

文章编号: 1005-0523(2026)03-0001-12



低空底层视觉综述

赵柯嘉, 孙一铭, 朱鹏飞

(东南大学自动化学院, 江苏 南京 210096)

摘要: 底层视觉技术对提升无人机(UAV)在复杂环境下的感知能力至关重要。然而, 低空场景特有的运动模糊、气象扰动及光照不足等耦合退化问题, 叠加 UAV 平台的算力约束与低空复杂物理环境, 严重制约了现有算法的鲁棒性与边缘端实时性。对此, 文章系统综述了低空底层视觉领域的研究进展, 聚焦退化恢复、信息增强与质量评估三大核心方向, 不仅深入剖析了超分辨率、恶劣天气退化去除、低光增强及多源融合等前沿方法的技术特点与应用价值, 还系统梳理了现有的量化评估体系, 并进一步指出未来需重点突破多模态协同、无监督/自监督学习等方向, 以推动低空智能感知技术的持续演进。

关键词: 低空底层视觉; UAV 感知; 图像恢复; 图像增强; 多源融合

中图分类号: V279; [U8]

文献标志码: A

本文引用格式: 赵柯嘉, 孙一铭, 朱鹏飞. 低空底层视觉综述[J]. 华东交通大学学报, 2026, 43(3): 1-12.

Research Review on Low-Altitude Low-Level Vision

Zhao Kejia, Sun Yiming, Zhu Pengfei

(School of Automation, Southeast University, Nanjing 210096, China)

Abstract: Low-level visual technology is crucial for enhancing the perception capability of unmanned aerial vehicles (UAV) in complex environments. However, the coupled degradation problems unique to low-altitude scenarios, such as motion blur, meteorological disturbances, and insufficient illumination, combined with the computational constraints of UAV platforms and the complex physical environment at low altitudes, severely restrict the robustness and edge-side real-time performance of existing algorithms. To address this, this paper systematically reviews the research progress in the low-altitude low-level vision field, focusing on three core directions: degradation recovery, information enhancement, and quality assessment. This paper not only deeply analyzes the technical characteristics and application value of cutting-edge methods such as super-resolution, degradation removal under adverse weather conditions, low-light enhancement, and multi-source fusion, but also systematically summarizes the existing quantitative evaluation systems. Furthermore, this paper points out that future efforts should focus on key breakthroughs in multimodal collaboration, unsupervised/self-supervised learning, etc., to drive the continuous advancement of low-altitude intelligent perception technology.

Key words: low-altitude low-level vision; UAV perception; image restoration; image enhancement; multi-source fusion

收稿日期: 2026-01-27

基金项目: 国家自然科学基金重点项目(62436002); 国家自然科学基金青年项目(62506073); 国家重点研发计划雄安新区科技创新专项(2025XAGG0039); 天津市杰出青年科学基金项目(23JCJQJC00270); 中国博士后科学基金国家资助博士后研究人员计划(GZB20250395)

Citation format: ZHAO K J, SUN Y M, ZHU P F. Research review on low-altitude low-level vision[J]. Journal of East China Jiaotong University, 2026, 43(3): 1–12.

随着低空经济的蓬勃发展,无人机(UAV)在复杂环境下的智能感知需求日益迫切。底层视觉技术作为提升其环境理解与任务执行可靠性的关键基础,其研究价值与应用重要性日益凸显。本文系统梳理了低空底层视觉在退化恢复、信息增强与质量评估三大方向的研究脉络,并深入剖析了超分辨率重建、低光增强及多源融合等前沿方法的技术特点与发展趋势。

1 低空底层视觉概述

1.1 研究背景与意义

低空飞行区域通常界定为海拔1 000 m以下的空域。近年来,党和国家高度重视低空经济的发展,在政策的有力推动下,低空经济呈现出蓬勃发展的态势。随着低空经济与UAV产业的快速发展,对底层视觉技术的需求也日益迫切。

底层视觉任务是计算机视觉领域的重要基础性组成部分,其通过对原始像素数据的处理,提取图像中的物理特征或改善图像质量,为高层语义理解任务提供高质量的数据支持。在低空经济场景下,底层视觉技术能够帮助UAV等飞行器更好地感知周围环境,实现精准定位、导航与控制,提高飞行的安全性和可靠性。例如,在UAV航拍图像小目标检测等下游关键任务中,高质量的底层视觉数据特征是保障算法精度与识别效率的核心基础^[1],对于推动低空经济的进一步发展具有重要意义。

1.2 底层视觉任务分类

底层视觉任务聚焦原始像素数据的预处理与增强,形成“退化恢复→信息增强→质量评估”的技术链条,共同解决低空场景中因环境干扰、设备限制导致的图像质量退化问题。根据任务目标与技术流程,本文将低空底层视觉任务系统划分为退化恢复任务、信息增强任务与质量评估体系三大类。底层视觉任务近10年部分典型算法如图1所示。

1.2.1 退化恢复任务

退化恢复任务旨在逆转成像过程中的物理失真,补偿由设备局限、环境干扰或传输损耗导致的

信息丢失,主要包括:①超分辨率重建;②天气退化去除;③运动与光学退化校正;④图像/视频压缩。

1.2.2 信息增强任务

不同于退化恢复的物理修复导向,信息增强任务聚焦于挖掘图像特征,提升其语义可辨识度与跨模态关联性,核心方向涵盖:①低光增强;②多源融合。

1.2.3 质量评估体系

质量评估体系构建了从评估框架到量化指标的闭环验证机制,兼顾像素级重构精度与人眼/机器的语义感知一致性,避免过度平滑或伪细节引入,以保障低空视觉数据的可靠性。

1.3 低空场景的独特挑战

低空底层视觉技术因所处物理环境与任务需求的深度融合而展现出显著的特殊性,面临环境退化复杂、任务耦合紧密及实时性要求高等多重挑战。在低空复杂大气环境中,多种退化因素常呈现非线性叠加,如雨雾、沙尘等悬浮粒子严重降低图像信噪比,而UAV受风扰产生的高频振动则导致空变运动模糊。此外,低空俯视视角下背景结构复杂、目标尺度小且遮挡严重,进一步增加感知难度^[2]。

低空作业通常需同步完成感知、定位与决策等多类任务,传统单任务优化模型难以应对其耦合需求。尽管已有研究通过引入注意力机制、设计轻量化架构或采用边云协同计算以提高时序一致性和任务协同能力,但在跨模态关联建模、运动目标跟踪等方面仍存在明显瓶颈。这些需求要求算法具备高度灵活性和场景适应性,进一步增加了低空底层视觉系统的设计复杂度。

2 退化恢复类任务

作为低空底层视觉的核心任务类型之一,退化恢复类任务致力于逆转成像链路中由设备局限、环境干扰或传输损耗导致的图像质量下降。本节将系统梳理涵盖超分辨率重建、低质量图像/视频复原、图像/视频压缩等关键方向的技术脉络。

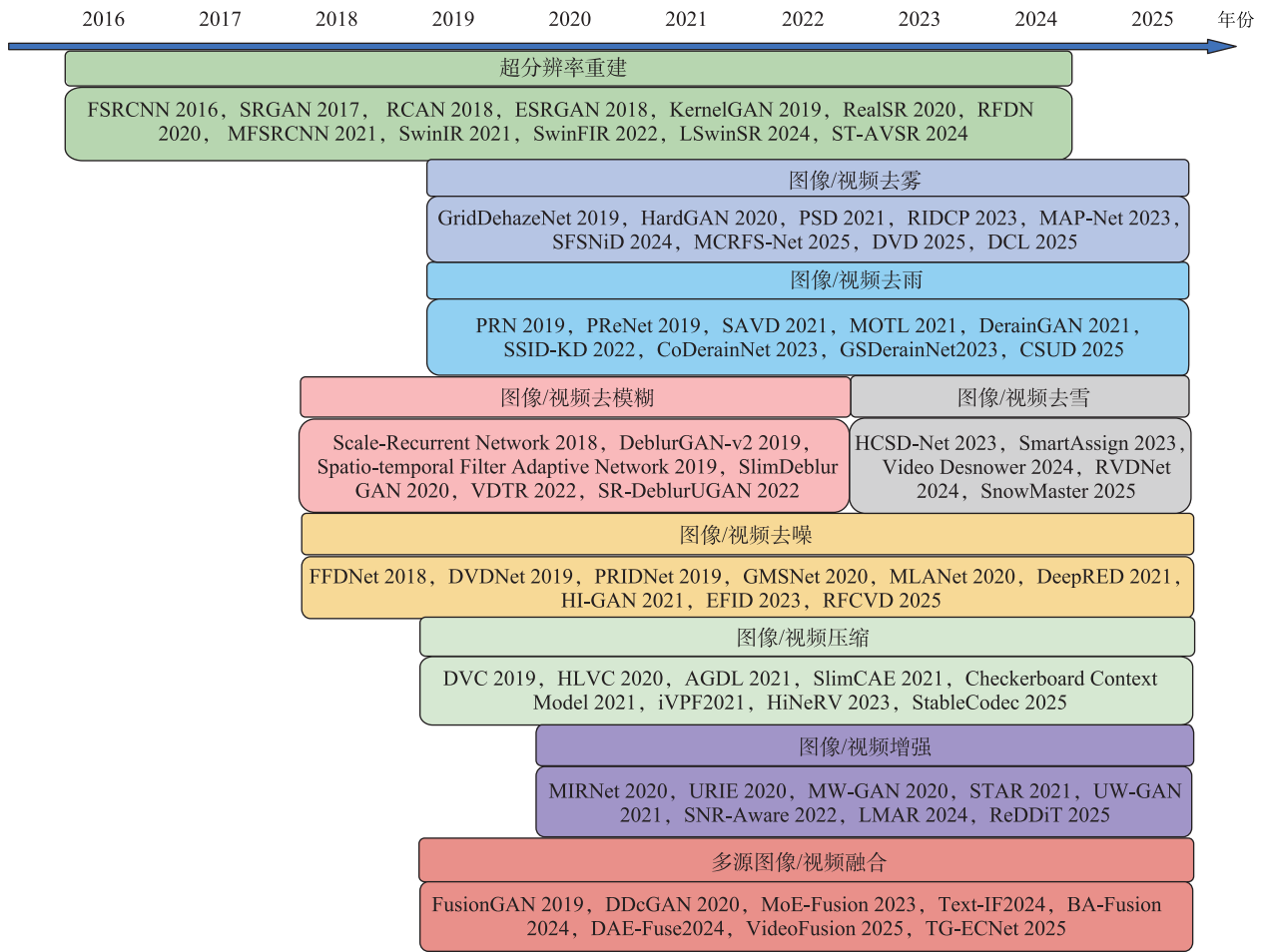


图1 底层视觉任务示意图

Fig. 1 Schematic diagram of low-level visual tasks

2.1 超分辨率重建

超分辨率重建(SR)是由低分辨率图像或视频序列重建出高分辨率图像或视频的关键技术。在低空UAV领域,受多种因素影响,UAV航拍图像常存在分辨率不足、细节模糊等问题,超分辨率技术可提升其清晰度与地物细节,获取更丰富、准确的关键信息。

近年来,变换器(Transformer)架构在超分辨率领域展现出突破性潜力。基于Swin Transformer的图像复原模型(SwinIR)将Swin Transformer的移位窗口自注意力机制引入图像复原,有效建模长程依赖关系,在多个基准任务中刷新性能记录^[3]。与单图超分辨率不同,视频超分辨率不仅利用目标帧自身的信息,还充分利用相邻帧之间包含的、具有细微差异的互补信息。为应对视频超分辨率中尺度灵活性不足的问题,Shang等^[4]提出了ST-AVSR算

法,创新性地引入结构与纹理先验信息。

UAV平台对算法的计算效率有严苛限制。针对远距离小目标探测需求,加速超分辨率卷积神经网络(MFSRCNN)设计四分支并行结构实现多尺度特征融合,其高频-低频特征兼容机制有效增强轮廓细节,在VisDrone数据集上提升目标检测置信度并满足实时性要求^[5]。轻量级Swin Transformer超分辨率网络(LSwinSR)则结合Swin Transformer的窗口注意力机制与线性复杂度设计,在嵌入式平台保持全局建模能力^[6]。为进一步压缩模型,残差特征蒸馏网络(RFDN)通过残差特征蒸馏模块迁移教师网络知识,在Jetson TX2平台实现低延迟推理^[7]。

总体而言,现有的超分辨率技术在低空场景中面临重建精度与计算开销的固有矛盾。基于Transformer的架构虽在纹理恢复和长程依赖建模上优势显著,但其庞大的参数量难以直接部署于微型

UAV;而轻量化卷积神经网络(CNN)虽能满足实时性要求,却往往在复杂地物(如农田、密集建筑)的边缘重建上表现乏力。未来,结合低空高频振动特性的频域先验知识,开发自适应动态网络,实现按需分配算力的轻量化模型,将是突破的关键。

2.2 恶劣天气退化去除

恶劣天气退化去除是指利用计算机视觉技术,从因恶劣天气导致质量下降的原始数据中恢复出清晰、高质量内容的过程,其核心在于纠正图像/视频在成像过程中产生的失真。

2.2.1 去雾任务

去雾任务旨在消除图像中由大气中悬浮颗粒(如雾、霾、烟尘)引起的散射效应,恢复场景的清晰度和色彩保真度,显著提升恶劣天气条件下的视觉质量和场景理解能力。

随着人工智能领域的发展,研究者开发了许多基于深度学习的图像去雾算法。但由于明显的域间隙,当应用于现实世界的含雾图像时,其性能会急剧下降。为克服这一挑战,目前已提出一些用于现实世界图像去雾的方法。例如,Chen等^[8]提出PSD去雾模型,利用三个统计图像先验进行无监督微调;Wu等^[9]利用预训练的向量量化生成对抗网络(VQGAN)中的潜在离散先验。

直接将图像去雾方法应用于视频会忽略视频中各帧之间所蕴含的时序信息,导致各帧的去雾结果不连续。Xu等^[10]设计了一个基于多范围时间对齐网络和物理先验的视频去雾框架MAP-Net,深度挖掘雾霾物理特性并实现时序信息的高效融合。2024年,Fan等^[11]开发了一种新的以深度为中心的学习框架,通过共享的深度估计网络将大气散射模型与亮度一致性约束相结合,利用相邻的去雾帧,通过亮度一致性约束来增强深度估计。

无人机(UAV)场景的去雾技术需应对高空成像的亮度不均问题。Su等^[12]提出的亮度均衡模型,先通过自适应伽马校正平衡图像全局亮度,再结合改进暗通道先验估计雾浓度,有效解决了无人机图像中天空与地面区域的亮度差异导致的去雾失衡问题。针对多旋翼无人机的动态成像特性,Li^[13]设计了基于透射率优化的去雾框架,通过引入运动模糊感知因子修正透射率估计,减少因机身抖动导致的局部雾浓度误判。面向感知任务的无人机去雾更注重场景语义保持,Qiu等^[14]提出超像素场景先

验引导的方法,将图像分割为具有一致语义的超像素块,基于块内雾分布特性自适应调整去雾强度,使去雾后的图像在作物分类、道路检测等下游任务中准确率提升。

纵观上述研究,基于物理先验的大气散射模型方法运算量小且可解释性强,但在浓雾或光照突变时易出现透射率估计失效;数据驱动的深度学习方法复原保真度高,但严重依赖特定训练数据分布,跨场景泛化能力受限。未来的研究应致力于将物理模型与无监督/半监督学习深度融合,摆脱对配对训练数据的依赖,并探索引入“视觉-语言”大模型(VLM)的零样本泛化能力,以应对开放域复杂气象条件下的去雾难题。

2.2.2 去雨任务

去雨任务专注于从受降雨影响的图像或视频中移除雨滴和雨痕,以恢复被雨水模糊或遮挡的细节,提升雨天户外视觉系统的性能。

深度学习方法通过端到端学习显著提升了去雨性能。Ren等^[15]提出的渐进式残差网络(PRN)与渐进式循环网络(PReNet),通过递归展开浅层ResNet并引入循环层挖掘跨阶段特征依赖,以极少的参数和简单的损失函数实现了优异的去雨性能,为后续研究确立了强基准。Ye等^[16]创新性地提出“雨纹-背景”解耦翻译框架(JRGR),通过共享编码器和分离解码器同步实现雨图生成与去雨,其循环一致性损失支持无配对数据训练,该框架的物理可解释性为后续研究奠定了重要基础。长序列一致性优化是视频去雨的关键挑战之一。自对齐视频去雨方法(SAVD)提出“透射-深度”一致性约束,基于大气散射模型推导透射率与深度关系,构建几何一致性损失对齐背景结构,该工作将物理先验嵌入视频修复,为无人机场景的几何约束提供技术范式^[17]。

无人机图像去雨任务需兼顾高频细节保留与算力约束。记忆导向迁移学习框架(MOTL)设计记忆导向迁移框架:预训练模型学习通用特征,记忆增强适配器存储无人机典型样本,知识蒸馏实现小样本微调,在极低的参数量级下实现了边缘设备上的低延迟实时推理^[18]。无监督方法通道一致性先验(CCP)^[19]利用雨纹RGB通道强度差异,构建通道间差分损失,通过自编码器实现无标签训练,解决标注数据匮乏问题。协同去雨网络(CoDeRain-Net)通过协同去雨与目标检测模块实现闭环优化,

在设计的双分支网络中,去雨分支输出清晰图像,检测分支反馈目标区域掩码,指导去雨分支聚焦感兴趣区域,在雨天无人机巡检中使绝缘子检测准确率得到提升^[20]。高斯-香农滤波器与深度学习的结合为无人机去雨提供了新路径,基于高斯-香农滤波器的去雨深度网络(GSDerainNet)先通过高斯-香农滤波器初步抑制高频雨纹,再利用卷积神经网络(CNN)学习残余雨雾的映射关系,该混合架构在保留地物边缘的同时,适配无人机低空成像的细节识别需求^[21]。生成对抗网络在无人机去雨的视觉逼真度提升上表现突出。基于 Wasserstein 生成对抗网络的单图去雨方法(DerainGAN)采用 Wasserstein 距离优化对抗损失,生成器通过 U-Net 结构输出去雨结果,判别器区分真实无雨图像与生成去雨图像,在无人机复杂背景中生成的去雨图像纹理更自然^[22]。

综上所述,当前的去雨算法已从单一的图像域处理向物理可解释的解耦框架演进,显著改善了静态细节的保留。然而,针对无人机视频去雨,如何在高速运动下保持帧间背景的几何一致性,同时避免将细小目标误判为雨纹予以滤除,仍是现有技术的盲区。后续研究可进一步挖掘雨滴的动态运动学轨迹,引入脉冲神经网络等低功耗、高时序敏感架构,实现高动态场景下的高效视频去雨。

2.2.3 去雪任务

去雪任务旨在去除图像中降雪造成的退化。去雪比去雨更复杂,因为雪花形态多样、分布更密集,且积雪会改变场景的纹理和结构信息,对算法的鲁棒性要求更高。

基于物理退化模型的去雪方法因其可解释性成为研究基石。深度展开去雪方法(DUSSP)率先将雪花形态学特征(snow shape prior)作为正则项嵌入迭代优化网络,开创的物理约束范式为后续研究指明了方向^[23]。面对局部积雪与飘雪共存的挑战,知识动态分配去雪方法(SmartAssign)提出知识动态分配机制,通过可微分阈值门控分离高密度积雪与低密度飘雪,标志着从全局处理到区域自适应的重要转变^[24]。视频去雪需额外解决时序一致性与运动模糊问题。随着运动建模需求深化,自适应视频去雪模型(video desnow)融合可变形卷积与 K 近邻(KNN)点云 Transformer,自适应融合模块聚合多帧信息,其运动感知架构显著提升动

态场景鲁棒性^[25]。

在无人机场景中,去雪技术与特定任务的结合更为紧密。Saleem 等^[26]通过去雪算法预处理无人机图像,消除雪花对太阳能电池板区域的干扰,再利用机器学习模型计算积雪覆盖度,使能量效率评估误差降低。Belmonte 等^[27]对不同时段无人机的航拍数据进行去雪处理,清晰化地物边界后结合时序分析,精准捕捉半干旱森林区域的雪盖持续时间变化,为生态研究提供可靠的数据支持。真实场景视频去雪域适应网络(RVDNet)设计两阶段域适应框架:合成数据预训练后,通过可变形卷积对齐真实无人机视频域,运动感知损失抑制云层移动伪影,为真实场景落地提供有效解决方案^[28]。

在图像恢复领域,复合退化场景的建模与处理一直是核心挑战。针对此问题,基于 Transformer 的图像复原框架(OneRestore)提出了通用恢复方法。该框架构建了多退化联合表示模型,将退化场景描述符与图像特征自适应融合,实现可控的退化感知恢复^[29]。

相比去雨,去雪任务因雪花形态多样及积雪遮挡而更具挑战性。现有方法通过引入形状先验或动态分配机制,已能较好应对飘雪干扰,但对于积雪严重改变地物原有纹理的深度遮挡问题,修复的语义连贯性依然欠佳。未来的发展方向在于构建“去雪-补全”的联合生成框架,并结合无人机的多视角时序观测优势,利用三维结构先验辅助被遮挡区域的内容推演。

2.3 运动与光学退化校正

运动与光学退化校正是指利用计算机视觉技术,从因相机抖动、目标快速移动或镜头对焦不准等导致质量下降的原始数据中恢复出清晰、高质量内容的过程。无人机(UAV)受限于飞行中的高频抖动、快速移动以及复杂光照下的传感器噪声,常获取模糊或噪点严重的影像,直接影响目标识别、精确定位与跟踪等任务的精度与可靠性。

2.3.1 去模糊任务

去模糊任务致力于纠正图像中因相机与拍摄对象的相对运动或镜头未能准确对焦而产生的整体或局部细节模糊。

基于真实模糊合成的去模糊方法(DRB)创新性地提出“模糊-清晰”对生成框架,通过物理模型合成逼真样本并设计域自适应模块对齐分布,大幅

提升了模型的泛化能力^[30]。视频去模糊需解决时序关联性问题。近年来,Transformer展现出显著优势,基于Transformer的视频去模糊方法(VDTR)首次将其引入视频去雾任务,时空注意力模块捕捉长程依赖,跨帧门控机制抑制闪烁伪影^[31]。

UAV图像面临高频细节丢失与算力约束的双重挑战。渐进式残差循环网络(PRRN)设计轻量化的“编码器-解码器”模块,利用门控渐进残差块大幅削减模型的计算开销,并通过特征循环结构有效抑制模糊残留^[32]。针对特定任务的UAV去模糊任务注重场景语义保持。轻量化生成对抗去模糊网络(SlimDeblurGAN)在着陆标记检测中,通过生成对抗网络学习模糊核与清晰特征,轻量化设计支持机载实时处理,验证了任务导向的实用价值^[33]。高速移动UAV的去模糊需针对性优化,Liu等^[34]引入运动模糊感知因子以调整去模糊强度,在高速公路航拍场景中抑制了因快速移动导致的远景模糊。融合超分辨率与去模糊的UAV图像处理更注重细节增强,融合超分辨率与去模糊的端到端UAV图像处理模型(SR-DeblurUGAN)通过双分支网络学习超分辨率特征与模糊核,在农田航拍中使作物分类准确率显著提升,体现了多任务融合的优势^[35]。

现有去模糊技术在应对低空特有的空间变化运动模糊时,往往陷入局部感受野受限或全局注意力计算过载的困境。此外,多数模型对UAV高频机械抖动引发的复杂模糊核估计不足。未来,探索融合事件相机等多模态传感器的异步视觉信息,利用其极高的时间分辨率辅助传统光学帧的模糊核估计与盲复原,将为低空极端运动模糊校正提供全新的技术路径。

2.3.2 去噪任务

去噪任务专注于去除图像在采集或传输过程中引入的随机干扰信号,同时尽可能保留原始图像的边缘、纹理等关键细节信息,避免过度平滑导致图像失真或产生伪影。

图像去噪的核心挑战在于平衡噪声抑制与细节保留。高效高分辨率图像复原Transformer模型(Restormer)为高效的高分辨率图像复原奠定了新范式,其构建的基于Transformer的架构引入多头转置注意力机制以并行处理通道维度关系,并设计门控前馈网络实现局部与全局特征的自适应融合,在多个去噪基准测试集上实现了卓越的性能与效率

平衡^[36]。

在UAV场景中,去噪技术为后续任务提供质量保障。Tian等^[37]将去噪预处理融入UAV绝缘子检测流程,通过抑制航拍图像中的随机噪声,提升DETR算法(一种端到端目标检测算法)对绝缘子的识别精度,使检测准确率显著提高。堆叠剪枝稀疏去噪自编码器与CNN相结合的方法,在UAV叶片损伤诊断中,先通过去噪自编码器去除振动信号中的噪声,再利用CNN提取损伤特征,使故障识别率显著提升^[38]。ID-SAbRUNet作为针对UAV逆合成孔径雷达(ISAR)图像的去噪网络,通过抑制图像中的干扰与噪声,增强目标散射特性,有效提升了UAV ISAR图像的解译精度^[39]。Azmat等^[40]将去噪技术融入UAV航拍的人体动作识别流程,通过抑制复杂背景中的噪声干扰,提升深度学习模型对人体动作特征的提取精度。

目前的去噪算法在标准高斯噪声数据集上的性能已趋于饱和,但面对UAV实际搭载的微型传感器在低照度、高感光度(ISO)条件下产生的复杂真实噪声,往往会出现过度平滑的问题,进而抹除高频有用信息。未来的突破点在于构建贴合低空物理成像链路的真实传感器噪声模型,并发展“任务驱动”的感知型去噪策略,即去噪的目标不再是单纯追求极致的信噪比,而是最大化下游机器视觉任务(如检测、追踪)的性能收益。

2.4 图像/视频压缩

图像/视频压缩是通过算法显著减小数据量的技术,其核心是在可接受的视觉质量损失下消除信息冗余。

图像压缩技术正经历从传统编码到神经表示的范式转变。注意力引导的深度学习图像压缩方法(AGDL)创新性地融合了压缩感知与视觉显著性机制,通过深度学习重建的显著骨架图指导码率分配,使Kodak数据集中人眼关注区域的峰值信噪比(PSNR)显著提升^[41]。在此基础上,可切换压缩比的可变参数压缩自编码器(SlimCAE)通过权重共享支持多档码率的动态调整,在移动端部署时能够在几乎不损失结构相似度的前提下,大幅缩减模型体积^[42]。视频压缩需平衡时域依赖性与计算复杂度,分层质量与递归增强视频压缩框架(HLVC)设计了三级增强结构:I帧压缩层用于初始化,双向深度压缩层利用时域参考,单运动压缩层共享运动矢

量,配合加权递归增强网络融合多级信息,有效降低了视频编码的码率开销^[43]。

UAV采集的高清视觉数据直接传输时带宽压力巨大,通过高效压缩可降低码率,保障5G回传的实时性与稳定性,满足电力巡检、灾害监测等应用的迫切需求。高效压缩显著降低了通信与存储成本,推动了UAV视联网的规模化落地。因此,UAV场景下的图像/视频压缩需平衡压缩效率与实时性。端到端智能视频压缩技术^[44]通过整合特征提取与压缩编码模块,在减少数据传输量的同时保证无人机航拍视频的关键细节清晰,适配UAV有限的传输带宽。将深度学习方法应用于UAV视频压缩,可针对航拍视频中动态背景与快速运动目标的特性优化压缩策略,在同等码率下有效提升重建视频的PSNR与客观视觉质量。Smirnov等^[45]对适用于遥控UAV的视频压缩算法进行了分析与优化,发现针对UAV抖动导致的帧间冗余,采用自适应帧间预测的压缩方法效果更优,能在减少数据量的同时保证视频流畅性。面向UAV实时视频分析的自适应兼容压缩方法通过动态调整压缩参数以适配不同场景需求,在目标检测场景中保留更多细节,在大范围监控场景中侧重提升压缩比,实现了压缩效率与分析精度的平衡^[46]。

智能视频压缩技术在极低码率下展现出超越传统标准的视觉保留能力,但其非对称的编解码算力需求(通常编码端计算量极大)与UAV端侧算力匮乏的现状存在严重错配。展望未来,神经视频编码亟需向轻量化编码、复杂解码的端云协同架构演变;同时,面向机器视觉的语义级压缩(仅保留感兴趣目标区域的特征)有望取代传统的像素级压缩,成为低空视联网图像传输的主流范式。

3 信息增强类任务

区别于退化恢复的“修复”导向,信息增强类任务聚焦于主动提升图像内容的可辨识度与可利用性,使其更适配于机器分析或人眼观察的特定需求。

3.1 低光增强

低光增强通过算法改善视觉内容的主观质量或特定信息的可辨识度。它不追求严格复原原始场景,而是根据应用需求优化观感或突出关键特征。其核心意义在于提升内容的视觉表现力与信

息传达效率,使图像/视频更清晰悦目,或更易于后续的人工观察与机器分析。

图像增强的核心挑战在于平衡真实感恢复与计算效率。多尺度图像复原网络(MIRNet)是一种多尺度并行特征融合架构,通过多尺度残差块(MRB)聚合高低分辨率的上下文信息,并结合通道与空间特征重校准机制,在LOL等数据集上实现了优异的视觉重建与低光增强效果^[47]。进一步地,信噪比感知(SNR-Aware)低光增强框架首次明确建模并利用了图像中的信噪比先验,有效避免了在极暗区域过度放大噪声^[48]。在单图增强基础上,She等^[49]提出的图像到视频模型聚焦实时视频增强需求,通过构建从单图到视频的增强映射关系,在保证帧间时序一致性的同时实现实时处理,为动态场景中快速运动物体的细节提升提供了有效方案。

UAV航拍常面临强烈光照变化、薄雾或霾导致的对比度下降、色彩失真等问题。应用增强技术可提升航拍数据的可用性与分析精度,支撑高效决策。Reale等^[50]将视频增强技术应用于UAV影像,通过增强关键区域特征提升影像解译性,在城市规划与灾害评估场景中,显著增强了UAV航拍影像中建筑物、道路等目标的可识别性。Sun等^[51]融合时空信息与光流特征以增强监控视频中的UAV目标,抑制复杂背景干扰,拓展了图像增强技术在目标检测下游任务中的应用。

现有的增强网络虽能有效提升暗部亮度和信噪比,但在处理光源直射、局部反光等高动态范围场景时,容易诱发色彩畸变与高光过曝。未来的演进趋势是摆脱单纯的像素值映射,转向基于物理光照重构的本征图像分解,并尝试引入多模态大模型的指令交互能力,实现根据具体勘察任务需求进行可控、自适应的照度增强。

3.2 多源图像/视频融合

多源图像/视频融合是指将来自不同传感器的图像或视频数据通过算法整合为单一的、信息更丰富且更可靠的输出。

2019年,Ma等^[52]开发了一种多模态图像融合生成对抗网络模型FusionGAN。该研究创新性地将生成对抗网络引入多模态图像融合任务,利用生成器同时保留红外目标显著性特征与可见光空间细节,并通过判别器的对抗性优化提升融合图像的

自然感知质量。为解决“可见光-红外”融合中的亮度不匹配问题, Sun 等^[53]提出了一种亮度自适应图像动态融合框架 BA-Fusion, 利用场景感知的亮度校正模块, 先通过灰度引导滤波精确估计参考亮度图, 再借助空间变换函数自适应调整红外与可见光图像的亮度贡献。

现有的多模态融合方法通常将图像融合技术直接扩展到视频融合任务中, 对“可见光-红外”视频对进行逐帧融合。这些方法丢失了固有的时间信息, 并且难以保持视频帧之间的时间一致性。一些研究者在保留时序信息的方向上进行了探索。2025年, Gong 等^[54]提出了一种名为 TemCoCo 的多模态视频融合模型, 它充分利用跨模态互补性和时间动态性, 从多模态输入生成时空连贯的视频。

在 UAV 场景下, 多源图像/视频融合算法围绕

特征对齐、场景适配、数据拓展与实用优化形成了递进式技术体系。Liu 等^[55]提出的 VL-MFL 算法作为基础, 通过多源特征提取模块同步处理 UAV 航拍光学图像与卫星遥感图像的局部及全局特征, 为后续融合任务奠定了特征一致性基础。在此基础上, Li 等^[56]提出的双流修复嵌入模糊特征融合网络 (DREB-Net) 针对高机动性场景设计了双流融合网络, 左分支恢复运动模糊图像特征, 右分支提取模糊区域边缘梯度, 经注意力加权后解决了传统方法对运动模糊的敏感性问题, 实现了融合策略对动态场景的适配。Sun 等^[57]提出的 TG-ECNet 算法则聚焦于退化多模态图像的“融合难题”, 其框架如图 2 所示。该框架减少了复原任务和融合任务之间的矛盾, 最大限度地降低了传统先复原再融合的级联结构中的信息损失。

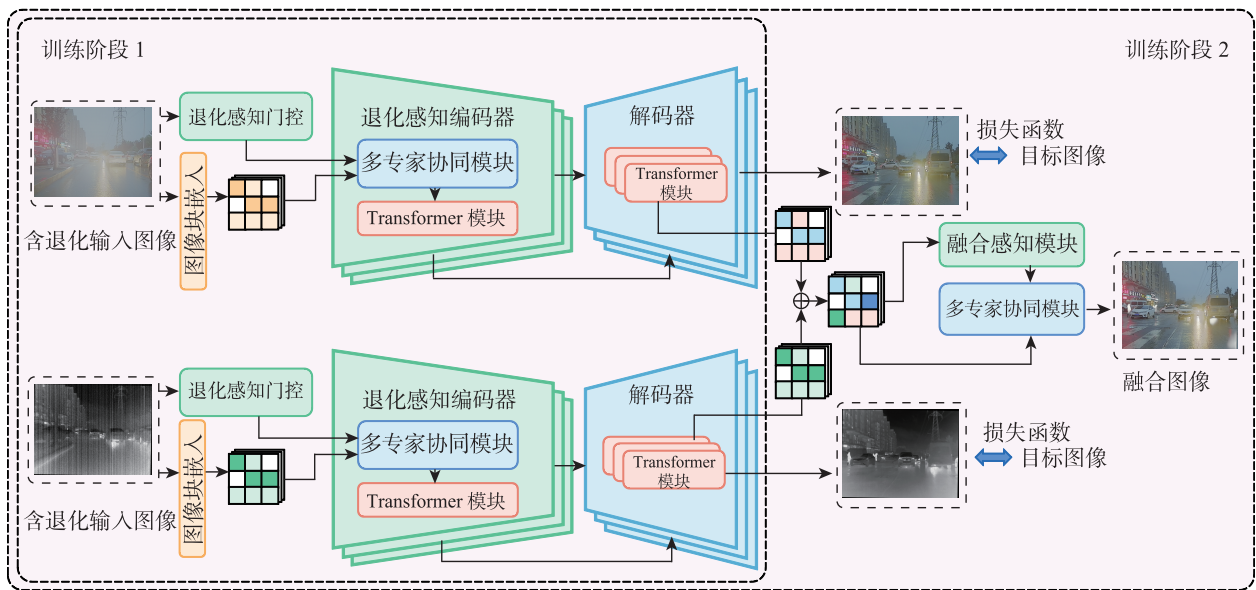


图 2 TG-ECNet 框架图

Fig. 2 Framework diagram of TG-ECNet

现阶段多源融合技术在特征表征层面已相当成熟, 但在实际低空部署中, 由 UAV 抖动导致的可见光与红外等模态间时空微小不对齐问题, 极易被现有深层网络放大为严重的重影等伪影。未来的研究必须打破“先配准、后融合”的传统串行级联思路, 走向抗畸变的联合对齐与融合一体化网络, 从源头解决跨传感器的空间配准误差问题。

4 底层视觉质量评估

质量评估体系为低空底层视觉任务提供了不

可或缺的量化标准与反馈机制。本节将系统阐述低空底层视觉领域的评估体系, 聚焦于评估框架和评估指标两大维度, 对现有评估体系进行详细论述。

4.1 典型评估框架

低空底层视觉的评估框架定义了量化算法性能的基本范式。针对数据获取复杂性与应用需求多样性, 需通过不同方法构建闭环验证机制, 为算法提供多维度的可靠判据。

基于 Transformer 的无参考图像质量评估框架 (TreS) 是 2021 年由 Golestaneh 等^[58]提出的针对无

参考情况的图像质量评估框架。该框架采用多尺度视觉 Transformer 作为主干网络,通过分层下采样捕获局部细节与全局语义的融合特征。这一设计能够避免 CNN 局部感受野的限制,提升对复杂失真的建模能力。模型采用自一致性约束,增强了对局部失真变化的鲁棒性,避免过拟合于特定区域。该框架在跨数据集测试中的泛化性提升显著。

多维度注意力无参考图像质量评估网络(MA-NIQA)是2022年由Yang等^[59]提出的针对无参考情况的评估框架。该框架通过引入“通道-空间”双维度注意力机制,显著提升了无参考图像质量评估的精度,尤其在处理复杂失真时表现卓越。其网络设计中采用预训练的 Vision Transformer(ViT-B/8)作

为特征提取层以提取多尺度特征,最后设计双分支预测结构,模拟人眼对图像不同区域的关注差异,使图像评价更符合人眼的感知规律。

4.2 评估指标

评估指标是评估框架的具体量化工具,用于精确衡量底层视觉算法的输出质量。低空场景的特殊性要求指标需协同考量三个维度:保真度指标量化信号层面的重建精度,反映图像底层结构的恢复程度;感知质量指标模拟人类视觉系统,以评估结果的视觉自然度和真实性,避免算法引入伪影或过度平滑;任务效用指标则直接关联高层应用性能,衡量处理结果对最终任务的贡献。底层视觉常用评估指标的分类情况如表1所示。

表1 底层视觉任务评估指标分类
Tab. 1 Classification of low-level visual task evaluation metrics

任务类型	项目	典型指标
退化恢复类任务	超分辨率重建	峰值信噪比 (PSNR)、结构相似性 (SSIM)、视觉信息保真度 (VIF)、均方误差 (MSE)
	去雾任务	PSNR、SSIM、MSE、色彩失真度量 (CDM)
	去雪任务	PSNR、SSIM
	去雨任务	PSNR、SSIM、均方根误差 (RMSE)
	去模糊任务	PSNR、SSIM、感知图像块相似度 (LPIPS)、多尺度结构相似性 (MS-SSIM)
	去噪任务	PSNR、SSIM
	修复任务	PSNR、SSIM、特征向量距离 (FID)、平均绝对误差 (MAE)、MSE
	压缩任务	PSNR、SSIM、相关系数 (CC)
信息增强类任务	增强任务	PSNR、SSIM、信息熵 (EN)、CC、LPIPS、FID
	多源融合	CC、MSE、PSNR、SSIM、空间频率 (SF)、VIF

5 结束语

低空底层视觉技术作为UAV智能感知的基石,目前已形成涵盖退化恢复、信息增强与质量评估的完整技术体系。然而,当前技术仍受限于极端气象条件下的性能衰减、边缘端多任务实时处理能力不足,以及多源数据时空对齐误差。此外,特定任务的高质量基准数据集的匮乏也严重制约了算法的针对性研发与评估。

为突破上述瓶颈,未来研究需聚焦多模态协同建模与自主演进。首先,应构建“气象-视觉”联合退化模型,将物理机理嵌入神经网络,并利用激光雷达辅助提升跨模态对齐精度,以应对浓烟、暴雨等极端环境。其次,应重点发展自监督与无监督学习,利用时空一致性约束与对比学习降低对标注数据的依赖。同时,引入开放世界“视觉-语言”模型,

将有助于低空系统在无先验场景下实现高鲁棒性的通用感知,为低空经济发展构建智能底座。

参考文献:

[1] 董华军,王宇栖. 基于改进YOLOv5s的无人机小目标检测算法研究[J]. 华东交通大学学报, 2024, 41(4): 118-126.
DONG H J, WANG Y Q. Research on UAV small target detection algorithm based on improved YOLOv5s[J]. Journal of East China Jiaotong University, 2024, 41(4): 118-126.

[2] 孙一铭,赵柯嘉,王硕,等. 面向无人机的低空视觉数据集研究综述[J]. 数据采集与处理, 2025, 40(2): 274-302.
SUN Y M, ZHAO K J, WANG S, et al. Research review on low-altitude visual datasets for unmanned aerial vehicles[J]. Journal of Data Acquisition & Processing, 2025, 40(2): 274-302.

[3] LIANG J Y, CAO J Z, SUN G L, et al. SwinIR: image res-

- toration using Swin transformer[C]//2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), October 11-17, 2021, Montreal, BC, Canada. New York: IEEE, 2021: 1833-1844.
- [4] SHANG W, REN D W, ZHANG W Y, et al. Arbitrary-scale video super-resolution with structural and textural priors[C]//LEONARDIS A, RICCI E, ROTH S, et al., eds. Computer Vision-ECCV 2024. Cham: Springer, 2025: 73-90.
- [5] 李志朋, 赵长明, 张海洋, 等. 多尺度融合超分辨率算法在无人机探测中的应用[J]. 应用光学, 2021, 42(3): 462-473.
- LI Z P, ZHAO C M, ZHANG H Y, et al. Application of multi-scale fusion super-resolution algorithm in UAV detection[J]. Journal of Applied Optics, 2021, 42(3): 462-473.
- [6] LI R, ZHAO X W. LSwinSR: UAV imagery super-resolution based on linear Swin transformer[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 5641013.
- [7] LIU J, TANG J, WU G S. Residual feature distillation network for lightweight image super-resolution[C]//BARTOLI A, FUSIELLO A, eds. Computer Vision-ECCV 2020 Workshops. Cham: Springer, 2020: 41-55.
- [8] CHEN Z Y, WANG Y C, YANG Y, et al. PSD: principled synthetic-to-real dehazing guided by physical priors[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE, 2021: 7176-7185.
- [9] WU R Q, DUAN Z P, GUO C L, et al. RIDCP: revitalizing real image dehazing via high-quality codebook priors[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 17-24, 2023, Vancouver, BC, Canada. New York: IEEE, 2023: 22282-22291.
- [10] XU J Q, HU X W, ZHU L, et al. Video dehazing via a multi-range temporal alignment network with physical prior[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 17-24, 2023, Vancouver, BC, Canada. New York: IEEE, 2023: 18053-18062.
- [11] FAN J K, WANG K, YAN Z Q, et al. Depth-centric dehazing and depth estimation from real-world hazy driving video[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2025, 39(3): 2852-2860.
- [12] SU J M, LIU L L, WAN Q T, et al. Dehazing research on brightness equalization model of drone image[J]. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2020, XLII-3/W10: 1289-1294.
- [13] LI Z L. A transmittance optimization-based framework for image dehazing on multi-rotor drones imaging[C]//2022 14th International Conference on Machine Learning and Computing (ICMLC), Guangzhou, China. New York: ACM, 2022: 550-554.
- [14] QIU Z F, GONG T Y, LIANG Z C, et al. Perception-oriented UAV image dehazing based on super-pixel scene prior[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 5913519.
- [15] REN D W, ZUO W M, HU Q H, et al. Progressive image deraining networks: a better and simpler baseline[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE, 2019: 39323941.
- [16] YE Y T, CHANG Y, ZHOU H Y, et al. Closing the loop: joint rain generation and removal via disentangled image translation[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE, 2021: 2053-2062.
- [17] YAN W D, TAN R T, YANG W H, et al. Self-aligned video deraining with transmission- depth consistency[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE, 2021: 11961-11971.
- [18] HUANG H B, YU A J, HE R. Memory oriented transfer learning for semi-supervised image deraining[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE, 2021: 7728-7737.
- [19] DONG G L, ZHENG T H, CAO Y, et al. Channel consistency prior and self-reconstruction strategy based unsupervised image deraining[C]//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 10-17, 2025, Nashville, TN, USA. New York: IEEE, 2025: 7469-7479.
- [20] XI Y, JIA W J, MIAO Q G, et al. CoDerainNet: collaborative deraining network for drone-view object detection in rainy weather conditions[J]. Remote Sensing, 2023, 15(6): 1487.
- [21] YAO Y J, SHI Z M, HU H W, et al. GSDerainNet: a deep network architecture based on a Gaussian Shannon filter for single image deraining[J]. Remote Sensing, 2023, 15(19): 4825.
- [22] YADAV S, MEHRA A, ROHMETRA H, et al. DerainGAN: single image deraining using Wasserstein GAN[J]. Multimedia Tools and Applications, 2021, 80(30): 36491-36507.
- [23] GUO X, WANG X, FU X Y, et al. Deep unfolding net-

- work for image desnowing with snow shape prior[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2025, 35(5): 4740-4752.
- [24] WANG Y L, MA C, LIU J Z. SmartAssign: learning a smart knowledge assignment strategy for deraining and desnowing[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 17-24, 2023, Vancouver, BC, Canada. New York: IEEE, 2023: 3677-3686.
- [25] LI Y X, DAI L. Video desnow: an adaptive feature fusion understanding video desnowing model with deformable convolution and KNN point cloud transformer[J]. *IEEE Access*, 2024, 12: 104354-104366.
- [26] SALEEM A, AWAD A, MAZEN A, et al. Estimating snow coverage percentage on solar panels using drone imagery and machine learning for enhanced energy efficiency[J]. *Energies*, 2025, 18(7): 1729.
- [27] BELMONTE A, SANKEY T, BIEDERMAN J, et al. UAV-based estimate of snow cover dynamics: optimizing semi-arid forest structure for snow persistence[J]. *Remote Sensing*, 2021, 13(5): 1036.
- [28] XUE T H, ZHOU G, HE R L, et al. RVDNet: a two-stage network for real-world video desnowing with domain adaptation[C]//ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), April 14-19, 2024, Seoul, Korea, Republic of. New York: IEEE, 2024: 3305-3309.
- [29] GUO Y, GAO Y, LU Y X, et al. OneRestore: a universal restoration framework for composite degradation[C]//Computer Vision- ECCV 2024. Cham: Springer Nature Switzerland, 2024: 255-272.
- [30] ZHANG K H, LUO W H, ZHONG Y R, et al. Deblurring by realistic blurring[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE, 2020: 2734-2743.
- [31] CAO M D, FAN Y B, ZHANG Y, et al. VDTR: video deblurring with transformer[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023, 33(1): 160-171.
- [32] ZHANG X Q, JIANG R H, WANG T, et al. Recursive neural network for video deblurring[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021, 31(8): 3025-3036.
- [33] TRUONG N Q, LEE Y W, OWAIS M, et al. SlimDeblurGAN-based motion deblurring and marker detection for autonomous drone landing[J]. *Sensors*, 2020, 20(14): 3918.
- [34] LIU Y, WANG J W, QIU T T, et al. An adaptive deblurring vehicle detection method for high-speed moving drones: resistance to shake[J]. *Entropy*, 2021, 23(10): 1358.
- [35] XIAO Y Z, ZHANG J D, CHEN W, et al. SR-DeblurUGAN: an end-to-end super-resolution and deblurring model with high performance[J]. *Drones*, 2022, 6(7): 162.
- [36] ZAMIR S W, ARORA A, KHAN S, et al. Restormer: efficient transformer for high-resolution image restoration [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 18-24, 2022, New Orleans, LA, USA. New York: IEEE, 2022: 5718-5729.
- [37] TIAN Y F, AHMAD R B, ABDULLAH N A B. Accurate and efficient insulator maintenance: a DETR algorithm for drone imagery[J]. *PLoS One*, 2025, 20(2): e0318225.
- [38] YANG P, WEN C W, GENG H L, et al. Intelligent fault diagnosis method for blade damage of quad-rotor UAV based on stacked pruning sparse denoising autoencoder and convolutional neural network[J]. *Machines*, 2021, 9(12): 360.
- [39] JIN S, BAE Y, LEE S. ID-SaBRUNet: deep neural network for disturbance suppression of drone ISAR images [J]. *IEEE Sensors Journal*, 2024, 24(9): 15551-15565.
- [40] AZMAT U, ALOTAIBI S S, ABDELHAQ M, et al. Aerial insights: deep learning-based human action recognition in drone imagery[J]. *IEEE Access*, 2023, 11: 83946-83961.
- [41] ZHANG X, WU X L. Attention-guided image compression by deep reconstruction of compressive sensed saliency skeleton[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE, 2021: 13349-13359.
- [42] YANG F, HERRANZ L, CHENG Y M, et al. Slimmable compressive autoencoders for practical neural image compression[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE, 2021: 4996-5005.
- [43] YANG R, MENTZER F, VAN GOOL L, et al. Learning for video compression with hierarchical quality and recurrent enhancement[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE, 2020: 6627-6636.
- [44] 叶枫, 董凡可, 贾川民. 端到端智能视频压缩技术及其在无人机中的应用[J]. *数据采集与处理*, 2025, 40(2): 303-319.
- YE F, DONG F K, JIA C M. End-to-end video compression

- sion technology and its application in unmanned aerial vehicles[J]. *Journal of Data Acquisition and Processing*, 2025, 40(2): 303-319.
- [45] SMIRNOV K, MOZHAEVA A. Analysis of video data compression algorithms and optimized for use in remotely controlled drones[J]. *Synchroinfo Journal*, 2023, 9(2): 9-16.
- [46] XU D, CHELLAPPA R, VAN GOOL L, et al. Guest editorial: special issue on deep learning for video analysis and compression[J]. *International Journal of Computer Vision*, 2021, 129(12): 3171-3173.
- [47] ZAMIR S W, ARORA A, KHAN S, et al. Learning enriched features for real image restoration and enhancement[C]//VEDALDI A, BISCHOF H, BROX T, et al., eds. *Computer Vision-ECCV 2020*. Cham: Springer, 2020: 492-511.
- [48] XU X G, WANG R X, FU C W, et al. SNR-aware low-light image enhancement[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 18-24, 2022, New Orleans, LA, USA. New York: IEEE, 2022: 17693-17703.
- [49] SHE D Y, XU K. An image-to-video model for real-time video enhancement[C]//Proceedings of the 30th ACM International Conference on Multimedia, Lisboa, Portugal. New York: ACM, 2022: 1837-1846.
- [50] REALE M J, MURPHY D P, CORNACCHIA M, et al. Video modification in drone and satellite imagery[C]//Disruptive Technologies in Information Sciences VIII, April 21-26, 2024. National Harbor, USA: SPIE, 2024: 41.
- [51] SUN Y, ZHI X Y, HAN H W, et al. Enhancing UAV detection in surveillance camera videos through spatiotemporal information and optical flow[J]. *Sensors*, 2023, 23(13): 6037.
- [52] MA J Y, YU W, LIANG P W, et al. FusionGAN: a generative adversarial network for infrared and visible image fusion[J]. *Information Fusion*, 2019, 48: 11-26.
- [53] SUN Y M, CAO B, ZHU P F, et al. Dynamic brightness adaptation for robust multi-modal image fusion[J]. *arXiv e-Prints*, 2024: arXiv:2411.04697.
- [54] GONG M Q, ZHANG H, YI X P, et al. TemCoCo: temporally consistent multi-modal video fusion with visual-semantic collaboration[J]. *arXiv e-Prints*, 2025: arXiv:2508.17817.
- [55] LIU G C, LI C, ZHANG S H, et al. VL-MFL: UAV visual localization based on multisource image feature learning[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 5618612.
- [56] LI Q P, ZHANG Y X, FANG L Y, et al. DREB-Net: dual-stream restoration embedding blur-feature fusion network for high-mobility UAV object detection[J]. *arXiv e-Prints*, 2024: arXiv:2410.17822.
- [57] SUN Y M, LI X, ZHU P F, et al. Task-gated multi-expert collaboration network for degraded multi-modal image fusion[C]//ICML'25: Proceedings of the 42nd International Conference on Machine Learning, July 19, 2025, Vancouver, Canada. Massachusetts, USA: MIT Press, 2025: 57571-57586.
- [58] GOLESTANEH S A, DADSETAN S, KITANI K M. No-reference image quality assessment via transformers, relative ranking, and self-consistency[C]//2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), January 3-8, 2022, Waikoloa, HI, USA. New York: IEEE, 2022: 3989-3999.
- [59] YANG S D, WU T H, SHI S W, et al. MANIQA: multi-dimension attention network for no-reference image quality assessment[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 19-20, 2022, New Orleans, LA, USA. New York: IEEE, 2022: 1190-1199.



第一作者:赵柯嘉(2003—),男,博士研究生,研究方向为低空智能感知。E-mail:230259113@seu.edu.cn。



通信作者:孙一铭(1996—),男,助理研究员,研究方向为无人机具身智能、多模态协同学习。E-mail:sunyiming@seu.edu.cn。

(责任编辑:姜红贵)