



(12) 发明专利

(10) 授权公告号 CN 112183537 B

(45) 授权公告日 2021.03.19

(21) 申请号 202011366763.X

G06N 3/04 (2006.01)

(22) 申请日 2020.11.30

(56) 对比文件

(65) 同一申请的已公布的文献号

CN 104050673 A, 2014.09.17

申请公布号 CN 112183537 A

CN 107204010 A, 2017.09.26

CN 109035319 A, 2018.12.18

(43) 申请公布日 2021.01.05

CN 104112275 A, 2014.10.22

(73) 专利权人 北京易真学思教育科技有限公司

US 2012050284 A1, 2012.03.01

地址 100144 北京市石景山区实兴大街30

审查员 郭放

号院3号楼2层A-2667房间

(72) 发明人 张蓓蓓 秦勇

(74) 专利代理机构 北京市铸成律师事务所

11313

代理人 阎敏 邓海鸿

(51) Int. Cl.

G06K 9/32 (2006.01)

G06K 9/62 (2006.01)

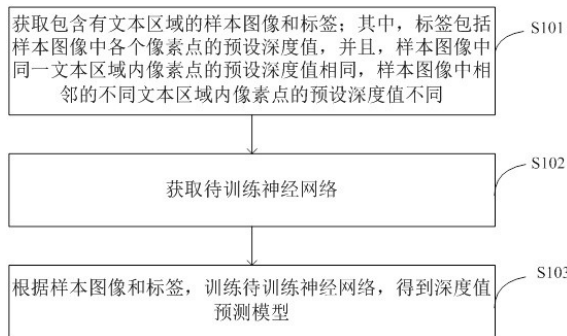
权利要求书3页 说明书12页 附图4页

(54) 发明名称

模型训练方法及装置、文本区域检测方法及装置

(57) 摘要

本申请提出模型训练方法及装置、文本区域检测方法及装置,其中,模型训练方法包括:获取包含有文本区域的样本图像和标签;其中,标签包括样本图像中各个像素点的预设深度值,并且,样本图像中同一文本区域内像素点的预设深度值相同,样本图像中相邻的不同文本区域内像素点的预设深度值不同;获取待训练神经网络;根据样本图像和标签,训练待训练神经网络,得到深度值预测模型。本申请有利于提高文本区域检测的准确率。



1. 一种模型训练方法,其特征在于,包括:

获取包含有文本区域的样本图像和标签;其中,所述标签包括所述样本图像中各个像素点的预设深度值,并且,所述样本图像中同一文本区域内像素点的预设深度值相同,所述样本图像中相邻的不同文本区域内像素点的预设深度值不同;所述预设深度值是仿照真实深度信息而设置的伪深度信息;

获取待训练神经网络;

根据所述样本图像和所述标签,训练所述待训练神经网络,得到深度值预测模型。

2. 根据权利要求1所述的方法,其特征在于,所述样本图像中非文本区域的像素点的预设深度值与文本区域的像素点的预设深度值不同。

3. 根据权利要求1或2所述的方法,其特征在于,所述待训练神经网络至少包括依次连接的特征提取网络、特征增强网络、卷积层和反卷积层;

其中,所述特征提取网络用于对输入所述待训练神经网络的带有所述标签的所述样本图像进行特征提取;所述特征增强网络用于对所述特征提取网络提取的特征进行增强处理;所述卷积层用于对所述特征增强网络增强后的特征进行卷积操作;所述反卷积层用于对所述卷积层卷积操作后的特征进行反卷积操作。

4. 根据权利要求3所述的方法,其特征在于,所述特征提取网络包括残差网络,和/或,所述特征增强网络包括特征金字塔增强模块。

5. 一种文本区域检测方法,其特征在于,包括:

获取待测图像;

将所述待测图像输入至深度值预测模型,得到所述待测图像中各个像素点的预测深度值;其中,所述深度值预测模型为采用权利要求1至4任一项所述的模型训练方法训练所得的模型;

根据所述待测图像中各个像素点的预测深度值,确定所述待测图像中的目标文本区域。

6. 根据权利要求5所述的方法,其特征在于,还包括:获取所述待测图像的初始文本区域;

所述根据所述待测图像中各个像素点的预测深度值,确定所述待测图像中的目标文本区域,包括:根据所述待测图像中各个像素点的预测深度值,对所述初始文本区域进行调整,得到所述待测图像的目标文本区域。

7. 根据权利要求6所述的方法,其特征在于,所述根据所述待测图像中各个像素点的预测深度值,对所述初始文本区域进行调整,得到所述待测图像的目标文本区域,包括:

基于文本区域中像素点的预测深度值相同的调整规则,并根据所述待测图像中各个像素点的预测深度值,对所述初始文本区域进行调整,得到所述待测图像的目标文本区域。

8. 根据权利要求7所述的方法,其特征在于,所述基于文本区域中像素点的预测深度值相同的调整规则,并根据所述待测图像中各个像素点的预测深度值,对所述初始文本区域进行调整,得到所述待测图像的目标文本区域,包括:

在所述初始文本区域中存在至少两种预测深度值的像素点的情况下,基于所述预测深度值对所述初始文本区域进行拆分,以使不同预测深度值的像素点位于不同的文本区域,得到所述待测图像的目标文本区域;

或者，

在所述初始文本区域中存在至少两种预测深度值的像素点，且所述至少两种预测深度值的像素点的占比均大于预设比例的情况下，对所述初始文本区域进行拆分，以使不同预测深度值的像素点位于不同的文本区域，得到所述待测图像的目标文本区域。

9. 根据权利要求6或7所述的方法，其特征在于，所述获取所述待测图像的初始文本区域，包括：

预测所述待测图像中各个像素点属于文本区域的概率；

根据所述待测图像中各个像素点属于文本区域的概率，确定所述待测图像的初始文本区域。

10. 根据权利要求9所述的方法，其特征在于，所述预测所述待测图像中各个像素点属于文本区域的概率，包括：

获取文本区域概率预测模型；

将所述待测图像输入至所述文本区域概率预测模型，得到所述待测图像中各个像素点属于文本区域的概率。

11. 一种模型训练装置，其特征在于，包括：

样本图像获取模块，用于获取包含有文本区域的样本图像和标签；其中，所述标签包括所述样本图像中各个像素点的预设深度值，并且，所述样本图像中同一文本区域内像素点的预设深度值相同，所述样本图像中相邻的不同文本区域内像素点的预设深度值不同；所述预设深度值是仿照真实深度信息而设置的伪深度信息；

待训练神经网络获取模块，用于获取待训练神经网络；

训练模块，用于根据所述样本图像和所述标签，训练所述待训练神经网络，得到深度值预测模型。

12. 根据权利要求11所述的装置，其特征在于，所述样本图像中非文本区域的像素点的预设深度值与文本区域的像素点的预设深度值不同。

13. 根据权利要求11或12所述的装置，其特征在于，所述待训练神经网络至少包括依次连接的特征提取网络、特征增强网络、卷积层和反卷积层；

其中，所述特征提取网络用于对输入所述待训练神经网络的带有所述标签的所述样本图像进行特征提取；所述特征增强网络用于对所述特征提取网络提取的特征进行增强处理；所述卷积层用于对所述特征增强网络增强后的特征进行卷积操作；所述反卷积层用于对所述卷积层卷积操作后的特征进行反卷积操作。

14. 根据权利要求13所述的装置，其特征在于，所述特征提取网络包括残差网络，和/或，所述特征增强网络包括特征金字塔增强模块。

15. 一种文本区域检测装置，其特征在于，包括：

待测图像获取模块，用于获取待测图像；

预测深度值获取模块，用于将所述待测图像输入至深度值预测模型，得到所述待测图像中各个像素点的预测深度值；其中，所述深度值预测模型为采用权利要求11至14任一项所述的模型训练装置训练所得的模型；

目标文本区域确定模块，用于根据所述待测图像中各个像素点的预测深度值，确定所述待测图像中的目标文本区域。

16. 根据权利要求15所述的装置,其特征在于,还包括:初始文本区域获取模块;
其中,所述初始文本区域获取模块,用于获取所述待测图像的初始文本区域;
所述目标文本区域确定模块,用于根据所述待测图像中各个像素点的预测深度值,对所述初始文本区域进行调整,得到所述待测图像的目标文本区域。

17. 根据权利要求16所述的装置,其特征在于,所述目标文本区域确定模块用于基于文本区域中像素点的预测深度值相同的调整规则,并根据所述待测图像中各个像素点的预测深度值,对所述初始文本区域进行调整,得到所述待测图像的目标文本区域。

18. 根据权利要求17所述的装置,其特征在于,所述目标文本区域确定模块,还用于:
在所述初始文本区域中存在至少两种预测深度值的像素点的情况下,基于所述预测深度值对所述初始文本区域进行拆分,以使不同预测深度值的像素点位于不同的文本区域,得到所述待测图像的目标文本区域;

或者,

在所述初始文本区域中存在至少两种预测深度值的像素点,且所述至少两种预测深度值的像素点的占比均大于预设比例的情况下,对所述初始文本区域进行拆分,以使不同预测深度值的像素点位于不同的文本区域,得到所述待测图像的目标文本区域。

19. 根据权利要求16或17所述的装置,其特征在于,所述初始文本区域获取模块,包括:
文本区域概率预测子模块,用于预测所述待测图像中各个像素点属于文本区域的概率;

初始文本区域确定子模块,用于根据所述待测图像中各个像素点属于文本区域的概率,确定所述待测图像的初始文本区域。

20. 根据权利要求19所述的装置,其特征在于,所述文本区域概率预测子模块,包括:
文本区域概率预测模型获取单元,用于获取文本区域概率预测模型;
文本区域概率确定单元,用于将所述待测图像输入至所述文本区域概率预测模型,得到所述待测图像中各个像素点属于文本区域的概率。

21. 一种电子设备,其特征在于,包括:处理器和存储器,所述存储器中存储指令,所述指令由处理器加载并执行,以实现如权利要求1至4任一项所述的模型训练方法或权利要求5至10任一项所述的文本区域检测方法。

22. 一种计算机可读存储介质,所述计算机可读存储介质内存储有计算机程序,所述计算机程序被处理器执行时实现如权利要求1至4任一项所述的模型训练方法或权利要求5至10任一项所述的文本区域检测方法。

模型训练方法及装置、文本区域检测方法及装置

技术领域

[0001] 本申请涉及人工智能,尤其涉及一种模型训练方法及装置、文本区域预测方法及装置。

背景技术

[0002] 文本检测应用范围广泛,是很多计算机视觉任务的重要步骤。文本检测的主要目的是定位文本行或字符在图像中的位置。随着深度学习的兴起,出现了大量专用于文本检测方法,但是针对文本非常密集的应用场景,比如小学生练习册,可能有百余个文本区域,且文本框之间间距小,此时,容易将多个邻近的文本框检测为同一文本框,降低了文本检测的准确率。

发明内容

[0003] 本申请实施例提供一种模型训练方法及装置、文本区域检测方法及装置,以解决相关技术存在的问题,技术方案如下:

[0004] 第一方面,本申请实施例提供了一种模型训练方法,包括:

[0005] 获取包含有文本区域的样本图像和标签;其中,标签包括样本图像中各个像素点的预设深度值,并且,样本图像中同一文本区域内像素点的预设深度值相同,样本图像中相邻的不同文本区域内像素点的预设深度值不同;

[0006] 获取待训练神经网络;

[0007] 根据样本图像和标签,训练待训练神经网络,得到深度值预测模型。

[0008] 第二方面,本申请实施例还提供一种文本区域检测方法,包括:

[0009] 获取待测图像;

[0010] 将待测图像输入至深度值预测模型,得到待测图像中各个像素点的预测深度值;其中,深度值预测模型为采用第一方面提供的模型训练方法训练所得的模型;

[0011] 根据待测图像中各个像素点的预测深度值,确定待测图像中的目标文本区域。

[0012] 第三方面,本申请实施例还提供一种模型训练装置,包括:

[0013] 样本图像获取模块,用于获取包含有文本区域的样本图像和标签;其中,标签包括样本图像中各个像素点的预设深度值,并且,样本图像中同一文本区域内像素点的预设深度值相同,样本图像中相邻的不同文本区域内像素点的预设深度值不同;

[0014] 待训练神经网络获取模块,用于获取待训练神经网络;

[0015] 训练模块,用于根据样本图像和标签,训练待训练神经网络,得到深度值预测模型。

[0016] 第四方面,本申请实施例还提供一种文本区域检测装置,包括:

[0017] 待测图像获取模块,用于获取待测图像;

[0018] 预测深度值获取模块,用于将待测图像输入至深度值预测模型,得到待测图像中各个像素点的预测深度值;其中,深度值预测模型为采用第三方面提供的模型训练装置训

练所得的模型；

[0019] 目标文本区域确定模块，用于根据待测图像中各个像素点的预测深度值，确定待测图像中的目标文本区域。

[0020] 第五方面，本申请实施例提供了一种电子设备，该电子设备包括：存储器和处理器。其中，该存储器和该处理器通过内部连接通路互相通信，该存储器用于存储指令，该处理器用于执行该存储器存储的指令，并且当该处理器执行该存储器存储的指令时，使得该处理器执行上述各方面任一种实施方式中的方法。

[0021] 第六方面，本申请实施例提供了一种计算机可读存储介质，计算机可读存储介质存储计算机程序，当计算机程序在计算机上运行时，上述各方面任一种实施方式中的方法被执行。

[0022] 上述技术方案中的优点或有益效果至少包括：通过本申请实施例的深度值预测模型所实现的文本检测，有利于提高文本区域检测的准确率。

[0023] 上述概述仅仅是为了说明书的目的，并不意图以任何方式进行限制。除上述描述的示意性的方面、实施方式和特征之外，通过参考附图和以下的详细描述，本申请进一步的方面、实施方式和特征将会是容易明白的。

附图说明

[0024] 在附图中，除非另外规定，否则贯穿多个附图相同的附图标记表示相同或相似的部件或元素。这些附图不一定是按照比例绘制的。应该理解，这些附图仅描绘了根据本申请公开的一些实施方式，而不应将其视为是对本申请范围的限制。

[0025] 图1为根据本申请实施例的模型训练方法的流程图；

[0026] 图2为根据本申请实施例的模型训练方法所提供的一样本图像的示例图；

[0027] 图3为根据本申请实施例的文本区域检测方法的流程图；

[0028] 图4为根据本申请实施例的文本区域检测方法的一示例的流程图；

[0029] 图5为根据本申请实施例的模型训练装置的结构框图；

[0030] 图6为根据本申请实施例的文本区域检测装置的结构框图；

[0031] 图7为根据本申请实施例的电子设备的结构框图。

具体实施方式

[0032] 在下文中，仅简单地描述了某些示例性实施例。正如本领域技术人员可认识到的那样，在不脱离本申请的精神或范围的情况下，可通过各种不同方式修改所描述的实施例。因此，附图和描述被认为本质上是示例性的而非限制性的。

[0033] 图1示出根据本申请一实施例的模型训练方法的流程图。如图1所示，该模型训练方法可以包括：

[0034] S101、获取包含有文本区域的样本图像和标签；其中，标签包括样本图像中各个像素点的预设深度值，并且，样本图像中同一文本区域内像素点的预设深度值相同，样本图像中相邻的不同文本区域内像素点的预设深度值不同；

[0035] S102、获取待训练神经网络；

[0036] S103、根据样本图像和标签，训练待训练神经网络，得到深度值预测模型。

[0037] 其中,深度值预测模型可以用于预测文本图像中各个像素点的预测深度值,以利用预测深度值确定文本图像中的文本区域。

[0038] 需要说明的是,对于一张二维图像来说,深度值指的是各个像素点到相机的距离。通常,在一张二维图像内得到的各个像素点的深度值大概率都是相同的(除非文本扭曲)。而在本申请实施例中,预设深度值以及后续的预测深度值,并非表征图像的真实深度信息或深度值,而是,仿照深度信息而设置的伪深度信息。如此,将相同的文本区域中伪深度信息设置为相同的预设深度值,将相邻的不同文本区域的伪深度信息设置为不同的预设深度值,为后续基于伪深度信息(也即预设深度值)对文本区域进行调整奠定了基础。对于各个样本图像,可以依据预设规律标注各个像素点的深度值。举例来说,如果一个文本区域没有跟其它文本区域相邻,则该文本区域的预设深度值为0.3。如果两个文本区域相邻,则面积较小的文本区域取较小的预设深度值,比如0.5,面积较大的文本区域取较大的预设深度值,比如0.7。如果3个文本区域连接在一起,按照文本区域面积由小到大,预设深度值依次为0.3、0.5和0.7。其它多个文本区域的情况均可依次类推。

[0039] 其中,文本区域中的文本可以是一种或多种字符的组合。如数学字符、数字、中文或英文等。用于训练的样本图像可以是练习册、课本、报刊或书籍等。

[0040] 参见图2示例,给出练习册的扫描图像作为样本图像为例,一道计算题对应占据一个文本区域,如图中示出的文本区域1和文本区域2作为例子。由于该练习册的计算题之间间距十分小,导致相邻的文本区域较难区分,增大了文本检测的难度。比如图2中,文本区域1和文本区域2这两个文本区域的边框相接在一起,此时容易将文本区域1和文本区域2识别为一个文本区域,导致文本检测效果较差。

[0041] 在本申请实施例中,为样本图像的各个像素点伪造深度值作为标签,将样本图像中相邻的不同文本区域通过不同的深度值进行体现。结合单目深度估计的思想,即通过一张图像来估计各个像素点的深度信息的思想,利用上述样本图像和标签训练用于预测深度值的模型。那么,通过该模型预测深度值时,相邻的不同文本区域的像素点得到不同的深度值。利用该模型所预测的深度值,可以帮助区分图像中相邻的不同文本区域,提高文本检测的抗干扰能力,而且,单目深度估计的计算较为简单。因此,通过本申请实施例的深度值预测模型所实现的文本检测,其准确率更高,检测效率更高。

[0042] 在一种实施方式中,样本图像中非文本区域的像素点的预设深度值与文本区域的像素点的预设深度值不同。

[0043] 例如,如果像素点所在不属于文本区域,其预设深度值设为0。如果像素点所在区域属于文本区域,其预设深度值的取值范围为(0,1]。当通过该深度值预测模型预测像素点的预测深度值取值为0,可以确定其属于非文本区域。通过上述实施方式,使得深度值预测模型的预测结果可以用于进一步区分文本区域和非文本区域。

[0044] 在一种实施方式中,待训练神经网络至少包括依次连接的特征提取网络、特征增强网络、卷积层和反卷积层。

[0045] 其中,特征提取网络用于对输入待训练神经网络的带有标签的样本图像进行特征提取;特征增强网络用于对特征提取网络提取的特征进行增强处理;卷积层用于对特征增强网络增强后的特征进行卷积操作;反卷积层用于对卷积层卷积操作后的特征进行反卷积操作。

[0046] 上述待训练神经网络,在前处理部分,采用特征提取网络和特征增强网络的结合,使得提取的特征更加精细化,能够更加充分地表征输入的文本图像。在后处理部分,采用卷积层和反卷积层的结合,处理简单,有利于提升检测速度。尤其是在密集文本场景中,文本区域数量非常多,检测速度受到文本区域数量的影响非常大,本申请实施例显著提高文本检测速度。

[0047] 进一步地,卷积层的数量为一层,反卷积层的数量为两层。

[0048] 进一步地,在将特征增强网络增强后的特征输入卷积层之前,还包括:对增强后的特征全部上采样到样本图像的1/4大小,并对上采样处理后的特征进行串联,以将串联后的特征输入卷积层。

[0049] 进一步地,上述串联后的特征在经过卷积层和反卷积层处理后,得到一个1通道的特征图,其大小与样本图像相同。

[0050] 在一种实施方式中,特征提取网络包括残差网络Resnet,和/或,特征增强网络包括特征金字塔增强模块FPPEM。

[0051] 具体地,残差网络Resnet对输入图像提取纹理、边缘、角点和语义信息等特征,这些特征可以由多组大小不同的多通道特征映射表征。然后,将提取得到的特征经过特征金字塔增强模块FPPEM,再次提取纹理、边缘、角点和语义信息等特征,使得提取的特征更加精细化,能够更加充分地表征输入文本图像。

[0052] 进一步地,残差网络Resnet采用Resnet18。

[0053] 进一步地,特征增强网络包括两个特征金字塔增强模块FPPEM。两个FPPEM模块串联连接,即第二个FPPEM模块是以第一个FPPEM模块的输出作为输入。

[0054] 以Resnet为Resnet18为例,得到的四组大小不同的多通道特征,按照从大到小的,从前往后的次序,依次称为第一组正向特征、第二组正向特征、第三组正向特征、第四组正向特征。将Resnet18的这四组正向特征输入第一个FPPEM模块。

[0055] 对应地,第一个FPPEM模块,对输入的四组正向特征进行特征增强处理,特征增强处理可以参照下述示例说明。

[0056] 第一步,根据四组正向特征确定四组反向特征。具体的,首先,将第四组正向特征视为第一组反向特征。然后,对第一组反向特征进行2倍上采样后与第三组正向特征按照通道逐点相加,对相加结果依次进行一次深度可分离卷积操作、一次卷积、批归一化和激活函数作用操作,得到第二组反向特征。依次类推,根据第二组反向特征和第三组正向特征,得到第三组反向特征;根据第三组反向特征和第一组正向特征得到第四组反向特征。由此得到四组反向特征。

[0057] 第二步,根据四组反向特征确定四组目标特征。具体的,首先,将第四组反向特征作为第一组目标特征。然后,对第一组目标特征进行2倍下采样后与第三组反向特征按照通道逐点相加,对相加结果依次进行一次深度可分离卷积操作、一次卷积、批归一化和激活函数作用操作,得到第二组目标特征。依次类推,根据第二组目标特征和第三组反向特征,得到第三组目标特征;根据第三组目标特征和第一组反向特征得到第四组目标特征。由此得到四组目标特征。

[0058] 第三步,输出四组目标特征。

[0059] 对于第二个FPPEM模块,参照上述第一个FPPEM模块的特征增强处理过程,对四组目

标特征进行特征增强处理,同样可以得到四组特征。

[0060] 在一种实施方式中,在步骤S103的训练过程中,可以使用目标损失函数进行训练优化。例如,损失函数可以采用L1损失函数。

[0061] 本申请实施例结合单目深度估计的思想,利用特征提取网络和特征增强网络的组合能够更好的提取图像特征的优势,以及采用卷积层加反卷积层的组合处理达到简化计算量。从而,在保证文本检测效果的前提下,实现了简单的处理思路和更少的计算量,网络训练更简单,模型更容易拟合,此外,提升密集文本检测速度并使检测操作更具鲁棒性。

[0062] 图3示出根据本申请一实施例的文本区域检测方法的流程图。如图3所示,该文本区域检测方法可以包括:

[0063] S301、获取待测图像。

[0064] S302、将待测图像输入至深度值预测模型,得到待测图像中各个像素点的预测深度值;其中,深度值预测模型为本申请实施例所提供的模型训练方法训练所得的模型,详细训练方法可以参照上述实施例的说明,此处不再赘述。

[0065] S303、根据待测图像中各个像素点的预测深度值,确定待测图像中的目标文本区域。

[0066] 而在本申请实施例中,模型训练的预设深度值以及后续模型预测的预测深度值,并非表征图像的真实深度信息或深度值,而是,仿照深度信息而设置的伪深度信息。如此,模型训练阶段将相同的文本区域中伪深度信息设置为相同的预设深度值,将不同的文本区域的伪深度信息设置为不同的预设深度值,为后续基于伪深度信息(也即预设深度值)对文本区域进行调整奠定了基础。

[0067] 在本申请实施例中,通过深度值预测模型预测深度值时,得到的预测结果中,相邻的不同文本区域的像素点得到不同的深度值。利用该模型所预测的深度值,可以帮助区分图像中相邻的不同文本区域,提高文本检测的抗干扰能力,而且,单目深度估计的计算较为简单。因此,通过本申请实施例的深度值预测模型所实现的文本检测,其鲁棒性更强,检测效率更高。

[0068] 在一种实施方式中,在步骤S302中,深度值预测模型以像素点预测深度值图的形式输出待测图像中各个像素点的预测深度值。像素点预测深度值图中的每个值表示对应位置像素点的预测深度值。

[0069] 在一种实施方式中,对于步骤S303,可以先将待测图像中各个像素点的预测深度值进行离散化处理,比如说,预测深度值小于0.3的全部设为1,大于等于0.3小于0.5的设为2,依次类推。然后,根据离散化后的像素点预测深度值,判断待测图像中的像素点有几种预测深度值,再根据同一目标文本区域内的预测深度值相同的原则,确定待测图像中目标文本区域。

[0070] 在一种实施方式中,本申请实施例提供的文本区域检测方法,还包括步骤:获取待测图像的初始文本区域。

[0071] 其中,初始文本区域可以理解为按照预设的初始文本区域确定方法所确定的待测图像中所有可能为文本区域的区域,如基于连通域的文本检测方法、基于滑动窗口的文本检测方法。

[0072] 对应地,步骤S303,包括:根据待测图像中各个像素点的预测深度值,对初始文本

区域进行调整,得到待测图像的目标文本区域。

[0073] 在上述实施方式中,利用不同预测深度值能够区分相邻的不同文本区域的特点,可以对初始文本区域进行进一步的划分,避免了将相邻的两个或多个真实文本区域误识别为一个初始文本区域的情况,提高了文本区域检测的准确度。

[0074] 进一步的,上述调整过程可以针对所有初始文本区域并行操作,有利于提高处理速度。

[0075] 在一种实施方式中,根据待测图像中各个像素点的预测深度值,对初始文本区域进行调整,得到待测图像的目标文本区域,包括:

[0076] 基于文本区域中像素点的预测深度值相同的调整规则,并根据待测图像中各个像素点的预测深度值,对初始文本区域进行调整,得到待测图像的目标文本区域。通过调整,使得得到的各个目标文本区域,同一目标文本区域中的预测深度值相同,不同的目标文本区域中的预测深度值不同。

[0077] 例如,对于一个初始文本区域,如果存在两种预测深度值的像素点,可以根据这两种深度值像素点对应得到两个目标文本区域;如果存在三种预测深度值的像素点,可以得到三个目标文本区域。可以依次类推其它情况。在一种实施方式中,基于文本区域中像素点的预测深度值相同的调整规则,并根据待测图像中各个像素点的预测深度值,对初始文本区域进行调整,得到待测图像的目标文本区域,包括:

[0078] 在初始文本区域中存在至少两种预测深度值的像素点的情况下,基于预测深度值对初始文本区域进行拆分,以使不同预测深度值的像素点位于不同的文本区域,得到待测图像的目标文本区域;

[0079] 或者,

[0080] 在初始文本区域中存在至少两种预测深度值的像素点,且至少两种预测深度值的像素点的占比均大于预设比例的情况下,对初始文本区域进行拆分,以使不同预测深度值的像素点位于不同的文本区域,得到待测图像的目标文本区域。

[0081] 举例说明,一个初始文本区域内包括两种预测深度值,若其中一种预测深度值对应的像素点数量不超过5%,那么就认为该初始文本区域为一个目标文本区域,而不拆分为两个目标文本区域,否则,比如,两种预测深度值的占比均大于5%,此时,认为初始文本区域包含有两个目标文本区域,对初始文本区域进行拆分,以将不同的预测深度值划分至不同的文本区域。通过设置容错机制,实现更加精准的调整。在一种实施方式中,获取待测图像的初始文本区域,包括:

[0082] 预测待测图像中各个像素点属于文本区域的概率;

[0083] 根据待测图像中各个像素点属于文本区域的概率,确定待测图像的初始文本区域。

[0084] 其中,待测图像中各个像素点属于文本区域的概率的数据形式,可以是文本区域概率图的形式。

[0085] 上述实施方式,利用各个像素点属于文本区域的概率,确定属于文本区域的像素点,从而确定初始文本区域。

[0086] 进一步地,根据待测图像中各个像素点属于文本区域的概率,确定待测图像的初始文本区域,包括:对待测图像中各个像素点属于文本区域的概率进行二值化处理,得到文

本区域二值图,根据文本区域二值图求取连通域,得到初始文本区域。例如,各个像素点的文本区域的概率的取值范围为[0,1],取值1表示属于文本区域,取值0表示不属于非文本区域。将取值大于0.5的全部设为1,取值小于等于0.5的全部设为0,得到文本区域二值图。将取值1的像素点连通后所得的连通域,作为初始文本区域。

[0087] 在一种实施方式中,预测待测图像中各个像素点属于文本区域的概率,包括:

[0088] 获取文本区域概率预测模型;

[0089] 将待测图像输入至文本区域概率预测模型,得到待测图像中各个像素点属于文本区域的概率。

[0090] 利用文本区域概率预测模型,有利于提高预测像素点属于文本区域概率的准确度。

[0091] 在一种示例中,文本区域概率预测模块的训练方法,可以包括:获取包含有文本区域的第二样本图像和第二标签;其中,第二标签包括第二样本图像的文本区域信息;获取第二待训练神经网络;根据第二样本图像和第二标签,训练第二待训练神经网络,得到文本区域预测概率模型。

[0092] 其中,参考深度值预测模型的神经网络结构,同样地,第二待训练神经网络可以至少包括依次连接的特征提取网络、特征增强网络、卷积层和反卷积层。此部分的详细内容可以参考前述实施例关于深度值预测模型的神经网络结构的说明,此处不再赘述。

[0093] 基于上述实施例的内容,下述为本申请实施例所提供的的一个示例,该示例将从训练阶段和测试阶段进行说明。

[0094] (1)训练阶段

[0095] 第一步,获取包含文本区域的样本图像、第一标签和第二标签。第一标签包括样本图像中各个像素点的预设深度值,第二标签包括样本图像的文本区域信息。

[0096] 以及,获取待训练神经网络,该待训练神经网络包括Resnet18、两个级联的FPFM模块和两个分支,每个分支均包括卷积层和反卷积层。其中,Resnet18作为基础网络模型,用于对输入的图像提取特征。两个FPFM模块用于对提取的特征进行特征增强处理。将处理后的特征上采样至原始图像的1/4大小后进行串联,串联后的特征分别输入两个分支。每个分支均用于对串联后的特征进行一次卷积和两次反卷积操作。

[0097] 第二步,利用样本图像、第一标签和第二标签,训练待训练神经网络,得到文本检测模型。其中,第一标签用于监督第一个分支的学习,以使第一分支输出像素点深度值图;第二标签用于监督第二个分支的学习,以使第二分支输出文本区域概率图。

[0098] 其中,对于第一个分支输出的像素点深度值图,可以使用L1损失函数进行训练优化。对于第二个分支输出的文本区域概率图,可以使用交并比值(Dice Loss)作为目标损失函数进行训练优化。交并比值 L_{tex} 的计算公式如下:

$$[0099] \quad L_{tex} = 1 - \frac{2 \sum_i P_{tex}(i) G_{tex}(i)}{\sum_i P_{tex}(i)^2 + \sum_i G_{tex}(i)^2}$$

[0100] 其中, $P_{tex}(i)$ 指图像分割结果中的第i个结果, $G_{tex}(i)$ 指第i个结果对应的标签数据。

[0101] (2)测试阶段

[0102] 参考图4,图4所示为测试阶段的流程示例图。

- [0103] 第一步,将文本图像输入Resnet18网络进行特征提取,得到特征1。
- [0104] 第二步,将特征1输入两个FPFM模块进行增强处理,得到特征2。特征2包括4组特征映射。
- [0105] 第三步,将特征2的4组特征映射的特征图大小全部上采样到原始的文本图像1/4大小,并将上采样后的4组特征映射串联在一起。
- [0106] 第四步,将第三步串联的特征映射输入第一分支,通过第一分支进行一次卷积操作,两次反卷积操作,得到像素点预测深度值图。像素点预测深度值图大小与原始的文本图像大小一致。
- [0107] 第五步,将第三步串联的特征映射输入第二分支,通过第二分支进行一次卷积操作,两次反卷积操作,得到初始文本区域概率图。初始文本区域概率图的大小与原始的文本图像大小一致。
- [0108] 第六步,将第四步得到的像素点预测深度值图进行离散化处理,得到像素点预测深度值离散图。
- [0109] 第七步,将第五步得到的初始文本区域概率图进行二值化处理,得到初始文本区域二值图。
- [0110] 第八步,根据第七步得到的初始文本区域二值图求取连通域,得到所有可能文本区域,即初始文本区域。
- [0111] 第九步,根据一个目标文本区域内像素点的深度值都相同这一规定,并行的对第八步得到的每个初始文本区域进行判断,如果一个初始文本区域包括两种或两种以上预测深度值,那么就根据不同预测深度值将这个文本区域划分为相应个数的目标文本区域。
- [0112] 当然,上述过程中,也可以设置容错机制,比如一个目标文本区域内包括两种预测深度值,如果第二种预测深度值对应的像素点数量不超过第一种5%,那么就认为还是一个目标文本区域。
- [0113] 需要说明的是,上述第一步至第九步的步骤顺序仅为一种示例,并非唯一方式。在实际应用过程中,上述部分步骤的顺序可以进行交换或者调整为并行,例如第四步和第五步可以交换顺序或者调整为并行,同样可以达到对应的技术效果。
- [0114] 图5示出根据本发明一实施例的模型训练装置的结构框图。如图5所示,该模型训练装置500可以包括:
- [0115] 样本图像获取模块501,用于获取包含有文本区域的样本图像和标签;其中,标签包括样本图像中各个像素点的预设深度值,并且,样本图像中同一文本区域内像素点的预设深度值相同,样本图像中相邻的不同文本区域内像素点的预设深度值不同;
- [0116] 待训练神经网络获取模块502,用于获取待训练神经网络;
- [0117] 训练模块503,用于根据样本图像和标签,训练待训练神经网络,得到深度值预测模型。
- [0118] 在一种实施方式中,样本图像中非文本区域的像素点的预设深度值与文本区域的像素点的预设深度值不同。
- [0119] 在一种实施方式中,待训练神经网络至少包括依次连接的特征提取网络、特征增强网络、卷积层和反卷积层;
- [0120] 其中,特征提取网络用于对输入待训练神经网络的带有标签的样本图像进行特征

提取;特征增强网络用于对特征提取网络提取的特征进行增强处理;卷积层用于对特征增强网络增强后的特征进行卷积操作;反卷积层用于对卷积层卷积操作后的特征进行反卷积操作。

[0121] 在一种实施方式中,特征提取网络包括残差网络,和/或,特征增强网络包括特征金字塔增强模块。

[0122] 图6示出根据本申请实施例的文本区域检测装置的结构框图。如图6所示,该文本区域检测装置600可以包括:

[0123] 待测图像获取模块601,用于获取待测图像;

[0124] 预测深度值获取模块602,用于将待测图像输入至深度值预测模型,得到待测图像中各个像素点的预测深度值;其中,深度值预测模型为采用本申请实施例所提供的模型训练装置训练所得的模型;

[0125] 目标文本区域确定模块603,用于根据待测图像中各个像素点的预测深度值,确定待测图像中的目标文本区域。

[0126] 在一种实施方式中,该文本区域检测装置600还包括:初始文本区域获取模块;

[0127] 其中,初始文本区域获取模块,用于获取待测图像的初始文本区域;

[0128] 目标文本区域确定模块用于根据待测图像中各个像素点的预测深度值,对初始文本区域进行调整,得到待测图像的目标文本区域。

[0129] 在一种实施方式中,目标文本区域确定模块用于基于文本区域中像素点的预测深度值相同的调整规则,并根据待测图像中各个像素点的预测深度值,对初始文本区域进行调整,得到待测图像的目标文本区域。

[0130] 在一种实施方式中,目标文本区域确定模块,还用于:

[0131] 在初始文本区域中存在至少两种预测深度值的像素点的情况下,基于预测深度值对初始文本区域进行拆分,以使不同预测深度值的像素点位于不同的文本区域,得到待测图像的目标文本区域;

[0132] 或者,

[0133] 在初始文本区域中存在至少两种预测深度值的像素点,且至少两种预测深度值的像素点的占比均大于预设比例的情况下,对初始文本区域进行拆分,以使不同预测深度值的像素点位于不同的文本区域,得到待测图像的目标文本区域。

[0134] 在一种实施方式中,初始文本区域获取模块,包括:

[0135] 文本区域概率预测子模块,用于预测待测图像中各个像素点属于文本区域的概率;

[0136] 初始文本区域确定子模块,用于根据待测图像中各个像素点属于文本区域的概率,确定待测图像的初始文本区域。

[0137] 在一种实施方式中,文本区域概率预测子模块,包括:

[0138] 文本区域概率预测模型获取单元,用于获取文本区域概率预测模型;

[0139] 文本区域概率确定单元,用于将待测图像输入至文本区域概率预测模型,得到待测图像中各个像素点属于文本区域的概率。

[0140] 本发明实施例各装置中的各模块的功能可以参见上述方法中的对应描述,在此不再赘述。

[0141] 图7示出根据本申请一实施例的电子设备的结构框图。如图7所示,该电子设备包括:存储器710和处理器720,存储器710内存储有可在处理器720上运行的计算机程序。处理器720执行该计算机程序时实现上述实施例中的模型训练方法和文本区域检测方法。存储器710和处理器720的数量可以为一个或多个。

[0142] 该电子设备还包括:

[0143] 通信接口730,用于与外界设备进行通信,进行数据交互传输。

[0144] 如果存储器710、处理器720和通信接口730独立实现,则存储器710、处理器720和通信接口730可以通过总线相互连接并完成相互间的通信。该总线可以是工业标准体系结构(Industry Standard Architecture,ISA)总线、外部设备互连(Peripheral Component Interconnect ,PCI)总线或扩展工业标准体系结构(Extended Industry Standard Architecture ,EISA)总线等。该总线可以分为地址总线、数据总线、控制总线等。为便于表示,图7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。

[0145] 可选的,在具体实现上,如果存储器710、处理器720及通信接口730集成在一块芯片上,则存储器710、处理器720及通信接口730可以通过内部接口完成相互间的通信。

[0146] 本申请实施例提供了一种计算机可读存储介质,其存储有计算机程序,该程序被处理器执行时实现本申请实施例中提供的方法。

[0147] 本申请实施例还提供了一种芯片,该芯片包括,包括处理器,用于从存储器中调用并运行存储器中存储的指令,使得安装有芯片的通信设备执行本申请实施例提供的方法。

[0148] 本申请实施例还提供了一种芯片,包括:输入接口、输出接口、处理器和存储器,输入接口、输出接口、处理器以及存储器之间通过内部连接通路相连,处理器用于执行存储器中的代码,当代码被执行时,处理器用于执行申请实施例提供的方法。

[0149] 应理解的是,上述处理器可以是中央处理器(Central Processing Unit,CPU),还可以是其他通用处理器、数字信号处理器(digital signal processing,DSP)、专用集成电路(application specific integrated circuit,ASIC)、现场可编程门阵列(fieldprogrammable gate array,FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者是任何常规的处理器等。值得说明的是,处理器可以是支持进阶精简指令集机器(advanced RISC machines,ARM)架构的处理器。

[0150] 进一步地,可选的,上述存储器可以包括只读存储器和随机存取存储器,还可以包括非易失性随机存取存储器。该存储器可以是易失性存储器或非易失性存储器,或可包括易失性和非易失性存储器两者。其中,非易失性存储器可以包括只读存储器(read-only memory,ROM)、可编程只读存储器(programmable ROM,PROM)、可擦除可编程只读存储器(erasable PROM,EPROM)、电可擦除可编程只读存储器(electrically EPROM,EEPROM)或闪存。易失性存储器可以包括随机存取存储器(random access memory,RAM),其用作外部高速缓存。通过示例性但不是限制性说明,许多形式的RAM可用。例如,静态随机存取存储器(static RAM,SRAM)、动态随机存取存储器(dynamic random access memory ,DRAM)、同步动态随机存取存储器(synchronous DRAM,SDRAM)、双倍数据速率同步动态随机存取存储器(double data rate SDRAM,DDR SDRAM)、增强型同步动态随机存取存储器(enhanced SDRAM,ESDRAM)、同步连接动态随机存取存储器(synchlink DRAM,SLDRAM)和直接内存总线

随机存取存储器(direct rambus RAM,DR RAM)。

[0151] 在上述实施例中,可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。当使用软件实现时,可以全部或部分地以计算机程序产品的形式实现。计算机程序产品包括一个或多个计算机指令。在计算机上加载和执行计算机程序指令时,全部或部分地产生按照本申请的流程或功能。计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。计算机指令可以存储在计算机可读存储介质中,或者从一个计算机可读存储介质向另一个计算机可读存储介质传输。

[0152] 在本说明书的描述中,参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包括于本申请的至少一个实施例或示例中。而且,描述的具体特征、结构、材料或者特点可以在任一个或多个实施例或示例中以合适的方式结合。此外,在不相互矛盾的情况下,本领域的技术人员可以将本说明书中描述的不同实施例或示例以及不同实施例或示例的特征进行结合和组合。

[0153] 此外,术语“第一”、“第二”仅用于描述目的,而不能理解为指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”的特征可以明示或隐含地包括至少一个该特征。在本申请的描述中,“多个”的含义是两个或两个以上,除非另有明确具体的限定。

[0154] 流程图中或在此以其他方式描述的任何过程或方法描述可以被理解为,表示包括一个或更多个用于实现特定逻辑功能或过程的步骤的可执行指令的代码的模块、片段或部分。并且本申请的优选实施方式的范围包括另外的实现,其中可以不按所示出或讨论的顺序,包括根据所涉及的功能按基本同时的方式或按相反的顺序,来执行功能。

[0155] 在流程图中表示或在此以其他方式描述的逻辑和/或步骤,例如,可以被认为是用于实现逻辑功能的可执行指令的定序列列表,可以具体实现在任何计算机可读介质中,以供指令执行系统、装置或设备(如基于计算机的系统、包括处理器的系统或其他可以从指令执行系统、装置或设备取指令并执行指令的系统)使用,或结合这些指令执行系统、装置或设备而使用。

[0156] 应理解的是,本申请的各部分可以用硬件、软件、固件或它们的组合来实现。在上述实施方式中,多个步骤或方法可以用存储在存储器中且由合适的指令执行系统执行的软件或固件来实现。上述实施例方法的全部或部分步骤是可以通程序来指令相关的硬件完成,该程序可以存储于一种计算机可读存储介质中,该程序在执行时,包括方法实施例的步骤之一或其组合。

[0157] 此外,在本申请各个实施例中的各功能单元可以集成在一个处理模块中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个模块中。上述集成的模块既可以采用硬件的形式实现,也可以采用软件功能模块的形式实现。上述集成的模块如果以软件功能模块的形式实现并作为独立的产品销售或使用,也可以存储在一个计算机可读存储介质中。该存储介质可以是只读存储器,磁盘或光盘等。

[0158] 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,可轻易想到其各种变化或替换,这些都应涵盖在本申请的保护范围之内。因此,本申请的保护范围应以权利要求的保护范

围为准。

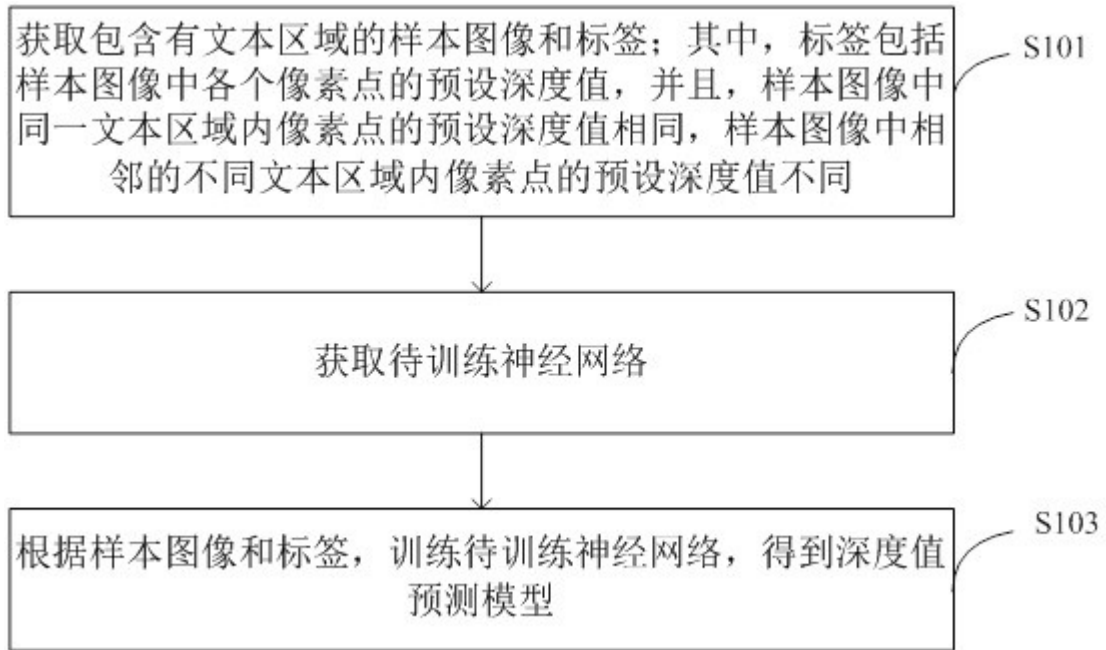


图1

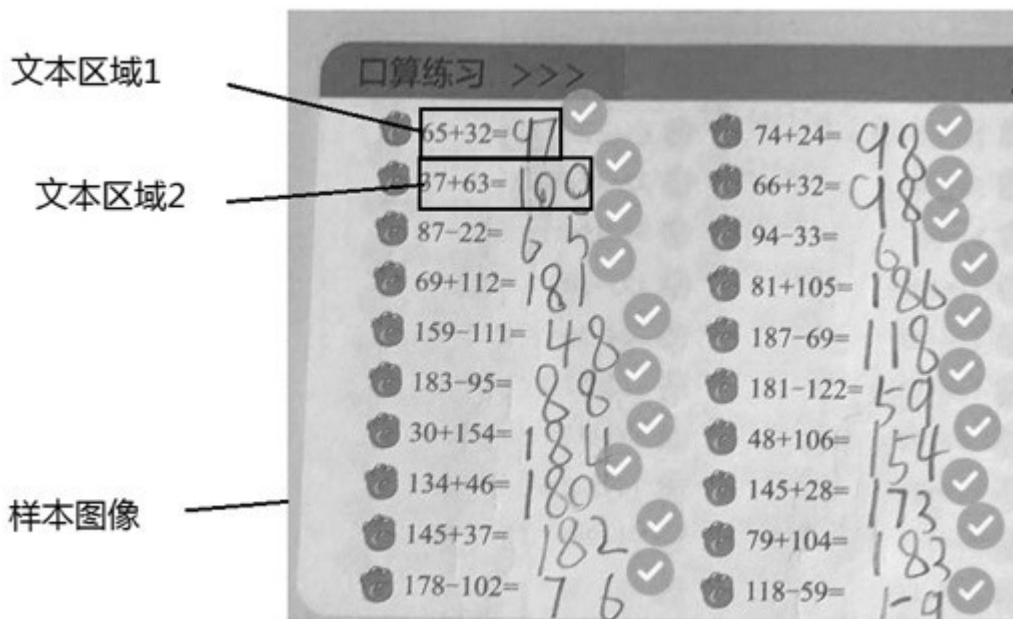


图2



图3

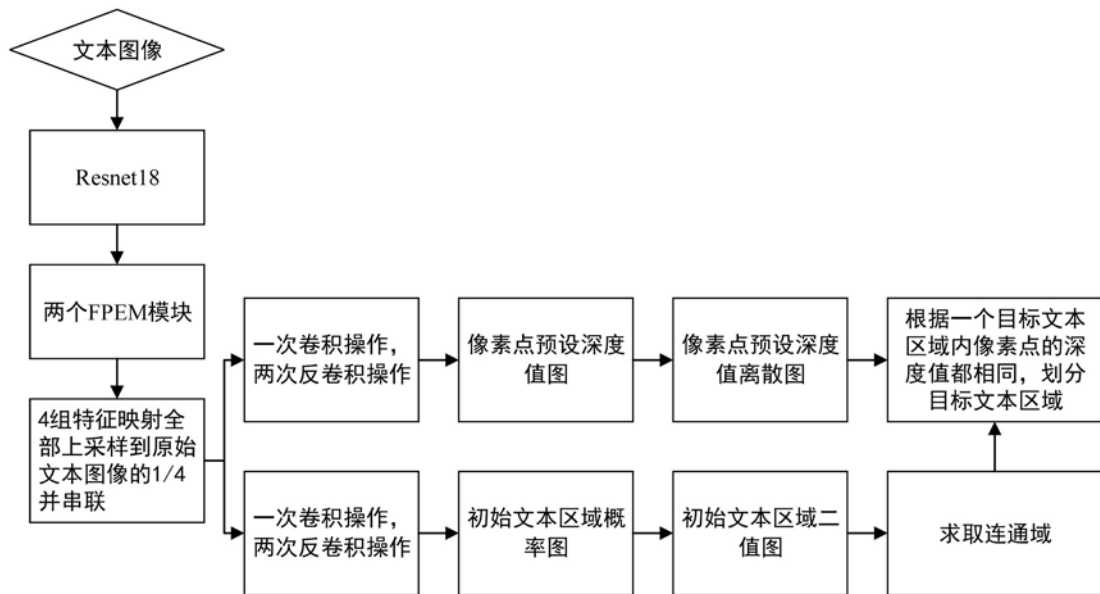


图4

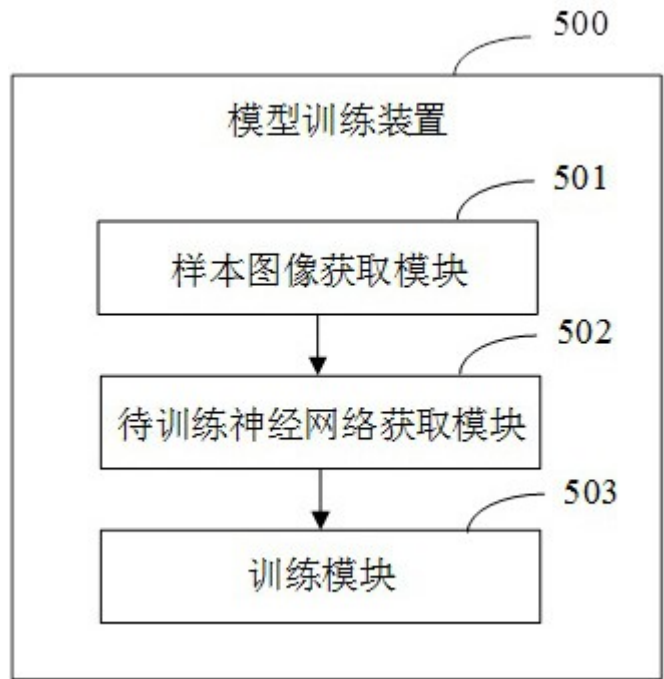


图5

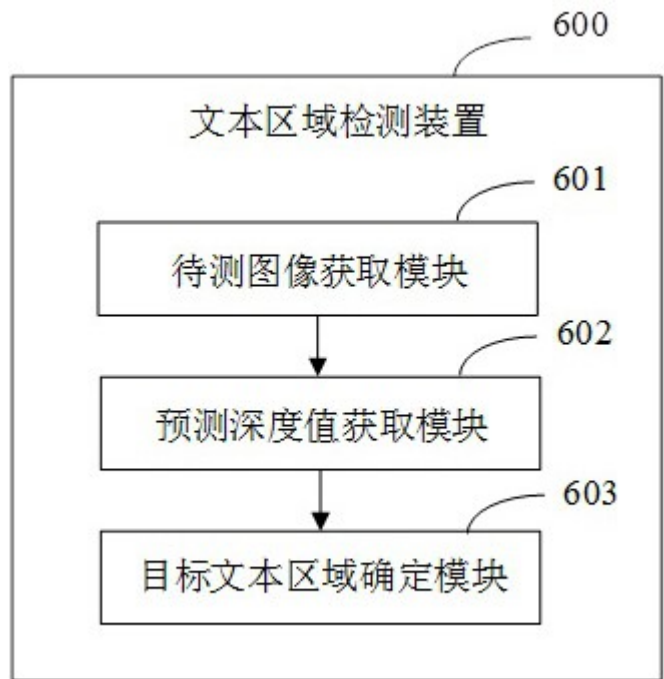


图6

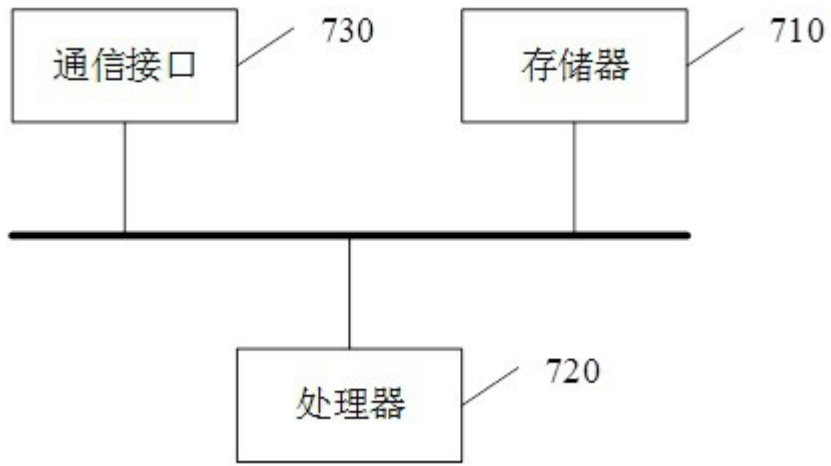


图7