

## 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, 叠加无人机平台的算力约束与高空复杂物理环境, 严重制约了现有算法的鲁棒性与边缘端实时性。对此, 本文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。本文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。为推动低空智能感知技术的持续演进, 本文进一步指出未来需重点突破多模态协同、无/自监督学习等方向, 推动低空智能感知技术的持续演进。

**关键词:** 低空底层视觉; 无人机感知; 图像恢复; 图像增强; 多源融合。

中图分类号: TP391.4

文献标志码: A

## 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 ability of unmanned aerial vehicles in complex environments. However, the coupled degradation problems such as motion blur, meteorological disturbances, and insufficient lighting in low-altitude scenarios, combined with the computational constraints of the unmanned aerial vehicle platform and the complex physical environment at high altitudes, severely restrict the robustness and 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 in adverse weather conditions, low-light enhancement, and multi-source fusion, but also systematically summarizes the existing quantitative evaluation systems. To promote the continuous evolution of low-altitude intelligent perception technology, this paper further points out that future efforts should focus on key breakthroughs in multi-modal collaboration, unsupervised/semi-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

随着低空经济的蓬勃发展, 无人机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也日益凸显。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、低光增强及多源融合等前沿方法的技术特点与发展趋势。本文通过构建“退化恢复-信息增强-评估验证”的完整技术体系分析框架, 为研究者提供清晰的理论参考, 并前瞻性地指出多模态协同、无/自监督学习等未来重点突破方向。

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(23JCJJC00270); 中国博士后科学基金国家资助博士后研究人员计划(GZB20250395); 江苏省卓越博士后计划(2025ZB294); 浙江省自然科学基金项目(LD24F020004)



### 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“评估框架-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其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，而无人机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<sup>[1]</sup>。

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。这些需求驱动算法需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衰减。本章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。

### 2.1 超分辨率重建

超分辨率重建（Super-Resolution Reconstruction, 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任务。在低空无人机领域，因多种因素影响，无人机航拍图像常分辨率不足、细节模糊，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，Transformer 架构在超分领域展现突破性潜力。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。与单图超分不同，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，天津大学任冬伟等人提出了 ST-AVSR 算法，创新性地引入结构与纹理先验信息。

无人机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，MFSRCNN（Multi-branch FSRCNN）<sup>[3]</sup>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。LSwinSR（Lightweight SwinSR）<sup>[4]</sup>则结合 Swin

Transformer 的窗口注意力与线性复杂度设计, 在嵌入式平台保持全局建模能力。为进一步压缩模型, RFDN (Residual Feature Distillation Network) 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。另一方面, 无人机图像常受大气扰动与运动模糊影响, 频域分析成为重要解决思路。SwinFIR-UAV 在 SwinIR 基础上增加快速傅里叶卷积模块, 其跨阶段融合机制分离处理低频特征与高频残差, 农田场景重建 SSIM 达 0.921, 病害识别准确率提升 11%。RingMo-SR 基于无人机基座模型 RingMo-Aerial, 通过频域特征增强与仿射对比学习预训练, 结合适配器微调实现超分-分割多任务协同, UAVDT 数据集多目标追踪 MOTA 提升 3.1%。

## 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、高质内容的过程, 其核心在于纠正图像/视频在形成过程中的失真。

### 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朦胧图像时, 其性能会急剧下降。为了克服这一挑战, 目前已经提出了一些用于现实世界图像除雾的工作。如, 刘东等人提出 PSD 去雾模型, 采用三个统计图像先验进行无监督微调, 而武睿祺等人在预训练的 VQGAN (Vector Quantized Generative Adversarial Network) 中利用了潜在离散先验。

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, 导致各个帧的去雾结果不连续。徐家奇等人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, 樊俊凯等人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。

无人机场景的去雾技术需应对高空成像的亮度不均问题。Su 等<sup>[5]</sup>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, 再结合改进的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 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<sup>[6]</sup>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。

纵观上述代表性研究, 基于物理先验的大气散射模型方法运算复杂度相对较低且泛化表现稳定, 十分契合算力受限的微型无人机平台, 但在遭遇浓雾或低空光照剧烈交变时容易发生透射率估计失效与色彩局部失真。相比之下, 基于深度学习的数据驱动框架能够大幅提升恶劣天气下复原图像的边缘细节与视觉保真度, 不过其高昂的端侧算力开销以及对特定训练数据分布的强依赖, 依然是制约其在低空开放场景中实现全天候部署的核心瓶颈。

### 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去雨性能。任冬伟等人提出的渐进式残差网络（Progressive ResNet, PRN）与渐进式循环网络（Progressive Recurrent Network, 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。JORDAR 创新性地提出雨纹-背景解耦翻译框架（Joint Rain Generation and Removal, 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重要基础。长序列一致性优化是视频去雨的关键挑战之一。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，使动态场景闪烁伪影减少 60%。此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。

无人机图像去雨任务需兼顾高频细节保留与算力约束。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，1.1M 参数量下 Jetson TX2 延迟仅 18ms。无监督方法 CCP 利用雨纹 RGB 通道强度差异，构建通道间差分损失，自编码器实现无标签训练，解决标注数据匮乏问题。Coderainnet<sup>[8]</sup>通过协同去雨与目标检测模块实现闭环优化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提升 19%。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，GSDerainNet<sup>[9]</sup>先通过高斯-香农滤波器初步抑制高频雨纹，再利用 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，使雨区 PSNR 提升 2.3dB，适配无人机低空成像的细节识别需求。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，DerainGAN<sup>[10]</sup>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，FID 指标较传统 GAN 降低 15%。

### 2.2.3 去雪任务

去雪任务目标是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。Deep Unfolding with Snow Shape Prior 率先将雪花形态学特征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方向。面对局部积雪与飘雪共存的挑战，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，标志从全局处理到区域自适应的重要转变。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，Video Desnowner 融合可变形卷积与 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。

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<sup>[11]</sup>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至 5%以内。Belmonte 等<sup>[12]</sup>对不同时段的无人机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数据支持。Gupta 等在

雪地地形的无人机-地面机器人协同导航中,对无人机获取的雪地图像进行去雪处理,提升地形特征清晰度,为自主导航提供可靠环境感知数据,增强复杂雪地场景中的导航鲁棒性。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,电力巡检视频 SSIM 达 0.896。该研究为真实场景落地提供有效解决方案。

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,OneRestore 提出了一种基于 Transformer 的通用恢复框架。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。

### 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、高质内容的过程。无人机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。

Deblurring by Realistic Blurring 创新性提出“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,在 RealBlur 数据集 SSIM 达 0.885,大幅提升模型泛化能力。视频去模糊需解决时序关联性。近年来 Transformer 展现出显著优势,VDTR<sup>[13]</sup>首次将其引入视频去模糊: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,在 REDS 数据集 PSNR 达 34.2 dB,刷新性能记录。

无人机图像面临高频丢失与算力约束双重挑战。Progressive Residual Recurrent Network (PRRN)<sup>[14]</sup>设计轻量编码器-解码器模块,门控渐进残差块减少 40%计算量,特征循环结构抑制模糊残留,在 AID 数据集 PSNR 达 30.80 dB,Jetson TX2 平台实现 35fps 实时处理。针对特定任务的无人机去模糊注重场景语义保持。SlimDeblurGAN<sup>[15]</sup>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,保持 92%的结构相似性,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。高速移动无人机的去模糊需针对性优化。Liu 等<sup>[16]</sup>引入运动模糊感知因子调整去模糊强度,在高速公路航拍中抑制了因快速移动导致的远景模糊,使道路检测准确率提升 9%。融合超分与去模糊的无人机图像处理更注重细节增强。SR-DeblurUGAN<sup>[17]</sup>通过双分支网络学习超分特征与模糊核,在农田航拍中使作物分类准确率提升 8%-12%,体现了多任务融合的优势。

#### 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。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,在多个去噪基准上实现了卓越的性能与效率平衡。

在无人机场景中,去噪技术为后续任务提供质量保障。Tian 等将去噪预处理融入无人机绝缘子检测

流程,通过抑制航拍图像中的随机噪声,提升 DETR 算法对绝缘子的识别精度,使检测准确率提高 6.3%。堆叠剪枝稀疏去噪自编码器与卷积神经网络结合的方法<sup>[18]</sup>,在无人机叶片损伤诊断中,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,使故障识别率提升至 98.2%。ID-SAbRUNet<sup>[19]</sup>作为针对无人机逆合成孔径雷达 (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,有效提升了无人机 ISAR 图像的解译精度。Azmat 等<sup>[20]</sup>将去噪技术融入无人机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,使识别准确率提高 5.2%。

## 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,核心是在可接受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。AGDL 创新融合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 数据集人眼关注区域的 PSNR 提升 1.2dB。在此基础上, SlimCAE 提出可切换压缩比架构,单个模型通过权重共享支持多档码率动态调整,在移动端部署时模型体积减少 70%而 MS-SSIM 仅下降 0.003。视频压缩需平衡时域依赖与计算复杂度。HLVC 设计三层级增强框架: I 帧压缩层初始化,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,实现 17% BD-Rate 节省。

无人机采集的高清视觉数据直接传输带宽压力巨大,通过高效压缩可降低码率,保障了 5G 回传的实时性与稳定性,满足电力巡检、灾害监测的迫切需求。这些应用显著降低了通信与存储成本,推动了无人机视联网的规模化落地。因此无人机场景的图像时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通过整合特征提取与压缩编码模块,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,适配无人机有限的传输带宽。将深度学习方法应用于无人机视频压缩,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,在同等码率下使重构视频的 PSNR 提升 2dB。Smirnov 等对适用于遥控无人机的视频压缩算法进行分析与优化,发现针对无人机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。面向无人机实时视频分析的自适应兼容压缩方法<sup>[21]</sup>,通过动态调整压缩参数适配不同场景需求——在目标检测场景保留更多细节,在大范围监控场景侧重提升压缩比,实现压缩效率与分析精度的平衡。

## 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。本章重点探讨低光增强和多源融合两大方向。

### 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。MIRNet 提出多路径特征融合架构,通过残差稠密块聚合局部与全局特征,结合通道注意力机制优化噪声与模糊的协同去除,在 RealBlur-J 数据集 PSNR 达 28.9 dB。进一步地, 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。在单图增强基础上, She 等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。

无人机常面临强烈光照变化、薄雾/霾导致对比度下降、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<sup>[22]</sup>将视频修改技术应用于无人机影像,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,显著增强无人机航拍影像中建筑物、道路等目标的可识别性。Sun 等<sup>[23]</sup>融合时空信息与光流特征增强监控视频中的无人机目标,抑制复杂背景干扰,使无人机检测准确率提升至 92.3%,拓展了图像增强技术在目标检测下游任务的应用。

### 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信息更丰富、更可靠的输出。

2019 年马佳义等<sup>[25]</sup>开发了 FusionGAN 融合模型,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,并通过判别器对抗性优化提升融合图像的自然感知质量。为解决红外-可见光融合中的亮度不匹配问题,孙一铭等<sup>[26]</sup>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。这些方法丢弃了固有的时间信息,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 年马佳义等提出了一种名为 VideoFusion 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,从多模态输入生成时空连贯的视频。

在无人机场景下,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<sup>[27]</sup>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无人机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 等的 DREB-net 针对高机动性场景设计双流融合网络,左流恢复运动模糊图像特征,右流提取模糊区域边缘梯度,经注意力加权后解决传统方法对运动模糊的敏感性,使高速飞行时的目标检测准确率较 VL-MFL 直接融合提升 9%,实现了融合策略对动态场景的适配。孙一铭等<sup>[28]</sup>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。

整体而言,现阶段低空多源融合技术呈现出明显的路线分化。其中生成对抗网络范式在凸显红外热目标与保留可见光背景纹理上极具优势,非常适合无人机夜间巡视或低能见度搜救任务,但极易受限于对抗训练的不稳定而产生空间局部伪影。相对而言,基于任务驱动与时空联合建模的融合架构虽然在应对无人机高速飞行与复杂退化耦合场景时展现出更强的特征鲁棒性与时序连贯性,但这类深度网络往往伴随较高的推理延时,对机载微型计算平台的边缘处理能力提出了严峻考验。

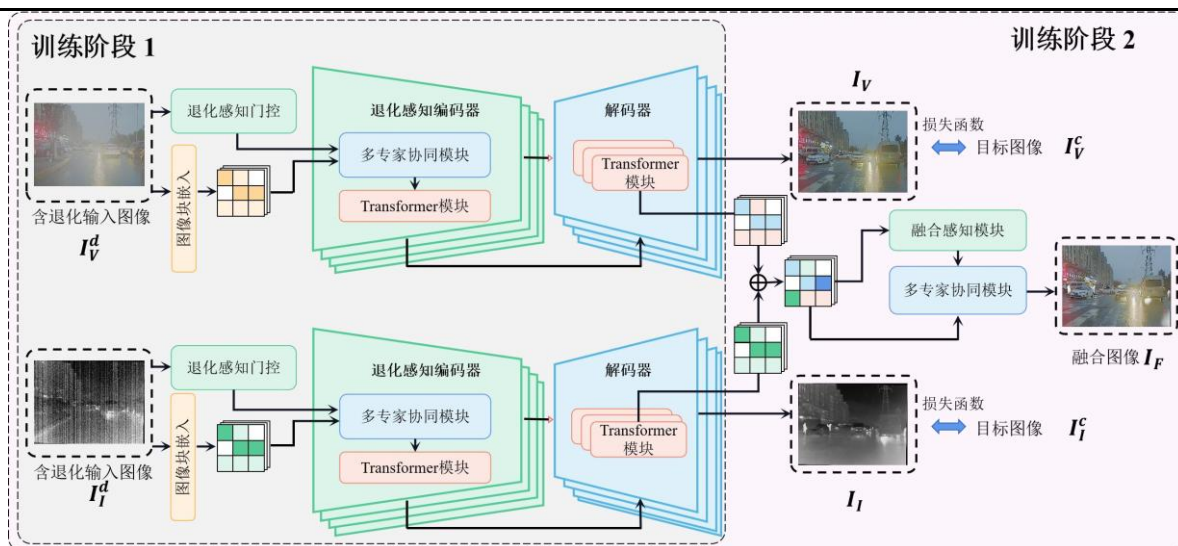


图 2 TG-ECNet 框架图<sup>[28]</sup>

Fig.2 Framework Diagram of TG-ECNet<sup>[28]</sup>

## 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。本章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，对现有的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，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。

TreS 是 2021 年 Sajad Golestaneh 等人提出的针对无参考图情况的图像评价框架，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，该设计能够避免 CNN 的局部感受野限制，提升对复杂失真的建模能力。模型采用自一致性约束，增强模型对局部失真变化的鲁棒性，避免过拟合特定区域。该框架在跨数据集测试中泛化性提升显著。

MANIQA 是 2022 年由杨思谛等人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 Vision Transformer (ViT-B/8) 作为特征提取层以提取多尺度特征，最后网络设计有双分支预测结构，模拟人眼对图像不同区域的关注差异，令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，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。底层视觉常用的评估指标分类情况如表 1 所示。

表 1 底层视觉任务评估指标分类

Tab.1 Classification of Baseline Visual Task Evaluation Metrics

任务类型	典型指标名称	
退化恢复类任务	超分辨率重建	PSNR SSIM VIF MSE
	去雾任务	PSNR SSIM MSE VCM
	去雪任务	PSNR SSIM
	去雨任务	PSNR SSIM RMSE
	去模糊任务	PSNR SSIM LPIPS MS-SSIM
	去噪任务	PSNR SSIM
	修复任务	PSNR SSIM FID MAE MSE
	压缩任务	PSNR SSIM CC
信息增强类任务	增强任务	PSNR SSIM EN CC LPIPS FID
	多源融合	CC MSE PSNR SSIM SF VIF

## 5 总结与展望

低空底层视觉技术作为无人机智能感知的基石，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下的性能衰减、边缘端多任务实时处理能力的不足，以及多源数据时空对齐误差。此外，特定任务高质量基准数据集的匮乏，亦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进化。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，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，将助力低空系统在无先验场景下实现高鲁棒的通用感知，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 Huajun, Wang Yuqi. 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 Yiming, Zhao Kejia, Wang Shuo, 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] 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 Zhipeng, Zhao Changming, Zhang Haiyang, et al. Application of multi-scale fusion super-resolution algorithm in UAV detection [J]. *Journal of Applied Optics*, 2021, 42 (3): 462-473.
- [4] Li Rui, Zhao Xiaowei. LSwinSR: UAV imagery super-resolution based on linear swin transformer [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 5641013.
- [5] 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.
- [6] Qiu Zifeng, Gong Tianyu, Liang Zichao, et al. Perception-oriented UAV image dehazing based on super-pixel scene prior [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 5913519.
- [7] Cui Xin, Wang Cong, Ren Dongwei, et al. Semi-supervised image deraining using knowledge distillation [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32 (12): 8327-8341.
- [8] Xi Yue, Jia Wenjing, Miao Qiguang, et al. CoDerainNet: collaborative deraining network for drone-view object detection in rainy weather conditions [J]. *Remote Sensing*, 2023, 15 (6): 1487.
- [9] Yao Yanji, Shi Zhimin, Hu Huiwen, et al. GSDerainNet: a deep network architecture based on a Gaussian Shannon filter for single image deraining [J]. *Remote Sensing*, 2023, 15 (19): 4825.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] Cao Mingdeng, Fan Yanbo, Zhang Yong, et al. VDTR: video deblurring with transformer [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023, 33 (1): 160-171.
- [14] Zhang Xiaoqin, Jiang Runhua, Wang Tao, et al. Recursive neural network for video deblurring [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021, 31 (8): 3025-3036.
- [15] 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.
- [16] Liu Yan, Wang Jingwen, Qiu Tiantian, et al. An adaptive deblurring vehicle detection method for high-speed moving drones: resistance to shake [J]. *Entropy*, 2021, 23 (10): 1358.
- [17] Xiao Yuzhen, Zhang Jidong, Chen Wei, et al. SR-DeblurUGAN: an end-to-end super-resolution and deblurring model with high performance [J]. *Drones*, 2022, 6 (7): 162.
- [18] Yang Pu, Wen Chenwan, Geng Huilin, 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.
- [19] 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.

- [20] 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.
- [21] Smirnov K, Mozhaeva A. Analysis of video data compression algorithms and optimized for use in remotely controlled drones [J] . Synchroninfo Journal, 2023, 9 (2): 9-16.
- [22] Reale M J, Murphy D, Cornacchia M, et al. Video modification in drone and satellite imagery[C]//Disruptive Technologies in Information Sciences VIII. SPIE, 2024, 13058: 391-400.
- [23] SUN Y, ZHI X, HAN H, et al. Enhancing UAV detection in surveillance camera videos through spatiotemporal information and optical flow[J]. Sensors, 2023, 23(13): 6037.
- [24] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.
- [25] MA J, YU W, LIANG P, et al. FusionGAN: A generative adversarial network for infrared and visible image fusion[J]. Information fusion, 2019, 48: 11-26.
- [26] Sun Y, Cao B, Zhu P, et al. Dynamic brightness adaptation for robust multi-modal image fusion[C]//Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence. 2024: 1317-1325.
- [27] LIU G, LI C, ZHANG S, et al. VL-MFL: UAV visual localization based on multisource image feature learning[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 1-12.
- [28] Sun Y, Li X, Zhu P, et al. Task-gated multi-expert collaboration network for degraded multi-modal image fusion[C]//International Conference on Machine Learning. PMLR, 2025: 57571-57586.