E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035
中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TP18;TP391.9　　**文献标识码**:A　　**文章编号**:1006-8961(2026)07-2358-23

论文引用格式:Zhang H, Tian Y L, Wang Y T, Gou C, Li X and Wang F Y. 2026. Parallel images: the theoretical foundation and multimodal vision applications. Journal of Image and Graphics, 31(7):2358-2380(张慧, 田永林, 王雨桐, 苟超, 李轩, 王飞跃. 2026. 平行图像:从理论框架到多模态智能视觉应用. 中国图象图形学报, 31(7):2358-2380)[DOI:10.11834/jig.250231]

平行图像:从理论框架到多模态智能视觉应用

张慧[1],田永林[2],王雨桐[2],苟超[3],李轩[4],王飞跃[2,5*]

1. 北京交通大学计算机科学与技术学院,北京　100044; 2. 中国科学院自动化研究所,北京　100190;
3. 中山大学智能工程学院,广州　510275; 4. 北京理工大学人工智能学院,珠海　519088;
5. 欧布达大学平行智能分布式自主科学中心,布达佩斯　H-1034,匈牙利

摘　要:在深度学习推动下,视觉感知系统在智能驾驶、安防监控和医疗诊断等领域取得显著进展。然而,现实场景中数据分布的极度不均衡、长尾样本的稀缺性及高昂的人工标注成本,已成为制约模型性能与泛化能力的关键瓶颈。平行图像作为基于平行系统理论发展而来的新型图像生成与建模方法体系,通过构建人工场景系统、开展计算实验推演及虚实平行执行,形成"建模—训练—反馈—优化"的闭环机制,为视觉感知系统提供了高质量、多样化、结构化的合成数据支撑。本文系统梳理了平行图像的理论基础与发展脉络,重点综述其在虚拟场景生成、多模态特征融合、虚实域迁移和异构知识驱动的平行推理等关键技术路径上的研究进展与应用探索。同时,结合当前生成式人工智能和大规模多模态基础模型的发展趋势,分析了平行图像在智能视觉系统演化中的融合潜力。最后,本文指出该领域面临的主要挑战,并对未来的研究方向和应用前景进行了展望。

关键词:平行图像;生成式人工智能;虚实融合;场景生成;数字孪生

Parallel images: the theoretical foundation and multimodal vision applications

Zhang Hui[1], Tian Yonglin[2], Wang Yutong[2], Gou Chao[3], Li Xuan[4], Wang Feiyue[2,5*]

1. School of Computer Science and Technology, Beijing Jiaotong University, Beijing　100044, China;
2. Institute of Automation, Chinese Academy of Sciences, Beijing　100190, China; 3. School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou　510275, China; 4. School of Artificial Intelligence, Beijing Institute of Technology, Zhuhai　519088, China; 5. DeSci Center of Parallel Intelligence, Obuda University, Budapest　H-1034, Hungary

Abstract: Driven by the rapid advancement of deep learning and large-scale computing, visual perception systems have achieved remarkable progress in a wide range of domains, including autonomous driving, intelligent transportation, security surveillance, medical diagnostics, industrial inspection, and human-robot interaction. Modern vision algorithms, empowered by massive datasets and increasingly sophisticated neural architectures, are now capable of performing object detection, semantic segmentation, scene understanding, and even causal reasoning with unprecedented accuracy. Despite this rapid growth, the development of visual intelligence still faces several critical bottlenecks. Chief among these challenges are the highly imbalanced data distributions that characterize real-world environments, the scarcity of long-tail or rare-event samples essential for robustness, and the substantial human and financial cost associated with large-scale

收稿日期:2025-05-28;**修回日期**:2025-12-19;**预印本日期**:2025-12-26

*通信作者:王飞跃　feiyue.wang@ia.ac.cn

基金项目:国家自然科学基金项目(62203040)

Supported by: National Natural Science Foundation of China(62203040)

manual annotation. These factors significantly hinder the performance, safety, and generalizability of deep perception systems, especially in complex, dynamic, or safety-critical scenarios. Parallel images technology, emerging as a novel image generation and modeling methodology grounded in parallel systems theory and the artificial systems, computational experiments, and parallel execution (ACP) framework, offers a promising pathway to address these limitations. The core idea behind parallel images is to construct controllable, high-fidelity artificial scene systems that reflect the structure, behavior, physics, and semantics of their real-world counterparts. Within these artificial systems, computational experiments can be conducted at scale, allowing for the controlled generation of diverse visual data that capture variations in illumination, geometry, environmental conditions, sensor characteristics, and task-specific factors. Through the interaction and iterative feedback between virtual and real environments, parallel images establish a closed-loop mechanism of "modeling-training-feedback-optimization", enabling perception models to evolve continuously, validate hypotheses, and improve performance under systematically generated variations. This closed-loop mechanism differentiates parallel images from traditional synthetic data generation in several important aspects. First, instead of passively producing static rendered images, parallel images emphasize dynamic parallelism, where virtual agents, environments, and tasks evolve in sync with real-world processes. Second, the approach integrates multimodal feedback, bridging visual, geometric, physical, and semantic modalities to ensure consistency and translatability across domains. Third, the framework supports scalable modeling of rare, dangerous, or expensive scenarios that are difficult or impossible to capture in real life, such as near-crash events in autonomous driving, rare diseases in medical imaging, or hazardous industrial operations. These capabilities make parallel images a powerful tool for enhancing the robustness, safety, and domain generalization of modern perception systems. This paper provides a comprehensive, systematic review of the theoretical foundations, methodological innovations, and developmental trajectory of parallel images technology. This paper begins by revisiting its roots in parallel intelligence and the ACP paradigm and detailing how artificial systems serve as controlled experimental platforms that complement real-world data collection. Then, this paper examines recent technical advances encompassing three major research directions aligned with the ACP framework: 1) multimodal data-driven virtual scene generation, which employs generative adversarial networks, diffusion models, neural radiance fields, and 3D Gaussian splatting to overcome data scarcity and annotation bottlenecks, enabling the creation of controllable, editable, and semantically consistent synthetic environments; 2) multiview feature fusion and virtual-real model transfer, aimed at addressing feature discrepancies and semantic misalignment across heterogeneous visual modalities through cross-modal alignment, multigranularity adaptive transfer, and domain-bridging strategies that enhance generalization and adaptability in hybrid virtual-real environments; 3) parallel reasoning through heterogeneous data and knowledge fusion, which integrates structured information extraction, external knowledge guidance, scene graphs, temporal logic, and large language models to advance perceptual understanding toward semantic-level reasoning and decision-making, thereby supporting continuous optimization and closed-loop evolution in complex scenes. Beyond summarizing technological developments, this paper also situates parallel images within the broader context of emerging trends in generative artificial intelligence and foundation models. With the rise of diffusion models, neural radiance fields, and large-scale multimodal models, parallel images are poised to integrate more deeply with generative simulation pipelines. This paper discusses how these innovations can strengthen the fidelity, controllability, and adaptability of artificial visual data and potentially enable new capabilities such as task-conditioned scene synthesis, human-AI cosimulation, interactive data generation, and closed-loop autonomous scenario exploration, and provides key capabilities for building general visual systems with continuous learning and feedback optimization. This paper provides crucial support for building general visual systems with continuous learning and feedback-driven optimization. Finally, this paper identifies several open challenges and future research directions that are essential for advancing the development of parallel images systems. These challenges include achieving high-quality expansion of virtual data, bridging the semantic gap between virtual and real domains, and enabling real-time, tightly coupled virtual-real interaction. Addressing these issues will require advances in intelligent generation models, self-supervised quality evaluation, unified data standards, causality-aware cross-domain alignment, and low-latency virtual-real collaboration supported by next-generation communication and sensing technologies. Solving these challenges will be critical for pushing forward the frontier of synthetic visual intelligence and unlocking the full potential of parallel images in real-world applications.

Key words: parallel images; generative artificial intelligence; virtual-real integration; scene generation; digital twin

0 引 言

近年来，随着人工智能（artificial intelligence，AI）特别是生成式人工智能（generative artificial intelligence）和大规模基础模型（foundation models）的持续突破，计算机视觉作为感知智能的核心支柱，正在加速其在各类复杂场景中的应用落地。从智能驾驶到城市安防，从智慧医疗到工业质检，视觉系统的识别、理解与推理能力不断迈向高精度与高智能。然而，尽管在算法结构、模型容量以及算力资源方面已取得显著进展，当前视觉感知系统仍面临“数据依赖性强、泛化能力弱、训练成本高”等核心困境，严重制约其在现实环境中的广泛部署与性能提升。

视觉模型的性能高度依赖于训练数据的多样性与标注质量。然而，在真实世界中，视觉数据呈现出明显的“长尾分布”特征：多数数据集中在常规场景如晴天、白天、规则交通中，而极端情境（如暴雨、雪夜、突发事故）下的图像数据极为稀缺。这种数据分布的不均衡，导致模型对高风险、低频事件的感知与响应能力显著下降，引发实际应用中的误判风险与安全问题。与此同时，传统数据采集与标注机制高度依赖人工操作，在面向图像分类、目标检测和语义分割等任务时，需执行帧级、框级乃至像素级的精细标注，造成标注效率瓶颈、人力资源紧张和质量控制难题。

尽管生成对抗网络（generative adversarial network，GAN）（Xu等，2024）、扩散模型（diffusion models）（Rombach等，2022）等生成式人工智能方法能够通过学习数据分布生成具有特定场景特征的合成视觉数据，但目前大多数生成模型仍集中于图像合成质量、风格控制等通用层面，难以生成具备高度物理一致性、语义完整性与任务可控性的专业级合成视觉数据，尤其在涉及多视角一致性、三维空间约束和动态交互建模等复杂场景时存在明显短板。此外，基础大模型，如CLIP（contrastive language-image pre-training）（Radford等，2021）、SAM（segment anything model）（Kirillov等，2023）和GPT-4V（generative pre-trained Transformer）（Wu等，2024）等虽然在多模态对齐、跨任务泛化方面展现出强大潜力，但在高精度、高安全等级的视觉感知应用中，其缺乏可解释性、不可控生成行为，以及训练数据透明性等问题依旧突出，亟待结合更具结构化、工程化的技术框架进行融合与校准。

与此相呼应，早在2004年，中国科学院自动化研究所王飞跃研究员便提出了平行系统理论（Wang，2007），将人工系统（artificial systems）、计算实验（computational experiments）与平行执行（parallel execution）有机结合，形成ACP（artificial systems，computational experiments，and parallel execution）方法。追溯其思想源头，1994年王飞跃研究员便提出“影子系统”概念（Wang，1994，2023），将模型视为数据生成器和可视化工具，这一构想为后续平行系统与平行智能的提出奠定了理论雏形（Yang等，2023a；Wang等，2024）。在平行智能理论的指导下发展出的“平行图像”（parallel images）（Wang，2020b；Yang等，2023b），通过在人工系统中构建与真实世界高度对应且可精确控制的虚拟场景，能够生成大规模、多样化并具备精细标注的合成视觉数据，为数据稀缺或不可采集场景下的模型训练、测试与对比评估提供了一种可控、可复制以及低成本的解决方案。

具体而言，平行图像围绕“人工场景生成—计算实验推演—虚实平行执行”三大核心模块展开：首先，借助虚拟仿真平台与三维建模技术构建多维异构、参数可控的人工视觉场景，涵盖天气、照明、视角和障碍物行为等多种变量组合；其次，基于生成数据进行计算实验，以验证与优化不同视觉模型在复杂场景下的鲁棒性与泛化能力；最后，结合现实世界中的反馈信号与性能指标，反向调整人工场景建模参数，实现虚拟—现实闭环优化与模型迭代升级。这一机制突破了传统“静态训练 + 一次部署”的范式，推动视觉系统逐步迈向“虚拟训练—现实适应—持续演进”的智能闭环。

在此基础上，随着多模态大模型的兴起，平行图像正迎来新的发展契机。一方面，生成式大模型能够作为虚拟场景构建与图像生成的核心引擎，基于文本、语义图或草图等多模态提示实现高质量、语义一致的图像合成；另一方面，平行图像又能提供结构化、标注完备的数据支撑，反哺大模型在特定任务和

应用场景中的微调与适配。二者的结合不仅提升了感知系统的表达力与适应性，更有望构建“任务驱动、语义约束、人机协同”的新一代视觉感知体系，进而推动智能形态由单纯依赖数据驱动，向模型驱动与知识引导并重的方向演进。

平行图像作为一种融合人工智能建模、虚拟仿真与现实反馈的系统性方法体系，可以有效应对数据稀缺、标注高成本与泛化困难等挑战。其与生成式人工智能、大模型的融合发展，不仅为感知系统带来更强的表达力与适应性，也为 AI 系统的可信性、可控性与系统性提升提供了坚实支撑。本文将围绕平行图像的核心概念、关键技术与典型应用进行系统性综述，并展望其在大模型时代的研究前景与应用潜力。

1　平行图像的发展历程

随着人工智能技术和平行理论在计算机视觉领域的深度融合与创新应用，平行图像已经发展出一套完整的理论方法体系和技术框架。如图 1 所示，本文从时间脉络与技术演进的角度，将平行图像的发展历程系统性地划分为 4 个关键阶段：探索试验阶段。理论形成阶段、扩展深化阶段和前沿突破阶段，各阶段间通过“实践验证—理论提炼—领域扩展—技术革新”的逻辑层层递进，共同推动平行图像从方法雏形走向范式化应用。

1）探索试验阶段。平行图像作为平行系统在图像领域的延伸，其早期探索与平行系统理论发展相辅相成。针对复杂系统（如社会经济、交通管理）研究中面临的预测不准、建模困难和传统控制方法局限等问题，平行系统（王飞跃，2004）作为一种创新性解决方案：通过构建人工系统实现计算实验与虚实互动，采用“多重世界”视角缩小建模鸿沟，并提升人工系统的主动性以实现动态管控。随后，Wang（2007）提出了 ACP 方法，为平行系统的研究构建了一套完整的计算理论和方法体系。其核心思想包括 3 个方面：首先，构建人工社会作为计算实验室，通过自下而上地建模模拟复杂系统的涌现行为，突破传统“单一世界”观，将人工系统视为现实的可能等价体；其次，采用“多重世界”视角，使人工系统与现实系统平行演化，实现复杂系统的可控、可重复分析；最后，建立虚实互动机制，通过动态对比与闭环反馈，推动系统的实验评估、决策优化与智能管控。

研究者积极探索平行系统和 ACP 方法在多个

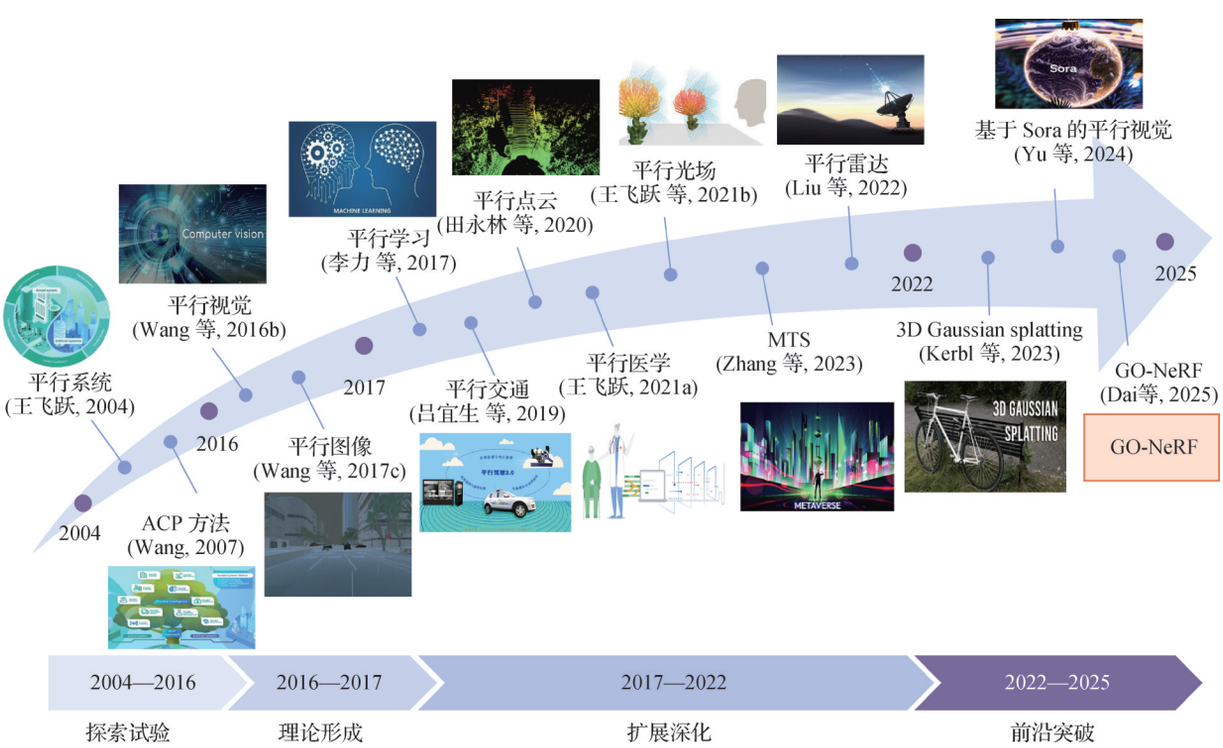

图 1　平行图像的发展历程

Fig. 1　The development history of parallel images

领域的创新应用。在交通管理领域，相关研究逐步从单一系统控制向多主体协同与复杂系统仿真方向发展。Li等人(2005)在IEEE国际智能车辆研讨会上系统综述了该领域的新进展与研究趋势，从智能车辆感知、车辆运动控制与通信，以及驾驶员与乘客辅助三大方向进行了深入探讨，并指出车与车、车与路侧设施、车与驾驶员之间的信息共享，以及不同厂商产品间的通信协议与互操作性，是当前亟待解决的关键问题。Wang(2010)通过整合实际交通系统、人工系统以及训练与评估模块，实现了计算实验与虚实互动的有机结合。该系统在济南、太仓等地的实际应用中显著提升了交通系统的应急响应能力和管理策略优化水平。Zhu等人(2011)通过构建人工交通系统，采用基于主体的建模方法，成功模拟了人口结构、活动计划、出行行为及恶劣天气等环境因素对交通决策的影响。这一方法有效解决了传统交通模拟在微观行为建模和间接因素考量方面的局限性。在公共安全领域，ACP方法同样展现出强大的适应能力，Duan等人(2013)通过构建融合多模型的人工社会，成功模拟了流行病传播过程并评估了不同干预策略的效果，为疫情防控提供了科学依据。在城市治理层面，Wang等人(2016a)将ACP方法应用于智能停车系统，通过模拟和优化不同管理策略，实现了城市停车资源的动态配置与高效调度。这些研究成果充分证明，ACP方法通过“建模—实验—反馈—优化”的闭环机制，为解决复杂系统的智能管控问题提供了通用性强、效率高且可持续的技术路径，也为平行图像从理论借鉴走向领域适配奠定了实践基础。

2)理论形成阶段。正是基于探索试验阶段在交通、公共安全等领域的大量实践，ACP方法在复杂系统建模与优化中的有效性得到了充分验证，同时也暴露了传统计算机视觉技术在数据稀缺、标注成本高以及模型泛化能力不足等问题上的局限。研究者开始思考将平行系统与ACP方法的核心逻辑迁移至计算机视觉领域。王坤峰等人(2016b)首次系统性地提出了基于ACP理论的智能视觉计算方法。该方法创新性地将平行系统的3个核心要素与计算机视觉技术深度融合：通过人工场景构建解决数据稀缺问题，借助计算实验优化模型性能，利用平行执行实现持续改进。2017年，平行图像理论(王坤峰等，2017c)被提出用于解决计算机视觉领域因实际图像数据采集困难、标注成本高而导致的模型泛化能力不足问题。其核心突破在于：(1)将传统图像生成从单一数据采集扩展为虚实协同的平行系统模式；(2)建立动态进化学习机制，使模型能在开放环境中持续自我更新。这些创新为自动驾驶、智能监控等动态场景下的视觉任务提供了全新解决方案，推动了计算机视觉从静态分析向动态进化的范式转变。

3)扩展深化阶段。在理论形成阶段确立了平行图像的核心逻辑后，平行系统理论及ACP方法的普适性得到进一步验证，开始向更多复杂领域延伸，形成领域特色关键技术，这些跨领域发展反过来又为平行图像提供了跨场景适配的参考经验。

当前，平行系统理论已成功拓展至多个重要领域：在机器学习领域创新提出了“平行学习”(李力等，2017)，通过软件定义的人工数据系统与真实数据平行演进，解决了小样本条件下的模型训练难题；在智能交通领域衍生出“平行交通”(吕宜生 等，2019)，通过构建人工交通系统与实际路网平行互动，实现了拥堵预测与信号控制的协同优化；在医疗健康领域发展为“平行医学”(王飞跃，2021a)，通过构建人工医学系统与实际医疗数据的平行互动，基于ACP方法将临床“小数据”扩展为合成“大数据”，并进一步提炼为用于精准诊疗的“深智能”，从而有效应对罕见病诊断与个性化治疗等小样本条件下的模型训练难题。这些领域突破不仅验证了ACP方法的普适价值，更为解决交通管理、疾病诊疗和机器学习等复杂系统问题提供了“虚实互动、平行进化”的新范式。

从平行图像的纵深化发展来看，其在平行视觉及多个视觉任务上的应用不断深化。在平行视觉领域，ACP方法的技术潜力同样日益凸显。李轩和王飞跃(2021)提出了面向智能驾驶的平行视觉感知框架。该框架基于ACP方法构建人工驾驶场景，通过合成数据训练与虚实交互优化视觉模型，有效提升了复杂环境下目标检测与语义分割等任务的性能与泛化能力。张慧等人(2021)聚焦平行视觉的技术实现，提出了利用合成数据训练视觉模型，通过全局/局部特征对齐和虚实交互提升模型性能，验证了其在目标检测和实例分割任务中的有效性。鉴于元宇宙在虚拟仿真与数据生成方面的强大能力，Zhang等人(2023)进一步提出了基于元宇宙的平行交通视

觉框架，通过构建虚拟交通空间、计算实验驱动的模型学习和虚实平行的反馈优化，解决智能交通系统中环境感知的泛化性、数据稀缺性和复杂场景适应性等挑战，拓展了平行视觉的应用边界。苟超等人(2024)在平行视觉框架下整合描述、指示和预测智能，实现了基于视觉的风险增强感知。

除此之外，在智能交通测试领域，Li 等人(2019a)通过虚拟—现实交互的平行测试系统，将真实场景数据转换为虚拟极端场景(如恶劣天气、紧急事件)，生成多样化测试任务，以高效、安全地评估和提升自动驾驶车辆的智能水平。在行人检测领域，Zhang等人(2020)通过平行视觉方法构建虚拟场景生成合成数据，结合在线学习机制，实现了对固定摄像头场景下行人检测模型的高效训练与动态优化。在医学图像分析领域，Shen等人(2021)通过构建病理组织的人工模型与真实病例数据平行互动，实现了小样本条件下的高精度诊断。为了突破传统二维图像在表达维度和真实感方面的局限，平行图像进一步融合三维重建、光场成像和雷达仿真等技术，衍生出平行点云(田永林 等，2020)、平行光场(王飞跃 等，2021b)和平行雷达(Liu等，2022)等新方向，结合光场成像对全光信息的捕捉能力与三维重建的空间建模优势，实现对复杂场景的多维度、高保真数字化重构。

4)前沿突破阶段。上述阶段的技术积累，使平行图像具备了场景适配与多维度感知的基础能力，但在高多样性人工场景构建效率、跨模态语义一致性以及实时辐射场渲染质量等方面仍存在局限，而多模态大模型的兴起与新型3D场景表示技术的突破，为解决这些关键问题提供了不同技术路径：一方面，生成式大模型大幅降低了高多样性人工场景的构建成本，通过接收多模态输入形式，能够生成逻辑连贯、风格统一和结构合理的二维图像与三维场景。Yu等人(2024)提出了基于视频生成模型Sora的平行视觉框架，通过将Sora与ACP平行系统理论结合，构建了虚实交互的闭环智能感知系统。该系统实现了从文本描述自动生成高保真驾驶场景数据，并利用反馈优化机制持续提升智能车辆在未知场景下的感知泛化能力。Zhang等人(2024)提出了基于神经辐射场(neural radiance fields，NeRF)的文本驱动3D场景生成方法，结合预训练文本—图像扩散模型生成初始内容与几何先验，并通过渐进式修复策略，实现了多视角一致的高保真场景合成。Dai等人(2025)提出了基于神经辐射场的虚拟现实内容生成方法，通过文本提示与用户指定区域生成高质量3D对象，并采用组合渲染与优化策略，实现了对象与场景的无缝融合。另一方面，Kerbl等人(2023)提出的3D高斯点渲染(3D Gaussian splatting)技术通过引入可优化的3D高斯表示与高效的GPU(graphics processing unit)栅格化渲染算法，在保证图像质量的同时大幅提升了训练和渲染速度，实现了高质量、实时的三维场景合成。该技术为平行图像中的实时高保真辐射场渲染提供了有效解决方案，已逐步应用于人工场景的生成与渲染环节，进一步完善了平行图像的技术体系。

借助大模型的泛化能力、知识表示与3D高斯点渲染的高效渲染能力，平行图像可在多个任务领域实现跨模态、高一致性的合成数据生成与实时高保真渲染，由此产生的图像数据在语义表达与结构组织上具备高度可控性，也便于精细化标注与可重复利用，为视觉模型在训练、验证与部署阶段提供稳定、丰富且高质量的数据与渲染支撑。通过多技术融合，平行图像不仅解决了数据获取、标注与实时渲染的瓶颈问题，更为构建可解释、可控制和可进化的新一代视觉感知体系奠定了坚实基础。

2 平行系统理论及ACP方法

平行系统理论旨在解决传统方法在社会经济、交通、医疗及视觉感知等复杂系统中面临的建模不准、实验困难与控制复杂等挑战。该理论将复杂系统“不可控、难预测”的特性转化为“可控、可推演”的计算问题，其核心是通过构建人工系统模拟现实系统的关键特征与动态演化规律，借助计算实验探索系统行为与干预策略之间的关联机制，最终通过虚实平行执行实现系统的闭环优化。

ACP方法是平行系统理论的核心实现框架，包括人工系统、计算实验和平行执行3个部分。人工系统(A环节)作为基础载体，通过对现实系统进行结构化、可编程的建模，构建出“高保真、可调控、可扩展”的虚拟环境，而非简单复制现实。计算实验(C环节)是在人工系统中进行模拟测试、策略优化与模型验证的核心步骤，相当于为复杂系统构建一

个“计算实验室”，以克服现实实验成本高、周期长和风险大的限制。平行执行（P环节）通过双向数据交互与协同演化，将计算实验的成果应用于现实系统，并利用现实反馈持续优化人工系统，形成虚实互促的长期优化机制。

ACP方法的3个环节环环相扣，人工系统为计算实验提供载体，计算实验为平行执行提供策略，平行执行反过来为人工系统提供更新依据，共同支撑“建模—实验—反馈—优化”的闭环流程，实现平行系统的动态优化目标。

3 平行图像整体框架

平行图像整体框架是平行系统理论与ACP方法在视觉领域的具体实践，将平行图像所面对的“复杂环境下视觉系统建模难、优化难”问题，转化为“可控、可推演”的计算问题。人工场景生成（A环节）通过构建色彩逼真的人工场景，模拟实际场景中可能出现的环境条件，为平行图像提供基础数据与场景载体；计算实验推演（C环节）利用人工场景进行计算实验，评价视觉算法有效性或优化模型参数，成为平行图像优化视觉模型的关键手段；虚实平行执行（P环节）利用视觉模型在实际与人工场景中协同运行，实现了训练评估的在线化、长期化，通过虚实数据交互推动平行图像系统持续优化。在平行图像整体框架中，如图2所示，人工场景生成、计算实验推演与虚实平行执行构成了一个由数据构建到模型优化再到闭环反馈链条，三者相辅相成，构建一个持续自我优化的智能视觉计算体系，为复杂环境下的智能感知与理解提供坚实支撑。

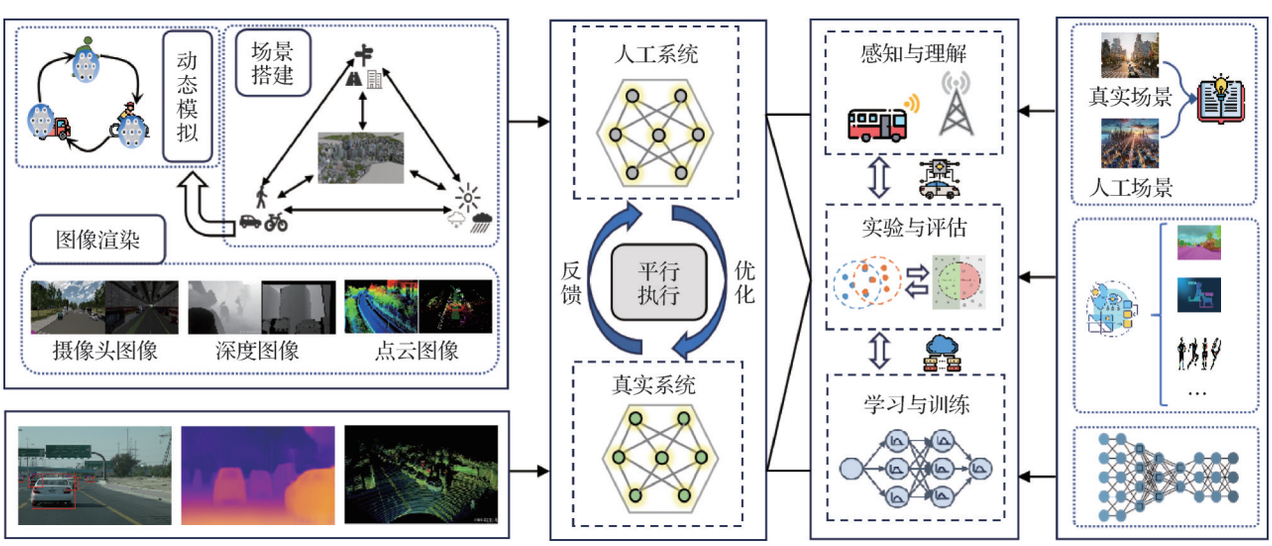

图2 平行图像整体框架

Fig. 2 Overall framework of parallel images

3.1 人工场景生成

人工场景生成作为平行图像整体框架的基础环节，其核心在于通过计算机图形学、虚拟现实和微观仿真等前沿技术，构建高度逼真且多样化的虚拟环境，以精准模拟复杂多变的实际场景。

人工场景生成的过程涵盖3个核心流程，分别为场景框架构建、动态行为模拟和图像渲染。其中，场景框架构建为动态行为模拟提供了坚实的物理与几何基础，动态行为模拟赋予环境以真实而复杂的交互动态，最终通过可编程渲染系统将这一虚拟环境转化为高保真、可标注的视觉数据。首先，场景框架构建依托三维建模工具和物理引擎创建场景拓扑并配置物体属性。在实际操作中，利用游戏引擎、仿真工具以及3D建模软件搭建场景，并从网络资源获取或自行创建海量三维模型，构建包含静态物体（建筑物、道路、植被等）和动态物体（行人、车辆等）以及自然环境要素（如雨、雪、雾等）的完整场景框架，为每个场景元素赋予精确的物理属性和几何特征。其次，在此静态基础上对行人、车辆等动态物体进行行为模拟，通过智能算法进行路径规划和障碍规避等复杂行为，并利用实时通信机制实现多主体协同交互。最后，通过可编程虚拟摄像机系统模拟真实拍

摄过程进行图像渲染与生成。该系统支持多角度、多参数的可控图像采集，结合光线追踪等先进渲染技术，生成具有完整标注信息的高保真图像数据集。

人工场景在数据多样性、标注准确性和场景覆盖度等方面具有显著优势，能有效解决真实场景下面临的数据采集成本高、标注效率低以及长尾场景覆盖不足等关键问题，为计算机视觉模型的训练和验证提供可靠的数据基础。

3.2 计算实验推演

计算实验推演是平行图像整体框架的核心环节，主要聚焦于使用生成的人工场景数据进行大规模计算实验来测试、优化和验证视觉算法的性能。

人工生成的虚拟数据能够高效模拟各类极端及长尾场景，这些场景在实际环境中难以大规模采集，从而为视觉模型的鲁棒性验证提供了多样化的测试条件。通过将虚拟数据与真实数据进行融合，可显著提升视觉模型的泛化性能，使其充分结合虚拟数据的覆盖广度与真实数据的分布真实性，从而提高视觉模型对各类场景的适应能力。考虑到人工场景数据和实际场景数据分布的差异，通过领域迁移技术的应用，模型首先在大量虚拟数据上学习通用特征表示，再通过实际数据进行针对性微调，显著提升了模型在实际应用场景中的适应能力。

针对不同的应用需求和研究目的，选择合适的视觉模型进行训练至关重要。如在虚拟交通场景中训练自动驾驶感知模型，可凭借其在特征提取与目标定位上的能力，有效适应不同交通密度、行人行为等复杂情况。在图像分类任务中，Vision Transformer、EfficientNet等架构凭借自注意力机制和更优的参效比，在合成数据训练中有更强的迁移性能，借助人工场景数据能更精准地判断图像类别；在语义/实例分割任务里，利用人工建筑场景训练分割模型，可大幅提高对道路、建筑、植被等元素的识别精度；在姿态估计与行为分析方面，通过在虚拟人体运动场景中训练动作识别模型，能够优化对复杂姿态的预测能力。在训练过程中，不断调整模型参数，使模型学习图像数据中的特征和模式，进而提高其在特定任务上的性能。

通过对模型在不同数据集上的性能评估，可以验证模型的泛化能力和对不同场景的适应性，发现模型存在的问题和不足，为后续改进提供依据。计算实验还可用于模型的比较和选择，将不同的视觉模型在相同数据集上训练和评估，对比性能指标，选择性能最优的模型用于实际应用。同时，通过计算实验探索不同模型结构、训练参数对模型性能的影响，为模型优化提供依据，如调整神经网络的层数、学习率等参数，观察模型性能变化，找到最优模型配置。

3.3 虚实平行执行

虚实平行执行是平行图像整体框架实现闭环的关键实践。它借助虚实互动的方式，基于实际场景优化反馈虚拟场景模型，并利用人工场景的模型引导实际场景，从而在线优化视觉模型，实现对复杂环境的智能感知与理解。

在平行执行过程中，通过将视觉模型部署到实际场景中进行测试和验证，收集模型在实际应用中的反馈信息，这些信息涵盖模型的预测结果与实际情况的偏差、在不同场景下的性能表现等多方面内容。依据这些反馈，对虚拟场景模型进行针对性调整和优化。针对智能驾驶场景中视觉模型在特定气象条件（如雾天）下出现的目标检测误判率偏高问题，可通过虚拟场景建模进行针对性优化。具体而言：1）基于实际测试反馈数据，在虚拟环境中构建相应的雾天场景模型，精确调节大气散射参数、光照强度及能见度等关键指标，确保虚拟场景的物理特性与实际环境保持高度一致；2）优化虚拟场景中动态目标的特征表征与行为模式，包括但不限于调整车辆在低能见度条件下的行驶速度分布、雾灯开启策略等关键参数，使虚拟环境中的目标特征与行为规律更符合实际观测数据。

此外，基于实际场景反馈优化后的虚拟场景模型，会进一步用于引导实际场景中的视觉模型。人工场景具有高度的可定制性和重复性，在虚拟场景中经过优化的模型参数、检测算法等，可以迁移到实际场景的视觉模型中。通过不断地进行这样的平行执行，即从实际场景获取反馈进而优化虚拟场景模型，再将优化后的虚拟场景模型应用到实际场景持续引导视觉模型，实现两者的协同进化。这使得视觉模型能够不断提升其性能和适应性，从而更加精准地实现对复杂环境的智能感知与理解，为实际应用提供更可靠的支持。

4 平行图像的关键技术

平行图像的核心目标是依据平行系统理论，构建以“人工场景—计算实验—平行执行”为框架的图像生成体系，如图3所示，通过场景构建、智能训练与反馈优化的闭环机制，为视觉感知系统持续提供高质量、多样化且高度结构化的合成数据支持。其关键技术主要包括多模态数据驱动的虚拟场景生成、多视域特征融合与虚实模型迁移，以及异构数据与知识融合的平行推理3个核心部分。其中，多模态数据驱动的虚拟场景生成通过使用生成对抗网络、扩散模型和神经辐射场等技术，来应对真实数据稀缺与标注成本高昂的瓶颈，生成可控、可编辑且语义一致的基础视觉数据；多视域特征融合与虚实模型迁移致力于解决多源跨域视觉数据中的特征差异与语义不对齐问题，借助跨模态对齐和多粒度自适应迁移策略，增强模型在虚实融合环境中的泛化与适应能力；异构数据与知识融合的平行推理进一步融合结构化信息提取与外部知识引导，推动视觉感知向语义理解与决策推理的深化演进，从而为复杂场景下的智能系统构建具备持续优化与闭环演进能力的支撑体系。

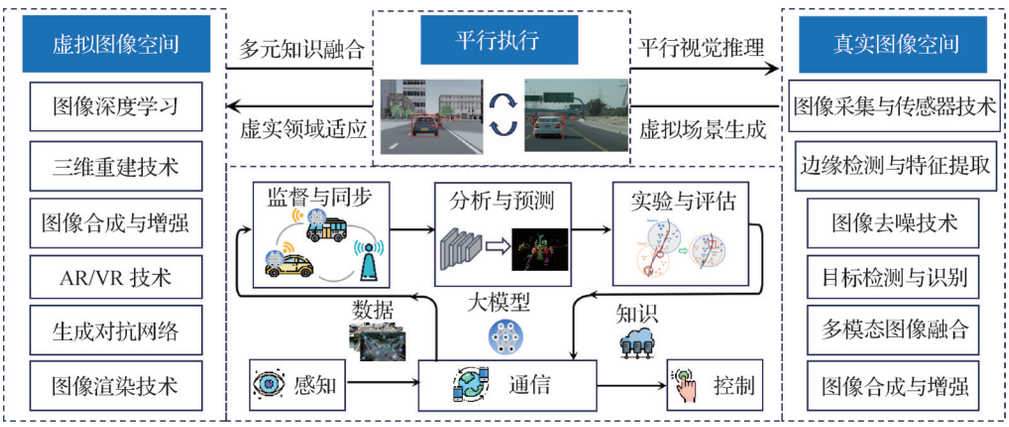

图3 平行图像关键技术

Fig. 3 Key technologies of parallel images

4.1 多模态数据驱动的虚拟场景生成

4.1.1 内容可编辑的虚拟场景生成

内容可编辑的虚拟场景生成利用深度学习、虚拟现实等生成合成类算法，依据输入的文本、图像和音频等数据，构建出可供用户进行内容编辑操作的虚拟场景。在平行图像系统中，此类技术进一步与软件定义的人工图像系统结合，能够基于实际场景中的“小数据”自动生成大规模、带标注的人工图像，构建适用于模型训练的平行图像“大数据”集合。以田永林等人(2020)提出的融合虚拟与现实数据的点云生成与三维模型演化闭环框架为例，该方法基于ACP理论，通过构建可编辑的虚拟场景，自动生成带标注的点云数据，并与真实点云融合形成混合数据集。在此基础上，利用计算实验对三维模型进行训练和评估，评估结果反过来指导虚拟场景的优化，实现数据生成与模型训练的协同进化。近年来，神经辐射场、扩散模型和3D高斯点渲染等生成式方法的出现，如表1所示，为平行图像系统中虚拟场景的生成提供了新的技术路径。这些方法不仅能够支持更高真实感、更强可控性的场景构建，还可实现更灵活的语义编辑与更高效率的视觉内容生成，进一步拓展了平行图像系统在视觉智能领域的应用潜力。

传统的3D重建与渲染方法主要依赖几何建模与纹理映射技术，通过构建精细的多边形网格并贴合真实纹理，生成逼真的三维场景。然而，此类方法面临数据获取困难、建模复杂度高以及渲染成本大的问题，尤其在处理动态场景或复杂材质时存在局限。为应对这些挑战，基于深度学习的方法逐步兴起，早期研究侧重于二维图像合成与样本生成，但难以直接适用于三维场景建模。在此背景下，Mildenhall等人(2021)提出了一种基于神经网络的三维场景表示与新视角图像合成方法NeRF。该方法将场景建模为一个连续的五维函数，输入为三维空间坐标(x, y, z)和视角方向(θ, φ)，输出为该点的颜色(RGB)和体积密度(σ)。通过多层感知机(multilayer perceptron，MLP)对该函数进行参数化，利用位

表 1　不同场景生成方法的差异对比

Table 1　Comparison of different scene generation methods

类别	方法	场景生成效率	真实感	编辑灵活性
神经辐射场	Zhang 等人(2024) Mildenhall 等人(2021)	生成速度较慢，依赖密集采样和复杂体积渲染，训练和推理时间较长	可生成高质量逼真场景，但对光照和材质变化敏感，局部细节可能出现伪影或不一致	编辑灵活性较低，对场景进行结构和外观上的大幅修改通常需要重新训练
扩散模型	Lee 等人(2024) Zhou 等人(2024)	生成速度较快，推理效率高于 NeRF，支持多样化的图像和场景合成	生成结果真实感强，细节丰富，风格可控，能够高度贴合自然图像分布，光学一致性较好	编辑灵活性较强，支持基于文本、草图等多模态条件编辑，局部和风格修改灵活，但对复杂结构编辑仍受限
3D 高斯点渲染	Kerbl 等人(2023) Chung 等人(2025) Zhao 等人(2025)	生成和渲染效率高，适用于实时及大规模场景合成，计算开销较低	在复杂场景中能保持较高视觉质量，但极细微结构和复杂反射效果可能略逊于 NeRF	编辑灵活性较强，支持较为灵活的场景调整，能够快速修改特定区域

置编码提升网络对高频细节的表达能力。在渲染阶段，沿相机光线采样多个点，输入神经网络获取颜色和密度信息，并通过体积渲染技术整合，生成对应视角的图像。训练过程中，优化目标为合成图像与真实图像之间的差异，采用可微分的渲染过程实现端到端的训练。此外，该方法引入了分层体积采样策略，提高了采样效率和渲染质量，在无需显式几何建模的情况下，实现了高质量的新视角图像合成。

NeRF 模型通过将三维场景表示为辐射场函数，实现了连续视角渲染和细节恢复，但其生成过程依赖于稠密视角数据，并且渲染速度较慢，难以实时应用。此外，NeRF 基于隐式表示的建模方法缺乏对语义控制的直接支持，限制了场景编辑和内容操控能力。扩散模型通过逐步扰动和还原数据分布，不仅具备强大的生成能力，还能够有效融合多模态信息(如文本和图像)，从而实现基于文本引导的场景编辑与生成。Lee 等人(2024)提出了一种针对真实户外场景的三维语义生成方法。该方法利用三平面表示，将三维场景数据压缩为 3 个正交的二维特征平面，以应对户外数据中常见的稀疏性问题。通过训练三平面自编码器，模型能够有效地从体素化场景中提取语义特征。在此基础上，作者构建了一个基于扩散模型的生成框架，对三平面特征进行逐步去噪，生成新的语义场景。此外，该方法引入了三平面操作机制，使得模型能够灵活地进行场景修复、扩展和语义补全等任务。现有方法缺乏对语义布局和对象间空间关系的有效建模，导致生成的场景在几何结构和纹理一致性上表现不足。针对此，Zhou 等人(2024)提出了一种基于布局引导的生成式高斯点渲染的文本到三维复杂场景生成方法，如图 4 所示。该方法首先利用大型语言模型(large language model，LLM)从文本描述中提取场景中的对象及其空间关系，生成初步的三维布局先验。随后，引入布局引导的三维高斯表示，将每个对象实例建模为一组各向异性的高斯分布，通过自适应几何控制模块，优化高斯的形状和空间分布，以确保场景中多对象之间的几何一致性和交互关系的准确性。此外，该方法采用组合优化机制，结合条件扩散模型，对布局先验进行迭代优化，使生成的三维场景在几何、纹理和尺度等方面与原始文本描述高度一致。

针对 NeRF 依赖密集采样和复杂的体积渲染计算，导致生成速度缓慢、渲染效率低的问题，以及扩散模型多基于二维图像或隐式表示，难以直接操控和编辑三维几何与空间关系、缺乏显式三维结构的局限，3D 高斯点渲染方法(Kerbl 等，2023)凭借显式场景表示与高效渲染特性，成为虚拟场景生成的重要补充技术。Chung 等人(2025)提出了能够从文本、RGB 图像和 RGBD 图像等多种输入生成 3D 场景的 LucidDreamer 方法。该方法将 3D 高斯点渲染技术与大规模扩散模型相结合，首先生成基于初始图像与深度信息的点云数据，随后通过“梦境”(dreaming)与“对齐”(alignment)两个步骤的迭代交替，生成多视角一致性的图像并逐步融合至点云中，在处理多视角场景时表现出较高的可靠性与效率。Zhao 等人(2025)提出了从文本描述中生成高质量的 3D 场景的 DreamScap 方法。该方法利用 3D 高斯点渲

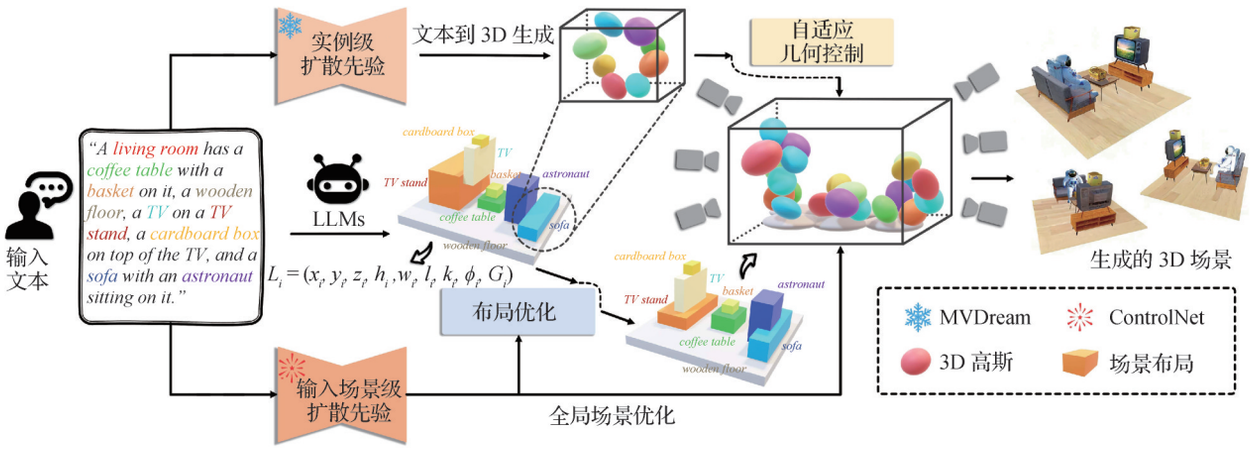

图4　基于布局引导的生成式高斯点渲染流程图(Zhou等,2024)

Fig. 4　Layout-guided generative Gaussian splatting pipeline (Zhou et al., 2024)

染作为3D表示,并引入3D高斯引导机制,通过大型语言模型从文本中编码语义原语、空间变换和场景相关性,实现从局部到全局的优化。在局部优化阶段,采用渐进式尺度控制技术,确保对象生成与整体场景尺度的一致性;在全局优化阶段,通过碰撞损失建模对象间的碰撞关系,减少LLM偏差,确保物理正确性。DreamScap不仅解决了三维生成中的多尺度问题,还进一步增强了场景的物理一致性和真实感,尤其在处理复杂的物理交互和细节方面,表现出优越的性能。

4.1.2　虚拟场景的风格属性变换

虚拟场景的风格属性变换旨在借助各类技术手段,对虚拟场景所呈现的视觉风格、艺术风格等属性实施转换操作,使其从初始的一种风格状态过渡到另一种风格状态。

最初基于卷积神经网络(convolutional neural network, CNN)的方法,通过对图像进行多层卷积操作,能够精准捕捉图像中的纹理、颜色以及其他细节特征。2016年,Gatys等人(2016)提出了一种基于卷积神经网络的图像风格迁移方法。该方法利用预训练的VGG-19(Visual Geometry Group-19)网络分别提取内容图像和风格图像的特征表示。内容特征通过网络中较高层的激活值表示,捕捉图像的整体结构信息;风格特征通过多个层的特征图的Gram矩阵表示,捕捉图像的纹理和颜色分布等统计特性。在优化过程中,以随机噪声图像为初始输入,通过最小化内容损失和风格损失的加权和,迭代更新生成图像,使其在保持内容图像结构的同时,呈现风格图像的艺术风格。该方法实现了内容与风格的有效分离与重组,生成了具有高感知质量的艺术风格图像。

生成对抗网络也常应用于图像风格转换中。生成对抗网络由生成器和判别器组成,二者通过不断对抗训练来提升性能。在风格迁移应用中,生成器负责生成具有特定风格的图像,判别器则判断生成图像是否符合目标风格以及是否足够真实。这种对抗机制使得GAN能够生成高质量的风格化图像。Zhu等人(2017)提出了一种无需成对数据的图像到图像转换方法CycleGAN。该方法引入了两个生成器($G: X\to Y$ 和 $F: Y\to X$)以及两个判别器,分别用于在源域X和目标域Y之间进行映射和判别。为了确保映射的可逆性和一致性,作者引入了循环一致性损失,即要求 $F(G(x))\approx x$ 和 $G(F(y))\approx y$。此外,采用对抗损失使生成的图像在目标域中难以被判别器区分,从而提高生成图像的真实性。通过联合优化对抗损失和循环一致性损失,CycleGAN能够在没有成对训练样本的情况下,实现高质量的图像风格迁移和域转换。

此外,在虚拟场景构建与风格变换过程中,三维重建技术主要用于获取场景的几何结构,为后续的风格化操作奠定基础。Höllein等人(2022)提出了一种针对室内三维重建场景的风格迁移方法。该方法通过对重建网格的显式纹理进行优化,实现了风格一致且无视角依赖的三维风格化效果。具体而言,利用深度和角度感知的优化策略,结合表面法线和深度信息,确保风格化纹理在整个场景中保持一致性。此外,采用多视角图像共同优化纹理,避免了传统二维风格迁移在三维场景中出现的拉伸和失真问题。该方法无须在推理阶段使用神经网络,优化

后的纹理可直接用于传统渲染管线，实现实时渲染。

4.1.3　多源数据驱动的模型训练与评估

在平行图像领域，多源数据驱动的模型训练与评估旨在利用从不同途径获取的、包含各类信息的数据，对相关模型进行训练与评估。

在实际视觉计算与测试领域，由于现实数据获取难度大、标注成本高、模型泛化能力有限，传统基于单一数据源的训练方法难以有效覆盖多样化场景和复杂环境。王坤峰等人（2016b）提出了一种基于ACP理论的智能视觉计算方法。该方法通过构建可控、可观、可重复的人工场景，模拟复杂的现实视觉环境，生成带有精确标注的虚拟图像数据。在此基础上，利用计算实验对视觉模型进行训练和评估，分析模型在复杂环境下的表现。最后，通过平行执行，将优化后的模型部署到实际系统中，实现虚实互动的闭环优化。该方法有效解决了现实数据获取困难、标注成本高以及模型泛化能力不足等问题，提升了视觉系统在复杂环境下的智能感知与理解能力。此外，王坤峰等人（2017c）还提出了一种名为“平行图像”的图像生成理论框架，旨在解决传统图像获取方法在数据规模、标注成本和多样性方面的限制。该方法的核心是软件定义的人工图像系统，通过从现实场景中获取少量图像“小数据”，输入人工图像系统，生成大量多样化的人工图像数据。具体实现包括图形渲染、图像风格迁移和生成式模型等技术手段。此外，框架引入了预测学习和集成学习的思想，分别用于模拟现实场景的动态变化和探索图像数据的空间分布，从而提高生成图像的真实性和多样性。最终，人工图像数据与真实图像数据相结合，构成平行图像“大数据”集合，用于视觉模型的训练与评估，实现虚实互动的视觉系统优化。

在自动驾驶测试面临场景复杂、突发情况多以及覆盖率不足的背景下，Li等人（2019a）提出了基于ACP理论的闭环平行测试框架，如图5所示。该方法首先构建语义任务空间，对驾驶任务进行抽象建模，定义原子任务并明确时空范围与复杂度。随后在虚拟环境中生成多样化测试场景，并与实车数据交互，形成虚实融合测试体系。通过融合虚实数据并借助计算实验，系统持续优化测试任务，提升覆盖率和挑战性。测试结果反馈至模型优化，增强车辆在复杂环境中的性能。该框架通过虚实整合显著提高了测试全面性与效率，克服了传统方法在效率、安全性和场景多样性方面的局限，推动了智能驾驶技术的发展。

4.2　多视域特征融合与虚实模型迁移

4.2.1　多任务的特征融合与语义信息交互

多任务的特征融合与信息交互旨在解决如何高效整合不同视角、条件或时间点获取的平行图像所包含的特征，同时推动这些图像在多个相关任务间达成语义信息的交互。

基于像素或帧序列的压缩编码方法往往难以有效捕捉和传输语义信息，导致带宽利用效率低且解码后图像缺乏语义一致性。为应对此类问题，Huang等人（2023）提出了基于深度强化学习的图像语义编码方法，旨在实现高效的图像语义通信。该方法首先定义图像的语义概念，包括类别、空间布局和视觉特征，并通过卷积神经网络提取这些语义信息。在编码阶段，采用强化学习策略进行语义比特分配，根据语义重要性自适应地分配量化级别，以优化传输率、语义保真度和感知质量之间的平衡。在解码阶段，利用生成对抗网络结合注意力机制，融合局部和全局特征，重建具有高感知质量和语义一致性的图像。

实现感知友好的视觉表达和精确的语义理解，是多模态图像融合的两大基本目标。然而，现有大多数方法主要关注融合图像的视觉质量，较少考虑其对下游高级视觉任务的支撑作用。同时，在视觉感知与语义理解任务中，如何提取和选择合适的特征仍然存在困难，现有多模态数据集在规模、标注质量和任务适配性等方面也存在一定不足。为解决这些问题，Liu等人（2023）提出了一种名为SegMiF的多交互特征学习框架，用于图像融合与语义分割的联合优化。该框架采用级联结构，包含融合子网络与分割子网络，并通过分层交互注意力模块实现模态特征与语义特征的深度融合。此外，引入动态权重因子，自动调整各任务的损失权重，平衡融合与分割任务间的优化目标。为支持该方法，作者构建了一个全时多模态基准数据集（full-time multi-modality benchmark，FMB），包含1 500对配准良好的红外与可见光图像，以及15类像素级注释，涵盖多种复杂环境。实验结果表明，SegMiF在融合图像质量与分割精度方面均优于现有方法，平均提升分割mIoU（mean intersection over union）达7.66%。

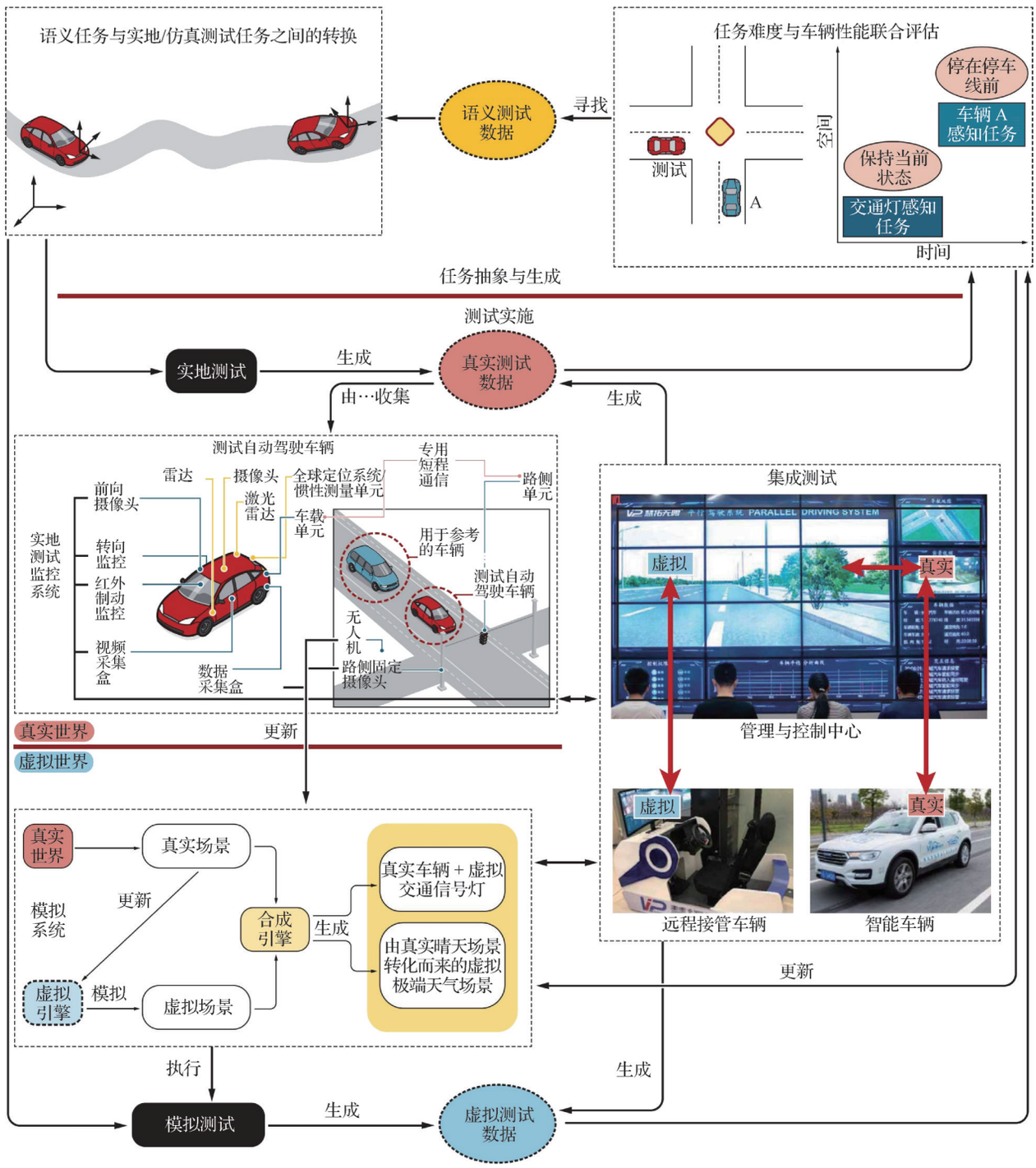

图 5　基于虚实交互的智能车辆平行测试框架示意图(Li 等,2019a)

Fig. 5　Schematic of a virtual-real interaction based parallel testing framework for intelligent vehicles (Li et al., 2019a)

4.2.2　多模态协同优化的虚实领域适应

多模态协同优化的虚实领域适应着力于解决如何对来自真实场景的平行图像数据与虚拟场景构建的数据进行协同处理与优化,使模型能够有效适应虚实领域之间存在的差异,包括但不限于图像特征、场景语义等方面的不同。

传统方法往往难以有效地从不同层次粒度提取和转移知识,无法很好地适应虚实领域之间在图像特征、场景语义等方面的差异。针对此问题,Min 等人(2020)提出了一种名为双粒度语义保持网络(bi-granularity semantic preserving network, BigSPN)的新型迁移学习方法,旨在解决在仅有基本层级标注的情况下识别细粒度子类别的问题。该方法通过构建两个专用的视觉编码器,分别处理基本类别和子类

别的视觉特征，并利用共享的语义解释器实现语义对齐。由于子类别缺乏图像标注，该方法引入了子级熵损失函数，通过多实例优化策略，使子类别的视觉表示与其语义嵌入对齐。最终，子类别的识别被转化为基于部件的视觉表示与语义嵌入之间的最近邻搜索问题。文章通过构建 CUB-HGTL、AWA2-HGTL 和 Flower-HGT 具有层级粒度的基准数据集，对所提出的 BigSPN 方法进行了系统性实验验证。结果表明，BigSPN 在细粒度子类别识别任务中显著优于现有方法，在 3 个数据集上的识别准确率分别提升 5.8%、6.8% 和 4.5%，充分证明了该方法在跨层级粒度迁移学习任务中的有效性和泛化能力。Zhou 等人（2022）提出了一种统一的多粒度对齐框架，旨在通过像素级、实例级和类别级的特征对齐，如图 6 所示，实现跨域目标检测中的域不变特征学习。该方法引入了全尺度门控融合模块，利用尺度感知卷积融合多尺度特征，增强实例级的判别能力。同时，设计了多粒度判别器，从不同粒度上区分源域和目标域的特征分布，确保特征对齐的全面性。此外，该方法提出了自适应指数移动平均策略，通过模型评估优化伪标签质量，缓解局部对齐误差，提高检测器的鲁棒性。该框架在 FCOS（fully convolutional one-stage object detection）和 Faster R-CNN 等检测器上进行了大量的实验，充分证明了其在多种跨域场景下的有效性。在天气变化适应任务中，基于 FCOS 检测器的方法相较于当前代表性的域适应方法在平均精度均值（mean average precision，mAP）指标上提升了 3.6%；在跨场景适应任务中，基于 Faster R-CNN 检测器的方法进一步将 mAP 指标提升了 5.3%。

由于数据来源多样，如何有效利用源领域知识向遥感目标领域迁移，以及如何对不同模态数据进行融合，是实现模型在虚实领域灵活适应的关键问题。Huang（2024）提出了一种高效遥感图像文本检索方法。该方法通过引入多模态门控适配器（multimodal gated adapter，MGA）和自适应三元组损失函数，实现了任务约束、模态对齐和单模态均匀对齐的统一优化。MGA 模块采用共享权重的自注意力机制，在图像和文本编码器之间建立跨模态交互，模拟人脑中多模态信息处理的层次结构。自适应三元组损失函数通过动态调整正负样本对的权重，缓解了同模态嵌入过度聚集的问题，提升了检索性能。该方法在无需外部数据的情况下，在 RSITMD 和 RSICD 等遥感数据集上仅通过少量可调参数，实现了与全微调模型相当甚至更优的性能，表明其在遥感多模态检索任务中的高效性和实用性。

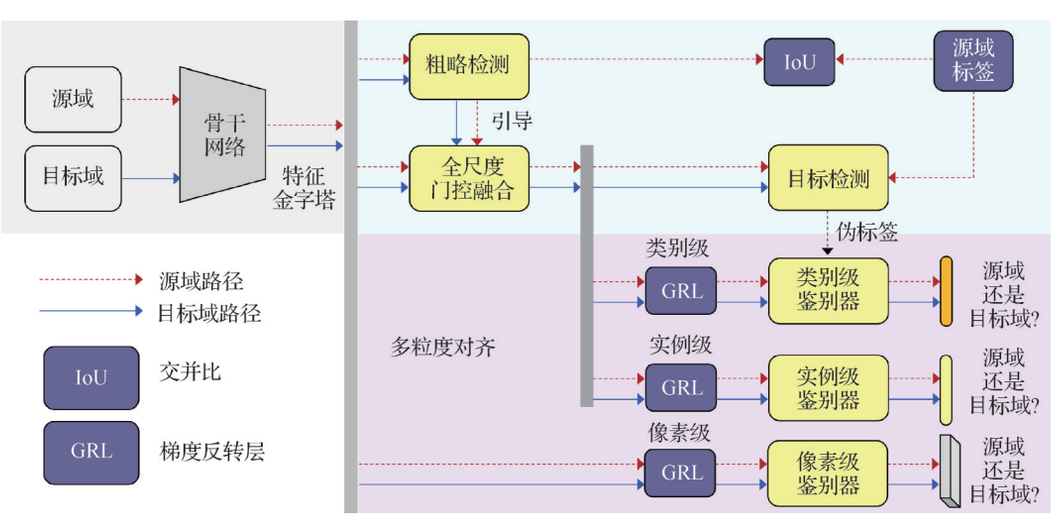

图 6　多粒度对齐的域适应框架（Zhou 等，2022）

Fig. 6　Multi-granularity alignment domain adaptation framework (Zhou et al., 2022)

4.3　异构数据与知识融合的平行推理

4.3.1　异构数据结构化信息提取与表征

在平行图像系统中，异构数据的结构化信息提取与表征是实现虚实交互与场景建模的关键环节。不同类型的数据（如点云、RGB 图像、深度图像和语义标签）在数据结构、分辨率和噪声特性上存在显著差异，如何有效融合并构建统一的表达框架，直接关系到系统对真实世界场景的全面感知与准确建模。

针对数据来源涉及多种模态，模态之间存在着数据结构和尺度的显著差异及数据缺失的问题，Hemker等人（2024）提出了一种用于整合异构生物医学数据的混合早期融合注意力学习网络（hybrid early-fusion attention learning network，HEALNet）。该方法结合了模态特异性结构信息和共享潜在空间中的跨模态交互，能够有效处理训练和推理过程中缺失的模态，并通过对原始数据的学习实现直观的模型解释。在TCGA（the cancer genome atlas）的4个癌症数据集上进行的多模态生存分析表明，HEALNet在性能上优于其他端到端训练的融合模型，显著提升了单模态和多模态基线模型的表现，并在缺失模态的情况下表现出较强的鲁棒性。

不同异构图像数据存在着数据特性差异大、难以有效融合和提取结构化信息的问题。传统方法无法充分挖掘这两种图像的互补信息，难以实现高质量的结构化信息提取与表征。Li等人（2025）提出了一种针对未配准红外—可见光图像融合的单阶段多模态融合监督框架该方法。该方法利用共享的浅层特征编码器提取红外和可见光图像的低层特征，并通过多头交叉注意力机制实现跨模态特征的对齐与融合。为了增强融合图像的结构和细节保留能力，该方法引入了多模态融合监督策略，结合结构保持损失和边缘保持损失，引导网络在保持源图像结构信息的同时，提升融合图像的视觉质量。此外，框架设计了多尺度特征融合模块，以适应不同分辨率下的图像融合需求。实验结果表明，该方法在多个未配准红外—可见光图像融合任务中表现出优越的性能，显著提升了融合图像的清晰度和细节保留能力。

4.3.2　多源数据融合驱动的视觉感知

多源数据融合驱动的视觉感知旨在充分利用异构传感器（如可见光、红外、深度、事件相机及语义标签等）所提供的互补信息，构建对真实场景更完整、更鲁棒的表征。通过在虚拟域和现实域中同步获取、对齐并融合多模态数据，可显著缓解单一数据源中的遮挡、噪声与视角局限问题，提升感知系统在复杂环境中的泛化能力与可解释性；同时，融合后的高维特征为后续任务（如目标检测、场景重建和风格迁移）提供了更丰富的先验和语义约束，为平行智能闭环优化奠定了坚实基础。

由于现实场景中数据存在长尾分布，部分类别的数据出现频率低，导致模型对这些类别感知能力欠佳。针对此类问题，Wang等人（2022）提出了一种基于平行视觉的长尾正则化方法，旨在解决自动驾驶中由于数据分布不均衡导致的长尾问题。该方法构建了平行视觉实现系统（parallel vision actualization system，PVAS），如图7所示，通过虚实交互和闭环优化，生成多样化的长尾场景数据，用于训练和评估自动驾驶模型。具体而言，PVAS利用虚拟环境模拟稀有或极端驾驶场景，并将生成的数据与真实世界的驾驶数据相结合，形成丰富的训练集。此外，系统引入了长尾正则化理论，通过调整模型的学习策略，使其在处理主流和长尾类别时表现更为均衡。实验结果表明，该方法在IVFC自动驾驶测试中有效提升了模型在长尾场景下的感知和决策能力，具有良好的泛化性能和实用价值。

图像数据模态丰富，但单纯依靠图像信息进行融合与理解存在局限，文本数据蕴含的语义信息未得到充分挖掘运用。Zhao等人（2024b）提出了一种图像融合新范式，首次引入显式的文本语义信息以指导多源图像的融合过程。该方法通过视觉语言模型（如BLIP2）从源图像中生成语义提示，并输入ChatGPT以获得详细的文本描述。这些描述在文本域中融合后，通过交叉注意力机制引导视觉特征的提取与融合，增强了特征提取和上下文理解能力。最终，融合后的视觉特征经解码器重建为高质量的融合图像。实验在红外—可见光、医学、多曝光和多焦点图像融合任务中取得了优异的性能，展示了该方法在多模态图像融合领域的广泛适用性。

4.3.3　多元知识引导的多任务平行推理

在复杂场景感知与认知任务中，单一任务和单一知识源往往难以全面捕捉多维度、多尺度的信息。多元知识引导的多任务平行推理方法通过整合来自不同领域和任务的异构知识，不仅能有效提升模型对多样化场景的理解与推理能力，还能在跨模态、多尺度信息的交互中实现全局信息与局部细节的统一表达。

在多模态学习领域，现有方法通常依赖于完整的多模态数据进行模型训练与推理。然而，在实际应用场景中，传感器故障、数据采集缺失或隐私保护需求等因素常常导致推理阶段的模态不完整性，从而限制了多模态模型的实际应用效果。传统的方法往往通过直接舍弃缺失模态或使用缺失模态填充策略进行推理，但这些方法无法充分挖掘已有模态中

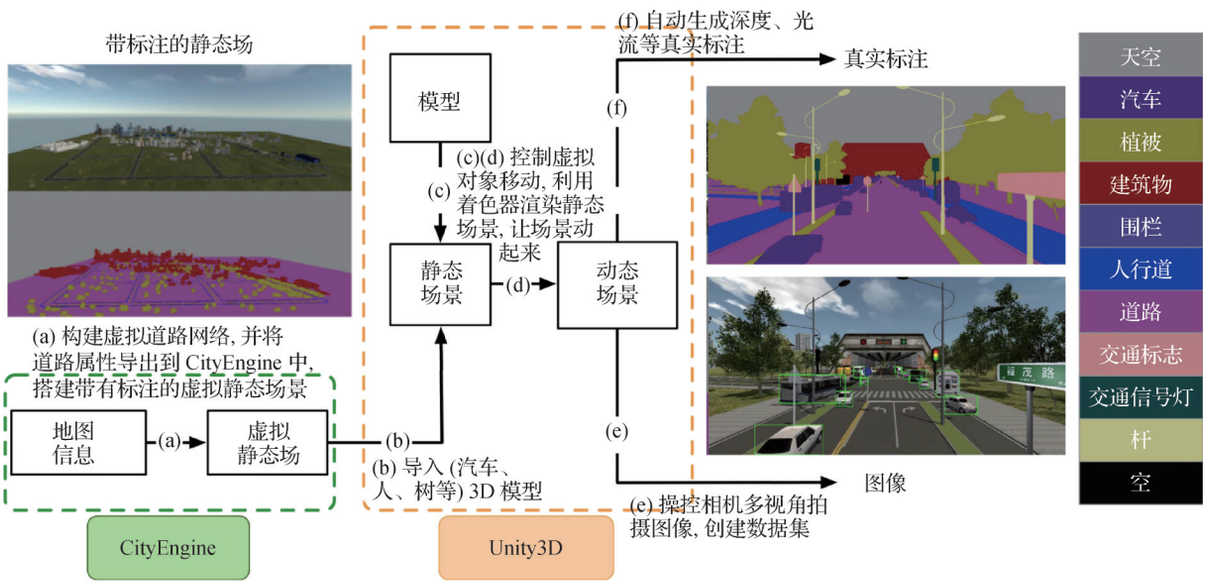

图 7　虚拟环境的构建流程及虚拟数据集的生成（Wang 等，2022）

Fig. 7　Construction process of virtual environments and generation of virtual datasets（Wang et al.，2022）

的判别性信息，导致模型性能下降。为此，Wei 等人（2024）提出了多模态幻觉框架，如图 8 所示，旨在解决训练阶段可获取多模态数据而推理阶段模态不完整的问题。该方法通过训练幻觉网络，使其在仅有部分模态输入的情况下，模拟完整多模态模型的表现，从而提升模型在实际应用中的鲁棒性。具体而言，作者引入了两种蒸馏策略：区域感知蒸馏和差异感知蒸馏。前者通过在多个区域建立并加权幻觉网络与多模态网络之间的响应映射，引导幻觉网络关注判别性区域；后者通过模仿多模态表示的局部样本间距离，使幻觉网络获得由多模态线索精化的类别判别能力。实验结果表明，该框架在多模态动作识别和人脸防伪等任务中，能够有效应对模态缺失问题，达到先进的性能水平。

随着视觉语言模型的发展，研究者们逐渐突破单一图像理解任务，开始探索多图像推理中的语义整合与跨图像关联分析。然而，现有的视觉语言模型大多针对单图像任务进行优化，对多图像输入间的关系建模缺乏有效机制，尤其是在多跳推理、多模态整合和复杂场景理解方面存在显著不足。此外，

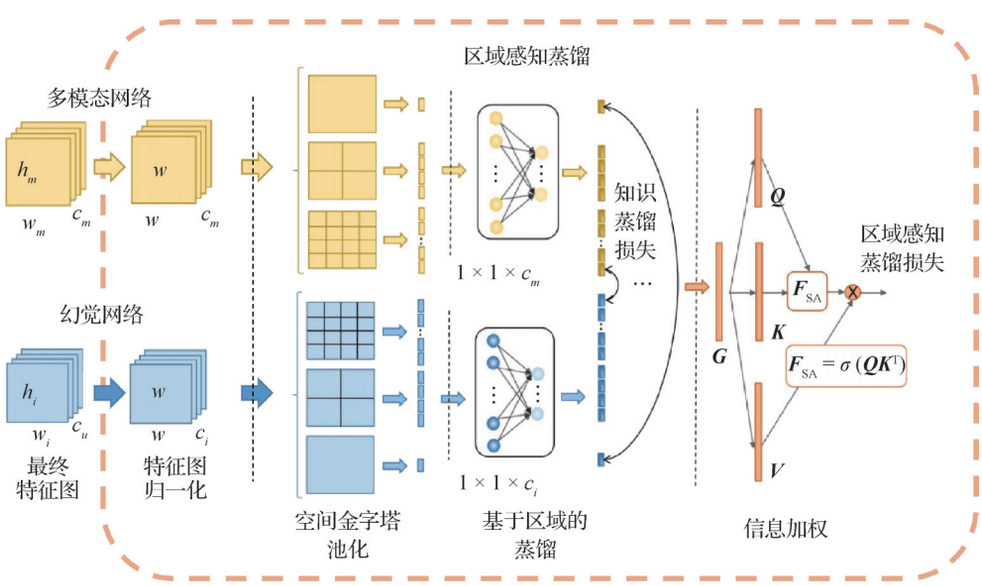

图 8　多模态幻觉框架（Wei 等，2024）

Fig. 8　Multimodal hallucination framework（Wei et al.，2024）

现有数据集主要聚焦于单图像问答或单视角物体检测，难以系统性评估模型在多图像语义关联与推理任务中的表现。基于此，Zhao等人（2024a）提出了多图像关系基准，旨在系统评估视觉语言模型（vision-language model，VLM）在多图像理解任务中的能力。该基准涵盖4个核心任务类型：感知、视觉世界知识、推理和多跳推理，全面测试模型在图像间比较、语义整合和跨图像推理等方面的表现。该方法通过构建多图像输入的问答任务，要求模型在复杂的图像组合中识别细节、推理关系并生成准确的自然语言回答。实验结果显示，尽管开源VLM在单图像任务中接近GPT-4V的性能，但在多图像推理任务上仍存在显著差距，甚至最先进的GPT-4V模型在该方法的测试中也面临挑战。

5 平行图像典型应用场景

平行图像作为平行智能理论在视觉计算领域的关键实践，已通过构建高逼真虚拟场景与合成数据有效支撑多个垂直领域中感知模型的训练、测试与优化。下文将分别以平行交通、平行医疗和平行矿山3个典型应用场景为代表，介绍平行图像的具体应用。

5.1 平行交通

为解决传统交通管理中系统级调控能力不足、交通数据采集成本高且场景覆盖有限，以及不同自动化水平车辆协同难度大等问题，平行交通系统（Wang，2010；吕宜生 等，2019）基于ACP理论，为系统化研究交通复杂问题提供了新途径。该系统致力于从交通“小数据”（如固定检测器采集的局部信息）生成覆盖多场景的交通“大数据”，并进一步提取出具有指导意义的交通“知识”（如交通流演化规律和最优信号控制策略），从而推动交通管理向数据驱动、虚实协同和全局智能的方向发展。在交通视觉感知方面，平行视觉（Wang等，2016a）融合虚拟交通影像数据（Li等，2019b，c）和交通领域专业知识，为小样本及极端场景下的交通目标检测任务提供了可解释的感知框架，有效应对暴雨、浓雾等恶劣条件下的车辆与车道线识别挑战。

对智能交通系统而言，从“单一物理调控”到“人机协同调控”，再到“虚实平行调控”的转变，已成为其发展的必然路径。城市交通网络作为典型的复杂系统，包含道路、车辆、行人及交通设施等多种要素，平行交通系统为高效管理交通网络、实现多类型车辆协同及应对突发交通事件提供了可行解决方案。以滨州和青岛的实践为例（Zhu等，2020），该系统通过构建与真实交通系统对应的人工交通系统，将交通基础设施和参与者全面数字化与平行化，从而为交通管理部门提供可靠、可预测且高效的调控方案。人工交通系统通过计算实验模拟不同交通策略的效果（如信号配时优化和车辆路径引导），评估其对交通流与通行效率的影响，进而支撑实际决策。物理系统中的检测器实时采集数据并反馈至人工系统，以动态优化模型；交通管理者负责全程监督，在策略调整与应急处理中拥有最高决策优先级。目前，平行驾驶（Wang等，2017b）、平行路径规划（Chen等，2019）和平行测试（Li等，2019a）等典型应用已广泛落地，有效突破了传统管理中场景覆盖有限、验证成本高的瓶颈，为筛选最优协同方案、保障系统高效安全运行提供了关键支撑。

5.2 平行医疗

医疗体系作为典型的复杂系统，涵盖患者、医护人员、医疗设备及医疗机构等多重要素。平行医疗系统（王飞跃 等，2021a）为精准优化诊疗流程、实现多主体协同以及应对复杂疾病诊疗提供了可行的解决方案。以青岛大学附属医院的平行痛风诊疗系统（Wang，2020a）和北京天坛医院的“天坛智慧脑”系统（Wang等，2021c）为例，该系统通过构建与真实医疗系统相对应的人工医疗系统，实现了医疗资源与诊疗流程的全面数字化与平行化，从而为医疗管理部门及医护人员提供可靠、可预测且高效的诊疗支持。人工医疗系统借助计算实验模拟不同诊疗策略的效果，例如痛风治疗的方案优化和手术流程规划，评估其对诊疗准确性、患者康复效率及医疗资源利用的影响，并为实际临床决策提供依据。物理医疗系统中的物联网设备实时采集患者生理数据与诊疗记录，并反馈至人工系统以动态调整模型参数；医护人员负责全程监督诊疗过程，在方案调整与应急处置中具有最终决策权。目前，针对典型医疗场景和需求，已涌现出平行手术（Wang等，2017a）、平行影像分析（Shen等，2021）和平行慢性病管理（Wang等，2020c）等一系列应用案例，有效弥补了传统医疗中场景覆盖不足和方案验证成本高的局限，为优化诊疗流程与资源调配、提升系统整体效率与安全性提

供了重要支撑。

5.3 平行矿山

矿山场景具有环境复杂、危险系数高以及真实数据采集难度大等特点，传统测试与训练方式难以满足自动驾驶技术规模化应用需求。在此背景下，相关研究借助平行图像技术构建虚拟测试环境与合成数据集，为矿山自动驾驶系统的性能提升提供关键支撑。Ai 等人（2024）提出了基于平行智能理论的自动驾驶矿山车辆测试平台 PMWorld。该平台通过构建高保真虚拟矿山环境并深度融合数字孪生技术，有效整合物理系统与虚拟系统，实现了对矿山自动驾驶车辆的高效、闭环测试。PMWorld 不仅能够模拟人类驾驶员和周边参与者的动态行为，还支持硬件在环、道路在环、虚实交互测试及调度系统协同测试等多种先进功能，为提升矿山自动驾驶系统的安全性、可靠性和运行效率提供了全面的评估与优化支撑。Ai 等人（2025）提出了首个面向露天矿自动驾驶任务的合成数据集 PMScenes。该数据集通过模拟极端天气条件及特殊场景（如车辆碰撞、异常入侵等），为复杂环境下的感知任务提供了大规模、多样化的合成数据支持。

6 平行图像与生成式大模型融合趋势

近年来，生成式人工智能与大规模多模态基础模型的迅猛发展，正重构人工智能技术体系。其中，基于扩散模型、生成对抗网络和神经辐射场的图像生成方法，在图像质量、语义一致性及多模态控制能力方面取得显著突破。同时，以 CLIP（Radford 等，2021）、SAM（Kirillov 等，2023）、GPT-4V（Wu 等，2024）为代表的多模态大模型，具备跨模态对齐、跨任务迁移与通用语义建模能力，正在推动视觉系统向“通用智能感知”迈进。

然而，这类基础大模型在特定任务场景中仍面临多重限制：1）缺乏场景物理可控性与真实感一致性；2）训练数据来源不透明，导致在高安全敏感领域存在泛化与鲁棒性问题；3）模型“黑箱化”程度高，缺乏针对现实反馈的快速适应机制。因此，如何构建具备可控生成、真实驱动以及任务导向特性的视觉感知机制，成为当前 AI 系统向产业级、工程级迈进的关键瓶颈。

在此背景下，平行图像技术作为一种具备“建模—训练—反馈—优化”闭环机制的视觉建模框架，与生成式大模型的结合展现出显著的协同潜力。具体地，其融合路径主要体现在以下 3 个方面：

1）通过大模型驱动虚拟场景内容生成，提升图像语义与风格控制能力。通过引入文本驱动的扩散模型、预训练语言—视觉大模型（如 CLIP-guided NeRF），可实现从自然语言、高层语义图或行为描述中生成高仿真图像、视频乃至 3D 场景，用于平行图像系统的“人工场景构建”阶段，从而显著增强其多样性与可定制性。

2）使用平行图像技术提供结构化、多模态和标注完备的合成数据，辅助大模型微调与评估。当前大模型普遍存在特定任务（如小目标检测、极端天气感知等）中的表现不稳定问题。平行图像可合成覆盖稀有场景、多模态传感器视图的高质量训练数据，结合虚实交互机制，用于对大模型进行有针对性的增强训练或部署前模拟评估。

3）构建基于大模型的虚实反馈学习机制，实现跨场景适应与演化优化。通过平行执行机制，将大模型部署至实际环境后采集的运行表现，作为反馈信号传入虚拟世界，实现基于生成模型的场景重构、参数微调与知识反演。该过程构成“生成—模拟—反馈—优化”的闭环，使视觉系统具备实时演化、自主适应能力。

平行图像技术与生成式人工智能、大规模多模态模型的融合，不仅能够在数据生成层面提升任务特异性与控制性，在模型训练层面补足大模型的“语义盲区”，更在系统演化层面构建起支持持续学习与反馈优化的闭环机制。未来，随着具备推理能力的多模态模型不断成熟，平行图像有望从“感知辅助”走向“知识驱动”的认知智能阶段，成为构建通用视觉系统的关键支撑技术。

7 技术挑战与未来展望

7.1 虚拟数据高质量扩展

尽管虚拟数据具备成本低、可控性强和可大规模生成等优势，其在质量、真实性与语义一致性方面仍面临诸多技术挑战。首先，当前的虚拟数据生成系统在视觉精度和物理逼真度方面取得了一定进展，但在语义完整性、多样性和细节真实感方面仍存在不足。许多生成的场景缺乏对边缘场景、极端天

气条件以及稀有事件等长尾分布数据的有效覆盖，难以支撑模型对复杂现实环境的全面理解与泛化能力的构建。其次，不同虚拟平台所生成的数据存在标准不一、格式不统一的问题，使得数据在跨平台共享和统一训练中的兼容性大打折扣。此外，虚拟数据生成仍严重依赖人工设计的规则与模板，难以自适应地生成符合特定任务语义需求的个性化数据，制约了其在细粒度学习和多任务场景下的应用价值。

为了解决上述问题，实现虚拟数据的高质量扩展，需要从多个方向开展深入研究与技术创新。一方面，应推动智能生成模型的演进，构建面向任务的高保真虚拟数据生成体系。具体而言，将生成对抗网络的细节生成能力、扩散模型的全局一致性优势与多模态大模型的语义理解能力相结合，通过输入真实场景的多维度数据，自动挖掘现实场景中的关键语义特征，并在生成过程中完成空间结构、动态语义和交互逻辑的联合建模，确保建筑布局合理、车辆行驶轨迹连贯以及行人避让车辆的行为模式等符合现实逻辑，从而全面提升虚拟数据与现实场景的还原度。另一方面，需要发展自监督和弱监督的质量评估机制，对生成数据的语义一致性、行为合理性以及与真实数据的相似性进行智能评估与筛选，提升数据质量控制的自动化水平。例如，可以基于真实数据构建语义一致性基准库，通过自监督学习让模型自动比对虚拟数据与基准库之间的语义偏差。同时，可引入弱监督信号，通过人工标注少量不合格虚拟数据样本，训练模型识别行为合理性问题，最终实现对生成数据的自动化筛选。

此外，未来还需构建开放、统一的虚拟数据扩展标准与共享框架，打通多源平台之间的数据壁垒，促进模型训练数据的标准化、高效化与可持续更新。最终，通过高质量虚拟数据与现实数据的协同演进，有望显著提升智能模型在复杂现实环境下的稳定性、可靠性与泛化能力，助力虚实协同智能系统迈向更高层次的发展阶段。

7.2 虚实跨域语义鸿沟

虚实跨域鸿沟广泛存在于自动驾驶、计算机视觉和机器人等领域，给相关技术的发展和应用带来了诸多挑战。这一鸿沟本质上源于虚拟与现实环境在诸多层面的显著差异，这些差异深刻影响着算法的有效性、模型的泛化能力以及系统的整体性能和适应性。以自动驾驶领域为例，借助虚拟环境，研发人员可以在相对安全、低成本的条件下对自动驾驶算法进行大量的实验和优化。然而，在虚拟环境中，道路标识的设计、车辆和行人的行为模式、光照和天气条件等因素，虽然在一定程度上模拟了现实，但在语义细节方面，与真实的交通场景存在着较大的偏差。这些差异使得基于虚拟数据训练的自动驾驶模型在面对现实交通场景时，难以准确理解和应对各种复杂情况，容易出现误判和错误决策，从而严重威胁到道路交通安全。

为了有效缓解虚实跨域语义鸿沟，需要在多个方面进行深入的研究和创新。首先，需重点发展融合因果推理机制的虚拟数据生成方法。这可以通过显式建模现实交通场景中的因果关系，提升虚拟环境中事件与行为的语义一致性和合理性。具体而言，可以探索基于因果推理的模型训练框架，结合虚拟数据生成与实际交通场景的因果规则，制定出可以有效减少虚实差异的方法。其次，应建立跨域语义对齐的评估体系，提出可量化的虚实差异指标。这些指标能够实现对虚拟数据和现实数据语义一致性的系统化验证，进而为虚拟环境与现实环境之间的语义对接提供可行的技术路径。

此外，应积极推进相关技术的产业落地路径规划，可与汽车制造商和交通管理部门等关键主体建立合作，共同推动自动驾驶仿真测试标准建设，促进虚拟训练模型与实际场景需求的有效对接，从而确保技术成果在真实应用中的可靠性、安全性及可推广性。

7.3 虚实协同实时交互

在实时性要求高的场景下，如沉浸式虚拟手术培训、实时体育赛事虚实互动直播，网络延迟、计算资源限制等问题导致虚实交互的流畅性与响应及时性难以保证。在虚拟手术培训中，医生操作虚拟器械与虚拟组织的交互反馈若存在延迟，将严重影响培训效果与真实感。同时，当前虚实协同交互的感知与反馈机制不够完善，例如触觉反馈在虚拟环境中难以精确模拟真实手术中器械与组织接触的复杂触感，视觉与触觉等多感知通道之间的协同也不够自然，降低了用户体验与交互效率。然而，随着 5G 等新一代通信技术的普及，网络延迟将大幅降低，为实时交互提供高速稳定的传输保障。同时，边缘计算、云计算技术的发展将缓解本地计算资源压力，提

升数据处理速度，实现虚实协同的实时渲染与交互响应。新型传感器技术的不断进步，为构建更为先进的多模态感知技术提供了坚实基础。尤其在触觉、力觉及温觉等关键感知领域，通过深度融合深度学习与多源传感器数据，能够显著优化触觉反馈的精确度与响应实时性，有效提升多感知通道之间的协同一致性。这一技术路径将有力完善虚实协同交互中的感知体系，最终实现更加自然、真实和沉浸式的交互体验。这将推动平行图像技术在医疗、娱乐以及工业制造等领域的实时交互应用实现质的飞跃，创造出更加沉浸式、高效的虚实协同体验。

8 结　语

平行图像作为融合人工建模、计算实验与虚实协同执行的综合性视觉数据生成与建模范式，正在成为智能视觉系统发展中的关键支撑技术。其基于平行系统理论构建的“建模—训练—反馈—优化”闭环机制，突破了传统数据依赖模式的限制，有效缓解了现实世界中数据稀缺、标注困难及样本分布不均衡等问题。本文系统回顾了平行图像的理论起点与技术路径，深入分析了其在虚拟场景生成、多模态融合、跨域迁移和知识驱动推理等方面的最新研究进展，并探讨了其与生成式人工智能及多模态基础模型的融合潜力。

尽管已有研究取得诸多成果，平行图像的发展仍面临多方面挑战，包括生成质量与真实感的持续提升、异构数据的高效融合机制、虚实协同运行的稳定性保障以及标准化平台与开放生态的建设等。未来，随着大模型、强化学习以及因果推理等技术的进一步发展，平行图像有望实现更加智能化、自适应和泛化的进化路径，赋能更多复杂多变的视觉任务场景，推动智能视觉系统向更高层次、更广领域持续拓展。

参考文献（References）

Ai Y F, Li X Q, Song R Q, Cui C L, Tian B, Chen L, et al. 2025. PMScenes: a parallel mine dataset for autonomous driving in surface mines. IEEE Intelligent Transportation Systems Magazine, 17(1): 30-44 [DOI: 10.1109/MITS.2024.3454275]

Ai Y F, Liu Y H, Gao Y, Zhao C, Cheng X, Han J P, et al. 2024. PMWorld: a parallel testing platform for autonomous driving in mines. IEEE Transactions on Intelligent Vehicles, 9(1): 1402-1411 [DOI: 10.1109/TIV.2023.3332739]

Chen L, Hu X M, Tian W, Wang H, Cao D P and Wang F Y. 2019. Parallel planning: a new motion planning framework for autonomous driving. IEEE/CAA Journal of Automatica Sinica, 6(1): 236-246 [DOI: 10.1109/JAS.2018.7511186]

Chung J, Lee S, Nam H, Lee J and Lee K M. 2025. LucidDreamer: domain-free generation of 3D Gaussian splatting scenes. IEEE Transactions on Visualization and Computer Graphics, 31(12): 10640-10651 [DOI: 10.1109/TVCG.2025.3611489]

Dai P, Tan F T, Yu X, Peng Y F, Zhang Y D and Qi X J. 2025. GO-NeRF: generating objects in neural radiance fields for virtual reality content creation. IEEE Transactions on Visualization and Computer Graphics, 31(5): 3087-3097 [DOI: 10.1109/TVCG.2025.3549558]

Duan W, Cao Z D, Wang Y Z, Zhu B, Zeng D, Wang F Y, et al. 2013. An ACP approach to public health emergency management: using a campus outbreak of H1N1 influenza as a case study. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 43(5): 1028-1041 [DOI: 10.1109/TSMC.2013.2256855]

Gatys L A, Ecker A S and Bethge M. 2016. Image style transfer using convolutional neural networks//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 2414-2423 [DOI: 10.1109/CVPR.2016.265]

Gou C, Liu X X, Guo Z P, Zhou Y C and Wang F Y. 2024. Enhanced risk perception method based on parallel vision for autonomous vehicles in safety-critical scenarios. Journal of Image and Graphics, 29(11): 3265-3279 (苟超, 刘欣欣, 郭子鹏, 周昱臣, 王飞跃. 2024. 无人驾驶突发紧要场景下基于平行视觉的风险增强感知方法. 中国图象图形学报, 29(11): 3265-3279) [DOI: 10.11834/jig.230748]

Hemker K, Simidjievski N and Jamnik M. 2024. HEALNet: multimodal fusion for heterogeneous biomedical data//Proceedings of the 38th Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates, Inc: 64479-64498 [DOI: 10.52202/079017-2057]

Höllein L, Johnson J and Nießner M. 2022. StyleMesh: style transfer for indoor 3D scene reconstructions//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 6188-6198 [DOI: 10.1109/CVPR52688.2022.00610]

Huang D L, Gao F F, Tao X M, Du Q Y and Lu J H. 2023. Toward semantic communications: deep learning-based image semantic coding. IEEE Journal on Selected Areas in Communications, 41(1): 55-71 [DOI: 10.1109/JSAC.2022.3221999]

Huang T J. 2024. Efficient remote sensing with harmonized transfer learning and modality alignment [EB/OL]. [2025-05-28]. https://arxiv.org/pdf/2404.18253.pdf

Kerbl B, Kopanas G, Leimkühler T and Drettakis G. 2023. 3D Gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics, 42(4): #139 [DOI: 110.1145/3592433]

Kirillov A, Mintun E, Ravi N, Mao H Z, Rolland C, Gustafson L, et al. 2023. Segment anything//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 3992-4003 [DOI: 10.1109/ICCV51070.2023.00371]

Lee J, Lee S, Jo C, Im W, Seon J and Yoon S E. 2024. SemCity: semantic scene generation with Triplane diffusion//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 28337-28347 [DOI: 10.1109/CVPR52733.2024.02677]

Li H F, Yang Z Y, Zhang Y F, Jia W, Yu Z T and Liu Y. 2025. MulFS-CAP: multimodal fusion-supervised cross-modality alignment perception for unregistered infrared-visible image fusion. IEEE Transactions on Pattern Analysis and Machine Intelligence, 47(5): 3673-3690 [DOI: 10.1109/TPAMI.2025.3535617]

Li L, Lin Y L, Cao D P, Zheng N N and Wang F Y. 2017. Parallel learning — A new framework for machine learning. Acta Automatica Sinica, 43(1): 1-8 (李力, 林懿伦, 曹东璞, 郑南宁, 王飞跃. 2017. 平行学习——机器学习的一个新型理论框架. 自动化学报, 43(1): 1-8) [DOI: 10.16383/j.aas.2017.y000001]

Li L, Song J Y, Wang F Y, Niehsen W and Zheng N N. 2005. IVS 05: new developments and research trends for intelligent vehicles. IEEE Intelligent Systems, 20(4): 10-14 [DOI: 10.1109/MIS.2005.73]

Li L, Wang X, Wang K F, Lin Y L, Xin J M, Chen L, et al. 2019a. Parallel testing of vehicle intelligence via virtual-real interaction. Science Robotics, 4(28): #4106 [DOI: 10.1126/scirobotics.aaw4106]

Li X and Wang F Y. 2021. Parallel visual perception for intelligent driving: basic concept, framework and application. Journal of Image and Graphics, 26(1): 67-81 (李轩, 王飞跃. 2021. 面向智能驾驶的平行视觉感知: 基本概念、框架与应用. 中国图象图形学报, 26(1): 67-81) [DOI: 10.11834/jig.200402]

Li X, Wang K F, Tian Y L, Yan L, Deng F and Wang F Y. 2019b. The ParallelEye dataset: a large collection of virtual images for traffic vision research. IEEE Transactions on Intelligent Transportation Systems, 20(6): 2072-2084 [DOI: 10.1109/TITS.2018.2857566]

Li X, Wang Y T, Yan L, Wang K F, Deng F and Wang F Y. 2019c. ParallelEye-CS: a new dataset of synthetic images for testing the visual intelligence of intelligent vehicles. IEEE Transactions on Vehicular Technology, 68(10): 9619-9631 [DOI: 10.1109/TVT.2019.2936227]

Liu J Y, Liu Z, Wu G Y, Ma L, Liu R S, Zhong W, et al. 2023. Multi-interactive feature learning and a full-time multi-modality benchmark for image fusion and segmentation//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 8081-8090 [DOI: 10.1109/ICCV51070.2023.00745]

Liu Y H, Shen Y, Fan L L, Tian Y L, Ai Y F, Tian B, et al. 2022. Parallel radars: from digital twins to digital intelligence for smart radar systems. Sensors, 22(24): #9930 [DOI: 10.3390/s22249930]

Lyu Y S, Chen Y Y, Jin J C, Li Z J, Ye P J and Zhu F H. 2019. Parallel transportation: virtual-real interaction for intelligent traffic management and control. Chinese Journal of Intelligent Science and Technology, 1(1): 21-33 (吕宜生, 陈圆圆, 金峻臣, 李镇江, 叶佩军, 朱凤华. 2019. 平行交通: 虚实互动的智能交通管理与控制. 智能科学与技术学报, 1(1): 21-33) [DOI: 10.11959/j.issn.2096-6652.201908]

Mildenhall B, Srinivasan P P, Tancik M, Barron J T, Ramamoorthi R and Ng R. 2021. NeRF: representing scenes as neural radiance fields for view synthesis. Communications of the ACM, 65(1): 99-106 [DOI: 10.1145/3503250]

Min S B, Xie H T, Yao H T, Deng X R, Zha Z J and Zhang Y D. 2020. Hierarchical granularity transfer learning//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: 11650-11661

Radford A, Kim J W, Hallacy C, Ramesh A, Goh G, Agarwal S, et al. 2021. Learning transferable visual models from natural language supervision//Proceedings of the 38th International Conference on Machine Learning. Online: [s.n.]: 8748-8763

Rombach R, Blattmann A, Lorenz D, Esser P and Ommer B. 2022. High-resolution image synthesis with latent diffusion models//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 10674-10685 [DOI: 10.1109/CVPR52688.2022.01042]

Shen T Y, Gou C, Wang J G, Huang J, He Y L, Xue H D, et al. 2021. Parallel medical imaging: an ACP-based approach for intelligent medical image recognition with small samples//Proceedings of the 1st IEEE International Conference on Digital Twins and Parallel Intelligence. Beijing, China: IEEE: 226-229 [DOI: 10.1109/DTPI52967.2021.9540120]

Tian Y L, Shen Y, Li Q and Wang F Y. 2020. Parallel point clouds: point clouds generation and 3D model evolution via virtual-real interaction. Acta Automatica Sinica, 46(12): 2572-2582 (田永林, 沈宇, 李强, 王飞跃. 2020. 平行点云: 虚实互动的点云生成与三维模型进化方法. 自动化学报, 46(12): 2572-2582) [DOI: 10.16383/j.aas.c200800]

Wang F Y. 1994. Shadow Systems: A New Concept for Nested and Embedded Co-Simulation for Intelligent Systems. University of Arizona

Wang F Y. 2004. Parallel system methods for management and control of complex systems. Control and Decision, 19(5): 485-489, 514 (王飞跃. 2004. 平行系统方法与复杂系统的管理和控制. 控制与决策, 19(5): 485-489, 514) [DOI: 10.3321/j.issn:1001-0920.2004.05.002]

Wang F Y. 2007. Toward a paradigm shift in social computing: the ACP approach. IEEE Intelligent Systems, 22(5): 65-67 [DOI: 10.

1109/MIS.2007.4338496］

Wang F Y. 2010. Parallel control and management for intelligent transportation systems: concepts, architectures, and applications. IEEE Transactions on Intelligent Transportation Systems, 11(3): 630-638［DOI: 10.1109/TITS.2010.2060218］

Wang F Y. 2020a. Parallel healthcare: robotic medical and health process automation for secured and smart social healthcares. IEEE Transactions on Computational Social Systems, 7(3): 581-586［DOI: 10.1109/TCSS.2020.2995282］

Wang F Y. 2020b. Parallel intelligence: belief and prescription for edge emergence and cloud convergence in CPSS. IEEE Transactions on Computational Social Systems, 7(5): 1105-1110［DOI: 10.1109/TCSS.2020.3029855］

Wang F Y. 2021a. Parallel medicine: from warmness of medicare to medicine of smartness. Chinese Journal of Intelligent Science and Technology, 3(1): 1-9 (王飞跃. 2021a. 平行医学: 从医学的温度到智慧的医学. 智能科学与技术学报, 3(1): 1-9)［DOI: 10.11959/j.issn.2096-6652.202101］

Wang F Y. 2023. Forward to the past: CASTLab's cyber-social-physical approach for ITS in 1999［history and perspectives］. IEEE Intelligent Transportation Systems Magazine, 15(4): 171-175［DOI: 10.1109/MITS.2023.3278158］

Wang F Y, Meng X B, Du S C and Geng Z. 2021b. Parallel light field: the framework and processes. Chinese Journal of Intelligent Science and Technology, 3(1): 110-122 (王飞跃, 孟祥冰, 杜思聪, 耿征. 2021. 平行光场: 基本框架与流程. 智能科学与技术学报, 3(1): 110-122)［DOI: 10.11959/j.issn.2096-6652.202112］

Wang F Y, Yang L Q, Yang J, Zhang Y L, Han S S and Zhao K. 2016a. Urban intelligent parking system based on the parallel theory//Proceedings of 2016 International Conference on Computing, Networking and Communications. Kauai, USA: IEEE: 1-5［DOI: 10.1109/ICCNC.2016.7440708］

Wang F Y, Zhang M, Meng X B, Wang R, Wang X, Zhang Z C, et al. 2017a. Parallel surgery: an ACP-based approach for intelligent operations. Pattern Recognition and Artificial Intelligence, 30(11): 961-970 (王飞跃, 张梅, 孟祥冰, 王蓉, 王晓, 张志成, 等. 2017. 平行手术: 基于ACP的智能手术计算方法. 模式识别与人工智能, 30(11): 961-970)［DOI: 10.16451/j.cnki.issn1003-6059.201711001］

Wang F Y, Zheng N N, Cao D P, Martinez C M, Li L and Liu T. 2017b. Parallel driving in CPSS: a unified approach for transport automation and vehicle intelligence. IEEE/CAA Journal of Automatica Sinica, 4(4): 577-587［DOI: 10.1109/JAS.2017.7510598］

Wang J, Wang X, Guo Y Y and Wang F Y. 2020c. A parallel medical diagnostic and treatment system for chronic diseases//2020 Chinese Automation Congress. Shanghai, China: IEEE: 7412-7416［DOI: 10.1109/CAC51589.2020.9327701］

Wang J G, Wang X, Shen T Y, Wang Y T, Li L, Tian Y L, et al. 2022. Parallel vision for long-tail regularization: initial results from IVFC autonomous driving testing. IEEE Transactions on Intelligent Vehicles, 7(2): 286-299［DOI: 10.1109/TIV.2022.3145035］

Wang K F, Gou C and Wang F Y. 2016b. Parallel vision: an ACP-based approach to intelligent vision computing. Acta Automatica Sinica, 42(10): 1490-1500 (王坤峰, 苟超, 王飞跃. 2016b. 平行视觉: 基于ACP的智能视觉计算方法. 自动化学报, 42(10): 1490-1500)［DOI: 10.16383/j.aas.2016.c160604］

Wang K F, Lu Y, Wang Y T, Xiong Z W and Wang F Y. 2017c. Parallel imaging: a new theoretical framework for image generation. Pattern Recognition and Artificial Intelligence, 30(7): 577-587 (王坤峰, 鲁越, 王雨桐, 熊子威, 王飞跃. 2017c. 平行图像: 图像生成的一个新型理论框架. 模式识别与人工智能, 30(7): 577-587)［DOI: 10.16451/j.cnki.issn1003-6059.201707001］

Wang X X, Yang J, Liu Y H, Wang Y T, Wang F Y, Kang M Z, et al. 2024. Parallel intelligence in three decades: a historical review and future perspective on ACP and cyber-physical-social systems. Artificial Intelligence Review, 57(9): #255［DOI: 10.1007/s10462-024-10861-9］

Wang Y J, Wang F Y, Wang X, Guan Z J, Ouyang L W, Wang J G, et al. 2021c. Parallel hospital: ACP-based hospital smart operating system//Proceedings of the 1st IEEE International Conference on Digital Twins and Parallel Intelligence. Beijing, China: IEEE: 474-477［DOI: 10.1109/DTPI52967.2021.9540207］

Wei S C, Luo C B, Luo Y and Xu J L. 2024. Privileged modality learning via multimodal hallucination. IEEE Transactions on Multimedia, 26: 1516-1527［DOI: 10.1109/TMM.2023.3282874］

Wu T, Yang G D, Li Z B, Zhang K, Liu Z W, Guibas L, et al. 2024. GPT-4V(ision) is a human-aligned evaluator for text-to-3D generation//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 22227-22238［DOI: 10.1109/CVPR52733.2024.02098］

Xu Y W, Zhao Y, Xiao Z S and Hou T B. 2024. UFOGen: you forward once large scale text-to-image generation via diffusion GANs//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 8196-8206［DOI: 10.1109/CVPR52733.2024.00783］

Yang J, Wang X, Tian Y L, Wang X and Wang F Y. 2023a. Parallel intelligence in CPSSs: being, becoming, and believing. IEEE Intelligent Systems, 38(6): 75-80［DOI: 10.1109/MIS.2023.3284694］

Yang J, Wang X, Wang Y T, Liu Z M, Li X S and Wang F Y. 2023b. Parallel intelligence and CPSS in 30 years: an ACP approach. Acta Automatica Sinica, 49(3): 614-634 (杨静, 王晓, 王雨桐, 刘忠民, 李小双, 王飞跃. 2023. 平行智能与CPSS: 三十年发展的回顾与展望. 自动化学报, 49(3): 614-634)［DOI: 10.16383/j.aas.c230015］

Yu H K, Liu X Y, Tian Y L, Wang Y T, Gou C and Wang F Y. 2024. Retracted: Sora-based parallel vision for smart sensing of intelligent vehicles: from foundation models to foundation intelligence.

IEEE Transactions on Intelligent Vehicles, 9(2): 3123-3126 [DOI: 10.1109/TIV.2024.3376575]

Zhang H, Li X and Wang F Y. 2021. The basic framework and key algorithms of parallel vision. Journal of Image and Graphics, 26(1): 82-92 (张慧, 李轩, 王飞跃. 2021. 平行视觉的基本框架与关键算法. 中国图象图形学报, 26(1): 82-92) [DOI: 10.11834/jig.200400]

Zhang H, Luo G Y, Li Y D and Wang F Y. 2023. Parallel vision for intelligent transportation systems in metaverse: challenges, solutions, and potential applications. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 53(6): 3400-3413 [DOI: 10.1109/TSMC.2022.3228314]

Zhang J B, Li X Y, Wan Z Y, Wang C and Liao J. 2024. Text2NeRF: text-driven 3D scene generation with neural radiance fields. IEEE Transactions on Visualization and Computer Graphics, 30(12): 7749-7762 [DOI: 10.1109/TVCG.2024.336150]

Zhang W W, Wang K F, Liu Y T, Lu Y and Wang F Y. 2020. A parallel vision approach to scene-specific pedestrian detection. Neurocomputing, 394: 114-126 [DOI: 10.1016/j.neucom.2019.03.095]

Zhao B C, Zong Y S, Zhang L T and Hospedales T. 2024a. Benchmarking multi-image understanding in vision and language models: perception, knowledge, reasoning, and multi-hop reasoning [EB/OL]. [2025-05-28]. https://arxiv.org/pdf/2406.12742.pdf

Zhao Y M, Yuan X N, Yang H Y and Huang D. 2025. DreamScape: 3D scene creation via Gaussian splatting joint correlation modeling// Proceedings of 2025 IEEE International Conference on Multimedia and Expo. Nantes, France: IEEE: 1-6 [DOI: 10.1109/ICME59968.2025.11210004]

Zhao Z X, Deng L L, Bai H W, Cui Y K, Zhang Z P, Zhang Y L, et al. 2024b. Image fusion via vision-language model//Proceedings of the 41st International Conference on Machine Learning. Vienna, Austria: JMLR.org: 60749-60765

Zhou W Z, Du D W, Zhang L B, Luo T J and Wu Y J. 2022. Multi-granularity alignment domain adaptation for object detection//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 9571-9580 [DOI: 10.1109/CVPR52688.2022.00936]

Zhou X Y, Ran X J, Xiong Y J, He J L, Lin Z W, Wang Y T, et al. 2024. GALA3D: towards text-to-3D complex scene generation via layout-guided generative Gaussian splatting//Proceedings of the 41st International Conference on Machine Learning. Vienna, Austria: JMLR.org: 62108-62118 [DOI: 10.5555/3692070.3694640]

Zhu F H, Lyu Y S, Chen Y Y, Wang X, Xiong G and Wang F Y. 2020. Parallel transportation systems: toward IoT-enabled smart urban traffic control and management. IEEE Transactions on Intelligent Transportation Systems, 21(10): 4063-4071 [DOI: 10.1109/TITS.2019.2934991]

Zhu F H, Wang F Y, Li R M, Lyu Y S and Chen S H. 2011. Modeling and analyzing transportation systems based on ACP approach//Proceedings of the 14th International IEEE Conference on Intelligent Transportation Systems. Washington, USA: IEEE: 2136-2141 [DOI: 10.1109/ITSC.2011.6083046]

Zhu J Y, Park T, Isola P and Efros A A. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 2242-2251 [DOI: 10.1109/ICCV.2017.244]

作者简介

张慧,女,副教授,主要研究方向为多智能体协同、多模态感知、具身智能和平行智能。E-mail:huizhang1@bjtu.edu.cn

王飞跃,通信作者,男,研究员,主要研究方向为智能系统和复杂系统的建模、分析与控制。E-mail:feiyue.wang@ia.ac.cn

田永林,男,助理研究员,主要研究方向为平行智能、自动驾驶和智能交通系统。E-mail: yonglin.tian@ia.ac.cn

王雨桐,女,副研究员,主要研究方向为计算机视觉和智能感知。E-mail:yutong.wang@ia.ac.cn

苟超,男,副教授,主要研究方向为模式识别与智能系统、人工智能、智能驾驶、计算机视觉。

E-mail:gouchao@mail.sysu.edu.cn

李轩,男,副教授,主要研究方向为智能无人驾驶、智能无人系统协同感知、生成式人工智能。E-mail:lixuan@bitzh.edu.cn