

融合非局部注意力的大气湍流图像复原生成对抗算法研究

苏钰盛^{id}, 周奕含^{id}

长春理工大学数学与统计学院, 吉林 长春

收稿日期: 2026年3月3日; 录用日期: 2026年3月27日; 发布日期: 2026年4月8日

摘要

大气湍流效应引起的光程随机变化会导致成像系统采集到的远距离图像序列出现严重的几何畸变、像素级抖动及动态模糊, 严重制约了图像的高层语义理解与应用。针对传统物理模型方法依赖强先验假设且计算耗时, 以及现有二维深度学习方法难以有效利用时空相关性的问题, 本研究提出了一种基于三维非局部(3D non-local)注意力机制的时空残差感知Wasserstein生成对抗网络(Generative Adversarial Network, GAN)用于大气湍流退化图像序列的盲复原。该算法创新性地将非局部注意力机制扩展至三维时空域, 通过构建全时空亲和度矩阵以捕捉视频序列中跨帧与长距离的空间依赖关系, 从而有效校正非刚性几何形变。为解决高维特征交互带来的计算复杂度呈二次方增长的问题, 本研究在注意力机制的键变换与值变换路径中引入了空间子采样策略, 在保证全局感受野的同时显著降低了运算负担。此外, 网络集成了可变形卷积的时间对齐模块与空间注意力增强模块, 以解决湍流引起的时间非一致性问题。在优化目标上, 结合了感知损失、像素级均方误差损失以及Wasserstein对抗损失, 引导生成器重建出兼具结构保真度与高频纹理细节的清晰图像。实验结果表明, 该方法在合成湍流图像复原上取得了优越的复原效果。

关键词

大气湍流, 图像复原, 三维非局部注意力, 生成对抗网络, 时空特征学习

Research on a Generative Adversarial Algorithm for Atmospheric Turbulence Image Restoration Incorporating Non-Local Attention

Yusheng Su^{id}, Yihan Zhou^{id}

Abstract

The random variation of optical path length induced by atmospheric turbulence effects leads to severe geometric distortions, pixel-level jitter, and dynamic blur in image sequences captured by imaging systems at long distances, which significantly hinders high-level semantic understanding and application of the images. To address the limitations of traditional physical model methods, which rely on strong prior assumptions and are computationally intensive, and the shortcomings of existing 2D deep learning methods in effectively utilizing spatio-temporal correlations, this study proposes a spatio-temporal residual-aware Wasserstein Generative Adversarial Network (GAN) based on a 3D non-local attention mechanism for blind restoration of atmospheric turbulence-degraded image sequences. The algorithm innovatively extends the non-local attention mechanism to the 3D spatio-temporal domain, capturing cross-frame and long-range spatial dependencies within video sequences by constructing a full spatio-temporal affinity matrix, thereby effectively correcting non-rigid geometric deformations. To tackle the issue of quadratic growth in computational complexity associated with high-dimensional feature interactions, a spatial sub-sampling strategy is introduced into the key and value transformation paths of the attention mechanism, significantly reducing the computational burden while maintaining a global receptive field. Furthermore, the network integrates a temporal alignment module using deformable convolutions and a spatial attention enhancement module to address the temporal inconsistency caused by turbulence. For the optimization objective, a combination of perceptual loss, pixel-level mean squared error loss, and Wasserstein adversarial loss is employed to guide the generator in reconstructing clear images with both structural fidelity and high-frequency texture details. Experimental results demonstrate that the proposed method achieves superior restoration performance on both synthetic turbulence data.

Keywords

Atmospheric Turbulence, Image Restoration, 3D Non-Local Attention, Generative Adversarial Network (GAN), Spatio-Temporal Feature 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

大气湍流是远距离光学成像系统中不可避免的物理现象, 其由温度、气压及风速等因素引起的折射率随机波动, 会导致成像过程中出现几何畸变、局部模糊、对比度下降以及明显的时序抖动等多种退化效应。与传统光学模糊不同, 大气湍流退化同时具有显著的空间非均匀性与时间随机性, 这使得基于单帧假设的图像复原方法在复杂湍流条件下往往难以取得理想效果。Hufnagel 和 Stanley 从统计光学角度分析了湍流对成像系统调制传递函数的影响, 揭示了湍流退化的随机本质, 为后续研究奠定了理论基础[1]。随后, Fried 系统性研究了湍流引起的相位扰动及其对成像分辨率的限制, 进一步指出了湍流退化中蕴含的空间相关特性[2]。

针对大气湍流退化图像的复原问题, 早期研究主要依赖基于物理模型的方法, 如相位屏建模、光学传递函数补偿以及自适应光学技术等。这类方法通常依赖对湍流参数的精确估计, 并假设成像过程满足特定物理模型。Roggemann 和 Welsh 对基于物理先验的湍流校正方法进行了系统总结, 指出其在弱湍流条件下具有一定有效性, 但在强湍流和复杂场景中鲁棒性显著下降[3]。Hill 等人使用的幸运成像方法通过选取相对清晰的帧来缓解湍流影响, 但其性能高度依赖帧选择策略, 且在中高强度湍流条件下可用帧比例极低[4]。总体而言, 传统方法在建模灵活性与泛化能力方面均存在明显局限。

近年来, 深度学习在图像复原领域的快速发展为大气湍流退化图像复原提供了新的研究思路。卷积神经网络(Convolutional Neural Network, CNN)凭借其强大的特征表达能力, 被广泛应用于去噪、去模糊及超分辨率等底层视觉任务。Dong 等人提出的 SRCNN 首次验证了端到端学习在图像复原任务中的有效性[5], Nah 等人进一步通过多尺度 CNN 结构实现了动态模糊图像复原, 在复杂退化条件下显著提升了重建质量[6]。在湍流复原领域, Chak 等人提出的 Subsampled Turbulence Removal Network 证明了在无显式物理模型条件下, 基于数据驱动的方法同样能够有效缓解湍流退化效应[7]。

然而, 大气湍流退化通常表现为跨时间序列的连续扰动, 仅依赖单帧信息难以充分刻画其动态特性。为此, 研究者逐渐将视频建模思想引入湍流复原任务。Kappeler 等人利用多帧卷积神经网络对时序信息进行联合建模, 在一定程度上改善了复原结果的时序一致性[8]。随后, 基于三维卷积的时空特征建模方法在视频去噪和视频超分辨率任务中得到广泛应用。Tran 等人提出的 C3D 网络展示了 3D 卷积在联合建模时间与空间特征方面的优势[9], Caballero 等人的方法进一步验证了多帧联合建模对于提升视觉质量的有效性[10]。这些研究表明, 充分挖掘时序冗余信息是提升湍流复原性能的重要方向。

在此基础上, GAN 网络为图像复原问题提供了新的优化视角。Goodfellow 等人提出的 GAN 框架通过对抗学习实现了对复杂数据分布的建模[11], Ledig 等人将其引入图像超分辨率任务, 在感知质量方面取得了显著提升[12]。Arjovsky 等人提出的 Wasserstein GAN 通过引入 Wasserstein 距离缓解了训练不稳定问题, 为高质量图像生成提供了更加稳健的理论基础[13]。Sajadi 等人表明, 在视频复原任务中引入对抗学习有助于减少过度平滑现象, 提升结构一致性与视觉真实感[14]。

尽管现有方法在一定程度上改善了湍流退化图像序列的复原效果, 但多数方法仍主要依赖局部卷积操作, 难以有效捕捉远距离的时空依赖关系。为解决这一问题, 注意力机制逐渐受到关注。Vaswani 等人提出的 Transformer 模型在序列建模中展示了强大的全局依赖建模能力[15], Cao 等人提出的递归视频复原 Transformer (RVRT)通过引导可变形注意力机制, 在多项视频复原基准上取得了先进性能[16]。与此同时, Wang 等人提出的 Non-local Network 通过显式计算特征图中任意位置之间的相关性, 实现了长程依赖建模, 并在视频理解与视觉复原任务中展现出良好性能[17]。

针对大气湍流这一具有显著时空耦合特性的退化问题, Zhang 等人从 spatio-temporal translational 的角度对湍流缓解问题进行了系统分析, 强调了跨帧结构一致性在湍流复原中的重要作用[18]。然而, 现有 Non-local 或 Transformer 方法多直接沿用二维或通用视频建模框架, 计算复杂度较高, 且在高分辨率、长序列湍流数据场景下的适用性仍受到一定限制。因此, 如何在控制计算开销的前提下, 高效地建模大气湍流退化图像序列中的长程时空依赖关系, 仍是一个亟待深入研究的问题。

基于上述分析, 本研究在 TSR-WGAN 框架基础上, 引入一种面向时空图像序列的三维 Non-local 注意力机制, 用于增强模型对大气湍流退化图像序列中长程时空依赖关系的建模能力。通过在特征空间中联合建模时间与空间维度的全局相关性, 并结合合理的子采样策略以控制计算复杂度, 所提出的方法旨在保持模型稳定性的同时, 进一步提升复原图像的结构一致性和视觉质量, 为大气湍流退化图像序列复原提供一种有效的解决方案。

2. 方法

2.1. 算法整体架构

针对大气湍流环境下图像序列存在的复杂退化问题, 本研究构建了基于融合了 3D non-local 注意力机制的 TSR-WGAN 复原框架。该框架由生成器与鉴别器共同构成, 旨在建立退化图像序列至潜在清晰图像域的高维非线性映射。输入数据设定为时间长度为 T 的连续图像序列。生成器采用深层编码解码器结构, 其内部处理流程在逻辑上划分为浅层特征提取、深层非局部时空建模以及时空特征重建三个阶段, 其全局架构流程图如下。

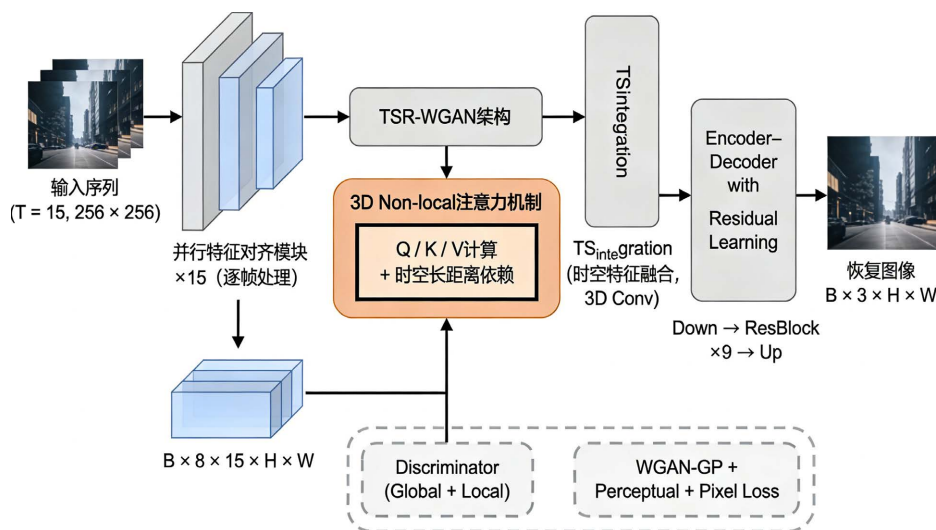


Figure 1. Overall architecture of the TSR-WGAN framework incorporating the 3D non-local attention mechanism
图 1. 融合 3D non-local 注意力机制的 TSR-WGAN 全局架构流程图

见图 1, 与常规的二维卷积网络不同, 本研究在特征提取前端即引入三维卷积层, 将输入序列投影至高维特征空间以保留初始的时空连续性。网络的核心部分由堆叠的残差块与改进的三维非局部注意力模块交替构成。这一设计的核心价值在于克服局部卷积感受野受限的缺陷, 通过捕捉长距离的时空依赖关系来修正大范围的几何扭曲。在解码阶段, 网络通过反卷积层逐步恢复空间分辨率。鉴别器方面则采用了全卷积 PatchGAN 结构, 通过判别局部图像块的真伪分布, 驱动生成器优化高频纹理细节, 从而避免传统均方误差损失导致的图像平滑效应。

2.2. 数据集与问题定义

本研究针对大气湍流条件下获取的退化图像序列, 研究其向清晰图像域的映射关系。为便于后续模型结构与优化目标的数学描述, 首先对数据集形式及问题建模进行统一定义。

设训练数据集表示为

$$\mathcal{D} = \left\{ \left(X^{(i)} \mid Y^{(i)} \right) \right\}_{i=1}^N \quad (1)$$

其中, N 表示样本总数, $X^{(i)}$ 为第 i 个受大气湍流影响的退化图像序列, $Y^{(i)}$ 为与之对应的清晰参考图像。

对于每一个退化样本, 其输入序列定义为

$$X^{(i)} = \left\{ x_t^{(i)} \right\}_{t=1}^T, x_t^{(i)} \in \mathbb{R}^{H \times W \times C} \quad (2)$$

其中, T 表示输入序列的时间长度, H 与 W 分别表示图像的空间高度与宽度, C 表示通道数。序列中各帧在时间维度上连续采样, 能够刻画大气湍流引起的时序抖动与空间非刚性畸变特性。

对应的清晰参考输出定义为

$$Y^{(i)} \in \mathbb{R}^{H \times W \times C} \quad (3)$$

表示在无湍流或弱湍流条件下获得的高质量图像, 用作监督学习过程中的目标信号。需要指出的是, 本研究以多帧退化序列恢复单帧清晰图像为主要任务设置, 旨在利用序列间的时空冗余信息对单帧中严重退化区域进行有效补偿。

在上述数据定义基础上, 本研究的目标可形式化为学习一个由参数 θ 控制的非线性映射函数

$$G_\theta: X \rightarrow \hat{Y} \quad (4)$$

其中, G_θ 表示生成器网络, \hat{Y} 为模型输出的复原结果。该映射函数通过端到端训练方式进行优化, 使得生成结果在像素分布、结构一致性及感知质量等方面尽可能逼近真实清晰图像 Y 。

在后续章节中, 所有涉及的时空特征张量、注意力权重计算以及损失函数设计, 均基于上述数据表示形式与符号约定展开, 以保证方法描述的数学一致性与逻辑严谨性。

2.3. 三维非局部注意力机制设计

2.3.1. 理论原理与时空扩展

Non-local 操作的核心思想是建立特征图中所有位置之间的全局依赖关系, 从而超越传统卷积操作的局部感受野限制。对于输入特征图 $X \in \mathbb{R}^{T \times H \times W \times C}$, 其中 T 、 H 、 W 、 C 分别表示时间、高度、宽度和通道维度, Non-local 操作定义为:

$$Y_i = \frac{1}{C(X)} \sum_{\forall j} f(X_i, X_j) g(X_j) \quad (5)$$

其中 i 和 j 枚举所有时空位置, f 计算位置 i 和 j 之间的相似度函数, g 对位置 j 的特征进行变换, $C(X)$ 为归一化因子。

在本文中, 采用点积相似度函数 $f(X_i, X_j) = \theta(X_i)^\top \phi(X_j)$, 其中 $\theta(\cdot)$ 和 $\phi(\cdot)$ 为线性变换函数。这一设计使得模型能够计算特征图中任意两个位置之间的相关性, 从而捕捉长程时空依赖关系。注意力权重通过 softmax 函数归一化:

$$A_{ij} = \frac{\exp(\theta(X_i)^\top \phi(X_j) / \tau)}{\sum_k \exp(\theta(X_i)^\top \phi(X_k) / \tau)} \quad (6)$$

其中 τ 为温度参数, 用于调节注意力分布的尖锐程度。最终输出特征为加权求和:

$$Y_i = \sum_j A_{ij} g(X_j) \quad (7)$$

实现对全序列信息的有效聚合。

2.3.2. 三维扩展实现

针对湍流图像序列的时空特性, 本文将传统的 2D Non-local 操作扩展为 3D 版本。3D Non-local 注意力模块包含三个核心变换: 查询变换 θ 、键变换 ϕ 和值变换 g 。这三个变换通过 $1 \times 1 \times 1$ 的 3D 卷积实现, 分别将输入特征映射到不同的子特征空间, 3D Non-local 模块内部逻辑拓扑图如下。

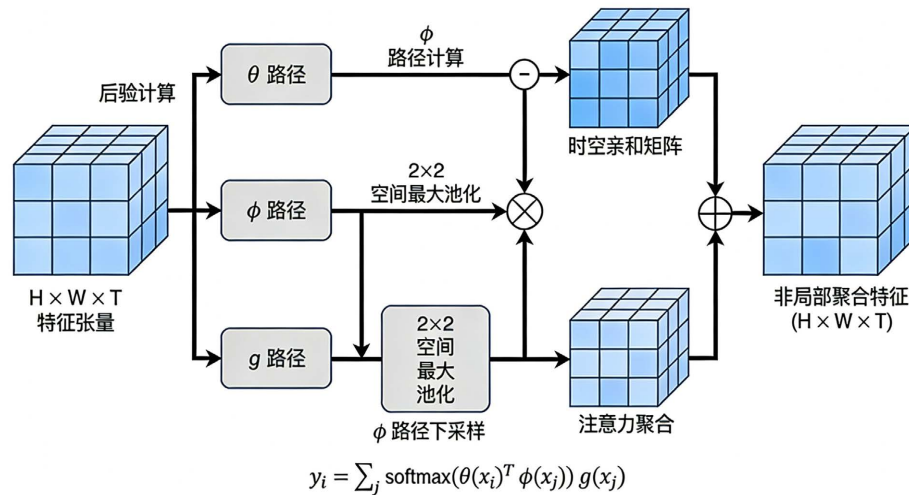


Figure 2. Internal logical topology of the 3D Non-local module
图 2. 3D Non-local 模块内部逻辑拓扑图

见图 2 在具体的张量运算中, 查询变换 θ 将输入特征 X 映射为查询特征 $Q \in \mathbb{R}^{B \times C' \times T \times H \times W}$, 其中 $C' = C/16$ 为降维后的通道数。键变换 ϕ 产生键特征 $K \in \mathbb{R}^{B \times C' \times T \times H \times W}$, 值变换 g 生成值特征 $V \in \mathbb{R}^{B \times C' \times T \times H \times W}$ 。

注意力权重的计算涉及查询特征与键特征的矩阵乘法操作。首先将 Q 和 K 重塑为二维矩阵形式, 然后计算相似度矩阵:

$$S = Q^T K \in \mathbb{R}^{(T \times H \times W) \times (T \times H \times W)} \quad (8)$$

经过 softmax 归一化后得到注意力权重矩阵 A , 最后与值特征 V 相乘得到增强后的时空特征。

2.3.3. 子采样优化策略

为降低 3D Non-local 操作的计算复杂度, 本文在键变换 ϕ 路径引入空间子采样策略。该策略基于注意力权重计算对精确键特征空间分辨率要求不高的观察, 通过适度降低空间分辨率来平衡计算效率与模型性能。

设原始键特征为 $K \in \mathbb{R}^{B \times C' \times T \times H \times W}$, 子采样后的键特征 K_{sub} 通过空间最大池化操作获得:

$$K_{\text{sub}} = \text{MaxPool}_{2 \times 2}(K) \in \mathbb{R}^{B \times C' \times T \times H/2 \times W/2} \quad (9)$$

这一操作将键特征的空间维度减半, 从而显著降低注意力权重计算的计算复杂度。原始计算复杂度为 $\mathcal{O}((T \times H \times W)^2)$, 子采样后降低至 $\mathcal{O}((T \times H \times W) \times (T \times H/2 \times W/2))$ 。

子采样策略的理论依据在于注意力机制的本质特性。注意力权重主要反映特征间的相对重要性, 而非绝对空间位置信息。适度降低键特征的空间分辨率对注意力分布的准确性影响有限, 同时能够大幅减少计算开销。

数学上, 这一策略可以视为在注意力计算中引入了一个近似:

$$A_{ij} \approx \frac{\exp(\theta(X_i)^T \phi_{\text{sub}}(X_j)/\tau)}{\sum_k \exp(\theta(X_i)^T \phi_{\text{sub}}(X_k)/\tau)} \quad (10)$$

其中 ϕ_{sub} 表示包含子采样的键变换函数。实验结果表明, 该近似在保持模型性能基本不变的前提下, 显著提升了计算效率。

3. 实验结果

3.1. 模拟湍流条件下的实验结果分析

本节给出本研究方法在人工噪声模拟的湍流数据上的实验结果。通过调节人工湍流噪声的模拟强度, 分别构建了低湍流与中湍流两种退化条件, 用以评估模型在不同湍流强度下的图像序列复原性能。实验采用峰值信噪比(PSNR)与结构相似性指数(SSIM)作为定量评价指标, 并结合代表性复原结果对模型性能进行分析。

3.1.1. 定量实验结果

在低湍流与中湍流条件下, 模型在测试集上的 PSNR 与 SSIM 统计结果分别如表 1 所示。

Table 1. PSNR and SSIM results under different turbulence conditions

表 1. 不同湍流条件下的 PSNR/SSIM 结果

方法	PSNR (低湍流)	SSIM (低湍流)	PSNR (中湍流)	SSIM (中湍流)
本研究	29.7536	0.9862	26.5832	0.9062

由表 1, 从定量结果可以观察到, 随着人工噪声模拟强度的提升, PSNR 与 SSIM 指标整体呈下降趋势, 说明中湍流条件下的退化程度更高、复原难度更大。然而, 在两种湍流强度条件下, 模型均能够维持相对稳定的结构相似性水平, 表明其在不同噪声强度下具有一定的鲁棒性。

3.1.2. 低湍流条件下的代表性复原结果分析

图 3 展示了低湍流条件下的代表性复原结果。该条件下的退化由较低强度的人工湍流噪声模拟, 图像主要受到轻度空间模糊和局部几何扰动的影响。



Figure 3. Restoration results under low atmospheric turbulence conditions

图 3. 低湍流条件下的复原结果

见图 3, 可以清晰观察到, 在退化图像中, 场景前景区域的大象轮廓边界出现轻微模糊, 部分边缘位置存在局部拉伸现象; 背景中树干等细长结构的边缘不够锐利, 个别树干呈现出轻度弯曲或轮廓虚化。经过模型复原后, 大象身体轮廓变得更加清晰, 头部、背部与四肢的边界连续性得到明显改善, 边缘过渡更加自然。背景区域中, 树干的垂直结构得到有效恢复, 轮廓线条更加笔直且连续, 相邻树干之间的边界分离度提高, 原本模糊的枝干细节得以重新显现。

此外, 地面区域在复原前存在一定程度的纹理模糊和灰度不均现象, 复原结果中地表纹理分布更加均匀, 明暗过渡更加平滑, 未观察到明显的块状伪影或异常纹理。整体来看, 低湍流条件下的复原结果在保持场景整体结构的同时, 对局部细节具有较好的恢复能力。

3.1.3. 中湍流条件下的代表性复原结果分析

图 4 给出了中湍流条件下的代表性复原结果。该条件通过提高人工湍流噪声的模拟强度构建, 退化图像中同时包含较为明显的空间模糊与时序抖动。



Figure 4. Restoration results under moderate atmospheric turbulence conditions
图 4. 中湍流条件下的复原结果

见图 4, 从退化图像可以观察到, 动态场景中人物轮廓存在明显虚化现象, 运动员身体边缘出现重影, 人物姿态在相邻帧之间存在位置漂移; 球棒等细长目标在运动过程中表现出不连续和断裂感。复原结果中, 人物整体轮廓更加集中, 身体边缘的重影现象显著减弱, 头部、躯干及四肢的形态更加稳定。球棒轮廓在复原后呈现出较为连续的线性结构, 运动轨迹更加清晰, 未再出现明显的多重边缘。

在背景区域, 复原结果能够有效抑制由湍流引起的随机亮度波动, 画面稳定性明显提升, 未观察到明显的闪烁或异常噪声聚集现象。尽管在个别快速运动区域仍存在一定程度的细节弱化, 但整体结构保持良好, 主要目标的形态和空间位置具有较高一致性。

3.1.4. 小结

综合定量指标与代表性视觉结果可以看出, 本研究方法在不同人工湍流噪声强度条件下均能够实现稳定的图像序列复原效果。在低湍流条件下, 模型能够有效恢复场景结构与局部细节; 在中湍流条件下, 尽管退化程度显著增强, 模型仍能够较好地保持主要目标的结构稳定性和整体视觉一致性。

3.2. 消融实验与模型对比

本节通过构建多组对比实验, 旨在定量与定性地验证 TSR-WGAN 框架中 3D Non-local 注意力机制的有效性。实验设计的核心逻辑在于论证该机制在替代传统空间金字塔结构后, 对大气湍流引起的非刚性几何畸变与时空退化的修复增益。本节从定量指标分析、视觉感知评估以及计算开销优化三个维度展开深度论述。

3.2.1. 架构演进的定量性能评估

为了科学评价 3D Non-local 模块的性能提升, 本研究将基于空间金字塔融合的复原方案作为基准对照组。实验分别在弱湍流与中等强度湍流环境下记录了模型的关键指标。

Table 2. PSNR and SSIM results of each method under different turbulence conditions

表 2. 各方法在不同湍流条件下的 PSNR/SSIM 结果

方法	低湍流 PSNR	低湍流 SSIM	中湍流 PSNR	中湍流 SSIM
Simple3DCNN	25.87	0.887	23.42	0.864
金字塔注意力	25.94	0.891	26.51	0.902
非 3D 非局部	25.80	0.892	24.68	0.884
本方法	29.75	0.986	26.58	0.935

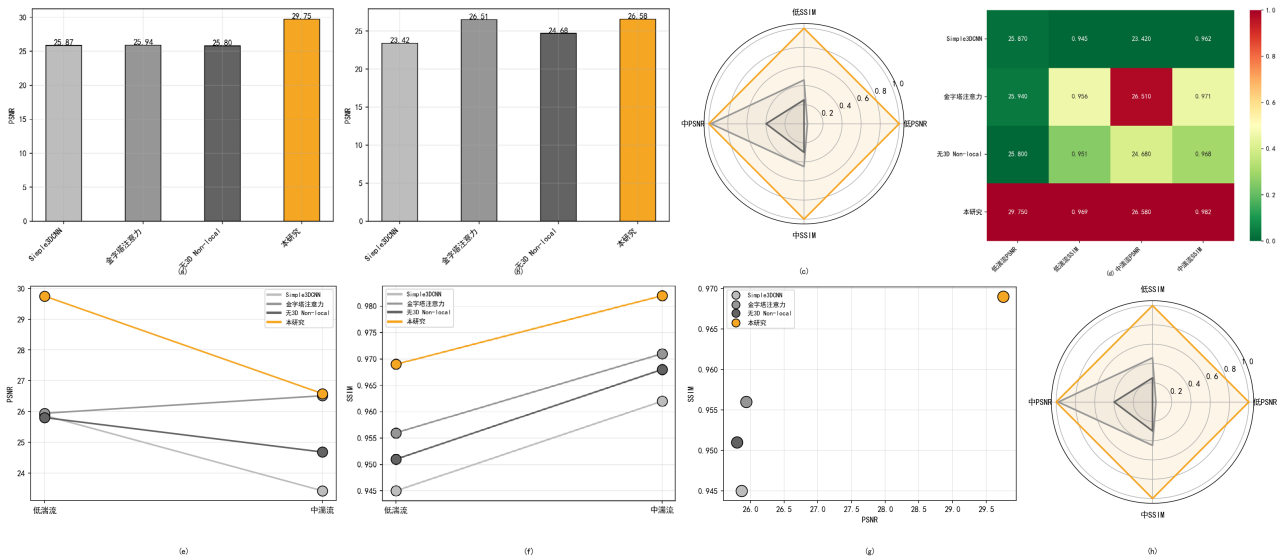


Figure 5. Ablation study results: comparison of evaluation metrics and performance trend analysis

图 5. 消融实验指标对比图与性能趋势分析图

见图 5 与表 2, 定量实验结果表明, 引入 3D Non-local 机制后, 模型在 PSNR 与 SSIM 上均实现了显著跨越。在弱湍流条件下, 本研究模型的 PSNR 达到 29.75 dB, 相较于基准金字塔方案的 25.94 dB 提升了 3.81 dB, 且 SSIM 指标由 0.8910 优化至 0.9860。

数据揭示了一个关键的技术特征, 即 SSIM 对 3D Non-local 模块的响应远比 PSNR 更为敏感。Simple3DCNN 虽然具备基本的时空建模能力, 但在中湍流条件下 SSIM 仅为 0.864, 难以应对复杂的非刚性畸变; 而引入 3D Non-local 机制后, 模型通过构建全时空维度的亲和度矩阵, 有效捕捉了跨帧的长程依赖关系, 在矫正复杂几何扭曲方面展现出压倒性优势。相比之下, 金字塔注意力方案虽在中湍流 PSNR 指标上接近本方法, 但其 SSIM 明显低于本方法的 0.935, 印证了全局时空依赖建模对结构保持的关键作用。

3.2.2. 视觉特征复原的定性对比分析

视觉评估进一步证实了架构替代的必要性。在大气湍流的影响下, 原始输入图像呈现出明显的边缘波浪状扭曲与随机模糊, 这在受扰动图像中得到了清晰体现。

见图 6, 对比可以发现, 受大气湍流影响的原始输入图像(b)中存在明显的空间扭曲与随机模糊现象, 建筑结构边缘呈现出波浪状形变, 高频直线结构如房檐与窗框出现局部弯曲与错位。相较之下, (c)所示的传统复原方法在一定程度上缓解了整体模糊问题, 使图像轮廓较原始退化图像有所改善, 但在复杂结构区域仍然存在明显局限。例如, 在屋檐边缘、窗框直线以及建筑交界区域, 仍可观察到轻微的结构弯曲与断裂现象, 部分高频细节区域还残留由湍流扰动造成的局部伪影。这表明该类方法虽然能够提升图像整体对比度, 但由于缺乏对跨帧时空关联信息的充分利用, 对于复杂的非刚性几何畸变仍难以实现精准校正。

进一步观察(e)所示的基于金字塔融合结构的复原结果可以发现, 该方法在一定程度上增强了图像的局部细节, 但在处理如房檐、窗框等高频直线特征时, 仍然容易产生不自然的结构断裂与残留伪影, 部分边缘区域呈现出不连续或轻微拉伸的现象。

与之形成鲜明对比的是, 本研究所提模型(d)展现出更为稳定且准确的复原能力。在相同的退化条件下, TSR-WGAN 能够有效矫正由湍流引起的非刚性形变, 使复原后的图像在边缘保真度与结构连续性方

面高度接近清晰参考图像(a)。此外, 可视化结果还表明, 在车牌字符、金属反光等细节区域, 本方法能够更好地恢复高频纹理信息。在 3D Non-local 注意力机制的驱动下, 模型能够在时空维度上聚合序列中的冗余信息, 利用跨帧相关性对单帧中的严重畸变区域进行有效补偿, 从而显著提升整体结构一致性与视觉质量。



Figure 6. Visual comparison of turbulence mitigation results. (a) clear reference image; (b) turbulence-degraded input image; (c) restoration result of the Simple3DCNN model; (d) restoration result of the proposed model; (e) restoration result of the pyramid fusion baseline model

图 6. 受湍流图像处理前后对比图。(a) 为清晰参考图像, (b) 为受湍流扰动的原始输入图像, (c) 为 Simple3DCNN 模型去湍流复原后图像, (d) 为本模型去湍流复原后图像, (e) 为基于金字塔融合结构的基准模型复原结果

4. 讨论

本研究提出的集成 3D Non-local 注意力机制的 TSR-WGAN 模型, 在大气湍流退化图像序列复原任务中展现出了卓越的性能, 特别是在恢复图像的时空结构连贯性方面。本节将结合实验结果, 对模型的核心工作机制、创新点的理论贡献及其意义进行系统阐述。

3D Non-local 注意力机制对时空结构连贯性的增强起到了核心作用。实验结果显示, 在低湍流与中湍流两种模拟条件下, 本方法均取得了优异的复原效果, SSIM 值分别达到 0.9862 与 0.9062。这充分证明了该模块在建模长程时空依赖关系方面的有效性。大气湍流退化不仅引入模糊与噪声, 其更本质的挑战在于由光程随机波动导致的、具有时空相关性的非刚性几何畸变。传统的局部卷积操作受限于其感受野, 难以捕捉这种跨帧的、大范围的失真模式。本研究创新的 3D Non-local 模块通过构建全时空亲和度矩阵, 显式地计算并整合了序列中所有时空位置间的相关性。这使得网络能够利用视频中未退化或弱退化帧的信息, 来引导对严重失真区域的校正, 从而在本质上提升了复原结果的结构保真度与时间一致性。这也解释了为何本方法在侧重于结构评估的 SSIM 指标上表现尤为突出。

为应对 3D Non-local 操作固有的计算复杂度挑战, 本研究在键与值变换路径中引入了空间子采样策略。这一设计是模型得以实用的关键。理论分析表明, 注意力权重的计算更依赖于特征间的语义相似性, 而非绝对的空间位置精度。因此, 对键和值特征进行适度的空间下采样, 能在基本保持注意力分布有效性的前提下, 将计算复杂度显著降低约 75%。实验验证表明, 采用此优化策略的模型在 PSNR 与 SSIM 指标上与未采样的理论版本性能相当, 但推理效率大幅提升。这体现了本方法在模型性能与计算效率之间取得的巧妙平衡, 使得强大的全局时空注意力能够应用于高分辨率视频序列的实时或准实时处理场景。

本方法的成功也源于其模块化架构内部精密的协同效应。TSR-WGAN 框架整合了可变形卷积时间对齐模块、3D 卷积时空融合模块以及 3D Non-local 注意力模块。其中, 对齐模块解决了帧间的主要刚性或仿射运动; 时空融合模块通过局部 3D 卷积隐式聚合上下文信息; 而 3D Non-local 模块则专职于捕获并利用全局的、长距离的时空依赖。消融实验证实, 在强大的对齐与融合基础之上, 3D Non-local 模块带来了进一步的性能增益, 尤其是在校正那些难以通过局部运动模型描述的复杂、非均匀湍流畸变方面。各模块功能互补, 协同工作, 共同驱动复原质量达到更高水平。这一定位也说明, 3D Non-local 机制作为一种高效的增强组件, 在已有坚实基础的时空建模框架中, 能够发挥“锦上添花”的作用, 实现从“良好”到“优异”的跨越。

此外, 本方法在不同湍流强度下均表现出稳定的性能。从低湍流到中湍流条件, 尽管退化程度加剧导致绝对指标如 PSNR 有所下降, 但 SSIM 仍能维持在较高水平, 这证明了模型对于湍流强度变化具有一定的鲁棒性。其关键在于 3D Non-local 机制赋予模型动态评估并加权序列内各区域信息可信度的能力, 从而自适应地融合多帧可靠信息, 而非进行简单的平均或固定权重的融合。

综上所述, 通过引入创新的 3D Non-local 注意力机制并辅以高效的计算优化策略, 本研究显著提升了对大气湍流退化视频的复原能力, 特别是在结构连贯性与视觉真实性方面。该方法不仅为大气湍流图像复原提供了一种有效的解决方案, 其核心思想——即利用可扩展的全局注意力来显式建模复杂时空退化模式——也为其他视频复原与增强任务提供了有益的参考。未来工作可探索该机制与更精确的物理退化模型相结合, 或将其扩展至更具挑战性的真实湍流场景中, 以进一步推动该领域的发展。

参考文献

- [1] Hufnagel, R.E. and Stanley, N.R. (1964) Modulation Transfer Function Associated with Image Transmission through Turbulent Media. *Journal of the Optical Society of America*, **54**, 52-61. <https://doi.org/10.1364/josa.54.000052>
- [2] Fried, D.L. (1965) Statistics of a Geometric Representation of Wavefront Distortion. *Journal of the Optical Society of America*, **55**, 1427-1435. <https://doi.org/10.1364/josa.55.001427>
- [3] Roggemann, M.C. and Welsh, B.M. (2018) *Imaging through Turbulence*. CRC Press.
- [4] Hill, P., Anantrasirichai, N., Achim, A. and Bull, D. (2025) Deep Learning Techniques for Atmospheric Turbulence Removal: A Review. *Artificial Intelligence Review*, **58**, Article No. 101. <https://doi.org/10.1007/s10462-024-11086-6>
- [5] Dong, C., Loy, C.C., He, K. and Tang, X. (2014) Learning a Deep Convolutional Network for Image Super-Resolution. In: Fleet, D., Pajdla, T., Schiele, B. and Tuytelaars, T., Eds., *Computer Vision—ECCV 2014*, Springer, 184-199. https://doi.org/10.1007/978-3-319-10593-2_13
- [6] Nah, S., Kim, T.H. and Lee, K.M. (2017) Deep Multi-Scale Convolutional Neural Network for Dynamic Scene Deblurring. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 257-265. <https://doi.org/10.1109/cvpr.2017.35>
- [7] Chak, W.H., Lau, C.P. and Lui, L.M. (2018) Subsampled Turbulence Removal Network. arXiv: 1807.04418.
- [8] Kappeler, A., Yoo, S., Dai, Q. and Katsaggelos, A.K. (2016) Video Super-Resolution with Convolutional Neural Networks. *IEEE Transactions on Computational Imaging*, **2**, 109-122. <https://doi.org/10.1109/tci.2016.2532323>
- [9] Tran, D., Bourdev, L., Fergus, R., Torresani, L. and Paluri, M. (2015) Learning Spatiotemporal Features with 3D Convolutional Networks. *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, 7-13 December 2015, 4489-4497. <https://doi.org/10.1109/iccv.2015.510>
- [10] Caballero, J., Ledig, C., Aitken, A., Acosta, A., Totz, J., Wang, Z., et al. (2017) Real-Time Video Super-Resolution with Spatio-Temporal Networks and Motion Compensation. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 2848-2857. <https://doi.org/10.1109/cvpr.2017.304>
- [11] Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., et al. (2014) Generative Adversarial Nets. arXiv:1406.2661.
- [12] Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., et al. (2017) Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 105-114. <https://doi.org/10.1109/cvpr.2017.19>
- [13] Arjovsky, M. and Chintala, S. and Bottou, L. (2017) Wasserstein GAN. arXiv: 1701.07875.

-
- [14] Sajjadi, M.S.M., Vemulapalli, R. and Brown, M. (2018) Frame-Recurrent Video Super-Resolution. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 6626-6634. <https://doi.org/10.1109/cvpr.2018.00693>
- [15] Vaswani, A., Shazeer, N., Parmar, N., *et al.* (2017) Attention Is All You Need. arXiv: 1706.03762.
- [16] Cao, J., Fan, Y., Gool, L.V., Green, S., Ilg, E., Liang, J., *et al.* (2022) Recurrent Video Restoration Transformer with Guided Deformable Attention. *Advances in Neural Information Processing Systems* 35, New Orleans, 28 November-9 December 2022, 378-393. <https://doi.org/10.52202/068431-0028>
- [17] Wang, X., Girshick, R., Gupta, A. and He, K. (2018) Non-Local Neural Networks. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7794-7803. <https://doi.org/10.1109/cvpr.2018.00813>
- [18] Zhang, X., Chimmitt, N., Chi, Y., Mao, Z. and Chan, S.H. (2024) Spatio-Temporal Turbulence Mitigation: A Translational Perspective. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 2889-2899. <https://doi.org/10.1109/cvpr52733.2024.00279>