

基于特征校准与频域解耦的视频隐式神经表示

余昊辉, 常青玲, 欧阳涛, 潘冠麟, 黄智鸿

五邑大学电子与信息工程学院, 广东 江门

收稿日期: 2026年1月1日; 录用日期: 2026年1月29日; 发布日期: 2026年2月9日

摘要

视频隐式神经表示(INR)为视频压缩提供了全新范式,但现有基于NeRV的方法仍受限于特征表示的各向同性冗余,以及深度网络固有的“频谱偏差”导致的高频纹理丢失。为此,本文提出一种基于特征校准与频域解耦的视频隐式神经表示网络(CF-NeRV)。首先,通过重构基础解码单元提出内容自适应特征校准模块(CFC-Block),在卷积前引入并发时空注意力机制以精准预校准显著特征并抑制背景冗余;其次,设计频域感知细化模块(FARM),通过显式频域解耦与自适应门控策略强制模型补偿高频残差信号;最后,引入误差感知解耦重建头(ERRH),通过主、副支路的专业化分工实现从特征到像素空间的高精度动态映射。在Bunny及UVG上的7个数据集上的实验结果表明,CF-NeRV的重建质量显著优于HNeRV、E-NeRV等主流模型,其中在UVG数据集上实现了1.29 dB的平均PSNR提升,充分验证了所提机制在复杂视频重建任务中的优越性与高效性。

关键词

视频隐式神经表示, 特征校准, 频域解耦, 频谱偏差

CF-NeRV: Video Implicit Neural Representation Based on Feature Calibration and Frequency Decoupling

Haohui Yu, Qingling Chang, Tao Ouyang, Guanlin Pan, Zhihong Huang

School of Electronics and Information Engineering, Wuyi University, Jiangmen Guangdong

Received: January 1, 2026; accepted: January 29, 2026; published: February 9, 2026

Abstract

Video Implicit Neural Representation (INR) has emerged as a promising paradigm for video compression. However, existing NeRV-based approaches are still limited by the isotropic redundancy in feature representation and the loss of high-frequency details caused by the inherent “spectral bias” of deep

文章引用: 余昊辉, 常青玲, 欧阳涛, 潘冠麟, 黄智鸿. 基于特征校准与频域解耦的视频隐式神经表示[J]. 计算机科学与应用, 2026, 16(2): 214-222. DOI: 10.12677/csa.2026.162053

neural networks. To address these challenges, we propose CF-NeRV, a video implicit neural representation network based on feature calibration and frequency decoupling. Specifically, we first reconstruct the basic decoding unit by introducing a Content-Adaptive Feature Calibration module (CFC-Block), which incorporates concurrent spatio-temporal attention mechanisms prior to convolution to accurately pre-calibrate salient features and suppress background redundancy. Secondly, a Frequency-Aware Refinement Module (FARM) is designed to force the model to compensate for missing high-frequency residual signals through an explicit frequency-domain decoupling strategy and adaptive gating mechanisms. Finally, an Error-Aware Decoupled Reconstruction head (ERRH) is introduced to achieve high-precision dynamic mapping from the feature space to the pixel space via a specialized functional division between the main and auxiliary branches. Experimental results on 7 benchmarks, including Bunny and UVG datasets, demonstrate that CF-NeRV significantly outperforms state-of-the-art models such as HNeRV and E-NeRV. Notably, our method achieves an average PSNR improvement of 1.29 dB on the UVG dataset, validating the superiority and efficiency of the proposed mechanism in complex video reconstruction tasks.

Keywords

Video Implicit Neural Representation, Feature Calibration, Frequency Decoupling, Spectral Bias

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

随着高清及超高清视频业务的爆发式增长，如何在极低带宽下实现高质量的视频存储与传输已成为计算机视觉与信号处理领域的核心课题。传统的视频编码标准(如 HEVC 和 VVC)虽然在工业界取得了巨大成功，但其高度依赖手工设计的预测与变换模块，难以充分挖掘复杂时空信号中的深度非线性冗余。近年来，隐式神经表示(Implicit Neural Representations, INR)以其特有的函数映射范式为视频压缩带来了全新的机遇。不同于传统像素域编码，视频神经表示将视频序列参数化为神经网络权重，通过优化连续函数的回归过程来实现视频的端到端表征与重建。

在这一演进过程中，NeRV 及其后续改进工作(如 HNeRV 和 E-NeRV)成功推动了从“坐标映射”到“帧映射”的范式迁移。通过引入内容自适应嵌入与卷积解码器，这些方法在率失真性能上展现出超越传统标准的潜力。然而，现有的视频神经表示框架在迈向更高保真度重建时仍面临关键阻碍。首先是特征表示的各向同性冗余：标准解码块倾向于对所有空域位置和通道维度进行均质化处理，缺乏对视频中动态显著区域的自适应重校准。其次是深度网络固有的“频谱偏差”(Spectral Bias)现象：模型在训练过程中倾向于优先拟合低频分量，导致细微的纹理细节在逐层上采样过程中被平滑，使得重建视频在复杂高频场景下出现不可忽视的模糊伪影。

针对上述挑战，本文提出了一种基于特征校准与频域解耦的视频隐式神经表示网络(CF-NeRV)。我们重构基础计算单元，提出内容自适应特征校准模块(CFC-Block)，该模块改变了传统先卷积后注意力的范式，通过引入并发的空间与通道注意力算子，在卷积操作前对特征流进行“预校准”，从而在源头上抑制任务无关的背景冗余，突出关键显著性特征。

针对高频细节丢失问题，我们设计了频域感知细化模块(FARM)。利用显式的频域解耦策略，该模块将特征分解为互补的高低频分量进行针对性增强，并引入自适应门控机制强制模型补偿丢失的高频残差信号。最后，我们引入误差感知的解耦重建头(ERRH)。通过主重建流与误差修正流的专业化分工，ERRH

实现了从特征空间到像素空间的高精度映射，通过像素级的动态修正进一步提升了重建精度。

本文的主要贡献可归纳如下：

- 1) 提出 CF-NeRV，一种旨在解决特征冗余与高频衰减的新型视频神经表示架构。
- 2) 设计 CFC-Block 和 FARM 模块，分别从时空特征预校准与显式频域细化的角度提升了模型的表征效率与图像保真度。
- 3) 在 Bunny 及 UVG 上的 7 个数据集上进行了广泛评估。实验结果表明，CF-NeRV 在多项评估指标上超过目前大部分现有的模型。例如，在 UVG 数据集上，本方法相比 HNeRV 实现了 1.29 dB 的平均 PSNR 提升，证明了所提机制在复杂视频重建任务中的优越性。

2. 相关工作

2.1. 隐式视频表示

隐式神经表示(INR)旨在利用连续函数将视觉信号参数化为神经网络权重。早期的坐标映射范式以时空坐标 (x, y, t) 为输入，利用多层感知机(MLP)回归像素连续场。然而，受限于推理复杂度随分辨率线性增长及全局一致性缺失，该范式难以支持高清视频的实时处理。随后，以 NeRV [1]为代表的帧映射范式(Frame-based)推动了架构演进，通过将时间索引映射为完整的视频特征图，显著提升了模型的表征效率与时空拓扑建模能力。

2.2. NeRV 系列方法

作为帧映射 INR 的基石，NeRV 首次利用卷积解码器实现了高效的视频重构。针对 NeRV 固定嵌入向量(Fixed Embeddings)造成的表征瓶颈，HNeRV [2]引入内容自适应机制与层级化架构，通过动态提取输入帧的潜在特征，大幅增强了模型对复杂时序变化的拟合精度。在此基础上，E-NeRV [3]通过对卷积算子的深度分解与架构轻量化，实现了在极低计算负载下的快速收敛。此外，FANeRV 与 HiNeRV 等工作进一步探索了多尺度细化与频率感知策略，旨在缓解复杂场景下的重建模糊问题。

2.3. 基于隐式神经表示的视频压缩

传统的视频压缩方法，如 H.264 [4]和 H.265 [5]，可以实现良好的重建效果，并且具有不错的解压缩速度。最近，视频神经表示通过对模型权重进行剪枝、量化及熵编码，将视频信息存储于比特流中，其性能核心在于率失真权衡(Rate-Distortion Trade-off)的优化。尽管现有方法在宏观结构上进行了诸多改良，但卷积解码器在逐层上采样过程中存在的时空特征混叠，以及深度网络固有的频谱偏差导致的高频细节衰减，仍是制约超高质量重建的关键瓶颈。

3. 方法

3.1. 内容自适应的特征校准 NeRV 模块

作为视频隐式神经表示网络的核心组件，NeRV 解码块(NeRV Block)主要承担着将低维潜层编码逐步映射回高维像素空间的上采样任务。在现有的主流架构(如 HNeRV、E-NeRV)中，标准的解码块(NeRVBlock)通常采用“卷积-归一化-激活”的级联结构。然而，这种各向同性的处理范式在处理高度压缩的视频特征时存在显著局限：它忽略了特征通道间的语义差异以及空间位置上的内容非均匀性。为了克服这一缺陷，受 scSE 机制启发，我们重构了基础解码单元，提出了一种内容自适应的特征校准 NeRV 模块(Content-Adaptive Feature Calibrated NeRV Block, CFC-Block)。

视频数据本质上具有极强的时空冗余性，且高频纹理信息往往稀疏分布于特定的空间区域(如物体边

缘)和语义通道中。传统的解码块在进行上采样操作时, 往往会对无效背景噪声和关键纹理特征进行无差别的放大, 导致重建图像出现伪影或模糊。为了直观展示特征流的重构过程, 我们在图 1 中对比了标准 NeRV 解码块与本文提出的 CFC-Block。不同于标准结构直接进行卷积上采样, CFC-Block 在卷积操作前显式嵌入了并行校准算子 Φ_{calib} , 具体处理流程被形式化定义为:

$$\mathbf{F}_{\text{out}} = \sigma\left(\mathcal{N}\left(\mathcal{F}_{\text{up}}\left(\Phi_{\text{calib}}\left(\mathbf{F}_{\text{in}}\right)\right)\right)\right) \quad (1)$$

其中, \mathbf{F}_{in} , \mathbf{F}_{out} 分别表示输入与输出特征; $\mathcal{F}_{\text{up}}(\cdot)$ 代表上采样卷积层; $\mathcal{N}(\cdot)$ 和 $\sigma(\cdot)$ 分为归一化层和激活函数; 而 $\Phi_{\text{calib}}(\cdot)$ 则是我们新引入的并行特征校准算子。这种设计遵循“先筛选, 后放大”的原则, 确保计算资源被集中分配至最具信息量的特征区域, 从而显著提升了模型的表征效率。为了实现高效的特征筛选, Φ_{calib} 算子采用了并行的双流架构, 分别从空间结构和通道语义两个维度对输入特征 $\mathbf{X} \in \mathbb{R}^{C \times H \times W}$ 进行重构。

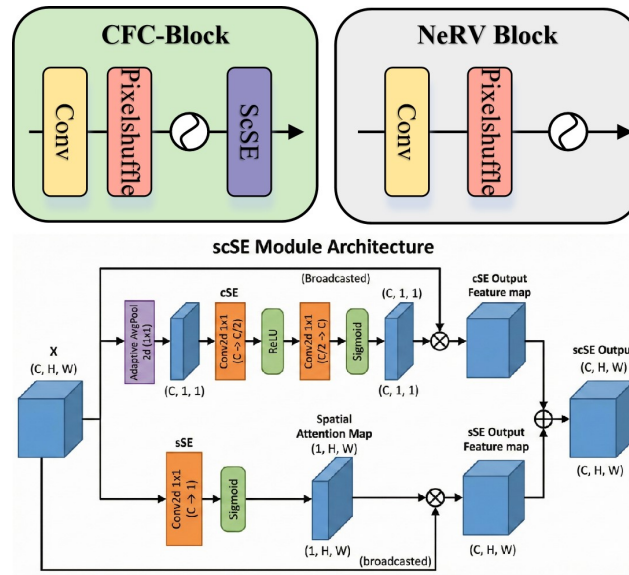


Figure 1. Structural comparison between NeRVBlock and the proposed CFC-Block
图 1. NeRVBlock 与 CFC-Block 结构对比

空间结构感知流。为了在解码早期恢复高频几何细节, 该分支致力于生成位置敏感的注意力掩码。不同于复杂的自注意力机制, 我们采用轻量级 1×1 卷积直接将特征图压缩为单通道权重图 $\mathbf{M}_s \in \mathbb{R}^{1 \times H \times W}$ 。该权重图通过 Sigmoid 激活后, 作为空间滤波器对特征进行加权:

$$\mathbf{U}_s = \mathbf{X} \odot \sigma(\mathbf{W}_s * \mathbf{X}) \quad (2)$$

这一过程显式地增强了边缘和纹理丰富区域的特征响应, 使得网络在随后的上采样过程中能够更好地保持几何结构的一致性。

通道语义选择流。考虑到隐式表示中的不同通道往往编码了不同层次的视觉模式, 该分支旨在抑制冗余通道的干扰。我们首先利用全局平均池化获取全局上下文描述符 $\mathbf{z} \in \mathbb{R}^C$, 随后通过两个全连接层(由 1×1 卷积实现)构建的瓶颈结构来捕获通道间的非线性依赖关系:

$$\mathbf{U}_c = \mathbf{X} \odot \sigma(\mathbf{W}_2 * \delta(\mathbf{W}_1 * \mathbf{z})) \quad (3)$$

其中 \mathbf{W}_1 和 \mathbf{W}_2 为卷积核权重, δ 为 ReLU 激活函数。该分支有效地实现了特征的通道级重标定。

最终, 校准后的特征 F_{calib} 由上述两个互补流的输出融合而成: $F_{\text{calib}} = U_s + U_c$ 。与现有的方法相比, 我们设计的 CFC-Block 具有以下核心优势, 内在的去噪能力, 通过将校准算子 Φ_{calib} 前置于卷积层, 我们在特征被上采样放大之前就滤除了无关噪声。这一设计思想与 HNeRV 中强调的高效特征交互一脉相承, 但以更显式的方式增强了信噪比。参数与性能的极佳平衡, 得益 1×1 卷积的轻量化特性, CFC-Block 在仅增加微乎其微参数量的前提下, 赋予了网络强大的内容自适应能力。这符合 E-NeRV 等近期工作所追求的高效率设计理念。

3.2. 单位频域解耦的纹理细化机制

尽管我们在基础解码单元中引入 CFC-Block 以增强特征的纯净度, 但深度神经网络在视频重建任务中仍面临着一个共性难题“频谱偏差”。现有的隐式表示模型倾向于优先拟合低频分量(如平滑背景和大尺度结构), 而高频分量(如细微纹理和边缘细节)往往在深层网络的逐层传递中发生衰减甚至丢失。为了在特征解码末端打破这一瓶颈, 显式地恢复高保真的细节信息, 我们提出了一种频域解耦的纹理细化模块(Frequency-Aware Refinement Module, FARM)。

不同于传统方法在混合特征空间中“隐式”地学习纹理, FARM 模块的核心理念在于分而治之。如图 2 所示, 该模块被置于解码器的末端, 接收来自深层网络的聚合特征 F_{dec} 。我们首先利用基于平均池化的低通滤波器 $\mathcal{LP}(\cdot)$, 显式地将输入特征解耦为低频结构分量 F_{low} 和高频残差分量 F_{high} :

$$F_{\text{low}} = \mathcal{LP}(F_{\text{dec}}), F_{\text{high}} = F_{\text{dec}} - F_{\text{low}} \quad (4)$$

这种基于减法的互补式分离策略, 确保了高频信息被完整地保留在残差流中, 避免了下采样操作带来的信息熵损失。

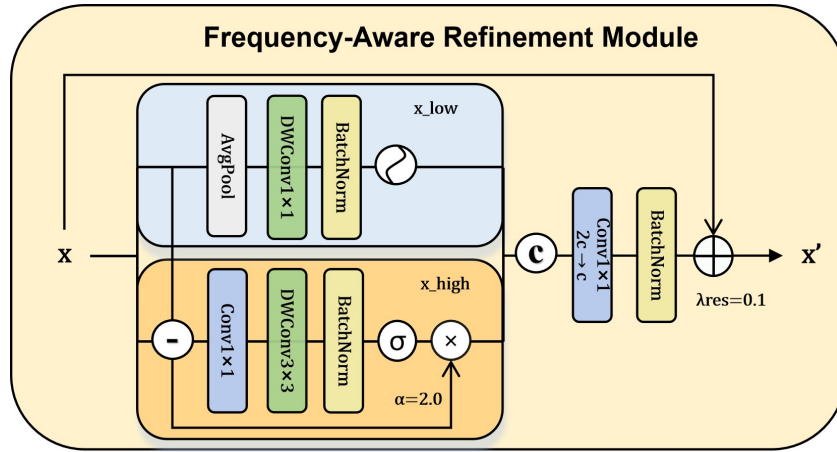


Figure 2. Flowchart of the proposed Frequency-Aware Refinement Module (FARM)
图 2. 频域解耦纹理细化模块流程图

针对分离后的不同频率分量, 我们设计了差异化的处理路径, 以适应其独特的视觉特。结构保持流。低频分量主要承载视频帧的整体布局和色调信息。为了增强其全局一致性, 我们采用膨胀卷积处理 F_{low} 。通过设置膨胀率 $d = 2$, 我们在不增加参数量的前提下有效扩大了感受野, 从而更好地捕捉大尺度的结构依赖关系, 确保重建图像的主体结构稳健。纹理增强流。高频分量包含了视频中最难以重建的边缘和纹理细节, 且往往具有稀疏性。为了从噪声中提取有效的高频信号, 我们设计了一种自适应高频门控机制。具体而言, F_{high} 首先经过深度卷积提取纹理特征, 随后通过 Sigmoid 激活函数生成高频注意力图 G_{high} 。

最终的纹理细化特征计算如下：

$$\tilde{F}_{\text{high}} = F_{\text{high}} \odot G_{\text{high}} \times \alpha \quad (5)$$

值得注意的是，我们在代码实现中引入了信号放大因子 $\alpha = 2.0$ 。这一设计动机在于对抗高频信号在网络传输中的自然衰减，强制网络关注微弱的纹理变化，从而显著提升了图像的锐度。

在分别完成结构和纹理的细化后，我们将两路特征进行拼接融合，并通过一个 1×1 卷积层进行通道整合。为了保证训练过程的稳定性并促进梯度的有效反向传播，我们借鉴 EDSR 等超分辨率网络的经验，引入了可学习的残差缩放系数 λ_{res} ：

$$F_{\text{refined}} = F_{\text{dec}} + \lambda_{\text{res}} \cdot \mathcal{F}_{\text{fusion}} \left(\left[\tilde{F}_{\text{low}}, \tilde{F}_{\text{high}} \right] \right) \quad (6)$$

其中， λ_{res} 初始化为 0.1。这种设计不仅加速了模型的收敛，还允许网络根据训练进程动态调整细化模块的贡献度，实现了从粗糙特征到精细纹理的平滑过渡。

通过引入 FARM 模块，我们的模型在物理层面实现了对视频信号频域特性的显式建模，有效地解决了 HNeRV 等基线模型在复杂纹理区域重建模糊的问题。

3.3. 公式误差感知的解耦重建机制

经过 CFC-Block 的时空预校准与 FARM 模块的频域细化后，解码器已生成了包含丰富纹理细节的高维特征表示。然而，如何将这隐式特征无损地映射回 RGB 像素空间仍是一个非平凡问题。现有的 NeRV 变体通常依赖单个卷积层作为输出头来直接生成最终图像。我们认为，这种单一映射策略迫使卷积核在恢复全局色彩与修正局部高频残差之间进行权衡，往往导致网络倾向于输出“平均化”的预测结果，从而产生模糊的视觉伪影。为了缓解这一映射瓶颈，我们提出了一种误差感知的解耦重建机制(ERRH)。

ERRH 的核心动机在于将复杂的图像重建任务分解为两个正交的子任务，基础内容逼近与重建误差修正。正如正文图 3 整体架构的输出阶段所示，我们并未采用单路卷积直接输出结果，通过引入并行的主头与误差头，我们允许网络在最后的映射阶段进行专业化分工。而是引入了并行的主头(Main Head)与误差头(Error Head)，允许网络在最后的映射阶段进行专业化分工。

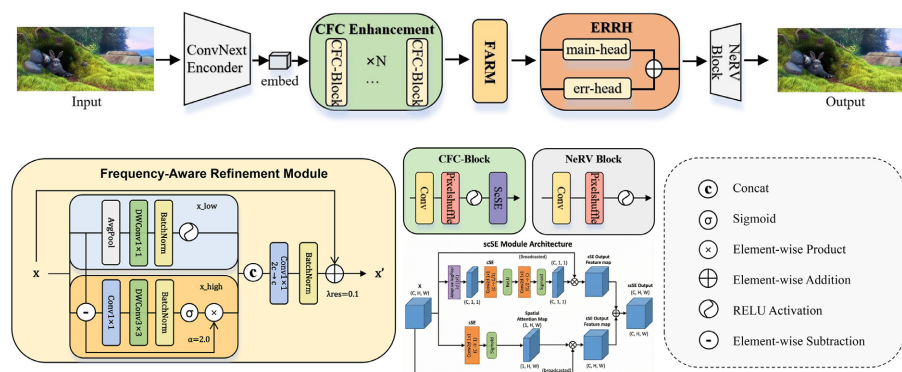


Figure 3. Overall network architecture diagram
图 3. 整体网络架构图

主重建流，该分支由一个标准的 3×3 卷积层构成，旨在生成图像的基础底色与主体结构。其优化目标是最小化全局均方误差，确保重建视频在色彩保真度和整体布局上的准确性。我们将该分支的输出记为 I_{main} 。误差修正流，该分支同样采用 3×3 卷积，但其学习目标专注于捕捉主重建流难以拟合的困难样

本，即那些细微的纹理偏差或边缘伪影。该分支并不直接输出完整的图像，而是预测一个残差图 I_{err} ，用于对 I_{main} 进行像素级的动态修补。为了将两个分支的输出有机结合，并保证训练初期的稳定性，我们设计了一个可学习的线性融合过程。最终的预测图像 I_{pred} 通过以下公式获得：

$$I_{\text{pred}} = \mathcal{A}(I_{\text{main}} + \beta \cdot I_{\text{err}}) \quad (7)$$

其中， $\mathcal{A}(\cdot)$ 表示将输出约束至像素值范围的激活函数(如 Sigmoid 或 Tanh)； β 是一个可学习的缩放系数。

在代码实现中，我们将 β 初始化为 0.1。这一初始化的物理意义在于实施软启动策略，在训练初期，网络主要依赖主头快速学习视频的大致轮廓，此时误差头的权重较低，避免了随机初始化带来的噪声干扰；随着训练的深入， β 会根据梯度流自动调整，逐渐增加误差头的贡献度，使其专注于精修那些主头无法完美重建的高频细节。

这种解耦重建机制构成了我们需要提出的完整方法论的最后一环。它与前文提出的 CFC-Block 和 FARM 模块形成了紧密的逻辑互补，CFC-Block 在编码端净化了特征输入；FARM 在解码中段恢复了频域细节；ERRH 在输出端通过残差修正确保了像素级的精准映射。三者共同协作使我们的模型在有效提升 PSNR 指标的同时，显著改善了视频帧的主观视觉质量。

4. 实验

4.1. 实验设置与实现细节

本文在 Bunny 及 UVG 上的 7 个数据集上评估 CF-NeRV 的性能。重建质量通过峰值信噪比(PSNR)和多尺度结构相似性(MS-SSIM)进行定量评估，压缩效率则采用每像素比特数(bpp)衡量。通过调整模型宽度，我们在不同码率下对比分析了各模型的率失真性能(Rate-Distortion Performance)。

所有模型均基于 PyTorch 框架在单张 NVIDIA RTX 3090 GPU 上完成训练。网络采用五级对称的编码器 - 解码器结构，下采样与上采样步长序列配置为[5, 4, 4, 2, 2]，特征维度自 64 逐级降至 16。卷积模块选用 ConvNeXt 结构配合像素混洗(Pixel Shuffle)算子，并统一采用 GELU 激活函数且不引入归一化层以保持原始信号动态范围。训练采 Adam 优化器，初始学习率为 1×10^{-3} ，批量大小为 2，共迭代 300 Epoch。实验每 30 个 Epoch 进行一次性能评测并选取最优指标进行分析，相关代码与权重将在稿件接收后公开发布。

4.2. 实验结果与分析

Bunny 数据集性能分析：表 1 展示了不同参数规模与训练轮数下各方法的重建精度对比。实验结果显示，CF-NeRV 在所有配置下均展现出更强的表征能力。在 0.75 M 的轻量化规模下，本文方法达到 33.59 dB 的 PSNR，优于 HNeRV (32.81 dB)和 ENeRV (30.95 dB)。当模型扩展至 3 M 参数时，PSNR 进一步提升至 38.48 dB，领先 HNeRV 约 1.05 dB。在训练效率方面，CF-NeRV 同样具有优势，在 1200 个 Epoch 时精度达到 34.93 dB。这验证了 CFC-Block 通过特征预校准显著提升了模型的表征容量，使其在有限参数下能更有效地拟合视觉内容。

UVG 数据集率失真性能：高清视频序列 UVG 上的评估结果如表 2 所示。相比基线方法，CF-NeRV 在各序列上均实现了性能增益，平均 PSNR 达到 33.98 dB，较 HNeRV 提升了 1.29 dB。特别是在 Beauty 和 Bosphorus 序列上，本方法的 PSNR 分别达到了 34.58 dB 和 35.67 dB。这些提升主要得益于 FARM 模块对频域信息的显式解耦以及 ERRH 重建头对预测误差的动态修正，有效地缓解了频谱偏差问题，通过精准捕捉高频纹理显著增强了图像的视觉保真度。

Table 1. PSNR comparison under different model scales and training epochs on the Bunny dataset**表 1.** Bunny 数据集下不同模型规模与训练轮数下的 PSNR 对比

Method	Size			Epoch		
	0.75 M	1.5 M	3 M	300	600	1200
NeRV	28.46	30.87	33.21	28.46	29.15	29.57
ENeRV	30.95	30.95	36.72	30.95	32.07	32.79
HNeRV	32.81	35.19	37.43	30.95	32.07	32.79
Ours	33.59	36.01	38.48	33.59	34.45	34.93

Table 2. Quantitative evaluation of reconstruction performance (PSNR) for video sequences in the UVG dataset**表 2.** UVG 数据集各视频序列的重建性能(PSNR)定量评估

Method	Beaut	Bosph	Honey	Jocke	Ready	Shake	Yacht	avg
NeRV	33.25	33.22	37.26	31.74	24.84	33.08	28.03	31.63
ENeRV	33.53	33.81	39.04	29.35	24.09	34.54	27.74	31.73
HNeRV	33.58	33.81	38.96	32.04	25.74	34.57	29.26	32.69
Ours	34.58	35.67	39.42	33.38	27.94	35.02	31.84	33.98

4.3. 消融实验

为了深入验证 CF-NeRV 中各核心组件对视频重建性能的贡献,我们以 HNeRV 为基准模型,在 UVG 数据集上进行了消融实验。如表 3 所示,我们通过逐步引入内容自适应特征校准模块(CFC-Block)、频域细化模块(FARM)以及误差感知重建头(ERRH)来构建不同的模型变体,并对其平均 PSNR 进行评估。

Table 3. Ablation study of CF-NeRV in terms of PSNR**表 3.** CFNeRV 在 PSNR 方面的消融研究

Method	Beaut	Bosph	Honey	Jocke	Ready	Shake	Yacht	avg
HNeRV	34.27	33.77	38.95	31.02	25.19	33.90	29.75	32.41
V1	34.29	33.91	39.01	31.25	25.35	33.92	29.95	32.53
V2	34.32	34.14	39.03	31.47	25.66	33.90	29.61	32.59
V3	34.32	34.10	39.03	31.46	25.68	33.94	30.17	32.67

特征预校准的有效性相较于基准模型(32.41 dB),引 CFC-Block 的变体 V1 在平均 PSNR 上提升至 32.53 dB。这一增益验证了第 3.1 节所述特征预校准机制的必要性,通过并行的时空校准算子,模型在卷积上采样前能够有效抑制无关的背景冗余,提升了特征表示的纯净度与显著性。

频域解耦细化的贡献在 V1 的基础上引入 FARM 模块后(变体 V2),平均 PSNR 进一步提升至 32.59 dB。该结果证明了显式频域解耦策略在缓解“频谱偏差”问题上的优势。FARM 模块通过对高低频分量的“分而治之”,强化了视频中稀疏的高频纹理残差,从而显著改善了细节区域的重建精度。

误差感知重建的作用最终模型 V3 (完整 CFNeRV)在加入 ERRH 模块后,平均 PSNR 达到了最优的 32.67 dB。ERRH 模块通过主重建流与误差修正流的专业化分工,实现了像素级的动态修正,补偿了主输出头难以拟合的映射误差。

综上所述,消融实验结果表明,本文提出的三个模块在功能上互补,协同推动了模型在保持轻量化特性的同时,实现了更优的视频表示性能。

5. 结论

本文提出了一种基于特征校准与频域解耦的新型视频隐式神经表示架构 CF-NeRV。针对现有方法在解码过程中存在的特征表达冗余及高频纹理衰减问题，我们从时空校准、频域解耦及解耦重建三个维度进行了系统性的优化。首先，通过引入内容自适应特征校准模块(CFC-Block)，模型实现了对时空显著性特征的精准捕获，在源头上抑制了各向同性卷积带来的信息冗余；其次，频域感知细化模块(FARM)通过显式的频率分流策略，有效地缓解了深度神经网络固有的“频谱偏差”现象，强制网络补偿了关键的高频残差信号；最后，误差感知的解耦重建头通过主、副支路的专业化分工，实现了从隐式特征到像素空间的高精度映射。

在 Bunny 及 UVG 上的 7 个数据集上的实验结果表明，CF-NeRV 在率失真性能与重建质量上均显著优于 HNeRV、E-NeRV 等现有主流方法。特别是在处理复杂纹理视频时，本方法展现出极强的视觉保真度与表征效率，例如在 UVG 序列上实现了 1.29 dB 的平均 PSNR 提升。本文的研究证明，通过显式引入内容感知与频域解耦机制，可以有效突破隐式视频表示在细节重构上的瓶颈。

在未来的工作中，我们将进一步探索 CF-NeRV 在更高分辨率视频(如 4K/8K)上的扩展性，并研究如何将时序一致性约束更深入地集成到解码架构中，以实现极低带宽下的高保真视频流传输。

基金项目

本研究得到了国家重点研发计划(项目编号：2022YFA1602003)“台中微子探测器智能监控”的支持。

参考文献

- [1] Chen, H., He, B., Wang, H., *et al.* (2021) Nerv: Neural Representations for Videos. *Advances in Neural Information Processing Systems*, **34**, 21557-21568.
- [2] Chen, H., Gwilliam, M., Lim, S.-N. and Shrivastava, A. (2023) HNeRV: A Hybrid Neural Representation for Videos. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, 17-24 June 2023, 10270-10279. <https://doi.org/10.1109/CVPR52729.2023.00990>
- [3] Li, Z.Z., Wang, M.M., Pi, H.J., Xu, K.C., Mei, J.B. and Liu, Y. (2022) E-nerv: Expedite Neural Video Representation with Disentangled Spatial-Temporal Context. *ECCV*, 2.
- [4] Wiegand, T., Sullivan, G.J., Bjontegaard, G. and Luthra, A. (2003) Overview of the H.264/AVC Video Coding Standard. *IEEE Transactions on Circuits and Systems for Video Technology*, **13**, 560-576. <https://doi.org/10.1109/TCSVT.2003.815165>
- [5] Sullivan, G.J., Ohm, J.-R., Han, W.-J. and Wiegand, T. (2012) Overview of the High Efficiency Video Coding (HEVC) Standard. *IEEE Transactions on Circuits and Systems for Video Technology*, **22**, 1649-1668. <https://doi.org/10.1109/TCSVT.2012.2221191>