

引用格式: 邓尔强, LOO Gowen, 朱国淞, 等. 基于改进光流的三维医学图像 Z 轴精度增强方法 [J]. 电子科技大学学报, 2025, 54(1): 91-102.
DENG E Q, LOO G, ZHU G S, et al. Enhanced Z-axis precision of 3D medical imaging via improved optical flow method[J]. Journal of University of Electronic Science and Technology of China, 2025, 54(1): 91-102.

基于改进光流的三维医学图像 Z 轴 精度增强方法



邓尔强, LOO Gowen, 朱国淞, 肖 鹏*

(电子科技大学 网络与数据安全四川省重点实验室, 成都 610054)

摘要: 三维医学图像在诊断和治疗中具有重要价值, 但由于切片结构导致纵轴方向分辨率较低, 给三维重建等任务带来一定挑战。针对三维医学图像在 z 轴方向上分辨率不足的问题, 提出了一种基于插帧技术的新方法。该方法通过生成中间切片来提升图像的连续性, 间接改善三维重建的质量。设计了一个紧凑型编码器-解码器网络, 融合中间光流估计和中间特征重建, 旨在克服传统图像处理方法中的模型失真和分辨率限制。此外, 该网络通过特征金字塔结构有效融合全局与局部信息, 并采用双重损失函数优化图像重建质量和特征空间的几何一致性。实验结果显示, 相比现有插帧方法, 该方法在图像质量、结构相似性、绝对误差和感知质量等方面实现了明显提升, 平均提高约 2.7% 到 5.8%。这些结果表明该插帧技术在脑肿瘤数据集上表现良好, 展示了在提升医学图像连续性和细节方面的潜在优势。该研究对提高临床诊断的准确性和治疗效果可以产生积极影响, 为医学图像处理领域的进一步研究提供了启示。

关键词: 三维医学图像; 图像插帧; 光流估计; 图像增强

中图分类号: TP183

文献标志码: A

DOI: 10.12178/1001-0548.2023292

Enhanced Z-axis precision of 3D medical imaging via improved optical flow method

DENG Erqiang, LOO Gowen, ZHU Guosong, and XIAO Peng*

(Network and Data Security Key Laboratory of Sichuan Province, University of Electronic Science and Technology of China, Chengdu 610054, China)

Abstract: Three-dimensional medical imaging plays a vital role in diagnosis and treatment. However, due to the structure of the slices, the resolution in the longitudinal direction is relatively low, which poses certain challenges for tasks such as three-dimensional reconstruction. This paper addresses the issue of insufficient resolution in the z-axis direction of three-dimensional medical images and proposes a novel method based on frame interpolation techniques. This method improves image continuity by generating intermediate slices, indirectly enhancing the quality of three-dimensional reconstruction. Aiming to overcome model distortion and resolution limitations inherent in traditional image processing methods, a compact encoder-decoder network is designed to integrate intermediate optical flow estimation and feature reconstruction. Furthermore, the network effectively merges global and local information through a feature pyramid structure and employs a dual loss function to optimize image reconstruction quality and the geometric consistency of the feature space. Experimental results demonstrate that, compared to existing frame interpolation methods, this approach significantly improves image quality, structural similarity, absolute error, and perceptual quality, with an average enhancement of approximately 2.7% to 5.8%. These results indicate that the frame interpolation technique proposed in this research performs well on brain tumor datasets, showcasing its potential advantages in enhancing the continuity and detail of medical images. This research offers a promising direction that could positively impact the accuracy of clinical diagnoses and treatment outcomes, providing new insights for further studies in the field of medical image processing.

Key words: three-dimensional image reconstruction; frame interpolation techniques; optical flow estimation; image enhancement

收稿日期: 2023-11-27

基金项目: 国家自然科学基金 (62372083)

作者简介: 邓尔强, 博士生, 主要从事深度学习、医学图像处理方面的研究。

*通信作者 E-mail: pongxiao@outlook.com

在现代医学诊断和治疗过程中,高质量的医学图像起着至关重要的作用。尤其在生成三维医学图像^[1]的过程中,原始二维图像的连续性和精确度对于确保最终三维重建的准确性和可靠性具有决定性影响。然而,传统的医学图像处理技术在处理 z 轴方向上稀疏分布的图像时,往往面临连续性不足的挑战。这种不足不仅影响了三维图像的整体质量,也可能导致在后续的诊断和治疗计划制定中出现误判。为了解决这一问题,本文提出了一种创新的插帧技术,专注于通过插入中间切片来改善医学图像的连续性,从而为生成更为准确的三维医学图像奠定基础。

本研究的动机源于提高医学图像处理技术的迫切需求,特别是在临床诊断和治疗决策中的应用。在肿瘤检测、心血管疾病评估和神经疾病诊断等关键领域,医学图像的连续性和精度直接关系到疾病的早期发现和治疗效果。如在脑肿瘤的诊断过程中,连续且详细的图像能够帮助医生更准确地定位肿瘤并评估其大小和形态,这对于制定手术计划或放疗方案具有重要意义。因此,本文采用了一种以插帧为基础的方法,将医学图像中相邻切片之间的补全视为一项关键任务。通过精确分析相邻切片,本方法预测并生成中间切片,以提高整体图像的连续性和细节表现。

本文的核心挑战在于如何精确地预测和重建这些中间切片。在这个过程中,光流技术发挥着至关重要的作用。光流技术能够描述逐像素的运动,从而帮助预测医学图像中相邻切片层间的组织结构变化^[2]。这一技术在计算机视觉领域已经得到了广泛应用,但在医学图像处理中仍面临许多挑战,如如何准确捕捉细小的解剖结构变化,以及如何处理由病变引起的异常组织运动。为了有效应对这些挑战,本文提出了一种集成了中间光流估计与特征重建的紧凑编码器-解码器网络结构。这种设计不仅简化了模型结构,还显著提高了处理效率,通过中间光流与特征的交互实现准确且迅速的中间切片补全。此外,考虑到医学图像的特殊性和复杂性,本研究特别强调全局和局部信息的融合,引入特征金字塔结构来捕捉图像中的细微变化和关键特征,这对于重建高质量的医学图像至关重要。

通过对医学图像数据集进行实验,验证了所提方法的有效性和优越性。实验结果表明,本方法在提升图像连续性、细节表现和整体质量方面均取得

了显著进步。特别是在处理具有复杂解剖结构的图像时,本方法展现出了更好的性能。此外,本文还探讨了该技术在不同临床应用场景中的潜在应用,如在神经外科手术规划和放射治疗中的应用前景,本研究为医学图像处理技术提供了一个有价值的探索。

1 相关工作

1.1 传统插帧技术

在医学图像处理的相关研究中,插值方法是一种基本技术。最邻近插值^[3]是最基础的插值方法之一,它通过将目标像素点的值设置为与其最近的源像素点的相同值来完成插值。尽管这种方法实现简单,但可能会导致插值结果出现阶梯效应,特别是在边缘区域。双线性插值^[4]则是利用4个最近邻点的加权平均来估计目标像素点的值,相比最邻近插值能够得到更平滑的插值效果。三次插值方法,如三次卷积插值^[5],进一步提高了插值的平滑度,通过考虑更多的邻域像素并使用三次多项式函数进行插值计算,使得重建图像在视觉上更为自然。

然而,这些传统方法通常不考虑图像中的结构信息,可能无法有效重建图像中的细节特征。为了克服这些限制,研究者开始探索基于模型的方法。在医学图像处理中,统计模型方法通过分析图像序列中的统计特征来估计和插入缺失的图像切片。这些方法,如自回归模型^[6]和隐马尔可夫模型^[7],依赖于图像数据的统计分布假设,从而预测未知切片的像素值。它们通过建立像素点之间的统计关系,利用已知的图像切片来推断缺失切片的内容。尽管统计模型能够在一定程度上反映图像的整体分布特征,但在处理高度复杂和非线性的医学图像时,可能难以准确捕捉到细节变化,尤其是在图像中存在大量细微结构和重要生理信息时。

为了提高医学图像插帧的准确性和质量,研究者开始探索更先进的方法,如基于深度学习的光流估计。光流法可以精确地模拟和重建图像间的动态变化,而深度学习则能够处理图像的复杂性和非线性特性,从而提高插帧的效果。光流法和深度学习的结合具有显著优势。深度学习可以通过学习大量的训练数据来捕捉更复杂的纹理和结构细节,从而提高光流估计的准确性,降低估计的偏差。此外,深度学习方法,如卷积神经网络(CNN)^[8],可以从数据中学习复杂的特征表示,从而更好地处理复杂和非线性的运动。

1.2 基于光流的插帧方法

视频帧插值 (video frame interpolation, VFI) 技术, 如文献 [9] 所述, 是通过在两个连续视频帧之间生成不存在的帧来提高视频流畅性的技术, 推动了利用光流来估计和合成中间帧的方法, 这种方法在动态视频场景中特别有用。

文献 [10] 使用深度学习技术对连续帧之间的中间帧进行插值, 以提高视频的时间分辨率, 这对于慢动作效果制作至关重要。此外, 文献 [11] 探索了一种新的 VFI 技术, 该技术使用自适应分离卷积网络来更精确地估计运动, 实验表明, 这种方法在保持视频质量的同时提升了插值的效率。在提高医学图像分辨率方面, 文献 [12] 显示, 基于光流的方法通过精确的运动估计和中间帧合成, 可以显著提升三维医学图像的质量。

随着这些方法的不断发展, 运动迭代特征学习的帧插值方法, 在提供平滑视觉体验方面的潜力日益显现^[13]。文献 [14] 突破了传统统计直觉, 即使在嘈杂的训练数据上也能保持良好的测试性能。该方法着重于利用深度学习来增强时序一致性, 这不仅提升了帧插值的准确性, 还增加了生成视频帧的时间连贯性。文献 [15] 提出了全对多场变换, 通过构建所有像素对的双向相关性体积和从粗流中导出的多组细粒度流场, 能够有效处理大运动和遮挡问题, 实现在多个基准测试中的先进性能。

目前, 基于光流的中间帧预测主要遵循两条主流思路: 1) 通过中间光流网络直接估计输入的两帧之间的中间光流, 进而指导输入帧像素通过中间帧合成网络映射以合成中间帧; 2) 估计输入帧之间的双向光流, 并通过编码器以及解码器网络近似或进一步细化以获得中间光流, 然后对输入帧及其特征进行编码, 并通过独立的合成网络合成中间帧纹理。然而, 尽管这两种方法已广泛应用于医学图像处理, 它们仍面临以下关键问题。

1) 现有基于光流的插帧算法通常将中间光流估计与中间帧特征合成分开到多个独立的编码器-解码器网络中, 导致中间帧与光流之间缺乏有效交互, 从而无法共同推断出准确的中间帧信息。

2) 这些方法采用多个编码器-解码器级联结构, 消耗大量的计算资源, 并显著延长推理时间, 造成了计算资源与时间上的冗余。

针对这些挑战, 本文提出了将原本分开的中间光流估计与中间特征重建过程整合到一个紧凑的编码器及解码器级联网络中。这种设计不仅简化了模

型结构, 而且显著提高了处理效率, 有效地利用了中间光流与中间切片特征之间的互动, 实现了准确且快速的中间切片补全。

2 方法

2.1 光流法在医学图像插帧中的适用性

光流法在视频插帧中的应用已经广泛且效果显著, 但在医学图像插帧的应用中面临着不同的挑战。视频插帧主要处理的是在二维平面上运动的物体, 如行驶的汽车和运动的人物, 其运动主要表现为平面内的平移和旋转。相比之下, 医学图像插帧涉及人体内部的组织结构, 如肿瘤、血管等, 这些结构在相邻切片间的变化包括形状、大小和位置的多维度变化, 为插帧过程带来了额外的复杂性。

要在医学图像插帧中应用光流法, 首先需要证明其在处理这两种不同类型图像时具有等价的适用性。在视频插帧问题中, 光流法通过考虑不同帧之间同一物体的像素集合间的映射关系来估计运动信息。以物体在两个相邻帧中的像素集合为例, 定义 A 与 B 如下:

$$\begin{aligned} A &= (x_i, y_i, t) \quad i = 1, 2, \dots, n \\ B &= (x_j, y_j, t + dt) \quad j = 1, 2, \dots, m \end{aligned} \quad (1)$$

式中, 时间 dt 差较小, 因此集合 A 与 B 在拓扑结构和亮度上具有相似性。通过求解映射 $f: A \rightarrow B$, 可得到插帧的解。

在相邻帧之间的像素运动通过以下方程表达:

$$I(x, y, t) = I(x + dx, y + dy, t + dt) \quad (2)$$

式中, I 是图像亮度函数, 描述了在特定时间 t 和空间位置 (x, y) 的像素亮度。假设物体的像素亮度值在 x 轴方向上发生 dx 的位移, 在 y 轴方向上发生 dy 的位移, 而时间轴方向上发生 dt 的变化时, 亮度值保持不变。这一假设是光流法估计物体在相邻帧之间像素运动的基础。

在医学图像插帧中, 如果考虑两个相邻切片的深度间隔较小, 有 1 个组织结构在第 1 个切片上由像素集合 A' 表示, 在第 2 个切片上由像素集合 B' 表示。由于这 2 个像素集合代表同一组织结构的连续部分, 它们在拓扑结构上具有相似性。因此, 可以定义一个映射关系 $f: A' \rightarrow B'$, 并在插帧操作中求解这个映射关系, 如下所示:

$$\begin{aligned} A' &= (x'_i, y'_i, z) \quad i = 1, 2, \dots, n \\ B' &= (x'_j, y'_j, z + dz) \quad j = 1, 2, \dots, m \end{aligned} \quad (3)$$

式中, dz 表示相邻切片间的深度差异。医学图像的像素集合亮度分布变化可以用光流法来捕捉和表达, 从而反映出组织结构在不同切片层的形状变化。

最终, 医学图像插帧问题可通过以下光流连续性方程来表达, 该方程假设在小的深度变化 dz 下, 组织结构的亮度在三维空间中保持不变:

$$I(x', y', z) = I(x' + dx', y' + dy', z + dz) \quad (4)$$

在上述方程中, (x', y', z) 表示第一个切片上某一像素的三维空间坐标, $(x' + dx', y' + dy', z + dz)$ 表示第 2 个切片上对应像素的新坐标。光流向量 (dx', dy', dz) 代表了从一个切片到另一个切片中相同物理结构的映射。该映射反映了切片间相应组织结构的空间变化, 包括形状变化和位置偏移。

通过求解上述方程, 获得一组光流向量, 这组向量隐式定义了相邻切片间像素的对应关系。该对应关系使得网络能够通过已知的两个切片预测出未知的中间切片, 从而实现医学图像的高精度插帧。

因此, 尽管光流法在医学图像插帧中面临的挑战与视频插帧不同, 但通过适当的方法和技术调整, 可以实现有效的插帧效果, 并为后续的医学图像分析提供更精确的图像信息。

2.2 误差与优化目标

深度神经网络模型 f 被特别设计以整合光流法, 旨在实现精确的医学图像插帧任务。网络模型 f 通过光流技术预测并模拟两个相邻已知切片 I_{i-1} 和 I_{i+1} 之间的像素运动, 从而生成中间切片的估计。因此, \hat{I}_i 定义为:

$$\hat{I}_i = f(I_{i-1}, I_{i+1}; \phi_{\text{flow}}) \quad (5)$$

式中, ϕ_{flow} 代表从相邻切片中提取的光流信息, 通过神经网络自动学习获得。这一设计允许模型在预测 \hat{I}_i 时能直接和充分地利用切片之间的动态变化和运动信息。

通过光流技术, 可以更精确地捕获和模拟医学图像中的细微运动, 从而在插帧过程中减小预测误差。单个帧的预测误差 E_i 表达为:

$$E_i = \|\hat{I}_i - I_i^{\text{gt}}\|_2^2 \quad (6)$$

式中, \hat{I}_i 是基于光流法预测的中间切片; I_i^{gt} 是相应的真实图像; L2 范数量化了预测图像与真实图像之间的差异, 可以作为优化网络性能的清晰目标。

目标是在整个训练集上最小化所有预测帧的累积误差。因此, 提出的数学优化问题可以表示为:

$$\min \sum_i \|\hat{I}_i - I_i^{\text{gt}}\|_2^2 \quad (7)$$

该式旨在找到一个最优的模型参数配置, 使得预测的插帧 \hat{I}_i 与相应的真实帧 I_i^{gt} 之间的差异最小。此优化目标反映了利用光流法减少插帧过程中预测误差的核心目的, 进而优化模型性能并提高插帧结果的质量。

2.3 切片补全网络设计

为了实现优化目标, 本研究提出了一种专门为医学图像设计的切片补全网络。该网络旨在充分利用医学图像中的全局与局部信息。医学图像插帧要求不仅在局部细节上保持高度精确, 同时也要在全局结构上保持一致性。为此, 本文采用了在医学图像分类与分割任务中验证有效的特征金字塔结构, 以实现这一目标。

本网络的结构设计如图 1 所示, 首先利用一个多层编码器从输入图像中提取金字塔特征。该编码器由 4 个卷积层构成, 记为 ϕ_m^g , 其中 g 表示层的级别, m 表示图像帧索引, 其中 0 代表初始帧, 1 代表后续帧。每层之后接一个 ReLU 激活函数, 以增强网络对非线性特征的捕捉能力, 这种从粗糙到精细的特征提取方式, 确保了全局和局部信息的有效捕获。

接下来, 网络在各个尺度上应用光流预测算法, 计算双向光流, 为相邻帧之间的运动提供估计。每个尺度生成的两张特征图分别代表了前后帧的特征信息。在特征金字塔的最底层, 最小尺度的特征图与模拟参数 t 一同输入到解码器 D^g 中, 其中 t 的取值范围是 0~1, 代表中间切片相对于前一帧的位置。

解码器的主要任务是捕获两张特征图之间的像素运动, 并生成代表从 t 到 0 和从 1 到 1 的双向光流图 F_i^g 。解码器同样负责输出一个特征图, 该特征图描绘了预测的中间帧在时间点 t 的状态。这一输出通过应用翘曲场 $\hat{\phi}_i^g$ 实现, 利用光流图指示的向量, 像素位置被翘曲至新的预测位置。为了补偿在下采样过程中可能丢失的信息, 双向光流图会被上采样至与上一层相同的尺度, 并和相应尺度的前后帧特征图结合。此外, 这些双向光流图还将通过另一个翘曲场 $\hat{\phi}_m^g$ 进行细化处理, 以确保最终像素位置的精确对齐。

在整个网络结构中, 通过逐层优化后的中间切片特征图, 网络能够更准确地预测组织分布, 同时

大幅减少伪影。这一迭代过程一直持续, 直到解码器输出的光流图与原始输入图像的分辨率匹配。最后, 通过合成网络将金字塔每一层的特征融合起

来, 生成在 t 位置的中间切片纹理特征, 作为网络的输出结果。

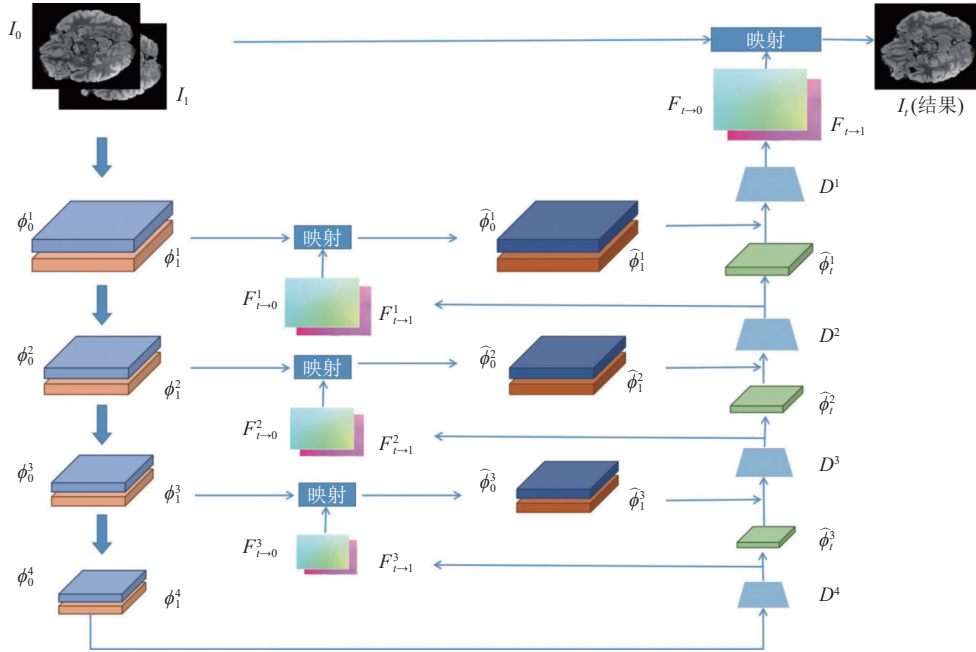


图 1 系统架构

这样的网络结构不仅能够提供中间光流的精确预测, 从而帮助精细重建中间帧特征, 同时, 优化后的中间帧特征也提供了锚点信息, 进一步增强了光流估计的精度。这种相互促进的机制显著提高了中间切片预测的准确性和效率。

2.4 损失函数的设计

为了确保网络结构的有效性和对中间切片的精确预测, 设计了一种综合损失函数来指导中间切片补全网络的训练。该损失函数考虑了图像重建质量和特征空间几何一致性两个方面。

图像重建损失是针对中间切片预测的主要损失函数, 其目标是 minimized 网络生成的切片与实际切片之间的差异, 表达如下:

$$L_r = \rho(\hat{I}_t - I_t^{\text{gt}}) + L_{\text{cen}}(\hat{I}_t, I_t^{\text{gt}}) \quad (8)$$

式中, \hat{I}_t 是网络输出的中间切片图像; I_t^{gt} 是对应的真实切片图像; ρ 是像素级损失如均方误差或绝对误差; L_{cen} 是中心损失函数, 定义为:

$$L_{\text{cen}}(\hat{I}, I^{\text{gt}}) = \sum_{p \in C} \|\hat{I}(p) - I^{\text{gt}}(p)\|_2^2 \quad (9)$$

式中, C 表示图像的中心区域; p 是该区域内的像素位置。这种设计旨在提升模型在医学图像中心区域 (通常包含关键诊断信息) 的预测性能。

特征空间几何一致性损失旨在保证重建的中间切片在特征层面上与真实切片保持一致, 可以用下式表示:

$$L_g = \sum_{k=1}^3 L_{\text{cen}}(\hat{\phi}_t^k, \phi_t^k) \quad (10)$$

式中, $\hat{\phi}_t^k$ 和 ϕ_t^k 分别代表在不同尺度 k 上网络预测的特征图与真实切片的特征图, 该损失函数强化了特征层次上的一致性, 有助于模型学习预测更加准确的细节和结构。

整体损失是上述两种损失的组合, 表示为:

$$L = L_r + \lambda L_g \quad (11)$$

式中, 超参数 λ 控制图像重建损失和特征空间几何一致性损失之间的平衡, 需要根据具体应用场景调整。这种损失函数的设计有利于在保持全局图像重建质量的同时, 特别强化中心区域和特征层面的准确性, 对于高精度医学图像处理任务至关重要。

3 实验与分析

3.1 实验环境

3.1.1 数据集

本研究使用了 BraTS2021^[16] 数据集, 这是一个多机构、多参数多模态核磁共振成像 (MRI) 数据

集, 包括训练集 (1 251 例)、验证集 (219 例) 和测试集 (530 例), 共计 2 000 例患者的 MRI 扫描结果。数据集包含 4 种模态数据: FLAIR、T1、T1ce 和 T2, 每个模态的数据大小为 $240 \times 240 \times 155$ 。重点关注 gliomas, 这是成人最常见的原发性脑肿瘤。该数据集的目标是自动分割脑肿瘤结构, 从而改进诊断、治疗计划和个体患者的随访。

3.1.2 模型训练

本研究对 BraTS2021 数据集进行了预处理操作: 去除图像中的噪声, 增强数据集的多样性 (通过随机旋转、平移和翻转等数据增强技术), 将数据归一化至 $[0, 1]$ 范围, 并将原始图像切割为固定大小的子图 (如 $64 \times 64 \times 64$) 以适应模型的输入。将 BraTS2021 数据集划分为训练集 (1 000 例)、验证集 (250 例) 和测试集 (750 例)。使用 Adam 优化算法, 学习率设置为 0.001, beta1 设置为 0.9, beta2 设置为 0.999。使用了批大小为 16 进行训练, 总共进行了 100 个 epoch。调整了学习率、批大小和模型深度等超参数来确定合适的参数。验证集上尝试了学习率从 0.000 1~0.01 的不同值, 最终选择了 0.001; 尝试了批大小分别为 8、16 和 32, 最终选择 16。

在本研究中, 超参数的选取经过了多次实验, 以确保模型的性能最优化。特别是在模型训练部分, 选择的超参数主要是学习率, 该参数对模型的训练效果和最终性能有着直接影响。

因此, 针对 T1 模态数据的学习率进行了一系列实验, 以确定其对模型性能的影响。如图 2 所示, 选择从 0.000 1~0.01 不等的学习率, 并观察模型在验证集上的表现。结果显示, 学习率为 0.001 时, 模型达到了最佳的平衡状态, 既能保证足够快的收敛速度, 又能避免过快收敛导致的过拟合问题。

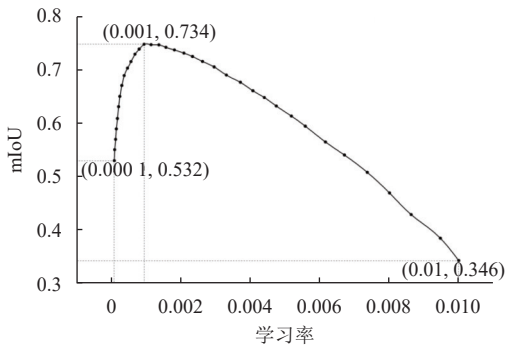


图 2 学习率对实验结果的影响

3.1.3 软硬件环境

本研究的实验在具有 NVIDIA Tesla V100S GPU

的高性能计算机上进行。操作系统采用 Ubuntu 22.04, 配套深度学习框架 PyTorch 1.12 版本, 来实现和训练本模型。所有实验结果均在相同的硬件和软件环境下获得, 以确保公平的比较。

3.1.4 评估指标

本研究采用了 PSNR^[17]、SSIM^[18]、MAE^[19] 和 UQI^[20] 这 4 种评估指标, 因为它们能够全面反映插帧方法在图像质量、结构相似性、绝对误差和感知质量方面的性能。

峰值信噪比 (PSNR) 通过计算原始图像与重建图像之间的均方误差 (MSE) 来评估图像质量, 较高的 PSNR 值表示重建图像与原始图像间的误差较小, 质量更优。如下所示:

$$\text{MSE} = \frac{1}{WH} \sum_{x=1}^W \sum_{y=1}^H |I(x,y) - K(x,y)|^2 \quad (12)$$

$$\text{PSNR} = 10 \log_{10} \left(\frac{\text{MAX}_I^2}{\text{MSE}} \right) \quad (13)$$

式中, $I(x,y)$ 表示原始图像在位置 (x,y) 的像素值; $K(x,y)$ 表示重建图像在位置 (x,y) 的像素值; W 和 H 分别表示图像的宽度和高度; MAX_I 表示图像的最大像素值 (如对于 8 位深度的图像, $\text{MAX}_I = 255$)。

结构相似性指数 (SSIM) 用于衡量两幅图像的结构、对比度和亮度相似性, 如下所示:

$$\text{SSIM}(I, K) = \frac{(2\mu_I\mu_K + C_1)(2\sigma_{IK} + C_2)}{(\mu_I^2 + \mu_K^2 + C_1)(\sigma_I^2 + \sigma_K^2 + C_2)} \quad (14)$$

式中, μ_I 和 μ_K 分别表示原始图像 I 和重建图像 K 的均值; σ_I^2 和 σ_K^2 分别表示原始图像 I 和重建图像 K 的方差; σ_{IK} 表示原始图像 I 和重建图像 K 的协方差; C_1 和 C_2 是常数, 用于避免分母为零的情况。

平均绝对误差 (MAE) 用于衡量原始图像与重建图像之间的平均绝对误差, 如下所示:

$$\text{MAE} = \frac{1}{WH} \sum_{x=1}^W \sum_{y=1}^H |I(x,y) - K(x,y)| \quad (15)$$

通用图像质量指数 (UQI) 考虑了图像的相关性、对比度和亮度差异, 如下所示:

$$\text{UQI} = \frac{4\sigma_{IK}\mu_I\mu_K}{(\sigma_I^2 + \sigma_K^2)(\mu_I^2 + \mu_K^2)} \quad (16)$$

式中, μ_I 和 μ_K 分别表示原始图像 I 和重建图像 K 的均值; σ_I^2 和 σ_K^2 分别表示原始图像 I 和重建图像 K 的方差; σ_{IK} 表示原始图像 I 和重建图像 K 的协方差; UQI 的值范围为 $[0, 1]$, 其中 1 表示两幅图像完全

相同, 而 0 表示两幅图像完全不相关。

3.2 消融实验

3.2.1 光流法消融实验

本研究通过消融实验对比了单向光流法和双向光流法在医学图像插帧任务上的性能。如图 3 所示, 从左至右展示了原始图像、单向光流法插帧结果、双向光流法插帧结果。

对比结果揭示了单向光流法在医学图像插帧过程中的特定局限性。单向光流法倾向于依赖单一方向的特征, 这虽然不会导致明显的模糊, 但却会使得重建图像在结构和纹理上与原始方向高度一致,

缺乏另一方向的特征继承。结果是, 虽然单向光流生成的图像在某些方面与原始图像保持了连贯性, 却在另一方向的特征表现上失去了复原的多样性和准确性。

相对而言, 双向光流法有效地综合了来自两个方向的特征信息, 实现了特征的均衡融合。这种方法不仅提高了结构和纹理的细节复原, 也使得生成的中间帧在局部细节上与原始图像更为接近, 无论是在结构上还是纹理上。这对于医学图像处理尤其重要, 因为准确的结构和纹理信息对于疾病的诊断和评估至关重要。

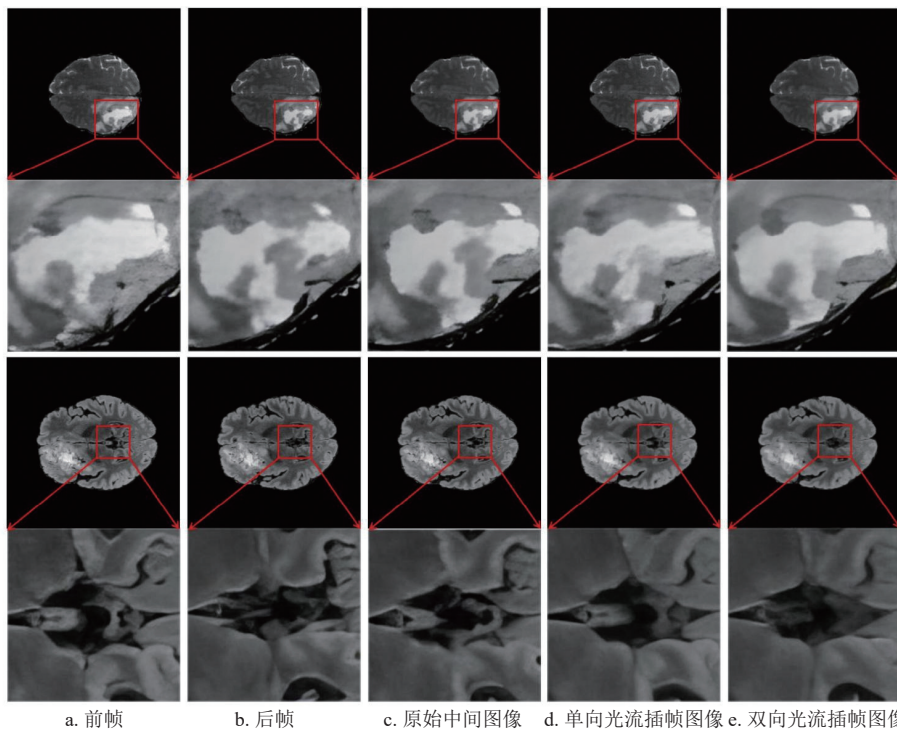


图 3 单向光流与双向光流效果对比

在进行了视觉对比后, 定量分析也支持了这些观察结果。双向光流法在结构相似性指数 (SSIM)、峰值信噪比 (PSNR) 以及绝对误差 (MAE) 等关键性能指标上均优于单向光流法。如表 1 所示, 双向光流法在医学图像插帧任务上提供了更准确的中间帧预测, 从而为后续的医学分析和诊断提供了更高质量的图像基础。

表 1 单向光流法与双向光流法对比

方法	PNSR/dB	SSIM	MAE	UQI
单向光流法	31.73	0.82	5.54	0.87
双向光流法	32.10	0.89	5.01	0.90

3.2.2 损失函数超参数实验

为了评估损失函数超参数 λ 在中间切片预测质

量上的影响, 进行了一系列的实验。超参数 λ 控制着图像重建损失与特征空间几何一致性损失之间的权衡, 对于模型的性能至关重要。

如图 4 所示, 实验通过改变 λ 的值来观察绝对误差 (MAE) 的变化。可以看到, 当 $\lambda=0.3$ 时, MAE 达到最低, 表明在此配置下模型取得了最佳的预测性能。当 λ 远离此值时, MAE 逐渐增加, 说明过小或过大的 λ 值都会降低模型的预测精度。特别地, λ 意味着仅依赖图像重建损失, 而忽略了特征空间的几何一致性, 导致性能下降; 而过高的 λ 值则过分强调特征层面的一致性, 而忽视了像素级的重建质量。

通过实验, 确定了适中的 λ 值对于平衡网络在

像素级重建和高级特征表示方面的性能至关重要，从而在医学图像插帧任务中取得优异结果。

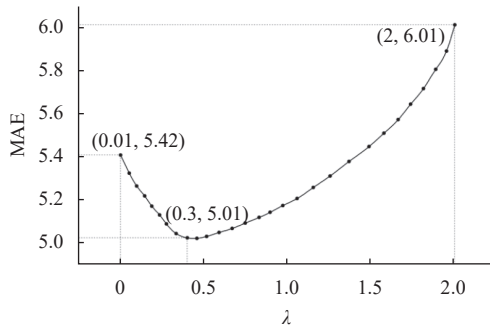


图 4 λ 对 MAE 的影响

3.3 其他方法对比

实验中选择了 CAIN^[21]、CDFI^[22]、EDSC^[23]、RIFE^[24] 和 AMT^[15] 作为对比模型，因为它们插帧任务中具有较好的性能和广泛的应用。

3.3.1 可视化对比

在本研究的可视化对比实验中，主要关注了坏死肿瘤核心、周围水肿区域和增强肿瘤区域这 3 个区域，因为这些区域对于脑肿瘤的诊断和治疗具有

重要意义。选取了具有代表性的灰度图像，并将不同插帧方法在这些区域生成的中间帧进行了可视化对比，以直观展示各种方法的性能差异。

如图 5 所示，本文提出的方法在恢复细节、保持边缘锐化以及运动估计方面表现出色。与其他方法相比，该方法在处理坏死肿瘤核心、周围水肿区域和增强肿瘤区域的图像时能够更好地保持原始图像的结构和质量。此外，该方法在图像纹理和对比度方面也表现出优异的性能，能够更准确地还原原始图像的灰度分布。在某些情况下，其他方法可能会导致模糊、重影或伪影等问题。这主要是由于这些方法在进行运动估计和插帧时未能准确地捕捉到肿瘤结构的信息。相比之下，本文方法更能准确地捕捉关键信息，从而避免上述问题。

通过对坏死肿瘤核心、周围水肿区域和增强肿瘤区域中生成的中间帧进行可视化对比实验，结果表明本文方法在恢复细节和图像质量方面相较于其他方法有明显改进。这些发现支持了该方法在脑肿瘤诊断和治疗中图像插帧任务的有效性和鲁棒性，为未来的研究提供了一个潜在的高质量解决方案。

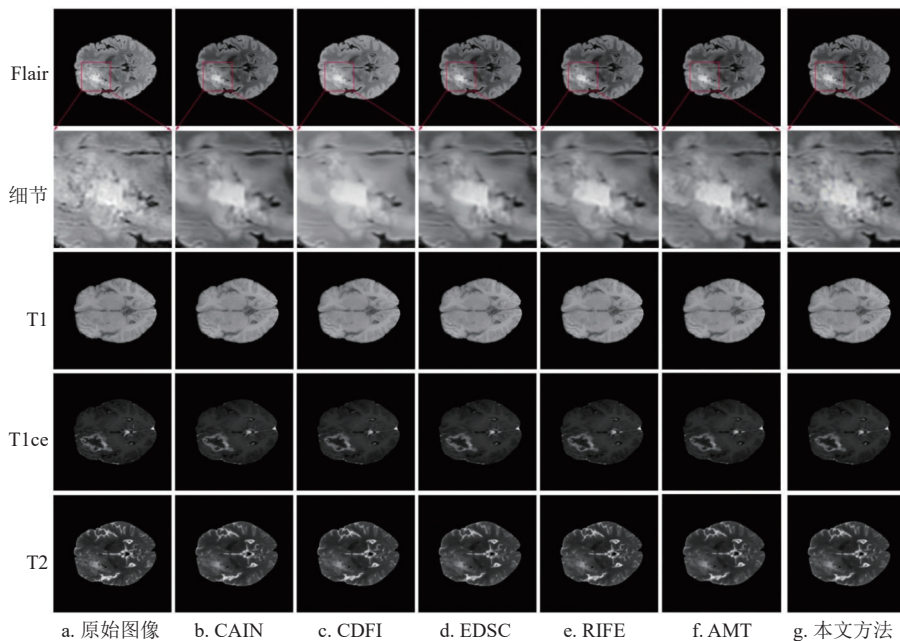


图 5 可视化效果对比

3.3.2 量化数据对比

从表 2 中，可以看出本研究的方法在 PSNR、SSIM、MAE 和 UQI 这 4 个评价指标上均优于 CAIN、CDFI、EDSC 和 RIFE，这表明该方法在图像质量、结构相似性、绝对误差和感知质量等方面的性能优于其他方法。

尤其是在 PSNR 和 SSIM 上，本文方法相较于其他方法有显著提升，这一结果表明，所提方法能有效生成与原始图像在视觉上更为接近的中间帧。此外，在 MAE 和 UQI 的评估中，该方法同样实现了较低的误差和更优的感知质量。

通过这些定量实验结果，可以得知本研究提出

的插帧方法在脑肿瘤数据集上具有较好的性能, 为脑肿瘤诊断和治疗提供了更高质量的中间帧。

表 2 各插帧方法性能对比

方法	PNSR/dB	SSIM	MAE	UQI
CAIN	31.25	0.87	5.32	0.88
CDFI	30.78	0.86	5.45	0.87
EDSC	30.43	0.85	5.68	0.86
RIFE	29.97	0.83	5.91	0.85
AMT	31.96	0.87	5.12	0.87
本文方法	32.10	0.89	5.01	0.90

3.4 分割性能对比

本研究选择了分割任务作为评估指标, 由于分割精度直接影响脑肿瘤的诊断和治疗效果。分割任务的目的是将图像中的像素分类到正确的区域, 如脑组织、肿瘤核心或肿瘤边缘。本实验分为两个部分: 1) 在 T1 模态下对不同插帧方法的详细性能进行比较; 2) 评估多模态下各插帧方法的平均分割性能。

分割性能的评价主要依赖于以下两个关键指标: Dice 系数和交并比 (IoU)。Dice 系数, 也称为 Dice 相似系数 (DSC), 是衡量两个样本相似性的统计工具, 定义为两倍的样本交集除以样本总数。在医学图像分割中, 通常用来衡量分割区域与真实标签的重叠程度。Dice 系数表示为:

$$\text{Dice} = \frac{2|X \cap Y|}{|X| + |Y|} \quad (17)$$

式中, X 表示预测的分割区域; Y 表示真实的标签区域。

交并比 (IoU), 是另一个衡量样本重叠度的

指标, 定义为样本交集除以样本并集。IoU 是分割任务中常用的性能指标, 尤其是在评估对象检测和实例分割算法时。其数学表达式为:

$$\text{IoU} = \frac{|X \cap Y|}{|X \cup Y|} \quad (18)$$

为了确保实验的准确性和可靠性, 选用了当前医学图像分割领域广泛应用且效果卓越的基线模型, 以不同插帧方法生成的中间帧为输入, 进行脑肿瘤分割。通过这些评价指标, 能够全面地衡量分割模型在医学图像中的性能。

3.4.1 T1 模态详细性能对比

在 T1 模态的分割性能对比中, 考察了使用 U-Net^[25]、SegNet^[26] 和 U-Net++^[27] 网络时, 不同插帧方法的分割效果。评估指标选择了均值 Dice 系数 (mDice), 该指标衡量了分割质量的整体性能, 并对每个类别背景 (BG)、非肿瘤 (NT)、肿瘤核心 (ED) 和增强肿瘤 (ET) 的表现进行了分析。如表 3 所示, 不同方法的性能有显著差异, 其中本文方法在所有网络架构中都显示出了较为优越的性能。

特别是在 U-Net 和 U-Net++ 这两个模型中, 本文方法相较于其他比较方法, 如 CAIN、CDFI、EDSC 和 RIFE, 显示出更高的 mDice 值, 这说明了其在分割肿瘤区域时的准确性更高。本文方法在肿瘤核心 (ED) 和增强肿瘤 (ET) 的分割上尤其表现良好, mDice 分别达到了 78.9% 和 77.6%, 而在其他网络架构上的表现也有明显优势, mDice 均超过了 80%。这些结果证实了本文方法在保留医学图像细节和提高分割精度方面的有效性。

表 3 各种方法在不同分割网络下的 mDice 对比 (T1 模态)

	SegNet					U-Net					UNet++					%
	整体性能	BG	NT	ED	ET	整体性能	BG	NT	ED	ET	整体性能	BG	NT	ED	ET	
CAIN	76.2	88.1	70.3	74.5	72.0	78.5	89.7	71.6	76.1	73.6	80.2	91.2	73.0	77.6	75.1	
CDFI	74.5	87.2	68.0	72.6	70.4	76.9	88.8	69.3	74.1	71.8	78.7	90.6	71.0	76.0	73.2	
EDSC	72.9	86.0	66.2	71.0	68.6	75.3	87.6	67.8	72.3	70.2	77.1	89.8	69.1	74.6	71.8	
RIFE	71.0	85.1	64.3	69.3	66.6	73.4	86.4	66.1	70.6	68.5	75.6	88.7	67.2	72.9	69.5	
AMT	77.4	89.3	70.8	76.3	73.2	79.0	90.1	72.4	77.8	75.6	80.2	91.4	74.0	78.5	76.9	
本文方法	80.1	90.5	73.5	78.9	77.6	82.3	92.0	74.7	80.3	79.1	84.2	93.7	75.8	82.0	81.2	

3.4.2 多种模态平均性能对比

跨模态的分割性能对比旨在评估不同插帧方法在各种医学成像模态中的泛化能力。在本实验中, 利用 Swin-UNet^[28] 分割网络, 比较了 CAIN、CDFI、EDSC、RIFE、AMT 以及本文方法在 T1、T2、T1CE 和 Flair 这 4 种模态下的性能。这些模态涵盖

了多种 MRI 成像条件, 反映了不同组织类型的对比度和结构细节。

在本实验的视觉分析部分, 对不同插帧方法在多模态医学图像分割任务中的性能进行了直观比较。如图 6 所示, 每行代表一种独特的成像模态, 包括 T1、T2、T1CE 和 Flair。对于每种模态, 从

左至右依次展示了原始图像、原始标签，以及使用 RIFE、EDSC、CDFI、CAIN、AMT 和本文方法生成的中间帧的分割结果。可以观察到，相较于

其他方法，本文方法生成的中间帧在边缘细节和关键区域的分割上更为精确，尤其是在肿瘤核心和增强区域。

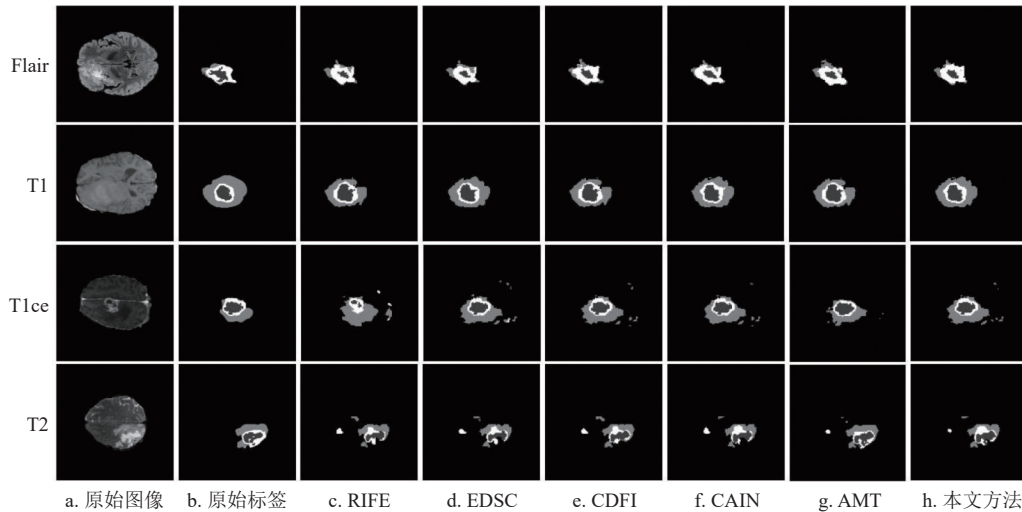


图 6 分割效果对比

实验的量化结果如表 4 所示，展示了在每种模式下，各种插帧方法的平均 Dice 系数 (mDice) 和平均交并比 (mIoU)。这两个评价指标均为分割

质量的重要度量，其中 mDice 强调了预测分割与真实标签的重叠程度，而 mIoU 评价了分割区域与实际标签的共同面积占比，更强调了预测准确性。

表 4 各插帧方法在不同模式下的分割性能对比

方法	T1		T2		TICE		Flair	
	mDice	mIoU	mDice	mIoU	mDice	mIoU	mDice	mIoU
CAIN	82.3	71.1	85.6	74.8	79.2	65.9	87.4	76.2
CDFI	80.1	68.9	84.4	73.5	77.5	63.7	86.0	75.1
EDSC	78.6	66.8	83.2	72.3	76.0	62.1	84.7	73.9
RIFE	76.9	64.7	81.7	70.8	74.3	60.4	83.2	72.6
AMT	82.6	71.6	86.2	75.3	79.6	88.0	87.9	76.8
本文方法	84.1	73.4	86.9	76.3	81.2	68.0	88.8	77.5

从表中可以看出，本文方法在所有模式中的表现均优于其他方法。特别是在 T1 模式下，本文方法的 mDice 达到了 84.1%，mIoU 达到了 73.4%；而在 Flair 模式中，mDice 高达 88.8%，mIoU 为 77.5%。这些结果说明，无论是对肿瘤的不同部分，还是在不同的成像条件下，本文方法都能提供更准确的分割结果。

3.5 实验分析

通过一系列综合实验，本研究证明了深度神经网络结合光流法在医学图像插帧任务中的有效性。在图像质量方面，所提出的方法在 PSNR 上达到了 32.10 dB，SSIM 为 0.89，这两个指标的提升分别反映了图像的清晰度和结构相似性得到显著改善。同时，MAE 降低到 5.01，UQI 提高到 0.90，

说明了本文方法在细节恢复和感知质量上的显著提升。这些指标的优异表现验证了本文方法在生成具有高度细节和准确边缘的中间帧方面的能力，特别是在捕捉复杂解剖结构及其变化方面显示出了对细微变化的敏感性，这在传统方法中往往难以实现。

进一步的量化评估显示，与现有的插帧方法相比，本文方法在提升图像连续性和准确性方面更为有效。在 T1 模式下的分割实验中，本文方法在 SegNet、U-Net、和 UNet++ 网络中的 mDice 分别达到了 80.1%、82.3% 和 84.2%，表现出显著优于现有方法的分割效果。这一结果彰显了中间帧质量对于脑肿瘤分割任务的重要性，也突出了本文方法在实际临床应用中的潜在价值。

在图像分割验证实验中，这些高质量的中间帧

进一步被用于脑肿瘤分割任务, 结果表明本文方法能够显著提高分割精度。在多模态分割性能对比中, 本文方法在 T1、T2、T1CE 和 Flair 模态下表现出一致的优越性, mDice 值分别为 84.1%、86.9%、81.2% 和 88.8%, 而 mIoU 值也分别达到了 73.4%、76.3%、68.0% 和 77.5%。这些数据不仅证实了本文方法在保持图像结构和细节方面的显著优势, 而且说明了其在不同成像条件下的稳定性和泛化能力。这些优越表现可归因于以下因素。

1) 光流估计的准确性: 通过改进的光流估计技术, 本研究能更准确地捕捉图像间微小的运动和变化。这对于医学图像而言尤为重要, 因为即使是微小的解剖结构变化也可能含有重要的临床信息。准确的光流估计使得模型能够在生成中间帧时更好地保留这些关键信息。

2) 特征融合机制: 本研究采用的编码器-解码器结构有效地融合了不同层次的特征信息, 强化了模型对图像细节的捕捉能力。特别是在处理解剖结构复杂或变化显著的图像时, 这种特征融合机制有助于维持图像的结构完整性。

3) 损失函数的设计: 本文采用的损失函数不仅关注重建图像的像素级准确性, 也强调结构和细节的保持。这使得模型在优化过程中更加注重于保留图像的结构特征, 从而在图像质量上获得更好的评价指标表现。

综上所述, 本研究提出的基于深度学习和光流的插帧技术在医学图像处理中显示出高效性和可靠性。未来工作将进一步探索这些技术因素对医学图像插帧性能的具体影响, 并且优化算法以适应更多样化的医学图像处理任务。

4 结束语

本研究提出了一种双向光流插帧算法, 用于解决三维医学图像 z 轴分辨率不足的问题。实验结果显示, 该算法在提升图像连续性、细节表现及分割准确率方面表现卓越, 尤其是在脑肿瘤诊断和治疗相关的医学图像处理中。所提出的紧凑型编码器-解码器结构网络为三维医学图像处理领域提供了新思路, 并期望对未来相关临床应用产生积极影响。

参考文献

- [1] SUN L, CHEN J, XU Y, et al. Hierarchical amortized gan for 3D high resolution medical image synthesis[J]. *IEEE Journal of Biomedical and Health Informatics*, 2022, 26(8): 3966-3975.
- [2] ZHANG J, XIE Y, WANG Y, et al. Inter-slice context residual learning for 3D medical image segmentation[J]. *IEEE Transactions on Medical Imaging*, 2020, 40(2): 661-672.
- [3] NI K S, NGUYEN T Q. An adaptable k-nearest neighbors algorithm for mmse image interpolation[J]. *IEEE Transactions on Image Processing*, 2009, 18(9): 1976-1987.
- [4] GRIBBON K T, BAILEY D G. A novel approach to real-time bilinear interpolation[C]//2nd IEEE International Workshop on Electronic Design, Test and Applications. Perth: IEEE, 2004: 126-131.
- [5] KEYS R. Cubic convolution interpolation for digital image processing[J]. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1981, 29(6): 1153-1160.
- [6] LIU X, ZHAO D, XIONG R, et al. Image interpolation via regularized local linear regression[J]. *IEEE Transactions on Image Processing*, 2011, 20(12): 3455-3469.
- [7] KINEBUCHI K, MURESAN D D, PARKS T W. Image interpolation using wavelet based hidden markov trees[C]//2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Salt Lake City: IEEE, 2001: 1957-1960.
- [8] O'SHEA K, NASH R. An introduction to convolutional neural networks[EB/OL]. [2023-12-19]. <http://arxiv.org/abs/1511.08458>.
- [9] BAKER S, SCHARSTEIN D, LEWIS J, et al. A database and evaluation methodology for optical flow[J]. *International Journal of Computer Vision*, 2011, 92: 1-31.
- [10] NIKLAUS S, MAI L, LIU F. Video frame interpolation via adaptive separable convolution[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 261-270.
- [11] PARIHAR A S, VARSHNEY D, PANDYA K, et al. A comprehensive survey on video frame interpolation techniques[J]. *The Visual Computer*, 2022: 1-25.
- [12] LI H D, YIN H, LIU Z H, et al. Enhanced spatialtemporal freedom for video frame interpolation[J]. *Applied Intelligence*, 2023, 53(9): 10535-10547.
- [13] SUN D, YANG X, LIU M Y, et al. Pwc-net: CNNs for optical flow using pyramid, warping, and cost volume [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 8934-8943.
- [14] LI X, CAO M, TANG Y, et al. Video frame interpolation via structure-motion based iterative fusion[EB/OL]. [2023-12-19]. <http://arxiv.org/abs/2015.05353>.
- [15] LI Z, ZHU Z L, HAN L H, et al. Amt: All-pairs multi-field transforms for efficient frame interpolation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2023: 9801-9810.
- [16] BAID U, GHODASARA S, MOHAN S, et al. The RSNA-ASNR-MICCAI brats 2021 benchmark on brain tumor segmentation and radiogenomic classification[EB/OL].

- [2023-12-19]. <http://arxiv.org/abs/2017.02134>.
- [17] CHOI M, KIM H, HAN B, et al. Channel attention is all you need for video frame interpolation [C]//Proceedings of the AAAI Conference on Artificial Intelligence. Vancouver: AAAI, 2020: 10663-10671.
- [18] DING T, LIANG L, ZHU Z, et al. Cdfi: Compression-driven network design for frame interpolation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2021: 8001-8011.
- [19] CHENG X, CHEN Z. Multiple video frame interpolation via enhanced deformable separable convolution[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(10): 7029-7045.
- [20] HUANG Z, ZHANG T, HENG W, et al. Real-time intermediate flow estimation for video frame interpolation[C]//European Conference on Computer Vision. Tel Aviv: Springer, 2022: 624-642.
- [21] KORHONEN J, YOU J. Peak signal-to-noise ratio revisited: Is simple beautiful?[C]//2012 4th International Workshop on Quality of Multimedia Experience. Melbourne: IEEE, 2012: 37-38.
- [22] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: From error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.
- [23] CHAI T, DRAXLER R R. Root mean square error (RMSE) or mean absolute error (MAE)[J]. Geoscientific Model Development Discussions, 2014, 7(1): 1525-1534.
- [24] WANG Z, BOVIK A C. A universal image quality index[J]. IEEE Signal Processing Letters, 2002, 9(3): 81-84.
- [25] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015. Munich: Springer, 2015: 234-241.
- [26] BADRINARAYANAN V, KENDALL A, CIPOLLA R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.
- [27] ZHOU Z, RAHMAN SIDDIQUEE M M, TAJBAKHS N, et al. Unet++: A nested unet architecture for medical image segmentation [C]//Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop. Granada: Springer, 2018: 3-11.
- [28] CAO H, WANG Y, CHEN J, et al. Swin-unet: Unet-like pure transformer for medical image segmentation [C]//European conference on computer vision. Tel Aviv: Springer, 2022: 205-218.

编辑 刘飞阳