

(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



(10) 国际公布号  
WO 2023/201975 A1

(43) 国际公布日  
2023年10月26日 (26.10.2023)

- (51) 国际专利分类号:  
G06V 30/40 (2022.01) G06V 30/18 (2022.01)  
G06V 30/418 (2022.01)
- (21) 国际申请号: PCT/CN2022/118852
- (22) 国际申请日: 2022年9月15日 (15.09.2022)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:  
202210407134.X 2022年4月19日 (19.04.2022) 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)。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)。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)。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: DIFFERENCE DESCRIPTION SENTENCE GENERATION METHOD AND APPARATUS, AND DEVICE AND MEDIUM

(54) 发明名称: 一种差异描述语句生成方法、装置、设备及介质

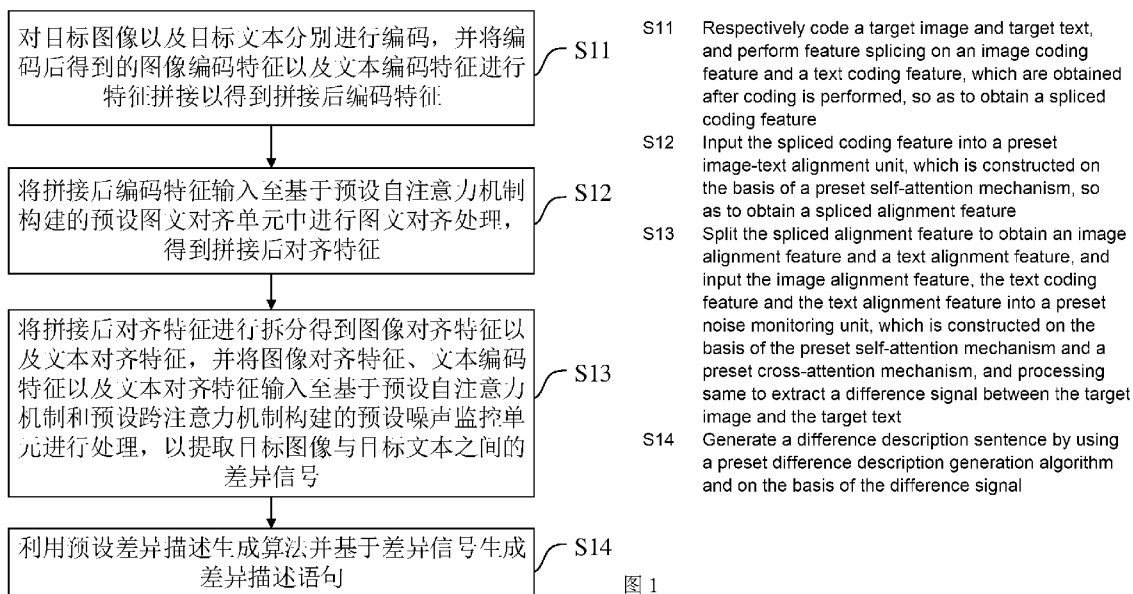


图1

(57) Abstract: The present application relates to the technical field of artificial intelligence. Disclosed are a difference description sentence generation method and apparatus, and a device and a medium. The method comprises: performing feature splicing on an image coding feature and a text coding feature; inputting a spliced coding feature into a preset image-text alignment unit, which is constructed on the basis of a preset self-attention mechanism, so as to obtain a spliced alignment feature; by using a preset noise monitoring unit, which is constructed on the basis of the preset self-attention mechanism and a preset cross-attention mechanism, respectively processing the text coding feature, and an image alignment feature and a text alignment feature, which are obtained by means of splitting the spliced

WO 2023/201975 A1

(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))。

---

alignment feature, so as to extract a difference signal; and generating a difference description sentence by using a preset difference description generation algorithm and on the basis of the difference signal.

(57) 摘要: 本申请公开了一种差异描述语句生成方法、装置、设备及介质, 涉及人工智能技术领域, 该方法包括: 对图像编码特征和文本编码特征进行特征拼接, 将拼接后编码特征输入至基于预设自注意力机制构建的预设图文对齐单元得到拼接后对齐特征, 利用基于预设自注意力机制和预设跨注意力机制构建的预设噪声监控单元分别对文本编码特征以及对拼接后对齐特征拆分后得到的图像对齐特征和文本对齐特征进行处理以提取差异信号; 利用预设差异描述生成算法并基于差异信号生成差异描述语句。

## 一种差异描述语句生成方法、装置、设备及介质

### 相关申请的交叉引用

本申请要求于 2022 年 04 月 19 日提交中国专利局，申请号为 202210407134.X，申请名称为“一种差异描述语句生成方法、装置、设备及介质”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

本申请涉及人工智能技术领域，特别涉及一种差异描述语句生成方法、装置、设备及介质。

### 背景技术

近年来，多模态成为人工智能领域中新兴的研究方向，例如视觉常识推理（Visual Commonsense Reasoning, VCR）、视觉问答（Visual Question Answering, VQA）等领域均成为行业重点研究方向。在多模态领域，现有课题都基于这样一个假设：人类语言在多模态过程中是绝对正确的，即人类语言必然会与图像匹配。然而，发明人意识到，对于人类而言，口误在所难免，虽然通常情况下，人类语言错误不会非常离谱，也就是说文本和图像本身之间是比较接近的，但是某个主语或定语的错误使用会使计算机模型不知道人类究竟想描述的问题是什么，在这种情况下，给计算机赋予与人类有针对性交互的能力是很有必要的。

为此，避免因人类语言错误导致的计算机无法正常推理，增强人机交互体验是本领域亟待解决的问题。

### 发明内容

第一方面，本申请公开了一种差异描述语句生成方法，包括：

对目标图像以及目标文本分别进行编码，并将编码后得到的图像编码特征以及文本编码特征进行特征拼接以得到拼接后编码特征；

将拼接后编码特征输入至基于预设自注意力机制构建的预设图文对齐单元中进行图文对齐处理，得到拼接后对齐特征；

将拼接后对齐特征进行拆分得到图像对齐特征以及文本对齐特征，并将图像对齐特征、文本编码特征以及文本对齐特征输入至基于预设自注意力机制和预设跨注意力机制构建的预设噪声监控单元进行处理，以提取目标图像与目标文本之间的差异信号；和  
利用预设差异描述生成算法并基于差异信号生成差异描述语句。

第二方面，本申请公开了一种差异描述语句生成装置，包括：

特征拼接模块，用于对目标图像以及目标文本分别进行编码，并将编码后得到的图像编码特征以及文本编码特征进行特征拼接以得到拼接后编码特征；

图文对齐模块，用于将拼接后编码特征输入至基于预设自注意力机制构建的预设图文对齐单元中进行图文对齐处理，得到拼接后对齐特征；

噪声监控模块，用于将拼接后对齐特征进行拆分得到图像对齐特征以及文本对齐特征，并将图像对齐特征、文本编码特征以及文本对齐特征输入至基于预设自注意力机制和预设跨注意力机制构建的预设噪声监控单元进行处理，以提取目标图像与目标文本之间的差异信号；和

差异描述语句生成模块，用于利用预设差异描述生成算法并基于差异信号生成差异描述语句。

第三方面，本申请公开了一种电子设备，存储器及一个或多个处理器，存储器中储存有计算机可读指令，上述计算机可读指令被上述一个或多个处理器执行时，使得上述一个或多个处理器执行上述公开的差异描述语句生成方法。

第四方面，本申请公开了一个或多个存储有计算机可读指令的非易失性计算机可读存储介质，上述计算机可读指令被一个或多个处理器执行时，使得上述一个或多个处理器执行前述公开的差异描述语句生成方法。

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或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请的实施例，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，还可以根据提供的附图获得其他的附图。

图1为本申请根据一个或多个实施例公开的一种差异描述语句生成方法流程图；

图2为本申请根据一个或多个实施例公开的一种具体的差异描述语句生成方法子流程图；

图3为本申请根据一个或多个实施例公开的一种具体的差异描述语句生成方法子流程图；

图4为一种传统的视觉问答样本结构示意图；

图5为本申请根据一个或多个实施例公开的一种视觉问答样本结构示意图；

图6为本申请根据一个或多个实施例公开的一种具体的差异描述语句生成方法流程图；

图7为本申请根据一个或多个实施例公开的一种图文对齐模块结构示意图；

图8为本申请根据一个或多个实施例公开的一种噪声监控模块结构示意图；

图9为本申请根据一个或多个实施例公开的一种描述生成模块结构示意图；

图10为本申请根据一个或多个实施例公开的一种差异描述语句生成装置结构示意图；

图11为本申请公开的一种电子设备结构图。

## 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，所描述的实施例仅仅是本申请一部分实施例，而不是全部的实施例。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，都属于本申请保护的范围。

在多模态领域，由于人类语言的错误会使得计算机模型不知道人类究竟想描述的问题是什么，进而使得计算机无法正常推理。

为此，本申请实施例提出一种差异描述语句生成方案，能够解决因人类语言错误导致的计算机无法正常推理的问题，进而增强人机交互体验。

本申请实施例公开了一种差异描述语句生成方法，参见图1所示，该方法以应用于计算机设备为例进行说明，该方法包括：

步骤S11：对目标图像以及目标文本分别进行编码，并将编码后得到的图像编码特征以及文本编码特征进行特征拼接以得到拼接后编码特征。

需要指出的是，对目标图像进行编码的图像编码器可以采用卷积神经网络ResNet（Residual Neural Network）、ResNext等，对目标文本进行编码的图像编码器可以采用

Roberta、Bert (Bidirectional Encoder Representation from Transformers) 等。在得到图像编码特征以及文本编码特征之后，需要按照预设比例对图像编码特征以及文本编码特征进行特征拼接以得到拼接后编码特征。

步骤S12：将拼接后编码特征输入至基于预设自注意力机制构建的预设图文对齐单元中进行图文对齐处理，得到拼接后对齐特征。

本实施例中，在将拼接后编码特征输入至基于预设自注意力机制构建的预设图文对齐单元中进行图文对齐处理之前，还包括：基于自注意力运算单元、层归一化单元以及特征相加单元构建图文对齐子单元；通过将第一预设数量的图文对齐子单元进行依次串接，以构建得到预设图文对齐单元。如此一来，将拼接后编码特征输入至基于预设自注意力机制构建的预设图文对齐单元中进行图文对齐处理，具体包括：确定预设图文对齐单元中的第一个图文对齐子单元为当前图文对齐子单元，并确定拼接后编码特征为待处理特征；将待处理特征输入至当前图文对齐子单元中；利用当前图文对齐子单元对待处理特征依次进行自注意力运算、层归一化运算以及特征相加运算，以得到相应的当前运算处理结果；判断当前图文对齐子单元是否为最后一个；若否，则将当前图文对齐子单元更新为下一个图文对齐子单元，将待处理特征更新为当前运算处理结果，并返回执行将待处理特征输入至当前图文对齐子单元中的步骤；若是，则确定当前运算处理结果为拼接后对齐特征。

需要指出的是，利用当前图文对齐子单元对待处理特征依次进行自注意力运算、层归一化运算以及特征相加运算具体可以包括：利用当前图文对齐子单元中的自注意力运算单元对待处理特征进行自注意力运算，得到第一运算特征；利用当前图文对齐子单元中的层归一化单元对第一运算特征进行层归一化处理，得到第二运算特征；利用当前图文对齐子单元中的特征相加单元对第二运算特征与待处理特征进行特征相加运算，以得到当前运算处理结果。

本实施例中，利用预设图文对齐单元可以学习到图像与文本之间的匹配关系，如此一来，对于图文特征不能对齐的部分就是错误部分，基于此，人类语言中与事实不符的部分可以被定位出来。

步骤S13：将拼接后对齐特征进行拆分得到图像对齐特征以及文本对齐特征，并将图像对齐特征、文本编码特征以及文本对齐特征输入至基于预设自注意力机制和预设跨注意力机制构建的预设噪声监控单元进行处理，以提取目标图像与目标文本之间的差异信号。

本实施例中，利用预设噪声监控单元中的预设自注意力机制与预设跨注意力机制对人类语言文本中与图像无法对齐的部分进行定位，如此一来，可以得到目标图像与目标文本之间的差异信号。

本实施例中，将拼接后对齐特征进行拆分得到图像对齐特征以及文本对齐特征，并输入至基于预设自注意力机制和预设跨注意力机制构建的预设噪声监控单元进行处理之前，还包括：基于自注意力运算单元、层归一化单元、特征相加单元、跨注意力运算单元以及注意力叠加单元构建噪声监控子单元；通过将第二预设数量的噪声监控子单元进行依次串接，以构建得到预设噪声监控单元。

**步骤S14：**利用预设差异描述生成算法并基于差异信号生成差异描述语句。

本实施例中，在得到差异信号之后，利用预设差异描述生成算法并基于差异信号生成差异描述语句，差异描述语句为当人类语言包含错误的时候，计算机可以给出的解释性描述语句。

可见，本申请提出一种差异描述语句生成方法，包括：对目标图像以及目标文本分别进行编码，并将编码后得到的图像编码特征以及文本编码特征进行特征拼接以得到拼接后编码特征；将拼接后编码特征输入至基于预设自注意力机制构建的预设图文对齐单元中进行图文对齐处理，得到拼接后对齐特征；将拼接后对齐特征进行拆分得到图像对齐特征以及文本对齐特征，并将图像对齐特征、文本编码特征以及文本对齐特征输入至基于预设自注意力机制和预设跨注意力机制构建的预设噪声监控单元进行处理，以提取目标图像与目标文本之间的差异信号；利用预设差异描述生成算法并基于差异信号生成差异描述语句。可见，本申请是基于预设自注意力机制和预设跨注意力机制进行图文差异噪声监测的，由于注意力机制自身的能够从复杂特征中提取重点信息的特点，使得在基于注意力机制进行图文差异噪声监测时，能够有效地监测出图像和文本之间的差异，由此提升了所监测到的图文差异噪声的准确性，接着通过对上述图文差异噪声进行相应的解释性描述，从而达到向产生错误语言文本的用户进行及时提醒的效果。综上所述，本申请通过上述技术方案，能够在出现错误语言文本的情况下保证计算机的正常推理功能，并增强了人机交互体验。

进一步的，本实施例针对前述实施例步骤S13中的基于预设自注意力机制和预设跨注意力机制构建的预设噪声监控单元提取目标图像与目标文本之间的差异信号的过程，进行详细的介绍和说明。具体的，参加图2所示，上述步骤S13包括：

**步骤S131：**确定预设噪声监控单元中的第一个噪声监控子单元为当前噪声监控子单元，并将图像对齐特征、文本编码特征以及文本对齐特征均确定为待处理特征。

步骤S132：将待处理特征输入至当前噪声监控子单元中。

步骤S133：利用当前噪声监控子单元对待处理特征依次进行自注意力运算、层归一化运算以及特征相加运算，以得到图像对齐特征、文本编码特征以及文本对齐特征各自对应的当前第一运算处理结果。

步骤S134：并对图像对齐特征、文本编码特征以及文本对齐特征各自对应的第一运算处理结果均分别进行跨注意力运算、层归一化运算以及特征相加运算，以得到图像对齐特征、文本编码特征以及文本对齐特征各自对应的当前第二运算处理结果。

需要指出的是，本实施例中，对文本编码特征对应的第一运算处理结果分别进行跨注意力运算、层归一化运算以及特征相加运算的过程具体包括：对文本编码特征对应的第一运算处理结果进行跨注意力运算，以得到相应的跨注意力运算结果；跨注意力运算包括文本编码特征对应的第一运算处理结果分别与图像对齐特征对应的第一运算处理结果以及文本对齐特征对应的第一运算处理结果之间的跨注意力运算；将跨注意力运算结果输入至基于预设公式构建注意力叠加单元进行相应的注意力叠加运算，以得到相应的注意力叠加运算结果；将注意力叠加运算结果依次进行层归一化运算以及特征相加运算后得到相应的第二运算处理结果；

其中，预设公式为：

$$f = \text{attn}_{\text{cross}}(f_{\text{align}}, g_{\text{raw}}) + \text{attn}_{\text{cross}}(g_{\text{align}}, g_{\text{raw}})$$

其中， $f$  表示注意力叠加运算结果， $f_{\text{align}}$  表示图像对齐特征， $g_{\text{align}}$  表示文本编码特征， $g_{\text{raw}}$  表示文本对齐特征， $\text{attn}_{\text{cross}}$  表示跨注意力运算。

步骤S135：判断当前噪声监控子单元是否为最后一个。

步骤S136：若否，则将当前噪声监控子单元更新为下一个噪声监控子单元，将待处理特征更新为图像对齐特征、文本编码特征以及文本对齐特征各自对应的当前第二运算处理结果，并返回执行将待处理特征输入至当前噪声监控子单元中的步骤。

步骤S137：若是，则确定当前第二运算处理结果为目标图像与目标文本之间的差异信号。

也即，本实施例基于预设自注意力机制和预设跨注意力机制进行图文差异噪声监测，由于注意力机制自身的能够从复杂特征中提取重点信息的特点，使得在基于注意力机制进行图文差异噪声监测时，能够有效地监测出图像和文本之间的差异。



进一步的，本实施例针对前述实施例步骤S14中的利用预设差异描述生成算法并基于差异信号生成差异描述语句的过程，进行详细的介绍和说明。具体的，参加图3所示，上述步骤S14包括：

步骤S141：将差异信号、第三预设数量的预设参考描述语句以及预测符输入至基于预设差异描述生成算法构建的差异描述生成单元中，并利用差异描述生成单元在与预测符对应的位置输出相应的预测结果以及与预测结果对应的分布概率。

步骤S142：将分布概率输入至预设文本生成器中计算每个预测结果的得分，并将每个预测结果的得分与预设参考描述语句的得分相加得到当前描述语句的得分。

步骤S143：对当前描述语句的得分进行降序排序，并按照排序先后顺序从降序排序结果中筛选出第三预设数量的当前描述语句作为新的预设参考描述语句。

步骤S144：判断预测结果是否为预设终止符。

步骤S145：若否，则返回执行将差异信号、第三预设数量的预设参考描述语句以及预测符输入至基于预设差异描述生成算法构建的差异描述生成单元中的步骤。

步骤S146：若是，则筛选出得分最高的当前描述语句作为差异描述语句。

需要指出的是，在第一次利用差异描述生成单元进行预测时，预设参考描述语句不包含任何内容，预设参考描述语句是在一次次循环预测中形成的，并且由于本实施例不改变特征维度，因此设定了预测符，以便在与预测符对应的位置输出相应的预测结果。

此外，由于对一个差异来说，可能存在多种描述都是正确的，因此本实施例设计了预设文本生成器用来存储中间文本，并计算每条描述语句的得分，然后通过对每条描述语句的得分进行排序，并基于排序结果确定出差异描述语句。如此一来，基于预设差异描述生成算法给出了相应的解释描述，解决了因人类语言错误导致的计算机无法正常推理的问题。

图4给出了当人类语言与图像匹配的情况下的视觉问答样本，图5给出了当人类语言包含错误的时候，本申请实施例能够给出解释性描述语句的视觉问答样本。

近两年来，由于transformer（一种NLP经典模型）结构的广泛应用，基于transformer的多模态网络结构在诸如视觉问答、图片描述（Image Caption）、视觉对话（Visual Dialog）等多模态任务中表现优异。由于transformer结构本身具有文本attention（注意力机制）的基因，其天然具有从复杂特征中提取重点的能力，这个能力使其在多模态领域脱颖而出。同时，对于差异的描述而言，attention结构同样有能力完成文本中错误的定位和解释，因此，本申请以transformer结构中的attention机制为灵感，进行改进和设计，实现差异描述语句的生成。另一方面，现有多模态任务，如短语定位（Phrase Grounding）、

相关描述理解 (Referring Expression Comprehension) 等均涉及到了文本和图像对齐的相关理论知识, 基于此本申请可以得到一个结论, 图像编码特征和文本编码特征是可以互相作为参考来对齐的, 那么对于图文不能完全匹配的样本, 其无法对齐的部分就是错误部分。基于此本申请认为: 在给定自然图像的前提下, 人类语言中与事实不符的内容根据attention机制可被定位, 对于这个差异的解释性描述也可被生成。

本申请公开了一种具体的差异描述语句生成方法流程图, 差异描述语句生成方法用来解决真实世界中因为人类语言错误导致多模态模型无法正常推理的问题, 此外, 本方法生成的描述可用来与人类沟通的性质作为本申请的附加功能; 差异描述语句生成方法流程参见图6所示, 包括图像和文本编码器、图文对齐模块、噪声监控模块、描述生成模块和生成文本存取器。

本申请所涉及图文对齐模块如图7所示, 首先, 对不同编码器的输出 (图像编码特征  $f_{raw}$  和文本编码特征  $g_{raw}$ ) 按照一定比例进行拼接, 然后将其输入到若干图文对齐子模块中进行编码, 并得到最终的图像对齐特征以及文本对齐特征, 本实施例中, 每个图文对齐子模块均包括自注意力层、层归一化层以及相加层, 在自注意力模块中, 基于第一预设公式创建预设自注意力机制。

第一预设公式为:

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

其中,  $f$  为每个自注意力运算单元的输入,  $W_q$ 、 $W_k$  以及  $W_v$  表示映

射矩阵,  $\text{soft max}(x) = \frac{e^x}{\sum_{j=1}^n e^x}$ ,  $\text{size}(f)$  表示维度。

本申请所涉及噪声监控模块如图8所示, 将图文对齐模块的输出特征按照拼接时的比例进行拆分, 并将拆分后得到的图像对齐特征、文本编码特征以及文本对齐特征输入到若干噪声监控子模块中进行编码, 以得到差异信号, 本实施例中, 每个噪声监控子模块均包括自注意力层、层归一化层、相加层、跨注意力层以及注意力叠层。需要指出的是, 在跨注意力层中进行跨注意力运算时, 基于第二预设公式确定预设跨注意力机制, 然后基于预设跨注意力机制进行相应运算。

第二预设公式为：

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

其中， $f$ ， $g$  分别表示跨注意力层中每次参与跨注意力运算的两个输入特征， $\text{size}(g)$  表示维度。

本申请通过设计噪声监控模块，将图像和文本之间的差异确定出来，输出的差异信号中带有图文之间不匹配的要素，如图5中展示出的“该女孩的裙子为白色”（图片）和“该女孩的裙子为红色”（文本），本模块的差异信号中则会带有“白色”、“女孩”、“裙子”、“红色”、“颜色”等信息。

本申请所涉及描述生成模块如图9所示，描述生成模块包括描述生成算法与生成文本存取器两部分。描述生成模块由若干个文本解码子模块级联组成，每个解码子模块包含一个自注意力层、一个层归一化层和一个相加层。描述生成算法的输入为三部分内容的拼接：差异信号，参考文本的固定编码和预测符的固定编码，输出的最后一个特征表示预测符对应位置的预测结果，是表示下一个字是什么字的概率分布。

本实施例中，需要指出的是，生成文本存取器包含四部分，分别是文本空间，用来存储当前合理的第三预设数量的差异描述语句（不要求是完整句子）；得分空间，用来存储第三预设数量的语句的相应得分；文本更新机制，用来将预设差异描述生成算法生成的新文字存储进文本空间，并删除掉得分低的相应文本；得分更新机制，用来对更新后的文本进行打分并存入得分空间中。

在一种具体实施方式中，假设第三预设数量为N，对于每次调用预设差异描述生成算法，将“文本空间”中的N条语句提取出来批量输入到预设差异描述生成算法中预测下一个文字，得到包含N×M个字和N×M个概率的概率分布；之后，将这个概率分布输入到预设文本生成器的“文本更新机制”中，从“得分空间”中读取N条语句的得分，并在“得分更新机制”中加上当前文字的N×M得分获得总得分并进行排序；将排序前N的文本存到“文本空间”中；将这N条文本更新后的得分存入“得分空间”，当预测文字为预设终止符时，结束循环，并筛选出得分最高的当前描述语句作为差异描述语句。

相应的，本申请实施例还公开了一种差异描述语句生成装置，参见图10所示，该装置包括：

特征拼接模块11，用于对目标图像以及目标文本分别进行编码，并将编码后得到的图像编码特征以及文本编码特征进行特征拼接以得到拼接后编码特征；

图文对齐模块12，用于将拼接后编码特征输入至基于预设自注意力机制构建的预设图文对齐单元中进行图文对齐处理，得到拼接后对齐特征；

噪声监控模块13，用于将拼接后对齐特征进行拆分得到图像对齐特征以及文本对齐特征，并将图像对齐特征、文本编码特征以及文本对齐特征输入至基于预设自注意力机制和预设跨注意力机制构建的预设噪声监控单元进行处理，以提取目标图像与目标文本之间的差异信号；

差异描述语句生成模块14，用于利用预设差异描述生成算法并基于差异信号生成差异描述语句。

其中，关于上述各个模块更加具体的工作过程可以参考前述实施例中公开的相应内容，在此不再进行赘述。

可见，本申请提出一种差异描述语句生成方法，包括：对目标图像以及目标文本分别进行编码，并将编码后得到的图像编码特征以及文本编码特征进行特征拼接以得到拼接后编码特征；将拼接后编码特征输入至基于预设自注意力机制构建的预设图文对齐单元中进行图文对齐处理，得到拼接后对齐特征；将拼接后对齐特征进行拆分得到图像对齐特征以及文本对齐特征，并将图像对齐特征、文本编码特征以及文本对齐特征输入至基于预设自注意力机制和预设跨注意力机制构建的预设噪声监控单元进行处理，以提取目标图像与目标文本之间的差异信号；利用预设差异描述生成算法并基于差异信号生成差异描述语句。可见，本申请是基于预设自注意力机制和预设跨注意力机制进行图文差异噪声监测的，由于注意力机制自身的能够从复杂特征中提取重点信息的特点，使得在基于注意力机制进行图文差异噪声监测时，能够有效地监测出图像和文本之间的差异，由此提升了所监测到的图文差异噪声的准确性，接着通过对上述图文差异噪声进行相应的解释性描述，从而达到向产生错误语言文本的用户进行及时提醒的效果。综上所述，本申请通过上述技术方案，能够在出现错误语言文本的情况下保证计算机的正常推理功能，并增强了人机交互体验。

进一步的，本申请实施例还提供了一种电子设备。图11是根据一示例性实施例示出的电子设备20结构图，图中的内容不能认为是对本申请的使用范围的任何限制。

图11为本申请实施例提供的一种电子设备20的结构示意图，该电子设备可以是计算机设备，该计算机设备可以是终端或服务器。该电子设备20，具体可以包括：至少一个处理器21、至少一个存储器22、显示屏23、输入输出接口24、通信接口25、电源26、和

通信总线27。其中，存储器22用于存储计算机可读指令，计算机可读指令由处理器21加载并执行，以实现前述任一实施例公开的差异描述语句生成方法中的相关步骤。另外，本实施例中的电子设备20具体可以为电子计算机。

本实施例中，电源26用于为电子设备20上的各硬件设备提供工作电压；通信接口25能够为电子设备20创建与外界设备之间的数据传输通道，其所遵循的通信协议是能够适用于本申请技术方案的任意通信协议，在此不对其进行具体限定；输入输出接口24，用于获取外界输入数据或向外界输出数据，其具体的接口类型可以根据具体应用需要进行选取，在此不进行具体限定。

另外，存储器22作为资源存储的载体，可以是只读存储器、随机存储器、磁盘或者光盘等，其上所存储的资源可以包括计算机可读指令221，存储方式可以是短暂存储或者永久存储。其中，计算机可读指令221除了包括能够用于完成前述任一实施例公开的由电子设备20执行的差异描述语句生成方法的计算机可读指令之外，还可以进一步包括能够用于完成其他特定工作的计算机可读指令。

进一步的，本申请实施例还公开了一种非易失性计算机可读存储介质，该非易失性计算机可读存储介质中存储有计算机可读指令，该计算机可读指令被一个或多个处理器执行时可实现上述任意一个实施例的公开的差异描述语句生成方法。

关于该方法的具体步骤可以参考前述实施例中公开的相应内容，在此不再进行赘述。

本申请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它实施例的不同之处，各个实施例之间相同或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。

结合本文中所公开的实施例描述的方法或算法的步骤可以直接用硬件、处理器执行的软件模块，或者二者的结合来实施。软件模块可以置于随机存储器（RAM）、内存、

只读存储器（ROM）、电可编程ROM、电可擦除可编程ROM、寄存器、硬盘、可移动磁盘、CD-ROM、或技术领域内所公知的任意其它形式的存储介质中。

最后，还需要说明的是，在本文中，诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括要素的过程、方法、物品或者设备中还存在另外的相同要素。

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机可读指令来指令相关的硬件来完成，上述的计算机可读指令可存储于一非易失性计算机可读存储介质中，该计算机可读指令在执行时，可包括如上述各方法的实施例的流程。其中，本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用，均可包括非易失性和/或易失性存储器。非易失性存储器可包括只读存储器（ROM）、可编程ROM（PROM）、电可编程ROM（EPROM）、电可擦除可编程ROM（EEPROM）或闪存。易失性存储器可包括随机存取存储器（RAM）或者外部高速缓冲存储器。作为说明而非局限，RAM以多种形式可得，诸如静态RAM（SRAM）、动态RAM（DRAM）、同步DRAM（SDRAM）、双数据率SDRAM（DDRSDRAM）、增强型SDRAM（ESDRAM）、同步链路（Synchlink）DRAM（SLDRAM）、存储器总线（Rambus）直接RAM（RDRAM）、直接存储器总线动态RAM（DRDRAM）、以及存储器总线动态RAM（RDRAM）等。

以上实施例的各技术特征可以进行任意的组合，为使描述简洁，未对上述实施例中的各个技术特征所有可能的组合都进行描述，然而，只要这些技术特征的组合不存在矛盾，都应当认为是本说明书记载的范围。

以上上述实施例仅表达了本申请的几种实施方式，其描述较为具体和详细，但并不能因此而理解为对申请专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本申请构思的前提下，还可以做出若干变形和改进，这些都属于本申请的保护范围。因此，本申请专利的保护范围应以所附权利要求为准。

## 权利要求书

1、一种差异描述语句生成方法，其特征在于，包括：

对目标图像以及目标文本分别进行编码，并将编码后得到的图像编码特征以及文本编码特征进行特征拼接以得到拼接后编码特征；

将所述拼接后编码特征输入至基于预设自注意力机制构建的预设图文对齐单元中进行图文对齐处理，得到拼接后对齐特征；

将所述拼接后对齐特征进行拆分得到图像对齐特征以及文本对齐特征，并将所述图像对齐特征、所述文本编码特征以及所述文本对齐特征输入至基于预设自注意力机制和预设跨注意力机制构建的预设噪声监控单元进行处理，以提取所述目标图像与所述目标文本之间的差异信号；和

利用预设差异描述生成算法并基于所述差异信号生成差异描述语句。

2、根据权利要求 1 所述的差异描述语句生成方法，其特征在于，所述对目标图像以及目标文本分别进行编码，包括：

采用图像编码器对所述目标图像和所述目标文本进行编码。

3、根据权利要求 2 所述的差异描述语句生成方法，其特征在于，对所述目标图像进行编码的图像编码器为卷积神经网络 ResNet 或深度神经网络 ResNext。

4、根据权利要求 2 所述的差异描述语句生成方法，其特征在于，对所述目标文本进行编码的图像编码器为 Roberta 或 Bert。

5、根据权利要求 1 所述的差异描述语句生成方法，其特征在于，所述将所述拼接后编码特征输入至基于预设自注意力机制构建的预设图文对齐单元中进行图文对齐处理之前，还包括：

基于自注意力运算单元、层归一化单元以及特征相加单元构建图文对齐子单元；和

通过将第一预设数量的所述图文对齐子单元进行依次串接，以构建得到所述预设图文对齐单元。

6、根据权利要求 5 所述的差异描述语句生成方法，其特征在于，所述将所述拼接后编码特征输入至基于预设自注意力机制构建的预设图文对齐单元中进行图文对齐处理，包括：

确定所述预设图文对齐单元中的第一个图文对齐子单元为当前图文对齐子单元，并确定所述拼接后编码特征为待处理特征；

将所述待处理特征输入至当前图文对齐子单元中；

利用当前图文对齐子单元对所述待处理特征依次进行自注意力运算、层归一化运算以及特征相加运算，以得到相应的当前运算处理结果；和

在当前图文对齐子单元不是最后一个时，将当前图文对齐子单元更新为下一个图文对齐子单元，将所述待处理特征更新为当前所述运算处理结果，并返回执行所述将所述待处理特征输入至当前图文对齐子单元中的步骤。

7、根据权利要求 6 所述的差异描述语句生成方法，其特征在于，所述方法还包括：

在当前图文对齐子单元是最后一个时，确定当前所述运算处理结果为所述拼接后对齐特征。

8、根据权利要求 6 所述的差异描述语句生成方法，其特征在于，所述利用当前图文对齐子单元对所述待处理特征依次进行自注意力运算、层归一化运算以及特征相加运算，包括：

利用当前图文对齐子单元中的所述自注意力运算单元对所述待处理特征进行自注意力运算，得到第一运算特征；

利用当前图文对齐子单元中的所述层归一化单元对所述第一运算特征进行层归一化处理，得到第二运算特征；和

利用当前图文对齐子单元中的所述特征相加单元对所述第二运算特征与所述待处理特征进行特征相加运算，以得到当前运算处理结果。

9、根据权利要求 1 所述的差异描述语句生成方法，其特征在于，所述将所述拼接后对齐特征进行拆分得到图像对齐特征以及文本对齐特征，并输入至基于预设自注意力机制和预设跨注意力机制构建的预设噪声监控单元进行处理之前，还包括：

基于自注意力运算单元、层归一化单元、特征相加单元、跨注意力运算单元以及注意力叠加单元构建噪声监控子单元；和

通过将第二预设数量的所述噪声监控子单元进行依次串接，以构建得到所述预设噪声监控单元。

10、根据权利要求 9 所述的差异描述语句生成方法，其特征在于，所述将所述拼接后对齐特征进行拆分得到图像对齐特征以及文本对齐特征，并将所述图像对齐特征、所述文本编码特征以及所述文本对齐特征输入至基于预设自注意力机制和预设跨注意力机制构建的预设噪声监控单元进行处理，以提取所述目标图像与所述目标文本之间的差异信号，包括：



确定所述预设噪声监控单元中的第一个噪声监控子单元为当前噪声监控子单元，并将所述图像对齐特征、所述文本编码特征以及所述文本对齐特征均确定为待处理特征；

将所述待处理特征输入至当前噪声监控子单元中；

利用当前噪声监控子单元对所述待处理特征依次进行自注意力运算、层归一化运算以及特征相加运算，以得到所述图像对齐特征、所述文本编码特征以及所述文本对齐特征各自对应的当前第一运算处理结果；并对所述图像对齐特征、所述文本编码特征以及所述文本对齐特征各自对应的第一运算处理结果均分别进行跨注意力运算、层归一化运算以及特征相加运算，以得到所述图像对齐特征、所述文本编码特征以及所述文本对齐特征各自对应的当前第二运算处理结果；和

在当前噪声监控子单元不是最后一个时，将当前噪声监控子单元更新为下一个噪声监控子单元，将所述待处理特征更新为所述图像对齐特征、所述文本编码特征以及所述文本对齐特征各自对应的当前第二运算处理结果，并返回执行所述将所述待处理特征输入至当前噪声监控子单元中的步骤。

11、根据权利要求 10 所述的差异描述语句生成方法，其特征在于，所述方法还包括：

在当前噪声监控子单元是最后一个时，确定当前所述第二运算处理结果为所述目标图像与所述目标文本之间的差异信号。

12、根据权利要求 10 所述的差异描述语句生成方法，其特征在于，对所述文本编码特征对应的第一运算处理结果分别进行跨注意力运算、层归一化运算以及特征相加运算的过程，包括：

对所述文本编码特征对应的第一运算处理结果进行跨注意力运算，以得到相应的跨注意力运算结果；将所述跨注意力运算结果输入至基于预设公式构建注意力叠加单元进行相应的注意力叠加运算，以得到相应的注意力叠加运算结果；和

将所述注意力叠加运算结果依次进行层归一化运算以及特征相加运算后得到相应的第二运算处理结果。

13、根据权利要求 12 所述的差异描述语句生成方法，其特征在于，所述跨注意力运算包括所述文本编码特征对应的第一运算处理结果分别与所述图像对齐特征对应的第一运算处理结果以及所述文本对齐特征对应的第一运算处理结果之间的跨注意力运算。

14、根据权利要求 12 所述的差异描述语句生成方法，其特征在于，所述预设公式为：

$$f = \text{attn}_{\text{cross}}(f_{\text{align}}, g_{\text{raw}}) + \text{attn}_{\text{cross}}(g_{\text{align}}, g_{\text{raw}});$$

其中,  $f$  表示所述注意力叠加运算结果,  $f_{\text{align}}$  表示所述图像对齐特征,

$g_{\text{align}}$  表示所述文本编码特征,  $g_{\text{raw}}$  表示所述文本对齐特征,  $\text{attn}_{\text{cross}}$  表示跨注意力运算。

15、根据权利要求 1 至 14 任一项所述的差异描述语句生成方法, 其特征在于, 所述利用预设差异描述生成算法并基于所述差异信号生成差异描述语句, 包括:

将所述差异信号、第三预设数量的预设参考描述语句以及预测符输入至基于预设差异描述生成算法构建的差异描述生成单元中, 并利用所述差异描述生成单元在与所述预测符对应的位置输出相应的预测结果以及与所述预测结果对应的分布概率;

将所述分布概率输入至预设文本生成器中计算每个所述预测结果的得分, 并将所述每个所述预测结果的得分与所述预设参考描述语句的得分相加得到当前描述语句的得分;

对所述当前描述语句的得分进行降序排序, 并按照排序先后顺序从降序排序结果中筛选出所述第三预设数量的所述当前描述语句作为新的所述预设参考描述语句; 和

在所述预测结果不是预设终止符时, 返回执行所述将所述差异信号、第三预设数量的预设参考描述语句以及预测符输入至基于预设差异描述生成算法构建的差异描述生成单元中的步骤。

16、根据权利要求 16 所述的差异描述语句生成方法, 其特征在于, 所述方法还包括:

在所述预测结果是预设终止符时, 筛选出得分最高的所述当前描述语句作为差异描述语句。

17、一种差异描述语句生成装置, 其特征在于, 包括:

特征拼接模块, 用于对目标图像以及目标文本分别进行编码, 并将编码后得到的图像编码特征以及文本编码特征进行特征拼接以得到拼接后编码特征;

图文对齐模块, 用于将所述拼接后编码特征输入至基于预设自注意力机制构建的预设图文对齐单元中进行图文对齐处理, 得到拼接后对齐特征;

噪声监控模块, 用于将所述拼接后对齐特征进行拆分得到图像对齐特征以及文本对齐特征, 并将所述图像对齐特征、所述文本编码特征以及所述文本对齐特征输入至基于

预设自注意力机制和预设跨注意力机制构建的预设噪声监控单元进行处理，以提取所述目标图像与所述目标文本之间的差异信号；和

差异描述语句生成模块，用于利用预设差异描述生成算法并基于所述差异信号生成差异描述语句。

18、根据权利要求 17 所述的差异描述语句生成装置，其特征在于，所述差异描述语句生成装置还用于基于自注意力运算单元、层归一化单元以及特征相加单元构建图文对齐子单元，通过将第一预设数量的所述图文对齐子单元进行依次串接，以构建得到所述预设图文对齐单元。

19、一种电子设备，其特征在于，包括存储器及一个或多个处理器，所述存储器中储存有计算机可读指令，所述计算机可读指令被所述一个或多个处理器执行时，使得所述一个或多个处理器执行如权利要求 1 至 16 任一项所述的差异描述语句生成方法。

20、一个或多个存储有计算机可读指令的非易失性计算机可读存储介质，其特征在于，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器执行如权利要求 1 至 16 任一项所述的差异描述语句生成方法。

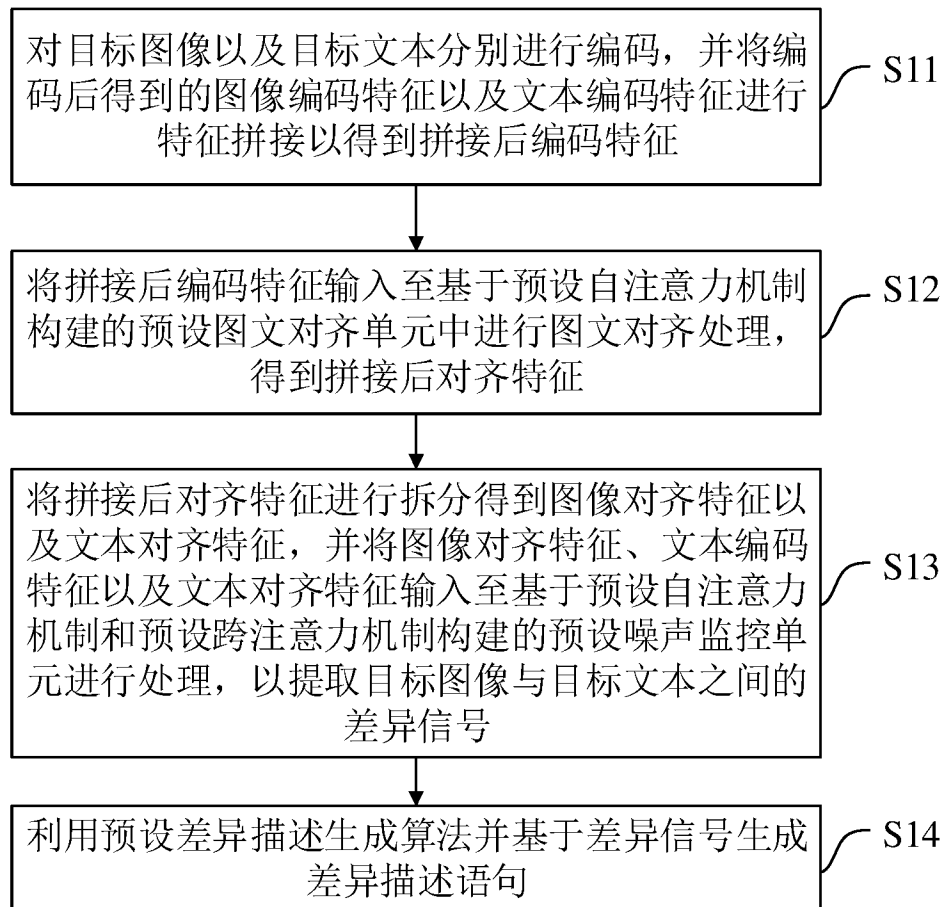


图 1

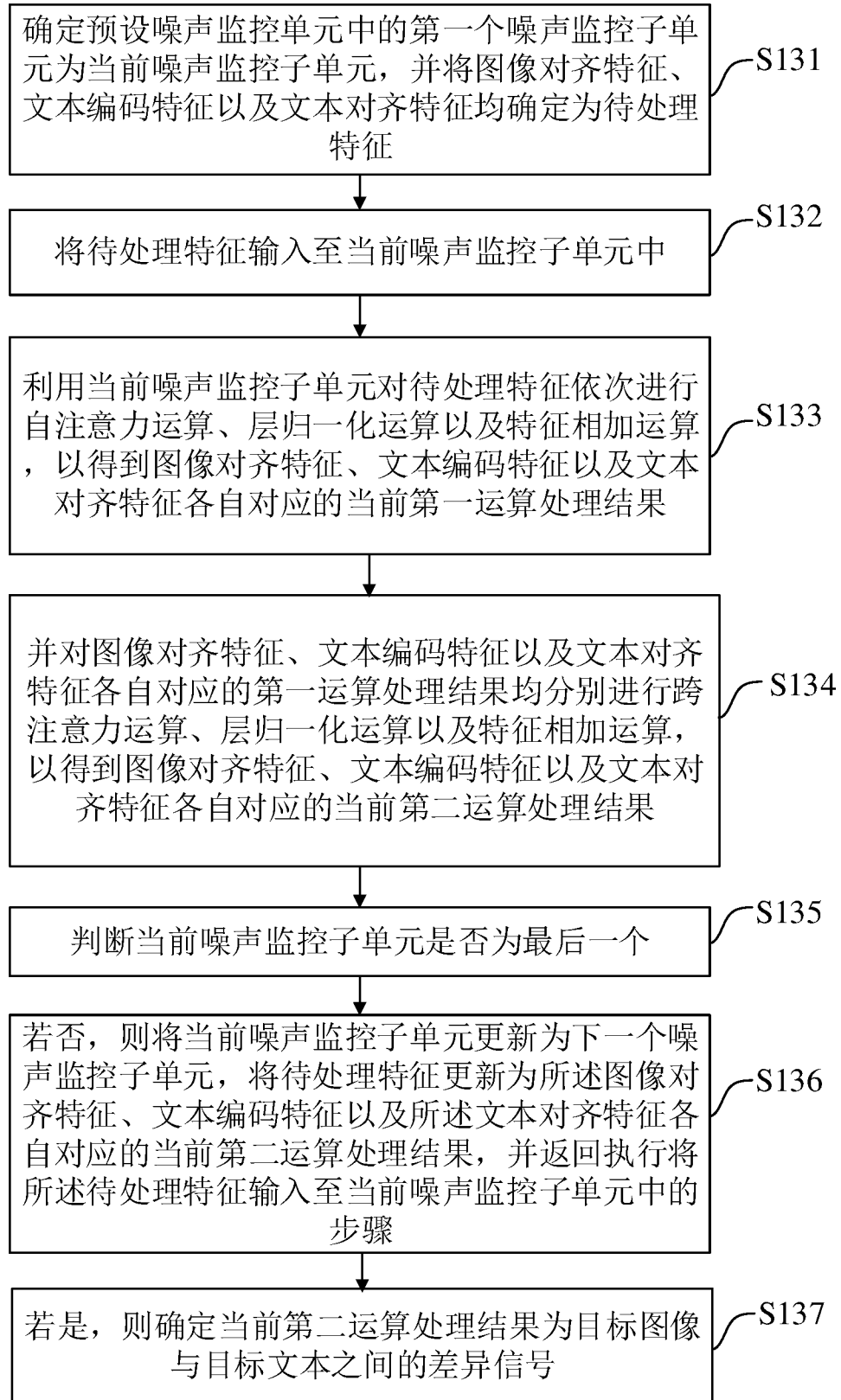


图 2

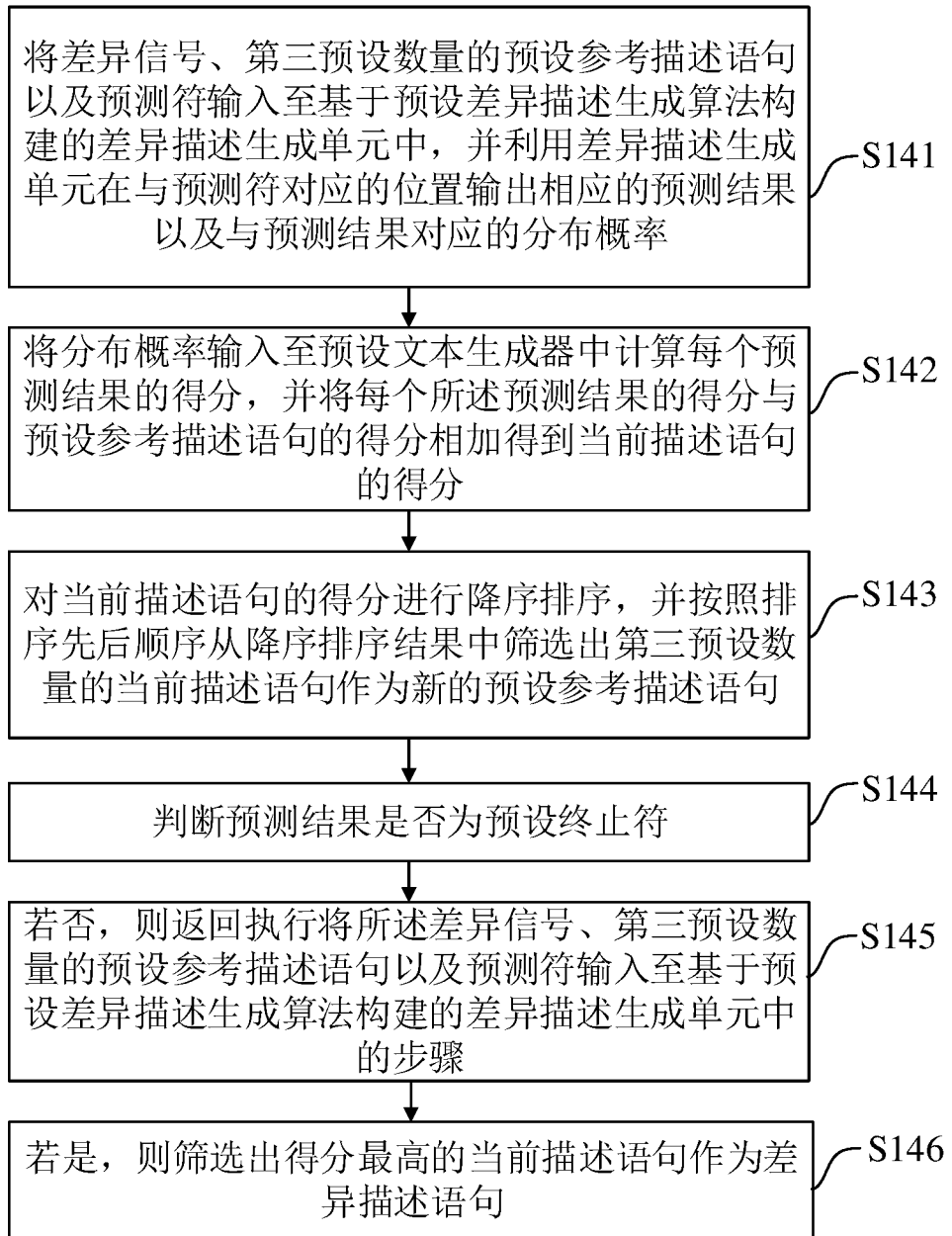


图 3



图 4



图 5

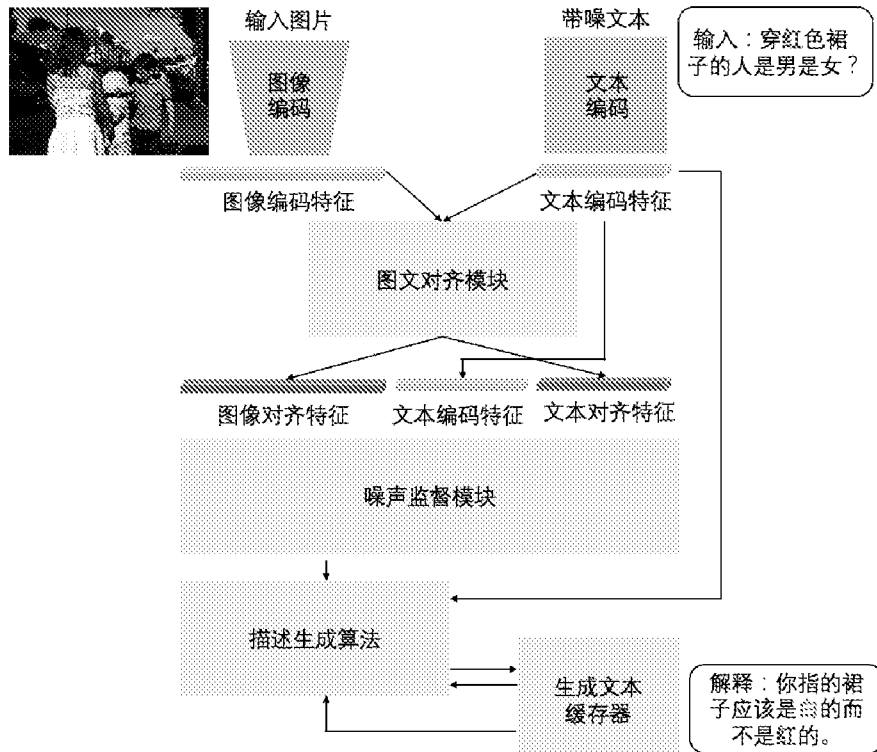


图 6

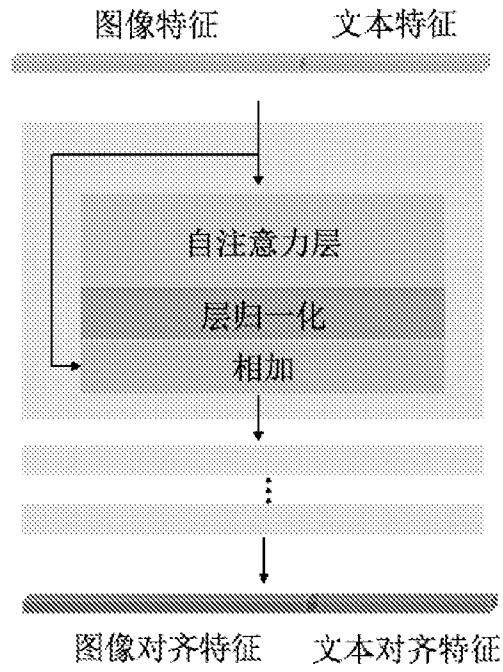


图 7

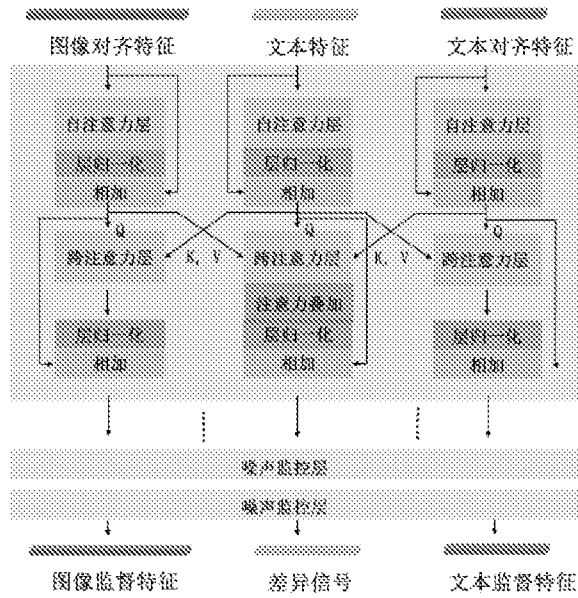


图 8



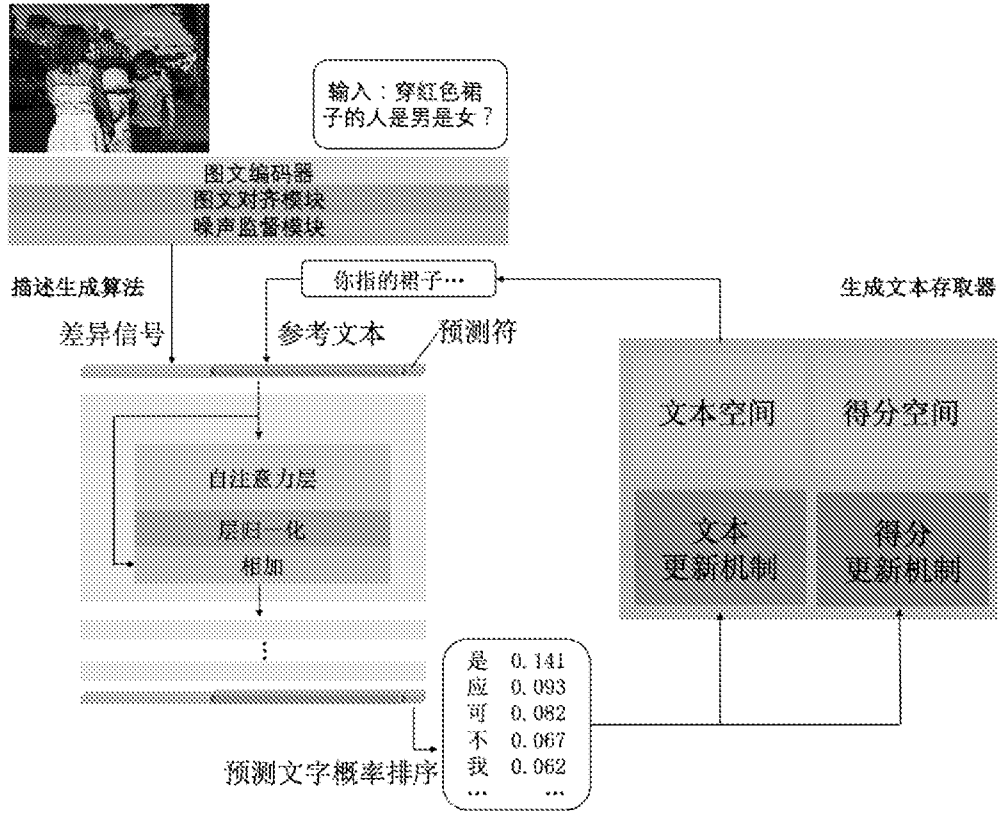


图 9

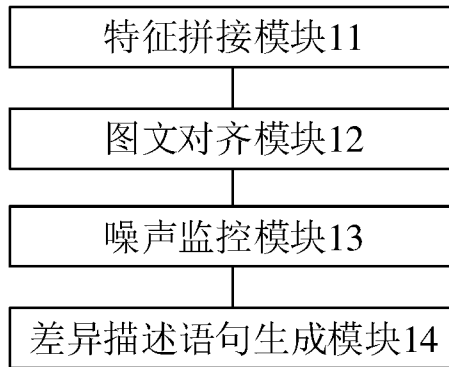


图 10

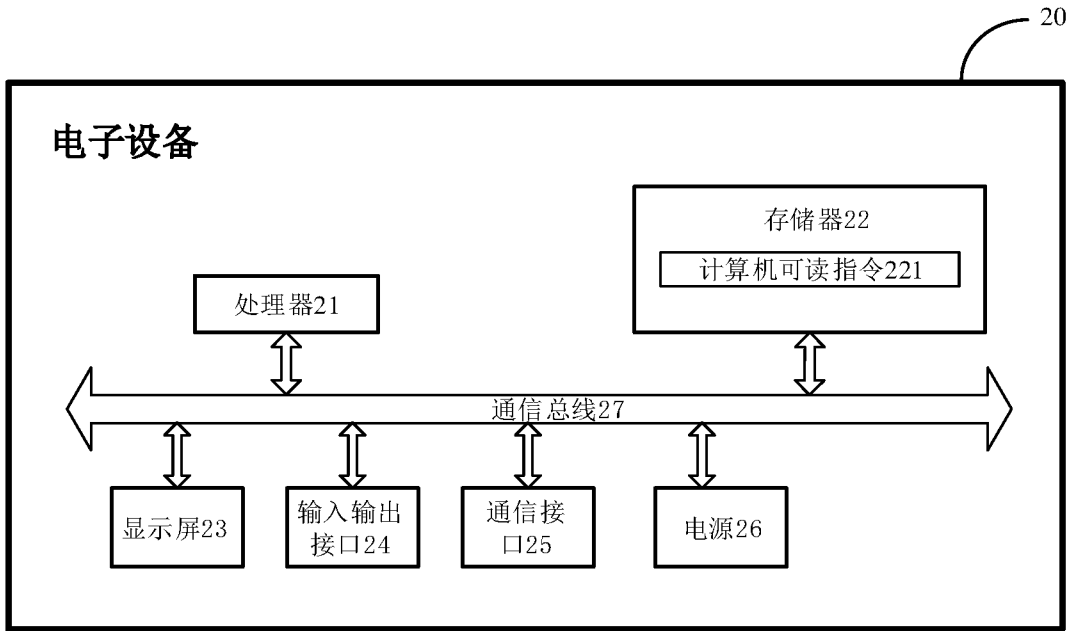


图 11

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/118852

<b>A. CLASSIFICATION OF SUBJECT MATTER</b>		
G06V 30/40(2022.01)i; G06V 30/418(2022.01)i; G06V 30/18(2022.01)i		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b>		
Minimum documentation searched (classification system followed by classification symbols) 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; VEN; DWPI; USTXT; WOTXT; EPTXT: 图像, 文本, 差异, 描述, 图文, 匹配, 编码, 自注意力, 信号, image, text, difference, description, teletext, matching, encode, attention, signal		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 114511860 A (SUZHOU INSPUR INTELLIGENT TECHNOLOGY CO., LTD.) 17 May 2022 (2022-05-17) claims 1-11, and description, paragraphs [0019]-[0074]	1-20
A	CN 114329034 A (WUHAN UNIVERSITY) 12 April 2022 (2022-04-12) description, paragraphs [0047]-[0090]	1-20
A	CN 113159095 A (HUAWEI TECHNOLOGIES CO., LTD.) 23 July 2021 (2021-07-23) entire document	1-20
A	US 2022092683 A1 (IBM) 24 March 2022 (2022-03-24) entire document	1-20
<input 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 <b>17 November 2022</b>		Date of mailing of the international search report <b>06 January 2023</b>
Name and mailing address of the ISA/CN <b>China National Intellectual Property Administration (ISA/CN) No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088, China</b> Facsimile No. (86-10)62019451		Authorized officer  Telephone No.

**INTERNATIONAL SEARCH REPORT**  
**Information on patent family members**

International application No. <b>PCT/CN2022/118852</b>
---

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	114511860	A	17 May 2022	CN	114511860	B	17 June 2022
CN	114329034	A	12 April 2022	None			
CN	113159095	A	23 July 2021	WO	2022161380	A1	04 August 2022
US	2022092683	A1	24 March 2022	None			

国际检索报告

国际申请号

PCT/CN2022/118852

<p><b>A. 主题的分类</b></p> <p>G06V 30/40(2022.01)i; G06V 30/418(2022.01)i; G06V 30/18(2022.01)i</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>B. 检索领域</b></p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>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;VEN;DWPI;USTXT;WOTXT;EPTXT; 图像, 文本, 差异, 描述, 图文, 匹配, 编码, 自注意力, 信号, image, text, difference, description, teletext, matching, encode, attention, signal</p>																	
<p><b>C. 相关文件</b></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>PX</td> <td>CN 114511860 A (苏州浪潮智能科技有限公司) 2022年5月17日 (2022 - 05 - 17) 权利要求第1-11项, 说明书第[0019]-[0074]段</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 114329034 A (武汉大学) 2022年4月12日 (2022 - 04 - 12) 说明书第[0047]-[0090]段</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 113159095 A (华为技术有限公司) 2021年7月23日 (2021 - 07 - 23) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>US 2022092683 A1 (IBM) 2022年3月24日 (2022 - 03 - 24) 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	PX	CN 114511860 A (苏州浪潮智能科技有限公司) 2022年5月17日 (2022 - 05 - 17) 权利要求第1-11项, 说明书第[0019]-[0074]段	1-20	A	CN 114329034 A (武汉大学) 2022年4月12日 (2022 - 04 - 12) 说明书第[0047]-[0090]段	1-20	A	CN 113159095 A (华为技术有限公司) 2021年7月23日 (2021 - 07 - 23) 全文	1-20	A	US 2022092683 A1 (IBM) 2022年3月24日 (2022 - 03 - 24) 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															
PX	CN 114511860 A (苏州浪潮智能科技有限公司) 2022年5月17日 (2022 - 05 - 17) 权利要求第1-11项, 说明书第[0019]-[0074]段	1-20															
A	CN 114329034 A (武汉大学) 2022年4月12日 (2022 - 04 - 12) 说明书第[0047]-[0090]段	1-20															
A	CN 113159095 A (华为技术有限公司) 2021年7月23日 (2021 - 07 - 23) 全文	1-20															
A	US 2022092683 A1 (IBM) 2022年3月24日 (2022 - 03 - 24) 全文	1-20															
<p><input type="checkbox"/> 其余文件在C栏的续页中列出。</p> <p><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>“&amp;” 同族专利的文件</p>																	
<p>国际检索实际完成的日期</p> <p>2022年11月17日</p>		<p>国际检索报告邮寄日期</p> <p>2023年1月6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>李娇</p> <p>电话号码 (86-512) 88995726</p>															

国际检索报告  
关于同族专利的信息

国际申请号  
PCT/CN2022/118852

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	114511860	A	2022年5月17日	CN	114511860	B	2022年6月17日
CN	114329034	A	2022年4月12日	无			
CN	113159095	A	2021年7月23日	WO	2022161380	A1	2022年8月4日
US	2022092683	A1	2022年3月24日	无			