

MTM3D:融合 Mamba 与改进 TTM 的三维医学图像分析模型

杨洋¹, 魏弘凯¹, 孙士杰², 胡红利¹, 王荣¹, 王天添¹

(1. 长安大学信息工程学院, 陕西 西安 710064; 2. 长安大学数据科学与人工智能研究院, 陕西 西安 710064)

摘要: 生物医学成像在诊断和治疗多种疾病中起着至关重要的作用。将深度学习方法应用于医学图像分析能够提高医学图像的可读性, 为临床决策提供更可靠的支持。然而, 传统的医学图像处理方法在有效捕获三维图像中的空间特征和复杂结构信息方面存在一定局限性, 尤其是在处理不同成像方式生成的复杂三维医学图像时, 模型的精度和泛化能力常常受限。针对这一挑战, 提出了一种 MTM3D 模型用于医学图像分类任务, 该模型结合了 Mamba 模型在复杂序列任务上的优异性能与改进令牌图灵机 (TTM) 网络的外部记忆存储功能。通过引入链式循环存储结构, MTM3D 能够在记忆单元中有效交互不同空间结构的特征, 从而提升对复杂空间关系的捕捉能力; 此外, Mamba 的引入进一步增强了记忆单元与处理单元的交互能力, 使模型具备更强的泛化能力, 在不同的医学图像数据集上表现出色。实验结果表明, MTM3D 在 MedMNIST v2 数据集上的医学图像理解能力表现优异。相比现有最佳的医学图像分析网络, MTM3D 的平均准确率 (ACC) 提升了 3.97%, 平均曲线下面积 (AUC) 提升了 2.00%, 充分展示了其在医学图像解读和协助医疗专业人员进行诊断与治疗规划中的巨大潜力。

关键词: 三维医学图像分类; Mamba; 令牌图灵机; Transformer; 记忆存储结构

中图分类号: TP391

文献标志码: A

DOI: 10.19678/j.issn.1000-3428.0070764

MTM3D: 3D Medical Image Analysis Model Integrating Mamba and Improved TTM

YANG Yang¹, WEI Hongkai¹, SUN Shijie², HU Hongli¹, WANG Rong¹, WANG Tiantian¹

(1. School of Information Engineering, Chang'an University, Xi'an 710064, Shaanxi, China;

2. School of Data Science and Artificial Intelligence, Chang'an University, Xi'an 710064, Shaanxi, China)

【Abstract】 Biomedical imaging is crucial in the diagnosis and treatment of various diseases. The application of deep learning methods to medical image analysis can enhance the readability of medical images and provide more reliable support for clinical decision-making. However, traditional medical image processing methods are limited in effectively capturing spatial features and complex structural information in 3D images, especially when handling complex 3D medical images generated by different imaging modalities. This often challenges the accuracy and generalization ability of the model. To address this challenge, an MTM3D model is proposed for medical image classification tasks. This model combines the excellent performance of the Mamba model in complex sequential tasks with the external memory storage function of the improved Token Turning Machine (TTM) network. By introducing a cyclic chain storage structure, MTM3D enables effective interaction of features from different spatial structures within memory units, thus enhancing its ability to capture complex spatial relationships. Furthermore, the incorporation of Mamba further strengthens the interaction between the memory and processing units, strengthening the generalization capability of the model and enhancing its performance across different medical imaging datasets. Experimental results on the MedMNIST v2 dataset demonstrate that MTM3D exhibits outstanding capabilities in understanding medical images. Compared with the current best medical image analysis networks, MTM3D improves the average Accuracy (ACC) by 3.97% and the average Area Under the Curve (AUC) by 2.00%, fully showcasing its tremendous potential in interpreting medical images and assisting healthcare professionals in diagnosis and treatment planning.

【Key words】 3D medical image classification; Mamba; Token Turing Machine (TTM); Transformer; memory storage structure

基金项目: 国家重点研发计划 (2023YFB4301800)。

作者简介: 杨洋, 男, 硕士研究生, 主研方向为模式识别、计算机视觉; 魏弘凯, 博士研究生; 孙士杰 (通信作者), 副教授、博士; 胡红利、王荣、王天添, 硕士研究生。

收稿日期: 2024-12-31

修回日期: 2025-03-26

E-mail: shijiesun@chd.edu.cn

0 引言

在医学技术迅猛发展的背景下,医学影像诊断^[1]已成为疾病治疗的核心手段,在早期诊断中发挥着关键作用。通过计算机断层扫描(CT)、X 射线等影像技术,医生能够及时识别肿瘤、血块、骨折等病变^[2-4],为疾病干预提供依据。然而,这些医学影像质量常受分辨率不足、噪声干扰和伪影影响,降低了清晰度并增加了解读难度。此外,医学影像的分析需要扎实的专业知识,这往往需要长时间的学习和实践积累,使其成为高成本的投入。

深度学习的引入为医学影像诊断带来了变革^[5-6]。卷积神经网络等深度学习算法^[7]在疾病识别中展现出巨大潜力^[8],并在某些情况下具备超越人类专家的诊断能力。然而,传统卷积神经网络在处理三维医学图像时,难以充分挖掘空间特征,且易出现过拟合问题^[9],这限制了其泛化能力^[10],给疾病的准确诊断带来了挑战。

在这一背景下,三维医学图像分析成为了一个重要的研究方向。三维图像包含更丰富的空间信息,需要模型具备更强的特征分析能力来捕捉其独特的空间结构特性^[11]。循环神经网络(RNN)、长短期记忆网络(LSTM)、Transformer 等模型^[12]通过捕捉特征序列的关联性,为三维图像分析提供了新视角。此外,神经图灵机(NTM)^[13]和令牌图灵机(TTM)^[14]等模型通过引入外部记忆机制,提升了精度和泛化能力。然而,TTM 因大量使用 Transformer Encoder 作为记忆增强单元,导致计算效率下降,推理性能受限,亟须进一步优化。相比之下,Mamba^[15]模型通过高效的计算性能和线性扩展能力,解决了 Transformer^[12]自注意力机制带来的高计算量问题。

本文的研究动机正是为了解决现有模型在三维医学图像分析中面临的计算效率和泛化能力问题。针对这一挑战,本文提出了一种融合 Mamba 模型和改进 TTM 网络的三维医学图像分析模型 MTM3D,该模型结合了 TTM 的外部记忆存储能力和 Mamba 面对复杂序列的高性能优势。通过改进 TTM 的存储结构,引入循环链式存储,使不同空间结构特征在记忆单元中循环交互,从而捕获更有价值的信息。同时,MTM3D 进一步融合了 Transformer 和 Mamba,增强了记忆单元与处理单元的交互能力,提升了模型的泛化性。

实验结果表明,MTM3D 在 MedMNIST v2 数据集^[16]上平均准确率(ACC)提升了 3.97%,平均

曲线下面积(AUC)提升了 2.00%。这些改进显著增强了模型的分析能力与泛化能力,使其在医学图像分析中表现出更高的准确性和鲁棒性,为医学影像分析领域提供了一个新的解决方案,具有广泛的应用前景。

1 相关工作

本文将相关工作细分为 3 个主要部分,以深入比较 TTM、Transformer 和 Mamba 这 3 种方法在处理特征序列方面的应用。

1.1 TTM

自 RNN 和 LSTM 被应用于医学图像分析以来^[17-19],其记忆处理机制经历了显著演变。这些模型使用循环结构来捕捉时空依赖关系,这对医学图像分割和诊断等任务至关重要。TTM^[14]通过引入外部存储单元实现了重大改进,这一概念已被其他模型^[20-21]采用,以提高它们在医学图像分析领域的能力。

TTM 是一种具有外部记忆单元的顺序自回归 Transformer 模型,能够保持恒定的计算复杂度来处理特征序列。该模型由外部记忆单元和记忆增强处理单元组成,每个组件都使用令牌(token)^[22-23]作为数据处理单位。

尽管 TTM 在处理复杂空间序列特征时表现优异,但其顺序存储结构和 Transformer Encoder 的低计算效率限制了其性能。

1.2 Transformer

Transformer 模型推动了自然语言处理领域的发展。Transformer 采用了编码器-解码器架构^[24-25],每层包括自注意力机制,能够捕捉输入序列中的全局信息,而非像 RNN 那样仅依赖前一个状态进行信息传递。在 Transformer 中,每个位置的词编码可以并行计算,突破了 RNN 序列化处理的局限性,使得模型在处理特征序列时具有更大优势。

然而,Transformer 中自注意力机制的计算量随着上下文长度的增加呈平方级增长,这限制了计算效率,并影响了其处理复杂特征序列任务的能力。

1.3 Mamba

Mamba^[15]模型作为一种新型的状态空间模型架构,采用选择性状态空间模型显著改进了传统的状态空间模型。在医学图像处理领域^[26],Mamba 模型的应用正逐渐增多^[27],这显现出其潜力和应用价值^[28]。

Mamba 模型的核心创新体现在其能够根据输入依赖的方式动态地调整状态空间模型参数,实现

信息的选择性传递或遗忘,从而提高了其处理离散型和信息密集型数据的效能与准确性,突破了传统模型的局限。此外,Mamba 模型运用了硬件感知的并行算法,提升了推理速度,相比传统 Transformer,其推理速度提升了 5 倍,计算量在序列长度扩展情况下也表现出线性缩放效果,保证了模型在长序列任务中的卓越性能。

2 网络设计

本文以 TTM 网络为基本框架,融合了 Mamba Block,创新性地提出了一种高效的三维医学图像分析方法。MTM3D 的核心亮点在于其存储结构的创新及 Mamba Block 的引入。

MTM3D 将传统的线性顺序存储结构改进为链式循环存储结构,这一改进不仅优化了特征序列的交互方式,还增强了模型处理复杂特征序列的能力。

MTM3D 的记忆增强处理单元设计为 5 种不同类型,分别为单一 Mamba Block 以及 Transformer Encoder 与 Mamba Block 的 4 种混合/融合模式,其中,MT 混合序变模式(MT-H)与 TM 混合序变模式(TM-H)通过动态调整两者的处理顺序实现了对特征序列的自适应;而 MT 融合序变模式(MT-F)与 TM 融合序变模式(TM-F)则进一步融合了两者,共同作用于特征序列。

2.1 模型结构

本文提出的模型结构如图 1 所示(彩色效果见《计算机工程》官网 HTML 版,下同),其中,记忆存储结构采用了链式循环记忆单元,取代了传统 TTM 模型中的顺序存储结构,然后通过三时态记忆融合机制将 3 个不同时间点的记忆块(过去时刻的 M^{t-1} 、当前时刻的 M^t 和未来时刻的 M^{t+1})进行整合,从而有效提取关键记忆信息。

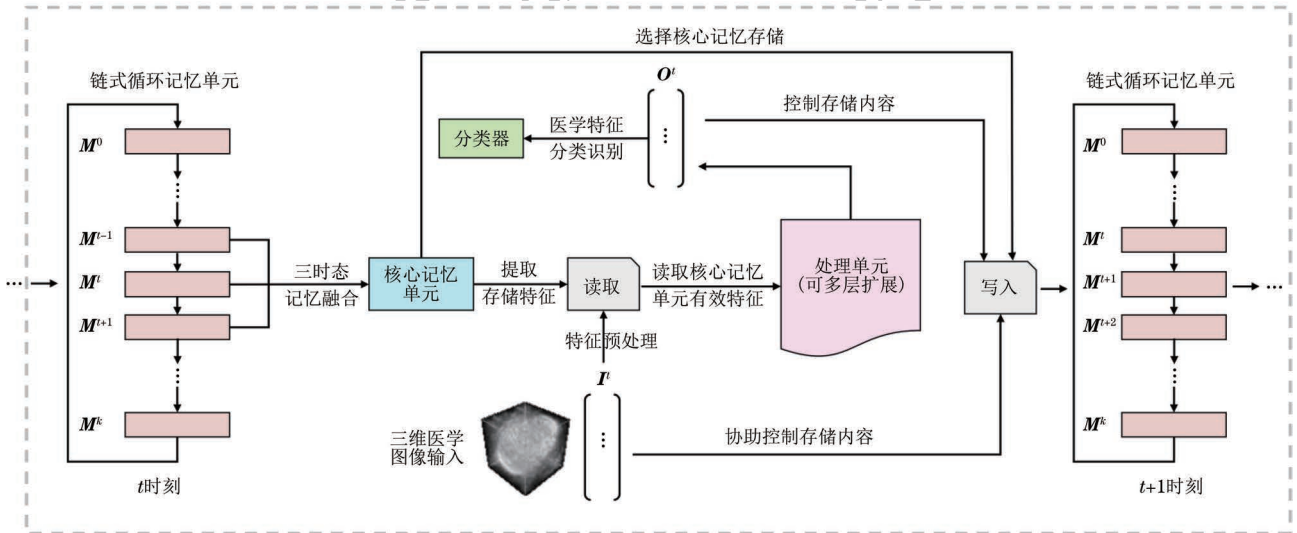


图 1 MTM3D 网络框架

Fig.1 MTM3D model framework

将三维医学图像 I^t 经过特征提取网络得到的特征与核心记忆单元中的有效特征进行融合,再通过记忆增强处理单元进行特征融合处理,从而更好地处理特征空间和上下文的全局关系。经过特征增强处理后得到输出控制向量 O^t , O^t 经过分类头进行最终分类得到分类结果。此外,特征结果也会和主干网络提取的图像特征以及三时态记忆融合得到的特征三者协同写入链式循环存储结构的下一时刻记忆单元,以实现存储结构中特征信息的持续更新和交互。

在记忆增强处理单元中,本文融合了 Mamba Block,并深入探索了 Mamba Block 与 Transformer Encoder 的 4 种混合/融合模式,提出了 5 种新型记忆增强处理单元模式,这些模式的应用提升了记忆

增强处理单元的处理能力。经过记忆增强处理的特征协同主干网络提取的特征序列和三时态记忆融合得到的特征,三者共同写入记忆单元。

此外,为了适应决策任务的需求,MTM3D 在每个输出控制向量后连接一个分类器,这使得模型能够对特征进行精准预测和分类。经过记忆增强处理单元处理的特征通过这个分类器,完成三维医学图像的分类感知任务。

MTM3D 能够综合考虑过去特征、当前特征以及对未来特征的预测,从而在三维医学图像分析中实现更精准的分类和预测。

2.2 链式循环存储结构和读写操作

2.2.1 链式循环存储结构

TTM 采用线性的顺序存储结构[如图 2(a)所

示],而 MTM3D 突破了传统 TTM 顺序存储结构的限制,引入了一种链式循环存储结构[如图 2(b)所示],这种结构将原本连续的存储空间 $M \in \mathbb{R}^{k \times d}$ 划分为多个独立的记忆存储区块,每个记忆存储区块由 k 个 d 维的特征向量组成。这些记忆存储区块通过链式连接,形成了一个循环链式存储结构。

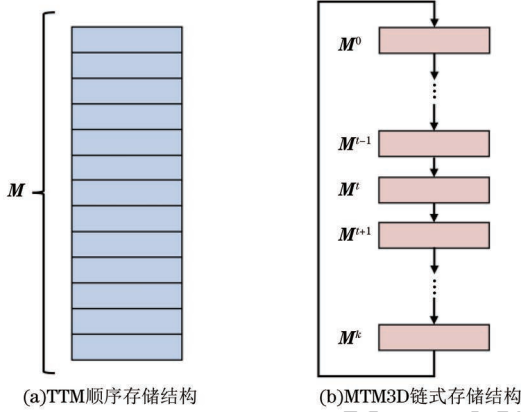


图 2 TTM 与 MTM3D 模型存储结构对比

Fig.2 Comparison of memory structures between TTM and MTM3D models

模型处理第 t 时刻的数据时,循环记忆单元中的第 $t-1$ 、 t 和 $t+1$ 时间步的记忆单元块被读取,将其特征拼接后输入线性变换层进行处理。这一过程有效处理了存储的特征之间的深层次关联性和空间信息,同时过滤和压缩了冗余信息,从而增强了对序列数据的理解和后续建模能力。

MTM3D 的链式循环存储结构允许模型循环访问和更新令牌特征,提高了数据处理效率,并增强了其对复杂空间特征序列的适应性和处理能力。这种设计使得 MTM3D 能够高效处理和利用外部存储的记忆特征,保持对数据动态变化的敏感性,显著提升了其在时空序列建模任务中的表现。

2.2.2 读写操作

MTM3D 的读写过程中运用了一维卷积,可以动态地将标记信息融合到循环链式记忆单元中。首先对输入数据 I 执行归一化,得到 I' 。随后,利用一维卷积和 GELU 激活函数来计算注意力权重矩阵 W 。为了确保权重矩阵的元素之和为 1,采用 Softmax 函数,得到归一化的权重矩阵 W' 。然后,将输入数据 I 通过一个令牌卷积层进行处理,以提取出令牌并构成令牌矩阵 F 。最终,将归一化的注意力权重张量 W' 与令牌矩阵 F 相乘,并借助 einsum 运算,对索引 k 进行求和进而实现加权聚合,从而得到综合的标记表示 Z ,如式(1)所示:

$$Z_{ij} = \sum_{k=1}^n W'_{ij} F_{kj} \quad (1)$$

式中: n 表示求和维度,它明确了求和操作的索引。通过这种方式,可将注意力机制的权重与标记矩阵中的元素相结合,为每个标记生成一种新的表示形式,这不仅增强了模型对特征的捕捉能力,也为后续的分析 and 处理提供了更为丰富和精准的信息。

2.3 融合 Mamba Block 的记忆增强处理单元

2.3.1 单一 Mamba Block 处理模式

MTM3D 模型通过记忆增强处理单元,实现了对循环链式存储单元中特征的高效提取,配合自适应的读写机制,确保了特征读写的精确性。这一系列过程不仅优化了特征的提取与存储,且通过动态调整读写操作,显著提升了模型在处理三维医学图像分类任务时的准确性。

在原有的 TTM 模型中,Transformer Encoder 作为其记忆增强处理单元,如图 3(a)所示,这种完全依赖 Transformer Encoder 的方式由于引入了自注意力机制导致较高的计算复杂度,在处理复杂数据时往往会导致计算量急剧增加。在本文提出的 MTM3D 中,第一种记忆增强处理单元模式采用了 Mamba Block 作为替代,如图 3(b)所示。这一改进增强了模型处理复杂空间结构信息的能力。不同于依赖多头注意力机制的 Transformer Encoder, Mamba Block 的设计基于状态空间模型。状态空间模型框架受到了连续系统理论的启发,提供了一种更为高效的方法来处理序列数据中的长期依赖性和动态变化。通过采用这种设计, Mamba Block 能够更有效地捕捉数据中的复杂模式。

在连续系统中,通过隐藏状态 $h(t) \in \mathbb{R}^N$,能够将一维函数或序列进行映射: $x(t) \in \mathbb{R} \rightarrow y(t) \in \mathbb{R}$ 。这一过程由演化参数 $A \in \mathbb{R}^{N \times N}$ 控制,同时使用 $B \in \mathbb{R}^{N \times 1}$ 和 $C \in \mathbb{R}^{1 \times N}$ 作为投影参数。连续系统的过程如式(2)所示:

$$h'(t) = Ah(t) + Bx(t), y(t) = Ch'(t) \quad (2)$$

Mamba 模型作为连续系统的离散版本,引入了时标参数 Δ 来将连续参数 A 和 B 转换为离散参数 $\bar{A}$ 和 $\bar{B}$,这一转换通常采用零阶保持(ZOH)方法,定义如下:

$$\bar{A} = \exp(\Delta A), \bar{B} = (\Delta A)^{-1} (\exp(\Delta A) - I) \cdot \Delta B \quad (3)$$

完成参数 $\bar{A}$ 和 $\bar{B}$ 的离散化后,式(2)的离散形式可以通过式(4)表示(其中步长为 Δ):

$$h_t = \bar{A}h_{t-1} + \bar{B}x_t, y_t = Ch_t \quad (4)$$

最终,模型通过全局卷积计算输出,如式(5)所示:

$$\bar{K} = (C\bar{B}, C\bar{A}\bar{B}, \dots, C\bar{A}^{M-1}\bar{B}), y = x * \bar{K} \quad (5)$$

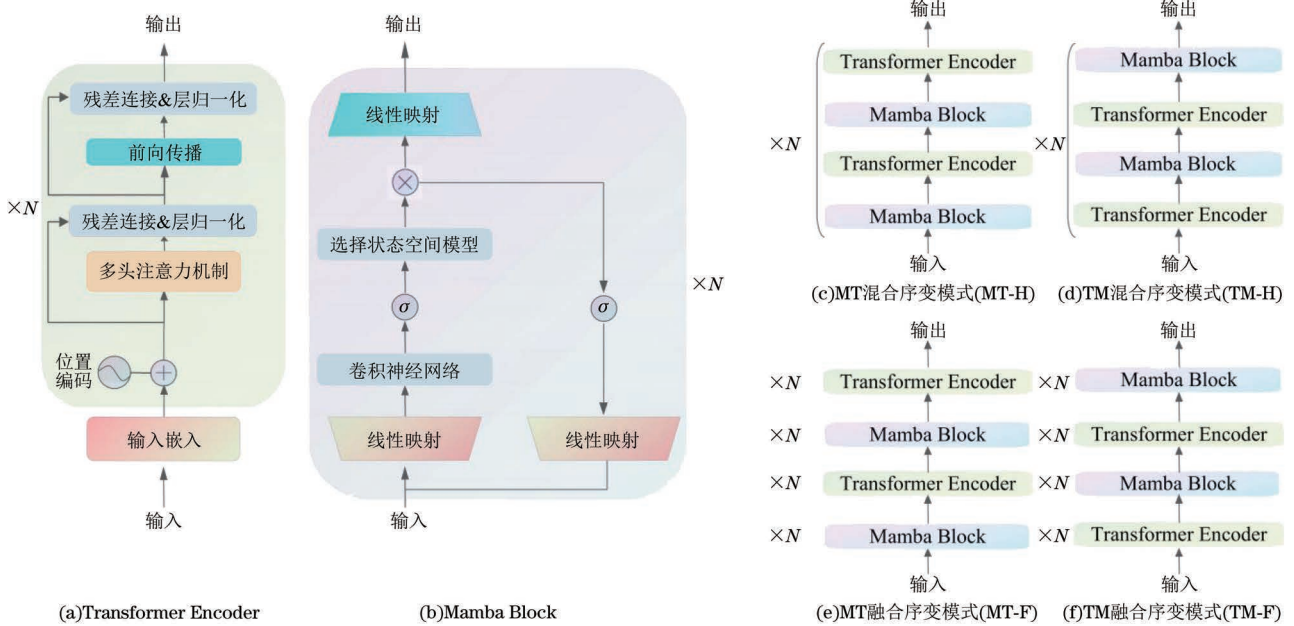


图 3 MTM3D 5 类记忆增强单元

Fig.3 Five types of memory-enhanced units in MTM3D

式中:输入序列 M 的长度为 x ; K 是一个结构化卷积核。

这种设计保持了模型对时空特征序列的敏感性,为三维医学图像分析提供了更精准的特征提取能力。

2.3.2 混合序列变换模式

Mamba 模型与 Transformer 模型各自具备独特的优势,为了充分发挥这两种模型的长处,本文在 MTM3D 模型中设计了 Mamba Block 与 Transformer Encoder 两种混合序列变换模式——MT 混合序变模式(MT-H)和 TM 混合序变模式(TM-H),如图 3(c)和图 3(d)所示。

MT 混合序变模式首先将读取到的特征送入 N 层混合层中进行处理。每一层混合层首先利用一个 Mamba Block 对特征进行初步提取和状态空间建模,随后由一个 Transformer Encoder 进一步整合空间信息和变换特征,从而生成更具表达力的特征表示。

TM 混合序变模式采取了相反的处理策略。该模式同样包含 N 层混合层,但每一层中 Transformer Encoder 位于 Mamba Block 之前。首先利用 Transformer Encoder 对输入特征进行空间信息的初步整合,随后再通过 Mamba Block 进行深度特征提取和融合,以强化模型对复杂数据模式的捕捉能力。

2.3.3 融合序列变换模式

在进一步探索 MTM3D 中 Mamba Block 与

Transformer 的序列变换时,本文还研究了另外两种有效的序列融合变换模式——MT 融合序变模式(MT-F)和 TM 融合序变模式(TM-F),如图 3(e)和图 3(f)所示。

MT 融合序变模式采用渐进式融合策略。首先通过 N 层 Mamba Block 对输入特征进行高效的提取和初步融合,这一过程充分利用了 Mamba Block 在特征提取和状态空间建模方面的优势。随后,这些经过初步处理的特征被送入 N 层 Transformer Encoder,进行更深层次的空间信息整合和变换。Transformer Encoder 的引入,凭借其自注意力机制和并行处理能力,进一步提升了特征之间的关联性,从而实现了更优的信息融合效果。

TM 融合序变模式采取了与之相反的策略。首先,利用 N 层 Transformer Encoder 对输入特征进行初步的空间信息捕获和整合,发挥 Transformer Encoder 在处理长距离依赖关系和空间信息方面的优势。然后,将这些初步整合的特征传递给 N 层 Mamba Block,通过状态空间模型框架进行深度融合,以强化模型对复杂数据模式的捕捉能力。

3 实验

本文将 MTM3D 模型衔接到 3D 医学图像分类的下游任务中,以评测模型的性能。在 MedMNIST v2 中的 6 个三维医学图像分类数据集上进行评测,与其他模型相比,MTM3D 的整体性能表现更出色。

3.1 实验数据及评估标准

本实验采用 MedMNIST v2 数据集,这是一个范围广泛的标准化生物学医学图像集合。它包括不同类型的医学影像,如 CT 图像、磁共振血管造影(MRA)图像和电子显微镜(EM)图像。在这些不同类型的数据集中,CT 图像提供高分辨率的结构信息,要求模型具备较强的细节提取能力和噪声容忍度,特别是在应对可能存在的伪影和数据缺失时;MRA 图像则主要用于血管结构的识别,其软组织

对比度和纹理特征要求模型具备较强的特征区分能力;而 EM 图像则为超高分辨率的细胞与亚细胞结构图像,模型必须能够捕捉其中微小细节并聚合局部信息。本文通过在这些不同类型的 3D 医学图像上进行分类任务以证明模型在面对多样化医学成像方式时的适应性与泛化能力。

数据集包括 6 个具备时空特征的三维医学图像分类数据集,可视化如图 4 所示,详细介绍如表 1 所示。

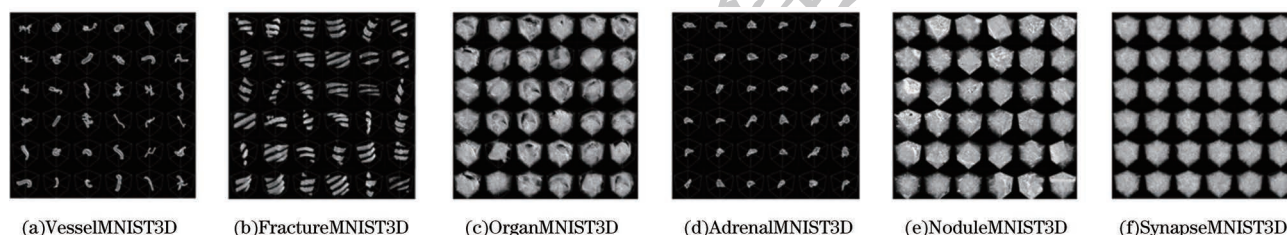


图 4 MedMNIST v2 数据集可视化

Fig. 4 Visualization of the MedMNIST v2 dataset

表 1 MedMNIST v2 数据集详细介绍

Table 1 Detailed introduction to the MedMNIST v2 dataset

单位:个

数据集	类型	来源	任务	样本数	训练/验证/测试数
OrganMNIST3D	器官数据	腹部 CT	多分类(11)	1 743	972/161/160
NoduleMNIST3D	肺结节数据	胸部 CT	二分类(2)	1 633	1 158/165/310
AdrenalMNIST3D	肾上腺数据	腹部 CT	二分类(2)	1 584	1 188/98/298
FractureMNIST3D	肋骨骨折数据	胸部 CT	多分类(3)	1 370	1 027/103/240
VesselMNIST3D	血管动脉瘤数据	脑部 MRA	二分类(2)	1 909	1 335/192/382
SynapseMNIST3D	神经突触数据	EM	二分类(2)	1 759	1 230/177/352

在 MedMNIST v2 数据集上,采用准确率(ACC)和 ROC 曲线下面积(AUC)来评估模型,其中 ACC 是推理过程中正确样本数量与推理样本总数的比值,而 AUC 用于更好地衡量模型的预测能力。此外,为了保持数据的一致性,采用官方的评估脚本进行评估。

3.2 实验细节

MTM3D 模型在 MedMNIST v2 中的 6 个三维医学图像分类数据集上进行训练和推理,将结果与已公布的在该数据集上性能较好的模型与部分主流模型进行比较。

在消融实验部分,将 TTM 网络中的 Transformer Encoder 结构分别替换为 Mamba Block、MT 混合序变、MT 融合序变、TM 混合序变以及 TM 融合序变结构,替换对应模块后的模型再分别进行训练和推理;同时,记忆单元结构由原来的顺序存储结构替换为本文提出的链式循环存储结构,其他实验参数均保持一致,然后在部分数据集上进行训练测试。

本文还对不同数据集上达到最佳推理性能的 MTM3D 模型的记忆增强处理单元参数配置进行研究。此外,通过数据集加噪对比实验来验证模型泛化能力的提升。

实验在 4 张 12 GB 的 NVIDIA TITAN Xp 显卡上进行,使用 PyTorch 2.1.0 和 Python 3.8 环境。经过前期初步验证,训练超参数选择了 Epoch 为 50, Batch Size 为 32。对于所有数据集,优化器采用 Adam,学习率设置为 0.001。所有不同参数配置的 MTM3D 模型中,Dropout 率统一设定为 0.2。

3.3 结果分析

MTM3D 模型在 6 个 MedMNIST v2 子数据集上与其他模型的性能测试结果如表 2 所示,表中加粗值表示该数据集上的最佳指标值,下划线值表示指标排行第二,下同。结果表明:在 OrganMNIST3D、AdrenalMNIST3D 和 VesselMNIST3D 数据集上,MTM3D 的 AUC 和 ACC 指标均达到了最优(SOTA)水平;在 FractureMNIST3D 数据集上,MTM3D 的 ACC 较现有最佳模型 Swin

Transformer 提升了 2.79%;在 AdrenalMNIST3D 数据集上,ACC 相较于 Swin Transformer 也提升了 6.77%;同样,MTM3D 在其他数据集上也表现优异。

表 2 MTM3D 与其他模型性能对比

Table 2 Performance comparison of MTM3D and other models													
方法	参数数量/10 ⁶	OrganMNIST3D		NoduleMNIST3D		FractureMNIST3D		AdrenalMNIST3D		VesselMNIST3D		SynapseMNIST3D	
		AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
ResNet-18+2.5D ^[29]	11	0.977	0.788	0.838	0.835	0.587	0.451	0.718	0.772	0.748	0.846	0.634	0.696
ResNet-18+3D ^[29]	33	<u>0.996</u>	<u>0.907</u>	0.863	0.844	0.712	0.508	0.827	0.721	0.874	0.877	0.820	<u>0.745</u>
ResNet-50+2.5D ^[29]	15	0.994	0.900	0.873	0.847	<u>0.714</u>	<u>0.497</u>	0.839	0.754	<u>0.930</u>	0.928	0.705	0.722
ResNet-50+3D ^[29]	44	0.974	0.769	0.835	0.848	<u>0.552</u>	0.397	0.732	0.763	0.751	0.877	0.669	0.735
Auto-sklearn ^[30]	14	0.994	0.889	0.886	0.841	0.750	0.517	0.828	0.758	0.912	0.858	0.719	0.709
AutoKeras ^[31]	12	0.977	0.814	0.914	<u>0.874</u>	0.628	0.453	0.828	0.802	0.910	0.915	0.631	0.730
FPVT ^[32]	15	0.979	0.804	0.844	<u>0.834</u>	0.642	0.458	0.804	0.705	0.773	<u>0.894</u>	0.538	0.724
ACS-Conv ^[33]	11	0.923	0.800	0.814	0.822	0.640	0.438	0.801	0.704	<u>0.770</u>	0.888	0.530	0.712
ViT-B/16 ^[34]	86	0.983	0.874	0.879	<u>0.863</u>	0.683	0.569	0.837	0.807	<u>0.877</u>	0.932	0.753	0.735
CLIP ^[35]	151	0.975	<u>0.832</u>	0.827	0.829	0.654	0.508	0.826	0.793	0.892	0.879	0.683	0.729
Swin Transformer ^[36]	28	0.991	0.905	0.892	0.871	0.689	<u>0.573</u>	<u>0.852</u>	<u>0.812</u>	0.895	<u>0.935</u>	<u>0.768</u>	0.736
MTM3D	10	0.998	0.950	<u>0.906</u>	0.875	0.703	0.589	0.890	0.867	0.935	0.944	0.765	0.797

除了性能指标对比分析外,本文还可可视化了模型 MTM3D 在部分数据集的推理结果,如图 5 所示。

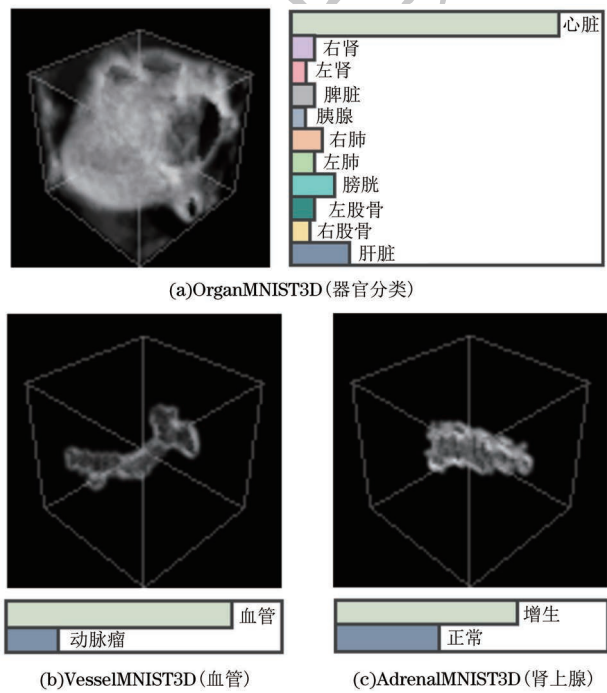


图 5 推理结果可视化

Fig.5 Visualization of inference results

此外,模型在 6 个数据集上的平均 AUC 和 ACC 指标如表 3 所示。与现有最佳模型相比,MTM3D 在平均 ACC 和 AUC 上均表现更优,尤其是在 ACC 指标上,MTM3D 相比现有最佳模型 Swin Transformer 性能提升幅度达 3.97%,AUC 超出现有最佳模型达 2.00%,这一结果充分体现了 MTM3D 模型相对于现有模型的优越性能。

表 3 模型在多数数据集上的指标均值

Table 3 Average metrics of multiple datasets		
方法	模型评估平均值	
	AUC	ACC
ResNet-18+2.5D ^[29]	0.750	0.731
ResNet-18+3D ^[29]	<u>0.849</u>	0.767
ResNet-50+2.5D ^[29]	0.843	0.775
ResNet-50+3D ^[29]	0.752	0.732
auto-sklearn ^[30]	0.848	0.762
AutoKeras ^[31]	0.815	0.765
FPVT ^[32]	0.763	0.737
ACS-Conv ^[33]	0.746	0.727
ViT-B/16 ^[34]	0.835	0.797
CLIP ^[35]	0.810	0.762
Swin Transformer ^[36]	0.848	<u>0.805</u>
MTM3D	0.866	0.837

MTM3D 结合了 TTM 与外部存储模块的交互能力和 Mamba 高效的序列处理能力,能够在处理三维医学图像的复杂分析任务时保持卓越的表现。模型通过动态学习和记忆机制,能够提取和存储医学图像中的关键特征,并在需要时快速检索和利用这些特征,从而提高预测和诊断的精度。因此,MTM3D 在医学图像分析中能够更好地学习和存储图像特征,尤其在多种医学问题的任务中表现优异。通过将 Transformer Encoder 替换为 5 种记忆增强处理单元模式,模型能够高效地与外部存储模块进行特征查询和交互。

3.4 消融实验

本部分通过探索不同的记忆增强处理单元模块

和记忆单元层数对于模型性能的影响,证明记忆增强处理单元模块的有效性,同时找出最优的模型结构和参数配置,以超越现有 SOTA 方法。此外,通过加噪对比实验来验证 Mamba Block 对泛化能力的提升影响。

3.4.1 记忆增强处理单元和存储结构的影响

记忆增强处理单元作为信息处理与存储的核心模块,对模型的学习和泛化能力具有重要影响。为验证不同记忆处理单元对模型推理和泛化能力的作用,

本文将原 TTM 的记忆处理单元替换为 Mamba Block 及图 3 中的 4 种混合/融合结构,并进行对比实验;同时,将存储结构也替换为链式循环存储结构,以考察其对模型性能的影响。

实验结果如表 4 和表 5 所示,表中实验基线采用 Transformer 作为记忆处理单元,顺序存储作为记忆模块存储结构;其他改进的处理单元和存储结构与之对比得到不同数据集的最佳参数配置。

表 4 不同记忆处理机制参数配置的性能表现(ACC)

Table 4 Performance of different memory processing mechanisms with varying parameter configurations (ACC)

存储结构	FractureMNIST3D						AdrenalMNIST3D					
	MTM3D						MTM3D					
	Transformer	Mamba Block	MT-H	MT-F	TM-H	TM-F	Transformer	Mamba Block	MT-H	MT-F	TM-H	TM-F
顺序	0.546	0.573	0.583	0.562	0.562	0.568	0.848	0.848	0.867	0.844	0.859	0.863
链式循环	0.547	0.557	0.583	0.589	0.557	0.568	0.852	0.855	0.855	0.856	0.848	0.852

存储结构	VesselMNIST3D						SynapseMNIST3D					
	MTM3D						MTM3D					
	Transformer	Mamba Block	MT-H	MT-F	TM-H	TM-F	Transformer	Mamba Block	MT-H	MT-F	TM-H	TM-F
顺序	0.929	0.938	0.944	0.938	0.934	0.938	0.781	0.781	0.775	0.791	0.788	0.784
链式循环	0.931	0.934	0.938	0.934	0.941	0.934	0.769	0.784	0.788	0.775	0.788	0.797

表 5 不同记忆处理机制参数配置的性能表现(AUC)

Table 5 Performance of different memory processing mechanisms with varying parameter configurations (AUC)

存储结构	FractureMNIST3D						AdrenalMNIST3D					
	MTM3D						MTM3D					
	Transformer	Mamba Block	MT-H	MT-F	TM-H	TM-F	Transformer	Mamba Block	MT-H	MT-F	TM-H	TM-F
顺序	0.662	0.669	0.684	0.677	0.675	0.658	0.881	0.872	0.884	0.889	0.883	0.879
链式循环	0.669	0.687	0.680	0.692	0.696	0.703	0.863	0.876	0.874	0.867	0.882	0.890

存储结构	VesselMNIST3D						SynapseMNIST3D					
	MTM3D						MTM3D					
	Transformer	Mamba Block	MT-H	MT-F	TM-H	TM-F	Transformer	Mamba Block	MT-H	MT-F	TM-H	TM-F
顺序	0.931	0.934	0.935	0.924	0.925	0.924	0.753	0.755	0.763	0.744	0.764	0.736
链式循环	0.911	0.921	0.910	0.910	0.918	0.935	0.751	0.740	0.754	0.765	0.760	0.750

当 Mamba Block 作为记忆增强处理单元时,模型推理效果显著优于原 TTM 中使用 Transformer 的情况;而采用 MTM3D 的 4 种混合/融合模式时,其推理效果又明显优于 Mamba Block。实验结果表明,引入 Mamba Block 使得模型的性能得到了提升。特别是在不同类型的数据集上,MTM3D 模型的推理效果普遍优于 TTM,表明引入 Mamba Block 能够显著提升模型的泛化能力和适应能力。

在 FractureMNIST3D 数据集上,MT 融合序变

模式结合链式循环存储结构的 ACC 超出基线 7.88%,TM 融合序变模式与链式循环存储结构的 AUC 超出基线 6.19%。在 AdrenalMNIST3D 数据集上,MT 混合序变结合顺序存储结构的 ACC 超出基线 2.24%,TM 融合序变模式结合链式循环存储结构的 AUC 超出基线 3.13%。这些实验结果表明,Mamba 不仅提升了推理能力,还显著提高了模型在不同数据集上的泛化能力,相比传统的 Transformer 模型,MTM3D 展现了更强的自适应学习能力。

3.4.2 记忆增强处理单元层数的影响

通过研究不同记忆单元层数对模型推理和泛化能力的影响,笔者发现层数对模型性能具有显著影响。通常,增加层数可以增加模型的学习深度,但过多层数可能导致训练困难和过拟合。因此,选择最佳层数对优化模型性能至关重要。

经过实验,本文确定了模型在 MedMNIST v2 数据集上达到最佳性能的参数配置,结果如表 6 所示。实验结果显示,在不同任务中,模型实现最佳性

能的参数配置具有多样性,证明了记忆单元层数对网络泛化和推理性能的影响。

在 OrganMNIST3D 数据集中,采用 12 层 MT 混合序变记忆处理单元和链式循环存储结构时,模型的 ACC 值达到 0.950,AUC 值高达 0.998,表现出良好的分类性能和预测不确定性处理能力。在其他数据集上也呈现类似规律,证明了合理选择记忆单元层数对提升 MTM3D 模型性能的重要性,其能强化模型在不同任务和数据集上的泛化能力和推理能力。

表 6 不同数据集对应模型的最佳参数

Table 6 Optimal parameters for model on different datasets

配置	OrganMNIST3D	NoduleMNIST3D	FractureMNIST3D	AdrenalMNIST3D	VesselMNIST3D	SynapseMNIST3D
存储结构	链式循环	链式循环	链式循环	顺序	顺序	链式循环
记忆处理单元类型	MT-H	Mamba	MT-F	MT-H	MT-H	TM-F
记忆处理单元层数/层	12	6	3	3	6	6
ACC	0.950	0.875	0.589	0.867	0.944	0.797
AUC	0.998	0.906	0.692	0.884	0.935	0.750

3.4.3 Mamba Block 分离消融实验

为分析 Mamba Block 的具体贡献,基于不同数据集的最佳配置参数,对模型中的 Mamba Block 进行分离消融实验。实验中,将表 6 中各数据集对应最佳记忆处理单元中的 Mamba Block 替换为 Transformer Encoder,其他参数配置保持不变。结

果如表 7 所示,表中,×表示将所有 Mamba Block 分离,将其替换为 Transformer Encoder,√表示不分离 Mamba Block。实验结果表明,引入 Mamba Block 的模型指标均有所提升,证明 Mamba Block 在提升模型对复杂三维医学图像理解能力方面具有关键作用。

表 7 Mamba Block 分离消融实验结果

Table 7 Results of Mamba Block ablation experiment

Mamba Block	指标	OrganMNIST3D	NoduleMNIST3D	FractureMNIST3D	AdrenalMNIST3D	VesselMNIST3D	SynapseMNIST3D
×	ACC	0.923	0.865	0.553	0.832	0.917	0.776
	AUC	0.983	0.882	0.689	0.876	0.929	0.743
√	ACC	0.950	0.875	0.589	0.867	0.944	0.797
	AUC	0.998	0.906	0.692	0.884	0.935	0.750

3.4.4 模型泛化性证明

为进一步验证 Mamba Block 对模型泛化能力的提升作用,在 AdrenalMNIST3D 数据集上进行了加噪实验以评估模型在处理带噪声输入数据时的鲁棒性与泛化能力。通过对比 TTM 和 MTM3D 模型的表现来证明在相同噪声干扰条件下,MTM3D 相较于 TTM 在泛化能力和鲁棒性方面具有更显著的优势。

实验结果如表 8 所示,其中,MTM3D 采用 TM 混合序变记忆处理单元和顺序存储结构。TTM 在加噪情况下,ACC 下降了 4.58%,而 MTM3D 仅下降了 3.23%。这说明引入 Mamba 使得模型的泛化能力得到显著提升,能够更好地适应噪声干扰,保持较高的性能,从而增强了模型在不同任务和数据集上的鲁棒性与适应性。

表 8 加噪对比实验结果

Table 8 Results of noise perturbation comparison experiment

加噪状态	MTM3D		TTM	
	ACC	AUC	ACC	AUC
不加噪	0.867	0.890	0.852	0.881
加噪	0.839	0.853	0.813	0.824

4 结束语

本文提出了 MTM3D 模型,这是一个基于 Mamba 和改进时空记忆结构 TTM 网络构建的三维医学图像分析模型。MTM3D 模型结合了 TTM 的外部记忆存储单元和 Mamba 模型在复杂序列任务中的优异性能,同时对存储结构进行了创新性改进,引入了链式循环存储结构,显著提升了模型的可扩展性和泛化能力。在 MedMNIST v2 数据集上的

实验结果证明了 MTM3D 在三维医学图像分类方面的卓越性能,凸显了其有效处理复杂三维医学图像的能力。此外,消融实验探究了不同记忆处理单元和存储结构对模型效能的影响,证明了 MTM3D 的有效性、灵活性和适应性。未来将继续探索 MTM3D 模型在更多医学图像分析任务中的应用,并进一步优化模型结构和参数配置,以实现更高的性能和更好的泛化能力。

参考文献

- [1] WEBB A. Introduction to biomedical imaging[M]. [S. l.]: John Wiley & Sons, 2022.
- [2] DOU Q, CHEN H, YU L Q, et al. Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks[J]. IEEE Transactions on Medical Imaging, 2016, 35(5): 1182-1195.
- [3] RASAL R, CASTRO D C, PAWLOWSKI N, et al. Deep structural causal shape models[C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2023: 400-432.
- [4] LIU F Z, CAI J Z, HUO Y K, et al. JSSR: a joint synthesis, segmentation, and registration system for 3D multi-modal image alignment of large-scale pathological CT scans[C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2020: 257-274.
- [5] 李翠云, 白静, 郑凉. 融合边缘增强注意力机制和 U-Net 网络的医学图像分割[J]. 图学学报, 2022, 43(2): 273-278.
LI C Y, BAI J, ZHENG L. A U-Net based contour enhanced attention for medical image segmentation [J]. Journal of Graphics, 2022, 43(2): 273-278. (in Chinese)
- [6] 张淑军, 彭中, 李辉. SAU-Net: 基于 U-Net 和自注意力机制的医学图像分割方法[J]. 电子学报, 2022, 50(10): 2433-2442.
ZHANG S J, PENG Z, LI H. SAU-Net: medical image segmentation method based on U-Net and self-attention[J]. Acta Electronica Sinica, 2022, 50(10): 2433-2442. (in Chinese)
- [7] 黄帅帅, 张清华, 靳朋仁, 等. 基于权重生成图卷积神经网络的链接预测方法[J]. 重庆邮电大学学报(自然科学版), 2025, 37(3): 375-386.
HUANG S S, ZHANG Q H, JIN P R, et al. Link prediction method based on weighted-generator graph convolutional networks[J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2025, 37(3): 375-386. (in Chinese)
- [8] 谢娟英, 张凯云. SOSNet: 一种非对称编码器-解码器结构的非小细胞肺癌 CT 图像分割模型[J]. 电子学报, 2024, 52(3): 824-837.
XIE J Y, ZHANG K Y. SOSNet: an asymmetric encoder-decoder structure model for automatic segmenting non-small cell lung cancer CT images [J]. Acta Electronica Sinica, 2024, 52(3): 824-837. (in Chinese)
- [9] RAZZAK M I, NAZ S, ZAIB A. Deep learning for medical image processing: overview, challenges and the future[M]//DEY N, ASHOUR A S, BORRA S, et al. Classification in BioApps: automation of decision making. Berlin, Germany: Springer, 2018: 323-350.
- [10] 王欣雨, 刘慧, 朱积成, 等. 基于高低频特征分解的深度多模态医学图像融合网络[J]. 图学学报, 2024, 45(1): 65-77.
WANG X Y, LIU H, ZHU J C, et al. Deep multimodal medical image fusion network based on high-low frequency feature decomposition[J]. Journal of Graphics, 2024, 45(1): 65-77. (in Chinese)
- [11] SAHA A, TUSHAR F I, FARYNA K, et al. Weakly supervised 3D classification of chest CT using aggregated multi-resolution deep segmentation features[C]//Proceedings of the Medical Imaging 2020: Computer-Aided Diagnosis. Houston, USA: SPIE, 2020: 39-44.
- [12] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York, USA: ACM Press, 2017: 6000-6010.
- [13] GRAVES A, WAYNE G, DANIHELKA I. Neural Turing machines[EB/OL]. [2024-10-12]. <https://arxiv.org/abs/1410.5401>.
- [14] RYOO M S, GOPALAKRISHNAN K, KAHATAPITIYA K, et al. Token Turing Machines[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2023: 19070-19081.
- [15] GU A, DAO T. Mamba: linear-time sequence modeling with selective state spaces [EB/OL]. [2024-10-12]. <https://arxiv.org/abs/2312.00752>.
- [16] YANG J C, SHI R, WEI D L, et al. MedMNIST v2—a large-scale lightweight benchmark for 2D and 3D biomedical image classification[J]. Scientific Data, 2023, 10(1): 41.
- [17] BORE J C, LI P Y, JIANG L, et al. A long short-term memory network for sparse spatiotemporal EEG source imaging[J]. IEEE Transactions on Medical Imaging, 2021, 40(12): 3787-3800.
- [18] CHEN X, LOWERISON M R, DONG Z J, et al. Localization free super-resolution microbubble velocimetry using a long short-term memory neural network[J]. IEEE Transactions on Medical Imaging, 2023, 42(8): 2374-2385.
- [19] ZHANG R, QIN B, ZHAO J, et al. Locating X-ray coronary angiogram keyframes via long short-term spatiotemporal attention with image-to-patch contrastive learning[J]. IEEE Transactions on Medical Imaging, 2024, 43(1): 51-63.
- [20] CHEN Y R, ZHANG H, WANG Y N, et al. MAMA net: multi-scale attention memory autoencoder network for anomaly detection [J]. IEEE Transactions on Medical Imaging, 2021, 40(3): 1032-1041.
- [21] WANG P Y, ZHANG H Q, ZHU M L, et al. MGIML: cancer grading with incomplete radiology-pathology data via memory learning and gradient homogenization [J]. IEEE Transactions on Medical Imaging, 2024, 43(6): 2113-2124.
- [22] RYOO M S, PIERGIOVANNI A, ARNAB A, et al. TokenLearner: what can 8 learned tokens do for images and videos? [EB/OL]. [2024-10-12]. <https://arxiv.org/abs/2106.11297>.
- [23] JAEGLE A, GIMENO F, BROCK A, et al. Perceiver: general perception with iterative attention[C]//Proceedings of International Conference on Machine Learning. [S. l.]: PMLR, 2021: 4651-4664.
- [24] AHMADI N, TSANG M Y, GU A N, et al. Transformer-based spatio-temporal analysis for classification of aortic stenosis severity from echocardiography cine series[J]. IEEE Transactions on Medical Imaging, 2023, 43(1): 366-376.
- [25] YANG Z Y, PAN J J, DAI J, et al. Self-supervised lightweight depth estimation in endoscopy combining CNN and transformer[J]. IEEE Transactions on Medical Imaging, 2024, 43(5): 1934-1944.
- [26] WU R K, LIU Y H, LIANG P C, et al. UltraLight VM-UNet: parallel vision mamba significantly reduces parameters for skin lesion segmentation [EB/OL]. [2024-10-12]. <https://arxiv.org/abs/2403.20035>.
- [27] RUAN J, XIANG S. VM-UNet: vision Mamba UNet for medical image segmentation [EB/OL]. [2024-10-12]. <https://arxiv.org/abs/2402.02491>.

- [28] ZHOU J Y, JIANG M Z, WU J D, et al. MGI: multimodal contrastive pre-training of genomic and medical imaging[EB/OL]. [2024-10-12]. <https://arxiv.org/abs/2406.00631>.
- [29] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE Press, 2016: 770-778.
- [30] FEURER M, KLEIN A, EGGENSPERGER K, et al. Efficient and robust automated machine learning [C] // Proceedings of NeurIPS 2015. [S. l.]: Curran Associates, Inc., 2015: 1-9.
- [31] JIN H F, SONG Q Q, HU X. Auto-keras: an efficient neural architecture search system [C] //Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM Press, 2019: 1946-1956.
- [32] LIU J W, LI Y, CAO G T, et al. Feature pyramid vision transformer for MedMNIST classification decathlon [C] // Proceedings of the International Joint Conference on Neural Networks (IJCNN). Padua, Italy: IEEE Press, 2022: 1-8.
- [33] YANG J C, HUANG X Y, HE Y, et al. Reinventing 2D convolutions for 3D images[J]. IEEE Journal of Biomedical and Health Informatics, 2021, 25(8): 3009-3018.
- [34] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: transformers for image recognition at scale[EB/OL]. [2024-10-12]. <https://arxiv.org/abs/2010.11929>.
- [35] RADFORD A, KIM J W, HALLACY C, et al. Learning transferable visual models from natural language supervision[C]// Proceedings of International Conference on Machine Learning. [S. l.]: PMLR, 2021: 8748-8763.
- [36] LIU Z, LIN Y T, CAO Y, et al. Swin Transformer: hierarchical vision transformer using shifted windows [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada: IEEE Press, 2022: 9992-10002.

文字编辑 金胡考
栏目编辑 赖玉玲