



(12) 发明专利申请

(10) 申请公布号 CN 115131604 A

(43) 申请公布日 2022. 09. 30

(21) 申请号 202210593162.5

G06K 9/62 (2022.01)

(22) 申请日 2022.05.27

(71) 申请人 腾讯科技(深圳)有限公司

地址 518044 广东省深圳市南山区高新区
科技中一路腾讯大厦35层

(72) 发明人 詹佳伟 刘俊

(74) 专利代理机构 北京同达信恒知识产权代理
有限公司 11291

专利代理师 朱佳

(51) Int. Cl.

G06V 10/764 (2022.01)

G06V 10/42 (2022.01)

G06V 10/44 (2022.01)

G06V 10/80 (2022.01)

G06V 10/82 (2022.01)

权利要求书3页 说明书19页 附图12页

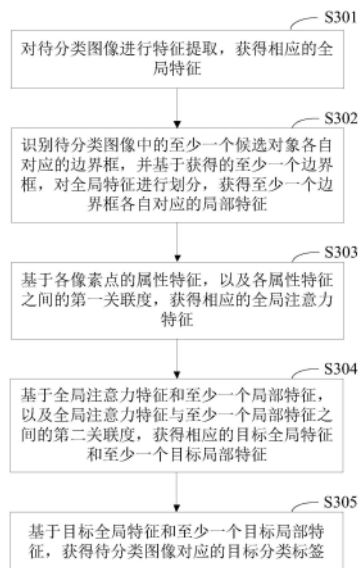
(54) 发明名称

一种多标签图像分类方法、装置、电子设备
及存储介质

(57) 摘要

本申请公开一种多标签图像分类方法、装置、电子设备及存储介质,涉及计算机技术领域。本申请实施例可应用于云技术、人工智能、智慧交通、辅助驾驶等各种场景中。在该方法中,对待分类图像进行特征提取,获得相应全局特征,并通过对待分类图像中候选对象进行识别获得的至少一个边界框,对全局特征进行划分,获得相应的至少一个局部特征,将通过对全局特征执行自注意力机制获得的全局注意力特征,与至少一个局部特征再次执行自注意力机制,获得相应的目标全局特征和至少一个目标局部特征,进而根据目标全局特征和至少一个目标局部特征,获得待分类图像对应的目标分类标签。与相关技术相比,可以有效地提高对图像进行多标签分类的准确度。

CN 115131604 A



1. 一种多标签图像分类方法,其特征在于,包括:

对待分类图像进行特征提取,获得相应全局特征;所述全局特征包含所述待分类图像中的各像素点的属性特征;

识别所述待分类图像中的至少一个候选对象各自对应的边界框,并基于获得的至少一个边界框,对所述全局特征进行划分,获得所述至少一个边界框各自对应的局部特征;

基于所述各像素点的属性特征,以及各属性特征之间的第一关联度,获得相应全局注意力特征;

基于所述全局注意力特征和至少一个局部特征,以及所述全局注意力特征与所述至少一个局部特征之间的第二关联度,获得相应的目标全局特征和至少一个目标局部特征;

基于所述目标全局特征和所述至少一个目标局部特征,获得所述待分类图像对应的目标分类标签。

2. 根据权利要求1所述的方法,其特征在于,所述基于所述各像素点的属性特征,以及各属性特征之间的第一关联度,获得相应全局注意力特征,包括:

针对所述各像素点,分别执行以下操作:将一个像素点作为查询像素点,所述各像素点中的其他像素点作为关键像素点;基于自注意力机制,将所述查询像素点的属性特征,分别与各关键像素点各自的属性特征进行关联度匹配,获得各属性特征之间的第一关联度;

根据各第一关联度,对所述各像素点的属性特征进行加权合并,获得相应全局注意力特征。

3. 根据权利要求1所述的方法,其特征在于,基于所述全局注意力特征和至少一个局部特征,以及所述全局注意力特征与所述至少一个局部特征之间的第二关联度,获得相应的目标全局特征和至少一个目标局部特征,包括:

针对至少一个局部特征,分别执行以下操作:基于自注意力机制,将一个局部特征中的各像素点的属性特征,分别与所述全局注意力特征中的各像素点的注意力特征进行关联度匹配,获得所述一个局部特征与所述全局注意力特征之间的第二关联度;

根据各第二关联度,对所述全局注意力特征和所述至少一个局部特征分别进行加权合并,获得相应的目标全局特征和至少一个目标局部特征。

4. 根据权利要求1、2或3所述的方法,其特征在于,所述基于所述目标全局特征和所述至少一个目标局部特征,获得所述待分类图像对应的目标分类标签,包括:

基于多标签分类模型,分别确定出所述目标全局特征对应的全局分类标签结果和所述至少一个目标局部特征各自对应的局部分类标签结果;

根据所述全局分类标签结果和各局部分类标签结果,获得所述待分类图像对应的目标分类标签。

5. 根据权利要求4所述的方法,其特征在于,所述基于多标签分类模型,分别确定出所述目标全局特征对应的全局分类标签结果和所述至少一个目标局部特征各自对应的局部分类标签结果,包括:

将所述目标全局特征输入多标签分类模型,获得所述目标全局特征分别属于各个候选分类的第一概率值,并将各第一概率值作为全局分类标签结果;

将所述至少一个目标局部特征,分别输入所述多标签分类模型,获得所述至少一个目标局部特征各自属于各个候选分类的第二概率值,并将各第二概率值作为局部分类标签结

果。

6. 根据权利要求5所述的方法,其特征在于,所述根据所述全局分类标签结果和各局部分类标签结果,获得所述待分类图像对应的目标分类标签,包括:

针对所述各个候选分类,分别执行以下操作:根据所述至少一个目标局部特征各自属于一个候选分类的第二概率值,将符合设定取值条件的第二概率值,与所述一个候选分类对应的第一概率值进行平均,确定所述一个候选分类对应的目标概率值;

根据所述各个候选分类各自对应的目标概率值,将所述目标概率值大于设定阈值的候选分类,作为所述待分类图像对应的目标分类标签。

7. 根据权利要求4所述的方法,其特征在于,所述多标签分类模型的训练过程包括:

获取训练数据集;所述训练数据集中包括多个图像样本,所述图像样本中标注有设定分类标签;

基于所述训练数据集,对所述多标签分类模型进行迭代训练,直到满足设定的收敛条件为止,其中,一次迭代训练过程包括:

基于从所述训练数据集中抽取的图像样本,获得所述图像样本对应的目标全局样本特征和至少一个目标局部样本特征;

通过所述多标签分类模型,根据所述目标全局样本特征和所述至少一个目标局部样本特征,确定所述图像样本对应的目标样本分类标签,并基于所述目标样本分类标签与所述设定分类标签确定的损失值,对所述多标签分类模型进行参数调整。

8. 根据权利要求7所述的方法,其特征在于,所述获得所述图像样本对应的目标全局样本特征和至少一个目标局部样本特征,包括:

对所述图像样本进行特征提取,获得相应全局样本特征,并基于对所述图像样本中的至少一个候选样本对象进行识别获得的至少一个边界框,对所述全局样本特征进行划分,获得所述至少一个样本边界框各自对应的局部样本特征;

根据所述全局样本特征和至少一个局部样本特征,获得相应的目标全局样本特征和至少一个目标局部样本特征。

9. 根据权利要求7所述的方法,其特征在于,所述通过所述多标签分类模型,根据所述目标全局样本特征和所述至少一个目标局部样本特征,确定所述图像样本对应的目标样本分类标签,包括:

通过所述多标签分类模型,确定所述目标全局样本特征分别属于各个候选样本分类的第一样本概率值,以及所述至少一个目标局部样本特征各自属于各个候选样本分类的第二样本概率值;

根据所述各个候选样本分类对应的第一样本概率值和第二样本概率值,确定所述图像样本对应的目标样本分类标签。

10. 一种多标签图像分类装置,其特征在于,包括:

全局特征提取单元,用于对待分类图像进行特征提取,获得相应全局特征;所述全局特征包含所述待分类图像中的各像素点的属性特征;

局部特征确定单元,用于识别所述待分类图像中的至少一个候选对象各自对应的边界框,并基于获得的至少一个边界框,对所述全局特征进行划分,获得所述至少一个边界框各自对应的局部特征;

全局注意力处理单元,用于基于所述各像素点的属性特征,以及各属性特征之间的第一关联度,获得相应全局注意力特征;

目标特征确定单元,用于基于所述全局注意力特征和至少一个局部特征,以及所述全局注意力特征与所述至少一个局部特征之间的第二关联度,获得相应的目标全局特征和至少一个目标局部特征;

多标签分类单元,基于所述目标全局特征和所述至少一个目标局部特征,获得所述待分类图像对应的目标分类标签。

11. 一种电子设备,其特征在於,其包括处理器和存储器,其中,所述存储器存储有程序代码,当所述程序代码被所述处理器执行时,使得所述处理器执行权利要求1~9中任一项所述方法的步骤。

12. 一种计算机可读存储介质,其特征在於,其包括程序代码,当所述程序代码在电子设备上运行时,所述程序代码用于使所述电子设备执行权利要求1~9中任一项所述方法的步骤。

13. 一种计算机程序产品,包括计算机程序/指令,其特征在於,该计算机程序/指令被处理器执行时实现权利要求1~9中任一项所述方法的步骤。

一种多标签图像分类方法、装置、电子设备及存储介质

技术领域

[0001] 本申请实施例涉及计算机技术领域,尤其涉及一种多标签图像分类方法、装置、电子设备及存储介质。

背景技术

[0002] 随着计算机视觉技术和多媒体技术的快速发展,多标签图像分类(Multi-label Image Classification)已经在图像检索、属性识别、自动图像注释等领域得到了广泛的应用。多标签图像分类主要用于对图像进行分类识别,从而将图像划归到一个或多个标签中,例如,通过对一个图像进行多标签分类,可以识别出该图像中的行人、动物和植物等。

[0003] 相关技术中,通常采用基于区域的多标签分类方法来完成图像的分类识别任务,该方法基于选择性搜索方法对待分类图像进行区域提取,确定出待分类图像中的至少一个候选区域,并采用卷积神经网络分别对每个候选区域进行特征提取,得到每个候选区域对应的候选特征,再分别将每个候选特征输入到分类模型中,得到待分类图像对应的多标签分类结果。

[0004] 上述方案中,通过选择性搜索方法可以得到大量的候选区域,由于候选区域之间往往不存在关联关系,从而在对图像进行多标签分类时,是分别对每个候选区域对应的候选特征单独进行分类的,因此得到的分类结果的准确度较低。例如,对于一个包含有“餐桌”和“椅子”的图像进行区域提取后,可以得到一个包含“餐桌”的候选区域和一个包含“椅子”的候选区域,由于包含“椅子”的候选区域中可能具有与“沙发”相似的特征,因此,在对包含“椅子”的候选区域对应的候选特征进行单独分类时,可能会将“椅子”误识成“沙发”,进而得到该图像中包含有“沙发”的结果。

发明内容

[0005] 为解决相关技术中存在的技术问题,本申请实施例提供一种多标签图像分类方法、装置、电子设备及存储介质,可以提高对图像进行多标签分类的准确度。

[0006] 为达到上述目的,本申请实施例的技术方案是这样实现的:

[0007] 一方面,本申请实施例提供了一种多标签图像分类方法,包括:

[0008] 对待分类图像进行特征提取,获得相应全局特征;所述全局特征包含所述待分类图像中的各像素点的属性特征;

[0009] 识别所述待分类图像中的至少一个候选对象各自对应的边界框,并基于获得的至少一个边界框,对所述全局特征进行划分,获得所述至少一个边界框各自对应的局部特征;

[0010] 基于所述各像素点的属性特征,以及各属性特征之间的第一关联度,获得相应全局注意力特征;

[0011] 基于所述全局注意力特征和至少一个局部特征,以及所述全局注意力特征与所述至少一个局部特征之间的第二关联度,获得相应的目标全局特征和至少一个目标局部特征;

[0012] 基于所述目标全局特征和所述至少一个目标局部特征,获得所述待分类图像对应的目标分类标签。

[0013] 一方面,本申请实施例还提供了一种多标签图像分类装置,包括:

[0014] 全局特征提取单元,用于对待分类图像进行特征提取,获得相应全局特征;所述全局特征包含所述待分类图像中的各像素点的属性特征;

[0015] 局部特征确定单元,用于识别所述待分类图像中的至少一个候选对象各自对应的边界框,并基于获得的至少一个边界框,对所述全局特征进行划分,获得所述至少一个边界框各自对应的局部特征;

[0016] 全局注意力处理单元,用于基于所述各像素点的属性特征,以及各属性特征之间的第一关联度,获得相应全局注意力特征;

[0017] 目标特征确定单元,用于基于所述全局注意力特征和至少一个局部特征,以及所述全局注意力特征与所述至少一个局部特征之间的第二关联度,获得相应的目标全局特征和至少一个目标局部特征;

[0018] 多标签分类单元,基于所述目标全局特征和所述至少一个目标局部特征,获得所述待分类图像对应的目标分类标签。

[0019] 可选的,所述全局注意力处理单元,具体用于:

[0020] 针对所述各像素点,分别执行以下操作:将一个像素点作为查询像素点,所述各像素点中的其他像素点作为关键像素点;基于自注意力机制,将所述查询像素点的属性特征,分别与各关键像素点各自的属性特征进行关联度匹配,获得各属性特征之间的第一关联度;

[0021] 根据各第一关联度,对所述各像素点的属性特征进行加权合并,获得相应全局注意力特征。

[0022] 可选的,所述目标特征确定单元,具体用于:

[0023] 针对至少一个局部特征,分别执行以下操作:基于自注意力机制,将一个局部特征中的各像素点的属性特征,分别与所述全局注意力特征中的各像素点的注意力特征进行关联度匹配,获得所述一个局部特征与所述全局注意力特征之间的第二关联度;

[0024] 根据各第二关联度,对所述全局注意力特征和所述至少一个局部特征分别进行加权合并,获得相应的目标全局特征和至少一个目标局部特征。

[0025] 可选的,所述多标签分类单元,具体用于:

[0026] 基于多标签分类模型,分别确定出所述目标全局特征对应的全局分类标签结果和所述至少一个目标局部特征各自对应的局部分类标签结果;

[0027] 根据所述全局分类标签结果和各局部分类标签结果,获得所述待分类图像对应的目标分类标签。

[0028] 可选的,所述多标签分类单元,还用于:

[0029] 将所述目标全局特征输入多标签分类模型,获得所述目标全局特征分别属于各个候选分类的第一概率值,并将各第一概率值作为全局分类标签结果;

[0030] 将所述至少一个目标局部特征,分别输入所述多标签分类模型,获得所述至少一个目标局部特征各自属于各个候选分类的第二概率值,并将各第二概率值作为局部分类标签结果。

[0031] 可选的,所述多标签分类单元,还用于:

[0032] 针对所述各个候选分类,分别执行以下操作:根据所述至少一个目标局部特征各自属于一个候选分类的第二概率值,将符合设定取值条件的第二概率值,与所述一个候选分类对应的第一概率值进行平均,确定所述一个候选分类对应的目标概率值;

[0033] 根据所述各个候选分类各自对应的目标概率值,将所述目标概率值大于设定阈值的候选分类,作为所述待分类图像对应的目标分类标签。

[0034] 可选的,所述装置还包括模型训练单元,用于:

[0035] 获取训练数据集;所述训练数据集中包括多个图像样本,所述图像样本中标注有设定分类标签;

[0036] 基于所述训练数据集,对所述多标签分类模型进行迭代训练,直到满足设定的收敛条件为止,其中,一次迭代训练过程包括:

[0037] 基于从所述训练数据集中抽取的图像样本,获得所述图像样本对应的目标全局样本特征和至少一个目标局部样本特征;

[0038] 通过所述多标签分类模型,根据所述目标全局样本特征和所述至少一个目标局部样本特征,确定所述图像样本对应的目标样本分类标签,并基于所述目标样本分类标签与所述设定分类标签确定的损失值,对所述多标签分类模型进行参数调整。

[0039] 可选的,所述模型训练单元,还用于:

[0040] 对所述图像样本进行特征提取,获得相应全局样本特征,并基于对所述图像样本中的至少一个候选样本对象进行识别获得的至少一个边界框,对所述全局样本特征进行划分,获得所述至少一个样本边界框各自对应的局部样本特征;

[0041] 根据所述全局样本特征和至少一个局部样本特征,获得相应的目标全局样本特征和至少一个目标局部样本特征。

[0042] 可选的,所述模型训练单元,还用于:

[0043] 通过所述多标签分类模型,确定所述目标全局样本特征分别属于各个候选样本分类的第一样本概率值,以及所述至少一个目标局部样本特征各自属于各个候选样本分类的第二样本概率值;

[0044] 根据所述各个候选样本分类对应的第一样本概率值和第二样本概率值,确定所述图像样本对应的目标样本分类标签。

[0045] 一方面,本申请实施例提供了一种计算机设备,包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,所述处理器执行所述程序时实现上述多标签图像分类方法的步骤。

[0046] 一方面,本申请实施例提供了一种计算机可读存储介质,其存储有可由计算机设备执行的计算机程序,当所述程序在计算机设备上运行时,使得所述计算机设备执行上述多标签图像分类方法的步骤。

[0047] 一方面,本申请实施例提供了一种计算机程序产品,所述计算机程序产品包括存储在计算机可读存储介质上的计算机程序,所述计算机程序包括程序指令,当所述程序指令被计算机执行时,使所述计算机执行上述多标签图像分类方法的步骤。

[0048] 本申请有益效果如下:

[0049] 本申请实施例提供的多标签图像分类方法、装置、电子设备及存储介质,对待分类

图像进行特征提取,获得相应全局特征,并基于对待分类图像中的至少一个候选对象进行识别获得的至少一个边界框,对全局特征进行划分,获得至少一个边界框各自对应的局部特征,基于各像素点的属性特征,以及各属性特征之间的第一关联度,获得相应全局注意力特征,并基于全局注意力特征和至少一个局部特征,以及全局注意力特征与至少一个局部特征之间的第二关联度,获得相应的目标全局特征和至少一个目标局部特征,根据目标全局特征和至少一个目标局部特征,获得待分类图像对应的目标分类标签。由于在得到图像的全局特征和局部特征后,通过对全局特征和局部特征进行自注意力处理,使得图像的全局与局部之间具有关联关系,从而在基于得到的目标全局特征和目标局部特征,对待分类图像进行分类时,可以得到更加准确的分类标签结果,提高对图像进行多标签分类的准确度。

[0050] 本申请的其它特征和优点将在随后的说明书中阐述,并且,部分地从说明书中变得显而易见,或者通过实施本申请而了解。本申请的目的和其他优点可通过在所写的说明书、权利要求书、以及附图中所特别指出的结构来实现和获得。

附图说明

[0051] 为了更清楚地说明本发明实施例中的技术方案,下面将对实施例描述中所需要使用的附图作简要介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域的普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

- [0052] 图1为相关技术中采用选择性搜索方法产生的候选区域示意图;
- [0053] 图2为本申请实施例提供的一种多标签图像分类方法的应用场景图;
- [0054] 图3为本申请实施例提供的一种多标签图像分类方法的流程示意图;
- [0055] 图4为本申请实施例提供的一种获得全局注意力特征的示意图;
- [0056] 图5为本申请实施例提供的一种获得目标全局特征和目标局部特征的示意图;
- [0057] 图6为本申请实施例提供的一种待分类图像的示意图;
- [0058] 图7为本申请实施例提供的一种多标签分类模型的训练流程示意图;
- [0059] 图8为本申请实施例提供的另一种多标签图像分类方法的流程示意图;
- [0060] 图9为本申请实施例提供的一种ResNet101模型的网络结构图;
- [0061] 图10为本申请实施例提供的一种RPN提取局部区域的过程示意图;
- [0062] 图11为本申请实施例提供的一种多标签分类模型的网络结构图;
- [0063] 图12a为本申请实施例提供的一种t-SNE可视化结果的示意图;
- [0064] 图12b为本申请实施例提供的一种多标签图像分类方法的具体场景示意图;
- [0065] 图13为本申请实施例提供的一种多标签图像分类装置的结构示意图;
- [0066] 图14为本申请实施例提供的另一种多标签图像分类装置的结构示意图;
- [0067] 图15为本申请实施例提供的一种电子设备的结构示意图;
- [0068] 图16为本申请实施例提供的另一种电子设备的结构示意图。

具体实施方式

[0069] 为使本申请实施例的目的、技术方案和优点更加清楚,下面将结合本申请实施例

中的附图,对本申请的技术方案进行清楚、完整地描述,显然,所描述的实施例是本申请技术方案的一部分实施例,而不是全部的实施例。基于本申请文件中记载的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本申请技术方案保护的范围。

[0070] 本申请的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象,而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换,以便这里描述的本发明的实施例能够在除了这里图示或描述的那些以外的顺序实施。

[0071] 以下对本申请实施例中的部分用语进行解释说明,以便于本领域技术人员理解。

[0072] 多标签图像分类:对图像进行分类识别,从而准确地将图像归入到一个或多个标签中。在多标签分类中,图像可以被分配到的类别数目没有约束,一个图像的标签数量不止一个,即一个图像可以对应多个标签。

[0073] 局部区域提议(Region Proposal):给定输入图像查找可以定位对象的所有可能位置,输出是对象的可能位置的边界框列表,这些边界框中包括的区域通常被称为感兴趣候选区域(Region of Interest,ROI)。

[0074] 注意力机制(Attention):源于对人类视觉的研究。在认知科学中,由于信息处理的瓶颈,人类会选择性地关注所有信息的一部分,同时忽略其他可见的信息。

[0075] 自注意力机制(Self-Attention)是注意力机制的变体,其减少了对外部信息的依赖,更擅长捕捉数据或特征的内部相关性。

[0076] 下文中所用的词语“示例性”的意思为“用作例子、实施例或说明性”。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。

[0077] 文中的术语“第一”、“第二”仅用于描述目的,而不能理解为明示或暗示相对重要性或者隐含指明所指示的技术特征的数量。由此,限定有“第一”、“第二”的特征可以明示或者隐含地包括一个或者更多个该特征,在本申请实施例的描述中,除非另有说明,“多个”的含义是两个或两个以上。

[0078] 本申请实施例涉及人工智能(Artificial Intelligence, AI)和机器学习(Machine Learning, ML)技术和自然语言处理(Nature Language processing, NLP),基于人工智能中的机器学习技术和自然语言处理技术而设计。

[0079] 人工智能(Artificial Intelligence, AI)是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能,感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。换句话说,人工智能是计算机科学的一个综合技术,它企图了解智能的实质,并生产出一种新的能以人类智能相似的方式做出反应的智能机器。人工智能也就是研究各种智能机器的设计原理与实现方法,使机器具有感知、推理与决策的功能。

[0080] 人工智能技术是一门综合学科,涉及领域广泛,既有硬件层面的技术也有软件层面的技术。人工智能基础技术一般包括如传感器、专用人工智能芯片、云计算、分布式存储、大数据处理技术、操作/交互系统、机电一体化等技术。人工智能软件技术主要包括计算机视觉技术、语音处理技术、自然语言处理技术以及机器学习/深度学习、自动驾驶、智慧交通等几大方向。

[0081] 机器学习(Machine Learning, ML)是一门多领域交叉学科,涉及概率论、统计学、

逼近论、凸分析、算法复杂度理论等多门学科。专门研究计算机怎样模拟或实现人类的学习行为,以获取新的知识或技能,重新组织已有的知识结构使之不断改善自身的性能。机器学习是人工智能的核心,是使计算机具有智能的根本途径,其应用遍及人工智能的各个领域。机器学习和深度学习通常包括人工神经网络、置信网络、强化学习、迁移学习、归纳学习、式教学习等技术。

[0082] 本申请实施例采用基于机器学习的多标签分类模型,根据待分类图像对应的目标全局特征和至少一个目标局部特征,获得待分类图像对应的目标分类标签。

[0083] 下面对本申请实施例的设计思想进行简要介绍:

[0084] 多标签图像分类主要用于对图像进行分类识别,从而将图像划归到一个或多个标签中。相关技术中,通常采用基于区域的多标签分类方法来完成图像的分类识别任务,该方法首先基于选择性搜索方法对待分类图像进行区域提取,产生至少一个候选区域,然后分别对每个候选区域进行特征提取,得到每个候选区域对应的候选特征,最后分别对每个候选特征进行分类,得到待分类图像对应的多标签分类结果。

[0085] 然而,上述方案既没有考虑到候选区域之间的关联关系,也没有考虑到局部和全局之间的关联关系,不能很好的利用整体与局部的信息,从而导致分类结果较差。

[0086] 上述方案具体存在以下缺陷:首先,为了达到较高的召回率,局部区域提取会产生如图1所示的大量的候选区域,通常是数以千计。这不仅对多标签学习来说是低效的,而且由于背景干扰和建议的不准确边界,也会降低性能。其次,许多类别之间的语义依赖关系被忽略了,这对多标签分类而言特别重要(例如,“猫”更有可能被误判为“狗”类别,而不是错误地与“雨伞”相关联,“牙刷”更有可能和“牙膏”一同出现,而不是和“飞机”相关联)。在一些先前的工作中,曾试图通过在基于CNN的模型之后附加一个RNN或LSTM结构来明确地捕捉类别依赖性来解决这个缺点。然而,这些模型只考虑了局部与局部之间的关联,并没有考虑到局部与全局的高阶相关性。因此,缺乏对图像全局信息的透彻理解,多标签信息不能有效地用于学习语义区域。此外,这些模型需要以复杂的迭代方式来协助学习过程,这对于模型的训练而言是低效的。

[0087] 因此,上述方案在对图像进行多标签分类时,是分别对每个候选区域对应的候选特征单独进行分类的,导致最终得到的分类结果的准确度较低。

[0088] 有鉴于此,本申请实施例提供一种多标签图像分类方法、装置、电子设备及存储介质,对待分类图像进行特征提取,获得相应的全局特征,并基于对待分类图像中的至少一个候选对象进行识别获得的至少一个边界框,对全局特征进行划分,获得至少一个边界框各自对应的局部特征,基于各像素点的属性特征,以及各属性特征之间的第一关联度,获得相应的全局注意力特征,并基于全局注意力特征和至少一个局部特征,以及全局注意力特征与至少一个局部特征之间的第二关联度,获得相应的目标全局特征和至少一个目标局部特征,根据目标全局特征和至少一个目标局部特征,获得待分类图像对应的目标分类标签。从而可以提高对图像进行多标签分类的准确度。

[0089] 以下结合说明书附图对本申请的优选实施例进行说明,应当理解,此处所描述的优选实施例仅用于说明和解释本申请,并不用于限定本申请,并且在不冲突的情况下,本申请实施例及实施例中的特征可以相互组合。

[0090] 参阅图2所示,为本申请实施例中应用场景示意图。该应用场景中至少包括终端设

备110和服务器130,可通过终端设备110登录应用操作界面120。终端设备110的数量可以是一个或多个,服务器130的数量也可以是一个或多个,本申请对终端设备110和服务器130的数量不做具体限定。终端设备110与服务器130之间可以通过通信网络进行通信。

[0091] 在本申请实施例中,终端设备110可以是便携设备(例如:手机、平板电脑、笔记本电脑等),也可以是计算机、智能屏或个人电脑(Personal Computer,PC)等。终端设备110包括但不限于手机、电脑、智能语音交互设备、智能家电、车载终端、飞行器等。

[0092] 服务器130可以是独立的物理服务器,也可以是多个物理服务器构成的服务器集群或者分布式系统,还可以是提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、内容分发网络(Content Delivery Network,CDN)、以及大数据和人工智能平台等基础云计算服务的云服务器。终端设备110与服务器130可以通过有线或无线通信方式进行直接或间接地连接,本申请在此不做限制。

[0093] 本申请实施例中的多标签图像分类方法可以由终端设备110执行,也可以由服务器130执行,还可以由终端设备110与服务器130交互执行。

[0094] 以服务器130执行本申请实施例中的多标签图像分类方法举例来说,包括以下步骤:

[0095] 目标对象通过终端设备110将待分类图像发送给服务器130,服务器130可以对待分类图像进行特征提取,获得相应的全局特征,并基于对待分类图像中的至少一个候选对象进行识别获得的至少一个边界框,对全局特征进行划分,获得至少一个边界框各自对应的局部特征,基于各像素点的属性特征,以及各属性特征之间的第一关联度,获得相应的全局注意力特征,并基于全局注意力特征和至少一个局部特征,以及全局注意力特征与至少一个局部特征之间的第二关联度,获得相应的目标全局特征和至少一个目标局部特征,最后根据目标全局特征和至少一个目标局部特征,获得待分类图像对应的目标分类标签。服务器在得到待分类图像对应的目标分类标签后,可以将目标分类标签发送给终端设备110,以使终端设备110向目标对象展示待分类图像对应的目标分类标签。

[0096] 应当说明的是,图2是对本申请的多标签图像分类方法的应用场景进行示例介绍,实际本申请实施例中的方法可以适用的应用场景并不限于此。并且,本申请实施例可应用于各种场景,包括但不限于云技术、人工智能、智慧交通、辅助驾驶等。

[0097] 为进一步说明本申请实施例提供的技术方案,下面结合附图以及具体实施方式对此进行详细的说明。虽然本申请实施例提供了如下述实施例或附图所示的方法操作步骤,但基于常规或者无需创造性的劳动在所述方法中可以包括更多或者更少的操作步骤。在逻辑上不存在必要因果关系的步骤中,这些步骤的执行顺序不限于本申请实施例提供的执行顺序。所述方法在实际的处理过程中或者装置执行时,可按照实施例或者附图所示的方法顺序执行或者并行执行。

[0098] 图3示出了本申请实施例提供的一种多标签图像分类方法的流程图,该方法可以由电子设备执行,该电子设备可以是图2中的终端设备110和/或服务器130。如图3所示,包括以下步骤:

[0099] 步骤S301,对待分类图像进行特征提取,获得相应的全局特征。

[0100] 其中,全局特征包含待分类图像中的各像素点的属性特征。

[0101] 在获取到待分类图像后,可以采用特征提取模型对待分类图像进行特征提取,获

得相应全局特征。

[0102] 步骤S302, 识别待分类图像中的至少一个候选对象各自对应的边界框, 并基于获得的至少一个边界框, 对全局特征进行划分, 获得至少一个边界框各自对应的局部特征。

[0103] 在获得待分类图像对应全局特征后, 可以对该全局特征进行局部区域提议操作, 以在待分类图像中识别出至少一个候选对象各自对应的边界框, 并基于获得的至少一个边界框, 将全局特征划分成至少一个局部特征, 其中, 每个局部特征对应一个边界框。

[0104] 步骤S303, 基于各像素点的属性特征, 以及各属性特征之间的第一关联度, 获得相应全局注意力特征。

[0105] 在获得待分类图像对应全局特征后, 对于全局特征中的各像素点, 如图4所示, 可以分别执行以下操作:

[0106] 将全局特征中的一个像素点作为查询像素点, 全局特征中的除该查询像素点以外的其他像素点作为关键像素点, 基于自注意力机制, 将查询像素点的属性特征, 分别与各关键像素点各自的属性特征进行关联度匹配, 获得各属性特征之间的第一关联度; 根据各第一关联度, 对各像素点的属性特征进行加权合并, 获得相应全局注意力特征。

[0107] 通过自注意力机制, 确定出全局特征中的各像素点的属性特征之间的关联关系, 并根据该关联关系, 对各像素点的属性特征进行加权合并, 获得全局注意力特征, 从而通过自注意力处理, 可以使得待分类图像中的各个候选对象对应的特征在全局特征中更加突出, 即使得在待分类图像的全局特征中, 更加关注于各个候选对象所在区域的特征信息, 进而提升全局特征的区分度。

[0108] 步骤S304, 基于全局注意力特征和至少一个局部特征, 以及全局注意力特征与至少一个局部特征之间的第二关联度, 获得相应的目标全局特征和至少一个目标局部特征。

[0109] 在通过对全局特征进行划分获得至少一个局部特征, 以及对全局特征进行自注意力机制获得全局注意力特征之后, 对于获得的至少一个局部特征, 如图5所示, 可以分别执行以下操作:

[0110] 基于自注意力机制, 将一个局部特征中的各像素点的属性特征, 分别与全局注意力特征中的各像素点的注意力特征进行关联度匹配, 获得一个局部特征与全局注意力特征之间的第二关联度, 根据各第二关联度, 对全局注意力特征和至少一个局部特征分别进行加权合并, 获得相应的目标全局特征和至少一个目标局部特征。

[0111] 通过自注意力机制, 对获得的全局注意力特征和至少一个局部特征进行处理, 可以将更多的信息从图像的全局传递到图像的各个局部, 有效地探索图像的全局和局部之间的互补信息, 获得全局与局部之间的关联关系, 从而使得通过自注意力处理后得到的目标全局特征中可以包含有图像的局部信息, 得到的目标局部特征中可以包含有图像的全局信息。同时待分类图像中包括的各个候选对象之间的相关关系也可以被包含于得到的目标全局特征和目标局部特征中, 从而可以以隐式的方式建立起标签依赖。

[0112] 步骤S305, 基于目标全局特征和至少一个目标局部特征, 获得待分类图像对应的目标分类标签。

[0113] 将目标全局特征输入多标签分类模型, 获得目标全局特征分别属于各个候选分类的第一概率值, 并将各第一概率值作为目标全局特征对应全局分类标签结果。

[0114] 将至少一个目标局部特征, 分别输入多标签分类模型, 获得至少一个目标局部特

征各自属于各个候选分类的第二概率值,并将各第二概率值作为至少一个目标局部特征各自对应的局部分类标签结果。

[0115] 根据全局分类标签结果和各局部分类标签结果,获得待分类图像对应的目标分类标签。具体地,针对各个候选分类,分别执行以下操作:根据至少一个目标局部特征各自属于一个候选分类的第二概率值,将符合设定取值条件的第二概率值,与一个候选分类对应的第一概率值进行平均,确定一个候选分类对应的目标概率值。

[0116] 在确定出各个候选分类各自对应的目标概率值后,可以根据各个候选分类各自对应的目标概率值,将目标概率值大于设定阈值的候选分类,作为待分类图像对应的目标分类标签。

[0117] 例如,待分类图像可以如图6所示,从图6中可以看出,该待分类图像中包含有一只蜜蜂和两朵花,则该待分类图像可以对应3个目标局部特征,其中,目标局部特征1中包含有蜜蜂的特征,目标局部特征2中包含有两朵花中的一朵花的特征,目标局部特征3中包含有两朵花中的另一朵花的特征。

[0118] 假设候选分类共有4个,分别为花、草、蜜蜂、鸟,将目标全局特征输入多标签分类模型,可以得到目标全局特征属于花的第一概率值为0.9,目标全局特征属于草的第一概率值为0.1,目标全局特征属于蜜蜂的第一概率值为0.88,目标全局特征属于鸟的第一概率值为0.12;将目标局部特征1输入多标签分类模型,可以得到目标局部特征1属于花的第二概率值为0.01,目标局部特征1属于草的第二概率值为0.01,目标局部特征1属于蜜蜂的第二概率值为0.96,目标局部特征1属于鸟的第二概率值为0.02;将目标局部特征2输入多标签分类模型,可以得到目标局部特征2属于花的第二概率值为0.94,目标局部特征2属于草的第二概率值为0.04,目标局部特征2属于蜜蜂的第二概率值为0.01,目标局部特征2属于鸟的第二概率值为0.01;将目标局部特征3输入多标签分类模型,可以得到目标局部特征3属于花的第二概率值为0.96,目标局部特征3属于草的第二概率值为0.02,目标局部特征3属于蜜蜂的第二概率值为0.01,目标局部特征3属于鸟的第二概率值为0.01。

[0119] 首先,对各个目标局部特征各自属于各个候选分类的第二概率值取最大值,即确定各个目标局部特征各自属于各个候选分类的最大第二概率值,可以得到各个目标局部特征各自属于花的最大第二概率值为0.96,各个目标局部特征各自属于草的最大第二概率值为0.04,各个目标局部特征各自属于蜜蜂的最大第二概率值为0.96,各个目标局部特征各自属于鸟的最大第二概率值为0.02。

[0120] 然后,将各个目标局部特征各自属于各个候选分类的最大第二概率值,分别与目标全局特征分别属于各个候选分类的第一概率值进行平均,可以得到待分类图像属于花的目标概率值为0.93,待分类图像属于草的目标概率值为0.07,待分类图像属于蜜蜂的目标概率值为0.92,待分类图像属于鸟的目标概率值为0.07。假设设定阈值为0.9,则可以确定待分类图像对应的目标分类标签为蜜蜂和花。

[0121] 上述步骤S305中所使用的多标签分类模型的训练过程可以如图7所示,下面结合图7,对该训练过程进行详细阐述。

[0122] 步骤S701,获取训练数据集。

[0123] 获取到的训练数据集中可以包括多个图像样本,每个图像样本中标注有设定分类标签。

[0124] 步骤S702,从训练数据集中抽取图像样本,并对抽取的图像样本进行特征提取,获得相应全局样本特征。

[0125] 在对多标签分类模型进行训练时,可以从训练数据集中抽取图像样本,并对抽取的图像样本进行特征提取,得到全局样本特征。

[0126] 步骤S703,基于对图像样本中的至少一个候选样本对象进行识别获得的至少一个边界框,对全局样本特征进行划分,获得至少一个样本边界框各自对应的局部样本特征。

[0127] 提供对全局样本特征中的至少一个候选样本对象进行边界识别,获得至少一个候选样本对象各自对应的边界框,并根据至少一个边界框,对全局样本特征进行划分,得到至少一个局部样本特征。

[0128] 步骤S704,根据全局样本特征和至少一个局部样本特征,获得相应的目标全局样本特征和至少一个目标局部样本特征。

[0129] 在获得抽取的图像样本对应的全局样本特征和至少一个局部样本特征后,可以先对全局样本特征执行自注意力机制,获得相应全局样本注意力特征,然后对全局样本注意力特征和至少一个局部样本特征执行自注意力机制,获得相应的目标全局样本特征和至少一个目标局部样本特征。

[0130] 步骤S705,将目标全局样本特征和至少一个目标局部样本特征,分别输入到多标签分类模型中,确定目标全局样本特征分别属于各个候选样本分类的第一样本概率值,以及至少一个目标局部样本特征各自属于各个候选样本分类的第二样本概率值。

[0131] 将目标全局样本特征输入到多标签分类模型中,可以得到目标全局样本特征分别属于各个候选样本分类的第一样本概率值。

[0132] 将至少一个目标局部样本特征分别输入到多标签分类模型中,可以得到至少一个目标局部样本特征各自属于各个候选样本分类的第二样本概率值。

[0133] 步骤S706,根据各个候选样本分类对应的第一样本概率值和第二样本概率值,确定图像样本对应的目标样本分类标签。

[0134] 在确定出目标全局样本特征分别属于各个候选样本分类的第一样本概率值,和至少一个目标局部样本特征各自属于各个候选样本分类的第二样本概率值后,针对各个候选样本分类,分别执行以下操作:根据至少一个目标局部样本特征各自属于一个候选样本分类的第二样本概率值,确定出至少一个目标局部样本特征各自属于一个候选样本分类的最大第二样本概率值,并将该候选样本分类的最大样本概率值,与该候选样本分类对应的第一样本概率值进行平均,确定该候选样本分类对应的目标样本概率值。

[0135] 在确定出各个候选样本分类各自对应的目标样本概率值后,可以将各个目标样本概率值中大于设定样本阈值的目标样本概率值对应的候选样本分类,作为图像样本对应的目标样本分类标签。

[0136] 步骤S707,基于目标样本分类标签与设定分类标签,确定相应的损失值。

[0137] 在计算损失值时,可以采用二进制交叉熵(Binary CrossEntropy, BCE)损失函数来计算损失值。具体地, BCE损失函数可以通过以下公式进行表示:

$$[0138] \quad \mathcal{L} = - \sum_i (y'_i \log(y_i) + (1 - y'_i) \log(1 - y_i))$$

[0139] 通常,损失值是判定实际的输出与期望的输出的接近程度。损失值越小,说明实际的输出越接近期望的输出。

[0140] 步骤S708,确定损失值是否收敛至预设的目标值;如果否,执行步骤S709;如果是,执行步骤S710。

[0141] 判断损失值是否收敛至预设的目标值,如果损失值小于或等于预设的目标值,或者,连续N次训练得到的损失值的变化幅度小于或等于预设的目标值时,认为损失值已收敛至预设的目标值,说明损失值收敛;否则,说明损失值尚未收敛。

[0142] 步骤S709,根据确定的损失值对多标签分类模型的参数进行调整。

[0143] 如果损失值未收敛,则对模型参数进行调整,调整模型参数后,返回执行步骤S702,继续下一轮的训练过程。

[0144] 步骤S710,结束训练得到已训练的多标签分类模型。

[0145] 如果损失值收敛,则将当前得到的多标签分类模型作为已训练的多标签分类模型。

[0146] 在一些实施例中,本申请提出的多标签图像分类方法还可以按照图8示出的过程进行实现,该过程可以由电子设备执行,该电子设备可以是图2中的终端设备110和/或服务器130。如图8所示,包括如下步骤:

[0147] 步骤S801,对待分类图像进行特征提取,获得相应全局特征。

[0148] 在获取到待分类图像后,可以将待分类图像输入到如残差网络101 (ResNet101)模型的特征提取模型中,进行待分类图像的高层语义特征提取,获得相应全局特征。

[0149] 当特征提取模型为ResNet101模型时,ResNet101模型的网络结构可以如图9所示。在图9中,输入ResNet101模型的特征需要先经过卷积层和全局最大池化层、再陆续经过三种不同的三个残差模块。

[0150] 需要注意的是,为了提升特征的通用性和泛化性能,在采用ResNet101模型对待分类图像进行特征提取,获得全局特征时,只选择Resnet101的前三个阶段作为特征提取模块,而第四个阶段可以被应用在对待分类图像进行多标签分类过程中,以帮助提取得到的特征能够更好的进行投影和学习语义信息。

[0151] 可选的,本申请实施例中的特征提取模型也可以为AlexNet、视觉几何组网络 (Visual Geometry Group Net, VGGNet) 和ResNet等。

[0152] 步骤S802,通过对待分类图像进行区域提取,得到的至少一个边界框对全局特征进行划分,获得至少一个局部特征。

[0153] 将待分类图像输入局部区域提议网络中,在待分类图像中提取出可能含有关键目标的边界框。本实施例中所使用的局部区域提议网络可以是预训练的区域提议网络 (Region Proposal Network) RPN。

[0154] 卷积特征图中的每个像素点都配备有k种anchors作为初始的检测框,进而去判断anchor是属于物体还是背景(即判断该anchor到底有没有覆盖目标),以及为属于物体的anchor进行第一次坐标修正。物体背景是二分类,所以分类分支可以得到 $2k$ 个scores;坐标修正是四个值(x,y,w,h),所以回归分支可以得到 $4k$ 个坐标值的属性。

[0155] 具体地,如图10所示为RPN进行局部区域提取的过程示意图,在图10中,用一个 3×3 的滑动窗口,遍历整个卷积特征图,在遍历的过程中,每个窗口中心按长宽比1:1、1:2、2:1

和面积 128×128 、 256×256 、 512×512 ,生成9种目标框(anchor),然后利用窗口分类层(cls层)对每个anchor做2分类(区分出时前景还是背景),利用位置精修层(reg层)确定每个anchor的坐标位置,最终可以得到结果 $2k$ 个分数和 $4k$ 个坐标。

[0156] 在提取出待分类图像中的至少一个边界框后,基于至少一个边界框可以对全局特征进行划分,获得至少一个局部特征。

[0157] 可选的,本申请实施例中所采用的局部区域提议网络可以是任意能够生成相关目标边界框的方法,如传统的二值化梯度幅值(Binarized Normed Gradients,BING)、EdgeBox,也可以是深度学习方法,如弱监督区域提议方法。

[0158] 步骤S803,对全局特征执行自注意力机制,获得相应全局注意力特征。

[0159] 由于全局特征和局部特征两个分支之间的信息传播常常被以前的工作所忽略,因此在本申请中提出了跨粒度注意力模块,用来有效的改进这一点。因为全局特征图和局部特征图在空间维度上是错位的,简单的融合并不能带来性能的提升。因此,可以借助于自注意力机制来实现局部到局部的交互,乃至全局到局部的高阶相关性。

[0160] 首先,在全局特征分支上执行自注意力机制(Self-attention),以捕捉非局部的依赖性,从而产生更高级的语义特征。在全局特征分支的自注意力机制中,可以令 $F \in \mathbb{R}^{C \times (H \times W)}$ 为全局特征,其中 H 、 W 和 C 分别为全局特征的高度、重量和通道数目。然后, $Q_g = \phi_g^q(F)$, $K_g = \phi_g^k(F)$, $V_g = \phi_g^v(F)$,其中 $\phi(\cdot)$ 表示一个线性投影,通过可学习的权重,可以将输入的全局特征 F 映射到同一维度的输出。

[0161] 因此,全局注意力图可以通过以下公式得到:

$$[0162] \quad A_g = \text{softmax} \left(\frac{Q_g^T K_g}{\sqrt{C}} \right), \quad A_g \in \mathbb{R}^{HW \times HW}$$

[0163] 其中,缩放因子 $\frac{1}{\sqrt{C}}$ 是为了避免点积后的结果太大导致溢出。对优化后的特征 $A_g V_g$ 沿空间维度进行最大池化(max-pooling)操作,就可以得到全局注意力特征,即 $F_g \in \mathbb{R}^{C \times 1}$ 。随着更多的信息从全局分支传递到局部分支,局部分支中的特征信息就可以更好地相互关联,隐含地建立相关信息。

[0164] 步骤S804,对全局注意力特征和至少一个局部特征,执行自注意力机制,获得相应的目标全局特征和至少一个目标局部特征。

[0165] 在跨粒度注意力模块的自注意力机制中,可以令 $F_l \in \mathbb{R}^{k_o \times C \times 1}$ 为至少一个局部特征, F_l 是通过一系列的可学习的投影得到的,其中, k_o 为局部特征的数量。通过新增一个额外的维度,可以将全局注意力特征 F_g 和 F_l 进行串联,得到 $F_{gl} \in \mathbb{R}^{(k_o+1) \times (C \times 1)}$ 。通过对串联得到的特征 F_{gl} 执行自注意力机制,可以得到跨粒度注意力特征,以便促进全局特征和局部特征之间的信息传播。与上述全局特征分支的自注意力机制中的公式类似,可以得到 $Q_{gl} = \phi_{gl}^q(F_{gl})$, $K_{gl} = \phi_{gl}^k(F_{gl})$, $V_{gl} = \phi_{gl}^v(F_{gl})$ 。

[0166] 因此,跨粒度注意图可以通过以下公式得到:

[0167] $A_{gl} = \text{softmax}(Q_{gl}K_{gl}^T)$, $A_{gl} \in \mathbb{R}^{(k_o+1) \times (k_o+1)}$

[0168] 从而通过 $A_{g1}V_{g1}$ 就可以得到包含有目标全局特征和至少一个目标局部特征的跨粒度注意力特征。一方面,在局部区域提议网络的帮助下,可以得到局部特征,这些局部特征中包含有全局特征分支所无法提供的更详细的信息。另一方面,局部特征没有全局特征的宏观视野,难以应对特殊的情况(如果可以了解到图像的整体环境,就拥有了一部分先验的知识,例如在卧室中,床出现的概率远高于汽车出现的概率)。

[0169] 总而言之,通过串联不同粒度的特征,将多标签图像分类问题简化到自注意力机制可以处理的范畴,从而实现了预期的目的和效果。通过全局注意力特征 F_g 和局部特征 F_l 的输入,跨粒度注意力模块可以有效地探索全局和局部之间的互补信息。同时,图像中包括的候选对象之间的相关关系也可以在所提取出的特征中被捕获,这意味着能够以隐式的方式建立标签依赖。因此,这两个突出优势保证了整体网络性能的提升。

[0170] 步骤S805,基于多标签分类模型,获得目标全局特征分别属于各个候选分类的第一概率值,以及至少一个目标局部特征各自属于各个候选分类的第二概率值。

[0171] 将目标全局特征和至少一个目标局部特征,分别输入到多标签分类模型中,基于多标签分类模型,分别获得目标全局特征分别属于各个候选分类的第一概率值,以及至少一个目标局部特征各自属于各个候选分类的第二概率值。

[0172] 其中,该多标签分类模型的网络结构可以如图11所示。输入多标签分类模型中的特征需要经过三个残差模块,再经过全局最大池化层和全连接层。图11的全连接层中的1000为类别示意,在不同的数据集中可以改成不同的类别数目,例如,在Microsoft COCO数据集中,分类类别数目可以为80,而在Pascal VOC 2007数据集中,分类类别数目可以为20。

[0173] 可选的,由于最大池化层和平均池化层具有较为相似的作用和效果,因此也可以将模型中的最大池化层替换为平均池化层。

[0174] 可选的,由于 1×1 的卷积层和全连接层本质上具有完全相同的作用和效果,因此,也可以将模型中的全连接层替换为 1×1 的卷积层。

[0175] 步骤S806,根据各个候选分类对应的第一概率值和第二概率值,确定各个候选分类对应的目标概率值,并基于目标概率值,确定待分类图像对应的目标分类标签。

[0176] 在得到目标全局特征分别属于各个候选分类的第一概率值,以及至少一个目标局部特征各自属于各个候选分类的第二概率值后,首先从至少一个目标局部特征中,选取出各个候选分类对应的最大第二概率值,然后将各个候选分类对应的最大第二概率值,分别与各个候选分类对应的第一概率值进行平均,得到各个候选分类对应的目标概率值。

[0177] 将各个候选分类对应的目标概率值中,对于设定阈值的候选分类,作为待分类图像对应的目标分类标签。

[0178] 在一种实施例中,本申请提出的多标签图像分类方法中的各个实现过程和模型的总体训练过程均在搭载Intel Xeon 8255C CPU和NVIDIA Tesla V100显卡的服务器上实现,采用8张V100显卡分布式并行训练,并进行推断结果的生成。编码采用Python 3.6.8,使用的深度学习框架为Pytorch 1.4.0与torchvision 0.5.0,opencv-python版本4.5.1,numpy版本1.16.1,scikit-learn版本0.23.0。

[0179] 在该实施例中,可以将本申请提出结合跨粒度注意力的多标签图像分类方法,与

相关技术中无跨粒度注意力的多标签图像分类方法进行比较,并采用t-分布随机邻域嵌入(t-Distributed Stochastic Neighbor Embedding,t-SNE)对比较结果进行可视化处理,t-SNE可视化的比较结果可以如图12a所示。在图12a中,每个点代表一个特定标签背景下的一个标签级特征,每种颜色代表一个类别。从图12a中可以看到,结合跨粒度注意力的特征,t-SNE可视化结果显得更加聚拢,具有更强的区分性,而没有跨粒度注意力的特征,显得则比较分散。因此,本申请提出的多标签图像分类方法与相关技术相比,对图像进行多标签分类的准确度较高。

[0180] 本申请提供的多标签图像分类方法,通过对待分类图像进行特征提取,获得相应的全局特征,通过对待分类图像中的至少一个候选对象进行识别获得的至少一个边界框,对全局特征进行划分,获得至少一个边界框各自对应的局部特征,基于各像素点的属性特征,以及各属性特征之间的第一关联度,获得相应的全局注意力特征,并基于全局注意力特征和至少一个局部特征,以及全局注意力特征与至少一个局部特征之间的第二关联度,获得相应的目标全局特征和至少一个目标局部特征,根据目标全局特征和至少一个目标局部特征,获得待分类图像对应的目标分类标签。该方法的整体框架结构明晰,各个模块均有较好的泛化能力,在流行的大规模多标签图像分类数据集上的实验证明了该方法的性能十分优越。该方法中所提出的跨粒度注意力机制模块可以建立起全局和局部之间的关联,使得特征包含对局部信息的关联和对图像全局信息的透彻理解,从而提升了全局特征和局部特征的区分度和质量。同时,该方法中的多标签分类模型的训练过程采用的是端到端的方式,不需要复杂的迭代方式也不需要更多的监督信息来协助学习过程,较为方便部署。

[0181] 如图12b所示,其为本申请实施例中所列举的一种多标签图像分类方法的具体场景示意图。目标对象可以通过终端设备向服务器发送待分类图像,服务器在接收到待分类图像后,可以对待分类图像进行特征提取,得到全局特征,并通过局部区域提取网络对待分类图像进行处理得到的两个边界框,对全局特征进行划分,可以得到局部特征1和局部特征2。通过将全局特征执行自注意力机制后得到的全局注意力特征,与局部特征1和局部特征2执行自注意力机制,得到目标全局特征、目标局部特征1和目标局部特征2。

[0182] 假设候选分类包括鸟、动物、木头、植物,根据目标全局特征,可以得到待分类图像属于鸟的概率为0.86、属于动物的概率为0.88、属于木头的概率为0.82、属于植物的概率为0.32;根据目标局部特征1,可以得到待分类图像属于鸟的概率为0.98、属于动物的概率为0.92、属于木头的概率为0.12、属于植物的概率为0.08;根据目标局部特征2,可以得到待分类图像属于鸟的概率为0.16、属于动物的概率为0.96、属于木头的概率为0.96、属于植物的概率为0.04。根据待分类图像属于各种候选分类的概率,最终可以确定待分类图像对应的目标分类标签为鸟、动物、木头。

[0183] 与图3所示的多标签图像分类方法基于同一发明构思,本申请实施例中还提供了一种多标签图像分类装置,该多标签图像分类装置可以布设在服务器或终端设备中。由于该装置是本申请多标签图像分类方法对应的装置,并且该装置解决问题的原理与该方法相似,因此该装置的实施可以参见上述方法的实施,重复之处不再赘述。

[0184] 图13示出了本申请实施例提供的一种多标签图像分类装置的结构示意图,如图13所示,该多标签图像分类装置包括全局特征提取单元1301、局部特征确定单元1302、全局注意力处理单元1303、目标特征确定单元1304和多标签分类单元1305。

[0185] 其中,全局特征提取单元1301,用于对待分类图像进行特征提取,获得相应全局特征;全局特征包含待分类图像中的各像素点的属性特征;

[0186] 局部特征确定单元1302,用于识别待分类图像中的至少一个候选对象各自对应的边界框,并基于获得的至少一个边界框,对全局特征进行划分,获得至少一个边界框各自对应的局部特征;

[0187] 全局注意力处理单元1303,用于基于各像素点的属性特征,以及各属性特征之间的第一关联度,获得相应全局注意力特征;

[0188] 目标特征确定单元1304,用于基于全局注意力特征和至少一个局部特征,以及全局注意力特征与至少一个局部特征之间的第二关联度,获得相应的目标全局特征和至少一个目标局部特征;

[0189] 多标签分类单元1305,基于目标全局特征和至少一个目标局部特征,获得待分类图像对应的目标分类标签。

[0190] 可选的,全局注意力处理单元1303,具体用于:

[0191] 针对各像素点,分别执行以下操作:将一个像素点作为查询像素点,各像素点中的其他像素点作为关键像素点;基于自注意力机制,将查询像素点的属性特征,分别与各关键像素点各自的属性特征进行关联度匹配,获得各属性特征之间的第一关联度;

[0192] 根据各第一关联度,对各像素点的属性特征进行加权合并,获得相应全局注意力特征。

[0193] 可选的,目标特征确定单元1304,具体用于:

[0194] 针对至少一个局部特征,分别执行以下操作:基于自注意力机制,将一个局部特征中的各像素点的属性特征,分别与全局注意力特征中的各像素点的注意力特征进行关联度匹配,获得一个局部特征与全局注意力特征之间的第二关联度;

[0195] 根据各第二关联度,对全局注意力特征和至少一个局部特征分别进行加权合并,获得相应的目标全局特征和至少一个目标局部特征。

[0196] 可选的,多标签分类单元1305,具体用于:

[0197] 基于多标签分类模型,分别确定出目标全局特征对应的全局分类标签结果和至少一个目标局部特征各自对应的局部分类标签结果;

[0198] 根据全局分类标签结果和各局部分类标签结果,获得待分类图像对应的目标分类标签。

[0199] 可选的,多标签分类单元1305,还用于:

[0200] 将目标全局特征输入多标签分类模型,获得目标全局特征分别属于各个候选分类的第一概率值,并将各第一概率值作为全局分类标签结果;

[0201] 将至少一个目标局部特征,分别输入多标签分类模型,获得至少一个目标局部特征各自属于各个候选分类的第二概率值,并将各第二概率值作为局部分类标签结果。

[0202] 可选的,多标签分类单元1305,还用于:

[0203] 针对各个候选分类,分别执行以下操作:根据至少一个目标局部特征各自属于一个候选分类的第二概率值,将符合设定取值条件的第二概率值,与一个候选分类对应的第一概率值进行平均,确定一个候选分类对应的目标概率值;

[0204] 根据各个候选分类各自对应的目标概率值,将目标概率值大于设定阈值的候选分

类,作为待分类图像对应的目标分类标签。

[0205] 可选的,如图14所示,上述装置还可以包括模型训练单元1401,用于:

[0206] 获取训练数据集;训练数据集中包括多个图像样本,图像样本中标注有设定分类标签;

[0207] 基于训练数据集,对多标签分类模型进行迭代训练,直到满足设定的收敛条件为止,其中,一次迭代训练过程包括:

[0208] 基于从训练数据集中抽取的图像样本,获得图像样本对应的目标全局样本特征和至少一个目标局部样本特征;

[0209] 通过多标签分类模型,根据目标全局样本特征和至少一个目标局部样本特征,确定图像样本对应的目标样本分类标签,并基于目标样本分类标签与设定分类标签确定的损失值,对多标签分类模型进行参数调整。

[0210] 可选的,模型训练单元1401,还用于:

[0211] 对图像样本进行特征提取,获得相应全局样本特征,并基于对图像样本中的至少一个候选样本对象进行识别获得的至少一个边界框,对全局样本特征进行划分,获得至少一个样本边界框各自对应的局部样本特征;

[0212] 根据全局样本特征和至少一个局部样本特征,获得相应的目标全局样本特征和至少一个目标局部样本特征。

[0213] 可选的,模型训练单元1401,还用于:

[0214] 通过多标签分类模型,确定目标全局样本特征分别属于各个候选样本分类的第一样本概率值,以及至少一个目标局部样本特征各自属于各个候选样本分类的第二样本概率值;

[0215] 根据各个候选样本分类对应的第一样本概率值和第二样本概率值,确定图像样本对应的目标样本分类标签。

[0216] 与上述方法实施例和装置实施例基于同一发明构思,本申请实施例中还提供了一种电子设备。

[0217] 在一种实施例中,电子设备可以是服务器,如图2所示的服务器130。在该实施例中,电子设备的结构可以如图15所示,包括存储器1501,通讯模块1503以及一个或多个处理器1502。

[0218] 存储器1501,用于存储处理器1502执行的计算机程序。存储器1501可主要包括存储程序区和存储数据区,其中,存储程序区可存储操作系统,以及运行即时通讯功能所需的程序等;存储数据区可存储各种即时通讯信息和操作指令集等。

[0219] 存储器1501可以是易失性存储器(volatile memory),例如随机存取存储器(random-access memory,RAM);存储器1501也可以是非易失性存储器(non-volatile memory),例如只读存储器,快闪存储器(flash memory),硬盘(hard disk drive,HDD)或固态硬盘(solid-state drive,SSD)、或者存储器1501是能够用于携带或存储具有指令或数据结构形式的期望的程序代码并能够由计算机存取的任何其他介质,但不限于此。存储器1501可以是上述存储器的组合。

[0220] 处理器1502,可以包括一个或多个中央处理单元(central processing unit,CPU)或者为数字处理单元等等。处理器1502,用于调用存储器1501中存储的计算机程序时

实现上述多标签图像分类方法。

[0221] 通讯模块1503用于与终端设备和其他电子设备进行通信。如果电子设备为服务器,则服务器可以通过通讯模块1503向终端设备发送的待分类图像对应的目标分类标签。

[0222] 本申请实施例中不限定上述存储器1501、通讯模块1503和处理器1502之间的具体连接介质。本公开实施例在图15中以存储器1501和处理器1502之间通过总线1504连接,总线1504在图15中以粗线表示,其它部件之间的连接方式,仅是进行示意性说明,并不引以为限。总线1504可以分为地址总线、数据总线、控制总线等。为便于表示,图15中仅用一条粗线表示,但并不表示仅有一根总线或一种类型的总线。

[0223] 在另一种实施例中,电子设备可以为手机、平板电脑、POS(Point of Sales,销售终端)、车载电脑、智能穿戴设备、PC机等任意电子设备,该电子设备也可以是图2中所示的终端设备110。

[0224] 图16示出了一种本申请实施例提供的电子设备的结构框图。如图16所示,该电子设备包括:射频(Radio Frequency,RF)电路1610、存储器1620、输入单元1630、显示单元1640、传感器1650、音频电路1660、无线保真(wireless fidelity,WiFi)模块1670、处理器1680等部件。本领域技术人员可以理解,图16中示出的电子设备结构并不构成对电子设备的限定,可以包括比图示更多或更少的部件,或者组合某些部件,或者不同的部件布置。

[0225] 下面结合图16对电子设备的各个构成部件进行具体的介绍:

[0226] RF电路1610可用于收发信息或通话过程中,信号的接收和发送,特别地,将基站的下行信息接收后,给处理器1680处理;另外,将设计上行的数据发送给基站。

[0227] 存储器1620可用于存储软件程序以及模块,如本申请实施例中的多标签图像分类方法以及装置对应的程序指令/模块,处理器1680通过运行存储在存储器1620的软件程序以及模块,从而执行电子设备的各种功能应用以及数据处理,如本申请实施例提供的多标签图像分类方法。存储器1620可主要包括存储程序区和存储数据区,其中,存储程序区可存储操作系统、至少一个应用的应用程序等;存储数据区可存储根据电子设备的使用所创建的数据等。此外,存储器1620可以包括高速随机存取存储器,还可以包括非易失性存储器,例如至少一个磁盘存储器件、闪存器件、或其他易失性固态存储器件。

[0228] 输入单元1630可用于接收目标对象输入的数字或字符信息,以及产生与终端的目标对象设置以及功能控制有关的键信号输入。

[0229] 可选的,输入单元1630可包括触控面板1631以及其他输入设备1632。

[0230] 其中,触控面板1631,也称为触摸屏,可收集目标对象在其上或附近的触摸操作(比如目标对象使用手指、触笔等任何适合的物体或附件在触控面板1631上或在触控面板1631附近的操作),并根据预先设定的程式实现相应的操作,如目标对象点击功能模块的快捷标识的操作等。可选的,触控面板1631可以包括触摸检测装置和触摸控制器两个部分。其中,触摸检测装置检测目标对象的触摸方位,并检测触摸操作带来的信号,将信号发送给触摸控制器;触摸控制器从触摸检测装置上接收触摸信息,并将它转换成触点坐标,再送给处理器1680,并能接收处理器1680发来的命令并加以执行。此外,可以采用电阻式、电容式、红外线以及表面声波等多种类型实现触控面板1631。

[0231] 可选的,其他输入设备1632可以包括但不限于物理键盘、功能键(比如音量控制按键、开关按键等)、轨迹球、鼠标、操作杆中的一种或多种。

[0232] 显示单元1640可用于显示由目标对象输入的信息或展示给目标对象的界面信息以及电子设备的各种菜单。显示单元1640即为终端设备的显示系统,用于呈现界面,如显示桌面、应用的操作界面或直播应用的操作界面等。

[0233] 显示单元1640可以包括显示面板1641。可选的,显示面板1641可以采用液晶显示屏(Liquid Crystal Display,LCD)、有机发光二极管(Organic Light-Emitting Diode,OLED)等形式来配置。

[0234] 进一步的,触控面板1631可覆盖显示面板1641,当触控面板1631检测到在其上或附近的触摸操作后,传送给处理器1680以确定触摸事件的类型,随后处理器1680根据触摸事件的类型在显示面板1641上提供相应的界面输出。

[0235] 虽然在图16中,触控面板1631与显示面板1641是作为两个独立的部件来实现电子设备的输入和输入功能,但是在某些实施例中,可以将触控面板1631与显示面板1641集成而实现终端的输入和输出功能。

[0236] 电子设备还可包括至少一种传感器1650,比如光传感器、运动传感器以及其他传感器。具体地,光传感器可包括环境光传感器及接近传感器,其中,环境光传感器可根据环境光线的明暗来调节显示面板1641的亮度,接近传感器可在电子设备移动到耳边时,关闭显示面板1641的背光。作为运动传感器的一种,加速度计传感器可检测各个方向上(一般为三轴)加速度的大小,静止时可检测出重力的大小及方向,可用于识别电子设备姿态的应用(比如横竖屏切换、相关游戏、磁力计姿态校准)、振动识别相关功能(比如计步器、敲击)等;至于电子设备还可配置的陀螺仪、气压计、湿度计、温度计、红外线传感器等其他传感器,在此不再赘述。

[0237] 音频电路1660、扬声器1661,传声器1662可提供对象与电子设备之间的音频接口。音频电路1660可将接收到的音频数据转换后的电信号,传输到扬声器1661,由扬声器1661转换为声音信号输出;另一方面,传声器1662将收集的声音信号转换为电信号,由音频电路1660接收后转换为音频数据,再将音频数据输出处理器1680处理后,经RF电路1610以发送给比如另一电子设备,或者将音频数据输出至存储器1620以便进一步处理。

[0238] WiFi属于短距离无线传输技术,电子设备通过WiFi模块1670可以帮助目标对象收发电子邮件、浏览网页和访问流式媒体等,它为对象提供了无线的宽带互联网访问。虽然图16示出了WiFi模块1670,但是可以理解的是,其并不属于电子设备的必须构成,完全可以根据需要在不改变发明的本质的范围内而省略。

[0239] 处理器1680是电子设备的控制中心,利用各种接口和线路连接整个电子设备的各个部分,通过运行或执行存储在存储器1620内的软件程序和/或模块,以及调用存储在存储器1620内的数据,执行电子设备的各种功能和处理数据。可选的,处理器1680可包括一个或多个处理单元;可选的,处理器1680可集成应用处理器和调制解调处理器,其中,应用处理器主要处理操作系统、应用及应用内部的功能模块等软件程序,如本申请实施例提供的多标签图像分类方法等。调制解调处理器主要处理无线通信。可以理解的是,上述调制解调处理器也可以不集成到处理器1680中。

[0240] 可以理解,图16所示的结构仅为示意,电子设备还可包括比图16中所示更多或者更少的组件,或者具有与图16所示不同的配置。图16中所示的各组件可以采用硬件、软件或其组合实现。

[0241] 根据本申请的一个方面,提供了一种计算机程序产品或计算机程序,该计算机程序产品或计算机程序包括计算机指令,该计算机指令存储在计算机可读存储介质中。计算机设备的处理器从计算机可读存储介质读取该计算机指令,处理器执行该计算机指令,使得该计算机设备执行上述实施例中的多标签图像分类方法。

[0242] 程序产品可以采用一个或多个可读介质的任意组合。可读介质可以是可读信号介质或者可读存储介质。可读存储介质例如可以是——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。可读存储介质的更具体的例子(非穷举的列表)包括:具有一个或多个导线的电连接、便携式盘、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)、光纤、便携式紧凑盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。

[0243] 以上所述,仅为本申请的具体实施方式,但本申请的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本申请揭露的技术范围内,可轻易想到变化或替换,都应涵盖在本申请的保护范围之内。



图1

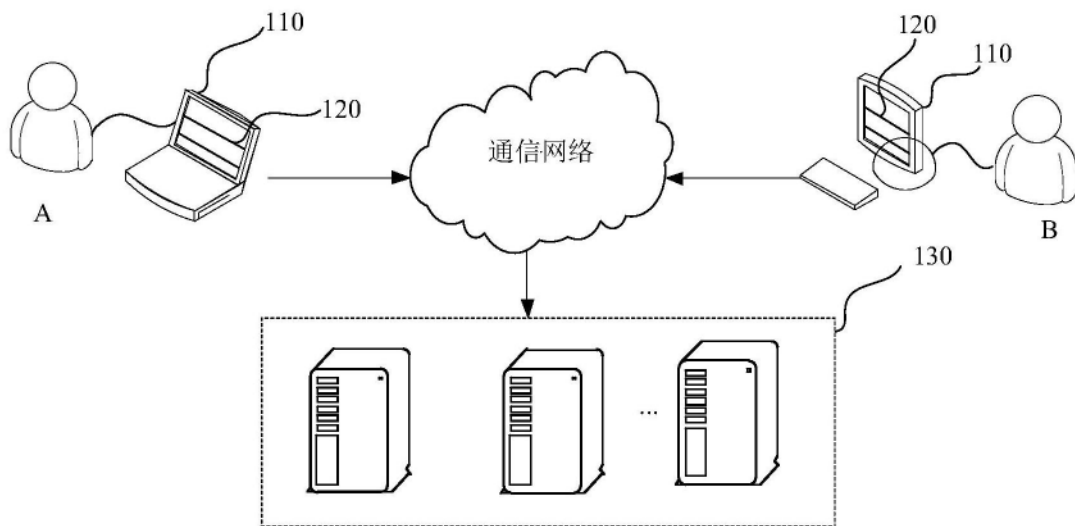


图2

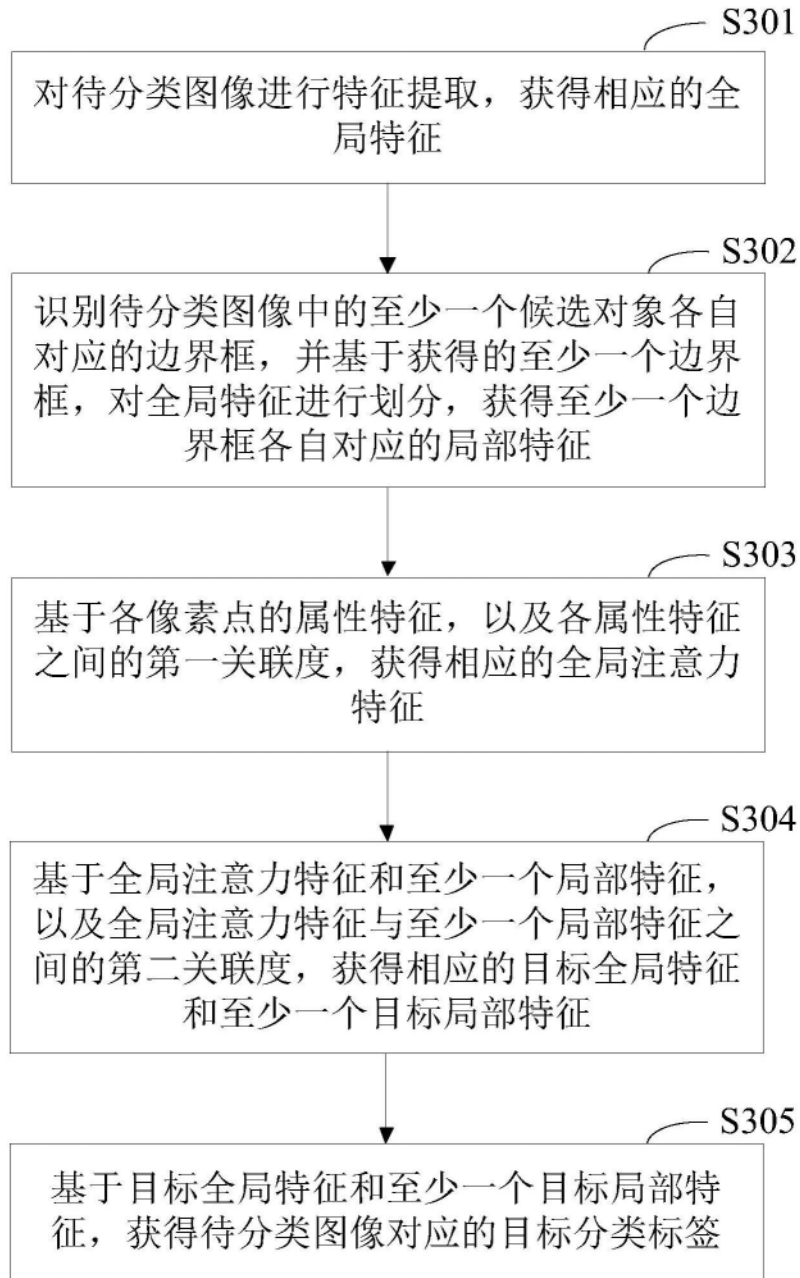


图3

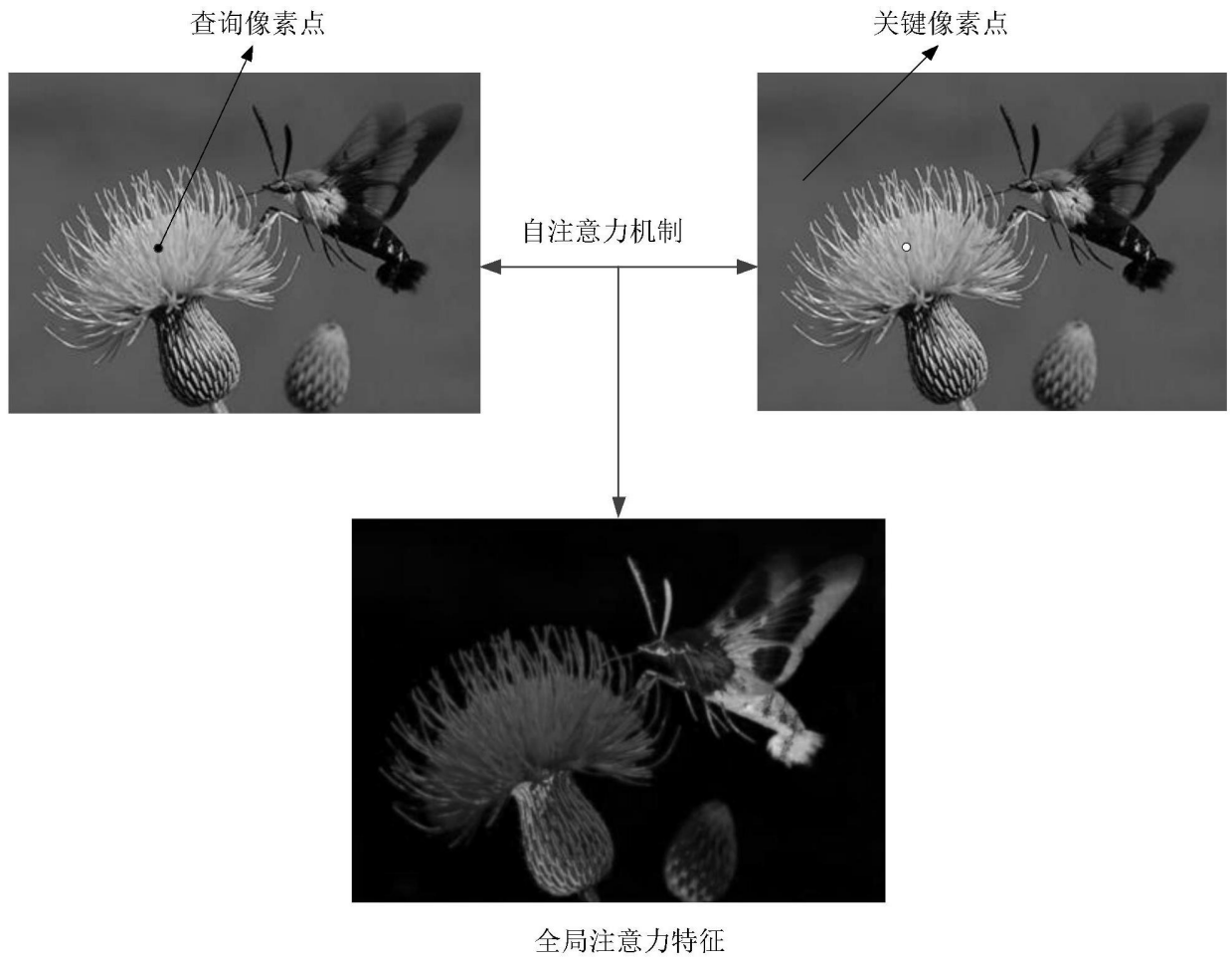


图4

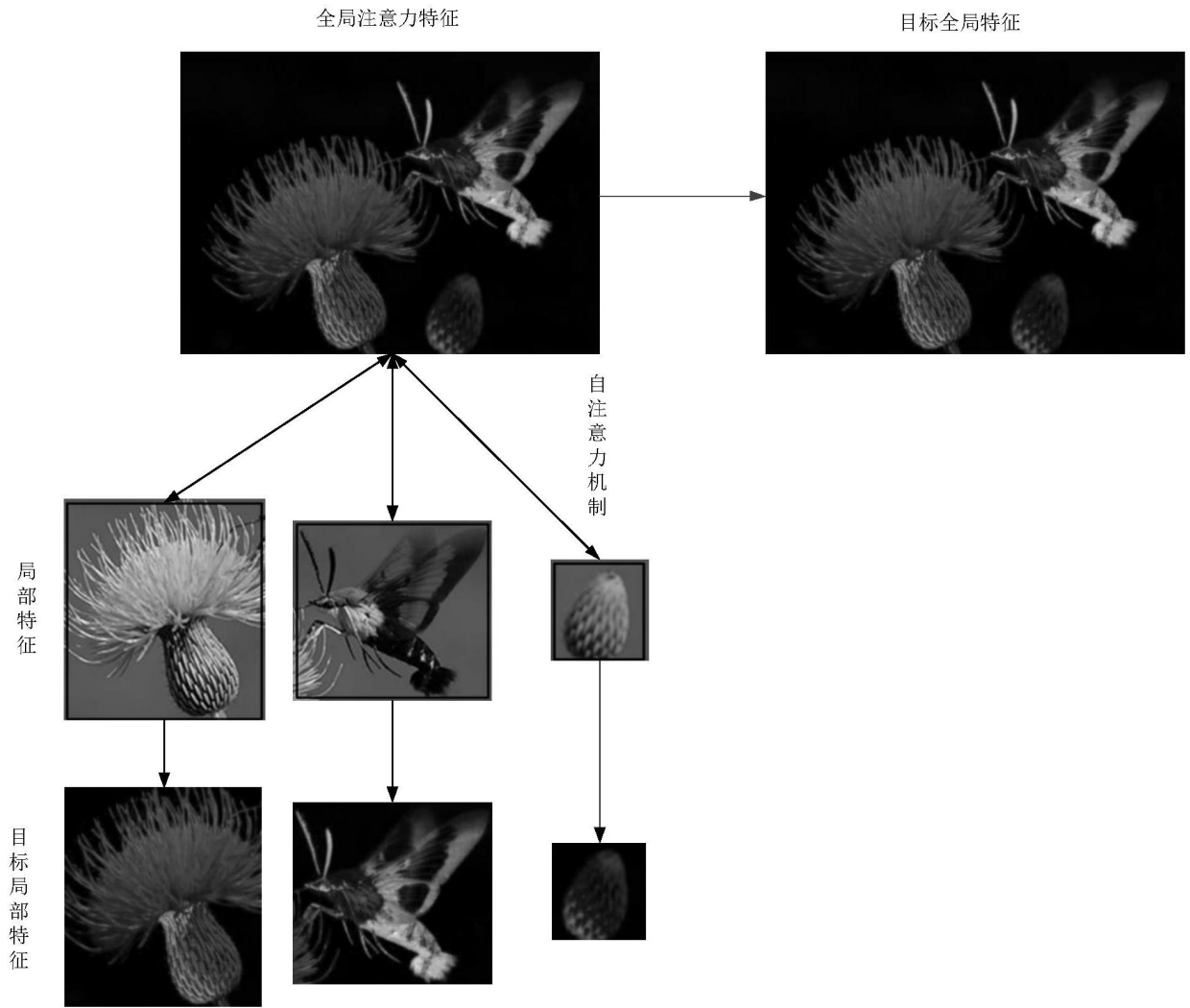


图5



图6

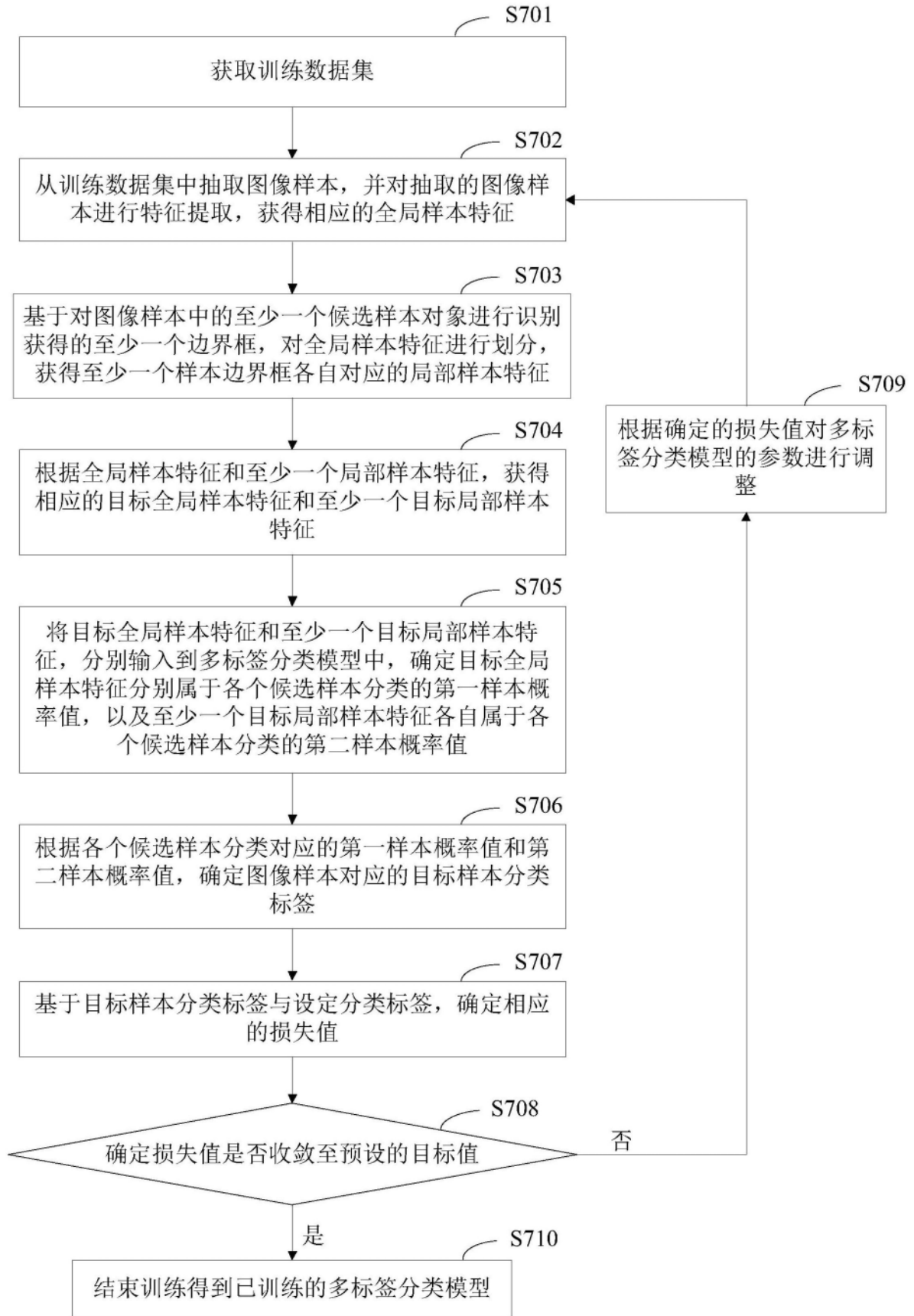


图7

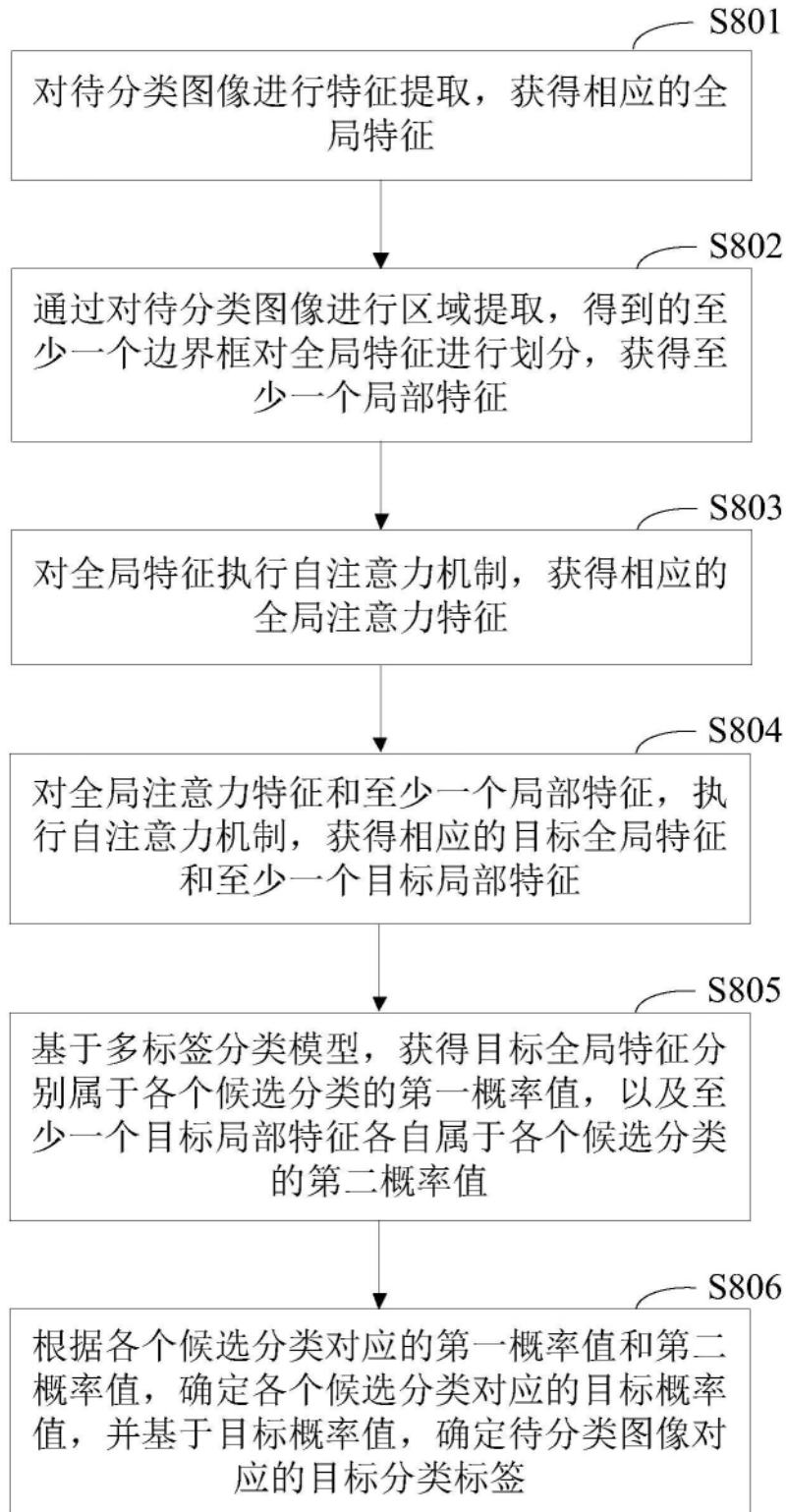


图8

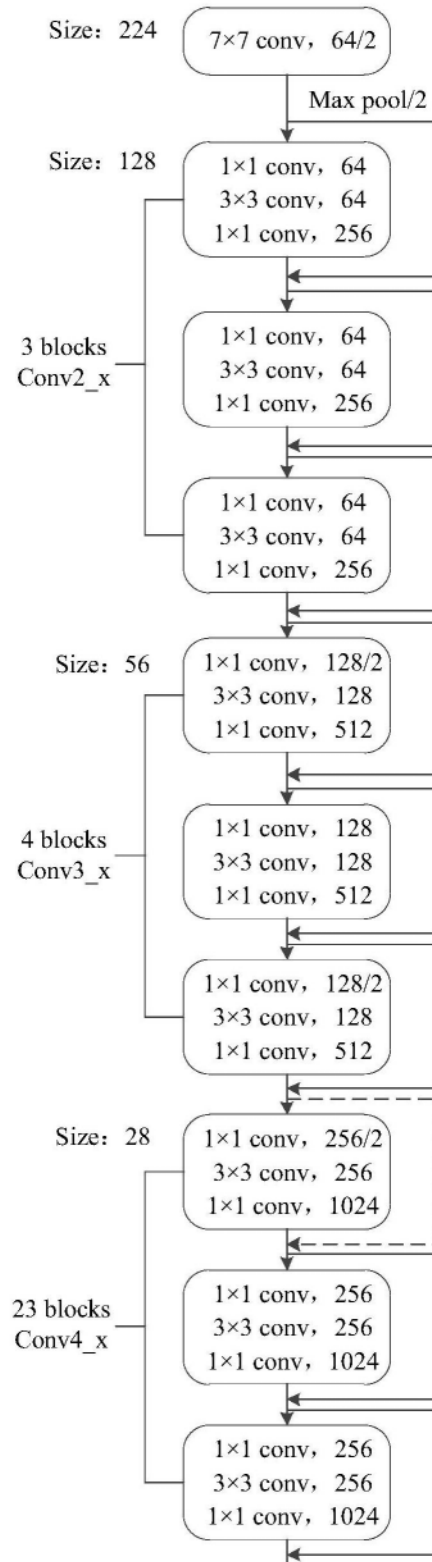


图9

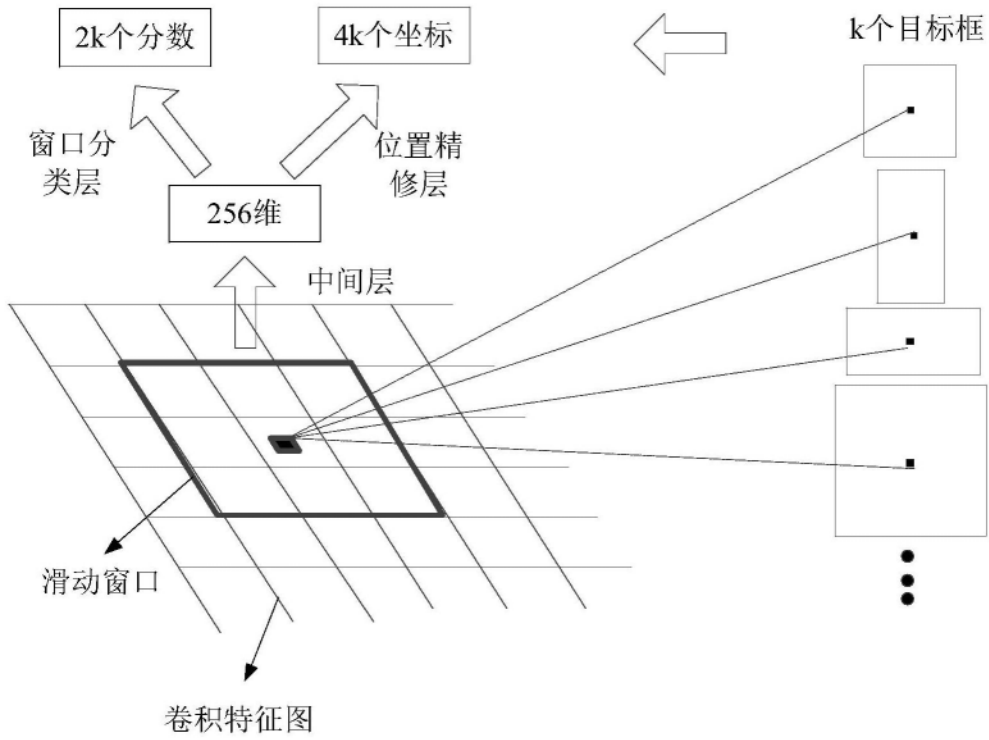


图10

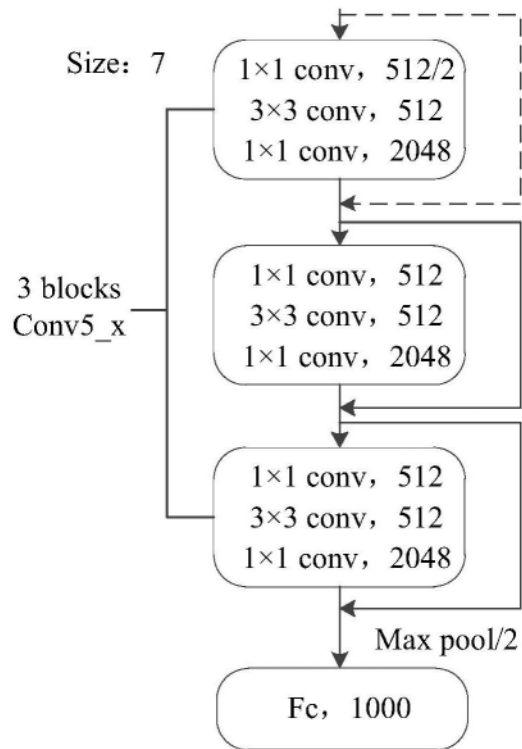
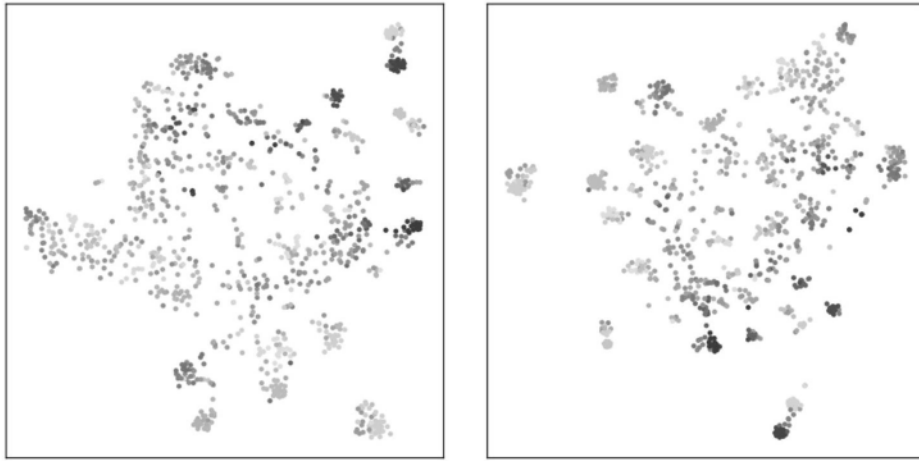


图11



(a) 无跨粒度注意力 (b) 结合跨粒度注意力

图12a

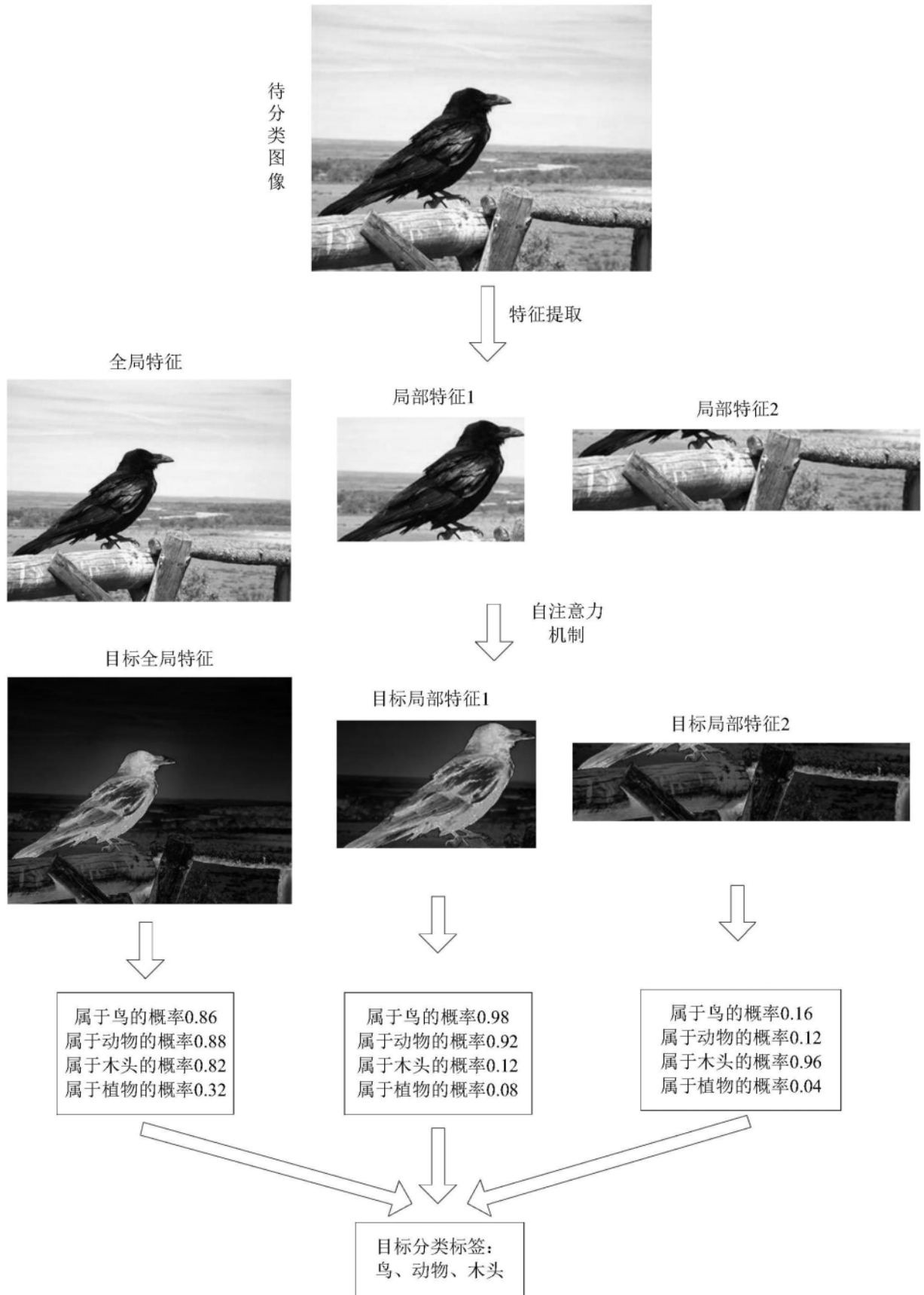


图12b

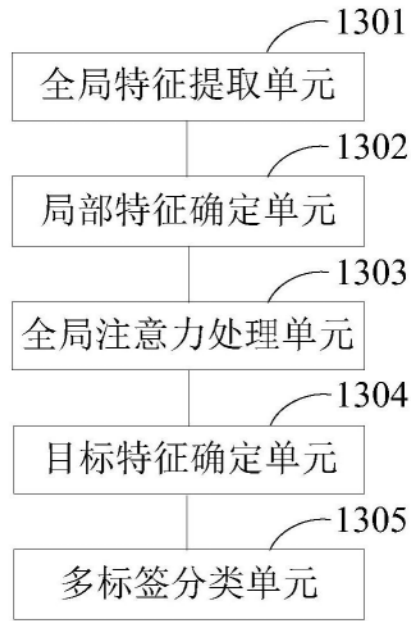


图13

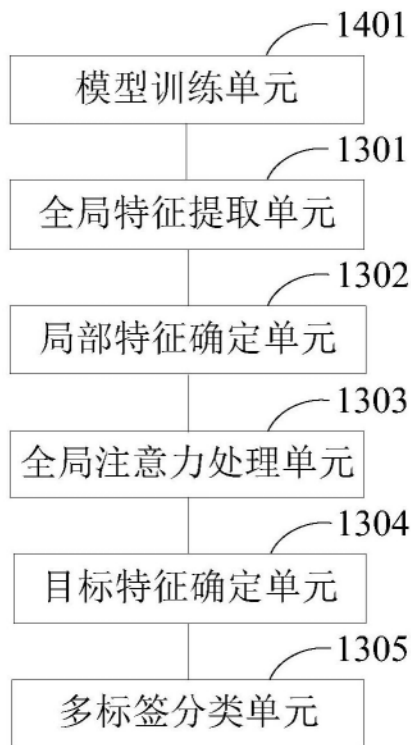


图14

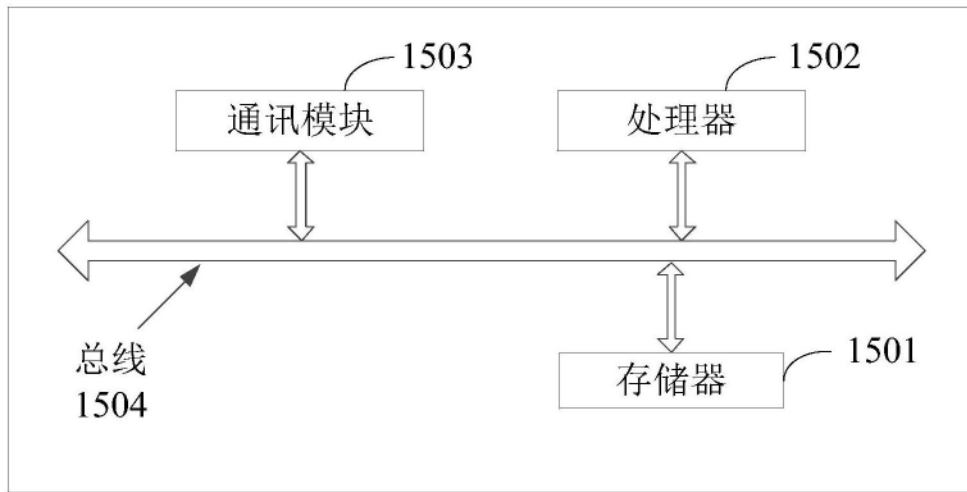


图15

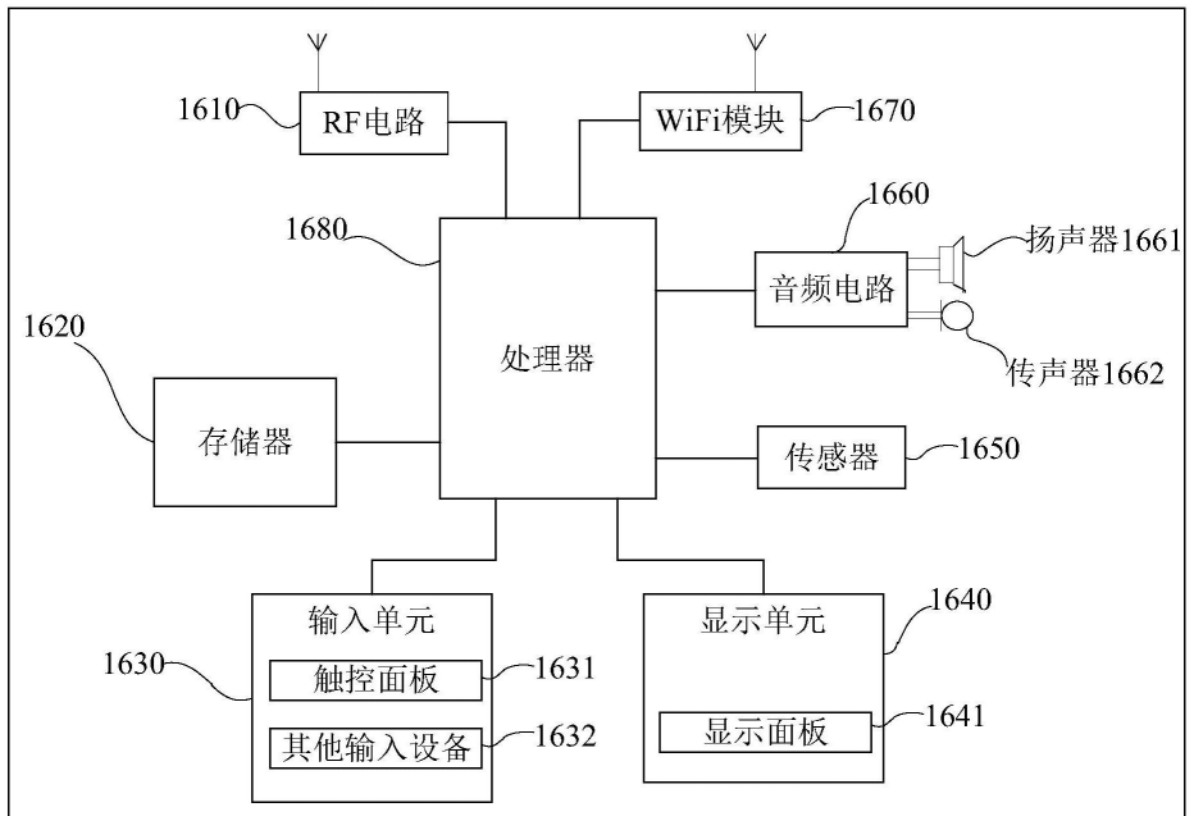


图16