

面向复杂退化场景的视频逆色调映射算法

张文友, 段盼君

合肥工业大学计算机与信息学院, 安徽 合肥

收稿日期: 2026年4月23日; 录用日期: 2026年5月22日; 发布日期: 2026年5月29日

摘要

针对现有视频逆色调映射方法对降质过程先验依赖较强且泛化能力不足的问题, 文章提出一种降质类型自适应的视频逆色调映射方法, 以提升复杂退化场景下的高动态范围重建性能。该方法首先构建包含多种色调映射算子的训练数据集, 并利用分类器对输入视频的降质特性进行预测, 通过嵌入编码将类别信息引入映射过程。在此基础上, 设计全局色彩变换模块实现初步动态范围扩展, 引入曝光引导的空间注意力机制对过曝与欠曝区域进行细节恢复, 同时采用时空特征协同对齐策略融合多帧上下文信息。实验结果表明, 在所构建的包含11类退化形式的视频数据集上, 所提方法在峰值信噪比、结构相似性及色度差异等指标上均优于现有主流方法。

关键词

逆色调映射, 动态范围, 递归神经网络, 自适应学习

Video Inverse Tone Mapping for Complex Degradation Scenarios

Wenyou Zhang, Panjun Duan

School of Computer Science and Information Engineering, Hefei University of Technology, Hefei Anhui

Received: April 23, 2026; accepted: May 22, 2026; published: May 29, 2026

Abstract

Existing inverse tone mapping methods heavily rely on prior knowledge of degradation processes, which limits their generalization across diverse scenarios. To address this issue, this paper proposes a degradation-type adaptive video inverse tone mapping network to improve performance under complex degradation conditions. Specifically, the proposed method constructs training data covering multiple tone mapping operators, employs a classifier to identify degradation characteristics of input videos, and embeds the predicted category information into the

mapping process via an embedding encoding mechanism. Within the mapping pipeline, a global color transformation module is designed to perform initial dynamic range expansion, while an exposure-guided spatial attention mechanism is introduced to restore over-exposed and under-exposed regions. In addition, a spatio-temporal feature collaborative alignment strategy is adopted to aggregate multi-frame contextual information. Experimental results demonstrate that, on the proposed video dataset containing 11 degradation types, the proposed method outperforms existing state-of-the-art approaches in terms of peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and chrominance difference.

Keywords

Inverse Tone Mapping, Dynamic Range, Recurrent Neural Network, Adaptive Learning

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着高动态范围(High Dynamic Range, HDR)显示设备的快速普及, HDR 视频凭借其更宽的亮度范围和更广的色域覆盖,能够显著提升视觉内容的真实感与沉浸感。然而,当前主流视频资源仍主要以标准动态范围(Standard Dynamic Range, SDR)格式进行存储与传输,如何将大规模存量 SDR 内容高效转换为适配现代显示设备的 HDR 格式,已成为超高清视频处理领域的重要研究问题。

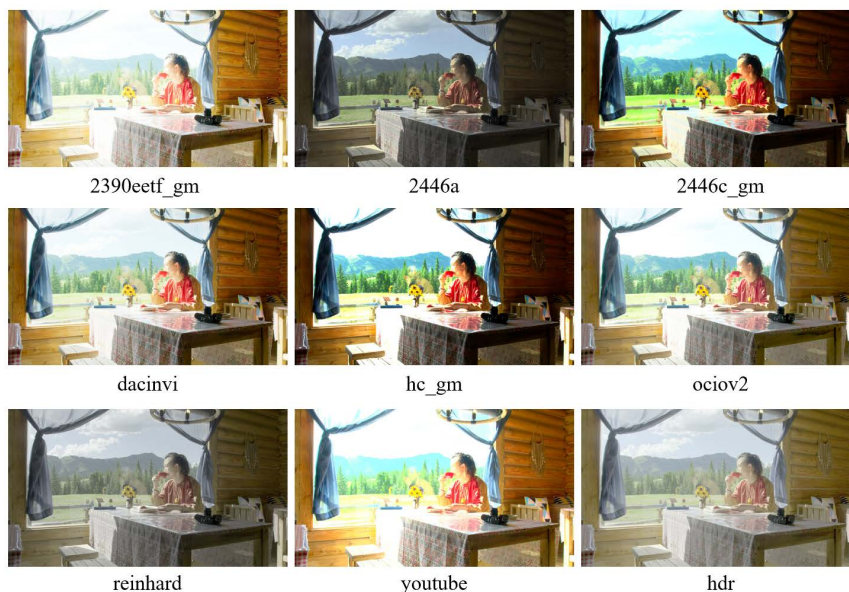


Figure 1. Visual examples of various degradation types in the HDRTV4K dataset [1]

图 1. HDRTV4K 数据集中多种退化类型视觉样例[1]

逆色调映射(Inverse Tone Mapping, ITM)技术[2]旨在将 SDR 内容重建为 HDR 表示,其核心目标在于恢复高动态范围信息并提升视觉质量。尽管近年来相关研究取得了一定进展,但现有方法大多针对静态图像设计,缺乏对视频序列中时序一致性的有效建模;同时,多数方法依赖于预定义或单一的退化假设

(仅使用一种色调映射算子(Tone Mapping Operator, TMO)进行降质处理), 难以应对真实场景中复杂多变的输入分布。此外, 在实际应用中, SDR 视频通常经历多种处理操作的叠加与耦合, 其生成过程复杂且不可控, 导致退化形式呈现出显著的多样性与不确定性。如图 1 所示, 不同来源的视频在亮度压缩方式、对比度分布以及色彩风格等方面存在明显差异, 这种由多种色调映射策略及后处理流程共同作用形成的复杂退化模式, 使得当前基于单一退化建模的学习方法难以获得良好的泛化能力。因此, 如何在复杂退化场景下实现鲁棒的视频逆色调映射, 成为当前亟待解决的关键问题。

基于此, 本文提出一种面向复杂退化场景的视频逆色调映射算法。首先, 构建了一个包含多种典型色调映射算子的多退化多帧视频数据集, 以覆盖多样化的退化模式; 其次, 引入退化感知机制通过预训练分类网络对输入视频的退化类型进行建模, 并结合位置编码将相关信息嵌入重建过程, 从而实现对不同退化形式的自适应处理。最后, 进一步地设计了全局色彩变换模块以实现跨域映射, 并结合曝光自适应的空间注意力机制, 有效恢复过曝与欠曝区域的细节信息。同时引入时空特征对齐模块, 实现帧间信息融合与隐状态传递, 从而提升视频重建结果的时序一致性与稳定性。实验结果表明, 所提出方法在色彩保真度、细节恢复能力以及复杂退化场景下的泛化性能方面均优于现有主流方法, 验证了其在实际应用中的有效性与鲁棒性。

2. 相关工作

2.1. 基于单帧的视频逆色调映射算法

早期基于深度学习的逆色调映射研究主要采用单帧单退化的视频数据训练网络模型。Chen 等人[3]构建了 HDRTV1K 数据集, 从 YouTube 平台采集 HDR 与 SDR 成对数据, 其中 SDR 版本经由平台内部色调映射与色彩压缩流程生成。该工作提出三阶段网络架构, 依次执行全局色彩映射、局部细节增强与高光区域重建。后续研究中, Chen 等人[4]对网络结构进行轻量化重构, 在保留成像先验约束的同时降低计算开销, 实现了效率与质量的双重提升。Xu 等人[5] [6]引入频率感知调制机制, 通过双频域建模捕获不同频率成分的互补信息, 有效提升了重建保真度。He 等人[7]提出层次化动态上下文映射网络, 逐层挖掘多尺度上下文依赖关系, 渐进式优化 SDR 到 HDR 的转换质量。Shao 等人[8]设计混合条件框架, 融合全局与局部约束以改善亮度一致性与细节真实感。Huang 等人[9]则在 IC_cp 感知均匀色彩空间中执行重建, 利用该空间对 HDR 域的适应性实现更精准的亮度与色度恢复。上述方法均假设退化类型比较单一且未利用时序信息, 在面对现实世界中风格多样的 SDR 视频内容时重建质量会显著下降。

2.2. 基于多帧的视频逆色调映射算法

为利用时序信息提升重建稳定性, 近期研究将逆色调映射扩展至多帧场景。Cao 等人[10]提出基于核预测的多帧交互网络, 显式建模帧间时序依赖。Yan 等人[11]将全局色彩映射与多帧特征交互相结合, 增强时序一致性与视觉保真度。Zhang 等人[12]设计小波域掩码图像建模框架, 在亮度与色度通道上充分挖掘帧间相关性, 实现色彩一致的 HDR 视频重建。然而, 这类方法虽然利用了时序信息, 但是和上类方法一样采用单一退化类型数据训练, 在面对风格多样的真实 SDR 视频时易出现色彩偏移。

2.3. 面向真实世界的视频逆色调映射算法

针对实际应用中 SDR 内容风格多样导致的泛化瓶颈, 近期研究开始关注真实场景下的重建问题。Guo 等人[1]构建包含多种 TMO 降质的 HDRTV4K 数据集, 并引入亮度分段网络优化高光区域重建。Xu 等人[13] [14]进一步融合真实 HDR 先验并解耦降质相关表征, 有效提升了异构 SDR 输入的处理能力。值得注意的是, 现有真实场景方法均只利用了视频单帧信息, 未能充分利用视频内容的时序连贯性。

综上所述, 现有方法或专注于单帧重建而忽略帧间关联, 或利用多帧信息但假设降质固定, 尚缺乏同时适应复杂降质类型与保持时序一致性的统一框架。本文提出降质自适应视频逆色调映射网络, 通过预训练分类器辨识输入降质特性并嵌入重建流程, 实现复杂退化场景泛化与帧间稳定的协同优化。

3. 算法介绍

针对真实 SDR 视频中退化形式复杂多样且不可控的问题, 本文提出一种面向复杂退化场景的视频逆色调映射方法——降质自适应视频逆色调映射网络(Degradation-Adaptive Video Inverse Tone Mapping Network, DAVITMNet)。该方法通过引入退化感知机制与时空协同建模策略, 实现复杂退化条件下的高质量 HDR 视频重建。

3.1. 方法概述

如图 2 所示, 从整体结构上看, 本文方法可划分为两个核心组成部分: 退化预测与 HDR 内容重建。

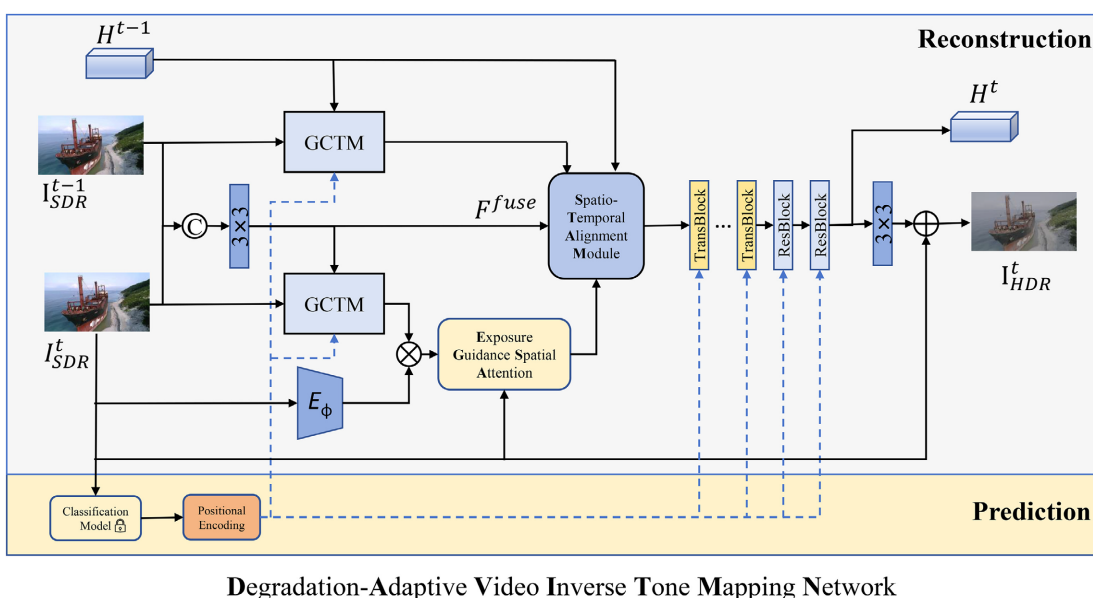


Figure 2. Architecture of the degradation-adaptive video inverse tone mapping network (DAVITMNet)
图 2. 降质自适应视频逆色调映射网络(DAVITMNet)结构图

首先, 在退化预测阶段本文通过分类网络 MobileNetv2 [15]从输入帧中提取全局统计特征, 包括亮度分布、对比度压缩及色彩风格等这些与退化相关的信息。在训练时将退化类型作为标签训练该分类网络, 推理时得到预测的退化类别标签并将其作为类别嵌入到整个 HDR 内容重建部分, 嵌入过程直接使用相加操作。该机制使本方法能够显式区分不同退化模式, 从而在统一框架下自适应调整映射策略, 有效缓解多种异构映射关系共存时的优化冲突问题。

其次, 在 HDR 内容重建阶段, 本文采用由粗到细的渐进式建模策略。首先进行全局动态范围扩展与色彩映射, 其次对局部极端曝光区域进行细节恢复, 最后结合时空对齐与特征融合机制, 充分利用帧间冗余信息, 实现高质量且时序一致的 HDR 视频重建。

整个方法采用循环架构处理视频序列, 对于第 t 帧 SDR 输入 I_{SDR}^t , 其对应的 HDR 帧 I_{HDR}^t 重建过程表示为:

$$I_{HDR}^t, H^t = F_{DAVITM} \left(I_{SDR}^{t-1}, I_{SDR}^t, H^{t-1}, c, \theta \right) \quad (1)$$

其中, I'_{HDR} 为重建的 HDR 帧, H^t 为当前帧隐状态特征, H^{t-1} 为前一帧传递的时序隐藏信息, θ 为网络可学习参数。符号 c 表示降质类别嵌入向量, 由预训练分类网络 F_D 预测并经位置编码 $\text{PE}(\cdot)$ 映射得到, 其中位置编码模块由两层 MLP 构成:

$$c = \text{PE}\left(F_D\left(I'_{\text{SDR}}\right)\right) \quad (2)$$

HDR 内容重建阶段整体包含 4 个主要功能模块: 全局色彩变换模块(Global Color Transformation Module, GCTM)执行初步动态范围扩展, 将 SDR 域映射至 HDR 域; 曝光指导空间注意力模块(Exposure-Guided Spatial Attention, EGSA)针对过曝与欠曝区域执行自适应细节修复; 时空特征对齐模块(Spatio-Temporal Alignment Module, STAM)聚合相邻帧信息, 补偿运动位移并融合时序上下文信息; 最后局部内容增强模块由级联 Transformer 块与残差块组成, 输出最终 HDR 结果。其中位置编码机制贯穿网络, 将降质类别信息嵌入各处理阶段, 引导网络针对不同退化特性执行自适应恢复。

3.2. 各模块详细结构

3.2.1. 全局色彩变换模块(GCTM)

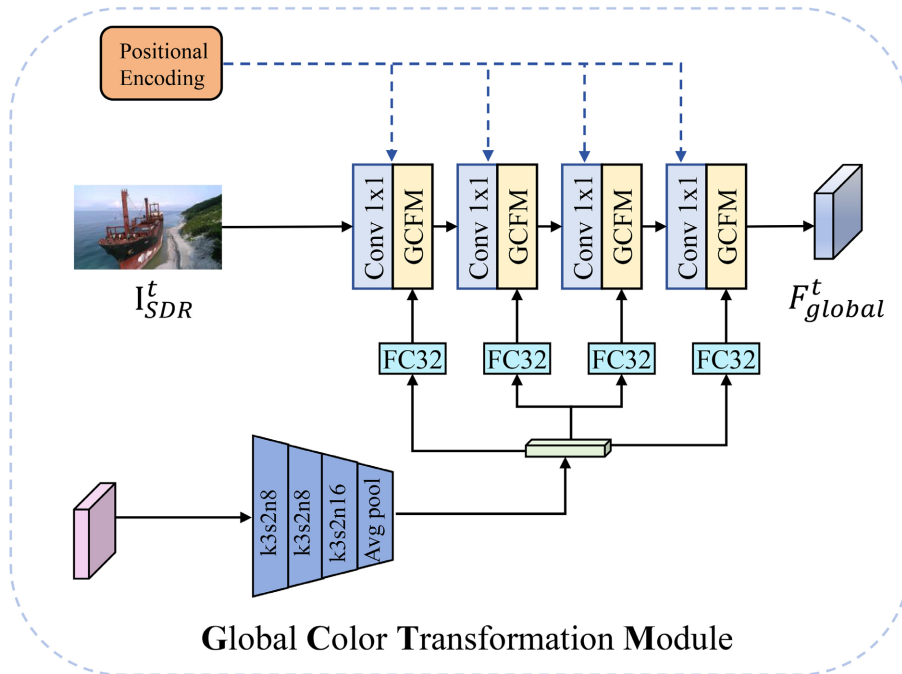


Figure 3. Structure of the global color transformation module (GCTM)
图 3. 全局色彩变换模块(GCTM)结构图

在 SDR 视频制作流程中, 很多操作都属于像素独立的全局操作, 例如色调映射与利用光电转换函数(OETF) [16]进行非线性化处理等, 其变换规律与像素空间位置无关, 仅依赖输入亮度值或全局统计量。因此本文设计全局色彩变换模块(GCTM)完成 SDR 到 HDR 域的初步迁移, 如图 3 所示, 该模块以 1×1 卷积层与激活函数构建主干, 执行像素独立的全局变换。 1×1 卷积在不扩展感受野的前提下实现通道维度变换与跨通道信息交互, 适合模拟色调映射与 OETF 的点点非线性特性。

考虑到部分色调映射算子(TMO) (如 Hable [17]、Reinhard [18]等)依赖整幅图像的统计特性(如均值、方差或直方图分布), 单纯像素独立变换不足以模拟其全局自适应行为。因此, 引入条件分支提取全局先

验, 通过 3 层下采样卷积(核尺寸 3×3 , 步长 2 后结合 ReLU 激活函数)逐步扩大感受野, 经全局平均池化压缩空间维度, 获取尺寸为 $N \times C \times 1 \times 1$ 的条件特征。该特征经全连接层生成调制参数, 通过全局通道特征调制(Global Channel Feature Modulation, GCFM)注入主干网络:

$$\text{GCFM}(x_i) = \alpha_i \otimes x_i + \beta_i \quad (3)$$

其中, x_i 为第 i 层 1×1 卷积输出的中间特征, α_i 与 β_i 分别为条件分支生成的缩放与偏移参数, \otimes 表示逐元素乘法。GCFM 机制使网络能够根据全局亮度分布自适应调整各通道响应, 增强对多样化 TMO 的拟合能力。

对于视频序列, 对当前帧和前一帧分别使用 GCTM 进行全局映射, 前一帧的 GCTM 计算流程: 利用其前一帧原始 SDR 视频帧 I_{SDR}^{t-1} 与历史隐状态 H^{t-1} 为输入, 生成全局特征 F_{global}^{t-1} , 隐状态传递确保时序一致性; 对于当前帧, 利用其当前帧 I_{SDR}^t 及两帧融合信息 F_{fuse} 为输入, 输出当前帧的全局特征 F_{global}^t 。同时将位置编码向量注入当前帧和前一帧的 GCTM, 使降质感知贯穿时序处理全过程。

3.2.2. 曝光指导空间注意力模块(EGSA)

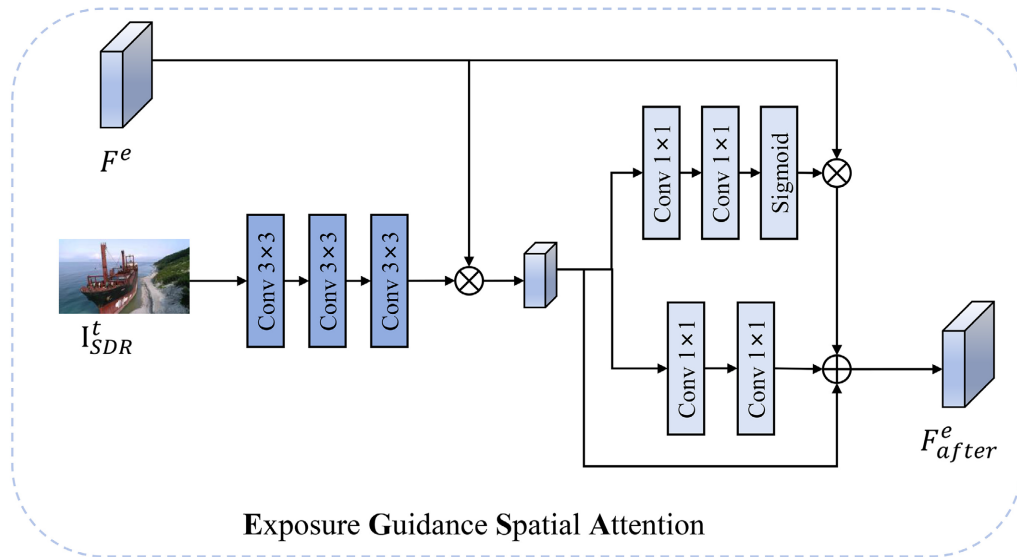


Figure 4. Structure of the exposure-guided spatial attention module (EGSA)
图 4. 曝光指导空间注意力模块(EGSA)结构图

经 GCTM 全局变换后, 图像已初步映射至 HDR 域, 但过曝与欠曝区域的细节信息仍待恢复。色调映射算子通常压缩高亮区域以避免裁切, 导致 SDR 图像中原本丰富的纹理信息丢失; 暗部区域则因量化精度不足而呈现噪声或色带。这些局部退化具有空间变异性, 需根据具体曝光程度执行自适应增强。因此本文提出曝光自适应空间注意力(EGSA), 其结构如图 4 所示, 核心思想为: 首先估计各像素的曝光程度, 生成曝光引导图; 随后根据引导图对全局特征执行空间自适应调制, 增强低曝光区域, 抑制过曝光区域。

具体实现分为 3 个阶段。第一阶段为曝光估计: 编码器 E_ϕ 由 4 层 3×3 卷积与激活函数组成, 逐步提取多尺度特征并预测软掩膜, 输出尺寸为 $N \times 1 \times H \times W$ 的曝光图, 数值越高表示该区域越可能为过曝或欠曝的极端曝光区域。该曝光图与 GCTM 输出的全局特征相乘, 得到曝光加权特征:

$$F_e = E_\phi(I_{\text{SDR}}^t) * F_{\text{global}}^t \quad (4)$$

其中, $*$ 表示逐元素乘法。第二阶段为特征融合: 当前帧 I_{SDR}^t 经 3 层 3×3 卷积提取特征后与 F_e 进行通道

调制, 生成中间融合特征 F_{mid}^e , 该过程融合原始输入的局部细节与全局变换的 HDR 结构。第三阶段为空间自适应增强: F_{mid}^e 并行通过 2 层 1×1 卷积生成空间调制系数, 一支经 Sigmoid 激活生成注意力权重, 另一支保持线性响应, 经残差相加融合得到最终增强特征 F_{after}^e :

$$F_{after}^e = F_{mid}^e + F_e \otimes f_{cl}(f_{cl}(F_{mid}^e)) + f_{cl}(f_{cl}(F_{mid}^e)) \quad (5)$$

其中, $f_{cl}(\cdot)$ 表示 1×1 卷积层, 第 2 个卷积后无激活函数以保留线性能力。该设计使网络能够根据局部亮度自适应调整注意力权重, 有效恢复极端曝光区域的纹理细节, 同时避免对正常曝光区域的过度处理。

3.2.3. 时空特征对齐模块(STAM)

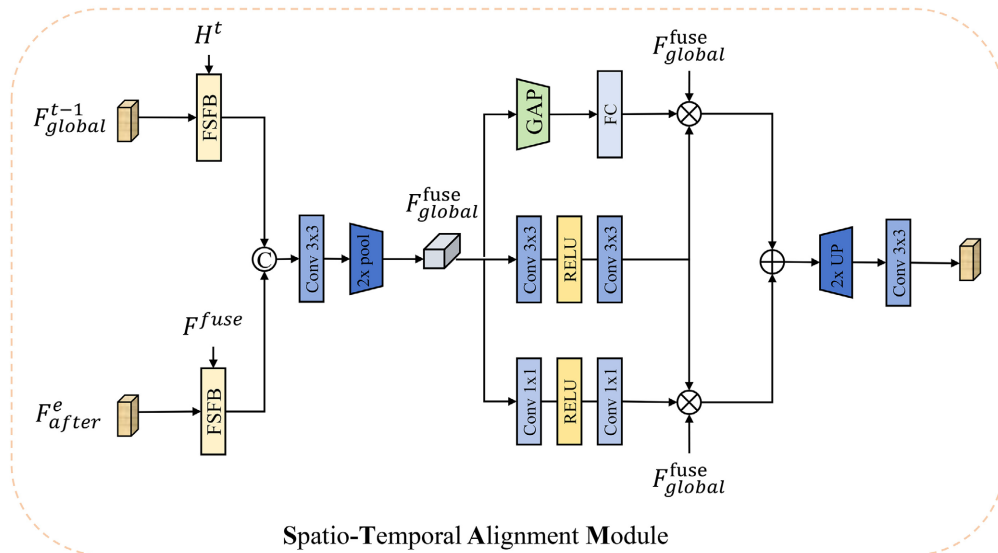


Figure 5. Structure of the spatio-temporal alignment module (STAM)
图 5. 时空特征对齐模块(STAM)结构图

视频序列具有显著的时序冗余性与连贯性, 相邻帧在场景内容、光照条件及色彩分布上高度相关。充分利用时序信息不仅能够提升重建稳定性, 还可借助帧间互补信息恢复单帧难以重建的细节。本文设计时空特征对齐模块(STAM), 对相邻帧特征执行显式对齐与深度融合。

如图 5 所示, 具体计算过程如下, 本文提出的对齐模块采用隐式对齐的策略, 该模块首先采用频率分离融合策略提取帧内低频信息, 通过频率分离融合块(Frequency Separation Fusion Block, FSFB) [19]将特征分解为低频基础层与高频细节层, 低频层包含场景的主要结构与色彩分布, 是帧间对齐的关键。随后在下采样空间(分辨率降低 1/2)执行时序对齐。时间注意力分解为静态与动态两个分支: 静态注意力由 2 个 3×3 卷积与激活函数组成, 捕获相邻帧的固定偏移关系; 动态注意力进一步细分为通道动态注意力与空间动态注意力。通道动态注意力经全局平均池化与线性层生成通道权重, 自适应选择对当前重建有益的通道; 空间动态注意力通过 2 个 1×1 卷积逐步压缩通道维度至 1 以生成空间权重图, 以此定位帧间对应区域。静动态注意力协同作用, 实现帧间特征的有效聚合与运动补偿, 对齐后的融合特征经 2 倍上采样接入 3×3 卷积恢复到原分辨率。

3.2.4. 局部内容增强模块

最后将经过 STAM 对齐后的特征送入局部内容重建模块增强局部细节, 如图 2 所示, 局部内容增强模块由 10 个 TransBlock (Transformer Block)块、2 个残差块(Residual Block, ResBlock)及 3×3 卷积级联构

成, 本文的 TransBlock 采用由多 DConv 头转置注意力(Multi-DConv Head Transposed Attention, MDTA) [20]结构建模长程依赖关系, 残差块则强化局部细节。远程跳跃连接将原始输入直接加到输出端, 保留输入的整体结构特性。退化类别信息经位置编码后的位置编码向量在重建阶段被再次注入, 确保降质类别信息全程参与重建决策, 最终输出 10 bit HDR 视频帧 I'_{HDR} 与更新后的隐状态 H' 。

4. 数据集制作

现有视频逆色调映射数据集 HDRTV1K [3]与 HDRTV4K [1]主要以帧级 SDR-HDR 配对数据的形式构建, 缺乏连续帧序列, 难以直接用于多帧模型的训练; 同时, 其包含的退化类型相对有限, 难以覆盖真实场景中复杂多样的 SDR 生成过程。为此, 本文在上述数据集基础上进一步扩展构建多帧视频数据集收集 4K 分辨率视频序列, 内容涵盖电视演示、山川湖泊、城市建筑、街道夜景及自然风光等多种典型场景, 以保证数据分布的多样性, 样例如图 6 所示。

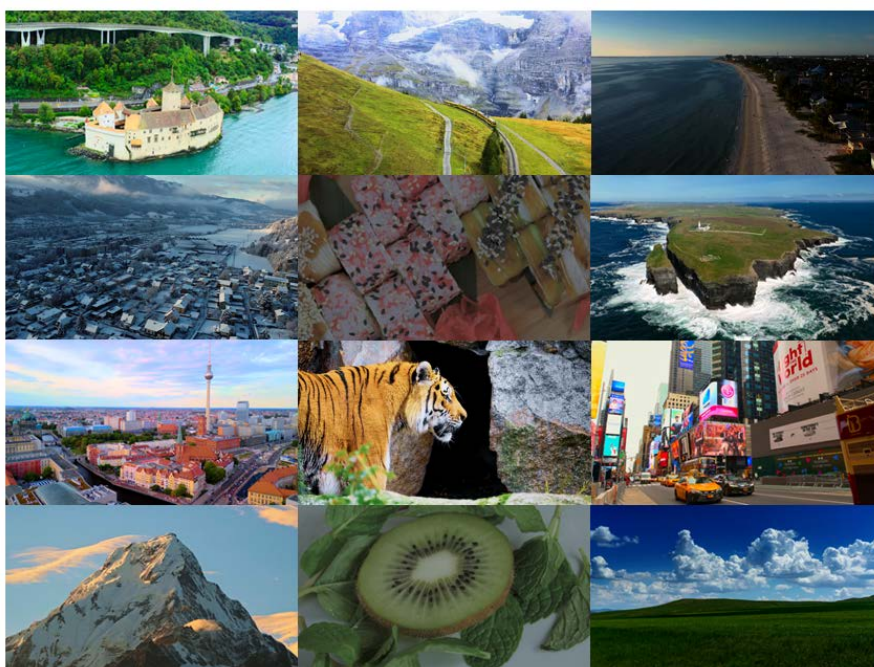


Figure 6. Visual examples of the proposed dataset

图 6. 所提出数据集的视觉样例

在数据预处理阶段, 首先对视频进行帧级采样。为避免训练样本间的内容冗余并增强数据多样性, 本文以固定时间间隔进行稀疏采样, 并从每个采样位置提取连续帧序列作为训练样本。具体地, 每隔一定帧数选取一段包含 10 帧的连续片段, 最终构建得到 468 组训练序列。

考虑到 4K 分辨率带来的计算与显存开销, 本文将原始分辨率由 3840×2160 统一缩放至 1920×1080 , 在保证视觉信息完整性的同时提高训练效率。为模拟真实世界中多样化的 SDR 生成过程, 本文采用 11 种典型色调映射算子(2446a [21]、2446c [22]、BestExposure [22]、DragoTMO [23]、FerwerdaTMO [24]、Hable [17]、LischinskiTMO [25]、Reinhard12 [26]、Reinhard [18]、Shan12 [27]、WardHistAdjTMO [28])对 HDR 数据进行退化处理, 并尽量保持各类退化类型在训练数据中的分布均衡, 从而构建对应的 SDR-HDR 配对训练集。

测试集的构建遵循与训练集一致的数据处理流程, 并额外选取 104 组视频序列作为测试样本, 每组

包含 5 帧连续图像, 用于评估模型在多帧输入条件下的重建性能。测试数据在内容上与训练集完全不重叠, 确保评估结果的客观性与泛化能力。

5. 实验

5.1. 模型训练

本文方法涉及两个阶段的训练: 降质分类器预训练与重建网络端到端训练。

分类器训练阶段: 从 468 组训练序列中抽取首帧, 经 11 种 TMO 降质后得到 5148 张训练图像(468 × 11), 测试集同理得到 1144 张图像(104 × 11)。输入图像被随机裁剪为 480 × 480 像素块, 经水平翻转、垂直翻转及 90° 旋转实现数据增强。优化器采用 Adam [29], 初始学习率设为 1×10^{-3} , 批量大小为 16, 共训练 400 个轮次(epoch)。

重建网络训练阶段: 加载上述分类器的预训练权重并固定参数, 直接使用其预测结果生成降质类别嵌入。输入 SDR 图像与对应 HDR 真值(Ground Truth, GT)随机裁剪为 256 × 256 像素块。优化器采用 Adam, 初始学习率为 4×10^{-4} , 利用余弦退火策略调整学习率, 在 50000、100000、150000 次迭代时重启学习率, 批量大小为 4, 共训练 300,000 次迭代。所有实验基于 PyTorch 框架在 NVIDIA TITAN V 图形处理器(Graphics Processing Unit, GPU)上完成。

为兼顾像素级精度与结构信息一致性, 本文损失函数由 L_1 损失与结构相似性(Structural Similarity Index Measure, SSIM)损失联合构成。逐像素 L_1 损失定义为:

$$L_1 = \|I_{\text{HDR}} - I_{\text{GT}}\|_1 \quad (6)$$

SSIM 损失关注图像结构信息, 其取值范围为[-1, 1], 越接近 1 表示两幅图像越相似。训练时采用 $1 - \text{SSIM}$ 作为损失项:

$$L_{\text{ssim}} = 1 - \text{SSIM}(I_{\text{HDR}}, I_{\text{GT}}) \quad (7)$$

对于包含 N_T 帧的视频序列, 总损失为:

$$L_{\text{total}} = \frac{1}{N_T - 1} \sum_{t=2}^{N_T} (L_1^t + L_{\text{ssim}}^t) \quad (8)$$

5.2. 实验结果

为验证本文方法的有效性, 选取多种当前先进的逆色调映射方法作为对比, 包括 HDCFM [7]、HDRTVNet++ [4]、LCATNet [9]、HDRTVDM [1]和 DFT [6]。为保证公平性, 大多数对比方法均在本文构建的数据集上按照其原始设置重新训练与测试。定量评估采用峰值信噪比(Peak Signal-to-Noise Ratio, PSNR)、结构相似性(Structural Similarity Index, SSIM)、谱残差相似性(Spectral Residual Similarity, SR-SIM) [30]、色差指标(ΔE_{TPP}) [31]以及 HDR 视觉差异预测(HDR Visual Difference Predictor 3, HDR-VDP3) [32]作为评价指标。其中, PSNR 用于衡量像素级重建精度, SSIM 与 SR-SIM 从结构与感知层面评估图像质量, ΔE_{TPP} 用于刻画色彩还原误差(值越小越好), HDR-VDP3 则基于人类视觉系统预测感知差异(值越高越好)。上述指标从重建精度、结构保持、感知质量及色彩保真度等多个维度对方法性能进行综合评价。

5.2.1. 定量结果比较

表 1 展示了各方法在本文所提出数据集上的定量对比结果。其中, “w/o train” 表示直接使用该方法公开的预训练权重进行测试。可以观察到, 在未进行再训练的情况下, 各方法整体性能均出现明显下降, 表明基于单一退化类型训练的模型在跨数据集场景下泛化能力有限, 难以有效应对真实环境中复杂多样的退

化模式。在此基础上, 从整体性能来看, 本文方法在 PSNR、SSIM、SR-SIM、 ΔE_{ITP} 以及 HDR-VDP3 等指标上均取得最优或次优表现, 表明其在重建精度、结构保持、感知质量及色彩还原等方面具有显著优势。

Table 1. Quantitative comparison results of different methods on the proposed dataset
表 1. 在所提出数据集上不同方法的定量比较结果

Method	PSNR \uparrow	SSIM \uparrow	SR-SIM \uparrow	ΔE_{ITP} \downarrow	HDR-VDP3 \uparrow	Params (K)	Time (s)
HDCFM	25.35	0.9397	0.9348	51.30	5.924	101	2.313
HDRTVNet++	25.19	0.9251	0.9915	54.04	8.120	591	0.018
HDRTVNet++ w/o train	14.94	0.7448	0.9334	140.25	5.780	591	0.018
LCATNet w/o train	19.88	0.8586	0.9890	82.94	7.611	759	0.869
LCATNet	27.91	<u>0.9578</u>	0.9949	36.25	<u>8.495</u>	759	0.869
HDRTVDM w/o train	19.12	0.8633	0.9876	101.01	7.549	402	0.026
HDRTVDM	<u>28.15</u>	0.9562	0.9940	34.74	8.338	402	0.026
DFT	26.43	0.9465	0.9924	42.50	7.771	135	0.044
Our	28.70	0.9607	<u>0.9948</u>	<u>35.64</u>	8.672	227	0.041

具体而言, 在像素级重建与结构保持方面, 本文方法在 PSNR 与 SSIM 指标上均表现领先, 说明其能够更准确地恢复 HDR 图像的亮度信息并保持良好的结构一致性。在感知质量相关指标上, 本文方法在 SR-SIM 与 HDR-VDP3 上同样具有竞争优势, 体现出更优的视觉感知效果与细节还原能力。在色彩还原方面, ΔE_{ITP} 指标反映了重建结果与参考图像之间的色差。本文方法在该指标上保持较低误差水平, 说明其在进行动态范围扩展的同时, 能够有效抑制色彩偏移, 实现较高的色彩保真度。此外, 从模型复杂度与推理效率来看, 本文方法在保持较高重建性能的同时, 具备较低参数规模与较快的推理速度, 体现出良好的效率与性能平衡。

5.2.2. 定性结果比较

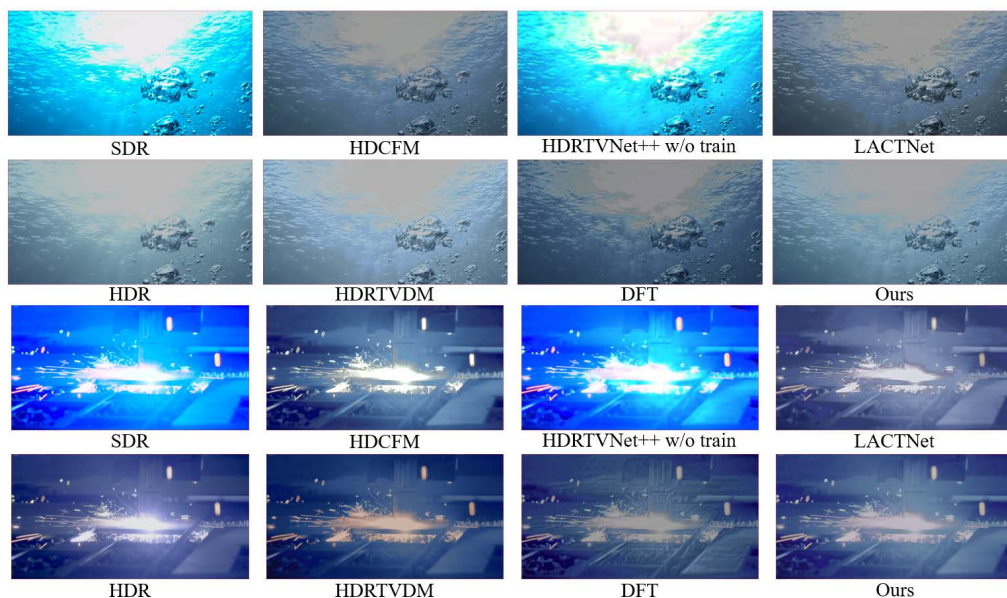


Figure 7. Qualitative comparison results on the proposed dataset
图 7. 在所提出数据集上的定性比较结果

图 7 展示了不同方法在典型高光场景下的 HDR 重建结果, 按照[3]中采用的可视化协议, 这些重建得到的 10-bit HDR 帧以 PNG (16-bit)格式直接呈现, 未进行任何额外后处理。需要说明的是, 由于采用基于 Gamma 的电光传输函数进行解码, 在常规 8-bit SDR 显示设备上显示时, HDR 图像可能会呈现偏暗的现象, 但这不影响与对比方法的比较。从整体视觉效果来看, 现有方法在复杂退化条件下对高光区域的处理普遍存在不足, 例如亮度恢复不充分、细节丢失或色彩偏移等问题。其中, 部分方法在高光区域易出现过度压制或过曝扩散现象, 而未进行再训练的模型(w/o train)由于缺乏对当前数据分布的适配, 重建结果中存在明显的亮度失衡与色彩偏移, 影响整体视觉体验。相比之下, 本文方法能够更加准确地恢复高光区域的亮度层次与细节结构, 同时保持色彩还原的准确性。

5.3. 消融实验

Table 2. Quantitative results of the ablation study

表 2. 消融实验定量结果

	PSNR \uparrow	SSIM \uparrow	SR-SIM \uparrow	$\Delta E_{ITP} \downarrow$	HDR-VDP3 \uparrow
w/o DE	26.62	0.9449	0.9889	41.78	8.230
w GT Labels	27.88	0.9548	0.9962	40.79	8.588
w/o GCTM	<u>27.67</u>	0.9504	0.9945	41.10	8.554
w/o EGSA	27.31	0.9553	<u>0.9954</u>	40.42	<u>8.605</u>
w/o E_ϕ	27.38	0.9545	0.9950	40.07	8.551
w/o STAM	27.09	0.9581	0.9945	42.23	8.667
Our	27.63	<u>0.9557</u>	0.9949	<u>40.10</u>	8.600

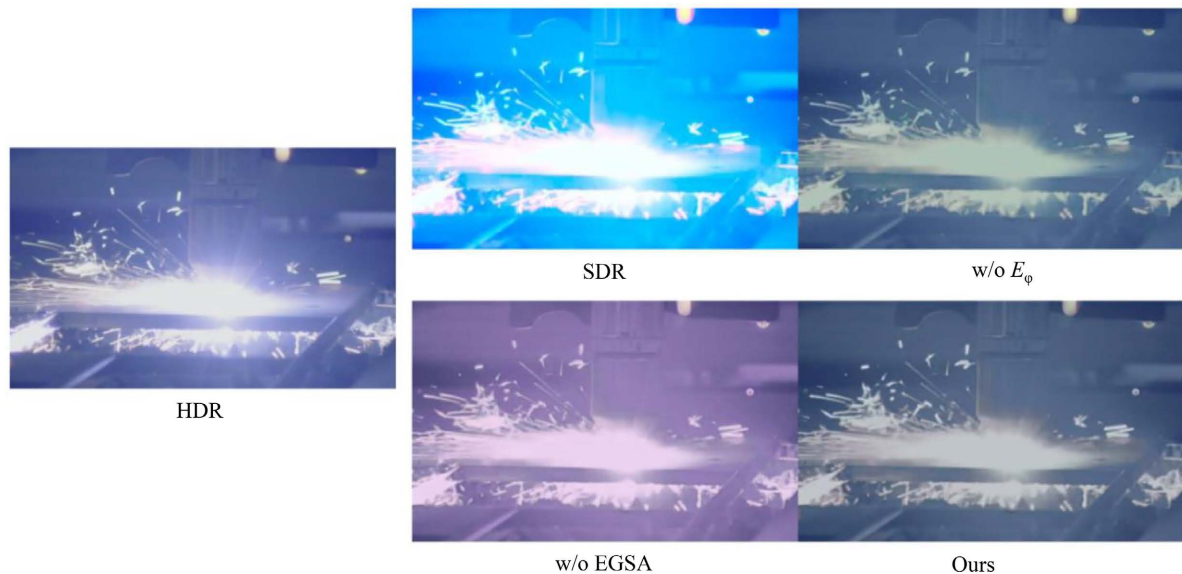


Figure 8. Results of the ablation experiment

图 8. 消融实验结果图

为验证本文提出的各个模块的贡献, 设计 6 组消融实验: w/o DE 代表移除降质嵌入模块, 直接使用可学习嵌入, 即不含任何降质感知模块的网络模型; w GT Labels 代表使用真实类别标签替代预测分布; w/o GCTM 以 2 层 3×3 卷积替换全局色彩变换; w/o EGSA 代表移除曝光指导空间注意力; w/o E_ϕ 代表

移除曝光估计编码器; w/o STAM 代表使用简单拼接替代时序对齐。如表 2 与图 8 所示, 完整网络取得最佳综合性能, 移除任一模块均导致指标下降, 验证了降质感知、全局变换、曝光自适应与时序对齐的必要性。从各模块的消融实验结果可以看出, 尽管本研究所提的方法并未在所有指标上都取得最优性能, 但整体而言, 该方法在各类指标间实现了更均衡的效果。

6. 结论

本文针对复杂退化场景下的逆色调映射问题, 提出了一种降质自适应逆色调映射网络。该方法构建涵盖 11 种典型色调映射算子的训练数据模拟真实世界 SDR 视频风格分布, 通过训练分类器辨识输入视频的降质特性, 并借助位置编码机制将类别信息嵌入重建流程; 同时在网络结构设计上, 本文设计了全局色彩变换模块完成初步动态范围扩展, 引入曝光指导空间注意力机制修复过曝与欠曝区域细节, 以及采用时空特征对齐策略聚合多帧信息确保时序一致性。实验结果表明, 在所提出多退化多帧视频数据集上, 本文方法在峰值信噪比、结构相似性、色度差异及感知质量等指标上均优于现有主流方法, 验证了降质类型感知机制与视频时序建模的有效性, 显著提升了视频逆色调映射的泛化能力。

参考文献

- [1] Guo, C., Fan, L., Xue, Z. and Jiang, X. (2023) Learning a Practical SDR-to-HDRTV Up-Conversion Using New Dataset and Degradation Models. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, 18-22 June 2023, 22231-22241. <https://doi.org/10.1109/cvpr52729.2023.02129>
- [2] Banterle, F., Ledda, P., Debattista, K. and Chalmers, A. (2006) Inverse Tone Mapping. *Proceedings of the 4th International Conference on Computer Graphics and Interactive Techniques in Australasia and Southeast Asia*, Kuala Lumpur, 29 November-2 December 2006, 349-356. <https://doi.org/10.1145/1174429.1174489>
- [3] Chen, X., Zhang, Z., Ren, J.S., Tian, L., Qiao, Y. and Dong, C. (2021) A New Journey from SDRTV to HDRTV. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 4500-4509. <https://doi.org/10.1109/iccv48922.2021.00446>
- [4] Chen, X., Li, Z., Zhang, Z., Ren, J.S., Liu, Y., He, J., et al. (2025) Towards Efficient SDRTV-to-HDRTV by Learning from Image Formation. *IEEE Transactions on Multimedia*, **27**, 8340-8354. <https://doi.org/10.1109/tmm.2025.3604961>
- [5] Xu, G., Hou, Q., Zhang, L. and Cheng, M. (2022) FMNet: Frequency-Aware Modulation Network for SDR-to-HDR Translation. *Proceedings of the 30th ACM International Conference on Multimedia*, Lisboa, 10-14 October 2022, 6425-6435. <https://doi.org/10.1145/3503161.3548016>
- [6] Xu, G., Hou, Q. and Cheng, M. (2024) Dual Frequency Transformer for Efficient SDR-to-HDR Translation. *Machine Intelligence Research*, **21**, 538-548. <https://doi.org/10.1007/s11633-023-1418-8>
- [7] He, G., Xu, K., Xu, L., Wu, C., Sun, M., Wen, X., et al. (2022) SDRTV-to-HDRTV via Hierarchical Dynamic Context Feature Mapping. *Proceedings of the 30th ACM International Conference on Multimedia*, Lisboa, 10-14 October 2022, 2890-2898. <https://doi.org/10.1145/3503161.3548043>
- [8] Shao, T., Zhai, D., Jiang, J. and Liu, X. (2022) Hybrid Conditional Deep Inverse Tone Mapping. *Proceedings of the 30th ACM International Conference on Multimedia*, Lisboa, 10-14 October 2022, 1016-1024. <https://doi.org/10.1145/3503161.3548129>
- [9] Huang, P., Cao, G., Zhou, F. and Qiu, G. (2023) Video Inverse Tone Mapping Network with Luma and Chroma Mapping. *Proceedings of the 31st ACM International Conference on Multimedia*, Ottawa, 29 October-3 November 2023, 1383-1391. <https://doi.org/10.1145/3581783.3612199>
- [10] Cao, G., Zhou, F., Yan, H., Wang, A. and Fan, L. (2022) KPN-MFI: A Kernel Prediction Network with Multi-Frame Interaction for Video Inverse Tone Mapping. *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, Vienna, 23-29 July 2022, 806-812. <https://doi.org/10.24963/ijcai.2022/113>
- [11] Yan, H., Zhang, H., Zhao, M., et al. (2025) Video Inverse Tone Mapping Network with GlobalColor Mapping and Multi-Frame Interaction. *IEEE Transactions on Consumer Electronics*, **71**, 2930-2943.
- [12] Zhang, Y., Ni, Z., Yang, W., et al. (2026) Wavelet-Domain Masked Image Modeling for Color-Consistent HDR Video Reconstruction.
- [13] He, G., Xu, K., Xu, L., Yu, W. and Wu, X. (2025) Beyond Feature Mapping GAP: Integrating Real HDRTV Priors for Superior SDRTV-to-HDRTV Conversion. *Proceedings of the Thirty-Fourth International Joint Conference on Artificial*

- Intelligence*, Montreal, 16-22 August 2025, 1089-1097. <https://doi.org/10.24963/ijcai.2025/122>
- [14] Xu, L., Wang, S., Xu, K., Zhang, L., He, G., Wang, W., *et al.* (2026) RealRep: Generalized SDR-to-HDR Conversion via Attribute-Disentangled Representation Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, **40**, 11305-11313. <https://doi.org/10.1609/aaai.v40i13.38111>
- [15] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L. (2018) MobileNetV2: Inverted Residuals and Linear Bottlenecks. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-22 June 2018, 4510-4520. <https://doi.org/10.1109/cvpr.2018.00474>
- [16] ITU-R (2018) Image Parameter Values for High Dynamic Range Television for Use in Production and International Programme Exchange: BT.2100-2.
- [17] Hable, J. (2010) Uncharted 2: HDR Lighting. *Proceedings of the Game Developers Conference*, San Francisco, 56.
- [18] Reinhard, E., Stark, M., Shirley, P. and Ferwerda, J. (2023) Photographic Tone Reproduction for Digital Images. In: *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, ACM, 661-670. <https://doi.org/10.1145/3596711.3596781>
- [19] Zhang, Y., Li, Q., Qi, M., Liu, D., Kong, J. and Wang, J. (2023) Multi-Scale Frequency Separation Network for Image Deblurring. *IEEE Transactions on Circuits and Systems for Video Technology*, **33**, 5525-5537. <https://doi.org/10.1109/tcsvt.2023.3259393>
- [20] Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S. and Yang, M. (2022) Restormer: Efficient Transformer for High-Resolution Image Restoration. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 19-20 June 2022, 5728-5739. <https://doi.org/10.1109/cvpr52688.2022.00564>
- [21] ITU-R (2021) Methods for Conversion of High Dynamic Range Content to Standard Dynamic Range Content and Vice-versa: BT.2446-1.
- [22] Banterle, F., Artusi, A., Debattista, K., *et al.* (2017) *Advanced High Dynamic Range Imaging*. 2nd Edition, AK Peters (CRC Press).
- [23] Drago, F., Myszkowski, K., Annen, T. and Chiba, N. (2003) Adaptive Logarithmic Mapping for Displaying High Contrast Scenes. *Computer Graphics Forum*, **22**, 419-426. <https://doi.org/10.1111/1467-8659.00689>
- [24] Ferwerda, J.A., Pattanaik, S.N., Shirley, P. and Greenberg, D.P. (1996) A Model of Visual Adaptation for Realistic Image Synthesis. *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, New Orleans, 4-9 August 1996, 249-258. <https://doi.org/10.1145/237170.237262>
- [25] Lischinski, D., Farbman, Z., Uyttendaele, M. and Szeliski, R. (2006) Interactive Local Adjustment of Tonal Values. *ACM Transactions on Graphics*, **25**, 646-653. <https://doi.org/10.1145/1141911.1141936>
- [26] Reinhard, E., Pouli, T., Kunkel, T., Long, B., Ballestad, A. and Damberg, G. (2012) Calibrated Image Appearance Reproduction. *ACM Transactions on Graphics*, **31**, 1-11. <https://doi.org/10.1145/2366145.2366220>
- [27] Shan, Q., DeRose, T. and Anderson, J. (2012) Tone Mapping High Dynamic Range Videos Using Wavelets. *Pixar Technical Memo*, **1**, 1-20.
- [28] Larson, G.W., Rushmeier, H. and Piatko, C. (1997) A Visibility Matching Tone Reproduction Operator for High Dynamic Range Scenes. *IEEE Transactions on Visualization and Computer Graphics*, **3**, 291-306. <https://doi.org/10.1109/2945.646233>
- [29] Kingma, D.P. and Ba, J. (2014) Adam: A Method for Stochastic Optimization.
- [30] Zhang, L. and Li, H. (2012) SR-SIM: A Fast and High Performance IQA Index Based on Spectral Residual. 2012 *19th IEEE International Conference on Image Processing*, Orlando, 30 September-3 October 2012, 1473-1476. <https://doi.org/10.1109/icip.2012.6467149>
- [31] ITU-R (2019) Objective Metric for the Assessment of the Potential Visibility of Colour Differences in Television: BT.2124-0.
- [32] Mantiuk, R.K., Hammou, D. and Hanji, P. (2023) HDR-VDP-3: A Multi-Metric for Predicting Image Differences, Quality and Contrast Distortions in High Dynamic Range and Regular Content.