基于内容自适应融合的海上航标监控视频流质量增强方法

孙宇，高曙，张炎鑫

（武汉理工大学人工智能学院，湖北 武汉 430070）

摘 要：海上多功能航标监控视频流具有静态海面背景占比高、动态船舶目标稀疏的鲜明场景特性，同时受标载通信网络低带宽限制，压缩后存在噪声叠加、块效应、振铃效应、帧模糊及帧间运动估计偏差等质量退化问题。现有通用视频质量增强方法难以兼顾单帧空域细节恢复与多帧时序信息利用，无法满足上述场景需求。为此，提出一种基于内容自适应融合的海上多功能航标监控视频流质量增强（CAF-VSMM）方法，构建双分支融合增强模型。基于多尺度特征与时空网络的单帧增强分支，通过虚拟帧挖掘单帧潜在时序信息，结合像素重排与残差网络实现多尺度空域特征融合，有效抑制压缩伪影与海上环境引发的空域质量退化；基于动态移位窗口与帧间运动补偿的多帧增强分支，高效建模长距离时空依赖，利用分层偏移估计与可变形卷积实现多船场景帧间精准对齐，缓解时域模糊与运动错位；基于动态特征感知的内容自适应融合（DFP-CAF）模块，依据监控视频流内容动态调整两分支特征权重，实现静态背景重细节、动态目标重时序的自适应增强。实验结果表明：在自建港口群航道监控视频数据集（PC-WMVD）上，该方法的峰值信噪比（PSNR）、结构相似性（SSIM）、视频多方法评估融合（VMAF）指标分别达 42.71 dB、99.33%、77.14%，较次优对比方法提升 3.90 dB、1.19 百分点、6.50 百分点；在公开数据集 JCT-VC 上分别达 28.66 dB、88.60%、55.40%，较次优对比方法提升 1.63 dB、5.42 百分点、3.19 百分点；模型参数量为 37.52×10^6、浮点运算量（FLOPs）为 51.74×10^9，处理帧率为 41.52 帧/s，满足航道监控实时性要求。该方法已在某港口群航道智能监控系统中示范应用，可显著提升船舶、航道标识等关键目标辨识度，有效支撑船舶目标跟踪等下游任务，验证了工程实用性与应用价值。

关键词：视频质量增强；多功能航标；时空网络；帧间运动补偿；单帧多帧融合；内容自适应融合

中图分类号：TP391.41 **文献标志码：**A **DOI：**10.19678/j.issn.1000-3428.0260594

Video Stream Quality Enhancement Method for Maritime Navigation Mark Monitoring Based on Content Adaptive Fusion

SUN Yu, GAO Shu, ZHANG Yanxin

(School of Artificial Intelligence, Wuhan University of Technology, Wuhan 430070, Hubei, China)

【Abstract】The video stream of maritime multifunctional navigation mark monitoring presents the distinctive scene characteristics of a dominant static sea surface background and sparse dynamic ship targets. Constrained by the low-bandwidth transmission of navigation mark-borne communication networks, compressed video streams experience severe quality degradation, including noise superposition, blocking artifacts, ringing artifacts, frame blurring, and inter-frame motion estimation bias. Existing general-purpose video quality enhancement methods are unable to jointly balance single-frame spatial detail recovery and multi-frame temporal information utilization, and therefore fall short of the requirements of the aforementioned scenario. To address this issue, a Content Adaptive Fusion-based Video Stream Quality Enhancement of Maritime Multifunctional Navigation Mark Monitoring (CAF-VSMM) method is proposed, in which a dual-branch fusion enhancement model is constructed. The single-frame enhancement branch, built on multiscale features and a spatio-temporal network, exploits the latent temporal information of a single-frame via virtual frames and integrates the pixel shuffle with a residual network to achieve multiscale spatial feature fusion, effectively suppressing compression artifacts and spatial degradation induced by the maritime environment. The multi-frame enhancement branch, based on a dynamic shifted window and inter-frame motion compensation, efficiently models long-range spatio-temporal dependencies and employs hierarchical offset estimation together with deformable convolution to achieve accurate inter-frame alignment in multiship scenes, thereby alleviating temporal blurring and motion misalignment. The Dynamic Feature Perception-based Content-Adaptive Fusion (DFP-CAF) module, based on dynamic feature perception, dynamically adjusts the feature weights of the two branches according to the video content, realizing an adaptive enhancement strategy that emphasizes detailed recovery for static backgrounds and temporal consistency for dynamic targets. Experimental

基金项目：国家重点研发计划（2023YFB2603800）。

作者简介：孙宇，男，硕士研究生，主研方向为计算机视觉；高曙（通信作者），教授、博士；张炎鑫，本科生。

收稿日期：2026-05-06 **修回日期：**2026-07-09 **E-mail：**gshu418@163.com

results show that, on the self-built Port Cluster Waterway Monitoring Video Dataset (PC-WMVD), the Peak Signal to Noise Ratio (PSNR), Structural Similarity (SSIM), and Video Multimethod Evaluation Fusion (VMAF) of the proposed method reach 42.71 dB, 99.33%, and 77.14%, outperforming the second-best method by 3.90 dB, 1.19 percentage points, and 6.50 percentage points, respectively; on the public dataset JCT-VC, they reach 28.66 dB, 88.60%, and 55.40%, exceeding the second-best method by 1.63 dB, 5.42 percentage points, and 3.19 percentage points, respectively. The model has 37.52×10^6 parameters and 51.74×10^9 Floating-Point Operations Per Second (FLOPs) and runs at 41.52 frames per second, meeting the real-time requirements of waterway monitoring. The proposed method is tested in an intelligent monitoring system for the waterways of port clusters, where it significantly improves the identifiability of key targets, such as ships and navigation marks, and effectively supports downstream tasks, such as ship tracking, verifying its engineering practicability and application value.

【Key words】 video quality enhancement; multi-functional navigation mark; spatio-temporal network; inter-frame motion compensation; single-frame and multi-frame fusion; content-adaptive fusion

0 引言

随着传感技术与边缘计算的快速发展，集成摄像头、雷达、船舶自动识别系统以及传感器（如风速风向仪、浊度传感器等）的多功能航标得到广泛应用，采集的航道监控视频流为船舶目标识别与跟踪[1]、船舶异常行为检测[2]等航道监控任务提供基础数据，极大提升了港口群航道智能监控能力，在“交通强国”战略与智慧海事建设中发挥重要作用。

海上多功能航标多部署于远岸航道水域，其数据传输依赖标载通信网络（一种依托新一代移动通信技术，在航标上集成部署的具备低时延、广覆盖、无缝连接能力，并支持多跳自组网以延伸通信距离的专用通信网络），由于航标能源供给依赖光伏发电，供电能力有限，因此航标上设备必须满足低功耗、低算力运行需求。同时，标载通信网络是一种低带宽的专用通信网络。因此，航标上采集的监控视频流数据通过此网络传输至岸基端时，受低功耗与低带宽双重约束，必须采用高效的视频压缩编码技术（如 H.264/AVC 标准[3]）对视频流进行压缩处理，以降低数据传输量与设备能耗。

另外，压缩过程中引入的块效应、振铃效应以及帧间运动估计偏差等失真问题，会导致视频质量退化，必须通过针对性的视频质量增强技术进行修复与优化；否则，压缩后的监控视频流数据将无法满足航道监控任务的实时感知与精准识别需求。因此，开展面向海上多功能航标监控视频流的质量增强方法研究，解决低带宽传输下压缩监控视频流的失真问题，不仅具有重要理论价值，而且更是推动“智慧港航”建设与工程应用落地的亟需。

视频质量增强旨在恢复视频压缩过程中损失的细节信息，缓解视频压缩产生的失真。为提高视频质量增强效果，目前增强方法从视频帧维度可分为视频单帧质量增强方法和视频多帧质量增强方法。

视频单帧增强方法与图像增强方法之间没有明确界限。LI 等[4]提出 WRFA-UNet，设计 JPEG 压缩伪影学习模块，结合空间通道注意力机制，提升多尺度区域的质量增强效果。IBRAHEEM 等[5]提出 RCNN，将 ResNet 的残差结构与深度卷积网络相结合，显著提升视频单帧质量。WANG 等[6]提出 PRCAFF-Net，融合密集块、渐进残差融合与三维卷积注意力模块，以提升恢复图像边缘纹理的效果，并高效去除高斯噪声。

视频多帧质量增强方法根据相邻帧之间是否需要对齐分为：1）无对齐的视频多帧质量增强方法；2）基于对齐的视频多帧质量增强方法。第 1）类方法通过构建端到端的深度学习框架，直接建模视频帧间的时空特征关联性以实现视频质量提升，无需对相邻帧执行显式对齐操作。尽管该类方法保持了网络结构的简洁高效，但缺乏针对大范围快速运动场景的处理能力，因此显式对齐仍然是不可避免的。第 2）类方法又可分为基于光流对齐[7]和基于可变形卷积网络（DCN）[8]对齐两大类。光流估计技术通过分析相邻帧之间像素点的差异来推断像素点的运动，面对目标运动场景时，易出现光流场求解偏差；可变形卷积通过自适应采样偏移量，突破传统卷积固定矩形采样网格的限制，对目标的复杂几何变换进行灵活精准的建模。基于上述理论，DENG 等[9]提出时空可变形融合网络，通过偏移预测网络来预测帧间特征关系，减少对齐误差。ZHANG 等[10]提出 STCF 网络，通过滑动窗口注意力机制，增强网络时序建模能力。此外，LIU 等[11]提出 MRDN 网络，采用金字塔结构提取多尺度特征，并引入边缘损失函数以增强目标边缘区域的质量。

尽管上述视频质量增强方法取得了阶段性进展，但仍存在诸多局限性：现有视频单帧质量增强方

法仅依赖空域信息，特征提取单一、多尺度信息融合低效，且无法有效挖掘航标监控视频流的时序关联性[12]，抑制噪声、块效应等空域失真问题的能力有限；视频多帧质量增强方法的运动补偿机制难以适配航道场景中船舶航行运动带来的帧间动态变化，易出现特征对齐精度不足，从而导致视频帧模糊、帧间运动估计偏差等时域质量退化问题，且基于全局注意力的时空建模方案计算复杂度高，无法满足航道监控任务的实时处理要求[13]；现有方法未考虑航标监控视频流“静态海面背景占比高、动态船舶目标稀疏”的场景特性，无法根据视频内容动态调整特征融合权重，难以兼顾视频单帧与多帧质量增强方法各自的优势[14]。

针对上述问题，本文提出基于内容自适应融合的海上多功能航标监控视频流质量增强(CAF-VSMM)方法，以提高监控视频流质量增强性能。本文主要贡献如下：

1)针对视频压缩引发的空域质量退化问题，设计基于多尺度特征和时空网络的单帧增强分支(MFSN-SEB)，通过虚拟帧构建伪视频序列以挖掘单帧潜在的时序依赖关系，结合像素重排与残差网络实现多尺度空域特征的分层提取，并基于动态卷积设计双阶段递进融合策略，完成多维度特征的深度融合，有效抑制噪声与块效应、振铃效应等压缩伪影，提升船舶、航道标识等目标的边缘清晰度与全局结构完整性。

2)针对视频压缩导致的时域质量退化问题，设计基于移位窗口和帧间运动补偿的多帧增强分支(SWIMCN-MEB)，通过动态移位窗口打破传统窗口的边界限制，实现跨窗口长距离时空依赖的高效建模，有效提升监控视频流的清晰度；同时基于邻域像素相关性构建端到端的分层偏移量估计网络，并结合可变形卷积实现海上多船航行场景下帧间特征的精准对齐，避免帧间运动估计偏差问题，提高监控视频流的时序一致性。

3)针对海上航标监控视频流“静态海面背景占比高、动态船舶目标稀疏”的场景特性，设计内容自适应融合模块，实现静态场景下空域特征主导、动态场景下时域特征主导的自适应融合，充分发挥视频单帧与多帧质量增强方法各自的优势，提高船舶、航道标识等目标的特征辨识度。

4)依托某港口群航道智能监控系统采集的数据制作港口群航道监控视频数据集(PC-WMVD)，为模型训练提供数据支撑。将所构建的模型部署至某港口群航道监控系统进行示范应用，实现监控视频流的实时质量增强处理，显著提升了船舶、航道标识等关键目标辨识度，有效支撑船舶目标跟踪等下游监控任务，验证所提方法的工程实用性与应用价值。

1 CAF-VSMM 模型架构

1.1 模型结构

CAF-VSMM 模型由单帧增强与多帧增强双分支、基于动态特征感知的内容自适应融合(DFP-CAF)模块组成，通过动态生成与视频内容自适应匹配的融合权重，实现单帧与多帧特征的高效融合，兼顾航标监控视频流中静态海面背景的空域细节恢复与动态船舶的时序信息利用，从而极大提升航道实时感知与目标精准监控能力。整体架构如图 1 所示(彩色效果见《计算机工程》官网 HTML 版，下同)。

1.2 MFSN-SEB

海上航标监控视频流经 H.264/AVC 标准压缩后，因单帧处理模式难以利用帧间的时序相关性，对噪声与块效应、振铃效应等压缩伪影的抑制能力不足。为此，构建 MFSN-SEB，它由基于虚拟帧与时空网络的时域特征提取子网络(VFSN-TFES)和基于像素重排与残差网络的多尺度空域特征提取子网络(PSRN-MSFES)构成。

1.2.1 VFSN-TFES

针对现有通用视频单帧质量增强方法难以利用时序信息修复视频压缩引发的块效应、振铃效应等压缩伪影问题，受文献[15]启发，通过投影梯度下降(PGD)[16]攻击算法对视频单帧 $\boldsymbol{X}_i \in \mathbb{R}^{H\times W\times C}$ ($i=1,2,\cdots,n$，其中 H、W、C 分别代表视频帧的高、宽、通道数)添加约束范围内的扰动，生成虚拟帧 $\boldsymbol{X}_i^{\text{adv}} \in \mathbb{R}^{H\times W\times C}$ ($i=1,2,\cdots,n$)以构建伪视频序列，模拟真实航标监控视频流的时序维度特征分布，解决三维卷积神经网络(CNN)难以直接应用于视频单帧场景的问题。由于 PGD 攻击算法的扰动影响，$\boldsymbol{X}_i$ 与 $\boldsymbol{X}_i^{\text{adv}}$ 的特征分布存在差异，因此通过 3×3 卷积提取初步特征 $\boldsymbol{P}_i^{\text{clean}} \in \mathbb{R}^{H\times W\times C}$ 与 $\boldsymbol{P}_i^{\text{adv}} \in \mathbb{R}^{H\times W\times C}$ ($i=1,2,\cdots,n$)后，分别使用主层(Main Layer)归一化与辅助层(Auxiliary Layer)归一化消除两类特征因生成机制不同而造成的分布偏移，避免因特征错位造成的时序依赖偏差。最后，借助 ReLU 激活函数三维卷积挖掘帧间潜在的时序依赖关系，输出潜在的时域特征 $\boldsymbol{P}_i \in \mathbb{R}^{H\times W\times C}$ ($i=1,2,\cdots,n$)，弥补单帧处理模式的缺陷。具体过程如式(1)～式(5)所示：

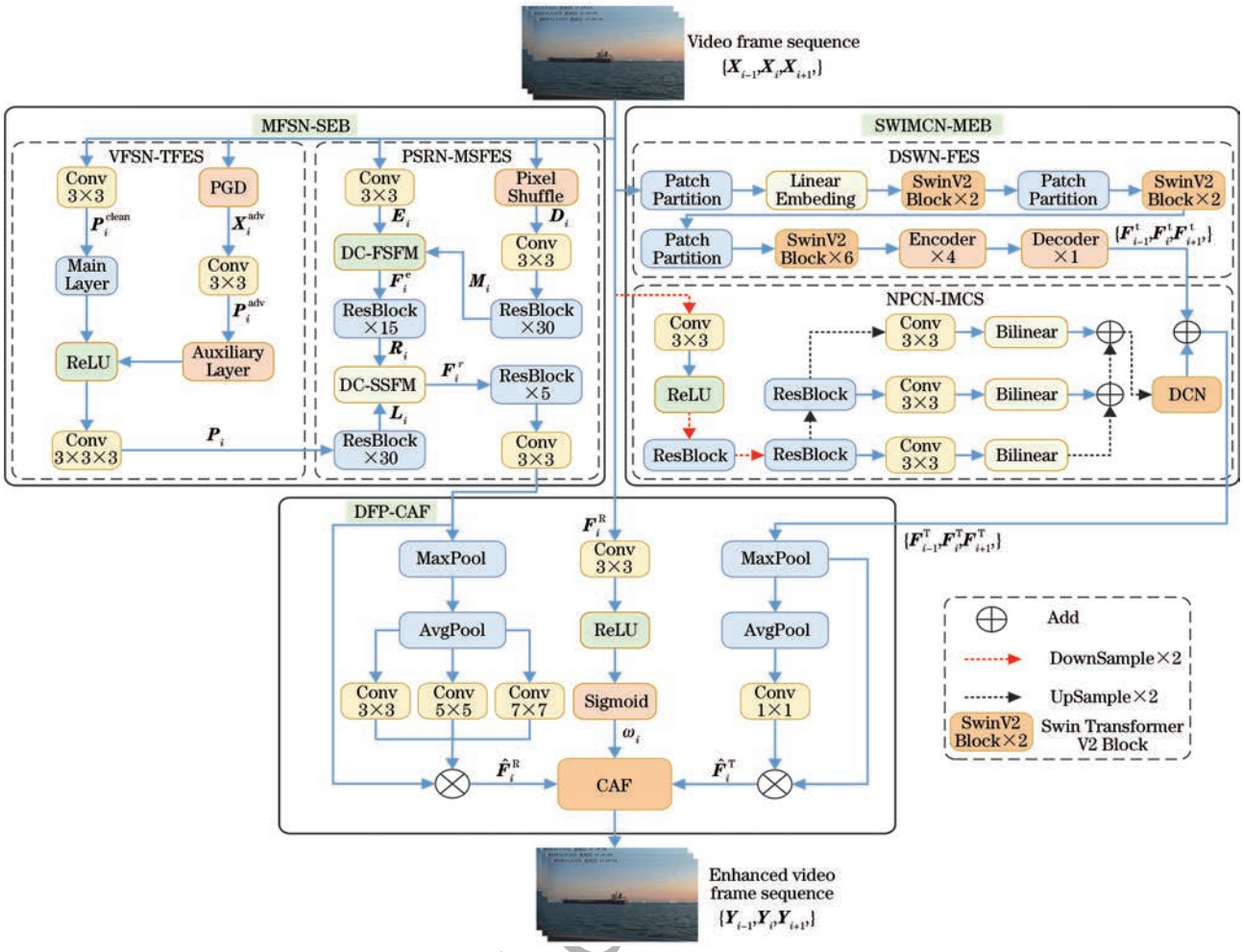


图 1 CAF-VSMM 模型架构

Fig. 1 CAF-VSMM model architecture

$$\boldsymbol{\delta}_0=\mu(-\varepsilon,\varepsilon) \tag{1}$$

$$\nabla_t=\nabla_{t-1}\xi(\mathrm{CNN}(\boldsymbol{\delta}_{t-1}),\boldsymbol{X}_i) \tag{2}$$

$$\boldsymbol{\delta}_t=\boldsymbol{\delta}_{t-1}+a\cdot\mathrm{sign}(\nabla_t) \tag{3}$$

$$\boldsymbol{\delta}_t=\mathrm{clip}(\boldsymbol{\delta}_t,\boldsymbol{X}_i-\varepsilon,\boldsymbol{X}_i+\varepsilon) \tag{4}$$

$$\boldsymbol{X}_i^{\mathrm{adv}}=\boldsymbol{X}_i+\boldsymbol{\delta}_T \tag{5}$$

式中：$\boldsymbol{\delta}$ 是扰动向量；ε 是扰动上限，为保证扰动幅度在人眼不可感知的范围内，同时避免虚拟帧与原始视频帧的时域关联性丢失，其值设置为 4[17-18]；$\mu(\cdot)$表示区间上的均匀分布函数，确保初始扰动幅度在约束范围内；∇_t是关于$\boldsymbol{\delta}_{t-1}$ 的梯度，t 是当前迭代次数；$\xi(\cdot)$表示分类损失函数；CNN$(\cdot)$表示卷积神经网络；a 是迭代步长；sign$(\cdot)$表示符号函数，确保扰动更新与梯度方向保持一致，并避免梯度数值过大导致扰动溢出；clip$(\cdot)$表示裁剪函数，将更新后的扰动约束至指定区间内；T 是总迭代次数，将最终扰动向量叠加至 $\boldsymbol{X}_i$，生成虚拟帧 $\boldsymbol{X}_i^{\mathrm{adv}}$。

1.2.2 PSRN-MSFES

针对噪声叠加以及船舶、航道标识等关键监控目标局部细节恢复能力不足等问题，PSRN-MSFES 采用像素重排替代传统卷积实现下采样操作，将航标监控视频单帧 $\boldsymbol{X}_i$ 相邻空间位置的特征整合到不同通道中，避免直接丢弃空间信息，更完整地保留监控视频单帧原始特征信息，以此得到下采样帧 $\boldsymbol{D}_i\in\mathbb{R}^{\frac{H}{4}\times\frac{W}{4}\times 4C}(i=1,2,\cdots,n)$。首先，对 $\boldsymbol{X}_i$ 执行 3×3 卷积操作完成初步特征 $\boldsymbol{E}_i\in\mathbb{R}^{H\times W\times C}(i=1,2,\cdots,n)$ 的提取；其次，对 $\boldsymbol{D}_i$ 执行 3×3 卷积操作后，再通过 30 个残差块构建的深度残差网络挖掘空域上下文信息，并缓解训练过程中出现的梯度消失问题，得到 $\boldsymbol{D}_i$ 对应的初步特征 $\boldsymbol{M}_i\in\mathbb{R}^{\frac{H}{4}\times\frac{W}{4}\times 4C}(i=1,2,\cdots,n)$。针对多尺度特征融合效率低的问题，设计双阶段递进融合策略（如图 2 所示），首先通过基于动态卷积的第一阶段融合模块（DC-FSFM）完成各尺度空域特征 $\boldsymbol{E}_i$、$\boldsymbol{M}_i$ 之间的互补融合，输出多尺度空域融合特征 $\boldsymbol{F}_i^{\mathrm{e}}\in\mathbb{R}^{\frac{H}{4}\times\frac{W}{4}\times 4C}(i=1,2,\cdots,n)$，经 15 个残差块构建的残差网络强化 $\boldsymbol{F}_i^{\mathrm{e}}$ 的空域细节，得到增强后的融合特征 $\boldsymbol{R}_i\in\mathbb{R}^{\frac{H}{2}\times\frac{W}{2}\times 2C}(i=1,2,\cdots,n)$；然后，

VFSN-TFES 输出的时域特征 $\boldsymbol{P}_i$ 经 30 个残差块构建的深度残差网络进行时域上下文特征细化处理，得到增强后的时域特征 $\boldsymbol{L}_i \in \mathbb{R}^{H\times W\times C}(i=1,2,\cdots,n)$，再利用基于动态卷积的第二阶段融合模块(DC-SSFM)将 $\boldsymbol{R}_i$ 与 $\boldsymbol{L}_i$ 进行深度融合，输出融合多维度信息的空域特征 $\boldsymbol{F}_i^{\mathrm{r}} \in \mathbb{R}^{H\times W\times C}(i=1,2,\cdots,n)$；最后，$\boldsymbol{F}_i^{\mathrm{r}}$ 经残差块和卷积处理输出单帧空域特征 $\boldsymbol{F}_i^{\mathrm{R}} \in \mathbb{R}^{H\times W\times C}(i=1,2,\cdots,n)$。

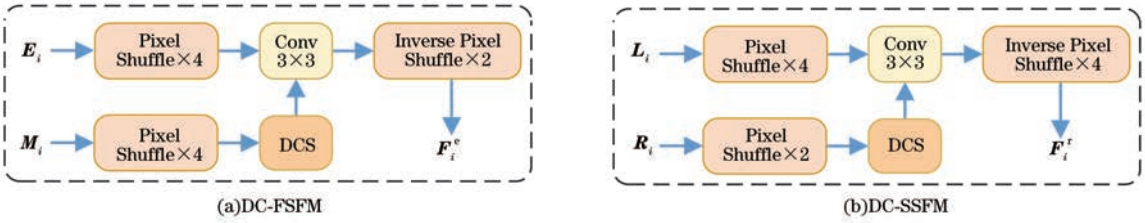


图 2 双阶段递进融合结构

Fig. 2 Two-stage progressive fusion structure

在双阶段递进融合的两个阶段 DC-FSFM 和 DC-SSFM 中，为避免简单拼接导致多尺度特征融合不充分，设计动态卷积子模块(DCS)，如图 3 所示。其通过平均池化(AvgPool)与最大池化(MaxPool)操作对输入特征 $\boldsymbol{F}$ 进行处理得到池化特征，利用 3×3 卷积进行特征变换，并借助自适应平均池化(Adaptive AvgPool)压缩空间维度，再经 1×1 卷积生成通道注意力权重向量 $\boldsymbol{\alpha}$，在 $\boldsymbol{\alpha}$ 的作用下实现多维度特征 $\boldsymbol{E}_i$、$\boldsymbol{M}_i$ 以及 $\boldsymbol{L}_i$ 之间的动态融合与空间维度恢复。

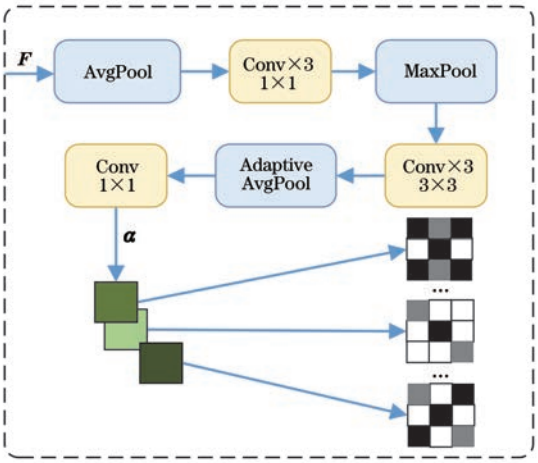


图 3 DCS 结构

Fig. 3 DCS structure

综上，VFSN-TFES 和 PSRN-MSFES 分别利用虚拟帧与时空网络和像素重排与残差网络等，在单帧输入条件下完成对航标监控视频流潜在时序信息的挖掘和多尺度空域特征的提取与融合，实现对视频压缩引发的噪声叠加、块效应以及振铃效应等空域质量退化的有效抑制，显著提升船舶、航道标识等关键监控目标的边缘清晰度与全局结构完整性，为后续基于 DFP-CAF 模块提供蕴含潜在时序信息的空域特征输入。

1.3 SWIMCN-MEB

海上航标监控视频流中存在大量船舶航行、目标遮挡等动态场景，同时因压缩引发帧模糊、帧间运动估计偏差等时域质量退化问题，现有视频多帧质量增强方法难以在该场景下对长距离时空依赖进行高效建模，帧间运动补偿精度不足容易导致特征对齐错位，产生运动伪影。为此，构建 SWIMCN-MEB，由基于动态移位窗口的特征提取子网络(DSWN-FES)和基于邻域像素相关性的帧间运动补偿子网络(NPCN-IMCS)构成。

1.3.1 DSWN-FES

现有通用视频多帧质量增强方法受限于其窗口机制，导致长距离依赖建模存在瓶颈，难以有效恢复视频中的模糊帧。DSWN-FES 通过动态移位窗口注意力机制，实现航标监控视频帧序列长距离时空依赖的高效建模，有效解决了传统 Transformer 全局注意力计算复杂度随监控视频流分辨率呈指数增长以及固定窗口注意力受边界约束而导致的跨窗口时空信息关联不充分等问题。首先，通过 Patch Partition 操作对视频帧序列 $\{\boldsymbol{X}_{i-1},\boldsymbol{X}_i,\boldsymbol{X}_{i+1}\} \in \mathbb{R}^{H\times W\times C}(i=1,2,\cdots,n)$ 进行分块处理，得到非重叠局部块；对局部块执行 Linear Embedding 操作，利用线性变换将分块特征映射至高维特征空间，解决了原始航标监控视频帧信息冗余度高、特征维度无法直接适配注意力机制建模的问题；通过多组 Patch Partition 操作和 Swin Transformer V2 Block[19] 单元实现对 $\{\boldsymbol{X}_{i-1},\boldsymbol{X}_i,\boldsymbol{X}_{i+1}\}$ 的长距离时空依赖高效建模。其次，为进一步细化局部特征，通过 Encoder 模块的层级化编码结构深度聚合帧间时序信息，逐步压缩特征空间维度，再利用 Decoder 模块对编码后的高维压缩特征进行维度恢复，以输出融合长距离时空关联信息与局部细节的时域特征 $\{\boldsymbol{F}_{i-1}^{\mathrm{t}},\boldsymbol{F}_i^{\mathrm{t}},\boldsymbol{F}_{i+1}^{\mathrm{t}}\} \in \mathbb{R}^{H\times W\times C}(i=1,2,\cdots,n)$。

1.3.2 NPCN-IMCS

针对视频压缩与海上多船航行场景引发的帧间运动估计偏差问题，受相邻像素偏移的强相关性[18-22]启发，在 NPCN-IMCS 中构建端到端的分层偏移量估计网络结构，并结合可变形卷积进一步对齐帧间特征，有效提高特征对齐精度。首先，通过 2×2 卷积对$\{\boldsymbol{X}_{i-1},\boldsymbol{X}_i,\boldsymbol{X}_{i+1}\}$执行 2 倍下采样操作，再经 3×3 卷积与 ReLU 激活函数聚焦于运动相关的特征信息，得到初步运动特征；为进一步深化运动特征表达，通过两次下采样操作扩大网络感受野，压缩特征空间维度以减少计算开销，并借助残差网络强化运动特征的非线性表达。其次，通过 2×2 转置卷积依次执行两次 2 倍上采样操作，逐步恢复特征空间维度，并使用 3×3 卷积与双线性插值(Bilinear)操作，对各尺度特征进行偏移量细化，得到不同尺度下的偏移特征，再通过逐元素相加操作融合各尺度偏移信息，并利用可变形卷积自适应调整卷积核采样位置，输出最终偏移量。将偏移量与 DSWN-FES 输出的时域特征执行逐元素相加操作，输出对齐后的多帧时域特征$\{\boldsymbol{F}_{i-1}^{\mathrm{T}},\boldsymbol{F}_i^{\mathrm{T}},\boldsymbol{F}_{i+1}^{\mathrm{T}}\}\in\mathbb{R}^{H\times W\times C}(i=1,2,\cdots,n)$，实现海上多船航行场景下的帧间特征的精准对齐，有效抑制运动伪影等时域质量退化，进而提升多功能航标监控视频流的时序连贯性。

综上，DSWN-FES 和 NPCN-IMCS 分别利用动态移位窗口和帧间运动补偿，在高效建模长距离时空依赖的同时，实现海上多船航行场景下帧间特征精准对齐，有效缓解了帧模糊、帧间运动估计偏差等时域质量退化问题，为后续 DFP-CAF 模块提供高效对齐的时域特征输入。

1.4 DFP-CAF

对于 MFSN-SEB 输出的空域特征$\boldsymbol{F}_i^{\mathrm{R}}$与 SWIMCN-MEB 输出的时域特征$\{\boldsymbol{F}_{i-1}^{\mathrm{T}},\boldsymbol{F}_i^{\mathrm{T}},\boldsymbol{F}_{i+1}^{\mathrm{T}}\}$，若仅使用简单拼接或固定权重加权的方式进行融合，将无法根据航标监控视频流中静态海面背景、动态船舶目标的内容差异自适应调整融合权重，导致 MFSN-SEB 的空域细节恢复优势与 SWIMCN-MEB 的时序信息利用优势难以充分互补[23-26]。为此，设计 DFP-CAF 模块以实现内容自适应融合，提高监控视频流的质量增强效果。首先，针对$\boldsymbol{F}_i^{\mathrm{R}}$，通过最大池化与平均池化操作聚合其中的空间维度关键信息，再经由 3×3、5×5、7×7 组成的多尺度卷积单元提取多尺度空域信息，并在通道维度上进行拼接得到增强后的空域特征$\hat{\boldsymbol{F}}_i^{\mathrm{R}}\in\mathbb{R}^{H\times W\times C}(i=1,2,\cdots,n)$。其次，针对$\{\boldsymbol{F}_{i-1}^{\mathrm{T}},\boldsymbol{F}_i^{\mathrm{T}},\boldsymbol{F}_{i+1}^{\mathrm{T}}\}$，通过最大池化与平均池化操作捕获帧间时序信息，再经 1×1 卷积整合帧间特征，得到增强后的时域特征$\hat{\boldsymbol{F}}_i^{\mathrm{T}}\in\mathbb{R}^{H\times W\times C}(i=1,2,\cdots,n)$。最后，为避免简单拼接无法适配监控视频流内容动态变化问题，通过 3×3 卷积提取监控视频帧序列$\{\boldsymbol{X}_{i-1},\boldsymbol{X}_i,\boldsymbol{X}_{i+1}\}$的全局时空特征，经 ReLU 激活函数强化非线性表达后，由 Sigmoid 函数映射至[0,1]区间，生成特征感知权重$\omega_i(i=1,2,\cdots,n)$，在ω_i指导下，根据视频内容变化自适应调整$\hat{\boldsymbol{F}}_i^{\mathrm{R}}$与$\hat{\boldsymbol{F}}_i^{\mathrm{T}}$的权重占比，实现基于监控视频流内容的特征自适应融合，如式(6)所示：

$$\boldsymbol{Y}_i=(1-\omega_i)\cdot\hat{\boldsymbol{F}}_i^{\mathrm{R}}+\omega_i\cdot\hat{\boldsymbol{F}}_i^{\mathrm{T}} \tag{6}$$

式中：·表示标量与特征图的向量乘操作；+表示逐元素加操作；$\boldsymbol{Y}_i\in\mathbb{R}^{H\times W\times C}(i=1,2,\cdots,n)$是增强后视频帧。重复上述过程得到增强后视频帧序列$\{\boldsymbol{Y}_{i-1},\boldsymbol{Y}_i,\boldsymbol{Y}_{i+1}\}\in\mathbb{R}^{H\times W\times C}(i=1,2,\cdots,n)$。

综上，DFP-CAF 通过双增强分支分别挖掘监控视频流单帧空域特征与视频流多帧时域特征，充分发挥 MFSN-SEB 与 SWIMCN-MEB 各自的优势，借助动态特征感知实现监控视频流中静态海面背景下单帧空域特征主导、动态船舶航行场景下多帧时域特征主导的内容自适应融合，有效解决了现有方法未考虑海上航标监控视频流场景特性问题，显著提升了船舶、航道标识等关键监控目标的特征辨识度。

1.5 CAF-VSMM 视频质量增强算法

为进一步说明 CAF-VSMM 各模块之间数据流转过程，并增强本文方法的可复现性[27-31]，给出 CAF-VSMM 视频质量增强算法描述，如算法 1 所示。首先，以压缩后监控视频帧序列$\{\boldsymbol{X}_{i-1},\boldsymbol{X}_i,\boldsymbol{X}_{i+1}\}$(其中$\boldsymbol{X}_i\in\mathbb{R}^{H\times W\times C}$)作为输入，利用 MFSN-SEB 提取当前帧的多尺度空域细节特征，同时通过 SWIMCN-MEB 对相邻帧进行运动补偿与时域特征对齐；然后，DFP-CAF 根据视频内容自适应预测空域特征与时域特征的融合权重，并完成两类特征的动态融合；最后，输出增强后的视频帧$\boldsymbol{Y}_i\in\mathbb{R}^{H\times W\times C}(i=1,2,\cdots,n)$。

算法 1 CAF-VSMM 视频质量增强算法

输入 压缩后监控视频帧序列$\{\boldsymbol{X}_{i-1},\boldsymbol{X}_i,\boldsymbol{X}_{i+1}\}$，其中$\boldsymbol{X}_i\in\mathbb{R}^{H\times W\times C}$

输出 增强后的视频帧$\boldsymbol{Y}_i\in\mathbb{R}^{H\times W\times C}$

1)$\hat{\boldsymbol{F}}_i^{\mathrm{R}}\leftarrow$MFSN-MEB($\boldsymbol{X}_i$)

//通过 MFSN-SEB 获取视频帧的空域特征

2)$\{\boldsymbol{F}_{i-1}^{\mathrm{T}},\boldsymbol{F}_i^{\mathrm{T}},\boldsymbol{F}_{i+1}^{\mathrm{T}}\}\leftarrow$SWIMCN-MEB($\boldsymbol{X}_{i-1},\boldsymbol{X}_i,\boldsymbol{X}_{i+1}$)

//通过 SWIMCN-MEB 获取视频的时域特征

3)($\hat{\boldsymbol{F}}_i^{\mathrm{R}},\hat{\boldsymbol{F}}_i^{\mathrm{T}},\omega_i$)$\leftarrow$DFP-CAF($\boldsymbol{F}_i^{\mathrm{R}}\{\boldsymbol{F}_{i-1}^{\mathrm{T}},\boldsymbol{F}_i^{\mathrm{T}},\boldsymbol{F}_{i+1}^{\mathrm{T}}\}$)

//通过 DFP-CAF 增强空域特征与时域特征以及动态
//特征感知权重

4)$\boldsymbol{Y}_i \leftarrow (1-\omega_i)\cdot\hat{\boldsymbol{F}}_i^{\mathrm{R}}+\omega_i\cdot\hat{\boldsymbol{F}}_i^{\mathrm{T}}$//对应式(6)

//根据动态特征感知权重,自适应融合增强后的空域
//特征与时域特征,并得到增强后的视频帧

5)return $\boldsymbol{Y}_i$

2 数据集构建

受限于标载通信网络带宽,多功能航标传输的监控视频流经压缩处理后存在压缩失真问题,无法直接作为视频质量增强任务所需的训练视频数据[32-34]。因此,在近海航标端(图 4 为海上多功能航标工作现场)部署高清摄像头,直接利用 4G 网络获取港口群航道水域未压缩的监控视频流,以此制作 PC-WMVD,为 CAF-VSMM 模型训练提供数据支撑。

图 4 海上多功能航标工作现场

Fig.4 Marine multi-functional navigation mark work site

PC-WMVD 构建步骤如下:

步骤 1 选取某港口群航道水域作为数据采集区域,如图 5 所示,采集不同时段与气象条件下监控视频流,共收集 108 个原始航道监控视频流,帧率固定为 25 帧/s,分辨率统一为 1 920×1 080 像素。

步骤 2 按照训练集与测试集 8∶2,对原始监控视频流进行划分。图 6 展示了数据集部分视频帧。

步骤 3 对原始航道监控视频流利用 H.264/AVC 标准进行压缩编码处理,并对其逐帧标注质量标签,构建"原始航道监控视频流-压缩后监控视频流-质量标签"的配对数据。

步骤 4 对 PC-WMVD 进行数据校验,确保数据集的完整性与有效性;通过帧级别的时间戳对齐,保证原始与压缩后监控视频流的时序一致性。

图 5 PC-WMVD 中不同时段与气象条件下的示例图

Fig.5 Example diagrams of PC-WMVD under different time periods and meteorological conditions

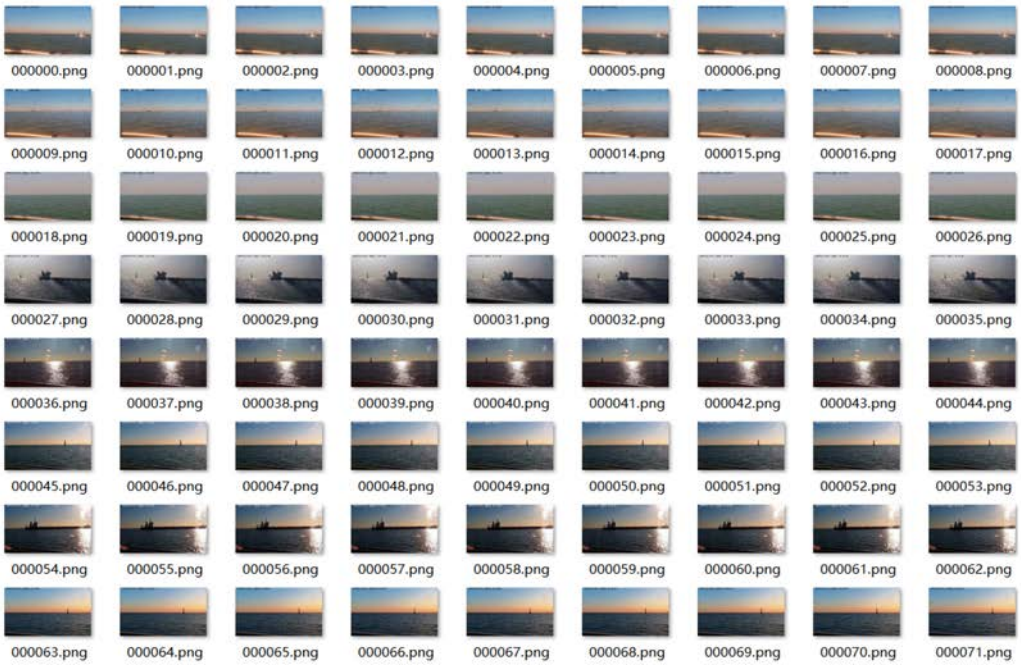


图 6 PC-WMVD 部分视频帧

Fig.6 Part of the video frames of PC-WMVD

3 实验与结果分析

3.1 实验环境和设置

3.1.1 实验环境

运行环境为某港口群智能监控系统岸基端服务平台，环境设置如表 1 所示。

表 1 实验环境

Table 1 Experimental environment

实验环境	配置信息
CPU	Intel® Core™ i7-10875H
GPU	RTX 4090D
显存/GB	24
操作系统	Windows 10
编程语言	Python 3.8
深度学习框架	PyTorch 1.10.0

3.1.2 实验设置

1)实验评价指标。

为验证 CAF-VSMM 的有效性，采用以下指标评价模型：(1)峰值信噪比(PSNR)，是衡量视频质量增强效果的经典客观指标，通过计算视频帧之间的均方误差(MSE)衡量视频帧间差异，其值大小与视频帧质量呈正相关，数值越高，增强后的视频帧越接近原始视频帧；(2)结构相似性(SSIM)，基于人眼视觉特性的客观评价指标，从亮度、对比度和结构 3 个维度评估视频帧的帧间相似度；(3)视频多方法评估融合(VMAF)，基于机器学习构建的视频主观感知质量评价指标，用于量化视频之间的人眼主观感知质量差异，评价结果与人类主观视觉感知具有高度一致性，有效弥补了 PSNR 和 SSIM 在衡量视频质量上的不足。此外，还采用参数量(Params)和浮点运算量(FLOPs)分别衡量模型内存资源消耗和推理运算开销。

2)实验参数设置。

CAF-VSMM 的实验训练参数设置如表 2 所示。其中，Batch Size 代表每批次训练的样本数量；Period 表示总训练次数；Epoch 表示训练轮数；Optimizer 为模型所采用的优化器，CAF-VSMM 使用 AdamW 优化器，通过权重衰减有效缓解过拟合；Learning rate 代表学习率初始值，学习率采用 CosineAnnealingRestartLR 调度器进行动态调整，通过周期性余弦函数在训练初期维持较大值以加快模型收敛，训练后期逐步衰减以细化参数更新，最小降至 1.0×10^{-7}，从而提升增强效果的稳定性与泛化能力；Weight decay 为权重衰减参数，用于训练过程中的正则化约束；QP 为量化参数，数值越高则编码压缩程度越强，输出码率越低。模型训练的损失函数采用 L1 Loss 与 Perceptual Loss 构成的混合损失函数，用于约束增强后视频帧与原始参考帧之间的特征差异。其中，L1 Loss 主要用于约束像素级的绝对误差，保证增强帧基础画质精度的一致性；Perceptual Loss 基于预训练的 VGG19 网络提取特征并计算特征差异，约束增强帧的高层视觉感知质量。

表 2 训练参数设置

Table 2 Training parameter setting

参数名	参数值
Batch Size	4
Period	150 000
Epoch	16
Optimizer	AdamW
Learning rate	2.0×10^{-4}
Weight decay	1.0×10^{-4}
QP	27

3.2 对比实验

3.2.1 客观指标评价结果

选择 AFD-Former[20]、GAN-PQ[21]、ConvNeXt[22]、UVE-Net[23]等作为对比模型，在自建数据集 PC-WMVD 和公开数据集 JCT-VC 上，使用本文 CAF-VSMM 模型以及对比模型进行视频质量增强，结果如表 3 所示，其中最优结果以粗体标明，次优结果以下划线标明，下同。

表 3 各模型在 PC-WMVD 和 JCT-VC 上的视频质量增强结果

Table 3 Video quality enhancement results of each model on PC-WMVD and JCT-VC

Model	PC-WMVD			JCT-VC			Params/10^6	FLOPs/10^9
	PSNR/dB	SSIM/%	VMAF/%	PSNR/dB	SSIM/%	VMAF/%		
AFD-Former	37.92	82.34	67.50	25.83	69.79	49.96	**16.56**	46.87
GAN-PQ	38.57	91.86	68.43	26.19	77.90	50.63	96.35	81.43
ConvNeXt	38.69	97.57	69.56	26.27	82.70	51.49	<u>30.70</u>	**43.80**
UVE-Net	<u>38.81</u>	<u>98.14</u>	<u>70.64</u>	<u>27.03</u>	<u>83.18</u>	<u>52.21</u>	35.47	<u>46.56</u>
CAF-VSMM	**42.71**	**99.33**	**77.14**	**28.66**	**88.60**	**55.40**	37.52	51.74

从表 3 中实验结果可知，CAF-VSMM 的视频质量增强效果均达到最佳，在 PC-WMVD 上 PSNR、SSIM 和 VMAF 较次优模型分别提升 3.90 dB、1.19 和 6.50 百分点，达到了 42.71 dB、99.33%和 77.14%，在 JCT-VC 上 PSNR、SSIM 和 VMAF 较次优模型分别提升 1.63 dB、5.42 和 3.19 百分点，达到了 28.66 dB、88.60% 和 55.40%，充分验证了 CAF-VSMM 对多功能航标监控视频流质量增强任务的有效性。此外，CAF-VSMM 的 Params 和 FLOPs 分别为 37.52×10^6 和 51.74×10^9，处于中等水平。

CAF-VSMM 相较于 4 种对比模型在 PC-WMVD 和 JCT-VC 上均有不同程度的性能提升，具体原因分析如下：

1）AFD-Former 通过 Transformer-CNN 构建双分支结构，其中 CNN 分支负责局部纹理特征的提取，Transformer 分支负责长距离依赖的建模，并借助 AFDU 实现分支权重的动态调整。但该权重调整机制仅聚焦于通道维度的非对称分配，未充分考虑视频帧间动态运动的建模需求。CAF-VSMM 通过帧内多尺度空域特征的分层提取与融合，并结合动态移位窗口机制打破窗口边界限制，高效建模跨窗口长距离时空依赖；此外，利用动态特征感知权重，充分发挥各分支网络的优势，因此结果更优。

2）GAN-PQ 依赖 ASFF 与 SFT 模块实现双帧特征的融合，虽通过改进的最近最少使用策略优化参考帧有效性，但仅聚焦双帧间的浅层信息互补，多尺度细节恢复效果有限。CAF-VSMM 通过单帧增强分支 MFSN-SEB 强化空域特征表达能力，再整合多维度特征，在细节修复与整体质量提升上更具优势。

3）ConvNeXt 通过大核卷积与通道转换提升单帧特征表达能力，但其设计更侧重视频帧内的空域特征提取，缺乏对帧间时序关联的有效建模，易导致增强后视频帧序列出现时序不一致问题。CAF-VSMM 通过动态移位窗口建模长距离时空依赖，结合邻域像素相关性运动补偿实现帧间特征的精准对齐，有效保障了视频的时序一致性。

4）UVE-Net 采用双分支架构，通过 AE-Net 处理中间帧，借助 FEGM 与 FRGM 指导相邻帧恢复，虽避免了帧对齐的计算开销，但仅依赖中间帧的间接引导，对多帧间运动关联的建模不足，帧间信息利用效率有限。CAF-VSMM 可精准捕捉物体不规则运动，实现多帧特征的高效对齐，结合单帧分支提取的多尺度空域特征，并基于动态特征感知自适应融合两类特征，充分挖掘多帧互补信息，更贴合人眼主观视觉感知。

3.2.2 可视化比较结果

图 7 分别展示了在自建数据集 PC-WMVD 和公开数据集 JCT-VC 中的测试视频序列。从图 7(a)中可以看出，相较于次优模型 UVE-Net，CAF-VSMM 能更有效地恢复船舶结构轮廓、静态背景的纹理以及亮度分布特征，支撑港口群航道智能监控场景下船舶目标识别等下游任务的数据需求；从图 7(b)中可以看出，CAF-VSMM 能够更清晰地恢复人物面部细节与衣物纹理的边缘质感。

图 7 CAF-VSMM 与次优模型 UVE-Net 在 PC-WMVD 和 JCT-VC 上可视化比较结果

Fig.7 Visual comparison results of CAF-VSMM and UVE-Net suboptimal model on PC-WMVD and JCT-VC

3.3 消融实验

为验证单帧增强分支 MFSN-SEB、多帧增强分支 SWIMCN-MEB 以及内容自适应融合模块 DFP-CAF 的有效性，在自建数据集 PC-WMVD 和公开数据集 JCT-VC 上进行消融实验，对比不同模块对视频质量增强效果的影响，结果如表 4 所示，其中

“√”表示使用该模块。仅添加 MFSN-SEB 的模型 A 在实验结果中的 Params 和 FLOPs 最低，分别达到了 16.42×10^6 和 21.34×10^9，在 PC-WMVD 上 PSNR、SSIM 和 VMAF 分别达到了 39.21 dB、98.61%和 70.82%，在 JCT-VC 上 PSNR、SSIM 和 VMAF 分别达到了 27.02 dB、86.58%和 52.45%，验证了单帧增强分支在抑制压缩伪影、恢复空域细节方面的有效性。同时，指标结果均低于其他模型，这也说明仅依靠单帧质量增强难以充分利用视频序列的时序信息，对动态船舶目标的细节恢复能力存在瓶颈。仅添加 SWIMCN-MEB 的模型 B，在 Params 和 FLOPs 略有增加的情况下，在 PC-WMVD 上 PSNR、SSIM 和 VMAF 较模型 A 分别提升了 0.64 dB、0.42 和 1.45 百分点，在 JCT-VC 上 PSNR、SSIM 和 VMAF 较模型 A 分别提升了 0.09 dB、0.81 和 1.53 百分点，表明多帧增强分支通过建模帧间时序依赖关系，能够有效利用相邻帧的互补信息，提升对船舶航行动态场景的增强性能。模型 C 采用简单拼接方式融合 MFSN-SEB 和 SWIMCN-MEB 的特征输出，相较于模型 A 与 B，在 PC-WMVD 上 PSNR、SSIM 和 VMAF 分别提升了 1.05 dB、0.46 和 1.95 百分点与 0.41 dB、0.04 和 0.50 百分点，在 JCT-VC 上 PSNR、SSIM 和 VMAF 分别提升了 0.90 dB、0.90 和 1.60 百分点与 0.81 dB、0.09 和 0.07 百分点，验证了双分支并行增强架构在视频质量增强任务中的协同增益，然而未根据视频内容动态调整融合权重，导致其性能提升幅度有限，且复杂度较单分支模型有所增加，Params 和 FLOPs 分别达到了 34.78×10^6 和 46.81×10^9，说明仅通过简单拼接方式难以充分发挥两类增强分支的互补优势。在模型 C 的基础上添加 DFP-CAF 的模型 D，在 PC-WMVD 上 PSNR、SSIM 和 VMAF 较模型 C 分别提升了 2.45 dB、0.26 和 4.37 百分点，在 JCT-VC 上 PSNR、SSIM 和 VMAF 较模型 C 分别提升了 0.74 dB、1.12 和 1.35 百分点，而 Params 和 FLOPs 仅有少量增加，在性能与效率之间实现了良好的平衡，充分验证了 DFP-CAF 能够根据航标监控视频流的场景特性动态调整空域与时域特征的融合权重，在静态背景区域强化单帧细节恢复，在动态目标区域强化多帧时序信息利用，有效解决了传统融合方式对本场景适配性不足的问题，实现了性能的显著提升。

表 4 在 PC-WMVD 和 JCT-VC 上的消融实验结果

Table 4 Ablation experiment results on PC-WMVD and JCT-VC

对照组	MFSN-SEB	SWIMCN-MEB	DFP-CAF	PC-WMVD			JCT-VC			Params/10^6	FLOPs/10^9
				PSNR/dB	SSIM/%	VMAF/%	PSNR/dB	SSIM/%	VMAF/%		
A	√			39.21	98.61	70.82	27.02	86.58	52.45	**16.42**	**21.34**
B		√		39.85	99.03	72.27	27.11	87.39	53.98	19.36	27.45
C	√	√		40.26	99.07	72.77	27.92	87.48	54.05	34.78	46.81
D	√	√	√	**42.71**	**99.33**	**77.14**	**28.66**	**88.60**	**55.40**	37.52	51.74

3.4 工程应用性能评估实验

为验证 CAF-VSMM 的工程应用价值，以某港口群航道智能监控系统为应用对象。利用自建数据集 PC-WMVD 对 CAF-VSMM 进行训练，并将预训练的 CAF-VSMM 部署至基于 Java 开发的港口群航道智能监控系统中，实现压缩后监控视频流的质量增强任务。其中，针对 Java 解释执行带来的性能瓶颈，该监控系统使用 JNI(Java Native Interface)技术调用 LibTorch 完成模型推理，利用 LibTorch 编译执行的低延时特性，降低监控视频流质量增强的时间开销，以满足航道监控场景下的实时性要求。分别从客观质量、实时性、可视化以及对下游任务影响(以船舶目标识别为例)等方面，全方位测试监控系统中 CAF-VSMM 的质量增强性能。

3.4.1 客观质量评价结果

PSNR、SSIM 和 VMAF 是评估视频质量增强效果的经典客观指标，如第 3.2 节和第 3.3 节所示，通过消融和对比实验，从多个维度验证了 CAF-VSMM 的有效性。同时，表 3 实验结果表明其在自建数据集 PC-WMVD 上，在以上 3 个指标上较次优模型 UVE-Net 分别提升 3.90 dB、1.19 和 6.50 百分点，达到了 42.71 dB、99.33%和 77.14%，在 JCT-VC 上较次优模型 UVE-Net 分别提升 1.63 dB、5.42 和 3.19 百分点，达到了 28.66 dB、88.60%和 55.40%。

3.4.2 实时性评估结果

为验证 CAF-VSMM 部署至港口群航道智能监控系统后，压缩监控视频流的质量增强处理速度能否满足系统实时性要求，使用帧率(FPS)衡量视频质量增强模型的实时处理能力，通过统计模型在单位时间内完成质量增强处理的视频帧数，反映模型的处理效率。视频帧处理速度示例如图 8 所示(仅

显示 3 次测速记录),经多次统计,部署后 CAF-VSMM 处理 25 帧视频帧的平均处理时间为 0.602 1 s,FPS 为 41.52 帧/s,处理速度大于监控系统最高帧率 25 帧/s,满足实时性要求。

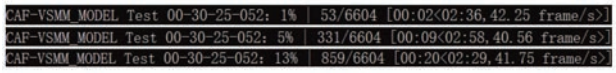
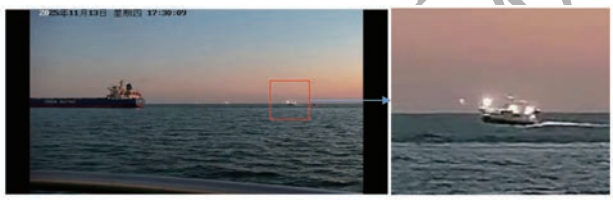


图 8 视频帧处理速度示例图

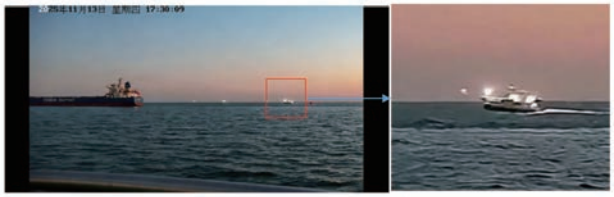
Fig.8 Example diagrams of video frame processing speed

3.4.3 质量增强可视化结果

图 9 分别展示了质量增强前后的监控视频流回放画面。增强前监控视频流中远处小型船舶目标轮廓模糊、航行灯光以及海面波纹细节被噪声掩盖,如图 9(a)中红框所示,经 CAF-VSMM 增强后,船体轮廓、航行灯光以及海面波纹的清晰度有所提升,如图 9(b)中红框所示。



(a)增强前



(b)增强后

图 9 监控视频流质量增强前后对比结果

Fig.9 Comparison results of surveillance video stream quality before and after enhancement

3.4.4 船舶目标识别可视化结果

为验证 CAF-VSMM 质量增强效果对下游任务存在积极影响,以船舶目标识别为例,采用 YOLOv8 对质量增强前后的监控视频流进行船舶目标识别,识别置信度阈值设置为 25%。图 10 分别展示了基于 CAF-VSMM 增强前后的目标识别结果。

从图 10(a)中可以看出,受压缩带来的失真影响,仅左侧大型船舶被成功识别,识别置信度为 83%;从图 10(b)中可以看出,监控视频流经 CAF-VSMM 增强后,左侧大型船舶的识别置信度提高了 2 百分点,达到 85%,同时右侧远处的小型船舶也被成功识别,识别置信度为 26%。由此可见,CAF-VSMM 质量增强效果对船舶目标识别等下游任务具有积极影响。

(a)增强前 (b)增强后

图 10 监控视频流质量增强前后的船舶目标识别可视化比较结果

Fig.10 Visual comparison results of ship target identification before and after enhancing the quality of surveillance video streams

3.5 失败案例分析

尽管 CAF-VSMM 在自建数据集 PC-WMVD 和公开数据集 JCT-VC 上的客观评价指标以及可视化结果均表现优异,并在港口群航道智能监控系统中取得了良好的工程应用效果,但在部分极端场景下仍存在一定的局限性。如图 11 中红框所示,在远距离微小船舶目标场景中,由于船舶目标仅占据较少像素,且经过低码率压缩后,大量高频纹理信息已经不可逆丢失,即使结合多帧信息进行细节恢复,增强后的远距离微小目标仍可能出现边缘模糊、轮廓不完整等现象。未来可进一步结合目标先验信息、自适应运动补偿等策略,提高模型在此类场景下的质量增强性能。

图 11 失败案例

Fig.11 A failure case

4 结束语

海上多功能航标监控视频流具有显著区别于通用监控视频的场景特性,其质量增强需同时应对低带宽压缩失真造成的空域与时域质量退化问题。为此,本文从空域增强(单帧)、时域建模(多帧)与内容自适应融合 3 个层面构建视频质量增强模型。在空域层面,通过强化多尺度特征表达并挖掘单帧中的潜在结构与时序依赖信息,有效提升海面区域的细节恢复能力并抑制压缩伪影;在时域层面,通过跨帧特征对齐与时空依赖建模,实现多船场景下关键目标区域的结构补偿与时序一致性恢复;在融合层面,

依据视频内容特征动态调整空域与时域信息的贡献权重，使静态海面区域以空域增强为主导、动态目标区域以时域信息补偿为主导，从而在统一框架下实现对静态背景与动态目标的差异化质量增强。实验结果表明，所提方法在 PSNR、SSIM 与 VMAF 3 项指标上较现有方法均取得明显增益，同时也达到了增强质量与实时性能的良好平衡，可有效满足航道监控场景下的实时处理需求。进一步的示范应用结果验证了该方法能够有效恢复海上多功能航标压缩监控视频流中船舶轮廓、纹理细节及亮度分布，增强船舶与航道标识等关键目标的可辨识性，为下游船舶目标识别、跟踪等任务提供可靠的数据支撑，具备良好的工程实用性与推广应用价值。未来工作将围绕以下方面继续深化：在保持质量增强效果的前提下，进一步开展模型轻量化研究；探索与下游船舶目标识别/异常行为检测任务的联合优化方法，使其更好地服务于智慧航道建设。

参考文献

[1] 张胜男. 基于深度学习的海上船舶目标识别与跟踪[D]. 大连：大连交通大学，2025.
ZHANG S N. Recognition and tracking of ship targets at sea based on deep learning[D]. Dalian: Dalian Jiaotong University, 2025. (in Chinese)

[2] 郑海林. 面向智慧海事监管的船舶异常行为检测研究[D]. 上海：上海海事大学，2025.
ZHENG H L. Research on ship abnormal behavior detection for intelligent maritime supervision[D]. Shanghai: Shanghai Maritime University, 2025. (in Chinese)

[3] KAU L J, TSENG C K, LEE M X. Perception-based H.264/AVC video coding for resource-constrained and low-bit-rate applications[J]. Sensors, 2025, 25(14): 4259.

[4] LI F Y, ZHAI H J, LIU T, et al. Learning compressed artifact for JPEG manipulation localization using wide-receptive-field network[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2024, 20(10): 1-23.

[5] IBRAHEEM M K, DVORKOVICH A V. Enhancing versatile video coding efficiency via post-processing of decoded frames using residual network integration in deep convolutional neural networks[C]//Proceedings of the 26th International Conference on Digital Signal Processing and Its Applications (DSPA). Washington D. C., USA: IEEE Press, 2024: 1-9.

[6] WANG T T, HU Z H, GUAN Y R. An efficient lightweight network for image denoising using progressive residual and convolutional attention feature fusion[J]. Scientific Reports, 2024, 14: 9554.

[7] 陈圣杰. 面向压缩视频质量增强的多帧对齐方法研究[D]. 成都：电子科技大学，2023.
CHEN S J. Research on multi-frame alignment method for compressed video quality enhancement[D]. Chengdu: University of Electronic Science and Technology of China, 2023. (in Chinese)

[8] 罗登晏. 基于深度学习的多帧压缩视频质量增强方法研究[D]. 成都：电子科技大学，2023.
LUO D Y. Deep learning methods for multi-frame quality enhancement on compressed video[D]. Chengdu: University of Electronic Science and Technology of China, 2023. (in Chinese)

[9] DENG J N, WANG L, PU S L, et al. Spatio-temporal deformable convolution for compressed video quality enhancement[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI Press, 2020: 10696-10703.

[10] ZHANG X J, YANG S, LUO W Y, et al. Video compression artifact reduction by fusing motion compensation and global context in a swin-CNN based parallel architecture[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI Press, 2023: 3489-3497.

[11] LIU J H, ZHOU M C, XIAO M. Deformable convolution dense network for compressed video quality enhancement[C]//Proceedings of the 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Washington D.C., USA: IEEE Press, 2022: 1930-1934.

[12] 魏柳. 多尺度压缩视频增强算法研究与应用[D]. 成都：电子科技大学，2024.
WEI L. Research and application of multi-scale compressed video enhancement algorithm[D]. Chengdu: University of Electronic Science and Technology of China, 2024. (in Chinese)

[13] ZHAO M Y, XU Y, ZHOU S G. Recursive fusion and deformable spatiotemporal attention for video compression artifact reduction[C]//Proceedings of the 29th ACM International Conference on Multimedia. New York, USA: ACM Press, 2021: 5646-5654.

[14] LI H, HE X H, XIONG S H, et al. A compressed video quality enhancement algorithm based on CNN and transformer hybrid network[J]. The Journal of Supercomputing, 2024, 81(1): 144.

[15] XIE C H, TAN M X, GONG B Q, et al. Adversarial examples improve image recognition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Washington D. C., USA: IEEE Press, 2020: 816-825.

[16] LI J, YU Z T, HE Z Q, et al. PGD-Imp: rethinking and unleashing potential of classic PGD with dual strategies for imperceptible adversarial attacks[C]//Proceedings of the 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Washington D. C., USA: IEEE Press, 2025: 1-5.

[17] MAENG J B, KWON H. Time-constrained adversarial attacks for video recognition models: temporally sparse but effective perturbations[J]. Machine Vision and Applications, 2026, 37(2): 22.

[18] HUANG Z J, SUN J, GUO X P. FastCNN: towards fast and accurate spatiotemporal network for HEVC compressed video enhancement[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2023, 19(3): 1-22.

[19] LIU Z, HU H, LIN Y T, et al. Swin Transformer V2: scaling up capacity and resolution[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Washington D. C., USA: IEEE Press, 2022: 11999-12009.

[20] ZHANG X, CAI N, ZHANG H, et al. AFD-former: a hybrid transformer with asymmetric flow division for synthesized view quality enhancement[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(8): 3786-3798.

[21] AGNOLUCCI L, GALTERI L, BERTINI M, et al. Perceptual quality improvement in videoconferencing using keyframes-based GAN[J]. IEEE Transactions on Multimedia, 2024, 26: 339-352.

[22] PARK S W, JUNG S H, SIM C B. NeXtSRGAN: enhancing super-resolution GAN with ConvNeXt discriminator for superior realism[J]. The Visual Computer, 2025, 41: 7141-7167.
[23] XIE Y F, KONG L W, CHEN K, et al. UVEB: a large-scale benchmark and baseline towards real-world underwater video enhancement [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Washington D. C., USA: IEEE Press, 2024: 22358-22367.
[24] YUE J, YE M, JI L P, et al. A survey of deep-learning-based compressed video quality enhancement [J]. IEEE Transactions on Broadcasting, 2025, 71(4): 977-992.
[25] YUAN Y, LI E Y, ZHANG J W, et al. PSTAN: a JND-aware pairwise spatio-temporal alignment network for compressed videos quality enhancement [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2026, 36(6): 8491-8505.
[26] XING Q L, ZHENG C, XU M, et al. Breaking the multi-enhancement bottleneck: domain-consistent quality enhancement for compressed images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2026, 48(4): 4692-4707.
[27] HUANG Z Y, CHAN Y L, KWONG N W, et al. Long short-term fusion by multi-scale distillation for screen content video quality enhancement[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2025, 35(8): 7762-7777.
[28] ZHU Q, HAO J H, DING Y K, et al. CPGA: coding priors-guided aggregation network for compressed video quality enhancement[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Washington D. C., USA: IEEE Press, 2024: 2964-2974.
[29] WANG W Z, DAI E Z. Multi-view video quality enhancement method based on multi-scale fusion convolutional neural network and visual saliency[J]. IEEE Access, 2024, 12: 33100-33108.
[30] SHALIN L A, KANSAL V, NIKITHA M, et al. Hybrid quantum-classical convolutional squeeze excitation neural network based HEVC video quality enhancement [C] // Proceedings of the 2nd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS). Washington D. C., USA: IEEE Press, 2024: 100-106.
[31] PEI S F, REN G M, ZHAO T S, et al. STDF-DSCS: a lightweight scheme for compressed video quality enhancement[C]// Proceedings of the 16th International Conference on Wireless Communications and Signal Processing (WCSP). Washington D. C., USA: IEEE Press, 2025: 1324-1329.
[32] GEHLOT S, SU G M. LViCAR: diffusion models for perceptual quality enhancement in video compression artifact reduction[C]//Proceedings of the 3rd International Workshop on Deep Multimodal Generation and Retrieval. New York, USA: ACM Press, 2025: 1-10.
[33] DESHPANDE S. Standards-based neural-network post-filters for improved video quality[C]//Proceedings of the 3rd Mile-High Video Conference. New York, USA: ACM Press, 2024: 75-81.
[34] CHOI W, YOON J. Real-time enhancement of low-quality video for constrained camera systems[C]//Proceedings of the 30th Annual International Conference on Mobile Computing and Networking. Washington D. C., USA: IEEE Press, 2024: 1659-1661.

文字编辑 陆燕菲
栏目编辑 赖玉玲