中图法分类号:TP391.4　　**文献标识码**:A　　**文章编号**:1006-8961(2025)06-1985-16

论文引用格式:Pan Y S, Ma H J, Xia Y and Zhang Y N. 2025. Application of content generation in medical images. Journal of Image and Graphics, 30(6):1985-2000(潘永生, 马豪杰, 夏勇, 张艳宁. 2025. 医学影像中的生成技术. 中国图象图形学报, 30(6):1985-2000)[DOI:10.11834/jig.240459]

医学影像中的生成技术

潘永生[1],马豪杰[2],夏勇[1,2,3*],张艳宁[1]

1. 西北工业大学计算机学院空天地海一体化大数据应用技术国家工程实验室,西安　710129;
2. 西北工业大学宁波研究院,宁波　315048;3. 西北工业大学深圳研究院,深圳　518057

摘　要:医学影像是一种利用各种成像技术捕捉人体内部结构和功能的医学诊断方法。这些技术可以提供关于人体解剖、生理和病理状态的视觉信息,在疾病诊断、治疗和预后预测中发挥着重要作用。由于不同类型或子类型的医学影像反映患者身体的不同信息,在医疗诊断时往往需要多种不同类型或子类型的医学影像来获取更加全面的信息,从而提高诊断准确率。然而在现实生活中,多模态影像数据获取面临采集时间长、费用高以及可能增加辐射剂量等困难。因此,学者期待能够使用图像处理技术进行跨模态医学影像合成,即使用某一种或一些模态的医学影像生成另一种或一些模态的医学影像。本文主要介绍医学图像领域中跨模态图像合成技术和跨模态医学影像合成的应用。跨模态医学影像合成虽然能为多模态影像诊断带来便利,但也存在一些技术挑战。例如合成影像和真实影像在诊断性能上具有明显差异,从而导致合成影像的临床失效问题,隐私和伦理问题会导致高质量多模态医学影像数据获取成本高的问题。同时,由于不同模态的影像数据在分辨率、对比度和图像质量上存在一定差异,这种差异会影响生成模型在生成过程中的一致性。如何解决不同模态之间的数据不一致性也是跨模态医学影像合成面临的挑战,研究者大多从模型本身入手,通过提高模型的表示能力或设计针对具体任务的约束条件来提高合成影像的质量,所开发的跨模态医学影像合成技术已应用于影像采集、重建、配准、分割、检测和诊断等环节,给许多问题带来新的解决思路和方法。

关键词:人工智能;医学影像;跨模态图像生成;深度学习;人工智能内容生成(AIGC)

Application of content generation in medical images

Pan Yongsheng[1], Ma Haojie[2], Xia Yong[1,2,3*], Zhang Yanning[1]

1. National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, School of Computer Science, Northwestern Polytechnical University, Xi'an　710129, China; 2. Ningbo Institute of Northwestern Polytechnical University, Ningbo　315048, China; 3. Research and Development Institute of Northwestern Polytechnical University in Shenzhen, Shenzhen　518057, China

Abstract: Medical imaging, a crucial tool for medical actions, utilizes various imaging techniques to capture the internal

收稿日期:2024-08-05;**修回日期**:2024-09-06;**预印本日期**:2024-09-13

*通信作者:夏勇　yxia@nwpu.edu.cn

基金项目:国家重点研发计划资助(2022YFC2009903,2022YFC2009900);国家自然科学基金项目(6240012686,62171377,62401475);中央高校基本科研业务费专项资金项目(D5000230376);宁波市医学影像临床医学研究中心项目(2021L003, 2022LYKFZD06);深圳市科技计划项目(JCYJ20220530161616036)

Supported by: National Key R&D Program of China (2022YFC2009903, 2022YFC2009900); National Natural Science Foundation of China (6240012686, 62171377, 62401475); Fundamental Research Funds for the Central Universities (D5000230376); Ningbo Clinical Research Center for Medical Imaging (2021L003, 2022LYKFZD06); Shenzhen Science and Technology Program (JCYJ20220530161616036)

structure and function of the human body. Common types of medical images include magnetic resonance imaging (MRI), computed tomography (CT), positron emission tomography (PET), plain X-rays, ultrasound, and optical imaging. The information obtained from these images varies due to differences in imaging principles. For example, MRI uses a strong magnetic field and radio waves to obtain images of the inside of the body and provides good information about soft tissues. CT uses X-rays and computerized processing to create images of cross-sections of internal body structures and is primarily used to image high-electron-density tissues (e. g. , bone); however, its capability to contrast soft tissues is somewhat limited. PET uses tracers labeled by radioisotopes to observe biological processes and functional activities within the body to image specific biological functions. At the same time, depending on the differences in imaging parameters and tracers, medical images of the same imaging type may also differ from different subtypes, such as T1- and T2-weighted MRI sequences, FDG-PET, and Aβ-PET. These medical imaging techniques provide visual information about the anatomical, physiological, and pathological states of the human body and play an important role in disease diagnosis, treatment, and prognosis prediction. Medical images of the same type or subtype are referred to as single modality, and medical images that contain different modalities are referred to as multiple modalities. Given that various types or subtypes of medical images respond to different information about the patients' body, multiple types/subtypes of medical images are often acquired to obtain more comprehensive information to improve diagnostic accuracy. However, multimodal image data acquisition faces difficulties such as long acquisition time, high cost, and possible increase in radiation dose. Therefore, generative techniques should be used for cross-modal medical image synthesis, that is, using medical images of one or some modalities to generate medical images of another or some other modalities. Although cross-modal medical image synthesis can facilitate multimodal image diagnosis, some technical challenges exist. For example, some information that can be captured in the target modality does not exist in the source modality due to different imaging principles of various imaging modalities. In this case, synthesized images of the target modality lack certain information, creating significant disparities in diagnostic performance between synthesized and real images, which leads to the problem of clinical failure. At the same time, privacy and ethical issues also contribute to the high cost of acquiring high-quality multimodal medical image data and the problem of missing data in cross-modal medical image synthesis. In addition, differences in resolution, contrast, and image quality between different modalities affect the consistency of image generation models during the generation process. Addressing these inconsistencies in data across different modalities poses a significant challenge for cross-modal medical image synthesis. The computational complexity and generalization ability of the model also need to be considered, as cross-modal medical image synthesis often requires complex models and many computational resources, which may limit the usefulness and scalability of cross-modal medical image synthesis methods. In addition to the training data that the model has already seen, whether the model can perform well on new or other different datasets should be considered. Most researchers start from the model itself and improve the quality of the synthesized images by improving the representation ability of the model or designing task-specific constraints. These developed cross-modal medical image synthesis techniques have been applied to image acquisition, reconstruction, alignment, segmentation, detection, and diagnosis, bringing new ideas and methods to solve many problems. This paper focuses on cross-modal image synthesis techniques and applications in the field of medical imaging. We will introduce existing cross-modal medical image synthesis techniques from three aspects: traditional synthesis methods, deep learning-based synthesis methods, and task-driven synthesis methods. Traditional synthesis methods usually divide an image into multiple small blocks and encode each block into a representation vector. This is done by establishing a mapping between the paired block representation vectors of different modalities and then generating the corresponding target modality block based on the encoding of the source modality block. The random forest-based approach treats image synthesis as a regression problem, assuming that the value of the target modal block or its central region is the dependent variable of the source modal block. This relationship can be obtained through a regression model. Dictionary learning-based methods assume that there exists a dictionary for each modality, that each image block can be obtained from a sparse representation of the elements in the dictionary, and that the image blocks corresponding to different modalities have the same dictionary encoding. Compared with traditional methods, deep learning-based cross-modal image synthesis methods can directly use large-scale parametric models to build mappings from source modal images to target modal images in an end-to-end manner. These methods automatically extract the representa-

tion features of an image or an image block in a data-driven manner without manually design the representation features. Given their ease of implementation and superior performance, deep learning-based techniques have become the leading method in cross-modal image synthesis. In this paper, we introduce them from simple CNN-based approach, encoder-decoder network-based approach, generative adversarial network approach, and diffusion model-based approach. Task-oriented cross-modal image synthesis methods consider that the synthesis task has a specific task bias and form a task-specific bias by adding a task-related design on the basis of a generalized technique. In this manner, the synthesized image preserves more information that contributes to the task and achieves a performance enhancement on the specific task. Such synthesis methods are presented in three categories: task-oriented biases, biases formed through network models, and image synthesis embedded in task models. Finally, we present the application scenarios of cross-modal medical image synthesis techniques and their application under their typical advantageous tasks.

Key words: artificial intelligence; medical imaging; cross-modal image synthesis; deep learning; artificial intelligence generated content(AIGC)

0　引　言

医学影像可以提供关于人体解剖、生理和病理状态的视觉信息,在疾病诊断、治疗和预后预测中发挥着重要的作用(Elangovan 和 Jeyaseelan, 2016;施俊 等,2020)。常见的医学影像类型包括磁共振影像(magnetic resonance imaging, MRI)、计算机断层成像(computed tomography, CT)、正电子发射成像(positron emission tomography, PET)、X 光平片、光学成像等,如图 1 所示。

由于成像原理的差异,这些影像获取的信息有所不同,例如 MRI 是通过使用强磁场和无线电波来获取身体内部的图像,能很好地提供有关软组织的信息(Vijayalaxmi 等,2015),CT 使用 X 射线和计算机处理技术创建人体内部结构的横截面的图像,主要用于成像高电子密度组织(如骨骼),但也可以提供一定程度的软组织对比度(Villarraga-Gómez 等,2019),PET 利用被放射性同位素标记的示踪剂来观察人体内部的生物过程和功能活动,从而成像特定的生物学功能。同时,根据成像参数和所用试剂的差异,同一种成像类型的医学影像也会存在差异,甚至形成不同的子类型。例如 MRI 就包括 T1 加权序列、T2 加权序列等(Tanitame 等,2017),PET 包括 FDG-PET、Aβ-PET 等(Pérez-Grijalba 等,2019)。

医学影像中同一类型或子类型的影像称为同一模态,同时包含不同模态的医学影像则称为多模态影像(Fard 等,2021)。相较于单模态影像,多模态医学影像能够同时包含不同模态数据间的多元化信息(罗娜 等,2022)。因此,往往需要多模态影像来获取更加全面的信息,从而提高诊断准确率。然而在现实生活中,采集多模态影像数据获取面临着时间长、费用高、可能增加辐射剂量等困难。人们期待能够使用图像处理技术进行跨模态医学影像合成,即使用某一种(或一些)模态的医学影像去生成另一种(或一些)模态的医学影像(Pan 等,2020)。而人工智能内容生成(artificial intelligence generated content, AIGC)能够根据给定的提示进行高效的图像生

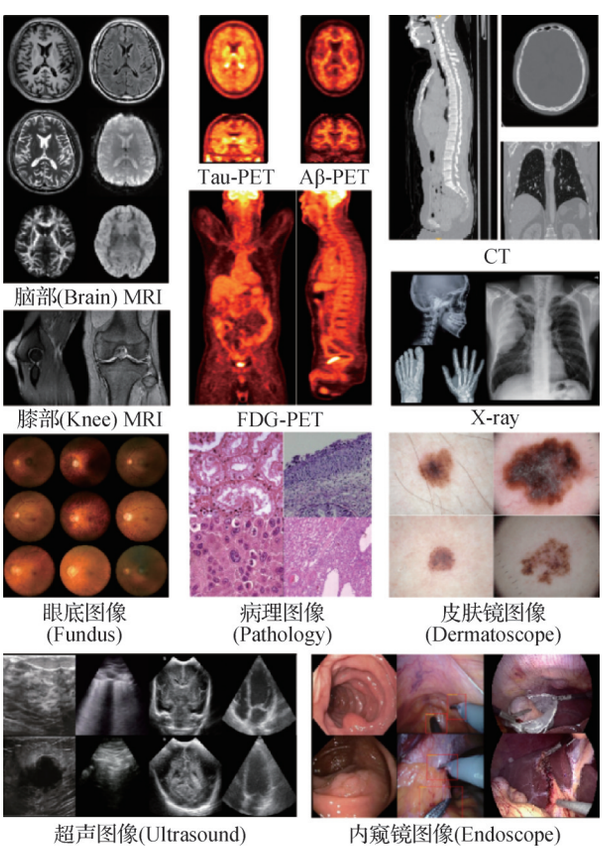

图 1　多种模态的医学影像

Fig. 1　Medical images of various modalities

成和编辑任务，因此许多科研工作者纷纷尝试使用人工智能生成技术进行跨模态图像生成来解决多模态影像缺失问题。

AIGC是指利用复杂的算法、模型和规则，从大规模数据集中学习映射规律，从而使模型能够创造出所需内容的人工智能技术（Cao等，2023；Foo等，2025；Di，2024）。AIGC技术涵盖多个领域，根据生成目标数据的形式可分为文本生成、音频生成、图像生成和视频生成。其中文本生成技术可用于新闻报道、媒体、营销文案创作、与用户进行文本交互聊天机器人等（Di，2024；Ren等，2023）。音频生成技术可用于将文本转换为语音的语音合成（语音播报）和将输入语音或文本转换成具有目标说话者特征语音的语音克隆（智能配音）等（Borsos等，2023；Božić和Horvat，2024）。图像生成技术可根据规则或描述进行艺术创作、将信息以更便于观察的影像形式呈现等任务（Singh和Raza，2021；Frolov等，2021）。视频生成技术可用于编辑视频内容、制作宣传片和广告片以及数字人控制等（Li等，2018；Bhagwatkar等，2020）。

AIGC技术中的生成模型主要分为单模态生成模型和多模态生成模型。单模态生成模型指输入和输出属于同一种模态，如聊天机器人的输入为文本输出也为文本，图像编辑输入为图像输出也为图像。单模态生成模型在设计和实现上更简单，且由于其专注于一种模态的数据可以更深入地学习该模态的特征，使得单模态生成模型在性能上往往高于多模态生成模型。但是单模态生成模型无法利用不同模态之间的互补信息，且有很强的数据依赖性，使得单模态生成模型的泛化能力很差。多模态模型生成的目标是通过从数据中学习不同模态间的关联性来建立统一的表示空间，能够融合来自不同模态的互补信息，从而提高模型的准确性和泛化能力。这种关联性有时是非常复杂的，特别是不同模态之间面临语义对齐挑战的时候，因此多模态表示空间与单模态相比更难学习，通常需要大量的、多样化的数据来训练模型。

本文主要围绕跨模态图像合成介绍AIGC技术在医学图像领域中的应用。

1 问题与挑战

跨模态生成是基于模态之间的语义一致性，实现不同模态数据形式上的相互转换，有助于提高不同模态间的迁移能力（刘华峰 等，2023）。其虽然能为多模态影像诊断带来便利，但也存在一些技术挑战，例如临床失效问题，即合成影像和真实影像在诊断性能上具有明显的差异（Pan等，2022）。这是因为，各类影像模态的成像原理不同，目标模态能采集的某些信息在源模态影像中并不存在，导致合成的目标模态影像依然不具有这些信息。同时，合成模型受到自身表示能力的限制，会产生一定信息损失，特别当其训练过程受到约束条件的偏置化引导时，将在合成影像中产生与偏置相关的信息损失。其次，受限于隐私和伦理问题，高质量的多模态医学影像数据获取成本高。这导致在跨模态医学影像合成中经常会面临数据缺失的问题。例如，在阿尔茨海默病神经影像（Alzheimer's disease neuroimaging initiative，ADNI-1）数据库（Jack等，2008）中，有800多名受试者有基线MRI扫描，但只有400受试者有基线PET数据。由于不同模态的影像数据在分辨率、对比度和图像质量上存在一定的差异，这种差异会影响生成模型在生成过程中的一致性，如何解决不同模态之间的数据不一致性也是跨模态医学影像合成所需要面临的挑战。除此之外，模型的计算复杂性和泛化能力也需要考虑到，由于跨模态医学影像合成往往需要复杂的模型和大量的计算资源，可能会限制跨模态医学影像合成方法的实用性和可扩展性。同时除了模型已经见过的训练数据之外，也应该考虑模型能否在新的或其他不同的数据集上拥有不错的性能。

当前，研究者大多从模型本身入手，通过提高模型的表示能力或设计针对具体任务的约束条件来提高合成影像的质量，所开发的跨模态医学影像合成技术已应用于影像采集、重建、配准、分割、检测和诊断等环节，给许多问题带来了新的解决思路和方法。

2 跨模态影像合成技术现状

2010年以来，跨模态影像合成受到研究者越来越多的关注，产生了许多合成方法。本文将这些方法大致分为3类，包括传统合成方法、基于深度学习的合成方法和任务驱动的合成方法。

2.1 问题陈述

为方便介绍，本文以两个模态为例介绍跨模态

图像合成方法。设$\boldsymbol{X}=\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_M\}$是源模态中的$M$幅图像，$\boldsymbol{Y}=\{\boldsymbol{y}_1,\cdots,\boldsymbol{y}_N\}$是目标模态中的$N$幅图像，当$M$和$N$足够大时，$\boldsymbol{X}$和$\boldsymbol{Y}$可近似源模态和目标模态图像服从的分布。跨模态图像合成假设存在一种映射$G:\boldsymbol{X}\rightarrow\boldsymbol{Y}$，具体为

$$\forall\,\boldsymbol{x}\sim\boldsymbol{X},\exists\,\boldsymbol{y}\sim\boldsymbol{Y}\ \text{s.t.}\ G(\boldsymbol{x})=\boldsymbol{y} \tag{1}$$

式中，$\boldsymbol{x}$表示源模态图像，$\boldsymbol{y}$表示目标模态图像，映射G可以通过某种优化方法由给定的数据$\boldsymbol{X}$和$\boldsymbol{Y}$近似地估计出来，此时其约束条件被松弛为某一种优化目标R，即

$$\hat{G}=\arg\min_{G}R(\boldsymbol{x},\boldsymbol{y};G)\quad\text{s.t.}\,\boldsymbol{x}\in\boldsymbol{X},\boldsymbol{y}\in\boldsymbol{Y} \tag{2}$$

目前所有的图像合成技术均围绕映射模型(G)和优化目标(R)的设计展开，两个方面相辅相成、齐头并进。其中典型的映射模型有字典学习、随机森林、卷积网络和Transformer等，优化目标有平均绝对误差、结构相似性和对抗损失，特征一致性等。

2.2　传统跨模态影像合成方法

这类方法通常将影像划分成多个小块，并将每个块编码成一个表示向量，通过建立不同模态的配对的块表示向量之间的映射，再根据源模态块的编码产生对应的目标模态块。主要关注表示向量的设计和映射模型的建立，模型的求解过程以类似“数据检索”的方式进行，优化目标通常使用平均均方误差等容易计算的指标。这类方法包括字典学习随机森林等。

基于字典学习的方法(Huang等，2018)假设每个模态存在一个字典，每个图像块均可由字典中元素的稀疏表示得到，不同模态对应的图像块具有相同的字典编码。进行跨模态影像合成时，为不同的模态设置统一的编码和不同的字典，并根据稀疏表示原理通过最小化联合重建误差来求解字典和编码。设对应于模态$\boldsymbol{X}$、$\boldsymbol{Y}$的字典分别为$\boldsymbol{\Phi}_X$、$\boldsymbol{\Phi}_Y$，$\boldsymbol{x}$、$\boldsymbol{y}$在$\boldsymbol{\Phi}_X$、$\boldsymbol{\Phi}_Y$上的编码系数(即表示向量)分别为$\boldsymbol{A}_x$、$\boldsymbol{A}_y$，它们之间的关系为$\boldsymbol{x}=\boldsymbol{\Phi}_X\boldsymbol{A}_x$和$\boldsymbol{y}=\boldsymbol{\Phi}_Y\boldsymbol{A}_y$。由于只有$\boldsymbol{x}$、$\boldsymbol{y}$是已知量，$\boldsymbol{\Phi}_X$、$\boldsymbol{\Phi}_Y$和$\boldsymbol{A}_x$、$\boldsymbol{A}_y$可由稀疏表示最小化重建误差求得，即

$$\min_{\boldsymbol{\Phi}_X,\boldsymbol{A}_x}\left\|\boldsymbol{x}-\boldsymbol{\Phi}_X\boldsymbol{A}_x\right\|_{\mathrm{F}}^2+\lambda\left\|\boldsymbol{A}_x\right\|_0,\ \forall\boldsymbol{x}\in\boldsymbol{X} \tag{3}$$

$$\min_{\boldsymbol{\Phi}_Y,\boldsymbol{A}_y}\left\|\boldsymbol{y}-\boldsymbol{\Phi}_Y\boldsymbol{A}_y\right\|_{\mathrm{F}}^2+\lambda\left\|\boldsymbol{A}_y\right\|_0,\ \forall\boldsymbol{y}\in\boldsymbol{Y} \tag{4}$$

式中，$\|\cdot\|_{\mathrm{F}}$是Frobenius范数，$\|\cdot\|_0$是L_0-范数并通常被松弛为L_1-范数，λ是平衡稀疏性和重建误差的权重参数(Wang等，2020)。如此，$\boldsymbol{x}$、$\boldsymbol{y}$便可由其编码$\boldsymbol{A}_x$、$\boldsymbol{A}_y$表示，$\boldsymbol{x}$和$\boldsymbol{y}$的映射简化为$\boldsymbol{A}_x$、$\boldsymbol{A}_y$的变换关系。在此基础上可直接令$\boldsymbol{A}_x=\boldsymbol{A}_y=\boldsymbol{A}$来联立求解一个相同的编码，即

$$\min_{\boldsymbol{\Phi}_X,\boldsymbol{\Phi}_Y,\boldsymbol{A}}\left\|\boldsymbol{x}-\boldsymbol{\Phi}_X\boldsymbol{A}\right\|_{\mathrm{F}}^2+\left\|\boldsymbol{y}-\boldsymbol{\Phi}_Y\boldsymbol{A}\right\|_{\mathrm{F}}^2+\lambda\left\|\boldsymbol{A}\right\|_1\quad\text{s.t.}\ \boldsymbol{x}\in\boldsymbol{X},\boldsymbol{y}\in\boldsymbol{Y} \tag{5}$$

也可通过求解额外的模态对应的投影变换P_X和P_Y，使得$\boldsymbol{A}_x$、$\boldsymbol{A}_y$在投影空间上具有相同的投影，即

$$\arg\min_{P_X,P_Y}\left\|P_X\boldsymbol{A}_x-P_Y\boldsymbol{A}_y\right\|\quad\text{s.t.}\ \boldsymbol{\Phi}_X\boldsymbol{A}_x\in\boldsymbol{X},\boldsymbol{\Phi}_Y\boldsymbol{A}_y\in\boldsymbol{Y} \tag{6}$$

在应用时，两者均需要先固定$\boldsymbol{\Phi}_X$、$\boldsymbol{\Phi}_Y$，并根据式(3)求出$\boldsymbol{A}_x$，然后令$\hat{\boldsymbol{A}}_y=\boldsymbol{A}_x$或$\hat{\boldsymbol{A}}_y=\boldsymbol{P}_Y^{-1}\boldsymbol{P}_X\boldsymbol{A}_x$，得出$\hat{\boldsymbol{y}}=\boldsymbol{\Phi}_Y\hat{\boldsymbol{A}}_y$。

基于随机森林的方法(Jog等，2017)将影像合成视为回归问题，假设目标模态块(或其中心点/中心区域)的值是源模态块的因变量，并且这种关系可以通过回归模型得到。这类方法需要首先使用其他方法编码每个图像块，因此非常受编码方式的影响。为了提高表示能力，随机森林使用的表示向量通常是由多个尺度的多种简单特征组合而成的，如空间位置、离散傅里叶系数、类haar特征和平均亮度等。随机森林回归器由一组回归树组成，每个回归树都是在训练集上随机选取的一个子集上学习得到的。其中树的每个内部节点选择一个特征来划分传入的训练样本来最大化信息增益，最终每棵树根据树中每个节点的划分将特征空间划分为多个区域。每个叶子节点的样本数小于限定的数量或限定的值时，其输出值为该节点所有样本目标值的平均或中值。在应用时，每一棵树均对输入的表示向量进行回归，最终的回归结果由所有树的结果通过某种加权平均方式产生。

2.3　基于深度学习的跨模态影像合成方法

随着深度学习的发展，跨模态影像合成研究已逐渐转移到深度学习框架中。从简单卷积神经网络(convolutional neural network，CNN)到变分自编码网络(variational auto-encoder，VAE)、U-Net、生成对抗网络(generative adversarial network，GAN)和扩散模型等，深度学习的各种技术都在跨模态医学影像合成中有所应用。与传统方法相比，此类方法可以直接使用大规模的参数化模型以端到端的方式建立从源模态影像到目标模态影像的映射，并以数据驱

动的方式自动提取图像(块)的表示特征,而不需要手工设计表示特征。由于其便于实现和性能优越,基于深度学习的跨模态影像合成技术目前已经占据了主导地位。

2.3.1 基于简单CNN的方法

基于简单CNN的方法早期常采用与传统方法类似的策略,首先将图像划分成一系列的小块,并使用每个源模态图像块预测目标模态图像块的中心值或中心区域的值。不同的是,这些方法通过堆叠多个卷积层来构建映射模型,并将特征提取和回归预测集成在一起,使用误差反向传播算法来优化模型参数。由于其数据驱动的学习方式,这些方法在训练样本充足的情况下能够达到比传统方法更好的性能。但是,这些方法的表示能力受到参数量和结构复杂性的限制(如使用的卷积层数量和卷积核大小)。因此,早期的CNN方法通常沿着增加模型复杂度的方向发展。Li等人(2014)只使用两层卷积建立MRI和PET之间的映射,Nie等人(2016)使用4层卷积建立MRI和CT之间的映射,而Xiang等人(2017)则进一步串联3个4层的卷积网络(共12层)建立T1-MRI和低剂量PET到高剂量PET的映射。然而,增加卷积层的个数在增强表示能力的同时也会增加计算复杂度;同时,由于依然需要逐块甚至逐像素计算目标值,这些方法的计算效率并不比传统方法更具优势。

2.3.2 基于编解码网络的方法

这类方法假设源模态和目标模态的影像在某一隐空间中存在共享的中间编码,因此其通常包含一个编码器和一个解码器,编码器将源模态图像(块)转换成中间编码,解码器将中间编码解码成目标模态图像(块)(Liu等,2017)。编码器通常采用多个带下采样的卷积网络来压缩源模态图像中的信息,解码器则采用多个带上采样的卷积网络来将压缩的信息恢复为目标模态图像。这类方法的优点是可以同时输出一个较大区域甚至整幅图像,不但避免了简单CNN方法中的逐像素计算,减少了计算代价,也有助于保留结构性信息,使合成图像具有更好的视觉效果。然而,这是一个“有损压缩”过程,存在信息丢失问题。例如,VAE模型常采用均值和方差的组合作为中间编码,虽然保留了源模态图像中的统计信息,但同时抑制了个体化信息。有两种策略可以补偿损失的信息。一种是简化编码器和解码器的结构来减少信息损失,并在编解码之间增加额外的特征转换单元(如加入多个残差模块)来增强信息转换能力(Pan等,2020;Pan和Xia,2021)。另一种是使用跳跃连接将编码器的中间特征同时作为解码器的输入来补偿缺失的信息(Pan和Xia,2021;Isensee等,2021)。需要注意的是,一般的编解码结构可以适用于配对或非配对的图像,而带跳跃连接的结构则只有在配对的图像上才有良好的性能(Yang等,2020a)。图2给出了用于影像生成的编解码网络及其变形的结构示意图。

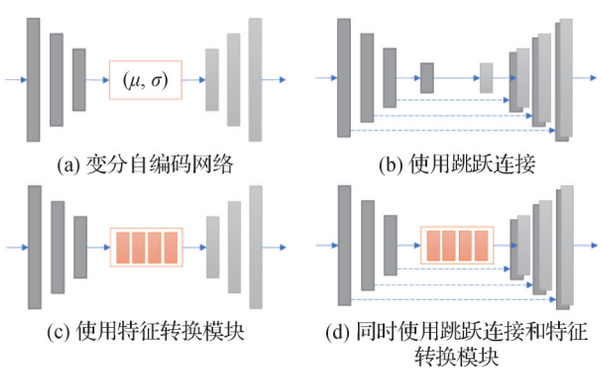

图2 编码网络及其变形

Fig. 2 Coding network and its variants ((a) variational auto-encoder; (b) with skip-connections; (c) with translation blocks; (d) with both skip-connections and translation blocks)

2.3.3 生成对抗网络的方法

简单CNN和编解码网络一般使用确定性的简单优化目标,如平均绝对误差(mean absolute error, MAE)、均方误差(mean squared error, MSE)等,从而会引入确定性偏置,即合成网络的优化方向始终朝向优化目标,导致与优化目标相背的信息传输被抑制,使得合成的图像在被抑制的信息方面呈现出平均的效果,虽然在MSE、峰值信噪比(peak signal-to-noise ratio, PSNR)和结构相似性(structural similarity, SSIM)等指标上表现优异,但看起来却非常模糊。为了克服这种确定性偏置的不利影响,出现了基于GAN的影像合成技术(Fard等,2021)。与前述的确定性目标不同,GAN使用一个判别网络——判别器(discriminator, D)来指导生成网络——生成器(generator, G)的学习过程:通过交替训练生成器和判别器,使两个网络以相互竞争的方式同步提高各自的能力,最终(在理想情况下)达到纳什均衡。这个过程中,生成器逐渐生成具有真实图像特征的合成图像,判别器不断提高对合成图像的鉴别能力。

假设将真实样本和合成样本输入判别器D时的判别标签分别为1和0，那么GAN的中的D和G通过不断迭代如下两个优化过程求解，具体为

$$\max_{D} \mathrm{E}_{x\sim X, y\sim Y}\left[\log\left(D(\boldsymbol{y})\right)+\log\left(1-D\left(G(\boldsymbol{x})\right)\right)\right] \tag{7}$$

$$\min_{G} \mathrm{E}_{x\sim Y}\left[\log\left(1-D\left(G(\boldsymbol{x})\right)\right)\right] \tag{8}$$

式中，$\boldsymbol{x}$表示源模态图像，$\boldsymbol{y}$表示目标模态图像，$\boldsymbol{X}$为源模态图像集合，$\boldsymbol{Y}$为目标模态图像集合。

相比于确定性的优化目标，判别器D的存在使得GAN可以直接优化似然度本身，而不是MAE等对数似然的下界。D的不断更新相当于不断变换生成网络G的优化目标，因此避免了引入确定性约束。由于需要训练两个网络，GAN在训练时相比于其他方法需要更多的计算时间和空间，但在应用时只需要运行生成网络，其时间复杂度和其他深度学习方法是接近的。

作为一种学习策略，任何可微的模型都可以用于构建判别器和生成器，但常用于影像合成的GAN通常使用图2中的一种结构作为生成器，而判别器则主要使用图3(a)所示的多层卷积结构。

训练GAN的理想状态是达到纳什均衡，这有时可以用梯度下降法或其衍生算法做到，但大部分时候是做不到的。目前尚无方法确保能达到纳什均衡，所以相比于使用确定性优化目标的方法，GAN的训练是不稳定的。为此，确定性优化目标不得不被重新考虑进来，如加入MAE、SSIM等。也可以使用感知损失(perceptual loss)提高GAN的稳定性，但感知损失依赖额外的网络来提取中间特征，并且只能在不同模态的配对图像上计算。如果在固定参数的预训练网络上计算，感知损失实际上相当于非常多但有限的确定性优化目标的组合，仍然会引入一定程度的确定性偏置。如果直接在判别网络上计算感知损失，也可以增强GAN的稳定性，此时多层感知损失也称为特征匹配损失(Wang等，2018)。图3给出了不同约束目标的对比示意图，通过组合不同的约束目标可以得到GAN的不同变形。

2.3.4　基于扩散模型(diffusion model)的方法

扩散模型是一种基于概率的生成模型，与GAN不稳定的训练过程相比，扩散模型具有更加简单的网络结构和更加稳定的训练过程。同时，扩散模型能够生成高质量、逼真的图像，有效避免了GAN中的模式崩溃问题。扩散模型可以看做是一个参数化

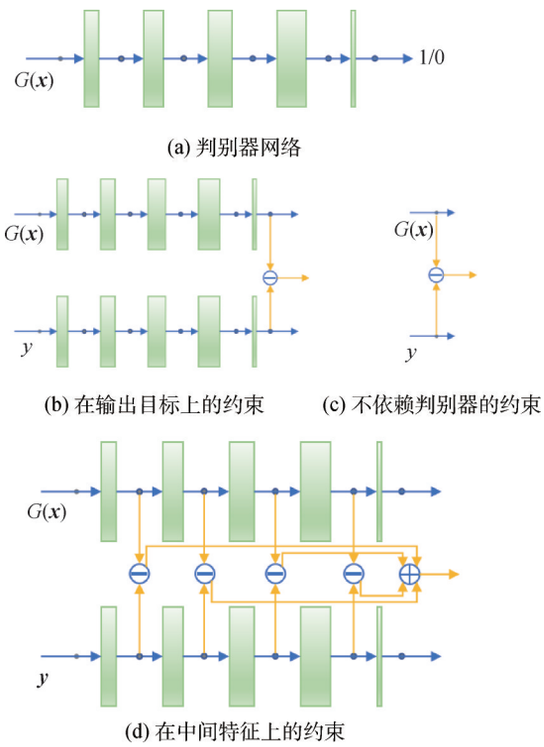

图3　不同约束目标的对比图

Fig. 3　Comparison for different constraint objectives ((a) discriminator network; (b) constraint on network output; (c) discriminator-independent constraint; (d) constraint on intermediate features)

的马尔可夫链，使用变分推理进行训练，在有限时间后产生与数据相匹配的样本(Ho等，2020)。其主要包括两个阶段：正向扩散阶段和逆向扩散阶段。正向扩散阶段是一个给原图逐步增加噪声的阶段。给定一个分布$q(\boldsymbol{x})$和采样次数$t=1,\cdots,T$，正向扩散阶段可以通过高斯转换$q(\boldsymbol{x}_t)=q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$添加高斯噪声生成一系列的中间变量$\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_T$，其中的高斯转换定义为

$$q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})=N\left(\boldsymbol{x}_t;\sqrt{1-\beta_t}\,\boldsymbol{x}_{t-1},\beta_t\boldsymbol{I}\right) \tag{9}$$

式中，$\beta_t\in(0,1)$是超参数，N表示高斯分布，$\boldsymbol{I}$为单位矩阵。利用概率链规则和马尔可夫性质，可以得到中间变量$\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_T$基于$\boldsymbol{x}_0$的联合分布，具体为

$$q(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_T|\boldsymbol{x}_0)=\prod_{t=1}^{T}q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) \tag{10}$$

进一步，可以直接根据初始状态得到t时的状态，具体为

$$q(\boldsymbol{x}_t|\boldsymbol{x}_0)=N\left(\boldsymbol{x}_t;\sqrt{\bar{\alpha}_t}\,\boldsymbol{x}_0,(1-\bar{\alpha}_t)\boldsymbol{I}\right) \tag{11}$$

式中，$\alpha_t=1-\beta_t$，$\bar{\alpha}_t=\prod_{s=0}^{t}\alpha_s$。由此对于给定的$\boldsymbol{x}_0$，其

t时的一个样本$\boldsymbol{x}_t$可以通过采样一个高斯噪声$\boldsymbol{\varepsilon} \in (0, 1)$并添加到$\boldsymbol{x}_0$上得到，具体为

$$\boldsymbol{x}_t = \sqrt{\bar{\alpha}_t}\,\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\,\boldsymbol{\varepsilon} \tag{12}$$

这个加噪的过程可以使用网络模型来预测添加的噪声，其优化目标为

$$\nabla_\theta \left\| \boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}_\theta\left(\sqrt{\bar{\alpha}_t}\,\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\,\boldsymbol{\varepsilon}, t\right)\right\|^2 \tag{13}$$

式中，$\boldsymbol{\varepsilon} \sim N(0, 1)$为高斯噪声，$\boldsymbol{\varepsilon}_\theta$为噪声预测模型，$\theta$为模型参数。这使得马尔可夫链可以通过去除噪声来从非结构化噪声向量生成新的数据样本。

逆扩散过程的高斯转换为

$$p\left(\boldsymbol{x}_{t-1}\middle|\boldsymbol{x}_t\right) = N\left(\boldsymbol{x}_{t-1}; \boldsymbol{\mu}_\theta\left(\boldsymbol{x}_t, t\right), \sum_\theta\left(\boldsymbol{x}_t, t\right)\right) \tag{14}$$

由此过程，可得

$$\boldsymbol{x}_{t-1} = \boldsymbol{\mu}_\theta\left(\boldsymbol{x}_t, t\right) + \sqrt{\sum_\theta\left(\boldsymbol{x}_t, t\right)}\,z = \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\varepsilon}_\theta\left(\boldsymbol{x}_t, t\right)\right) + \sigma_t z \tag{15}$$

式中，z为服从$N(0, 1)$的高斯噪声，通过T次迭代去噪，逆向扩散过程便可从$\boldsymbol{x}_t$得到去噪后的图像$\boldsymbol{x}_0$。基础的扩散模型推理时间一般较长，这可以通过一些快速扩散方法来降低模型推理时间（Özbey等，2023），但这些快速方法通常不能提高生成图像的质量。

基本的扩散模型的信息输入只有噪声，因此在应对跨模态转换时需要进行一些改进来加入其他模态的信息。其中最简单的思路是通过将其他模态信息作为条件直接集成到去噪过程中，从而建立一个条件扩散模型。如将其他模态数据送入到一个可训练或已经预训练好的条件特异性编码器中得到对应的条件嵌入向量，将得到的嵌入向量和训练过程中模型提取的特征混合来融合不同模态的信息。沿着这种思路，Lyu和Wang（2022）结合条件扩散模型（Ho等，2020）和得分匹配模型（Song等，2021）来实现医学成像中CT和MRI图像之间的转换。为了使扩散模型在有限的计算资源上进行训练，同时保持其质量和灵活性，Rombach等人（2022）通过将图像转移到潜空间并在潜空间中对特征进行前向扩散和逆向扩散过程（其中，图像和潜空间的转换则需要学习一个自编码模型）来保证两个空间在感知上等价。

另一种思路是通过适当的监督来指导反向去噪过程，或以较低的成本对已经训练好的扩散模型进行微调，从而避免重新训练新的扩散模型。在这条思路上，Dhariwal和Nichol（2021）将分类器网络的优化步骤作为条件信息纳入采样过程来增强预训练的扩散模型，不过生成图像的质量受限于分类器的性能。Zhu等人（2023）在二维切片自编码器网络中引入一个由一系列二维断层组成的网络，并利用三维数据对该网络进行微调，使得二维扩散模型能够扩展到其对应的三维立体影像。

从原理上可以看出，扩散模型相比于其他模型的特点在于增加了生成影像数据的多样性和视觉保真度，而不是提高生成图像的有用性，因此可以预计在对于以诊断为目的跨模态医学影像生成上无法表现出相对其他模型的显著优势。其中一个特点是越复杂的扩散模型越无法生成与目标一致的影像。

2.4 面向任务的跨模态影像合成方法

许多跨模态影像合成技术并没有考虑具体问题的特点。因此，有的方法可能会产生“幻觉”，即倾向于合成训练数据中显著存在的图像模式，但该模式可能并不是下游任务所需要的（Cohen等，2018）。为了解决该问题，需要在通用技术的基础上添加与任务相关的设计，形成对具体任务的偏置，使合成的图像保存更多有助于任务的信息，在具体任务上取得性能提升。前述感知损失就是一种这样的设计，其目的是为了平衡合成图像中的空间结构、形状与颜色、纹理等的保持与丢失。在医学影像合成中，更多的是需要合成的影像对后续的诊断、分割和配准等任务有所帮助，这通常很难通过通用的合成方法实现，因为这些方法实际上主要包含与具体任务无关的偏置，这种无关偏置主导了模型的学习过程，使朝向具体任务的偏置得到抑制，进而产生“无用”的合成影像。这也是合成的医学影像在实际应用中很难得到认可的一个重要原因。为了提高合成的影像的“有用性”，针对具体问题设计专门的合成模型是一种有效的手段。因此，产生了一系列面向任务的影像合成方法（Pan等，2021a，2022）。

2.4.1 面向任务的偏置

设某种图像合成模型的驱动目标$\boldsymbol{R}$可以分解为与任务相关的部分$\boldsymbol{R}_r$和与任务无关的部分$\boldsymbol{R}_i$，即

$$\boldsymbol{R}\left(\boldsymbol{x},\boldsymbol{y}; \boldsymbol{G}\right) = \boldsymbol{R}_r\left(\boldsymbol{x},\boldsymbol{y}; \boldsymbol{G}\right) + \boldsymbol{R}_i\left(\boldsymbol{x},\boldsymbol{y}; \boldsymbol{G}\right) \tag{16}$$

式中，$\boldsymbol{R}_r$正交于$\boldsymbol{R}_i$，使用梯度下降算法时，对应的图

像合成模型的优化方向为

$$\frac{\partial \boldsymbol{R}}{\partial \boldsymbol{G}}=\frac{\partial \boldsymbol{R}_r}{\partial \boldsymbol{G}}+\frac{\partial \boldsymbol{R}_i}{\partial \boldsymbol{G}} \tag{17}$$

同样地，$\frac{\partial \boldsymbol{R}_r}{\partial \boldsymbol{G}}$正交于$\frac{\partial \boldsymbol{R}_i}{\partial \boldsymbol{G}}$。由此可以看出，当$\left\| \frac{\partial \boldsymbol{R}_i}{\partial \boldsymbol{G}} \right\| > 0$时，优化方向会偏向任务无关的方向，合成模型合成的图像相对于任务的价值降低；当$\left\| \frac{\partial \boldsymbol{R}_i}{\partial \boldsymbol{G}} \right\| \gg \left\| \frac{\partial \boldsymbol{R}_r}{\partial \boldsymbol{G}} \right\|$时，优化方向朝向与任务无关的方向，此时合成的图像完全没有价值。在Cohen等人(2018)给出的例子中，GAN合成的影像虽然看起来更加真实，但在一些问题下甚至不如只使用MAE损失有用，原因就在于判别器的优化方向在任务无关的方向上的分量大于在任务相关方向上的分量。因此，如果需要将合成的影像用于某一下游任务，必须突出任务相关的分量。为了增加任务相关方向的优化分量，最直接的做法就是使用具体的任务模型作为影像合成目标，即让合成影像具有与真实影像在一个任务相关模型上相近的输出。设$\boldsymbol{F}: \boldsymbol{X} \to \boldsymbol{Y}$是面向某一任务(如分割、分类、配准等)的方法模型，其输入为合成模型$\boldsymbol{G}$的目标模态$\boldsymbol{y} \in \boldsymbol{Y}$，输出为$\boldsymbol{y}$对应的任务标签$\boldsymbol{f}$，该模型的求解目标为$\boldsymbol{R}_s(\boldsymbol{y}, \boldsymbol{f}; \boldsymbol{F}): \boldsymbol{f}=\boldsymbol{F}(\boldsymbol{y})$，那么将$\boldsymbol{R}_f(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{G})=\boldsymbol{R}_s(\boldsymbol{G}(\boldsymbol{x}), \boldsymbol{f}; \boldsymbol{F})$加入到合成模型的优化目标中，形成扩展的优化目标，具体为

$$\tilde{\boldsymbol{R}}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{G})=\boldsymbol{R}_r(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{G})+\boldsymbol{R}_i(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{G})+\boldsymbol{R}_f(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{G}) \tag{18}$$

式中，$\tilde{\boldsymbol{R}}$表示扩展的优化目标，$\boldsymbol{R}_f$表示任务相关的目标，通过对上述公式求偏导即可产生一个面向任务的偏置$\frac{\partial \boldsymbol{R}_f}{\partial \boldsymbol{G}}$。实验表明，这样的偏置可以显著提高合成模型对任务的适用性，将合成的影像用于训练对下游任务有一定的帮助。比如在风格转换任务中，使用多尺度结构相似性约束作为优化目标，可以显著改善图像超分辨和去噪后的失真(即提高了结构相似性)(Malhotra等，2021)；在从其他模态合成CT影像的任务中，使用阈值分割约束即可提高合成的CT影像中不同组织(骨头、软组织、脂肪和气体等)间的差异性(Liu等，2017)。

2.4.2　通过网络模型形成偏置

对于很多实际问题，往往需要使用一个复杂的模型来得到近似的解决方案。例如，用一个卷积网络来进行病灶分割、生存期预测等。尽管这些模型很可能无法达到足够好的性能，但只要它们产生的结果与真实结果之间的误差在可接受范围，依然可以用它们来为影像合成模型产生面向任务的偏置。同时，除了上述$\boldsymbol{R}_f(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{G})=\boldsymbol{R}_s(\boldsymbol{G}(\boldsymbol{x}), \boldsymbol{f}; \boldsymbol{F})$的偏置目标外，还可以通过替换其中的$\boldsymbol{F}$得到偏置目标的另一种形式：$\boldsymbol{R}_f(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{G})=\boldsymbol{R}_s(\boldsymbol{y}, \boldsymbol{F}(\boldsymbol{G}(\boldsymbol{x})); \boldsymbol{F})$。与前一种形式相比，这种形式不需要引入训练样本的监督信息，故而在求解合成模型时可以保持与原来相同的数据量。此外，如果任务模型过于复杂，还可以使用图3(d)的方式，通过提取任务模型中间层的特征，并使用特征一致性约束来减少计算量。在之前的工作中，使用的大都是这样一种形式(Pan等，2021a，2022；Pan和Xia，2021)。需要注意的是，根据网络的特点，浅层的特征对任务的表达能力通常弱于深层的特征，因此使用浅层的特征形成的对任务的偏置通常也要比深层的特征弱一些。

2.4.3　嵌入任务模型中的影像合成

虽然使用与任务相关的偏置能够使合成的影像更加适合该任务，但并不总是容易得到一个好的任务模型。因此也希望合成的影像能帮助提高任务模型的性能(Pan等，2021a)。此时可以通过优化如下模型来利用从$\boldsymbol{X}$合成的图像提高任务模型$\boldsymbol{F}$，具体为

$$\min_{\boldsymbol{F}} \boldsymbol{R}_s(\boldsymbol{G}(\boldsymbol{x}), \boldsymbol{f}; \boldsymbol{F}), \boldsymbol{x} \in \boldsymbol{X} \tag{19}$$

此时，$\boldsymbol{Y}$中的图像也可以通过下式同时加入到对$\boldsymbol{F}$的优化中，具体为

$$\min_{\boldsymbol{F}} \boldsymbol{R}_s(\boldsymbol{y}, \boldsymbol{f}; \boldsymbol{F}), \boldsymbol{y} \in \boldsymbol{Y} \tag{20}$$

此时的合成模型$\boldsymbol{G}$通过优化下式得到，以使其产生对任务的偏置，具体为

$$\min_{\boldsymbol{G}} \tilde{\boldsymbol{R}}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{G}) \tag{21}$$

在优化过程中，$\boldsymbol{F}$和$\boldsymbol{G}$的求解联合进行，其中$\boldsymbol{R}_f$项可以根据目标模态图像和标签的数量选择。这个模型为许多任务提供了新的解决思路。以跨模态配准为例，同模态(如T1-MRI和T1-MRI)之间的配准已经有许多成熟的工具，但跨模态(如FDG-PET和T1-MRI)的配准($\boldsymbol{F}$)依然是一个具有挑战的课题。在这个问题上，可以根据FDG-PET合成T1-MRI(Yang等，2020b)，进而用同模态配准方法得到配准参数，再应用于FDG-PET；也可以根据T1-MRI合成FDG-PET($\boldsymbol{G}$)，进而通过联合求解$\boldsymbol{F}$和$\boldsymbol{G}$得到FDG-

PET和T1-MRI之间的模型。

另外，有的任务模型期望同时使用多种模态来达到更好的性能（Pan等，2021a；Wei等，2021），但其中部分训练样本只有其中一种模态的影像（$\boldsymbol{X}$），也即是另一种模态图像（$\boldsymbol{Y}$）缺失了，此时我们希望使用合成模型（$\boldsymbol{G}$）来合成缺失的影像，并同时提高多模态任务模型（记为$\tilde{\boldsymbol{R}}_s\left(\boldsymbol{x},\boldsymbol{y},\boldsymbol{f};\ \boldsymbol{F}\right),\boldsymbol{f}=\boldsymbol{F}\left(\boldsymbol{x},\boldsymbol{y}\right)$），这种情况依然可以通过联合优化$\boldsymbol{F}$和$\boldsymbol{G}$实现，只需要将此时的优化目标改为

$$\boldsymbol{R}_f\left(\boldsymbol{x},\boldsymbol{y};\boldsymbol{G}\right)=\tilde{\boldsymbol{R}}_s\left(\boldsymbol{x},\boldsymbol{G}\left(\boldsymbol{x}\right),\boldsymbol{f};\boldsymbol{F}\right) \tag{22}$$

以阿尔茨海默病的诊断为例，通常认为同时使用MRI和PET是达到更好性能的有效手段，但许多样本没有采集PET影像。因此，多模态MRI-PET诊断模型面临数据缺失的问题。使用上述方法便可补全缺失数据，利用所有的样本训练多模态诊断模型。

3 跨模态医学影像合成的应用

跨模态影像合成技术在成像、重建、配准、分割、预测、诊断等医学影像智能计算的各个任务中都有应用，同时也涉及到MRI、PET、CT和X-光平片等影像模态。在成像过程中，受到患者条件或设备的限制，会存在一些模态的影像难以获得或获得到的影像质量不佳。在这种情况下，跨模态影像合成技术可以通过合成高质量的影像来辅助医生进行更准确的诊断。在医学影像的重建过程中，使用跨模态影像合成技术合成的影像可以用于三维结构的重建，提供更全面的解剖信息。同时，跨模态影像合成技术在影像配准中也十分重要，它能够将来自不同模态的医学影像精确对齐，以便于比较和分析。此外，在分割任务中，跨模态影像合成技术合成的影像可以提供额外的对比度或特征，帮助算法更准确地识别和分割出感兴趣的区域。在疾病的发展预测和诊断中，跨模态影像合成技术可以结合不同模态的优势，提供更全面的疾病特征。例如，PET影像可以提供关于代谢活动的信息，而MRI可以提供详细的解剖结构信息，两者的合成影像可以为医生提供更全面的诊断依据。

当前，跨模态医学影像合成的应用场景包括但不限于影像转换、数据扩充、数据统一、隐私保护和可信智能等。在机器学习模型的训练中，数据的多样性和数量对模型性能有重要影响。跨模态影像合成技术可以通过生成新的合成影像来扩充数据集，提高模型的泛化能力。同时，它还可以帮助实现不同模态数据的统一，简化数据处理流程。而在医学研究和数据分析中，保护患者隐私至关重要。跨模态影像合成技术可以创建去标识化的合成影像，用于研究和分析，而不泄露患者的个人信息。

本节将从影像转换的应用领域和典型的优势任务两个方面对跨模态医学影像的应用进行介绍。

3.1 影像转换的应用领域

传统的医学影像合成方法通常都受限于其特定的转换模式，即仅能在特定的模态之间实现影像转换。而跨模态医学影像合成不仅限于合成特定的模态，还能够实现任意缺失模态的合成，具有更高的灵活性。同时跨模态医学影像合成技术相较于传统的医学影像合成方法能够合成更加高质量的医学影像。跨模态医学影像合成技术在影像转换领域的应用逐渐显现出其在提高诊断准确性、优化治疗计划和加速临床研究中的潜力。从影像转换的目的来说，这些应用大致可以分为影像替代和影像补全两类。而从影像转换数据的角度来看，这些应用又可以大致分为不同模态之间的图像转换和同一模态的两种不同成像参数之间的转换。如不同模态之间的图像转换方法有使用CT影像合成MR影像，同一模态的不同对比度成像之间的转换方法有使用T1WI影像合成T2WI影像。本节从两种不同的角度介绍跨模态医学影像合成技术在影像转换领域的应用。

3.1.1 影像的等效替代

由于不同的模态影像可能存在互补的信息，在现实诊断中往往需要获取全面的影像数据。而由于设备的可用性、患者自身的条件和成像成本等问题，导致某些模态的数据可能难以获得。为了解决这一问题，可以采用影像的等效替代方法，可以尝试利用其他可获得模态的影像合成出该模态的影像。例如，PET成像通常依赖CT影像进行辐射衰减校正，但在PET/MRI设备中无法采集CT影像，此时可以利用MRI影像或未校正的PET影像合成CT影像，用以完成PET校正（Bahrami等，2020）。同时，有些模态的影像采集过程需要依赖耗费高或耗时长的成像方式，难以在临床中广泛的使用。例如，脑血容量（cerebral blood volume，CBV）是评价颅内占位性病

变最有用的参数，但 CBV 的测量依赖于血流灌注成像技术，存在成像时间长、成本高、给患者带来极大不适等明显缺点。考虑到CBV 影像中的信息可能也部分存在于其他MRI序列中，可以尝试利用多个MRI序列来合成CBV影像，从而获得一个近似的结果（Pan等，2021b）。

3.1.2 缺失影像的补全

对于有些任务，使用多种模态的影像相互配合，可以达到更好的精度。例如，同时使用 MRI 和 PET 影像可以建立更加准确的阿尔茨海默病诊断模型（Pan等，2021a，2022），同时使用多种 MRI 序列可以提升胶质瘤分割精度，以便显现更加完整的病灶面貌（Jia等，2020）。但由于病人意愿或条件限制等原因，数据不完整的情况非常普遍。这时，可以通过影像合成技术使用已有模态的影像合成缺失模态的影像，从而利用所有的样本来训练模型，并可以将该模型应用于可能存在缺失模态影像的样本上。需要注意的是，影像替代和影像补全虽然技术上是相似的，但在目的上有着本质的区别。前者假设源模态包含目标模态的完整信息，希望挖掘源模态与目标模态的共有信息，并将这种信息以目标模态的模式呈现出来。后者则假设不同模态中的信息是互补的，希望合成的目标模态影像在共有信息的基础上呈现出与已有模态影像互补的信息，只有这样合成影像才能在下游任务中发挥正向的作用。例如，在CT影像中，不同组织通常对应不同的HU（Hounsfiled unit）值，因此在使用MRI影像合成CT影像时，希望MRI影像所反映的组织分布信息以HU值的形式呈现。而在脑肿瘤的分割中，不同模态的影像都只有肿瘤区域的一部分信息，合成的影像只有提供本模态特有的与其他模态互补的信息，才能提高分割精度。这是很难做到的，即便使用面向任务的合成技术也很难满足期待，多模态模型性能的提升通常是得益于样本数量的增加或数据多样性的提高。

3.1.3 多源数据的统一化

医学影像通常呈现出多样化的特点，即使同一种模态的影像，也可能存在分辨率、噪声和数值分布等方面较大的差异。不同的医疗机构、研究中心可能因为使用的设备、成像协议和数据记录标准的不同，也会导致所得到的相同类型的数据在数据格式和质量上存在差异。而这些因为成像设备、成像协议等导致的数据分布上存在差异的数据集称为跨中心数据集。受到跨中心数据集异质性的影响，在将训练数据集上获得的模型应用于跨中心数据时，其性能会出现难以预测的下降。如果将这种富含多样性的影像视为广义的多模态影像，就可以使用跨模态影像合成技术将多中心的数据变成统一的风格，从而提高任务模型跨中心的迁移能力（Lei等，2022）。有时在某一模态上的标注比其他模态容易，如果能够建立不同模态的对应关系，则可以更容易地获得在所有模态上的标注（Taleb等，2021；Yamashita等，2021；Su等，2023）。比如，在多模态和多序列的医学图像上只需要在一种模态或序列上的标注就可以训练全模态/全序列的分割模型（Su等，2023），将非医学图像的风格应用于组织病理学图像则可能提高图像分类方法在肿瘤分类上的能力（Yamashita等，2021）。

3.2 典型的优势任务

医学影像的生成主要目的是提升医学影像处理流程中各个任务的性能，包括成像、重建、配准和分割等。本节将围绕这几个典型的优势任务介绍跨模态医学影像的应用。

3.2.1 成像

医学影像领域的成像任务是指通过各种成像技术获取人体内部结构和功能的信号以便于进行诊断、治疗活动。有时因为患者状况或技术设备的限制，某些医学影像难以获得或获得到的影像质量不佳，在这种情况下使用跨模态影像合成技术则有可能创造出高质量的图像，帮助医生做出更精确的诊断。例如MRI的多序列配合能得到更充分的疾病信息，但由于采集时间较长或采集条件受限，只能采集厚层的或不完整的多序列数据，如果某些序列能够使用其他序列生成出来，则可以节省时间，采集更高质量的影像数据（Han等，2018）。MRI引导放疗仍然需要CT图像来计算剂量分布，而MRI和CT要使用两个设备分开采集，显著延长了诊疗周期，如果使用MRI或其他模态的影像生成与之配对的CT影像则可以获得配对的MRI/CT数据，缩减诊疗流程（Armanious等，2019；Touati等，2021；Dovletov等，2022；Zhou等，2023）。

3.2.2 重建

医学影像重建中可以通过深度学习技术来提高低质量的图像，从而降低噪声，提高图像的清晰度、对比度和分辨率。跨模态医学影像合成技术在图像

重建领域的应用主要在于实现局部图像修复和加速图像重建，从而生产高质量的临床使用的医疗图像，同时降低成本和患者风险。例如，使用传统的滤波反投影算法进行CT影像重建会由于X射线视场（Field of view，FOV）受限出现严重的截断伪影，此时可以使用生成网络从截断的正弦图获取扩展的正弦图来改善重建图像的质量（Ketola等，2021）。MRI扫描时间较长且对时间分辨率要求高，快速成像模式采集的降采样的MRI质量较差，此时可以通过生成技术来估计缺失的K空间样本来保持重建影像在解剖学意义上的质量（Shaul等，2020）。MRI和CT影像的分辨率受到成像设备和采集条件的限制，其质量可能无法满足诊疗的精确性，此时可以使用生成技术实现影像的超分辨，从而为临床诊疗提供更可靠和有效的基础（Xiao等，2023）。CT、PET影像存在辐射危害，但低剂量影像通常含有明显的噪声和伪影，此时可以使用生成技术从低剂量的CT/PET影像生成标准剂量的CT/PET，使得低剂量情况下可以重建出与标准剂量一致的影像（Zhang等，2021；Xu等，2017），从而保证诊断结果的一致性。

3.2.3　配准

医学图像配准是指通过计算和应用空间变换将不同时间或不同条件下获取的医学图像精确对齐，使它们在解剖结构和功能特征上相互匹配，从而便于医生进行比较分析、疾病监测、治疗规划和手术导航。传统的配准方法通过优化均方误差、归一化互信息等目标函数来迭代优化变形场，在三维的医学影像上非常耗时。对于配准问题可以使用生成技术获得配准后的图像和相应的变形场，从而避免了耗时的多次迭代（Mahapatra等，2018）。同时，由于传统配准的迭代方式和扩散模型的采样相似，因此可以用扩散模型来模拟配准的形变过程（Kim等，2022），虽然这种模拟没有速度上的优势，但其可学习方式有望获得更强的配准能力。传统的配准算法在跨模态影像上常常效果不佳甚至失败，如果能使用跨模态生成技术将不同模态的影像转变成相同模态的影像，则有望提高跨模态配准的成功率以及配准精度（Song等，2022）。

3.2.4　分割

医学图像分割是医学图像处理中的一个关键步骤，其目的是将图像划分为不同的区域或对象，分离出特定的解剖结构（如器官、血管和肿瘤等），以便进一步分析和研究。现有的分割方法可能存在标注数据不足、分割精度受限等问题。这些问题在一些场景下有望通过跨模态医学影像合成技术解决。例如将有分割标注的模态转换为无分割标注的模态，产生适用于无分割标注模态的训练数据集，从而在没有目标模态影像的手动标注的情况下，训练用于目标成像模态的分割网络（Huo等，2019）。某些分割任务特定的模态影像上才能达到较好的精度，此时也可以将其他模态影像转换成目标模态，从而获得更好的结果（Zhou等，2019）。某些疾病在增强影像上比普通影像上更明显，此时可以从增强影像上生成普通影像，并通过对比增强影像和生成的普通影像获得病灶区域（Liu等，2022）。

4　结语与展望

本文介绍人工智能内容生成技术在医学影像领域的跨模态合成问题上的方法及其应用。首先介绍该项研究的意义和技术难点，然后回顾相关技术的发展和应用场景。目前，跨模态医学影像合成已得到广泛研究，并在医学影像智能计算的诸多方向中有所应用，给许多问题的解决带来新思路。但是现有的跨模态医学影像合成技术仍面临许多亟待解决的问题：1）虽然合成的影像质量在不断提高，但当前技术合成的影像和真实影像仍然存在较大差异，面临明显的临床失效风险；2）现有的跨模态生成模型网络结构仍然有较大改进空间，如基于U-Net的跨模态合成模型受限于固定的感受野，导致捕获的信息有限，基于扩散模型的合成技术由于引入随机噪声经常会破坏输入图像的原始数据分布和空间结构；3）与自然图像相比，一些医学影像数据收集十分困难，跨模态医学影像合成技术经常面临数据集过小的问题。展望未来，跨模态医学影像生成可能在以下方向发展：1）受限于医学影像数据量小的问题，将无监督学习或小样本学习的思路引入到生成模型的训练中；2）将各种各样的应用目标融入合成模型的设计中，从而产生面向特定应用的生成模型。其中，如何融入目标任务相关的知识、如何设计与目标任务相关的模型、如何对影像合成模型进行专业化的改进等，都值得进一步深入研究。

参考文献(References)

Armanious K, Jiang C M, Abdulatif S, Küstner T, Gatidis S and Yang B. 2019. Unsupervised medical image translation using cycle-MedGAN//Proceedings of the 27th European Signal Processing Conference. A Coruna, Spain: IEEE: 1-5 [DOI: 10.23919/EUSIPCO.2019.8902799]

Bahrami A, Karimian A, Fatemizadeh E, Arabi H and Zaidi H. 2020. A new deep convolutional neural network design with efficient learning capability: application to CT image synthesis from MRI. Medical Physics, 47(10): 5158-5171 [DOI: 10.1002/mp.14418]

Bhagwatkar R, Bachu S, Fitter K, Kulkarni A and Chiddarwar S. 2020. A review of video generation approaches//Proceedings of 2020 International Conference on Power, Instrumentation, Control and Computing (PICC). Thrissur, India: IEEE: 1-5 [DOI: 10.1109/PICC51425.2020.9362485]

Borsos Z, Marinier R, Vincent D, Kharitonov E, Pietquin O, Sharifi M, Roblek D, Teboul O, Grangier D, Tagliasacchi M and Zeghidour N. 2023. AudioLM: a language modeling approach to audio generation. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 31: 2523-2533 [DOI: 10.1109/TASLP.2023.3288409]

Božić M and Horvat M. 2024. A survey of deep learning audio generation methods [EB/OL]. [2024-08-05]. https://arxiv.org/pdf/2406.00146.pdf

Cao Y H, Li S Y, Liu Y X, Yan Z L, Dai Y T, Yu P S and Sun L C. 2023. A comprehensive survey of AI-generated content (AIGC): a history of generative AI from GAN to chatGPT. arXiv preprint arXiv: 2303.04226 [DOI: 10.48550/arXiv.2303.04226]

Cohen J P, Luck M and Honari S. 2018. Distribution matching losses can hallucinate features in medical image translation//Proceedings of the 21st International Conference on Medical Image Computing and Computer Assisted Intervention — MICCAI 2018. Granada, Spain: Springer: 529-536 [DOI: 10.1007/978-3-030-00928-1_60]

Dhariwal P and Nichol A. 2021. Diffusion models beat GANs on image synthesis//Proceedings of the 35th International Conference on Neural Information Processing Systems. Online: Curran Associates Inc.: 8780-8794

Di J. 2024. Principles of AIGC technology and its application in new media micro-video creation. Applied Mathematics and Nonlinear Sciences, 9(1): 1-13 [DOI: 10.2478/amns-2024-1393]

Dovletov G, Pham D D, Lörcks S, Pauli J, Gratz M and Quick H H. 2022. Grad-CAM guided U-Net for MRI-based pseudo-CT synthesis//Proceedings of the 44th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Glasgow, United Kingdom: IEEE: 2071-2075 [DOI: 10.1109/EMBC48229.2022.9871994]

Elangovan A and Jeyaseelan T. 2016. Medical imaging modalities: a survey//Proceedings of 2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS). Pudukkottai, India: IEEE: 1-4 [DOI: 10.1109/ICETETS.2016.7603066]

Fard A S, Reutens D C and Vegh V. 2021. CNNs and GANs in MRI-based cross-modality medical image estimation [EB/OL]. [2024-08-05]. https://arxiv.org/pdf/2106.02198.pdf

Foo L G, Rahmani H and Liu J. 2025. AI-generated content (AIGC) for various data modalities: a survey [EB/OL]. [2024-08-05]. https://arxiv.org/pdf/2308.14177.pdf

Frolov S, Hinz T, Raue F, Hees J and Dengel A. 2021. Adversarial text-to-image synthesis: a review. Neural Networks, 144: 187-209 [DOI: 10.1016/j.neunet.2021.07.019]

Han C, Hayashi H, Rundo L, Araki R, Shimoda W, Muramatsu S, Furukawa Y, Mauri G and Nakayama H. 2018. GAN-based synthetic brain MR image generation//Proceedings of the 15th IEEE International Symposium on Biomedical Imaging (ISBI 2018). Washington, USA: IEEE: 734-738 [DOI: 10.1109/ISBI.2018.8363678]

Ho J, Jain A and Abbeel P. 2020. Denoising diffusion probabilistic models//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: 6840-6851

Huang Y W, Shao L and Frangi A F. 2018. Cross-modality image synthesis via weakly coupled and geometry co-regularized joint dictionary learning. IEEE Transactions on Medical Imaging, 37(3): 815-827 [DOI: 10.1109/TMI.2017.2781192]

Huo Y K, Xu Z B, Moon H, Bao S X, Assad A, Moyo T K, Savona M R, Abramson R G and Landman B A. 2019. SynSeg-Net: synthetic segmentation without target modality ground truth. IEEE Transactions on Medical Imaging, 38(4): 1016-1025 [DOI: 10.1109/TMI.2018.2876633]

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]

Jack C R Jr, Bernstein M A, Fox N C, Thompson P, Alexander G, Harvey D, Borowski B, Britson P J, Whitwell J L, Ward C, Dale A M, Felmlee J P, Gunter J L, Hill D L G, Killiany R, Schuff N, Fox-Bosetti S, Lin C, Studholme C, DeCarli C S, Krueger G, Ward H A, Metzger G J, Scott K T, Mallozzi R, Blezek D, Levy J, Debbins J P, Fleisher A S, Albert M, Green R, Bartzokis G, Glover G, Mugler J and Weiner M W. 2008. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. Journal of Magnetic Resonance Imaging, 27(4): 685-691 [DOI: 10.1002/jmri.21049]

Jia H Z, Xia Y, Cai W D and Huang H. 2020. Learning high-resolution and efficient non-local features for brain glioma segmentation in

MR images//Proceedings of the 23rd International Conference on Medical Image Computing and Computer Assisted Intervention — MICCAI 2020. Lima, Peru: Springer: 480-490 [DOI: 10.1007/978-3-030-59719-1_47]

Jog A, Carass A, Roy S, Pham D L and Prince J L. 2017. Random forest regression for magnetic resonance image synthesis. Medical Image Analysis, 35: 475-488 [DOI: 10.1016/j. media. 2016. 08.009]

Ketola J H J, Heino H, Juntunen M A K, Nieminen M T, Siltanen S and Inkinen S I. 2021. Generative adversarial networks improve interior computed tomography angiography reconstruction. Biomedical Physics and Engineering Express, 7(6): #065041 [DOI: 10.1088/2057-1976/ac31cb]

Kim B, Han I and Ye J C. 2022. DiffuseMorph: unsupervised deformable image registration using diffusion model//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 347-364 [DOI: 10.1007/978-3-031-19821-2_20]

Lei H J, Liu W X, Xie H, Zhao B J, Yue G H and Lei B Y. 2022. Unsupervised domain adaptation based image synthesis and feature alignment for joint optic disc and cup segmentation. IEEE Journal of Biomedical and Health Informatics, 26(1): 90-102 [DOI: 10.1109/JBHI.2021.3085770]

Li R J, Zhang W L, Suk H I, Wang L, Li J, Shen D G and Ji S W. 2014. Deep learning based imaging data completion for improved brain disease diagnosis//Proceedings of the 17th International Conference on Medical Image Computing and Computer-Assisted Intervention. Boston, USA: Springer: 305-312 [DOI: 10.1007/978-3-319-10443-0_39]

Li Y T, Min M, Shen D H, Carlson D and Carin L. 2018. Video generation from text//Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans, USA: AAAI: 7065-7072 [DOI: 10.1609/aaai.v32i1.12233]

Liu H F, Chen J J, Li L, Bao B K, Li Z C, Liu J Y and Nie L Q. 2023. Cross-modal representation learning and generation. Journal of Image and Graphics, 28(6): 1608-1629 (刘华峰, 陈静静, 李亮, 鲍秉坤, 李泽超, 刘家瑛, 聂礼强. 2023. 跨模态表征与生成技术. 中国图象图形学报, 28(6): 1608-1629) [DOI: 10.11834/jig.230035]

Liu M Y, Breuel T and Kautz J. 2017. Unsupervised image-to-image translation networks//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 700-708

Liu Y, Mu F H, Shi Y, Cheng J, Li C and Chen X. 2022. Brain tumor segmentation in multimodal MRI via pixel-level and feature-level image fusion. Frontiers in Neuroscience, 16: #1000587 [DOI: 10.3389/fnins.2022.1000587]

Luo N, Song M, Yang Z Y and Jiang T Z. 2022. Survey on cross-modal fusion based on brain atlas data. Journal of Image and Graphics, 27(6): 2036-2056 (罗娜, 宋明, 杨正宜, 蒋田仔. 2022. 跨模态脑图谱数据融合研究进展. 中国图象图形学报, 27(6): 2036-2056) [DOI: 10.11834/jig.220036.]

Lyu Q and Wang G. 2022. Conversion between CT and MRI images using diffusion and score-matching models [EB/OL]. [2024-08-05]. https://arxiv.org/pdf/2209.12104.pdf

Mahapatra D, Antony B, Sedai S and Garnavi R. 2018. Deformable medical image registration using generative adversarial networks// The 15th IEEE International Symposium on Biomedical Imaging. Washington, USA: IEEE: 1449-1453 [DOI: 10.1109/ISBI.2018.8363845]

Malhotra R, Sharma K, Kumar K and Rath N. 2021. Integrating SSIM in GANs to generate high-quality brain MRI images//Proceedings of 2021 International Conference on Data Engineering and Communication Technology. Singapore, Singapore: Springer: 419-426 [DOI10.1007/978-981-16-0081-4_42]

Nie D, Cao X H, Gao Y Z, Wang L and Shen D G. 2016. Estimating CT image from MRI data using 3D fully convolutional networks//Proceedings of the 1st International Workshop on Deep Learning and Data Labeling for Medical Applications, DLMIA 2016. Athens, Greece: Springer: 170-178 [DOI: 10.1007/978-3-319-46976-8_18]

Özbey M, Dalmaz O, Dar S U H, Bedel H A, Özturk Ş, Güngör A and Çukur T. 2023. Unsupervised medical image translation with adversarial diffusion models [EB/OL]. [2024-08-05]. https://arxiv.org/pdf/2207.08208.pdf

Pan Y S, Chen Y T, Shen D G and Xia Y. 2021a. Collaborative image synthesis and disease diagnosis for classification of neurodegenerative disorders with incomplete multi-modal neuroimages//Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention. Strasbourg, France: Springer: 480-489 [DOI: 10.1007/978-3-030-87240-3_46]

Pan Y S, Huang J Y, Wang B, Zhao P, Liu Y C and Xia Y. 2021b. Cerebral blood volume prediction based on multi-modality magnetic resonance imaging//Proceedings of the 6th International Workshop on Simulation and Synthesis in Medical Imaging. Strasbourg, France: Springer: 121-130 [DOI: 10.1007/978-3-030-87592-3_12]

Pan Y S, Liu M X, Lian C F, Xia Y and Shen D G. 2020. Spatially-constrained fisher representation for brain disease identification with incomplete multi-modal neuroimages. IEEE Transactions on Medical Imaging, 39(9): 2965-2975 [DOI: 10.1109/TMI.2020.2983085]

Pan Y S, Liu M X, Xia Y and Shen D G. 2022. Disease-image-specific learning for diagnosis-oriented neuroimage synthesis with incomplete multi-modality data. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(10): 6839-6853 [DOI: 10.1109/TPAMI.2021.3091214]

Pan Y S and Xia Y. 2021. Ultimate reconstruction: understand your bones from orthogonal views//2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). Nice, France: IEEE: 1155-1158 [DOI: 10.1109/ISBI48211.2021.9433758]

Pérez-Grijalba V, Arbizu J, Romero J, Prieto E, Pesini P, Sarasa L, Guillen F, Monleóonl, San-Jos Fl, MartJnez-Lage P, Munuera J, Hernández I, Buendíuen, Sotolongo-Grau O, Alegret M, Ruiz A, T rraga L, Boada M, Sarasa M, and The AB255 Study Group. 2019. Plasma Aβ42/40 ratio alone or combined with FDG-PET can accurately predict amyloid-PET positivity: a cross-sectional analysis from the AB255 Study. Alzheimer's Research and Therapy, 11(1): #96 [DOI: 10.1186/s13195-019-0549-1]

Ren X F, Tong L L, Zeng J and Zhang C. 2023. AIGC scenario analysis and research on technology roadmap of Internet industry application. China Communications, 20(10): 292-304 [DOI: 10.23919/JCC.fa.2023-0359.202310]

Rombach R, Blattmann A, Lorenz D, Esser P and Ommer B. 2022. High-resolution image synthesis with latent diffusion models//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 10674-10685 [DOI: 10.1109/CVPR52688.2022.01042]

Shaul R, David I, Shitrit O and Riklin Raviv T. 2020. Subsampled brain MRI reconstruction by generative adversarial neural networks. Medical Image Analysis, 65: #101747 [DOI: 10.1016/j.media.2020.101747]

Shi J, Wang L L, Wang S S, Chen Y X, Wang Q, Wei D M, Liang S J, Peng J L, Yi J J, Liu S F, Ni D, Wang M L, Zhang D Q and Shen D G. 2020. Applications of deep learning in medical imaging: a survey. Journal of Image and Graphics, 25(10): 1953-1981 (施俊, 汪琳琳, 王珊珊, 陈艳霞, 王乾, 魏冬铭, 梁淑君, 彭佳林, 易佳锦, 刘盛锋, 倪东, 王明亮, 张道强, 沈定刚. 2020. 深度学习在医学影像中的应用综述. 中国图象图形学报, 25(10): 1953-1981) [DOI: 10.11834/jig.200255]

Singh N K and Raza K. 2021. Medical image generation using generative adversarial networks: a review//Patgiri R, Biswas A and Roy P, eds. Health Informatics: A Computational Perspective in Healthcare. Singapore: Springer: 77-96 [DOI: 10.1007/978-981-15-9735-0_5]

Song X R, Chao H Q, Xu X A, Guo H T, Xu S, Turkbey B, Wood B J, Sanford T, Wang G and Yan P K. 2022. Cross-modal attention for multi-modal image registration. Medical Image Analysis, 82: #102612 [DOI: 10.1016/j.media.2022.102612]

Song Y, Sohl-Dickstein J, Kingma D P, Kumar A, Ermon S and Poole B. 2021. Score-based generative modeling through stochastic differential equations [EB/OL]. [2024-08-05]. https://arxiv.org/pdf/2011.13456.pdf

Su Z X, Yao K, Yang X, Huang K Z, Wang Q F and Sun J. 2023. Rethinking data augmentation for single-source domain generalization in medical image segmentation//Proceedings of the 37th AAAI Conference on Artificial Intelligence. Washington, USA: AAAI: 2366-2374 [DOI: 10.1609/aaai.v37i2.25332]

Taleb A, Lippert C, Klein T and Nabi M. 2021. Multimodal self-supervised learning for medical image analysis//Proceedings of the 27th International Conference on Information Processing in Medical Imaging. Virtual Event: Springer: 661-673 [DOI: 10.1007/978-3-030-78191-0_51]

Tanitame N, Tanitame K and Awai K. 2017. Clinical utility of optimized three-dimensional T1-, T2-, and T2*-weighted sequences in spinal magnetic resonance imaging. Japanese Journal of Radiology, 35(4): 135-144 [DOI: 10.1007/s11604-017-0621-3]

Touati R, Le W T and Kadoury S. 2021. A feature invariant generative adversarial network for head and neck MRI/CT image synthesis. Physics in Medicine and Biology, 66(9): #095001 [DOI: 10.1088/1361-6560/abf1bb]

Vijayalaxmi, Fatahi M and Speck O. 2015. Magnetic resonance imaging (MRI): a review of genetic damage investigations. Mutation Research/Reviews in Mutation Research, 764: 51-63 [DOI: 10.1016/j.mrrev.2015.02.002]

Villarraga-Gómez H, Herazo E L and Smith S T. 2019. X-ray computed tomography: from medical imaging to dimensional metrology. Precision Engineering, 60: 544-569 [DOI: 10.1016/j. precisioneng.2019.06.007]

Wang S Y, Wu W W, Feng J, Liu F L and Yu H Y. 2020. Low-dose spectral CT reconstruction based on image-gradient L_0-norm and adaptive spectral PICCS. Physics in Medicine and Biology, 65(24): #245005 [DOI: 10.1088/1361-6560/aba7cf]

Wang T C, Liu M Y, Zhu J Y, Tao A, Kautz J and Catanzaro B. 2018. High-resolution image synthesis and semantic manipulation with conditional GANs//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 8798-8807 [DOI: 10.1109/CVPR.2018.00917]

Wei J, Pan Y, Xia Y S and Shen D G. 2021. Learning to synthesize 7T MRI from 3T MRI with few data by deformable augmentation//12th International Workshop on Machine Learning in Medical Imaging. Strasbourg, France: Springer: 70-79 [DOI: 10.1007/978-3-030-87589-3_8]

Xiang L, Qiao Y, Nie D, An L, Lin W L, Wang Q and Shen D G. 2017. Deep auto-context convolutional neural networks for standard-dose PET image estimation from low-dose PET/MRI. Neurocomputing, 267: 406-416 [DOI: 10.1016/j.neucom.2017.06.048]

Xiao Y Y, Chen C X, Wang L, Yu J, Fu X, Zou Y, Lin Z and Wang K P. 2023. A novel hybrid generative adversarial network for CT and MRI super-resolution reconstruction. Physics in Medicine and Biology, 68(13): #135007 [DOI: 10.1088/1361-6560/acdc7e]

Xu J S, Gong E H, Pauly J and Zaharchuk G. 2017. 200× low-dose PET reconstruction using deep learning [EB/OL]. [2024-08-05]. https://arxiv.org/pdf/1712.04119.pdf

Yamashita R, Long J, Banda S, Shen J and Rubin D L. 2021. Learning domain-agnostic visual representation for computational pathology using medically-irrelevant style transfer augmentation. IEEE Transactions on Medical Imaging, 40(12): 3945-3954 [DOI: 10.1109/TMI.2021.3101985]

Yang H, Qian P J and Fan C. 2020a. An indirect multimodal image registration and completion method guided by image synthesis. Computational and Mathematical Methods in Medicine, 2020: #2684851 [DOI: 10.1155/2020/2684851]

Yang Q Y, Li N N, Zhao Z X, Fan X Y, Chang E I C and Xu Y. 2020b. MRI cross-modality image-to-image translation. Scientific Reports, 10(1): #3753 [DOI: 10.1038/s41598-020-60520-6]

Zhang Y K, Hu D L, Zhao Q L, Quan G T, Liu J, Liu Q G, Zhang Y, Coatrieux G, Chen Y and Yu H Y. 2021. CLEAR: comprehensive learning enabled adversarial reconstruction for subtle structure enhanced low-dose CT imaging. IEEE Transactions on Medical Imaging, 40(11): 3089-3101 [DOI: 10.1109/TMI.2021.3097808]

Zhou T X, Ruan S and Canu S. 2019. A review: deep learning for medical image segmentation using multi-modality fusion. Array, 3-4: #100004 [DOI: 10.1016/j.array.2019.100004]

Zhou X R, Cai W W, Cai J J, Xiao F, Qi M K, Liu J W, Zhou L H, Li Y B and Song T. 2023. Multimodality MRI synchronous construction based deep learning framework for MRI-guided radiotherapy synthetic CT generation. Computers in Biology and Medicine, 162: #107054 [DOI: 10.1016/j.compbiomed.2023.107054]

Zhu L T, Xue Z Y, Jin Z C, Liu X, He J Z, Liu Z W and Yu L Q. 2023. Make-a-volume: leveraging latent diffusion models for cross-modality 3D brain MRI synthesis [EB/OL]. [2024-08-05]. https://arxiv.org/pdf/2307.10094.pdf

作者简介

潘永生，男，教授，主要研究方向为医学影像健康计算、跨模态影像生成、模式识别与机器学习。

E-mail: yspan@nwpu.edu.cn

夏勇，通信作者，男，教授，主要研究方向为医学影像健康计算、跨模态影像生成、模式识别与机器学习。

E-mail：yxia@nwpu.edu.cn

马豪杰，男，硕士研究生，主要研究方向为计算机视觉、跨模态影像生成、模式识别与机器学习。

E-mail：a2233539443@gmail.com

张艳宁，女，教授，主要研究方向为图像处理与计算机视觉、模式识别与人工智能、机器学习。

E-mail：ynzhang@nwpu.edu.cn