中图法分类号:TP391　　**文献标识码**: A　　**文章编号**: 1006-8961(XXXX)XX-0001-33

论文引用格式:Li Xiaohui, Wang Yutong, Zhang Yuhan, Niu Sijie, Chen Qiang. Generative prediction of longitudinal medical imaging: A survey[J/OL]. Journal of Image and Graphics, XXXX:1-33. DOI: 10.11834/jig.260414. (李孝辉, 汪语桐, 张语涵, 牛四杰, 陈强. 纵向医学图像生成式预测综述[J/OL]. 中国图象图形学报,XXXX:1-33. DOI: 10.11834/jig.260414.)[DOI:10.11834/jig.260414]

纵向医学图像生成式预测综述

李孝辉[1],汪语桐[1],张语涵[2],牛四杰[3],陈强[1]

1. 南京理工大学计算机科学与工程学院,南京　210094; 2. 深圳大学医学部生物医学工程学院,深圳　518060; 3. 济南大学信息科学与工程学院,济南　250022

摘　要: 对疾病演变、疗效评估与器官发育的动态监测是精准医疗的核心需求。纵向医学成像(Longitudinal medical imaging)通过长期序列采集,为捕捉疾病跨时空演变与预后趋势提供了关键无创依据。近年来,由基线历史图像定向合成未来指定时间步影像或形态流形的纵向预测生成任务备受关注,涌现出多种技术路线。据作者所知,本文是首篇专门面向纵向医学图像生成式预测的综述,旨在厘清该领域的底层逻辑、技术边界与演进脉络。现有方法可归纳为三大体系:1)时序建模方法:以循环神经网络(recurrent neural networks, RNN)、长短期记忆网络(long short-term memory, LSTM)和视觉Transformer为代表,通过专用时序架构捕获时间依赖,建模直观但高维生成能力有限;2)深度生成模型:作为主导路线,涵盖生成对抗网络(generative adversarial networks, GAN)、扩散模型(diffusion model)/流匹配(flow matching)及变分自编码器(variational autoencoder, VAE),通过学习数据分布实现条件生成,合成清晰度与细节最优;3)物理/先验引导模型:借助反应扩散方程(reaction diffusion)、神经微分方程(differential equation)和物理信息神经网络(physical information neural networks, PINN)等将生物学规律编码为约束,在可解释性与小样本效率上优势显著。本文首先对任务进行严格数学定义,梳理公开随访数据集与多维评估指标;继而对现有方法开展系统分类与多维对比,总结各临床领域的方法偏好与演进脉络,剖析其结构特点与局限;最后探讨该领域的关键挑战与未来趋势。

关键词: 纵向医学图像;未来图像预测;疾病进展建模;时序模型;生成模型;物理模型

Generative prediction of longitudinal medical imaging: A survey

Li Xiaohui[1], Wang Yutong[1], Zhang Yuhan[2], Niu Sijie[3], Chen Qiang[1]

1. School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China; 2. School of Biomedical Engineering, Shenzhen University Medical School, Shenzhen 518060, China; 3. School of Information Science and Engineering, University of Jinan, Jinan 250022, China

Abstract: The dynamic monitoring of spatiotemporal disease pathology, individualized therapeutic response assessment, and long-term organ development constitutes the scientific foundation and core clinical demand of modern precision medicine. Longitudinal medical imaging, which involves the acquisition of sequential image scans from the same patient over extended follow-up intervals, provides indispensable non-invasive evidence for capturing dynamic evolutionary patterns and prognostic transition trends across multiple spatial and temporal dimensions. However, accurately extrapolating a patient′s future anatomical structures, organ deformations, and localized lesion evolutions remains one of the most formidable fron-

收稿日期:2026-07-10;**修回日期**:2026-11-18

基金项目:国家自然科学基金重点研发计划(92370109);国家自然科学基金(62172223);江苏省自然科学基金(BK20252033)

Supported by: Major Research Plan of the National Natural Science Foundation of China (92370109); National Natural Science Foundation of China (62172223); Jiangsu Provincial Natural Science Foundation (BK20252033)

© 中国图象图形学报版权所有

tends of research within the joint communities of computer vision, deep learning, and medical image analysis, particularly when dealing with the clinical reality of sparse, irregularly sampled, and incomplete follow-up histories. In recent years, the task of image-to-image longitudinal prediction and generation, which aims to directionally synthesize high-fidelity future images or shape manifolds at arbitrary specified time steps based on one or multiple baseline historical scans, has emerged as a highly active research area, yielding a diverse methodology landscape. To the best of our knowledge, this paper presents the first comprehensive survey specifically dedicated to generative prediction in longitudinal medical imaging, aiming to clarify the underlying logic, core technical boundaries, and methodological evolution of this rapidly expanding field. Structurally, we establish a robust taxonomy categorized by three core technical routes, which systematically evaluates the existing literature comprising 77 core research papers published between 2016 and 2026. The first technical route, temporal modeling methods, focuses on explicit sequence modeling architectures such as recurrent neural networks (RNNs), long short-term memory (LSTM) networks, and spatial-temporal Vision Transformers; these approaches are specifically engineered to capture chronological trajectories and historical momentum across discrete scan intervals. While these temporal architectures offer intuitive alignment and clear modeling pipelines for capturing time-variant trends, their capacity to synthesize complex, high-dimensional 3D/4D anatomical structures without structural fidelity collapse remains relatively constrained, often suffering from blurred generations or severe error accumulation during long-term autoregressive rollouts. The second and currently dominant methods, Deep generative models (constituting approximately 73% of the published literature), represents a massive paradigm shift toward probability distribution learning, encompassing generative adversarial networks (GANs), variational autoencoders (VAEs), and recent state-of-the-art continuous-time formulations including latent diffusion models and flow matching frameworks. By mapping high-dimensional image manifolds into well-behaved latent spaces, these deep generative approaches achieve unparalleled performance in synthesizing fine-grained diagnostic details, sharp structural boundaries, and high-frequency textural patterns, while offering exceptional conditional flexibility to incorporate arbitrary time-deltas or clinical metadata. The third technical route, Physics and Prior-Guided Models, bridges data-driven deep learning and domain-specific mechanistic knowledge by mathematically integrating biological growth laws, cell reaction-diffusion partial differential equations (PDEs), biomechanical hyperelastic deformation mechanics, neural ordinary differential equations (NeuralODEs), and physics-informed neural networks (PINNs) directly into the neural network optimization loop. By imposing hard or soft scientific constraints on the latent velocity fields or continuous implicit neural representations (INR), this paradigm demonstrates unique and unparalleled advantages in terms of clinical interpretability, biophysically compliant boundary preservation, and extreme data efficiency under low-resource clinical cohorts or rare disease scenarios where paired longitudinal sequences are exceptionally scarce. Beyond this core taxonomy, this survey provides a rigorous mathematical formalization of longitudinal image-to-image prediction tasks across varying temporal conditions and extensively catalogs the currently available open-source longitudinal imaging databases, such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) for neurodegeneration, alongside localized ophthalmology and oncology registries, while shedding light on the institutional silo effects that restrict data sharing. We critically analyze the deep-seated tension between traditional computer vision statistics (such as PSNR, SSIM, and FID) and true clinical utility, highlighting the dangerous gap between superficial visual realism and authentic anatomical or pathological truth. Finally, we provide an in-depth discussion on the open challenges and prospective future directions of the field, outlining the necessary transition toward continuous-time longitudinal foundation models. We emphasize the emerging frontier of multi-modal clinical data integration, where voxel arrays are jointly embedded with non-structured electronic health records (EHR) and micro-level genomics to capture holistic patient trajectories. Crucially, we project the evolution of these models from passive, natural-history simulations into active, counterfactual inference engines capable of simulating personalized "what-if" therapeutic scenarios, such as varying radiotherapy dosages or drug intervention frequencies, to directly assist in clinical decision-making. Ultimately, we hope this survey not only serves as an insightful methodological navigation for researchers entering this rapidly evolving field but also inspires collaborative efforts across the multidisciplinary community to drive tangible progress toward clinically impactful longitudinal generation.

Key words: longitudinal medical image; future image prediction; disease progression modeling; temporal models; generative models; physics-guided models

© 中国图象图形学报版权所有

论文引用格式：Li X H, Wang Y T, Zhang Y H, Niu S J and Chen Q. 2026. Generative prediction of longitudinal medical imaging: A survey. Journal of Image and Graphics, XXXX:XXXX-XXXX(李孝辉，汪语桐，张语涵，牛四杰，陈强. 2026. 纵向医学图像生成式预测综述. 中国图象图形学报，XXXX:XXXX-XXXX)[DOI: 10.11834/jig.260414]

0 引　言

医学影像在现代临床诊疗中扮演着不可替代的角色。从阿尔茨海默病(Alzheimer's disease, AD)的脑萎缩评估到黄斑变性的视网膜进展监测，从肺结节的生长跟踪到肿瘤放疗中的疗效评价，对同一患者在不同时间点采集的纵向图像序列为理解疾病的时间演化规律提供了影像学依据。纵向磁共振成像(magnetic resonance imaging, MRI)扫描可揭示AD患者进行性脑萎缩的进展轨迹(Puglisi等，2025；Litrico等，2024)；纵向光学相干层析成像(optical coherence tomography, OCT)成像能够捕获新生血管性年龄性黄斑变性(age-related macular degeneration, AMD)和糖尿病黄斑水肿(diabetic macular edema, DME)患者视网膜病理的演变过程，为抗血管内皮生长因子(anti-vascular endothelial growth factor, anti-VEGF)治疗方案的制定提供影像学依据(Zhang等，2023；Li等，2026c)；放疗期间的序列计算机断层扫描(computed tomography, CT)成像则有助于跟踪肿瘤反应，支持自适应治疗计划的调整(Liu等，2025b；Mantegna等，2026)。然而，临床实践中对纵向影像的分析仍主要依赖于医生的视觉比较和主观判断，难以实现精准、定量的进展预测。如何从已有的历史影像数据中学习疾病进展规律，进而生成患者未来的解剖结构图像，已成为医学图像分析领域亟待解决的重要问题。

传统的预测任务主要集中在分类(Li等，2021；Xu等，2022；Luo等，2025)和回归(Rohm等，2018；Xu等，2021)上，例如预测疾病分期标签或临床评分。这些方法虽能提供一定的决策支持，却无法呈现未来解剖结构的直观图像。近年来，随着生成模型技术的迅速发展，一种新的任务范式逐渐兴起：医学图像纵向预测生成。该任务的目标是基于同一患者的一张或多张历史医学图像，合成其未来的解剖结构图像，并可选择性融入治疗信息、临床指标等多模态数据。与分类和回归相比，图像生成能够提供更直观的疾病进展可视化，对临床治疗方案的制定具有更直接的参考价值。

从技术演进的脉络来看，纵向预测生成方法经历了多阶段的建模机制演变。早期研究主要借助反应扩散方程等物理模型描述病灶的生长规律(Zhu等，2017b；Zhang等，2018)，将疾病进展建模为偏微分方程(partial differential equation, PDE)控制的过程，可解释性强，但难以刻画复杂的非线性进展模式。随着深度学习技术的引入，循环神经网络(recurrent neural networks, RNN)、卷积长短期记忆网络(convolutional long short-term memory network, ConvLSTM)等递归架构被用于建模影像序列的时序依赖关系(Zhang等，2020)。这类方法通过递归连接捕捉相邻时间点之间的变化模式，在肿瘤生长预测和器官形变预测等任务中取得了良好效果，但递归结构的固有限制使其在长程依赖建模上表现不足。Transformer自注意力机制的引入突破了这一瓶颈，使得模型能够在任意时间点之间建立直接关联，显著提升了对不规则采样间隔和长时间跨度的建模能力(Ma等，2023；Wang等，2022)。与此同时，生成对抗网络(generative adversarial networks, GAN)，凭借其对抗训练机制大幅提升了生成图像的逼真度，在2020至2023年间成为纵向预测生成的主导技术路线，被广泛应用于脑肿瘤生长预测(Elazab等，2020)、眼科疾病进展预测(Li等，2026c；Pham等，2022)等任务。然而，对抗训练的不稳定性和时序约束机制的缺乏促使学界探索新的技术路线。自2023年起，扩散模型(diffusion models)凭借其稳定的训练过程和优越的生成质量迅速崛起(Yoon等，2023；Litrico等，2024)，其迭代去噪机制天然适合融入多层次的条件信号，已成为当前的主流技术方向。流匹配(flow matching)等连续时间生成方法进一步解决了扩散模型推理速度慢的问题(Disch等，2025；Chen等，2026)，代表了该领域的最新发展趋势。

在相邻研究领域中，现有的医学图像生成综述(Yi等，2019；Pan等，2019；Kazerouni等，2023)主要聚焦于静态图像合成、跨模态转换或数据增强，缺乏对时间轴上连续演化依赖、长程时空拓扑形变以及非均匀采样机制的系统讨论。尚未有工作将纵向预测与图像生成的交叉作为一个独立的研究方向进行

© 中国图象图形学报版权所有

系统梳理。不同研究在方法框架、时序建模策略、输入数据配置和评估标准上差异显著，使得研究者难以快速把握领域全貌和技术前沿。鉴于此，本文首次对医学图像纵向预测生成方向进行了系统性梳理与综述。

为确保本综述的系统性与完备性，本文制定了“多数据库布尔关键词初检”与“双向引文追踪”相结合的系统检索策略。核心数据库涵盖 Google Scholar、IEEE Xplore 以及 PubMed，检索年份跨度限定为 2016 年 1 月 1 日至 2026 年 6 月 30 日。检索策略由“时序随访模态”与“生成式预测任务”两组主题词通过布尔逻辑算子组合构建，设定的通用逻辑检索式为：("longitudinal medical image" OR "sequential medical images" OR "spatiotemporal evolution") AND ("prediction" OR "forecasting" OR "future image prediction" OR "future image generation" OR "progression modeling")，并针对不同检索平台进行了相应的语法字段适配。文献筛选流程由两位作者独立并行执行，包含跨库去重、题录/摘要初筛、全文精读复核以及双向引文追踪四个阶段；当两位作者在文献纳入判定上出现分歧时，交由通讯作者共同复审全文并达成裁决共识。在通过关键词初筛建立核心候选文献池后，本文进一步开展双向引文追踪：一方面对初选文献的参考文献列表进行反向追溯，另一方面利用学术索引工具对初选文献的后续被引记录进行正向追踪，以此最大程度捕获跨学科及最新发表的相关文献。

在全文精读复查阶段，本文制定了严格的操作性纳入与排除标准：1）纳入标准：必须聚焦于“图像到图像”或“序列到图像”的未来纵向预测生成任务，输出端须为二维/三维高分辨率解剖图像、局部病理体积或连续形态流形；模型须针对纵向时序依赖演化、潜空间连续采样路径、时空形变几何对齐或生物物理机制约束，提出了实质性的网络架构改良、微分方程动力学建模或新型目标损失函数设计；文献须发表于医学图像分析、模式识别及计算机视觉领域的同行评议期刊、权威国际会议，或 Arxiv 预印本。2）排除标准：仅输出疾病分期标签、生存概率、临床量表评分或病灶标量体积的一维非生成式分类与回归预测研究；缺乏针对时序演化或生成机制的实质性算法创新，仅将通用开源生成网络直接微调于私有数据的应用性文章；不含沿时间轴前瞻随访信息、仅进行基线静态跨模态影像风格迁移的研究。经过去重、多轮纳排核验以及双向引文追踪验证，本综述最终聚焦并深度纳入 77 篇核心文献展开系统性的技术解构与横向对比，覆盖脑神经退行、眼底病变、肺部病灶以及肿瘤与放疗等主要临床应用方向。

图 1 总结了该领域的总体框架，涵盖了典型临床应用场景、纵向预测生成任务定义，以及三类核心技术路线。本文的主要贡献包括：1）提出了一个三分类体系，将现有方法按核心建模机制划分为时序建模方法、深度生成模型和物理/先验引导模型，并明确了各建模机制的技术特征与优缺点；2）系统梳理了该领域常用的公开数据集和评估指标，分析了不同应用领域的方法偏好与技术演进脉络；3）揭示了该领域目前存在的不规则时间采样、纵向数据稀缺和标准化基准缺失等关键挑战，并对该领域的未来发展趋势和应用前景进行了深入分析和讨论。

1 分类体系与问题形式化

1.1 问题形式化定义

医学图像纵向预测生成的核心任务可形式化如下。给定在时间点 $\{t_1, t_2, \cdots, t_N\}$ 采集的一组基线图像 $\mathcal{X}=\{x_1, x_2, \cdots, x_N\}$（其中 $N \geqslant 1$），以及可选的协变量 c（如年龄、性别、治疗信息、临床生物标志物等），任务目标是在未来时间点 t_{N+1} 生成对应的医学图像 y_{N+1}。生成的图像 y_{N+1} 应近似真实条件分布 $p(x_{N+1}|\mathcal{X}, c)$。该形式化定义涵盖了该领域的核心要素：输入图像序列的数量（N）、协变量信息（c）以及预测的时间跨度（$\Delta t = t_{N+1} - t_N$）。其中 N 的取值直接决定了模型可用的时序信息量，Δt 则影响预测的不确定性程度。

现有方法中输入数据的配置在多个维度上有所不同。基于输入输出的时间结构，可将现有工作归纳为以下三种常见设定：

1）单时间点预测（$N = 1$）：使用单一基线图像预测单一未来图像。这是最简单的形式，常见于基于条件 GAN（Mirza 等，2014）的方法，如利用术前面部图像预测减压术后外观（Yoo 等，2020），或从基线眼底图像预测早期 AMD 的眼底变化（Pham 等，2022）。

2）多时间点预测（$N > 1$）：使用多张历史图像预测一张或多张未来图像。该设定编码了时间进展信

© 中国图象图形学报版权所有

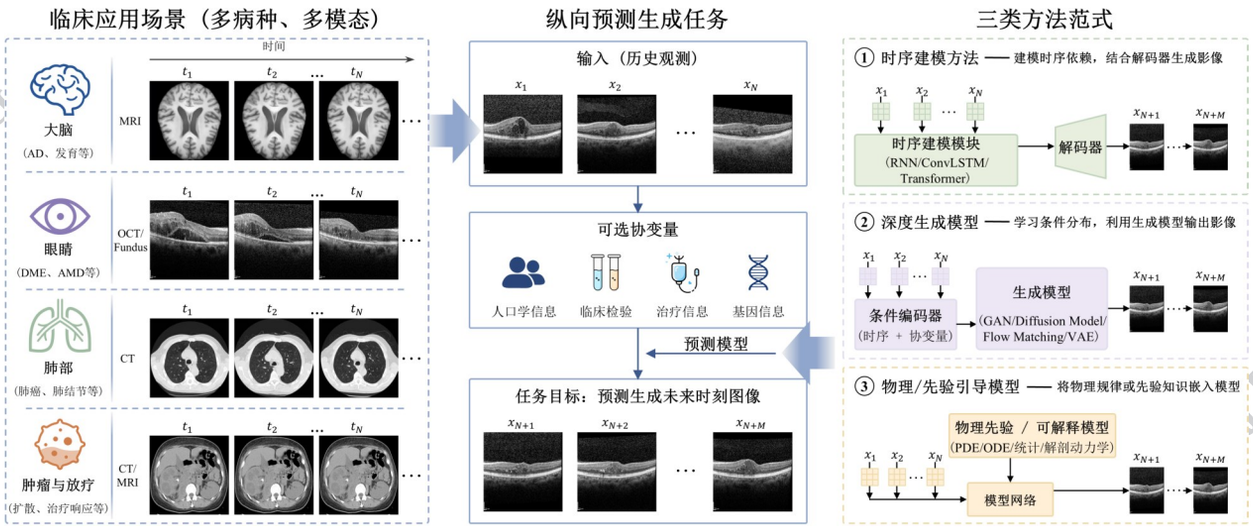

图 1　纵向医学图像预测生成的总体框架及研究机制概览

Fig. 1　Overview of longitudinal medical image prediction and generation: clinical applications, task formulation, and methodological routes

息，体现了纵向预测的核心特点。ConvLSTM 等方法通过递归处理连续时间点的影像序列以预测肿瘤生长（Zhang 等，2020），序列感知扩散模型（sequence-aware diffusion model，SADM）则将有序时间戳图像序列作为条件生成未来图像（Yoon 等，2023）。

3）连续时间预测：不局限于离散固定的时间步，而是构建可在任意未来物理时刻评估的连续病程演化轨迹。在此设定下，需要区分两类不同的连续性实现机制：（1）物理动力学连续积分：以神经常微分方程（neural ordinary differential equations，Neural ODE）为代表，直接在真实物理时间轴（t）上将疾病进展建模为连续演化系统，通过数值求解器在任意请求的随访时间点对解剖状态进行导数积分求值（Lachinov 等，2024）；（2）连续时间条件化生成：以扩散模型或流匹配为代表，此类方法主要通过将任意连续随访跨度 $\Delta t = t_{N+1} - t_N$ 作为显式标量嵌入注入去噪或速度场网络，从而实现对任意连续随访节点的生成控制。此外，部分前沿工作（如 Disch 等，2025）则通过差值流匹配结合稀疏填充策略，实现了对包含不规则时间间隔序列的统一端到端建模。

如图 2 所示，除了时间配置，输入数据在模态（单模态与多模态成像）、规整性（等间隔与非等间隔时间点）以及条件信息（仅图像、图像加临床协变量、是否包含治疗信息）方面也存在差异。近年来，处理不规则采样和多模态融合的趋势显著增强，反映了临床实际中成像安排难以严格规律的现实。

1.2　分类体系

纵向医学图像预测生成本质上是一个兼具时间序列建模、高维空间生成与生物机理合规的复杂综合任务。从底层逻辑剖析，现有研究的方法学构建实际上依托于三个正交的方法学维度：1）时序建模维度：描述网络如何在离散或连续轴上感知时序依赖、累积历史演变动量并处理不规则随访间隔（如基于 RNN、ConvLSTM 或 Transformer）；2）深度生成范式维度：描述网络如何学习从低维潜在表征或条件输入向高维像素/体素分布的映射机制，决定了影像的生成保真度与纹理细节（如基于 GAN、VAE、扩散模型或流匹配）；3）领域物理与生物先验维度：描述网络如何将机体生长定律、反应扩散偏微分方程、生物力学变形或临床统计图模型作为归纳偏置嵌入优化过程，提供可解释性与生物合规性约束。

在实际研究中，上述三个维度并非相互排斥的孤立技术，而是高度交织并可协同共存于同一复合模型中（例如将自注意力时序模块作为扩散模型的引导条件，并引入常微分方程正则项约束形态演化）。为了在保持多维视角的同时建立清晰、可索引的文献全貌，本文摒弃以“技术发表时间”或单一组件作为分类依据的做法，而是以文献所声明的核心方法学贡献与消融实验验证的重心作为基准归类依据：1）当文献的核心创新与消融对比集中于突破长程依赖、记忆瓶颈或时空特征解耦时，将其归入“时序建模方法”；2）当文献的理论推进主要聚焦于数据

© 中国图象图形学报版权所有

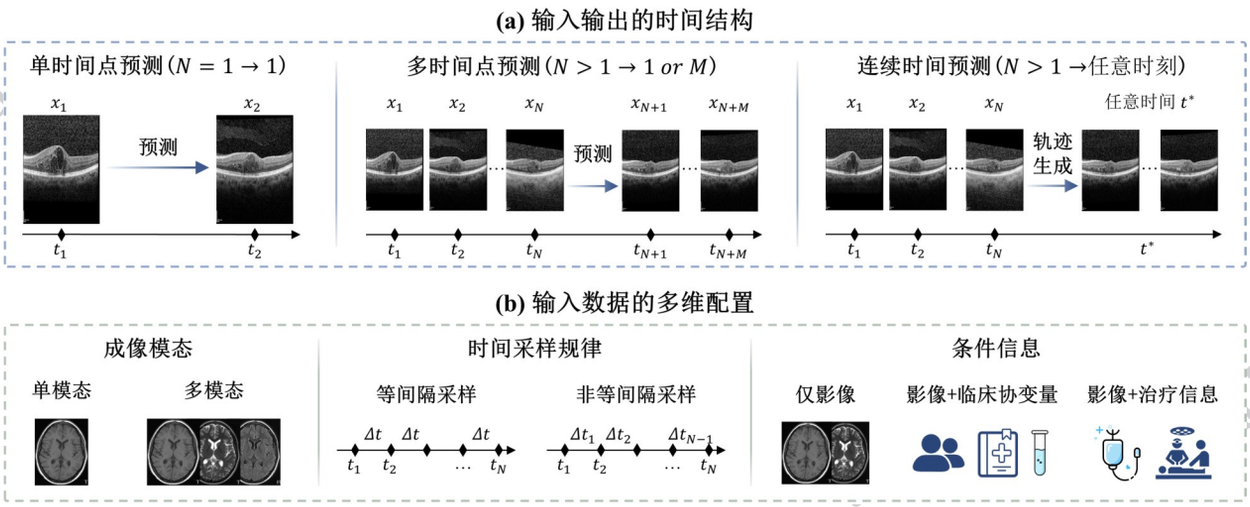

图2　纵向医学图像预测任务的典型设定与输入配置

Fig. 2　Typical prediction settings and input configurations for longitudinal medical image prediction ((a) temporal structure of inputs and outputs; (b) multi-dimensional configuration of input data)

分布拟合路径的重构、潜空间连续采样流设计或生成对抗/去噪保真度的提升时，将其归入“深度生成模型”；3）当文献的核心主旨在于通过数学解析方程、生物组织力学或概率图先验克服纯数据驱动的“黑盒”不可解释性与小样本泛化困境时，将其归入“物理/先验引导模型”。表1汇总了三大核心技术路线的技术特征与代表性文献，并在后续各节中深入探讨多维度技术在各自主导路线下的复合与演进趋势。

2　公开数据集与评估指标

2.1　公开可用数据集

纵向医学影像数据集的建设成本高昂且周期漫长，导致公开可用的纵向数据集在医学图像预测生成领域中极为稀缺。表2汇总了相关研究中涉及的公开数据集。鉴于多数数据集并非针对纵向医学影像任务专门采集，表中统计的患者数、随访数以及时间间隔等关键基准数据均提取自带有“__*”标记的参考文献；“-”则表示相关文献中缺乏确切的数据记录，或样本间的时间间隔存在较大异质性。

值得注意的是，相当数量的研究（尤其是眼科和放疗方向）依赖于与医疗机构合作的私有数据集，这也是当前制约该领域标准化评估和可重复性研究的主要瓶颈之一。Alzheimer's Disease Neuroimaging Initiative（ADNI）（Jack等，2008）是当前使用最为广泛的公开数据集，尤其集中于阿尔茨海默病的进展建模研究。

2.2　评估指标

该领域不存在单一的标准化基准，因为每项研究通常针对特定的疾病、成像模态和临床场景。主要指标的选择取决于具体任务。表3汇总了常用评估指标及代表来源文献。传统视觉指标与临床特异性指标在评估纵向预测生成任务时各有侧重。

以峰值信噪比（peak signal-to-noise ratio，PSNR）、结构相似性指数（structural similarity index，SSIM）、均方误差（mean square error，MSE）等为代表的像素级对齐指标具有坚实的像素级或体素级数学物理基础，能够稳定地约束生成影像的宏观解剖结构不发生扭曲或产生严重伪影，是保证合成质量的底线基石。然而，在纵向预测语境下，此类指标极易陷入“恒等映射作弊”的陷阱。由于患者在纵向随访期间的大部分解剖背景保持静态，模型即使不做任何进展外推、直接“拷贝”基线图像，也能凭借高比例的静态背景获取极高的评分；相反，若模型准确预测了局部病灶的演变趋势，但只要在空间上产生了微小的像素偏移，就会遭受严厉的空间惩罚，这导致像素级指标在实际评估中过度偏爱“保守、不作为”的模型。

而以Fréchet Inception距离（Fréchet inception distance，FID）和学习型感知图像块相似度（learned perceptual image patch similarity，LPIPS）为代表的分

© 中国图象图形学报版权所有

表1　三大核心技术路线及代表性文献

Table 1　Three core methodological routes and representative literature

类别	子类	核心思想	文献
时序建模方法	循环神经网络/长短期记忆网络	通过递归连接显式建模时空演化	Zhang等人(2020)、Wang等人(2020)、Li等人(2024)、Weng等人(2024)、Li等人(2026a)、Dao等人(2025)、Zhang等人(2021)
	Transformer/注意力	通过自注意力捕获长程时序依赖	Wang等人(2022)、Ma等人(2023)、Xiao等人(2026)、Gong等人(2025)
深度生成模型	对抗生成网络	对抗训练生成逼真图像，条件机制约束预测	Song等人(2019)、Elazab等人(2020)、Yoo等人(2020)、Ning等人(2020)、Peng等人(2021)、Pham等人(2022)、Huang等人(2023)、Zhang等人(2023)、Zhao等人(2024a)、Qiao等人(2025)、Luo等人(2026)、Li等人(2026c)、Sampathkumar等人(2026)、Wen等人(2026)
	扩散模型/流匹配	迭代去噪生成高质量图像；连续向量场实现可逆生成	Yoon等人(2023)、Litrico等人(2024)、Zhang等人(2024)、Puglisi等人(2024)、Zhao等人(2024b)、Gao等人(2024)、Liu等人(2025b)、Puglisi等人(2025)、Cho等人(2025)、Zhang等人(2025)、Jiang等人(2025)、Yang等人(2025)、Disch等人(2025)、Liu等人(2025a)、Wu等人(2025)、Lei等人(2025)、Favero等人(2025)、Laslo等人(2025)、Fan等人(2025)、Mouadden等人(2025)、Wang等人(2025)、Mantegna等人(2026)、Litrico等人(2026)、Song等人(2026)、Chen等人(2026)、Das等人(2026)、Li等人(2026b)
	变分/自编码器	潜在空间中建模疾病演化，提供不确定性量化	Ravi等人(2019)、Zhang等人(2019)、Xiao等人(2020)、Xia等人(2021)、Xiong等人(2022)、Sauty等人(2022)、Marti等人(2023)、Fang等人(2023)、Zhou等人(2023)、Yuan等人(2024)、Chen等人(2025)、Fang等人(2025)、Zhao等人(2025)、Dannecker等人(2025)、Tang等人(2026)
物理/先验引导	反应扩散/微分方程	偏微分方程或神经常微分方程建模病灶生长	Zhu等人(2017)、Zhang等人(2018)、Disch等人(2023)、Lachinov等人(2024)、Lachinov等人(2026)
	统计与混合模型	可解释的统计组件或概率模型学习进展模式	Niu等人(2016)、Amiri等人(2019)、Meng等人(2021)、Kim等人(2023)、Hong等人(2025)

布级特征指标能够跨越像素级严格对齐的限制，从高维特征空间精准捕捉预测影像的视觉逼真度与高频纹理细节，对条件扩散模型等高度非线性生成架构的合成质量具有极强的量化能力。但无法忽视的是，由于FID等属于静态且无序的集合度量，它们仅能评估某一特定未来时间步生成的静态数据是否真实，无法有效反映4D（3D+时间）形态流形演化的时序连续性。一个时序模型即便产生了解剖拓扑前后错乱、病灶动态忽大忽小等完全违背生物学规律的荒谬序列，只要单帧图像足够逼真，其FID依然可能表现完美。

为了克服纯计算机视觉指标与临床效用的系统性脱节，学界正越来越多地引入下游临床任务和动态生物标志物作为临床进展指标。这类指标通常利用预训练的分割网络提取预测病灶的Dice系数（Dice similarity coefficient）以评估关键病理语义的保留度，或直接计算生长率和病灶体积变化绝对值，从而绕开了不改变的静态解剖背景，专注于量化模型对“时间动量”和疾病演化斜率的捕获能力，因而最具临床推演的参考价值。不过，此类指标的计算方式相对复杂，且极度依赖第三方预训练工具在异质数据上的鲁棒性；更关键的是，由于不同疾病与成像模态的临床标志物（如脑萎缩体积与视网膜积液）完全不同，该类指标目前仍缺乏通用的标准化计算范式，导致不同研究之间难以进行跨疾病、跨模态的横向基准对比。

为克服既有评估方案过于原则化、评估指标极易被静态解剖背景稀释的缺陷，本综述建议未来纵向预测生成研究应采纳一套标准化的最低评估基准组合，涵盖以下五项核心维度：1）基线图像复制对照：将最后一次可观测图像（$\hat{x}_{N+1} = x_N$）作为基准对照，若模型各项评分无法显著超越该下限，表明其仅学习到了静态背景记忆而非真实的病理外推机制；2）病灶局域指标：将传统SSIM、PSNR及MSE的计算范围强制掩模化并聚焦于局部动态演化区域（如脑室扩张带、视网膜积液区及肿瘤浸润边缘），剔除

© 中国图象图形学报版权所有

表2　文献中使用的公开数据集

Table 2　Public datasets used in the literature

数据集	器官/疾病	数据模态	患者数	随访数	时间间隔	代表文献
ADNI (Jack等,2008)	脑/AD	MRI, PET, 临床	1369	5768	~12个月	Ravi等人(2019)、Ning等人(2020)、Xia等人(2021)、Disch等人(2023)、Marti等人(2023)、Lachinov等人(2024)、Puglisi等人(2025)、Yang等人(2025)、Cho等人(2025)、Dao等人(2025)*、Hong等人(2025)、Chen等人(2026)、Das等人(2026)
OASIS-3 (LaMontagne等,2019)	脑/老化与AD	MRI, PET, 临床	634	2535	~6个月	Litrico等人(2024)*、Puglisi等人(2025)、Cho等人(2025)、Chen等人(2026)
AIBL (Ellis等,2009)	脑/AD	MRI, PET, 临床	242	–	~18个月	Puglisi等人(2025)*、Chen等人(2026)
NACC (Beekly等,2007)	脑/AD	MRI, PET, 临床	962	2257	–	Puglisi等人(2025)*
BraTS-Seg (Baheti等,2021)	脑/脑肿瘤	多模态MRI	160	1760	–	Elazab等人(2020)、Zhou等人(2023)、Luo等人(2026)*
LUMIERE (Suter等,2022)	脑/细胞瘤	多模态MRI	37	132	–	Liu等人(2025b)*
IBSI (Hazlett等,2017)	脑/婴儿发育	多模态MRI	289	578	~6个月	Peng等人(2021)*
dHCP (Makropoulos等,2018)	脑/婴儿发育	多模态MRI	115	230	~3周	Dannecker等人(2025)*
OLIVES (Prabhushankar等,2022)	视网膜/DME	OCT, 临床	55	513	~4周	Li等人(2026b)*、Li等人(2026c)、Sampathkumar等人(2026)
SIGF (Li等,2020)	视网膜/青光眼	眼底照相	–	405	–	Zhang等人(2024)*、Fan等人(2025)
NLST (NLSTRT等,2011)	肺/肺癌	CT, X光	1065	3195	–	Xiao等人(2020)、Ma等人(2023)*、Fang等人(2023)、Lei等人(2025)、Xiao等人(2026)
MIMIC-CXR (Johnson等,2019)	胸/胸片	X光, 临床	–	–	–	Mouadden等人(2025)*

注:关键基准数据(患者数、随访数、时间间隔等)提取自标注“__*”的参考文献;“–”代表文献无明确记录或时间间隔异质性高;正子断层扫描(positron emission tomography, PET)。

不变解剖结构的评分稀释;3)疾病进展变化量误差(Δ – Error):重点量化预测变化量 $\Delta\hat{y} = \hat{x}_{N+1} - x_N$ 与真实变化量 $\Delta y = x_{N+1} - x_N$ 之间的差异,直接评估模型对病变‘速度与加速度’演化动量的外推精度;4)按预测时间跨度分层评估:杜绝将所有时间步混合计算均值的做法,必须按短程(如≤3个月)、中程(如6~12个月)及长程(如≥24个月)时间跨度(Δt)分层展示评估结果,以揭示误差随时间外推的非线性累积轨迹;5)多样本预测的不确定性统计校准:对输出分布特征的生成模型,不仅需展示多次随

© 中国图象图形学报版权所有

机采样的方差图，还必须通过预期校准误差（expected calibration error，ECE）或负对数似然（negative log-likelihood，NLL）等指标验证不确定性与真实预测误差的空间相关性，以确保置信度的临床可用性。

表3　常用评估指标

Table 3　Common evaluation metrics

类别	具体指标	说明	代表性文献
图像相似度	PSNR, SSIM, MSE, MAE	预测与真实图像之间的像素级相似度	Lei等人(2025)、Puglisi等人(2025)
特征相似度	FID, LPIPS	预测图像集与真实图像集的特征空间差异	Li等人(2026b)、Lei等人(2025)
临床效用	Dice系数，临床评分误差	预测图像的下游任务性能	Zhang等人(2020)、Liu等人(2025b)
疾病进展	生长率误差，体积误差	预测图像中疾病相关生物标志物的准确性	Li等人(2026c)、Zhang等人(2018)

3　基于时序建模方法的纵向预测生成

时序建模（temporal modeling）方法作为本综述分类体系中的第一类方法，共计纳入11篇论文，约占全部77篇综述文献的14%，详细对比统计如表4所示。该类方法的核心前提是：纵向预测的关键挑战在于对时序依赖关系的精确捕获，因此解决方案的核心应围绕设计专门用于序列数据建模的架构。与后文将讨论的深度生成模型不同，时序建模方法并非追求生成框架本身的创新，而是通过强化时序建模能力来提升预测质量，图像生成则通过将时序引擎与图像解码器相耦合来实现。在这一建模机制下，现有方法可进一步细分为两类技术路线：一是以ConvLSTM为代表的RNN方法，其通过递归连接显式建模时空演化过程；二是以自注意力机制为核心的Transformer/注意力方法，其通过建立任意时间点之间的直接关联来克服递归模型的记忆瓶颈。两类方法在时序建模的侧重点上各有不同，反映了该领域从递归建模向注意力机制演进的技术趋势。

3.1　基于RNN/LSTM的方法

循环神经网络，特别是LSTM（Hochreiter等，1997）和ConvLSTM（Shi等，2015）及其变体，为医学图像纵向预测提供了最为直观的时序建模框架。其核心思想在于通过递归连接模拟临床随访序列的时序特性，通用架构如图3所示：网络的隐藏状态在不同时间步之间持续传递，使模型能够逐步累积对疾病演化过程的内部表征，并基于累积的历史信息外推出未来的解剖状态。这一方向的发展脉络清晰地展示了从早期的概念验证，到面向多样化临床场景的专门化适配，再到解决临床真实世界中不规则采样和模态缺失等现实挑战的演进过程。

在技术发展初期，Zhang等人（2020）的工作为该类方法在纵向预测领域的应用奠定了重要基础。该研究提出了ST-ConvLSTM（Spatio-Temporal Conv-LSTM），将标准的ConvLSTM扩展至能够处理4D（3D+时间）纵向影像数据的架构。不同于此前主要依赖手工设计特征或传统数学模型的肿瘤生长预测方法，ST-ConvLSTM通过卷积递归单元同时编码三维空间特征和帧间时间转换，从而在单一网络中实现了对肿瘤静态形态和动态变化过程的联合学习。在4D CT纵向数据集上的实验表明，该方法在预测精度上显著优于传统线性模型、常规ConvLSTM和GAN基线，有力地论证了递归架构捕获临床进展动态的可行性。然而，该工作仅通过特征拼接引入时间间隔，未能摆脱离散递归机制对连续演化动力学的建模局限，且刚性的输入架构难以应对临床常见的模态缺失问题。

随着时空递归建模可行性的确立，ConvLSTM框架凭借其架构灵活性，很快被引入到自适应放疗等需要敏捷预测的临床场景中。Wang等人（2020）针对放疗引发的食管炎早期监测需求，提出了食管预测网络（esophagus prediction network，EP-net）。该方法采用两阶段级联架构：首先在patch级别利用CNN从每周MRI扫描中捕获食管对辐射的局部响应模式，随后通过门控循环单元（gated recurrent

© 中国图象图形学报版权所有

表4　时序建模方法文献统计

Table 4　Literature review on time series modeling methods

文献	任务	输入序列 N	不规则采样	多模态	数据规模	评估指标
Zhang等人(2020)	肿瘤生长预后	>1	是	是	33例	Dice,RVD,RMSE,diff. HU
Wang等人(2020)	食管空间变化预测	>1	否	是	11例	Dice,RMSSD,Precision,Jacobian maps
Zhang等人(2021)	GA生长区域预测	>1	是	否	55例	Dice,CC
Wang等人(2022)	肿瘤生长预测	>1	是	否	1325例	Dice,RMSE,Sensitivity,Specificity
Ma等人(2023)	肿瘤生长预测	>1	是	否	1065例	Dice,RMSE,Sensitivity
Li等人(2024)	视盘旁萎缩预测	>1	否	否	342例	Precision,Sensitivity,Accuracy,F1 score
Weng等人(2024)	腹部运动预测	>1	否	否	17例	SSIM,PSNR,MSE,NCC
Dao等人(2025)	AD进展预测	>1	是	是	1369例	PSNR,MSE,Precision,Sensitivity,Accuracy,AUC
Gong等人(2025)	肺部呼吸形变预测	>1	否	是	–	SSIM,PSNR
Li等人(2026a)	眼底图像预测	>1	是	否	584例	SSIM,PSNR,MSE,F1 Score
Xiao等人(2026)	肺肿瘤生长预测	>1	否	是	2800例	FID,Dice,RMSE,diff.HU,Sensitivity,Precision

注:"–"代表文献无明确记录;相对体积差异(relative volume difference, RVD),均方根误差(root mean square error, RMSE),亨氏单位差值(difference in hounsfield units, diff. HU),均方根表面距离(root mean square surface distance, RMSSD),雅可比映射(Jacobian maps),相关系数(correlation coefficient, CC),归一化相关系数(normalized correlation coefficient, NCC),曲线下面积(area under the curve, AUC)。

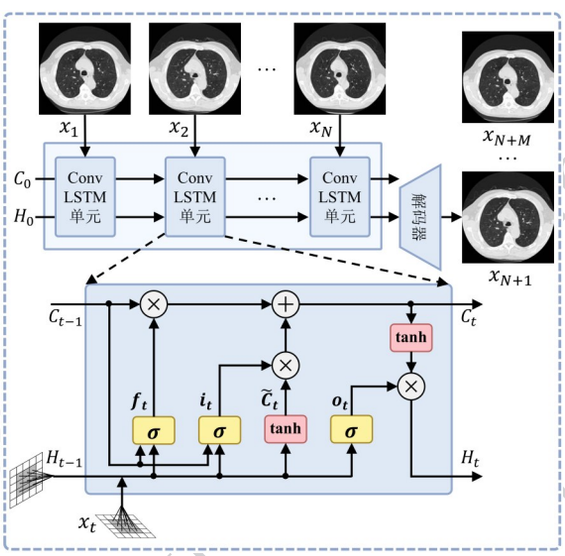

图3　基于RNN/ConvLSTM的通用架构示意图

Fig. 3　General architecture based on RNN/ConvLSTM

unit, GRU)编码-解码器解析这些特征的时间序列演化规律,最终重建出未来数周的患者特异性食管轮廓。实验表明,该模型不仅在食管形态预测上实现了较高的解剖保真度,且展现出一定的跨模态泛化能力,即MRI训练的模型可直接用于CBCT(cone-beam CT)数据。然而,该研究受限于小样本验证(仅11例患者),模型存在过拟合风险,且高度依赖手动勾画的初始轮廓。Weng等人(2024)则将递归预测推向了更具时效性的放疗场景:MR-Linac系统中的腹部运动实时补偿。该研究利用ConvLSTM构建了多对一架构,基于极短窗口的连续cine-MRI预测未来数秒内的腹部器官位置,并系统评估了该模型在不同腹部器官(胃、肝、胰腺和肾等)间的迁移能力。不过,由于其聚焦于瞬时的呼吸运动预测,其技术路线更侧重于实时运动补偿,而非面向长期疾病进展的宏观时序外推。

在需要长期随访追踪的眼科和神经退行性疾病领域,临床数据往往伴随着更为复杂的非理想采样问题,这促使研究者开始针对"不规则时间采样"和"数据缺失"进行架构层面更深度的改良。在眼科视盘旁萎缩的像素级进展预测任务中,Li等人(2024)通过设计时序记忆模块与跨层RNN状态传递机制,实现了历史随访信息的充分利用以及高层语义对底层细节特征的指导学习,并引入预定采样(scheduled-sampling)策略来缓解临床序列中的缺失数据问题。然而,该方法仍默认随访时间点为等间隔排列。为了打破这一理想化假设,该团队在后续

© 中国图象图形学报版权所有

工作(Li等,2026a)中进行了针对性演进,将ODE运动估计引入预测框架,将离散的不规则时间点映射到连续时间域,使模型高效适配任意间隔的随访安排;同时配合风格迁移模块消除了设备采集条件差异带来的非病理伪影。这一工作体现了从假设等间隔到适应任意不规则间隔的重要演进,但ODE模块的引入也显著提升了系统的计算复杂度。

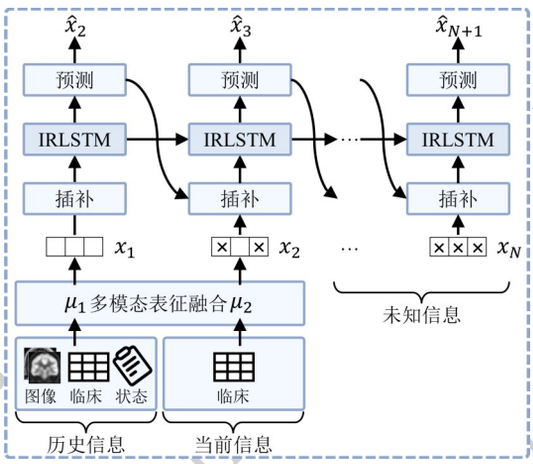

图4　LMDP-Net架构图(经许可修改自Dao等, 2025)

Fig. 4　Architecture of LMDP-Net (adapted from Dao et al., 2025 with permission)

与此同时,神经退行性疾病(如AD)的纵向预测则面临着"异构模态缺失"与"超长时序依赖"的双重挑战。Dao等人(2025)提出了面向多模态AD进展预测的递归框架(longitudinal multimodal disease progression prediction network, LMDP-Net),如图4所示:通过将MRI、PET及临床异构指标约束至一个公共表示空间,实现了模态缺失下的灵活融合;同时,作者针对RNN门控在长期序列中易出现的"全保留或全遗忘"信息流偏差设计了辅助控制函数,提升了模型在长跨度时间序列下的稳健性。针对同样的不规则采样痛点,Zhang等人(2021)在地图样萎缩(geographic atrophy, GA)进展预测中选择了另一种显式约束路径:他们将时间间隔因子作为条件特征直接集成到双向LSTM中,使网络能够感知相邻随访之间的实际时间差,并利用CNN精细化模块恢复空间细节。其系统性的消融实验设计,也为后续不规则采样研究提供了实验参考,但其时间因子的离散化编码方式,相比于连续时间的ODE外推仍存在一定局限。

优势与局限:RNN/LSTM方法提供了直观的时序建模能力和良好的可解释性,其递归结构天然契合纵向成像的序列特性。从最初基于ConvLSTM的肿瘤生长预测,到逐步扩展至放疗、眼科和神经退行性疾病等多个临床方向,这一技术路线在约五年的发展历程中展现出了良好的场景适应性。然而,该类方法也面临着三项固有的底层挑战:1)长程依赖捕获困难:递归计算路径拉长了时间反向传播的链条,梯度消失问题难以避免,限制了模型的有效时间感受野;2)并行训练效率低下:时间步上的串行依赖机制限制了图形处理单元的并行计算效率,使得大模型规模和海量数据拓扑下的训练吞吐量遭遇瓶颈;3)时间连续性适配不足:传统循环单元在缺乏显式动态适配或外部分支机制(如时间因子、ODE)的情况下,难以自然、优雅地处理现实临床中高度非均匀采样的离散时间点。这些固有的架构约束,共同促使该领域在近年开始转向探索具有更强并行计算能力和长程关联建模优势的变体及替代方案。

3.2　基于Transformer/注意力的方法

基于Transformer(Vaswani等,2017;Dosovitskiy等,2021;Liu Z等,2021)的架构为医学图像纵向预测提供了与RNN截然不同的时序建模思路。自注意力机制打破了RNN固有的信息压缩瓶颈:模型不再需要将全长历史信息强行编码为固定长度的隐藏状态,而是可以在序列内任意时间点之间直接建立动态注意力关联,从而有选择地加权并提取对当前预测最有价值的历史特征,其通用架构如图5所示。这一特性在处理长跨度、非均匀时间间隔的纵向随访序列时展现出独特的架构优势。

在Transformer引入该领域的早期探索中,Wang等人(2022)提出了SDC-Transformer(static-dynamic coordinated Transformer)。其核心思路是将纵向CT序列中的表征信息解耦为不随时间变化的静态解剖特征,以及由疾病驱动的动态进展特征,并通过专门设计的局部自适应模块与动态生长估计模块分别对其进行协同处理。这一"分离-协同"的显式解耦策略在肺肿瘤生长预测任务中验证了自注意力机制的有效性。Ma等人(2023)在技术路径上做出了重要推进,提出了LCTformer(coordinated longitudinal Transformer)。该工作并未盲目用Transformer完全替代RNN,而是将两者并行融合在统一框架内:由ConvLSTM捕获肿瘤的空间解剖位置和局部瞬时时序模式,而由Transformer负责捕获CT序列间的长程

© 中国图象图形学报版权所有

动态关联。在NLST（national lung screening trial）数据集上的实验结果表明，这种混合架构的性能优于单一架构，从实验层面证实了两个架构家族在时空建模中存在互补协同关系，而非简单的替代关系。

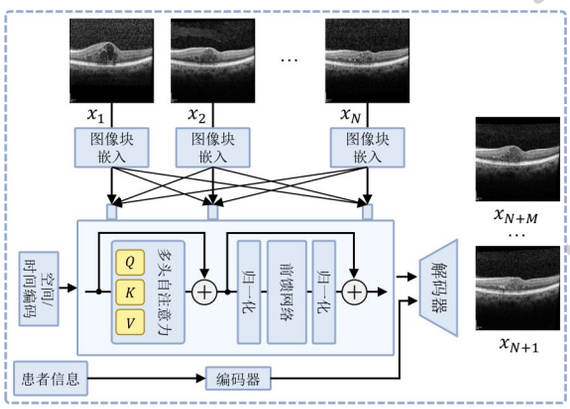

图5　基于Transformer/注意力的通用架构示意图

Fig. 5　General architecture based on Transformer/Attention

随着两类架构融合技术的成熟，研究者的视野开始从纯影像建模拓展至更多维度的临床异构信息。Gong等人（2025）将Transformer的应用场景拓展至4D CT呼吸运动影像的周期性预测生成。该研究提出的KSPNet（knowledge-guided dynamic-static feature synergistic prediction network）通过双向体素流注意力模块协同建模解剖结构与运动形变场，并引入光流引导的运动先验知识以增强时序校正能力。这一工作的学术启示在于：当外部物理或解剖领域知识（如运动速度场、方向场）能够被显式编码时，Transformer在周期性运动建模中的表征能力可以得到进一步增强。但受制于其2D切片级的处理方式，自回归预测策略中误差随时间步长累积的固有问题仍有待解决。Xiao等人（2026）则提出了ST-ConvTransformer（spatio-temporal convolutional transformer），如图6所示：通过共享自注意力的双流架构，同时编码影像的时空特征与患者的临床文本指标（如吸烟史等），并利用对比学习实现多模态语义对齐。该工作推进了Transformer方法在多模态临床信息融合维度的延伸。

优势与局限：Transformer方法通过自注意力机制赋予了时序纵向预测前所未有的架构灵活性，体现在任意时间步之间的直接关联能力、对非均匀采样间隔的高效适配，以及对异构多模态信息的灵活融合边界。从SDC-Transformer的静动态特征解耦，

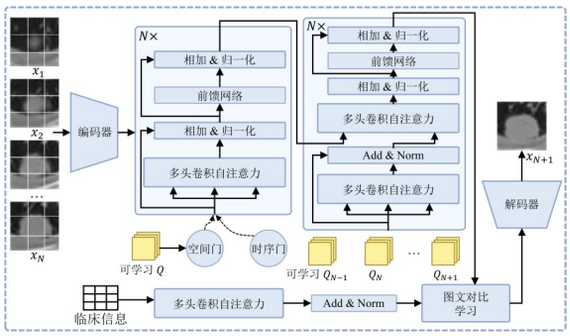

图6　ST-ConvTransformer架构图（经许可修改自Xiao等，2026）

Fig. 6　Architecture of ST-ConvTransformer (adapted from Xiao et al., 2026 with permission)

到LCTformer的递归-注意力协同，再到多模态融合与外部物理知识的引入，该技术路线在近几年的发展中呈现出从“单一架构创新”向“多维信息扩展”的清晰演进脉络。然而，该类建模机制在医学影像领域的深度落地仍面临两项固有约束：1）计算复杂度与高维数据的冲突：自注意力机制的计算与内存复杂度随序列长度和空间分辨率呈二次方增长，在直接处理高分辨率3D（甚至4D）医学图像体积时会带来极其严峻的计算开销；2）数据饥渴特性与纵向稀缺性的矛盾：Transformer通常缺乏强烈的归纳偏置，其优异的表征能力高度依赖大规模数据集，这与医学纵向序列数据获取成本高、样本量稀缺的现实形成了突出矛盾。这些瓶颈也深刻解释了为什么在当前的临床预测生成领域，纯Transformer架构并未完全取代RNN，而是更多地以“混合架构”或“协同建模”的形式与其他底层网络（如ConvLSTM、Diffusion等）互补存在。

4　基于深度生成模型的纵向预测生成

深度生成模型是本综述分类体系中规模最大的一部分，共纳入56篇论文，约占全部77篇综述文献的73%，详细对比统计如表5所示。与前一类型将时序架构设计作为核心关注点不同，深度生成模型将问题的重心置于"生成"本身：核心目标是学习从历史观测到未来图像的条件数据分布，而生成框架的选择，如GAN、Diffusion Models与Flow Matching、VAE，从根本上决定了时序信息被整合的方式以及最终预测图像的质量特征。在这一技术路线中，时序建模维度与物理先验维度通常退居为条件控制信

© 中国图象图形学报版权所有

号或隐空间正则项，而生成机制本身的分布学习效率与细节重建能力成为方法学的第一焦点。三类生成方法在纵向预测任务中呈现出清晰的演进脉络：GAN凭借对抗训练在2020至2023年间占据主导地位，扩散/流匹配模型自2023年起迅速崛起并成为当前最主流的技术路线，而VAE方法则在不确定性量化和小样本场景中保持着独特的应用价值。

表5　深度生成方法文献统计

Table 5　Literature review on deep generation methods

文献	任务	输入序列 N	不规则采样	多模态	数据规模	评估指标
Ravi等人(2019)	AD进展模拟	>1	否	否	1055例	SSIM，临床专家
Zhang等人(2019)	GA分割与生长预测	>1	是	否	29例	Dice，OR
Song等人(2019)	人脸老化/去龄化预测	1	是	否	82例	SSIM，PSNR
Xiao等人(2020)	肺肿瘤生长预测	>1	是	是	881例	Dice，RMSE，Precision，Sensitivity
Elazab等人(2020)	脑胶质瘤生长预测	>1	是	是	18例	Dice，RVD，RMSE，JI，Sensitivity，Specificity
Yoo等人(2020)	眼眶术后外观预测	1	–	否	109例	SSIM，MAE
Ning等人(2020)	AD进展预测	>1	是	否	1461例	SSIM，PSNR，MAE，RMSE
Peng等人(2021)	婴儿脑发育预测	1	否	是	289例	LPIPS，Dice，AVD，ASD
Xia等人(2021)	脑老化模拟	1	是	否	824例	SSIM，PSNR，MSE，PAD

注：**续表5 深度生成方法文献统计**

Continued Table 5 Literature review on deep generation methods

4.1　基于GAN的方法

生成对抗网络(Goodfellow等，2014)通过引入对抗训练机制能够合成高分辨率且感知逼真的图像，因而率先被引入到医学图像的纵向预测任务中，其一般架构如图7所示。条件GAN(Mirza等，2014)、CycleGAN(Zhu等，2017a)和StyleGAN(Karras等，2019)等模型的提出进一步拓展了其在不同场景下的适用性。然而，将GAN应用于纵向设定的核心挑战在于如何强制实施时序一致性约束，即确保预测的未来图像是历史序列信息的合理延续，而非仅在外观上逼真但缺乏时间连续性的独立静态图像。早期的方法通常通过将生成器条件化在单一基线图像上来实现这一目的，但该策略很快暴露了其时序动态建模能力上的局限，从而促使研究者从空间解耦、先验引导及多模态融合等多个方向探索改进路径。

在技术探索的初期，Elazab等人(2020)提出的GP-GAN(growth prediction GAN)是该类方法在纵向预测领域的早期尝试之一。该研究采用堆叠式3D GAN模块在BraTS-Seg(brain tumor segmentation)数据集上实现了由纵向MRI预测脑肿瘤生长的任务。每个GAN模块以当前时间点的T1 MRI图像及分割特征图(肿瘤边界、白质、灰质、脑脊液)为条件，预测下一时间点的肿瘤状态；多个GAN级联以模拟肿瘤随时间的渐进生长过程。几乎在同一时期，Yoo等人(2020)将条件GAN应用于眼眶减压手术后的面部外观预测这一不同的临床场景。这两项工作共同验证了GAN在纵向预测任务中的可行性，但也清晰地暴露了其固有局限：GP-GAN的级联架构虽实现了多时间步的递进预测，但逐模块的误差累积制约了长程预测的稳定性。而Yoo等的方法由于受限于极小样本量与低分辨率(64×64)输出，导致合成结果的临床真实度评估仍不尽如人意。

为了提升图像生成质量并增强预测结果的临床相关性，研究者开始引入领域先验知识，相关尝试首先在眼科影像领域取得进展。Pham等人(2022)提出的MuMo-GAN(multi-modal GAN)通过引入玻璃膜疣(drusen)分割掩码作为额外条件，引导生成器聚焦于早期AMD的关键病理特征，并设计了双判别器分别监督全局图像质量和局部病变演变的合理性。该工作证明了病理先验掩码能够有效嵌入GAN架

© 中国图象图形学报版权所有

文献	任务	输入序列 N	不规则采样	多模态	数据规模	评估指标
Xiong等人(2022)	乳腺癌骨溶解预测	>1	否	否	111例	SSIM,PSNR,LPIPS
Pham等人(2022)	AMD眼底图像预测	1	是	是	1263例	Dice,FID,Accuracy,Sensitivity,Specificity
Sauty等人(2022)	纵向神经影像进展	>1	是	否	–	–
Marti等人(2023)	AD进展预测	>1	是	是	897例	MAE
Huang等人(2023)	婴儿脑发育预测	1	是	否	30例	SSIM,PSNR
Zhang等人(2023)	nAMD治疗反应预测	1	否	否	22例	SSIM,PSNR,LPIPS
Yoon等人(2023)	纵向医学图像预测	>=1	是	否	3001例	SSIM,PSNR,RMSE
Fang等人(2023)	肺结节生长预测	1	是	否	1168例	SSIM,PSNR,MSE,Dice
Zhou等人(2023)	肿瘤复发位置预测	1	是	是	67例	Dice,HD,Sensitivity
Zhao等人(2024a)	nAMD治疗反应预测	1	否	是	186例	FID,SWD,临床专家
Litrico等人(2024)	AD进展预测	1	是	否	634例	SSIM,PSNR,MAE
Zhang等人(2024)	青光眼图像预测	>1	是	否	405例	Accuracy,Sensitivity,Specificity,AUC
Puglisi等人(2024)	AD进展预测	>=1	是	是	2805例	SSIM,MSE,MAE
Zhao等人(2024b)	青光眼图像预测	>1	是	是	264例	SSIM,PSNR,LPIPS,AMS,VCDR
Gao等人(2024)	乳腺癌治疗反应预测	1	是	是	434例	LPIPS,Sensitivity,Specificity,AUC,
Yuan等人(2024)	婴儿皮层发育预测	1	是	否	209例	MAE,PCC
Qiao等人(2025)	DME进展预测	1	否	是	4658例	SSIM,PSNR,FID,AUC
Liu等人(2025a)	疾病进展轨迹预测	>1	是	否	302例	SSIM,PSNR,MAE,MSE,Dice,HD
Puglisi等人(2025)	AD进展预测	>=1	是	是	2805例	SSIM,MSE,MAE
Cho等人(2025)	AD进展预测	>1	是	否	1741例	RMSE,WD,JSD
Zhang等人(2025)	婴儿脑发育预测	1	否	否	196例	SSIM,PSNR,Dice
Jiang等人(2025)	肺纤维化进展预测	>1	否	是	400例	SSIM,PSNR,RMSE
Yang等人(2025)	AD进展预测	1	是	是	–	SSIM,PSNR,MAE
Disch等人(2025)	4D医学图像预测	>1	是	否	373例	SSIM,PSNR,RMSE
Liu等人(2025b)	胶质瘤生长预测	>1	是	否	60例	SSIM,PSNR,MSE,Dice,RVD
Wu等人(2025)	肺癌LDCT图像预测	1	否	否	1226例	SSIM,LPIPS,FID,AUC
Lei等人(2025)	肺癌LDCT图像预测	>1	否	否	363例	SSIM,PSNR,LPIPS,FID
Chen等人(2025)	头颈癌放疗变化预测	>1	否	是	121例	SSIM,PSNR,MSE,Dice,ASD
Fang等人(2025)	婴儿脑发育预测	1	是	是	152例	SSIM,PSNR,Dice
Zhao等人(2025)	肺纤维化进展预测	1	是	否	317例	SSIM,PSNR,MSE
Favero等人(2025)	多发性硬化症病变掩码预测	>1	否	是	2131例	MSE,AUC,AP,Precision,Sensitivity,F1 Score

注:**续表5 深度生成方法文献统计**

Continued Table 5 Literature review on deep generation methods

构以提高预测的临床保真度。Zhao等人(2024a)则提出了生物标志物感知的非对称双分支GAN,通过并行编码器分别处理OCT影像与临床生物标志物(视网膜液体、色素上皮脱离、视网膜下高反射物质

© 中国图象图形学报版权所有

文献	任务	输入序列 N	不规则采样	多模态	数据规模	评估指标
Laslo等人(2025)	脑肿瘤生长预测	>1	是	是	37例	SSIM，RMSE，HD95
Fan等人(2025)	青光眼图像预测	>1	是	否	405例	SSIM，PSNR，LPIPS，MSE，FID
Luo等人(2026)	术后结果预测	1	否	否	572例	SSIM，PSNR，MAE，NMI
Li等人(2026c)	DME治疗反应预测	1	否	否	70例	SSIM，PSNR，LPIPS，MSE，FID，HAD，TC
Sampathkumar 等人(2026)	视网膜OCT图像与视功能纵向预测	>1	否	是	96例	SSIM，PSNR，FID，MAE，F1 Score
Wen等人(2026)	AMD治疗反应预测	1	否	是	3000例	SSIM，PSNR，MAE，Accuracy，AUC，F1 Score
Mantegna等人(2026)	肺癌放疗进展预测	>1	是	是	222例	SSIM，MSE，FID，VD
Litrico等人(2026)	AD进展预测	1	是	否	1596例	SSIM，MSE，MAE
Song等人(2026)	肺结节进展预测	>1	否	是	1226例	LPIPS，FID，AUC
Chen等人(2026)	AD进展预测	1	是	否	–	SSIM，PSNR，MAE，Δ-RMAE
Das等人(2026)	AD进展预测	1	是	否	952例	SSIM，PSNR，MSE，FID，MAE
Dannecker等人(2025)	围产期脑发育预测	1	是	是	115例	SSIM，PSNR，Dice
Mouadden等人(2025)	间质性肺病进展预测	>1	是	否	230例	Dice，VD
Wang等人(2025)	婴儿脑发育预测	1	否	是	284例	SSIM，PSNR，MSE，Dice
Li等人(2026b)	DME治疗反应预测	>1	否	否	33例	SSIM，PSNR，LPIPS，Dice，LVD
Tang等人(2026)	肺结节生长预测	>1	是	否	103例	SSIM，PSNR，MSE，Dice，Sensitivity，AUC，PPV

注：“–”代表文献无明确记录；重叠率(overlap ratio，OR)，杰卡德指数(Jaccard index，JI)，平均绝对误差(mean absolute error，MAE)，绝对体积差(absolute volume difference，AVD)，平均对称表面距离(average symmetric surface distance，ASD)，预测年龄差(predicted age difference，PAD)，豪斯多夫距离(Hausdorff distance，HD)，切片瓦瑟斯坦距离(sliced Wasserstein distance，SWD)，属性匹配得分(attribute matching score，AMS)，垂直杯盘比(vertical cup-to-disc ratio，VCDR)，皮尔逊相关系数(Pearson's correlation coefficient，PCC)，瓦瑟斯坦距离(wasserstein distance，WD)，詹森-香农散度(Jensen-Shannon divergence，JSD)，平均精度(average precision，AP)，95% 豪斯多夫距离(95% Hausdorff distance，HD95)，归一化互信息(normalized mutual information，NMI)，高度平均差(height average difference，HAD)，趋势预测一致性(trend prediction consistency，TC)，体积差异度(volume discrepancy，VD)，病灶体积差(lesion volume difference，LVD)，阳性预测值(positive predictive value，PPV)。

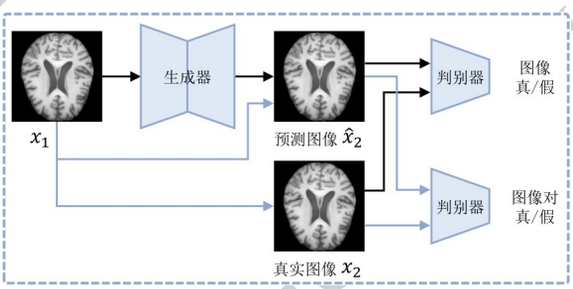

图7　基于GAN的通用架构示意图

Fig. 7　General architecture based on GAN

等)，并在融合前利用自适应记忆批归一化保留患者的个体特征。临床指标的显式融入显著提升了抗VEGF治疗响应预测的准确性。在提升图像生成质量上，Huang等人(2023)聚焦于婴幼儿脑MRI的跨时间点预测，针对出生后第一年内脑组织对比度快速翻转的特异性挑战，提出了MGAN(metamorphic GAN)架构。该模型引入空间-频率迁移模块，利用小波变换将特征解耦为低频组织对比度信息与高频解剖细节信息进行分别处理，并通过不确定性量化机制为预测结果提供可信度估计。在临床验证维度

© 中国图象图形学报版权所有

上，Qiao 等人（2025）取得了显著进展，其设计的 DRForecastGAN（diabetic retinopathy forecast GAN）在包含 2969 例训练和 1689 例外部测试的多中心数据集上进行了外部验证，并通过集成配准网络有效缓解了随访图像间的空间不对齐问题。

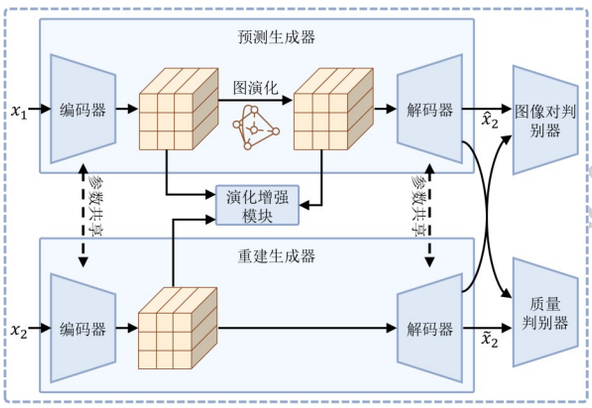

图 8　SHENet 架构图（经许可修改自 Zhang 等，2023）

Fig. 8　Architecture of SHENet (adapted from Zhang et al., 2023 with permission)

在 GAN 架构的演进方面，Zhang 等人（2023）提出的 SHENet（single-horizon disease evolution network）选择在高维潜在空间进行改进，通过图演化模块在解码前预测未来图像的深度特征，再恢复为视网膜 OCT 图像，如图 8 所示。这一潜在空间演化策略避开了直接生成高分辨率图像的训练难度。Sampathkumar 等人（2026）将 StyleGAN 架构引入纵向预测，利用其潜在空间的风格调制机制解耦高级解剖属性与像素级的随机变化，从而在跨时间点预测中自然地保持了视网膜结构的底层一致性。Luo 等人（2026）的 TraceTrans 则试图从几何形变的角度弥补 GAN 的不足：该模型同时预测空间变形场和图像转换结果，通过显式的体素位移约束确保生成图像与输入图像之间的物理对应关系，使模型在合成术后影像的同时能够追溯每个体素的变形来源，在一定程度上提升了深度网络的可解释性与临床可信度。进一步地，既往方法的“单时刻输入”架构设定导致模型无法感知疾病进展的个体化速度与加速度，Li 等人（2026）的 MT-DENet（multi-temporal disease evolution network）将输入从单时刻基线扩展到多时序序列，利用级联图演化模块从多次治疗前随访中提取患者特异性的长期病变演化趋势，并将治疗效果的先验知识融入网络硬性约束中。随着从单时刻向多时刻纵向序列的演进，GAN 架构得以更有效地应对复杂、非线性临床病程，为实现长程追踪与连续外推提供了更具扩展性的解决思路。

优势与局限：在 2020 至 2023 年间，GAN 是纵向预测生成领域最为主流的技术路线，在眼科、脑肿瘤和脑发育等多个方向上积累了丰富的方法学创新与临床验证。然而，从早期 GP-GAN 到后期 TraceTrans、MT-DENet 的演进脉络也清晰地揭示了 GAN 方法的固有边界：1）训练动力学不稳定：对抗博弈过程容易导致模式崩溃，在高维 3D/4D 医学图像生成中训练收敛极具挑战；2）时序一致性非内生：时序连贯性与解剖合理性往往需要通过外部辅助模块（如配准网络、形变场、风格解耦）进行显式硬性约束，而非对抗生成过程的天然属性；3）多时序序列拟合能力瓶颈：尽管部分后期工作引入了多时刻输入，但由于 GAN 本质上是在像素或静态潜空间上进行条件映射，其底层架构在处理高度非均匀采样、长程复杂的时空序列依赖时，往往面临显著的能力瓶颈。

4.2　基于扩散模型与 Flow Matching 的方法

扩散模型（Ho 等，2020；Song 等，2021；Peebles 等，2023）自 2023 年起迅速崛起，已成为纵向预测生成领域最主流的技术路线。潜空间扩散模型（latent diffusion model，LDM）（Rombach 等，2022）通过在潜空间中执行扩散过程显著提升了高维图像合成的效率。其在纵向生成任务中的应用，主要得益于其三项底层的架构与数学特性：首先，其渐进式迭代去噪机制展现出良好的灵活性，能够支持多尺度复杂条件控制信号的融入；其次，其分布拟合机制在一定程度上缓解了对抗博弈所带来的训练不稳定与模式崩溃问题，从而有利于高维影像合成边界的拓展；此外，其在数学上建立在常微分方程或随机微分方程的连续概率流框架之上。需要指出的是，标准流匹配（Lipman 等，2023）与校正流（Liu 等，2022）中的速度场积分时间 $\tau \in [0, 1]$ 描述的是从先验分布向目标图像分布的确定性传输路径，而非患者疾病演化的物理时间轴。为了实现纵向病程的连续推演，现有研究通常采取两种机制：一是将连续物理时间跨度 Δt 作为显式外在条件，指导内层生成速度场 $v_\theta(x_\tau, \tau;\ \Delta t)$ 的定向去噪积分；二是通过轨迹流匹配（trajectory flow matching，TFM）或潜空间演化场，将速度流直接参数化在真实的病程发展轴上，突破了

© 中国图象图形学报版权所有

传统离散预测框架对固定采样步长依赖的局限。基础架构如图9所示。纵观该类建模机制的技术演进脉络，可以发现其在方法学上展现出了清晰的演进趋势：即从序列条件扩散逐步向潜空间表征拓展、从纯影像驱动向异构多模态融合延伸，并实现了从离散时间步预测向连续时空轨迹建模的跨越。

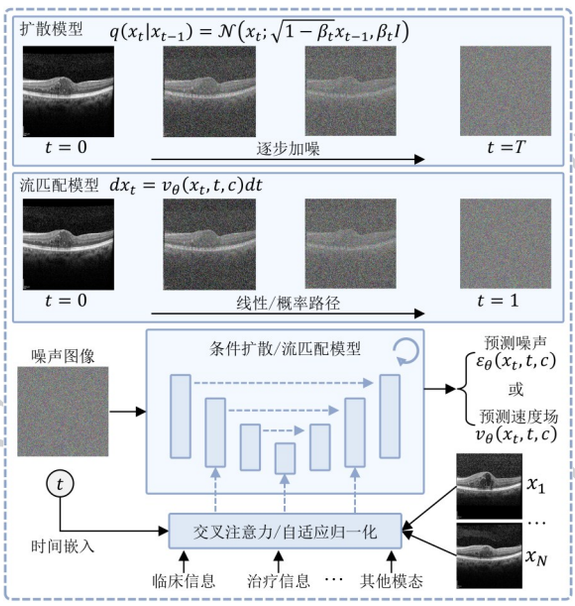

图9　基于扩散模型和流匹配模型的通用架构示意图

Fig. 9　General architecture based on diffusion model and flow matching

在技术发展的开端，Yoon等人(2023)提出的SADM奠定了基础。该工作首次将时序编码器与条件扩散模型相结合，其核心设计在于利用基于ViViT(video vision transformer)的时序网络对多张历史3D影像体积进行时空自注意力编码，进而生成全局条件信号以引导扩散模型合成未来的4D时空张量。这一工作有力地推动了序列条件扩散基本架构的建立与普及，并支持自回归采样与缺失帧的鲁棒处理；然而，该方法在一定程度上仍缺乏对患者个体特异性临床元数据的条件化整合，且推理过程也受限于扩散模型经典去噪框架所固有的慢采样瓶颈。为了降低高维直接生成的复杂度，Litrico等人(2024)在TADM(temporally-aware diffusion model)中引入了截然不同的改进思路：模型不再直接合成完整的未来图像，而是聚焦于预测基线与随访影像之间的“强度残差图”，并以年龄差作为条件，有效缓解了模型对年龄均衡数据集的依赖。该团队在后续的改进工作(Litrico等，2026)中进一步深化了这一路线，通过引入双向时间正则化强制前向与反向预测之间的一致性，利用循环一致性约束显著缓解了长跨度预测中的误差累积。

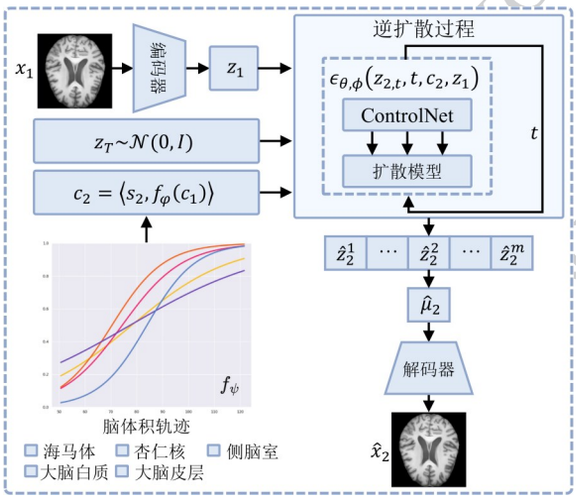

图10　BrLP架构图(经许可修改自Puglisi等，2024)

Fig. 10　Architecture of BrLP (adapted from Puglisi et al., 2024 with permission)

由于3D/4D高分辨率医学影像的底层数据体积巨大，直接在像素空间进行长时间步的扩散去噪会带来不可承受的计算开销，这促使该领域向LDM发生重大转折。Puglisi等人(2024，2025)提出的BrLP(brain latent progression)代表了这一时期的里程碑工作。其方法架构如图10所示：该模型在预训练自编码器的低维潜在空间中运行，结合ControlNet以保持个体特异性的解剖结构，并巧妙地引入辅助机制将AD相关脑区体积变化的生物学先验显式转化为扩散条件。针对多时间步预测中时空一致性劣化问题，该工作提出的潜空间平均稳定化技术，在包含万余张MRI的大规模多中心队列上验证了显著的保真度提升。然而，BrLP高达553M的参数量以及辅助模型误差级联传播的风险仍是其隐忧。为了进一步提升潜空间预测的鲁棒性，Zhang等人(2024)在C2F-LDM中设计了Coarse-to-Fine的双阶段潜扩散策略，其时序自适应编码器不仅能灵活处理眼底图像的不规则访视间隔，还在潜空间内同步实现了青光眼分类概率与眼底影像可视化的双重输出。与此同时，Das等人(2026)的Align-cDAE(attention-aligned conditional diffusion auto-encoder)则深入到了交叉注意力机制的改良上，通过在潜空间中引入显式对齐损失，强制扩散网络聚焦于脑室扩大等与

© 中国图象图形学报版权所有

疾病进展强相关的局部区域，并实现了进展特征与解剖身份特征在子空间中的深度解耦。

随着潜空间生成架构的日趋成熟，如何跨越单一影像模态、引入更丰富的临床多模态异构信号以贴合真实临床决策，成为近年的研究热点。Liu等人（2025b）率先在脑胶质瘤生长预测中实现了“治疗感知”的扩散机制，将放射剂量、化疗方案等非影像治疗信息作为显式控制条件融入扩散模型去噪过程，联合分割与生成网络同步输出肿瘤掩码和多参数MRI，从而赋予了模型评估反事实医疗决策（即“若采取不同治疗方案，肿瘤将如何演变”）的前瞻性推演能力。Gao等人（2024）则延伸至文本模态，其LIMRA（longitudinal image-report alignment）模型在潜空间中对齐不同时间点的放射学报告文本嵌入与乳腺X线特征，利用报告中的深层语义指导乳腺癌新辅助治疗反应的治中影像预测。在此基础上，Song等人（2026）的NAMD（nodule-aligned multi-modal diffusion）模型进一步将大语言模型（large language model，LLM）MedGemma与扩散模型深度整合，通过LLM将复杂的非结构化电子健康档案（electronic health record，EHR）转化为语义条件流，配合结节对齐的临床属性潜空间，在肺结节一年后进展预测任务上实现了逼近真实随访影像的诊断性能。

在实际临床场景中，纵向序列往往充斥着不规则采样、缺失随访以及配对数据极端稀缺等异质性痛点。为了解决这些制约临床落地的现实问题，研究者们从多个维度推进了扩散模型的场景适配能力。在处理非均匀采样与缺失访视方面，Fan等人（2025）提出的tHPM-LDM（continuous-time historical and population memory-conditioned LDM）引入了基于Sinkhorn-Knopp聚类的群体记忆模块，通过从训练数据中检索相似典型的进展模式，为缺失特定访视节点的个体个体化预测提供群体级先验参考。Mouadden等人（2025）的CLDM（conditional LDM）则凭借时序Transformer的灵活设计，构建了能够同时兼容2D与3D纵向影像的高通用性潜扩散框架。在克服数据稀缺瓶颈方面，Zhang等人（2025）的DDM（development-driven diffusion model）模型针对胎儿脑随访数据极难获取的现实，创新性地从未配对的临床横截面胎儿MRI中构造伪配对数据，成功训练出跨孕周的相邻发育轨迹预测模型。为了在保持连续轨迹的同时恢复空间高频细节，Li等人（2026b）在视网膜病变预测中构筑了两阶段级联框架，先利用隐式神经表示（implicit neural representations，INR）将连续时空坐标映射为微观体素强度，再通过对抗条件扩散精细化INR生成的过平滑纹理，但两阶段级联的设计也带来了较高的训练复杂度。此外，Laslo等人（2025）则展示了物理机制与数据驱动融合的潜力，将描述肿瘤生长动力学和放疗细胞死亡效应的ODE模型与梯度引导扩散模型相结合，在样本极度受限（仅37例患者）的罕见儿童脑干胶质瘤预测上实现了生物学机理对深度生成的纠偏。

近期，针对扩散模型固有的计算复杂度高、采样速度慢等瓶颈，生成架构本身迎来了新一轮的变革。Yang等人（2025）推出的MambaControl，将具有线性计算复杂度的Mamba状态空间模型（Gu等，2024）引入扩散框架以替代传统UNet（Ronneberger等，2015），在长序列纵向预测中，仅用215.5M的精简参数量便在时序关联与解剖控制上超越了拥有553M参数的BrLP模型，为高维长程纵向建模开辟了低能耗的新路径。而流匹配与校正流（rectified flow）的引入，则标志着该领域正式从离散去噪走向了高效的连续时间轨迹建模。Wang等人（2025）的FDRF（flexibly distilled 3D rectified flow）在3D矫正流框架下实施柔性知识蒸馏，双流联合预测婴幼儿脑MRI与组织分割图，并通过单步蒸馏技术实现了0.08秒/例的极速推理。Disch等人（2025）提出的TFM则通过“差异建模”将流匹配的构建空间从复杂的完整图像分布聚焦于微观的变化分布，结合稀疏填充策略，在心脏、脑卒中及脑肿瘤三个异构基准上超越了现有扩散基线。为了解决传统自编码器潜在空间语义无序的问题，Liu等人（2025a）的ImageFlowNet首先学习患者和时间点的多尺度联合表示空间，然后在位置参数化的共享空间中学习疾病流场轨迹，支持确定性与随机性的双模态预测；Chen等人（2026）的Δ-LFM则通过学习患者特异性的潜空间对齐，强制患者轨迹沿疾病严重度轴单调排列，使流匹配能够在结构化的流形中生成更具解剖连贯性的轨迹。此外，Wu等人（2025）则将相关性自编码器与潜空间流匹配深度融合，显式挖掘基线与随访之间的特征交织，成功从基线低剂量CT中合成了用于早期肺癌诊断的一年期虚拟随访图。

优势与局限：扩散模型与流匹配在短短三年内

© 中国图象图形学报版权所有

实现了爆发式增长，并迅速确立了其在纵向预测生成领域的地位。其核心优势在于极其优异的条件化边界与高保真的图像合成质量。从早期的序列条件扩散进化到潜空间先验深度集成，再到如今的连续时间流匹配，其技术迭代速度与对复杂异构数据的处理弹性优于同期的GAN与时序模型。然而，此类方法走向临床部署仍面临不可忽视的底层瓶颈：1）即便流匹配与蒸馏技术已将采样步数压缩至十步甚至单步级别，但对于需要高频实时响应的临床交互场景，其计算开销依然偏高；2）高维潜空间模型（如BrLP等）庞大的参数量对边缘医疗设备的显存不友好；3）其无偏置的强拟合能力依然极度依赖大规模多中心纵向数据的训练。目前，这些瓶颈正通过轻量化状态空间模型、位置参数化流场以及跨时空差异建模等前沿路径被逐步缓解。

4.3 基于VAE/AE的方法

尽管GAN与扩散模型在图像合成的视觉保真度上占据了领先地位，但基于VAE（Kingma等，2014）和向量量化VAE（vector quantized VAE，VQ-VAE）（Van den Oord等，2017）及自动编码器（auto-encoder，AE）的方法在纵向预测生成领域仍然维系着独特的学术生态位，如图11所示：其一，概率潜空间的显式构造为病程演化中的不确定性建模提供了原则性的理论框架，能够显式解耦并表征病理进展的随机扰动；其二，紧凑而结构化的潜在流形天然适合嵌入解剖先验与生物生长动力学约束；其三，编解码架构计算开销较低且具备良好的表征压缩能力，常作为两阶段生成模型或多任务协同的轻量化基座。总体而言，基于VAE/AE的方法在该领域的发展脉络，呈现出一条从基础重构框架到嵌入物理/生物学先验、从单一任务映射到多任务协同，并最终与隐式神经表示及流匹配等前沿范式深度融合的演进轨迹。

在早期探索阶段，研究者们主要聚焦于验证紧凑的潜空间表征在时序外推任务中的可行性，并倾向于将预测与传统的下游临床任务相整合。Zhang等人（2019）率先构建了多尺度CNN，在同一网络框架内联合进行GA病灶的当前分割与未来位置预测。该工作表明分割与预测任务共享特征提取器能够实现多任务的相互增强，但其缺乏专门的时序建模机制，导致长期依赖捕获能力存在先天不足。为了引入更稳健的时序拟合能力，Xiong等人（2022）提出的T-VAE（temporal VAE）将VAE的分布拟合特长与LSTM的时序递推能力深度耦合。如图12所示，该模型利用先验估计器从历史序列中学习骨溶解进展的潜在分布，并通过未来预测器进行随机采样，从而赋予了预测结果以原则性的多样性。其设计的边缘感知损失有效克服了CT影像中前景稀疏带来的类不平衡挑战，输出的不确定性图更是直观地为临床提供了预测置信度的空间分布。然而，T-VAE在向3D高维空间扩展时暴露出了明显的过拟合风险，且当病变在预测节点突发（即前三周无任何前驱征兆）时，模型的预测精度会显著退化。

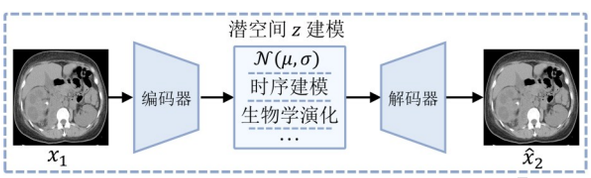

图11　基于VAE/AE的通用架构示意图

Fig. 11　General architecture based on VAE and AE

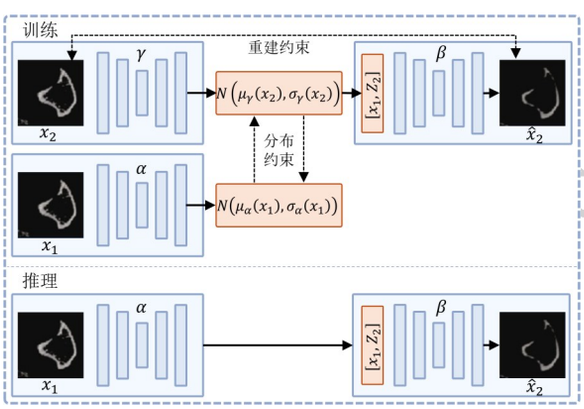

图12　T-VAE架构图（经许可修改自Xiong等，2022）

Fig. 12　Architecture of T-VAE (adapted from Xiong et al., 2022 with permission)

随着基础重构框架的局限性逐步显现，自2023年起，研究者们开始摒弃纯数据驱动的盲目映射，转而探索将确定性的生物学演化规律或临床迁移知识作为强约束硬性嵌入到自编码器的潜空间中。Fang等人（2023）做出了代表性尝试，他们将经典的肿瘤生长数学模型，Gompertz生长曲线直接整合至AE的潜在流形中，强制约束潜变量遵循“早期指数增长、后期趋于饱和”的生物合理轨迹。这种数据驱动与物理机制相融合的“灰盒”思路在肺结节数据集上展现出优异的解剖学合理性，但Gompertz固定的数学

© 中国图象图形学报版权所有

公式在面对某些高度异质性的非典型结节生长模式时仍显得缺乏弹性。与此同时，针对小样本临床数据的预测难题，Zhou等人（2023）则选择引入临床迁移先验。该团队在BraTS-Seg数据集上预训练多模态分割网络以提取鲁棒的解剖特征，结合多尺度空间/模态注意力校准与非线性相关学习，在仅包含67例患者的小样本队列中实现了脑肿瘤术后复发位置的联合分割与外推预测，为个性化放疗规划提供了具有参考价值的解剖学轮廓线索。

进一步的，纵向预测的研究视野从单一病灶的局部演变逐渐拓宽至人体器官的宏观长程发育（如婴幼儿脑发育），这使得如何保持患者个体身份一致性并实现连续时间查询成为了新的技术焦点。Yuan等人（2024）针对婴幼儿皮层形态学发育预测中“不同历史输入推演同一目标时点导致结果冲突”的轨迹不一致痛点，提出了三元组解耦AE。该模型通过空间注意力将潜空间特征显式分离为“身份相关”与“年龄相关”两个独立子空间，并协同应用三元组损失与动态时间规整损失，首次在保障个体特异性的同时实现了跨多时间点输入的轨迹单调一致性。随后，Fang等人（2025）面对婴幼儿脑MRI随访数据中更为复杂的非线性演变，提出了更为精细的三阶段年龄感知生成框架。该方法将复杂的发育过程进行解剖学降维，依次建模为脑体积扩张（全局仿射）、皮层折叠（局部形变场预测）以及髓鞘化对比度演变（对比合成）三个递进阶段，并以时间注意力模块实现了对0至24个月任意月龄的定向控制生成。不过，这两项工作目前仍主要局限于皮层表面属性图或单模态影像，未能充分挖掘多模态序列的联合诊断价值。

近期，基于AE的纵向预测方法呈现出与其他前沿生成方法深度融合与跨界交织的趋势。为了在离散潜空间中实现优雅的连续时间建模，Zhao等人（2025）将VQ-AE与NeuralODE结合，提出了4D VQ-GAN架构，在离散的向量量化码本上利用NeuralODE成功捕获了特发性肺纤维化的连续时序动态，支持任意采样随访点的CT体积重建。为了缓解现代深层生成网络对高算力的需求，Tang等人（2026）提出的轻量化NGP-net（nodule growth prediction network）则通过极限精简AE架构，以远低于扩散模型替代方案的计算开销和极速的推理响应，在肺结节生长预测任务上证明了轻量化边缘端部署的临床可行性。此外，Dannecker等人（2025）尝试将INR引入脑发育时序预测，通过神经网络将连续的时空坐标映射为微观体素强度；这种摆脱离散时间步限制、支持全连续时间自由查询的技术路径，展现出良好的发展前景，也为自编码器技术路线的演进提供了具有参考价值的前沿方向。

优势与局限：AE与VAE方法的技术轨迹清晰地揭示了其与现代扩散机制相互交融的发展趋势。其概率潜空间为纵向预测提供了显式的不确定性建模途径。然而，在方法学与临床应用层面，生成样本的多样性与经过统计校准的预测置信度存在明确的概念界限。部分研究主要通过潜在空间的多重随机采样呈现差异化的预测序列，这在本质上更多反映了模型解空间的分布覆盖度；而在临床决策支持场景下，引入预期校准误差、负对数似然或方差-残差相关性等评估机制，是将多样性采样转化为可靠置信边界的关键保障。一般而言，当模型输出的方差分布与解剖变异程度或病理演变前沿表现出经过验证的统计关联时，不确定性估计方能为临床风险评估提供可解释的参考依据。在当前的方法学格局中，编解码器架构越来越多地作为两阶段LDM的底层第一阶段运行，完美实现了“前级负责表示效率（VAE/AE），后级负责高保真生成（扩散/流匹配）”的黄金组合。随着纯粹的自编码预测工作在近两年显著向混合架构过渡，VAE曾经独占的概率建模与不确定性量化优势也逐步被现代潜空间连续生成流形所吸收。但其在轻量化临床部署、多任务协同（联合分割与预测）的生态位上，依然保留着不可替代的方法学价值。

5 基于物理/先验引导的纵向预测生成

物理和先验引导模型是本文分类体系中的第三个核心建模机制，共计纳入10篇论文，约占全部77篇综述文献的13%，详细对比统计如表6所示。与前述将时序架构设计或对抗/扩散生成机制作为核心驱动力的纯数据驱动路线不同，这类方法在建模逻辑上进行了重要调整：它们不再试图完全从影像数据中黑盒式地学习映射分布，而是尝试将医学、生物学及临床实践中的领域知识（如物理定律、解剖动力学、ODE/PDE生长模型）作为显式约束或底层机

© 中国图象图形学报版权所有

表6　物理/先验方法文献统计

Table 6　Literature review on physics and priors methods

文献	任务	输入序列N	不规则采样	多模态	数据规模	评估指标
Niu等人(2016)	GA生长区域预测	>1	是	否	29例	Dice,Sensitivity,Specificity,PPV,NPV,CC
Zhu等人(2017)	CNV生长预测	>1	否	否	7例	Dice,TPVF,FPVF
Zhang等人(2018)	胰腺肿瘤生长预测	>1	是	是	11例	Dice,Sensitivity,Precision,RVD
Amiri等人(2019)	脑肿瘤生长预测	>1	否	是	2例	Dice
Meng等人(2021)	CNV生长预测	>1	否	否	6例	Dice,RVD,TPR,FPR
Disch等人(2023)	AD进展轨迹建模	>1	是	否	–	SSIM,MSE
Kim等人(2023)	腹主动脉瘤生长预测	>1	是	是	54例	RMSE,RE
Lachinov等人(2024)	GA进展预测	1	是	否	817例	Dice,HD95,MAE
Hong等人(2025)	AD进展预测	>1	是	否	1785例	SSIM,PSNR,MSE,Accuracy,Precision,F1 score
Lachinov等人(2026)	GA生长区域预测	>1	是	否	75例	Dice,MSE,RE,F1 score,Sensitivity,Specificity

注:“–”代表文献无明确记录;阴性预测值(negative predictive value, NPV),真阳性体积占比(true positive volume fraction, TPVF),假阳性体积占比(false positive volume fraction, FPVF),真阳性率(true positive rate, TPR),假阳性率(false positive rate, FPR),相对误差(relative error, RE)。

制融入到网络中,其通用架构如图13所示。这类方法的主要优势在于,有望提升模型的临床可解释性与小样本利用效率;而其局限性则在于,对高度异质且复杂的生物学进程可能存在过度简化假设的风险。在具体实现上,目前主要聚焦于两条技术路线:一是基于反应扩散方程与神经常微分方程的方法,尝试将疾病进展建模为受PDE或ODE控制的物理演化过程;二是基于统计与混合模型的方法,利用可解释的概率组件或统计框架来捕捉进展模式。

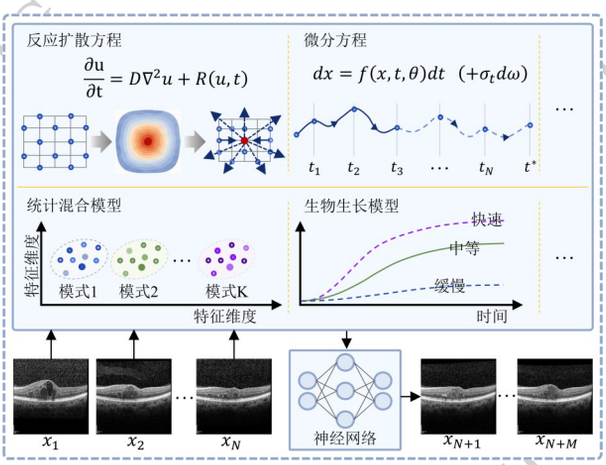

图13　基于物理/先验引导的通用架构示意图

Fig. 13　General architecture guided by physics and priors

5.1　基于反应扩散/微分方程的方法

纵向医学图像外推预测的最早尝试,往往直接诉诸于物理学和生物学中成熟的数学演化方程。其基本思想是将病理组织或解剖结构的演变,特别是肿瘤细胞的侵袭与病灶斑块的扩张,形式化为由PDE或ODE控制的时空动态传输过程。Zhu等人(2017b)将脉络膜新生血管(choroidal neovascularization, CNV)的3D生长过程建模为一个典型的反应扩散控制系统,其中反应项采用Logistic模型描述CNV病灶在空间中的增殖,扩散项则刻画病灶细胞向周围健康视网膜组织的渗透与侵袭,同时该方程还首次显式引入了治疗项以模拟抗VEGF药物对血管生长的动态抑制效应。最终在长达一年的病程终点预测中达到了76.40%的平均Dice表现。这一物理约束框架在临床可解释性方面展现出了良好的优势:临床医生可以通过检查方程中各项系数的物理含义来评估疾病进展的剧烈程度。然而,该工作也暴露了早期纯物理模型的局限:对样本量的极度敏感性(仅依赖7例患者进行验证)使得统计效力严重受限;数值求解极度依赖半自动分割的前期质量,导致分割误差极易向下游外推级联放大;且有限元法在高维3D体积数据上的计算复杂度高昂,无法满足现代大规模多中心临床数据的处理需求。

© 中国图象图形学报版权所有

为了缓解有限元数值求解的计算需求，研究者们开始尝试引入CNN作为代理模型来加速或替代物理方程的求解。Zhang等人(2018)提出的卷积侵袭和扩张网络在“物理机理”与“数据驱动”之间架起了一座精妙的桥梁。该工作没有直接求解复杂的PDE，而是顺应医学图像分析中由CNN提取局部解剖特征的技术趋势，通过两个解耦的深度CNN子网络分别对应物理方程中的核心项：侵袭网络通过融合代谢率、细胞密度和肿瘤边界等多模态影像特征来直接预测细胞的空间侵袭模式；而扩张网络则从肿瘤团块的宏观运动中建模质量效应导致的解剖形变，克服了纯物理方法无法泛化的缺陷。不过，这种将肿瘤侵袭与物理扩张强行分离的二分法假设，在处理真实世界中高度时空耦合、彼此交织的病理进展时，依然表现出了一定的简化局限。

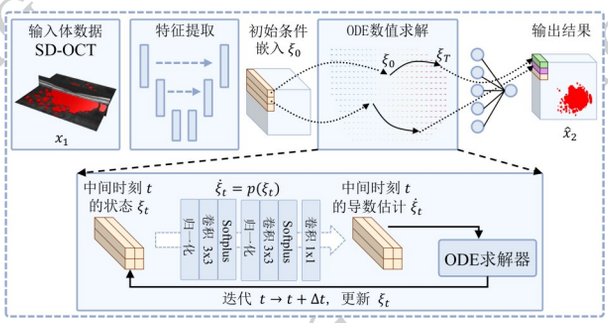

图14　NODE架构图(经许可修改自Lachinov等，2024)

Fig. 14　Architecture of NODE (adapted from Lachinov et al., 2024 with permission, ©2024 IEEE)

Lachinov等人(2024)将NeuralODE引入疾病进展建模，在方法论上实现了从离散物理求解器到连续时间学习的跃迁。如图14所示，其核心设计在于将时空疾病进展形式化为一个时间不变的物理过程，通过NeuralODE学习像素级变化的导数，从而在处理单次基线扫描后即可在任意请求的未来时间点输出目标解剖结构的分割。在967个OCT体和2823个脑MRI的大规模跨疾病验证中，该模型展示了NeuralODE在不规则随访场景下的优势：ODE求解器可在任意时间点求值，无需特殊适配即可处理可变间隔的临床数据。然而，该方法在进展速率随病程变化的疾病中可能引入偏差，且长时间ODE外推面临误差累积问题。在该工作的基础上，Lachinov等人(2026)进一步提出了PINN-EM，用INR连续编码空间坐标与时间轴到病变强度的映射，同时将经典反应扩散PDE转化为可微的“软约束项”来动态指导INR在外推未知时间域时的演化轨迹。其迭代期望-最大化算法的训练策略将非线性PDE系数恢复问题解耦为交替的系数拟合和数据拟合步骤，使模型在GA的新发病变区域预测中超越了后验知识模型。

5.2　基于统计模型与混合方法

相较于基于反应扩散/微分方程的物理流体或细胞扩散假设，另一条互补的技术路线则致力于探索统计概率与多物理机制混合的方法。这类方法并不预设极度严苛的解析方程形式，而是从多维临床数据中挖掘潜在的演变模式，同时通过构建可解释的统计组件或生物力学模型来保持预测链路的完全透明度。

在深度学习尚未应用于该领域之前，如何从高维影像中提炼确定性的临床风险因子并建立概率映射，是早期统计方法的关注核心。Niu等人(2016)作为这一探索时期的早期代表，系统性地从纵向视网膜SD-OCT扫描中提取了多维度的定量结构标志物，包括各视网膜层的厚度和反射率、网状假性玻璃膜疣的存在、光感受器丢失等临床公认风险因素。他们利用随机森林分类器构建预测矩阵，实现了对GA未来新生与扩张概率区域的定量评估。该方法的优势在于完全透明：每个生物标志物对预测的贡献可直接量化评估。但其输出的概率区域而非完整生成图像限制了临床信息量，且对大规模自动化特征提取的精度存在严重的级联依赖。Amiri等人(2019)从层次统计推理的角度提出了隐马尔可夫诱导的动态层次贝叶斯网络，在宏观组织尺度、中观细胞簇尺度以及微观细胞尺度三个空间跨度上编码了肿瘤细胞的复杂生物学交互。利用隐马尔可夫的转移概率捕获群体增长模式，再以贝叶斯信念传播实现患者级别的个性化动态预测。这种概率图模型框架天然支持不确定性建模和缺失数据处理，但其在2例脑肿瘤患者的极小样本上验证不足。

随着临床对解剖结构形态学演化精准度的要求日益提高，研究者们意识到，单一的生物成分扩散方程无法完美复现真实器官所承受的复杂物理形变。因此，从2021年起，该领域走向了多物理机制融合(如细胞增殖与生物力学耦合)的跨界发展阶段。Meng等人(2021)回归到CNV这一眼科进展难题，在纯反应扩散方程的基础上引入了超弹性生物力学模

© 中国图象图形学报版权所有

型。该框架的核心突破在于通过“质量效应”建立双向耦合：反应扩散方程负责定量描述CNV细胞在空间中的浓度跃迁与药物抑制，而超弹性力学方程则动态模拟CNV快速生长过程中对周围宿主视网膜组织造成的非线性大变形挤压与机械交互。然而，机制的叠加也反向拉高了模型对图像配准质量及半自动初始分割的依赖，导致系统对边缘模糊或伴有伪影的高噪点影像表现出了一定的鲁棒性脆弱。

随后，为了缓解繁重的物理机制求解并与现代深度学习融合，Kim 等人（2023）尝试优化传统物理方法在线求解的计算管线，探讨了将计算流体动力学模拟作为前置“物理特征提取器”的可行性，以此获取瘤壁剪切应力、血流流态等高阶血流动力学指标。该方法通过参数化映射机制，将非结构化的流体力学网格数据转化为适合标准 CNN 吞吐的规则张量表征，从而构建了从多物理特征到腹主动脉瘤未来局部生长率的端到端映射架构。这一尝试为缓解计算流体力学在实时临床部署中的计算瓶颈提供了新的思路，并在方法学上暗示了流体力学先验对大血管疾病进展外推的潜在临床价值。

此外，Hong等人（2025）提出的时间参数估计模型（normal inverse gamma distribution，T-NIG）为统计与混合方法的研究提供了新的视角。该方法尝试将贝叶斯统计中的正态逆伽马共轭分布融入到连续时间参数的估计过程中。通过坐标邻域检测特征，将时间跨度作为显式变量引入分布函数，以探索解剖结构在不同临床随访间隔下的动态退化规律；同时，该研究对纵向预测中由于数据稀缺引起的“认知不确定性”与由于影像噪声带来的“偶然不确定性”进行了显式解耦与量化尝试。实验表明，在仅需两个随机时间点基线图像的条件下，模型即可生成中间时点的解剖结构或外推未来的AD脑萎缩影像，一定程度上缓解了传统统计模型在高保真全维度图像生成上的局限性，为混合引导技术路线的发展提供了有益的探索思路。

优势与局限：物理与先验引导的建模机制通过将人类医学领域的确定性知识（如微分方程、生物力学机构及概率统计）引入纵向预测，在一定程度上缓解了纯数据驱动机制的“黑盒”属性。此类方法的主要优势在于能够提供良好的临床可解释性与更高的小样本数据利用效率，尝试建立更具透明度的预测机理，从而为数据稀缺或非均匀采样的复杂临床随访研究提供有益的解决思路。然而，这一路线的落地也面临着一些技术瓶颈：1）解析物理方程或多物理力学场的显式硬编码往往高度依赖高精度的前置影像配准与病灶分割，且级联数值求解的计算复杂度较高，对高维3D/4D体积数据的扩展性仍需进一步优化；2）刚性的数学公式与宏观的统计假设在面对高度异质、突变且复杂的个体化真实病理进展时，容易存在过度简化的局限。这促使其在近期呈现出向PINN与INR结合的演进趋势，旨在保持物理合规性的同时，进一步释放深度生成模型的架构弹性。

6 纵向预测生成的挑战与未来方向

纵向医学图像预测与生成作为连接计算机视觉与精准医疗的前沿交叉方向，在过去十年间经历了从纯解析物理模型到深度生成建模的重大跃迁。然而，要将这些方法学创新真正转化为能够辅助临床全周期管理的数字孪生系统，该领域仍面临着一系列深层的方法学瓶颈与技术挑战。

6.1 现存关键挑战

6.1.1 评估指标的系统性脱节与临床效用割裂

当前纵向预测生成领域最显著的瓶颈在于缺乏标准化的基准评估体系。由于不同研究采用的数据集、随访周期和预测跨度存在显著差异，导致跨研究的横向基准对比极为困难。更关键的是，目前普遍采用的PSNR、SSIM、FID等计算机视觉指标仅能评估像素级的图像重建质量，无法直接反映临床诊断的有效性。在实际诊断场景中，一个FID分数很低的生成图像可能在视网膜无血管区或脑肿瘤边界产生解剖学上不合理的虚假结构——这种”视觉上逼真而病理上虚假”的现象在临床应用中可能造成严重后果。因此，如何构建一套整合解剖形态一致性、病灶体积演变精度和下游临床任务性能的复合评估框架，已成为该技术走向临床的首要方法学难题。

6.1.2 纵向影像数据集的极端稀缺与孤岛效应

相较于横向医学影像数据的爆发式增长，具备长周期、多节点、高质量配对的纵向随访数据集在公开领域依然极为稀缺。除阿尔茨海默病（如ADNI队列）因起步较早而拥有相对完善的开源多中心数据外，眼科、肿瘤学和胎儿发育等核心方向的纵向研究大多依赖特定机构的私有临床数据。这种数据孤岛效应不仅限制了研究的可重复性，更使得深度模

© 中国图象图形学报版权所有

型在面对跨中心扫描设备差异、临床协议多样性和病理突变异质性时，泛化性能显著下降。在极小样本量和高度偏置的数据条件下，复杂的深度网络难以学习到稳健且普适的疾病演化规律。

6.1.3　时空维度的非均匀采样与长程外推误差累积

真实临床工作流中的纵向随访本质上是不受控且高度离散的。患者因个体差异、经济因素或依从性波动，导致随访时间间隔往往极不规则，且关键访视节点可能缺失。虽然以 NeuralODE 为代表的动力学系统天然支持连续物理时间的积分求解，而 Flow Matching 亦可通过连续时间嵌入或轨迹流构建任意时刻的外推映射，但在高维 3D/4D 影像空间的长程推演中，求解器与生成网络依然面临深层挑战：1）物理动力学求解的误差累积：对于直接沿病程时间轴进行连续状态积分的 ODE 求解器，离散化近似误差与病理漂移会随外推时间跨度 Δt 的增加而呈非线性级联放大，导致远期器官外生结构发生严重解剖失真；2）生成采样流与条件时间外推的泛化衰减：对于采用时间条件化的流匹配与扩散网络，当输入的随访时间间隔 Δt 超出训练集分布时，条件嵌入在隐空间的表示能力将迅速衰退，容易引发生成解剖拓扑的模糊与异常重构。

6.1.4　纵向空间配准误差与病理拓扑演化的双重挑战

纵向预测生成的基础前提是建立跨时间节点的高精度空间对应关系，以准确提取和度量解剖结构的局部演变。然而，在真实临床随访场景中，由于患者扫描体位差异、呼吸及器官生理运动以及设备硬件差异等因素，多时刻影像之间往往存在显著的空间偏移与非刚性形变。现有基于相似性度量或深度学习的配准方法在面对跨中心、多设备以及长时间跨度数据时仍存在泛化能力不足的问题，容易产生局部错配与形变伪影，并进一步导致前级配准误差在后续预测生成过程中的级联传播。更严峻的挑战在于，长期疾病进展通常伴随着显著的拓扑与形态结构变化（如肿瘤侵袭性外生生长，以及视网膜或脑组织的严重萎缩与重构等），这些变化可能超出传统连续形变模型对于结构保持性的基本假设。在此情况下，过度配准可能通过强制约束空间一致性而错误消除真实的病理进展信息，而配准不足则会使采集偏移与疾病演化特征深度混杂，导致模型难以学习稳定可靠的时空变化规律。因此，如何在维持个体解剖结构时序一致性的同时，自适应保留高异质性的病理结构演变特征，也是限制纵向预测生成领域进一步发展的瓶颈之一。

6.1.5　个体特异性特征保留与群体规律泛化之间的内在张力

纵向预测生成的核心挑战在于完美调和患者个体特异性表征（即解剖学身份保持）与群体级疾病进展共性（即病理演化趋势）之间的内在张力。若模型过度聚焦于捕获当前个体的精细解剖结构与当前基线特征，则极易陷入局部过拟合，从而在面对长程外推时彻底失去对未来非线性突发病变（如新发病灶的突然外生型扩散）的预测能力；反之，若模型过度倾向于提炼和顺应宏观群体的统计学疾病进展规律，则往往会抹杀患者个体化的病程斜率与异质性，导致最终生成的未来随访影像走向解剖结构的均值化与平庸化，无法有效区分高风险快速进展型与良性稳定型的个体差异，这严重背离了精准医疗的初衷。

6.2　开放问题与未来方向

6.2.1　跨中心医学纵向生成标准化评估基准的建立

构建多中心、大样本、覆盖多器官与多疾病种类的开源纵向影像基准（Benchmark）数据集，是加速该领域技术规范化迭代的坚实基石。未来的演进方向不仅在于促进高质量开源随访数据本身的释放，更在于制定出一套国际统一的纵向预测生成测试协议，明确规范标准基线输入的时间步长、外推预测窗口期的划分标准，并提供标准化的模型代码本。具体而言，未来基准的制定应满足上述最低评估基准组合：将强制比对基线复制下限、局部病理感兴趣区评估、演化变化量（Δ）残差测算、按外推时间步长分层分析，以及预测不确定性的统计校准度，纳入排行榜与模型评测的标准流水线中，使方法学开发者能够在一个高度透明、公正且贴合临床实际需求的流水线上进行公平比较，推动社区走出单纯依赖计算机视觉指标的评估模式。

6.2.2　跨时空维度的连续时间纵向基础大模型探索

面对高维纵向数据获取成本高昂且标注极度稀缺的现实，探索数据高效型的纵向基础大模型代表了该领域的必然趋势。研究者将不再孤立地依赖单

© 中国图象图形学报版权所有

中心的配对序列，而是首先利用海量无标签的横截面静态影像实施超大规模的自监督预训练，通过构建时空掩码自编码器或跨模态对比表征来提炼出具备强解剖归纳偏置的底层通用视觉骨架。随后，通过引入具有线性计算复杂度的状态空间模型或参数化自适应时间Transformer，将离散多时间点的三维高维影像与非结构化的电子健康档案以及微观的多组学数据进行跨模态联合统一嵌入，打造出能够跨疾病、跨器官自适应对齐不规则序列的基础大模型，从而实现低数据依赖下的高质量时空轨迹外推。

6.2.3　生物物理生长机制与连续时间深度生成的深层融合

纯数据驱动的生成模型在长程时空外推时，由于缺乏物理和生物边界的科学制约，容易生成反常识的解剖伪影。将确定性的生物流体力学机制、肿瘤细胞反应扩散偏微分方程与现代连续时间生成框架（如流匹配、隐式神经表示）进行深层融合，是确保模型外推合规性并提升临床可信度的有效路径。未来的发展方向在于将物理或生物学演化规律重构为可微的、动态演进的物理信息软约束项，或者直接作为控制潜在演化速度场的速度矢量导数。通过让隐式神经表示直接在全连续时空坐标上求值，配合物理规律对整个像素/潜空间的平滑规约，使得系统能够在数据极端稀缺或严重不规则采样的临床场景中，始终为模型外推提供符合生物学常理的硬性规约边界，最终实现符合生物学规律的高保真预测。

6.2.4　多模态时空信息融合与临床知识增强的纵向生成框架探索

当前纵向医学影像生成模型大多依赖单一影像序列进行疾病演化建模，而真实临床决策过程通常受到影像表型、患者基础信息、治疗方案以及分子生物标志物等多维因素共同驱动。单模态影像虽然能够提供丰富的空间结构信息，但难以充分解释疾病进展机制及个体间异质性的演变规律。因此，未来纵向生成模型将逐渐由单纯影像序列预测转向影像、临床文本、实验室指标、组学信息及治疗历史等多模态信息的统一时空建模。其核心挑战在于不同模态之间存在数据分布差异、采样频率不一致以及时间语义异构等问题，需要发展具备动态对齐能力的跨模态表示学习方法。未来可结合多模态基础模型、跨模态注意力机制以及医学知识约束，实现影像表型与临床语义信息在连续时间空间中的深度融合，使模型从“预测影像变化”进一步迈向“理解疾病演化机制”，实现更加精准、个体化和可解释的未来状态推演。

6.2.5　基于反事实推理的治疗响应模拟与个性化临床推演

医学图像纵向预测生成的终极临床价值是辅助个性化精准医疗的决策制定，因而将模型从单纯的“自然病程静态模拟”升级为具备“治疗响应干预推演”能力的反事实模型，是该领域最具转化医学价值的方向之一。未来的研究将致力于在连续生成框架内显式地将异质性的治疗干预方案（如放疗靶区剂量分布、手术切除路径选择、靶向药物给药频次）参数化为网络的动态控制条件，利用反事实推理回答"若采取不同治疗，患者未来解剖结构将如何演变"这一核心临床问题，从而真正帮助医生动态优化医疗方案，跨越从技术到临床应用的鸿沟。

6.2.6　面向疾病动态演进的医学病理世界模型构建与前瞻推演

近年来，“世界模型（world models）”在具身智能、通用视频生成与强化学习领域引起了广泛关注，其核心思想在于促使人工智能智能体对外部物理环境的因果律与时空转移规律形成内部抽象表征，进而基于当前观测与施加动作预判环境的未来状态。在方法学本质上，医学图像的纵向预测生成与通用世界模型具有深层次的结构同构性——该任务并非孤立的像素级风格转换，而是在构建一个理解人体内部解剖与病理演变规律的“病理生理世界模型（pathological world model）”。在这一构想下，患者的历史影像及多组学特征对应于系统的初始环境观测状态（state s_t），复杂的疾病自然进展机制被内生编码为潜在环境的状态转移函数（$P(s_{t+1}|s_t)$），而手术切除、放射治疗靶区剂量分布及药物给药频次等临床干预方案，则严格对应于对该生理系统施加的外部动作（action a_t）。

然而，相较于自动驾驶与仿真游戏等物理世界模型，医学病理世界模型的落地构建面临着显著的领域特异性挑战与理论边界：1）极端稀疏与不规则的动态观测：物理世界模型通常建立在高帧率连续视频流的基础之上，而纵向医学影像随访跨越数月乃至数年，采样极不规则且节点频繁缺失；这要求医学世界模型必须具备跨尺度、非均匀连续时间积分

© 中国图象图形学报版权所有

能力,摆脱离散马尔可夫决策过程的局限;2)生物病理过程的不可逆性与个体异质性:物理世界中的碰撞、位移往往满足刚体动力学守恒与时间可逆性,而组织的病理性坏死、纤维化与微环境免疫重塑通常属于热力学与生物学上的不可逆过程,且不同机体对同种临床"动作"的反馈轨迹高度异质,对世界模型的隐空间表达能力提出了极大挑战;3)长程反事实推演中的误差级联与安全性制约:在通用世界模型中,轻微的多步推演幻觉通常不致命;但医学世界模型若用于前瞻性"治疗方案仿真",在潜空间进行长程自回归推演时累积的时空解剖误差或虚构病变,将直接误导临床决策。

未来面向医学病理世界模型的技术突破,亟需从三个维度协同推进:一是探索联合嵌入预测架构(如 JEPA 范式),不再强求在像素空间逐体素重构高维噪声,而是直接在紧凑的高层抽象解剖流形中建模疾病对外界干预的因果动态;二是融合生物可微物理与流体力学引擎,将反应扩散偏微分方程和超弹性组织力学模型作为确定性"生物模拟先验",规约世界模型推演的状态转移边界,避免反常识拓扑撕裂;三是发展基于主动干预的动态仿真数字孪生体,使世界模型最终具备"若增减放疗剂量,肿瘤坏死边界将如何退缩"的前瞻仿真能力,真正实现从被动拟合病程向具备自主因果推理能力的医学数字孪生的跨越。

6.2.7　面向临床部署的纵向生成模型安全性与可信性保障

随着纵向医学影像生成模型逐步走向临床辅助决策,其安全性、可靠性与可解释性将成为关键问题。不同于分类或分割任务,生成模型可能产生视觉逼真的虚假解剖结构或错误病理变化,而这类难以察觉的幻觉信息可能对临床判断造成潜在风险。此外,医学影像中的设备差异、患者群体偏倚及隐私约束也可能导致模型跨中心应用时出现性能下降。因此,未来研究需要建立涵盖数据质量控制、不确定性量化、生成结果验证和异常检测的系统性安全机制。通过概率生成建模、置信度估计等方法,模型可主动识别超出训练分布的病例并提供可靠性反馈;结合可解释人工智能技术,则可进一步增强生成轨迹和疾病演化依据的透明度。同时,需要通过前瞻性临床验证、风险评估和持续性能监测,推动纵向生成模型由"高性能生成"向"可信赖医疗智能"转变。

6.2.8　前瞻性多中心临床工作流验证与转化医学落地

从实验室的 Dice 系数与 PSNR 等指标走向真实的临床部署,是检验纵向预测生成技术生命力的最终标准。未来的研究应跨越单纯的方法学架构修补,坚定地走向跨中心的前瞻性临床研究与院内影像归档与通信系统(picture archiving and communication system, PACS)系统的边缘端无缝嵌入。通过将轻量化的预测生成模块集成到日常放射学或眼科阅片的工作流中,严格、定量地评估虚拟未来随访影像在协助低资历医生规避高风险漏诊率、精准勾画放射治疗靶区,以及提升患者对病程演变感知度与依从性等方面的实际临床价值,用真实的转化医学证据链完成从学术算法到临床应用的闭环。

7　结　语

据作者所知,本工作是首篇专门针对纵向医学图像生成式预测进行系统梳理与深入解构的综述。研究纳入并深度分析了近十年间发表的 77 篇核心文献,尝试构建了涵盖"时序建模方法"、"深度生成模型"和"物理/先验引导模型"的三分类体系。通过对底层架构、条件化机制及物理约束等多个维度的交叉分析,本综述梳理了该领域从早期纯解析物理模型经生成对抗网络向连续时间潜空间扩散与流匹配演进的建模机制转移倾向。统计表明,深度生成模型目前在已发表工作中占据了较高比例(约 73%),在一定程度上反映了学界在高质量合成、无偏分布拟合和跨模态条件化层面的技术选择。此外,本综述提炼的临床现实痛点(如不规则采样、长程误差累积、个体身份保持与群体演化规律之间的张力等),也有望为后续的方法学攻坚提供有益的参考。

尽管本综述力求对前沿文献进行系统的检索与分类,但受限于该领域的快速迭代,本工作仍存在一定的局限性。首先,在任务范畴上,本综述主要将边界限定于二维图像或三维体数据级别的纵向预测生成,因而未深入涵盖仅输出疾病分期标签或临床指标的纯分类与回归工作;这种对生成任务的聚焦,可能在一定程度上减弱了对部分轻量级时序预测方法的覆盖度。其次,在文献检索方面,本工作主要涵盖了英文同行评议期刊、计算机视觉会议论文以及主

© 中国图象图形学报版权所有

流预印本，鉴于该领域广泛跨越眼科、神经影像和实体肿瘤等多个相对独立的临床方向，尽管进行了多轮关键词检索，仍不排除存在少数发表于特定医学专刊或新兴交叉会议的纵向影像预测原型未被完全纳入的情况。

医学图像纵向预测生成正处于向个体化数字孪生与精准医疗转化的重要节点。未来的潜在发展方向之一，在于将高分辨率三维影像、非结构化临床文本及多组学数据进行深度跨模态整合，以构建能够自适应对齐不规则随访序列的纵向基础大模型。同时，将确定性的生物医学机理以柔性约束形式融入流匹配等连续时间生成框架中，推动数据与物理的双驱交互，有利于进一步提升长程外推的解剖学合规性。最终，该任务范式的核心临床价值有望从“自然病程的被动模拟”向“动态干预的反事实推演”延伸通过将治疗方案参数化为显式控制条件，辅助医生具备在前瞻性窗口期内推演不同疗效路径的反事实决策能力。随着算法向临床工作流的逐步转化部署，该技术在辅助规避漏诊风险、优化放疗规划等方面的潜在价值将进一步释放。我们希望本综述能为该交叉领域的研究者提供一份清晰的方法学导航，并凝聚多学科社区的协同努力，共同推动先进纵向预测生成技术的演进。

参考文献(References)

Amiri S, Mahjoub M A, Vento M and Percannella G. 2019. HMDHBN: hidden Markov inducing a dynamic hierarchical Bayesian network for tumor growth prediction//Proceedings of the 18th International Conference on Computer Analysis of Images and Patterns. Salerno, Italy, Cham: Springer: 3-14 [DOI: 10.1007/978-3-030-29888-3_1]

Baheti B, Waldmannstetter D, Chakrabarty S, Akbari H, Bilello M, Wiestler B, et al. 2021. The brain tumor sequence registration challenge: Establishing correspondence between pre-operative and follow-up MRI scans of diffuse glioma patients [EB/OL]. [2021-12-13]. https://arxiv.org/pdf/2112.06979.pdf

Beekly D L, Ramos E M, Lee W W, Deitrich W D, Jacka M E, Wu J, et al. 2007. The National Alzheimer's Coordinating Center (NACC) database: the uniform data set. Alzheimer Disease & Associated Disorders, 21 (3): 249-258 [DOI: 10.1097/WAD.0b013e318142774e]

Chen H, Yin R, Chen Y, Chen Q and Li C. 2026. Learning patient-specific disease dynamics with latent flow matching for longitudinal imaging generation [EB/OL]. [2026-02-13]. https://arxiv.org/pdf/2512.09185.pdf

Chen M, Wang K, Dohopolski M, Morgan H, Sher D and Wang J. 2025. TransAnaNet: transformer-based anatomy change prediction network for head and neck cancer radiotherapy. Medical Physics, 52(5): 3015-3029 [DOI: 10.1002/mp.17655]

Cho H, Wei Z, Lee S, Dan T, Wu G and Kim W H. 2025. Conditional diffusion with ordinal regression: longitudinal data generation for neurodegenerative disease studies//Proceedings of the 13th International Conference on Learning Representations. Singapore, IMLS: 2025: 33329-33350

Dannecker M and Rueckert D. 2025. Predicting longitudinal brain development via implicit neural representations//Proceedings of the 28th International Conference on Medical Image Computing and Computer-Assisted Intervention. Daejeon, Republic of Korea, Cham: Springer: 477-487 [DOI: 10.1007/978-3-032-05141-7_46]

Dao D P, Yang H J, Kim J and Ho N H. 2025. Longitudinal Alzheimer disease progression prediction with modality uncertainty and optimization of information flow. IEEE Journal of Biomedical and Health Informatics, 29 (1): 259-272 [DOI: 10.1109/JBHI.2024.3472462]

Das A, Ram K and Sivaprakasam M. 2026. Align-cDAE: Alzheimer disease progression modeling with attention-aligned conditional diffusion auto-encoder//Proceedings of the 39th International Symposium on Computer-Based Medical Systems, Limassol, Cyprus, IEEE: 1048-1054 [DOI: 10.1109/CBMS69103.2026.00215]

Disch N and Zimmerer D. 2023. A holistic vision: modeling patient trajectories in longitudinal medical imaging//Proceedings of the Symbiosis of Deep Learning and Differential Equations III. Louisiana, USA

Disch N A, Kirchhoff Y, Peretzke R, Rokuss M, Roy S, Ulrich C, et al. 2025. Temporal flow matching for learning spatio-temporal trajectories in 4D longitudinal medical imaging [EB/OL]. [2025-08-29]. https://arxiv.org/pdf/2508.21580.pdf

Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. 2021. An image is worth 16x16 words: transformers for image recognition at scale [EB/OL]. [2021-06-03]. http://arxiv.org/pdf/2010.11929.pdf

Elazab A, Wang C, Gardezi S J S, Bai H, Hu Q, Wang T, et al. 2020. GP-GAN: brain tumor growth prediction using stacked 3D generative adversarial networks from longitudinal MR images. Neural Networks, 132: 321-332 [DOI: 10.1016/j.neunet.2020.09.004]

Ellis K A, Bush A I, Darby D, De Fazio D, Foster J, Hudson P, et al. 2009. The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's dis-

© 中国图象图形学报版权所有

ease. International psychogeriatrics, 21(4): 672-687 [DOI: 10.1017/S1041610209009405]

Fan Y, Xie J, Luo Y, Meng Y, Madhusudhan S, Lip G Y H, et al. 2025. tHPM-LDM: integrating individual historical record with population memory in latent diffusion-based glaucoma forecasting// Proceedings of the 28th International Conference on Medical Image Computing and Computer-Assisted Intervention. Daejeon, Republic of Korea, Cham: Springer: 619-629 [DOI: 10.1007/978-3-032-04927-8_59]

Fang J, Wang J, Li A, Yan Y, Liu H, Li J, et al. 2023. Parameterized Gompertz-guided morphological autoencoder for predicting pulmonary nodule growth. IEEE Transactions on Medical Imaging, 42(12): 3610-3622 [DOI: 10.1109/TMI.2023.3297209]

Fang Y, Xiong H, Huang J, Wang X, Shen Z, Liu F, et al. 2025. Longitudinal image synthesis for early brain development with age awareness//Proceedings of the 22nd International Symposium on Biomedical Imaging. Texas, USA, IEEE: 1-5 [DOI: 10.1109/ISBI60581.2025.10981151]

Favero G M, Luo G Y, Fathi N, Szeto J, Arnold D L, Nichyporuk B, et al. 2025. Spatio-temporal conditional diffusion models for forecasting future multiple sclerosis lesion masks conditioned on treatments// Proceedings of the International Workshop on Learning with Longitudinal Medical Images and Data. Daejeon, Republic of Korea, Cham: Springer: 56-67 [DOI: 10.1007/978-3-032-16128-4_6]

Gao Y, Zhou H Y, Wang X, Zhang T, Han L, Lu C, et al. 2024. Improving neoadjuvant therapy response prediction by integrating longitudinal mammogram generation with cross-modal radiological reports: a vision-language alignment-guided model//Proceedings of the 27th International Conference on Medical Image Computing and Computer-Assisted Intervention. Marrakesh, Morocco, Cham: Springer: 133-143 [DOI: 10.1007/978-3-031-72378-0_13]

Gong J, Liu J, Wang S, Fu C, Zhang Z, Wu X, et al. 2025. Knowledge-guided dynamic-static feature synergy for 4DCT medical image generation//Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine. Wuhan, China, IEEE: 6248-6255 [DOI: 10.1109/BIBM66473.2025.11356695]

Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. 2014. Generative adversarial nets//Advances in Neural Information Processing Systems. Quebec, Canada, 27: 2672-2680 [DOI: 10.5555/2969033.2969125]

Gu A and Dao T. 2024. Mamba: linear-time sequence modeling with selective state spaces [EB/OL]. [2024-05-31]. https://arxiv.org/pdf/2312.00752.pdf

Hazlett H, Gu H, Munsell B C, Kim S H, Styner M, Wolff J J, et al. 2017. Early brain development in infants at high risk for autism spectrum disorder. Nature, 542: 348 - 351 [DOI: 10.1038/nature21369]

Ho J, Jain A and Abbeel P. 2020. Denoising diffusion probabilistic models//Advances in Neural Information Processing Systems. Virtual Event, 33: 6840-6851 [DOI: 10.5555/3495724.3496298]

Hochreiter S and Schmidhuber J. 1997. Long short-term memory. Neural Computation, 9(8): 1735-1780 [DOI: 10.1162/neco.1997.9.8.1735]

Hong X, Sun X, Li Y and Chen Y W. 2025. Long-term Alzheimer disease prediction: a novel image generation method using temporal parameter estimation with normal inverse gamma distribution on uneven time series [EB/OL]. [2025-11-26]. https://arxiv.org/pdf/2511.21057.pdf

Huang Y, Ahmad S, Han L, Wang S, Wu Z, Lin W, et al. 2023. Longitudinal prediction of postnatal brain magnetic resonance images via a metamorphic generative adversarial network. Pattern Recognition, 143: 109715 [DOI: 10.1016/j.patcog.2023.109715]

Jack Jr C R, Bernstein M A, Fox N C, Thompson P, Alexander G, Harvey D, et al. 2008. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine, 27(4): 685-691 [DOI: 10.1002/jmri.21049]

Jiang C, Xing X, Nan Y, Fang Y, Zhang S, Walsh S, et al. 2025. A lung structure and function information-guided residual diffusion model for predicting idiopathic pulmonary fibrosis progression. Medical Image Analysis, 96: 103604 [DOI: 10.1016/j.media.2025.103604]

Johnson A E, Pollard T J, Berkowitz S J, Greenbaum N R, Lungren M P, Deng C Y, et al. 2019. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. Scientific data, 6(1): 317 [DOI: 10.6084/m9.figshare.10303823]

Karras T, Laine S and Aila T. 2019. A style-based generator architecture for generative adversarial networks//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. California, USA, IEEE: 4401-4410 [DOI: 10.1109/CVPR.2019.00453]

Kazerouni A, Aghdam E K, Heidari M, Azad R, Fayyaz M, Hacihaliloglu I, et al. 2023. Diffusion models in medical imaging: a comprehensive survey. Medical Image Analysis, 88: 102846 [DOI: 10.1016/j.media.2023.102846]

Kim S, Jiang Z, Zambrano B A, Jang Y, Baek S, Yoo S, et al. 2023. Deep learning on multiphysical features and hemodynamic modeling for abdominal aortic aneurysm growth prediction. IEEE Transactions on Medical Imaging, 42(1): 196-208 [DOI: 10.1109/TMI.2022.3206142]

Kingma D P and Welling M. 2014. Auto-encoding variational Bayes//Proceedings of the International Conference on Learning Representations. Alberta, Canada, ICLR: [DOI: 10.48550/arXiv.1312.6114]

Lachinov D, Chakravarty A, Grechenig C, Schmidt-Erfurth U and Bogunovic H. 2024. Learning spatio-temporal model of disease progression with NeuralODEs from longitudinal volumetric data. IEEE

© 中国图象图形学报版权所有

Transactions on Medical Imaging, 43(3): 1165-1179 [DOI: 10.1109/TMI.2023.3330576]

Lachinov D, Pinetz T and Bogunovic H. 2026. PINN-EM: physics-guided disease progression model of geographic atrophy. IEEE Transactions on Biomedical Engineering, 1-12 [DOI: 10.1109/TBME.2026.3655164]

LaMontagne P J, Benzinger T L, Morris J C, Keefe S, Hornbeck R, Xiong C, et al. 2019. OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease. medrxiv, 2019-12 [DOI: 10.1101/2019.12.13.19014902]

Laslo D, Georgiou E, Linguraru M G, Rauschecker A M, Muller S, Jutzeler C R, et al. 2025. Mechanistic learning with guided diffusion models to predict spatio-temporal brain tumor growth//Proceedings of the 3rd International Workshop on Longitudinal Image Data. Daejeon, Republic of Korea, Cham: Springer: 68-79 [DOI: 10.1007/978-3-032-16128-4_7]

Lei K Z, Lee S Y and Huang C R. 2025. Temporal feature-guided latent diffusion network for generating future low-dose CT images in early cancer monitoring//Proceedings of the 10th International Conference on Intelligent Informatics and Biomedical Sciences. Okinawa, Japan, IEEE: 10: 1-8 [DOI: 10.1109/ICIIBMS66230.2025.11316735]

Li C, Li Z, Feng B, Chen Y, Liu Z, Zhang S, et al. 2021. Stroke image prediction method based on active contour model and radiomics. Modern Electronics Technique, 44(2): 169-173 (李昌林, 李智, 冯宝, 陈业航, 刘壮盛, 张绍荣等. 2021. 基于活动轮廓模型和影像组学的脑卒中图像预测方法. 现代电子技术, 44(2): 169-173) [DOI: 10.16652/j.issn.1004-373x.2021.02.036]

Li L, Wang X, Xu M, Liu H and Chen X. 2020. DeepGF: Glaucoma forecast using the sequential fundus images//Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Virtual Event, Cham: Springer: 626-635 [DOI: 10.1007/978-3-030-59722-1_60]

Li M, Zhang W, Zhao H, Xu Y, Xu J and Li H. 2024. A spatiotemporal convolution recurrent neural network for pixel-level peripapillary atrophy prediction using sequential fundus images. Applied Soft Computing, 155: 111431 [DOI: 10.1016/j.asoc.2024.111431]

Li M, Zhao H, Zhang W, Xu J and Li H. 2026a. A spatiotemporal model for future fundus image prediction with irregularly sampled sequential data. Applied Soft Computing, 186: 114062 [DOI: 10.1016/j.asoc.2025.114062]

Li X, Huang K, Zhang Y, Yuan S, Niu S and Chen Q. 2026b. MT-DENet: prediction of post-therapy OCT images in diabetic macular edema by multi-temporal disease evolution network. Biomedical Signal Processing and Control, 112: 108721 [DOI: 10.1016/j.bspc.2025.108721]

Li X, Zhang Y, Niu S, Yuan S and Chen Q. 2026c. From discrete to continuous: a spatiotemporal evolution-aware adversarial diffusion framework for retinal disease progression prediction. Neurocomputing, 660: 131873 [DOI: 10.1016/j.neucom.2025.131873]

Lipman Y, Chen R T Q, Ben-Hamu H, Nickel M and Le M. 2023. Flow matching for generative modeling//Proceedings of the International Conference on Learning Representations. Kigali, Rwanda, IMLS: [DOI: 10.48550/arXiv.2210.02747]

Litrico M, Guarnera F, Giuffrida M V, Ravi D and Battiato S. 2024. TADM: temporally-aware diffusion model for neurodegenerative progression on brain MRI//Proceedings of the 27th International Conference on Medical Image Computing and Computer-Assisted Intervention. Marrakesh, Morocco, Cham: Springer: 444-453 [DOI: 10.1007/978-3-031-72069-7_42]

Litrico M, Guarnera F, Giuffrida M V, Ravi D and Battiato S. 2026. Temporally-aware diffusion model for brain progression modelling with bidirectional temporal regularisation. Computerized Medical Imaging and Graphics, 127: 102688 [DOI: 10.1016/j.compmedimag.2025.102688]

Liu C, Xu K, Shen L L, Huguet G, Wang Z, Tong A, et al. 2025a. ImageFlowNet: forecasting multiscale image-level trajectories of disease progression with irregularly-sampled longitudinal medical images//Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Hyderabad, India, IEEE: 1-5 [DOI: 10.1109/ICASSP49660.2025.10890535]

Liu Q, Fuster-Garcia E, Thokle Hovden I, MacIntosh B J, Grodem E O S, Brandal P, et al. 2025b. Treatment-aware diffusion probabilistic model for longitudinal MRI generation and diffuse glioma growth prediction. IEEE Transactions on Medical Imaging, 44(6): 2449-2462 [DOI: 10.1109/TMI.2025.3533038]

Liu X, Gong C and Liu Q. 2022. Flow straight and fast: learning to generate and transfer data with rectified flow[EB/OL].[2022-09-07]. https://arxiv.org/pdf/2209.03003.pdf

Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. 2021. Swin transformer: hierarchical vision transformer using shifted windows//Proceedings of the IEEE/CVF International Conference on Computer Vision. Virtual Event, IEEE: 10012-10022 [DOI: 10.1109/ICCV48922.2021.00986]

Luo X, Li H, Cheng X, Zhao H, Hu Y, Song X, et al. 2026. TraceTrans: translation and spatial tracing for surgical prediction//Proceedings of the AAAI Conference on Artificial Intelligence. Singapore, AAAI: 40(9): 7698-7706 [DOI: 10.1609/aaai.v40i9.37712]

Luo Y, Jiang H and Xu F. 2025. Predicting response to neoadjuvant therapy in breast cancer using deep learning on primary core Needle biopsy slides. Chinese General Practice, 28(19): 2407-2413 (罗云昭, 蒋宏传 和 徐峰. 2025. 基于深度学习模型辅助穿刺病理图像预测乳腺癌新辅助治疗疗效的研究. 中国全科医学, 28(19): 2407-2413) [DOI: 10.12114/j.issn.1007-9572.2023.0512]

Ma M, Zhang X, Li Y, Wang X, Zhang R, Wang Y, et al. 2023. Conv-LSTM coordinated longitudinal transformer under spatio-temporal

© 中国图象图形学报版权所有

features for tumor growth prediction. Computers in Biology and Medicine, 164: 107313 [DOI: 10.1016/j. compbiomed. 2023. 107313]

Makropoulos A, Robinson E C, Schuh A, Wright R, Fitzgibbon S, Bozek J, et al. 2018. The developing human connectome project: A minimal processing pipeline for neonatal cortical surface reconstruction. Neuroimage, 173: 88-112 [DOI: 10.1016/j.neuroimage. 2018.01.054]

Mantegna M, Ayllon E M, Caragliano A N, Di Feola F, Tacconi C, Fiore M, et al. 2026. Virtual treatment for radiotherapy: multimodal generative prediction of longitudinal NSCLC tumor progression. Computerized Medical Imaging and Graphics, 132: 102774 [DOI: 10.1016/j.compmedimag.2026.102774]

Marti-Juan G, Lorenzi M, Piella G and Alzheimer's Disease Neuroimaging Initiative. 2023. MC-RVAE: multi-channel recurrent variational autoencoder for multimodal Alzheimer's disease progression modelling. NeuroImage, 268: 119892 [DOI: 10.1016/ j.neuroimage.2023.119892]

Meng Q, Zuo C, Shi F, Zhu W, Xiang D, Chen H and Chen X. 2021. Three-dimensional choroid neovascularization growth prediction from longitudinal retinal OCT images based on a hybrid model. Pattern Recognition Letters, 146: 108-114 [DOI: 10.1016/j. patrec. 2021.03.006]

Mirza M and Osindero S. 2014. Conditional generative adversarial nets [EB/OL]. [2014-11-06]. https://arxiv.org/pdf/1411.1784.pdf

Mouadden N, Laousy O, Marini R, Ong V, Revel M P, Chassagnon G, et al. 2025. Conditional latent diffusion models for irregularly spaced longitudinal radiological data//Proceedings of the 28th International Conference on Medical Image Computing and Computer-Assisted Intervention. Daejeon, Republic of Korea, Cham: Springer, 106-115 [DOI: 10.1007/978-3-032-05182-0_11]

National Lung Screening Trial Research Team. 2011. Reduced lung-cancer mortality with low-dose computed tomographic screening. New England Journal of Medicine, 365(5): 395-409 [DOI: 10. 1056/ NEJMoa1102873]

Ning Z, Zhang Y, Pan Y, Zhong T, Liu M and Shen D. 2020. LDGAN: Longitudinal-diagnostic generative adversarial network for disease progression prediction with missing structural MRI//Proceedings of the International Workshop on Machine Learning in Medical Imaging. Virtual Event, Cham: Springer, 12436: 170-179 [DOI: 10. 1007/978-3-030-59861-7_18]

Niu S, De Sisternes L, Chen Q, Rubin D L and Leng T. 2016. Fully automated prediction of geographic atrophy growth using quantitative spectral-domain optical coherence tomography biomarkers. Ophthalmology, 123(8): 1737-1745 [DOI: 10.1016/ j. ophtha. 2016.04.042]

Pan Y, Ma H, Xia Y and Zhang Y. 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]

Peebles W and Xie S. 2023. Scalable diffusion models with transformers// Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris, France, IEEE: 4195-4205 [DOI: 10.1109/ ICCV51070.2023.00387]

Peng L, Lin L, Lin Y, Chen Y, Mo Z, Vlasova RM, et al. 2021. Longitudinal prediction of infant MR images with multi-contrast perceptual adversarial learning. Frontiers in Neuroscience, 15: 653213 [DOI: 10.3389/fnins.2021.653213]

Pham Q, Ahn S, Shin J and Song S J. 2022. Generating future fundus images for early age-related macular degeneration based on generative adversarial networks. Computer Methods and Programs in Biomedicine, 216: 106648 [DOI: 10.1016/j.cmpb.2022.106648]

Prabhushankar M, Kokilepersaud K, Logan Y Y, Trejo Corona S, AlRegib G and Wykoff C. 2022. Olives dataset: Ophthalmic labels for investigating visual eye semantics//Advances in Neural Information Processing Systems. Louisiana, USA, 35: 9201-9216 [DOI: 10.52202/068431-0669]

Puglisi L, Alexander D C, Ravi D, Linguraru M G and Dou Q. 2024. Enhancing spatiotemporal disease progression models via latent diffusion and prior knowledge//Proceedings of the 27th International Conference on Medical Image Computing and Computer-Assisted Intervention. Marrakesh, Morocco, Cham: Springer: 173-183 [DOI: 10.1007/978-3-031-72069-7_17]

Puglisi L, Alexander D C and Ravi D. 2025. Brain latent progression: individual-based spatiotemporal disease progression on 3D brain MRIs via latent diffusion. Medical Image Analysis, 103: 103734 [DOI: 10.1016/j.media.2025.103734]

Qiao H, Tang F, Zhou H, Cai Y, Guo K, Wang J, et al. 2025. Forecasting the diabetic retinopathy progression using generative adversarial networks. Communications Medicine, 5(1): 368 [DOI: 10. 1038/ s43856-025-01092-2]

Ravi D, Alexander D C, Oxtoby N P and Alzheimer's Disease Neuroimaging Initiative. 2019. Degenerative adversarial neuroimage nets: generating images that mimic disease orogression //Proceedings of the 22th International Conference on Medical Image Computing and Computer Assisted Intervention. Shenzhen, China, Cham: Springer, 11766:164-172 [DOI: 10.1007/978-3-030-32248-9_19]

Rohm M, Tresp V, Müller M, Kern C, Manakov I, Weiss M, et al. 2018. Predicting visual acuity by using machine learning in patients treated for neovascular age-related macular degeneration. Ophthalmology, 125(7): 1028-1036. [DOI: 10.1016/j. ophtha. 2017.12.034]

Rombach R, Blattmann A, Lorenz D, Esser P and Ommer B. 2022. High-resolution image synthesis with latent diffusion models// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Louisiana, USA, IEEE: 10684-10695 [DOI: 10.

© 中国图象图形学报版权所有

1109/CVPR52688.2022.01042］

Ronneberger O，Fischer P and Brox T. 2015. U-Net：convolutional networks for biomedical image segmentation//Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich，Germany，Cham：Springer：234-241［DOI：10.1007/978-3-319-24574-4_28］

Sauty B and Durrleman S. 2022. Longitudinal variational autoencoders learn a riemannian progression model for imaging data//Proceedings of the Geometric Deep Learning in Medical Image Analysis（Extended abstracts）. Amsterdam，The Netherlands，PMLR

Sampathkumar A and Kowerko D. 2026. Longitudinal forecasting of retinal structure and function using a multimodal StyleGAN-based architecture. Bioengineering，13(2)：149［DOI：10.3390/ bioengineering13020149］

Shi X，Chen Z，Wang H，Yeung D Y and Wong W K. 2015. Convolutional LSTM network：a machine learning approach for precipitation nowcasting//Advances in Neural Information Processing Systems. Quebec，Canada，28：802-810［DOI：10.5555/2969636.2969724］

Song H and Wu X. 2019. High-quality image generation model for face aging/de-aging. Journal of Image and Graphics，24(4)：592-602（宋昊泽，吴小俊. 2019. 人脸老化/去龄化的高质量图像生成模型. 中国图象图形学报，24(4)：592-602）［DOI：10.11834/jig.180272］

Song J，Wang Y，Zhou C and Shen L. 2026. Nodule-aligned latent space learning with LLM-driven multimodal diffusion for lung nodule progression prediction［EB/OL］.［2026-03-16］. https://arxiv.org/pdf/2603.15932.pdf

Song J，Meng C and Ermon S. 2021. Denoising diffusion implicit models//Proceedings of the International Conference on Learning Representations. Virtual Event［DOI：10.48550/arXiv.2010.02502］

Suter Y，Knecht U，Valenzuela W，Notter M，Hewer E，Schucht P，et al. 2022. The LUMIERE dataset：longitudinal glioblastoma MRI with expert RANO evaluation. Scientific data，9(1)：768［DOI：10.1038/ s41597-022-01881-7］

Tang X，Luo Z，Liu F，Huang W and Zou J. 2026. NGP-net：a lightweight growth prediction network for pulmonary nodules. IEEE Transactions on Medical Imaging，45(5)：2468-2480［DOI：10.1109/TMI.2026.3656184］

Van den Oord A，Vinyals O and Kavukcuoglu K. 2017. Neural discrete representation learning//Advances in Neural Information Processing Systems. California，USA，30：6309-6318［DOI：10.5555/3235684. 3236022］

Vaswani A，Shazeer N，Parmar N，Uszkoreit J，Jones L，Gomez A N，et al. 2017. Attention is all you need//Advances in Neural Information Processing Systems. California，USA，30：5998-6008［DOI：10.5555/3295222.3295349］

Wang C，R Alam S，Zhang S，Hu Y C，Nadeem S，Tyagi N，et al. 2020. Predicting spatial esophageal changes in a multimodal longitudinal imaging study via a convolutional recurrent neural network. Physics in Medicine and Biology，65(23)：235027［DOI：10.1088/1361-6560/abb1d9］

Wang H，Ren Z，Chang H，Qiu X，Wang F，Lian C and Ma J. 2025. Flexibly distilled 3D rectified flow with anatomical constraints for developmental infant brain MRI prediction//Proceedings of the 28th International Conference on Medical Image Computing and Computer-Assisted Intervention. Daejeon，Republic of Korea，Cham：Springer：228-237［DOI：10.1007/978-3-032-05182-0_23］

Wang H，Xiao N，Zhang J，Yang W，Ma Y，Suo Y，et al. 2022. Static-dynamic coordinated transformer for tumor longitudinal growth prediction. Computers in Biology and Medicine，148：105922［DOI：10.1016/j.compbiomed.2022.105922］

Wen Y，Zeng Y，Bi L，Zhao X，Shi W，Fu H，et al. 2026. M2Net：multimodal multitask mutual learning for Anti-VEGF efficacy prediction. IEEE Transactions on Medical Imaging［DOI：10.1109/TMI.2026.3684331］

Weng J，Bhupathiraju S H V，Samant T，Dresner A，Wu J and Samant S S. 2024. Convolutional LSTM model for cine image prediction of abdominal motion. Physics in Medicine and Biology，69(8)：85024［DOI：10.1088/1361-6560/ad3722］

Wu Y，Wang Y，Zhang Q，Zhou C and Ying L. 2025. Early lung cancer diagnosis from virtual follow-up LDCT generation via correlational autoencoder and latent flow matching［EB/OL］.［2025-11-22］. https://arxiv.org/pdf/2511.18185.pdf

Xia T，Chartsias A，Wang C，Tsaftaris S A and Alzheimer's Disease Neuroimaging Initiative. 2021. Learning to synthesise the ageing brain without longitudinal data. Medical Image Analysis，73：102169［DOI：10.1016/j.media.2021.102169］

Xiao N，Li A，Qiang Y，Zhao J，Geng Y and Li L. 2026. Lung tumor growth prediction of follow-up via spatio-temporal convolutional transformer. Biomedical Signal Processing and Control，119：109858［DOI：10.1016/j.bspc.2026.109858］

Xiao N，Qiang Y，Zhao Z，Zhao J and Lian J. 2020. Tumour growth prediction of follow-up lung cancer via conditional recurrent variational autoencoder. IET Image Processing，14(15)：3975-3981［DOI：10.1049/iet-ipr.2020.0496］

Xiong W，Yeung N，Wang S，Liao H，Wang L and Luo J. 2022. Breast cancer induced bone osteolysis prediction using temporal variational autoencoders. BME Frontiers，2022：9763284［DOI：10.34133/ 2022/9763284］

Xu F，Wan C，Zhao L，Liu S，Hong J，Xiang Y，et al. 2021. Predicting post-therapeutic visual acuity and OCT images in patients with central serous chorioretinopathy by artificial intelligence. Frontiers in Bioengineering and Biotechnology，9：649221.［DOI：10.3389/fbioe.2021.649221］

Xu K，Feng B，Chen Y，Liu Y，Zhou H and Chen X. 2022. Thymoma CT image prediction method based on deep learning and improved

© 中国图象图形学报版权所有

extreme learning machine ensemble learning. Computer Science, 49(z2): 443-448 (徐坤财, 冯宝, 陈业航, 刘昱, 周皓阳和陈相猛. 2022. 结合深度学习与改进的极限学习机的集成学习胸腺瘤CT图像预测方法. 计算机科学, 49(z2): 443-448) [DOI: 10.11896/jsjkx.211200097]

Yang H, Tan Z, Tan S, Yang W, Cai K, Chen C and Sun Y. 2025. MambaControl: anatomy graph-enhanced Mamba ControlNet with Fourier refinement for diffusion-based disease trajectory prediction [EB/OL]. [2025-05-15]. https://arxiv.org/pdf/2505.09965.pdf

Yi X, Walia E and Babyn P. 2019. Generative adversarial network in medical imaging: A review. Medical Image Analysis, 58: 101552 [DOI: 10.1016/j.media.2019.101552]

Yoo T K, Choi J Y and Kim H K. 2020. A generative adversarial network approach to predicting postoperative appearance after orbital decompression surgery for thyroid eye disease. Computers in Biology and Medicine, 118: 103628 [DOI: 10.1016/j.compbiomed.2020.103628]

Yoon J, Zhang C, Suk H I, Guo J and Li X. 2023. SADM: sequence-aware diffusion model for longitudinal medical image generation// Proceedings of the 28th International Conference on Information Processing in Medical Imaging. San Carlos de Bariloche, Argentina, Cham: Springer: 388-400 [DOI: 10.1007/978-3-031-34048-2_30]

Yuan X, Cheng J, Hu D, Wu Z, Wang L, Lin W, et al. 2024. Longitudinally consistent individualized prediction of infant cortical morphological development//Proceedings of the 27th International Conference on Medical Image Computing and Computer-Assisted Intervention. Marrakesh, Morocco, Cham: Springer: 447-457 [DOI: 10.1007/978-3-031-72086-4_42]

Zhang K, Chen G, Huang S, Zhu F, Ding Z and Shen D. 2025. Development-driven diffusion model for longitudinal prediction of fetal brain MRI with unpaired data. IEEE Transactions on Medical Imaging, 44(9): 3641-3653 [DOI: 10.1109/TMI.2024.3496860]

Zhang L, Lu L, Wang X, Zhu R M, Bagheri M, Summers R M and Yao J. 2020. Spatio-temporal convolutional LSTMs for tumor growth prediction by learning 4D longitudinal patient data. IEEE Transactions on Medical Imaging, 39(4): 1114-1126 [DOI: 10.1109/TMI.2019.2943841]

Zhang L, Lu L, Summers R M, Kebebew E and Yao J. 2018. Convolutional invasion and expansion networks for tumor growth prediction. IEEE Transactions on Medical Imaging, 37(2): 638-648 [DOI: 10.1109/TMI.2017.2774044]

Zhang Y, Huang K, Li M, Yuan S and Chen Q. 2023. Learn single-horizon disease evolution for predictive generation of post-therapeutic neovascular age-related macular degeneration. Computer Methods and Programs in Biomedicine, 230: 107364 [DOI: 10.1016/j.cmpb.2023.107364]

Zhang Y, Huang K, Yang X, Ma X, Wu J and Wang N. 2024. Coarse-to-fine latent diffusion model for glaucoma forecast on sequential fundus images//Proceedings of the 27th International Conference on Medical Image Computing and Computer-Assisted Intervention. Marrakesh, Morocco, Cham: Springer: 166-176 [DOI: 10.1007/978-3-031-72086-4_16]

Zhang Y, Ji Z, Niu S, Leng T, Rubin D L and Chen Q. 2019. A multi-scale deep convolutional neural network for joint segmentation and prediction of geographic atrophy in SD-OCT images//Proceedings of the IEEE 16th International Symposium on Biomedical Imaging. Venice, Italy, IEEE: 565-568 [DOI: 10.1109/ISBI.2019.8759253]

Zhang Y, Zhang X, Ji Z, Niu S, Leng T, Rubin D L, et al. 2021. An integrated time adaptive geographic atrophy prediction model for SD-OCT images. Medical Image Analysis, 68: 101893 [DOI: 10.1016/j.media.2020.101893]

Zhao A, Xu M, Shahin A H, Wuyts W, Jones M G and Jacob J. 2025. 4D VQ-GAN: synthesising medical scans at any time point for personalised disease progression modelling of idiopathic pulmonary fibrosis [EB/OL]. [2025-02-08]. https://arxiv.org/pdf/2502.05713.pdf

Zhao P, Song X, Xi X, Nie X, Meng X, Qu Y and Yin Y. 2024a. Biomarkers-aware asymmetric bibranch GAN with adaptive memory batch normalization for prediction of anti-VEGF treatment response. IEEE Journal of Biomedical and Health Informatics, 28(1): 557-568 [DOI: 10.1109/JBHI.2023.3302989]

Zhao Z, Yang J, Faghihroohi S, Zhao Y, Zapp D, Huang K, et al. 2024b. Extrapolating prospective glaucoma fundus images through diffusion in irregular longitudinal sequences//Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine. Lisbon, Portugal, IEEE: 4032-4035 [DOI: 10.1109/BIBM62325.2024.10822368]

Zhou T, Noeuveglise A, Modzelewski R, Ghazouani F, Thureau S, Fontanilles M and Ruan S. 2023. Prediction of brain tumor recurrence location based on multi-modal fusion and nonlinear correlation learning. Computerized Medical Imaging and Graphics, 106: 102218 [DOI: 10.1016/j.compmedimag.2023.102218]

Zhu J Y, Park T, Isola P and Efros A A. 2017a. Unpaired image-to-image translation using cycle-consistent adversarial networks//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, IEEE: 2223-2232 [DOI: 10.1109/ICCV.2017.244]

Zhu S, Shi F, Xiang D, Zhu W, Chen H and Chen X. 2017b. Choroid neovascularization growth prediction with treatment based on reaction-diffusion model in 3-D OCT images. IEEE Journal of Biomedical and Health Informatics, 21(6): 1667-1674 [DOI: 10.1109/JBHI.2017.2702603]

作者简介

李孝辉,男,博士研究生,主要研究方向为医学影像时序预测生成。E-mail: xiaohuili@njust.edu.cn

© 中国图象图形学报版权所有

汪语桐，男，硕士研究生，主要研究方向为医学影像分析与生成。E-mail：1241060229 73@njust.edu.cn
张语涵，男，博士，主要研究方向为医学影像分析与多模态融合。E-mail：zhangyuhan@szu.edu.cn
牛四杰，男，教授，主要研究方向为模式识别与医学影像分析。E-mail：ise_niusj@ujn.edu.cn
陈强，男，教授，主要研究方向为模式识别与医学影像分析。E-mail：chen2qiang@njust.edu.cn

© 中国图象图形学报版权所有