ZHANG W, TANG Y, et al. Residual feature aggregation network for image super-resolution; proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, F, 2020 [C].
- [41] XU K, QIN M, SUN F, et al. Learning in the frequency domain; proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, F, 2020 [C].
- [42] LIM B, SON S, KIM H, et al. Enhanced deep residual networks for single image super-resolution; proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition

- workshops, F, 2017 [C].
- [43] JOHNSON J, ALAHI A, FEI-FEI L. Perceptual losses for real-time style transfer and super-resolution; proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14, F, 2016 [C]. Springer.
- [44] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv:14091556, 2014.
- [45] DONG C, LOY C C, HE K, et al. Image super-resolution using deep convolutional networks [J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 38(2): 295-307.
- [46] DONG C, LOY C C, TANG X. Accelerating the super-resolution convolutional neural network; proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14, F, 2016 [C]. Springer.
- [47] ZHANG Y, TIAN Y, KONG Y, et al. Residual dense network for image super-resolution; proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, F, 2018 [C].
- [48] ZHANG Y, LI K, LI K, et al. Image super-resolution using very deep residual channel attention networks; proceedings of the Proceedings of the European conference on computer vision (ECCV), F, 2018 [C].
- [49] WANG X, YU K, WU S, et al. Esrgan: Enhanced super-resolution generative adversarial networks; proceedings of the Proceedings of the European conference on computer vision (ECCV) workshops, F, 2018 [C].
- [50] WANG L, CHEN S, LIU L, et al. Axial super-resolution optical coherence tomography via complex-valued network [J]. Physics in Medicine & Biology, 2023, 68(23): 235016.
- [51] CAI Q, LI J, LI H, et al. TDPN: Texture and detail-preserving network for single image super-resolution [J]. IEEE Transactions on Image Processing, 2022, 31: 2375-89.
- [52] LIANG J, CAO J, SUN G, et al. Swinir: Image restoration using swin transformer; proceedings of the Proceedings of the IEEE/CVF international conference on computer vision, F, 2021 [C].
- [53] LU Z, LI J, LIU H, et al. Transformer for single image super-resolution; proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, F, 2022 [C].
- [54] ODENA A, DUMOULIN V, OLAH C. Deconvolution and checkerboard artifacts [J]. Distill, 2016, 1(10): e3.

致 谢

时光荏苒，岁月如梭，四年的大学生活即将画上圆满的句号。在这段充实而宝贵的时光中，我不仅收获了知识，更收获了成长与感悟。在本论文完成之际，我怀着无比感激的心情向所有在我学习和生活中给予帮助和支持的人表示诚挚的谢意。

首先，我要衷心感谢我的指导老师赵占芳老师，感谢您在百忙之中给予我耐心细致的指导。从选题思路的确立到论文写作的每一处细节，您严谨的治学态度和精益求精的工作精神让我受益匪浅。在我遇到困难和迷茫时，您总是耐心解答，给予我信心和方向。此外，感谢河北地质大学这片充满活力的学术沃土，感谢各位任课老师在课堂上认真授课、倾囊相授，是您们的教诲为我打下了坚实的专业基础。

同时，我要感谢与我并肩作战的同学和朋友，感谢你们在学习、生活中的陪伴和鼓励，让我在求学路上不再孤单。我们一起拼搏、一起欢笑，这段共同经历将成为我人生中最珍贵的回忆。

最后，衷心感谢我的父母和家人，感谢你们一直以来无私的关爱与默默的支持，是你们给了我坚定的后盾和无限的动力。

再次感谢所有帮助和支持过我的人，愿我们今后的道路皆宽广，前程似锦！