



(12)发明专利申请

(10)申请公布号 CN 111368840 A

(43)申请公布日 2020.07.03

(21)申请号 202010104574.9

(22)申请日 2020.02.20

(71)申请人 中国建设银行股份有限公司
地址 100033 北京市西城区金融大街25号
申请人 建信金融科技有限责任公司

(72)发明人 吴昀蓁 熊博颖 易显维

(74)专利代理机构 中原信达知识产权代理有限
责任公司 11219
代理人 张一军 张效荣

(51) Int. Cl.
G06K 9/34(2006.01)
G06K 9/62(2006.01)

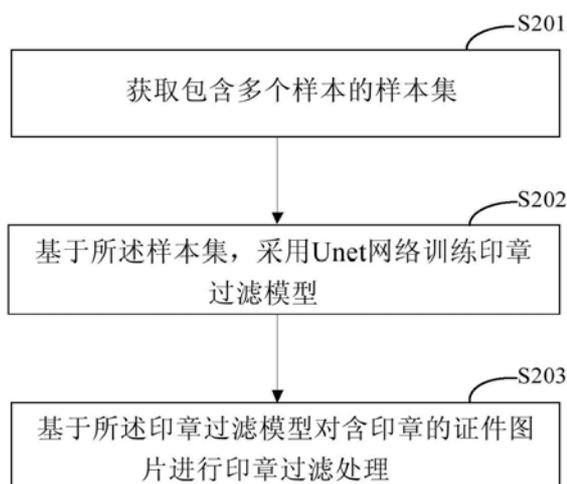
权利要求书2页 说明书9页 附图4页

(54)发明名称

证件图片处理的方法和装置

(57)摘要

本发明公开了证件图片处理的方法和装置，涉及计算机技术领域。该方法的一具体实施方式包括：获取包含多个样本的样本集；基于所述样本集，采用Unet网络训练印章过滤模型；基于所述印章过滤模型对含印章的证件图片进行印章过滤处理。该实施方式能够在不破坏证件图片中信息的前提下去除证件图片中的印章，便于后续识别证件图片中的栏位。



1. 一种证件图片处理的方法,其特征在于,包括:

获取包含多个样本的样本集;

基于所述样本集,采用Unet网络训练印章过滤模型;

基于所述印章过滤模型对含印章的证件图片进行印章过滤处理。

2. 如权利要求1所述的方法,其特征在于,获取包含多个样本的样本集,包括:获取多个不含印章的证件图片,在每个所述不含印章的证件图片中添加印章得到含印章的证件图片;

基于所述样本集,采用Unet网络训练训练印章过滤模型,包括:以所述含印章的证件图片作为模型输入、以所述不含印章的证件图片作为模型输出,训练印章过滤模型。

3. 如权利要求1所述的方法,其特征在于,在每个所述不含印章的证件图片中添加印章得到含印章的证件图片,包括:

采用PS软件从印章模板图片中截取印章图片,利用opencv程序将所述印章图片添加到每个不含印章的证件图片中,得到与每个不含印章的证件图片对应的含印章的证件图片。

4. 如权利要求1所述的方法,其特征在于,对含印章的证件图片进行印章过滤处理之后,还包括:

对证件图片中的所有像素点进行聚类,得到背景类像素点和前景类像素点,将背景类像素点的像素值置为0、前景类像素点的像素值置为1得到二值化图像;

采用连通域算法确定所述二值化图像中的各个连通域以及每个所述连通域的位置信息;

根据预设的证件配置信息以及每个所述连通域的位置信息,确定每个所述连通域对应的栏位。

5. 如权利要求4所述的方法,其特征在于,采用K-means算法对证件图片中的所有像素点进行聚类;所述连通域算法为:Two-Pass方法或Seed-Filling方法。

6. 如权利要求4所述的方法,其特征在于,所述证件配置信息中包括:所述证件图片中各个栏位的栏位位置信息;

根据预设的证件配置信息以及每个所述连通域的位置信息,确定每个所述连通域对应的栏位,包括:对于任一连通域,将与所述任一连通域的位置信息匹配的栏位位置信息所对应的栏位,作为所述任一连通域的栏位。

7. 如权利要求1-6任一所述的方法,其特征在于,所述证件为身份证。

8. 一种证件图片处理的装置,其特征在于,包括:

样本获取模块,获取包含多个样本的样本集;

模型训练模块,基于所述样本集,采用Unet网络训练印章过滤模型;

印章过滤模块,基于所述印章过滤模型对含印章的证件图片进行印章过滤处理。

9. 如权利要求8所述的装置,其特征在于,所述样本获取模块获取包含多个样本的样本集,包括:获取多个不含印章的证件图片,在每个所述不含印章的证件图片中添加印章得到含印章的证件图片;

所述模型训练模块基于所述样本集,采用Unet网络训练训练印章过滤模型,包括:以所述含印章的证件图片作为模型输入、以所述不含印章的证件图片作为模型输出,训练印章过滤模型。

10. 如权利要求8所述的装置,其特征在於,所述样本获取模块在每个所述不含印章的证件图片中添加印章得到含印章的证件图片,包括:

采用PS软件从印章模板图片中截取印章图片,利用opencv程序将所述印章图片添加到每个不含印章的证件图片中,得到与每个不含印章的证件图片对应的含印章的证件图片。

11. 如权利要求8所述的装置,其特征在於,还包括:栏位识别模块,用于在对含印章的证件图片进行印章过滤处理之后,

对证件图片中的所有像素点进行聚类,得到背景类像素点和前景类像素点,将背景类像素点的像素值置为0、前景类像素点的像素值置为1得到二值化图像;

采用连通域算法确定所述二值化图像中的各个连通域以及每个所述连通域的位置信息;

根据预设的证件配置信息以及每个所述连通域的位置信息,确定每个所述连通域对应的栏位。

12. 如权利要求11所述的装置,其特征在於,所述栏位识别模块采用K-means算法对证件图片中的所有像素点进行聚类;所述连通域算法为:Two-Pass方法或Seed-Filling方法。

13. 如权利要求11所述的装置,其特征在於,所述证件配置信息中包括:所述证件图片中各个栏位的栏位位置信息;

所述栏位识别模块根据预设的证件配置信息以及每个所述连通域的位置信息,确定每个所述连通域对应的栏位,包括:对于任一连通域,将与所述任一连通域的位置信息匹配的栏位位置信息所对应的栏位,作为所述任一连通域的栏位。

14. 如权利要求8-13任一所述的装置,其特征在於,所述证件为身份证。

15. 一种证件图片处理的电子设备,其特征在於,包括:

一个或多个处理器;

存储装置,用于存储一个或多个程序,

当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现如权利要求1-7中任一所述的方法。

16. 一种计算机可读介质,其上存储有计算机程序,其特征在於,所述程序被处理器执行时实现如权利要求1-7中任一所述的方法。

证件图片处理的方法和装置

技术领域

[0001] 本发明涉及计算机技术领域,尤其涉及一种证件图片处理的方法和装置。

背景技术

[0002] 证件图片中含有印章时,若印章覆盖在证件的栏位上,则在做证识别之前需要去除证件图片中加盖的印章。现有技术主要使用目标检测来确定证件图片中的印章位置,在图像识别之前根据识别出的印章位置抹掉证件图片上的印章。

[0003] 在实现本发明过程中,发明人发现现有技术中至少存在如下问题:

[0004] 往往会破坏证件图片中的信息,导致印章位置在经过图像处理之后会模糊不清。

发明内容

[0005] 有鉴于此,本发明实施例提供一种证件图片处理的方法和装置,能够在不破坏证件图片中信息的前提下去除证件图片中的印章,便于后续识别证件图片中的栏位。

[0006] 为实现上述目的,根据本发明实施例的一个方面,提供了一种证件图片处理的方法,包括:

[0007] 获取包含多个样本的样本集;

[0008] 基于所述样本集,采用Unet网络训练印章过滤模型;

[0009] 基于所述印章过滤模型对含印章的证件图片进行印章过滤处理。

[0010] 可选地,获取包含多个样本的样本集,包括:获取多个不含印章的证件图片,在每个所述不含印章的证件图片中添加印章得到含印章的证件图片;

[0011] 基于所述样本集,采用Unet网络训练训练印章过滤模型,包括:以所述含印章的证件图片作为模型输入、以所述不含印章的证件图片作为模型输出,训练印章过滤模型。

[0012] 可选地,在每个所述不含印章的证件图片中添加印章得到含印章的证件图片,包括:

[0013] 采用PS软件从印章模板图片中截取印章图片,利用opencv程序将所述印章图片添加到每个不含印章的证件图片中,得到与每个不含印章的证件图片对应的含印章的证件图片。

[0014] 可选地,对含印章的证件图片进行印章过滤处理之后,还包括:

[0015] 对证件图片中的所有像素点进行聚类,得到背景类像素点和前景类像素点,将背景类像素点的像素值置为0、前景类像素点的像素值置为1得到二值化图像;

[0016] 采用连通域算法确定所述二值化图像中的各个连通域以及每个所述连通域的位置信息;

[0017] 根据预设的证件配置信息以及每个所述连通域的位置信息,确定每个所述连通域对应的栏位。

[0018] 可选地,采用K-means算法对证件图片中的所有像素点进行聚类;所述连通域算法为:Two-Pass方法或Seed-Filling方法。

- [0019] 可选地,所述证件配置信息中包括:所述证件图片中各个栏位的栏位位置信息;
- [0020] 根据预设的证件配置信息以及每个所述连通域的位置信息,确定每个所述连通域对应的栏位,包括:对于任一连通域,将与所述任一连通域的位置信息匹配的栏位位置信息所对应的栏位,作为所述任一连通域的栏位。
- [0021] 可选地,所述证件为身份证。
- [0022] 根据本发明实施例的第二方面,提供一种证件图片处理的装置,包括:
- [0023] 样本获取模块,获取包含多个样本的样本集;
- [0024] 模型训练模块,基于所述样本集,采用Unet网络训练印章过滤模型;
- [0025] 印章过滤模块,基于所述印章过滤模型对含印章的证件图片进行印章过滤处理。
- [0026] 可选地,所述样本获取模块获取包含多个样本的样本集,包括:获取多个不含印章的证件图片,在每个所述不含印章的证件图片中添加印章得到含印章的证件图片;
- [0027] 所述模型训练模块基于所述样本集,采用Unet网络训练训练印章过滤模型,包括:以所述含印章的证件图片作为模型输入、以所述不含印章的证件图片作为模型输出,训练印章过滤模型。
- [0028] 可选地,所述样本获取模块在每个所述不含印章的证件图片中添加印章得到含印章的证件图片,包括:
- [0029] 采用PS软件从印章模板图片中截取印章图片,利用opencv程序将所述印章图片添加到每个不含印章的证件图片中,得到与每个不含印章的证件图片对应的含印章的证件图片。
- [0030] 可选地,本发明实施例的装置还包括:栏位识别模块,用于在对含印章的证件图片进行印章过滤处理之后,
- [0031] 对证件图片中的所有像素点进行聚类,得到背景类像素点和前景类像素点,将背景类像素点的像素值置为0、前景类像素点的像素值置为1得到二值化图像;
- [0032] 采用连通域算法确定所述二值化图像中的各个连通域以及每个所述连通域的位置信息;
- [0033] 根据预设的证件配置信息以及每个所述连通域的位置信息,确定每个所述连通域对应的栏位。
- [0034] 可选地,所述栏位识别模块采用K-means算法对证件图片中的所有像素点进行聚类;所述连通域算法为:Two-Pass方法或Seed-Filling方法。
- [0035] 可选地,所述证件配置信息中包括:所述证件图片中各个栏位的栏位位置信息;
- [0036] 所述栏位识别模块根据预设的证件配置信息以及每个所述连通域的位置信息,确定每个所述连通域对应的栏位,包括:对于任一连通域,将与所述任一连通域的位置信息匹配的栏位位置信息所对应的栏位,作为所述任一连通域的栏位。
- [0037] 可选地,所述证件为身份证。
- [0038] 根据本发明实施例的第三方面,提供一种证件图片处理的电子设备,包括:
- [0039] 一个或多个处理器;
- [0040] 存储装置,用于存储一个或多个程序,
- [0041] 当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现本发明实施例第一方面提供的方法。

[0042] 根据本发明实施例的第四方面,提供一种计算机可读介质,其上存储有计算机程序,所述程序被处理器执行时实现本发明实施例第一方面提供的方法。

[0043] 上述发明中的一个实施例具有如下优点或有益效果:采用Unet网络训练印章过滤模型、基于训练得到的印章过滤模型对含印章的证件图片进行印章过滤处理,能够在不破坏证件图片中信息的前提下去除证件图片中的印章,便于后续识别证件图片中的栏位。

[0044] 上述的非惯用的可选方式所具有的进一步效果将在下文中结合具体实施方式加以说明。

附图说明

[0045] 附图用于更好地理解本发明,不构成对本发明的不当限定。其中:

[0046] 图1是本发明可选实施例中证件图片处理的方法的主要流程示意图;

[0047] 图2是本发明可选实施例中进行印章过滤处理的主要流程示意图;

[0048] 图3是本发明可选实施例中Unet网络结构的示意图;

[0049] 图4是本发明可选实施例中进行栏位识别的主要流程示意图;

[0050] 图5是本发明实施例的证件图片处理的装置的主要模块的示意图;

[0051] 图6是本发明实施例可以应用于其中的示例性系统架构图;

[0052] 图7是适于用来实现本发明实施例的终端设备或服务器的计算机系统的结构示意图。

具体实施方式

[0053] 以下结合附图对本发明的示范性实施例做出说明,其中包括本发明实施例的各种细节以助于理解,应当将它们认为仅仅是示范性的。因此,本领域普通技术人员应当认识到,可以对这里描述的实施例做出各种改变和修改,而不会背离本发明的范围和精神。同样,为了清楚和简明,以下的描述中省略了对公知功能和结构的描述。

[0054] 根据本发明实施例的一个方面,提供了一种证件图片处理的方法。

[0055] 本发明实施例的证件图片处理方法,可以包括对证件图片进行印章过滤处理,也可以包括对证件图片进行印章过滤处理和栏位识别处理。

[0056] 图1是本发明可选实施例中证件图片处理的方法的主要流程示意图,如图1所示,证件图片处理的方法包括:步骤S101、获取含印章的证件图片;步骤S102、对含印章的证件图片进行印章过滤处理;步骤S103、对步骤S102处理后的证件图片进行栏位识别处理。

[0057] 以下结合图2对进行印章过滤处理的主要流程进行示例性说明。如图2所示,包括:步骤S201、步骤S202、步骤S203。

[0058] 步骤S201、获取包含多个样本的样本集。

[0059] 样本用于训练印章过滤模型,每个样本包括两张图片:不含印章的证件图片和含印章的证件图片。

[0060] 可选地,获取包含多个样本的样本集,包括:获取多个不含印章的证件图片,在每个所述不含印章的证件图片中添加印章得到含印章的证件图片。如此,能够保证同一个样本中含印章的证件图片与不含印章的证件图片的一致性,进而提高模型训练的准确性。

[0061] 可选地,在每个所述不含印章的证件图片中添加印章得到含印章的证件图片,包

括:采用PS (Photoshop, 一个图像处理软件) 软件从印章模板图片中截取印章图片, 利用opencv (开源计算机视觉库) 程序将所述印章图片添加到每个不含印章的证件图片中, 得到与每个不含印章的证件图片对应的含印章的证件图片。采用PS软件获取印章图片, 能够提高印章图片的清晰度。采用opencv程序将印章图片添加到每个不含印章的证件图片中, 能够提高样本获取的速度。

[0062] 步骤S202、基于所述样本集, 采用Unet网络 (基于FCNs的分割网络) 训练印章过滤模型。采用Unet网络训练印章过滤模型, 准确性好。

[0063] 可选地, 采用Unet网络训练印章过滤模型包括: 以所述含印章的证件图片作为模型输入、以所述不含印章的证件图片作为模型输出, 训练印章过滤模型。本实施例从模型的输入端输入有印章的证件图片, 将无印章的证件图片作为学习目标训练印章过滤模型, 能够提高模型训练的准确性。

[0064] Unet网络的具体结构可以根据实际情况进行选择设定, 本发明实施例对此不作限定。在图3示出的可选实施例中, 使用keras (一个高层神经网络API) 深度学习框架对Unet网络进行描述, 具体的算子表达形式及代码示例性为:

```
[0065] inpt=Input(shape=(input_size_1,input_size_2,3))
[0066] conv1=Conv2d_BN(inpt,8,(3,3))
[0067] conv1=Conv2d_BN(conv1,8,(3,3))
[0068] pool1=MaxPooling2D(pool_size=(2,2),strides=(2,2),padding='same')
(conv1)
[0069] conv2=Conv2d_BN(pool1,16,(3,3))
[0070] conv2=Conv2d_BN(conv2,16,(3,3))
[0071] pool2=MaxPooling2D(pool_size=(2,2),strides=(2,2),padding='same')
(conv2)
[0072] conv3=Conv2d_BN(pool2,32,(3,3))
[0073] conv3=Conv2d_BN(conv3,32,(3,3))
[0074] pool3=MaxPooling2D(pool_size=(2,2),strides=(2,2),padding='same')
(conv3)
[0075] conv4=Conv2d_BN(pool3,64,(3,3))
[0076] conv4=Conv2d_BN(conv4,64,(3,3))
[0077] pool4=MaxPooling2D(pool_size=(2,2),strides=(2,2),padding='same')
(conv4)
[0078] conv5=Conv2d_BN(pool4,128,(3,3))
[0079] #conv5=Dropout(0.1)(conv5)
[0080] conv5=Conv2d_BN(conv5,128,(3,3))
[0081] #conv5=Dropout(0.1)(conv5)
[0082] convt1=Conv2dT_BN(conv5,64,(3,3))
[0083] concat1=concatenate([conv4,convt1],axis=3)
[0084] #concat1=Dropout(0.1)(concat1)
[0085] conv6=Conv2d_BN(concat1,64,(3,3))
```

```
[0086] conv6=Conv2d_BN(conv6,64,(3,3))
[0087] convt2=Conv2dT_BN(conv6,32,(3,3))
[0088] concat2=concatenate([conv3,convt2],axis=3)
[0089] #concat2=Dropout(0.1)(concat2)
[0090] conv7=Conv2d_BN(concat2,32,(3,3))
[0091] conv7=Conv2d_BN(conv7,32,(3,3))
[0092] convt3=Conv2dT_BN(conv7,16,(3,3))
[0093] concat3=concatenate([conv2,convt3],axis=3)
[0094] #concat3=Dropout(0.1)(concat3)
[0095] conv8=Conv2d_BN(concat3,16,(3,3))
[0096] conv8=Conv2d_BN(conv8,16,(3,3))
[0097] convt4=Conv2dT_BN(conv8,8,(3,3))
[0098] concat4=concatenate([conv1,convt4],axis=3)
[0099] #concat4=Dropout(0.1)(concat4)
[0100] conv9=Conv2d_BN(concat4,8,(3,3))
[0101] conv9=Conv2d_BN(conv9,8,(3,3))
[0102] #conv9=Dropout(0.1)(conv9)
[0103] outpt=Conv2D(filters=3,kernel_size=(1,1),strides=(1,1),padding='same',activation='sigmoid')(conv9)
```

[0104] 步骤S203、基于所述印章过滤模型对含印章的证件图片进行印章过滤处理。基于步骤S202训练得到的印章过滤模型对含印章的证件图片进行印章过滤处理，能够在不破坏证件图片中信息的前提下去除证件图片中的印章，便于后续识别证件图片中的栏位。

[0105] 在可选的实施例中，进行印章过滤处理之后，还可以进一步进行栏位识别处理。栏位是指证件中的信息项。栏位识别处理即识别证件图片中各个栏位的信息。

[0106] 以下结合图4对进行栏位识别处理的主要流程进行示例性说明。如图4所示，在对含印章的证件图片进行印章过滤处理之后，进行栏位识别处理的流程包括：

[0107] 步骤S401、对证件图片中的所有像素点进行聚类，得到背景类像素点和前景类像素点，将背景类像素点的像素值置为0、前景类像素点的像素值置为1得到二值化图像；

[0108] 步骤S402、采用连通域算法确定所述二值化图像中的各个连通域以及每个所述连通域的位置信息；

[0109] 步骤S403、根据预设的证件配置信息以及每个所述连通域的位置信息，确定每个所述连通域对应的栏位。

[0110] 将物理或抽象对象的集合分成由类似的对象组成的多个类的过程称为聚类。通过聚类和二值化处理，能够避免由于证件图片曝光过度而导致的无法识别或识别准确性低的情况发生。

[0111] 聚类方法可以根据实际情况进行选择设定，例如K-medoids算法、CLARANS算法。可选地，采用K-means算法对证件图片中的所有像素点进行聚类。K-means算法的步骤为：(1) 选择初始化的k个样本作为初始聚类中心；(2) 针对数据集中每个样本计算它到k个聚类中心的距离并将其分到距离最小的聚类中心所对应的类中；(3) 针对每个类别，重新计算它

的聚类中心(即属于该类的所有样本的质心);(4)重复上面(2)、(3)两步操作,直到达到中止条件(迭代次数、最小误差变化等)。K-means算法的聚类效果好,处理大数据集的时候可以保证较好的伸缩性,且算法复杂度低。

[0112] 连通域(Connected Component)一般是指图像中具有相同像素值且位置相邻的前景像素点组成的图像区域(Region, Blob)。连通域算法是指用于确定出连通域的算法。连通域算法可以根据实际情况进行选择设定,可选地,所述连通域算法为:Two-Pass方法或Seed-Filling方法。采用连通域法能够快速确定各个证件图片中的各个连通域。每个连通域对应一个栏位。

[0113] 栏位一般包括栏位名称以及栏位内容。以身份证中的姓名栏为例,“姓名”为栏位名称,“姓名”之后的“XXX”为栏位内容;以身份证中的性别栏为例,“性别”为栏位名称,“男”或“女”为栏位内容。本发明实施例在确定各个连通域时,可以获取包括栏位名称和栏位内容的连通域,也可以进获取包括栏位内容的连通域。

[0114] 证件配置信息是指用于描述证件图片中各个栏位的位置的信息。以身份证的正面图片为例,证件配置信息可以为:第一行为姓名栏,第二行左侧为性别栏、右侧为民族栏,第三行左侧为出生年栏、中间为出生月栏、右侧为出生日栏,第四行为住址栏,第五行为公民身份证号码栏。本发明实施例的栏位识别方法尤其适用于具有特定格式的证件,例如身份证、户口本、银行卡、护照、车牌等。

[0115] 可选地,所述证件配置信息中包括:所述证件图片中各个栏位的栏位位置信息。根据预设的证件配置信息以及每个所述连通域的位置信息,确定每个所述连通域对应的栏位,包括:对于任一连通域,将与所述任一连通域的位置信息匹配的栏位位置信息所对应的栏位,作为所述任一连通域的栏位。

[0116] 示例性地,以前文中的证件配置信息为例,基于预设的证件配置信息,可以将第一行的连通域作为姓名栏,第二行左侧的连通域作为性别栏、右侧的连通域作为民族栏,第三行左侧的连通域作为出生年栏、中间的连通域作为出生月栏、右侧的连通域作为出生日栏,第四行的连通域作为住址栏,第五行的连通域作为公民身份证号码栏。如此,可以根据识别出的栏位进行后续分析,例如分析各个栏位内的内容。

[0117] 本发明实施例的栏位识别方法通过对证件图片中的所有像素点进行聚类、根据聚类得到的二值化图片进行栏位识别,能够避免由于图片曝光较高而导致的栏位识别不准确或无法识别的情况发生,提高证件栏位识别的准确性。

[0118] 本发明实施例的证件可以是身份证、户口本、银行卡、护照、车牌等。证件图片是指证件的复印件或扫描件等。

[0119] 根据本发明实施例的第二方面,提供一种实现上述方法的装置。

[0120] 图5是本发明实施例的证件图片处理的装置的主要模块的示意图,如图5所示,证件图片处理的装置500包括:

[0121] 样本获取模块501,获取包含多个样本的样本集;

[0122] 模型训练模块502,基于所述样本集,采用Unet网络训练印章过滤模型;

[0123] 印章过滤模块503,基于所述印章过滤模型对含印章的证件图片进行印章过滤处理。

[0124] 可选地,所述样本获取模块获取包含多个样本的样本集,包括:获取多个不含印章

的证件图片,在每个所述不含印章的证件图片中添加印章得到含印章的证件图片;

[0125] 所述模型训练模块基于所述样本集,采用Unet网络训练印章过滤模型,包括:以所述含印章的证件图片作为模型输入、以所述不含印章的证件图片作为模型输出,训练印章过滤模型。

[0126] 可选地,所述样本获取模块在每个所述不含印章的证件图片中添加印章得到含印章的证件图片,包括:

[0127] 采用PS软件从印章模板图片中截取印章图片,利用opencv程序将所述印章图片添加到每个不含印章的证件图片中,得到与每个不含印章的证件图片对应的含印章的证件图片。

[0128] 可选地,本发明实施例的装置还包括:栏位识别模块,用于在对含印章的证件图片进行印章过滤处理之后,

[0129] 对证件图片中的所有像素点进行聚类,得到背景类像素点和前景类像素点,将背景类像素点的像素值置为0、前景类像素点的像素值置为1得到二值化图像;

[0130] 采用连通域算法确定所述二值化图像中的各个连通域以及每个所述连通域的位置信息;

[0131] 根据预设的证件配置信息以及每个所述连通域的位置信息,确定每个所述连通域对应的栏位。

[0132] 可选地,所述栏位识别模块采用K-means算法对证件图片中的所有像素点进行聚类;所述连通域算法为:Two-Pass方法或Seed-Filling方法。

[0133] 可选地,所述证件配置信息中包括:所述证件图片中各个栏位的栏位位置信息;

[0134] 所述栏位识别模块根据预设的证件配置信息以及每个所述连通域的位置信息,确定每个所述连通域对应的栏位,包括:对于任一连通域,将与所述任一连通域的位置信息匹配的栏位位置信息所对应的栏位,作为所述任一连通域的栏位。

[0135] 可选地,所述证件为身份证。

[0136] 根据本发明实施例的第三方面,提供一种证件图片处理的电子设备,包括:

[0137] 一个或多个处理器;

[0138] 存储装置,用于存储一个或多个程序,

[0139] 当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现本发明实施例第一方面提供的方法。

[0140] 根据本发明实施例的第四方面,提供一种计算机可读介质,其上存储有计算机程序,所述程序被处理器执行时实现本发明实施例第一方面提供的方法。

[0141] 图6示出了可以应用本发明实施例的证件图片处理的方法或证件图片处理的装置的示例性系统架构600。

[0142] 如图5所示,系统架构500可以包括终端设备601、602、603,网络604和服务器605。网络504用以在终端设备601、602、603和服务器605之间提供通信链路的介质。网络604可以包括各种连接类型,例如有线、无线通信链路或者光纤电缆等等。

[0143] 用户可以使用终端设备601、602、603通过网络604与服务器605交互,以接收或发送消息等。终端设备601、602、603上可以安装有各种通讯客户端应用,例如购物类应用、网页浏览器应用、搜索类应用、即时通信工具、邮箱客户端、社交平台软件等(仅为示例)。

[0144] 终端设备601、602、603可以是具有显示屏并且支持网页浏览的各种电子设备,包括但不限于智能手机、平板电脑、膝上型便携计算机和台式计算机等等。

[0145] 服务器605可以是提供各种服务的服务器,例如对用户利用终端设备601、602、603所浏览的购物类网站提供支持的后台管理服务器(仅为示例)。后台管理服务器可以对接收到的印章过滤请求或栏位识别请求等数据进行分析等处理,并将处理结果(例如去除印章的证件图片、栏位识别结果信息—仅为示例)反馈给终端设备。

[0146] 需要说明的是,本发明实施例所提供的证件图片处理的方法一般由服务器605执行,相应地,证件图片处理的装置一般设置于服务器605中。

[0147] 应该理解,图6中的终端设备、网络和服务器的数目仅仅是示意性的。根据实现需要,可以具有任意数目的终端设备、网络和服务器。

[0148] 下面参考图7,其示出了适于用来实现本发明实施例的终端设备的计算机系统700的结构示意图。图7示出的终端设备仅仅是一个示例,不应对本发明实施例的功能和使用范围带来任何限制。

[0149] 如图7所示,计算机系统700包括中央处理单元(CPU)701,其可以根据存储在只读存储器(ROM)702中的程序或者从存储部分708加载到随机访问存储器(RAM)703中的程序而执行各种适当的动作和处理。在RAM 703中,还存储有系统700操作所需的各种程序和数据。CPU 701、ROM 702以及RAM 703通过总线704彼此相连。输入/输出(I/O)接口705也连接至总线704。

[0150] 以下部件连接至I/O接口705:包括键盘、鼠标等的输入部分706;包括诸如阴极射线管(CRT)、液晶显示器(LCD)等以及扬声器等的输出部分707;包括硬盘等的存储部分708;以及包括诸如LAN卡、调制解调器等网络接口卡的通信部分709。通信部分709经由诸如因特网的网络执行通信处理。驱动器710也根据需要连接至I/O接口705。可拆卸介质711,诸如磁盘、光盘、磁光盘、半导体存储器等等,根据需要安装在驱动器710上,以便于从其上读出的计算机程序根据需要被安装入存储部分708。

[0151] 特别地,根据本发明公开的实施例,上文参考流程图描述的过程可以被实现为计算机软件程序。例如,本发明公开的实施例包括一种计算机程序产品,其包括承载在计算机可读介质上的计算机程序,该计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中,该计算机程序可以通过通信部分709从网络上被下载和安装,和/或从可拆卸介质711被安装。在该计算机程序被中央处理单元(CPU)701执行时,执行本发明的系统中限定的上述功能。

[0152] 需要说明的是,本发明所示的计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机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。计算机可读存储介质的更具体的例子可以包括但不限于:具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑磁盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本发明中,计算机可读存储介质可以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。而在本发明中,计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号,

其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式,包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读介质,该计算机可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输,包括但不限于:无线、电线、光缆、RF等等,或者上述的任意合适的组合。

[0153] 附图中的流程图和框图,图示了按照本发明各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分,上述模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意,在有些作为替换的实现中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个接连地表示的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意,框图或流程图中的每个方框、以及框图或流程图中的方框的组合,可以用执行规定的功能或操作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0154] 描述于本发明实施例中所涉及到的模块可以通过软件的方式实现,也可以通过硬件的方式来实现。所描述的模块也可以设置在处理器中,例如,可以描述为:一种处理器包括:样本获取模块,获取包含多个样本的样本集;模型训练模块,基于所述样本集,采用Unet网络训练印章过滤模型;印章过滤模块,基于所述印章过滤模型对含印章的证件图片进行印章过滤处理。其中,这些模块的名称在某种情况下并不构成对该模块本身的限定,例如,样本获取模块还可以被描述为“基于所述印章过滤模型对含印章的证件图片进行印章过滤处理的模块”。

[0155] 作为另一方面,本发明还提供了一种计算机可读介质,该计算机可读介质可以是上述实施例中描述的设备中所包含的;也可以是单独存在,而未装配入该设备中。上述计算机可读介质承载有一个或者多个程序,当上述一个或者多个程序被一个该设备执行时,使得该设备包括:获取包含多个样本的样本集;基于所述样本集,采用Unet网络训练印章过滤模型;基于所述印章过滤模型对含印章的证件图片进行印章过滤处理。

[0156] 根据本发明实施例的技术方案,采用Unet网络训练印章过滤模型、基于训练得到的印章过滤模型对含印章的证件图片进行印章过滤处理,能够在不破坏证件图片中信息的前提下去除证件图片中的印章,便于后续识别证件图片中的栏位。

[0157] 上述具体实施方式,并不构成对本发明保护范围的限制。本领域技术人员应该明白的是,取决于设计要求和因素,可以发生各种各样的修改、组合、子组合和替代。任何在本发明的精神和原则之内所作的修改、等同替换和改进等,均应包含在本发明保护范围之内。

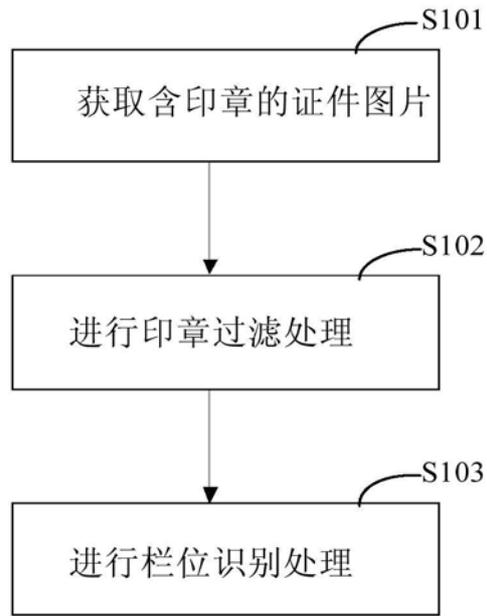


图1

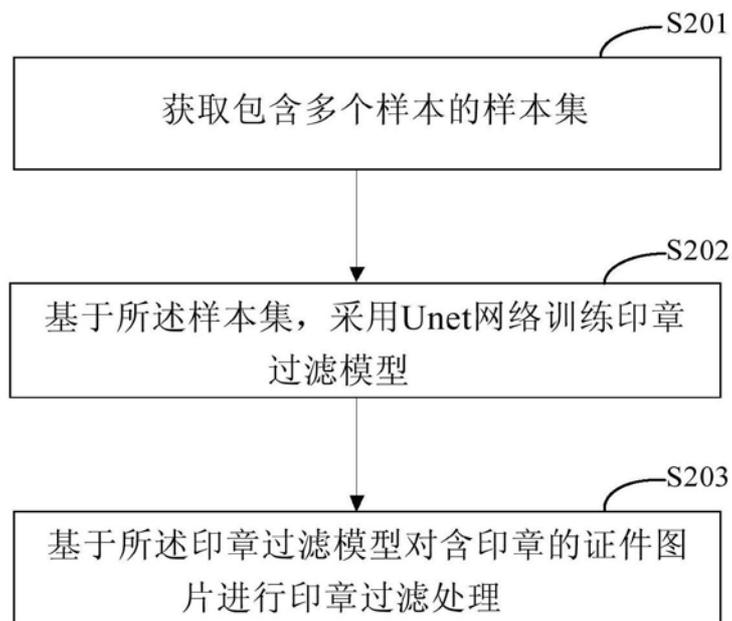


图2

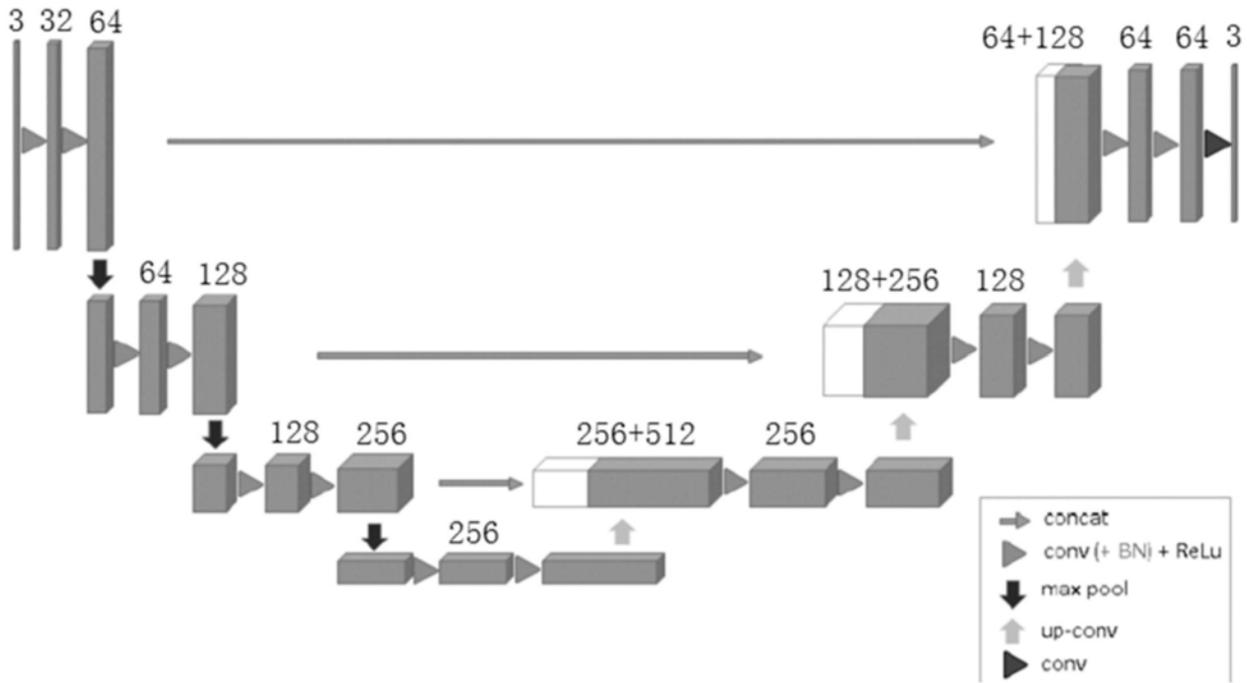


图3

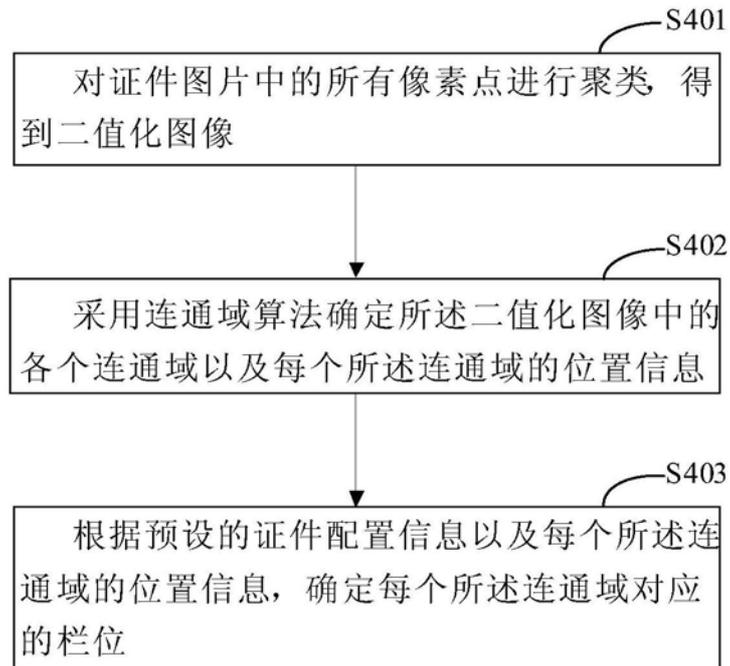


图4

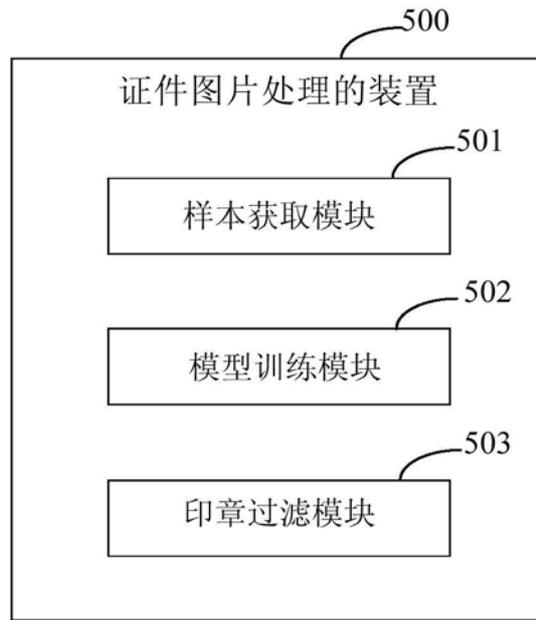


图5

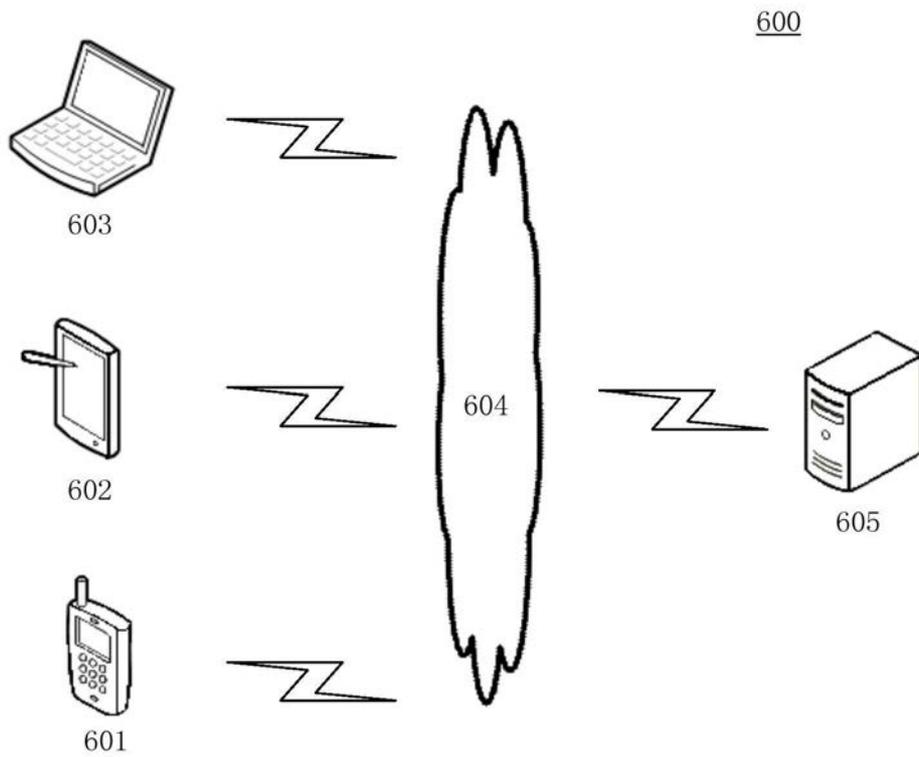


图6

700

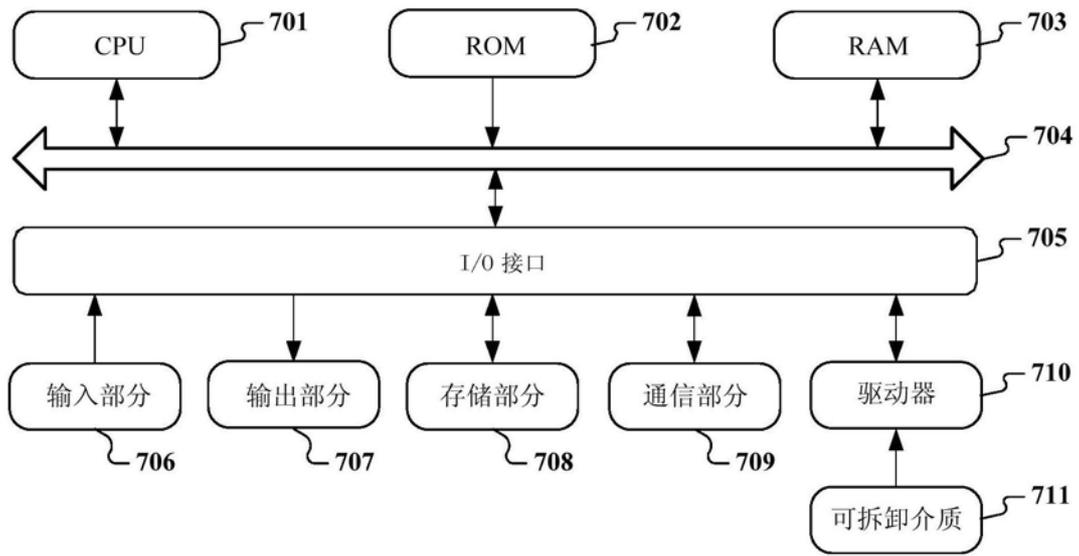


图7