

(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
2022 年 5 月 5 日 (05.05.2022)



(10) 国际公布号
WO 2022/089115 A1

- (51) 国际专利分类号:
G06T 7/11 (2017.01) **G06K 9/46** (2006.01)
- (21) 国际申请号: PCT/CN2021/120815
- (22) 国际申请日: 2021 年 9 月 27 日 (27.09.2021)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202011197790.9 2020 年 10 月 30 日 (30.10.2020) CN
- (71) 申请人: 北京有竹居网络技术有限公司(BEIJING YOUZHUJU NETWORK TECHNOLOGY CO. LTD.) [CN/CN]; 中国北京市平谷区林荫北街 13 号信息大厦 802 室, Beijing 101299 (CN)。
- (72) 发明人: 孔涛(KONG, Tao); 中国北京市海淀区知春路 63 号中国卫星通信大厦今日头条小邮局, Beijing 100190 (CN)。 荆雅(JING, Ya); 中国北京市海淀区知春路 63 号中国卫星通信大厦今日头条小邮局, Beijing 100190 (CN)。 李磊(LI, Lei); 中国北京市海淀区知春路 63 号中国卫星通信大厦今日头条小邮局, Beijing 100190 (CN)。
- (74) 代理人: 北京品源专利代理有限公司(BEYOND ATTORNEYS AT LAW); 中国北京市海淀区莲花池东路 39 号西金大厦 6 层, Beijing 100036 (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, IT,

(54) Title: IMAGE SEGMENTATION METHOD AND APPARATUS, AND DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 图像分割方法、装置、设备及存储介质

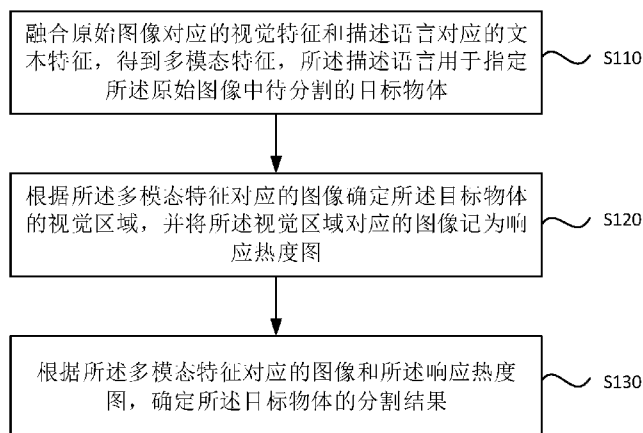


图 1

- S110 Fuse a visual feature corresponding to an original image and a text feature corresponding to a description language to obtain a multi-modal feature, the description language being used for designating a target object to be segmented in the original image
- S120 Determine a visual region of the target object according to an image corresponding to the multi-modal feature, and record an image corresponding to the visual region as a response heat map
- S130 Determine a segmentation result of the target object according to the image corresponding to the multi-modal feature and the response heat map

(57) Abstract: Disclosed are an image segmentation method and apparatus, and a device, and a storage medium. The image segmentation method comprises: fusing a visual feature corresponding to an original image and a text feature corresponding to a description language to obtain a multi-modal feature, the description language being used for designating a target object to be segmented in the original image; determining a visual region of the target object according to an image corresponding to the multi-modal feature, and recording an image corresponding to the visual region as a response heat map; and determining a segmentation result of the target object according to the image corresponding to the multi-modal feature and the response heat map.



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))。

(57) 摘要: 本文公开了一种图像分割方法、装置、设备及存储介质。该图像分割方法包括: 融合原始图像对应的视觉特征和描述语言对应的文本特征, 得到多模态特征, 描述语言用于指定原始图像中待分割的目标物体; 根据多模态特征对应的图像确定目标物体的视觉区域, 并将视觉区域对应的图像记为响应热度图; 根据多模态特征对应的图像和响应热度图, 确定目标物体的分割结果。

图像分割方法、装置、设备及存储介质

本申请要求在2020年10月30日提交中国专利局、申请号为202011197790.9的中国专利申请的优先权，该申请的全部内容通过引用结合在本申请中。

技术领域

本公开涉及图像处理技术领域，例如涉及一种图像分割方法、装置、设备及存储介质。

背景技术

语言指示下的图像分割是跨模态学习中的一项非常重要的技术，又被称为指代性图像分割，语言指示下的图像分割的目标是分割出图像中语言指定的物体。语言指示下的图像分割由于需要消除图像与语言描述之间的语义鸿沟问题，因而更加具有挑战性。

发明内容

本公开提供一种图像分割方法、装置、设备及存储介质，能够在描述语言的指示下有效分割出图像中的指定物体。

本公开提供了一种图像分割方法，包括：

融合原始图像对应的视觉特征和描述语言对应的文本特征，得到多模态特征，其中，所述描述语言用于指定所述原始图像中待分割的目标物体；

根据所述多模态特征对应的图像确定所述目标物体的视觉区域，并将所述视觉区域对应的图像记为响应热度图；

根据所述多模态特征对应的图像和所述响应热度图，确定所述目标物体的分割结果。

本公开还提供了一种图像分割装置，包括：

融合模块，设置为融合原始图像对应的视觉特征和描述语言对应的文本特征，得到多模态特征，其中，所述描述语言用于指定所述原始图像中待分割的目标物体；

视觉区域确定模块，设置为根据所述多模态特征对应的图像确定所述目标物体的视觉区域，并将所述视觉区域对应的图像记为响应热度图；

分割结果确定模块，设置为根据所述多模态特征对应的图像和所述响应热

度图，确定所述目标物体的分割结果。

本公开还提供了一种电子设备，包括：

一个或多个处理器；

存储器，设置为存储一个或多个程序；

当所述一个或多个程序被所述一个或多个处理器执行时实现上述的图像分割方法。

本公开还提供了一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现上述的图像分割方法。

附图说明

图 1 为本公开实施例一提供的一种图像分割方法的流程图；

图 2 为本公开实施例二提供的一种图像分割方法的流程图；

图 3 为本公开实施例二提供的一种图像分割模型的结构示意图；

图 4 为本公开实施例二提供的一种图像分割方法的实现流程图；

图 5 为本公开实施例二提供的一种原始图像的示意图；

图 6 为本公开实施例二提供的一种分割结果的示意图；

图 7 为采用传统方法得到的分割结果的示意图；

图 8 为本公开实施例二提供的一种利用本公开实施例提供的图像分割方法和相关技术对同一图像进行分割的结果对比示意图；

图 9 为本公开实施例三提供的一种图像分割装置的结构图；

图 10 为本公开实施例四提供的一种电子设备的结构图。

具体实施方式

下面将参照附图描述本公开的实施例。虽然附图中显示了一些实施例，然而本公开可以通过多种形式来实现，而且不应该被解释为限于这里阐述的实施例，提供这些实施例是为了更加透彻和完整地理解本公开。

本公开的方法实施方式中记载的多个步骤可以按照不同的顺序执行，和/或并行执行。此外，方法实施方式可以包括附加的步骤和/或省略执行示出的步骤。本公开的范围在此方面不受限制。

本文使用的术语“包括”及其变形是开放性包括，即“包括但不限于”。术语“基于”是“至少部分地基于”。术语“一个实施例”表示“至少一个实施例”；术语“另一

实施例”表示“至少一个另外的实施例”；术语“一些实施例”表示“至少一些实施例”。其他术语的相关定义将在下文描述中给出。

本公开中提及的“第一”、“第二”等概念仅用于对不同的装置、模块、单元或操作进行区分，并非用于限定这些装置、模块、单元或操作所执行的功能的顺序或者相互依存关系。

本公开中提及的“一个”、“多个”的修饰是示意性而非限制性的，除非在上下文另有指出，否则应该理解为“一个或多个”。

本公开实施方式中的多个装置之间所交互的消息或者信息的名称仅用于说明性的目的，而并不是用于对这些消息或信息的范围进行限制。

实施例一

图 1 为本公开实施例一提供的一种图像分割方法的流程图，本实施例可适用于在语言指示下的图像中进行物体分割的情况，可以应用于交互时的图像编辑或人机交互等领域，也可以应用于语言驱动的图像物体检测或语言驱动的图像理解等领域。该方法可以由图像分割装置来执行，该装置可以采用软件和/或硬件的方式实现，该装置可以配置在电子设备中，该电子设备可以是具备图像数据处理功能的终端，例如可以是手机、平板、笔记本等移动终端，也可以是台式机等固定终端或服务器。如图 1 所示，该方法包括如下步骤：

S110、融合原始图像对应的视觉特征和描述语言对应的文本特征，得到多模态特征，所述描述语言用于指定所述原始图像中待分割的目标物体。

原始图像可以是包含至少一个物体的图像，可以通过摄像头或扫描仪等方式获取，也可以从已有的图像库中选择。视觉特征可以是原始图像在设定分辨率下对应的图像特征，视觉特征实际上是一个图像，本实施例对分辨率的大小不进行限定。在一个示例中，可以通过视觉特征提取网络提取原始图像在设定分辨率下的图像特征，得到对应的视觉特征图像。视觉特征提取网络可以采用 Darknet 网络结构或其他可以用于提取视觉特征的网络结构。Darknet 是一个开源的深度学习框架，结构简单，没有依赖项，利用其提取视觉特征时灵活性更好。

描述语言可以是指代性语言描述对应的文本，用于指定分割原始图像中的目标物体。描述语言可以包括图像信息、目标物体的位置信息以及目标物体的外表信息等。图像信息用于确定需要分割的图像的信息，例如可以包括图像的名称或编号等唯一标识该图像的信息。位置信息用于初步确定目标物体在原始图像中的位置。外表信息用于确定最终的目标物体。例如描述语言可以是图像 A 中手拿羽毛球拍、穿红色衣服的用户。文本特征可以是反映描述语言含义的特

征，一般可以采用向量的形式表示。可选的，可以通过文本特征提取网络提取描述语言的文本特征，本实施例对文本特征提取网络的结构不进行限定，例如可以采用门控循环单元（Gated Recurrent Unit, GRU）网络，GRU是循环神经网络的一种。

多模态特征即融合多个模态的特征得到的融合特征，本实施例将视觉特征和文本特征融合得到多模态特征，实现了跨模态的特征表示，消除了图像和描述语言之间的语义鸿沟。可选的，可以将视觉特征和文本特征对应位置的数据叉乘得到视觉特征和文本特征的融合特征，也即多模态特征。多模态特征实际上也是一个图像，多模态特征也可以称为多模态特征图像或多模态特征对应的图像，即包含多模态特征的图像。

S120、根据所述多模态特征对应的图像确定所述目标物体的视觉区域，并将所述视觉区域对应的图像记为响应热度图。

视觉区域为目标物体所在的区域，例如目标物体为披萨时，视觉区域为披萨所在的区域。本实施例在分割目标物体时，首先确定目标物体的视觉区域，在视觉区域的基础上分割目标物体，可以有效降低图像分割的复杂度。可选的，可以对多模态特征对应的图像进行滤波，消除非目标物体对目标物体的干扰，得到目标物体的视觉区域。本实施例将滤波后的图像也即视觉区域对应的图像记为响应热度图，该图表示了目标物体的位置信息，每一个区域对应有一个响应值，响应值越大，表示该区域存在目标物体的可能性越大。可选的，可以将响应值大于设定阈值的区域作为目标物体的视觉区域，并进行高亮显示，本实施例对设定阈值的大小不进行限定。

S130、根据所述多模态特征对应的图像和所述响应热度图，确定所述目标物体的分割结果。

可选的，可以通过图像分割模型结合多模态特征对应的图像和响应热度图确定分割结果。图像分割模型用于确定目标物体的分割结果，该模型的结构可以根据需要设定，例如可以包括卷积层和上采样层，卷积层用于对输入的图像执行卷积操作，上采样层用于对卷积结果进行上采样，得到分割结果，分割结果所对应图像的大小与原始图像中真实分割结果的大小相同。本实施例的图像分割模型以多模态特征对应的图像和响应热度图为输入，应用之前，可以将多模态特征对应的图像和响应热度图输入图像分割模型，对图像分割模型进行训练，以调整卷积层和上采样层的参数。可选的，可以确定图像分割模型输出的分割结果相对于原始图像对应的真实分割结果的损失值，当损失值小于设定阈值时，训练结束，并将损失值小于设定阈值时所对应的模型作为本实施例用于分割目标物体的图像分割模型。

本公开实施例一提供一种图像分割方法，通过融合原始图像对应的视觉特征和描述语言对应的文本特征，得到多模态特征，所述描述语言用于指定所述原始图像中待分割的目标物体；根据所述多模态特征对应的图像确定所述目标物体的视觉区域，并将所述视觉区域对应的图像记为响应热度图；根据所述多模态特征对应的图像和所述响应热度图，确定所述目标物体的分割结果。该方法对图像分割过程进行了分解，首先基于融合得到的多模态特征对应的图像确定目标物体的视觉区域，得到响应热度图，然后根据多模态特征对应的图像和响应热度图确定分割结果，有效消除了图像和描述语言之间的语义鸿沟，分割出了描述语言指定的目标物体。

实施例二

图 2 为本公开实施例二提供的一种图像分割方法的流程图，本实施例是在上述实施例的基础上进行说明，参考图 2，该方法可以包括如下步骤：

S210、融合原始图像对应的视觉特征和描述语言对应的文本特征，得到多模态特征。

视觉特征的数量可以是一个或多个，为了提高分割结果的准确度，可以包括在至少两个分辨率下分别从原始图像中提取的视觉特征。视觉特征的数量过少会影响分割结果的准确度，过多会增加计算量。本实施例以三个视觉特征为例，在提高分割结果准确度的同时可以减小计算量。本实施例的视觉特征可以包括原始图像（ $R^{H \times W \times 3}$ ）在第一分辨率下提取的第一视觉特征 $F_{v1} \in R^{\frac{H}{32} \times \frac{W}{32} \times d_1}$ 、原始图像在第二分辨率下提取的第二视觉特征 $F_{v2} \in R^{\frac{H}{16} \times \frac{W}{16} \times d_2}$ 和原始图像在第三分辨率下提取的第三视觉特征 $F_{v3} \in R^{\frac{H}{8} \times \frac{W}{8} \times d_3}$ ，其中，第一分辨率 < 第二分辨率 < 第三分辨率。第一分辨率、第二分辨率和第三分辨率的数值可以根据情况设定，本实施例以第一分辨率为原始图像的分辨率的 $\frac{1}{32}$ 、第二分辨率为原始图像的分辨率的 $\frac{1}{16}$ 、第三分辨率为原始图像的分辨率的 $\frac{1}{8}$ 为例。 H 和 W 分别为原始图像的长和宽， d_i 为视觉特征所对应图像的维度，本实施例中 $i=1,2,3$ 。

在一个示例中，可以通过如下方式得到多模态特征：

按照分辨率的大小顺序排列所述至少两个视觉特征，得到排序结果；通过映射矩阵将所述文本特征映射到所述排序结果中的第一分辨率对应的第一视觉特征所在的特征空间，所述第一分辨率的值最小；拼接所述第一视觉特征和映射后的文本特征，得到第一拼接特征；对所述第一拼接特征进行上采样，并将上采样后的第一拼接特征与所述排序结果中的第二分辨率对应的第二视觉特征

进行拼接，得到第二拼接特征，循环执行上采样和拼接的操作，直至将上采样后的拼接特征与所述排序结果中的分辨率最大的视觉特征拼接，得到多模态特征为止，所述第二分辨率大于所述第一分辨率，所述第二分辨率为除所述第一分辨率之外最小的分辨率。

考虑到文本特征和视觉特征的长度不同，为了保证融合效果，可以先将文本特征和视觉特征映射到同一个特征空间。例如可以将视觉特征映射到文本特征所在的空间，也可以将文本特征映射到视觉特征所在的空间，还可以将文本特征和视觉特征映射到其他特征空间。实施例以将文本特征映射到视觉特征所在的特征空间为例，可以简化映射过程，降低计算量。

当视觉特征存在多个时，其对应的分辨率也不同，对应的图像大小也不同，为了保证拼接的有效性，本实施例基于多个视觉特征的分辨率由小到大的顺序对多个视觉特征进行排序，得到排序结果。在拼接特征时，按照排序结果中的视觉特征的分辨率由小到大的顺序进行，即先拼接分辨率最小的视觉特征和文本特征，然后对拼接结果进行上采样，再和排序结果中下一个分辨率对应的视觉特征进行拼接，直至和分辨率最大的视觉特征拼接结束为止。

以视觉特征包括三个为例，分别为第一视觉特征、第二视觉特征和第三视觉特征，第一视觉特征、第二视觉特征和第三视觉特征分别对应第一分辨率、第二分辨率和第三分辨率，其中，第一分辨率<第二分辨率<第三分辨率。可以通过叉乘的方式计算第一视觉特征和映射后的文本特征对应位置的值，实现第一视觉特征和映射后的文本特征的拼接，得到第一拼接特征。如上所述，第一视觉特征的分辨率<第二视觉特征的分辨率<第三视觉特征的分辨率，也即第一拼接特征的分辨率小于第二视觉特征的分辨率，为了实现相同分辨率下的特征拼接，本实施例对第一拼接特征进行上采样，使上采样得到的拼接特征的分辨率与第二视觉特征的分辨率相同，然后执行与第一拼接特征类似的拼接操作，拼接第二视觉特征和上采样后的第一拼接特征，得到第二拼接特征，然后对第二拼接特征进行上采样，并将上采样后的第二拼接特征与第三视觉特征进行拼接，得到第三拼接特征，也即多模态特征。本实施例将文本特征与视觉特征融合，实现了特征的跨模态表示，在后续分割目标物体时，可以提高分割结果的准确度。

S220、对所述多模态特征对应的图像进行相关性滤波，得到所述目标物体的视觉区域。

在一个示例中，可以根据文本特征生成一个卷积核，根据该卷积核对多模态特征对应的图像进行卷积操作，实现对多模态特征图像的相关性滤波，得到每个区域对应的响应值，响应值越大的区域表示存在目标物体的可能性越大，

根据响应值即可得到对应的视觉区域。

S230、将所述视觉区域对应的图像记为响应热度图。

S240、将所述多模态特征对应的图像和所述响应热度图输入目标图像分割模型，获取所述目标图像分割模型输出的结果，作为所述目标物体的分割结果。

本实施例中，根据视觉区域，设计了一个初始图像分割模型，以获取更加准确的分割结果。示例性的，参考图 3，图 3 为本公开实施例二提供的一种初始图像分割模型的结构示意图，该图像分割模型包括输入层、并行的第一卷积层、拼接层、第二卷积层、上采样层和输出层，第一卷积层的数量可以根据情况设定，图 3 以 5 个第一卷积层为例，从而可以更好地捕捉图像在不同尺度上的内容。每个第一卷积层对应一个采样率，即分别对输入图像执行 5 种不同采样率的第一卷积操作，得到 5 种卷积结果。拼接层用于拼接这 5 种卷积结果。第二卷积层用于对拼接后的结果再次执行卷积操作。上采样层用于保证图像分割模型输出的分割结果的分辨率与原始图像真实分割结果的分辨率一致。

在应用图像分割模型之前，可以对初始图像分割模型中第一卷积层、拼接层、第二卷积层和上采样层的参数进行训练，得到目标图像分割模型。训练过程如下：

获取样本图像和样本描述语言，并提取所述样本图像的样本视觉特征和所述样本描述语言的样本文本特征；融合所述样本视觉特征和所述样本文本特征，得到样本多模态特征；根据所述样本多模态特征对应的图像确定所述样本目标物体的样本视觉区域，并将所述样本视觉区域对应的图像记为样本响应热度图；根据所述样本多模态特征对应的图像和所述样本响应热度图训练初始图像分割模型，得到所述目标图像分割模型。

本实施例对样本图像和样本描述语言的数量不进行限定，为了提高图像分割模型的准确度，可以选择多组样本图像和多组样本描述语言；然后提取并拼接样本图像的样本视觉特征和样本描述语言的样本文本特征，得到样本多模态特征；对多模态特征进行相关性滤波得到样本响应热度图；由此可以根据样本多模态特征对应的图像和样本响应热度图训练初始图像分割模型，得到目标图像分割模型。特征的提取、拼接和滤波过程可以参考上述实施例，此处不再赘述。

可以按照图 3 所示的结构，将样本多模态特征对应的图像和样本响应热度图输入初始图像分割模型，得到样本多模态特征对应的图像和样本响应热度图的多个第一卷积结果；拼接多个第一卷积结果，得到拼接结果；对拼接结果进行第二卷积操作，得到第二卷积结果；对第二卷积结果进行上采样，得到样本

分割结果；确定样本分割结果相对样本图像的真实分割结果的损失值；当损失值小于设定阈值时，停止训练初始图像分割模型，并将损失值小于设定阈值的图像分割模型作为目标图像分割模型；当损失值不小于设定阈值时，继续训练初始图像分割模型，直至损失值小于设定阈值。可选的，可以采用如下损失函数确定样本分割结果相对样本图像的真实分割结果的损失值：

$$L = \sum_{l=1}^{\frac{H}{4} \times \frac{W}{4}} [y_l \log(p_l) + (1 - y_l) \log(1 - p_l)]$$

其中， L 为样本分割结果相对样本图像的真实分割结果的损失值， y_l 为原始图像下采样后的真实分割结果中每个区域的元素值， p_l 为样本分割结果中每个区域的元素值。设定阈值的大小可以根据情况设定，例如可以为0.5，即当 $L < 0.5$ 时，训练结束。

示例性的，参考图4，图4为本公开实施例二提供的一种图像分割方法的实现流程图。首先获取原始图像和描述语言，然后提取原始图像在不同层级的视觉特征和描述语言对应的文本特征，图4以三个层级为例，分别对应三种分辨率，然后拼接第一视觉特征 F_{v1} 与映射后的文本特征得到第一拼接特征 F_{m1} ，对第一拼接特征 F_{m1} 上采样之后与第二视觉特征 F_{v2} 进行拼接，得到第二拼接特征 F_{m2} ，对第二拼接特征 F_{m2} 上采样之后与第三视觉特征 F_{v3} 进行拼接，得到第三拼接特征 F_{m3} ，也即多模态特征。然后对多模态特征 F_{m3} 进行相关性滤波，得到响应热度图，将响应热度图和多模态特征 F_{m3} 对应的图像输入目标图像分割模型，即可得到目标物体的分割结果，简单有效。

示例性的，参考图5，图5为本公开实施例二提供的一种原始图像的示意图，假定语言文本为“Pizza Nearest”，即分割出距离最近的披萨，按照上述实施例提供的图像分割方法，可以得到如图6所示的分割结果。利用传统方法得到的分割结果如图7所示。示例性的，参考图8，图8为本公开实施例二提供的一种利用本公开实施例提供的图像分割方法和相关技术对同一图像进行分割的结果对比示意图。其中，第一列分别为三种原始图像，第二列分别为采用本公开实施例提供的方法分割得到的物体，第三列和第四列分别为采用相关技术分割得到的物体。由图6和8可以看出，利用本公开实施例提供的图像分割方法分割得到的物体更接近其真实结果，提高了图像分割结果的准确度。

本公开实施例二提供一种图像分割方法，在上述实施例的基础上，对图像分割的过程进行分解，先初步确定目标物体的视觉区域，然后构建初始图像分割模型，简化了初始图像分割模型的复杂度，利用多模态特征图像和响应热度图训练初始图像分割模型，得到目标图像分割模型，进而利用目标图像分割模型得到分割结果，有效消除了图像和描述语言之间的语义鸿沟，也在一定程度

上提高了分割结果的准确度。

实施例三

图 9 为本公开实施例三提供的一种图像分割装置的结构图，该装置可以执行上述实施例所述的图像分割方法，该装置可以集成在电子设备中，参考图 9，该装置可以包括：

融合模块 31，设置为融合原始图像对应的视觉特征和描述语言对应的文本特征，得到多模态特征，所述描述语言用于指定所述原始图像中待分割的目标物体；视觉区域确定模块 32，设置为根据所述多模态特征对应的图像确定所述目标物体的视觉区域，并将所述视觉区域对应的图像记为响应热度图；分割结果确定模块 33，设置为根据所述多模态特征对应的图像和所述响应热度图，确定所述目标物体的分割结果。

本公开实施例提供一种图像分割装置，通过融合原始图像对应的视觉特征和描述语言对应的文本特征，得到多模态特征，所述描述语言用于指定所述原始图像中待分割的目标物体；根据所述多模态特征对应的图像确定所述目标物体的视觉区域，并将所述视觉区域对应的图像记为响应热度图；根据所述多模态特征对应的图像和所述响应热度图，确定所述目标物体的分割结果。该装置对图像分割过程进行了分解，首先基于融合得到的多模态特征对应的图像确定目标物体的视觉区域，得到响应热度图，然后根据多模态特征对应的图像和响应热度图确定分割结果，有效消除了图像和描述语言之间的语义鸿沟，分割出了描述语言指定的目标物体。

在上述实施例的基础上，视觉区域确定模块 32，包括：

滤波单元，设置为对所述多模态特征对应的图像进行相关性滤波，得到所述目标物体的视觉区域。

在上述实施例的基础上，滤波单元，设置为：

根据所述文本特征确定卷积核；根据所述卷积核对所述多模态特征对应的图像进行卷积操作，得到所述目标物体的视觉区域。

在上述实施例的基础上，所述视觉特征包括在至少两个分辨率下分别从所述原始图像中提取的视觉特征。

在上述实施例的基础上，融合模块 31，设置为：

按照分辨率的大小顺序排列所述至少两个视觉特征，得到排序结果；通过映射矩阵将所述文本特征映射到所述排序结果中的第一分辨率对应的第一视觉特征所在的特征空间，所述第一分辨率的值最小；拼接所述第一视觉特征和映

射后的文本特征，得到第一拼接特征；对所述第一拼接特征进行上采样，并将上采样后的第一拼接特征与所述排序结果中的第二分辨率对应的第二视觉特征进行拼接，得到第二拼接特征，循环执行上采样和拼接的操作，直至将上采样后的拼接特征与所述排序结果中的分辨率最大的视觉特征拼接，得到多模态特征为止，所述第二分辨率大于所述第一分辨率，所述第二分辨率为除所述第一分辨率之外最小的分辨率。

在上述实施例的基础上，分割结果确定模块 33，设置为：

将所述多模态特征对应的图像和所述响应热度图输入目标图像分割模型，获取所述目标图像分割模型输出的结果，作为所述目标物体的分割结果。

在上述实施例的基础上，所述目标图像分割模型的训练过程如下：

获取样本图像和样本描述语言，并提取所述样本图像的样本视觉特征和所述样本描述语言的样本文本特征，所述样本描述语言用于指定所述样本图像中待分割的样本目标物体；融合所述样本视觉特征和所述样本文本特征，得到样本多模态特征；根据所述样本多模态特征对应的图像确定所述样本目标物体的样本视觉区域，并将所述样本视觉区域对应的图像记为样本响应热度图；根据所述样本多模态特征对应的图像和所述样本响应热度图训练初始图像分割模型，得到所述目标图像分割模型。

在上述实施例的基础上，所述根据所述样本多模态特征对应的图像和所述样本响应热度图训练初始图像分割模型，得到所述目标图像分割模型，包括：

将所述样本多模态特征对应的图像和所述样本响应热度图输入所述初始图像分割模型，得到所述样本多模态特征对应的图像和所述样本响应热度图的多个第一卷积结果，所述多个第一卷积结果由对所述样本多模态特征对应的图像和所述样本响应热度图在不同采样率下执行第一卷积操作得到；拼接所述多个第一卷积结果，得到拼接结果；对所述拼接结果进行第二卷积操作，得到第二卷积结果；对所述第二卷积结果进行上采样，得到样本分割结果；确定所述样本分割结果相对所述样本图像的真实分割结果的损失值；当所述损失值小于设定阈值时，停止训练所述初始图像分割模型，并将损失值小于设定阈值的图像分割模型作为所述目标图像分割模型；当所述损失值不小于设定阈值时，继续训练所述初始图像分割模型，直至损失值小于设定阈值。

本公开实施例提供的图像分割装置与上述实施例提供的图像分割方法属于同一构思，未在本实施例中详尽描述的技术细节可参见上述实施例，并且本实施例具备执行图像分割方法相同的效果。

实施例四

下面参考图 10，其示出了适于用来实现本公开实施例的电子设备 600 的结构示意图。本公开实施例中的电子设备可以包括但不限于诸如移动电话、笔记本电脑、数字广播接收器、个人数字助理（Personal Digital Assistant，PDA）、平板电脑（PAD）、便携式多媒体播放器（Portable Media Player，PMP）、车载终端（例如车载导航终端）等等的移动终端以及诸如数字（Television，TV）、台式计算机以及服务器等等的固定终端。图 10 示出的电子设备仅仅是一个示例，不应对本公开实施例的功能和使用范围带来任何限制。

如图 10 所示，电子设备 600 可以包括处理装置（例如中央处理器、图形处理器等）601，其可以根据存储在只读存储器（Read-Only Memory，ROM）602 中的程序或者从存储装置 608 加载到随机访问存储器（Random Access Memory，RAM）603 中的程序而执行多种适当的动作和处理。在 RAM 603 中，还存储有电子设备 600 操作所需的多种程序和数据。处理装置 601、ROM 602 以及 RAM 603 通过总线 604 彼此相连。输入/输出（Input/Output，I/O）接口 605 也连接至总线 604。

通常，以下装置可以连接至 I/O 接口 605：包括例如触摸屏、触摸板、键盘、鼠标、摄像头、麦克风、加速度计、陀螺仪等的输入装置 606；包括例如液晶显示器（Liquid Crystal Display，LCD）、扬声器、振动器等输出装置 607；包括例如磁带、硬盘等的存储装置 608；以及通信装置 609。通信装置 609 可以允许电子设备 600 与其他设备进行无线或有线通信以交换数据。虽然图 10 示出了具有多种装置的电子设备 600，但是并不要求实施或具备所有示出的装置。可以替代地实施或具备更多或更少的装置。

根据本公开的实施例，上文参考流程图描述的过程可以被实现为计算机软件程序。例如，本公开的实施例包括一种计算机程序产品，其包括承载在非暂态计算机可读介质上的计算机程序，该计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中，该计算机程序可以通过通信装置 609 从网络上被下载和安装，或者从存储装置 608 被安装，或者从 ROM 602 被安装。在该计算机程序被处理装置 601 执行时，执行本公开实施例的方法中限定的上述功能。

实施例五

本公开上述的计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件，或者任意以上的组合。计算机可读存储介质的例子可以包括但不限于：具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、RAM、ROM、可擦式可编程只

读存储器 (Erasable Programmable Read-Only Memory, EPROM 或闪存)、光纤、便携式紧凑磁盘只读存储器 (Compact Disc Read-Only Memory, CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本公开中, 计算机可读存储介质可以是任何包含或存储程序的有形介质, 该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。而在本公开中, 计算机可读信号介质可以包括在基带中或者作为载波一部分传播的数据信号, 其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式, 包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读信号介质还可以是计算机可读存储介质以外的任何计算机可读介质, 该计算机可读信号介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输, 包括但不限于: 电线、光缆、射频 (Radio Frequency, RF) 等等, 或者上述的任意合适的组合。

在一些实施方式中, 客户端、服务器可以利用诸如超文本传输协议 (HyperText Transfer Protocol, HTTP) 之类的任何当前已知或未来研发的网络协议进行通信, 并且可以与任意形式或介质的数字数据通信 (例如, 通信网络) 互连。通信网络的示例包括局域网 (Local Area Network, LAN), 广域网 (Wide Area Network, WAN), 网际网 (例如, 互联网) 以及端对端网络 (例如, ad hoc 端对端网络), 以及任何当前已知或未来研发的网络。

上述计算机可读介质可以是上述电子设备中所包含的; 也可以是单独存在, 而未装配入该电子设备中。

上述计算机可读介质承载有一个或者多个程序, 当上述一个或者多个程序被该电子设备执行时, 使得该电子设备: 融合原始图像对应的视觉特征和描述语言对应的文本特征, 得到多模态特征, 所述描述语言用于指定所述原始图像中待分割的目标物体; 根据所述多模态特征对应的图像确定所述目标物体的视觉区域, 并将所述视觉区域对应的图像记为响应热度图; 根据所述多模态特征对应的图像和所述响应热度图, 确定所述目标物体的分割结果。

可以以一种或多种程序设计语言或其组合来编写用于执行本公开的操作的计算机程序代码, 上述程序设计语言包括但不限于面向对象的程序设计语言——诸如 Java、Smalltalk、C++, 还包括常规的过程式程序设计语言——诸如“C”语言或类似的程序设计语言。程序代码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 LAN 或 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）。

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

描述于本公开实施例中所涉及到的模块可以通过软件的方式实现，也可以通过硬件的方式来实现。其中，模块的名称在一种情况下并不构成对该模块本身的限定，例如，拼接模块还可以被描述为“拼接原始图像对应的视觉特征和语言文本对应的文本特征，得到多模态特征的模块”。

本文中以上描述的功能可以至少部分地由一个或多个硬件逻辑部件来执行。例如，非限制性地，可以使用的示范类型的硬件逻辑部件包括：现场可编程门阵列（Field Programmable Gate Array, FPGA）、专用集成电路（Application Specific Integrated Circuit, ASIC）、专用标准产品（Application Specific Standard Parts, ASSP）、片上系统（System on Chip, SOC）、复杂可编程逻辑设备（Complex Programmable Logic Device, CPLD）等等。

在本公开的上下文中，机器可读介质可以是有形的介质，其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备，或者上述内容的任何合适组合。机器可读存储介质的示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、RAM、ROM、EPROM 或快闪存储器、光纤、CD-ROM、光学储存设备、磁储存设备、或上述内容的任何合适组合。

根据本公开的一个或多个实施例，本公开提供了一种图像分割方法，包括：

融合原始图像对应的视觉特征和描述语言对应的文本特征，得到多模态特征，其中，所述描述语言用于指定所述原始图像中待分割的目标物体；根据所述多模态特征对应的图像确定所述目标物体的视觉区域，并将所述视觉区域对

应的图像记为响应热度图；根据所述多模态特征对应的图像和所述响应热度图，确定所述目标物体的分割结果。

根据本公开的一个或多个实施例，本公开提供的图像分割方法中，所述根据所述多模态特征对应的图像确定所述目标物体的视觉区域，包括：

对所述多模态特征对应的图像进行相关性滤波，得到所述目标物体的视觉区域。

根据本公开的一个或多个实施例，本公开提供的图像分割方法中，所述对所述多模态特征对应的图像进行相关性滤波，得到所述目标物体的视觉区域，包括：

根据所述文本特征确定卷积核；根据所述卷积核对所述多模态特征对应的图像进行卷积操作，得到所述目标物体的视觉区域。

根据本公开的一个或多个实施例，本公开提供的图像分割方法中，所述视觉特征包括在至少两个分辨率下分别从所述原始图像中提取的视觉特征。

根据本公开的一个或多个实施例，本公开提供的图像分割方法中，所述融合原始图像对应的视觉特征和描述语言对应的文本特征，得到多模态特征，包括：

按照分辨率的大小顺序排列所述至少两个视觉特征，得到排序结果；通过映射矩阵将所述文本特征映射到所述排序结果中的第一分辨率对应的第一视觉特征所在的特征空间，所述第一分辨率的值最小；拼接所述第一视觉特征和映射后的文本特征，得到第一拼接特征；对所述第一拼接特征进行上采样，并将上采样后的第一拼接特征与所述排序结果中的第二分辨率对应的第二视觉特征进行拼接，得到第二拼接特征，循环执行上采样和拼接的操作，直至将上采样后的拼接特征与所述排序结果中的分辨率最大的视觉特征拼接，得到多模态特征为止，所述第二分辨率大于所述第一分辨率，所述第二分辨率为除所述第一分辨率之外最小的分辨率。

根据本公开的一个或多个实施例，本公开提供的图像分割方法中，所述根据所述多模态特征对应的图像和所述响应热度图，确定所述目标物体的分割结果，包括：

将所述多模态特征对应的图像和所述响应热度图输入目标图像分割模型，获取所述目标图像分割模型输出的结果，作为所述目标物体的分割结果。

根据本公开的一个或多个实施例，本公开提供的图像分割方法中，所述目标图像分割模型的训练过程如下：

获取样本图像和样本描述语言，并提取所述样本图像的样本视觉特征和所述样本描述语言的样本文本特征，其中，所述样本描述语言用于指定所述样本图像中待分割的样本目标物体；融合所述样本视觉特征和所述样本文本特征，得到样本多模态特征；根据所述样本多模态特征对应的图像确定所述样本目标物体的样本视觉区域，并将所述样本视觉区域对应的图像记为样本响应热度图；根据所述样本多模态特征对应的图像和所述样本响应热度图训练初始图像分割模型，得到所述目标图像分割模型。

根据本公开的一个或多个实施例，本公开提供的图像分割方法中，所述根据所述样本多模态特征对应的图像和所述样本响应热度图训练初始图像分割模型，得到所述目标图像分割模型，包括：

将所述样本多模态特征对应的图像和所述样本响应热度图输入所述初始图像分割模型，得到所述样本多模态特征对应的图像和所述样本响应热度图的多个第一卷积结果，其中，所述多个第一卷积结果由对所述样本多模态特征对应的图像和所述样本响应热度图在不同采样率下执行第一卷积操作得到；拼接所述多个第一卷积结果，得到拼接结果；对所述拼接结果进行第二卷积操作，得到第二卷积结果；对所述第二卷积结果进行上采样，得到样本分割结果；确定所述样本分割结果相对所述样本图像的真实分割结果的损失值；当所述损失值小于设定阈值时，停止训练所述初始图像分割模型，并将损失值小于设定阈值的图像分割模型作为所述目标图像分割模型；当所述损失值不小于设定阈值时，继续训练所述初始图像分割模型，直至损失值小于设定阈值。

根据本公开的一个或多个实施例，本公开提供了一种图像分割装置，包括：

融合模块，设置为融合原始图像对应的视觉特征和描述语言对应的文本特征，得到多模态特征，其中，所述描述语言用于指定所述原始图像中待分割的目标物体；视觉区域确定模块，设置为根据所述多模态特征对应的图像确定所述目标物体的视觉区域，并将所述视觉区域对应的图像记为响应热度图；分割结果确定模块，设置为根据所述多模态特征对应的图像和所述响应热度图，确定所述目标物体的分割结果。

根据本公开的一个或多个实施例，本公开提供了一种电子设备，包括：

一个或多个处理器；存储器，设置为存储一个或多个程序；当所述一个或多个程序被所述一个或多个处理器执行时实现如本公开任一实施例提供的图像分割方法。

根据本公开的一个或多个实施例，本公开提供了一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现如本公开任一实施例提供

的图像分割方法。

此外，虽然采用特定次序描绘了多个操作，但是这不应当理解为要求这些操作以所示出的特定次序或以顺序次序执行来执行。在一定环境下，多任务和并行处理可能是有利的。同样地，虽然在上面论述中包含了多个实现细节，但是这些不应当被解释为对本公开的范围的限制。在单独的实施例的上下文中描述的一些特征还可以组合地实现在单个实施例中。相反地，在单个实施例的上下文中描述的多种特征也可以单独地或以任何合适的子组合的方式实现在多个实施例中。

权利要求书

1、一种图像分割方法，包括：

融合原始图像对应的视觉特征和描述语言对应的文本特征，得到多模态特征，其中，所述描述语言用于指定所述原始图像中待分割的目标物体；

根据所述多模态特征对应的图像确定所述目标物体的视觉区域，并将所述视觉区域对应的图像记为响应热度图；

根据所述多模态特征对应的图像和所述响应热度图，确定所述目标物体的分割结果。

2、根据权利要求1所述的方法，其中，所述根据所述多模态特征对应的图像确定所述目标物体的视觉区域，包括：

对所述多模态特征对应的图像进行相关性滤波，得到所述目标物体的视觉区域。

3、根据权利要求2所述的方法，其中，所述对所述多模态特征对应的图像进行相关性滤波，得到所述目标物体的视觉区域，包括：

根据所述文本特征确定卷积核；

根据所述卷积核对所述多模态特征对应的图像进行卷积操作，得到所述目标物体的视觉区域。

4、根据权利要求1所述的方法，其中，所述视觉特征包括在至少两个分辨率下分别从所述原始图像中提取的视觉特征。

5、根据权利要求4所述的方法，其中，所述融合原始图像对应的视觉特征和描述语言对应的文本特征，得到多模态特征，包括：

按照分辨率的大小顺序排列所述至少两个视觉特征，得到排序结果；通过映射矩阵将所述文本特征映射到所述排序结果中的第一分辨率对应的第一视觉特征所在的特征空间，其中，所述第一分辨率的值最小；

拼接所述第一视觉特征和映射后的文本特征，得到第一拼接特征；

对所述第一拼接特征进行上采样，并将上采样后的第一拼接特征与所述排序结果中的第二分辨率对应的第二视觉特征进行拼接，得到第二拼接特征，循环执行上采样和拼接的操作，直至将上采样后的拼接特征与所述排序结果中的分辨率最大的视觉特征拼接，得到所述多模态特征为止，其中，所述第二分辨率大于所述第一分辨率，所述第二分辨率为除所述第一分辨率之外最小的分辨率。

6、根据权利要求1-5中任一项所述的方法，其中，所述根据所述多模态特

征对应的图像和所述响应热度图，确定所述目标物体的分割结果，包括：

将所述多模态特征对应的图像和所述响应热度图输入目标图像分割模型，获取所述目标图像分割模型输出的结果，作为所述目标物体的分割结果。

7、根据权利要求6所述的方法，其中，所述目标图像分割模型的训练过程如下：

获取样本图像和样本描述语言，并提取所述样本图像的样本视觉特征和所述样本描述语言的样本文本特征，其中，所述样本描述语言用于指定所述样本图像中待分割的样本目标物体；

融合所述样本视觉特征和所述样本文本特征，得到样本多模态特征；

根据所述样本多模态特征对应的图像确定所述样本目标物体的样本视觉区域，并将所述样本视觉区域对应的图像记为样本响应热度图；

根据所述样本多模态特征对应的图像和所述样本响应热度图训练初始图像分割模型，得到所述目标图像分割模型。

8、根据权利要求7所述的方法，其中，所述根据所述样本多模态特征对应的图像和所述样本响应热度图训练初始图像分割模型，得到所述目标图像分割模型，包括：

将所述样本多模态特征对应的图像和所述样本响应热度图输入所述初始图像分割模型，得到所述样本多模态特征对应的图像和所述样本响应热度图的多个第一卷积结果，其中，所述多个第一卷积结果由对所述样本多模态特征对应的图像和所述样本响应热度图在不同采样率下执行第一卷积操作得到；

拼接所述多个第一卷积结果，得到拼接结果；

对所述拼接结果进行第二卷积操作，得到第二卷积结果；

对所述第二卷积结果进行上采样，得到样本分割结果；

确定所述样本分割结果相对所述样本图像的真实分割结果的损失值；

在所述损失值小于设定阈值的情况下，停止训练所述初始图像分割模型，并将损失值小于设定阈值的图像分割模型作为所述目标图像分割模型；在所述损失值不小于设定阈值的情况下，继续训练所述初始图像分割模型，直至损失值小于所述设定阈值。

9、一种图像分割装置，包括：

融合模块，设置为融合原始图像对应的视觉特征和描述语言对应的文本特征，得到多模态特征，其中，所述描述语言用于指定所述原始图像中待分割的

目标物体；

视觉区域确定模块，设置为根据所述多模态特征对应的图像确定所述目标物体的视觉区域，并将所述视觉区域对应的图像记为响应热度图；

分割结果确定模块，设置为根据所述多模态特征对应的图像和所述响应热度图，确定所述目标物体的分割结果。

10、一种电子设备，包括：

至少一个处理器；

存储器，设置为存储至少一个程序；

当所述至少一个程序被所述至少一个处理器执行时实现如权利要求 1-8 中任一项所述的图像分割方法。

11、一种计算机可读存储介质，存储有计算机程序，其中，所述程序被处理器执行时实现如权利要求 1-8 中任一项所述的图像分割方法。

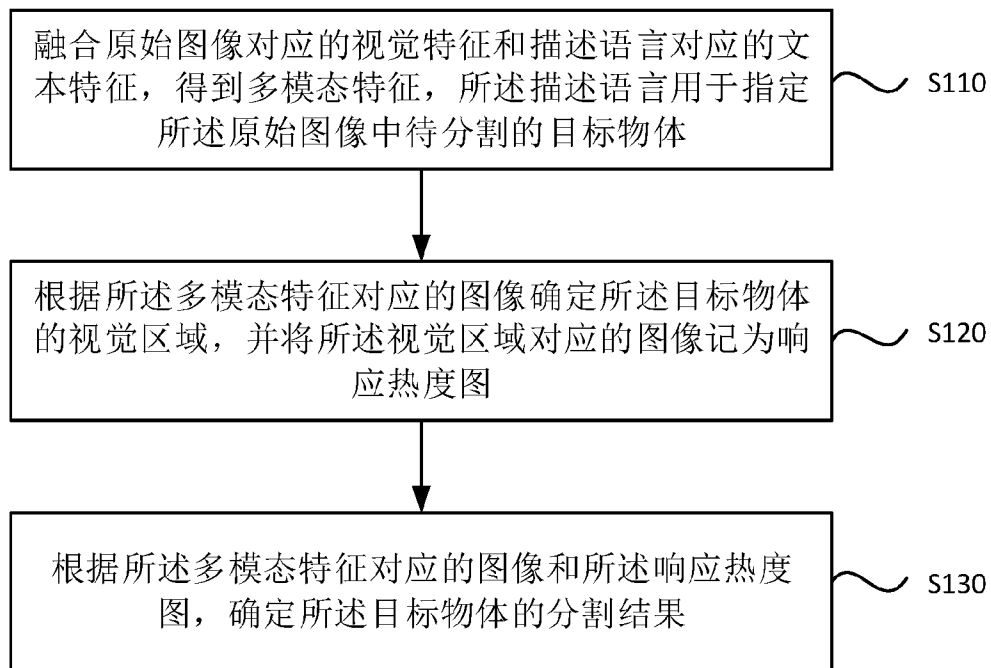


图 1

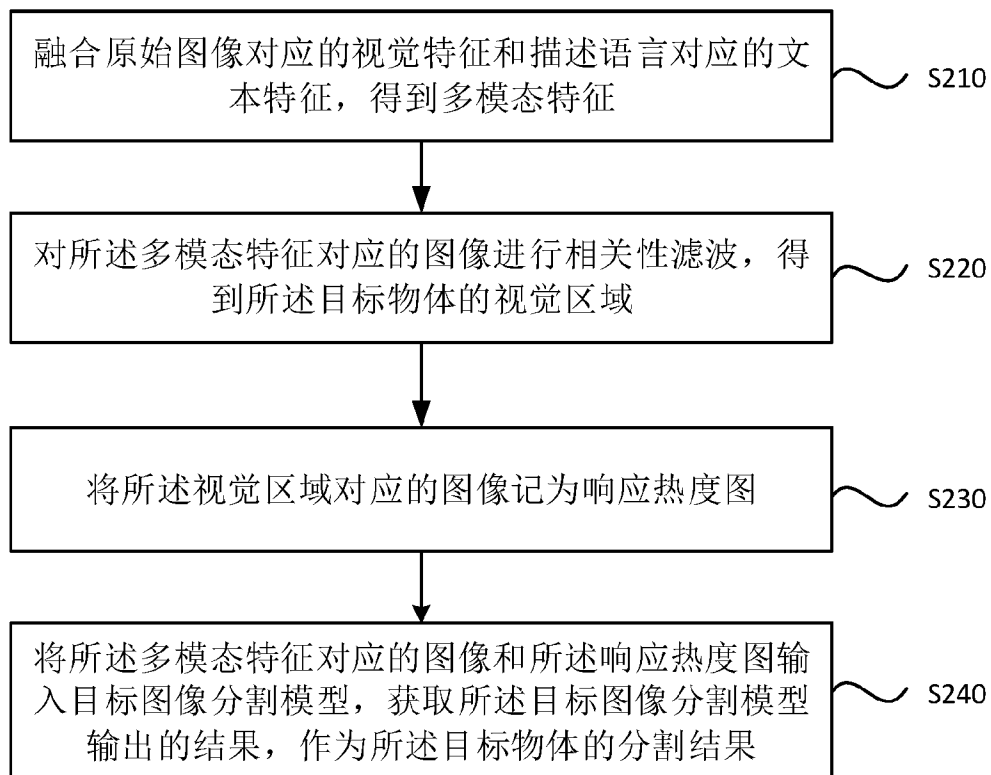


图 2

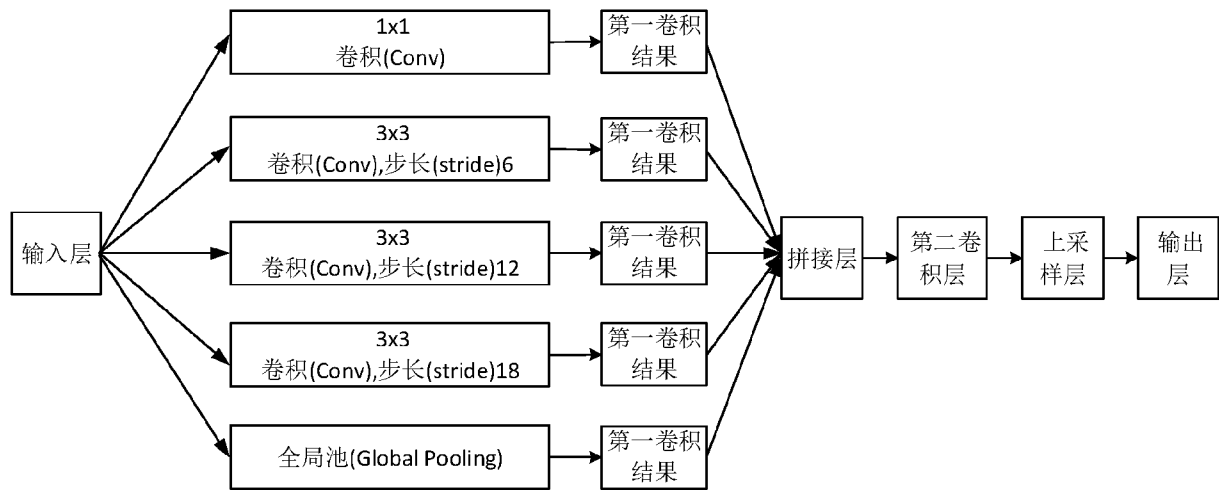


图 3

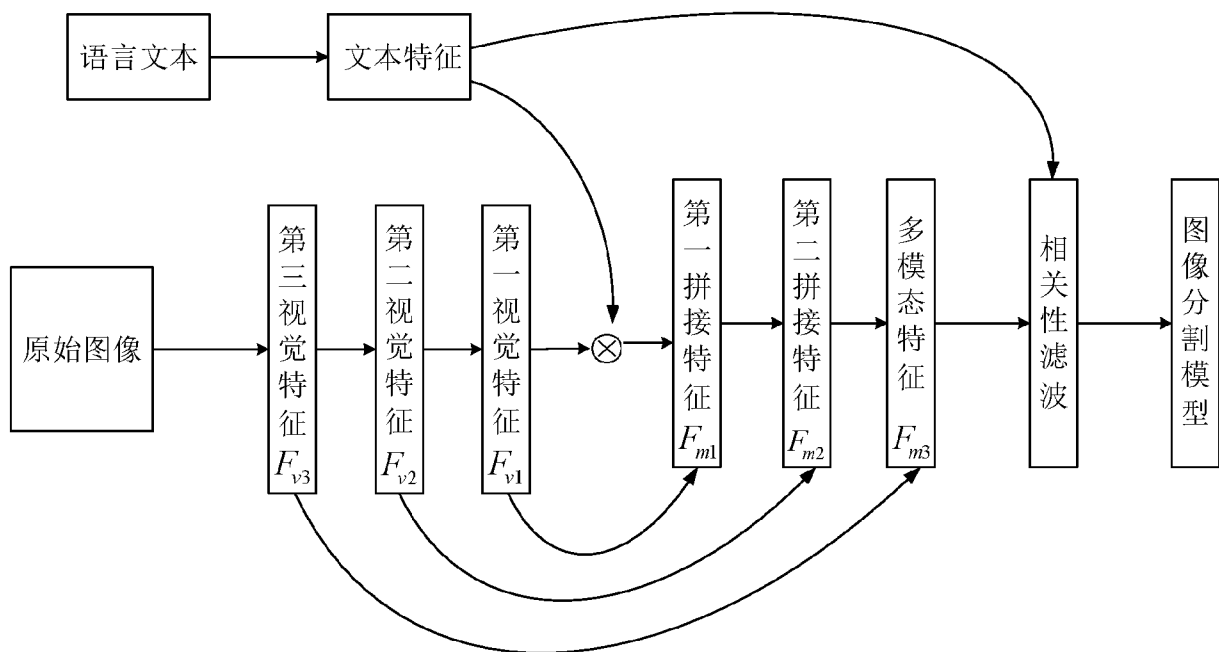


图 4

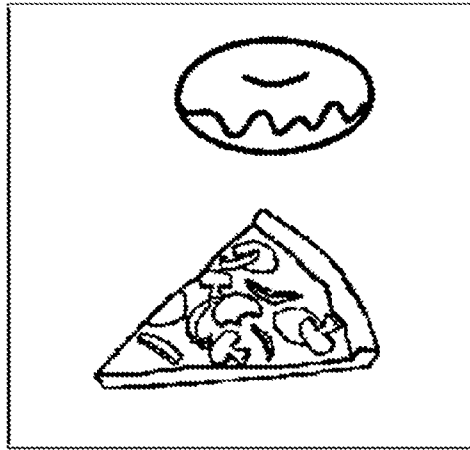


图 5

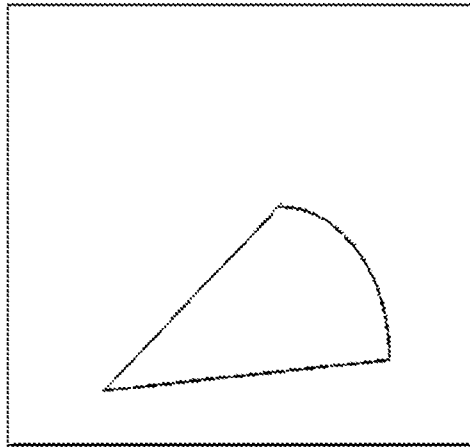


图 6

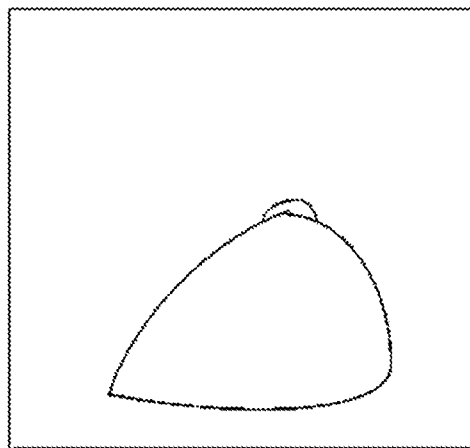


图 7

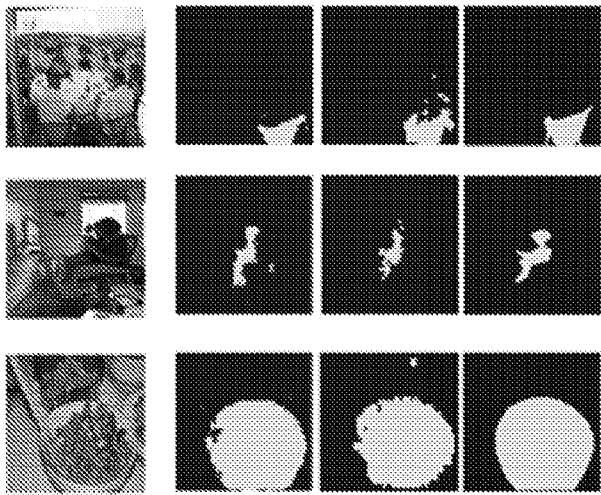


图 8

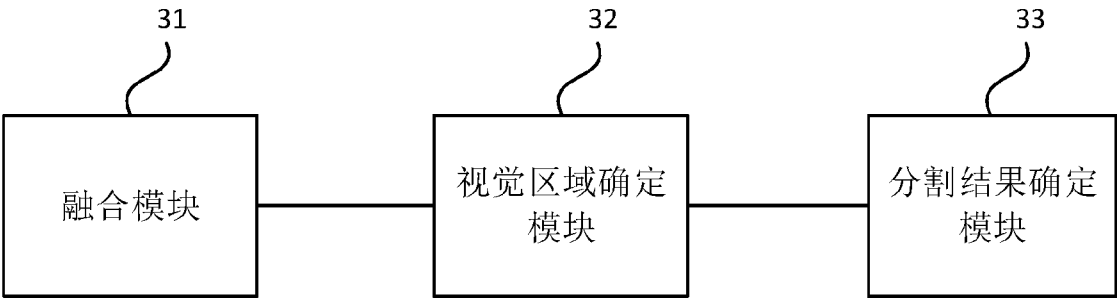


图 9

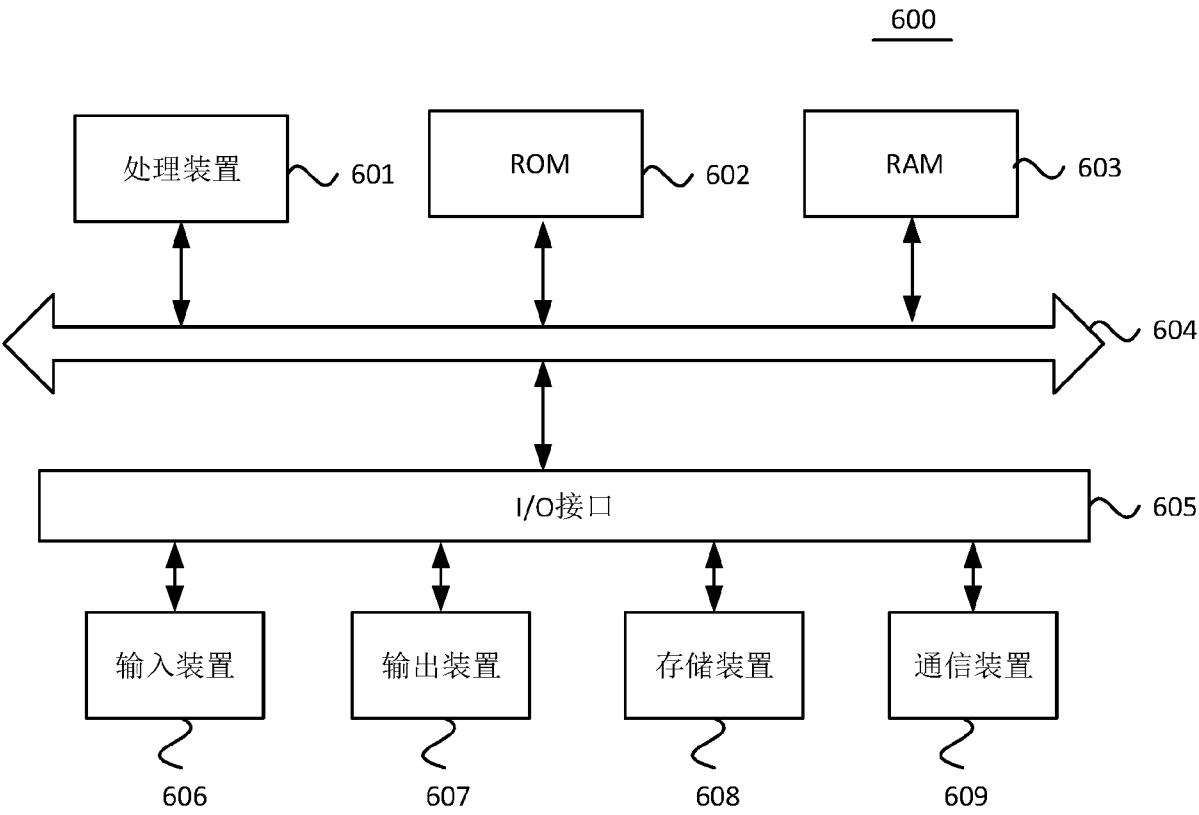


图 10

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2021/120815

A. CLASSIFICATION OF SUBJECT MATTER

G06T 7/11(2017.01)i; G06K 9/46(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)

G06T; 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)

CNPAT, EPODOC, WPI, CNKI, 百度, IEEE: 有竹居, 指代, 图像, 分割, 文本, 视觉, 融合, 多模态, 响应, 热度图, referring image segment, text, visual, fusion, multi-modal, response, hotmap

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 112184738 A (BEIJING YOUZHUJU NETWORK TECHNOLOGY CO. LTD.) 05 January 2021 (2021-01-05) claims 1-11	1-11
PX	JING, Ya et al. "Locate then Segment: A Strong Pipeline for Referring Image Segmentation" <i>arXiv:2103.16284v1</i> , 30 March 2021 (2021-03-30), pp. 1-10	1-11
X	HU, Ronghang et al. "Segmentation from Natural Language Expressions" <i>arXiv:1603.06180v1</i> , 20 March 2016 (2016-03-20), pp. 1-25	1-3, 6-11
Y	HU, Ronghang et al. "Segmentation from Natural Language Expressions" <i>arXiv:1603.06180v1</i> , 20 March 2016 (2016-03-20), pp. 1-25	4-5
Y	LUO, Gen et al. "Multi-task Collaborative Network for Joint Referring Expression Comprehension and Segmentation" <i>2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)</i> , 05 August 2020 (2020-08-05), pp. 10031-10040	4-5



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

14 December 2021

Date of mailing of the international search report

30 December 2021

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/CN2021/120815

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 110555337 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 10 December 2019 (2019-12-10) entire document	1-11
A	CN 110390289 A (SOOCHOW UNIVERSITY) 29 October 2019 (2019-10-29) entire document	1-11
A	US 2018268548 A1 (ADOBE SYSTEMS INCORPORATED) 20 September 2018 (2018-09-20) entire document	1-11
A	WANG, Fei et al. "Residual Attention Network for Image Classification" <i>2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)</i> , 09 November 2017 (2017-11-09), pp. 6450-6458	1-11

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2021/120815

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	112184738	A	05 January 2021	None			
CN	110555337	A	10 December 2019	None			
CN	110390289	A	29 October 2019	None			
US	2018268548	A1	20 September 2018	AU	2018200165	A1	04 October 2018
				AU	2018200165	B2	23 September 2021
				DE	102018000299	A1	20 September 2018
				CN	108573257	A	25 September 2018
				US	2019035083	A1	31 January 2019
				US	10410351	B2	10 September 2019
				GB	201800842	D0	07 March 2018
				GB	2560621	A	19 September 2018
				GB	2560621	B	26 August 2020
				US	10089742	B1	02 October 2018

国际检索报告

国际申请号

PCT/CN2021/120815

A. 主题的分类

G06T 7/11(2017.01)i; G06K 9/46(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)

G06T; 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))

CNPAT, EPDOC, WPI, CNKI, 百度, IEEE: 有竹居, 指代, 图像, 分割, 文本, 视觉, 融合, 多模态, 响应, 热度图, referring image segment, text, visual, fusion, multi-modal, response, heatmap

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 112184738 A (北京有竹居网络技术有限公司) 2021年1月5日 (2021 - 01 - 05) 权利要求1-11	1-11
PX	JING, Ya 等. "Locate then Segment: A Strong Pipeline for Referring Image Segmentation" arXiv:2103.16284v1, 2021年3月30日 (2021 - 03 - 30), 第1-10页	1-11
X	HU, Ronghang 等. "Segmentation from Natural Language Expressions" arXiv:1603.06180v1, 2016年3月20日 (2016 - 03 - 20), 第1-25页	1-3、6-11
Y	HU, Ronghang 等. "Segmentation from Natural Language Expressions" arXiv:1603.06180v1, 2016年3月20日 (2016 - 03 - 20), 第1-25页	4-5
Y	LUO, Gen 等. "Multi-task Collaborative Network for Joint Referring Expression Comprehension and Segmentation" 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020年8月5日 (2020 - 08 - 05), 第10031-10040页	4-5

☒ 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

2021年12月14日

国际检索报告邮寄日期

2021年12月30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

授权官员

刘梦瑶

电话号码 86-(10)-53961396

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	CN 110555337 A (腾讯科技深圳有限公司) 2019年12月10日 (2019 - 12 - 10) 全文	1-11
A	CN 110390289 A (苏州大学) 2019年10月29日 (2019 - 10 - 29) 全文	1-11
A	US 2018268548 A1 (ADOBE SYSTEMS INCORPORATED) 2018年9月20日 (2018 - 09 - 20) 全文	1-11
A	WANG, Fei 等. "Residual Attention Network for Image Classification" 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017年11月9日 (2017 - 11 - 09), 第6450-6458页	1-11

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2021/120815

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	112184738	A	2021年1月5日	无			
CN	110555337	A	2019年12月10日	无			
CN	110390289	A	2019年10月29日	无			
US	2018268548	A1	2018年9月20日	AU	2018200165	A1	2018年10月4日
				AU	2018200165	B2	2021年9月23日
				DE	102018000299	A1	2018年9月20日
				CN	108573257	A	2018年9月25日
				US	2019035083	A1	2019年1月31日
				US	10410351	B2	2019年9月10日
				GB	201800842	D0	2018年3月7日
				GB	2560621	A	2018年9月19日
				GB	2560621	B	2020年8月26日
				US	10089742	B1	2018年10月2日