

(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
2025 年 6 月 19 日 (19.06.2025)



(10) 国际公布号
WO 2025/123765 A1

(51) 国际专利分类号:

G06N 3/0455 (2023.01) **G06N 3/09** (2023.01)

G06N 3/0464 (2023.01) **G06N 5/04** (2023.01)

G06N 3/0475 (2023.01)

(21) 国际申请号: PCT/CN2024/113737

(22) 国际申请日: 2024 年 8 月 21 日 (21.08.2024)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:

202311736419.9 2023 年 12 月 15 日 (15.12.2023) CN

(71) 申请人: 阿里巴巴 (中国) 有限公司 (**ALIBABA (CHINA) CO., LTD**) [CN/CN]; 中国浙江省杭州市滨江区长河街道网商路 699 号 4 号楼 5 楼 508 室 310052 (CN)。

(72) 发明人: 江泽胤子 (**JIANG, Zeyinzi**); 中国浙江省杭州市西湖区三墩镇灯彩街 1008 号 310024 (CN)。毛超杰 (**MAO, Chaojie**); 中国浙江省杭州市西湖区三墩镇灯彩街 1008 号 310024 (CN)。潘玉霖 (**PAN,**

Yulin); 中国浙江省杭州市西湖区三墩镇灯彩街 1008 号 310024 (CN)。韩振 (**HAN, Zhen**); 中国浙江省杭州市西湖区三墩镇灯彩街 1008 号 310024 (CN)。张靖峰 (**ZHANG, Jingfeng**); 中国浙江省杭州市西湖区三墩镇灯彩街 1008 号 310024 (CN)。

(74) 代理人: 北京博浩百睿知识产权代理有限公司 (**KSK IP AGENCY**); 中国北京市海淀区知春路甲 48 号盈都大厦 C 座 1 单元 10E 100098 (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。

(54) **Title:** DATA PROCESSING METHOD, INTELLIGENT INTERACTION METHOD, MODEL TRAINING AND DEVELOPMENT METHOD, AND DEVICE AND MEDIUM

(54) 发明名称: 数据处理、智能交互、模型训练和开发方法、设备和介质

[图 2A]

获取预先训练的生成式模型, 生成式模型包括残差结构, 残差结构包括编码层与解码层之间的跳跃连接, 编码层与解码层之间形成至少一个调优子网络

S210

至少将描述数据输入到生成式模型, 得到描述数据对应的生成数据

S220

S210 Acquire a pre-trained generative model, wherein the generative model comprises a residual structure, the residual structure comprises a skip connection between an encoding layer and a decoding layer, and at least one tuning sub-network is formed between the encoding layer and the decoding layer

S220 At least input description data into the generative model, so as to obtain generated data corresponding to the description data

(57) **Abstract:** A data processing method, an intelligent interaction method, a model training and development method, and a device and a medium. The data processing method comprises: acquiring a pre-trained generative model, wherein the generative model comprises a residual structure, the residual structure comprises a skip connection between an encoding layer and a decoding layer, and at least one tuning sub-network is formed between the encoding layer and the decoding layer; and at least inputting description data into the generative model, so as to obtain generated data corresponding to the description data. Since a residual structure comprises a skip connection between an encoding layer and a decoding layer, and at least one tuning sub-network is formed between the encoding layer and the decoding layer, the at least one tuning sub-network can be connected between the encoding layer and the decoding layer, which form the skip connection, so as to efficiently train the residual structure to obtain a generative model having higher prediction reliability, and thus the generative model is used to execute more reliable data processing.

[见续页]

(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: 一种数据处理、智能交互、模型训练和开发方法、设备和介质。数据处理方法包括: 获取预先训练的生成式模型, 生成式模型包括残差结构, 残差结构包括编码层与解码层之间的跳跃连接, 编码层与解码层之间形成至少一个调优子网络; 至少将描述数据输入到生成式模型, 得到描述数据对应的生成数据。残差结构包括编码层与解码层之间的跳跃连接, 编码层与解码层之间形成至少一个调优子网络, 因此, 至少一个调优子网络能够通过连接到跳跃连接的编码层与解码层之间, 高效地训练残差结构, 进而得到预测可靠性更高的生成式模型, 从而采用生成式模型执行更可靠的数据处理。

说明书

发明名称: 数据处理、智能交互、 模型训练和开发方法、设备和介质

[0001] 本申请要求于2023年12月15日提交中国专利局、申请号为202311736419.9、申请名称“数据处理、智能交互、模型训练和开发方法、设备和介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

[0002] 本申请实施例涉及计算机技术领域，尤其涉及一种数据处理、智能交互、模型训练和开发方法、设备和介质。

背景技术

[0003] 在人工智能领域中，诸如大语言模型等生成式模型通过生成式网络构建，诸如扩散生成式网络具有有效的网络结构，通过样本训练之后能够表现出极优的泛化能力。基于大规模数据预训练下的基础生成式模型能够针对诸如图像生成、图像编辑、图像变换等一系列下游任务及应用进行微调训练。

[0004] 随着生成式模型规模的参数数量不断增长，对这样的模型进行各项任务适配的方法的成本变得相当昂贵，而如果采用低成本的训练方式，则训练效果较差，相应地采用训练后的生成式模型进行数据处理时，可靠性较差。因此，生成式模型的训练效率以及采用生成式模型进行数据处理的可靠性均存在提升的空间。

[0005] 发明内容

[0006] 有鉴于此，本申请实施例提供一种数据处理、智能交互、模型训练和开发方法、设备和介质，以至少部分解决上述问题。

[0007] 根据本申请实施例的第一方面，提供了一种数据处理方法，包括：获取预先训练的生成式模型，所述生成式模型包括残差结构，所述残差结构包括编码层与解码层之间的跳跃连接，所述编码层与所述解码层之间形成至少一个调优子网

络；至少将描述数据输入到所述生成式模型，得到所述描述数据对应的生成数据。

[0008] 根据本申请实施例的第二方面，提供了一种智能交互方法，包括：获取交互界面中输入的第一描述数据；至少将所述第一描述数据输入到预先训练的生成式模型，得到所述第一描述数据对应的第一生成数据，其中，所述第一生成数据包括图像、视频帧、视频中的至少一者，所述生成式模型包括残差结构，所述残差结构包括编码层与解码层之间的跳跃连接，所述编码层与所述解码层之间形成至少一个调优子网络。在所述问答界面中展示所述第一生成数据。

[0009] 根据本申请实施例的第三方面，提供了一种模型训练方法，包括：获取初始训练的生成式网络，所述生成式网络包括编码器和解码器，所述编码器的编码层与所述解码器的解码层之间形成跳跃连接；在所述初始训练的生成式网络的所述编码层与所述解码层之间连接所述跳跃连接的至少一个调优子网络，得到待微调生成式网络，其中，所述编码层的输出连接到所述至少一个调优子网络的输入，所述至少一个调优子网络的输出连接到所述解码层的输入；基于训练样本，对所述待微调生成式网络进行微调训练，得到生成式模型。

[0010] 根据本申请实施例的第四方面，提供了一种应用开发方法，包括：创建应用程序的用户界面模块，所述用户界面模块被配置成至少基于用户操作数据生成描述数据，并且基于所述描述数据的生成数据返回呈现数据；获取生成式模型的调用接口，所述生成式模型根据第三方面所述的模型训练方法得到，所述调用接口被配置成在被调用时返回所述生成数据；至少将所述生成式模型的调用接口嵌入到所述用户界面模块中。

[0011] 根据本申请实施例的第五方面，提供了一种数据处理装置包括：获取模块，设置为获取预先训练的生成式模型，所述生成式模型包括残差结构，所述残差结构包括编码层与解码层之间的跳跃连接，所述编码层与所述解码层之间形成至少一个调优子网络；生成模块，设置为至少将描述数据输入到所述生成式模型，得到所述描述数据对应的生成数据。

[0012] 根据本申请实施例的第六方面，提供了一种智能交互装置包括：获取模块，设置为获取交互界面中输入的第一描述数据；生成模块，设置为至少将所述第一

描述数据输入到预先训练的生成式模型，得到所述第一描述数据对应的第一生成数据，其中，所述第一生成数据包括图像、视频帧、视频中的至少一者，所述生成式模型包括残差结构，所述残差结构包括编码层与解码层之间的跳跃连接，所述编码层与所述解码层之间形成至少一个调优子网络；展示模块，设置为在所述问答界面中展示所述第一生成数据。

[0013] 根据本申请实施例的第七方面，提供了一种模型训练装置包括：模块获取模块，设置为获取初始训练的生成式网络，生成式网络包括编码器和解码器，编码器的编码层与解码器的解码层之间形成跳跃连接；模型调整模块，设置为在所述初始训练的生成式网络的所述编码层与所述解码层之间连接所述跳跃连接的至少一个调优子网络，得到待微调生成式网络，其中，所述编码层的输出连接到所述至少一个调优子网络的输入，所述至少一个调优子网络的输出连接到所述解码层的输入；模型训练模块，设置为基于训练样本，对所述待微调生成式网络进行微调训练，得到生成式模型。

[0014] 根据本申请实施例的第八方面，提供了一种应用开发装置包括：创建模块，设置为创建应用程序的用户界面模块，所述用户界面模块被配置成至少基于用户操作数据生成描述数据，并且基于所述描述数据的生成数据返回呈现数据；获取模块，设置为获取生成式模型的调用接口，所述生成式模型根据第三方面所述的模型训练方法得到，所述调用接口被配置成在被调用时返回所述生成数据；嵌入模块，设置为至少将所述生成式模型的调用接口嵌入到所述用户界面模块中。

[0015] 根据本申请实施例的第九方面，提供了一种电子设备，包括：处理器、存储器、通信接口和通信总线，所述处理器、所述存储器和所述通信接口通过所述通信总线完成相互间的通信；所述存储器用于存放至少一可执行指令，所述可执行指令使所述处理器执行如第一方面所述方法对应的操作。

[0016] 根据本申请实施例的第十方面，提供了一种计算机存储介质，其上存储有计算机程序，该程序被处理器执行时实现如第一方面所述的方法。

[0017] 在本实施例中，预先训练的生成式模型包括残差结构，残差结构包括编码层与解码层之间的跳跃连接，编码层与解码层之间形成至少一个调优子网络，因此

，至少一个调优子网络能够通过连接到跳跃连接的编码层与解码层之间，高效地训练残差结构，进而得到预测可靠性更高的生成式模型，从而采用生成式模型能够执行更可靠的数据处理。

附图说明

[0018] 为了更清楚地说明本申请实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请实施例中记载的一些实施例，对于本领域普通技术人员来讲，还可以根据这些附图获得其他的附图。

[0019] 图1为根据一些示例的生成式网络的结构示意图。

[0020] 图2A为根据本申请的另一些实施例的数据处理方法的步骤流程图。

[0021] 图2B为图2实施例的一些示例的生成式网络的部分结构示意图。

[0022] 图2C为图2实施例的另一些示例的生成式网络的部分结构示意图。

[0023] 图2D为根据本申请的另一些实施例的数据处理方法的步骤流程图。

[0024] 图2E为图2实施例的另一些示例的生成式网络的部分结构示意图。

[0025] 图3为根据本申请的另一些实施例的智能交互方法的步骤流程图。

[0026] 图4A为根据本申请的另一些实施例的模型训练方法的步骤流程图。

[0027] 图4B为根据本申请的另一些实施例的模型训练方法的步骤流程图。

[0028] 图5为根据本申请的另一些实施例的应用开发方法的步骤流程图。

[0029] 图6为根据本申请的另一些实施例的数据处理装置的示意性框图。

[0030] 图7为根据本申请的另一些实施例的智能交互装置的示意性框图。

[0031] 图8为根据本申请的另一些实施例的模型训练装置的示意性框图。

[0032] 图9为根据本申请的另一些实施例的应用开发装置的示意性框图。

[0033] 图10为根据本申请的另一些实施例的电子设备的结构示意图。

具体实施方式

[0034] 为了使本领域的人员更好地理解本申请实施例中的技术方案，下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅是本申请实施例一部分实施例，而不是全部的实施例。

基于本申请实施例中的实施例，本领域普通技术人员所获得的所有其他实施例，都应当属于本申请实施例保护的范围。

[0035] 下面结合本申请实施例附图进一步说明本申请实施例具体实现。

[0036] 诸如大语言模型等生成式模型通过生成式网络构建，例如，图1所示的生成式网络包括编码器110和解码器120，编码器110的输出可以直接连接到解码器120的输入，编码器110的输出也可以通过采用注意力机制的上下文对齐层(未示出)连接到解码器120的输入。图1的生成式网络可以为生成式扩散网络，其具有编码层与解码层之间的跳跃连接(Skip Connection, SC)的结构。跳跃连接可以实现为一个或多个，例如，图1所示的跳跃连接#1、跳跃连接#2、...、跳跃连接#N。各个跳跃连接在编码器110与解码器120之间形成了残差结构，从而使生成式网络获得了更优良的泛化能力。然而，随着生成式模型规模的参数数量不断增长对这样的模型进行各项任务适配的方法的成本变得相当昂贵，本申请的各个实施例提供了一系列的技术方案能够提高微调训练的效率并且减少生成式网络的微调训练开销。

[0037] 图2A为根据本申请的另一一些实施例的数据处理方法的步骤流程图。本实施例的方案可以适用于任意适当的具有数据处理能力的电子设备，包括但不限于：服务器、移动终端(如手机、PAD等)和PC机等。例如，在模型训练(training)阶段，可以利用配置有CPU(处理单元的示例)+GPU(加速单元的示例)架构的计算设备(例如，数据中心)基于训练样本对编码器解码器模型进行训练。诸如数据中心的计算设备可以部署在诸如专有云、私有云、或混合云的云服务器中。相应地，在推理(inference)阶段，也可以利用配置有CPU(处理单元的示例)+GPU(加速单元的示例)架构的计算设备进行推理运算。具体地，图2A的数据处理方法，包括：

[0038] S210：获取预先训练的生成式模型，生成式模型包括残差结构，残差结构包括编码层与解码层之间的跳跃连接，编码层与解码层之间形成至少一个调优子网络。

[0039] 应理解，生成式模型可以通过生成式网络训练得到，生成式网络可以是生成扩散网络。生成式模型可以是根据描述文本来生成图像的生成式模型，也可以是根据文本来生成文本的文本处理模型，例如，知识问答模型或翻译模型。

- [0040] 还应理解，生成式模型可以包括编码器和解码器，编码器与解码器之间可以具有一个或多个跳跃连接(Skip Connection, SC)的残差结构，例如，U-Net结构。编码器中的网络层为编码层解码器中的网络层为解码层，跳跃连接所连接的编码层和解码层对应。编码器与解码器之间可以采用注意力机制等形式包括一个或多个网络层(例如，全连接层)的上下文对齐结构。
- [0041] S220：至少将描述数据输入到生成式模型，得到描述数据对应的生成数据。
- [0042] 应理解，描述数据可以描述文本或者能够转换成描述文本的描述数据，例如，可以将语音描述数据转换成对应的描述文本，输入到生成式模型中。对于具体的转换方式，可以将语音描述数据输入到预先训练的文本转换模型中，得到描述文本，也可以将文本转换模型连接到生成式模型的编码器的输入侧。可替代地，还可以将生成式模型训练成接收语音描述数据作为输入，也就是说，在对生成式模型进行端到端训练时，训练样本的输入侧样本为语音样本。
- [0043] 还应理解，生成数据可以为描述文本(例如，在知识问答场景或翻译场景中)；生成数据也可以为图片或视频帧，或者基于视频帧序列的视频。这取决于生成式模型的训练阶段所采用的端到端训练的样本类型，如果输出侧样本为视频帧序列，则推理阶段的生成数据为视频，如果输出侧样本为图片或视频帧，则推理阶段的生成数据为图片或视频帧。
- [0044] 在本实施例中，预先训练的生成式模型包括残差结构，残差结构包括编码层与解码层之间的跳跃连接，编码层与解码层之间形成至少一个调优子网络，因此，至少一个调优子网络能够通过连接到跳跃连接的编码层与解码层之间，高效地训练残差结构，进而得到预测可靠性更高的生成式模型，从而采用生成式模型能够执行更可靠的数据处理。
- [0045] 在另一些示例中，各种条件图像数据分别输入到各个条件子网络。条件图像可以为描述文本和多个附加条件，不同图像条件用于描述不同类型的图像特征，多种类型的图像特征数据包括但不限于指示图像边缘检测特征，图像深度特征、图像分割特征、图像颜色特征、图像扩展特征、图像修复特征等不同图像特征。具体地，将描述文本输入到生成式模型的编码器，将每个附加条件分别输入到多个跳跃连接的对应调优子网络中，得到输出的生成图像。

[0046] 具体而言，在如图2B所示的残差结构(U-Net结构)中，编码层的输出连接到每个调优子网络30的输入，每个调优子网络的输出连接到解码层的输入。也就是说，每个调优子网络的输出与编码层通过跳跃连接的输入进行融合作为跳跃连接的解码层的输入。即： $O_j^{sc}(X_{N-j}) = T_j(X_{N-j}) + X_{N-j}$ 。

[0047] 其中， O_j^{sc} 为残差结构中的第j层的输出， X_{N-j} 为跳跃连接的输入， T_j 为调优子网络的输出，N为残差结构中编码器的总层数。

[0048] 在另一些示例中，残差结构还包括至少一个条件子网络，至少一个条件子网络对应于至少一个调优子网络。作为至少将描述数据输入到生成式模型，得到描述数据对应的生成数据的示例，可以将描述数据输入到生成式模型的编码层，将条件数据输入到生成式模型的至少一个条件子网络，从生成式模型的解码层获得描述数据对应的生成数据。

[0049] 具体地，至少一个调优子网络分别对应至少一个条件子网络，每个条件子网络的第一输出连接到对应的调优子网络的输入与编码层的输入之间，每个条件子网络的第二输出连接在对应的调优子网络的输出与解码层的输入之间。

[0050] 更具体地，每个条件子网络的第二输出与对应的调优子网络的输出融合后，经由该条件子网络在多个条件子网络中的条件权重进行处理，输出到解码层的输入。

[0051] 在另一些示例中，描述数据为描述文本，生成数据包括多个视频帧。数据处理方法还包括：将多个视频帧进行时序编码处理，得到描述文本的生成视频。

[0052] 图2D示出了根据本申请的另一些实施例的数据处理方法。

[0053] S280：获取预先训练的生成式模型。

[0054] S290：将描述数据输入到生成式模型的编码层，将条件数据输入到生成式模型的至少一个条件子网络，从生成式模型的解码层获得描述数据对应的生成数据。

[0055] 应理解，图2D的实施例与图2A的实施例相同或类似的部分，此处不再赘述。例如，步骤S280与步骤S210类似。

[0056] 进一步地，本实施例的生成式模型可以包括附带条件子网络的调优子网络。例如，如图2C所示，至少一个调优子网络30分别对应至少一个条件子网络40，每

个条件子网络40的第一输出连接在对应的调优子网络30的输入与编码层的输入之间，每个条件子网络40的第二输出连接在对应的调优子网络30的输出与解码层的输入之间。

[0057] 即：

$O_j^{CSC}(X_{N-j}, C_j) = \sum_{m=1}^M \alpha^m (T_j^m (X_{N-j} + C_j^m) + C_j^m) + X_{N-j}$, 其中, O_j^{CSC} 为残差结构中的第j层的调优子网络处理后的输出；m为第m个条件，遍历1到M之间的值； X_{N-j} 为输入； T_j^m 为第j层第m个条件的调优子网络， α^m 为第m个条件的条件权重系数； C_j 为第j层第m个条件编码后的输入， C_j 为 $\{C_{j0} \cdots C_{jm}\}$ 的条件集合。

[0058] 其中，M个条件为在微调训练时与编码器的输入数据的补充条件数据。例如，在编码器的输入为描述文本时，M个条件可以为对描述文本进行补充的条件图像。

[0059] 更进一步地，如图2E所示，对于各个条件子网络(例如，M个条件子网络)，条件子网络的数目可以与调优子网络的数目一致，也就是说，第m条件子网络对应于第m调优子网络。

[0060] 更具体地，编码器的多个编码层与解码器的多个解码层之间分别形成N个跳跃连接，每个跳跃连接对应M个调优子网络和M个条件子网络，第m个条件子网络对应于第m个调优子网络。第m个条件子网络包括从输入侧到输出侧依次设置的N个网络层，第n个网络层对应于第n跳跃连接，第n个网络层的输出连接到第n跳跃连接的第m个调优子网络的输入。其中，N为大于1的整数，M为大于1的整数， $1 \leq n \leq N$ ， $1 \leq m \leq M$ 。

[0061] 进一步地，第n网络层的输出连接到第n跳跃连接的编码层。应理解，每个网络层可以由诸如下采样层的线性层和激活层等实现，第n网络层的输出的数据维度与第n跳跃连接的编码层的数据维度对齐或一致，从而提高了生成式模型的训练过程的可靠性，也提高了采用生成式模型进行数据处理的可靠性。

[0062] 在数据处理方法的一些变型例中，包括未附带条件子网络的调优子网络的生成式模型的生成图像可以作为包括附带条件子网络的调优子网络的生成式模型的条件图像，从而通过描述文本生成了条件图像，进一步采用附带条件子网络的

调优子网络的生成式模型，将条件图像与进一步的描述文本两者得到生成图像

。

[0063] 在数据处理方法的另一些变型例中，未附带条件子网络的调优子网络的生成式模型的生成图像可以作为包括附带条件子网络的调优子网络的生成式模型的编码器的输入数据，从而采用所附带的条件子网络进一步得到更加匹配描述文本的生成图像。

[0064] 下面将结合图3详细描述本申请的另一些实施例的基于生成式模型的智能交互方法的步骤流程图。图3的智能交互方法包括：

[0065] S310：获取交互界面中输入的第一描述数据。

[0066] S320：至少将第一描述数据输入到预先训练的生成式模型，得到第一描述数据对应的第一生成数据，其中，第一生成数据包括图像、视频帧、视频中的至少一者，生成式模型包括残差结构，残差结构包括编码层与解码层之间的跳跃连接，编码层与解码层之间形成至少一个调优子网络。

[0067] S330：在问答界面中展示第一生成数据。

[0068] 在本实施例中，通过用户界面来接收用户输入的描述数据并且向用户展示数据处理结果，实现了快捷的交互，此外，本实施例对用户输入的描述数据的类型不作限定，方便用户能够输入诸如文本描述数据或语音描述数据等各种类型的描述数据。此外，用户界面所呈现的生成数据包括不限于文本、图片或视频。因此，本实施例通过结合用户界面的便捷交互方式以及生成式模型的数据处理能力，为用户提供了智能信息服务，例如，可以应用于智能助手、虚拟专家等智能产品中。

[0069] 在一些示例中，智能交互方法还包括：从交互界面中获取针对第一生成数据进行修改的第二描述数据；至少将第二描述数据输入到预先训练的生成式模型，生成第二描述数据对应的第二生成数据，作为第一生成数据的修改结果。

[0070] 通过上述数据修改流程，在第一生成数据的初始生成流程之后，以第一描述数据作为上下文数据，基于第二描述数据作为修改描述数据，对第一生成数据进行了调整，得到了第二生成描述。

- [0071] 应理解，此处的第一生成数据和第二生成数据的类型可以一致，例如，对语音数据进行修改得到了修改后的语音数据，对文本数据进行修改得到了修改后的文本数据，对视频数据进行修改得到了修改后的视频数据。
- [0072] 进一步地，第一描述数据和第二描述数据的类型可以相同，也可以不同。例如，在另一些示例中，第一生成数据为图片或视频帧。从交互界面中获取针对第一生成数据进行修改的第二描述数据，包括：从交互界面中获取针对图片或视频的修改标记；识别修改标记，生成第二描述数据，或者，结合修改标记以及图片或视频的修改描述文本，生成第二描述数据。
- [0073] 也就是说，可以采用与第一描述数据的相同类型的第二描述数据，作为第一描述数据的补充和修改提示，也可以采用与第一描述数据的不同类型的第二描述数据，作为第一描述数据的补充和修改提示。例如，在第一描述数据为文本数据时，可以进一步补充文本数据来修改第一生成数据，也可以采用与文本数据不同的语音数据/操作数据来作为文本数据的补充或修改提示，操作数据可以是对上述图片或视频帧中所标记的目标区域，也就是说，目标区域作为待修改的区域或者修改权重较大的区域，通过这样的操作，结合了操作数据的第二描述数据由于提供了更多了修改权重的信息，能够进一步提高修改的精度和可靠性。
- [0074] 具体地，图4A示出了根据本申请的一些实施例的模型训练方法。图4A的模型训练方法包括：
- [0075] S410：获取初始训练的生成式网络，生成式网络包括编码器和解码器，编码器的编码层与解码器的解码层之间形成跳跃连接。
- [0076] 应理解，初始训练的生成式网络可以是经由预训练得到的生成式网络，例如，生成式扩散网络。
- [0077] S420：在初始训练的生成式网络的编码层与解码层之间连接跳跃连接的至少一个调优子网络，得到待微调生成式网络，其中，编码层的输出连接到至少一个调优子网络的输入，至少一个调优子网络的输出连接到解码层的输入。
- [0078] 应理解，部分或全部跳跃连接的编码层与解码层之间可以连接至少一个调优子网络，调优子网络形成的结构可以被称为SC-Tuner。调优子网络至少包括线性

层和激活层，例如，调优子网络可以形成有第一线性层、激活层和第二线性层形成，激活层设置在第一线性层与第二线性层之间。此外，还可以在第一线性层的输入侧和/或第二线性层的输出侧增加其他激活层，也可以调优子网的任意位置增加其他线性层。上述的线性层可以为全连接层、也可以为卷积层。在待微调生成式网络的输出为图像数据的情况下，调优子网络中的线性层可以是二维数据，在待微调生成式网络的输出为文本数据的情况下，调优子网络中的线性层可以是一维数据，这是因为，调优子网络连接在跳跃连接的编码层与解码层之间，只需与解码器的输出数据的维度相同。

[0079] S430：基于训练样本，对待微调生成式网络进行微调训练，得到生成式模型。

[0080] 应理解，对待微调生成式网络的微调训练可以是监督训练，训练样本可以包括输入侧样本和输出侧样本，输入侧样本可以为注入到编码器中的特定编码层的文本数据，输入侧样本还可以包括编码器的输入的图像样本，输出侧样本可以为文本数据，也可以为诸如图片或视频帧的图像数据。

[0081] 在本申请实施例的方案中，待微调生成式网络的至少一个调优子网络连接在跳跃结构所连接的编码层与解码层之间，编码层的输出连接到至少一个调优子网络的输入，至少一个调优子网络的输出连接到解码层的输入，使得在通过微调训练生成式网络时，通过至少一个调优子网络进一步地优化跳跃结构，提高了微调训练的效率。此外，至少一个调优子网络未处于编码器中或解码器中，无需在微调训练中过多地进行参数调整，节省了微调训练的计算开销。

[0082] 进一步地，作为基于训练样本，对待微调生成式网络进行微调训练的示例，可以基于训练样本，对待微调生成式网络进行多次迭代训练，得到生成式模型，其中，在每次迭代训练中，确定训练样本中的输入侧样本经过待微调生成式网络的正向传播输出与训练样本中的输入侧样本之间的差异，并且在保持编码器的各个参数的情况下，通过差异在解码器和至少一个调优子网络中的梯度反向传播，调整解码器和至少一个调优子网络中的各个参数，即，在每次迭代训练中，无需对编码器中的参数进行训练，对编码器与至少一个调优子网络的训练过程进行了解耦，从而显著地减小了微调训练中的计算开销。

- [0083] 具体地，输入侧样本可以为描述文本，输出侧样本可以为生成图像，例如，描述文本用于描述生成图像，从而能够使训练后的生成式模型能够更有效地适配初始训练的生成式网络的下游任务。
- [0084] 例如，以稳定扩散生成式网络为例，诸如U-Net等残差结构中的往往包含12层跳跃连接，在每一层的跳跃连接上加上调优子网络，可针对不同的任务进行相应高效训练。相比于诸如LoRA等其他微调训练方法，能够以更小的调优参数和内存消耗进行训练。也就是说，在高效的生成式调优或少样本任务调优上的应用各个调优子网络，可以适用到少量样本的条件下进行快速定制化的迁移训练，诸如学习到特定的人物特征、特定的图像风格等图像特征。
- [0085] 在调优子网络未附带条件子网络(即，SC-Tuner)的情况下，作为对待微调生成式网络进行微调训练的示例，可以将输入侧样本作为编码器的输入，将输出侧样本作为解码器的输出，对至少一个调优子网络中的网络参数进行调整，得到生成式模型，从而可靠地训练了未附带条件子网络的调优子网络。也就是说，调优子网络的输入侧的约束和输出侧的约束来自于调优子网络对应的跳跃连接的编码层和解码层，相应地，输入侧样本至少包括描述文本，输出侧样本至少包括图像样本，图像样本包括但不限于图片和视频帧。
- [0086] 进一步地，每个条件子网络的第二输出与对应的调优子网络的输出融合后，经由该条件子网络在多个条件子网络中的条件权重进行处理，输出到解码层的输入。应理解，由于多个条件子网络对应的条件的重要性可以是不同，一条件子网络在多个条件子网络中的条件权重指示该条件子网络中的重要性。此外，条件子网络的条件权重也可以实现为诸如全连接层的线性层，连接在各个调优子网络与对应的编码层之间。至少一个调优子网络及其条件子网络对应于跳跃连接，附带条件子网络的调优子网络也可以被称为CSC-Tuner。
- [0087] 进一步地，不同的条件子网络的功能不同，即，多个条件子网络分别用于输出多种图像特征数据，例如，多种图像特征数据包括但不限于指示图像边缘检测特征，图像深度特征、图像分割特征、图像颜色特征、图像扩展特征、图像修复特征等不同图像特征。

- [0088] 不失一般性地，在模型训练框架中，可以配置用于微调训练残差结构的组件，通过将跳跃连接和相关的条件输入到SC-Tuner(即，不附带条件子网络的调优子网络)和CSC-Tuner(即，附带条件子网络的调优子网络)中，并和原始的相应残差结构的解码层中的输入进行拼接送入下一解码层中。
- [0089] 即， $g_{j+1} = G_j[O_j(X_{N-j-1} + C_j); g_j]$ ，其中， g_j 为第j层的残差结构的解码层的输入； G_j 为第j层的残差结构的解码层操作后的输出； O_j 为残差结构中的第j层的调优子网络处理后的输出。
- [0090] 编码器120的多个编码层与解码器120的多个解码层之间分别形成多个跳跃连接130，条件子网络40包括从输入侧到输出侧依次设置的多个网络层，多个网络层分别对应于多个跳跃连接。
- [0091] 具体而言，如图2E所示，对于各个条件子网络(例如，M个条件子网络)，条件子网络的数目可以与调优子网络的数目一致，也就是说，第m条件子网络对应于第m调优子网络。
- [0092] 更具体地，编码器的多个编码层与解码器的多个解码层之间分别形成N个跳跃连接，每个跳跃连接对应M个调优子网络和M个条件子网络，第m个条件子网络对应于第m个调优子网络。第m个条件子网络包括从输入侧到输出侧依次设置的N个网络层，第n个网络层对应于第n跳跃连接，第n个网络层的输出连接到第n跳跃连接的第m个调优子网络的输入。其中，N为大于1的整数，M为大于1的整数， $1 \leq n \leq N$ ， $1 \leq m \leq M$ 。
- [0093] 进一步地，第n网络层的输出连接到第n跳跃连接的编码层。应理解，每个网络层可以由诸如下采样层的线性层和激活层等实现，第n网络层的输出的数据维度与第n跳跃连接的编码层的数据维度对齐或一致，从而提高了生成式模型的训练过程的可靠性。
- [0094] 也就是说，第n+1跳跃连接的第m条件子网络输出连接到第n+1网络层的输入，得到第n跳跃连接的第m条件子网络。即，第n+1网络层的输出作为第n跳跃连接的第m条件子网络的输出。
- [0095] 图4B示出了根据本申请的另一些实施例的模型训练方法，包括：

- [0096] S470: 获取初始训练的生成式网络, 生成式网络包括编码器和解码器, 编码器的编码层与解码器的解码层之间形成跳跃连接。
- [0097] S480: 在初始训练的生成式网络的编码层与解码层之间连接跳跃连接的至少一个调优子网络和至少一个条件子网络, 得到待微调生成式网络, 其中, 编码层的输出连接到至少一个调优子网络的输入, 至少一个调优子网络的输出连接到解码层的输入。至少一个调优子网络分别对应至少一个条件子网络, 每个条件子网络的第一输出连接到对应的调优子网络的输入与编码层的输入之间, 每个条件子网络的第二输出连接在对应的调优子网络的输出与解码层的输入之间。
- [0098] S490: 将输入侧样本作为编码器的输入, 将输出侧样本作为解码器的输出, 将条件样本作为至少一个条件子网络的输入, 对至少一个调优子网络和至少一个条件子网络中的网络参数进行调整, 得到生成式模型。
- [0099] 在本实施例中, 待微调生成式网络的网络结构可以参考图2C和图2E, 此处不再赘述。例如, 步骤S470与步骤S410类似。
- [0100] 也就是说, 在本实施例中, 在调优子网络附带条件子网络(即, CSC-Tuner)的情况下, 作为对待微调生成式网络进行微调训练的示例, 至少一个调优子网络分别对应至少一个条件子网络, 可以将输入侧样本作为编码器的输入, 将输出侧样本作为解码器的输出, 将条件样本作为至少一个条件子网络的输入, 对至少一个调优子网络和至少一个条件子网络中的网络参数进行调整, 得到生成式模型。也就是说, 在条件控制生成式图像处理任务中, 将条件先通过级联的网络层进行编码(即, C_j), 作为调优子网络的一个输入, 与调优子网络从编码层的另一输入独立地应用于调优子网络, 从而提升了调优子网络的训练效果。
- [0101] 下面将结合图5来详细描述根据本申请的另一些实施例的应用开发方法。图5的应用开发方法可以应用于PAAS云服务或IAAS云服务, 即, 生成式模型在PAAS云服务或IAAS云服务的服务端执行推理过程(其训练过程可以在该服务端执行或在不同的服务端执行后再部署到该服务端), 并且向PAAS云服务或IAAS云服务的租户以虚拟机的方式开放生成式模型的调用接口, 即, 生成式模型的调用接口提供给一个虚拟机或多个虚拟机, 租户可以基于调用接口在虚拟中执行应用程序的开发。

- [0102] 具体地，应用开发方法包括：
- [0103] S510：创建应用程序的用户界面模块，用户界面模块被配置成至少基于用户操作数据生成描述数据，并且基于描述数据的生成数据返回呈现数据。
- [0104] S520：获取生成式模型的调用接口，生成式模型根据模型训练方法得到，调用接口被配置成在被调用时返回生成数据；
- [0105] S530：至少将生成式模型的调用接口嵌入到用户界面模块中。
- [0106] 在另一些示例中，用户界面模块还被配置成调用服务模块返回呈现数据，服务模块被配置成对生成数据进行数据处理，得到呈现数据，调用接口被配置成在被服务模块调用时返回生成数据。
- [0107] 具体地，用户界面模块用于实现应用程序的与用户交互的界面，可以用于接收用户的输入操作并且向用户提供输出结果，应用程序还包括至少一个服务模块，服务模块与用户界面模块对应，能够执行基于用户的输入进行数据处理，并且得到处理结果。服务模块的处理过程可以完全由生成式模型执行，即，服务模块的输入直接或通过简单预处理作为生成式模型的输入，生成式模型的输出直接或者通过简单的后续处理作为服务模块的输出。可替代地，服务模块的处理过程可以部分有生成式模型执行，即，基于生成式模型的推理能力增强服务模块的数据处理能力，这时，服务模块的功能函数被配置为调用生成式模型的调用接口，或者，生成式模型通过调用接口返回的输出数据作为服务模块的功能函数的输入。
- [0108] 例如，应用程序为内容编辑类应用程序，其中包括诸如字幕生成、文案生成、配音生成等服务模块。相应地，用户界面模块中可以提供内容编辑界面。进一步地，通过嵌入生成式模型的调用接口，可以从调用接口返回生成的字幕数据、文案数据或配音数据等，服务模块中的功能函数能够进一步对字幕数据、文案数据或配音数据等进行基于时间戳的对齐处理，并且将对齐处理结果整体作为服务模块的输出，返回到用户界面模块将对齐处理结果以可视化的方式提供给内容编辑界面，使用户实现了应用程序的高效开发。此外，将虚拟机与生成式模型的调用接口打包提供给租户，在提高应用程序的开发效率的同时还降低了本地开发的资源要求，并且提高了开发的灵活性。

[0109] 下面将结合图6描述根据本申请的另一些实施例的数据处理装置。图6的数据处理装置与数据处理方法对应，数据处理装置包括：

[0110] 获取模块610，设置为获取预先训练的生成式模型，所述生成式模型包括残差结构，所述残差结构包括编码层与解码层之间的跳跃连接，所述编码层与所述解码层之间形成至少一个调优子网络；

[0111] 生成模块620，设置为至少将描述数据输入到所述生成式模型，得到所述描述数据对应的生成数据。

[0112] 在另一些示例中，所述残差结构还包括至少一个条件子网络，所述至少一个条件子网络对应于至少一个调优子网络；生成模块设置为：将描述数据输入到所述生成式模型的编码层，将条件数据输入到所述生成式模型的至少一个条件子网络，从所述生成式模型的解码层获得所述描述数据对应的生成数据。

[0113] 在另一些示例中，所述至少一个调优子网络分别对应至少一个条件子网络，每个条件子网络的第一输出连接到对应的调优子网络的输入与所述编码层的输入之间，每个条件子网络的第二输出连接在对应的调优子网络的输出与所述解码层的输入之间。

[0114] 在另一些示例中，每个条件子网络的第二输出与对应的调优子网络的输出融合后，经由该条件子网络在所述多个条件子网络中的条件权重进行处理，输出到所述解码层的输入。

[0115] 在另一些示例中，所述描述数据为描述文本，所述生成数据包括多个视频帧，数据处理装置还包括：视频处理模块，设置为将所述多个视频帧进行时序编码处理，得到所述描述文本的生成视频。

[0116] 下面将结合图7详细描述根据本申请的另一些实施例的智能交互装置。智能交互装置与智能交互方法对应，具体地，智能交互装置包括：

[0117] 获取模块710，设置为获取交互界面中输入的第一描述数据；

[0118] 生成模块720，设置为至少将所述第一描述数据输入到预先训练的生成式模型，得到所述第一描述数据对应的第一生成数据，其中，所述第一生成数据包括图像、视频帧、视频中的至少一者，所述生成式模型包括残差结构，所述残

差结构包括编码层与解码层之间的跳跃连接，所述编码层与所述解码层之间形成至少一个调优子网络。

[0119] 展示模块730，设置为在所述问答界面中展示所述第一生成数据。

[0120] 在另一些示例中，获取模块还设置为：从所述交互界面中获取针对所述第一生成数据进行修改的第二描述数据。生成模块还设置为：至少将所述第二描述数据输入到预先训练的生成式模型，生成所述第二描述数据对应的第二生成数据，作为所述第一生成数据的修改结果。

[0121] 在另一些示例中，所述第一生成数据为图片或视频帧。获取模块还设置为：从所述交互界面中获取针对所述图片或视频的修改标记，识别所述修改标记，生成所述第二描述数据，或者，结合所述修改标记以及所述图片或视频的修改描述文本，生成所述第二描述数据。

[0122] 下面将结合图8详细描述根据本申请的另一些实施例的模型训练装置。本实施例的方案可以适用于任意适当的具有数据处理能力的电子设备，包括但不限于：服务器、移动终端(如手机、PAD等)和PC机等。例如，在模型训练(training)阶段，可以利用配置有CPU(处理单元的示例)+GPU(加速单元的示例)架构的计算设备(例如，数据中心)基于训练样本对编码器解码器模型进行训练。诸如数据中心的计算设备可以部署在诸如专有云、私有云、或混合云的云服务器中。相应地，在推理(inference)阶段，也可以利用配置有CPU(处理单元的示例)+GPU(加速单元的示例)架构的计算设备进行推理运算。

[0123] 具体地，图8的模型训练装置对应模型训练方法对应，模型训练装置包括：

[0124] 模块获取模块810，设置为获取初始训练的生成式网络，生成式网络包括编码器和解码器，编码器的编码层与解码器的解码层之间形成跳跃连接。

[0125] 模型调整模块820，设置为在所述初始训练的生成式网络的所述编码层与所述解码层之间连接所述跳跃连接的至少一个调优子网络，得到待微调生成式网络，其中，所述编码层的输出连接到所述至少一个调优子网络的输入，所述至少一个调优子网络的输出连接到所述解码层的输入。

[0126] 模型训练模块830，设置为基于训练样本，对所述待微调生成式网络进行微调训练，得到生成式模型。

- [0127] 在本申请实施例的方案中，待微调生成式网络的至少一个调优子网络连接在跳跃结构所连接的编码层与解码层之间，编码层的输出连接到至少一个调优子网络的输入，至少一个调优子网络的输出连接到解码层的输入，使得在通过微调训练生成式网络时，通过至少一个调优子网络进一步地优化跳跃结构，提高了微调训练的效率，此外，至少一个调优子网络未处于编码器中或解码器中，无需在微调训练中过多地进行参数调整，节省了微调训练的计算开销。
- [0128] 在另一些示例中，基于训练样本，对所述待微调生成式网络进行微调训练，包括：基于训练样本，对所述待微调生成式网络进行多次迭代训练，得到生成式模型，其中，在每次迭代训练中，确定所述训练样本中的输入侧样本经过所述待微调生成式网络的正向传播输出与所述训练样本中的输入侧样本之间的差异，并且在保持编码器的各个参数的情况下，通过所述差异在所述解码器和所述至少一个调优子网络中的梯度反向传播，调整所述解码器和所述至少一个调优子网络中的各个参数。
- [0129] 在另一些示例中，基于训练样本，对所述待微调生成式网络进行微调训练，得到生成式模型，包括：将输入侧样本作为所述编码器的输入，将输出侧样本作为所述解码器的输出，对至少一个调优子网络中的网络参数进行调整，得到生成式模型。
- [0130] 在另一些示例中，至少一个调优子网络分别对应至少一个条件子网络，每个条件子网络的第一输出连接在对应的调优子网络的输入与所述编码层的输入之间，每个条件子网络的第二输出连接在对应的调优子网络的输出与所述解码层的输入之间；基于训练样本，对所述待微调生成式网络进行微调训练，得到生成式模型，包括：将输入侧样本作为所述编码器的输入，将输出侧样本作为所述解码器的输出，将条件样本作为至少一个条件子网络的输入，对至少一个调优子网络和所述至少一个条件子网络中的网络参数进行调整，得到生成式模型。
- [0131] 在另一些示例中，每个条件子网络的第二输出与对应的调优子网络的输出融合后，经由该条件子网络在所述多个条件子网络中的条件权重进行处理，输出到所述解码层的输入。

[0132] 在另一些示例中，编码器的多个编码层与所述解码器的多个解码层之间分别形成N个跳跃连接，每个跳跃连接对应M个调优子网络和M个条件子网络，第m个条件子网络对应于第m个调优子网络。第m个条件子网络包括从输入侧到输出侧依次设置的N个网络层，第n个网络层对应于第n跳跃连接，第n个网络层的输出连接到第n跳跃连接的第m个调优子网络的输入。

[0133] 下面将结合图9描述根据本申请的另一些实施例的应用开发装置。应用开发装置与应用开发方法对应，具体地，应用开发装置包括：

[0134] 创建模块910，设置为创建应用程序的用户界面模块，所述用户界面模块被配置成至少基于用户操作数据生成描述数据，并且基于所述描述数据的生成数据返回呈现数据；

[0135] 获取模块920，设置为获取生成式模型的调用接口，所述生成式模型根据模型训练方法得到，所述调用接口被配置成在被调用时返回所述生成数据；

[0136] 嵌入模块930，设置为至少将所述生成式模型的调用接口嵌入到所述用户界面模块中。

[0137] 在另一些示例中，所述用户界面模块还被配置成调用服务模块返回所述呈现数据，所述服务模块被配置成对所述生成数据进行数据处理，得到所述呈现数据，所述调用接口被配置成在被所述服务模块调用时返回所述生成数据。

[0138] 上述的模型训练装置或数据处理装置中各模块的具体实现可以参见上述方法实施例中的相应步骤对应的描述，并具有相应的有益效果，在此不赘述。所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的装置和模块的具体工作过程，可以参考前述方法实施例中的对应过程描述，在此不再赘述。

[0139] 参照图10，示出了根据本申请的另一实施例的一种电子设备的结构示意图，本申请具体实施例并不对电子设备的具体实现做限定。

[0140] 如图10所示，该电子设备可以包括：用于执行程序1010的处理器(processor)1002、通信接口(Communications Interface)1004、存储器(memory)1006、以及通信总线1008。

[0141] 处理器、通信接口、以及存储器通过通信总线完成相互间的通信。

- [0142] 通信接口，用于与其它电子设备或服务器进行通信。
- [0143] 处理器，用于执行程序，具体可以执行上述方法实施例中的相关步骤。
- [0144] 具体地，程序可以包括程序代码，该程序代码包括计算机操作指令。
- [0145] 处理器可能是CPU，或者是特定集成电路ASIC(Application Specific Integrated Circuit)，或者是被配置成实施本申请实施例的一个或多个集成电路。智能设备包括的一个或多个处理器，可以是同一类型的处理器，如一个或多个CPU；也可以是不同类型的处理器，如一个或多个CPU以及一个或多个ASIC。
- [0146] 存储器，用于存放程序。存储器可能包含高速RAM存储器，也可能还包括非易失性存储器(non-volatile memory)，例如至少一个磁盘存储器。
- [0147] 程序可包括多条计算机指令，程序具体可以通过多条计算机指令使得处理器执行前述多个方法实施例中任一实施例所描述的模型训练方法或数据处理方法对应的操作。
- [0148] 程序中各步骤的具体实现可以参见上述方法实施例中的相应步骤和单元中对应的描述，并具有相应的有益效果，在此不赘述。所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的设备和模块的具体工作过程，可以参考前述方法实施例中的对应过程描述，在此不再赘述。
- [0149] 本申请实施例还提供了一种计算机存储介质，其上存储有计算机程序，该程序被处理器执行时实现前述多个方法实施例中任一实施例所描述的方法。该计算机存储介质包括但不限于：只读光盘(Compact Disc Read-Only Memory, CD-ROM)、随机存储器(Random Access Memory, RAM)、软盘、硬盘或磁光盘等。
- [0150] 本申请实施例还提供了一种计算机程序产品，包括计算机指令，该计算机指令指示计算设备执行上述多个方法实施例中的模型训练方法或数据处理方法对应的操作。
- [0151] 此外，需要说明的是，本申请实施例所涉及到的与用户有关的信息(包括但不限于用户设备信息、用户个人信息等)和数据(包括但不限于用于对模型进行训练的样本数据、用于分析的数据、存储的数据、展示的数据等)，均为经用户授权或者经过各方充分授权的信息和数据，并且相关数据的收集、使用和处理需要遵守相关规定和标准，并提供有相应的操作入口，供用户选择授权或者拒绝。

- [0152] 需要指出, 根据实施的需要, 可将本申请实施例中描述的各个部件/步骤拆分为更多部件/步骤, 也可将两个或多个部件/步骤或者部件/步骤的部分操作组合成新的部件/步骤, 以实现本申请实施例的目的。
- [0153] 上述根据本申请实施例的方法可在硬件、固件中实现, 或者被实现为可存储在记录介质(诸如CD-ROM、RAM、软盘、硬盘或磁光盘)中的软件或计算机代码, 或者被实现通过网络下载的原始存储在远程记录介质或非暂时机器可读介质中并将被存储在本地记录介质中的计算机代码, 从而在此描述的方法可被存储在使用通用计算机、专用处理器或者可编程或专用硬件(诸如专用集成电路(Application Specific Integrated Circuit, ASIC)或现场可编程门阵列(Field Programmable Gate Array, FPGA))的记录介质上的这样的软件处理。可以理解, 计算机、处理器、微处理器控制器或可编程硬件包括可存储或接收软件或计算机代码的存储组件(例如, 随机存储器(Random Access Memory, RAM)、只读存储器(Read-Only Memory, ROM)、闪存等), 当所述软件或计算机代码被计算机、处理器或硬件访问且执行时, 实现在此描述的方法。此外, 当通用计算机访问用于实现在此示出的方法的代码时, 代码的执行将通用计算机转换为用于执行在此示出的方法的专用计算机。
- [0154] 本领域普通技术人员可以意识到, 结合本文中所公开的实施例描述的各示例的单元及方法步骤, 能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行, 取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能, 但是这种实现不应认为超出本申请实施例的范围。
- [0155] 以上实施方式仅用于说明本申请实施例, 而并非对本申请实施例的限制, 有关技术领域的普通技术人员, 在不脱离本申请实施例的精神和范围的情况下, 还可以做出各种变化和变型, 因此所有等同的技术方案也属于本申请实施例的范畴, 本申请实施例的专利保护范围应由权利要求限定。

工业实用性

- [0156] 本申请实施例所提供的预先训练的生成式模型包括残差结构, 残差结构包括编码层与解码层之间的跳跃连接, 编码层与解码层之间形成至少一个调优子网络

，因此，至少一个调优子网络能够通过连接到跳跃连接的编码层与解码层之间，高效地训练残差结构，进而得到预测可靠性更高的生成式模型，从而采用生成式模型能够执行更可靠的数据处理。

权利要求书

- [权利要求 1] 一种数据处理方法，包括：
- 获取预先训练的生成式模型，所述生成式模型包括残差结构，所述残差结构包括编码层与解码层之间的跳跃连接，所述编码层与所述解码层之间形成至少一个调优子网络；
- 至少将描述数据输入到所述生成式模型，得到所述描述数据对应的生成数据。
- [权利要求 2] 根据权利要求1所述的数据处理方法，其中，所述残差结构还包括至少一个条件子网络，所述至少一个条件子网络对应于至少一个调优子网络；
- 至少将描述数据输入到所述生成式模型，得到所述描述数据对应的生成数据，包括：
- 将描述数据输入到所述生成式模型的编码层，将条件数据输入到所述生成式模型的至少一个条件子网络，从所述生成式模型的解码层获得所述描述数据对应的生成数据。
- [权利要求 3] 根据权利要求2所述的数据处理方法，其中，所述至少一个调优子网络分别对应至少一个条件子网络，每个条件子网络的第一输出连接到对应的调优子网络的输入与所述编码层的输入之间，每个条件子网络的第二输出连接在对应的调优子网络的输出与所述解码层的输入之间。
- [权利要求 4] 根据权利要求3所述的数据处理方法，其中，每个条件子网络的第二输出与对应的调优子网络的输出融合后，经由该条件子网络在所述多个条件子网络中的条件权重进行处理，输出到所述解码层的输入。
- [权利要求 5] 根据权利要求1所述的数据处理方法，其中，所述描述数据为描述文本，所述生成数据包括多个视频帧，
- 所述方法还包括：

将所述多个视频帧进行时序编码处理，得到所述描述文本的生成视频。

[权利要求 6]

一种智能交互方法，包括：

获取交互界面中输入的第一描述数据；

至少将所述第一描述数据输入到预先训练的生成式模型，得到所述第一描述数据对应的第一生成数据，其中，所述第一生成数据包括图像、视频帧、视频中的至少一者，所述生成式模型包括残差结构，所述残差结构包括编码层与解码层之间的跳跃连接，所述编码层与所述解码层之间形成至少一个调优子网络；

在所述问答界面中展示所述第一生成数据。

[权利要求 7]

根据权利要求6所述的智能交互方法，其中，所述方法还包括：

从所述交互界面中获取针对所述第一生成数据进行修改的第二描述数据；

至少将所述第二描述数据输入到预先训练的生成式模型，生成所述第二描述数据对应的第二生成数据，作为所述第一生成数据的修改结果。

[权利要求 8]

根据权利要求6所述的智能交互方法，其中，所述第一生成数据为图片或视频帧，

从所述交互界面中获取针对所述第一生成数据进行修改的第二描述数据，包括：

从所述交互界面中获取针对所述图片或视频的修改标记；

识别所述修改标记，生成所述第二描述数据，或者，结合所述修改标记以及所述图片或视频的修改描述文本，生成所述第二描述数据。

[权利要求 9]

一种模型训练方法，包括：

获取初始训练的生成式网络，所述生成式网络包括编码器和解码器，所述编码器的编码层与所述解码器的解码层之间形成跳跃连接；

在所述初始训练的生成式网络的所述编码层与所述解码层之间连接所述跳跃连接的至少一个调优子网络，得到待微调生成式网络，其中，所述编码层的输出连接到所述至少一个调优子网络的输入，所述至少一个调优子网络的输出连接到所述解码层的输入；
基于训练样本，对所述待微调生成式网络进行微调训练，得到生成式模型。

[权利要求 10]

根据权利要求9所述的训练方法，其中，基于训练样本，对所述待微调生成式网络进行微调训练，包括：

基于训练样本，对所述待微调生成式网络进行多次迭代训练，得到生成式模型，其中，在每次迭代训练中，确定所述训练样本中的输入侧样本经过所述待微调生成式网络的正向传播输出与所述训练样本中的输入侧样本之间的差异，并且在保持编码器的各个参数的情况下，通过所述差异在所述解码器和所述至少一个调优子网络中的梯度反向传播，调整所述解码器和所述至少一个调优子网络中的各个参数。

[权利要求 11]

根据权利要求9所述的训练方法，其中，基于训练样本，对所述待微调生成式网络进行微调训练，得到生成式模型，包括：

将输入侧样本作为所述编码器的输入，将输出侧样本作为所述解码器的输出，对所述至少一个调优子网络中的网络参数进行调整，得到生成式模型。

[权利要求 12]

根据权利要求9所述的训练方法，其中，所述至少一个调优子网络分别对应至少一个条件子网络，每个条件子网络的第一输出连接到对应的调优子网络的输入与所述编码层的输入之间，每个条件子网络的第二输出连接在对应的调优子网络的输出与所述解码层的输入之间；

基于训练样本，对所述待微调生成式网络进行微调训练，得到生成式模型，包括：

将输入侧样本作为所述编码器的输入，将输出侧样本作为所述解码器的输出，将条件样本作为至少一个条件子网络的输入，对所述至少一个调优子网络和所述至少一个条件子网络中的网络参数进行调整，得到生成式模型。

[权利要求 13] 根据权利要求12所述的训练方法，其中，每个条件子网络的第二输出与对应的调优子网络的输出融合后，经由该条件子网络在所述多个条件子网络中的条件权重进行处理，输出到所述解码层的输入。

[权利要求 14] 一种应用开发方法，包括：
创建应用程序的用户界面模块，所述用户界面模块被配置成至少基于用户操作数据生成描述数据，并且基于所述描述数据的生成数据返回呈现数据；

获取生成式模型的调用接口，所述生成式模型根据权利要求9-13中任一项所述的模型训练方法得到，所述调用接口被配置成在被调用时返回所述生成数据；

至少将所述生成式模型的调用接口嵌入到所述用户界面模块中。

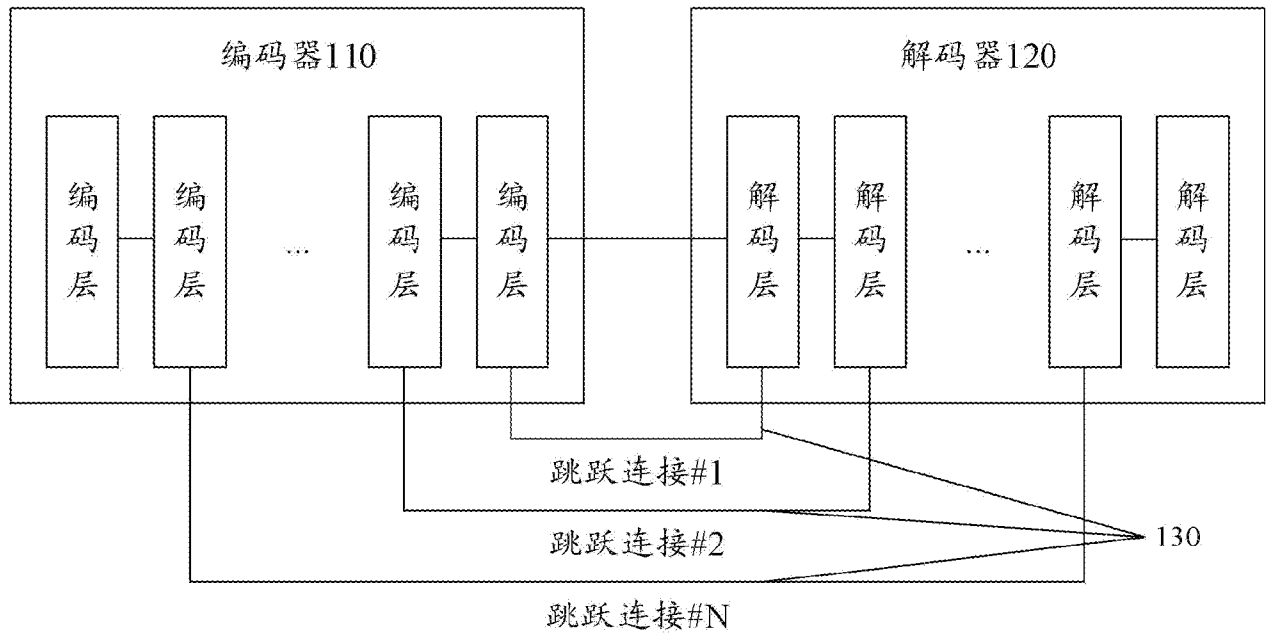
[权利要求 15] 根据权利要求14所述的方法，其中，所述用户界面模块还被配置成调用服务模块返回所述呈现数据，所述服务模块被配置成对所述生成数据进行数据处理，得到所述呈现数据，所述调用接口被配置成在被所述服务模块调用时返回所述生成数据。

[权利要求 16] 一种电子设备，包括：处理器、存储器、通信接口和通信总线，所述处理器、所述存储器和所述通信接口通过所述通信总线完成相互间的通信；

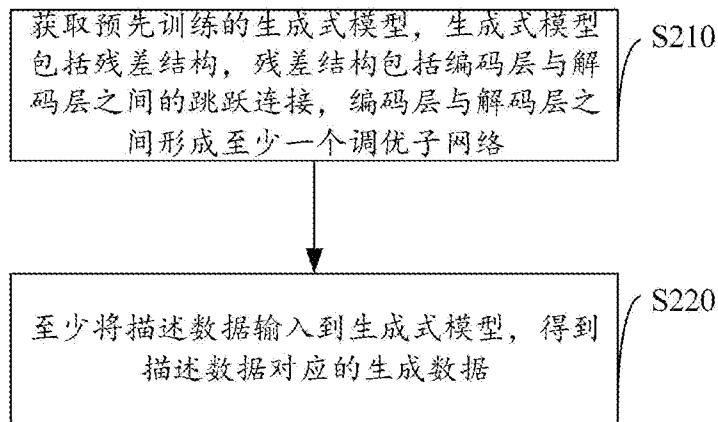
所述存储器用于存放至少一可执行指令，所述可执行指令使所述处理器执行根据权利要求1-15中任一项所述的方法对应的操作。

[权利要求 17] 一种计算机存储介质，其上存储有计算机程序，该程序被处理器执行时实现根据权利要求1-15中任一项所述的方法。

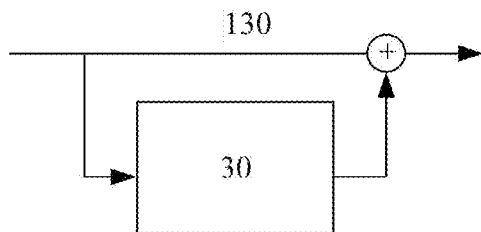
[图 1]



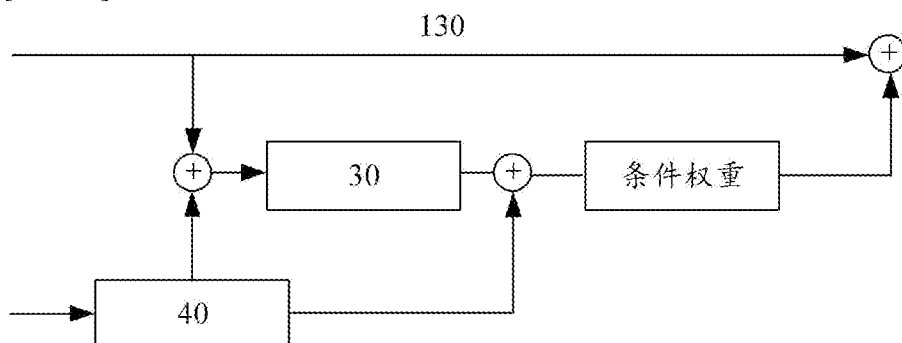
[图 2A]



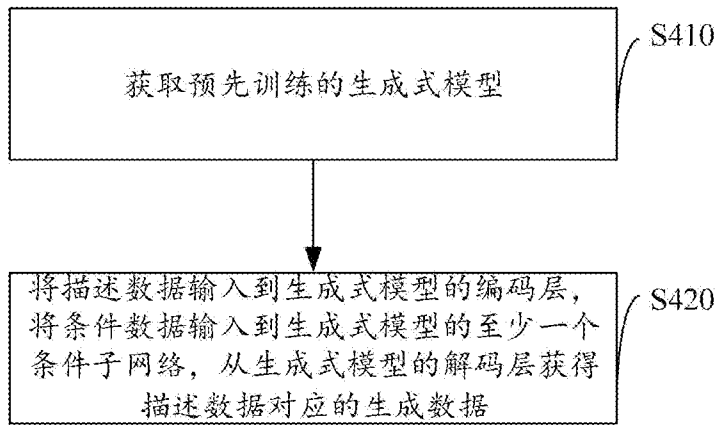
[图 2B]



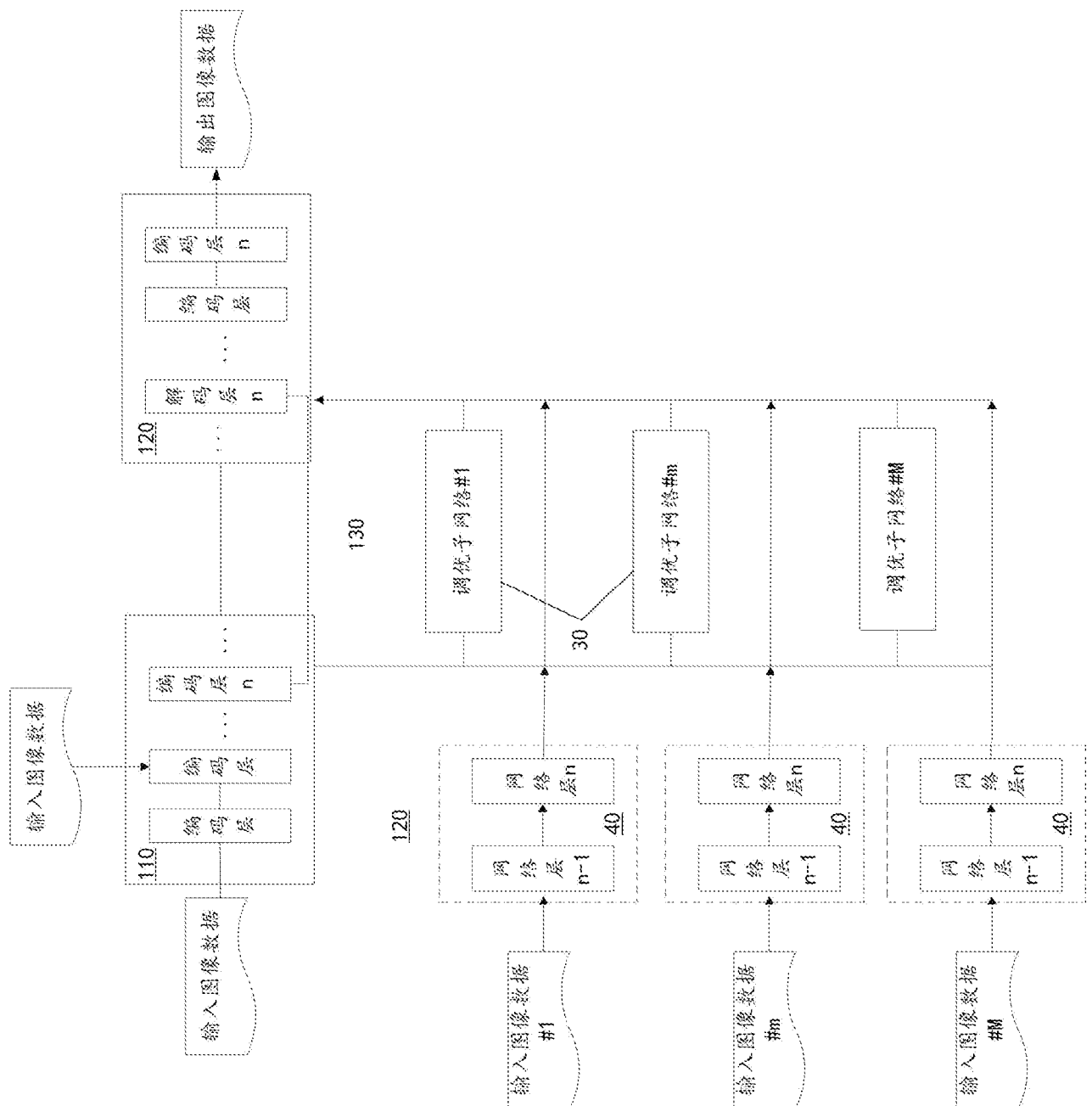
[图 2C]



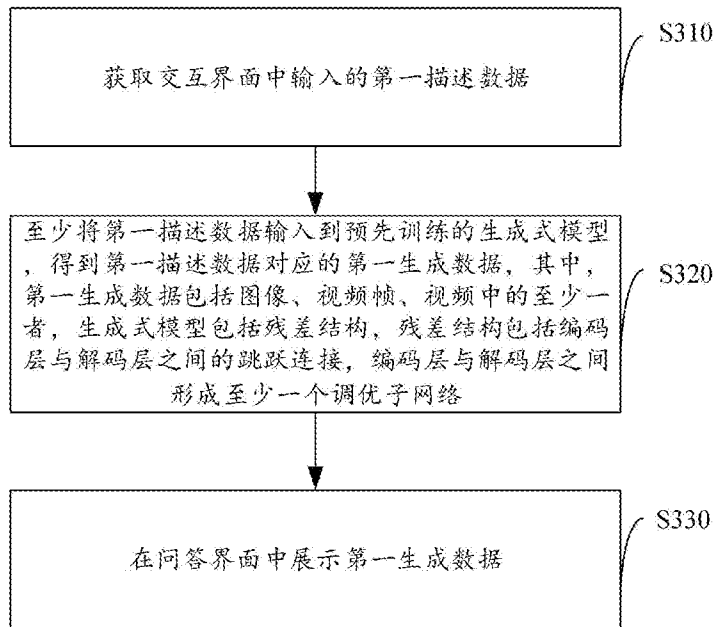
[图 2D]



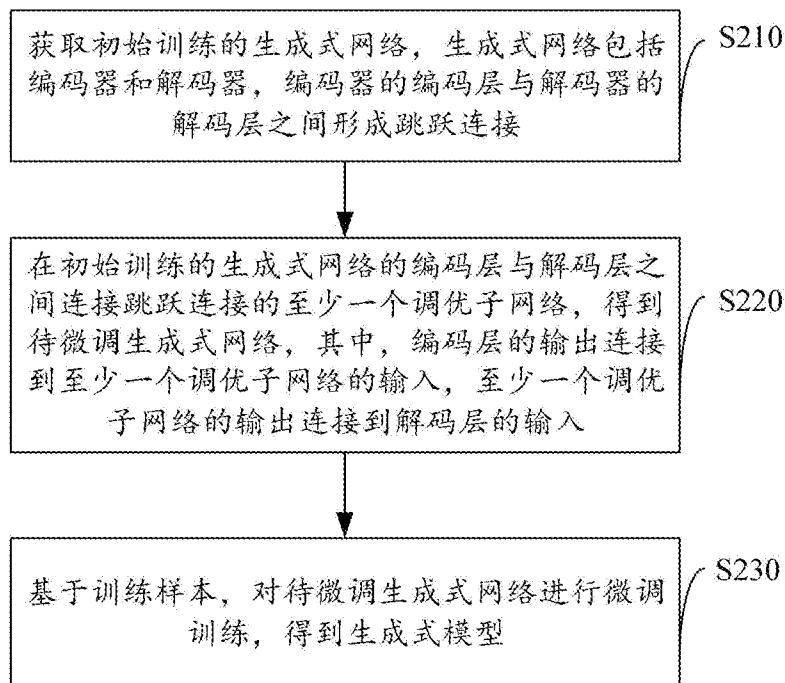
[图 2E]



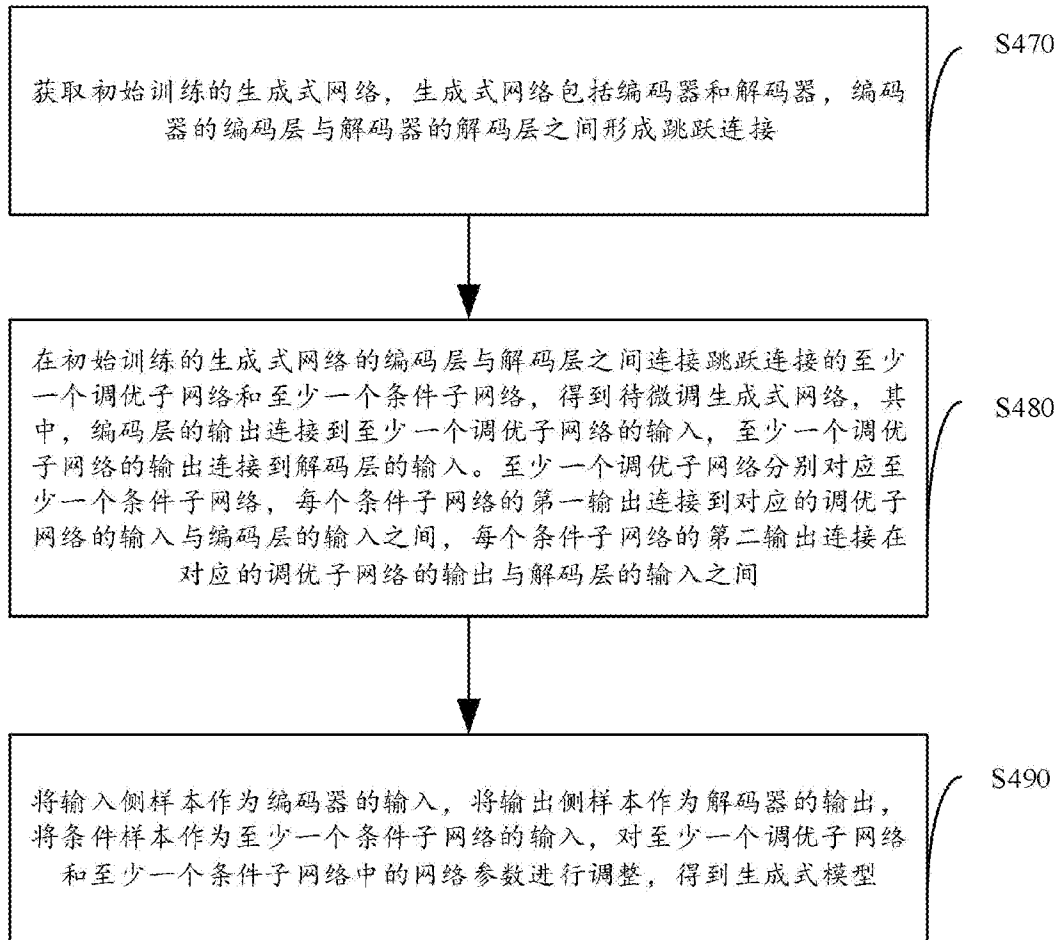
[图 3]



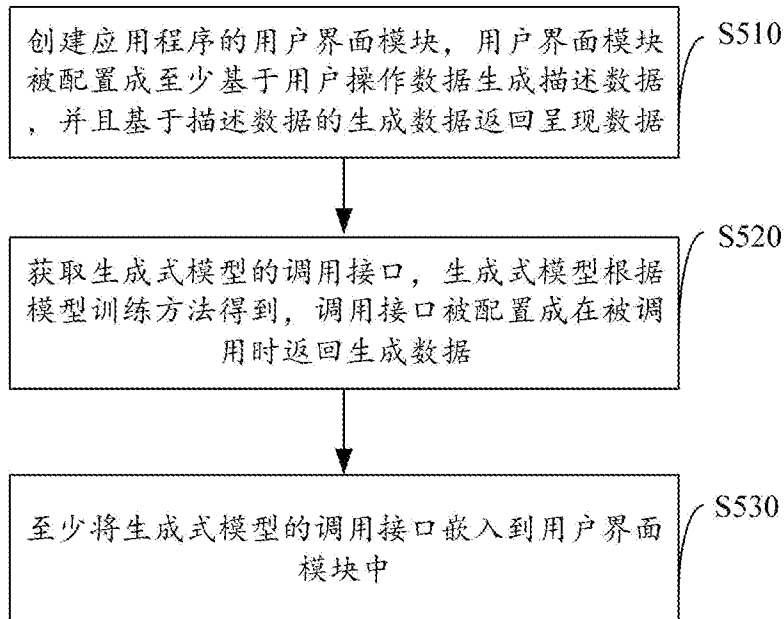
[图 4A]



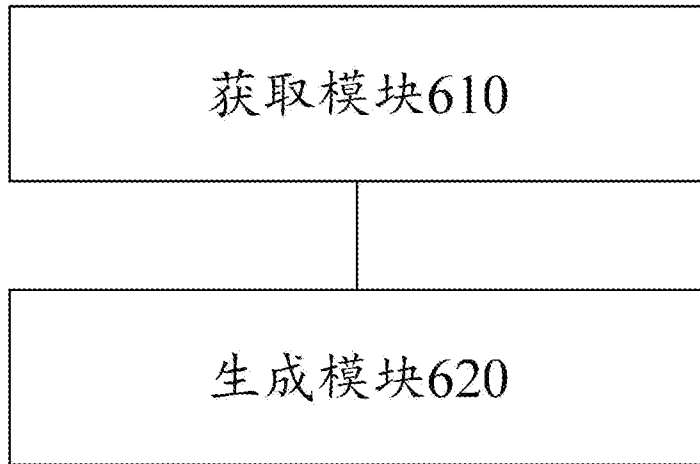
[图 4B]



[图 5]



[图 6]



[图 7]



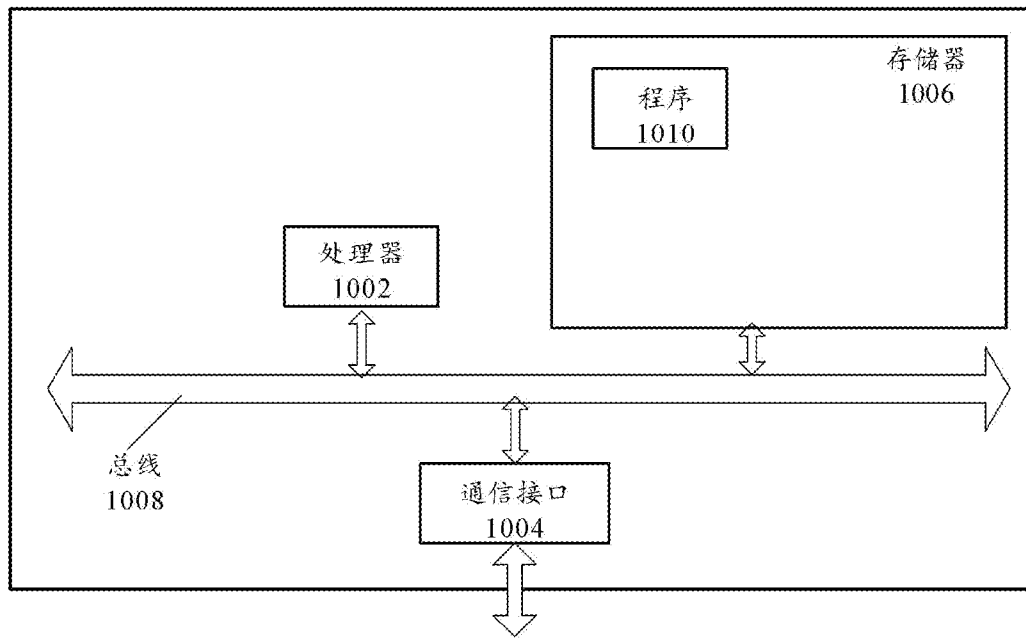
[图 8]



[图 9]



[图 10]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2024/113737

A. CLASSIFICATION OF SUBJECT MATTER G06N 3/0455(2023.01)i; G06N 3/0464(2023.01)i; G06N 3/0475(2023.01)i; G06N 3/09(2023.01)i; G06N 5/04(2023.01)i 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) IPC: G06N 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) CNTXT; CNABS; CNKI; VEN; USTXT; WOTXT; EPTXT: 生成式, 模型, 残差结构, 编码, 解码, 跳跃连接, 调优, 子网络, U-Net, generative, model, residual structure, coding, decoding, skip connection, tuning, sub-network		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 115482242 A (SOUTHWEST UNIVERSITY OF SCIENCE AND TECHNOLOGY) 16 December 2022 (2022-12-16) description, paragraphs [0003]-[0005]	1, 5, 6, 9-11, 14-17
A	CN 115482242 A (SOUTHWEST UNIVERSITY OF SCIENCE AND TECHNOLOGY) 16 December 2022 (2022-12-16) description, paragraphs [0003]-[0005]	2-4, 7, 8, 12, 13
A	CN 117114063 A (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.) 24 November 2023 (2023-11-24) entire document	1-17
A	CN 111612790 A (HANGZHOU DIANZI UNIVERSITY) 01 September 2020 (2020-09-01) entire document	1-17
☐ 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 22 October 2024		Date of mailing of the international search report 02 November 2024
Name and mailing address of the ISA/CN China National Intellectual Property Administration (ISA/CN) China No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088		Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2024/113737

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)		Publication date (day/month/year)
CN	115482242	A	16 December 2022	None		
CN	117114063	A	24 November 2023	None		
CN	111612790	A	01 September 2020	CN	111612790	B 17 October 2023

A. 主题的分类

G06N 3/0455(2023.01)i; G06N 3/0464(2023.01)i; G06N 3/0475(2023.01)i; G06N 3/09(2023.01)i; G06N 5/04(2023.01)i

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)

IPC: G06N

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))

CNTEXT;CNABS;CNKI;VEN;USTXT;WOTXT;EPTXT: 生成式, 模型, 残差结构, 编码, 解码, 跳跃连接, 调优, 子网络, U-Net, generative, model, residual structure, coding, decoding, skip connection, tuning, sub-network

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
X	CN 115482242 A (西南科技大学) 2022年12月16日 (2022 - 12 - 16) 说明书第[0003]-[0005]段	1, 5, 6, 9-11, 14-17
A	CN 115482242 A (西南科技大学) 2022年12月16日 (2022 - 12 - 16) 说明书第[0003]-[0005]段	2-4, 7, 8, 12, 13
A	CN 117114063 A (北京百度网讯科技有限公司) 2023年11月24日 (2023 - 11 - 24) 全文	1-17
A	CN 111612790 A (杭州电子科技大学) 2020年9月1日 (2020 - 09 - 01) 全文	1-17

☐ 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

2024年10月22日

国际检索报告邮寄日期

2024年11月2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

授权官员

徐晓艳

电话号码 (+86) 0512-88995723

PCT/ISA/210 表(第2页) (2022年7月)

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2024/113737

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	115482242	A	2022年12月16日	无	
CN	117114063	A	2023年11月24日	无	
CN	111612790	A	2020年9月1日	CN	111612790 B 2023年10月17日