· 综述 ·

基于深度学习的医学图像融合方法的研究进展

张毅　刘柳　王梦　谢文晖
上海交通大学医学院附属胸科医院核医学科，上海 200030
通信作者：谢文晖，Email：shxknuclear@126.com

【摘要】 医学图像融合技术通过结合不同模态的医学图像优势，提供更全面、精准的影像信息，广泛应用于临床诊断与疾病研究中。传统融合方法依赖信号处理技术，但其局限性显著，尤其在多模态图像的融合中难以提取深层次语义信息。随着深度学习的引入，基于卷积神经网络（CNN）、生成对抗网络（GAN）、深层特征提取器、自注意力机制等技术，图像融合的效果得到了显著提升。然而，深度学习方法仍面临数据稀缺、模态异质性、计算资源需求高、模式崩塌与训练不稳定等挑战。该文分析了上述问题，并探讨了相应的解决方案以及未来可能的发展方向。

【关键词】 深度学习；图像处理，计算机辅助；发展趋势
基金项目：国家自然科学基金（82372007，82272044，82202219，82302239）；上海市探索者计划（23TS1400900）
DOI：10.3760/cma.j.cn321828-20241201-00414

Research progress of medical image fusion methods based on deep learning

Zhang Yi, Liu Liu, Wang Meng, Xie Wenhui
Department of Nuclear Medicine, Shanghai Chest Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200030, China
Corresponding author: Xie Wenhui, Email: shxknuclear@126.com

【**Abstract**】 Medical image fusion technology combines the advantages of different modal medical images to provide more comprehensive and precise imaging information, widely applied in clinical diagnosis and disease research. Traditional fusion methods rely on signal processing techniques, but they have significant limitations, especially in multi-modal image fusion, where extracting deep semantic information is challenging. With the introduction of deep learning, image fusion effects have been significantly enhanced through technologies such as convolutional neural networks (CNN), generative adversarial networks (GAN), deep feature extractors, and self-attention mechanisms. However, deep learning methods still face challenges such as data scarcity, modal heterogeneity, high computational resource requirements, mode collapse, and training instability. This paper analyzes these issues and explores corresponding solutions as well as potential future development directions.

【**Key words**】 Deep learning; Image processing, computer-assisted; Trends
Fund program: National Natural Science Foundation of China (82372007, 82272044, 82202219, 82302239); Shanghai Explorer Program (23TS1400900)
DOI：10.3760/cma.j.cn321828-20241201-00414

医学影像学是临床医学诊断的主要技术之一，也是当前医学研究的一个重要方向。由于成像原理和特点的不同，不同模式的医学图像通常有其特定的应用场景。例如，CT 图像可清晰地显示密度高的器官（如骨骼），但不适合捕捉软组织的细节；MRI 则适用于软组织的无创高分辨率检查，但显示骨骼清晰度不如 CT；PET 能够提供关于器官功能的信息，特别是代谢和生理活动方面的细节；而 SPECT 主要用于揭示特定位置的血流和功能状态，尤其是在心血管和脑血流评估中具有重要作用。然而，单模态影像往往无法提供足够的信息，因此，近年来许多研究者致力于开发多模态医学图像融合方法。例如，PET/MR 图像同时提供位置和功能信息；SPECT/MR 图像结合了软组织图像与血流信息；PET/CT 图像则同时显示了功能和解剖信息。多模态医学图像由于其可靠性和全面性，在疾病诊断、临床分期、健康检查等方面具有广泛的应用。

传统的图像融合方法主要依赖于信号处理和数学变换技术，如拉普拉斯金字塔、小波变换和主成分分析等，其通过提取图像的空间、频率或统计特征，结合人工设计的规则，将多模态图像的关键信息进行组合[1]。尽管这些方法实现简单、计算效率较高，但仍存在显著的局限性：特征提取能力有限，难以捕捉图像深层次的语义信息，尤其是在多模态图像融合中效果不佳[2]。此外，这些方法对复杂场景的适应性较弱，在噪声或伪影较多的情况下易受干扰。而人工设计的特征和权重分配不仅增加了主观性，也限制了算法的通用性[3]。

深度学习基于数据驱动，通过构建神经网络自动学习图像的多层次特征，从而全面捕捉模态间的互补信息。诸如卷积神经网络（convolutional neural networks, CNN）、生成对抗

网络(generative adversarial networks, GAN)、深层特征提取器[如用于生物医学图像分割的深度学习网络架构(convolutional networks for biomedical image segmentation, U-Net)、视觉几何组提出的 19 层 CNN(visual geometry group 19-layer CNN, VGG-19)和残差网络(residual network, ResNet)]和自注意力机制(如 Transformer)GAN,均能够有效满足多种模态图像的融合需求。本文就基于深度学习的图像融合方法的研究进展进行综述。

1. CNN。为提高深度网络训练效率,Xia 等[4]进一步优化了深度 CNN 的训练流程。通过 He 初始化方法(He initialization)优化卷积核,并利用反向传播训练基本单元,以堆叠自编码器的方式逐层训练深度神经网络,从底层改进 CNN 模型的训练性能,显著提升了 PET/CT 图像的融合质量。在此基础上,Li 等[5]进一步关注融合过程中边界模糊的问题,提出深度回归配对学习方法,将源图像输入由卷积块和残差块构成的 CNN 中,提取浅层与深层特征后生成加权映射,通过点乘和加权求和实现融合图像的生成。此种方法强化了边界细节的表达能力,使得融合图像在细节和语义信息上更为均衡。此外,Wang 等[6]提出结合非下采样轮廓波变换与 CNN 的融合方法,通过将图像分解为低频和高频子带分别处理,显著提升融合效果,但计算成本较高。Xia 等[4]进一步采用拉普拉斯和高斯滤波器对图像进行分解,通过反向传播训练 CNN 进行融合,增强了图像对比度与清晰度,但对滤波器依赖性较强,通用性受到限制。Liu 等[7]结合连体 CNN 与多尺度金字塔策略,设计了一种符合人类视觉处理方式的 SPECT/CT 融合算法,但该方法需针对不同模态进行大量调参。Wang 等[8]则提出对比金字塔与 CNN 相结合的通用算法,通过局部相似性策略提升融合质量,但时间复杂度较高。最后,Xu 和 Ma[9]针对信息失真提出一种无监督的增强医学图像融合网络,通过浅层与深层特征约束增强信息保存,尤其适用于功能性和结构性图像融合,但其无监督框架可能在特定任务中限制精确性。

2. GAN。初期,基于经典 GAN 的图像融合方法直接采用标准的 GAN 架构,包括一个生成器和一个判别器。生成器利用 2 幅源图像的特征信息生成融合图像,鉴别器则用于鉴别生成图像和原图像的真伪。生成器和鉴别器相互循环,直到生成器获得最佳结果。在实际应用中,Ma 等[10]首次将 GAN 用于图像融合并取得了较好的效果,但在保留 2 幅源图像的不同细节方面存在问题。Wang 等[11]提出了基于 GAN 的形变不变跨域信息融合医学图像合成框架,在网络中集成了一个改进的可变形卷积层,避免了循环 GAN 损失和图像对准损失之间的冲突,并提出了相关的形变不变循环一致性损失函数,对不同域的图像合成表现出更好的鲁棒性。冯莉娟等[12]则使用 GAN 在一定程度上改善了儿童低剂量 PET/CT 图像质量。Guo 等[13]提出一种基于条件 GAN 的新型图像融合技术,该模型特别设计了一个连体网络结构的生成器,以处理图像融合任务中的双输入和单输出的需求。Wang 等[14]则针对条件 GAN 的颜色信息损失的问题提出了用于多焦点图像融合的新型 GAN,使用 6 个基于注意力的并行网络从彩色图像中提取所有通道的特征,但依然存在 2 个关键问题需解决:一是如何避免模式崩塌;二是如何平衡生成器和鉴别器的训练程度,提高训练过程的稳定性。研究者针对这 2 个问题提出了改进方案,例如:Radford 等[15]提出的深度卷积 GAN,该方法通过对鉴别器和生成器架构进行实验优化,最终找到了较为理想的网络架构设置,但这一方法只能在一定程度上缓解模式崩塌问题并降低训练难度。后来 Arjovsky 等[16]提出的瓦瑟斯坦 GAN(Wasserstein GAN, WGAN)彻底解决了 GAN 训练不稳定的问题,并同时使模式崩塌问题得到基本解决。而与 WGAN 在同一年提出的最小二乘 GAN(least square GAN, LSGAN)仅通过将 GAN 的目标函数由交叉熵损失换成最小二乘损失,便同时解决了上述 2 个问题[17]。

3.深层特征提取器。U-Net 被认为是能够解决 CNN 效率低和对齐问题的方案,该方案会使得图像语义不会被忽略。为了更好地提取和保留图像细节,同时减少训练难度,Fan 等[18]提出了一种基于 U-Net 的医学图像融合方法,可以利用数据增强来训练少量样本,解决了语义丢失问题。除此之外,U-Net 还被应用于改善短帧 PET 图像质量[19]。除上述方法之外,VGG-19 在特征提取上具有较强的优势,Zhou 等[20]利用 VGG-19 作为特征提取器,提取源图像的低级和高级特征。另有研究微调了模型用来提取图像的深层特征,保留了更多源图像的细节信息[21]。而王钰帏等[22]则去掉输入层,将剩下的整个网络作为一个固定的特征提取器。同样,Li 等[23]使用 ResNet 从源图像中提取深度特征,然后利用零相分量分析准则对深度特征进行归一化处理,并获得初始权重图。与 VGG-19 相比,ResNet 可以实现更好的融合性能。

4.自注意力机制。Transformer 是一种基于完全自注意力机制的深度学习模型,其取代了传统的循环神经网络,并凭借其平行化处理能力显著提升了训练效率。在融合 PET 和 CT 等多模态信息时,Transformer 架构的自注意力机制能够在编码器部分为不同模态建立关联,统一特征的权重,确保多模态特征能够协同发挥作用。Song 等[24]利用 Transformer 的全局特征提取能力和稠密连接网络的局部特征强化机制提出了一种新型深度学习网络模型,该模型在减少特征损失和降低边缘模糊风险方面表现出色。Liu 等[25]创新地引入了一种混合网络,将 CNN 和 Transformer 模块结合在一起,能够同时建模长距离和短距离的依赖关系。周涛等[26]则将此种方法应用于肺部肿瘤 PET/CT 跨模态医学图像融合,实现了在挖掘源图像局部信息的同时,也能学习特征之间的全局交互信息。

表 1 是多种深度学习方法的详细对比,包括优缺点以及实际的应用效果。

5.挑战与展望。面对医学影像特有的成像物理特性与临床需求,深度融合方法仍面临多重挑战,主要体现在以下几个方面:(1)数据稀缺与模态异质性问题严重制约了模型的泛化能力。核医学图像如 PET、SPECT 通常伴随高噪声、低分辨率和不均匀伪影等特征,而与 CT 等解剖图像相比,核医学图像在对比度、细节表达及数据分布上存在显著差异。上述差异使得深度模型在跨模态信息对齐过程中面临困难,特别是在图像特征的提取与配准时,模型的泛化能力受到限制。当前缺乏大规模、标准化的公开 PET/CT 或 SPECT/CT 数据集,多中心采集协议和注射剂量差异进一步加剧了数据异质性。对此,可以通过风格迁移[如循环GAN(CycleGAN)]

表 1　深度学习方法的对比

方法	优点	缺点	实际效果
CNN	对局部结构特征(边缘、纹理)提取能力强;模型结构成熟,训练与推理速度快,适合部署;易于与传统图像处理方法结合	缺乏全局依赖建模,难以捕捉远距离区域间的语义关联;对模态间强差异性数据(如 PET/CT)融合效果不佳	适用于组织边界清晰、结构清楚的医学图像融合任务,缺乏全局信息时可能存在质量问题
GAN	能生成高质量融合图像,保留细节能力强;适用于提升图像对比度、增强模糊图像信息;具备强大的非线性建模能力,适合复杂分布	模型训练不稳定,容易模式崩溃;对网络设计和参数调节较为敏感;训练时间长	在PET/CT、PET/MR 等图像质量差异较大的模态融合中表现优异;可提升肿瘤区域、病灶边界等细节的可视化质量
深度特征提取器			
U-Net	编码器提取深层语义信息,解码器恢复细节,适合图像密集预测任务;可同时保留全局结构和局部细节;强大的端到端融合能力	网络较深,计算复杂;可能导致过拟合,尤其是在数据集较小的情况下	适合 PET/MR 等图像的空间配准与结构融合;在保留解剖结构的同时融合功能信息,常用于肿瘤定位或术前规划
VGG-19	网络深度大,特征提取能力强;能更好区分模态间的图像细节与语义层次差异;在图像分类任务中表现优秀,能够捕捉高质量的特征,适用于大规模医学图像任务	模型庞大,计算成本高,融合图像可能出现过平滑;对数据的预处理和调优要求高	常作为预训练模型用于提取医学图像的深层表征;可用于引导结构或纹理融合,但需结合其他结构增强模块
ResNet	深层网络解决了梯度消失问题,便于训练深层模型;提高特征学习能力	网络加深带来计算资源需求增加;可能不适合对精细对齐要求极高的融合任务	在复杂模态融合(如 PET/MR)中表现稳定,能兼顾结构、功能、代谢等多源信息,适用于高精度医学图像分析
自注意力机制(Transformer)	高效捕捉全局依赖关系,适合多模态融合任务;注意力机制有助于关键区域聚焦(如病灶)	计算开销大,训练时间长;对高质量标注数据依赖大,训练资源要求高	在多模态医学图像融合中展现出色的全局和局部特征提取能力,适用于复杂任务,如肿瘤识别与精准治疗规划

注:CNN 为卷积神经网络,GAN 为生成对抗网络,U-Net 为用于生物医学图像分割的深度学习网络架构,VGG-19 为视觉几何组提出的 19 层 CNN,ResNet 为残差网络

或特征对齐网络进行模态统一处理,同时采用模拟系统生成合成数据或利用联合学习机制提升模型的跨中心泛化能力。(2)深度融合模型训练成本高、部署效率低也是限制其临床推广的重要因素。处理三维或全身图像时,基于 U-Net、GAN 或 Transformer 架构的模型通常需要大量的显存和计算资源,这使得其难以满足临床实时性和稳定性的要求。例如,训练过程中需要几百 GB 的显存,并且推理时也需要高性能硬件支持。解决方案包括网络结构的轻量化,如用轻量卷积网络或通道重排网络替代主干网络,或引入知识蒸馏让小模型学到大模型的表达能力。在部署时,也可采用模型剪枝、量化等方式,压缩模型大小并提升推理速度。为了满足实时性,部分研究提出了 ROI 优先处理机制,即仅对关键区域进行高精度融合,节省计算资源。(3)模式崩塌与训练不稳定在基于 GAN 的融合模型中较为常见。由于 PET、SPECT 图像对比度低、结构信息不明确,传统 GAN 中的判别器难以准确判断图像质量,因此生成器可能会产生模糊或伪结构图像,导致模式崩塌。为了解决这一问题,可以采用更稳定的生成模型,如 WGAN[16]、LSGAN[17],甚至扩散模型[27]。此外,还可以通过在训练机制中引入结构一致性或物理先验,例如结合 CT 与 PET 的物理关系进行联合损失计算,或增加解剖一致性约束,从而减少模式崩塌和伪影问题。

医学图像融合能结合多模态信息,提高诊断精度,尤其在肿瘤检测和疾病诊断中有重要应用。尽管深度学习在提升图像融合效果方面取得了显著进展,但仍面临数据稀缺、计算资源需求大、模式崩塌等挑战。本文提出了风格迁移、特征对齐、模型轻量化等方法来应对上述问题。未来研究将聚焦提升跨模态融合的鲁棒性和模型稳定性,并探索基于物理先验的信息融合策略,推动医学图像融合技术更好地应用于临床诊断。

利益冲突　所有作者声明无利益冲突

作者贡献声明　张毅:文献调研、论文撰写;刘柳:研究指导、论文修改;王梦:论文修改、数据收集;谢文晖:研究指导

参　考　文　献

[1] 黄渝萍,李伟生.医学图像融合方法综述[J].中国图象图形学报, 2023, 28(1): 118-143. DOI:10.11834/jig.220603.
Huang YP, Li WS. A review of medical image fusion methods[J]. J Image Graphics, 2023, 28(1): 118-143. DOI:10.11834/jig.220603.

[2] Zhang Y, Liu Y, Sun P, et al. IFCNN: a general image fusion framework based on convolutional neural network[J]. Inf Fusion, 2020, 54: 99-118. DOI:10.1016/j.inffus.2019.07.011.

[3] He D, Li W, Wang G, et al. MMIF-INet: multimodal medical image fusion by invertible network[J]. Inf Fusion, 2025, 114: 102666. DOI:10.1016/ j.inffus.2024.102666.

[4] Xia KJ, Yin HS, Wang JQ. A novel improved deep convolutional neural network model for medical image fusion[J]. Cluster Comput, 2019, 22(S1): 1515-1527. DOI:10.1007/s10586-018-2026-1.

[5] Li J, Guo X, Lu G, et al. DRPL: deep regression pair learning for multi-focus image fusion[J]. IEEE Trans Image Process, 2020, 29: 4816-4831. DOI: 10.1109/TIP.2020.2976190.

[6] Wang Z, Li X, Duan H, et al. Medical image fusion based on convolutional neural networks and non-subsampled contourlet transform [J]. Expert Syst Appl, 2021, 171: 114574. DOI:10.1016/j.eswa.2021.114574.

[7] Liu Y, Chen X, Cheng J, et al. A medical image fusion method based on convolutional neural networks[C]. 2017 20th Int Conf Inf Fusion, 2017: 1-7. DOI:10.23919/ICIF.2017.8009769.

[8] Wang K, Zheng M, Wei H, et al. Multi-modality medical image fusion using convolutional neural network and contrast pyramid[J]. Sensors (Basel), 2020, 20(8): 2169. DOI:10.3390/s20082169.

[9] Xu H, Ma J. EMFusion: an unsupervised enhanced medical image fusion network[J]. Inf Fusion, 2021, 76: 177-186. DOI:10.1016/j.inffus.2021.06.001.

[10] Ma J, Yu W, Liang P, et al. FusionGAN: a generative adversarial network for infrared and visible image fusion[J]. Inf Fusion, 2019, 48: 11-26. DOI:10.1016/j.inffus.2018.09.004.

[11] Wang C, Yang G, Papanastasiou G, et al. DiCyc: GAN-based deformation invariant cross-domain information fusion for medical image synthesis[J]. Inf Fusion, 2021, 67: 147-160. DOI:10.1016/j.inffus.2020.10.015.

[12] 冯莉娟,马欢,鲁霞,等.基于生成对抗网络改善儿童低剂量 PET 图像质量的研究[J].中华核医学与分子影像杂志, 2022, 42(12): 708-712. DOI:10.3760/cma.j.cn321828-20220705-00212.
Feng LJ, Ma H, Lu X, et al. Study on improving the quality of low-dose PET images of children based on generative adversarial networks[J]. Chin J Nucl Med Mol Imaging, 2022, 42(12): 708-712. DOI:10.3760/cma.j.cn321828-20220705-00212.

[13] Guo X, Nie R, Cao J, et al. FuseGAN: learning to fuse multi-focus image via conditional generative adversarial network [J]. IEEE Trans Multimedia, 2019, 21(8): 1982-1996. DOI: 10.1109/TMM.2019.2895292.

[14] Wang Y, Xu S, Liu J, et al. MFIF-GAN: a new generative adversarial network for multi-focus image fusion[J]. Signal Process Image Commun, 2021, 96: 116295. DOI: 10.1016/j.image.2021.116295.

[15] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. IEICE Trans Fundam Electron Commun Comput Sci, 2015. DOI: 10.48550/arXiv.1511.06434.

[16] Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks[C]. Proc 34th Int Conf Mach Learn, 2017: 214-223. DOI:10.48550/arXiv.1701.07875.

[17] Mao X, Li Q, Xie H, et al. Least squares generative adversarial networks[C]. Proc IEEE Int Conf Comput Vis (ICCV), 2017: 2813-2821. DOI:10.1109/ICCV.2017.304.

[18] Fan F, Huang Y, Wang L, et al. A semantic-based medical image fusion approach[J]. Signal Image Video Process, 2019. DOI:10.48550/arXiv.1906.00225.

[19] 胡琳君,胡奕奕,郭彬威,等.深度学习重建方法改善快速采集 PET 图像质量的临床研究[J].中华核医学与分子影像杂志, 2021, 41(10): 602-606. DOI:10.3760/cma.j.cn321828-20210514-00164.
Hu LJ, Hu YY, Guo BW, et al. Clinical study of deep learning reconstruction to improve the quality of rapidly acquired PET images [J]. Chin J Nucl Med Mol Imaging, 2021, 41(10): 602-606. DOI:10.3760/cma.j.cn321828-20210514-00164.

[20] Zhou J, Ren K, Wan M, et al. An infrared and visible image fusion method based on VGG-19 network[J]. Optik, 2021, 248: 168084. DOI:10.1016/j.ijleo.2021.168084.

[21] Lu Y, Qiu Y, Gao Q, et al. Infrared and visible image fusion based on tight frame learning via VGG19 network [J]. Digit Signal Process, 2022, 131: 103745. DOI:10.1016/j.dsp.2022.103745.

[22] 王钰帏,王雷,郭新萍,等.基于复剪切波变换与 VGG19 模型的医学图像融合方法[J].山东理工大学学报(自然科学版), 2024, 38(4): 53-60. DOI: 10.3969/j.issn.1672-6197.2024.04.009.
Wang YW, Wang L, Guo XP, et al. The medical image fusion method based on the complex shearlet transform and the VGG19 model[J]. J Shandong Univ Technol(Natural Sci Edition), 2024, 38(4): 53-60. DOI:10.3969/j.issn.1672-6197.2024.04.009.

[23] Li H, Wu XJ, Durrani TS. Infrared and visible image fusion with ResNet and zero-phase component analysis[J]. Infrared Phys Technol, 2019, 102: 103039. DOI:10.1016/j.infrared.2019.103039.

[24] Song Y, Dai Y, Liu W, et al. DesTrans: a medical image fusion method based on Transformer and improved DenseNet[J]. Comput Biol Med, 2024, 174: 108463. DOI:10.1016/j.compbiomed.2024.108463.

[25] Liu Y, Zang Y, Zhou D, et al. An improved hybrid network with a transformer module for medical image fusion[J]. IEEE J Biomed Health Inform, 2023, 27(7): 3489-3500. DOI: 10.1109/JBHI.2023.3264819.

[26] 周涛,程倩茹,张祥祥,等.基于 DCIF-GAN 的肺部肿瘤 PET/CT 跨模态医学图像融合[J].光学精密工程, 2024, 32(2): 221-236. DOI:10.37188/OPE.20243202.0221.
Zhou T, Cheng QR, Zhang XX, et al. PET/CT cross-modal medical image fusion of lung tumors based on DCIF-GAN[J]. Opt Precision Eng, 2024, 32(2): 221-236. DOI: 10.37188/OPE.20243202.0221.

[27] Sun J. MedFusion-TransNet: multi-modal fusion via transformer for enhanced medical image segmentation[J]. Front Med (Lausanne), 2025, 12: 1557449. DOI:10.3389/fmed.2025.1557449.

(收稿日期:2024-12-01)