

(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
2023年10月19日 (19.10.2023)

(10) 国际公布号
WO 2023/197512 A1

- (51) 国际专利分类号:
G06F 40/126 (2020.01) *G06K 9/62* (2022.01)
G06F 40/232 (2020.01) *G06V 10/74* (2022.01)
- (21) 国际申请号: PCT/CN2022/116249
- (22) 国际申请日: 2022年8月31日 (31.08.2022)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202210371375.3 2022年4月11日 (11.04.2022) CN
- (71) 申请人: 苏州浪潮智能科技有限公司
(INSPUR SUZHOU INTELLIGENT TECHNOLOGY
CO., LTD.) [CN/CN]; 中国江苏省苏州市吴中

区吴中经济开发区郭巷街道官浦路1号9幢, Jiangsu 215000 (CN)。

- (72) 发明人: 李晓川(LI, Xiaochuan); 中国江苏省苏州市吴中区吴中经济开发区郭巷街道官浦路1号9幢, Jiangsu 215000 (CN)。 赵雅倩(ZHAO, Yaqian); 中国江苏省苏州市吴中区吴中经济开发区郭巷街道官浦路1号9幢, Jiangsu 215000 (CN)。 李仁刚(LI, Rengang); 中国江苏省苏州市吴中区吴中经济开发区郭巷街道官浦路1号9幢, Jiangsu 215000 (CN)。 郭振华(GUO, Zhenhua); 中国江苏省苏州市吴中区吴中经济开发区郭巷街道官浦路1号9幢, Jiangsu 215000 (CN)。 范宝余(FAN, Baoyu); 中国江苏省苏州市吴中区吴中经济开发区郭巷街道官浦路1号9幢, Jiangsu 215000 (CN)。

(54) Title: TEXT ERROR CORRECTION METHOD AND APPARATUS, AND ELECTRONIC DEVICE AND MEDIUM

(54) 发明名称: 一种文本纠错方法、装置、电子设备和介质

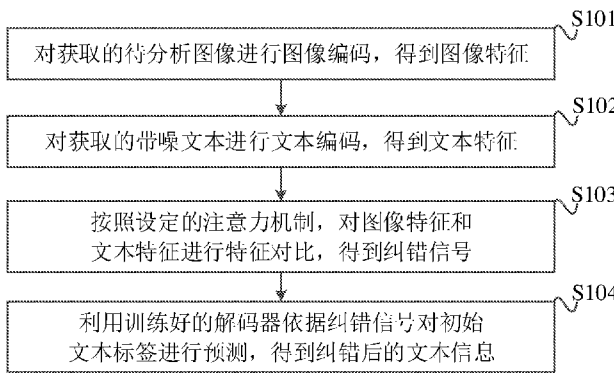


图 1

- S101 Perform image encoding on an acquired image to be analyzed, so as to obtain image features
- S102 Perform text encoding on acquired noisy text, so as to obtain text features
- S103 Perform feature comparison on the image features and the text features according to a set attention mechanism, so as to obtain an error correction signal
- S104 According to the error correction signal, predict an initial text label by using a trained decoder, so as to obtain error-corrected text information

(57) Abstract: A text error correction method and apparatus, and an electronic device and a medium. The text error correction method comprises: performing image encoding on an acquired image to be analyzed, so as to obtain image features (S101); performing text encoding on acquired noisy text, so as to obtain text features (S102); performing feature comparison on the image features and the text features according to a set attention mechanism, so as to obtain an error correction signal (S103); and according to the error correction signal, predicting an initial text label by using a trained decoder, so as to obtain error-corrected text information (S104).

(74) 代理人: 北京市万慧达律师事务所 (BEIJING WANHUIDA LAW FIRM); 中国北京市海淀区中关村南大街1号友谊宾馆颐园写字楼2楼, Beijing 100873 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, 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, SC, 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))。

(57) 摘要: 一种文本纠错方法、装置、电子设备和介质, 文本纠错方法包括: 对获取的待分析图像进行图像编码, 得到图像特征(S101); 对获取的带噪文本进行文本编码, 得到文本特征(S102); 按照设定的注意力机制, 对图像特征和文本特征进行特征对比, 得到纠错信号(S103); 以及利用训练好的解码器依据纠错信号对初始文本标签进行预测, 得到纠错后的文本信息(S104)。

一种文本纠错方法、装置、电子设备和介质

相关申请的交叉引用

本申请要求于 2022 年 04 月 11 日提交中国专利局，申请号为 202210371375.3，申请名称为“一种文本纠错方法、装置、电子设备和介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及一种文本纠错方法、装置、电子设备和计算机可读存储介质。

背景技术

近年来，多模态（Multi Modal，MM）成为人工智能领域中新兴的研究方向，像视觉常识推理（Visual Commonsense Reasoning，VCR）、视觉问答（Visual Question Answering，VQA）等领域均成为行业重点研究课题。然而在多模态领域，现有课题基本都是假定人类语言在多模态过程中是绝对正确的。然而对现实世界中人类而言，口误在所难免。通过实验发现，将现有多模态任务中的人类文本替换为口误文本时，原有模型的性能会大幅衰减。

以依据文本确定出图像中文本所描述的物品在图像中的位置为例，经实验测试发现，当输入为标准文本时，模型可以输出正确的坐标框；当输入为带噪的文本即模拟人类语言口误所产生的文本时，模型输出的坐标框出现了错误。在真实世界中，由于口误造成的文本语言错误是在所难免的。因此，发明人意识到，对于多模态任务而言，模型对文本语言错误的抗噪能力成为本领域亟待研究的课题之一。

可见，如何提升多模态任务的抗噪能力，是本领域技术人员需要解决的问题。

发明内容

根据本申请公开的各种实施例，提供一种文本纠错方法、装置、电子设备和计算机可读存储介质。

一种文本纠错方法，包括：

对获取的待分析图像进行图像编码，得到图像特征；

对获取的带噪文本进行文本编码，得到文本特征；

按照设定的注意力机制，对所述图像特征和所述文本特征进行特征对比，得到纠错信号；以及

利用训练好的解码器依据所述纠错信号对初始文本标签进行预测，得到纠错后的文本信息。

一种文本纠错装置，包括：

图像编码单元，用于对获取的待分析图像进行图像编码，得到图像特征；

文本编码单元，用于对获取的带噪文本进行文本编码，得到文本特征；

特征对比单元，用于按照设定的注意力机制，对所述图像特征和所述文本特征进行特征对比，得到纠错信号；以及

预测单元，用于利用训练好的解码器依据所述纠错信号对初始文本标签进行预测，得到纠错后的文本信息。

一种电子设备，包括：

存储器，用于存储计算机可读指令；以及

处理器，用于执行所述计算机可读指令以实现如上述文本纠错方法的步骤。

一种计算机可读存储介质，所述计算机可读存储介质上存储有计算机可读指令，所述计算机可读指令被一个或多个处理器执行时实现上述任一项的文本纠错方法的步骤。

本申请的一个或多个实施例的细节在下面的附图和描述中提出。本申请的其它特征和优点将从说明书、附图以及权利要求书变得明显。

附图说明

为了更清楚地说明本申请实施例，下面将对实施例中所需要使用的附图做简单的介绍，显而易见地，下面描述中的附图仅仅是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图 1 为本申请一个或多个实施例的文本纠错方法的流程示意图；

图 2 为本申请一个或多个实施例的自注意力机制对应的网络结构的示意图；

图 3 为本申请一个或多个实施例的对齐特征和文本特征进行分析的网络结构的示意图；

图 4 为本申请一个或多个实施例的文本纠错装置的结构示意图；

图 5 为本申请一个或多个实施例的电子设备的结构示意图；

图 6 为本申请一个或多个实施例的计算机可读存储介质的结构示意图。

具体实施方式

下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅是本申请一部分实施例，而不是全部实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下，所获得的所有其他实施例，都属于本申请保护范围。

本申请的说明书和权利要求书及上述附图中的术语“包括”和“具有”以及他们任何变形，意图在于覆盖不排他的包含。例如包含了一系列步骤或单元的过程、方法、系统、产品或设备没有限定于已列出的步骤或单元，而是可包括没有列出的步骤或单元。

为了使本技术领域的人员更好地理解本申请方案，下面结合附图和具体实施方式对本申请作进一步的详细说明。

接下来，详细介绍本申请的实施例所提供的文本纠错方法。图 1 为本申请一个或多个实施例的文本纠错方法的流程示意图，该方法包括：

S101：对获取的待分析图像进行图像编码，得到图像特征。

带噪文本是以文字形式描述目标物，待分析图像可以是包含有目标物的图像。为了实现对待分析图像中目标物的着重分析，可以对待分析图像进行编码。编码得到的图像特征反映了待分析图像中与目标物强相关的特征。图像编码方式属于较为成熟的技术，在此不做赘述。

S102：对获取的带噪文本进行文本编码，得到文本特征。

带噪文本可以是包含有错误描述信息的文本。举例说明，待分析图像中包含有一个穿白色衣服的女孩，带噪文本描述的是“一个穿绿色衣服的女孩”。

图像特征一般以矩阵的形式呈现，为了实现图像特征与带噪文本的对比，需要对带噪文本进行文本编码，以将带噪文本转换为文本特征的形式。带噪文本包含有多少个字符，文本特征对应的会包含多少个特征。

S103：按照设定的注意力机制，对图像特征和文本特征进行特征对比，得到纠错信号。

在本申请实施例中，为了基于图像特征对文本特征中的错误描述信息进行修正，可以采用注意力机制，分析图像特征和文本特征中存在差异的特征。

注意力机制可以包括自注意力机制和跨注意力机制。

在一个或多个实施例中，可以按照自注意力机制，对图像特征和文本特征进行关联性分析，得到对齐特征。按照自注意力机制和跨注意力机制，对对齐特征和文本特征进行分析，得到纠错信号。

其中，对齐特征可以包括图像特征和文本特征的对对应关系。

通过自注意力机制可以充分学习图像特征和文本特征之间的对应关系。自注意力机制对应的网络结构的示意图如图 2 所示，自注意力机制对应的网络结构包括一个自注意力层、一个层归一化和一个相加模块。将图像特征和文本特征进行拼接之后，可以将其输入到自注意力机制对应的网络结构中进行编码，从而得到最终的对齐特征。

获取纠错信号是实现文本纠错的关键步骤，对齐特征和文本特征进行分析的网络结构的示意图如图 3 所示，按照自注意力机制对对齐特征 f 和文本特征 g 各自进行注意力分析，可以得到对齐特征的自注意力特征和文本特征的自注意力特征。将对齐特征的自注意力特征和文本特征的自注意力特征进行跨注意力分析，可以得到跨注意力向量。图 3 中为了区分对齐特征和文本特征各自对应的两个分支，将对齐特征所在分支包含跨注意力分析标记为跨注意力层 A，将文本特征所在分支包含跨注意力分析标记为跨注意力层 B。通过对文本特征所在分支的跨注意力向量进行层归一化、相加和纠错处理，可以最终得到纠错信号。其中，纠错处理可以基于若干个纠错层叠加实现。

S104：利用训练好的解码器依据纠错信号对初始文本标签进行预测，得到纠错后的文本信息。

在本申请实施例中，可以利用一些已知正确文本信息的图像，预先对解码器进行训练。在具体实现中，可以收集历史图像，以及历史图像对应的历史带噪文本和正确文本。按照上述 S101 至 S103 的操作，对历史图像及其对应的历史带噪文本进行处理，从而得到历史纠错信号。在获取到历史纠错信号之后，可以利用历史纠错信号和正确文本对解码器进行训练，以得到训练好的解码器。

需要说明的是，在得到训练好的解码器之后，后续直接利用训练好的解码器依据纠错信号对初始文本标签进行预测即可，无需每次预测时都对解码器进行训练。

初始文本标签可以包括起始符号，在本申请实施例中，可以对纠错信号以及初始文本标签进行自注意力分析，确定出与初始文本标签相邻的下一个字符；将下一个字符添加至初始文本标签，并返回对纠错信号以及初始文本标签进行自注意力分析，确定出与初始文本标签相邻的下一个字符的步骤，直至下一个字符为结束字符，则将当前的初始文本标签作为纠错后的文本信息。

举例说明，假设带噪文本包含的是“穿绿色裙子的女孩”，待分析图像中包含的是一个穿白色裙子的女孩。初始文本标签可以为包含一个初始符号“start”的字符，利用训练好的解码器依据纠错信号对初始文本标签进行预测，可以依次得到“穿”、“白”、“色”“裙”、“子”、“的”、“女”、“孩”，循环使用解码器预测下一个字符，直达产生结束符号

“end”表示该预测过程结束，此时得到的“穿白色裙子的女孩”即为纠错后的文本信息。

由上述技术方案可以看出，对获取的待分析图像进行图像编码，得到图像特征；图像特征反映了待分析图像中与目标物强相关的特征。带噪文本是以文字形式对目标物进行描述。带噪文本中存在错误的描述信息，为了实现对带噪文本的纠错，可以对获取的带噪文本进行文本编码，得到文本特征。按照设定的注意力机制，对图像特征和文本特征进行特征对比，得到纠错信号。纠错信号包含了文本特征和图像特征存在差异的特征，以及带噪文本所表征的文本信息。利用训练好的解码器依据纠错信号对初始文本标签进行预测，可以得到纠错后的文本信息。在该技术方案中，通过图像所表征的特征对带噪文本进行修正，可以得到包含正确信息的文本，降低了带噪文本中错误的描述信息对模型性能的影响，提升了多模态任务的抗噪能力。

在一个或多个实施例中，自注意力机制有其对应的注意力计算公式，可以按照如下公式（1），确定出图像特征和文本特征的自注意力向量；其中，自注意力向量可以包含图像特征的每维特征与文本特征的每维特征的关联特征；

$$attention(f) = \text{soft max} \left(\frac{(W_q \cdot f)^T \times (W_k \cdot f)}{\sqrt{\text{size}(f)}} \right) \times (W_v \cdot f) \quad (1);$$

其中， $\text{soft max}(x) = \frac{e^x}{\sum_{j=1}^n e^x}$ ， x 表示 $\left(\frac{(W_q \cdot f)^T \times (W_k \cdot f)}{\sqrt{\text{size}(f)}} \right)$ ， f 表示拼接后的图像特征和文本特征， W_q 、 W_k 、 W_v 均为模型训练得到的模型参数；以及

对自注意力向量进行层归一化和相加处理，可以得到对齐特征。

对于对齐特征和文本特征的分析过程，可以包括按照自注意力机制，对对齐特征进行注意力分析，得到对齐特征的自注意力特征；按照自注意力机制，对文本特征进行注意力分析，得到文本特征的自注意力特征；按照如下公式（2），确定出对齐特征的自注意力特征和文本特征的自注意力特征之间的跨注意力向量，

$$attention_A(f, g) = \text{soft max} \left(\frac{(W_q \cdot f)^T \times (W_k \cdot g)}{\sqrt{\text{size}(g)}} \right) \times (W_v \cdot g) \quad (2);$$

其中， f 表示对齐特征的自注意力向量， g 表示文本特征的自注意力向量， W_q 、 W_k 、 W_v 均为模型训练得到的模型参数；以及

对跨注意力向量进行层归一化、相加和纠错处理，得到纠错信号。

考虑到通常情况下，带噪文本中需要纠错的字是非常少的，如果一句话大多数文字都有错误，将无法通过正确的文字判断哪里错了进而纠错。另一方面，纠错信号代表句子纠错的方向，因此需要控制绝大多数文字的特征在这个方向上为零，因此在本申请实施例中可以设计阈注意力机制，控制文字纠错信号的生成。也即除了按照上述公式（2）计算跨注意力向量外，在本申请实施例中，也可以设置阈注意力机制，其对应的公式包括公式（3）和公式（4）。

在具体实现中，可以按照如下公式（3）和（4），确定出对齐特征的自注意力特征和文本特征的自注意力特征之间的跨注意力向量，

$$attention_B(g, f) = thresh - relu \left(\text{soft max} \left(\frac{(W_q \cdot g)^T \times (W_k \cdot f)}{\sqrt{size(g)}} \right) \times (W_v \cdot f) \right) \quad (3);$$

$$thresh - relu(x) = \begin{cases} 0, & x < thresh \\ x, & x \geq thresh \end{cases} \quad (4);$$

其中， x 表示 $\left(\text{soft max} \left(\frac{(W_q \cdot g)^T \times (W_k \cdot f)}{\sqrt{size(g)}} \right) \times (W_v \cdot f) \right)$ ， f 表示对齐特征的自注意力向量， g 表示文本特征的自注意力向量， W_q 、 W_k 、 W_v 均为模型训练得到的模型参数， $thresh$ 表示设定的阈值；以及

对跨注意力向量进行层归一化、相加和纠错处理，得到纠错信号。

在本申请实施例中，采用阈注意力机制用来生成纠错信号，可以实现对图像特征强相关的文本特征进一步加强，对图像特征弱相关的文本特征进行削弱，从而达到修正目的。

图 4 为本申请实施例提供的文本纠错装置的结构示意图，包括图像编码单元 41、文本编码单元 42、特征对比单元 43 和预测单元 44；其中，

图像编码单元 41，用于对获取的待分析图像进行图像编码，得到图像特征；

文本编码单元 42，用于对获取的带噪文本进行文本编码，得到文本特征；

特征对比单元 43，用于按照设定的注意力机制，对图像特征和文本特征进行特征对比，得到纠错信号；以及

预测单元 44，用于利用训练好的解码器依据纠错信号对初始文本标签进行预测，得到纠错后的文本信息。

在一个或多个实施例中，注意力机制包括自注意力机制和跨注意力机制；

特征对比单元包括第一分析子单元和第二分析子单元；

第一分析子单元，用于按照自注意力机制，对图像特征和文本特征进行关联性分析，得到对齐特征；其中，对齐特征包括图像特征和文本特征的对应关系；以及

第二分析子单元，用于按照自注意力机制和跨注意力机制，对对齐特征和文本特征进行分析，得到纠错信号。

在一个或多个实施例中，第一分析子单元用于按照如下公式，确定图像特征和文本特征的自注意力向量；其中，自注意力向量包含图像特征的每维特征与文本特征的每维特征的关联特征；

$$attention(f) = \text{soft max} \left(\frac{(W_q \cdot f)^T \times (W_k \cdot f)}{\sqrt{\text{size}(f)}} \right) \times (W_v \cdot f) ;$$

$$\text{soft max}(x) = \frac{e^x}{\sum_{j=1}^n e^{x_j}}, \quad x \text{ 表示 } \left(\frac{(W_q \cdot f)^T \times (W_k \cdot f)}{\sqrt{\text{size}(f)}} \right),$$

其中，

f 表示拼接后的图像特征和文本特征， W_q 、 W_k 、 W_v 均为模型训练得到的模型参数；

对自注意力向量进行层归一化和相加处理，得到对齐特征。

在一个或多个实施例中，第二分析子单元用于按照自注意力机制，对对齐特征进行注意力分析，得到对齐特征的自注意力特征；

按照自注意力机制，对文本特征进行注意力分析，得到文本特征的自注意力特征；

按照如下公式，确定出对齐特征的自注意力特征和文本特征的自注意力特征之间的跨注意力向量，

$$attention_A(f, g) = \text{soft max} \left(\frac{(W_q \cdot f)^T \times (W_k \cdot g)}{\sqrt{\text{size}(g)}} \right) \times (W_v \cdot g) ;$$

其中， f 表示对齐特征的自注意力向量， g 表示文本特征的自注意力向量， W_q 、 W_k 、

W_v 均为模型训练得到的模型参数；以及

对跨注意力向量进行层归一化、相加和纠错处理，得到纠错信号。

在一个或多个实施例中，第二分析子单元用于按照自注意力机制，对对齐特征进行注意力分析，得到对齐特征的自注意力特征；

按照自注意力机制，对文本特征进行注意力分析，得到文本特征的自注意力特征；

按照如下公式，确定出对齐特征的自注意力特征和文本特征的自注意力特征之间的跨注意力向量，

$$attention_g(g, f) = thresh - relu \left(\text{soft max} \left(\frac{(W_q \cdot g)^T \times (W_k \cdot f)}{\sqrt{size(g)}} \right) \times (W_v \cdot f) \right);$$

$$thresh - relu(x) = \begin{cases} 0, & x < thresh \\ x, & x \geq thresh \end{cases};$$

其中， x 表示 $\left(\text{soft max} \left(\frac{(W_q \cdot g)^T \times (W_k \cdot f)}{\sqrt{size(g)}} \right) \times (W_v \cdot f) \right)$ ， f 表示对齐特征的自注意力向量， g 表示文本特征的自注意力向量， W_q 、 W_k 、 W_v 均为模型训练得到的模型参数， $thresh$ 表示设定的阈值；以及

对跨注意力向量进行层归一化、相加和纠错处理，得到纠错信号。

在一个或多个实施例中，初始文本标签包括起始符号；

预测单元包括确定子单元和添加子单元；

确定子单元，用于对纠错信号以及初始文本标签进行自注意力分析，确定出与初始文本标签相邻的下一个字符；以及

添加子单元，用于将下一个字符添加至初始文本标签，并返回对纠错信号以及初始文本标签进行自注意力分析，确定出与初始文本标签相邻的下一个字符的步骤，直至下一个字符为结束字符，则将当前的初始文本标签作为纠错后的文本信息。

在一个或多个实施例中，针对于解码器的训练过程，装置包括获取单元和训练单元；获取单元，用于获取历史纠错信号及其对应的正确文本；以及

训练单元，用于利用历史纠错信号和正确文本对解码器进行训练，以得到训练好的解码器。

关于文本纠错装置的具体限定可以参见上文中对于文本纠错方法的限定，在此不再赘述。上述文本纠错装置中的各个单元可全部或部分通过软件、硬件及其组合来实现。上述各单元可以硬件形式内嵌于或独立于计算机设备中的处理器中，也可以以软件形式存储于计算机设备中的存储器中，以便于处理器调用执行以上各个单元对应的操作。

由上述技术方案可以看出，对获取的待分析图像进行图像编码，得到图像特征；图像特征反映了待分析图像中与目标物强相关的特征。带噪文本是以文字形式对目标物进行描述。带噪文本中存在错误的描述信息，为了实现对带噪文本的纠错，可以对获取的带噪文本进行文本编码，得到文本特征。按照设定的注意力机制，对图像特征和文本特征进行特征对比，得到纠错信号。纠错信号包含了文本特征和图像特征存在差异的特征，以及带噪文本所表征的文本信息。利用训练好的解码器依据纠错信号对初始文本标签进行预测，可以得到纠错后的文本信息。在该技术方案中，通过图像所表征的特征对带噪文本进行修正，可以得到包含正确信息的文本，降低了带噪文本中错误的描述信息对模型性能的影响，提升了多模态任务的抗噪能力。

图 5 为本申请实施例提供的电子设备的结构示意图，如图 5 所示，电子设备包括：

存储器 20，用于存储计算机可读指令 201；以及

一个或多个处理器 21，用于执行计算机可读指令 201 时实现如上述的任一实施例的文本纠错方法的步骤。

本实施例的电子设备的可以包括但不限于智能手机、平板电脑、笔记本电脑或台式电脑等。

其中，处理器 21 可以包括一个或多个处理核心，比如 4 核心处理器、8 核心处理器等。处理器 21 可以采用 DSP (Digital Signal Processing, 数字信号处理)、FPGA (Field Programmable Gate Array, 现场可编程门阵列)、PLA (Programmable Logic Array, 可编程逻辑阵列) 中的至少一种硬件形式来实现。处理器 21 也可以包括主处理器和协处理器，主处理器是用于对在唤醒状态下的数据进行处理的处理单元，也称 CPU (Central Processing Unit, 中央处理器)；协处理器是用于对在待机状态下的数据进行处理的低功耗处理单元。在一些实施例中，处理器 21 可以在集成有 GPU (Graphics Processing Unit, 图像处理单元)，GPU 用于负责显示屏所需要显示的内容的渲染和绘制。一些实施例中，处理器 21 还可以包括 AI (Artificial Intelligence, 人工智能) 处理单元，该 AI 处理单元用于处理有关机器学习的计算操作。

存储器 20 可以包括一个或多个计算机可读存储介质，该计算机可读存储介质可以是非暂态的。存储器 20 还可包括高速随机存取存储器，以及非易失性存储器，比如一个或

多个磁盘存储设备、闪存存储设备。本实施例中，存储器 20 至少用于存储以下计算机可读指令 201，其中，该计算机可读指令 201 被处理器 21 加载并执行之后，能够实现前述任一实施例公开的文本纠错方法的相关步骤。另外，存储器 20 所存储的资源还可以包括操作系统 202 和数据 203 等，存储方式可以是短暂存储或者永久存储。其中，操作系统 202 可以包括 Windows、Unix、Linux 等。数据 203 可以包括但不限于图像特征、文本特征、注意力机制等。

在一些实施例中，电子设备还可包括有显示屏 22、输入输出接口 23、通信接口 24、电源 25 以及通信总线 26。

本领域技术人员可以理解，图 5 中示出的结构并不构成对电子设备的限定，可以包括比图示更多或更少的组件。

可以理解的是，如果上述实施例中的文本纠错方法以软件功能单元的形式实现并作为独立的产品销售或使用，可以存储在一个计算机可读存储介质中。基于这样的理解，本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，执行本申请各个实施例方法的全部或部分步骤。而前述的存储介质包括：U 盘、移动硬盘、只读存储器（Read-Only Memory，ROM）、随机存取存储器（Random Access Memory，RAM）、电可擦除可编程 ROM、寄存器、硬盘、可移动磁盘、CD-ROM、磁碟或者光盘等各种可以存储程序代码的介质。

在一个或多个实施例中，本申请实施例还提供一种计算机可读存储介质，参考图 6 所示，计算机可读存储介质 60 上存储有计算机可读指令 61，计算机可读指令 61 被一个或多个处理器执行时实现如上述的任一实施例的文本纠错方法的步骤。

本申请实施例所述计算机可读存储介质的各功能模块的功能可根据上述方法实施例中的方法具体实现，其具体实现过程可以参照上述方法实施例的相关描述，此处不再赘述。

以上对本申请实施例所提供的一种文本纠错方法、装置、电子设备和计算机可读存储介质进行了详细介绍。说明书中各个实施例采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似部分互相参见即可。对于实施例公开的装置而言，由于其与实施例公开的方法相对应，所以描述的比较简单，相关之处参见方法部分说明即可。

专业人员还可以进一步意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、计算机软件或者二者的结合来实现，为了清楚地说明硬件

和软件的可互换性，在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本申请的范围。

以上对本申请所提供的一种文本纠错方法、装置、电子设备和计算机可读存储介质进行了详细介绍。本文中应用了具体个例对本申请的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本申请的方法及其核心思想。应当指出，对于本技术领域的普通技术人员来说，在不脱离本申请原理的前提下，还可以对本申请进行若干改进和修饰，这些改进和修饰也落入本申请权利要求的保护范围内。

权利要求书

- 1、一种文本纠错方法，其特征在于，包括：
对获取的待分析图像进行图像编码，得到图像特征；
对获取的带噪文本进行文本编码，得到文本特征；
按照设定的注意力机制，对所述图像特征和所述文本特征进行特征对比，得到纠错信号；以及
利用训练好的解码器依据所述纠错信号对初始文本标签进行预测，得到纠错后的文本信息。
- 2、根据权利要求 1 所述的方法，其特征在于，所述文本特征的个数与所述带噪文本包含的字符数相同。
- 3、根据权利要求 1 或 2 所述的方法，其特征在于，所述注意力机制包括自注意力机制和跨注意力机制，所述按照设定的注意力机制，对所述图像特征和所述文本特征进行特征对比，得到纠错信号，包括：
按照所述自注意力机制，对所述图像特征和所述文本特征进行关联性分析，得到对齐特征；以及
按照所述自注意力机制和所述跨注意力机制，对所述对齐特征和所述文本特征进行分析，得到纠错信号。
- 4、根据权利要求 3 所述的方法，其特征在于，所述对齐特征包括所述图像特征和所述文本特征的对应关系。
- 5、根据权利要求 3 或 4 所述的方法，其特征在于，所述自注意力机制包括一个自注意力层、一个层归一化模块和一个相加模块。
- 6、根据权利要求 3 或 4 或 5 所述的方法，其特征在于，所述按照所述自注意力机制，对所述图像特征和所述文本特征进行关联性分析，得到对齐特征，包括：
将所述图像特征和所述文本特征进行拼接，将拼接后的所述图像特征和所述文本特征输入至所述自注意力机制中进行编码，得到所述自注意力机制输出的所述对齐特征。
- 7、根据权利要求 3 或 4 或 5 所述的方法，其特征在于，所述按照所述自注意力机制，对所述图像特征和所述文本特征进行关联性分析，得到对齐特征，包括：
确定所述图像特征和所述文本特征的自注意力向量；以及
对所述自注意力向量进行层归一化和相加处理，得到对齐特征。
- 8、根据权利要求 7 所述的方法，其特征在于，所述自注意力向量包含所述图像特征的每维特征与所述文本特征的每维特征的关联特征。

9、根据权利要求 8 所述的方法，其特征在于，所述确定所述图像特征和所述文本特征的自注意力向量，包括：

按照如下公式确定所述图像特征和所述文本特征的自注意力向量，

$$attention(f) = \text{soft max} \left(\frac{(W_q \cdot f)^T \times (W_k \cdot f)}{\sqrt{\text{size}(f)}} \right) \times (W_v \cdot f) ;$$

$$\text{soft max}(x) = \frac{e^x}{\sum_{j=1}^n e^x}$$

其中， f 表示拼接后的所述图像特征和所述文本特征， W_q 、 W_k 、 W_v 均为模型训练得到的模型参数。

10、根据权利要求 3 或 4 或 5 所述的方法，其特征在于，所述按照所述自注意力机制和所述跨注意力机制，对所述对齐特征和所述文本特征进行分析，得到纠错信号，包括：

按照所述自注意力机制，对所述对齐特征进行注意力分析，得到所述对齐特征的自注意力特征；

按照所述自注意力机制，对所述文本特征进行注意力分析，得到所述文本特征的自注意力特征；

确定出所述对齐特征的自注意力特征和所述文本特征的自注意力特征之间的跨注意力向量；以及

对所述跨注意力向量进行层归一化、相加和纠错处理，得到纠错信号。

11、根据权利要求 10 所述的方法，其特征在于，所述纠错处理基于多个纠错层的叠加实现。

12、根据权利要求 10 所述的方法，其特征在于，所述确定所述对齐特征的自注意力特征和所述文本特征的自注意力特征之间的跨注意力向量，包括：

按照如下公式确定所述对齐特征的自注意力特征和所述文本特征的自注意力特征之间的跨注意力向量，

$$attention_A(f, g) = \text{soft max} \left(\frac{(W_q \cdot f)^T \times (W_k \cdot g)}{\sqrt{\text{size}(g)}} \right) \times (W_v \cdot g) ;$$

其中， f 表示对齐特征的自注意力向量， g 表示文本特征的自注意力向量， W_q 、 W_k 、

W_v 均为模型训练得到的模型参数。

13、根据权利要求 10 所述的方法，其特征在于，所述确定所述对齐特征的自注意力特征和所述文本特征的自注意力特征之间的跨注意力向量，包括：

设置阈注意力机制，通过所述阈注意力机制确定所述对齐特征的自注意力特征和所述文本特征的自注意力特征之间的跨注意力向量。

14、根据权利要求 10 所述的方法，其特征在于，所述确定所述对齐特征的自注意力特征和所述文本特征的自注意力特征之间的跨注意力向量，包括：

按照如下公式确定所述对齐特征的自注意力特征和所述文本特征的自注意力特征之间的跨注意力向量，

$$attention_B(g, f) = thresh - relu \left(\text{soft max} \left(\frac{(W_q \cdot g)^T \times (W_k \cdot f)}{\sqrt{\text{size}(g)}} \right) \times (W_v \cdot f) \right);$$

$$thresh - relu(x) = \begin{cases} 0, & x < thresh \\ x, & x \geq thresh \end{cases};$$

其中， f 表示对齐特征的自注意力向量， g 表示文本特征的自注意力向量， W_q 、 W_k 、 W_v 均为模型训练得到的模型参数， $thresh$ 表示设定的阈值。

15、根据权利要求 1 至 14 任意一项所述的方法，其特征在于，所述初始文本标签包括起始符号，所述利用训练好的解码器依据所述纠错信号对初始文本标签进行预测，得到纠错后的文本信息，包括：

对所述纠错信号以及所述初始文本标签进行自注意力分析，确定出与所述初始文本标签相邻的下一个字符；以及

将所述下一个字符添加至所述初始文本标签，并返回所述对所述纠错信号以及所述初始文本标签进行自注意力分析，确定出与所述初始文本标签相邻的下一个字符的步骤，直至所述下一个字符为结束字符，则将当前的初始文本标签作为纠错后的文本信息。

16、根据权利要求 1 至 15 任意一项所述的方法，其特征在于，所述方法还包括：对所述解码器进行训练。

17、根据权利要求 16 所述的方法，其特征在于，所述对所述解码器进行训练，包括：获取历史纠错信号及其对应的正确文本；以及

利用所述历史纠错信号和所述正确文本对所述解码器进行训练，以得到训练好的解

码器。

18、一种文本纠错装置，其特征在于，包括：

图像编码单元，用于对获取的待分析图像进行图像编码，得到图像特征；

文本编码单元，用于对获取的带噪文本进行文本编码，得到文本特征；

特征对比单元，用于按照设定的注意力机制，对所述图像特征和所述文本特征进行特征对比，得到纠错信号；以及

预测单元，用于利用训练好的解码器依据所述纠错信号对初始文本标签进行预测，得到纠错后的文本信息。

19、一种电子设备，其特征在于，包括：

存储器，用于存储计算机可读指令；以及

一个或多个处理器，用于执行所述计算机可读指令以实现如权利要求 1 至 17 任意一项所述的方法的步骤。

20、一种计算机可读存储介质，其特征在于，所述计算机可读存储介质上存储有计算机可读指令，所述计算机可读指令被一个或多个处理器执行时实现如权利要求 1 至 17 任意一项所述的方法的步骤。

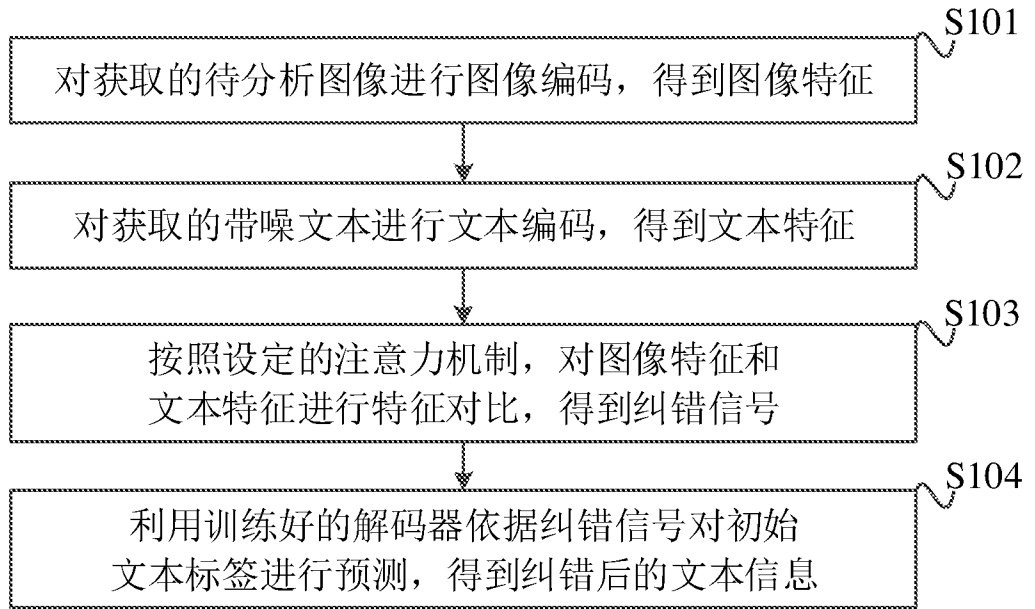


图 1

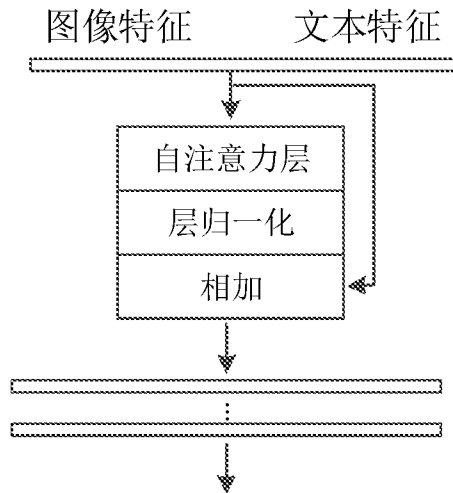


图 2

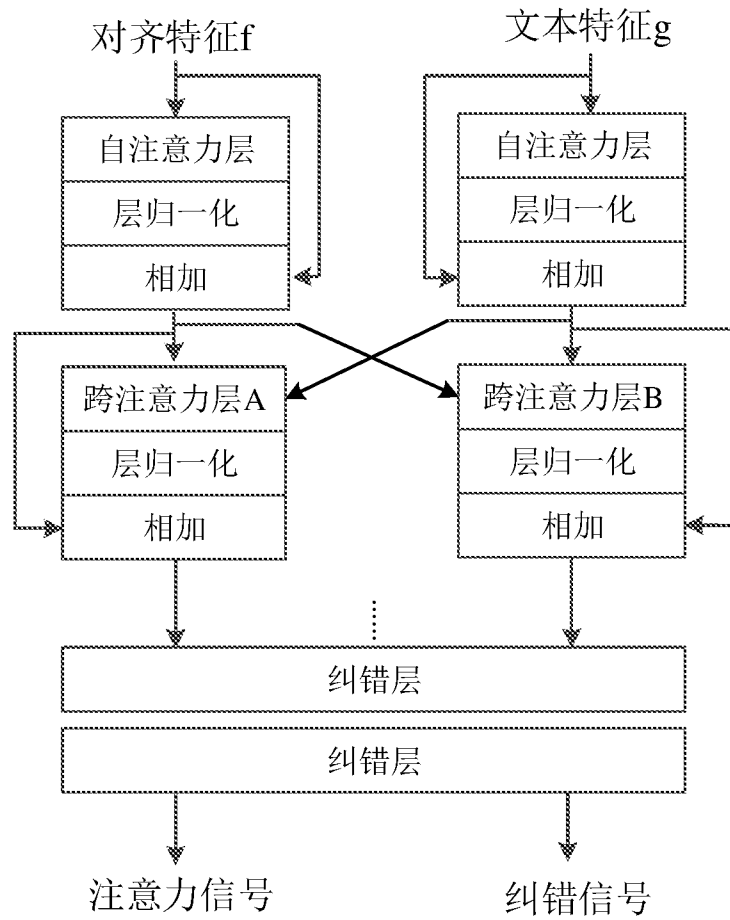


图 3

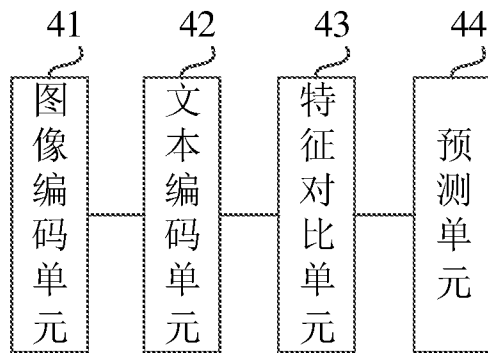


图 4

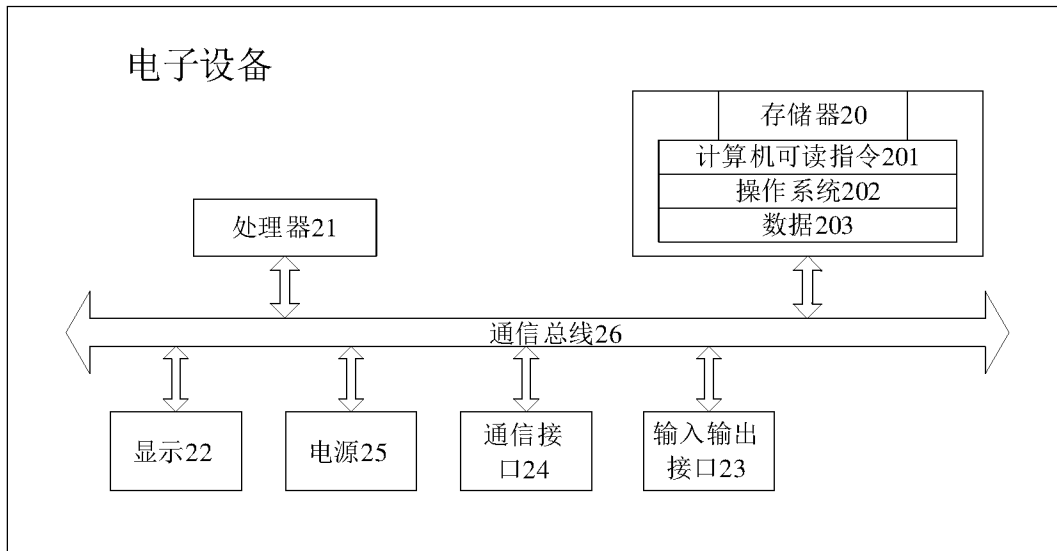


图 5

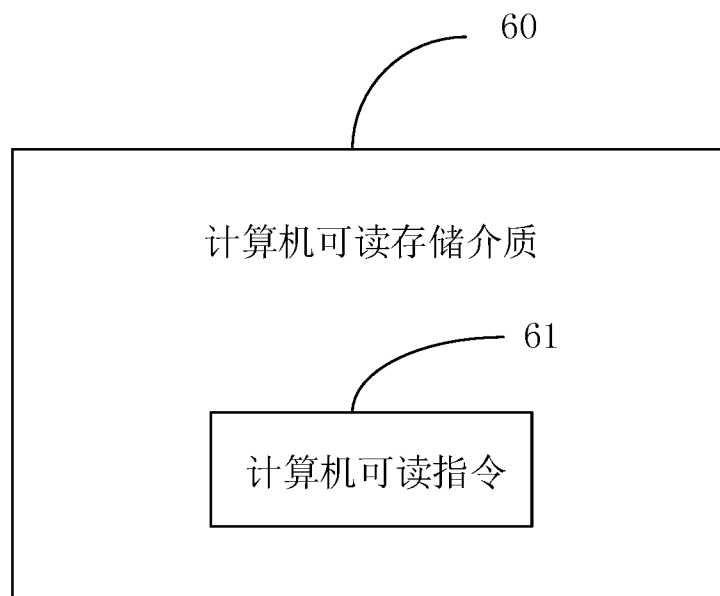


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/116249

A. CLASSIFICATION OF SUBJECT MATTER		
G06F 40/126(2020.01)i; G06F 40/232(2020.01)n; G06K 9/62(2022.01)n; G06V 10/74(2022.01)n		
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) G06F; G06K; G06V		
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) CNABS; CNTXT; CNKI; 万方, WANFANG; 百度, BAIDU; DWPI; SIPOABS; VEN; ENTXT; IEEE; 注意力, 参数, 归一, 去噪, 纠错, 公式, 文本, 自注意力, 跨注意力, 图像, 抗噪, 降噪, 编码, 预测, attention, text, image, noise, error, correct, reduce, code, predict		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 114241279 A (IFLYTEK INTERNET (BEIJING) INFORMATION TECHNOLOGY CO., LTD. et al.) 25 March 2022 (2022-03-25) description, paragraphs 3-148	1, 2, 15-20
PX	CN 114462356 A (SUZHOU INSPUR INTELLIGENT TECHNOLOGY CO., LTD.) 10 May 2022 (2022-05-10) claims 1-10, and description, paragraphs 28-28	1-20
A	CN 113743101 A (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.) 03 December 2021 (2021-12-03) entire document	1-20
A	CN 112633290 A (BEIJING CENTURY TAL EDUCATION TECHNOLOGY CO., LTD.) 09 April 2021 (2021-04-09) entire document	1-20
A	CN 112905827 A (UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA) 04 June 2021 (2021-06-04) entire document	1-20
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> 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 27 September 2022		Date of mailing of the international search report 19 October 2022
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

International application No.

PCT/CN2022/116249

C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 112632912 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 09 April 2021 (2021-04-09) entire document	1-20
A	WO 2021232589 A1 (PINGAN INTERNATIONAL SMART CITY TECHNOLOGY CO., LTD.) 25 November 2021 (2021-11-25) entire document	1-20
A	US 5761686 A (XEROX CORP.) 02 June 1998 (1998-06-02) entire document	1-20

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2022/116249

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)		Publication date (day/month/year)
CN	114241279	A	25 March 2022	None		
CN	114462356	A	10 May 2022	None		
CN	113743101	A	03 December 2021	None		
CN	112633290	A	09 April 2021	None		
CN	112905827	A	04 June 2021	None		
CN	112632912	A	09 April 2021	WO	2022126897	23 June 2022
WO	2021232589	A1	25 November 2021	CN	111737458	02 October 2020
US	5761686	A	02 June 1998	JP	H10124289	A 15 May 1998

国际检索报告

国际申请号

PCT/CN2022/116249

<p>A. 主题的分类</p> <p>G06F 40/126(2020.01)i; G06F 40/232(2020.01)n; G06K 9/62(2022.01)n; G06V 10/74(2022.01)n</p> <p>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</p>																										
<p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>G06F; G06K; G06V</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>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))</p> <p>CNABS;CNTXT;CNKI;万方;百度;DWPI;SIPOABS;VEN;ENTXT;IEEE:注意力, 参数, 归一, 去噪, 纠错, 公式, 文本, 自注意力, 跨注意力, 图像, 抗噪, 降噪, 编码, 预测, attention, text, image, noise, error, correct, reduce, code, predict</p>																										
<p>C. 相关文件</p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>CN 114241279 A (中科讯飞互联北京信息科技有限公司等) 2022年3月25日 (2022 - 03 - 25) 说明书第3-148段</td> <td>1, 2, 15-20</td> </tr> <tr> <td>PX</td> <td>CN 114462356 A (苏州浪潮智能科技有限公司) 2022年5月10日 (2022 - 05 - 10) 权利要求1-10, 说明书第28-28段</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 113743101 A (北京百度网讯科技有限公司) 2021年12月3日 (2021 - 12 - 03) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 112633290 A (北京世纪好未来教育科技有限公司) 2021年4月9日 (2021 - 04 - 09) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 112905827 A (中国科学技术大学) 2021年6月4日 (2021 - 06 - 04) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 112632912 A (平安科技深圳有限公司) 2021年4月9日 (2021 - 04 - 09) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>WO 2021232589 A1 (PINGAN INTERNATIONAL SMART CITY TECHNOLOGY CO., LTD.) 2021年11月25日 (2021 - 11 - 25) 全文</td> <td>1-20</td> </tr> </tbody> </table>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	CN 114241279 A (中科讯飞互联北京信息科技有限公司等) 2022年3月25日 (2022 - 03 - 25) 说明书第3-148段	1, 2, 15-20	PX	CN 114462356 A (苏州浪潮智能科技有限公司) 2022年5月10日 (2022 - 05 - 10) 权利要求1-10, 说明书第28-28段	1-20	A	CN 113743101 A (北京百度网讯科技有限公司) 2021年12月3日 (2021 - 12 - 03) 全文	1-20	A	CN 112633290 A (北京世纪好未来教育科技有限公司) 2021年4月9日 (2021 - 04 - 09) 全文	1-20	A	CN 112905827 A (中国科学技术大学) 2021年6月4日 (2021 - 06 - 04) 全文	1-20	A	CN 112632912 A (平安科技深圳有限公司) 2021年4月9日 (2021 - 04 - 09) 全文	1-20	A	WO 2021232589 A1 (PINGAN INTERNATIONAL SMART CITY TECHNOLOGY CO., LTD.) 2021年11月25日 (2021 - 11 - 25) 全文	1-20
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																								
X	CN 114241279 A (中科讯飞互联北京信息科技有限公司等) 2022年3月25日 (2022 - 03 - 25) 说明书第3-148段	1, 2, 15-20																								
PX	CN 114462356 A (苏州浪潮智能科技有限公司) 2022年5月10日 (2022 - 05 - 10) 权利要求1-10, 说明书第28-28段	1-20																								
A	CN 113743101 A (北京百度网讯科技有限公司) 2021年12月3日 (2021 - 12 - 03) 全文	1-20																								
A	CN 112633290 A (北京世纪好未来教育科技有限公司) 2021年4月9日 (2021 - 04 - 09) 全文	1-20																								
A	CN 112905827 A (中国科学技术大学) 2021年6月4日 (2021 - 06 - 04) 全文	1-20																								
A	CN 112632912 A (平安科技深圳有限公司) 2021年4月9日 (2021 - 04 - 09) 全文	1-20																								
A	WO 2021232589 A1 (PINGAN INTERNATIONAL SMART CITY TECHNOLOGY CO., LTD.) 2021年11月25日 (2021 - 11 - 25) 全文	1-20																								
<p><input checked="" type="checkbox"/> 其余文件在C栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。</p>																										
<p>* 引用文件的具体类型:</p> <p>“A” 认为不特别相关的表示了现有技术一般状态的文件</p> <p>“E” 在国际申请日的当天或之后公布的在先申请或专利</p> <p>“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)</p> <p>“O” 涉及口头公开、使用、展览或其他方式公开的文件</p> <p>“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</p> <p>“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</p> <p>“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</p> <p>“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</p> <p>“&” 同族专利的文件</p>																										
<p>国际检索实际完成的日期</p> <p>2022年9月27日</p>		<p>国际检索报告邮寄日期</p> <p>2022年10月19日</p>																								
<p>ISA/CN的名称和邮寄地址</p> <p>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</p> <p>传真号 (86-10)62019451</p>		<p>受权官员</p> <p>陈响</p> <p>电话号码 (86-27)59371802</p>																								

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
A	US 5761686 A (XEROX CORP.) 1998年6月2日 (1998 - 06 - 02) 全文	1-20

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2022/116249

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	114241279	A	2022年3月25日	无			
CN	114462356	A	2022年5月10日	无			
CN	113743101	A	2021年12月3日	无			
CN	112633290	A	2021年4月9日	无			
CN	112905827	A	2021年6月4日	无			
CN	112632912	A	2021年4月9日	WO	2022126897		2022年6月23日
WO	2021232589	A1	2021年11月25日	CN	111737458		2020年10月2日
US	5761686	A	1998年6月2日	JP	H10124289	A	1998年5月15日