



(12) 发明专利申请

(10) 申请公布号 CN 113298091 A

(43) 申请公布日 2021.08.24

(21) 申请号 202110573055.1

(22) 申请日 2021.05.25

(71) 申请人 商汤集团有限公司

地址 中国香港新界沙田香港科学园科技大道西一号核心大楼第二座2楼226-230室

(72) 发明人 陈博宇 李楚鸣

(74) 专利代理机构 北京林达刘知识产权代理事务所(普通合伙) 11277

代理人 刘新宇

(51) Int. Cl.

G06K 9/46 (2006.01)

G06N 3/04 (2006.01)

G06N 3/08 (2006.01)

权利要求书3页 说明书17页 附图5页

(54) 发明名称

图像处理方法及装置、电子设备和存储介质

(57) 摘要

本公开涉及一种图像处理方法及装置、电子设备和存储介质,所述方法包括:确定目标图像对应的第一图像块特征,以及将所述第一图像块特征划分为第二图像块特征和第三图像块特征;基于全局注意力机制对所述第二图像块特征进行特征强化,得到第四图像块特征,以及基于局部注意力机制对所述第三图像块特征进行特征强化,得到第五图像块特征;根据所述第四图像块特征和所述第五图像块特征,确定所述目标图像对应的目标图像块特征;根据所述目标图像块特征,对所述目标图像进行目标图像处理操作,得到图像处理结果。本公开实施例可以提高图像处理结果的精度。



1. 一种图像处理方法,其特征在于,包括:

确定目标图像对应的第一图像块特征,以及将所述第一图像块特征划分为第二图像块特征和第三图像块特征;

基于全局注意力机制对所述第二图像块特征进行特征强化,得到第四图像块特征,以及基于局部注意力机制对所述第三图像块特征进行特征强化,得到第五图像块特征;

根据所述第四图像块特征和所述第五图像块特征,确定所述目标图像对应的目标图像块特征;

根据所述目标图像块特征,对所述目标图像进行目标图像处理操作,得到图像处理结果。

2. 根据权利要求1所述的方法,其特征在于,所述第一图像块特征对应的通道数是目标通道数;

所述将所述第一图像块特征划分为第二图像块特征和第三图像块特征,包括:

根据所述目标通道数和目标比例,将所述第一图像块特征在通道维度上进行划分,得到所述第二图像块特征和所述第三图像块特征。

3. 根据权利要求1或2所述的方法,其特征在于,所述基于全局注意力机制对所述第二图像块特征进行特征强化,得到第四图像块特征,包括:

根据所述第二图像块特征,确定第一特征向量、第二特征向量和第三特征向量;

根据所述第一特征向量和所述第二特征向量,确定注意力特征图;

根据所述注意力特征图和所述第三特征向量,确定所述第四图像块特征。

4. 根据权利要求1至3中任意一项所述的方法,其特征在于,所述基于局部注意力机制对所述第三图像块特征进行特征强化,得到第五图像块特征,包括:

根据目标卷积核和第一通道扩展率,对所述第三图像块特征进行卷积处理,得到所述第五图像块特征。

5. 根据权利要求1至4中任意一项所述的方法,其特征在于,所述根据所述第四图像块特征和所述第五图像块特征,确定所述目标图像对应的目标图像块特征,包括:

根据第二通道扩展率,对所述第四图像块特征和所述第五图像块特征进行特征转换,得到所述目标图像块特征。

6. 根据权利要求1至5中任意一项所述的方法,其特征在于,所述图像处理方法通过自注意力神经网络实现,其中,所述自注意力神经网络中包括至少一个注意力模块,至少一个所述注意力模块中包括全局注意力子模块、局部注意力子模块和前馈子模块;

所述基于全局注意力机制对所述第二图像块特征进行特征强化,得到第四图像块特征,包括:

利用所述全局注意力子模块,基于全局注意力机制对所述第二图像块特征进行特征强化,得到所述第四图像块特征;

所述基于局部注意力机制对所述第三图像块特征进行特征强化,得到第五图像块特征,包括:

利用所述局部注意力子模块,基于局部注意力机制对所述第三图像块特征进行特征强化,得到所述第五图像块特征;

所述根据所述第四图像块特征和所述第五图像块特征,确定所述目标图像对应的目标

图像块特征,包括:

利用所述前馈子模块,对所述第四图像块特征和所述第五图像块特征进行特征转换,得到所述目标图像块特征。

7. 根据权利要求6所述的方法,其特征在于,所述方法还包括:

构建第一网络结构搜索空间和第二网络结构搜索空间,其中,所述第一网络结构搜索空间中包括多个模块分布超参数,所述第二网络结构搜索空间中包括多个模块结构超参数;

根据所述第一网络结构搜索空间,从所述多个模块分布超参数中确定目标模块分布超参数,其中,所述目标模块超参数用于指示各注意力模块中包括的所述全局注意力子模块的个数,以及所述局部注意力子模块的个数;

根据所述目标模块分布超参数和所述第二网络结构搜索空间,从所述多个模块结构超参数中确定目标模块结构超参数,其中,所述目标模块结构超参数用于指示各注意力模块对应的目标通道数、各局部注意力子模块对应的目标卷积核和第一通道扩展率、以及各前馈模块对应的第二通道扩展率;

根据所述目标模块分布超参数和所述目标模块结构超参数,构建所述自注意力神经网络。

8. 根据权利要求7所述的方法,其特征在于,所述根据所述第一网络结构搜索空间,从所述多个模块分布超参数中确定目标模块分布超参数,包括:

根据所述第一网络结构搜索空间,构建第一超级网络,其中,所述第一超级网络中包括根据所述多个模块分布超参数构建的多个第一可选网络结构;

通过对所述第一超级网络进行网络训练,从所述多个第一可选网络结构中确定第一目标网络结构;

将所述第一目标网络结构对应的模块分布超参数确定为所述目标模块分布超参数。

9. 根据权利要求7或8所述的方法,其特征在于,所述根据所述目标模块分布超参数和所述第二网络结构搜索空间,从所述多个模块结构超参数中确定目标模块结构超参数,包括:

根据所述目标模块分布超参数和所述第二网络结构搜索空间,构建第二超级网络,其中,所述第二超级网络中包括根据所述多个模块结构超参数构建的多个第二可选网络结构;

通过对所述第二超级网络进行网络训练,从所述多个第二可选网络结构中确定第二目标网络结构;

将所述第二目标网络结构对应的模块结构超参数确定为所述目标模块结构超参数。

10. 根据权利要求7至9中任意一项所述的方法,其特征在于,相同所述注意力模块中,各所述局部注意力子模块对应相同的所述目标卷积核和所述第一通道扩展率。

11. 一种图像处理装置,其特征在于,包括:

特征确定模块,用于确定目标图像对应的第一图像块特征,以及将所述第一图像块特征划分为第二图像块特征和第三图像块特征;

注意力模块,用于基于全局注意力机制对所述第二图像块特征进行特征强化,得到第四图像块特征,以及基于局部注意力机制对所述第三图像块特征进行特征强化,得到第五

图像块特征；

第一确定模块,用于根据所述第四图像块特征和所述第五图像块特征,确定所述目标图像对应的目标图像块特征；

目标图像处理模块,用于根据所述目标图像块特征,对所述目标图像进行目标图像处理操作,得到图像处理结果。

12. 一种电子设备,其特征在于,包括:

处理器;

用于存储处理器可执行指令的存储器;

其中,所述处理器被配置为调用所述存储器存储的指令,以执行权利要求1至10中任意一项所述的方法。

13. 一种计算机可读存储介质,其上存储有计算机程序指令,其特征在于,所述计算机程序指令被处理器执行时实现权利要求1至10中任意一项所述的方法。

图像处理方法及装置、电子设备和存储介质

技术领域

[0001] 本公开涉及计算机技术领域,尤其涉及一种图像处理方法及装置、电子设备和存储介质。

背景技术

[0002] 近期,自注意力网络在自然语言处理中得到了广泛的应用,自注意力网络通过建立特征之间的联系来进行特征的强化,从而提升网络的最终性能。随着视觉自注意力网络的提出,自注意力网络在计算机视觉领域也得到了大规模应用,展现出极大的潜力。然而,现有的视觉自注意力网络的设计只是简单地照搬了自然语言处理中的设计,并没有针对计算机视觉的特征进行改进,使得视觉自注意力网络的性能较差。

发明内容

[0003] 本公开提出了一种图像处理方法及装置、电子设备和存储介质的技术方案。

[0004] 根据本公开的一方面,提供了一种图像处理方法,包括:确定目标图像对应的第一图像块特征,以及将所述第一图像块特征划分为第二图像块特征和第三图像块特征;基于全局注意力机制对所述第二图像块特征进行特征强化,得到第四图像块特征,以及基于局部注意力机制对所述第三图像块特征进行特征强化,得到第五图像块特征;根据所述第四图像块特征和所述第五图像块特征,确定所述目标图像对应的目标图像块特征;根据所述目标图像块特征,对所述目标图像进行目标图像处理操作,得到图像处理结果。

[0005] 在一种可能的实现方式中,所述第一图像块特征对应的通道数是目标通道数;所述将所述第一图像块特征划分为第二图像块特征和第三图像块特征,包括:根据所述目标通道数和目标比例,将所述第一图像块特征在通道维度上进行划分,得到所述第二图像块特征和所述第三图像块特征。

[0006] 在一种可能的实现方式中,所述基于全局注意力机制对所述第二图像块特征进行特征强化,得到第四图像块特征,包括:根据所述第二图像块特征,确定第一特征向量、第二特征向量和第三特征向量;根据所述第一特征向量和所述第二特征向量,确定注意力特征图;根据所述注意力特征图和所述第三特征向量,确定所述第四图像块特征。

[0007] 在一种可能的实现方式中,所述基于局部注意力机制对所述第三图像块特征进行特征强化,得到第五图像块特征,包括:根据目标卷积核和第一通道扩展率,对所述第三图像块特征进行卷积处理,得到所述第五图像块特征。

[0008] 在一种可能的实现方式中,所述根据所述第四图像块特征和所述第五图像块特征,确定所述目标图像对应的目标图像块特征,包括:根据第二通道扩展率,对所述第四图像块特征和所述第五图像块特征进行特征转换,得到所述目标图像块特征。

[0009] 在一种可能的实现方式中,所述图像处理方法通过自注意力神经网络实现,其中,所述自注意力神经网络中包括至少一个注意力模块,至少一个所述注意力模块中包括全局注意力子模块、局部注意力子模块和前馈子模块;所述基于全局注意力机制对所述第二图

像块特征进行特征强化,得到第四图像块特征,包括:利用所述全局注意力子模块,基于全局注意力机制对所述第二图像块特征进行特征强化,得到所述第四图像块特征;所述基于局部注意力机制对所述第三图像块特征进行特征强化,得到第五图像块特征,包括:利用所述局部注意力子模块,基于局部注意力机制对所述第三图像块特征进行特征强化,得到所述第五图像块特征;所述根据所述第四图像块特征和所述第五图像块特征,确定所述目标图像对应的目标图像块特征,包括:利用所述前馈子模块,对所述第四图像块特征和所述第五图像块特征进行特征转换,得到所述目标图像块特征。

[0010] 在一种可能的实现方式中,所述方法还包括:构建第一网络结构搜索空间和第二网络结构搜索空间,其中,所述第一网络结构搜索空间中包括多个模块分布超参数,所述第二网络结构搜索空间中包括多个模块结构超参数;根据所述第一网络结构搜索空间,从所述多个模块分布超参数中确定目标模块分布超参数,其中,所述目标模块超参数用于指示各注意力模块中包括的所述全局注意力子模块的个数,以及所述局部注意力子模块的个数;根据所述目标模块分布超参数和所述第二网络结构搜索空间,从所述多个模块结构超参数中确定目标模块结构超参数,其中,所述目标模块结构超参数用于指示各注意力模块对应的目标通道数、各局部注意力子模块对应的目标卷积核和第一通道扩展率、以及各前馈模块对应的第二通道扩展率;根据所述目标模块分布超参数和所述目标模块结构超参数,构建所述自注意力神经网络。

[0011] 在一种可能的实现方式中,所述根据所述第一网络结构搜索空间,从所述多个模块分布超参数中确定目标模块分布超参数,包括:根据所述第一网络结构搜索空间,构建第一超级网络,其中,所述第一超级网络中包括根据所述多个模块分布超参数构建的多个第一可选网络结构;通过对所述第一超级网络进行网络训练,从所述多个第一可选网络结构中确定第一目标网络结构;将所述第一目标网络结构对应的模块分布超参数确定为所述目标模块分布超参数。

[0012] 在一种可能的实现方式中,所述根据所述目标模块分布超参数和所述第二网络结构搜索空间,从所述多个模块结构超参数中确定目标模块结构超参数,包括:根据所述目标模块分布超参数和所述第二网络结构搜索空间,构建第二超级网络,其中,所述第二超级网络中包括根据所述多个模块结构超参数构建的多个第二可选网络结构;通过对所述第二超级网络进行网络训练,从所述多个第二可选网络结构中确定第二目标网络结构;将所述第二目标网络结构对应的模块结构超参数确定为所述目标模块结构超参数。

[0013] 在一种可能的实现方式中,相同所述注意力模块中,各所述局部注意力子模块对应相同的所述目标卷积核和所述第一通道扩展率。

[0014] 根据本公开的一方面,提供了一种图像处理装置,包括:特征确定模块,用于确定目标图像对应的第一图像块特征,以及将所述第一图像块特征划分为第二图像块特征和第三图像块特征;注意力模块,用于基于全局注意力机制对所述第二图像块特征进行特征强化,得到第四图像块特征,以及基于局部注意力机制对所述第三图像块特征进行特征强化,得到第五图像块特征;确定模块,用于根据所述第四图像块特征和所述第五图像块特征,确定所述目标图像对应的目标图像块特征;目标图像处理模块,用于根据所述目标图像块特征,对所述目标图像进行目标图像处理操作,得到图像处理结果。

[0015] 根据本公开的一方面,提供了一种电子设备,包括:处理器;用于存储处理器可执

行指令的存储器；其中，所述处理器被配置为调用所述存储器存储的指令，以执行上述方法。

[0016] 根据本公开的一方面，提供了一种计算机可读存储介质，其上存储有计算机程序指令，所述计算机程序指令被处理器执行时实现上述方法。

[0017] 在本公开实施例中，确定目标图像对应的第一图像块特征，以及将第一图像块特征划分为第二图像块特征和第三图像块特征；基于全局注意力机制对第二图像块特征进行特征强化，得到第四图像块特征，以及基于局部注意力机制对第三图像块特征进行特征强化，得到第五图像块特征；根据第四图像块特征和第五图像块特征，确定目标图像对应的目标图像块特征；根据目标图像块特征，对目标图像进行目标图像处理操作，得到图像处理结果。分别基于全局注意力机制和局部注意力机制对图像块特征进行特征强化，以使得特征强化之后的目标图像块特征既包括全局信息又包括局部信息，有效提高了目标图像块特征的语义表达能力，进而在利用具有较高语义表达能力的目标图像块特征进行目标图像处理操作之后，可以提高图像处理结果的精度。

[0018] 应当理解的是，以上的一般描述和后文的细节描述仅是示例性和解释性的，而非限制本公开。根据下面参考附图对示例性实施例的详细说明，本公开的其它特征及方面将变得清楚。

附图说明

[0019] 此处的附图被并入说明书中并构成本说明书的一部分，这些附图示出了符合本公开的实施例，并与说明书一起用于说明本公开的技术方案。

[0020] 图1示出根据本公开实施例的一种图像处理方法的流程图；

[0021] 图2示出根据本公开实施例的确定目标图像对应的多个第一图像块特征的示意图；

[0022] 图3示出相关技术中的一种自注意力神经网络的网络结构图；

[0023] 图4示出根据本公开实施例的一种自注意力神经网络的网络结构图；

[0024] 图5示出根据本公开实施例的分层级的网络结构搜索的示意图；

[0025] 图6示出根据本公开实施例的一种图像处理装置的框图；

[0026] 图7示出根据本公开实施例的一种电子设备的框图；

[0027] 图8示出根据本公开实施例的一种电子设备的框图。

具体实施方式

[0028] 以下将参考附图详细说明本公开的各种示例性实施例、特征和方面。附图中相同的附图标记表示功能相同或相似的元件。尽管在附图中示出了实施例的各种方面，但是除非特别指出，不必按比例绘制附图。

[0029] 在这里专用的词“示例性”意为“用作例子、实施例或说明性”。这里作为“示例性”所说明的任何实施例不必解释为优于或好于其它实施例。

[0030] 本文中术语“和/或”，仅仅是一种描述关联对象的关联关系，表示可以存在三种关系，例如，A和/或B，可以表示：单独存在A，同时存在A和B，单独存在B这三种情况。另外，本文中术语“至少一种”表示多种中的任意一种或多种中的至少两种的任意组合，例如，包括A、

B、C中的至少一种,可以表示包括从A、B和C构成的集合中选择的任意一个或多个元素。

[0031] 另外,为了更好地说明本公开,在下文的具体实施方式中给出了众多的具体细节。本领域技术人员应当理解,没有某些具体细节,本公开同样可以实施。在一些实例中,对于本领域技术人员熟知的方法、手段、元件和电路未作详细描述,以便于凸显本公开的主旨。

[0032] 图1示出根据本公开实施例的一种图像处理方法的流程图。该图像处理方法可以由终端设备或服务器等电子设备执行,终端设备可以为用户设备(User Equipment, UE)、移动设备、用户终端、蜂窝电话、无绳电话、个人数字助理(Personal Digital Assistant, PDA)、手持设备、计算设备、车载设备、可穿戴设备等,所述方法可以通过处理器调用存储器中存储的计算机可读指令的方式来实现。或者,可通过服务器执行所述方法。如图1所示,该图像处理方法可以包括:

[0033] 在步骤S11中,确定目标图像对应的第一图像块特征,以及将第一图像块特征划分为第二图像块特征和第三图像块特征。

[0034] 这里的目标图像是需要进行图像处理的待处理图像。为了更好地获取目标图像的内部相关性,可以对目标图像进行分割,将目标图像划分为多个图像块,通过对各图像块进行特征提取,可以得到目标图像对应的多个第一图像块特征。多个图像块的个数,可以根据实际情况确定,本公开对此不做具体限定。

[0035] 图2示出根据本公开实施例的确定目标图像对应的第一图像块特征的示意图。如图2所示,对目标图像进行分割,将目标图像划分为L个图像块,对L个图像块分别进行特征提取,得到L个第一图像块特征。

[0036] 为了便于对多个第一图像块特征进行后续处理,可以将多个第一图像块特征转换为第一图像块特征序列。将多个第一图像块特征转换为第一图像块特征序列的具体方式可以根据实际情况确定,本公开对此不做具体限定。

[0037] 将第一图像块特征划分为第二图像块特征和第三图像块特征两个部分,以使得后续可以分别基于全局注意力机制和局部注意力机制,对第二图像块特征和第三图像块特征进行特征强化。后文会结合本公开可能的实现方式,对将第一图像块特征划分为第二图像块特征和第三图像块特征的过程进行详细描述,此处不作赘述。

[0038] 在步骤S12中,基于全局注意力机制对第二图像块特征进行特征强化,得到第四图像块特征,以及基于局部注意力机制对第三图像块特征进行特征强化,得到第五图像块特征。

[0039] 全局注意力机制可以是指确定全部特征之间的全局注意力关系,进而根据该全局注意力关系确定新特征的过程。因此,基于全局注意力机制,确定各第二图像块特征之间的全局注意力关系,进而基于该全局注意力关系确定第四图像块特征,以实现特征强化。其中,由于第四图像块特征是基于全局注意力关系确定得到的,因此,第四图像块特征包括目标图像的全局信息。后文会基于本公开可能的实现方式,对基于全局注意力进行特征强化的过程进行详细描述,此处不作赘述。

[0040] 局部注意力机制可以是指确定局部相邻特征之间的局部注意力关系,进而根据该局部注意力关系确定新特征的过程。因此,基于局部注意力机制,确定局部相邻第三图像块特征之间的局部注意力关系,进而基于该局部注意力关系确定第五图像块特征,以实现特征强化。其中,由于第五图像块特征是基于局部注意力关系确定得到的,因此,第五图像块

特征包括目标图像的局部信息。后文会基于本公开可能的实现方式,对基于局部注意力进行特征强化的过程进行详细描述,此处不作赘述。

[0041] 在步骤S13中,根据第四图像块特征和第五图像块特征,确定目标图像对应的目标图像块特征。

[0042] 由于第四图像块特征包括目标图像的全局信息,第五图像块特征包括目标图像的局部信息,因此,根据第四图像块特征和第五图像块特征确定的目标图像块特征,既包括目标图像的全局信息,又包括目标图像的局部信息,具有较高的语义表达能力。

[0043] 在步骤S14中,根据目标图像块特征,对目标图像进行目标图像处理操作,得到图像处理结果。

[0044] 由于目标图像块特征既包括全局信息又包括局部信息,具有较高的语义表达能力,因此,根据图像处理需求,利用目标图像块特征进行目标图像处理操作,可以得到精度较高的图像处理结果。

[0045] 在本公开实施例中,确定目标图像对应的多个第一图像块特征,以及将第一图像块特征划分为第二图像块特征和第三图像块特征;基于全局注意力机制对第二图像块特征进行特征强化,得到第四图像块特征,以及基于局部注意力机制对第三图像块特征进行特征强化,得到第五图像块特征;根据第四图像块特征和第五图像块特征,确定目标图像对应的目标图像块特征;根据目标图像块特征,对目标图像进行目标图像处理操作,得到图像处理结果。分别基于全局注意力机制和局部注意力机制对图像块特征进行特征强化,以使得特征强化之后的目标图像块特征既包括全局信息又包括局部信息,有效提高了目标图像块特征的语义表达能力,进而在利用具有较高语义表达能力的目标图像块特征进行目标图像处理操作之后,可以提高图像处理结果的精度。

[0046] 在一种可能的实现方式中,第一图像块特征对应的通道数是目标通道数;将第一图像块特征划分为第二图像块特征和第三图像块特征,包括:根据目标通道数和目标比例,将第一图像块特征在通道维度上进行划分,得到第二图像块特征和第三图像块特征。

[0047] 通过对第一图像块特征在通道维度上进行划分,可以得到用于进行全局注意力处理的第二图像块特征,以及用于进行局部注意力处理的第三图像块特征,第二图像块特征和第三图像块特征的个数符合目标比例,为后续特征强化过程做好准备工作。目标通道数和目标比例的具体取值可以根据实际情况确定,本公开对此不做具体限定。

[0048] 仍以上述L个第一图像块特征为例,目标通道数是 d_k ,目标比例是 $a:b$ 。因此,可以对L个第一图像块特征沿通道维度平均分成N个部分,将N个部分中的 $N \times \frac{a}{a+b}$ 个部分确定

为 $N \times \frac{a}{a+b}$ 个第二图像块特征组,以及将N个部分中的 $N \times \frac{b}{a+b}$ 个部分确定为 $N \times \frac{b}{a+b}$ 个

第三图像块特征组。其中,各第二图像块特征组中包括L个第二图像块特征,各第二图像块特征对应的通道数是 d_k/N ;各第三图像块特征组中包括L个第三图像块特征,各第三图像块特征对应的通道数是 d_k/N 。N的具体取值可以根据实际情况确定,本公开对此不做具体限定。

[0049] 在一种可能的实现方式中,基于全局注意力机制对第二图像块特征进行特征强化,得到第四图像块特征,包括:根据第二图像块特征,确定第一特征向量、第二特征向量和

第三特征向量;根据第一特征向量和第二特征向量,确定注意力特征图;根据注意力特征图和第三特征向量,确定多个第二图像块特征。

[0050] 通过确定注意力特征图,可以得到各第二图像块特征之间的全局注意力关系,进而根据该注意力特征图可以实现基于全局注意力机制的特征强化,有效得到包括全局信息的第四图像块特征。

[0051] 仍以上述 $N \times \frac{a}{a+b}$ 个第二图像块特征组为例,针对任一个第二图像块特征组,将该第二图像块特征组中包括的L个第二图像块特征,转换为三个不同的特征向量:第一特征向量Q、第二特征向量K和第三特征向量V,其中,第一特征向量Q、第二特征向量K和第三特征向量V对应的特征通道数均是 d_k/N ,进而利用下述公式(1),基于全局注意力机制进行特征强化:

$$[0052] \quad Att(Q, K, V) = Softmat \left(\frac{Q \cdot (K^T)}{\sqrt{d_k/N}} \right) \cdot V \quad (1)。$$

[0053] 其中,Softmax(\cdot)表示归一化函数, \cdot 表示向量点积。如上述公式(1)所示,利用特征向量Q和特征向量K进行点积操作,可以得到点积结果 $Q \cdot (K^T)$,利用特征通道数 d_k/N 以及归一化函数Softmax(\cdot),对点积结果 $Q \cdot (K^T)$ 进行归一化操作,可以得到用于指示全局

注意力关系的注意力特征图 $Softmat \left(\frac{Q \cdot (K^T)}{\sqrt{d_k/N}} \right)$,利用注意力特征图 $Softmat \left(\frac{Q \cdot (K^T)}{\sqrt{d_k/N}} \right)$ 与特

征向量V进行点积操作得到Att(Q,K,V),Att(Q,K,V)是基于全局注意力机制进行特征强化后得到的该第二图像块特征组对应的输出结果。

[0054] 基于上述方式分别得到 $N \times \frac{a}{a+b}$ 个第二图像块特征组对应的输出结果,进而将 $N \times \frac{a}{a+b}$ 个第二图像块特征组对应的输出结果在通道维度进行融合,得到基于全局注意力

机制进行特征强化之后的L个第四图像块特征。

[0055] 在一种可能的实现方式中,基于局部注意力机制对第三图像块特征进行特征强化,得到多第五图像块特征,包括:对第三图像块特征进行卷积处理,得到第五图像块特征。

[0056] 通过对第三图像块特征进行卷积处理,以实现基于局部注意力机制的特征强化,根据卷积结果可以有效得到包括局部信息的多个第五图像块特征。由于第三图像块特征是一维的,因为,可以对第三图像块特征进行一维卷积处理,得到第五图像块特征。

[0057] 仍以上述 $N \times \frac{b}{a+b}$ 个第三图像块特征组为例,针对任一个第三图像块特征组,对该第三图像块特征组中包括的L个第三图像块特征进行一维卷积处理,得到该第三图像块特征组对应的卷积结果。

[0058] 在一种可能的实现方式中,基于局部注意力机制对第三图像块特征进行特征强化,得到第五图像块特征,包括:根据目标卷积核和第一通道扩展率,对第三图像块特征进

行卷积处理,得到第五图像块特征。

[0059] 由于第三图像块特征是一维的,因为,可以根据目标卷积核和第一通道扩展率,对第三图像块特征进行一维卷积处理,得到第五图像块特征。

[0060] 仍以上述 $N \times \frac{b}{a+b}$ 个第三图像块特征组为例,针对任一个第三图像块特征组,根据目标卷积核和第一通道扩展率,对该第三图像块特征组中包括的L个第三图像块特征进行卷积处理,具体包括:根据第一通道扩展率E,对L个第三图像块特征进行一维逐点卷积(pointwise卷积)处理,得到第一卷积子结果,其中,第一卷积结果对应的通道数是 $d_k/N \times E$;根据目标卷积核K,对第一卷积结果进行一维逐层卷积(depthwise卷积)处理,得到第二卷积子结果,其中,第二卷积子结果对应的通道数是 $d_k/N \times E$;根据第一通道扩展率E,对第二卷积子结果进行一维pointwise卷积处理,得到该第三图像块特征组对应的卷积结果,其中,该第三图像块特征组对应的卷积结果的通道数是 d_k/N 。第一通道扩展率E和目标卷积核K的具体取值可以根据实际情况确定,本公开对此不做具体限定。

[0061] 基于上述方式分别得到 $N \times \frac{b}{a+b}$ 个第三图像块特征组对应的卷积结果,进而将 $N \times \frac{b}{a+b}$ 个第三图像块特征组对应的卷积结果在通道维度进行融合,得到基于局部注意力机制进行特征强化之后的L个第五图像块特征。

[0062] 在一种可能的实现方式中,基于局部注意力机制对第三图像块进行特征强化,除了可以采用上述一维卷积处理的方式之外,还可以通过缩小处理数据大小的方式来实现。

仍以上述 $N \times \frac{b}{a+b}$ 个第三图像块特征组为例,针对任一个第三图像块特征组,仅对该第三图像块特征组中包括的L个第三图像块特征中的部分第三图像块特征(例如,L/2个第三图像块特征),进行上述公式(1)所示的特征强化,从而通过缩小处理数据大小的方式,实现基于局部注意力机制的特征强化。基于局部注意力机制对第三图像块进行特征强化还可以根据实际需要采用其它处理形式,本公开对此不做具体限定。

[0063] 在一种可能的实现方式中,根据第四图像块特征和第五图像块特征,确定目标图像对应的目标图像块特征,包括:根据第二通道扩展率,对第四图像块特征和第五图像块特征进行特征转换,得到目标图像块特征。

[0064] 通过第二通道扩展率,对第四图像块特征和第五图像块特征进行进一步地特征转换,得到既包括全局信息,又包括局部信息的目标图像块特征,有效提高了目标图像块特征的语义表达能力。

[0065] 仍以上述L个第四图像块特征和L个第五图像块特征,将L个第四图像块特征和L个第五图像块特征在通道维度进行融合,得到融合结果,进而利用第二通道扩展率 d_z ,对融合结果进行特征转换,以使得将融合结果的通道数转换为,与特征强化之前的第一图像块特征对应的目标通道数相同,从而得到通道数是 d_k 的L个目标图像块特征。第二通道扩展率 d_z 的具体取值可以根据实际情况确定,本公开对此不做具体限定。

[0066] 在一种可能的实现方式中,目标图像处理操作包括下述之一:目标检测、目标跟踪、图像识别、图像分类。

[0067] 利用具有更高的语义表达能力的目标图像块特征进行目标图像处理操作,从而可以得到精度较高的图像处理结果。

[0068] 例如,根据目标图像块特征,对目标图像进行图像分类,得到目标图像对应的具有较高分类精度的图像分类结果。目标图像处理操作除了可以包括上述目标检测、目标跟踪、图像识别、图像分类之外,还可以根据实际图像处理需求,包括其它图像处理操作,本公开对此不做具体限定。

[0069] 在一种可能的实现方式中,该图像处理方法通过自注意力神经网络实现,其中,自注意力神经网络中包括至少一个注意力模块,至少一个注意力模块中包括全局注意力子模块、局部注意力子模块和前馈子模块;基于全局注意力机制对第二图像块特征进行特征强化,得到第四图像块特征,包括:利用全局注意力子模块,基于全局注意力机制对第二图像块特征进行特征强化,得到第四图像块特征;基于局部注意力机制对第三图像块特征进行特征强化,得到第五图像块特征,包括:利用局部注意力子模块,基于局部注意力机制对第三图像块特征进行特征强化,得到第五图像块特征;根据第四图像块特征和第五图像块特征,确定目标图像对应的目标图像块特征,包括:利用前馈子模块,对第四图像块特征和第五图像块特征进行特征转换,得到目标图像块特征。

[0070] 通过在自注意力神经网络中设置注意力模块,注意力模块中包括全局注意力子模块、局部注意力子模块和前馈子模块,以使得可以利用全局注意力子模块进行基于全局注意力机制的特征强化,利用局部注意力子模块进行基于局部注意力机制的特征强化,以及利用前馈模块对基于全局注意力机制和局部注意力机制进行特征强化之后的图像块特征进一步进行特征转换,以得到既包括全局信息又包括局部信息的目标图像块特征,有效提高了目标图像块特征的语义表达能力,进而有效提高了自注意力神经网络的网络性能。

[0071] 图3示出相关技术中的一种自注意力神经网络的网络结构图。如图3所示,自注意力神经网络中包括M个注意力模块(G-Block_1至G-Block_M),各注意力模块均采用多头注意力(Multi-Head Attention,MHA)机制。如图3所示,各注意力模块中包括G=3个全局注意力子模块(即各全局注意力模块中包括3个注意力头)和一个前馈子模块(FeedForward Neural Network,FFN)。

[0072] 仍以上述L个第一图像块特征为例,针对G-Block_1,在G-Block_1对应的目标通道数是 d_k 的情况下,由于G-Block_1中包括G=3个全局注意力子模块,因此,将L个第一图像块特征沿通道维度平均分成 $N=3$ 个部分,每个部分包括通道数是 $d_k/3$ 的L个图像块特征,进而将3部分通道数是 $d_k/3$ 的L个图像块特征分别发送至每个全局注意力子模块,以在每个全局注意力子模块中分别执行上述公式(1)所示的基于全局注意力机制的特征强化,得到每个全局注意力子模块的输出结果,进而将每个全局注意力子模块的输出结果在通道维度进行融合后,输入FFN子模块进行进一步特征转换之后,得到G-Block_1的输出结果,该输出结果为通道数是 d_k 的L个图像块特征;进而将G-Block_1输出的通道数是 d_k 的L个图像块特征,输入G-Block_2,重复执行上述特征强化过程,直至得到G-Block_M的输出结果。

[0073] 如图3所示的相关技术中的自注意力神经网络,在每个注意力模块中仅包括全局注意力子模块和FFN子模块,因此,相关技术中的自注意力神经网络仅能够实现基于全局注意力机制的特征强化,导致忽略了目标图像中的局部信息,使得自注意力神经网络的网络性能较差。

[0074] 图4示出根据本公开实施例的一种自注意力神经网络的网络结构图。如图4所示,本公开提出的自注意力神经网络包括M个注意力模块(GL-Block_1至GL-Block_M)。至少一个注意力模块中包括G个全局注意力子模块、L个局部注意力子模块和一个FFN子模块。结合图3和图4,G-Block_2中包括G=3个全局注意力子模块,将其中的2个全局注意力子模块替换为两个局部注意力子模块之后,可以得到GL-Block_2,也就是说,GL-Block_2中包括G=1个全局注意力子模块、L=2个局部注意力子模块和一个FFN子模块。

[0075] 仍以上述L个第一图像块特征为例,针对GL-Block_1,在G-Block_1对应的目标通道数是 d_k ,且GL-Block_1中包括G=1个全局注意力子模块、L=2个局部注意力子模块的情况下,将L个第一图像块特征沿通道维度平均分成N=3个部分,每个部分包括通道数是 $d_k/3$ 的L个图像块特征,进而将其中一部分通道数是 $d_k/3$ 的L个图像块特征发送至全局注意力子模块,以及将剩余两部分通道数是 $d_k/3$ 的L个图像块特征分别发送至两个局部注意力子模块。

[0076] 针对GL-Block_1,在G=1个全局注意力子模块中,执行上述公式(1)所示的基于全局注意力机制的特征强化,得到全局注意力子模块的输出结果;在L=2个局部注意力子模块中,分别通过执行一维卷积处理,实现基于局部注意力机制的特征强化,得到L=2个局部注意力子模块的输出结果;进而将1个全局注意力子模块和2个局部注意力子模块的输出结果在通道维度进行融合后,输入FFN子模块进行进一步特征转换之后,得到GL-Block_1的输出结果,该输出结果为通道数是 d_k 的L个图像块特征;进而将GL-Block_1输出的通道数是 d_k 的L个图像块特征,输入G-Block_2,以及根据G-Block_2中全局注意力子模块和局部注意力子模块的分布情况,重复执行上述特征强化过程,直至得到GL-Block_M的输出结果。

[0077] GL-Block_M的输出结果为经过M次全局-局部特征强化之后的图像块特征,即为上述目标图像对应的目标图像块特征,如图4所示,自注意力神经网络中还包括图像处理模块(图中所示Head模块),将GL-Block_M输出的目标图像块特征输入Head模块进行相应的目标图像处理操作,以得到图像处理结果。Head模块可以根据实际图像处理需求进行设置,例如,可以是图像分类模块、目标检测模块、图像识别模块、图像跟踪模块等,本公开对此不做具体限定。

[0078] 通过在自注意力神经网络中引入局部注意力子模块,以使得在特征强化过程中既可以利用全局信息,又可以利用局部信息,以使得最终得到的目标图像块特征具有较高的语义表达能力,进而在利用目标图像特征对目标图像进行目标图像处理操作之后,可以得到精度较高的图像处理结果,从而有效提高自注意力神经网络的网络性能。

[0079] 在一种可能的实现方式中,局部注意力子模块可以包括两个pointwise卷积层,以及两个pointwise卷积层之间的一个depthwise卷积层。其中,第一个pointwise卷积层对应第一通道扩展率E,以实现通道数进行E倍扩展,之后,depthwise卷积层对应目标卷积核K,不会改变通道数,最后一个pointwise卷积层将通道数恢复到输入通道数。局部注意力子模块基于局部注意力机制的特征强化过程与上述相关描述类似,此处不再赘述。

[0080] 在一种可能的实现方式中,FFN子模块对应第二通道扩展率 d_z ,用于对各全局注意力子模块和各局部注意力子模块的输出结果进行一步的特征转换。FFN子模块的特征转换过程与上述相关描述类似,此处不再赘述。

[0081] 如图4所示的自注意力神经网络,在构建该自注意力神经网络时,各注意力模块对

应的模块分布参数(各注意力模块中包括的全局注意力子模块的个数 G ,以及局部注意力子模块的个数 L)、各注意力模块对应的模块结构超参数(各注意力模块对应的目标通道数 d_k 、各局部注意力子模块对应的目标卷积核 K 和第一通道扩展率 E 、以及各FFN子模块对应的第二通道扩展率 d_z),都是需要考虑的网络超参数。

[0082] 在一种可能的实现方式中,该图像处理方法还包括:构建第一网络结构搜索空间和第二网络结构搜索空间,其中,第一网络结构搜索空间中包括多个模块分布超参数,第二网络结构搜索空间中包括多个模块结构超参数;所述第一网络结构搜索空间,从多个模块分布超参数中确定目标模块分布超参数,其中,目标模块超参数用于指示各注意力模块中包括的全局注意力子模块的个数,以及局部注意力子模块的个数;根据目标模块分布超参数和第二网络结构搜索空间,从多个模块结构超参数中确定目标模块结构超参数,其中,目标模块结构超参数用于指示各注意力模块对应的目标通道数、各局部注意力子模块对应的目标卷积核和第一通道扩展率、以及各前馈模块对应的第二通道扩展率;根据目标模块分布超参数和目标模块结构超参数,构建自注意力神经网络。

[0083] 通过构建第一网络结构搜索空间和第二网络结构搜索空间,可以实现基于第一网络结构搜索空间和第二网络结构搜索空间,对目标模块分布超参数和目标结构超参数,进行分层级的网络结构搜索,以使得可以降低各层级网络结构搜索对应的搜索空间大小,有效提高搜索效率,从而快速构建得到自注意力神经网络。

[0084] 表1示出第一网络结构搜索空间和第二网络结构搜索空间的示例。

[0085] 表1

第一网络结构搜索空间	(G, L)	$(0, 3), (1, 2), (2, 1), (3, 0)$	
[0086] 第二网络结构搜索空间	局部注意力子模块	K	17, 31, 45
		E	1, 2, 4
	注意力模块	d_k	96, 192, 384
	FFN子模块	d_z	2, 4, 6

[0087] 如表1所示,第一网络结构搜索空间中包括各注意力模块中可能的全局注意力子模块和局部注意力子模块分布情况。例如, $(G, L) = (1, 2)$ 用于指示注意力模块中包括 $G=1$ 个全局注意力子模块和 $L=2$ 个局部注意力子模块。

[0088] 第二网络结构搜索空间中包括各注意力模块对应的模块结构参数。例如,各注意力模块对应的目标通道数 d_k (可选项包括96, 192, 384)、各局部注意力子模块对应的目标卷积核 K (可选项包括17, 31, 45)和第一通道扩展率 E (可选项包括1, 2, 4)、以及各FFN子模块对应的第二通道扩展率 d_z (可选项包括2, 4, 6)。

[0089] 模块分布超参数 (G, L) 和模块结构超参数 (K, E, d_k, d_z) 的可选项除了可以包括上述表1所示可选项以外,还可以根据实际情况确定其它可选项,各超参数对应的可选项的个数也可以根据实际情况确定,本公开对此不做具体限定。

[0090] 图5示出根据本公开实施例的分层级的网络结构搜索的示意图。如图5所示,首先,在第一网络结构搜索空间中进行高层级的网络结构搜索,从多个模块分布超参数中为每个自注意力模块确定目标模块分布超参数 (G, L) 。然后,固定每个自注意力模块中的目标模块

分布超参数(G,L)不变,并在第二网络结构搜索空间中进行低层级的网络结构搜索,从多个模块结构超参数中为每个自注意力模块确定目标模块结构超参数(K, E, d_k, d_v)。

[0091] 在一种可能的实现方式中,根据第一网络结构搜索空间,从多个模块分布超参数中确定目标模块分布超参数,包括:根据第一网络结构搜索空间,构建第一超级网络,其中,第一超级网络中包括根据多个模块分布超参数构建的多个第一可选网络结构;通过对第一超级网络进行网络训练,从多个第一可选网络结构中确定第一目标网络结构;将第一目标网络结构对应的模块分布超参数确定为目标模块分布超参数。

[0092] 根据第一网络结构搜索空间中包括的多个模块分布超参数,可以构建得到多个第一可选网络结构,进而可以构建包括多个第一可选网络结构的第一超级网络,通过对第一超级网络进行训练,以实现在第一网络结构搜索空间中进行高层级的网络结构搜索,从多个第一可选网络结构中确定出性能最优的目标第一网络结构,并将目标第一网络结构对应的模块分布超参数确定为目标模块分布超参数。

[0093] 例如,对第一超级网络进行训练,得到训练后的多个第一可选网络架构;基于进化算法对训练后的多个第一可选网络架构进行准确性验证,并从中选择出第一预设数目的准确性排名最优的第一可选网络架构;对选择出的该预设数目的第一可选网络架构进行再次训练,以使得从中确定出网络性能符合需求的目标第一可选网络架构,进而将目标第一可选网络架构对应的模块分布超参数确定为目标模块分布超参数。第一预设数目的大小可以根据实际情况确定,本公开对此不做具体限定。

[0094] 如图5所示,在表1所示的第一网络结构搜索空间中进行上述高层级的网络结构搜索之后,得到GL-Block_1对应的目标模块分布超参数($G=1, L=2$);GL-Block_2对应的目标模块分布超参数($G=3, L=0$);……;GL-Block_M对应的目标模块分布超参数($G=0, L=3$)。

[0095] 在一种可能的实现方式中,根据目标模块分布超参数和第二网络结构搜索空间,从多个模块结构超参数中确定目标模块结构超参数,包括:根据目标模块分布超参数和第二网络结构搜索空间,构建第二超级网络,其中,第二超级网络中包括根据多个模块结构超参数构建的多个第二可选网络结构;通过对第二超级网络进行网络训练,从多个第二可选网络结构中确定第二目标网络结构;将第二目标网络结构对应的模块结构超参数确定为目标模块结构超参数。

[0096] 固定各注意力模块对应的目标模块分布超参数不变,进而根据第二网络结构搜索空间中包括的多个模块结构超参数,可以构建得到多个第二可选网络结构,进而可以构建包括多个第二可选网络结构的第二超级网络,通过对第二超级网络进行训练,以实现在第二网络结构搜索空间中进行低层级的网络结构搜索,从多个第二可选网络结构中确定出性能最优的目标第二网络结构,并将目标第二网络结构对应的模块结构超参数确定为目标模块结构超参数。

[0097] 在一种可能的实现方式中,相同注意力模块中,各局部注意力子模块对应相同的目标卷积核和第一通道扩展率。

[0098] 自注意力神经网络的构建原则可以包括:针对相同注意力模块,该注意力模块中的各局部注意力子模块对应相同的目标卷积核K和第一通道扩展率E。例如,针对上述GL-Block_1,GL-Block_1中包括L=2个局部注意力子模块,该L=2个局部注意力子模块对应相同的目标卷积核K和第一通道扩展率E。基于自注意力神经网络的构建原则,可以删除第二

超级网络中包括的不符合该原则的第二可选网络结构,以减少第二网络搜索空间的大小,提高后续对目标第二网络结构的搜索效率。

[0099] 在基于自注意力神经网络的构建原则对第二网络结构搜索空间进行缩小之后,第二超级网络中包括的都是符合自注意力神经网络的构建原则的第二可选网络结构。通过对第二超级网络进行训练,以实现在第二网络结构搜索空间中进行低层级的网络结构搜索,确定目标模块结构超参数。

[0100] 例如,对第二超级网络进行训练,得到训练后的多个第二可选网络架构;基于进化算法对训练后的多个第二可选网络架构进行准确性验证,并从中选择出第二预设数目的准确性排名最优的第二可选网络架构;对选择出的该预设数目的第二可选网络架构进行再次训练,以使得从中确定出网络性能符合需求的目标第二可选网络架构,进而将目标第二可选网络架构对应的模块结构超参数确定为目标结构超参数。第一预设数目的大小可以根据实际情况确定,本公开对此不做具体限定。

[0101] 如图5所示,在表1所示的第二网络结构搜索空间中进行上述低层级的网络结构搜索之后,得到GL-Block_1对应的目标模块结构超参数($K=17, E=4, d_k=92, d_z=2$);GL-Block_2对应的目标模块结构超参数($K=45, E=1$);……;GL-Block_M对应的目标模块结构超参数($d_k=192, d_z=3$)。

[0102] 根据高层级的网络结构搜索得到的目标模块分布超参数,以及根据低层级的网络结构搜索得到的目标模块结构超参数,可以构建得到自注意力神经网络。自注意力神经网络中包括至少一个自注意力模块,至少一个自注意力模块中包括全局注意力子模块和局部注意力子模块,以使得自注意力神经网络在进行特征强化过程中,既可以利用全局信息,又可以利用局部信息,从而有效提高自注意力神经网络的网络性能。

[0103] 根据图像处理需要,本公开的自注意力神经网络可以应用于目标检测、目标跟踪、图像识别、图像分类等图像处理任务,本公开对此不做具体限定。

[0104] 可以理解,本公开提及的上述各个方法实施例,在不违背原理逻辑的情况下,均可以彼此相互结合形成结合后的实施例,限于篇幅,本公开不再赘述。本领域技术人员可以理解,在具体实施方式的上述方法中,各步骤的具体执行顺序应当以其功能和可能的内在逻辑确定。

[0105] 此外,本公开还提供了图像处理装置、电子设备、计算机可读存储介质、程序,上述均可用来实现本公开提供的任一种图像处理方法,相应技术方案和描述和参见方法部分的相应记载,不再赘述。

[0106] 图6示出根据本公开实施例的一种图像处理装置的框图。如图6所示,装置60包括:

[0107] 特征确定模块61,用于确定目标图像对应的第一图像块特征,以及将第一图像块特征划分为第二图像块特征和第三图像块特征;

[0108] 注意力模块62,用于基于全局注意力机制对第二图像块特征进行特征强化,得到第四图像块特征,以及基于局部注意力机制对第三图像块特征进行特征强化,得到第五图像块特征;

[0109] 第一确定模块63,用于根据第四图像块特征和第五图像块特征,确定目标图像对应的目标图像块特征;

[0110] 目标图像处理模块64,用于根据目标图像块特征,对目标图像进行目标图像处理

操作,得到图像处理结果。

[0111] 在一种可能的实现方式中,第一图像块特征对应的通道数是目标通道数;

[0112] 特征确定模块61,具体用于:

[0113] 根据目标通道数和目标比例,将第一图像块特征在通道维度上进行划分,得到第二图像块特征和第三图像块特征。

[0114] 在一种可能的实现方式中,注意力模块62,包括:全局注意力子模块,具体用于:

[0115] 根据第二图像块特征,确定第一特征向量、第二特征向量和第三特征向量;

[0116] 根据第一特征向量和所述第二特征向量,确定注意力特征图;

[0117] 根据注意力特征图和第三特征向量,确定第四图像块特征。

[0118] 在一种可能的实现方式中,注意力模块62,还包括:局部注意力子模块,具体用于:

[0119] 根据目标卷积核和第一通道扩展率,对第三图像块特征进行卷积处理,得到第五图像块特征。

[0120] 在一种可能的实现方式中,第一确定模块63,具体用于:

[0121] 根据第二通道扩展率,对第四图像块特征和第五图像块特征进行特征转换,得到目标图像块特征。

[0122] 在一种可能的实现方式中,装置60通过自注意力神经网络实现图像处理方法,其中,自注意力神经网络中包括至少一个注意力模块62,至少一个注意力模块62中包括全局注意力子模块、局部注意力子模块和前馈子模块;

[0123] 全局注意力子模块,用于基于全局注意力机制对第二图像块特征进行特征强化,得到第四图像块特征;

[0124] 局部注意力子模块,用于基于局部注意力机制对第三图像块特征进行特征强化,得到第五图像块特征;

[0125] 前馈子模块,用于对第四图像块特征和第五图像块特征进行特征转换,得到目标图像块特征。

[0126] 在一种可能的实现方式中,装置60还包括:

[0127] 搜索空间构建模块,用于构建第一网络结构搜索空间和第二网络结构搜索空间,其中,第一网络结构搜索空间中包括多个模块分布超参数,第二网络结构搜索空间中包括多个模块结构超参数;

[0128] 第二确定模块,用于根据第一网络结构搜索空间,从多个模块分布超参数中确定目标模块分布超参数,其中,目标模块超参数用于指示各注意力模块中包括的全局注意力子模块的个数,以及局部注意力子模块的个数;

[0129] 第三确定模块,用于根据目标模块分布超参数和第二网络结构搜索空间,从多个模块结构超参数中确定目标模块结构超参数,其中,目标模块结构超参数用于指示各注意力模块对应的目标通道数、各局部注意力子模块对应的目标卷积核和第一通道扩展率、以及各前馈模块对应的第二通道扩展率;

[0130] 网络构建模块,用于根据目标模块分布超参数和目标模块结构超参数,构建自注意力神经网络。

[0131] 在一种可能的实现方式中,第二确定模块,具体用于:

[0132] 根据第一网络结构搜索空间,构建第一超级网络,其中,第一超级网络中包括根据

多个模块分布超参数构建的多个第一可选网络结构；

[0133] 通过对第一超级网络进行网络训练，从多个第一可选网络结构中确定第一目标网络结构；

[0134] 将第一目标网络结构对应的模块分布超参数确定为目标模块分布超参数。

[0135] 在一种可能的实现方式中，第三确定模块，具体用于：

[0136] 根据目标模块分布超参数和第二网络结构搜索空间，构建第二超级网络，其中，第二超级网络中包括根据多个模块结构超参数构建的多个第二可选网络结构；

[0137] 通过对第二超级网络进行网络训练，从多个第二可选网络结构中确定第二目标网络结构；

[0138] 将第二目标网络结构对应的模块结构超参数确定为目标模块结构超参数。

[0139] 在一种可能的实现方式中，相同注意力模块中，各局部注意力子模块对应相同的目标卷积核和第一通道扩展率。

[0140] 在一些实施例中，本公开实施例提供的装置具有的功能或包含的模块可以用于执行上文方法实施例描述的方法，其具体实现可以参照上文方法实施例的描述，为了简洁，这里不再赘述。

[0141] 本公开实施例还提出一种计算机可读存储介质，其上存储有计算机程序指令，所述计算机程序指令被处理器执行时实现上述方法。计算机可读存储介质可以是易失性或非易失性计算机可读存储介质。

[0142] 本公开实施例还提出一种电子设备，包括：处理器；用于存储处理器可执行指令的存储器；其中，所述处理器被配置为调用所述存储器存储的指令，以执行上述方法。

[0143] 本公开实施例还提供了一种计算机程序产品，包括计算机可读代码，或者承载有计算机可读代码的非易失性计算机可读存储介质，当所述计算机可读代码在电子设备的处理器中运行时，所述电子设备中的处理器执行上述方法。

[0144] 电子设备可以被提供为终端、服务器或其它形态的设备。

[0145] 图7示出根据本公开实施例的一种电子设备的框图。如图7所示，电子设备800可以是移动电话，计算机，数字广播终端，消息收发设备，游戏控制台，平板设备，医疗设备，健身设备，个人数字助理等终端。

[0146] 参照图7，电子设备800可以包括以下一个或多个组件：处理组件802，存储器804，电源组件806，多媒体组件808，音频组件810，输入/输出(I/O)的接口812，传感器组件814，以及通信组件816。

[0147] 处理组件802通常控制电子设备800的整体操作，诸如与显示，电话呼叫，数据通信，相机操作和记录操作相关联的操作。处理组件802可以包括一个或多个处理器820来执行指令，以完成上述的方法的全部或部分步骤。此外，处理组件802可以包括一个或多个模块，便于处理组件802和其他组件之间的交互。例如，处理组件802可以包括多媒体模块，以方便多媒体组件808和处理组件802之间的交互。

[0148] 存储器804被配置为存储各种类型的数据以支持在电子设备800的操作。这些数据的示例包括用于在电子设备800上操作的任何应用程序或方法的指令，联系人数据，电话簿数据，消息，图片，视频等。存储器804可以由任何类型的易失性或非易失性存储设备或者它们的组合实现，如静态随机存取存储器(SRAM)，电可擦除可编程只读存储器(EEPROM)，可擦

除可编程只读存储器 (EPROM), 可编程只读存储器 (PROM), 只读存储器 (ROM), 磁存储器, 快闪存储器, 磁盘或光盘。

[0149] 电源组件806为电子设备800的各种组件提供电力。电源组件806可以包括电源管理系统, 一个或多个电源, 及其他与为电子设备800生成、管理和分配电力相关联的组件。

[0150] 多媒体组件808包括在所述电子设备800和用户之间的提供一个输出接口的屏幕。在一些实施例中, 屏幕可以包括液晶显示器 (LCD) 和触摸面板 (TP)。如果屏幕包括触摸面板, 屏幕可以被实现为触摸屏, 以接收来自用户的输入信号。触摸面板包括一个或多个触摸传感器以感测触摸、滑动和触摸面板上的手势。所述触摸传感器可以不仅感测触摸或滑动动作的边界, 而且还检测与所述触摸或滑动操作相关的持续时间和压力。在一些实施例中, 多媒体组件808包括一个前置摄像头和/或后置摄像头。当电子设备800处于操作模式, 如拍摄模式或视频模式时, 前置摄像头和/或后置摄像头可以接收外部的多媒体数据。每个前置摄像头和后置摄像头可以是一个固定的光学透镜系统或具有焦距和光学变焦能力。

[0151] 音频组件810被配置为输出和/或输入音频信号。例如, 音频组件810包括一个麦克风 (MIC), 当电子设备800处于操作模式, 如呼叫模式、记录模式和语音识别模式时, 麦克风被配置为接收外部音频信号。所接收的音频信号可以被进一步存储在存储器804或经由通信组件816发送。在一些实施例中, 音频组件810还包括一个扬声器, 用于输出音频信号。

[0152] I/O接口812为处理组件802和外围接口模块之间提供接口, 上述外围接口模块可以是键盘, 点击轮, 按钮等。这些按钮可包括但不限于: 主页按钮、音量按钮、启动按钮和锁定按钮。

[0153] 传感器组件814包括一个或多个传感器, 用于为电子设备800提供各个方面的状态评估。例如, 传感器组件814可以检测到电子设备800的打开/关闭状态, 组件的相对定位, 例如所述组件为电子设备800的显示器和小键盘, 传感器组件814还可以检测电子设备800或电子设备800一个组件的位置改变, 用户与电子设备800接触的存在或不存在, 电子设备800方位或加速/减速和电子设备800的温度变化。传感器组件814可以包括接近传感器, 被配置用来在没有任何的物理接触时检测附近物体的存在。传感器组件814还可以包括光传感器, 如互补金属氧化物半导体 (CMOS) 或电荷耦合装置 (CCD) 图像传感器, 用于在成像应用中使用。在一些实施例中, 该传感器组件814还可以包括加速度传感器, 陀螺仪传感器, 磁传感器, 压力传感器或温度传感器。

[0154] 通信组件816被配置为便于电子设备800和其他设备之间有线或无线方式的通信。电子设备800可以接入基于通信标准的无线网络, 如无线网络 (WiFi), 第二代移动通信技术 (2G) 或第三代移动通信技术 (3G), 或它们的组合。在一个示例性实施例中, 通信组件816经由广播信道接收来自外部广播管理系统的广播信号或广播相关信息。在一个示例性实施例中, 所述通信组件816还包括近场通信 (NFC) 模块, 以促进短程通信。例如, 在NFC模块可基于射频识别 (RFID) 技术, 红外数据协会 (IrDA) 技术, 超宽带 (UWB) 技术, 蓝牙 (BT) 技术和其他技术来实现。

[0155] 在示例性实施例中, 电子设备800可以被一个或多个应用专用集成电路 (ASIC)、数字信号处理器 (DSP)、数字信号处理设备 (DSPD)、可编程逻辑器件 (PLD)、现场可编程门阵列 (FPGA)、控制器、微控制器、微处理器或其他电子元件实现, 用于执行上述方法。

[0156] 在示例性实施例中, 还提供了一种非易失性计算机可读存储介质, 例如包括计算

机程序指令的存储器804,上述计算机程序指令可由电子设备800的处理器820执行以完成上述方法。

[0157] 图8示出根据本公开实施例的一种电子设备的框图。如图8所示,电子设备1900可以被提供为一服务器。参照图8,电子设备1900包括处理组件1922,其进一步包括一个或多个处理器,以及由存储器1932所代表的存储器资源,用于存储可由处理组件1922的执行的指令,例如应用程序。存储器1932中存储的应用程序可以包括一个或一个以上的每一个对应于一组指令的模块。此外,处理组件1922被配置为执行指令,以执行上述方法。

[0158] 电子设备1900还可以包括一个电源组件1926被配置为执行电子设备1900的电源管理,一个有线或无线网络接口1950被配置为将电子设备1900连接到网络,和一个输入输出(I/O)接口1958。电子设备1900可以操作基于存储在存储器1932的操作系统,例如微软服务器操作系统(Windows Server™),苹果公司推出的基于图形用户界面操作系统(Mac OS X™),多用户多进程的计算机操作系统(Unix™),自由和开放原代码的类Unix操作系统(Linux™),开放原代码的类Unix操作系统(FreeBSD™)或类似。

[0159] 在示例性实施例中,还提供了一种非易失性计算机可读存储介质,例如包括计算机程序指令的存储器1932,上述计算机程序指令可由电子设备1900的处理组件1922执行以完成上述方法。

[0160] 本公开可以是系统、方法和/或计算机程序产品。计算机程序产品可以包括计算机可读存储介质,其上载有用于使处理器实现本公开的各个方面的计算机可读程序指令。

[0161] 计算机可读存储介质可以是保持和存储由指令执行设备使用的指令的有形设备。计算机可读存储介质例如可以是(但不限于)电存储设备、磁存储设备、光存储设备、电磁存储设备、半导体存储设备或者上述的任意合适的组合。计算机可读存储介质的更具体的例子(非穷举的列表)包括:便携式计算机盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、静态随机存取存储器(SRAM)、便携式压缩盘只读存储器(CD-ROM)、数字多功能盘(DVD)、记忆棒、软盘、机械编码设备、例如其上存储有指令的打孔卡或凹槽内凸起结构、以及上述的任意合适的组合。这里所使用的计算机可读存储介质不被解释为瞬时信号本身,诸如无线电波或者其他自由传播的电磁波、通过波导或其他传输媒介传播的电磁波(例如,通过光纤电缆的光脉冲)、或者通过电线传输的信号。

[0162] 这里所描述的计算机可读程序指令可以从计算机可读存储介质下载到各个计算/处理设备,或者通过网络、例如因特网、局域网、广域网和/或无线网下载到外部计算机或外部存储设备。网络可以包括铜传输电缆、光纤传输、无线传输、路由器、防火墙、交换机、网关计算机和/或边缘服务器。每个计算/处理设备中的网络适配卡或者网络接口从网络接收计算机可读程序指令,并转发该计算机可读程序指令,以供存储在各个计算/处理设备中的计算机可读存储介质中。

[0163] 用于执行本公开操作的计算机程序指令可以是汇编指令、指令集架构(ISA)指令、机器指令、机器相关指令、微代码、固件指令、状态设置数据、或者以一种或多种编程语言的任意组合编写的源代码或目标代码,所述编程语言包括面向对象的编程语言—诸如Smalltalk、C++等,以及常规的过程式编程语言—诸如“C”语言或类似的编程语言。计算机可读程序指令可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独

立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中,远程计算机可以通过任意种类的网络—包括局域网(LAN)或广域网(WAN)—连接到用户计算机,或者,可以连接到外部计算机(例如利用因特网服务提供商来通过因特网连接)。在一些实施例中,通过利用计算机可读程序指令的状态信息来个性化定制电子电路,例如可编程逻辑电路、现场可编程门阵列(FPGA)或可编程逻辑阵列(PLA),该电子电路可以执行计算机可读程序指令,从而实现本公开的各个方面。

[0164] 这里参照根据本公开实施例的方法、装置(系统)和计算机程序产品的流程图和/或框图描述了本公开的各个方面。应当理解,流程图和/或框图的每个方框以及流程图和/或框图中各方框的组合,都可以由计算机可读程序指令实现。

[0165] 这些计算机可读程序指令可以提供给通用计算机、专用计算机或其它可编程数据处理装置的处理器,从而生产出一种机器,使得这些指令在通过计算机或其它可编程数据处理装置的处理器执行时,产生了实现流程图和/或框图中的一个或多个方框中规定的功能/动作的装置。也可以把这些计算机可读程序指令存储在计算机可读存储介质中,这些指令使得计算机、可编程数据处理装置和/或其他设备以特定方式工作,从而,存储有指令的计算机可读介质则包括一个制品,其包括实现流程图和/或框图中的一个或多个方框中规定的功能/动作的各个方面的指令。

[0166] 也可以把计算机可读程序指令加载到计算机、其它可编程数据处理装置、或其它设备上,使得在计算机、其它可编程数据处理装置或其它设备上执行一系列操作步骤,以产生计算机实现的过程,从而使得在计算机、其它可编程数据处理装置、或其它设备上执行的指令实现流程图和/或框图中的一个或多个方框中规定的功能/动作。

[0167] 附图中的流程图和框图显示了根据本公开的多个实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段或指令的一部分,所述模块、程序段或指令的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。在有些作为替换的实现中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个连续的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意的,框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合,可以用执行规定的功能或动作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0168] 该计算机程序产品可以具体通过硬件、软件或其结合的方式实现。在一个可选实施例中,所述计算机程序产品具体体现为计算机存储介质,在另一个可选实施例中,计算机程序产品具体体现为软件产品,例如软件开发包(Software Development Kit, SDK)等等。

[0169] 以上已经描述了本公开的各实施例,上述说明是示例性的,并非穷尽性的,并且也不限于所披露的各实施例。在不偏离所说明的各实施例的范围和精神的情况下,对于本技术领域的普通技术人员来说许多修改和变更都是显而易见的。本文中术语的选择,旨在最好地解释各实施例的原理、实际应用或对市场中的技术的改进,或者使本技术领域的其它普通技术人员能理解本文披露的各实施例。

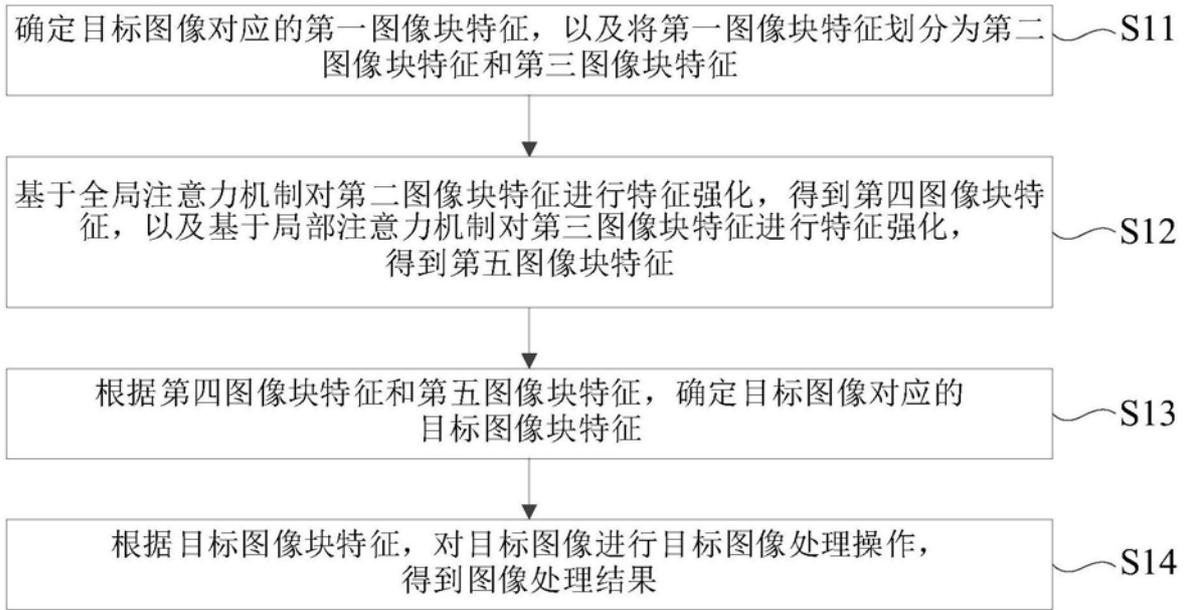


图1

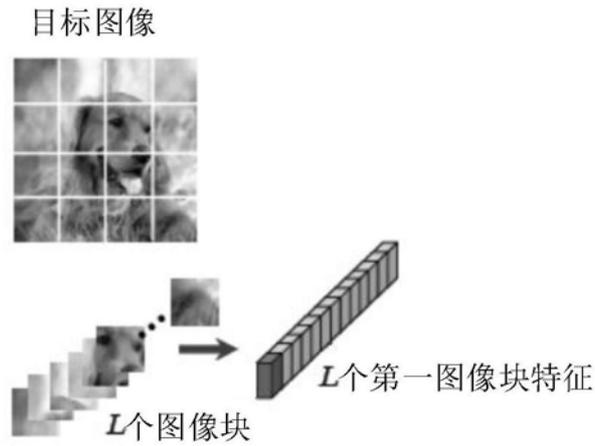


图2

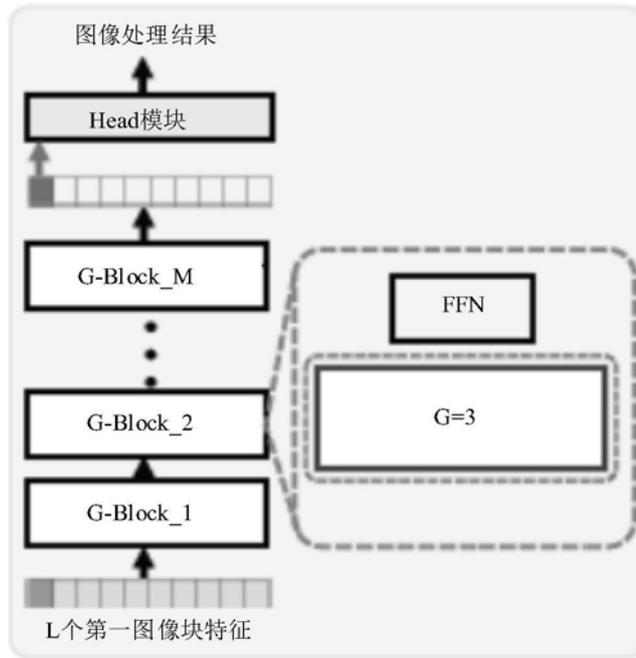


图3

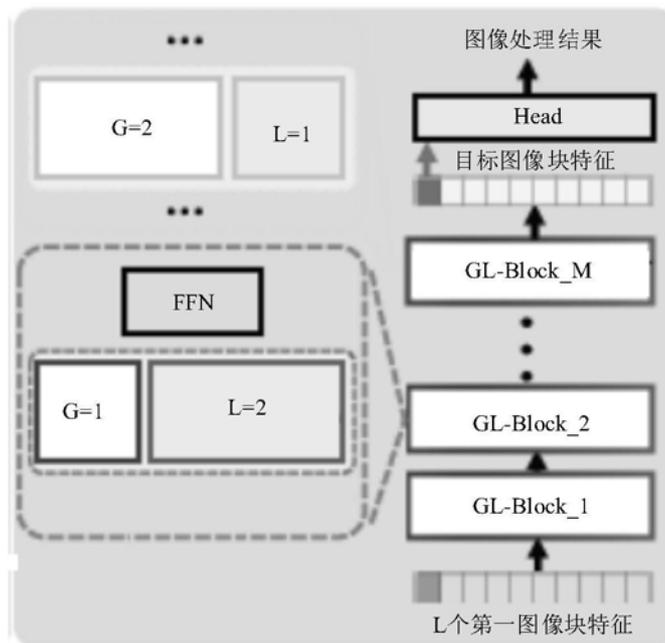


图4

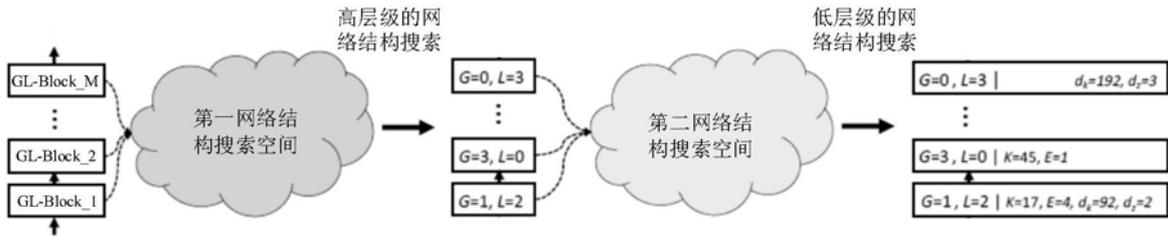


图5

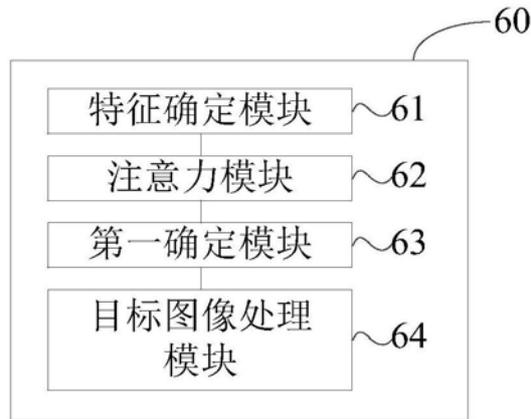


图6

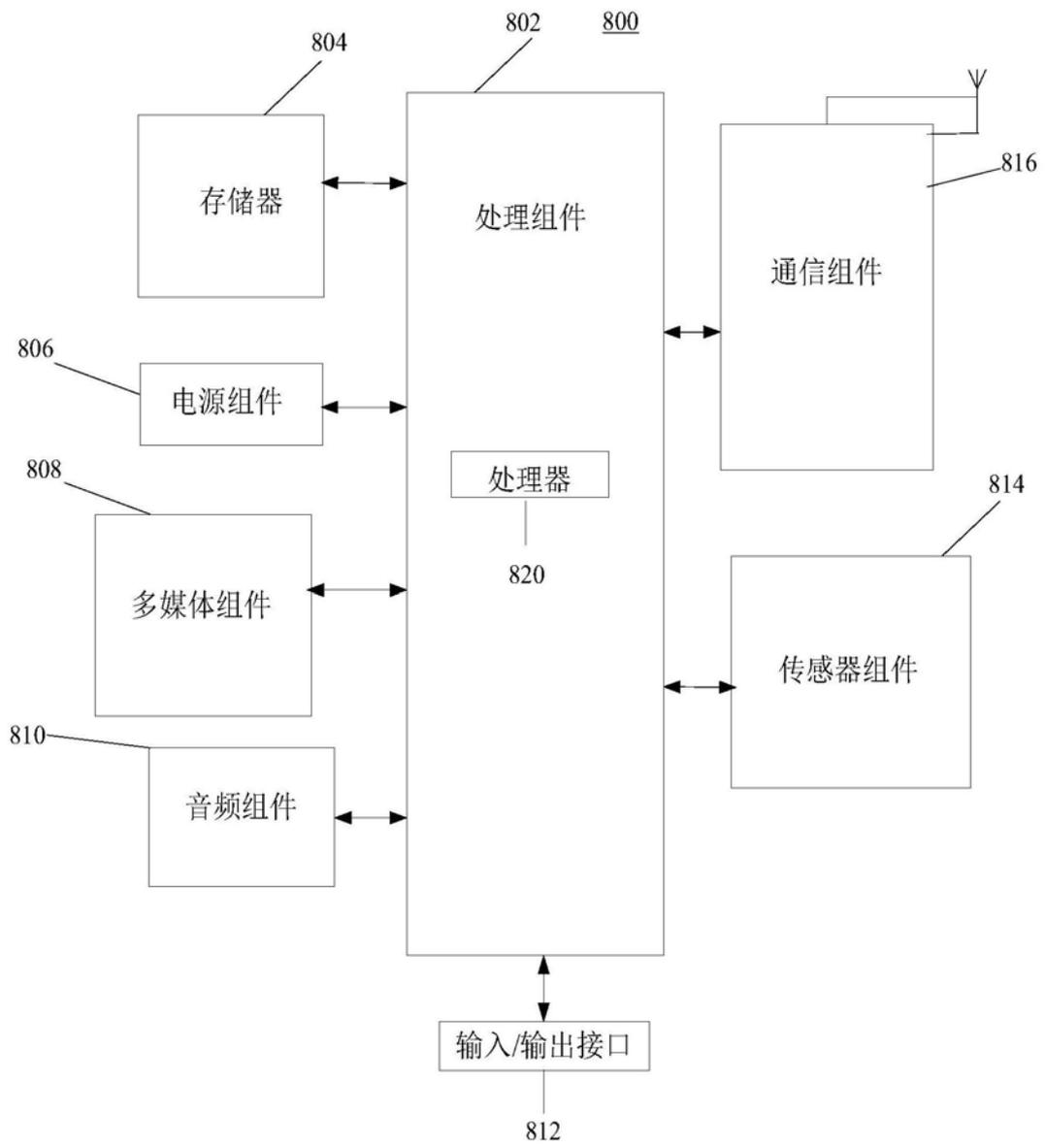


图7

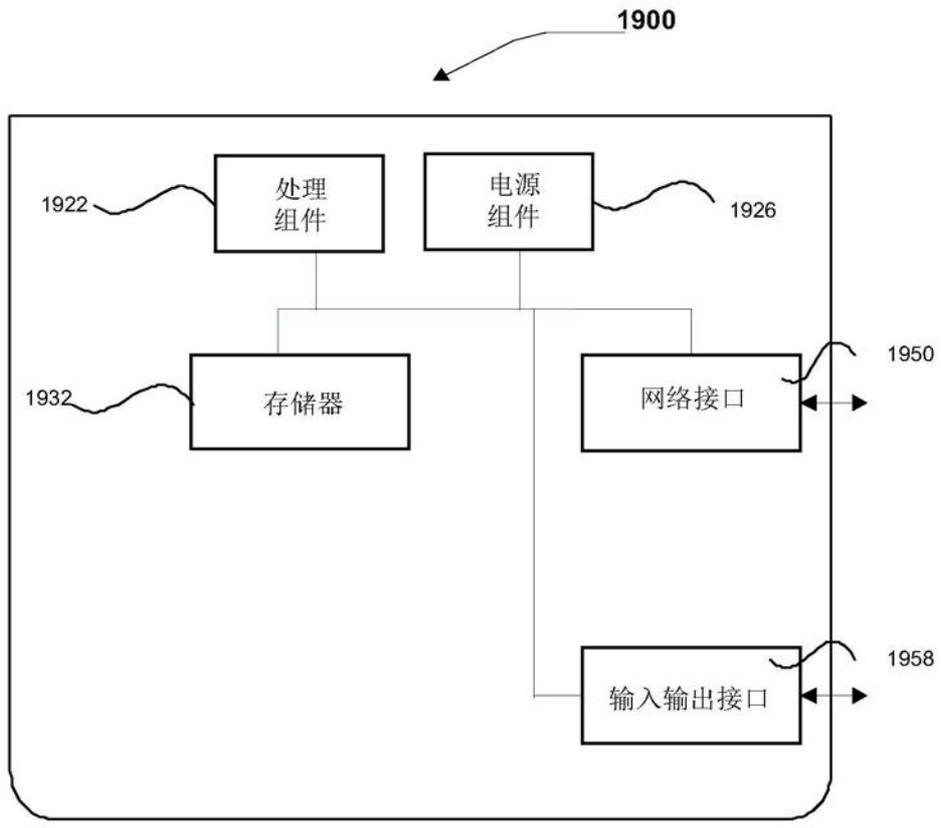


图8