

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 10 月 8 日 (08.10.2020)



(10) 国际公布号
WO 2020/199704 A1

- (51) 国际专利分类号:
G06K 9/34 (2006.01) **G06K 9/62** (2006.01)
- (21) 国际申请号: PCT/CN2020/070568
- (22) 国际申请日: 2020 年 1 月 7 日 (07.01.2020)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201910267233.0 2019年4月3日 (03.04.2019) CN
- (71) 申请人: 北京市商汤科技开发有限公司(BEIJING SENSETIME TECHNOLOGY DEVELOPMENT CO., LTD.) [CN/CN]; 中国北京市海淀区北四环西路58号11层1101-1117室, Beijing 100080 (CN)。
- (72) 发明人: 刘 学 博 (LIU, Xuebo); 中国北京市海淀区北四环西路58号11层1101-1117室, Beijing 100080 (CN)。
- (74) 代理人: 北京博思佳知识产权代理有限公司 (BEIJING BESTIPR INTELLECTUAL PROPERTY LAW CORPORATION); 中国北京市海淀区上地三街9号嘉华大厦B座409, Beijing 100085 (CN)。
- (81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB,

GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

- (84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告(条约第21条(3))。

(54) Title: TEXT RECOGNITION

(54) 发明名称: 文本识别

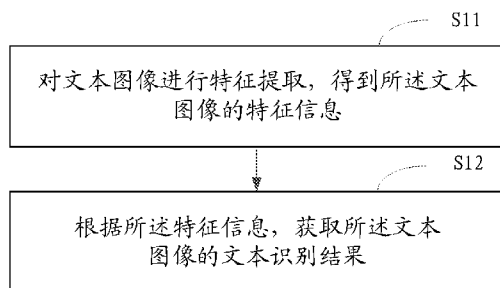


图 1

S11 Perform feature extraction on a text image to obtain feature information of the text image
S12 Obtain a text recognition result of the text image according to the feature information

(57) Abstract: The present application relates to a text recognition method and apparatus, an electronic device, and a storage medium. The method comprises: performing feature extraction on a text image to obtain feature information of the text image; and obtaining a text recognition result of the text image according to the feature information, wherein the text image comprises at least two characters, the feature information comprises a text correlation feature, and the text correlation feature is used for indicating the correlation between the characters in the text image.

(57) 摘要: 本公开涉及一种文本识别方法及装置、电子设备和存储介质, 所述方法包括: 对文本图像进行特征提取, 得到所述文本图像的特征信息; 根据所述特征信息, 获取所述文本图像的文本识别结果; 其中, 所述文本图像中包括至少两个字符, 所述特征信息包括文本关联特征, 所述文本关联特征用于表示所述文本图像中的字符之间的关联性。

WO 2020/199704 A1

文本识别

技术领域

[01]本公开涉及图像处理技术，尤其涉及文本识别。

背景技术

- 5 [02]在对图像中的文本进行识别过程中，往往存在待识别图像中文本分布不均匀的情况。例如，沿图像的水平方向分布有多个字符，沿竖直方向分布有单个字符，导致文本分布不均匀。通常的文本识别方法无法很好地处理这种类型的图像。

发明内容

[03]本公开提出了一种文本识别技术方案。

- 10 [04]根据本公开的一方面，提供了一种文本识别方法，包括：对文本图像进行特征提取，得到所述文本图像的特征信息；根据所述特征信息，获取所述文本图像的文本识别结果；其中，所述文本图像中包括至少两个字符，所述特征信息包括文本关联特征，所述文本关联特征用于表示所述文本图像中的字符之间的关联性。

- 15 [05]在一种可能的实现方式中，所述对文本图像进行特征提取，得到所述文本图像的特征信息，包括：通过至少一个第一卷积层对所述文本图像进行特征提取处理，得到所述文本图像的文本关联特征，其中，所述第一卷积层的卷积核尺寸为 $P \times Q$ ， P 、 Q 为整数，且 $Q > P \geq 1$ 。

- 20 [06]在一种可能的实现方式中，所述特征信息还包括文本结构特征；所述对文本图像进行特征提取，得到所述文本图像的特征信息，包括：通过至少一个第二卷积层对所述文本图像进行特征提取处理，得到所述文本图像的文本结构特征，其中，所述第二卷积层的卷积核尺寸为 $N \times N$ ， N 为大于 1 的整数。

[07]在一种可能的实现方式中，所述根据所述特征信息，获取所述文本图像的文本识别结果，包括：对所述文本关联特征和所述特征信息包括的文本结构特征进行融合处理，得到融合特征；根据所述融合特征，获取所述文本图像的文本识别结果。

- 25 [08]在一种可能的实现方式中，所述方法通过神经网络实现，所述神经网络中的编码网络包括多个网络块，每个网络块包括卷积核尺寸为 $P \times Q$ 的第一卷积层和卷积核尺寸为

$N \times N$ 的第二卷积层，其中，所述第一卷积层和所述第二卷积层的输入端分别与所述网络块的输入端连接。

[09]在一种可能的实现方式中，所述对所述文本关联特征和所述文本结构特征进行融合处理，得到融合特征，包括：对所述多个网络块中第一网络块的第一卷积层输出的文本关联特征和所述第一网络块的第二卷积层输出的文本结构特征进行融合，得到所述第一网络块的融合特征。

[10]所述根据所述融合特征，获取所述文本图像的文本识别结果，包括：对所述第一网络块的融合特征和所述第一网络块的输入信息进行残差处理，得到所述第一网络块的输出信息；基于所述第一网络块的输出信息，得到所述文本识别结果。

[11]在一种可能的实现方式中，所述神经网络中的编码网络包括下采样网络以及与所述下采样网络的输出端连接的多级特征提取网络，其中，每级特征提取网络包括至少一个所述网络块以及与所述至少一个网络块的输出端连接的下采样模块。

[12]在一种可能的实现方式中，所述神经网络为卷积神经网络。

[13]在一种可能的实现方式中，所述对文本图像进行特征提取，得到所述文本图像的特征信息，包括：对所述文本图像进行下采样处理，得到下采样结果；对所述下采样结果进行特征提取，得到所述文本图像的特征信息。

[14]根据本公开的另一方面，提供了一种文本识别装置，包括：特征提取模块，用于对文本图像进行特征提取，得到所述文本图像的特征信息；结果获取模块，用于根据所述特征信息，获取所述文本图像的文本识别结果；其中，所述文本图像中包括至少两个字符，所述特征信息包括文本关联特征，所述文本关联特征用于表示所述文本图像中的字符之间的关联性。

[15]根据本公开的另一方面，提供了一种电子设备，包括：处理器；用于存储处理器可执行指令的存储介质；其中，所述处理器被配置为调用所述存储介质存储的指令，以执行上述文本识别方法。

[16]根据本公开的另一方面，提供了一种机器可读存储介质，其上存储有机器可执行指令，所述机器可执行指令被处理器执行时实现上述文本识别方法。

[17]根据本公开实施例的文本识别方法，能够提取表示图像中字符之间的关联性的文本关联特征，根据包括文本关联特征的特征信息获取图像的文本识别结果，从而提高文本识别的准确性。

[18]应当理解的是，以上的一般描述和后文的细节描述仅是示例性和解释性的，而非限制本公开。根据下面参考附图对示例性实施例的详细说明，本公开的其它特征及方面将变得清楚。

附图说明

5 [19]此处的附图被并入说明书中并构成本说明书的一部分，这些附图示出了符合本公开的实施例，并与说明书一起用于说明本公开的技术方案。

[20]图 1 示出根据本公开实施例的文本识别方法的流程图。

[21]图 2 示出根据本公开实施例的网络块的示意图。

[22]图 3 示出根据本公开实施例的编码网络的示意图。

10 [23]图 4 示出根据本公开实施例的文本识别装置的框图。

[24]图 5 示出根据本公开实施例的一种电子设备的框图。

[25]图 6 示出根据本公开实施例的一种电子设备的框图。

具体实施方式

[26]以下将参考附图详细说明本公开的各种示例性实施例、特征和方面。附图中相同的
15 附图标记表示功能相同或相似的元件。除非特别指出，不必按比例绘制附图。

[27]在这里专用的词“示例性”意为“用作例子、实施例或说明性”。“示例性实施例”不必解释为优于或好于其它实施例。

[28]文本中术语“和/或”，仅仅用于描述关联对象的关联关系，表示可以存在多种关系。例如，A 和/或 B，可以表示：单独存在 A，同时存在 A 和 B，单独存在 B 这三种情况。

20 另外，文本中术语“至少一种”表示多种中的任意一种或多种中的至少两种的任意组合。例如，A、B、C 中的至少一种，可以表示从 A、B 和 C 构成的集合中选择的任意一个或多个元素。

[29]另外，为了更好地说明本公开，在下文的具体实施方式中给出了众多的具体细节。本领域技术人员应当理解，没有某些具体细节，本公开同样可以实施。在一些实例中，
25 对于本领域技术人员熟知的方法、手段、元件和电路未作详细描述，以便于凸显本公开的主旨。

[30]图 1 示出根据本公开实施例的文本识别方法的流程图。该文本识别方法可以由终端设备或其它设备执行，其中，终端设备可以为用户设备（User Equipment, UE）、移动设备、用户终端、终端、蜂窝电话、无绳电话、个人数字处理（Personal Digital Assistant, PDA）、手持设备、计算设备、车载设备、可穿戴设备等。

5 [31]如图 1 所示，所述方法包括：

步骤 S11，对文本图像进行特征提取，得到所述文本图像的特征信息；

步骤 S12，根据所述特征信息，获取所述文本图像的文本识别结果；

其中，所述文本图像中包括至少两个字符，所述特征信息包括文本关联特征，所述文本关联特征用于表示所述文本图像中的字符之间的关联性。

10 [32]根据本公开实施例的文本识别方法，能够提取包括文本关联特征的特征信息，其中，该文本关联特征表示图像中文本字符之间的关联性，并根据该特征信息获取图像的文本识别结果，从而提高文本识别的准确性。

[33]举例来说，文本图像可以是由图像采集设备（例如摄像头）采集的、包括字符的图像，例如在线身份验证的场景下拍摄的、包括字符的证件图像。文本图像也可以是从互
15 联网下载、用户上传或以其他方式获取的、包括字符的图像。本公开对文本图像的来源及类型不作限制。

[34]另外，在本文中提到的“字符”可以包括任意文本字符，例如文字、字母、数字、符号等，在本公开中不对“字符”的类型进行限制。

[35]在一些实施例中，在步骤 S11 中对文本图像进行特征提取，得到文本图像的特征信息，该特征信息可包括文本关联特征，用于表示文本图像中的文本字符之间的关联性，
20 例如，各个字符的分布次序、某几个字符同时出现的概率等。

[36]在一些实施例中，步骤 S11 包括：通过至少一个第一卷积层对所述文本图像进行特征提取处理，得到所述文本图像的文本关联特征，其中，所述第一卷积层的卷积核尺寸为 $P \times Q$ ， P 、 Q 为整数，且 $Q > P \geq 1$ 。

25 [37]举例来说，文本图像中可包括至少两个字符，在不同方向上字符可能分布不均匀，例如沿水平方向分布有多个字符，沿竖直方向分布有单个字符。在该情况下，进行特征提取的卷积层可采用在不同方向上尺寸不对称的卷积核，以更好地提取字符较多的方向上的文本关联特征。

[38] 在一些实施例中，通过卷积核尺寸为 $P \times Q$ 的至少一个第一卷积层对文本图像进行特征提取处理，以便适应字符分布不均匀的图像。在文本图像中水平方向的字符数量大于竖直方向的字符数量时，可以设定 $Q > P \geq 1$ ，以便更好地提取水平方向（横向）的语义信息（文本关联特征）。在一些实施例中， Q 与 P 之间的差别大于某一阈值。例如，

5 文本图像中的字符为横向排布（例如，单排）的多个文字时，第一卷积层可以采用 1×5 、 1×7 、 1×9 等尺寸的卷积核。

[39] 在一些实施例中，在文本图像中水平方向的字符数量小于竖直方向的字符数量时，可以设定 $P > Q \geq 1$ ，以便更好地提取竖直方向（纵向）的语义信息（文本关联特征）。例如，文本图像中的字符为纵向排布（例如，单列）的多个文字时，第一卷积层可以采

10 用 5×1 、 7×1 、 9×1 等尺寸的卷积核。本公开对第一卷积层的层数以及卷积核的具体尺寸不作限制。

[40] 通过这种方式，能够更好地提取文本图像中的字符较多的方向上的文本关联特征，从而提高文本识别的准确性。

[41] 在一些实施例中，所述特征信息还包括文本结构特征；步骤 S11 包括：通过至少一个第二卷积层对所述文本图像进行特征提取处理，得到所述文本图像的文本结构特征，其中，所述第二卷积层的卷积核尺寸为 $N \times N$ ， N 为大于 1 的整数。

15

[42] 举例来说，文本图像的特征信息还包括文本结构特征，用于表示文本的空间结构信息，例如字符的结构、形状、笔画粗细、字体类型或字体角度等信息。在该情况下，进行特征提取的卷积层可采用在不同方向上尺寸对称的卷积核，以更好地提取文本图像中的各个字符的空间结构信息，得到文本图像的文本结构特征。

20

[43] 在一些实施例中，通过卷积核尺寸为 $N \times N$ 的至少一个第二卷积层对文本图像进行特征提取处理，得到文本图像的文本结构特征， N 为大于 1 的整数。其中， N 例如可取值为 2、3、5 等，也即第二卷积层可采用 2×2 、 3×3 、 5×5 等尺寸的卷积核。本公开对第二卷积层的层数以及卷积核的具体尺寸不作限制。通过这种方式，能够提取文本图

25 像中的字符的文本结构特征，从而提高文本识别的准确性。

[44] 在一些实施例中，所述对文本图像进行特征提取，得到所述文本图像的特征信息，包括：

对所述文本图像进行下采样处理，得到下采样结果；

对所述下采样结果进行特征提取，得到所述文本图像的特征信息。

[45]举例来说，在对文本图像特征提取之前，首先通过下采样网络对文本图像进行下采样处理。该下采样网络包括至少一个卷积层，该卷积层的卷积核尺寸例如为 3×3 。将下采样结果分别输入至少一个第一卷积层和至少一个第二卷积层进行特征提取，得到文本图像的文本关联特征和文本结构特征。通过下采样处理，可进一步降低特征提取的计算量，提高网络的运行速度；同时避免数据分布不均衡对特征提取产生的影响。

[46]在一些实施例中，根据在步骤 S11 中得到的特征信息，可在步骤 S12 中获取所述文本图像的文本识别结果。

[47]在一些实施例中，文本识别结果是对特征信息进行分类处理之后的结果。文本识别结果例如为针对文本图像中各个字符的具有最大预测概率的预测结果字符。例如，将文本图像中位置 1、2、3、4 处的字符预测为“很多文字”。文本识别结果还例如为文本图像中各个字符的预测概率。例如，当文本图像中位置 1、2、3、4 处为“很多文字”四个汉字时，其对应的文本识别结果包括：将位置 1 的字符预测为“根”的概率为 85%，预测为“很”的概率为 98%；将位置 2 的字符预测为“夕”的概率为 60%，预测为“多”的概率为 90%；将位置 3 的字符预测为“纹”的概率为 65%，预测为“文”的概率为 94%；将位置 4 的字符预测为“写”的概率为 70%，预测为“字”的预测概率为 90%。本公开对文本识别结果的表示形式不作限制。

[48]在一些实施例中，可仅根据文本关联特征来获取文本识别结果，也可根据文本关联特征和文本结构特征来获取文本识别结果。本公开对此不作限制。

[49]在一些实施例中，步骤 S12 包括：

对所述文本关联特征和所述特征信息包括的文本结构特征进行融合处理，得到融合特征；

根据所述融合特征，获取所述文本图像的文本识别结果。

[50]在本公开实施例中，可以通过具有不同卷积核尺寸的不同卷积层分别对文本图像进行卷积处理以获得文本图像的文本关联特征和文本结构特征。然后，对得到的文本关联特征和文本结构特征进行融合，得到融合特征。该“融合”处理例如可以为将该不同卷积层输出的结果逐像素进行相加的操作。进而，根据融合特征获取文本图像的文本识别结果。获取的融合特征能够更全面地指示文本信息，从而提高文本识别的准确性。

[51]在一些实施例中，所述文本识别方法通过神经网络实现，所述神经网络中的编码网络包括多个网络块，每个网络块包括卷积核尺寸为 $P \times Q$ 的第一卷积层和卷积核尺寸为

$N \times N$ 第二卷积层，其中，所述第一卷积层和所述第二卷积层的输入端分别与所述网络块的输入端连接。

[52] 在一些实施例中，所述神经网络例如为卷积神经网络，本公开对神经网络的具体类型不作限制。

5 [53] 举例来说，该神经网络可包括编码网络，编码网络包括多个网络块，每个网络块包括卷积核尺寸为 $P \times Q$ 的第一卷积层和卷积核尺寸为 $N \times N$ 第二卷积层，分别用于提取文本图像的文本关联特征和文本结构特征。其中，所述第一卷积层和所述第二卷积层的输入端分别与所述网络块的输入端连接，以使网络块的输入信息能够分别被输入第一卷积层和第二卷积层进行特征提取。

10 [54] 在一些实施例中，在第一卷积层和第二卷积层之前，可以分别设置有卷积核尺寸例如为 1×1 的第三卷积层，对网络块的输入信息进行降维处理；将降维后的输入信息分别输入第一卷积层和第二卷积层进行特征提取，从而有效减少特征提取的计算量。

[55] 在一些实施例中，所述对所述文本关联特征和所述文本结构特征进行融合处理，得到融合特征的步骤，包括：对所述网络块的第一卷积层输出的文本关联特征和所述网络
15 块的第二卷积层输出的文本结构特征进行融合，得到所述网络块的融合特征。

[56] 所述根据所述融合特征，获取所述文本图像的文本识别结果的步骤，包括：对所述网络块的融合特征和所述网络块的输入信息进行残差处理，得到所述网络块的输出信息；基于所述网络块的输出信息，得到所述文本识别结果。

[57] 举例来说，对于任意一个网络块，可将网络块的第一卷积层输出的文本关联特征和
20 网络块的第二卷积层输出的文本结构特征进行融合，得到所述网络块的融合特征，获取的融合特征能够更全面地指示文本信息。

[58] 在一些实施例中，对网络块的融合特征和网络块的输入信息进行残差处理，得到网络块的输出信息；进而根据网络块的输出信息得到文本识别结果。这里的“残差处理”利用了与 ResNet (Residual Neural Network) 中的残差学习类似的技术。通过使用残差连
25 接，每个网络块只需要学习输出的融合特征和输入信息之间的差值（网络块的输出信息），而不需要学习全部特征，使学习更容易收敛，从而减小网络块的计算量，并使得网络块更易于训练。

[59] 图 2 示出根据本公开实施例的网络块的示意图。如图 2 所示，该网络块包括卷积核尺寸为 1×1 的第三卷积层 21、卷积核尺寸为 1×7 的第一卷积层 22 以及卷积核尺寸为

3×3 的第二卷积层 23。网络块的输入信息 24 分别输入两个第三卷积层 21 中进行降维处理，从而减少特征提取的计算量。将降维后的输入信息分别输入第一卷积层 22 和第二卷积层 23 进行特征提取，得到网络块的文本关联特征和文本结构特征。

[60] 在一些实施例中，对网络块的第一卷积层输出的文本关联特征和网络块的第二卷积层输出的文本结构特征进行融合，得到所述网络块的融合特征，从而更全面地指示文本信息。对网络块的融合特征与网络块的输入信息进行残差处理，得到网络块的输出信息 25。根据网络块的输出信息，可获取文本图像的文本识别结果。

[61] 在一些实施例中，所述神经网络中的编码网络包括下采样网络以及与所述下采样网络的输出端连接的多级特征提取网络，其中，每级特征提取网络包括至少一个所述网络块以及与所述至少一个网络块的输出端连接的下采样模块。

[62] 举例来说，可通过多级特征提取网络对文本图像进行特征提取。在该情况下，神经网络中的编码网络包括下采样网络以及与所述下采样网络的输出端连接的多级特征提取网络。将文本图像输入下采样网络（包括至少一个卷积层）进行下采样处理，输出下采样结果；将下采样结果输入多级特征提取网络进行特征提取，可得到文本图像的特征信息。

[63] 在一些实施例中，将文本图像的下采样结果输入到第一级特征提取网络中进行特征提取，输出第一级特征提取网络的输出信息；再将第一级特征提取网络的输出信息输入第二级特征提取网络中，输出第二级特征提取网络的输出信息；以此类推，可将最后一级特征提取网络的输出信息作为编码网络最终的输出信息。

[64] 其中，每级特征提取网络包括至少一个所述网络块以及与所述至少一个网络块的输出端连接的下采样模块。该下采样模块包括至少一个卷积层，可在每个网络块的输出端连接下采样模块，也可在每级特征提取网络的最后一个网络块的输出端连接下采样模块。这样，每级特征提取网络的输出信息都会经过下采样再被输入到下一级特征提取网络，从而降低特征尺寸，减小计算量。

[65] 图 3 示出根据本公开实施例的编码网络的示意图。如图 3 所示，编码网络包括下采样网络 31 以及与下采样网络的输出端连接的五级特征提取网络 32、33、34、35、36，其中第一级特征提取网络 32 至第五级特征提取网络 36 分别包括 1、3、3、3、2 个网络块，每级特征提取网络的最后一个网络块的输出端连接下采样模块。

[66] 在一些实施例中，文本图像输入下采样网络 31 进行下采样处理，输出下采样结果；

下采样结果输入到第一级特征提取网络 32（网络块+下采样模块）中进行特征提取，输出第一级特征提取网络 32 的输出信息；第一级特征提取网络 32 的输出信息输入到第二级特征提取网络 33 中，依次经由三个网络块以及下采样模块处理，输出第二级特征提取网络 33 的输出信息；以此类推，将第五级特征提取网络 36 的输出信息作为编码网络最终的输出信息。

[67]通过下采样网络及多级特征提取网络进行特征提取，可形成瓶颈（bottleneck）结构，能够提高文字识别的效果，显著减小计算量，在网络训练过程中更容易收敛，降低了训练难度。

[68]在一些可能的实现方式中，所述方法还包括：对所述文本图像进行预处理，得到预处理后的文本图像。

[69]在本公开的实现方式中，所述文本图像可以是包括多行或多列的文本图像，预处理操作可以是将包括了多行或多列的文本图像分割为单行或单列的文本图像，进而开始识别。

[70]在一些可能的实现方式中，所述预处理操作可以是归一化处理、几何变换处理和图像增强处理等操作。

[71]在一些实施例，可根据预设的训练集对神经网络中的编码网络进行训练。在训练过程中，使用联结时序分类损失对编码网络进行监督学习，对图片每个部分的预测结果进行分类，分类结果与真实结果越接近损失越小。在满足训练条件时，可得到训练后的编码网络。本公开对编码网络的损失函数的选取及具体训练方式不作限制。

[72]根据本公开实施例的文本识别方法，能够通过卷积核尺寸不对称的卷积层提取表示图像中字符之间的关联性的文本关联特征，提高了特征提取的效果并减小了不必要的计算量；能够分别提取文本关联特征以及字符的文本结构特征，实现了深度神经网络的并行化，显著减少运算时间。

[73]根据本公开实施例的文本识别方法，采用了利用残差连接以及瓶颈结构的多级特征提取网络的网络结构，不需要递归神经网络就可以很好地捕捉图像中的文本信息，能够得到很好的识别结果，大大减少了计算量；并且该网络结构易于训练，能够快速完成训练过程。

[74]根据本公开实施例的文本识别方法可应用于身份认证，内容审核，图片检索，图片翻译等使用场景中，实现文本识别。例如，在身份验证的使用场景中，通过该方法提取

身份证、银行卡、驾驶证等各种类型的证件图像中的文字内容，以便完成身份验证；在内容审核的使用场景中，通过该方法提取对社交网络中用户上传的图像中的文字内容，识别图像中是否包含非法信息，例如暴力相关的文本等。

[75]可以理解，本公开提及的上述各个方法实施例，在不违背原理逻辑的情况下，均可以彼此相互结合形成结合后的实施例，限于篇幅，本公开不再赘述。本领域技术人员可以理解，在具体实施方式的上述方法中，各步骤的具体执行顺序应当以其功能和可能的内在逻辑确定。

[76]此外，本公开还提供了文本识别装置、电子设备、计算机可读存储介质、程序，上述均可用来实现本公开提供的任一种文本识别方法，相应技术方案和描述和参见方法部分的相应记载，不再赘述。

[77]图 4 示出根据本公开实施例的文本识别装置的框图，如图 4 所示，所述文本识别装置包括：

特征提取模块 41，用于对文本图像进行特征提取，得到所述文本图像的特征信息；结果获取模块 42，用于根据所述特征信息，获取所述文本图像的文本识别结果；其中，所述文本图像中包括至少两个字符，所述特征信息包括文本关联特征，所述文本关联特征用于表示所述文本图像中的字符之间的关联性。

[78]在一些实施例中，所述特征提取模块包括：第一提取子模块，用于通过至少一个第一卷积层对所述文本图像进行特征提取处理，得到所述文本图像的文本关联特征，其中，所述第一卷积层的卷积核尺寸为 $P \times Q$ ， P 、 Q 为整数，且 $Q > P \geq 1$ 。

[79]在一些实施例中，所述特征信息还包括文本结构特征；所述特征提取模块包括：第二提取子模块，用于通过至少一个第二卷积层对所述文本图像进行特征提取处理，得到所述文本图像的文本结构特征，其中，所述第二卷积层的卷积核尺寸为 $N \times N$ ， N 为大于 1 的整数。

[80]在一些实施例中，所述结果获取模块包括：融合子模块，用于对所述文本关联特征和所述特征信息包括的文本结构特征进行融合处理，得到融合特征；结果获取子模块，用于根据所述融合特征，获取所述文本图像的文本识别结果。

[81]在一些实施例中，所述装置适用于神经网络，所述神经网络中的编码网络包括多个网络块，每个网络块包括卷积核尺寸为 $P \times Q$ 的第一卷积层和卷积核尺寸为 $N \times N$ 的第二卷积层，其中，所述第一卷积层和所述第二卷积层的输入端分别与所述网络块的输入

端连接。

[82] 在一些实施例中，所述装置适用于神经网络，所述神经网络中的编码网络包括多个网络块，所述融合子模块用于：对所述多个网络块中第一网络块的第一卷积层输出的文本关联特征和所述第一网络块的第二卷积层输出的文本结构特征进行融合，得到所述第一网络块的融合特征。

[83] 所述结果获取子模块用于：对所述第一网络块的融合特征和所述第一网络块的输入信息进行残差处理，得到所述第一网络块的输出信息；基于所述第一网络块的输出信息，得到所述文本识别结果。

[84] 在一些实施例中，所述神经网络中的编码网络包括下采样网络以及与所述下采样网络的输出端连接的多级特征提取网络，其中，每级特征提取网络包括至少一个所述网络块以及与所述至少一个网络块的输出端连接的下采样模块。

[85] 在一些实施例中，所述神经网络为卷积神经网络。

[86] 在一些实施例中，所述特征提取模块包括：下采样子模块，用于对所述文本图像进行下采样处理，得到下采样结果；第三提取子模块，用于对所述下采样结果进行特征提取，得到所述文本图像的特征信息。

[87] 在一些实施例中，本公开实施例提供的装置具有的功能或包含的模块可以用于执行上文方法实施例描述的方法，其具体实现可以参照上文方法实施例的描述，为了简洁，这里不再赘述

[88] 本公开实施例还提出一种机器可读存储介质，其上存储有机器可执行指令，所述机器可执行指令被处理器执行时实现上述方法。机器可读存储介质可以是非易失性机器可读存储介质。

[89] 本公开实施例还提出一种电子设备，包括：处理器；用于存储处理器可执行指令的存储介质；其中，所述处理器被配置为调用所述存储介质存储的指令，以执行上述方法。

[90] 电子设备可以被提供为终端、服务器或其它形态的设备。

[91] 图 5 示出根据本公开实施例的一种电子设备 800 的框图。例如，电子设备 800 可以是移动电话，计算机，数字广播终端，消息收发设备，游戏控制台，平板设备，医疗设备，健身设备，个人数字助理等终端。

[92] 参照图 5，电子设备 800 可以包括以下一个或多个组件：处理组件 802，存储介质

804, 电源组件 806, 多媒体组件 808, 音频组件 810, 输入/输出 (I/O) 接口 812, 传感器组件 814, 以及通信组件 816。

[93] 处理组件 802 通常控制电子设备 800 的整体操作, 诸如与显示, 电话呼叫, 数据通信, 相机操作和记录操作相关联的操作。处理组件 802 可以包括一个或多个处理器 820 来执行指令, 以完成上述的方法的全部或部分步骤。此外, 处理组件 802 可以包括一个或多个模块, 便于处理组件 802 和其他组件之间的交互。例如, 处理组件 802 可以包括多媒体模块, 以方便多媒体组件 808 和处理组件 802 之间的交互。

[94] 存储介质 804 被配置为存储各种类型的数据以支持在电子设备 800 的操作。这些数据的示例包括用于在电子设备 800 上操作的任何应用程序或方法的指令, 联系人数据, 电话簿数据, 消息, 图片, 视频等。存储介质 804 可以由任何类型的易失性或非易失性存储设备或者它们的组合实现, 如静态随机存取存储器 (SRAM), 电可擦除可编程只读存储器 (EEPROM), 可擦除可编程只读存储器 (EPROM), 可编程只读存储器 (PROM), 只读存储器 (ROM), 磁存储器, 快闪存储器, 磁盘或光盘。

[95] 电源组件 806 为电子设备 800 的各种组件提供电力。电源组件 806 可以包括电源管理系统, 一个或多个电源, 及其他与为电子设备 800 生成、管理和分配电力相关联的组件。

[96] 多媒体组件 808 包括提供所述电子设备 800 和用户之间的输出接口的屏幕。在一些实施例中, 屏幕可以包括液晶显示器 (LCD) 和触摸面板 (TP)。如果屏幕包括触摸面板, 屏幕可以被实现为触摸屏, 以接收来自用户的输入信号。触摸面板包括一个或多个触摸传感器以感测触摸、滑动和触摸面板上的手势。所述触摸传感器可以不仅感测触摸或滑动动作的边界, 而且还检测与所述触摸或滑动操作相关的持续时间和压力。在一些实施例中, 多媒体组件 808 包括一个前置摄像头和/或后置摄像头。当电子设备 800 处于操作模式, 如拍摄模式或视频模式时, 前置摄像头和/或后置摄像头可以接收外部的多媒体数据。每个前置摄像头和后置摄像头可以是一个固定的光学透镜系统或具有焦距和光学变焦能力。

[97] 音频组件 810 被配置为输出和/或输入音频信号。例如, 音频组件 810 包括一个麦克风 (MIC), 当电子设备 800 处于操作模式, 如呼叫模式、记录模式和语音识别模式时, 麦克风被配置为接收外部音频信号。所接收的音频信号可以被进一步存储在存储器 804 或经由通信组件 816 发送。在一些实施例中, 音频组件 810 还包括一个扬声器, 用于输出音频信号。

[98] I/O 接口 812 为处理组件 802 和外围接口模块之间提供接口，上述外围接口模块可以是键盘，点击轮，按钮等。这些按钮可包括但不限于：主页按钮、音量按钮、启动按钮和锁定按钮。

5 [99] 传感器组件 814 包括一个或多个传感器，用于为电子设备 800 提供各个方面的状态评估。例如，传感器组件 814 可以检测到电子设备 800 的打开/关闭状态，组件的相对定位，例如所述组件为电子设备 800 的显示器和小键盘，传感器组件 814 还可以检测电子设备 800 或电子设备 800 一个组件的位置改变，用户与电子设备 800 接触的存在或不存在，电子设备 800 方位或加速/减速和电子设备 800 的温度变化。传感器组件 814 可以包括接近传感器，被配置用来在没有任何的物理接触时检测附近物体的存在。传感器组件
10 814 还可以包括光传感器，如 CMOS 或 CCD 图像传感器，用于在成像应用中使用。在一些实施例中，该传感器组件 814 还可以包括加速度传感器，陀螺仪传感器，磁传感器，压力传感器或温度传感器。

[100] 通信组件 816 被配置为便于电子设备 800 和其他设备之间有线或无线方式的通信。电子设备 800 可以接入基于通信标准的无线网络，如 WiFi，2G 或 3G，或它们的组合。
15 在一个示例性实施例中，通信组件 816 经由广播信道接收来自外部广播管理系统的广播信号或广播相关信息。在一个示例性实施例中，所述通信组件 816 还包括近场通信（NFC）模块，以促进短程通信。例如，在 NFC 模块可基于射频识别（RFID）技术，红外数据协会（IrDA）技术，超宽带（UWB）技术，蓝牙（BT）技术和其他技术来实现。

20 [101] 在示例性实施例中，电子设备 800 可以被一个或多个应用专用集成电路（ASIC）、数字信号处理器（DSP）、数字信号处理设备（DSPD）、可编程逻辑器件（PLD）、现场可编程门阵列（FPGA）、控制器、微控制器、微处理器或其他电子元件实现，用于执行上述方法。

[102] 在示例性实施例中，还提供了一种非易失性机器可读存储介质，例如包括机器
25 可执行指令的存储介质 804，上述机器可执行指令可由电子设备 800 的处理器 820 执行以完成上述方法。

[103] 图 6 示出根据本公开实施例的一种电子设备 1900 的框图。例如，电子设备 1900 可以被提供为一服务器。参照图 6，电子设备 1900 包括处理组件 1922，其进一步包括一个或多个处理器，以及由存储器 1932 所代表的存储器资源，用于存储可由处理组件
30 1922 的执行的指令，例如应用程序。存储器 1932 中存储的应用程序可以包括一个或一

个以上的每一个对应于一组指令的模块。此外，处理组件 1922 被配置为执行指令，以执行上述方法。

[104] 电子设备 1900 还可以包括一个电源组件 1926 被配置为执行电子设备 1900 的电源管理，一个有线或无线网络接口 1950 被配置为将电子设备 1900 连接到网络，和一个输入输出（I/O）接口 1958。电子设备 1900 可以操作基于存储在存储器 1932 的操作系统，例如 Windows ServerTM，Mac OS XTM，UnixTM，LinuxTM，FreeBSDTM 或类似。

[105] 在示例性实施例中，还提供了一种非易失性机器可读存储介质，例如包括计算机程序指令的存储器 1932，上述计算机程序指令可由电子设备 1900 的处理组件 1922 执行以完成上述方法。

[106] 本公开可以是系统、方法和/或计算机程序产品。计算机程序产品可以包括计算机可读存储介质，其上载有用于使处理器实现本公开的各个方面的计算机可读程序指令。

[107] 计算机可读存储介质可以是可以保持和存储由指令执行设备使用的指令的有形设备。计算机可读存储介质例如可以是——但不限于——电存储设备、磁存储设备、光存储设备、电磁存储设备、半导体存储设备或者上述的任意合适的组合。计算机可读存储介质的更具体的例子（非穷举的列表）包括：便携式计算机盘、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）、静态随机存取存储器（SRAM）、便携式压缩盘只读存储器（CD-ROM）、数字多功能盘（DVD）、记忆棒、软盘、机械编码设备、例如其上存储有指令的打孔卡或凹槽内凸起结构、以及上述的任意合适的组合。这里所使用的计算机可读存储介质不被解释为瞬时信号本身，诸如无线电波或者其他自由传播的电磁波、通过波导或其他传输媒介传播的电磁波（例如，通过光纤电缆的光脉冲）、或者通过电线传输的电信号。

[108] 这里所描述的计算机可读程序指令可以从计算机可读存储介质下载到各个计算/处理设备，或者通过网络、例如因特网、局域网、广域网和/或无线网下载到外部计算机或外部存储设备。网络可以包括铜传输电缆、光纤传输、无线传输、路由器、防火墙、交换机、网关计算机和/或边缘服务器。每个计算/处理设备中的网络适配卡或者网络接口从网络接收计算机可读程序指令，并转发该计算机可读程序指令，以供存储在各个计算/处理设备中的计算机可读存储介质中。

[109] 用于执行本公开操作的计算机程序指令可以是汇编指令、指令集架构（ISA）指

令、机器指令、机器相关指令、微代码、固件指令、状态设置数据、或者以一种或多种编程语言的任意组合编写的源代码或目标代码，所述编程语言包括面向对象的编程语言—诸如 Smalltalk、C++等，以及常规的过程式编程语言—诸如“C”语言或类似的编程语言。计算机可读程序指令可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中，远程计算机可以通过任意种类的网络—包括局域网(LAN)或广域网(WAN)—连接到用户计算机，或者，可以连接到外部计算机(例如利用因特网服务提供商来通过因特网连接)。在一些实施例中，通过利用计算机可读程序指令的状态信息来个性化定制电子电路，例如可编程逻辑电路、现场可编程门阵列(FPGA)或可编程逻辑阵列(PLA)，该电子电路可以执行计算机可读程序指令，从而实现本公开的各个方面。

[110] 这里参照根据本公开实施例的方法、装置(系统)和计算机程序产品的流程图和/或框图描述了本公开的各个方面。应当理解，流程图和/或框图的每个方框以及流程图和/或框图中各方框的组合，都可以由计算机可读程序指令实现。

[111] 这些计算机可读程序指令可以提供给通用计算机、专用计算机或其它可编程数据处理装置的处理器，从而生产出一种机器，使得这些指令在通过计算机或其它可编程数据处理装置的处理器执行时，产生了实现流程图和/或框图中的一个或多个方框中规定的功能/动作的装置。也可以把这些计算机可读程序指令存储在计算机可读存储介质中，这些指令使得计算机、可编程数据处理装置和/或其他设备以特定方式工作，从而，存储有指令的计算机可读介质则包括一个制造品，其包括实现流程图和/或框图中的一个或多个方框中规定的功能/动作的各个方面的指令。

[112] 也可以把计算机可读程序指令加载到计算机、其它可编程数据处理装置、或其它设备上，使得在计算机、其它可编程数据处理装置或其它设备上执行一系列操作步骤，以产生计算机实现的过程，从而使得在计算机、其它可编程数据处理装置、或其它设备上执行的指令实现流程图和/或框图中的一个或多个方框中规定的功能/动作。

[113] 附图中的流程图和框图显示了根据本公开的多个实施例的系统、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、程序段或指令的一部分，所述模块、程序段或指令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。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，两个连续的方框实际上可

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。也要注意的，框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合，可以用执行规定的功能或动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。

- 5 [114] 以上已经描述了本公开的各实施例，上述说明是示例性的，并非穷尽性的，并且也不限于所披露的各实施例。在不偏离所说明的各实施例的范围和精神的情况下，对于本技术领域的普通技术人员来说许多修改和变更都是显而易见的。文本中所用术语的选择，旨在最好地解释各实施例的原理、实际应用或对市场中的技术的改进，或者使本技术领域的其它普通技术人员能理解文本披露的各实施例。

权利要求书

1、一种文本识别方法，其特征在于，包括：

对文本图像进行特征提取，得到所述文本图像的特征信息；

根据所述特征信息，获取所述文本图像的文本识别结果；

5 其中，所述文本图像中包括至少两个字符，所述特征信息包括文本关联特征，所述文本关联特征用于表示所述文本图像中的字符之间的关联性。

2、根据权利要求 1 所述的方法，其特征在于，所述对文本图像进行特征提取，得到所述文本图像的特征信息，包括：

10 通过至少一个第一卷积层对所述文本图像进行特征提取处理，得到所述文本图像的文本关联特征，其中，所述第一卷积层的卷积核尺寸为 $P \times Q$ ， P 、 Q 为整数，且 $Q > P \geq 1$ 。

3、根据权利要求 1 或 2 所述的方法，其特征在于，所述特征信息还包括文本结构特征；

所述对文本图像进行特征提取，得到所述文本图像的特征信息，包括：

15 通过至少一个第二卷积层对所述文本图像进行特征提取处理，得到所述文本图像的文本结构特征，其中，所述第二卷积层的卷积核尺寸为 $N \times N$ ， N 为大于 1 的整数。

4、根据权利要求 1-3 中任意一项所述的方法，其特征在于，所述根据所述特征信息，获取所述文本图像的文本识别结果，包括：

20 对所述文本关联特征和所述特征信息包括的文本结构特征进行融合处理，得到融合特征；

根据所述融合特征，获取所述文本图像的文本识别结果。

25 5、根据权利要求 1-4 中任意一项所述的方法，其特征在于，所述方法通过神经网络实现，所述神经网络中的编码网络包括多个网络块，每个网络块包括卷积核尺寸为 $P \times Q$ 的第一卷积层和卷积核尺寸为 $N \times N$ 的第二卷积层，其中，所述第一卷积层和所述第二卷积层的输入端分别与所述网络块的输入端连接。

6、根据权利要求 4 所述的方法，其特征在于，所述方法通过神经网络实现，所述神经网络中的编码网络包括多个网络块，

所述对所述文本关联特征和所述文本结构特征进行融合处理，得到融合特征，包括：

30 对所述多个网络块中第一网络块的第一卷积层输出的文本关联特征和所述第一网络块的第二卷积层输出的文本结构特征进行融合，得到所述第一网络块的融合特征；

所述根据所述融合特征，获取所述文本图像的文本识别结果，包括：

对所述第一网络块的融合特征和所述第一网络块的输入信息进行残差处理，得到所述第一网络块的输出信息；

基于所述第一网络块的输出信息，得到所述文本识别结果。

7、根据权利要求 5 或 6 所述的方法，其特征在于，所述神经网络中的编码网络包括下采样网络以及与所述下采样网络的输出端连接的多级特征提取网络，其中，每级特征提取网络包括至少一个所述网络块以及与所述至少一个网络块的输出端连接的下采样模块。

8、根据权利要求 5-7 中任意一项所述的方法，其特征在于，所述神经网络为卷积神经网络。

9、根据权利要求 1 至 8 中任一项所述的方法，其特征在于，所述对文本图像进行特征提取，得到所述文本图像的特征信息，包括：

对所述文本图像进行下采样处理，得到下采样结果；

对所述下采样结果进行特征提取，得到所述文本图像的特征信息。

10、一种文本识别装置，其特征在于，包括：

特征提取模块，用于对文本图像进行特征提取，得到所述文本图像的特征信息；

结果获取模块，用于根据所述特征信息，获取所述文本图像的文本识别结果；

其中，所述文本图像中包括至少两个字符，所述特征信息包括文本关联特征，所述文本关联特征用于表示所述文本图像中的字符之间的关联性。

11、根据权利要求 10 所述的装置，其特征在于，所述特征提取模块包括：

第一提取子模块，用于通过至少一个第一卷积层对所述文本图像进行特征提取处理，得到所述文本图像的文本关联特征，其中，所述第一卷积层的卷积核尺寸为 $P \times Q$ ， P 、 Q 为整数，且 $Q > P \geq 1$ 。

12、根据权利要求 10 或 11 所述的装置，其特征在于，所述特征信息还包括文本结构特征；

所述特征提取模块包括：

第二提取子模块，用于通过至少一个第二卷积层对所述文本图像进行特征提取处理，得到所述文本图像的文本结构特征，其中，所述第二卷积层的卷积核尺寸为 $N \times N$ ， N 为大于 1 的整数。

13、根据权利要求 10-12 中任意一项所述的装置，其特征在于，所述结果获取模块包括：

融合子模块，用于对所述文本关联特征和所述特征信息包括的文本结构特征进行融合处理，得到融合特征；

结果获取子模块，用于根据所述融合特征，获取所述文本图像的文本识别结果。

14、根据权利要求 10-13 中任意一项所述的装置，其特征在于，所述装置适用于神经网络，所述神经网络中的编码网络包括多个网络块，每个网络块包括卷积核尺寸为 P

$\times Q$ 的第一卷积层和卷积核尺寸为 $N \times N$ 的第二卷积层, 其中, 所述第一卷积层和所述第二卷积层的输入端分别与所述网络块的输入端连接。

15、根据权利要求 13 所述的装置, 其特征在于, 所述装置适用于神经网络, 所述神经网络中的编码网络包括多个网络块, 所述融合子模块用于:

5 对所述多个网络块中第一网络块的第一卷积层输出的文本关联特征和所述第一网络块的第二卷积层输出的文本结构特征进行融合, 得到所述第一网络块的融合特征;

所述结果获取子模块用于:

对所述第一网络块的融合特征和所述第一网络块的输入信息进行残差处理, 得到所述第一网络块的输出信息;

10 基于所述第一网络块的输出信息, 得到所述文本识别结果。

16、根据权利要求 14 或 15 所述的装置, 其特征在于, 所述神经网络中的编码网络包括下采样网络以及与所述下采样网络的输出端连接的多级特征提取网络, 其中, 每级特征提取网络包括至少一个所述网络块以及与所述至少一个网络块的输出端连接的下采样模块。

15 17、根据权利要求 14-16 中任意一项所述的装置, 其特征在于, 所述神经网络为卷积神经网络。

18、根据权利要求 10 至 17 中任一项所述的装置, 其特征在于, 所述特征提取模块包括:

下采样子模块, 用于对所述文本图像进行下采样处理, 得到下采样结果;

20 第三提取子模块, 用于对所述下采样结果进行特征提取, 得到所述文本图像的特征信息。

19、一种电子设备, 其特征在于, 包括:

处理器;

用于存储处理器可执行指令的存储介质;

25 其中, 所述处理器被配置为调用所述存储介质存储的指令, 以执行权利要求 1 至 9 中任意一项所述的方法。

20、一种机器可读存储介质, 其上存储有机器可执行指令, 其特征在于, 所述机器可执行指令被处理器执行时实现权利要求 1 至 9 中任意一项所述的方法。

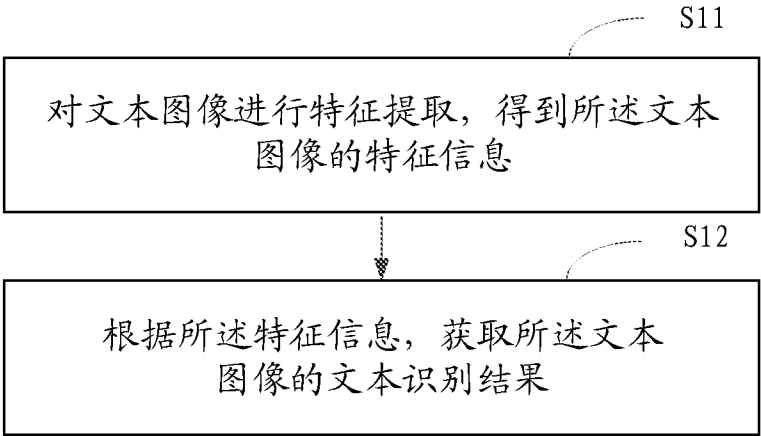


图 1

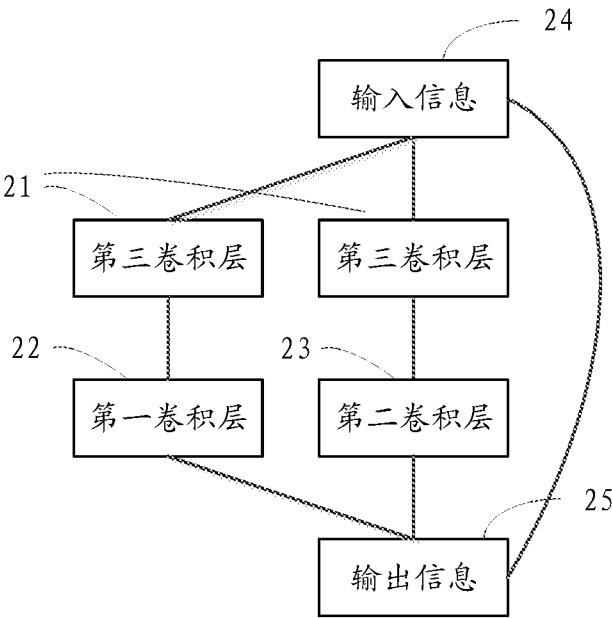


图 2

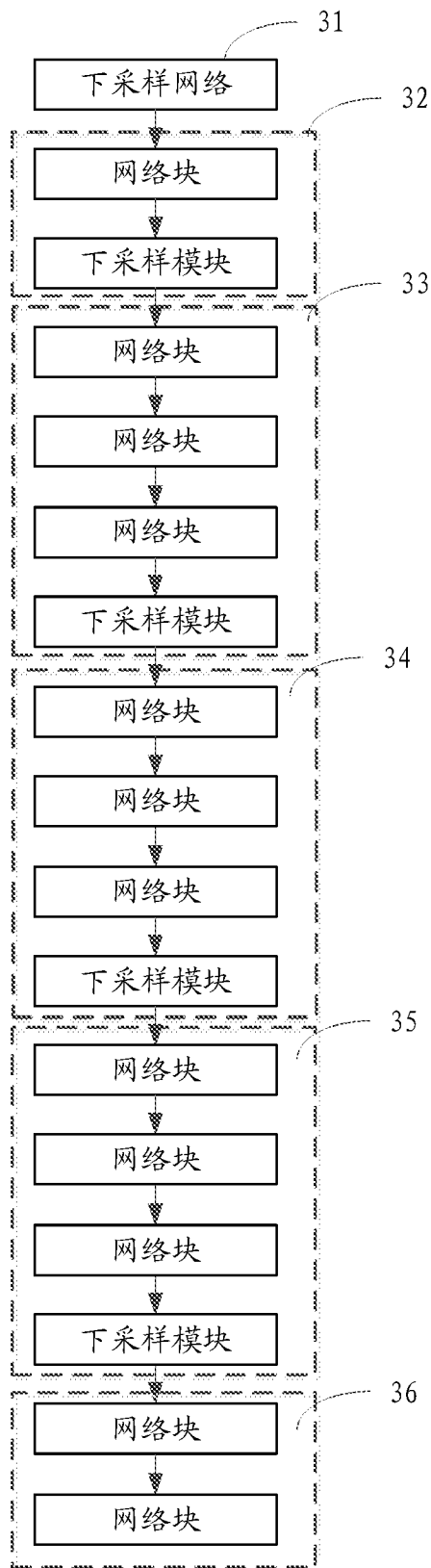


图 3

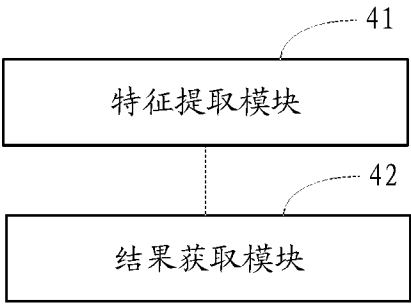


图 4

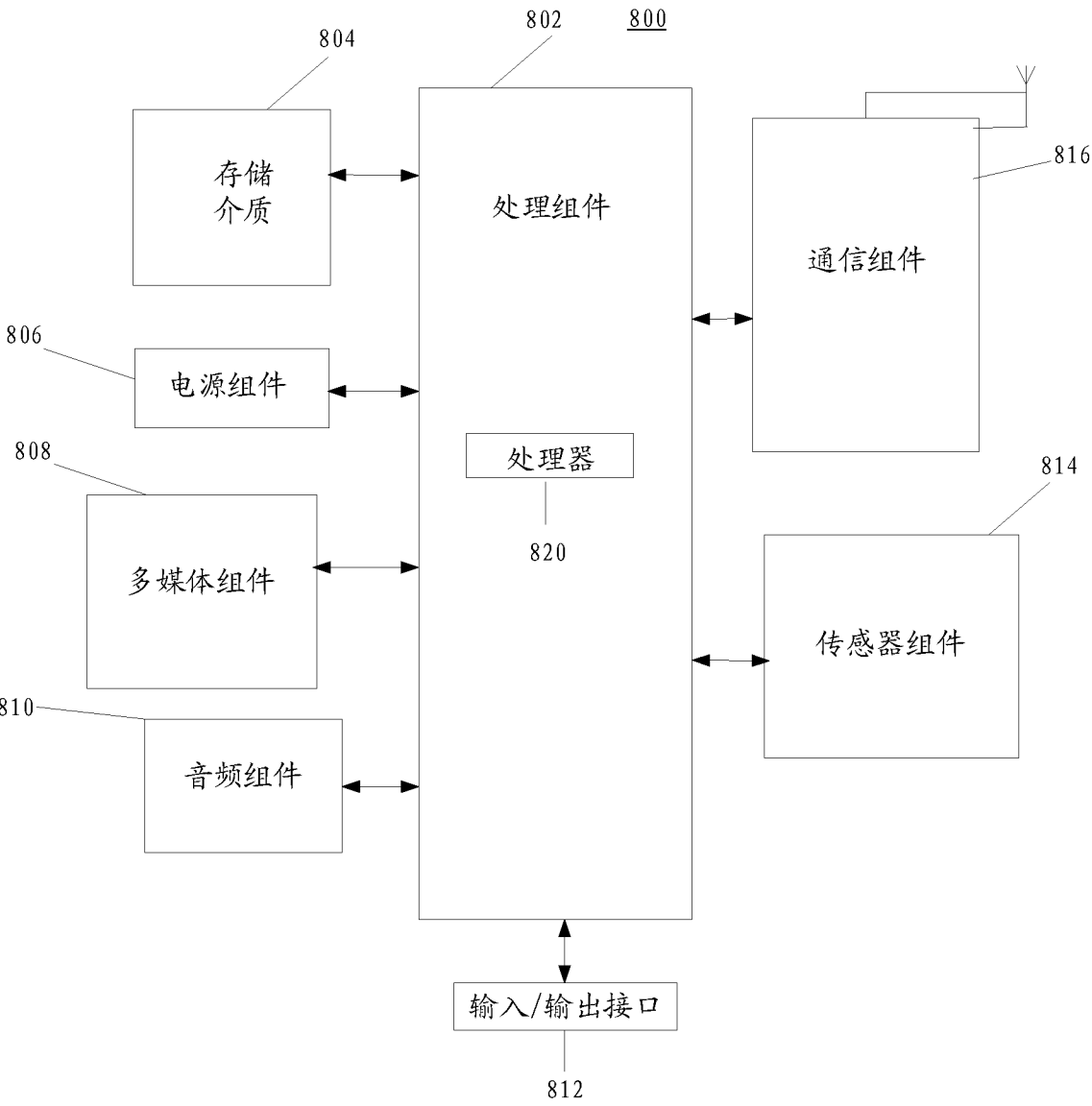


图 5

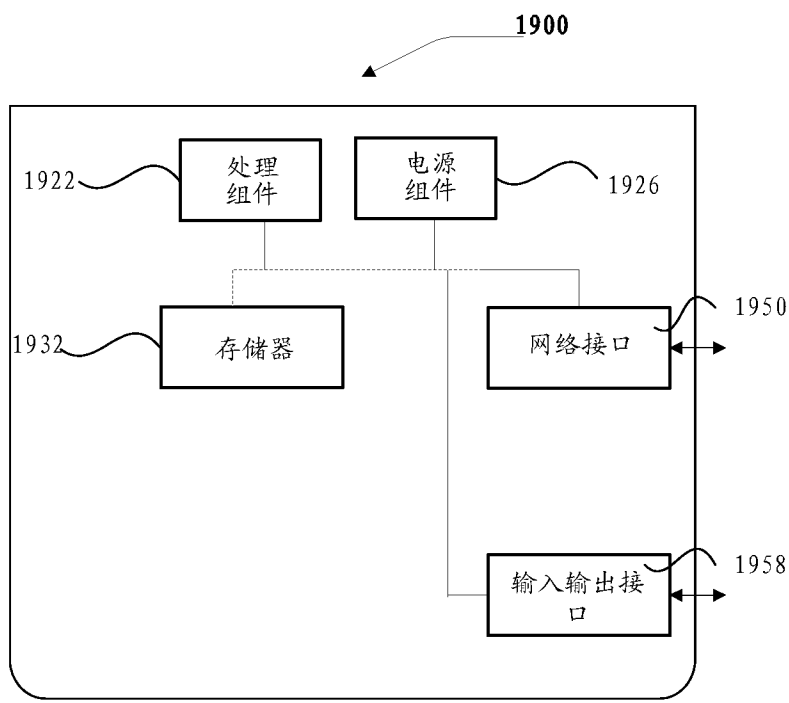


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/070568

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/34(2006.01)i; G06K 9/62(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPTXT; USTXT; VEN; WOTXT; CNABS; CNTXT; CNKI: 光学字符识别, 文字, 文本, 字符, 检测, 识别, 特征, 关联, 联系, 频率, 顺序, 次序, 卷积神经网络, OCR, character, recognise, recognize, identify, characteristic, feature, association, relation, frequency, order, CNN

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 109635810 A (BEIJING SCIENCE AND TECHNOLOGY CO., THREE FAST ONLINE) 16 April 2019 (2019-04-16) description, paragraphs [0006]-[0010], and figure 1	1, 10, 19, 20
Y	CN 105930842 A (SHENZHEN LONGRISE SCIENCE & TECHNOLOGY CO., LTD.) 07 September 2016 (2016-09-07) description, paragraphs [0036]-[0043] and [0053]-[0061], and figures 1 and 3	1-20
Y	CN 108764226 A (SF TECHNOLOGY CO., LTD.) 06 November 2018 (2018-11-06) description, paragraphs [0032], [0042]-[0046], [0116] and [0122], and figure 5	1-20
A	CN 105335754 A (BEIJING XIAOMI TECHNOLOGY CO., LTD.) 17 February 2016 (2016-02-17) entire document	1-20
A	US 2002085758 A1 (AYSHI, Mohammed Abu et al.) 04 July 2002 (2002-07-04) entire document	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

01 April 2020

Date of mailing of the international search report

09 April 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2020/070568

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	109635810	A	16 April 2019	None			
CN	105930842	A	07 September 2016	None			
CN	108764226	A	06 November 2018	None			
CN	105335754	A	17 February 2016	None			
US	2002085758	A1	04 July 2002	US	7010166	B2	07 March 2006

国际检索报告

国际申请号

PCT/CN2020/070568

A. 主题的分类

G06K 9/34(2006.01)i; G06K 9/62(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06K

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

EPTXT;USTXT;VEN;WOTXT;CNABS;CNTXT;CNKI:光学字符识别, 文字, 文本, 字符, 检测, 识别, 特征, 关联, 联系, 频率, 顺序, 次序, 卷积神经网络, OCR, character, recognise, recognize, identify, characteristic, feature, association, relation, frequency, order, CNN

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 109635810 A (北京三快在线科技有限公司) 2019年 4月 16日 (2019 - 04 - 16) 说明书第6-10段、图1	1, 10, 19-20
Y	CN 105930842 A (深圳市永兴元科技有限公司) 2016年 9月 7日 (2016 - 09 - 07) 说明书36-43、53-61段, 图1、3	1-20
Y	CN 108764226 A (顺丰科技有限公司) 2018年 11月 6日 (2018 - 11 - 06) 说明书第32、42-46、116、122段, 图5	1-20
A	CN 105335754 A (小米科技有限责任公司) 2016年 2月 17日 (2016 - 02 - 17) 全文	1-20
A	US 2002085758 A1 (AYSHI, Mohammed Abu 等) 2002年 7月 4日 (2002 - 07 - 04) 全文	1-20

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2020年 4月 1日

国际检索报告邮寄日期

2020年 4月 9日

ISA/CN的名称和邮寄地址

中国国家知识产权局(ISA/CN)

中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

梁静静

电话号码 86-(20)-28958273

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/070568

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	109635810	A	2019年 4月 16日	无	
CN	105930842	A	2016年 9月 7日	无	
CN	108764226	A	2018年 11月 6日	无	
CN	105335754	A	2016年 2月 17日	无	
US	2002085758	A1	2002年 7月 4日	US 7010166 B2	2006年 3月 7日