

分类号：

单位代码：10140

密 级：公开

学 号：4032032366

遼寧大學

# 硕 士 学 位 论 文

中文题目： 基于人类视觉特性的全景图像质量评价方法研究

Research on Omnidirectional Image Quality Evaluation

英文题目： Method Based on Human Visual Characteristics

论文作者： 殷 晓 华

指导教师： 刘 允 副教授

专 业： 计算机应用技术

完成时间： 二〇二三年五月

申请辽宁大学硕士学位论文

基于人类视觉特性的全景图像质量评价方  
法研究

Research on Omnidirectional Image Quality Evaluation  
Method Based on Human Visual Characteristics

作 者： 殷 晓 华

指导教师： 刘 允 副教授

专 业： 计 算 机 应 用 技 术

答辩日期： 2023 年 5 月 18 日

二〇二三年五月·中国辽宁

## 辽宁大学学位论文原创性声明

本人郑重声明：所呈交的学位论文是本人在导师的指导下独立完成的。论文中取得的研究成果除加以标注的内容外，不包含其他个人或集体已经发表或撰写过的研究成果，不包含本人为获得其他学位而使用过的成果。对本文的研究做出重要贡献的个人和集体均已在文中进行了标注，并表示谢意。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：殷晓华

2023年 6 月 2日

## 学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交学位论文的原件、复印件和电子版，允许学位论文被查阅和借阅。本人授权辽宁大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编学位论文。同时授权中国学术期刊（光盘版）电子杂志社将本学位论文收录到《中国博士学位论文全文数据库》和《中国优秀硕士学位论文全文数据库》并通过网络向社会公众提供信息服务。学校须按照授权对学位论文进行管理，不得超越授权对学位论文进行任意处理。

保密（ ），在\_\_\_\_年后解密适用本授权书。（保密：请在括号内划“√”）

授权人签名：殷晓华 指导教师签名：刘 侃

日期：2023年 6 月 2日 日期：2023年 6 月 2日

## 摘 要

现如今，随着 5G 时代的到来和虚拟现实技术的不断完善为沉浸式用户体验提供了强有力的技术支撑，全景图像也逐渐引起了研究者的广泛关注。随着元宇宙概念的诞生，全景图像再次被推向热潮。然而全景图像在采集、压缩、传输及存储中会产生各种各样的噪声及损失，这会影响人们的观看体验甚至是对图像的理解。为了提升用户的体验，衡量全景图像的质量以得到优化指标成为了当前学者的一项重要研究课题。

通常情况下，全景图像是由观察者通过头戴式设备来观看的。在佩戴该设备的情况下，观察者的视觉范围具有一定的局限性。另外，负责处理视觉信息的人类视觉系统（Human Visual System, HVS）在感知图像时其注意力机制会发挥关键的作用，这提升了大脑感知世界的效率。相对于传统图像，全景图像有着更复杂的视觉信息，因此研究视觉注意力机制是衡量全景图像质量必不可少的一环。

因此，本文就全景图像显著性检测和全景图像质量评价两方面展开研究，主要研究内容如下：

一、考虑到现有的显著性检测模型没有考虑全景图像的特点，因此本次研究提出了一种针对全景图像的显著性分析模型，利用超像素技术将全景图像以像素组作为基本的处理单元，这相比于像素级方法有效减少了图像中的冗余信息，并为显著性检测提供更有用的空间结构。然后基于格式塔理论，本文利用了布尔图拓扑结构的尺度不变性，通过计算不同阈值的布尔图来分离图像的前景和后景，利用不同的前景-后景信息检测得到全景图像的显著性区域，将其通过  $L_2$  范数进行归一化，最后利用线性组合方法得到最终的高斯模糊图。本文为了充分地考虑全景图像的特性，建立了一种针对全景图像显著性的准确分析模型。

二、受 HVS 分层感知的启发，可以将视觉质量下降描述为一个分层过程，即对输入的视觉信号，其感知过程是从局部细节逐步过渡到全局语义的。因此，为了更好的感知全景图像的质量，本文模拟了 HVS 的感知过程。首先衡量最先受到影响的细节信息，利用二维离散小波变换分离出全景图像中能反映细节信

息的高频分量，结合灰度共现矩阵提取图像的细节特征；同时利用信息熵衡量通过显著性算法得到显著性信息，用以表示全景图像中重要的局部损失；最后，本工作利用自然场景统计方法分析了全局信息对全景图像整体语义的表达。除此之外，本文还考虑了不同颜色通道之间的相关性和依赖性，通过计算全景图像的跨通道局部二值模式来表达图像中的颜色信息。通过结合全景图像中不同的特征，建立了一种全面感知全景图像质量的评价模型。

为了验证算法的性能，本文对所提模型在全景图像数据集上利用支持向量回归（Support Vector Regression, SVR）技术进行了验证。通过与其他模型的对比，实验结果表明本文提出的全景图像显著性建模模型能有效地检测全景图像的显著性区域，并且提出的全景图像质量评价模型在两个公开可用的数据集上准确地预测出了与视觉感知保持一致的质量分数。

**关键词：**全景图像，无参考质量评价，HVS，视觉显著性，SVR

## ABSTRACT

Nowadays, with the arrival of 5G era and the continuous improvement of virtual reality technology providing strong technical support for immersive user experience, panoramic image has gradually attracted wide attention of researchers. With the birth of the concept of metacomes, panoramic images have been pushed to the craze again. However, all kinds of noise and loss will occur in the acquisition, compression, transmission and storage of panoramic image, which will affect people's viewing experience and even their understanding of the image. In order to improve the user experience, it has become an important research topic for scholars to measure the quality of panoramic image to get the optimization index.

Typically, panoramic images are viewed by an observer through a headset. When the device is worn, the visual range of the observer has certain limitations. In addition, the attention mechanism of the Human Visual System (HVS), which is responsible for processing visual information, plays a key role in perceiving images, which improves the efficiency of the brain. Compared with traditional images, panoramic images have more complex visual information, so the study of visual attention mechanism is an indispensable part of measuring the quality of panoramic images.

Therefore, this paper carries out research on the significance detection of panoramic images and the quality evaluation of panoramic images. The main research contents are as follows:

1. Considering that the existing significance detection model does not consider the characteristics of panoramic images, a significance analysis model for panoramic images is proposed in this study. Superpixel technology is used to take the pixel group as the basic processing unit of panoramic images. Compared with the pixel-level method, this method can effectively reduce the redundant information in the image and provide a more useful spatial structure for significance detection. Then, based on the Gestalt theory, this paper makes use of the scale invariance of the topology structure of the Boolean graph, separates the foreground and rear view of the image by calculating the Boolean graph of different threshold values, detects the

significance region of the panoramic image by using different foreground-rear information, normalizes it by L2 norm, and finally obtains the final Gaussian fuzzy graph by using the linear combination method. In order to fully consider the characteristics of panoramic images, an accurate analysis model for the saliency of panoramic images is established in this paper.

2. Inspired by the hierarchical perception of HVS, visual quality degradation can be described as a hierarchical process, that is, the perception process of input visual signals is gradually transitioning from local details to global semantics. Therefore, in order to better perceive the quality of panoramic images, this paper simulates the perception process of HVS. Firstly, the first affected detail information was measured, and the high frequency component reflecting the detail information of the panoramic image was separated by using the two-dimensional discrete wavelet transform, and the detail features of the image were extracted by combining the gray co-occurrence matrix. At the same time, the significance information is obtained by the significance algorithm using the measure of information entropy, which is used to represent the important local loss in the panoramic image. Finally, the natural scene statistics method is used to analyze the representation of global information on the whole semantic of panoramic images. In addition, this paper also considers the correlation and dependence between different color channels, and expresses the color information in the panoramic image by calculating the cross-channel local binary mode. By combining different features of panoramic image, a comprehensive perception of panoramic image quality evaluation model is established.

In order to verify the performance of the proposed algorithm, the proposed model was tested using Support Vector Regression (SVR) technique on a panoramic image dataset. By comparison with other models, the experimental results show that the proposed panoramic image saliency modeling model can effectively detect the saliency region of the panoramic image, and the proposed panoramic image quality evaluation model can accurately predict the quality score consistent with the visual perception on the two publicly available data sets.

**Keywords:** Omnidirectional image, no reference quality evaluation, HVS, visual salience, SVR

## 目 录

第 1 章 绪论.....	1
1.1 研究背景与意义 .....	1
1.2 国内外研究现状 .....	4
1.2.1 图像显著性图建模方法.....	4
1.2.2 全景图像质量评价方法.....	5
1.3 本文主要研究内容 .....	7
1.4 本文章节安排 .....	8
第 2 章全景图像质量评价相关理论背景 .....	10
2.1 人类视觉感知机制 .....	10
2.1.1 视觉显著性注意力机制.....	10
2.1.2 视觉的分层感知机理 .....	11
2.2 全景图像数据集 .....	12
2.3 相关性能评价指标 .....	16
2.3.1 显著性图建模方法的性能评价指标.....	16
2.3.2 全景图像质量评价方法的性能评价指标.....	17
2.4 本章小结 .....	18
第 3 章 基于视觉感知特性的全景图像显著性建模方法 .....	19
3.1 问题描述 .....	20
3.2 算法思想 .....	20
3.3 显著性图建模方法 .....	22
3.3.1 全景图像的超像素分割算法.....	23
3.3.2 基于全景超像素图像的布尔映射算法.....	26
3.3.3 基于注意力图的显著性建模算法.....	28
3.4 算法的实现流程 .....	30



3.5 实验结果与分析 .....	31
3.6 本章小结 .....	35
第4章 基于多维视觉特征的全景图像质量评价方法 .....	36
4.1 问题描述 .....	36
4.2 算法思想和流程 .....	37
4.3 全景图像视觉感知特征提取 .....	38
4.3.1 基于灰度共现矩阵的频域细节特征.....	38
4.3.2 基于信息熵的局部显著性特征.....	43
4.3.3 基于自然场景统计特性的全局语义特征.....	43
4.3.4 基于跨通道局部二值模式的颜色特征.....	46
4.4 质量回归与预测 .....	49
4.5 实验结果与分析 .....	50
4.6 本章小结 .....	56
第5章 总结与展望 .....	57
5.1 总结 .....	57
5.2 展望 .....	58
致谢.....	59
参考文献.....	60
攻读学位期间发表的学术论文及参加科研情况 .....	67

## 图 表 目 录

## 图目录

图 1-1 观看轨迹图.....	2
图 1-2 全景图像的投影示例.....	3
图 1-3 原始图像和 4 种类型失真图像示例.....	3
图 2-1 视觉显著性介绍示例.....	11
图 2-2 分层感知过程图.....	12
图 2-3 CMP 投影规则 .....	13
图 2-4 CMP 投影结果图 .....	13
图 2-5 等矩形投影结果图.....	14
图 2-6 分段球面投影结果图.....	14
图 2-7 CVIQD2018 数据集图像示例 .....	15
图 2-8 OIQA 数据集图像示例 .....	16
图 3-1 视觉显著性注意力机制示例图.....	19
图 3-2 基于视觉感知特性的全景图像显著性建模方法流程.....	21
图 3-3 传统图像和全景图像及其显著性图示例.....	22
图 3-4 像素与聚类中心关联示意图.....	24
图 3-5 不同聚类中心数作用下的图像分割效果.....	25
图 3-6 图像前景-后景介绍示意图.....	26
图 3-7 不同阈值下的布尔图.....	28
图 3-8 输入图像与其高斯模糊图.....	29
图 3-9 四种不同检测算法在 AUC 指标下的性能对比.....	32
图 3-10 全景图像在 4 种建模方法计算下的显著性图.....	34
图 4-1 图像模糊示例.....	36
图 4-2 基于视觉显著性特性的全景图像质量评价方法流程图.....	38
图 4-3 原始图像与受损图像在空间域上的差异对比图.....	39
图 4-4 DWT 变换子带图.....	40
图 4-5 DWT 实现过程.....	40

图 4-6 原始图像和不同受损类型下高频图像的直方图 .....	42
图 4-7 自然图像和人工图像示例 .....	44
图 4-8 不同扭曲类型下全景图像的 MSCN 系数分布曲线图 .....	46
图 4-9 颜色信息主导图像内容的示例 .....	47
图 4-10 LBP 的计算规则 .....	48
图 4-11 R-G-B 序列下 CCLBP 的构建过程。 .....	48
图 4-12 SVR 回归原理示意图 .....	50
图 4-13 预测分数和主观分数的性能拟合散点图 .....	55

## 表目录

表 3-1 四种不同显著性检测算法在 4 种指标下的性能对比 .....	33
表 3-2 不同聚类中心数下的显著性检测算法的性能对比 .....	33
表 3-3 不同聚类中心数下的显著性检测算法的计算复杂度对比 .....	33
表 4-1 全景图像 CVIQD2018 数据集上的整体性能对比 .....	51
表 4-2 OIQA 数据集上的性能比较 .....	52
表 4-3 单个失真类型在 CVIQD2018 数据集上的性能比较 .....	53
表 4-4 计算复杂度 .....	53
表 4-5 不同特征的性能比较 .....	56

## 第 1 章 绪论

全景图像质量评价旨在确定全景图像的质量受损情况，以便提高图像质量，拓宽应用范围，提升用户的观看体验。无参考全景图像质量评估的最终目标是设计一个全面的感知模型。该模型可以在没有先验知识的情况下预测受损的全景图像质量。本章主要对全景图质量评价方法的研究背景意义，国内外的研究现状进行了简要的介绍，简单的概述了本文的研究内容和主要贡献以及本文的总体结构安排。

### 1.1 研究背景与意义

网络和传输技术的快速发展推动了各种多媒体应用和广播服务的发展。由于越来越多容易访问的多媒体服务的出现，最终用户期望得到的服务体验质量 (Quality of Experience, QoE) 越来越高。因此，如何测量和改善终端用户的感知体验质量成为服务提供商面临的一个紧迫问题，目前越来越多的研究者已经提出了一些模型来测量和增强感知质量<sup>[1-5]</sup>。

虚拟现实 (Virtual Reality, VR) 是近年来的一个非常热门的研究课题<sup>[6-7]</sup>。作为 VR 的一个重要组成部分，全景图像和视频围绕一个中心位置记录所有可能方向的视图，它们在头戴式显示器 (Head Mounted Device, HMD) 的帮助下为用户提供真实场景的沉浸式体验<sup>[8]</sup>。人通过头戴式设备的观看轨迹如图 1-1 所示。具体地说，其观看轨迹是从球体的内中心点向外扩展，这也就意味着在佩戴 HMD 设备时，某一时间点能看到的范围是有一定限制的。一般来说，其有效的视觉范围在水平上是 110 度到 120 度之间，垂直方向在 60 度的范围内，其它的地方是看不到的。这就意味着在人们观看全景图像时，不同的区域对观察者的影响存在着一定的差异。

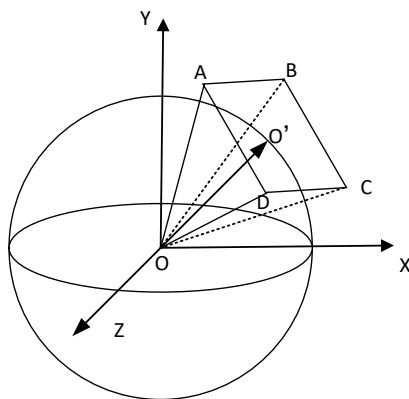


图 1-1 观看轨迹图

除了在佩戴 HMD 设备时不同区域的内容对感受的影响有很大的不同之外，人眼在观看普通的平面图像时也会有一定的侧重点，备受关注的区域会对人的感官影响更大。与此同时，注意力区域的受损信息更容易被捕捉到，进而影响人的观看体验，这是由 HVS 决定的。视觉注意力机制在观看图像时起到了至关重要的作用，因此考虑视觉注意力机制是模拟人类感知外界信息过程中必不可少的关键因素。

另外，由于全景图像观看环境的特殊性，在通常情况下，全景图像的像素信息都是按照一定的规则投影到平面上进行展示、传输、处理和存储的<sup>[9]</sup>，其投影示例图由图 1-2 展示。这是因为目前全景图像的发展还未成熟，关于如何对全景图像编码问题至今没有完成一个统一的、准确的方案。因此，目前对于全景图像的研究大都是利用传统图像的编码技术，即将全景图像映射到 2 维平面上，然后再进行后续的处理。随着全景图像方面的研究增多，多种投影格式被相继提出并被广泛使用到各个方面。例如 cube map projection (CMP) 格式，equiangular cube map (EAC) 格式，Equirectangular projection (ERP) 格式等。其中 ERP 格式（等矩形格式）是由 the motion picture expert group (MPEG) 开发并且由于该格式具有观看方便、容易生成和存储的特点，所以成为默认的投影方式，在行业中使用也最为广泛，等矩形格式投影的示意图如图 1-2 所示。

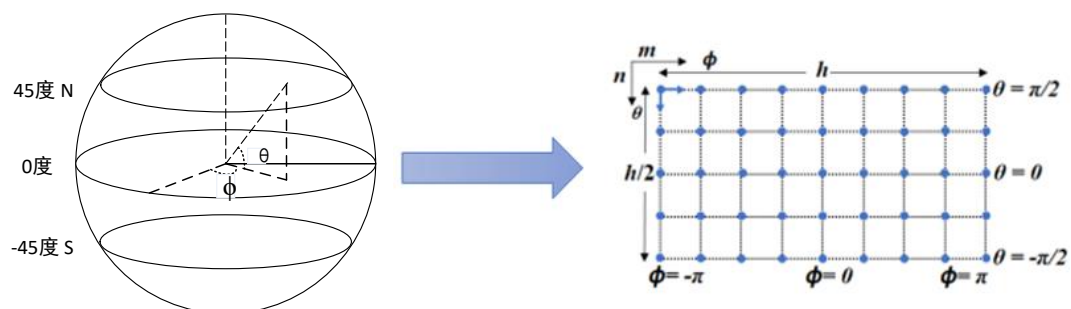


图 1-2 全景图像的投影示例

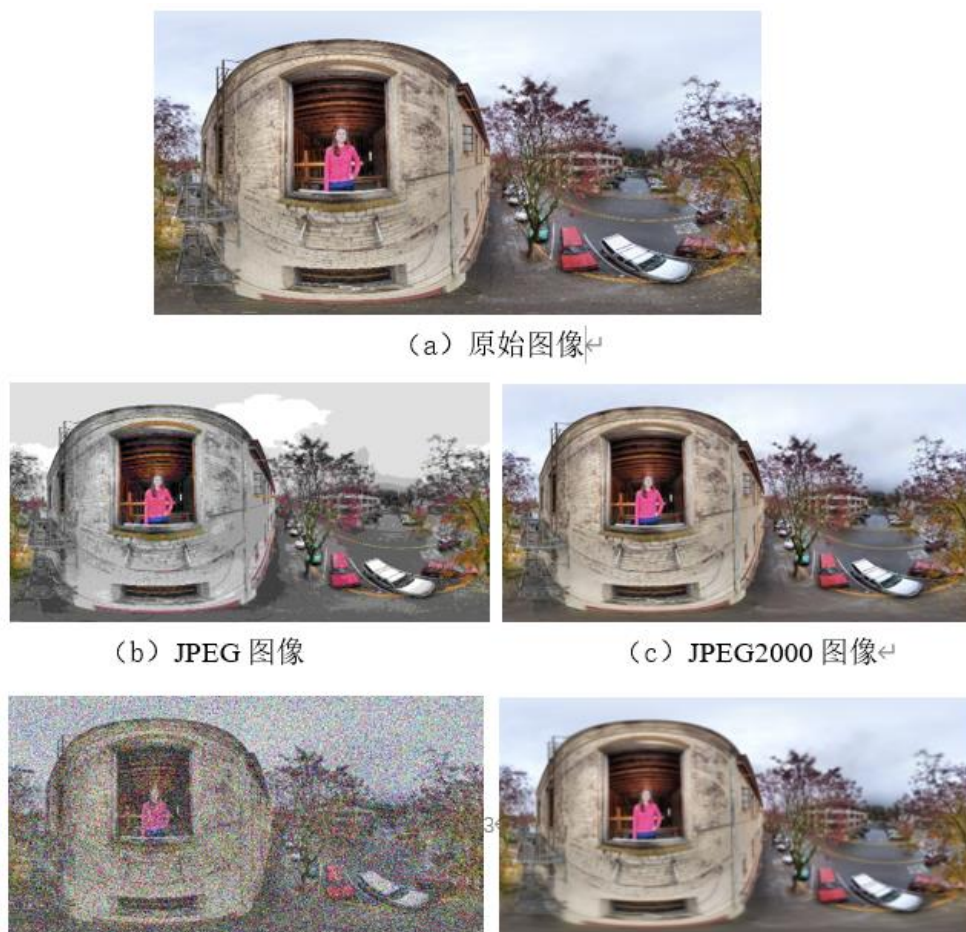


图 1-3 原始图像和 4 种类型失真图像示例

图像的失真会对用户的体验造成严重的影响，甚至会产生对图像内容的误解。不同类型的失真给人的观感也存在一定的差异，例如图 1-3 所示，原始图像在没有经过压缩的情况下除了投影导致的 ERP 图像两极的拉伸扭曲外，其它

地方的图像内容都被精准的表达和展示。不难看出经过压缩、噪声和模糊的全景图像展示了不同的不自然效果造成了图像信息的丢失和改变，进而破坏图像的完整性，使得用户体验差，因此如何提高用户体验成为研究者的重要课题。

提高用户观看全景图像体验的前提是如何度量全景图像的质量，因此发展全景图像的前提必须了解当前全景图像的质量，才能进一步的研究如何提高。由于人是最终对全景图像质量进行评估的评价者，因此观看者的感受是评价图像的最终评判者。但通过将全景图像展示在观察者面前，让观察者根据主观印象直接给出相应的质量分数的方法效率极低并且人之间的差异会导致同一幅图像的质量出现不同的评价结果。因此为了符合大众的评判水平，需要综合所有观察者的结果来作为最终的图像质量值，这样就耗费了大量的人力物力和时间成本。很多情况下都是将主观评价的分数作为设计客观算法的参考和衡量的依据。受此启发，衡量全景图像的质量必须以人的感受出发，模拟人眼观看图像的过程和特性，进而评价出与人的评分保持高度一致的质量分数。因此，在提高全景图像质量之前非常有必要对全景图像的质量进行评估进而得到提高质量的关键指标。

## 1.2 国内外研究现状

### 1.2.1 图像显著性图建模方法

在计算机中，视觉显著性的研究很早就开始于二十世纪九十年代，往往显著性检测的研究对象针对于图像和视频。根据生物学中显著性机制的处理过程，Itti<sup>[10]</sup>提出在计算机中模拟该过程行为进而使计算机学会近似生物学的显著性机制使其处理过程大致符合人眼的注意顺序，然而该模型的过程复杂、效率低、计算量高并且检测性能差。考虑到 Itti 模型的缺点，Harel 等人<sup>[11]</sup>在 Itti 模型的基础上进行了一定的改进，具体地说，Harel 提出的 GBVS 模型是利用了马尔科夫随机场来计算特征显著图，其相对于 Itti 模型在性能上有了很大的提升，但是仍存在计算量大的问题。

侯等人<sup>[12]</sup>考虑到变换域的特点，提出变换域中谱残差的概念。通过使用傅里叶变换中图像的对数谱表示来计算图像的显著性值。该思想从一个新的角度出发从信息角度分析了图像的组成部分，具有计算速度快的显著特点但是其检

测的显著性区域不明显。Achanta 等人<sup>[13]</sup>从图像的基本单元像素出发,通过像素的亮度来确定像素的显著性值,最终总结每个像素的显著性质进而得到显著性映射。该模型对边缘信息敏感并且能保证显著性映射的分辨率,然而该模型无法对显著性目标所在的中心区域进行明显的突出。Achanta 等人还针对颜色空间中的蕴含的信息和滤波在图像中的作用分析了图像的显著性值<sup>[14]</sup>,相比其它显著性计算模型,该模型构造简单,计算速度比较快,大大提升了显著性检测的效率。

近年来已经开发了大量的算法来检测平面 2D 图像中的显著区域,但对 360 度图像显著性的研究有限。方等人<sup>[15]</sup>认为有效的视觉注意力建模是帮助提高 VR/AR 数据整体体验质量的关键因素。因此在全景图像的显著性研究中,他们根据格式塔理论的图像基础法则,提出了一种 360 度图像的基于质量保证变换的显著性检测模型,通过考虑 LAB 颜色空间中呈现的图像的互补信息提取全景图像的底层细节特征计算显著性预测。

### 1.2.2 全景图像质量评价方法

图像质量评估 (IQA) 的目的是预测与人类视觉感知一致的质量<sup>[16-18]</sup>。国内外学者都开展了许多研究和探索。其中包括美国的得克萨斯大学奥斯汀分校、加拿大的滑铁卢大学,以及我国的上海交通大学等等。

图像质量评估方法根据是否有人直接参与到图像质量评价过程中分为两类:主观图像质量评价方法和客观图像质量评估方法<sup>[19]</sup>。基于主观评估实验,通过要求人们对 OI 质量和数据处理进行评分。但是该方法存在致命的缺点即缺乏现实应用性,因此目前都将作为设计客观算法的参考和衡量的依据并基于此建立了一些图像数据集。

考虑到实际应用,客观 IQA 方法引起了研究人员的广泛关注<sup>[20-24]</sup>,包括完全参考 (FR)、减少参考 (RR) 和无参考/盲 (NR/B) 图像质量评价方法<sup>[25]</sup>。其中客观无参考 (NR) 图像质量评估是指使用算法对图像进行自动质量评估,使得算法在对质量进行预测之前接收的唯一信息是其质量被评估的失真图像。在分类的另一端是全参考算法,该算法不仅需要失真图像作为输入,还需要“干净”的原始参考图像作为输入来评估失真图像的质量。在这两个极端之间的折中部分,减少参考方法除了失真图像之外,还具有关于参考图像 (例如,



水印)的一些信息,但不具有实际参考图像本身。与2D图像质量评估方法的分类类似,全景图像质量评价(OIQA)模型也分为FR-OIQA、RR-OIQA和NR-OIQA<sup>[26]</sup>。

对于FR-OIQA模型,它们同样需要参考图像的所有信息。通过使用2D IQA模型的许多指标,如峰值信噪比(PSNR)<sup>[27]</sup>、结构相似性(SSIM)<sup>[28]</sup>、视觉信息保真度(VIF)<sup>[29]</sup>、视觉显著性诱导(VSI)<sup>[30]</sup>、梯度幅度相似性偏差(GMSD)<sup>[31]</sup>、特征相似性(FSIM)<sup>[32]</sup>等扩展到OIQA领域评价全景图像的质量。其中SSIM主要原理是计算参考图像和待测图像的结构相似度进而衡量图像的质量;FSIM模型考虑了相位一致性对图像局部结构的作用,即图像的相位不会因为图像的失真或扭曲而发生变化,以此利用图像的该特性分析图像的特征信息。然而,这些FR模型无法通过简单地将2D IQA模型扩展到OIQA区域来实现良好的性能。

考虑到VR环境中观看条件的重要性,Sui等人<sup>[33]</sup>建立了一个FR-OIQA模型,以评估由具有不同观看条件的多个视频表示的全景图像的感知质量。受全景图像特性的启发,Yu等人<sup>[34]</sup>提出了一种基于PSNR的S-PSNR方法来评估OI的质量。Zakharchenko等人<sup>[35]</sup>建立了一个CPP-PSNR模型,以评估球面全景视频的质量。此外,Zhou等人<sup>[36]</sup>建立了加权SSIM(WS-SSIM)模型,以解决基于SSIM的空间转换引起的偏差传播。

全参考图像质量评价模型已经取得了不错的进展,但在很多实际情况中,经常没有办法获取到完整的参考图像的信息。因此,学者和研究人员考虑利用可获得的部分参考信息研究图像的质量,因此许多基于部分参考信息减少参考图像质量评价算法被相继提出。例如,考虑到图像梯度和纹理信息在失真图像中表现,Maalouf等人<sup>[37]</sup>基于这些底层图像信息结合图像的受损类型提出了一种考虑群变换特性的RR-IQA模型。张等人<sup>[38]</sup>认为图像的锐度图可以有效分析图像的质量,因此基于该图特征提出了减少参考图像质量模型。Sun等人<sup>[39]</sup>提出图像质量的退化可以通过分析图像中的信息增益以及图像的纹理相似度进行感知。离散小波变换(DWT)被文献<sup>[40]</sup>认为可以应用于计算图像的局部梯度进而利用梯度信息对图像的质量进行评估。虽然减少参考的质量评价模型也取得了一定的成果,但参考信息的获取仍然是学者在研究过程中存在的难题。

考虑到在实际情况无法获得原始图像的参考信息,NR-OIQA模型引起了研究者的广泛关注。Zheng等人<sup>[41]</sup>提出了一种采用分段球面投影(SSP)格式

的 OI 的 BOIQA 模型，其中双极区域和赤道区域是处理 OI 失真的关键。通过使用深度学习技术，Sun 等人<sup>[42]</sup>开发了 NR-OIQA 模型 MC360IQA，以测量 OI 的质量。基于立方体图投影（CMP）格式，Jiang 等人<sup>[43]</sup>介绍了热点图，以分析 OI 中的受损区域和细节，并建立了一个完整的名为 CPBQA 的质量模型。一种盲质量模型<sup>[44]</sup>通过量化重影和结构不一致伪影，将卷积稀疏编码（CSC）用于拼接全景图像。考虑到视口质量的重要性，Xu 等人<sup>[45]</sup>基于图卷积网络（VGCN）建立了盲 OIQA 模型。Zhou 等人<sup>[46]</sup>还认为，具有不同视口的图像质量是影响全景图像质量的重要因素，因此提出了基于 ERP 格式的具有不同视口图像的 DDAMN 模型，以感知全向图像的质量。受感知建模视角的启发，Jiang 等人<sup>[47]</sup>提出了一种基于颜色全向失真（COD）单元的基于多角度投影的 BOIQ（MP-BOIQA）模型。

在质量评价中，众多研究者考虑到人眼视觉特性发挥的重要作用，因此将视觉显著性机制考虑到衡量图像质量的指标中。例如 Chetouania 等人<sup>[25]</sup>根据经典的 GBVS 算法计算待评价的显著性图，提出利用路径扫描器得到图像的多个注视点，并根据注视点对图像进行分割，然后利用神经网络对输入图像计算其质量分数。随着研究者在视觉注意力机制方面取得成果的逐渐增多，Jabar 等人<sup>[48]</sup>提出全景图像的质量不应该只考虑整体图像的失真信息，还应该重点感知显著性区域的扭曲细节。

相比于需要参考图像参与的评价方法，无参考的评价方法由于没有参考信息实现起来难度更大，但在现实生活中失真图像没有任何参考信息的标注，因此无参考质量评价方法更适合于实际应用<sup>[49]</sup>。尽管上述无参考 OIQA 模型实现了比 FR 和 RR 模型更好的性能，但很多模型并没有考虑人类视觉注意力机制导致提取了不准确的质量感知特征进而影响了它们的表现性能。为了解决这个问题，本文需要考虑具有代表性的视觉特征并检测全景图像中的显著性区域并分析其中的失真信息来构建一种有效衡量全景图像质量的评价模型。

### 1.3 本文主要研究内容

VR 技术的发展和元宇宙的提出，全景图像的发展受到大众的广泛关注，作为全景图像发展的关键指标图像的质量成为了学者的重要研究对象。本文针

对全景图像的质量提出了一种基于视觉显著性特征的全景图像质量评价方法。该方法分析了人类视觉感知机制，模拟了视觉感知过程，考虑了全景图像 ERP 投影的适用性和应用广泛性，利用基于 ERP 格式的两个公开可用的全景图像数据集对模型进行了训练和测试，并利用 SVR 对全景图像进行准确地预测。实验结果表明本文提出的评价模型通过考虑多层次感知特性分析图像中的受损信息进而评价质量的方法做出了稳定可靠的性能表现。

本文的研究内容和主要贡献主要概括为以下几点：

(1) 根据 HVS 中的分层感知结构，本文导出分解的高频子带，并使用灰度共现矩阵来反映全景图像中的低级别细节信息，这使本文的质量评价模型在捕获细节失真方面具有更好的准确性。

(2) 现有的显著性检测模型没有考虑全景图像的特点并且针对全景图像提出的显著性检测方法没有考虑到模型的计算复杂度，因此本文利用超像素技术将信息量巨大的全景图像以像素组作为基本的处理单元，这减少了全景图像中的冗余信息，并为显著性检测提供更有用的空间结构。基于格式塔理论，本文利用布尔图的拓扑结构信息的特性，利用布尔图来分析全景图像的显著性区域。这可以有效地降低本工作模型的计算复杂性，该超像素反映了人类视觉在有意义的视觉区域中感知全方位图像的事实。

(3) 除了特定特征和局部显著性特征外，对投影的 ERP 图提取基于自然场景统计的全局特征作为感知视觉分层过程的高级特征，准确地捕捉了受损全景图像的自然度，并进一步提高了所提出的模型的整体性能。

(4) 由于分层特征不能表达全景图像中失真的颜色信息，因此设计了一种新的跨通道局部二值模式描述符来计算具有不同通道的像素之间的颜色依赖关系，从而捕获通道间的颜色信息。

## 1.4 本文章节安排

本文的章节将在基于显著性特征的全景图像质量评价算法上进行了如下安排：

第一章：绪论，首先介绍了本文研究的背景与意义，其次分别概述了图像显著性图建模方法的研究现状和国内外关于全景图像质量评价算法的研究现状，然后对本文的研究内容和主要贡献做了简单的介绍，最后简要阐述了本文的章节安排。

第二章：全景图像质量评价相关理论背景，首先通过从视觉注意力机制和视觉的分层感知机理两个方面详细介绍了人类视觉感知机制，然后重点分析了全景图像的数据集理论，然后阐述了本工作应用到的相关性能指标，最后对本第二章进行了简单的小结。

第三章：基于视觉感知特性的全景图像显著性建模方法研究，首先对本章算法的主要思想和流程做了简单的概述，其次着重介绍的显著性图的建模方法，包括详细介绍全景图像的超像素分割算法和基于布尔映射的显著性算法的搭建流程和细节，然后针对该算法进行了实验结果与性能做出的准确的分析和比较，最终对该章节算法的主要内容做了简单的总结。

第四章：基于多维视觉特征的全景图像质量评价方法的研究，首先介绍了基于视觉显著性特征的无参考全景图像质量评价方法的主要思想和流程，然后从基于灰度共现矩阵的频域细节特征、基于信息熵的局部显著性特征、基于自认场景统计的全局语义特征和基于跨通道的局部二值模式的颜色特征分析了全景图像的视觉感知特征，然后详细介绍了用于全景图像的质量回归、预测算法，然后对该方法的实验结果和性能比较进行了阐述，最后进行了总结。

第五章：总结与展望，总结了本文的主要工作并对下一步的研究方向内容进行了展望。

## 第2章全景图像质量评价相关理论背景

### 2.1 人类视觉感知机制

近几十年来，世界逐步进入信息时代。其中，视觉信息的比重占有80%以上。数字图像和视频由于其表现形式和视觉效果，正逐步成为信息传播的主要媒介。作为人体感官器官之一视觉系统是接受图像、视频的主要载体，是负责感知处理视觉信息的模块，也是人们认识客观世界最主要的途径。

人类是通过感觉器官在人脑中形成的直接反映来感知客观事物的。其中视觉感受器官包含晶状体、视网膜和中央凹等几个重要的构造，其中晶状体通过肌肉收缩和扩张以控制视觉焦距，视网膜则是成像的载体，而中央凹是视觉敏感区域。视觉是通过感光细胞进行感知的视觉信号的，其中中央凹区域只有视锥细胞的分布，视锥细胞能感受到强光并且对颜色有分辨功能，随着与中央的距离增大，视杆细胞逐渐增多，它主要负责感受弱光并且无法分辨颜色信息，但其对光的敏感度比视锥细胞强万倍。

#### 2.1.1 视觉显著性注意力机制

中央凹是处于视网膜上最重要的部分，其具有最高视力的特点，相对而言，其它地方的视力相对要弱，这在视觉研究中很少被提及。为此，本节举个例子解释该现象，当人们仔细盯着电脑屏幕时，虽然除了屏幕之前的东西也可以被模糊的看到，但是除了屏幕之外的信息人们几乎看不清楚具体的细节。其原因就是，整个视网膜上投射了人们所能观察到的所有信息成像，而能看到的视觉范围是有限的并且受到视网膜覆盖角度的限制，只有刚好投射到中央凹的成像内容是可以被清楚地看清，人的大脑可以控制眼球的运动使人们想看的内容刚好聚焦与中央凹部分。

基于上述视觉原理，人们可以清楚的了解到投射到中央凹的内容往往是人们在某个时间点需要重点关注的信息。在人们浏览或者观看图像或者世界时，侧重点区域时一直存在的，即显著性区域是人们着重关注的地方，因此视觉注意力机制在人类了解世界时发挥了重要的作用。如图2-1展示了视觉注意机制

的示例，当人们在观察这两幅图像时，首先主要到的蓝色的鸭子，而忽略其它的鸭子也存在很大的不同。在面对大量的视觉信息时，人类视觉系统会有选择性地处理接受到的视觉信息，其显著性注意力机制会发挥非常重要的作用，在复杂且丰富的视觉场景下，注意力机制会自动选择场景中重要信息，即实现将复杂的视觉场景自动进行简化，而大脑只需要对注意力机制简化后的感兴趣的内容进行高效处理，这极大的提高了大脑在面对大量信息时，感知世界内容的效率。尤其适用于拥有庞大信息量的全景图像并对图像处理过程起到提高效率的作用。

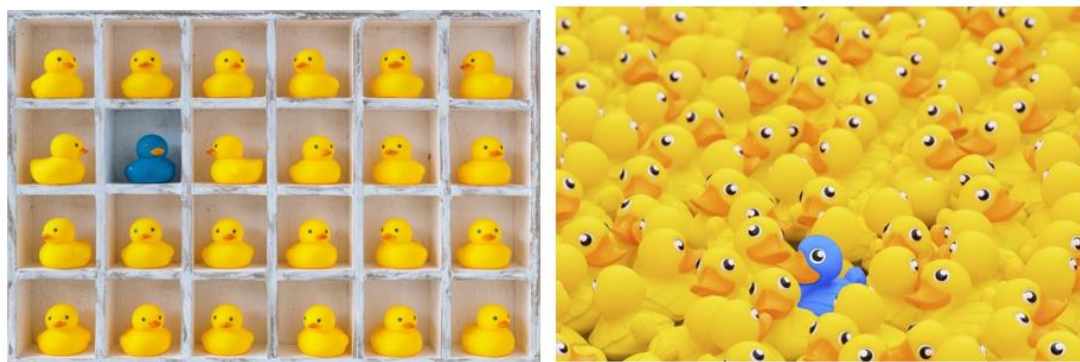


图 2-1 视觉显著性介绍示例

### 2.1.2 视觉的分层感知机理

神经科学研究表明，人类的视觉感知是一种分层结构，即信息逐步向更高层次抽象的过程。对于输入的视觉信号，视觉感知从简单的局部细节到复杂的全局语义信息进行分层处理<sup>[50]</sup>。此外，神经科学家发现，提取不同级别视觉信息的视觉皮层区域也存在不同<sup>[51]</sup>，视觉信息主要可分为三类：由V1、V2区域所处理的低级视觉特征、中间颞区和V3、V4区域所提取的中级特征，以及前额和颞下区域处理的高级特征。根据经验，不同程度的失真会导致分层特征的个别退化。深度学习中的卷积神经网络就是借鉴了HVS层次结构的结果，尤其借鉴了低层级中神经节细胞在视网膜内的滤波器的作用进而取得了不错的成果。

在视觉感知过程中，分层处理过程如图2-2所示。首先感知到图像受损的通常是由V1、V2区域所处理的低级视觉特征，该特征着重关注图像中的细节损失。随着损失程度的增加，细节扭曲逐渐导致局部区域发生扭曲，导致观看体验差，而局部显著性特征着重突出了受关注的图像区域存在的受损情况，该类

中级特征在视觉皮层上被中间颞区和V3、V4区域所提取。最终，图像的局部区域扭曲程度增大的影响使得图像的高级全局语义特征遭到破坏，影响图像的整体表达，影响观看者对图像的理解。

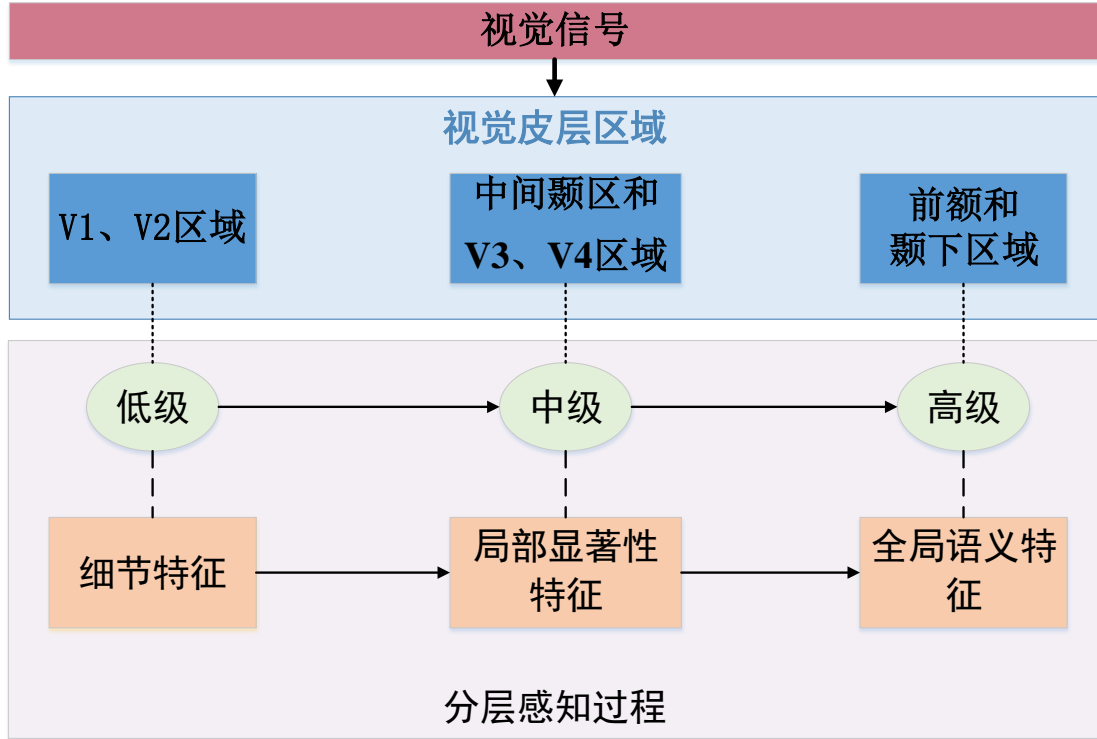


图 2-2 分层感知过程图

## 2.2 全景图像数据集

全景图像因为其更广阔的视野范围和更高的分辨率，提升了观看的用户体验，同时大大提升了用户的满意度。全景图像的展示方式与传统图像有这很大的差异，传统图像的视野范围固定且有限，因此其展示性强。然而在 VR 系统中，全景图像被捕获并以球形形式表示，但是，这种形式不容易被存储或直接传输。因此，运动图像专家组（MPEG）开发了用于编码、存储、传输和呈现全向媒体的全向媒体应用格式（OMAF）<sup>[52]</sup>。在 OMAF 中，以将原始球面信号投影到二维（2D）平面上，以便根据现有编码标准进行压缩和传输<sup>[53]</sup>。详细介绍几个主流的投影格式如下所示。

## (1) 立方体映射投影 (cube map projection, CMP)

CMP 通过透视的形式实现从球面到立方体面的映射，具体的操作是将球体图像与立方体内切，然后将球体投影到立方体的六个平面上（图 2-3 展示了 CMP 投影的方式），实现球体到平面的投影，其实质就是简单的坐标比例缩放。该投影方式的规则简单但是其均匀性较差并且从下图可明显地看出在平面上存在一些未被使用的区域，造成了一定的资源浪费。立方体映射投影结果图如图 2-4 所示。

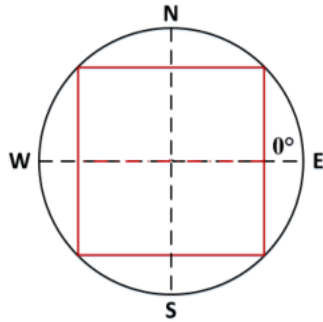


图 2-3 CMP 投影规则

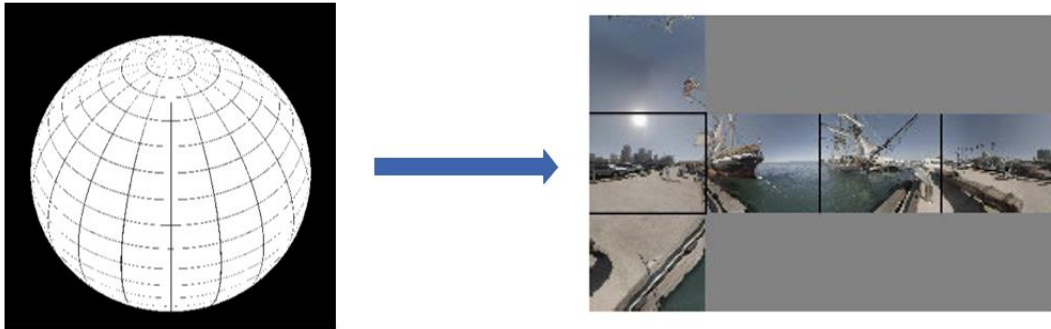


图 2-4 CMP 投影结果图

## (2) 等矩形投影 (Equirectangular projection, ERP)

ERP 因其简单、方便、应用范围广逐渐成为了当前主流的投影方式。在 OMAF 中等矩形投影 (ERP) 也被用作默认投影。其投影规则为在平面宽长比为 1:2 的矩形区域内按照目标分辨率进行均匀的像素格划分，获得球面网格。具体地说，将球面形式图像的经纬线分别投影为平面形式的垂直线和水平线，赤道上的像素投影到平面形式图像的中线上，两极区域投影并延伸到平面图像的上下两个边缘，这就造成了等矩形投影的两级区域因为拉伸而造



成失真程度大的弊端。该投影规则示意图 2-5 如下。

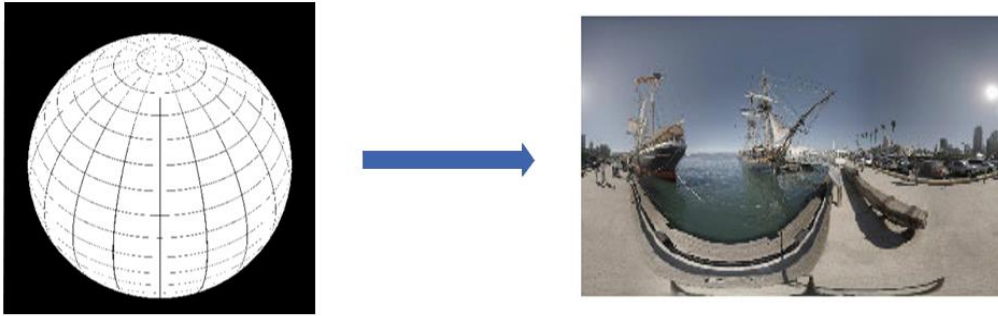


图 2-5 等矩形投影结果图

### (3) 分段球面投影 (Segmented Spherical Projection, SSP)

针对等矩形投影的弊端，SSP 通过将两极区域映射为两个圆来消除 ERP 图像的双极区域中的拉伸失真问题（图 2-6 展示了 SSP 投影的方式）。同时，它在 ERP 格式的赤道区域保留了直观和连续边界的优点。将球面赤道区域的像素按照条带状投影成矩形，而将两级区域的像素投影在圆内的方式大大提高了计算复杂性。

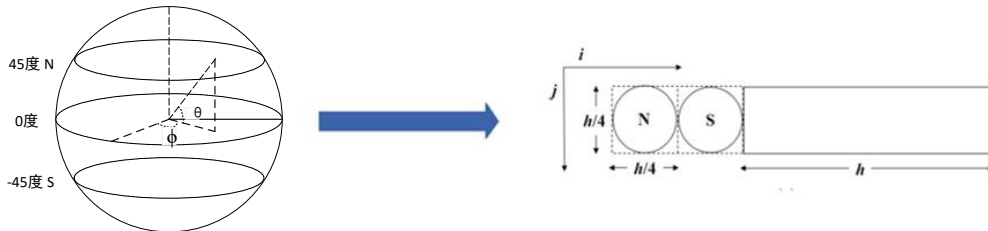


图 2-6 分段球面投影结果图

分段球面投影结果图评估全景图像质量算法的表现性能和可靠性，需要利用数据集来进行训练和测试验证。一个完整的用于客观评价图像质量的数据集需要包含大量的图像数据以及其对应的主观质量评分。目前来说，由于全景图像的发展正处于初级阶段，因此可使用的数据集也相对匮乏并需要进一步的发展和推进。至今，在全景图像质量评价方面没有一个完全统一的数据集系统。本文涉及了两个公开可靠的数据集。每个数据集都包含大量的全景图像（各种类型的失真全景图像以及它们的参考图像）和每个全景图像的主观分数值，详细介绍如下：

(1) 全景图像质量评价数据集 CVIQD2018<sup>[54]</sup>

该数据集提供了由三种失真类型(JPEG、AVC、HEVC)构成的 544 张全向图像, 其中 538 张受损全景图像来源于 16 张参考图像, 其中 12 幅全景图像是通过 Insta360 4K 球形 VR 摄像机拍摄得到的, 其余 4 幅则是从 JVET 的测试视频中获取的, 并且这些图像的分辨率被统一固定为 4096x2048。该数据集使用三种经典的压缩算法: JPEG、AVC 和 HEVC 来生成失真图像。其中受到 JPEG 类型压缩的全景图像以间隔为 5, 从 0 到 50 共 11 个压缩级别的形式展示, 然而在 AVC 和 HEVC 压缩类型下的全景图像以 2 为间隔从 30 设置到 50 共达到 11 个压缩级别的形式存在, 因此该数据集中每一幅原始图像都各自生成了包含三种不同失真类型的 33 个受损图像。通过进行单刺激(SS)主观实验, 提供每张图像的 MOS 值作为基准。下图 2-7 展示了该全景图像数据集的图像示例。

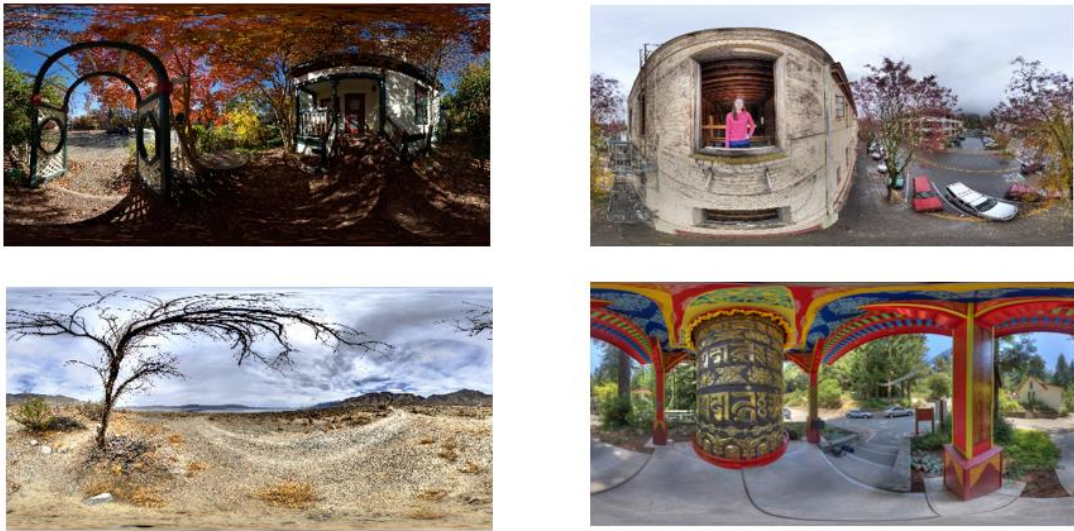


图 2-7 CVIQD2018 数据集图像示例

(2) 全景图像质量评价数据集 OIQA<sup>[55]</sup>

该数据集包含了 320 张失真全景图像和 16 张参考图像, 这些失真图像分为被 JPEG、JPEG2000、高斯白噪声和高斯模糊在 5 个压缩级别下生成的。原始图像是由专业摄影师拍摄得到的。该数据集的所有全景图像的分辨率都在 11332x5666 和 13320x6660 范围之间。除 JPEG、JPEG2000 外, 高斯白噪声和高斯的失真并不是直接在矩形的参考图像的基础上进行压缩的, 而是先对的全景图像进行分割切块, 然后在分割后的图像块上进行单独增加不同的噪声, 最

后再将已有噪声的全景图像块进行拼接得到矩形受损的全景图像。与 CVIQD2018 相似，主观测试采用 SS 方法，并提供主观值(MOS)值该全景图像数据集的示例也被展示在图 2-8 所示。

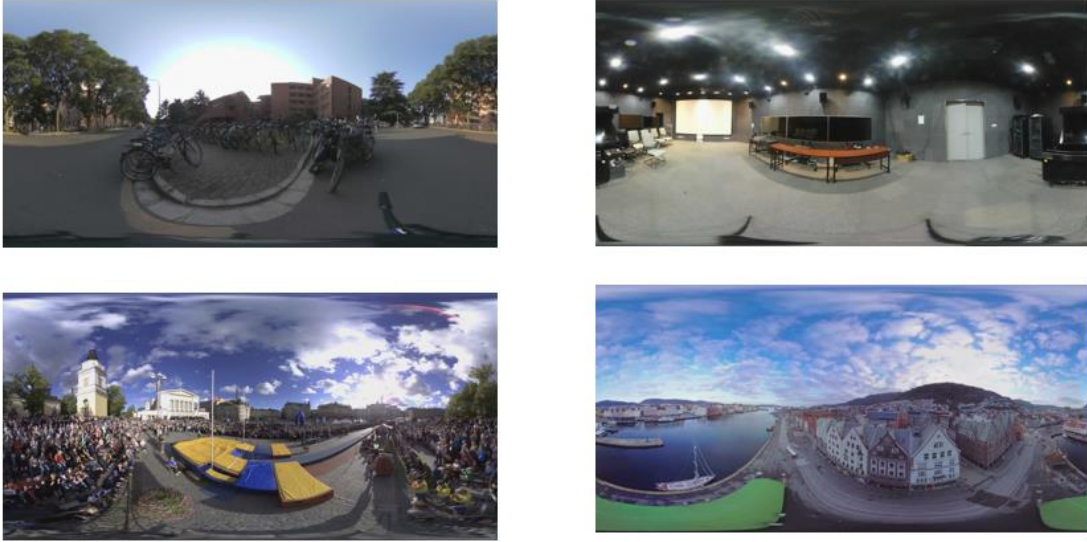


图 2-8 OIQA 数据集图像示例

与 OIQA 数据集相比，CVIQD2018 数据集的样本图像数量相对多，因此它对模型的性能有更好的代表性。相反，OIQA 数据集相比于 CVIQD2018 数据集有更多类型的失真类型，因此 OIQA 数据集有更好的失真多样性，但这也增加了准确评估全景图像质量的难度。

## 2.3 相关性能评价指标

### 2.3.1 显著性图建模方法的性能评价指标

本文使用了 5 种经典的性能指标：ROC、Area Under the Curve(AUC)<sup>[56]</sup>、皮尔逊相关系数 (CC)<sup>[56]</sup>、KL 散度<sup>[56]</sup>和 JS 散度。其中 AUC 表示 ROC 曲线下的面积，ROC (Receiver Operating Characteristic) 曲线通常被人们用于评价模型的预测能力，ROC 曲线表示真正率 (TPR) 和假正率 (FPR) 的比值。因此 ROC 曲线下的面积越大，表示预测性能越好。

CC 用于度量两个变量之间的相关程度，因此 CC 值越大，表示显著性检测

算法越准确。

$$CC = \frac{\sum_{i=1}^N (s_i - \mu_{s_i})(o_i - \mu_{o_i})}{\sqrt{\sum_{i=1}^N (s_i - \mu_{s_i})^2 \sum_{i=1}^N (o_i - \mu_{o_i})^2}} \quad (2.1)$$

其中  $s_i$  和  $o_i$  表示第  $i$  个主观质量值和映射的客观质量值。 $\mu_{s_i}$  和  $\mu_{o_i}$  分别表示  $s_i$  和  $o_i$  的相应平均值。

KL 散度和 JS 散度一般用于度量两种分布之间的距离，表示两种度量的离散程度，其值越大表示，显著性检测算法性能越差。

$$KL(p||q) = \sum p(x) \log \frac{p(x)}{q(x)} \quad (2.2)$$

$$JS(P_1||P_2) = \frac{1}{2} KL(P_1||\frac{P_1+P_2}{2}) + \frac{1}{2} KL(P_2||\frac{P_1+P_2}{2}) \quad (2.3)$$

### 2.3.2 全景图像质量评价方法的性能评价指标

为了可视化模型的性能，视频质量专家组(VQEG)<sup>[57]</sup>提出计算映射的分数和主观值之间的关系以表示模型的性能，包括皮尔逊线性相关系数(PLCC)、斯皮尔曼秩相关系数(SRCC)、肯德尔秩相关系数(KRCC)和均方根误差(RMSE)。这四个具有不同性能的统计指标总结了 IQA 算法的性能。

PLCC 表示精度，其值越大算法的性能越好。PLCC 由下式计算得到：

$$PLCC = \frac{\sum_{i=1}^N (s_i - \mu_{s_i})(o_i - \mu_{o_i})}{\sqrt{\sum_{i=1}^N (s_i - \mu_{s_i})^2 \sum_{i=1}^N (o_i - \mu_{o_i})^2}} \quad (2.4)$$

SRCC 表示预测单调性，其值越大表明算法的性能越好：

$$SRCC = 1 - \frac{6 \sum_{i=1}^N d_i^2}{N(N^2-1)} \quad (2.5)$$

其中  $N$  是图像样本的数量。 $d_i$  表示第  $i$  幅图像的主观评价和客观评价之间的等级差异。

KRCC 表示相关性，其值越大代表算法有更好的性能表现：

$$KRCC = \frac{2(N_c - N_d)}{N(N-1)} \quad (2.6)$$

$N_c$  是数据集中的一致对的个数（也叫作“和谐对”，指变量大小顺序相同的两个样本观测值，即  $x$  等级高低顺序与  $y$  等级高低顺序相同，否则称为“不和谐”或“不一致”）， $N_d$  是数据集中的不一致对个数。

RMSE 表示预测误差，其值越小表示算法性能误差越小集性能越好。

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (s_i - o_i)^2}{N}} \quad (2.7)$$

## 2.4 本章小结

本章首先介绍了 HVS 中的注意力机制和视觉分层感知处理的过程和特点，然后对目前应用较为广泛的全景图像的投影格式进行了简单的介绍；其次，详细介绍了本文使用的两个可靠的全景图像的数据集；最后对本文所涉及到的性能指标进行了简要的概述。通过本章对基于显著性特征的全景图像质量评价相关工作的介绍为后续工作的展开提供了有针对性的依据。



### 第3章 基于视觉感知特性的全景图像显著性建模方法

人类往往每天都需要感受世界上的海量信息，其中，人类视觉系统中的视觉注意力可以让大脑免受大量信息的干扰，具有选择性的挑选图像中最重要的部分，而在观看视觉场景时忽略不重要的部分。本节简单的介绍视觉注意力机制的作用（如图 3-1 示例），当人们看到这张图片时会下意识的把注意力集中在人所在的附近区域，而忽略背景中天空和地面信息。也就是说，在观看这幅图像的时候，并非能够对图片的所有信息给予相同的关注度，而是将注意力着重放在某个局部区域。



图 3-1 视觉显著性注意力机制示例图

计算机视觉中的各种研究试图通过刺激人类视觉系统中的注意力机制来建立视觉注意力的计算模型<sup>[58]</sup>。随着元宇宙概念的诞生，全景图像的应用及发展再次被推向热潮。相比于传统的 2D 图像，全景图像有着更加庞大的视觉信息，就使得人类视觉系统中的显著性机制在处理全景图像的过程中发挥了及其重要的作用。因此迫切需要基于全景图像的特性提出针对性显著性算法，进而提高感知全景图像的效率。

### 3.1 问题描述

人类视觉系统的显著性注意力机制在图像处理中发挥了重要的作用，受此启发，越来越多的研究者们将注意力机制作为处理显著性建模的理论支撑。当图像的前景和背景的颜色信息相差很大时（例如图 3-1 示例），仅通过颜色信息即可准确地检测显著性区域，然而当处理复杂场景时，例如前景和背景信息相似度极大或者前景内容极为复杂时，仅依靠三种颜色通道图像（RGB）信息检测显著性区域的方法无法获得令人满意的结果。

基于上述问题，研究者们引入了空间结构的深度图(Depth)以对颜色通道的信息进行深层信息的补充。这些模型采用了多尺度残差运算<sup>[59]</sup>、协作学习框架<sup>[60]</sup>、跨模态融合<sup>[61]</sup>、多分支主干监督网络<sup>[62]</sup>、挑层卷积神经网络<sup>[63]</sup>等策略。然而这些显著性建模方法仅局限于研究传统的 2D 图像，没有考虑全景图像的发展及其视野广阔信息量大的特性，因而应用到全景图像处理中时无法取得令人满意的结果并且会产生巨大的计算量。

### 3.2 算法思想

考虑到人类视觉特性对显著性区域的信息更加敏感，因此，获取全景图像的显著性区域为大脑感知外界世界提供了一种更加有针对性和更加高效的方式。本章结合全景图像的特点和人类视觉系统的感知特性为全景图像设计了一种显著性分析算法，基于计算复杂度和全景图像的特性，本文引入了超像素的概念处理全景图像的冗余信息，并且结合人类视觉特性对全景图像进行了显著性建模，在一定程度上降低了处理全景图像时的计算复杂度并且提高了显著性检测的准确性，其流程图如图 3-2 所示。

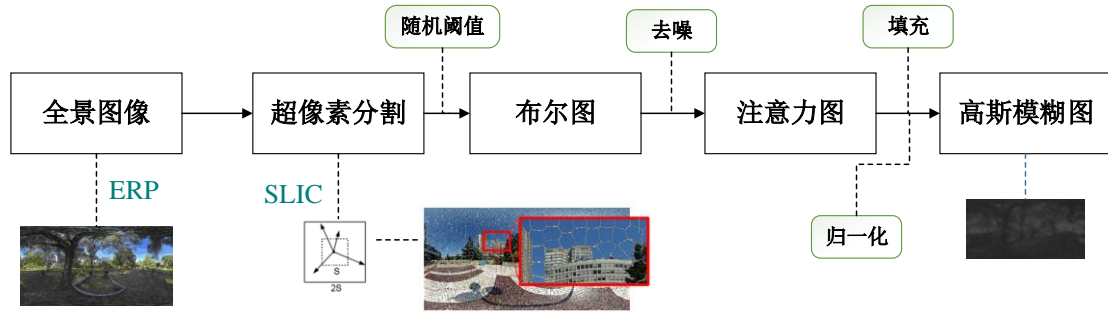


图 3-2 基于视觉感知特性的全景图像显著性建模方法流程

三维世界每个方向的场景内容都被全景图像展示在一个完整的球面上，全景图像的表达方式注定了全景图像拥有更加广阔的视野范围和更加庞大的信息量体系。为了全景图像具有应用广泛性，通常情况下都是将全景图像投影到二维平面中，以平面的形式展示的全景图像包含了视觉视觉场景的整体语义内容。图像的整体语义在显著性建模中扮演了重要的角色，划分显著性区域是以语义为基础，以物体为主体的。另外，像素之间的关系并不是单一的形式，像素与像素之间可能存在着的联系也可能不同像素之间差异的相差巨大。因此分析像素之间的关系对于全景图像语义的处理有种至关重要的作用。

相比于基于像素处理分析全景图像的方法，超像素级方法更能结合像素之间的关系。全景图像相比于传统的 2D 图像有更大的信息量，去除全景图像中像素的冗余信息是显著性建模必不可少的关键一环。将几个相邻的具有一定共性的像素通过一定的计算利用某个信息量表示多个像素的共有特性，为全景图像显著性建模提供更加有用的空间结构，这不仅极大程度上去除了全景图像中的冗余信息，大大提高了全景图像的处理效率，还模拟了大脑处理视觉信息的过程，降低了视觉注意模型的计算复杂度。

格式塔心理学试图理解在一个明显混乱的世界中获得和保持有意义的感知能力背后的规律。根据这一理论，有八个格式塔定律决定了视觉系统如何自动将元素分组为模式，分别为：接近性、相似性、封闭性、对称性、共同命运、连续性以及良好的格式塔和过去的经验。格式塔的封闭性定律是全景图像的显著性建模的一个重要因素。格式塔封闭性定律指出，个体的感知填补了不完整形状的视觉空白。例如，对于一个轮廓上有间隙的图像，人类仍然能够识别出完整的图像内容。显然兴趣点比无法识别图像语义内容的点对人们理解更有用。



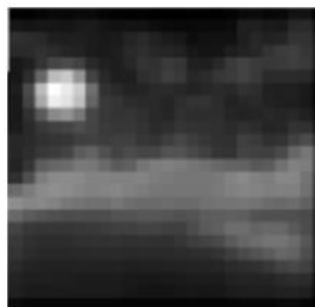
格式塔理论强调研究的整体性，利用整体性理解部分，因此该理论常常被应用于分离图像的前景和后景，进而分析图像中的注意力区域。图像中的信息受多个因素的影响，为了不受其它不利因素的印象，分析全景图像的布尔映射，有效地挖掘了全景图像中的本质信息。另外，对得到的布尔映射处理其中的噪声和不利因素计算得到全景图像的注意力信息，不仅根据图像中的深层信息分析了其显著性区域，还为后续的全景图像的处理降低了复杂度。

### 3.3 显著性图建模方法

全景图像的分辨率通常是传统图像的数几倍，这使得其存储和传输都变得十分困难。然而，人的视觉注意力机制具有一种选择性的注意能力，使得大脑在接收视觉信息时分得清主次，即当人们在面对一个场景时，能够自动地处理其感兴趣区域，而选择性地忽略不感兴趣的区域。因此，本文有必要对全景图像中的信息进行显著性检测，合理地减少全景图像中的视觉冗余信息。检测全景图像的显著性区域，不仅能提高压缩全景图像的效率，减少传输图像时的带宽，而且在编辑全景图像的过程中也起到了至关重要的作用。



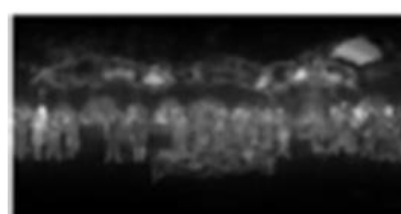
(a) 传统图像



(b) 传统图像的显著性图



(c) 全景图像



(d) 全景图像的显著性图

图 3-3 传统图像和全景图像及其显著性图示例

近年来，针对传统图像的显著性检测技术已经相对成熟，许多优秀的学者针对不同的问题提出了很多检测模型。这些模型同样可以直接应用到全景图像中，但全景图像独特的观看路径和观看方式使得这些应用于传统图像上的显著性检测表现出不理想的性能效果，因此非常需要根据全景图像的特点构建适合全景图像的显著性检测方法。如图 3-3 分别为传统图像及其显著性映射和全景图像及其显著性映射。从图中看出传统图像和全景图像显著性区域的侧重点不同，即传统图像更倾向于注意物体，全景图像更倾向于人们的观看轨迹。

显著性目标检测一般分为自顶向下和自底向上两种方式。其中自顶向下是一种有监督分类学习的目标驱动方式，受到主观意识的支配，学习目标的高层语义特征，通过判别得到指定目标的显著性图，依赖于具体的任务。而自底向上的显著性检测基于图像本身，是利用数据驱动的前注意(pre-attentive)机制从输入图像中提取的不同底层特征，计算与周围具有较强对比度或明显不同的区域，判断目标区域和周围区域的差异，进而计算目标区域的显著性。本文以研究自底向上的显著性目标检测问题。

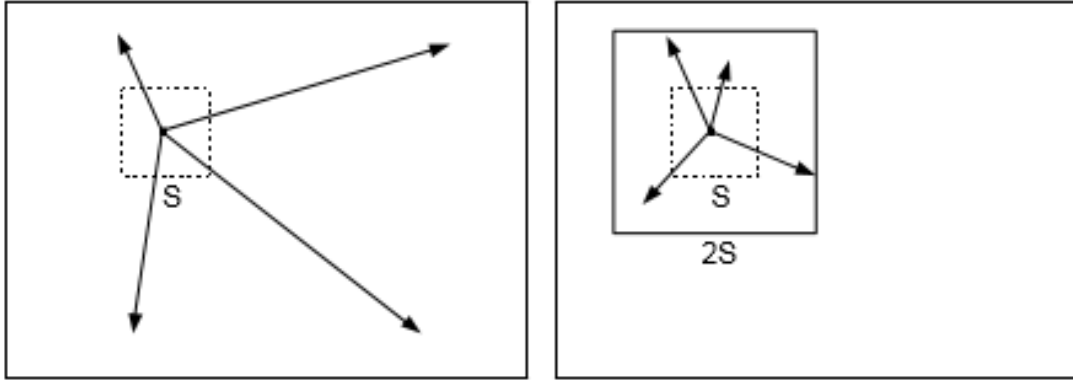
### 3.3.1 全景图像的超像素分割算法

全景图像因为其拥有复杂的信息在处理过程中往往会导致相对于普通图像更高的计算复杂度。为了解决全景图像的冗余信息的问题，本节利用超像素技术以像素组为处理单位对全景图像进行预处理，对全景图像进行小区域的分割成多个子图像。由 Achanta 和 Radhakrishna 提出的简单的线性迭代聚类 (simple linear iterative, SLIC)<sup>[64]</sup>方法因为其计算思想的简单性和实现的方便性被广泛用于 2D 图像中。考虑到该算法的简单有效性，本文利用基于 K 均值聚类方法产生感知局部和边缘的超像素分割算法预处理信息量巨大的全景图像。其唯一的参数为聚类中心  $k$ ，该方法大大降低了像素数的线性复杂度。

超像素分割经常被用于计算图像显著性区域的预处理过程。首先利用像素数  $N$  和唯一参数聚类中心  $k$  计算的到分割的大小均匀的块，该过程通过如下公式计算得到：

$$S = \sqrt{N/k} \quad (3.1)$$

在传统的 K 均值聚类中，每个像素都必须与所有聚类中心进行比较。传统和自适应 K-means 算法的搜索区域如图 3-4 (a) 所示。



(a) 传统的 k 聚类关联

(b) SLIC 算法关联

图 3-4 像素与聚类中心关联示意图

如图 3-4 (b) 所示，SLIC 仅计算从每个聚类中心到  $2S \times 2S$  区域内的像素的距离。期望的超像素大小仅为  $S \times S$ ，由较小的正方形表示。这种方法不仅减少了距离计算，而且使得 SLIC 的复杂性与超像素的数量无关。通过限制搜索区域的大小显著地减少了距离计算的数量，并且导致相对于常规 k-means 聚类的显著的速度优势，其中每个像素必须与所有聚类中心比较。这只能通过引入距离测量  $D$  来实现，该距离测量  $D$  确定每个像素的最近聚类中心。

如果简单地将  $D$  定义为空间中的 5D 欧氏距离将导致出现不同超像素大小的聚类行为不一致的问题。对于大尺寸的超像素，其空间距离大于颜色接近度，因此相比于颜色特性，空间接近度对该超像素更重要，这会产生较为紧凑的超像素，导致这些超像素不能很好地粘附在图像边界上。相反，对于较小的超像素，颜色接近度相比于空间接近度更能发挥作用。为了将这两个距离组结成一个单独的度量，通过聚类内的最大距离  $N_s$  和  $N_c$  来归一化空间相关性和颜色相关性：

$$d_c = \sqrt{(l_j - l_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2} \quad (3.2)$$

$$d_s = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \quad (3.3)$$

$$D' = \sqrt{\left(\frac{d_c}{N_c}\right)^2 + \left(\frac{d_s}{N_s}\right)^2} \quad (3.4)$$

其中，颜色接近度的最大距离 $N_c$ 被定义为常数：

$$N_c = S = \sqrt{\frac{N}{K}} \quad (3.5)$$

在 2D 图像中，SLIC 算法已经被反复验证聚类中心数设置为大于 200 以更好地坚持图像边界。考虑到全景图像的信息量远远大于 2D 图像，因此，本节以 1000 作为最低的聚类中心数以及时间间隔，发现随着聚类数的增加，分割的效果越来越好，但同时也增加了计算复杂度，因此本节选择了在保证分割效果的基础上，复杂度最低的聚类中心数： $K=2000$ 。本文选择展示在三种不同聚类中心数作用下的图像分割效果，如图 3-5 所示。

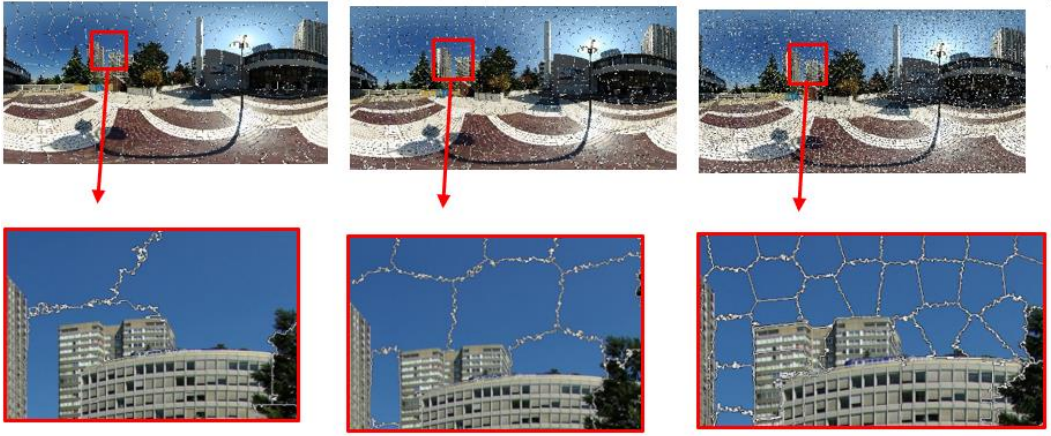


图 3-5 不同聚类中心数作用下的图像分割效果

在图 3-5 中从左到右，聚类中心数分别为 500，1000，2000。从图中可以看出当聚类中心数小于 1000 时，分割超像素过大，其颜色接近度被忽略，图像的语义整体性遭到破坏，因此该聚类中心数不能应用于分割全景图像。当聚类中心数设置为 1000 时，全景图像的边界信息得到了基本的保证，但是对于语义信息没能到达更加精准的效果，这就验证了本节中提到的对于全景图像，聚类中心数设置为大于 1000 才能更好地保护全景图像的边界的整体性并且保留其表达的语义信息。

随着聚类中心数的增加，全景图像的语义差异会被更详细展示，然而其分割效果存在一定程度的冗余，计算复杂度也呈线性增加，因此，在本次工作中将聚类中心数设置为 2000。

### 3.3.2 基于全景超像素图像的布尔映射算法

正如格式塔心理学研究所表明的，人物比背景元素更容易被关注<sup>[65-66]</sup>，人物背景分配可以在没有集中注意力的情况下进行<sup>[67]</sup>。神经科学研究结果表明，猴子和人类大脑中参与形状感知的某些反应严重依赖于图像-背景分配<sup>[68-69]</sup>，这表明这一过程可能在视觉系统的早期开始出现。因此，该部分在分割后的全景图像的基础上利用格式塔理论分离全景图像的前景和后景信息以更有针对性的获取目标图像的显著性区域。在一定的场景下，如下图 3-6 所示，前景图像一般包含了图像中需要突出表达的内容，而背景信息一般相对应于前景物体起到衬托的作用的部分。在一般的摄影规则中，前景区域大多指的是该图像中的主体景象，其对整个图像起到了核心概括主体的作用，而后景信息中的天空和地面大多是为了突出丰富图像的主体部分。当前景和背景相差较大时，图像就非常容易的凸显成人们需要着重注意的对象，例如在拍摄人物图像中，人主体往往被展示在前景中，而其它的风光、环境等信息皆为背景信息被用于衬托人物。



图 3-6 图像前景-后景介绍示意图

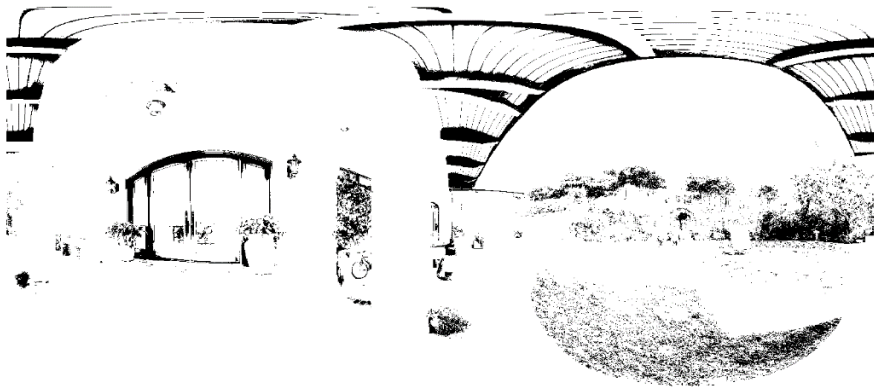
另外，本文引入布尔图的概念并利用其特性分析全景图像的显著性建模方法。简单的来说，布尔图是指利用布尔运算处理后的图像，布尔运算本质上是为了检测真假值，因此其结果只有真-假两种情况，通常来说，人们利用 1 和 0 来代表运算的真假。利用布尔运算处理全景图像的过程可以简单地表示为判断全景图像每个像素的真假性，其判断的依据可以引入阈值的概念，阈值是指引导一个事件产生的最高临界值，因此海量数据可以根据阈值简单的分为两类。



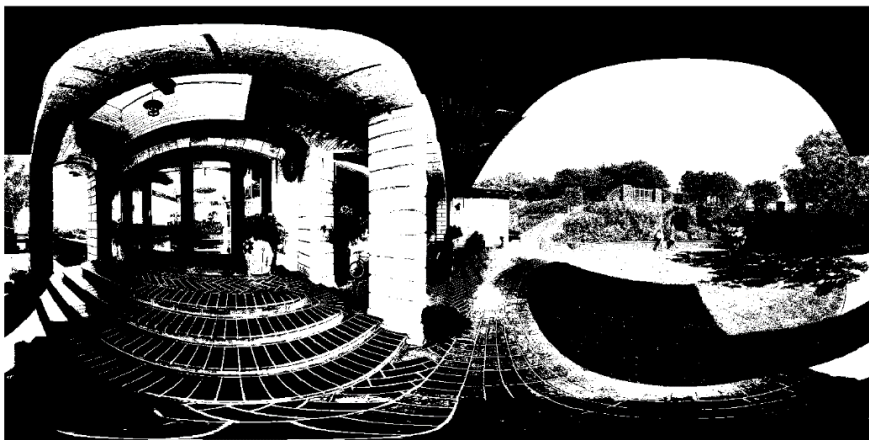
不同的阈值会产生不同的分类，因此设置不同阈值即可获得不同的布尔图。本节通过图 3-7 展示了在全景图像不同阈值作用下的布尔映射。



(a) 全景图像



(b) 低阈值的布尔图 (0.21)



(c) 中阈值的布尔图 (0.49)



(d) 高阈值的布尔图 (0.79)

图 3-7 不同阈值下的布尔图

在该部分工作中，通过对输入的分割后的全景图像设置随机阈值的方式获取一系列的布尔图  $B = \{B_1, B_2, \dots, B_n\}$ ，最主要的思想是通过分析布尔图的拓扑结构来计算显著性图的过程，这主要考虑到拓扑结构信息是尺度不变的，对视觉注意力有很大的影响。由如下公式表示：

$$B_i = \text{Thresh}(I', \theta), i \in n \quad (3.6)$$

其中， $I'$ 表示分割后的全景图像， $\theta$ 表示设置的阈值，在本工作中，阈值的设置根据分割后全景图像的方差特性设定，方差能准确表示图像的对比度，进而使自底向上的显著性检测方法更好的表达图像的语义像素的分布特性。阈值的范围为： $[-60, 99]$ ，步长为：5。

### 3.3.3 基于注意力图的显著性建模算法

基于 3.3.2 中的算法，本文得到了若干布尔映射，对每个布尔映射，本文利用带有核  $\omega_0$  的开放操作来去除每个布尔映射中存在的噪声，公式如下：

$$B_i \circ \omega_0 = (B_i \ominus \omega_0) \oplus \omega_0 \quad (3.7)$$

其中  $\circ$  表示开放操作， $\oplus$  表示膨胀运算， $\ominus$  表示腐蚀运算。 $\omega_0$  核设置为 5。

格式塔心理学研究表明，例如尺寸、包围性、凸性和对称性等因素可能会影响图像-背景隔离<sup>[70]</sup>。包围性的本质是图像和地面之间的包围拓扑关系。为了测量包围性，本文通过一组布尔映射来表征图像。本节利用边界跟踪的方式处

理数字化二值图像，边界追踪方法通过利用二值化图像的像素值分析二值图像的边界，它从 1-连通域和 0-连通域之间的边界上导出系列坐标点，其中 1-连通域和 0-连通域分别是由 1-像素，0-像素连接组成的。如果 0-连通域  $B$  包含图片的边框，就把  $B$  称为背景，否则就是一个孔洞。该方法由于其应用广泛性被深入研究，其中包括了图像处理中关于图像的识别、分析和数据压缩等。

在该工作中，通过二值图像处理技术有效地计算注意力图，以激活具有封闭外部的区域。原则上被包围的区域是更可能被感知为图像的<sup>[70]</sup>。包围性布尔映射中的包围性可以被很好地定义为具有闭合外部轮廓的连接区域（值为 1 或 0）的属性。根据此定义，只有连接到图像边界的区域不被包围。为了计算注意力图，本工作将 1 分配给被包围区域的并集，将 0 分配给图的其余部分。该操作可以通过使用“泛洪填充”算法来实现。该算法随机选中某个点，对其附近的像素点进行填充，直到封闭区域内的像素都被填充成相同的颜色为止，这有效地屏蔽了连接到图像边界的所有像素。

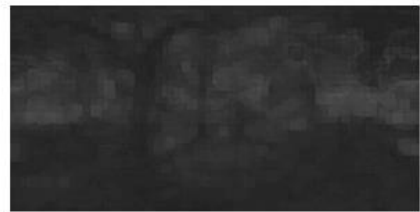
在线性组合步骤之前，需要对生成的注意力图进行归一化，以便具有小的集中活动区域的注意力图将得到更多的强调。对于眼睛注视预测，本工作使用简单的 L2 归一化，即将矢量化地图除以其 L2 范数，以强调具有小活动区域的注意力地图。与 L1 标准化相比，L2 标准化对具有极小活动区域的注意力图不太敏感，否则将主导融合过程。L2 范数归一化处理的实质是对向量  $Y1$  的所有维度数据  $y1, y2, \dots, yn$  都除以  $\|Y1\|^2$ ，以得到一个新向量  $Y2$  的过程，即：

$$Y2 = (\frac{y1}{\|Y1\|^2}, \frac{y2}{\|Y1\|^2}, \dots, \frac{yn}{\|Y1\|^2}) \quad (3.8)$$

$$\|Y1\|^2 = \sqrt{y1^2 + y2^2 + \dots + yn^2} \quad (3.9)$$



(a) 输入图像



(b) 高斯模糊图

图 3-8 输入图像与其高斯模糊图



为了进一步惩罚具有小的分散活动区域的注意力图，本节在归一化之前扩展了具有核宽度 $\omega_1=7$ 的注意力图。最终，通过本方法计算得到的显著性映射如图 3-8 所示。

### 3.4 算法的实现流程

本文介绍的显著性建模方法考虑了全景图像特殊的观看路径和观看方式，解决了全景图像由于信息量大而导致的计算复杂度高的问题，准确地突出了人眼观看全景图像时的重点关注区域。本章的主要实现流程算法 1 所示：

---

#### 算法 1 显著性图建模

---

**输入：**

全景图像  $I$

**输出：**

全景图像的显著性图  $S$

```

1: for 每个颜色通道图  $\{I_k: k=1,2,3\}$  in Lab 颜色空间
2:    $I'_k = \text{SLIC}(I_k)$ 
3:   for  $\theta = -60:\delta:255$ 
4:      $B_i = \text{THRESH}(I'_k, \theta)$ 
5:      $B_i = \text{OPENING}(B_i, \omega_0)$ 
6:     if  $B_i$  是连通区域
7:        $A_i = 1$ 
8:        $A_i = \text{DILATION}(A_i, \omega_1)$ 
9:        $A_i = \text{NORMALIZE}(A_i)$ 
10:  $S = \frac{1}{n} \sum_{i=1}^n A_i$ 
11: return  $S$ 

```

---

算法 1 中的第 2 行，SLIC 表示 3.3.1 节中使用的超像素分割技术，其聚类中心依据全景图像的特点、分割效果和计算复杂度设置为 2000；第三行的  $\theta$  表示 3.3.2 节中设置的阈值， $\delta$  表示步长，在本文中步长取值为 5；第 4 行表示的是阈值函数；算法 1 的第 5 行利用带有核 $\omega_0=5$ 的开放操作来去除每个布尔映射中存在的噪声；6-7 行操作在本章中使用了“泛洪填充”算法来实现；第 8 行利用的是具有核宽度 $\omega_1=7$ 的闭运算；在本章节中 NORMALIZE 函数使用的是 L2 范数归一化以强调具有小的集中活动区域的注意力图。最终  $S$  表示输出的全景图像的显著性图。

### 3.5 实验结果与分析

为了验证本文使用的计算全景图像显著性检测方法的性能，本节结合三种应用广泛的、经典的显著性检测方法进行综合比较，本文的运行环境是在 Intel(R)Core(TM)i5 CPU, 16GB RAM, 64 位系统 PC 端上进行的。比较的算法既有 2D 图像的算法也包括了全景图像的显著性检测算法，参与比较的算法包括 SP<sup>[12]</sup>、BMS<sup>[71]</sup>、360-saliency<sup>[15]</sup>，其中前两者为经典的 2 维图像显著性检测方法，最后一个则为性能表现优异的全景图像显著性检测方法。

本文使用的是 5 种经典的性能指标：ROC、Area Under the Curve(AUC)、皮尔逊相关系数(CC)、KL 散度和 JS 散度。

AUC 表示 ROC 曲线下的面积，ROC (Receiver Operating Characteristic) 曲线被广泛用于显著性检测模型的性能评估。显著性检测模型计算的显著性图可以分为显著点和非显著点，来自眼睛跟踪数据的注视图包括目标点和背景点。真正率 (TPR) 被定义为从计算显著性检测模型中落入显著点的目标点的百分比，而假正率 (FPR) 则被定义为落入显著点中的背景点的百分比。ROC 曲线通过可视化真正率 (TPR) 和假正率 (FPR) 的比值表达显著性检测模型的性能。具体而言，TPR 和 FPR 定义为：

$$TPR = \frac{TP}{TP+FN} \quad (3.10)$$

$$FPR = \frac{FP}{TN+FP} \quad (3.11)$$

其中，TP, FP, TN 和 FN 分别表示真正例、假正例、真反例和假反例，并且 TP+FP+TN+FN=样例总数。因此 ROC 曲线越高，显著性模型对显著性预测的预测性能越好。ROC 曲线下面积 (AUC) 可用于定量测量显著性表现。AUC 值越大，显著性预测的性能越优异。

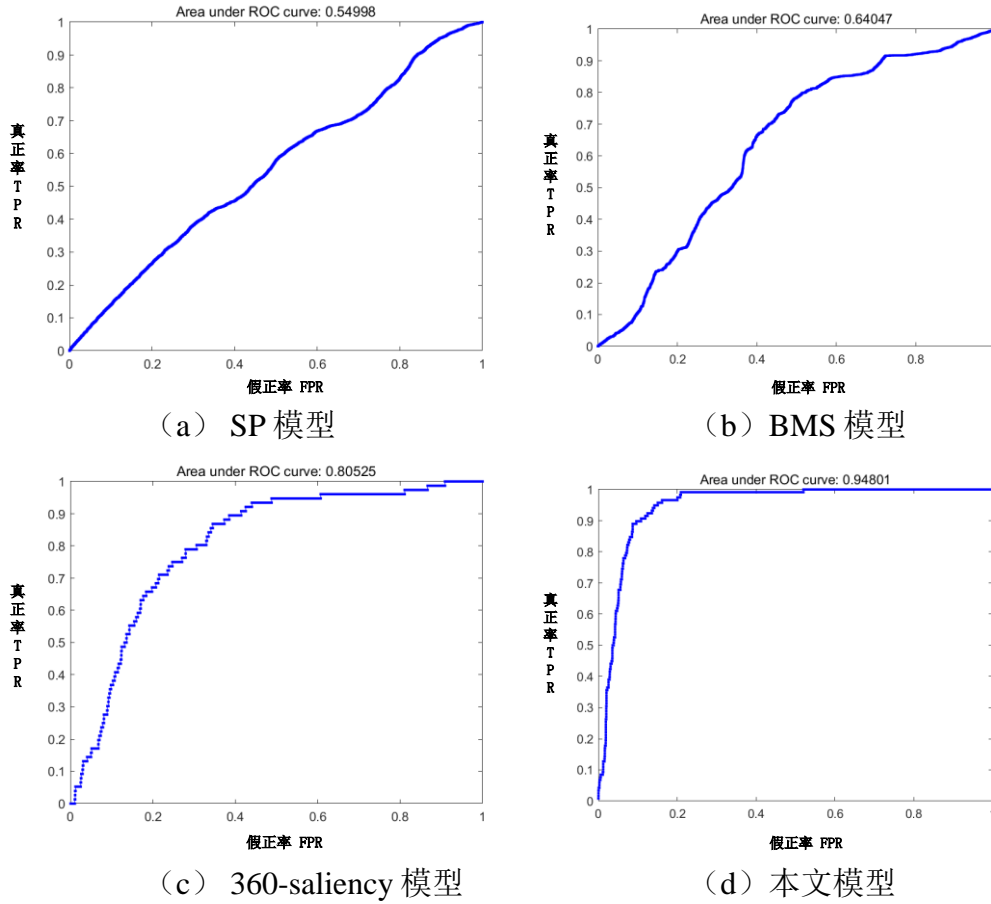


图 3-9 四种不同检测算法在 AUC 指标下的性能对比

比较模型的 ROC 曲线如图 3-9 所示，用于衡量所提出的显著性检测模型的性能。从这个图中我们可以看到，在比较的模型中，本文所提出的模型相比于不考虑全景图像特性的基于 2D 显著性建模方法表现出了更好的性能，这也验证了全景图像与传统图像的差异性以及全景图像特性的不可忽略性。然而本文的算法相比于 360-saliency 算法没有达到更令人满意的效果。

CC（如式 2.1 所示）提供了显著性图和注视图之间的线性相关性程度的度量。在本实验中，本工作使用两个变量之间的 Pearson 相关系数来计算显著性检测模型的 CC。它被定义为显著图和注视图的协方差除以它们的标准差的乘积，因此 CC 值越大，表示显著性检测算法越准确。

KL 散度（如式 2.2 所示）和 JS 散度（如式 2.3 所示）一般用于度量两种分布之间的距离，表示两种度量的离散程度。其值越大表示，显著性检测算法性能越差。

表 3-1 四种不同显著性检测算法在 4 种指标下的性能对比

评估算法	ROC	CC	KL	JS
SP	0.5500	0.0026	14.2296	0.6912
BMS	0.6405	0.0029	10.6879	0.6910
360-saliency	0.9480	0.3277	1.9811	0.4466
本文算法	<b>0.8053</b>	<b>0.1995</b>	<b>2.6725</b>	<b>0.5536</b>

如表 3-1 所示, 本文提出的显著性检测方法的性能优于基于 2 维图像提出的显著性算法。具体地, 经典的 SP 算法在全景图像的显著性检测表现出较差的性能, 这主要是该算法没有考虑全景图像的特性。其次 BMS 算法在直接应用到全景图像时表现出了相对于 SP 算法更好的性能, 但是在 CC 和 KL 指标上表现性能不足以较为准确地检测全景图像的显著性区域。尽管 360-saliency 显著性检测算法排名第一, 但是该算法并没有考虑计算的复杂度, 这大大提升了该工作的计算难度。

另外, 本节分别在表 3-2 和表 3-3 中展示了不同聚类中心数的分割方案对显著性建模方法性能的影响及其计算复杂度。增加聚类中心的数量, 显著性建模方法的性能在 CC、KL、JS 指标上得到了优化, 在 ROC 指标上略微降低, 但其计算复杂度了显著增大。本文利用平均处理一幅全景图像的时间作为衡量显著性建模计算复杂度的标准。

表 3-2 不同聚类中心数下的显著性检测算法的性能对比

聚类中心数	ROC	CC	KL	JS
500	0.8061	0.1940	2.6833	0.5549
1000	0.8086	0.1976	2.6753	0.5541
<b>2000</b>	<b>0.8053</b>	<b>0.1995</b>	<b>2.6725</b>	<b>0.5536</b>
3000	0.8021	0.1989	2.6722	0.5536

表 3-3 不同聚类中心数下的显著性检测算法的计算复杂度对比

聚类中心数	时间 (s)
1000	28.80
<b>2000</b>	<b>30.35</b>
3000	40.71

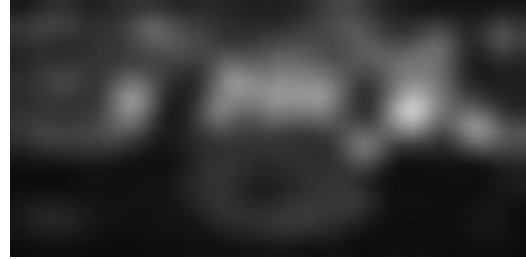
为了保证人们在显著性建模方法性能时视觉体验,本节在图 3-10 中展示了 SP、BMS、360-saliency 和本章算法的显著性图。其中 SP 算法的显著性图在视觉上的体验极差;相比 SP 算法,BMS 算法的显著性图展示了全景图像的模糊的大体轮廓;360-saliency 算法的着重突出了全景图像的赤道区域;本章算法的显著性图相比于其它的算法来说具有略强的视觉体验。



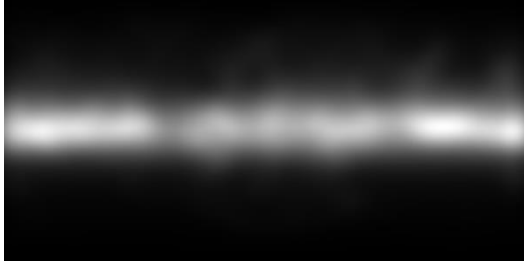
(a) 全景图像



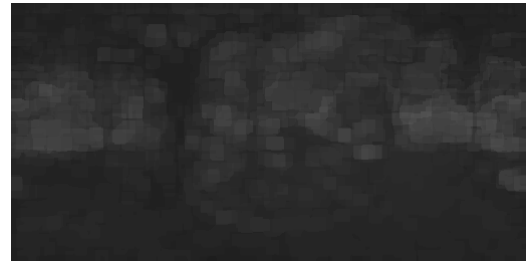
(b) SP



(c) BMS



(d) 360-saliency



(e) 本文算法

图 3-10 全景图像在 4 种建模方法计算下的显著性图

总的来说,本文的显著性检测算法在所有性能指标上排名第二,但无论是哪个性能指标都略低于排名第一的 360-saliency 算法,并且性能远其它两种算法表现优异,这意味着本章的算法不仅能较为准确的检测出全景图像的显著性区域以满足本工作的需求,并降低了检测注意力区域的计算复杂度。

### 3.6 本章小结

本章根据全景图像的特点和人类视觉系统的注意力机制设计了一种基于全景图像的显著性检测算法。首先在对于全景图像利用超像素分割技术去除全景图像中存在的大量冗余信息，然后结合格式塔理论计算一系列的布尔映射进而分离全景图像的前景-后景信息；进一步地，考虑到布尔映射中的拓扑结构具有不变性进而计算全景图像中的显著性区域，得到全景图像的显著性映射。本文在 5 种性能指标下展示了本文算法的表现性能，实验结果显示本文设计的方法可以达到比较好的性能并能准确地检测出全景图像的显著性区域。

## 第4章 基于多维视觉特征的全景图像质量评价方法

随着虚拟现实（VR）设备的逐渐问世以及 5G 的发展，给人带来了沉浸式用户体验，具有身临其境的特点，人们也慢慢地了解到全景的概念。元宇宙概念的诞生，全景图像再次被推向热潮。相比于普通场景，全景图像有着更加广阔的视野范围。热力学之父曾经指出：“如果人们不能衡量它，人们就不能提升它。”因此，衡量全景图像的质量，得到优化的指标，才能提升用户的体验。本文基于上一章提出的针对全景图像的显著性检测算法进一步地综合其它三种算法提取的特征衡量全景图像的质量好坏问题。一般来说，人们在接受到视觉信息时，都倾向于看到清晰明亮色彩鲜明的图像，这会提升人的观感，使人心情愉悦，让人赏心悦目，但是往往有时候看到的视觉效果并不令人满意，例如下图 4-1 所示的模糊图像，尤其是在注意力区域的模糊对人的影响更大，直接性地降低了人们的观看体验。



图 4-1 图像模糊示例

### 4.1 问题描述

全景图像在采集、存储、传输、压缩以及处理中经常需要进行展示格式的转换。因此不可避免的会产生噪声、压缩损失和拉伸，尤其是在全景图像的双极区域，存在严重的拉伸现象。这往往会影响人们的观看体验甚至影响人们对图像的理解。因此衡量全景图像的质量是未来全景图像发展必不可少的重要一环，也为后续全景图像的处理提供了关键的指标。在质量评价的过程中，提取的视觉感知特征对于是否能准确评估全景图像的质量起着至关重要的作用，另

外，因为人是最终的质量评价者，因此挖掘贴近人类视觉特性的特征是本文研究的重点。目前，全景图像质量评价模型存在以下问题：

（1）主观的全景图像质量评价方法利用人直接对全景图像进行观看并做出评价，该方法虽评价准确，但十分耗时耗力。并且全景图像的客观评价方法中的全参考和半参考评价方法没有考虑到现实情况中参考图像不可获取的问题。

（2）全景图像直接应用传统的评价方法虽然计算复杂度低，但没有考虑全景图像的观看特点和观看轨迹。

（3）基于全景图像的特点提出的全景图像质量评价方法通常忽略了全景图像信息量大而造成了评价模型计算复杂度高。

（4）全景图像由于其观看方式的特殊性，观看者在某一时间点所能看到的视野范围有一定的限制，并且观看者具有很高的观看行为一致性，然而在评价模型中经常忽略其显著性特点对质量评价的关键作用。

（5）全景图像的颜色信息经常在评价模型中被忽视，而当前对于颜色信息的研究仅存在于挖掘单通道颜色信息的特点，没有考虑不同颜色通道之间的相关性和依赖性。

## 4.2 算法思想和流程

一般来说，大多数能有效感知全景图像质量的视觉特征都是基于人类视觉系统的，这与人类大脑的视觉信息保持高度一致。神经科学研究表明人类视觉感知是一个分层结构并且人脑的视觉神经元在提取不同的视觉信息时是分区域运行的<sup>[50]</sup>。视觉信息在大脑处理过程中大体上可以被分为三种不同的策略。受人类视觉系统分层感知的启发可以将视觉质量下降描述为一个分层过程，即输入的视觉信号从局部细节逐步过渡到全局语义。为了更好的感知全景图像的质量，本章模拟了人类视觉系统的感知过程。

从经验来看，不同程度的扭曲会导致个体对等级特征的退化。因此，在本章提出的模型中，首先将层次特征指定为细节特征、局部显著特征和全局特征。如下图 4-2 所示，首先衡量最先受到影响的细节信息，然后利用第三章中的显著性建模方法检测全景图像的显著性区域，衡量重要的局部损失，最后，本文分析了全局信息对全景图像的整体语义表达。另外本章还考虑了颜色信息对全景图像的影响，因为颜色信息经常被研究者所忽视，然而对全景图像的质量有



不可忽视的作用。值得一提的是，提取图像的哪些特征是这类模型的一个关键阶段，具体地说提取的特征越符合人类视觉特性，评价模型就越能达到更好的性能。本章通过提取不同的全景图像特征并使用 SVR 技术评估图像质量，算法流程如图 4-2 所示。

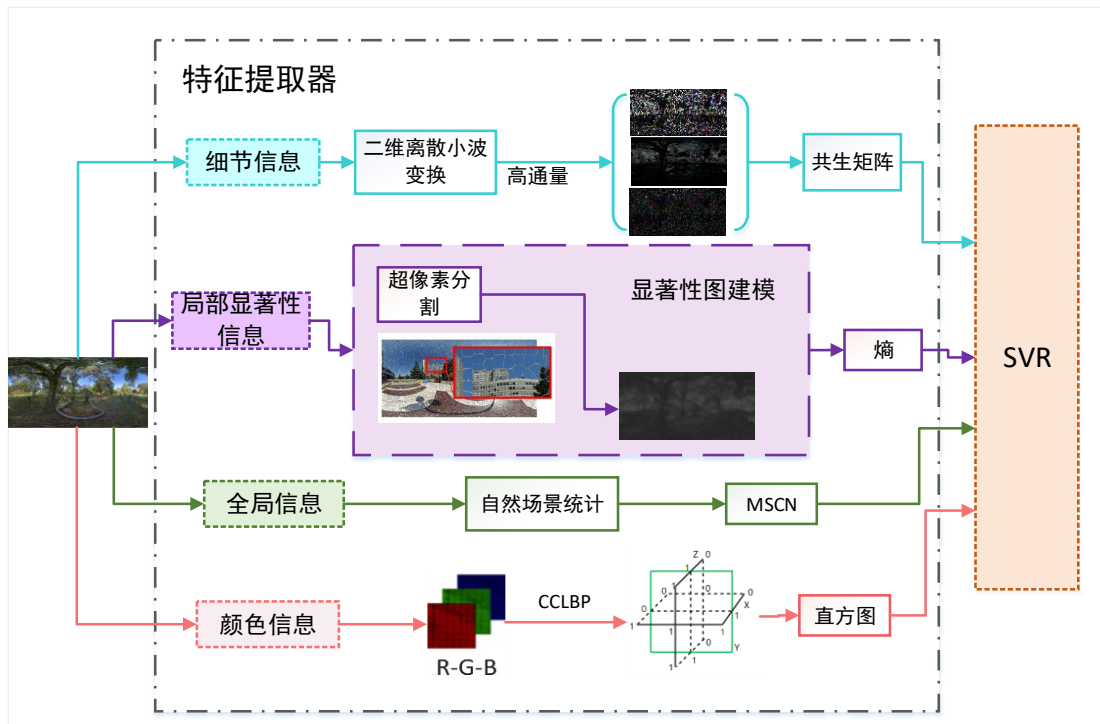


图 4-2 基于视觉显著性特性的全景图像质量评价方法流程图

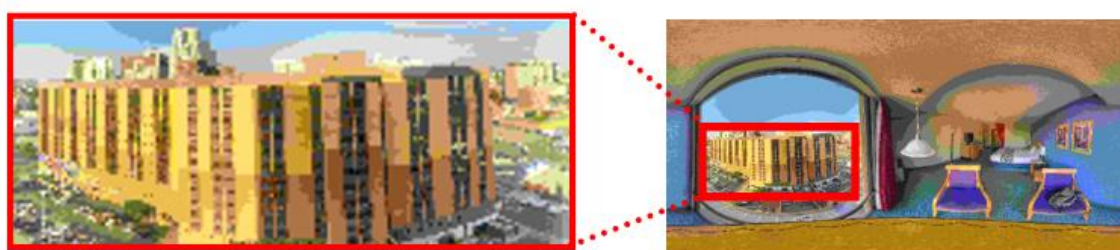
### 4.3 全景图像视觉感知特征提取

#### 4.3.1 基于灰度共现矩阵的频域细节特征

根据神经科学中的层次感知过程，图像首先受到噪声影响的是一些细节信息，这启发了本工作对全景图像中细节损失的探索。在空间域中，图像的细节损失往往不易被肉眼察觉，也就是说人们对空间域中的细节损失并不敏感。如下图 4-3 所示，可以清楚地观察到细节受损的图像与原始图像整体上视觉效果的区别并不大，但是从左侧放大的图像中中可以明显地看出图像受到的细节损失。



(a) 原始图像



(b) 受损图像

图 4-3 原始图像与受损图像在空间域上的差异对比图

然而，图像中的细节损失在频率域中会被轻易地察觉到，因此，为了模拟人类视觉系统机制，本文采用了具有高低频分解功能的标准正交离散小波变换(DWT)将全景图像转换到频域<sup>[72]</sup>，其中，分解得到的高频部分（高通带）充分包含了损坏的全景图像的细节<sup>[73]</sup>。对于二维的图像来说，经过一次离散小波变换后图像会被分解为 4 幅图像分别为是原始图像的低频分量图如图 4-4 分解模型的左上角所表示部分，其左下角、右上角和右下角分别为原始图像的在垂直、水平和对角不同方向上的高频分量。

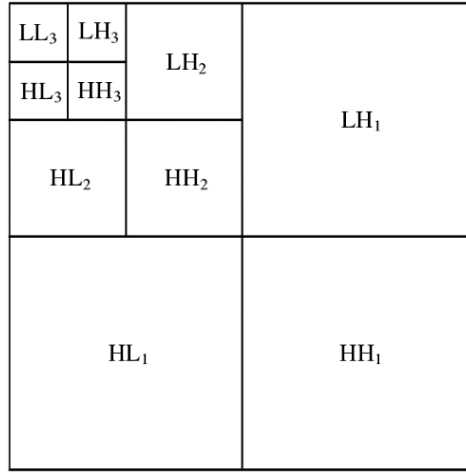


图 4-4 DWT 变换子带图

图 4-5 展示了 DWT 的具体分解过程，其中， $x[m,n]$  表示输入的离散信号，在本工作中代表输入的全景图像。 $g[]$  代表了低通滤波器（low pass filter），其作用为过滤输入的信号中的高频部分而输出低频部分。 $h[]$  是可以过滤信号中低频信息而保留高频信息的高通滤波器（high pass filter）。 $\downarrow N$  代表降采样滤波器，其中本文将降采样滤波器  $N$  设置为 2。

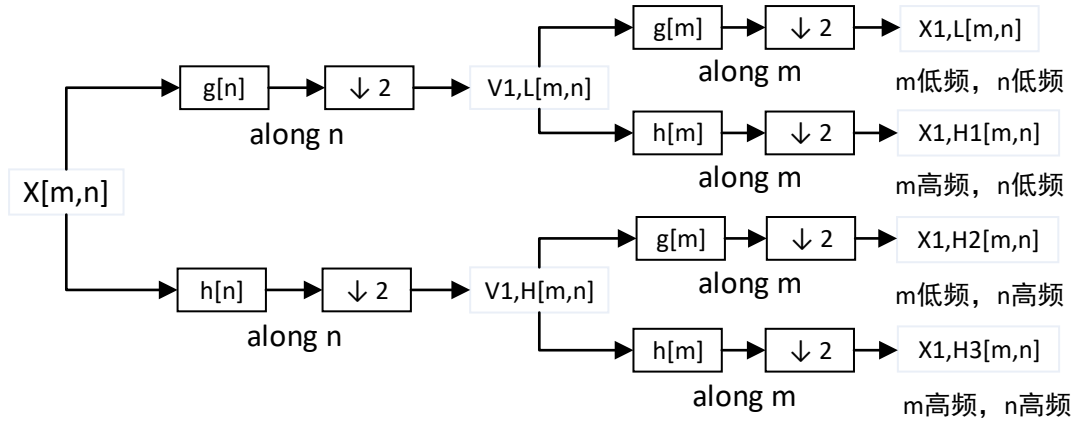


图 4-5 DWT 实现过程

与传统的特征提取不同，该算法采用高频共现矩阵<sup>[74]</sup>来强调特征，能够准确捕捉高频图像中相邻像素的相关性。本文从四个不同的方向（水平、垂直、主对角和副对角）计算 8 维的高频共现矩阵，并且使用两个统计量来表示全景图像中的细节特征，分别为具有衡量图像纹理的是否均匀分布作用的能量  $I_e$  和能图像纹理分布密度的同质性  $I_m$ ，并定义为如下公式：

$$I_e = \sum_{i=1}^N \sum_{j=1}^N p_{d,\delta}(i,j)^2 \quad (4.1)$$

其中,  $d$  为矩阵的计算距离,  $\delta$  表示矩阵的方向参数,  $p$  为  $G$  函数的归一化值, 计算公式为:

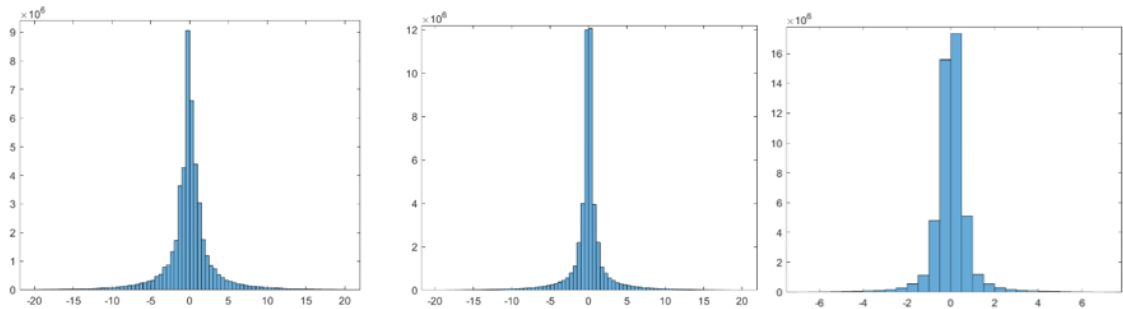
$$p_{d,\delta}(i,j) = \frac{g_{d,\delta}(i,j)}{\sum_{i=1}^N \sum_{j=1}^N g_{d,\delta}(i,j)} \quad (4.2)$$

$g_{d,\delta}(i,j)$  为像素对的共现矩阵值。

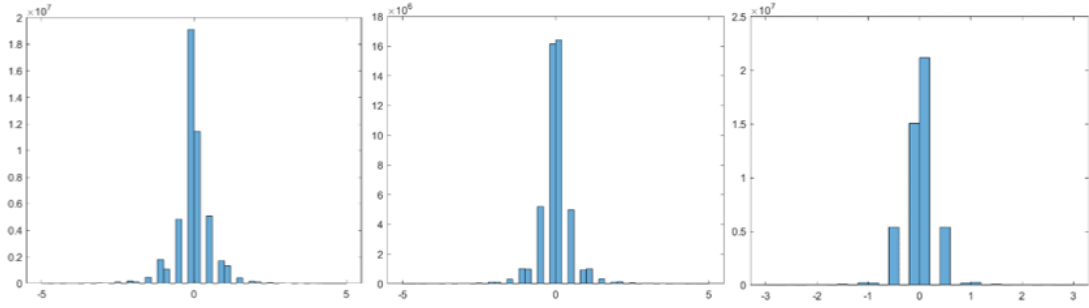
$$I_m = \sum_{i=1}^N \sum_{j=1}^N \frac{p_{d,\delta}(i,j)}{1+|i-j|^2} \quad (4.3)$$

此外, 高频图像的方差也是一种能够精确捕捉全景图像中失真细节的有效工具。因此, 具有反映图像高频部分大小作用的方差也被提取为表示全景图像细节信息的一部分。

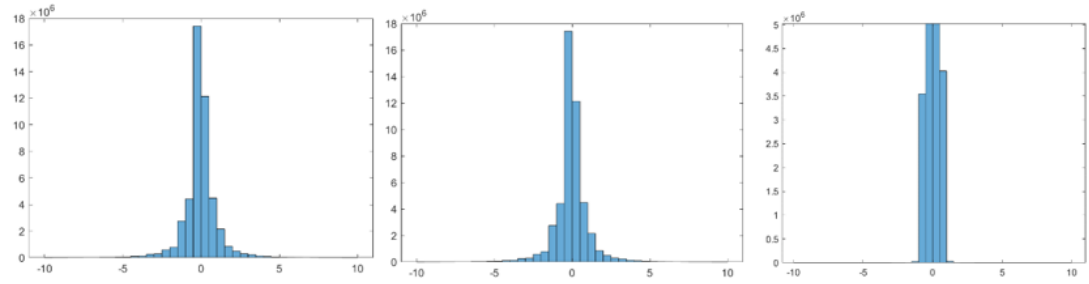
为了验证本节方法的有效性, 本工作在图 4-6 中展示了基于参考图像和 JPEG、JPEG2K、WN 和 BLUR 四种不同类型损坏全景图像的高频分量直方图, 其重点分析了图像不同的像素值的分布情况。横坐标代表的是图像的像素值, 纵坐标表示的是相同像素值在该图像中出现的次数。该图证明了在评估全景图像质量方面的有效性。其中, (a) 为原始图像高频分量的直方图, (b) 为 JPEG 图像高频分量的直方图, (c) 为 JPEG2K 图像高频分量的直方图, (d) 为 WN 图像高频分量的直方图, (e) 为 BLUR 图像高频分量的直方图, 从左到右分别为水平、垂直和对角方向。最终本章从全景图像分解到的高频图中提取出 27 个维度特征 ((4 个方向 \* 2 个统计量 + 1 个方差) \* 3 个方向) 作为全景图像的细节特征。



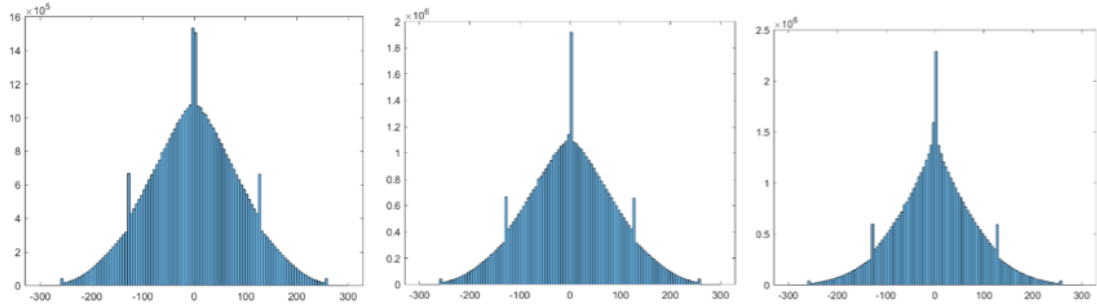
(a) 原始图像



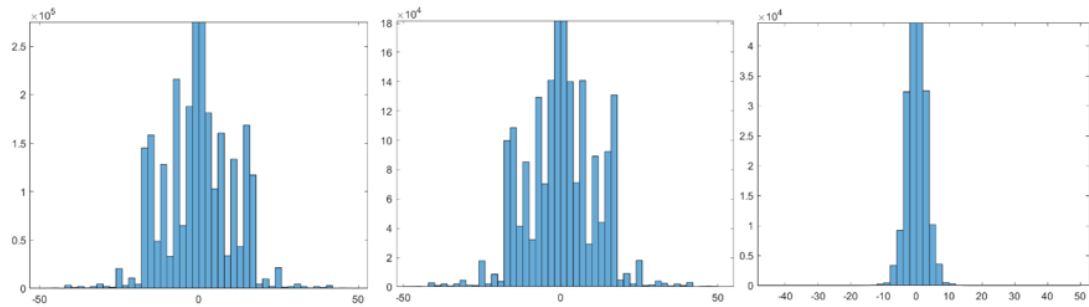
(b) JPEG 图像



(c) JPEG2K 图像



(d) WN 图像



(e) BLUR 图像

图 4-6 原始图像和不同受损类型下高频图像的直方图

### 4.3.2 基于信息熵的局部显著性特征

在本节中，基于上一章中通过超像素分割技术和显著性检测算法获得了包含全景图像视觉显著性区域的显著性映射，分析了在细节受损后全景图像逐步遭受到损失的局部显著性信息并作为分层处理中的中级特征。由于在全景图像质量评价中，熵能显著地感知图像退化的类型和程度<sup>[75-76]</sup>，而显著映射中包含了图像中重要的视觉信息，因此本章挖掘基于熵的显著信息来提供本节局部显著性特征，计算公式如下：

$$E(i) = \sum_{j \in i} p(j) \log_2 p(j) \quad (4.4)$$

其中 $j$ 是显著图 $i$ 中的像素， $p(j)$ 表示经验概率密度。最后，本文共提取 1 维特征作为局部显著性信息。

### 4.3.3 基于自然场景统计特性的全局语义特征

根据 HVS 的理论层次感知过程，随着失真程度的增加，作为高级特征的全局语义会遭到破坏，迫使图像变得不自然，破坏了图像的整体表达。自然场景统计（Natural Scene Statistics, NSS）<sup>[77]</sup>特征是图像中最本质的属性，可以敏感的感知全景图像中的“自然性”，图像的自然性特征是有效感知图像整体质量表达的关键因素，它可以有效地代表全景图像的全局语义。

自然图像（如图 4-7（a））并非人们印象中的以自然环境为主题的自然图像，而是指任何通过光学相机捕捉到的并且尚未经过人工处理的图像，最近网络上比较火的 AI 绘图（如图 4-7（b）人工合成图像）就是完全相反于自然图像的存在图 4-7 中展示了自然图像和 AI 图像。自然图像与非自然图像的本质区别是其具有一定规则的统计特性，当图像遭受的噪声或压缩损失时，自然图像中特有的统计特性会发生不同类型和程度改变。



(a) 自然图像



(b) 人工合成图像

图 4-7 自然图像和人工图像示例

由于全景图像也是由光学相机拍摄得到的，因此其同样具有自然图像有规则的统计特性。大多数真实世界的失真过程会扰乱一些统计数据，并产生不自然的全向图像。因此，研究全景图像的统计特性，建立全向图像的统计模型是了解全景图像全局信息的有效途径<sup>[78-82]</sup>。NSS 模型通过测量全景图像的不自然程度，可以准确地表示全景图像的全局语义。此外，将 NSS 特征提取为全景图像中的全局信息，可以有效提高模型的效率。所采用的 NSS 方法首先进行归一化，具体如下：

$$I' = \frac{I(x,y) - \mu(x,y)}{\sigma(x,y) + 1} \quad (4.5)$$

其中  $x$  和  $y$  表示图像  $I$  的像素坐标， $I'$  为  $(x, y)$  处计算得到的 MSCN (Mean Subtracted Contrast Normalized) 系数，其标准差用  $\sigma$  表示， $\mu$  为以  $(x, y)$  为中心的全景图像的局部均值。众所周知，MSCN 的系数服从正态分布，即一旦图像受到损失，其 MSCN 系数分布会发生相应的变化导致偏离正态分布，该系数一般由零均值广义高斯分布(the zero-mean generalized Gaussian distribution, GGD)进行拟合，具体表现为受损图像的 GGD 参数会发生变化，其计算公式如下：

$$GGD(x, \alpha, \sigma^2) = \frac{\alpha}{2\beta\gamma(\frac{1}{\alpha})} e^{-\left(\frac{x}{\sigma}\right)^{\frac{\gamma(\frac{3}{\alpha})}{\gamma(\frac{1}{\alpha})}}} \quad (4.6)$$

其中，这里  $\gamma()$  代表伽马函数，定义为：

$$\gamma(x) = \int_0^{\infty} \varphi^{x-1} e^{-\varphi} d\varphi, \quad x > 0 \quad (4.7)$$

其中  $\alpha$  和  $\beta$  为 GGD 参数，可以有效反映图像失真。因此，本章将 GGD 参数作为质量感知特征元素进行估计，以捕捉图像的全局失真。从 MSCN 归一化畸变图像中获得的参数  $\alpha$  和  $\sigma^2$  作为全局特征的第一部分。此外，MSCN 的相邻系



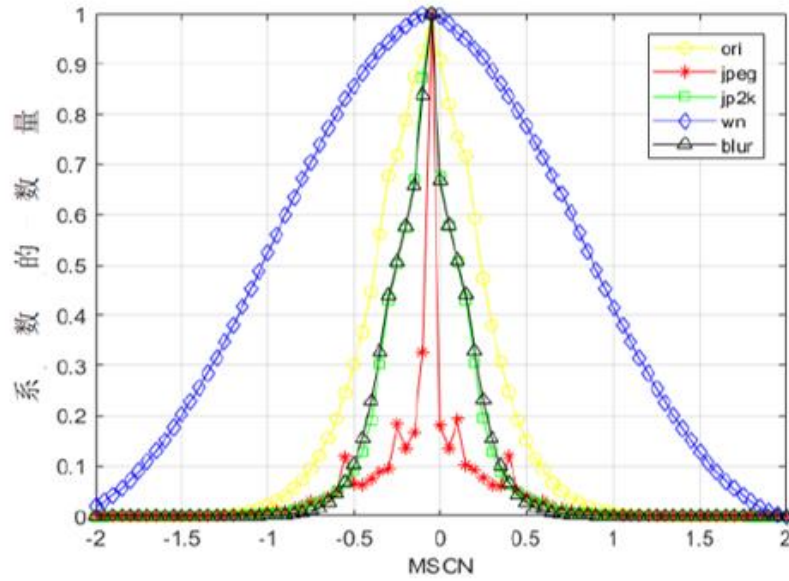
数的乘积同样能反映全景图像的全局信息，因此本文利用相邻的 MSCN 系数包括水平方向、垂直方向、主对角线方向和次对角线方向来量化全景图像的全局失真信息。因此，零模非对称广义高斯分布(AGGD)<sup>[83]</sup>表示为相邻的 MSCN 系数，如下所示：

$$AGGD(x, \tau, \alpha, \delta) = \begin{cases} \frac{\tau}{(\alpha + \beta)\gamma(\frac{1}{\tau})} e^{-\left(\frac{-x}{\alpha}\right)^\tau}, & \forall x \leq 0 \\ \frac{\tau}{(\alpha + \beta)\gamma(\frac{1}{\tau})} e^{-\left(\frac{-x}{\beta}\right)^\tau}, & \forall x \geq 0 \end{cases} \quad (4.8)$$

AGGD 的均值定义为：

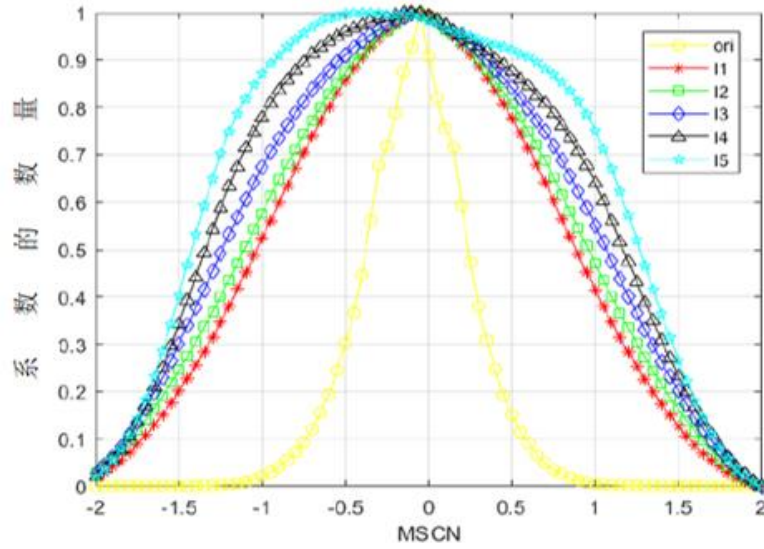
$$\delta = (\beta - \alpha) \frac{\gamma(\frac{2}{\tau})}{\gamma(\frac{1}{\tau})} \quad (4.9)$$

其中，参数 $(\tau, \alpha, \beta, \delta)$ 也是 NSS 特性。考虑到全景图像中的多尺度信息，本节总共提取了 36 维特征(16 个 AGGD 参数(4 个参数 \* 4 个方向)+2 个 GGD 分布参数)\*2 个尺度)作为提出模型的第二部分全局特征。



(a) 原始图像及 4 种扭曲类型的 MSCN 系数分布





(b) 5种程度高斯模糊失真的系数分布

图 4-8 不同扭曲类型下全景图像的 MSCN 系数分布曲线图

由于 MSCN 系数分布的特殊性，即失真会导致其偏离正态分布，因此本文利用该特性可视化全景图像在遭受不同噪声种类和等级后的系数分布变化。图 4-8 给出了原始图像及不同畸变类型下的 MSCN 系数分布图。图 (a) 展示了原始图像以及分别在 jpeg (Joint Photographic Experts Group)、jpeg2k、高斯白噪声 (White Noise, wn) 高斯噪声、高斯模糊 (Gaussian Blur, blur) 下的系数分布曲线。在(b)图中，本文选取 wn 类型在不同扭曲等级下的系数分布图，其中 l1 到 l5 表示扭曲程度逐渐增大。由图 4-8 可知，MSCN 系数分布的方差因受损类型的变化而变化，并且随着扭曲程度的增加，MSCN 系数分布的均值和方差皆相应的发生改变。综上所述，本节验证了 MSCN 系数能准确捕捉失真的类型和扭曲等级，进而衡量全景图像中全局信息的扭曲。

#### 4.3.4 基于跨通道局部二值模式的颜色特征

人类在感知图像的第一瞬间，颜色信息占感知内容的 80%<sup>[84]</sup>。此外，考虑到全景图像中颜色信息无法通过层次特征准确地被表达，且全景图像失真时产生的颜色信息是不自然的，这是以往研究中经常被研究者忽视的问题但却是图像质量评价(IQA)中重要的质量评价指标。因此，本工作的模型有针对性地提取了全景图像中上的颜色特征。分析颜色特征还可以有助于分析图像中的内容，

如下图 4-9 中示例，(a) 图像中在 RGB 三种不同通道中，红色通道的大部分信息表示的是苹果的内容，而 G 通道则表达该图中的背景信息，人们可以轻易地知道 B 通道在在途中的作用并不重要。当 B 通道在三种颜色通道中占据重要作用时，可以联想到，该图最有可能的是像 (b) 展示的以蓝天为主的图像。

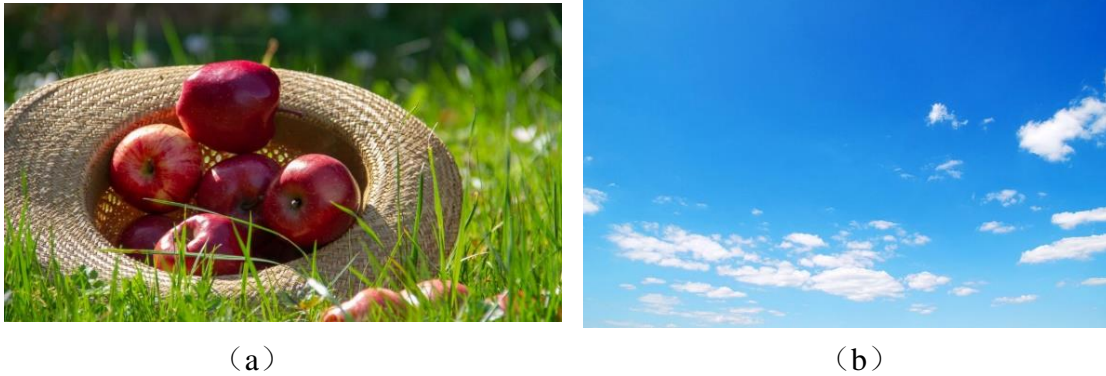


图 4-9 颜色信息主导图像内容的示例

局部二进制模式 (LBP) 方法<sup>[85]</sup>是一种具有旋转和灰度不变性的描述图像局部纹理特征的经典算子，正因为其优点和有效性相比其他描述符来说更易于实现、理解，计算复杂度低，因此被广泛使用。LBP 被定义为一个由一个中心点及其周围的邻域组成的矩阵，通常被用于描述单通道颜色特征，其定义由下式定义：

$$LBP_{P,R}(i,j) = \sum_{n=1}^P \delta(r_n - r_c) 2^{n-1} \quad (4.10)$$

$$\delta(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (4.11)$$

其中， $R$  是采样半径， $P$  表示半径为  $R$  的圆上采样点的数量。 $r_c$  表示中心点  $(i, j)$  处的像素值， $r_n$  是局部邻域中第  $n$  个采样点的像素值（以  $r_c$  为中心， $R$  为半径  $(n=1, 2, \dots, P)$ ）。以  $r_c$  为阈值，将周围邻域中的  $r_n$  逐一进行比较和二值化。如果  $r_n$  大于或等于  $r_c$ ，则编码为 1，否则编码为 0。然后，1 和 0 的序列形成局部二进制模式，该过程如图 4-10 所示。

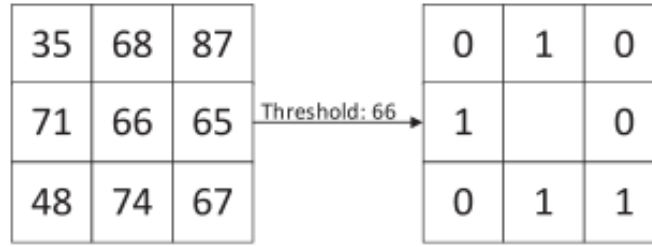


图 4-10 LBP 的计算规则

假设图像的大小为  $m \times n$ 。在识别每个像素的 LBP 码之后，构建直方图  $h$  来表示纹理图像，其中  $k$  是最大 LBP 值。

$$H(x) = \sum_{i=1}^M \sum_{j=1}^N h(LBP_{P,R}(i,j), k) \quad (4.12)$$

$$h(x,y) = \begin{cases} 1, & x = y \\ 0, & otherwise \end{cases} \quad (4.13)$$

考虑到利用不同颜色多通道之间的相关性和依赖性构建的通道间特征可以有效提高特征识别能力，与单通道颜色信息的处理不同，通道间特征能充分体现不同颜色通道之间的相关性和依赖性，因此本文基于不同通道间的关系分析全景图像中蕴含的颜色特征。另外，考虑到本文采用的公开数据集的色彩特点，本节按照 R-G-B 信道序列采用跨通道局部二值模式(cross-channel local binary pattern, CCLBP)<sup>[86]</sup>对全景图像中的每个像素通道间特征进行编码，计算不同图像中像素之间的颜色关系以捕获跨信道信息，这使得所提出的模型能够更有效、更全面地评价全景图像的质量。

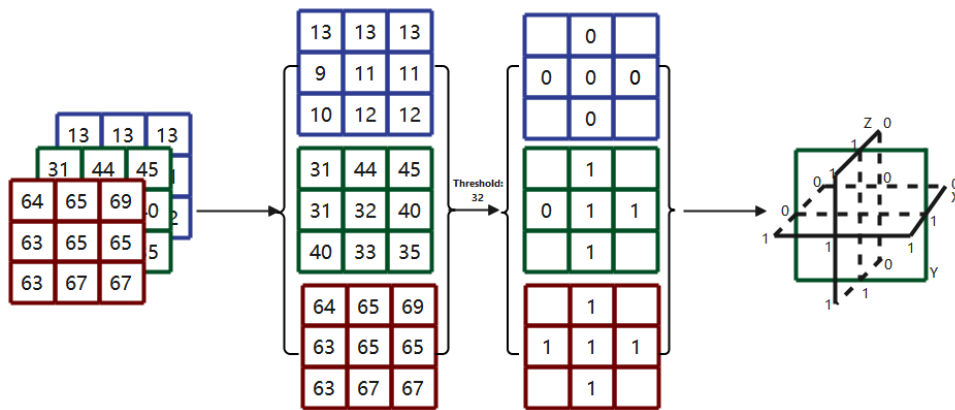


图 4-11 R-G-B 序列下 CCLBP 的构建过程。

图 4-11 展示了基于 R-G-B 序列的 CCLBP 的构建过程。首先，在半径  $r = 1$  的情况下，局部  $3 \times 3 \times 3$  立方体中总共有 18 个邻域点，本章将局部立方体分成三个不同的颜色通道。本节以 G 通道的中心点作为阈值，在局部立方体的三个正交平面上执行二值化操作，其中三个正交平面分别被称为平面 X、平面 Y 和平面 Z。从三个正交平面中，本节得到了一个 CCLBP 码。具体而言，平面 Y 的 LBP 码表示单通道局部纹理特征，然而，平面 X 和 Z 的 LBP 代码表示多通道局部纹理特性。这三个本地特性构成了本文的 CCLBP 描述符，其定义如下：

$$CCLBP_{f,r,m}(i,j) = \sum_{n=1}^m \eta_f(p_n - p_c) 2^{n-1} \quad (4.14)$$

$$\eta_f(k) = \begin{cases} 1, & k \geq 0 \\ 0, & k < 0 \end{cases} \quad (4.15)$$

其中  $r$  表示采样半径， $m$  表示半径为  $r$  的圆上采样点的个数。 $f$  表示三个正交平面的索引 ( $f = 1: x, 2: y, 3: z$ )，并将  $\eta_f()$  定义为阈值函数。 $p_n$  为中心像素的像素值， $p_c$  为以  $p_n$  为中心的邻域像素值。类似于<sup>[87]</sup>，最后将跨通道信息的前 20 个直方图特征提取为颜色特征。

#### 4.4 质量回归与预测

通过上几小节本文已经获取了全景图像的三种层次感知特征和颜色特征，本节将对使用的回归预测算法进行简单的介绍。作为支持向量机 (Support Vector Machine, SVM)<sup>[88]</sup> 的一个重要分支，SVR 是在经常被用于图像质量评价的回归模型，一般应用于数据的拟合。

在本文中将提取到的视觉感知特征和主观质量分数输入 SVR，利用输入的特征值计算图像的预测质量，即获得最终客观评价分数。评价算法的好坏根据预测分数值与输入的主观分数值的拟合程度加以判定，本文中利用 2.3.2 中的性能评价指标直观地表达主观分数与利用 SVR 计算的客观分数之间的拟合程度。

如图 4-12 展示了 SVR 的基本原理。距离  $\varepsilon$  被定义为容忍偏差，即在间隔带内的点与实线之间的可以忽略不计的偏差值，损失  $\zeta$  计算了间隔带外的点到虚线的距离，最后通过最小化总损失以及最大化间隔来优化模型。

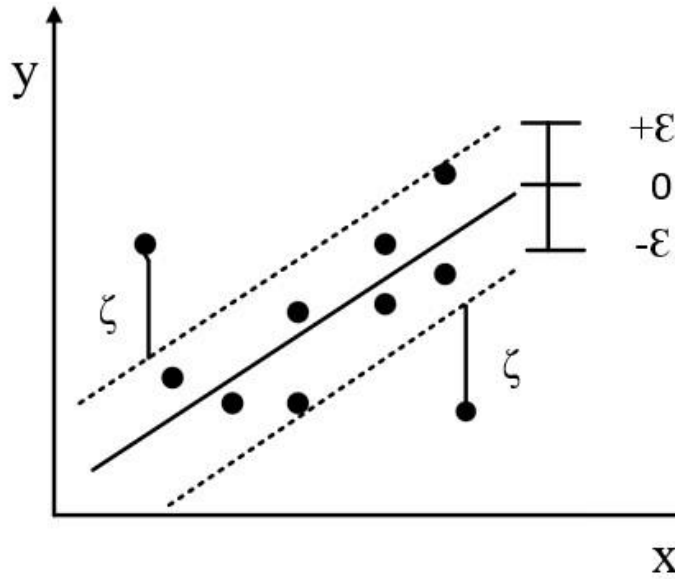


图 4-12 SVR 回归原理示意图

在本文的实验中，使用径向基函数核(radial basis function, RBF)执行 $\varepsilon$ -SVR ( $\varepsilon$ -SVR 的参数(cost, gamma)设置为 12 和 -8)来作为模型的质量评估器，从四个不同方面有效地分析了全景图像失真，并将特征转换为质量分数，很好地符合主观感知。本工作将每个数据集随机分为两个不重叠的部分：80%的全景图像作为训练样本，其余是测试样本，以便更好地训练所提出的预测模型。另外，此过程被迭代 1000 次以避免计算过程中出现的偶然性。

#### 4.5 实验结果与分析

为了验证本章提出的基于视觉显著性特征的无参考全景图像质量评价算法的性能，本节利用两个公开可靠的数据集在算法的性能和稳定性上进行了综合比较。本文运行的环境是在 Intel(R)Core(TM)i5 CPU, 16GB RAM, 64 位系统 PC 端上进行的。将本章的评价算法与许多已被提出的 FR-IQA、FR-OIQA、NR-IQA 和 NR-OIQA 模型进行了比较。所选择的 FR-IQA 方法包括 PSNR、SSIM、VIF、VSI、GMSD、FSIM 和 FSIMc。为了表达 FR-OIQA 模型的性能，本章采用了四种算法，分别是 CPP - PSNR、S-PSNR、WS-SSIM 和 WS-PSNR<sup>[89]</sup>。其中采用的 NR-IQA 方法包括 BRISQUE、NIQE、BMPRI<sup>[90]</sup>、DB-

CNN<sup>[91]</sup>、MEON<sup>[92]</sup>、CSC<sup>[30]</sup>、ASY-PIQA<sup>[93]</sup>、SSP-BOIQA<sup>[41]</sup>、MC360IQA<sup>[42]</sup>和DDAMN<sup>[46]</sup>模型。

表 4-1 全景图像 CVIQD2018 数据集上的整体性能对比

评估算法	SRCC	KRCC	PLCC	RMSE
PSNR	0.7731	0.5698	0.2585	13.8028
SSIM	0.6747	0.4819	0.6901	10.3403
VIF	0.8347	0.6428	0.8368	7.8230
VSI	0.8929	0.7115	0.9130	5.8277
GMSD	-0.8503	-0.6527	0.8647	7.1864
FSIM	0.9150	0.7425	0.9248	5.3094
FSIMc	0.9161	0.7448	0.9298	5.2952
CPP-PSNR	0.8256	0.6268	0.8397	7.5900
S-PSNR	0.7997	0.5999	0.8330	7.6747
WS-SSIM	0.9116	0.7445	0.9293	5.2626
BRISQUE	0.9192	0.7628	0.9377	4.8705
NIQE	-0.8168	-0.6200	0.8454	7.6316
BMPRI	0.7470	NA	0.7919	8.5258
DB-CNN	0.9308	NA	0.9366	4.9311
MECN	0.5670	NA	0.6650	10.5100
MC360IQA	0.9139	NA	0.9506	3.0905
SSP-BOIQA	0.8614	0.6954	0.9077	6.1178
CSC	0.7560	NA	0.7320	9.3870
DDAMN	0.9110	NA	0.9020	6.1170
本文算法	<b>0.9536</b>	<b>0.8172</b>	<b>0.9569</b>	<b>4.1671</b>

本节在表 4-1 和表 4-2 中展示了本章提出的模型和其他算法分别在 CVIQ2018 和 OIQA 数据集上的性能对比结果。其中对比模型的结果均来自其出处文章的实验结果值,“NA”是不可用数据的标志。从两个表中可以看出,本章的模型算法在 OIQA 数据集上的性能指标比 CVIQD2018 数据集的性能指标更好,这就证实了本章的算法更适用于需要更好的稳定性和普遍适用性的预测模型、以及全景图像损伤程度和类型不确定和多样的实际应用。从表中可以清楚地知道本文的算法在两个数据集上的性能指标均在前两名。在 CVIQD2018 数据库中,本模型在 PLCC (如式 2.4 所示)、SRCC (如式 2.5 所示)和 KRCC (如式 2.6 所示)指标下的性能都达到了第一位,RSME (如式 2.7 所示)性能指标排名第二,与其他方法相比具有非常强的竞争力。值得一提的是,在 OIQA 数据集中,各项指标都是最好的,具有较强竞争优势的 BRISQUE 模型与

文的模型的性能存在着较大的差距。综上所述，本章提出的模型在评估全景图像质量方面具有优异的竞争力和无可比拟的稳定性，这就意味着它在预测全景图像的质量时能极其接近人类的感知质量。

表 4-2 OIQA 数据集上的性能比较

评估算法	SRCC	KRCC	PLCC	RMSE
PSNR	0.5634	0.3940	0.5828	1.7916
SSIM	0.8573	0.6598	0.8679	1.0953
VIF	0.8305	0.6413	0.8331	1.2194
VSI	0.9165	0.7388	0.9191	0.8684
GMSD	-0.3186	-0.2229	0.3166	2.0913
FSIM	0.9225	0.7482	0.9274	0.8250
FSIMc	0.9240	0.7496	0.9284	0.8194
CPP-PSNR	0.5182	0.3502	0.5186	1.8078
S-PSNR	0.5303	0.3588	0.5319	1.7904
WS-PSNR	0.5032	0.3414	0.5044	1.8256
BRISQUE	0.9247	0.7728	0.9374	0.7367
NIQE	-0.8155	-0.6186	0.8155	1.2760
BMPRI	0.8331	NA	0.8424	1.1261
DB-CNN	0.8653	NA	0.8852	0.9717
MEON	0.7170	NA	0.7490	9.5360
ASY-PIQA	0.7150	0.5270	0.7408	1.4264
MC360IQA	0.9187	NA	0.9247	4.6247
CSC	0.7070	NA	0.7640	9.2940
DDAMN	0.9230	NA	0.8990	6.3960
本文算法	<b>0.9614</b>	<b>0.8364</b>	<b>0.9620</b>	<b>0.5819</b>

另外，由于每个数据集中的全景图像失真类型不同，本节验证了本章的模型与其他模型在特定失真类型下预测全景图像质量的性能，并分别在表 4-3 中展示了在 CVIQD2018 数据集上的性能对比结果。总的来说，本章算法相比于其它模型来说优势突出，具有更好的稳定性。具体而言，与其他方法相比，例如 BRISQUE 和 DDAMN 在 JPEG 类型，本章介绍的模型在 JPEG 类型中具有出色的性能，这表明它对任何扭曲类型都有积极的影响。通过以上分析，可以得出本章模型对预测各种类型的受损全景图像都具有竞争力和准确性。

图 4-13 显示了本文模型在 CVIQD2018 数据集和 OIQA 数据集上预测得到的客观值和数据集提供的主观值之间拟合性的散点图。算法的总体性能可以通

过散点图的收敛性来表示好坏。具体来说,散点图的收敛性越好,说明模型拥有越优异的性能。图的纵轴表示提供的主观评分,而横轴为模型预测得到的客观评分。图 4-13 (a)和(c)表示两个数据集的总体性能,(b)和(d)表示不同损坏类型下的算法性能。从图的收敛性中来看,本文算法通过模拟视觉感知特征,准确地评估了全景图像质量,并预测了与主观评分保持高度一致的客观值分数。

表 4-3 单个失真类型在 CVIQD2018 数据集上的性能比较

评估算法	JPEG		AVC		HEVC	
	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC
PSNR	0.888	0.7637	0.7842	0.7842	0.7468	0.745
SSIM	0.8392	0.7439	0.5729	0.5782	0.5746	0.5761
VIF	0.9743	0.9417	0.944	0.9422	0.9394	0.9379
VSI	0.9625	0.9198	0.9157	0.9133	0.8814	0.871
GMSD	0.9659	-0.9129	0.8574	-0.8604	0.863	-0.8634
FSIM	0.9711	0.9286	0.9460	0.9430	0.9318	0.9303
FSIMc	0.9719	0.9322	0.9467	0.9453	0.9308	0.9297
WS-PSNR	0.8772	0.752	0.7708	0.769	0.7428	0.7386
CPP-PSNR	0.8802	0.7604	0.7748	0.7726	0.7469	0.743
S-PSNR	0.8886	0.7729	0.7854	0.7815	0.7578	0.754
BRISQUE	<b>0.9818</b>	0.9541	0.9428	0.9377	0.8991	0.9015
NIQE	0.8965	-0.8427	0.8166	-0.8127	0.7835	-0.7766
BMPRI	0.9464	0.9031	0.7745	0.7714	0.7548	0.7644
MEON	0.8080	0.566	0.5990	0.5740	0.7830	0.7820
MC360IQA	0.9746	0.9316	0.9461	0.9244	0.9126	0.8985
CSC	0.7520	0.6840	0.7760	0.7090	0.7500	0.7140
DDAMN	0.9570	<b>0.9610</b>	0.9530	0.9490	0.9290	0.9140
本文算法	0.9799	0.9519	<b>0.9756</b>	<b>0.9689</b>	<b>0.9475</b>	<b>0.9395</b>

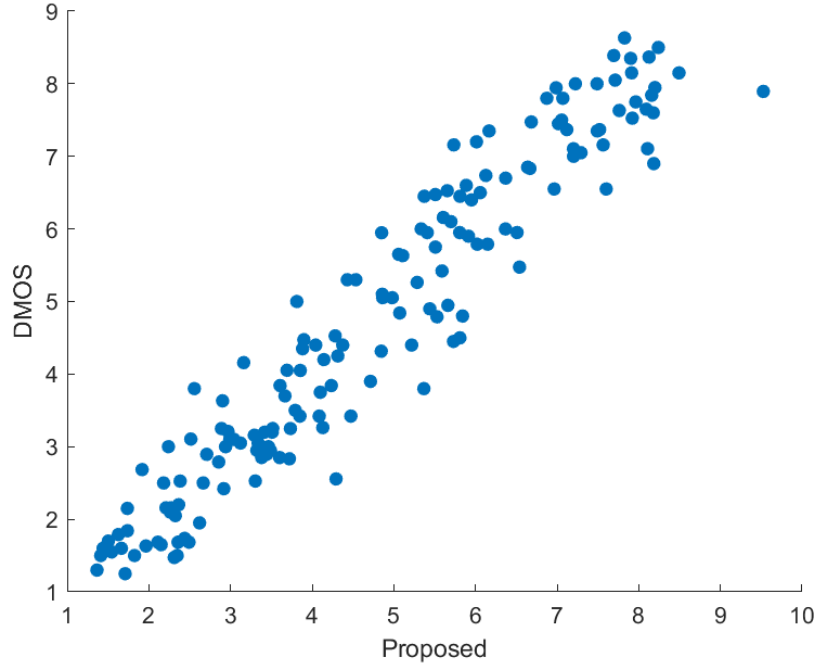
表 4-4 计算复杂度

评估算法	时间 (s)
BRISQUE	2.19
MC360IQA	4645.84
本文算法	<b>159.81</b>

为了评估我们提出的模型的计算复杂度,我们计算了 CVIQD2018 中一幅全景图像的四个模型中的平均运行时间,如表 4-4 所示。由表可以看出,本文模型的计算成本相对 2D IQA 模型较高,而相对 OIQA 模型计算成本较低。但我

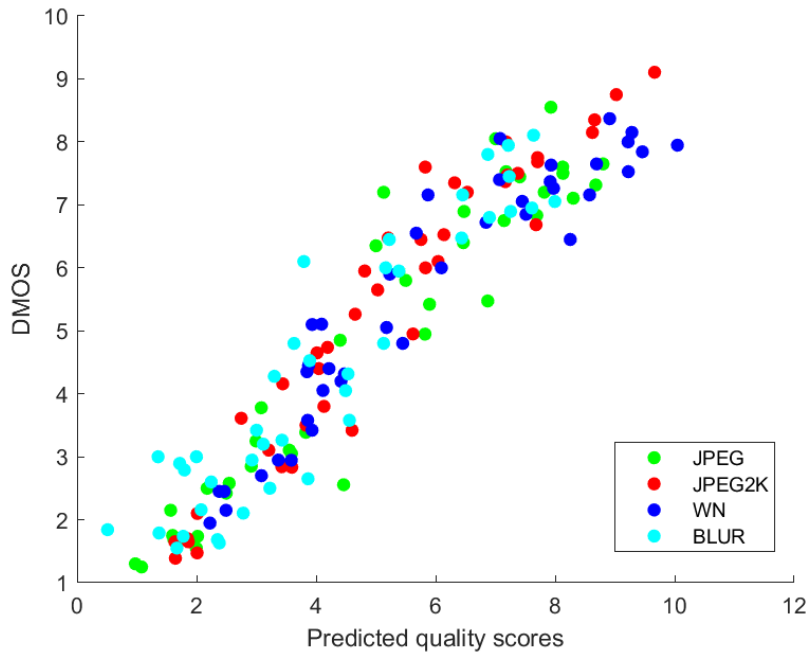


们的模型性能优于所有 2D IQA 模型和 OIQA 模型，包括 MC360IQA 模型。综

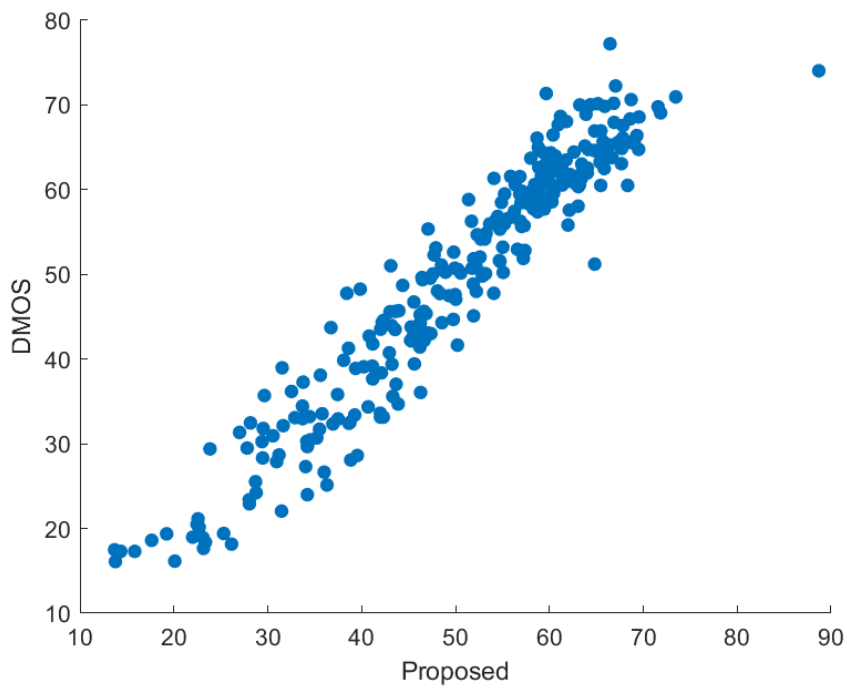


上所述，该模型在计算复杂度上有较大优势，比其他模型更适合实际应用。

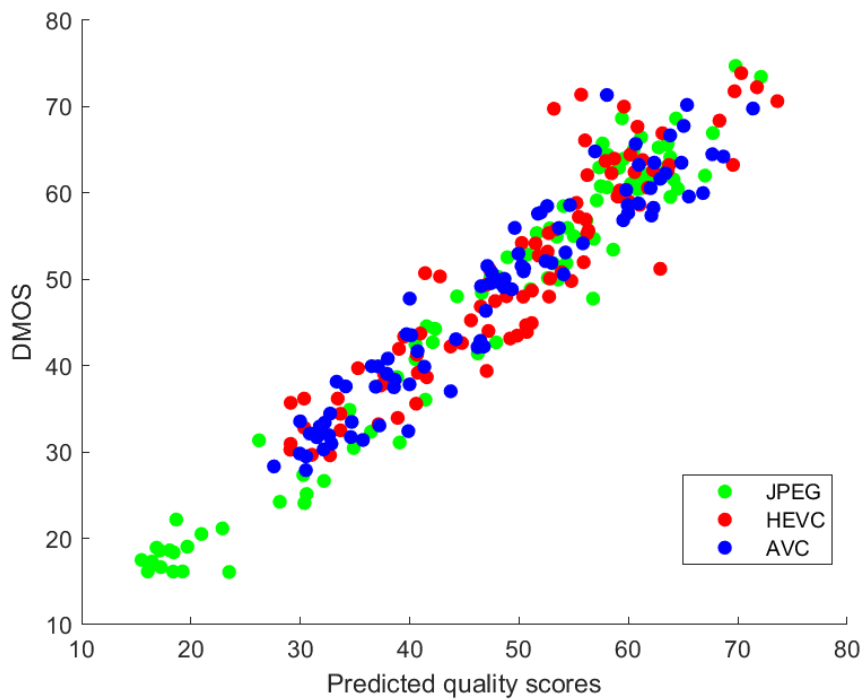
(a) 在 OIQA 数据集上的整体性能



(b) OIQA 数据集上分噪声性能



(c) CVIQD2018 数据集上的整体性能



(d) CVIQD2018 数据集上的分噪声性能

图 4-13 预测分数和主观分数的性能拟合散点图

本章提出的算法分别获取了细节特征、局部显著性特征、全局特征和颜色特征四个完整的特征。为了检验各个特征成分的有效性，本节设计了几个消融实验，分别采用方案 A、方案 B、方案 C、方案 D 四种不同方案。相应的结果如表 4-5 所示。从表中可以看出，使用任何特征不完全的方案，模型的性能都比使用四种完整特征信息差。其中，方案 A 仅排除了细节信息，其结果表明细节描述符在反映全景图像隐藏的扭曲细节信息方面非常有效。方案 B 证明了忽略局部显著性特征降低了所提模型评估全景图像质量的全面性。方案 C 忽略了全局信息，从该方案的性能中可以总结出，全局特征可以有效提高本文模型的准确度。方案 D 对无颜色信息的全景图像质量进行了预测，表明颜色信息对提高模型的整体性能起着至关重要的作用。总之，该实验充分验证了模型算法和本章所有分析的特征的有效性。

表 4-5 不同特征的性能比较

方案	OIQA			CVIQD2018		
	PLCC	KRCC	SRCC	PLCC	KRCC	SRCC
方案 A	0.9508	0.8114	0.9491	0.9512	0.8036	0.9468
方案 B	0.9620	0.8356	0.9611	0.9555	0.8141	0.9518
方案 C	0.9360	0.7910	0.9393	0.9501	0.8009	0.9440
方案 D	0.9565	0.8207	0.9553	0.9513	0.8052	0.9484
综合	<b>0.9620</b>	<b>0.8364</b>	<b>0.9614</b>	<b>0.9569</b>	<b>0.8172</b>	<b>0.9536</b>

## 4.6 本章小结

本章在考虑全景图像四个方面内容损失的基础上，提出了一种较新的评估全景图像的模型算法。首先本章提取了感知全向图像是否失真的细节信息，利用基于频域小波变换和基于灰度共现的底层特征捕捉失真图像的细节变化。考虑到注意力区域在全景图像中的重要作用，采用基于布尔映射并结合 SLIC 算法提取局部中层显著特征。本章利用层次感知特征的高级全局特征，通过自然场景统计算法来表示被扭曲的全局语义，准确地反映了全景图像的不自然性。基于人眼对颜色信息的敏感感知性，采用基于 CCLBP 的视觉感知颜色纹理描述符，挖掘不同通道颜色像素之间的相关性和依赖关系。最后，基于 SVR 提取的四种类型特征，并将预测得分进行映射，结果验证了本文的模型对全景图像的畸变非常敏感。

## 第 5 章 总结与展望

### 5.1 总结

近年来,随着计算机视觉的飞速发展,基于虚拟现实(VR)技术的全景图像逐渐被大众所熟知,并引起了广大学者们的关注。然而,全景图像在处理、传输、压缩、存储等过程中都有一定程度的退化。与二维图像表示不同,全景图像是根据头戴式显示器(HMD)的原理生成的,这在某种程度上造成了有限的视口。此外,当人们使用 HMD 设备观看时,容易捕捉到扭曲的视场,这影响了人们对体验的满意度。因此,如何更准确有效地评价全景图像的质量成为未来必须解决的关键问题。

由于人是图像质量的最终评价者,而负责处理视觉信息的 HVS 在感知图像时其注意力机制会发挥关键的作用,因此研究视觉注意力机制对全景图像的工作机理以提高处理全景图像的效率是衡量全景图像质量、推动全景图像发展的必不可少的一环。因此,本文分析针对全景图像的特点以及人眼的视觉特性的视觉显著性模型,以准确地捕捉全景图像中重要的信息。本文提出利用超像素分割技术对全景图像进行语义分割,去除图像中的冗余信息,极大地降低计算复杂度,另外本文认为布尔映射的拖布结构的不变性在显著性区域的检测中扮演着重要的角色,因此基于布尔映射本文提出了一种新型的显著性检测算法,这极有效的帮助计算机准确的获取与人眼视觉系统保持一致的视觉限制性区域。

NR-OIQA 的最终目标是设计一种综合感知方法,能够在没有先验知识的情况下准确评估受损的全向图像质量。然而,由于缺乏基于非神经科学或非生物学的模型,大多数现有研究无法达到可信的准确性。受此启发,本文考虑 HVS 的层次感知特征,包括细节信息、局部显著性信息、全局信息和常被研究者忽视的颜色信息,提出了一种新颖的基于视觉感知和神经科学的 OIQA 模型。根据神经科学中的分层过程,应用基于高频共现矩阵和基于方差的细节特征来感知频域中首先感受到扭曲的细节信息。基于熵的全景显著图与简单线性迭代聚类算法的结合,强调了丰富的对质量敏感的局部显著信息。利用全局全景统计模型表示全局语义失真作为人类感知高阶特征。采用视觉感知的颜色纹理描述

符和跨通道局部二进制模式来映射颜色信息，该纹理描述符有效地反映了像素在不同颜色通道之间的相关性和依赖性。最后，提取上述特征并结合主观评分，通过 SVR 对客观质量评分进行评价。实验表明，该方法在 CVIQD2018 和 OIQA 数据集上具有更高的精度和更强的稳定性。

## 5.2 展望

本文基于全景图像的格式特点和人类视觉注意力机制设计了基于视觉显著性特性的无参考全景图像质量评价算法，实验结果表明本文提出的算法能够准确的评价全景图像的视觉感知质量。但关于本文的研究工作依然存在可提高的空间，未来还需要努力落实的要点如下：

（1）本文基于全景图像的特点提出了计算全景图像显著性映射的算法。但是在不考虑算法计算复杂度的条件下，该算法的实验的性能并未达到最优的结果，这表明该算法对于显著性区域的计算并没有达到最理想的效果，因此如何在保证低的计算复杂度的基础上更加贴合人类视觉注意力机制以达到更加准确地检测全景图像的显著性区域，是接下来需要着重努力的方向。

（2）本文提出的全景图像质量评价方法是基于传统的提取手工特征的方式，虽然该方式一定去得了一定的研究成果，但深度学习的发展使得机器拥有了自己的学习能力，这大大降低了人工投入思想的成本也逐渐被研究者所青睐。基于其极强的自我学习能力，将其应用到全景图像的质量评价方面将成为下一阶段研究课题的重点所在。

## 致谢

今天，终于鼓足勇气开始这最难以下笔的部分，这也意味着我即将在辽大求学的道路上走到了终点。

读研三年，疫情两年半，研究生生活没那么快乐，研究和人生一样需要不断的调整和修改，我们都是一边修船一边开船。但我很庆幸能在辽宁大学这个大家庭欣赏人生风景。

“老师就像一盏灯，为我们照亮前方的路。”每次遇到诸如此类的话，我都无法苟同，我不知道老师到底是有多么无私才会“燃烧自己，照亮别人”，可是现在我们的刘允老师确实让我懂得什么是照亮前方路的灯，又是如何燃烧自己，照亮别人。如果对于教师的描述只是学识渊博、见解独特、工作严谨，那我觉得这对刘允老师描述的太片面、太抽象，应该描述在数不清的深夜和假期处理我们的科研问题时的奉献，还应该描述耐心帮我们解决棘手问题的效率，安抚我们情绪的强大力量……幸遇恩师，方知何为大学者。我所羡慕的三十而立女性，于我而言亦师亦友。“不怕做错事、把眼光放长远”，每每感到焦虑的时候，小刘老师的话总能给我力量和指引。我好幸运，遇到了照亮我人生路的恩师，可是……

山水一程，三生有幸，感谢我的室友们，陪伴彼此的生活。给予我持续温暖的大哥，总能在我最难过的时候如和煦春风给我源源不断的安抚；给予我快乐的萍萍，总能在我艰苦的时候伴我左右；感谢夸赞鼓励我的雪莉和晶晶总能让我在自我怀疑时重回自信，顶着寒风给被隔离的我送温暖；感谢小崔每每都能为我们创造干净整洁的生活环境。何其有幸，遇到你们，给我在这三年留下无数温暖。

感谢我的同门祖亮，帮助我解决科研问题，给我带来新奇的人生见解，在每次科研疲惫的时候只要转头就能看到他无言的陪伴，这给予了我极大的力量。

感谢陪我一起走过很多年的朋友们，好的友情是永远心里留着原本的位置。

二十年的求学路，感谢我的家人给予我的支持、尊重我的选择。这世界上最爱我的人，无以为报，唯愿健康。

感谢没那么优秀不太自信但尽最大努力进步的自己。

最后感谢所有参与评审和答辩的专家老师们，是你们的帮助让我的论文更加严谨。

悲伤和希望都是一束光，而我知道这些光终将汇聚成深夜的萤火。人世间山水迢迢，路遥马急。希望此后有机会用双脚去丈量书里的世界，不辜负读书吃过的苦，终有所成。

殷晓华

二〇二三年五月

## 参考文献

- [1]. Z. Luo, L. Song, S. Zheng, & N. Ling. H.264/advanced video control perceptual optimization coding based on JND-directed coefficient suppression[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2013, 23(6):935–948.
- [2]. J. Zou, H. Xiong, C. Li, L. Song, Z. He, & T. Chen. Prioritized flow optimization with multi-path and network coding based routing for scalable multirate multicasting[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2011, 21(3): 259–273.
- [3]. X. Liu, D. Zhai, J. Zhou, S. Wang, D. Zhao, & H. Gao. Sparsity-based image error concealment via adaptive dual dictionary learning and regularization[J]. IEEE Transactions Image Processing, 2017, 26(2): 782–796.
- [4]. X. Liu, D. Zhao, J. Zhou, W. Gao, & H. Sun. Image interpolation via graph-based Bayesian label propagation[J]. IEEE Transactions Image Processing, 2014, 23(3): 1084–1096.
- [5]. J. Zhou, O. C. Au, G. Zhai, Y. Y. Tang, & X. Liu. Scalable compression of stream cipher encrypted images through contextadaptive sampling[J]. IEEE Transactions Information Forensics Security, 2014, 9(11): 1857–1868.
- [6]. H. G. Kim, H. T. Lim, & Y. M. Ro. Deep virtual reality image quality assessment with human perception guider for omnidirectional image[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2020, 30(4): 917–928.
- [7]. M. Chen, Y. Jin, T. Goodall, X. Yu, & A. Conrad Bovik. Study of 3D virtual reality picture quality[J]. IEEE Journal of Selected Topics in Signal Processing, 2020, 14(1): 89–102. 2020.
- [8]. Y. Kusuno, J. Takahashi, & Y. Yu. A method localizing an omnidirectional image in pre-constructed 3d wireframe map[C]. IEEE/SICE International Symposium on System Integration, 2019.
- [9]. Y. Kusuno, J Takahashi, & Y. Yu. A Method Localizing an Omnidirectional Image in Pre-Constructed 3D Wireframe Map[C]. IEEE/SICE International Symposium on System Integration, 2019.
- [10]. L. Itti, C. Koch, & E. Niebur. A model of saliency based visual attention for rapid scene analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11):1254-1259.
- [11]. J. Harel, C. Koch, & P. Perona. Graph-based visual saliency[C]. Advances in Neural Information Processing Systems, 2006, 19:545-552.
- [12]. X. Hou, & L. Zhang. Saliency Detection: A spectral residual approach[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2007, 1-8.
- [13]. R. Achanta, F. Estrada, P. Wils & S. Süsstrunk. Salient Region Detection and Segmentation[C], International Conference on Computer Vision Systems (ICVS '08),

- Springer Lecture Notes in Computer Science, 2008, 5008:66-75.
- [14]. R. Achanta, S. Hemami, F. Estrada, & S. Süsstrunk. Frequency-tuned salient region detection[C]. IEEE International Conference on Computer Vision and Pattern Recognition, 2009, 1597-1604.
- [15]. Y. Fang, X. Zhang, & I. Nevrez. A novel superpixel-based saliency detection model for 360-degree images[J]. Signal Processing Image Communication, 69(2018):S0923596518307379-.
- [16]. A. D. Angelis, A. Moschitta, F. Russo, & P. Carbone. A Vector Approach for Image Quality Assessment and Some Metrological Considerations[J]. IEEE Transactions on Instrumentation and Measurement, 2009, 58(1): 14-25.
- [17]. 卢鹏, 刘楷贇, 邹国良, 王振华, 郑宗生. 基于多特征融合和卷积神经网络的无参考图像质量评价[J]. 液晶与显示, 2022, 37(1): 66-76.
- [18]. 杨春玲, 杨雅静. 基于多尺度特征逐层融合深度神经网络的无参考图像质量评价方法[J]. 华南理工大学学报:自然科学版, 2022, 50(4): 81-89+141.
- [19]. R. W. Buccigrossi, & E. P. Simoncelli. Image compression via joint statistical characterization in the wavelet domain[J]. IEEE Transactions on Image Processing, 1999, 8(12):1688-1701[J].
- [20]. C. Chen, H. Zhao, H. Yang, C. Peng, & T. Yu. Full Reference Screen Content Image Quality Assessment by Fusing Multi-level Structure Similarity [J]. ACM Transactions on Multimedia Computing Communications and Applications, 2020, 1(1).
- [21]. C. Chen, H. Zhao, H. Yang, T. Yu, C. Peng, & H. Qin. Full-Reference Screen Content Image Quality Assessment by Fusing Multilevel Structure Similarity[J]. ACM Transactions on Multimedia Computing Communications and Applications, 2021, 17(3).
- [22]. Z. Chen, Z. Wei, & W. Li. Blind Stereoscopic Video Quality Assessment: From Depth Perception to Overall Experience[J]. IEEE Transaction Image Processing, 2017, 2(2017): 721-734.
- [23]. 丁盈秋, 杨杨, 成茗, 张卫明. 基于彩色三要素的无参考对比度失真图像质量评价方法[J]. 北京航空航天大学学报, 2022, 48(8): 1418-1427.
- [24]. 王茜, 郑斌军, 孔玲君, 顾萍. 基于视觉显著性和感知相似性的全参考图像质量评价方法 [J]. 包装工程, 2022, 43(9): 239-248.
- [25]. A. Chetouani, & L. Li. On the use of a scanpath predictor and convolutional neural network for blind image quality assessment[J]. Signal Processing Image Communication, 2020, 89 (2020): 115963.
- [26]. Y. Li, F. Yang, W. Wan, J. Wang, M. Gao, J. Zhang, & J. Sun. No-Reference Stereoscopic Image Quality Assessment Based on Visual Attention and Perception[J]. IEEE Access, 2019, 7 (2019): 46706-46716.
- [27]. Y. Kusuno, J Takahashi, & Y. Yu. A Method Localizing an Omnidirectional Image in Pre-Constructed 3D Wireframe Map[C]. In 2019 IEEE/SICE International Symposium on System Integration, 2019.
- [28]. N. E. Lasmar, Y. Stitou, & Y. Berthoumieu. Multiscale skewed heavy tailed model for texture



- analysis[C]. In IEEE International Conference on Image Processing, 2010.
- [29].J. Lei, B. Wang, Y. Fang, W. Lin, P. L. Callet, N. Ling, & C. Hou. A Universal Framework for Salient Object Detection[J]. IEEE Transactions on Multimedia, 2016, 18(9) :1783–1795.
- [30].S. Ling, G. Cheung, & P. L. Callet. No-Reference Quality Assessment for Stitched Panoramic Images Using Convolutional Sparse Coding and Compound Feature Selection[C]. In 2018 IEEE International Conference on Multimedia and Expo (ICME), 2018.
- [31].W. Xue, L. Zhang, X. Mou, & A. C. Bovik. Gradient Magnitude Similarity Deviation: A Highly Efficient Perceptual Image Quality Index[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2014.
- [32].L. Zhang, L. Zhang, X. Mou, & D. Zhang. FSIM: A Feature Similarity Index for Image Quality Assessment[J]. IEEE Transactions on Image Processing, 2011, 20(8):2378–2386.
- [33].X. Sui, K. Ma, Y. Yao, & Y. Fang. Perceptual Quality Assessment of Omnidirectional Images as Moving Camera Videos[J]. IEEE Transactions on Visualization and Computer Graphics, 2021, 99 (2021): 1–1.
- [34].M. Yu, H. Lakshman, & B. Girod. A Framework to Evaluate Omnidirectional Video Coding Schemes[C]. In IEEE International Symposium on Mixed & Augmented Reality, 2015.
- [35].V. Zakharchenko, K. P. Choi, & J. H. Park. Quality metric for spherical panoramic video[C]. In SPIE Optical Engineering + Applications, 2016.
- [36].Y. Zhou, M. Yu, H. Ma, H. Shao, & G. Jiang. Weighted-to-Spherically-Uniform SSIM Objective Quality Evaluation for Panoramic Video[C]. In 2018 14th IEEE International Conference on Signal Processing (ICSP), 2018.
- [37].A. Maalouf, M. C. Larabi, & C. Fernandez-Maloigne. A grouplet-based reduced reference image quality assessment[C]. International Workshop on Quality of Multimedia Experience. IEEE, 2009.
- [38].Y. Zhang, T. D. Phan, & D. M. Chandler. Reduced-reference image quality assessment based on distortion families of local perceived sharpness[J]. Signal Processing Image Communication, 2017, 55:130-145.
- [39].L. Tang, K. Sun, L. Liu, G. Wang, & Y. Liu. A reduced-reference quality assessment metric for super-resolution reconstructed images with information gain and texture similarity[J]. Signal Processing: Image Communication, 2019, 79:32-39.
- [40].SeyedAlireza, Golestaneh, J. Lina, & Karam. Reduced-Reference Quality Assessment Based on the Entropy of DWT Coefficients of Locally Weighted Gradient Magnitudes[J]. IEEE Transactions on Image Processing, 2016, 25(11):5293-5303.
- [41].X. Zheng, G. Jiang, M. Yu, & H. Jiang. Segmented Spherical Projection Based Blind Omnidirectional Image Quality Assessment[J]. IEEE Access, 2020, 8 (2020): 1–1.
- [42].W. Sun, W. Luo, X. Min, G. Zhai, & S. Ma. MC360IQA: The Multi-Channel CNN for Blind 360-Degree Image Quality Assessment[C]. In 2019 IEEE International Symposium on Circuits and Systems (ISCAS), 2019.
- [43].H. Jiang, G. Jiang, M. Yu, Y. Zhang, Y. Yang, Z. Peng, F. Chen, & Q. Zhang. Cubemap-based

- perception-driven blind quality assessment for 360-degree images[J]. *IEEE Transactions on Image Processing*, 30: 2364–2377.
- [44]. G. Ling, P. Cheung, & Le Callet. No-reference quality assessment for stitched panoramic images using convolutional sparse coding and compound feature selection[C]. In *2018 IEEE International Conference on Multimedia and Expo (ICME)*, 2018.
- [45]. J. Xu, W. Zhou, & Z. Chen. Blind omnidirectional image quality assessment with viewport oriented graph convolutional networks[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021, 31(5): 1724–1737.
- [46]. Y. Zhou, Y. Sun, L. Li, K. Gu, & Y. Fang. Omnidirectional image quality assessment by distortion discrimination assisted multi-stream network[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(4): 1767–1777.
- [47]. H. Jiang, G. Jiang, M. Yu, T. Luo, & H. Xu. Multi-angle projection based blind omnidirectional image quality assessment[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(7): 4211–4223.
- [48]. F. Jabar, J. M. Ascenso, & M. P. Queluz. Objective Assessment of Perceived Geometric Distortions in Viewport Rendering of 360° Images[J]. *IEEE Journal of Selected Topics in Signal Processing*, 2019, 99: 1-1.
- [49]. C. G. Bampis, P. Gupta, R. Soundararajan & A. C. Bovik. SpEED-QA: Spatial Efficient Entropic Differencing for Image and Video Quality[J]. *IEEE Signal Processing Letters*, 2017, 24(9): 1333-1337.
- [50]. M. Riesenhuber, & T. Poggio. Hierarchical models of object recognition in cortex[J]. *Nature Neuroscience*, 1999, 2(11): 1019-25.
- [51]. S. Hochstein, & M. Ahissar. View from the top: hierarchies and reverse hierarchies in the visual system[J]. *Neuron*, 2002, 36(5): 791–804.
- [52]. B. Choi, & E. Yip. OMAF PACK-VE: Stereoscopic 360 Video Format and Packing[C], Standard ISO/IEC JTC1/SC29/WG11, MPEG/M39842v3, 2017.
- [53]. Z. Chen, Y. Li, & Y. Zhang. Recent advances in omnidirectional video coding for virtual reality: Projection and evaluation[J]. *Signal Process*, 2018, 146: 66–78.
- [54]. W. Sun, K. Gu, S. Ma, W. Zhu, N. Liu, & G. Zhai. A Large-Scale Compressed 360-Degree Spherical Image Database: From Subjective Quality Evaluation to Objective Model Comparison[C]. In *2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP)*, 2018.
- [55]. H. Duan, G. Zhai, X. Min, Y. Zhu, & X. Yang. Perceptual Quality Assessment of Omnidirectional Images[C] In *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2018.
- [56]. J. Gutiérrez, E. David, Y. Rai, & P. L. Callet. Toolbox and Dataset for the Development of Saliency and Scanpath Models for Omnidirectional/360° Still Images[J]. Elsevier, 2018.
- [57]. Video Quality Experts Group et al. Final report from the video quality experts group on the validation of objective models of video quality assessment[R], phase II. VQEG, 2003.

- [58].A. Borji, & L. Itti, State-of-the-art in visual attention modeling[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1): 185–207.
- [59].S. Chen, & Y. Fu. Progressively guided alternate refinement network for RGB-D salient object detection[C], 2020,520-538
- [60].W. Ji, J. Li, M Zhang, Y. Piao, & H. Lu. Accurate RGB-D salient object detection via collaborative learning[C]. ECCV 2020, 52-69
- [61].李琦,戴蒙,张晴,刘云翔.基于注意力感知的 RGB-D 显著性检测[J].计算机工程与设计,2023,44(2):511-517.
- [62].王卫兵,张晓琢,邓强.多分支主干监督网络下的 RGB-D 图像显著性检测[J].哈尔滨理工大学学报,2022,27(4):39-45.
- [63].陈曦涛,訾玲玲,张雪曼.采用跳层卷积神经网络的 RGB-D 图像显著性检测[J].计算机工程与应用,2022,58(2):252-258.
- [64].R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, & S. S. S. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2012, 34(11): 2274–2282.
- [65].V. Mazza, M. Turatto, & C. Umiltà. Foreground–background segmentation and attention: A change blindness study[J]. Psychological Research, 2005, 69(3):201–210.
- [66].E. Rubin. Figure & ground. Readings in Perception[R]. The University Series in, 1958, 194–203.
- [67].R. Kimchi, & M. A. Peterson. Figure-ground segmentation can occur without attention[J]. Psychological Science, 2008, 19(7):660–668.
- [68].G. C. Baylis, & J. Driver. Shape-coding in it cells generalizes over contrast and mirror reversal, but not figureground reversal[J]. Nature Neuroscience, 2001, 4: 937–942.
- [69].Z. Kourtzi, & N. Kanwisher. Representation of perceived object shape by the human lateral occipital complex[J]. Science, 2001, 293(5534):1506–1509.
- [70].S. E. Palmer. Vision science: Photons to phenomenology[J]. The MIT press, 1999.
- [71].J Zhang, & S Sclaroff. Saliency Detection: A Boolean Map Approach[C]. In 2013 IEEE International Conference on Computer Vision, 2013.
- [72].S. G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1989, 11(4).
- [73].Y. Yang, Y. Que, S. Huang, & P. Lin. Multiple Visual Features Measurement With Gradient Domain Guided Filtering for Multisensor Image Fusion[J]. IEEE Transactions on Instrumentation and Measurement, 2017, 66(4): 691–703.
- [74].R. M. Haralick, K. Shanmugam, & I. Dinstein. Textural Features for Image Classification[J]. Studies in Media and Communication SMC, 1973, 3(6): 10–621.
- [75].Z. Chen, Z. Wei, & W. Li. Blind Stereoscopic Video Quality Assessment: From Depth Perception to Overall Experience[J]. IEEE Trans Image Proces, 2017, 2 (2017): 721–734.
- [76].Y. Li, F. Yang, W. Wan, J. Wang, M. Gao, J. Zhang, & J. Sun. No-Reference Stereoscopic

- Image Quality Assessment Based on Visual Attention and Perception[J]. IEEE Access, 2019, 7: 46706–46716.
- [77].H. R. Sheikh, A. C. Bovik, & L. K. Cormack. No-reference quality assessment using natural scene statistics: JPEG2000[J]. IEEE Transactions on Image Processing, 2005, 14(11): 1918–1927.
- [78].R. W. Buccigrossi, & E. P. Simoncelli. Image compression via joint statistical characterization in the wavelet domain[J]. IEEE Transactions on Image Processing, 1999, 8(12): 1688–1701.
- [79].E. Doi, & M. S. Lewicki. Relations between the statistical regularities of natural images and the response properties of the early visual system[J]. Japanese Cognitive Science Society, SIG P&P, 2005:1-8.
- [80].Daniel L Ruderman. The statistics of natural images[J]. Network Computation in Neural Systems, 2009.
- [81].E. P. Simoncelli. Statistical Models for Images: Compression[J]. Restoration and Synthesis, 1999.
- [82].M. J. Wainwright, E. P. Simoncelli, & A. S. Willsky. Random cascades on wavelet trees and their use in analyzing and modeling natural images[C]. In Wavelet Applications in Signal & Image Processing VIII, 2000.
- [83].N. E. Lasmar, Y. Stitou, & Y. Berthoumieu. Multiscale skewed heavy tailed model for texture analysis[C]. In IEEE International Conference on Image Processing, 2010.
- [84].L. He, X. Gao, L. Wen, X. Li, & D. Tao. Image quality assessment based on S-CIELAB model[J]. Signal Image & Video Processing, 2011, 5(3): 83–290.
- [85].T. Ojala, M. Pietikainen, & T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24: 971–987.
- [86].X Shu, Z. Song, J Shi, S. Huang, & X. J. Wu. Multiple channels local binary pattern for color texture representation and classification[J]. Signal Processing Image Communication, 2021, 116392.
- [87].X. Zheng, G. Jiang, M. Yu, & H. Jiang. Segmented Spherical Projection Based Blind Omnidirectional Image Quality Assessment[J]. IEEE Access, 2020, 8: 1–1.
- [88].C. Chung, & L. Jen. LIBSVM: A library for support vector machines[J]. ACM Transactions on Intelligent Systems and Technology, 2011, 2(3):1-39.
- [89].Y. Sun, A. Lu, & L. Yu. Weighted-to-Spherically-Uniform Quality Evaluation for Omnidirectional Video[J]. IEEE Signal Processing Letters (2017), 1408–1412.
- [90].X. Min, G. Zhai, K. Gu, Y. Liu, & X. Yang. Blind Image Quality Estimation via Distortion Aggravation[J]. IEEE Transactions on Broadcasting, 2018, 1–10.
- [91].W. Zhang, K. Ma, J. Yan, D. Deng, & Z. Wang. Blind Image Quality Assessment Using A Deep Bilinear Convolutional Neural Network[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2019, 1–1.

- [92]. K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, & W. Zuo. End-to-End Blind Image Quality Assessment Using Deep Neural Networks[J]. IEEE Transactions on Image Processing, 2018, 27(3): 1202–1213.
- [93]. Y. Xia, Y. Wang, & Y. Peng. Blind Panoramic Image Quality Assessment via the Asymmetric Mechanism of Human Brain[C]. In 2019 IEEE Visual Communications and Image Processing (VCIP), 2019.

## 攻读学位期间发表的学术论文及参加科研情况

### 发表论文

#### 一、发表论文、出版专著

1. Y. Liu, **X. Yin\***, T. Tang, G. Yue, & Y. Wang. A No-reference Panoramic Image Quality Assessment With Hierarchical Perception and Color Features. Journal of Visual Communication and Image Representation, under review.
2. Y. Liu, **X. Yin\***, Y. Wang, Z. Yin, & Z. Zheng. HVS-Based Perception-Driven No-Reference Omnidirectional Image Quality Assessment. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1-11.
3. Y. Liu, **X. Yin\***, Z. Wan, G. Yue, & Z. Zheng. Toward A No-reference Omnidirectional Image Quality Evaluation by Using Multi-perceptual Features. ACM Trans. Multimedia Comput. Commun, 2023, 19(2): 1-19.
4. Y. Liu, **X. Yin\***, G. Yue, Z. Zheng, J. Jiang, Q. He & X. Li. Blind Omnidirectional Image Quality Assessment with Representative Features and Viewport Oriented Statistical Features. Journal of Visual Communication and Image Representation, 2023, 19.
5. 刘允, **殷晓华**. 一种无参考全面感知全景图像质量的评价方法, 2022-07-11, 中国, CN202210808412.2 (专利).

#### 二、科研项目

国家自然科学基金青年基金, 61901205, 2020.1.1-2022.12.31