(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2024 年 10 月 24 日 (24.10.2024)

WIPO | PCT

(10) 国际公布号
WO 2024/216672 A1

(51) 国际专利分类号:
G06T 19/20 (2011.01) *G06N 3/0455* (2023.01)
G16H 30/00 (2018.01) *G06V 10/26* (2022.01)
G06N 3/0464 (2023.01)

(21) 国际申请号: PCT/CN2023/091750

(22) 国际申请日: 2023 年 4 月 28 日 (28.04.2023)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202310465972.7 2023年4月20日 (20.04.2023) CN

(71) 申请人: 中国科学院深圳先进技术研究院 **(SHENZHEN INSTITUTES OF ADVANCED TECHNOLOGY CHINESE ACADEMY OF SCIENCES)** [CN/CN]; 中国广东省深圳市南山区深圳大学城学苑大道 1068 号, Guangdong 518055 (CN)。

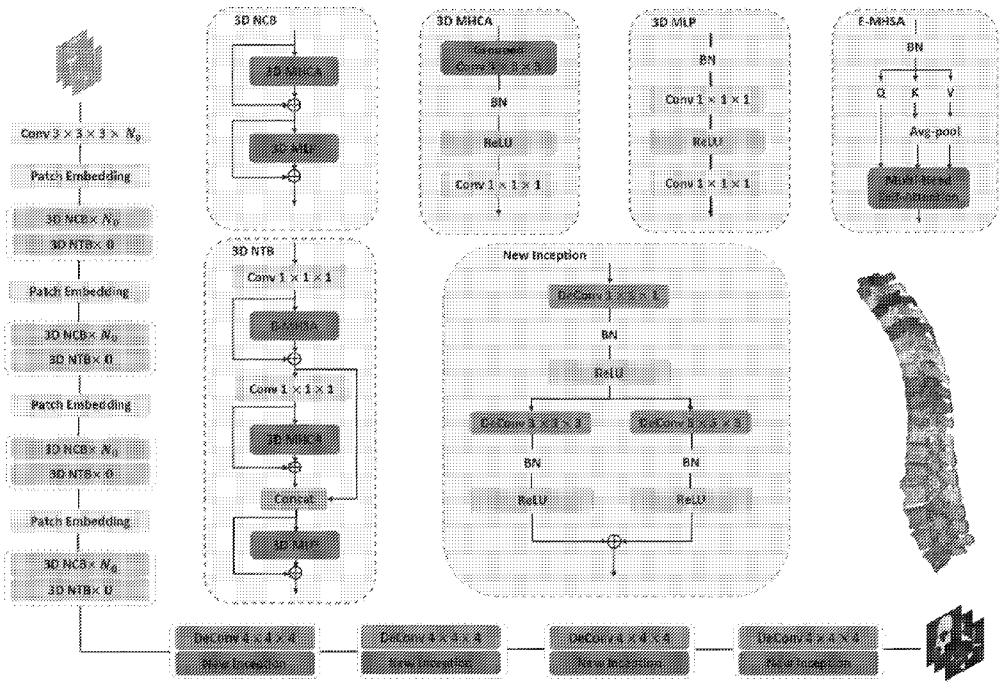
(72) 发明人:胡颖**(HU, Ying)**; 中国广东省深圳市南山区深圳大学城学苑大道1068号, Guangdong 518055 (CN)。 谭志强**(TAN, Zhiqiang)**; 中国广东省深圳市南山区深圳大学城学苑大道1068号, Guangdong 518055 (CN)。 王琼**(WANG, Qiong)**; 中国广东省深圳市南山区深圳大学城学苑大道1068号, Guangdong 518055 (CN)。 赵宇**(ZHAO, Yu)**; 中国广东省深圳市南山区深圳大学城学苑大道1068号, Guangdong 518055 (CN)。 李世博**(LI, Shibo)**; 中国广东省深圳市南山区深圳大学城学苑大道1068号, Guangdong 518055 (CN)。

(74) 代理人: 北京中巡通大知识产权代理有限公司 **(BEIJING ZHONG XUN TONG DA INTELLECTUAL PROPERTY AGENCY CO., LTD.)**;

(54) **Title:** 3D SEGMENTATION METHOD AND SYSTEM BASED ON DIFFUSION MODEL AND 3D TRANSFORMER, DEVICE, AND MEDIUM

(54) 发明名称: 基于扩散模型及3D TRANSFORMER的三维分割方法、系统、设备及介质

[图 1]



(57) **Abstract:** Disclosed are a 3D segmentation method and system based on a diffusion model and a 3D transformer, a device, and a medium. The method comprises: obtaining an image to be segmented; and inputting said image into a 3D segmentation network model based on a diffusion model and a 3D transformer to obtain a 3D segmentation result for said image. By means of the method and system, the device, and the medium, a medical image can undergo 3D segmentation.

[见续页]

WO 2024/216672 A1

中国北京市西城区西直门外大街1号院2号楼7层7C1, Beijing 100044 (CN)。

(81) 指定国(除另有指明，要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国(除另有指明，要求每一种可提供的地区保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:
— 包括国际检索报告(条约第21条(3))。

(57) 摘要: 本发明公开了一种基于扩散模型及3D Transformer的三维分割方法、系统、设备及介质，包括：获取待分割图像；将待分割图像输入到基于扩散模型及3D Transformer的三维分割网络模型中，得到待分割图像的三维分割结果，方法、系统、设备及介质能够对医学影像进行三维分割。

说明书

发明名称：基于扩散模型及3D Transformer的三维分割方法、系统、设备及介质

技术领域

[0001] 本发明属于图像处理技术领域，涉及一种基于扩散模型及3D Transformer的三维分割方法、系统、设备及介质。

背景技术

[0002] 医学影像数据通常是三维数据，如CT、MRI等。具体来说，对我们所关注的脊柱侧弯本身，其就是三维畸形。若从二维的脊柱侧弯图像分割模型去提取出脊柱区域也有两种方式，第一种则是从X光数据进行分割，但分割得到的数据时二维层面的分割图，第二种则是从CT的每一张切片数据进行分割，但是这种方法需要将脊柱侧弯的CT影像数据的所有切片提取出来，然后单独一张张地输入到模型进行训练，在这种情况下会存在一个效率问题。另一个重要的问题则是通过单张切片数据进行训练丢失了CT影像之间的上下文关联关系。因此，当前的研究动态开始着眼于利用三维卷积来直接进行三维分割。

[0003] 医学影像一直都是医学领域重要的分支，其提供给医生病人的组织器官信息，对一些病变情况进行分析诊断。实际上，医学影像分割，一直都是医学领域一个瓶颈问题。从技术角度来说，正常以及病变组织的三维重建、定量分析等操作都是建立在影像分割的基础上，医学影像分割通过将图像中的感兴趣目标区域分割出来，以便为后续的一系列分析性操作打下基础。因此，分割的好坏程度直接影响了后续一系列的操作。从另外一个角度来看，医学影像分割的准确性对医生判断疾病的具体情况并做出诊断以及后续一些影像上的操作具有至关重要的意义。医学图像具有模糊、不均匀、个体差异、复杂多样等特点，也正是由于这个特点使得医学影像的分割显得尤为重要。深度学习技术发展至今，许多领域都依托于深度学习技术得到了长足的发展，医学影像分割也不例外，Ronneberger等人在ISBI cell tracking challenge 2015 使用U-net网络

（Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[J]. Springer, Cham, 2015.）以较大的优势赢得了比赛，并成为医学影像领域广为人知的模型。

[0004] Lee等人（Lee H, Kim J, Lee S, Kim H, Ro Y. Structure Boundary Preserving Segmentation for Medical Image With Ambiguous Boundary[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, 4816-4825.）提出了一种新颖的结构边界保留分割框架。此外，为了将专家知识嵌入全自动分割中，该论文提出了一种新颖的形状边界感知评估器（SBE），其具有专家指示的地面真实结构信息。所提出的SBE可以基于结构边界关键点向分割网络提供反馈。另一方面，医学影像通常难以获取，当前许多学者也致力于研究小样本甚至零样本在医学影像中的应用。Khoi等人（Nguyen K, Todorovic S. FAPIS: A Few-shot Anchor-free Part-based Instance Segmenter[J]. 2021.）提出了一个小样本无锚实例分割框架，通过对训练对象类别之间的共享隐藏对象部分进行明确建模，从而促进测试中新样本的小样本学习，该方法在基准COCO20i数据集上得到很好的评估结果。针对脊柱分割任务，AL Arif等人利用U-net与FCN设计出一个全自动分割颈柱的框架，将颈柱分割任务分成了三个子任务，通过分别制作相应的数据集完成了任务并且达到了很好的效果。并且其在DICE测度下，精度能达到95.7%（Al, Arif, S, et al. Fully automatic cervical vertebrae segmentation framework for X-ray images[J]. Computer Methods and Programs in Biomedicine: An International Journal Devoted to the Development, Implementation and Exchange of Computing Methodology and Software Systems in Biomedical Research and Medical Practice, 2018, 157:95-111.）。Yu等人通过整合不确定性估计来进行半监督学习，通过结合蒙特卡洛法来进行不确定性图构建并指导模型进行逐步学习，可以得到不错的效果（Yu L, Wang S, Li X, et al. Uncertainty-aware Self-ensembling Model for Semi-supervised 3D Left Atrium Segmentation[J]. 2019.）。Sun等人（Lsa B, Cl A, Xd A , et al. Few-shot medical image

segmentation using a global correlation network with discriminative embedding[J]. Computers in Biology and Medicine, 2021, 140.）利用片段式(episodic)训练的深度卷积网络构造了一个小样本图像分割器。基于医学图像的空间一致性和规律性，其开发了一个高效的全局相关模块来捕获支持和查询图像之间的相关性，并将其纳入到全局相关网络中，其实验结果在小样本医学分割中得到了不错的结果验证。

[0005] 在过去的几年时间中，卷积神经网络在医学影像处理中取得了非常大的成就。其中，尤为突出的是基于U型结构的收缩路径和扩展路径以及跳跃连接的深度卷积神经网络在各类医学图像处理任务中得到了广泛的应用。然而，尽管卷积神经网络已经取得了很好的性能，但由于卷积运算的局部性，使得该模型不能很好地学习全局信息以及上下文语义交互信息。目前已有一些学者开始结合Transformer模型（空间变换模型）来进行医学影像分割任务（Chen J, Lu Y, Yu Q, et al. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation[J]. 2021.; Cao H, Wang Y, Chen J, et al. Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation[J]. 2021.）。相比卷积神经网络，Transformer模型位置关联操作不受限，建模能力强，通用性强，可扩展性强，能更好的进行并行运算。另一方面，Transformer模型计算两个位置之间的关联所需的操作次数不随距离增长。而且其中的自注意力机制可以产生更具可解释性的模型，其中注意力机制的各个注意头(attention head)可以学会执行不同的任务，有更强的建模能力。另外，Transformer 可以建立图节点之间的关系，具有较好的通用性：无论节点多么异构，它们之间的关系都可以通过投影到一个可以比较的空间里计算相似度来建立。

[0006] 但从当前的研究现状来看，目前的Transformer模型主要基于二维影像，目前还没有针对三维影像构建的医学影像Transformer模型，不能进行医学影像的三维分割。

发明概述

技术问题

[0007] 本发明的目的在于克服上述现有技术的缺点，提供了一种基于扩散模型及3D Transformer的三维分割方法、系统、设备及介质，该方法、系统、设备及介质能够对医学影像进行三维分割。

技术解决方案

[0008] 为达到上述目的，本发明采用如下技术方案：

[0009] 本发明一方面，本发明提供了一种基于扩散模型及3D Transformer的三维分割方法，包括：

[0010] 获取待分割图像；

[0011] 将所述待分割图像输入到基于扩散模型及3D Transformer的三维分割网络模型中，得到待分割图像的三维分割结果。

[0012] 本发明所述基于扩散模型及3D Transformer的三维分割方法进一步的改进在于：

[0013] 所述基于扩散模型及3D Transformer的三维分割网络模型由CNN-Transformer混合架构、解码器及扩散模型构成。

[0014] 所述CNN-Transformer 混合架构由3D NCB模块及3D NTB模块堆叠而成，其中，通过所述3D NCB模块学习待分割图像中的局部表示；通过所述3D NTB模块捕获轻量级机制中的多频信号。

[0015] 所述3D NCB模块通过3D MHCA模块及3D MLP 模块构建而成。

[0016] 所述解码器基于Inception结构的三维反卷积形式构建而成。

[0017] 所述Inception结构由一个核大小为1×1×1的三维point-wise卷积层及两个核大小为3×3×3和5×5×5的三维反卷积层组成。

[0018] 所述扩散模型通过对语义标签添加随机噪声，并学习语义标签上的随机噪声分量，再对所述语义标签进行去噪，得到相关的语义信息。

[0019] 本发明二方面，本发明提供了一种基于扩散模型及3D Transformer的三维分割系统，包括：

[0020] 获取模块，用于获取待分割图像；

[0021] 分割模块，用于将所述待分割图像输入到基于扩散模型及3D Transformer的三维分割网络模型中，得到待分割图像的三维分割结果。

[0022] 本发明三方面，本发明提供了一种计算机设备，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序，所述处理器执行所述计算机程序时实现所述基于扩散模型及3D Transformer的三维分割方法的步骤。

[0023] 本发明四方面，本发明提供了一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序被处理器执行时实现所述基于扩散模型及3D Transformer的三维分割方法的步骤。

有益效果

[0024] 本发明具有以下有益效果：

[0025] 本发明所述的基于扩散模型及3D Transformer的三维分割方法、系统、设备及介质在具体操作时，采用扩散模型及3D Transformer构建而成的三维分割网络模型进行待分割图像的三维分割，其中，本发明基于三维的Transformer模块用于后续的医学影像三维分割任务，实现医学影像的三维分割，同时通过扩散模型可以很好的捕捉到像素级的语义信息，使得三维分割网络模型能够更好的分割出病灶区域。

附图说明

[0026] 图1为本发明中基于扩散模型及3D Transformer的三维分割网络模型的框架图。

[0027] 图2为本发明的方法框架示意图。

[0028] 图3为本发明中扩散过程的示意图。

[0029] 图4a为仿真实验中真实图像。

[0030] 图4b为本发明的分割效果图。

[0031] 图4c为基于Transformer的分割模型的分割效果图。

[0032] 图4d为基于卷积网络的分割模型的分割效果图。

[0033] 图5为本发明的系统结构图。

本发明的实施方式

[0034] 为了使本技术领域的人员更好地理解本发明方案，下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分的实施例，不是全部的实施例，而并非要限制本发明公开的范围。此外，在以下说明中，省略了对公知结构和技术的描述，以避免不必要的混淆本发明公开的概念。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都应当属于本发明保护的范围。

[0035] 在附图中示出了根据本发明公开实施例的结构示意图。这些图并非是按比例绘制的，其中为了清楚表达的目的，放大了某些细节，并且可能省略了某些细节。图中所示出的各种区域、层的形状及它们之间的相对大小、位置关系仅是示例性的，实际中可能由于制造公差或技术限制而有所偏差，并且本领域技术人员根据实际所需可以另外设计具有不同形状、大小、相对位置的区域/层。

[0036] 实施例一

[0037] 参考图1及图2，本发明所述的基于扩散模型及3D Transformer的三维分割方法包括以下步骤：

[0038] 1）构建基于扩散模型及3D Transformer的三维分割网络模型；

[0039] 步骤1）的具体过程为：

[0040] 本发明将Transformer结构及扩散模型加入到现有的三维分割网络模型中，以形成基于扩散模型及3D Transformer的三维分割网络模型，具体为：

[0041] 所述Tansformer结构改进于Next-ViT结构，遵循分层金字塔网络结构，在每一个阶段都配置一个patch 嵌入层及一系列卷积或 Transformer块，将其三维空间分辨率逐步降低为原来的 1/32，然后将通道维度按不同阶段进行扩展。

[0042] 本发明中通过堆叠3D NCB模块及3D NTB模块构建新型CNN-Transformer 混合架构，需要说明的是，本发明将NCB模块及NTB模块进行重组，并将NCB模块及NTB模块改为三维模块，即3D NCB模块及3D NTB模块，以便模拟脊柱CT序列影像数据中的短期和长期依赖关系，其中，3D NTB模块中还能够进行局部和全局信息的融合，进一步提高建模能力。

[0043] 需要说明的是，所述3D NCB模块通过3D MHCA模块及3D MLP 模块构建而成，所述3D NCB模块能够在保证BottleNeck 块的部署优势的同时，获得作为Transformer 块的突出性能，3D NCB模块遵循Transformer的一般架构。与此同时，本发明采用多头卷积注意力（MHCA）作为具有部署友好卷积操作的高效token 混合器。

[0044] 进一步，3D MHCA模块：为解决现有基于注意力的token混合器具有高延迟的问题，本发明采用新的具有高效卷积操作的注意力机制，即3D卷积注意力(3D CA)，以实现快速推理速度。同时，受 MHSA 中有效多头设计的启发，本发明采用多头范式构建卷积注意力机制，该范式联合关注来自不同位置的不同表示子空间的信息，以实现有效的局部表示学习，所述多头卷积注意力(MHCA)的表示为：

[0045] $$MHCA=Concat\left(CA_1(x_1),CA_2(x_2),\ldots,CA_n(x_n)\right)W^P \quad (1)$$

[0046] 参考图1，在工作时，MHCA从n个并行表示子空间中捕获信息，为促进多个头部之间的信息交互，MHCA配备有一个投影层W^P， $x=[x_1,x_2,\cdots,x_n]$表示将输入特征 x在通道维度上划分为多头形式，3D CA作为一种单头卷积注意力机制可以表示为：

[0047] $$CA(X)=O\left(w_i(T_m,T_n)\right) \quad \text{where } T_{(m,n)}\in x \quad (2)$$

[0048] 其中，T_m及T_n为输入特征 x中相邻的Token，O代表具有可训练参数 w与输入Token之间的内积运算，3D CA可以通过迭代优化可训练参数 w来学习局部感受野中不同Token 之间的关联性。

[0049] 进一步，3D NTB模块：所述3D NCB模块可以有效地学习影像中的局部表示，但影像的全局信息的捕获同样需要解决。Transformer 模块具有较强的捕获低频信号的能力，从而提供全局信息，例如，影像中特征形状和结构。但是Transformer 模块可能会在一定程度上恶化高频信息，例如，局部纹理信息。不同频率段的信号在视觉系统中是必不可少的，并将以特定的方式进行融合，以提取更基本和明显的特征。因此，本发明将Next Transformer Block(NTB)模

块改为3D NTB模块，以捕获轻量级机制中的多频信号。此外，所述3D NTB模块作为多频信号混频器，可以进一步提升整体建模能力，在工作时，所述3D NTB模块采用高效的多头自注意力机制(E-MHSA)捕获低频信号，所述多头自注意力机制表示为：

[0050] $$E\text{-}MHCA=Concat\left(SA_1(x_1),SA_2(x_2),\ldots,SA_n(x_n)\right)W^P \quad (3)$$

[0051] 其中， $x=[x_1,x_2,\cdots,x_n]$ 表示将输入的特征值 x在通道维度上划分为多头形式，SA作为一种空间缩减自注意力算子表示为：

[0052] $$SA(X)=Attention\left(X\cdot W^Q,P_s(X\cdot W^K),P_s(X\cdot W^V)\right) \quad (4)$$

[0053] 其中， P_s 表示为一种具有Stride的平均池化操作，主要用于在注意力操作之前对空间维度进行降采样，从而减少计算成本。

[0054] 进一步，New Inception模块：本发明在三维分割网络模型中加入Inception结构的三维反卷积形式(New Inception)来作为三维分割网络模型的解码器，所述Inception结构由一个核大小为1×1×1的三维point-wise卷积及两个核大小为3×3×3和5×5×5的三维反卷积组成。所述反卷积模块的三维point-wise卷积层能够在相同大小的生成场上叠加更多的反卷积，从而获得待分割图像中更多的细节，以便进行更为细化的分割。另外，三维point-wise卷积层在降维中也起着至关重要的作用。进行三维反卷积运算将产生大量的计算量，通过加入三维point-wise卷积层可以有效减少输入特征的数量，从而提高计算效率。New Inception模块由两个分支网络组成，每个分支网络使用不同大小的滤波器进行反卷积，其中，所述分支网络可以产生不同规模的信息。

[0055] New Inception模块使用的原理是将稀疏矩阵分解成密集矩阵进行计算，将特征维数分解为多个密集分布的子特征集，其中，高度相关的特征聚集在一起，不相关的特征则被削弱，最后将各子特征在特征维度上进行拼接，并与输入维度保持一致，以降低计算成本，保证最终的训练结果不受影响。

[0056] 进一步，本发明将扩散模型加入到所述三维分割网络模型中，用于进行语义分割，其中，所述扩散模型在工作时，通过对语义标签添加随机噪声，并学习语

义标签上的随机噪声分量，然后对语义标签进行去噪，得到相关的语义信息，从而使得三维分割网络模型能够更好地分割出病灶区域。

[0057] 需要说明的是，参考图3，所述扩散模型的灵感主要来源于非平衡热力学，设定一个扩散步骤的马尔可夫链，通过逐步向真实数据中添加随机噪声（前向过程），然后学习反向扩散过程（逆扩散过程），从噪声中得到所需的数据样本结果。

[0058] 所述扩散模型分为前向过程以及逆向过程，其中，前向过程不包含可学习参数，其值随着t不断增大，最终值的分布为各向独立的高斯分布。例如，给出真实数据的分布为 $x_0 \sim q(x)$，在前向过程中通过逐步加入一个小的高斯噪声，一共加入 T步，从而产生一系列加入有噪声的样本 $x_1, x_2, x_3, \ldots, x_n$，加入噪声的均值和方差由 φ_t 决定，其值在(0,1)范围内，而且给定条件 $\varphi_1 < \varphi_2 < \ldots < \varphi_t$，可以看出，向扩散过程中所添加的噪声量越来越大，因此前向过程的定义为：

[0059] $$q(x_t|x_{t-1}) = \mathcal{N}\left(x_t; \sqrt{1-\varphi_t}(x_t, t); \varphi_t I\right) \quad (5)$$

[0060] 扩散过程中的逆向过程是从随机高斯噪声中恢复出原始数据，由于正向过程中每次所添加的噪声值很小，所以可以将 $p(x_{t-1}|x_t)$ 作为一个高斯分布，其过程可以使用深度神经网络进行拟合。因此，该逆向过程也可以是一个马尔科夫链过程，因此，所述逆向过程表示为：

[0061] $$p(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta x_{t-1}, \sum_\theta (x_t, t)) \quad (6)$$

[0062] 2）获取待分割图像，将所述待分割图像输入到所述基于扩散模型及3D Transformer的三维分割网络模型，得到待分割图像的脊柱三维分割结果。

[0063] 仿真实验

[0064] 本实验分别构建当前主流的基于卷积神经网络的分割网络以及本发明，对脊椎图像进行分割，分割结果如图4a、图4b、图4c及图4d所示，由图4a、图4b、图4c及图4d可知，本发明构建的分割网络模型所得到的分割结果更为精确。

[0065] 需要说明的是，本发明不仅可以用于脊柱三维分割，作为一种通用型算法，本发明还可以用于其他医学影像分割任务，如乳腺癌分割，肝脏分割等。

[0066] 实施例二

[0067] 参考图5，本发明所述的基于扩散模型及3D Transformer的三维分割系统，包括：

[0068] 获取模块1，用于获取待分割图像；

[0069] 分割模块2，用于将所述待分割图像输入到基于扩散模型及3D Transformer的三维分割网络模型中，得到待分割图像的三维分割结果。

[0070] 实施例三

[0071] 一种计算机设备，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序，所述处理器执行所述计算机程序时实现所述基于扩散模型及3D Transformer的三维分割方法的步骤，其中，所述存储器可能包含内存，例如高速随机存储器，也可能还包括非易失性存储器，例如，至少一个磁盘存储器等；处理器、网络接口、存储器通过内部总线互相连接，该内部总线可以是工业标准体系结构总线、外设部件互连标准总线、扩展工业标准结构总线等，总线可以分为地址总线、数据总线、控制总线等。存储器用于存放程序，具体地，程序可以包括程序代码、所述程序代码包括计算机操作指令。存储器可以包括内存和非易失性存储器，并向处理器提供指令和数据。

[0072] 实施例四

[0073] 一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序被处理器执行时实现所述基于扩散模型及3D Transformer的三维分割方法的步骤，具体地，所述计算机可读存储介质包括但不限于例如易失性存储器和/或非易失性存储器。所述易失性存储器可以包括随机存储存储器（RAM）和/或高速缓冲存储器（cache）等。所述非易失性存储器可以包括只读存储器（ROM）、硬盘、闪存、光盘、磁盘等。

[0074] 本领域内的技术人员应明白，本申请的实施例可提供为方法、系统、或计算机程序产品。因此，本申请可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质（包括但不限于磁盘存储器、CD-ROM、光学存储器等）上实施的计算机程序产品的形式。

[0075] 本申请是参照根据本申请实施例的方法、设备（系统）、和计算机程序产品的流程图和／或方框图来描述的。应理解可由计算机程序指令实现流程图和／或方框图中的每一流程和／或方框、以及流程图和／或方框图中的流程和／或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和／或方框图一个方框或多个方框中指定的功能的装置。

[0076] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一个流程或多个流程和／或方框图一个方框或多个方框中指定的功能。

[0077] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上，使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和／或方框图一个方框或多个方框中指定的功能的步骤。

[0078] 最后应当说明的是：以上实施例仅用以说明本发明的技术方案而非对其限制，尽管参照上述实施例对本发明进行了详细的说明，所属领域的普通技术人员应当理解：依然可以对本发明的具体实施方式进行修改或者等同替换，而未脱离本发明精神和范围的任何修改或者等同替换，其均应涵盖在本发明的权利要求保护范围之内。

权利要求书

[权利要求 1] 一种基于扩散模型及3D Transformer的三维分割方法，其特征在于，包括：

获取待分割图像；

将所述待分割图像输入到基于扩散模型及3D Transformer的三维分割网络模型中，得到待分割图像的三维分割结果。

[权利要求 2] 根据权利要求1所述的基于扩散模型及3D Transformer的三维分割方法，其特征在于，所述基于扩散模型及3D Transformer的三维分割网络模型由CNN-Transformer 混合架构、解码器及扩散模型构成。

[权利要求 3] 根据权利要求2所述的基于扩散模型及3D Transformer的三维分割方法，其特征在于，所述CNN-Transformer 混合架构由3D NCB模块及3D NTB模块堆叠而成，其中，通过所述3D NCB模块学习待分割图像中的局部表示；通过所述3D NTB模块捕获轻量级机制中的多频信号。

[权利要求 4] 根据权利要求3所述的基于扩散模型及3D Transformer的三维分割方法，其特征在于，所述3D NCB模块通过3D MHCA模块及3D MLP 模块构建而成。

[权利要求 5] 根据权利要求2所述的基于扩散模型及3D Transformer的三维分割方法，其特征在于，所述解码器基于Inception结构的三维反卷积形式构建而成。

[权利要求 6] 根据权利要求5所述的基于扩散模型及3D Transformer的三维分割方法，其特征在于，所述Inception结构由一个核大小为1×1×1的三维point-wise卷积层及两个核大小为3×3×3和5×5×5的三维反卷积层组成。

[权利要求 7] 根据权利要求2所述的基于扩散模型及3D Transformer的三维分割方法，其特征在于，所述扩散模型通过对语义标签添加随机噪声，

并学习语义标签上的随机噪声分量，再对所述语义标签进行去噪，得到相关的语义信息。

[权利要求 8] 一种基于扩散模型及3D Transformer的三维分割系统，其特征在于，包括：

获取模块（1），用于获取待分割图像；

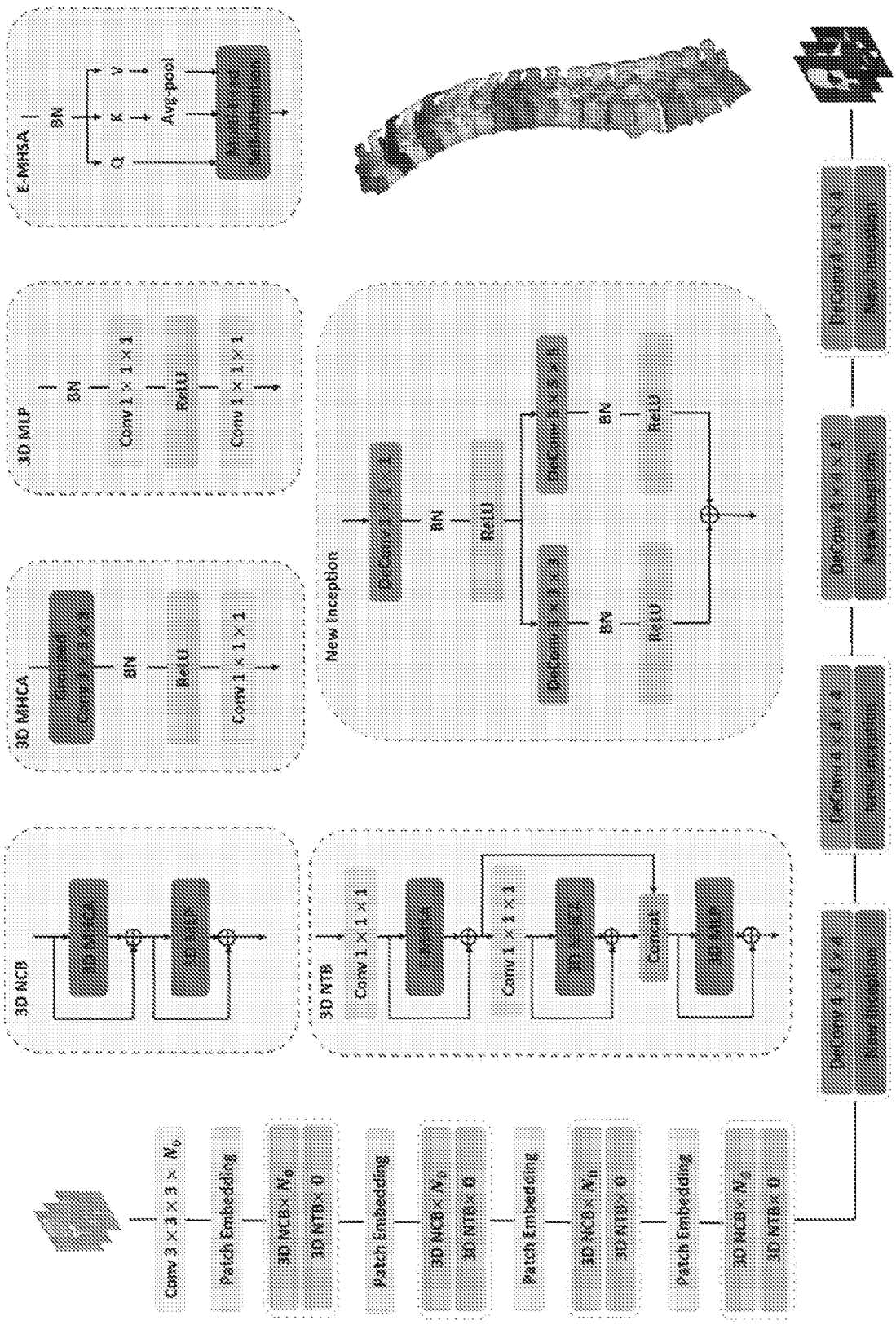
分割模块（2），用于将所述待分割图像输入到基于扩散模型及3D Transformer的三维分割网络模型中，得到待分割图像的三维分割结果。

[权利要求 9] 一种计算机设备，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序，其特征在于，所述处理器执行所述计算机程序时实现如权利要求1-7任一项所述基于扩散模型及3D Transformer的三维分割方法的步骤。

[权利要求 10] 一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，其特征在于，所述计算机程序被处理器执行时实现如权利要求1-7任一项所述基于扩散模型及3D Transformer的三维分割方法的步骤。

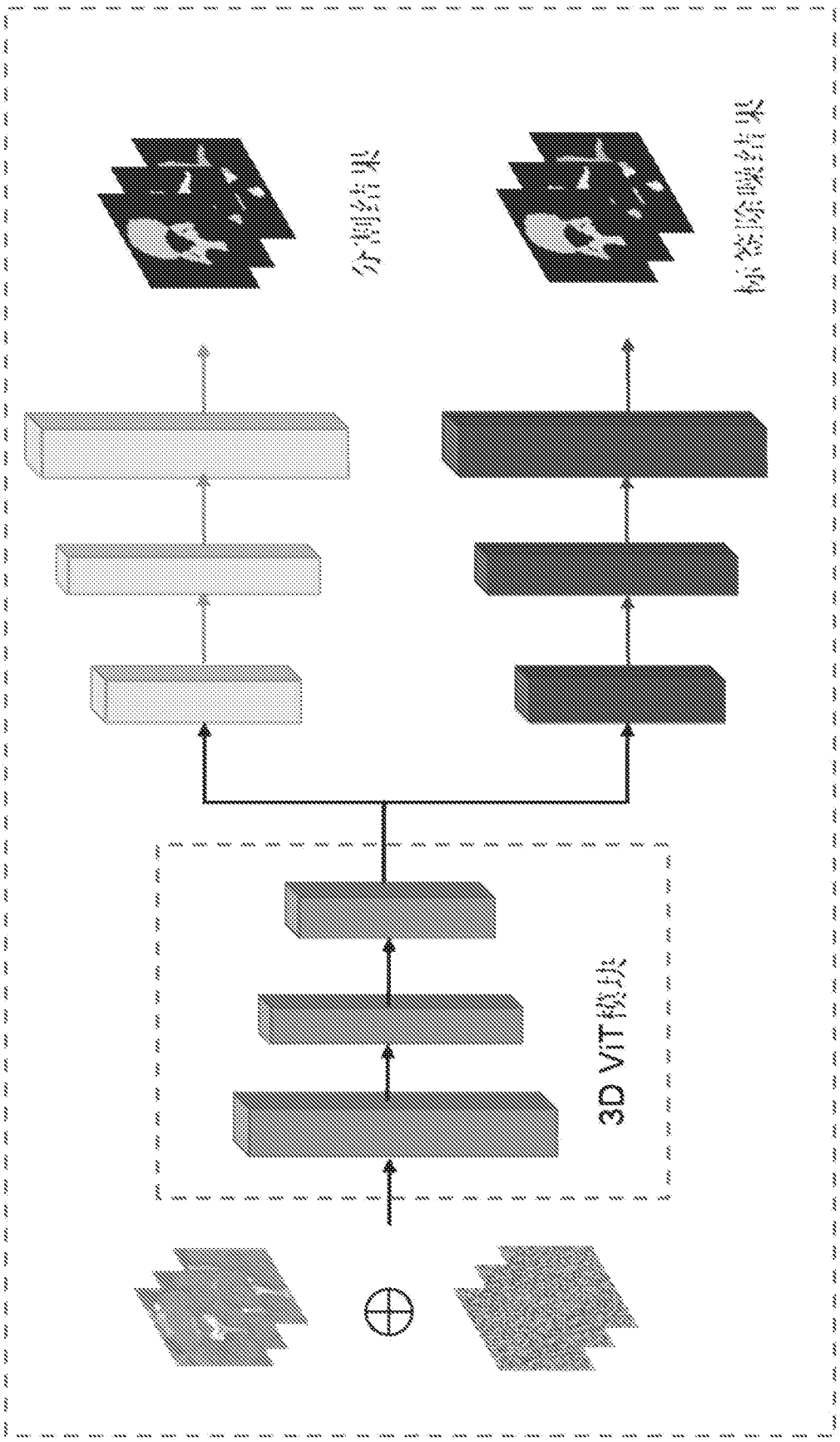
WO 2024/216672 PCT/CN2023/091750

[图 1]



细则 26.
24.07.2023

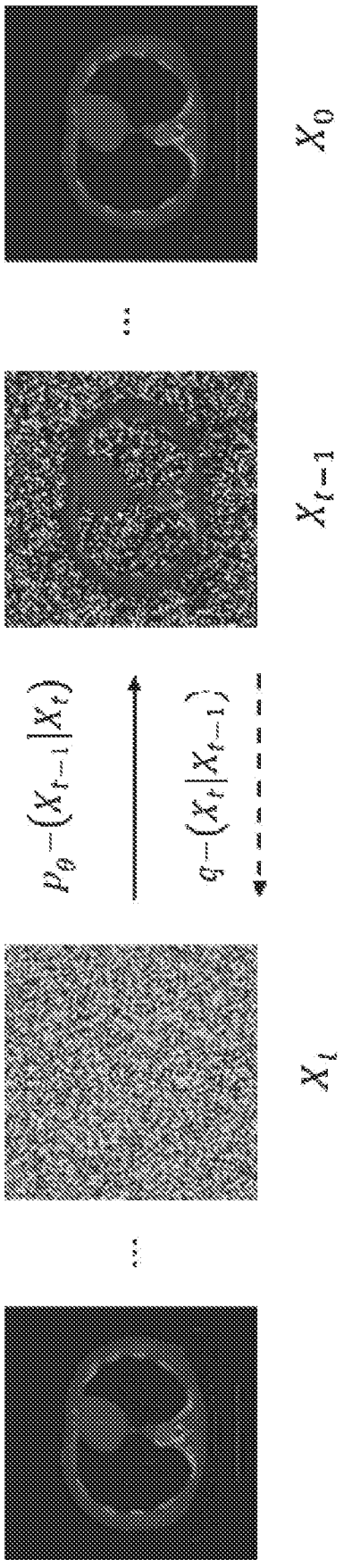
[图 2]



细则 26.
24.07.2023

细则 26.
24.07.2023

[图 3]

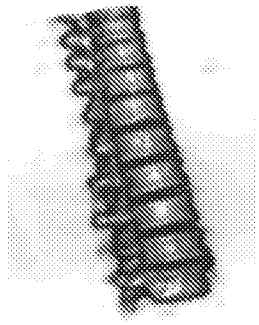


[图 4a]



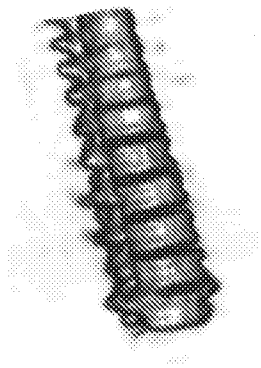
细则 26, 24.07.2023

[图 4b]



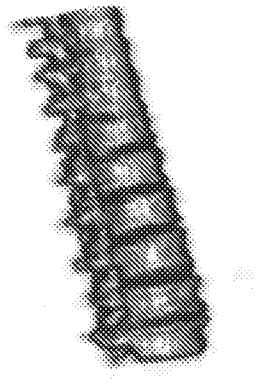
细则 26, 24.07.2023

[图 4c]



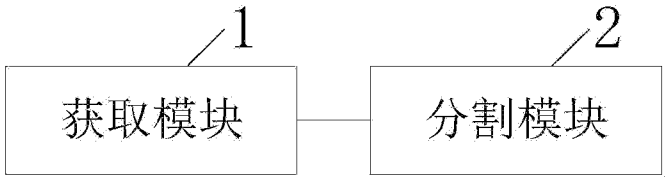
细则 26, 24.07.2023

[图 4d]



细则 26, 24.07.2023

[图 5]



细则 26, 24.07.2023

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2023/091750

A. CLASSIFICATION OF SUBJECT MATTER

G06T19/20(2011.01)i; G16H30/00(2018.01)n; G06N3/0464(2023.01)n; G06N3/0455(2023.01)n; G06V10/26(2022.01)n

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06T G16H G06N G06V

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CJFD, CNTXT, DWPI, ENTXT, ENTXTC, VEN, WPABS, WPABSC, semantic scholar, 百度学术, BAIDU SCHOLAR: 分割, 扩散, 模型, 三维, 解码, 噪声, 语义, 自注意力, 3D, transformer, viT, segmentation, self-attentioin, diffusion, model, NCB, NTB, inception, decode, semantic

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 115880312 A (CHONGQING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS) 31 March 2023 (2023-03-31) description, paragraphs 2-120, and figures 1-10	1-10
Y	CN 115409733 A (SHANDONG UNIVERSITY OF FINANCE AND ECONOMICS et al.) 29 November 2022 (2022-11-29) description, paragraphs 2-17, and figures 1-4	1-10,
Y	LI, Jiashi et al. "Next-ViT: Next Generation Vision Transformer for Efficient Deployment in Realistic Industrial Scenarios" *ArXiv, abs/2207.05501,* 16 August 2022 (2022-08-16), pages 1-14 abstract, and section 3	3-4
A	CN 115239637 A (SHENZHEN INSTITUTES OF ADVANCED TECHNOLOGY, CHINESE ACADEMY OF SCIENCES) 25 October 2022 (2022-10-25) entire document	1-10

☑ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
17 November 2023	**21 December 2023**

Name and mailing address of the ISA/CN	Authorized officer
China National Intellectual Property Administration (ISA/CN) **China No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088**	Telephone No.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 115457009 A (FUJIAN AGRICULTURE AND FORESTRY UNIVERSITY) 09 December 2022 (2022-12-09) entire document	1-10
A	CN 115018809 A (HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY) 06 September 2022 (2022-09-06) entire document	1-10
A	CN 115619810 A (PEKING UNION MEDICAL COLLEGE HOSPITAL, CHINESE ACADEMY OF MEDICAL SCIENCES) 17 January 2023 (2023-01-17) entire document	1-10
A	CN 115375698 A (CHONGQING RESEARCH INSTITUTE OF WUHAN UNIVERSITY OF TECHNOLOGY) 22 November 2022 (2022-11-22) entire document	1-10
A	WO 2022242131 A1 (SHENZHEN INSTITUTES OF ADVANCED TECHNOLOGY, CHINESE ACADEMY OF SCIENCES) 24 November 2022 (2022-11-24) entire document	1-10
A	WO 2023059699 A1 (GOOGLE L.L.C.) 13 April 2023 (2023-04-13) entire document	1-10

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2023/091750

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 115880312 A	31 March 2023	None	
CN 115409733 A	29 November 2022	CN 115409733 B	02 June 2023
CN 115239637 A	25 October 2022	None	
CN 115457009 A	09 December 2022	None	
CN 115018809 A	06 September 2022	None	
CN 115619810 A	17 January 2023	None	
CN 115375698 A	22 November 2022	None	
WO 2022242131 A1	24 November 2022	CN 113159056 A	23 July 2021
WO 2023059699 A1	13 April 2023	None	

国际检索报告

国际申请号
PCT/CN2023/091750

A. 主题的分类

G06T19/20(2011.01)i；G16H30/00(2018.01)n；G06N3/0464(2023.01)n；G06N3/0455(2023.01)n；G06V10/26(2022.01)n

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06T G16H G06N G06V

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用））

CJFD,CNTXT,DWPI,ENTXT,ENTXTC,VEN,WPABS,WPABSC,semantic scholar,百度学术:分割,扩散,模型,三维,解码,噪声,语义,自注意力,3D,transformer,viT,segmentation,self-attentioin,diffusion,model,NCB,NTB,inception,decode,semantic

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
Y	CN 115880312 A (重庆邮电大学) 2023年3月31日 (2023 - 03 - 31) 说明书第2-120段、附图1-10	1-10
Y	CN 115409733 A (山东财经大学等) 2022年11月29日 (2022 - 11 - 29) 说明书第2-17段，附图1-4	1-10
Y	Li, Jiashi et al. "Next-ViT: Next Generation Vision Transformer for Efficient Deployment in-Realistic Industrial Scenarios" ArXiv, abs/2207.05501, 2022年8月16日 (2022 - 08 - 16)，第1-14页 摘要，第3节	3-4
A	CN 115239637 A (中国科学院深圳先进技术研究院) 2022年10月25日 (2022 - 10 - 25) 全文	1-10
A	CN 115457009 A (福建农林大学) 2022年12月9日 (2022 - 12 - 09) 全文	1-10
A	CN 115018809 A (华中科技大学) 2022年9月6日 (2022 - 09 - 06) 全文	1-10

☑ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

"A" 认为不特别相关的表示了现有技术一般状态的文件

"D" 申请人在国际申请中引证的文件

"E" 在国际申请日的当天或之后公布的在先申请或专利

"L" 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

"O" 涉及口头公开、使用、展览或其他方式公开的文件

"P" 公布日先于国际申请日但迟于所要求的优先权日的文件

"T" 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

"X" 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

"Y" 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

"&" 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
2023年11月17日	2023年12月21日
ISA/CN的名称和邮寄地址 中国国家知识产权局 中国北京市海淀区蓟门桥西土城路6号 100088	受权官员 黄剑飞 电话号码 (+86) 027-59371789

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
A	CN 115619810 A (中国医学科学院北京协和医院) 2023年1月17日 (2023 - 01 - 17) 全文	1-10
A	CN 115375698 A (武汉理工大学重庆研究院) 2022年11月22日 (2022 - 11 - 22) 全文	1-10
A	WO 2022242131 A1 (SHENZHEN INST OF ADV TECH CAS) 2022年11月24日 (2022 - 11 - 24) 全文	1-10
A	WO 2023059699 A1 (GOOGLE LLC.) 2023年4月13日 (2023 - 04 - 13) 全文	1-10

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2023/091750

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	115880312	A	2023年3月31日	无			
CN	115409733	A	2022年11月29日	CN	115409733	B	2023年6月2日
CN	115239637	A	2022年10月25日	无			
CN	115457009	A	2022年12月9日	无			
CN	115018809	A	2022年9月6日	无			
CN	115619810	A	2023年1月17日	无			
CN	115375698	A	2022年11月22日	无			
WO	2022242131	A1	2022年11月24日	CN	113159056	A	2021年7月23日
WO	2023059699	A1	2023年4月13日	无			