(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 115409855 B
(45)授权公告日 2023.07.07

(21)申请号 202211147197.2
(22)申请日 2022.09.20
(65)同一申请的已公布的文献号
申请公布号 CN 115409855 A
(43)申请公布日 2022.11.29
(73)专利权人 北京百度网讯科技有限公司
地址 100085 北京市海淀区上地十街10号百度大厦2层
(72)发明人 王云浩 王晓迪 张滨 李超 辛颖 韩树民
(74)专利代理机构 中科专利商标代理有限责任公司 11021
专利代理师 李世阳
(51)Int.Cl.
G06T 7/11(2017.01)
G06T 5/00(2006.01)
G06T 7/187(2017.01)
(56)对比文件
US 2021375458 A1,2021.12.02
US 2021264190 A1,2021.08.26
CN 114708380 A,2022.07.05
CN 113674334 A,2021.11.19
WO 2022116771 A1,2022.06.09
CN 114550313 A,2022.05.27
CN 114743017 A,2022.07.12
US 2020074243 A1,2020.03.05
JP 2022063236 A,2022.04.21
CN 114092774 A,2022.02.25
雷鹏程;刘丛;唐坚刚;彭敦陆.分层特征融合注意力网络图像超分辨率重建.中国图象图形学报.2020,(09),全文.
赵小虎;尹良飞;赵成龙.基于全局-局部特征和自适应注意力机制的图像语义描述算法.浙江大学学报(工学版).2020,(01),全文.
兰红;刘秦邑.图注意力网络的场景图到图像生成模型.中国图象图形学报.2020,(08),全文.
袁晓冬;李超;盛浩.小波多分辨分量相关性图像融合方法.北京航空航天大学学报.2013,(06),全文.
审查员 舒婷

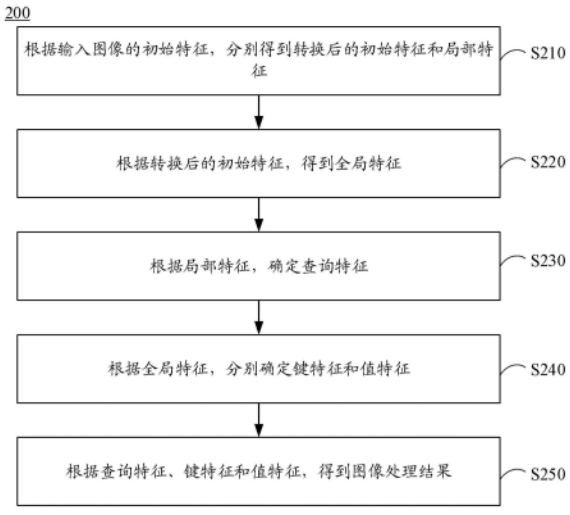
权利要求书4页 说明书14页 附图6页

(54)发明名称

图像处理方法、装置、电子设备和存储介质

(57)摘要

本公开提供了一种图像处理方法,涉及人工智能技术领域,尤其涉及计算机视觉技术领域和深度学习技术领域。具体实现方案为:根据输入图像的初始特征,分别得到转换后的初始特征和局部特征;根据转换后的初始特征,得到全局特征;根据局部特征,确定查询特征,其中,查询特征的元素数量与局部特征的元素数量之间的差异小于或等于预设差异阈值;根据全局特征,分别确定键特征和值特征;以及根据查询特征、键特征和值特征,得到图像处理结果。本公开还提供了一种图像处理装置、电子设备和存储介质。



CN 115409855 B

1.一种图像处理方法,包括:

根据输入图像的初始特征,分别得到转换后的初始特征和局部特征;

根据所述转换后的初始特征,得到全局特征;

根据所述局部特征,确定查询特征,其中,所述查询特征的元素数量与所述局部特征的元素数量之间的差异小于或等于预设差异阈值;

根据所述全局特征,分别确定键特征和值特征;

根据所述查询特征、所述键特征和所述值特征,得到注意力特征;以及

根据所述注意力特征,得到所述图像处理结果,

其中,所述初始特征为N级初始特征,所述N级初始特征中每级初始特征均为至少一个,N为大于1的整数,所述N级初始特征的第n级初始特征为M个,M为大于1的整数,n为大于或等于1且小于或等于N的整数,m为大于或等于1且小于M的整数,

其中,所述根据所述查询特征、所述键特征和所述值特征,得到注意力特征包括:根据第n级第m个查询特征、第n级第m个键特征和第n级第m个值特征,得到第n级第m个注意力特征,作为第n级第m+1个初始特征;

其中,所述根据输入图像的初始特征,分别得到转换后的初始特征和局部特征包括:对第n级第m+1个初始特征进行重组处理,得到第n级第m+1个重组后的初始特征;对所述第n级第m+1个重组后的初始特征进行下采样处理,得到第n级第m+1个转换后的初始特征;

其中,所述根据所述转换后的初始特征,得到全局特征包括:对第n级第m+1个转换后的初始特征进行双线性插值处理,得到第n级第m+1个插值后的初始特征;根据处理参数矩阵和第n级第m+1个转换后的初始特征,得到第n级第m+1个处理后初始特征;根据偏置参数、第n级第m+1个插值后初始特征和第n级第m+1个处理后初始特征,得到第n级第m+1个全局特征。

2.根据权利要求1所述的方法,其中,所述根据所述转换后的初始特征,得到全局特征还包括:

对所述转换后的初始特征进行插值处理,得到所述全局特征。

3.根据权利要求1所述的方法,其中,所述根据所述查询特征、所述键特征和所述值特征,得到注意力特征包括:

根据所述查询特征和所述键特征,得到融合特征;

对所述融合特征进行缩放,得到缩放后的融合特征;

利用预设函数处理所述缩放后的融合特征,得到处理后的融合特征;以及

根据所述值特征和所述处理后的融合特征,得到所述注意力特征。

4.根据权利要求1所述的方法,其中,所述根据所述局部特征,确定查询特征包括:

根据第一参数矩阵和所述局部特征,得到所述查询特征。

5. 根据权利要求1所述的方法,其中,所述根据所述全局特征,分别确定键特征和值特征包括:

根据第二参数矩阵和所述全局特征,得到所述键特征;以及

根据第三参数矩阵和所述全局特征,得到所述值特征。

6.根据权利要求1所述的方法,其中,所述根据输入图像的初始特征,分别得到转换后的初始特征和局部特征还包括:

根据第n级第m个初始特征，分别得到第n级第m个转换后的初始特征和第n级第m个局部特征。

7.根据权利要求6所述的方法，其中所述根据所述查询特征、所述键特征和所述值特征，得到注意力特征包括：

根据第n级第M个查询特征、第n级第M个键特征和第n级第M个值特征，得到第n级第M个注意力特征。

8.根据权利要求7所述的方法，其中，n为小于N的整数，所述根据输入图像的初始特征，分别得到转换后的初始特征和局部特征包括：

根据第n级第M个注意力特征，得到第n+1级第1个初始特征。

9.根据权利要求7所述的方法，其中，所述N级初始特征的第N级初始特征为J个，J为大于或等于1的整数，

所述根据所述注意力特征，得到所述图像处理结果包括：

根据第N级第J个注意力特征，得到所述图像处理结果。

10.根据权利要求1所述的方法，其中，所述图像处理结果包括目标对象检测结果、分类结果和分割结果中的至少一个。

11.根据权利要求1所述的方法，其中，所述局部特征的元素数量与所述初始特征的元素数量一致。

12.一种图像处理装置，包括：

第一获得模块，用于根据输入图像的初始特征，分别得到转换后的初始特征和局部特征；

第二获得模块，用于根据所述转换后的初始特征，得到全局特征；

第一确定模块，用于根据所述局部特征，确定查询特征，其中，所述查询特征的元素数量与所述局部特征的元素数量之间的差异小于或等于预设差异阈值；

第二确定模块，用于根据所述全局特征，分别确定键特征和值特征；以及

第三获得模块，用于根据所述查询特征、所述键特征和所述值特征，得到图像处理结果，

其中，所述第三获得模块包括：

第二获得子模块，用于根据所述查询特征、所述键特征和所述值特征，得到注意力特征；以及

第三获得子模块，用于根据所述注意力特征，得到所述图像处理结果，

其中，所述初始特征为N级初始特征，所述N级初始特征中每级初始特征均为至少一个，N为大于1的整数，所述N级初始特征的第n级初始特征为M个，M为大于1的整数，n为大于或等于1且小于或等于N的整数，m为大于或等于1且小于M的整数，

其中，所述第二获得子模块包括：第五获得单元，用于根据第n级第m个查询特征、第n级第m个键特征和第n级第m个值特征，得到第n级第m个注意力特征，作为第n级第m+1个初始特征；

其中，所述第一获得模块还用于：对第n级第m+1个初始特征进行重组处理，得到第n级第m+1个重组后的初始特征；对所述第n级第m+1个重组后的初始特征进行下采样处理，得到第n级第m+1个转换后的初始特征；

其中，所述第二获得模块还用于：对第n级第m+1个转换后的初始特征进行双线性插值处理，得到第n级第m+1个插值后的初始特征；根据处理参数矩阵和第n级第m+1个转换后的初始特征，得到第n级第m+1个处理后初始特征；根据偏置参数、第n级第m+1个插值后初始特征和第n级第m+1个处理后初始特征，得到第n级第m+1个全局特征。

13.根据权利要求12所述的装置，其中，所述第二获得模块包括：

插值子模块，用于对所述转换后的初始特征进行插值处理，得到所述全局特征。

14.根据权利要求12所述的装置，其中，所述第二获得子模块包括：

第三获得单元，用于根据所述查询特征和所述键特征，得到融合特征；

缩放单元，用于对所述融合特征进行缩放，得到缩放后的融合特征；

处理单元，用于利用预设函数处理所述缩放后的融合特征，得到处理后的融合特征；以及

第四获得单元，用于根据所述值特征和所述处理后的融合特征，得到所述注意力特征。

15.根据权利要求12所述的装置，其中，所述第一确定模块包括：

第一确定子模块，用于根据第一参数矩阵和所述局部特征，得到所述查询特征。

16. 根据权利要求12所述的装置，其中，所述第二确定子模块包括：

第二确定子模块，用于根据第二参数矩阵和所述全局特征，得到所述键特征；以及

第三确定子模块，用于根据第三参数矩阵和所述全局特征，得到所述值特征。

17.根据权利要求12所述的装置，其中，所述第一获得模块还用于：

根据第n级第m个初始特征，分别得到第n级第m个转换后的初始特征和第n级第m个局部特征。

18.根据权利要求17所述的装置，其中，所述第二获得子模块包括：

第六获得单元，用于根据第n级第M个查询特征、第n级第M个键特征和第n级第M个值特征，得到第n级第M个注意力特征。

19.根据权利要求18所述的装置，其中，n为小于N的整数，所述第一获得模块包括：

第五获得子模块，用于根据第n级第M个注意力特征，得到第n+1级第1个初始特征。

20.根据权利要求18所述的装置，其中，所述N级初始特征的第N级初始特征为J个，J为大于或等于1的整数，

所述第三获得子模块包括：

第七获得单元，用于根据第N级第J个注意力特征，得到所述图像处理结果。

21.根据权利要求12所述的装置，其中，所述图像处理结果包括目标对象检测结果、分类结果和分割结果中的至少一个。

22.根据权利要求12所述的装置，其中，所述局部特征的元素数量与所述初始特征的元素数量一致。

23. 一种电子设备，包括：

至少一个处理器；以及

与所述至少一个处理器通信连接的存储器；其中，

所述存储器存储有可被所述至少一个处理器执行的指令，所述指令被所述至少一个处理器执行，以使所述至少一个处理器能够执行权利要求1至11中任一项所述的方法。

24.一种存储有计算机指令的非瞬时计算机可读存储介质，其中，所述计算机指令用于

使所述计算机执行根据权利要求1至11中任一项所述的方法。

图像处理方法、装置、电子设备和存储介质

技术领域

[0001] 本公开涉及人工智能技术领域，尤其涉及计算机视觉技术领域和深度学习技术领域，可应用于目标检测、对象分类和图像分割等场景下。更具体地，本公开提供了一种图像处理方法、装置、电子设备和存储介质。

背景技术

[0002] 随着人工智能技术的发展，深度学习模型广泛地应用于执行图像识别、物体检测、图像分割等任务。基于应用在自然语言处理领域的Transformer模型，可以得到应用于计算机视觉领域的视觉Transformer(Vision Transformer)模型。在利用大量训练样本对视觉Transformer模型进行预训练之后，经预训练的视觉Transformer模型可以具有良好的图像识别性能。

发明内容

[0003] 本公开提供了一种图像处理方法、装置、设备以及存储介质。

[0004] 根据本公开的一方面，提供了一种图像处理方法，该方法包括：根据输入图像的初始特征，分别得到转换后的初始特征和局部特征；根据转换后的初始特征，得到全局特征；根据局部特征，确定查询特征，其中，查询特征的元素数量与局部特征的元素数量之间的差异小于或等于预设差异阈值；根据全局特征，分别确定键特征和值特征；以及根据查询特征、键特征和值特征，得到图像处理结果。

[0005] 根据本公开的另一方面，提供了一种图像处理装置，该装置包括：第一获得模块，用于根据输入图像的初始特征，分别得到转换后的初始特征和局部特征；第二获得模块，用于根据转换后的初始特征，得到全局特征；第一确定模块，用于根据局部特征，确定查询特征，其中，查询特征的元素数量与局部特征的元素数量之间的差异小于或等于预设差异阈值；第二确定模块，用于根据全局特征，分别确定键特征和值特征；以及第三获得模块，用于根据查询特征、键特征和值特征，得到图像处理结果。

[0006] 根据本公开的另一方面，提供了一种电子设备，包括：至少一个处理器；以及与至少一个处理器通信连接的存储器；其中，存储器存储有可被至少一个处理器执行的指令，指令被至少一个处理器执行，以使至少一个处理器能够执行根据本公开提供的方法。

[0007] 根据本公开的另一方面，提供了一种存储有计算机指令的非瞬时计算机可读存储介质，该计算机指令用于使计算机执行根据本公开提供的方法。

[0008] 根据本公开的另一方面，提供了一种计算机程序产品，包括计算机程序，该计算机程序在被处理器执行时实现根据本公开提供的方法。

[0009] 应当理解，本部分所描述的内容并非旨在标识本公开的实施例的关键或重要特征，也不用于限制本公开的范围。本公开的其它特征将通过以下的说明书而变得容易理解。

附图说明

[0010] 附图用于更好地理解本方案,不构成对本公开的限定。其中:

[0011] 图1是根据本公开的一个实施例的可以应用图像处理方法和装置的示例性系统架构示意图;

[0012] 图2是根据本公开的一个实施例的图像处理方法的流程图;

[0013] 图3A是根据本公开的一个实施例的局部全局Transformer模块的示意图;

[0014] 图3B是根据本公开的一个实施例的全局注意力单元的原理图;

[0015] 图4是根据本公开的一个实施例的图像处理模型的示意图;

[0016] 图5是根据本公开的一个实施例的图像处理方法的效果的示意图;

[0017] 图6是根据本公开的另一个实施例的图像处理方法的效果的示意图;

[0018] 图7是根据本公开的一个实施例的图像处理装置的框图;以及

[0019] 图8是根据本公开的一个实施例的可以应用图像处理方法的电子设备的框图。

具体实施方式

[0020] 以下结合附图对本公开的示范性实施例做出说明,其中包括本公开实施例的各种细节以助于理解,应当将它们认为仅仅是示范性的。因此,本领域普通技术人员应当认识到,可以对这里描述的实施例做出各种改变和修改,而不会背离本公开的范围和精神。同样,为了清楚和简明,以下的描述中省略了对公知功能和结构的描述。

[0021] 基于视觉Transformer模型的图像处理方法的应用范围在不断地增加。视觉Transformer模型可以应用于处理计算机视觉领域的大部分任务。

[0022] 视觉Transformer模型可以基于全局注意力机制进行图像处理。全局注意力的计算复杂度较高,在图像的分辨率较高的情况下,视觉Transformer模型需要消耗较高的硬件资源。可以利用局部注意力机制替代全局注意力机制,以提高视觉Transformer模型的效率。例如,可变窗口的Transformer(Shifted windows Transformer,Swin Transformer)模型可以基于不同的窗口以及窗口之间的移位来进行图像处理。然而,窗口之间的信息交互是不足的,可变窗口的Transformer模型的性能仍有不足。

[0023] 可以将Transformer模型同卷积结合,以提高视觉Transformer模型的效率。例如,卷积视觉Transformer(Convolutional vision Transformer,CvT)模型可以将Transformer模型的自注意力模块的线性投影转换为卷积投影。卷积自注意力网络(Convolution and self-Attention Networks,CoAtNets)可以将深度可分离卷积和自注意力机制统一起来,并可以将卷积层和自注意力层有效地堆叠起来。然而,卷积处理与Transformer模型的相关处理之间可能存在一些冲突。利用卷积处理可以获得局部特征,利用自注意力机制可以获得全局特征,但结合了卷积和Transformer模型的混合模型难以训练,也难以使用自监督的方式进行训练。

[0024] 图1是根据本公开一个实施例的可以应用图像处理方法和装置的示例性系统架构示意图。需要注意的是,图1所示仅为可以应用本公开实施例的系统架构的示例,以帮助本领域技术人员理解本公开的技术内容,但并不意味着本公开实施例不可以用于其他设备、系统、环境或场景。

[0025] 如图1所示,根据该实施例的系统架构100可以包括终端设备101、102、103,网络

104和服务器105。网络104用以在终端设备101、102、103和服务器105之间提供通信链路的介质。网络104可以包括各种连接类型，例如有线和/或无线通信链路等等。

[0026] 用户可以使用终端设备101、102、103通过网络104与服务器105交互，以接收或发送消息等。终端设备101、102、103可以是具有显示屏并且支持网页浏览的各种电子设备，包括但不限于智能手机、平板电脑、膝上型便携计算机和台式计算机等等。

[0027] 服务器105可以是提供各种服务的服务器，例如对用户利用终端设备101、102、103所浏览的网站提供支持的后台管理服务器(仅为示例)。后台管理服务器可以对接收到的用户请求等数据进行分析等处理，并将处理结果(例如根据用户请求获取或生成的网页、信息、或数据等)反馈给终端设备。

[0028] 需要说明的是，本公开实施例所提供的图像处理方法一般可以由服务器105执行。相应地，本公开实施例所提供的图像处理装置一般可以设置于服务器105中。本公开实施例所提供的图像处理方法也可以由不同于服务器105且能够与终端设备101、102、103和/或服务器105通信的服务器或服务器集群执行。相应地，本公开实施例所提供的图像处理装置也可以设置于不同于服务器105且能够与终端设备101、102、103和/或服务器105通信的服务器或服务器集群中。

[0029] 图2是根据本公开的一个实施例的图像处理方法的流程图。

[0030] 如图2所示，该方法200可以包括操作S210至操作S250。

[0031] 在操作S210，根据输入图像的初始特征，分别得到转换后的初始特征和局部特征。

[0032] 在本公开实施例中，可以对输入图像进行特征提取，得到初始特征。例如，可以利用块融合(Patch Merging)操作对输入图像进行特征提取。又例如，可以对输入图像进行N级特征提取，得到N级初始特征。可以理解，N可以为大于或等于1的整数。

[0033] 在本公开实施例中，可以利用局部注意力机制(Local Attention)处理初始特征，得到局部特征。例如，基于窗口注意力机制(Windows Attention)，可以利用可变窗口的Transformer模型处理初始特征，得到局部特征。又例如，基于十字型窗口注意力机制(Cross-Shaped Window self-attention)，可以利用十字型窗口的Transformer(Cross-Shaped Window self-attention Transformer，CSwin Transformer)模型处理初始特征，得到局部特征。局部特征可以是输入图像的细粒度的特征表达。

[0034] 在本公开实施例中，可以利用各种方式对初始特征进行转换，得到转换后的初始特征。例如，可以利用重组(Reshape)操作对初始特征进行转换，以便得到转换后的初始特征。又例如，也可以对初始特征进行下采样，以便得到转换后的初始特征。

[0035] 在操作S220，根据转换后的初始特征，得到全局特征。

[0036] 在本公开实施例中，根据转换后的初始特征，可以利用各种方式，得到全局特征。例如，基于全局注意力机制处理转换后的初始特征，可以得到全局特征。

[0037] 在操作S230，根据局部特征，确定查询特征。

[0038] 例如，根据局部特征和第一参数信息，可以确定查询特征。

[0039] 在本公开实施例中，查询特征的元素数量与局部特征的元素数量之间的差异小于或等于预设差异阈值。例如，查询特征的元素数量可以与局部特征的元素数量一致。

[0040] 在操作S240，根据全局特征，分别确定键特征和键特征。

[0041] 例如，根据全局特征和第二参数信息，可以确定键特征。又例如，根据全局特征和

第三参数信息,可以确定值特征。

[0042] 在操作S250,根据查询特征、键特征和值特征,得到图像处理结果。

[0043] 在本公开实施例中,可以将查询特征、键特征和值特征融合,再利用全连接模块得到图像处理结果。例如,图像处理结果可以为目标检测结果。又例如,可以利用各种方式将查询特征至值特征融合。

[0044] 通过本公开实施例,根据初始特征,分别获得了局部特征和全局特征。即,利用双分支结构分别获取了局部特征和全局特征。可以充分地获取图像的局部和全局的信息。根据局部特征确定的查询特征,二者之间元素数量的差异较小,使得查询特征可以保留与图像相关的更多的信息,以提高图像处理精度。

[0045] 在本公开的另一些实施例中,查询特征元素数量和局部特征的元素数量之间可以存在差异。在本公开实施例中,查询特征元素数量和局部特征的元素数量之间的差异可以小于预设差异阈值。预设差异阈值可以是一个较小的值。例如,查询特征元素数量可以略大于局部特征的元素数量。又例如,查询特征元素数量可以略小于局部特征的元素数量。

[0046] 下面将结合相关实施例对本公开提供的图像处理方法的原理进行详细说明。在本公开实施例中,可以利用一个局部全局Transformer模块(Local Global Transformer Block,LoGo Transformer Block)来实现本公开提供的方法。

[0047] 图3A是根据本公开的一个实施例的局部全局Transformer模块的示意图。

[0048] 如图3A所示,局部全局Transformer模块可以包括局部注意力单元311、全局注意力单元312和交叉注意力单元313。

[0049] 图3B是根据本公开的一个实施例的全局注意力单元的原理图。

[0050] 在一些实施例中,在例如上述的操作S210的一些实施方式中,可以根据初始特征,得到局部特征。例如,如图3A所示,可以利用局部注意力单元311处理初始特征X 301,得到局部特征X_L。初始特征X 301的尺寸可以为h×w×c,局部特征X_L的尺寸可以为c×L_local。L_local可以等于h×w。可以理解,局部注意力单元311可以作为上述双分支结构的一个分支。

[0051] 在本公开实施例中,局部特征的元素数量可以同初始特征的元素数量一致。例如,初始特征可以包括h×w×c个元素,局部特征可以包括h×w×c个元素,二者的元素数量一致,使得局部特征可以保留初始特征的大部分信息,以便提高图像处理的精度。

[0052] 在一些实施例中,在上述的操作S210的一些实施方式中,可以根据初始特征,得到全局特征。例如,可以利用全局注意力单元312处理初始特征X 301,得到全局特征X_G。可以理解,全局注意力单元312可以作为上述的双分支结构的一个分支。下面将进行详细说明。

[0053] 在本公开实施例中,根据初始特征,可以得到重组后的初始特征。对重组后的初始特征进行下采样处理,可以得到转换后的初始特征。例如,如图3B所示,初始特征X 301的尺寸可以为h×w×c。可以对初始特征X301进行重组处理,得到重组后的初始特征X_r 303。重组后的初始特征X_r 303的尺寸可以为c×Li,Li可以等于h×w。在一个示例中,重组后的初始特征X_r 303可以包括h×w个长度为c的子特征。又例如,可以对重组后的初始特征X_r 303进行下采样处理,得到转换后的初始特征X_t304。转换后的初始特征X_t 304的尺寸可以为c×Lo,Lo可以等于h×w÷i。在一个示例中,转换后的初始特征X_t 304可以包括h×w÷i个长度为c的子特征。在一个示例中,h可以为224,w可以为224,c可以为3,i可以为14。h、

w、c、i均为大于1的整数。可以理解，在图像处理过程中，局部注意力单元311和全局注意力单元312可以对初始特征进行一次或多次重组(Reshape)操作，图3A和图3B中初始特征X 301的形状仅为示意。

[0054] 在一些实施例中，在上述的操作S220的一些实施方式中，可以对转换后的初始特征进行插值处理，得到全局特征。

[0055] 在本公开实施例中，对转换后的初始特征进行双线性插值处理，可以得到插值后的初始特征。例如，可以对转换后的初始特征X_t 304进行双线性插值处理，得到插值后的初始特征。

[0056] 在本公开实施例中，根据处理参数矩阵和转换后的初始特征，可以得到处理后的初始特征。例如，基于矩阵乘法操作，可以将处理参数矩阵A_T和转换后的初始特征X_t 304相乘，可以得到处理后的初始特征。

[0057] 在本公开实施例中，根据偏置参数、插值后的初始特征和处理后的初始特征，得到全局特征X_G。例如，可以通过以下公式得到全局特征：

[0058] X_G＝interpolation(X_t)+X_t×A_T+b (公式一)

[0059] interpolation(X_t)可以为插值后的初始特征。X_t×A_T可以为处理后的初始特征。B可以为偏置参数。全局特征X_G的尺寸可以c×L_global。上述的L_local可以为L_global的i倍。

[0060] 通过本公开实施例，利用初始特征，分别得到了局部特征和全局特征，可以充分地获取局部特征的token(标记)与全局特征的token之间的相关性，有助于提升局部特征的表达能力，可以自适应地利用局部特征与全局特征之间的关系进行训练或推理。此外，通过本公开实施例，利用初始特征，分别得到了局部特征和全局特征，有助于实现局部特征和全局特征之间的高度对齐和耦合。此外，通过本公开实施例，对转换后的初始特征进行了插值处理，与视觉Transformer模型的全局注意力机制相比，获取全局特征的速度更快，提高了局部全局Transformer模块的处理效率。

[0061] 在一些实施例中，在上述的操作S230的一些实施方式中，根据局部特征，确定查询特征可以包括：根据第一参数矩阵和局部特征，确定查询特征。例如，上述的第一参数信息可以实现为第一参数矩阵。例如，基于矩阵乘法操作，将第一参数矩阵Wq_local和局部特征X_L相乘，得到查询特征Q_L。在一个示例中，可以通过以下公式得到查询特征：

[0062] Q_L＝X_L×Wq_local (公式二)

[0063] 通过本公开实施例，利用局部特征确定了查询特征，如上述，局部特征的元素数量与初始特征的元素数量一致，使得局部特征保留初始特征的大部分信息。由此，若查询特征的元素数量与局部特征的元素数量一致，查询特征也可以保留初始特征的大部分信息。

[0064] 在一些实施例中，在上述的操作S240的一些实施方式中，根据全局特征，分别确定键特征和值特征可以包括：根据第二参数矩阵和全局特征，确定键特征。例如，上述的第二参数信息可以实现为第二参数矩阵。例如，基于矩阵乘法操作，将第二参数矩阵Wk_global和全局特征X_G相乘，得到键特征K_G。在一个示例中，可以通过以下公式得到键特征：

[0065] K_G＝X_G×Wk_global (公式三)

[0066] 在一些实施例中，在上述的操作S240的一些实施方式中，根据全局特征，分别确定键特征和值特征可以包括：根据第三参数矩阵和全局特征，确定值特征。例如，上述的第三

参数信息可以实现为第三参数矩阵。例如，基于矩阵乘法操作，将第三参数矩阵Wv_global和全局特征X_G相乘，得到值特征V_G。在一个示例中，可以通过以下公式得到值特征：

[0067] V_G＝X_G×Wv_global （公式四）

[0068] 在一些实施例中，在上述的操作S250的一些实施方式中，根据查询特征、键特征和值特征，得到图像处理结果可以包括：根据查询特征、键特征和值特征，可以得到注意力特征。

[0069] 在本公开实施例中，根据查询特征、键特征和值特征，可以得到注意力特征可以包括：根据查询特征和键特征，得到融合特征。对融合特征进行缩放，得到缩放后的融合特征。利用预设函数处理缩放后的融合特征，得到处理后的融合特征。根据值特征和处理后的融合特征，得到注意力特征。

[0070] 通过本公开实施例，利用查询特征和键特征得到了融合特征，进而得到了注意力特征，可以使得注意力特征的元素数量与查询特征的元素数量保持一致，进而可以使得注意力特征的元素数量与初始特征的元素数量保持一致。进而，注意力特征可以高效地从初始特征中获取有效的信息，有助于提高模型的处理效率和精度。

[0071] 例如，可以通过以下公式得到注意力特征Cross_Att 302：

[0072] $$Cross\_Att = softmax(\frac{Q\_L \times K^T\_G}{\sqrt{d}})V\_G$$ （公式五）

[0073] softmax()为预设函数。K^T_G为键特征的转置。$\sqrt{d}$为缩放参数。

[0074] 可以理解，Q_L×K^T_G可以作为融合特征。$\frac{Q_L \times K^T_G}{\sqrt{d}}$可以作为缩放后的融合特征。$softmax(\frac{Q_L \times K^T_G}{\sqrt{d}})$可以作为处理后的融合特征。

[0075] 在一些实施例中，在上述的操作S250的一些实施方式中，根据查询特征、键特征和值特征，得到图像处理结果还可以包括：根据注意力特征，得到图像处理结果。例如，可以利用一个全连接网络处理注意力特征Cross_Att，以便得到图像处理结果。

[0076] 在本公开实施例中，图像处理结果可以为目标检测结果、分类结果和分割结果中的至少一个。例如，可以利用不同的全连接层处理注意力特征，得到目标检测结果、分类结果或分割结果。

[0077] 可以理解，上文结合1个局部全局Transformer模块对本公开的方法进行了详细说明。但本公开中，局部全局Transformer模块的数量可以为多个，下面将结合相关实施例进行详细说明。

[0078] 在一些实施例中，初始特征可以为N级初始特征。N可以为大于1的整数。下面将结合图4进行详细说明。

[0079] 图4是根据本公开的一个实施例的图像处理模型的示意图。

[0080] 如图4所示，图像处理模型400可以包括4个图像处理级。4个图像处理级可以分别为第1个图像处理级Stage_400_1、第2个图像处理级Stage_400_2、第3个图像处理级Stage_400_3以及第4个图像处理级Stage_400_4。第1个图像处理级Stage_400_1可以包括M个局部全局Transformer模块410_1。第2个图像处理级Stage_400_2可以包括J个局部全局Transformer模块410_2。可以理解，第1个图像处理级Stage_400_1可以处理第1级初始特

征。第2个图像处理级Stage_400_2可以处理第2级初始特征。本实施例中，N＝2。又例如，M为大于1的整数，J为大于1的整数。也可以理解，上述关于局部全局Transformer模块310的详细描述，同样适用于本实施例的M个局部全局Transformer模块410_1中的每个模块和J个局部全局Transformer模块410_2中的每个模块。

[0081] 在一些实施例中，N级初始特征中每级初始特征均为至少一个。

[0082] 在本公开实施例中，N级初始特征的第n级初始特征可以为M个。例如，M个局部全局Transformer模块410_1可以处理第1级的M个初始特征。在一个示例中，M可以为2。m的取值可以分别为1、2。

[0083] 在本公开实施例中，N级初始特征的第N级初始特征可以为J个。例如，J个局部全局Transformer模块410_2可以处理第2级的J个初始特征。在一个示例中，J可以为2。j的取值可以分别为1、2。

[0084] 在一些实施例中，在上述的操作S210的另一些实施方式中，根据输入图像的初始特征，分别得到全局特征和局部特征可以包括：根据第n级第m个初始特征，可以分别得到第n级第m个全局特征和第n级第m个局部特征。

[0085] 如图4所示，利用块融合模块431对输入图像进行块融合(Patch Merging)处理，可以得到第1级第1个初始特征X1_1。可以利用M个局部全局Transformer模块410_1的第1个局部全局Transformer模块处理第1级第1个初始特征X1_1。

[0086] 在一些实施例中，可以利用第1个局部全局Transformer模块的局部注意力单元处理第1级第1个初始特征X1_1，得到第1级第1个局部特征X1_L1。第1级第1个局部特征X1_L1的元素数量可以与第1级第1个初始特征X1_1的元素数量一致。

[0087] 在本公开实施例中，可以利用第1个局部全局Transformer模块的全局注意力单元处理第1级第1个初始特征X1_1，得到第1级第1个全局特征X1_G1。例如，可以对第1级第1个初始特征X1_1进行重组处理，得到第1级第1个重组后的初始特征X1_r1。可以对第1级第1个重组后的初始特征X1_r1进行下采样处理，得到第1级第1个转换后的初始特征X1_t1。可以对第1级第1个转换后的初始特征X1_t1进行双线性插值处理，得到第1级第1个插值后的初始特征。基于矩阵乘法操作，可以将处理参数矩阵A_T和第1级第1个转换后的初始特征X1_t1相乘，可以得到第1级第1个处理后的初始特征。根据偏置参数、第1级第1个插值后的初始特征和第1级第1个处理后的初始特征，利用上述的公式一，可以得到第1级第1个全局特征X1_G1。

[0088] 在本公开实施例中，可以利用第1个局部全局Transformer模块的交叉注意力单元处理第1级第1个全局特征X1_G1和第1级第1个局部特征X1_L1。

[0089] 例如，基于矩阵乘法操作，将第一参数矩阵Wq_local和第1级第1个局部特征X1_L1相乘，得到第1级第1个查询特征Q1_L1。第1级第1个查询特征Q1_L1的元素数量可以与第1级第1个局部特征X1_L1的元素数量一致。

[0090] 例如，基于矩阵乘法操作，将第二参数矩阵Wk_global和第1级第1个全局特征X1_G1相乘，得到第1级第1个键特征K1_G1。

[0091] 例如，基于矩阵乘法操作，将第三参数矩阵Wv_global和第1级第1个全局特征X1_G1相乘，得到第1级第1个值特征V1_G1。

[0092] 在本公开实施例中，根据第n级第m个查询特征、第n级第m个键特征和第n级第m个

值特征，可以得到第n级第m个注意力特征，作为第n级第m+1个初始特征。例如，在n＝1且m＝1的情况下，根据第1级第1个查询特征Q1_L1和第1级第1个键特征K1_G1，可以得到第1级第1个融合特征。对第1级第1个融合特征进行缩放，可以得到第1级第1个缩放后的融合特征。利用预设函数处理第1级第1个缩放后的融合特征，可以得到第1级第1个处理后的融合特征。根据第1级第1个值特征V1_G1和第1级第1个处理后的融合特征，可以得到第1级第1个注意力特征Cross_Attl_1。通过本公开实施例，第1级第1个注意力特征Cross_Attl_1的元素数量可以与第1级第1个查询特征Q1_L1的元素数量一致，也可以与第1级第1个初始特征X1_1的元素数量一致。

[0093] 可以将第1级第1个注意力特征Cross_Attl_1作为第1级第2个初始特征X1_2。又例如，如上述，在M＝2的情况下，第1级第2个初始特征X1_2可以为第1级第M个初始特征X1_M。

[0094] 在一些实施例中，可以利用M个局部全局Transformer模块410_1的第M个局部全局Transformer模块处理第1级第M个初始特征X1_M。

[0095] 在本公开实施例中，可以利用第M个局部全局Transformer模块的局部注意力单元处理第1级第M个初始特征X1_M，得到第1级第M个局部特征X1_LM。第1级第M个局部特征X1_LM的元素数量可以和第1级第M个初始特征X1_M的元素数量一致。由此，第1级第M个局部特征X1_LM的元素数量可以与第1级第1个初始特征X1_1的元素数量一致。

[0096] 在本公开实施例中，可以利用第M个局部全局Transformer模块的全局注意力单元处理第1级第M个初始特征X1_M。例如，可以对第1级第M个初始特征X1_M进行重组处理，得到第1级第M个重组后的初始特征X1_rM。可以理解，可以对第1级第M个初始特征X1_M进行至少一次重组处理。接下来，可以对第1级第M个重组后的初始特征X1_rM进行下采样处理，得到第1级第M个转换后的初始特征X1_tM。可以对第1级第M个转换后的初始特征X1_tM进行双线性插值处理，得到第1级第M个插值后的初始特征。基于矩阵乘法操作，可以将处理参数矩阵A_T和第1级第M个转换后的初始特征X1_tM相乘，可以得到第1级第M个处理后的初始特征。根据偏置参数、第1级第M个插值后的初始特征和第1级第M个处理后的初始特征，利用上述的公式一，可以得到全局特征X1_GM。

[0097] 在本公开实施例中，可以利用第M个局部全局Transformer模块的交叉注意力单元处理第1级第M个全局特征X1_GM和第1级第M个局部特征X1_LM。

[0098] 例如，基于矩阵乘法操作，可以将第一参数矩阵Wq_local和第1级第M个局部特征X1_LM相乘，得到第1级第M个查询特征Q1_LM。第1级第M个查询特征Q1_LM的元素数量可以与第1级第M个局部特征X1_LM的元素数量一致。

[0099] 例如，基于矩阵乘法操作，可以将第二参数矩阵Wk_global和第1级第M个全局特征X1_GM相乘，得到第1级第M个键特征K1_GM。

[0100] 例如，基于矩阵乘法操作，可以将第三参数矩阵Wv_global和第1级第M个全局特征X1_GM相乘，得到第1级第M个值特征V1_GM。

[0101] 在本公开实施例中，根据第n级第M个查询特征、第n级第M个键特征和第n级第M个值特征，可以得到第n级第M个注意力特征。例如，根据第1级第M个查询特征Q1_LM和第1级第M个键特征K1_GM，可以得到第1级第M个融合特征。对第1级第M个融合特征进行缩放，可以得到第1级第M个缩放后的融合特征。利用预设函数处理第1级第M个缩放后的融合特征，可以得到第1级第M个处理后的融合特征。根据第1级第M个值特征V1_GM和第1级第M个处理后的

融合特征，可以得到第1级第M个注意力特征Cross_Att1_M。通过本公开实施例，第1级第M个注意力特征Cross_Att1_M的元素数量可以与第1级第M个查询特征Q1_LM的元素数量一致，也可以与第1级第M个初始特征X1_M的元素数量一致，进而与第1级第1个初始特征的元素数量一致。由此，在第1个图像处理级中，可以保留第1级第1个初始特征的大部分信息，以便提高图像处理的精度。

[0102] 在一些实施例中，第1级第M个注意力特征Cross_Att1_M可以作为第1个图像处理级Stage_400_1的输出。下面将利用第2个图像处理级Stage_400_2处理第1级第M个注意力特征Cross_Attl_M。

[0103] 在一些实施例中，在n小于N的情况下，根据输入图像的初始特征，分别得到全局特征和局部特征可以包括：根据第n级第M个注意力特征，得到第n+1级第1个初始特征。例如，利用第2个图像处理级Stage_400_2的块融合模块432处理第1级第M个注意力特征Cross_Att1_M，可以得到第2级第1个初始特征。此外，如上述，N＝2。第2级第1个初始特征也可以为第N级第1个初始特征。可以理解，在经块融合模块432处理之后，第N级第1个初始特征的元素数量与第1级的2个初始特征的元素数量不一致。

[0104] 可以理解，J个局部全局Transformer模块410_2对第2级J个初始特征的处理方式，与M个局部全局Transformer模块410_1对第1级的M个初始特征的处理方式相同或类似，本公开在此不再赘述。

[0105] 在本公开实施例中，在利用J个局部全局Transformer模块410_2处理第N级第1个初始特征之后，可以得到第N级第J个注意力特征Cross_Att2_J。

[0106] 可以理解，上文对第1个图像处理级Stage_400_1和第2个图像处理级Stage_400_2的原理进行了详细说明。接下来，将对第3个图像处理级Stage_400_3和第4个图像处理级Stage_400_4的原理进行详细说明。

[0107] 可以理解，与第1级的M个初始特征相比，第N级第J个注意力特征的尺寸较小。利用基于全局注意力机制的视觉Transformer模块处理第N级第J个注意力特征，也可以获得较高的处理效率，同时获得较高的处理精度。

[0108] 例如，利用第3个图像处理级Stage_400_3的块融合模块433处理第N级第J个注意力特征Cross_Att2_J，可以得到第3级第1个初始特征。接下来，可以利用Y个视觉Transformer模块420_1处理第3级第1个初始特征，得到第3级第Y个全局注意力特征G_Att3_Y。在一个示例中，Y＝6。

[0109] 又例如，利用第4个图像处理级Stage_400_4的块融合模块434处理第3级第Y个注意力特征G_Att3_Y，可以得到第4级第1个初始特征。接下来，可以利用Z个视觉Transformer模块420_2处理第4级第1个初始特征，得到第4级第Z个全局注意力特征G_Att4_Z。在一个示例中，Z＝2。

[0110] 又例如，可以利用一个全连接网络处理第4级第Z个全局注意力特征G_Att4_Z，得到图像处理结果。

[0111] 通过本公开实施例，利用多个局部全局Transformer模块处理输入图像的初始特征，可以更加充分地获取图像的局部和全局的信息。在特征的尺寸下降之后，基于全局注意力机制进一步处理输入图像的特征，可以在保证模型处理效率的情况下，进一步提高模型的精度。

[0112] 可以理解，上述的多个局部全局Transformer模块之间的处理参数矩阵、第一参数矩阵、第二参数矩阵、第三参数矩阵等可以是一致的。但本公开不限于此，多个局部全局Transformer模块之间的处理参数矩阵、第一参数矩阵、第二参数矩阵、第三参数矩阵等也可以是不一致的。每个局部全局Transformer模块的各个参数矩阵可以在训练阶段确定。

[0113] 可以理解，上文以图像处理模型包括两个由局部全局Transformer模块构成的图像处理级为示例，对本公开的模型进行了详细说明。但本公开不限于此，图像处理模型的4个图像处理级均可以由局部全局Transformer模块构成。

[0114] 可以理解，图像处理模型可以包括任意数量的图像处理级，本公开对此不进行限制。

[0115] 可以理解，上文对本公开的图像处理方法进行了详细说明，下面将结合相关实施例对本公开的图像处理方法的效果进行详细说明。

[0116] 图5是根据本公开的一个实施例的图像处理方法的效果的示意图。

[0117] 基于ImageNet-1k数据集，利用本公开的图像处理模型执行分类任务，得到分类结果，并确定本公开的图像处理模型的分类结果的精度。

[0118] 如图5所示，LoGo-S、LoGo-B、LoGo-L分别为小型图像处理模型、中型图像处理模型和大型图像处理模型。Train Size为训练样本尺寸。Test Size为测试样本尺寸。Params(M)为单位是百万的参数量。FLOPs(G)为单位是十亿次的每秒所执行的浮点运算次数(FLoating-point OPerations per second)。

[0119] 可以将本公开的图像处理模块同以下模型进行比较：高效数据图像Transformer(Data-efficient image Transformer,DeiT)模型、可变窗口的Transformer模型、十字型窗口的Transformer模型、卷积自注意力网络、焦点稀疏卷积网络(Focal sparse convolutional networks,Focal)模型、混编Transformer(Shuffle Transformer,Shuffle)模型、基于交叉注意力机制的视觉Transformer(Cross-Attention multi-scale Vision Transformer,CrossVit)以及基于标签的视觉Transformer模型(LV-ViT)。

[0120] 如图5所示，LoGo-S、LoGo-B、LoGo-L的分类结果精度分别为83.7％、85.0％、85.9％。与参数量接近的不同模型相比，LoGo-S、LoGo-B、LoGo-L均获得了当前模型最佳效果(State Of The Art,SOTA)。

[0121] 图6是根据本公开的另一个实施例的图像处理方法的效果的示意图。

[0122] 基于另一数据集，利用本公开的图像处理模型执行目标检测任务，得到目标检测结果，并确定本公开的图像处理模型的目标检测结果的精度。例如，可先利用ImageNet-1k数据集对本公开的图像处理模型进行预训练，再利用预训练后的图像处理模型执行目标检测任务。

[0123] 可以将本公开的图像处理模块同以下模型进行比较：残差网络(Residual Networks,Res)、金字塔型视觉Transformer(Pyramid Vision Transformer,PVT)模型、车道实例检测(Video Instance Lane Detection,ViL)模型、可变窗口的Transformer模型以及十字型窗口的Transformer模型等。

[0124] 在执行目标检测任务时，LoGo-S、LoGo-B、LoGo-L的检测框平均精度(bounding box mean Average Precision,bbox mAP)分别为47.0％、49.6％、50.7％。与参数量接近的不同模型相比，LoGo-S、LoGo-B、LoGo-L均获得了当前模型最佳效果。

[0125] 此外，在执行分割任务时，LoGo-S、LoGo-B、LoGo-L的分割平均精度(segment mean Average Precision，segm mAP)分别为42.7%、44.6%、45.4%。与参数量接近的不同模型相比，LoGo-S、LoGo-B、LoGo-L均获得了当前模型最佳效果。

[0126] 图7是根据本公开的一个实施例的图像处理装置的框图。

[0127] 如图7所示，该装置700可以包括第一获得模块710、第二获得模块720、第一确定模块730、第二确定模块740和第三获得模块750。

[0128] 第一获得模块710，用于根据输入图像的初始特征，分别得到转换后的初始特征和局部特征。

[0129] 第二获得模块720，用于根据转换后的初始特征，得到全局特征。

[0130] 第一确定模块730，用于根据局部特征，确定查询特征。

[0131] 第二确定模块740，用于根据全局特征，分别确定键特征和值特征。例如，查询特征的元素数量与局部特征的元素数量之间的差异小于或等于预设差异阈值。

[0132] 第三获得模块750，用于根据查询特征、键特征和值特征，得到图像处理结果。

[0133] 在一些实施例中，第二获得模块包括：插值子模块，用于对转换后的初始特征进行插值处理，得到全局特征。

[0134] 在一些实施例中，第一获得模块包括：第一获得子模块，用于根据初始特征，得到重组后的初始特征；以及下采样子模块，用于对重组后的初始特征进行下采样处理，得到转换后的初始特征。

[0135] 在一些实施例中，插值子模块包括：双线性插值单元，用于对转换后的初始特征进行双线性插值处理，得到插值后的初始特征；第一获得单元，用于根据处理参数矩阵和转换后的初始特征，得到处理后的初始特征；以及第二获得单元，用于根据偏置参数、插值后的初始特征和处理后的初始特征，得到全局特征。

[0136] 在一些实施例中，第三获得单元包括：第二获得子模块，用于根据查询特征、键特征和值特征，得到注意力特征；以及第三获得子模块，用于根据注意力特征，得到图像处理结果。

[0137] 在一些实施例中，第二获得子模块包括：第三获得单元，用于根据查询特征和键特征，得到融合特征；缩放单元，用于对融合特征进行缩放，得到缩放后的融合特征；处理单元，用于利用预设函数处理缩放后的融合特征，得到处理后的融合特征；以及第四获得单元，用于根据值特征和处理后的融合特征，得到注意力特征。

[0138] 在一些实施例中，第一确定模块包括：第一确定子模块，用于根据第一参数矩阵和局部特征，得到查询特征。

[0139] 在一些实施例中，第二确定子模块包括：第二确定子模块，用于根据第二参数矩阵和全局特征，得到键特征；以及第三确定子模块，用于根据第三参数矩阵和全局特征，得到值特征。

[0140] 在一些实施例中，初始特征为N级初始特征，N级初始特征中每级初始特征均为至少一个，N为大于1的整数，N级初始特征的第n级初始特征为M个，M为大于1的整数，n为大于或等于1且小于或等于N的整数，m为大于或等于1且小于或等于M的整数，第一获得模块包括：第四获得子模块，用于根据第n级第m个初始特征，分别得到第n级第m个全局特征和第n级第m个局部特征。

[0141] 在一些实施例中，m为小于M的整数，第二获得子模块包括：第五获得单元，用于根据第n级第m个查询特征、第n级第m个键特征和第n级第m个值特征，得到第n级第m个注意力特征，作为第n级第m+1个初始特征；第六获得单元，用于根据第n级第M个查询特征、第n级第M个键特征和第n级第M个值特征，得到第n级第M个注意力特征。

[0142] 在一些实施例中，n为小于N的整数，第一获得模块包括：第五获得子模块，用于根据第n级第M个注意力特征，得到第n+1级第1个初始特征。

[0143] 在一些实施例中，N级初始特征的第N级初始特征为J个，J为大于或等于1的整数，第三获得子模块包括：第七获得单元，用于根据第N级第J个注意力特征，得到图像处理结果。

[0144] 在一些实施例中，图像处理结果包括目标对象检测结果、分类结果和分割结果中的至少一个。

[0145] 在一些实施例中，局部特征的元素数量与初始特征的元素数量一致。

[0146] 本公开的技术方案中，所涉及的用户个人信息的收集、存储、使用、加工、传输、提供和公开等处理，均符合相关法律法规的规定，且不违背公序良俗。

[0147] 根据本公开的实施例，本公开还提供了一种电子设备、一种可读存储介质和一种计算机程序产品。

[0148] 图8示出了可以用来实施本公开的实施例的示例电子设备800的示意性框图。电子设备旨在表示各种形式的数字计算机，诸如，膝上型计算机、台式计算机、工作台、个人数字助理、服务器、刀片式服务器、大型计算机、和其它适合的计算机。电子设备还可以表示各种形式的移动装置，诸如，个人数字处理、蜂窝电话、智能电话、可穿戴设备和其它类似的计算装置。本文所示的部件、它们的连接和关系、以及它们的功能仅仅作为示例，并且不意在限制本文中描述的和/或者要求的本公开的实现。

[0149] 如图8所示，设备800包括计算单元801，其可以根据存储在只读存储器(ROM)802中的计算机程序或者从存储单元808加载到随机访问存储器(RAM)803中的计算机程序，来执行各种适当的动作和处理。在RAM 803中，还可存储设备800操作所需的各种程序和数据。计算单元801、ROM 802以及RAM 803通过总线804彼此相连。输入/输出(I/O)接口805也连接至总线804。

[0150] 设备800中的多个部件连接至I/O接口805，包括：输入单元806，例如键盘、鼠标等；输出单元807，例如各种类型的显示器、扬声器等；存储单元808，例如磁盘、光盘等；以及通信单元809，例如网卡、调制解调器、无线通信收发机等。通信单元809允许设备800通过诸如因特网的计算机网络和/或各种电信网络与其他设备交换信息/数据。

[0151] 计算单元801可以是各种具有处理和计算能力的通用和/或专用处理组件。计算单元801的一些示例包括但不限于中央处理单元(CPU)、图形处理单元(GPU)、各种专用的人工智能(AI)计算芯片、各种运行机器学习模型算法的计算单元、数字信号处理器(DSP)、以及任何适当的处理器、控制器、微控制器等。计算单元801执行上文所描述的各个方法和处理，例如图像处理方法。例如，在一些实施例中，图像处理方法可被实现为计算机软件程序，其被有形地包含于机器可读介质，例如存储单元808。在一些实施例中，计算机程序的部分或者全部可以经由ROM 802和/或通信单元809而被载入和/或安装到设备800上。当计算机程序加载到RAM 803并由计算单元801执行时，可以执行上文描述的图像处理方法的一个或多

个步骤。备选地，在其他实施例中，计算单元801可以通过其他任何适当的方式(例如，借助于固件)而被配置为执行图像处理方法。

[0152] 本文中以上描述的系统和技术的各种实施方式可以在数字电子电路系统、集成电路系统、场可编程门阵列(FPGA)、专用集成电路(ASIC)、专用标准产品(ASSP)、芯片上系统的系统(SOC)、复杂可编程逻辑设备(CPLD)、计算机硬件、固件、软件、和/或它们的组合中实现。这些各种实施方式可以包括：实施在一个或者多个计算机程序中，该一个或者多个计算机程序可在包括至少一个可编程处理器的可编程系统上执行和/或解释，该可编程处理器可以是专用或者通用可编程处理器，可以从存储系统、至少一个输入装置、和至少一个输出装置接收数据和指令，并且将数据和指令传输至该存储系统、该至少一个输入装置、和该至少一个输出装置。

[0153] 用于实施本公开的方法的程序代码可以采用一个或多个编程语言的任何组合来编写。这些程序代码可以提供给通用计算机、专用计算机或其他可编程数据处理装置的处理器或控制器，使得程序代码当由处理器或控制器执行时使流程图和/或框图中所规定的功能/操作被实施。程序代码可以完全在机器上执行、部分地在机器上执行，作为独立软件包部分地在机器上执行且部分地在远程机器上执行或完全在远程机器或服务器上执行。

[0154] 在本公开的上下文中，机器可读介质可以是有形的介质，其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备，或者上述内容的任何合适组合。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦除可编程只读存储器(EPROM或快闪存储器)、光纤、便捷式紧凑盘只读存储器(CD-ROM)、光学储存设备、磁储存设备、或上述内容的任何合适组合。

[0155] 为了提供与用户的交互，可以在计算机上实施此处描述的系统和技术，该计算机具有：用于向用户显示信息的显示装置(例如，CRT(阴极射线管)显示器或者LCD(液晶显示器))；以及键盘和指向装置(例如，鼠标或者轨迹球)，用户可以通过该键盘和该指向装置来将输入提供给计算机。其它种类的装置还可以用于提供与用户的交互；例如，提供给用户的反馈可以是任何形式的传感反馈(例如，视觉反馈、听觉反馈、或者触觉反馈)；并且可以用任何形式(包括声输入、语音输入或者、触觉输入)来接收来自用户的输入。

[0156] 可以将此处描述的系统和技术实施在包括后台部件的计算系统(例如，作为数据服务器)、或者包括中间件部件的计算系统(例如，应用服务器)、或者包括前端部件的计算系统(例如，具有图形用户界面或者网络浏览器的用户计算机，用户可以通过该图形用户界面或者该网络浏览器来与此处描述的系统和技术的实施方式交互)、或者包括这种后台部件、中间件部件、或者前端部件的任何组合的计算系统中。可以通过任何形式或者介质的数字数据通信(例如，通信网络)来将系统的部件相互连接。通信网络的示例包括：局域网(LAN)、广域网(WAN)和互联网。

[0157] 计算机系统可以包括客户端和服务器。客户端和服务器一般远离彼此并且通常通过通信网络进行交互。通过在相应的计算机上运行并且彼此具有客户端-服务器关系的计算机程序来产生客户端和服务器的关系。

[0158] 应该理解，可以使用上面所示的各种形式的流程，重新排序、增加或删除步骤。例如，本公开中记载的各步骤可以并行地执行也可以顺序地执行也可以不同的次序执行，只要能够实现本公开公开的技术方案所期望的结果，本文在此不进行限制。

[0159] 上述具体实施方式，并不构成对本公开保护范围的限制。本领域技术人员应该明白的是，根据设计要求和其他因素，可以进行各种修改、组合、子组合和替代。任何在本公开的精神和原则之内所作的修改、等同替换和改进等，均应包含在本公开保护范围之内。

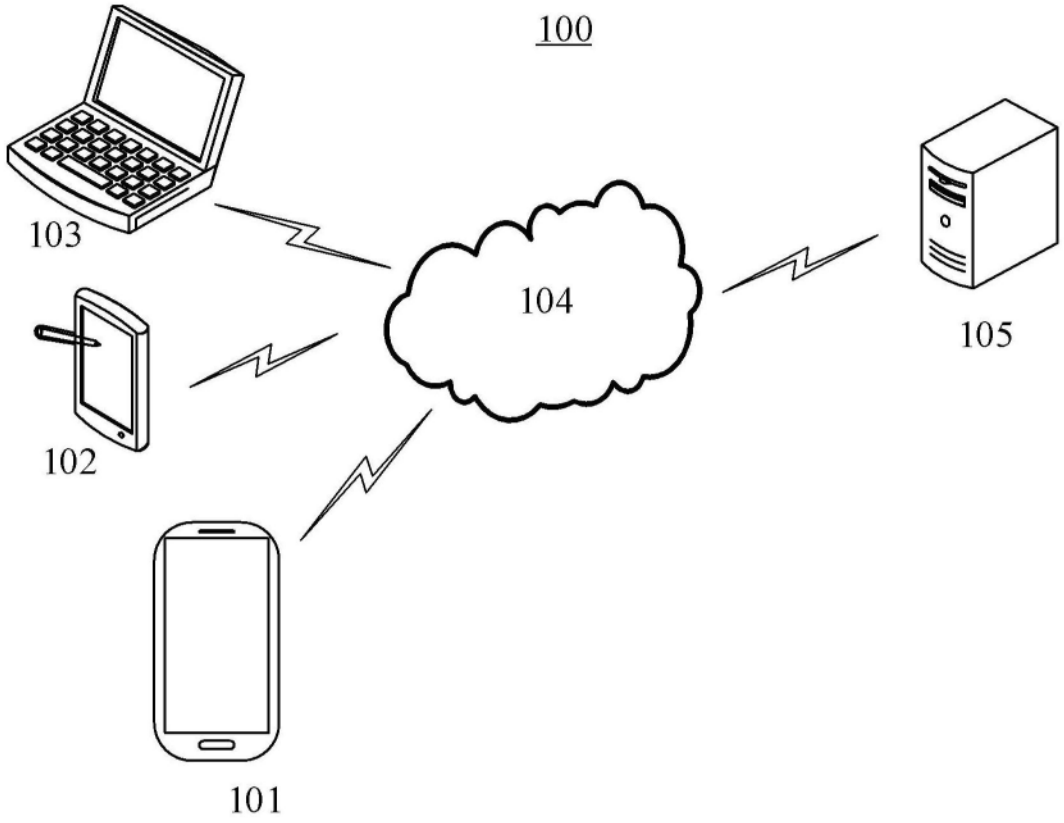


图1

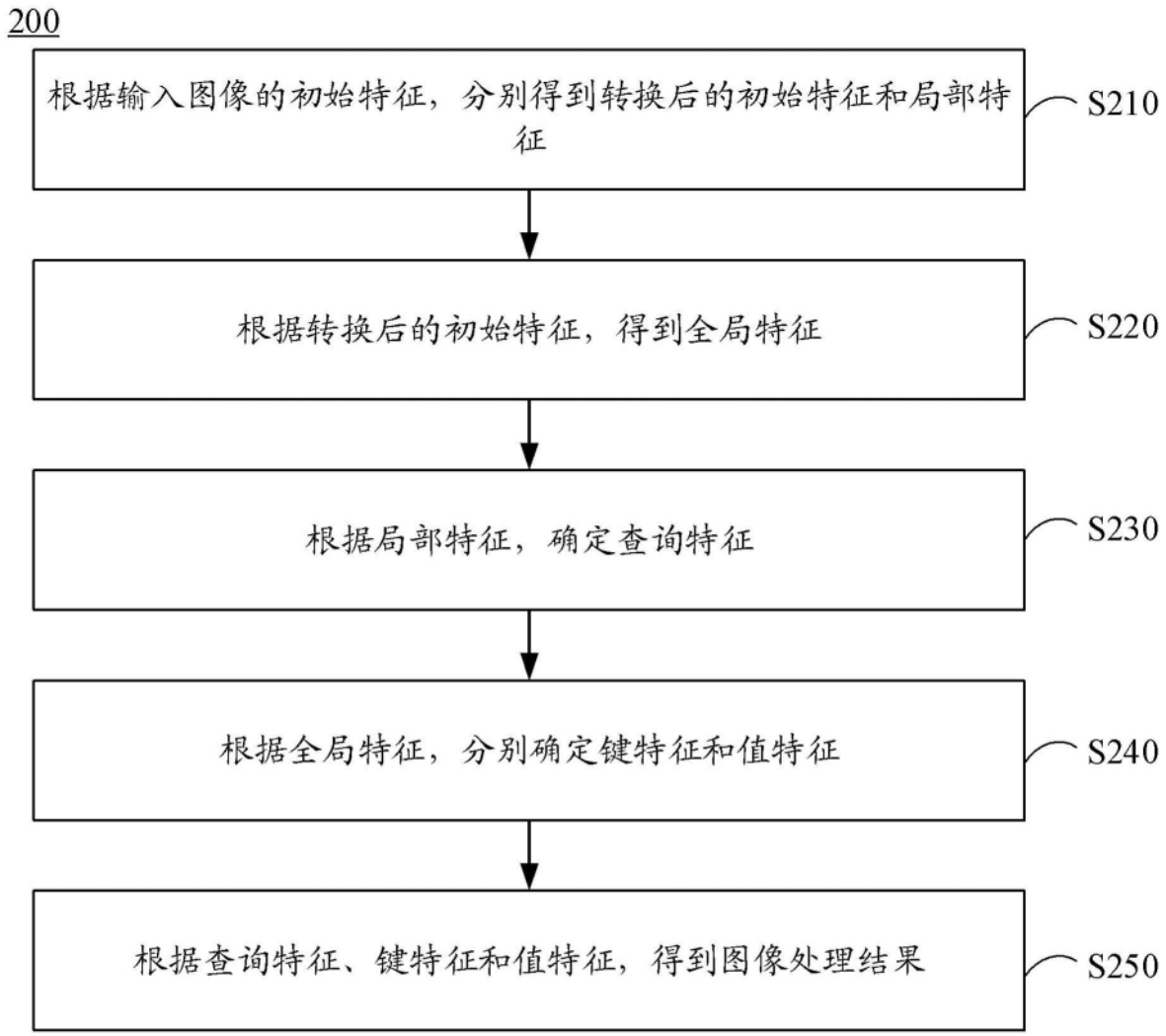

200
根据输入图像的初始特征，分别得到转换后的初始特征和局部特征
S210
根据转换后的初始特征，得到全局特征
S220
根据局部特征，确定查询特征
S230
根据全局特征，分别确定键特征和值特征
S240
根据查询特征、键特征和值特征，得到图像处理结果
S250

图2

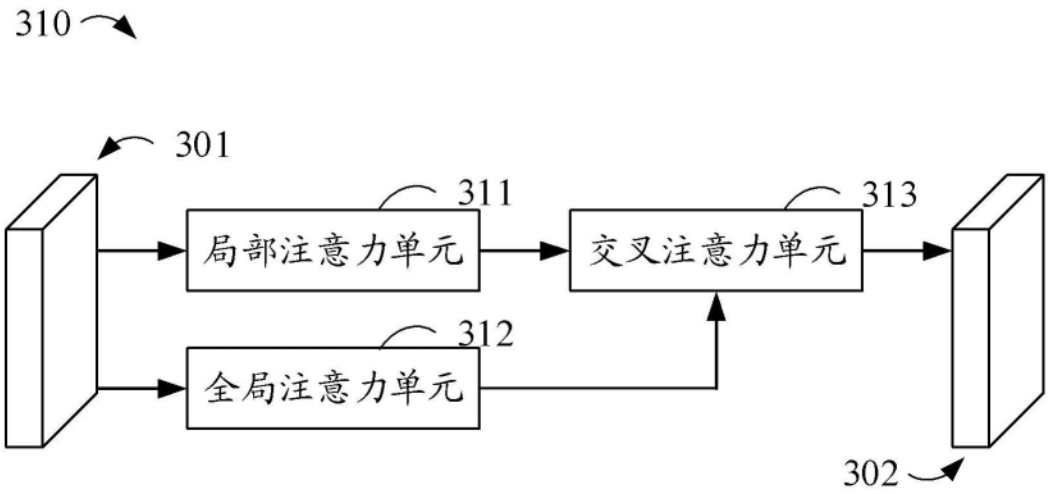

310
301
311
局部注意力单元
313
交叉注意力单元
312
全局注意力单元
302

图3A

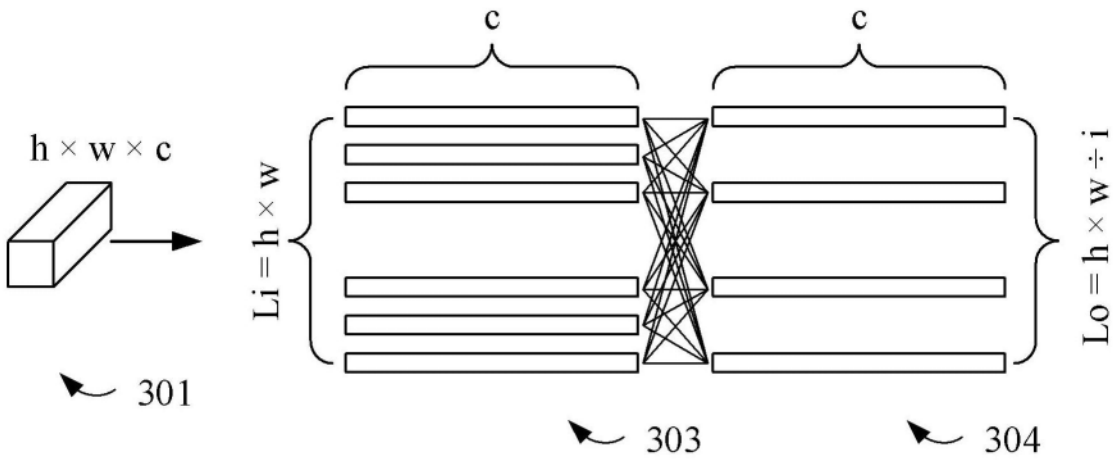

c
c
h × w × c
Li = h × w
Lo = h × w ÷ i
301
303
304

图3B

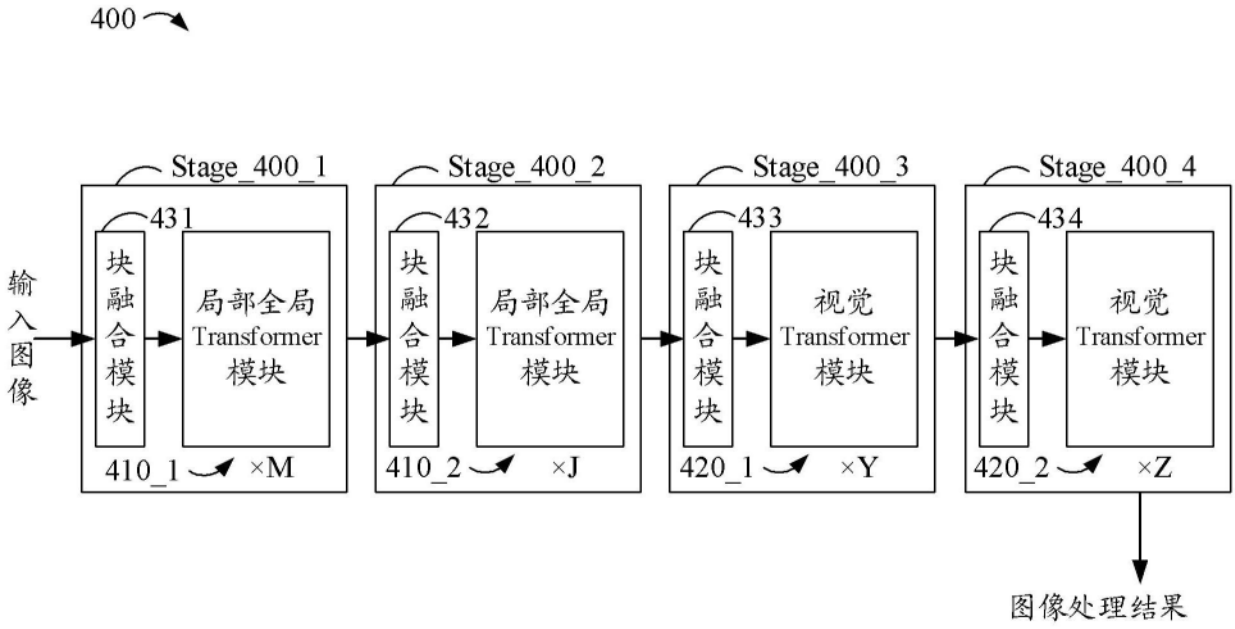

400
Stage_400_1
Stage_400_2
Stage_400_3
Stage_400_4
431
432
433
434
输入图像
块融合模块
局部全局Transformer模块
视觉Transformer模块
410_1 ×M
410_2 ×J
420_1 ×Y
420_2 ×Z
图像处理结果

图4

Models	Train Size	Test Size	Params(M)	FLOPs(G)	Top1(%)
DeiT-S [31]	224^2	224^2	22	4.6	79.8
Swin-T [23]	224^2	224^2	29	4.5	81.3
CrossViT-15 [3]	224^2	224^2	27	5.8	81.5
CoAtNet-0 [8]	224^2	224^2	25	4.6	81.6
Focal-T [41]	224^2	224^2	29	4.9	82.2
DsNet-S [25]	224^2	224^2	23	3.5	82.3
Shuffle-T [17]	224^2	224^2	29	4.6	82.5
Cswin-T [10]	224^2	224^2	23	4.3	82.7
LV-ViT-S [19]	224^2	224^2	26	6.6	83.3
LoGo-S	224^2	224^2	23	4.5	**83.7**
CrossViT-18 [3]	224^2	224^2	44	9.5	82.8
Swin-S [23]	224^2	224^2	50	8.7	83.0
DsNet-B [25]	224^2	224^2	49	8.4	83.1
Twins-SVT-B [7]	224^2	224^2	56	8.3	83.2
CoAtNet-1 [8]	224^2	224^2	42	8.4	83.3
Shuffle-S [17]	224^2	224^2	50	8.9	83.5
Focal-S [41]	224^2	224^2	51	9.1	83.5
Cswin-S [10]	224^2	224^2	35	8.9	83.6
LV-ViT-M [19]	224^2	224^2	56	16	84.1
LoGo-B	224^2	224^2	53	9.8	**85.0**
DeiT-B [31]	224^2	224^2	86	17.5	81.8
CrossViT-B [3]	224^2	224^2	105	21.2	82.2
Swin-B [23]	224^2	224^2	88	15.4	83.5
Focal-B [41]	224^2	224^2	90	16.0	83.8
Shuffle-B [17]	224^2	224^2	88	15.6	84
CSwin-B [10]	224^2	224^2	78	15.0	84.2
CoAtNet-3 [8]	224^2	224^2	168	34.7	84.5
CaiT-M36 [32]	224^2	384^2	271	247.8	85.1
LV-ViT-L [19]	288^2	288^2	150	59.0	85.3
LoGo-L	224^2	224^2	105	22.6	**85.9**

图5

Backbone	Params (M)	FLOPs (G)	Mask R-CNN 1x schedule					
			AP^b	AP^b_{50}	AP^b_{75}	AP^m	AP^m_{50}	AP^m_{75}
Res50 [14]	44	260	38.0	58.6	41.4	34.4	55.1	36.7
PVT-S [36]	44	245	40.4	62.9	43.8	37.8	60.1	40.3
ViL-S [47]	45	218	44.9	67.1	49.3	41.	64.2	44.1
TwinsP-S [7]	44	245	42.9	65.8	47.1	40.4	62.7	42.9
Twins-S [7]	44	228	43.4	66.0	47.3	40.3	63.2	
Swin-T [23]	48	264	42.2	64.6	46.2	39.1	64.6	42.0
CSwin-T [10]	42	279	46.7	68.6	51.3	42.2	65.6	45.4
LoGo-S	41	256	**47.0**	**69.5**	**51.6**	**42.7**	**66.5**	**46.1**
Res101 [14]	63	336	40.4	61.1	44.2	36.4	57.7	38.8
X101-32 [39]	63	340	41.9	62.5	45.9	37.5	59.4	40.2
PVT-M [36]	64	302	42.0	64.4	45.6	39.0	61.6	42.1
ViL-M [47]	60	261	43.4	–	–	39.7	–	–
TwinsP-B [7]	64	302	44.6	66.7	48.9	40.9	63.8	44.2
Twins-B [7]	76	340	45.2	67.6	49.3	41.5	64.5	44.8
Swin-S [23]	69	354	44.8	66.6	48.9	40.9	63.4	44.2
CSwin-S [10]	54	342	47.9	70.1	52.6	43.2	67.1	46.2
LoGo-B	71	354	**49.6**	**71.4**	**54.7**	**44.6**	**68.6**	**48.4**
X101-64 [39]	101	493	42.8	63.8	47.3	38.4	60.6	41.3
PVT-L [36]	81	364	42.9	65.0	46.6	39.5	61.9	42.5
ViL-B [47]	76	365	45.1	–	–	41.0	–	–
TwinsP-L [7]	81	364	45.4	–	–	41.5	–	–
Twins-L [7]	111	474	45.9	–	–	41.6	–	–
Swin-B [23]	107	496	46.9	–	–	42.3	–	–
CSwin-B [10]	97	526	48.7	70.4	53.9	43.9	67.8	47.3
LoGo-L	122	609	**50.7**	**72.4**	**55.6**	**45.4**	**69.7**	**49.2**

图6

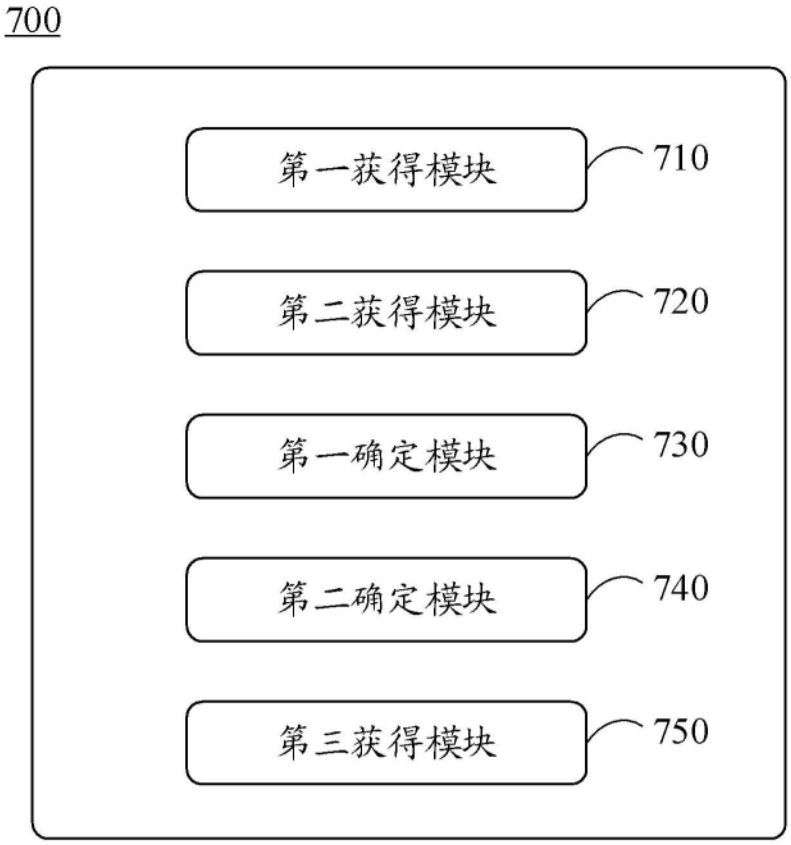

700
第一获得模块
710
第二获得模块
720
第一确定模块
730
第二确定模块
740
第三获得模块
750

图7

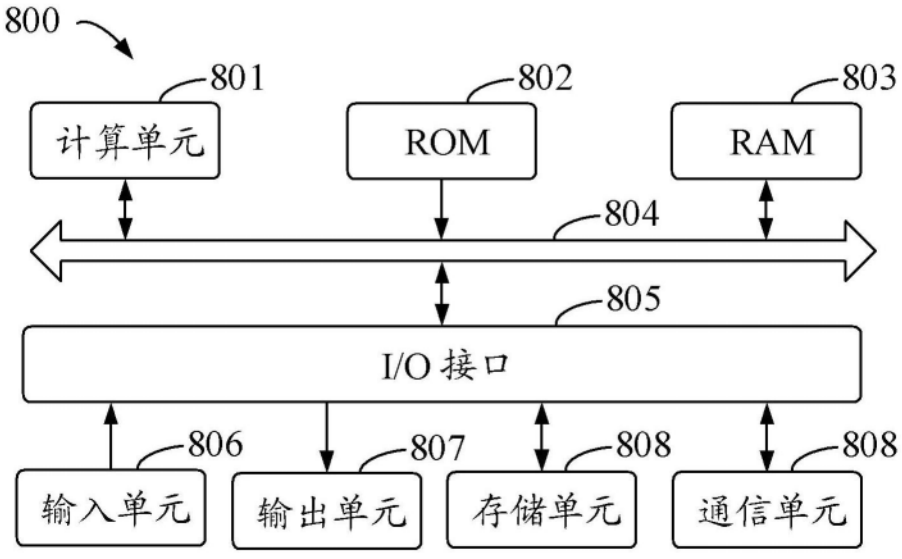

800
801
计算单元
802
ROM
803
RAM
804
805
I/O 接口
806
输入单元
807
输出单元
808
存储单元
808
通信单元

图8