



(12) 发明专利申请

(10) 申请公布号 CN 113436222 A

(43) 申请公布日 2021.09.24

(21) 申请号 202110597964.9

G06N 3/04 (2006.01)

(22) 申请日 2021.05.31

G06N 3/08 (2006.01)

(71) 申请人 新东方教育科技集团有限公司  
地址 100080 北京市海淀区海淀中街6号9层

(72) 发明人 麻凯利 马志国 张飞飞

(74) 专利代理机构 北京市柳沈律师事务所  
11105

代理人 侯鉴玻

(51) Int. Cl.

G06T 7/136 (2017.01)

G06T 7/194 (2017.01)

G06T 7/11 (2017.01)

G06K 9/20 (2006.01)

G06K 9/34 (2006.01)

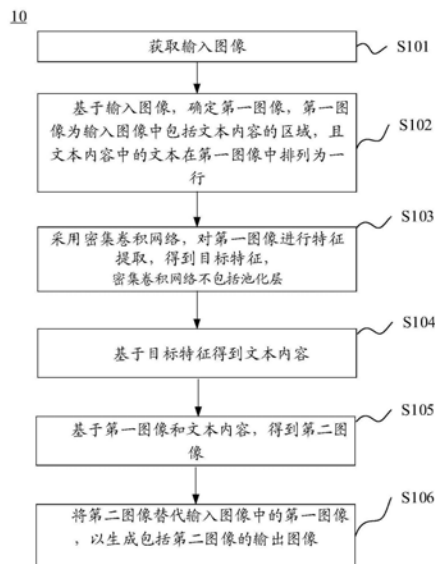
权利要求书3页 说明书16页 附图9页

(54) 发明名称

图像处理方法和图像处理装置、电子设备及存储介质

(57) 摘要

一种图像处理方法、图像处理装置、电子设备和存储介质。该图像处理方法包括：获取输入图像；基于输入图像，确定第一图像，第一图像为输入图像中包括文本内容的区域，且文本内容中的文本在第一图像中排列为一行；采用密集卷积网络，对第一图像进行特征提取，得到目标特征，该密集卷积网络不包括池化层；基于目标特征得到文本内容；基于第一图像和文本内容，得到第二图像，将第二图像替代输入图像中的第一图像，以生成包括第二图像的输出图像。该图像处理方法可以在保留输入图像的背景的同时，使得输入图像中的文本内容更规整、更清晰、分辨率更高，从而增强文本内容的观看质量；还可以较大程度地保留第一图像的特征信息，以提高文本识别的准确性。



1. 一种图像处理方法,包括:
  - 获取输入图像;
  - 基于所述输入图像,确定第一图像,其中,所述第一图像为所述输入图像中包括文本内容的区域,且所述文本内容中的文本在所述第一图像中排列为一行;
  - 采用密集卷积网络,对所述第一图像进行特征提取,得到目标特征,其中,所述密集卷积网络不包括池化层;
  - 基于所述目标特征得到所述文本内容;
  - 基于所述第一图像和所述文本内容,得到第二图像;
  - 将所述第二图像替代所述输入图像中的第一图像,以生成包括所述第二图像的输出图像。
2. 根据权利要求1所述的方法,其中,在所述密集卷积网络的卷积核中,至少一个卷积核在第一方向上的步长大于在第二方向上的步长。
3. 根据权利要求2所述的方法,其中,在所述密集卷积网络的卷积核中,至少一个卷积核在所述第一方向上的步长为2,并且在所述第二方向上的步长为1。
4. 根据权利要求2或3所述的方法,其中,所述第一方向为高度方向,所述第二方向为宽度方向。
5. 根据权利要求1所述的方法,其中,所述输入图像包括教学场景的图像,所述文本内容包括板书内容。
6. 根据权利要求1所述的方法,其中,基于所述第一图像和所述文本内容,得到所述第二图像,包括:
  - 基于所述第一图像,确定所述第一图像中的文本区域和非文本区域;
  - 基于所述非文本区域和所述文本区域,确定所述第二图像的背景图像;
  - 基于所述文本内容和所述背景图像,得到所述第二图像。
7. 根据权利要求6所述的方法,其中,基于所述第一图像,确定所述第一图像中的文本区域和非文本区域,包括:
  - 对所述第一图像进行灰度化处理,得到灰度图像;
  - 对所述灰度图像进行二值化处理,得到二值图像;
  - 根据所述二值图像确定所述第一图像中的文本区域和非文本区域。
8. 根据权利要求7所述的方法,其中,对所述灰度图像进行二值化处理,得到所述二值图像,包括:
  - 采用自适应阈值算法,对所述灰度图像进行二值化处理,得到所述二值图像。
9. 根据权利要求6所述的方法,其中,基于所述非文本区域和所述文本区域,得到所述第二图像的背景图像,包括:
  - 基于所述非文本区域的各个像素值,得到第一像素值;
  - 基于所述第一像素值和所述文本区域,得到所述背景图像。
10. 根据权利要求9所述的方法,其中,所述第一像素值包括所述非文本区域的各个像素值的平均值、最大值、最小值中之一。
11. 根据权利要求6所述的方法,其中,基于所述文本内容和所述背景图像,得到所述第二图像,包括:

基于所述文本内容和所述背景图像的大小,得到前景图像,其中,所述前景图像包括所述文本内容;

将所述前景图像覆盖在所述背景图像上,得到所述第二图像。

12. 根据权利要求11所述的方法,其中,所述文本内容对应的文字图像在所述前景图像中均匀排列。

13. 根据权利要求1所述的方法,其中,所述第二图像中的文本内容用印刷体表示。

14. 根据权利要求1所述的方法,其中,所述第一图像中的文本内容用手写体表示。

15. 根据权利要求1所述的方法,还包括:

获取包括多帧所述输出图像的图像序列;

抽取所述图像序列中的至少一帧输出图像;

响应于检测到所述至少一帧输出图像中存在满足第一预定条件的输出图像,将满足所述第一预定条件的输出图像作为目标帧输出图像存储,并存储与所述目标帧输出图像对应的文本内容,以用于对所述图像序列进行定位。

16. 根据权利要求15所述的方法,其中,所述第一预定条件包括:

所述目标帧输出图像对应的文本内容与检测到的前一帧输出图像对应的文本内容的文本相似度超过第一阈值,或者

在基于所述输入图像确定的第一图像包括至少一个第一图像的情况下,所述目标帧输出图像包括的第一图像的数量与检测到的前一帧输出图像包括的第一图像的数量之间的差值超过第二阈值。

17. 根据权利要求16所述的方法,其中,所述文本相似度包括归一化编辑距离。

18. 根据权利要求1所述的方法,其中,基于所述输入图像,确定所述第一图像,包括:

对所述输入图像进行文本区域检测,得到包括所述文本内容的第一图像。

19. 根据权利要求18所述的方法,其中,对所述输入图像进行文本区域检测,得到包括所述文本内容的第一图像,包括:

采用基于分割的可导的二值化网络,对所述输入图像进行文本区域检测,得到包括所述文本内容的第一图像。

20. 根据权利要求1所述的方法,其中,基于所述目标特征得到所述文本内容,包括:

通过循环神经网络,基于所述目标特征得到序列特征;

采用第一处理方法,基于所述序列特征得到所述文本内容。

21. 根据权利要求20所述的方法,其中,所述循环神经网络包括双向循环神经网络。

22. 根据权利要求20或21所述的方法,其中,所述第一处理方法包括连接时间分类方法。

23. 根据权利要求1所述的方法,还包括:

对所述第一图像进行预处理,

其中,所述预处理包括仿射变换、薄板样条变换中的至少一个。

24. 一种图像处理装置,包括:

获取单元,配置为获取输入图像;

确定单元,配置为基于所述输入图像,确定第一图像,其中,所述第一图像为所述输入图像中包括文本内容的区域,且所述文本内容中的文本在所述第一图像中排列为一行;

特征提取单元,配置为采用密集卷积网络,对所述第一图像进行特征提取,得到目标特征,其中,所述密集卷积网络不包括池化层;

识别单元,配置为基于所述目标特征得到所述文本内容;

融合单元,配置为基于所述第一图像和所述文本内容,得到第二图像;

生成单元,配置为将所述第二图像替代所述输入图像中的第一图像,以生成包括所述第二图像的输出图像。

25. 一种电子设备,包括:

存储器,用于非暂时性存储计算机可读指令;以及

处理器,用于运行所述计算机可读指令,其中,所述计算机可读指令被所述处理器运行时执行根据权利要求1-23任一项所述的图像处理方法。

26. 一种非瞬时性存储介质,非暂时性地存储计算机可读指令,其中,当所述非暂时性计算机可读指令由计算机执行时,执行根据权利要求1-23任一项所述的图像处理方法的指令。

## 图像处理方法、图像处理装置、电子设备及存储介质

### 技术领域

[0001] 本公开的实施例涉及一种图像处理方法、图像处理装置、电子设备和非瞬时性存储介质。

### 背景技术

[0002] 随着教育信息化的不断发展,视频课程在教学过程中的应用日益广泛。例如,在进行面授课堂教学时,可以利用录播系统实现对课堂教学内容的录制以形成课堂教学视频,由此,学生可以在线观看课堂教学视频以补习、复习相关教学内容。另外,课堂教学视频还广泛应用于教学评估、示范课录制、教学观摩、远程教学等。

### 发明内容

[0003] 本公开至少一个实施例提供一种图像处理方法,包括:获取输入图像;基于所述输入图像,确定第一图像,所述第一图像为所述输入图像中包括文本内容的区域,且所述文本内容中的文本在所述第一图像中排列为一行;采用密集卷积网络,对所述第一图像进行特征提取,得到目标特征,所述密集卷积网络不包括池化层;基于所述目标特征得到所述文本内容;基于所述第一图像和所述文本内容,得到第二图像;将所述第二图像替代所述输入图像中的第一图像,以生成包括所述第二图像的输出图像。

[0004] 例如,在本公开至少一个实施例提供的图像处理方法中,在所述密集卷积网络的卷积核中,至少一个卷积核在第一方向上的步长大于在第二方向上的步长。

[0005] 例如,在本公开至少一个实施例提供的图像处理方法中,在所述密集卷积网络的卷积核中,至少一个卷积核在所述第一方向上的步长为2,并且在所述第二方向上的步长为1。

[0006] 例如,在本公开至少一个实施例提供的图像处理方法中,所述第一方向为高度方向,所述第二方向为宽度方向。

[0007] 例如,在本公开至少一个实施例提供的图像处理方法中,所述输入图像包括教学场景的图像,所述文本内容包括板书内容。

[0008] 例如,在本公开至少一个实施例提供的图像处理方法中,基于所述第一图像和所述文本内容,得到第二图像,包括:基于所述第一图像,确定所述第一图像中的文本区域和非文本区域;基于所述非文本区域和所述文本区域,确定所述第二图像的背景图像;基于所述文本内容和所述背景图像,得到所述第二图像。

[0009] 例如,在本公开至少一个实施例提供的图像处理方法中,基于所述第一图像,确定所述第一图像中的文本区域和非文本区域,包括:对所述第一图像进行灰度化处理,得到灰度图像;对所述灰度图像进行二值化处理,得到二值图像;根据所述二值图像确定所述第一图像中的文本区域和非文本区域。

[0010] 例如,在本公开至少一个实施例提供的图像处理方法中,对所述灰度图像进行二值化处理,得到二值图像,包括:采用自适应阈值算法,对所述灰度图像进行二值化处理,得

到所述二值图像。

[0011] 例如,在本公开至少一个实施例提供的图像处理方法中,基于所述非文本区域和所述文本区域,得到所述第二图像的背景图像,包括:基于所述非文本区域的各个像素值,得到第一像素值;基于所述第一像素值和所述文本区域,得到所述背景图像。

[0012] 例如,在本公开至少一个实施例提供的图像处理方法中,所述第一像素值包括所述非文本区域的各个像素值的平均值、最大值、最小值中之一。

[0013] 例如,在本公开至少一个实施例提供的图像处理方法中,基于所述文本内容和所述背景图像,得到所述第二图像,包括:基于所述文本内容和所述背景图像的大小,得到前景图像。所述前景图像包括所述文本内容;将所述前景图像覆盖在所述背景图像上,得到所述第二图像。

[0014] 例如,在本公开至少一个实施例提供的图像处理方法中,所述文本内容对应的文字图像在所述前景图像中均匀排列。

[0015] 例如,在本公开至少一个实施例提供的图像处理方法中,所述第二图像中的文本内容用印刷体表示。

[0016] 例如,在本公开至少一个实施例提供的图像处理方法中,所述第一图像中的文本内容用手写体表示。

[0017] 例如,本公开至少一个实施例提供的图像处理方法还包括:获取包括多帧所述输出图像的图像序列;抽取所述图像序列中的至少一帧输出图像;响应于检测到所述至少一帧输出图像中存在满足第一预定条件的输出图像,将满足所述第一预定条件的输出图像作为目标帧输出图像存储,并存储与所述目标帧输出图像对应的文本内容,以用于对所述图像序列进行定位。

[0018] 例如,在本公开至少一个实施例提供的图像处理方法中,所述第一预定条件包括:所述目标帧输出图像对应的文本内容与检测到的前一帧输出图像对应的文本内容的文本相似度超过第一阈值,或者在基于所述输入图像确定的第一图像包括至少一个第一图像的情况下,所述目标帧输出图像包括的第一图像的数量与检测到的前一帧输出图像包括的第一图像的数量之间的差超过第二阈值。

[0019] 例如,在本公开至少一个实施例提供的图像处理方法中,所述文本相似度包括归一化编辑距离。

[0020] 例如,在本公开至少一个实施例提供的图像处理方法中,基于所述输入图像,确定第一图像,包括:对所述输入图像进行文本区域检测,得到包括所述文本内容的第一图像。

[0021] 例如,在本公开至少一个实施例提供的图像处理方法中,对所述输入图像进行文本区域检测,得到包括所述文本内容的第一图像,包括采用基于分割的可导的二值化网络,对所述输入图像进行文本区域检测,得到包括所述文本内容的第一图像。

[0022] 例如,在本公开至少一个实施例提供的图像处理方法中,于所述目标特征得到所述文本内容,包括:通过循环神经网络,基于所述目标特征得到序列特征;采用第一处理方法,基于所述序列特征得到所述文本内容。

[0023] 例如,在本公开至少一个实施例提供的图像处理方法中,所述循环神经网络包括双向循环神经网络。

[0024] 例如,在本公开至少一个实施例提供的图像处理方法中,所述第一处理方法包括

连接时间分类方法。

[0025] 例如,在本公开至少一个实施例提供的图像处理方法还包括:对所述第一图像进行预处理,所述预处理包括:仿射变换、薄板样条变换中的至少一个。

[0026] 例如,本公开至少一个实施例还提供一种图像处理装置,包括:获取单元,配置为获取输入图像;确定单元,配置为基于所述输入图像,确定第一图像,所述第一图像为所述输入图像中包括文本内容的区域,且所述文本内容中的文本在所述第一图像中排列为一行;特征提取单元,配置为采用密集卷积网络,对所述第一图像进行特征提取,得到目标特征,密集卷积网络不包括池化层;识别单元,配置为基于所述目标特征得到所述文本内容;融合单元,配置为基于所述第一图像和所述文本内容,得到第二图像;生成单元,配置为将所述第二图像替代所述输入图像中的第一图像,以生成包括所述第二图像的输出图像。

[0027] 例如,本公开至少一个实施例还提供一种电子设备,包括:存储器,用于非暂时性存储计算机可读指令;以及处理器,用于运行所述计算机可读指令,所述计算机可读指令被所述处理器运行时执行本公开任一实施例所述的图像处理方法。

[0028] 例如,本公开至少一个实施例还提供一种非瞬时性存储介质,非暂时性地存储计算机可读指令,当所述非暂时性计算机可读指令由计算机执行时,执行本公开任一实施例所述的图像处理方法的指令。

## 附图说明

[0029] 为了更清楚地说明本公开实施例的技术方案,下面将对实施例的附图作简单地介绍,显而易见地,下面描述的附图仅仅涉及本公开的一些实施例,而非对本公开的限制。

[0030] 图1为本公开至少一个实施例提供的关于教学场景的一帧视频图像;

[0031] 图2为本公开至少一个实施例提供的一种图像处理方法的流程图;

[0032] 图3为本公开至少一个实施例提供的一种输入图像经过文本检测后的结果示意图;

[0033] 图4为本公开至少一个实施例提供的一种第一图像的示意图;

[0034] 图5为本公开至少一个实施例提供的一种经过预处理后的第一图像的示意图;

[0035] 图6为本公开至少一个实施例提供的一种文本识别方法的示意图;

[0036] 图7为本公开至少一个实施例提供的一种连接时间分类方法的示意图;

[0037] 图8为本公开至少一个实施例提供的一种密集卷积网络(DenseNet)与常规DenseNet-121的结构对比表;

[0038] 图9为本公开至少一个实施例提供的一种密度分布不均匀的第一图像;

[0039] 图10为本公开至少一个实施例提供的一种经过处理后的第一图像;

[0040] 图11为本公开至少一个实施例提供的对应步骤S105的流程示意图;

[0041] 图12为本公开至少一个实施例提供的对应步骤S201的流程示意图;

[0042] 图13为本公开至少一个实施例提供的一种第二图像的示意图;

[0043] 图14为本公开至少一个实施例提供的另一种图像处理方法的流程图;

[0044] 图15为本公开至少一个实施例提供的一种应用于教学场景的图像处理方法的示意图;

[0045] 图16为本公开至少一个实施例提供的一种图像处理装置的示意框图;

- [0046] 图17为本公开至少一个实施例提供的一种电子设备的示意框图；
- [0047] 图18为本公开至少一个实施例提供的另一种电子设备的示意框图；以及
- [0048] 图19为本公开至少一个实施例提供的一种存储介质的示意图。

### 具体实施方式

[0049] 为使本公开实施例的目的、技术方案和优点更加清楚，下面将结合附图，对本公开实施例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公开的一部分实施例，而不是全部的实施例。基于所描述的本公开的实施例，本领域普通技术人员在无需创造性劳动的前提下所获得的所有其他实施例，都属于本公开保护的范围。

[0050] 除非另外定义，本公开使用的技术术语或者科学术语应当为本公开所属领域内具有一般技能的人士所理解的通常意义。本公开中使用的“第一”、“第二”以及类似的词语并不表示任何顺序、数量或者重要性，而只是用来区分不同的组成部分。同样，“一个”、“一”或者“该”等类似词语也不表示数量限制，而是表示存在至少一个。“包括”或者“包含”等类似的词语意指出现该词前面的元件或者物件涵盖出现在该词后面列举的元件或者物件及其等同，而不排除其他元件或者物件。“连接”或者“相连”等类似的词语并非限定于物理的或者机械的连接，而是可以包括电性的连接，不管是直接的还是间接的。“上”、“下”、“左”、“右”等仅用于表示相对位置关系，当被描述对象的绝对位置改变后，则该相对位置关系也可能相应地改变。

[0051] 在录制课堂教学视频时，通常使用摄像头对面授课教室的讲台方向进行视频采集。在实际的课堂教学现场中，不同教室中摄录黑板部分的摄像头分辨率不同，不同设备分辨率差异较大。例如，在部分教室中，限于低带宽的线路等原因，一些设备会将视频压缩得体积较小、质量较差，以便传输。复杂教室场景中由于黑板一般较宽，黑板中间部分和两边部分距离摄像头距离差别较大，因而聚焦程度不同，容易造成视频中黑板局部的板书文字质量较差的现象。例如，图1为本公开至少一个实施例中提供的关于教学场景的一帧视频图像，如图1所示，该视频图像的分辨率较低，尤其是左边的板书区域，该板书区域中的板书文字难以看清。

[0052] 为了解决上述视频图像中板书文字质量较差、难以看清的问题，通常使用的方法主要包括三种。第一种方法是对视频图像中的区域进行无差别的盲修复，诸如锐化、边缘增强、直方图均衡化、使用拉普拉斯算子等方法。然而，在这种方法中，执行无差别的盲修复时往往对于阴影、不同背景等复杂场景的修复效果不理想，并且对全图整个区域的调整可能使得非文字区域的质量下降。第二种方法是在分离文字和背景时采用固定阈值等方式。然而，在这种方法中，处理文字区域时采用各种固定阈值，则难以适应不同场景的教室（例如，黑板或者白板等）以及不同场景的光线（例如，黑板部分一半被阳光直射而另一半没有被阳光直射），鲁棒性较差，难以准确的将手写板书和背景准确的区分开。第三种方法是针对文字区域使用基于深度学习的图像增强模型。然而，在这种方法中，针对性的增强了文字区域像素，由于是基于像素级别的增强，鲁棒性十分依赖模型性能，因此并没有提供文字部分的语义信息。

[0053] 经本申请的发明人研究发现，若基于新背景重建板书信息，则长视频中一旦有某些元素漏检等失误会影响视频观看体验，而在原视频上增强板书内容则可避免这种缺陷。



[0054] 本公开至少一个实施例提供一种图像处理方法,包括:获取输入图像;基于输入图像,确定第一图像,第一图像为输入图像中包括文本内容的区域,且文本内容中的文本在第一图像中排列为一行;采用密集卷积网络,对第一图像进行特征提取,得到目标特征,该密集卷积网络不包括池化层;基于目标特征得到文本内容;基于第一图像和文本内容,得到第二图像,将第二图像替代输入图像中的第一图像,以生成包括第二图像的输出图像。

[0055] 本公开至少一个实施例还提供对应于上述图像处理方法的图像处理装置、电子设备以及非瞬时性存储介质。

[0056] 本公开至少一个实施例提供的图像处理方法,通过生成包括第二图像的输出图像,可以在保留输入图像的背景的同时,使得输入图像中的文本内容更加规整、更清晰、分辨率更高,从而增强文本内容的观看质量。通过采用不包括池化层的密集卷积网络,较大程度地保留了第一图像的特征信息,从而提高文本识别的准确性。

[0057] 下面结合附图,通过几个示例或实施例对根据本公开的至少一个实施例提供的图像处理方法进行非限制性的说明,如下面所描述的,在不相互抵触的情况下这些具体示例或实施例中不同特征可以相互组合,从而得到新的示例或实施例,这些新的示例或实施例也都属于本公开保护的范畴。为了保持本公开实施例的以下说明清楚且简明,本公开省略了已知功能和已知部件的详细说明。当本公开实施例的任一部件在一个以上的附图中出现时,该部件在每个附图中由相同或类似的参考标号表示。

[0058] 图2为本公开至少一个实施例提供的一种图像处理方法的流程图。例如,如图2所示,图像处理方法10可以应用于计算设备,该计算设备包括具有计算功能的任何电子设备,例如可以为智能手机、笔记本电脑、平板电脑、台式计算机、服务器等,本公开的实施例对此不作限制。例如,该计算设备具有中央处理单元(Central Processing Unit,CPU)或图形处理单元(Graphics Processing Unit,GPU),该计算设备还包括存储器。该存储器例如为非易失性存储器(例如只读存储器(Read Only Memory,ROM)),其上存储有操作系统的代码。例如,存储器上还存储有代码或指令,通过运行这些代码或指令,可以实现本公开实施例提供的图像处理方法。

[0059] 例如,如图2所示,该图像处理方法10可以包括以下步骤S101至步骤S106。需要说明的是,在本公开的实施例中,步骤S101-步骤S106可以顺序执行,也可以按调整后的其他次序执行,步骤S101-步骤S106中的部分或全部操作还可以并行执行,本公开的实施例对各个步骤的执行顺序不作限制,可以根据实际情况调整。例如,在一些示例中,实施本公开至少一个实施例提供的图像处理方法10可以选择地执行步骤S101-S106中的部分步骤,也可以执行除了步骤S101-S106以外的一些附加步骤,本公开的实施例对此不做具体限制。

[0060] 步骤S101:获取输入图像。

[0061] 例如,在本公开至少一个实施例中,输入图像可以通过相机拍摄得到,通过从网络(例如,因特网)下载得到,或者通过从本地读取得到等,本公开的实施例对此不作限制。例如,输入图像中包括本文内容。例如,文本内容可以是指文字、字母、符号、图形等内容,本公开的实施例对此不作限制。

[0062] 例如,在本公开至少一个实施例中,输入图像可以包括教学场景的输入图像,文本内容可以包括板书内容。例如,输入图像可以为如图1所示的图像,在通常的教学场景中,教师在讲台上进行教学,并辅以各种板书。例如,在本公开的实施例中,板书是指可以用于展

示教师的教学内容的载体,包括黑板、白板、PPT的投影区域等;例如,在板书中,教师可以通过文字、图形、字符等板书内容(即文本内容)对教学内容进行阐释。应当理解的是,对于黑板、白板、PPT的投影区域等板书而言,无论其上是否具有具体的板书内容,其均被视为板书。

[0063] 例如,教学场景的输入图像可以包括通过摄像头(例如,摄像机的摄像头、智能手机的摄像头等)对上述教学场景的讲台方向进行拍摄得到的照片和视频图像等。例如,教学场景的输入图像的画面通常包括上述板书的画面(即板书区域)。

[0064] 例如,在一些实施例中,输入图像可以为彩色图像。例如,彩色图像包括但不限于具有三种颜色通道的彩色图像等。例如,该三种颜色通道包括第一颜色通道、第二颜色通道和第三颜色通道。例如,该三种颜色通道分别对应于三原色。例如,在一些实施例中,第一颜色通道为红色(R)通道,第二颜色通道为绿色(G)通道,第三颜色通道为蓝色(B)通道,即上述彩色图像可以为RGB格式的彩色图像,需要说明的是,本公开的实施例包括但不限于此。例如,在另一些实施例中,输入图像也可以为灰度图像。

[0065] 步骤S102:基于输入图像,确定第一图像,第一图像为输入图像中包括文本内容的区域,且文本内容中的文本在第一图像中排列为一行。

[0066] 例如,在本公开至少一个实施例中,第一图像可以是指文本内容在输入图像中所出现的区域的图像,且文本内容中的文本在第一图像中排列为一行。例如,在一些实施例中,第一图像可以是仅覆盖一行文本长度的文本框。通常,对于复杂场景的文本识别,首先要定位文本所在位置,即进行文本区域检测,得到文本区域图像(例如,第一图像),以便后续更高效准确地执行文本识别操作。例如,输入图像为包括教学场景的图像时,图像中的板书中可能存在多行文本内容。本公开实施例提供的图像处理方法以板书中的一行文本内容为对象进行处理,当需要对板书中的多行文本内容进行处理时,只需要针对各行文本内容分别执行本公开实施例提供的图像处理方法中相应的操作,即可以实现对板书中所有的文本内容的处理。

[0067] 例如,在本公开至少一个实施例中,步骤S102可以包括:对输入图像进行文本检测,得到包括文本内容的第一图像。

[0068] 图3为本公开至少一个实施例提供的一种输入图像经过文本检测后的结果示意图。例如,输入图像可以包括一行或多行文本。例如,在一个示例中,输入图像包括一行文本,例如仅包括一个第一图像,在这种情况下,可以对检测到的第一图像执行后续文本识别操作。例如,在另一个示例中,如图3所示,输入图像可以包括多行文本,例如多个第一图像,在这种情况下,可以依次对多个第一图像执行文本识别操作,输出多个第一图像的文本内容,本公开的实施例对此不做限制。

[0069] 需要说明的是,在本公开的实施例中,可以采用各种常规的文本检测方法对输入图像进行文本区域检测,本公开的实施例对此不作限制,只要能得到输入图像中包括文本内容的第一图像即可。

[0070] 目前,文本检测方法大致分为两类:基于回归的方法和基于分割的方法。通常,对于复杂场景(例如,教学场景),使用基于分割的场景文本检测方法通常较为准确。例如,在至少一个实施例中,可以采用基于分割的可导二值化网络(differentiable binarization network),对输入图像进行文本区域检测,得到包括文本内容的第一图像。

[0071] 例如,在一些实施例中,将输入图像输入基于分割的可导二值化网络后,经过特征提取和上采样融合等操作后得到特征图,然后使用特征图预测出概率图(probability map)和阈值图(threshold map),通过概率图和阈值图计算出近似二值图,从而通过概率图或近似二值图来推断出第一图像(例如,文本区域或者文本框)。在上述方法中,概率图用于输出每个像素是文本的概率,提供文本所在的像素区域。采用基于分割的可导二值化网络方法的优点在于对每一个像素点进行自适应二值化,二值化阈值由网络学习得到,彻底将二值化这一步骤加入到网络里一起训练,这样最终的输出对于阈值就会具有较好的鲁棒性。由于阈值图的存在,该二值化网络可以很好的学习到边界,因此,对倾斜、扭曲等文本检测效果较好。

[0072] 例如,采用基于分割的可导二值化网络方法,可以输出多边形的文本区域,可以克服基于矩形框的贴合不准的问题。当然,文本区域的形状也可以是矩形、阶梯形等其他形状,本公开的实施例对此不作限制。

[0073] 例如,在一些实施例中,可以采用连接文本提议网络(Connectionist Text Proposal Network,CTPN)来实现对输入图像的文本检测。例如,CTPN包括检测小尺度文本框、循环连接文本框、以及文本行边细化等步骤。需要说明的是,在本公开的实施例中,对文本检测的方法不作限制。

[0074] 例如,在一些实施例中,如图3中示出的位于灰色四边形框内的区域的图像可以作为第一图像。例如,图3所示的输入图像包括六行文字(即,六个四边形文本框),则基于图3的输入图像可以检测到六个第一图像,以便后续依次对这六个第一图像执行文本识别操作。

[0075] 例如,在一些实施例中,第一图像(文本框或者文本检测区域)的各边分别与最近的文字的距离在一定阈值范围内,例如,阈值为2~5mm。例如,当文本框是矩形时,矩形的上、下边与一行文字的最上、下边缘的距离不超过一定阈值,矩形的左、右边与一行文字的两侧边缘的距离不超过一定阈值。例如,阈值是可以根据需要调整的,即第一图像的大小是可以限定的,根据使用的文本检测方法调整。

[0076] 例如,在本公开至少一个实施例中,可以对经过步骤S102得到的一个或多个第一图像进行文本识别,以确定每个第一图像中的文本内容。需要说明的是,可以采用各种常规的文本识别方法,本公开的实施例对此不作限制,只要能基于第一图像,识别出相应的文本内容即可。

[0077] 例如,在本公开至少一个实施例中,为了提高文本识别的准确率,在对第一图像进行文本识别之前,可以先对第一图像进行预处理。

[0078] 图4为本公开至少一个实施例提供的一种第一图像的示意图,图5为本公开至少一个实施例提供的一种经过预处理后的第一图像的示意图。例如,在一个示例中,在对输入图像进行文本区域检测后,得到一种如图4所示的第一图像,如图4所示,第一图像为一种倾斜的四边形文本区域。为了提高后续文本识别操作的识别精度,将该倾斜的四边形文本区域经过例如,仿射变换、薄板样条(thin plate spline,TPS)变换等预处理操作后转换为矩形区域,如图5所示。例如,仿射变换是指一个向量空间进行一次线性变换并接上一个平移,变换为另一个向量空间,仿射变换可以允许图形任意倾斜,而且允许图形在两个方向上任意伸缩。例如,薄板样条(TPS)变换通常基于2D插值,常应用于图像配准方面。例如,在两张图

像中找出N个匹配点,应用TPS可以将这N个点形变到对应位置,同时给出了整个空间的形变。例如,薄板样条(TPS)变换可以根据自适应的采样点进一步对局部扭曲进行矫正,提高后续文本识别的准确性。

[0079] 需要说明的是,预处理操作除了上述仿射变换、薄板样条变换以外,还可以包括其他操作,例如透视变换等,本公开的实施例对此不作限制。

[0080] 例如,在本公开至少一个实施例中,可以采用神经网络模型来实现对第一图像的文本识别,当然,也可以采用其他常规的文本识别方法,本公开的实施例对此不作限制,根据实际需求来调整。

[0081] 例如,在一个示例中,采用如图6所示的神经网络模型来进行文本识别。例如,该神经网络模型的网络结构可以由三部分组成,自下而上地包括卷积层、循环层和转录层。例如,卷积层可以采用任意一种卷积神经网络(CNN),如图6所示的底部CNN,用于对第一图像(例如,图5中的第一图像)提取特征,得到卷积特征图。例如,在一个示例中,第一图像的大小为(32,100,3),通过CNN以后,转换为(1,25,512)大小的卷积特征矩阵(卷积特征图),这里是以(高度,宽度,通道)的形式表示大小。

[0082] 为了将特征输入到循环层,需要从上述CNN输出的特征图中提取特征向量序列,每一个特征向量在特征图上按列从左到右生成,例如,在上述示例中,从(1,25,512)大小的卷积特征矩阵中提取的每一个特征向量包含512维特征,这些特征向量就构成特征序列。该特征序列可以作为循环层的输入,每个特征向量作为循环神经网络(RNN)在一个时间步T的输入。在上述(1,25,512)大小的卷积特征矩阵的示例中,一共有25个时间步T。例如,循环层可以采用任意一种循环神经网络(RNN),例如双向循环神经网络(Bi-directional RNN, BiRNN)、双向长短期记忆(Bi-directional Long Short Term Memory, BiLSTM)网络等,如图6所示的RNN,用于对特征序列进行预测,对特征序列中每个特征向量进行学习,然后输出序列特征(例如,所有标签的概率分布)。例如,一个时间步有一个输入特征向量,对于每个输入特征向量输出一个序列特征,例如概率向量。

[0083] 例如,转录层可以采用连接时间分类(Connectionist Temporal Classification, CTC)等第一处理方法,用于根据RNN输出的多个序列特征(概率向量)预测具有最高概率组合的标签序列,从而得到文本内容。例如,标签可以是字符,也可以是文字,这可以根据实际需求来设置。上述使用的CTC方法可以解决输入长度不固定与输出字符长度不固定的匹配问题。

[0084] 例如,图7为本公开至少一个实施例提供的一种CTC方法的示意图。CTC方法可以通过预测空白字符来解决对齐问题。例如,在一个示例中,如图7所示,CTC方法可以包括:(1)预测结果重合并重复字符;(2)去掉空白字符;(3)输出剩余字符。如图7所示,例如,预测出多个h,由于多个h中间没有空白字符,则多个h会被合并,即空白字符之间的重复字符会被合并,而原本就是重复的字符之间必须要有空白字符才能不被合并。例如,图中“hello”的“l”和“l”之间有空白符,因而最终合并后具有两个“l”。因此,假设某一字符串具有8个“h”,那么在步骤(2)之前会变成“h-h-h-h-h-h-h-h”,至少需要模型能预测15个字符的能力才能预测出重复的8个“h”,否则一些“h”就会被合并。

[0085] 需要说明的是,为了简洁,此处仅仅对CTC方法进行简单介绍,具体请参考CTC方法的相关文献。

[0086] 例如,通过采用图6所述的文本识别方法,基于图5中的第一图像可以识别出第一图像中的文本内容为“第十四次课小古文阅读(二)”。需要说明的是,在本公开的实施例中,也可以采用其他的文本识别方法,只要能基于第一图像,得到对应的文本内容即可。

[0087] 鉴于通常检测到的第一图像(文本框或文本区域)长宽不固定,且在大部分场景下(例如,教学场景),第一图像对应的是手写文本,因此经常会出现第一图像内文字密度不均匀,或第一图像内有较大的空格区域等现象,若直接采用一些常规文本识别模型,通常效果不理想。因此,在本公开至少一个实施例中,基于图6所示的神经网络模型,对部分网络结构进行调整。

[0088] 步骤S103:采用密集卷积网络,对第一图像进行特征提取,得到目标特征,该密集卷积网络不包括池化层。

[0089] 例如,在至少一个实施例中,上述卷积层中的卷积神经网络可以采用密集卷积网络(DenseNet),以进行图像特征提取。

[0090] 例如,常规的卷积神经网络,参数只能一层一层的向下传播,而密集卷积网络在保证网络中层与层之间最大程度的信息传输的前提下,直接将所有层连接起来,从而更好地利用特征中的信息,提高模型识别精度。例如,相比使用残差网络(ResNet),密集卷积网络进一步消灭了梯度消失的问题,提高了特征的复用性,同时也减少了网络参数量。

[0091] 图8为本公开至少一个实施例提供的一种密集卷积网络(DenseNet)与常规DenseNet-121的结构对比表。

[0092] 例如,在本公开至少一个实施例中,如图8所示,相比于常规的密集卷积网络(例如,图8中的DenseNet-121),调整后的密集卷积网络(例如,图8中的DenseNet)不包括池化层。例如,去掉了常规DenseNet-121中的单独的池化层,仅仅依靠卷积核步长(stride)来对各个层的输出维度进行调整,以免网络过快地丢掉像素中的特征,更大程度上保留第一图像的像素信息,从而提高文本识别的准确性。

[0093] 例如,在本公开至少一个实施例中,在密集卷积网络的卷积核中,至少一个卷积核在第一方向上的步长大于在第二方向上的步长,使得在第一方向上的特征图维度降低得比在第二方向上更快。例如,至少一个卷积核在第一方向上的步长为2,并且在第二方向上的步长为1。例如,第一方向为高度方向,并且第二方向为宽度方向。需要说明的是,在第一方向上的步长和在第二方向上的步长也可以是其他具体数值,本公开的实施例对此不作限制。

[0094] 例如,在本公开至少一个实施例中,密集卷积网络中各个层的卷积核步长均可以采用高度步长为2,宽度步长为1的设置,本公开的实施例对此不作限制,可以根据实际需求来设置。

[0095] 例如,在本公开至少一个实施例中,如图8所示,常规的DenseNet-121网络在结构上包括1-8层,假设输入维度为 $224 \times 224$ ,则最终输出维度为 $7 \times 7$ 。基于本公开至少一个实施例提供的一种DenseNet网络,假设输入维度为 $224 \times 32$ ,最终输出维度为 $56 \times 1$ 。结合图8容易看出,相比于DenseNet-121,本公开至少一个实施例提供的DenseNet网络不包括第2层池化层,从而在网络初期可以减少维度变小的程度,提高网络对像素信息的保留,从而提高文本识别的准确性。因为通过有参数的卷积操作来减小特征图维度,相比于通过没有参数的池化操作,更能保留特征信息。

[0096] 例如,相比于常规DenseNet-121,本公开至少一个实施例提供的DenseNet网络在第4层和第8层的过渡层模块中采用的卷积核步长为(1,2)来替换(2,2),即在宽度方向上步长为1,在高度方向上步长为2。这样,只降低高度方向上像素的维度,不降低宽度方向上像素的维度,提高了网络模型对密集文字的识别能力,从而提高文本识别的准确性。

[0097] 例如,通过上述调整后的密集卷积网络,最终输出维度为 $56 \times 1$ ,则该模型对单个样本可识别出的文字数量理论上在0-56个。在一些情况下(例如,单行字符全都相同的情况下),由于前述CTC方法中重复字符之间需要空白字符才不被合并,因此,可识别字符数量为0-28个。而通过常规密集卷积网络(例如,图8中的DenseNet-121),使用(2,2)的卷积核步长对同样的维度为 $224 \times 32$ 的图片进行操作,当高度为1时,宽度为7,理论上能识别的字符数为0-7个,但是在一些情况下(例如,单行字符全都相同的情况下),只能识别4个字符,如果要提高单行识别数量则需提高输入分辨率,或者增大输入图片的宽度。此外,即便理论上识别的字符数量满足实际需求,由于手写板书在文本行(第一图像)中密度不均匀,或有较大空格时,理论上最大识别7个字符的模型也难以预测出7个字符。

[0098] 图9为本公开至少一个实施例提供的一种密度分布不均匀的第一图像,图10为本公开至少一个实施例提供的一种经过处理后的第一图像。

[0099] 例如,基于图9所示的第一图像,基于上述CTC模型进行预测时,会把特征向量划分为均匀的N个区域,最终预测出至多N个字符。当文本均匀时识别结果较好,但当文字不均匀时,识别结果可能会较差。例如,在一个示例中,当模型在某一分辨率下,例如, $224 \times 32$ ,输出维度为 $7 \times 1$ 。如图10所示,经过均匀划分后的文本区域中有三个空格区域,最多只能预测4个有效字符。因此,上述输出维度为 $56 \times 1$ 的模型具有足够的分辨率用于区分字符。

[0100] 步骤S104:基于目标特征得到文本内容。

[0101] 例如,结合如图6所示的基于图像的文本识别方法,步骤S104可以包括通过循环神经网络(例如,双向LSTM),基于目标特征得到序列特征;以及采用第一处理方法(例如,CTC方法),基于序列特征得到文本内容。例如,在一个示例中,经过图8中所示的DenseNet后,可以将高度维度为1,宽度为T的特征向量依次输入双向LSTM中,提取序列特征,然后采用CTC方法预测得到相应的文本内容。需要说明的是,关于步骤S104的详细说明可以参考对图6的文本识别方法的相关描述,在此不再赘述。

[0102] 步骤S105:基于第一图像和文本内容,得到第二图像。

[0103] 例如,在本公开的实施例中,通过步骤S105,可以达到在原来场景的纹理背景下逼真地更换文本的效果,例如,将字迹潦草的手写文本更换为字迹规整、大小合适的打印字体,提高了视频中文本内容的观看质量。

[0104] 图11为本公开至少一个实施例提供的对应步骤S105的流程示意图,图12为本公开至少一个实施例提供的对应步骤S201的流程示意图,图13为本公开至少一个实施例提供的一种第二图像的示意图。

[0105] 例如,在本公开至少一个实施例中,如图11所示,对于步骤S105,基于第一图像和文本内容,得到第二图像可以包括以下步骤S201-步骤S203。

[0106] 步骤S201:基于第一图像,确定第一图像中的文本区域和非文本区域。

[0107] 例如,在本公开至少一个实施例中,对于第一图像(例如,文本区域或者文本框),需要将文本区域和非文本区域分离。例如,如图12所示,对于步骤S201,可以包括以下步骤

S301-步骤S303。

[0108] 步骤S301:对第一图像进行灰度化处理,得到灰度图像。

[0109] 例如,在一些实施例中,输入图像为彩色图像(例如,RGB格式的彩色图像),相应地,第一图像也为彩色图像;在此情况下,可以通过常用的转换公式将彩色图像(例如,第一图像)转换为灰度图像。以将RGB格式的彩色图像转换为灰度图像为例,可以通过下述转换公式进行转换:

[0110]  $Gray = R * 0.299 + G * 0.587 + B * 0.114,$

[0111] 其中,Gray表示灰度图像的亮度信息,R、G和B分别表示RGB格式的彩色图像的红色信息(即红色通道的数据信息)、绿色信息(即绿色通道的数据信息)和蓝色信息(即蓝色通道的数据信息)。当然,也可以采用其他转换公式以将彩色图像转换为灰度图像,本公开的实施例对此不作限制。

[0112] 例如,在另一些实施例中,输入图像为灰度图像,相应地,第一图像也为灰度图像;在此情况下,可以直接将第一图像作为灰度图像。

[0113] 步骤S302:对灰度图像进行二值化处理,得到二值图像。

[0114] 例如,在本公开至少一个实施例中,可以选择合适的值作为阈值,对灰度图像进行二值化处理,以得到二值图像。例如,在一个示例中,选择灰度图像中所有像素的灰度平均值作为阈值。需要说明的是,本公开包括但不限于此。例如,在实际应用中,还可以采用其他任意可行的方式确定二值化处理中的阈值。

[0115] 例如,在本公开至少一个实施例中,为了适应复杂场景的图像,例如教学场景的图像(通常在图像的不同区域具有不同的照明条件),可以采用自适应阈值算法,对灰度图像进行二值化处理,得到二值图像。例如,自适应阈值算法可以较好地适应局部像素块。自适应阈值算法基于某一像素周围的小区域来确定该像素的阈值。因此,对于同一图像的不同区域,可以获得不同的阈值,这为光照度变化的图像提供了更好的二值化结果。

[0116] 需要说明的是,可以采用各种常规的生成自适应阈值的方法,本公开的实施例对此不作限制。

[0117] 步骤S303:基于二值图像确定第一图像中的文本区域和非文本区域。

[0118] 例如,在本公开至少一个实施例中,第一图像中文本像素的数量往往是少于非文本像素的数量,因此,基于二值图像,容易区分第一图像中的文本区域和非文本区域。例如,对于教学场景的输入图像,还需要判断场景中是白板还是黑板。例如,通过判断二值图像中两个类别的像素差,可以判断是白板还是黑板。

[0119] 步骤S202:基于非文本区域和文本区域,确定第二图像的背景图像。

[0120] 例如,在本公开至少一个实施例中,为了使第二图像的背景图像与第一图像的背景图像基本一致,第二图像的背景图像保留第一图像的非文本区域,并用第一像素值填充第一图像的文本区域。例如,步骤S202可以包括:基于非文本区域的各个像素值,得到第一像素值;基于第一像素值和文本区域,得到背景图像。例如,在一个示例中,采用第一像素值填充文本区域,以得到第二图像的背景图像。例如,第一像素值可以是第一图像的非文本区域的各个像素值的平均值、最大值、最小值等,当然,本公开的实施例对此不作限制。

[0121] 步骤S203:基于文本内容和背景图像,得到第二图像。

[0122] 例如,在本公开至少一个实施例中,步骤S203包括基于文本内容和背景图像的尺

寸,得到第二图像的前景图像,前景图像包括文本内容;将前景图像覆盖在背景图像上,得到第二图像,如图13所示。例如,在一些实施例中,如图13所示,文本内容对应的文字图像在第二图像的前景图像中均匀排列。

[0123] 例如,在一些实施例中,基于检测到的文本区域或文本框尺寸(例如,第一图像或者背景图像的尺寸等),以及识别出来的文本内容的字符数,可以确定要渲染的文本所需字体大小,并将文本均匀排布在文本区域或文本框范围内,从而得到第二图像的前景图像。例如,前景图像的尺寸与背景图像的尺寸相同。例如,通过将识别到的文本内容转换成合适大小的打印字体(例如,宋体、楷体等),可以避免由于输入图像中手写文本字迹潦草以及字体过小等导致的难以看清的问题。

[0124] 例如,在一些实施例中,将前景图像覆盖在背景图像上以得到第二图像可以通过将前景图像对应的像素块与背景图像对应的像素块相乘来实施,本公开的实施例对此不作限制。

[0125] 例如,在一个示例中,先将第一图像中的手写笔迹区域(即文本区域)使用第一像素值填充,从而抹去手写痕迹。例如,以白板背景、黑色文字为例,假设第一像素值为 $(r, g, b)$ ,第一图像的宽度为 $w$ ,第一图像的高度为 $h$ ,以该第一像素值填充文本区域中的像素,得到背景图像。经过文本识别后得到第一图像中字符串数量为 $N$ ,由于第一图像的宽度为 $w$ ,则根据 $w/N$ 来得到每个字符渲染后的像素宽度,生成一个宽度为 $w$ ,高度为 $h$ ,大小与背景图像相同,且三通道像素值为1的三通道图像,在该三通道图像上通过加载预设字体(如宋体、楷体等打印字体)渲染出黑色的字符串,得到前景图像。将背景图像与相同大小的前景图像对应像素相乘,即可得到类似图13效果的渲染图,即第二图像。

[0126] 步骤S106:将第二图像替代输入图像中的第一图像,以生成包括第二图像的输出图像。

[0127] 例如,比较图4或图5的第一图像和图13的第二图像,可以看出第二图像中文本内容的观看质量得到有效提升。例如,相较于第一图像,第二图像有效消除了第一图像中手写字迹周边的像素扩散现象;将第一图像中大小不一、排布不均匀的手写文字重新均匀排布;将第一图像中有个人特色的手写字体转化为标准字体,例如打印字体;第二图像中字体颜色与背景对比度更高,因为第二图像中的文字从字体库中矢量图重新渲染得到,可以选择与背景对比度更高的颜色,且渲染后的文字内部像素值更均匀统一。因此,图像处理方法10可以在保留输入图像的背景的前提下,将更规整、更高分辨率的打印体文字替换原手写字,从而提高文本内容的观看质量。通过采用调整后的密集卷积网络,提高了文本识别的准确性。

[0128] 例如,在教学场景下,一节课的完整视频通常较长,可能几十分钟甚至一两个小时,视频变化往往较小。视频变化一般集中在板书、老师的位置、以及PPT投影的变化。由于板书部分是讲课内容的凝练和总结,因此,可以将板书内容对应的文字信息作为检索或者定位视频位置的信息,有助于视频定位。这样,当用户或者学生期望从某节课的视频中准确找到某个知识点的讲解来重听,只需要使用相关的关键词进行搜索,有针对性的观看视频,从而减少了在长视频中人工查阅以定位某个具体知识点的视频讲解的时间花费。

[0129] 图14为本公开至少一个实施例提供的另一种图像处理方法的流程图。例如,为了实现上述视频定位的功能,如图14所示,本公开至少一个实施例还提供一种图像处理方法



100,该图像处理方法100包括步骤S401-S409。

[0130] 需要说明的是,图14中步骤S401-步骤S406与图2中步骤S101-步骤S106基本一致,实现的技术效果也基本一致,为了简洁,本公开的实施例对步骤S401-步骤S406不再重复描述,可以参考步骤S101-步骤S106的相关内容。

[0131] 步骤S407:获取包括多帧输出图像的图像序列。

[0132] 例如,可以采用前述步骤S401至步骤S406实现生成输出图像的操作,从而获取包括多帧输出图像的图像序列,具体实现过程和细节可以参考前述相关描述,在此不再重复赘述。

[0133] 步骤S408:抽取图像序列中的至少一帧输出图像。

[0134] 例如,在一些实施例中,可以对连续的图像序列每隔预定时间(例如30秒)抽取一帧输出图像,从而得到至少一帧输出图像。又例如,也可以从获得的图像序列中随机抽取例如10~20帧输出图像,本公开包括但不限于此。

[0135] 步骤S409:响应于检测到至少一帧输出图像中存在满足第一预定条件的输出图像,将满足第一预定条件的输出图像作为目标帧输出图像存储,并存储与目标帧输出图像对应的文本内容,以用于对图像序列进行定位。

[0136] 例如,在一些实施例中,第一预定条件可以是目标帧输出图像对应的文本内容与检测到的前一帧输出图像对应的文本内容的文本相似度超过第一阈值。

[0137] 例如,在一些实施例中,文本相似度可以用归一化编辑距离来计算,当如,文本相似度还可以用TF/IDF(词频/逆文档概率)、余弦距离、Simhash距离等其他算法来计算,本公开的实施例对此不作限制。

[0138] 例如,在一个示例中,第一预定条件可以是目标帧输出图像对应的文本内容与检测到的前一帧输出图像对应的文本内容的归一化编辑距离超过10%。当然,本公开的实施例对第一阈值的设置不作限制,可以根据实际情况调整。

[0139] 例如,在一些实施例中,在基于输入图像确定的第一图像包括至少一个第一图像的情况下,第一预定条件可以是目标帧输出图像包括的第一图像的数量与检测到的前一帧输出图像包括的第一图像的数量之间的差值超过第二阈值。

[0140] 例如,在一些实施例中,当目标帧输出图像包括的第一图像的数量比检测到的前一帧输出图像包括的第一图像的数量多了2个时,对目标帧输出图像以及目标帧输出图像的文本内容进行一次存档,以作为后续用于图像序列定位或搜索的内容。当然,本公开的实施例对第二阈值的设置不作限制,可以根据实际情况调整。

[0141] 因此,通过上述图像处理方法100,除了增强视频图像中的文本内容外,还可以实现对图像序列进行定位或搜索的功能。例如,对于教学视频,可以将老师的板书进行结构化存储(例如,包括板书位置、板书内容、板书所在视频帧),便于用户(例如,学生)通过知识点搜索视频位置。

[0142] 图15为本公开至少一个实施例提供的应用于教学场景的图像处理方法的示意图。例如,在一个示例中,输入教学视频(例如图像序列),从中获取单帧的输入图像;基于该输入图像,采用任意文本检测方法(例如,本公开的实施例所提供的文本检测方法)来确定第一图像;采用任意文本识别方法(例如,本公开的实施例所提供的文本识别方法)来确定第一图像对应的文本内容,从而可以向用户提供教学视频的板书内容检索或定位功能。例如,

还可以去除第一图像中原本的文本内容,得到背景图像,再结合所识别到的文本内容,可以得到重新渲染后的第二图像,将第二图像替代第一图像,可以实现板书内容增强的图像帧,从而增强板书内容的观看质量。

[0143] 由于图15中各个步骤都可以参见上文中对图像处理方法10和100的相关描述,为了简洁,此处不再赘述。

[0144] 需要说明的是,在本公开的实施例中,上述图像处理方法10/100的流程可以包括更多或更少的操作,这些操作可以顺序执行或并行执行。虽然上文描述的图像处理方法的流程包括特定顺序出现的多个操作,但是应该清楚地了解,多个操作的顺序并不受限制。上文描述的图像处理方法可以执行一次,也可以按照预定条件执行多次。

[0145] 本公开的实施例提供的图像处理方法10/100,通过生成包括第二图像的输出图像,可以在保留输入图像的背景的同时,使得输入图像中的文本内容更加规整、更清晰、分辨率更高,从而增强文本内容的观看质量;采用了调整后的密集卷积网络,更大程度地保留了第一图像的特征信息,提高了文本识别的准确性;还可以实现对图像序列进行定位或搜索的功能。

[0146] 本公开至少一个实施例还提供一种图像处理装置。该图像处理装置可以在保留输入图像的背景的同时,使得输入图像中的文本内容更加规整、更清晰、分辨率更高,从而增强输入图像中文本内容的观看质量,还可以提高文本识别的准确性。

[0147] 图16为本公开至少一个实施例提供的一种图像处理装置的示意框图。如图16所示,该图像处理装置20包括获取单元21、确定单元22、特征提取单元23、识别单元24、融合单元25和生成单元26。例如,该图像处理装置20可以应用于视频录制软件、智慧教室软件等应用程序中,也可以应用于任意的需要进行书写和显示的设备或系统中,本公开的实施例对此不作限制。

[0148] 获取单元21配置为获取输入图像。例如,获取单元21可以执行如图2所示的图像处理方法的步骤S101。确定单元22配置为基于输入图像,确定第一图像,第一图像为输入图像中包括文本内容的区域,且文本内容中的文本在第一图像中排列为一行。例如,确定单元22可以执行如图2所示的图像处理方法的步骤S102。特征提取单元23配置为采用密集卷积网络,对第一图像进行特征提取,得到目标特征,该密集卷积网络不包括池化层。例如,特征提取单元23可以执行如图2所示的图像处理方法的步骤S103。识别单元24配置为基于目标特征得到文本内容。例如,识别单元24可以执行如图2所示的图像处理方法的步骤S104。融合单元25配置为基于第一图像和文本内容,得到第二图像。例如,融合单元25可以执行如图2所示的图像处理方法的步骤S105。生成单元26配置为将第二图像替代输入图像中的第一图像,以生成包括第二图像的输出图像。例如,生成单元26可以执行如图2所示的图像处理方法的步骤S106。

[0149] 例如,获取单元21、确定单元22、特征提取单元23、识别单元24、融合单元25和生成单元26可以为硬件、软件、固件以及它们的任意可行的组合。例如,获取单元21、确定单元22、特征提取单元23、识别单元24、融合单元25和生成单元26可以为专用或通用的电路、芯片或装置等,也可以为处理器和存储器的结合。关于获取单元21、确定单元22、特征提取单元23、识别单元24、融合单元25和生成单元26的具体实现形式,本公开的实施例对此不作限制。

[0150] 需要说明的是,本公开的实施例中,图像处理装置20的各个单元与前述的图像处理方法10的各个步骤对应,关于该图像处理装置20的具体功能可以参考上文中图像处理方法的相关描述,此处不再赘述。图16所示的图像处理装置20的组件和结构只是示例性的,而非限制性的,根据需要,该图像处理装置20还可以包括其他组件和结构。例如,在一些示例中,图像处理装置20还可以包括图像序列获取单元、抽取单元和判断单元等。例如,该图像序列获取单元配置为获取包括多帧输出图像的图像序列。该抽取单元配置为抽取图像序列中的至少一帧输出图像。判断单元配置为响应于检测到至少一帧输出图像中存在满足第一预定条件的输出图像,将满足第一预定条件的输出图像作为目标帧输出图像存储,并存储与目标帧输出图像对应的文本内容,以用于对图像序列进行定位。也即是,图像序列获取单元、抽取单元和判断单元可以执行如图14所示的步骤S407-S409。

[0151] 本公开至少一个实施例还提供一种电子设备,该电子设备包括处理器和存储器,一个或多个计算机程序模块被存储在该存储器中并被配置为由该处理器执行,一个或多个计算机程序模块包括用于实现本公开任一实施例提供的图像处理方法的指令。该电子设备可以在保留输入图像的背景的同时,使得输入图像中的文本内容更加规整、更清晰、分辨率更高,从而增强文本内容的观看质量;还可以较大程度地保留了第一图像的特征信息,提高文本识别的准确性。

[0152] 图17为本公开至少一个实施例提供的一种电子设备的示意框图。如图17所示,该电子设备30包括存储器32和处理器31。

[0153] 存储器32用于存储非暂时性计算机可读指令(例如一个或多个计算机程序模块)。处理器31用于运行非暂时性计算机可读指令,非暂时性计算机可读指令被处理器31运行时可以执行上文所述的图像处理方法中的一个或多个步骤。存储器32和处理器31可以通过总线系统和/或其它形式的连接机构(未示出)互连。例如,该电子设备30可以采用Windows、Android等操作系统,根据本公开实施例的图像处理方法通过运行于该操作系统中的应用程序实现,或者通过安装在该操作系统中的浏览器访问云端服务器提供的网站而实现。

[0154] 例如,处理器31可以是中央处理单元(CPU)、图形处理单元(GPU)、数字信号处理器(DSP)或者具有数据处理能力和/或程序执行能力的其它形式的处理单元,例如现场可编程门阵列(FPGA)等;例如,中央处理单元(CPU)可以为X86或ARM架构等。处理器31可以为通用处理器或专用处理器,可以控制电子设备30中的其它组件以执行期望的功能。

[0155] 例如,存储器32可以包括一个或多个计算机程序产品的任意组合,计算机程序产品可以包括各种形式的计算机可读存储介质,例如易失性存储器和/或非易失性存储器。易失性存储器例如可以包括随机存取存储器(RAM)和/或高速缓冲存储器(cache)等。非易失性存储器例如可以包括只读存储器(ROM)、硬盘、可擦除可编程只读存储器(EPROM)、便携式紧致盘只读存储器(CD-ROM)、USB存储器、闪存等。在计算机可读存储介质上可以存储一个或多个计算机程序模块,处理器31可以运行一个或多个计算机程序模块,以实现电子设备30的各种功能。在计算机可读存储介质中还可以存储各种应用程序和各种数据以及应用程序使用和/或产生的各种数据等。

[0156] 需要说明的是,本公开的实施例中,电子设备30的具体功能和技术效果可以参考上文中关于图像处理方法10/100的描述,此处不再赘述。

[0157] 图18为本公开至少一个实施例提供的另一种电子设备的示意框图。该电子设备40

例如适于用来实施本公开实施例提供的图像处理方法。需要注意的是,图18示出的电子设备40仅仅是一个示例,其不会对本公开实施例的功能和使用范围带来任何限制。

[0158] 如图18所示,电子设备40可以包括处理装置(例如中央处理器、图形处理器等)41,其可以根据存储在只读存储器(ROM)42中的程序或者从存储装置48加载到随机访问存储器(RAM)43中的程序而执行各种适当的动作和处理。在RAM 43中,还存储有电子设备40操作所需的各种程序和数据。处理装置41、ROM 42以及RAM 43通过总线44彼此相连。输入/输出(I/O)接口45也连接至总线44。

[0159] 通常,以下装置可以连接至I/O接口45:包括例如触摸屏、触摸板、键盘、鼠标、摄像头、麦克风、加速度计、陀螺仪等的输入装置46;包括例如液晶显示器(LCD)、扬声器、振动器等的输出装置47;包括例如磁带、硬盘等的存储装置48;以及通信装置49。通信装置49可以允许电子设备40与其他电子设备进行无线或有线通信以交换数据。虽然图18示出了具有各种装置的电子设备40,但应理解的是,并不要求实施或具备所有示出的装置,电子设备40可以替代地实施或具备更多或更少的装置。

[0160] 例如,根据本公开的实施例,图2和图14所示的图像处理方法10/100可以被实现为计算机软件程序。例如,本公开的实施例包括一种计算机程序产品,其包括承载在非暂态计算机可读介质上的计算机程序,该计算机程序包括用于执行上述图像处理方法的程序代码。在这样的实施例中,该计算机程序可以通过通信装置49从网络上被下载和安装,或者从存储装置48安装,或者从ROM 42安装。在该计算机程序被处理装置41执行时,可以实现本公开实施例提供的图像处理方法中限定的功能。

[0161] 本公开至少一个实施例还提供一种存储介质,用于存储非暂时性计算机可读指令,当该非暂时性计算机可读指令由计算机执行时可以实现本公开任一实施例所述的图像处理方法。利用该存储介质,可以在保留输入图像的背景的同时,使得输入图像中的文本内容更加规整、更清晰、分辨率更高,从而增强输入图像中的文本内容的观看质量,还可以较大程度地保留第一图像的特征信息,提高文本识别的准确性。

[0162] 图19为本公开至少一个实施例提供的一种存储介质的示意图。如图19所示,存储介质50用于存储非暂时性计算机可读指令51。例如,当非暂时性计算机可读指令51由计算机执行时可以执行根据上文所述的图像处理方法10/100中的一个或多个步骤。

[0163] 例如,该存储介质50可以应用于上述电子设备中。例如,存储介质50可以为图17所示的电子设备30中的存储器32。例如,关于存储介质50的相关说明可以参考图17所示的电子设备30中的存储器32的相应描述,此处不再赘述。

[0164] 有以下几点需要说明:

[0165] (1) 本公开实施例附图只涉及到本公开实施例涉及到的结构,其他结构可参考通常设计。

[0166] (2) 在不冲突的情况下,本公开的实施例及实施例中的特征可以相互组合以得到新的实施例。

[0167] 以上所述,仅为本公开的具体实施方式,但本公开的保护范围并不局限于此,本公开的保护范围应以所述权利要求的保护范围为准。



图1

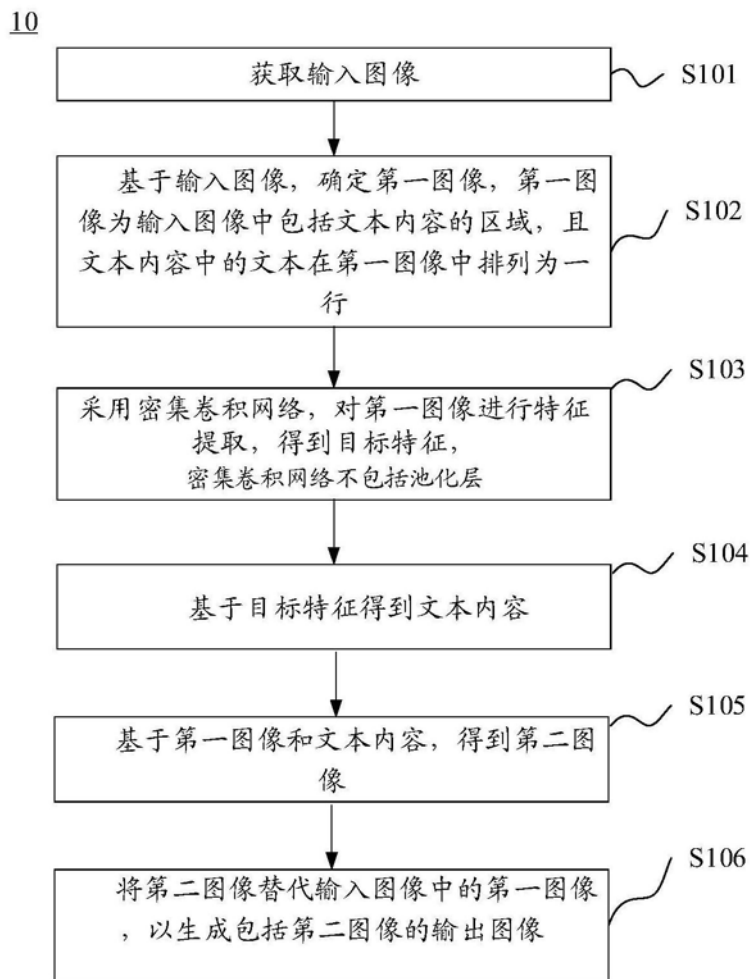


图2



图3

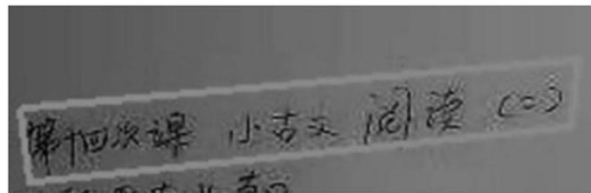


图4

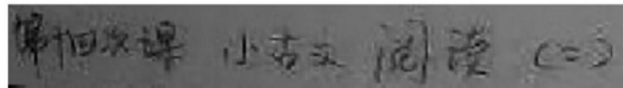


图5

第十四次课 小古文 阅读 (二)

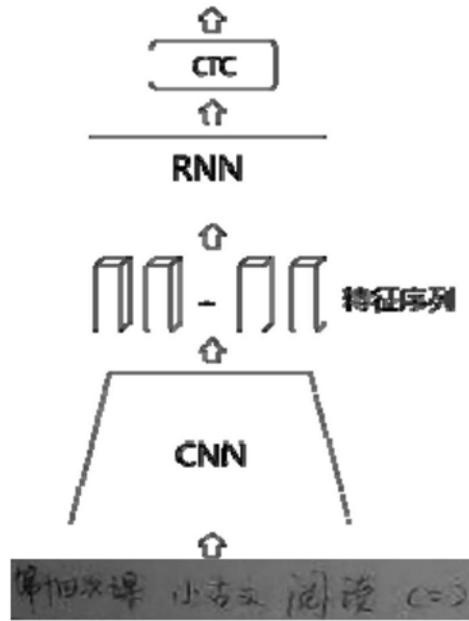


图6

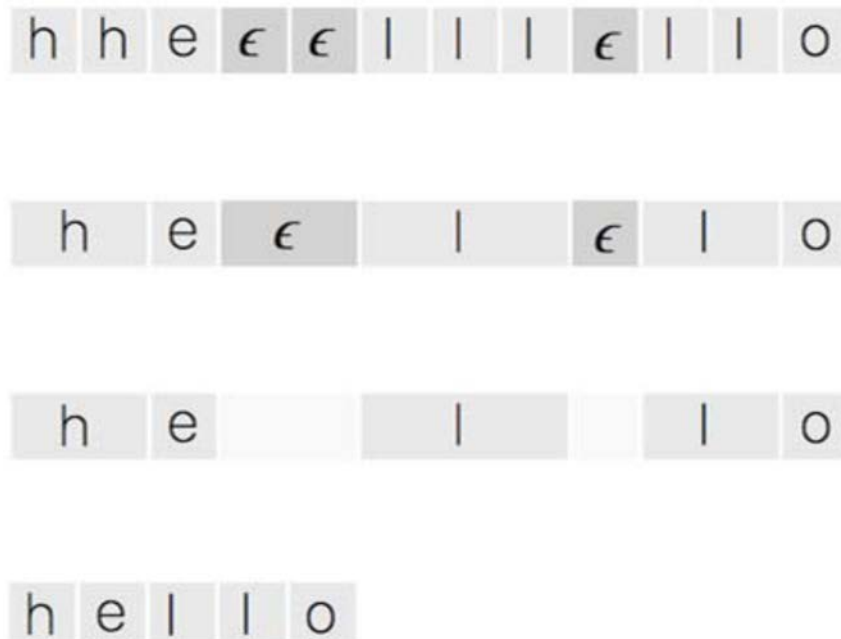


图7

	网络模块	常规 DenseNet-121		DenseNet	
		输出维度	模型操作	输出维度	模型操作
1	卷积层	112x112	Conv,stride2	112x16	Conv,stride2
2	池化层	56x56	Stride2	无	无
3	密集块 1	56x56	Conv	112x16	Conv
4	过渡层 1	28x28	Conv,stride(2,2)	112x8	Conv,stride(1,2)
5	密集块 2	28x28	Conv	112x8	Conv
6	过渡层 2	14x14	Conv,stride(2,2)	56x4	Conv,stride(2,2)
7	密集块 3	14x14	Conv	56x4	Conv
8	过渡层 3	7x7	Conv,stride(2,2)	56x2	Conv,stride(1,2)
9	过渡层 4			56x1	Conv,stride(1,2)

图8



图9



图10

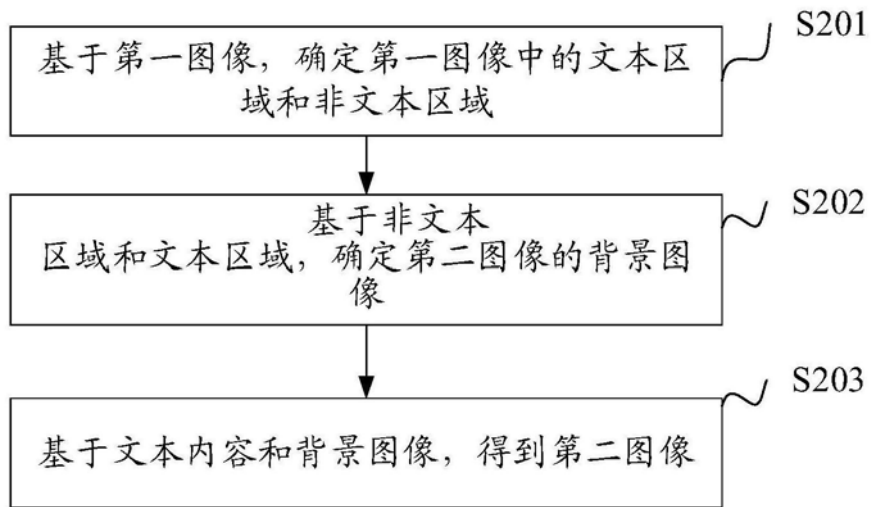


图11



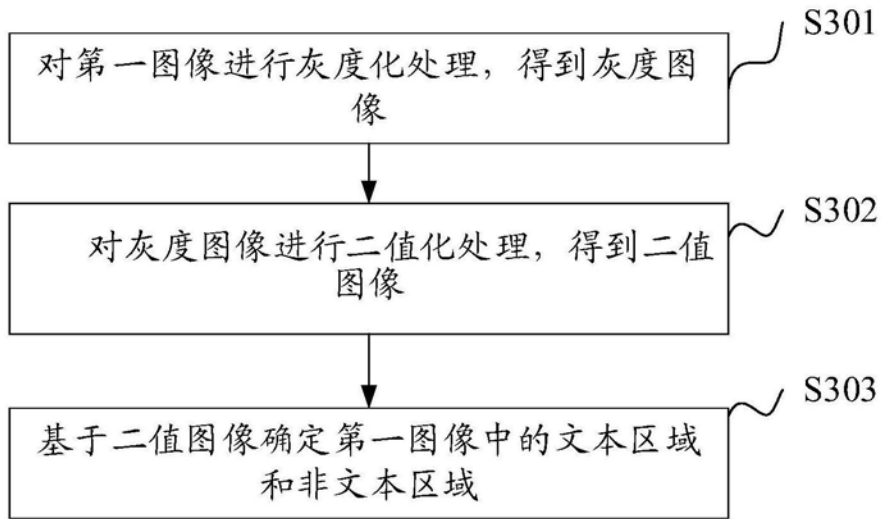


图12

第十四次课 小古文阅读（二）

图13

100

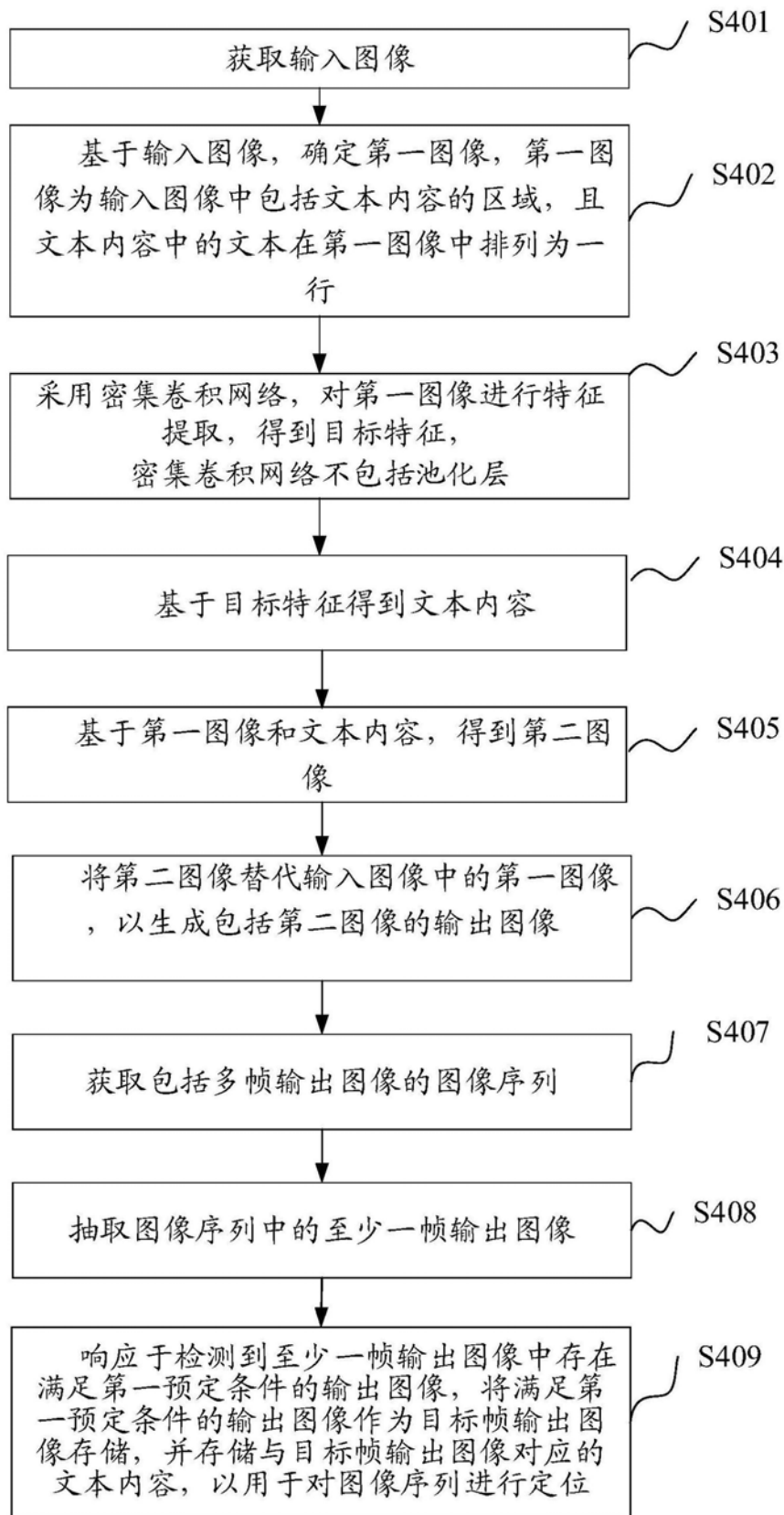


图14

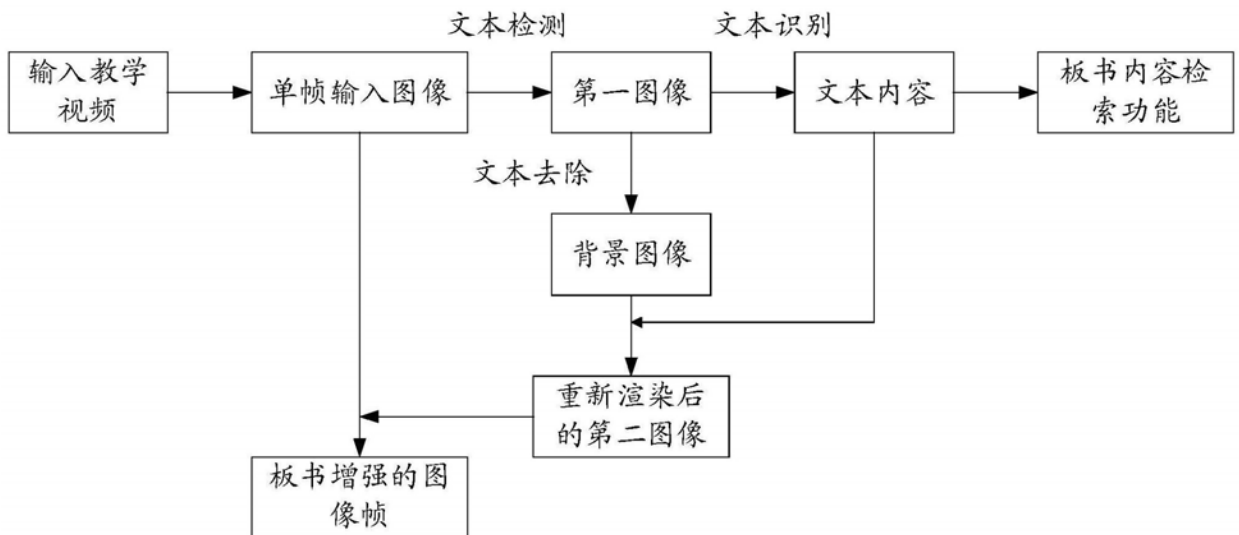


图15

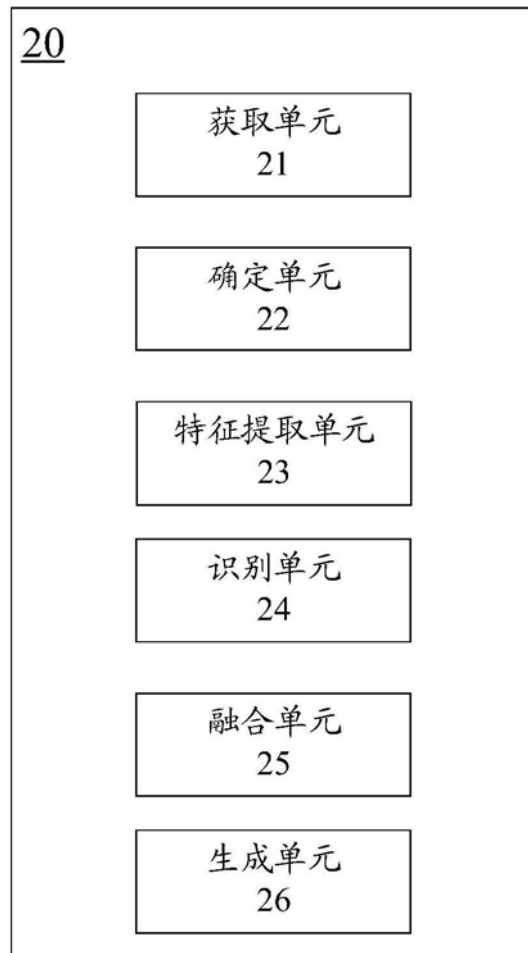


图16

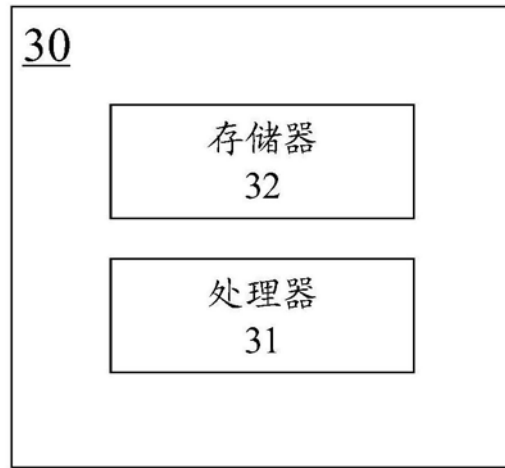


图17

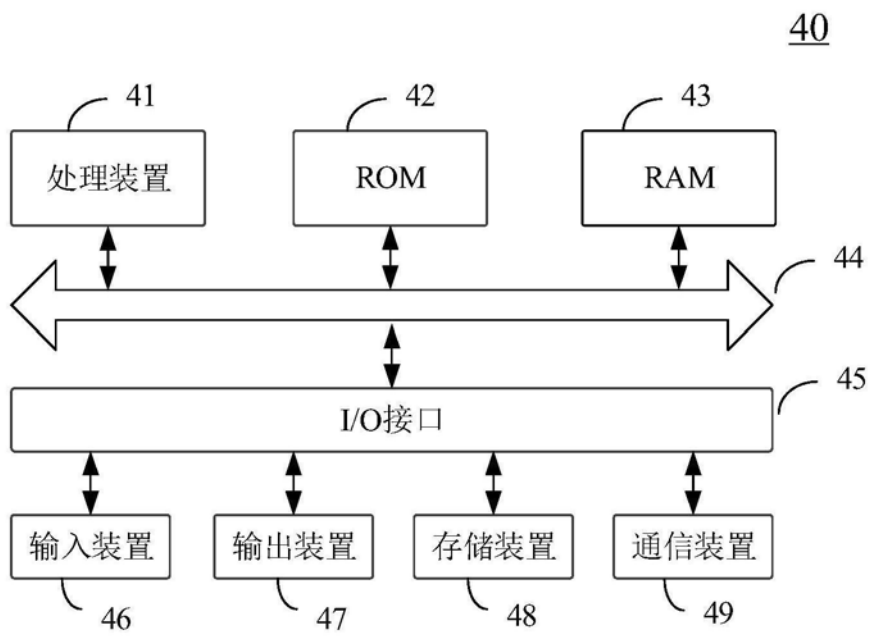


图18

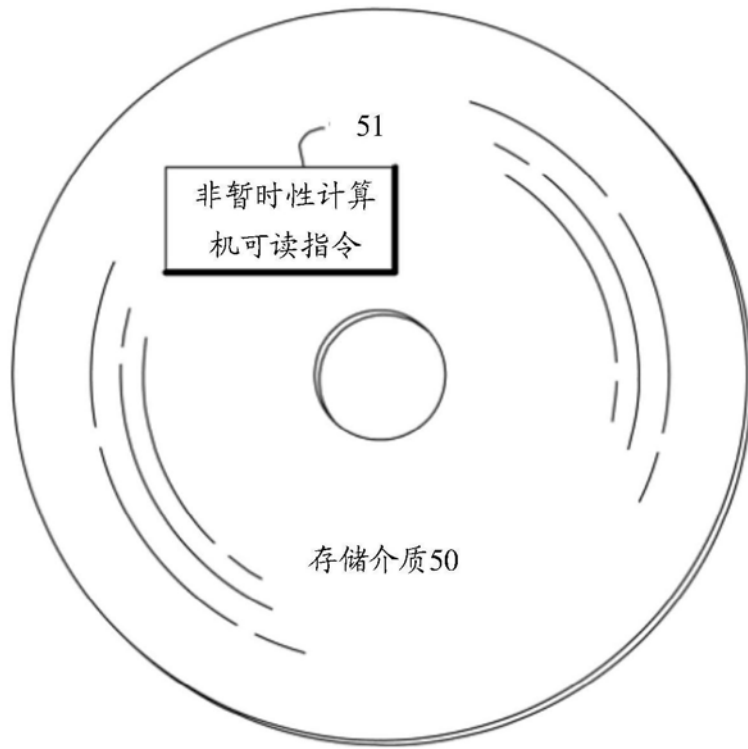


图19