

学校代码: 10284  
分类号: TP391.4  
密 级: 公开  
U D C: 004.932.2  
学 号: MG21230082



南京大学

# 硕士学位论文

论文题目 高光谱图像光源估计  
及其应用

作者姓名 王乙卜  
专业名称 信号与信息处理  
研究方向 人工智能与类脑视觉  
导师姓名 曹汛 教授

2024年5月24日

答辩委员会主席 王少尉 教授

评 阅 人 王 健 教授

张丽敏 教授

马 展 教授

论文答辩日期 2024年5月24日

研究生签名：王少尉

导师签名：曹讯

# **Hyperspectral Image Illumination Estimation and Applications**

by

**Yibo Wang**

Supervised by

**Professor Xun Cao**

A dissertation submitted to

the graduate school of Nanjing University

in partial fulfilment of the requirements for the degree of

**MASTER**

in

Signal and Information Processing



School of Electronic Science and Engineering

Nanjing University

May 24, 2024



## 南京大学学位论文原创性声明

本人郑重声明，所提交的学位论文是本人在导师指导下独立进行科学研究所取得的成果。除本论文中已经注明引用的内容外，本论文不包含其他个人或集体已经发表或撰写过的研究成果，也不包含为获得南京大学或其他教育机构的学位证书而使用过的材料。对本文的研究做出重要贡献的个人和集体，均已在论文的致谢部分明确标明。本人郑重申明愿承担本声明的法律责任。

研究生签名：\_\_\_\_\_

日期：\_\_\_\_\_



# 南京大学研究生毕业论文中文摘要首页用纸

毕业论文题目： 高光谱图像光源估计及其应用

信号与信息处理 专业 2021 级硕士生姓名： 王乙卜

指导教师（姓名、职称）： 曹汛 教授

## 摘 要

在不同光源下拍摄的 RGB 图像的整体颜色会随着光源改变，对于具有更丰富的光谱特征的高光谱图像，光源的影响则更为显著。光源的光谱特征可能混淆物质原有的光谱特征，为了得到物质本征的光谱特征，需要获取光源的光谱并对高光谱图像进行校正，而光源估计则是一种间接获取光源光谱的方法，在无法准确测量光源光谱的场景中有着重要的作用。

本文围绕高光谱图像的光源光谱估计问题及其应用展开研究。首先预测高光谱图像的光源光谱，然后使用光源光谱引导 RGB 图像的动态范围压缩，最后使用光源光谱对 RGB 图像进行校正并用于材质分割。主要工作和创新点包括以下几个方面：

1、光源光谱估计。构建了真实而非仿真的高光谱光源估计数据集；提出了光谱联合空间自注意机制，用于突出重要的光谱通道和图像中有效区域的作用；提出了平滑约束的损失函数，避免了不实际的病态结果，从而提升了光源估计的准确度。

2、光源光谱引导的图像增强。构建了 RGB 图像与高光谱图像对齐的图像增强数据集；提出了局部亮度适应法，调节高动态范围 RGB 图像不同区域的亮度；提出了光谱感知自注意机制，将光源光谱嵌入到 RGB 特征图中，提升了 RGB 图像色调映射的效果。

3、光源光谱校正的材质分割。构建了 RGB 图像与高光谱图像对齐的材质分割数据集；提出了光谱联合 RGB 分解模型，使用光源光谱校正 RGB 图像的反射率，有效提高了 RGB 图像的材质分割精度。

**关键词：**高光谱图像；光源估计；图像增强；材质分割



# 南京大学研究生毕业论文英文摘要首页用纸

THESIS: Hyperspectral Image Illumination Estimation and Applications

---

SPECIALIZATION: Signal and Information Processing

---

POSTGRADUATE: Yibo Wang

---

MENTOR: Professor Xun Cao

---

## ABSTRACT

The overall color of RGB images shot under different illuminations will change with the illumination. For hyperspectral images with richer spectral characteristics, the impact of the illumination is more significant. The spectral characteristics of the illumination may confuse the original spectral characteristics of the substance. In order to obtain the intrinsic spectral characteristics of a substance, it is necessary to obtain the spectrum of the illumination and correct the hyperspectral image. Illumination estimation is an indirect method to obtain the illumination spectrum, which plays an important role in scenes where the illumination spectrum cannot be accurately measured.

This paper focuses on the problem of illumination spectrum estimation for hyperspectral images and applications. Firstly the illumination spectrum of the hyperspectral image is predicted, then the illumination spectrum is used to guide the image enhancement of the RGB image, and finally the illumination spectrum is used to correct the RGB image for material segmentation. The main work and innovation points include the following aspects:

1. Illumination spectrum estimation. A real rather than simulated hyperspectral illumination estimation data set is constructed. The spectral joint spatial self-attention mechanism is proposed to highlight the important spectral channels and effective areas in the hyperspectral image. A smoothing constrained loss function is proposed to avoid unrealistic ill-conditioned results, improving the accuracy of illumination estimation.

2. Image Enhancement guided by illumination spectrum. An image enhancement data set in which RGB images and hyperspectral images are aligned is constructed. The

## **ABSTRACT**

---

local brightness adaptation method is proposed to adjust the brightness of different areas of High Dynamic Range (HDR) RGB images. The Spectral Perception Self-Attention (SPSA) mechanism is proposed to embed the illumination spectrum into the RGB feature, improving the tone-mapping performance of RGB images.

3. Material segmentation with illumination spectrum correction. A material segmentation data set in which RGB images and hyperspectral images are aligned is constructed. The spectral joint RGB decomposition model is proposed, using the illumination spectrum to correct the reflectance of RGB images, effectively improving material segmentation accuracy of RGB images.

**KEYWORDS:** Hyperspectral Image; Illumination Estimation; Image Enhancement; Material Segmentation

# 目 录

<b>第一章 绪论</b> . . . . .	1
1.1 研究的背景及意义 . . . . .	1
1.2 国内外研究现状及存在的问题 . . . . .	3
1.2.1 国内外研究现状 . . . . .	3
1.2.2 研究存在的问题 . . . . .	5
1.3 论文研究内容及章节安排 . . . . .	6
 <b>第二章 光源光谱估计</b> . . . . .	9
2.1 引言 . . . . .	9
2.2 相关研究 . . . . .	11
2.2.1 基于学习的光源估计 . . . . .	11
2.2.2 用于高光谱图像的 Transformer . . . . .	11
2.3 本文构建的高光谱光源估计数据集 . . . . .	12
2.3.1 真实数据集 . . . . .	13
2.3.2 仿真数据集 . . . . .	14
2.4 本文提出的方法 . . . . .	15
2.4.1 光谱联合空间自注意 . . . . .	16
2.4.2 平滑约束的损失函数 . . . . .	19
2.5 实验过程与结果分析 . . . . .	19
2.5.1 实验设置 . . . . .	19
2.5.2 评价指标 . . . . .	20
2.5.3 仿真场景实验 . . . . .	22
2.5.4 真实场景实验 . . . . .	23
2.5.5 消融实验 . . . . .	23
2.6 本章小结 . . . . .	25

<b>第三章 光源光谱引导的图像增强</b>	27
3.1 引言	27
3.2 相关研究	29
3.2.1 高动态范围图像的亮度调整	29
3.2.2 基于学习的色调映射	29
3.3 本文构建的图像增强数据集	30
3.3.1 双相机系统及其对齐	31
3.3.2 光源的影响及阴影图	32
3.4 本文提出的方法	34
3.4.1 局部亮度适应	35
3.4.2 光谱感知自注意	36
3.5 实验过程与结果分析	37
3.5.1 评价指标	37
3.5.2 定性分析	38
3.5.3 定量分析	40
3.6 本章小结	42
 <b>第四章 光源光谱校正的材质分割</b>	45
4.1 引言	45
4.2 相关研究	46
4.2.1 基于阴影和反射率先验的本征分解	46
4.2.2 基于 Transformer 的语义分割	46
4.3 本文构建的材质分割数据集	47
4.4 本文提出的方法	49
4.4.1 反射率校正预处理	50
4.4.2 光谱联合 RGB 分解模型	52
4.5 实验过程与结果分析	53
4.5.1 评价指标	53
4.5.2 定性分析	54
4.5.3 定量分析	54

4.6 本章小结 . . . . .	57
<b>第五章 总结与展望 . . . . .</b>	<b>59</b>
5.1 总结 . . . . .	59
5.2 展望 . . . . .	60
<b>参考文献 . . . . .</b>	<b>61</b>
<b>致 谢 . . . . .</b>	<b>75</b>
<b>攻读硕士学位期间研究成果 . . . . .</b>	<b>77</b>
<b>学位论文出版授权书 . . . . .</b>	<b>79</b>



# 第一章 绪论

## 1.1 研究的背景及意义

光源是影响图像拍摄的因素之一，在不同的室内光源下，或昼夜变化的太阳光下，相机捕获到的图像的整体颜色有很大差异。这一问题在 RGB 图像领域作为颜色恒常问题<sup>[1-6]</sup>进行研究。在专业摄影时，往往使用标准色卡来校正光源、镜头、传感器等多方面的影响。即使在同一拍摄场景，不同时间的光源光谱也不相同，反映在 RGB 图像上则如图1-1所示，对于刻画了纳米级精细光谱特征的高光谱图像 (HyperSpectral Image, HSI)，光源光谱的影响会更加显著。



图 1-1 一天内太阳光色度的变化 (图片来源于 The Lighting Practice)

与仅包含红绿蓝三个谱段 (Red, Green, Blue) 的 RGB 图像相比，HSI 包含数百个谱段，跨越电磁波谱的紫外、可见光、近红外、中波甚至长波红外区域。一幅 HSI 相当于数百幅灰度图像，不仅记录了场景中物体丰富的空间结构，同时图像中每一个像元都对应一条与场景中物体材质相关的光谱曲线，使得 HSI 能够检测和分析不同物质间细微的光谱特征差异。因此，HSI 广泛应用于现实世界的各个方面，成为环境、农业、地质和矿产勘探、生物、医学和军事侦查等<sup>[7-11]</sup>领域研究的有效技术手段，发挥着越来越重要的作用。

光谱成像过程受到成像机理、大气条件、环境光照等因素的影响，导致 HSI 的光谱信息、环境光照及其他信息相互耦合<sup>[12]</sup>。因此，有必要研究影响光谱成像过程的各种因素，环境中的光源就是其中之一。

若拍摄场景的光源条件不可控，即光源光谱未经标定或测量，则需要从图像中估计未知的光源光谱，该任务被称为光源光谱估计。关于高光谱成像技术中的光源估计问题目前研究尚少，相关解决方案仍未被深入探索<sup>[13]</sup>。为进一步推进高光谱领域的深度学习算法研究，有必要简化光源光谱测量流程或尽可能准确地估计光源光谱，保证光源条件不可控场景下的高光谱图像的有效性，从而高效地构建大规模高光谱图像数据集。

排除影响光谱成像过程的各种因素后，高光谱图像中的每一个像元就对应于该位置的物质的光谱本征信息。

高光谱图像本征信息分解不仅仅是将输入图像分解为光照与颜色，同样也是将图像分解为与物体几何特性相关的空间光照图，以及与物体材质直接相关的光谱反射率图，分别对应于图像增强任务及材质分割任务。然而，现有的高光谱图像本征分解工作大多专注于提出一些高光谱本征分解框架，最后在公共数据集上验证该本征分解方法的性能优劣，没有将分解后得到的本征图像进一步推广至现实场景应用<sup>[12]</sup>。一些研究者已经证实，相比于高光谱原始图像，高光谱本征图像在矿物勘探、变化检测领域都为解译模型带来了性能上的提升，但在军事侦察等其他领域暂时还未出现高光谱本征图像的成熟应用先例<sup>[14-15]</sup>。为进一步促成科研成果的高效转化落地，有必要围绕下游应用开展高光谱图像本征信息分解研究，贴合实际场景对本征分解方法进行适应性改进，从而较大幅度地发挥高光谱图像本征信息分解方法的潜能。

综上所述，本文重点研究从高光谱图像中估计光源光谱与分解光谱图像的关键技术，以提高光源估计的准确度，并将分解的光谱图像用于下游任务。本文的研究成果将对大规模高光谱图像数据集的构建和高光谱图像本征信息分解的应用提供方法借鉴。估计并校正光源光谱，能够尽可能地保证光源条件不可控场景下的高光谱图像有效性，提高高光谱图像数据集构建的效率。作为高光谱图像本征信息分解的输出，空间光照图能够用于引导图像进行亮度调整；同时，光谱反射率图能够用于辅助图像进行材质分割。

## 1.2 国内外研究现状及存在的问题

### 1.2.1 国内外研究现状

#### 1. 光源光谱估计研究现状

高光谱图像可以分解为光源光谱和反射率光谱，光源光谱估计则专注于从高光谱图像中提取光源光谱。

光源光谱估计方法可以大致分为两类：基于统计的方法和基于学习的方法。在深度学习普及之前，人们通过统计图像中的像素点来估计光源光谱；随着深度学习算法的开发和算力的提高，人们更倾向于训练模型来预测光源光谱。

首先是基于统计的方法，这类方法是基于捕获场景的统计特性并利用这些先验提供的假设来估计光谱。最开始的光源光谱估计方法受 RGB 图像的光源估计启发，一些经典方法假设光源颜色可以直接通过以下方式获得：图像中所有像素颜色的均值 (Gray-World)<sup>[16]</sup>，或具有最大光谱响应的像素 (Max-RGB)<sup>[17]</sup>，或图像中主要目标的边缘像素 (Gray-Edge)<sup>[2]</sup>。这些假设适用于颜色恒常性问题<sup>[1-6]</sup>，也适用于光谱的光源估计任务并表现良好<sup>[18-19]</sup>。通过这些假设，可以估计出光源每个波长通道的光谱功率分布 (Spectral Power Distribution, SPD)。此外，Khan 等人<sup>[20]</sup>和 Su 等人<sup>[21]</sup>分别提出，从多光谱或高光谱图像中捕获的镜面反射可以提供有关场景光源的信息，但这些方法依赖于非朗伯表面来提取镜面反射分量，不适用于仅含有朗伯表面的场景。Zheng 等人<sup>[22]</sup>提出了另一类方法，通过利用光谱反射率的低秩特性来分离光源光谱和反射率。然而，在更一般的场景中，仅仅依靠单一假设来实现准确的光源光谱估计往往是不够的，许多基于统计的方法都是对所有场景使用同一组固定的超参数，使得估计的准确性对超参数设定和成像条件高度敏感。

然后是基于学习的方法，一开始，深度学习技术被用于对 RGB 图像中的光源颜色进行估计<sup>[4,23-25]</sup>，这些方法利用大规模图像数据集，相比基于统计的方法具有更高的精度。后续的研究将基于学习的模型扩展到高光谱领域，Robles 等人<sup>[26]</sup>利用卷积神经网络 (Convolutional Neural Networks, CNN) 来估计光源光谱，为了获取图像整体和局部光源信息，采用多尺度输入作为前端网络。Li 等人<sup>[27]</sup>将光源估计作为一个矩阵分解问题求解，并采用深度展开网络来优化矩阵分解，将高光谱图像分解为光源和反射率。

## 2. 光源光谱引导的图像增强研究现状

图像增强的目的是改善被掩盖的图像细节，突出图像中有用的信息和感兴趣的特征<sup>[28]</sup>，图像增强的广泛研究为许多先进的研究领域做出了贡献<sup>[29]</sup>，例如大气科学、天文学、海洋学、遥感成像、医学成像等。

图像增强任务包含的研究范围很广，对于不同的图像增强目的则对应着不同的方法。Dileep 和 Murthy<sup>[30]</sup>为了消除图像中存在的照明效应，使用 Retinex 算法和同态滤波 (Homomorphic Filter) 算法，实现了低对比度图像的色彩渲染和动态范围压缩。

更多与光源相关的图像增强工作聚焦于图像色彩校正方法。Ancuti 等人<sup>[31]</sup>提出了一种无监督的全局和局部色彩增强方法，考虑图像中的颜色空间分布，提取具有不同光照条件的图像中的视觉信息，结果表明该方法在颜色恒常校正和色调平衡方面具有良好的性能。Shi 等人<sup>[32]</sup>使用归一化伽玛变换和对比度限制的自适应直方图均衡 (Adaptive Histogram Equalization, AHE) 来校正图像颜色失真，在此方法中，输入的 RGB 图像首先转换为 Lab 色彩空间，将归一化伽马校正算法应用于增强图像 L 分量的对比度，使用基于灰色世界的颜色校正函数增强 a 和 b 分量，由于使用了 Lab 色彩空间，增强后的图像没有颜色偏差。Afifi 等人<sup>[33]</sup>提出的方法将色偏向量 (Color Cast Vector) 与颜色校正函数连接起来，通过现成的照度估计函数来估计色偏向量，对每个色偏向量进行聚类，采用最合适的聚类对应的颜色校正函数来生成颜色校正图像。Tai 等人<sup>[34]</sup>提出的方法应用了两种常用的颜色增强假设：灰色世界和颜色直方图拉伸函数的宽色域特性，可以很好地保持色彩对比度，避免过度拉伸问题，能够保留原始图像的自然度，尤其适用于色彩失真明显的图像。

## 3. 光源光谱校正的材质分割研究现状

材质分割是对图像中的材质进行像素级的识别<sup>[35]</sup>，与语义分割相似，但是可能需要将同一语义进一步分为不同的材质，例如将树分为树干和树叶。

材质分割数据集为材质分割的研究提供了数据基础。Dana 等人<sup>[36]</sup>构建了一个在 205 个光源种类及其方向组合中拍摄的 61 个不同纹理样本的数据集，每张图像仅包含一种材质，用于研究光源种类和光源方向对材质的纹理特征的影响。Bell 等人<sup>[37]</sup>构建了一个包含 23 种材质的 200 万个 64×64 分辨率图像块的数据集，基于 CNN 和条件随机场 (Conditional Random Field, CRF)，将大量的

样本用于材质分割模型的训练。Schwartz 和 Nishino<sup>[38]</sup>在 PASCAL VOC<sup>[39]</sup>, MS COCO<sup>[40]</sup>和 ImageNet<sup>[41]</sup>三个大规模数据集的基础上进行材质分割的标注，引入了一种深度网络架构，整合了局部材质特征和全局语义信息。Liang 等人<sup>[35]</sup>构建了包含 RGB 图像、近红外图像和偏振图像的多模态材质分割数据集，并用语义分割的结果作为图像分区域学习的依据。

由于语义分割与材质分割的相似性，前者为后者的网络设计提供了思路。语义分割在引入了全卷积网络后，用卷积层代替全连接层，在分割精度上有了显著进步<sup>[42]</sup>。Chen 等人<sup>[43]</sup>提出了动态区域感知卷积 (Dynamic Region-Aware Convolution, DRConv)，自动地将多个卷积核分配给特征表示相似的空间区域，在提升性能的同时，还维持了一定的空间不变性。Deeplabv2<sup>[44]</sup>采用空洞空间卷积池化金字塔 (Atrous Spatial Pyramid Pooling, ASPP) 来提取多尺度特征。Deeplabv3+<sup>[45]</sup>则采用编码器加解码器的结构，进一步提升了图像分割的精度。

### 1.2.2 研究存在的问题

根据现有的研究成果，可以发现以下几个方面的不足。

1、由于从单个图像中恢复像素级光源光谱是一个约束不足的问题，现有的方法通常假设整个场景具有均匀的光源分布。基于统计的光源估计方法通常假设场景具有特定的统计分布特征，例如光照的分布服从某种概率模型。然而，真实世界中的场景往往具有非常复杂的光照条件，这些假设可能无法完全捕捉到场景的真实特征，从而导致基于统计的方法的适用范围受到限制。

2、缺少将光源估计与各种视觉任务相结合的研究，已有的光源估计研究主要解决颜色稳定性问题，如色彩校正，白平衡等。在本征分解、重照明等任务中则将光源估计与其他要素相叠加，并没有单独考量光源估计的准确性。

## 1.3 论文研究内容及章节安排

本文围绕高光谱图像的光源光谱问题，基于深度学习理论，并结合自注意机制、损失函数设计、局域与全局分析、反射率分解等方法，从光源光谱的估计出发，以图像增强和材质分割为例，探索光源光谱在低层次问题和高层次问题的应用。

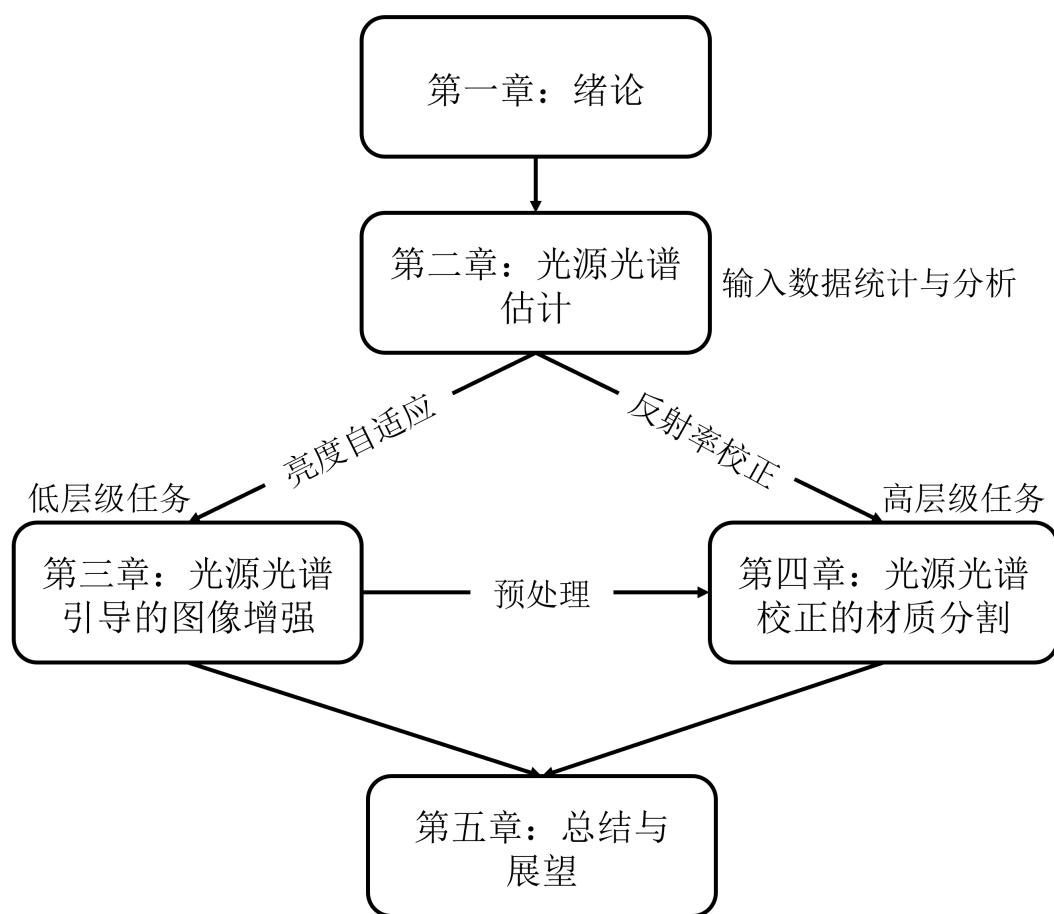


图 1-2 论文组织结构图

本文的组织结构如图1-2所示。

论文章节安排具体内容如下：

第一章，绪论。主要介绍了高光谱图像中光源光谱的重要性、研究现状以及目前光源光谱研究的发展趋势和存在的不足。最后给出了本文研究内容的框架和章节的分布及每章的贡献介绍。

第二章，光源光谱估计。利用深度学习的方法对高光谱图像进行光源光谱估计，构建仿真数据集并拍摄真实数据集，提出了光谱联合空间自注意机制以及基于光源特性的损失函数。

第三章，光源光谱引导的图像增强。利用第二章估计的光源光谱结果来引导 RGB 图像的高动态范围增强，拍摄并构建了高光谱与 RGB 对齐数据集，提出了局部亮度适应法和光谱感知自注意机制。

第四章，光源光谱校正的材质分割。利用第二章估计的光源光谱结果来校

正反射率，并利用第三章的图像增强方法对 RGB 图像进行预处理，在第三章的数据集的基础上进行材质分割标注，以构建高光谱-RGB 材质分割数据集，提出了光谱联合 RGB 分解模型。

第五章，总结与展望。对本文的研究内容及研究结论加以归纳总结；阐述本论文研究内容的欠缺之处，并对下一步的研究内容及方向做出展望。



## 第二章 光源光谱估计

### 2.1 引言

关于高光谱成像技术中的光源估计目前研究尚少，相关解决方案仍未被深入探索<sup>[13]</sup>。若光源光谱无法估计，则高光谱图像的光谱反射率重构将面临巨大挑战。

光照的变化对高光谱技术的实际应用也是一项艰巨的挑战<sup>[46]</sup>，为此，高光谱图像通常需要使用已知的色卡或白板作为参考<sup>[27]</sup>或使用专用测量设备<sup>[47]</sup>进行校准，以确保光谱反射率的测量准确且一致。

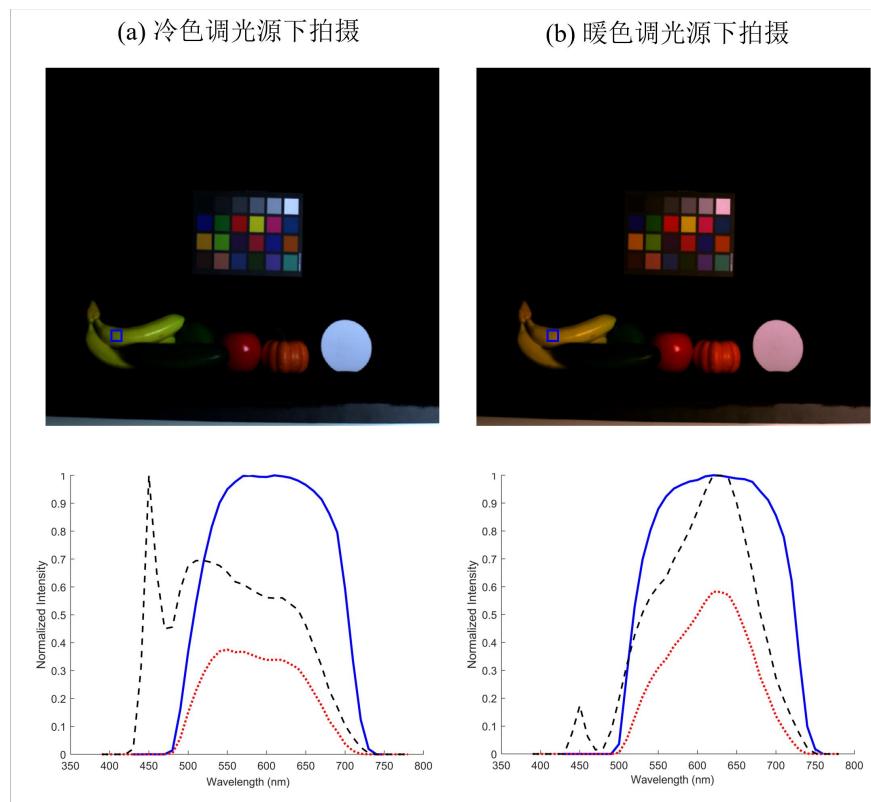


图 2-1 一些塑料水果模型在冷色调和暖色调光下的图像，相关色温分别为 (a)6500K 和 (b)2500K，以及从蓝框区域反射的相应光谱，黑色虚线代表从白板获得的(归一化)光源光谱，红色虚线代表方框区域的测量光谱，蓝色线代表其校正的(归一化)反射率

如图 2-1 所示，同一个物体在不同的光源下呈现出不同的光谱曲线，红色曲

线代表的测量值明显与黑色虚线代表的光源光谱曲线具有形态上的相似性，体现出类似的能量分布，渲染为 RGB 图像则表现为场景中的物体与光源颜色相和谐。然而同样是香蕉的光谱测量值表现出不同：在冷色调光源下的测量值呈现出 500nm 和 640nm 两处较小的峰，而在暖色调光源下的测量值呈现出 630nm 一处较强的峰。将光谱测量值逐波段除以光源光谱再进行归一化处理，就得到了对应的反射率，这是物质的本征属性，是可以用于区分物质的“基因”。如图2-1中蓝线所示，不同光源下香蕉的光谱曲线在校正后变得一致。

上述的校准方法假设场景具有全局一致的光源光谱，在实际场景中，光源光谱的强度和光谱曲线可能随着空间位置而改变。

对于非全局一致的光源光谱估计，定义高光谱图像矩阵为  $\mathbf{I} \in \mathbb{R}^{H \times W \times C}$ ，其反射率矩阵为  $\mathbf{R} \in \mathbb{R}^{H \times W \times C}$ ，其中  $H$  是高度， $W$  是宽度， $C$  是通道数。光源光谱矩阵为  $\mathbf{L} \in \mathbb{R}^{H \times W \times C}$ ，满足  $\mathbf{I} = \mathbf{R} \odot \mathbf{L}$ ，其中  $\odot$  表示哈达玛积。对于全局一致的光源光谱估计，光源光谱退化为向量  $L \in \mathbb{R}^{H \times W \times C}$ ，满足  $\mathbf{I} = \mathbf{R} \cdot L$ ，其中  $\cdot$  表示点积。由此可知，无论是否假设场景具有全局一致的光源光谱，从高光谱图像中估计出光源光谱都是一个病态问题。

具体来说，光源光谱估计方法主要分为两类：基于统计的方法<sup>[18-22]</sup>和基于学习的方法<sup>[26-27]</sup>。在基于统计的方法中，镜面反射分解<sup>[22]</sup>针对场景中的非朗伯表面，从镜面反射中提取高光谱图像的光源光谱。在基于学习的方法中，深度展开网络<sup>[27]</sup>将光源估计作为一个约束矩阵分解问题求解，并采用深度展开网络来优化矩阵分解，将高光谱图像分解为光源光谱和反射率。

近年来，基于 Transformer 的网络也被用于光谱相关的研究并取得了良好的效果<sup>[48-50]</sup>，得益于其全局感知能力，可以应用于全局光源光谱估计。

在本章中，本文以 Transformer 网络为基础，提出了光谱联合空间自注意机制，通过分阶段处理的方式避免了高维数据的“维数灾难”，用于突出重要的光谱通道和图像中有效区域的作用；提出了平滑约束的损失函数，避免了不实际的病态结果，从而提升了光源估计的准确度。

## 2.2 相关研究

### 2.2.1 基于学习的光源估计

随着机器学习的快速发展，多种机器学习算法，尤其是深度学习算法也被应用于光源估计。Cardei 等人<sup>[51]</sup>用一个多层次的神经网络来学习图像的色度直方图和光源色度间的关系。Wang 等人<sup>[52]</sup>提出了一种基于支持向量回归的光源估计算法，并表明图像的导数比图像的原始像素值更有效。Bianco 等人<sup>[53]</sup>将图像分为小的图像块，使用卷积神经网络进行局部的光源估计。

除了直接将机器学习方法用于光源估计问题，将光源估计问题转化为其他问题也是一种思路。Barron<sup>[54]</sup>将光源估计问题重新表述为一个在色度空间中的二维空间定位任务，从而将目标检测和结构化预测等技术应用于光源估计。Lou 等人<sup>[55]</sup>将光源估计问题看做回归问题，提出了使用深度神经网络进行光源估计的新框架。Shi 等人<sup>[56]</sup>提出的方法将光源估计作为两阶段的任务，首先用一个深度网络提出光源估计的多种可能，然后用另一个深度网络从中选择最合适的光源估计值。Oh 和 Kim<sup>[57]</sup>将光源聚类为几种类别，从而将光源估计问题简化为分类问题并用深度学习进行类别预测。

### 2.2.2 用于高光谱图像的 Transformer

Transformer 最初被用于自然语言处理<sup>[58]</sup>，现在已被广泛应用于计算机视觉，包括高光谱图像的研究。Hong 等人<sup>[59]</sup>引入了一种基于 Transformer 的骨干(backbone) 网络，通过生成群级光谱编码(spectral embeddings)，从附近的高光谱波段捕获光谱局部信息。Jia 等人<sup>[60]</sup>提出了一种多尺度 Transformer 来编码空间-光谱信息。Yang 等人<sup>[61]</sup>使用三维卷积投影模块来编码局部空间-光谱细节并用一个卷积置换器(conv-permutator)沿着高、宽、光谱三个维度获取信息。Li 等人<sup>[62]</sup>提出了一个 CNN-Transformer 框架，其中全局特征使用 Transformer 生成，局部特征使用 CNN 生成，将这些特征相结合，并利用具有空间损失和光谱损失的损失函数进行训练。

另外，还有一些专门针对光谱和空间自注意的设计。Liu 等人<sup>[63]</sup>对高光谱图像分两路分别提取光谱自注意和空间自注意，然后将两路的特征通过串接的方式融合。Zhong 等人<sup>[64]</sup>提出了两个超参数搜索空间，一个用于搜索光谱自注意

模块与空间自注意模块的模块内的参数配置，另一个用于搜索这两种模块的分布顺序。

### 2.3 本文构建的高光谱光源估计数据集

基于学习的光源光谱估计算法能表现出更好的效果，但是需要大量数据来驱动模型的学习。为了支撑基于学习的光源光谱估计算法，本文构建了高光谱图像的光源光谱估计数据集，其中包括真实数据集和仿真数据集。

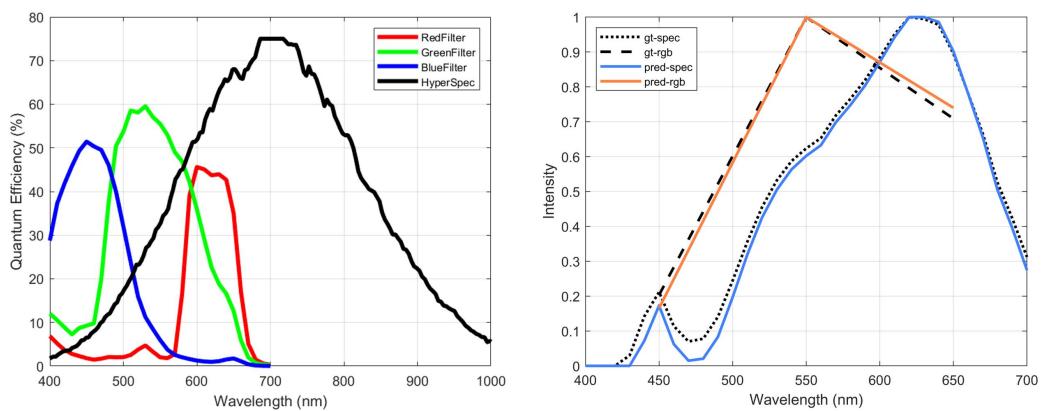


图 2-2 左图：RGB 和高光谱相机的量子效率 右图：RGB 和高光谱域中光源估计的结果

高光谱图像与 RGB 图像的差别不仅仅是通道数量的差别，高光谱相机的量子效率 (Quantum Efficiency) 曲线也与 RGB 相机的不同，如图2-2左图所示。RGB 相机仅集成了三个宽谱波段，而高光谱相机则在更连续的光谱维度中采样信息，且高光谱相机的波段范围更大，涵盖 700nm 以上的近红外波段。

高光谱图像比 RGB 具有更丰富的光谱信息<sup>[65]</sup>，反映在单点的反射率或场景的光源光谱上时，RGB 图像的光源光谱是一条只有三个采样点的折线，而光谱图像的光源光谱可以绘制为曲线。如图 2-2 右图所示，黑色点线表示光源光谱真值曲线，黑色虚线表示光源 RGB 真值折线，蓝色曲线表示预测的光源光谱，橙色折线表示预测的光源光谱积分到 RGB 三通道后的颜色。这表明了，如果将预测的光源光谱积分到 RGB 的三个通道中，则预测的 RGB 光源更容易接近真实光源。此外，更多的光谱通道数会导致每个独立的光谱通道图像的信噪比降低，常见的做法是对多个高光谱通道的图像求平均，得到信噪比更高的多光谱图像。因此，高光谱光源估计问题比 RGB 域中的光源估计问题更加不确定和困难。

### 2.3.1 真实数据集

在现有的高光谱图像数据集中，为了凸显数据的多样性，光源的类型和强度对在同一场景获取的高光谱图像的影响尚未得到充分的研究。因此，本文构建了一个在受控场景中切换光源类型和强度的数据集。

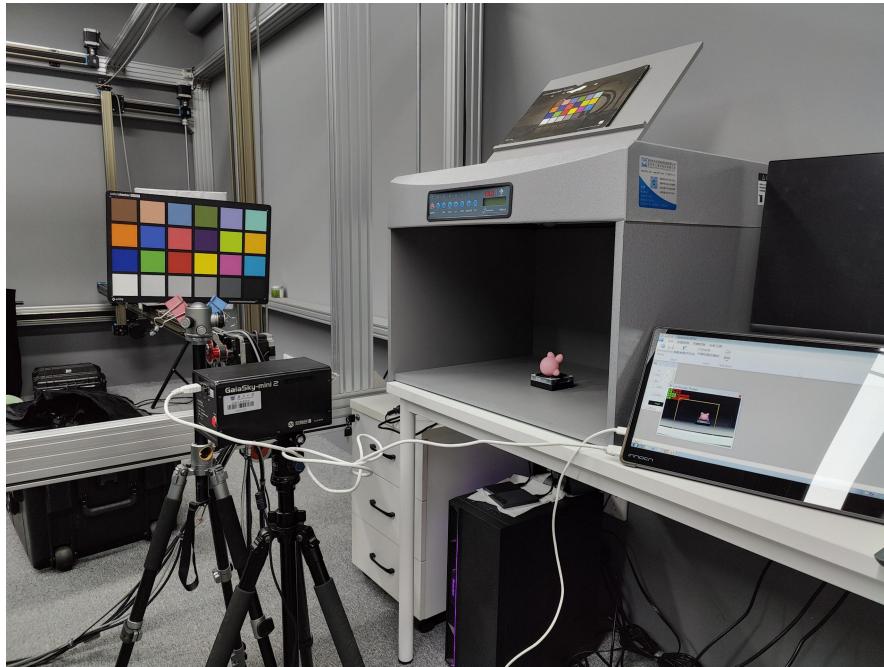


图 2-3 高光谱图像采集系统

本文的数据集是使用推扫式高光谱相机 GaiaSky-Mini2 捕获的，拍摄时的设置及参数值如表2-1所示。高光谱图像原始尺寸宽×高×通道数为 1057×960×176，光谱波段范围从 390nm 到 1000nm。由于本文使用的室内光源不包含近红外波段，将获取到的图像中大于 770nm 波段的通道图像舍弃，并对多个高光谱通道的图像求平均，减少波段数量，提高信噪比。处理后图像尺寸变为宽×高×通道数为 1057×960×38，范围从 400nm 到 770nm，间隔 10nm。

表 2-1 光源光谱估计数据集拍摄设置及参数值

拍摄设置	参数值
高光谱相机镜头焦距	16mm
高光谱相机数值孔径	F/2.8
相机距拍摄画面中心距离	0.5m
光源距拍摄画面中心距离	0.3m
光源照射方向	单侧水平照射

该数据集包含 18 个场景，全部在暗室环境中捕获，以确保光源条件可控。场景中的物体包括真假绿植、真假食物、玩偶、木质品、金属制品、石制品、皮

肤、毛发等。本文使用了 5 种不同的光源，挑选自常见的室内照明，色温范围为 2500K 至 6500K，如图2-4所示。每个光源都以 4 个强度级别（从 100lux 到 1000lux）捕获，总共生成 360 张高光谱图像。该数据集已经公布，可用于非商业目的的学术研究。

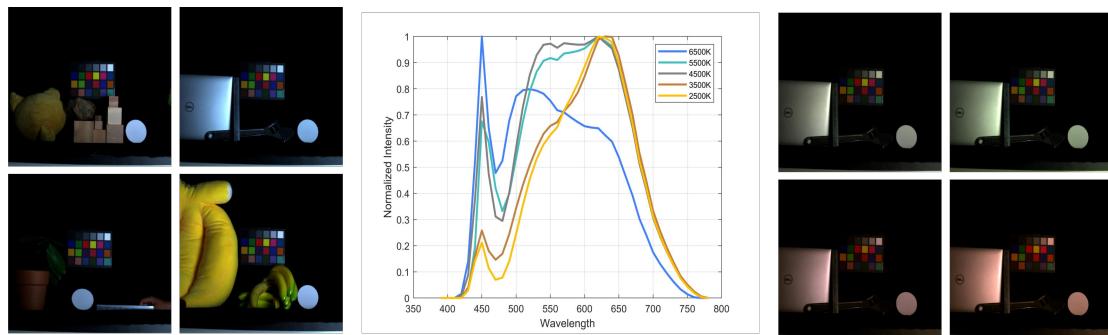


图 2-4 本文构建的数据集 左：光源色温为 6500K 的部分场景的 RGB 渲染图 中：光源光谱 右：同一场景在光源色温范围为 5500K 至 2500K 时的 RGB 渲染图

如表2-2所示，本文提出的真实数据集与常用的高光谱图像数据集相比，有更大光谱波段范围，校正方式包括标准 24 色卡和白板，仅在室内光源下拍摄。

表 2-2 高光谱图像或光源数据集的比较

发布时间	名称	数量	分辨率	波段	校正方式	光源
2002	SFU <sup>[66]</sup>	<b>1995</b>	<b>1×1×101</b>	380-780nm	直接测量	室内 & 室外
2008	CAVE <sup>[67]</sup>	32	512×512×31	400-700nm	色卡	室内
2011	Harvard <sup>[68]</sup>	75	1040×1392×31	420-720nm	无	室内 & 室外
2014	NUS <sup>[69]</sup>	64	984×1312×31	<b>400-1000nm</b>	色卡	室内 & 室外
2015	Foster <sup>[70]</sup>	30	1018×1339×33	400-720nm	漫反射球	室外
2017	KAIST <sup>[71]</sup>	30	<b>3376×2704×31</b>	420-720nm	色卡	室内
2021	KAUST <sup>[27]</sup>	<b>409</b>	512×512×31	400-730nm	白板	室内 & 室外
2022	NTIRE <sup>[72]</sup>	<b>1000</b>	482×512×31	400-700nm	无	室内 & 室外
2024	本数据集	<b>360</b>	960×1057×38	400-770nm	色卡 + 白板	室内

### 2.3.2 仿真数据集

为了进一步增强训练场景的多样性，特别是对室外场景的补充，本文使用 KAUST<sup>[27]</sup> 提出的高光谱反射率数据集和 SFU<sup>[66]</sup> 提出的光源光谱数据集合成仿真数据集。KAUST 的数据集由 400 张高光谱图像组成，包括室内场景和室外场景，尺寸为  $512 \times 512 \times 31$ ，光谱波段范围从 400nm 到 700nm，光谱分辨率为 10nm。SFU 的数据集包含 102 条光源光谱曲线，包括室内光源和太阳光，每条曲线有 101 个波段，光谱波段范围从 380nm 到 780nm，光谱分辨率为 4nm。本数

据集截取其中的 400nm 至 700nm 范围，以 10nm 间隔对其进行下采样，以获得 31 个波段的光源曲线。按照  $\mathbf{I} = \mathbf{R} \cdot \mathbf{L}$ ，其中  $\cdot$  表示点积，将处理后的 SFU 数据集中的光源光谱  $\mathbf{L}$  与 KAUST 数据集提供的反射率  $\mathbf{R}$  相乘，以合成仿真的高光谱图像  $\mathbf{I}$ ，构成仿真数据集。仿真的高光谱图像  $\mathbf{I} \in \mathbb{R}^{H \times W \times C}$ ，其反射率图像为  $\mathbf{R} \in \mathbb{R}^{H \times W \times C}$ ，其中  $H$  是高度， $W$  是宽度， $C$  是通道数，光源光谱为  $\mathbf{L} \in \mathbb{R}^C$ ，意味着这是一个全局一致的光源光谱估计数据集。筛选掉一些重复场景后，总共使用了 384 张高光谱图像进行训练和验证。

## 2.4 本文提出的方法

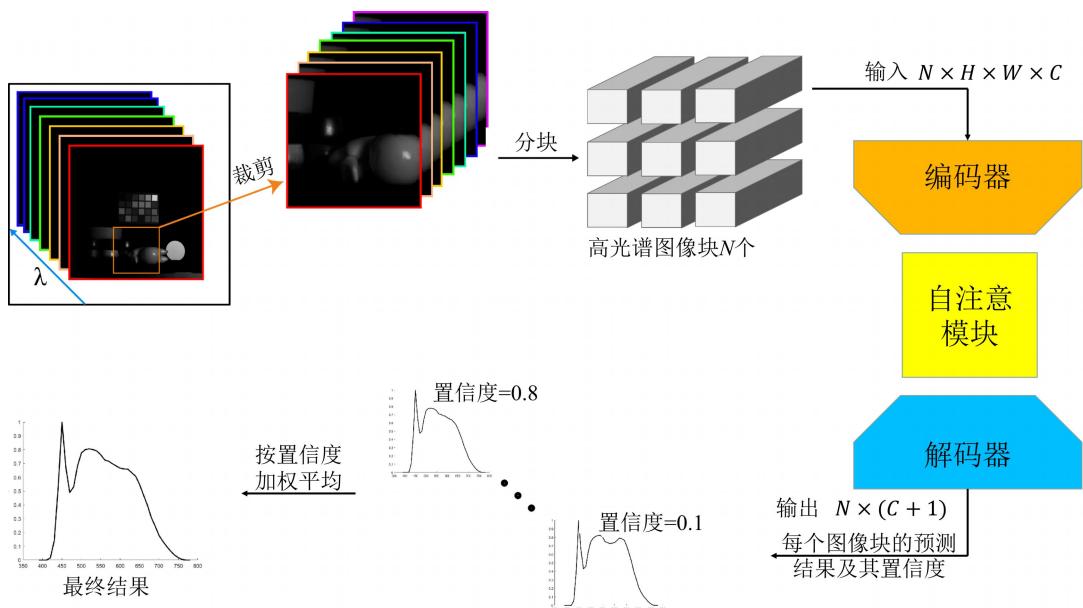


图 2-5 算法处理流程

对于全局光源估计的任务，本文设计了一个算法处理流程，如图2-5所示。采用 Transformer 的编码器和解码器，网络参数设置如表2-3所示，并在其中加入光谱联合空间自注意模块，在损失函数中添加平滑约束。

考虑到高分辨率的高光谱图像在模型训练时会占用大量显存空间，本文将输入的高光谱图像首先沿空间维度分割成大小相等的高光谱图像块，并且假设每个高光谱图像块包含完整的光源光谱。这些高光谱图像块将被输入到模型中，模型分别输出相应的预测光源光谱和置信权重。最后根据置信度权重进行加权平均，得到预测的全局光源光谱。

表 2-3 光源光谱估计网络参数设置

网络参数	参数值
等边图像尺寸	512×512
等边图像块尺寸	8×8
为每一个图像块编码的长度	31
编码器深度	6
多头注意力中头的数目	2
多层感知器中隐含层的维度	512
随机数种子编号	42

在训练期间，将每个图像块与真实测量值进行比较以计算损失。在测试过程中，将预测的全局光源光谱与真实测量值进行比较，进而获得角误差和均方差。在训练和测试时，使用 0-1 掩膜将图像中的色卡和白板遮盖，避免模型直接从色卡或白板中学习光源光谱。

#### 2.4.1 光谱联合空间自注意

考虑到 Transformer 在捕获非局部远程依赖性方面的有效性及其在其他视觉任务中优越的性能，本文的目标是探索 Transformer 在高光谱光源估计中的潜力。然而，直接将 Transformer 应用到高光谱光源估计时存在两个主要问题。

第一个问题是原始的 Transformer 在空间维度上的建模依赖于远程关系<sup>[48]</sup>。另一个问题是高维数据的参数量太大，无法直接输入到自注意模块中。

本文设计了包含四个步骤的自注意模块，如图2-6所示，特征图和自注意使用不同的颜色加以区分。该模块包含两方面的自注意操作：通道自注意和空间自注意。通道自注意用于增强或抑制不同通道的信息，而空间自注意用于强调或削弱特定位置的信息。

对于通道自注意，首先通过全局平均池化，计算输入特征的全局平均值，然后使用两个全连接层计算通道权重。第一个全连接层使用 ReLU 激活函数，第二个全连接层使用 Sigmoid 激活函数，输出通道自注意。这种方法允许网络专注于特定的通道。

如图2-6(a) 所示，将从特征图中提取通道自注意的过程可以表示为：

$$f_c[u] = \frac{1}{M \times N} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} F[m, n, u], \quad (2-1)$$

$$a_c[u] = \text{Sigmoid}(\omega_2[u] \text{ReLU}(\omega_1[u] f_c[u] + b_1[u]) + b_2[u]).$$

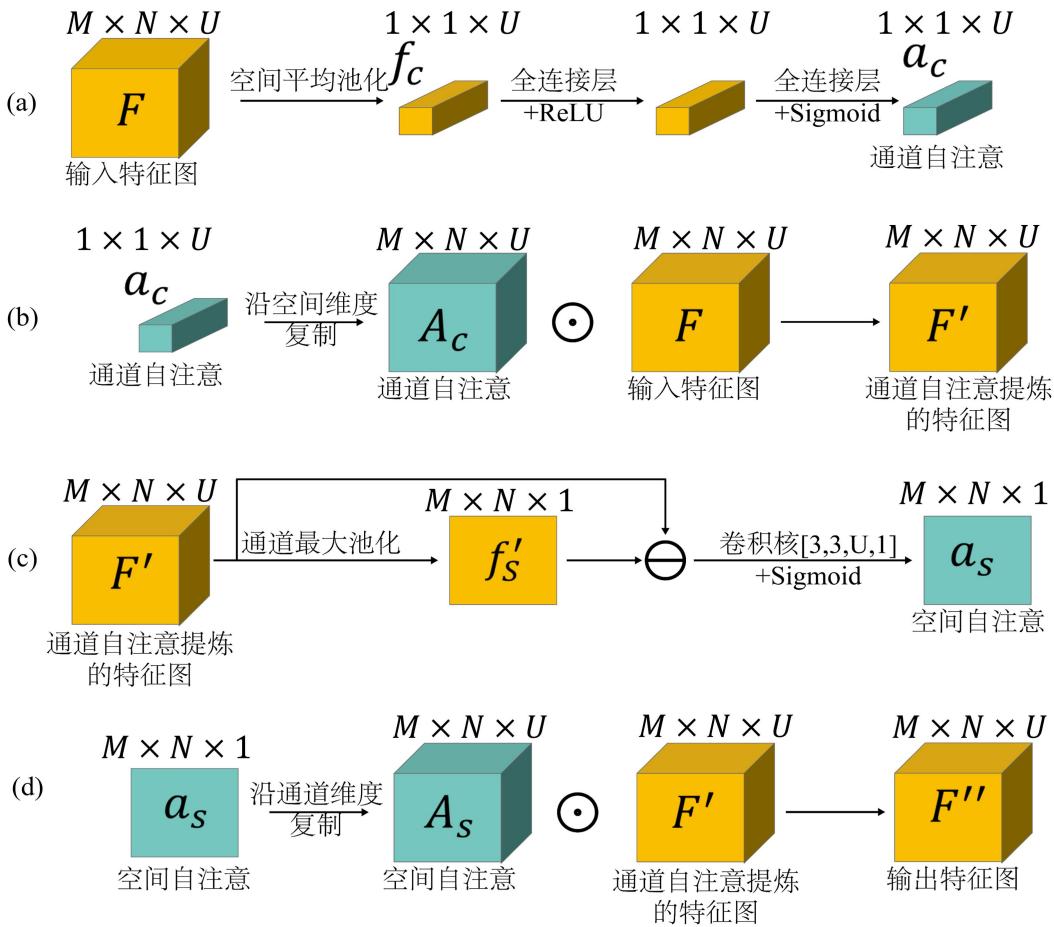


图 2-6 光谱联合空间自注意模块 (a) 提取通道自注意 (b) 将通道自注意作用于特征图  
(c) 提取空间自注意 (d) 将空间自注意作用于特征图

其中  $F$  为特征图，它的维度为  $M \times N \times U$ ， $M$  是高度， $N$  是宽度， $U$  是通道数， $F[m, n, u]$  表示矩阵  $F$  中  $[m, n, u]$  位置的元素， $f_c$  为在沿空间维度平均池化后的特征，它的维度等于通道数  $U$ 。 $w_1, b_1, w_2, b_2$  分别为第一个和第二个全连接层的权重和偏置， $a_c$  表示维度为通道数  $U$  的通道自注意。

如图2-6(b)所示，将通道自注意作用于特征图的过程可以表示为：

$$\begin{aligned} A_c &= a_c \otimes M \otimes N, \\ F' &= A_c \odot F. \end{aligned} \tag{2-2}$$

其中  $\otimes$  为 Kronecker 积， $A_c$  为  $a_c$  沿着空间维度复制得到的维度为  $M \times N \times U$  的通道自注意。 $\odot$  表示矩阵的像素乘积， $F'$  为经过通道自注意提炼后的特征图。

对于空间自注意，首先计算通道自注意的最大值，然后使用 3x3 的卷积核

进行卷积操作。输入特征减去通道最大值得到的差值被输入到卷积运算中，然后通过 Sigmoid 函数进行缩放，从而产生空间自注意。这种方法允许网络专注于图像中的特定区域。

如图2-6(c) 所示，从特征图中提取空间自注意的过程可以表示为：

$$\begin{aligned} f'_s[m, n] &= \max_u(F'[m, n, u]), \\ \Delta F' &= f_s \otimes U - F', \\ a_s[m, n] &= \text{Sigmoid} \left( \sum_{i=0}^2 \sum_{j=0}^2 \sum_{u=0}^{U-1} w_3[i, j, u] \Delta F'[m+i, n+j, u] + b_3[i, j, u] \right). \end{aligned} \quad (2-3)$$

其中  $f'_s$  为  $F'$  沿着通道维度最大值池化的结果，将  $f'_s$  沿着通道维度复制，维度对齐于  $F'$  后与之作差的结果为  $\Delta F'$ 。维度为  $M \times N$  的空间自注意  $a_s$  为  $\Delta F'$  经过  $3 \times 3 \times U \times 1$  的卷积核以步长为 1 卷积并使用 Sigmoid 函数激活后得到，该卷积核在通道维度输入为  $U$ ，输出为 1，其权重为  $w_3$ ，偏置为  $b_3$ 。

如图2-6(d) 所示，将空间自注意作用于特征图的过程可以表示为：

$$\begin{aligned} A_s &= a_s \otimes U, \\ F'' &= A_s \odot F'. \end{aligned} \quad (2-4)$$

其中  $A_s$  为  $a_s$  沿着通道维度复制得到的维度为  $M \times N \times U$  的空间自注意。 $F''$  为光谱联合空间自注意模块最终输出的特征图。

该自注意模块可以嵌入深度学习网络中，以增强网络对特定通道和位置信息的感知。

## 2.4.2 平滑约束的损失函数

在之前的研究中，基于角误差的损失函数通常优于基于均方差的损失函数。此外，本文引入了平滑损失以适应自然光和白色节能灯的光谱特征。平滑损失通过计算光源光谱的二阶导数来衡量，总体损失函数可以表示为：

$$\ell = e_{angular} + \eta \ell_{smooth}. \quad (2-5)$$

其中  $e_{angular}$  为预测的光源光谱与光源光谱真实值的角误差， $\ell_{smooth}$  为平滑损失项， $\eta$  为平滑系数。

平滑损失项通过以下方式计算：

$$\begin{aligned}\Delta^2[L](\lambda) &= L(\lambda+2) - 2L(\lambda+1) + L(\lambda), \\ \ell_{smooth} &= \frac{1}{C} \|\Delta^2[L](\lambda)\|^2.\end{aligned}\tag{2-6}$$

其中  $C$  表示光谱通道数， $\lambda$  是离散的光谱通道。 $\|\cdot\|$  表示 L-2 范数， $\Delta^2$  表示光源  $L$  的二阶导数。

## 2.5 实验过程与结果分析

### 2.5.1 实验设置

本文的深度学习算法训练和测试，均在一台搭载了 Intel Xeon Gold 6510 处理器 (2.70 GHz)、125GB 内存、2 块 NVIDIA RTX3090 显卡 (每块显卡显存 24GB) 的远程服务器中完成。对于软件系统，在 AMD64 架构的 Linux 系统 Ubuntu 20.04.6 LTS 平台下运行 Docker 25.0.0 完成容器搭建。在容器中，在 Ubuntu 18.04.5 LTS 平台下使用 conda 4.10.1 管理深度学习环境，主要包括 python 3.9.6 和 pytorch 1.8.1，使用 cuda 11.1 将模型部署于显卡。

实验结果的数据分析在一台搭载了 12 代 Intel Core i7-12800HX 处理器 (2.00 GHz)、64GB 内存、NVIDIA RTX A1000 显卡 (显存 4GB) 的笔记本电脑中完成。对于软件系统，在 64 位 Windows 11 Pro 23H2 系统下运行 MATLAB R2022a。

训练和测试阶段的参数设置如表2-4所示。由于本文构建的真实数据集的每个场景包括 20 张在不同光源种类及强度的光源下拍摄，数据间的相似度较高，在此数据集内训练并测试不足以说明模型的泛化性。因此，本文在仿真场景实验与真实场景实验中同样使用仿真数据集中的 288 张高光谱图像训练，在测试时分别在各自的数据集中测试。

### 2.5.2 评价指标

光源估计通常分为两种类型：全局光源估计和像素级光源估计。

表 2-4 光源光谱估计实验设置及参数值

实验参数	参数值
深度学习框架	PyTorch
优化器	Adam
初始学习率	0.001
权重衰减	0.0005
训练轮数	3000
批大小	16
用于训练的样本数量	288
用于测试的样本数量	仿真场景 96, 真实场景 360

本文在设计实验时，专注于讨论全局光源估计，这种方法可用作白板等光源参考的有效替代。

本文使用角误差  $e_{angular}$  和均方差  $e_{msquare}$  两种评价指标对光源光谱估计的结果进行评价，角误差的计算方式为：

$$e_{angular}(L, \hat{L}) = \arccos \left( \frac{\langle L, \hat{L} \rangle}{\|L\| \|\hat{L}\|} \right). \quad (2-7)$$

均方差的计算方式为：

$$e_{msquare}(L, \hat{L}) = \frac{1}{C} \sum_{\lambda=1}^C (L(\lambda) - \hat{L}(\lambda))^2. \quad (2-8)$$

其中  $L$  是预测的光源光谱， $\hat{L}$  是光源光谱真实值，本文中角误差统一用角度制表示。

光源光谱估计旨在最小化预测光源矢量与真实光源矢量之间的误差。需要注意的是，在高光谱图像处理中，图像通常被全局归一化，从而获得无单位的反射率图，仅保留像素之间的相对强度而非绝对强度，因此，不受光源绝对强度影响的角误差更适用于高光谱图像光源估计结果的评估。

### 2.5.3 仿真场景实验

在仿真场景实验中，本文使用仿真数据集共 384 张高光谱图像中的 75% 用于训练，剩余的 25% 用于测试。将本文提出的方法与其他方法进行对比，包括基于统计的方法 (Spectral Gray World、Max Spectral、Spectral Gray Edge<sup>[18]</sup> 和镜面反射分解<sup>[22]</sup>) 以及基于学习的方法 (PixelWise Illuminant Recovery (PWIR)<sup>[26]</sup> 和深度展开网络<sup>[27]</sup>)。PWIR 中使用了浅层的 AlexNet，为了更公平地对比，额外测

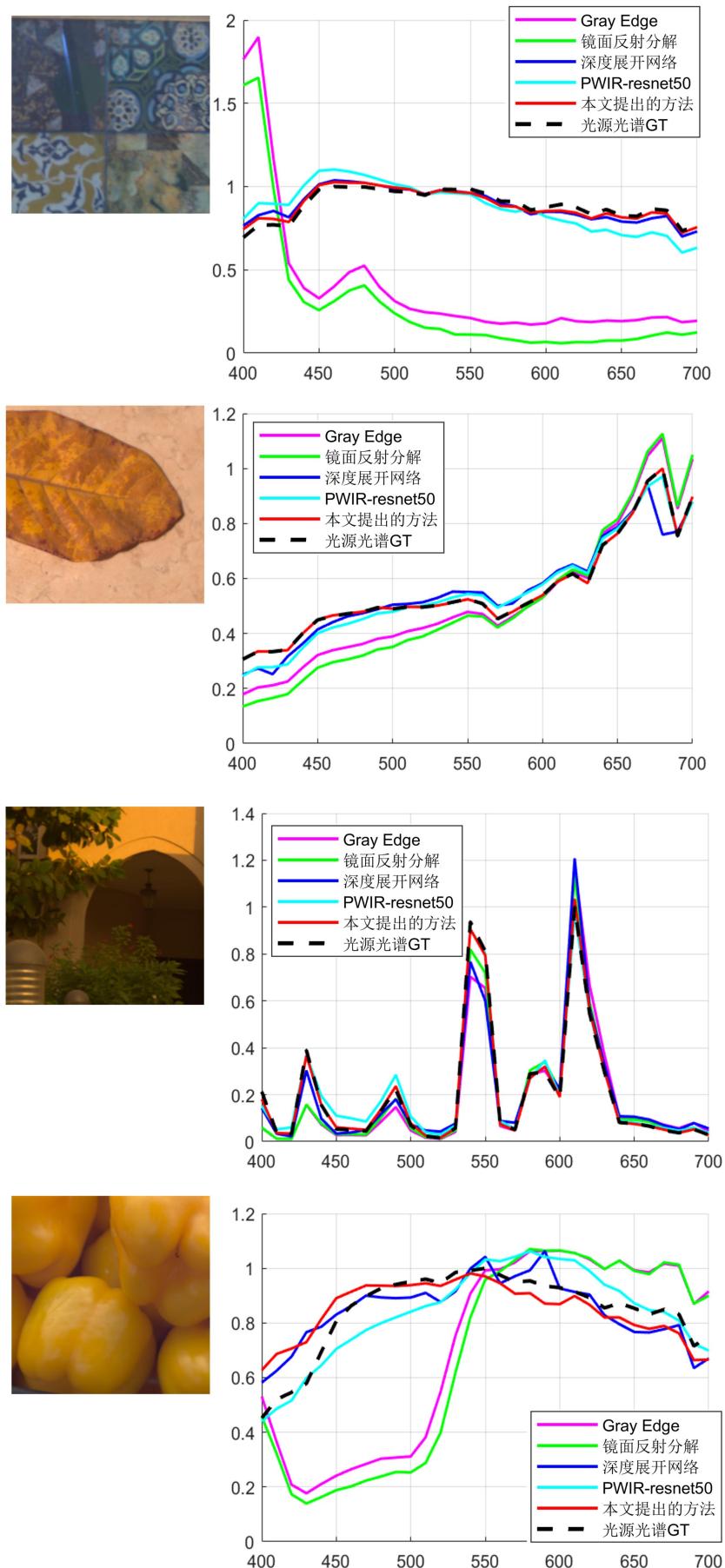


图 2-7 仿真场景实验结果可视化

试了用 resnet50 取代 AlexNet 的效果。其余对比算法的参数设置均固定采用相应论文中的默认设置。

表 2-5 不同算法在仿真数据集上的比较

方法	角误差/度				均方差			
	平均值	中位值	最小值	最大值	平均值	中位值	最小值	最大值
Gray World	25.19	23.45	0.845	73.59	0.154	0.142	0.015	0.312
Max Spectral	22.95	21.17	0.662	65.71	0.141	0.121	0.011	0.300
Gray Edge	22.94	21.08	0.716	65.75	0.142	0.125	0.012	0.298
镜面反射分解	25.16	23.57	0.805	69.01	0.148	0.129	0.015	0.308
深度展开网络	19.14	16.53	3.952	56.51	4.318	3.871	2.305	6.832
PWIR-alexnet	8.649	6.976	0.607	49.48	2.362	1.824	0.493	5.209
PWIR-resnet50	5.415	4.440	0.608	22.68	0.018	0.005	<b>0.000</b>	0.060
本文提出的方法	<b>3.841</b>	<b>2.709</b>	<b>0.088</b>	<b>24.05</b>	<b>0.014</b>	<b>0.001</b>	<b>0.000</b>	<b>0.050</b>

如表2-5所示，在合成数据集上，本文提出的方法得到的光源光谱估计的准确性在角误差  $e_{angular}$  和均方差  $e_{msquare}$  指标上都显着优于以前的方法。对于仿真场景实验，在本文提出的方法得到的结果中，平均角误差约为 3.8 度，相比平均角误差约为 5.4 度的 PWIR-resnet50 方法降低了约 29%；而基于统计的方法得到的平均角误差约为 25 度左右。

如图2-7所示，基于统计的方法估计出的光源光谱曲线更容易偏离真实的光源光谱且具有类似的趋势，而基于学习的方法的估计具有另一种类似的趋势，并且基于学习的方法普遍优于基于统计的方法。值得注意的是，由于引入了平滑损失，本文提出的方法在光源曲线中具有峰值的光源（例如 LED）和具有平滑光谱曲线的光源（例如自然光、白炽灯）的两种情况下的光源光谱估计都表现良好。从如图2-7的第 2、3 行可以看到，本文提出的方法对光谱曲线中带有尖峰的光源估计得更准，而对第 1、4 行中不带有尖峰的光源的估计有明显的偏差，可能的原因是带有尖峰的曲线的特征更易于表示。

#### 2.5.4 真实场景实验

以往的研究普遍缺少真实场景的实验，原因在于：一、室外场景很难在改变光源条件的同时保证场景中物体位置不变化；二、购买多种光源的成本较高且拍摄更复杂。

为了更有效的评估各种光源估计方法，本实验在室内用 5 种不同的光源进行拍摄，虽然光源种类不如仿真实验的 101 种光源多，但都是实际室内光源中

经常使用的，有一定的代表性。

表 2-6 不同算法在真实数据集上的比较

方法	角误差/度				均方差			
	平均值	中位值	最小值	最大值	平均值	中位值	最小值	最大值
Gray World	29.32	27.29	0.983	85.65	0.177	0.163	0.017	0.358
Max Spectral	23.9	22.04	0.689	68.43	0.143	0.122	0.011	0.304
Gray Edge	36.82	33.83	1.149	105.5	0.195	0.171	0.016	0.409
镜面反射分解	33.49	31.37	1.071	91.85	0.18	0.156	0.018	0.374
深度展开网络	25.14	21.71	5.190	74.22	0.159	0.142	0.084	0.251
PWIR-alexnet	14.18	11.43	0.995	81.12	0.092	0.071	0.004	0.202
PWIR-resnet50	6.693	5.487	0.751	28.03	0.039	0.030	0.019	0.13
本文提出的方法	<b>4.485</b>	<b>3.163</b>	<b>0.102</b>	<b>28.08</b>	<b>0.016</b>	<b>0.015</b>	<b>0.002</b>	<b>0.057</b>

如表2-6所示，本文提出的方法在真实数据集上仍然是最有效的。本文构建的数据集是在同一个暗室实验室场景中捕获的，使用本文构建的数据集中的一些样本进行重新训练容易过拟合。因此，本文直接使用在仿真数据集上训练的模型进行测试。数据集之间的差异导致预测的角误差增加，但仍在可接受的范围内。对于真实场景实验，在本文提出的方法得到的结果中，平均角误差约为 4.4 度，相比平均角误差约为 6.6 度的 PWIR-resnet50 方法降低了约 32%；而基于统计的方法得到的平均角误差约为 30 度左右。

## 2.5.5 消融实验

为了充分研究网络结构和损失函数设计对本文提出的方法的作用，本文对仿真数据集进行了 8 次消融实验。

表 2-7 高光谱光源估计的消融研究

是否加入自注意	损失函数类型	是否加入平滑损失	角误差/度	均方差
×	均方差	×	5.53	0.0176
×	均方差	✓	4.67	0.0129
×	角误差	×	4.51	0.0165
×	角误差	✓	4.44	0.0238
✓	均方差	×	4.86	0.0150
✓	均方差	✓	4.73	0.0175
✓	角误差	×	3.99	0.0162
✓	角误差	✓	<b>3.84</b>	<b>0.0143</b>

### 光谱联合空间自注意的影响

在关于是否加入自注意机制的消融实验中，不加入自注意的对比实验将特征图跳过光谱联合空间自注意模块，直接输入解码器。消融实验的定量结果如

表2-7所示，在使用角误差损失函数以及加入平滑损失的情况下，自注意模块可以将角误差从 4.44 度降低到 3.84 度，降低了 13.5%，均方差降低了 39.9%。

### 损失函数类型的影响

在关于损失函数类型的消融实验中，损失函数类型为均方差时，总损失函数表达式2-5中默认使用的  $e_{angular}$  被  $e_{msquare}$  代替。角度损失函数比均方更适合光源估计，在加入自注意和平滑损失的情况下，使用角误差损失函数相比使用均方差损失函数，光源光谱估计结果的角误差可以减少 18.8%，均方差也可以减少 18.2%。

可能的原因在于，不同波段的误差对角误差的贡献不同，而对均方差的贡献相同。例如，假设，

$$GT = [1, 3, 2] ,$$

$$pred_a = [1, 3, 3] ,$$

$$pred_b = [2, 3, 2] .$$

与光源光谱真值  $GT$  相比，第一种光源光谱预测  $pred_a$  在第三个波段存在误差，第二种光源光谱预测  $pred_b$  在第一个波段存在误差，误差值同样为 1，计算出来的均方差相同而角误差不同。

$$e_{angular}(GT, pred_a) = 11.17 ,$$

$$e_{angular}(GT, pred_b) = 13.51 ,$$

$$e_{msquare}(GT, pred_a) = 0.33 ,$$

$$e_{msquare}(GT, pred_b) = 0.33 .$$

因此，角误差相比于均方差更适合描述高光谱光源估计的准确性，作为损失函数也更有效果。

### 平滑损失的影响

在关于是否加入平滑损失的消融实验中，不加入平滑损失时，总损失函数表达式2-5中的平滑系数  $\eta$  从 0.5 改为 0。在加入自注意且使用角误差损失函数的情况下，引入平滑损失后，角误差总体上可降低 3.7%，但对光谱曲线平滑的光

源(自然光、白炽灯)有更明显的影响，在图2-7第四行展示的场景中，引入平滑约束的损失函数使得角误差从7.81度减少到4.99度，减少了36.1%。

## 2.6 本章小结

在本章中，为了解决从单张高光谱图像中获取光源光谱的问题，本文提出了一种端到端网络来解决高光谱光源估计问题。

本文提出的方法使用更深的主干网络，引入通道和空间自注意以及平滑损失，显著优于以前的方法。本方法在仿真图像和真实图像上的良好性能展示了其有效性、灵活性以及泛化能力。本文构建了一个具有不同光源的真实高光谱图像的大型数据集进行测试，可以将其用于未来的训练和分析工作。相较于传统方法在估计的光源光谱的准确度上取得了明显的进步。这主要归功于本章方法在以下几个方面的研究：

1. 提出了光谱联合空间自注意机制，作用于编码器输出的特征图，用于突出重要的光谱通道和图像中有效区域的作用。
2. 通过分析常见光源光谱曲线的平滑性特点，提出了平滑约束的损失函数，引入到模型训练过程中，避免了不实际的病态结果，从而提升了光源估计的准确度。

本章虽然在全局的光源光谱估计上取得了进步，但是对于局部光源光谱估计仍缺少研究，所以在未来的研究工作中需要补充和改进，进而设计出对于像素级光源光谱进行准确估计的算法。希望本文提出的无白板光源校准方法能够将现有的未校正的高光谱数据集校正为准确的高光谱反射率数据集，并更好地用于其他视觉任务。为此，后文将联合高光谱图像光源光谱估计和本征分解，并结合图像增强和材质分割任务，进一步探索高光谱成像技术的应用落地。



## 第三章 光源光谱引导的图像增强

### 3.1 引言

图像增强任务包含的研究范围很广，涉及图像处理流程的各个部分。

如图3-1所示，一个标准的图像处理流程<sup>[73]</sup>包括拜耳 (Bayer) 解码过程和后续的增强恢复过程。增强方法有曝光校正<sup>[74]</sup>、白平衡<sup>[75]</sup>、对比度增强<sup>[76]</sup>、局部和全局的色调映射<sup>[77]</sup>等，而这些方法均会受场景中光源的影响。本文在第二章中进行的光源光谱估计，将被用来引导图像增强。

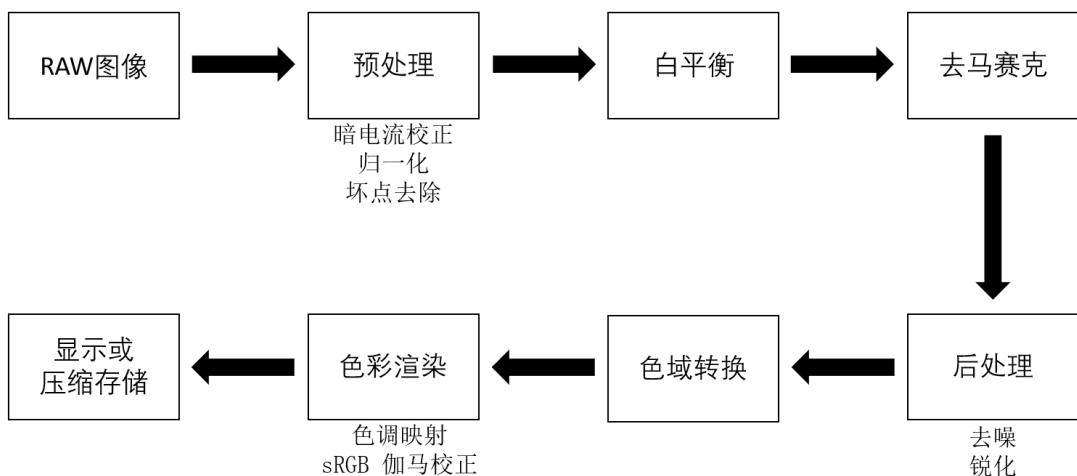


图 3-1 图像处理流程图 不同的相机硬件实现可能会有所不同，但是处理顺序类似

相机在生成全彩图像之前，需要先预处理从传感器获得的原始数据，去除噪点和伪影，一个常用的预处理步骤是缺陷像素校正，基本方法为通过对其邻域中准确记录的数据进行插值来估计缺陷位置的正确像素值。下一步是进行白平衡，校正在不同光源下引入的颜色偏差。在下一步进行去马赛克时，利用像素邻域信息来估计未测量的像素颜色，但这个过程会引入大量的预测值，就需要在后续的图像处理流程中进行去噪和锐化。人眼对尖锐的边缘高度敏感，且在水平和垂直方向的敏感度高于对角线方向，这也是许多相机厂商在进行图像锐化时会遵循的规律<sup>[78]</sup>。相机与人眼对颜色的敏感度并不相同，相机在图像处理

时需要将 RGB 值转换到更符合人眼视觉的色域空间。色彩渲染是为了图像显示而设计的，这个过程涉及动态范围压缩和色调映射，最常见的渲染空间是 sRGB 颜色空间<sup>[79]</sup>。

本研究中的图像增强任务处于图像处理中色彩渲染这一环节，涉及高动态范围 (High Dynamic Range, HDR) 场景的色调范围调整。动态范围是图像中最大亮度与最小亮度的比值，大于  $2^{16}$  称为高动态范围，大于  $2^{10}$  称为增强动态范围 (Enhanced Dynamic Range, EDR)，小于等于  $2^{10}$  称为低动态范围 (Low Dynamic Range, LDR)。通常将 HDR 图像和 EDR 图像统称为 HDR 图像，而常见的 8 位图像则是 LDR 图像<sup>[80]</sup>。受硬件限制，现有的技术不能使图像显示像真实场景一样具有大的动态范围，不经过任何处理而直接显示 HDR 图像将导致场景的暗区和亮区的细节丢失，通常可以用色调映射的方法解决这个问题。色调映射是一种在保持视觉效果基本不变的前提下将高动态范围图像映射到常规低动态显示设备上进行显示的技术<sup>[81]</sup>，目的是使得动态范围压缩后显示出的图像更符合人类视觉系统 (Human Visual System, HVS) 的感知。

色调映射常用的方法是直方图均衡化<sup>[82]</sup>，从全局上提高整幅图像的对比度。然而，在处理高动态范围的场景时，全局操作往往会出现不足。在深度学习时代，基于颜色变换的方法通常以全局-局部协同的方式用于高分辨率图像的实时处理。根据获取颜色变换函数的不同方式，这些方法可以分为仿射变换矩阵<sup>[83-84]</sup>、多层感知器 (MultiLayer Perceptron, MLP)<sup>[85]</sup>、三维查找表 (3D Look Up Table, 3D LUT)<sup>[86-88]</sup>等。开创性的工作是高动态范围网络 (HDRNet)<sup>[83]</sup>，它在低分辨率下学习系数的双边网格 (Bilateral Grid)，并在全分辨率下执行从输入到输出的转换，从而在不降低色调映射的效果的前提下提高算法的运行速度。自适应三维查找表<sup>[87]</sup>将仿射变换矩阵替换为图像自适应的三维查找表，实现了快速、鲁棒的图像增强。

在本章中，本文使用的高光谱相机的光谱响应范围包含可见光和近红外，利用近红外波段代表室外太阳光照在图像中不同局部的强度，从而跨模态地去除 RGB 图像中光源强度的影响，更好地引导 RGB 图像的增强。本文将高光谱图像中的近红外波段近似地看做是本征信息分解输出的光照信息并与 RGB 图像共享，提出了局部亮度适应法，在保证颜色不变的情况下将高动态范围图像各区域的亮度压缩到低动态范围的亮度，便于色调映射。本文选择 HDRNet 作为图

像增强的基线，关键的挑战在于多模态信息的融合，包括将 HDRNet 与额外的低空间分辨率光谱图像集成。为此，本文提出了光谱感知自注意机制，将光源光谱嵌入到 RGB 特征图中，提升了 RGB 图像色调映射的效果，使得高动态范围图像压缩到低动态范围后仍保留丰富的色彩信息。

## 3.2 相关研究

### 3.2.1 高动态范围图像的亮度调整

高动态范围图像的亮度调整受人类视觉系统的启发，Drago 等人<sup>[89]</sup>对图像亮度的动态范围进行对数压缩，通过使用较小的对数底数来更好地增亮暗像素，并通过使用更大的对数底数来压缩较亮的像素。Kim 等人<sup>[90]</sup>在此基础上加入了非线性比例因子，更好地适配人眼视网膜的响应。Lee 等人<sup>[91]</sup>为了保留 HDR 图像中最暗和最亮区域的细节，提出了一种非对称的 S 型曲线进行亮度调整。

图像亮度直方图是用于调整图像亮度的常见的依据。Duan 等人<sup>[92]</sup>通过直方图均衡映射与线性映射的加权平均值来构造直方图，两个映射相互补充，以提供更好的亮度调整性能。出于人的视网膜在暗光条件下对亮度变化更敏感这一经验，Khan 等人<sup>[93]</sup>构造了非均匀的亮度直方图，较暗的像素值对应较窄的宽度。Lee 等人<sup>[94]</sup>将亮度值聚类为固定数量的组，对每一组亮度自适应地确定一个用于色调映射的参数。Oskarsson<sup>[95]</sup>采用动态规划方法求解聚类问题的全局最优解，并提出了对 HDR 图像亮度值的聚类。Yao 等人<sup>[96]</sup>通过在亮度直方图上使用高斯混合模型 (Gaussian Mixture Model, GMM)，估计了 HDR 图像的暗、亮区域的高斯分布，分别对暗区域和亮区域用高斯分布确定的参数进行亮度调整。

### 3.2.2 基于学习的色调映射

色调映射的研究广泛应用于深度学习算法。Rana 等人<sup>[97]</sup>提出了一种生成对抗网络来对各种场景的 HDR 图片进行高分辨率的色调映射，但是需要耗费大量的计算资源和时间。为解决这一问题，Cao 等人<sup>[98]</sup>用低分辨率的图像进行训练，对高分辨率和低分辨率的色调映射分别提出了优化后的架构。Panetta 等人<sup>[99]</sup>引入了一个注意力引导模块来克服色调映射时的伪影问题，并使用大规模低光照图像数据集扩充了色调映射的训练集。此外，Goswami 等人<sup>[100]</sup>还将 HDR 图像

按场景分类，并使用全卷积网络对图像进行语义分割，对于不同的类别单独学习色调映射。

由于真实配对的 HDR-LDR 色调映射图像对较少，半监督的方法也被应用于色调映射。Zhang 等人<sup>[101]</sup>在原有的由 HDR-LDR 图像对计算的损失函数上增加了其他两个仅由 LDR 图像计算的损失项，并利用未配对的 LDR 数据，使得训练时可以减少对配对的 HDR-LDR 数据的依赖。Guo 等人<sup>[102]</sup>则在损失函数中引入了基于客观的图像质量评估 (Image Quality Assessment, IQA) 指标的无监督项，进一步减少了色调映射对大规模数据集的依赖。

### 3.3 本文构建的图像增强数据集

为了支撑本文提出的图像增强算法，本文构建了高光谱与 RGB 图像的图像增强数据集。

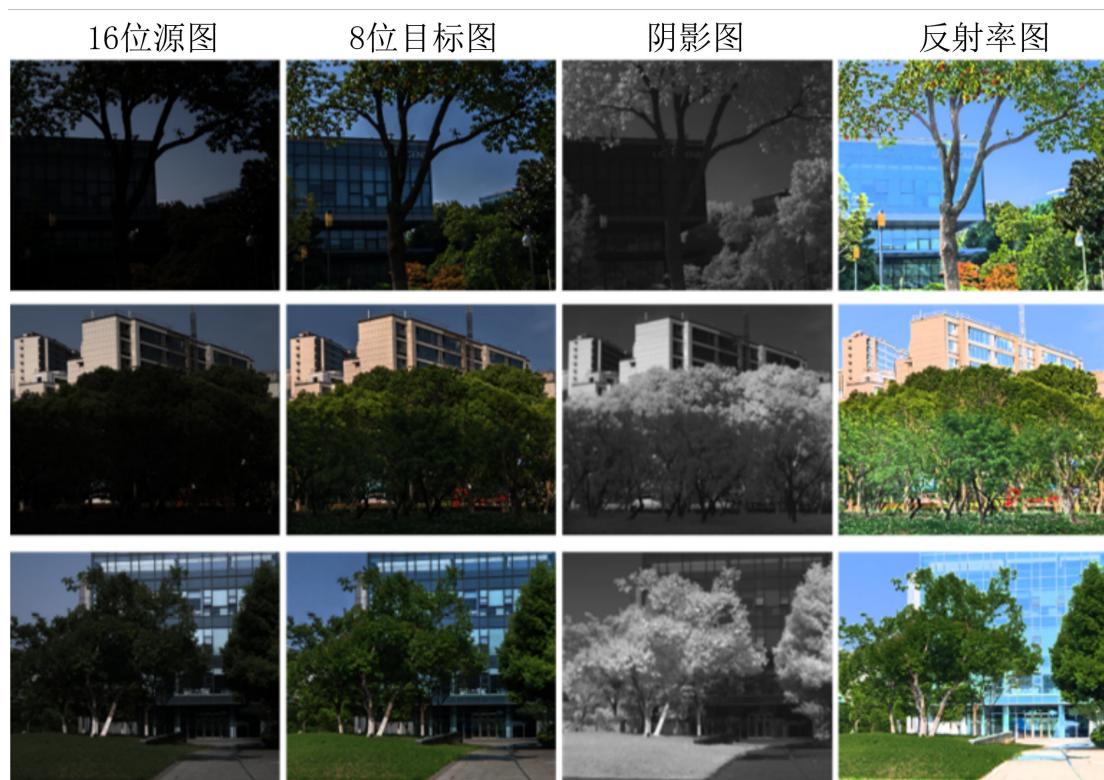


图 3-2 高光谱与 RGB 图像的图像增强数据集

据本文的调查，同时包含对齐的高光谱和 RGB 图像的图像增强数据集尚未被发现。为了弥补这一数据集缺口，本实验用智能手机和高光谱相机进行联合拍摄，构建了高质量的高光谱与 RGB 对齐的图像增强数据集，数据集包含 200

组图像，称为 Mobile-Spec 数据集。如图3-2所示，每组图像同时具有 16 位 RGB 源图、8 位 RGB 目标图、高光谱图像及其对应的阴影、反射率图像。

以上数据集面对的图像增强任务包含 RGB 图像动态范围压缩和色调映射，16 位 RGB 图像作为图像增强的输入，8 位 RGB 图像作为图像增强的输出。

为了保证 Mobile-Spec 数据集的高质量，本数据集的数据经过了严格的筛选。首先，对 16 位源图仅保留了高动态范围样本，动态范围由图像中最大像素值和最小像素值之间的比值确定。其次，滤除了 RGB 与高光谱图像之间存在显著对齐误差的样本。第三，考虑色差、清晰度、噪声、伪影等因素，对 8 位目标图进行主观评价，剔除了视觉效果不理想的样本。最后，为了增加数据集内容的多样性，剔除了相似度高的场景。

### 3.3.1 双相机系统及其对齐

本文构建的 Mobile-Spec 数据集是用高端商用智能手机和推扫式高光谱相机组成的双相机系统采集的。该推扫式高光谱相机的光谱范围为 400-1000 nm，包含 176 个光谱通道。16 位源图是智能手机图像处理过程中多个曝光帧的融合，8 位目标图代表相应的色调映射结果。智能手机和推扫式高光谱相机在进行数据采集时，处于尽可能相近的位置，并使用三脚架进行固定，在使用高光谱相机推扫拍摄的过程中同步完成智能手机的拍摄。

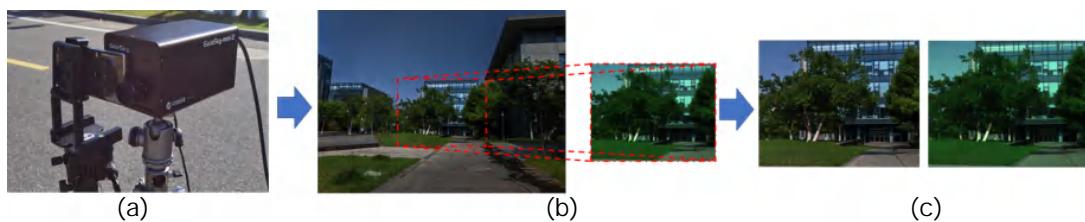


图 3-3 双相机系统的配准过程 (a) 包括智能手机和高光谱相机的双相机系统 (b)RGB 图像与高光谱图像的重叠区域由 SIFT<sup>[103]</sup> 算法检测得到，对智能手机捕获的 RGB 图像进行仿射变换，使之与高光谱图像渲染出的 RGB 图像对齐 (c) 对齐后的 RGB 和高光谱图像，其中高光谱图像被渲染为 RGB 形式

本实验采集到的高光谱图像分辨率为  $1057 \times 960 \times 176$ ，16 位源图、8 位目标图的分辨率均为  $4096 \times 3072 \times 3$ 。值得注意的是，高光谱图像与 RGB 图像在视场上存在相当大的差异。如图3-3所示，为了将具有宽视角的 RGB 图像对齐到高光谱图像上，本文采用了一种自动的对齐算法，其中包括尺度不变特征变

换 (Scale Invariant Feature Transform, SIFT)<sup>[103]</sup> 和估计最优的投影变换 (Projective Transformation) 算法。经过上述自动对齐算法，本实验将 RGB 图像对齐到高光谱图像上，使两种图像的空间分辨率均为  $1057 \times 960$ 。

### 3.3.2 光源的影响及阴影图

如图3-4所示，阳光在可见光范围内包含连续的宽谱，延伸到紫外 (UltraViolet, UV) 和红外 (InfraRed, IR) 区域，其强度在可见光光谱中相对均匀。相反，LED 光源的光谱特征是在蓝色区域有一个明显的峰值，其强度在近红外波长处迅速减弱。

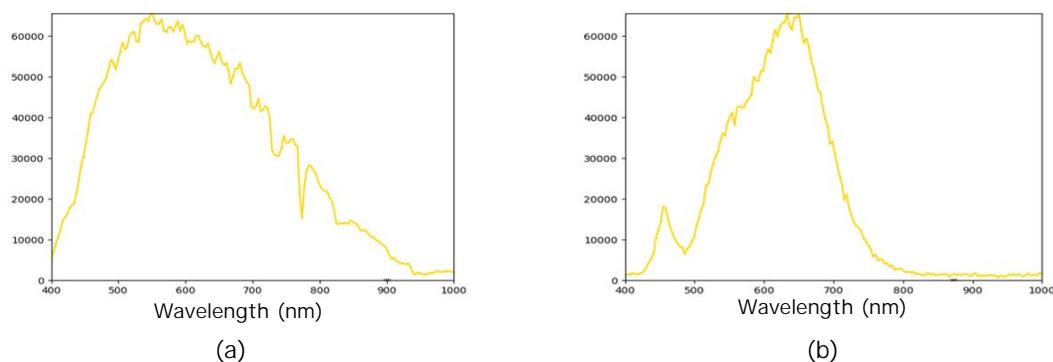


图 3-4 室外光与室内光的光谱区别 (a) 太阳光光谱曲线 (b) LED 光源光谱曲线

如图3-5，可以发现，不同的材质在太阳光下的近红外反射强度是接近的，可以近似地认为，高光谱图像中近红外部分只反映太阳光的强度，而可见光部分则反映颜色。

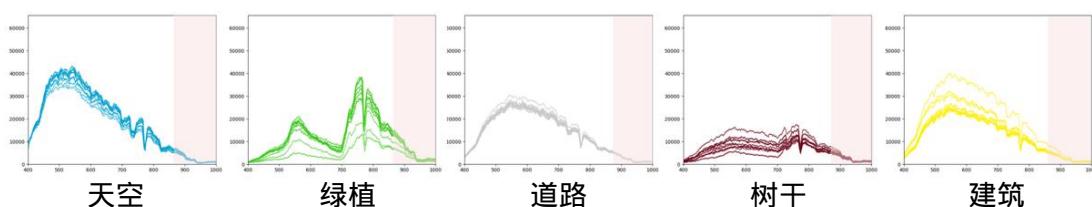


图 3-5 不同材质种类在太阳光下的光谱曲线，其中近红外波段表现出一致的趋势

进一步地，本文在 24 色标准色卡 (如图3-7) 上测试各种颜色在室内和室外光下的光谱曲线，图3-6(a) 反映出在太阳光下各种颜色在近红外波段同样表现出类似的趋势，而图3-6(b) 则表明在 LED 光源下各种颜色的近红外波段响应几乎为零。

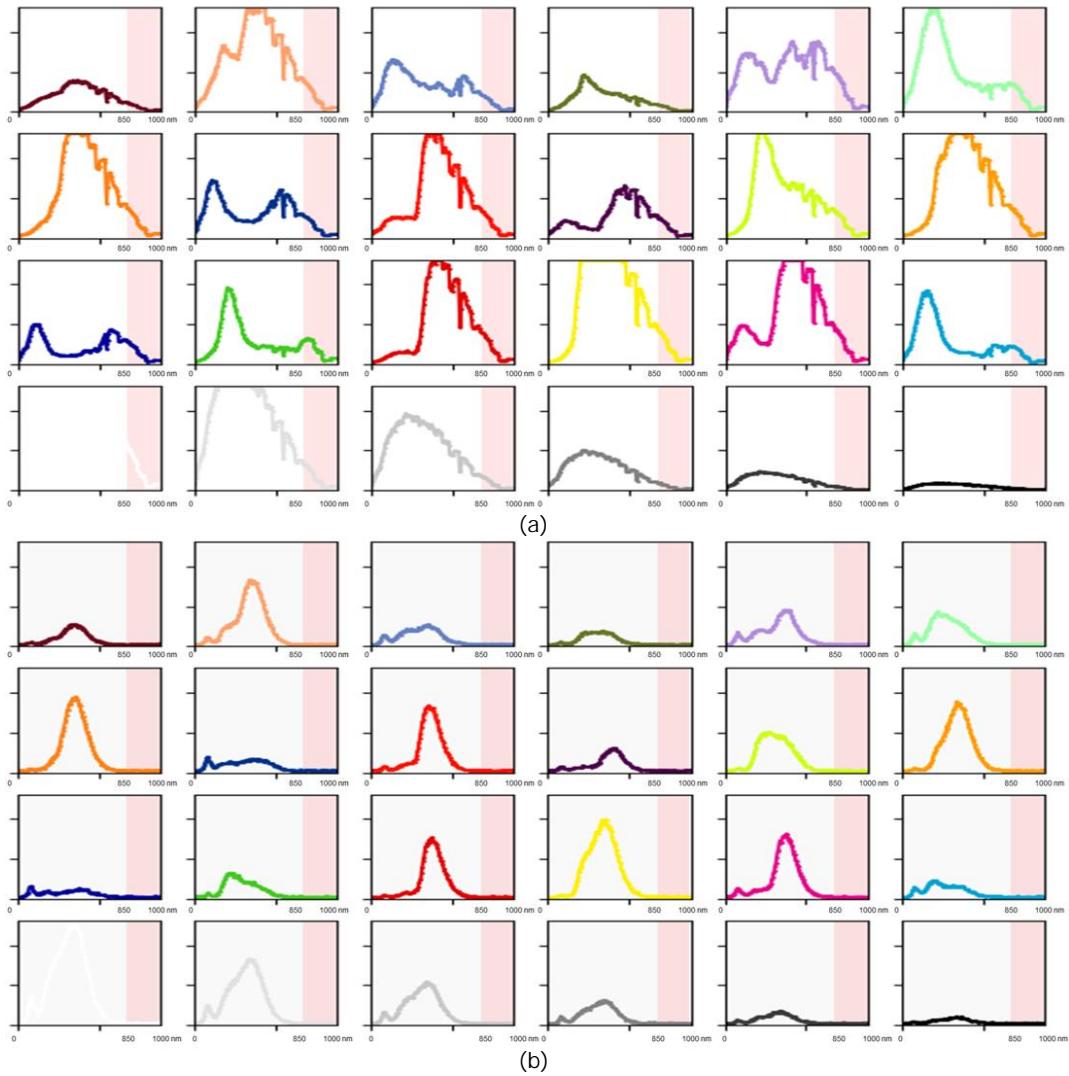


图 3-6 24 色标准色卡的每个颜色在不同光源下的光谱曲线，其中标红的区域为近红外波段中 850-1000nm 的部分 (a) 室外太阳光 (b) 室内 LED 光源

因此，近红外图像包含亮度信息且不包含可见光颜色信息，可以充当阴影图<sup>[104]</sup>，本实验通过对高光谱图像中 850nm 到 1000nm 的近红外波段进行平均来逼近阴影图。基于 Retinex 理论<sup>[105-106]</sup>，使用采集到的高光谱图像除以阴影图得到反射率图。

这样，本文就解决了真实阴影图难以获取的问题，从而避免使用算法进行本征分解，而是使用更可靠的直接测量的方法来获得阴影图。

为了利用太阳光在近红外的辐射以获得高质量的阴影图，并避免复杂的室内光源对数据集采集的不利影响，本文将 Mobile-Spec 数据集的采集场景限制在室外。



图 3-7 实验用到的 24 色标准色卡

### 3.4 本文提出的方法

在 HDRNet<sup>[83]</sup>的研究基础上，本文提出了局部亮度适应来调整高动态范围 RGB 图像不同区域的亮度，并提出光谱感知自注意机制，将光源光谱嵌入到 RGB 图像增强的特征图中，提高色调映射的效果。

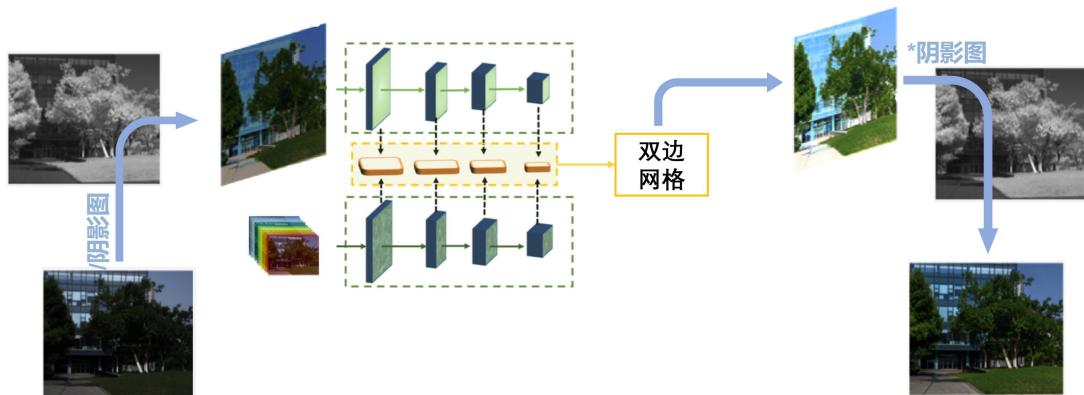


图 3-8 本文提出的图像增强方法示意图

本文提出的图像增强方法示意图如图3-8所示，本文参考 HDRNet 的方法进行图像增强，训练了一个双边网格用于色调映射。其中，代表光源强度的阴影图不参与色调映射过程，称为阴影先验。将额外引入的低分辨率光谱反射率图作为反射率先验，进一步引导双边网格的训练。浅蓝色箭头表示主要的图像处理流程，“/阴影图”表示使用逐像素的除法去除阴影图得到反射率图，“\* 阴影图”

表示使用逐像素的乘法将阴影图重新叠加到反射率图上，从而得到最终的图像增强结果。

### 3.4.1 局部亮度适应

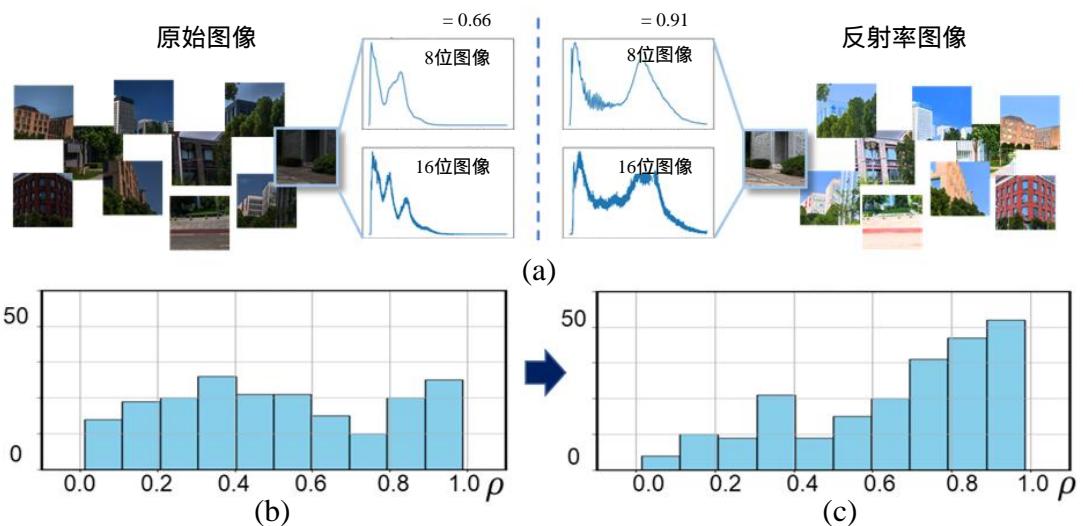


图 3-9 局部亮度适应的作用 (a)一对原始图像和反射率图像的亮度分布统计 (b)(c)8 位图像与对应的 16 位图像的亮度分布的皮尔逊相关系数 (Pearson correlation coefficient) $\rho$  在原始图像集和反射率图像集中的直方图

为了在色调增强过程中同时对高亮区域和阴影区域进行调整，HDRNet 中的双边网格深度设置为 8，可以容纳 8 种不同的亮度级别。然而，这种设计在处理高动态范围场景中的各种局部亮度变化时可能不具有足够的适应性。

本文认为，在这种情况下，去除阴影分量会简化色调映射过程，为模型的学习带来好处。如图3-9 (a) 所示，16 位输入和 8 位目标之间的像素值分布向量在原始图像中表现出 0.66 的皮尔逊相关系数  $\rho$ ，在反射率图像中增加到 0.91。系数  $\rho$  越大，表示 16 位图像与 8 位图像的亮度特征相似度越大，相互间的映射关系也就越简单。通过对 Mobile-Spec 数据集中  $\rho$  分布的统计发现，与原始图像集(图3-9 (b)) 相比，反射率图像集(图3-9 (c)) 具有更大比例的高  $\rho$  样本。

这一现象表明，将原始图像中的阴影分量去除，得到反射率图像后，可能会降低颜色映射学习的难度。

在本文提出的方法中，为了将阴影分量  $S$  转换为具有 8 个亮度等级的亮度图  $\hat{S}$ ，本文提出了一个局部亮度自适应模块，包含两层卷积和反卷积。该轻量级模块将缓解局部高动态范围区域亮度的剧烈变化，从而进一步增强自适应性。

通过以下等式可以将图像中各个局部不同的光源强度去除：

$$\begin{aligned} R_{rgb} &= I_{rgb}/\hat{S}, \\ R_{hsi} &= I_{hsi}/\hat{S}. \end{aligned} \quad (3-1)$$

其中， $I_{rgb}$  和  $I_{hsi}$  分别为输入的 16 位 RGB 图像和低空间分辨率的高光谱图像， $R_{rgb}$  和  $R_{hsi}$  分别为去除光源强度后的对应的反射率图像。

随后，16 位图像的反射率图像  $R_{rgb}$  和低空间分辨率光谱图像的反射率图像  $R_{hsi}$  将被送入双边网格的系数预测部分，用于预测图像增强的颜色变换系数。

### 3.4.2 光谱感知自注意

从原始图像转换到反射率图像后，本文利用低空间分辨率光谱图像的反射率图像  $R_{hsi}$  来增强双边网格的颜色映射学习。

本文认为，由于双边网格系数的预测是在更低的空间分辨率下进行的，因此，在空间维度上的下采样不会给颜色映射造成明显的退化。由此，本文设计了光谱感知自注意 (Spectral Perception Self-Attention, SPSA) 模块，利用  $R_{hsi}$  引导  $R_{rgb}$  在颜色映射学习中产生更好的双边系数。 $R_{hsi}$  与  $R_{rgb}$  在空间分辨率上相互配准，均为  $256 \times 256$ ，用于 HDRNet 中的低分辨率双边网格的系数预测。

SPSA 模块对  $R_{hsi}$  和  $R_{rgb}$  进行卷积层堆叠处理，生成分层特征图。将这些特征图  $\mathcal{F}_{R_{msi}}$ 、 $\mathcal{F}_{R_{rgb}}$  沿通道维拼接，形成交叉谱特征图  $\mathcal{F}_R$ 。然后将  $1 \times 1$  卷积  $W_1$  和  $3 \times 3$  深度卷积  $W_3$  应用到  $\mathcal{F}_R$  中生成查询  $Q$ 、密钥  $K$  和值  $V$ ：

$$\begin{aligned} Q &= W_1^Q W_3^Q \mathcal{F}_R, \\ K &= W_1^K W_3^K \mathcal{F}_R, \\ V &= W_1^V W_3^V \mathcal{F}_R. \end{aligned} \quad (3-2)$$

$K$  和  $Q$  被重塑为  $\hat{K} \in \mathbb{R}^{C \times HW}$  和  $\hat{Q} \in \mathbb{R}^{HW \times C}$ ，它们的点积交互产生了光谱通道间的自注意  $A \in \mathbb{R}^{C \times C}$ 。光谱感知自注意  $A$  对重构  $\hat{V} \in \mathbb{R}^{HW \times C}$  的不同通道的重要性进行了重新加权。将卷积  $W_3$  沿通道维度拆分为  $W_3^{hsi}$  和  $W_3^{rgb}$ ，对应于  $\mathcal{F}_{R_{msi}}$  和  $\mathcal{F}_{R_{rgb}}$  的通道，然后重塑为  $\hat{W}_3^{hsi}$  和  $\hat{W}_3^{rgb}$  以适应  $\hat{V} \cdot A$  的形状。最

终，SPSA 模块可表示为：

$$\begin{aligned} A &= \text{softmax}(\sigma \hat{K} \cdot \hat{Q}), \\ \hat{\mathcal{F}}_{R_{hsi}} &= \hat{W}_3^{hsi} \hat{V} \cdot A + W_3^{hsi} \mathcal{F}_{R_{hsi}}, \\ \hat{\mathcal{F}}_{R_{rgb}} &= \hat{W}_3^{rgb} \hat{V} \cdot A + W_3^{rgb} \mathcal{F}_{R_{rgb}}. \end{aligned} \quad (3-3)$$

$\sigma$  是一个可学习的缩放参数，用来控制  $\hat{K}$  与  $\hat{Q}$  点积的大小。SPSA 模块作为残差学习，通过自适应的方式将交叉谱特征  $\hat{V} \cdot A$  与原始特征  $\mathcal{F}_{R_{hsi}}, \mathcal{F}_{R_{rgb}}$  进行加权融合。SPSA 模块的融合输出  $\hat{\mathcal{F}}_{R_{hsi}}, \hat{\mathcal{F}}_{R_{rgb}}$  以递进的方式馈入下一层。

## 3.5 实验过程与结果分析

在本章研究中，实验设置所使用的硬件平台和相关软件同第二章实验设置部分。

在 Mobile-Spec 数据集中，80% 的样本分配给训练集，剩下的 20% 分配给测试集，批大小 (Batch Size) 为 4，学习率为 0.0001。网络参数的优化是通过最小化 L2 损失来实现的。在训练过程中，输入图像从  $1057 \times 960$  被裁剪为  $512 \times 512$  大小，这些图像在低分辨率流中被进一步下采样为  $256 \times 256$ 。

### 3.5.1 评价指标

本文采用峰值信噪比 (Peak Signal-to-Noise Ratio, PSNR)、结构相似性 (Structural SIMilarity, SSIM)<sup>[107]</sup> 和色差  $\Delta E^*$  共 3 个指标来验证图像增强方法的有效性。

图像增强的结果与目标图计算出的 PSNR 越大，表示图像增强的效果越好。对于空间分辨率为  $m \times n$  的两幅图像  $I(i, j)$  和  $K(i, j)$ ，PSNR 的计算方式如下所示：

$$\begin{aligned} MSE &= \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2, \\ PSNR &= 10 \cdot \log_{10} \left( \frac{MAX_I^2}{MSE} \right) = 20 \cdot \log_{10} \left( \frac{MAX_I}{\sqrt{MSE}} \right). \end{aligned} \quad (3-4)$$

其中  $MAX_I$  为图像可能的最大像素值，对于 8 位图像则为 255。

SSIM 对样本  $x$  和  $y$  的三个方面的相似性进行比较：亮度 (luminance)  $l(x, y)$ 、对比度 (contrast)  $c(x, y)$  和结构 (structure)  $s(x, y)$ ，SSIM 值越大表示样本间的相似度越高，计算方式如下：

$$\begin{aligned} l(x, y) &= \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1}, \\ c(x, y) &= \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}, \\ s(x, y) &= \frac{\sigma_{xy} + c_3}{\rho_x\sigma_y + c_3}. \end{aligned} \quad (3-5)$$

其中  $\mu_x$  和  $\mu_y$  分别为样本  $x$  和  $y$  的均值， $\sigma_x^2$  和  $\sigma_y^2$  分别为样本  $x$  和  $y$  的方差， $\sigma_{xy}$  为样本  $x$  和  $y$  的协方差， $c_1 = (k_1 MAX_I)^2$ ， $c_2 = (k_2 MAX_I)^2$ 。一般来说， $c_3 = c_2/2$ ， $k_1 = 0.01$ ， $k_2 = 0.03$ ，本文也采用上述默认值。

SSIM 将上述三个方面的相似性相结合，计算方式如下所示：

$$SSIM(x, y) = [l(x, y)^\alpha \cdot c(x, y)^\beta \cdot s(x, y)^\gamma]. \quad (3-6)$$

其中  $\alpha$ ， $\beta$  和  $\gamma$  根据上述三个方面的相似性的重要性确定，本文将  $\alpha$ ， $\beta$  和  $\gamma$  均设为 1，由等式3-5和等式3-6可以得到：

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}. \quad (3-7)$$

$\Delta E^*$  衡量 CIELAB 空间中两种颜色的差异，图像增强的结果与目标图间计算出的  $\Delta E^*$  值越小表示图像增强的颜色精度越高、图像质量越好。对于 CIELAB 空间中的两种颜色  $(L_1^*, a_1^*, b_1^*)$  和  $(L_2^*, a_2^*, b_2^*)$ ， $\Delta E^*$  的计算方式如下所示：

$$\Delta E^* = \sqrt{(L_2^* - L_1^*)^2 + (a_2^* - a_1^*)^2 + (b_2^* - b_1^*)^2}. \quad (3-8)$$

### 3.5.2 定性分析

图3-10给出了代表性例子的定性比较。本文提出的方法得到的结果在绿植区域表现出增强的对比度和更宽的动态范围，而其他方法增强出的色彩则显得比较暗淡。本文提出的方法得到的结果在墙面区域与目标图的颜色偏差较小，表

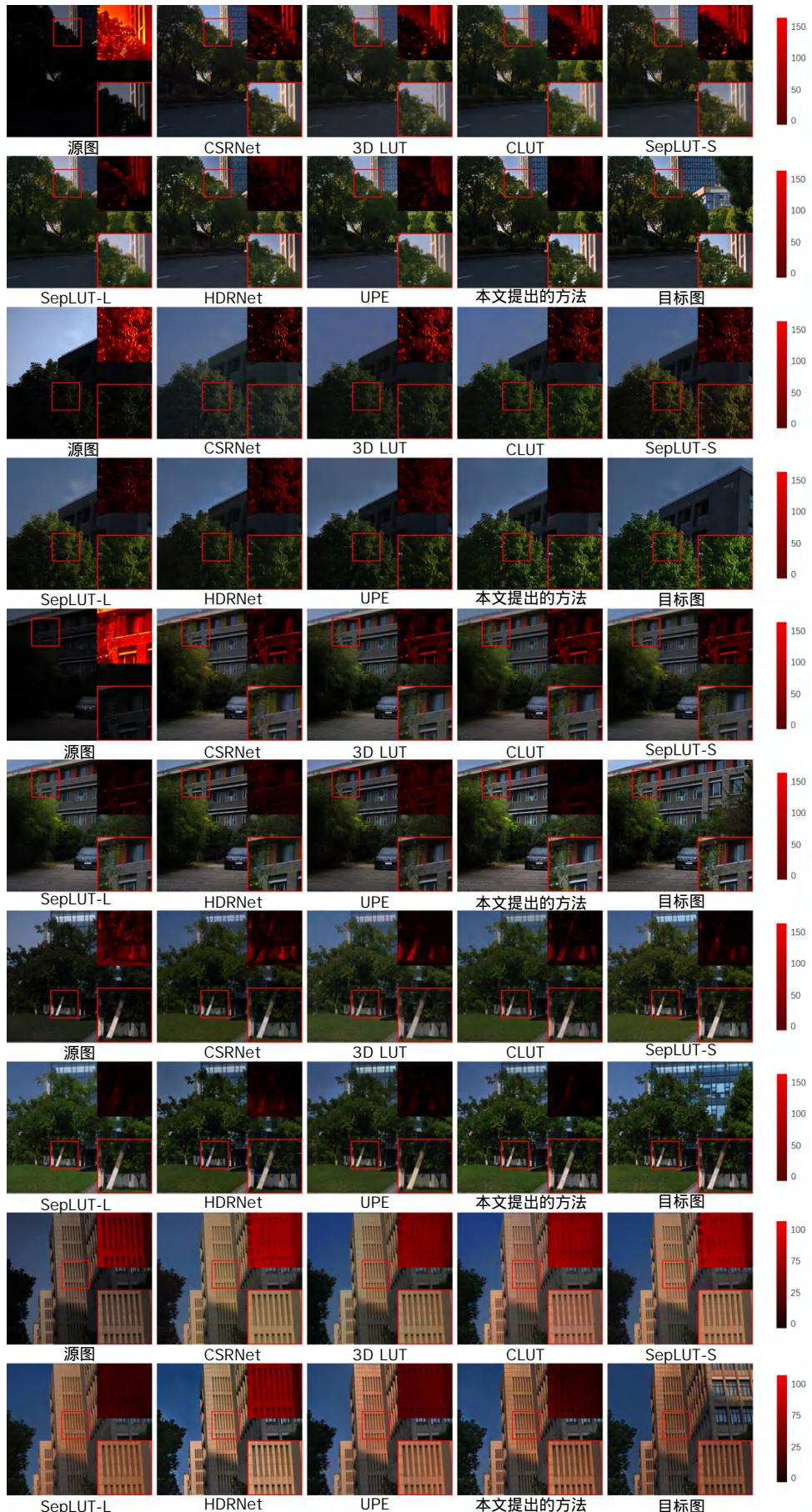


图 3-10 定性比较 每张图片右上角显示的误差图说明了结果与目标图的差异

明光谱信息集成对于提高摄影真实感方面产生了明显的提升。

从以上样本中，可以得出以下结论：(1) 在局部亮度剧烈变化的区域，例如具有不同光影分布的阳光下的树叶，本文提出的方法通过去除图像的阴影分量  $S$ ，自适应地进行高动态范围的场景的色调调整。(2) 将低空间分辨率光谱图像的反射率分量  $R$  与扩展的颜色通道相结合，使得用本文提出的方法得到的色调增强结果相比于目标图有最小的颜色偏差。但考虑到 Mobile-Spec 数据集以天空和绿植为主，蓝色和绿色等冷色调占据了大部分颜色，导致对红色等较暖色调的映射学习不足，今后需要在数据集中扩充暖色调样本。

引入额外的低空间分辨率光谱图像可以在一定程度上缓解不同颜色之间学习的不平衡性。与其他方法相比，本文提出的方法获得了更准确和更符合人眼感受的颜色。

由上述实验结果可以看出，本文提出的方法在阴影和反射率先验的引导下，对于高动态范围的场景，获得了更加准确和令人愉悦的颜色。

定性比较强调了将图像分解为阴影、反射率的有效性。该框架为色调增强提供了明确的引导，并克服了光谱图像的内在复杂性。通过对低空间分辨率光谱图像在色调增强任务中的探索，可以为光谱信息在移动摄影中的应用奠定基础。

### 3.5.3 定量分析

表 3-1 不同图像增强算法的定量比较

	CSRNet	3D LUT	CLUT	SepLUT-S	SepLUT-L	HDRNet	UPE	本方法
PSNR↑	26.34	27.52	27.30	27.57	28.08	27.75	28.19	<b>29.83</b>
SSIM↑	0.923	0.926	0.938	0.933	0.944	0.939	0.946	<b>0.967</b>
$\Delta E^* \downarrow$	6.44	5.39	4.63	5.37	4.26	5.12	4.79	<b>3.60</b>

将本文提出的方法与现有的基于颜色变换的方法进行了图像增强任务的对比分析。在表3-1中，基于 MLP 的方法，条件序列修饰网络 (Conditional Sequential Retouching Network, CSRNet)<sup>[85]</sup> 的 PSNR 略低，为 26.34dB。相比之下，基于 LUT 的方法，如 3D LUT<sup>[87]</sup>、CLUT<sup>[88]</sup>、SepLUT-S<sup>[86]</sup> 等，PSNR 相近，在 27.30-27.57dB 之间。值得注意的是，具有较大参数的 SepLUT-L 表现出了优越的性能，PSNR 达到了 28.08dB，这是由于 SepLUT-L 同时利用 1D 和 3D LUT 进行增强，其中 1D LUT 以图像自适应的方式调整图像对比度，实现了更均匀的对比度分布。此外，

UPE<sup>[84]</sup>利用平滑损失约束，估计了一个图像到光源的映射，该映射专为增强不同光源条件下的图像而设计，并取得了良好的处理性能，PSNR 达到了 28.19dB。

由表3-1可知，本文方法的图像增强性能达到了目前领先的水平，先前已有的方法在学习 HDR 场景下准确的色调映射方面面临挑战，而本文提出的方法通过有效地利用阴影和反射率先验来提供明确的引导，PSNR 达到了 29.83dB，比先前最优的方法提高了 5.8%。

为了避免分解不准确的影响，本实验将阴影先验和反射率先验（记为  $S^*$ ,  $R^*$ ）的理想值单独整合到 HDRNet 基线中进行验证。

### 阴影先验 $S^*$

表 3-2 阴影先验的影响

	PSNR↑	SSIM↑	$\Delta E^* \downarrow$		PSNR↑	SSIM↑	$\Delta E^* \downarrow$
HDRNet	27.75	0.939	5.12	HDRNet+ $S^*$	28.68	0.957	4.29

从表3-2可以看出，HDRNet 基线的 PSNR 为 27.75dB，而引入阴影先验  $S^*$  后使得本文提出的方法得出的 PSNR 提高到 28.68dB。值得注意的是，在这个实验中，HDRNet 的结构保持不变，只是将输入的 RGB 图像转换到反射率空间，就得到了上述的明显提升。

这一发现表明，使用阴影先验  $S^*$  将原始图像转换到反射率图像是一种简单而有效的色调增强任务设计，在不同的高动态范围场景中表现出普遍的适应性。

### 反射率先验 $R^*$

鉴于光谱成像能力在移动设备上的实际限制，本文开展了消融研究，以考察低空间分辨率光谱图像的光谱和空间分辨率配置对图像增强的影响。首先将空间分辨率保持在默认设置  $16 \times 16$ ，并提取不同光谱范围的波段。低空间分辨率光谱图像有 10 个波段，波长范围为 400-1000nm，每个波段带宽为 60nm。值得注意的是，400-760nm 包含 6 个可见光波段，而 760-1000nm 包含 4 个近红外波段。如表3-3所示，使用越靠近红色通道的波长作为图像增强的引导，得到的 PSNR 越高。可能的解释是，本文构建的数据集的组成中占较大面积的类别是天空和绿植，而红色物体的面积和数量相对有限。因此，在训练过程中，HDRNet 基线倾向于优先考虑蓝色和绿色波长而不是红色波长，其中，使用全部 10 个通道 (400-1000nm) 组合的效果最好，PSNR 达到了 29.68dB。

在本文看来，细粒度的光谱信息增强了双边网格的颜色感知能力，并补偿

**表 3-3 反射率先验在不同波段的影响**

波段范围	PSNR↑	SSIM↑	$\Delta E^* \downarrow$
400-520 nm	28.74	0.957	4.00
520-640 nm	29.09	0.964	3.82
640-760 nm	29.19	0.962	3.92
400-760 nm	29.24	0.967	3.72
760-1000 nm	29.07	0.965	3.88
400-1000 nm	29.68	0.968	3.55

了不同颜色通道的不平衡学习。

**表 3-4 不同空间分辨率的反射率先验的影响**

空间分辨率	PSNR↑	SSIM↑	$\Delta E^* \downarrow$
1×1	29.05	0.963	4.06
4×4	29.21	0.963	3.80
16×16	29.68	0.968	3.55
64×64	29.81	0.969	3.47
256×256	29.78	0.967	3.49

在下一个实验中, 使用全部 10 个通道, 空间分辨率从  $1\times 1$  逐步调整到  $256\times 256$ 。如表3-4所示, 随着空间分辨率的提高, 各项评价指标有提高的趋势。考虑到光谱图像在移动设备上空间分辨率的限制, 本文认为低空间分辨率光谱图像的最优参数配置可能与双边网格的空间分辨率 ( $16\times 16$ ) 对齐, 也就是低分辨率图像流输入的空间分辨率 ( $256\times 256$ ) 对应的 16 倍下采样。

实验结果表明, 光谱图像  $16\times 16$  的空间分辨率配置在引导 RGB 图像增强时产生了令人满意的结果, 在更高的空间分辨率上的实验结果有明显的边际效应。

### 3.6 本章小结

先前的方法在学习高动态范围场景下准确的色调增强方面面临挑战, 而本文提出的方法通过有效地利用阴影和反射率先验来提供明确的引导。通过大量的实验和测试, 证明了本文提出的方法在动态范围压缩和色调映射上相对于已有的方法有所提升。这主要归功于本章方法在以下几个方面的研究:

1. 利用室外太阳光近红外波段近似出的阴影图, 提出了局部亮度适应法, 调节高动态范围 RGB 图像不同区域的亮度。
2. 通过将光源光谱作为图像增强网络的额外输入, 提出了光谱感知自注意机制, 将光源光谱嵌入到 RGB 特征图中, 提升了 RGB 图像色调映射的效果。

本章虽然在图像的动态范围压缩和色调映射上进行了研究，在图像锐度增强、去噪、除雾等方面还缺少研究，值得在未来进一步研究光源光谱对这些方面的辅助效果。本方法使用推扫式高光谱相机获取的图像引导 RGB 的图像增强，在商用高端智能手机上目前还无法实现高精度的光谱图像采集，需要进一步探索受限的光谱成像对 RGB 图像增强的帮助。



## 第四章 光源光谱校正的材质分割

### 4.1 引言

材质分割是一个重要的研究课题，从视觉外观中识别材料对计算机视觉任务至关重要，特别是那些涉及与现实世界交互的任务。

在高光谱图像研究中，分割和分类之间的边界并不清楚。图像分类通常是指给整个图像分配一个或多个标签的任务，而图像分割是指为图像中的每个像素分配一个类别。大量关于高光谱图像分类的研究实际上都是在进行语义分割。例如，高光谱图像分类<sup>[108-110]</sup>中最先进的方法通过分割模型来进行逐像素的分类。

高光谱图像分类问题是根据每个像素向量的光谱或光谱空间特性为其分配一个类别标签的任务，这与计算机视觉中的语义分割任务密切相关，相应地，高光谱图像的材质分类任务也与材质分割任务高度相关。对于与材质相关的高光谱图像中的每个像素，都对应一个由反射率数据组成的向量。这些向量形成了一个宽度为  $w$ 、高度为  $h$  和波段数为  $c$  的高光谱数据立方体 ( $w \times h \times c$ )。在 RGB 图像中  $c = 3$ ，而在高光谱图像中  $c$  可以高达数百。由于不同的材质有它们本征的光谱反射率特征，可以利用这些信息将高光谱图像的像素进行分类。

鉴于高光谱图像具有更精细的光谱细节特征以及图谱合一等优势，其被广泛应用于分类、检测和分割等任务。Ahmad 等人<sup>[111]</sup>回顾了从传统方法到深度学习模型的高光谱图像分类，Su 等人<sup>[112]</sup>回顾了高光谱目标定位中的异常检测方法，Huang 等人<sup>[113]</sup>探索了在城市道路场景的高光谱图像的语义分割，Liang 等人<sup>[35]</sup>在 RGB、偏振和近红外图像中实现了多模态的材质分割。尽管现有的基于高光谱图像的分类、检测和分割方法已经取得了显着的成效，但是其中大多数方法都忽略了场景光源条件的影响。

在本章中，本文使用光源光谱校正的方式，得到不同材质的本征光谱反射率特征，从而将高光谱图像更有效地运用于材质分割。为支撑本文的深度学习

算法，本文提出了 RGB 图像与高光谱图像对齐的材质分割数据集。考虑到跨模态数据总量的平衡，本文使用高空间分辨率的 RGB 图像和低空间分辨率且具有多通道的光谱图像相结合的方式进行材质分割，从而将高光谱图像光谱特征更细的优势带给 RGB 图像，提高材质分割的准确度。

## 4.2 相关研究

### 4.2.1 基于阴影和反射率先验的本征分解

光谱反射率与物体材质直接相关，使用基于阴影和反射率先验的本征分解能将与光源和形状无关的反射率提取出来。

一种研究思路是用物体的形状及其造成的阴影作为约束条件。Barron 和 Malik<sup>[114-116]</sup>提出了三个假设：(1) 物体形状的曲率变化较小，表面较为光滑；(2) 表面法线的取向是各向同性的，物体表面可以朝向任何一个方向；(3) 在被遮挡物体的边界附近，表面法线的方向朝向外部；根据这三个假设进行图像本征分解。Chen 等人<sup>[117]</sup>对 RGB-深度 (RGB-D) 图像进行本征分解，通过 RGB 图像和深度图像消除预测的阴影的歧义。Jeon 等人<sup>[118]</sup>假定光源平滑变化且物体表面为朗伯表面，据此确定物体表面法线，作为本征分解的约束条件。

另一种研究思路是从反射率本身出发进行本征分解。Shen 等人<sup>[119]</sup>利用物体表面色调的局部连续性作为约束，求解物体的反射率。Zhou 等人<sup>[120]</sup>提出了一种数据驱动的图像本征分解方法，通过在人工标注的反射率数据集上训练一个模型来预测反射率。Zoran 等人<sup>[121]</sup>提出了从灰度图像中分辨光照变化的方法。Shen 和 Yeo<sup>[122]</sup>利用纹理的局部连续性将原始图像分离为阴影和反射率。

在本章中，本文利用第三章提出的数据集中阴影图作为训练样本，将本征分解和材质分割两个问题同时求解。

### 4.2.2 基于 Transformer 的语义分割

材质分割作为与语义分割相接近的研究问题，因为目前相应的数据集匮乏，一种做法是通过特殊的设计来实现特定材质的区分，而另一种做法是将语义分割的方法迁移到材质分割任务中。

近年来，学者将基于 Transformer 的方法应用于语义分割，取得了显著的

成效。Zheng 等人<sup>[123]</sup>将语义分割作为一个序列到序列的预测任务，采用一个纯 Transformer 作为其分割模型的编码器，而不使用任何卷积层，增大了模型的感受野，从全局提取信息，提高了语义分割整体的准确度。Liu 等人<sup>[124]</sup>通过层次特征映射将 Transformer 中计算自注意的复杂度从平方复杂度降低到线性复杂度，并将滑动窗口法改进为移位窗口法，大幅度降低延迟。Strudel 等人<sup>[125]</sup>提出的方法包含一个在 ImageNet 上预训练的 ViT<sup>[126]</sup>(Vision Transformer) 主干网络，并引入一个掩码 Transformer 作为解码器，使用预训练好的模型在中等大小的图像分割数据集上进行微调，提高了训练效率。Wang 等人<sup>[127-128]</sup>通过引入一个渐进收缩的金字塔主干网络来降低计算成本，同时输出粒度更细的分割结果，从而克服了 ViT 的问题。Ranftl 等人<sup>[129]</sup>提出了一种基于 Transformer 的编码器-解码器结构的密集预测模型，能够在 Transformer 架构的所有阶段中保持空间分辨率，为密集预测任务保留了细粒度信息。

与成熟的卷积神经网络不同，Transformer 仍处于发展的早期阶段，且需要比 CNN 更多的训练数据，因此对于材质分割，尤其是高光谱图像的材质分割这种缺少数据集的任务来说，Transformer 的潜力还未完全发挥出来。

为此，本章利用 Transformer 在语义分割中的最新研究，结合第二章相关研究中介绍的用于高光谱图像的 Transformer 中光谱自注意的提取方法，将 RGB 图像与高光谱图像相结合，并使用第二章中得到的光源光谱进行反射率校正，有效提高材质分割的精度。

### 4.3 本文构建的材质分割数据集

为了支撑本文提出的材质分割算法，本文对第三章中提到的 Mobile-Spec 数据集进行了材质分割的标注。Mobile-Spec 数据集中的阴影图仍被保留，用于 RGB 图像的反射率校正。

如图4-1展示了本文构建的数据集中的材质分割图像，其中分割类别被指定为绿植、树干、建筑、道路、天空和其他，这些类别反映了室外场景中最常见的对象。本文的材质分割标注具有极高的细粒度，特别是在绿植区域内单个叶片的细节方面。

值得注意的是，在本数据集的标注中，树冠上的树叶与地面上的草和灌木

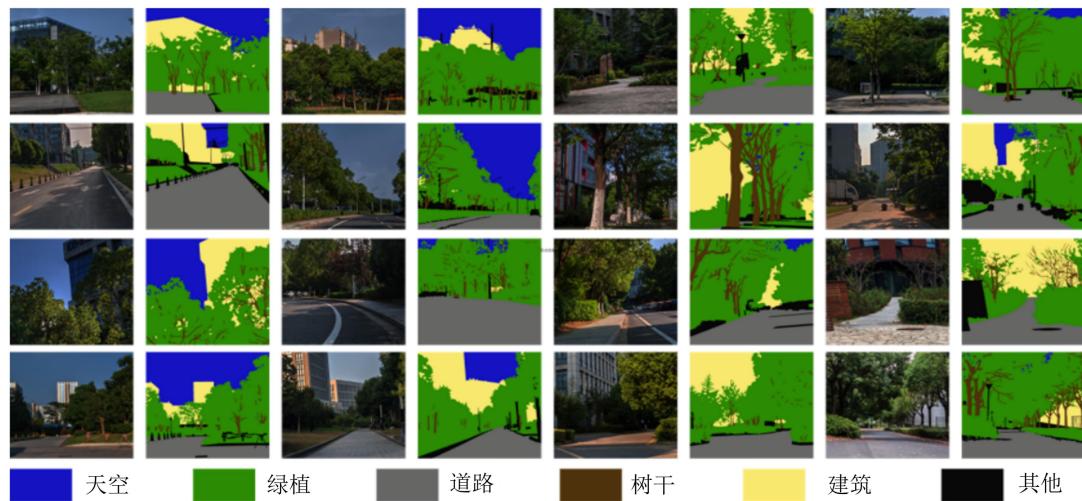


图 4-1 高光谱材质分割数据集

一起被标注为“绿植”，本文认为绿植的光谱特征比较相近，这样的标注方法是区别于语义分割的。不像语义分割中将树干与树叶标注为“树”这一类，本数据集单独对树干进行标注，用于区分其与绿植不同的光谱响应。对于道路和建筑的区分，有些比较矮的围墙与道路相连，本数据集将可供人行走的区域标注为道路，不可供人行走的区域标注为建筑。

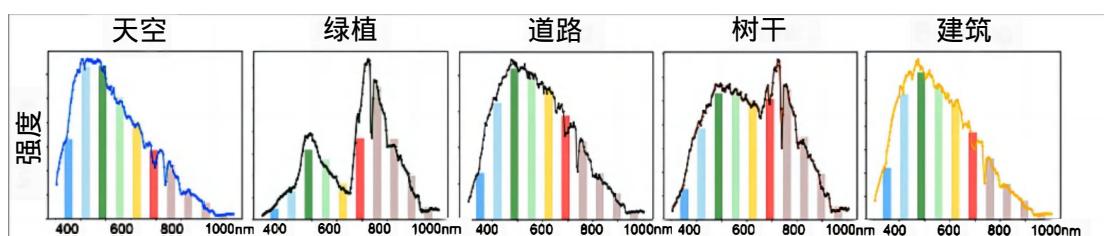


图 4-2 不同材质类别的光谱响应曲线

如图4-2所示，不同的材质类别有着不同的光谱响应，这为材质分割提供了细致的光谱信息依据。其中，蓝色波段(450 - 520 nm)是天空的强反射带；绿色波段(520 - 600 nm)处绿植有一个明显的峰；红色波段(630 - 690 nm)是绿植的主要吸收带；近红外波段(760 - 900 nm)是绿植和树干的强反射带。

本文采用 t-随机邻域嵌入(t-distributed Stochastic Neighbor Embedding, t-SNE)<sup>[130]</sup>来分析复杂的光谱模式和不同光谱波段之间的关系。如图4-3所示，与 RGB 图像相比，代表不同类别的特征在高光谱图像的 t-SNE 可视化中表现出更密集的聚集。这一现象表明，光谱通道数的增加使得对不同材质的辨别变得更加简单，因此，低空间分辨率光谱图像有助于材质分割任务。

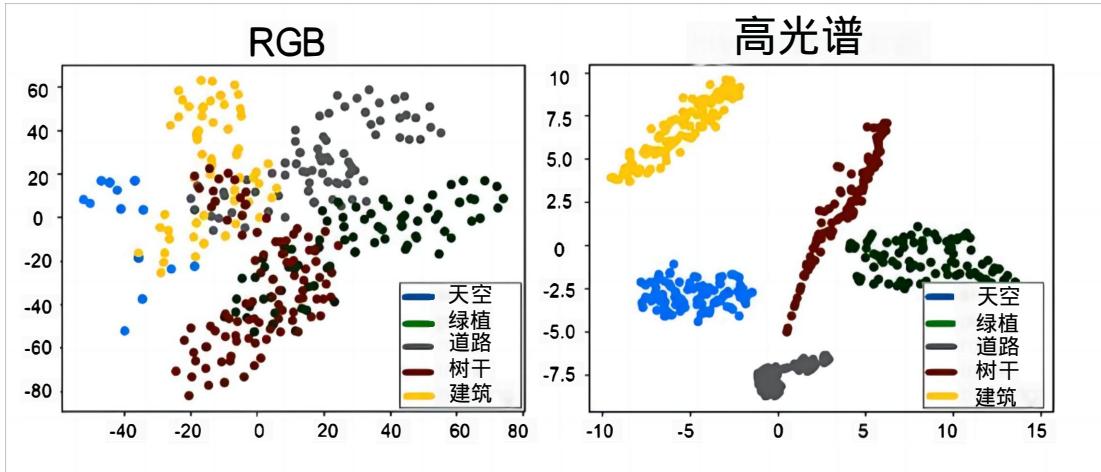


图 4-3 不同材质类别的在 RGB 域和高光谱域的聚类可视化

## 4.4 本文提出的方法

在 RGB 语义分割模型的基础上，本文引入光谱信息作为第二路输入来辅助材质分割。对第二路的光谱信息，首先进行反射率校正的预处理，去除光源和阴影的影响，再使用光谱联合 RGB 分解模型进一步分解 RGB 图像的阴影和反射率，将 RGB 图像的反射率输入材质分割网络，提高材质分割的准确度。

为了将低空间分辨率光谱图像有效地集成到高空间分辨率的 RGB 图像分割的模型中，本文首先分析了高光谱相机和 RGB 相机在成像机制上的差异，高光谱相机成像模型<sup>[131-132]</sup>可以表示如下：

$$I_{k,x} = \int_{400\text{nm}}^{1000\text{nm}} C_k(\lambda) L(\lambda) S(x) R(\lambda, x) d\lambda, k = 1, 2, 3 \dots \quad (4-1)$$

$I_{k,x}$  表示第  $k$  波段位置  $x$  处的像素强度， $C_k(\lambda)$  表示相机响应函数， $L(\lambda)$  表示光源的光谱曲线， $S(x)$  表示阴影， $R(\lambda, x)$  表示场景的反射率。对于 RGB 相机，波段数量仅有 3 个，且波段范围在可见光范围内。RGB 相机与高光谱相机的核心区别在于  $C_k(\lambda)$ ，由于  $C_k(\lambda)$  具有更高的光谱采样率，高光谱图像具有细粒度的颜色通道。

需要注意的是，本文将光源对光谱数据的影响分解为光源光谱曲线  $L(\lambda)$  和阴影  $S(x)$  两项。光源光谱曲线  $L(\lambda)$  通常由光源估计获得，用于自动白平衡，而阴影  $S(x)$  和反射率  $R(\lambda, x)$  两项的作用却很少被研究，本文同时使用光源曲线  $L(\lambda)$  和阴影  $S(x)$  来对高光谱和 RGB 图像进行校正，并将得到的反射率  $R(\lambda, x)$  用于材质分割。

#### 4.4.1 反射率校正预处理

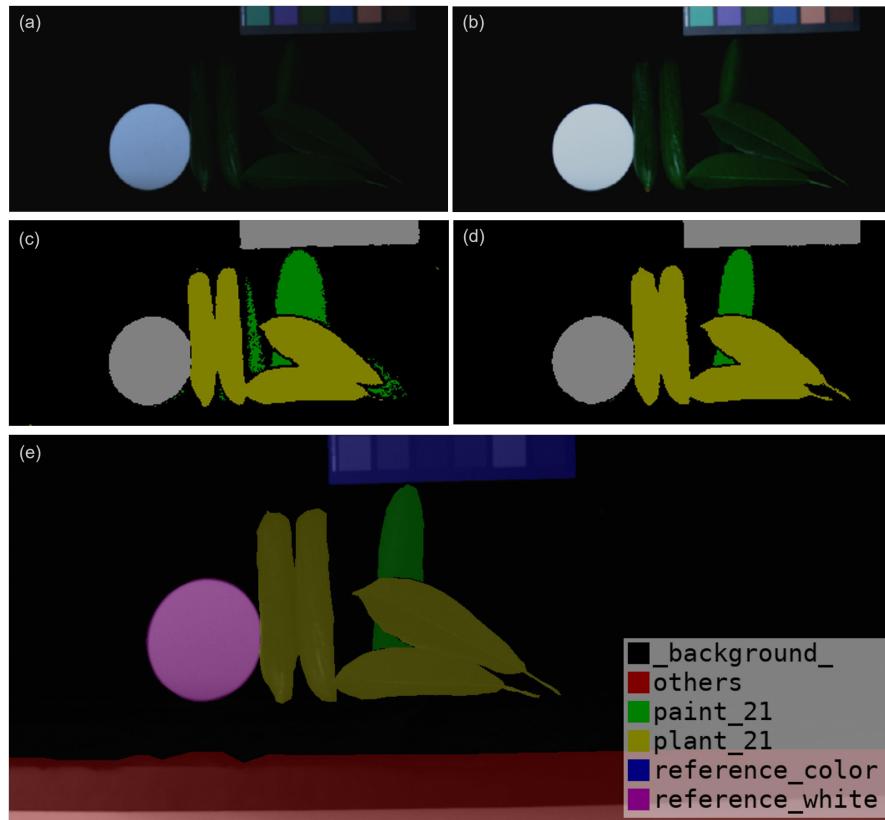


图 4-4 对高光谱图像的无监督分割，初步验证光源校正的作用 (a) 未经校正的图像 (b) 校正后的图像 (c)(d) 分别对应 (a)(b) 的分割结果 (e) 人工标注

光谱反射率是与光源和形状无关的、物质的本征光谱信息，用于检测和分析不同物质间细微的光谱特征差异，为了更好地将高光谱图像用于材质分割，本文对高光谱图像进行反射率校正。反射率校正采用以下公式：

$$R(k, x) = I(k, x) / (L(k) \otimes S(x)). \quad (4-2)$$

其中  $k$  和  $x$  仍然用于表示波段位置和像素位置， $\otimes$  为外积，目的是将每个波段的光源光谱曲线  $L(k)$  用在阴影  $S(x)$  的每个像素上，上式中的除法是矩阵的逐像素除法，从拍摄的高光谱图像的每个波段的每个像素  $I(k, x)$  中提取出反射率  $R(k, x)$ 。光源光谱曲线  $L(k)$  由第二章的光源光谱估计得到，阴影  $S(x)$  由第三章的数据集得到，如果缺少阴影图，则使用全 1 矩阵代替阴影图，默认场景具有全局一致的阴影。

如图4-4，通过对校正前后的高光谱图像无监督分割结果，可以看到，未经校正的高光谱图像中深颜色的黄瓜与深色背景难以区分，导致了未经校正的

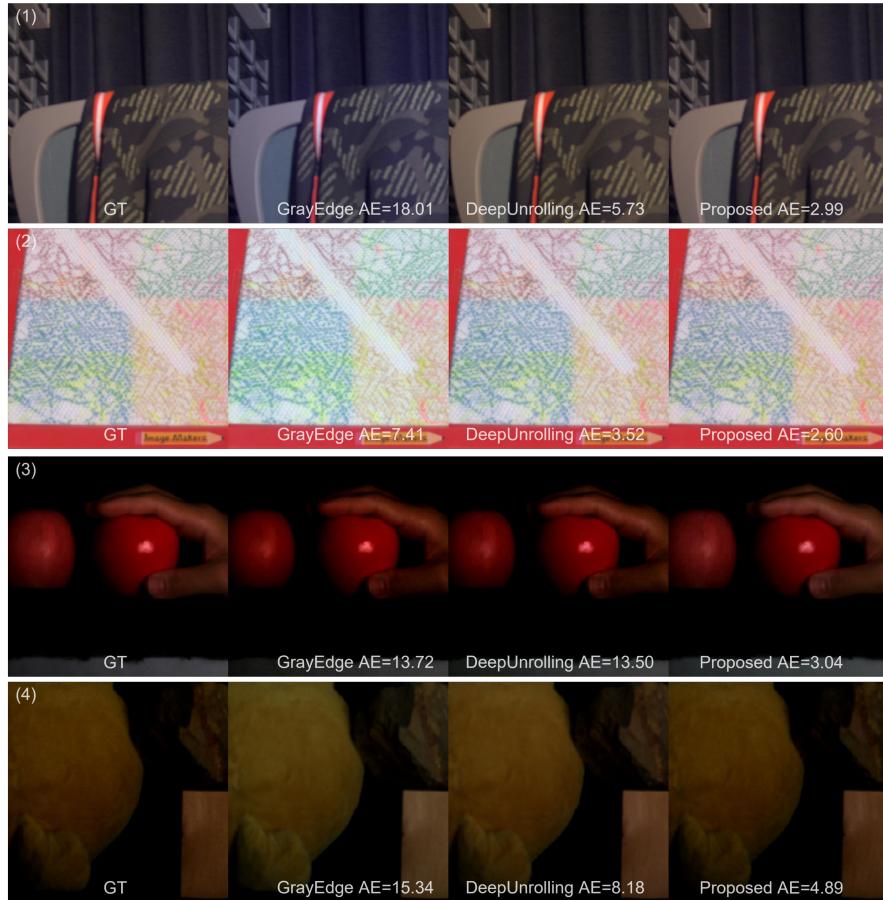


图 4-5 使用不同光源估计算法用于校正的可视化对比 (1)(2) 仿真数据集 (3)(4) 真实数据集

高光谱图像无监督分割的效果不佳；而校正后的高光谱图像提升了同样为深色的物体与背景的对比度，有效地提高了高光谱图像无监督分割的效果，证明了光源校正对于图像分割的有效性。

因此，需要准确的光源校正方法来对材质分割等任务中使用的图像进行反射率校正预处理。如图4-5，展示了经不同的光源估计方法校正后的高光谱图像的RGB渲染图。得益于本文在第二章中提出的更准确的光源估计方法，使用本文提出的光源估计方法获得的光源光谱能够更有效地帮助图像分割。

#### 4.4.2 光谱联合 RGB 分解模型

为了将高光谱图像的反射率校正用在 RGB 图像上从而对 RGB 图像进行更准确的材质分割，本文使用光谱联合 RGB 分解模型，利用低空间分辨率光谱图像中的近红外和可见光波段辅助材质分割，低空间分辨率的光谱图像与高空间分辨率的 RGB 图像正好在光谱和空间两个维度上互补。

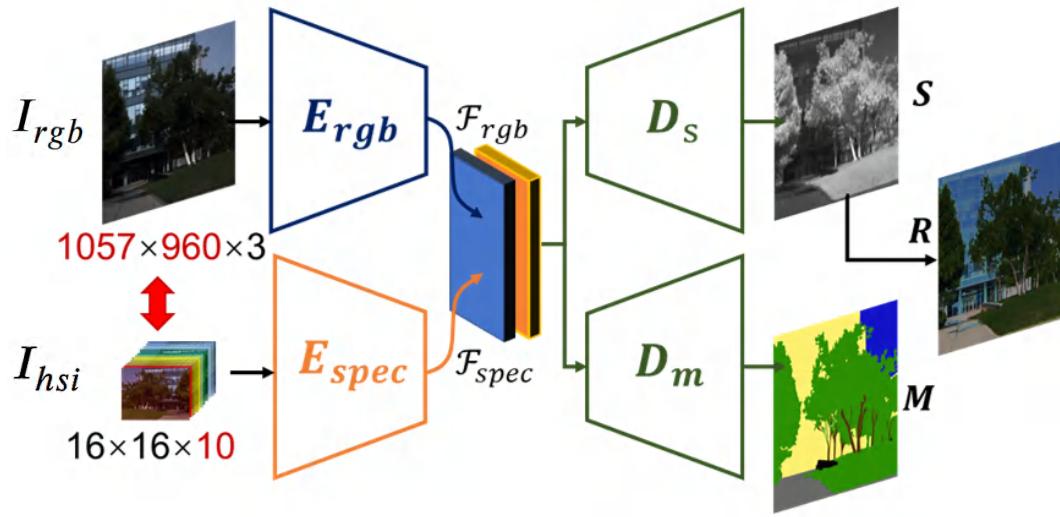


图 4-6 光谱联合 RGB 分解模型示意图

如图4-6所示，光谱联合 RGB 分解模型融合了两个独立的编码器和解码器，输入为高空间分辨率的 RGB 图像和低空间分辨率的光谱图像，输出为阴影图  $S$  和材质分割结果  $M$ ，阴影图  $S$  可与光源光谱结合，进行反射率校正，得到反射率图  $R$ 。

具体来说，将低空间分辨率光谱图像调整为与 RGB 图像相同的空间分辨率，并预训练 RGB 编码器  $E_{rgb}$  和光谱编码器  $E_{spec}$ ，将 RGB 图像  $I_{rgb}$  和低空间分辨率光谱图像  $I_{hsr}$  投影到相同的隐空间进行特征的对齐，得到 RGB 图像的特征  $\mathcal{F}_{rgb}$  和高光谱图像的特征  $\mathcal{F}_{spec}$ ：

$$\begin{aligned}\mathcal{F}_{rgb} &= E_{rgb}(I_{rgb}), \\ \mathcal{F}_{spec} &= E_{spec}(I_{hsr}).\end{aligned}\tag{4-3}$$

然后通过串接操作将  $\mathcal{F}_{rgb}$  和  $\mathcal{F}_{spec}$  融合在一起，以共享互补的特征。

基于融合后的特征，利用解码器  $D_m$  和  $D_s$  独立预测材质分割  $M$  和阴影  $S$ 。光谱联合 RGB 分解模型的解码器可以表示为：

$$M, S = D_{m,s}(\text{concat}(\mathcal{F}_{rgb}, \mathcal{F}_{spec})).\tag{4-4}$$

分解得到的阴影  $S$  可以根据等式4-2反过来对 RGB 图像进行反射率校正，并得到 RGB 图像的反射率图像  $R$ 。

光谱联合 RGB 分解模型在训练时的损失函数包括材质分割相关的损失项和阴影图分解的损失项，表示为如下等式：

$$\ell = \eta \ell_{mIoU}(M, M_{GT}) + (1 - \eta) \ell_{L1}(S, S_{GT}). \quad (4-5)$$

其中  $M$  和  $S$  为预测的材质分割图和阴影图， $M_{GT}$  和  $S_{GT}$  材质分割真值标注和阴影图真值， $\ell_{mIoU}$  为平均交并比损失， $\ell_{L1}$  为 L1 损失， $\eta$  为权重系数。当  $\eta = 1$  时，光谱联合 RGB 分解模型可以看做纯材质分割模型；当  $\eta = 0$  时，光谱联合 RGB 分解模型可以看做本征信息分解模型。

$$\begin{aligned} \text{IoU}_i(M, M_{GT}) &= \frac{|M(i) \cap M_{GT}(i)|}{|M(i) \cup M_{GT}(i)|}, \\ \text{mIoU}(M, M_{GT}) &= \frac{\sum_{i=0}^n \text{IoU}_i(M, M_{GT})}{n + 1}, \\ \ell_{mIoU}(M, M_{GT}) &= 1 - \text{mIoU}(M, M_{GT}). \end{aligned} \quad (4-6)$$

其中  $n + 1$  表示材质分割的总类别数， $\text{IoU}_i$  表示第  $i$  类材质的交并比。

$$\ell_{L1}(S, S_{GT}) = |S - S_{GT}|. \quad (4-7)$$

## 4.5 实验过程与结果分析

在本章研究中，实验设置所使用的硬件平台和相关软件同第二章实验设置部分。

### 4.5.1 评价指标

对于材质分割的结果，本文采用平均交并比 (mean Intersection-over-Union, mIoU) 作为评价指标对材质分割的准确度进行衡量。

如下等式定义了分割图 A 与分割图 B 的平均交并比：

$$\begin{aligned} \text{IoU}_i(A, B) &= \frac{|A(i) \cap B(i)|}{|A(i) \cup B(i)|}, \\ \text{mIoU}(A, B) &= \frac{\sum_{i=0}^n \text{IoU}_i(A, B)}{n + 1}. \end{aligned} \quad (4-8)$$

其中,  $|X|$  为集合  $X$  的大小,  $\cap$  和  $\cup$  分别为集合的交集和并集。将一张图像的每个像素(按照一定顺序)的标签视为一个元素, 则每张图像中所有像素的标签代表一个有序集。预测的材质分割图与标注的材质分割图之间的 mIoU 越大, 材质分割的准确度越高。由此, 预测结果和人工标注可以使用 mIoU 度量进行比较。

#### 4.5.2 定性分析

如图4-7所示, 本文给出了在不同场景下的材质分割预测与人工标注的对比。

在第一行图像中, 得益于阴影图的校正, 阴影下的树干得到了较好的分割。在第二行图像和第五行图像中, 分割失败的树干上同时有颜色的变化(白漆)和亮度的变化, 即使进行了阴影图的校正仍未准确分割。在第二行图像和第三行图像中, 同样得益于阴影图的校正, 地面上的亮度变化并没有影响道路的分割。在第四行图像中, 建筑的镂空部分原本应该是天空类, 但本文提出的算法没能区分玻璃反射的天空与真实的天空, 从场景图中可以看到天空与玻璃间有亮度的差异, 但是本文使用阴影图进行校正后的反射率图(可参考图3-2)中天空和玻璃具有相似的亮度, 仅从颜色上很难区分两者。

定性分析强调了使用阴影图进行反射率校正对材质分割的有效性。其优点在于减少了阴影变化对同一材质分割类别的整体一致性的影响, 缺点在于使得原本需要用亮度进行区分的类别变得更难区分。

#### 4.5.3 定量分析

在实际应用中, 高光谱的波段数量和空间分辨率往往受到限制, 因此有必要考量不同波段数量以及空间分辨率下的光源光谱辅助的效果。更少的波段数量和更低的空间分辨率通过从高空间分辨率的高光谱图像中下采样得到。

如图4-8所示, 随着光谱波段数量的降低, 材质分割结果交并比总体上从 90% 降低到了 80%, 而引入的光源光谱的空间分辨率对材质分割结果的影响并不大。这或许是因为光源光谱是通过间接的方式辅助材质分割, 网络中起主要作用的依然是高分辨率的 RGB 图像, 但是光谱波段数量会影响反射率校正的精度, 从而比较明显地影响 RGB 图像的分解以及材质分割。

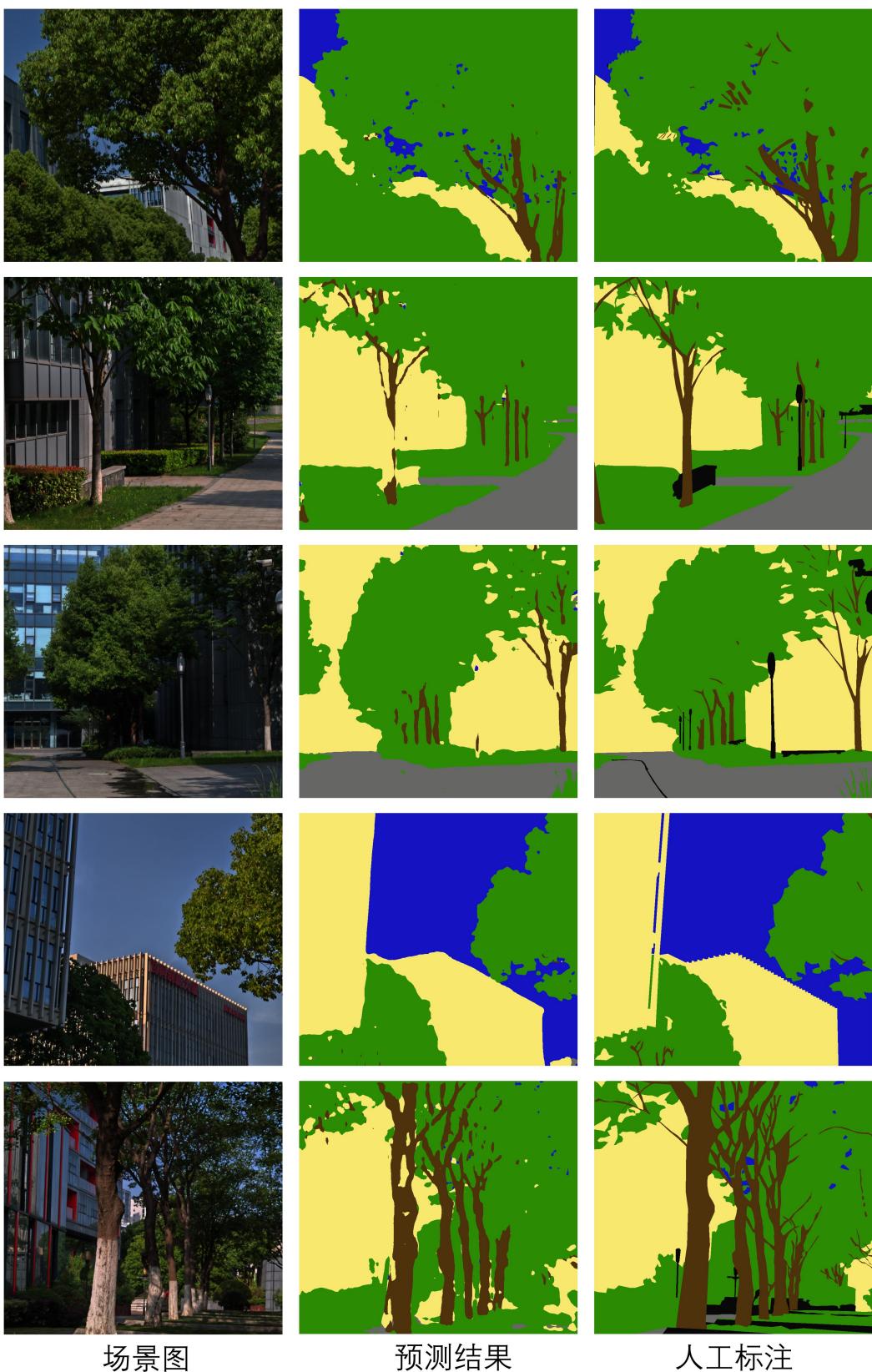


图 4-7 材质分割结果图

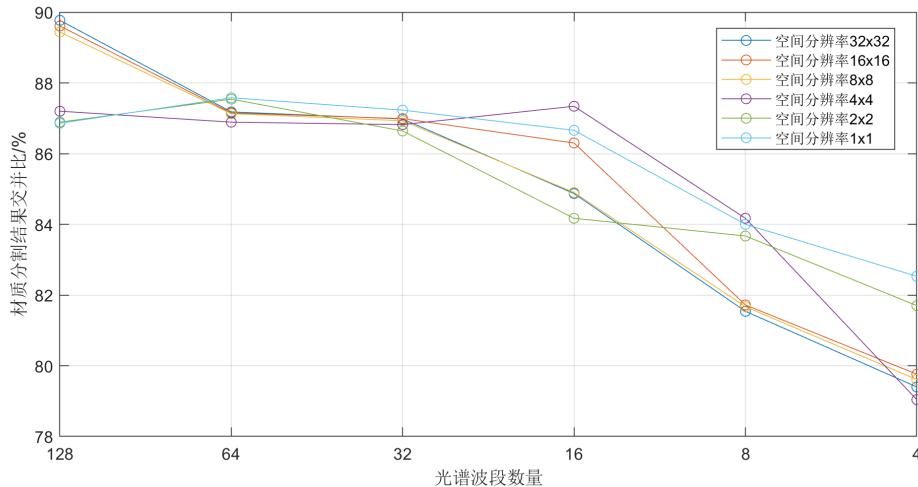


图 4-8 引入光源光谱波段数量以及空间分辨率对材质分割结果的影响

由此可知，在该任务中，光谱波段数量的作用大于光谱空间分辨率的作用，这或许对于实际应用中光谱传感器的设计有一定的引导作用。

表 4-1 引入不同程度的光谱辅助下的材质分割交并比(%)对比

通道选择	无	3 通道 RGB			16 通道光谱			128 通道光谱			
		1×1	4×4	16×16	1×1	4×4	16×16	1×1	4×4	16×16	
空间分辨率	无	95.21	94.69	94.69	94.8	95.14	96.04	95.82	95.59	95.27	<b>96.28</b>
天空	95.21	94.69	94.69	94.8	95.14	96.04	95.82	95.59	95.27	<b>96.28</b>	
建筑	71.5	73.8	74.52	<b>78.58</b>	78.46	76.92	77.34	77.79	77.91	76.79	
道路	53.16	67.41	78.39	76.15	77.86	84.69	82.09	75.1	80.04	<b>86.71</b>	
绿植	88.14	90.98	93.11	92.86	92.81	94.06	93.87	92	93.62	<b>94.51</b>	
树干	90.94	92.38	90.9	91.17	93.11	92.46	92.92	92.71	<b>93.11</b>	92.2	
平均交并比	79.79	83.85	86.32	86.71	87.47	88.83	88.40	86.63	87.99	<b>89.29</b>	

此外，不同类别的分割精度也呈现出很大差异，如表4-1所示，天空类别无需光谱辅助就能达到 95% 左右的材质分割交并比；对于建筑类别，16×16 的空间分辨率的额外辅助的情况下，3 通道 RGB 辅助的效果甚至可能优于更多通道的光谱辅助的结果，分别为 78.58% 和 76.79%；光谱辅助对道路类别的影响最大，在无辅助的情况下为 53.16%，在 16×16×128 的光谱的辅助下达到了 86.71%；对于绿植和树干类，引入额外的光谱辅助则有轻微的提升。

对于比较复杂的类别，例如建筑，光谱的引入几乎不起作用。这是因为建筑的复杂材质，如玻璃、砖块以及粉刷墙面的光谱反射率差别过大，反而在分割时起到反效果。而对于道路这类在 RGB 图像中色彩各异、在特定波段的光谱反射率却比较稳定的类别，则尤其适合用光谱辅助。

## 4.6 本章小结

诸如材质分割等高层级的任务会受图像处理等低层级的任务的影响，在本章中，本文利用材质分割这个任务证明了光源校正的必要性和有效性，相较于未引入光源光谱校正的结果取得了明显的进步。这主要归功于本章方法在以下几个方面的研究：

1. 通过将光源光谱作为材质分割网络的额外输入，提出了光谱联合 RGB 分解模型来利用光谱辅助 RGB 图像进行材质分割和阴影图预测，并推理反射率图。
2. 利用光源光谱估计的结果来进行反射率校正，区分同样深色物体的反射率，从数据输入层面增加不同材质类别间的特征差异。

实验结果表明，更多的光源光谱波段、更高的空间分辨率的光谱辅助更有利 RGB 的材质分割。本方法的不足之处在于目前尚无法完美解决“同物易谱”、“异物同谱”的问题，在反射率校正的精度方面仍有提升空间。



## 第五章 总结与展望

### 5.1 总结

本文围绕高光谱图像的光源光谱估计问题及其应用展开研究。总体可以分为三个部分：其一是高光谱图像的光源光谱估计任务本身，对给定的高光谱图像预测其光源光谱；其二是光源光谱引导的图像增强，使用光源光谱引导高动态范围 RGB 图像的增强；其三是光源光谱校正的材质分割，使用光源光谱进行反射率校正并辅助 RGB 图像的材质分割。具体工作如下：

一、归纳并探讨现有的光源估计算法，分析了基于统计和基于学习的光源估计算法的优劣，在此基础上，本文构建了真实而非仿真的高光谱光源估计数据集并对算法进行了改进。本文提出了光谱联合空间自注意机制，在增强了特征图对特定通道和位置信息的感知的同时，有效缓解了基于学习的光源光谱估计任务中高光谱数据计算量大、网络参数多、训练慢的问题。此外，将损失函数改进为平滑约束的损失函数，有效避免了不实际的病态结果，从而提升了光源估计的准确度。

二、在预测出高光谱图像的光源光谱的基础上，本文探索其对 RGB 图像增强的作用。构建了高动态范围 RGB 图像与高光谱图像对齐的图像增强数据集。在此基础上，提出了局部亮度适应法，在室外太阳光下使用近红外光源光谱修正 RGB 图像的局部亮度，为 16 位源图动态范围压缩到 8 位目标图提供了有效的预处理。另外，将低空间分辨率的光谱图像作为 RGB 图像增强的第二路输入，提出了光谱感知自注意机制，将光谱特征嵌入到 RGB 特征图中，从而提升了 RGB 图像增强的效果。

三、在将光源光谱应用于低层级任务的基础上，本文进一步探索其在高层级任务的作用。本文提出的高光谱材质分割数据集可以为进一步的材质分割提供精确的反射率。基于本文上述的光源光谱估计算法，对用于材质分割的图像进行反射率校正预处理，在此基础上，将低空间分辨率的光谱反射率图像与高

空间分辨率的 RGB 图像使用光谱联合 RGB 分解模型联合分解，提高了 RGB 图像的材质分割精度。

## 5.2 展望

目前，高光谱成像在民用和工业界应用还不够广泛，技术尚不够成熟。现有的高光谱成像系统还难以嵌入到小型的、轻便的或手持的移动端设备中，且高维度数据的计算受制于有限的算力。在未来，高维视觉信息的感知与计算会越来越普遍，相应的算法也会与时俱进。同时，因个人的眼界、学识和研究时间等方面的限制，本论文尚存在一些不足之处，有待于在今后研究工作中进一步完善。下面对本文涉及的研究方向进行展望：

一、在光源校正流程高度自动化，高光谱采集设备的使用普及化以及高维数据的传输便捷化后，通用的大规模高光谱图像数据集将会逐渐被构建，并应用于各种高光谱图像相关任务的预训练。

二、本文用于 RGB 图像增强的低空间分辨率光谱图像的空间分辨率建议至少为  $16 \times 16$ ，这在目前的智能手机上尚未实现。随着未来光谱成像元件与系统的小型化、集成化的发展，有望在小型的、轻便的或手持的移动端实现高分辨率的光谱成像，从而将实验中的更高分辨率下的结果转化为现实应用。

三、光源光谱校正的材质分割在未来可能应用于自动驾驶，考虑到驾驶过程中环境光的差异，仿照人眼的自适应性，合适的光源校正或许更有利于视觉模型的认知。

## 参考文献

- [1] GIJSENIJ A, GEVERS T, VAN DE WEIJER J. Computational color constancy: Survey and experiments[J]. IEEE transactions on image processing, 2011, 20(9): 2475-2489.
- [2] VAN DE WEIJER J, GEVERS T, GIJSENIJ A. Edge-based color constancy[J]. Ieee Transactions on Image Processing, 2007, 16(9): 2207-2214.
- [3] JOZE H R V, DREW M S. Exemplar-Based Color Constancy and Multiple Illumination[J]. Ieee Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(5): 860-873.
- [4] HU Y, WANG B, LIN S, et al. FC4: Fully Convolutional Color Constancy with Confidence-weighted Pooling[C] // IEEE Conference on Computer Vision and Pattern Recognition: 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 330-339.
- [5] GIJSENIJ A, LU R, GEVERS T. Color Constancy for Multiple Light Sources [J]. Ieee Transactions on Image Processing, 2012, 21(2): 697-707.
- [6] GAO S B, REN Y Z, ZHANG M, et al. Combining Bottom-Up and Top-Down Visual Mechanisms for Color Constancy Under Varying Illumination[J]. Ieee Transactions on Image Processing, 2019, 28(9): 4387-4400.
- [7] TEKE M, DEVECI H S, HALILOĞLU O, et al. A short survey of hyperspectral remote sensing applications in agriculture[C] // 2013 6th international conference on recent advances in space technologies (RAST). 2013: 171-176.
- [8] LU G, FEI B. Medical hyperspectral imaging: a review[J]. Journal of biomedical optics, 2014, 19(1): 010901-010901.
- [9] ADÃO T, HRUŠKA J, PÁDUA L, et al. Hyperspectral imaging: A review

- on UAV-based sensors, data processing and applications for agriculture and forestry[J]. Remote sensing, 2017, 9(11): 1110.
- [10] KHAN M J, KHAN H S, YOUSAF A, et al. Modern trends in hyperspectral image analysis: A review[J]. Ieee Access, 2018, 6: 14118-14129.
- [11] LU B, DAO P D, LIU J, et al. Recent advances of hyperspectral imaging technology and applications in agriculture[J]. Remote Sensing, 2020, 12(16): 2659.
- [12] 李树涛, 吴琼, 康旭东. 高光谱遥感图像本征信息分解前沿与挑战[J]. 测绘学报, 2023, 52(7): 1059.
- [13] 鲁洋, 徐海松. 基于单幅多光谱图像的照明光谱估计方法[J]. 光学学报, 2022, 42(7): 274-282.
- [14] GAO G, LIU B, ZHANG X, et al. Multitemporal intrinsic image decomposition with temporal–spatial energy constraints for remote sensing image analysis[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-16.
- [15] DUAN P, LAI J, GHAMISI P, et al. Component decomposition-based hyperspectral resolution enhancement for mineral mapping[J]. Remote Sensing, 2020, 12(18): 2903.
- [16] BUCHSBAUM G. A SPATIAL PROCESSOR MODEL FOR OBJECT COLOR-PERCEPTION[J]. Journal of the Franklin Institute-Engineering and Applied Mathematics, 1980, 310(1): 1-26.
- [17] LAND E H, MCCANN J J. LIGHTNESS AND RETINEX THEORY[J]. Journal of the Optical Society of America, 1971, 61(1).
- [18] KHAN H A, THOMAS J B, HARDEBERG J Y, et al. Illuminant estimation in multispectral imaging[J]. Journal of the Optical Society of America a-Optics Image Science and Vision, 2017, 34(7): 1085-1098.
- [19] KHAN H A, THOMAS J B, HARDEBERG J Y, et al. Spectral Adaptation Transform for Multispectral Constancy[J]. Journal of Imaging Science and Technology, 2018, 62(2).

- [20] KHAN H A, THOMAS J B, HARDEBERG J Y. Towards Highlight Based Illuminant Estimation in Multispectral Images[C]//Lecture Notes in Computer Science: 8th International Conference on Image and Signal Processing (ICISP): vol. 10884. 2018: 517-525.
- [21] SU T, ZHOU Y, YU Y, et al. Illumination separation of non-Lambertian scenes from a single hyperspectral image[J]. Optics Express, 2018, 26(20): 26167-26178.
- [22] ZHENG Y, SATO I, SATO Y, et al. Illumination and Reflectance Spectra Separation of a Hyperspectral Image Meets Low-Rank Matrix Factorization[C]//IEEE Conference on Computer Vision and Pattern Recognition: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015: 1779-1787.
- [23] SHI W, LOY C C, TANG X. Deep Specialized Network for Illuminant Estimation[C]//Lecture Notes in Computer Science: 14th European Conference on Computer Vision (ECCV): vol. 9908. 2016: 371-387.
- [24] AFIFI M, BROWN M S, Ieee. Deep White-Balance Editing[C]//IEEE Conference on Computer Vision and Pattern Recognition: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020: 1394-1403.
- [25] XU B, LIU J, HOU X, et al. End-to-End Illuminant Estimation based on Deep Metric Learning[C]//IEEE Conference on Computer Vision and Pattern Recognition: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020: 3613-3622.
- [26] ROBLES-KELLY A, WEI R, Ieee. A Convolutional Neural Network for Pixelwise Illuminant Recovery in Colour and Spectral Images[C]//International Conference on Pattern Recognition: 24th International Conference on Pattern Recognition (ICPR). 2018: 109-114.
- [27] LI Y, FU Q, HEIDRICH W. Multispectral illumination estimation using deep unrolling network[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 2672-2681.

- [28] LIU X, PEDERSEN M, WANG R. Survey of natural image enhancement techniques: Classification, evaluation, challenges, and perspectives[J]. *Digital Signal Processing*, 2022, 127: 103547.
- [29] MAINI R, AGGARWAL H. A comprehensive review of image enhancement techniques[J]. arXiv preprint arXiv:1003.4053, 2010.
- [30] DILEEP M, MURTHY A S. A comparison between different colour image contrast enhancement algorithms[C] // 2011 International Conference on Emerging Trends in Electrical and Computer Technology. 2011: 708-712.
- [31] RIZZI A, GATTA C, MARINI D. A new algorithm for unsupervised global and local color correction[J]. *Pattern Recognition Letters*, 2003, 24(11): 1663-1677.
- [32] SHI Z, FENG Y, ZHAO M, et al. Normalised gamma transformation-based contrast-limited adaptive histogram equalisation with colour correction for sand-dust image enhancement[J]. *IET Image Processing*, 2020, 14(4): 747-756.
- [33] AFIFI M, BROWN M S. Interactive white balancing for camera-rendered images[J]. arXiv preprint arXiv:2009.12632, 2020.
- [34] TAI S C, LIAO T W, CHANG Y Y, et al. Automatic white balance algorithm through the average equalization and threshold[C] // 2012 8th international conference on information science and digital content technology (ICIDT2012): vol. 3. 2012: 571-576.
- [35] LIANG Y, WAKAKI R, NOBUHARA S, et al. Multimodal material segmentation[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 19800-19808.
- [36] DANA K J, VAN GINNEKEN B, NAYAR S K, et al. Reflectance and texture of real-world surfaces[J]. *ACM Transactions On Graphics (TOG)*, 1999, 18(1): 1-34.
- [37] BELL S, UPCHURCH P, SNAVELY N, et al. Material recognition in the wild with the materials in context database[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3479-3487.

- [38] SCHWARTZ G, NISHINO K. Recognizing material properties from images[J]. IEEE transactions on pattern analysis and machine intelligence, 2019, 42(8): 1981-1995.
- [39] EVERINGHAM M, VAN GOOL L, WILLIAMS C K, et al. The pascal visual object classes (voc) challenge[J]. International journal of computer vision, 2010, 88: 303-338.
- [40] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: Common objects in context[C]//Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. 2014: 740-755.
- [41] RUSSAKOVSKY O, DENG J, SU H, et al. Imagenet large scale visual recognition challenge[J]. International journal of computer vision, 2015, 115: 211-252.
- [42] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.
- [43] CHEN J, WANG X, GUO Z, et al. Dynamic region-aware convolution[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 8064-8073.
- [44] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 40(4): 834-848.
- [45] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 801-818.
- [46] 曹汛, 周凯来, 戴琼海. 计算光谱成像的前沿进展[J]. 中国计算机学会通讯, 2020, 16(9): 11-16.
- [47] ALVAREZ-CORTES S, KUNKEL T, MASIA B. Practical Low-Cost Recovery

- of Spectral Power Distributions[C]//Computer Graphics Forum: vol. 35: 1. 2016: 166-178.
- [48] CAI Y, LIN J, HU X, et al. Mask-guided spectral-wise transformer for efficient hyperspectral image reconstruction[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 17502-17511.
- [49] ROY S K, DERIA A, SHAH C, et al. Spectral-spatial morphological attention transformer for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-15.
- [50] LI M, LIU J, FU Y, et al. Spectral enhanced rectangle transformer for hyperspectral image denoising[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 5805-5814.
- [51] CARDEI V C, FUNT B, BARNARD K. Estimating the scene illumination chromaticity by using a neural network[J]. JOSA a, 2002, 19(12): 2374-2386.
- [52] WANG N, XU D, LI B. Edge-based color constancy via support vector regression[J]. IEICE transactions on information and systems, 2009, 92(11): 2279-2282.
- [53] BIANCO S, CUSANO C, SCHETTINI R. Color constancy using CNNs[C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2015: 81-89.
- [54] BARRON J T. Convolutional color constancy[C]//Proceedings of the IEEE International Conference on Computer Vision. 2015: 379-387.
- [55] LOU Z, GEVERS T, HU N, et al. Color Constancy by Deep Learning.[C]//BMVC. 2015: 76-1.
- [56] SHI W, LOY C C, TANG X. Deep specialized network for illuminant estimation [C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14. 2016: 371-387.
- [57] OH S W, KIM S J. Approaching the computational color constancy as a classi-

- fication problem through deep learning[J]. Pattern Recognition, 2017, 61: 405-416.
- [58] VASWANI A, SHAZER N, PARMAR N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.
- [59] HONG D, HAN Z, YAO J, et al. SpectralFormer: Rethinking hyperspectral image classification with transformers[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-15.
- [60] JIA S, WANG Y. Multiscale convolutional transformer with center mask pre-training for hyperspectral image classification[J]. arXiv preprint arXiv:2203.04771, 2022.
- [61] YANG X, CAO W, LU Y, et al. Hyperspectral image transformer classification networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-15.
- [62] LI S, GUO Q, LI A. Pan-sharpening based on CNN+ pyramid transformer by using no-reference loss[J]. Remote Sensing, 2022, 14(3): 624.
- [63] LIU B, YU A, GAO K, et al. DSS-TRM: Deep spatial-spectral transformer for hyperspectral image classification[J]. European Journal of Remote Sensing, 2022, 55(1): 103-114.
- [64] ZHONG Z, LI Y, MA L, et al. Spectral-spatial transformer network for hyperspectral image classification: A factorized architecture search framework[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-15.
- [65] HARDEBERG J Y, SCHMITT F, BRETTEL H. Multispectral color image capture using a liquid crystal tunable filter[J]. Optical engineering, 2002, 41(10): 2532-2548.
- [66] BARNARD K, MARTIN L, FUNT B, et al. A data set for color research[J]. Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour

- Centre Foundation, Colour Society of Australia, Centre Français de la Couleur, 2002, 27(3): 147-151.
- [67] YASUMA F, MITSUNAGA T, ISO D, et al. Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum[J]. IEEE transactions on image processing, 2010, 19(9): 2241-2253.
- [68] CHAKRABARTI A, ZICKLER T. Statistics of Real-World Hyperspectral Images[C]// Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2011: 193-200.
- [69] NGUYEN R M, PRASAD D K, BROWN M S. Training-based spectral reconstruction from a single RGB image[C]// Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part VII 13. 2014: 186-201.
- [70] NASCIMENTO S M, AMANO K, FOSTER D H. Spatial distributions of local illumination color in natural scenes[J]. Vision research, 2016, 120: 39-44.
- [71] CHOI I, KIM M, GUTIERREZ D, et al. High-quality hyperspectral reconstruction using a spectral prior[R]. 2017.
- [72] ARAD B, TIMOFTE R, YAHEL R, et al. Ntire 2022 spectral recovery challenge and data set[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 863-881.
- [73] KARAIMER H C, BROWN M S. A software platform for manipulating the camera imaging pipeline[C]// Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. 2016: 429-444.
- [74] YUAN L, SUN J. Automatic exposure correction of consumer photographs[C]// Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7–13, 2012, Proceedings, Part IV 12. 2012: 771-785.
- [75] GIJSENIJ A, GEVERS T, VAN DE WEIJER J. Computational color con-

- stancy: Survey and experiments[J]. IEEE transactions on image processing, 2011, 20(9): 2475-2489.
- [76] CAI J, GU S, ZHANG L. Learning a deep single image contrast enhancer from multi-exposure images[J]. IEEE Transactions on Image Processing, 2018, 27(4): 2049-2062.
- [77] MANTIUK R, DALY S, KEROFSKY L. Display adaptive tone mapping[G] // ACM SIGGRAPH 2008 papers. 2008: 1-10.
- [78] RAMANATH R, SNYDER W E, YOO Y, et al. Color image processing pipeline[J]. IEEE Signal processing magazine, 2005, 22(1): 34-43.
- [79] COMMISSION I E, et al. Multimedia systems and equipment-Color measurement and management-Part 2-1[J]. Color management-Default RGB color space-sRGB, 1999.
- [80] HAN X, KHAN I R, RAHARDJA S. High dynamic range image tone mapping: Literature review and performance benchmark[J]. Digital Signal Processing, 2023: 104015.
- [81] 朱仲杰, 崔伟锋, 白永强, 等. 宏微观信息增强与色彩校正的高效色调映射[J]. 中国图象图形学报, 2023, 28(9): 2833-2843.
- [82] KIM Y T. Contrast enhancement using brightness preserving bi-histogram equalization[J]. IEEE transactions on Consumer Electronics, 1997, 43(1): 1-8.
- [83] GHARBI M, CHEN J, BARRON J T, et al. Deep bilateral learning for real-time image enhancement[J]. ACM Transactions on Graphics (TOG), 2017, 36(4): 1-12.
- [84] WANG R, ZHANG Q, FU C W, et al. Underexposed photo enhancement using deep illumination estimation[C] // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 6849-6857.
- [85] HE J, LIU Y, QIAO Y, et al. Conditional sequential modulation for efficient global image retouching[C] // Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16.

2020: 679-695.

- [86] YANG C, JIN M, XU Y, et al. Seplut: Separable image-adaptive lookup tables for real-time image enhancement[C] // European Conference on Computer Vision. 2022: 201-217.
- [87] ZENG H, CAI J, LI L, et al. Learning image-adaptive 3d lookup tables for high performance photo enhancement in real-time[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 44(4): 2058-2073.
- [88] ZHANG F, ZENG H, ZHANG T, et al. Clut-net: Learning adaptively compressed representations of 3dluts for lightweight image enhancement[C] // Proceedings of the 30th ACM International Conference on Multimedia. 2022: 6493-6501.
- [89] DRAGO F, MYSZKOWSKI K, ANNEN T, et al. Adaptive logarithmic mapping for displaying high contrast scenes[C] // Computer graphics forum: vol. 22: 3. 2003: 419-426.
- [90] KIM M H, KAUTZ J. Consistent tone reproduction[C] // The Tenth IASTED International Conference on Computer Graphics and Imaging (CGIM 2008). 2008: 152-159.
- [91] LEE D H, FAN M, KIM S W, et al. High dynamic range image tone mapping based on asymmetric model of retinal adaptation[J]. Signal Processing: Image Communication, 2018, 68: 120-128.
- [92] DUAN J, BRESSAN M, DANCE C, et al. Tone-mapping high dynamic range images by novel histogram adjustment[J]. Pattern Recognition, 2010, 43(5): 1847-1862.
- [93] KHAN I R, RAHARDJA S, KHAN M M, et al. A tone-mapping technique based on histogram using a sensitivity model of the human visual system[J]. IEEE Transactions on Industrial Electronics, 2017, 65(4): 3469-3479.
- [94] LEE J W, PARK R H, CHANG S. Local tone mapping using the K-means algorithm and automatic gamma setting[J]. IEEE Transactions on Consumer

- Electronics, 2011, 57(1): 209-217.
- [95] OSKARSSON M. Temporally consistent tone mapping of images and video using optimal k-means clustering[J]. Journal of mathematical imaging and vision, 2017, 57: 225-238.
- [96] YAO J, CAO X, ZHAO Q, et al. Robust subspace clustering via penalized mixture of Gaussians[J]. Neurocomputing, 2018, 278: 4-11.
- [97] RANA A, SINGH P, VALENZISE G, et al. Deep tone mapping operator for high dynamic range images[J]. IEEE Transactions on Image Processing, 2019, 29: 1285-1298.
- [98] CAO X, LAI K, YANUSHKEVICH S N, et al. Adversarial and adaptive tone mapping operator for high dynamic range images[C] / / 2020 IEEE Symposium Series on Computational Intelligence (SSCI). 2020: 1814-1821.
- [99] PANETTA K, KEZEBOU L, OLUDARE V, et al. Tmo-net: A parameter-free tone mapping operator using generative adversarial network, and performance benchmarking on large scale hdr dataset[J]. IEEE Access, 2021, 9: 39500-39517.
- [100] GOSWAMI A, PETROVICH M, HAUSER W, et al. Tone mapping operators: Progressing towards semantic-awareness[C] / / 2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). 2020: 1-6.
- [101] ZHANG N, ZHAO Y, WANG C, et al. A real-time semi-supervised deep tone mapping network[J]. IEEE Transactions on Multimedia, 2021, 24: 2815-2827.
- [102] GUO C, JIANG X. Deep tone-mapping operator using image quality assessment inspired semi-supervised learning[J]. IEEE Access, 2021, 9: 73873-73889.
- [103] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International journal of computer vision, 2004, 60: 91-110.
- [104] CHENG Z, ZHENG Y, YOU S, et al. Non-local intrinsic decomposition with near-infrared priors[C] / / Proceedings of the IEEE/CVF international confer-

- ence on computer vision. 2019: 2521-2530.
- [105] BARROW H, TENENBAUM J, HANSON A, et al. Recovering intrinsic scene characteristics[J]. Comput. vis. syst, 1978, 2(3-26): 2.
- [106] LAND E H, MCCANN J J. Lightness and retinex theory[J]. Josa, 1971, 61(1): 1-11.
- [107] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE transactions on image processing, 2004, 13(4): 600-612.
- [108] ROY S K, KRISHNA G, DUBEY S R, et al. HybridSN: Exploring 3-D–2-D CNN feature hierarchy for hyperspectral image classification[J]. IEEE Geoscience and Remote Sensing Letters, 2019, 17(2): 277-281.
- [109] ROY S K, MANNA S, SONG T, et al. Attention-based adaptive spectral–spatial kernel ResNet for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(9): 7831-7843.
- [110] DAI M, WARD W O, MEYERS G, et al. Residential building facade segmentation in the urban environment[J]. Building and Environment, 2021, 199: 107921.
- [111] AHMAD M, SHABBIR S, ROY S K, et al. Hyperspectral image classification—Traditional to deep models: A survey for future prospects[J]. IEEE journal of selected topics in applied earth observations and remote sensing, 2021, 15: 968-999.
- [112] SU H, WU Z, ZHANG H, et al. Hyperspectral anomaly detection: A survey[J]. IEEE Geoscience and Remote Sensing Magazine, 2021, 10(1): 64-90.
- [113] HUANG Y, SHEN Q, FU Y, et al. Weakly-supervised semantic segmentation in cityscape via hyperspectral image[C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 1117-1126.
- [114] BARRON J T, MALIK J. Shape, albedo, and illumination from a single image of an unknown object[C] // 2012 IEEE Conference on Computer Vision and

- Pattern Recognition. 2012; 334-341.
- [115] BARRON J T, MALIK J. Intrinsic scene properties from a single rgb-d image [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013: 17-24.
- [116] BARRON J T, MALIK J. Shape, illumination, and reflectance from shading[J]. IEEE transactions on pattern analysis and machine intelligence, 2014, 37(8): 1670-1687.
- [117] CHEN Q, KOLTUN V. A simple model for intrinsic image decomposition with depth cues[C] // Proceedings of the IEEE international conference on computer vision. 2013: 241-248.
- [118] JEON J, CHO S, TONG X, et al. Intrinsic image decomposition using structure-texture separation and surface normals[C] // Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part VII 13. 2014: 218-233.
- [119] SHEN J, YANG X, LI X, et al. Intrinsic image decomposition using optimization and user scribbles[J]. IEEE transactions on cybernetics, 2013, 43(2): 425-436.
- [120] ZHOU T, KRAHENBUHL P, EFROS A A. Learning data-driven reflectance priors for intrinsic image decomposition[C] // Proceedings of the IEEE international conference on computer vision. 2015: 3469-3477.
- [121] ZORAN D, ISOLA P, KRISHNAN D, et al. Learning ordinal relationships for mid-level vision[C] // Proceedings of the IEEE international conference on computer vision. 2015: 388-396.
- [122] SHEN L, YEO C. Intrinsic images decomposition using a local and global sparse representation of reflectance[C] // CVPR 2011. 2011: 697-704.
- [123] ZHENG S, LU J, ZHAO H, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers[C] // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 6881-

6890.

- [124] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 10012-10022.
- [125] STRUDEL R, GARCIA R, LAPTEV I, et al. Segmenter: Transformer for semantic segmentation[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 7262-7272.
- [126] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.
- [127] WANG W, XIE E, LI X, et al. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 568-578.
- [128] WANG W, XIE E, LI X, et al. Pvt v2: Improved baselines with pyramid vision transformer[J]. Computational Visual Media, 2022, 8(3): 415-424.
- [129] RANFTL R, BOCHKOVSKIY A, KOLTUN V. Vision transformers for dense prediction[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 12179-12188.
- [130] VAN DER MAATEN L, HINTON G. Visualizing data using t-SNE.[J]. Journal of machine learning research, 2008, 9(11).
- [131] CHEN X, ZHU W, ZHAO Y, et al. Intrinsic decomposition from a single spectral image[J]. Applied optics, 2017, 56(20): 5676-5684.
- [132] JIANG J, LIU D, GU J, et al. What is the space of spectral sensitivity functions for digital color cameras?[C]//2013 IEEE Workshop on Applications of Computer Vision (WACV). 2013: 168-179.

## 致 谢

回首三年的硕士研究生生涯，非常感谢在我身边默默关注、支持我的老师、同学和父母，优秀的你们给我做出了最好的榜样，激励着我朝更高、更远的目标前进。

感谢我的导师曹汛教授，曹老师学识渊博、工作严谨，是我引以为豪的导师和榜样。曹老师不但指导和帮助我进行科学研究、论文写作，而且在平时的学习和生活中对我悉心教导，关怀有加。曹老师给予了我很多学术指导以及丰富的科研资源，带领我们一众研究生参加杭州、天津等地的学术会议，开阔了我的学术视野，坚定了我继续从事学术研究的信心。衷心感谢曹汛老师三年来对我的关心和指导！

感谢沈秋副教授，沈老师拓宽了我的研究课题，在我的科研过程中给出了许多宝贵的意见和建议。沈老师在帮助我修改专利和论文的过程中，表现出敏锐的科研洞察力和精益求精的学术精神，让我对科研有了更深的理解。

感谢南京大学计算成像实验室 (CITE Lab) 的师兄：陈林森、字崇德、邓智威、黄尔齐、李昀谦、祖永祥、周凯来、蔡李靖、庄义昱、金周宇等。尤其感谢周凯来师兄引领我发表国际顶会论文、参与全国性竞赛，让我领会了要注重学术研究和科研活动过程中的每一步细节。

感谢和我同级进入 CITE Lab 实验室的同学以及陆续到来的学弟学妹：吕涛、施展、孙旭森、吴萌华、袁庆博、刁政宇、叶浩、杨富强、徐志成、董祥宇、张梦雅、黄成龙、胡李豪、李世桥、李齐平、张丁源等。感谢你们和我一同融入实验室，尤其感谢吕涛、施展和叶浩，我们共同进行学术研究和讨论，我们一起泛舟湖上的快乐时光是我宝贵的记忆。

感谢我的父亲和母亲，是他们让我无后顾之忧地投入学习与学术研究，探索人生与科研的乐趣，让我拥抱世界、走向未来。

王乙卜

2024 年 5 月于南京大学



## 攻读硕士学位期间研究成果

- [1] T Lv, H Ye, Q Yuan, Z Shi, **Y Wang**, S Wang, X Cao, “Aperture Diffraction for Compact Snapshot Spectral Imaging” . IEEE/CVF International Conference on Computer Vision (ICCV), 2023.
- [2] Z Shi, H Ye, T Lv, **Y Wang**, X Cao, “Compact Self-adaptive Coding for Spectral Compressive Sensing” . IEEE International Conference on Computational Photography (ICCP), 2023.
- [3] K Zhou, **Y Wang**, T Lv, Y Li, L Chen, Q Shen, X Cao, “Explore Spatio-temporal Aggregation for Insubstantial Object Detection: Benchmark Dataset andBaseline” . IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.
- [4] 王乙卜，沈秋，曹汛。一种基于高光谱的自动驾驶场景分割方法，中国发明专利，已公开，公开号：CN115661818A
- [5] 王乙卜，杨富强，徐志成。MBIST 自动规划分组算法，第五届集成电路 EDA 设计精英挑战赛三等奖，2023
- [6] 吕涛，施展，王乙卜，霍城城，叶浩，徐志成，董祥宇，俞瀚洋，王琳翔，黄尔齐，周凯来，指导教师曹汛，沈秋。基于国产解决方案的天然气管道沿线危险目标检测关键技术与工业化实践，第九届中国国际大学生创新大赛金奖，2023
- [7] 吕涛，王乙卜，施展，周凯来，指导教师曹汛。化工气体泄漏光谱视频智能检测算法，第三届中国研究生人工智能创新大赛一等奖，2021



## 学位论文出版授权书

本人完全同意《中国优秀博硕士学位论文全文数据库出版章程》（以下简称“章程”），愿意将本人的学位论文提交“中国学术期刊（光盘版）电子杂志社”在《中国博士学位论文全文数据库》、《中国优秀硕士学位论文全文数据库》中全文发表。《中国博士学位论文全文数据库》、《中国优秀硕士学位论文全文数据库》可以以电子、网络及其他数字媒体形式公开出版，并同意编入《中国知识资源总库》，在《中国博硕士学位论文评价数据库》中使用和在互联网上传播，同意按“章程”规定享受相关权益。

作者签名：\_\_\_\_\_

\_\_\_\_年\_\_\_\_月\_\_\_\_日

论文题名	高光谱图像光源估计及其应用				
研究生学号	MG21230082	所在院系	电子科学与 工程学院	学位年度	2024
论文级别	<input checked="" type="checkbox"/> 学术学位硕士 <input type="checkbox"/> 专业学位硕士 <input type="checkbox"/> 学术学位博士 <input type="checkbox"/> 专业学位博士				
作者 Email					
导师姓名	曹汛				

论文涉密情况：

不保密

保密，保密期（\_\_\_\_\_年\_\_\_\_月\_\_\_\_日至\_\_\_\_\_年\_\_\_\_月\_\_\_\_日）

