



(12) 发明专利申请

(10) 申请公布号 CN 114677402 A

(43) 申请公布日 2022.06.28

(21) 申请号 202210360153.1

G06N 3/08 (2006.01)

(22) 申请日 2022.04.07

(71) 申请人 北京智源人工智能研究院

地址 100083 北京市海淀区中关村东路1号
院8号楼三层B201D-1

申请人 中国人民大学

(72) 发明人 金楚浩 宋睿华 许洪腾 卢志武
曹岗 文继荣

(74) 专利代理机构 北京辰权知识产权代理有限公司 11619

专利代理师 李小朋

(51) Int. Cl.

G06T 7/187 (2017.01)

G06T 11/60 (2006.01)

G06N 3/04 (2006.01)

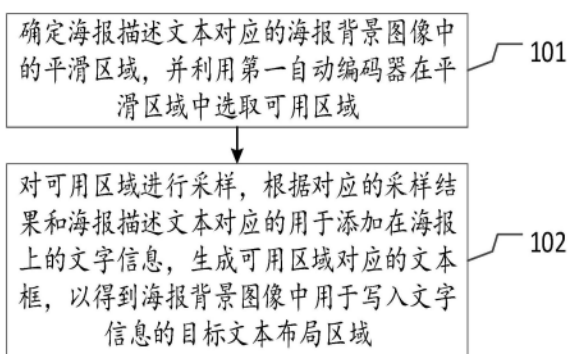
权利要求书2页 说明书19页 附图5页

(54) 发明名称

海报文本布局、海报生成方法及相关装置

(57) 摘要

本申请公开了一种海报文本布局、海报生成方法及相关装置。通过应用本申请的技术方案，可以从海报背景图像中的平滑区域中利用一个级联的自动编码器选取其中的可用区域，并在后续依据另一个级联自动编码器对该可用区域进行迭代细化后自动确定其中的文本框并将文字信息填入其中，从而智能的生成海报图像。进而避免了相关技术中，仅依靠人工设计生成海报所面临的无法满足高质量海报的大量需求的问题。



1. 一种海报文本布局方法,其特征在于,包括:

确定海报描述文本对应的海报背景图像中的平滑区域,并利用第一自动编码器在所述平滑区域中选取可用区域;

对所述可用区域进行采样,根据对应的采样结果和所述海报描述文本对应的用于添加在海报上的文字信息,生成所述可用区域对应的文本框,以得到所述海报背景图像中用于写入所述文字信息的目标文本布局区域。

2. 如权利要求1所述的海报文本布局方法,其特征在于,所述确定海报描述文本对应的海报背景图像中的平滑区域,包括:

在所述海报背景图像中选取多个候选框;以及,利用光谱残差算法,生成所述海报背景图像对应的显著图像,其中,各个所述候选框中存在相互重叠的候选框;

基于所述显著图像的显著平均值、所述候选框的像素数以及预设的偏移量,为每个候选框确定与其对应的候选值;

选取所述候选值低于预设候选阈值的候选框作为目标候选框,并基于所述目标候选框,确定所述海报背景图像中的平滑区域,其中每个目标候选框在所述海报背景图像中的位置各不相同。

3. 如权利要求2所述的海报文本布局方法,其特征在于,所述选取所述候选值低于预设候选阈值的候选框作为目标候选框,并基于所述目标候选框,确定所述海报背景图像中的平滑区域,包括:

基于非极大值抑制方法选取所述候选值低于预设候选阈值的候选框作为目标候选框;

确定所有目标候选框所在所述海报背景图像中的候选框区域;

将所述海报背景图像中的所述候选框区域转换为二值图,并将转换后的二值图区域作为所述海报背景图像中的平滑区域。

4. 如权利要求1所述的海报文本布局方法,其特征在于,所述利用第一自动编码器在所述平滑区域中选取可用区域,包括:

将所述平滑区域对应的各个目标候选框分别作为所述第一自动编码器的编码器端输入,得到编码器输出结果,其中所述编码器端由堆叠的CNN所构建得到;

将所述编码器输出结果与位置嵌入图进行连接后,作为所述第一自动编码器的解码器端输入,得到所述平滑区域中作为初始文本布局结果的可用区域,其中所述解码器端由堆叠的Transposed-CNN所构建得到。

5. 如权利要求1所述的海报文本布局方法,其特征在于,在所述生成所述可用区域对应的文本框之后,还包括:

利用第二自动编码器,以自回归方式对各个所述可用区域各自对应的文本框分别进行细化布局处理,得到各个所述可用区域各自对应的目标文本框,以形成所述海报背景图像中用于写入所述文字信息的目标文本布局区域。

6. 如权利要求5所述的海报文本布局方法,其特征在于,

所述第一自动编码器和所述第二自动编码器均为具有级联自动编码架构的布局预测器。

7. 如权利要求6所述的海报文本布局方法,其特征在于,还包括:

获取多个样本图像,其中每个样本图像中包含标注有对应文本描述字段的文本区域;

提取每个样本图像中的样本背景图像,并利用平滑区域检测器确定所述样本背景图像中的样本平滑图像区域;

将所述文本描述字段、所述文本区域、所述样本背景图像以及所述样本平滑图像区域合并为编码器训练数据集;

利用所述编码器训练数据集分别独立训练所述第一自动编码器以及所述第二自动编码器,以使所述第一自动编码器用于预测文本布局的概率分布以及使所述第二自动编码器用于细化布局本框。

8.如权利要求5至7任一项所述的海报文本布局方法,其特征在于,所述第二自动编码器还包括:

利用自监督学习策略训练得到,其中所述第二自动编码器的编码器端由堆叠的CNN所构建得到,所述第二自动编码器的解码器端由2层的双向LSTM所构建得到。

9.一种海报生成方法,其特征在于,包括:

基于权利要求1至8任一项所述的海报文本布局方法获取目标文本布局区域;

将所述文字信息填充在所述文本布局区域中以生成所述文字信息对应的目标海报。

10.如权利要求9所述的海报生成方法,其特征在于,在所述获取目标文本布局区域之前,还包括:

获取海报描述文本,所述海报描述文本中包含有用于添加在海报上的文字信息;

基于预训练的视觉文本模型,在预获取的候选背景图像集合中选取与所述文字信息弱相关匹配的候选背景图像作为海报背景图像。

11.如权利要求9所述的海报生成方法,其特征在于,所述将所述文字信息填充在所述文本布局区域中以生成所述文字信息对应的海报,包括:

提取所述文字信息的文字特征,以及检测所述文本布局区域的背景颜色;

基于所述文本布局区域的背景颜色,从预设的文本样式数据库中选取与所述文字特征以及所述背景颜色相匹配的目标文本样式;

将所述文字信息按照所述目标文本样式进行转换处理后填充在所述文本布局区域中,得到所述文字信息对应的目标海报。

12.一种海报文本布局装置,其特征在于,包括:

确定模块,确定海报描述文本对应的海报背景图像中的平滑区域,并利用第一自动编码器在所述平滑区域中选取可用区域;

生成模块,对所述可用区域进行采样,根据对应的采样结果和所述海报描述文本对应的用于添加在海报上的文字信息,生成所述可用区域对应的文本框,以得到所述海报背景图像中用于写入所述文字信息的目标文本布局区域。

13.一种电子设备,其特征在于,包括:

存储器,用于存储可执行指令;以及,

处理器,用于与所述存储器以执行所述可执行指令从而完成权利要求1-8中任一所述海报文本布局方法的操作,或者完成权利要求9-11中任一所述海报生成方法的操作。

14.一种计算机可读存储介质,用于存储计算机可读的指令,其特征在于,所述指令被执行时执行权利要求1-8中任一所述海报文本布局方法的操作,或者完成权利要求9-11中任一所述海报生成方法的操作。

海报文本布局、海报生成方法及相关装置

技术领域

[0001] 本申请中涉及图像数据处理技术,尤其是一种海报文本布局、海报生成方法及相关装置。

背景技术

[0002] 海报作为一种兼具艺术性和功能性的媒介,已广泛应用于许多商业和非商业场景中,用于宣传和传播信息。例如,电子商务平台使用有吸引力的海报来宣传其自身的商品。

[0003] 相关技术中,现有的海报生成方式通常是通过专业的美工将各个商品图像及信息经过处理后排版形成。其具体过程包括人工选取海报的背景图像、人工在海报中进行文本布局以及人工确定海报上文本样式等过程。

[0004] 然而,这样一个耗时且主观的人工设计生成海报的过程无法满足现实应用中对高质量海报的大量且快速增长的需求,从而降低了信息传播的效率。因此,如何设计一种可以利用预训练模型自动实现上述海报生成过程的技术方案,成为了需要解决的问题。

发明内容

[0005] 本申请实施例提供一种海报文本布局、海报生成方法及相关装置。用以解决相关技术中存在的,只能依靠人工生成海报所导致的降低信息传播效率的问题。

[0006] 其中,根据本申请实施例的一个方面,提供一种海报文本布局方法,包括:

[0007] 确定海报描述文本对应的海报背景图像中的平滑区域,并利用第一自动编码器在所述平滑区域中选取可用区域;

[0008] 对所述可用区域进行采样,根据对应的采样结果和所述海报描述文本对应的用于添加在海报上的文字信息,生成所述可用区域对应的文本框,以得到所述海报背景图像中用于写入所述文字信息的目标文本布局区域。

[0009] 可选地,在基于本申请上述方法的另一个实施例中,所述确定海报描述文本对应的海报背景图像中的平滑区域,包括:

[0010] 在所述海报背景图像中选取多个候选框;以及,利用光谱残差算法,生成所述海报背景图像对应的显著图像,其中,各个所述候选框中存在相互重叠的候选框;

[0011] 基于所述显著图像的显著平均值、所述候选框的像素数以及预设的偏移量,为每个候选框确定与其对应的候选值;

[0012] 选取所述候选值低于预设候选阈值的候选框作为目标候选框,并基于所述目标候选框,确定所述海报背景图像中的平滑区域,其中每个目标候选框在所述海报背景图像中的位置各不重叠。

[0013] 可选地,在基于本申请上述方法的另一个实施例中,所述选取所述候选值低于预设候选阈值的候选框作为目标候选框,并基于所述目标候选框,确定所述海报背景图像中的平滑区域,包括:

[0014] 基于非极大值抑制方法选取所述候选值低于预设候选阈值的候选框作为目标候

选框；

[0015] 确定所有目标候选框所在所述海报背景图像中的候选框区域；

[0016] 将所述海报背景图像中的所述候选框区域转换为二值图，并将转换后的二值图区域作为所述海报背景图像中的平滑区域。

[0017] 可选地，在基于本申请上述方法的另一个实施例中，所述利用第一自动编码器在所述平滑区域中选取可用区域，包括：

[0018] 将所述平滑区域对应的各个目标候选框分别作为所述第一自动编码器的编码器端输入，得到编码器输出结果，其中所述编码器端由堆叠的CNN所构建得到；

[0019] 将所述编码器输出结果与位置嵌入图进行连接后，作为所述第一自动编码器的解码器端输入，得到所述平滑区域中作为初始文本布局结果的可用区域，其中所述解码器端由堆叠的Transposed-CNN所构建得到。

[0020] 可选地，在基于本申请上述方法的另一个实施例中，在所述生成所述可用区域对应的文本框之后，还包括：

[0021] 利用第二自动编码器，以自回归方式对各个所述可用区域各自对应的文本框分别进行细化布局处理，得到各个所述可用区域各自对应的目标文本框，以形成所述海报背景图像中用于写入所述文字信息的目标文本布局区域。

[0022] 可选地，在基于本申请上述方法的另一个实施例中，所述第一自动编码器和所述第二自动编码器均为具有级联自动编码架构的布局预测器。

[0023] 可选地，在基于本申请上述方法的另一个实施例中，获取多个样本图像，其中每个样本图像中包含标注有对应文本描述字段的文本区域；

[0024] 提取每个样本图像中的样本背景图像，并利用平滑区域检测器确定所述样本背景图像中的样本平滑图像区域；

[0025] 将所述文本描述字段、所述文本区域、所述样本背景图像以及所述样本平滑图像区域合并为编码器训练数据集；

[0026] 利用所述编码器训练数据集分别独立训练所述第一自动编码器以及所述第二自动编码器，以使所述第一自动编码器用于预测文本布局的概率分布以及使所述第二自动编码器用于细化布局本框。

[0027] 可选地，在基于本申请上述方法的另一个实施例中，所述第二自动编码器

[0028] 利用自监督学习策略训练得到，其中所述第二自动编码器的编码器端由堆叠的CNN所构建得到，所述第二自动编码器的解码器端由2层的双向LSTM所构建得到。

[0029] 其中，根据本申请实施例的一个方面，提供一种海报生成方法，包括：

[0030] 基于上述的海报文本布局方法获取目标文本布局区域；

[0031] 将所述文字信息填充在所述文本布局区域中以生成所述文字信息对应的目标海报。

[0032] 可选地，在基于本申请上述方法的另一个实施例中，在所述获取目标文本布局区域之前，还包括：

[0033] 获取海报描述文本，所述海报描述文本中包含有用于添加在海报上的文字信息；

[0034] 基于预训练的视觉文本模型，在预获取的候选背景图像集合中选取与所述文字信息弱相关匹配的候选背景图像作为海报背景图像。

[0035] 可选地,在基于本申请上述方法的另一个实施例中,所述将所述文字信息填充在所述文本布局区域中以生成所述文字信息对应的海报,包括:

[0036] 提取所述文字信息的文字特征,以及检测所述文本布局区域的背景颜色;

[0037] 基于所述文本布局区域的背景颜色,从预设的文本样式数据库中选取与所述文字特征以及所述背景颜色相匹配的目标文本样式;

[0038] 将所述文字信息按照所述目标文本样式进行转换处理后填充在所述文本布局区域中,得到所述文字信息对应的目标海报。

[0039] 其中,根据本申请实施例的又一个方面,提供一种海报的生成装置,其特征在于,包括:

[0040] 确定模块,确定海报描述文本对应的海报背景图像中的平滑区域,并利用第一自动编码器在所述平滑区域中选取可用区域;

[0041] 生成模块,对所述可用区域进行采样,根据对应的采样结果和所述海报描述文本对应的用于添加在海报上的文字信息,生成所述可用区域对应的文本框,以得到所述海报背景图像中用于写入所述文字信息的目标文本布局区域。

[0042] 根据本申请实施例的又一个方面,提供一种电子设备,包括:

[0043] 存储器,用于存储可执行指令;以及

[0044] 显示器,用于与所述存储器以执行所述可执行指令从而完成上述任一所述海报的生成方法的操作。

[0045] 根据本申请实施例的还一个方面,提供一种计算机可读存储介质,用于存储计算机可读的指令,所述指令被执行时执行上述任一所述海报的生成方法的操作。

[0046] 本申请中,可以确定海报描述文本对应的海报背景图像中的平滑区域,并利用第一自动编码器在平滑区域中选取可用区域;对可用区域进行采样,根据对应的采样结果和海报描述文本对应的用于添加在海报上的文字信息,生成可用区域对应的文本框,以得到海报背景图像中用于写入文字信息的目标文本布局区域。通过应用本申请的技术方案,可以从海报背景图像中的平滑区域中利用一个级联的自动编码器选取其中的可用区域,并在后续依据另一个级联自动编码器对该可用区域进行迭代细化后自动确定其中的文本框并将文字信息填入其中,从而智能的生成海报图像。进而避免了相关技术中,仅依靠人工设计生成海报所面临的无法满足高质量海报的大量需求的问题。

[0047] 下面通过附图和实施例,对本申请的技术方案做进一步的详细描述。

附图说明

[0048] 构成说明书的一部分的附图描述了本申请的实施例,并且连同描述一起用于解释本申请的原理。

[0049] 参照附图,根据下面的详细描述,可以更加清楚地理解本申请,其中:

[0050] 图1为本申请提出的一种海报文本布局的方法示意图;

[0051] 图2为本申请提出的一种海报文本布局的流程示意图;

[0052] 图3为本申请提出的一种海报的生成的流程示意图;

[0053] 图4为本申请提出的一种海报背景图像选取的实例参照图;

[0054] 图5为本申请提出的一种文本布局的方法与现有技术流程的文本布局方法的对比

示意图；

[0055] 图6为本申请提出的一种海报背景图像选取的流程示意图；

[0056] 图7为本申请提出的一种海报生成方法的整体流程示意图；

[0057] 图8为本申请提出的一种海报文本布局的电子装置的结构示意图；

[0058] 图9为本申请提出的一种海报文本布局的电子设备的结构示意图。

具体实施方式

[0059] 现在将参照附图来详细描述本申请的各种示例性实施例。应注意到：除非另外具体说明，否则在这些实施例中阐述的部件和步骤的相对布置、数字表达式和数值不限制本申请的范围。

[0060] 同时，应当明白，为了便于描述，附图中所示出的各个部分的尺寸并不是按照实际的比例关系绘制的。

[0061] 以下对至少一个示例性实施例的描述实际上仅仅是说明性的，不作为对本申请及其应用或使用的任何限制。

[0062] 对于相关领域普通技术人员已知的技术、方法和设备可能不作详细讨论，但在适当情况下，所述技术、方法和设备应当被视为说明书的一部分。

[0063] 应注意到：相似的标号和字母在下面的附图中表示类似项，因此，一旦某一项在一个附图中被定义，则在随后的附图中不需要对其进行进一步讨论。

[0064] 另外，本申请各个实施例之间的技术方案可以相互结合，但是必须是以本领域普通技术人员能够实现为基础，当技术方案的结合出现相互矛盾或无法实现时应当认为这种技术方案的结合不存在，也不在本申请要求的保护范围之内。

[0065] 需要说明的是，本申请实施例中所有方向性指示（诸如上、下、左、右、前、后……）仅用于解释在某一特定姿态（如附图所示）下各部件之间的相对位置关系、运动情况等，如果该特定姿态发生改变时，则该方向性指示也相应地随之改变。

[0066] 下面结合图1-图7来描述根据本申请示例性实施方式的用于进行海报的生成方法。需要注意的是，下述应用场景仅是为了便于理解本申请的精神和原理而示出，本申请的实施方式在此方面不受任何限制。相反，本申请的实施方式可以应用于适用的任何场景。

[0067] 本申请还提出一种海报文本布局、海报生成方法及相关装置。

[0068] 图1示意性地示出了根据本申请实施方式的一种海报的生成方法的流程示意图。如图1所示，该方法包括：

[0069] S101，确定海报描述文本对应的海报背景图像中的平滑区域，并利用第一自动编码器在平滑区域中选取可用区域。

[0070] 相关技术中，海报作为一种兼具艺术性和功能性的媒介，已广泛应用于许多商业和非商业场景中，用于宣传和传播信息。例如，电子商务平台使用有吸引力的海报来宣传他们的商品。诸如会议等的社交活动的网站通常装饰有精美且内容丰富的海报。这些高质量的海报是通过将风格化的文本嵌入到合适的背景图像中生成的，该生成过程需要大量的人工编辑工作和难以量化的艺术设计经验。然而，这样一个耗时且主观的人工设计过程无法满足现实应用中对高质量海报的大量且快速增长的需求，从而降低了信息传播的效率，导致宣传效果欠佳。

[0071] 一种方式中,海报生成的过程至少包含有三个步骤,其中包括:

[0072] 步骤1:选取海报的背景图像;

[0073] 步骤2:在海报背景图像中进行文本布局;

[0074] 步骤3:确定海报上文本的样式并将其填入到背景图像中,从而生成最终的海报图像。

[0075] 基于上述过程,其中针对步骤1,为了实现海报的背景图像的自动搜索,现有的经典检索方法一般通过将海报内容文本与背景图像标注文本匹配来搜索合适的图像。

[0076] 然而,大相关技术中的方法通常只考虑单模态特征,因此通常无法弥合视觉和文本模态之间的语义鸿沟,并会产生偏差较大的检索结果。也即背景图像的选取与用户本次感兴趣的方向并不匹配。

[0077] 另外,针对步骤2,相关技术中为了进行文本布局预测,常规的基于规则的方法通常从有限数量的预定义布局模板中选择文本的布局,但是其灵活性较差。从而导致在布置文本时均忽略了背景图像的内容,因此可能导致海报生成的效果并不理想。

[0078] 进一步的,为了解决上述存在的在海报背景图像中文本框选取不匹配的问题,本申请提出一种海报文本布局的技术方案。

[0079] 概括来说,本申请可以首先确定海报描述文本对应的海报背景图像中的平滑区域,并利用第一自动编码器在该平滑区域中其中的选取可用区域(也即可用区域为平滑区域中的部分区域)。另外,本申请实施例还可以对该可用区域进行采样,以使后续根据对应的采样结果和海报描述文本对应的用于添加在海报上的文字信息,生成可用区域对应的文本框(也即文本框是在可用区域中得到的),从而最后得到海报背景图像中用于写入所述文字信息的目标文本布局区域。

[0080] 一种方式中,本申请还可以首先采用大规模预训练的视觉文本模型,并根据给定的目标海报的描述文本在众多候选图像中选择作为目标海报背景的图像,然后通过多个级联自动编码器在该背景图像上迭代布置文本,最后通过基于匹配的方法对文本进行风格化处理以合成目标海报图像。

[0081] 一种方式中,本申请在确定海报描述文本对应的海报背景图像中的平滑区域的过程中,可以通过如下步骤实现:

[0082] 本申请需要首先在背景图像I中生成若干个不同大小的区域(候选框),这些区域中存在重叠,记为 $\{A_i\}_{i=1}^{N_A}$ 。然后应用光谱残差方法生成背景图像I对应的显著图(saliency map),表示为S。

[0083] 其中,对于每个候选框,通过计算显著图S的平均值来为每个候选框分别分配一个

候选值 v_i 。即
$$v_i = \frac{1}{|A_i|} (\lambda + \sum_{p \in A_i} S(p))$$

[0084] 其中,S(p)是指文本布局预测结果p处的显著性, $|A_i|$ 是 A_i 中的像素数, λ 是对候选框尺寸敏感的偏移量。

[0085] 其中,在一种优选方案中,本申请可以选取候选值低于预设候选阈值的候选框作为目标候选框。并确定目标候选框所在海报背景图像中的候选框区域,再将海报背景图像中的候选框区域转换为二值图。最后,将转换后的二值图区域作为海报背景图像中的平滑

图像区域。

[0086] 例如,如图2所示,图像(a)的较中心部分为背景图像I的显著图S,在实际应用中可以采用第一颜色(例如为蓝色)来显示,图像(a)中的较边缘矩形区域部分即为通过本步骤识别得到的平滑区域图A,实际应用中可以采用明显区别于第一颜色的第二颜色(例如为红色)来显示。

[0087] 进一步的,本申请需要利用第一自动编码器在平滑图像区域中确定至少一个可用区域。

[0088] 一种方式中,本申请可以基于上述步骤得到的平滑区域A,采用第一级联自动编码器 g_1 来预测用来进行文本布局的概率分布,表示为 L 。对于每个像素 p , $L(p) \in [0, 1]$ 表示 p 属于一个文本框的概率。

[0089] 其中,本申请所采用的第一级联自动编码器 g_1 拥有一个自动编码架构,其编码器 f_1 是堆叠的CNN,解码器 h_1 是堆叠的Transposed-CNN。而后通过将编码器 f_1 的输出与可学习的位置嵌入图(表示为 E)连接起来,来构造解码器 h_1 的输入。

[0090] 例如图2中的图像(b)中的偏白部分显示了海报背景图像I中可以进行文本布局的概率分布预测结果 L ,其中颜色越亮则布局概率越高。

[0091] S102,对可用区域进行采样,根据对应的采样结果和海报描述文本对应的用于添加在海报上的文字信息,生成可用区域对应的文本框,以得到海报背景图像中用于写入文字信息的目标文本布局区域。

[0092] 一种方式中,本申请可以利用第二自动编码器,以自回归方式对至少一个可用区域进行迭代细化布局,得到可用区域对应的文本框。

[0093] 其中,在上述步骤中得到的布局预测概率分布结果 L 即为可以进行文本布局的初始区域筛选,还需要对其进行细化以将各个可用区域形成明确的文本框。

[0094] 即,首先以 L 为基础,通过从 L 中采样各个来初始化布局。将未归一化的作为第 i 个文本框的左上角坐标并初始化该框,其大小由对应的文本信息的长度及其属性来决定,最终在背景图像I中确定用于写入各个文本信息的各个文本框的位置及大小,即文本布局预测结果。

[0095] 在一种具体实现方式中,本申请可以利用第二级联赛自动编码器对可用区域进行文本布局细化处理以得到最终的文本布局区域,其中包括:

[0096] $P^{(k+1)} = g_2(\text{Concat}(A, L), P^{(k)}), k=0, \dots, K-1$ 。

[0097] 其中, k 为迭代次数, g_2 为第二自动编码器, A 为平滑图像区域, L 为可用区域的概率分布, P 为海报背景图像中的像素点。

[0098] 需要说明的是,对于第二级联赛自动编码器,它的编码器可以是堆叠的CNN,它的解码器可以是一个2层的双向LSTM。

[0099] 在一种优选方式中,上述步骤中用于可用区域布局预测的第一级联自动编码器和上述步骤中用于文本布局细化的第二级级联自动编码器均可以采用具有级联自动编码架构的布局预测器,从而实现可以有效模仿手动图像编辑过程的目的。

[0100] 本申请中,可以确定海报描述文本对应的海报背景图像中的平滑区域,并利用第一自动编码器在平滑区域中选取可用区域;对可用区域进行采样,根据对应的采样结果和海报描述文本对应的用于添加在海报上的文字信息,生成可用区域对应的文本框,以得到

海报背景图像中用于写入文字信息的目标文本布局区域。通过应用本申请的技术方案,可以从海报背景图像中的平滑区域中利用一个级联的自动编码器选取其中的可用区域,并在后续依据另一个级联自动编码器对该可用区域进行迭代细化后自动确定其中的文本框并将文字信息填入其中,从而智能的生成海报图像。进而避免了相关技术中,仅依靠人工设计生成海报所面临的无法满足高质量海报的大量需求的问题。

[0101] 可选地,在基于本申请上述方法的另一个实施例中,确定海报描述文本对应的海报背景图像中的平滑区域,包括:

[0102] 在海报背景图像中选取多个候选框;以及,利用光谱残差算法,生成海报背景图像对应的显著图像,其中,各个候选框中存在相互重叠的候选框;

[0103] 基于显著图像的显著平均值、候选框的像素数以及预设的偏移量,为每个候选框确定与其对应的候选值;

[0104] 选取候选值低于预设候选阈值的候选框作为目标候选框,并基于目标候选框,确定海报背景图像中的平滑区域,其中每个目标候选框在海报背景图像中的位置各不重叠。

[0105] 可选地,在基于本申请上述方法的另一个实施例中,选取候选值低于预设候选阈值的候选框作为目标候选框,并基于目标候选框,确定海报背景图像中的平滑区域,包括:

[0106] 基于非极大值抑制方法选取候选值低于预设候选阈值的候选框作为目标候选框;

[0107] 确定所有目标候选框所在海报背景图像中的候选框区域;

[0108] 将海报背景图像中的候选框区域转换为二值图,并将转换后的二值图区域作为海报背景图像中的平滑区域。

[0109] 一种方式中,本申请在确定海报描述文本对应的海报背景图像中的平滑区域的过程中,可以通过如下步骤实现:

[0110] 本申请需要首先在背景图像I中生成若干个不同大小的区域(候选框),这些区域中存在重叠,记为 $\{A_i\}_{i=1}^{N_A}$ 。然后应用光谱残差方法生成背景图像I对应的显著图(saliency map),表示为S。

[0111] 其中,对于每个候选框,通过计算显著图S的平均值来为每个候选框分别分配一个

候选值 v_i 。即
$$v_i = \frac{1}{|A_i|} (\lambda + \sum_{p \in A_i} S(p))$$

[0112] 其中,S(p)是指文本布局预测结果p处的显著性, $|A_i|$ 是 A_i 中的像素数, λ 是对候选框尺寸敏感的偏移量。

[0113] 其中,在一种优选方案中,本申请可以选取候选值低于预设候选阈值的候选框作为目标候选框。并确定目标候选框所在海报背景图像中的候选框区域,再将海报背景图像中的候选框区域转换为二值图。最后,将转换后的二值图区域作为海报背景图像中的平滑图像区域。

[0114] 需要说明的是,预设候选阈值可以根据不同的图像进行适应性设置,即:可以根据当前的背景图像I的各个区域各自对应的值的平均值来设置阈值,例如可以将阈值自适应地设置为 $1.4 \times \text{mean}\{\}$,并应用非极大值抑制方法NMS来确保从前述生成的区域中最终选取的各个区域之间不重叠(例如,在前期生成了1000个候选框,这些框可能会有重叠。但需要确保最终选取的各个区域之间不重叠)。

[0115] 可选地,在基于本申请上述方法的另一个实施例中,利用第一自动编码器在平滑区域中选取可用区域,包括:

[0116] 将平滑区域对应的各个目标候选框分别作为第一自动编码器的编码器端输入,得到编码器输出结果,其中编码器端由堆叠的CNN所构建得到;

[0117] 将编码器输出结果与位置嵌入图进行连接后,作为第一自动编码器的解码器端输入,得到平滑区域中作为初始文本布局结果的可用区域,其中解码器端由堆叠的Transposed-CNN所构建得到。

[0118] 其中,本申请所采用的第一级联自动编码器 g_1 拥有一个自动编码架构,其编码器 f_1 是堆叠的CNN,解码器 h_1 是堆叠的Transposed-CNN。而后通过将编码器 f_1 的输出与可学习的位置嵌入图(表示为 E)连接起来,来构造解码器 h_1 的输入。

[0119] 具体的,本申请可以通过将编码器 f_1 的输出与可学习的位置嵌入图(表示为 E)连接起来,来构造解码器 h_1 的输入,因此得到背景图像 I 中可以进行文本布局的概率分布预测结果 L (即初始文本布局结果的可用区域)。其中,公式如下:

[0120] $L = g_1(A) = h_1(\text{Concat}(f_1(A), E))$ 。

[0121] 其中, A 代表平滑区域, E 代表位置嵌入图。

[0122] 可选地,在基于本申请上述方法的另一个实施例中,在生成可用区域对应的文本框之后,还包括:

[0123] 利用第二自动编码器,以自回归方式对各个可用区域各自对应的文本框分别进行细化布局处理,得到各个可用区域各自对应的目标文本框,以形成海报背景图像中用于写入文字信息的目标文本布局区域。

[0124] 利用第二自动编码器,以自回归方式对至少一个可用区域进行迭代细化布局,得到目标文本布局区域。

[0125] 一种方式中,本申请可以在上述步骤中得到的布局预测概率分布结果 L 即为可以进行文本布局的初始区域筛选,还需要对其进行细化以将各个可用区域形成明确的文本框。

[0126] 即,首先以 L 为基础,通过从 L 中采样各个来初始化布局。将未归一化的作为第 i 个文本框的左上角坐标并初始化该框,其大小由对应的文本信息的长度及其属性来决定,最终在背景图像 I 中确定用于写入各个文本信息的各个文本框的位置及大小,即文本布局预测结果。

[0127] 在一种具体实现方式中,本申请可以利用第二级联赛自动编码器对可用区域进行文本布局细化处理以得到最终的文本布局区域,其中包括:

[0128] $P^{(k+1)} = g_2(\text{Concat}(A, L), P^{(k)})$, $k=0, \dots, K-1$ 。

[0129] 其中, k 为迭代次数, g_2 为第二自动编码器, A 为平滑图像区域, L 为可用区域的概率分布, P 为海报背景图像中的像素点。

[0130] 需要说明的是,对于第二级联赛自动编码器,它的编码器可以是堆叠的CNN,它的解码器可以是一个2层的双向LSTM。

[0131] 可选地,在基于本申请上述方法的另一个实施例中,第一自动编码器和第二自动编码器均为具有级联自动编码架构的布局预测器。

[0132] 在一种优选方式中,本申请提及的用于可用区域布局预测的第一级联自动编码器

和本申请提及的用于文本布局细化的第二级联赛自动编码器均可以采用具有级联自动编码架构的布局预测器,从而实现可以有效模仿手动图像编辑过程的目的。

[0133] 可选地,在基于本申请上述方法的另一个实施例中,获取多个样本图像,其中每个样本图像中包含标注有对应文本描述字段的文本区域;

[0134] 提取每个样本图像中的样本背景图像,并利用平滑区域检测器确定样本背景图像中的样本平滑图像区域;

[0135] 将文本描述字段、文本区域、样本背景图像以及样本平滑图像区域合并为编码器训练数据集;

[0136] 利用编码器训练数据集分别独立训练第一自动编码器以及第二自动编码器,以使第一自动编码器用于预测文本布局的概率分布以及使第二自动编码器用于细化布局本框。

[0137] 其中,本申请实施例对于构建编码器训练数据集的过程来说,可以包括如下步骤:

[0138] 步骤1:一种方式中,可以从图片素材网站等处收集多种类别的海报图片,以形成样本图像集合。

[0139] 步骤2:对于样本图像集合中的每张样本图像,首先应用OCR工具来检测其文本 \hat{T}' 和相应的文本框 $\hat{P}=[\hat{p}]$,并根据这些文本框生成对应的二进制布局图,表示为 \hat{L} 。

[0140] 步骤3:通过二元布局图对其对应的海报图像文字区域进行遮挡,并通过图像修复方法来填充被遮挡的区域,以分别得到样本图像集合D中的各个样本图像各自对应的样本背景图像 \hat{I} 。

[0141] 步骤4:应用平滑区域检测器可以相应地获得各个样本背景图像 \hat{I} 分别对应的样本平滑区域图 \hat{A} 。

[0142] 综上,本申请提出的编码器训练数据集D可以表示为:

$$[0143] \quad D = \left\{ (\hat{T}'_n, \hat{P}_n, \hat{L}_n, \hat{A}_n) \right\}_{n=1}^N。$$

[0144] 其中, \hat{T}' 代表文本描述字段、 \hat{P} 代表文本区域、 \hat{I} 代表样本背景图像以及 \hat{A} 代表样本平滑图像区域。

[0145] 可选地,在基于本申请上述方法的另一个实施例中,第二自动编码器

[0146] 利用自监督学习策略训练得到,其中第二自动编码器的编码器端由堆叠的CNN所构建得到,第二自动编码器的解码器端由2层的双向LSTM所构建得到。

[0147] 进一步的,对于构建第一自动编码器 g_1 来说,可以基于下述公式得到:

$$[0148] \quad \min g_1 \frac{1}{N} \sum_{n=1}^N \left\| g_1(\hat{A}_n) - \hat{L}_n \right\|_2^2。$$

[0149] 更进一步的,对于构建第二自动编码器 g_2 来说,可以基于下述公式得到:

$$[0150] \quad \min g_2 \frac{1}{N} \sum_{n=1}^N \left\| g_2(\text{Concat}(\hat{A}_n, \hat{L}_n), P_n^{(0)}) - \hat{p}_n \right\|_F^2。$$

[0151] 其中, $|\hat{p}_n|$ 表示第n个海报图像的文本框数。

[0152] 需要说明的是,本申请在训练得到第二自动编码器的过程中,可以利用自监督学习策略来训练第二自动编码器。

[0153] 一种方式中,可以通过 $p_{i,n}^{(0)}$ 至 $Uniform(\hat{p}_{i,n} - \Delta, \hat{p}_{i,n} + \Delta), \forall \hat{p}_{i,n} \in \hat{P}_n$ 来对初始位置 $P_n^{(0)} = [p_{i,n}^{(0)}]$ 进行采样。其中,扰动 $\Delta = [0.1, 0.1]^T$ 控制初始位置和目标位置之间的方差。

$P_n^{(0)}$ 基于真实布局 \hat{P}_n 采样,实现了自监督机制。

[0154] 需要注意的是,本申请实施例中利用编码器训练数据集分别独立训练第一自动编码器以及所述第二自动编码器的过程中,除了在初始模型构建阶段会用到,在后续更新、优化模型时同样也会持续对模型进行训练。

[0155] 也即该步骤可以在S101之前执行以利用编码器训练数据集分别独立训练第一自动编码器以及第二自动编码器。也可以在S102、S103或其他任意时间点执行,从而实现以根据积累的更新训练数据对两个自动编码器进行优化”

[0156] 通过应用本申请的技术方案,可以从海报背景图像中的平滑区域中利用一个级联的自动编码器选取其中的可用区域,并在后续依据另一个级联自动编码器对该可用区域进行迭代细化后自动确定其中的文本框并将文字信息填入其中,从而智能的生成海报图像。进而避免了相关技术中,仅依靠人工设计生成海报所面临的无法满足高质量海报的大量需求的问题。

[0157] 图3示意性地示出了根据本申请实施方式的一种海报的生成方法的流程示意图。如图3所示,该方法包括:

[0158] S201,基于上述的海报文本布局方法获取目标文本布局区域。

[0159] 相关技术中,海报生成的过程至少包含有三个步骤,其中包括:

[0160] 步骤1:选取海报的背景图像;

[0161] 步骤2:在海报背景图像中进行文本布局;

[0162] 步骤3:确定海报上文本的样式并将其填入到背景图像中,从而生成最终的海报图像。

[0163] 基于上述过程,其中针对步骤1,为了实现海报的背景图像的自动搜索,现有的经典检索方法一般通过将海报内容文本与背景图像标注文本匹配来搜索合适的图像。

[0164] 然而,大相关技术中的方法通常只考虑单模态特征,因此通常无法弥合视觉和文本模态之间的语义鸿沟,并会产生偏差较大的检索结果。也即背景图像的选取与用户本次感兴趣的方向并不匹配。

[0165] 另外,针对步骤2,相关技术中为了进行文本布局预测,常规的基于规则的方法通常从有限数量的预定义布局模板中选择文本的布局,但是其灵活性较差。从而导致在布置文本时均忽略了背景图像的内容,因此可能导致海报生成的效果并不理想。

[0166] 进一步的,为了解决上述存在的海报背景图像选取与用户本次感兴趣的方向不匹配的问题,本申请提出一种基于Text2Poster的预训练的视觉文本模型来进行海报背景图像选取方法的技术方案。

[0167] 概括来说,本申请可以首先获取大规模预训练的视觉文本模型,并根据用户给定的海报描述文本在众多候选背景图像中选择与海报描述文本相匹配的候选背景图像作为海报背景图像,然后通过级联自动编码器在该图像上迭代布置文本,最后通过基于匹配的

方法对文本进行风格化处理以合成目标海报。

[0168] 一种方式中,本申请还可以通过弱监督和自我监督的学习策略来优化框架的每个模块,从而减少对标签数据的依赖。从而实现在基于数据驱动框架Text2Poster的技术方案在生成海报的质量方面可以表现的更加优秀的目的。

[0169] 具体的,本申请可以首先获取用户生成的,包含有用于添加在海报上的文字信息的海报描述文本。可以理解的,该海报描述文本即为用户本次想要生成的海报信息的描述问题。其中可以包括多个参数,例如包括标题信息,场景信息,实施方式信息,人物信息等等。

[0170] S202,基于预训练的视觉文本模型,在预获取的候选背景图像集合中选取与文字信息弱相关匹配的候选背景图像作为海报背景图像。

[0171] 可选的,本申请提供的一种海报生成方法中,首先需要基于一种海报背景图像选取的方法,并针对该海报背景图像选取方法实现从多个背景海报图像中选取目标背景海报图像,以由此生成对应的目标海报图像。具体来说,本申请中的海报背景图像选取的方法可以包括如下步骤:

[0172] 获取海报描述文本,所述海报描述文本中包含有用于添加在海报上的文字信息;

[0173] 基于预训练的视觉文本模型,在预获取的候选背景图像集合中选取与所述文字信息弱相关匹配的候选背景图像作为海报背景图像。

[0174] 可选的,所述基于预训练的视觉文本模型,在预获取的候选背景图像集合中选取与所述文字信息弱相关匹配的候选背景图像作为海报背景图像,包括:

[0175] 基于所述预训练的文本编码器,以使该预训练的文本编码器对所述文字信息进行特征提取,得到文字特征;以及,

[0176] 基于所述预训练的图像编码器对每个所述候选背景图像分别进行特征提取,得到各个所述候选背景图像分别对应的图像特征;

[0177] 计算所述文字特征与各个所述候选背景图像分别对应的图像特征之间的弱相关特征相似度,将其中与所述文字特征的弱相关特征相似度最高的候选背景图像作为所述海报背景图像。

[0178] 可选的,所述预训练的文本编码器包括:中文预训练模型中的RoBERTa-Large的编码器;

[0179] 所述基于所述预训练的文本编码器,以使该预训练的文本编码器对所述文字信息进行特征提取,得到文字特征,包括:

[0180] 将所述文字信息输入所述中文预训练模型中的RoBERTa-Large的编码器,以使该编码器对所述文字信息进行特征提取并输出对应的文字特征。

[0181] 可选的,所述预训练的图像编码器包括:预训练的Faster R-CNN和EfficientNet;

[0182] 所述基于所述预训练的图像编码器对每个所述候选背景图像分别进行特征提取,得到各个所述候选背景图像分别对应的图像特征,包括:

[0183] 将各个所述候选背景图像输入所述预训练的图像编码器,以使所述预训练的Faster R-CNN对各个所述候选背景图像进行视觉对象检测处理,再基于所述EfficientNet对经所述视觉对象检测处理后的各个所述候选背景图像进行特征提取,以得到各个所述候选背景图像分别对应的图像特征。

[0184] 可选的,计算所述文字特征与各个所述候选背景图像分别对应的图像特征之间的弱相关特征相似度,将其中与所述待比对文字特征的弱相关特征相似度最高的候选背景图像作为所述海报背景图像,包括:

[0185] 基于预设的弱监督学习方法和对比学习方法CPC构建的InfoNCE损失函数,确定所述文字特征与各个所述候选背景图像分别对应的图像特征之间的弱相关特征相似度。

[0186] 可选的,本申请提供的一种海报生成方法中,还需要基于一种用于选取海报背景图像的视觉文本模型训练方法,以由此实现从多个背景海报图像中选取目标背景海报图像的目的。具体来说,本申请中的用于选取海报背景图像的视觉文本模型训练方法可以包括如下步骤:

[0187] 获取多个弱相关图像文本对,其中,所述弱相关图像文本对用于表示一组弱相关的候选背景图像与历史文本信息;

[0188] 基于多个所述弱相关图像文本对对预设的视觉文本模型进行预训练及训练,得到用于选取与文字信息之间弱相关的海报背景图像的视觉文本模型。

[0189] 可选的,预训练的视觉文本模型包括:包含有预训练的图像编码器和预训练的文本编码器的BriVL;

[0190] 所述预训练的文本编码器包括:中文预训练模型中的RoBERTa-Large的编码器,所述预训练的图像编码器包括:预训练的Faster R-CNN和EfficientNet;

[0191] 所述RoBERTa-Large的编码器用于对各个所述历史文字信息进行特征提取并输出对应的文字特征;

[0192] 所述预训练的Faster R-CNN用于对各个所述候选背景图像进行视觉对象检测处理,所述EfficientNet用于对经所述视觉对象检测处理后的各个所述候选背景图像进行特征提取,以得到各个所述候选背景图像分别对应的图像特征;

[0193] 所述BriVL基于预设的弱监督学习方法和对比学习方法CPC构建的InfoNCE损失函数,确定所述文字特征与各个所述候选背景图像分别对应的图像特征之间的弱相关特征相似度。

[0194] 一种方式中,为了提高海报的高质量性,本申请实施例中可以在海报生成过程中在进行背景图像的检索时,旨在查找与文本信息(即海报描述文本中包含的用于添加在海报上的文字信息)具备弱相关匹配的图像。

[0195] 具体来说,弱相关匹配的方式即为具备隐喻性匹配的方式。例如,在根据短语“鲍勃和爱丽丝的婚礼”检索背景图像时,本申请倾向于找到一些具有更多隐喻性的图像,而不是寻找特定婚礼场景(即非强相关性)的图像,例如一张显示蓝天下的白色教堂的隐喻爱情的图片。

[0196] 可选的,图4为利用本申请的提出的,旨在查找与文本信息(即海报描述文本中包含的用于添加在海报上的文字信息)具备弱相关匹配的背景图像的示意图。其中示出了各个文本信息下对应选取的背景图像。

[0197] 进一步的,为了实现在候选背景图像集合中选取与文字信息弱相关匹配的候选背景图像作为海报背景图像的目的。本申请可以利用预训练视觉文本模型之一的BriVL,基于文本信息从候选图像中选取背景图像。

[0198] 具体来说,例如如图5所示,BriVL即为本申请提出的视觉文本模型,可以看出其由

一个图像编码器和一个文本编码器组成,分别表示为 f_I (即图像编码器)和 f_T (即文本编码器)。

[0199] 其中,图像编码器 f_I 首先利用预训练的Faster R-CNN模型来检测视觉对象,然后应用EfficientNet模型作为其视觉主框架来提取每个背景候选图像对应的图像特征。

[0200] 另外,文本编码器 f_T 则可以使用中文预训练模型中的RoBERTa-Large的编码器作为其文本主框架。基于上述主框架模型的输出,BriVL堆叠多层Transformer以导出文字信息对应的文字特征。

[0201] 需要说明的是,本申请中的视觉文本模型BriVL需要预先在从网络收集的多个(例如为3000万个)弱相关的“样本图像-文本对”上进行了训练,因此能够满足本申请提出的海报生成结果的弱相关艺术性。

[0202] 可以理解的,本申请应用弱监督学习策略和InfoNCE损失函数将文本的特征与图像的特征对齐。其中,对比学习方法CPC(Contrastive Predictive Coding)构建的损失函数即为InfoNCE,其中的NCE是指噪声对比估计(Noise Contrastive Estimation)。

[0203] 一种方式中,本申请可以从图片素材网站等对象处收集多张高质量图像作为我们的候选背景图像作为候选背景图像集合。以使后续基于本申请提及的视觉文本模型BriVL实现海报背景图像的选取。

[0204] 具体来说,由于本申请提取到了各个候选背景图像分别对应的图像特征,以及,文字信息对应的文字特征。因此本申请可以通过分别计算文字特征与各个候选背景图像分别对应的图像特征之间的弱相关特征相似度,并将其中与文字特征的弱相关特征相似度最高的候选背景图像作为海报背景图像。

[0205] 可选的,图5为利用本申请的提出的海报文本布局方法与现有技术中其他文本布局方法的对比示意图。其中示出了各个海报文本布局下对应的目标海报图像。

[0206] 本申请中,可以获取用户生成的海报描述文本,海报描述文本中包含有用于添加在海报上的文字信息;获取候选背景图像集合,并基于视觉文本模型,在候选背景图像集合中选取与文字信息弱相关匹配的候选背景图像作为海报背景图像,其中视觉文本模型由多个标注有与其弱相关文本对的样本图像训练得到;在海报背景图像中确定文本布局区域后,将文字信息填充在文本布局区域中,得到目标海报图像。

[0207] 通过应用本申请的技术方案,可以将多个弱相关图像文本对对预设的视觉文本模型进行训练,并通过该训练得到的视觉文本模型来自动选取与用户感兴趣的文字信息之间弱相关的海报背景图像。进而后续基于该自动选取的海报背景图像生成最终的海报图像。进而避免了相关技术中,仅依靠人工设计生成海报所面临的无法满足高质量海报的大量需求的问题。

[0208] 可选地,在基于本申请上述方法的另一个实施例中,在获取目标文本布局区域之前,还包括:

[0209] 获取海报描述文本,海报描述文本中包含有用于添加在海报上的文字信息;

[0210] 基于预训练的视觉文本模型,在预获取的候选背景图像集合中选取与文字信息弱相关匹配的候选背景图像作为海报背景图像。

[0211] 仍然以上述图6进行举例说明,其中BriVL即为本申请提出的视觉文本模型,可以看出其由一个图像编码器和一个文本编码器组成,分别表示为 f_I (即图像编码器)和 f_T (即文

本编码器)。

[0212] 其中,图像编码器 f_I 首先利用预训练的Faster R-CNN模型来检测视觉对象,然后应用EfficientNet模型作为其视觉主框架来提取每个背景候选图像对应的图像特征。

[0213] 另外,文本编码器 f_T 则可以使用中文预训练模型中的RoBERTa-Large的编码器作为其文本主框架。基于上述主框架模型的输出,BriVL堆叠多层Transformer以导出文字信息对应的文字特征。

[0214] 具体来说,本申请实施例中在得到了各个候选背景图像分别对应的图像特征以及海报描述文本中包含的文字信息对应的文字特征之后,即可以将该多个特征转换为对应的编码向量值来表示。

[0215] 例如,当获取到文字特征的编码向量 r_T ,和每个候选背景图像的编码向量 r_{I_j} 之后,

即可以用 $r_T = f_T(U_i T_i), \{r_{I_j} = f_I(I_j)\}_{j=1}^{N_I}$ 来表示。

[0216] 一种方式中,本申请实施例可以计算 r_T 与每个 r_{I_j} 之间的余弦相似度,并将其中相似度最高的候选背景图像选定为海报背景图像I。

[0217] 其中,计算特征向量之间的余弦相似度的公式可以为:

$$[0218] \quad I = \arg \max_{I_j \in I'} \frac{r_T^T r_{I_j}}{\|r_T\|_2 \|r_{I_j}\|_2}。$$

[0219] 可选地,在基于本申请上述方法的另一个实施例中,预训练的文本编码器包括:中文预训练模型中的RoBERTa-Large的编码器;

[0220] 基于预训练的文本编码器,以使该预训练的文本编码器对文字信息进行特征提取,得到文字特征,包括:

[0221] 将文字信息输入中文预训练模型中的RoBERTa-Large的编码器,以使该编码器对文字信息进行特征提取并输出对应的文字特征。

[0222] 可选地,在基于本申请上述方法的另一个实施例中,预训练的图像编码器包括:预训练的Faster R-CNN和EfficientNet;

[0223] 基于预训练的图像编码器对每个候选背景图像分别进行特征提取,得到各个候选背景图像分别对应的图像特征,包括:

[0224] 将各个候选背景图像输入预训练的图像编码器,以使预训练的Faster R-CNN对各个候选背景图像进行视觉对象检测处理,再基于EfficientNet对经视觉对象检测处理后的各个候选背景图像进行特征提取,以得到各个候选背景图像分别对应的图像特征。

[0225] 可选地,在基于本申请上述方法的另一个实施例中,计算文字特征与各个候选背景图像分别对应的图像特征之间的弱相关特征相似度,将其中与待比对文字特征的弱相关特征相似度最高的候选背景图像作为海报背景图像,包括:

[0226] 基于预设的弱监督学习方法和对比学习方法CPC构建的InfoNCE损失函数,确定文字特征与各个候选背景图像分别对应的图像特征之间的弱相关特征相似度。

[0227] 需要说明的是,本申请中的视觉文本模型BriVL需要预先在从网络收集的多个(例如为3000万个)弱相关的“样本图像-文本对”上进行了训练,因此能够满足本申请提出的海报生成结果的弱相关艺术性。

[0228] 如图7所示,为本申请提出的海报生成方法的整体流程架构图。一种方式中,在本申请利用候选背景图像的图像特征以及文字特征进行弱相关匹配的过程中,可以将其中与待比对文字特征的弱相关特征相似度最高的候选背景图像作为海报背景图像。具体而言,本申请需要应用弱监督学习策略和InfoNCE损失函数将文字特征与每个候选背景图像的图像特征进行特征对齐的操作。并在特征对齐后,在将二者进行若相关性匹配。

[0229] 其中,对于InfoNCE损失函数来说,本申请可以利用对比学习方法CPC(Contrastive Predictive Coding)构建的损失函数作为该InfoNCE,其中的NCE是指噪声对比估计(Noise Contrastive Estimation)。

[0230] 可选地,在基于本申请上述方法的另一个实施例中,将文字信息填充在文本布局区域中以生成文字信息对应的海报,包括:

[0231] 提取文字信息的文字特征,以及检测文本布局区域的背景颜色;

[0232] 基于文本布局区域的背景颜色,从预设的文本样式数据库中选取与文字特征以及背景颜色相匹配的目标文本样式;

[0233] 将文字信息按照目标文本样式进行转换处理后填充在文本布局区域中,得到文字信息对应的目标海报。

[0234] 进一步的,本申请可以对于海报描述文本中的每个文字信息 $T_i \in T'$,均需要提取其对应的文字特征 $r = f_T(T)$,并获得文字特征的背景颜色为 $c^T = I(p)$ 。其中 p_i 是文本信息 T_i 对应的文本布局区域。

[0235] 另外,基于 (r, c^T) ,可以在余弦相似度下从预设的文本样式数据库F中查找与该文字特征以及背景颜色相匹配的目标文本样式,从而确定各个 T_i 的颜色和字体。并将其填入到对应的文本布局区域中。

[0236] A) g_1 中使用的每个卷积层均包含16个大小为 9×9 的卷积核; g_1 的编码器最终输出一个64维的特征向量。

[0237] B) g_2 的编码器的每个卷积层包含64个大小为 5×5 的卷积核; g_2 的2层双向LSTM(解码器)的隐藏层的维度设置为200。

[0238] C) 在训练上述两个自动编码器时,可以划分训练集和验证集,例如可以将数据集D分成138013个训练用海报图像和16000个验证用海报图像。每张海报图像的大小都可以调整为 300×400 。并可以利用Adam算法来优化自动编码器 g_1 和 g_2 的模型,学习率为0.05,批量大小为512。在四个V100 GPU上,分别训练两个自动编码器4(g_1)和48(g_2)小时。

[0239] 最后,为了证明本申请提出的Text2Poster框架的有效性,以下部分还包括对应的验证,具体说明如下:

[0240] 1) 针对图像检索的验证:

[0241] 首先验证(i)图像检索方法的合理性。为了便于对比,除了本申请提供的BriVL方法之外,本申请还参考了(a)应用unsplash.com的搜索引擎和(b)将输入文本与本申请图像检索库中的图像标签进行匹配这两种搜索方式作为对照组。

[0242] 其中,本申请显示了通过三种不同搜索方法获得的具有代表性的检索结果。相较于其他两种方式,采用本申请提供的BriVL方法检索到的图像确实包含与输入文本对应的隐喻。例如,针对给定文本“Campus Charity Sale”(校园慈善义卖),unsplash.com的搜索引擎更倾向于寻找具有明确概念的图像,如“sale”和“campus”,而本申请提供的BriVL方法

检索的则是具有生长的树木、彩虹和彩色手印的图像,其内容隐喻但适合“Campus Charity Sale”(校园慈善义卖)的语义。即使对于像“See the world together”(一起看世界)和“Dreams never stop”(梦想永不停息)这样具有挑战性的抽象描述,本申请提供的BriVL方法仍然可以找到合适的图像。

[0243] 在主观评估中,给定50个文本查询,本申请通过不同的方法检索每个查询的前5个图像。本申请邀请了三名志愿者对检索到的图像质量进行评分,从0(非常差)到4(非常好)。Unsplash.com的搜索引擎的得分平均值和标准差为 2.17 ± 0.10 ,基于标签的匹配方法的得分平均值和标准差为 1.64 ± 0.16 ,本申请提供的BriVL方法的得分平均值和标准差为 2.38 ± 0.13 ,这进一步证明了本申请方法的优越性。

[0244] 2) 针对文本布局预测的验证:

[0245] 本申请定量和定性地评估了本申请提出的布局预测器,并将其与以下对照组进行比较:

[0246] A) 最先进的基于学习的方法,LayoutGAN++;

[0247] B) 最先进的基于规则的方法IUI和DeSal;

[0248] C) <https://luban.aliyun.com>上的商业海报生成器LUBAN。

[0249] 此外,为了证明本申请的迭代布局优化策略的有用性,对于本申请的布局预测器,本申请将K分别设置为1、5和30。

[0250] 本申请从huaban.com收集16,000张海报来构建参考数据集,并准备三个背景图像集:Unsplash2K、Unsplash10K和PSD1.6K。Unsplash2K和Unsplash10K分别包含来自unsplash.com的2,000和10,000个背景图像。PSD1.6K包含从PSD格式的海报文件中提取的1,637张背景图像。对于每个图像集,本申请通过各种方法在背景图像上布置输入文本并生成海报。本申请按照LayoutGAN++中的工作计算海报和参考数据集之间的初始距离FID。表1中的结果表明,本申请的方法始终优于上述对照组,并且其性能随着K的增加而提高,这验证了本申请迭代细化策略的合理性。

[0251] 表1

Method	FID			Aesthetics Score
	PSD1.6K	Unsplash2K	Unsplash10K	
LayoutGAN++	55.46	75.63	58.11	1.50 ± 0.25
IUI	44.39	68.85	53.24	2.38 ± 0.26
DeSal	43.47	70.30	55.25	2.37 ± 0.41
Ours ($K = 1$)	41.83	71.64	55.20	2.27 ± 0.33
Ours ($K = 5$)	39.28	68.77	52.78	2.54 ± 0.21
Ours ($K = 30$)	39.09	67.94	52.75	2.56 ± 0.19

[0253] 表1显示了各种布局预测方法的客观和主观评价。C)中的LUBAN仅提供收费服务,无法进行大规模的量化评价。

[0254] 此外,本申请还手动选择了50个文本集,每个文本集都包含一个标题和几个子标题或描述。

[0255] 对于每种布局预测方法,本申请首先为每组文本检索五个背景图像,并相应地生成250张海报。本申请要求三名志愿者对这些生成的海报从0(非常差)到4(非常好)的布局美学进行评分。对于每种方法,得分的平均值和标准差如表1所示。

[0256] 可选的,在本申请的另外一种实施方式中,如图8所示,本申请还提供一种海报的生成装置。其中包括:

[0257] 确定模块301,确定海报描述文本对应的海报背景图像中的平滑区域,并利用第一自动编码器在所述平滑区域中选取可用区域;

[0258] 生成模块302,对所述可用区域进行采样,根据对应的采样结果和所述海报描述文本对应的用于添加在海报上的文字信息,生成所述可用区域对应的文本框,以得到所述海报背景图像中用于写入所述文字信息的目标文本布局区域。

[0259] 本申请中,可以确定海报描述文本对应的海报背景图像中的平滑区域,并利用第一自动编码器在平滑区域中选取可用区域;对可用区域进行采样,根据对应的采样结果和海报描述文本对应的用于添加在海报上的文字信息,生成可用区域对应的文本框,以得到海报背景图像中用于写入文字信息的目标文本布局区域。通过应用本申请的技术方案,可以从海报背景图像中的平滑区域中利用一个级联的自动编码器选取其中的可用区域,并在后续依据另一个级联自动编码器对该可用区域进行迭代细化后自动确定其中的文本框并将文字信息填入其中,从而智能的生成海报图像。进而避免了相关技术中,仅依靠人工设计生成海报所面临的无法满足高质量海报的大量需求的问题。

[0260] 在本申请的另外一种实施方式中,确定模块301,被配置执行的步骤包括:

[0261] 在所述海报背景图像中选取多个候选框;以及,利用光谱残差算法,生成所述海报背景图像对应的显著图像,其中,各个所述候选框中存在相互重叠的候选框;

[0262] 基于所述显著图像的显著平均值、所述候选框的像素数以及预设的偏移量,为每个候选框确定与其对应的候选值;

[0263] 选取所述候选值低于预设候选阈值的候选框作为目标候选框,并基于所述目标候选框,确定所述海报背景图像中的平滑区域,其中每个目标候选框在所述海报背景图像中的位置各不重叠。

[0264] 在本申请的另外一种实施方式中,确定模块301,被配置执行的步骤包括:

[0265] 基于非极大值抑制方法选取所述候选值低于预设候选阈值的候选框作为目标候选框;

[0266] 确定所有目标候选框所在所述海报背景图像中的候选框区域;

[0267] 将所述海报背景图像中的所述候选框区域转换为二值图,并将转换后的二值图区域作为所述海报背景图像中的平滑区域。

[0268] 在本申请的另外一种实施方式中,确定模块301,被配置执行的步骤包括:

[0269] 将所述平滑区域对应的各个目标候选框分别作为所述第一自动编码器的编码器端输入,得到编码器输出结果,其中所述编码器端由堆叠的CNN所构建得到;

[0270] 将所述编码器输出结果与位置嵌入图进行连接后,作为所述第一自动编码器的解码器端输入,得到所述平滑区域中作为初始文本布局结果的可用区域,其中所述解码器端由堆叠的Transposed-CNN所构建得到。

[0271] 在本申请的另外一种实施方式中,生成模块302,被配置执行的步骤包括:

[0272] 利用第二自动编码器,以自回归方式对各个所述可用区域各自对应的文本框分别进行细化布局处理,得到各个所述可用区域各自对应的目标文本框,以形成所述海报背景图像中用于写入所述文字信息的目标文本布局区域。

[0273] 在本申请的另外一种实施方式中,确定模块301,被配置执行的步骤包括:

[0274] 所述第一自动编码器和所述第二自动编码器均为具有级联自动编码架构的布局预测器。

[0275] 在本申请的另外一种实施方式中,生成模块302,被配置执行的步骤包括:

[0276] 获取多个样本图像,其中每个样本图像中包含标注有对应文本描述字段的文本区域;

[0277] 提取每个样本图像中的样本背景图像,并利用平滑区域检测器确定所述样本背景图像中的样本平滑图像区域;

[0278] 将所述文本描述字段、所述文本区域、所述样本背景图像以及所述样本平滑图像区域合并为编码器训练数据集;

[0279] 利用所述编码器训练数据集分别独立训练所述第一自动编码器以及所述第二自动编码器,以使所述第一自动编码器用于预测文本布局的概率分布以及使所述第二自动编码器用于细化布局本框。

[0280] 在本申请的另外一种实施方式中,生成模块302,被配置执行的步骤包括:

[0281] 利用自监督学习策略训练得到,其中所述第二自动编码器的编码器端由堆叠的CNN所构建得到,所述第二自动编码器的解码器端由2层的双向LSTM所构建得到。

[0282] 图9是根据一示例性实施例示出的一种电子设备的逻辑结构框图。例如,电子设备400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。

[0283] 在示例性实施例中,还提供了一种包括指令的非临时性计算机可读存储介质,例如包括指令的存储器,上述指令可由电子设备处理器执行以完成上述海报的生成方法,该方法包括:确定海报描述文本对应的海报背景图像中的平滑区域,并利用第一自动编码器在所述平滑区域中选取可用区域;对所述可用区域进行采样,根据对应的采样结果和所述海报描述文本对应的用于添加在海报上的文字信息,生成所述可用区域对应的文本框,以得到所述海报背景图像中用于写入所述文字信息的目标文本布局区域。可选地,上述指令还可以由电子设备的处理器执行以完成上述示例性实施例中所涉及的其他步骤。例如,非临时性计算机可读存储介质可以是ROM、随机存取存储器(RAM)、CD-ROM、磁带、软盘和光数据存储设备等。

[0284] 在示例性实施例中,还提供了一种应用程序/计算机程序产品,包括一条或多条指令,该一条或多条指令可以由电子设备的处理器执行,以完成上述海报的生成方法,该方法包括:确定海报描述文本对应的海报背景图像中的平滑区域,并利用第一自动编码器在所述平滑区域中选取可用区域;对所述可用区域进行采样,根据对应的采样结果和所述海报描述文本对应的用于添加在海报上的文字信息,生成所述可用区域对应的文本框,以得到所述海报背景图像中用于写入所述文字信息的目标文本布局区域。可选地,上述指令还可以由电子设备的处理器执行以完成上述示例性实施例中所涉及的其他步骤。

[0285] 图9为电子设备400的示例图。本领域技术人员可以理解,示意图9仅仅是电子设备400的示例,并不构成对电子设备400的限定,可以包括比图示更多或更少的部件,或者组合某些部件,或者不同的部件,例如电子设备400还可以包括输入输出设备、网络接入设备、总线等。

[0286] 所称处理器402可以是中央处理单元(Central Processing Unit,CPU),还可以是其他通用处理器、数字信号处理器(Digital Signal Processor,DSP)、专用集成电路(Application Specific Integrated Circuit,ASIC)、现场可编程门阵列(Field-Programmable Gate Array,FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器402也可以是任何常规的处理器等,处理器402是电子设备400的控制中心,利用各种接口和线路连接整个电子设备400的各个部分。

[0287] 存储器401可用于存储计算机可读指令403,处理器402通过运行或执行存储在存储器401内的计算机可读指令或模块,以及调用存储在存储器401内的数据,实现电子设备400的各种功能。存储器401可主要包括存储程序区和存储数据区,其中,存储程序区可存储操作系统、至少一个功能所需的应用程序(比如声音播放功能、图像播放功能等)等;存储数据区可存储根据电子设备400的使用所创建的数据等。此外,存储器401可以包括硬盘、内存、插接式硬盘,智能存储卡(Smart Media Card,SMC),安全数字(Secure Digital,SD)卡,闪存卡(Flash Card)、至少一个磁盘存储器件、闪存器件、只读存储器(Read-Only Memory,ROM)、随机存取存储器(Random Access Memory,RAM)或其他非易失性/易失性存储器件。

[0288] 电子设备400集成的模块如果以软件功能模块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,本发明实现上述实施例方法中的全部或部分流程,也可以通过计算机可读指令来指令相关的硬件来完成上述提及的计算机可读指令可存储于一计算机可读存储介质中,该计算机可读指令在被处理器执行时,可实现上述各个方法实施例的步骤。

[0289] 本领域技术人员在考虑说明书及实践这里公开的发明后,将容易想到本申请的其它实施方案。本申请旨在涵盖本申请的任何变型、用途或者适应性变化,这些变型、用途或者适应性变化遵循本申请的一般性原理并包括本申请未公开的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的,本申请的真正范围和精神由下面的权利要求指出。

[0290] 应当理解的是,本申请并不局限于上面已经描述并在附图中示出的精确结构,并且可以在不脱离其范围进行各种修改和改变。本申请的范围仅由所附的权利要求来限制。

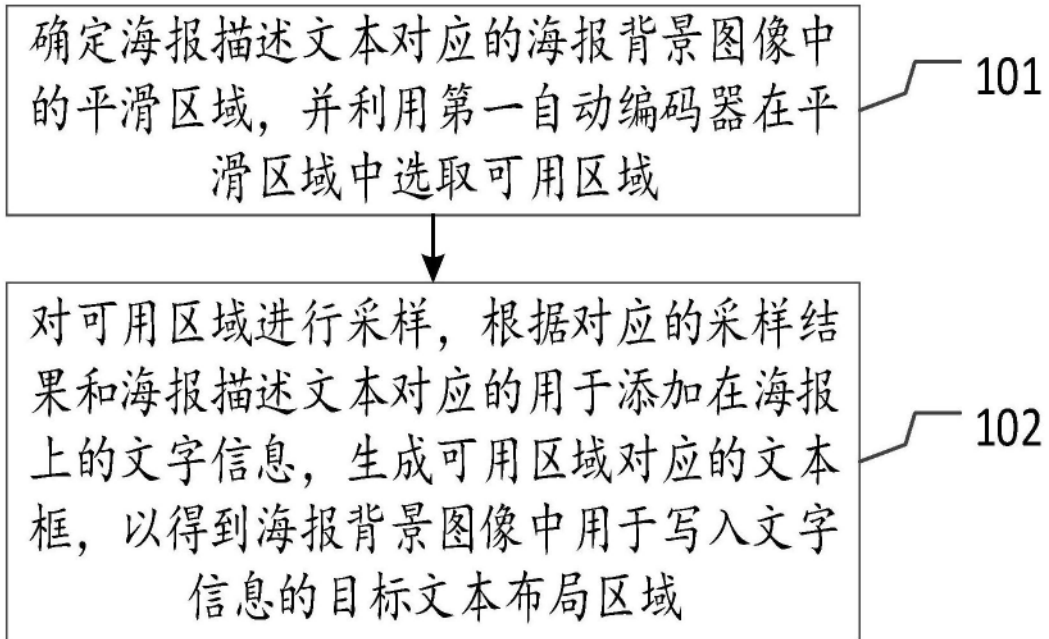


图1

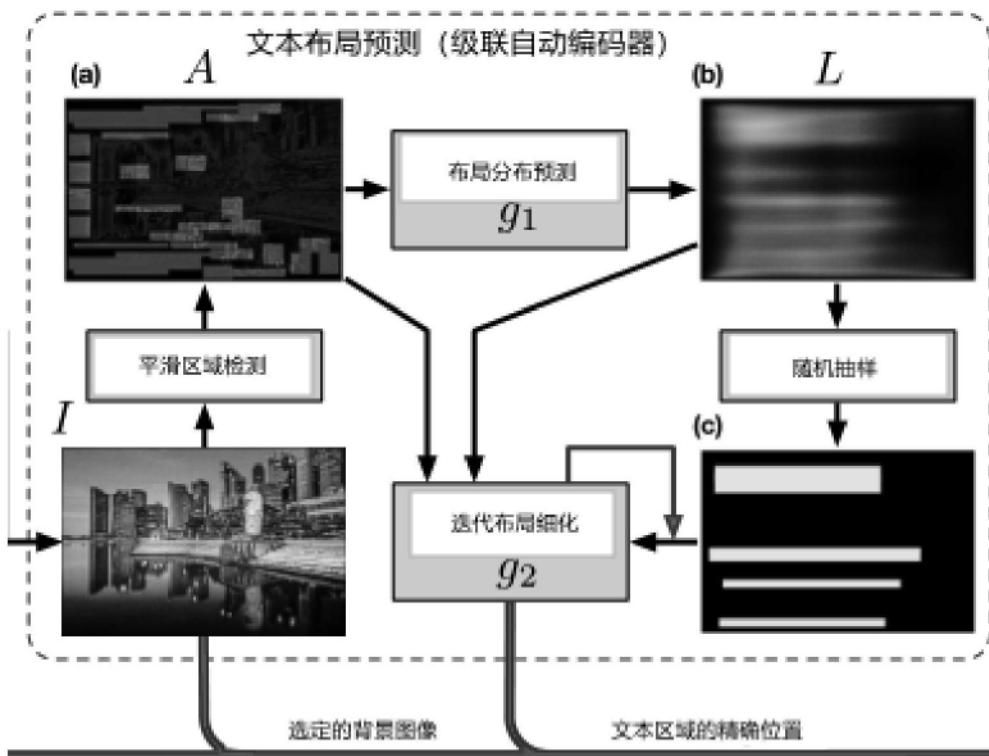


图2

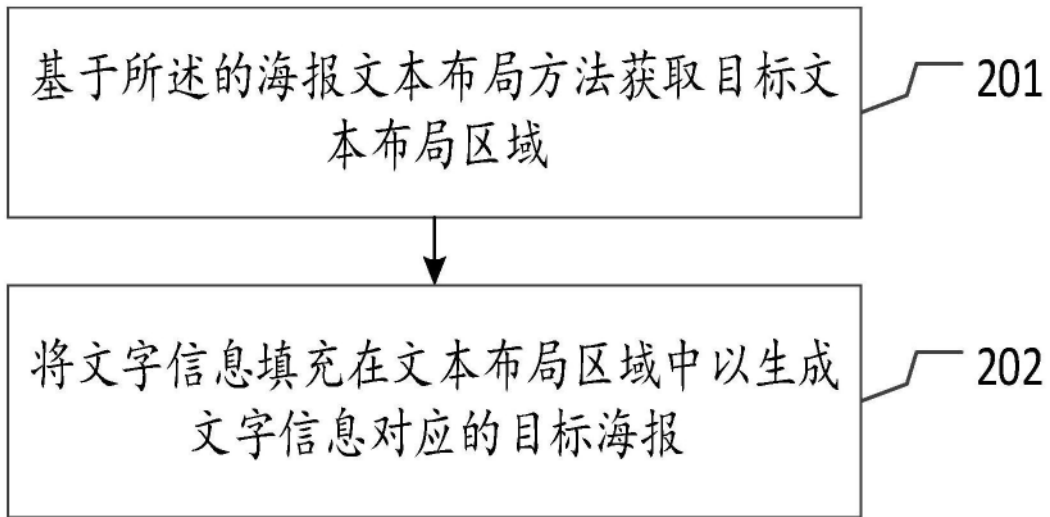


图3



图4

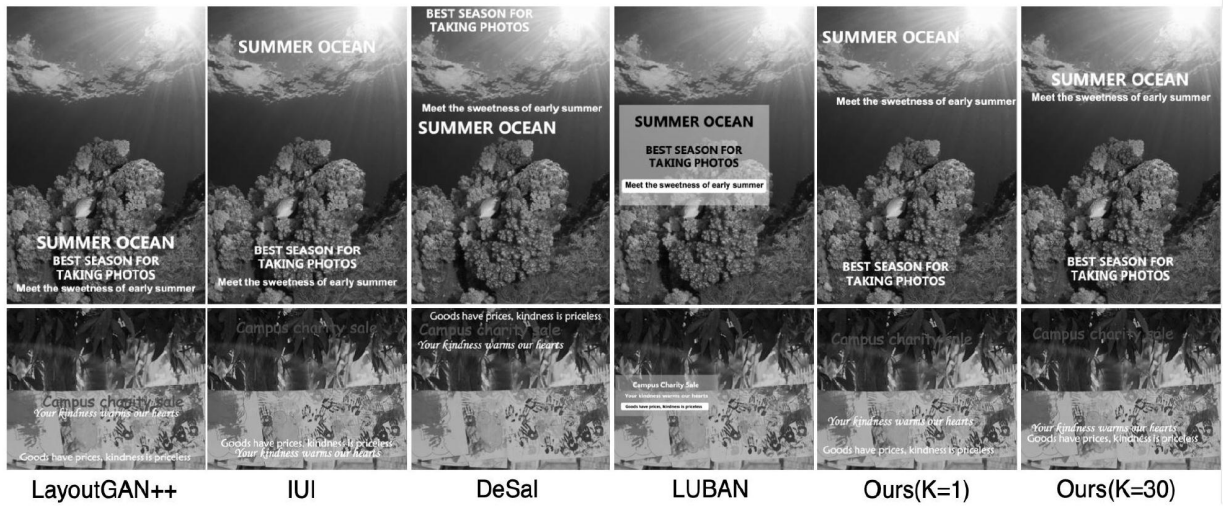


图5

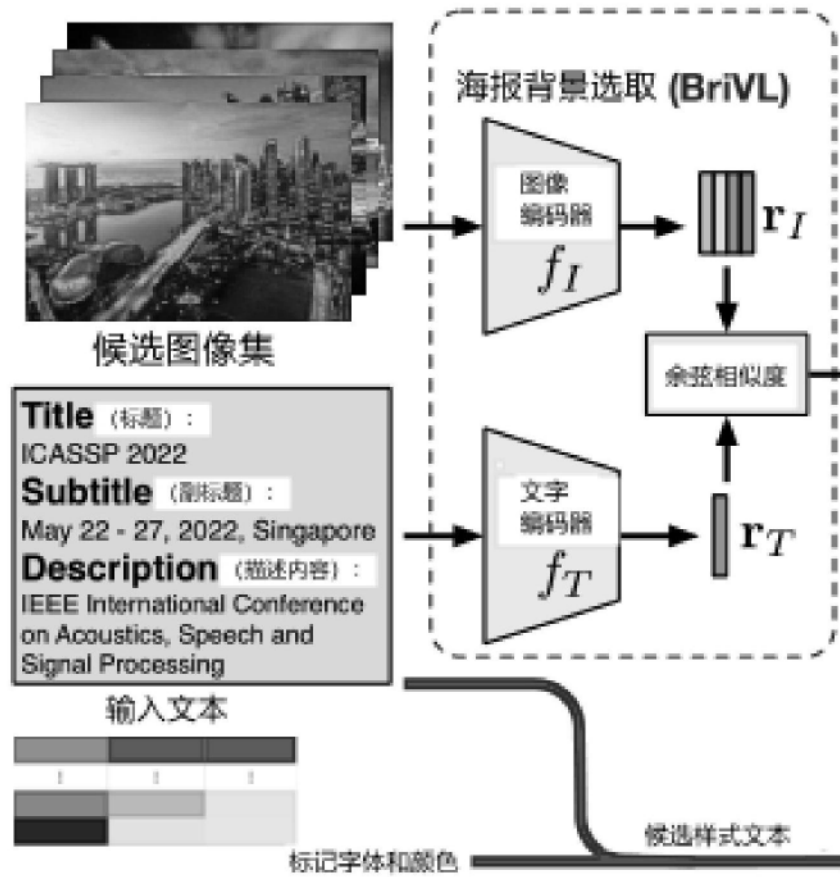


图6

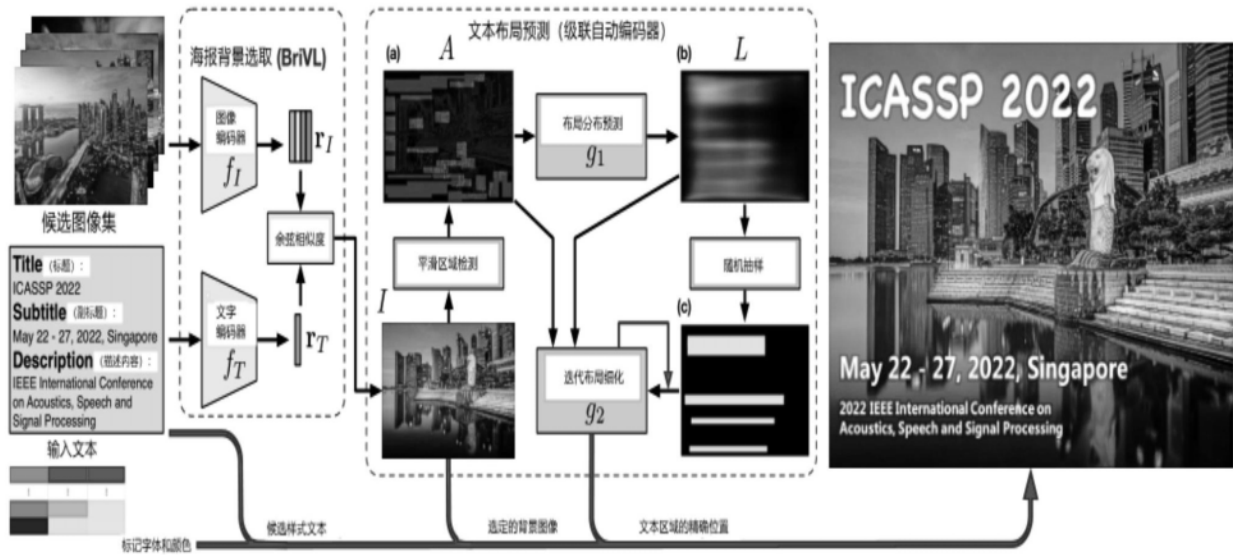


图7

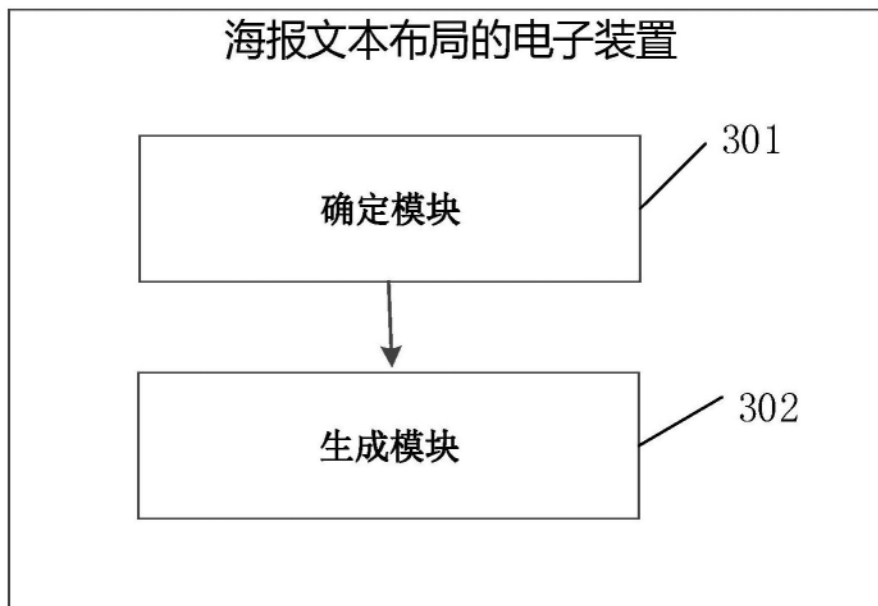


图8

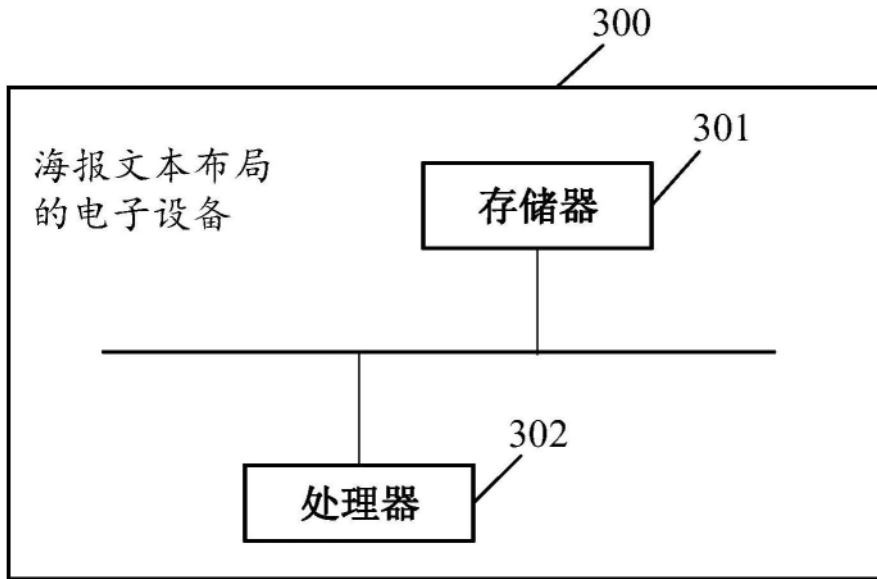


图9