

天津大学

本科生毕业设计



题目：基于双目相机的图像联合超分辨率与去噪研究

学 院 电气自动化与信息工程学院

专 业 通信工程

年 级 2020

姓 名 何宇正

学 号 3020234384

指导教师 岳焕景

独创性声明

本人声明：所呈交的毕业设计（论文），是本人在指导教师指导下，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，本毕业设计（论文）中不包含任何他人已经发表或撰写过的研究成果。对本毕业设计（论文）所涉及的研究工作做出贡献的其他个人和集体，均已在论文中作了明确的说明。本毕业设计（论文）原创性声明的法律责任由本人承担。

论文作者签名：

年 月 日

本人声明：本毕业设计（论文）是本人指导学生完成的研究成果，已经审阅过论文的全部内容。

论文指导教师签名：

年 月 日

摘 要

超分辨率作为一项用于提升图像质量的图像处理技术，在社会生活中具有广泛应用。其中，基于双目镜头的超分辨率利用一对镜头（广角镜头与长焦镜头）捕捉的图像进行超分辨率重建，其拥有天然参考图的优势弥补了基于参考图像超分辨率难以获取参考图的缺点。同时，由于现阶段大多数超分辨率研究只关注无噪的情况，而真实图片往往含有大量噪声，因此本课题对双目镜头联合超分辨率与去噪展开研究，其中如何在无干净样本条件下训练超分辨率网络的去噪能力成为主要挑战。本课题的主要工作与贡献如下：

一是构建了一个带噪数据集：其中 450 对图像样本均在光照不良条件下使用双目镜头设备拍摄；二是设计了一种去噪损失函数：其利用预训练的盲去噪网络处理超分辨率重建结果，通过计算与优化重建图片去噪前后噪声的减小幅度，间接训练超分辨率网络的去噪能力。实验结果表明，噪声会显著降低超分辨率重建图像的质量，而本实验设计的噪声损失函数的引入能有效滤除噪声，提高了超分辨率网络在复杂环境下的实用性和泛用性。

关键词： 超分辨率技术，双目镜头超分辨率，去噪能力，损失函数

ABSTRACT

Super-resolution, as an image processing technology aimed at enhancing image quality, has extensive applications in societal life. Among these, dual-lens super-resolution utilizes images captured by a pair of lenses (a wide-angle and a telephoto lens) to perform super-resolution reconstruction. This approach benefits from having a natural reference image, which addresses the challenge often faced by reference-based super-resolution techniques in obtaining suitable reference images. Currently, most research in super-resolution focuses on noise-free scenarios, while real-world images typically contain significant noise. Thus, this project investigates the integration of super-resolution and denoising using dual-lens systems, with the primary challenge being how to train a super-resolution network's denoising capabilities without clean samples. The main work and contributions of this project are as follows:

Firstly, a noisy dataset was constructed, comprising 450 pairs of image samples taken under poor lighting conditions using dual-lens equipment. Secondly, a denoising loss function was designed: it uses a pre-trained blind denoising network to process the super-resolution reconstruction results. By calculating and optimizing the reduction in noise before and after denoising the reconstructed images, the super-resolution network's denoising capability is indirectly trained. Experimental results demonstrate that noise significantly reduces the quality of super-resolution reconstructed images. The introduction of the designed noise loss function effectively filters out noise, enhancing the practicality and generalizability of the super-resolution network in complex environments.

KEY WORDS: Super-resolution, Dual-lens super-resolution, Denoising capability, Loss function

目 录

第一章 绪论	1
第二章 研究历史与现状	2
2.1 图像超分辨率	2
2.1.1 单图像超分辨率	2
2.1.2 基于参考图像的超分辨率	3
2.1.3 基于双目镜头的超分辨率	4
2.2 图像去噪	4
2.2.1 噪声类型	4
2.2.2 去噪网络	6
2.3 性能评价指标	7
第三章 数据与网络设计	9
3.1 数据处理	9
3.1.1 数据获取	9
3.1.2 粗略对齐	8
3.1.3 精确对齐	10
3.1.4 色彩校正	11
3.1.5 数据裁剪	11
3.2 神经网络架构	11
3.2.1 中心对齐	12
3.2.2 无核匹配与角落对齐	12
3.2.3 单图像超分辨率编码器	13
3.2.4 自适应融合模块	13
3.3 损失函数	14
3.3.1 重建损失	14
3.3.2 感知损失与对抗损失	14
3.3.3 噪声损失	15
第四章 实验与分析	17

4.1	数据准备	17
4.2	不同损失函数对比实验	19
4.3	实验与结果比较	20
4.3.1	性能指标对比	21
4.3.2	视觉效果对比	21
4.4	实验结果分析与总结	22
第五章	总结与展望	26
5.1	课题总结	26
5.2	未来展望	26
5.3	设计方案经济成本分析与对社会环境的影响	26
参考文献		28
致 谢		30

第一章 绪论

在当今的信息科技时代，高质量的图像数据对于各种应用领域具有至关重要的作用，涵盖医疗成像、卫星图像解析和媒体消费等多个方面。其中图像超分辨率（Image Super-Resolution）和图像去噪（Image Denoising）已成为获取高质量图像数据的核心技术。

图像超分辨率作为一项能提升图像清晰度的图像处理技术，在医疗、安全及娱乐等领域有广泛应用。例如，在医疗领域，图像超分辨率技术能提高 CT、MRI 扫描图像的清晰度，帮助医生识别细微的解剖结构和潜在病变；在安全监控领域，图像超分辨率技术可以用来增强监控摄像头捕捉的低分辨率图像，比如放大并清晰地显示人脸、车牌等关键信息，从而提升公共安全和助力犯罪侦破；在日常娱乐领域，超分辨率技术可以用来提升老旧影片的图像质量，使其达到现代高清标准，满足现代观众的观看需求。

图像去噪技术是图像处理中的另一项关键技术，旨在从图像中移除噪声同时保持图像的重要细节和结构，其在医疗、天文及遥感等领域中拥有广泛应用。例如，在医疗领域，图像去噪技术有助于减少医疗成像噪声和伪影，从而促进更准确的诊断；在天文领域，由于遥远星体的光线非常微弱，拍摄图像常常带有大量噪声，而应用图像去噪技术能清晰显示天体图像；在遥感成像领域，图像去噪技术可以减弱因大气扰动、光照条件变化等带来的噪声影响，提高由卫星或航空摄像获得的图像质量。

尽管图像超分辨率和去噪技术在各自领域内已展现出显著成效，但目前研究往往将两者分开处理，例如仅使用无噪样本训练超分辨率网络。因此，本研究旨在探讨将图像超分辨率与去噪技术结合的可能性与效果，通过在超分辨率神经网络中引入用于去噪的损失函数，同时训练网络的超分辨率重建与去噪能力。本文的结构安排如下：第一章为绪论；第二章介绍相关研究和评价指标；第三章介绍数据处理和网络设计；第四章分析实验设计和结果；第五章总结课题并展望未来。

第二章 研究历史与现状

2.1 图像超分辨率

图像的分辨率指的是图像的像素密度，分辨率高代表像素密度高，能够提供更多细节。而图像的超分辨率则是从低分辨率（Low-Resolution，简称 LR）图像中恢复出高分辨率（High-Resolution，简称 HR）图像的过程。

现有的图像超分辨率方法可大致分为基于插值、基于重构以及基于学习的方法。由于深度学习在近年来的迅速发展，各种专注处理图像的神经网络（如卷积神经网络）应运而生，各种优点如自动提取特征、适应性强使得基于学习的方法取得了巨大成功，因此本节将着重介绍基于深度学习的方法。

2.1.1 单图像超分辨率

单图像超分辨率是旨在从单个低分辨率图像恢复高分辨率图像的过程，而如何从有限的像素信息推断出更多的高频细节成为单图像超分辨率的主要挑战。

超分辨率卷积神经网络（Super-Resolution Convolutional Neural Network，简称 SRCNN）的诞生首次打开了深度学习在该领域应用的大门^[1]。它采取了简单的三层卷积层作为网络结构，通过一个端到端的学习过程，实现了比当时所有传统方法都优越的超分辨率性能。

卷积神经网络（Convolutional Neural Network，简称 CNN）是一种专注处理图像数据的网络结构，它克服了原始将图像数据转成一维后丧失空间信息的缺点，利用卷积层直接从二维图像中提取特征。SRCNN 正是采用三层卷积层去拟合低分辨率与高分辨率图像对之间的非线性映射，充分利用 CNN 的图像特征提取优势，在实现高性能的同时还维持快速的处理速度。自 SRCNN 被设计出后，各种以 CNN 为基础的单图像超分辨率网络被设计出来，例如快速超分辨率卷积神经网络（Fast Super-Resolution Convolutional Neural Network，简称 FSRCNN）在网络结构上相较于以往的 SRCNN 进行了优化，将所有卷积层嵌入更小的网络中，以降低计算复杂度和模型参数数量^[19]；亚像素卷积神经网络（Efficient Sub-Pixel Convolutional Neural Network，简称 ESPCN）则利用亚像素卷积技术在输出层直接生成高分辨率图像，进一步提高了重建速度^[20]。

随着研究进行，各种新的网络结构被加入到超分辨率网络中。其中，超分辨率生成对抗网络（Super-Resolution Generative Adversarial Network，简称 SRGAN）

不仅引入了残差块去解决梯度消失与网络退化问题，而且还采用了生成对抗网络（Generative Adversarial Network，简称 GAN）结构在生成判别、互相博弈的训练过程中实现生成高质量图像^[2]。这种生成对抗的机制确保网络不断调整策略以提高输出质量，实现更高的视觉保真度。与此同时，SRGAN 新定义了感知损失函数，旨在恢复图像的高频信息，重建人眼感知度更高的图像细节。这种损失函数通过衡量图像之间的感知相似度，引导生成网络生成更加接近人眼所感知的高质量图像，而不仅仅依赖于像素级别的误差。此外还有许多单图像超分辨率网络在各种超分辨率任务背景下取得优秀成果，这些网络在网络结构及损失函数设计方面各有所异，同时网络复杂度和注意力机制的引入等方面也大有不同。

2.1.2 基于参考图像的超分辨率

基于参考图像的超分辨率利用额外的高分辨率图像作为参考（Reference，简称 Ref）图像，辅助低分辨率图像进行超分辨率重建。由于神经网络能利用的信息增多，该超分辨率方法往往比单图像超分辨率效果更好，但同时面临着更多挑战，比如如何获取相似的参考图像、图像特征如何匹配以及复杂度的提高导致实时性差等等。

其中低分辨率图像与参考图像的特征匹配是核心挑战，其匹配效果将直接影响网络的输出效果，因此各种神经网络大多基于特征匹配算法进行创新。一种基于检索的超分辨率网络（Landmark）首次引入参考图像，其采用一种名为全局注册的方法实现低分辨率图像与参考图像在空间上的对齐，以提高后续匹配精度^[3]。除此之外，Landmark 还采用结构感知标准优化匹配精度，在计算像素级差异的同时，还考虑了图像块之间的结构差异。

后续越来越多的图像匹配算法被开发出来，如密集补丁匹配^[21]、光流法^[22]等。其中密集补丁匹配（Dense Patch Matching）是一种利用图像块或补丁进行匹配的策略，通过将目标图像划分为一系列小的补丁，并在参考图像中搜索最相似的补丁，以实现特征的一致性匹配，即保持图像细节和纹理的一致性；而光流法（Optical Flow）是一种计算机视觉中的运动估计技术，通过跟踪图像中像素的移动路径来估计场景中物体的运动方向和速度，在超分辨率应用中表现出独特优势。光流法可以在一系列图像帧中保持运动的连续性，因此其可以通过跟踪图像中的运动路径为超分辨率任务提供可靠的参考。后续随着越来越多图像匹配算法的开发，各种各样的适用该超分辨率方法的任务场景也逐渐被发掘出来。

2.1.3 基于双目镜头的超分辨率

基于双目镜头的超分辨率是基于参考图像超分辨率的实际应用场景之一，其拥有天然参考图的优势弥补了常规应用场景下参考图难以获取的不足。其利用一组镜头拍摄同一场景得到低分辨率-高分辨率图像对，在此过程中，由长焦镜头拍摄的清晰图像被视作高分辨率图像，由广角镜头拍摄的模糊图像被视作低分辨率图像。双镜头超分辨率网络（Dual-Lens Super-Resolution，简称 DCSR）首次探讨了能否利用长焦镜头拍摄的图像来提高广角图像分辨率的问题，同时被成功证明该方法的可行性^[4]。DCSR 使用了对齐注意力模块来缓解不匹配的问题，同时设计了一种自监督适应方案使模型更加适应现实的图像输入。通过在研究者自己构建的数据集和公共数据集上的广泛训练与验证，DCSR 被证明实现了优异的超分辨率重建性能。

尽管 DCSR 是双目镜头超分辨率领域的一个良好开端，但数据对齐与匹配等问题仍需要更先进的算法去解决。同时双目镜头使用不同焦距拍摄导致的视场差距和分辨率差异成为新的问题。为解决数据对齐与匹配的问题，无核双目超分辨率网络（Kernel-Free Dual-Lens Super-Resolution，简称 KeDuSR）采用一种无核匹配策略，使得低分辨率图像的角落与中心区域进行匹配，巧妙避免了视场与分辨率差异问题^[5]。同时 KeDuSR 的训练过程是在完全由人工拍摄的现实图像上进行的，因此在性能领先的同时，也拥有更高的泛用能力。

2.2 图像去噪

图像噪声是指在图像中出现的与原始图像无关的、不可预期的随机或非随机的干扰信号，通常表现为图像中的颗粒、杂点或纹理，通常会导致图像细节丢失，质量下降。图像噪声的来源包括传感器的物理限制、信号传输过程中的干扰以及图像处理算法中的误差等，其中常见的噪声类型包括高斯噪声、椒盐噪声和泊松噪声等。

2.2.1 噪声类型

高斯噪声是图像处理领域最常见的噪声类型，其统计分布符合高斯分布特性，即概率密度函数为正态分布曲线，且通常情况下均值为零，标准差反应噪声强度。

在图像处理中高斯噪声主要由传感器的热噪声、传输过程中信号受到的干扰或量化误差引起，会导致图像出现细节模糊、颗粒感增加等质量问题。在以深度学习为基础的图像去噪方法里，研究人员通常在干净图像中添加特定强度的高斯噪声，以获取受到污染的训练数据和测试数据。图 2-1 为原始图像与受到高斯噪

声污染图像:

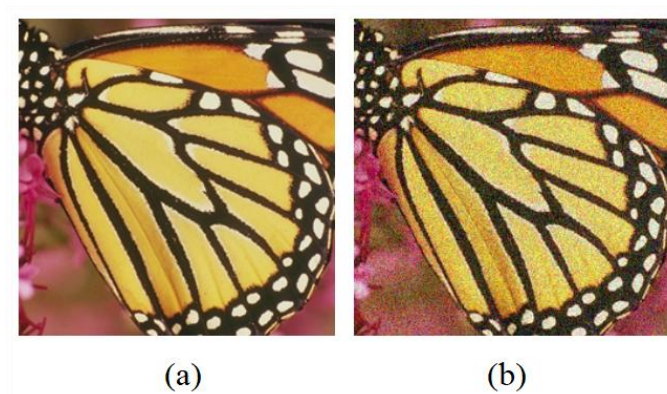


图 2-1 高斯噪声示例

(a) 原始图像; (b) 受到高斯噪声污染图像

椒盐噪声是另一种常见的噪声类型,也被称为脉冲噪声或散斑噪声,呈现为图像中的黑色或白色像素点,并且出现位置具有随机性。具体而言,这些像素点值达到最大值或最小值后,便表现出黑白斑点形式,其通常由图像传感器的故障、数据传输过程中的错误或恶劣环境中的干扰引起,会引起图像质量的显著下降,尤其会影响图像细节及纹理。图 2-2 为原始图像与受到椒盐噪声污染图像:

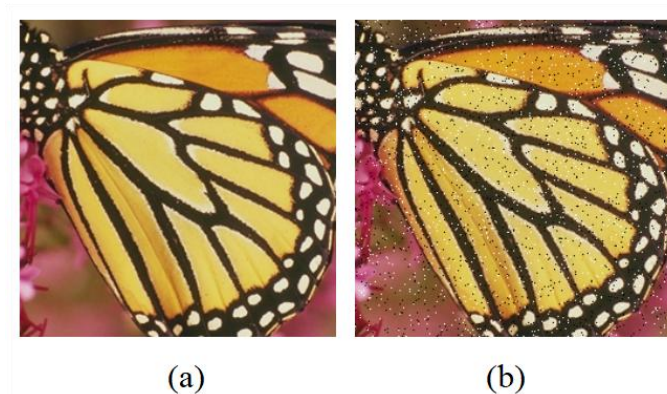


图 2-2 椒盐噪声示例

(a) 原始图像; (b) 受到椒盐噪声污染图像

泊松噪声因其分布符合泊松分布而得名,其存在的根本原因在于光是由离散的光子构成的,其强度大小与光源强度和传感器原理直接相关,因此泊松噪声通常发生在信号计数的过程中,如传感器记录光信号的过程。图 2-3 为原始图像与受到泊松噪声污染图像:

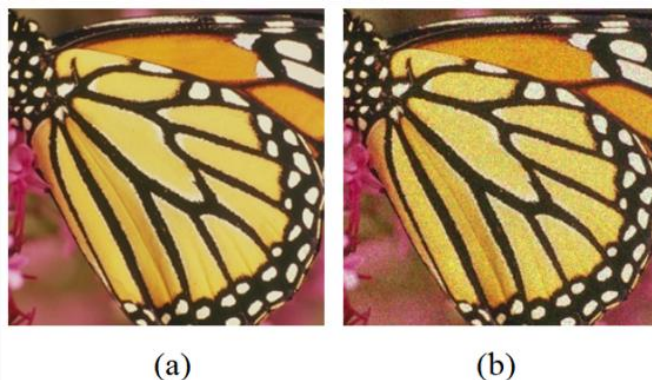


图 2-3 泊松噪声示例

(a) 原始图像；(b) 受到泊松噪声污染图像

通常情况下，在光源中心泊松噪声强度很高，但由于信号强度最高，导致实际信噪比维持在较高水平，因此噪声的视觉效果并不明显。在远离光源区域，由于光照弱导致噪声强度很低，因此肉眼也无法明显辨别泊松噪声的存在。但是在光源附近，由于相比光源中心信号强度大幅下降，而噪声强度仍维持较高水平，导致信噪比较高，因此噪声表现较为明显。在光照不良条件下拍摄的图片会存在大量泊松噪声，尤其在光源附近噪声表现尤为明显，图像细节也会因此受到影响，而由于其噪声类型更符合实际噪声分布，因此在自然图像降噪的过程中需要首先考虑对泊松噪声进行滤除。

2.2.2 去噪网络

图像去噪作为图像处理领域里的另一项重要任务，旨在从受噪声污染的图像中恢复出干净的原始图像。近年来深度学习在图像去噪领域取得了显著的进展，去噪卷积神经网络（Denoising Convolutional Neural Network，简称 DnCNN）便是首个图像去噪神经网络，其采取了残差学习策略和深层网络结构，能捕获复杂的噪声模式，同时允许网络直接学习输入图像与干净图像的差异^[6]。快速去噪卷积神经网络（Fast and Flexible Denoising Convolutional Network，简称 FFDNet）在 DnCNN 的基础上进行了升级，因其使用额外一个可控参数作为输入，因此 FFDNet 对不同强度噪声的适应能力与计算量均要优于 DnCNN^[18]。后来研究人员意识到，仅在添加了高斯噪声的图片上进行训练并不能很好对自然图片降噪，因此新的网络结构的提出、自然图像数据集的拍摄以及新的损失函数的设计等问题亟待解决。因此在 DnCNN 与 FFDNet 被提出后，各种类型的神经网络架构也被设计出应用于图像去噪任务，这些网络各自发挥其优势，适用于不同的应用场景和噪声类型。

除了去噪性能，图像去噪同样会遇到许多其它困难，如数据依赖性，细节模糊等等。数据依赖性主要体现在很难获得干净的自然图像，为解决这个问题，新的 Noise2Noise 去噪思想被提出^[7]。这种自监督去噪技术旨在仅从噪声图像中恢复干净图像，并给出了非常严谨的数学证明，在一定程度上缓解了部分应用场景下的数据稀少问题。由于这种方法需要一些数学前提，如需已知图片的噪声模式等，导致其应用范围非常有限。而细节模糊的问题几乎存在所有的去噪网络中，因为噪声可能与图像的细节纹理有相似的特性，在去噪过程中难以区分，容易导致细节丢失或留下残余噪声，因此设计去噪神经网络时往往需要在去噪能力与细节模糊程度之间做出权衡。

2.3 性能评价指标

图像超分辨率与去噪任务通常需要评价网络输出图像与标签图像的相似度，不同的指标可以从不同的角度评价图片的质量和相似性。以下是两种常用的评价指标：

峰值信噪比（Peak Signal-to-Noise Ratio，简称 PSNR）用于计算像素级的误差。PSNR 的计算是在均方误差(Mean Squared Error，简称 MSE)的基础上进行的，其表达式分别如下：

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [(I_{(i,j)} - K_{(i,j)})]^2 \quad (2-1)$$

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right) = 20 \cdot \log_{10} \left(\frac{MAX_I}{\sqrt{MSE}} \right) \quad (2-2)$$

公式（2-1）中 I 、 K 分别表示两张图像，其尺寸大小均为 $m \times n$ ，一张为高质量参考图像，另一张为进行评估的图像，通常是进行重建后的图像。较高的 PSNR 通常表示两幅图像之间的相似度较高。尽管 PSNR 能提供一种简单直接的误差度量，但它无法充分捕捉图像中的感知质量。

结构相似度（Structural Similarity Index Measure，简称 SSIM）是另一种广泛使用的图像相似度评价指标，其在图像质量的衡量上，更侧重评价图像的结构信息，其表达式如下：

$$SSIM_{(x,y)} = [l_{(x,y)}]^\alpha [c_{(x,y)}]^\beta [s_{(x,y)}]^\gamma \quad (2-3)$$

其中， x, y 为两个图像信号， $l_{(x,y)}$ 为亮度对比， $c_{(x,y)}$ 为对比度对比， $s_{(x,y)}$ 为结构对比， α, β, γ 为常数。将亮度、对比度、结构计算公式代入原始表达式中可得到最终的 SSIM 表达式：

$$SSIM_{(x,y)} = \frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)} \quad (2-4)$$

其中 μ_x 、 μ_y 代表两张图像的平均亮度； σ_x^2 、 σ_y^2 代表两张图像的方差，即对比度； σ_{xy} 代表两张图像的协方差，即结构相似度； c_1 、 c_2 为避免分母为 0 的常数。

SSIM 通过模仿人类视觉系统，从三个角度对图像进行评估，使得其评价值更高的结果更接近人类视觉的判断。通常情况下，虽然 SSIM 评估结果趋势与 PSNR 大体一致，但 SSIM 通常被认为在很多场景下比 PSNR 更能反映真实的图像质量。

除了 PSNR 和 SSIM 以外，其他评价指标也在图像处理领域广泛使用，例如均方误差（MSE）、均方根值误差（Root Mean Square Error，简称 RMSE）、峰度相似度（Gradient Magnitude Similarity Deviation，简称 GMSD）等。这些指标可以根据特定场景或需求，提供不同的图像相似度评估方法，因此选择适当的评价指标应取决于研究的目标和图像特性，以确保评估结果能够充分反映图像的真实质量和相似程度。

第三章 数据与网络设计

本节将以无核匹配双目超分辨率网络（KeDuSR）为基础，依次介绍数据处理方法、网络构架以及损失函数设计原理。

3.1 数据处理

3.1.1 数据获取

为了提高网络的泛用能力，KeDuSR 使用真实世界拍摄的数据集进行网络训练。但是在动态场景中采集对齐良好的数据图像是一项颇具挑战性的任务，因为场景中移动的对象可能会导致不同镜头之间的图像发生错位。以往的数据集如 CameraFusion^[4]采用镜头切换的方式获取数据，然而这种方式会因为两个镜头拍摄的时间间隔而产生较大的对齐误差，特别是在存在运动物体或动态场景时，两个镜头捕获的图像无法保持一致性。

现代智能手机大多装配了两个及以上的摄像头，并能通过一些应用程序同时调动一组摄像头进行拍摄。为了克服这一问题，本实验将使用 iPhone 13 手机，通过 DoubleTake 应用程序在采集数据的同时激活广角和长焦两个镜头。这种同步捕捉的方式能够确保两幅图像的拍摄时间点完全一致，从而在动态场景中避免因镜头错位导致的对齐问题。这种数据采集方法不仅提高了配对图像之间的对齐准确性，还能够更真实地模拟实际应用中的双镜头拍摄场景，为后续的图像超分辨率训练和评估提供更精确的数据基础。

但由于一组镜头在空间上是错位的，对同一场景的拍摄会导致些许视角的差异，并且由于镜头焦距的不同，因此在数据输入神经网络之前还需要进行一些列数据处理工作。

3.1.2 粗略对齐

在获得了真实拍摄的低分辨率-高分辨率图像对后，首先需要对其进行粗略对齐处理，包括尺度不变特征变换（Scale-Invariant Feature Transform, 简称 SIFT）^[8]和随机抽样一致性（Random Sample Consensus, 简称 RANSAC）^[9]。

SIFT 算法作为一种计算机视觉的特征提取算法，旨在侦测与提取图像的局部特征，使这些特征在不同尺度空间中保持一致，并确保特征在旋转、缩放和光照等变化下依然能够稳定识别。

本实验首先需要提取广角与长焦图像中的关键特征点。由于 SIFT 算法具有尺度不变和旋转不变的特性，能够可靠地检测出广角与长焦图像对中的显著特征，因此首先采用 SIFT 提取特征点。随后，对提取的特征点进行匹配，识别出在广角与长焦图像对中共同出现的特征点对。

RANSAC 算法是从一组含有外点的数据中正确估计数学模型参数的迭代算法，在本实验中被应用于在匹配的特征点对中剔除错误匹配的点对的过程，从而得到更为准确的特征点匹配结果。这些被筛选出的可靠的匹配点随后被应用于计算广角与长焦镜头图像之间的单应矩阵（Homography Matrix）。

最后利用计算出的单应矩阵，将长焦镜头的图像变换到广角镜头的视角，以便实现初步的几何对齐。这一粗略对齐过程确保了长焦镜头图像与广角镜头图像具有相似的视角，为后续的精确对齐和色彩校正奠定基础。

3.1.3 精确对齐

在双镜头图像对齐过程中，粗略对齐能够初步校正长焦与广角镜头之间的视角差异，但由于两个镜头的位置和焦距不同，仍可能产生细微的透视差异与视差。因此，在粗略对齐的基础上，本实验将使用 DeepFlow 光流算法对图像进行进一步的精确对齐^[10]。光流算法的目标是在不同视角的图像间估计像素级的运动矢量场，通过识别和追踪图像中特征的移动，推测像素的运动轨迹，从而将图像对齐。

本实验使用的 DeepFlow 算法在传统光流技术的基础上，结合了深度卷积网络的多尺度特征提取能力，生成高质量的光流矢量场。首先，该算法提取长焦与广角图像的特征点，计算并匹配两者的特征，以估计出每个像素的运动矢量场。随之利用矢量场的计算结果，将长焦图像中的每个像素移动到与广角镜头对应的位置，从而实现像素级的精确对齐。这一过程能够纠正两个镜头之间的视差与透视误差，确保两幅图像之间的几何一致性。

应用这种精确对齐策略能获得在动态场景下同样稳定的低分辨率-高分辨率数据对，为后续的图像超分辨率和融合任务提供高质量的数据基础。精确对齐过程不仅确保了多镜头图像在所有细节上的几何一致性，还减少了后续图像处理中的错位问题，进一步提高了图像的清晰度和稳定性。

3.1.4 色彩校正

在双镜头图像对齐和融合的过程中，通常会遇到不同镜头由于传感器特性或拍摄条件差异而导致的色彩不一致问题。尤其在高分辨率和低分辨率镜头之间，光学元件和图像传感器的不同可能造成明显的色彩差异。为确保两个镜头捕获的图像在融合过程中保持色彩一致性，需要进行精确的色彩校正。

此过程主要针对每个图像通道采用线性缩放系数来进行校正^[11]。具体而言，对于每个通道的像素值，通过计算高-低分辨率图像对之间的线性关系，确定该通道的缩放系数。这一系数能够有效纠正高-低分辨率图像对之间的亮度、饱和度差异，使两个图像的颜色在相同场景中保持一致。此外，这种线性缩放还能校正不同镜头的白平衡问题，确保拍摄对象的真实色彩得以还原。

由于色彩差异会影响后续的图像特征提取和融合，因此对图像进行精确的色彩校正有助于提高学习模型的训练效果。在高分辨率与低分辨率图像的匹配和超分辨率任务中，准确的色彩一致性确保了模型能够充分利用多镜头图像信息，以获得更高的重建精度和视觉质量。最终，通过线性缩放每个通道的系数来实现高-低分辨率图像对的色彩校正，确保多镜头图像在后续超分辨率任务中能够保持一致的色彩表现，为模型训练提供更可靠的输入数据。

3.1.5 数据裁剪

为了获得能直接用于网络监督学习的低-高分辨率图像对以及参考图像，首先将历经对齐与色彩校正后的广角镜头图像与长焦镜头图像的重叠区域裁剪出来，以确保图像的几何位置一致。

这一过程将裁剪对齐后的长焦镜头图像的中心区域作为直接用于训练的参考图像。这是因为长焦镜头具有较高的分辨率和更狭窄的视野，其中心区域通常是最清晰的部分，能作为参考图像辅助低分辨率图像的超分辨率重建过程。而原始未经处理的长焦镜头图像也需要进行相同的裁剪步骤，以获得用于评估性能的标签。通过这种裁剪和对齐方法，最终可以获得高质量的低分辨率-参考高分辨率-标签高分辨率成对数据（LR-Ref-HR），使模型能够根据参考图像更加准确地学习从低分辨率到高分辨率的映射关系，为双目镜头超分辨率任务提供更精确的训练和测试数据。

3.2 神经网络架构

在 KeDuSR 框架中，构建了一种全面的双镜头超分辨率架构，以处理广角低分辨率和长焦高分辨率相机之间的重叠视场。该框架分别处理图像对的中心区域和角落区域：参考图像首先经过中心区域的对齐，然后利用无核匹配获得的索引对参考图像进行角落区域的对齐，最后通过自适应融合模块和重建模块将变换后的参考特征和低分辨率特征相结合，获得最终的超分辨率重建的结果图像。

3.2.1 中心对齐

在 KeDuSR 框架的中心对齐模块中，为实现高分辨率参考图像（ I_{Ref} ）与

低分辨率输入图像的中心区域 (I_{LRC}) 的精确对齐, 首先需要进行全局的初步对齐, 然后通过局部细化对齐进一步优化。首先, 在参考图像和低分辨率图像之间找到重叠的特征点。通过 SIFT 算法检测并匹配两幅图像的关键点, 随后 RANSAC 算法被用于剔除噪声和异常点, 只保留真实有效的匹配点, 以计算两幅图像之间的单应矩阵。单应矩阵描述了从 I_{Ref} 到 I_{LRC} 的全局变换关系, 应用于参考图像可生成粗略对齐的参考图像 $\overline{I_{Ref}}$ 。

利用该单应矩阵变换, 将低分辨率图像中对应于 $\overline{I_{Ref}}$ 的区域裁剪出来, 标记为低分辨率的中心区域 I_{LRC} 。尽管通过单应矩阵实现了全局对齐, 但仍存在细微的对齐误差, 因此需要利用流引导的可变形卷积网络 (DCN) 进一步细化局部对齐效果^[12]。具体而言, 先将 $\overline{I_{Ref}}$ 下采样到与 I_{LRC} 相同的分辨率, 然后通过预训练的 SpyNet 模型计算两者之间的光流^[13]。通过残差块提取 I_{LRC} 和 $\overline{I_{Ref}}$ 的特征, 生成特征图 F_{LRC} 和 $\overline{I_{Ref}}$ 。最终, 通过上采样的光流引导 DCN 对齐 F_{LRC} 与 $\overline{I_{Ref}}$, 生成精确对齐的参考特征。

这套中心对齐流程确保了高分辨率参考图像的特征与低分辨率输入图像的中心区域充分对齐, 为后续的角度对齐和自适应融合提供高质量的参考特征。

3.2.2 无核匹配与角落对齐

无核匹配与角落对齐是为了解决低分辨率图像 (I_{LR}) 和参考图像 (I_{Ref}) 之间存在的分辨率差异而提出的策略。传统方法直接在低分辨率图像的角落区域和参考图像之间进行图块匹配, 但两者的分辨率差异较大, 且简单的下采样无法准确模拟两台摄像机之间的映射关系, 因此无法解决这种问题。研究人员针对该问题提出了 KernelGAN^[14]或概率降解模型 (PDM)^[15]来学习这种映射关系, 并利用学习所获得的核对参考图像进行降解。然而, 这些方法依赖于特定摄像机的特点, 导致通用性不足, 无法适用其余摄像机。

KeDuSR 的研究发现角落和中心区域之间存在非局部相似性, 低分辨率图像的角落区域可以在中心区域中找到相似的图块。因此, 无核匹配策略旨在利用低分辨率图像的中心区域作为参考, 与角落区域的查询图块进行匹配。在这个过程中, 利用余弦相似度计算图块之间的相似性分数, 生成索引映射和置信度映射。然后根据匹配的索引, 从参考特征中提取对应的高分辨率图块, 以实现角落区域的对齐特征。实验证明, 这种匹配策略在高分辨率和低分辨率域之间的匹配结果一致性较好, 能够有效恢复图像细节。

另外, KeDuSR 通过在 VGG 特征空间中提取 I_{LR} 和 I_{LRC} 的特征来进行无核匹配, 将其分割成密集的 3×3 图块。根据各图块之间的余弦相似度, 能够找到每个查询图块与其最匹配的中心区域图块, 并生成匹配索引和置信度。在参考特征的

对齐结果中，角落区域的特征可以通过中心区域的图块匹配得到，从而充分利用高分辨率参考特征来增强低分辨率图像的细节恢复效果。

3.2.3 单图像超分辨率编码器

单图像超分辨率（Single Image Super-Resolution，简称 SISR）编码器作为 KeDuSR 网络中的关键模块，用于将输入低分辨率图像的超分辨率重建与从参考图像中进行纹理迁移的任务相互分离。实验结果显示直接将这两项任务结合在一起会导致干扰并降低网络性能，因此，SISR 编码器被设计成一个独立模块，以确保对低分辨率图像的特征进行高效提取。

该模块由 24 个带有通道注意力机制的残差块组成，可以有效提取低分辨率图像的特征，并通过上采样使其与参考特征的尺寸保持一致。提取得到的特征被命名为 F_{LR} ，与两阶段的训练方式不同，该 SISR 编码器作为整个网络的重要组成部分进行了整体训练，确保了与其他模块的有效协作和特征融合，从而最终实现高质量的图像超分辨率重建。

3.2.4 自适应融合模块

在 KeDuSR 网络中，对齐后的参考特征 \tilde{F}_{Ref} 与低分辨率特征 F_{LR} 通过自适应融合模块（AdaFusion）进行融合。利用无核匹配过程中生成的置信度映射指导融合过程，确保从 \tilde{F}_{Ref} 的高频形式（ \tilde{F}_{hf}^{Ref} ）提取的高频细节可以补充低分辨率图像的特征。这一模块由残差块组成，同时结合了空间和通道注意力机制，通过高效的特征融合为最终的超分辨率重建输出 I^{SR} 提供准确且丰富的细节信息。

完整的 KeDuSR 网络架构如图 3-1 所示：

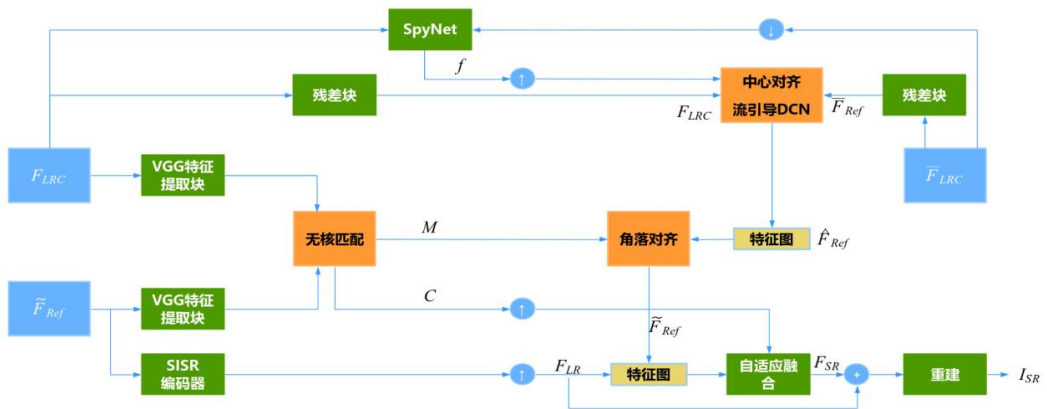


图 3-1 KeDuSR 完整网络架构

在图示结构中，无核匹配在低分辨率图像角落和中心区域之间进行，以生成

索引映射 M 和置信度映射 C 。然后利用中心对齐和角落对齐的方法，可以获得参考图像的对齐高分辨率特征图 \tilde{F}_{Ref} 。通过自适应融合模块，将对齐的参考特征与低分辨率图像特征 F_{LR} 进行融合，最终生成超分辨率重建结果 I_{SR} 。

3.3 损失函数

KeDuSR 采用混合损失函数来训练网络，包括重建损失、感知损失、对抗损失以及自定义的噪声损失。

3.3.1 重建损失

重建损失函数作为图像重建的主要优化目标，直接在低分辨率与高分辨率图像之间进行计算，且应赋予最大的权重。重建损失的计算采用 Charbonnier 损失，其计算方法如下：

$$\mathcal{L}_{ch} = \sqrt{\|I_{HR} - I_{SR}\|_2^2 + \varepsilon} \quad (3-1)$$

其中 $\varepsilon = 1 \times 10^{-6}$ ，以确保自变量为零时的可微性，缓解梯度消失问题。 I_{HR} 代表高分辨率图像，也就是标签值。 I_{SR} 代表超分辨率重建结果，Charbonnier 损失函数通过计算并降低这二者的差异，来实现图像重建效果。对比 L2 损失，该损失函数的梯度更加平滑，并且由于根号的存在，使得其对图像异常值的敏感性降低。

3.3.2 感知损失与对抗损失

将感知与对抗损失函数用作优化目标是为了使重建的图像拥有更好的视觉效果，其中感知损失（perceptual loss）的计算方法如下^[2]：

$$\mathcal{L}_{per} = \|\Phi_i(I_{HR}) - \Phi_i(I_{SR})\|_2 \quad (3-2)$$

公式（3-2）中 Φ_i 表示 VGG19 网络的第 i 层。VGG19 是一种预训练好的深度卷积神经网络，通过使用 VGG19 逐层提取特征并进行比较计算，感知损失能有效重建图像的高频细节。对抗损失（adversarial loss）引入了额外的一组生成对抗网络进行计算，计算结果以 \mathcal{L}_{adv} 进行表示^[16]。

3.3.3 噪声损失

由于本项目的研究目标之一是在超分辨率重建过程中引入去噪功能，且同时由于没有干净的数据样本，因此需要设计新的用于减少噪声的损失函数。借鉴感

知与对抗损失函数的计算都引入了额外的网络，本实验定义噪声损失的计算将引入的额外的去噪网络。

卷积盲去噪网络（Convolutional Blind Denoising Network，简称 CBDNet）^[17] 针对以往去噪模型在真实噪声上应用效果差的问题，首次提出对更接近实际噪声的泊松噪声进行滤除，使得去噪不再局限于较理想化高斯噪声。传统的去噪网络性能很大程度上取决于合成噪声与实际噪声是否匹配，因此 CBDNet 的去噪过程分为两个阶段，第一阶段利用噪声估计子网络进行估计噪声强度，第二阶段使用非盲去噪自网络将噪声估计结果与带噪图像一并作为输入进行非盲去噪。

对于数据集，CBDNet 在由人工合成的噪声图像和真实拍摄的自然图像组成的混合数据集上进行训练，同时采用由重建损失、非对称损失以及全变分损失组成的混合损失函数作为优化目标。相比其它去噪神经网络，CBDNet 做到了性能与不足的平衡：它在能有效移除噪声的同时，最大程度减少细节模糊等问题。图 3-2 为图像通过 CBDNet 去噪前后的结果：

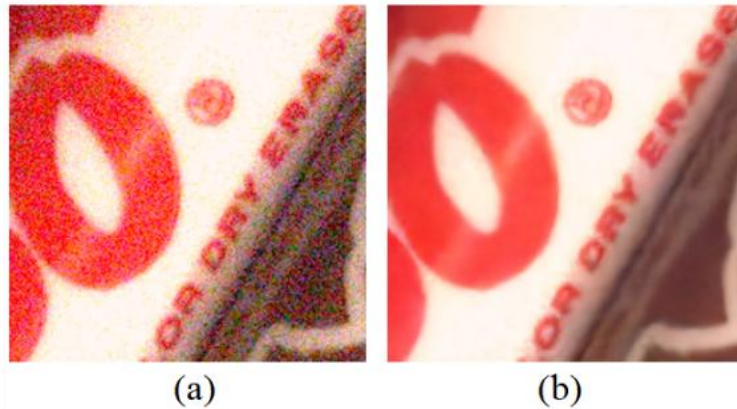


图 3-2 图像通过 CBDNet 前后对比

(a) 原始噪声图像；(b) 通过 CBDNet 去噪后的图像

图 3-2 清楚地反应了图片经过 CBDNet 前后的噪声滤除效果，同时图像的原始信息极大程度被保留，缓解了普遍存在于去噪网络中的细节模糊问题，因此本实验对噪声损失的设计将引入 CBDNet。

本实验采用的噪声损失（noise loss）定义如下：将 KeDuSR 超分辨率重建结果 I_{SR} 通过一个预训练的 CBDNet，计算 I_{SR} 通过网络前后的像素级误差。同时借鉴 Charbonnier 损失的表达形式，为误差结果加上微小常量 $\varepsilon = 1 \times 10^{-6}$ 并开根号，其表达式如下所示：

$$\mathcal{L}_{no} = \sqrt{\|I_{SR1} - I_{SR2}\|_2^2 + \varepsilon} \quad (3-3)$$

其中 I_{SR1} 、 I_{SR2} 分别代表通过 CBDNet 前后的图像，其采用了一种间接的去噪逻辑：如果原 KeDuSR 拥有良好的去噪功能，则 I_{SR1} 理应含有很少的噪声，通过去噪网络后噪声降低的幅度很小，而 \mathcal{L}_{no} 通过计算前后差值，其结果理应很小；反之，如果原 KeDuSR 去噪功能很弱，则 I_{SR1} 含有较多的噪声，这将给去噪网络更大“发挥”的空间，其通过去噪网络后的噪声降低幅度很大，导致 \mathcal{L}_{no} 计算值偏大。最终，去噪损失通过优化重建结果 I_{SR} 经过 CBDNet 前后的噪声减小幅度，间接迫使原 KeDuSR 超分辨率网络的去噪性能得到提高。

最终，给不同的损失函数赋予不同的权重后，得到的混合损失表达式如下：

$$\mathcal{L} = \mathcal{L}_{ch} + \lambda_1 \mathcal{L}_{per} + \lambda_2 \mathcal{L}_{adv} + \lambda_3 \mathcal{L}_{no} \quad (3-4)$$

其中， λ_1 、 λ_2 、 λ_3 作为权重超参数，需要反复实验确定最佳取值。

第四章 实验与分析

4.1 数据准备

为了研究超分辨率网络的去噪性能，首先需要对数据集进行更新。用于 KeDuSR 训练的原始数据集（Dual-Real）一共包含 380 对训练数据与 40 对测试数据，均在阳光充足的条件下拍摄，因此噪声表现非常微弱^[5]。图 4-1 为 Dual-Real 数据集的样本示例：



图 4-1 Dual-Real 数据集的样本示例

(a) 广角镜头拍摄的 LR 图像；(b) 长焦镜头拍摄的 HR 图像

为了引入噪声，我们在光照条件不好的情况下重新拍摄了 450 组数据，其中阴天拍摄占比 20%，傍晚拍摄占比 60%，夜晚拍摄占比 20%，并将该带噪的数据集命名为 Dual-Real-Noise。图 4-2 列举了 Dual-Real-Noise 数据集的样本示例：

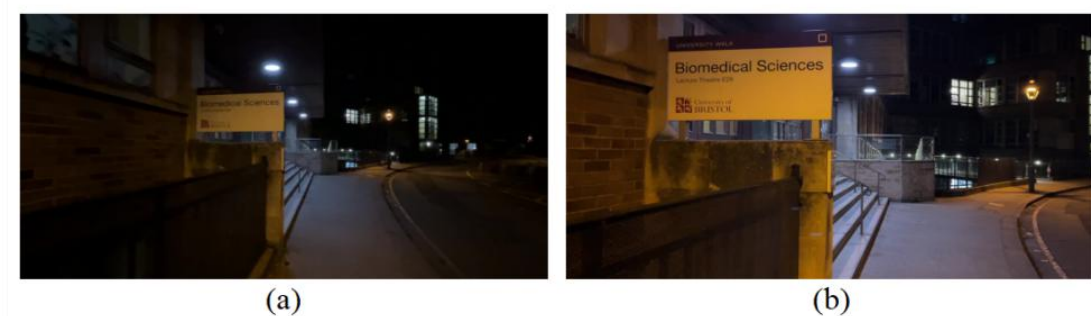


图 4-2 Dual-Real-Noise 数据集的样本示例

(a) 广角镜头拍摄的 LR 图像；(b) 长焦镜头拍摄的 HR 图像

随后我们采用了 KeDuSR 的数据处理方法对 Dual-Real-Noise 中的样本进行数据处理，包括粗对齐、精对齐、色彩校正与裁剪等操作。在这一过程中我们移

除了不匹配的图像共 24 对，剩余匹配良好的 426 对数据被划分成 380 对训练数据与 46 对测试数据。

为了进一步训练网络的去噪能力，对数据进行以下额外的处理：首先在低分辨率输入图像上（LR）加上幅度很小的泊松噪声，以拉开与标签图像（HR）的噪声水平差距，提高网络的去噪能力学习空间，同时加入的噪声幅度很小是为了确保不能影响超分辨率中的无核匹配过程；其次对参考图像（Ref）进行去噪处理，让其通过 CBDNet，减少其噪声水平；最后在原来的输入数据里额外增加一个输入，即标签图像（HR）经过 CBDNet 去噪后的干净的高分辨率标签图像（HR_noise_removed）。

最终得到处理过的数据如表 4-1 所示：

表 4-1 数据集构成

训练数据集（train）		测试数据集（test）	
英文简称	中文含义	英文简称	中文含义
LR_noise	低分辨率图像加入噪声	LR	低分辨率图像
Ref_full_noise_removed	参考图像经过去噪处理	LR_center	低分辨率图像中心区域
HR	高分辨率标签图像	Ref_SIFT	匹配后的参考图像
HR_noise_removed	高分辨率去噪标签图像	HR	高分辨率标签图像

在网络利用训练数据集进行训练时，LR_noise、Ref_full_noise_removed 用作网络输入以生成重建图像 SR；HR、HR_noise_removed 用作标签值进行性能评估以及损失值计算。图 4-3 展示了一组经过数据处理后的低分辨率图像（LR_noise）和参考图像（Ref_full_noise_removed）：

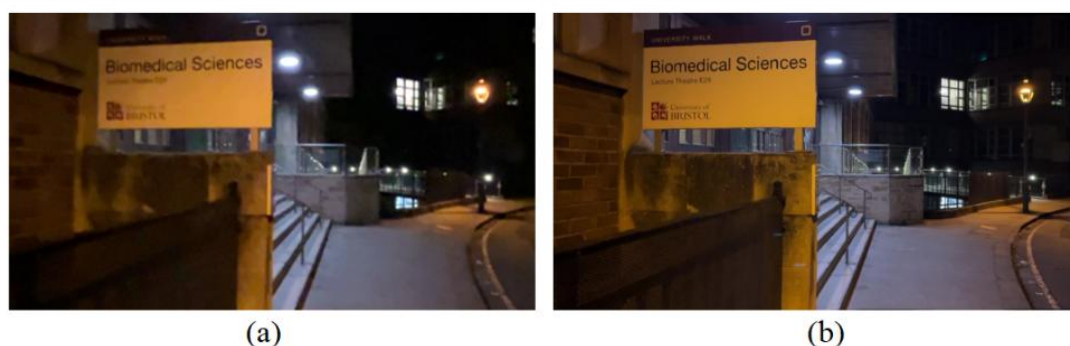


图 4-3 KeDuSR 网络输入图像示例

(a) LR_noise; (b) Ref_full_noise_removed

图 4-3 所示的一组图像数据可被直接输入至 KeDuSR 网络中，KeDuSR 将根据（b）图所示的参考图像对（a）图所示的低分辨率图像进行超分辨率重建。

4.2 不同损失函数对比实验

为方便后续性能评估，我们在训练 KeDuSR 网络同时，也同步训练了 DCSR 网络。同时为了比较各部分损失函数的作用，设计了以下对照实验，以分别比较不同网络结构、不同损失函数条件下的网络超分辨率与去噪性能：

表 4-2 实验分类

实验名称	网络结构	损失函数
实验一	KeDuSR	仅使用 \mathcal{L}_{ch}
实验二		使用 \mathcal{L}_{ch} 、 \mathcal{L}_{per} 、 \mathcal{L}_{adv}
实验三		使用 \mathcal{L}_{ch} 、 \mathcal{L}_{per} 、 \mathcal{L}_{adv} 、 \mathcal{L}_{no}
实验四	DCSR	仅使用 \mathcal{L}_{ch}
实验五		使用 \mathcal{L}_{ch} 、 \mathcal{L}_{per} 、 \mathcal{L}_{adv}

其中 \mathcal{L}_{ch} 代表重建损失函数，它计算超分辨率重建结果 SR 与 HR 图像之间的差异。 \mathcal{L}_{per} 代表感知损失函数， \mathcal{L}_{adv} 代表对抗损失函数，它们计算 SR 与经历去噪的 HR_noise_removed 之间的差异，这里使用去噪的标签值是因为噪声表现为高频形式的条形状、小块状，对于利用 \mathcal{L}_{per} 、 \mathcal{L}_{adv} 恢复图像细节不利。

对于噪声损失 \mathcal{L}_{no} 做出以下改进，在原来间接减少 SR 噪声的基础上，增加了 SR 与 HR_noise_removed 之间的差异计算，以直接保证网络的去噪能力。其表达式如下：

$$\mathcal{L}'_{no} = \sqrt{\|I_{SR1} - I_{SR2}\|_2^2 + \varepsilon} + \sqrt{\|I_{SR1} - I_{HR_{noi_rem}}\|_2^2 + \varepsilon} \quad (4-1)$$

其中 I_{SR1} 表示超分辨率重建结果， I_{SR2} 表示 I_{SR1} 经过去噪网络后的结果， $I_{HR_{noi_rem}}$ 表示经过去噪处理的高分辨率标签图像， ε 为微小常量。最终的混合损失函数表达式如下所示：

$$\mathcal{L} = \mathcal{L}_{ch} + \lambda_1 \mathcal{L}_{per} + \lambda_2 \mathcal{L}_{adv} + \lambda_3 \mathcal{L}'_{no} \quad (4-2)$$

调整参数为 $\lambda_1 = 0.005$ ， $\lambda_2 = 1 \times 10^{-10}$ ， $\lambda_3 = 0.8$ 。其中重建损失 \mathcal{L}_{ch} 系数设置为 1，因为重建图像是主要任务，应赋予最高的权重；噪声损失权重设置较高，是因为其表达式的第二项与超分辨率重建过程拥有同样的目标，但由于标签进行了去噪，避免不了微弱的细节模糊问题，因此权重不能高于重建损失的权重。

4.3 实验与结果比较

根据表 4-2 的实验分类，使用不同的网络结构与损失函数组合进行训练。实验的相关设置如下：批量大小 (batch size) 设置为 4，图像块尺寸大小 (patch size) 设置为 128×128 ，初始学习率 (learning rate) 设置为 10^{-4} ，并逐步衰减至 10^{-6} ，优化器 (optimizer) 采用 Adam 优化器。所有实验都 Nvidia GeForce RTX 3090 GPU 上使用 PyTorch 进行，且在每次实验训练完成后，保存训练的最佳模型权重。

4.3.1 性能指标对比

依据 PSNR 和 SSIM 评价指标，记录五次实验的指标结果，绘制表格如下：

表 4-3 实验指标结果

实验名称	PSNR (dB)	SSIM
实验一	25.605944	0.828641
实验二	25.217181	0.803758
实验三	25.356210	0.813623
实验四	24.442584	0.786661
实验五	24.154373	0.764157

实验一~三采用 KeDuSR 网络结构，其中实验一仅使用重建损失，由于重建损失直接考虑图片之间的像素级差异，其指标最终保持在最高水平。在加入了提升视觉效果的感知损失与对抗损失后，PSNR 与 SSIM 的值有小幅下降。在继续加入自定义的噪声损失后，PSNR 上升约 0.139dB，SSIM 上升约 0.987%。

实验四、五采用 DCSR 网络结构，分别使用与实验一、二相同的损失函数进行试验。在仅使用重建损失时，KeDuSR 比 DCSR 的 PSNR 指标高出约 1.163dB，SSIM 指标高出 4.2%；在使用重建损失、感知损失与对抗损失后，KeDuSR 比 DCSR 的 PSNR 指标高出约 1.063dB，SSIM 指标高出约 3.97%。

4.3.2 视觉效果对比

本节于测试数据集中选取了对比明显的几组结果进行视觉比较。其中选取的对比图像简称以及含义如表 4-4 所示：

表 4-4 实验视觉结果分类

图片简称	图片含义
LR	原始低分辨率输入全局图像
LR_detail	原始低分辨率输入细节图像
SR_1	实验一结果 (KeDuSR 、 \mathcal{L}_{ch})
SR_2	实验二结果 (KeDuSR 、 \mathcal{L}_{ch} 、 \mathcal{L}_{per} 、 \mathcal{L}_{adv})
SR_3	实验三结果 (KeDuSR 、 \mathcal{L}_{ch} 、 \mathcal{L}_{per} 、 \mathcal{L}_{adv} 、 \mathcal{L}'_{no})
SR_4	实验五结果 (DCSR 、 \mathcal{L}_{ch} 、 \mathcal{L}_{per} 、 \mathcal{L}_{adv})
HR_detail	高分辨率标签细节图像

选取的视觉对比图如图 4-4 所示：

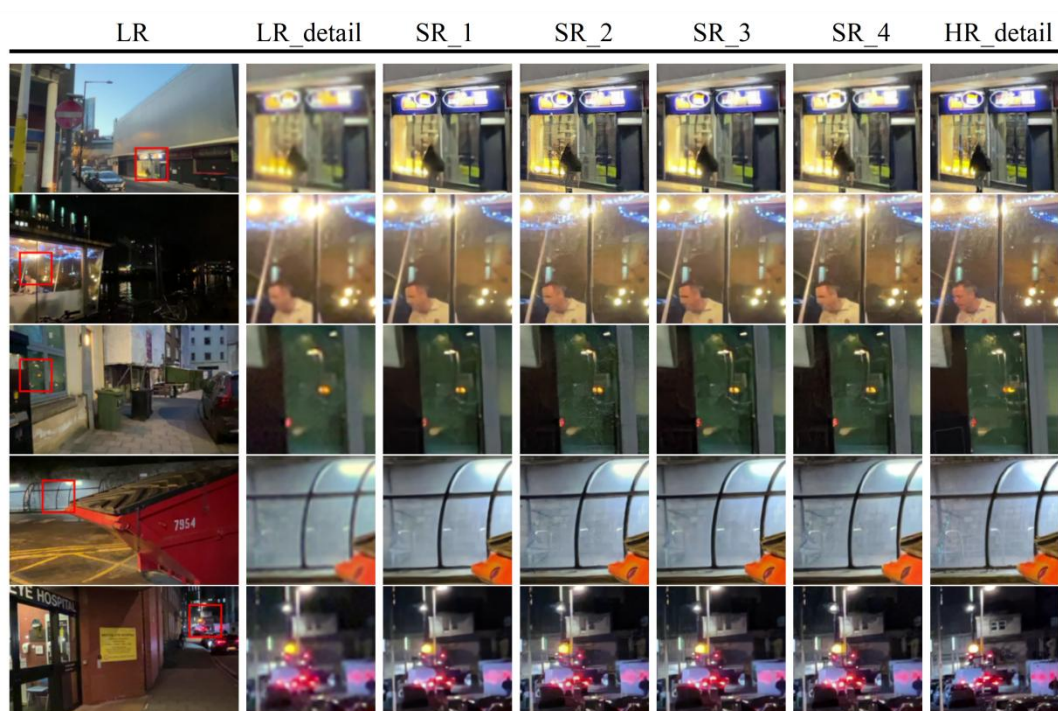


图 4-4 视觉效果对比

4.4 实验结果分析与总结

本实验更新原本的数据集 (Dual-Real) 为带噪数据集 (Dual-Real-Noise)，并采用两种网络结构 (KeDuSR、DCSR)、多种损失函数组合 (\mathcal{L}_{ch} 、 \mathcal{L}_{per} 、 \mathcal{L}_{adv} 、 \mathcal{L}'_{no}) 进行多次实验，以证明 KeDuSR 网络结构的优越性以及自定义的噪声损失函数 \mathcal{L}'_{no} 的有效性。根据指标结果表 4-3 所示，仅利用重建损失进行训练，通过

减少超分辨率重建结果与标签图像的像素级差异，可以实现较高的 PSNR 与 SSIM 值，能成功实现图像重建的功能。

图 4-5 进一步从视觉上比较了引入重建损失训练前后的结果：

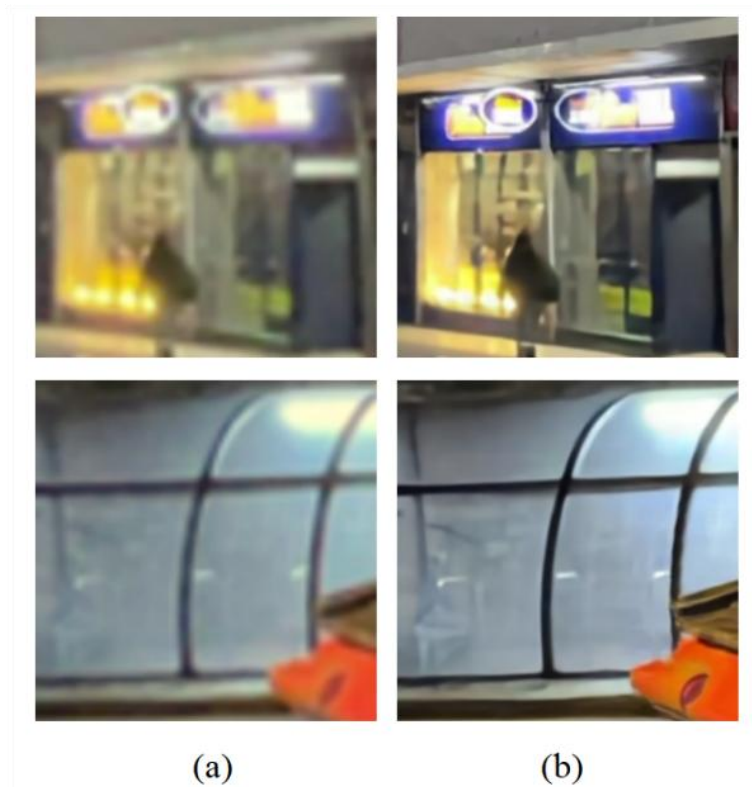


图 4-5 引入重建损失效果对比

(a) 低分辨率输入图像；(b) 使用 \mathcal{L}_{ch} 超分辨率重建结果

图 4-5 中 (b) 图的超分辨率重建结果图明显比 (a) 图的输入图像更清晰，无论是图片的整体质量还是细节纹理均得到显著提高。

但由于人类感官关心图像的高频信息，因此引入了感知损失与对抗损失去恢复图像的纹理细节，以实现更高的感官质量。在引入这部分的损失函数之后，PSNR 与 SSIM 均有小幅下降，但图像细节明显恢复，感官质量得到提高，图 4-6 对比了引入感知损失与对抗损失前后的视觉效果：

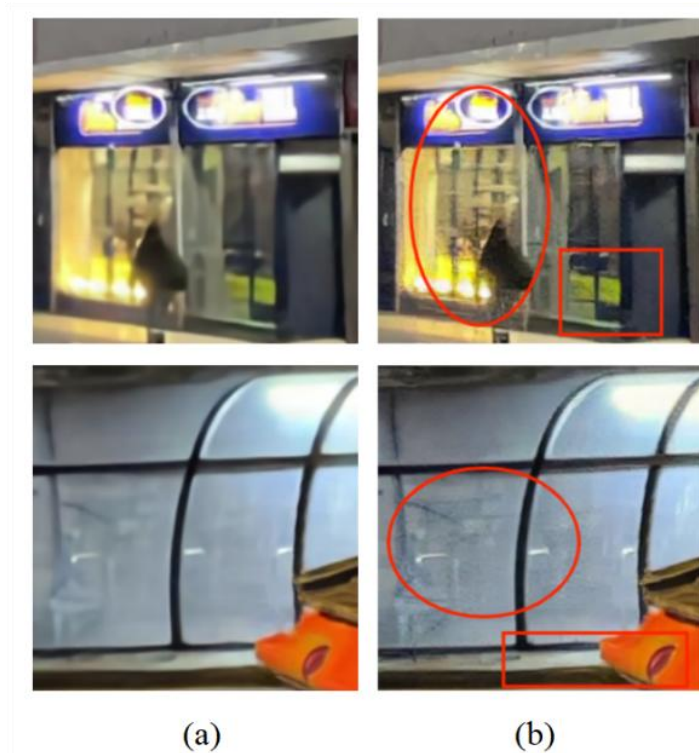


图 4-6 引入感知、对抗损失效果对比

(a) 仅使用 \mathcal{L}_{ch} 超分辨率重建结果；(b) 引入 \mathcal{L}_{per} 、 \mathcal{L}_{adv} 超分辨率重建结果

在图 4-6 中，引入感知、对抗损失函数后的超分辨率重建结果图像纹理细节质量得到了提高，如图（b）中红色方框部分的线条形状物体更清晰，同时噪声变得更加明显，如图（b）中红色椭圆框部分有明显的噪点。出现这一现象的原因除了本身的数据均含有噪声外，感知、对抗损失的存在也会恢复一部分高频的噪声信息。因此，为了使得超分辨率网络性能更加泛用，需要引入用于减少噪声的损失函数。

在加入自定义的噪声损失函数并重新训练网络结构后，PSNR 与 SSIM 指标均有小幅上升，并且视觉效果上噪点得到了明显的滤除，同时极大程度上缓解了普遍存在于去噪网络中的细节模糊问题。图 4-7 对比了引入噪声损失前后的视觉效果，其中引入噪声损失的网络训练结果明显有效滤除了噪声，如图 4-7（b）中红色方框所示，噪点数量大幅减少，同时图像细节等部分并未受到太大影响。

最后为了证明 KeDuSR 网络结构的优越性，实验也对目前性能第二好的网络 DCSR 进行了训练。性能指标表 4-4 表明使用 DCSR 训练的 PSNR 指标低于 KeDuSR 约 1dB，SSIM 指标低约 4%。同时图 4-8 对比了使用相同损失函数下 KeDuSR 与 DCSR 的超分辨率重建结果。



图 4-7 引入噪声损失效果对比

(a) 使用 \mathcal{L}_{ch} 、 \mathcal{L}_{per} 、 \mathcal{L}_{adv} 超分辨率重建结果；(b) 引入 \mathcal{L}'_{no} 超分辨率重建结果

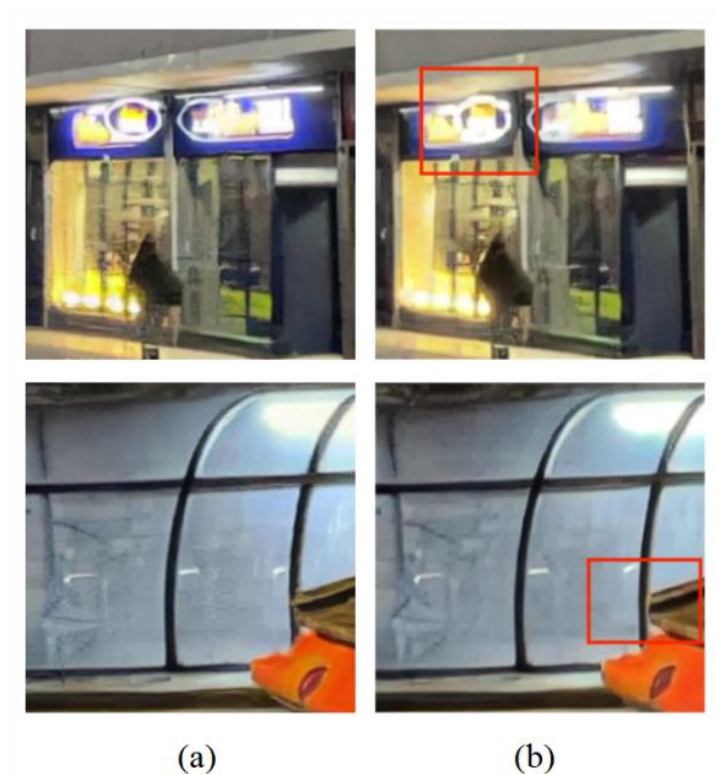


图 4-8 KeDuSR 与 DCSR 效果对比

(a) KeDuSR 超分辨率重建结果；(b) DCSR 超分辨率重建结果

图 4-8 在使用同样的重建、感知、对抗损失函数的条件下，对比了 KeDuSR 与 DCSR 两种网络的超分辨率重建性能。观察结果可知，使用 DCSR 的图像重建、细节恢复效果差于 KeDuSR，如图 4-8（b）中红色方框所示，细节模糊程度明显更大，同时结合 PSNR 与 SSIM 评价结果可知：KeDuSR 的网络结构优于 DCSR。

最终，通过多次实验得到的性能指标结果，结合视觉质量可以得出：KeDuSR 网络结构在超分辨率重建任务上的性能领先 DCSR，其利用得无核匹配方法具有可行性；同时，实验设计的噪声损失函数能在网络进行超分辨率的同时，有效训练其去噪能力，进一步增强了网络的泛用性。

第五章 总结与展望

5.1 课题总结

本课题旨在研究双目镜头联合超分辨率与去噪。为了考察噪声对超分辨率过程的影响，我们更新了原始干净数据集为带噪数据集；为了克服在无干净样本条件下训练超分辨率网络去噪能力的难题，我们借助额外的盲去噪网络定义用于减少噪声的损失函数，并对数据集进行了相关数据处理工作。指标结果与视觉结果均表明：噪声会显著影响超分辨率重建过程，而自定义的噪声损失函数能有效训练超分辨率网络的去噪能力。这些创新显著提高了模型在复杂环境下的实用性和泛用性。

5.2 未来展望

本课题虽然成功实现了联合超分辨率与去噪的研究目标，但仍有一些不足之处：一是模型训练难度大，收敛偏慢，其定义的噪声损失需要额外的去噪网络辅助计算，计算量偏大；二是网络结构复杂，参数量庞大，应用的时长偏长，实时性差；三是个别复杂的测试图像出现了轻微的伪影，影响了图片观感。

后续的工作可以围绕以下方向进行：一是探索轻量化的网络结构，减少参数量，提高模型应用的实时性；二是探索更高效的噪声损失函数，进一步加强模型在超分辨率重建过程中的识别噪声、滤除噪声的能力，同时尽可能减少其计算量以降低模型训练难度；三是继续探索更先进的图像匹配算法，为双目镜头超分辨率领域提供更简洁高效的数据处理方法。

5.3 设计方案经济成本分析与对社会环境影响

在环境保护与可持续发展方面，本实验采用纯软件模拟的方式进行。实施过程不会产生任何污染物或环境损害，对人类社会的可持续发展不存在威胁，也不会对人类健康造成任何损害。

在经济成本方面，由于本实验需要高性能计算资源，本课题实际租用远程 GPU 进行实验操作，因此产生了一定的租用费用。尽管如此，远程 GPU 的租用使得实验可以高效进行，并且相比购买昂贵硬件，其成本相对较低，因此这部分费用在课题预算内是可控且合理的。

在社会安全及法律方面，本课题在实际应用中具有一定的现实意义，可以在特定条件下发挥重要作用。实验过程中基本不存在安全隐患问题，所有实验内容均严格遵守国家法律法规。

参考文献

- [1] Dong C, Loy C C, He K, et al. Image super-resolution using deep convolutional networks [J]. arXiv preprint arXiv:1501.00092, 2015: 1-12.
- [2] Ledig C, Theis L, Huszar F, et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network [J]. arXiv preprint arXiv:1609.04802, 2017: 1-9.
- [3] Yue H, Sun X, Yang J, et al. Landmark Image Super-Resolution by Retrieving Web Images [J]. IEEE Transactions on Image Processing, 2013: 4865-4878.
- [4] Wang T, Xie J, Sun W, et al. Dual-Camera Super-Resolution with Aligned Attention Modules [J]. arXiv preprint arXiv:2109.01349, 2021: 1-8.
- [5] Yue H, Cui Z, Li K, et al. KeDuSR: Real-World Dual-Lens Super-Resolution via Kernel-Free Matching [J]. arXiv preprint arXiv:2312.17050, 2024: 1-7.
- [6] Zhang K, Zuo W, Chen Y, et al. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising [C]. IEEE Transactions on Image Processing, 2017: 3142-3155.
- [7] Lehtinen J, Munkberg J, Hasselgren J, et al. Noise2Noise: Learning Image Restoration without Clean Data [J]. arXiv preprint arXiv:1803.04189, 2018: 1-8.
- [8] Lowe D G. Distinctive Image Features from Scale-Invariant Keypoints [J]. International Journal of Computer Vision, 2004, 60: 91-110.
- [9] Fischler M A, Bolles R C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography [C]. Readings in Computer Vision, 1987: 726-740.
- [10] Weinzaepfel P, Revaud J, Harchaoui Z, et al. DeepFlow: Large Displacement Optical Flow with Deep Matching [C]. 2013 IEEE International Conference on Computer Vision, 2013: 1385-1392.
- [11] Yue H, Zhang Z, Yang J. Real-RawVSR: Real-World Raw Video Super-Resolution with a Benchmark Dataset [J]. arXiv preprint arXiv:2209.12475, 2022: 1-14.

- [12] Chan K C K, Zhou S, Xu X, et al. BasicVSR++: Improving Video Super-Resolution with Enhanced Propagation and Alignment [J]. arXiv preprint arXiv:2104.13371, 2021: 1-8.
- [13] Ranjan A, Black M J. Optical Flow Estimation using a Spatial Pyramid Network [J]. arXiv preprint arXiv:1611.00850, 2016: 1-8.
- [14] Bell-Kligler S, Shocher A, Irani M. Blind Super-Resolution Kernel Estimation using an Internal-GAN [J]. arXiv preprint arXiv:1909.06581, 2020: 1-8.
- [15] Luo Z, Huang Y, Li S, et al. Learning the Degradation Distribution for Blind Image Super-Resolution [J]. arXiv preprint arXiv:2203.04962, 2024: 1-8.
- [16] Jolicoeur-Martineau A. The relativistic discriminator: a key element missing from standard GAN [J]. arXiv preprint arXiv:1807.00734, 2018: 1-12.
- [17] Guo S, Yan Z, Zhang K, et al. Toward Convolutional Blind Denoising of Real Photographs [J]. arXiv preprint arXiv:1807.04686, 2019: 1-8.
- [18] Zhang K, Zuo W, Zhang L. FFDNet: Toward a Fast and Flexible Solution for CNN-Based Image Denoising [J]. IEEE Transactions on Image Processing, 2018, 27(9): 4608–4622.
- [19] Dong C, Loy C C, Tang X. Accelerating the Super-Resolution Convolutional Neural Network [J]. arXiv preprint arXiv:1608.00367, 2016: 1-15.
- [20] Shi W, Caballero J, Huszár F, et al. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network [J]. arXiv preprint arXiv:1609.05158, 2016: 1-8.
- [21] Zheng H, Ji M, Han L, et al. Learning Cross-scale Correspondence and Patch-based Synthesis for Reference-based Super-Resolution [C]. British Machine Vision Conference, 2017: 1-10.
- [22] Zheng H, Ji M, Wang H, et al. CrossNet: An End-to-end Reference-based Super Resolution Network using Cross-scale Warping [J]. arXiv preprint arXiv:1807.10547, 2018: 1-14.

致 谢

毕业将至，纵使万千思绪涌动，到落笔时也唯有“感谢”二字。

感谢天津大学给予的优良的学风环境与先进的学术资源，在这里学习与成长的经历不仅丰富了我的学术知识，为我的毕业设计的完成提供坚实的知识储备，更塑造了我独立思考、追求卓越的学术精神。

感谢我的导师岳焕景老师，她不仅具有深厚的学术功底与广博的专业知识，还具有和蔼可亲的态度和无私的指导精神。每一次的学术探讨，她总能准确指出我的问题，同时提出建设性的建议。

感谢我的学长崔子凡同学，他对该研究领域了解全面、独到，在我的研究遇到瓶颈时总能及时分享他的研究经验，帮助我扫清了很多研究中遇到的困难。

最后感谢所有的知识分子，你们每次研究成果的累积，铸就了如今辉煌的人类文明；你们探索与求知的精神，将指引我继续前行。