

中图分类号：TP309
学科分类号：083900

论文编号：1028716 24-B003

博士学位论文

面向可信视觉任务的 不变表征方法研究

研究生姓名	祁树仁
学科、专业	网络空间安全
研究方向	可信人工智能
指导教师	张玉书 教授

南京航空航天大学

研究生院 计算机科学与技术学院

二〇二四年六月

Nanjing University of Aeronautics and Astronautics
The Graduate School
College of Computer Science and Technology

Invariant Representations for Trustworthy Vision Tasks

A Thesis in
Cyberspace Security

by

Shuren Qi

Advised by

Prof. Yushu Zhang

Submitted in Partial Fulfillment
of the Requirements
for the Degree of
Doctor of Philosophy

June, 2024

承诺书

本人声明所呈交的博士学位论文是本人在导师指导下进行的研究工作及取得的研究成果。除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得南京航空航天大学或其他教育机构的学位或证书而使用过的材料。

本人授权南京航空航天大学可以将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、汇编学位论文。

（保密的学位论文在解密后适用本承诺书）

作者签名：_____

日 期：_____

摘 要

开发稳健、可解释的视觉系统是迈向可信人工智能的关键步骤。为此，一类重要的研究范式将视觉任务及其物理系统所固有的不变性或对称性结构抽象出来，作为视觉表征环节的设计原则，从而在基础上改善整个系统的稳健性和可解释性。该类方法被称为“不变表征”，在几何深度学习、人工智能驱动的科学发现、以及信息安全和取证等可信人工智能领域得到了广泛的研究与应用。

目前的不变表征方法仍面临以下两方面的研究难点。在理论层面，经典不变量通常基于图像的全局表征假设，而对于视觉任务中更具信息量的图像局部和层次表征，缺乏相应的理论扩展；在实践层面，经典不变量通常用于小规模视觉任务，而对于具有不变性或对称性先验的更大规模视觉任务，缺乏相应的实践扩展。

围绕以上难点，本文系统性分析了前人的不变表征研究，进一步提出了全局、局部和层次不变表征的新方法，并成功应用于多种信息安全和取证的可信任务。主要内容概括如下：

分析经典不变量。本文首先深入分析了经典不变量方法，为后续的三个创新性课题提供了理论和实践基础。首先，在理论层面回顾不变量的发展历史并梳理其中的核心知识，重点分析近期在快速/精确计算、稳健性/不变性优化、定义扩展和应用的成功经验，这些内容将在后续课题中得到进一步的理论扩展。此外，在实践层面提供了系统性的开源实现 MomentToolbox，并基于此给出了广泛的数值评估结果，这些内容将在后续课题中得到进一步的实践扩展。

完善全局不变量。针对经典不变量在含噪环境中的适用性问题，本文提出了一种具有时频分析能力的噪声不变表征方法。首先，将频域不变量推广为时频域不变量，表现出新的时频判别性，该推广基于分数阶多项式基和傅里叶基的研究。此外，还将几何不变量推广为几何与噪声联合不变量，从而更适用于非理想成像条件，该推广基于不变子空间的研究。以上内容是对全局不变量理论的完善，与基于数据增广训练或数据增强预处理的深度表征相比，其在小规模的经典视觉任务以及版权保护的可靠任务中表现出稳健性、解释性和效率优势。

设计局部不变量。针对经典不变量在局部视觉任务中的适用性问题，本文提出了一种局部不变表征方法。首先，将经典不变量的定义从全局扩展到具有尺度空间的局部，形式化了其在几何不变性、等变性和协变性方面的核心结论。此外，推导出了高效的离散计算策略，具有低误差、常数阶复杂性以及基函数通用性的特点。以上内容设计了新的局部不变量理论，与先进的稠密局部表征相比，其在中等规模的经典视觉任务以及人为篡改取证的可靠任务中表现出稳健性、解释性和效率优势。

探索层次不变量。针对经典不变量在大规模视觉任务中的适用性问题，本文提出了一种层次不变表征方法。以不变量理论重新审视卷积神经网络中的典型模块，形式化了实现层次不变性的一般蓝图和具体定义，将上述全局和局部不变量的基本理论统一推广到层次假设。群论分析证明，提出方法在其任意中间层都表现出连续且直接的几何对称性。以上内容探索了新的层次不变量理论，与现有的经典不变量、学习表征网络、不变表征网络相比，其在更大规模的经典视觉任务以及机器生成取证的可信任任务中能够更好地权衡不变性和判别性，同时表现出解释性和效率优势。

关键词：可信人工智能，稳健性，可解释性，不变性，对称性

ABSTRACT

Developing robust and interpretable vision systems is a crucial step towards trustworthy artificial intelligence (AI). In this regard, a promising paradigm considers embedding task-required invariant/symmetry structures in the visual representation, thereby improving the robustness and interpretability of the system fundamentally. Such methods, called *invariant representations*, have been widely investigated in the trustworthy fields like geometric deep learning, AI for science, security and forensic.

Current community of invariant representations faces the following research challenges. At the theoretical level, classical invariants are based on certain global assumptions. As for more informative local and hierarchical invariants in computer vision, there is a challenge on corresponding theoretical expansion. At the practical level, classical invariants are often used in certain small-scale vision tasks. As for larger-scale vision tasks with symmetry prior, there is a challenge on corresponding practical expansion.

Focusing on the above challenges, we review previous research and present new methods for global, local and hierarchical invariant representations, which have also been successfully applied to security and forensic tasks. The main contents are summarized as follows:

Analyzing classical invariants. This dissertation begins with an in-depth analysis of classical invariants, providing theoretical and practical foundations for three subsequent topics. At the theoretical level, the history of invariants is reviewed and the core methods are sorted out, focusing on recent successes in fast/accurate computation, robustness/invariance optimization, definitional extensions, and applications. Such efforts will be theoretically extended in following topics. At the practical level, a systematic open-source implementation, MomentToolbox, is provided with extensive numerical results. Such efforts will be practically extended in following topics.

Refining global invariants. The first topic is to refine the established theory of global invariants. Here, we generalize the classical frequency-only invariants into a new tool of time-frequency analysis, thus exhibiting higher flexibility and discriminability. The core lies in my study on polynomial or Fourier bases with fractional orders. Also, we generalize the classical geometric invariants to noise invariants, thus being more practical in non-ideal imaging conditions. The core lies in my study on invariant subspaces w.r.t. signal degradation. Such efforts complement the theory of global invariants, with robustness, interpretability and efficiency benefits in small-scale visual tasks and copyright protections, than the deep representations based on augmentation training or enhancement preprocessing.

Designing local invariants. The second topic is to generalize the fundamental theory of global in-

variants to the local case. We propose local invariant representation, by extending the definition of classical orthogonal moments from the global to the local with scale space. The mathematical analysis explains the important properties, e.g., the representation robustness based on invariance, equivariance, and covariance. Also, we derive a fine-tuned discrete computation strategy, which is characterized by low error, constant complexity, and generality for arbitrary basis functions. Such efforts led to a new theory of local invariants, exhibiting robustness, interpretability and efficiency benefits in mid-scale vision tasks and hand-crafted forgery forensics, than state-of-the-art dense representations.

Exploring hierarchical invariants. The third topic is to generalize the fundamental theory of global and local invariants to the hierarchical case. We propose hierarchical invariant representation, by rethinking the typical modules of convolutional neural networks (CNN). We formalize a blueprint for hierarchical invariance and define new modules with their compositions to fulfill the blueprint. The group theory shows the continuous and one-shot equivariance at each intermediate layer. Such efforts led to a new theory of hierarchical invariants, with better trade-off between invariance and discriminability than traditional invariants and CNN in larger-scale vision tasks and AI-generated forgery forensics, also exhibiting interpretability and efficiency benefits.

Keywords: Trustworthy artificial intelligence, robustness, interpretability, invariance, symmetry

目 录

第一章 绪论	1
1.1 研究背景和意义	1
1.2 研究现状和难点	2
1.3 本文工作和创新	3
第二章 分析经典不变量	5
2.1 历史回顾	5
2.2 预备知识和经典方法	6
2.3 研究方向和最新进展	12
2.4 开源实现和数值结果	26
2.5 本章小结	36
第三章 完善全局不变量	37
3.1 引言	37
3.2 预备和基础	39
3.3 隐式路径	41
3.4 显式路径	49
3.5 实验和应用	52
3.6 本章小结	62
第四章 设计局部不变量	63
4.1 引言	63
4.2 预备和基础	65
4.3 局部不变量理论	66
4.4 局部不变量实践	74
4.5 实验和应用	76
4.6 本章小结	86
第五章 探索层次不变量	87
5.1 引言	87
5.2 预备和基础	89
5.3 层次不变量理论	91
5.4 层次不变性实践	98

5.5 实验和应用	101
5.6 本章小结	111
第六章 总结与展望	113
6.1 本文总结	113
6.2 未来展望	114
参考文献	117
致谢	137
在学期间的学术论文与研究成果	139

图表清单

图 1.1	本文研究框架。	3
图 2.1	图像矩方法分类。	6
图 2.2	正交圆谐矩的径向基函数示例。	11
图 2.3	图像矩和矩不变量的研究方向。	13
图 2.4	数字图像的矩形区域与基函数的单位圆区域之间的映射关系。	13
图 2.5	极坐标像素平铺方案。	15
图 2.6	单位圆中的对称点。	17
图 2.7	不同圆谐正交矩的平均计算误差和分解时间。	28
图 2.8	不同圆谐正交矩的重构图像样例，其中 $K = \{0, 5, \dots, 20\}$ （从左到右）。	29
图 2.9	不同圆谐正交矩的均方重构误差和结构相似性。	31
图 2.10	训练图像示例。	32
图 2.11	测试图像示例，其中旋转角度为 $\{0^\circ, 30^\circ, \dots, 120^\circ\}$ （从左到右）以及噪声方差为 $\{0, 0.05, 0.1\}$ （从上到下），同时包含尺度和位置的随机扰动。	32
图 3.1	拉东变换示意图：(a) 图像函数 $f(x, y)$ ，与原点距离 r 且与 y 轴夹角为 θ 的直线 $L(r, \theta)$ ；(b) 沿着该直线的线积分， $\int_{L(r, \theta)} f(x, y) dx dy$ ，即对应拉东变换 $\mathbf{R}_f(r, \theta)$ 。	40
图 3.2	引入分数阶参数 $\alpha \in \mathbb{R}^+$ 对径向自变量的影响，其中 $r_{\text{new}} \in [0, 1]$ 替换了 $r_{\text{old}} \in [0, 1]$ ，基于 $r_{\text{old}} = r_{\text{new}}^\alpha$ 。	45
图 3.3	时频判别性示意图：在不同参数设置下，角向基函数 $A_m(\theta)$ 、谐波径向基函数 $R_n^\alpha(r)$ 、以及二维基函数 $V_{nm}^\alpha(r, \theta)$ 的相位分别用红色、蓝色和黑色边框绘制；其中， n 和 m 分别编码径向和角度频率性质，而新引入的 α 则编码径向空间性质（之前的整数阶方法无法实现，即 $\alpha = 1$ ）。	46
图 3.4	计算过程示意图：首先将原始图像 f 投影到拉东空间中得到 $\mathbf{R}_f$ ，然后在不同参数 (α, n, m) 下计算 $\mathbf{R}_f$ 与基函数 V_{nm}^α 的内积，得到 FMR 系数 $M_{nm}^\alpha = \langle \mathbf{R}_f, V_{nm}^\alpha \rangle$ ；其中，内积的计算和基函数值的估计可以分别通过傅里叶和递归策略高效地实现，并且以上变换是可逆的。	47
图 3.5	隐式路径和显式路径关系示意图：显式路径可直接从原始图像 f 导出 FMR 系数 M_{nm}^α ，无需拉东变换 $\mathbf{R}_f$ 作为中间步骤。	49
图 3.6	含高斯噪声的 CSIQ 图像集和含泊松噪声的 CIDIQ 图像集：从右到左噪声强度增加。	52

图 3.7	CSIQ 和 CIDIQ 图像集上谐波和多项式 FMR 的幅度直方图：从右到左对应不同图像，从前到后对应噪声强度增加。.....	52
图 3.8	含噪图像重构示例：数值是重构版本相对于原始版本的均方重构误差和结构相似性。.....	54
图 3.9	不同方法下含噪图像的重构版本与原始版本的均方重构误差和结构相似性。.....	54
图 3.10	自然畸变下的模板匹配示例：(a) 和 (h) 是在噪声环境下的任意朝向模板；(b) 和 (i) 是通过荧光显微镜和智能手机拍摄的含有真实成像噪声的图像，左上与右下对应干净和含噪可视化；(c)~(g) 和 (j)~(n) 是通过不同表征得到的结果，红色和绿色对应匹配模版和真实目标可视化。.....	58
图 3.11	零水印方法示例：不同劣化图像及其峰值信噪比，相应的版权编码及其误码率。.	60
图 3.12	不同零水印方法在稳健性实验中的误码率箱线图。.....	61
图 4.1	定义扩展的示意图。.....	67
图 4.2	几何变换示意图。.....	68
图 4.3	实践流程示意图。.....	75
图 4.4	不同计算方法的精确性和复杂性对比。.....	76
图 4.5	不同稠密局部表征方法的模式检测示例。.....	77
图 4.6	不同稠密局部表征方法的模式匹配示例。.....	79
图 4.7	不同检测方法用于复制-粘贴篡改基准的结果示例。.....	81
图 4.8	不同检测方法在复制-缩放-粘贴稳健性实验中的结果示例。.....	83
图 4.9	不同取证方法在综合稳健性实验中的取证分数线箱图。.....	84
图 4.10	不同感知哈希方法用于内容认证的结果示例。.....	84
图 5.1	层次不变性蓝图：图像信息能够以几何可控的方式通过每个中间层，并在最后一层以紧凑的设计生成具有足够信息量的不变特征。.....	92
图 5.2	具有 $\mathfrak{G}_1$ 不变性的单尺度实践：树状网络编码了一组 HIR 路径，其中蓝色和黑色节点分别表示表征单元（具有不同参数）和恒等函数，连线表示节点之间的级联关系。.....	98
图 5.3	具有 $\mathfrak{G}_0$ 不变性的多尺度实践：多尺度 HIR 网络基于尺度分离先验，其中缩放协变性转变为多尺度网络之间的线性平移模式，可通过从多尺度上对应节点的特征池化得到缩放不变表征。.....	99
图 5.4	计算机视觉和模式识别实验中的图像集示例。.....	101
图 5.5	数字取证和篡改检测应用中的图像集示例。.....	106
图 5.6	对抗性扰动 UP 基准上不同取证方法的取证分数 (F1, %) 对比。.....	106

图 5.7	人工智能生成内容 SD 1.5 基准上不同取证方法的取证分数 (F1, %) 对比。.....	109
图 6.1	本文研究框架。.....	114
表 2.1	正交圆谐矩的径向基函数汇总。.....	7
表 2.2	正交圆谐矩的径向基函数性质。.....	10
表 2.3	对称点的笛卡尔坐标和极坐标。.....	18
表 2.4	不同圆谐正交矩阶数集合 $S_{nm}(K)$ 的定义和基数。.....	27
表 2.5	不同圆谐正交矩不变量的分类分数 (%)。.....	34
表 2.6	学习方法和矩不变量方法的分类分数 (%)。.....	35
表 3.1	本章方法与相关方法的概念性对比。.....	38
表 3.2	矩不变量经典方法、FM 方法和 FMR 方法的分类分数 (%)。.....	56
表 3.3	学习方法和矩不变量方法的分类分数 (%)。.....	57
表 3.4	真实荧光显微镜图像上不同表征的模板匹配分数 (%)。.....	59
表 3.5	真实手机图像上不同表征的模板匹配分数 (%)。.....	59
表 4.1	本章方法与相关方法的概念性对比。.....	64
表 4.2	不同稠密局部表征方法在模式检测实验中的 F1 分数 (%)。.....	78
表 4.3	不同稠密局部表征方法在模式匹配实验中的重复率分数 (%)。.....	80
表 4.4	不同检测方法在多种复制-粘贴篡改基准上的取证分数 (%)。.....	82
表 4.5	多种检测方法在 FAU 复制-粘贴篡改基准上的取证分数 (%)。.....	82
表 4.6	复制-缩放-粘贴稳健性实验中引入本章方法前后的取证分数和匹配性能。.....	83
表 4.7	多种检测方法在 FAU 复制-缩放-移动稳健性实验中的取证分数 (%)。.....	83
表 4.8	不同感知哈希方法在稳健性实验中的取证分数 (%)。.....	85
表 5.1	本章方法与相关方法的概念性对比。.....	88
表 5.2	小规模纹理基准上不同表征方法的分类分数 (%) 和运行时间 (秒)。.....	102
表 5.3	中等规模数码基准上不同表征方法的分类分数 (%) 和运行时间 (秒)。.....	103
表 5.4	大规模寄生虫显微基准上不同表征方法的分类分数 (%) 和运行时间 (秒, 对应 8/2)。.....	104
表 5.5	不同表征方法在多种对抗性扰动基准上的取证分数 (F1, %)。.....	107
表 5.6	不同表征方法在对抗性扰动实际 (混合) 基准上的取证分数 (%)。.....	108
表 5.7	不同表征方法在多种 AIGC 基准上的取证分数 (F1, %)。.....	110
表 5.8	不同表征方法在多种 AIGC 基准上的取证稳健性分数 (F1, %)。.....	111
表 5.9	不同表征方法在 AIGC 实际 (混合) 基准上的取证分数 (%)。.....	112

第一章 绪论

本文以可信人工智能背景，研究不变表征方法及其在计算机视觉问题中的应用。本章首先介绍研究背景和意义，随后阐明该领域的现状和难点，最后描述文章的整体结构和创新点。

1.1 研究背景和意义

可信人工智能。在人工智能（Artificial Intelligence, AI）研究中，可信性^[1,2]是重要内容之一，目前已成为社会公众所普遍关心的问题：中国等 28 国签署的全球性声明《布莱切利宣言》强调“人工智能应该以安全、以人为中心、值得信赖和负责任的方式设计、开发、部署和使用”。

回顾人工智能近年来的发展，随着深度机器学习方法在感知信息理解^[3]和棋类游戏^[4]等应用中取得突破性进展，其场景正在自然拓展到信任攸关的领域，包括但不限于科学问题求解^[5]、生物识别和身份认证^[6]、医疗诊断^[7]、自动驾驶和机器人^[8]、以及信息安全和取证^[9]。显然，在上述可信场景中，人工智能系统的稳健性和可解释性是至关重要的：一般而言，稳健性是指系统的性能对于输入的类内变化是稳定的；可解释性是指系统的行为可以被人类理解或预测。

不变表征与计算机视觉。为实现稳健和可解释的人工智能系统，一类重要的研究范式将任务背后的不变性或对称性先验抽象出来并作为特征表示（简称表征）的设计原则，被称为“不变表征”。该方法已在诸多可信领域中成功应用，并发展出了几何深度学习^[10]等新兴学科方向。

让我们将上述不变表征的一般性讨论具体到本文所关注的计算机视觉（Computer Vision, CV）领域。表征在视觉系统中扮演基础作用，即知识发现或数据处理算法一般建立在视觉数据（如图像）的“有意义”特征之上，而非直接建立于原始数据。因而，视觉系统的性能在很大程度上取决于视觉表征的好坏；在 Kaggle 等知名平台的视觉竞赛中，大量的实践经验也表明表征对结果的影响往往比算法更重要。为此，众多视觉领域都对如何获取有效的表征进行了研究^[11]，其中共同的成功经验在于视觉表征的设计必须依赖于视觉任务的先验知识。在此方面，存在一类最为基础和通用的先验知识，即自然世界的对称性先验——最早由数学家克莱因在几何学的“爱尔兰根纲领”中正式定义^[12]，后经数学及物理学家诺特、外尔、杨振宁等人发展^[13]，已成为现代物理学的核心概念，被广泛认为是主导物理学法则的一种基本规律。在几何学和物理学文献中，对称性可以被非正式地描述为：“系统的对称性是一种变换使所述系统的某些属性保持不变”。计算机视觉任务作为自然世界的一种模拟，其中对称性先验同样无处不在，例如平移作为图像目标检测系统的对称性，即目标的类别在平移变换下是不变的^[14]。

因此，在视觉表征中嵌入任务所需的不变性或对称性结构是实现计算机视觉系统稳健性和可解释性的一种重要的原则性保障。

1.2 研究现状和难点

回顾计算机视觉领域，对不变性或对称性的利用几乎贯穿领域本身的发展历史，宏观上跨越手工表征和学习表征两大研究路径^[15]：

手工表征。对平移、旋转、缩放等几何变换的不变性一直是研究者所考虑的基本原则，涵盖该领域几乎有所经典方法：从经典不变量为代表的全局特征^[16]，到尺度不变特征变换（Scale-Invariant Feature Transform, SIFT）等为代表的局部稀疏特征^[17]，再到 DAISY 为代表的局部稠密特征^[18]。在深度学习时代之前，这些方法不仅是图像处理等底层任务的主流解决方案，同时还与视觉词袋模型等联合用于图像理解的高层任务。然而，这些手工表征在设计上固定的、依赖于完备或欠完备的变换基，因而在现今的超大规模视觉任务（如 ImageNet 分类）中无法提供足够的判别性水平^[19]。

学习表征。基于深度学习的人工神经网络通过可学习非线性变换的级联实现了强大的判别性水平。其中，卷积神经网络（Convolutional Neural Networks, CNN）在超大规模视觉任务实现了显著优于手工表征性能的突破性进展。尽管深度学习领域的典型观点认为表征应形成于数据驱动的学习而非人工的设计^[3]，但 CNN 的成功很大程度上基于对不变性或对称性的利用^[14,20]。具体而言，即使有万能逼近定理的保障，全连接神经网络对图像数据的直接表征仍是不稳定且低效的，这是维度诅咒的典型案列；相反，CNN 引入了对称性先验（卷积的平移对称性^[21]），因而在很大程度上缓解了维度诅咒问题。另一方面，这些高度非线性方法仍然表现出严重的稳健性和解释性问题，例如，缺乏定制化引入任务先验（包括除平移外的其他几何对称性）的有效手段^[22]、对抗性扰动的存在及其解释^[23,24]等。因而，经典学习表征不能良好扩展到信任和安全攸关的视觉任务^[25]。

宏观难点。综合上述研究现状可知，本文宏观上面临如下困难：不变性方法和判别性方法分别在手工和学习路径上近乎独立的发展，这将导致研究者在迈向稳健和可解释的可信视觉任务时面临二者棘手的权衡。

具体难点。受以上宏观分析启发，本文聚焦于经典不变量理论的进一步发展，力图改善不变量的在现实世界可信视觉任务中的判别性。通过深入分析经典不变量的研究现状（在第二章展开），本文研究面临的困难又可具体为以下两方面判别性问题：

- 理论上，经典不变量通常基于图像的全局表征假设，而对于视觉任务中更具信息量的图像局部和层次表征，缺乏相应的理论扩展；
- 实践上，经典不变量通常用于小规模视觉任务，而对于具有不变性或对称性先验的更大规模视觉任务，缺乏相应的实践扩展。

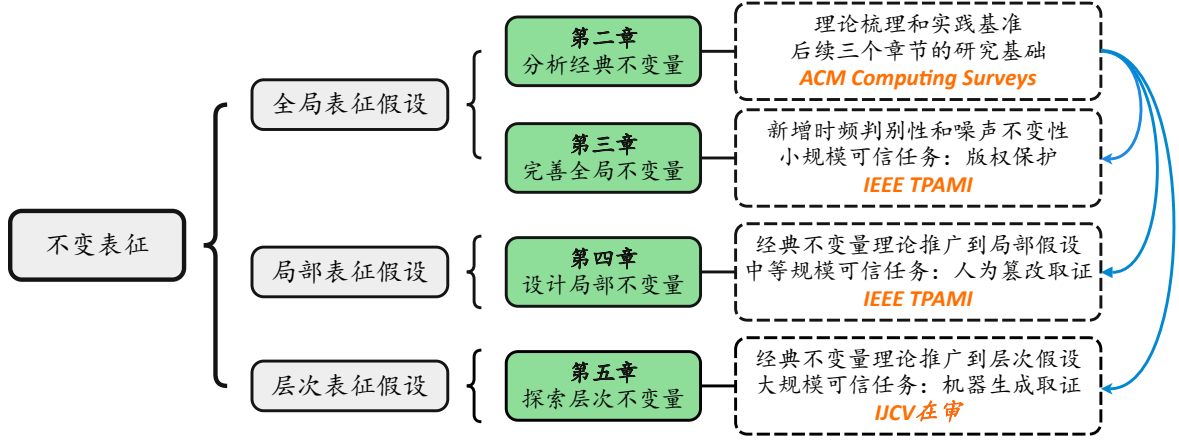


图 1.1 本文研究框架。

1.3 本文工作和创新

围绕以上难点，本文系统性分析了前人的不变表征研究，进一步提出了全局、局部和层次不变表征的新方法，并成功应用于多种信息安全和取证的可信任任务，如图 1.1 所示。

本文的主要研究成果及其创新性概括如下：

分析经典不变量。本文首先深入分析了经典不变量方法，为后续的三个创新性课题提供了理论和实践基础。首先，在理论层面回顾不变量的发展历史并梳理其中的核心知识，重点分析近期在快速/精确计算、稳健性/不变性优化、定义扩展和应用的成功经验，这些内容将在后续课题中得到进一步的理论扩展。此外，在实践层面提供了系统性的开源实现 MomentToolbox，并基于此给出了广泛的数值评估结果，这些内容将在后续课题中得到进一步的实践扩展。

完善全局不变量。针对经典不变量在含噪环境中的适用性问题，本文提出了一种具有时频分析能力的噪声不变表征方法。首先，将频域不变量推广为时频域不变量，表现出新的时频判别性，该推广基于分数阶多项式基和傅里叶基的研究。此外，还将几何不变量推广为几何与噪声联合不变量，从而更适用于非理想成像条件，该推广基于不变子空间的研究。以上内容是对全局不变量理论的完善，与基于数据增广训练或数据增强预处理的深度表征相比，其在小规模的经典视觉任务以及版权保护的信任任务中表现出稳健性、解释性和效率优势。

设计局部不变量。针对经典不变量在局部视觉任务中的适用性问题，本文提出了一种局部不变表征方法。首先，将经典不变量的定义从全局扩展到具有尺度空间的局部，形式化了其在几何不变性、等变性和协变性方面的核心结论。此外，推导出了高效的离散计算策略，具有低误差、常数阶复杂性以及基函数通用性的特点。以上内容设计了新的局部不变量理论，与先进的稠密局部表征相比，其在中等规模的经典视觉任务以及人为篡改取证的信任任务中表现出稳健性、解释性和效率优势。

探索层次不变量。针对经典不变量在大规模视觉任务中的适用性问题，本文提出了一种层

次不变表征方法。以不变量理论重新审视卷积神经网络中的典型模块，形式化了实现层次不变性的一般蓝图和具体定义，将上述全局和局部不变量的基本理论统一推广到层次假设。群论分析证明，提出方法在其任意中间层都表现出连续且直接的几何对称性。以上内容探索了新的层次不变量理论，与现有的经典不变量、学习表征网络、不变表征网络相比，其在更大规模的经典视觉任务以及机器生成取证的可信任任务中能够更好地权衡不变性和判别性，同时表现出解释性和效率优势。

本文的组织结构概括如下：

- 第一章绪论，概括本文的背景、难点以及工作创新。
- 第二章分析经典不变量，展开论述该领域的研究现状，其理论梳理和实践基准作为后续三个创新性课题的研究基础。
- 第三章完善全局不变量，将全局不变量推广到时频域含噪假设，由此改善不变表征方法在非理想成像条件下的适用性。
- 第四章设计局部不变量，将全局不变量的基本理论推广到更广义的局部假设，由此改善不变表征方法在局部视觉任务中的适用性。
- 第五章探索层次不变量，将上述全局和局部不变量的基本理论统一推广到层次假设，由此改善不变表征方法在大规模视觉任务中的适用性。
- 第六章总结，回顾全文内容并展望未来工作。

第二章 分析经典不变量

本学位论文的理论来源主要是经典不变量，即图像矩和矩不变量，该理论涵盖了现有图像不变表征方法的主要实现手段。本章全面分析经典不变量的基本概念和各个研究方向的研究现状，并首次给出系统性的开源测试基准和软件实现。这些理论和实践内容将在后续三个章节中得到进一步扩展。

2.1 历史回顾

1962 年，Hu 首次在计算机视觉和模式识别领域引入了经典不变量方法，其源自 Hilbert 的代数不变量研究^[26]。此后，经过 60 余年的研究，已经开发出了许多基于经典不变量的图像表示技术，在其领域实现了不同程度的成功：

- 1998 年，Mukundan 等人^[27]对当时的主要论文进行了综述，主要总结了几种经典矩函数的理论研究。
- 2006 年，Pawlak^[28]对不变量的重构和计算方面进行了全面综述，特别是引入了误差分析的系统性方法。
- 2007 年，Shu 等人^[29-31]分别对经典不变量的数学定义、不变性增强和数值计算进行了简要综述。
- 2009 年，Flusser 等人^[32]发表了经典不变量的首个专著，系统性地描述了二维假设（图像）下各种几何不变量的理论，这对经典不变量的发展做出了重大的贡献。该专著的扩展版本^[33]将二维假设进一步扩展至更高维（图形）。
- 2011 年，Hoang^[34]在其博士论文中具体综述了旋转不变量方法，涵盖了理论基础、表征性质和数值计算。上述大部分综述未涵盖近 10 年的最新进展。
- 2014 年，Papakostas 等人^[35]筛选出经典不变量研究的里程碑工作，并首次对各个研究方向的具体进展进行综述。然而，作者并未对这些新进展的理论基础进行抽象和总结。
- 2019 年，Kaur 等人^[36]对经典不变量进行了比较性综述，但仍然缺乏对各个方向最新进展的深入分析。上述综述文献的共同缺陷在于缺乏开源的测试基准和软件实现，这限制了该领域的进一步发展。

在本章，我们贡献是给出了不变量研究的系统性现状分析，特别是涵盖最新进展的深入梳理以及开源的测试基准和软件实现，这将对以上文献的有益补充，同时作为后续三个创新性课题的研究基础。本章从图像矩领域的基本理论和经典方法开始，特别关注背后的动机和成功经验（2.2 节）。随后，我们对各研究方向的最新进展进行了讨论，包括快速/准确计算、稳健

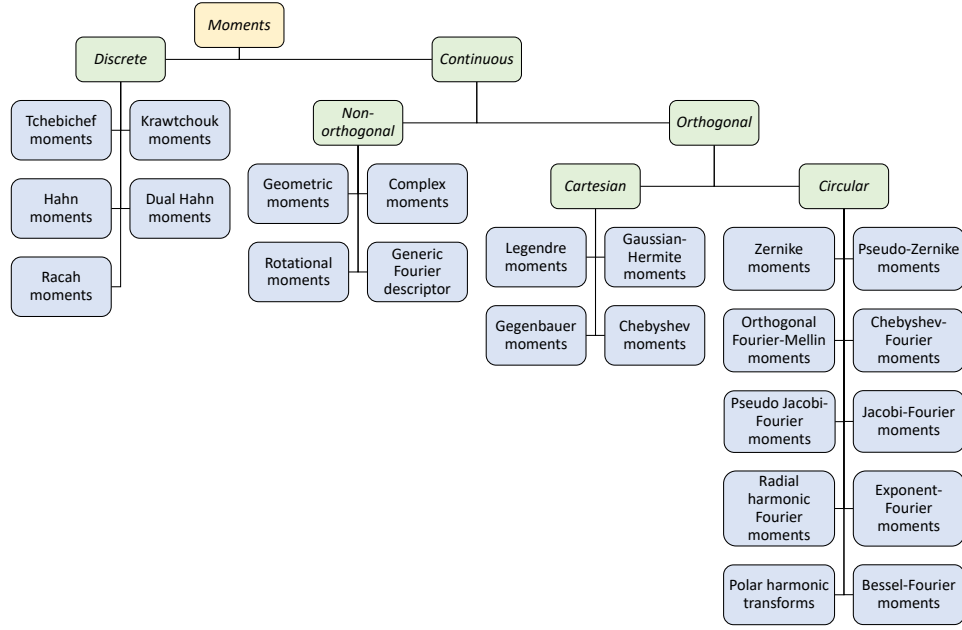


图 2.1 图像矩方法分类。

性/不变性优化、定义扩展和应用（2.3 节）。此外，我们报告了这些方法的比较结果以及开源实现（2.4 节）。最后，对本章内容进行总结（2.5 节）。

2.2 预备知识和经典方法

数学上，图像矩一般被定义为图像函数 f 与 $(n+m)$ 阶的基函数 V_{nm} 在区域 D 上的内积 $\langle f, V_{nm} \rangle$ [32]：

$$\langle f, V_{nm} \rangle = \iint_D V_{nm}^*(x, y) f(x, y) dx dy, \quad (2.1)$$

其中星号 $*$ 表示复共轭。直观上，图像矩集合 $\langle f, V_{nm} \rangle$ 是 f 在一组基函数 $\{V_{nm} : (n, m) \in \mathbb{Z}^2\}$ 形成的子空间上的投影 [34]。事实上，存在无限的基函数集合满足上述基本定义，研究者通常需要选择一组特殊的基函数 V_{nm} 使得 $\langle f, V_{nm} \rangle$ 表现出有益的图像表征性质。因而，根据基函数的数学性质，图像矩家族可以分为不同的类别，如图 2.1 所示。

首先，根据基函数是否满足正交性，图像矩可以分为正交矩和非正交矩。正交性指的是基函数集中的任意两个不同的基函数 V_{nm} 和 $V_{n'm'}$ 线性无关（或几何术语“垂直”），使得所形成的图像矩集合没有冗余信息。数学上，当满足以下条件时 V_{nm} 和 $V_{n'm'}$ 正交：

$$\langle V_{nm}, V_{n'm'} \rangle = \iint_D V_{nm}(x, y) V_{n'm'}^*(x, y) dx dy = \delta_{nn'} \delta_{mm'}, \quad (2.2)$$

表 2.1 正交圆谐矩的径向基函数汇总。

正交圆谐矩	径向基函数定义
ZM	$R_{nm}^{(ZM)}(r) = \sqrt{\frac{n+1}{\pi}} \sum_{k=0}^{\frac{n- m }{2}} \frac{(-1)^k (n-k)! r^{n-2k}}{k! (\frac{n+ m }{2}-k)! (\frac{n- m }{2}-k)!}$
PZM	$R_{nm}^{(PZM)}(r) = \sqrt{\frac{n+1}{\pi}} \sum_{k=0}^{n- m } \frac{(-1)^k (2n+1-k)! r^{n-k}}{k! (n+ m +1-k)! (n- m -k)!}$
OFMM	$R_n^{(OFMM)}(r) = \sqrt{\frac{n+1}{\pi}} \sum_{k=0}^n \frac{(-1)^{n+k} (n+k+1)! r^k}{k! (n-k)! (k+1)!}$
CHFM	$R_n^{(CHFM)}(r) = \frac{2}{\pi} \left(\frac{1-r}{r}\right)^{\frac{1}{4}} \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} \frac{(-1)^k (n-k)! (4r-2)^{n-2k}}{k! (n-2k)!}$
PJFM	$R_n^{(PJFM)}(r) = \sqrt{\frac{(n+2)(r-r^2)}{\pi(n+3)(n+1)}} \sum_{k=0}^n \frac{(-1)^{n+k} (n+k+3)! r^k}{k! (n-k)! (k+2)!}$
JFM	$R_n^{(JFM)}(p, q, r) = \sqrt{\frac{r^{q-2}(1-r)^{p-q}(p+2n) \cdot \Gamma(q+n) \cdot n!}{2\pi \Gamma(p+n) \cdot \Gamma(p-q+n+1)}} \sum_{k=0}^n \frac{(-1)^k \Gamma(p+n+k) r^k}{k! (n-k)! \Gamma(q+k)}$
RHFM	$R_n^{(RHFM)}(r) = \begin{cases} \frac{1}{\sqrt{2\pi r}} & n = 0 \\ \sqrt{\frac{1}{\pi r}} \sin(\pi(n+1)r) & n > 0 \text{ \& } n \text{ odd} \\ \sqrt{\frac{1}{\pi r}} \cos(\pi n r) & n > 0 \text{ \& } n \text{ even} \end{cases}$
EFM	$R_n^{(EFM)}(r) = \frac{1}{\sqrt{2\pi r}} \exp(j2n\pi r)$
PCET	$R_n^{(PCET)}(r) = \frac{1}{\sqrt{\pi}} \exp(j2n\pi r^2)$
PCT	$R_n^{(PCT)}(r) = \begin{cases} \frac{1}{\sqrt{\pi}} & n = 0 \\ \sqrt{\frac{2}{\pi}} \cos(n\pi r^2) & n > 0 \end{cases}$
PST	$R_n^{(PST)}(r) = \sqrt{\frac{2}{\pi}} \sin(n\pi r^2)$
BFM	$R_n^{(BFM)}(r) = \frac{1}{\sqrt{\pi} J_{v+1}(\lambda_n)} J_v(\lambda_n r), J_v(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k! \Gamma(v+k+1)} \left(\frac{x}{2}\right)^{v+2k}$

其中 δ_{ij} 是克罗内克函数，定义为：

$$\delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}. \quad (2.3)$$

非正交矩中，知名方法包括几何矩^[32] (Geometric Moments)、旋转矩^[37] (Rotational Moments)、复数矩^[38] (Complex Moments) 和广义傅立叶描述子^[39] (Generic Fourier Descriptor)。由于基函数的非正交性，基于这些矩的不变表征存在较高的信息冗余，缺乏可逆性（难以实现图像重构）和完备性（表征的判别性不足）。因此，正交性是设计基函数时需满足的一般性要求。

其次，根据基函数是否连续，图像矩可以分为连续矩和离散矩。在二维（图像）假设下，连续和离散矩的基函数分别定义在二维的实数值和整数值空间中，即域 $D \in \mathbb{R}^2$ 和 $D \in \mathbb{Z}^2$ 。当计算数字图像的连续矩时，通常需要引入适当的离散化以近似连续积分，并随之产生不同程度的计算误差。在后续章节，我们将进一步描述此类误差的原因和解决方案。对于离散矩而言，如离散切比雪夫矩^[40] (Tchebichef Moments)、克拉夫丘克矩^[41] (Krawtchouk Moments)、哈恩矩^[42] (Hahn Moments)、双哈恩矩^[43] (Dual Hahn Moments)、拉卡矩^[44] (Racah Moments)，其计算则

不涉及任何近似误差。因此，离散矩更适用于高精度的图像处理任务，例如图像重构、压缩和去噪。

最后，根据定义基函数的坐标系，图像矩可以分为笛卡尔矩和圆谐矩。对于连续矩而言，笛卡尔矩的基函数定义在 $D = \{(x, y) : x \in (-\infty, +\infty), y \in (-\infty, +\infty)\}$ 或 $D = \{(x, y) : x \in [-1, 1], y \in [-1, 1]\}$ 上，而圆谐矩的定义域为 $D = \{(r, \theta) : r \in [0, +\infty), \theta \in [0, 2\pi)\}$ 或 $D = \{(r, \theta) : r \in [0, 1], \theta \in [0, 2\pi)\}$ （即单位圆）。根据 Bhatia 等人的证明^[45]，对图像关于原点 $(x, y) = (0, 0)$ 旋转不变的图像矩需定义在极坐标，且具有以下形式：

$$V_{nm}(r \cos \theta, r \sin \theta) \equiv V_{nm}(r, \theta) = R_n(r)A_m(\theta), \quad (2.4)$$

其中角向基函数 $A_m(\theta) = \exp(jm\theta)$ ($j = \sqrt{-1}$)，径向基函数 $R_n(r)$ 可以是任意形式^[34]。令 f_R 为原始图像 f 关于中心的旋转版本，当 V_{nm} 符合上述形式时，必然存在一个函数 I （除常值函数外）使得

$$I(\{f, V_{nm}\}) \equiv I(\{f_R, V_{nm}\}), \quad (2.5)$$

即满足旋转不变性。这种旋转不变量是一种经典的矩不变量。

对于笛卡尔矩而言，如勒让德矩^[46]（Legendre Moments）、高斯-埃尔米特矩^[47]（Gaussian-Hermite Moments）、盖根鲍尔矩^[48]（Gegenbauer Moments）、连续切比雪夫矩^[49]（Chebyshev Moments）以及上述离散矩，在实现旋转不变性方面存在困难。由于数字图像定义在笛卡尔坐标系而圆谐矩定义在极坐标系，计算数字图像的圆谐矩时通常需要引入坐标转换，并由此产生不同程度的计算误差。在后续章节，我们将进一步描述此类误差的原因和解决方案。

基于上述预备知识可知，正交性和旋转不变性结构是设计经典不变量方法的重要原则。因此，接下来将对正交圆谐矩进行综述。

从公式（2.4）可以得知，圆谐矩的基函数是可分离的，则在正交圆谐矩中公式（2.2）可以重写为：

$$\begin{aligned} \langle V_{nm}, V_{n'm'} \rangle &= \int_0^{2\pi} \int_0^1 R_n(r)A_m(\theta)R_{n'}^*(r)A_{m'}^*(\theta)rdrd\theta \\ &= \int_0^{2\pi} A_m(\theta)A_{m'}^*(\theta)d\theta \int_0^1 R_n(r)R_{n'}^*(r)rdr \\ &= 2\pi\delta_{mm'} \int_0^1 R_n(r)R_{n'}^*(r)rdr. \end{aligned} \quad (2.6)$$

由于 $\langle V_{nm}, V_{n'm'} \rangle = \delta_{nn'} \delta_{mm'}$ ，径向基函数 $R_n(r)$ 应满足以下加权正交性条件：

$$\int_0^1 R_n(r) R_{n'}^*(r) r dr = \frac{1}{2\pi} \delta_{nn'}. \quad (2.7)$$

公式 (2.7) 是在设计正交圆谐矩时需满足的一般性要求。角向基函数 $A_m(\theta)$ 由于 Bhatia 等人^[45] 的证明，具有固定形式 $\exp(jm\theta)$ ，这意味着现有正交圆谐矩方法的主要差异在于径向基函数的定义。在这方面，主要有三种类型的正交函数可用作定义，包括雅可比多项式、谐波函数和本征函数。接下来，我们将简要介绍这三组方法，并给出它们的径向基函数的归一化定义，即直接满足公式 (2.7)。

2.2.1 雅可比多项式

以雅可比多项式定义径向基函数的正交圆谐矩主要包括：泽尔尼克矩^[46] (Zernike Moments, ZM)、伪泽尔尼克矩^[37] (Pseudo-Zernike Moments, PZM)、正交傅里叶-梅林矩^[50] (Orthogonal Fourier-Mellin Moments, OFMM)、切比雪夫-傅里叶矩^[51] (Chebyshev-Fourier Moments, CHFM)、伪雅可比-傅里叶矩^[52] (Pseudo Jacobi-Fourier Moments, PJFM)、雅可比-傅里叶矩^[53] (Jacobi-Fourier Moments, JFM)。表 2.1 汇总了这些方法径向基函数的定义。

注意，JFM 的径向基函数是直接从原始雅可比多项式^[45] 构造，而 ZM、PZM、OFMM、CHFM 和 PJFM 的径向基函数实际上都是雅可比多项式的特殊情况。因此，JFM 是上述方法的推广定义：通过设置参数 p 和 q 的值，可以直接从 $R_{nm}^{(JFM)}(p, q, r)$ 获得 $R_{nm}^{(OFMM)}(r)$ 、 $R_{nm}^{(CHFM)}(r)$ 和 $R_{nm}^{(PJFM)}(r)$ ； $R_{nm}^{(JFM)}(p, q, r)$ 与 $R_{nm}^{(ZM)}(r)/R_{nm}^{(PZM)}(r)$ 之间的关系更加复杂，请参考 Hoang 等人的工作^[54] 以获取更多细节。这里，参数设置总结如下：

- ZM —— $p = |m| + 1, q = |m| + 1$;
- PZM —— $p = 2|m| + 2, q = 2|m| + 2$;
- OFMM —— $p = 2, q = 2$;
- CHFM —— $p = 2, q = 1.5$;
- PJFM —— $p = 4, q = 3$;

基于雅可比多项式的正交圆谐矩，特别是标志性的 ZM，在光学、数字图像处理和模式识别领域具有悠久的历史。然而，从上述定义可以看出，这些径向基函数依赖于阶乘或伽马项的求和，导致高昂的计算成本。此外，阶乘或伽马函数往往会引起计算误差，产生数值不稳定性。在 2.3 节中，我们将进一步描述该问题的成因和解决方案。

表 2.2 正交圆谐矩的径向基函数性质。

正交圆谐矩	参数	复杂度	稳定性	零点数量	零点分布
ZM	$n \in \mathbb{N}, m \in \mathbb{Z}, m \leq n,$ $n - m = \text{偶数}$	高	差	$(n - m)/2$	有偏
PZM	$n \in \mathbb{N}, m \in \mathbb{Z}, m \leq n$	高	差	$n - m $	有偏
OFMM	$n \in \mathbb{N}, m \in \mathbb{Z}$	高	差	n	均匀
CHFM	$n \in \mathbb{N}, m \in \mathbb{Z}$	高	差	n	均匀
PJFM	$n \in \mathbb{N}, m \in \mathbb{Z}$	高	差	n	均匀
JFM	$n \in \mathbb{N}, m \in \mathbb{Z}, p, q \in \mathbb{R},$ $p - q > -1, q > 0$	高	差	n	均匀
RHFM	$n \in \mathbb{N}, m \in \mathbb{Z}$	低	中	n	均匀
EFM	$n \in \mathbb{Z}, m \in \mathbb{Z}$	低	中	$2n$	均匀
PCET	$n \in \mathbb{Z}, m \in \mathbb{Z}$	低	好	$2n$	有偏
PCT	$n \in \mathbb{N}, m \in \mathbb{Z}$	低	好	n	有偏
PST	$n \in \mathbb{N}^+, m \in \mathbb{Z}$	低	好	$n - 1$	有偏
BFM	$n \in \mathbb{N}, m \in \mathbb{Z}, v \in \mathbb{R},$ $\lambda_n = J_v(x)$ 的第 n 个零点	很高	中	n	均匀

2.2.2 谐波函数

以谐波函数定义径向基函数的正交圆谐矩主要包括：圆谐-傅里叶矩^[55]（Radial Harmonic Fourier Moments, RHFM）、指数-傅里叶矩^[56]（Exponent-Fourier Moments, EFM）和极坐标谐波变换^[57]（Polar Harmonic Transforms, PHT）。这里，PHT 包括三种不同的变换：极坐标复指数变换（Polar Complex Exponential Transform, PCET）、极坐标余弦变换（Polar Cosine Transform, PCT）和极坐标正弦变换（Polar Sine Transform, PST）。表 2.1 汇总了这些方法径向基函数的定义。

可以看出，RHFM、EFM 和 PHT 属于调和分析方法，基于复指数函数 $\{\exp(j2n\pi r) : n \in \mathbb{Z}\}$ 和三角函数 $\{1, \cos(2n\pi r), \sin(2n\pi r) : n \in \mathbb{Z}^+\}$ 。上述径向基函数定义密切相关，可以通过欧拉公式 $\exp(j\alpha) = \cos(\alpha) + j\sin(\alpha)$ 和变量替换 $dr = \frac{1}{2}dr^2$ 实现相互转化。计算方面，与基于雅可比多项式的正交圆谐矩相比，此类方法不涉及任何复杂的阶乘或伽玛项以及求和，意味着更低的时间复杂性和更好的数值稳定性。

2.2.3 本征函数

目前，基于本征函数的正交圆谐矩还比较少，主要包括贝塞尔-傅里叶矩^[58]（Bessel-Fourier Moments, BFM），其径向基函数定义见表 2.1。可以看到，该径向基函数依赖于无穷级数和阶乘/伽玛项，还需要根据第一类贝塞尔函数 $J_v(x)$ 的第 n 个零点确定参数 λ_n 。因此，与基于雅可比多项式和谐波函数的方法相比，此方法在理论上具有更高的复杂性。

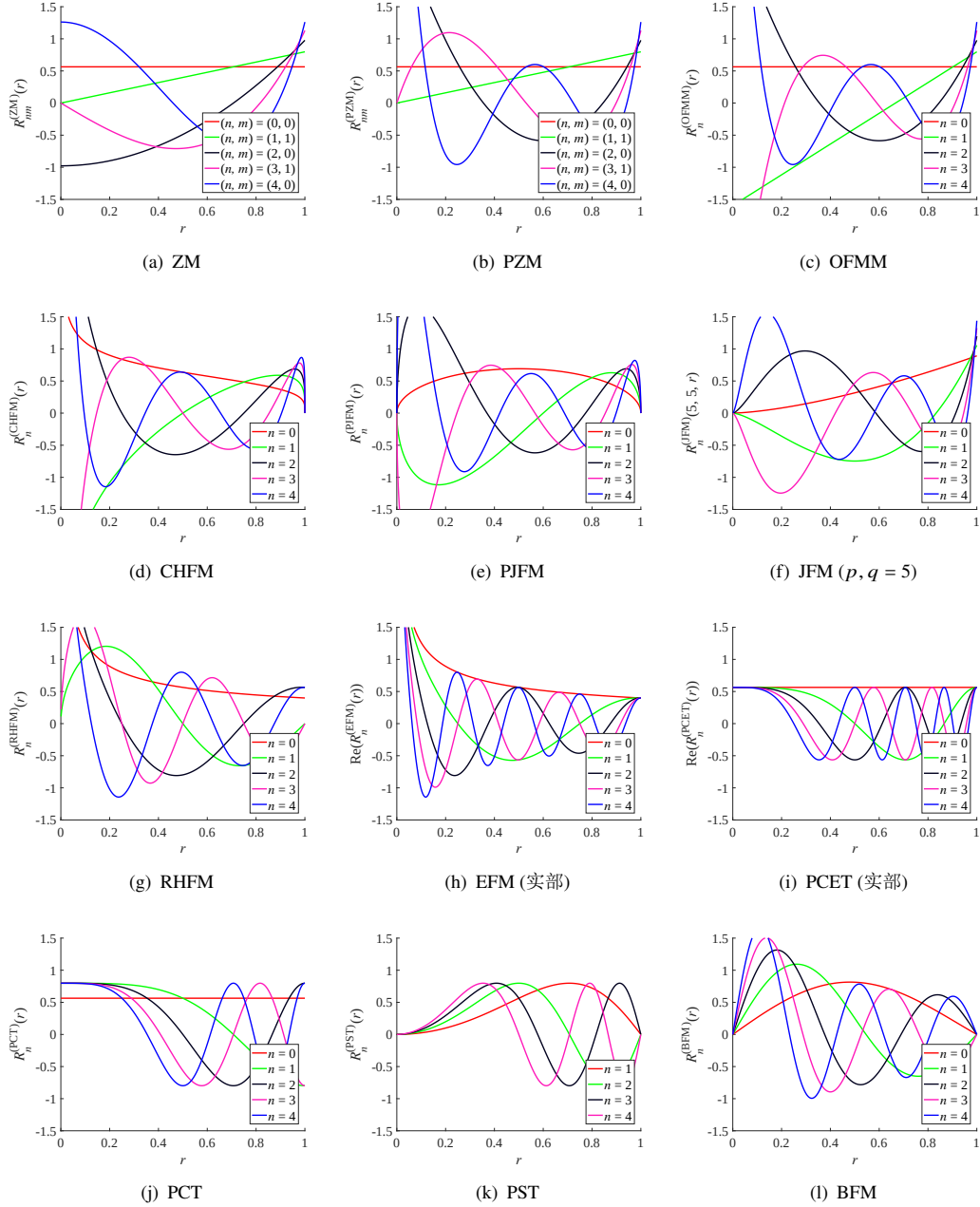


图 2.2 正交圆谐矩的径向基函数示例。

2.2.4 总结和讨论

为了更好地可视化，我们在图 2.2 中给出了正交圆谐矩的径向基函数示例。此外，我们还在表 2.2 中分析了这些基函数的数学属性，包括参数、计算复杂度、数值稳定性、零点数量和零点分布。基函数的复杂度（仅考虑基于定义的直接计算）分为低、高和非常高，取决于定义中是否涉及阶乘/伽玛项、求和/级数运算和求根过程。数值稳定性分为差、中和好，取决于函数是否包含阶乘/伽玛项和非常高的函数绝对值（特别是无界函数）。径向基函数的零点数量与捕

捉图像频率信息的能力有关^[34,57]。径向基函数的零点分布是一项关键属性，它与图像平面上的表征重点有关^[34,57]。当判别性信息在图像平面上均匀分布时，对某些区域的不合理强调将对表征的判别性产生负面影响，被称为信息抑制问题^[38]。

从表 2.2 可以看出，几乎没有一种正交圆谐矩方法同时具有低计算复杂度、良好数值稳定性以及无偏零点分布。这一事实促进了正交圆谐矩的新设计和改进，这些内容将在第 2.3 节具体讨论。

2.3 研究方向和最新进展

现有的工作主要集中在优化和扩展上述经典方法。如图 2.3 所示，目前的研究方向主要包括：精确计算、快速计算、稳健性/不变性优化、定义扩展、以及应用。本节将分别介绍上述方向和最新进展。

2.3.1 精确计算

作为本节和下一节的共同数学背景，我们在这里首先给出计算图像矩的一般过程。

回到公式 (2.1)，为了在离散笛卡尔网格上计算数字图像 $\{f(i, j) : (i, j) \in [1, 2, \dots, N]^2\}$ 的矩 $\langle f, V_{nm} \rangle$ ，首先需要统一 $f(i, j)$ 和 $V_{nm}(x, y)$ 的定义域，即设计两者之间的一个映射关系：

$$(i, j) \rightarrow (x_i, y_j), \quad (2.8)$$

其中以 (i, j) 为中心的像素区域 $[i - \frac{\Delta i}{2}, i + \frac{\Delta i}{2}] \times [j - \frac{\Delta j}{2}, j + \frac{\Delta j}{2}]$ 被映射到以 (x_i, y_j) 为中心的区域 $[x_i - \frac{\Delta x_i}{2}, x_i + \frac{\Delta x_i}{2}] \times [y_j - \frac{\Delta y_j}{2}, y_j + \frac{\Delta y_j}{2}]$ 。基于坐标映射 $(i, j) \rightarrow (x_i, y_j)$ ，公式 (2.1) 可以重写为如下离散形式：

$$\langle f, V_{nm} \rangle = \sum_{(x_i, y_j) \in D} h_{nm}(x_i, y_j) f(i, j), \quad (2.9)$$

其中 $h_{nm}(x_i, y_j)$ 是基函数 V_{nm} 在映射像素区域 $[x_i - \frac{\Delta x_i}{2}, x_i + \frac{\Delta x_i}{2}] \times [y_j - \frac{\Delta y_j}{2}, y_j + \frac{\Delta y_j}{2}]$ 上的积分值，定义为：

$$h_{nm}(x_i, y_j) = \int_{x_i - \frac{\Delta x_i}{2}}^{x_i + \frac{\Delta x_i}{2}} \int_{y_j - \frac{\Delta y_j}{2}}^{y_j + \frac{\Delta y_j}{2}} V_{nm}^*(x, y) dx dy. \quad (2.10)$$

一般来说，正交圆谐矩的计算受到几何误差、数值积分误差和表示误差（主要是数值不稳定性）的影响。这些误差将严重限制表征质量，特别是对于描述图像细节的高阶矩。因此，精确计算策略对于正交圆谐矩的实际应用至关重要。

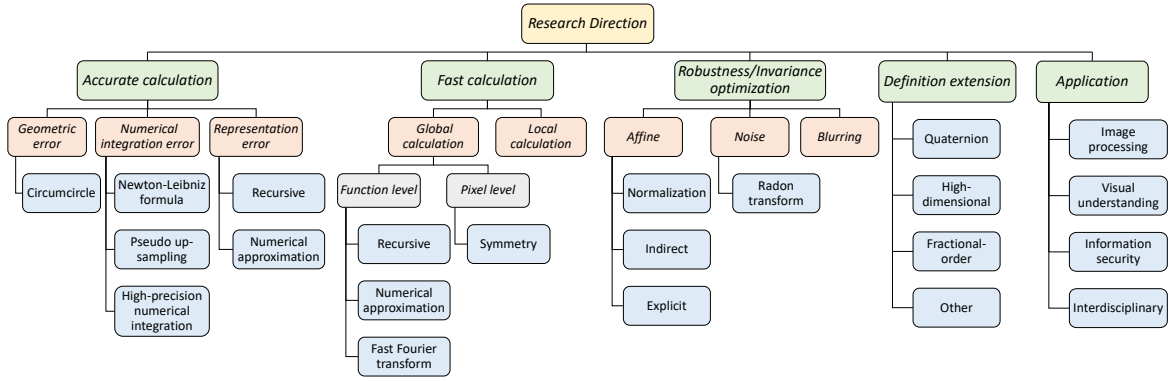


图 2.3 图像矩和矩不变量的研究方向。

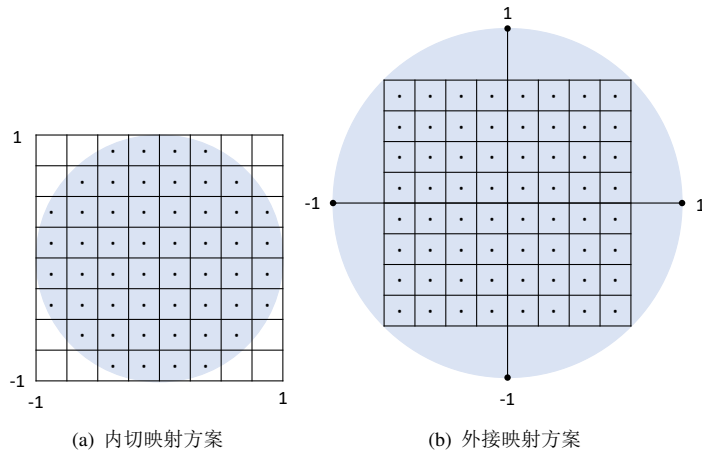


图 2.4 数字图像的矩形区域与基函数的单位圆区域之间的映射关系。

2.3.1.1 几何误差

当图像域映射到基函数域时可能会产生几何误差，即公式 (2.8)。由于数字图像通常定义于矩形区域而非单位圆，所以在计算圆谐矩时几何误差很常见。

关于矩形区域和单位圆之间的映射关系，文献中存在着如图 2.4 所示的内切映射方案^[59]：

$$\begin{cases} i \rightarrow x_i = \frac{2i-N}{N} = i\Delta x_i - 1 \\ j \rightarrow y_j = \frac{2j-N}{N} = j\Delta y_j - 1 \end{cases}, \quad (2.11)$$

和外接映射方案^[60]：

$$\begin{cases} i \rightarrow x_i = \frac{2i-N}{\sqrt{2}N} = i\Delta x_i - \frac{\sqrt{2}}{2} \\ j \rightarrow y_j = \frac{2j-N}{\sqrt{2}N} = j\Delta y_j - \frac{\sqrt{2}}{2} \end{cases}. \quad (2.12)$$

当使用内切映射方案时，存在 (i, j) 使得 $(x_i, y_j) \notin D$ ，即部分像素完全映射到基函数的单位圆区域之外，同时存在 (i, j) 使得 $[x_i - \frac{\Delta x_i}{2}, x_i + \frac{\Delta x_i}{2}] \times [y_j - \frac{\Delta y_j}{2}, y_j + \frac{\Delta y_j}{2}] - D \neq \emptyset$ ，即部分像素

区域与基函数的单位圆区域部分相交。在这两种情况下，会出现几何误差。

至于外接映射方案，它能完全避免上述两种情况因此不会产生几何误差。然而，这以表征能力的损失为代价^[34]，因为包含图像判别性信息的区域仅占整个单位圆的 $\frac{2}{\pi}$ 。

2.3.1.2 数值积分误差

当计算连续基函数的积分时可能会产生数值积分误差，即公式 (2.10)。在这里，离散矩的基函数 V_{nm} 在区间 $[x_i - \frac{\Delta x_i}{2}, x_i + \frac{\Delta x_i}{2}] \times [y_j - \frac{\Delta y_j}{2}, y_j + \frac{\Delta y_j}{2}]$ 上通常是常数。因此，很容易得出 $h_{nm}(x_i, y_j) = V_{nm}^*(x_i, y_j) \Delta x_i \Delta y_j$ ，即离散矩不涉及数值积分误差^[40]。而对于连续矩，由于基函数 V_{nm} 在区间 $[x_i - \frac{\Delta x_i}{2}, x_i + \frac{\Delta x_i}{2}] \times [y_j - \frac{\Delta y_j}{2}, y_j + \frac{\Delta y_j}{2}]$ 上连续，求解 $h_{nm}(x_i, y_j)$ 就需要一些积分技巧。

作为最简单的情况， $h_{nm}(x_i, y_j)$ 存在解析解，即公式 (2.10) 可以直接以牛顿-莱布尼兹公式^[61,62] 求解。这样的计算也不涉及数值积分误差。

作为更一般的情况，鉴于许多基函数 V_{nm} 定义的复杂性，往往很难确定解析解，通常需要一些近似算法给出 $h_{nm}(x_i, y_j)$ 的数值解。最常用的近似算法是零阶近似^[59]，类似于离散矩的计算过程：

$$h_{nm}(x_i, y_j) \simeq V_{nm}^*(x_i, y_j) \Delta x_i \Delta y_j. \quad (2.13)$$

一般来说，数值积分方法的精度与区间面积 $\Delta x_i \Delta y_j$ 成反比。因此，通过将单个像素区域进一步细分为多个更小的积分区间（例如， 3×3 个子区间）可以很直接地获得更高的精度。这种策略通常被称为伪上采样^[63,64]。当对这些子区间使用零阶近似时，积分可以表示为：

$$h_{nm}(x_i, y_j) \simeq \sum_{(a,b)} V_{nm}^*(u_a, v_b) \Delta u_a \Delta v_b, \quad (2.14)$$

其中 (u_a, v_b) 是采样点，有 $u_a \in [x_i - \frac{\Delta x_i}{2}, x_i + \frac{\Delta x_i}{2}]$ ， $v_b \in [y_j - \frac{\Delta y_j}{2}, y_j + \frac{\Delta y_j}{2}]$ 。除了这种简单方法外，还可以使用其他更复杂的高精度数值积分策略^[65-67] 来计算公式 (2.10)，如高斯法和辛普森法。这类方法的一般定义是：

$$h_{nm}(x_i, y_j) \simeq \sum_{(a,b)} w_{ab} V_{nm}^*(u_a, v_b) \Delta x_i \Delta y_j, \quad (2.15)$$

其中， w_{ab} 是与采样点 (u_a, v_b) 对应的权重，由具体的高精度数值积分策略给出。

值得注意的是，我们以上讨论的内容都基于笛卡尔坐标系，其普遍适用于各种矩的计算。相应地，还存在一种基于极坐标系的计算方案，专门适用于正交圆谐矩，称为极坐标像素平铺^[68,69]。

首先，该方案对数字图像 $f(i, j)$ 进行重采样，得到如图 2.5 所示的离散极坐标网格 (r_u, θ_{uv}) 。

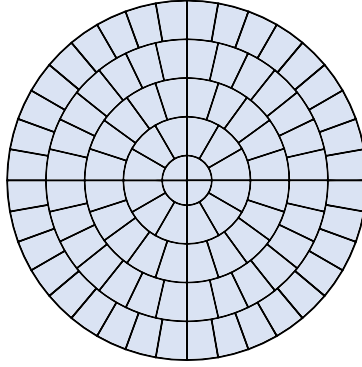


图 2.5 极坐标像素平铺方案。

然后，按照类似于公式 (2.9) 和 (2.10) 的方式进行计算：

$$\langle f, V_{nm} \rangle = \sum_{(r_u, \theta_{uv}) \in D} h_{nm}(r_u, \theta_{uv}) f(r_u, \theta_{uv}), \quad (2.16)$$

其中

$$h_{nm}(r_u, \theta_{uv}) = \int_{r_u}^{r_{u+1}} \int_{\theta_{uv}}^{\theta_{u(v+1)}} V_{nm}^*(r, \theta) r dr d\theta = \int_{r_u}^{r_{u+1}} R_n(r) r dr \int_{\theta_{uv}}^{\theta_{u(v+1)}} \exp(-jm\theta) d\theta. \quad (2.17)$$

注意，公式 (2.17) 具有一个明显的优势，即将 $h_{nm}(r_u, \theta_{uv})$ 的计算分为两个独立的部分^[70]：1) $\int_{r_u}^{r_{u+1}} R_n(r) r dr$ 可以通过伪上采样或高精度数值积分方法进行近似；2) $\int_{\theta_{uv}}^{\theta_{u(v+1)}} \exp(-jm\theta) d\theta$ 可以直接通过牛顿-莱布尼茨公式进行精确积分，得到：

$$\int_{\theta_{uv}}^{\theta_{u(v+1)}} \exp(-jm\theta) d\theta = \begin{cases} \frac{j[\exp(-jm\theta_{u(v+1)}) - \exp(-jm\theta_{uv})]}{m} & m \neq 0 \\ \theta_{u(v+1)} - \theta_{uv} & m = 0 \end{cases}. \quad (2.18)$$

除了上述的优势之外，极坐标像素平铺方案在几何误差和数值积分误差方面与笛卡尔方案具有类似的特性，这里不再赘述。

2.3.1.3 表示误差

表示误差由数值计算系统的精度限制所引起，在计算的全过程中都会发生，即公式 (2.8)、(2.9) 和 (2.10)。表示误差又可以进一步分为上溢误差，下溢误差和舍入误差^[34]。常见的数值不稳定性主要归因于舍入误差和上溢误差。

基于雅各比多项式的基函数包含阶乘或伽玛项。当阶数较大时，这些项的实际值可能超出数值计算系统的表示范围，从而导致舍入误差甚至上溢误差。为了避免直接计算阶乘或伽玛项，

可以使用递归策略^[71-73]或数值逼近策略^[74,75]。递归策略基于不同阶数基函数之间的递归关系,该关系由 $a! = a \cdot (a-1)!$ 或 $\Gamma(a) = a \cdot \Gamma(a-1)$ 导出,即可以使用几个低阶基函数直接推导出高阶基函数,不涉及大数的阶乘或伽玛计算。数值逼近策略则引入适当的逼近公式,如斯特林公式,实现了无阶乘或伽玛的计算。

除了阶乘或伽玛项之外,一些基函数在某些点上的绝对值非常高(甚至无穷),这也可能超出数值计算系统的表示范围。例如,尽管 RHFM 和 EFM 的定义不涉及任何阶乘或伽玛项,但它们的基函数绝对值在零点处是无穷大的(见图 2.2),这将导致舍入误差甚至上溢误差。

2.3.1.4 精确计算的最新进展

在正交圆谐矩的精确计算领域,Xin 等人的极坐标像素平铺^[68]可能是最具影响力和代表性的工作。最近的论文通常将极坐标像素平铺与其他技术相结合,包括外接映射、伪上采样、高精度数值积分、牛顿-莱布尼茨公式和递归策略等,以实现最先进的计算精度。

在这方面,Sáez-Landete^[76]整合了 Camacho-Bello 等人^[70]和 Upneja 等人^[73]的工作,为 JFM 提供了精确计算方案。注意,由于 JFM 的广义性质,该方案也直接适用于其他基于雅各比多项式的正交圆谐矩,如 OFMM 和 CHFM。基于极坐标像素平铺,可以使用高斯积分、伪上采样和递归关系来计算复杂的径向部分,而角度部分则通过牛顿-莱布尼茨公式直接精确计算。Hosny 等人^[77]也提出了一种类似的方法用于 PHT,其中径向部分也通过牛顿-莱布尼茨公式进行精确积分。

这两个工作提供了几近完美的精度,分别用于计算基于雅各比多项式和基于谐波函数的正交圆谐矩。唯一的缺点是在将图像重采样到离散极坐标网格^[68]的过程中引入的误差。如果使用非线性插值方法,例如双立方插值法,对于图像表征任务来说,插值误差基本可以忽略不计。另一方面,如果给定的任务对这种误差敏感,其他更复杂的数学工具,如伪极坐标^[78],对于设计更精确计算^[79]具有指导意义。

2.3.2 快速计算

在正交圆谐矩的计算中,当需要大量矩、需要高分辨图像、需要处理大量图像、或需要高精度计算时,整体的复杂性可能会变得很高。由于这些要求在实际应用中很常见,因此需要发展正交圆谐矩的快速计算方案。根据应用场景的不同,对计算速度进行优化又可以分为两类:全局计算和局部计算。

2.3.2.1 全局计算

全局计算是指对从整幅图像计算矩。在这种情况下,矩的数量和图像的分辨率是影响时间复杂度的主要因素。以基于零阶近似的直接计算为例,一个 n 阶图像矩的时间成本来自于:

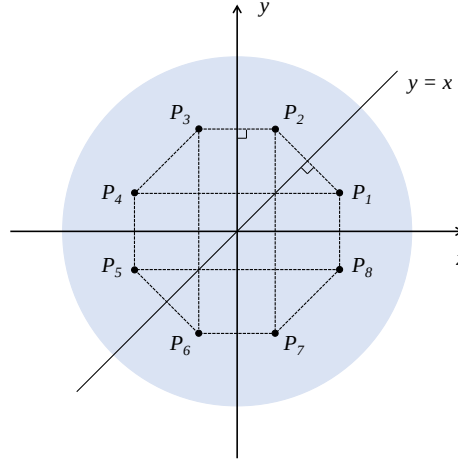


图 2.6 单位圆中的对称点。

- 对于公式 (2.10)，需要在 $N \times N$ 个采样点上估计基函数 V_{nm} 的值，即 $\{V_{nm}^*(x_i, y_j) : (i, j) \in [1, 2, \dots, N]^2\}$ 。其时间复杂度与基函数的定义有关。例如，基于雅可比多项式的基函数需要进行 $O(nN^2)$ 次加法，基于谐波函数的基函数则不涉这样的加法运算。
- 对于公式 (2.9)，需要计算基函数 V_{nm} 与数字图像 f 在 $N \times N$ 个采样点上的内积，即 $\langle f(i, j), V_{nm}^*(x_i, y_j) \Delta x_i \Delta y_j \rangle$ 。其时间复杂度是 $O(N^2)$ 次乘法和 $O(N^2)$ 次加法。

注意，上述计算仅针对单个矩。令 K 为一个整数常数，若计算集合 $\{(n, m) : |n|, |m| \leq K\}$ 中所有阶数的矩，总复杂度将扩大 $O(K^2)$ 倍。此外，如果采用高精度数值积分策略（例如高斯法），公式 (2.10) 的计算成本还会大幅上升。为了降低公式 (2.9) 和 (2.10) 的复杂度，现有的方法分别在函数层面和像素层面进行加速设计。

- 函数层面：对于基于雅可比多项式的矩，如果使用递归方法或数值逼近来估计基函数的值，则计算公式 (2.10) 中的加法（来自求和）和乘法（来自阶乘）数量可以大大减少^[71-75]。例如，对于多项式 $R_n(\sqrt{x_i^2 + y_j^2})$ 的递归计算需要 $O(N^2)$ 次加法，少于直接计算的 $O(nN^2)$ 次加法。如果需要 K 个阶数为 $\{0, 1, \dots, K\}$ 的多项式 $R_n(\sqrt{x_i^2 + y_j^2})$ ，直接计算需要 $O(K^2 N^2)$ 次加法，而递归计算仅需要 $O(KN^2)$ 次加法。对于基于谐波函数的矩，除了采用类似的递归计算^[80-83]，更有效的策略是利用它与傅里叶变换^[84-88]之间的内在联系。一旦确定了两者的显式关系，就可以使用快速傅里叶变换算法来计算矩，能够将公式 (2.9) 中最耗时的乘法从 $O(N^2)$ 次减少到 $O(N \log N)$ 次。
- 像素层面：主要借助定义域中基函数的对称性和反对称性来简化计算^[89,90]。具体地说，全部 $N \times N$ 个采样点上的基函数值可以完全由少数特殊的采样点推导出来，从而降低了公式 (2.9) 和 (2.10) 的复杂度。在图 2.6 中，我们给出了单位圆中对称点的示例。对于点 P_1 来说，存在七个关于坐标轴、原点和 $y = x$ 对称的点 $P_2, P_3, P_4, P_5, P_6, P_7, P_8$ 。它们的笛卡尔坐标和极坐标列在表 2.3 中。可以看出，所有这些点都具有相同的径向坐标和关联

表 2.3 对称点的笛卡尔坐标和极坐标。

对称点	对称轴	笛卡尔坐标	极坐标
P_1		(x, y)	(r, θ)
P_2	$y = x$	(y, x)	$(r, \frac{\pi}{2} - \theta)$
P_3	$y = x, y$ 轴	$(-y, x)$	$(r, \frac{\pi}{2} + \theta)$
P_4	y 轴	$(-x, y)$	$(r, \pi - \theta)$
P_5	原点	$(-x, -y)$	$(r, \pi + \theta)$
P_6	原点, $y = x$	$(-y, -x)$	$(r, \frac{3\pi}{2} - \theta)$
P_7	$y = x, x$ 轴	$(y, -x)$	$(r, \frac{3\pi}{2} + \theta)$
P_8	x 轴	$(x, -y)$	$(r, 2\pi - \theta)$

的角度坐标。基于复指数函数的数学性质，即三角恒等式，坐标上的这种相关性可以直接转换为基函数值的相关性。因此，这个观察可以将公式 (2.9) 和 (2.10) 的计算复杂度大约降低至 $1/8$ 。以上对称性和反对称性在所有正交圆谐矩中都成立，因此这种策略非常通用且易于与其他快速算法结合使用。

2.3.2.2 局部计算

局部计算是指对图像的部分（如密集的图像分块或特征点的兴趣区域）计算矩。在这种场景中，这些图像片段的分辨率通常较小，使用少量的矩就可以达到表征效果。作为推高时间复杂度的主要原因，这些图像片段的数量通常非常大，例如密集图像分块的数量可达 10^5 甚至 10^6 数量级。因此，常用于全局计算的加速方法通常不能很好地解决局部计算问题。目前，该问题还没有得到充分研究。最有希望的路径是在图像片段的循环处理中寻找并压缩冗余操作。在这方面，研究者尝试利用基函数的有用特性（例如文献^[91,92]中的平移特性）和特殊的数据结构（例如文献^[93]中的复值积分图）来提高计算效率。

2.3.2.3 快速计算的最新进展

对于基于雅各比多项式的正交圆谐矩，最先进的快速全局计算通常结合递归和对称性。Upneja 等人提出了一种 JFM 的快速算法^[73]（读者可同时参考^[70,76]），需要 $O(\frac{1}{8}(KN^2 + K^2N^2))$ 次加法和 $O(\frac{1}{8}K^2N^2)$ 次乘法来计算集合 $\{(n, m) : |n|, |m| \leq K\}$ 中所有阶数的矩。考虑到典型计算系统中乘法通常比加法更耗时，因此总复杂性由 $O(\frac{1}{8}K^2N^2) = O(K^2N^2)$ 次乘法主导。很明显，当需要更高的图像分辨率（即 N 增加）或更多的矩（即 K 增加）时，计算成本仍以平方阶增加。

对于基于谐波函数的正交圆谐矩，最先进的快速全局计算通常基于快速傅里叶变换。此类思路可以追溯到 2014 年，Ping 等人^[87]通过 EFM 与傅里叶变换之间的关系引入了快速傅里叶变换。随后，类似的思路被用于 RHFM^[85]和 PCET^[86]。最近，Yang 等人^[84]将这类技术进一步推广到了一种广义形式，其具有 $O(M^2 \log M) = O(N^2 \log N)$ 的乘法复杂度，其中 M 是采样参数并且 $M \propto N$ 。注意，变量 K 并未出现在上述复杂度公式中，即相对于最大阶数 K 的常数阶

计算复杂度。这意味着当 K 稍微增加, 使得 K^2 大于 $\log N$ 时, 该方法的耗时将显著低于基于递归和对称性的策略^[84]。除了提高效率外, 该方法还已被证明 (理论和实验上) 可以避免数值不稳定性, 并在 M 较大时提供相当高的计算精度。

对于快速局部计算, Bera 等人^[93] 做出了重要贡献。作者引入了一种特殊的数据结构, 即复值积分图, 将密集图像分块的 ZM 计算降低到相对于块尺寸的常数阶计算复杂度。作者还引入另一种数据结构, 即查找表, 实现了进一步的效率提升。虽然复值积分图策略并非首次提出, 例如 SURF 已使用它来加速 SIFT 的计算^[94], 但这仍是首次用于正交圆谐矩的计算: 作者通过复杂的推导将 ZM 的定义改写成积分图形式, 基于复平面表示和二项式展开。然而, 正如作者所讨论的, 这种快速算法似乎只适用于 ZM , 而不适用于其他具有更好表征能力 (即没有信息抑制问题) 的矩, 如 OFMM。目前, 正交圆谐矩的研究仍然缺乏常数阶且通用的快速局部计算策略, 发展此类策略将极大地推动现实应用。

2.3.3 稳健性/不变性优化

作为本节的数学背景, 我们首先给出了图像表征中不变性、稳健性和判别性的一般描述。

若存在一个函数 $\mathcal{R}$, 使得原始图像 f 和其劣化版本 $\mathcal{D}(f)$, 其中 $\mathcal{D}$ 是给定的劣化操作, 对于任何 f 均满足:

$$\mathcal{R}(f) \equiv \mathcal{R}(\mathcal{D}(f)), \quad (2.19)$$

则表征 $\mathcal{R}$ 对于劣化 $\mathcal{D}$ 满足不变性。

除不变性外, 还有两个与之相关的表征术语:

- 等变性—— $\mathcal{R}(\mathcal{D}(f)) \equiv \mathcal{D}(\mathcal{R}(f))$;
- 协变性—— $\mathcal{R}(\mathcal{D}(f)) \equiv \mathcal{D}'(\mathcal{R}(f))$ 。

其中 $\mathcal{D}'$ 是关于 $\mathcal{D}$ 可预测的复合函数。事实上, 协变性是不变性和等变性的推广术语, 是一种非常广义的对称性关系。在协变性或等变性成立的前提下, 构造不变表征是可行的, 基于输入和输出之间可预测的对称性关系。

将函数 $\mathcal{L} : \mathbf{X} \times \mathbf{X} \rightarrow [0, +\infty)$ 定义为集合 $\mathbf{X}$ 上的距离度量, 其对于任意 $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbf{X}$ 均满足:

- 同一性—— $\mathcal{L}(\mathbf{x}, \mathbf{y}) = 0 \Leftrightarrow \mathbf{x} = \mathbf{y}$;
- 对称性—— $\mathcal{L}(\mathbf{x}, \mathbf{y}) = \mathcal{L}(\mathbf{y}, \mathbf{x})$;
- 三角不等式—— $\mathcal{L}(\mathbf{x}, \mathbf{y}) \leq \mathcal{L}(\mathbf{x}, \mathbf{z}) + \mathcal{L}(\mathbf{z}, \mathbf{y})$ 。

对于给定距离度量 $\mathcal{L}$, 稳健性要求表征 $\mathcal{R}$ 的类内距离:

$$\mathcal{L}(\mathcal{R}(f), \mathcal{R}(\mathcal{D}(f))), \quad (2.20)$$

应该足够小。另一方面，假设图像 g 在语义上与图像 f 不同，判别性要求表征 $\mathcal{R}$ 的类间距离：

$$\mathcal{L}(\mathcal{R}(f), \mathcal{R}(g)), \quad (2.21)$$

应该足够大。可以看到，不变性是最理想的稳健性，即 $\mathcal{L}(\mathcal{R}(f), \mathcal{R}(\mathcal{D}(f))) = 0$ 当且仅当公式 (2.19) 成立，基于 $\mathcal{L}$ 的同一性。注意，基于矩的表征 $\mathcal{I}$ 是一类特殊的表征 $\mathcal{R}$ ，依赖于给定的基函数 V_{nm} ，可以写成：

$$\mathcal{R}(f) = \mathcal{I}(\{\langle f, V_{nm} \rangle\}). \quad (2.22)$$

实践中，几乎不可能设计出能够在各种劣化条件下均保持不变或稳健性的表征。更确切地说，只有完全牺牲判别性的表征才有可能实现这一点^[32]。因此，对于表征 $\mathcal{R}$ 的设计必须依赖于劣化 $\mathcal{D}$ 的某些假设。在此方面，面向仿射变换、噪声和模糊攻击的矩不变量已被相继提出。

2.3.3.1 仿射变换

仿射变换是一类图像函数定义域的变换，即劣化算子 $\mathcal{D}$ 是从像素坐标 (x, y) 到新坐标 (x', y') 的映射 $\mathcal{D} : (x, y) \rightarrow (x', y')$ ，定义如下：

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} a_{00} & a_{01} & t_x \\ a_{10} & a_{11} & t_y \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}, \quad (2.23)$$

其中仿射参数 $(a_{00}, a_{01}, a_{10}, a_{11}, t_x, t_y)$ 可以编码旋转、缩放、平移、错切、翻转以及上述变换的线性组合。在矩不变量的研究中，有三类相互关联的思路可以实现对仿射变换 $\mathcal{D}$ 的不变性：标准化方法、间接方法和显式方法^[32,33,95]。

- 标准化方法^[96,97]：其思路在于将输入图像 $\mathcal{D}(f)$ 转换为某种“标准”形式 f_{ref} ，使得标准化后图像的表征 $\mathcal{R}(f_{\text{ref}})$ 对仿射变换 $\mathcal{D}$ 不变。显然，这种方法的不变性来自于额外的标准化技术。更具体地说，标准化可被视为一个函数 $\mathcal{N}$ ，其对于任意图像 f 和任意仿射变换 $\mathcal{D}$ 满足 $\mathcal{N}(\mathcal{D}(f)) = \mathcal{N}(f) = f_{\text{ref}}$ 。注意，这里标准化是概念性的，无需在数字图像上实际执行矫正过程^[32]。可以看出，标准化结果将从根本上影响不变性。多年来，研究者在几何校正^[98,99]和图像配准^[100,101]等任务中长期追求标准化方法的设计。其中，最简单的方法之一是基于几何矩的方法，基于低阶几何矩估计并标准化图像的“质心”、“尺度因子”和“主轴”，从而分别实现对平移、缩放和旋转的不变性。基于几何矩的标准化可参阅 Flusser 等人的专著^[32,33]。其他类型的标准化方法可参阅 Zitova 等人的论文^[101]。
- 间接方法^[102,103]：其思路在于利用经典几何矩不变量导出复杂的正交矩不变量。几何矩的仿射不变量理论已在早期文献^[32,33,104]中给出。因此，可以通过正交矩和几何矩（非正

交)的代数关系将正交矩不变量表示为几何矩不变量的线性组合。显然,这种方法的不变性来自于几何矩不变量。注意,这种转化关系是普遍的,因为雅可比多项式是由序列 $\{x^0, x^1, \dots, x^n\}$ 的线性组合形成,谐波函数则可以通过泰勒级数以类似形式写出。在理论上,使用几何矩进行图像标准化然后基于正交矩进行描述,实际上等价于通过几何矩不变量的线性组合导出正交矩不变量。而在实践中,两类方法经常表现出完全不同的数值稳定性和计算复杂性^[32]。

- 显式方法^[105-107]: 其思路在于直接从正交矩计算不变量。数学上,这种方法通过设计 V_{nm} 和 I 来显式地满足恒等式 $I(\{\langle f, V_{nm} \rangle\}) = I(\{\langle \mathcal{D}(f), V_{nm} \rangle\})$ 。显然,相比于上述两类方法,这类方法的不变性是“直接”的。在这类方法中,一个常见的例子是显式构造圆谐正交矩的旋转不变性。基于角向基函数 $A_m(\theta) = \exp(jm\theta)$ 和傅里叶平移定理,旋转图像 $f_R(r, \theta) = f(r, \theta + \phi)$ 的圆谐正交矩是 $M'_{nm} = \exp(jm\phi)M_{nm}$ 。因此,图像旋转在这里只影响相位。为了抵消相位变化,传统方法基于幅度运算: $|M'_{nm}| = |\exp(jm\phi)M_{nm}| = |M_{nm}|$ 。但这种简单方法舍弃了太多的图像信息:只保留了幅度,而且忽略了更具信息量的相位^[108,109]。为此,Flusser^[110]提出了包含相位信息的复数值旋转不变量。令 $L \geq 1$, $n_i \in \mathbb{N}$, $m_i, k_i \in \mathbb{Z}$, $i = 1, \dots, L$, 并且满足 $\sum_{i=1}^L m_i k_i = 0$, 则圆谐正交矩的复数值旋转不变量可以定义为 $inv = \prod_{i=1}^L [M_{n_i m_i}]^{k_i}$ ^[110]。实践中,标准化方法对于许多场景来说可能不存在可行性,例如密集图像分块的表征,此外标准化方法也会引入误差。对于间接方法,则需要分配较长的时间来计算多项式系数。相反,显式方法由于其直接的结构更适用于现实场景,同时具有更高的计算精确性和效率^[79]。

2.3.3.2 噪声

与仿射变换不同,噪声作用于图像函数的值域,即劣化算子 $\mathcal{D}$ 将原始强度函数 $f(x, y)$ 映射到新的强度函数 $f_\eta(x, y)$: $\mathcal{D}: f(x, y) \rightarrow f_\eta(x, y)$ 。数学上,基于特定的噪声函数 $\eta(x, y)$, 常见的加性噪声描述为:

$$f_\eta(x, y) = f(x, y) + \eta(x, y), \quad (2.24)$$

乘性噪声描述为:

$$f_\eta(x, y) = f(x, y) \times \eta(x, y). \quad (2.25)$$

对于噪声稳健性或不变性,很难采用与仿射不变量中类似的标准化方法,因为图像的逆问题通常是病态或不适定的^[111,112]。目前,一种常见的策略是通过特定的变换 $\mathcal{T}$, 例如拉东变换,将图像转换到具有更高信噪比的新空间^[113-115]。基于这个新空间,设计 V_{nm} 和 I 以实现稳健或不变表征 $\mathcal{R}$:

$$I(\{\langle \mathcal{T}(f), V_{nm} \rangle\}) \simeq I(\{\langle \mathcal{T}(\mathcal{D}(f)), V_{nm} \rangle\}). \quad (2.26)$$

基于拉东变换的图像表征具有加性噪声的稳健性^[116]。图像函数 $f(x, y)$ 的拉东变换被定义为：

$$\mathbf{R}_f(r, \theta) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \delta_{r, x \cos \theta + y \sin \theta} dx dy, \quad (2.27)$$

其中, δ_{ij} 是公式 (2.3) 定义的克罗内克函数, $r = x \cos \theta + y \sin \theta$ 是以距离参数 r (相对于原点) 和角度参数 θ (相对于 y 轴) 定义的直线。公式 (2.24) 定义的噪声图像的拉东变换可以重写为：

$$\mathbf{R}_{f_\eta}(r, \theta) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [f(x, y) + \eta(x, y)] \delta_{r, x \cos \theta + y \sin \theta} dx dy = \mathbf{R}_f(r, \theta) + \mathbf{R}_\eta(r, \theta). \quad (2.28)$$

在连续域中, 噪声的拉东变换 $\mathbf{R}_\eta(r, \theta)$ 等于噪声的平均值。假设平均值为零, 即白噪声, 我们有：

$$\mathbf{R}_{f_\eta}(r, \theta) = \mathbf{R}_f(r, \theta). \quad (2.29)$$

这意味着零均值加性噪声对图像的拉东变换没有影响。至于乘性噪声 (2.25), 一种常见策略是使用对数变换作为预处理, 将图像和噪声再次分离为两个加性部分, 即 $\log[f(x, y) \times \eta(x, y)] = \log[f(x, y)] + \log[\eta(x, y)]$ 。注意, 由于离散图像和噪声存在采样和量化, 公式 (2.29) 的完美恒等性在实践中不成立。对于离散情况, 在拉东空间的信噪比仍然显著高于原始图像, 意味着更好的稳健性。相关理论分析和实验请参阅 Hoang 等人的论文^[114] 和 Jafari-Khouzani 等人的论文^[116]。

2.3.3.3 模糊

与噪声类似, 模糊也作用于强度域, 即 $\mathcal{D} : f(x, y) \rightarrow f_\mu(x, y)$ 。数学上, 基于特定的点扩散函数 $\mu(x, y)$, 图像模糊可以被描述为卷积运算：

$$f_\mu(x, y) = f(x, y) \otimes \mu(x, y). \quad (2.30)$$

因此, 模糊不变性也可以称为卷积不变性。对于图像模糊, 标准化实际上意味着盲去卷积, 这同样是一个病态或不适应问题^[117,118]。目前, 模糊不变量通常依赖于特定变换, 比如空间、频率或群投影操作, 其主要思想也可通过公式 (2.26) 来表达。相关理论分析和实验请参阅 Flusser 等人的论文^[119-122]。

2.3.3.4 稳健性/不变性优化的最新进展

二维和三维的仿射不变量已是非常成熟的理论^[123-126]。在 Flusser 等人的专著^[32,33] 中, 详细推导了基于几何矩和复数矩的显式方法和标准化方法, 这种不变量还可以通过转换关系扩展到正交矩。因此, Flusser 等人的仿射不变量理论非常系统化和广义化。对于该理论的扩展^[127] 是

目前领域的主要方向：1) 扩展到其他变换群，例如反射^[128]、射影^[129]、莫比乌斯^[130]和手性^[131]变换下的不变量。2) 扩展到其他数据维度或格式，例如彩色图像^[132]、曲线^[133]、曲面^[134]、矢量场^[135,136]和张量场^[137-139]的不变量。作为一项值得关注的工作，Li 等人^[140]深入探讨了不变量的结构：他们提出了两个基本的生成函数，可以编码几何矩不变量并进一步构造新不变量，就像 DNA 编码蛋白质一样；他们还发现，Hu 的矩不变量可以进一步分解为一个更简单的原始不变量，可能意味着不变量研究的新视角^[141]。

图像噪声和模糊的逆问题通常比仿射变换的逆问题更难解决，标准化方法，即去噪和去模糊，是非常具有挑战性的任务。此类方法通常是低效的且不稳定的，会引入额外的误差^[119]。对于 CNN 这类学习表征，由于缺乏结构上的不变性，使得对训练中未见的噪声和模糊操作（即使只是参数不同）敏感，最近的一项工作^[142]通过广泛的实验证实了这个结论。最常见的解决策略是数据增广，然而这非常消耗计算资源。在不变量研究中，一个主要目标是简化有关退化的假设。例如，通过对点扩散函数施加假设作为先验知识，推导出与运动模糊^[143]、两轴情况下的轴对称模糊^[144]、圆对称模糊^[145,146]、任意 N 折对称模糊^[120]、圆对称高斯模糊^[121]和各向异性高斯模糊^[119]的不变量。最近，Flusser 等人^[147]极大地发展了上述方法，提出了模糊不变量的通用理论，不依赖于点扩散函数的先验知识。

2.3.4 定义扩展

从不同的应用场景和优化目标出发，对矩和矩不变量的经典定义进行数学扩展也是热门课题。接下来，我们将介绍一些常见的定义扩展。

2.3.4.1 四元数

数学上，灰度图像函数 $f(x, y)$ 可以被定义为从二维图像平面到一维强度值的映射，即 $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ ；彩色图像函数 $\mathbf{f}(x, y) = \{f_{\text{red}}(x, y), f_{\text{green}}(x, y), f_{\text{blue}}(x, y)\}$ 可以被定义为从二维图像平面到三维强度值的映射，即 $\mathbf{f} : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ 。这里， $f_{\text{red}}(x, y)$, $f_{\text{green}}(x, y)$, 和 $f_{\text{blue}}(x, y)$ 分别是 RGB 颜色模型中的红色、绿色和蓝色分量。需要注意的是，其他的三通道颜色模型也适用于本节分析。

前文所列出的矩和矩不变量均是为灰度图像而设计的。处理彩色图像时，有两种直接的策略^[148,149]：1) 灰度化方法，直接计算灰度版本的矩；2) 逐通道方法，分别计算每个颜色通道的矩。然而，灰度化方法会丢失重要的颜色信息；逐通道方法很难产生紧凑的表征，并忽略了颜色通道之间的相关性。

为了解决上述问题，一个常用的策略是将彩色图像 $\mathbf{f}(x, y)$ 定义为从二维图像平面到四元数强度值 $f_{\text{red}}(x, y)\mathbf{i} + f_{\text{green}}(x, y)\mathbf{j} + f_{\text{blue}}(x, y)\mathbf{k}$ 的映射，即 $\mathbf{f} : \mathbb{R}^2 \rightarrow \mathbb{H}$ 。相应地，基函数 V_{nm} 也从实数域 $\mathbb{R}$ 或复数域 $\mathbb{C}$ 扩展到四元数域 $\mathbb{H}$ ，得到新的基函数 $V_{nm} \in \mathbb{H}$ ，将原始内积 $\langle f, V_{nm} \rangle$ 扩展

为 $\langle f, V_{nm} \rangle$ [84,150,151]。相关理论分析和实验请参阅 Chen 等人的论文 [150,151]。

2.3.4.2 高维

四元数扩展作用于图像和基函数的值域，而高维扩展作用于它们的定义域。前文所列出的矩和矩不变量均是为二维图像设计的。当处理高维空间 $\mathbb{R}^d$ （主要为三维图形， $d = 3$ ）中定义的函数时，基函数 V_{nm} 的定义域 D 也应扩展至 $\mathbb{R}^d$ [33]。

对于笛卡尔矩，这种扩展相当直接：基于一个一维正交多项式集，我们可以通过在每个方向/变量上使用相同的多项式集来生成 d 维正交基函数 [152]。

对于圆谐矩，高维扩展则没有这样直接。常见的策略是用阶数 $l \in \mathbb{N}$ 和 $m \in \mathbb{Z}$ 的球谐函数 $Y_{ml}(\theta, \varphi)$ 替代角向基函数 $A_m(\theta) = \exp(jm\theta)$ [153,154]：

$$Y_{ml}(\theta, \varphi) = \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} L_{ml}(\cos \theta) \exp(jm\varphi), \quad (2.31)$$

其中 $\theta \in [0, \pi)$ 且 $\varphi \in [0, 2\pi)$ 分别为球坐标系的极角和方位角， $|m| \leq l$ ，而 L_{ml} 表示勒让德函数：

$$L_{ml}(x) = (-1)^m 2^l (1-x^2)^{\frac{m}{2}} \sum_{k=m}^l \frac{k! x^{k-m}}{(k-m)!} \binom{l}{k} \binom{\frac{l+k+1}{2}}{l}. \quad (2.32)$$

球谐函数 $Y_{ml}(\theta, \varphi)$ 满足正交性条件：

$$\langle Y_{ml}, Y_{m'l'} \rangle = \int_0^{2\pi} \int_0^\pi Y_{ml}(\theta, \varphi) Y_{m'l'}^*(\theta, \varphi) \sin \theta d\theta d\varphi = \delta_{mm'} \delta_{ll'}. \quad (2.33)$$

因此，可以组合球谐函数 $Y_{ml}(\theta, \varphi)$ 和径向基函数 $R_n(r)$ 来构建球坐标中的三维正交基函数，只需轻微的归一化调整 [155]。

2.3.4.3 分数阶

前文所列出的矩和矩不变量均基于整数阶域 $(n, m) \in \mathbb{Z}^2$ 。最近，出现了一类扩展经典矩和矩不变量的阶数域的思路 [156–158]。通过适当的变量替换，例如 $r := r^\alpha \in [0, 1]$ （圆谐矩）或 $x := x^\alpha, y := y^\alpha \in [0, 1]$ （笛卡尔矩），引入分数阶参数 $\alpha \in \mathbb{R}$ 。这种新定义中阶数可以扩展到实数域，例如 $\alpha n \in \mathbb{R}$ （圆谐矩）或 $(\alpha n, \alpha m) \in \mathbb{R}^2$ （笛卡尔矩）。

注意，这种分数阶矩不仅仅是经典矩的数学扩展，还具有独特的时频分析能力 [159]。具体而言，通过改变分数阶参数的值能够控制基函数的零点分布。根据对信息抑制问题的研究，基函数的零点分布是一个非常重要的属性，其与矩在图像平面上的描述重点直接相关。因此，分数阶矩有利于解决信息抑制问题和提取图像局部特征 [156–158]。

2.3.4.4 定义扩展的最新进展

矩和矩不变量的四元数扩展已成为彩色图像处理中的常见策略^[148,160-164]。此外，四元数理论本身也越来越多地应用于相关领域，包括其他手工表征^[165-167]和学习表征^[168-171]。以下是近期在矩和矩不变量方面的一些进展。**Chen** 等人^[172]采用四元数代数表示 RGB-D（即 RGB 和深度）图像，其中四元数数实部编码深度分量。**Yamni** 等人^[173]提出了一种用于彩色立体图像表征的八元数矩。相关研究的理论来源主要是超复数代数^[174]。

矩和矩不变量的高维扩展在很早就已出现^[155]，并在过去十年中引起了更大关注^[152,175-180]，这与三维图形相关设备和技术（如医学成像^[181]和计算机图形学^[182]）的快速发展有关。此领域的最新工作主要集中与三维矩的新定义^[152,175]、不变量^[179,180]和精确/快速计算^[176-178]。注意，二维矩精确/快速计算中的大部分思路（2.3.1 和 2.3.2 节）可以自然地推广到三维情况，三维矩不变量的重要工作已在 2.3.3 节中分析。

与上述两类工作相比，矩和矩不变量的分数阶扩展在近期吸引了更多关注^[183-185]。在这方面，我们发现的最早工作由 **Hoang** 等人^[156,157]提出：通过分数阶扩展，将基于谐波函数的圆谐正交矩（2.2.2 节）扩展为广义极坐标谐波变换（Generic PHT, GPHT）。其中，一个重要贡献是首次发现和分析了分数阶扩展的时频特性，及其在图像表征中的潜力。**Xiao** 等人^[158]则给出了一种分数阶扩展的通用框架，适用于基于雅可比多项式的圆谐矩和笛卡尔矩。大多数后续论文^[186-193]遵循了 **Hoang** 等人^[156,157]和 **Xiao** 等人^[158]的方法。上述工作都关注于解决信息抑制问题或提取局部图像信息上。最近，**Yang** 等人^[159]将基于雅可比多项式的圆谐正交矩（2.2.1 节）扩展为分数阶雅可比-傅里叶矩（Fractional-order JFM, FJFM）。更重要的是，作者发现上述方法在图像全局表征中存在一个共同的缺陷，即使用固定的分数阶参数（经实验确定）进行全局表征只能带来轻微的精度提升，并尝试利用分数阶矩的时频特性来改善全局表征的判别性。

2.3.5 应用

除上述理论研究，矩和矩不变量的应用研究也十分广泛。这些应用涵盖了图像处理、计算机视觉和信息安全等许多熟悉问题，还包括一些交叉学科的前沿领域。

- 图像处理——对于低级视觉任务，矩和矩不变量是常用技术。应用涉及图像处理的不同分支：低级特征检测和表征（如边缘^[194]和特征点^[195]检测、感兴趣区域^[196]、纹理^[72]和光流^[197]表征），图像恢复（如去噪^[198]、去模糊^[199]和超分辨率^[200]），图像配准^[201]，图像压缩、编码和通信^[202]以及图像质量评估^[203]。类似地，在视频分析^[204]和计算机图形学^[205]中，矩和矩不变量也被用于低级任务。
- 视觉理解——对于高级视觉任务，矩和矩不变量已经在计算机视觉的不同分支中得到少量应用：图像分类^[206]、实例检索^[207]、目标检测^[208]和语义分割^[209]；同时也在模式识别

的不同分支中得到应用：行为识别^[210]、文本识别^[211]、生物特征识别^[212]和情感识别^[213]。

- 信息安全——矩和矩不变量对视觉媒体的信息安全研究产生了重要影响，矩和矩不变量的稳健性特别适用于与安全研究的对抗性假设。应用涉及视觉媒体信息安全的不同分支：数字水印^[214]、隐写术^[215]、感知哈希^[216]和被动取证（例如，复制-粘贴^[217]和拼接^[172]检测）。
- 跨学科——鉴于视觉信息处理和理解是许多学科中的基础性任务，矩和矩不变量也因此自然扩展到这些领域。跨学科应用涉及：医学（如医学成像^[218]）、地理学（如遥感^[219]）、机器人学（如视觉伺服^[220]）、物理学（如光学^[221]和流体力学^[135]）、化学（如分析化学^[222]）、生物学（如蛋白质结构^[223]）和材料科学（如原子环境^[224]）。

从以上应用中可以看到：1）矩和矩不变量的理论研究具有很强的普适性，适用于诸多小规模的应用领域；2）矩和矩不变量常在要求高效性、稳健性和可解释性的场景中发挥关键作用。

2.3.6 总结和讨论

综合以上经典不变量的研究现状，可以总结出以下两方面的判别性问题：

- 理论上，经典不变量通常基于图像的全局表征假设，即基础性的 2.2 节以及 2.3.1~2.3.4 节，而对于视觉任务中更具信息量的图像局部和层次表征，缺乏相应的理论扩展；
- 实践上，经典不变量通常用于小规模视觉任务，即 2.3.5 节，而对于具有不变性或对称性先验的更大规模视觉任务，缺乏相应的实践扩展。

2.4 开源实现和数值结果

在本节中，我们将给出矩和矩不变量的开一个系统性的开源软件包，涵盖部分精确/快速计算、稳健性/不变性优化和定义扩展策略。这个软件包被称为 MomentToolbox，可在我们的开源链接 <https://github.com/ShurenQi/MomentToolbox> 下载源代码。

在这个统一的软件基础上，我们将通过矩的计算、图像重构和模式识别实验分别评估方法的精确性/复杂度、表征能力以及稳健性/不变性。

2.4.1 矩的计算：精确性和复杂性

我们使用不同方法计算一个 128×128 单位灰度图像的矩和矩不变量，即 $\{f_{\text{uni}}(x, y) = 1 : (x, y) \in D\}$ ，来评估精确性和复杂性。

回到公式 (2.1) 和 (2.4)，可以很容易地推导出该图像的圆谐正交矩为

$$\langle f_{\text{uni}}, V_{nm} \rangle = \iint_D V_{nm}^*(x, y) f_{\text{uni}}(x, y) dx dy = \begin{cases} 0 & m \neq 0 \\ 2\pi \int_0^1 R_n^*(r) r dr & m = 0 \end{cases}. \quad (2.34)$$

表 2.4 不同圆谐正交矩阶数集合 $S_{nm}(K)$ 的定义和基数。

圆谐正交矩	$S_{nm}(K)$	$ S_{nm}(K) $
ZM ^[46]	$n - m = \text{even}, m \leq n \leq K$	$(K+1)(K+2)/2$
PZM ^[37]	$ m \leq n \leq K$	$(K+1)^2$
OFMM ^[50] , CHFM ^[51] , PJFM ^[52] , JFM ^[53] , RHFM ^[55] , PCT ^[57] , BFM ^[58] , FJFM ^[159] , GRHFM ^[157] , GPCET ^[157] EFM ^[56] , PCET ^[57] , GPCET ^[157] PST ^[57] , GPST ^[157]	$0 \leq n, m \leq K$	$(K+1)(2K+1)$
	$ n , m \leq K$	$(2K+1)^2$
	$1 \leq n \leq K, m \leq K$	$K(2K+1)$

上述公式意味着，对于任意 $(n, m) \in \{(n, m) : (n, m) \in \mathbb{Z}^2, m \neq 0\}$ ，其对应矩 $\langle f_{\text{uni}}, V_{nm} \rangle$ 的理论值是 0。但在实际操作中，这种恒等关系通常不成立，源于计算方法的误差^[225,226]。事实上，2.3.1 节列举的误差，即几何误差、数值积分误差和表示误差，都可能导致这种现象发生。

根据上述事实，计算精度可以通过一种简单的度量标准来评估，称为平均计算误差（Average Calculation Error, ACE）：

$$\text{ACE} = \frac{\sum_{(n,m) \in \{S_{nm}(K), m \neq 0\}} |M_{nm}^{(f_{\text{uni}})}|}{|S_{nm}(K)|} \quad (2.35)$$

其中 $M_{nm}^{(f_{\text{uni}})}$ 是通过特定计算方法得到的单位灰度图像 f_{uni} 的圆谐正交矩值， $S_{nm}(K)$ 是基于最大阶数 K 所产生的 (n, m) 阶数采样集合。这里，各种圆谐正交矩阶数集合 $S_{nm}(K)$ 的定义和基数在表 2.4 中列出。

计算复杂性则通过阶数集合 $S_{nm}(K)$ 上所用的计算时间来评估，称为分解时间（Decomposition Time, DT）：

$$\text{DT} = \sum_{(n,m) \in S_{nm}(K)} \text{time}(M_{nm}^{(f_{\text{uni}})}) \quad (2.36)$$

其中 $\text{time}(M_{nm}^{(f_{\text{uni}})})$ 表示计算矩 $M_{nm}^{(f_{\text{uni}})}$ 所经过的时间。为公平起见，所有方法都在单线程模式下运行，如 Matlab 中的 -singleCompThread 选项。

对比方法包括：

- 雅可比多项式经典方法的直接计算（ZM^[46]、PZM^[37]、OFMM^[50]、CHFM^[51]、PJFM^[52] 和 JFM^[53]）；
- 谐波函数经典方法的直接计算（RHFM^[55]、EFM^[56]、PCET^[57]、PCT^[57] 和 PST^[57]）；
- 本征函数经典方法的直接计算（BFM^[58]）；
- 雅可比多项式分数阶方法的直接计算（FJFM^[159]）；
- 谐波函数分数阶方法的直接计算（GRHFM^[157]、GPCET^[157]、GPCT^[157] 和 GPST^[157]）；
- 雅可比多项式分数阶方法的递归计算（FJFM^[159]）；
- 谐波函数分数阶方法的快速傅里叶变换计算（GPCET^[84]）。

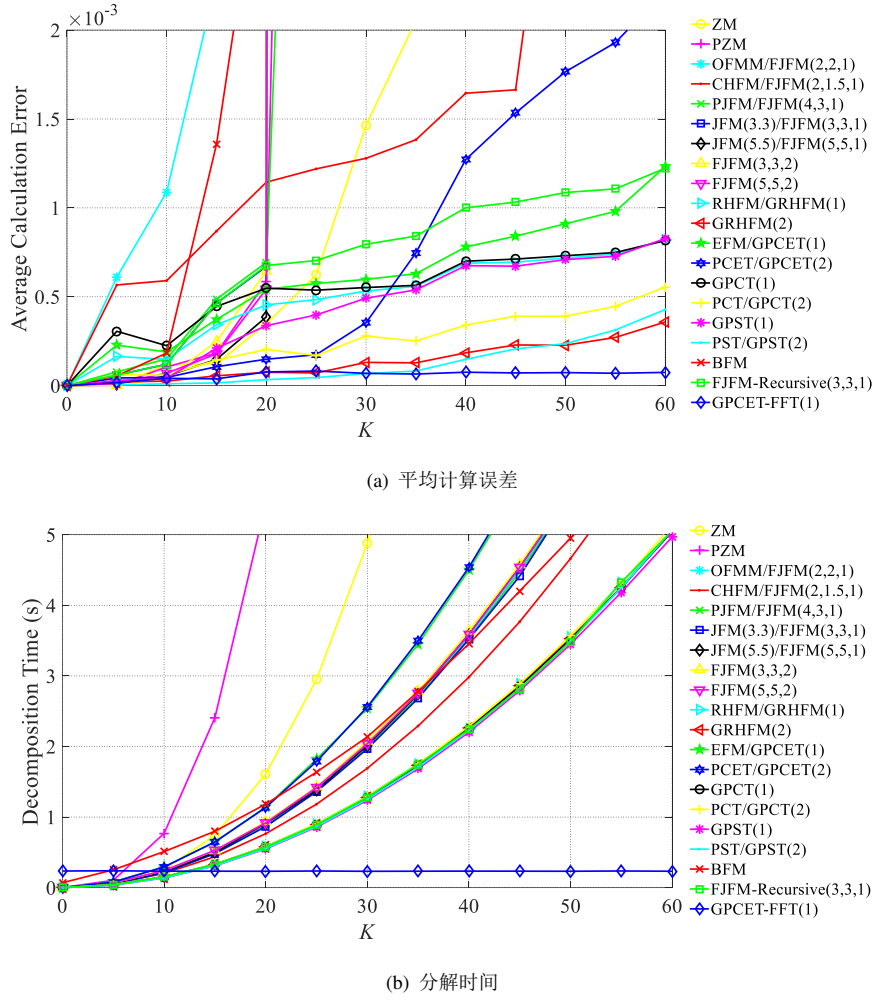


图 2.7 不同圆谐正交矩的平均计算误差和分解时间。

这里, JFM 具有参数 (p, q) , FJFM 具有参数 (p, q, α) , GRHF/GRHF/GPCET/GPCT/GPST 具有参数 α 。注意 FJFM 和 GPCET 实际上是现有的基于雅可比多项式的圆谐正交矩 [37,46,50–53,158,186–189] 和基于谐波函数的圆谐正交矩 [55–57,157,191–193,227] 的扩展版本。换句话说, FJFM 和 GPCET 的精确/快速计算可以直接用于其特殊版本, 只需适当设置参数的值。有关分析请参阅 [84,159]。

图 2.7 展示了所有上述比较方法的平均计算误差和分解时间。可以观察到:

- 精确性: 1) 当 K 增加到一定值后, 雅可比多项式方法的误差曲线会出现突然上升。显然, 这是由阶乘或伽玛项导致的数值不稳定性问题。可以预期, 基于递归的计算具有更好的精度由于不涉及大数的阶乘或伽玛, 因而在图 2.7 中没有观察到类似的数值不稳定性现象。2) 谐波函数方法主要问题是数值积分误差, 根据采样定理, 随着 K 的增加, 数值积分误差也会增加, 这与我们在图 2.7 中观察到的现象一致。注意, 某些谐波函数方法的径向基函数是无界的, 例如 RHF 和 EFM, 这意味着数值不稳定性问题并导致二者的计算精度较差。由于伪上采样和极坐标定义方式, 基于快速傅里叶变换的计算表现出几

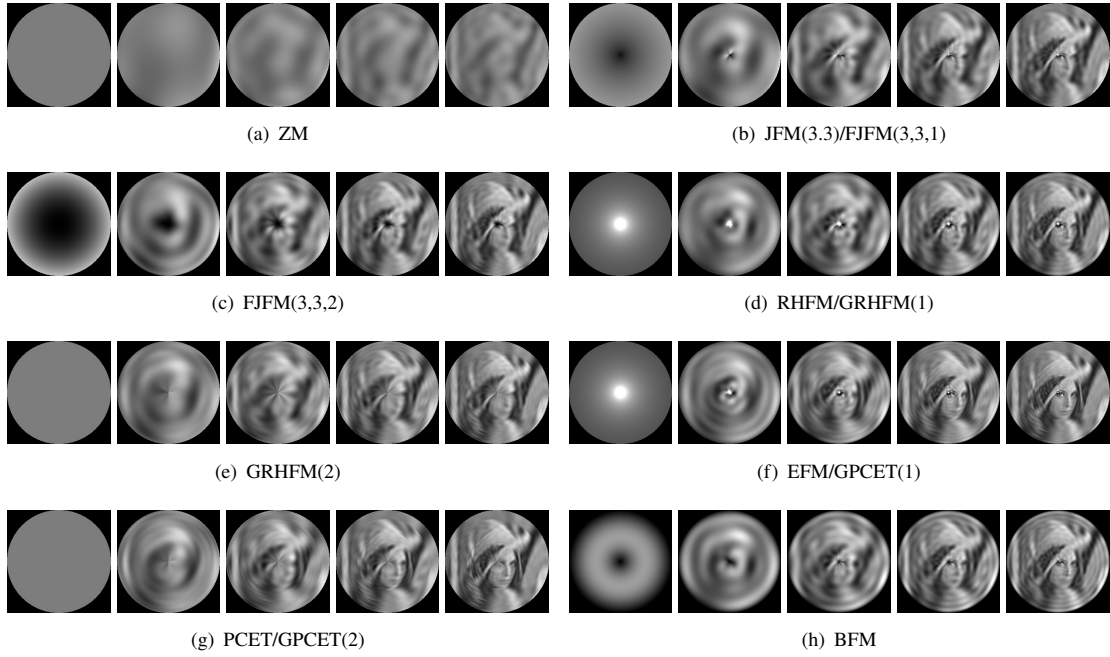


图 2.8 不同圆谐正交矩的重构图像样例，其中 $K = \{0, 5, \dots, 20\}$ （从左到右）。

乎具有恒定的误差曲线，同时避免了无界性导致的数值不稳定现象。3) 对于本征函数方法，虽然我们使用了快速且稳定的 Matlab 内置函数 `besselj` 来计算径向基函数，但其误差曲线仍然急剧上升。事实上，对于较大的 K ，其误差介于雅可比多项式方法和谐波函数方法之间（没有数值不稳定性）。

- 复杂性：1) 雅可比多项式方法的复杂性一般较高，特别是当 K 较大时，这是因为径向基函数中复杂的阶乘或伽玛项以及求和。通过递归策略，这部分计算可以大大减少，展现出与谐波函数方法类似的复杂性。2) 谐波函数方法的计算成本主要来自内积，这是因为径向基函数相对简单。注意，对于相同的 K ，EFM 和 PCET 的矩的数量是其他谐波函数方法的两倍（表 2.4），因此二者的计算时间也高出约两倍，这是符合预期的。若使用基于快速傅里叶变换的计算，内积计算将大大减少，并且其复杂性与 K 无关。3) 本征函数方法表现出与雅可比多项式方法类似的复杂性。注意，如果直接根据定义计算而不使用快速的 `besselj`，其复杂性将明显提高，由于阶乘或伽玛项、无穷级数和求根操作。

评注：在直接计算场景中，谐波函数方法具有更好的计算效率和精度相比于其他两类方法。递归策略可以有效地克服雅可比多项式方法的数值稳定性问题并具有更好的效率，应该在实际应用中推广。快速傅里叶变换策略进一步优化了谐波函数方法的效率和精度，在整个矩和矩不变量家族中表现出最优异的计算性能。

2.4.2 图像重构: 表征能力

我们使用不同方法重构 256×256 灰度图像 “Lena” 来评估表征能力。

回到公式 (2.1) 和 (2.4)，由于 V_{nm} 的正交性，可以从圆谐正交矩集合 $\{M_{nm} = \langle f, V_{nm} \rangle\}$ 重构出图像：

$$\hat{f}(x, y) = \sum_{(n,m) \in S_{nm}(K)} M_{nm} V_{nm}(x, y) \quad (2.37)$$

其中， $\hat{f}$ 是原始图像 f 的重构（近似）版本， $S_{nm}(K) \in \mathbb{Z}^2$ 在表 2.4 中定义。该过程的误差通过均方重构误差（Mean-Square Reconstruction Error，MSRE）^[37] 评估：

$$\text{MSRE} = \frac{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [f(x, y) - \hat{f}(x, y)]^2 dx dy}{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [f(x, y)]^2 dx dy}, \quad (2.38)$$

同时，结构化相似性（Structural SIMilarity，SSIM）^[228] 用作保真度评估指标：

$$\text{SSIM} = \frac{(2\mu_f \mu_{\hat{f}} + C_1)(2\sigma_{f, \hat{f}} + C_2)}{(\mu_f^2 + \mu_{\hat{f}}^2 + C_1)(\sigma_f^2 + \sigma_{\hat{f}}^2 + C_2)}, \quad (2.39)$$

其中， $\mu_{\bullet}$ 、 $\sigma_{\bullet}$ 和 $\sigma_{\bullet, \bullet}$ 分别表示均值、标准差和协方差；常量 C_1 和 C_2 被设置为 $C_1 = (0.01 \times 255)^2$ 和 $C_2 = (0.03 \times 255)^2$ 。

对比方法包括：

- 雅可比多项式经典方法（ZM^[46]、PZM^[37]、OFMM^[50]、CHFMM^[51]、PJFM^[52] 和 JFM^[53]）；
- 谐波函数经典方法（RHFM^[55]、EFM^[56]、和 PCET^[57]）；
- 本征函数经典方法（BFM^[58]）；
- 雅可比多项式分数阶方法（FJFM^[159]）；
- 谐波函数分数阶方法（GPCET^[157]）。

为了更好地反映图像重构和表征的真实能力，我们在这里全部采用了基于递归或快速傅里叶变换的精确计算而非直接计算，抑制重构结果中计算误差的影响。具体地，ZM 和 PZM 分别采用 q -递归方法^[60] 和 p -递归方法^[90] 进行实现；其他雅可比多项式方法和谐波函数方法分别采用通用递归计算^[159] 和通用快速傅里叶变换计算^[84]；BFM 则通过稳定的 Matlab 内置函数 `besselj` 计算。

图 2.8 展示了相对于 $K = \{0, 5, \dots, 20\}$ 的一些重构图像样例，上述比较方法的均方重构误差和结构相似性值在图 2.9 给出。可以观察到：

- 更多矩值可以更好地重构/表征图像。从重构图像的可视结果和相应的量化分数来看，总体上 K 值越大重构效果越好。一些例外，例如 OFMM，其均方重构误差曲线在 K 增加到

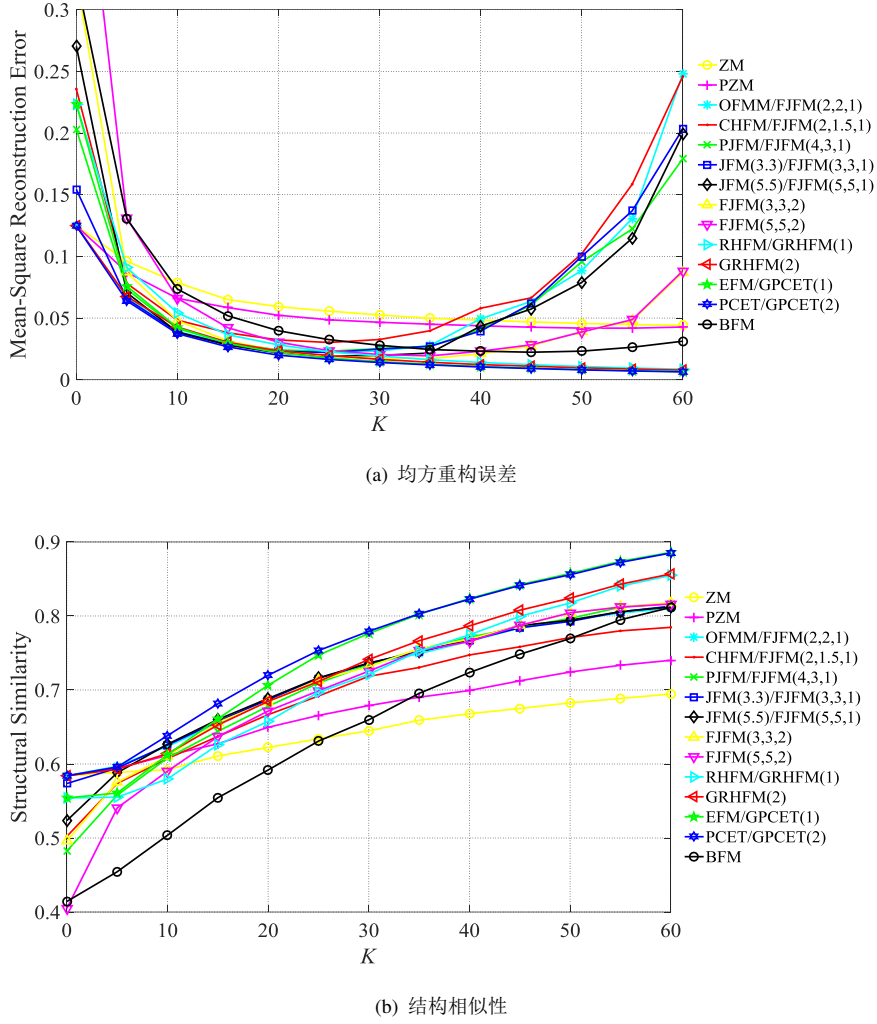


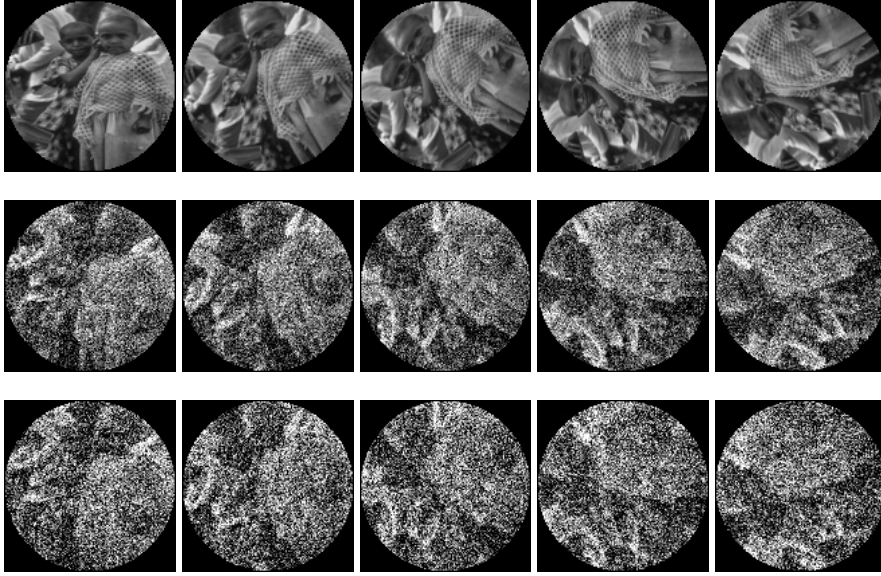
图 2.9 不同圆谐正交矩的均方重构误差和结构相似性。

一定程度时开始上升，这主要是由于在 $r \approx 1$ 处的数值不稳定性。

- 低阶矩主要反映图像的整体灰度/强度信息，而高阶矩则反映更复杂的纹理/形状信息。从重构图像的可视结果来看，总体上随着 K 值增加，图像大体轮廓先重构出来而后精细内容才被重构。从稳健性的角度来看，我们可以推断高阶矩（对应于图像的高频成分）对噪声、模糊和采样/量化效应等的稳健性更低。
- 重构/表征质量与基函数的零点数量密切相关。在雅可比多项式方法中，ZM 和 PZM 由于具有较少的基函数零点，其重构图像的量化分数要比其他方法差。类似的现象也见于谐波函数方法中的 RHFM 和 GRHFM。
- 对于分数阶方法，在 $\alpha \approx 1$ 时，重构质量相对较好，这应该归因于径向基函数零点的均匀分布。当 $\alpha > 1 / \alpha < 1$ 时，很难重构图像的内/外部分；相反，当 $\alpha < 1 / \alpha > 1$ 时，则可以通过更少的矩来重构内/外部分。简洁起见，这里值展示了 $\alpha > 1$ 的情况，可以使用



图 2.10 训练图像示例。


 图 2.11 测试图像示例，其中旋转角度为 $\{0^\circ, 30^\circ, \dots, 120^\circ\}$ （从左到右）以及噪声方差为 $\{0, 0.05, 0.1\}$ （从上到下），同时包含尺度和位置的随机扰动。

MomentToolbox 验证其他情况。

评注: 在雅可比多项式方法中，除了 ZM 和 PZM 外，其他方法的重构/表征效果均很相似，这是由于它们的基函数具有类似的数学特性。事实上，ZM 和 PZM 需要更高阶的矩来获得与其他方法相近的性能，同时以稳健性为代价。谐波函数方法，特别是 EFM，PCET 和 GPCET，在表征能力上优于其他方法。注意，分数阶方法的时频分析能力对于某些场景（如解决信息抑制问题和提取图像局部特征）是优异的。

2.4.3 模式识别: 稳健性和不变性

我们使用不同方法分类劣化图像（仿射变换和噪声）来评估评估稳健性和不变性。

在训练集方面，我们从 COREL 数据集^[229]中选择了 100 幅图像并统一调整为 128×128 的

大小, 如图 2.10 所示, 这些训练图像的特征向量将作为分类的基准。在测试集方面, 我们将训练集中的每个图像以角度 $\{0^\circ, 10^\circ, \dots, 350^\circ\}$ 旋转, 以反映任意朝向; 此外, 尺度和位置分别被随机扰动了 $\{10, 12, \dots, 18\}$ 和 $\{0, 1, \dots, 4\}$ 个像素, 以模拟感兴趣区域的检测误差。随后, 每个图像都被加入方差为 $\{0, 0.05, \dots, 0.3\}$ 的高斯白噪声。因此, 测试图像同时包含仿射和信号劣化操作。基于此, 每个训练图像对应 $36 \times 7 = 252$ 幅测试图像, 测试集中共计 $252 \times 100 = 25200$ 幅图像, 如图 2.11 所示。

特征方面, 我们使用圆谐正交矩的旋转不变量来描述训练集和测试集中的所有图像, 定义为:

$$\mathbf{v}(f) = \{|\langle f, V_{nm} \rangle| : (n, m) \in S_{nm}(K)\}, \quad (2.40)$$

其中, $S_{nm}(K) \in \mathbb{Z}^2$ 在表 2.4 中定义。回到公式 (2.4), 容易验证当 V_{nm} 具有形式 $R_n(r) \exp(jm\theta)$ 时公式 (2.40) 满足旋转不变性。分类器方面, 我们使用简单的最小距离分类器 (基于欧几里德距离), 以更好地反映表征本身的稳健性和不变性。指标分数方面, 分类的准确率由正确分类百分比 (Correct Classification Percentages, CCP)^[58] 评估:

$$\text{CCP} = \frac{\# \text{ correctly classified samples}}{\# \text{ total samples in the dataset}}. \quad (2.41)$$

对比方法包括:

- 雅可比多项式经典方法 (ZM^[46]、PZM^[37]、OFMM^[50]、CHFM^[51]、PJFM^[52] 和 JFM^[53]);
- 谐波函数经典方法 (RHFM^[55]、EFM^[56]、和 PCET^[57]);
- 本征函数经典方法 (BFM^[58]);
- 雅可比多项式分数阶方法 (FJFM^[159]);
- 谐波函数分数阶方法 (GPCET^[157]);
- 两种轻量级学习方法, PCANet^[230] 和 CBF^[231], 分别受降维描述子 PCA 和纹理描述子 LBP 的启发;
- 两种深度学习方法, GoogLeNet^[232] 和 ResNet-50^[233], 已在高层视觉任务广泛应用。

注意, 为了进一步证明矩和矩不变量在小规模稳健可解释图像表征中的有效性, 我们将与学习表征进行比较。对于 PCANet 和 CBF, 我们使用原始论文中的公共参数设置, 为了避免内存溢出, PCANet 中输入图像的分辨率被降低到 64×64 。对于 GoogLeNet 和 ResNet-50, 预训练模型形成于超大规模图像集 ImageNet, 基于迁移学习策略进行重新训练以适应本文图像集。相关细节请参阅 MomentToolbox 代码和原始论文^[230-233]。

表 2.5 列出了不同矩不变量方法的分类分数对比, 基于 $K = \{10, 20\}$ 。为了更好地可视化, 对于每个噪声方差, 相对较高/较低的分类分数用蓝色/红色底纹表示, 并用粗体标出最高的分类分数。此外, 表 2.6 列出了与学习表征的对比, 其中分类分数平均值 (表 2.5, $K = 20$) 用于代

表 2.5 不同圆谐正交矩不变量的分类分数 (%)。

矩不变量	$K = 10$, 高斯白噪声 (方差)						
	0	0.05	0.1	0.15	0.2	0.25	0.3
ZM	89.03	68.78	42.00	26.67	17.47	12.31	9.06
PZM	92.17	76.33	45.36	30.44	18.14	12.94	9.47
OFMM/FJFM(2,2,1)	97.25	90.00	72.42	47.11	26.53	15.36	8.78
CHFM/FJFM(2,1.5,1)	98.56	93.86	78.03	60.94	40.22	25.42	16.50
PJFM/FJFM(4,3,1)	97.94	92.64	79.08	58.28	37.50	22.67	13.81
JFM(3.3)/FJFM(3,3,1)	97.92	93.81	77.19	55.94	35.44	22.39	13.89
JFM(5.5)/FJFM(5,5,1)	97.75	93.53	78.86	59.67	38.56	24.92	16.86
FJFM(3,3,2)	97.36	93.22	76.92	56.83	38.11	24.25	16.50
FJFM(5,5,2)	96.58	91.47	75.64	56.22	37.53	22.44	15.22
RHFM/GRHFM(1)	98.17	94.22	77.81	57.72	35.33	21.94	13.61
GRHFM(2)	97.42	89.03	68.17	43.17	25.78	14.78	9.72
EFM/GPCET(1)	99.56	97.06	86.47	61.78	37.33	21.94	13.44
PCET/GPCET(2)	98.75	92.97	79.00	52.53	29.67	16.92	10.50
GPCT(1)	98.64	95.25	80.39	60.22	39.33	24.75	16.53
PCT/GPCT(2)	97.17	88.44	70.14	43.81	25.08	15.44	9.22
GPST(1)	98.89	95.72	79.61	60.50	38.72	23.53	13.75
PST/GPST(2)	96.94	89.69	70.47	46.50	25.56	15.11	9.06
BFM	98.39	94.28	79.33	60.06	37.47	21.53	13.28

矩不变量	$K = 20$, 高斯白噪声 (方差)						
	0	0.05	0.1	0.15	0.2	0.25	0.3
ZM	89.97	70.86	43.03	27.36	16.50	12.28	9.19
PZM	92.67	76.53	46.36	30.58	18.44	12.92	9.03
OFMM/FJFM(2,2,1)	97.78	90.78	74.08	49.47	29.97	16.61	10.00
CHFM/FJFM(2,1.5,1)	98.50	94.36	78.83	61.39	41.36	26.53	17.61
PJFM/FJFM(4,3,1)	98.17	93.75	80.28	60.97	40.78	23.53	15.69
JFM(3.3)/FJFM(3,3,1)	98.19	94.22	78.44	57.22	37.53	22.69	14.42
JFM(5.5)/FJFM(5,5,1)	97.94	94.11	79.83	60.94	41.31	26.69	18.11
FJFM(3,3,2)	97.67	93.83	78.78	59.44	41.06	26.47	17.58
FJFM(5,5,2)	97.08	92.89	77.33	60.86	40.94	26.92	17.89
RHFM/GRHFM(1)	98.44	94.97	79.78	60.67	39.08	24.67	14.86
GRHFM(2)	97.81	90.78	70.06	45.89	27.64	16.78	10.39
EFM/GPCET(1)	99.47	97.86	87.97	66.83	42.47	25.56	16.33
PCET/GPCET(2)	99.11	94.56	80.42	55.42	33.19	21.61	14.42
GPCT(1)	98.50	95.17	81.83	62.00	42.11	26.64	17.94
PCT/GPCT(2)	97.53	89.72	70.89	46.06	27.36	16.75	10.31
GPST(1)	98.89	96.25	81.00	62.14	41.58	24.67	15.42
PST/GPST(2)	97.28	91.17	73.11	50.36	28.81	17.92	10.53
BFM	98.50	94.69	81.08	61.44	40.75	23.75	14.75

表矩不变量方法的整体性能。可以观察到：

- 表 2.5 中列出的所有矩不变量方法在干净图像上都取得了较高的分数，这是因为保障的旋转不变性和正交性。直观地，分类分数与噪声方差呈负相关、与 K 呈正相关。因此，分类分数受方差或 K 的影响程度反映了对其噪声的稳健性。

表 2.6 学习方法和矩不变量方法的分类分数 (%)。

表征方法	高斯白噪声 (方差)						
	0	0.05	0.1	0.15	0.2	0.25	0.3
PCANet (直接训练)	19.28	7.58	4.64	3.03	2.08	2.03	1.56
PCANet (旋转增广训练)	99.97	14.25	4.33	2.19	1.69	1.25	1.19
CBFD (直接训练)	22.58	3.72	2.08	1.22	1.19	1.14	1.08
CBFD (旋转增广训练)	98.92	3.89	2.50	1.75	1.53	1.64	1.11
GoogLeNet (旋转增广训练)	96.17	6.28	2.50	1.39	1.22	0.89	1.11
ResNet-50 (旋转增广训练)	99.33	3.03	1.61	1.00	1.17	0.78	1.03
矩不变量 (平均分数)	97.42	91.47	74.62	54.39	35.05	21.83	14.14

- 分类分数与基函数的零点数量/分布密切相关。如表 2.5 所示, ZM 和 PZM 的性能低于其他方法, 这是因为它们具有更少且有偏的零点。并且, 分数阶矩不变量的最佳分数通常在 $\alpha \simeq 1$ 附近, 这是因为它们具有均匀的零点分布。
- 总体而言, 对于给定方差和 K , EFM/GPCET(1), GPCT(1), GPST(1), CHFM/FJFM(2,1.5,1) 和 JFM(5.5)/FJFM(5,5,1) 表现出更高的分类分数, 尤其是 EFM/GPCET(1)。这意味着它们在全局表征中具有更好的判别性和稳健性综合性能。
- 表 2.6 中 PCANet 和 CBFD 的直接训练, 即直接使用与矩不变量方法相同的训练集, 二者的分类分数较差即使对于干净图像, 这是因为缺乏旋转不变性。
- 在学习表征中, 应对几何变体的常见思路是数据增广, 即在训练集中包含相应的几何变换版本。表 2.6 中 PCANet 和 CBFD 的旋转增广训练取得了性能提升 (尤其是在干净图像上), 但随着噪声方差增加, 分数急剧下降, 与直接训练相差不大。注意, 旋转增广训练将训练时间扩大近 36 倍。
- 大规模深度表征 GoogLeNet 和 ResNet-50 在样本非常少的情况下很难训练, 因此表 2.6 中二者均基于旋转增广训练, 并伴随着高昂的计算代价。从分数来看, 上述类似现象也发生在 GoogLeNet 和 ResNet-50 上: 对于干净测试图像, 分类分数令人满意, 这是因为这些图像与训练数据高度对齐; 然而, 模型对噪声仍然很敏感, 在噪声图像上的分数急剧下降。
- 矩不变量实现了整体更高的分类分数, 相对于各个噪声方差下的学习表征方法。注意, 这是直接训练的基础上实现的, 其时间消耗远远小于学习表征方法。

评注: 这些共性现象证明了, 矩不变量在几何变换和信号损失下的图像识别问题上具有一定优势。矩不变量的不变性和正交性使其在小规模稳健可解释视觉问题上表现出独特性和实用价值。

2.5 本章小结

本章对经典不变量的理论和实践研究现状进行了系统性分析，是对现有综述文献的有益补充，同时作为后续三个创新性课题的研究基础。

本章工作在理论层面的要点如下：

- 我们回顾该领域的发展历史（2.1 节）并梳理其中的核心知识和经典方法（2.2 节），由此抽象出如下核心的研究目标：对仿射变换和信号损失的不变性/鲁棒性，对大量的模式的判别性，以及合理的计算复杂性和准确性。
- 基于这些观察，我们重点分析近期在快速/精确计算、稳健性/不变性优化、定义扩展和应用的最新进展和成功经验，同时总结出经典不变量所面临的判别性瓶颈问题（2.3 节）。

本章工作在实践层面的要点如下：

- 我们提供了系统性的开源实现 MomentToolbox，涵盖了矩和矩不变量的经典方法和一些改进策略。
- 基于统一的软件基准，我们通过矩的计算、图像重构和模式识别实验分别评估了众多矩和矩不变量方法的精确性/复杂性、表征能力和稳健性/不变性（2.4 节）。

本章工作已总结成文：

- Shuren Qi, Yushu Zhang[✉], Chao Wang, et al. A survey of orthogonal moments for image representation: Theory, implementation, and evaluation. **ACM Computing Surveys**, 2023, 55(1): 1-35.

第三章 完善全局不变量

针对不变表征方法在含噪环境中的适用性问题，本学位论文的第一个课题是完善全局不变量，为全局不变量带来了新的时频判别性以及噪声不变性。本章将分别从理论（定义、性质和计算）和应用（经典和可信视觉任务）层面介绍所提出的全局不变量新方法。

3.1 引言

信号损失图像的处理和理解是许多计算机视觉任务中一个基础且具有挑战性的问题^[234]。其中，一类主要的信号损失模型是图像噪声：观察到的版本，即含噪图像 f_η ，是理想场景 f 受到噪声 η 干扰后的结果，即 $f_\eta = f + \eta$ 。

以现代视角来看，图像噪声可能在图像生命周期的三个阶段中发生：1）拍摄噪声是由非理想的摄影过程引起，例如在低光照环境下使用较快的快门速度^[235]；2）传输噪声是由非理想的信道传输过程引起，例如在线社交媒体中的图像预处理^[236]；3）编辑噪声是一种主动的人为扰动，敌手可由编辑软件引入干扰以误导视觉系统，例如对抗性扰动^[237]。

因此，出于有效性和安全性的原则^[238]，人工视觉系统应该对图像噪声稳健。

3.1.1 研究动机

稳健的视觉系统依赖于稳健的数据表征^[15,239]。为了实现噪声稳健的图像表征，以 CNN 为代表的学习表征范式通常诉诸于数据增广训练和去噪预处理这两类策略^[240]。

在训练阶段，增广策略直接在训练集中添加含噪声图像样本以促使 CNN 习得噪声稳健性。当训练和测试具有一致的噪声模式时，该方法在经验上是有效的；相反，对于训练中未见的噪声模式（即使是不同参数下的同类噪声），CNN 表征对噪声的稳健性通常会失效^[241]。另外，这种增广训练显著增加了训练的成本，同时缺乏可解释性上的保证。

在测试阶段，图像去噪可以作为 CNN 表征的预处理，即从观测 f_η 中估计出 $\hat{f}$ 以逼近理想场景 f 。这种去噪模块既有助于提升可解释性，又可与现有表示直接兼容而无需重训练。然而，由于逆问题的病态和不适定性质^[242,243]，去噪结果通常是不稳定的并会引入新的误差，事实上 CNN 对这种图像恢复误差也是敏感的^[142,244]。效率方面，在测试阶段对于每个样本执行去噪预处理所产生的计算成本也非常可观。

另一种范式旨在直接从劣化图像版本中获得稳健表征，无需大规模学习或预处理。

这种内在稳健性的想法可以追溯到上世纪 60 年代，Hu 基于 Hilbert 的代数研究提出了几何不变量^[26]。经过 60 余年的研究，矩和矩不变量已成为解决小规模稳健视觉问题的经典表征策

表 3.1 本章方法与相关方法的概念性对比。

方法	拉东统计量 [245,246]	拉东正交变换 [114,247,248]	图像分数阶矩 [80,159]	拉东整数阶矩 [113]	本章方法
广义性			✓		✓
旋转不变性		✓	✓	✓	✓
噪声稳健性	✓	✓		✓	✓
时频判别性			✓		✓

略^[16]。在这个领域，早期工作（包括 Hu 的不变量）虽然实现多个几何变换群的不变性，但对噪声非常敏感，因为基函数的非正交性^[46]。为此，后续相继发展了一系列正交的矩和矩不变量，包括第二章所介绍的 ZM^[46]、OFMM^[50]、CHFM^[51]、JFM^[53]、EFM^[56]、PCET^[57] 和 BFM^[58]。在实际应用中，这些正交不变量被广泛验证具有以下有益性质：几何不变性以及一定程度的噪声稳健性。然而，以傅里叶分析视角来看，这种稳健性是受限的：只有捕捉低频信息的低阶矩和矩不变量是稳定的，而其余仍对噪声敏感^[159]。特别是当需要更多矩来实现更具判别性的表征时，时常面对一个棘手的矛盾，即判别性和噪声稳健性之间的矛盾。为了缓解这种矛盾，研究者试图将正交矩和矩不变量从图像域扩展到具有较高信噪比的变换域，如拉东变换^[113]。此类拉东矩（Moments in Radon space, MR）已被证明在噪声稳健性方面明显改善，同时保持图像域正交矩和矩不变量的优点，即正交性和几何不变性。

动机。数学上，拉东域的矩和矩不变量研究只限制在阶数为整数的情况。从实际应用来看，这种数学性质将导致其在表征的灵活性和判别性方面受限，特别是表征中图像的空间信息被完全忽略。

3.1.2 核心结论

本章围绕“严重噪声干扰下的图像全局表征”课题展开，尝试给出更系统性的研究结果。我们提出了一种新的时频判别图像表征方法，同时具有噪声稳健性和几何不变性，且无需学习或去噪过程。其在各类小规模稳健视觉问题中具有应用潜力，特别适合对抗性的信息安全和取证场景。

本章工作是对矩和矩不变量研究在理论和应用层面上的扩展。

理论。我们首次将拉东域的矩和矩不变量方法从整数阶空间推广到分数阶空间。数学上，我们分别以隐式和显式方式定义了分数阶拉东矩（Fractional-order Moments in Radon space, FMR）及其不变量，形式化分析了其表征性质。注意，FMR 是一个更广义的设计将诸多拉东域表征囊括为特例，同时保持了有益的噪声稳健性和几何不变性。更重要的是，FMR 中引入的分数阶参数提供了前文诸多表征所不具有的时频判别性。

应用。我们在挑战性噪声干扰下的模式识别实验中全面评估了 FMR 的性能，涵盖与下列

研究路径的对比：图像域矩和矩不变量、轻量和深度学习表征、以及数据增广训练和去噪预处理。注意，这样全面的噪声稳健性评估鲜见于相关工作，尤其涵盖与学习表征路径的对比。此外，我们还在面向可信场景的现实世界应用中评估了 FMR 的性能：自然畸变下的模板匹配和用于版权保护的零水印。上述基准实验和真实世界应用都表明 FMR 在噪声稳健性、旋转不变性和时频判别性方面具有独特性和实用性。

3.2 预备和基础

本节介绍预备概念，即相关的拉东域表征、分数阶矩和矩不变量、以及作为另一种范式的图像去噪；并且介绍数学基础，即拉东变换定义、符号和命题。

3.2.1 预备概念

拉东域表征。拉东变换 (Radon Transform) 是一种经典的图像投影运算，广泛应用于医学 (成像重构)、工程学 (电子显微镜)、物理学 (光学) 和材料科学 (晶体学) 中。在计算机视觉和模式识别中，拉东变换由于其正交性、低复杂度、抗噪性和对几何变换的有益性质而受到关注。最初，一些基于拉东变换的表征方法，例如 R -signature^[249]、 Φ -signature^[250]、RCF^[245] 和 HRT^[246]，是拉东变换系数的简单统计 (例如求和^[249,250]、最大值^[245] 或直方图^[246])。一般来说，这些非正交表征会大量丢弃判别性信息，限制了其在更大规模视觉任务中的应用。因此，其他基于拉东变换的表征方法，如 R2DFM^[247]、RWF^[248]、RMF^[114] 和 MR^[113]，则借助正交变换 (如傅里叶变换^[247]、小波变换^[248]、梅林变换^[114] 或正交矩^[113])，以实现更具判别性的表征。在这里，MR 表征框架可被视作是 R2DFM 和 RMF 在基本定义和不变性分析方面的推广。注意，对于上述基于正交变换的拉东域表征，其定义仅限制在整数阶空间。后面的分析将表明这种数学上的限制将导致其在表征的灵活性和判别性方面受限，特别是表征中图像的空间信息被完全忽略。

分数阶矩和矩不变量。分数阶矩 (Fractional-order Moments, FM) 和矩不变量的研究主要起源于信息抑制问题。数学上，基函数的零点分布与表征在图像平面上描述重点密切相关^[38]。信息抑制问题即矩和矩不变量表征强调图像中不具有判别性信息的部分，由于有偏的零点分布所造成。在计算机视觉和模式识别中，Kan 等人首先在 ZM 和 OFMM 的实证对比中研究了信息抑制问题^[251]。Hoang 等人提出了第一类分数阶矩和矩不变量，即 GPHT^[80]，涵盖一系列基于谐波函数的圆谐正交矩定义。其能够通过改变分数阶参数来控制零点分布，从而有助于解决信息抑制问题并提取局部特征。之后，一系列分数阶矩和矩不变量相继提出^[158,186,188,189]。最近，Yang 等人提出了 FJFM^[159]，涵盖一系列基于雅可比多项式的圆谐正交矩定义。作者通过其时频判别性的系统性分析发现，将具有基函数零点分布互补的低阶矩和矩不变量组合使用可以捕捉图像中互补的稳健特征，从而能够缓解稳健性和判别性之间的矛盾。其中，GPHT 和 FJFM 是更通

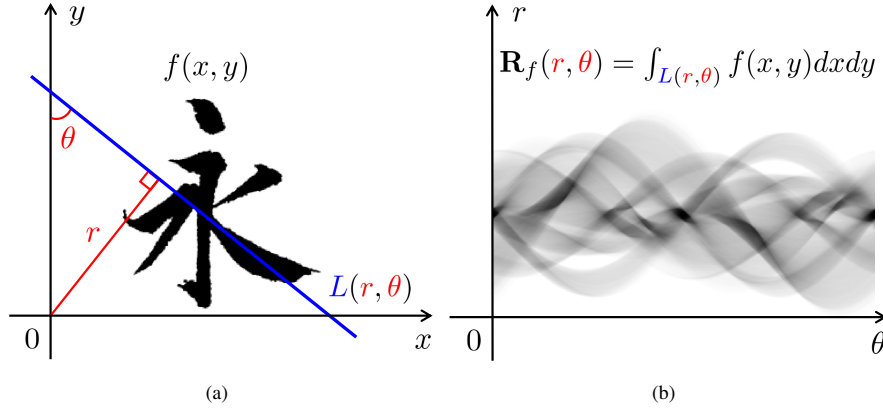


图 3.1 拉东变换示意图：(a) 图像函数 $f(x, y)$ ，与原点距离 r 且与 y 轴夹角为 θ 的直线 $L(r, \theta)$ ；(b) 沿着该直线的线积分， $\int_{L(r, \theta)} f(x, y) dx dy$ ，即对应拉东变换 $\mathbf{R}_f(r, \theta)$ 。

用的定义，可作为分数阶矩和矩不变量研究的统一数学工具。需要注意的是，对于上述分数阶矩和矩不变量，其定义仅限制在原始图像空间。而在拉东空间中，此类方法的定义、噪声稳健性、旋转不变性、时频判别性以及应用还尚未得到研究。

图像去噪。与直接增强图像表征稳健性不同，图像增强和恢复，如图像去噪^[252]，是另一大类研究范式。为了从含噪观测中最大近似恢复其原始场景，研究者通常借助关于噪声和场景的先验知识^[242, 243]，鉴于逆问题的病态和不适定性。手工时代，去噪算法在设计上主要受噪声和场景信号的分布模式所启发。在这方面，里程碑是 DM3D^[253]，基于以下先验知识的显式建模：自然图像在空间域的局部平滑性和非局部自相似性，以及噪声在频率域的统计规律。深度学习时代，普遍认为此类分布模式可以通过深度表示来隐式地建模，以覆盖更复杂和更真实的分布先验。在这方面，为去噪专门设计的深度网络，如 DnCNN^[112]、FFDNet^[254] 和 SRMD^[255]，能够自适应地学习从含噪图像到干净图像的映射，实现了人类感知视角下的先进结果。对于上述手工和深度去噪方法，其性能通常以所得图像与干净图像的视觉相似性来评价^[252]，包括自动的质量评估或基于用户体验的打分。然而在文献中，去噪预处理和后续的视觉任务很少组合起来进行优化与评价，导致去噪算法走向应用层面存在一定程度的困难^[142, 244]。具体而言，这种困难表现为以下形式：1) 由于缺乏噪声先验而导致的适用性问题，2) 由去噪过程引入误差而导致的准确性和稳健性问题，3) 以及由预处理导致的效率问题。

3.2.2 数学基础

定义 3.1：拉东变换。图像函数 f 的拉东变换，记作 $\mathbf{R}_f$ ，定义为沿直线 L 的线积分^[256]：

$$\mathbf{R}_f(r, \theta) = \int_{L(r, \theta)} f(x, y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(r - x \cos \theta - y \sin \theta) dx dy, \quad (3.1)$$

其中, $L(r, \theta) = \{(x, y) \in \mathbb{R}^2 \text{ s.t. } r - x \cos \theta - y \sin \theta = 0\}$ 是由距离参数 r (相对于原点) 和角度参数 θ (相对于 y 轴) 定义的直线; δ 是克罗内克函数: $\delta(\cdot) = [\cdot = 0]$ 。

在图 3.1 中, 我们给出了上述参数或术语的几何图示, 方便读者对拉东变换形成直观理解。

命题 3.1: 拉东变换对旋转的协变性。 令 f_R 为图像 f 的旋转版本, 其中旋转的角度为 φ 并围绕图像中心, 则 f_R 的拉东变化是 $\mathbf{R}_f$ 关于 θ 的循环平移版本, 且移动相同的 φ ^[257]:

$$\mathbf{R}_{f_R}(r, \theta) = \mathbf{R}_f(r, \theta + \varphi), \quad (3.2)$$

上述公式即拉东变换对旋转的协变性。

命题 3.2: 拉东变换对加性白噪声的稳健性。 令 f_η 为图像 f 关于加性白噪声 η 的含噪版本, 即 $f_\eta = f + \eta$, 则在连续域假设下 f_η 和 f 拉东变换满足恒等关系^[116]:

$$\mathbf{R}_{f_\eta}(r, \theta) = \mathbf{R}_f(r, \theta) + \mathbf{R}_\eta(r, \theta) = \mathbf{R}_f(r, \theta), \quad (3.3)$$

其中, $\mathbf{R}_\eta(r, \theta) = 0$ 因为加性噪声的拉东变换正比于噪声的期望值。注意, 对于实际的离散域假设, 上述恒等关系不再成立。令 f_η 和 f 经过采样和量化的数字图像, 其中二者尺寸均为 $N \times M$ 像素、图像 f 的均值为 μ 、噪声 η 的方差为 σ^2 , 则含噪图像与其拉东变换在离散域中的信噪比 (Signal-to-Noise Ratio, SNR) 满足如下关系^[258]:

$$\text{SNR}_{\mathbf{R}_{f_\eta}(\cdot, \theta)} = \text{SNR}_{f_\eta} + \frac{\mu^2(c(\theta) - 1)}{\sigma^2}, \quad (3.4)$$

其中, $\mathbf{R}_{f_\eta}(\cdot, \theta)$ 是含噪图像 f_η 沿角度参数 θ 所形成的拉东变换 (一维投影), $c(\theta)$ 是关于上述过程的常数项, 取值大致在区间 $[\min(N, M), \max(N, M)]$ 中。

该命题表明, 拉东投影后产生的信噪比增量为 $\mu^2(c(\theta) - 1)/\sigma^2$, 在实践中其取值通常很大, 意味着噪声稳健性的大幅度提升。直观起见, 下面给出了一个相关的例子。

举例 3.1: 假设图像像素的均值为 $\mu = 0.5$, 噪声方差为 $\sigma^2 = 0.1$ (二者均基于归一化灰度), 并且图像尺寸为 $N = M = 256$, 则拉东投影产生的信噪比增量为 $\mu^2(c(\theta) - 1)/\sigma^2 \approx 637.5$ 。

3.3 隐式路径

本节介绍构建 FMR 的隐式路径。我们从数学定义开始, 随后深入分析派生出的有益特性, 最后设计了精确且快速的计算策略。

3.3.1 隐式定义

数学上, 本章工作的一般理论基于以下定义, 即圆谐基函数与图像拉东变换之间的内积。

定义 3.2: **FMR 的隐式定义**。对于图像函数 f , 其 FMR 定义:

$$\langle \mathbf{R}_f, V_{nm}^\alpha \rangle = \iint_D [V_{nm}^\alpha(r, \theta)]^* \mathbf{R}_f(r, \theta) r dr d\theta, \quad (3.5)$$

其中, $\mathbf{R}_f$ 是 f 的拉东变换, 由公式 (3.1) 定义, V_{nm}^α 是具有阶数参数 $(n, m) \in \mathbb{Z}^2$ 和分数阶参数 $\alpha \in \mathbb{R}^+$ 的基函数, 定义域为 $D \in \mathbb{R}^2$, $*$ 表示复共轭。这里, 我们对基函数的定义施加了两个有益的约束条件, 即正交性和旋转不变性, 引导出以下基函数形式:

$$V_{nm}^\alpha(r, \theta) = R_n^\alpha(r) A_m(\theta), \quad (3.6)$$

其中, 角向基函数 $A_m(\theta) = \exp(jm\theta)$ ($j = \sqrt{-1}$), 径向基函数 $R_n^\alpha(r)$ 满足加权正交条件 $\int_0^1 R_n^\alpha(r) [R_{n'}^\alpha(r)]^* r dr = \frac{1}{2\pi} \delta_{nn'}$, 定义域为单位圆区域 $D = \{(r, \theta) : r \in [0, 1], \theta \in [0, 2\pi)\}$ 。

在上述定义中, R_n^α 的具体形式尚未给出, 实际上它可以由数学文献中几乎任何一种正交函数定义^[16]。为了使图像表征及其离散实现表现出良好性质, 同时不失一般性, 我们介绍了如下两类分数阶正交函数作为径向基函数, 分别基于谐波和多项式定义。

定义 3.3: **谐波径向基函数**。基于傅里叶分析中复指数函数^[80] 定义 R_n^α , 得到:

$$R_n^\alpha(r) = \sqrt{\frac{\alpha r^{\alpha-2}}{2\pi}} \exp(j2n\pi r^\alpha), \quad (3.7)$$

对于任意 α , 上述定义均满足加权正交条件。

定义 3.4: **多项式径向基函数**。基于经典正交多项式中的雅可比多项式^[159] 定义 R_n^α , 得到:

$$R_n^\alpha(p, q, r) = \sqrt{\frac{\alpha r^{\alpha q-2} (1-r^\alpha)^{p-q} (p+2n)\Gamma(q+n)n!}{2\pi\Gamma(p+n)\Gamma(p-q+n+1)}} \sum_{k=0}^n \frac{(-1)^k \Gamma(p+n+k) r^{\alpha k}}{k!(n-k)!\Gamma(q+k)}, \quad (3.8)$$

其中, $n \in \mathbb{N}$, 多项式参数 $p, q \in \mathbb{R}$ 应满足: $p - q > -1$, $q > 0$; 同样地, 对于任意 α 、 p 和 q , 上述定义均满足加权正交条件。

定义的总结: 本章后续将由谐波或多项式径向基函数定义的 FMR 分别简称为谐波/多项式 FMR。通过定义 1~4, 我们已经形式化了本章工作中最基础的部分: 在拉东空间中设计了一个通用的表示框架, 其满足正交性和旋转不变性约束, 并可选择谐波或多项式作为径向基函数。

3.3.2 性质

接下来将讨论从上述定义中得出的重要性质, 包括广义性、旋转不变性、噪声稳健性和时频判别性。

性质 3.1: 广义性。通过引入分数阶参数 α , FMR 将基于谐波或多项式径向基函数的拉东

整数阶矩均囊括为特例。

评注：现有拉东矩的相关工作，例如^[113]，都是 FMR 在固定参数 $\alpha = 1$ 或 $\alpha = 2$ 时的特例。注意，通过改变 α 值可得到一系列拉东域表征，这一系列表征既具有共同的稳健性或不变性优势（即性质 3.2 和 3.3），同时又在判别性信息上表现出互补性（即性质 3.4）。

性质 3.2：旋转不变性。令 f_R 为图像 f 的旋转版本，其中旋转的角度为 φ 并围绕图像中心，则必定存在除常值函数外的函数 $\mathcal{I}$ 使得下列等式成立：

$$\mathcal{I}(\{\langle \mathbf{R}_{f_R}, V_{nm}^\alpha \rangle\}) \equiv \mathcal{I}(\{\langle \mathbf{R}_f, V_{nm}^\alpha \rangle\}), \quad (3.9)$$

即对任意输入图像和 FMR 参数满足旋转不变性。

证明：基于命题 3.1 和定义 3.2，可以验证旋转图像与原始图像的 FMR，即 $\langle \mathbf{R}_{f_R}, V_{nm}^\alpha \rangle$ 和 $\langle \mathbf{R}_f, V_{nm}^\alpha \rangle$ ，之间只有与旋转角 φ 相关的相位差异：

$$\begin{aligned} & \langle \mathbf{R}_{f_R}(r, \theta), V_{nm}^\alpha(r, \theta) \rangle \\ &= \langle \mathbf{R}_f(r, \theta + \varphi), V_{nm}^\alpha(r, \theta) \rangle \\ &= \langle \mathbf{R}_f(r, \theta'), V_{nm}^\alpha(r, \theta' - \varphi) \rangle \\ &= \langle \mathbf{R}_f(r, \theta'), V_{nm}^\alpha(r, \theta') \rangle [A_m(-\varphi)]^* \\ &= \langle \mathbf{R}_f, V_{nm}^\alpha \rangle A_m(\varphi), \end{aligned} \quad (3.10)$$

其中， $\theta' = \theta + \varphi$ 。注意，第一个等号和第二个等号分别基于 $\mathbf{R}_{f_R}(r, \theta) = \mathbf{R}_f(r, \theta + \varphi)$ 和 $A_m(\theta) = \exp(jm\theta)$ 。因此，我们应该寻找函数 $\mathcal{I}$ 以适当抵消额外相位项 $A_m(\varphi)$ 的影响。例如，一个朴素的思路是直接利用幅度运算去除表征的相位信息：

$$\mathcal{I}(\langle \mathbf{R}, V_{nm}^\alpha \rangle) \triangleq |\langle \mathbf{R}, V_{nm}^\alpha \rangle|, \quad (3.11)$$

显然满足 $|\langle \mathbf{R}_{f_R}, V_{nm}^\alpha \rangle| = |\langle \mathbf{R}_f, V_{nm}^\alpha \rangle A_m(\varphi)| = |\langle \mathbf{R}_f, V_{nm}^\alpha \rangle|$ 。为了实现更具判别性和信息量的表征，另一种方法将幅度运算推广到更灵活的相位抵消，例如^[110]：

$$\mathcal{I}(\langle \mathbf{R}, V_{nm}^\alpha \rangle) \triangleq \prod_{i=1}^L (\langle \mathbf{R}, V_{nm_i}^\alpha \rangle)^{k_i}, \quad (3.12)$$

其中， $L \geq 1$ ， $m_i, k_i \in \mathbb{Z}$ ， $i = 1, \dots, L$ ，满足 $\sum_{i=1}^L m_i k_i = 0$ 。 □

性质 3.3：对加性白噪声的稳健性。令 f_η 为图像 f 关于加性白噪声 η 的含噪版本，即 $f_\eta =$

$f + \eta$ ，则拉东域 FMR 表征相比于其对应的图像域表征 FM 更稳健：

$$\|\{\langle \mathbf{R}_{f_\eta}, V_{nm}^\alpha \rangle\} - \{\langle \mathbf{R}_f, V_{nm}^\alpha \rangle\}\|_l \leq \|\{\langle f_\eta, V_{nm}^\alpha \rangle\} - \{\langle f, V_{nm}^\alpha \rangle\}\|_l, \quad (3.13)$$

其中，稳健性以 l 范数下的表征类内距离衡量，并且当且仅当 $\eta = 0$ 时等号成立。

证明：连续域假设下，根据命题 3.2 和定义 3.2，我们可以将公式 (3.13) 的左侧重写为：

$$\begin{aligned} & \|\{\langle \mathbf{R}_{f_\eta}, V_{nm}^\alpha \rangle\} - \{\langle \mathbf{R}_f, V_{nm}^\alpha \rangle\}\|_l \\ &= \|\{\langle \mathbf{R}_{f_\eta} - \mathbf{R}_f, V_{nm}^\alpha \rangle\}\|_l \\ &= \|\{\langle 0, V_{nm}^\alpha \rangle\}\|_l = 0, \end{aligned} \quad (3.14)$$

其中第一个等号源自内积的线性性质，第二个等号源自公式 (3.3)。同样地，公式 (3.13) 的右侧可重写为：

$$\begin{aligned} & \|\{\langle f_\eta, V_{nm}^\alpha \rangle\} - \{\langle f, V_{nm}^\alpha \rangle\}\|_l \\ &= \|\{\langle f_\eta - f, V_{nm}^\alpha \rangle\}\|_l \\ &= \|\{\langle \eta, V_{nm}^\alpha \rangle\}\|_l \geq 0, \end{aligned} \quad (3.15)$$

在这里我们可以断言 $\|\{\langle \eta, V_{nm}^\alpha \rangle\}\|_l \geq 0$ ，当且仅当 $\eta = 0$ 时等号成立。不失一般性，我们假设存在一个非零的 η ，使得 $\|\{\langle \eta, V_{nm}^\alpha \rangle\}\|_l = 0$ 。基于这个假设，可知 $\eta \notin \{V_{nm}^\alpha\}$ ，因为如果 $\eta \in \{V_{nm}^\alpha\}$ ，则 $\|\{\langle \eta, V_{nm}^\alpha \rangle\}\|_l = \|\delta\|_l \neq 0$ 。综上所述，该假设意味着 η 是一个非零元素，不属于基函数集合，并且正交于该集合中的任何元素以确保 $\|\{\langle \eta, V_{nm}^\alpha \rangle\}\|_l = 0$ 。这显然与 Hilbert 空间中基函数的完备性和正交性相矛盾。因此该假设不成立。可以注意到公式 (3.14) 和 (3.15) 的正确性实际上保证了原始性质 (3.13) 的正确性。 $\square$

尽管离散域中等式 (3.14) 不再成立，但 (3.4) 仍然保证了 FMR 的稳健性，因为噪声图像 f_η 的信噪比显著低于其拉东变换 $\mathbf{R}_{f_\eta}$ 的信噪比。

性质 3.4：时频判别性。在 FMR 表征中，图像信息的频率和空间特性可分别用参数 (n, m) 和 α 来控制。

评注：在矩和矩不变量的研究中，已经发现表征信息的频率和空间特性分别取决于基函数零点的数量和位置。对于 FMR，基函数零点的数量和位置可以分别通过 (n, m) 和 α 显式地调整，这是实现时频判别性的核心。

- 数量方面，径向零点数量和径向零点数量分别与 n 和 m 成正比关系，源自谐波函数和多项式函数的固有性质。
- 位置方面，当 $\alpha < 1$ 时径向零点位置会向 0 偏移，当 $\alpha = 1$ 时径向零点位置通常是均匀

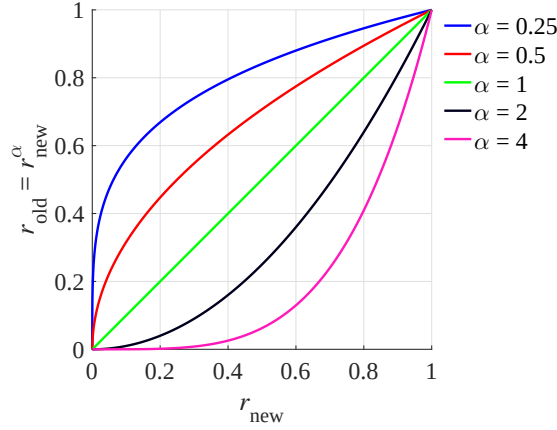


图 3.2 引入分数阶参数 $\alpha \in \mathbb{R}^+$ 对径向自变量的影响，其中 $r_{\text{new}} \in [0, 1]$ 替换了 $r_{\text{old}} \in [0, 1]$ ，基于 $r_{\text{old}} = r_{\text{new}}^\alpha$ 。

的，而当 $\alpha > 1$ 时径向零点位置会向 1 偏移，其中 α 与 1 的偏差越大，零点分布就越向一边偏斜^[159]。原因在于，在定义域 $(0, 1)$ 上，当 α 小于/大于 1 时 r^α 始终大于/小于 r^1 ，这对应着零点的左/右移动，如图 3.2 所示。

注意，符号 V_{nm}^α 的上标和下标分别用于表示空间域和频率域的参数。

直观起见，我们在图 3.3 中以不同的参数展示了 V_{nm}^α 的相位情况： $n = 5$ ， $m = \{2, 4, 8\}$ ， $\alpha = \{1/2, 1, 2\}$ 。从上述理论分析我们可以预期，改变 m 会改变基函数在角向上的零点数量，进而对应于频率性质的变化。实际上 n 也会有类似的现象，尽管在图 3.3 中没有展示出来。至于引入的 α ，改变其值的会改变基函数在径向上零点的分布，进而对应于空间性质的变化。这种时频判别性在以前的整数阶方法中无法实现，该方法只对应于 $\alpha = 1$ 的情况。

性质的总结：值得强调的是，FMR 相对于理论相关方法表现出独特性。如表 3.1 所示，与其他四类理论相关的表征方法相比，我们的方法是一个更全面的框架：同时具有上述性质 3.1 ~ 3.4。其中，图像域矩和矩不变量不具备性质 3.3，拉东域的整数阶矩和正交变换不具备性质 3.1 和 3.4，而其他的拉东域表征进一步丧失了性质 3.2。

3.3.3 计算

以上 3.3.1 和 3.3.2 节已对 FMR 的定义和性质进行了描述，均在连续域假设下。接下来，我们将重点考虑 FMR 在离散域假设下的数值计算，考虑其准确性和复杂性。

对于谐波 FMR，我们引入以下基于傅里叶变换的定义，这可进一步引出计算高效和数值稳定的计算。

定理 3.1：谐波 FMR 的傅里叶定义——解析公式。由定义 3.2 和 3.3 给出的谐波 FMR 可被重写为以下傅里叶形式^[259]：

$$\langle \mathbf{R}_f, V_{nm}^\alpha \rangle = 2\pi \mathcal{F}(\mathcal{S}(\gamma, \vartheta)), \quad (3.16)$$

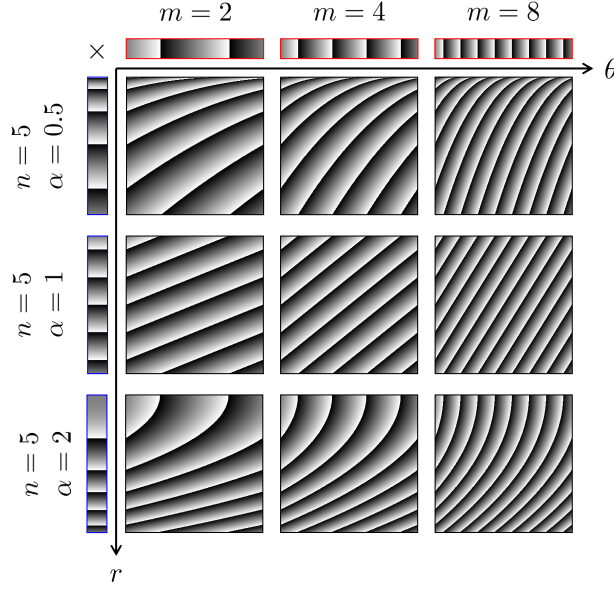


图 3.3 时频判别性示意图：在不同参数设置下，角向基函数 $A_m(\theta)$ 、谐波径向基函数 $R_n^\alpha(r)$ 、以及二维基函数 $V_{nm}^\alpha(r, \theta)$ 的相位分别用红色、蓝色和黑色边框绘制；其中， n 和 m 分别编码径向和角度频率性质，而新引入的 α 则编码径向空间性质（之前的整数阶方法无法实现，即 $\alpha = 1$ ）。

其中， $\gamma = r^\alpha$ ， $\vartheta = \theta/2\pi$ ， $\mathcal{S}(\gamma, \vartheta) = \sqrt{\gamma^{\alpha/2-1}/2\pi\alpha} \times \underbrace{\mathbf{R}_f(\gamma^{\frac{1}{\alpha}})}_r, \underbrace{2\pi\vartheta}_\theta$ 。

证明：将谐波径向基函数（3.7）代入定义 3.2 的公式（3.5）和（3.6），则谐波 FMR 可展开为：

$$\begin{aligned}
 \langle \mathcal{R}_f, V_{nm}^\alpha \rangle &= \int_0^1 \int_0^{2\pi} [R_n^\alpha(r) A_m(\theta)]^* \mathcal{R}_f(r, \theta) r dr d\theta \\
 &= \int_0^1 \int_0^{2\pi} \sqrt{\frac{\alpha r^{\alpha-2}}{2\pi}} \exp(-j2n\pi r^\alpha) \exp(-jm\theta) \mathcal{R}_f(r, \theta) r dr d\theta \\
 &= \int_0^1 \int_0^{2\pi} \sqrt{\frac{\alpha r^\alpha}{2\pi}} \exp(-j2n\pi r^\alpha) \exp(-jm\theta) \mathcal{R}_f(r, \theta) dr d\theta \\
 &= 2\pi \int_0^1 \int_0^1 \mathcal{S}(\gamma, \vartheta) \exp(-jn2\pi\gamma) \exp(-jm2\pi\vartheta) d\gamma d\vartheta \\
 &= 2\pi \mathcal{F}(\mathcal{S}(\gamma, \vartheta)),
 \end{aligned}$$

其中，符号 γ ， ϑ 和 $\mathcal{S}(\gamma, \vartheta)$ 如前定义；最后一个等号源自二维傅里叶变换 $\mathcal{F}$ 的定义。 $\square$

我们给出以下引理作为解析公式（3.16）对应的离散版本，并引入快速傅里叶变换实现高效和数值稳定的计算。

引理 3.1：谐波 FMR 的傅里叶定义——数值公式。基于傅里叶基函数和 $\mathcal{S}(\gamma, \vartheta)$ 的离散采

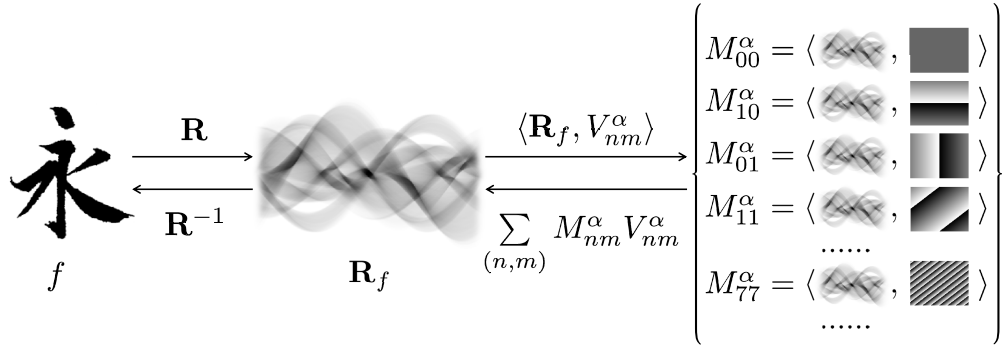


图 3.4 计算过程示意图：首先将原始图像 f 投影到拉东空间中得到 $\mathbf{R}_f$ ，然后在不同参数 (α, n, m) 下计算 $\mathbf{R}_f$ 与基函数 V_{nm}^α 的内积，得到 FMR 系数 $M_{nm}^\alpha = \langle \mathbf{R}_f, V_{nm}^\alpha \rangle$ ；其中，内积的计算和基函数值的估计可以分别通过傅里叶和递归策略高效地实现，并且以上变换是可逆的。

样和零阶近似，解析公式 (3.16) 可重写为近似的离散形式^[259]：

$$\langle \mathbf{R}_f, V_{nm}^\alpha \rangle \simeq \frac{2\pi}{M^2} \mathcal{F}[S[u, v]], \quad (3.17)$$

其中， $S[u, v] \triangleq S(\gamma_u, \vartheta_v) = \sqrt{\frac{\gamma_u^{\alpha/2-1}}{2\pi\alpha}} \mathbf{R}_f(r_u, \theta_v)$ （方括号表示使用离散自变量）基于如下采样：

$$\begin{cases} \gamma_u = r_u^\alpha = \frac{u}{M} \\ \vartheta_v = \frac{\theta_v}{2\pi} = \frac{v}{M} \end{cases}, \begin{cases} \Delta\gamma_u = \frac{1}{M} \\ \Delta\vartheta_v = \frac{1}{M} \end{cases}, \begin{cases} r_u = (\frac{u}{M})^{\frac{1}{\alpha}} \\ \theta_v = \frac{2\pi v}{M} \end{cases} \quad (3.18)$$

其中， $(u, v) \in \{0, 1, \dots, M-1\}^2$ ， M 是与采样率相关的常数。在我们的实现中，快速傅里叶变换算法用以计算公式 (3.17)。

评注：结合解析定理 3.1 和相应的数值引理 3.1，上述谐波 FMR 的实现策略比基于定义 3.2 和 3.3 的直接实现更加稳定和快速：

- 稳定性方面，直接实现和上述傅里叶实现分别建立在笛卡尔坐标系和极坐标系中。基于转换关系 $dxdy = r dr d\theta$ ，这两种实现中径向部分具有完全不同的有界性：对于 $\alpha \in (0, 2)$ ，当 $(x, y) \rightarrow (0, 0)$ 时， $\lim_{(x,y) \rightarrow (0,0)} |R_n^\alpha(\sqrt{x^2 + y^2})| = \infty$ ；对于 $\alpha \in (0, +\infty)$ ，当 $r \rightarrow 0$ 时， $\lim_{r \rightarrow 0} |r \cdot R_n^\alpha(r)| = 0$ 。显然，傅里叶实现在 $r \simeq 0$ 时具有更好的稳定性。
- 复杂性方面，考虑计算一组 $\langle \mathbf{R}_f, V_{nm}^\alpha \rangle$ ，其中 $(n, m) \in \{-K, \dots, 0, 1, \dots, K\}^2$ 且 α 固定，图像大小为 $N \times N$ 且采样参数 $M \propto N$ 。直接实现需要 $O(N^2(2K+1)^2) = O(N^2K^2)$ 次乘法，因为要在笛卡尔采样上逐点相乘。傅里叶实现则需要 $O(M^2 \log M) = O(N^2 \log N)$ 次乘法，其中 K 对计算复杂性没有影响，即相对于阶数采样数的常数阶计算。由于 K^2 和 $\log N$ 的增长特性显著不同，傅里叶实现具有更好的计算效率。

对于多项式 FMR，我们引入以下基于递归的定义，这可进一步引出计算高效和数值稳定的计算。

定理 3.2: 多项式 FMR 的递归定义——解析公式。由定义 3.4 给出的多项式径向基函数可被重写为以下递归形式^[159]:

$$R_n^\alpha(p, q, r) = \sqrt{\frac{(p+2n)\alpha r^{\alpha q-1}(1-r^\alpha)^{p-q}}{r}} C_n(p, q) P_n^\alpha(p, q, r), \quad (3.19)$$

其中, $P_n^\alpha(p, q, r)$ 被递归定义为:

$$P_n^\alpha = (L_1 r^\alpha + L_2) P_{n-1}^\alpha + L_3 P_{n-2}^\alpha, n \geq 2, \quad (3.20)$$

具有如下因子 $L_1 = -(2n+p-1)(2n+p-2)/(n(q+n-1))$, $L_2 = (p+2n-2) + L_1(n-1)(q+n-2)/(p+2n-3)$, $L_3 = (p+2n-4)(p+2n-3)/2 + L_1(q+n-3)(n-2)/2 - (p+2n-4)L_2$; 递归初始值为 $P_0 = \Gamma(p)/\Gamma(q)$, $P_1 = (1-r^\alpha(p+1)/q)\Gamma(p+1)/\Gamma(q)$ 。类似的, $C_n(p, q)$ 也被递归定义为:

$$C_n = \sqrt{\frac{n(q+n-1)}{(p+n-1)(p-q+n)}} C_{n-1}, n \geq 1, \quad (3.21)$$

递归初始值为 $C_0 = \sqrt{\Gamma(q)/(\Gamma(p)\Gamma(p-q+1))}$ 。

注意, 定理 3.2 的证明很直接, 仅依赖性质 $a! = a \cdot (a-1)!$ 和 $\Gamma(a) = a \cdot \Gamma(a-1)$ 。此外, 将上述解析公式转换为数值公式也非常直接, 即上述递归关系在采样上直接成立。因此, 本章将不再赘述定理 3.2 的证明和数值版本。

评注: 基于定理 3.2, 上述多项式 FMR 的实现策略比基于定义 3.2 和 3.4 的直接实现更加稳定和快速:

- 稳定性方面, 公式 (3.8) 的阶乘/伽玛项是导致数值不稳定的主要因素。例如, 在一个典型的浮点算术标准中, 当 $a > 21$ 时 $a!$ 将超出表示范围并发生数值不稳定性。对比公式 (3.8) 直接实现和上述递归实现, 可以注意到递归实现中避免了大数的阶乘/伽玛项, 因而其稳定性更好。
- 复杂性方面, 考虑计算 $R_n^\alpha(p, q, r)$ 在 $N \times N$ 采样点 (即数字图像像素) 上的基函数值, 其中 n 和 α 均固定。直接实现需要 $O(nN^2)$ 次加法, 因为在公式 (3.8) 中需要对 $N \times N$ 个采样点进行 n 次求和。递归实现则从已经得到的 R_{n-1}^α 和 R_{n-2}^α 直接导出, 仅需 $O(N^2)$ 次加法运算来进行一次求和。若考虑计算一组 $R_n^\alpha(p, q, r)$, 其中 $n \in \{0, 1, \dots, K\}^2$, α 固定。直接和递归实现的加法复杂度分别为 $O(K^2N^2)$ 和 $O(KN^2)$ 。显然, 递归实现的计算效率更好。

计算的总结: 直观起见, 我们回顾整个实现过程, 如图 3.4 所示。原始图像首先通过拉东变换 (3.2.2 节) 得到投影, 然后投影分解为正交矩 (3.3 节)。对于分解过程中的内积的计算和基函数值的估计, 分别采用傅里叶变换 (定理 3.1 与引理 3.1) 和递归公式 (定理 3.2) 进行稳

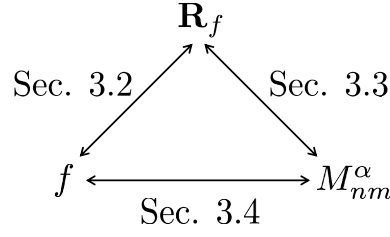


图 3.5 隐式路径和显式路径关系示意图：显式路径可直接从原始图像 f 导出 FMR 系数 M_{nm}^α ，无需拉东变换 $\mathbf{R}_f$ 作为中间步骤。

定和快速的实现。

3.4 显式路径

本节介绍构建 FMR 的显式路径。与上述必须通过拉东变换的间接定义不同，我们试图给出直接从图像到 FMR 的显式定义。

3.4.1 从隐式到显式

从隐式路径和显式路径定义 FMR 的关系如图 3.5 所示。数学上，3.2.2 节中原始图像 f 到拉东投影 $\mathbf{R}_f$ 是通过基 $\mathcal{B}_1 = \delta(r - x \cos \theta - y \sin \theta)$ 实现的；3.3 节中拉东投影 $\mathbf{R}_f$ 到分数阶圆谐正交矩 M_{nm}^α 是通过基 $\mathcal{B}_2 = V_{nm}^\alpha(r, \theta)$ 实现的。因此，一个自然的思路是通过级联基函数 $\mathcal{B} = \mathcal{B}_1 \circ \mathcal{B}_2$ ，使得原始图像 f 可以直接映射到 FMR 系数 M_{nm}^α 。

注意，以下形式化定义主要作为补充性的数学材料，可供进一步研究参考。本章的分析和实验仍基于 3.2.2 和 3.3 节的隐式定义。

3.4.2 显式定义

对于谐波 FMR，我们引入以下显式定义，可解释为两个积分项的线性组合。

定理 3.3：谐波 FMR 的显式定义。由定义 3.2 和 3.3 给出的谐波 FMR 可被重写为如下显式形式^[113]：

$$\langle \mathbf{R}_f, V_{nm}^\alpha \rangle = \sum_{k=0}^{\infty} W_1^*(n, k) \sum_{t=0}^{\infty} \binom{\alpha k + \frac{\alpha}{2}}{t} \Theta_{\alpha k + \frac{\alpha}{2} - t, t} G_{\alpha k + \frac{\alpha}{2} - t, t}, \quad (3.22)$$

其中，系数项 W_1 定义为：

$$W_1(n, k) = \frac{\sqrt{\alpha}(j2n\pi)^k}{\sqrt{2\pi}k!}, \quad (3.23)$$

两个积分项 Θ 和 G 分别定义为：

$$\Theta_{\xi_1, \xi_2} = \int_{\theta} A_m^*(\theta) (\cos \theta)^{\xi_1} (\sin \theta)^{\xi_2} d\theta, \quad (3.24)$$

$$G_{\xi_1, \xi_2} = \int_x \int_y f(x, y) x^{\xi_1} y^{\xi_2} dx dy, \quad (3.25)$$

其中 G 也就是矩和矩不变量的经典概念几何矩^[26]。

证明：我们首先将 (3.7) 中的谐波径向基函数重写为关于自变量 r 和系数 W_1 的多项式形式：

$$R_n^\alpha(r) = \sum_{k=0}^{\infty} W_1(n, k) r^{\alpha k + \frac{\alpha}{2} - 1}, \quad (3.26)$$

上述等价形式由指数函数的泰勒级数导出，即 $\exp(\cdot) = \sum_{k=0}^{\infty} \cdot^k / k!$ 。

接下来，谐波 FMR 隐式定义中的 $\mathbf{R}_f$ 和 R_n^α 将分别展开，得出以下显式形式：

$$\begin{aligned} \langle \mathcal{R}_f, V_{nm}^\alpha \rangle &= \int_r \int_\theta [R_n^\alpha(r) A_m(\theta)]^* \mathcal{R}_f(r, \theta) r dr d\theta \\ &= \int_r \int_\theta [R_n^\alpha(r) A_m(\theta)]^* \int_x \int_y f(x, y) \delta(r - x \cos \theta - y \sin \theta) dx dy r dr d\theta \\ &= \int_r \int_\theta \int_x \int_y f(x, y) A_m^*(\theta) \delta(r - x \cos \theta - y \sin \theta) \sum_{k=0}^{\infty} W_1^*(n, k) r^{\alpha k + \frac{\alpha}{2}} dx dy r dr d\theta \\ &= \sum_{k=0}^{\infty} W_1^*(n, k) \int_\theta \int_x \int_y f(x, y) A_m^*(\theta) (x \cos \theta + y \sin \theta)^{\alpha k + \frac{\alpha}{2}} dx dy d\theta \\ &= \sum_{k=0}^{\infty} W_1^*(n, k) \sum_{t=0}^{\infty} \binom{\alpha k + \frac{\alpha}{2}}{t} \int_\theta \int_x \int_y f(x, y) A_m^*(\theta) (x \cos \theta)^{\alpha k + \frac{\alpha}{2} - t} (y \sin \theta)^t dx dy d\theta \\ &= \sum_{k=0}^{\infty} W_1^*(n, k) \sum_{t=0}^{\infty} \binom{\alpha k + \frac{\alpha}{2}}{t} \int_x \int_y f(x, y) x^{\alpha k + \frac{\alpha}{2} - t} y^t dx dy \int_\theta A_m^*(\theta) (\cos \theta)^{\alpha k + \frac{\alpha}{2} - t} (\sin \theta)^t d\theta \\ &= \sum_{k=0}^{\infty} W_1^*(n, k) \sum_{t=0}^{\infty} \binom{\alpha k + \frac{\alpha}{2}}{t} \Theta_{\alpha k + \frac{\alpha}{2} - t, t} G_{\alpha k + \frac{\alpha}{2} - t, t}, \end{aligned} \quad (3.27)$$

其中，第五个等号源自牛顿广义二项式定理。 $\square$

定理 3.4：多项式 FMR 的显式定义。由定义 3.2 和 3.4 给出的多项式 FMR 可被重写为如下显式形式^[113]：

$$\langle \mathbf{R}_f, V_{nm}^\alpha \rangle = \sum_{k=0}^n W_2(\alpha, p, q, n, k) \sum_{s=0}^{\infty} W_3(p, q, s) \sum_{t=0}^{\infty} \binom{\alpha(s+k+\frac{q}{2})}{t} \Theta_{\alpha(s+k+\frac{q}{2})-t, t} G_{\alpha(s+k+\frac{q}{2})-t, t}, \quad (3.28)$$

其中，系数项 W_2 和 W_3 分别定义为：

$$W_2(\alpha, p, q, n, k) = \sqrt{\frac{\alpha(p+2n)\Gamma(q+n)n!}{2\pi\Gamma(p+n)\Gamma(p-q+n+1)}} \frac{(-1)^k \Gamma(p+n+k)}{k!(n-k)!\Gamma(q+k)}, \quad (3.29)$$

$$W_3(p, q, s) = (-1)^s \binom{\frac{p-q}{2}}{s}, \quad (3.30)$$

其中符号 Θ 和 G 如前定义。

证明：我们首先将 (3.8) 中的多项式径向基函数重写为如下等价形式：

$$R_n^\alpha(r) = \sum_{k=0}^n W_2(\alpha, p, q, n, k) \sum_{s=0}^{\infty} W_3(p, q, s) r^{\alpha(s+k+\frac{q}{2})-1}, \quad (3.31)$$

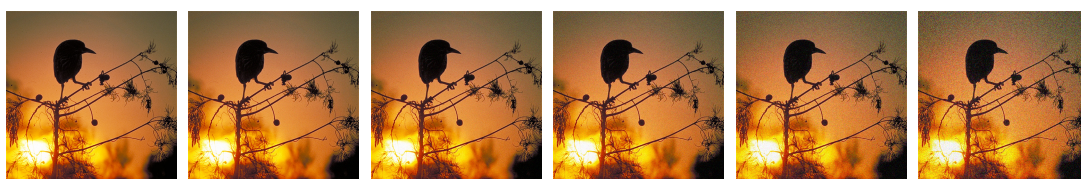
其中，与 r 有关的项均被归并，基于牛顿广义二项式定理。

接下来，多项式 FMR 隐式定义中的 $\mathbf{R}_f$ 和 R_n^α 将分别展开，得出以下显式形式：

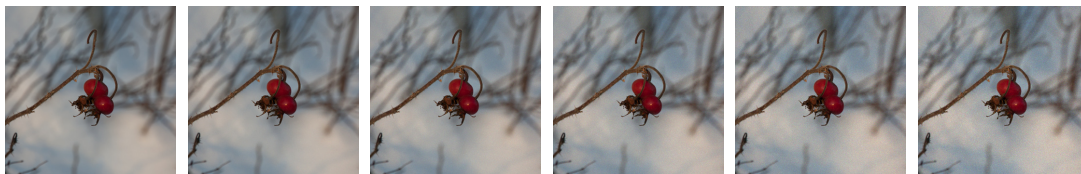
$$\begin{aligned} \langle \mathcal{R}_f, V_{nm}^\alpha \rangle &= \int_r \int_\theta [R_n^\alpha(r) A_m(\theta)]^* \mathcal{R}_f(r, \theta) r dr d\theta \\ &= \int_r \int_\theta [R_n^\alpha(r) A_m(\theta)]^* \int_x \int_y f(x, y) \delta(r - x \cos \theta - y \sin \theta) dx dy r dr d\theta, \\ &= \int_r \int_\theta \int_x \int_y f(x, y) A_m^*(\theta) \delta(r - x \cos \theta - y \sin \theta) \sum_{k=0}^n W_2(\alpha, p, q, n, k) \\ &\quad \times \sum_{s=0}^{\infty} W_3(p, q, s) r^{\alpha(s+k+\frac{q}{2})} dx dy r dr d\theta \\ &= \sum_{k=0}^n W_2(\alpha, p, q, n, k) \sum_{s=0}^{\infty} W_3(p, q, s) \int_\theta \int_x \int_y f(x, y) A_m^*(\theta) (x \cos \theta + y \sin \theta)^{\alpha(s+k+\frac{q}{2})} dx dy d\theta \\ &= \sum_{k=0}^n W_2(\alpha, p, q, n, k) \sum_{s=0}^{\infty} W_3(p, q, s) \sum_{t=0}^{\infty} \binom{\alpha(s+k+\frac{q}{2})}{t} \\ &\quad \times \int_\theta \int_x \int_y f(x, y) A_m^*(\theta) (x \cos \theta)^{\alpha(s+k+\frac{q}{2})-t} (y \sin \theta)^t dx dy d\theta \\ &= \sum_{k=0}^n W_2(\alpha, p, q, n, k) \sum_{s=0}^{\infty} W_3(p, q, s) \sum_{t=0}^{\infty} \binom{\alpha(s+k+\frac{q}{2})}{t} \int_x \int_y f(x, y) x^{\alpha(s+k+\frac{q}{2})-t} y^t dx dy \\ &\quad \times \int_\theta A_m^*(\theta) (\cos \theta)^{\alpha(s+k+\frac{q}{2})-t} (\sin \theta)^t d\theta \\ &= \sum_{k=0}^n W_2(\alpha, p, q, n, k) \sum_{s=0}^{\infty} W_3(p, q, s) \sum_{t=0}^{\infty} \binom{\alpha(s+k+\frac{q}{2})}{t} \Theta_{\alpha(s+k+\frac{q}{2})-t, t} G_{\alpha(s+k+\frac{q}{2})-t, t}, \end{aligned} \quad (3.32)$$

类比定理 3.3 的证明。 $\square$

显式定义的总结：本节给出的显式定义提供了一种理解 FMR 的深入角度。其中一个有价值的观察是 FMR 可以被解释为几何矩的无穷线性组合，从而与矩和矩不变量的基本概念建立了联系。值得注意的是，在相关工作^[113]中拉东域的整数阶矩均为几何矩的有限线性组合，这

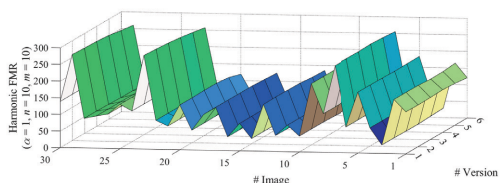


(a) 含高斯噪声的 CSIQ 图像集

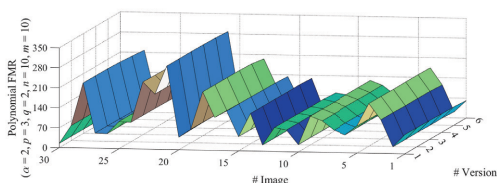


(b) 含泊松噪声的 CIDIQ 图像集

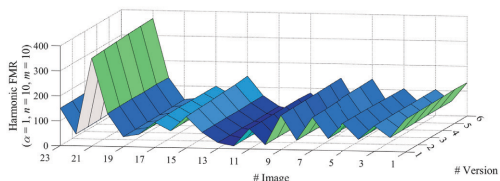
图 3.6 含高斯噪声的 CSIQ 图像集和含泊松噪声的 CIDIQ 图像集：从右到左噪声强度增加。



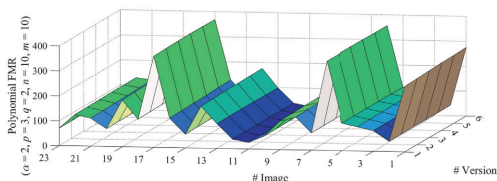
(a) 高斯噪声 CSIQ 图像集上的谐波 FMR 幅度



(b) 高斯噪声 CSIQ 图像集上的多项式 FMR 幅度



(c) 泊松噪声 CIDIQ 图像集上的谐波 FMR 幅度



(d) 泊松噪声 CIDIQ 图像集上的多项式 FMR 幅度

图 3.7 CSIQ 和 CIDIQ 图像集上谐波和多项式 FMR 的幅度直方图：从右到左对应不同图像，从前到后对应噪声强度增加。

种区别也从另一角度反映了 FMR 的广义性。

3.5 实验和应用

本节将全面评估所提出的 FMR 的性能，涵盖仿真实验和实际应用层面的对比。

在仿真实验层面，我们首先给出了 FMR 的直方图和重构图像，作为噪声稳健性和判别性的直观可视化结果。在量化结果方面，我们还在含有严重噪声的图像上进行了具有挑战性的模式识别实验，并与先进学习表征方法（及其数据增广训练和去噪预处理策略）全面比较。

在实际应用层面，我们考虑 FMR 在两类稳健视觉任务中的具体应用。这里，首先应用于自然畸变下的模板匹配任务，涵盖如医学图像分析的可信场景。然后，还应用于图像版权保护的零水印算法，其反映了信息安全场景的内在对抗性。注意，图像表征的噪声稳健性和几何不变

性对于实现上述应用目标至关重要。

3.5.1 特征直方图

在该实验中，我们通过可视化不同噪声条件下的特征直方图来验证 FMR 的噪声稳健性。同时，实验还考虑了不同图像之间的判别性，从而直观展示这种稳健特征的非平凡性（即不是常值函数）。

如图 3.6 所示，实验在知名的劣化图像集 CSIQ^[260] 和 CIDIQ^[261] 上进行，分别涉及高斯噪声和泊松噪声。这两个图像数据集分别包含 30 和 23 幅原始图像，每个原始图像都有 5 个不同强度的噪声版本，共计 318 幅图像。

如图 3.7 所示，我们给出了谐波/多项式 FMR 在清晰图像和含噪图像上的幅度直方图。其中，谐波 FMR 的参数为 $\alpha = 1$ ， $n = 10$ 和 $m = 10$ ，多项式 FMR 的参数为 $\alpha = 2$ ， $p = 3$ ， $q = 2$ ， $n = 10$ 和 $m = 10$ 。从直方图中可以看出，对于每个图像的含噪序列，FMR 特征几乎保持恒定，但对于不同的图像，FMR 特征则表现出显著差异。上述现象表明，FMR 对高斯噪声和泊松噪声实现了良好的稳健性，而且 FMR 对于不同图像具有足够的判别性，不是平凡不变量。

3.5.2 图像重构

在该实验中，我们从含噪图像的 FMR 表征重构图像以可视化表征对应于何种图像信息，这里我们可以直观地了解表征对图像内容的描述程度以及噪声对特征的干扰程度。

基于 V_{nm}^α 的正交性，可从 FMR 系数 $\{M_{nm}^\alpha = \langle \mathbf{R}_f, V_{nm}^\alpha \rangle\}$ 重构出图像的拉东投影：

$$\widehat{\mathbf{R}}_f = \sum_{(n,m) \in S_{nm}(K)} M_{nm}^\alpha V_{nm}^\alpha \quad (3.33)$$

其中， $\widehat{\mathbf{R}}_f$ 是原始拉东投影 $\mathbf{R}_f$ 的重构版本，而 $S_{nm}(K) = \{(n, m) : |n|, |m| \leq K\}$ 是由常数 K 限制的阶数 (n, m) 的采样集合。得到 $\widehat{\mathbf{R}}_f$ 后，可通过逆拉东变换进一步得到原始图像的重构版本 $\hat{f}$ 。这里，逆拉东变换是一项成熟研究并且已经内置在典型的编程语言中——技术上一般是通过傅里叶变换实现^[262]。

实验在数学家傅里叶和雅可比的头像上进行，并添加方差为 0.2 的高强度高斯噪声。注意，为了得到合理的对比基线，我们引入了拉东域 FMR 直接对应的图像域 FM： $\langle f, V_{nm}^\alpha \rangle$ ^[80,159]。定量结果方面，我们计算了含噪图像的重构版本与原始版本之间的均方重构误差^[37]或结构相似度^[228]。

在图 3.8 中，我们展示了通过 FM 和 FMR 的重构结果，其中谐波 FM 和 FMR 的参数为 $\alpha = 1$ ， $K = 30$ ；多项式 FM 和 FMR 的参数为 $\alpha = 2$ ， $p = 3$ ， $q = 2$ ， $K = 50$ 。注意，FMR 和 FM 在这里的设置非常一致，唯一的区别在于是否涉及拉东投影。从视觉上看，所提出的 FMR

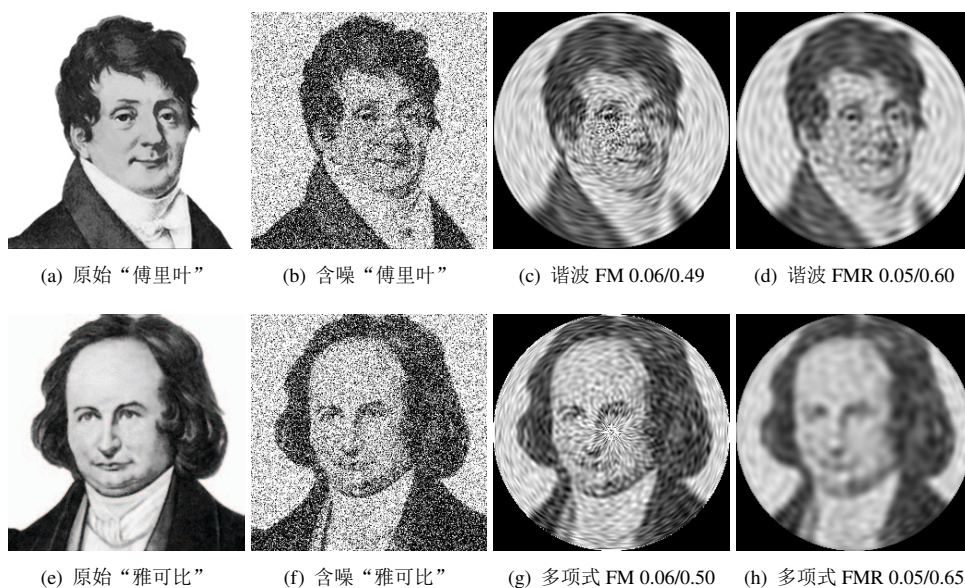


图 3.8 含噪图像重构示例：数值是重构版本相对于原始版本的均方重构误差和结构相似性。

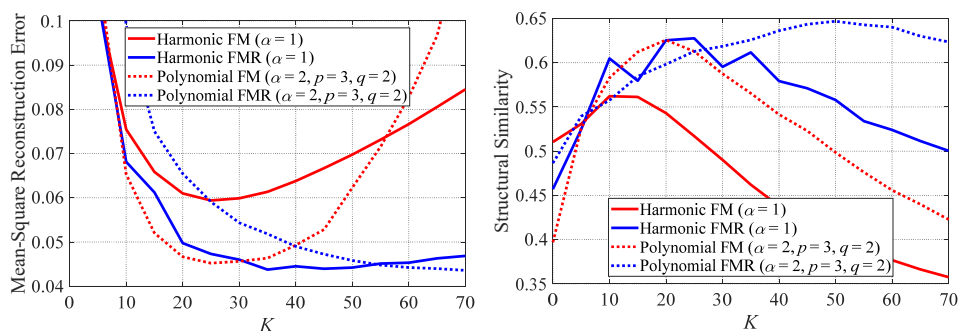


图 3.9 不同方法下含噪图像的重构版本与原始版本的均方重构误差和结构相似性。

与 FM 一样能很好地重构干净图像的主要内容，同时比 FM 重构的噪声更少。相应的数值结果也支持这一观察，FMR 相比 FM 表现出较低的误差和较高的相似度。以上实验结果验证了研究动机：拉东投影确实增强了噪声稳健性，即完全符合性质 3.3 的预测。

在图 3.9 中，我们提供了更全面的定量结果。误差和相似度曲线显示，FM（红线）的视觉效果在超过特定 K 值后开始大幅下降，由于更多噪声被重构，而我们的 FMR（蓝线）明显减缓了这一趋势，特别是在较高的 K 值下表现出显著的性能优势。这些现象表明：1）高阶的图像域 FM 对噪声敏感，这阻碍了高阶矩被用于提升表征判别性；2）高阶的拉东域 FMR 则表现出更好的噪声稳健性，可视为性质 3.3 的进一步定量证据。

3.5.3 模式识别

在该实验中，我们对任意朝向的含噪图像进行模式识别实验，以直接验证了 FMR 的旋转不变性、噪声稳健性和判别性水平。其中，拉东域 FMR 首先与其直接对应的图像域 FM 进行了全面对比，涵盖多个对应的参数设置。然后，特别与一些学习表征展开对比，涵盖基于数据增广训练和去噪预处理的稳健性增强策略。基于上述一系列对比，可定位出 FMR 相对于上述表征方法在小规模稳健视觉问题中的性能水平。

类似于 2.4.3 节，实验图像来自 COREL 图像集^[229]，选择了 100 幅图像并将它们的尺寸归一化为 128×128 。对于非学习的 FM 和 FMR，分类器将只在这些原始图像的表征上进行训练。测试将在原始图像的劣化版本上进行，劣化操作包括方差为 $\{0, 0.05, \dots, 0.3\}$ 的高斯噪声和角度为 $\{0^\circ, 10^\circ, \dots, 350^\circ\}$ 的旋转。因此，测试集总计包含 25200 幅图像。对于学习表征，每个类别仅有一个训练图像一般来说不足以形成良好表示，因此我们还在训练集的每个类别中添加了 36 个旋转样本和 36 个含噪样本，作为数据增广训练策略。显然，上述增广会使得学习表征的训练规模相对于非学习的 FM 和 FMR 增加 36 倍（旋转增广）或 72 倍（旋转和噪声联合增广）。此外，还为学习表征引入了测试图像的去噪预处理，去噪过程由深度神经网络 DnCNN 实现^[112]。注意，这里采用朴素的最小距离分类器，以反映表征本身的性能水平。

对比方法包括：

- 经典方法，ART^[263]、GFD^[39] 和 ZM^[46]；
- FM 方法，FJFM(p, q, α)^[159] 和 GPCET(α)^[80]，涵盖整数阶特例 OFMM^[50]、CHFM^[51]、JFM^[53]、EFM^[56] 和 PCET^[57]；
- FMR 方法，上述所有 FM 表征的拉东域扩展版本；
- 轻量学习方法，PCANet^[230] 和 CBF^[231]，以及旋转/噪声增广或去噪预处理^[112]；
- 深度学习方法，GoogLeNet^[232] 和 ResNet-50^[233]，以及旋转/噪声增广或去噪预处理^[112]。

对于上述基于矩的方法，我们将集合 $S_{nm}(K) = \{(n, m) : |n|, |m| \leq K\}$ 上得到的所有矩的幅度值（旋转不变量）作为分类特征。由于 FMR 的高阶系数比 FM 的高阶系数噪声稳健性更强（3.5.2 节），我们在 FMR 特征上引入因子 $(n+1)$ 以平衡特征的值域范围并增强判别性。注意，该策略对于 FM 可能会有害。轻量学习表征遵循其原始论文中的参数设置。深度学习表征则基于 ImageNet 预训练和迁移学习重训练。

在表 3.2 中，我们列出了基于矩的表征方法在 $K = \{10, 20\}$ 时的正确分类百分比。其中，蓝色/红色底纹表示排名较高/较低的分数，而最佳分数则用加粗表示。可以观察到：

- 经典不变量的分数通常较差，主要原因是非正交/非完备的基函数所导致的低判别性。
- FM 不变量在不同的参数设置下的分数表现相对一致，尤其是在严重噪声情况下明显优于经典矩不变量。

表 3.2 矩不变量经典方法、FM 方法和 FMR 方法的分类分数 (%)。

矩不变量	$K = 10$, 高斯白噪声 (方差)						
	0	0.05	0.1	0.15	0.2	0.25	0.3
ART	100	99.17	88	74.53	50.14	30.14	19.19
GFD	100	99.28	89.86	74.75	50.14	28.67	16.86
ZM	100	88.83	57.19	35.17	25.92	18.64	12.86
OFMM/FJFM(2,2,1)	100	100	95.28	82.94	60.89	36.03	22.36
CHFM/FJFM(2,1.5,1)	100	100	97.78	86.78	70.97	51.19	36.22
JFM(3.3)/FJFM(3,3,1)	100	100	97.08	85.92	71.92	47.31	30.61
FJFM(3,3,2)	100	100	98.56	86.11	75.92	51.86	36.75
EFM/GPCET(1)	100	100	97.28	90.47	68.25	46.39	28.83
PCET/GPCET(2)	100	100	95.36	83.33	60.17	35.5	22.06
GPCET(1.5)	100	100	96.92	84.78	67	43.72	27.72
OFMMR/FJFMR(2,2,1)	100	100	98.89	91.69	77.56	52.89	32.97
CHFMR/FJFMR(2,1.5,1)	100	100	98.11	92.03	77.75	53.31	32.75
JFMR(3.3)/FJFMR(3,3,1)	100	100	99	93.39	78.89	54.42	33.39
FJFMR(3,3,2)	100	100	98.61	86.19	71.67	47.47	33
EFMR/GPCETR(1)	100	100	99	92.56	82.36	56.47	36.47
PCETR/GPCETR(2)	100	100	98.94	89.36	73.53	48.08	31.03
GPCETR(1.5)	100	100	98.83	92.61	78.44	50.42	31.78
矩不变量	$K = 20$, 高斯白噪声 (方差)						
	0	0.05	0.1	0.15	0.2	0.25	0.3
ART	100	99.39	93.28	80.14	59.22	37.42	23.67
GFD	100	99.28	93.69	80.03	60.67	38.67	23.53
ZM	100	90.36	61.97	37.83	26.67	19.64	13.53
OFMM/FJFM(2,2,1)	100	100	97.06	88.44	71.33	48.22	30.31
CHFM/FJFM(2,1.5,1)	100	100	98	91.81	80.14	59.14	45.31
JFM(3.3)/FJFM(3,3,1)	100	100	97.61	90.78	78.39	57.83	36.86
FJFM(3,3,2)	100	100	98.64	91.19	82.81	63.89	44.83
EFM/GPCET(1)	100	100	97.08	91.28	78.06	56.42	40.56
PCET/GPCET(2)	100	100	96	87.61	71.06	45.92	29.69
GPCET(1.5)	100	100	97	88.94	75.89	52.64	34.17
OFMMR/FJFMR(2,2,1)	100	100	100	97.89	95.53	84.89	57.47
CHFMR/FJFMR(2,1.5,1)	100	100	100	98.06	95.72	85.53	57.19
JFMR(3.3)/FJFMR(3,3,1)	100	100	100	98.33	96.42	84.5	57.42
FJFMR(3,3,2)	100	100	99.67	98	90	78.17	52.28
EFMR/GPCETR(1)	100	100	99.06	97.14	92.72	79.78	58.5
PCETR/GPCETR(2)	100	100	99.06	95.08	88.94	72.14	52.44
GPCETR(1.5)	100	100	98.92	95.42	90.78	75.64	52

表 3.3 学习方法和矩不变量方法的分类分数 (%)。

表征方法	训练规模	预处理	高斯白噪声 (方差)						
			0	0.05	0.1	0.15	0.2	0.25	0.3
PCANet (直接训练)	×1		19.64	12.31	9.06	6.72	5.31	4.56	4.44
PCANet (旋转增广)	×36		100	51.94	23.22	11.67	7.64	5.03	4.69
PCANet (旋转和噪声增广)	×48		100	38.25	13.72	6.25	3.72	2.58	1.97
PCANet (旋转增广和去噪)	×36	✓	97	68.53	39.08	24.19	16.19	12.75	9.14
CBFD (直接训练)	×1		24.94	4.27	1.92	1.81	1.28	1.17	0.92
CBFD (旋转增广)	×36		100	4.53	2.56	1.97	1.53	1.47	1.78
CBFD (旋转和噪声增广)	×48		100	11.97	4.61	2	1.56	1.53	1.42
CBFD (旋转增广和去噪)	×36	✓	90.5	15.75	9.08	5.58	4.17	3.33	2.67
GoogLeNet (旋转增广)	×36		99.97	6.08	2.22	1.89	1.42	1.28	1.33
GoogLeNet (旋转和噪声增广)	×72		99.97	69.19	52.47	39.22	28.25	19.58	13.14
GoogLeNet (旋转增广和去噪)	×36	✓	100	44.69	19.81	8.42	4.28	3.28	2.75
ResNet-50 (旋转增广)	×36		100	3.44	1.28	1.03	1.03	1.19	1.25
ResNet-50 (旋转和噪声增广)	×72		100	78.06	66.33	59.58	54.22	49.19	44.47
ResNet-50 (旋转增广和去噪)	×36	✓	100	51.81	23.78	11.47	6.44	4.86	4
FM (平均结果)	×1		100	100	97.12	87.88	72.34	49.72	33.31
FM (最佳结果)	×1		100	100	98.64	91.81	82.81	63.89	45.31
FMR (平均结果)	×1		100	100	99.15	94.13	85.02	65.98	44.19
FMR (最佳结果)	×1		100	100	100	98.33	96.42	85.53	58.5

- FMR 不变量在几乎所有参数设置、所有 K 值和所有噪声方差下均实现了相对于 FM 的分数提升，表现出超越经典和 FM 不变量的最佳性能。

上述结果为 FMR 的旋转不变性、噪声稳健性和判别性水平提供了数值证据，相较于这些最相关工作具有独特性和实用性。

在表 3.3 中，我们还列出了与学习表征的分类分数对比，这里 FM 和 FMR 的分数以表 3.2 中的平均/最佳值为代表。可以观察到：

- 使用矩不变量相同的训练图像集时，轻量学习表征即使在干净图像上也近乎无效，主要由于 PCANet 和 CBFD 不具备内在的旋转不变性。
- 在引入旋转增广训练后，所有学习表征都在干净图像上实现了 ~100% 的分数，但训练规模也扩大了 36 倍。另一方面，这些方法的分数在噪声方差增加时急剧下降，说明缺乏噪声稳健性。
- 当同时引入旋转和噪声增广训练后，深度学习表征的稳健性显著提高，而轻量级学习表征则没有这种现象，可能是由于模型容量的差异所致。注意，这以训练规模扩大 72 倍为代价，这显然是会导致效率显著下降。
- 去噪预处理带来的分数提升未达到一般预期。显然，图像恢复和识别研究之间仍然存在

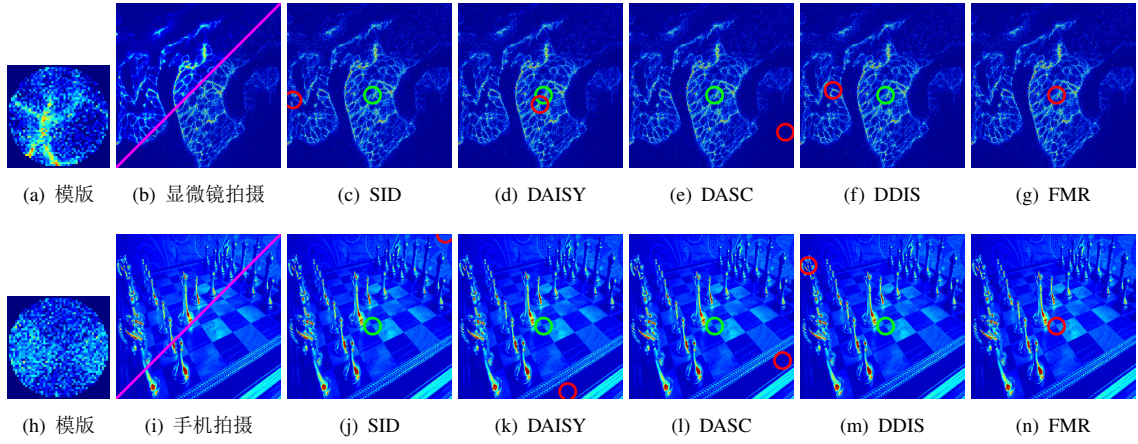


图 3.10 自然畸变下的模板匹配示例：(a) 和 (h) 是在噪声环境下的任意朝向模板；(b) 和 (i) 是通过荧光显微镜和智能手机拍摄的含有真实成像噪声的图像，左上与右下对应干净和含噪可视化；(c)~(g) 和 (j)~(n) 是通过不同表征得到的结果，红色和绿色对应匹配模版和真实目标可视化。

鸿沟，新引入的去噪误差强烈干扰了学习表征以及随后的识别。我们注意到类似现象也出现在一些最近的研究中^[142,244]。

- 相比之下，FM 和 FMR 表征在旋转不变性、噪声稳健性和计算效率方面表现显著优于上述学习方法，即使在平均分数上也具有优势。进一步的，FMR 表征在总体的性能表现上超越了 FM 表征以及轻量和深度学习表征方法。

上述结果表明所提出的 FMR 在小规模稳健视觉问题中具有竞争力，即使在现今的深度学习时代。

3.5.4 模板匹配

模板匹配即在图像中寻找给定的模板^[264]，是一类重要的底层视觉任务并且在许多图像处理问题以及可信场景（如医学图像分析）中扮演基础性的角色。

该任务的主要难题是模板和目标在信号和几何层面的自然畸变，这种不一致性常会导致模板匹配算法出现稳健性问题。接下来，我们考虑通过直接应用 FMR 来改进自然畸变下模板匹配的稳健性，并与该领域的先进方案进行对比。

如图 3.10 所示，实验分别在荧光显微镜^[265] 和智能手机^[266] 拍摄的真实图像上进行。注意，这些图像由于光照环境和成像硬件的固有限制，常会表现出严重的成像噪音。此外，这些实验图像还表现出强烈的非局部自相似性，例如相似的细胞，这也带来了判别性方面的挑战。从噪声环境下的任意定向模板匹配的角度考虑，实验将基于以下四类模板变体：

- 理想模版，良好成像且朝向对齐的模版；
- 旋转模版，良好成像但未经朝向对齐的模版；
- 含噪模版，成像不佳但经过朝向对齐的模版；

表 3.4 真实荧光显微镜图像上不同表征的模板匹配分数 (%)。

表征方法	理想模版	旋转模版	含噪模版	旋转并含噪模版
SID	50	38	31	20
DAISY	73	11	50	11
DASC	80	7	22	2
DDIS	68	67	65	65
FMR	99	99	96	93

表 3.5 真实手机图像上不同表征的模板匹配分数 (%)。

表征方法	理想模版	旋转模版	含噪模版	旋转并含噪模版
SID	81	64	27	20
DAISY	99	13	7	0
DASC	99	10	5	1
DDIS	71	71	71	70
FMR	100	99	89	83

- 旋转并含噪模版，成像不佳且未经朝向对齐的模版；

理想模版是在每张干净图像上随机选择 20 个区域，其中有 30 对干净和含噪图像，因此共计形成了 600 个理想模版。进一步的，通过对上述理想模版施加任意角度旋转和/或方差为 0.01 的高斯噪声，形成了共计 2400 个模版。实验算法将比较模版的全局表征和待分析图像 (512×512 像素) 的局部表征，其最小表征距离即对应模版匹配结果。因此，2400 个模版中，每个模版都对应着 $512 \times 512 = 262144$ 个潜在目标。

对比方法包括：

- 几何不变表征 SID^[267]，具有和 FM 类似的旋转不变性设计；
- 先进的稠密纹理表征 DAISY^[18] 和 DASC^[268]，具有几何或信号层面的稳健性设计；
- 先进的深度学习表征 DDIS^[269]，具有专用于模版匹配的稳健性设计。

在表 3.4 和 3.5 中，我们分别列出了上述表征在显微镜图像和手机图像上的模版匹配准确性得分。

可以观察到，这些方法在四个实验协议上表现出显著的性能差异，侧面反映出自然畸变模版匹配的困难性。由于对图像噪声的敏感性，SID 的匹配准确性偏低，即使是在仅包含旋转的实验协议上。注意，这一现象也说明了在不变表征研究中引入类似 FMR 设计的必要性。相比之下，DAISY 和 DASC 在理想协议中表现出一致良好的性能，主要因为具有更高的判别性。然而，二者均缺乏内在的几何不变性，导致在含有旋转的实验协议上均失效。专门为模版匹配设计的深度表征 DDIS 在四个协议上达到了相当一致的分数的表现。然而，其稳健性可能以判别性为代

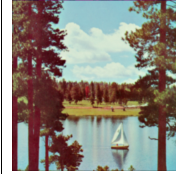
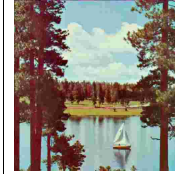
					
Original / PSNR	Rotation / 8.3	Gauss. Noise / 9.7	S. & P. Noise / 11.9	Blurring / 26.6	JPEG / 24.4
					
FM: BER % ↓	1.3	11.7	9.1	2.4	2.6
					
FMR: BER % ↓	0.9	2.9	3.1	0.6	0.5

图 3.11 零水印方法示例：不同劣化图像及其峰值信噪比，相应的版权编码及其误码率。

价，分数的整体水平较低，反映了在深度表征中嵌入不变结构的困难性。所提出的 **FMR** 在两个数据集和四个协议上均表现出显著更好且更稳定的分数，这一现象表明 **FMR** 在真实的噪声图像上实现了理论预期的旋转不变性和噪声稳健性，并且对于模板匹配任务具有足够的判别性。

3.5.5 零水印

与传统视觉任务不同，视觉信息安全任务具有很强的对抗性：研究者提出的安全技术与敌手的攻击方法此消彼长^[238]。在这种对抗性任务中，图像表征的对成像噪声、传输噪声和编辑噪声的稳健性是算法有效性和安全性的重要基础。

接下来考虑将 **FMR** 应用于图像版权保护的零水印技术。零水印技术是传统水印技术的一种替代方案，其不将版权信息嵌入到图像中，而是由图像内容和版权信息生成二进制序列（即零水印）并记录在可信数据库（主要是区块链）^[259]。技术上，这种零水印是原始图像的感知哈希编码（即内容导出的二进制序列）和版权编码（如作者身份的二进制序列）的异或。在验证过程中，如果测试图像是原始图像的内容保持副本，则从测试图像提取的感知哈希编码与零水印作异或将恢复出版权编码。

该任务的主要难题是确保原始图像的内容保持版本（例如含噪版本）能够正确地通过零水印的验证。我们尝试以 **FMR** 作为零水印算法的表征环节，以实现旋转不变和噪声稳健的版权保护。公平起见，算法遵循相同的框架，只在图像表征上有所不同，从而更好地反映出表征在零水印任务中的性能水平；同时，特征向量的维度与版权编码相同，表征参数（如 n, m, p, q, α ）随机生成。

在图 3.11 中，我们展示了零水印算法在不同表征和攻击下的稳健性，这里以谐波 **FMR** 和 **FM** 为例。其中，原始图像劣化形成内容保持的变体，然后该变体通过零水印算法进行验证并提取版权编码。引入峰值信噪比（Peak Signal-to-Noise Ratio, PSNR）和误码率（Bit Error Ratio,

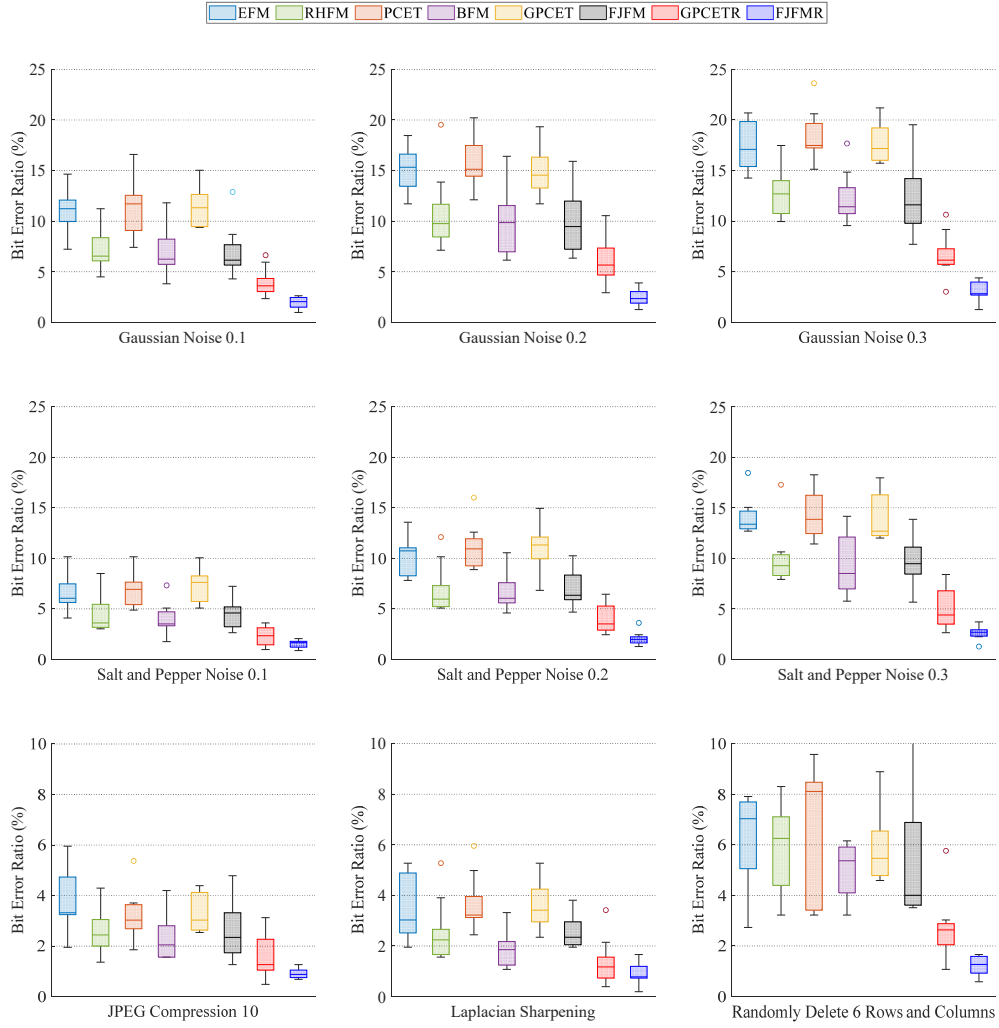


图 3.12 不同零水印方法在稳健性实验中的误码率箱线图。

BER), 分别用于衡量图像变体的保真度和提取到的版权编码的错误率。

可以观察到, 由 **FMR** 提取的版权编码的误码率一直较低, 尤其是对于峰值信噪比低至 10 附近的严重噪声。这种相对于 **FM** 的优势与理论预期 (性质 3.3) 一致, 并且与模式识别实验 (3.5.3 节) 也一致。注意, 尽管在连续域假设下 **FM** 和 **FMR** 都具有旋转不变性, 但离散域旋转中的重采样/重量化误差可能会影响这种不变性。

在图 3.12 中, 我们给出了对应的稳健性分数箱线图, 涵盖更全面的表征和攻击并揭示出误码率分数的分布特性。实验囊括了零水印领域的几个主要竞争算法, 包括基于 **EFM**^[270]、**RHFM**^[271]、**PCET**^[272]、**BFM**^[273]、**GPCET**^[259] 和 **FJFM**^[159] 表征的零水印算法。在这里, 谐波 **FMR** 和多项式 **FMR** 的算法也被命名为 **GPCETR** 和 **FJFMR**, 以与上述符号和第 3.5.3 节保持一致。

可以观察到, 与零水印领域的现有表征相比, 所提出的 **FMR** (尤其是 **FJFMR**) 实现了误码率的平均值和方差均较低, 且该现象在不同类型和强度的攻击下始终一致。因此, 这些结果可

视为 FMR 在零水印任务中独特性和实用性的数值证据。

3.6 本章小结

本章对严重噪声干扰下的全局不变量进行了系统性研究，提出了名为“分数阶拉东矩 FMR”的全局不变量新方法，具有广义性、噪声稳健性、旋转不变性和时频判别性的特点。

本章工作在理论层面的要点如下：

- 定义方面，详细讨论了如何分别通过隐式路径（3.3.1 节）和显式路径（3.4 节）来构造 FMR。其中，隐式定义更直观，有助于推导表征性质和设计高效计算。显式定义则提供了深入理解 FMR 的角度，其可以被解释为几何矩的无穷线性组合，为未来工作提供了数学工具。
- 性质方面，全面分析了 FMR 的广义性、旋转不变性、噪声稳健性和时频判别性（3.3.2 节）。这些有益性质使得 FMR 区别于一些理论相关的方法，揭示了 FMR 在视觉任务中的独特性和实用性（表格 3.1）。
- 计算方面，通过傅里叶变换和递归公式（3.3.3 节）设计了计算效率高且数值稳定的离散实现方案。

本章工作在教育层面的要点如下：

- 仿真实验方面，FMR 的噪声稳健性和判别性首先通过特征直方图和图像重构进行可视化（3.5.1 和 3.5.2 节）。接下来在严重噪声图像上进行模式识别实验，全面评估了拉东域 FMR、其直接对应的图像域 FM、以及一些学习表征方法，特别涵盖目前流行的数据增广训练和去噪预处理策略（3.5.3 节）。实验结果充分验证了本研究在旋转不变性、噪声稳健性和时频判别性方面的正确性。
- 实际应用方面，FMR 直接应用于自然畸变下的模板匹配（3.5.4 节）和面向版权保护的零水印技术（3.5.5 节）。FMR 在这两个任务中整体性地提升了对各类信号和几何劣化的稳健性，在此方面直接优于两个任务中的常用表征方法。总体而言，FMR 分数表现证明了其在小规模稳健视觉问题中的应用潜力。

本章工作已总结成文：

- [Shuren Qi](#), [Yushu Zhang](#)[✉], [Chao Wang](#), et al. Representing noisy image without denoising. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 2024, Early Access.

第四章 设计局部不变量

针对不变表征方法在局部视觉任务中的适用性问题，本学位论文的第二个课题是设计局部不变量，将全局不变量的基本理论推广到更广义的局部假设。本章将从一类局部视觉任务（图像取证）的实际需求出发，在理论（定义、性质和计算）、实践（设置原则）和应用（经典和可信视觉任务）层面介绍所提出的局部不变量新方法。

4.1 引言

虚假视觉媒体正在成为社会中的一种现实威胁，在新闻报道、学术出版、司法调查和社交网络等重要场景中存在滥用风险。为此，验证视觉数据真实性和完整性的取证研究已引起了学术界的广泛关注^[274]。

图像取证分析作为一类视觉任务，其有效性依赖于同样依赖于合适的图像表征^[275]：图像中的操纵痕迹被合理地表示为特征，进而实现有效的真伪判定。因此，在取证研究中一种常见做法是直接继承经典视觉任务的表征技术。另一方面，图像取证同时作为一类可信任务，稳健性和可解释性是至关重要的。具体而言，取证任务和经典视觉任务中的表征研究尽管在概念上有相似之处，但在设计目标上有很大不同。

面向取证任务的图像表征应该满足以下基本要求：判别性，表征的图像信息足够丰富，足以支持区分真实数据和虚假数据；稳健性，表征对敌手可能引入的几何或信号劣化操作稳定；稠密性，表征能够捕捉图像平面上任意位置的信息，因为篡改内容可能出现在包括背景在内的任何图像区域；可解释性，表征应该有可靠的理论保证，因为取证结果应作为证据性的判断标准，其中因果关系远比相关关系更重要；计算精确性和复杂性，表征应该有计算上合理的离散实现，特别是满足稠密性要求的前提下。

4.1.1 研究动机

从以上要求出发，可知目前经典视觉任务（例如图像分类）中流行的表征方法往往不是取证任务的最佳选择，主要由于其在稳健性、可解释性和稠密性方面的不足。

深度学习表征，例如 CNN，在各类高层视觉任务中表现出先进的性能，主要归功于可学习非线性变换的级联能够提供强大的数据自适应能力^[276]。尽管深度学习表征在判别性方面具有优势，但两个固有特性使其不能很好适应取证任务。首先稳健性问题^[234]，特别是对于几何变化^[142]和对抗样本^[277]，这可能被敌手引入以误导取证算法，即反取证^[274]。其次是可解释性问题^[278]，黑盒决策机制会降低取证结果的可信度，特别是对于司法取证这类关键场景^[274]。目

表 4.1 本章方法与相关方法的概念性对比。

方法	启发式 局部不变量 ^[267]	笛卡尔矩个例扩展为 局部不变量 ^[91]	圆谐矩个例扩展为 局部不变量 ^[93]	本章方法
广义性				✓
旋转不变性	✓		✓	✓
常数阶复杂性		✓	✓	✓

前，深度学习表征的稳健性和可解释性研究正在发展^[279,280]，但远未达到取证任务的要求。

接下来分析手工表征。基于特征点的局部表征，例如 **SIFT**，是特征工程中流行且有竞争力的方法。其中，特征点检测子和描述子的主要设计目标是在各种几何变换下具有高度可重复性^[15]。然而，这些特征点在图像平面上是稀疏的，会导致一些区域的信息丢失。对于需要稠密表征的任务而言，例如细粒度分类^[281]和本章关注的取证任务^[274]，这种稀疏性是不利的。值得注意的是，目前的一些研究表明 **SIFT** 类的特征点可以通过微小的像素修改来进行特征点的删除或注入^[282]。显然这种性质可能被敌手用于反取证中，类似于深度学习表征的对抗性攻击。

出于上述原因，取证研究人员更青睐稠密的局部表征，例如 **DAISY**^[18]。这种基于规则采样网格形成的表征（含图像平面采样和多尺度采样）不仅保证了稠密性，还保证了特征之间简单的空间关系：这种规则的空间模式可广泛用于建模马尔可夫过程、构造空间池化特征、以及加速特征匹配等场景；对于特征点而言，其空间模式更为任意，在上述场景中需要更复杂的处理过程^[283]。稠密局部表征面临的主要挑战是几何不变性和计算效率：表征在几何变换下通常不稳定，即使是常见的旋转、缩放和翻转；此外，密集采样导致计算成本较高，这也阻碍了引入更耗时的不变性结构。

4.1.2 核心结论

本章围绕“取证任务中的图像局部表征”课题展开，尝试给出更系统性的研究结果。尽我们所知，这是该方面的首个理论性工作，此前很少将其视为一个相对独立的研究课题。我们将全局不变量的基本理论推广到更广义的局部假设，提出适用于取证任务的图像表征“原则性”设计。

本章工作是对矩和矩不变量研究在理论、实践和应用层面上的扩展。

理论。我们提出了新的局部不变表征（Local Invariant Representation, **LIR**）理论，涵盖定义扩展、性质分析和高效计算。首先，我们通过将圆谐正交矩的定义从全局扩展到具有尺度空间的局部，提出了 **LIR** 方法，并形式化讨论了其在几何不变性、等变性和协变性方面的核心性质。此外，我们推导出了高效的离散计算策略，具有低误差、常数阶复杂性以及基函数通用性的特点。因此，**LIR** 能够满足取证任务的核心要求，尤其是稳健性，可解释性和稠密性方面。

实践。我们给出了 **LIR** 的实践指南，旨在帮助研究者更好地将上述理论付诸于具体应用场

景。其中，我们特别是明确了 LIR 的一般实践流程以及在不同类似的取证任务中的设置原则。

应用。我们在各类仿真实验和实际应用中全面评估了 LIR 的性能。在两类经典视觉任务中，即稠密模式检测和匹配，LIR 相对于现有稠密局部表征的先进方法实现了更优的整体性能，特别是在挑战性的几何或信号劣化场景。在两类可信视觉任务中，即复制-粘贴篡改检测（被动取证）和感知哈希（主动取证），LIR 相对于该领域先进取证方法实现了取证分数和效率的提升。

4.2 预备和基础

本节介绍预备概念，即相关的稠密局部表征以及图像取证和篡改检测；并且介绍数学基础，即矩和矩不变量的全局定义和符号（详见 2.2 节）。

4.2.1 预备概念

稠密局部表征。以往工作中，稠密局部表征的设计通常扩展自频域变换、纹理描述子和图像矩方法。频域变换方面，研究者设计了基于离散沃尔什-哈达玛变换^[284]和哈尔小波^[285]的稠密局部表征，并用于稠密模式检测。这些研究侧重于计算复杂度和对噪声攻击的稳健性，而对几何不变性考虑较少^[264]。纹理描述子方面，稠密局部表征 DAISY^[18]和 DASC^[268]通过梯度分布和自相似性来建模图像的局部结构。这些方法被考虑为本章工作的间接竞争方法，由于具有类似的稳健性目标，侧重于稠密模式匹配（立体匹配）任务中的由于成像因素引起的几何（例如视角）和信号（例如模糊）劣化。图像矩方面，研究者设计了基于启发式^[267]、笛卡尔矩^[91]和圆谐矩^[93]的稠密局部表征，并用于稠密模式检测和图像取证任务中。这些方法被考虑为本章工作的直接竞争方法，由于具有类似的稳健性目标和实现路径。在表 4.1 中，我们总结了本章工作相对于这些直接竞争方法的概念性对比。本章后续内容将会证明，LIR 的表征性质和高效计算对于一类基函数而言是完全通用的：具有旋转不变性，其高效计算表现出相对于局部表征尺度的常数阶复杂性

图像取证和篡改检测。过去十几年，虚假图像内容主要通过类似于 Photoshop 的编辑软件产生，涉及复制-粘贴^[286]、拼接^[287]和修补^[288]等操作。近年来，随着深度学习的兴起，以 Deepfake^[289]为代表的人工智能生成技术使得人们以更加灵活的方式创造逼真的虚假图像内容。虚假内容检测方面，当前的方案主要分为主动取证和被动取证。主动取证需要在图像分发前，对图像进行某种信息嵌入（例如数字水印^[290]）或信息提取（例如感知哈希^[291]）。基于上述信息，主动取证方法理论上可以检测任何类型篡改，但显然需要额外的事前处理成本并且不能用于已分发的图像。与之相对应的被动取证则不依赖于图像分发前的处理，仅使用待分析图像本身的信息。当篡改发生时自然数据的一致性特征将被不可避免的破坏，因此被动取证旨在挖掘数字^[292]、物理^[293]或语义^[294]等层面的不一致性。这类方法的主要缺陷是泛化性和稳健性问题：被动取证通常只能检测预设的篡改类型，并且被动场景下学得的泛化特征往往十分脆弱。因此，有必要

设计一类面向取证任务的图像表征，以从基础上提高主动和被动方法的性能。

4.2.2 数学基础

数学上，图像矩的一般理论基于以下内积定义 $\langle f, V_{nm} \rangle$ ^[32]:

$$\langle f, V_{nm} \rangle = \iint_D V_{nm}^*(x, y) f(x, y) dx dy, \quad (4.1)$$

其中， f 为图像函数， V_{nm} 为阶数 $(n, m) \in \mathbb{Z}^2$ 的基函数，二者均定义于 $D \in \mathbb{R}^2$ ， $*$ 表示复共轭。

在图像表征中，通常对基函数的显式形式施加两个限制。

正交性是实现信号分析中诸多有益特性的核心，如信息保持和去相关性。当任意两个基函数 V_{nm} 和 $V_{n'm'}$ 满足以下条件时，则符合正交性：

$$\langle V_{nm}, V_{n'm'} \rangle = \iint_D V_{nm}(x, y) V_{n'm'}^*(x, y) dx dy = \delta_{nn'} \delta_{mm'}, \quad (4.2)$$

其中， δ 为克罗内克函数。一组正交函数的完备性意味着它的线性生成空间在 Hilbert 空间中稠密。

旋转不变性是几何分析中是一种理想的结构。这需要基函数 V_{nm} 在极坐标 (r, θ) 中表示，并且具有以下形式：

$$V_{nm}(\underbrace{r \cos \theta}_x, \underbrace{r \sin \theta}_y) \equiv V_{nm}(r, \theta) = R_n(r) A_m(\theta), \quad (4.3)$$

其中，角向基函数 $A_m(\theta) = \exp(jm\theta)$ ($j = \sqrt{-1}$)，径向基函数 $R_n(r)$ 可以是任意形式^[295]。我们在极坐标下施加上述两个约束条件，可以导出公式 (4.1) 的旋转不变且正交形式，即圆谐正交矩：

$$\langle f, V_{nm} \rangle = \iint_D R_n^*(r) A_m^*(\theta) f(r, \theta) r dr d\theta, \quad (4.4)$$

其中， $R_n(r)$ 满足加权正交条件： $\int_0^1 R_n(r) R_{n'}^*(r) r dr = \frac{1}{2\pi} \delta_{nn'}$ ，并且定义域为单位圆区域： $D = \{(r, \theta) : r \in [0, 1], \theta \in [0, 2\pi)\}$ 。

4.3 局部不变量理论

本节介绍 LIR 的理论方面。我们从数学定义开始，随后深入分析派生出的有益特性，最后设计了精确且快速的计算策略。

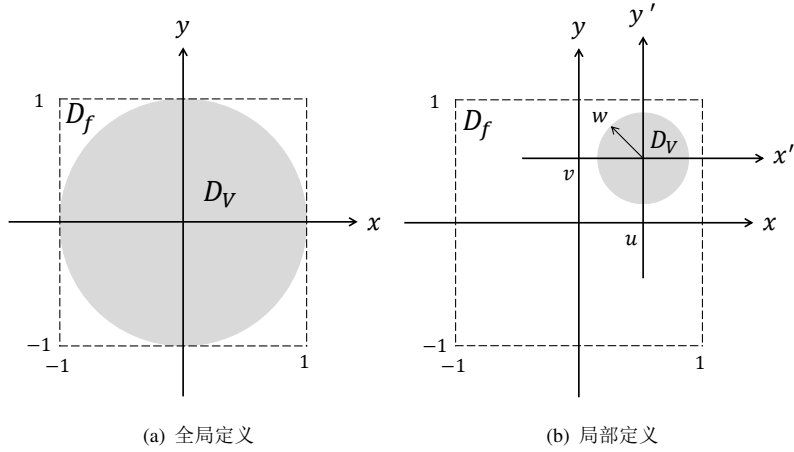


图 4.1 定义扩展的示意图。

4.3.1 定义扩展和基本公式

经典矩和矩不变量的一个重要属性是全局定义：图像函数 f 和基函数 V_{nm} 建立于统一的坐标系，更准确地说，二者有相同的原点和相似的尺度。

理论上，这种全局定义使得经典矩和矩不变量无法用于局部表征，并阻碍了其在诸多实际视觉问题中的应用。对于图像取证任务，特别是篡改区域定位，通常需要逐像素的局部表征以满足稠密性要求。显然，经典定义（4.1）和（4.4）无法实现。

接下来，我们旨在通过推广经典矩和矩不变量的定义来满足取证任务中的稠密性要求。考虑为基函数 V_{nm} 引入局部坐标系 (x', y') ，它是全局坐标系 (x, y) 的平移和缩放版本，具有平移向量 (u, v) 和缩放因子 w 。两个坐标系之间的显式关系如下：

$$(x', y') = \frac{(x, y) - (u, v)}{w}, \quad (4.5)$$

相应的极坐标可写为：

$$\begin{cases} r' = \sqrt{(x')^2 + (y')^2} = \frac{1}{w} \sqrt{(x-u)^2 + (y-v)^2} \\ \theta' = \arctan(\frac{y'}{x'}) = \arctan(\frac{y-v}{x-u}) \end{cases} \quad (4.6)$$

定义 4.1：局部定义。 基于局部极坐标（4.6），可以导出圆谐正交矩（4.4）的局部定义形式：

$$\langle f, V_{nm}^{uvw} \rangle = \iint_D \underbrace{R_n^* \left(\frac{\sqrt{(x-u)^2 + (y-v)^2}}{w} \right)}_{(V_{nm}^{uvw}(x,y))^*} \underbrace{A_m^* \left(\arctan \left(\frac{y-v}{x-u} \right) \right)}_{\theta'} f(x, y) dx dy, \quad (4.7)$$

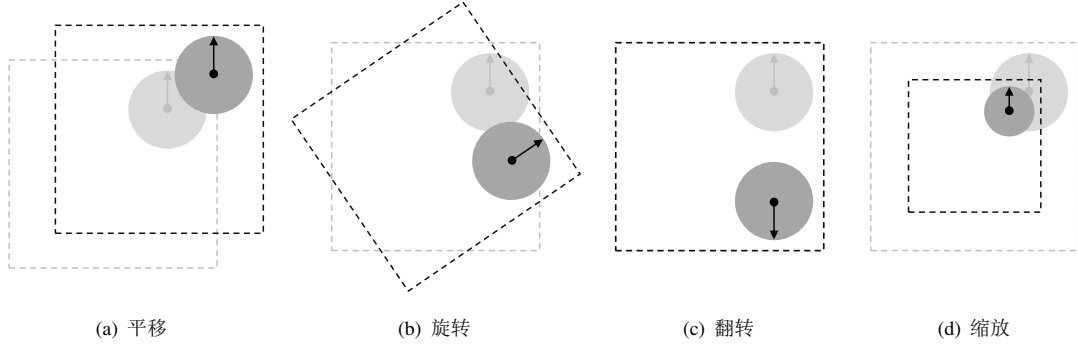


图 4.2 几何变换示意图。

其中,定义域是以 (u, v) 为圆心并以 w 为半径的圆形区域: $D = \{(x, y) : (x - u)^2 + (y - v)^2 \leq w^2\}$ 。

由此得到的公式 (4.7) 在构建 LIR 中扮演基础性作用, 因此本章中将其称为**基本公式**。符号 V_{nm}^{uvw} 的上标和下标分别表示空间域和频率域的参数。如图 4.1 所示, 这个新定义允许图像的定义域 D_f 和基函数的定义域 D_v 建立于不同的坐标系。从定义 4.1 中可以直接得到两个命题, 即广义性和局部表征能力。

命题 4.1: 广义性。经典矩和矩不变量的全局定义 (4.4) 是 LIR 基本公式 (4.7) 的特例, 具有固定的位置参数 $(u, v) = (0, 0)$ 和尺度参数 $w = 1$ 。

因此, LIR 可被视为矩和矩不变量研究的统一数学框架。

命题 4.2: 局部表征能力。基于 LIR 基本公式 (4.7), 可以分别通过调整位置参数 (u, v) 和尺度参数 w 来控制局部表征区域 D 的中心和尺度。

局部表示能力是经典矩和矩不变量所不具有的性质, 因此 LIR 在图像处理和视觉理解的局部 (像素级) 任务中具有潜在应用。

定义的总结: 数学上, 相关工作^[93,267]是基本公式 (4.7) 的特例, 具有固定的 $R_n(r)$ 。这意味着我们的表征是一种更通用的框架, 并且接下来对 (4.7) 性质和计算的讨论也是完全通用的。而相关工作^[91]的笛卡尔定义与公式 (4.3) 不一致, 因此也与基本公式 (4.7) 不一致, 如下文所述这种笛卡尔定义限制了其实现旋转不变性。

4.3.2 关键性质和表征公式

表征的稳健性对取证算法非常关键, 因为敌手可能会引入基于几何变换和信号损失的反取证攻击^[274]。

接下来, 我们通过分析 LIR 基本公式 (4.7) 在不同图像变换下的重要表示特性, 旨在满足取证任务中的稳健性要求。注意, 这种局部稳健性在基于全局假设的矩和矩不变量中尚未得到研究。

令 f_T 为图像函数 f 的平移版本, 具有偏移量 $(\Delta x, \Delta y)$, 可写为 $f_T(x, y) = f(x + \Delta x, y + \Delta y)$,

如图 4.2 所示。

性质 4.1: 平移等变性。将 f_T 代入基本公式 (4.7)，可以验证图像平移操作只影响表征参数 (u, v) ，如下：

$$\langle f_T(x, y), V_{nm}^{uvw}(x, y) \rangle = \langle f(x, y), V_{nm}^{(u+\Delta x)(v+\Delta y)w}(x, y) \rangle, \quad (4.8)$$

其中，相同的偏移量 $(\Delta x, \Delta y)$ 也出现在表征的位置参数 (u, v) 中，这意味着平移等变性。

证明：公式 (4.8) 的正确性基于以下推导过程：

$$\begin{aligned} & \langle f_T(x, y), V_{nm}^{uvw}(x, y) \rangle \\ &= \iint_D R_n^* \left(\frac{1}{w} \sqrt{(x-u)^2 + (y-v)^2} \right) A_m^* \left(\arctan \left(\frac{y-v}{x-u} \right) \right) f(x+\Delta x, y+\Delta y) dx dy \\ &= \iint_D R_n^* \left(\frac{1}{w} \sqrt{(x-\Delta x-u)^2 + (y-\Delta y-v)^2} \right) A_m^* \left(\arctan \left(\frac{y-\Delta y-v}{x-\Delta x-u} \right) \right) f(x, y) dx dy \\ &= \iint_D R_n^* \left(\frac{1}{w} \sqrt{(x-(u+\Delta x))^2 + (y-(v+\Delta y))^2} \right) A_m^* \left(\arctan \left(\frac{y-(v+\Delta y)}{x-(u+\Delta x)} \right) \right) f(x, y) dx dy \\ &= \langle f(x, y), V_{nm}^{(u+\Delta x)(v+\Delta y)w}(x, y) \rangle. \end{aligned} \quad (4.9)$$

□

命题 4.3: 平移不变表征。由平移等变性保障，可以通过特征匹配在 (u, v) 域检索给定模式的平移版本（局部不变），还可以通过 (u, v) 域的置换不变映射（如平均或最大池化）生成平移不变特征（全局不变）。

平移等变性在接下来推导旋转、翻转和缩放的性质时也是至关重要的。

令 f_R 为图像函数 f 围绕中心的旋转版本，具有旋转角 ϕ ，可写为 $f_R(r, \theta) = f(r, \theta + \phi)$ ，其中方便起见使用极坐标。注意，考虑 f_R 和 f 中任意一对相对应的圆形区域，它们的几何关系事实上可以建模的中心对齐旋转和平移的复合变换，为如图 4.2 所示。由于平移等变性的保障，接下来的分析可仅关注中心对齐旋转而不失一般性，即将位置参数固定为 $(u, v) = (0, 0)$ 。

性质 4.2: 旋转协变性。将 f_R 代入基本公式 (4.7) 并重写为极坐标形式，其中限制位置参数 $(u, v) = (0, 0)$ ，可以验证图像旋转操作只影响表征的相位，如下：

$$\langle f_R(r, \theta), V_{nm}^{uvw}(r', \theta') \rangle = \langle f(r, \theta), V_{nm}^{uvw}(r', \theta') \rangle A_m^*(-\phi), \quad (4.10)$$

其中，相同的旋转角 ϕ 也出现在表征的相位中，这意味着旋转协变性。

证明：公式 (4.10) 的正确性基于以下推导过程：

$$\begin{aligned}
 & \langle f_R(r, \theta), V_{nm}^{uvw}(r', \theta') \rangle \\
 &= \iint_D R_n^*(r') A_m^*(\theta') f(r, \theta + \phi) r' dr' d\theta' \\
 &= \frac{1}{w^2} \int_0^{2\pi} \int_0^w R_n^*\left(\frac{r}{w}\right) A_m^*(\theta) f(r, \theta + \phi) r dr d\theta \\
 &= \frac{1}{w^2} \int_0^{2\pi} \int_0^w R_n^*\left(\frac{r}{w}\right) A_m^*(\theta - \phi) f(r, \theta) r dr d\theta \quad (4.11) \\
 &= \frac{1}{w^2} \int_0^{2\pi} \int_0^w R_n^*\left(\frac{r}{w}\right) A_m^*(\theta) A_m^*(-\phi) f(r, \theta) r dr d\theta \\
 &= \frac{1}{w^2} \langle f(r, \theta), V_{nm}^{uvw}(r', \theta') \rangle w^2 A_m^*(-\phi) \\
 &= \langle f(r, \theta), V_{nm}^{uvw}(r', \theta') \rangle A_m^*(-\phi).
 \end{aligned}$$

□

命题 4.4：旋转不变表征。由相位域的旋转协变性保障，旋转不变性将在幅度域直接成立：

$$|\langle f_R(r, \theta), V_{nm}^{uvw}(r', \theta') \rangle| = |\langle f(r, \theta), V_{nm}^{uvw}(r', \theta') \rangle|.$$

在实现旋转不变性方面，更广义且信息保持的策略是采用适当的相位抵消从而排除项 $A_m^*(-\phi)$ 的影响。这样得到的特征通常是复数值，并含有相位信息。另外，也可以从相位中提取旋转角度，即角度估计。相位抵消和角度估计的更多细节可参阅文献^[110]。

命题 4.5：翻转不变表征。类比旋转协变性的推导，水平翻转版本 $f_{HF}(r, \theta) = f(r, \pi - \theta)$ 代入基本公式 (4.7) 可得 $(-1)^m (\langle f(r, \theta), V_{nm}^{uvw}(r', \theta') \rangle)^*$ 。因此水平翻转不变性在幅度域也成立。进一步基于中心对齐旋转和水平翻转的复合变换，可将水平不变性推广为任意翻转不变性。

令 f_S 为图像函数 f 的缩放版本，具有缩放因子 s ，可写为 $f_S(x, y) = f(sx, sy)$ 。同样地，考虑 f_S 和 f 中任意一对相对应的圆形区域，它们的几何关系事实上可以建模的中心对齐缩放和平移的复合变换，为如图 4.2 所示。由于平移等变性的保障，接下来的分析可仅关注中心对齐缩放而不失一般性，即将位置参数固定为 $(u, v) = (0, 0)$ 。

性质 4.3：缩放协变性。将 f_S 代入基本公式 (4.7)，其中限制位置参数 $(u, v) = (0, 0)$ ，可以验证图像缩放操作只影响表征参数 w ，如下：

$$\langle f_S(x, y), V_{nm}^{uvw}(x, y) \rangle = \langle f(x, y), V_{nm}^{uv(ws)}(x, y) \rangle, \quad (4.12)$$

其中，相同的因子 s 也出现在表征的尺度参数 w 中，这意味着缩放协变性。

证明：公式 (4.12) 的正确性基于以下推导过程：

$$\begin{aligned}
 & \langle f_S(x, y), V_{nm}^{uvw}(x, y) \rangle \\
 &= \iint_{x^2+y^2 \leq w^2} R_n^*\left(\frac{1}{w}\sqrt{x^2+y^2}\right) A_m^*\left(\arctan\left(\frac{y}{x}\right)\right) f(sx, sy) dx dy \\
 &= \iint_{x^2+y^2 \leq w^2 s^2} R_n^*\left(\frac{1}{w}\sqrt{\left(\frac{x}{s}\right)^2 + \left(\frac{y}{s}\right)^2}\right) A_m^*\left(\arctan\left(\frac{y}{x}\right)\right) f(x, y) d\frac{x}{s} d\frac{y}{s} \\
 &= \frac{1}{s^2} \iint_{x^2+y^2 \leq w^2 s^2} R_n^*\left(\frac{1}{ws}\sqrt{x^2+y^2}\right) A_m^*\left(\arctan\left(\frac{y}{x}\right)\right) f(x, y) dx dy \\
 &= \frac{1}{s^2} \langle f(x, y), V_{nm}^{uv(ws)}(x, y) \rangle s^2 \\
 &= \langle f(x, y), V_{nm}^{uv(ws)}(x, y) \rangle.
 \end{aligned} \tag{4.13}$$

□

命题 4.6：缩放不变表征。由缩放协变性保障，可以通过特征匹配在 w 域检索给定模式的缩放版本（局部不变），还可以通过 w 域的置换不变映射（如平均或最大池化）生成缩放不变特征（全局不变）。

注意，由于对尺度进行稠密采样是困难的，上述缩放不变性在实践中往往会降级为某种稳健性。除了置换不变映射之外，尺度选择也是实现缩放不变性的常见方法，因而稠密尺度选择方法^[296]也可以与 LIR 结合实现缩放不变性。

继承自经典矩和矩不变量理论， V_{nm}^{uvw} 的集合形成了 Hilbert 空间的正交完备基，这种信息保持结构保障了捕捉信息的唯一性和独立性^[16,32]，从而提高了判别性。另外，低阶 (n, m) 表征在信号损失（如噪声、模糊和采样/量化效应）下更加稳健，其中信号损失操作主要影响图像高频分量，对应高阶矩^[16,32]。为了让更稳健的信息优先用于表征，阶数 (n, m) 的集合可以由特定范数和常数 K 约束：

$$S_{nm}(K) = \{(n, m) : \|(n, m)\|_\ell \leq K\}, \tag{4.14}$$

其中 $\|\cdot\|_\ell$ 表示 ℓ 范数，文献中通常取为 ℓ_1 或 ℓ_∞ 。

性质公式 (4.8)、(4.10) 和 (4.12) 为构建局部不变量提供了可解释的指导，因此在本章中称为**表征公式**。注意，出于判别性和效率的考虑，在表征中实现上述全部类型的不变性并不明智，应根据应用需求选择所需的不变性。

性质的总结：相关工作^[91,93,267]几乎没有讨论平移、翻转、缩放和信号损失下的表征属性。相关工作^[93,267]也具有公式 (4.8) 的旋转协变性，因而可以构造旋转不变表征，而基于笛卡尔定义的相关工作^[91]则不具有上述性质。

4.3.3 精确计算和快速计算

连续域的基本公式和表示公式已在前文给出。接下来，我们将专注于数字图像中的离散实现，讨论 LIR 的精度和效率问题并提供相应的解决方案。

面向数字一致性的取证依赖于高阶统计特征，例如 PRNU 方法^[292]。在这种场景中，计算误差甚至可能主导取证性能。如后文所述，LIR 的直接实现并不适用于这种高精度取证场景。为此，我们旨在通过减小数值积分误差和表示误差，以满足取证任务中的计算精度要求。

考虑在离散笛卡尔坐标系网格上定义数字图像 $\{f(i, j) : (i, j) \in \{1, 2, \dots, M\} \times \{1, 2, \dots, N\}\}$ 。方便起见，我们引入离散变量 (i, j) 的连续版本 $(x, y) \in [1, M] \times [1, N]$ 。其中，一个以 (i, j) 为中心的像素区域被定义为 $D_{ij} = \{(x, y) \in [i - \frac{\Delta i}{2}, i + \frac{\Delta i}{2}] \times [j - \frac{\Delta j}{2}, j + \frac{\Delta j}{2}]\}$ ；在该区域上， f 的值是恒定的且等于 $f(i, j)$ 。

定义 4.2：离散定义。基于上述符号，基本公式（4.7）可以重写为离散形式，如下：

$$\langle f, V_{nm}^{uvw} \rangle = \sum_{(i,j) \text{ s.t. } D_{ij} \cap D \neq \emptyset} h_{nm}^{uvw}(i, j) f(i, j), \quad (4.15)$$

其中， $h_{nm}^{uvw}(i, j)$ 是 $(V_{nm}^{uvw})^*$ 在像素区域 D_{ij} 与定义域 D 交集上的积分值：

$$h_{nm}^{uvw}(i, j) = \iint_{D_{ij} \cap D} (V_{nm}^{uvw}(x, y))^* dx dy. \quad (4.16)$$

由于复杂函数的连续积分，LIR 的计算精度主要受到公式（4.16）的影响。具体地说，我们需要确定数值积分策略并计算采样点处基函数的值。接下来将分别讨论这两个方面。

对于数值积分策略，零阶近似是一种常用的策略，直接设置 $h_{nm}^{uvw}(i, j) \simeq (V_{nm}^{uvw}(i, j))^* \frac{\Delta i \Delta j}{w^2}$ 。数学上，其误差取决于函数的频率，因此当阶数 (n, m) 较大或者尺度参数 w 较小时误差可能非常显著。注意，这两种情况在 LIR 中均很常见。因此，我们建议根据应用的要求引入高精度数值积分方法，比如伪上采样^[63]和高斯法^[66]。这些策略可以统一表述为如下定义。

定义 4.3：基于高精度数值积分的精确计算。基于 L 维求积公式，可得到公式（4.16）的数值近似为：

$$h_{nm}^{uvw}(i, j) \simeq \sum_{(a,b)} c_{ab} (V_{nm}^{uvw}(x_a, y_b))^* \frac{\Delta i \Delta j}{w^2}, \quad (4.17)$$

其中， $(x_a, y_b) \in D_{ij}$ 为采样点，具有相应的权重 c_{ab} ，有 $\#\{(a, b)\} = L$ 。

命题 4.7：精确性分析。给定 (n, m) 时公式（4.17）的近似误差为 $O((\frac{\Delta i \Delta j}{w^2})^{L+1})$ ，当 $L > 1$ 时相对于零阶近似误差 $O((\frac{\Delta i \Delta j}{w^2})^2)$ 更精确。

计算基函数的采样值时，以数值不稳定为代表的表示误差是另一类精度问题，主要由雅可比多项式中的阶乘或伽马项引起。我们建议使用递归策略^[159]，直接从几个低阶基函数导出高

阶基函数，避免大数的阶乘或伽马计算。

离散公式 (4.15)、(4.16) 和 (4.17) 提供了 LIR 的精确数值近似，因此在本章中称为**精确计算公式**。

从取证任务的稠密性和稳健性要求出发，我们倾向于对 LIR 参数 (u, v) 和 w 进行充分采样。在这种设置下，基于定义的直接计算会表现出相当高的复杂性。为此，我们旨在引入一些有益的计算定理和数据结构，以满足取证任务中的计算效率要求。

定义 4.4: 卷积定义。 考虑 (u, v) 在离散网格 $\{1, 2, \dots, M\} \times \{1, 2, \dots, N\}$ 上的稠密采样，离散公式 (4.15) 可以重写为卷积形式，如下：

$$\langle f, V_{nm}^{uvw} \rangle = f(i, j) \otimes (H_{nm}^w(i, j))^T, \quad (4.18)$$

其中， $\otimes$ 表示卷积运算， $(\cdot)^T$ 表示矩阵转置； H_{nm}^w 是由 h_{nm}^{uvw} 定义的卷积核：

$$H_{nm}^w(i, j) = \{h_{nm}^{uvw}(i, j) : u, v = w, (i, j) \text{ s.t. } D_{ij} \cap D \neq \emptyset\}. \quad (4.19)$$

这里，公式 (4.15) 中的稠密内积运算被转换为公式 (4.18) 中的卷积。注意，我们使用术语“卷积”而不是类似的术语“互相关”是为了方便后续讨论。

给定 (n, m) 时公式 (4.18) 的计算复杂度由以下参量共同决定：核 H_{nm}^w 的大小（即 $4w^2$ ）、参数 (u, v) 采样集合 S_{uv} 的大小（记为 $\#_{uv} = |S_{uv}|$ ）、以及参数 w 采样集合 S_w 的大小（记为 $\#_w = |S_w|$ ）。总共需要 $\Theta(4 \sum_{w \in S_w} w^2 \#_{uv})$ 或 $O(w_{\max}^2 \#_{uv} \#_w)$ 次乘法运算，其中 $w_{\max}$ 是 S_w 中的最大值。至于公式 (4.19)，其复杂度几乎可以忽略不计，因为核函数只需要在给定 n 、 m 和 w 的情况下计算一次，无需 (u, v) 上的密集计算。

接下来将介绍公式 (4.18) 的快速计算策略。

定义 4.5: 基于卷积定理的快速计算。 引入傅里叶变换的卷积定理，将公式 (4.18) 中的空间域卷积可以转换为频域乘积^[297]：

$$\langle f, V_{nm}^{uvw} \rangle = \mathcal{F}^{-1}(\mathcal{F}(f) \odot \mathcal{F}((H_{nm}^w)^T)), \quad (4.20)$$

其中， $\mathcal{F}$ 表示傅立叶变换， $\odot$ 表示逐点乘法。

命题 4.8: 卷积定理的复杂性分析。 若采用快速傅里叶变换实现公式 (4.20)，则乘法复杂度降为 $\Theta(\#_w(3\#_{uv} \log \#_{uv} + \#_{uv}))$ 或 $O(\#_w \#_{uv} \log \#_{uv})$ 。

注意，尺度参数 w 对于复杂度没有影响，即相对于表征尺度的常数阶计算复杂度。理论上，当 $w_{\max}$ 足够大使得 $w_{\max}^2$ 大于 $\log(\#_{uv})$ 时，基于快速傅里叶变换的实现将比公式 (4.18) 的直接计算更快速。实际上上述结论非常重要，因为对 w 的采样通常涉及一些较大的值以满足缩放

不变性。

接下来将指出定义 4.5 中可进一步加速的环节。

定义 4.6: 基于缩放定理的快速计算。引入傅里叶变换的缩放定理, 则频域系数项 $\mathcal{F}((H_{nm}^w)^T)$ 可由已经计算过的项 $\mathcal{F}((H_{nm}^{w_0})^T)$ 直接导出:

$$\mathcal{F}((H_{nm}^w)^T) = \left(\frac{w}{w_0}\right)^2 \mathcal{F}((H_{nm}^{w_0})^T) \left(\frac{w\xi_i}{w_0}, \frac{w\xi_j}{w_0}\right), \quad (4.21)$$

其中, (ξ_i, ξ_j) 是频率自变量。

命题 4.9: 缩放定理的复杂性分析。对于给定的 (n, m) 以及一对 w 和 w_0 , 公式 (4.21) 用插值法代替了快速傅里叶变换, 并因此节省 $\Theta(\#_{uv} \log \#_{uv})$ 次乘法。

除卷积和缩放定理外, 我们还考虑一些实际的加速技巧。

命题 4.10: 其他快速计算及其复杂性分析。回到公式 (4.20), 可以观察到项 $\mathcal{F}((H_{nm}^w)^T)$ 实际上与输入 f 的内容无关, 只与其大小相关。因此, 可以预先计算出一个索引为 (n, m, w) 的查找表, 包含尺寸固定为 $M_0 \times N_0$ 的一组 $\mathcal{F}((H_{nm}^w)^T)$, 同时将数字图像 f 的尺寸归一化为 $M_0 \times N_0$ 。因此, 对于给定的 (n, m) 和 w , 查找表策略可节省 $\Theta(\#_{uv} \log \#_{uv})$ 次乘法。

离散公式 (4.18)、(4.19)、(4.20) 和 (4.21) 提供了 LIR 的高效数值实现, 因此在本章中称为快速计算公式。

计算的总结: 在相关工作^[91,93,267]中, 几乎没有讨论过精确计算。相关工作^[91,93]提出了类似的常数阶计算策略, 但这些快速计算都依赖于特定的基函数。相比之下, 本章工作给出了通用的快速计算框架, 不受基函数具体定义的限制。注意, 精确计算策略 (4.17) 主要适用于 w 较小的情况, 而快速计算策略 (4.20) 则适用于 w 较大的情况。

4.4 局部不变量实践

本节介绍 LIR 的实践方面。我们给出了一般实践流程的可视化, 并提供了面向取证任务的设置原则。

4.4.1 实践示例

为了使 4.3 节的理论转化为直观认知, 我们在图 4.3 中给出了 LIR 在纹理图像上的实践示例。

计算方面。回到公式 (4.17) 和 (4.19), 首先可由不同阶数 (n, m) 和尺度 w 的基函数 V_{nm}^{uvw} 导出一组卷积核 H_{nm}^w 。图 4.3 展示了一组 H_{nm}^w 的实部和虚部。随后, 计算图像 f 与转置核 $(H_{nm}^w)^T$ 的空域卷积 (4.18) 或与之等价的频域乘积 (4.20), 由此将图像 f 分解为沿多频率和多尺度的稠密系数 $\langle f, V_{nm}^{uvw} \rangle$ 。

表征方面。如图 4.3 所示, 这些系数的幅度捕捉了主要的结构信息 (线条模式), 而相位提

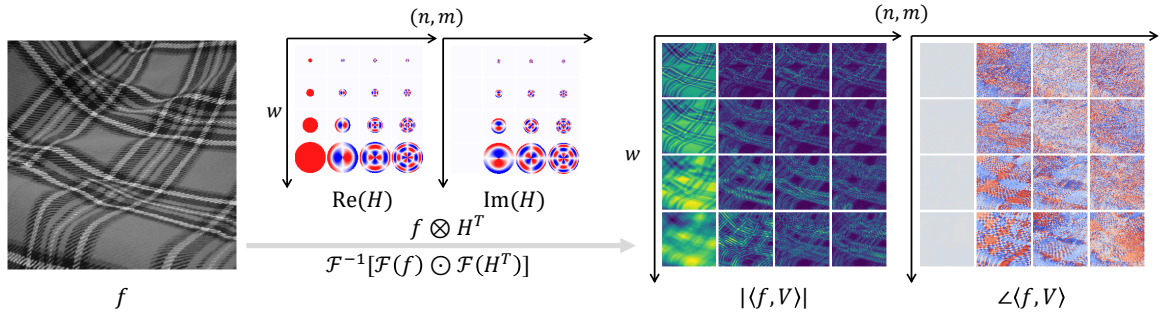


图 4.3 实践流程示意图。

供了非常丰富的细节（纹理特征）。为了获得稳健表征，可以通过系数的幅度（对旋转和翻转不变）和池化（对平移和缩放不变）构造相应的不变量，参照表征公式（4.8）、（4.10）和（4.12）。在实践中，基于幅度的 LIR 通常可以满足要求；若需提高判别性，可以通过适当的相位抵消将相位信息也编码进特征。

4.4.2 实践原则

接下来，我们提供参数和径向基函数的设置原则，主要面向在取证任务中的实践。

鉴于 LIR 参数 (n, m) 、 (u, v) 和 w 的物理意义，其设置主要参考取证问题的数学性质。

- 参数 (n, m) 控制所表征图像信息的频率。图像取证分析关注的语义、物理和数字一致性特征分别对应低频、中频和高频信息^[274]，因此可分别采用低阶、中阶和高阶 (n, m) 进行有效表征。
- 参数 (u, v) 控制所表征的图像信息的位置。理论上， (u, v) 的逐像素采样可最大程度满足取证任务的稠密性要求，避免篡改区域的漏检。在实践中，可采用间隔采样以降低计算复杂性，涉及复杂性和稠密性要求的权衡。
- 参数 w 控制所表征图像信息的尺度。对于不需尺度不变性的取证任务， w 可以直接取一固定值。相反，对于需要尺度不变性的任务， w 应该被充分采样。特别应该注意的是，如果已知待分析模式的尺度先验，应该在设置中引入这些知识以避免不必要的采样。

至于 LIR 中径向基函数的定义，其设置主要考虑表征能力和计算精度/效率。

- 表征能力主要受径向基函数的零点分布影响。一般而言，在 $[0, 1]$ 上零点分布均匀的径向基函数是最优选择。另一种可行的方法是将具有互补零点分布的径向基函数组合使用，以捕捉互补的图像信息^[159]。
- 计算精度/效率主要受径向基函数的基本数学性质影响。其中，复杂度与是否涉及阶乘/伽马项、求和/级数运算和求根过程有关，数值稳定性与阶乘/伽马项和函数有界性有关。

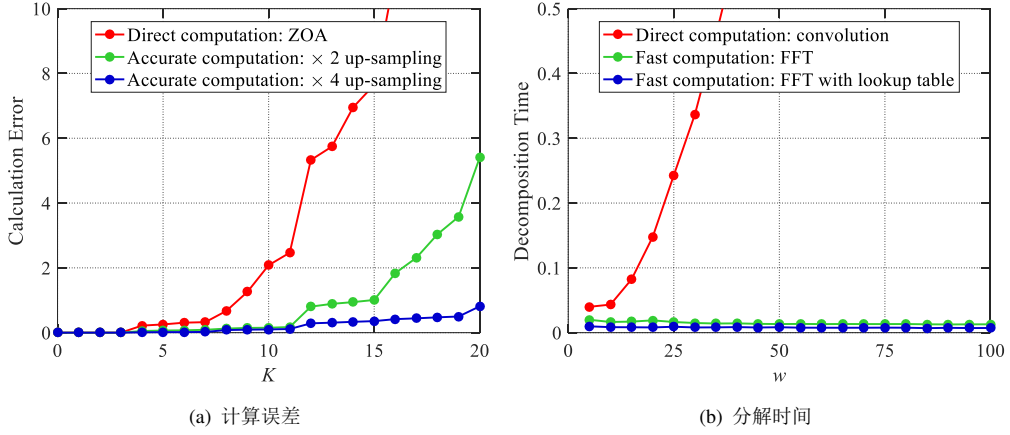


图 4.4 不同计算方法的精确性和复杂性对比。

4.5 实验和应用

本节将全面评估所提出的 LIR 的性能，涵盖仿真实验和实际应用层面的对比。

在仿真实验层面，我们提供了精确性/复杂性和稳健性/不变性的数值结果。对于精确性和复杂性，我们比较了 LIR 的几种计算策略。对于稳健性和不变性，我们在两类经典视觉任务中，即稠密模式检测和匹配，分别测试了 LIR 与先进稠密局部表征方法在多种几何或信号劣化下的综合性能。

在实际应用层面，我们分别考虑了被动取证路径中复制-粘贴篡改检测的和主动取证路径中的感知哈希。其中，LIR 作为特征模块直接应用于上述算法，提供了与先进取证方法在典型场景和稳健性场景下的取证精度和效率对比。

为了简洁并凸显方法的有效性，一组简单的余弦函数被选为径向基函数用于全部实验，即 PCT^[57] 的 LIR 扩展版本。

4.5.1 精确性和复杂性

我们采用 2.4.1 节的精确性和复杂性指标，分别评估 4.3.3 节中给出的几种 LIR 计算策略。

精确性。比较方法包括基于零阶近似的直接计算和基于伪上采样的精确计算，详见定义 4.2 和 4.3 以及命题 4.7。理论上，数值积分误差分别与函数频率和采样数成正比和反比关系，因此实验考虑 $w = 8$ 和 $K \in \{0, 1, \dots, 20\}$ 。

在图 4.4 (a) 中，我们给出了计算误差的对比结果。可以观察到，上采样策略的误差要小于直接计算，特别是对于更高的上采样率和更高的阶数 K 。显然，在 w 较小和 (n, m) 较大的情况下，直接计算的 LIR 系数可能会包含相当显著的误差，从而干扰后续任务。上述数据支撑了我们关于误差的理论分析结论，验证了准确计算公式在实际场景中的有效性。

复杂性。比较方法包括基于卷积的直接计算和基于快速傅里叶变换的快速计算，详见定义

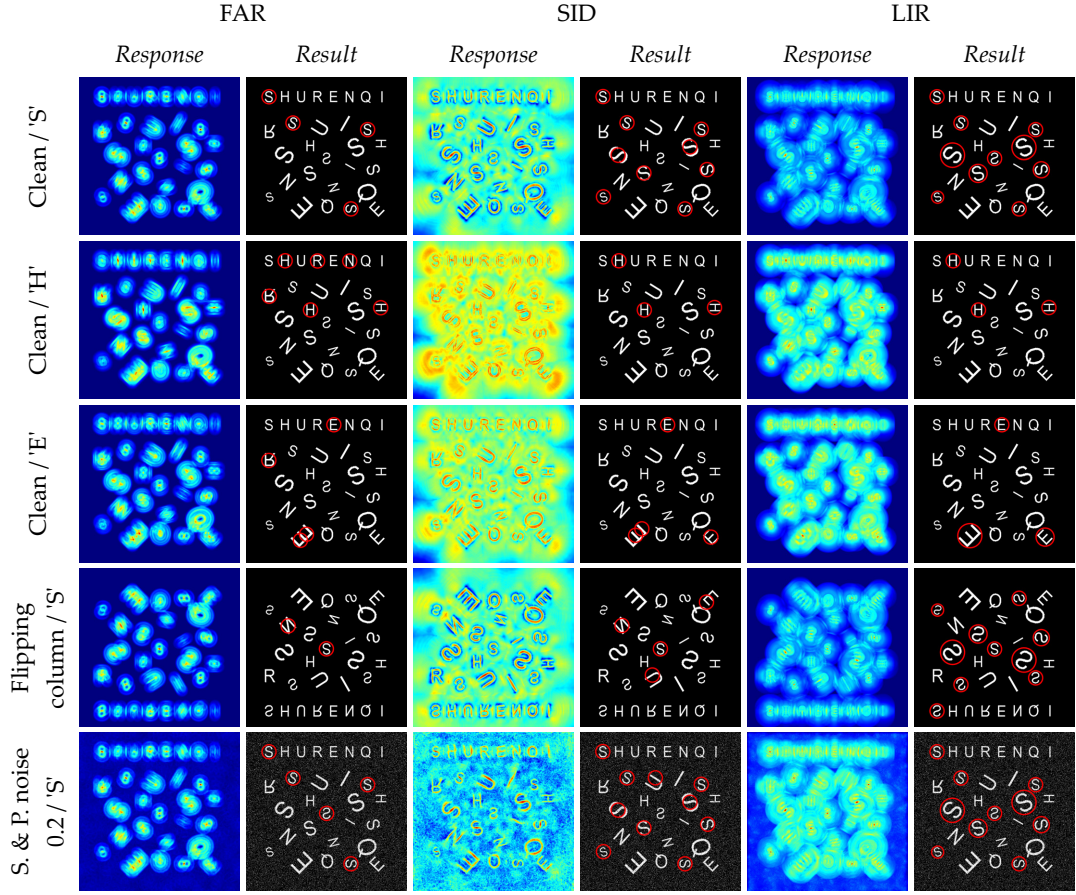


图 4.5 不同稠密局部表征方法的模式检测示例。

4.4~4.6 以及命题 4.8~4.10。实验考虑计算大小为 512×512 图像的稠密 LIR 系数，即 $(u, v) \in \{1, 2, \dots, 512\}^2$ ，其中阶数固定为 $(n, m) = (1, 1)$ ，尺度参数为 $w \in \{5, 10, \dots, 200\}$ 。

在图 4.4 (b) 中，我们给出了分解时间的对比结果。随着 w 的增加，直接计算的时间曲线急剧上升，而基于快速傅里叶变换的快速计算则表现出几乎平直的时间曲线，引入查找表技巧进一步减少了其时间。显然，该观察支撑了我们关于复杂性的理论分析结论。注意，这里只考虑固定 (n, m, w) 的计算时间，而在实际场景中通常需要一组 (n, m, w) ，上述计算策略之间的速度差距将进一步扩大。

4.5.2 稳健性和不变性

接下来将通过两个经典视觉任务，即稠密模式检测和匹配，评估 LIR 的稳健性和不变性水平。注意，这两个任务在表征层面具有不同的特性：检测通常使用大尺度、多尺度的复杂表征，因为结果的稀疏性；而匹配通常使用小尺度、单尺度的紧凑表征，因为结果的稠密性。

注意，实验中的算法框架非常简单，因为目标是反映不同表征在两种任务框架中的性能水平，而不是直接在这些任务上取得最先进的结果。

表 4.2 不同稠密局部表征方法在模式检测实验中的 F1 分数 (%)。

表征方法	FAR ^[298]	SID ^[267]	LIR
干净	61.54	92.68	97.56
旋转 20°	60.00	73.08	100.00
旋转 45°	56.41	56.25	97.56
翻转列	34.48	31.58	97.56
翻转行	34.48	26.47	97.56
缩放 0.8	25.64	35.05	74.07
缩放 0.5	19.35	19.18	45.61
高斯噪声 0.01	61.54	92.31	100.00
高斯噪声 0.02	61.54	92.68	93.33
椒盐噪声 0.01	60.00	90.48	97.56
椒盐噪声 0.02	60.00	92.68	100.00
均值滤波 7 × 7	43.24	45.57	61.76
均值滤波 9 × 9	40.00	37.11	60.61
高斯滤波 7 × 7	45.71	50.75	84.00
高斯滤波 9 × 9	43.24	46.34	59.15
中值滤波 7 × 7	60.00	92.68	100.00
中值滤波 9 × 9	61.22	92.68	97.67
JPEG 压缩 10	61.54	92.68	100.00
JPEG 压缩 5	61.54	92.68	100.00
拉普拉斯锐化	57.89	92.68	100.00
分数平均值 ↑	50.47	67.28	88.20
分数标准差 ↓	13.26	27.26	17.12

4.5.2.1 稠密模式检测

在该实验中，我们考虑从一个包含多个字母及其旋转/缩放/翻转版本的 800×800 图像中检测给定模式，如图 4.5 所示。此外，实验还涉及图像全局的几何和信号劣化攻击，作为更具挑战性的实验协议。

算法设计。 给定模式是从原始干净图像上直接取出的字母，具有固定的朝向和尺度。给定模式的特征向量作为基准，用于和整幅图像的稠密特征进行对比：通过计算特征间的欧氏距离，将距离较小的位置视为模式检测结果。我们使用 F1 分数来评估检测精度。

对比方法包括具有不变性的稠密局部表征方法：FAR^[298] 和 SID^[267]。其中，FAR 是专门设计用于旋转不变模式检测的表征，其在严重噪声条件下的具有稳健性；SID 则具有内在的旋转和缩放不变性，并且无需尺度和方向的池化或选择。

我们的 LIR 实现细节如下。在公式 (4.14) 中，设置最大阶数 $K = 5$ ，基于无穷范数，使得 LIR 在尺度空间中每个位置 (u, v, w) 上具有 36 维特征。考虑位置参数 (u, v) 为图像网格上的稠密采样，尺度参数 w 从 30 增加到 120 并含 10 个采样点。对于 FAR 和 SID，我们使用其原始论文中的常规参数设置。注意，FAR 和 SID 都定义在单尺度上，在每个平面位置上的特征维数分

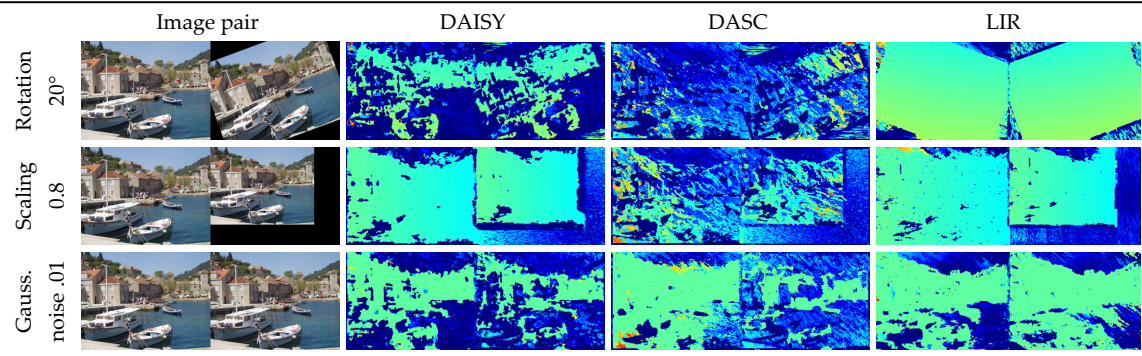


图 4.6 不同稠密局部表征方法的模式匹配示例。

别为 21 和 1008。采样方面，FAR 是稠密采样，而 SID 采用 2 个像素的采样间隔，为了降低时间/空间成本（避免内存溢出）。

稳健性与不变性。图 4.5 给出了特征响应图和最终检测结果的示例，表 4.2 列出了对比方法在各种实验协议下的 F1 分数。可以观察到，在劣化操作下，稠密局部表征要在保持判别性的同时实现稳健性是一项挑战性任务。FAR 表现出旋转不变性，并且对于严重噪声具有强稳健性。但是 FAR 无法应对尺度变化，并且在滤波类操作下特征不稳定。SID 的对数极坐标定义使得其在单一尺度上实现旋转和缩放不变性。然而，响应图表明 SID 特征在排除不相关字母方面较弱，即判别性较低。另外，FAR 和 SID 都不具备翻转不变性。与之相对，所提出的 LIR 对于旋转、翻转和缩放具有更强的稳健性，这应该归功于协变性的充分利用。此外，由于 LIR 基函数的正交性和完备性，能够很好地在信号劣化环境中良好区分相关和无关模式。

效率。所提出的 LIR 保持了合理的时间成本：FAR，SID 和 LIR 的 CPU 运行时间分别约为 2 秒，35 秒和 6 秒，注意我们的方法是多尺度表征。

4.5.2.2 稠密模式匹配

在该实验中，我们考虑在原始图像与其劣化版本之间建立稠密的局部对应关系，如图 4.6 所示。我们从 INRIA Holidays 图像集^[299]中选择了 10 张图像以提供平均分数，其中图像尺寸归一化为 1000×1333 。

算法设计。实验基于如下通用框架：PatchMatch^[300]用于匹配稠密特征，RANSAC^[301]用于排除错误匹配。这里，稠密特征的规则空间关系使得 PatchMatch 能够以较高效率实现几何不变的特征匹配^[302]。我们使用重复率分数^[15]来评估匹配精度。

对比方法包括具有不变性的稠密局部表征方法：DAISY^[18]和 DASC^[268]。其中，DAISY 是为宽基线立体匹配而设计，因此考虑了几何不变性（透视畸变）；DASC 进一步发展了 DAISY 的思想，额外关注光度变化下的稳健性。

我们的 LIR 实现细节如下。选择了最大阶数为 3 的一组 (n, m) ，使得 LIR 在尺度空间中每

表 4.3 不同稠密局部表征方法在模式匹配实验中的重复率分数 (%)。

表征方法	DAISY ^[18]	DASC ^[268]	LIR
干净	95.58	94.48	94.26
旋转 20°	28.67	0.16	85.01
旋转 45°	0.30	0.35	76.61
翻转列	0.23	0.15	92.81
翻转行	0.48	0.13	91.03
缩放 0.8	72.07	7.54	49.36
缩放 1.3	74.79	3.20	32.27
高斯噪声 0.01	23.66	31.43	35.60
高斯噪声 0.02	15.54	20.73	21.98
椒盐噪声 0.01	60.17	63.68	58.96
椒盐噪声 0.02	45.44	47.59	45.43
均值滤波 5 × 5	52.56	68.09	60.53
均值滤波 7 × 7	27.58	29.63	18.77
高斯滤波 5 × 5	54.81	69.70	64.47
高斯滤波 7 × 7	32.87	41.91	29.95
中值滤波 5 × 5	54.09	72.93	59.88
中值滤波 7 × 7	28.84	42.30	32.21
JPEG 压缩 10	48.49	40.25	35.97
JPEG 压缩 5	28.15	17.02	13.69
拉普拉斯锐化	43.75	79.58	47.39
分数平均值 ↑	39.40	36.54	52.31
分数标准差 ↓	25.10	29.58	24.84

个位置 (u, v, w) 上具有 10 维特征。考虑位置参数 (u, v) 为图像网格上的稠密采样，尺度参数 w 从 8 增加到 32 并含 10 个采样点。对于 DAISY 和 DASC，我们使用其原始论文中的常规参数设置。注意，DAISY 和 DASC 在每个位置的特征维度分别为 200 和 128，远多于 LIR 的维度。

稳健性和不变性。图 4.6 给出稠密局部对应的示例，表 4.3 列出了对比方法在各种实验协议下的重复率分数。几何变换方面，DAISY 和 DASC 很难保持稳定的匹配，特别是对于翻转和大角度旋转；DAISY 在缩放方面表现出良好的重复率，而 DASC 则不具有这样的性质。信号损失方面，DASC 重复率的整体水平要优于 DAISY，这与 DASC 的设计目标一致。与这些为匹配任务专门设计的描表征相比，所提出的 LIR 能够提供更好的几何变换稳健性和相近的信号损失稳健性。

效率。DAISY、DASC 和 LIR 的特征提取时间分别约为 6 秒、106 秒和 6 秒，特征匹配时间分别约为 79 秒、77 秒和 20 秒。基于所提出的常数阶计算，LIR 的特征提取非常快速，与以效率著称的 DAISY 相当。更重要的是由于基函数正交性，LIR 的特征向量更加紧凑，从而在匹配过程中节省了大量时间。

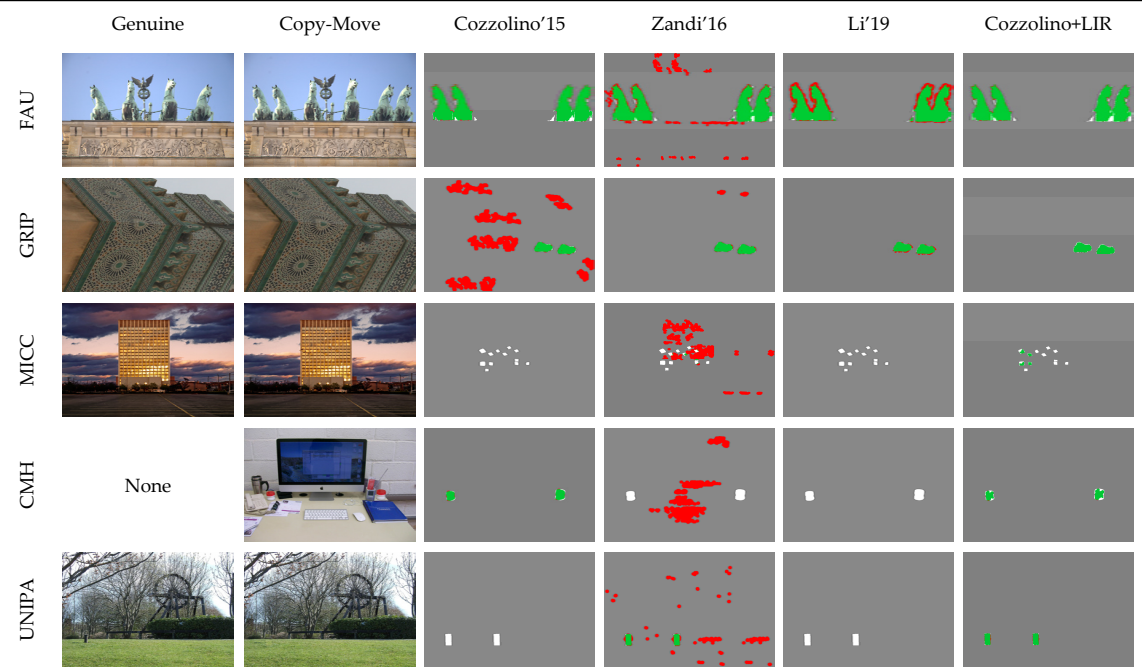


图 4.7 不同检测方法用于复制-粘贴篡改基准的结果示例。

4.5.3 被动取证：复制-粘贴篡改检测

复制-粘贴是图像篡改中最基本的操作之一，从图像中复制并粘贴某些区域以强调或掩饰特定的语义对象。复制-粘贴篡改的被动取证算法一般从图像中提取稀疏或稠密的局部特征，并通过匹配这些特征来揭示潜在的复制-粘贴行为。显然，提取特征的质量对性能有重要影响。整体上，稀疏方法更高效且几何不变性更好，但取证精确性较低；相反，稠密方法更准确，但效率较低且几何不变性较差。上述现象与我们在 4.1 节中的分析一致。

算法设计。由 Cozzolino 等人^[217]提出的复制-粘贴取证算法是一种代表性的稠密方法，依赖于圆谐正交矩（旋转不变）。理论上，其特征提取模块可被视为 LIR 的一种特例，具有固定的尺度。可以预见，该方法在实践中对旋转、翻转和信号损失具有稳定性，但对尺度变化敏感。因此，我们主要将 LIR 框架引入该算法中以改善其缩放稳健性，即原 PCT 特征提取模块被其 LIR 扩展版本替代。我们的 LIR 实现细节如下，选择了一组最大阶数为 3 的 (n, m) ，在尺度空间的每个位置 (u, v, w) 上生成 10 维特征，考虑位置参数 (u, v) 在图像网格上的稠密采样，尺度参数 w 从 8 增加到 32 并含 10 个采样点。为了实现具有缩放稳健性的紧凑表征，我们引入了跨尺度的平均池化。作为一种常见的技巧，低分辨率图像将被等比例放大（长边固定为 2000 像素），以抑制参数灵敏度。注意，我们在分数计算中排除了篡改区域和背景之间的边界像素^[217,286]。

复制-粘贴基准。我们首先在五个复制-粘贴取证基准上进行了定量对比：FAU^[286]、GRIP^[217]、MICC^[305]、CMH^[306] 和 UNIPA^[307]。这里，FAU、GRIP 和 UNIPA 仅包含刚性的复制-粘贴操作，而 MICC 和 CMH 则带有进一步的攻击（例如缩放和旋转）以实现令人信服的视觉效果。对比方

表 4.4 不同检测方法在多种复制-粘贴篡改基准上的取证分数 (%)。

取证方法		Cozzolino'15 ^[217]	Zandi'16 ^[303]	Li'19 ^[304]	Cozzolino+LIR
Precision	FAU	93.08	79.59	89.44	96.75
	GRIP	93.01	81.90	92.7	96.45
	MICC	92.02	66.06	88.77	94.30
	CMH	82.98	54.96	85.26	88.62
	UNIPA	85.37	73.07	86.18	96.5
Recall	FAU	91.94	95.09	88.94	92.44
	GRIP	96.41	98.64	97.43	95.89
	MICC	89.07	74.98	86.53	88.76
	CMH	78.99	65.27	71.98	75.91
	UNIPA	91.76	99.05	93.95	95.03
F1	FAU	91.89	83.39	88.63	93.62
	GRIP	93.85	86.67	94.75	95.74
	MICC	89.22	66.41	86.73	89.82
	CMH	80.12	58.25	76.35	79.50
	UNIPA	88.31	81.01	89.30	95.66

法包括基础的 Cozzolino'15^[217] 及其 LIR 版本, 最先进的稀疏方法: Silva'15^[306]、Zandi'16^[303] 和 Li'19^[304], 以及最先进的稠密方法: Ryu'13^[308]、Li'13^[309]、Emam'16^[310]、Pun'18^[311]、Bi'18^[312]、Wu'18^[313] 和 Zhong'20^[314]。其中, Wu'18 和 Zhong'20 基于深度学习网络。注意, Ryu'13、Li'13、Emam'16、Pun'18 和 Bi'18 中的特征提取模块也可视为 LIR 的特例。

图 4.7 给出了上述基准上的复制-粘贴取证示例, 表 4.4 列出了流行的开源算法^[217,303,304] 在以上全部基准上的取证分数, 表 4.5 则列出了其他(闭源)对比方法在 FAU 基准上的取证分数。可以观察到, 总体上稠密方法, 特别是 Cozzolino'15 及其 LIR 扩展, 提供了比稀疏方法更高的取证精度。稀疏方法中, Li'19 在性能上优于 Zandi'16, 这主要归因于其更稠密的特征点。上述现象证实了表征稠密性对于取证任务的重要性。通过引入 LIR, Cozzolino'15 在大多数基准上的分数进一步提高(在 CMH 上基本持平), 这意味着 LIR 的实用性而不仅仅是一种数学上的扩展。这种性能提升在很大程度上归因于 LIR 的多尺度结构, 使得特征在同时具有缩放稳健性和足够的信息量。

复制-缩放-粘贴稳健性。我们接下来评估基础的 Cozzolino'15 和其 LIR 版本对缩放的稳健性。实验在 FAU 基准上进行并采用复制-缩放-粘贴操作, 其中缩放因子设为 0.8。对比方法包括最先进的稠密方法 Ryu'13^[308]、Li'13^[309]、Emam'16^[310]、Pun'18^[311]、Bi'18^[312]、Wu'18^[313] 和

表 4.5 多种检测方法在 FAU 复制-粘贴篡改基准上的取证分数 (%)。

取证方法	Ryu'13 ^[308]	Li'13 ^[309]	Silva'15 ^[306]	Emam'16 ^[310]	Pun'18 ^[311]	Bi'18 ^[312]	Wu'18 ^[313]	Zhong'20 ^[314]	Cozzolino+LIR
Precision	95.02	58.05	88.02	-	91.07	90.55	44.59	75.61	96.75
Recall	88.15	92.26	89.72	-	90.21	91.66	31.60	74.13	92.44
F1	91.50	71.21	88.95	84.91	90.33	91.07	37.11	74.82	93.62

本表中对比方法的分数摘录自文献^[314]。

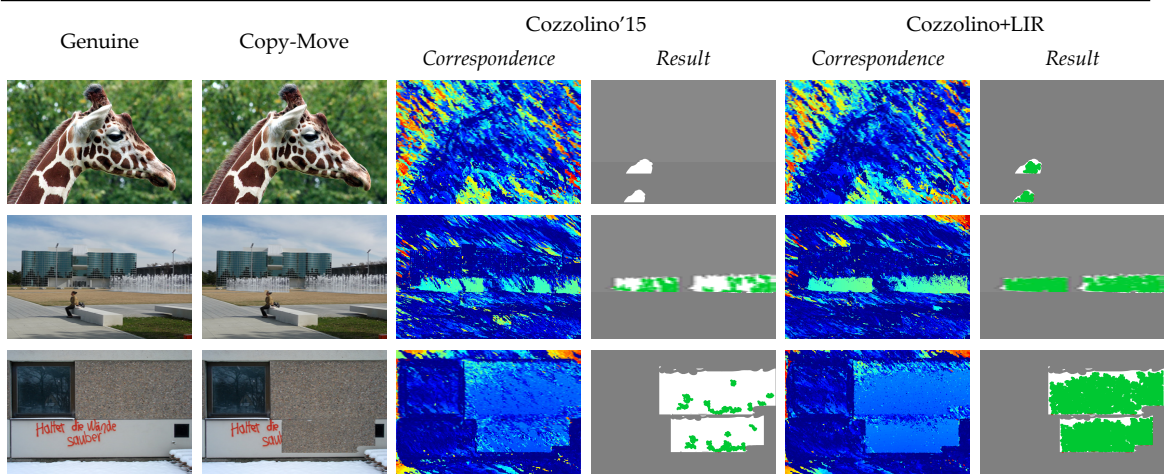


图 4.8 不同检测方法在复制-缩放-粘贴稳健性实验中的结果示例。

表 4.6 复制-缩放-粘贴稳健性实验中引入本章方法前后的取证分数和匹配性能。

取证方法	Cozzolino'15 ^[217]	Cozzolino+LIR	引入本章方法的增益率
Precision	84.47	90.86	7.56
Recall	46.06	63.32	37.47
F1	56.15	71.18	26.77
平均匹配数	144519.75	270730.23	87.33

Zhong'20^[314]。

图 4.8 给出了复制-缩放-粘贴的取证示例，相应的取证分数总结在表 4.6 和 4.7 中。在这种缩放场景中，Cozzolino'15 和其 LIR 版本之间的性能差距显著增大，其他稠密方法的性能也严重降低。其中，LIR 的引入提升了召回率（增益率约 37%）和精确率（增益率约 8%），从而实现了更高的 F1 分数（增益率约 27%）。更重要的是，LIR 引入后每张图像的平均匹配数量增加了近 2 倍。与其他稠密方法相比，LIR 的性能优势更为明显。

综合稳健性。我们最后在 FAU 基准上评估了全面攻击下的稳健性，包括对整幅图像的 6 种信号损失攻击和对操作区域的 4 种几何变换攻击。对比方法包括 Cozzolino'15^[217]、Zandi'16^[303]和 Li'19^[304]。为了揭示取证分数的分布特性（尤其是分数的平均和偏差性质），我们在图 4.9 给出了相应的箱线图。

可以观察到，与基础的 Cozzolino'15 相比，所提出的 LIR 版本在实现更高或保持相似的平

表 4.7 多种检测方法在 FAU 复制-缩放-移动稳健性实验中的取证分数（%）。

Method	Ryu'13 ^[308]	Li'13 ^[309]	Emam'16 ^[310]	Pun'18 ^[311]	Bi'18 ^[312]	Wu'18 ^[313]	Zhong'20 ^[314]	Cozzolino+LIR
Precision	< 25	< 25	-	41.48	62.09	34.84	68.05	90.86
Recall	< 20	24.95	-	39.69	< 20	20.12	64.59	63.32
F1	< 20	20.97	< 20	40.51	22.91	25.22	64.67	71.18

本表中对比方法的分数摘录自文献^[314]。

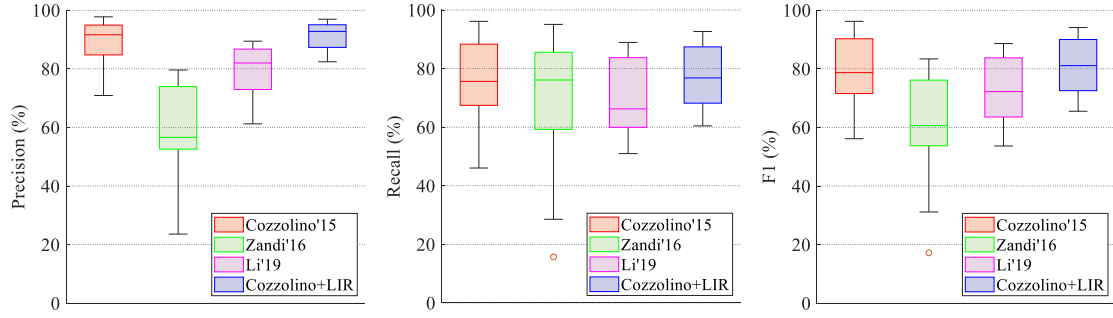


图 4.9 不同取证方法在综合稳健性实验中的取证分数线箱图。

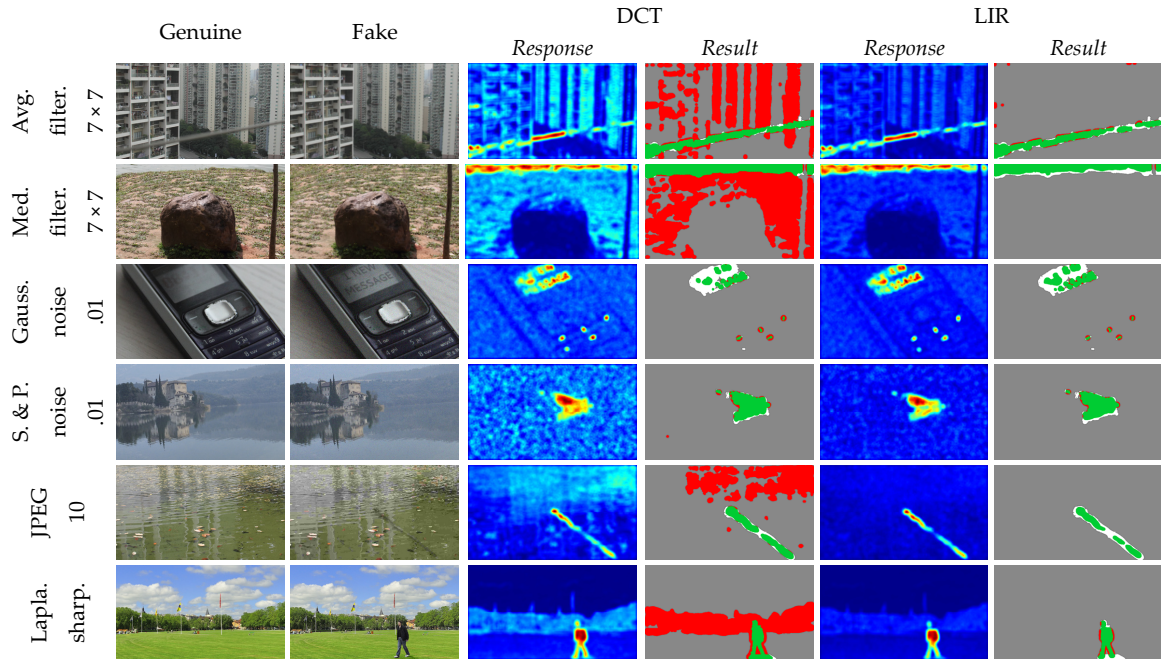


图 4.10 不同感知哈希方法用于内容认证的结果示例。

均分数的同时，表现出明显较低的分数的偏差。而对于 Zandi'16 和 Li'19 来说，所提出的 LIR 版本在分数的平均和偏差方面的优势更为明显。这些共性现象表明，通过引入 LIR，取证算法的稳定性得到了显著改善（特别是在缩放方面），同时不牺牲精确性的平均水平。

上述实验证据支持了我们对几何变换稳健性的理论预期，验证了 LIR 在被动取证场景中的实用性。我们想要指出的是，上述 LIR 应用是非常直接的，通过结合稠密尺度选择方法可以实现更精细的缩放不变性设计。

4.5.4 主动取证：感知哈希

感知哈希是一种用于内容认证的主动取证方法，其主要思想是生成反映图像内容的二进制哈希序列，通过哈希比较实现内容认证。显然这要求，图像表征应该对内容保持的操作具有稳

表 4.8 不同感知哈希方法在稳健性实验中的取证分数 (%)。

取证方法	DCT ^[315,316]			LIR		
	Precision	Recall	F1	Precision	Recall	F1
干净	87.24	77.7	79.85	89.03	71.97	76.95
高斯噪声 0.01	87.02	74.81	77.8	88.63	70.13	75.46
高斯噪声 0.02	86.74	74.2	77.21	87.8	70.05	74.68
椒盐噪声 0.01	86.38	77.83	79.52	88.61	71.88	76.68
椒盐噪声 0.02	85.73	77.66	78.97	88.29	71.65	76.35
均值滤波 7×7	55.43	75.33	53.43	78.8	68.09	68.06
均值滤波 14×14	37.87	74.02	38.45	52.89	65.55	50.18
高斯滤波 7×7	59.49	75.48	56.72	80.43	68.73	69.41
高斯滤波 14×14	49.86	74.97	49.03	71.47	68.17	63.35
中值滤波 7×7	60.61	74.91	57.84	82.43	68.51	70.17
中值滤波 14×14	44.72	69.49	43.13	66.83	61.51	56.72
JPEG 压缩 10	84.22	76.08	76.28	88.61	71.02	76.15
JPEG 压缩 5	77.12	75.57	69.86	87.55	70.88	75.29
拉普拉斯锐化	55.36	80.17	54.55	78.76	72.79	69.64
分数平均值 $\uparrow$	68.41	75.59	63.76	80.72	69.35	69.94
分数标准差 $\downarrow$	17.53	2.36	14.32	10.26	2.89	7.88

健性，同时对不同的视觉内容具有判别性，且哈希序列具有足够紧凑。上述要求与我们在 4.1 节中的分析一致。

算法设计。目前，最先进的感知哈希算法通常遵循一个综合了稀疏和稠密特征的通用框架：稀疏特征（如 SIFT）用于几何校正，稠密特征（如稠密 DCT）用于篡改区域定位^[315,316]。然而，像 DCT 这样的稠密特征在表征能力和哈希紧凑性两方面存在缺陷。在某些信号损失场景，如模糊和压缩，其取证结果相当不稳定。受此启发，我们以 LIR 替换这类稠密 DCT，以改善其信号损失稳健性并缩短哈希序列。我们的 LIR 实现细节如下，考虑位置参数 (u, v) 在图像网格上的 8 像素间隔采样（为了哈希紧凑性），在公式 (4.14) 中设置最大阶数 $K = 3$ ，基于无穷范数，使得 LIR 在尺度空间中每个位置 (u, v, w) 上具有 16 维特征。由于存在几何校正，尺度参数 w 仅从 8 增加到 12 并含 3 个采样点，引入跨尺度的平均池化。在实践中，哈希距离比较的阈值很重要，公平性起见我们使用大津算法^[317]自适应地确定所有比较方法的阈值。

哈希稳健性。我们取证基准图像集 RTD^[318]上进行实验，该基准包含修补、拼接、复制-粘贴操纵以及其他微妙的手工润饰。由于感知哈希的固有特性（主动取证），在这些干净图像上直接取证不具有挑战性。因此，实验主要考虑在全局信号损失下更具挑战性的协议。对比方法包括基础的 SIFT-DCT 框架和我们的 SIFT-LIR 版本。请注意，该基础框架涵盖了几种最先进的感知哈希算法，例如 Zhang'20^[315]、Wang'15^[316]、Hao'21^[319] 和 Biswas'20^[320]。

图 4.10 给出了感知哈希内容认证的示例，相应的取证分数总结在表 4.8 中。可以观察到，即使在含有几何校正的取证场景中 LIR 仍然表现出一定优势。从响应图可以看出，某些攻击下

DCT 很难在实现稳健性的同时保持判别性，倾向于在内容保持处理后的区域中给出误报。与基础的 SIFT-DCT 框架相比，LIR 在干净和噪声图像上提供了类似的取证分数，同时在模糊、压缩和锐化方面表现出更强的稳健性。分数的平均值和标准差层面也凸显了 LIR 的性能优势，特别是在精确率（与误报有关）方面。总体而言，LIR 在整体定位精确性（即平均值）以及在不同攻击下的稳定性（即标准差）方面表现更好。

哈希紧凑性。每个位置上的 DCT 和 LIR 特征的维度分别为 32 和 16，这意味着在存储和传输成本方面实现了近 50% 的节省。

上述实验证据支持了我们对信号损失稳健性和表征紧凑性的理论预期，验证了 LIR 在主动取证场景中的有用性。

4.6 本章小结

本章对面向取证场景的局部不变量进行了系统性研究，将全局不变量的基本理论推广到更广义的局部假设，提出了名为“局部不变表征 LIR”的新方法。

本章工作在理论层面的要点如下：

- 定义方面，从基本公式上推广了矩和矩不变量理论，给出了具有局部表征能力的新定义 LIR（4.3.1 节）。
- 性质方面，形式化讨论了 LIR 在几何不变性、等变性和协变性方面的核心性质，作为形成不变表征的指导原则（4.3.2 节）。
- 计算方面，推导出了 LIR 的高效计算策略，具有低误差、常数阶复杂性以及基函数通用性的特点（4.3.3 节）。

本章工作在实践层面的要点如下：

- 给出了 LIR 的实践指南，旨在帮助研究者更好地将上述理论付诸于具体应用场景，特别是明确了在取证任务中的设置原则（4.4 节）。

本章工作在教育层面的要点如下：

- 仿真实验方面，即稠密模式检测和匹配，LIR 相对于现有稠密局部表征的先进方法实现了更优的整体性能，特别是在挑战性的几何或信号劣化场景（4.5.2 节）。
- 实际应用方面，即复制-粘贴篡改检测（被动取证）和感知哈希（主动取证），LIR 相对于该领域先进取证方法实现了取证分数和效率的提升（4.5.3 和 4.5.4 节）。

本章工作已总结成文：

- Shuren Qi, Yushu Zhang[✉], Chao Wang, et al. A principled design of image representation: Towards forensic tasks. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 2022, 45(5): 5337-5354.

第五章 探索层次不变量

针对不变表征方法在大规模视觉任务中的适用性问题，本学位论文的第三个课题是探索层次不变量，以在更大规模假设下更好地权衡不变性和判别性。本章以重新审视卷积网络中的典型模块开端，在理论（蓝图、定义、性质和计算）、实践（设置原则和数据自适应）和应用（经典和可信视觉任务）层面介绍所提出的层次不变量新方法。

5.1 引言

稳健和可解释的图像表征是迈向可信人工智能的关键技术之一。在此方面，一条成功路径是在图像表征的设计中嵌入任务所需的不变性结构。然而，从历史的角度，不变性和判别性分别在手工（如 SIFT）和学习（如 CNN）研究范式上近乎独立的发展，仍然缺乏良好兼顾不变性和判别性的一般性设计，这限制了稳健和可解释的图像表征在更大规模视觉问题中的应用。

5.1.1 研究动机

面向大规模的可信场景，最新进展试图给出一些更先进的不变性设计以同时满足判别性、稳健性和解释性需求。

为了增强手工表征的判别性，研究者尝试引入学习表征的成功经验，特别是深度级联和过完备字典。在这方面最具代表性的工作是散射网络^[321]（Scattering Networks），作者将经典的小波变换拓展为深度级联的过完备表征。与典型的 CNN 不同：1) 框架上，卷积层由固定的小波基定义，采用基于模运算的非线性，没有中间的池化操作，表征不需学习；2) 表征性质上，提供了平移等变性以及相对于非线性形变的稳健性。后续理论工作进一步探讨了多种几何不变性^[323]、更一般的数学形式化^[279]、以及在改善先进 CNN 技术的效率、解释性、稳健性方面的潜力^[324]等。应用方面，在音频^[325]，图像^[321,323,324]和图^[326]数据的多种任务中，其中的一些甚至是跨学科的^[327,328]，提供了有竞争力的结果。鉴于类似的设计目标和实现路径，我们将散射网络作为本章工作的主要竞争者。

为了增强学习表征的稳健性和解释性，研究者尝试引入手工表征的成功经验，特别是固有的不变/等变结构。在这方面最具代表性的工作是等变网络^[322]（Equivariant Networks），作者将经典的卷积变换推广为对称群上的新定义。与典型的 CNN 不同：1) 框架上，卷积变换是学得的但具有先验启发的新结构，一般涉及卷积的重参数化以实现可控的对称性；2) 表征性质上，通常旨在实现平移和旋转联合等变性。后续理论工作进一步从多种数学理论出发，包括可操纵滤波^[329]、谐波分析^[330]、尺度空间^[331]、李群^[332]、B 样条插值^[333]，完善了旋转^[329,330,334]、翻转^[329]、

表 5.1 本章方法与相关方法的概念性对比。

方法	经典不变量 ^[16]	经典卷积网络 ^[21]	散射网络 ^[321]	等变网络 ^[322]	本章方法
判别性		✓	✓	✓	✓
稳健性	✓		✓	✓	✓
可解释性	✓		✓	✓	✓
高效性	✓				✓

缩放^[331,335]以及它们的联合^[336]的等变性。应用方面，在底层视觉任务^[337]，特别是具有几何先验的科学发现^[338,339]中，扮演关键性角色。鉴于类似的设计目标但在学习路径上，我们将等变网络作为本章工作的次要竞争者。

动机。尽管从非常不同的理论出发，上述方法在实现效率和表征能力方面表现出一些共同的问题。对于上述方法一个朴素的技术阐释是^[336]：输入特征图与同一基函数的多个对称性版本进行卷积以获得多通道特征，其中输入上的失真（比如旋转）将对应于通道之间的循环移位，而不变特征形成于跨通道的池化。1) 实现效率方面，这种并行设计将导致计算规模成倍扩张（取决于对称性版本的采样率），特别是对于学习表征而言，新增的可学习参数也使得训练的难度增加。2) 表征能力方面，基函数对称性版本的离散采样引起不变性和判别性的棘手权衡：提高采样率意味着更好的不变性，但高昂计算成本将限制模型的整体规模，进而抑制判别性的提高。

5.1.2 核心结论

本章围绕“稳健和可解释的图像层次表征”课题展开，尝试给出更系统性的研究结果。如表 5.1 总结的，我们的方法可以被考虑为经典不变量^[16]和经典 CNN^[21]在稳健性和判别性上两个极端的有效折衷；与最新的散射网络^[321]和等变网络^[322]相比，我们的方法具有更简洁的不变性结构，其中等变性是连续的、直接的、并且跨层成立，没有复杂的对称性采样、并行结构和跨通道池化。因此，我们的方法表现出更好的实现效率，并能更灵活的扩大网络规模（更深或更宽）以增加表征容量，在效率、不变性和判别性的具体实践中实现了良好平衡。

本章工作是对矩和矩不变量研究在理论、实践和应用层面上的扩展。

理论。我们将全局和局部不变量的基本理论统一推广到层次假设，称为层次不变表征（Hierarchical Invariant Representation, HIR）。从层次不变性的理想蓝图出发，我们形式化地描述了如何以简洁的层次结构生成过完备的不变量，以更好权衡不变性和判别性。注意，HIR 在每一中间层上都表现出连续且直接的平移、旋转和翻转等变性，尽我们所知目前没有竞争算法能实现这一点。我们还提供了 HIR 的一些快速、精确数值实现，这些策略对于任意基函数通用。

实践。我们探索了层次不变性理论的实践灵活性，其涵盖多种具体设计以更好适应给定视觉任务。我们将该理论被具体化为一类网络，涉及网络拓扑、层和参数的设置原则，特别是设

计了新颖的频率池化方法以满足解释性、不变性和判别性的共同要求。我们还提供了数据自适应策略，基于 HIR 的过完备特性，对给定任务更具判别性的特征可以通过类似网络架构搜索的方式形成。

应用。我们在多个仿真实验和现实应用中验证了 HIR 的有效性。针对典型纹理、数字和寄生虫图像集的分类实验，HIR 展现出不同问题规模和几何变体下先进的精度、不变性以及效率。针对大规模、现实场景的数字取证任务，即对抗性扰动^[340]和人工智能生成内容^[341]（Artificial Intelligence Generated Content, AIGC）取证，HIR 的直接应用也展现出深度学习时代有竞争力的判别性水平。

5.2 预备和基础

完整性起见，我们简要回顾的矩和矩不变量的全局和局部定义以及几何对称性质，涵盖本章预备和基础的概念、符号和定义（详见 2.2 和 4.3 节）。

5.2.1 全局和局部表征

一般而言，经典的矩和矩不变量是图像的全局表征，其理论建立于如下定义^[16]：

$$\langle f, V_{nm} \rangle = \iint_D V_{nm}^*(x, y) f(x, y) dx dy, \quad (5.1)$$

其中， f 为图像函数， V_{nm} 为阶数 $(n, m) \in \mathbb{Z}^2$ 的基函数，二者均定义于 $D \in \mathbb{R}^2$ ， $*$ 表示复共轭。注意，公式（5.1）中 f 和 V_{nm} 的定义域具有相同/类似的位置和尺度，意味着表征信息的全局性。

考虑图像的稀疏性和几何先验，两个典型的约束经常施加于 V_{nm} 的显式定义，即正交性和旋转不变性，导出如下极坐标形式：

$$\langle f, V_{nm} \rangle = \iint_D R_n^*(r) A_m^*(\theta) f(r, \theta) r dr d\theta, \quad (5.2)$$

其中 $V_{nm}(\underbrace{r \cos \theta}_x, \underbrace{r \sin \theta}_y) \equiv V_{nm}(r, \theta)$ 被分离为角向基函数 $A_m(\theta) = \exp(jm\theta)$ ($j = \sqrt{-1}$) 和径向基函数 R_n 的积， R_n 需满足加权正交条件 $\int_0^1 R_n(r) R_{n'}^*(r) r dr = \frac{1}{2\pi} \delta_{nn'}$ 。注意，公式（5.2）中基函数 $V_{nm} = R_n A_m$ 在 D 上正交，且 $\langle f, V_{nm} \rangle$ 的幅度对 f 的旋转操作不变。

在我们最近的工作中，矩和矩不变量被拓展为图像的局部表征，其理论建立于如下定义^[238]：

$$\langle f, V_{nm}^{uvw} \rangle = \iint_D R_n^*(r') A_m^*(\theta') f(x, y) dx dy, \quad (5.3)$$

其中，新的基函数 V_{nm}^{uvw} 引入了位置参数 (u, v) 和尺度参数 w ，可以被理解为全局基函数 V_{nm} 的平移和缩放版本，具有如下坐标变换：

$$\begin{cases} r' = \frac{1}{w} \sqrt{(x-u)^2 + (y-v)^2} \\ \theta' = \arctan(\frac{y-v}{x-u}) \end{cases}, \quad (5.4)$$

其中，定义域是以 (u, v) 为圆心、 w 为半径的圆形区域： $D = \{(x, y) : (x-u)^2 + (y-v)^2 \leq w^2\}$ 。注意，公式 (5.3) 允许 V_{nm}^{uvw} 的定义域建立在与 f 不同的位置和尺度上，意味着表征信息的局部性。同时，经典定义 (5.2) 实际上是新定义 (5.3) 的一个特例，其中 $(u, v) = (0, 0)$ ， $w = 1$ 。

5.2.2 不变性、等变性和协变性

从不变性、等变性和协变性术语和局部表征 (5.3) 出发，可以验证表征对图像的平移、旋转、翻转和缩放具有如下结论。

由图像平移可以导出：

$$\langle f(x + \Delta x, y + \Delta y), V_{nm}^{uvw}(x, y) \rangle = \langle f(x, y), V_{nm}^{(u+\Delta x)(v+\Delta y)w}(x, y) \rangle, \quad (5.5)$$

其中， $(\Delta x, \Delta y)$ 为平移向量。可见同样的 $(\Delta x, \Delta y)$ 出现在表征参数 (u, v) 中，意味着表征对平移具有等变性。由于平移等变性的成立，下面关于旋转、翻转和缩放的讨论将只考虑中心对齐的情况，即设定 $(u, v) = (0, 0)$ 而不失一般性。

由图像旋转可以导出：

$$\langle f(r, \theta + \phi), V_{nm}^{uvw}(r', \theta') \rangle = \langle f(r, \theta), V_{nm}^{uvw}(r', \theta') \rangle A_m^*(-\phi), \quad (5.6)$$

其中限制位置参数 $(u, v) = (0, 0)$ ， ϕ 为旋转角度，可见同样的 ϕ 出现在表征系数的相位域中，意味着表征对旋转具有协变性。易知，当取表征系数的幅度时上述协变性将特化为不变性：

$$|\langle f(r, \theta + \phi), V_{nm}^{uvw}(r', \theta') \rangle| = |\langle f(r, \theta), V_{nm}^{uvw}(r', \theta') \rangle|。$$

由图像翻转可以导出：

$$\langle f(r, -\theta), V_{nm}^{uvw}(r', \theta') \rangle = (\langle f(r, \theta), V_{nm}^{uvw}(r', \theta') \rangle)^*, \quad (5.7)$$

其中限制位置参数 $(u, v) = (0, 0)$ ， $f(r, -\theta)$ 为纵向翻转版本。可见类似于旋转的协变性，纵向翻转同样仅影响表征系数的相位。对于其他朝向的翻转，结论是类似的，可由旋转和纵向翻转的复合导出。易知，当取表征系数的幅度时可以导出旋转和翻转联合不变性。

由图像缩放可以导出：

$$\langle f(sx, sy), V_{nm}^{uvw}(x, y) \rangle = \langle f(x, y), V_{nm}^{uv(ws)}(x, y) \rangle, \quad (5.8)$$

其中限制位置参数 $(u, v) = (0, 0)$, s 为缩放因子, 可见同样的 s 出现在表征参数 w 中, 意味着表征对缩放具有协变性。

对于 $(u, v) \neq (0, 0)$ 时的表征性质, 可由平移与中心对齐旋转、翻转和缩放的复合导出。因此, 局部表征 (5.3) 的幅度在 (u, v) 域中对于任何平移、旋转和翻转具有联合等变性, 同时在 w 域中对于任何缩放具有协变性。

5.3 层次不变量理论

本节介绍 HIR 的理论方面。我们首先以不变性视角重新审视 CNN 表征的典型模块, 给出了实现层次不变性的形式化蓝图。我们随后定义了新的模块及其级联以满足层次不变性蓝图, 同时提供了表示性质的分析和高效的数值实现。最后, 作为本章理论的总结, 我们讨论了层次不变性与现有类似概念之间的批判和发展关系。

5.3.1 层次不变性的蓝图

我们的目标是实现一种类似 CNN 的层次表征, 其中图像信息能够以几何可控的方式通过每个中间层, 并在最后一层以紧凑的设计生成具有足够信息量的不变特征。本章将这种理想的表征结构称为层次不变性。

受上述目标启发, 接下来我们反思了 CNN 的几种典型模块, 并形式化了这些模块在层次不变性中的理想蓝图, 如图 5.1 所示。

形式化 5.1: 层次不变性蓝图。 特征图 (含输入图像) 的集合记为 $X(\Omega, H) \triangleq \{M(i, j; k) : \Omega \rightarrow H\}$, 这里 $(i, j) \in \Omega$ 为定义域 (离散网格), $\mathbb{C}^k \in H$ 为通道 (复数值域); Ω' 和 H' 分别标记 Ω 和 H 经某种运算后的变体, 例如下采样版本; $\mathfrak{G}$ 为一个群建模 Ω 上所有感兴趣的对称性。我们为层次不变性形式化如下模块:

- 卷积层 $\mathbf{C} : X(\Omega, H) \rightarrow X(\Omega', H')$ 通过卷积运算捕捉局部特征。这里, $\mathbf{C}$ 的几何设计原则是对称群 $\mathfrak{G}$ 上的协变性, 即存在可以预测的 $\mathbf{g}'$ 使得 $\mathbf{C}(\mathbf{g}M) = \mathbf{g}'\mathbf{C}(M)$ 对于任意 $\mathbf{g} \in \mathfrak{G}$ 和 $M \in X$ 成立, 如果 $\mathbf{g}' = \mathbf{g}$ 则特化为等变性。
- 非线性层 $\mathbf{S} : X(\Omega, H) \rightarrow X(\Omega, H')$ 在特征中引入非线性以改善判别性, 其具有逐元素的非线性运算 σ 使得 $(\mathbf{S}M)(i, j) = \sigma(M(i, j))$, 其中卷积层和非线性层的级联 $\mathbf{S} \circ \mathbf{C}$ 仍然保持协变性由于 σ 的逐元素作用。
- 局部池化层 $\mathbf{P} : X(\Omega, H) \rightarrow X(\Omega', H)$ 下采样特征的平面维度以降低运算复杂性, 其中

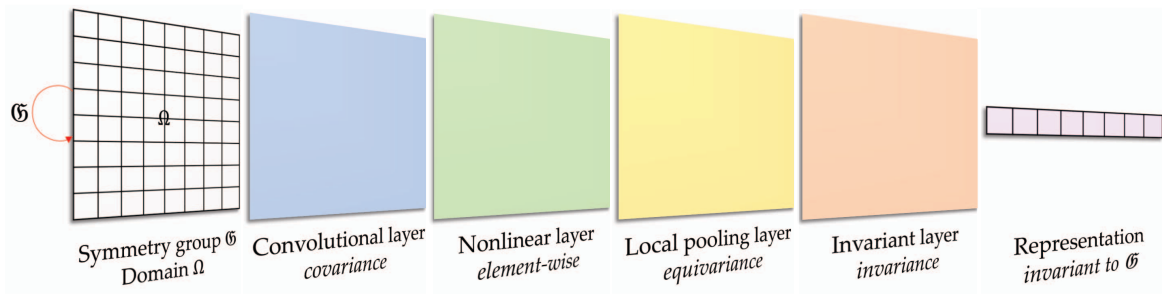


图 5.1 层次不变性蓝图：图像信息能够以几何可控的方式通过每个中间层，并在最后一层以紧凑的设计生成具有足够信息量的不变特征。

$\Omega' \subseteq \Omega$ 。这里， $\mathbf{C}$ 的几何设计原则是对由 $\mathbf{C}$ 产生的任意 $\mathbf{g}'$ 的近似等变性，即 $\mathbf{P}(\mathbf{g}'M) \simeq \mathbf{g}'\mathbf{P}(M)$ ，以允许后续的卷积层 $\mathbf{C}$ 继续捕捉协变特征，其中级联 $\mathbf{P} \circ \mathbf{S} \circ \mathbf{C}$ 仍然保持协变性。

- 不变层 $\mathbf{I} : X(\Omega, H) \rightarrow Y$ 获得最终的向量化表征通过某种平面维度上的全局池化。这里， $\mathbf{I}$ 的几何设计原则是对称群 $\mathcal{G}$ 上的不变性，即 $\mathbf{I}(\mathbf{g}M) = \mathbf{I}(M)$ 对于任意 $\mathbf{g} \in \mathcal{G}$ 和 $M \in X$ 成立。

从层次不变性蓝图出发，我们可以通过上述模块的有序级联构造 HIR。

5.3.2 层和路径的定义

从 5.2 节矩和矩不变量理论中，我们将给出层 $\mathbf{C}$ 、 $\mathbf{S}$ 、 $\mathbf{P}$ 和 $\mathbf{I}$ 的定义以满足层次不变性蓝图。

定义 5.1：卷积层。 对于定义在 $\Omega = \{1, 2, \dots, \#_i\} \times \{1, 2, \dots, \#_j\}$ 和 $H = \mathbb{C}^{\#_k}$ 上的输入特征图 $M(i, j; k)$ ，卷积层 $\mathbf{C}$ 被定义为逐通道的局部协变表征，基于公式 (5.3) 和 (5.4)：

$$\mathbf{C}M \triangleq \langle M, V_{nm}^{uvw} \rangle = M(i, j; k) \otimes (H_{nm}^w(i, j))^T, \quad (5.9)$$

其中， $\otimes$ 为 Ω 上的卷积运算， $(\cdot)^T$ 为矩阵转置， H_{nm}^w 是卷积核具有如下定义：

$$H_{nm}^w(i, j) = \{h_{nm}^{uvw}(i, j) : u, v = w, (i, j) \text{ s.t. } D_{ij} \cap D \neq \emptyset\}, \quad (5.10)$$

其中， h_{nm}^{uvw} 为如下积分值：

$$h_{nm}^{uvw}(i, j) = \iint_{D_{ij} \cap D} (V_{nm}^{uvw}(x, y))^* dx dy, \quad (5.11)$$

其中，以 (i, j) 为中心的像素区域记为 $D_{ij} = \{(x, y) \in [i - \frac{\Delta_i}{2}, i + \frac{\Delta_i}{2}] \times [j - \frac{\Delta_j}{2}, j + \frac{\Delta_j}{2}]\}$ 。

评注：在定义 5.1 中，卷积层 $\mathbf{C}$ 仅由 w 和 (n, m) 指定，分别控制 $\mathbf{C}M$ 的表征尺度和表征频率属性。注意，由公式 (5.9) 定义的卷积层 $\mathbf{C}$ 将直接继承 5.2.2 节介绍的表征性质 (5.5) ~ (5.8)，可被视群 $\mathcal{G}$ 上的线性协变层，该群建模了对 Ω 上的所有平移、旋转、翻转和缩放对称性。为了

方便后续讨论，我们标记 $\mathfrak{G} = \mathfrak{G}_1 \times \mathfrak{G}_2$ ，其中 $\mathfrak{G}_1$ 是平移/旋转/翻转对称群， $\mathfrak{G}_2$ 是缩放对称群。

定义 5.2: 非线性层。 对于定义在 $\Omega = \{1, 2, \dots, \#_i\} \times \{1, 2, \dots, \#_j\}$ 和 $H = \mathbb{C}^{\#_k}$ 上的输入特征图 $M(i, j; k)$ ，非线性层 $\mathbf{S}$ 被定义为逐通道的幅度运算：

$$\mathbf{S}M = \sigma(M(i, j)) \triangleq |M(i, j; k)|, \quad (5.12)$$

其中， $M(i, j; k)$ 具有复数值域，上式可被显式的写为 $\sqrt{(\text{Re}M(i, j; k))^2 + (\text{Im}M(i, j; k))^2}$ 。

评注：基于定义 5.2 和 5.2.2 节，卷积层和非线性层的级联将满足平移、旋转和翻转联合等变性，即 $\mathbf{S} \circ \mathbf{C}(\mathbf{g}_1 M) = \mathbf{g}_1 \mathbf{S} \circ \mathbf{C}(M)$ 对于任意 $\mathbf{g}_1 \in \mathfrak{G}_1$ 和 $M \in X$ 成立。因此，由公式 (5.12) 定义的非线性层不仅实现了特征的去线性化，而且还将卷积层 $\mathbf{C}$ 对旋转和翻转产生的协变 $\mathbf{g}'_1$ 即特化为更易处理的等变性 $\mathbf{g}_1$ 。此外，级联 $\mathbf{S} \circ \mathbf{C}$ 保持 $\mathbf{C}$ 的缩放协变性，由于 $\mathbf{S}$ 的逐元素运算，即 $\mathbf{S} \circ \mathbf{C}(\mathbf{g}_2 M) = \mathbf{g}'_2 \mathbf{S} \circ \mathbf{C}(M)$ 对于任意 $\mathbf{g}_2 \in \mathfrak{G}_2$ 和 $M \in X$ 成立。

定义 5.3: 局部池化层。 对于定义在 $\Omega = \{1, 2, \dots, \#_i\} \times \{1, 2, \dots, \#_j\}$ 和 $H = \mathbb{C}^{\#_k}$ 上的输入特征图 $M(i, j; k)$ ，局部池化层 $\mathbf{P}$ 被定义为恒等运算：

$$\mathbf{P}M = M. \quad (5.13)$$

评注：基于相关研究，网络中的下采样操作将不同程度的损伤等变性，即近似 $\mathbf{P}(\mathbf{g}M) \simeq \mathbf{g}\mathbf{P}(M)$ ，特别是对于更大的池化尺度和更深的网络深度，导致计算规模和等变性的权衡。由于 HIR 是非学习的且具有高效的不变性结构，无需典型 CNN 的大规模训练，我们在计算成本可以接受的情况下直接忽略下采样操作，简单设置 $\mathbf{P} = \text{id}$ 。另外，能更好地平衡计算规模和等变性的池化设计也可以用于定义 $\mathbf{P}$ ，详见于 Zhang 的工作^[21]。显然，级联 $\mathbf{P} \circ \mathbf{S} \circ \mathbf{C}$ 具有 $\mathbf{S} \circ \mathbf{C}$ 相同的表征性质。

定义 5.4: 不变层。 对于定义在 $\Omega = \{1, 2, \dots, \#_i\} \times \{1, 2, \dots, \#_j\}$ 和 $H = \mathbb{C}^{\#_k}$ 上的输入特征图 $M(i, j; k)$ ，不变层 $\mathbf{I}$ 被定义为逐通道的全局不变表征，基于公式 (5.1) 和 (5.2)：

$$\mathbf{I}M = \mathcal{I}(\{\langle M(i, j; k), V_{nm}(x_i, y_j) \rangle\}), \quad (5.14)$$

其中， $\mathcal{I}$ 是一个特殊的变换将图像矩映射为全局不变量，相对于感兴趣对称群 $\mathfrak{G}_0 \subseteq \mathfrak{G}_1 \times \mathfrak{G}_2$ 和任意 $M \in X$ 。

评注：在定义 5.4 中，我们并未给出具体的 $\mathcal{I}$ ，以允许接下来讨论的一般性；在 5.4.1 节我们给出 $\mathcal{I}$ 的具体定义，针对本文考虑的应用场景。注意，基于定义 5.1~5.3 以及 5.2.2 节，通过 $\mathbf{P} \circ \mathbf{S} \circ \mathbf{C}$ 及其级联得到的 M 将保持原始图像 f 的几何结构（具体的说是平移、旋转和翻转对称性）。因此， $\mathcal{I}$ （相对于深度特征图 M ）的设计思路是与全局不变量理论（相对于原始图像 f ）

是非常类似的，具有广泛的潜在设计^[16]。

定义 5.5：路径。基于定义 5.1~5.4，我们定义 HIR 的一个路径为 $p = (\lambda_{[1]}, \lambda_{[2]}, \dots, \lambda_{[L]})$ ，其中 $\lambda_{[z]} = (n, m, w)_{[z]}$ 指定排序为 z 的卷积层的参数。沿着路径 p 的 HIR， $\mathcal{R}_p$ ，被定义为如下有序级联运算并具有对应的参数 $p = (\lambda_{[1]}, \lambda_{[2]}, \dots, \lambda_{[L]})$ ：

$$\mathcal{R}_p \triangleq \mathbf{I} \circ \mathbf{P}_{[L]} \circ \mathbf{S}_{[L]} \circ \mathbf{C}_{[L]} \circ \dots \circ \mathbf{P}_{[1]} \circ \mathbf{S}_{[1]} \circ \mathbf{C}_{[1]}. \quad (5.15)$$

评注：通过定义 5.5，我们实际上将矩和矩不变量的全局表征和局部表征理论进一步统一到一个层次表征框架中。注意，不变层 $\mathbf{I}$ 之前的层为表征 $\mathcal{R}_p$ 提供了几何结构保持的特性。这里，全局表征 (5.15) 面向图像级视觉任务，如图像分类；至于像素级视觉任务，如图像分割，我们可以保留空间维度通过去掉最后的不变层。在下一节，我们将详尽分析定义 5.5 的表征性质。

5.3.3 表征性质

在典型 CNN 中，图像信息与学得表征的关系是高度非线性的，难以理解或预测。而对于 HIR，我们可以给出如下形式化的结论关于图像与表征的几何对称性关系，意味着良好的解释性和稳健性。

性质 5.1：平移、旋转和翻转联合等变性。对表征单元 $\mathbf{U} \triangleq \mathbf{P} \circ \mathbf{S} \circ \mathbf{C}$ ，具有任意的卷积层参数 λ ，其任意次级联都满足平移、旋转和翻转联合等变性（忽略边际效应和采样近似），即下列恒等式：

$$\mathbf{U}_{[L]} \circ \dots \circ \mathbf{U}_{[2]} \circ \mathbf{U}_{[1]}(\mathbf{g}_1 M) \equiv \mathbf{g}_1 \mathbf{U}_{[L]} \circ \dots \circ \mathbf{U}_{[2]} \circ \mathbf{U}_{[1]}(M). \quad (5.16)$$

对于任意级联长度 $L \geq 1$ 、任意 $\mathbf{g}_1 \in \mathfrak{G}_1$ 和 $M \in X$ 均成立，其中 $\mathfrak{G}_1$ 为平移、旋转和翻转对称群。

证明：首先，让我们审视一个表征单元 $\mathbf{U}$ 在 $\mathfrak{G}_1$ 上的行为：

$$\begin{aligned} \mathbf{U}(\mathbf{g}_1 M) &= \mathbf{P} \circ \mathbf{S} \circ \mathbf{C}(\mathbf{g}_1 M) \\ &= \mathbf{P} \circ \mathbf{S} \circ \mathbf{g}'_1 \mathbf{C}(M) \\ &= \mathbf{P} \circ \mathbf{g}_1 \mathbf{S} \circ \mathbf{C}(M) \\ &= \mathbf{g}_1 \mathbf{P} \circ \mathbf{S} \circ \mathbf{C}(M) \\ &= \mathbf{g}_1 \mathbf{U}(M), \end{aligned} \quad (5.17)$$

其中第二个等号出自卷积层 $\mathbf{C}$ 对旋转和翻转的协变性 (5.6) 和 (5.7)， $\mathbf{g}'_1$ 是一个可预测的操作作用在 $\mathbf{C}(M)$ 的相位域；第三个等号出自非线性层 $\mathbf{S}$ 对协变性 $\mathbf{g}'_1$ 的特化，这里取幅度运算去除了额外的相位变化，导致单纯的等变性 $\mathbf{g}_1$ ；第四个等号出自局部池化层 $\mathbf{P}$ 的恒等变换，当采用下采样的 $\mathbf{P}$ 时将导致第四个等号变为近似相等。

这里, $\mathbf{U}(\mathbf{g}_1 M) = \mathbf{g}_1 \mathbf{U}(M)$ 意味着表征单元 $\mathbf{U}$ 可视为等变层, 相对于任意 $\mathbf{g}_1 \in \mathfrak{G}_1$ 和 $M \in X$ 成立——换句话说, 对于 $M \in X$ 其上单个 $\mathbf{U}$ 和 $\mathbf{g}_1$ 的操作是可交换; 并且, 标记 $M_{[l]} \triangleq \mathbf{U}_{[l]} \circ \cdots \circ \mathbf{U}_{[1]}(M) = \mathbf{U}_{[l]} M_{[l-1]}$, 有 $M_{[l]} \in X$ 对于 $l \in \{1, 2, \dots, L\}$ 成立。因而, $\mathbf{g}_1$ 可以和 $\mathbf{U}$ 的任意级联交换顺序, 意味着性质 5.1 的正确性。□

性质 5.2: 缩放协变性。对表征单元 $\mathbf{U}$, 我们指定其卷积层的尺度参数为 w 并记为 $\mathbf{U}^w \triangleq \mathbf{P} \circ \mathbf{S} \circ \mathbf{C}^w$, 其任意次级联都满足缩放协变性 (忽略边际效应和采样近似), 即下列恒等式:

$$\begin{aligned} & \mathbf{U}_{[L]}^w \circ \cdots \circ \mathbf{U}_{[2]}^w \circ \mathbf{U}_{[1]}^w(\mathbf{g}_2 M) \\ & \equiv \mathbf{g}_2' \mathbf{U}_{[L]}^w \circ \cdots \circ \mathbf{U}_{[2]}^w \circ \mathbf{U}_{[1]}^w(M) \\ & = \mathbf{g}_2 \mathbf{U}_{[L]}^{ws} \circ \cdots \circ \mathbf{U}_{[2]}^{ws} \circ \mathbf{U}_{[1]}^{ws}(M), \end{aligned} \quad (5.18)$$

对于任意级联长度 $L \geq 1$ 、任意 $\mathbf{g}_2 \in \mathfrak{G}_2$ 和 $M \in X$ 成立, 其中 $\mathbf{g}_2'$ 为可预测操作, 与 $\mathbf{g}_2$ 相关, 可重写为显式形式 $\mathbf{g}_2' \mathbf{U}^w \triangleq \mathbf{g}_2 \mathbf{U}^{ws}$, s 为对应于 $\mathbf{g}_2$ 的图像缩放因子, $\mathfrak{G}_2$ 为缩放对称群。

证明: 首先, 让我们审视一个表征单元 $\mathbf{U}^w$ 在 $\mathfrak{G}_2$ 上的行为:

$$\begin{aligned} \mathbf{U}^w(\mathbf{g}_2 M) &= \mathbf{P} \circ \mathbf{S} \circ \mathbf{C}^w(\mathbf{g}_2 M) \\ &= \mathbf{P} \circ \mathbf{S} \circ \mathbf{g}_2' \mathbf{C}^w(M) \\ &= \mathbf{P} \circ \mathbf{S} \circ \mathbf{g}_2 \mathbf{C}^{ws}(M) \\ &= \mathbf{P} \circ \mathbf{g}_2 \mathbf{S} \circ \mathbf{C}^{ws}(M) \\ &= \mathbf{g}_2 \mathbf{P} \circ \mathbf{S} \circ \mathbf{C}^{ws}(M) \\ &= \mathbf{g}_2 \mathbf{U}^{ws}(M) \\ &= \mathbf{g}_2' \mathbf{U}^w(M), \end{aligned} \quad (5.19)$$

其中第二个等号出自卷积层 $\mathbf{C}$ 对缩放的协变性 (5.8), $\mathbf{g}_2'$ 是一个可预测的操作同时作用于 $\mathbf{C}^w(M)$ 的 Ω 域 (即同样的缩放 $\mathbf{g}_2$) 和参数 w 域 (即因子 s); 第四个和第五个等号分别出自非线性层 $\mathbf{S}$ 的逐元素性和局部池化层 $\mathbf{P}$ 的恒等性。

这里, $\mathbf{U}^w(\mathbf{g}_2 M) = \mathbf{g}_2' \mathbf{U}^w(M)$ 意味着表征单元 $\mathbf{U}^w$ 可视为协变层, 相对于任意 $\mathbf{g}_2 \in \mathfrak{G}_2$ 和 $M \in X$ 成立——换句话说, 对于 $M \in X$ 其上单个 $\mathbf{U}^w$ 和 $\mathbf{g}_2$ 的操作是可交换, 但需改变尺度参数为 ws ; 并且有 $M_{[l]} \in X$ 对于 $l \in \{1, 2, \dots, L\}$ 成立。因而, $\mathbf{g}_2$ 可以和 $\mathbf{U}^w$ 的任意级联交换顺序同时改变其尺度参数为 ws , 意味着性质 5.2 的正确性。□

性质 5.3: 层次不变性。对表征单元 $\mathbf{U}$ 的任意次级联, 设计感兴趣对称群 $\mathfrak{G}_0 \subseteq \mathfrak{G}_1 \times \mathfrak{G}_2$ 的全局不变映射 $\mathcal{I}$ 在实践中是可行的, 由于性质 5.1 和 5.2 保证的输入图像和深度表征 (任意层) 之间可预测的几何对称性关系。更具体的说, 基于定义 5.4, 我们假设存在 $\mathcal{I}$ 使得 $\mathbf{I}(\mathbf{g}_0' M) = \mathbf{I}(M)$

对于任意 $\mathbf{g}_0 \in \mathfrak{G}_0$ 和 $M \in X$ 成立，其中 $\mathbf{g}'$ 为相对于 $\mathbf{g}$ 和 $\mathbf{U}$ 的可预测操作，即不变性在单层上成立，则有不变性：

$$\mathbf{I}(\mathbf{g}'_0 M)_{[L]} \equiv \mathbf{I}M_{[L]}, \quad (5.20)$$

对于任意级联长度 $L \geq 1$ 成立。

证明：我们可将公式 (5.20) 重写为：

$$\begin{aligned} \mathbf{I}(\mathbf{g}_0 M)_{[L]} &= \mathbf{I} \circ \mathbf{U}_{[L]} \circ \cdots \circ \mathbf{U}_{[2]} \circ \mathbf{U}_{[1]}(\mathbf{g}_0 M) \\ &= \mathbf{I}(\mathbf{g}'_0 \mathbf{U}_{[L]} \circ \cdots \circ \mathbf{U}_{[2]} \circ \mathbf{U}_{[1]}(M)) \\ &= \mathbf{I} \circ \mathbf{U}_{[L]} \circ \cdots \circ \mathbf{U}_{[2]} \circ \mathbf{U}_{[1]}(M) \\ &= \mathbf{I}M_{[L]}, \end{aligned} \quad (5.21)$$

其中第二个等号出自性质 5.1 和 5.2 的保证， $\mathbf{g}_0 \in \mathfrak{G}_0 \subseteq \mathfrak{G}_1 \times \mathfrak{G}_2$ ， $\mathbf{g}'_0$ 与 $\mathbf{g}_1$ 和 $\mathbf{g}'_2$ 相关；第三个等号出自我们的假设， $\mathbf{I}(\mathbf{g}'_0 M) = \mathbf{I}(M)$ 对于任意 $\mathbf{g}_0 \in \mathfrak{G}_0$ 和 $M \in X$ 成立，并且 $M_{[l]} \in X$ 对于任意 $l \in \{1, 2, \dots, L\}$ 成立。 $\square$

5.3.4 快速和精确计算

至此，我们已经形式化了本文的核心成分，即 **HIR** 的定义和性质。在本节，我们将补充介绍 **HIR** 的数值实现，特别是定义 5.1 的快速、精确计算（详见 4.3.3 节）。请注意本节的讨论是非常一般性的，并不对基函数的具体定义进行限制。

定义 5.6：快速计算。让我们考虑引入卷积定理作为定义 5.1 的快速计算，使得公式 (5.9) 的空间域卷积可以转换为如下频率域乘积形式^[238]：

$$\mathbf{C}M = \mathcal{F}^{-1}(\mathcal{F}(M(i, j; k)) \odot \mathcal{F}((H_{nm}^w(i, j))^T)), \quad (5.22)$$

其中， $\mathcal{F}$ 为傅里叶变换， $\odot$ 为逐点乘积。

性质 5.4：复杂性分析。在定义 5.1 中，稠密二维卷积 (5.9) 主导了计算效率。对于定义在 $\Omega = \{1, 2, \dots, \#_i\} \times \{1, 2, \dots, \#_j\}$ 和 $\mathbf{H} = \mathbb{C}^{\#_k}$ 的输入特征图 $M(i, j; k)$ ，假设需要计算给定频率 (n, m) 和通道 k 下的 $\mathbf{C}M$ ，其具有尺度参数 $w \in S_w$ ，记特征图采样数为 $\#_{ij} = \#_i \#_j$ 、尺度采样数为 $\#_w = |S_w|$ 。通过定义 5.6 和快速傅里叶变换，我们可以以乘法复杂度 $\mathcal{O}(\#_w \#_{ij} \log \#_{ij})$ 实现上述计算，与之相对定义 5.1 的直接计算需要 $\mathcal{O}(\#_w \#_{ij} w_{\max}^2)$ ，其中 $w_{\max}$ 为 S_w 中的最大尺度。注意，去除相同项后上述复杂度中平方级和对数级增长的巨大差异，即当 $w_{\max}^2 > \log \#_{ij}$ 时，定义 5.6 将表现出更优的效率。

定义 5.7: 精确计算。让我们考虑引入高精度数值积分作为定义 5.1 的精确计算, 使得公式 (5.11) 的二维连续积分可以近似为如下求和形式^[238]:

$$h_{nm}^{uvw}(i, j) \simeq \sum_{(a,b) \in S_{ab}} c_{ab} (V_{nm}^{uvw}(x_a, y_b))^* \frac{\Delta i \Delta j}{w^2}, \quad (5.23)$$

其中, 数值积分采样集合 S_{ab} 编码了积分点 $(x_a, y_b) \in D_{ij}$ 以及相应的积分权重 c_{ab} , 它们由特定的数值积分策略指定。

性质 5.5: 精确性分析。在定义 5.1 中, 复杂函数连续积分 (5.11) 主导了计算精度。假设需要计算给定频率 (n, m) 和位置 (u, v) 的 h_{nm}^{uvw} , 记积分采样点数量为 $\#_{ab} = |S_{ab}|$ 。基于定义 5.7 的计算表现出 $O((\frac{\Delta i \Delta j}{w^2})^{\#_{ab}+1})$ 的近似误差, 与之相对定义 5.1 的直接计算 (零阶近似) 误差为 $O((\frac{\Delta i \Delta j}{w^2})^2)$ 。注意, 当一个像素区域内的采样点多于一个时, 即 $\#_{ab} > 1$, 定义 5.7 将表现出更优的精度。

5.3.5 与相关工作的对比

作为表 5.1 的进一步解释, 有必要总结本节并强调与典型相关工作的理论关系。

- **经典不变量。**我们的工作推广了这一理论, 将其全局和局部理论统一到一个层次不变表征理论中。更具体的, 我们基于局部不变量理论^[238]形式化了层 **C**、**S** 和 **P** 的具体定义 (定义 5.1 ~ 5.3), 特别论证了等变/协变性可以在特定的级联下跨层保持 (性质 5.1 和 5.2)。而且, 我们还基于全局不变量理论^[16]形式化了层 **I** 的具体定义 (定义 5.4), 特别是论证了图像域全局不变性的成功经验可以被直接地用于等变/协变的深度表征域 (性质 5.3)。在我们的层次不变表征框架 (定义 5.5) 下, 经典全局^[16] 和局部^[238] 不变量理论可视为特例, 即 **I** f 和 **I** $\circ$ **S** $\circ$ **C** f 。
- **经典卷积网络。**我们的工作具有类似的深度级联框架, 但在几何对称性方面具有更好性质, 能够产生稳健和可解释的图像表征。更具体的, 我们在不变量的构造中引入了 CNN 在判别性方面的成功经验^[21], 即深度级联和过完备表征。更重要的是, 我们批判性分析了 CNN 的基本模块 (形式化 5.1), 使得提出层次表征相对于原始图像表现出完全透明的几何对称性关系 (性质 5.1~5.3), 在信任攸关任务中可作为高度非线性的黑盒 CNN 的有效替代。
- **散射网络。**我们的工作在设计上同样基于显式变换 (小波) 的深度级联^[321], 因而概念上与本章工作相类似。然而, 从小波基构造旋转不变性是复杂的, 需要多朝向小波的并行卷积和跨通道池化, 提高朝向采样将导致复杂度将指数级提高。而我们的方法受益于经典不变量理论, 旋转等变性是连续 (任意朝向) 且直接 (无需池化) 的 (性质 5.1), 具有更优的效率并容易扩大网

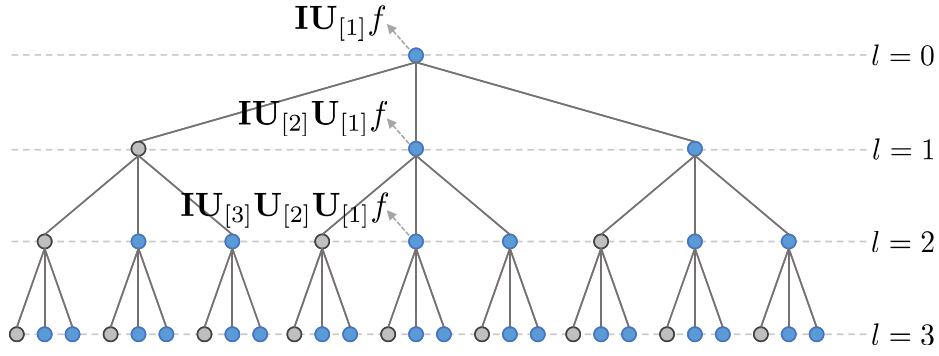


图 5.2 具有 $\mathfrak{G}_1$ 不变性的单尺度实践：树状网络编码了一组 HIR 路径，其中蓝色和黑色节点分别表示表征单元（具有不同参数）和恒等函数，连线表示节点之间的级联关系。

络规模以改善表征容量。

- **等变网络。**我们的工作是非学习的，同时在实现连续和联合不变性方面更紧凑。作为次要竞争者，等变网络在设计上同样由群论保障^[322]，因而概念上与本章工作相类似。然而，等变网络中的卷积层是学得，导致不同程度的数据依赖。特别是其具有和散射网络类似的并行结构，导致指数级复杂性和优化困难。尽管等变网络是一种非常广义的设计，但就连续和联合不变表征（性质 5.1~5.3）而言，提出方法具有更优的效率并容易扩大网络规模以改善表征容量。

5.4 层次不变性实践

本节介绍 HIR 的实践方面。我们首先将上述理论具体化为网络架构，涉及网络拓扑、层和参数的设置细节。我们还讨论了如何赋予上述框架以数据自适应性，通过特征/架构选择或级联学习模块。注意本节内容仅作为层次不变性的一个可行的实践，基于第 5.5 节应用的一般建模，面向其他应用的层次不变性实践是开放性。

5.4.1 网络架构

基于性质 5.1 ~ 5.3，我们面向感兴趣对称群 $\mathfrak{G}_0 = \mathfrak{G}_1 \times \mathfrak{G}_2$ 提出层次不变性的一个具体实践架构，其具有树状的拓扑结构并可拓展到多尺度上。

单尺度网络。让我们首先介绍单个尺度上的拓扑结构，即所有涉及的卷积层具有共同的尺度参数 w ，其表现出 $\mathfrak{G}_1$ 上的不变性（性质 5.1 和 5.2）。如图 5.2 所示，我们将路径的集合组织为树状的网络结构：1) 蓝色节点对应表征单元 $\mathbf{U}$ 具有不同参数 (n, m) ，黑色节点对应恒等变换；2) 线表示表征单元之间的级联关系，在路径中排序 l 相同的全部节点被绘制在同一级 l ，执行顺序为从上到下。注意，每个节点对应的特征图将输入不变层 $\mathbf{I}$ 以形成该路径下的不变表征，该网络整体的表征即所有路径表征的集合。这里，前序节点对应的阶数参数总是小于后续节点（基于特定的范数），因此任一路径均表现出阶数上升的趋势。这一设计是使得图像的主要信息

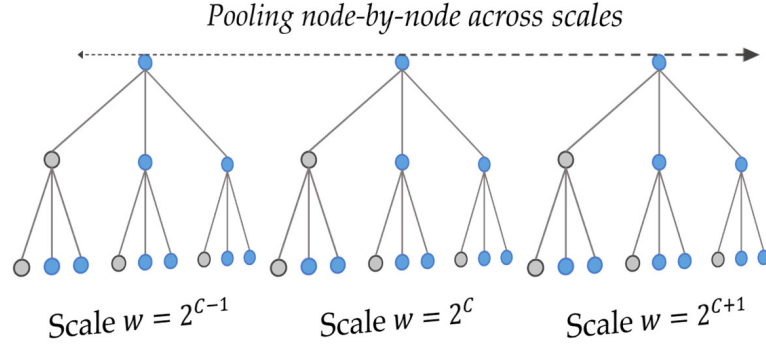


图 5.3 具有 $\mathfrak{G}_0$ 不变性的多尺度实践：多尺度 HIR 网络基于尺度分离先验，其中缩放协变性转变为多尺度网络之间的线性平移模式，可通过从多尺度上对应节点的特征池化得到缩放不变表征。

通过早期节点，以允许后续节点捕捉丰富特征。另外，恒等变换（黑色节点）也被引入网络，作为一种跳跃连接技巧，允许更多图像信息传递至深层。在本文中，处于同一级 l 的所有表征单元的阶数参数分别取自集合 $\{(n, m) : n + m = l, (n, m) \in \mathbf{N}^2\}$ ，即这些参数具有共同的 ℓ_1 范数。

多尺度网络。接下来，让我们考虑上述网络的多尺度版本，引入尺度分离先验，将不变性拓展到对称群到 $\mathfrak{G}_0$ （性质 5.3）。如图 5.3 所示，多个如前所述的单尺度网络被引入：1) 它们具有完全相同的树状结构和节点阶数参数，但每个网络的尺度参数不同，采样自集合 $\{w : w = 2^t, t \in \mathbf{Z}\}$ ，即将缩放因子的指数增长模式（关于 w ）转化为尺度间的线性平移模式（关于 t ）。直接地，上述网络所形成的表征集合即可视为图像的多尺度表征（具有 $\mathfrak{G}_1$ 不变性），可用于具有多尺度结构的视觉任务。进一步地，鉴于性质 5.3 的协变关系，我们可以形成缩放不变表征（具有 $\mathfrak{G}_0$ 不变性），通过一系列多尺度对应节点上的特征图池化。注意，在实践中我们无法对尺度进行完全和稠密的采样，因而上述缩放不变性是受限的。

径向基函数。在定义 3.3 和 3.4 中，我们已经介绍了两类广义的径向基函数，分别基于谐波函数家族和多项式函数家族。二者可以被用于定义（5.2）中的 R_n ，均满足 5.2.1 节提及的正交条件。为了简洁并凸显方法的有效性，一组简单的余弦函数被选为径向基函数用于全部实验，即 PCT^[57] 的 HIR 扩展版本。

不变层。图像域全局不变量的一系列策略^[16,32]均可被用于定义（5.14）中的 $\mathcal{I}$ ，基于深度特征图的等变和协变行为（性质 5.1 ~ 5.3）。为了简洁并提供取证应用的解释性，我们设计了基于频率池化的全局不变量并用于全部实验。

对于公式（5.14），我们首先令傅里叶基作为 $V_{nm}(x_i, y_j)$ 。注意傅里叶变换在信号处理社区已经被充分地研究，可视为良好的解释基础。随后，基于傅里叶变换的阶数/频率采样 $(n, m) \in$

$[-K, K]^2$ ，我们定义 $\mathcal{I}$ 为极坐标的频带积分：

$$\mathcal{I}(\{\langle M, V_{nm} \rangle\}) \triangleq \{I_i = \sum_{(n,m) \in \mathcal{B}_i} \{\langle M, V_{nm} \rangle\} : i = 1, 2, \dots, \#_B\}, \quad (5.24)$$

其中， $\mathcal{B}_i = \{(n, m) : \sqrt{2Ki} - 1/\#_B \leq \|(n, m)\|_2 \leq \sqrt{2Ki}/\#_B\}$ 为 l_2 范数下的第 i 频带，具有频带数量 $\#_B$ 。

由此，我们可以声明由上述频带积分导出的特征 $\{I_i : i = 1, 2, \dots, \#_B\}$ 直接满足 $\mathfrak{G}_1$ 不变性，鉴于性质 5.1 以及傅里叶变换的平移、旋转和翻转性质。至于缩放， $\mathcal{I}$ 兼容单尺度和多尺度网络：对于前者，提供一定程度的鲁棒性对于 $\mathfrak{G}_2$ ，以频带宽度为上限，鉴于性质 5.2 以及傅里叶变换的缩放性质；对于后者，缩放协变性已在输入 $\mathcal{I}$ 之前得到消解，因此将满足联合不变性对于 $\mathfrak{G}_0 = \mathfrak{G}_1 \times \mathfrak{G}_2$ 。

注意，常用的平均池化事实上是公式 (5.24) 的特例，其中 $K = 0$ 和 $\#_B = 1$ 。因此上述频带积分 $\mathcal{I}$ 可被视为一种推广的全局池化操作，以兼顾可解释性、不变性和判别性的共同要求。

5.4.2 数据自适应性

鉴于 HIR 的手工设计性质，其固定特征集合对于输入数据的分布不具有自适应性。面向更大规模视觉任务，我们提出以下策略赋予其数据自适应性，以达到学习表征类似的判别性水平。注意，相比于典型（即完备或欠完备）的不变量，提出表征由于其局部和层次结构而表现出高度的过完备性，这是以下数据自适应策略的基础。

特征/架构选择。对于给定任务的判别性特征可以以选择的方式形成，受网络架构搜索^[342]（Neural Architecture Search, NAS）启发。首先，我们可以构造涵盖一个大规模（深和宽）的树状网络，涵盖广泛路径集合和参数取值，类似 NAS 中超网^[343]的概念。随后，基于给定任务下的训练数据集，我们可以执行特征集合和标签的相关性分析排序判别特征（以及相应的路径），类似 NAS 中的架构采样和评价阶段^[343]。基于此，在应用环节，我们可以对上述初始网络进行大幅度的简化使得路径主要覆盖该任务下的判别特征而非无关特征，允许高效、可解释且任务判别的不变表征。

级联学习模块。对于给定任务的判别性特征也可以以级联学习模块的方式形成，受混合表征学习^[324]（Hybrid Representation Learning, HRL）启发。这里，我们的主要思想是将学习的 CNN 的浅层用固定的层次不变表征代替，允许判别性特征形成于一个具有几何对称性的空间。根据相关论证^[324]，该策略能够在大规模分类基准上实现与典型 CNN 匹敌的判别性水平，同时表现出明显更好的训练紧凑性。在本文的全部实验和应用中，我们仍然采用朴素的特征/架构选择策略以强调 HIR 本身的能力，同时还考虑到级联的学习 CNN 对不变性和解释性存在不同程度的削弱。

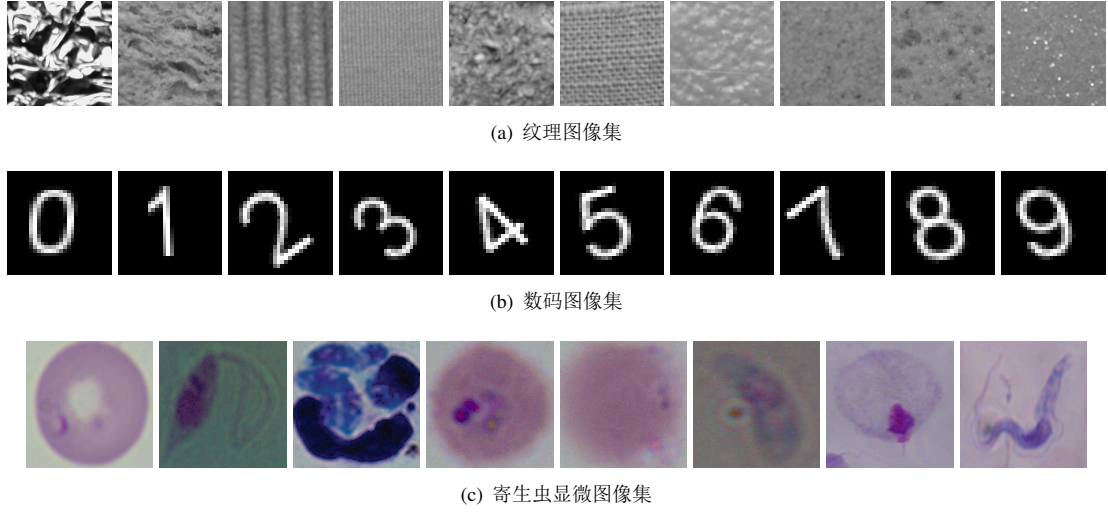


图 5.4 计算机视觉和模式识别实验中的图像集示例。

5.5 实验和应用

本节将全面地评估 HIR 的判别性、稳健性和效率，分别涵盖 5.5.1 节的仿真实验和 5.5.2 节的现实应用。在这里，主要目的是检验前文理论承诺的性质，并确定其在深度学习时代的判别能力。

5.5.1 计算机视觉和模式识别

仿真实验方面，我们在典型的纹理、数字和寄生虫图像集上分别进行了 HIR 的分类实验，以基准评估其表征能力。注意，这一系列实验从不同问题规模和几何变体的下全面考察了前文理论承诺的性质，并提供了与多种手工和深度的图像表征方法的对比。

这些实验中，HIR 均基于第 5.4 节的实践策略，基于单尺度网络，尺度参数 $w = 10$ ，级联层数 $L = 6$ ；不变层（5.24）被设置为特化的平均池化： $K = 0$ 和 $\#_B = 1$ ，以便与基于平均池化的深度表征进行公平对比。注意，这部分实验并未采用 5.4.2 节的数据自适应策略，以直接反映表征原始的判别能力。特征均输入一个经典的 PCA 分类器，该分类器使用训练集的特征集合训练。这里，除非另有说明，训练集和测试集在原始图像库上分别按 80% 和 20% 比例随机采样形成，无任何交叉。

对比方法包括：

- 经典完备表征：离散余弦变换（Cosine）作为全局表征；
- 经典过完备表征：离散小波变换^[344]（Wavelet）和克拉夫丘克矩^[41]（Krawtchouk）作为局部表征，具有不同的时频分辨率；
- 先进过完备表征：1) 经典 CNN，包括直接学习 CNN（SimpleNet），迁移学习 AlexNet^[345] 和 VGG16^[346]，同时涉及数据增广训练技巧（“+” 标记）；2) 不变 CNN，即散射网络^[321]

表 5.2 小规模纹理基准上不同表征方法的分类分数(%)和运行时间(秒)。

表征方法	时间	直接测试		F1	随机朝向和翻转变体测试		
	GPU†	Precision	Recall		Precision	Recall	F1
经典表征							
Cosine	5	70.74	67.50	66.85	69.65	66.25	65.30
Wavelet	6	69.43	64.38	64.68	62.34	58.13	57.82
Krawtchouk	5	70.67	67.50	66.30	64.41	60.00	59.55
学习表征							
SimpleNet	52†	70.33	67.50	67.09	54.63	43.13	41.31
SimpleNet+	52†	46.93	49.38	46.06	47.18	48.13	44.93
AlexNet	42†	98.82	98.75	98.75	91.69	91.25	91.28
AlexNet+	41†	87.61	84.38	84.05	88.37	85.63	85.76
VGGNet	266†	99.41	99.38	99.37	92.18	91.25	91.37
VGGNet+	609†	91.34	90.00	89.81	92.15	91.25	91.08
不变表征							
ScatterNet	42	98.89	98.75	98.75	84.98	83.13	83.08
HIR	27	96.98	96.88	96.87	96.32	96.25	96.23

(ScatterNet) 和我们的 HIR。总体而言, 它们均可视为一类层次不变表征, 但具有不同的不变性水平, 其中典型的 CNN 仅对平移具有不变性。

5.5.1.1 纹理

如图 5.4 所示, 该实验在 KTH-TIPS¹数据集上进行, 这是一个经典的纹理图像分类基准。该数据集共有 10 个类别, 每个类别包含 81 个实例, 总大小为 $10 \times 81 = 810$, 因此属于小规模视觉问题。

如表 5.2 所示, 我们列出了各类表征在该基准上的三种性能分数, 以及时间耗用, 即手工表征的 CPU 特征时间和学习表征的 GPU 训练时间。注意实验协议还考虑了随机朝向 (即 $\{0, 90, 180, 270\}$ 度旋转) 或翻转 (即 x 或 y 轴向翻转) 的变体测试。可以观察到:

- 经典的完备或过完备表征未能实现令人满意的判别性水平, 即使在这样小规模问题的直接测试协议下。
- 相反, 学习 CNN 家族整体上实现了显著更高的精度水平由于其高度过完备性和数据自适应性, 特别是基于大规模训练和迁移学习的 AlexNet 和 VGGNet; 而在其中, SimpleNet 的性能相对较差, 这体现出学习表征对于网络规模以及训练方式的高度依赖性。在变体测试协议下, 基于直接训练的 CNN 家族表现出明显的性能下降, 表明其学得表征对于纹理的自然几何变化缺乏内在的不变性。在引入基于朝向和翻转的数据增广训练后, CNN 表征的分类分数更加稳定, 但似乎以判别性为代价。这一现象的主要原因是过少的训练

¹<https://www.csc.kth.se/cvap/databases/kth-tips/index.html>

表 5.3 中等规模数码基准上不同表征方法的分类分数 (%) 和运行时间 (秒)。

表征方法	时间	直接测试			随机平移和旋转变体测试		
	GPU†	Precision	Recall	F1	Precision	Recall	F1
经典表征							
Cosine	15	45.68	45.35	45.43	32.50	31.20	30.83
Wavelet	16	67.11	66.75	66.75	38.31	35.25	35.44
Krawtchouk	15	71.73	69.85	69.69	27.05	26.30	25.83
学习表征							
SimpleNet	535†	98.60	98.60	98.60	35.42	33.50	33.72
SimpleNet+	551†	52.70	48.90	48.82	54.26	50.75	50.78
AlexNet	393†	100	100	100	66.18	64.45	64.27
AlexNet+	392†	93.07	92.00	91.80	94.23	93.10	92.93
VGGNet	3610†	100	100	100	70.74	70.25	69.93
VGGNet+	7731†	95.98	95.70	95.68	95.53	95.20	95.13
不变表征							
ScatterNet	115	98.96	98.95	98.95	57.20	56.95	56.23
HIR	57	97.48	97.45	97.45	95.05	94.95	94.98

数据量。此外，上述学习 CNN 家族计算资源的耗用相当可观对于此小规模问题，并且观察到其性能对训练过程的不稳定性。

- 相比之下，散射网络在无需训练和增广的情况下提供了相对良好的判别性和稳健性水平，表明其将经典小波拓展至深度完备表征的成功性。特别是相对于表中原始小波的判别能力。
- 我们的研究进一步扩大了这种范式的成功。所提出的 HIR 实现了与学习 CNN 家族类似的判别水平，同时相比于散射网络在变体测试中表现出更优的稳健性能。注意，这种优势建立于我们的更紧凑和高效的表征框架，表现出更低的计算时间相比于上述两类工作。

5.5.1.2 数码

如图 5.4 所示，该实验在数码图像数据集¹上进行，这是一个类似于 MNIST 的基准。该数据集共有“0”到“9”的 10 个类别，每个类别包含 1000 个实例，模拟了非常丰富的字体差异和几何变化，总规模是 $10 \times 1000 = 10000$ ，因此属于中等规模视觉问题。

如表 5.3 所示，我们列出了各类表征在该基准上的三种性能分数以及时间耗用。注意实验协议还考虑了随机平移（即 $-2 \sim 2$ 像素的轴向平移）或随机旋转（即 $0 \sim 360$ 角度旋转）的变体测试。可以观察到：

- 随着问题规模的扩大，完备的 Cosine 表现出明显的性能下降，而过完备的 Wavelet 和 Krawtchouk 更稳定，说明过完备性对于判别水平非常重要。在稳健性方面，此类经典表

¹<https://ww2.mathworks.cn/help/deeplearning/ug/data-sets-for-deep-learning.html>

表 5.4 大规模寄生虫显微基准上不同表征方法的分类分数 (%) 和运行时间 (秒, 对应 8/2)。

表征方法	时间	训练/测试 = 8/2			训练/测试 = 1/9		
	GPU†	Precision	Recall	F1	Precision	Recall	F1
经典表征							
Cosine	37	36.19	32.60	29.85	49.40	41.97	43.80
Wavelet	39	41.68	45.20	41.79	53.69	47.97	49.27
Krawtchouk	42	66.56	69.49	67.21	71.60	57.88	61.10
学习表征							
SimpleNet	2244†	90.15	89.25	89.65	84.51	76.14	78.84
AlexNet	1796†	98.87	98.40	98.63	95.92	94.69	95.27
VGGNet	9184†	99.24	98.97	99.11	97.95	97.37	97.65
不变表征							
ScatterNet	1277	68.41	69.71	67.55	72.52	63.30	65.70
HIR	823	88.73	92.18	90.10	91.26	88.76	89.85

征对随机平移和旋转的变体测试均无法实现稳定的分类分数，这也从侧面反映该实验协议的挑战性。

- 整体上，学习 CNN 家族的性能表现仍然符合 5.5.1.1 节的分析，进一步证实了其对训练的高度依赖性和对几何变体的不稳定性。在引入基于平移和旋转的数据增广训练后，CNN 表征的稳健性明显增强。然而理论上，获得的不变性不受保障对于未见数据分布（即便是未见参数的同类变体）。特别应注意其计算规模的快速扩张，VGGNet 的迁移学习甚至需要约 2 个小时的 GPU 时间。
- 随着问题规模的扩大，非学习的散射网络依然提供了良好的判别性水平，进一步验证了该策略的成功性。然而，其对于平移和旋转变体的分数偏低（甚至低于未增广的学习 CNN），未能实现其预期的稳健性。
- 在中等规模问题中，我们的 HIR 在内在稳健性和紧凑性的约束下仍然实现了学习 CNN 的判别性水平，因而显著优于直接竞争者散射网络。这里有一个值得注意的实验现象，HIR 是唯一在直接训练中表现出对类别“6”到“9”混淆的方法，同时对其他类实现了 ~100% 的分类分数，证实了理论预期的旋转不变性和高度判别性。

5.5.1.3 寄生虫

如图 5.4 所示，该实验在显微图像数据集¹上进行，这是一个用于寄生虫图像分类的典型基准。该数据集共有 6 个寄生虫类和 2 个宿主类，每类实例数量各异，在成像、背景、形态和几何层面表现出真实的多样性，总规模是 34298，因此属于现实世界的大规模视觉问题。

如表 5.4 所示，我们列出了各类表征在该基准上的三种性能分数以及时间耗用。注意实验

¹<https://data.mendeley.com/datasets/38jtn4nzs6/3>

协议还考虑了不同的训练-测试比例以分析训练数据依赖性以及样本效率。可以观察到：

- 在此大规模问题中，经典表征的分类分数进一步下降，表明浅层手工表征严重受限的判别性水平。另一方面，当训练样本的减少时，它们的性能对于相对稳定，甚至在小样本的 1/9 情况表现更好，表明其在样本效率上的优势。
- 在学习 CNN 家族中，基于直接训练的 SimpleNet 表现出明显的数据依赖性。具体而言，其在 8/2 情况下实现 ~90% 的分数水平（与 HIR 类似），而在 1/9 情况下分数水平下降明显（低于 HIR）。相比之下，AlexNet 和 VGGNet 实现了最优的判别性和稳定性水平，表明迁移策略有效继承了 ImageNet 上的预训练先验。另一方面，大规模预训练和迁移计算的成本仍是相当可观的，并且对于给定数据域不具有受保障的稳健性和迁移性。
- 尽管优于原始小波，散射网络未能提供与学习 CNN 类似的判别性水平，这显然限制了该类稳健和可解释表征在更大规模任务上的应用。这里，一系列手工表征在更大规模判别性方面的普遍失效，可视为本章动机的重要证据。
- 在大规模问题中，我们的 HIR 实现了与 SimpleNet 类似的判别性，并且显著优于直接竞争者散射网络。同时，HIR 对于训练数据的减少并不敏感，在训练数据依赖性以及样本效率方面优于学习 CNN。注意，HIR 固定特征的判别性水平仍低于基于大规模预训练的迁移学习策略。因此在接下来的现实应用中，我们将基于 5.4.2 节的策略赋予数据自适应性。

5.5.2 数字取证和篡改检测

现实应用方面，我们将 HIR 应用于大规模的数字取证任务，即对抗性扰动和 AIGC 的检测，以直接检验其在稳健和可解释的现实任务中的有效性。注意这一“即插即用”策略不仅与同类表征进行对比，还将与目前的一系列取证解决方案（包括精心设计的深度取证）直接竞争。

这些应用中，HIR 均基于第 5.4 节的实践策略，基于单尺度网络，尺度参数 $w = 10$ ，级联层数 $L = 7$ ；不变层 (5.24) 被设置为 $K = \#_{ij}/2$ 和 $\#_B = 30$ ，为了提升对数字一致性缺陷的判别性。注意，这里采用 5.4.2 节的特征/架构选择策略，分别为 AIGC 和对抗性扰动保留任务最相关的前 500 和 1000 维特征，提高对现实应用的数据自适应性和判别性。得到的特征均输入经典的 NN 或 SVM 分类器以评估特征对分类器的敏感性，两个分类器使用训练集的特征集合训练。这里，除非另有说明，训练集和测试集在原始图像库上分别按 50% 和 50% 比例随机采样形成，无任何交叉。

对比方法包括：

- 沿用 5.5.1 节的表征策略用于直接检测；
- 更多深度表征里程碑用于直接检测，包括 GoogLeNet^[232]、ResNet^[233]、DenseNet^[347]、InceptionNet^[348] 和 MobileNet^[349]；



图 5.5 数字取证和篡改检测应用中的图像集示例。

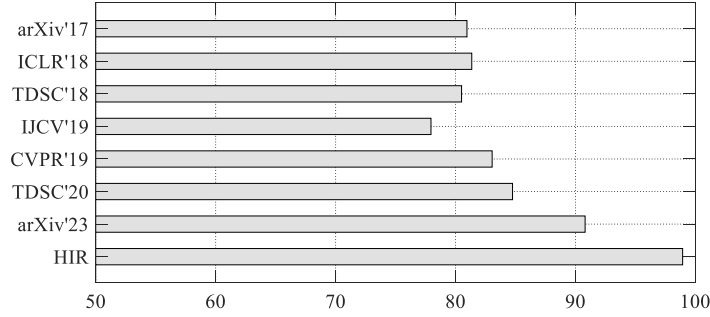


图 5.6 对抗性扰动 UP 基准上不同取证方法的取证分数 (F1, %) 对比。

- 为对抗性扰动取证专门设计的 arXiv'17^[350]、ICLR'18^[351]、TDSC'18^[352]、IJCV'19^[353]、CVPR'19^[354]、TDSC'20^[355] 和 arXiv'23^[356]；
- 为 AIGC 取证专门设计的 ECCV'20^[357]、CVPR'20a^[358] 和 CVPR'20b^[359]。

5.5.2.1 对抗性扰动

如图 5.5 所示, 6 种对抗性扰动, 即 BIM^[360]、CW^[361]、Damage^[362]、FGSM^[363]、PGD^[364] 和 UP^[365], 被分别施加于图像分类经典基准 ImageNet¹, 由此形成了 6 个基准, 每个基准包含 5000 张干净图像和 5000 张相应的扰动版本。鉴于对抗性扰动本身的丰富变化, 以及基础的 ImageNet 的高度多样性, 该问题表现出现实世界级别的判别性挑战。

在图 5.6 中, 我们首先在基础且实际的 UP 基准上提供了与对抗性扰动取证现有方案的对比。尽管具有固定的扰动模式, 但仍有很多竞争方法无法实现良好的取证分数。这些方法通常基于非完备的表征, 因而无法全面地捕捉扰动模式。而基于过完备表征的 arXiv'23 以及提出的 HIR 均实现了 >90% 的取证分数, 进一步揭示了表征在取证任务的基础性地位。接下来, 我们将在表征层面全面地对比相关竞争策略。

¹<https://www.image-net.org/>

表 5.5 不同表征方法在多种对抗性扰动基准上的取证分数 (F1, %)。

表征方法	BIM	CW	DAmage	FGSM	PGD	UP	平均分数	最差分数
经典表征								
Cosine NN	34.63	33.19	90.78	39.80	34.69	2.22	39.22	2.22
Cosine SVM	79.57	83.34	97.26	78.24	79.22	96.68	85.72	78.24
Wavelet NN	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Wavelet SVM	72.83	82.09	97.77	78.21	71.80	95.87	83.10	71.80
Krawtchouk NN	66.43	66.49	90.86	66.43	66.44	0.00	59.44	0.00
Krawtchouk SVM	0.00	55.87	0.00	56.44	0.00	70.37	30.45	0.00
学习表征								
SimpleNet	4.24	3.24	92.13	49.89	33.13	99.86	47.08	3.24
AlexNet	90.20	72.72	96.63	94.61	90.91	98.45	90.59	72.72
VGGNet	96.04	62.50	99.08	98.12	96.99	99.15	91.98	62.50
GoogLeNet	90.29	80.04	97.09	95.29	89.94	98.75	91.90	80.04
ResNet	90.22	75.59	97.35	94.66	90.17	98.40	91.07	75.59
DenseNet	98.93	90.19	99.34	99.23	98.85	99.76	97.72	90.19
InceptionNet	98.70	85.14	97.38	97.32	98.66	99.41	96.10	85.14
MobileNet	92.51	82.67	97.37	96.81	92.10	98.19	93.27	82.67
不变表征								
ScatterNet NN	81.30	70.23	95.27	91.17	82.65	94.64	85.88	70.23
ScatterNet SVM	84.40	69.49	96.77	90.57	83.86	95.12	86.70	69.49
HIR NN	89.66	84.92	98.89	93.26	90.08	97.78	92.43	84.92
HIR SVM	92.30	89.10	99.30	95.96	91.60	98.93	94.53	89.10

在表 5.5 中, 我们分别在 6 种扰动基准上训练并测试表征方法, 给出了相应的 F1 分数以及平均得分和最差得分的统计数据。该协议相比于固定扰动的取证表现出更丰富的类内变化。可以观察到:

- 自然数据和扰动版本的频率差异被证明是一种富有成效的取证线索。在此大规模问题上经典频域和时频域表征实现了高于一般预期的取证精度。然而, 这些特征表现出严重的分类器依赖性, 可能是因为在特征空间中受限的可分离性, 导致必须诉诸复杂分类策略。
- 在学习 CNN 家族中, 除直接训练的 SimpleNet 外, 其他大规模的网络均表现出 >90% 的平均取证分数, 特别是 DenseNet 和 InceptionNet。这说明迁移学习策略善于在数据充足且训练-测试高度对齐的情况下捕捉判别性特征。从攻击的角度看, CW 上的取证更具挑战性, 主导了最差分数, 由于其多变且微弱的扰动模式。
- 与原始小波相比, 散射网络实现了相近的取证分数, 但分类器稳定性明显更优。这说明特征分离性方面的改善。然而, 散射网络平均取证分数未达到 90%, 未能提供与学习 CNN 类似的判别性水平。
- 我们的 HIR 对分类器的变化非常稳健, 在此基础上实现了学习 CNN 的判别性水平, 即与 MobileNet 类似, 略低于 DenseNet 和 InceptionNet, 且显著优于直接竞争者散射网络。

表 5.6 不同表征方法在对抗性扰动实际（混合）基准上的取证分数（%）。

表征方法	训练/测试 = 5/5			训练/测试 = 1/9		
	Precision	Recall	F1	Precision	Recall	F1
经典表征						
Cosine NN	0.00	0.00	0.00	0.00	0.00	0.00
Cosine SVM	79.08	73.33	76.10	81.13	68.79	74.45
Wavelet NN	0.00	0.00	0.00	0.00	0.00	0.00
Wavelet SVM	77.53	66.95	71.85	76.05	61.13	67.78
Krawtchouk NN	50.53	15.22	23.40	50.00	15.10	23.20
Krawtchouk SVM	50.03	65.34	56.67	49.75	48.77	49.26
学习表征						
SimpleNet	47.31	48.11	47.71	50.59	63.63	56.36
AlexNet	81.46	87.35	84.30	72.24	61.36	66.35
VGGNet	81.41	90.04	85.51	78.83	75.35	77.05
GoogLeNet	82.74	85.46	84.08	63.35	57.74	60.42
ResNet	80.93	84.70	82.77	68.48	66.64	67.55
DenseNet	87.92	93.25	90.51	82.07	83.96	83.00
InceptionNet	84.60	90.92	87.65	69.58	70.77	70.17
MobileNet	83.07	88.07	85.50	68.73	69.50	69.11
不变表征						
ScatterNet NN	69.85	68.94	69.39	74.93	77.31	76.10
ScatterNet SVM	75.70	72.07	73.84	76.42	78.63	77.51
HIR NN	81.27	80.68	80.98	79.09	82.17	80.60
HIR SVM	86.20	86.06	86.13	83.42	83.29	83.35

因此，HIR 可视为稳健性、解释性、判别性综合性能更好的表征策略，特别其效率优势将在下面的挑战性实验协议中得到进一步突显。

在表 5.6 中，我们以 6 种对抗性扰动混合的形式训练并测试表征方法，给出了两种训练-测试比例下的三种取证分数。注意，该实验协议更具挑战性由于更复杂的类内变化，并且更加符合真实世界取证场景。可以观察到：

- 经典表征对于不同分类器仍表现出严重的性能波动，与之前的观察一致。我们还注意到相比于表 5.5 的情况，这些手工策略性能有所下降，由于该混合协议导致的判别性挑战。另一方面，它们的表现相对于训练样本减少非常稳定，进一步验证了手工表征在效率方面的固有优势。
- 随着进入混合基准，学习 CNN 家族产生了一致且较大的性能下降，特别是对于小样本的 1/9 情况。该现象是学习表征（即使基于迁移策略）数据依赖性的直接证据。实际上，对抗性扰动的真实世界取证情况往往是攻击类型多样且部分攻击的训练样本偏少。因而，数据依赖的学习取证算法在实践中常表现出耗时的（重）训练，同时对代表性不足的扰动模式难以保证其基本功能。

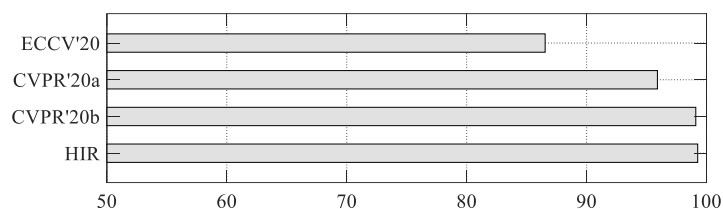


图 5.7 人工智能生成内容 SD 1.5 基准上不同取证方法的取证分数 (F1, %) 对比。

- 散射网络基本继承了表 5.5 中的判别性水平和分类器稳定性。值得一提的是其在小样本的 1/9 情况的分数优于大部分经典和学习表征，反映其在判别性和样本效率方面的综合性能较优。
- 在这个挑战性的实验协议中，我们手工的 HIR 依然实现了与学习 CNN 类似的判别性水平，并优于散射网络。更重要的是，HIR 对训练样本的依赖性明显低于学习 CNN，意味着在实践中可以更好应对代表性不足的扰动模式。在接下来更大规模的 AIGC 取证任务中，HIR 在稳健性、解释性、判别性和效率方面的综合性能优势相比于学习 CNN 将进一步突显。

5.5.2.2 人工智能生成内容

如图 5.5 所示，8 种 AIGC 方法，即 ADM^[366]、BGAN^[367]、GLIDE^[368]、Midjourney¹、SD 1.4^[369]、SD 1.5^[369]、VQDM^[370] 和 Wukong²，被分别用于生成与 ImageNet 内容类似的图像，由此形成了 8 个基准，每个基准包含 6000 张真实图像和 6000 张合成图像。鉴于生成内容和自然内容自身高度的多样性，且二者的差异是非常多变的，该问题表现出更高水平的判别性挑战。

在图 5.7 中，我们首先提供了与先进 AIGC 取证解决方案的对比，在基础的 SD1.5 基准上。可以看到这些基于深度网络和特征增强的方法均实现了良好的取证精度。这说明对典型的学习表征网络而言，数据充足且训练-测试高度对齐的取证场景不构成挑战。这里，手工的 HIR 也实现了 ~100% 的取证分数，说明在此场景下表现出类似的判别能力。接下来，我们将主要在表征层面全面地对比相关竞争策略，也提供了一些 AIGC 取证先进解决方案的分数作为参考。

在表 5.7 中，我们分别在 8 种 AIGC 基准上训练并测试竞争表征，给出了相应的 F1 分数以及平均得分和最差得分的统计数据。同时在表 5.8 中，我们还考虑了测试集的随机朝向/翻转版本，以反映最基础的几何稳健性需求。可以观察到：

- 人工智能生成数据的频率线索仍然十分有效。基于 SVM 分类器的经典表征实现了一致较好的取证精度和几何稳健性，说明包括扩散模型在内的先进生成器仍表现出某种固定的频率瑕疵。另一方面，这些手工特征对分类器敏感，与 5.5.2.1 节的观察相同。

¹<https://www.midjourney.com/home>

²<https://xihe.mindspore.cn/modelzoo/wukong>

表 5.7 不同表征方法在多种 AIGC 基准上的取证分数 (F1, %)。

表征方法	ADM	BGAN	GLIDE	Midjourney	SD1.4	SD1.5	VQDM	Wukong	平均分数	最差分数
经典表征										
Cosine NN	0.00	0.00	0.00	0.00	0.00	65.09	0.00	0.00	8.14	0.00
Cosine SVM	99.19	99.95	99.57	89.02	99.10	98.80	99.46	99.11	98.03	89.02
Wavelet NN	0.00	0.00	0.00	0.00	0.00	2.08	0.00	0.00	0.26	0.00
Wavelet SVM	99.98	99.70	99.87	85.55	98.63	99.04	99.97	99.38	97.76	85.55
Krawtchouk NN	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Krawtchouk SVM	99.75	99.60	98.49	62.44	74.11	77.98	93.63	76.00	85.25	62.44
学习表征										
SimpleNet	98.25	97.87	92.98	68.00	73.52	74.37	74.88	76.32	82.02	68.00
AlexNet	94.45	98.99	98.26	81.52	87.96	88.66	84.24	88.62	90.34	81.52
VGGNet	99.40	99.38	98.57	86.44	89.97	91.86	93.60	90.09	93.66	86.44
GoogLeNet	80.30	99.18	98.16	75.00	82.77	82.44	86.75	82.32	85.87	75.00
ResNet	98.78	99.14	97.78	87.41	89.88	90.85	88.53	88.80	92.65	87.41
DenseNet	99.63	99.60	98.57	93.08	93.79	94.50	95.01	92.55	95.84	92.55
InceptionNet	97.69	99.41	98.32	90.07	89.40	92.55	92.72	88.35	93.56	88.35
MobileNet	90.08	99.28	97.95	87.49	88.51	90.75	87.74	88.29	91.26	87.49
不变表征										
ScatterNet NN	99.10	99.63	98.44	79.47	89.26	89.95	96.71	89.07	92.70	79.47
ScatterNet SVM	99.18	99.67	99.05	85.21	95.85	95.58	97.02	94.60	95.77	85.21
HIR NN	99.92	99.97	99.83	92.63	98.57	98.97	99.92	98.58	98.55	92.63
HIR SVM	99.90	99.92	99.78	92.18	99.07	99.26	99.87	99.42	98.68	92.18

- 除 SimpleNet 和 GoogLeNet 外的学习表征实现了 >90% 的平均取证分数。进一步证实学习 CNN 家族在训练充足且高度对齐场景中具有良好的判别性。然而在表 5.8 的稳健性测试中，这些方法均表现出不同程度的性能劣化（在平均和最差统计层面）。可见即使自然且微小的数据分布偏移（例如翻转）也可能强烈干扰学习表征的取证能力。更糟糕的是这种干扰是高度黑盒且难以预测的，一个例子是 ADM 上的性能波动明显高于其他。
- 散射网络在这里具有与学习表征类似的判别性和稳健性水平，同时在分类器稳定性方面优于原始小波。注意其并表现出理论预计的不变性，因而相对于学习 CNN 未能实现更高的稳健性分数。
- 在判别性、几何稳健性和分类器稳健性三方面，我们的 HIR 实现了更好的综合性能相对于经典表征、散射网络以及学习 CNN 家族。这与我们的理论预期相符，即 HIR 兼顾了手工表征和学习表征的优势。其效率优势将在下面的挑战性实验协议中得到进一步凸显。

在表 5.9 中，我们以 8 种 AIGC 图像混合的形式训练并测试表征方法，给出了两种训练-测试比例下的三种取证分数。注意该实验协议更具挑战性由于更复杂的类内变化，并且更加符合真实世界取证场景。可以观察到：

- 经典表征在此混合基准上表现出良好的判别性以及训练样本减少一贯的稳定性，验证了其在小样本场景下的性能优势。与之前的观察一致，这些方法对于不同分类器仍表现出严重的性能波动。
- 与之相对，学习 CNN 家族的取证分数明显下降，在小样本的 1/9 情况的平均 F1 分数仅

表 5.8 不同表征方法在多种 AIGC 基准上的取证稳健性分数 (F1, %)。

表征方法	随机朝向和翻转变体测试								平均分数	最差分数
	ADM	BGAN	GLIDE	Midjourney	SD1.4	SD1.5	VQDM	Wukong		
经典表征										
Cosine SVM	99.16	99.95	99.55	88.07	99.05	98.72	99.36	99.11	97.87	88.07
Wavelet SVM	99.95	99.80	99.85	82.83	99.08	98.71	99.95	99.16	97.42	82.83
Krawtchouk SVM	70.90	99.58	98.57	64.90	77.28	76.58	94.44	76.72	82.37	64.90
学习表征										
SimpleNet	77.72	95.04	92.99	65.25	74.52	74.72	73.62	76.91	78.85	65.25
AlexNet	81.82	99.08	97.99	77.45	85.87	87.93	83.29	86.78	87.53	77.45
VGGNet	76.18	99.40	98.41	82.23	89.26	89.30	88.79	89.00	89.07	76.18
GoogLeNet	80.62	99.30	98.09	73.13	81.29	81.92	85.36	82.31	85.25	73.13
ResNet	85.62	99.28	97.60	81.82	85.60	86.84	87.94	85.21	88.74	81.82
DenseNet	84.57	99.56	98.66	88.57	91.16	91.60	94.26	89.47	92.23	84.57
InceptionNet	91.99	99.30	98.56	87.04	85.72	88.85	92.19	85.24	91.11	85.24
MobileNet	85.14	99.36	97.78	84.54	86.74	88.80	87.50	86.56	89.55	84.54
不变表征										
ScatterNet SVM	92.61	99.67	99.18	82.75	88.81	89.40	97.05	83.83	91.66	82.75
HIR SVM	99.88	99.87	99.77	91.99	99.02	99.25	99.83	99.45	98.63	91.99

为 ~70%。这进一步说明基于学习表征的取证算法在应对真实世界场景上的不足，即数据依赖问题。

- 散射网络继承了表 5.7 中的判别性水平和分类器稳定性，并且在小样本的 1/9 情况的分数优于全部学习表征，说明其较优的综合性能。
- 在这里，HIR 实现了最高的取证精度相比于经典表征、散射网络以及学习 CNN 家族。其判别能力足以支撑 1/9 小样本情况下的 8 种 AIGC 图像的混合检测，意味着取证场景中良好的实用性。值得注意的是为 AIGC 取证专门设计的 ECCV'20、CVPR'20a 和 CVPR'20b 在 1/9 情况下的 F1 分数分别为 79.67%、73.74% 和 76.89%。综合以上，HIR 以“即插即用”的方式在三种经典视觉任务和两种取证任务中实现了一致良好的判别性、稳健性和效率，没有一类实验中涉及的表征方法和取证方案能实现这一点。

5.6 本章小结

本章对面向更大规模视觉任务的层次不变量进行了系统性研究，将全局和局部不变量的基本理论统一推广到层次假设，实现了判别性和不变性的有效折衷，提出了名为“层次不变表征 HIR”的新方法。与相关工作相比，HIR 的特点可总结为：原则性和可解释性的设计、高效的不变结构，以及深度学习时代有竞争力的判别性水平。

本章工作在理论层面的要点如下：

- 形式化了层次不变性的蓝图，并重新审视了 CNN 的几种典型模块（5.3.1 节）。
- 定义了新的模块及其级联以满足层次不变性蓝图（5.3.2 节），特别是提供了关于图像与表征的几何对称性关系的形式化结论（5.3.3 节）。
- 探讨了上述新概念与典型概念之间的批判和发展关系，凸显了 HIR 在迈向稳健和可解释

表 5.9 不同表征方法在 AIGC 实际（混合）基准上的取证分数（%）。

表征方法	训练/测试 = 5/5			训练/测试 = 1/9		
	Precision	Recall	F1	Precision	Recall	F1
经典表征						
Cosine NN	0.00	0.00	0.00	0.00	0.00	0.00
Cosine SVM	94.95	94.57	94.76	94.36	91.06	92.68
Wavelet NN	48.70	94.17	64.20	48.69	94.13	64.18
Wavelet SVM	94.03	94.57	94.30	83.55	93.48	88.24
Krawtchouk NN	0.00	0.00	0.00	0.00	0.00	0.00
Krawtchouk SVM	75.24	74.77	75.00	71.56	68.57	70.03
学习表征						
SimpleNet	61.79	40.70	49.08	56.40	60.48	58.37
AlexNet	80.76	77.63	79.16	71.83	72.50	72.17
VGGNet	84.75	86.67	85.70	72.45	72.37	72.41
GoogLeNet	74.15	80.40	77.15	67.84	68.83	68.33
ResNet	85.10	83.03	84.06	76.88	73.67	75.24
DenseNet	86.83	85.23	86.02	76.84	75.37	76.10
InceptionNet	82.69	86.63	84.62	68.62	68.56	68.59
MobileNet	81.54	82.47	82.00	68.52	68.57	68.55
不变表征						
ScatterNet NN	83.68	83.73	83.71	79.37	79.70	79.53
ScatterNet SVM	90.31	85.17	87.67	85.28	79.70	82.40
HIR NN	96.79	96.47	96.63	95.66	93.04	94.33
HIR SVM	96.92	96.37	96.64	95.21	94.26	94.73

表征方面的独特性（5.3.5 节）。

本章工作在实践层面的要点如下：

- 提供了层次不变性理论的一个具体框架，涉及网络拓扑、层和参数的实践原则（5.4.1 节）。
- 挖掘了上述框架的数据自适应潜力，通过特征/架构选择或级联学习模块（5.4.2 节）。

本章工作在教育层面的要点如下：

- 在典型的纹理、数字和寄生虫图像集上分别进行了 HIR 的模式分类实验，从不同问题规模和几何变体的下全面考察了理论承诺的性质（5.5.1 节）。
- 在大规模的对抗性扰动和 AIGC 取证任务中直接检验 HIR 的现实有效性（5.5.2 节）。结果表明，HIR 以“即插即用”的方式实现了一致良好的判别性、稳健性、效率和解释性，表现出有竞争力的综合性能相比于现有表征方法和取证方案。

本章工作已总结成文：

- Shuren Qi, Yushu Zhang[✉], Chao Wang, et al. Hierarchical invariance for robust and interpretable vision tasks at larger scales. **International Journal of Computer Vision**, Under Review.

第六章 总结与展望

本文围绕可信人工智能场景，以视觉表征的不变性和对称性先验为切入点，系统性分析了前人的不变表征研究，进一步提出了全局、局部和层次不变表征的新方法，并成功应用于多种信息安全和取证的可信任任务。本章首先回顾全文的核心内容及其重要性，随后指出更长时间周期的研究展望。

6.1 本文总结

稳健和可解释的计算机视觉系统是可信人工智能的重要组成部分。受数学和物理学的“爱尔兰根纲领”启发，目前一批研究者认为：在视觉表征中嵌入任务所需的不变性或对称性结构是实现计算机视觉系统稳健性和可解释性的一种重要的原则性保障。如图 6.1 所示，本文由此展开：

第二章分析了经典不变量，作为本文的理论来源。理论层面，回顾该领域的发展历史（2.1 节）并梳理其中的核心知识和经典方法（2.2 节），重点分析近期在快速/精确计算、稳健性/不变性优化、定义扩展和应用的最新进展和成功经验（2.3 节）。实践层面，提供了系统性的开源实现 MomentToolbox，并且评估了众多经典不变量方法的精确性/复杂性、表征能力和稳健性/不变性（2.4 节）。第二章是对现有综述文献的有益补充，并作为后续章节创新性内容的知识基础。

第三章完善了全局不变量，为全局不变量带来了新的时频判别性以及噪声不变性（表 3.1）。方法命名为“分数阶拉东矩 FMR”。理论层面，首先详细讨论了如何分别通过隐式路径（3.3.1 节）和显式路径（3.4 节）来构造 FMR，前者直观、后者深入；随后，全面分析了 FMR 的广义性、旋转不变性、噪声稳健性和时频判别性（3.3.2 节）；最后，通过傅里叶变换和递归公式（3.3.3 节）设计了计算效率高且数值稳定的离散实现方案。应用层面，首先通过含噪模式分类实验充分验证了上述理论预期的性质；还将 FMR 直接应用于自然畸变下的模板匹配（3.5.4 节）和面向版权保护的零水印技术（3.5.5 节），FMR 整体性地提升了对各类信号和几何劣化的稳健性，证明了其在小规模可信视觉问题中的应用潜力。

第四章设计了局部不变量，将全局不变量的基本理论推广到更广义的局部假设（表 4.1）。方法命名为“局部不变表征 LIR”。理论层面，首先从基本公式上推广了矩和矩不变量理论，给出了具有局部表征能力的新定义 LIR（4.3.1 节）；随后，形式化讨论了 LIR 在几何不变性、等变性和协变性方面的核心性质，作为形成不变表征的指导原则（4.3.2 节）；最后，推导出了 LIR 的高效计算策略，具有低误差、常数阶复杂性以及基函数通用性的特点（4.3.3 节）。实践层面，给出了 LIR 的实践指南，特别是明确了在取证任务中的设置原则（4.4 节）。应用层面，首先通

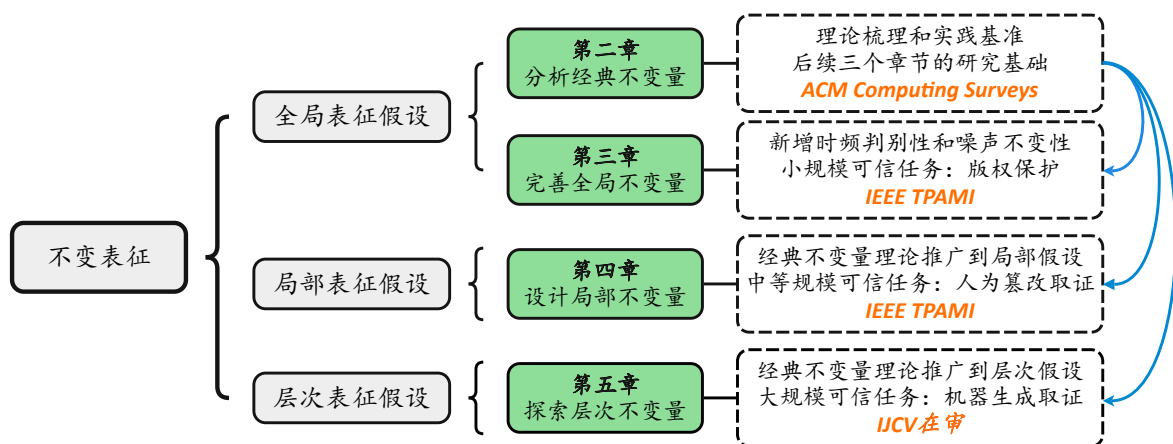


图 6.1 本文研究框架。

过稠密模式检测和匹配实验验证了上述理论预期的性质；还将 LIR 直接应用于复制-粘贴篡改检测（被动取证）和感知哈希（主动取证），LIR 实现了取证分数和效率的提升（4.5.3 和 4.5.4 节），证明了其在中等规模可信视觉问题中的应用潜力。

第五章探索了层次不变量，在更大规模的视觉任务中更好地权衡了不变性和判别性（表 5.1）。方法命名为“层次不变表征 HIR”。理论层面，首先形式化了层次不变性的蓝图，并重新审视了 CNN 的几种典型模块（5.3.1 节）；随后，定义了新的模块及其级联以满足层次不变性蓝图（5.3.2 节），特别是提供了关于图像与表征的几何对称性关系的形式化结论（5.3.3 节）；最后，探讨了上述新概念与典型概念之间的批判和发展关系，凸显了 HIR 在迈向稳健和可解释表征方面的独特性（5.3.5 节）。实践层面，提供了层次不变性理论的一个具体框架，涉及网络拓扑、层和参数的实践原则（5.4.1 节）；并且挖掘了上述框架的数据自适应潜力，通过特征/架构选择或级联学习模块（5.4.2 节）。应用层面，首先通过纹理、数字和寄生虫显微图像的模式分类实验，从不同问题规模和几何变体的下全面考察了上述理论预期的性质（5.5.1 节）；还将 HIR 直接应用于大规模的对抗性扰动和 AIGC 取证任务中，HIR 以“即插即用”的方式实现了一致良好的判别性、稳健性、效率和解释性，证明了其在大规模可信视觉问题中的应用潜力（5.5.2 节）。

总而言之，本文工作是经典不变量方法在理论、实践和应用层面上的扩展，尤其聚焦于增强不变表征的判别性。本文所提出的主要概念，即局部不变量和层次不变量，在不变量的理论研究中具有独创性，更加适用于多种实际的可信任务。

6.2 未来展望

宏观来看，在计算机视觉系统中嵌入物理世界的不变性和对称性先验的研究方面，所面临的主要困难是知识驱动和数据驱动的折衷融合问题。本文的未来工作旨在解决该问题：

- 知识驱动路径擅长以“构造”的方式定义具有不变结构的视觉表征（即满足物理先验），

但数学建模通常过于理想化，在实际的大规模问题中不能实现高精度的任务功能。为此，主要的研究目标是提升手工表征的判别性，借鉴数据驱动路径的成功经验，本文的层次不变量即属于此方面的初步尝试。为了打破判别性瓶颈，研究者应主要关注更广义的数学建模方法，将现有不变性的成功设计推广到更一般的假设上，即扩大不变量所在特征空间以允许潜在判别性特征的进一步选择。

- 数据驱动路径擅长以“学习”的方式自适应形成给定任务下高度判别的视觉表征（即满足任务功能），但学习过程通常不符合物理先验（或者说学习逼近先验的效率过低且不受保障），不能自动拓展至信任和安全攸关的场景。为此，主要的研究目标是赋予学习表征以不变性，借鉴知识驱动路径的成功经验。为了打破不变性瓶颈，研究者应主要关注可学习的不变结构，在深度网络的工程实践中抽象出其成功特质进行数学刻画，并且考虑物理先验进行再设计以限制网络过多的参数和过高的自由度。

参考文献

- [1] Wing J M. Trustworthy AI[J]. Commun. ACM, 2021, 64(10):64–71.
- [2] 张钹, 朱军, 苏航. 迈向第三代人工智能 [J]. 中国科学: 信息科学, 2020, 50(9):1281–1302.
- [3] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553):436–444.
- [4] Silver D, Schrittwieser J, Simonyan K, et al. Mastering the game of go without human knowledge[J]. Nature, 2017, 550(7676):354–359.
- [5] Wang H, Fu T, Du Y, et al. Scientific discovery in the age of artificial intelligence[J]. Nature, 2023, 620(7972):47–60.
- [6] Sundararajan K, Woodard D L. Deep learning for biometrics: A survey[J]. ACM Comput. Surv., 2018, 51(3):1–34.
- [7] Warnat-Herresthal S, Schultze H, Shastry K L, et al. Swarm learning for decentralized and confidential clinical machine learning[J]. Nature, 2021, 594(7862):265–270.
- [8] Feng S, Sun H, Yan X, et al. Dense reinforcement learning for safety validation of autonomous vehicles[J]. Nature, 2023, 615(7953):620–627.
- [9] Vosoughi S, Roy D, Aral S. The spread of true and false news online[J]. Science, 2018, 359(6380):1146–1151.
- [10] Bronstein M M, Bruna J, LeCun Y, et al. Geometric deep learning: going beyond euclidean data[J]. IEEE Signal Process. Mag., 2017, 34(4):18–42.
- [11] Bengio Y, Courville A, Vincent P. Representation learning: A review and new perspectives[J]. IEEE Trans. Pattern Anal. Mach. Intell., 2013, 35(8):1798–1828.
- [12] Klein F. A comparative review of recent researches in geometry[J]. Bull. Am. Math. Soc., 1893, 2(10):215–249.
- [13] Weyl H. Symmetry[M]. Princeton University Press, 2015.
- [14] Fukushima K, Miyake S. Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position[J]. Pattern Recognit., 1982, 15(6):455–469.
- [15] Balntas V, Lenc K, Vedaldi A, et al. H-Patches: A benchmark and evaluation of handcrafted and learned local descriptors.[J]. IEEE Trans. Pattern Anal. Mach. Intell., 2019, 42(11):2825–2841.
- [16] Qi S, Zhang Y, Wang C, et al. A survey of orthogonal moments for image representation: theory, implementation, and evaluation[J]. ACM Comput. Surv., 2021, 55(1):1–35.
- [17] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. Int. J. Comput. Vis., 2004, 60:91–110.
- [18] Tola E, Lepetit V, Fua P. Daisy: An efficient dense descriptor applied to wide-baseline stereo[J]. IEEE Trans. Pattern Anal. Mach. Intell., 2009, 32(5):815–830.
- [19] Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge[J]. Int. J. Comput. Vis., 2015, 115:211–252.
- [20] 周飞燕, 金林鹏, 董军, et al. 卷积神经网络研究综述 [J]. 计算机学报, 2017, 40(6):1229–1251.
- [21] Zhang R. Making convolutional networks shift-invariant again[C]. Proceedings of Int. Conf. Mach. Learn., 2019. 7324–7334.

- [22] Garrido Q, Assran M, Ballas N, et al. Learning and leveraging world models in visual representation learning[J]. arXiv preprint arXiv:2403.00504, 2024..
- [23] Buckner C. Understanding adversarial examples requires a theory of artefacts for deep learning[J]. *Nature Mach. Intell.*, 2020, 2(12):731–736.
- [24] 董胤蓬, 苏航, 朱军. 面向对抗样本的深度神经网络可解释性分析 [J]. *自动化学报*, 2022, 48(1):75–86.
- [25] Taddeo M, McCutcheon T, Floridi L. Trusting artificial intelligence in cybersecurity is a double-edged sword[J]. *Nature Mach. Intell.*, 2019, 1(12):557–560.
- [26] Hu M K. Visual pattern recognition by moment invariants[J]. *IRE Trans. Inf. Theory*, 1962, 8(2):179–187.
- [27] Mukundan R, Ramakrishnan K. *Moment Functions in Image Analysis: Theory and Applications*[M]. World Scientific, 1998.
- [28] Pawlak M. *Image Analysis by Moments: Reconstruction and Computational Aspects*[M]. Oficyna Wydawnicza Politechniki Wrocawskiej, 2006.
- [29] Shu H, Luo L, Coatrieux J L. Moment-based approaches in imaging part 1: Basic features[J]. *IEEE Eng. Med. Biol.*, 2007, 26(5):70–74.
- [30] Shu H, Luo L, Coatrieux J L. Moment-based approaches in imaging part 2: Invariance[J]. *IEEE Eng. Med. Biol.*, 2008, 27(1):81–83.
- [31] Shu H, Luo L, Coatrieux J L. Moment-based approaches in imaging part 3: Computational considerations[J]. *IEEE Eng. Med. Biol.*, 2008, 27(3):89–91.
- [32] Flusser J, Zitova B, Suk T. *Moments and Moment Invariants in Pattern Recognition*[M]. John & Sons: Wiley, 2009.
- [33] Flusser J, Suk T, Zitov B. *2D and 3D Image Analysis by Moments*[M]. John & Sons: Wiley, 2016.
- [34] Hoang T V. *Image representations for pattern recognition*[D]. Dept. Comput. Sci., Nancy 2 Univ., Nancy, France, 2011.
- [35] Papakostas G A. Over 50 years of image moments and moment invariants[M]. . In: Papakostas G A, (eds.). *Proceedings of Moments and Moment Invariants: Theory and Applications*. Science Gate, 2014: 3–32.
- [36] Kaur P, Pannu H S, Malhi A K. Comprehensive study of continuous orthogonal moments-a systematic review[J]. *ACM Comput. Surv.*, 2019, 52:4.
- [37] Teh C H, Chin R T. On image analysis by the methods of moments[J]. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1988, 10(4):496–513.
- [38] Abu-Mostafa Y S, Psaltis D. Recognitive aspects of moment invariants[J]. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1984, 6(6):698–706.
- [39] Zhang D, Lu G. Shape-based image retrieval using generic Fourier descriptor[J]. *Signal Process.-Image Commun.*, 2002, 17(10):825–848.
- [40] Mukundan R, Ong S H, Lee P A. Image analysis by Tchebichef moments[J]. *IEEE Trans. Image Process.*, 2001, 10(9):1357–1364.
- [41] Yap P T, Paramesran R, Ong S H. Image analysis by Krawtchouk moments[J]. *IEEE Trans. Image Process.*, 2003, 12(11):1367–1377.
- [42] Yap P T, Paramesran R, Ong S H. Image analysis using Hahn moments[J]. *IEEE Trans. Pattern*

- Anal. Mach. Intell., 2007, 29(11):2057–2062.
- [43] Zhu H, Shu H, Zhou J, et al. Image analysis by discrete orthogonal dual Hahn moments[J]. Pattern Recognit. Lett., 2007, 28(13):1688–1704.
- [44] Zhu H, Shu H, Liang J, et al. Image analysis by discrete orthogonal Racah moments[J]. Signal Process., 2007, 87(4):687–708.
- [45] Bhatia A B, Wolf E. On the circle polynomials of Zernike and related orthogonal sets[J]. Math. Camb. Philos. Soc., 1954. 40–48.
- [46] Teague M R. Image analysis via the general theory of moments[J]. J. Opt. Soc. Am. A., 1980, 70(8):920–30.
- [47] Shen J. Orthogonal Gaussian-Hermite moments for image characterization[C]. Proceedings of SPIE Intell. Rob. Comput. Vis., 1997. 224–233.
- [48] Hosny K M. Image representation using accurate orthogonal Gegenbauer moments[J]. Pattern Recognit. Lett., 2011, 32(6):795–804.
- [49] Zhu H. Image representation using separable two-dimensional continuous and discrete orthogonal moments[J]. Pattern Recognit., 2012, 45(4):1540–1558.
- [50] Sheng Y, Shen L. Orthogonal Fourier-Mellin moments for invariant pattern recognition[J]. J. Opt. Soc. Am. A., 1994, 11(6):1748–57.
- [51] Ping Z, Wu R, Sheng Y. Image description with Chebyshev-Fourier moments[J]. J. Opt. Soc. Am. A., 2002, 19(9):1748–1754.
- [52] Amu G, Hasi S, Yang X, et al. Image analysis by pseudo-Jacobi ($p=4$, $q=3$)-Fourier moments[J]. Appl. Optics, 2004, 43(10):2093–2101.
- [53] Ping Z, Ren H, Zou J, et al. Generic orthogonal moments: Jacobi-Fourier moments for invariant image description[J]. Pattern Recognit., 2007, 40(4):1245–1254.
- [54] Hoang T V, Tabbone S. Errata and comments on 'Generic orthogonal moments: Jacobi-Fourier moments for invariant image description'[J]. Pattern Recognit., 2013, 46(11):3148–3155.
- [55] Ren H, Ping Z, Bo W, et al. Multidistortion-invariant image recognition with radial harmonic Fourier moments[J]. J. Opt. Soc. Am. A., 2003, 20(4):631–637.
- [56] Hu H, Zhang Y, Shao C, et al. Orthogonal moments based on exponent functions: Exponent-Fourier moments[J]. Pattern Recognit., 2014, 47(8):2596–2606.
- [57] Yap P T, Jiang X, Kot A C. Two-dimensional polar harmonic transforms for invariant image representation[J]. IEEE Trans. Pattern Anal. Mach. Intell., 2010, 32(7):1259–1270.
- [58] Xiao B, Ma J, Wang X. Image analysis by Bessel-Fourier moments[J]. Pattern Recognit., 2010, 43(8):2620–2629.
- [59] Liao S, Pawlak M. On the accuracy of Zernike moments for image analysis[J]. IEEE Trans. Pattern Anal. Mach. Intell., 1998, 20(12):1358–1364.
- [60] Chong C W, Raveendran P, R Mukundan . comparative analysis of algorithms for fast computation of Zernike moments[J]. Pattern Recognit., 2003, 36(3):731–742.
- [61] Wee C Y, Paramesran R. On the computational aspects of Zernike moments[J]. Image Vis. Comput., 2007, 25(6):967–980.
- [62] Singh C, Walia E. Algorithms for fast computation of Zernike moments and their numerical stability[J]. Image Vis. Comput., 2011, 29(4):251–259.
- [63] Wang X, Liao S. Image reconstruction from orthogonal Fourier-Mellin moments[C]. Proceedings

- of Int. Conf. Image Analysis and Recognit., 2013. 687–694.
- [64] Nwali M, Liao S. A new fast algorithm to compute continuous moments defined in a rectangular region[J]. Pattern Recognit., 2019, 89:151–160.
- [65] Prata A, Rusch W V. Algorithm for computation of Zernike polynomials expansion coefficients[J]. Appl. Optics, 1989, 28(4):749–54.
- [66] Singh C, Upneja R. Accurate calculation of high order pseudo-Zernike moments and their numerical stability[J]. Digit. Signal Prog., 2014, 27:95–106.
- [67] Singh C, Upneja R. Accurate computation of orthogonal Fourier-Mellin moments[J]. J. Math. Imaging Vis., 2012, 44(3):411–431.
- [68] Xin Y, Pawlak M, Liao S. Accurate computation of Zernike moments in polar coordinates[J]. IEEE Trans. Image Process., 2007, 16(2):581–587.
- [69] Hosny K M, Shouman M A, Salam H M A. Fast computation of orthogonal Fourier-Mellin moments in polar coordinates[J]. J. Real-Time Image Process., 2011, 6(2):73–80.
- [70] Camacho-Bello C, Toxqui-Quitl C, Padilla-Vivanco A, et al. High-precision and fast computation of Jacobi-Fourier moments for image description[J]. J. Opt. Soc. Am. A., 2014, 31(1):124–134.
- [71] Mukundan R, Ramakrishnan K R. Fast computation of Legendre and Zernike moments[J]. Pattern Recognit., 1995, 28(9):1433–42.
- [72] Ruberto C D, Putzu L, Rodriguez G. Fast and accurate computation of orthogonal moments for texture analysis[J]. Pattern Recognit., 2018, 83:498–510.
- [73] Upneja R, Singh C. Fast computation of Jacobi-Fourier moments for invariant image recognition[J]. Pattern Recognit., 2015, 48(5):1836–1843.
- [74] Papakostas G A, Boutalis Y S, Karras D A, et al. A new class of Zernike moments for computer vision applications[J]. Inf. Sci., 2007, 177(13):2802–2819.
- [75] Papakostas G A, Boutalis Y S, Karras D A, et al. Modified factorial-free direct methods for Zernike and pseudo-Zernike moment computation[J]. IEEE Trans. Instrum. Meas., 2009, 58(7):2121–2131.
- [76] Saez-Landete J. Comments on 'Fast computation of Jacobi-Fourier moments for invariant image recognition'[J]. Pattern Recognit., 2017, 67:16–22.
- [77] Hosny K M, Darwish M M. A kernel-based method for fast and accurate computation of PHT in polar coordinates[J]. J. Real-Time Image Process., 2019, 16(4):1235–1247.
- [78] Averbuch A, Coifman R R, Donoho D L, et al. Fast and accurate polar Fourier transform[J]. Appl. Comput. Harmon. Anal., 2006, 21(2):145–167.
- [79] Yang H, Qi S, Wang C, et al. Image analysis by log-polar Exponent-Fourier moments[J]. Pattern Recognit., 2020, 101.
- [80] Hoang T V, Tabbone S. Fast generic polar harmonic transforms[J]. IEEE Trans. Image Process., 2014, 23(7):2961–2971.
- [81] Singh C, Ranade S K. A high capacity image adaptive watermarking scheme with radial harmonic Fourier moments[J]. Digit. Signal Prog., 2013, 23(5):1470–1482.
- [82] Hoang T V, Tabbone S. Fast computation of orthogonal polar harmonic transforms[C]. Proceedings of Int. Conf. Pattern Recognit., 2012. 3160–3163.
- [83] Singh C, Kaur A. Fast computation of polar harmonic transforms[J]. J. Real-Time Image Process., 2015, 10(1):59–66.

- [84] Yang H, Qi S, Niu P, et al. Color image zero-watermarking based on fast quaternion generic polar complex exponential transform[J]. *Signal Process.-Image Commun.*, 2020, 82.
- [85] Wang C, Wang X, Xia Z. Geometrically invariant image watermarking based on fast Radial Harmonic Fourier Moments[J]. *Signal Process.-Image Commun.*, 2016, 45:10–23.
- [86] Wang X, Wang C, Yang H, et al. Robust and effective multiple copy-move forgeries detection and localization[J]. *Pattern Anal. Appl.*, 2021..
- [87] Ping Z, Jiang Y, Zhou S, et al. FFT algorithm of complex exponent moments and its application in image recognition[J]. *SPIE Int. Conf. Digit. Image Process.*, 2014. 4177–4180.
- [88] Singh S P, Urooj S. Accurate and fast computation of Exponent-Fourier moment[J]. *Arab. J. Sci. Eng.*, 2017, 42(8):3299–3306.
- [89] Hwang S K, Kim W Y. A novel approach to the fast computation of Zernike moments[J]. *Pattern Recognit.*, 2006, 39(11):2065–2076.
- [90] Al-Rawi M S. Fast computation of pseudo Zernike moments[J]. *J. Real-Time Image Process.*, 2010, 5(1):3–10.
- [91] Chen B, Coatrieux G, Wu J, et al. Fast computation of sliding discrete Tchebichef moments and its application in duplicated regions detection[J]. *IEEE Trans. Signal Process.*, 2015, 63(20):5424–5436.
- [92] Martinez J, Thomas F. Efficient computation of local geometric moments[J]. *IEEE Trans. Image Process.*, 2002, 11(9):1102–1111.
- [93] Bera A, Klesk P, Sychel D. Constant-time calculation of Zernike moments for detection with rotational invariance[J]. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2019, 41(3):537–551.
- [94] Bay H, Tuytelaars T, Gool L V. SURF: Speeded up robust features[C]. *Proceedings of Eur. Conf. Comput. Vision.*, 2006. 404–417.
- [95] Benouni R, Batioua I, Zenkour K, et al. Fast and accurate computation of Racah moment invariants for image classification[J]. *Pattern Recognit.*, 2019, 91:100–110.
- [96] Pei S, Lin C. Image normalization for pattern recognition[J]. *Image Vis. Comput.*, 1995, 13(10):711–723.
- [97] Dinggang S, Ip H H S. Generalized affine invariant image normalization[J]. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1997, 19(5):431–40.
- [98] Dongseok S, Pollard J K, Muller J P. Accurate geometric correction of ATSR images[J]. *IEEE Trans. Geosci. Remote Sensing.*, 1997, 35(4):997–1006.
- [99] Wang C, Wang X, Zhang C, et al. Geometric correction based color image watermarking using fuzzy least squares support vector machine and Bessel K form distribution[J]. *Signal Process.*, 2017, 134:197–208.
- [100] Brown L G. A survey of image registration techniques[J]. *ACM Comput. Surv.*, 1992, 24(4):325–376.
- [101] Zitova B, Flusser J. Image registration methods: A survey[J]. *Image Vis. Comput.*, 2003, 21(11):977–1000.
- [102] Chong C W, Raveendran P, Mukundan R. Translation invariants of Zernike moments[J]. *Pattern Recognit.*, 2003, 36(8):1765–1773.
- [103] Karakasis E G, Papakostas G A, Koulouriotis D E, et al. Generalized dual Hahn moment invariants[J]. *Pattern Recognit.*, 2013, 46(7):1998–2014.

- [104] Flusser J, Suk T. Pattern recognition by affine moment invariants[J]. *Pattern Recognit.*, 1993, 26(1):167–74.
- [105] Chong C W, Raveendran P, Mukundan R. Translation and scale invariants of Legendre moments[J]. *Pattern Recognit.*, 2004, 37(1):119–129.
- [106] Belkasim S, Hassan E, Obeidi T. Explicit invariance of Cartesian Zernike moments[J]. *Pattern Recognit. Lett.*, 2007, 28(15):1969–1980.
- [107] Zhu H, Shu H, Xia T, et al. Translation and scale invariants of Tchebichef moments[J]. *Pattern Recognit.*, 2007, 40(9):2530–2542.
- [108] Oppenheim A V, Lim J S. The importance of phase in signals[J]. *Proc. IEEE*, 1981, 69(5):529–41.
- [109] Revaud J, Lavoue G, Baskurt A. Improving Zernike moments comparison for optimal similarity and rotation angle retrieval[J]. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2009, 31(4):627–636.
- [110] Flusser J. On the independence of rotation moment invariants[J]. *Pattern Recognit.*, 2000, 33(9):1405–1410.
- [111] Shao L, Yan R, Li X, et al. From heuristic optimization to dictionary learning: A review and comprehensive comparison of image denoising algorithms[J]. *IEEE T. Cybern.*, 2014, 44(7):1001–1013.
- [112] Zhang K, Zuo W, Chen Y, et al. Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising[J]. *IEEE Trans. Image Process.*, 2017, 26(7):3142–3155.
- [113] Xiao B, Cui J, Qin H, et al. Moments and moment invariants in the Radon space[J]. *Pattern Recognit.*, 2015, 48(9):2772–2784.
- [114] Hoang T V, Tabbone S. Invariant pattern recognition using the RFM descriptor[J]. *Pattern Recognit.*, 2012, 45(1):271–284.
- [115] Miao Q, Liu J, Li W, et al. Three novel invariant moments based on radon and polar harmonic transforms[J]. *Opt. Commun.*, 2012, 285(6):1044–1048.
- [116] Jafari-Khouzani K, Soltanian-Zadeh H. Rotation-invariant multiresolution texture analysis using Radon and wavelet transforms[J]. *IEEE Trans. Image Process.*, 2005, 14(6):783–795.
- [117] Kundur D, Hatzinakos D. Blind image deconvolution[J]. *IEEE Signal Process. Mag.*, 1996, 13(3):43–64.
- [118] Campisi P, Egiazarian K. *Blind Image Deconvolution: Theory and Applications*[M]. CRC press, 2017.
- [119] Kostkova J, Flusser J, Lebl M, et al. Handling Gaussian blur without deconvolution[J]. *Pattern Recognit.*, 2020, 103.
- [120] Flusser J, Suk T, Boldys J, et al. Projection operators and moment invariants to image blurring[J]. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2015, 37(4):786–802.
- [121] Flusser J, Farokhi S, Hoschl C, et al. Recognition of images degraded by Gaussian blur[J]. *IEEE Trans. Image Process.*, 2016, 25(2):790–806.
- [122] Pedone M, Flusser J, Heikkila J. Registration of images with N-Fold dihedral blur[J]. *IEEE Trans. Image Process.*, 2015, 24(3):1036–1045.
- [123] Suk T, Flusser J. Tensor method for constructing 3D moment invariants[C]. *Proceedings of Int. Conf. Comput. Anal. Images Patterns*, 2011. 212–219.
- [124] Suk T, Flusser J. Graph method for generating affine moment invariants[C]. *Proceedings of Int. Conf. Pattern Recognit.*, 2004. 192–195.

- [125] Lo C H, Don H S. 3D moment forms their construction and application to object identification and positioning[J]. IEEE Trans. Pattern Anal. Mach. Intell., 1989, 11(10):1053–64.
- [126] Mamistvalov G. N-dimensional moment invariants and conceptual mathematical theory of recognition n-dimensional solids[J]. IEEE Trans. Pattern Anal. Mach. Intell., 1998, 20(8):819–831.
- [127] Xu D, Li H. Geometric moment invariants[J]. Pattern Recognit., 2008, 41(1):240–249.
- [128] Li E, Li H. Reflection invariant and symmetry detection[J]. arXiv preprint, 2017..
- [129] Li E, Mo H, Xu D, et al. Image projective invariants[J]. IEEE Trans. Pattern Anal. Mach. Intell., 2019, 41(5):1144–1157.
- [130] He Z, Hanlin M, You H, et al. Differential and integral invariants under Mobius transformation[C]. Proceedings of Chinese Conf. Pattern Recognit. Comput. Vis., 2018. 280–291.
- [131] Zhang H, Mo H, Hao Y, et al. Fast and efficient calculations of structural invariants of chirality[J]. Pattern Recognit. Lett., 2019, 128:270–277.
- [132] You H, Hanlin M, Qi L, et al. Dual affine moment invariants[J]. arXiv preprint, 2019..
- [133] Xu D, Li H. 3D curve moment invariants for curve recognition[C]. Proceedings of Intell. Comput. Signal Process. Pattern Recognit., 2006. 572–577.
- [134] Xu D, Li H. 3D surface moment invariants[C]. Proceedings of Int. Conf. Pattern Recognit., 2006. 173–176.
- [135] Kostkova J, Suk T, Flusser J. Affine invariants of vector fields[J]. IEEE Trans. Pattern Anal. Mach. Intell., 2019..
- [136] Kostkova J, Suk T, Flusser J. Affine moment invariants of vector fields[C]. Proceedings of Int. Conf. Image Process., 2018. 1338–1342.
- [137] Langbein M, Hagen H. A generalization of moment invariants on 2D vector fields to tensor fields of arbitrary order and dimension[C]. Proceedings of Int. Symp. Vis. Comput., 2009. 1151–1160.
- [138] Bujack R, Kasten J, Hotz I, et al. Moment invariants for 3D flow fields via normalization[C]. Proceedings of IEEE Pacific Vis. Symp., 2015. 9–16.
- [139] Bujack R, Hagen H. Moment invariants for multi-dimensional data[M]. . In: Ozerslan E, Schultz T, Hotz I, (eds.). Proceedings of Modelling, Analysis, and Visualization of Anisotropy. Mathematics and Visualization, 2017: 43–64.
- [140] Li E, Huang Y, Xu D, et al. Shape DNA: Basic generating functions for geometric moment invariants[J]. arXiv preprint, 2017..
- [141] Li E, Li H. Isomorphism between differential and moment invariants under affine transform[J]. arXiv preprint, 2017..
- [142] Pei Y, Huang Y, Zou Q, et al. Effects of image degradation and degradation removal to CNN-based image classification[J]. IEEE Trans. Pattern Anal. Mach. Intell., 2019..
- [143] Flusser J, Suk T, Saic S. Recognition of images degraded by linear motion blur without restoration[M]. . In: Kropatsch W, Klette R, Solina F, et al., (eds.). Proceedings of Theoretical Foundations of Computer Vision. Computing Supplement, 1996: 37–51.
- [144] Flusser J, Suk T, Saic S. Image features invariant with respect to blur[J]. Pattern Recognit., 1995, 28(11):1723–32.
- [145] Flusser J, Zitova B. Invariants to convolution with circularly symmetric PSF[C]. Proceedings of Int. Conf. Pattern Recognit., 2004. 11–14.
- [146] Chen B, Shu H, Zhang H, et al. Combined invariants to similarity transformation and to blur

- using orthogonal Zernike moments[J]. *IEEE Trans. Image Process.*, 2011, 20(2):345–360.
- [147] Flusser J, Lébl M, Šroubek F, et al. Blur invariants for image recognition[J]. *Int. J. Comput. Vis.*, 2023, 131(9):2298–2315.
- [148] Karakasis E G, Papakostas G A, Koulouriotis D E, et al. A unified methodology for computing accurate quaternion color moments and moment invariants[J]. *IEEE Trans. Image Process.*, 2014, 23(2):596–611.
- [149] Hosny K M, Darwish M M. New set of multi-channel orthogonal moments for color image representation and recognition[J]. *Pattern Recognit*, 2019, 88:153–173.
- [150] Chen B, Shu H, Zhang H, et al. Quaternion Zernike moments and their invariants for color image analysis and object recognition[J]. *Signal Process.*, 2012, 92(2):308–318.
- [151] Chen B, Shu H, Coatrieux G, et al. Color image analysis by quaternion-type moments[J]. *J. Math. Imaging Vis.*, 2015, 51(1):124–144.
- [152] Batioua I, Benouini R, Zenkouar K, et al. 3D image representation using separable discrete orthogonal moments[J]. *Procedia Comput. Sci.*, 2019, 148:389–398.
- [153] Kazhdan M, Funkhouser T, Rusinkiewicz S. Rotation invariant spherical harmonic representation of 3D shape descriptors[C]. *Proceedings of Symp. Geom. Process.*, 2003. 156–164.
- [154] Novotni M, Klein R. Shape retrieval using 3D Zernike descriptors[J]. *Comput.-Aided Des.*, 2004, 36(11):1047–1062.
- [155] Canterakis N. 3D Zernike moments and Zernike affine invariants for 3D image analysis and recognition[C]. *Proceedings of Scandinavian Conf. Image Analysis*, 1999. 85–93.
- [156] Hoang T V, Tabbone S. Generic polar harmonic transforms for invariant image description[C]. *Proceedings of Int. Conf. Image Process.*, 2011. 829–832.
- [157] Hoang T V, Tabbone S. Generic polar harmonic transforms for invariant image representation[J]. *Image Vis. Comput.*, 2014, 32(8):497–509.
- [158] Xiao B, Li L, Li Y, et al. Image analysis by fractional-order orthogonal moments[J]. *Inf. Sci.*, 2017, 382:135–149.
- [159] Yang H, Qi S, Tian J, et al. Robust and discriminative image representation: Fractional-order Jacobi-Fourier moments[J]. *Pattern Recognit.*, 2021, 115.
- [160] Yang T, Ma J, Miao Y, et al. Quaternion weighted spherical Bessel-Fourier moment and its invariant for color image reconstruction and object recognition[J]. *Inf. Sci.*, 2019, 505:388–405.
- [161] Wang C, Wang X, Li Y, et al. Quaternion polar polar harmonic Fourier moments for color images[J]. *Inf. Sci.*, 2018, 450:141–156.
- [162] Wang X, Li W, Yang H, et al. Invariant quaternion radial harmonic Fourier moments for color image retrieval[J]. *Opt. Laser Technol.*, 2015, 66:78–88.
- [163] Wang X, Niu P, Yang H, et al. A new robust color image watermarking using local quaternion exponent moments[J]. *Inf. Sci.*, 2014, 277:731–754.
- [164] Shao Z, Shu H, Wu J, et al. Quaternion Bessel-Fourier moments and their invariant descriptors for object reconstruction and recognition[J]. *Pattern Recognit.*, 2014, 47(2):603–611.
- [165] Wang J, Li T, Luo X, et al. Identifying computer generated images based on quaternion central moments in color quaternion wavelet domain[J]. *IEEE Trans. Circuits Syst. Video Technol.*, 2019, 29(9):2775–2785.
- [166] Said S, Bihan N L, Sangwine S J. Fast complexified quaternion Fourier transform[J]. *IEEE Trans.*

- Signal Process., 2008, 56(4):1522–1531.
- [167] Bulow T, Sommer G. Hypercomplex signals-a novel extension of the analytic signal to the multidimensional case[J]. IEEE Trans. Signal Process., 2001, 49(11):2844–2852.
 - [168] Chen Y, Xiao X, Zhou Y. Low-rank quaternion approximation for color image processing[J]. IEEE Trans. Image Process., 2020, 29:1426–1439.
 - [169] Zhang S, Tay Y, Yao L, et al. Quaternion knowledge graph embeddings[C]. Proceedings of Adv. Neural Inf. Process. Syst., 2019. 2731–2741.
 - [170] Liu Y, Zheng Y, Lu J, et al. Constrained quaternion-variable convex optimization: a quaternion-valued recurrent neural network approach[J]. IEEE Trans. Neural Netw. Learn. Syst., 2020, 31(3):1022–1035.
 - [171] Zhu X, Xu Y, Xu H, et al. Quaternion convolutional neural networks[C]. Proceedings of Eur. Conf. Comput. Vision, 2018. 645–661.
 - [172] Chen B, Qi X, Sun X, et al. Quaternion pseudo-Zernike moments combining both of RGB information and depth information for color image splicing detection[J]. J. Vis. Commun. Image Represent., 2017, 49:283–290.
 - [173] Yamni M, Karmouni H, Sayyouri M, et al. Novel octonion moments for color stereo image analysis[J]. Digit. Signal Prog., 2021, 108.
 - [174] Xu D, Mandic D P. The theory of quaternion matrix derivatives[J]. IEEE Trans. Signal Process., 2015, 63(6):1543–1556.
 - [175] Batioua I, Benouini R, Zenkour K, et al. 3D image analysis by separable discrete orthogonal moments based on Krawtchouk and Tchebichef polynomials[J]. Pattern Recognit., 2017, 71:264–277.
 - [176] Pozo J M, Villa-Uriol M C, Frangi A F. Efficient 3D geometric and Zernike moments computation from unstructured surface meshes[J]. IEEE Trans. Pattern Anal. Mach. Intell., 2011, 33(3):471–484.
 - [177] Karmouni H, Jahid T, Sayyouri M, et al. Fast 3D image reconstruction by cuboids and 3D Charlier’s moments[J]. J. Real-Time Image Process., 2020, 17(4):949–965.
 - [178] Daoui A, Yamni M, Ogri O E, et al. New algorithm for large-sized 2D and 3D image reconstruction using higher-order Hahn moments[J]. Circuits Syst. Signal Process., 2020, 39(9):4552–4577.
 - [179] Benouini R, Batioua I, Zenkour K, et al. Efficient 3D object classification by using direct Krawtchouk moment invariants[J]. Multimed. Tools Appl., 2018, 77(20):27517–27542.
 - [180] Batioua I, Benouini R, Zenkour K. Image recognition using new set of separable three-dimensional discrete orthogonal moment invariants[J]. Multimed. Tools Appl., 2020, 79(19-20):13217–13245.
 - [181] Suetens P. Fundamentals of Medical Imaging[M]. Cambridge University Press, 2017.
 - [182] Marschner S, Shirley P. Fundamentals of computer graphics[M]. CRC Press, 2018.
 - [183] Xiao B, Luo J, Bi X, et al. Fractional discrete Tchebyshev moments and their applications in image encryption and watermarking[J]. Inf. Sci., 2020, 516:545–559.
 - [184] Yamni M, Daoui A, Ogri O E, et al. Fractional Charlier moments for image reconstruction and image watermarking[J]. Signal Process., 2020, 171.
 - [185] Chen B, Yu M, Su Q, et al. Fractional quaternion cosine transform and its application in color image copy-move forgery detection[J]. Multimed. Tools Appl., 2019, 78(7):8057–8073.

- [186] Benouini R, Batioua I, Zenkour K, et al. Fractional-order orthogonal Chebyshev moments and moment invariants for image representation and pattern recognition[J]. *Pattern Recognit.*, 2019, 86:332–343.
- [187] Hosny M, Darwish M M, Aboelenen T. New fractional-order Legendre-Fourier moments for pattern recognition applications[J]. *Pattern Recognit.*, 2020, 103.
- [188] Zhang H, Li Z, Liu Y. Fractional orthogonal Fourier-Mellin moments for pattern recognition[C]. *Proceedings of Chinese Conf. Pattern Recognit.*, 2016. 766–778.
- [189] Chen B, Yu M, Su Q, et al. Fractional quaternion Zernike moments for robust color image copy-move forgery detection[J]. *IEEE Access*, 2018, 6:56637–56646.
- [190] Hosny M, Darwish M M, Aboelenen T. Novel fractional-order generic Jacobi-Fourier moments for image analysis[J]. *Signal Process.*, 2020, 172.
- [191] Hosny M, Darwish M M, Aboelenen T. Novel fractional-order polar harmonic transforms for gray-scale and color image analysis[J]. *J. Frankl. Inst.-Eng. Appl. Math.*, 2020, 357(4):2533–2560.
- [192] Hosny M, Darwish M M, Eltoukhy M M. Novel multi-channel fractional-order radial harmonic Fourier moments for color image analysis[J]. *IEEE Access*, 2020, 8:40732–40743.
- [193] Hosny M, Elaziz M A, Darwish M M. Color face recognition using novel fractional-order multi-channel exponent moments[J]. *Neural Comput. Appl.*, 2020..
- [194] Qu Y, Cui C, Chen S, et al. A fast subpixel edge detection method using Sobel-Zernike moments operator[J]. *Image Vis. Comput.*, 2005, 23(1):11–17.
- [195] Iscen A, Tolias G, Gosselin P H, et al. A comparison of dense region detectors for image search and fine-grained classification[J]. *IEEE Trans. Image Process.*, 2015, 24(8):2369–2381.
- [196] Chen Z, Sun S K. A Zernike moment phase-based descriptor for local image representation and matching[J]. *IEEE Trans. Image Process.*, 2010, 19(1):205–219.
- [197] Schlemmer M, Heringer M, Morr F, et al. Moment invariants for the analysis of 2D flow fields[J]. *IEEE Trans. Vis. Comput. Graph.*, 2007, 13(6):1743–1750.
- [198] Kumar A, Ahmad M O, Swamy M N S. Tchebichef and adaptive steerable-based total variation model for image denoising[J]. *IEEE Trans. Image Process.*, 2019, 28(6):2921–2935.
- [199] Kumar A. Deblurring of motion blurred images using histogram of oriented gradients and geometric moments[J]. *Signal Process.-Image Commun.*, 2017, 55:55–65.
- [200] Gao X, Wang Q, Li X, et al. Zernike-moment-based image super resolution[J]. *IEEE Trans. Image Process.*, 2011, 20(10):2738–2747.
- [201] Yang Z, Cohen F S. Cross-weighted moments and affine invariants for image registration and matching[J]. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1999, 21(8):804–814.
- [202] Xiao G L, Zhang Y, Li W, et al. Lossless image compression based on integer discrete Tchebichef transform[J]. *Neurocomput.*, 2016, 214:587–593.
- [203] Li L, Lin W, Wang X, et al. No-reference image blur assessment based on discrete orthogonal moments[J]. *IEEE T. Cybern.*, 2016, 46(1):39–50.
- [204] Liang M, Du J, Li L, et al. Video super-resolution reconstruction based on deep learning and spatio-temporal feature self-similarity[J]. *IEEE Trans. Knowl. Data Eng.*, 2020..
- [205] Bronstein M, Bronstein M M, Guibas L J, et al. Shape Google: Geometric words and expressions for invariant shape retrieval[J]. *ACM Trans. Graph.*, 2011, 30:1.

- [206] Murtagh F, Starck J L. Wavelet and curvelet moments for image classification: Application to aggregate mixture grading[J]. *Pattern Recognit. Lett.*, 2008, 29(10):1557–1564.
- [207] Paschos G, Radev I, Prabakar N. Image content-based retrieval using chromaticity moments[J]. *IEEE Trans. Knowl. Data Eng.*, 2003, 15(5):1069–1072.
- [208] Zhao L, Davis L S. Closely coupled object detection and segmentation[C]. *Proceedings of IEEE Int. Conf. Comput. Vis.*, 2005. 454–461.
- [209] Wang X, Wu Z, Chen L, et al. Pixel classification based color image segmentation using quaternion exponent moments[J]. *Neural Netw.*, 2016, 74:1–13.
- [210] Ioannidis D, Tzovaras D, Damousis I G, et al. Gait recognition using compact feature extraction transforms and depth information[J]. *IEEE Trans. Inf. Forensic Secur.*, 2007, 2(3):623–630.
- [211] Vincent N, Ogier J M. Shall deep learning be the mandatory future of document analysis problems?[J]. *Pattern Recognit.*, 2019, 86:281–289.
- [212] Liu M, Yap P T. Invariant representation of orientation fields for fingerprint indexing[J]. *Pattern Recognit.*, 2012, 45(7):2532–2542.
- [213] Lajevardi S M, Hussain Z M. Higher order orthogonal moments for invariant facial expression recognition[J]. *Digit. Signal Prog.*, 2010, 20(6):1771–1779.
- [214] Tsougenis D, Papakostas G A, Koulouriotis D E, et al. Performance evaluation of moment-based watermarking methods: A review[J]. *J. Syst. Softw.*, 2012, 85(8):1864–1884.
- [215] Zhang Y, Luo X, Guo Y, et al. Zernike moment-based spatial image steganography resisting scaling attack and statistic detection[J]. *IEEE Access*, 2019, 7:24282–24289.
- [216] Zhao Y, Wang S, Zhang X, et al. Robust hashing for image authentication using Zernike moments and local features[J]. *IEEE Trans. Inf. Forensic Secur.*, 2013, 8(1):55–63.
- [217] Cozzolino D, Poggi G, Verdoliva L. Efficient dense-field copy-move forgery detection[J]. *IEEE Trans. Inf. Forensic Secur.*, 2015, 10(11):2284–2297.
- [218] Dai X, Shu H, Luo L, et al. Reconstruction of tomographic images from limited range projections using discrete Radon transform and Tchebichef moments[J]. *Pattern Recognit.*, 2010, 43(3):1152–1164.
- [219] Dai X, Khorram S. A feature-based image registration algorithm using improved chain-code representation combined with invariant moments[J]. *IEEE Trans. Geosci. Remote Sensing*, 1999, 37(5):2351–2362.
- [220] Chaumette F. Image moments: A general and useful set of features for visual servoing[J]. *IEEE Trans. Robot. Autom.*, 2004, 20(4):713–723.
- [221] Casasent D, Psaltis D. New optical transforms for pattern recognition[J]. *IEEE*, 1977, 65(1):77–84.
- [222] Zhai L, Li B, Chen J, et al. Chemical image moments and their applications[J]. *Trac-Trends Anal. Chem.*, 2018, 103:119–125.
- [223] Sit A, Shin W H, Kihara D. Three-dimensional Krawtchouk descriptors for protein local surface shape comparison[J]. *Pattern Recognit.*, 2019, 93:534–545.
- [224] Uhrin M. Through the eyes of a descriptor: Constructing complete, invertible, descriptions of atomic environments[J]. *arXiv preprint*, 2021..
- [225] Xin Y, Liao S, Pawlak M. Circularly orthogonal moments for geometrically robust image watermarking[J]. *Pattern Recognit.*, 2007, 40(12):3740–3752.

- [226] Li L, Li S, Abraham A, et al. Geometrically invariant image watermarking using polar harmonic transforms[J]. *Inf. Sci.*, 2012, 199:1–19.
- [227] Wang C, Wang X, Xia Z, et al. Image description with polar harmonic Fourier moments[J]. *IEEE Trans. Circuits Syst. Video Technol.*, 2020, 30(12):4440–4452.
- [228] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: From error visibility to structural similarity[J]. *IEEE Trans. Image Process.*, 2004, 13(4):600–612.
- [229] Wang Z, Li J, Wiederhold G. SIMPLiCity: Semantics-sensitive integrated matching for picture libraries[J]. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2001, 23(9):947–963.
- [230] Chan T H, Jia K, Gao S, et al. PCANet: A simple deep learning baseline for image classification[J]. *IEEE Trans. Image Process.*, 2015, 24(12):5017–5032.
- [231] Lu J, Liong V E, Zhou X, et al. Learning compact binary face descriptor for face recognition[J]. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2015, 37(10):2041–2056.
- [232] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]. *Proceedings of IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015. 1–9.
- [233] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. *Proceedings of IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016. 770–778.
- [234] Zhang X, Liu C, Suen C Y. Towards robust pattern recognition: A review[J]. *Proc. IEEE*, 2020, 108(6):894–922.
- [235] Chen C, Xiong Z, Tian X, et al. Real-world image denoising with deep boosting[J]. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2019, 42(12):3071–3087.
- [236] Wu H, Zhou J, Tian J, et al. Robust image forgery detection against transmission over online social networks[J]. *IEEE Trans. Inf. Forensic Secur.*, 2022, 17:443–456.
- [237] Biggio B, Roli F. Wild patterns: Ten years after the rise of adversarial machine learning[J]. *Pattern Recognit.*, 2018, 84:317–331.
- [238] Qi S, Zhang Y, Wang C, et al. A principled design of image representation: Towards forensic tasks[J]. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2022, 45(5):5337–5354.
- [239] Xie L, Wang J, Lin W, et al. Towards reversal-invariant image representation[J]. *Int. J. Comput. Vis.*, 2017, 123(2):226–250.
- [240] Liu D, Cheng B, Wang Z, et al. Enhance visual recognition under adverse conditions via deep networks[J]. *IEEE Trans. Image Process.*, 2019, 28(9):4401–4412.
- [241] Zhou K, Liu Z, Qiao Y, et al. Domain generalization: A survey[J]. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2022..
- [242] Li S, Ren W, Wang F, et al. A comprehensive benchmark analysis of single image deraining: Current challenges and future perspectives[J]. *Int. J. Comput. Vis.*, 2021, 129(4):1301–1322.
- [243] Genzel M, Macdonald J, Marz M. Solving inverse problems with deep neural networks—robustness included?[J]. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2022, 45(1):1119–1134.
- [244] VidalMata R G, Banerjee S, RichardWebster B, et al. Bridging the gap between computational photography and visual recognition[J]. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2020, 43(12):4272–4290.
- [245] Chen Y W, Chen Y Q. Invariant description and retrieval of planar shapes using radon composite features[J]. *IEEE Trans. Signal Process.*, 2008, 56(10):4762–4771.
- [246] Tabbone S, Terrades O R, Barrat S. Histogram of Radon transform: A useful descriptor for shape

- retrieval[C]. Proceedings of Int. Conf. Pattern Recognit., 2008. 1–4.
- [247] Wang X, Xiao B, Ma J, et al. Scaling and rotation invariant analysis approach to object recognition based on Radon and Fourier-Mellin transforms[J]. Pattern Recognit., 2007, 40(12):3503–3508.
- [248] Chen G, Bui T D, Krzyzak A. Invariant pattern recognition using Radon, dual-tree complex wavelet and Fourier transforms[J]. Pattern Recognit., 2009, 42(9):2013–2019.
- [249] Tabbone S, Wendling L, Salmon J P. A new shape descriptor defined on the Radon transform[J]. Comput. Vis. Image Underst., 2006, 102(1):42–51.
- [250] Nacereddine N, Tabbone S, Ziou D, et al. Shape-based image retrieval using a new descriptor based on the Radon and wavelet transforms[C]. Proceedings of Int. Conf. Pattern Recognit., 2010. 1997–2000.
- [251] Kan C, Srinath M D. Invariant character recognition with Zernike and orthogonal Fourier-Mellin moments[J]. Pattern Recognit., 2002, 35(1):143–154.
- [252] Cui Y, Ren W, Cao X, et al. Image Restoration Via Frequency Selection[J]. IEEE Trans. Pattern Anal. Mach. Intell., 2023..
- [253] Dabov K, Foi A, Katkovnik V, et al. Image denoising by sparse 3-D transform-domain collaborative filtering[J]. IEEE Trans. Image Process., 2007, 16(8):2080–2095.
- [254] Zhang K, Zuo W, Zhang L. FFDNet: Toward a fast and flexible solution for CNN-based image denoising[J]. IEEE Trans. Image Process., 2018, 27(9):4608–4622.
- [255] Zhang K, Zuo W, Zhang L. Learning a single convolutional super-resolution network for multiple degradations[C]. Proceedings of IEEE Conf. Comput. Vis. Pattern Recognit., 2018. 3262–3271.
- [256] Deans S R. The Radon transform and some of its applications[M]. Courier Corporation, 2007.
- [257] Hjouj F, Kammler D W. Identification of reflected, scaled, translated, and rotated objects from their Radon projections[J]. IEEE Trans. Image Process., 2008, 17(3):301–310.
- [258] Hoang T V, Tabbone S. The generalization of the R-transform for invariant pattern representation[J]. Pattern Recognit., 2012, 45(6):2145–2163.
- [259] Yang H, Qi S, Niu P, et al. Color image zero-watermarking based on fast quaternion generic polar complex exponential transform[J]. Signal Process. - Image Commun., 2020, 82:115747.
- [260] Larson E C, Chandler D M. Most apparent distortion: Full-reference image quality assessment and the role of strategy[J]. J. Electron. Imaging, 2010, 19(1):011006.
- [261] Liu X, Pedersen M, Hardeberg J Y. CID: IQ – a new image quality database[C]. Proceedings of Int. Conf. Image Signal Process., 2014. 193–202.
- [262] Press W H. Discrete Radon transform has an exact, fast inverse and generalizes to operations other than sums along lines[J]. Natl. Acad. Sci., 2006, 103(51):19249–19254.
- [263] Manjunath B S, Salembier P, Sikora T. Introduction to MPEG-7: Multimedia content description interface[M]. John Wiley & Sons, 2002.
- [264] Ouyang W, Tombari F, Mattoccia S, et al. Performance evaluation of full search equivalent pattern matching algorithms[J]. IEEE Trans. Pattern Anal. Mach. Intell., 2011, 34(1):127–143.
- [265] Zhang Y, Zhu Y, Nichols E, et al. A poisson-gaussian denoising dataset with real fluorescence microscopy images[C]. Proceedings of IEEE Conf. Comput. Vis. Pattern Recognit., 2019. 11710–11718.
- [266] Abdelhamed A, Lin S, Brown M S. A high-quality denoising dataset for smartphone cameras[C]. Proceedings of IEEE Conf. Comput. Vis. Pattern Recognit., 2018. 1692–1700.

- [267] Kokkinos I, Yuille A. Scale invariance without scale selection[C]. Proceedings of IEEE Conf. Comput. Vis. Pattern Recognit., 2008. 1–8.
- [268] Kim S, Min D, Ham B, et al. DASC: Robust dense descriptor for multi-modal and multi-spectral correspondence estimation[J]. IEEE Trans. Pattern Anal. Mach. Intell., 2016, 39(9):1712–1729.
- [269] Talmi I, Mechrez R, Zelnik-Manor L. Template matching with deformable diversity similarity[C]. Proceedings of IEEE Conf. Comput. Vis. Pattern Recognit., 2017. 175–183.
- [270] Wang C, Wang X, Xia Z, et al. Geometrically resilient color image zero-watermarking algorithm based on quaternion exponent moments[J]. J. Vis. Commun. Image Represent., 2016, 41:247–259.
- [271] Wang C, Wang X, Xia Z, et al. Ternary radial harmonic Fourier moments based robust stereo image zero-watermarking algorithm[J]. Inf. Sci., 2019, 470:109–120.
- [272] Wang C, Wang X, Chen X, et al. Robust zero-watermarking algorithm based on polar complex exponential transform and logistic mapping[J]. Multimed. Tools Appl., 2017, 76(24):26355–26376.
- [273] Gao G, Jiang G. Bessel-Fourier moment-based robust image zero-watermarking[J]. Multimed. Tools Appl., 2015, 74(3):841–858.
- [274] Verdoliva L. Media forensics and deepfakes: An overview[J]. IEEE J. Sel. Top. Signal Process., 2020, 14(5):910–932.
- [275] Simon H A. The sciences of the artificial[M]. MIT press, 2019.
- [276] Gu J, Wang Z, Kuen J, et al. Recent advances in convolutional neural networks[J]. Pattern Recognit., 2018, 77:354–377.
- [277] Yuan X, He P, Zhu Q, et al. Adversarial examples: Attacks and defenses for deep learning[J]. IEEE Trans. Neural Netw. Learn. Syst., 2019, 30(9):2805–2824.
- [278] Zhang Y, Tiño P, Leonardis A, et al. A survey on neural network interpretability[J]. IEEE Trans. Emerg. Top. Comput. Intell., 2021, 5(5):726–742.
- [279] Wiatowski T, Bölcskei H. A mathematical theory of deep convolutional neural networks for feature extraction[J]. IEEE Trans. Inf. Theory, 2017, 64(3):1845–1866.
- [280] Bronstein M M, Bruna J, Cohen T, et al. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges[J]. arXiv preprint arXiv:2104.13478, 2021..
- [281] Iscen A, Tolias G, Gosselin P H, et al. A comparison of dense region detectors for image search and fine-grained classification[J]. IEEE Trans. Image Process., 2015, 24(8):2369–2381.
- [282] Li Y, Zhou J, Cheng A. SIFT keypoint removal via directed graph construction for color images[J]. IEEE Trans. Inf. Forensic Secur., 2017, 12(12):2971–2985.
- [283] Tuytelaars T. Dense interest points[C]. Proceedings of IEEE Conf. Comput. Vis. Pattern Recognit., 2010. 2281–2288.
- [284] Ouyang W, Cham W K. Fast algorithm for Walsh Hadamard transform on sliding windows[J]. IEEE Trans. Pattern Anal. Mach. Intell., 2009, 32(1):165–171.
- [285] Ouyang W, Zhao T, Cham W K, et al. Fast full-search-equivalent pattern matching using asymmetric Haar wavelet packets[J]. IEEE Trans. Circuits Syst. Video Technol., 2016, 28(4):819–833.
- [286] Christlein V, Riess C, Jordan J, et al. An evaluation of popular copy-move forgery detection approaches[J]. IEEE Trans. Inf. Forensic Secur., 2012, 7(6):1841–1854.
- [287] Liu Y, Zhu X, Zhao X, et al. Adversarial learning for constrained image splicing detection and

- localization based on atrous convolution[J]. *IEEE Trans. Inf. Forensic Secur.*, 2019, 14(10):2551–2566.
- [288] Li H, Luo W, Huang J. Localization of diffusion-based inpainting in digital images[J]. *IEEE Trans. Inf. Forensic Secur.*, 2017, 12(12):3050–3064.
- [289] Mirsky Y, Lee W. The creation and detection of deepfakes: A survey[J]. *ACM Comput. Surv.*, 2021, 54(1):1–41.
- [290] Cox I, Miller M, Bloom J, et al. Digital watermarking and steganography[M]. Morgan kaufmann, 2007.
- [291] Du L, Ho A T, Cong R. Perceptual hashing for image authentication: A survey[J]. *Signal Process.-Image Commun.*, 2020, 81:115713.
- [292] Cozzolino D, Verdoliva L. Noiseprint: A CNN-based camera model fingerprint[J]. *IEEE Trans. Inf. Forensic Secur.*, 2019, 15:144–159.
- [293] Matern F, Riess C, Stamminger M. Gradient-based illumination description for image forgery detection[J]. *IEEE Trans. Inf. Forensic Secur.*, 2019, 15:1303–1317.
- [294] Qi H, Guo Q, Juefei-Xu F, et al. DeepRhythm: Exposing deepfakes with attentional visual heart-beat rhythms[C]. *Proceedings of ACM Int. Conf. Multimed.*, 2020. 4318–4327.
- [295] Bhatia A, Wolf E. On the circle polynomials of Zernike and related orthogonal sets[C]. *Proceedings of Math. Camb. Philos. Soc.*, volume 50. Cambridge University Press, 1954. 40–48.
- [296] Lindeberg T. Dense scale selection over space, time, and space-time[J]. *SIAM J. Imaging Sci.*, 2018, 11(1):407–441.
- [297] Sze V, Chen Y H, Yang T J, et al. Efficient processing of deep neural networks: A tutorial and survey[J]. *IEEE*, 2017, 105(12):2295–2329.
- [298] Zhao T, Blu T. The Fourier-Argand representation: an optimal basis of steerable patterns[J]. *IEEE Trans. Image Process.*, 2020, 29:6357–6371.
- [299] Jegou H, Douze M, Schmid C. Hamming embedding and weak geometric consistency for large scale image search[C]. *Proceedings of Eur. Conf. Comput. Vision*, 2008. 304–317.
- [300] Barnes C, Shechtman E, Finkelstein A, et al. PatchMatch: A randomized correspondence algorithm for structural image editing[J]. *ACM Trans. Graph.*, 2009, 28(3):24.
- [301] Fischler M A, Bolles R C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography[J]. *Commun. ACM*, 1981, 24(6):381–395.
- [302] Barnes C, Shechtman E, Goldman D B, et al. The generalized PatchMatch correspondence algorithm[C]. *Proceedings of Eur. Conf. Comput. Vision*, 2010. 29–43.
- [303] Zandi M, Mahmoudi-Aznavah A, Talebpour A. Iterative copy-move forgery detection based on a new interest point detector[J]. *IEEE Trans. Inf. Forensic Secur.*, 2016, 11(11):2499–2512.
- [304] Li Y, Zhou J. Fast and effective image copy-move forgery detection via hierarchical feature point matching[J]. *IEEE Trans. Inf. Forensic Secur.*, 2019, 14(5):1307–1322.
- [305] Amerini I, Ballan L, Caldelli R, et al. Copy-move forgery detection and localization by means of robust clustering with J-Linkage[J]. *Signal Process.-Image Commun.*, 2013, 28(6):659–669.
- [306] Silva E, Carvalho T, Ferreira A, et al. Going deeper into copy-move forgery detection: Exploring image telltales via multi-scale analysis and voting processes[J]. *J. Vis. Commun. Image Represent.*, 2015, 29:16–32.
- [307] Ardizzone E, Bruno A, Mazzola G. Copy-move forgery detection by matching triangles of key-

- points[J]. IEEE Trans. Inf. Forensic Secur., 2015, 10(10):2084–2094.
- [308] Ryu S J, Kirchner M, Lee M J, et al. Rotation invariant localization of duplicated image regions based on Zernike moments[J]. IEEE Trans. Inf. Forensic Secur., 2013, 8(8):1355–1370.
- [309] Li Y. Image copy-move forgery detection based on polar cosine transform and approximate nearest neighbor searching[J]. Forensic Sci. Int., 2013, 224(1-3):59–67.
- [310] Emam M, Han Q, Niu X. PCET based copy-move forgery detection in images under geometric transforms[J]. Multimedia Tools Appl., 2016, 75(18):11513–11527.
- [311] Pun C M, Chung J L. A two-stage localization for copy-move forgery detection[J]. Inf. Sci., 2018, 463:33–55.
- [312] Bi X, Pun C M. Fast copy-move forgery detection using local bidirectional coherency error refinement[J]. Pattern Recognit., 2018, 81:161–175.
- [313] Wu Y, Abd-Almageed W, Natarajan P. Busternet: Detecting copy-move image forgery with source/target localization[C]. Proceedings of Eur. Conf. Comput. Vision, 2018. 168–184.
- [314] Zhong J L, Pun C M. An end-to-end dense-inceptionnet for image copy-move forgery detection[J]. IEEE Trans. Inf. Forensic Secur., 2020, 15:2134–2146.
- [315] Zheng Y, Cao Y, Chang C H. A PUF-based data-device hash for tampered image detection and source camera identification[J]. IEEE Trans. Inf. Forensic Secur., 2020, 15:620–634.
- [316] Wang X, Pang K, Zhou X, et al. A visual model-based perceptual image hash for content authentication[J]. IEEE Trans. Inf. Forensic Secur., 2015, 10(7):1336–1349.
- [317] Otsu N. A threshold selection method from gray-level histograms[J]. IEEE Trans. Sys. Man Cyber., 1979, 9(1):62–66.
- [318] Korus P, Huang J. Evaluation of random field models in multi-modal unsupervised tampering localization[C]. Proceedings of IEEE Int. Workshop Inf. Forensic Secur., 2016. 1–6.
- [319] Hao Q, Luo L, Jan S T, et al. It's not what it looks like: Manipulating perceptual hashing based applications[C]. Proceedings of ACM Conf. Comput. Commun. Secur., 2021. 69–85.
- [320] Biswas R, Gonzalez-Castro V, Fidalgo E, et al. Perceptual image hashing based on frequency dominant neighborhood structure applied to tor domains recognition[J]. Neurocomput., 2020, 383:24–38.
- [321] Bruna J, Mallat S. Invariant scattering convolution networks[J]. IEEE Trans. Pattern Anal. Mach. Intell., 2013, 35(8):1872–1886.
- [322] Cohen T, Welling M. Group equivariant convolutional networks[C]. Proceedings of Int. Conf. Mach. Learn., 2016. 2990–2999.
- [323] Sifre L, Mallat S. Rotation, scaling and deformation invariant scattering for texture discrimination[C]. Proceedings of IEEE Conf. Comput. Vis. Pattern Recognit., 2013. 1233–1240.
- [324] Oyallon E, Zagoruyko S, Huang G, et al. Scattering networks for hybrid representation learning[J]. IEEE Trans. Pattern Anal. Mach. Intell., 2018, 41(9):2208–2221.
- [325] Andén J, Mallat S. Deep scattering spectrum[J]. IEEE Trans. Signal Process., 2014, 62(16):4114–4128.
- [326] Chen X, Cheng X, Mallat S. Unsupervised deep Haar scattering on graphs[J]. Adv. Neural Inf. Process. Syst., 2014, 27.
- [327] Yu S. Evolving scattering networks for engineering disorder[J]. Nature Comput. Sci., 2023, 3(2):128–138.

- [328] Cheng S, Ting Y S, Ménard B, et al. A new approach to observational cosmology using the scattering transform[J]. *Mon. Not. R. Astron. Soc.*, 2020, 499(4):5902–5914.
- [329] Cohen T S, Welling M. Steerable CNNs[C]. *Proceedings of Int. Conf. Learn. Representations*, 2016.
- [330] Worrall D E, Garbin S J, Turmukhambetov D, et al. Harmonic networks: Deep translation and rotation equivariance[C]. *Proceedings of IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017. 5028–5037.
- [331] Worrall D, Welling M. Deep scale-spaces: Equivariance over scale[J]. *Adv. Neural Inf. Process. Syst.*, 2019, 32.
- [332] Finzi M, Stanton S, Izmailov P, et al. Generalizing convolutional neural networks for equivariance to Lie groups on arbitrary continuous data[C]. *Proceedings of Int. Conf. Mach. Learn.*, 2020. 3165–3176.
- [333] Bekkers E J. B-Spline CNNs on Lie groups[C]. *Proceedings of Int. Conf. Learn. Representations*, 2019.
- [334] Weiler M, Hamprecht F A, Storath M. Learning steerable filters for rotation equivariant CNNs[C]. *Proceedings of IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018. 849–858.
- [335] Sosnovik I, Szmaja M, Smeulders A. Scale-Equivariant Steerable Networks[C]. *Proceedings of Int. Conf. Learn. Representations*, 2019.
- [336] Sun Z, Blu T. Empowering Networks With Scale and Rotation Equivariance Using A Similarity Convolution[C]. *Proceedings of Int. Conf. Learn. Representations*, 2022.
- [337] Xie Q, Zhao Q, Xu Z, et al. Fourier series expansion based filter parametrization for equivariant convolutions[J]. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2022, 45(4):4537–4551.
- [338] Atz K, Grisoni F, Schneider G. Geometric deep learning on molecular representations[J]. *Nature Mach. Intell.*, 2021, 3(12):1023–1032.
- [339] Townshend R J, Eismann S, Watkins A M, et al. Geometric deep learning of RNA structure[J]. *Science*, 2021, 373(6558):1047–1051.
- [340] Goodfellow I, McDaniel P, Papernot N. Making machine learning robust against adversarial inputs[J]. *Commun. ACM*, 2018, 61(7):56–66.
- [341] Zhan F, Yu Y, Wu R, et al. Multimodal Image Synthesis and Editing: The Generative AI Era[J]. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2023..
- [342] Bender G, Kindermans P J, Zoph B, et al. Understanding and simplifying one-shot architecture search[C]. *Proceedings of Int. Conf. Mach. Learn.*, 2018. 550–559.
- [343] Guo M, Yang Y, Xu R, et al. When NAS meets robustness: In search of robust architectures against adversarial attacks[C]. *Proceedings of IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020. 631–640.
- [344] Mallat S G. A theory for multiresolution signal decomposition: The wavelet representation[J]. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1989, 11(7):674–693.
- [345] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. *Commun. ACM*, 2017, 60(6):84–90.
- [346] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. *arXiv preprint arXiv:1409.1556*, 2014..
- [347] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]. *Pro-*

- ceedings of IEEE Conf. Comput. Vis. Pattern Recognit., 2017. 4700–4708.
- [348] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]. Proceedings of IEEE Conf. Comput. Vis. Pattern Recognit., 2016. 2818–2826.
- [349] Sandler M, Howard A, Zhu M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]. Proceedings of IEEE Conf. Comput. Vis. Pattern Recognit., 2018. 4510–4520.
- [350] Feinman R, Curtin R R, Shintre S, et al. Detecting adversarial samples from artifacts[J]. arXiv preprint arXiv:1703.00410, 2017..
- [351] Liang S, Li Y, Srikant R. Enhancing The Reliability of Out-of-distribution Image Detection in Neural Networks[C]. Proceedings of Int. Conf. Learn. Representations, 2018.
- [352] Liang B, Li H, Su M, et al. Detecting adversarial image examples in deep neural networks with adaptive noise reduction[J]. IEEE Trans. Dependable Secure Comput., 2018, 18(1):72–85.
- [353] Goswami G, Agarwal A, Ratha N, et al. Detecting and mitigating adversarial perturbations for robust face recognition[J]. Int. J. Comput. Vis., 2019, 127:719–742.
- [354] Liu J, Zhang W, Zhang Y, et al. Detection based defense against adversarial examples from the steganalysis point of view[C]. Proceedings of IEEE Conf. Comput. Vis. Pattern Recognit., 2019. 4825–4834.
- [355] Agarwal A, Singh R, Vatsa M, et al. Image transformation-based defense against adversarial perturbation on deep learning models[J]. IEEE Trans. Dependable Secure Comput., 2020, 18(5):2106–2121.
- [356] Wang C, Qi S, Huang Z, et al. Towards an accurate and secure detector against adversarial perturbations[J]. arXiv preprint arXiv:2305.10856, 2023..
- [357] Qian Y, Yin G, Sheng L, et al. Thinking in frequency: Face forgery detection by mining frequency-aware clues[C]. Proceedings of Eur. Conf. Comput. Vis., 2020. 86–103.
- [358] Wang S Y, Wang O, Zhang R, et al. CNN-generated images are surprisingly easy to spot... for now[C]. Proceedings of IEEE Conf. Comput. Vis. Pattern Recognit., 2020. 8695–8704.
- [359] Liu Z, Qi X, Torr P H. Global texture enhancement for fake face detection in the wild[C]. Proceedings of IEEE Conf. Comput. Vis. Pattern Recognit., 2020. 8060–8069.
- [360] Kurakin A, Goodfellow I J, Bengio S. Adversarial examples in the physical world[M]. . Proceedings of Artificial Intelligence Safety and Security. Chapman and Hall/CRC, 2018: 99–112.
- [361] Carlini N, Wagner D. Towards evaluating the robustness of neural networks[C]. Proceedings of IEEE Symp. Secur. Privacy, 2017. 39–57.
- [362] Chen S, He Z, Sun C, et al. Universal adversarial attack on attention and the resulting dataset DamageNet[J]. IEEE Trans. Pattern Anal. Mach. Intell., 2020, 44(4):2188–2197.
- [363] Goodfellow I J, Shlens J, Szegedy C. Explaining and harnessing adversarial examples[J]. arXiv preprint arXiv:1412.6572, 2014..
- [364] Madry A, Makelov A, Schmidt L, et al. Towards deep learning models resistant to adversarial attacks[C]. Proceedings of Int. Conf. Learn. Representations, 2018.
- [365] Moosavi-Dezfooli S M, Fawzi A, Fawzi O, et al. Universal adversarial perturbations[C]. Proceedings of IEEE Conf. Comput. Vis. Pattern Recognit., 2017. 1765–1773.
- [366] Dhariwal P, Nichol A. Diffusion models beat GANs on image synthesis[J]. Adv. Neural Inf. Process. Syst., 2021, 34:8780–8794.
- [367] Brock A, Donahue J, Simonyan K. Large scale GAN training for high fidelity natural image

- synthesis[J]. arXiv preprint arXiv:1809.11096, 2018..
- [368] Nichol A, Dhariwal P, Ramesh A, et al. Glide: Towards photorealistic image generation and editing with text-guided diffusion models[J]. arXiv preprint arXiv:2112.10741, 2021..
- [369] Rombach R, Blattmann A, Lorenz D, et al. High-resolution image synthesis with latent diffusion models[C]. Proceedings of IEEE Conf. Comput. Vis. Pattern Recognit., 2022. 10684–10695.
- [370] Gu S, Chen D, Bao J, et al. Vector quantized diffusion model for text-to-image synthesis[C]. Proceedings of IEEE Conf. Comput. Vis. Pattern Recognit., 2022. 10696–10706.

致谢

文章千古事，得失寸心知。行文至此，我将为我四年的博士生涯，乃至十余年的学生生涯画上一个句号。从文学到理工、从辽河到长江、从表面到深层、从浮躁到平和，这些年在自己的折腾、导师的引领和家人的陪伴下，我逐渐找到了人生的方向，化解了很多困惑和不平。穿过脑海中繁杂的细节和生僻的知识，他们对我的帮助难以清晰地回溯：不具体于某件事、亦不是可量化的某种利益，但他们的影响以社会关系总和的形式清晰地重塑了我。

在此，我要向所有传道、授业、解惑于我的“师者”，表示最诚挚的感激。

首先我要感谢我的导师张玉书教授。回顾博士生涯，张老师对我一直坚持鼓励、平和、民主的教育模式——经师易遇、人师难遭，我想我是非常幸运的。报考张老师的博士时，竞争十分激烈，当时我的条件并不出众，但张老师还是选择了我，用张老师的话说是：“缘分”、“第一印象对了”，非常感谢张老师在我的不善言辞中感受到我对科研的真心，我相信是因为张老师对科研保有初心才能产生这种可贵的共振。顺利入学后，我一方面初生牛犊一心想搞大文章，另一方面也很自卑显得畏首畏尾，张老师在无数次交流中不断地鼓励我大胆实践，并以平和、民主的方式调整我的错误。幸而入学第一篇综述文章顺利被顶刊录用，审稿人给出了非常积极的评价，张老师很高兴并感叹博士录取时选择，我则松了一口气：终于初步地不负师恩。进入状态后平稳地开展全局、局部和层次不变量三个主要工作，期间即使有不利的外界消息，张老师也一直积极鼓励我并未丧失信心，时常与我探讨未来的研究方向，为我打开眼界并寻找新的机会。我很坚信，如果没有张老师亦师亦友的教育，我可能仍然做着微小的重复性工作甚至为此沾沾自喜，张老师让我懂得学海无涯，也让我懂得勇敢和吃苦是学海唯一的通路。四年以来，此种宝贵经历，无法一时言尽，我将倍加珍惜并铭记于心。

我要感谢周建涛教授。周老师对我的科研方法产生了深刻影响。局部不变量的研究是整个博士期间的重要节点，当时我已经感到这个方向（已有少量启发式方法提出）可能存在某种的共同的理论基础，但囿于认知上的惰性，我所形成的理论体系有些混乱。周老师明确指出这一点，在多次修改轮回中迫使我深入思考这些抽象概念之间的关系，如何使用恰当的自然或数学语言描述，如何浓缩出独立且简明的定义、命题、性质等。周老师给我施加的这种痛苦思考极大地历练了我，形成了从表象抽象出理论的形式化思维方式。这种思维工具极其有力，没有周老师的严格要求我很可能得过且过，无法走出直觉世界。同时，在此方面我也要感谢戴华老师和他的抽象代数课程，尽管每周赶去明故宫校区早课很辛苦，但戴老师敬业地讲述让我有幸一窥数学宏伟的大厦，深深震撼于形式化思维的巨大力量。如果没有戴老师的抽象代数课，我几乎无法完成层次不变量的研究，而这是我博士期间最满意的工作。学业生涯第一门和最后一门

数学课都是“加减乘除”，但一路走来思维方式却已经发生了多次巨变，由衷地感慨学无止境、所知有限。

我要感谢操晓春教授。操老师的组会开阔了我的科研知识面。刚一参加操老师的组会就见识了一流水平的研究工作，顿时感觉非常紧张；虽然被安排了和两位顶尖博士生一样的汇报时长，但我的内容相比之下在深度上显得很薄弱。这样明显的差距给人压力同时催人奋进，此后我开始更加深挖研究工作，争取在下次汇报时尽可能减小差距。之后的汇报我相对自如一些，甚至研究体系得到了老师积极的评价；操老师的一些问题也给了我很多启示，其中层次不变量的研究就呼应了操老师早期的一些提问。感谢操老师提供的交流机会，希望以后能赶上操老师课题组顶尖博士生的研究水平。

我要感谢我的同门。感谢同为博士生的赵若宇、肖祥立、汪涛和田苗，感谢你们的并肩工作，特别是赵若宇在我遇到学业和生活问题时候给予我帮助。感谢作为硕士生的师弟师妹们，谈情、傅志彬、陈诺、翁静、石博文、张璐、孙元元和沈梦欣，我们在科研上的精诚合作，特别是和谈情、傅志彬、陈诺一起成功发表了数篇论文，感谢你们与我一同努力、相互鼓励，共同度过了难忘的时光。感谢张老师课题组的所有同学，同窗之谊，定当铭记。

我要感谢我的父母。他们从小到大一直给予我无微不至的关爱，一直支持我的求学生涯。我的父亲十分聪明、为人豁达，在改革开放前自学化工，几乎靠一己之力推动油漆厂的技术升级，改革开放后又跑通原料和销售渠道，但受限于时代最终未能进入大学。我的父亲极端厌恶应试教育和形式主义，当我被题海淹没时他说“能用计算器就不要手算”、“掌握思想”，这种宽松且实事求是的教育风格使我不仅没有厌学，反而一直保持了可贵的好奇心和求知欲；虽然高考有所失利，但进入研究生阶段后的我深刻理解了我父亲教育的远见和勇气。我的父亲非常关心我博士期间的研究进展，在论文发表时甚至高兴得在家里放起了鞭炮，我想这可能是对他未能上大学的遗憾的一点宽慰吧。我的母亲对我信任且包容，我所做的人生决定她几乎毫无保留的支持，一直是我最强有力的后盾和支持。在读博期间，我的母亲给了我无限的关怀和鼓励，让我有勇气和信心去追求自己的梦想。哀哀父母，生我劬劳，进入而立之年的我将承担起更多责任。希望你们永远健康、幸福！

祁树仁

2024 年春于南京

在学期间的学术论文与研究成果

博士学位论文相关的主要学术论文

1. Shuren Qi, Yushu Zhang[✉], Chao Wang, et al. A survey of orthogonal moments for image representation: Theory, implementation, and evaluation. *ACM Computing Surveys (CSUR)*, 2023, 55(1): 1-35. (第二章, 计算机学科顶级期刊, IF 16.6, ESI 高被引论文)
2. Shuren Qi, Yushu Zhang[✉], Chao Wang, et al. Representing noisy image without denoising. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2024, Early Access. (第三章, 人工智能领域顶级期刊, IF 23.6, CCF-A)
3. Shuren Qi, Yushu Zhang[✉], Chao Wang, et al. A principled design of image representation: Towards forensic tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2022, 45(5): 5337-5354. (第四章, 人工智能领域顶级期刊, IF 23.6, CCF-A)
4. Shuren Qi, Yushu Zhang[✉], Chao Wang, et al. Hierarchical invariance for robust and interpretable vision tasks at larger scales. *International Journal of Computer Vision (IJCV)*, *Under Review*. (第五章, 计算机视觉领域顶级期刊, IF 19.5, CCF-A)

攻读博士学位期间完成的其他学术论文

1. Shuren Qi, Qing Tan, Yushu Zhang[✉], et al. Learning to predict forgery probability from PRNU. *Under Review*.
2. Shuren Qi, Nuo Chen, Yushu Zhang[✉], et al. Beyond colors: Exposing the recoloring manipulation over the spatial dimensions. *Under Review*.
3. Chao Wang, Shuren Qi[✉], Zhiqiu Huang, et al. Spatial-frequency discriminability for revealing adversarial perturbations. *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, *Major Revision*.
4. Yushu Zhang, Zhibin Fu, Shuren Qi[✉], et al. PS-Net: A learning strategy for accurately exposing the professional Photoshop inpainting. *IEEE Transactions on Neural Networks and Learning Systems (TNNLS)*, 10.1109/TNNLS.2023.3272733, 2023.
5. Yushu Zhang, Qing Tan, Shuren Qi[✉], et al. PRNU-based image forgery localization with deep multi-scale fusion. *ACM Transactions on Multimedia Computing Communications and Applications (TOMM)*, 2023, 19(2): 1-20.
6. Yushu Zhang, Nuo Chen, Shuren Qi[✉], et al. Detection of recolored image by texture features in

- chrominance components. *ACM Transactions on Multimedia Computing Communications and Applications (TOMM)*, 2023, 19(3): 1-23.
7. Yushu Zhang, Zhibin Fu, Shuren Qi[✉], et al. Localization of inpainting forgery with feature enhancement network. *IEEE Transactions on Big Data (TBD)*, 2023, 9(3): 936-948.
 8. Chao Wang, Zhiqiu Huang[✉], Shuren Qi, et al. Shrinking the semantic gap: Spatial pooling of local moment invariants for copy-move forgery detection. *IEEE Transactions on Information Forensics and Security (TIFS)*, 2023, 18: 1064-1079.
 9. Wenying Wen, Ziyue Yuan, Shuren Qi, et al. PPM-SEM: A privacy-preserving mechanism for sharing electronic patient records and medical images in telemedicine. *IEEE Transactions on Multimedia (TMM)*, 10.1109/TMM.2023.3339588, 2023.
 10. Wenying Wen, Haigang Huang, Shuren Qi, et al. Joint coverless steganography and image transformation for covert communication of secret messages. *IEEE Transactions on Network Science and Engineering (TNSE)*, 10.1109/TNSE.2024.3354941, 2024.
 11. Tao Wang, Yushu Zhang[✉], Shuren Qi, et al. Security and privacy on generative data in AIGC: A survey. *Under Review*.

攻读博士学位期间申请的专利

1. 祁树仁, 张玉书, 王超. 一种稠密域图像鲁棒表征方法: 202210355254.X (公开), 2022.
2. 王超, 黄志球, 祁树仁, 沈国华. 一种基于视觉词袋模型的图像复制-粘贴篡改检测算法: 202210456583.3 (公开), 2022.
3. 谈情, 张玉书, 祁树仁. 一种基于 PRNU 的深度多尺度融合图像取证方法: 202111330526.2 (公开), 2022.
4. 傅志彬, 张玉书, 祁树仁. 一种基于特征增强神经网络的图像修复取证方法: 202111327201.9 (公开), 2021.
5. 陈诺, 张玉书, 祁树仁. 一种基于空间相关性的重新着色图像取证方法: 202111392793.2 (公开), 2021.

攻读博士学位期间获得的主要奖励与荣誉

1. 江苏省研究生科研与实践创新计划, 15000 元, 江苏省教育厅, 2022.
2. 浦芯精英奖学金, 10000 元, 南京市浦口区政府, 2023.
3. 三好研究生和科研创新先进个人, 连续三年, 南京航空航天大学, 2021-2023.
4. 隐者联盟公众号年度论文奖, 最佳论文, 2022.
5. 中国图象图形学报研究生学术论坛, 优秀报告, 2023.