光电技术及应用

DOI: 10.16818/j.issn1001－5868.20260202002

基于物理噪声与时空变换器的RAW视频去噪方法

付政权*,徐本源,谢志华,张维刚
(重庆光电技术研究所,重庆 400060)

摘 要: 针对微光视频中由泊松－高斯噪声引起的低信噪比以及噪声随机分布导致的时空不稳定问题,文章提出一种噪声模型引导的时空网络(NM-STNet)。该方法通过噪声感知编码器将物理噪声统计特性嵌入特征学习过程,并利用基于变换器的时空融合模块,在无需显式运动估计的情况下实现多帧特征的自适应对齐与聚合。此外,引入基于物理模型的噪声合成及矫正策略,以提升有限训练数据条件下的泛化能力。实验结果表明,NM-STNet在真实弱光视频序列上取得了优于现有方法的去噪性能与时空一致性。

关键词: 低照度成像;深度学习去噪;低信噪比;RAW视频数据集
中图分类号: TP391.41 **文章编号:** 1001－5868(2026)03－0536－08

RAW Video Denoising via Physics-Guided Noise Modeling and Spatiotemporal Transformers

FU Zhengquan, XU Benyuan, XIE Zhihua, ZHANG Weigang
(**Chongqing Optoelectronics Research Institute, Chongqing 400060, CHN**)

Abstract: To address the low signal-to-noise ratio induced by Poisson－Gaussian noise and the spatiotemporal instability resulting from the random distribution of noise in low-light videos, this paper proposes a Noise Model-guided Spatiotemporal Network (NM-STNet). The method embeds physical noise statistics into the feature learning process via a noise-aware encoder and employs a Transformer-based spatiotemporal fusion module to achieve adaptive alignment and aggregation of multi-frame features without explicit motion estimation. Furthermore, a physics-based noise synthesis and correction strategy is introduced to enhance generalization capabilities under conditions of limited training data. Experimental results demonstrate that NM-STNet significantly outperforms existing methods in terms of both denoising performance and spatiotemporal consistency on real-world low-light video sequences.

Key words: low-light imaging; deep learning denoising; low signal-to-noise ratio; RAW video dataset

0 引言

随着图像传感器制造工艺的进步以及计算摄影技术的不断发展,像元尺寸小于1 μm的图像传感器也能通过多帧叠加及深度学习修复实现高质量的图片成像。然而,视频去噪这一典型应用仍然面临三类核心挑战:首先,真实传感器噪声呈现出由散粒噪声和读出噪声共同构成的泊松－高斯复合特性[1-3],其统计分布与信号强度密切相关[4-5];其次,场景和物体运动引入的帧间变化使视频序列具有强时间相关性,容易引发时空不一致问题[6-7];最后,真实弱光视频难以获取高质量的无噪参考,限制了大规模配对数据集的构建[8-10]。

针对上述问题,传统图像与视频去噪方法(如

收稿日期:2026－02－02.
* 通信作者:付政权

BM3D[11-13]和 WNNM[14]）主要依赖固定的先验，难以刻画视频中的时序动态特性。基于深度学习的去噪方法虽取得了一定进展[15-18]，但大多局限于单帧图像建模[16]，或依赖显式的运动补偿（如光流估计）[18]，在复杂或不规则运动条件下容易产生重影伪影。真实弱光视频中的噪声与合成训练中普遍假设的加性高斯白噪声不同，其遵循泊松－高斯分布。其中，散粒噪声的方差随场景亮度变化，而读出噪声则来源于成像电路[3]。在缺乏显式噪声建模的情况下，现有神经网络方法容易过拟合于特定数据集的统计特性，并在真实传感器条件下的表现不佳[19]。同时，跨帧的简单时序聚合往往会放大未对齐结构，易引发细节丢失或时间闪烁现象[20]。

基于此，本文提出一种噪声模型引导的时空网络（NM-STNet），将物理噪声建模与深度时空特征学习显式集成于统一框架中。针对不同传感器相关噪声水平和空间变化信噪比，设计了一个噪声感知编码器，将物理启发的泊松-高斯噪声模型嵌入到特征提取链路。进一步引入基于变换器的融合单元，无需显式运动估计即可自适应对齐和聚合跨多帧的特征，即使在动态运动和遮挡场景下也能保持时间稳定性。最后为扩展有限视频数据集的多样性和鲁棒性，提出散粒噪声增强和暗影校正来合成真实的噪声统计特性。

1　相关工作

1.1　RAW 成像中的噪声建模

噪声建模是低级图像复原中的一个基本组成部分。与经过色调映射和压缩的 sRGB 图像不同，RAW 数据保留了线性的传感器响应，其噪声统计主要由散粒噪声和读出噪声主导。泊松-高斯噪声模型已被广泛用于描述这种分布，其中散粒噪声遵循与信号强度成正比的泊松过程，而读出噪声近似为高斯分布且与信号无关。

传统参数拟合类方法通常依赖校准数据估计传感器噪声水平，但这类方法在变化的 ISO 或光照条件下常常失效。虽然近期研究已探索基于深度学习的噪声建模，代表性方案包括将物理先验展开到神经架构中，或从真实配对数据集中学习噪声感知先验。然而，由于大规模 RAW 视频数据集的可用性有限，此类模型在时间一致性和跨传感器泛化方面仍然受限。

本文提出的 NM-STNet 在特征编码过程中显式集成泊松-高斯噪声先验，通过噪声模型引导的编码器学习解耦信号相关的变化，从而能在不同曝光条件和传感器类型下提供鲁棒的表征。所提出 NM-STNet 整体训练架构如图 1 所示。

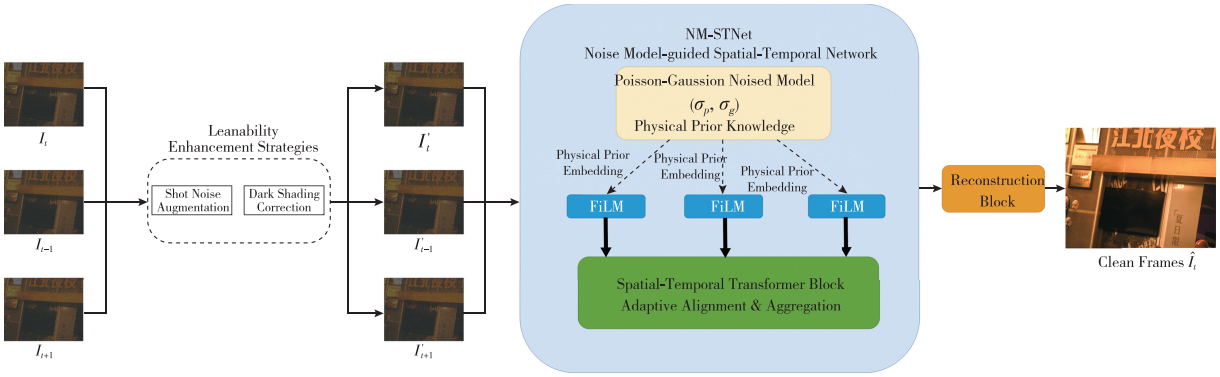


图 1　NM-STNet 整体框架图

1.2　视频去噪网络

深度学习的出现使得基于 CNN 的视频去噪器（如 ViDeNN[21]，DVDnet[22]，FastDVDnet[23]）取得了显著进展。其中，FastDVDnet 引入了一个两阶段 CNN pipeline 来联合处理多帧，既避免了显式的光流计算，又保持了时间上的平滑性。然而，这些方法通常在合成的高斯噪声上进行训练，并假设跨帧的噪声水平均匀，这会导致在 RAW 视频序列中性能下降。相比之下，NM-STNet 通过将物理噪声先验直接嵌入到时序融合过程中，能够联合推理传感器噪声和运动动态，从而使网络在低光照和高 ISO 条件下优于纯数据驱动的方法。

1.3　基于变换器的时空融合

变换器因其全局注意力机制，在视觉任务中显示出巨大潜力。对于视频去噪，基于变换器的架构（如 BasicVSR＋＋[24]，RVRT[25]）通过利用跨帧的

长程依赖性，展示了出色的时序建模能力。然而，在噪声RAW输入下，注意力机制的计算成本和不稳定性仍然是重大挑战。为缓解这个问题，轻量级混合设计被提出，将CNN编码器与时序变换器相结合，以平衡准确性和效率。

在NM-STNet中引入一个时空变换器模块（STTM），通过使用多尺度注意力窗口和噪声感知门控，以选择性融合相邻帧的信息。与传统的自注意力不同，所提出的STTM由估计的噪声方差引导，确保时间不一致或低信噪比区域获得自适应加权。这种物理信息引导的注意力机制可在增强时间一致性的同时，抑制传感器噪声跨帧的传播。

2 算法设计

2.1 噪声感知FiLM编码器

当前深度学习图像降噪方法仍采用隐式的噪声建模，仅通过大量样本分布学习来获取噪声与信号。然而，真实成像系统中的噪声往往呈现出明显非均匀性与信号相关性。尤其是在CMOS/CCD图像传感器中，由光子计数统计产生的泊松噪声与读出电路引入的高斯噪声同时存在，使得噪声幅值随亮度变化而动态波动，并在空间上表现出不同的噪声水平。传统卷积网络无法直接将这种物理先验嵌入特征提取过程，在高噪声、低照度或噪声不均匀的成像条件下，往往出现特征提取不稳定、纹理恢复失真等问题。

为解决上述问题，本文使用一种物理驱动的噪声感知FiLM编码器[26]，作为RAW视频去噪框架的首阶段特征提取模块。其核心是将成像系统的传感器噪声模型参数（如读出噪声标准差δ_r、光子增益k、曝光时间$t_{\exp}$、ISO增益g等）作为显式条件输入，通过特征线性调制（FiLM）机制对卷积特征进行逐通道仿射调制，使网络在底层表征阶段即具备对物理噪声强度与空间分布的感知能力，FiLM架构如图2所示。FiLM的数学形式如下：

$$FiLM(F;c)=\gamma(c)\cdot F+\beta(c) \tag{1}$$

式中，$F\in R^{C\times H\times W}$表示特征图；$\gamma(c)$，$\beta(c)$分别表示每个通道的缩放和平移参数；$\cdot$是通道维度广播乘法；$c=[\delta_r,k,t_{\exp},g]\in R^4$为条件向量，包括物理参数：读出噪声$\delta_r$，光子增益$k$，曝光时间$t_{\exp}$，ISO增益$g$。输出与输入$F$完全同形，每个通道的激活值已被条件$c$动态重加权和偏移。

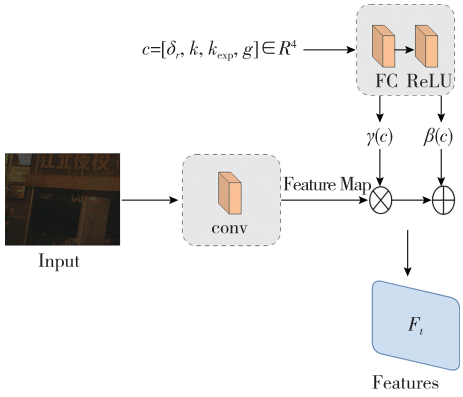


图2 FiLM架构示意图

为消除不同物理量纲带来的数值差异，对各参数进行线性归一化，将其映射至[0,1]区间。例如，对于ISO范围[100,1 600]，归一化形式为：$\hat{g}=(g-100)/(1\,600-100)$。其余参数依此类推，基于其在自研传感器中的实际取值范围进行标准化。此种线性调制方式具有三个显著优势：一是结构轻量化，不引入大量额外参数；二是能够直接将噪声模型映射到特征空间，实现显式噪声控制；三是不改变卷积核本身，仅调整特征响应，与现有网络结构兼容性极佳。

本模块以端到端方式联合优化：FiLM Generator、卷积编码器与后续降噪模块共享梯度，在训练中自适应学习“根据物理噪声参数调整特征响应”。这不仅避免了传统两阶段方法（先估计噪声再去噪）的误差累积问题，还使模型具备跨设备、跨曝光设置的泛化能力——即使面对未见过的传感器或极端ISO，也能基于物理参数合理调制特征。

实验结果表明，在多个真实低光RAW视频数据集上引入该编码器后，模型在小于1 lux照度、高ISO（不低于1 600）及空间噪声剧烈变化场景下的PSNR与视觉质量均显著优于现有盲去噪与条件去噪方法。综上，噪声感知FiLM编码器首次实现了“物理噪声参数→特征调制”的直接通路，为低光RAW视频去噪提供了兼具物理可解释性与深度学习表达力的新范式。

2.2 基于变换器的跨帧特征融合单元

在噪声感知FiLM编码器生成稳定的单帧表征之后，如何有效利用多帧之间的冗余信息，进一步提升重建质量与时间一致性，是视频降噪与低照度增强任务中的关键环节。现有多帧方法普遍采用显式运动估计（如光流、特征金字塔配准或基于块匹配的局部搜索）来实现不同帧在空间上与参考帧

对齐，再进行特征融合。然而，低照度场景中的多种因素使得显式对齐方法难以稳定工作：极端噪声干扰导致光流估计不稳定甚至完全失效；非刚体运动与快速动态变化使传统光流假设难以成立；遮挡与半遮挡区域在跨帧间不存在一致的空间对应关系，直接对齐会引入无效或错误的像素信息；真实图像信号处理（ISP）下的噪声非均匀性进一步加剧了空间对应关系的混乱。

针对上述问题，所提出的跨帧融合单元基于Transformer[27]的自注意力机制，实现了一种无需显式运动估计的隐式特征级对齐方式。如图3所示，该模块通过在特征空间中直接学习多帧对应关系，使网络能够根据结构一致性、亮度变化和局部运动趋势自适应决定每个区域的融合权重，从而实现跨帧动态信息的稳健聚合。

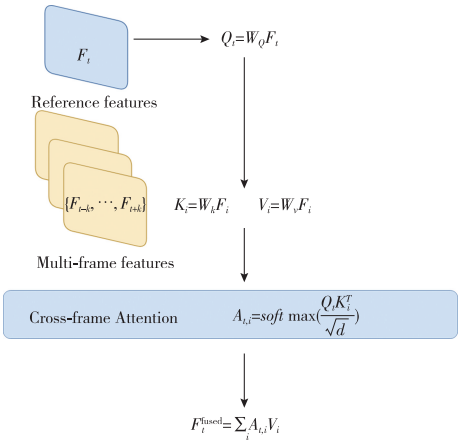


图3　跨帧特征融合单元框架

2.2.1　跨帧关系建模机制

对于参考帧 t 的特征表示 F_t，以及时间邻域内的多帧特征 $\{F_{t-k},\cdots,F_{t+k}\}$，本模块首先将其映射到查询（$\boldsymbol{Q}$）、键（$\boldsymbol{K}$）和值（$V$）空间：

$$\boldsymbol{Q}_t=\boldsymbol{W}_Q\boldsymbol{F}_t,\boldsymbol{K}_i=\boldsymbol{W}_k\boldsymbol{F}_i,\boldsymbol{V}_i=\boldsymbol{W}_v\boldsymbol{F}_i \tag{2}$$

式中，权重 $\boldsymbol{W}_Q,\boldsymbol{W}_k,\boldsymbol{W}_v$ 均为可学习的线性投影矩阵。随后，通过跨帧自注意力计算相似度：

$$\boldsymbol{A}_{t,i}=soft\max\left(\frac{\boldsymbol{Q}_t\boldsymbol{K}_i^T}{\sqrt{d}}\right) \tag{3}$$

式中，d 表示特征维度的缩放因子，注意力权重 $\boldsymbol{A}_{t,i}$ 表示参考帧在融合过程中从多帧借用信息的空间位置。最终，多帧融合输出为：

$$\boldsymbol{F}_t^{\text{fused}}=\sum_i\boldsymbol{A}_{t,i}\boldsymbol{V}_i \tag{4}$$

这一机制使网络在特征域中寻找匹配区域，而不依赖像素级对齐，因此能够应对复杂运动模式和严重噪声干扰。

2.2.2　无需显式对齐的自适应融合优势

所提出的变换器融合单元具备以下优势：一是抗噪性强，不受高噪声影响。传统光流在极端噪声下会错误估计运动矢量，而本方法直接在高维特征空间计算相似度，FiLM编码器提供的噪声稳定特征使注意力计算更可靠。二是自动适配不同区域的运动状态。对于静态区域，注意力权重会集中在所有帧的空间一致位置，从而提升信号质量；对于动态区域，注意力会自动跟随运动轨迹选择最匹配的位置；对于遮挡区域，注意力会降低被遮挡区域的权重以避免错误聚合。三是保持时序一致性，显著减少闪烁。由于注意力机制建立了跨帧的全局关联，网络可稳定输出在时间上连续的结构与亮度，极大减少局部瑕疵导致的时间抖动。四是可高度扩展。无需手工定义运动模型Transformer不依赖任何几何假设，适用于目标变形、视角变化、柔性物体或非结构化场景。

2.2.3　与整体降噪框架的协同作用

噪声感知FiLM编码器与Transformer融合单元形成高度互补的结构，FiLM使特征提取阶段即具备噪声稳健性，Transformer融合则将这种稳定特征在时间域充分扩展与聚合。二者协同使网络在动态与弱光环境中仍能获得高质量的结构特征。最终输出的跨帧融合特征兼具低噪声特性与时间一致性，为解码器图像重构提供有力支持。

2.3　真实降噪的数据合成与校正

尽管现有深度视频降噪方法在合成高斯噪声上表现优异，但在真实低光视频上性能显著下降。原因在于真实传感器噪声是信号相关的，主要由散粒噪声和读出噪声组成[1]；固定模式噪声（FPN），如暗电流非均匀性、热点像素，在长时间曝光下尤为显著；标注的真实低光视频数据集极度稀缺（如SID、DRV规模小、设备单一），难以覆盖不同ISO、温度、传感器型号下的噪声分布。

因此，仅依赖有限真实数据训练的模型（即使配备噪声感知FiLM和跨帧Transformer）无法泛化到未见成像条件。为此，提出一种物理驱动、可学习、端到端兼容的增强策略，在训练阶段主动合成具有真实统计特性的噪声视频，从而提升模型鲁棒性与泛化能力，如图4所示。

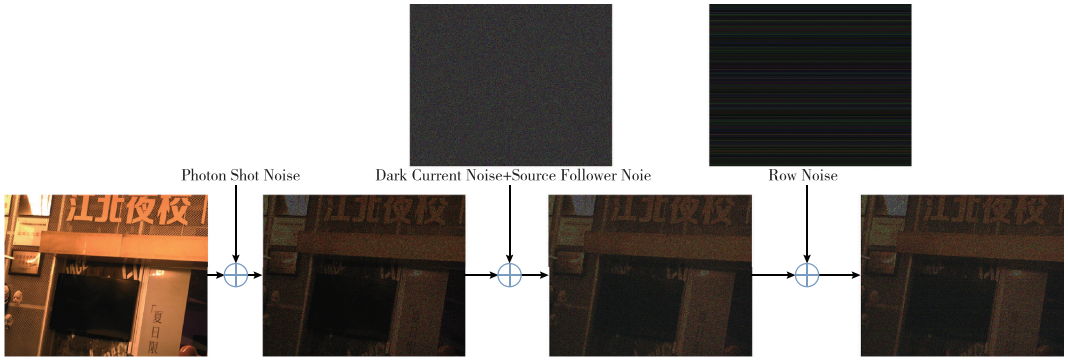


图4 噪声注入模型

2.3.1 物理成像模型基础

采用广泛接受的相机响应函数来建模从辐照度x到观测值z的映射：

$$z=\frac{1}{g}\cdot Poisson(g\cdot x+d)+n_{\text{read}} \tag{5}$$

式中，$x\in R^{+}$：理想无噪辐照度；$g>0$：模拟增益（与ISO正相关）；$d\in R^{+}$：暗电流随曝光时间与温度增长；$n_{\text{read}}\sim\boldsymbol{N}(0,\delta_{\text{read}}^{2})$：读出噪声，方差$\delta_{\text{read}}^{2}$通常与ISO相关。

该模型揭示了噪声的非平稳性：方差$V(z)=x+\frac{d}{g}+\frac{\delta_{\text{read}}^{2}}{g^{2}}$依赖于信号强度$x$和成像参数$(g,d,\delta_{\text{read}})$。

2.3.2 散粒噪声增强

在给定干净视频帧x前提下，使用如下参数合成带噪帧z：增益$g\sim U(g_{\min},g_{\max})$（对应ISO范围）；暗电流均值$\overline{d}\sim U(0,d_{\max})$；读出噪声标准差$\delta_{\text{read}}\sim U(\delta_{\min},\delta_{\max})$。最后生成空间变化的暗电流图：

$$d=\overline{d}\cdot(1+\alpha\cdot\varepsilon),\varepsilon\sim N(0,I) \tag{6}$$

式中，$\alpha\in[0,0.2]$控制暗电流非均匀性强度。应用物理成像模型公式(5)生成z。该方法可覆盖从“轻微低光”到“极端夜景”的连续噪声谱，远超真实数据集范围。

传统泊松采样不可微，阻碍梯度回传。采用Anscombe近似和高斯重参数化实现可微合成：

$$z\approx\frac{1}{g}\left(gx+d+\sqrt{gx+d}\cdot\xi_{1}+\delta_{\text{read}}\cdot\xi_{2}\right),$$
$$\xi_{1},\xi_{2}\sim N(0,1) \tag{7}$$

该近似在$gx+d>5$时高度准确，且完全可微，允许噪声合成模块嵌入训练流程。

2.3.3 暗影校正

真实传感器在无光照条件下仍会输出非零值（暗帧D），其包含全局偏移和空间相关FPN（如热点、温漂）。基于自研sensor采集构建暗帧字典$\{D_i\}_{i=1}^{K}$，具体操作为：在训练时随机选择一个暗帧D_i，缩放至当前曝光条件后，将其加到合成信号中：$z\leftarrow z+D_i$；同时提供“伪干净标签”：$x_{\text{corr}}=x+D_i$（即模型需学会去除暗影）；在推理时（无标签）引入轻量级暗影估计头，从输入帧预测$\widehat{D}$；将$\widehat{D}$作为FiLM编码器的额外条件输入，引导网络抑制FPN。

3 实验设计

为全面验证所提方法在真实视频去噪任务中的有效性，围绕自研传感器采集的数据构建了一套完整的评估体系。所有实验均基于自研传感器测试集，该数据集包含多场景、多光照条件下采集的配对带噪-干净视频序列，具有真实的噪声分布和复杂的动态特性。图片均在室外路灯或月光环境下采集，照度范围约为0.1～3 lux。相机光圈为F/1.2，焦距25 mm。为模拟超低光照环境，噪声图片曝光时间为1/300，增益为ISO 100/200/400/800/1 600五档相机增益。参考图片使用较长曝光时间1/25 s，增益使用原生最高增益ISO 1 600。所有对比基线方法（包括BM4D（video），FastDVD和BasicVSR++）均在自研训练集上重新训练或微调。

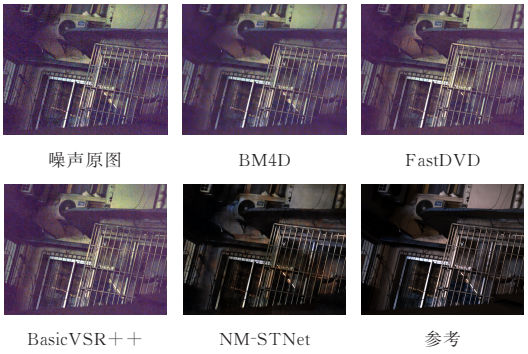
3.1 主结果对比

在自研传感器[28]测试集上评估了主流视频去噪方法的峰值信噪比（PSNR）与结构相似性（SSIM）指标，结果如表1所示。可以看出，尽管现有方法在合成噪声或SIDD等静态图像数据集上表现良好，但在真实视频噪声下性能显著下降。相比之下，所提出的完整模型在PSNR和SSIM指标上均取得最优结果，对真实传感器噪声的有强适应能

力(见图5)。

表1　不同方法降噪结果

方法	*PSNR*	*SSIM*
BM4D	28.88	0.787
FastDVD	32.36	0.874
BasicVSR++	35.53	0.921
NM-STNet	40.13	0.975



噪声原图　BM4D　FastDVD

BasicVSR++　NM-STNet　参考

图5　降噪算法效果对比

3.2　消融实验

3.2.1　数据策略消融

如表2所示，对比了不同训练数据来源对模型性能的影响。仅使用公开合成噪声(高斯+泊松)训练的模型泛化能力最弱；引入SIDD Video等公开真实数据集可提升性能，但仍受限于域偏移。使用自研传感器采集的有限配对数据训练后，性能明显改善；而进一步结合提出的生成式真实感带噪视频扩展策略(即利用少量配对数据驱动噪声模拟器生成大规模逼真训练样本)，模型达到最高PSNR(见图6)。这表明当真实配对数据稀缺时，在第2.3节中所提出的数据生成框架可作为高效的数据增强或半监督训练手段，显著提升模型鲁棒性。

表2　消融实验对比

方法	*PSNR*	*LTV*
w/o FiLM	31.56	0.025
w/o Cross-Frame	32.36	0.038
w/光流对齐+CNN融合	35.53	0.018
NM-STNet (FiLM+Transformer)	40.13	0.016

3.2.2　模块消融

为进一步评价关键模块对去噪质量与时序稳定性的贡献，设计了移除FiLM条件调制、完全忽略跨帧信息、采用传统光流对齐+CNN融合的方案三种实验方法。结果表明，移除FiLM条件调制导致性能下降且画面闪烁加剧；完全忽略跨帧信息则造成严重性能损失与剧烈时序抖动(见图7灯光)。采



P+G　+SIDD

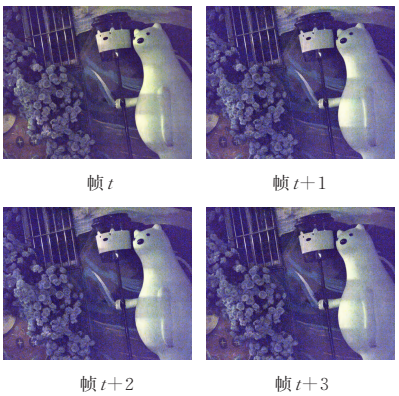
+Paired　+Augment

图6　数据消融实验对比

用传统光流对齐+CNN融合的方案虽有一定效果，但在复杂运动下易失效。相比之下，所提出的FiLM+Transformer联合架构不仅获得最高PSNR，还在亮度时序方差上表现最优，显著提升视觉流畅度。亮度时序方差定义如下：

$$LTV=\frac{1}{T}\sum_{t=1}^{T-1}\left\| Y_{t+1}-Y_t \right\|_1 \tag{8}$$

式中，Y_t表示第t帧的平均亮度值，$\left\| Y_{t+1}-Y_t \right\|_1$表示相邻两帧之间的亮度差异的绝对值之和。



帧t　帧$t+1$

帧$t+2$　帧$t+3$

图7　时序抖动

4　总结

本文针对低照度视频成像中噪声强度高、统计特性复杂以及多帧对齐困难等关键问题，提出一种面向真实传感器噪声的视频降噪方法。首先，从物理成像机理出发，构建了噪声模型引导的学习框架，设计了噪声感知编码器，以提升对传感器相关噪声的建模能力与鲁棒性。其次，引入基于变换器的时空特征融合模块，避免了传统显式运动估计在高噪声和复杂运动场景下的不稳定问题，在存在快速运动和遮挡的情况下仍能有效保持时间一致性。

最后，提出可学习性增强策略，通过散粒噪声增强与暗影校正联合建模真实弱光条件下的噪声统计特性，显著扩展了训练数据的多样性，提升了模型在不同设备、不同曝光条件下的泛化性能。综合实验结果表明，所提方法在低照度场景下相比传统方法峰值信噪比提升12.9%，结构相似性提升5.8%，亮度时序方差提升11%，对真实场景下的视频降噪与低照度增强具有一定的参考价值。

参考文献：

[1] Foi A, Trimeche M, Katkovnik V, et al. Practical poissonian-Gaussian noise modeling and fitting for single-image raw-data [J]. IEEE Transactions on Image Processing, 2008, 17(10): 1737-1754.

[2] 刘业琦，刘戈扬，祝晓笑. 一种适用于昼夜兼容CMOS图像传感器的动态范围拓展方法[J]. 半导体光电，2022，43(4)：760-764.
Liu Yeqi, Liu Geyang, Zhu Xiaoxiao. A dynamic range expansion method for day night compatible CMOS image sensor [J]. Semiconductor Optoelectronics, 2022, 43(4): 760-764.

[3] 张为森，马艳华，曲　杨，等. 一种应用于CMOS图像传感器的列级高精度ADC设计[J]. 半导体光电，2022，43(5)：867-872.
Zhang Weisen, Ma Yanhua, Qu Yang, et al. A column-level high resolution ADC design for CMOS image sensor [J]. Semiconductor Optoelectronics, 2022, 43(5): 867-872.

[4] Zhang Y, Zhu Y, Nichols E, et al. A Poisson-Gaussian denoising dataset with real fluorescence microscopy images [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 11702-11710.

[5] Liu C, Freeman W T, Szeliski R, et al. Noise estimation from a single image[C]//2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR′06), 2006: 901-908.

[6] Wu P, Zhou X, Pang G, et al. Weakly supervised video anomaly detection and localization with spatio-temporal prompts [C]//Proceedings of the 32nd ACM International Conference on Multimedia, 2024: 9301-9310.

[7] Dai T, Liu Y, Wang Y, et al. Video demoireing with relation-based temporal consistency [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 2114 - 2123.

[8] Zhang Y, Qin H, Wang X, et al. Rethinking noise synthesis and modeling in raw denoising [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2022: 4573-4581.

[9] Brooks T, Mildenhall B, Xue T, et al. Unprocessing images for learned raw denoising [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 11028-11037.

[10] Feng H, Wang L, Wang Y, et al. Learnability enhancement for low-light raw denoising: where paired real data meets noise modeling [C]//Proceedings of the 30th ACM International Conference on Multimedia, 2022: 1436-1444.

[11] Buades A, Coll B, Morel J M. A non-local algorithm for image denoising [C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR′05), 2005: 60-65.

[12] Dabov K, Foi A, Egiazarian K. Video denoising by sparse 3D transform-domain collaborative filtering [C]//2007 15th European Signal Processing Conference, 2015: 145-149.

[13] Dabov K, Foi A, Katkovnik V, et al. Image denoising by sparse 3-D transform-domain collaborative filtering[J]. IEEE Transactions on Image Processing, 2007, 16(8): 2080-2095.

[14] Gu S, Zhang L, Zuo W, et al. Weighted nuclear norm minimization with application to image denoising [C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014: 2862-2869.

[15] Zhang K, Zuo W, Chen Y, et al. Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising [J]. IEEE Transactions on Image Processing, 2017, 26(7): 3142-3155.

[16] Lefkimmiatis S. Non-local color image denoising with convolutional neural networks [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 5882-5891.

[17] Chen Y, Pock T. Trainable nonlinear reaction diffusion: a flexible framework for fast and effective image restoration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1256-1272.

[18] Liu H, Lu M, Ma Z, et al. Neural video coding using multiscale motion compensation and spatiotemporal context model [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(8): 3182-3196.

[19] Tian C, Xu Y, Fei L, et al. Deep learning for image denoising: a survey [C]//Genetic and Evolutionary Computing, 2019: 563-572.

[20] Qu L, Liu Z, Zhou S, et al. BurstDeflicker: a benchmark dataset for flicker removal in dynamic scenes [EB/OL]. 2025: arXiv: 2510.09996.

[21] Claus M, van Gemert J. ViDeNN: deep blind video denoising [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2019: 1843-1852.

[22] Tassano M, Delon J, Veit T. DVDNET: a fast network for deep video denoising [C]//2019 IEEE International Conference on Image Processing (ICIP), 2019: 1805-1809.

[23] Tassano M, Delon J, Veit T. FastDVDnet: towards real-time deep video denoising without flow estimation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern

Recognition（CVPR），2020：1351-1360.

[24] Chan K C K，Zhou S，Xu X，et al. BasicVSR++：improving video super-resolution with enhanced propagation and alignment [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition（CVPR），2022：5962-5971.

[25] Cao J，Fan Y，Gool L V，et al. Recurrent video restoration transformer with guided deformable attention [C]//Advances in Neural Information Processing Systems 35. Neural Information Processing Systems Foundation，2022：378-393.

[26] Perez E，Strub F，De Vries H，et al. FiLM：visual reasoning with a general conditioning layer [C]// Proceedings of the AAAI Conference on Artificial Intelligence，2018：3942-3951.

[27] 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），2022：5718-5729.

[28] 张维刚，谢志华，车　驰，等. 基于CIS的星敏感成像组件设计与实现[J]. 半导体光电，2025，46(2)：360-366.
Zhang Weigang，Xie Zhihua，Che Chi，et al. Design of star-sensor photoelectric imaging module [J]. Semiconductor Optoelectronics，2025，46(2)：360-366.

作者简介：

付政权(1993—)，男，硕士研究生，主要从事图像处理算法、嵌入式开发等方面的研究。