

文章编号 1004-924X(2026)16-2523-13

基于零样本学习的核电站水下视频去噪方法

刘颖^{1*}, 杨杰^{1,2}, 罗杰^{1,2}, 刘顺^{1,2}, 李天昊¹, 肖晓南¹, 汪昱成²

(1. 苏州天河中电电力工程技术有限公司, 江苏 苏州 215028;

2. 国核电站运行服务技术有限公司, 上海 200233)

摘要: 针对真实核电站水下环境中高能粒子辐射噪声引起的视频图像严重退化问题, 提出一种基于零样本学习的水下视频去噪方法, 以降低对成对清晰样本和专门训练数据的依赖, 并提升水下检修视觉信息的可用性。以预训练图像复原扩散模型为基础, 对低质量噪声视频进行分批处理, 并选取关键帧作为参考输入; 在去噪早期, 通过层次化潜在变形与光流对齐建立全局结构约束和局部运动约束; 在去噪中后期, 引入混合流引导的空间感知标记合并策略, 在下采样阶段利用光流匹配保持运动一致性, 在上采样阶段结合空间约束相似度保留结构细节, 并采用合并比例退火机制动态平衡时间一致性与细节保真。在采集的真实核电站水下噪声视频数据集上, 所提方法的 MaxVQA, CLIPVQA 和 E_{warp} 分别达到 34.2, 0.97 和 1.2, 优于对比视频复原与去噪方法。该方法可在缺乏标注训练数据和清晰参考视频的核电站水下场景中实现有效视频去噪, 为水下光学成像质量提升、图像智能分析和设备缺陷检测提供技术支撑。

关键词: 核辐射; 零样本学习; 水下视频去噪; 扩散模型; 标记合并

中图分类号: TP394.1; TH691.9 **文献标识码:** A

doi: 10.37188/OPE.20263416.2523

CSTR: 32169.14.OPE.20263416.2523

Zero-shot underwater video denoising for nuclear power plant inspection

LIU Ying^{1*}, YANG Jie^{1,2}, LUO Jie^{1,2}, LIU Shun^{1,2}, LI Tianhao¹, XIAO Xiaonan¹, WANG Yucheng²

(1. Suzhou Tianhe Zhongdian Power Engineering Technology Co., Ltd., Suzhou 215028, China;

2. State Nuclear Power Plant Service Co., Ltd., Shanghai 200233, China)

* Corresponding author, E-mail: ahtly0830@163.com

Abstract: To address severe video degradation caused by high-energy particle radiation in real underwater environments of nuclear power plants, this study proposed a zero-shot underwater video denoising method that reduced the dependence on paired clean samples and task-specific training data, and improved the usability of visual information for underwater inspection. A pre-trained image restoration diffusion model was used as the backbone. Low-quality noisy videos were processed in batches, and key frames were selected as reference inputs. In the early denoising stage, hierarchical latent deformation and optical-flow alignment were used to impose global structural constraints and local motion constraints. In the middle and later stages, a hybrid flow-guided spatial-aware token merging strategy was introduced. Optical-flow matching was used in the downsampling stage to maintain motion consistency, spatially constrained similarity was used

收稿日期: 2026-06-08; 修订日期: 2026-07-20.

基金项目: 江苏省青年科技人才托举工程项目 (No. JSTJ-2025-560)

in the upsampling stage to preserve structural details, and a merging-ratio annealing mechanism was adopted to dynamically balance temporal consistency and detail fidelity. On the collected underwater noisy video dataset from a nuclear power plant, the proposed method achieved MaxVQA, CLIPVQA, and Ewarp values of 34.2, 0.97, and 1.2, respectively, outperforming the compared video restoration and denoising methods. The proposed method enables effective video denoising in underwater nuclear power plant scenarios where annotated training data and clean reference videos are difficult to obtain, and provides technical support for underwater optical imaging enhancement, intelligent image analysis, and equipment defect detection.

Key words: nuclear radiation; zero-shot learning; underwater video denoising; diffusion model; token merging

1 引言

在全球能源结构加速重构与气候治理紧迫性日益凸显的双重背景下,核能产业已成为各国破解能源安全困局、实现碳中和目标的关键战略支点。作为全球最大能源消费国和清洁能源投资主体,中国核电产业实现了跨越式发展:从1985年秦山核电站破土动工填补国内空白,到2024年建成在运及核准在建核电机组总装机容量突破1.13亿千瓦,装机规模跃居全球首位,形成了完整的自主三代核电技术体系^[1]。而在核电站发展事业中,运维检修是保障核能安全高效利用的核心环节。通过预防性维护、智能监测与精准检修,可及时消除设备隐患,显著降低机组故障率,确保反应堆安全屏障万无一失;同时优化机组性能,提升发电效率并延长设备寿命。

在现代核电站设计中,水因其出色的冷却性能与辐射屏蔽能力,成为不可或缺的介质,核燃料转运、反应堆运行及乏燃料贮存等关键流程均需在水下完成。然而,站内水体环境复杂,常具有酸性强、辐射剂量高等特点,致使工作人员无法直接进入现场作业。因此,搭载图像采集模块(通常包含CCD或CMOS摄像头)的检修机器人,在日常风险排查中承担重要作用^[2]。然而,核电站运行中产生的大量高能粒子(如中子, α 、 β 、 γ 射线^[3]等),也给机器人的视觉系统带来了严峻挑战。这些粒子进入图像采集模块后,会与CMOS阵列或CCD器件发生相互作用,引发剂量效应与单粒子效应^[4],导致感光电荷异常增加,从而使采集到的图像中出现大量强烈且不规则的斑块噪声。受此影响,机器人所获图像质量显著下

降,设备缺陷细节被噪声掩盖,最终降低了其在强辐射环境下的精准感知与故障识别能力。

与一般自然图像去噪相比,核电站水下视频退化具有明显的场景特殊性。一方面,水体对光线的吸收和散射会造成低对比度、照明不均及细节衰减;另一方面,高能粒子与CCD/CMOS感光器件相互作用,会引起局部感光电荷异常,形成位置、强度和尺度均不规则的高亮斑点、条带或斑块噪声。因此,该场景并非单一平稳噪声退化,而是水下光学退化与辐射诱发非规则噪声共同作用的复合退化。与此同时,辐射噪声在连续帧中的随机出现容易干扰特征匹配和运动估计。

为应对上述核辐射环境导致的图像退化问题,并保障核电站水下检修工作顺利开展,研究人员主要聚焦于两类技术路径:抗辐射加固技术与抗辐射图像去噪技术,以缓解高能粒子冲击引起的图像噪声退化。其中,抗辐射加固技术主要基于集成电路版图设计层面的加固策略,旨在提升图像采集模块自身的抗单粒子干扰能力;而抗辐射图像去噪技术,则通过分析辐射噪声的形成机制与特征,运用图像复原方法在保留关键图像结构与边缘信息的同时提升图像质量。由于加固技术通常成本较高、工艺复杂,相比之下,图像去噪技术更具有经济可行性与广泛的学术研究价值。

在抗辐射图像去噪研究领域,早期的传统方法主要通过空间域内借助局部或非局部滤波器,利用相邻像素或整幅图像像素之间的相似性来去除噪声。例如,王姮等^[5]提出了一种结合自适应中值滤波与小波变换的核辐射图像去噪方法。为进一步利用退化图像的全局信息,Dov-

ganich等^[6]采用非局部均值滤波算法对核辐射退化图像进行处理,以更好地保留图像边缘结构与细节信息。除滤波方法外,也有研究基于稀疏表示的正则化框架进行图像去噪。例如,Sagheer等^[7]将全变分正则化引入辐射图像去噪中;黄小莉等^[8]则融合非凸二阶全变分与重叠组稀疏正则化,提出一种混合二阶全变分的抗核辐射图像去噪方法。针对核辐射斑块噪声的冲击脉冲特性,陈明举等^[9]通过最大后验概率估计推导出相应的保真项与正则化项,构建了一种基于变分模型的斑块噪声去除方法。然而,这些传统方法大多依赖于人工设计的先验知识,在高噪声水平下去噪性能有限,且通常需针对不同图像手动调整参数,因而难以适应多样化的实际应用场景。

传统图像去噪方法常因泛化能力有限,在处理未知样本时复原效果难以达到预期。随着人工智能技术的快速发展,基于深度学习的图像去噪方法应运而生,并取得了显著进展。例如,孙跃文等^[10]将残差神经网络结构引入核噪声图像的去除中,通过训练辐射图像样本建立噪声图像与无干扰图像间的映射关系,从而实现对核辐射图像的噪声抑制;针对核辐射噪声图像所特有的强噪声干扰问题,孙鑫等^[11]设计了一种包含噪声学习单元与纹理学习单元的双分支去噪网络,在提升去噪效率的同时,有效保留了图像的细节纹理信息。黄小莉等^[12]提出了一种结合并行多尺度特征提取与残差特征提取的降噪网络,并引入空间注意力机制以增强特征表征,从而显著提升了噪声图像的视觉质量。此外,鉴于真实核辐射场景中噪声图像样本稀缺,谢明洁等^[13]采用生成对抗网络学习核辐射图像的噪声分布,进而生成清晰图像所对应的合成噪声退化图像,为模型训练提供了丰富的配对数据。针对现有去噪网络在高噪声水平下清晰度不足的问题,谢

明洁等^[13]随即又提出了一种融合边缘信息的上下迭代采样辐射噪声图像去噪网络。然而,现有深度去噪方法大多依赖成对清晰图像或特定场景数据训练。在真实核电站水下辐射环境中,场景条件复杂、噪声类型难以预先建模,且清晰参考图像难以获取,高质量标注数据严重不足。因此,如何在缺乏额外训练数据的条件下实现稳定的视频去噪,仍是核电站水下智能视觉应

用面临的关键问题。

近年来,扩散模型(Diffusion Model)凭借其强大的生成建模能力,在图像复原领域取得了突破性进展。与传统判别式去噪网络^[13]不同,扩散模型通过逐步学习噪声分布的逆扩散过程,可在复杂退化条件下恢复更加自然且细节丰富的图像。Saharia等^[14]提出的SR3首次证明了扩散模型能够有效用于图像超分辨率与去噪任务;随后,IR-SDE^[15]利用随机微分方程统一描述图像退化与恢复过程,实现了真实图像去噪和去模糊;DiffBIR进一步结合预训练扩散先验,实现了无需针对特定退化重新训练即可完成盲图像去噪,在复杂未知退化场景下表现出优异的泛化能力^[16]。这些研究表明,扩散模型蕴含的通用复原先验有望拓展至核电站水下视频去噪任务。

除单帧图像复原外,近年来研究人员进一步利用视频帧间的运动关联和时序冗余提升视频复原性能。Liang等^[17]提出的VRT模型通过并行建模空间信息和时间信息实现视频去噪、去模糊和超分辨率等多类复原任务。Li等^[18]提出的Shift-Net利用分组时空移位操作促进跨帧信息交互,在较低计算复杂度下实现视频复原。Ghasemabadi等^[19]通过截断因果历史建模利用前序帧信息设计了TURTLE模型,提高了长视频序列的复原效率和时间稳定性。针对视频噪声退化,Fu等^[20]将时间建模模块作为插件引入预训练图像去噪器,在不依赖成对视频训练数据的条件下实现无监督视频去噪。Yue等^[21]提出的RViDeformer利用多分支空间—时间注意力对原始视频中的空间噪声和帧间相关性进行联合建模,增强了复杂噪声条件下的视频去噪能力。上述视频复原与去噪方法主要面向高斯噪声、真实传感器噪声、模糊和低分辨率等通用退化,其与核电站水下场景中辐射诱发强斑块噪声、水下光学退化相差甚远。

从工程应用角度看,现有辐射去噪方法还多聚焦于单帧图像或离线处理,对水下拍摄过程中的视频时序一致性关注不足。如何充分利用预训练图像复原先验,同时保持跨帧一致性和局部结构细节,仍是核电站水下视频去噪面临的关键问题。视频编辑方法VidToMe^[22]提出了基于标记合并的策略,通过在扩散模型内部合并具有相

似语义的视觉标记,在保持视频时间一致性方面取得了较好的效果。受此启发,本文在现有扩散模型框架基础上,以标记合并策略为核心切入点,对视频去噪过程中标记对应关系建立、跨帧特征传播及不同去噪阶段的调度机制进行改进,从而提升真实核电站水下噪声视频的时序一致性和细节恢复能力。

对于核电站水下巡检而言,图像复原不仅关系到单帧视觉清晰度,还直接影响后续缺陷识别、目标定位和状态判读的可靠性。为此,本文基于真实核电站水下视频数据,提出一种基于扩散模型的零样本水下视频去噪网络(Zero-Shot Underwater Video Denoising Network, ZS-UVDN)。本文主要贡献如下:

(1)实现预训练图像复原扩散模型向水下核辐射视频去噪任务的零样本迁移,无需额外训练即可兼顾时序连贯性与空间细节保真度;

(2)提出层次化潜在变形机制与混合流引导的空间感知标记合并策略,在不同去噪阶段分别强化全局结构约束、局部运动对齐和细节保持;

(3)基于真实核电站水下采集视频开展验证,结果表明本文方法较现有视频复原与去噪方

法具有更好的感知质量、时间一致性和场景适应性。

2 ZS-UVDN 网络构建

2.1 网络总体描述

核电站水下辐射噪声具有局部强响应和非规则分布等特点,易污染视觉特征并干扰跨帧匹配;同时,受现场作业条件限制,成对清晰视频难以获取。为此,本文提出零样本水下视频去噪网络 ZS-UVDN,以预训练图像复原扩散模型为基础,无需额外训练,从而降低对特定噪声模型和现场成对数据的依赖。

如图 1 所示,ZS-UVDN 采用分批处理方式,并选取关键帧作为参考。在去噪初期,通过关键帧间变形和批次内特征传播构建层次化潜在变形,以建立全局与局部结构约束;在主体去噪阶段,引入混合流引导的空间感知标记合并,分别利用光流和空间约束相似度完成下采样与上采样特征对齐。同时,采用合并比例退火机制,逐步实现由时间一致性保持向细节恢复的过渡,从而兼顾异常噪声抑制、跨帧稳定性和局部结构保真。

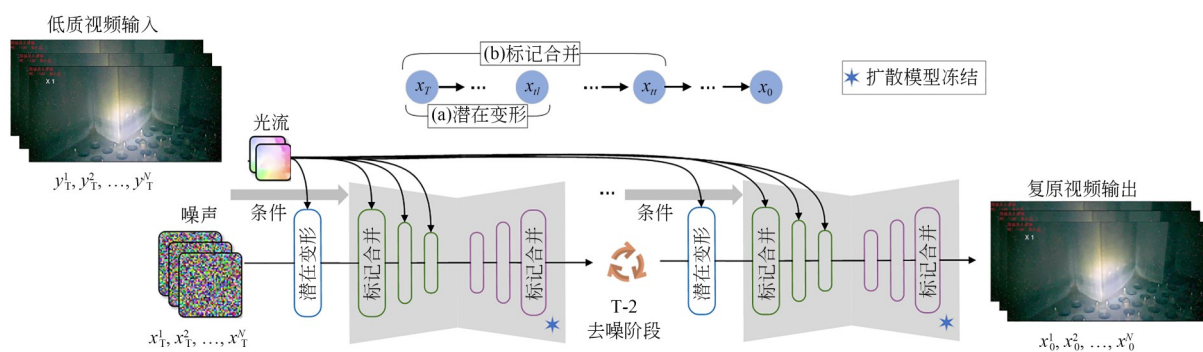


图 1 零样本学习水下视频去噪方法框图

Fig. 1 Block diagram of the zero-shot learning underwater video denoising method

2.2 用于视频编辑的扩散模型

扩散模型的前向过程通常是在 T 个时间步内逐步向干净图像 x_0 添加噪声:

$$\begin{aligned} x_t &= \sqrt{\alpha_t} x_{t-1} + \sqrt{1 - \alpha_t} \epsilon_{t-1} \Rightarrow \\ x_t &= \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon_t, \end{aligned} \quad (1)$$

其中: $t \sim [1, T]$, $\epsilon_t, \epsilon \sim N(0, I)$, $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$ 。基于 U-Net 结构的去噪器 ϵ_θ 能够学习估计并去

除噪声,其逆向过程对 x_t 逐步去噪,最终恢复得到 x_0 [23-24]。

近期视频编辑方法(如 VidToMe [22])通过在帧块内合并相似标记来保持时间一致性。对于给定标记块 $T \in \mathbf{R}^{B \times A \times C}$ (其中 $A = \omega \times h$),标记被划分为源标记 T_{src} 和目标标记 T_{tar} 。标记间相似度计算如式(2)所示:

$$s(T_{\text{src}}, T_{\text{tar}}) = \frac{T_{\text{src}} \cdot T_{\text{tar}}}{\|T_{\text{src}}\| \|T_{\text{tar}}\|}, c = \max_{\{t \in T_{\text{tar}}\}} (s(T_{\text{src}}, t)), \quad (2)$$

其中: $s(\cdot, \cdot)$ 表示余弦相似度, c 表示对应关系。标记合并与解合并操作定义如式(3)所示:

$$T_{\text{merge}} = M(T_{\text{src}}, T_{\text{tar}}, c, r), T_{\text{unmerge}} = u(T_{\text{merge}}, c). \quad (3)$$

然而,现有标记合并策略直接用于核电站水下辐射噪声视频仍存在局限:辐射诱发的局部强响应和非规则斑块易污染潜在特征,导致下采样阶段基于余弦相似度的标记匹配产生错误对应;仅依赖局部相似性或单尺度传播,难以兼顾全局结构与局部运动一致性;持续采用较高合并比例还可能传播异常特征并造成设备边缘和缺陷纹理过度平滑。为此,本文设计层次化潜在变形、混合流引导的空间感知标记合并及合并比例退火策略。

2.3 层次化潜在变形

考虑到真实设备结构在连续帧间具有较强的时序连续性,而局部辐射异常的位置、强度和形态具有较强随机性,本文设计层次化潜在变形模块,以利用跨帧稳定结构抑制随机异常对复原结果的干扰。该模块在潜在空间的两个层级协同工作:第一层级通过关键帧间变形优化视频全局一致性,第二层级将变形后的潜在特征传播至当前处理批次内的各帧,以保障局部一致性。如图2所示,该分层结构可同时提供全局和局部尺度的形状引导。对于去噪步骤 t 下第 i 个关键帧的预测潜在特征 $\hat{x}_{t \rightarrow 0}^i$,首先通过关键帧变形建立全局一致性:

$$\hat{x}_{t \rightarrow 0}^i \leftarrow M_{ji} \cdot \hat{x}_{t \rightarrow 0}^j + (1 - M_{ji}) \cdot W(\hat{x}_{t \rightarrow 0}^j, f_{ji}), \quad (4)$$

其中: $j = i - 1$ 代表前一关键帧。光流场 f_{ji} 和遮

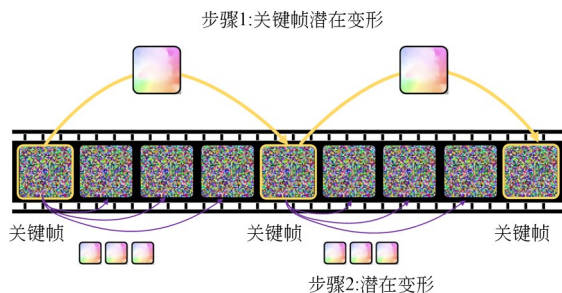


图2 层次化潜在变形示意图

Fig. 2 Schematic diagram of hierarchical latent deformation

挡掩码 M_{ji} 由GMFlow^[25]从低质量帧 l_{qj} 计算至低质量帧 l_{qi} 计算得到。该公式将原始潜在特征与前一关键帧的变形特征进行融合,并由遮挡掩码动态调节融合权重,以降低遮挡区域或变形不可靠区域对复原结果的影响。

完成全局对齐后,本文将变形后的潜在特征传播至当前批次的所有帧,以构建局部一致性。与仅依赖相邻帧变形的的方法相比,所提层次化策略可在去噪初期使对应点在全局和局部时间尺度上保持相近的潜在特征表示,从而提升复杂运动场景下的视频结构连贯性。

针对严重退化区域可能产生的变形误差,本文引入前后向一致性检查和选择性特征传播机制。该设计可在光流估计失效或遮挡发生时降低错误传播风险,使复原过程较单尺度变形方法具有更强鲁棒性。实验结果表明,该层次化策略能够降低时间伪影,并有效保留复原视频中的细节纹理。

2.4 混合流引导的空间感知标记合并

层次化潜在变形能够在去噪早期利用跨帧稳定结构抑制随机辐射异常,但若在中后期持续对潜在特征进行强制传播,可能造成异常信息扩散和细节模糊。为此,本文提出混合流引导的空间感知标记合并方法,在语义信息更丰富的特征块空间中平衡时间一致性与细节保留。其核心思想是:在不同去噪阶段和不同网络深度中,采用与特征可靠性相匹配的对应关系建模方式。

在去噪初期,潜在特征受噪声影响较大,传统余弦相似度度量尤其在U-Net下采样模块中容易失效。此时,基于低质量输入估计的光流可提供更稳定的运动引导。随着去噪过程推进,光流匹配与相似度匹配呈现互补关系,因此本文采用混合策略。如图3所示,针对下采样模块,本文结合光流引导的对应关系和前后向一致性检查,以提高匹配可靠性,计算公式如式(5)所示:

$$\sigma = e^{(-\|f_{\text{src} \rightarrow \text{tar}}(X(T_{\text{src}})) + f_{\text{tar} \rightarrow \text{src}}(X(T_{\text{src}})) + f_{\text{src} \rightarrow \text{tar}}(X(T_{\text{src}})))\|_2^2)}, \quad (5)$$

其中: σ 表示置信度得分, $X(T_{\text{src}})$ 表示源标记 T_{src} 的空间位置, $f_{\text{src} \rightarrow \text{tar}}, f_{\text{tar} \rightarrow \text{src}}$ 分别代表正向光流和反向光流。该置信度得分用于指导标记合并过程,如公式(6)所示:

$$T_{\text{merge}} = M(T_{\text{src}}, T_{\text{tar}}, f_{\text{src} \rightarrow \text{tar}}, \sigma, r). \quad (6)$$

针对上采样模块,本文通过增强空间感知机

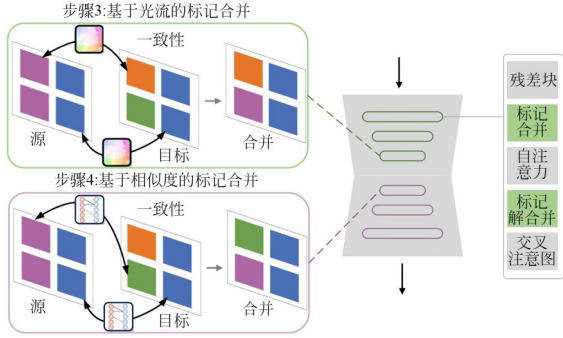


图 3 空间感知标记合并操作

Fig. 3 Spatial-aware token merging operation

制优化余弦相似度匹配,以减少纹理均匀区域的错误匹配。具体而言,采用基于空间邻近性的加权方式动态调整相似度得分:

$$s'_{ij} = s_{ij} \cdot e^{-\tau}, \text{ 其中 } \tau = \left\lfloor \frac{\|X(i) - X(j)\|_2^2}{R} \right\rfloor, \quad (7)$$

其中: $X(i), X(j)$ 分别表示标记的空间位置, R 为预设影响半径。为避免填充区域引入伪影,本文在特征块合并前剔除填充区域,并在解合并后恢复相应填充。为在去噪过程中兼顾复原质量与计算稳定性,本文进一步采用合并比例退火策略,即逐步降低合并比例:

$$r = r \cdot \cos\left(\frac{\pi}{2} \cdot \max\left(\min\left(\delta \cdot \frac{i - i_{\text{beg}}}{i_{\text{end}} - i_{\text{beg}}}, 1\right), 0\right)\right), \quad (8)$$

其中:参数 i_{beg} 和 i_{end} 分别表示退火周期的起始阶段和结束阶段,参数 δ 用于调控退火速度。该策略可在去噪早期增强时间一致性约束,并在后期逐步释放细节表达能力,从而避免过度平滑,并较逐帧处理方式更好地维持时间维度连贯性。

2.5 调度

本文方法在完整去噪过程中对各组件进行协同调度。早期阶段通过关键帧间特征变形和批次内特征传播,利用层次化潜在变形建立全局与局部一致性;主体去噪阶段在各注意力层前启用混合空间感知标记合并,并根据网络深度自适应调整匹配方式:下采样模块以光流引导为主,上采样模块以相似度匹配为主。同时,退火策略动态调节合并比例,使模型由一致性优先逐步转向细节保留优先。该渐进式调度既降低过度平滑风险,又增强了跨帧连贯性。

3 实验与分析

3.1 数据介绍及实验设置

目前,针对水下核辐射噪声视频的公开研究和数据资源仍较少。为验证本文算法,实验数据来源于本团队在国内某核电大修现场利用自研耐辐照摄像机采集的真实水下辐射噪声视频。受高能粒子影响,采集视频中单帧图像的典型噪声分布如图4所示。

真实核电站水下作业数据具有采集条件受限、噪声形态复杂和清晰参考视频难以同步获得等特点。受检修安全、辐射防护和现场作业窗口限制,本文不依赖成对清晰视频构建监督训练集,而是直接利用真实退化视频评估零样本复原能力。该设置更接近核电站水下巡检中的实际应用需求,也凸显了无需额外训练的工程意义。

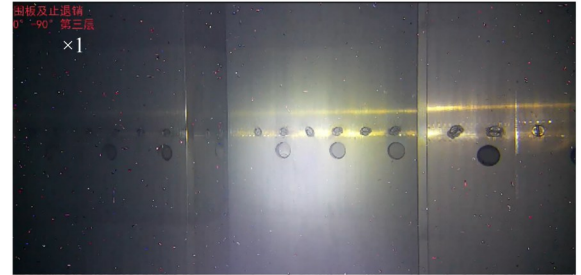
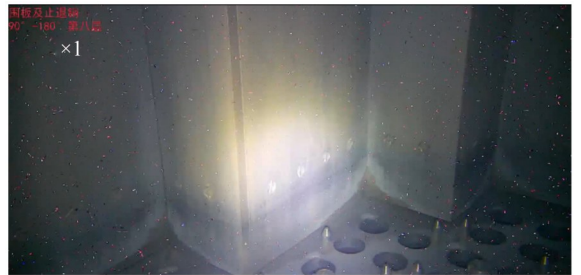
(a) 帧1
(a) Frame 1(b) 帧2
(b) Frame 2

图 4 水下辐射噪声视频中的任意两帧

Fig. 4 Two arbitrary frames from an underwater radiated noise video

本文采用PyTorch框架实现所提零样本水下视频去噪方法,实验在单张NVIDIA RTX 4090 GPU上完成,并基于CUDA 11.3环境部

署。扩散模型采用 DiffBIR 架构^[16],其扩散过程使用线性噪声调度策略:起始噪声强度设为 0.000 85,结束噪声强度设为 0.012 0,总扩散时间步数为 1 000。

为保证对比实验的可重复性与公平性,所有方法均在相同硬件环境和相同输入图像块尺寸下进行测试;对比算法采用公开实现或原文推荐参数,不针对本文数据额外训练或调参。由于真实场景缺少同条件清晰参考视频,本文采用无参考感知质量评价与时间一致性评价相结合的方式评估复原效果。

3.2 客观性能评价指标

考虑到核电站水下真实退化视频难以获得逐帧配准的清晰参考图像,本文采用互补客观指标评价复原视频的感知质量和时间一致性,以避免单一指标无法同时反映视觉质量和跨帧稳定性的问题:

(1) 感知质量指标采用 MaxVQA^[26] 和 CLIPVQA^[27]。MaxVQA 通过联合表征视频时空多维属性获得可解释的质量评价结果;CLIPVQA 利用大规模视觉-语言预训练模型的先验知识,对无参考视频质量评价具有较好的跨场景泛化能力。两者可从不同角度反映复原视频的视觉自然度和感知清晰度。

(2) 时间一致性指标采用形变误差 E_{warp} 。该指标用于量化视频运动估计或变形操作中预测位置与实际位置之间的偏差,常用于评

价运动补偿和时序对齐精度。 E_{warp} 数值越低,说明复原后视频在跨帧运动一致性方面越稳定。

3.3 定性实验结果对比

为评估所提方法的去噪性能,本文选取多种主流视频复原和去噪方法进行对比,包括 VRT^[17],TURTLE^[19],RViDeformer^[21],TAP^[20] 和 Shift-Net^[18]。其中,VRT,TURTLE 和 Shift-Net 主要面向通用视频复原任务,TAP 和 RViDeformer 主要面向视频去噪任务。上述方法覆盖了基于 Transformer 的视频复原、无监督视频去噪和预训练图像去噪器迁移等不同技术路线,可用于检验本文方法在真实核辐射水下噪声场景中的适应性。为保证比较公平,所有方法均在统一裁剪为 256×256 的图像块上进行测试。

图 5 和图 6 给出了本文方法与对比方法在连续核噪声视频帧上的视觉复原结果。可以看出,在水下核辐射噪声环境中,VRT,TURTLE 和 Shift-Net 对强斑块噪声的抑制能力有限,说明通用复原模型难以直接适应核辐射噪声的非规则退化特征;TAP 和 RViDeformer 虽能减少部分噪声,但容易造成多帧间对比度下降或引入伪影,导致帧间内容不一致。相比之下,本文方法在抑制噪声的同时更好地保持了跨帧内容一致性和局部结构细节,更适合服务于后续水下缺陷识别与状态判读任务。

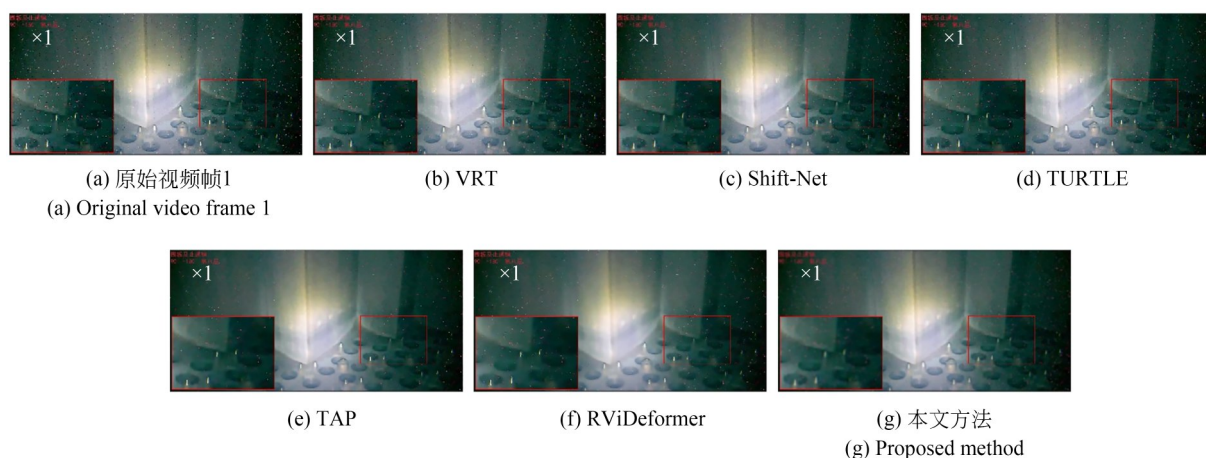


图 5 水下核辐射噪声视频帧 1 的恢复性能对比

Fig. 5 Restoration performance comparison on Frame 1 of the underwater nuclear radiation noise video

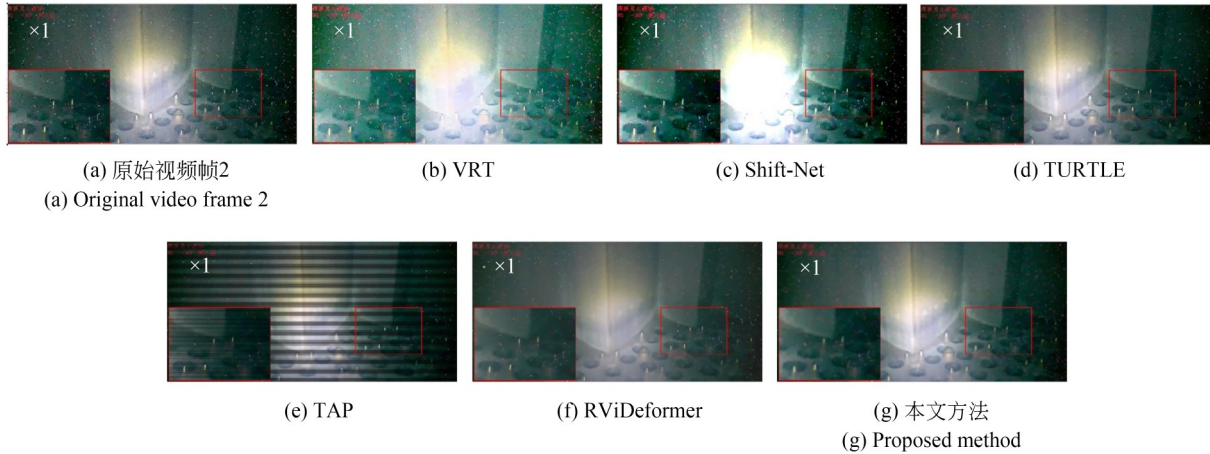


图 6 水下核辐射噪声视频帧 2 的恢复性能对比

Fig. 6 Restoration performance comparison on Frame 2 of the underwater nuclear radiation noise video

3.4 定量实验结果对比

客观质量评价具有量化一致、评估高效和可重复等特点,可为视频去噪算法性能比较提供依据。表 1 和图 7 列出了水下视频受核辐射噪声影响时,不同复原方法的客观质量评价结果。需要说明的是,在缺少清晰参考视频的真实工程场景中,无参考质量指标更适合用于横向比较,但其结果仍需结合可视化复原效果和时序一致性共同判断。

由表 1 可知,Shift-Net 对核辐射噪声的鲁棒性较弱,其 MaxVQA 和 CLIPVQA 分别为 28.7 和 0.83, E_{warp} 误差为 3.3。VRT 和 TURTLE 通过自注意力建模或循环推理提升时间一致性,CLIPVQA 分别达到 0.85 和 0.88,但 MaxVQA 均低于 31,表明通用视频复原方法在强辐射噪声下的帧内细节保留仍存在局限。

表 1 不同方法的性能对比

Tab. 1 Performance comparison of different methods

对比方法	MaxVQA	CLIP-VQA	E_{warp}
TURTLE	30.1	0.88	2.8
TAP	31.2	0.91	2.5
RViDeformer	32.0	0.93	2.1
Shift-Net	28.7	0.83	3.3
VRT	29.3	0.85	3.0
本文方法	34.2	0.97	1.2

相比之下,面向噪声退化的视频去噪方法整体性能优于通用视频复原方法。TAP 利用跨噪声帧时间信息补充空间去噪能力,MaxVQA 和 CLIPVQA 分别为 31.2 和 0.91, E_{warp} 为 2.5。RViDeformer 借助多分支空间/时间自注意力建模,在噪声视频中取得 MaxVQA 为 32.0, CLIPVQA 为 0.93, E_{warp} 为 2.1 的结果。本文方

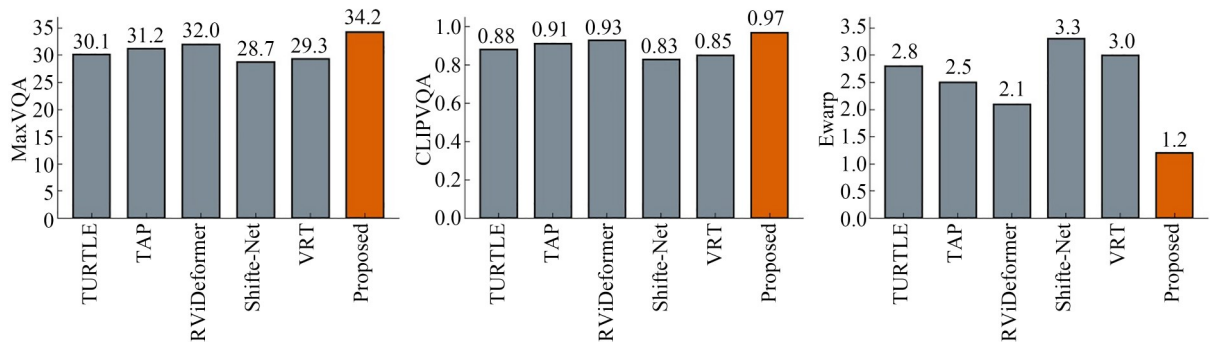


图 7 不同视频复原与去噪方法的客观性能对比

Fig. 7 Objective performance comparison of different video restoration and denoising methods

法进一步将 MaxVQA 和 CLIPVQA 提升至 34.2 和 0.97,并将 E_{warp} 降至 1.2,表明所提混合空间感知机制不仅改善了单帧感知质量,也增强了跨帧时序稳定性。图 8 进一步显示,本文方法位于高感知质量和低时序误差区域,验证了其在视觉真实性、细节保留和时间一致性方面的综合优势。

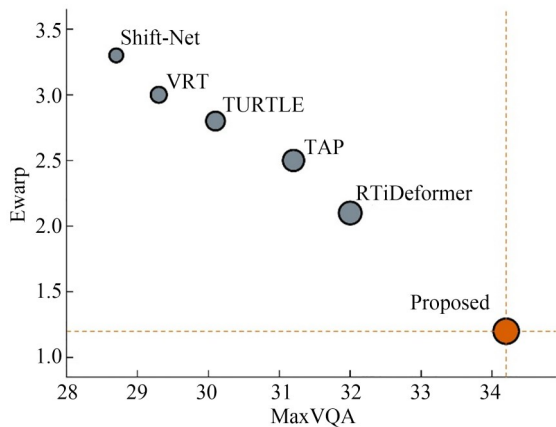


图 8 感知质量与时间一致性的权衡
Fig. 8 Trade-off between perceptual quality and temporal consistency

3.5 消融实验

为比较跨帧标记对应关系的不同识别策略,本文在 U-Net 架构的不同阶段对光流与余弦相似度的组合方式进行评估,结果如表 2 所示。实验结果表明,在下采样模块中采用光流匹配、在上采样模块中采用余弦相似度匹配时,各项指标表现最佳,说明不同对应机制在网络不同深度层具有不同适用性。进一步引入空间感知机制后,方法性能继续提升,表明该机制能够减少纹理缺失区域的错误匹配,并增强局部特征对应的一致性。

表 2 不同一致性匹配方法的性能对比
Tab. 2 Performance comparison of different consistency matching methods

下采样模块	上采样模块	空间感知	MaxVQA	CLIP-VQA	E_{warp}
光流	光流	—	28.6	0.82	2.4
余弦相似度	余弦相似度	—	25.3	0.76	3.1
余弦相似度	光流	—	29.1	0.88	2.0
光流	余弦相似度	—	30.5	0.91	2.1
光流	余弦相似度	✓	32.8	0.95	1.8

为分析层次化潜在变形和混合流引导空间感知标记合并两个核心组件在不同去噪阶段的部署效果,本文开展消融实验,结果如表 3 所示。实验表明,在中期或后期阶段继续应用潜在变形会导致性能下降,说明潜在空间操作更适用于去噪早期;而标记级操作需要贯穿主要去噪过程,以维持细节表达和时间一致性。其原因在于,随着去噪推进,潜在空间语义信息逐渐增强,直接操作潜在表征可能引发语义偏移或细节损失。因此,去噪约束应由早期的潜在空间结构约束逐步过渡到标记层面的精细约束。

标记合并比率 r 影响时间一致性与细节保留之间的权衡。为分析该超参数的作用,本文比较了固定合并比率与渐进式退火策略(r 由 0.9 逐步降至 0)的性能差异,结果如表 4 和图 9 所示。较高固定比例(0.9, 0.6)有助于增强时间一致性并提升 CLIPVQA,但会降低 MaxVQA;较低固定比例(0.3)更利于细节保留,因此 MaxVQA 较高。相比固定比例策略,退火策略可在去噪早期利用较高合并比例强化跨帧一致性,在后期逐步降低合并比例以释放细节表达能力,从而在一致

表 3 不同去噪阶段的定量对比
Tab. 3 Quantitative comparison across various denoising phases

层次化潜在变形			混合空间感知标记合并			MaxVQA	CLIP-VQA	E_{warp}
早期	中期	后期	早期	中期	后期			
—	—	—	—	—	—	32.5	0.97	2.9
✓	—	—	✓	—	—	31.7	0.94	2.5
✓	✓	—	✓	✓	✓	26.3	0.81	2.2
✓	✓	✓	✓	✓	✓	29.2	0.87	2.0
✓	—	—	✓	✓	✓	32.8	0.95	1.8

表 4 不同合并比例下的性能对比

Tab. 4 Performance comparison under different merging ratios

合并比例 r	MaxVQA	CLIP-VQA
0.3	34.2	0.91
0.6	32.5	0.93
0.9	31.8	0.96
0.6 $\rightarrow$ 0	32.7	0.92
0.9 $\rightarrow$ 0	32.8	0.95

性约束和细节保真之间取得更稳健的平衡。

3.6 跨场景与跨退化泛化性能研究

3.6.1 不同核电站水下作业场景的跨场景性能
为验证本文方法在不同核电站水下场景中

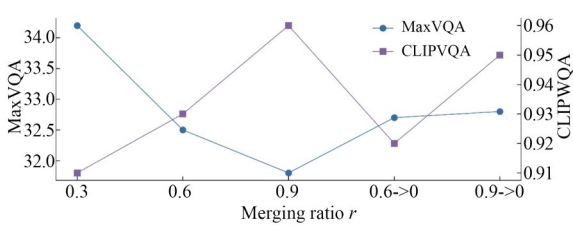


图 9 对应合并比例的消融研究

Fig. 9 Ablation study of correspondence merging ratios

的泛化性能,进一步选取另一段现场采集的核辐射噪声视频进行对比实验。不同于前述大面积彩色斑块噪声,该视频主要包含由单次高能粒子辐射及摄像机后续图像处理形成的密集细小白色亮点。图 10 和图 11 展示了各方法在连续视频帧上的复原结果,以验证所提方法对不同辐射噪声形态的适应能力。

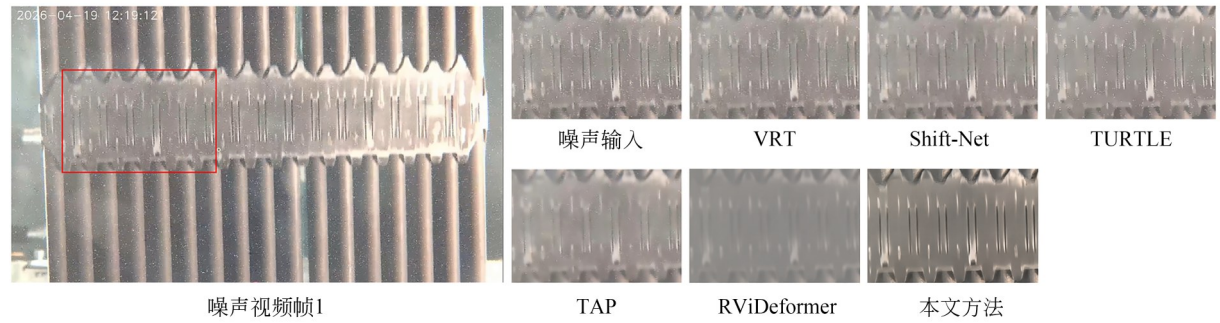


图 10 跨场景噪声视频帧 1 的恢复性能对比

Fig. 10 Restoration performance comparison on Frame 1 under cross-scenario noise

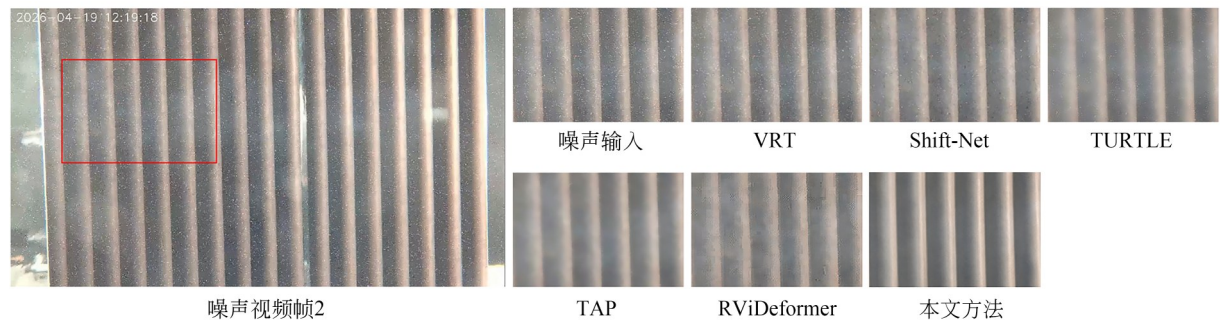


图 11 跨场景噪声视频帧 2 的恢复性能对比

Fig. 11 Restoration performance comparison on Frame 2 under cross-scenario noise

如图 10 所示,VRT,Shift-Net 和 TURTLE 仍存在残余噪声,TAP 易造成过度平滑与细节丢失,RViDeformer 则产生对比度失真。相比之下,本文方法在有效去噪的同时,更好地保留了目标

边缘、轮廓及表面细节,整体清晰度与结构完整性更优。在图 11 中,VRT,Shift-Net 和 TURTLE 仍存在不同程度的残余噪声,结构边缘也略显模糊;TAP 去噪较强,但出现明显过度平滑;

RViDeformer虽降低了噪声,却引入一定的块状伪影。相比之下,本文方法能够更充分地抑制噪声,并较好地地保持纵向结构的连续性、边缘清晰度与局部对比度。

3. 6. 2 一般视频去噪任务的跨退化性能

尽管本文方法主要面向核电站水下辐射噪声视频,但其时序建模与标记合并机制不依赖特定的辐射噪声模型,因而具备向常规视频去噪任务迁移的潜力。为验证其跨退化适用性,本文在

不进行额外训练和参数更新的条件下,选用Set8公开数据集^[28]开展零样本去噪实验。该数据集包含8段视频,其中4段来自Derf测试媒体合集,另外4段由GoPro相机采集。实验分别添加标准差为 $\sigma=30,40,50$ 的高斯噪声,以评估所提方法对不同噪声强度的鲁棒性。采用峰值信噪比(PSNR)和结构相似性指数(SSIM)^[29]评价去噪帧与对应清晰参考帧之间的重建质量,指标值越高,表明去噪性能越好。具体结果见表5。

表 5 Set8去噪数据集不同方法的性能对比
Tab. 5 Performance comparison of different methods on the Set8 denoising dataset

对比方法	Set8, $\sigma=30$		Set8, $\sigma=40$		Set8, $\sigma=50$	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
TURTLE	32.22	0.900	31.17	0.880	30.29	0.855
TAP	32.76	0.910	31.40	0.888	30.50	0.862
RViDeformer	33.10	0.918	32.00	0.898	31.10	0.875
Shift-Net	33.50	0.925	32.30	0.905	31.40	0.882
VRT	33.35	0.920	32.20	0.902	31.32	0.880
本文方法	34.00	0.930	32.90	0.912	32.05	0.890

由表5可知,随着高斯噪声标准差由30增大至50,各方法的PSNR和SSIM均逐渐下降。相比之下,本文方法在三种噪声水平下均取得最优结果。具体而言,相较于性能最优的对比方法Shift-Net,本文方法在 $\sigma=30,40,50$ 时的PSNR分别提升0.50,0.60和0.65 dB,SSIM分别提升0.005,0.007和0.008。随着噪声强度增大,本文方法的性能优势进一步扩大。结果表明,所提层次化时序建模和空间感知标记合并能够有效利用跨帧冗余信息,在常规高斯噪声视频中仍具有较好的噪声抑制与结构恢复能力,体现出良好的跨退化迁移能力和鲁棒性。

4 结 论

针对真实核电站水下环境中高能粒子辐射噪声导致的视频质量退化问题,本文提出一种基于零样本学习的水下视频去噪方法。该方法利用预训练图像复原扩散模型,无需额外训练即可通过层次化潜在变形和混合流引导的空间感知标记合并,实现噪声抑制、细节保留和时间一致

性维护。在真实核电站水下噪声视频数据集上的实验结果表明,本文方法的MaxVQA和CLIPVQA分别达到34.2和0.97, E_{warp} 降至1.2, 优于现有主流视频复原与去噪方法。该研究为缺乏标注数据条件下的核电站水下光学成像增强提供了可行方案,可为后续设备缺陷检测和智能运维应用提供更可靠的视觉输入。需要指出的是,受真实检修现场数据获取条件限制,本文尚需在更多机组环境、更多水下作业场景和更长视频序列上进一步验证;同时,算法的端侧部署效率和实时处理能力仍需在工程系统中持续优化。

作者贡献声明:

- 刘颖:方法的提出,论文构思和撰写;
- 杨杰:实验的设计及数据整理和分析;
- 罗杰:论文审核与编辑写作;
- 刘顺:项目管理与资源协调;
- 李天昊:实验数据分析;
- 肖晓南:论文指导;
- 汪昱成:数据采集。

参考文献:

- [1] 王璐. 中国核电实现跨越发展“核能+人工智能”打开新空间[N]. 经济参考报, 2025-03-31(004).
WANG L. China's Nuclear Power Achieves Leapfrog Development: "Nuclear Energy + Artificial Intelligence" Opens New Space [N]. *Economic Information Daily*, 2025-03-31(004). (in Chinese)
- [2] 邓豪. 基于时空关联特征的 γ 辐射图像去噪方法研究[D]. 绵阳: 西南科技大学, 2024.
HANG D. *Researched on γ Radiation Image Denoising Method Based on Spatiotemporal Correlation Features*[D]. Mianyang: Southwest University of Science and Technology, 2024. (in Chinese)
- [3] 杨中晔, 全权, 刘杨, 等. 核辐射对无人系统的影响研究[J]. 人工智能, 2021(4): 127-136.
YANG Z Y, QUAN Q, LIU Y, *et al.* Study on the influence of nuclear radiation on unmanned system[J]. *AI-View*, 2021(4): 127-136. (in Chinese)
- [4] 杨康. 核辐射场景中视觉SLAM定位技术研究[D]. 绵阳: 西南科技大学, 2019.
YANG K. *Research on Visual SLAM Positioning Technology in Nuclear Radiation Scene*[D]. Mianyang: Southwest University of Science and Technology, 2019. (in Chinese)
- [5] 王矩, 桑瑞娟, 张华, 等. 一种强辐射环境监测下图像降噪的新方法[J]. 传感器与微系统, 2011, 30(11): 59-61.
WANG H, SANG R J, ZHANG H, *et al.* A new image denoising method for monitoring in intense radioactive environment[J]. *Transducer and Microsystem Technologies*, 2011, 30(11): 59-61. (in Chinese)
- [6] DOVGANICH A A, KRYLOV A S. A nonlocal image denoising algorithm using the structural similarity metric[J]. *Programming and Computer Software*, 2019, 45(4): 141-146.
- [7] MOHD SAGHEER S V, GEORGE S N. Denoising of low-dose CT images via low-rank tensor modeling and total variation regularization[J]. *Artificial Intelligence in Medicine*, 2019, 94: 1-17.
- [8] 黄小莉, 陈春梅, 刘桂华. 混合二阶全变分的抗核辐射图像降噪方法[J]. 重庆邮电大学学报(自然科学版), 2022, 34(4): 585-594.
HUANG X L, CHEN C M, LIU G H. A hybrid second-order total variational noise reduction method for radiation-resistant images [J]. *Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition)*, 2022, 34(4): 585-594. (in Chinese)
- [9] 陈明举. 基于稀疏表示与变分法的核辐射干扰图像增强技术研究[D]. 绵阳: 西南科技大学, 2020.
CHEN M J. *Research on Image Enhancement Technology of Nuclear Radiation Interference Based on Sparse Representation and Variational Method* [D]. Mianyang: Southwest University of Science and Technology, 2020. (in Chinese)
- [10] 孙跃文, 刘洪, 丛鹏, 等. 基于卷积神经网络的辐射图像降噪方法研究[J]. 原子能科学技术, 2017, 51(9): 1678-1682.
SUN Y W, LIU H, CONG P, *et al.* Radiation image denoising based on convolutional neural network [J]. *Atomic Energy Science and Technology*, 2017, 51(9): 1678-1682. (in Chinese)
- [11] 孙鑫. 核辐射场景的图像去噪与作业目标检测技术研究[D]. 绵阳: 西南科技大学, 2021.
SUN X. *Research on Image Denoising and Operational Object Detection Technology in Nuclear Radiation Scenes* [D]. Mianyang: Southwest University of Science and Technology, 2021. (in Chinese)
- [12] 黄小莉. 核辐射干扰图像降噪技术研究[D]. 绵阳: 西南科技大学, 2022.
HUANG X L. *Research on Image Denoising in Nuclear Radiation Scenes* [D]. Mianyang: Southwest University of Science and Technology, 2022. (in Chinese)
- [13] 谢明洁. 基于深度学习的辐射干扰噪声视频图像去噪方法及应用研究[D]. 杭州: 浙江大学, 2023.
XIE M J. *Research on Radiation Contaminated Noise Image and Video Denoising Methods and Applications Based on Deep Learning* [D]. Hangzhou: Zhejiang University, 2023. (in Chinese)
- [14] SAHARIA C, HO J, CHAN W, *et al.* Image super-resolution via iterative refinement [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(4): 4713-4726.
- [15] LUO Z W, GUSTAFSSON F K, ZHAO Z, *et al.* Image Restoration with Mean-reverting Stochastic Differential Equations [EB/OL]. 2023: arXiv: 2301.11699. <https://arxiv.org/abs/2301.11699>
- [16] LIN X Q, HE J W, CHEN Z Y, *et al.* DiffBIR: toward blind image restoration with generative dif-

- fusion prior[C]. *Computer Vision-ECCV* 2024. Cham: Springer, 2025: 430-448.
- [17] LIANG J Y, CAO J Z, FAN Y C, *et al.* VRT: a video restoration transformer[J]. *IEEE Transactions on Image Processing*, 2024, 33: 2171-2182.
- [18] LI D S, SHI X Y, ZHANG Y, *et al.* A simple baseline for video restoration with grouped spatial-temporal shift[C]. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 17-24, 2023, Vancouver, BC, Canada. IEEE, 2023: 9822-9832.
- [19] GHASEMABADI A, JANJUA M, SALAMEH M, *et al.* Learning truncated causal history model for video restoration[C]. *Advances in Neural Information Processing Systems* 37. December 10-15, 2024. Vancouver, BC, Canada. Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2024: 27584-27615.
- [20] FU Z X, GUO L Q, WANG C, *et al.* Temporal as a Plugin: Unsupervised Video Denoising with Pre-trained Image Denoisers[M]. *Computer Vision - ECCV* 2024. Cham: Springer Nature Switzerland, 2024: 349-367.
- [21] YUE H J, CAO C, LIAO L, *et al.* RViDeformer: efficient raw video denoising transformer with a larger benchmark dataset[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2025, 35(9): 8929-8944.
- [22] LI X R, MA C, YANG X K, *et al.* VidToMe: Video Token Merging for Zero-shot Video Editing[C]. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 16-22, 2024, Seattle, WA, USA. IEEE, 2024: 7486-7495.
- [23] TENG H, QUAN Y H, WANG C Y, *et al.* Fingerprinting Denoising Diffusion Probabilistic Models[C]. 2025 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 10-17, 2025, Nashville, TN, USA. IEEE, 2025: 28811-28820.
- [24] YAN L Q, LI X B, ZHANG J H, *et al.* F-DDIM: a featurized denoising diffusion implicit model for facial image steganography[C]. *Proceedings of the 33rd ACM International Conference on Multimedia*. Dublin Ireland. ACM, 2025: 8488-8496.
- [25] XU H F, ZHANG J, CAI J F, *et al.* GMFlow: learning optical flow via global matching[C]. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 18-24, 2022, New Orleans, LA, USA. IEEE, 2022: 8111-8120.
- [26] WU H N, ZHANG E L, LIAO L, *et al.* Towards explainable in-the-wild video quality assessment: a database and a language-prompted approach[C]. *Proceedings of the 31st ACM International Conference on Multimedia*. Ottawa ON Canada. ACM, 2023: 1045-1054.
- [27] XING F C, LI M J, WANG Y G, *et al.* CLIPVQA: video quality assessment via CLIP[J]. *IEEE Transactions on Broadcasting*, 2025, 71(1): 291-306.
- [28] TASSANO M, DELON J, VEIT T. FastDVD-net: towards real-time deep video denoising without flow estimation[C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 13-19, 2020. Seattle, WA, USA. IEEE, 2020: 1351-1360.
- [29] WANG Z, BOVIK A C, SHEIKH H R, *et al.* Image quality assessment: from error visibility to structural similarity[J]. *IEEE Transactions on Image Processing*, 2004, 13(4): 600-612.

通讯作者:



刘 颖(1995—),男,安徽安庆人,硕士研究生,2021年于上海海事大学获得硕士学位,主要从事核工业智能监测方面的研究。E-mail: ahtly0830@163.com

作者简介:



杨 杰(1992—),男,陕西渭南人,硕士,2025年于西北工业大学获得硕士学位,主要从事核工业智能监测方面的研究。E-mail: 329832332@qq.com