



(12) 发明专利

(10) 授权公告号 CN 110874618 B

(45) 授权公告日 2020.11.27

(21) 申请号 202010057171.3

(22) 申请日 2020.01.19

(65) 同一申请的已公布的文献号
申请公布号 CN 110874618 A

(43) 申请公布日 2020.03.10

(73) 专利权人 同盾控股有限公司
地址 311121 浙江省杭州市余杭区五常街
道文一西路998号18幢704室

(72) 发明人 朱帆 邱君华 李宏宇 李晓林

(74) 专利代理机构 杭州创智卓英知识产权代理
事务所(普通合伙) 33324
代理人 刘丽琴

(51) Int. Cl.
G06K 9/62 (2006.01)
G06N 3/04 (2006.01)

(56) 对比文件

CN 105184265 A, 2015.12.23

CN 107239802 A, 2017.10.10

审查员 韩丹华

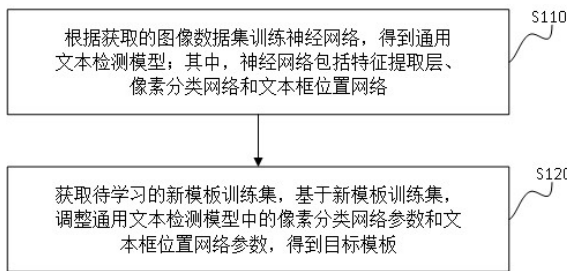
权利要求书2页 说明书10页 附图4页

(54) 发明名称

基于小样本的OCR模板学习方法、装置、电子设备
及介质

(57) 摘要

本发明公开了一种基于小样本的OCR模板学习方法,涉及图像模板处理技术领域,该方法包括以下步骤:根据获取的图像数据集训练神经网络得到通用文本检测模型;其中,神经网络包括特征提取层、像素分类网络和文本框位置网络;获取待学习的新模板训练集,基于新模板训练集调整通用文本检测模型中的像素分类网络参数和文本框位置网络参数,获得目标模板。该方法无需从头训练新的OCR模型,也无需大量的新模板训练数据,基于通用文本检测模型自动迁移学习新模板,提高数据处理效率,减少模型学习时间,提升开发效率,节约开发成本,提高信息管理效率,适用多种新图像模板的定制开发。本发明还公开基于小样本的OCR模板学习装置、电子设备和计算机存储介质。



1. 一种基于小样本的OCR模板学习方法,其特征在于:包括以下步骤:

获取图像数据集,根据所述图像数据集训练神经网络,得到通用文本检测模型,所述通用文本检测模型用于学习各种不同类型的OCR新模板;其中,所述神经网络包括特征提取层、像素分类网络和文本框位置网络;

获取待学习的新模板训练集,固定所述通用文本检测模型中的特征提取层参数,基于所述新模板训练集调整所述通用文本检测模型中的像素分类网络参数和文本框位置网络参数,获得目标模板,其中,所述新模板训练集包括新模板图像、像素分类标签值和文本框位置标签值;

其中,获取待学习的新模板训练集,包括:获取待学习的新模板图像;所述新模板图像的数量大于或等于一张;

基于所述新模板训练集调整所述通用文本检测模型中的像素分类网络参数和文本框位置网络参数,包括:所述新模板图像通过所述通用文本检测模型中的特征提取层计算得到新模板特征;所述新模板特征通过所述通用文本检测模型中的像素分类网络和文本框位置网络计算,分别得到第二置信度预测值和第二位置预测值;根据像素分类损失函数对所述第二置信度预测值和所述像素分类标签值进行计算得到分类损失,根据所述分类损失和误差反传算法调整所述像素分类网络参数;根据文本框位置损失函数对所述第二位置预测值和所述文本框位置标签值进行计算得到位置损失,根据所述位置损失和误差反传算法调整所述文本框位置网络参数,其中,所述文本框位置损失函数包括文本框位置边框损失函数和中心点距离损失函数;

根据文本框位置损失函数对所述第二位置预测值和所述文本框位置标签值进行计算得到位置损失,包括:根据所述第二位置预测值和所述文本框位置标签值得到区域交集和区域并集,根据所述文本框位置边框损失函数对所述区域交集和所述区域并集进行除法计算得到边框损失;根据所述第二位置预测值和所述文本框位置标签值得到两个中心坐标,根据所述中心点距离损失函数对所述两个中心坐标进行计算得到中心坐标损失;将所述边框损失和所述中心坐标损失相加得到所述位置损失。

2. 如权利要求1所述的一种基于小样本的OCR模板学习方法,其特征在于:所述图像数据集包括文本图像数据和文本模板数据,根据所述图像数据集训练神经网络,得到通用文本检测模型,包括:

通过所述特征提取层对所述文本图像数据计算,得到图像特征;

通过所述像素分类网络和所述文本框位置网络对所述图像特征计算,分别得到第一置信度预测值和第一位置预测值;

根据所述第一置信度预测值、所述第一位置预测值和所述文本模板数据训练所述神经网络,得到所述通用文本检测模型。

3. 如权利要求1所述的一种基于小样本的OCR模板学习方法,其特征在于:获取待学习的新模板训练集,还包括:

获取所述新模板图像待提取的文本区域;

根据预设的标注规则,计算所述文本区域的位置坐标,根据所述位置坐标,得到像素分类标签值和文本框位置标签值。

4. 如权利要求1-3任一项所述的一种基于小样本的OCR模板学习方法,其特征在于:获

得目标模板,包括:

当参数调整达到预设的调整目标时,得到完成调整的文本检测模型,根据调整后的所述文本检测模型,得到目标置信度和目标位置,根据所述目标置信度和所述目标位置,生成所述目标模板。

5. 一种基于小样本的OCR模板学习装置,其特征在于,其包括:

通用文本模型建立模块,用于获取图像数据集,根据所述图像数据集训练神经网络,得到通用文本检测模型,所述通用文本检测模型用于学习各种不同类型的OCR新模板;其中,所述神经网络包括特征提取层、像素分类网络和文本框位置网络;

目标模板学习模块,用于获取待学习的新模板训练集,固定所述通用文本检测模型中的特征提取层参数,基于所述新模板训练集调整所述通用文本检测模型中的像素分类网络参数和文本框位置网络参数,获得目标模板,其中,所述新模板训练集包括新模板图像、像素分类标签值和文本框位置标签值;其中,获取待学习的新模板训练集,包括:获取待学习的新模板图像;所述新模板图像的数量大于或等于一张;基于所述新模板训练集调整所述通用文本检测模型中的像素分类网络参数和文本框位置网络参数,包括:所述新模板图像通过所述通用文本检测模型中的特征提取层计算得到新模板特征;所述新模板特征通过所述通用文本检测模型中的像素分类网络和文本框位置网络计算,分别得到第二置信度预测值和第二位置预测值;根据像素分类损失函数对所述第二置信度预测值和所述像素分类标签值进行计算得到分类损失,根据所述分类损失和误差反传算法调整所述像素分类网络参数;根据文本框位置损失函数对所述第二位置预测值和所述文本框位置标签值进行计算得到位置损失,根据所述位置损失和误差反传算法调整所述文本框位置网络参数,其中,所述文本框位置损失函数包括文本框位置边框损失函数和中心点距离损失函数;根据文本框位置损失函数对所述第二位置预测值和所述文本框位置标签值得到区域交集和区域并集,根据所述文本框位置边框损失函数对所述区域交集和所述区域并集进行除法计算得到边框损失;根据所述第二位置预测值和所述文本框位置标签值得到两个中心坐标,根据所述中心点距离损失函数对所述两个中心坐标进行计算得到中心坐标损失;将所述边框损失和所述中心坐标损失相加得到所述位置损失。

6. 一种电子设备,其包括处理器、存储介质以及计算机程序,所述计算机程序存储于存储介质中,其特征在于,所述计算机程序被处理器执行时执行权利要求1至4任一项所述的基于小样本的OCR模板学习方法。

7. 一种计算机存储介质,其上存储有计算机程序,其特征在于:所述计算机程序被处理器执行时实现权利要求1至4任一项所述的基于小样本的OCR模板学习方法。

基于小样本的OCR模板学习方法、装置、电子设备及介质

技术领域

[0001] 本发明涉及图像模板处理技术领域,尤其涉及一种基于小样本的OCR模板学习方法、装置、电子设备及存储介质。

背景技术

[0002] OCR(Optical Character Recognition,光学字符识别)技术是指对输入图像进行分析识别处理,获取图像中文本信息的过程,是当前应用最为广泛的人工智能技术之一。自动化提取固定版式的票据、卡证、文件等图片OCR模板,可应用于纸质文件分类归档、信息统计分析、关键内容提取等众多场景,能够有效地降低人工录入成本,提升信息管理效率。

[0003] 然而随着OCR应用场景的日益增多,通用的OCR模板已经无法满足各种不同的应用场景对特定版式的OCR结构化需求。在对不同版式的证件、金融票据、医疗单据进行OCR处理时,由于这些不同类型图片中所包含的内容、排版、样式等信息差异较大,采用通用的OCR模板进行文本检测,获取到的检测结果往往不够理想。因此需要使用不同的OCR模板,用来检测对应的图片版式,但是由于个人信息隐私保护、数据传输安全等原因,实际能够获取的版式图像数据量通常是很有有限的,而少量的图像数据无法支撑OCR深度模型从头开始训练,也就无法通过模型获得OCR目标模板。

发明内容

[0004] 为了克服现有技术的不足,本发明的目的之一在于提供一种基于小样本的OCR模板学习方法,其基于待学习的新模板,调整通用文本检测模型中的像素分类网络参数和文本框位置网络参数,从而获得目标模板。

[0005] 本发明的目的之一采用以下技术方案实现:

[0006] 获取图像数据集,根据所述图像数据集训练神经网络,得到通用文本检测模型;其中,所述神经网络包括特征提取层、像素分类网络和文本框位置网络;

[0007] 获取待学习的新模板训练集,固定所述通用文本检测模型中的特征提取层参数,基于所述新模板训练集调整所述通用文本检测模型中的像素分类网络参数和文本框位置网络参数,获得目标模板。

[0008] 进一步地,所述图像数据集包括文本图像数据和文本模板数据,根据所述图像数据集训练神经网络,得到通用文本检测模型,包括:

[0009] 通过所述特征提取层对所述文本图像数据计算,得到图像特征;

[0010] 通过所述像素分类网络和所述文本框位置网络对所述图像特征计算,分别得到第一置信度预测值和第一位置预测值;

[0011] 根据所述第一置信度预测值、所述第一位置预测值和所述文本模板数据训练所述神经网络,得到所述通用文本检测模型。

[0012] 进一步地,获取待学习的新模板训练集,包括:

[0013] 获取待学习的新模板图像与所述新模板图像待提取的文本区域;

[0014] 根据预设的标注规则,计算所述文本区域的位置坐标,根据所述位置坐标,得到像素分类标签值和文本框位置标签值,所述新模板训练集包括所述新模板图像、所述像素分类标签值和所述文本框位置标签值。

[0015] 进一步地,基于所述新模板训练集调整所述通用文本检测模型中的像素分类网络参数和文本框位置网络参数,包括:

[0016] 所述新模板图像通过所述通用文本检测模型中的特征提取层计算,得到新模板特征;

[0017] 所述新模板特征通过所述通用文本检测模型中的像素分类网络和文本框位置网络计算,分别得到第二置信度预测值和第二位置预测值;

[0018] 根据所述第二置信度预测值、所述第二位置预测值、所述像素分类标签值和所述文本框位置标签值,调整所述像素分类网络参数和所述文本框位置网络参数。

[0019] 进一步地,调整所述像素分类网络参数和所述文本框位置网络参数,包括:

[0020] 根据像素分类损失函数,对所述第二置信度预测值和所述像素分类标签值进行计算,得到分类损失,根据所述分类损失和误差反传算法,调整所述像素分类网络参数;

[0021] 根据文本框位置损失函数,对所述第二位置预测值和所述文本框位置标签值进行计算,得到位置损失,根据所述位置损失和误差反传算法,调整所述文本框位置网络参数。

[0022] 进一步地,所述文本框位置损失函数包括文本框位置边框损失函数和中心点距离损失函数,对所述第二位置预测值和所述文本框位置标签值进行计算,得到位置损失,包括:

[0023] 根据所述第二位置预测值和所述文本框位置标签值,得到区域交集和区域并集,根据所述文本框位置边框损失函数,对所述区域交集和所述区域并集进行除法计算,得到边框损失;

[0024] 根据所述第二位置预测值和所述文本框位置标签值,得到两个中心坐标,根据所述中心点距离损失函数,对所述两个中心坐标进行计算,得到中心坐标损失;

[0025] 将所述边框损失和所述中心坐标损失相加,得到所述位置损失。

[0026] 进一步地,获得目标模板,包括:

[0027] 当参数调整达到预设的调整目标时,得到完成调整的文本检测模型,根据调整后的所述文本检测模型,得到目标置信度和目标位置,根据所述目标置信度和所述目标位置,生成所述目标模板。

[0028] 本发明的目的之二在于提供一种基于小样本的OCR模板学习装置,其基于待学习的新模板,调整通用文本检测模型中的像素分类网络参数和文本框位置网络参数,从而获得目标模板。

[0029] 本发明的目的之二采用以下技术方案实现:

[0030] 一种基于小样本的OCR模板学习装置,其包括:

[0031] 通用文本模型建立模块,用于获取图像数据集,根据所述图像数据集训练神经网络,得到通用文本检测模型;其中,所述神经网络包括特征提取层、像素分类网络和文本框位置网络;

[0032] 目标模板学习模块,用于获取待学习的新模板训练集,固定所述通用文本检测模型中的特征提取层参数,基于所述新模板训练集调整所述通用文本检测模型中的像素分类

网络参数和文本框位置网络参数,获得目标模板。

[0033] 本发明的目的之三在于提供执行发明目的之一的电子设备,其包括处理器、存储介质以及计算机程序,所述计算机程序存储于存储介质中,所述计算机程序被处理器执行时本发明目的之一的基于小样本的OCR模板学习方法。

[0034] 本发明的目的之四在于提供存储发明目的之一的计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时实现本发明目的之一的基于小样本的OCR模板学习方法。

[0035] 相比现有技术,本发明的有益效果在于:

[0036] 本发明将通用文本检测模型作为基础模型,模型结构和特征提取层参数固定不变,根据待学习的新模板训练集数据,自动调整像素分类网络参数和文本框位置网络参数,无需从头训练新的OCR模型,也无需大量的新模板训练集数据,即可获得定制的目标模板,用以检测新的图片版式;该方法基于用户需要的新模板数据,通用文本检测模型自动迁移学习新模板,有效提高处理分析新模板数据的效率,减少模型学习时间,提升开发效率,节约开发成本,提高信息管理效率,适用于多种新图像模板的定制开发。

附图说明

[0037] 图1为本发明实施例一的基于小样本的OCR模板学习方法流程图;

[0038] 图2为本发明实施例二的新模板训练集的获取流程图;

[0039] 图3为本发明实施例三的目标模板的学习流程图;

[0040] 图4为本发明实施例四的票据图像的示意图;

[0041] 图5为本发明实施例四的票据模板学习流程图;

[0042] 图6为本发明实施例五的订单图像的示意图;

[0043] 图7为本发明实施例五的订单模板学习流程图;

[0044] 图8为本发明实施例六的基于小样本的OCR模板学习装置的结构框图;

[0045] 图9为本发明实施例七的电子设备的结构框图。

具体实施方式

[0046] 以下将结合附图,对本发明进行更为详细的描述,需要说明的是,以下参照附图对本发明进行的描述仅是示意性的,而非限制性的。各个不同实施例之间可以进行相互组合,以构成未在以下描述中示出的其他实施例。

[0047] 实施例一

[0048] 实施例一提供了一种基于小样本的OCR模板学习方法,旨在将通用文本检测模型作为基础模型,根据待学习的新模板训练集数据,自动调整像素分类网络参数和文本框位置网络参数,从而获得目标模板。该模板学习方法将通用文本检测模型作为基础模型,模型结构和特征提取层参数固定不变,无需从头训练新的OCR模型,也无需大量的新模板训练集数据,基于通用文本检测模型自动迁移学习新模板数据,即可获得定制的目标模板,用以检测新的图片版式。这种方法能够减少模型学习新模板数据的时间,有效提高处理分析新模板数据的效率,节约开发成本,提升开发效率,能够为各种图片版式提供定制模板。

[0049] 请参照图1所示,一种基于小样本的OCR模板学习方法,包括以下步骤:

[0050] S110、获取图像数据集,根据图像数据集训练神经网络,得到通用文本检测模型;其中,神经网络包括特征提取层、像素分类网络和文本框位置网络。

[0051] 图像数据集包括文本图像数据和文本模板数据。文本图像数据可以从已有的图像数据库中获取的文本图像数据,也可以是即时采集的文本图像数据,在此不做限定。即时采集的文本图像可以通过拍照、扫描等方式获取,比如通过具有自动调整拍摄角度功能的高清拍摄电子设备采集图像。将图像数据作为神经网络的输入,以训练神经网络模型。

[0052] 文本模板数据是对图像数据标注得到的标签数据。标注每个图像数据的文本模板数据,包括图像中文本框的位置坐标和根据标注的文本框得到相应的像素标签值,将该位置坐标和该像素标签值作为神经网络输入的标签值,用以训练神经网络模型。

[0053] 本实施例中,将全卷积神经网络作为训练对象。将图像数据输入全卷积神经网络中的特征提取层进行计算,得到图像特征。特征提取层采用具有特征提取功能的网络结构,在本实施例中,采用基础卷积神经网络VGG16中的前五层卷积结构,作为特征提取层。VGG16卷积神经网络是牛津大学VGG组提出的,VGG16卷积神经网络分为16层,由13个卷积层加上3个全连接层叠加而成,采用连续的几个3x3的卷积核代替较大卷积核,在保证具有相同感知野的条件下,提升网络的深度,在一定程度上提升了神经网络的效果。

[0054] 全卷积神经网络包括两个分支网络,两个分支网络共享特征提取层输出的图像特征。通过像素分类网络和文本框位置网络,分别对图像特征进行像素点的分类和文本框的位置回归。由像素分类网络对所有像素点的图像特征进行逐个分类,判断每个像素点是否为文本,输出每个像素点属于文本的置信度值,由所有像素点的置信度值构成第一置信度预测值。由文本框位置网络对文本框位置进行回归计算,输出文本框的四个顶点坐标,将顶点坐标作为第一位置预测值。

[0055] 图像数据经过全卷积神经网络计算,得到第一置信度预测值和第一位置预测值。根据第一置信度预测值、第一位置预测值和图像数据对应标注的文本模板数据,得到损失误差,将损失误差反向传播到全卷积神经网络的特征提取层,基于梯度下降法训练全卷积神经网络,得到通用文本检测模型。基于该通用文本检测模型能够迁移学习新模板,无需大量的新模板训练数据,即可获得定制的目标模板。该通用文本检测模型可用于学习各种不同类型的OCR新模板。

[0056] S120、获取待学习的新模板训练集,固定通用文本检测模型中的特征提取层参数,基于新模板训练集调整通用文本检测模型中的像素分类网络参数和文本框位置网络参数,获得目标模板。

[0057] 根据实际场景中用户对OCR模板的需求,获取需要检测的新模板图像。新模板图像可以从用户的图像数据库中获取的,也可以是按照用户需求采集的图像,在此不做限定。通过OCR检测方法获取新模板图像中的文本区域。

[0058] 为了从新模板图像中获得用户需要的结构化信息,根据预设的标注规则,比如标注字段、字段标注格式、标注文件输出类型等规则对新模板图像进行标注,只提取用户所需的关键模板信息,提高了数据处理效率。

[0059] 根据标注规则,计算文本区域中的待检测文本框的坐标位置,将待检测文本框的最小外接矩形的四个顶点坐标,作为新模板图像的文本框位置标签值。根据文本框的最小外接矩形,计算图像中所有像素点的分类标签值,作为像素分类标签值。优选地,对图像中

的像素点进行二分类,像素点的像素值分为非零值和零值,由所有像素点的像素值构成像素分类标签值。

[0060] 新模板训练集包括但不限于新模板图像、像素分类标签值和文本框位置标签值等新模板信息,并将新模板信息按照预设的标注规则输出。新模板训练集包括了用户需要的OCR新模板的结构化信息,只需获取数量大于或等于一张的新模板图像,就可以实现基于通用文本检测模型的迁移学习,不受新模板数量的限制,并且减少了学习时间,提升OCR新模板的开发效率。

[0061] 基于通用文本检测模型迁移学习新模板训练集,将通用文本检测模型作为基础模型,通用文本检测模型的模型结构不变,同时固定通用文本检测模型中的特征提取层参数,仅仅调整通用文本检测模型中的像素分类网络和文本框位置网络的网络参数。当像素分类网络参数和文本框位置网络参数达到预设的调整目标时,两个网络参数的调整完成,获得调整后的文本检测模型。新模板数据的学习无需从头训练新的OCR模型,减少了模型训练时间,提高新模板的开发效率。

[0062] 调整目标可以是预设的调整次数阈值,当参数的调整次数达到调整次数阈值时,就得到完成调整的文本检测模型;调整目标也可以是预设的变化范围阈值,当参数的变化趋于稳定,在最近的N次变化中参数的最大值与最小值之差小于或等于变化范围阈值时,完成模型参数的调整,最近的变化次数N根据用户需求决定。调整目标根据用户实际需要预设,在此不做限定。

[0063] 通过调整后的文本检测模型中的像素分类网络和文本框位置网络,分别得到目标置信度和目标位置。由目标置信度和目标位置构成目标模板,就可以确定根据该目标位置定位的文本框,并根据目标置信度判断该文本框是否属于待检测文本。使得调整后的文本检测模型,能够准确地检测到新模板图像中的待检测文本,不检测用户不关注的文本内容,从而提高数据管理效率。

[0064] 当需要学习多类OCR新模板时,可以共享通用文本检测模型的模型结构和特征提取层参数,通过对像素分类网络参数和文本框位置网络参数的自动调整,完成迁移学习,获得定制的目标模板。无需对每类新模板从头训练新的OCR模型,缩短模板的开发周期,节约开发成本。

[0065] 实施例二

[0066] 实施例二是在实施例一基础上进行的改进,对获取的新模板图像,根据预设的标注规则,计算图像对应的标签值,获得新模板训练集。只要提取用户需要的关键模板数据,就能得到学习目标模板所需的结构化信息,进行通用文本检测模型的迁移学习,提高了数据处理效率。

[0067] 标注规则包括但不限于标注字段、字段标注格式、标注文件输出类型等信息。本实施例中,标注规则为标注字段、字段标注格式、输出类型。新模板训练集的获取,请参照图2所示,包括以下步骤:

[0068] S210、获取图像中文本区域的标注字段。标注字段用户关注的图像信息,是在相同模板的不同图像中,位置固定不变的字段,比如常用发票模板中的“货物或应税劳务、服务名称”、“金额”、“税额”等字段。

[0069] S220、计算文本框位置标签值。分别将图像的下边沿、左边沿作为X轴、Y轴,建立平

面坐标图。根据文本区域在图像中所在的位置,计算文本区域最小外接矩形的四个顶点坐标,将该顶点坐标作为图像对应的文本框位置标签值。

[0070] S230、计算像素分类标签值。本实施例中,对所有像素点进行二分类,文本图像中有标注字段的区域对应像素值非零,非标注区域对应像素值为零。计算文本区域中的所有像素点的标签值,将所有像素点对应的标签值作为像素分类标签值。

[0071] S240、按照字段标注格式和输出类型,输出标注文件。字段标注格式为“标注字段:文本框位置标签值;像素分类标签值”,其中文本框位置标签值对应标注字段所在文本区域的位置坐标,像素分类标签值对应文本区域中所有像素点的分类标签值。

[0072] 由单张或多张同类模板图像获得的新模板训练集,新模板训练集包括图像数据和标注文件。标注文件的输出类型包括但不限于txt、json、csv等计算机程序可读文件,便于通用文本检测模型读取新模板训练集,进而进行迁移学习。

[0073] 实施例三

[0074] 实施例三是在实施例一或/和实施例二的基础上进行的改进,根据新模板图像,调整通用文本检测模型中的像素分类网络和文本框位置网络的网络参数,进而获得调整后的文本检测模型。使得对目标模板的学习,无需从模型结构构造或特征提取开始,缩短了新模板的开发周期,提高开发效率,降低开发成本。

[0075] 请参照图3所示,目标模板的学习包括以下步骤:

[0076] S310、计算新模板图像对应的第二预测值。

[0077] 固定通用文本检测模型的模型结构和特征提取层的参数,新模板图像通过通用文本检测模型中的特征提取层计算,得到新模板特征。新模板特征通过通用文本检测模型中的像素分类网络和文本框位置网络计算,分别得到第二置信度预测值和第二位置预测值。两个预测值分别用于调整像素分类网络和文本框位置网络的网络参数。

[0078] S320、根据像素分类损失函数,计算分类损失。

[0079] 对第二置信度预测值和像素分类标签值进行正则化计算,将正则化计算后的第二置信度预测值和像素分类标签值,输入像素分类损失函数,得到分类损失。分类损失统计的是像素分类网络输出的第二置信度预测值与对应图像的分类标签之间的误差,用以调整像素分类网络参数。像素分类损失函数不限于L1损失函数、L2损失函数、Huber损失函数中的一种。本实施例中,像素分类损失函数为L1损失函数。

[0080] S330、根据文本框位置损失函数,计算位置损失。

[0081] 根据文本框位置损失函数,对第二位置预测值和文本框位置标签值进行计算,得到位置损失。根据第二位置预测值,获得相应的预测文本区域S1,根据文本框位置标签值,获得相应的图像中的真实文本区域S0。优选地,文本框位置损失函数包括文本框位置边框损失函数和中心点距离损失函数。

[0082] 计算预测文本区域S1和真实文本区域S0的区域交集和区域并集,将区域交集和区域并集输入文本框位置边框损失函数,进行除法计算,得到边框损失。文本框位置边框损失函数不限于IoU损失函数、GIoU损失函数中的一种。本实施例中,采用IoU损失函数计算边框

$$\text{损失: } L_{IoU} = 1 - \frac{|S1 \cap S0|}{|S1 \cup S0|}。$$

[0083] 分别计算预测文本区域S1的中心坐标M1和真实文本区域S0的中心坐标M0,将中心

坐标M1 (x_{pred}, y_{pred}) 和中心坐标M0 (x_{gt}, y_{gt}) 输入中心点距离损失函数进行计算,得到中心坐标损失,其中, x_{pred}, y_{pred} 为预测文本区域S1的中心坐标M1的预测值, x_{gt}, y_{gt} 为真实文本区域S0的中心坐标M0的标签值。中心坐标损失函数不限于L1损失函数、L2损失函数、Huber损失函数中的一种。本实施例中,采用L2损失函数计算中心坐标损失: $L(M1, M0) = (x_{pred} - x_{gt})^2 + (y_{pred} - y_{gt})^2$ 。

[0084] 将边框损失和中心坐标损失相加,得到位置损失。位置损失度量的是预测文本区域与真实文本区域之间的几何误差,用以调整文本框位置网络参数。

[0085] S340、根据分类损失误差和位置损失误差,调整两个网络的参数,根据调整后的文本检测模型,得到目标模板。

[0086] 分别将分类损失和位置损失反向传播到像素分类网络和文本框位置网络中,进行多次迭代,每次迭代更新两个分支网络的网络参数,使得两个分支网络不断降低这两部分损失误差,使得位置损失度量的几何误差越来越小,进而确保文本框位置网络能够准确地检测到待检测文本的所在位置,使得分类损失统计的误差越来越小,确保像素分类网络能够回归出待检测文本。

[0087] 当分类损失和位置损失同时达到预设的调整目标时,参数调整完成,获得调整后的文本检测模型。调整目标可以是两个损失的值均达到损失误差阈值,也可以是两个损失的迭代次数均达到迭代次数阈值,还可以是两个损失的值的变化范围小于变化范围阈值,但不限于上述三种调整目标。

[0088] 根据调整后的文本检测模型,获得用户需要的对应目标模板。根据目标模板能够准确地检测到用户关注的文本内容,不对不关注的无用文本内容进行检测,提高了信息管理效率。

[0089] 当需要多类OCR新模板时,同样可以使用通用文本检测模型的模型结构和特征提取层参数,只要调整像素分类网络和文本框位置网络的网络参数,就可以获得相应的目标模板。通用文本检测模型自动学习新模板,不需要从头训练新的OCR模型,节约了模板开发成本。

[0090] 实施例四

[0091] 实施例四是对实施例二的应用说明,现有用户需要新增一个OCR票据模板,提供如图4所示的票据图像,用户关注票据中对应“货物或应税劳务、服务名称”、“金额”、“税额”的各项字段内容。当前没有通用文本检测模型,请参照图5所示,该票据模板学习包括以下步骤:

[0092] S410、利用图像数据集训练全卷积神经网络,得到通用文本检测模型。

[0093] 图像数据集包括文本图像数据和文本模板数据,文本图像数据是从已有的图像数据库中获取的,文本模板数据由该文本图像数据标注得到。将文本图像数据输入全卷积神经网络中的特征提取层进行计算,得到图像特征。图像特征经过两个分支网络计算,得到第一置信度预测值和第一位置预测值。

[0094] 根据第一置信度预测值、第一位置预测值和文本模板数据,训练全卷积神经网络,得到通用文本检测模型detect.model。该通用文本检测模型可用于学习多种新的OCR模板。

[0095] S420、根据预设的标注规则,获取票据图像的标注文件。

[0096] 标注规则为标注字段、字段标注格式,其中,标注字段为用户关注的“货物或应税

劳务、服务名称”、“金额”、“税额”，字段标注格式为“字段名称:文本框位置标签值”，文本框位置标签值为字段所在的文本区域最小外接矩形的四个顶点坐标。图4所示的单张图像对应的标注文件为：

[0097] 货物或应税劳务、服务名称:391,893,842,893,391,983,842,983;

[0098] 金额:2705,903,2816,903,2705,995,2816,995;

[0099] 税额:3001,901,3187,901,3001,990,3187,990;

[0100] 其中,第一行表示”货物或应税劳务、服务名称”对应文本所在图像中的所外最小外接矩形的四个顶点在图像中的像素点坐标为[(391,893),(842,893),(391,983),(842,983)]。第二、三行和第一行的表示含义相同。

[0101] S430、迁移学习得到票据图像的OCR模板。

[0102] 固定步骤S410中的通用文本检测模型detect.model的模型结构和特征提取层参数,根据如图4所示的图像数据和步骤S420中的标注数据,调整通用文本检测模型detect.model中的像素分类网络参数和文本框位置网络参数。当参数调整的次数达到调整次数阈值时,通过调整后的文本检测模型,获得如图4所示的此类票据图像的OCR模板。

[0103] 实施例五

[0104] 实施例五是在实施例三的应用说明,现有用户需要新增一个OCR订单模板,该订单模板包括但不限于网约车订单、餐饮订单、酒店订单的模板,获取用户提供的如图6所示的订单图像,用户关注订单图像中对应“订单编号”、“支付时间”的各项字段内容。如果当前已有通用文本检测模型,如实施例四中的通用文本检测模型detect.model,请参照图7所示,该订单模板学习包括以下步骤:

[0105] S510、根据预设的标注规则,获取订单图像的标注文件。

[0106] 标注规则为标注字段、字段标注格式,其中,标注字段为用户关注的“订单编号”、“支付时间”,字段标注格式为“字段名称:文本框位置标签值;像素分类标签值”,文本框位置标签值为字段所在的文本区域最小外接矩形的四个顶点坐标。图6所示的单张图像对应的标注文件为:

[0107] 订单编号:100,802,436,802,100,983,842,983;像素分类标签值X1;

[0108] 支付时间:100,913,436,913,100,1015,913,1015;像素分类标签值X2;

[0109] 其中,第一行表示“订单编号”对应文本所在图像中的所外最小外接矩形的四个顶点在图像中的像素点坐标为[(391,893),(842,893),(391,983),(842,983)],根据四个顶点坐标获得订单真实区域。第二行和第一行的表示含义相同,通用获得支付真实区域。像素分类标签值X1、像素分类标签值X2分别对“订单编号”、“支付时间”对应的最小外接矩形中所有像素点的标签值,在此不一一写出。

[0110] S520、计算订单图像的分类损失和位置损失。

[0111] 将如图6所示的的订单图像输入detect.model,经过特征提取层计算,得到订单特征。订单特征通过像素分类网络和文本框位置网络计算,得到第二置信度预测值和第二位置预测值。根据第二位置预测值,获得相应的订单预测区域和支付预测区域。

[0112] 对第二置信度预测值和步骤S510中获得像素分类标签值X1、像素分类标签值X2正则化后,输入L1损失函数,得到分类损失。

[0113] 计算订单预测区域和订单真实区域的区域交集和区域并集,计算支付预测区域和

支付真实区域的区域交集和区域并集,将两组区域交集和区域并集输入IoU损失函数计算,得到相应的订单边框损失和支付边框损失,将两个边框损失相加,得到该订单图像的边框损失。

[0114] 计算订单预测区域的中心坐标M11和支付预测区域的中心坐标M12,采用L2损失函数,计算中心坐标M11与订单真实区域的中心坐标M01的订单中心损失,以及中心坐标M12与订单真实区域的中心坐标M02的支付中心损失。将订单中心损失和支付中心损失相加,得到该订单图像的中心坐标损失。将中心坐标损失与该订单图像的边框损失相加,得到位置损失。

[0115] S530、迁移学习得到订单图像的OCR模板。

[0116] 分别将分类损失和位置损失反向传播到detect.model的像素分类网络和文本框位置网络中,进行多次迭代,每次迭代更新两个分支网络的网络参数。当两部分损失的值均达到预设的误差阈值时,获得调整后的文本检测模型,同时得到如图6所示的此类订单图像的OCR模板。

[0117] 订单图像的OCR模板无需从头训练新的OCR模型,只要将通用文本检测模型detect.model作为基础模型,调整detect.model中的两个网络的参数,就能学习该订单图像,得到相应的OCR模板。

[0118] 实施例六

[0119] 实施例六公开了一种对应上述实施例的一种基于小样本的OCR模板学习装置,为上述实施例的虚拟装置结构,请参照图8所示,包括:

[0120] 通用文本模型建立模块610,用于获取图像数据集,根据所述图像数据集训练神经网络,得到通用文本检测模型;其中,所述神经网络包括特征提取层、像素分类网络和文本框位置网络;

[0121] 目标模板学习模块620,用于获取待学习的新模板训练集,固定所述通用文本检测模型中的特征提取层参数,基于所述新模板训练集调整所述通用文本检测模型中的像素分类网络参数和文本框位置网络参数,获得目标模板。

[0122] 实施例七

[0123] 图9为本发明实施例七提供的一种电子设备的结构示意图,如图9所示,该电子设备包括处理器710、存储器720、输入装置730和输出装置740;计算机设备中处理器710的数量可以是一个或多个,图9中以一个处理器710为例;电子设备中的处理器710、存储器720、输入装置730和输出装置740可以通过总线或其他方式连接,图9中以通过总线连接为例。

[0124] 存储器720作为一种计算机可读存储介质,可用于存储软件程序、计算机可执行程序以及模块,如本发明实施例中的基于小样本的OCR模板学习方法对应的程序指令/模块(例如,基于小样本的OCR模板学习装置中的通用文本模型建立模块610和目标模板学习模块620)。处理器710通过运行存储在存储器720中的软件程序、指令以及模块,从而执行电子设备的各种功能应用以及数据处理,即实现上述实施例一至实施例五的基于小样本的OCR模板学习方法。

[0125] 存储器720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、至少一个功能所需的应用程序;存储数据区可存储根据终端的使用所创建的数据等。此外,存储器720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、或其他非易失性固态存储器件。在一些实例中,存储器720可进一步包括相对于处理器710远程设置的存储器,这些远程存储器可以通过网络连接至电子设备。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

[0126] 输入装置730可用于接收图像数据集、新模板训练集等。输出装置740可包括显示屏等显示设备。

[0127] 实施例八

[0128] 本发明实施例八还提供一种包含计算机可执行指令的存储介质,所述计算机可执行指令在由计算机处理器执行时用于执行基于小样本的OCR模板学习方法,该方法包括:

[0129] 获取图像数据集,根据所述图像数据集训练神经网络,得到通用文本检测模型;其中,所述神经网络包括特征提取层、像素分类网络和文本框位置网络;

[0130] 获取待学习的新模板训练集,固定所述通用文本检测模型中的特征提取层参数,基于所述新模板训练集调整所述通用文本检测模型中的像素分类网络参数和文本框位置网络参数,获得目标模板。

[0131] 当然,本发明实施例所提供的一种包含计算机可执行指令的存储介质,其计算机可执行指令不限于如上所述的方法操作。

[0132] 通过以上关于实施方式的描述,所属领域的技术人员可以清楚地了解到,本发明可借助软件及必需的通用硬件来实现,当然也可以通过硬件实现,但很多情况下前者是更佳的实施方式。基于这样的理解,本发明的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品可以存储在计算机可读存储介质中,如计算机的软盘、只读存储器(Read-Only Memory,ROM)、随机存取存储器(Random Access Memory,RAM)、闪存(FLASH)、硬盘或光盘等,包括若干指令用以使得一台电子设备(可以是手机,个人计算机,服务器,或者网络设备等)执行本发明各个实施例所述的方法。

[0133] 值得注意的是,上述基于小样本的OCR模板学习装置实施例中,所包括的各个单元和模块只是按照功能逻辑进行划分的,但并不局限于上述的划分,只要能够实现相应的功能即可;另外,各功能单元的具体名称也只是为了便于相互区分,并不用于限制本发明的保护范围。

[0134] 对本领域的技术人员来说,可根据以上描述的技术方案以及构思,做出其它各种相应的改变以及形变,而所有的这些改变以及形变都应该属于本发明权利要求的保护范围之内。

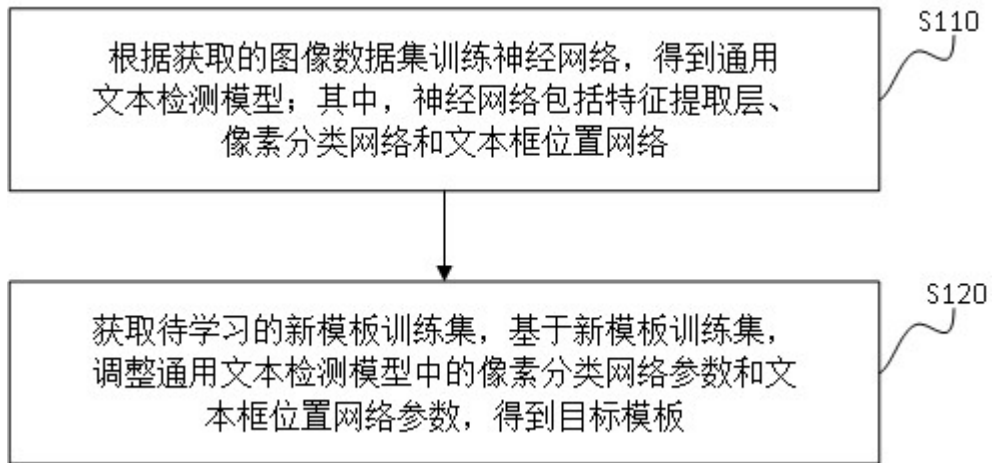


图1

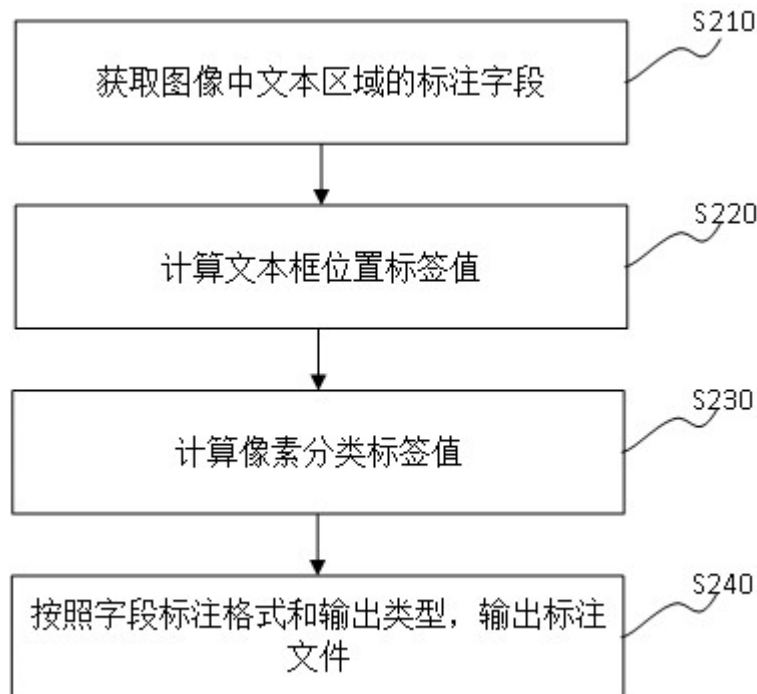


图2

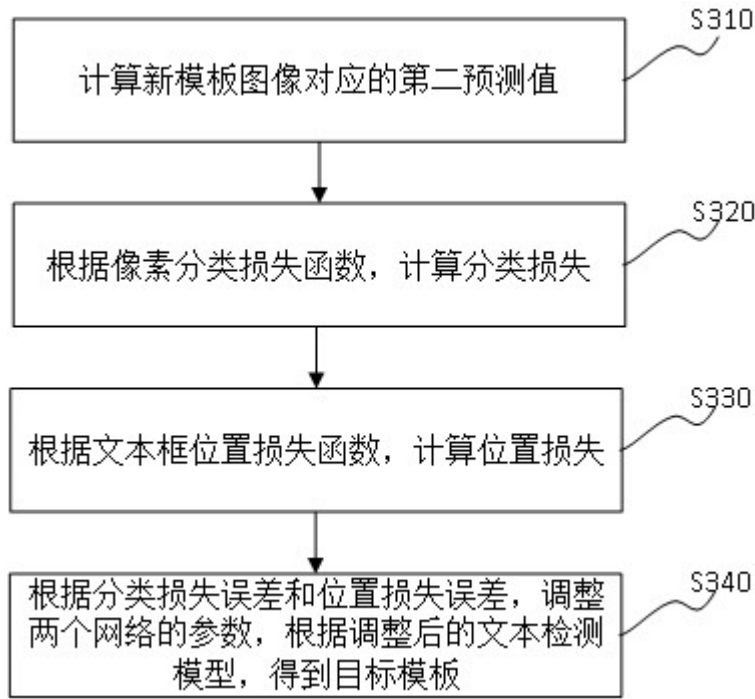


图3

X省增值税普通发票					发票号/开票日期/校验码				
购买方	名称:				密码区				
	纳税人识别号:								
	地址、电话:								
	开户行及账号:								
货物或应税劳务、服务名称		规格型号	单位	数量	单价	金额	税率	税额	
合计									
价税合计 (大写)		(小写)							
销售方	名称:				备注				
	纳税人识别号:								
	地址、电话:								
	开户行及账号:								
收款人:	复核:	开票人:	销售方: (章)						

图4

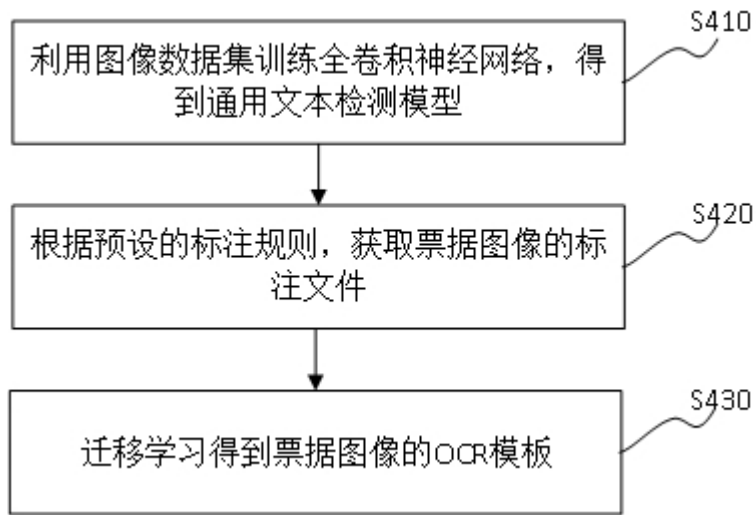


图5

订单详情	
发票类型:	
发票抬头:	
发票内容:	
订单编号:	
下单时间:	
支付方式:	
支付时间:	

图6

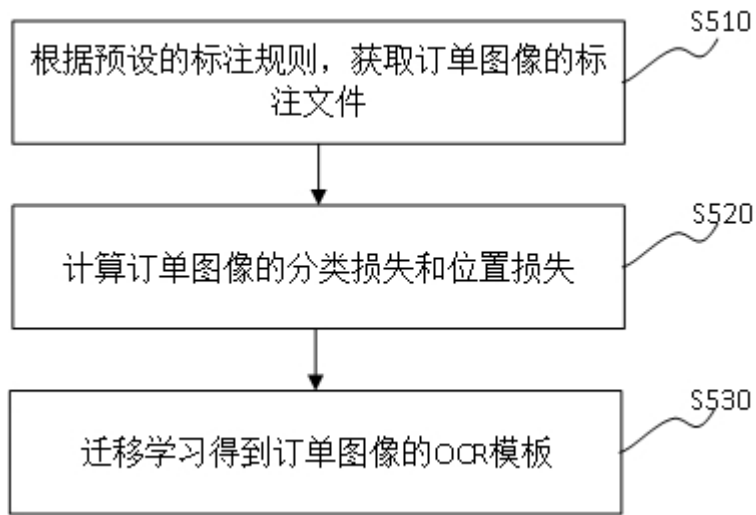


图7

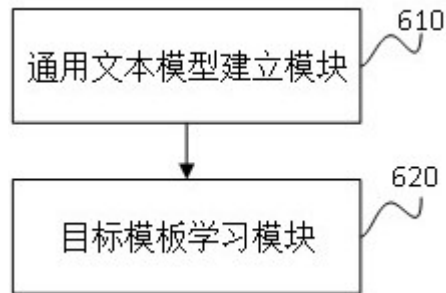


图8

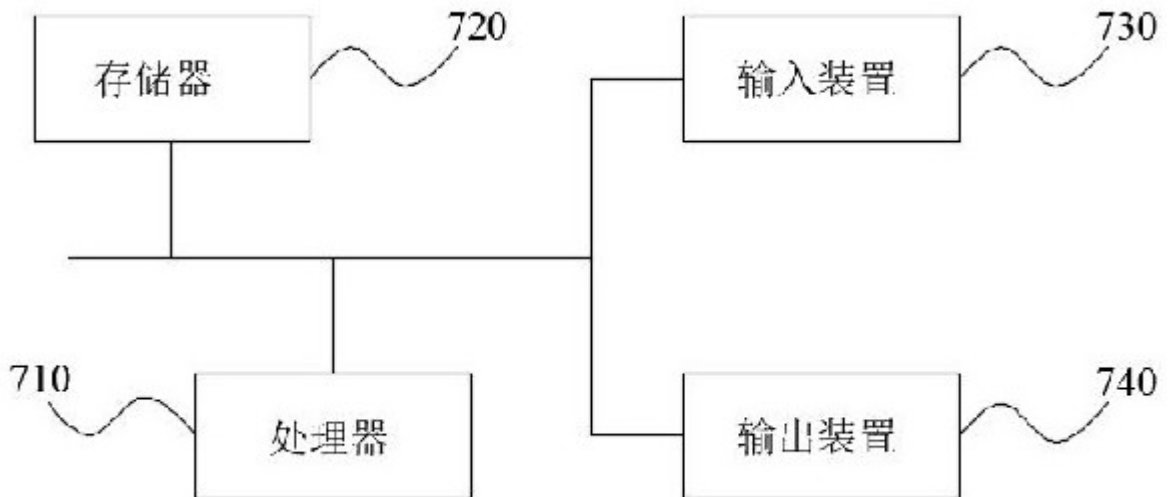


图9