中图法分类号:TP18; TP391　　**文献标识码**: A　　**文章编号**: 1006-8961(XXXX)XX-0001-34

论文引用格式:Yu Zitong, Ye Shuo, Lin Xun, Wang Jun, Ma Hui, Guo Dan, Zou Bochao, Hu Ruimin, Shan Caifeng, Liu Qingshan, Zhao Yao, Cao Xiaochun, Shan Shiguang. Subtle Visual Computing: A Survey[J/OL]. Journal of Image and Graphics, XXXX: 1-34. DOI: 10.11834/jig.260599. (余梓彤, 叶硕, 林讯, 王俊, 马辉, 郭丹, 邹博超, 胡瑞敏, 单彩峰, 刘青山, 赵耀, 操晓春, 山世光. 微视觉计算研究综述[J/OL]. 中国图象图形学报,XXXX:1-34. DOI: 10.11834/jig.260599.)[DOI:10.11834/jig.260599]

微视觉计算研究综述

余梓彤[1],叶硕[1],林讯[1],王俊[1],马辉[1],郭丹[2],邹博超[3],胡瑞敏[4],单彩峰[5],刘青山[6],赵耀[7],操晓春[8],山世光[9]

1. 大湾区大学,广东东莞·523000; 2. 合肥工业大学,安徽合肥·230009; 3. 北京科技大学,北京·100000; 4. 武汉大学,湖北武汉·430000; 5. 南京大学,江苏南京·210000; 6. 南京邮电大学,江苏南京·210000; 7. 北京交通大学,北京·100000; 8. 中山大学,广东深圳·518107; 9. 中国科学院计算技术研究所, 北京·100190

摘　要: 现实视觉场景中,决定识别、判断与决策的关键信息往往并非显著目标或主导语义,而是以信号微弱、空间微小或依赖上下文关系等形式存在,呈现出“视觉显著性低而任务判别价值高”的共同特征。现有研究围绕微表情、微小目标、工业与医学异常、伪装目标、生成内容鉴别、生物特征防伪及基于远程光电容积脉搏波(remote photoplethysmography,rPPG)的非接触式生理信号测量等任务形成了大量方法,但不同方向长期独立发展,尚缺乏对这类低显著视觉证据的统一认识与系统建模。为此,本文从视觉证据的微弱性、稀疏性与隐现性三类典型外在表现出发,并进一步将其归结为观测临界性、表征易失性与判别欠定性三个相互关联的核心困难,系统提出微视觉计算的概念与研究框架,将其概括为从复杂视觉信息中发现、恢复、表征、关联,以支撑精细感知、状态理解与可靠决策的视觉计算过程。在方法层面,本文从输入层的信号增强与数据生成、模型层的精细判别与稳健表征,以及数据与任务层的多模态与多任务统一学习三个层次梳理共性技术机制,并分析不同方法在微弱信息保持、局部证据发现、时空关系建模、跨域泛化和知识迁移等方面的特点与局限。在应用层面,本文以观测临界性、表征易失性和判别欠定性为统一分析维度,归纳十类代表性微视觉任务及其典型数据集,并通过统一或可比的评测协议分析不同核心困难在代表性算法中的具体表现。在此基础上,本文进一步从微视觉基础模型与统一表征学习、微视觉生成与理解一体化、微视觉应用由显式目标识别走向潜在状态感知,以及开放环境下的可信微视觉计算四个方面讨论未来发展趋势,推动微视觉计算由分散的任务建模走向统一能力构建,由单纯的微弱信息识别进一步走向对复杂场景中潜在状态的理解与可信决策。本文旨在建立贯通不同任务、场景与技术路线的微视觉计算统一视角,为低显著高价值视觉信息的系统研究提供概念框架、方法体系与发展参考。

关键词: 微视觉计算;微弱视觉信号;低可观测性;多模态学习;视觉基础模型

Subtle Visual Computing: A Survey

Yu Zitong[1], Ye Shuo[1], Lin Xun[1], Wang Jun[1], Ma Hui[1], Guo Dan[2], Zou Bochao[3], Hu Ruimin[4], Shan Caifeng[5], Liu Qingshan[6], Zhao Yao[7], Cao Xiaochun[8], Shan Shiguang[9]

1. Great Bay University, Dongguan 523000, Guangdong, China; 2. Hefei University of Technology, Hefei 230009, Anhui, China; 3. University of Science and Technology Beijing, Beijing 100000, China; 4. Wuhan University, Wuhan 430000, Hubei, China; 5. Nanjing University, Nanjing 210000, Jiangsu, China; 6. Nanjing University of Posts and Telecommunications, Nanjing 210000, Jiangsu, China;

收稿日期:2026-09-20;**修回日期**:2026-11-16

基金项目:国家自然科学基金项目(项目编号:62576076,62441619,62411540034,62676132, 62576032, 62625602, U22A2035, 62572359)

Supported by: National Natural Science Foundation of China (Grant No. 62576076,62441619,62411540034,62676132, 62576032, 62625602, U22A2035,62572359)

© 中国图象图形学报版权所有

7. Beijing Jiaotong University, Beijing 100000, China; 8. Sun Yat-sen University, Shenzhen 518107, Guangdong, China; 9. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China.

Abstract: Modern visual intelligence is predominantly developed around salient objects and dominant semantic patterns, implicitly assuming that task-relevant evidence can be sufficiently captured by imaging systems and reliably preserved throughout compression, downsampling, feature aggregation, and semantic abstraction. In many real-world scenarios, however, the evidence that ultimately determines recognition, interpretation, and decision-making is neither visually prominent nor information-redundant. Instead, it may appear as weak signals, spatially or temporally sparse structures, subtle local variations, or context-dependent relationships. Such evidence is characterized by low visual saliency but high task relevance, and can easily be overwhelmed by noise, suppressed by dominant visual patterns, lost during representation learning, or misinterpreted without appropriate contextual reasoning. Although substantial progress has been made in areas such as micro-expression recognition, tiny object detection, industrial and medical anomaly detection, camouflaged object perception, biometric anti-spoofing, generated-content detection, subtle action understanding, and remote photoplethysmography (rPPG)-based contactless physiological measurement, these directions have largely evolved independently, resulting in fragmented problem definitions, methodological taxonomies, and evaluation paradigms. To bridge these isolated research lines, this survey introduces Subtle Visual Computing (SVC) as a unified perspective for studying low-observability yet high-task-value visual evidence. We first characterize subtle visual evidence through three recurring manifestations, weak magnitude, sparse spatiotemporal support, and context-dependent observability, and further abstract them into three interconnected computational challenges: observability criticality, where task-relevant evidence approaches the limit of reliable sensing and sampling; representational fragility, where already observed evidence is easily weakened, diluted, or discarded during feature transformation; and discriminative underdetermination, where isolated observations remain insufficient for reliable inference without contextual, temporal, structural, or multimodal relationships. On this basis, we formulate SVC as a complete computational chain spanning evidence acquisition, enhancement, representation, localization, relational reasoning, verification, and decision-making. Rather than organizing existing methods solely by application, we develop a mechanism-oriented taxonomy at three levels: input-level signal enhancement and data generation, model-level fine-grained discrimination and robust representation, and data- and task-level multimodal and multitask unified learning. We further examine ten representative SVC task families, their benchmark datasets, evaluation protocols, and representative algorithms, using the three core challenges as common analytical dimensions rather than mutually exclusive task categories. The resulting synthesis reveals that advances across seemingly different SVC tasks are governed by a shared requirement: task-critical evidence must remain observable, preservable, and discriminative throughout the entire sensing-to-decision pipeline. Input enhancement, reconstruction, temporal amplification, frequency analysis, and generative augmentation can improve the accessibility of subtle evidence, but may also amplify irrelevant noise or introduce artificial cues. At the representation level, high-resolution and multiscale modeling, local evidence discovery, spatiotemporal learning, contextual reasoning, domain generalization, and foundation-model adaptation increasingly shift the focus from simply strengthening weak responses to selectively preserving task-relevant information under complex interference. Multimodal and multitask learning further enable complementary sensing and knowledge transfer across tasks and environments, while introducing new challenges related to modality alignment, optimization conflict, and negative transfer. Meanwhile, benchmark evolution shows a clear transition from single-modality, static, controlled, and closed-set settings toward multimodal, dynamic, natural, and open-world scenarios with richer contextual and relational information. Comparative results across representative tasks further indicate that although current methods can achieve strong performance under controlled or in-domain conditions, substantial degradation remains under cross-dataset evaluation, complex backgrounds, distribution shifts, and open environments, suggesting that observability criticality, representational fragility, and discriminative underdetermination often interact rather than occur in isolation. SVC should therefore not be regarded as a simple collection of tasks involving small objects, weak signals, or fine-grained differences. Instead, it represents a broader class of visual computing problems centered on the reliable acquisition, preservation, interpretation, and utilization of low-observability, high-task-value visual evidence. This unified view connects previously separated

© 中国图象图形学报版权所有

research areas through a common sensing-to-decision perspective and provides a basis for cross-task knowledge transfer and general-purpose model development. Looking forward, we identify four particularly promising directions: developing high-resolution, long-temporal, and multimodal SVC foundation models with unified representations; integrating subtle visual generation, restoration, amplification, and understanding within a closed-loop framework; extending visual intelligence from explicit object recognition toward the perception of latent physical, physiological, behavioral, and semantic states; and building trustworthy SVC systems capable of robust reasoning, uncertainty awareness, and continual adaptation in open environments. Together, these directions may drive subtle visual computing from fragmented task-specific modeling toward unified visual capabilities for fine-grained perception, latent-state understanding, and reliable decision-making.

Key words: subtle visual computing; subtle visual signals; low observability; multimodal learning; vision foundation models

论文引用格式：Yu Z T, Ye S, Lin X, Wang J, Ma H, Guo D, Zou B C, Hu R M, Shan C F, Liu Q S, Zhao Y, Cao X C, Shan S G. Subtle Visual Computing: A Survey [J]. Journal of Image and Graphics. [余梓彤，叶硕，林讯，王俊，马辉，郭丹，邹博超，胡瑞敏，单彩峰，刘青山，赵耀，操晓春，山世光. 微视觉计算研究综述[J]. 中国图像图形学报.] [DOI:10.11834/jig.260599]

0 引　言

随着机器视觉、智能传感器和深度学习技术的快速发展，视觉计算系统已广泛应用于工业制造、医疗健康、公共安全、智能交通、遥感监测和人机交互等领域。传统视觉计算通常面向轮廓清晰、尺度适中且语义显著的目标，通过提取纹理、形状及高层语义特征，实现目标检测、识别与场景理解（文浪等，2025）。然而，在大量真实应用中，影响最终判断与决策的关键信息并不总是以显著目标或强语义模式的形式出现。受目标属性、成像条件、环境干扰和场景关系等因素影响，许多具有重要判别价值的视觉证据往往表现出信号微弱、空间支撑有限或难以直接观测等特征，因而容易在图像采集、特征提取和语义推理过程中被噪声淹没、被复杂背景掩盖，或被占主导地位的视觉模式所抑制（王晓燕等，2025）。如何有效获取、表征和理解这类细微但关键的视觉信息，已成为视觉计算进一步走向精细化、可靠化与智能化所面临的重要问题。

近年来，围绕上述问题已经形成了多个较为活跃的研究方向，包括微小目标检测、细粒度视觉识别、工业与医学异常检测、微表情与微动作识别、基于rPPG的非接触式生理信号测量、伪装目标检测以及生成内容鉴别等。针对不同类型的视觉困难，研究者提出了超分辨率重建、多尺度特征融合、局部区域学习、运动放大、频域分析、上下文建模和多模态信息融合等方法，推动了视觉计算向细微局部证据、弱动态变化和隐含关系信息的理解不断拓展。如图1所示，上述任务虽然涉及不同研究对象和应用场景，但其关键视觉证据通常呈现微弱性、稀疏性和隐现性等共性特征。然而，现有研究大多围绕特定任务独立展开，在问题定义、技术方法和评价体系方面仍存在一定分割，尚未充分揭示不同任务在关键证据获取、保持与判别过程中的内在联系。

为此，本文围绕微视觉计算的概念体系、技术方法和应用发展，对相关研究进行系统梳理。首先，本文从观测临界性、表征易失性和判别欠定性三个相互关联的层面界定微视觉问题的核心困难，并分析不同属性之间的联系与区别，在此基础上给出微视觉计算的总体技术框架。其次，本文总结微视觉计算中的典型算法，重点分析信号增强、多尺度表征、局部判别学习、时空建模、上下文推理和基础模型迁移等技术的基本思想、适用条件与主要局限。随后，本文归纳微视觉计算的主要应用场景、代表性任务、常用数据集和评价指标，并对不同方向中的代表性算法性能进行整理与比较。最后，本文讨论当前研究在数据获取、标注成本、跨域泛化、统一建模、计算效率和可信解释等方面面临的挑战，并对微视觉基础模型、多模态协同感知和开放环境学习等未来方向进行展望。

1 微视觉问题定义与应用场景

1.1 微视觉信息与微视觉计算

现有视觉计算通常围绕显著对象和主导语义展

© 中国图象图形学报版权所有

开，并隐含假设任务相关证据已经被成像系统充分采样，能够在后续的压缩、下采样、特征聚合和语义抽象过程中得到稳定保留。在这一假设下，模型可以通过抑制局部扰动、压缩冗余信息和学习语义不变性获得稳健表征。然而，在一类重要的视觉任务中，决定最终识别、判断与决策的证据并不具有足够的观测裕度与信息冗余，其可能表现为变化幅度较低、空间或时间支撑有限，或者只有在特定上下文关系中才具有明确的判别意义。此类证据虽然真实存在于观测数据中，却可能在从物理采集到计算决策的不同阶段被淹没、削弱或误释。因此，这类问题的核心困难不仅在于高层语义建模，更在于任务关键证据在整个视觉计算链条中的有效获取、保持与解释。

从视觉证据的外在表现来看，微视觉信息通常呈现“弱、稀、隐”三类特征：“弱”指证据幅度、对比度或统计偏移较低；“稀”指证据仅占据有限的空间区域、时间区间或观测次数；“隐”指其判别意义难以通过孤立观测直接确定，而依赖上下文关系、历史状态或先验知识。三者并非彼此独立，而可在同一任务中交叉存在。进一步从任务关键证据由物理观测传递至计算决策的过程来看，本文将微视觉问题的核心困难概括为**观测临界性、表征易失性和判别欠定性**三个相互关联的层面。

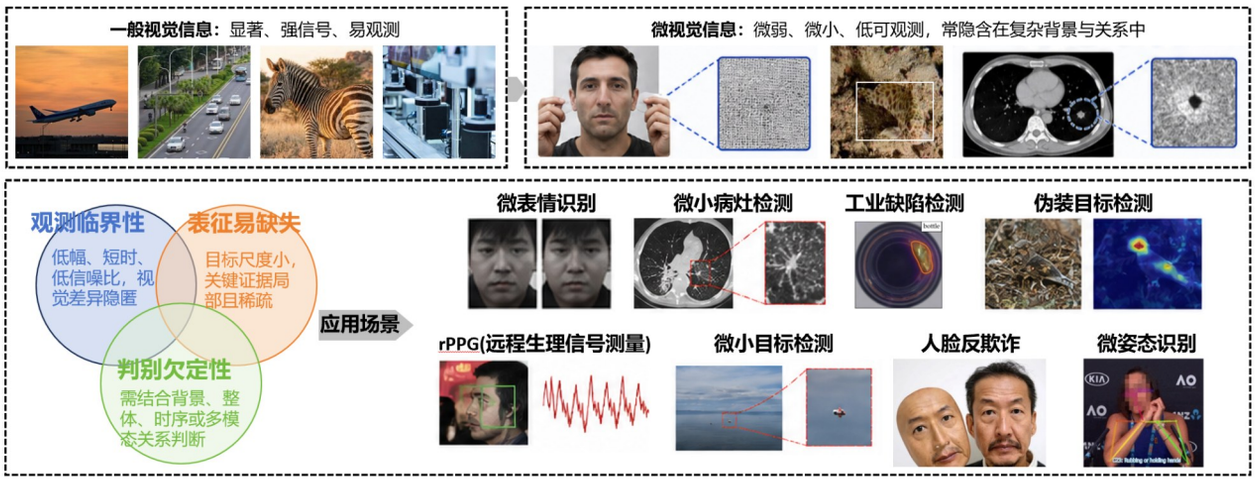


图1　微视觉计算的概念、核心属性与典型应用。

Fig. 1　Concept, Core Characteristics, and Typical Applications of Subtle Visual Computing.

观测临界性。观测临界性描述任务关键证据在感知与采样阶段接近稳定获取边界的状态。此类证据相对于成像噪声、环境变化或有限采样能力通常具有较低的可观测裕度，其细微变化可能难以被稳定记录，或难以从背景干扰中可靠分离。其既可能来源于低对比度、低幅度和低信噪比等较低的证据--干扰比，也可能来源于有限的空间尺度、较短的持续时间或稀疏的采样机会。例如，微小目标和微小病灶的有限像素支撑会限制纹理与结构信息，而微表情及rPPG信号中的低幅动态则对时间分辨率和采样稳定性提出更高要求。因此，微视觉任务首先面临的是相关证据能否被可靠观测，而不仅仅是后续模型是否具有足够的表征能力。

表征易失性。表征易失性描述任务证据虽然已经存在于原始观测中，但在由输入数据向深层特征传递的过程中容易被削弱、稀释或覆盖。由于微视觉证据通常具有有限的时空支撑、较低的信息冗余，或主要体现为细微局部与统计差异，其在图像压缩、尺度变换、空间与时间下采样、特征聚合和语义抽象过程中更容易遭受信息损失。例如，连续下采样可能使微小目标和局部缺陷的边缘与纹理逐渐消失，时序聚合则可能削弱微表情和微动作中的瞬时变化。表征易失并不意味着证据一定完全消失，也可能表现为与背景信息混合、被强响应区域覆盖，或者在高层特征中失去足够的区分性。因此，微视觉建模不仅需要发现任务相关信息，还需要保证其在多层计算过程中得到有效保持和传递。

判别欠定性。判别欠定性描述任务证据即使能够被稳定观测并在模型中得到保留，仅依靠孤立外观仍可能不足以支持可靠判断的情况。其根源在于

© 中国图象图形学报版权所有

部分微视觉证据具有明显的关系依赖性，同一局部模式在不同背景、状态或语义条件下可能具有不同的任务含义。例如，伪装目标的局部颜色与纹理可能与背景高度一致，需要通过目标--背景关系才能形成有效区分；某些生成或篡改区域本身具有较高视觉真实性，其异常则主要表现为跨区域结构不连续、跨帧状态不一致或局部证据与全局语义之间的冲突。因此，对于此类问题，上下文与关系信息并非简单的辅助特征，而是提高证据判别确定性的重要依据。

上述三个层面分别对应任务关键证据在视觉计算过程中**“能否被可靠获取”**、**“能否被有效保持”以及“能否被稳定判别”**三个基本问题。三者并非互斥类别，而可能在同一任务中共同存在。例如，微小病灶的识别不仅受到有限尺度和低对比度的影响，还需要避免特征提取过程中的局部信息损失，并结合器官位置及周围组织结构提高判断可靠性。由此，“弱、稀、隐”刻画微视觉证据的外在表现，三个核心困难则揭示其在视觉计算过程中的内在挑战。

微视觉还具有明显的任务相对性和观测相对性。同一视觉模式在不同任务目标、成像设备、输入尺度、采样频率和观测时长下可能具有不同的可观测程度。例如，某一病灶在整幅病理切片中可能仅占少量像素，而在高倍局部图像中则能够清晰辨识；某一生理变化在单帧图像中几乎不可观察，但通过长时间序列聚合后可以形成稳定估计。因此，微视觉信息并不是图像内容的一种固定类别，而是由视觉证据、任务目标、观测条件和计算过程共同决定的关系属性。

基于上述分析，本文将**微视觉信息**定义为：在给定任务、成像、观测与计算条件下，对正确认知与决策具有关键贡献，但由于观测裕度有限、表征稳定性不足或判别关系欠定而处于有效可观测性边界，容易在感知、传输、表征或推理过程中被淹没、削弱、丢失或误释的高任务价值视觉证据。

相应地，本文将**微视觉计算**定义为：面向处于有效可观测性边界但具有高任务价值的视觉证据，研究其感知、增强、表征、发现、关联、验证与决策的视觉计算理论与方法体系。其目标并非无条件放大所有细微变化，也不是单纯处理物理尺度较小的视觉对象，而是在复杂干扰条件下实现对任务相关微视觉证据的可靠获取、稳定保持与有效解释，同时保持对无关噪声、成像波动和分布变化的鲁棒性，使原本难以利用的信息转化为可感知、可表征、可验证和可决策的任务证据。

从这一意义上看，小目标检测、微弱运动分析、细粒度识别、伪装目标检测、微小病灶分析、远程生理测量及视觉内容取证等并非微视觉计算的同义概念，而是分别由观测临界性、表征易失性或判别欠定性等不同机制主导的典型问题。微视觉计算所关注的是如何围绕低可观测、高任务价值的视觉证据，形成从可靠获取、稳定表征到关系理解与任务决策的完整计算链条。需要强调的是，微视觉计算与细粒度视觉识别、小目标感知、弱信号分析及异常检测等已有研究方向既存在交叉，也具有不同的问题边界。细粒度视觉识别主要关注类别间细微差异的区分，小目标感知侧重有限空间尺度下的目标定位与识别，弱信号分析强调低幅度或低信噪比信息的提取，而异常检测则主要面向偏离正常分布的模式发现。这些研究通常由特定任务目标或数据属性界定，而微视觉计算并不以目标尺度、信号强度或任务类型作为唯一划分依据，而是以任务关键视觉证据在特定观测与计算条件下是否面临可靠获取、有效保持或稳定判别的困难作为核心判定标准。因此，并非所有细粒度识别、小目标检测或异常检测问题都必然属于典型微视觉问题；只有当任务结果主要依赖低可观测、易丢失或关系依赖的关键证据时，才具有明确的微视觉计算属性。相较于已有任务级定义，微视觉计算的价值在于从观测临界性、表征易失性和判别欠定性三个维度揭示不同任务的共同困难，并建立贯通证据获取、特征保持与关系判别的统一分析框架，为跨任务方法复用、知识迁移和统一模型设计提供依据。

1.2 应用场景

微视觉计算已广泛涉及身份安全、工业制造、医疗健康、公共安全和人机交互等领域。不同应用的研究对象与决策目标存在差异，其关键证据在观测、表征和判别过程中面临的困难也各有侧重。本节结合十类代表性任务，重点分析不同应用中的关键视觉证据、主要技术挑战及其具体表现，以揭示微视觉计算在不同场景中的共性与差异。

生物特征防伪。生物特征防伪通过局部反射、纹理连续性、深度结构、眨眼及成像伪影等细微信息区分真实人体与照片、视频回放和面具等攻击。由

© 中国图象图形学报版权所有

于不同攻击介质可能具有相似外观，且设备与环境变化容易干扰局部特征，因此需要结合细粒度纹理、时空活体信息和跨域不变表征，提高对未知攻击的识别能力。

人脸 Deepfake 与通用人工智能生成内容（artificial intelligence-generated content，AIGC）检测。该任务旨在区分真实内容与由人脸操纵、生成对抗网络、扩散模型及其他生成系统产生的伪造内容。随着生成质量提高，传统显著伪影逐渐减少，判别依据更多表现为局部纹理、频率、结构连续性、时序一致性及语义关系中的细微异常。这些线索既可能因压缩和后处理而衰减，也往往需要结合跨区域、跨帧或局部--整体关系才能可靠判断。

工业异常检测。工业异常通常表现为划痕、裂纹、凹陷、污染或装配偏差等局部异常，并可能被材料纹理和光照变化掩盖。除细微外观缺陷外，部件缺失、错误排列及装配关系异常还需要结合整体结构进行判断。因此，该任务既关注局部异常的精确定位，也强调异常样本稀缺条件下的开放异常识别与逻辑关系建模。

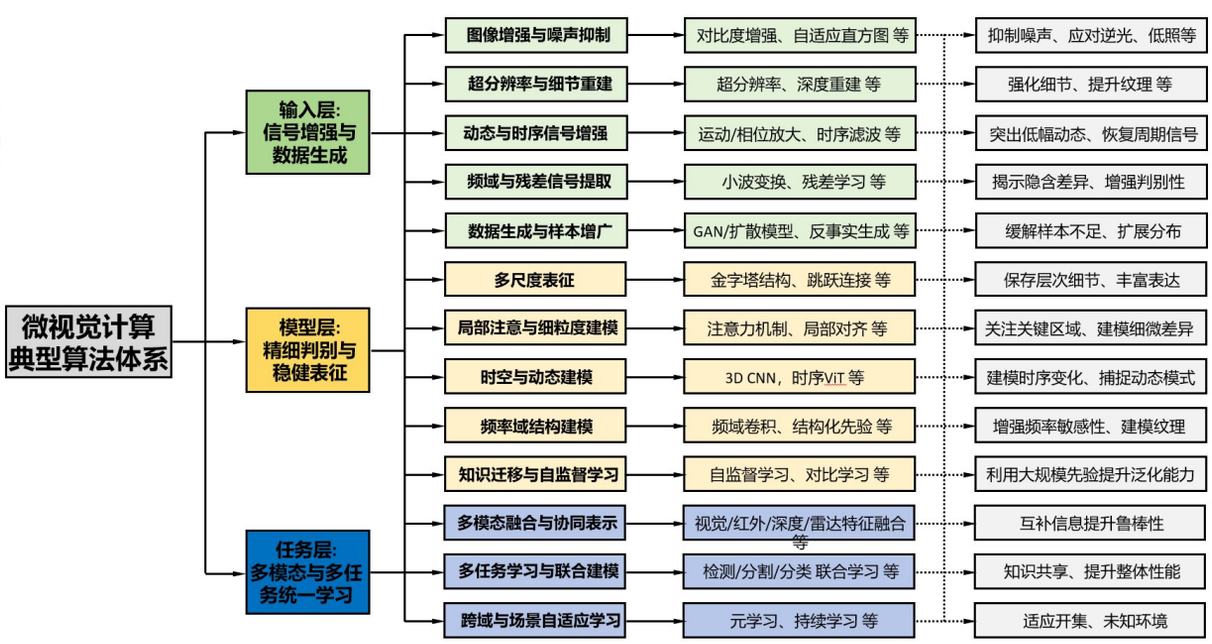
医学异常检测。微小病灶、早期组织变化和低对比异常通常仅占有限区域，并与正常组织具有相似外观。其识别不仅依赖细微形态与纹理信息，还需要结合器官位置、周围组织结构及临床先验，减少正常组织变化与病理异常之间的混淆，因此对检测敏感性、定位精度和结果可解释性提出较高要求。

微表情识别。微表情具有持续时间短、运动幅度低和局部化等特点，关键线索主要集中于眉眼、鼻翼和嘴角等区域。其主要困难在于捕获短时肌肉运动的细微变化，并区分真实表情活动与头部运动、光照波动及个体差异，因此需要结合局部运动增强、关键帧定位和细粒度时空建模。

伪装目标检测。伪装目标与背景在颜色、纹理和结构上高度相似，使目标轮廓与区域归属难以直接确定。该任务的关键在于从局部纹理差异、边界连续性及整体形态中积累有效线索，并结合目标--背景关系和多尺度上下文实现目标区域的准确分离。

微姿态与微动作识别。微姿态与微动作表现为头部、肢体或身体局部区域中幅度较小、持续时间较短的姿态和动作变化。相似动作之间往往仅存在局部关节位移、运动方向或时序节奏上的细微差异，因此需要结合身体结构、局部运动轨迹和多时间尺度特征进行识别。

基于 rPPG 的非接触式生理信号测量。该任务通过摄像头采集人体皮肤视频，分析血液容积变化引起的微弱肤色波动，提取远程光电容积脉搏波（remote photoplethysmography，rPPG），并据此估计心率等生理指标。由于目标信号容易受到光照变化、头部运动、相机噪声和视频压缩的干扰，该任务重点关注微弱生理信号的稳定提取、非生理扰动抑制及连续测量的准确性。



© 中国图象图形学报版权所有

欺骗行为检测。欺骗行为检测利用面部微表情、细微动作、生理变化和多模态行为线索辅助判断真实与欺骗状态。由于单一行为线索缺乏稳定的欺骗特异性，相同表现在不同个体和语境下可能具有不同含义，因此需要结合行为时序、语境信息及多模态证据进行综合分析，并谨慎评估判断的不确定性。

微小目标检测。微小目标通常仅由少量像素表示，并伴随纹理不足、背景干扰、遮挡和密集分布等问题。该任务主要关注有限像素条件下的目标结构保持、精确定位及密集实例区分，需要通过高分辨率表征、多尺度特征融合和上下文建模缓解小目标特征不足与背景混淆问题。

图2 微视觉计算典型算法体系。

Fig. 2. Representative Algorithmic Framework for Subtle Visual Computing.

2 技术概览与典型算法

2.1 微视觉计算技术概览

在前述问题定义的基础上，微视觉计算的技术体系可进一步划分为感知与采集、增强与重建、判别表征、证据定位、关系推理和任务决策等环节。不同环节分别承担原始信息获取、信号质量改善、特征提取、关键证据发现、关系理解及决策输出等功能，并可根据具体任务需求进行联合优化。

感知与采集。感知与采集是微视觉计算的输入基础，其目标是在成像和采样阶段尽可能保留任务相关的原始信息。对于微小目标和局部结构，需要提高空间分辨率与局部采样能力；对于短时、低幅动态，需要保证足够的时间分辨率与采样稳定性；对于可见光条件下难以区分的目标，可引入热红外、深度、雷达、事件相机和多光谱等补充模态。合理的成像方式、传感器配置和采样策略能够降低信息在进入计算模型之前的不可逆损失。

增强与重建。增强与重建旨在改善微视觉信息的可辨识程度，缓解噪声干扰、分辨率不足和信号衰减造成的信息缺失。典型方法包括图像去噪、对比度增强、超分辨率重建、运动放大、频率增强和多帧融合等。其中，超分辨率与细节重建侧重恢复有限空间区域中的结构信息，运动放大和时序滤波用于突出低幅动态，频率分析则有助于揭示空间域中不易直接观察的细微变化。该阶段不仅需要增强目标证据，还应避免引入虚假细节或放大无关噪声。

判别表征。判别表征的目标是将原始微视觉信息转化为稳定且具有任务区分能力的特征。由于局部细节和弱信号容易在深层网络的下采样与特征聚合过程中被抑制，模型通常需要兼顾全局语义保持与局部信息保真。常见技术包括高分辨率表示、多尺度特征融合、局部注意力、时空特征建模和频率表征等。近年来，视觉基础模型和生成模型提供的多层级先验，也为微视觉信息的迁移表征和细粒度建模提供了新的技术基础。

证据定位。微视觉证据通常只存在于少量空间区域、局部部件或有限时间片段中，因此需要从大范围背景和冗余观测中准确筛选任务相关位置。相关方法包括候选区域生成、关键点定位、局部部件发现、显著区域搜索、关键帧选择、稀疏注意力和弱监督定位等。证据定位不仅回答关键视觉信息“是否存在”，还进一步确定其“位于何处”以及“在何时出现”，是连接特征表征与后续语义理解的重要环节。

关系推理。对于低可观测视觉问题，局部证据本身往往不足以支持可靠判断，还需要分析其与周围环境及其他信息之间的关系。关系推理主要涉及目标--背景关系、对象--对象关系、局部--整体一致性、时序状态变化以及跨模态对应关系。上下文建模、图结构学习、Transformer、因果推理和视觉语言模型等方法能够将孤立的局部证据置于更完整的场景和语义结构中，从而揭示其隐含意义并减少由背景偏差和表面相关性造成的误判。

任务决策。任务决策将微视觉表征和推理结果转化为检测、分割、分类、异常识别、状态估计和行为理解等具体输出。由于微视觉证据通常不完整、不稳定，并容易受到噪声、遮挡和分布变化影响，最终决策除追求预测精度外，还需要关注鲁棒性与可信性。置信度校准、不确定性估计、开放集识别和可解释分析等技术，可用于评估证据充分性、识别未知情况并提高模型在真实环境中的决策可靠性。

上述环节共同构成微视觉计算的技术流程，为不同任务中的证据获取、特征建模与关系理解提供系统支撑。后续章节将进一步从共性技术机制出发，梳理代表性算法及其在典型应用场景中的具体实现。

基于方法的作用阶段与优化目标，本文将微视觉计算中的典型算法归纳为三个层次：**输入层的信**

© 中国图象图形学报版权所有

号增强与数据生成、模型层的精细判别与稳健表征，以及数据与任务层的多模态与多任务统一学习，整体框架如图2所示。该分类不以具体任务或应用场景为边界，而是从微视觉证据处理过程中的共性技术机制出发，系统梳理不同研究方向之间可复用的方法路径。具体而言，输入层方法主要通过图像增强、细节重建、时序信号强化、频域分析和样本生成，提高微弱或微小证据的可获得性与可辨识性；模型层方法侧重利用多尺度表征、局部注意、时空建模、结构建模和基础模型迁移，增强细微证据的保持、区分与稳健表达；数据与任务层方法则进一步通过多模态协同、多任务联合和跨域适应，实现不同数据源、任务与场景之间的信息互补和知识迁移。三类方法分别对应微视觉证据的输入改善、特征建模与知识共享，共同构成由证据增强到任务决策的完整技术体系。从核心问题与技术路径的对应关系来看，信号增强与数据生成主要针对观测临界性，通过信号强化、细节重建和样本扩充提高微弱证据的可观测性与可利用性；精细判别与稳健表征主要针对表征易失性，通过多尺度特征融合、局部差异强化和时空建模减少关键证据在特征提取过程中的损失；多模态与多任务统一学习则主要面向判别欠定性，通过跨模态信息互补、上下文关联和跨任务知识共享提高有限证据条件下的判别可靠性。需要指出的是，上述对应关系并非相互独立或严格的一一映射，例如，多模态融合也能够缓解单一传感器的观测局限，而上下文关系建模同样可以在模型表征阶段降低判别不确定性。因此，三类技术路线分别从不同层次切入，又相互补充，共同应对微视觉计算中的三个核心困难。

2.2 信号增强与数据生成

微视觉信息在原始观测中通常存在对比度低、信噪比不足、空间分辨率有限或动态变化幅度较小等问题，使任务相关证据难以被稳定提取。为改善微视觉信息的可辨识程度，现有研究主要从信号增强、信息重建、频域变换和样本生成等方面开展工作。此类方法作用于原始输入或中间信号空间，通过恢复（Bergmann 等，2019；Bergmann 等，2022；Fan 等，2020）、强化（Zhao 等，2024；Hu 等，2021；Li 等，2013）或扩充任务相关信息（Ge 等，2024；Wang 等，2023），为后续特征学习提供更充分的数据基础。根据处理对象与技术目标，可将其概括为图像增强与噪声抑制（Bergmann 等，2019；Bergmann 等，2022；Zou 等，2022；Fan 等，2020；Zhao 等，2024；Hu 等，2021）、超分辨率与细节重建（Wang 等，2021；Zhu 等，2018；Bao 等，2024；Zhu 等，2025）、弱动态与时序信号增强（Yan 等，2013；Yan 等，2014；Davison 等，2018；Li 等，2013；Ben 等，2021；Li 等，2022；Zhao 等，2023；Sribabu 等，2026；De Haan 等，2013；Chen 等，2018；Yu 等，2019；Yu 等，2022；Zou 等，2024；Qian 等，2026）、频域与残差信号提取（Qian 等，2020；Liu 等，2021；Wang 等，2023），以及数据生成与样本扩充（Ge 等，2024；Ojha 等，2023；Zhu 等，2023；Wang 等，2023；Yan 等，2025）等方向。

图像增强与噪声抑制。对于低照度、低对比度及复杂成像条件下的微视觉信息，灰度变换、直方图均衡、局部对比度增强、图像去噪和照明校正等方法被广泛用于改善输入质量。全局增强方法能够调整图像整体亮度和对比度，但可能同时放大背景噪声并削弱局部差异；局部自适应方法则根据区域统计特征突出弱纹理、细边界和异常结构，更适合处理亮度分布不均或局部证据稀疏的场景（Bergmann 等，2019；Bergmann 等，2022；Zou 等，2022；Fan 等，2020；Zhao 等，2024；Hu 等，2021）。近年来，基于卷积神经网络（convolutional neural network，CNN）、Transformer 和生成模型的学习式方法（Ge 等，2024；Wang 等，2023）进一步建立退化观测与高质量图像之间的映射，并通过任务监督保留具有判别价值的局部信息。与传统增强方法相比，学习式方法具有更强的自适应能力，但其性能通常依赖于退化模型、训练数据和实际成像条件之间的一致性。

超分辨率与细节重建。当微小目标或局部证据仅由少量像素表示时，成像过程、图像压缩和网络下采样会造成纹理、边缘及内部结构的显著损失。超分辨率方法（Wang 等，2021；Zhu 等，2018；Bao 等，2024；Zhu 等，2025）通过单帧恢复、多帧融合或生成式先验重建高频细节，为微小目标检测、细粒度识别和微小病灶分析提供更充分的空间信息。与面向视觉质量的一般图像复原不同，微视觉任务更加关注重建信息对下游识别结果的实际贡献。因此，部分研究将超分辨率模块与检测（Yu 等，2020；Guo 等，2023）、分割（Yin 等，2024；Roth 等，2022）或分类（Roth 等，2022；Deng 和 Li，2022）网络进行联合优化（Wang 等，2024；Ge 等，2024），使细节恢复过程直接

© 中国图象图形学报版权所有

受到任务目标约束。此类方法需要重点避免生成不存在的结构或纹理，以免虚假细节干扰后续判断。

弱动态与时序信号增强。微表情、细微动作、远程生理信号和微振动等任务的关键信息主要表现为短时、低幅或弱周期变化（Yan 等，2013；Yan 等，2014；Davison 等，2018；Li 等，2013；Ben 等，2021；Li 等，2022；Zhao 等，2023；Sribabu 等，2026）。运动放大、时域滤波、频带分解和相位分析等方法通过增强帧间变化，提高弱动态信号的可观测程度。对于基于 rPPG 的非接触式生理信号测量，CHROM（De Haan 和 Jeanne，2013）利用色度通道的线性组合分离脉搏信号，DeepPhys（Chen 和 McDuff，2018）和 PhysNet（Yu 等，2019）采用卷积注意力与时序网络建模肤色变化，PhysFormer（Yu 等，2022）进一步引入时序差分 Transformer 捕捉远程光电容积脉搏波的时域自相似性，RhythmMamba（Zou 等，2024）则将状态空间模型用于长序列生理信号的高效建模，Freq-Phys（Qian 等，2026）显式利用生理频带先验抑制运动与光照噪声。传统方法通常依据预设频段或运动假设进行信号放大，容易受到相机抖动、头部运动和环境变化的干扰。深度学习方法则通过时序卷积、循环网络和注意力机制学习任务相关动态，自适应地区分有效变化与随机噪声。该类方法的关键在于，在放大目标动态的同时抑制非目标运动，避免噪声、背景扰动和压缩伪影被同步增强。

频域与残差信号提取。部分微视觉证据在空间域中表现并不明显，却可能在频谱分布、相位结构或高频残差中呈现更稳定的差异。工业缺陷、图像篡改和生成内容中的异常通常伴随局部频率偏移、边界不连续、周期纹理异常或生成残差。相关方法通过傅里叶变换（Qian 等，2020）、小波分解（Liu 等，2021）、高通滤波（Qian 等，2020；Liu 等，2021）和可学习频率模块（Wang 等，2023），将视觉信号分解至不同频段，并强化与任务相关的频率成分。空间域与频域联合建模（Qian 等，2020；Liu 等，2021；Wang 等，2023）能够兼顾目标语义与细微异常，但压缩方式、传感器噪声和图像处理操作同样会改变频率分布。因此，频域方法需要区分任务相关异常与成像链路产生的非语义干扰，以避免模型依赖特定数据集或压缩条件中的偶然特征。

数据生成与样本扩充。微视觉任务通常面临有效样本数量有限、类别分布不平衡、困难样本稀缺及极端条件覆盖不足等问题。传统数据增强通过裁剪、缩放、颜色扰动、几何变换和样本混合扩展训练分布，生成式方法则能够进一步合成微小目标、局部缺陷、低幅动态和不同环境条件下的视觉样本（Ge 等，2024；Ojha 等，2023；Zhu 等，2023；Wang 等，2023；Yan 等，2025）。生成对抗网络和扩散模型还可模拟目标由清晰到退化、由显著到隐蔽或由正常到异常的连续变化，为模型提供更丰富的边界样本和困难样本（Ge 等，2024；Wang 等，2023）。除直接增加样本数量外，生成方法还可用于类别均衡、反事实构造和跨域风格迁移（Ojha 等，2023；Zhu 等，2023；Yan 等，2025）。然而，生成数据可能包含不真实的纹理、结构偏差或模型特有伪影，需要结合质量评估、语义约束、真实性筛选及真实样本联合训练，防止模型将生成器痕迹误认为任务依据。

总体而言，信号增强与数据生成方法通过改善观测质量、恢复局部细节、突出弱动态并扩展训练分布，提高微视觉证据的可获得性和可辨识性。然而，此类方法的有效性仍受到原始信息完整程度和生成质量的限制。过度增强可能放大无关噪声，重建和生成过程也可能引入虚假证据。因此，当前研究正逐渐由独立的数据预处理转向任务驱动的联合优化，使增强、重建与生成过程受到下游识别目标的约束，从而在信息强化与真实性保持之间取得平衡。

2.3 精细判别与稳健表征

经过输入增强与信息重建后，微视觉证据仍可能在网络下采样、特征聚合和高层语义抽象过程中被逐渐削弱。尤其当局部证据所占区域有限、动态变化幅度较低或目标与背景高度相似时，常规模型容易受到显著区域和主导语义的干扰。模型层方法的核心目标是在保持整体结构与语义信息的同时，对局部细节、弱动态变化和隐含关系进行精确建模，并提高所得表征在光照、背景、设备和场景变化下的稳定性。相关研究主要包括高分辨率与多尺度表征、局部差异强化、关键区域建模、时空动态建模、上下文关系推理、泛化学习以及基础模型适配等方向。

高分辨率与多尺度表征。微视觉证据在不同空间尺度上具有不同的表征需求。浅层特征保留较多纹理、边缘和位置信息，但语义区分能力有限；深层特征具有更强的抽象表达能力，却可能因连续下采样而丢失细小结构。特征金字塔、多尺度卷积、跨层连接和高分辨率网络通过融合不同层级的空间细节

© 中国图象图形学报版权所有

与语义信息，缓解局部保真与全局理解之间的矛盾。在视频理解中，TSM（Lin等，2019）通过时序移位模块实现高效率的多尺度时序建模；Video Swin Transformer（Liu等，2022）采用分层注意力机制捕获不同尺度的时空依赖。在微小目标检测中，RFLA（Xu等，2022）通过高斯感受野标签分配适配微小目标的尺度特性，NWD-RKA（Xu等，2022）利用归一化Wasserstein距离缓解微小边界框的位置敏感性，DQ-DETR（Huang等，2024）结合动态查询与密度引导特征增强提升密集微小目标的检测性能。从技术特点来看，高分辨率表征更侧重减少空间下采样造成的细节损失，适用于微小目标和局部结构的精确感知；多尺度融合则强调不同层级特征之间的语义互补，在目标尺度变化较大或背景复杂的场景中具有优势。二者的共同挑战在于如何以合理的计算开销实现关键尺度与有效特征的自适应选择，避免冗余特征融合稀释微弱判别信息。

局部细节与差异强化。许多微视觉任务的判别依据并非显著的绝对响应，而是相邻区域之间的细微变化。差分卷积、局部对比学习、边缘增强、高频卷积和梯度建模等方法通过显式刻画中心与邻域、目标与背景或不同局部区域之间的相对差异，强化弱边界、纹理突变和细小结构变化。在人脸反欺诈中，NAS-FAS（Yu等，2020）通过静态-动态中心差分网络搜索增强局部纹理差异的建模能力；SSDG-R（Jia等，2020）和SSAN-R（Wang等，2022）通过风格解耦与组装强化域不变局部特征。在图像异常检测中，PatchCore（Roth等，2022）利用预训练特征的局部块记忆库实现异常定位，部分研究进一步设计多方向、多尺度或自适应差分算子，以适应裂纹、轮廓、斑点和局部伪影等不同形态。从技术机制来看，差分卷积、边缘增强及梯度建模等方法通过显式局部算子突出细微结构变化，具有较强的局部差异感知能力；局部对比学习和预训练特征匹配则更侧重利用学习得到的特征分布识别细微异常，对复杂外观变化具有更灵活的适应能力。然而，前者容易受到噪声和无关高频信息的干扰，后者则依赖训练样本或预训练表征的覆盖能力，因此需要在局部细节敏感性与语义判别稳定性之间取得平衡。

关键区域与部件建模。在细粒度识别、工业缺陷检测、医学病灶分析和生成内容鉴别中，目标整体外观往往无法提供充分的区分信息，真正具有任务价值的证据可能仅分布于少量局部区域。区域提议、关键点检测、部件发现、显著区域搜索和注意力机制通过定位高价值区域，将有限的模型容量集中于细微判别证据。在微表情识别中，LBP-TOP（Zhao和Pietikainen，2007）通过局部二值模式在三个正交平面上的统计特征捕获面部关键区域的时空纹理变化；SLSTT（Zhang等，2022）联合短距离与长距离时空关系建模局部面部运动；HTNet（Wang等，2024）利用层次化Transformer提取局部肌肉运动及其全局语义关系；Ra-ICL（Cai等，2026）通过多实例运动表示和区域一致性约束挖掘分散的弱运动线索。弱监督方法能够在缺少像素级或部件级标注时，根据图像级标签自动发现关键位置；多区域或多部件方法则通过互补特征聚合，避免判断过度依赖单一局部。该类方法需要同时解决定位准确性与区域多样性问题，防止模型关注偶然相关的背景模式，或因注意力塌缩而遗漏其他有效证据。

时空动态建模。对于微表情、微姿态与微动作、远程生理信号和微振动分析，判别信息不仅存在于单帧外观中，更体现在连续观测之间的低幅变化及其演化规律中。光流、动态纹理、三维卷积、循环网络和时序Transformer可用于建模短时运动、周期变化和长期依赖。在微姿态与微动作识别中，MANet（Guo等，2024）通过多尺度注意力网络捕获细微身体运动；MSF-Mamba-M+（Li等，2026）提出运动感知状态融合与多尺度时空建模增强相似微姿态之间的区分能力；VideoMamba（Li等，2024）将状态空间模型引入视频理解，实现对连续时序信息的高效建模。从时间建模范围来看，光流、局部时序卷积及关键帧选择更侧重捕获短时、局部的细微运动，适用于微表情等瞬态变化分析；时序Transformer和状态空间模型则更强调长时依赖与动态演化关系，适用于连续动作及周期性信号建模。不同方法均需处理微弱动态与相机抖动、主体整体运动及环境扰动之间的混淆，并在细微运动保真、长时关系建模和计算效率之间进行权衡。

上下文与关系表征。微视觉证据的判别意义通常取决于其与背景、其他对象、整体结构和先验知识之间的关系。上下文建模通过联合表示目标区域与周围环境，判断局部证据是否符合整体场景；图神经网络和Transformer则能够刻画对象之间、区域之间以及不同时刻之间的依赖关系。在伪装目标检测

© 中国图象图形学报版权所有

中，SINet-v2（Fan等，2021）采用先搜索后识别的两阶段策略捕获目标-背景上下文差异；ZoomNet（Pang等，2022）通过混合尺度三元组网络增强不同尺寸目标的上下文感知；FSPNet（Huang等，2023）利用特征收缩金字塔与Transformer全局关系建模累积微弱但具有判别性的目标线索；CamoFormer-P（Yin等，2024）将注意力头分别用于前景、背景和全局关系建模，更有效地区分高度相似的目标与周围环境。对于伪装目标，模型需要分析目标纹理与背景结构之间的细微差异；对于上下文异常和生成内容鉴别，局部区域本身可能具有较高真实性，其异常主要体现为物理结构、空间连续性或语义逻辑上的不一致。关系表征由此将难以直接观测的隐含差异转化为可学习、可比较的判别信号。

不变性与泛化学习。微视觉证据通常对光照、姿态、设备、压缩方式和背景变化较为敏感。若模型主要依赖训练数据中的背景共现、设备噪声或压缩模式，便可能在跨数据集和真实环境中出现明显性能下降。域适应、域泛化、对比学习、特征解耦和因果学习通过约束不同环境下的表示一致性，分离任务相关因素与环境相关因素。在人脸反欺诈领域，MADDG（Shao等，2019）采用多对抗判别深度域泛化；RFMeta（Shao等，2020）通过正则化细粒度元学习提升跨域泛化能力；D2AM（Chen等，2021）学习可泛化的混合域表示；SA-FAS（Sun等，2023）从可分离性与对齐角度重新审视域泛化；GAC-FAS（Le和Woo，2024）利用梯度对齐增强跨域泛化；TTDG-V（Zhou等，2024）探索测试时域泛化策略。从泛化机制来看，域对齐与对抗学习主要通过减小不同环境之间的特征分布差异提高迁移能力，元学习侧重模拟域变化以增强模型的快速适应能力，而特征解耦与因果约束则尝试分离任务相关因素和环境干扰因素，以获得更加稳定的判别依据。然而，微视觉任务中的关键证据往往表现为细微的局部差异，过强的分布对齐或不变性约束可能同时削弱有效判别信息，因此需要在环境鲁棒性与微弱证据保真之间建立合理平衡。

基础模型与生成式表征。大规模视觉及视觉语言基础模型具有丰富的语义知识和较强的迁移能力，但其预训练目标通常偏向显著目标、整体语义和自然图像中的主导模式，未必能够直接感知低对比度边界、细小部件和局部异常。现有研究通过提示学习、参数高效微调、局部特征注入、多层特征融合和高分辨率适配，将基础模型知识迁移至微视觉任务。CLIP-SA（Ma等，2026）利用对比语言－图像预训练模型（Contrastive Language－Image Pre-training，CLIP）引导的语义对齐增强人脸反欺诈的跨域泛化能力；DGPDL（Liu等，2026）通过域引导的提示分布学习实现更强的泛化性能。VideoLLaMA2（Cheng等，2024）和Qwen2. 5-7B-VL（Bai等，2025）将多模态大模型扩展至视频理解与推理任务，为微视觉分析提供语义先验。生成式模型，特别是扩散模型，在逐步去噪过程中形成由整体结构到局部细节的多层级表示，为弱边界、细粒度结构和局部伪影的统一建模提供了新的可能。然而，如何避免高层语义先验覆盖底层微弱证据，并控制模型规模、计算成本和任务适配难度，仍是基础模型应用于微视觉计算需要解决的问题。

总体而言，精细判别与稳健表征决定了微视觉证据能否在模型内部得到有效保留、准确区分和稳定利用。其关键并非单纯增加网络深度或参数规模，而是针对微弱信号、有限空间和上下文依赖建立具有选择性的表征机制，并使所得特征能够在复杂干扰和分布变化下保持可靠性。

2.4　多模态与多任务统一学习

单一模态通常只能反映视觉场景的部分属性，单一任务监督也难以充分揭示微视觉信息的多层次结构。可见光图像包含丰富的颜色与纹理信息，但容易受到光照、遮挡和背景干扰；热红外能够反映温度差异，却缺少细致纹理；深度、雷达、事件相机和多光谱数据能够补充几何、距离、动态或光谱信息，但在空间分辨率、采样频率和数据分布方面存在明显差异。与此同时，微小目标检测、细粒度识别、伪装目标感知和局部异常分析虽然具有不同的输出目标，却共享对局部结构、弱动态和上下文关系的建模需求。因此，从模态、任务和场景层面实现信息互补与知识共享，是提升微视觉模型泛化能力、数据利用效率和部署适应性的重要方向。

多模态信息互补。多模态学习通过联合利用可见光、热红外、深度、雷达、事件相机和多光谱等数据，弥补单一传感器在特定环境下的信息缺失。在工业异常检测中，MVTec 3D-AD（Bergmann等，2021）同时提供RGB图像与3D点云，Real-IAD D3（Zhu等，2025）进一步同步提供高分辨率RGB图像、

© 中国图象图形学报版权所有

伪三维信息与微米级三维点云；Real3D-AD（Liu等，2023）专注于纯几何形态的异常检测；Eyecandies（Bonfiglioli等，2022）融合RGB与深度图像覆盖外观与几何异常；MulSen-AD（Li等，2025）集成RGB、红外热像和3D点云三种传感器数据；InspecSafe-V1（Liu等，2026）进一步涵盖可见光、红外、深度、雷达、音频、气体和温湿度共7种模态，为多传感器融合的工业安全评估提供评测平台。在人脸反欺诈中，CASIA-SURF CeFA（Liu等，2021）同时考虑可见光、深度和红外信息并引入跨族裔评价。例如，可见光与热红外融合能够提高低照度、夜间和复杂天气条件下的目标感知能力；RGB与深度信息结合有助于区分颜色相近但空间结构不同的对象；事件数据则能够补充普通视频在高速运动、低延迟和高动态范围场景中的观测不足。根据融合发生的位置，多模态方法可分为输入级、特征级和决策级融合。输入级融合保留较完整的原始对应关系，但对配准误差较为敏感；特征级融合具有较强的交互建模能力，但结构和计算复杂度较高；决策级融合相对灵活稳健，却可能无法充分利用模态之间的细粒度关联。

跨模态对齐与共享表示。不同模态在成像机制、空间尺度、采样频率和统计分布上存在差异，未经对齐的直接拼接容易造成位置错位与语义混淆。跨模态对齐通过空间配准（Bergmann等，2021；Zhu等，2025；Zhu等，2018）、时间同步（Yan等，2014；Davison等，2018；Stricker等，2014；Hu等，2021）、特征映射（Liu等，2021）和对比学习（Guo等，2023；Huang等，2026），将不同来源的信息投影至具有一致语义的表示空间。交叉注意力（Guo等，2023；Ma等，2026）、双向交互（Huang等，2026）和动态门控（Li等，2025）模块可进一步根据场景条件估计各模态的有效性，选择性地利用互补信息，降低低质量模态对最终预测的干扰。针对实际系统中可能出现的模态缺失、传感器故障和异步采样问题（Liu等，2026），模态补全（Wang等，2021）、知识蒸馏（Liu等，2021）和缺失模态鲁棒学习（Zhou等，2024）也逐渐受到关注。

多任务共享与知识迁移。检测、分割、分类、定位、边界提取和状态估计等任务虽然具有不同的输出形式，但往往共享纹理、轮廓、空间结构和上下文信息。多任务学习通过共享骨干网络（Guo等，2024；Liu等，2021）、层级特征（Yin等，2024；Fan等，2021）或部分参数（Cai等，2024；Ma等，2026；Liu等，2026），使不同监督信号相互补充。例如，分割任务能够为分类提供更精确的目标区域（Le等，2019；Fan等，2020），边界检测有助于提升微小目标和低对比度区域的定位质量（Roth等，2022；Deng和Li，2022），重建任务则能够约束模型保留局部细节（Deng和Li，2022；Zavrtanik等，2021）。辅助任务还可在精细标注不足时提供额外监督，提高特征学习的稳定性。然而，不同任务对于特征尺度和语义抽象程度的需求并不完全一致，可能产生梯度冲突和参数竞争，因此需要通过损失加权（Guo等，2023）、梯度平衡（Liu等，2025）、任务分支（Liu，2024）和参数解耦（Cai等，2024）实现有效协调。

跨场景统一表征。微小目标、细粒度类别、伪装目标和局部伪影在视觉表现上存在明显差异，但其共同目标均是发现容易被显著模式掩盖的任务关键证据。跨场景统一学习尝试利用共享特征空间（You等，2022；Guo等，2023）、通用查询（Huang等，2024）、文本条件提示（Li等，2026）和统一预测接口（Li等，2026），将多个微视觉场景纳入同一模型。共享主干负责学习具有迁移能力的基础结构和局部模式（Guo等，2024；Li等，2025），任务提示（Li等，2026）、适配器（Cai等，2024；Ma等，2026）或独立预测头（Li等，2026；Li等，2025）则用于保留不同数据集的语义体系与输出要求。相比针对每个任务分别训练独立模型，统一表征能够减少重复训练和部署成本，并促进边界、纹理、部件和上下文知识在不同任务之间迁移。但统一并不意味着完全共享，任务间差异仍需要通过条件化模块和任务特定参数加以保留。

任务平衡与负迁移抑制。多任务和多数据集联合训练通常面临数据规模不均衡、类别空间异构、标注形式不同和优化目标冲突等问题。大规模或易学习的数据集可能主导模型更新，使小样本任务和困难任务无法得到充分训练；某一任务所需要的特征也可能对其他任务产生干扰。常见解决方法包括任务均衡采样（Liu等，2021；Guo等，2024）、动态损失加权（Guo等，2023）、梯度投影与解耦（Liu等，2025）、稀疏任务路由（Liu，2024）以及参数高效适配（Cai等，2024；Ma等，2026；Liu等，2026）。部分方法还根据任务相似度动态确定共享范围，使关联程度较高的任务共享更多参数，而差异较大的任务保持

© 中国图象图形学报版权所有

相对独立。统一模型只有在知识共享收益大于任务干扰时才具有实际优势，因此需要在公共表示能力与任务特异性之间进行动态权衡。

开放环境与持续适应。真实应用中的类别体系、传感器配置、任务需求和数据分布会随时间持续变化，统一模型需要具备吸收新知识并适应新环境的能力。开放集识别和开放词汇学习用于识别训练阶段未出现的类别或异常模式（Ojha等，2023；Zhu等，2023；Yan等，2025），增量学习（Wang等，2021；Zhou等，2024）和持续学习（Shao等，2020；Wang等，2022）则使模型能够在较少重新训练的条件下学习新任务、新设备和新场景，同时减少对已有知识的遗忘。视觉语言模型所提供的语义对齐和开放词汇能力，为微视觉任务突破固定标签空间提供了新的基础（Ma等，2026；Liu等，2026；Huang等，2026）。然而，微视觉类别之间往往差异细微，单纯依赖文本语义仍难以完成精确区分，因此需要进一步结合局部视觉提示、细粒度描述和不确定性估计，提高开放环境中的识别可靠性。

3 代表性数据集

现有微视觉数据集的演进不仅体现为样本规模和任务类别的扩展，更反映了研究重点由受控条件下的单一视觉证据逐步转向复杂环境中的微视觉证据获取、稳定表征与关系判别。不同任务的数据集虽然具有不同的数据形式和标注方式，但总体呈现出由单一模态向多模态、由静态向动态、由受控环境向自然场景、由封闭类别向开放分布，以及由孤立局部证据向上下文和关系信息扩展的发展趋势。

表1　工业图像异常检测与定位的代表性数据集

Table 1　Representative datasets for industrial image anomaly detection and localization

数据集	数据形式	类别数	数据规模	数据集链接
MVTec AD(Bergmann等，2019)	RGB图像	15	5,354张图像	https://www.mvtec.com/research-teaching/datasets/mvtec-ad
MVTec LOCO AD(Bergmann等，2022)	RGB图像	5	3,644张图像	https://www.mvtec.com/research-teaching/datasets/mvtec-loco-ad
VisA(Zou等，2022)	RGB图像	12	10,821张图像	https://registry.opendata.aws/visa/
MVTec 3D-AD(Bergmann等，2021)	RGB图像，3D点云	10	4,147组样本	https://www.mvtec.com/research-teaching/datasets/mvtec-3d-ad
Real-IAD(Wang等，2024)	RGB图像	30	约15万张图像	https://realiad4ad.github.io/Real-IAD/
Real-IAD D3(Zhu等，2025)	RGB图像，3D点云	20	8,450组样本	https://realiad4ad.github.io/Real-IAD_D3/
Real3D-AD(Liu等，2023)	3D点云	12	1,254组样本	https://github.com/m-3lab/real3d-ad
Eyecandies(Bonfiglioli等，2022)	RGB图像，深度	10	15,000张图像	https://eyecan-ai.github.io/eyecandies/
MulSen-AD(Li等，2025)	RGB图像，红外	15	2,035组样本	https://github.com/MapleBoat/MulSen-AD
InspecSafe-V1(Liu等，2026)	RGB图像，深度，红外等	5	5,013个实例	https://github.com/liuzy0708/InspecSafe

工业图像异常检测与定位。工业异常数据集的发展体现了由局部低显著缺陷向复杂结构和关系异常的逐步扩展，代表性数据集见表1。MVTec AD（Bergmann等，2019）主要关注划痕、裂纹等局部异常，为微小低对比证据的观测与保持建立了经典基准；MVTec LOCO AD（Bergmann等，2022）进一步引

© 中国图象图形学报版权所有

入部件缺失、错误排列等逻辑异常，使问题由局部外观差异扩展至结构关系判别。VisA（Zou等，2022）和Real-IAD（Wang等，2024）强化复杂多实例与真实多视角场景，而Real-IAD D3（Zhu等，2025）、Real3D-AD（Liu等，2023）、Eyecandies（Bonfiglioli等，2022）、MulSen-AD（Li等，2025）和InspecSafe-V1（Liu等，2026）进一步引入几何和多传感器信息。总体而言，该方向由改善局部异常的可观测与表征，逐步走向利用多模态和结构关系缓解复杂异常的判别欠定性。

微表情识别。微表情数据集首先需要解决短时、低幅运动的观测临界性，随后逐渐关注这些细微动态在长视频和复杂场景中的保持与判别，代表性数据集见表2。SMIC（Li等，2013）、CASME（Yan等，2013）、CASME II（Yan等，2014）和SAMM（Davison等，2018）主要通过提高帧率和空间分辨率增强微弱面部运动的可观测性；CAS（ME）2（Qu等，2018）、MMEW（Ben等，2021）和CAS（ME）3（Li等，2022）进一步扩展至长视频、宏微表情共存及定位识别联合设置，使表征易失问题更加突出。4DME（Li等，2022）、DFME（Zhao等，2023）和BlendME（Sribabu等，2026）则从三维几何、细粒度动态及Blendshape时序等角度增加互补证据，以提升复杂条件下的稳定判别能力。

表2　微表情识别的代表性数据集

Table 2　Representative datasets for micro-expression recognition

数据集	数据形式	类别数	数据规模	数据集链接
SMIC(Li等，2013)	RGB/红外视频	3	164个片段	https://doi.org/10.1109/FG.2013.6553717
CASME(Yan等，2013)	RGB视频	8	195个片段	https://doi.org/10.1109/FG.2013.6553799
CASME II(Yan等，2014)	RGB视频	5	247个片段	https://casme.psych.ac.cn/casme/e2
SAMM(Davison等，2018)	RGB视频	7	159个片段	https://e-space.mmu.ac.uk/617069/
CAS(ME)2(Qu等，2018)	RGB视频	–	87个片段	https://doi.org/10.1109/TAFFC.2017.2654440
MMEW(Ben等，2021)	RGB视频	7	1200个片段	https://github.com/benxianyeteam/MMEW-Dataset
4DME(Li等，2022)	4D多模态序列	–	–	–
CAS(ME)3(Li等，2022)	RGB-D视频	–	4599个片段	https://casme.psych.ac.cn/casme/e4
DFME(Zhao等，2023)	RGB视频	–	–	https://mea-lab-421.github.io/
BlendME(Sribabu等，2026)	Blendshape时序	–	112个片段	https://swethasree-sribabu.github.io/BlendME

微姿态与微动作识别。微姿态与微动作数据集逐渐由局部、短时动作观测扩展至长视频中的证据保持及高层状态理解，代表性数据集见表3。iMiGUE（Liu等，2021）和SMG（Chen等，2023）关注自然或受控场景中的细微身体行为，MA-52（Guo等，2024）进一步扩展至全身微动作及身体区域标注；iMiGUE-3K（Wang等，2026）将问题拓展至大规模长视频，使稀疏动作在冗余时序中的稳定表征成为新的挑战。MMA-52（Li等，2025）、DEEMO（Li等，2025）和MA-Bench（Li等，2026）进一步引入多标签、多模态和视觉问答设置，使研究由动作本身的识别逐步走向结合上下文与多模态关系的状态理解。

人脸反欺诈。人脸反欺诈数据集的发展体现了由受控条件下的细微伪造痕迹识别向开放环境下可靠判别的转变，代表性数据集见表4。CASIA-FASD（Zhang等，2012）、Replay-Attack（Chingovska等，2012）和MSU-MFSD（Wen等，2015）主要围绕打印、回放和设备变化构建可观测的攻击差异；OULU-NPU（Boulkenafet等，2017）和SiW（Liu等，2018）进一步增加姿态、光照和环境变化，对微弱活体证据的稳定表征提出更高要求。SiW-M（Liu等，2019）、CelebA-Spoof（Zhang等，2020）和CASIA-SURF CeFA

© 中国图象图形学报版权所有

表3　微姿态与微动作识别的代表性数据集

Table 3　Representative datasets for micro-posture and micro-action recognition

数据集	数据形式	类别数	数据规模	数据集链接
iMiGUE (Liu等,2021)	RGB 视频	32	18,499个片段	https://github.com/linuxsino/iMiGUE
SMG (Chen等,2023)	RGB 视频	16	3,712个片段	https://oulurepo.oulu.fi/handle/10024/45261
MA-52 (Guo等,2024)	RGB 视频	52	22,422个片段	https://github.com/VUT-HFUT/Micro-Action
iMiGUE-3K (Wang等,2026)	RGB 视频	32	约3,400段视频	https://arxiv.org/abs/2605.17179
MMA-52 (Li等,2025)	RGB 视频	52	6,528段视频	https://github.com/VUT-HFUT/Micro-Action
DEEMO (Li等,2025)	RGB 视频	37	24,722个身体语言实例	https://arxiv.org/abs/2504.19549
MA-Bench (Li等,2026)	4D 多模态序列	52	20,510个片段	https://github.com/MA-Bench/MA-Bench

表4　人脸反欺诈的代表性数据集

Table 4　Representative datasets for face anti-spoofing

数据集	数据形式	类别数	数据规模	数据集链接
CASIA-FASD(Zhang等,2012)	RGB 视频	2	600段	http://www.cbsr.ia.ac.cn/english/FaceAntiSpoofDatabases.asp
Replay-Attack(Chingovska等,2012)	RGB 视频	2	1,300段	https://www.idiap.ch/en/scientific-research/data/replayattack
MSU-MFSD(Wen等,2015)	RGB 视频	2	280段	https://cvlab.cse.msu.edu/project-face-anti.html
OULU-NPU(Boulkenafet等,2017)	RGB 视频	2	4,950段	https://sites.google.com/site/oulunpudatabase
SiW(Liu等,2018)	RGB 视频	2	4,478段	https://cvlab.cse.msu.edu/siw-spoof-in-the-wild-database.html
SiW-M(Liu等,2019)	RGB 视频	2	1,630段	https://cvlab.cse.msu.edu/project-face-anti.html
CelebA-Spoof(Zhang等,2020)	RGB 图像	2	625,537	https://github.com/ZhangYuanhan-AI/CelebA-Spoof
CASIA-SURF CeFA(Liu等,2021)	RGB/深度/红外	2	23,538段	https://doi.org/10.1109/WACV48630.2021.00122

(Liu等,2021)进一步覆盖未知攻击、细粒度属性、多模态与跨族裔场景,使评测重点由局部纹理差异转向跨攻击、跨设备和跨人群条件下的判别确定性。

人脸Deepfake与通用AIGC检测。随着生成质量提高,相关数据集逐渐由明显伪影识别转向对微弱、易失和关系依赖证据的综合评价,代表性数据集见表5。FaceForensics++(Rossler等,2019)、Celeb-DF v2(Li等,2020)和DFDC(Dolhansky等,2020)主要推动不同操纵方式和生成质量下的伪造检测;DeeperForensics-1.0(Jiang等,2020)与WildDeepfake(Zi等,2020)进一步引入压缩、噪声和真实网络分布,使伪造证据的表征易失性更加突出,FakeAVCeleb(Khalid等,2021)则扩展至视听一致性。UniversalFakeDetect(Ojha等,2023)、GenImage(Zhu等,2023)、DIRE(Wang等,2023)、WildFake(Hong和Zhang,2024)、Chameleon(Yan等,2025)和AIGCDetectBenchmark(Zhong等,2023)进一步面向未知生成器和高真实性内容,使问题逐渐由局部伪影发现转向跨区域、跨模态及语义一致性的可靠判别。

医学图像异常检测。医学异常检测数据集覆盖MRI、CT、OCT、X光和病理等多种成像模态,其核心挑战由微小、弱对比病灶的可靠观测进一步延伸至跨模态表征和上下文判别,代表性数据集见表6。BMAD(Bao等,2024)通过统一预处理和数据划分,将BraTS 2021(Menze等,2014)、BTCV与LiTS(Landman等,2015;Bilic等,2023)、RESC(Chen等,2020)、OCT2017(Kermany等,2018)、RSNA Pneumonia(Shih等,2019)及Camelyon16(Ehteshami Bejnordi等,

© 中国图象图形学报版权所有

表5　人脸Deepfake与通用AIGC检测的代表性数据集

Table 5　Representative datasets for face deepfake and general AIGC detection

数据集	数据形式	类别数	数据规模	数据集链接
FaceForensics++(Rossler等,2019)	人脸RGB视频	2	5,000段	https://github.com/ondyari/FaceForensics
Celeb-DF v2(Li等,2020)	人脸RGB视频	2	6,229段	https://github.com/yuezunli/celeb-deepfakeforensics
DFDC(Dolhansky等,2020)	人脸RGB视频	2	128,154段	https://ai.meta.com/datasets/dfdc/
DeeperForensics-1.0(Jiang等,2020)	人脸RGB视频	2	60,000段	https://github.com/EndlessSora/DeeperForensics-1.0
WildDeepfake(Zi等,2020)	人脸RGB视频	2	7,314段	https://github.com/OpenTAI/wild-deepfake
FakeAVCeleb(Khalid等,2021)	人脸RGB视频	2	20,000段	https://github.com/DASH-Lab/FakeAVCeleb
UniversalFakeDetect(Ojha等,2023)	通用RGB图像	2	–	https://github.com/WisconsinAIVision/UniversalFakeDetect
GenImage(Zhu等,2023)	通用RGB图像	2	133万张	https://github.com/GenImage-Dataset/GenImage
DIRE(Wang等,2023)	通用RGB图像	2	–	https://github.com/ZhendongWang6/DIRE
WildFake(Hong和Zhang,2024)	通用RGB图像	2	–	https://ojs.aaai.org/index.php/AAAI/article/view/32363
Chameleon(Yan等,2025)	通用RGB图像	2	11,000张	https://shilinyan99.github.io/AIDE/index_aide.html
AIGCDetectBenchmark(Zhong等,2023)	通用RGB图像	2	–	https://github.com/Ekko-zn/AIGCDetectBenchmark

表6　医学图像异常检测的代表性数据集

Table 6　Representative datasets for medical image anomaly detection

数据集	数据形式	类别数	数据规模	数据集链接
BraTS 2021(Menze等,2014)	MRI图像	1	11,298张	https://www.synapse.org/brats2021
BTCV + LiTS(Landman等, 2015; Bilic等, 2023)	CT图像	1	3,201张	https://github.com/DorisBao/BMAD
RESC(Bao等,2024)	OCT图像	1	6,217张	https://github.com/DorisBao/BMAD
OCT2017(Kermany等,2018)	OCT图像	4	27,315张	https://data.mendeley.com/datasets/rscbjbr9sj
RSNA Pneumonia(Shih等,2019)	X光图像	1	26,684张	https://www.kaggle.com/c/rsna-pneumonia-detection-challenge
Camelyon16(Ehteshami Bejnordi等,2017)	病理图像	1	7,321张	https://camelyon16.grand-challenge.org/Data/

© 中国图象图形学报版权所有

表7　欺骗行为检测的代表性数据集

Table 7　Representative datasets for deception detection

数据集	数据形式	类别数	数据规模	数据集链接
Real-Life Trial(Pérez-Rosas等,2015)	RGB视频	2	121个片段	https://public.websites.umich.edu/~zmohamed/resources.html
Bag-of-Lies(Gupta等,2019)	RGB视频	2	325个片段	https://doi.org/10.1109/CVPRW.2019.00016
Box of Lies(Soldner等,2019)	RGB视频	2	25个片段	https://aclanthology.org/N19-1065/
MU3D(Lloyd等,2019)	RGB视频	2	320个片段	https://ngdc.cncb.ac.cn/databasecommons/database/id/6354
DOLOS(Guo等,2023)	RGB视频	2	1,675个片段	https://rose1.ntu.edu.sg/dataset/DOLOS/
MDPE(Cai等,2025)	RGB视频	2	104小时	https://huggingface.co/datasets/MDPEdataset/MDPE_Dataset
T4-Deception(Huang等,2026)	RGB视频	2	1,695个片段	https://arxiv.org/abs/2603.23916

表8　基于rPPG的非接触式生理信号测量的代表性数据集

Table 8　Representative datasets for rPPG-based contactless physiological signal measurement

数据集	数据形式	类别数	数据规模	数据集链接
PURE(Stricker等,2014)	RGB视频	–	60	https://www.tu-ilmenau.de/en/neurob/data-sets-code/pulse
COHFACE(Zhao等,2024)	RGB视频	–	160	https://www.idiap.ch/en/scientific-research/data/cohface
UBFC-rPPG(Hu等,2021)	RGB视频	–	42	https://sites.google.com/view/ybenezeth/ubfcrppg
VIPL-HR(Niu等,2018)	RGB/红外视频	–	3,130	https://vipl.ict.ac.cn/view_database.php?id=15
MR-NIRP(Nowara等,2020)	RGB/红外视频	–	–	https://computationalimaging.rice.edu/mr-nirp-dataset/
UBFC-Phys(Sabour等,2021)	RGB视频	–	168	https://sites.google.com/view/ybenezeth/ubfc-phys
MMPD(Tang等,2023)	RGB视频	–	660	https://github.com/ubicomplab/rPPG-Toolbox

表9　微小目标检测的代表性数据集

Table 9　Representative datasets for tiny object detection

数据集	数据形式	类别数	数据规模	数据集链接
AI-TOD(Wang等,2021)	RGB图像	8	28,036	https://github.com/jwwangchn/AI-TOD
AI-TOD-v2(Xu等,2022)	RGB图像	8	28,036	https://eis-vipg.github.io/AI-TOD-v2/
TinyPerson(Yu等,2020)	RGB图像	1	1,610	https://github.com/ucas-vg/TinyBenchmark
SODA-A(Duan等,2022)	RGB图像	9	2,513	https://shaunyuan22.github.io/SODA/
VisDrone(Zhu等,2018)	RGB图像	10	10,209	https://github.com/VisDrone/VisDrone-Dataset
DOTA-v2.0(Ding等,2021)	RGB图像	18	11,268	https://captain-whu.github.io/DOTA/

2017)组织为可比异常检测设置。不同数据分别覆盖微小肿瘤、弱对比病灶、视网膜异常和病理微转移等场景,使研究不仅关注病灶能否被观测和保持,也更加重视结合解剖位置、组织结构及跨模态信息降低异常判断的不确定性。

欺骗行为检测。欺骗数据集的发展突出体现了微弱证据由单一观测向关系推理的转变,代表性数据集见表7。Real-Life Trial(Pérez-Rosas等,2015)和Box of Lies(Soldner等,2019)主要关注自然交互中的细微行为变化,Bag-of-Lies(Gupta等,2019)进一步同步视频、语音、视线和脑电等信息,MU3D(Lloyd等,2019)通过同一参与者真假陈述减弱身份干扰。DOLOS(Guo等,2023)、MDPE(Cai等,2025)和T4-Deception(Huang等,2026)进一步扩展人物、情境、

© 中国图象图形学报版权所有

个体属性及文化背景，使问题由稀疏局部线索的发现逐步转向结合个体、语境和多模态关系缓解判别欠定性。

基于rPPG的非接触式生理信号测量。远程生理数据集主要用于检验低幅周期信号在复杂环境中的观测和保持能力，代表性数据集见表8。PURE（Stricker等，2014）、COHFACE（Zhao等，2024）和UBFC-rPPG（Hu等，2021）逐步引入头部运动、光照和自然生理波动，直接增加微弱信号的观测难度；VIPL-HR（Niu等，2018）和MR-NIRP（Nowara等，2020）进一步考虑多设备、近红外和驾驶环境，而UBFC-Phys（Sabour等，2021）与MMPD（Tang等，2023）扩展至压力、移动设备和肤色变化。由此，研究重点由受控条件下的周期信号获取逐步转向如何在运动、光照和设备扰动下保持真实生理成分，并避免将非生理变化误判为目标信号。

微小目标检测。微小目标数据集主要通过降低目标尺度、增加实例密度和复杂背景持续逼近视觉信息的观测与表征边界，代表性数据集见表9。AI-TOD（Wang等，2021）和AI-TOD-v2（Xu等，2022）聚焦航空影像中的极小目标及其标注质量，TinyPerson（Yu等，2020）强调远距离密集人体，SODA-A（Duan等，2022）和DOTA-v2.0（Ding等，2021）进一步增加高分辨率、任意方向和密集实例，VisDrone（Zhu等，2018）则引入动态平台、遮挡和复杂交通背景。相关基准的发展使问题由“目标是否能够被看到”逐步延伸至“有限目标证据能否在尺度变换和背景竞争中被稳定保持并正确分离”。

表10　伪装目标检测与分割的代表性数据集

Table 10　Representative datasets for camouflaged object detection and segmentation

数据集	数据形式	类别数	数据规模	数据集链接
CAMO(Le等，2019)	RGB图像	43	1,250张	https://arxiv.org/abs/1811.10200
COD10K-v2(Fan等，2020)	RGB图像	68	10,000张	https://arxiv.org/abs/2002.03779
NC4K(Lv等，2021)	RGB图像	37	4,121张	COD-Rank-Localize-and-Segment·GitHub
CHAMELEON(Skurowski等，2018)	RGB图像	–	76张	https://github.com/DengPingFan/SINet
GBU-UCOD(Wu等，2026)	水下RGB图像	–	–	https://github.com/Wuwenji18/GBU-UCOD
MoCA(Lamdouar等，2020)	RGB视频	67	141段	https://www.robots.ox.ac.uk/~vgg/data/MoCA/
MoCA-Mask(Cheng等，2022)	RGB视频	67	87段	https://xueliancheng.github.io/SLT-Net-project/
YUV20K(Liu等，2026)	RGB视频	47	91段	https://github.com/K1NSA/YUV20K
R2C7K(Zhang等，2025)	RGB图像	64	6,615张	https://github.com/zhangxuying1004/RefCOD

伪装目标检测与分割。伪装目标数据集同时体现了观测临界性、表征易失性与判别欠定性的耦合，代表性数据集见表10。CAMO（Le等，2019）、COD10K（Fan等，2020）、NC4K（Lv等，2021）和CHAMELEON（Skurowski等，2018）主要覆盖不同复杂度的静态伪装场景，GBU-UCOD（Wu等，2026）进一步强化水下低对比和复杂纹理条件下的观测与表征困难。MoCA（Lamdouar等，2020）、MoCA-Mask（Cheng等，2022）和YUV20K（Liu等，2026）将评价扩展至动态场景中的运动、时序对应和背景干扰，而R2C7K（Zhang等，2025）通过引入清晰参考图像为高度欠定的目标--背景关系提供外部判别依据。由此，相关基准逐步由静态低对比目标发现走向时空关系与外部知识协同下的可靠伪装理解。

4　代表性算法性能

在前述微视觉问题定义、典型应用、技术方法及代表性数据集分析的基础上，本节进一步对不同微视觉任务中的代表性算法进行性能比较。针对各任务的评价目标与数据特点，分别介绍主要评估指标、数据集及评估协议，并结合典型算法的核心技术机制分析其性能差异，重点讨论观测临界性、表征易失性和判别欠定性在实际评测中的具体表现与现有方

© 中国图象图形学报版权所有

法的局限。

4.1 工业图像异常检测与定位

评估指标。工业异常检测通常同时评价图像级异常识别和像素级异常定位性能。图像级受试者工作特征曲线下面积(Image-level Area Under the Receiver Operating Characteristic Curve，I-AUROC)衡量模型区分正常图像与异常图像的能力；像素级受试者工作特征曲线下面积(Pixel-level AUROC，P-AUROC)衡量异常像素与正常像素的区分能力；像素级区域重叠曲线下面积(Pixel-level AUPRO，P-AUPRO)则以异常区域为基本单位评价预测区域与真实缺陷区域的重叠程度，能够减弱不同缺陷面积带来的影响。上述指标均为数值越高表示性能越好。

数据集与评估协议。本节选取MVTec AD(Bergmann等，2019)、VisA(Zou等，2022)和Real-IAD(Wang等，2024)进行比较。MVTec AD覆盖多类工业物体与纹理，是工业异常检测中应用最广泛的标准基准；VisA包含更丰富的物体结构和细粒度缺陷；Real-IAD来源于真实工业生产环境，在成像条件、产品类型和异常形态方面更加复杂。为提高结果的可比性，表11采用统一多类别无监督异常检测(multi-class unsupervised anomaly detection，MUAD)设置，即同一数据集中的所有产品类别共同训练一个统一模型，而不是为每个类别分别训练独立模型。表中方法采用统一的任务定义、数据划分和指标实现，同时保留各方法自身的网络结构与推荐训练配置。

性能分析。表11显示，Dinomaly在三个数据集及全部指标上均取得最优结果。从技术机制来看，UniAD通过统一重建框架学习不同产品类别的正常特征分布，ReContrast利用特征对比与重建增强异常区分能力，而Dinomaly结合预训练视觉特征与特征重建机制，提高多类别场景下正常与异常模式的区分能力。MVTec AD上多数方法的I-AUROC和P-AUROC已超过95%，而在结构更复杂的VisA和真实生产环境Real-IAD上性能下降更明显，说明场景复杂化后局部异常的稳定表征与关系判别仍是主要瓶颈。Real-IAD上Dinomaly的I-AUROC和P-AUPRO分别达到89.3%和93.9%，但不同指标之间仍存在差异，因此评价时需同时关注图像级识别、像素级区分与区域级定位质量。

表11 统一多类别无监督异常检测设置下代表性方法在MVTec AD、VisA和Real-IAD上的性能比较

Table 11 Performance comparison of representative methods on MVTec AD, VisA, and Real-IAD under the unified multi-class unsupervised anomaly detection setting

方法	MVTec AD			VisA			Real-IAD		
	I-AUROC	P-AUROC	P-AUPRO	I-AUROC	P-AUROC	P-AUPRO	I-AUROC	P-AUROC	P-AUPRO
RD4AD(Deng和Li，2022)	94.6	96.1	91.1	92.4	98.1	91.8	82.4	97.3	89.6
SimpleNet(Liu等，2023)	95.3	96.9	86.5	87.2	96.8	81.4	57.2	75.7	39.0
UniAD(You等，2022)	96.5	96.8	90.7	88.8	98.3	85.5	83.0	97.3	86.7
ReContrast(Guo等，2023)	98.3	97.1	93.2	95.5	98.5	91.9	86.4	97.8	91.8
DiAD(He等，2024)	97.2	96.8	90.7	86.8	96.0	75.2	75.6	88.0	58.1
MambaAD(He等，2024)	98.6	97.7	93.1	94.3	98.5	91.0	86.3	98.5	90.5
Dinomaly(Guo等，2025)	99.6	98.4	94.8	98.7	98.7	94.5	89.3	98.8	93.9

4.2 微表情识别

评估指标。微表情数据通常存在样本数量有限和类别分布不均衡的问题，因此仅采用总体准确率可能高估模型对多数类别的识别能力。本节采用非加权F1分数(Unweighted F1-score，UF1)和非加权平均召回率(Unweighted Average Recall，UAR)评价模型性能。UF1首先分别计算各类别的F1分数，再对所有类别进行等权平均，综合衡量模型的分类精确率与召回率；UAR则对各类别召回率进行等权平均，能够更直观地反映模型对少数类别的识别能力。两项指标均为数值越高表示性能越好。

数据集与评估协议。本节选取CASME II(Yan

© 中国图象图形学报版权所有

等，2014）、SAMM（Davison 等，2018）和 SMIC-HS（Li 等，2013）三个经典自发微表情数据集。由于不同数据集采用的原始情绪类别体系并不完全一致，实验按照通用设置将样本映射为消极、积极和惊讶三类，并在每个数据集上分别采用留一被试交叉验证（Leave-One-Subject-Out，LOSO）。在每轮实验中，以一名被试的全部样本作为测试集，其余被试样本用于训练，从而避免同一被试同时出现在训练集和测试集中。表 12 汇总了相同三分类标签映射与 LOSO 协议下的代表性结果。

性能分析。表 12 显示，微表情识别已由人工时空纹理描述逐步转向局部运动与时空关系联合建模。LBP-TOP 主要依赖人工设计的时空纹理特征，SLSTT 和 HTNet 通过时空关系建模增强局部肌肉运动的表征能力，而 Ra-ICL 进一步利用多实例运动表示与区域一致性约束，挖掘分散于不同面部区域的细微动态线索。性能方面，Ra-ICL 在三个数据集上均取得最优结果，在 SAMM 上的 UF1 和 UAR 较 HTNet 分别提高 5.37 和 7.61 个百分点，表明对分散弱运动证据的有效建模有助于缓解短时、低幅信息的表征易失性。然而，不同数据集间仍存在明显性能波动，说明采集条件、被试差异和运动强度变化仍会影响微弱动态的稳定判别。

表 12　代表性微表情识别方法在 3 个数据集上的性能比较

Table 12　Performance comparison of representative micro-expression recognition methods on three datasets

方法	CASME II		SAMM		SMIC-HS	
	UF1	UAR	UF1	UAR	UF1	UAR
LBP-TOP(Zhao 等，2007)	70.26	74.29	39.54	41.02	20.00	52.80
Bi-WOOF(Liong 等，2018)	78.05	80.26	52.11	51.39	57.27	58.29
SLSTT(Zhang 等，2022)	90.10	88.50	71.50	64.30	74.00	72.00
HTNet(Wang 等，2024)	95.32	95.16	81.31	81.24	80.49	79.05
Ra-ICL(Cai 等，2026)	96.20	96.20	86.68	88.85	81.79	82.74

4.3　微姿态与微动作识别

评估指标。微姿态与微动作识别通常采用 Top-1 准确率评价模型的整体分类性能，即预测概率最高的类别与真实类别一致的样本比例。考虑到微动作数据通常具有较明显的长尾分布，本节进一步在 MA-52 上采用宏平均 F1 分数（Macro-F1），即分别计算各类别的 F1 分数并进行等权平均，以降低多数类别样本数量对整体结果的主导作用。Top-1 准确率和 Macro-F1 均为数值越高表示性能越好。

数据集与评估协议。本节选取 iMiGUE（Liu 等，2021）、SMG（Chen 等，2023）和 MA-52（Guo 等，2024）进行比较。iMiGUE 主要采集运动员新闻发布会中的非刻意上半身微姿态，强调身份无关的细粒度行为识别；SMG 通过访谈式任务采集自然产生的全身微姿态，包含更丰富的身体区域和动作变化；MA-52 则在心理访谈场景中采集坐姿微动作，并同时提供身体区域层面的粗粒度标签和动作类别层面的细粒度标签。MA-52 结果采用细粒度 52 类识别协议，报告 Top-1 准确率和 Macro-F1。

性能分析。表 13 显示，不同模型在 iMiGUE、SMG 和 MA-52 上的表现存在明显差异。Video-Mamba 和 MSF-Mamba-M+通过连续时序与多尺度运动建模提升了对局部弱动作的保持能力，Qwen3-VL-8B+GMoT 进一步体现出运动感知视觉 Token 与语义先验结合的潜力。MA-52 上 MANet 的 Top-1 和 Macro-F1 分别为 61.33% 和 49.22%，两者差距表明长尾类别及外观高度相似动作仍难稳定区分。整体上，该任务的瓶颈已由“能否捕获微弱动作”进一步转向长时序中的证据保持与细粒度判别。

4.4　人脸反欺诈

评估指标。人脸反欺诈需要同时控制将攻击样本错误接受为真实人脸和将真实人脸错误拒绝为攻击样本的风险。本节采用半总错误率（Half Total Error Rate，HTER）和受试者工作特征曲线下面积（Area Under the ROC Curve，AUC）评价模型性能。HTER 为错误接受率与错误拒绝率的平均值，数值越低表示模型的综合判别错误越少；AUC 衡量模型在不同判别阈值下区分真实人脸与呈现攻击的整体

© 中国图象图形学报版权所有

表13　代表性微姿态与微动作识别方法在3个数据集上的性能比较

Table 13　Performance comparison of representative micro-posture and micro-action recognition methods on iMiGUE, SMG, and MA-52

方法	iMiGUE	SMG	MA-52	
	Top-1	Top-1	Top-1	Macro-F1
TSM(Lin等,2019)	56.23	58.69	56.75	40.19
Video SwinT(Liu等,2022)	60.11	48.03	57.23	38.53
MANet(Guo等,2024)	58.08	48.69	61.33	49.22
VideoMamba(Li等,2024)	58.13	53.28	--	--
MSF-Mamba-M+(Li等,2026)	62.98	60.13	--	--
Qwen3-VL-8B+GMoT(Wang等,2026)	67.32	73.11	--	--

能力,数值越高表示性能越好。

数据集与评估协议。本节采用MSU-MFSD(Wen等,2015)、CASIA-FASD(Zhang等,2012)、Replay-Attack(Chingovska等,2012)和OULU-NPU(Boulkenafet等,2017)四个人脸反欺诈数据集进行跨数据集域泛化实验。四个数据集在采集设备、光照环境、被试构成以及攻击类型等方面存在明显的域差异,能够有效评估模型在未知目标域上的泛化能力。具体而言,实验采用多源域训练、单一未见目标域测试的设置,每次选取其中三个数据集作为源域进行训练,并将剩余一个数据集作为未见目标域进行测试,整个过程中不使用目标域数据参与模型训练。例如,OCI→M表示在OULU-NPU、CASIA-FASD和Replay-Attack上进行训练,并直接在未见的MSU-MFSD上进行测试。类似地,通过依次将M、C、I和O作为未见目标域,综合评估不同方法在跨数据集场景下的域泛化性能。实验采用HTER和AUC作为评价指标,其中较低的HTER和较高的AUC表示更优的跨域泛化性能。

性能分析。表14显示,跨数据集人脸反欺诈性能随着域泛化方法的发展持续提升:平均HTER由MADDG的23.09%降至SSDG-R的11.28%,GAC-FAS和DCRN进一步降至约6.5%。这表明设备、环境和攻击介质变化会使局部活体证据产生明显表征偏移,单纯依赖视觉域不变学习仍存在局限。CLIP-SA利用CLIP提供的语义先验增强活体与攻击特征的对齐与区分,DGPDL则通过域引导的提示分布学习,提高不同环境下语义表征的适应能力。两者的平均HTER分别降至3.75%和2.72%,说明语义先验能够为未知域中的活体与攻击判断提供更稳定的辅助约束。由此可见,人脸反欺诈的核心挑战已由局部伪造痕迹检测进一步转向跨域条件下微弱证据的稳定保持与可靠判别。

4.5　人脸Deepfake与通用AIGC检测

评估指标。人脸Deepfake检测采用受试者工作特征曲线下面积(Area Under the ROC Curve,AUC)评价模型区分真实帧与伪造帧的能力。表中报告帧级AUC,即将视频帧作为独立样本进行预测,数值越高表示检测性能越好。通用AIGC图像检测采用平均分类准确率(Accuracy,Acc),即模型正确区分真实图像与生成图像的样本比例,并对不同生成器上的结果取平均。由于两类任务的数据形式、生成方式与评价协议不同,两个面板中的数值不进行横向比较。

数据集与评估协议。人脸Deepfake检测选取FaceForensics++(Rossler等,2019)、Celeb-DF v2(Li等,2020)和DFDC(Dolhansky等,2020)。所有方法均在FaceForensics++的c23轻度压缩版本上训练,并在FaceForensics++上进行域内测试,同时在Celeb-DF v2和DFDC上直接进行跨数据集测试,不使用目标数据集进行微调。通用AIGC检测选取AIGCDetectBenchmark、GenImage(Zhu等,2023)和Chameleon(Yan等,2025)。在AIGCDetectBenchmark上,模型主要采用ProGAN生成图像训练,其中DIRE对应使用相同ProGAN数据训练的DIRE-G;在GenImage上,模型采用SD v1.4生成图像训练,并测试其他生成器;在Chameleon上,模型采用完整GenImage训练集训练,以评价对高真实感网络生成图像

© 中国图象图形学报版权所有

表14　在MSU-MFSD、CASIA-FASD、ReplayAttack和OULU-NPU数据集上的跨数据集域泛化实验结果

Table 14　Cross-dataset domain generalization results on MSU-MFSD, CASIA-FASD, Replay-Attack, and OULU-NPU

方法	OCI→M		OMI→C		OCM→I		ICM→O		avg.
	HTER	AUC	HTER	AUC	HTER	AUC	HTER	AUC	HTER
MADDG (Shao等,2019)	17.69	88.06	24.50	84.51	22.19	84.99	27.98	80.02	23.09
DR-MD-Net (Wang等,2020)	17.02	90.10	19.68	87.43	20.87	86.72	25.02	81.47	20.64
RFMeta (Shao等,2020)	13.89	93.98	20.27	88.16	17.30	90.48	16.45	91.16	16.97
NAS-FAS (Yu等,2020)	19.53	88.63	16.54	90.18	14.51	93.84	13.80	93.43	16.09
D2AM (Chen等,2021)	12.70	95.66	20.98	85.58	15.43	91.22	15.27	90.87	16.09
SDA (Wang等,2021)	15.40	91.80	24.50	84.40	15.60	90.10	23.10	84.30	19.65
ANRL (Liu等,2021)	10.83	96.75	17.83	89.26	16.03	91.04	15.67	91.90	15.09
SSDG-R (Jia等,2020)	7.38	97.17	10.44	95.94	11.71	96.59	15.61	91.54	11.28
SSAN-R (Wang等,2022)	6.67	98.75	10.00	96.67	8.88	96.79	13.72	93.63	9.81
PatchNet (Wang等,2022)	7.10	98.46	11.33	94.58	13.40	95.67	11.82	95.07	10.91
SA-FAS (Sun等,2023)	5.95	96.55	8.78	95.37	6.58	97.54	10.0	96.23	7.82
TTDG-V (Zhou等,2024)	4.16	98.48	7.59	98.18	9.62	98.18	10.0	96.15	7.84
GAC-FAS (Le和Woo,2024)	5.00	97.56	8.20	95.16	4.29	98.87	8.60	97.16	6.52
DiffFAS (Ge等,2024)	2.86	98.41	10.1	96.32	6.36	97.89	8.11	97.27	6.86
ViT-S-Adapter (Cai等,2024)	2.90	99.48	7.37	97.63	8.54	97.17	8.02	97.69	6.74
DCRN (Liu等,2025)	4.05	99.12	7.38	97.57	6.17	98.22	8.33	98.17	6.48
CA-MoEiT (Liu等,2024)	2.88	98.76	7.89	97.70	6.18	98.94	9.72	96.22	6.67
CLIP-SA (Ma等,2026)	3.87	99.30	2.47	99.20	4.24	98.87	4.42	99.17	3.75
DGPDL (Liu等,2026)	4.76	98.63	0.66	99.88	3.50	99.35	1.97	99.84	2.72

的泛化能力。

性能分析。表15表明,人脸Deepfake方法在FaceForensics++域内的AUC均超过96%,但迁移至Celeb-DF v2和DFDC后明显下降,反映出局部纹理、频率和伪影线索容易随压缩方式、数据来源和生成算法变化而失效。SELFI通过引入身份相关辅助信息改善跨数据集表现,在Celeb-DF v2和DFDC上分别较次优方法提高7.4和7.7个百分点。通用AIGC检测呈现相同趋势。AIDE在AIGCDetectBenchmark、GenImage和Chameleon上均取得最优结果,但在高真实性Chameleon上的准确率仍仅为63.89%。这说明当显式生成伪影减弱后,仅依赖局部异常已不足以稳定判别,模型仍需进一步结合跨区域关系、全局语义及对未知生成器和后处理扰动的鲁棒表征。

4.6　医学图像异常检测

评估指标。医学图像异常检测旨在判断输入图像是否包含病灶或其他异常结构。本节采用图像级受试者工作特征曲线下面积(Image-level Area Under the ROC Curve,I-AUROC)评价模型性能。该指标通过计算不同判别阈值下真正例率与假正例率形成的ROC曲线面积,衡量模型区分正常样本与异常样本的整体能力。I-AUROC不依赖单一分类阈值,数值越高表示模型的异常识别能力越强。表中结果为BMAD基准在三个不同随机种子下重复实验所得的平均值。

数据集与评估协议。本节采用BMAD统一医学异常检测基准中的BraTS 2021、BTCV+LiTS、RESC和OCT2017。BraTS 2021由脑部MRI体数据的二维轴向切片构成,将含有脑肿瘤的切片视为异常样本;BTCV+LiTS使用BTCV中的无病灶肝脏CT切片进

© 中国图象图形学报版权所有

表15　人脸Deepfake与通用AIGC检测方法的性能比较

Table 15　Performance comparison of representative methods for face Deepfake and General AIGC detection

方法	FaceForensics++	Celeb-DF v2	DFDC
(a) 人脸Deepfake检测：帧级AUC			
Xception (Chollet，2017)	96.4	73.7	70.8
F3-Net (Qian等，2020)	96.4	73.5	70.2
SPSL (Liu等，2021)	96.1	76.5	70.4
RECCE (Cao等，2022)	96.2	73.2	71.3
UCF (Yan等，2023)	97.1	75.3	71.9
SELFI (Kim等，2025)	98.0	83.9	79.6
方法	**AIGCDetectBenchmark**	**GenImage**	**Chameleon**
(b) 通用AIGC图像检测：平均准确率			
UnivFD (Ojha等，2023)	78.43	73.29	60.42
DIRE (Wang等，2023)	68.68	70.66	57.83
NPR (Tan等，2024)	82.91	--	57.81
AIDE (Yan等，2025)	92.77	86.88	63.89

行训练，并采用LiTS中的正常与肿瘤切片进行测试，因而同时包含数据来源和病理状态差异；RESC主要包含视网膜水肿异常并提供像素级标注；OCT2017包含正常视网膜以及CNV、DME和DRUSEN三类异常。本节仅比较四个数据集共同支持的图像级异常检测结果。BMAD重新组织上述医学数据集，并规定相互独立的训练集、验证集和测试集。所有方法仅使用正常样本进行训练，通过验证集选择模型或调整训练配置，最终在测试集上进行评价。BMAD优先采用各方法原始论文推荐的模型结构和超参数，并使用统一的数据划分、模型选择流程及评价指标进行复现。

性能分析。表16显示，不同方法在四个医学数据集上的相对表现差异明显。PatchCore在BraTS 2021、RESC和OCT2017上取得最优结果，但在BTCV+LiTS上的AUROC仅为60.28%；SimpleNet在该数据集上达到72.28%，仍明显低于其他场景。跨来源肝脏CT中的设备、灰度和解剖差异表明，病灶证据不仅需要被有效观测和保持，还需避免将成像域差异误判为异常。总体而言，现有方法尚难在不同模态和器官上保持一致优势。后续评价应进一步关注跨模态泛化、微小病灶敏感性以及图像级异常判断与病灶级定位的一致性，以同时检验微弱证据的稳定表征与可靠判别。

表16　代表性医学图像异常检测方法的图像级比较

Table 16　Image-level comparison of representative medical image anomaly detection methods

方法	BraTS 2021	BTCV +LiTS	RESC	OCT2017
DRAEM (Zavrtanik等，2021)	62.35	69.95	83.22	88.03
UTRAD (Chen等，2022)	82.92	55.81	89.39	96.78
SimpleNet (Liu等，2023)	82.52	72.28	76.15	94.68
RD4AD (Deng和Li，2022)	89.45	60.38	87.77	97.30
PatchCore (Roth等，2022)	91.65	60.28	91.55	98.57
CFA (Lee等，2022)	84.38	62.00	69.90	79.47

4.7　欺骗行为检测

评估指标。欺骗行为检测通常被建模为真实与欺骗之间的二分类任务。本节采用准确率(Accuracy，Acc)和F1分数评价模型性能。准确率表示被正确分类的样本占全部测试样本的比例；F1分数为精确率与召回率的调和平均值，能够综合反映模型识别欺骗样本时的查准率和查全率。两项指标均为数值越高表示性能越好。由于原始基准未将表中F1明确标注为宏平均或微平均，本文沿用原论文的指标名称，不额外改变其定义。

数据集与评估协议。本节选取DOLOS(Guo等，2023)、Bag-of-Lies(Gupta等，2019)和MU3D(Lloyd

© 中国图象图形学报版权所有

等,2019)进行比较。DOLOS包含游戏节目中的故事虚构行为,场景和人物表现相对自然;Bag-of-Lies通过实验室图像描述任务诱发真实与欺骗陈述;MU3D则记录被试围绕社会关系进行真实或虚假评价的视频。三个数据集分别包含1,675、325和320个样本。实验采用域内被试独立交叉验证,其中DOLOS和Bag-of-Lies使用三折协议,MU3D使用四折协议。表17仅保留音视频输入方法,并按照完整监督训练和LoRA微调分别比较。

性能分析。表17显示,完整监督方法具有明显的数据集依赖性,而DecepGPT在统一音视频输入和LoRA微调设置下取得最佳结果,相较VideoLLaMA2在DOLOS、Bag-of-Lies和MU3D上的准确率分别提高19.75、15.81和9.41个百分点,说明多模态融合与结构化推理有助于整合分散的微弱行为线索。但即使最优方法在Bag-of-Lies和MU3D上的准确率也仅为63.46%和61.25%。这说明欺骗行为的局部视觉或语音线索本身具有较强欠定性,其含义高度依赖个体、语境与多模态关系,跨被试、跨数据集和跨文化泛化仍是关键挑战。

表17　统一域内交叉验证协议下音视频欺骗行为检测方法的性能比较

Table 17　Performance comparison of audio-visual Deception detection methods under a unified in-domain cross-validation protocol

方法	DOLOS		Bag-of-Lies		MU3D	
	Acc	**F1**	**Acc**	**F1**	**Acc**	**F1**
(a) 完整监督训练						
LieNet(Karnati等,2022)	56.50	69.72	59.78	58.14	53.48	33.62
PECL(Guo等,2023)	64.75	71.20	59.51	51.06	55.31	60.07
(b) LoRA参数高效微调						
VideoLLaMA2(Cheng等,2024)	53.48	56.12	47.65	49.34	51.84	54.62
Video-SALMONN2(Tang等,2025)	52.63	55.44	46.82	48.51	51.06	53.79
Qwen2.5-7B-VL(Bai等,2025)	46.10	52.37	44.24	41.15	53.86	56.24
DecepGPT(Huang等,2026)	73.23	76.13	63.46	63.72	61.25	67.22

4.8　远程生理信号测量

评估指标。基于rPPG的非接触式生理信号测量通过分析摄像头采集的面部视频中由血液容积变化引起的微弱肤色波动,提取rPPG信号并估计心率。本节采用平均绝对误差(Mean Absolute Error, MAE)评价心率估计性能,其计算预测心率与真实心率绝对误差的平均值,单位为每分钟心跳次数(beats per minute,bpm)。MAE越低,表示预测心率与接触式传感器测得的参考心率越接近。

数据集与评估协议。本节选取PURE(Stricker等,2014)、UBFC-rPPG(Hu等,2021)和MMPD(Tang等,2023)进行跨数据集评价。PURE包含静止、说话和不同幅度头部运动等受控场景;UBFC-rPPG主要在相对稳定的室内环境下采集;MMPD使用智能手机采集,包含不同肤色、光照和运动条件,场景复杂度更高。对于学习型方法,模型仅在箭头左侧的源数据集上训练,并直接在右侧未见目标数据集上测试,不使用目标域样本进行微调。所有数据划分均以被试身份为单位,以避免训练集与测试集之间的身份泄漏。CHROM为无需数据训练的传统信号处理方法,因此箭头所表示的源训练集不适用于该方法;其在相同目标数据集上的结果保持一致。

性能分析。表18显示,现有方法在PURE与UBFC-rPPG之间迁移时误差较低,但迁移方向不同仍会产生明显差异;当目标域变为包含复杂光照、运动和肤色变化的MMPD时,所有方法误差均显著增大,说明微弱脉搏信号在真实环境中仍面临突出的观测临界性和表征不稳定问题。从技术机制来看,传统CHROM方法主要通过颜色空间变换提取脉搏相关变化,深度学习方法则利用时空特征建模增强动态信号表征,而FreqPhys进一步引入生理频带先验与主频结构约束,提高有效周期成分的提取能力并抑制非生理扰动。FreqPhys在四种迁移设置下均取得最优结果,体现了显式生理先验对跨域稳定测

© 中国图象图形学报版权所有

量的积极作用。然而，其在MMPD上的MAE仍达到8.91--10.11 bpm，表明复杂环境下真实生理成分与非生理扰动的可靠分离仍有较大提升空间。

表18　跨数据集远程心率估计方法的MAE比较

Table 18　MAE comparison of cross-dataset remote heart rate estimation methods

方法	PURE→UBFC-rPPG	UBFC-rPPG→PURE	PURE→MMPD	UBFC-rPPG→MMPD
CHROM(De Haan和Jeanne，2013)	4.06	5.77	13.66	13.66
DeepPhys(Chen和McDuff，2018)	1.21	5.54	16.92	17.50
PhysNet(Yu等，2019)	1.63	9.36	13.22	10.24
PhysFormer(Yu等，2022)	1.44	12.92	14.57	12.10
RhythmMamba(Zou等，2024)	0.95	1.98	10.44	10.63
FreqPhys(Qian等，2026)	0.43	0.95	10.11	8.91

4.9　微小目标检测

评估指标。微小目标检测通常采用COCO风格的平均精度（Average Precision，AP）评价检测性能。AP是在交并比阈值从0.50到0.95、间隔为0.05的多个条件下计算检测精度并取平均，综合反映目标分类与定位能力；AP50和AP75分别表示交并比阈值为0.50和0.75时的平均精度。AP50对目标框定位误差相对宽容，而AP75要求预测框与真实框具有更高重叠度。三项指标均为数值越高表示性能越好。

数据集与评估协议。本节选取AI-TOD-v2（Xu等，2022）和VisDrone（Zhu等，2018）进行比较。AI-TOD-v2针对原始AI-TOD中存在的漏标和位置偏差进行了重新标注，目标通常仅占据少量像素，且不同图像中的目标数量差异显著；VisDrone由无人机在不同城市、天气、光照和飞行高度下采集，包含行人、车辆和自行车等密集小目标。AI-TOD-v2上的方法均在trainval划分上训练并在test划分上评价，VisDrone上的方法均在train划分上训练并在val划分上评价。

性能分析。表19显示，通用Faster R-CNN在两个数据集上的性能明显较低，而NWD-RKA、RFLA和DQ-DETR通过改进相似度度量、标签分配、多尺度特征和动态查询显著提升微小目标检测性能。DQ-DETR在AI-TOD-v2和VisDrone上的AP分别达到30.2%和37.0%，说明针对有限像素支撑进行特征保持和密度适配能够有效改善检测。但AI-TOD-v2上各方法的AP75仍显著低于AP50，例如DQ-DETR分别为22.3%和68.6%。这表明现有方法虽能发现更多微小目标，但有限像素造成的观测临界性和边界信息易失仍直接制约精确定位，密集实例之间的稳定区分也仍需进一步提升。

表19　代表性微小目标检测方法在两个数据集上的性能比较

Table 19　Performance comparison of representative tiny object detection methods on two datasets

方法	AI-TOD-v2			VisDrone		
	AP	**AP50**	**AP75**	**AP**	**AP50**	**AP75**
Faster R-CNN(Ren等，2015)	12.8	29.9	9.4	21.4	40.7	19.9
NWD-RKA(Xu等，2022)	24.7	57.4	17.1	--	--	--
RFLA(Xu等，2022)	25.7	58.9	18.8	--	--	--
DNTR(Liu等，2024)	--	--	--	34.4	57.9	35.3
DINO-DETR(Zhang等，2022)	25.9	61.3	17.5	35.8	58.3	36.8
DQ-DETR(Huang等，2024)	30.2	68.6	22.3	37.0	60.9	37.9

© 中国图象图形学报版权所有

4.10 伪装目标检测与分割

评估指标。伪装目标检测通常被建模为前景与背景的二值分割任务。本节采用结构度量(Structure-measure,S_α)和平均绝对误差(Mean Absolute Error,MAE)评价模型性能。S_α从目标级和区域级两个方面衡量预测掩膜与真实掩膜之间的结构相似性,数值越高表示目标整体结构和区域分布越准确;MAE计算预测图与真实掩膜之间逐像素绝对误差的平均值,数值越低表示预测结果中的漏检和误检越少。两项指标分别反映结构一致性和像素级误差,具有较好的互补性。

数据集与评估协议。本节选取CAMO(Le等,2019)、COD10K(Fan等,2020)和NC4K(Lv等,2021)进行比较。按照标准静态伪装目标检测协议,模型采用CAMO训练集中的1,000张图像和COD10K训练集中的3,040张图像进行训练,并分别在CAMO测试集、COD10K测试集和NC4K上进行评价。三个测试集分别包含250、2,026和4,121张图像。表中方法采用相同的数据划分和评价指标,但保留各自的骨干网络及训练配置。

性能分析。表20显示,多尺度特征融合与上下文关系建模能够有效提升伪装目标的结构完整性与定位精度。从技术机制来看,ZoomNet利用不同尺度的互补信息增强伪装目标感知,FSPNet通过特征收缩金字塔与Transformer关系建模逐步积累目标线索,而CamoFormer-P进一步利用前景、背景及全局关系建模,提高高度相似目标与复杂背景之间的区分能力。性能方面,CamoFormer-P在三个数据集的全部指标上均取得最优结果,表明局部细节保持与显式关系建模的协同作用有助于缓解伪装目标的判别欠定性。然而,不同数据集之间仍存在明显性能差异,说明复杂背景、尺度变化和目标--背景高相似性仍会共同影响微弱边界的稳定表征与可靠判别。

表20 标准静态伪装目标检测协议下代表性方法在三个测试集上的性能比较

Table 20 Performance comparison of representative methods on three test sets under the standard static camouflaged object detection protocol

方法	CAMO		COD10K		NC4K	
	S_α	MAE	S_α	MAE	S_α	MAE
SINet-v2(Fan等,2021)	0.820	0.070	0.815	0.037	0.847	0.048
ZoomNet(Pang等,2022)	0.820	0.066	0.838	0.029	0.853	0.043
FSPNet(Huang等,2023)	0.857	0.050	0.851	0.026	0.879	0.035
Camouflageator(He等,2024)	0.829	0.066	0.843	0.028	0.869	0.041
CamoFormer-P(Yin等,2024)	0.872	0.046	0.869	0.023	0.892	0.030

总体来看,不同任务中的性能变化共同表明,微视觉方法的进步并非只体现在单一任务精度提升,而在于能否在复杂环境中同时保证关键证据可获取、可保持并可判别。现有方法在受控或同域条件下已取得较高性能,但跨数据集、复杂背景和开放环境下的明显退化说明,观测临界性、表征易失性和判别欠定性仍常以耦合形式共同限制模型可靠性。

5 展 望

微视觉计算的核心目标,是从处于观测边界、易受干扰的视觉信息中发现具有任务价值的微视觉证据。随着视觉基础模型、生成式人工智能和多模态学习的发展,微视觉计算正在由面向特定任务的局部特征提取,进一步走向统一表征、生成与理解协同以及开放场景下的潜在状态推理。未来研究可重点围绕以下几个方向展开。

微视觉基础模型与统一表征学习。从方法模型来看,现有微视觉任务大多仍采用针对特定任务独立设计模型的研究范式,例如微表情识别关注短时面部运动,伪装目标检测强调弱边界与前景--背景差异,生成内容检测关注局部伪影和频率异常,不同任务之间的知识共享仍然有限。然而,这些任务所依赖的信息在本质上具有共同特征,即需要从高维视觉信息中发现低显著、局部化且容易被主体语义掩盖的判别证据。现有视觉基础模型主要依赖大规

© 中国图象图形学报版权所有

模自然图像和显著语义信息进行预训练，其空间下采样、Token压缩和时序采样过程可能导致微小区域、弱纹理以及瞬时动态信息的损失。因此，未来有必要构建面向微视觉信息的基础模型，通过高分辨率空间建模、长时序动态建模以及跨尺度特征交互（杜仕荣等，2024），统一学习空间、时间和频率域中的细粒度表征。在此基础上，可进一步结合视觉、语言、生理信号和环境信息，使模型由单一的“微弱特征检测器”发展为能够完成微弱证据发现、定位、关联、推理和解释的通用微视觉模型，并借鉴已有医学视觉基础模型在跨任务知识迁移方面的研究进展（Zhou等，2023），探索不同微视觉任务之间知识迁移与能力涌现的可能性。

从核心瓶颈来看，高分辨率、长时序与跨尺度统一建模主要针对表征易失性，旨在减少微弱证据在空间压缩、时序采样和语义抽象过程中的损失；多模态基础模型则侧重通过异构信息互补与跨模态语义关联缓解判别欠定性，使单一视觉证据不足时仍能够形成更加可靠的判断依据。

微视觉生成与理解一体化。从信号增强和微视觉数据生成来看，微视觉计算长期受到信号微弱、样本稀缺和高质量标注成本高等问题制约。传统方法主要通过图像增强、数据扩增、图像融合或特征增强提高微弱信息的可辨识性（唐霖峰等，2023），但增强与理解通常作为相互独立的过程，容易产生与真实任务无关甚至破坏原始微弱证据的视觉变化。生成式模型的发展为这一问题提供了新的研究范式，其中，面向微弱视觉信号的放大与显化本身也可视为生成式应用的重要分支，即通过生成模型对难以直接观测的微小运动、细微纹理变化、局部形变及生理波动等信息进行恢复、重建与可控放大，使潜在的微弱证据转化为更易感知和分析的视觉表征。未来可进一步探索微视觉生成与理解的一体化学习，使生成模型不仅用于扩充样本数量，还能够在任务知识约束下主动发现、恢复、放大和生成具有判别价值的微弱视觉证据。例如，可利用扩散模型构造具有可控强度、位置、持续时间和语义属性的微表情、微动作、微小缺陷、局部伪影及生理变化，并通过生成过程中的多尺度中间表征辅助下游理解。进一步地，理解模型也可以反向约束生成过程，使生成内容保持物理合理性、时序一致性和任务相关性，形成“微弱信息发现--恢复与放大--受控生成--判别学习--反馈优化”的闭环。由此，数据生成将由传统的样本扩充工具逐步转变为探索微视觉信息空间和提升模型认知能力的重要手段。

从核心瓶颈来看，生成与理解一体化主要面向观测临界性，通过任务约束下的信号恢复、细节重建和可控放大，提高低可观测证据的可辨识性。同时，该方向也需要关注表征易失性与生成真实性问题，避免在增强和重建过程中丢失关键细节或引入缺乏真实观测依据的虚假证据。

微视觉应用由显式目标识别走向潜在状态感知。从广泛应用来看，微视觉计算未来的重要发展趋势之一，是由“识别可见对象”进一步走向“利用微弱视觉证据推断不可直接观察的潜在状态”。在人因计算方面，可通过长期积累的面部微动作、注视、眨眼、姿态和行为节律，研究表观性格、认知负荷、疲劳、压力以及共情状态，实现更加细粒度的人机交互与情感智能。在健康领域，除已有的远程心率、呼吸和皮肤状态估计外，还可进一步探索从面部颜色、纹理、局部脂肪沉积、软组织形态及其空间分布等微弱表型中发现代谢异常和潜在疾病风险，使视觉感知由外观分析拓展至非接触式数字健康筛查。例如，脂肪性肝病等代谢相关疾病虽然主要发生于内部器官，但其发生发展往往伴随脂质代谢、糖代谢、氨基酸代谢及能量代谢状态的系统性改变，这些变化可能通过面部脂肪分布、肤色、皮肤纹理和局部组织形态等外在表型产生间接映射。因此，基于面部影像建立疾病风险分数，不应仅关注预测准确率，还可进一步结合血浆代谢组学等分子层信息，为视觉模型捕获的风险信号寻找“分子锚点”。通过分析面部风险分数与代谢物丰度之间的连续关联，并考察相关代谢物在糖脂代谢、氨基酸代谢和能量代谢通路中的富集情况，可以验证模型所利用的面部微弱表型是否对应真实的系统代谢扰动，从而为“由面识别潜在脂肪肝风险”等任务提供超越统计相关性的生物学解释。进一步地，可将面部表型、临床指标、代谢组、蛋白组乃至基因组信息进行跨尺度关联，建立“视觉微表型—分子表型—疾病状态”的多层级映射关系，使微视觉计算由单纯的疾病风险预测进一步发展为具有生物学依据的数字表型分析。类似思想还可以扩展至皮肤屏障状态、疼痛评估、神经行为异常以及长期健康变化监测。在教育与科研场景中，微视觉计算可结合图像局部异常、文档编辑痕迹、实

© 中国图象图形学报版权所有

验图像重复与篡改模式以及生成式内容特征，辅助开展学术不端与科研图像完整性检测。在具身智能中，机器人还需要感知人的微姿态、视线变化、接触形变、物体微小位移和操作过程中的细微状态变化，从而由显式指令理解进一步发展为对人、物体与环境隐性状态的连续感知。此外，该技术还可拓展至植物胁迫、食品新鲜度、材料疲劳、文物劣化和动物健康等场景，通过长期微变化分析发现显著异常出现之前的早期征兆。

从核心瓶颈来看，潜在状态感知主要面向判别欠定性，即单一或局部微弱视觉证据往往无法直接确定其所对应的生理、行为或环境状态。未来需要通过时序信息积累、多模态证据关联及领域先验约束，建立视觉微表型与潜在状态之间更加可靠的映射关系，并通过独立观测与外部证据验证其有效性。

开放环境下的可信微视觉计算。随着微视觉计算由目标识别进一步拓展至性格、心理、生理和疾病风险等潜在属性推断，模型输出与个体权益之间的关系将更加紧密，也对算法可信性提出了更高要求。微弱视觉证据本身具有较高的不确定性，其与最终状态之间通常并非简单的一一对应关系，因此模型需要避免将相关性直接解释为因果关系。未来应重点研究跨设备、跨场景、跨人群和跨时间条件下的鲁棒微视觉学习，并建立包含域外检测、不确定性估计、置信度校准和证据可解释性的评价体系。特别是在心理状态、人格属性和健康风险等应用中，模型应能够回答“依据了哪些微弱证据”“这些证据在何时何地出现”以及“当前证据是否足以支持判断”，并在证据不足时主动降低预测置信度。同时，还需要关注身份、年龄、性别、肤色和文化背景等因素可能形成的捷径相关性以及长期视觉数据采集带来的隐私问题。只有实现由高性能识别向可解释、可校准和可审计决策的转变，微视觉计算才能真正从实验室任务走向开放环境中的长期应用。从核心瓶颈来看，开放环境下的可信微视觉计算重点关注分布变化对微弱证据稳定性与判别可靠性的影响。一方面，成像条件、设备和环境变化可能进一步加剧表征易失性；另一方面，未知场景与个体差异可能增加证据解释的歧义性，使判别欠定性更加突出。因此，持续适应、不确定性估计与可信决策应协同发展，以提高微视觉模型在开放环境中的长期可靠性。

6 结 语

本文围绕微视觉计算这一新兴研究方向，对其基本概念、核心困难、技术体系、典型任务、数据集、代表性算法及未来趋势进行了系统梳理。本文从微弱性、稀疏性与隐现性三类外在表现出发，将微视觉计算进一步归结为观测临界性、表征易失性和判别欠定性三个相互关联的核心困难，并以此贯通共性技术、典型应用、数据集演进与算法性能分析。相关任务虽然具有不同的对象和目标，但均可从关键证据能否被可靠获取、有效保持和稳定判别三个层面进行理解。由此，微视觉计算并非既有应用方向的简单集合，而是围绕低可观测、高任务价值视觉证据形成的一类共性视觉计算问题。最后探讨了未来发展，强调高分辨率、长时序、多模态微视觉基础模型的趋势。

参考文献（References）

Wen Lang, Gou Guanglei, Bai Ruifeng, Miao Wanyu. 2025. Bifurcated attention and feature interaction for few-shot fine-grained learning. Journal of Image and Graphics, 30(5): 1419-1432DOI: 10.11834/jig.240429. (文浪, 苟光磊, 白瑞峰, 缪宛谕. 2025. 双分支注意和特征交互的小样本细粒度学习. 中国图象图形学报, 30(5): 1419-1432)DOI: 10.11834/jig.240429.

Wang Xiaoyan, Liang Wenhui, Li Jie, Mu Jianhong, Wang Xiyu. 2025. Fusion of bounding box Gaussian modeling and feature aggregation distribution for fine-grained recognition of remote sensing aircraft images. Journal of Image and Graphics, 30(01): 0282-0296DOI: 10.11834/jig.230862. (王晓燕, 梁文辉, 李杰, 牟建宏, 王禧钰. 2025. 融合边界框高斯建模与特征聚合分发的遥感飞机细粒度识别. 中国图象图形学报, 30(01): 0282-0296)DOI: 10.11834/jig.230862.

Bergmann P, Fauser M, Sattlegger D and Steger C. 2019. MVTec AD—A comprehensive real-world dataset for unsupervised anomaly detection//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR): 9584-9592

Bergmann P, Batzner K, Fauser M, Sattlegger D and Steger C. 2022. Beyond dents and scratches: Logical constraints in unsupervised anomaly detection and localization. International Journal of Computer Vision, 130(4): 947-969

Fan D P, Ji G P, Sun G, Cheng M M, Shen J and Shao L. 2020. Camouflaged object detection//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR): 2774-2784

© 中国图象图形学报版权所有

Zhao P, Sun Q, Tian X, Yang Y, Tao S, Cheng J, et al. 2024. Toward motion robustness: a masked attention regularization framework in remote photoplethysmography//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW): 7829-7838

Hu C, Zhang K Y, Yao T, Ding S, Li J, Huang F, et al. 2021. An end-to-end efficient framework for remote physiological signal sensing//2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW): 2378-2384

Li X, Pfister T, Huang X, Zhao G and Pietikäinen M. 2013. A spontaneous micro-expression database: Inducement, collection and baseline//2013 10th IEEE International Conference and Workshops on Automatic face and gesture recognition (fg): 1-6

Ge X, Liu X, Yu Z, Shi J, Qi C, Li J, et al. 2024. Difffas: face anti-spoofing via generative diffusion models//European Conference on Computer Vision: 144-161

Wang Z, Bao J, Zhou W, Wang W, Hu H, Chen H, et al. 2023. Dire for diffusion-generated image detection//2023 IEEE/CVF International Conference on Computer Vision (ICCV): 22388-22398

Zou Y, Jeong J, Pemula L, Zhang D and Dabeer O. 2022. SPot-the-Difference Self-Supervised Pre-training for Anomaly Detection and Segmentation//European Conference on Computer Vision: 392-408

Wang J, Yang W, Guo H, Zhang R and Xia G S. 2021. Tiny object detection in aerial images//2020 25th international conference on pattern recognition (ICPR): 3791-3798

Zhu P, Wen L, Bian X, Ling H and Hu Q. 2018. Vision meets drones: A challenge//arXiv preprint arXiv:1804.07437

Bao J, Sun H, Deng H, He Y, Zhang Z and Li X. 2024. Bmad: Benchmarks for medical anomaly detection//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW): 4042-4053

Zhu W, Wang L, Zhou Z, Wang C, Pan Y, Zhang R, et al. 2025. Real-IAD D3: A Real-World 2D/Pseudo-3D/3D Dataset for Industrial Anomaly Detection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition: 15214-15223

Yan W J, Wu Q, Liu Y J, Wang S J and Fu X. 2013. CASME Database: A Dataset of Spontaneous Micro-Expressions Collected From Neutralized Faces//2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG): 1-7

Yan W J, Li X, Wang S J, Zhao G, Liu Y J, Chen Y H, et al. 2014. CASME II: An improved spontaneous micro-expression database and the baseline evaluation. PloS one, 9(1): e86041

Davison A K, Lansley C, Costen N, Tan K and Yap M H. 2018. SAMM: A Spontaneous Micro-Facial Movement Dataset. IEEE Transactions on Affective Computing, 9(1): 116-129

Ben X, Ren Y, Zhang J, Wang S J, Kpalma K, Meng W, et al. 2021. Video-based facial micro-expression analysis: A survey of datasets, features and algorithms. IEEE transactions on pattern analysis and machine intelligence, 44(9): 5826-5846

Li J, Dong Z, Lu S, Wang S J, Yan W J, Ma Y, et al. 2022. CAS (ME) 3: A third generation facial spontaneous micro-expression database with depth information and high ecological validity. IEEE transactions on pattern analysis and machine intelligence, 45(3): 2782-2800

Zhao S, Tang H, Mao X, Liu S, Zhang Y, Wang H, et al. 2023. Dfme: A new benchmark for dynamic facial micro-expression recognition. IEEE Transactions on Affective Computing, 15(3): 1371-1386

Sribabu S, Raj V, KV G and Muniyandi M. 2026. BlendME: A VR Blendshape Microexpression Dataset with Retrospective Self-Annotation//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops: 1392-1400

De Haan G and Jeanne V. 2013. Robust pulse rate from chrominance-based rPPG. IEEE transactions on biomedical engineering, 60 (10): 2878-2886

Chen W and McDuff D. 2018. Deepphys: Video-based physiological measurement using convolutional attention networks//european conference on Computer Vision: 356-373

Yu Z, Li X and Zhao G. 2019. Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks//arXiv preprint arXiv:1905.02419

Yu Z, Shen Y, Shi J, Zhao H, Torr P H S and Zhao G. 2022. Physformer: Facial video-based physiological measurement with temporal difference transformer//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition: 4186-4196

Zou B, Guo Z, Hu X and Ma H. 2024. Rhythmmamba: Fast remote physiological measurement with arbitrary length videos//arXiv preprint arXiv:2404.06483. 7

Qian W, Guo D, Zhou J, Zou B, Yu Z and Wang M. 2026. Freqphys: Repurposing implicit physiological frequency prior for robust remote photoplethysmography//arXiv preprint arXiv:2604.00534

Qian Y, Yin G, Sheng L, Chen Z and Shao J. 2020. Thinking in frequency: Face forgery detection by mining frequency-aware clues//European conference on computer vision: 86-103

Liu H, Li X, Zhou W, Chen Y, He Y, Xue H, et al. 2021. Spatial-phase shallow learning: rethinking face forgery detection in frequency domain//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR): 772-781

Ojha U, Li Y and Lee Y J. 2023. Towards universal fake image detectors that generalize across generative models//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR): 24480-24489

Zhu M, Chen H, Yan Q, Huang X, Lin G, Li W, et al. 2023. Genimage: A million-scale benchmark for detecting ai-generated image//Advances in neural information processing systems. 36: 77771-77782

Yan S, Li O, Cai J, Hao Y, Jiang X, Hu Y, et al. 2025. A sanity check for ai-generated image detection//International Conference on Learning Representations. 2025: 70702-70720

© 中国图象图形学报版权所有

Yu Z, Wan J, Qin Y, Li X, Li S Z and Zhao G. 2020. NAS-FAS: Static-Dynamic Central Difference Network Search for Face Anti-Spoofing//TPAMI

Guo X, Selvaraj N M, Yu Z, Kong A W K, Shen B and Kot A. 2023. Audio-visual deception detection: Dolos dataset and parameter-efficient crossmodal learning//2023 IEEE/CVF International Conference on Computer Vision (ICCV): 22078-22088

Yin B, Zhang X, Fan D P, Jiao S, Cheng M M, Van Gool L, et al. 2024. Camoformer: Masked separable attention for camouflaged object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(12): 10362-10374

Roth K, Pemula L, Zepeda J, Schölkopf B, Brox T and Gehler P. 2022. Towards total recall in industrial anomaly detection//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR): 14298-14308

Deng H and Li X. 2022. Anomaly detection via reverse distillation from one-class embedding//2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR): 9727-9736

Wang Z, Zhang K, Luo W and Sankaranarayana R. 2024. Htnet for micro-expression recognition. Neurocomputing, 602: 128196

Lin J, Gan C and Han S. 2019. Tsm: Temporal shift module for efficient video understanding//Proceedings of the IEEE/CVF international conference on computer vision: 7083-7093

Liu Z, Ning J, Cao Y, Wei Y, Zhang Z, Lin S, et al. 2022. Video swin transformer//2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR): 3192-3201

Xu C, Wang J, Yang W, Yu H, Yu L and Xia G S. 2022. RFLA: Gaussian receptive field based label assignment for tiny object detection//European conference on computer vision: 526-543

Xu C, Wang J, Yang W, Yu H, Yu L and Xia G S. 2022. Detecting tiny objects in aerial images: A normalized Wasserstein distance and a new benchmark. ISPRS Journal of Photogrammetry and Remote Sensing, 190: 79-93

Huang Y X, Liu H I, Shuai H H and Cheng W H. 2024. Dq-detr: Detr with dynamic query for tiny object detection//European Conference on Computer Vision: 290-305

Jia Y, Zhang J, Shan S and Chen X. 2020. Single-side domain generalization for face anti-spoofing//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition: 8484-8493

Wang Z, Wang Z, Yu Z, Deng W, Li J, Gao T, et al. 2022. Domain Generalization via Shuffled Style Assembly for Face Anti-Spoofing//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition: 4123-4133

Zhao G and Pietikainen M. 2007. Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE transactions on pattern analysis and machine intelligence, 29(6): 915-928

Zhang L, Hong X, Arandjelović O and Zhao G. 2022. Short and long range relation based spatio-temporal transformer for micro-expression recognition. IEEE Transactions on Affective Computing, 13(4): 1973-1985

Cai Y, Chen C L, Xu S, Liu H and Zhang T. 2026. Region-Aware Instance Consistency Learning for Micro-Expression Recognition//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition: 34428-34438

Guo D, Li K, Hu B, Zhang Y and Wang M. 2024. Benchmarking Micro-action Recognition: Dataset, Methods, and Applications. IEEE Transactions on Circuits and Systems for Video Technology, 34(7): 6238-6252

Li D, Shao J, Xing B, Gao R, Wen B, Kälviäinen H, et al. 2026. MSF-mamba: Motion-aware state fusion mamba for efficient micro-gesture recognition. IEEE Transactions on Multimedia

Li K, Li X, Wang Y, He Y, Wang Y, Wang L, et al. 2024. Video-mamba: State space model for efficient video understanding//European conference on computer vision: 237-255

Fan D P, Ji G P, Cheng M M and Shao L. 2021. Concealed object detection. IEEE transactions on pattern analysis and machine intelligence, 44(10): 6024-6042

Pang Y, Zhao X, Xiang T Z, Zhang L and Lu H. 2022. Zoom in and out: A mixed-scale triplet network for camouflaged object detection//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR): 2150-2160

Huang Z, Dai H, Xiang T Z, Wang S, Chen H X, Qin J, et al. 2023. Feature shrinkage pyramid for camouflaged object detection with transformers//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR): 5557-5566

Shao R, Lan X, Li J and Yuen P C. 2019. Multi-Adversarial Discriminative Deep Domain Generalization for Face Presentation Attack Detection//CVPR

Shao R, Lan X and Yuen P C. 2020. Regularized fine-grained meta face anti-spoofing//Proceedings of the AAAI Conference on Artificial Intelligence. 34(07): 11974-11981

Chen Z, Yao T, Sheng K, Ding S, Tai Y, Li J, et al. 2021. Generalizable representation learning for mixture domain face anti-spoofing//Proceedings of the AAAI Conference on Artificial Intelligence. 35(2): 1132-1139

Sun Y, Liu Y, Liu X, Li Y and Chu W S. 2023. Rethinking Domain Generalization for Face Anti-spoofing: Separability and Alignment//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition: 24563-24574

Le B M and Woo S S. 2024. Gradient Alignment for Cross-Domain Face Anti-Spoofing//CVPR

Zhou Q, Zhang K Y, Yao T, Lu X, Ding S and Ma L. 2024. Test-time domain generalization for face anti-spoofing//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition: 175-187

Ma H, Liu A, Xue S, Diao Y, Liang D, Liang H, et al. 2026. CLIP-SA: CLIP-Guided Semantic Alignment for Generalizable Face Anti-

© 中国图象图形学报版权所有

spoofing. IEEE Transactions on Multimedia

Liu A, Lin X, Zhi R, Liang Y, Zhu X, Cai Z, et al. 2026. DGPDL: domain-guided prompt distribution learning for generalizable face anti-spoofing. IEEE Transactions on Pattern Analysis and Machine Intelligence

Cheng Z, Leng S, Zhang H, Xin Y, Li X, Chen G, et al. 2024. Video-llama 2: Advancing spatial-temporal modeling and audio understanding in video-llms//arXiv preprint arXiv:2406.07476

Bai S, Chen K, Liu X, Wang J, Ge W, Song S, et al. 2025. Qwen2.5-VL Technical Report//arXiv preprint arXiv:2502.13923

Bergmann P, Jin X, Sattlegger D and Steger C. 2021. The mvtec 3d-ad dataset for unsupervised 3d anomaly detection and localization//arXiv preprint arXiv:2112.09045

Liu J, Xie G, Chen R, Li X, Wang J, Liu Y, et al. 2023. Real3d-ad: A dataset of point cloud anomaly detection. Advances in Neural Information Processing Systems, 36: 30402-30415

Bonfiglioli L, Toschi M, Silvestri D, Fioraio N and De Gregorio D. 2022. The eyecandies dataset for unsupervised multimodal anomaly detection and localization//Asian Conference on Computer Vision: 459-475

Li W, Zheng B, Xu X, Gan J, Lu F, Li X, et al. 2025. Multi-sensor object anomaly detection: Unifying appearance, geometry, and internal properties//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR): 9984-9993

Liu Z, Liu S, Min J, Zhang Z, Cen J, Han P, et al. 2026. InspecSafe-V1: A Multimodal Benchmark for Safety Assessment in Industrial Inspection Scenarios//arXiv preprint arXiv:2601.21173

Liu A, Tan Z, Wan J, Escalera S, Guo G and Li S Z. 2021. Casia-surf cefa: A benchmark for multi-modal cross-ethnicity face anti-spoofing//2021 IEEE Winter Conference on Applications of Computer Vision (WACV): 1178-1186

Stricker R, Müller S and Gross H M. 2014. Non-contact video-based pulse rate measurement on a mobile service robot//The 23rd IEEE international symposium on robot and human interactive communication: 1056-1062

Huang J, Zhu D, Yu Z, Ma H, Zhang J, Zhu C, et al. 2026. DeceptGPT: Schema-Driven Deception Detection with Multicultural Datasets and Robust Multimodal Learning//arXiv preprint arXiv:2603.23916

Li D, Xing B, Liu X, Xia B, Wen B and Kälviäinen H. 2025. Deemo: De-identity multimodal emotion recognition and reasoning//Proceedings of the 33rd ACM International Conference on Multimedia: 5707-5716

Wang J, Zhang J, Bian Y, Cai Y, Wang C and Pu S. 2021. Self-domain adaptation for face anti-spoofing//Proceedings of the AAAI Conference on Artificial Intelligence. 35(4): 2746-2754

Liu S, Zhang K Y, Yao T, Bi M, Ding S, Li J, et al. 2021. Adaptive normalized representation learning for generalizable face anti-spoofing//Proceedings of the 29th ACM International Conference on Multimedia: 1469-1477

Liu X, Shi H, Chen H, Yu Z, Li X and Zhao G. 2021. imigue: An identity-free video dataset for micro-gesture understanding and emotion analysis//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR): 10626-10637

Cai R, Yu Z, Kong C, Li H, Chen C, Hu Y, et al. 2024. S-adapter: Generalizing vision transformer for face anti-spoofing with statistical tokens. TIFS

Le T N, Nguyen T V, Nie Z, Tran M T and Sugimoto A. 2019. Anabranch network for camouflaged object segmentation. Computer vision and image understanding, 184: 45-56

Zavrtanik V, Kristan M and Skočaj D. 2021. DRAEM-A discriminatively trained reconstruction embedding for surface anomaly detection//2021 IEEE/CVF International Conference on Computer Vision (ICCV): 8310-8319

Liu Y, Li Z and Wu L. 2025. Dual consistency regularization for generalized face anti-spoofing. IEEE Transactions on Information Forensics and Security

Liu A. 2024. CA-MoEiT: Generalizable Face Anti-spoofing via Dual Cross-Attention and Semi-fixed Mixture-of-Expert. International Journal of Computer Vision: 1-14

You Z, Cui L, Shen Y, Yang K, Lu X, Zheng Y, et al. 2022. A unified model for multi-class anomaly detection//Advances in Neural Information Processing Systems. 35: 4571-4584

Guo J, Lu S, Jia L, Zhang W and Li H. 2023. Recontrast: Domain-specific anomaly detection via contrastive reconstruction//Advances in Neural Information Processing Systems. 36: 10721-10740

Li K, Gu J, Wang F, Wu Z, Fan H and Guo D. 2026. Ma-bench: Towards fine-grained micro-action understanding//arXiv preprint arXiv:2603.26586

Li K, Liu P, Guo D, Wang F, Wu Z, Fan H, et al. 2025. Mmad: Multi-label micro-action detection in videos//2025 IEEE/CVF International Conference on Computer Vision (ICCV): 13225-13236

Wang C, Zhu W, Gao B B, Gan Z, Zhang J, Gu Z, et al. 2024. Real-IAD: A Real-World Multi-View Dataset for Benchmarking Versatile Industrial Anomaly Detection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition: 22883-22892

Qu F, Wang S J, Yan W J, Li H, Wu S and Fu X. 2018. CAS(ME)2: A Database for Spontaneous Macro-Expression and Micro-Expression Spotting and Recognition. IEEE Transactions on Affective Computing, 9(4): 424-436

Li X, Cheng S, Li Y, Behzad M, Shen J, Zafeiriou S, et al. 2022. 4DME: A spontaneous 4D micro-expression dataset with multimodalities. IEEE Transactions on Affective Computing, 14(4): 3031-3047

Chen H, Shi H, Liu X, Li X and Zhao G. 2023. Smg: A micro-gesture dataset towards spontaneous body gestures for emotional stress state analysis. International Journal of Computer Vision, 131(6): 1346-

© 中国图象图形学报版权所有

1366

Wang C, Chen H, Wei H, Yang Y, Chen Y and Zhao G. 2026. imigue-3k: A large-scale benchmark for micro-gesture analysis with self-supervised learning//arXiv preprint arXiv:2605.17179

Zhang Z, Yan J, Liu S, Lei Z, Yi D and Li S Z. 2012. A face antispoofing database with diverse attacks//2012 5th IAPR international conference on Biometrics (ICB): 26-31

Chingovska I, Anjos A and Marcel S. 2012. On the effectiveness of local binary patterns in face anti-spoofing//2012 BIOSIG-proceedings of the international conference of biometrics special interest group (BIOSIG): 1-7

Wen D, Han H and Jain A K. 2015. Face spoof detection with image distortion analysis. IEEE Transactions on Information Forensics and Security, 10(4): 746-761

Liu Y, Jourabloo A and Liu X. 2018. Learning deep models for face anti-spoofing: Binary or auxiliary supervision//2018 IEEE/CVF Conference on computer vision and pattern recognition: 389-398

Liu Y, Stehouwer J, Jourabloo A and Liu X. 2019. Deep tree learning for zero-shot face anti-spoofing//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR): 4675-4684

Zhang Y, Yin Z, Li Y, Yin G, Yan J, Shao J, et al. 2020. Celeba-spoof: Large-scale face anti-spoofing dataset with rich annotations// European conference on computer vision: 70-85

Boulkenafet Z, Komulainen J, Li L, Feng X and Hadid A. 2017. OULU-NPU: A mobile face presentation attack database with real-world variations//2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017): 612-618

Rossler A, Cozzolino D, Verdoliva L, Riess C, Thies J and Nießner M. 2019. Faceforensics++ : Learning to detect manipulated facial images//Proceedings of the IEEE/CVF international conference on computer vision: 1-11

Li Y, Yang X, Sun P, Qi H and Lyu S. 2020. Celeb-df: A large-scale challenging dataset for deepfake forensics//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR): 3204-3213

Dolhansky B, Bitton J, Pflaum B, Lu J, Howes R, Wang M, et al. 2020. The deepfake detection challenge (dfdc) dataset//arXiv preprint arXiv:2006.07397

Jiang L, Li R, Wu W, Qian C and Loy C C. 2020. Deeperforensics-1.0: A large-scale dataset for real-world face forgery detection//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR): 2886-2895

Zi B, Chang M, Chen J, Ma X and Jiang Y G. 2020. Wilddeepfake: A challenging real-world dataset for deepfake detection//Proceedings of the 28th ACM international conference on multimedia: 2382-2390

Khalid H, Tariq S, Kim M and Woo S S. 2021. FakeAVCeleb: A novel audio-video multimodal deepfake dataset//arXiv preprint arXiv: 2108.05080

Hong Y and Zhang J. 2024. Wildfake: A large-scale challenging dataset for ai-generated images detection//arXiv preprint arXiv:2402.11843

Zhong N, Xu Y, Li S, Qian Z and Zhang X. 2023. Patchcraft: Exploring texture patch for efficient ai-generated image detection//arXiv preprint arXiv:2311.12397

Menze B H, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, et al. 2014. The multimodal brain tumor image segmentation benchmark (BRATS). IEEE transactions on medical imaging, 34 (10): 1993-2024

Landman B, Xu Z, Igelsias J, Styner M, Langerak T and Klein A. 2015. Miccai multi-atlas labeling beyond the cranial vault-workshop and challenge//Proc. MICCAI Multi-Atlas Labeling Beyond Cranial Vault—Workshop Challenge. 5: 12

Bilic P, Christ P, Li H B, Vorontsov E, Ben-Cohen A, Kaissis G, et al. 2023. The liver tumor segmentation benchmark (lits). Medical image analysis, 84: 102680

Chen C, Yuan J, Lu Y, Liu Y, Su H, Yuan S, et al. 2020. OoDAnalyzer: Interactive analysis of out-of-distribution samples. IEEE transactions on visualization and computer graphics, 27(7): 3335-3349

Kermany D S, Goldbaum M, Cai W, Valentim C C S, Liang H, Baxter S L, et al. 2018. Identifying medical diagnoses and treatable diseases by image-based deep learning. cell, 172(5): 1122-1131

Shih G, Wu C C, Halabi S S, Kohli M D, Prevedello L M, Cook T S, et al. 2019. Augmenting the national institutes of health chest radiograph dataset with expert annotations of possible pneumonia. Radiology: Artificial Intelligence, 1(1): e180041

Ehteshami Bejnordi B, Veta M, Johannes van Diest P, Van Ginneken B, Karssemeijer N, Litjens G, et al. 2017. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. Jama, 318(22): 2199-2210

Pérez-Rosas V, Abouelenien M, Mihalcea R and Burzo M. 2015. Deception detection using real-life trial data//Proceedings of the 2015 ACM on international conference on multimodal interaction: 59-66

Soldner F, Pérez-Rosas V and Mihalcea R. 2019. Box of lies: Multimodal deception detection in dialogues//Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers): 1768-1777

Gupta V, Agarwal M, Arora M, Chakraborty T, Singh R and Vatsa M. 2019. Bag-of-lies: A multimodal dataset for deception detection// 2019 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW): 83-90

Lloyd E P, Deska J C, Hugenberg K, McConnell A R, Humphrey B T and Kunstman J W. 2019. Miami University deception detection database. Behavior research methods, 51(1): 429-439

Cai C, Liang S, Liu X, Zhu K, Wen Z, Tao J, et al. 2025. Mdpe: A multimodal deception dataset with personality and emotional characteristics//Proceedings of the 33rd ACM International Conference

© 中国图象图形学报版权所有

on Multimedia：12957-12964

Niu X，Han H，Shan S and Chen X. 2018. VIPL-HR：A multi-modal database for pulse estimation from less-constrained face video// Asian conference on computer vision：562-576

Nowara E M，Marks T K，Mansour H and Veeraraghavan A. 2020. Near-infrared imaging photoplethysmography during driving. IEEE transactions on intelligent transportation systems，23(4)：3589-3600

Sabour R M，Benezeth Y，De Oliveira P，Chappe J and Yang F. 2021. Ubfc-phys：A multimodal database for psychophysiological studies of social stress. IEEE Transactions on Affective Computing，14(1)：622-636

Tang J，Chen K，Wang Y，Shi Y，Patel S，McDuff D，et al. 2023. Mmpd：Multi-domain mobile video physiology dataset//2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)：1-5

Yu X，Gong Y，Jiang N，Ye Q and Han Z. 2020. Scale match for tiny person detection//2020 IEEE winter conference on applications of computer vision (WACV)：1246-1254

Duan R，Deng H，Tian M，Deng Y and Lin J. 2022. SODA：A large-scale open site object detection dataset for deep learning in construction. Automation in Construction，142：104499

Ding J，Xue N，Xia G S，Bai X，Yang W，Yang M Y，et al. 2021. Object detection in aerial images：A large-scale benchmark and challenges. IEEE transactions on pattern analysis and machine intelligence，44(11)：7778-7796

Lv Y，Zhang J，Dai Y，Li A，Liu B，Barnes N，et al. 2021. Simultaneously localize，segment and rank the camouflaged objects//2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR)：11586-11596

Skurowski P，Abdulameer H，Błaszczyk J，Depta T，Kornacki A and Kozieł P. 2018. Animal camouflage analysis：Chameleon database. Unpublished manuscript，2(6)：7

Wu W，Ye S，Liu Y，He J，Wang Z and Yu Z. 2026. High-Resolution Underwater Camouflaged Object Detection：GBU-UCOD Dataset and Topology-Aware and Frequency-Decoupled Networks//arXiv preprint arXiv：2602.03591

Lamdouar H，Yang C，Xie W and Zisserman A. 2020. Betrayed by motion：Camouflaged object discovery via motion segmentation// Asian conference on computer vision：488-503

Cheng X，Xiong H，Fan D P，Zhong Y，Harandi M，Drummond T，et al. 2022. Implicit motion handling for video camouflaged object detection//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)：13854-13863

Liu Y，Ye S，Hao C and Yu Z. 2026. YUV20K：A Complexity-Driven Benchmark and Trajectory-Aware Alignment Model for Video Camouflaged Object Detection//arXiv preprint arXiv：2604.09985

Zhang X，Yin B，Lin Z，Hou Q，Fan D P and Cheng M M. 2025. Referring camouflaged object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence，47(5)：3597-3610

Liu Z，Zhou Y，Xu Y and Wang Z. 2023. Simplenet：A simple network for image anomaly detection and localization//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)：20402-20411

He H，Zhang J，Chen H，Chen X，Li Z，Chen X，et al. 2024. A diffusion-based framework for multi-class anomaly detection//Proceedings of the AAAI conference on artificial intelligence. 38(8)：8472-8480

He H，Bai Y，Zhang J，He Q，Chen H，Gan Z，et al. 2024. Mambaad：Exploring state space models for multi-class unsupervised anomaly detection//Advances in Neural Information Processing Systems. 37：71162-71187

Guo J，Lu S，Zhang W，Chen F，Li H and Liao H. 2025. Dinomaly：The less is more philosophy in multi-class unsupervised anomaly detection//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)：20405-20415

Liong S T，See J，Wong K and Phan R C W. 2018. Less is more：Micro-expression recognition from video using apex frame. Signal Processing：Image Communication，62：82-92

Wang T，Xia W，Ma H，Song Z，Zhang J，Wang Z，et al. 2026. GMoT：Gated Motion-Aware Tokenization for Fine-Grained Micro-Gesture Video Reasoning with Multimodal LLMs//arXiv preprint arXiv：2607.16322

Wang G，Han H，Shan S and Chen X. 2020. Unsupervised Adversarial Domain Adaptation for Cross-Domain Face Presentation Attack Detection. TIFS

Wang C Y，Lu Y D，Yang S T and Lai S H. 2022. Patchnet：A simple face anti-spoofing framework via fine-grained patch recognition// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition：20281-20290

Chollet F. 2017. Xception：Deep learning with depthwise separable convolutions//2017 IEEE conference on computer vision and pattern recognition (CVPR)：1800-1807

Cao J，Ma C，Yao T，Chen S，Ding S and Yang X. 2022. End-to-end reconstruction-classification learning for face forgery detection// 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)：4103-4112

Yan Z，Zhang Y，Fan Y and Wu B. 2023. Ucf：Uncovering common features for generalizable deepfake detection//2023 IEEE/CVF International Conference on Computer Vision (ICCV)：22355-22366

Kim Y，Jang M，Kwon M J，Lee W and Kim C. 2025. SELFI：Selective Fusion of Identity for Generalizable Deepfake Detection//arXiv preprint arXiv：2506.17592

Tan C，Zhao Y，Wei S，Gu G，Liu P and Wei Y. 2024. Rethinking the up-sampling operations in cnn-based generative network for generalizable deepfake detection//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition：28130-28139

Chen L，You Z，Zhang N，Xi J and Le X. 2022. UTRAD：Anomaly detection and localization with U-transformer. Neural Networks，

© 中国图象图形学报版权所有

147: 53-62

Lee S, Lee S and Song B C. 2022. Cfa: Coupled-hypersphere-based feature adaptation for target-oriented anomaly localization. IEEE Access, 10: 78446-78454

Karnati M, Seal A, Yazidi A and Krejcar O. 2021. LieNet: a deep convolution neural network framework for detecting deception. IEEE transactions on cognitive and developmental systems, 14(3): 971-984

Tang C, Li Y, Yang Y, Zhuang J, Sun G, Li W, et al. 2025. video-salmonn 2: Captioning-enhanced audio-visual large language models//arXiv e-prints: arXiv-2506

Ren S, He K, Girshick R and Sun J. 2016. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE transactions on pattern analysis and machine intelligence, 39(6): 1137-1149

Zhu H, Xu C, Yang W, Zhang R, Zhang Y and Xia G S. 2024. Robust tiny object detection in aerial images amidst label noise//arXiv preprint arXiv:2401.08056

Zhang H, Li F, Liu S, Zhang L, Su H, Zhu J, et al. 2022. Dino: Detr with improved denoising anchor boxes for end-to-end object detection//arXiv preprint arXiv:2203.03605

He C, Li K, Zhang Y, Zhang Y, You C, Li X, et al. 2024. Strategic preys make acute predators: Enhancing camouflaged object detectors by generating camouflaged objects//International Conference on Learning Representations. 2024: 36657-36675

Zhou Y, Chia M A, Wagner S K, et al. 2023. A foundation model for generalizable disease detection from retinal images. Nature, 622: 156-164. DOI: 10.1038/s41586-023-06555-x.

Linfeng Tang, Hao Zhang, Han Xu, Jiayi Ma. Deep learning-based image fusion: a survey[J]. Journal of Image and Graphics, 2023, 28(1): 3-36. (DOI:10.11834/jig.220422.

唐霖峰, 张浩, 徐涵, 马佳义. 基于深度学习的图像融合方法综述[J]. 中国图象图形学报, 2023, 28(1): 3-36). DOI: 10.11834/jig.220422.

Du Shirong, Fan Yingle, Cai Zhefei, Fang Tao. 2024. A contour detection method combing dual visual pathway and scale information fusion. Journal of Image and Graphics, 29(12):3657-3669DOI:10.11834/jig.230761. (杜仕荣, 范影乐, 蔡哲飞, 房涛. 2024. 结合双视通路与尺度信息融合的轮廓检测方法. 中国图象图形学报, 29(12):3657-3669)DOI:10.11834/jig.230761.

作者简介

余梓彤,男,长聘副教授,通讯作者,主要研究方向:微视觉计算。E-mail:yuzitong@gbu.edu.cn

叶硕,男,博士后,主要研究方向:微视觉计算。E-mail:shuoye.ke@gmail.com

林讯,男,博士后,主要研究方向:微视觉计算。E-mail:linxun@buaa.edu.cn

王俊,男,博士后,主要研究方向:微视觉计算。E-mail:wangjunsdnu@gmail.com

马辉,男,博士后,主要研究方向:微视觉计算。E-mail:mah@gbu.edu.cn

郭丹,女,教授,主要研究方向:微姿态识别。E-mail:guodan@hfut.edu.cn

邹博超,男,教授,主要研究方向:人机协同与情感计算。E-mail:zoubochao@gmail.com

胡瑞敏,男,教授,主要研究方向:步态识别。E-mail:hrm1964@163.com

单彩峰,男,教授,主要研究方向:情感计算。E-mail:cfshan@nju.edu.cn

刘青山,男,教授,主要研究方向:智能医学影像。E-mail:qsliu@njupt.edu.cn

赵耀,男,教授,主要研究方向:数字水印与数字取证。E-mail:yzhao@bjtu.edu.cn

操晓春,男,教授,主要研究方向:人工智能安全。E-mail:caoxiaochun@mail.sysu.edu.cn

山世光,男,教授,主要研究方向:基于视觉的情感计算和心理健康评估。E-mail:sgshan@ict.ac.cn

© 中国图象图形学报版权所有