

中图法分类号: TP391.41; R445 文献标识码: A 文章编号: 1006-8961(2025)11-3583-21

论文引用格式: Ai B, Sun Z, Hou Y S and Sun M C. 2025. Bidirectional feedback iterative optimization network for joint reconstruction and segmentation of photoacoustic tomographic images. Journal of Image and Graphics, 30(11):3583-3603(艾冰, 孙正, 侯英飒, 孙美晨. 2025. 面向光声层析图像联合重建与分割的双向反馈迭代优化网络. 中国图象图形学报, 30(11):3583-3603)[DOI:10.11834/jig.240743]

面向光声层析图像联合重建与分割的 双向反馈迭代优化网络

艾冰¹, 孙正^{1,2*}, 侯英飒¹, 孙美晨¹

1. 华北电力大学电子与通信工程系, 保定 071003; 2. 华北电力大学河北省电力物联网技术重点实验室, 保定 071003

摘要: 目的 光声层析成像(photoacoustic tomography, PAT)是一种多物理场耦合的新型成像技术。目前, 主流PAT应用是在图像重建后进行图像分割和目标识别, 重建中存在物理模型不合理假设、噪声和数据不完备等问题, 直接影响后续分割精度。为了解决这一问题, 本文提出一种联合图像重建和分割的深度学习框架。方法 将图像重建和分割两个任务结合起来, 充分利用二者之间的关联: 在分割过程中, 将重建结果作为先验知识, 更好地恢复和保留目标的形状和边界; 分割结果又为重建提供约束和引导, 使得重建结果更贴近真实目标, 从而同步完成PAT图像重建和分割。结果 仿真、仿体和在体实验结果表明, 与两步法相比, 采用所提方法重建图像的结构相似度指标可提高约10.01%, 峰值信噪比可提高约12.15%, 图像分割的DICE指标可提高约13.27%, Jaccard指数可提高约6.08%, 平均表面对称距离可降低约16.55%。和其他联合重建与分割方法相比, 所得方法重建图像的结构相似度可提高约3.65%, 峰值信噪比可提高约4.96%; 图像分割的DICE指标可提高约2.73%, Jaccard指数可提高约3.85%, 平均表面对称距离可降低约5.99%。结论 本文方法通过实现PAT图像的联合重建与分割, 有效减少了重建误差积累对于分割精度的影响, 提高PAT的成像质量和分割精度, 展现出良好的应用前景。

关键词: 光声层析图像(PAT); 图像重建; 图像分割; 深度学习; 联合分析

Bidirectional feedback iterative optimization network for joint reconstruction and segmentation of photoacoustic tomographic images

Ai Bing¹, Sun Zheng^{1,2*}, Hou Yingsa¹, Sun Meichen¹

1. Department of Electronic and Communication Engineering, North China Electric Power University, Baoding 071003, China;

2. Hebei Key Laboratory of Power Internet of Things Technology, North China Electric Power University, Baoding 071003, China

Abstract: Objective Photoacoustic tomography (PAT) is an emerging imaging technique that combines optical contrast with acoustic resolution, and it demonstrates great potential in the field of biomedical applications in recent years. Its non-invasive nature, high contrast, and high resolution make it an important tool in medical diagnosis and research. However, due to limitations in the imaging principles and the complexity of biological tissues, photoacoustic images often exhibit blurring, artifacts, and low contrast, which restrict their accuracy and reliability in clinical applications. Improving image quality is crucial for the further advancement of PAT technology. Current research generally treats image reconstruction and segmentation as two separate steps. First, the inverse problem is solved to reconstruct a 2D or 3D image from the photo-

收稿日期: 2024-12-12; 修回日期: 2025-04-02; 预印本日期: 2025-04-09

* 通信作者: 孙正 sunzheng@ncepu.edu.cn

基金项目: 国家自然科学基金项目(62071181)

Supported by: National Natural Science Foundation of China (62071181)

acoustic signals collected by the ultrasonic detectors, which realizes the “signal-to-image” conversion. Then, regions of interest (e. g. , specific tissues or structures) are segmented or identified from the reconstructed images, which realizes the “image-to-feature” conversion. This decoupled processing framework shows substantial limitations in practical applications. In the image reconstruction phase, errors inevitably arise due to inaccuracy of the physical model, unreasonable assumptions (e. g. , the acoustically homogeneous medium and constant speed of sound), noise interference, and incomplete data acquisition. These errors accumulate in subsequent segmentation and target recognition tasks, which greatly reduces the accuracy of automated analysis. This inadequacy, in turn, limits the potential of PAT in fields such as precision medicine and real-time diagnosis. To address this issue, this study proposes an end-to-end deep learning framework that innovatively integrates image reconstruction and segmentation tasks into a unified network. Through the interaction and collaborative optimization between the two tasks, the framework fully leverages their potential correlations to considerably improve the overall quality of PAT images and achieve higher accuracy in image segmentation. **Method** This study presents an end-to-end joint reconstruction and segmentation deep learning framework, which aims to enhance the quality of PAT image reconstruction and segmentation accuracy through iterative optimization. The framework consists of two core modules: the reconstruction and segmentation modules. The reconstruction module is used to generate an initial image using deep neural networks while minimizing reconstruction errors caused by factors such as incomplete physical models or noises. The segmentation module performs further image processing based on the reconstruction module, with its task being the precise segmentation of the regions of interest. To capture the details of the target regions in the image, the segmentation module adopts a context-aware mechanism, which combines global and local features of the image. Specifically, the segmentation module employs a hybrid architecture that combines fully convolutional networks (FCNs) and DenseNet. The encoder-decoder structure of the FCN enables pixel-level classification and extracts low-level and high-level features from the image, while DenseNet enhances the utilization of information from previous layers through dense connections, which improves feature representation capability. The combination of FCN and DenseNet allows the segmentation module to better process complex image information and achieve higher precision semantic segmentation. The key aspect of the framework is the establishment of a bidirectional feedback mechanism between the reconstruction and segmentation modules, which creates a deep collaboration loop. In this mechanism, the output of the segmentation module serves as a constraint that guides the reconstruction module to refine the restoration of the target region, while the image provided by the reconstruction module offers important prior information to the segmentation module. Ultimately, the segmentation of the target regions is further optimized. Through this bidirectional feedback, the two modules promote each other during the iterative process, which continuously improves the accuracy of reconstruction and segmentation. In each iteration, the reconstruction module corrects reconstruction errors with the feedback from the segmentation module, while the segmentation module extracts more accurate features from the progressively optimized reconstructed image, which improves segmentation precision. Unlike traditional methods, this framework avoids the issue of accumulation of errors due to independent tasks through deep collaboration. In conventional approaches, image reconstruction and segmentation are typically processed separately, which leads to a lack of information sharing between the two tasks and ultimately affects the overall accuracy of the results. On the contrary, the proposed framework optimizes both modules jointly. This way ensures that the output of each module provides higher-quality input for the next, which ensures synchronized improvement of reconstruction and segmentation tasks. Ultimately, through multiple iterations, this deep integration of information not only greatly enhances image clarity but also substantially improves segmentation accuracy. **Result** To validate the effectiveness of the proposed method, several simulation experiments, phantom studies, and in vivo experiments were conducted. Experimental results demonstrate remarkable advantages over the two-step scheme in image reconstruction and segmentation tasks. In simulation experiments, the proposed method improved the structural similarity index (SSIM) of the reconstructed images by up to 10.01% and the peak signal-to-noise ratio (PSNR) by up to 12.15%. In segmentation tasks, the DICE coefficient and the Jaccard index increased by up to 13.27% and 6.08%, respectively, while the average symmetric surface distance (ASSD) was reduced by up to 16.55%. In phantom experiments, the method exhibited robust performance even in complex scenarios (e. g. , heterogeneous media and variable speed of sound). Compared with other joint reconstruction and segmentation methods, the SSIM and PSNR of the reconstructed images can be enhanced by approximately 3.65% and 4.96%, respec-

tively; the DICE, Jaccard index, and ASSD for image segmentation can be raised by nearly 2.73%, 3.85%, and reduced by nearly 5.99%, respectively. Further in vivo experiments validated the applicability of this method in real-world biological imaging. In tumor detection and vascular imaging tasks, the images generated by the proposed method exhibited sharp boundaries, accurate shapes, and significantly improved segmentation integrity of the target regions, which exceed those of traditional methods. Therefore, the proposed method has strong robustness and generalizability. **Conclusion** This study proposes an end-to-end deep learning framework that jointly performs image reconstruction and segmentation tasks, which effectively solves the critical issue of error accumulation in traditional two-step methods. Compared with existing approaches, the proposed method better restores the shape and boundaries of the target region, reduces information loss and blurring, and demonstrates great performance improvements in simulation, phantom, and in vivo experiments. The study shows that the proposed method has significant theoretical importance and demonstrates great potential in practical applications. Future research will further explore the application potential of this framework in real-time processing and large-scale datasets to promote the clinical translation of PAT into medical diagnosis and treatment.

Key words: photoacoustic tomography (PAT); image reconstruction; image segmentation; deep learning; joint analysis

0 引言

光声层析成像(photoacoustic tomography, PAT)是一种新型非电离成像技术,融合了超声和光学成像的优势,实现非侵入、高分辨和高对比度的生物组织成像(Oh等,2006;Lungu等,2007;Yao等,2013;Yao和Wang,2018)。该技术利用短脉冲激光照射目标,组织吸收光能后产生宽带超声波(即光声波),位于目标周围的超声探测器采集光声波,经图像重建后生成初始声压、光吸收能量、光学特性参数和发色团浓度等的分布图(Poudel等,2019;Yang等,2019)。

图像重建是PAT技术的核心,目的是从探测器采集的光声信号中恢复出目标的形态、结构和功能成分信息。传统方法包括延迟求和(delay and sum, DAS)(Xu等,2003)、时间反演(time reversal, TR)(Bal和Uhlmann,2010)、滤波反投影(filtered back-projection, FBP)(Xu和Wang,2005)和迭代重建(Paltauf等,2002)等,这些方法通常基于理想场景的假设,如完备测量、恒定声速和均匀照明,在实际应用中容易受到有限角度扫描、声速不均匀性以及非均匀光通量等因素的影响,导致重建图像质量下降。例如,DAS因忽略了声速的空间异质性会产生相位失配噪声,导致图像边缘模糊;FBP在有限角度扫描时易受条纹伪影干扰;TR在处理不均匀介质时可能产生不稳定解;迭代重建算法虽然能提高重建精度,但计算复杂度较高。通过改进传统方法,可以提高图像重建的准确性和鲁棒性。例如,利用对光声信

号的稀疏表示优化重建过程可以有效降低噪声干扰(Liu等,2019)或通过迭代求解前向成像模型的逆问题,同步重建初始声压和声速分布图,解决恒定声速假设导致的图像质量下降问题(Sun和Sun,2021)。此外,压缩感知有助于缓解因稀疏采样和有限角度扫描导致的测量数据不完备问题(Arridge等,2016;Sun和Yan,2020b)。

近年来,基于深度学习的图像重建技术受到关注,现有方法分为图像域到图像域的转换(Antholzer等,2019;Farnia等,2020;Sun和Yan,2020a)和信号域到图像域的转换(Dey等,2024;Li等,2022;Guo等,2022;Hakimnejad等,2024;Madhumathy和Pandey,2022;Yang等,2021;Joseph等,2021;Lu等,2021;Sun等,2021)两类。前者结合卷积神经网络(convolutional neural network, CNN)与传统方法,先用传统方法重建初始图像,再用CNN优化图像,以抑制噪声、增强边缘信息并减少伪影;后者则直接根据声压信号重建高质量图像。例如,自平衡双机制(self-balancing dual mechanism, SBDM)方法通过低分辨率重建和高分辨率增强两个模块实现噪声自适应的信号处理(Dey等,2024);深度立体几何网络(deep projection geometry network, DPG-Net)利用深度学习代替传统投影几何模型,通过学习信号传播关系进行精确重建(Li等,2022);自适应采样网络(adaptive sampling network, AS-Net)根据图像的边缘和细节优化采样位置,抑制高噪声区域,提高重建效率(Guo等,2022);密集光声网络(dense photoacoustic network, DensePANet)通过多层特征共享实现多尺度特征融合,增强对比度(Hakimnejad等,2024)。

此外, CNN 直接学习光声信号与图像间的映射关系 (Madhumathy 和 Pandey, 2022; Yang 等, 2021), 而生成对抗网络 (generative adversarial network, GAN) 则可通过对抗训练减少伪影, 进行图像重建 (Joseph 等, 2021; Lu 等, 2021)。深度梯度下降 (deep gradient decent, DGD) 网络结合了传统梯度下降方法和 CNN, 利用输入信号的梯度信息减少稀疏数据对图像质量的影响 (Sun 等, 2021)。

PAT 图像的精确分割对于提升其诊断和量化能力至关重要。然而, 由于光声成像过程中噪声干扰、组织散射和设备分辨率等因素的影响, 分割任务面临较大挑战。传统方法如阈值分割 (Sun 等, 2020)、边缘检测 (Mandal 等, 2015; Zhao 等, 2018)、基于 Hessian 矩阵的方法 (卢学敏, 2021) 和区域生长 (Zhao 等, 2014) 虽然操作简便, 但存在对低质量图像的分割难度大、泛化性差和难以实现语义级分割等问题。例如, 阈值分割依赖于固定或自适应阈值, 在低信噪比的光声图像中难以准确区分目标区域与背景; 边缘检测方法易受噪声影响, 可能导致虚假边界检测或边缘断裂; Hessian 矩阵方法虽然能增强线状结构, 但在低对比度区域容易误检; 区域生长方法则对初始种子点选择敏感, 可能因噪声导致过分割或欠分割。此外, 传统方法主要依赖局部特征进行分割, 缺乏对全局语义信息的理解, 难以实现精准的组织识别和语义级分割。近年来, 深度学习在光声图像分割中取得了显著进展, 尤其是 U-Net 及其变体 (周涛 等, 2021)。例如, 无新网络 U-Net (no new U-Net, nnU-Net) 能根据输入数据自适应调整参数和结构以获得最佳分割效果 (Isensee 等, 2021), 但其计算成本较高, 对硬件要求较高; U-Mamba (U-shaped Mamba) 模型适合多模态图像融合分割 (Ma 等, 2024); 全局上下文扩展的 U-Net (global context extended UNet, GCxt-UNet) 结合 U-Net、图卷积网络 (graph convolutional network, GCN) 和 Transformer 的优势, 可以更高效地提取空间特征 (Alrfou 和 Zhao, 2024); 直接训练 U-Net 使其学习光声图像中不同组织的特征表示, 可以用于提取皮肤和其覆盖血管产生的光声信号振幅之间的关系特征 (Ly 等, 2022); 稀疏 U-Net 可用于多光谱图像中的血管分割 (Chlis 等, 2020); 利用 U-Net 可以实现光声—超声双模态图像中目标边界的提取 (Lafci 等, 2020, 2021)。除了 U-Net, 其他网络如胶囊网络 (capsule network,

CapsNet) (Lalonde 等, 2021)、分割网络 (segmentation network, SegNet) (Badrinarayanan 等, 2017) 和基于分组注意力的医学图像分割模型 (medical image segmentation module based on group attention, GAU-Net) (张学峰 等, 2023) 也展现出优势。CapsNet 通过胶囊层保留和传递特征位置信息, 但计算量大; SegNet 适用于边缘和细节精度要求较低的任务, 结构简洁, 计算效率高; GAU-Net 能够有效地增强模型对局部和全局重要特征的提取能力, 但在小目标分割上仍表现不足。尽管深度学习方法在光声图像分割中取得了显著进展, 但仍存在计算复杂度高、对低信噪比数据适应性差、边界模糊、跨模态适应性不足及泛化能力有限等问题。nnU-Net 和 GCxt-UNet 计算资源需求大, 限制了其在资源受限环境中的应用; U-Net 及其变体在低信噪比条件下可能误识别噪声为目标 (假阳性) 或遗漏目标 (假阴性); GCxt-UNet 虽然结合 GCN 和 Transformer 增强特征提取, 但可能导致边界过度平滑, 影响精细结构分割; 光声—超声双模态任务中, 因不同模态的物理特性差异, 跨模态适应性较差; CapsNet 虽然保留了特征位置信息, 但计算复杂度高, 泛化能力受限。此外, 深度学习模型可能累积标注误差、背景噪声误识别、边界模糊伪影、设备特性伪影及跨模态误分类等, 这些因素均可能影响最终分割结果的可靠性。

目前, 主流 PAT 应用研究中是将图像分割和目标识别作为图像后处理的步骤完成的, 即采用两步策略: 首先, 根据探测器采集的光声信号重建图像, 实现信号到图像的转换; 然后, 从图像中提取或识别出感兴趣目标, 实现图像到特征的转换。在图像重建过程中, 不准确的物理模型、对成像场景的不合理假设 (如恒定声速、均匀介质、理想点探测器和均匀光覆盖等)、测量数据中的噪声或不完备性等因素都会导致误差, 造成图像模糊、特征丢失和整体质量恶化。这种误差会累积到后续图像处理步骤中, 图像的低对比度、伪影和模糊会导致在图像域中进行的自动分割和识别变得复杂, 最终影响分割的准确性和目标识别的精度。

联合图像重建与分割旨在充分利用重建和分割任务之间的相互关系, 重建受益于分割获得的目标信息, 而分割通过考虑激发场的异质性和成像区域的组织特性, 可以增强重建性能 (Corona 等, 2019)。例如, Boink 等人 (2020) 利用部分学习原对偶

(learned primal dual, L-PD)算法,对图像域中的原始算子和数据域中的对偶算子进行迭代更新,首次实现了PAT图像的同时重建和分割。L-PD方法依赖于训练数据的多样性和质量,此外在处理复杂结构或不均匀数据分布时,易受模型泛化能力的限制,可能导致分割和重建质量的下降。孙美晨(2023)提出一种端到端的深度学习框架(image reconstruction and identification network, IRI-Net),同步完成PAT图像重建与目标识别。IRI-Net中的重建模块利用DGD网络将光声信号转换为光吸收能量图像,优化模块采用U-Net进一步提升图像质量,识别模块采用ResNet-50进行特征提取和分类识别,最终完成从信号到类别的映射。然而,该框架设计对计算资源要求较高,尤其在处理复杂组织和结构时,训练时间较长。

本文提出一种面向PAT图像联合重建与分割的双向反馈迭代优化网络,将图像重建和分割两个任务结合起来,充分利用二者之间的关联,实现从光声信号测量数据到重建图像和分割后图像之间的端到端映射。在图像分割过程中,将图像重建结果作为先验知识,更好地恢复和保留目标的形状和边界。同时,分割结果也为图像重建提供更准确的约束和引导,使得重建结果更贴近真实目标。

1 PAT图像联合重建与分割方法

1.1 联合重建与分割网络的架构

如图1所示,本文提出的PAT图像联合重建与分割框架由重建模块和分割模块组成,采用循环迭代和双向反馈机制进行信息交互,以优化重建质量和分割精度。重建模块基于全卷积神经网络搭建,

以探测器采集的声压数据和低质量图像作为输入,用于实现图像重建。分割模块采用全连接网络(fully connected network, FCN)(Sun 和 Wang, 2018)和密集连接网络(dense connection network, DenseNet)(Peng 等, 2023)相结合的架构,实现基于像素的语义分割。第1个重建模块的输入是声压数据和用常规方法DAS重建的低质量图像,其输出图像随后进入第1个分割模块进行初步分割。然后,分割结果被输入下一个重建模块,与原始声压数据结合后用于优化重建质量,生成更高质量的PAT图像。随后,优化后的重建图像再次输入下一个分割模块,使其能够获得更精确的目标边界。该过程持续迭代,每次的分割结果都用于调整下一阶段的重建过程,同时优化分割任务自身,使系统在多个循环中逐步收敛,最终实现同步优化的高质量重建和精确分割。这种迭代式的连接使得每个阶段的重建和分割结果相互影响。具体而言,重建模块在每次迭代中都会接收来自前一轮分割结果中获取的信息,逐步纠正重建中的误差,以优化组织结构的恢复效果。而分割模块能够从不断优化的重建图像中获得更准确的图像特征,使得其预测结果更精确地匹配优化后的重建图像,减少误分割或漏分割的情况。通过这种循环迭代的方式,每个模块的输出都为下一个模块提供更高质量的输入数据,使得整体重建和分割效果在多个循环中逐步提升,在低信噪比环境下有效应对噪声干扰,提高目标组织的可识别性,最终同步实现高质量的图像重建和精确的图像分割。

如图2所示,重建模块包括声压测量数据和低质量图像两个分支,分别对声压数据和低质量图像进行特征提取和表示学习,逐渐减小特征图的空间

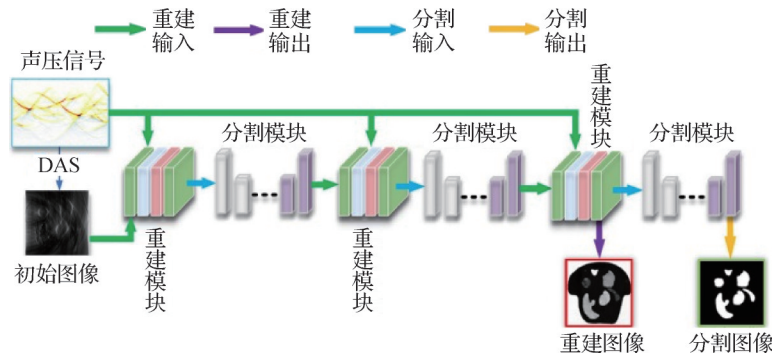


图1 PAT图像联合重建与分割网络结构

Fig. 1 Architecture of the joint reconstruction and segmentation network for PAT images

尺寸,同时增加通道数,以提取更高级别的特征,捕捉到原始数据中的重要结构信息。最后,采用逐元素加权求和的方式,将测量数据分支和低质量图像分支输出的特征图融合成一个综合和丰富的特征表示,以充分利用声压测量数据和低质量图像中隐含的信息,提高重建图像的准确性和清晰度。

如图3所示,分割模块将FCN和DenseNet结合起来,充分利用两个网络结构的优势和两者之间的互补性,实现对图像的语义分割。标签图像(即掩膜)是由GCxt-UNet(Alrfou和Zhao,2024)分割生成的高精度图像,其中包含图像的边界类别信息,有助于区分图像中的不同目标或区域,通过比较输出与真实掩膜之间的匹配度,判断模型的准确性和有效性。FCN用于对输入图像进行像素级分类,得到初

步分割结果。FCN包含编码器(提取特征)和解码器(上采样)两部分。编码器用于逐步提取图像中的特征信息。解码器通过上采样操作将编码器得到的特征图还原到原始输入图像的大小,并进行分类。DenseNet是一种密集连接的网络结构,用于从原始图像中提取丰富的特征信息。它的每一层都将之前所有层的输出特征图连接起来作为输入,这种密集连接机制使得网络能够更好地利用前面层的特征信息,并将其传递给后续层,从而提高网络的特征表征能力。FCN的输出与DenseNet输出的特征图进行逐元素的融合和拼接,得到更准确的分割结果。

1.2 联合损失

利用Adam梯度下降算法优化网络参数。为了在训练过程中同时优化重建和分割模块的参数,使

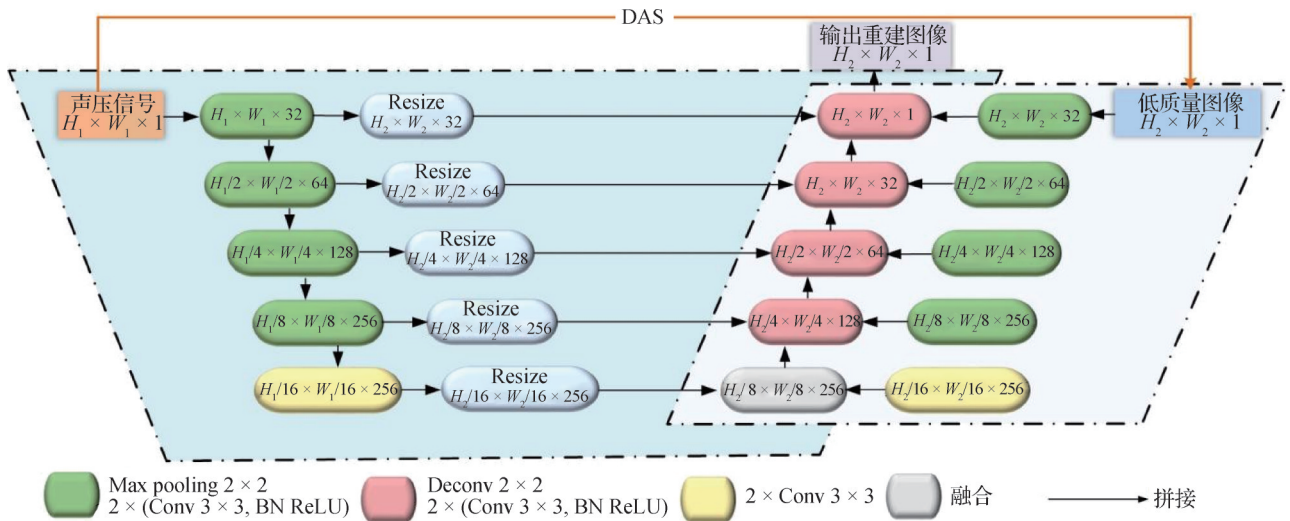


图2 重建模块的结构
Fig. 2 Structure of the reconstruction module

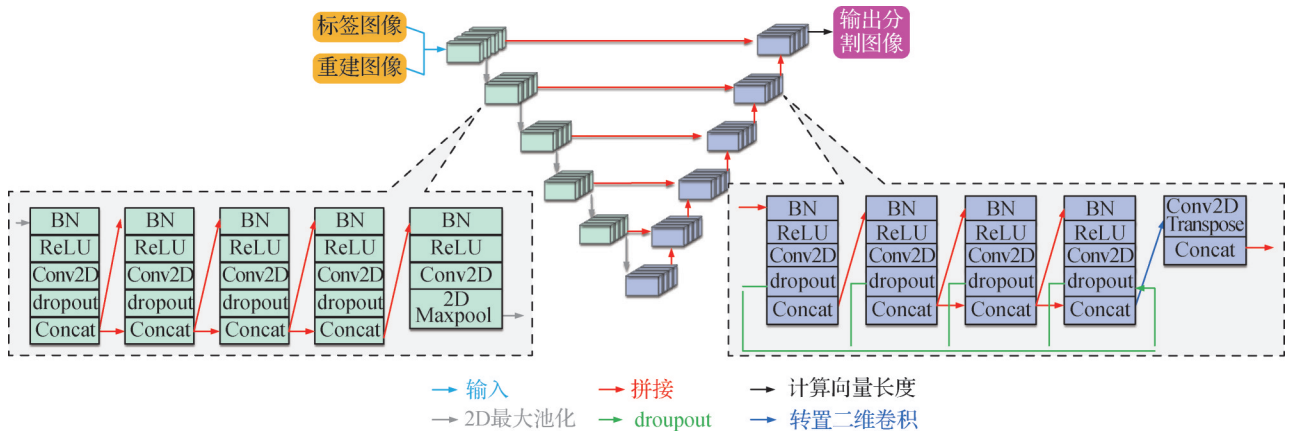


图3 分割模块的结构
Fig. 3 Structure of the segmentation module

得分割可以促进重建,达到更好的联合效果,将网络的总损失函数定义为

$$\mathcal{L} = \sum_{t=1}^T (\lambda_{\text{rec}} \mathcal{L}_{\text{rec},t} + \lambda_{\text{seg}} \mathcal{L}_{\text{seg},t} + \lambda_{\text{cross}} \mathcal{L}_{\text{cross},t}) \quad (1)$$

式中, T 是迭代次数, $\mathcal{L}_{\text{rec},t}$ 、 $\mathcal{L}_{\text{seg},t}$ 和 $\mathcal{L}_{\text{cross},t}$ 分别是第 t 次迭代的图像重建损失、图像分割损失和交叉约束损失, λ_{rec} 、 λ_{seg} 和 λ_{cross} 是权重系数,用于平衡不同损失项在总损失中的贡献。

采用均方误差损失定义图像重建损失,同时考虑到低质量图像分支含有目标的大多数纹理信息,因此在该分支添加深度监督损失约束其图像特征,具体为

$$\mathcal{L}_{\text{rec},t} = \mathcal{L}_{\text{rms},t} + \lambda_{\text{aux}} \mathcal{L}_{\text{aux},t} \quad (2)$$

式中, $\mathcal{L}_{\text{rms},t}$ 是第 t 次迭代的均方误差损失, $\mathcal{L}_{\text{aux},t}$ 是深度监督损失。具体为

$$\mathcal{L}_{\text{rms},t} = \frac{1}{n} \sum_{i=1}^n \|y_{i,t} - g_{i,t}\|_2^2 \quad (3)$$

$$\mathcal{L}_{\text{aux},t} = \frac{1}{n} \sum_{i=1}^n \|z_{i,t} * \kappa^T - R(g_{i,t})\|_2^2 \quad (4)$$

式中, λ_{aux} 是 $\mathcal{L}_{\text{aux},t}$ 的超参数, n 是训练集中的样本总数, $y_{i,t}$ 和 $g_{i,t}$ 分别表示第 t 次迭代中第 i 个样本输入重建模块之后,该模块的实际输出图像和期望输出图像, $\|\cdot\|_2$ 是 2 范数, $R(\cdot)$ 是尺寸调整(resize)函数, $z_{i,t}$ 是第 t 次迭代中低质量图像分支的输出,经过一个卷积层后通道数变为 1, κ 是卷积核。

图像分割损失采用加权二元交叉熵(weighted binary cross entropy, WBCE)损失定义(Jadon, 2021),具体为

$$\mathcal{L}_{\text{seg},t} = \frac{1}{n} \sum_{i=1}^n \left\{ -\frac{1}{w \cdot h} \sum_w \sum_h [\beta p_{i,t} \log q_{i,t} + (1 - p_{i,t}) \log (1 - q_{i,t})] \right\} \quad (5)$$

式中, w 和 h 分别是卷积核的宽度和高度, β 是正负样本的权重因子, $p_{i,t}$ 和 $q_{i,t}$ 分别是第 t 次迭代中的分割掩膜图像和图像分割模块输出的图像。

交叉约束损失 $\mathcal{L}_{\text{cross},t}$ 定义为

$$\mathcal{L}_{\text{cross},t} = \frac{1}{n} \sum_{i=1}^n \|S(y_{i,t}) - g_{i,t}\|^2 \quad (6)$$

式中, $S(\cdot)$ 表示图像分割模块的操作。

损失函数的计算流程如图 4 所示。权重系数 λ_{rec} 、 λ_{seg} 和 λ_{cross} 对网络模型的训练和优化至关重要。为了在图像重建与分割两个任务之间保持合理的权

衡,本文采用了动态学习策略,使这些参数在训练过程中自适应调整,而非固定设定。在训练阶段,整个网络模型(包括编码器和解码器)的参数(权重和偏置)通过反向传播算法不断优化,基于损失函数计算出的误差进行调整。每次迭代时,网络根据当前的梯度信息动态调整重建损失、分割损失和交叉约束损失的权重,从而在最小化整体损失的同时,避免某一子任务主导优化,确保各任务的平衡。这种动态学习机制使网络能够在减少输入数据与重构数据之间的差异的同时,保留关键的图像结构信息,进而有效学习输入数据的特征表示并准确重构原始数据。此外,通过实验优化和交叉验证,分析不同初始化权重对模型性能的影响,确保训练过程稳定收敛,从而达到光声成像重建与分割任务的最优效果。

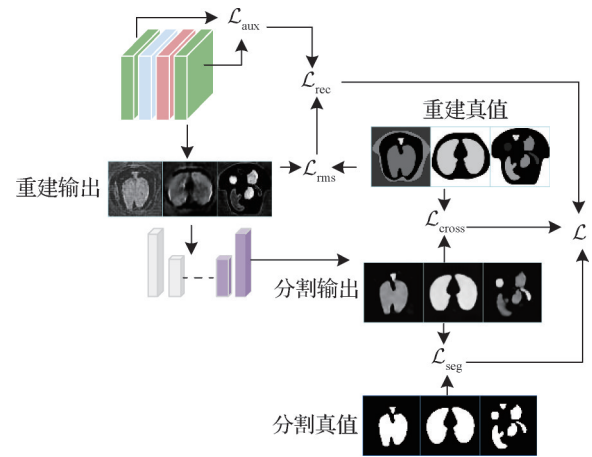


图4 网络损失函数的计算流程

Fig. 4 The calculation process of the network loss function

2 实验及结果分析

2.1 实验数据集

分别构建仿真、体模和在体数据集,用于训练、验证和测试本文搭建的网络模型。

2.1.1 仿真数据

建立包含不同组织成分的数值小鼠仿体,如图 5 所示。参照小鼠不同组织和器官的光学(Cox and Beard, 2005; Stabin 等, 2006)和声学特性参数(全杰等, 1989)设定数值仿体中不同组织成分的特性参数,如表 1 所示。

采用蒙特卡罗方法进行光问题的前向仿真,得到数值仿体各切片内的光辐射通量分布,进而得

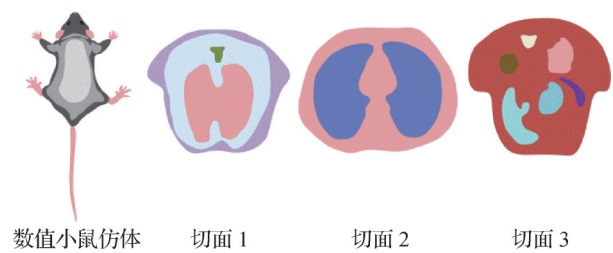


图5 数值小鼠仿体及其切面示例
Fig. 5 Example of the numerical mouse phantom and its cross-sections

到光吸收能量密度的空间分布。利用 MATLAB k-wave 工具包进行声问题的前向仿真,得到各切片内的仿真声压信号。每个切片中成像区域的大小是 38.4 mm × 38.4 mm,探测器采用线性阵列或环形阵列,包含 128 或 256 个阵元,换能器的中心频率为 4 MHz、5 MHz、6 MHz 或 7 MHz,带宽范围为 50%~80%。对仿真光声信号数据进行归一化,以确保模型训练的稳定性和有效性。

为了模拟真实实验数据,在仿真光声信号中

加入了随机高斯白噪声,具体参数如下:均值为 0,方差为 $10^{-8} \sim 10^{-5}$ (对应于信号功率的 $10^{-4}\% \sim 0.1\%$),采样时间为 25 ns (与实验数据的采样频率一致),信噪比为 30~60 dB。这种随机变化的噪声更符合实验环境中的噪声特性,有助于提高仿真数据的真实性。此外,选择高斯白噪声模型是因为实际设备中的噪声(如电子器件的热噪声、探测器的固有噪声、光学散射噪声及背景信号干扰等)具有高斯分布特性,且在工作频段内功率谱密度近似平坦。为了确保模拟噪声与设备真实噪声的一致性,本文参考了设备的实际噪声功率,使仿真中的噪声方差与真实噪声水平一致。信噪比设定为 30 dB~60 dB 的主要依据是实验中的光声信号质量可能受环境因素、探测器性能及激光功率等影响而有所波动。此外,该设定有助于平衡仿真数据的噪声水平,使其既不过于理想化,也不过度劣化信号质量,确保网络模型能够适应实验数据的噪声特征。

表 1 数值小鼠仿体胸腹部切片中不同组织成分的特性参数

Table 1 Property parameters of different tissue types in thoracoabdominal slices of numerical mouse phantoms

器官名称	平均折射率	光吸收系数/cm ⁻¹	光散射系数/cm ⁻¹	各向异性因子	声速/(m/s)	密度/(kg/L)
心脏	1.38	0.82	145	0.93	1 590	1.050
肺	1.36	0.69	120	0.92	1 570	1.045
肾	1.35	0.79	200	0.88	1 575	1.055
脾	1.39	0.55	140	0.94	1 580	1.045
肠	1.38	0.12	90	0.96	1 565	1.048
肝脏	1.36	0.77	95	0.90	1 600	1.055
背景	1.35	0.22	4	0.89	1 530	1.010

2. 1. 2 体模实验数据

采用明胶和琼脂等材料制作实体模型,其中包含不同数量、位置和材料的嵌入物,模拟具有不同光吸收系数的吸收体,部分实物照片如图 6 所示,所用材料及对应的光吸收系数如表 2 所示。采用 MSOT inVision 256 多光谱光声断层扫描成像系统(ithera medical gmbh, munich, germany)扫描体模,该系统采用 Nd: YAG 激光器输出的全波波长(680 nm~900 nm)可调谐激光作为光源,重复率是 10 Hz,脉冲宽度是 7 ns,最大脉冲能量为 120 mJ。超声探测装置由 256 个中心频率为 5 MHz、带宽为 60% 的聚焦超声换能器阵元排列成覆盖角度为 270° 的环形阵

列,采样频率为 40 MSps (即每秒 4 000 万次采样)。沿体模长轴方向进行扫描,扫描长度约为 10 mm,切片间距为 1 mm,共采集了 14 个实体模型的 412 个切片的光声信号数据。

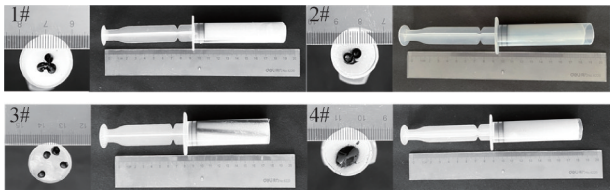


图 6 4 个实验体模的实物照片
Fig. 6 Photographs of four experimental phantoms

表 2 体模材料以及波长为 680 nm 时的光吸收系数
Table 2 Materials used for experimental phantoms and their optical absorption coefficients at a 680 nm

编号	外壳		嵌入物	
	材料	光吸收系数 /cm ⁻¹	材料	光吸收系数 /cm ⁻¹
#1	1% 低熔点琼脂+5% 蓖麻油+10% 乳化剂	0.28	2% 琼脂+5% 蓖麻油+10% 乳化剂+印度墨水	0.32
#2	1% 低熔点琼脂+5% 蓖麻油+10% 乳化剂	0.28	2% 琼脂+5% 蓖麻油+10% 乳化剂+印度墨水	0.32
#3	2% 低熔点琼脂+5% 蓖麻油+10% 乳化剂	0.28	1% 琼脂+5% 蓖麻油+10% 乳化剂+印度墨水	0.32
#4	1% 低熔点琼脂+5% 蓖麻油+10% 乳化剂	0.28	2% 琼脂+5% 蓖麻油+10% 乳化剂+印度墨水	0.32

2. 1. 3 在体实验数据

在体实验数据是利用 MSOT inVision 128 多光谱光声断层扫描系统对麻醉后的活体小鼠进行全身扫描采集的。成像系统采用 Nd:YAG 激光器输出的 715 nm、730 nm、760 nm、800 nm、850 nm 可调谐激光作为光源,在 730 nm 处的最大入射脉冲能量约为 70 mJ,脉冲重复率是 10 Hz,脉冲宽度是 8 ns。超声探测装置由 128 个中心频率为 5 MHz、带宽为 60% 的聚焦超声换能器阵元排列成曲率半径为 40 mm、覆盖角度为 270° 的环形阵列,采样频率为 40 MSps。实验数据由中国医学科学院药物研究所提供,实验过程严格遵守动物伦理标准,确保实验的合规性。实验装置如图 7 所示。在对仰卧和俯卧位的雄性裸鼠进行全身扫描的过程中共采集了 394 个头部、胸部和腹部切片的光声信号。

2. 1. 4 构建数据集

对于仿真数据、体模和在体实验数据,首先分别采用常规 DAS 方法(Xu 等,2003)和最新的光声图像重建网络 DensePANet (Hakimnejad 等,2024)根据各

切片中的声压信号矩阵重建出低质量和高质量初始声压分布图,然后采用最新的医学图像分割网络 GCxt-UNet(Alrfou 和 Zhao,2024)分割各切片图像。对于各个样本,以各切片中的声压信号矩阵和 DAS 重建的低质量图像作为样本的输入,以 DensePANet 重建的高质量图像和 GCxt-UNet 分割的图像作为样本的期望输出,构造数据集。

对于仿真、体模和在体数据集,将所有样本随机打乱之后,分别按照 8:2 的比例划分成训练集和测试集。为了避免产生过拟合,对训练集中的样本图像进行随机旋转(−180°~180°)、随机平移(不超过图像宽度的 10%)和翻转,扩充后的仿真、体模和在体训练集中分别包含 6 083、8 984 和 4 572 个样本。

2. 2 方法的性能评价

2. 2. 1 评价手段和基线方法

本文采用仿真、仿体和在体测试集,通过多角度、多层次的测试,验证所提方法在理想场景和实际应用场景中的有效性和鲁棒性。选择 L-PD(Boink 等,2020)和 IRI-Net(孙美晨,2023)作为基线方法,评估不同方法在 PAT 图像联合重建与分割方面的性能。此外,设计消融实验,验证网络架构的合理性,如表 3 所示。其中,单独消融是指分别与不同的图像重建方法和图像分割方法进行比较,评估本文所提框架中重建模块和分割模块的性能;组合消融是指采用不同重建模块和分割模块按照迭代连接的方式组合成多种联合网络,进而比较本文方法与这些联合网络。

2. 2. 2 评价指标

采用结构相似度(structural similarity, SSIM)和峰值信噪比(peak signal to noise ratio, PSNR)对重建图像质量进行客观评价,SSIM 和 PSNR 越大,说明重

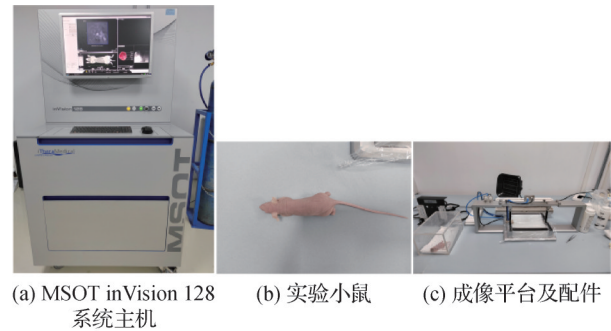


图 7 活体小鼠光声层析扫描实验装置
Fig. 7 PAT experimental setup for living mice
(a) MSOT inVision 128 system console; (b) experimental mouse; (c) imaging platform and related accessories

表3 消融实验设计细节

Table 3 Details of the ablation experiment design

方法	重建模块	分割模块	实验目的
单独消融	SBDM	—	评价重建图像清晰度、背景噪声抑制效果及重建速度
	DPG-Net	—	
	AS-Net	—	
	—	nnU-Net	评价图像分割准确性、边缘保留能力和低对比度区域处理能力
	—	U-Mamba	
	—	CapsNet	
组合消融	本文	GCxt-UNet	验证本文网络结构的合理性和优越性
		SegNet	
	DGD	本文	
		GCxt-UNet	
		SegNet	
	DensePANet	本文	
		SegNet	

注：“—”表示缺少实验。

建图像的失真越小,重建质量越高。

采用DICE指标、Jaccard指数、平均表面对称距离(average symmetric surface distance, ASSD)(Heimann和Meinzer,2009)定量评价图像分割精度。DICE指标和Jaccard指数值越大,ASSD值越小,说明图像分割质量越好。

2.2.3 方法的实现环境

本文实验中,采用MATLAB(R2022a, the Math Works, Inc.)编程实现光学和声学正问题的前向仿真,所用计算机硬件配置为AMD Ryzen 7 5800H CPU、16 GB RAM、Radeon Graphics 3.20 GHz,运行

Windows 11 64位操作系统。卷积神经网络的搭建、训练和测试的实现环境是Ubuntu 20.04,搭载的GPU是NVIDIA GeForce RTX 4090,显存为24 GB,编程语言为Python 3.7,深度学习框架为TensorFlow 2.6.0与Keras 2.0。

2.2.4 超参数设置

采用联合损失函数 $\mathcal{L}$ 对本文方法进行训练,设置学习率为0.000 1,批次处理大小batchsize = 16, epoch = 600,损失函数曲线如图8所示。

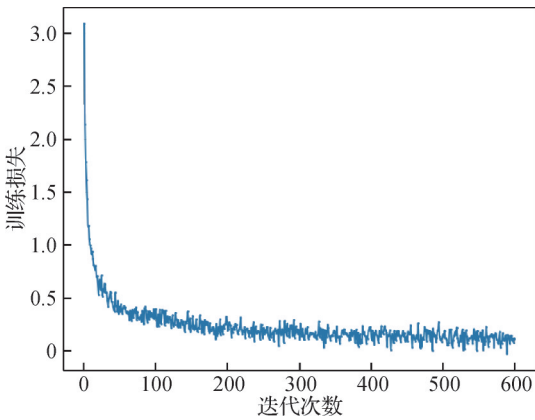


图8 训练网络的损失函数曲线

Fig. 8 Training loss of the network

2.3 实验结果

2.3.1 仿真实验结果

采用本文方法对仿真测试集的实验结果示例如图9所示。由图9可知,经过初次迭代后,重建图像的质量较差,细节模糊,但是随着迭代次数增加,图像中的细节逐渐清晰。每次分割都对重建图像质量产生了正反馈作用,使得目标结构更加清晰和准确。

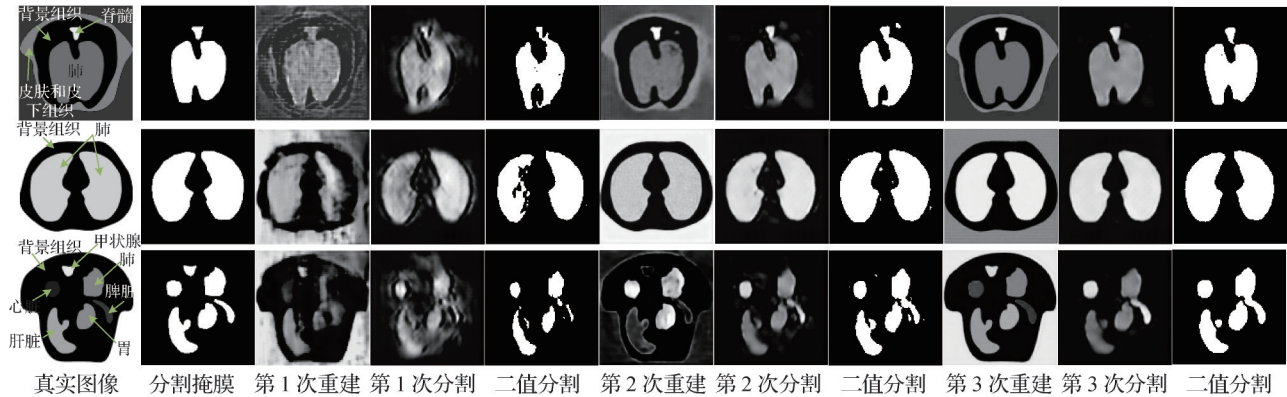


图9 数值小鼠仿体的PAT图像重建与分割结果

Fig. 9 PAT image reconstruction and segmentation results of the numerical mouse phantom

3次迭代以后,图像质量有了明显提高。这一结果表明,正反馈机制在图像重建过程中起了关键作用,通过多次迭代和分割处理,图像清晰度得到了显著提升,有效增强了图像细节。

根据图10和表4所示的评价指标可知,在3种联合重建和分割方法中,IRI-Net的分割性能较好,能够精准地识别并分割图像中的目标区域,但是其重建性能相对较差,这表现在其重建的图像质量和细节还原方面不及其他方法。L-PD的重建性能较好,能够很好地保留图像的结构信息和细节,但是分割表现不佳,准确性和一致性不及IRI-Net和本文方法。本文方法的综合表现最佳,不仅在图像重建中能够有效保留图像结构和细节,还在分割任务中显示出高精度和一致性。

对于单独消融实验,针对图像重建任务,AS-Net虽然架构简单,计算复杂度低,模型训练和推理速度

快,在硬件资源有限的情况下表现出色,但其受限于真值的质量,对复杂场景的适应性弱,需要更多的预处理工作来确保输入图像的质量。DPG-Net能够较好地保留图像结构信息,但其依赖训练数据的先验信息,因而增加了处理时间。SBDM基于概率模型,其重建图像质量接近本文方法,在生成图像时具有较强的鲁棒性和灵活性,但是训练时间长,计算复杂度高,对硬件资源的需求大。此外,模型的生成结果存在随机性,影响结果的一致性。本文方法通过将重建模块与分割模块以循环迭代的方式连接起来,使得重建和分割任务相互促进,提升了整体性能。针对图像分割任务,CapsNet能够捕捉到复杂图像特征,但因其采用胶囊结构,在处理大量数据时的训练时间长,计算资源消耗大。而且其结构较为复杂,对模型参数和超参数的设置非常敏感,因此调整和优化较为困难。nnU-Net可以获得较高的分割精度,能

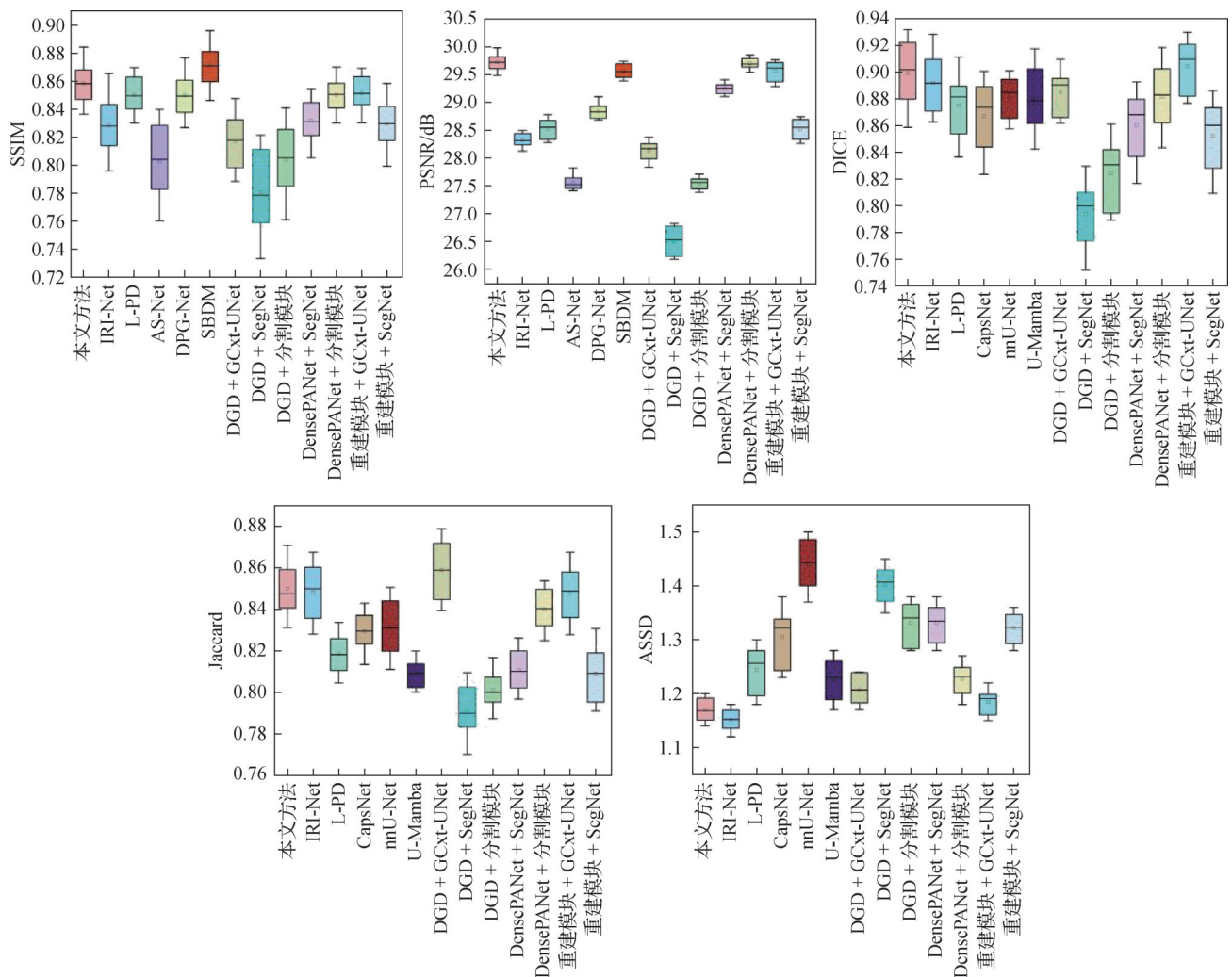


图10 不同方法对仿真数据集全部样本各项评价指标的测试结果

Fig. 10 Test results of evaluation metrics for all samples in the simulation dataset using different methods

表4 本文方法和基线方法对仿真数据集中单个样本进行图像重建与分割的评价指标结果

Table 4 Evaluation metrics for image reconstruction and segmentation of a single sample in the simulation dataset using the proposed method and baseline methods, respectively

方法		SSIM	PSNR/dB	DICE	Jaccard	ASSD	GPU 运行时间/s	训练时间/h
本文		0.858 4	29.724 6	0.898 8	0.850 1	1.170 0	23.73	29.67
IRI-Net		0.828 2	28.319 4	0.891 8	0.848 3	1.151 8	24.68	31.34
L-PD		0.850 1	28.517 3	0.874 9	0.818 6	1.244 5	28.13	34.97
图像重建	AS-Net	0.802 1	27.553 4	—	—	—	15.69	19.36
	DPG-Net	0.849 9	28.832 5	—	—	—	17.91	22.89
	SBDM	0.870 9	29.556 5	—	—	—	19.82	25.07
图像分割	CapsNet	—	—	0.866 9	0.829 4	1.305 3	19.12	23.43
	nnU-Net	—	—	0.880 9	0.831 1	1.439 5	18.52	23.65
	U-Mamba	—	—	0.880 1	0.809 1	1.225 5	18.11	22.83
DGD + GCxt-UNet		0.816 9	28.128 9	0.885 1	0.858 9	1.207 0	22.80	28.53
DGD + SegNet		0.780 3	26.503 2	0.793 5	0.791 4	1.402 1	20.07	25.10
DGD + 本文分割模块		0.803 7	27.538 6	0.824 1	0.801 4	1.331 7	20.96	26.24
DensePANet + SegNet		0.831 8	29.240 3	0.859 8	0.810 9	1.331 2	23.84	31.05
DensePANet + 本文分割模块		0.850 2	29.700 6	0.881 4	0.839 9	1.227 3	26.46	31.94
本文重建模块 + GCxt-UNet		0.851 3	29.557 5	0.904 2	0.847 8	1.184 5	24.45	30.81
本文重建模块 + SegNet		0.829 5	28.517 0	0.851 8	0.808 9	1.322 2	22.82	28.53

注:加粗字体表示各列最优结果(最后两列中是联合方法的最优结果)。“—”表示缺少实验。

够适应不同的数据集和任务,但在处理图像边界区域时,ASSD数值较高,这表明模型在捕捉边界细节时存在不足,导致细节丢失。U-Mamba在处理大规模图像数据时表现出一定优势,然而其泛化能力有限,尽管U-Mamba对体模数据集测试结果的评估指标优于本文方法,但对在体数据集的测试结果明显不如本文方法。

对于组合消融实验,本文方法表现出相对均衡的性能,在重建和分割任务中均表现优异。相对而言,“DensePANet+SegNet”组合在重建方面表现优秀,虽然DensePANet是高性能图像重建模块,但由于SegNet是低性能分割模块,导致其分割性能不佳。该结果表明,分割模块的性能对整体网络的表现有重要影响。“DensePANet + 本文分割模块”和“本文重建模块 + GCxt-UNet”组合也表现出较高的性能,但由于网络复杂度较高,资源消耗明显增加,容易导致过拟合和梯度爆炸等,增加了模型训练的难度和不稳定性,在实际应用中可能带来资源分配和计算效率方面的挑战。此外,“DGD + SegNet”组合在重建和分割任务中的表现均不理想,主要原因是该组

合中的两个模块均为低性能模块,无法满足高精度任务的需求。

2.3.2 体模实验结果

对体模测试集的实验结果如图11所示,评价指标如图12和表5所示。由图11和图12可知,迭代优化策略对模型性能的提升至关重要,通过多次迭代,网络的输出可以逐步逼近最优解,在重建和分割任务中表现出更优的性能。由表5可知,由于IRI-Net的特征提取层能够更好地捕捉目标区域的细节,所以其分割指标较优。L-PD能够有效抑制噪声并恢复图像细节,因此其在图像重建方面具有优势。本文方法通过综合优化,同时进行噪声抑制和细节恢复,因而在重建和分割任务中均表现优异。DensePANet与SegNet或GCxt-UNet的组合在重建和分割任务中的表现存在明显差异。DensePANet具有良好的特征提取和细节恢复能力,因此重建图像的指标较优。然而,由于其网络结构复杂且包含大量参数,因此训练和推理时间较长。与SegNet结合后,尽管联合网络在重建任务中表现优异,但由于SegNet在特征融合和细节处理方面的

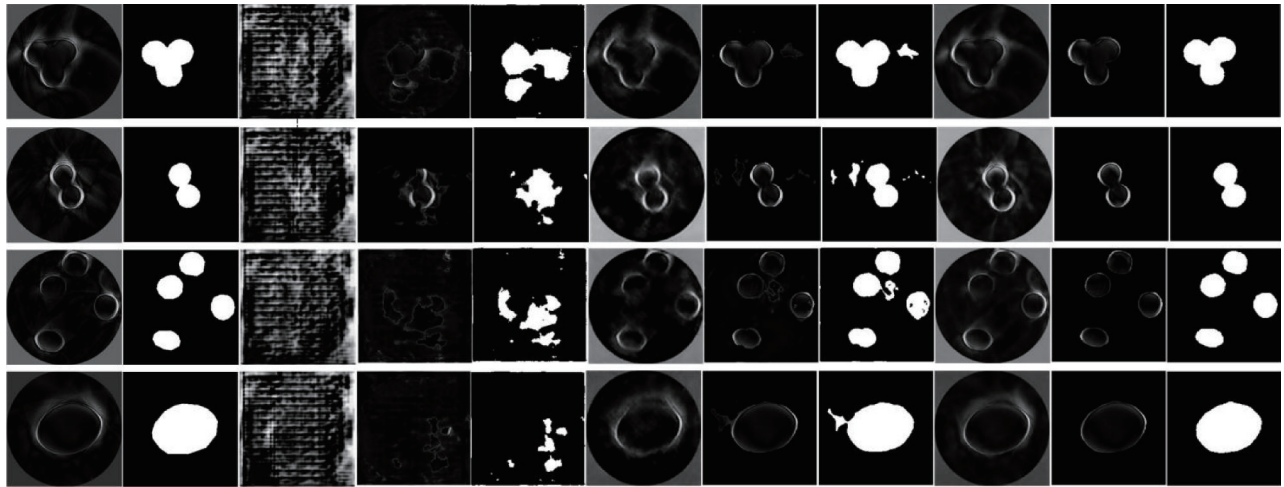


图 11 采用不同方法得到的体模 PAT 图像重建与分割结果

Fig. 11 PAT image reconstruction and segmentation results of the phantom obtained using different methods

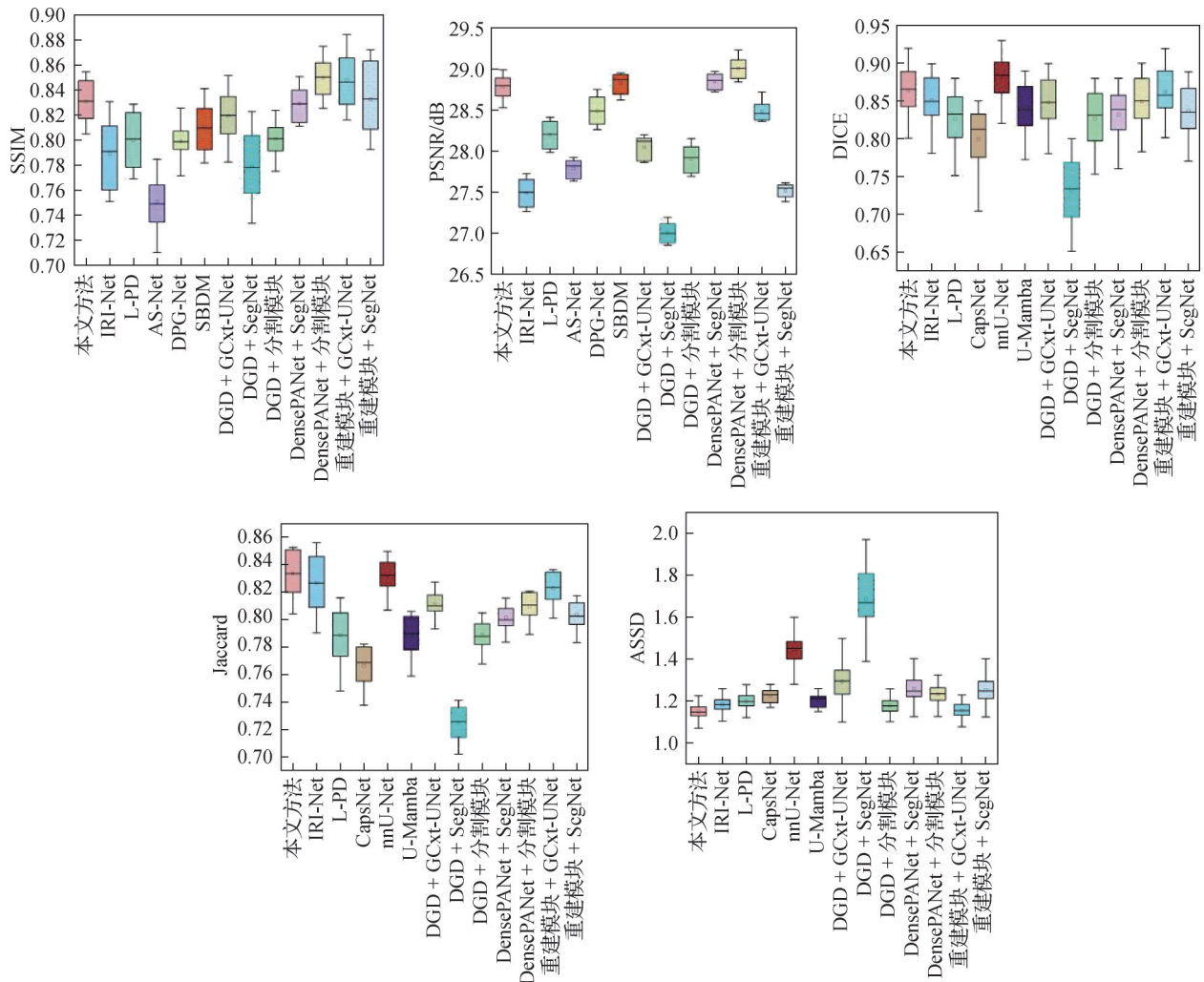


图 12 不同方法对体模数据集全部样本各项评价指标的测试结果

Fig. 12 Test results of evaluation metrics for all samples in the phantom dataset using different methods

不足,使得分割性能较差。与 GCxt-UNet 结合时, GCxt-UNet 的多尺度特征融合和上下文信息增强能力提升了分割性能,使得联合网络的整体表现较优。DGD 是在优化过程中逐步减小梯度,具有较好的收敛性和稳定性,能够有效减少重建图像中的伪

影和噪声,提高图像清晰度。因此, DGD 与 SegNet 的组合在重建任务中表现较好,但在分割任务中表现一般;而与 GCxt-UNet 结合时, GCxt-UNet 在分割任务中的高性能弥补了 DGD 的不足,使得分割性能有所提升。

表 5 本文方法和基线方法对体模数据集中单个样本进行图像重建与分割的各项评价指标结果
Table 5 Evaluation metrics for image reconstruction and segmentation of a single sample in the phantom dataset using both the proposed method and baseline methods

方法		SSIM	PSNR/dB	DICE	Jaccard	ASSD	GPU 运行时间/s	训练时间/h
本文		0.830 7	28.782 9	0.864 1	0.831 5	1.149 9	25.04	43.80
IRI-Net		0.788 8	27.485 1	0.850 9	0.823 5	1.184 2	28.76	47.18
L-PD		0.799 7	28.195 8	0.826 6	0.786 9	1.200 2	30.44	50.53
图像重建	AS-Net	0.750 4	27.784 1	—	—	—	13.88	29.35
	DPG-Net	0.799 2	28.489 9	—	—	—	16.68	33.60
	SBDM	0.810 2	28.821 2	—	—	—	18.12	37.06
图像分割	CapsNet	—	—	0.799 5	0.768 8	1.225 5	18.48	35.85
	nnU-Net	—	—	0.880 2	0.832 4	1.444 4	17.58	34.73
	U-Mamba	—	—	0. 839 7	0.790 7	1.203 1	17.78	34.48
DGD + GCxt-UNet		0.819 0	28.044 0	0.848 1	0.812 2	1.292 6	23.56	42.11
DGD + SegNet		0.778 7	27.044 0	0.732 0	0.724 6	1.687 6	19.51	37.05
DGD + 本文分割模块		0.800 8	27.904 1	0.826 6	0.790 1	1.178 5	20.86	38.73
DensePANet + SegNet		0.822 1	28.845 8	0.832 0	0.802 5	1.260 6	25.13	45.51
DensePANet + 本文分割模块		0.829 8	29.056 0	0.849 0	0.811 2	1.232 1	27.74	47.18
本文重建模块 + GCxt-UNet		0.828 1	28.483 3	0.862 4	0.824 2	1.156 5	26.39	45.46
本文重建模块 + SegNet		0.822 3	27.514 6	0.836 9	0.804 7	1.253 9	23.69	42.12

注:加粗字体表示各列最优结果(最后两列中是联合方法的最优结果)。“—”表示缺少实验。

2. 3. 3 在体实验结果

采用不同方法对活体小鼠胸腹切片的光声信号进行图像重建和分割的部分结果如图 13 所示,评价指标如图 14 和表 6 所示。在体成像数据中包含大量由生理活动(如血流、呼吸和心跳等)所致的噪声和干扰,增加了数据集标注、信号处理和图像重建的难度。对数据集的标注误差和不一致性,会影响深度学习模型的训练效果和图像重建与分割的准确性。此外,实际生物组织的结构和成分复杂且多变,包含血管、脂肪和肌肉等多种不同类型的组织或器官,它们的光学和声学特性具有高度差异性。这种复杂性和异质性在仿真和体模数据集中通常得不到充分体现,因而深度学习模型在这些数据集上可以较容易

地学习和提取特征。但是在处理在体数据时,网络模型的特征提取和识别能力受到限制,导致图像重建和分割性能下降。尽管如此,与基线和消融方法相比,本文方法在处理较复杂的成像信号时仍然显示了优势。

2. 4 讨论

本节结合仿真、体模和在体实验结果,分别讨论迭代次数、网络输入和初始图像生成方法对本文方法性能的影响。

2. 4. 1 迭代次数的影响

本节在网络模型同时输入声压信号和初始低质量图像且初始图像采用 DAS 生成的情况下,讨论网络迭代次数对方法性能的影响。本文方法在仿真测

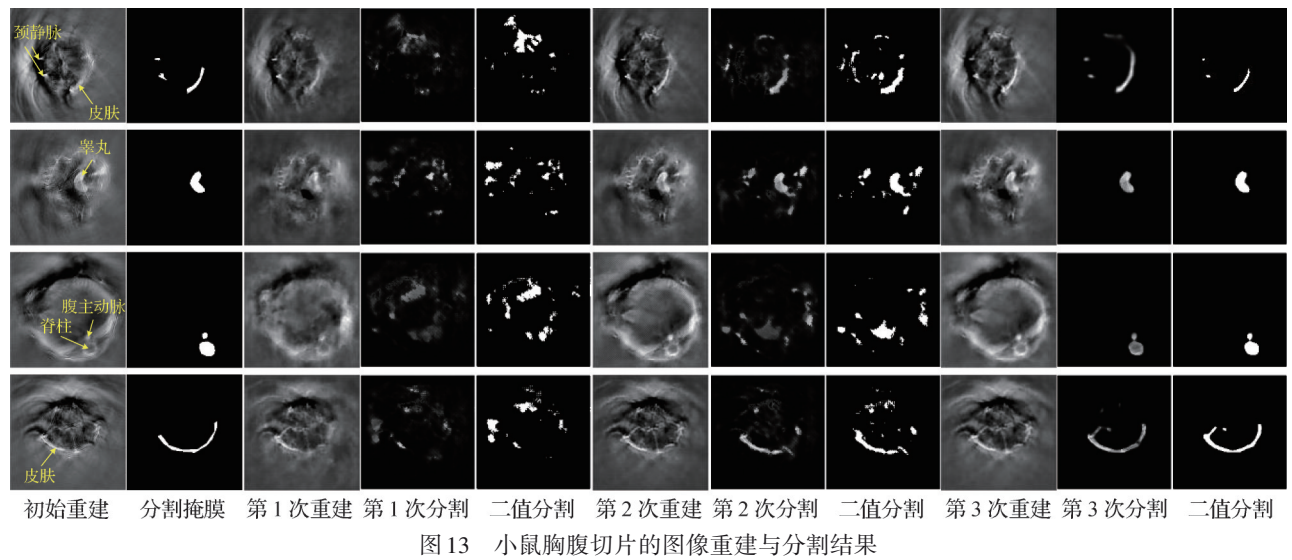


Fig. 13 Image reconstruction and segmentation results of mouse thoracic and abdominal slices

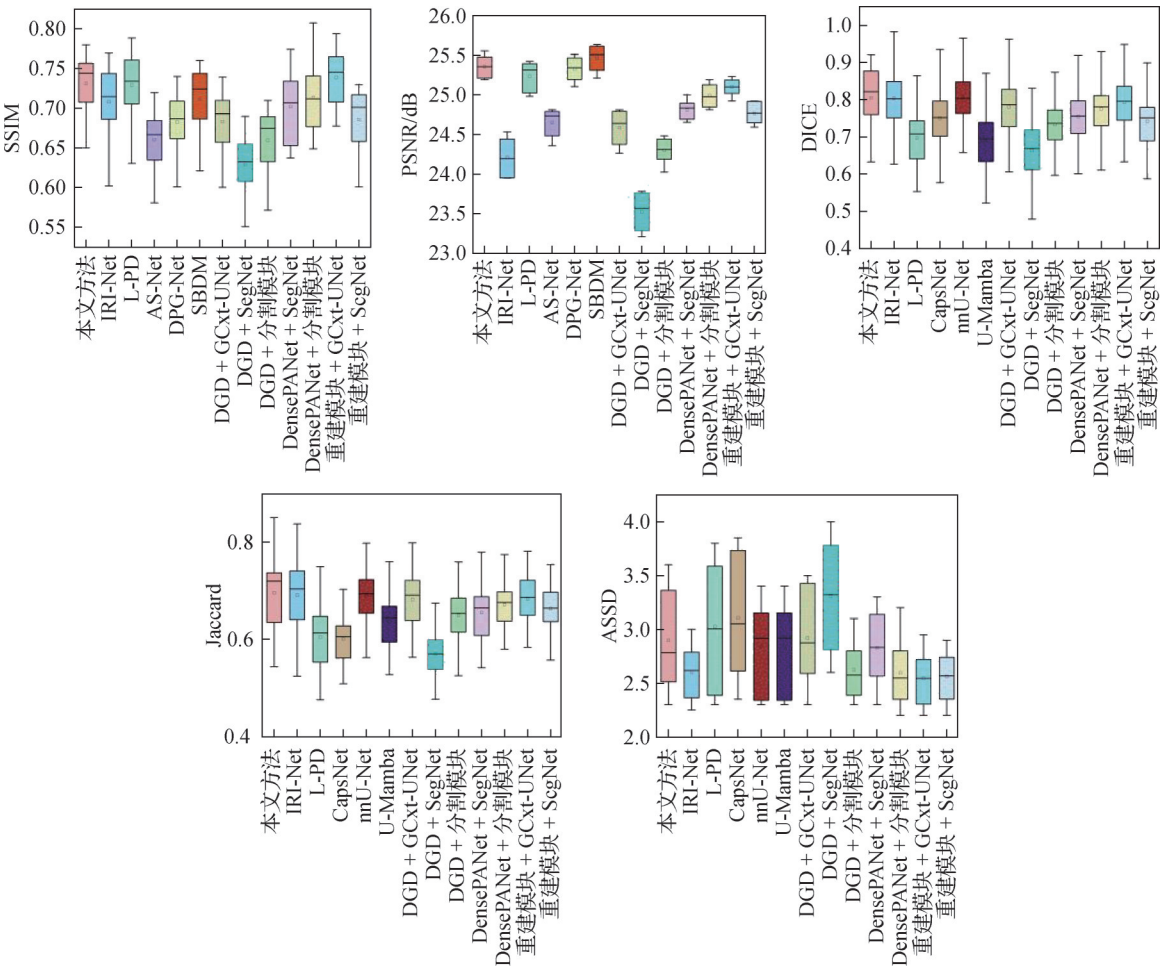


图 14 不同方法对在体数据集全部样本各项评价指标的测试结果

Fig. 14 Test results of evaluation metrics for all samples in the in vivo dataset using different methods

试集上迭代 2 次、3 次、4 次的图像重建和分割结果及各项评价指标结果分别如图 15 和图 16 所示,迭代次数、网络输入和初始图像生成方法对本文方法训练

时长和处理时间的影响如表 7 所示。由图 15、图 16 和表 7 可知,与迭代 2 次相比,迭代 3 次之后重建图像的质量改善明显,SSIM 和 PSNR 分别提高约 15%

表 6 对在体数据集中单个样本进行图像重建与分割的评价指标

Table 6 Evaluation metrics for image reconstruction and segmentation of a single sample in the in vivo dataset							
方法	SSIM	PSNR/dB	DICE	Jaccard	ASSD	GPU 运行时间/s	训练时间/h
本文	0.731 0	25.352 7	0.804 4	0.695 9	2.547 0	27.82	22.29
IRI-Net	0.778 2	24.209 6	0.803 7	0.690 8	2.602 0	30.58	25.71
L-PD	0.728 9	25.232 3	0.696 5	0.604 4	3.026 1	32.91	28.65
图像重建	AS-Net	0.660 2	24.647 3	—	—	21.03	13.29
	DPG-Net	0.682 9	25.323 6	—	—	23.94	17.43
	SBDM	0.711 4	25.457 4	—	—	24.52	18.53
图像分割	CapsNet	—	—	0.689 8	0.600 7	3.105 9	18.57
	nnU-Net	—	—	0.805 7	0.689 6	3.059 0	18.00
	U-Mamba	—	—	0.709 8	0.636 8	2.841 9	17.86
DGD + GCxt-UNet							
DGD + SegNet							
DGD + 本文分割模块							
DensePANet + SegNet							
DensePANet + 本文分割模块							
本文重建模块 + GCxt-UNet							
本文重建模块 + SegNet							

注:加粗字体表示各列最优结果(最后两列中是联合方法的最优结果)。“—”表示缺少实验。

和 18%, DICE 和 Jaccard 分别提高约 14% 和 12%, ASSD 降低的 17%, 但 2 次迭代的处理时间约是 3 次迭代的 0.55 倍。与迭代 3 次相比, 迭代 4 次之后重建图像的质量改善不明显, SSIM 和 PSNR 分别提高约 2.5% 和 1.8%, DICE 和 Jaccard 分别提高约 0.75% 和 0.85%, ASSD 降低约 1.02%, 但 4 次迭代的处理时间约是 3 次迭代的 1.77 倍。为了在图像质量和处理效率之间取得平衡, 本文在搭建网络时最终选择迭代 3 次作为最优结果。

2.4.2 网络输入的影响

本节讨论在迭代 3 次且初始图像采用 DAS 生成的情况下, 网络的输入对方法性能的影响。对于双分支网络架构, 每个分支有独立的编码器, 分别处理声压信号和低质量图像, 这两个分支在独立对各自的输入数据编码后, 在网络的中间层进行融合, 并通过解码器逐步还原出高质量的输出图像。这种设计的优势在于, 能够有效地利用来自不同数据源的互补信息, 从而提高最终的输出质量。然而, 当网络仅接收单一类型的数据(即仅声压信号或仅低质量图像)时, 网络结构保留的单一分支将独立承担编码任务, 而原本用于处理另一种输入数据的分支将失去

表 7 迭代次数、网络输入和初始图像生成方法对本文方法训练时长和处理时间的影响

Table 7 The impact of iteration count, network input, and initial image generation method on the training and processing time of the proposed method			
影响方法性能的因素		GPU 运行时间/s	训练时间/h
迭代次数	2 次	15.30	12.26
	3 次	27.82	22.29
	4 次	49.24	39.46
网络输入	声压信号+低质量图像	27.82	22.29
	仅声压信号	27.09	22.11
	仅低质量图像	27.35	22.07
初始图像生成方法	DAS	25.04	43.6
	FBP	26.06	44.51
	TR	25.48	44.13
	迭代法	26.89	45.21

注:加粗字体表示各列最优结果。

作用。在这种情况下, 网络可以简化为传统的 U-Net 结构, 即保留一个分支并去掉另一个分支, 从而使网

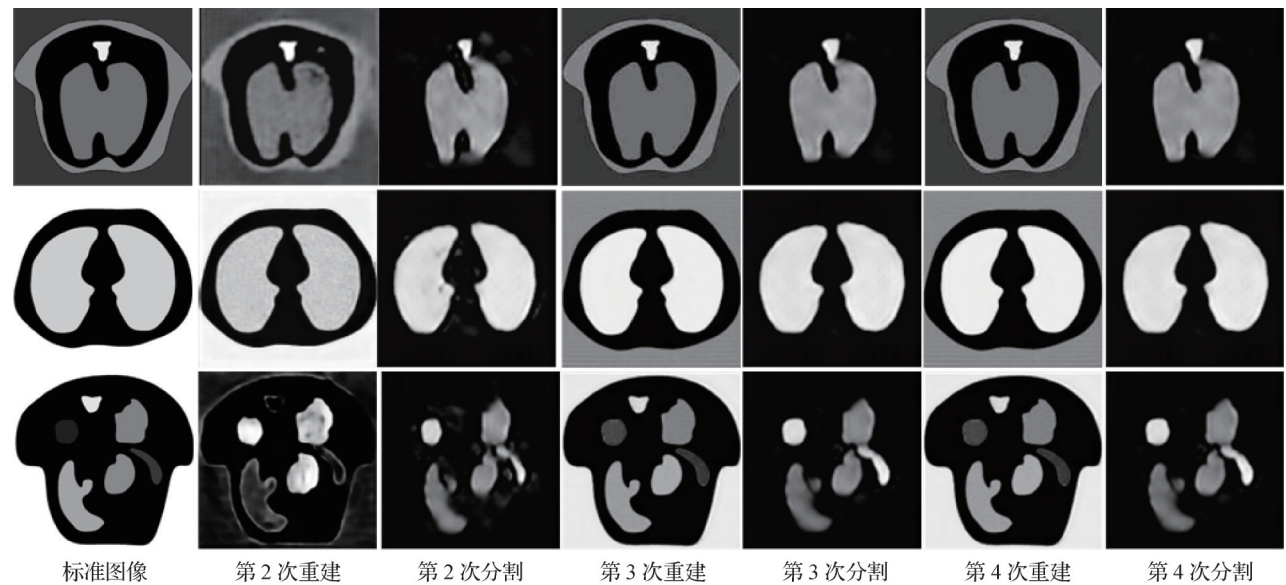


图 15 本文方法在仿真测试集上迭代 2、3、4 次的图像重建和分割结果

Fig. 15 Image reconstruction and segmentation results using the proposed method with two, three, and four iterations for the simulation test set

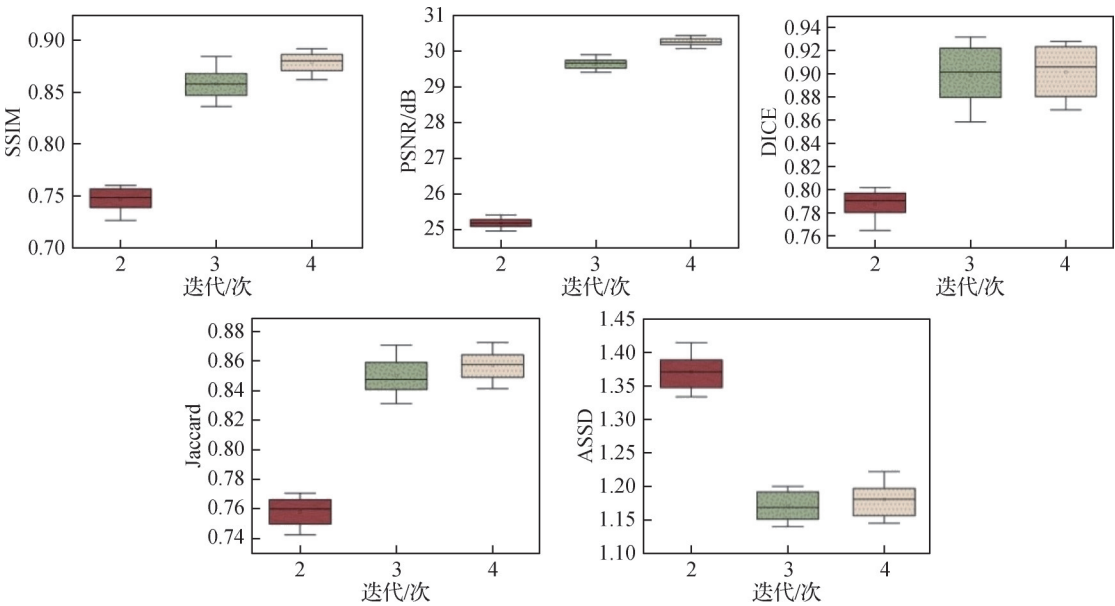


图 16 本文方法在仿真测试集上迭代 2、3、4 次的各项指标结果

Fig. 16 Evaluation metrics results using the proposed method with two, three, and four iterations for the simulation test set

络结构与输入数据相匹配。例如,如果仅使用声压信号作为输入,则可以去除处理低质量图像的分支,反之亦然。

图 17 和图 18 为网络输入不同时,对在体测试数据的图像重建结果和各项评价指标结果。由图可知,当网络同时输入声压测量数据和低质量图像时,重建图像质量显著优于仅输入探测器数据或低质量图像的结果,处理时间则相差不大(如表 7 所示)。这一优势源于本文方法的重建模块能够充分利用两个

输入数据源的信息,有效整合探测器测量数据信息与初始低质量图像的结构信息,显著提高图像的重建质量和细节表现。仅使用声压测量数据的重建结果虽然优于仅使用低质量图像作为输入的结果,但评价指标仍然不及本文方法,这表明虽然声压测量数据中包含了对重建图像至关重要的信息,但是单独使用这些数据进行图像重建仍难以达到最优质量。相较之下,本文方法通过综合两种数据源的优势,提高图像重建的准确性,有效减少了重建过程中可能出现

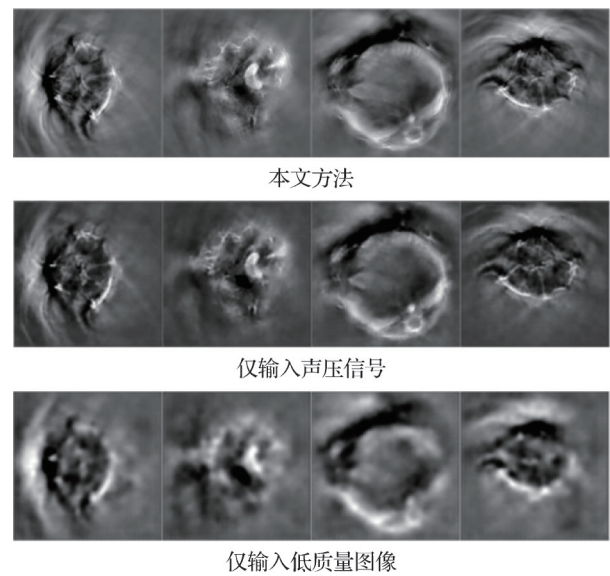


图 17 当网络输入不同时,在体测试样本的图像重建结果
Fig. 17 Image reconstruction results for samples in the in vivo test set with different network inputs

的伪影,显示出更强的抗噪和抗失真的能力。

2.4.3 初始图像质量的影响

在本文搭建的网络模型中,第一个重建模块的

输入是声压测量数据和用常规方法根据原始声压信号重建的低质量图像。本节在迭代3次且网络同时输入声压测量信号和低质量图像的情况下,结合体模实验结果,讨论分别采用DAS、TR、FBP和迭代法生成低质量图像对本文方法性能的影响。由图19可知,初始重建方法对本文方法所得最终图像质量和分割精度影响不大,但是初始重建方法越复杂,处理时间越长,如表7所示。综合考虑处理效率和图像质量,本文最终选择DAS作为初始重建方法。

3 结 论

本文提出一种端到端的深度学习框架,用于光声层析图像的联合重建与分割。与传统的先重建后分割方法不同,该框架实现了重建与分割的同步优化,避免了重建误差在分割阶段的累积。框架包括图像重建模块和图像分割模块,通过双向反馈机制和学习迭代优化策略,使得两者相互促进,从而提升了图像质量和分割精度。

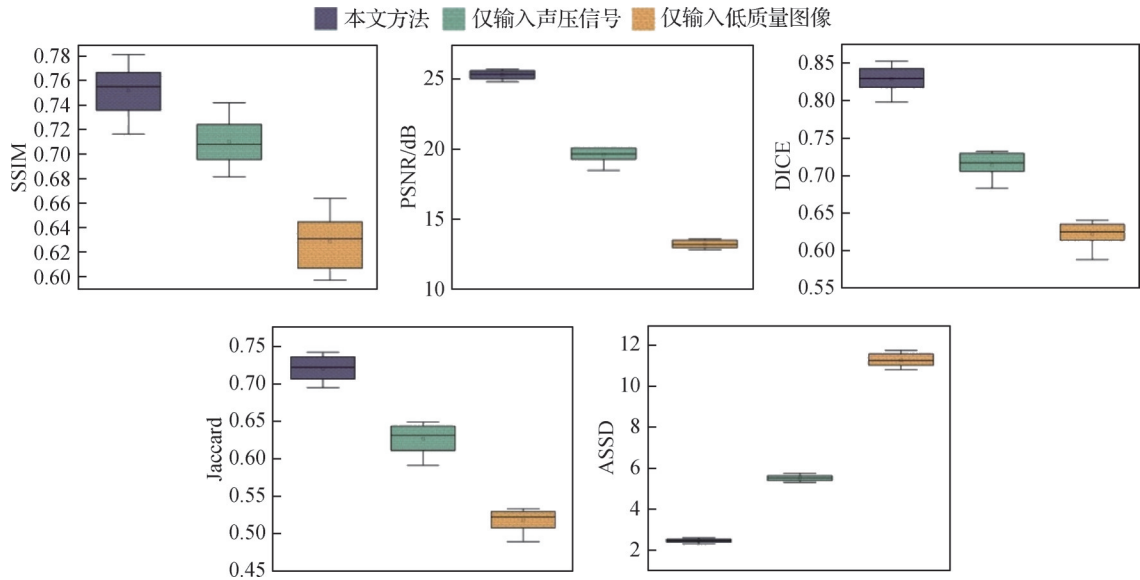


图 18 当网络输入不同时,在体测试样本的各项评价指标结果
Fig. 18 Evaluation metrics results for samples in the in vivo test set with different network inputs

在图像重建模块中,通过充分利用声压测量提供的物理信息和低质量图像的视觉信息,提高重建精度与细节还原能力。图像分割模块结合了FCN和DenseNet结构,实现了像素级的语义分割。FCN用于对重建图像进行像素级分类,得到初步分割结果;DenseNet则提取包括局部和全局语义信息的丰

富特征,通过融合FCN与DenseNet输出的特征图,生成最终的分割结果。

为了减少误差传播并提升精度,框架引入了重建模块与分割模块之间的双向反馈机制。在该机制下,初步重建图像通过分割模块进行处理后,反馈至重建模块以优化重建过程。重建模块根据反馈信息

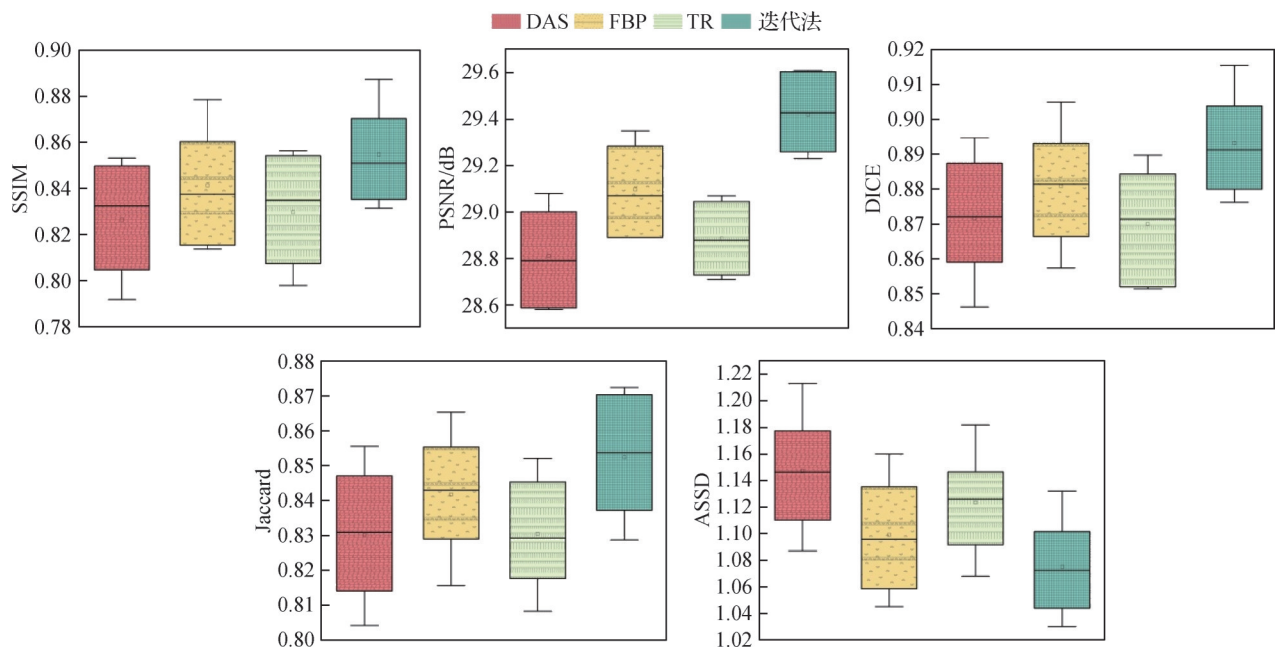


图19 使用不同方法生成初始图像时,采用本文方法对体模测试集中全部样本的图像重建和分割指标统计结果

Fig. 19 Evaluation metrics for image reconstruction and segmentation of all samples in the phantom test set using the proposed method with initial images generated by different methods

逐步修正误差,提高图像重建精度,而分割模块则基于优化后的重建图像获得更清晰、准确的特征,从而提高分割精度。通过模块间的协同优化,整个框架能够在多次迭代中逐步提升重建与分割效果,最终实现高质量的图像重建与精确的分割。

对仿真、体模和在体数据集的测试结果表明,该方法在仿真数据集上能精确重建和分割目标区域,在体模和在体数据集上展现出较强的抗噪能力和对复杂组织结构的适应性,整体性能优于传统的独立重建与分割方法。然而,在对活体小鼠全身扫描的应用场景中,该方法在处理接近或重叠的复杂结构时可能会丢失部分细节。未来工作将集中于优化网络结构和训练策略,以提升方法的鲁棒性并减少计算资源消耗。

参考文献(References)

- Alrfou K and Zhao T. 2024. GCTX-UNet: efficient network for medical image segmentation [EB/OL]. [2024-12-12]. <https://arxiv.org/pdf/2406.05891.pdf>
- Antholzer S, Haltmeier M and Schwab J. 2019. Deep learning for photoacoustic tomography from sparse data. *Inverse Problems in Science and Engineering*, 27 (7): 987-1005 [DOI: 10.1080/17415977.2018.1518444]
- Arridge S, Beard P, Betcke M, Cox B, Huynh N, Lucka F, Ogunlade O and Zhang E. 2016. Accelerated high-resolution photoacoustic tomography via compressed sensing. *Physics in Medicine and Biology*, 61(24): 8908-8940 [DOI: 10.1088/1361-6560/61/24/8908]
- Badrinarayanan V, Kendall A and Cipolla R. 2017. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12): 2481-2495 [DOI: 10.1109/TPAMI.2016.2644615]
- Bal G and Uhlmann G. 2010. Inverse diffusion theory of photoacoustics. *Inverse Problems*, 26(8): #085010 [DOI: 10.1088/0266-5611/26/8/085010]
- Boink Y E, Manohar S and Brune C. 2020. A partially-learned algorithm for joint photo-acoustic reconstruction and segmentation. *IEEE Transactions on Medical Imaging*, 39(1): 129-139 [DOI: 10.1109/TMI.2019.2922026]
- Chlis N K, Karlas A, Fasoula N A, Kallmayer M, Eckstein H H, Theis F J, Ntziachristos V and Marr C. 2020. A sparse deep learning approach for automatic segmentation of human vasculature in multi-spectral optoacoustic tomography. *Photoacoustics*, 20: #100203 [DOI: 10.1016/j.pacs.2020.100203]
- Corona V, Benning M, Ehrhardt M J, Gladden L F, Mair R, Reci A, Sederman A J, Reichelt S and Schönlieb C B. 2019. Enhancing joint reconstruction and segmentation with non-convex Bregman iteration. *Inverse Problems*, 35(5): #055001 [DOI: 10.1088/1361-6420/ab0b77]
- Cox B T and Beard P C. 2005. Fast calculation of pulsed photoacoustic fields in fluids using k -space methods. *The Journal of the Acoustical Society of America*, 117(6): 3616-3627 [DOI: 10.1121/1.1920227]
- Dey S, Saha S, Feng B T, Cui M X, Delisle L, Leong O, Wang L V

- and Bouman K L. 2024. Score-based diffusion models for photoacoustic tomography image reconstruction [EB/OL]. [2024-03-30]. <https://arxiv.org/pdf/2404.00471.pdf>
- Farnia P, Mohammadi M, Najafzadeh E, Alimohamadi M, Makkiabadi B and Ahmadian A. 2020. High-quality photoacoustic image reconstruction based on deep convolutional neural network: towards intra-operative photoacoustic imaging. *Biomedical Physics and Engineering Express*, 6(4): #045019 [DOI: 10.1088/2057-1976/ab9a10]
- Guo M J, Lan H R, Yang C C, Liu J and Gao F. 2022. AS-Net: fast photoacoustic reconstruction with multi-feature fusion from sparse data. *IEEE Transactions on Computational Imaging*, 8: 215-223 [DOI: 10.1109/tci.2022.3155379]
- Hakimnejad H, Azimifar Z and Goshtasbi N. 2024. DensePANet: an improved generative adversarial network for photoacoustic tomography image reconstruction from sparse data [EB/OL]. [2024-04-19]. <https://arxiv.org/pdf/2404.13101.pdf>
- Heimann T and Meinzer H P. 2009. Statistical shape models for 3D medical image segmentation: a review. *Medical Image Analysis*, 13(4): 543-563 [DOI: 10.1016/j.media.2009.05.004]
- Isensee F, Jaeger P F, Kohl S A A, Petersen J and Maier-Hein K H. 2021. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2): 203-211 [DOI: 10.1038/s41592-020-01008-z]
- Jadon S. 2021. SemSegLoss: a python package of loss functions for semantic segmentation. *Software Impacts*, 9: #100078 [DOI: 10.1016/j.simpa.2021.100078]
- Joseph F K, Arora A, Kancharla P, Singh M K A, Steenbergen W and Channappayya S S. 2021. Generative adversarial network-based photoacoustic image reconstruction from bandlimited and limited-view data//Proceedings of 2021 SPIE International Conference on Photons Plus Ultrasound: Imaging and Sensing. [s.l.]: SPIE: #1164235 [DOI: 10.1117/12.2577750]
- Lafci B, Mercep E, Morscher S, Deán-Ben X L and Razansky D. 2020. Efficient segmentation of multi-modal optoacoustic and ultrasound images using convolutional neural networks//Proceedings of 2020 SPIE International Conference on Photons Plus Ultrasound: Imaging and Sensing. San Francisco, USA: SPIE: #112402N [DOI: 10.1117/12.2543970]
- Lafci B, Mercep E, Morscher S, Deán-Ben X L and Razansky D. 2021. Deep learning for automatic segmentation of hybrid optoacoustic ultrasound (OPUS) images. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 68(3): 688-696 [DOI: 10.1109/TUFFC.2020.3022324]
- Lalonde R, Xu Z Y, Irmakci I, Jain S and Bagci U. 2021. Capsules for biomedical image segmentation. *Medical Image Analysis*, 68: #101889 [DOI: 10.1016/j.media.2020.101889]
- Li J, Guo S W, Meng X T, Lai Z C and Han S K. 2022. DPG-Net: densely progressive-growing network for point cloud completion. *Neurocomputing*, 491: 1-13 [DOI: 10.1016/j.neucom.2022.03.060]
- Liu F Y, Gong X J, Wang L V, Guan J J, Song L and Meng J. 2019. Dictionary learning sparse-sampling reconstruction method for in-vivo 3D photoacoustic computed tomography. *Biomedical Optics Express*, 10(4): 1660-1677 [DOI: 10.1364/BOE.10.001660]
- Lu T, Chen T T, Gao F, Sun B, Ntziachristos V and Li J. 2021. LV-GAN: a deep learning approach for limited-view optoacoustic imaging based on hybrid datasets. *Journal of Biophotonics*, 14(2): #e202000325 [DOI: 10.1002/jbio.202000325]
- Lu X M. 2021. Research and Quantitative Analysis of 3D Segmentation of Vascular Photoacoustic Images Based on the Hessian Matrix. Guiyang, China: Guizhou University (卢学敏. 2021. 基于Hessian矩阵的血管光声图像三维分割研究及量化分析. 贵阳: 贵州大学) [DOI: 10.27047/d.cnki.ggudu.2021.003078]
- Lungu G F, Li M L, Xie X Y, Wang L V and Stoica G. 2007. In vivo imaging and characterization of hypoxia-induced neovascularization and tumor invasion. *International Journal of Oncology*, 30(1): 45-54 [DOI: 10.3892/ijo.30.1.45]
- Ly C D, Nguyen V T, Vo T H, Mondal S, Park S, Choi J, Vu T T H, Kim C S and Oh J. 2022. Full-view in vivo skin and blood vessels profile segmentation in photoacoustic imaging based on deep learning. *Photoacoustics*, 25: #100310 [DOI: 10.1016/j.pacs.2021.100310]
- Ma J, Li F F and Wang B. 2024. U-Mamba: enhancing long-range dependency for biomedical image segmentation [EB/OL]. [2024-01-09]. <https://arxiv.org/pdf/2401.04722.pdf>
- Madhumathy P and Pandey D. 2022. Deep learning based photo acoustic imaging for non-invasive imaging. *Multimedia Tools and Applications*, 81(5): 7501-7518 [DOI: 10.1007/s11042-022-11903-6]
- Mandal S, Viswanath P S, Yesashwini N, Deán-Ben X L and Razansky D. 2015. Multiscale edge detection and parametric shape modeling for boundary delineation in optoacoustic images//Proceedings of the 37th IEEE Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Milan, Italy: IEEE: 707-710 [DOI: 10.1109/EMBC.2015.7318460]
- Oh J T, Li M L, Zhang H F, Maslov K and Wang L V. 2006. Three-dimensional imaging of skin melanoma in vivo by dual-wavelength photoacoustic microscopy. *Journal of Biomedical Optics*, 11(3): #034032 [DOI: 10.1117/1.2210907]
- Paltauf G, Viator J A, Pohl S A and Jacques S L. 2002. Iterative reconstruction algorithm for optoacoustic imaging. *The Journal of the Acoustical Society of America*, 112(4): 1536-1544 [DOI: 10.1121/1.1501898]
- Peng D F, Zhai C C, Zhang Y J and Guan H Y. 2023. High-resolution optical remote sensing image change detection based on dense connection and attention feature fusion network. *The Photogrammetric Record*, 38(184): 498-519 [DOI: 10.1111/phor.12462]
- Poudel J, Yang L and Anastasio M A. 2019. A survey of computational frameworks for solving the acoustic inverse problem in three-dimensional photoacoustic computed tomography. *Physics in Medicine and Biology*, 64(14): #14TR01 [DOI: 10.1088/1361-6560/ab2017]
- Stabin M G, Peterson T E, Holburn G E and Emmons M A. 2006. Voxel-

- based mouse and rat models for internal dose calculations. *The Journal of Nuclear Medicine*, 47(4): 655-659
- Sun M C. 2023. Research on Photoacoustic Tomography Image Enhancement and Tissue Characterization. Baoding: North China Electric Power University (孙美晨. 2023. 光声层析图像增强与组织标定方法研究. 保定: 华北电力大学) [DOI: 10.27139/d.cnki.ghbdu.2023.000412]
- Sun M J, Li C, Chen N B, Zhao H X, Ma L Y, Liu C B, Shen Y, Lin R Q and Gong X J. 2020. Full three-dimensional segmentation and quantification of tumor vessels for photoacoustic images. *Photoacoustics*, 20: #100212 [DOI: 10.1016/j.pacs.2020.100212]
- Sun W W and Wang R S. 2018. Fully convolutional networks for semantic segmentation of very high resolution remotely sensed images combined with DSM. *IEEE Geoscience and Remote Sensing Letters*, 15(3): 474-478 [DOI: 10.1109/LGRS.2018.2795531]
- Sun Z and Sun L S. 2021. Simultaneous reconstruction of optical absorption property and speed of sound in intravascular photoacoustic tomography. *Inverse Problems in Science and Engineering*, 29(12): 1764-1788 [DOI: 10.1080/17415977.2021.1879805]
- Sun Z, Wang X Y and Yan X Y. 2021. An iterative gradient convolutional neural network and its application in endoscopic photoacoustic image formation from incomplete acoustic measurement. *Neural Computing and Applications*, 33(14): 8555-8574 [DOI: 10.1007/s00521-020-05607-x]
- Sun Z and Yan X Y. 2020a. A deep learning method for limited-view intravascular photoacoustic image reconstruction. *Journal of Medical Imaging and Health Informatics*, 10(11): 2707-2713 [DOI: 10.1166/jmihi.2020.3204]
- Sun Z and Yan X Y. 2020b. Image reconstruction based on compressed sensing for sparse-data endoscopic photoacoustic tomography. *Computers in Biology and Medicine*, 116: #103587 [DOI: 10.1016/j.combiomed.2019.103587]
- Tong J, Dong Y W and Sun Y C. 1989. Relations between acoustic non-linearity parameter and sound speed and tissue composition. *Chinese Journal of Biomedical Engineering*, 8(3): 190-195 (全杰, 董彦武, 孙永臣. 1989. 非线性声参量B/A与组织声速及组份的关系. 中国生物医学工程学报, 8(3): 190-195)
- Xu M H and Wang L V. 2005. Universal back-projection algorithm for photoacoustic computed tomography. *Physical Review E*, 71: #016706 [DOI: 10.1117/12.589146]
- Xu M H, Xu Y and Wang L V. 2003. Time-domain reconstruction algorithms and numerical simulations for thermoacoustic tomography in various geometries. *IEEE Transactions on Biomedical Engineering*, 50(9): 1086-1099 [DOI: 10.1109/TBME.2003.816081]
- Yang C C, Lan H R, Gao F and Gao F. 2021. Review of deep learning for photoacoustic imaging. *Photoacoustics*, 21: #100215 [DOI: 10.1016/j.pacs.2020.100215]
- Yang L, Park S, Anis F, Su R, Oraevsky A A and Anastasio M A. 2019. Analysis of the use of unmatched backward operators in iterative image reconstruction with application to three-dimensional optoacoustic tomography. *IEEE Transactions on Computational Imaging*, 5(3): 437-449 [DOI: 10.1109/TCI.2019.2895217]
- Yao J J and Wang L V. 2018. Recent progress in photoacoustic molecular imaging. *Current Opinion in Chemical Biology*, 45: 104-112 [DOI: 10.1016/j.cbpa.2018.03.016]
- Yao J J, Xia J, Maslov K I, Nasirivanaki M, Tsytarev V, Demchenko A V and Wang L V. 2013. Noninvasive photoacoustic computed tomography of mouse brain metabolism in vivo. *NeuroImage*, 64: 257-266 [DOI: 10.1016/j.neuroimage.2012.08.054]
- Zhang X F, Zhang S, Zhang D H and Liu R. 2023. Group attention-based medical image segmentation model. *Journal of Image and Graphics*, 28(10): 3231-3242 (张学峰, 张胜, 张冬晖, 刘瑞. 2023. 引入分组注意力的医学图像分割模型. 中国图象图形学报, 28(10): 3231-3242) [DOI: 10.11834/jig.220748]
- Zhao H X, Wang G S, Lin R Q, Gong X J, Song L, Li T, Wang W J, Zhang K Y, Qian X Q, Zhang H X, Li L, Liu Z C and Liu C B. 2018. Three-dimensional hessian matrix-based quantitative vascular imaging of rat iris with optical-resolution photoacoustic microscopy in vivo. *Journal of Biomedical Optics*, 23(4): #046006 [DOI: 10.1117/1.JBO.23.4.046006]
- Zhao Y Q, Wang X H, Wang X F and Shih F Y. 2014. Retinal vessels segmentation based on level set and region growing. *Pattern Recognition*, 47(7): 2437-2446 [DOI: 10.1016/j.patcog.2014.01.006]
- Zhou T, Dong Y L, Huo B Q, Liu S and Ma Z J. 2021. U-Net and its applications in medical image segmentation: a review. *Journal of Image and Graphics*, 26(9): 2058-2077 (周涛, 董雅丽, 霍兵强, 刘珊, 马宗军. 2021. U-Net网络医学图像分割应用综述. 中国图象图形学报, 26(9): 2058-2077) [DOI: 10.11834/JIG.200704]

作者简介

艾冰, 女, 硕士研究生, 主要研究方向为深度学习和光声图像重建与分割技术。E-mail: aibing1232022@163.com

孙正, 通信作者, 女, 教授, 主要研究方向为多模态成像技术、图像重建和反问题求解。E-mail: sunzheng@ncepu.edu.cn

候英飒, 女, 硕士研究生, 主要研究方向为深度学习和光声图像重建技术。E-mail: houyingsa@163.com

孙美晨, 女, 硕士研究生, 主要研究方向为深度学习和图像重建技术。E-mail: 32013393@qq.com