

基于改进 V-Net 的脊柱 X 射线重建 CT 图像方法

高 天¹ 石征锦¹ 武 晨¹

GAO Tian SHI Zhengjin WU Chen

摘 要

脊柱作为人体支撑的核心要素,对维护身体健康和功能至关重要,手术机器人的出现为复杂脊柱手术提供了一种新的解决方案。但是 2D X 射线图像只能在有限的投影视图中为机器人系统提供重叠的解剖信息,这意味着它无法直观地显示完整的全视图解剖信息和精确的立体结构,而提供 3D 图像的术中 CT 扫描技术又增加了患者和医护人员受到辐射暴露的风险。为了在不增加时间成本且最大程度减少辐射的前提下,能够实时为手术机器人导航和准确定位提供更全面的解剖结构,提出一种用于脊柱 X 射线重建 CT 图像的 V 形卷积注意力网络。所提出的网络通过编码器和解码器之间的任务一致性以减小特征映射之间的语义差异,同时利用通道注意力机制来迫使网络关注重要特征区域,可有效减小冗余特征信息,从而提高网络训练效率。实验得出 CT 图像中脊柱结构的结构相似性指数 (structural similarity index, SSIM) 值为 0.786,峰值信噪比 (peak signal to noise ratio, PSNR) 值为 34.60 dB,证明通过 X 射线图像进行精准的 3D 重建为手术机器人提供图像支持拥有巨大潜力。

关键词

脊柱;手术机器人;改进 V-Net;注意力机制;2D 图像;3D 重建

doi: 10.3969/j.issn.1672-9528.2024.02.016

0 引言

近年来,医疗技术的迅猛发展引领脊柱手术领域的革新,脊柱手术机器人作为一项前沿技术备受瞩目和研究。与此同时,依赖于准确的图像数据进行手术导航和定位的机器人系统也更加凸显了术中实时成像的重要性。

目前,手术室通常依赖 X 射线成像来帮助外科医生在手术中进行导航和定位,尽管这种方法速度较快,但在复杂手术情境下,2D 图像提供的重叠解剖信息极大地限制了医生对脊柱结构的全面理解。而术中 CT 扫描虽然能够提供高质量的 3D 图像,但往往需要更多的时间和资源,并引发患者及医护人员对放射线暴露的担忧。这些技术的限制在一定程度上降低了机器人辅助手术的精确性和效率。

现阶段,深度学习方法的崛起为解决 3D 重建任务提供了更高效的途径。众多研究表明,深度学习技术能够通过多层神经网络提取那些人眼难以察觉的微小图像特征,与传统的人工提取特征方法相较,基于深度学习的 3D 重建方法呈现出高效和高精度的显著优势。

为了在提供详尽 3D 解剖信息的同时降低放射线暴露的风险,本文提出了一种 V 形卷积注意力网络,可以直接准确

高效地将单平面脊柱 2D X 射线转化为 3D CT 图像,从而为手术机器人系统提供更精准的图像支持,减少手术风险并提高手术成功率。

1 数据处理

本研究实验数据均来自中国医科大学附属盛京医院^[1-2]。由于获取与脊柱 CT 图像相匹配的 X 射线图像存在极大的困难,采用数字重建投影 (digitally reconstructed radiograph, DRR)^[3] 来模拟 X 射线图像作为模型的输入。对数据进行了重采样和图像切割等处理,以确保图像的一致性和适用性。为解决数据量不足的问题,通过对一部分数据进行了旋转、水平和垂直镜像翻转操作,共获得了 2000 个代表患者不同场景的 DRR,其中 1900 个用于训练,100 个用于测试。此外,为了增强 DRR 对比度,采用全局直方图均衡化^[4] 来重新平均分配图像中的每个像素,从而使图像中的脊柱结构更加清晰可见。

2 网络结构

2.1 整体网络结构

大量研究表明,V-Net 中采用卷积层替代传统的池化层的操作,使得网络在下采样过程中有效地保留了更多的细节信息,能够学习输入数据的局部特征表示有助于更好地捕捉

1. 沈阳理工大学 辽宁沈阳 110159

和还原图像的空间结构,这对重建任务来说是非常必要的。所以在本项研究中,提出了一种V形卷积注意力网络来完成2D图像到3D图像的重建,同时利用一种高效的自适应通道注意力机制来迫使神经网络模型更加专注于关键特征通道,以达到提升网络训练效率和重建图像精度的目的。网络整体结构如图1所示。

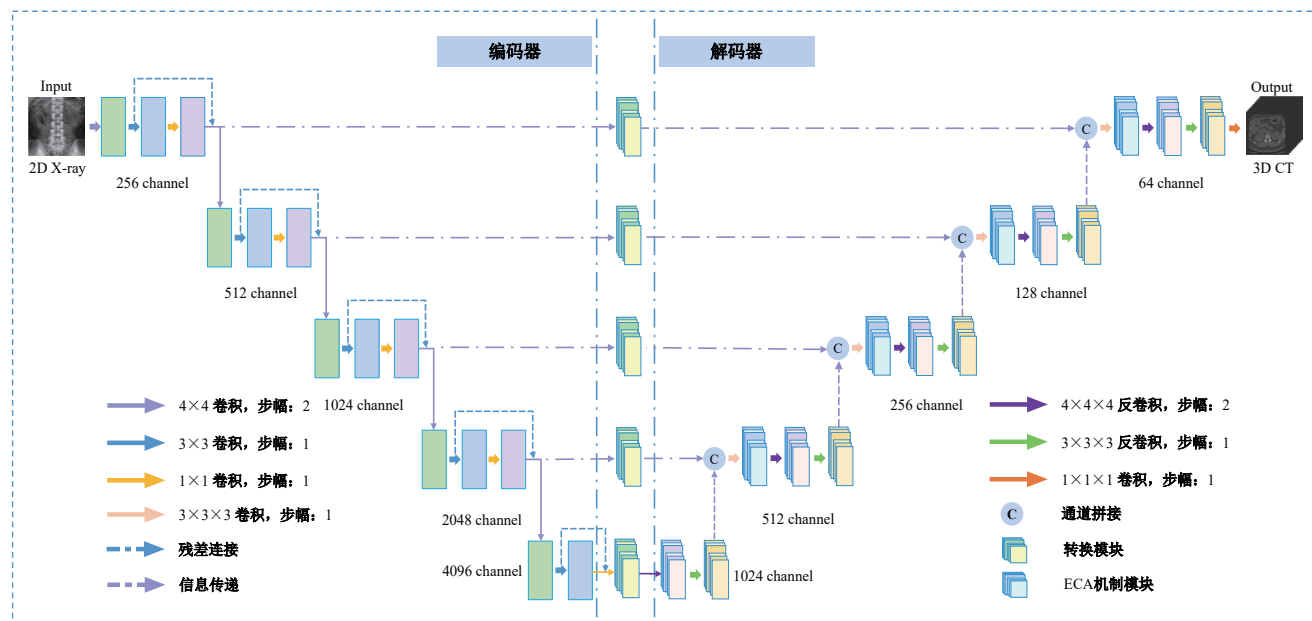


图1 网络整体结构图

V-Net为对称的编码器-解码器结构,其中3D卷积核、反卷积核的尺寸分别为 $5 \times 5 \times 5$ 、 $2 \times 2 \times 2$,每个卷积块首尾采用残差连接结构。重建网络由改进V-Net网络和ECA机制模块构成,为了适应输入的2D图像和出于内存考虑,将编码器的3D下采样卷积块替换为2D特征提取下采样卷积块。通过核尺寸为 4×4 步幅为2、 3×3 步幅为1和 1×1 步幅为1的卷积层,可以在网络捕捉较大范围内特征的同时,进一步提取局部细小特征,从而提高重建图像的质量。此外,在跳跃连接中引入了一系列的转换模块。如图2所示,通过将2D特征图重塑为3D形式,再利用核尺寸为 $1 \times 1 \times 1$ 步幅为1的反卷积层稳固特征空间形状,实现了2D到3D的特征映射变换。

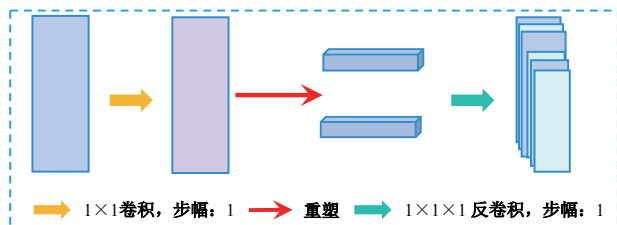


图2 转换模块结构图

首先,在解码器中,将前一解码层的输出与编码器提

取的特征相融合,以产生更为精细和更大规模的解码结果。其次,利用核大小为 $3 \times 3 \times 3$ 步幅为1的卷积层融合跳跃连接特征,以产生更为细化解码结果。然后通过ECA模块自适应地加权不同通道的特征图,提高网络的计算效率的同时提高重建图像的精度。最后使用核大小分别为 $4 \times 4 \times 4$ 和 $3 \times 3 \times 3$ 的反卷积层,用于放大特征图的尺寸

和稳固空间形状。将通道数进行了数倍的扩展,使网络可以适应复杂的数据分布,从而提高特征的丰富性。为了加速网络训练,在每一层后都加入批量规范化和并用Relu函数激活。

在重建网络的最后,通过一个大小为 $1 \times 1 \times 1$ 步幅为1的卷积层构成一个输出变换模块,并去掉了通道的维度,输出与重构图像形状拟合的预测脊柱CT图像。

2.2 ECA 模块

V-Net存在的一个主要问题是高阶和低阶特征的串联会导致特征信息冗余,因此有必要通过一种通道注意力机制来确保网络更专注于关键通道特征^[5]。挤压激励(squeeze-and-excitation, SE)^[6]机制是最典型的跨通道注意力模块之一,它通过全局平均池化获取特征映射的全局分布,并引入两层密集神经网络获得通道权值,自适应地加权不同通道的特征图,以提高网络的计算效率。虽然SE模块被广泛应用于各种网络,但其中的降维操作可能导致信息丢失,而参数数量巨大的全连接层也增加了模型的计算成本。ECA(efficient channel attention)^[7]机制利用1D卷积代替了全连接层,在避免信息丢失的同时又大量节省了模型的计算成本。ECA模块整体结构如图3所示。

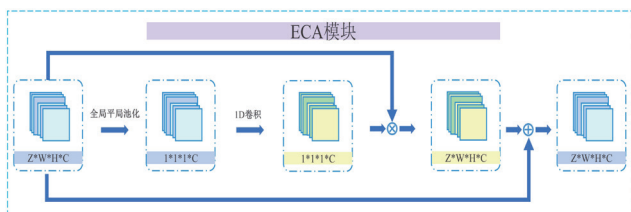


图3 ECA 模块结构图

输入特征的深度为 Z ，高度为 H ，宽度为 W ，通道数为 C 。该模块通过应用全局平均池化层获得每个通道的全局信息^[8]，以便更好地理解各个通道的重要性。随后采用自适应卷积核的1D卷积的方式来实现适当的全连接跨通道交互，以获取初始特征映射的通道权值。这个过程中，每次卷积操作仅与一部分通道发生交互，从而避免了降维操作并有效减少了参数计算量。卷积核与通道数映射关系可表示为：

$$C = \phi(k) = 2^{(\gamma * k - b)} \quad (1)$$

$$k = \psi(C) = \left\lfloor \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}} \quad (2)$$

式中： γ 和 b 为预先设定的参数值， odd 代表取正奇数。将学习到的权重应用于每个通道的特征图，使得模型能够自适应地放大或缩小不同通道的贡献，以提高关键特征信息的重要性。ECA模块输出可表示为：

$$G = F_{\text{gap}}(I) = \frac{1}{Z \times H \times W} \sum_{i=1}^Z \sum_{j=1}^H \sum_{k=1}^W I(i, j, k) \quad (3)$$

$$O = \left(\frac{1}{1 + e^{-f(G)}} \times I \right) + I \quad (4)$$

式中： G 为全局平均池化输出， I 为输入特征图， $f(\cdot)$ 代表1D卷积， O 为输出特征图。

3 损失函数和超参数

3.1 损失函数

损失函数作为模型的优化方向是非常重要的。图像重建是一种像素级的任务，选用均方误差（mean absolute error, MSE）作为损失函数。MSE以平方误差的平均值作为度量标准，衡量了模型的预测与实际值之间的差异，有助于促使模型生成平滑的输出，从而减少了输出中的不连续性，MSE计算公式为：

$$\ell_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2 \quad (5)$$

式中： n 为样本个数， y_i 为真实图像， y'_i 为预测图像。

3.2 超参数设置

使用PyTorch库来实现网络，并使用Adam^[9]作为优化器，通过最小化损失函数并使用反向传播迭代更新网络

参数，初始学习率为0.001，所有的实验都将批次大小设置为1。

4 实验结果与分析

4.1 性能指标

使用脊柱数据集进行实验，综合考虑了SSIM、PSNR以及推理速度等指标，从多个角度全面评估了重建网络的性能。SSIM和PSNR定义为：

$$\text{SSIM}(y_{CT}, y'_{CT}) = \frac{(2\mu_{y_{CT}}\mu_{y'_{CT}} + c_1)(2\text{cov}(y_{CT}, y'_{CT}) + c_2)}{(\mu_{y_{CT}}^2 + \mu_{y'_{CT}}^2 + c_1)(\sigma_{y_{CT}}^2 + \sigma_{y'_{CT}}^2 + c_2)} \quad (6)$$

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}_{y_{CT}}^2}{\text{MSE}} \right) \quad (7)$$

式中： y_{CT} 为真实图像像素值， y'_{CT} 为预测图像像素值， μ 为平均值， σ 为标准差， $\text{cov}(\cdot)$ 为协方差， c 为稳定除法变量， MSE 为均方误差。

4.2 对比试验

为了验证ECA模块具有加速网络模型训练效率的作用，分别对带有ECA模块和不带有ECA模块的网络进行30次迭代训练，损失下降曲线如图4所示。从图中可以看出带有ECA模块的网络损失下降得更快也更平稳，从而验证了ECA模块具有加速网络训练效率的作用。

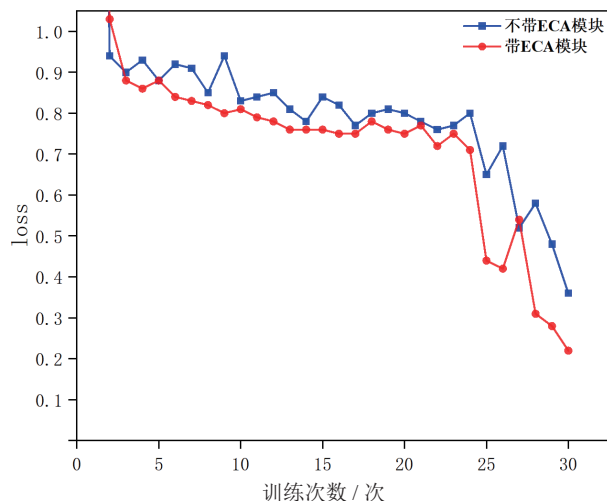


图4 引入ECA模块前后损失曲线对比图

将本文方法与现有的重建方法进行了比较，以评估该方法在图像重建任务中的性能表现。使用相同的数据集，从各个性能指标全面评估了各个方法的表现情况，由于肌肉组织在的研究中并不具备参考价值，在评估中将其从图像中去除，对比实验性能指标和结果图分别如表1和图5所示。从表1中可以看出，相比其他方法，本文的方法实现了更高精度的脊柱CT图像的重建，其中SSIM值平均提高了6.4%，PSNR值平均提高了0.80 dB。从图5中可以看出，它在视觉上实现

了清晰无重叠的全视图脊柱结构，具有清晰的分布、形状和清晰可读的解剖纹理。

表 1 性能指标

方法	SSIM	PSNR	推理速度
Shen 等 ^[10]	0.736±0.025	33.70±0.32	0.52 s/患者
Ying 等 ^[11]	0.742±0.027	33.91±0.28	0.86 s/患者
本文方法	0.786±0.011	34.60±0.17	0.74 s/患者

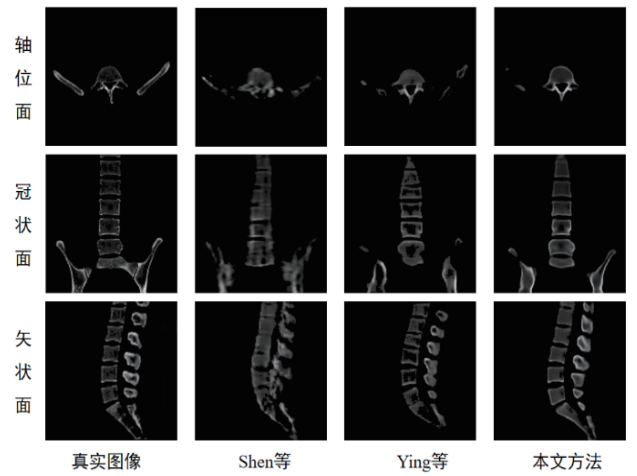


图 5 对比试验结果图

5 结语

本文提出了一种用于脊柱 X 射线重建 CT 图像的 V 形卷积注意力网络。对比实验结果表明，重建图像中脊柱结构的 SSIM、PSNR 值分别平均提升了 6.4% 和 0.80 dB，能够在不浪费时间和资源的前提下避免临床操作中可能存在的 2D 图像信息不完全以及术中 CT 带来的辐射问题，为手术机器人在术中定位和导航方面提供有力的图像支持。然而该方法仍然存在一些局限性，例如，在脊柱部分椎体重建上的还原度可能较差，并且椎体间的互锁关系表现不佳。未来的研究将继续致力于克服这些局限性，推动手术机器人技术的进一步发展。

参考文献:

[1] 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.

[2] 孟令帅. 妊娠合并心力衰竭相关因素分析 [D]. 沈阳: 中国医科大学, 2022.

[3] MILICKOVIC N, BALTAS D, GIANNOULI S, et al. CT imaging based digitally reconstructed radiographs and their ap-

plication in brachytherapy[J]. Physics in medicine & biology, 2000, 45(10): 2787.

[4] SENTHILKUMARAN N, THIMMIARAJA J. Histogram equalization for image enhancement using MRI brain images[C]//2014 World congress on computing and communication technologies.Piscataway:IEEE,2014: 80-83.

[5] 王佩. 基于特征融合与信息增强的目标检测算法研究 [D]. 赣州: 江西理工大学, 2022.

[6] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE conference on computer vision and pattern recognition.Piscataway:IEEE,2018: 7132-7141.

[7] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.Piscataway:IEEE,2020: 11534-11542.

[8] 闫道申, 邵冬梅, 许铭洋, 等. 基于高效注意力机制的说话人识别 [J]. 警察技术, 2023(4):47-51.

[9] KINGMA D P, BA J. Adam: A method for stochastic optimization[EB/OL].(2014-11-22)[2023-09-06].<https://arxiv.org/abs/1412.6980>.

[10] SHEN L, ZHAO W, XING L. Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning[J]. Nature biomedical engineering, 2019, 3(11): 880-888.

[11] YING X, GUO H, MA K, et al. X2CT-GAN: reconstructing CT from biplanar X-rays with generative adversarial networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.Piscataway:IEEE, 2019: 10619-10628.

【作者简介】

高天（1998—），男，辽宁朝阳人，硕士，研究方向：复杂系统综合自动化技术。

石征锦（1963—），男，辽宁锦州人，硕士，教授，研究方向：先进控制理论及应用、现代检测与自动化装置。

武晨（1998—），女，山东泰安人，硕士，研究方向：检测技术与自动化装置。

（收稿日期：2023-11-17）