

(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 2 月 23 日 (23.02.2023)



(10) 国际公布号
WO 2023/020005 A1

- (51) 国际专利分类号:
G06F 16/583 (2019.01) **G06V 10/40** (2022.01)
G06K 9/62 (2022.01)
- (21) 国际申请号: PCT/CN2022/089626
- (22) 国际申请日: 2022 年 4 月 27 日 (27.04.2022)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202110945344.X 2021年8月17日 (17.08.2021) CN
- (71) 申请人: 北京百度网讯科技有限公司 (BEIJING BAIDU NETCOM SCIENCE TECHNOLOGY CO., LTD.) [CN/CN]; 中国北京

市海淀区上地十街 10 号百度大厦 2 层, Beijing 100085 (CN)。

(72) 发明人: 陈玥蓉(CHEN, Yuerong); 中国北京市海淀区上地十街 10 号百度大厦 2 层, Beijing 100085 (CN)。 姚锟(YAO, Kun); 中国北京市海淀区上地十街 10 号百度大厦 2 层, Beijing 100085 (CN)。 孙逸鹏(SUN, Yipeng); 中国北京市海淀区上地十街 10 号百度大厦 2 层, Beijing 100085 (CN)。 韩钧宇(HAN, Junyu); 中国北京市海淀区上地十街 10 号百度大厦 2 层, Beijing 100085 (CN)。 刘经拓(LIU, Jingtuo); 中国北京市海淀区上地十街 10 号百度大厦 2 层, Beijing 100085 (CN)。

(74) 代理人: 北京市汉坤律师事务所(BEIJING HAN KUN LAW OFFICES); 中国北京市东城区东长安街 1 号东方广场 C1 座 9 层, Beijing 100738 (CN)。

(54) Title: NEURAL NETWORK MODEL TRAINING METHOD, IMAGE RETRIEVAL METHOD, DEVICE, AND MEDIUM

(54) 发明名称: 神经网络模型的训练方法、图像检索方法、设备和介质

100

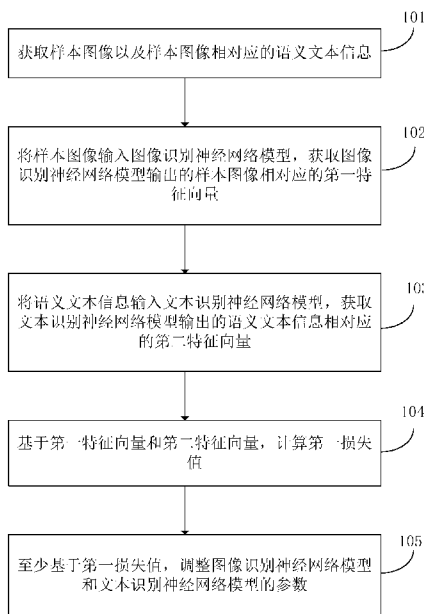


图 1

- 101 Obtain a sample image and semantic text information corresponding to the sample image
- 102 Input the sample image into an image recognition neural network model to obtain a first feature vector corresponding to the sample image and output by the image recognition neural network model
- 103 Input the semantic text information into a text recognition neural network model to obtain a second feature vector corresponding to the semantic text information and output by the text recognition neural network model
- 104 Calculate a first loss value on the basis of the first feature vector and a second feature vector
- 105 Adjust parameters of the image recognition neural network model and the text recognition neural network model at least on the basis of the first loss value

(57) Abstract: The present disclosure relates to the technical field of artificial intelligence, and in particular to the technical field of computer vision and deep learning, and provides an image retrieval method and apparatus, a device, and a medium. The present disclosure can be used in scenarios such as image processing and image recognition. The image retrieval method comprises: inputting an image to be detected into an image recognition neural network model to obtain an image feature vector output by the image recognition neural network model; and determining, on the basis of the image feature vector of the image to be detected, a first matching image

(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, 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, 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))。

set of said image from a database.

(57) 摘要: 本公开提供了一种图像检索方法、装置、设备和介质, 涉及人工智能技术领域, 具体为计算机视觉和深度学习技术领域, 可应用于图像处理和图像识别等场景。该图像检索方法包括: 将待检测图像输入图像识别神经网络模型, 获取图像识别神经网络模型输出的图像特征向量; 基于待检测图像的图像特征向量, 从数据库中确定待检测图像的第一匹配图像集。

神经网络模型的训练方法、图像检索方法、设备和介质

相关申请的交叉引用

本申请要求于 2021 年 8 月 17 日提交的中国专利申请 202110945344.X 的优先权，其全部内容通过引用整体结合在本申请中。

技术领域

本公开涉及人工智能技术领域，尤其涉及计算机视觉和深度学习技术，可应用于图像处理和图像识别等场景，具体涉及一种神经网络模型的训练方法、图像检索的方法、装置、电子设备、计算机可读存储介质和计算机程序产品。

背景技术

人工智能是研究使计算机来模拟人的某些思维过程和智能行为(如学习、推理、思考、规划等)的学科，既有硬件层面的技术也有软件层面的技术。人工智能硬件技术一般包括如传感器、专用人工智能芯片、云计算、分布式存储、大数据处理等技术;人工智能软件技术主要包括计算机视觉技术、语音识别技术、自然语言处理技术以及机器学习/深度学习、大数据处理技术、知识图谱技术等几大方向。

随着互联网的普及，网络购物的优点更加突出，日益成为一种重要的购物形式。与此同时，在购物软件中搜索产品对于用户来说是一个突出的需求。现在的方法除了通过关键词搜索外，利用图片搜索产品也是当前的主要方法。

在此部分中描述的方法不一定是之前已经设想到或采用的方法。除非另有指明，否则不应假定此部分中描述的任何方法仅因其包括在此部分中就被认为是现有技术。类似地，除非另有指明，否则此部分中提及的问题不应认为在任何现有技术中已被公认。

发明内容

本公开提供了一种神经网络模型的训练方法、图像检索方法、装置、电子设备、计算机可读存储介质和计算机程序产品。

根据本公开的一方面，提供了一种神经网络模型的训练方法，所述神经网络模型包括图像识别神经网络模型和文本识别神经网络模型，该方法包括：获取样本图像以及样本图像相对应的语义文本信息；将样本图像输入图像识别神经网络模型，获取图像识别神经网络模型输出的样本图像相对应的第一特征向量；将语义文本信息输入文本识别神经网络模型，获取文本识别神经网络模型输出的语义文本信息相对应的第二特征向量；基于第一特征向量和第二特征向量，计算第一损失值；以及至少基于第一损失值，调整图像识别神经网络模型和文本识别神经网络模型的参数。

根据本公开的另一方面，提供了一种基于神经网络模型的图像检索方法，该神经网络模型为通过上述的训练方法进行训练来得到的，该神经网络模型包括图像识别神经网络模型和文本识别神经网络模型，该方法包括：将待检测图像输入图像识别神经网络模型，获取图像识别神经网络模型输出的图像特征向量；以及基于待检测图像的图像特征向量，从数据库中确定待检测图像的第一匹配图像集。

根据本公开的另一方面，提供了一种神经网络模型的训练装置，神经网络模型包括图像识别神经网络模型和文本识别神经网络模型，该装置包括：第一获取单元，被配置用于获取样本图像以及样本图像相对应的语义文本信息；第二获取单元，被配置用于将样本图像输入图像识别神经网络模型，获取图像识别神经网络模型输出的样本图像相对应的第一特征向量；第三获取单元，被配置用于将语义文本信息输入文本识别神经网络模型，获取文本识别神经网络模型输出的语义文本信息相对应的第二特征向量；计算单元，被配置为基于第一特征向量和第二特征向量，计算第一损失值；以及调参单元，被配置为至少基于第一损失值，调整图像识别神经网络模型的参数。

根据本公开的另一方面，提供了一种基于神经网络模型的图像检索装置，该神经网络模型为通过上述的训练方法进行训练来得到的，该神经网络模型包括图像识别神经网络模型和文本识别神经网络模型，该装置包括：第一获取单元，被配置用于将待检测图像输入图像识别神经网络模型，获取图像识别神经网络模型输出的图像特征向量，其中图像特征提取网络模型为通过上

述的训练方法进行训练来得到的；以及第一确定单元，被配置用于基于待检测图像的图像特征向量，从数据库中确定待检测图像的第一匹配图像集。

根据本公开的另一方面，提供了一种电子设备，包括：至少一个处理器；以及与至少一个处理器通信连接的存储器；其中存储器存储有可被至少一个处理器执行的指令，这些指令被至少一个处理器执行，以使至少一个处理器能够执行上述图像检索方法或神经网络模型的训练方法。

根据本公开的另一方面，提供了一种存储有计算机指令的非瞬时计算机可读存储介质，其中，计算机指令用于使计算机执行上述图像检索方法或神经网络模型的训练方法。

根据本公开的另一方面，提供了一种计算机程序产品，包括计算机程序，其中，计算机程序在被处理器执行时实现上述图像检索方法或神经网络模型的训练方法。

根据本公开的一个或多个实施例，利用样本图像以及与样本图像对应的语义文本信息对图像识别神经网络和文本识别神经网络进行训练，能够使得图像识别神经网络学习到图像的语义信息。在实际应用的时候，利用训练好的图像识别神经网络来获取图像特征，由于该图像特征包含了图像的语义信息，从而能够提高图像检索结果的准确性。

应当理解，本部分所描述的内容并非旨在标识本公开的实施例的关键或重要特征，也不用于限制本公开的范围。本公开的其它特征将通过以下的说明书而变得容易理解。

附图说明

附图示例性地示出了实施例并且构成说明书的一部分，与说明书的文字描述一起用于讲解实施例的示例性实施方式。所示出的实施例仅出于例示的目的，并不限制权利要求的范围。在所有附图中，相同的附图标记指代类似但不一定相同的要素。

图 1 示出了根据本公开的实施例的神经网络模型的训练方法的流程图；

图 2 示出了根据本公开的实施例的另一种神经网络模型的训练方法的流程图；

图 3 示出了根据本公开的实施例的另一种神经网络模型的训练方法的流程图；

图 4 示出了根据本公开的实施例的图像检索方法的流程图；

图 5 示出了根据本公开的实施例的另一种图像检索方法的流程图；

图 6 示出了根据本公开的实施例的另一种图像检索方法的流程图；

图 7 示出了根据本公开的实施例的另一种图像检索方法的流程图；

图 8 示出了根据本公开的实施例的另一种图像检索方法的流程图；

图 9 示出了根据本公开的实施例的神经网络模型的训练装置的结构框图；

图 10 示出了根据本公开示例性实施例的图像检索装置的结构框图；以及

图 11 示出了能够用于实现本公开的实施例的示例性电子设备的结构框图。

具体实施方式

以下结合附图对本公开的示范性实施例做出说明，其中包括本公开实施例的各种细节以助于理解，应当将它们认为仅仅是示范性的。因此，本领域普通技术人员应当认识到，可以对这里描述的实施例做出各种改变和修改，而不会背离本公开的范围。同样，为了清楚和简明，以下的描述中省略了对公知功能和结构的描述。

在本公开中，除非另有说明，否则使用术语“第一”、“第二”等来描述各种要素不意图限定这些要素的位置关系、时序关系或重要性关系，这种术语只是用于将一个元件与另一元件区分开。在一些示例中，第一要素和第二要素可以指向该要素的同一实例，而在某些情况下，基于上下文的描述，它们也可以指代不同实例。

在本公开中对各种所述示例的描述中所使用的术语只是为了描述特定示例的目的，而并非旨在进行限制。除非上下文另外明确地表明，如果不特意限定要素的数量，则该要素可以是一个也可以是多个。此外，本公开中所使用的术语“和/或”涵盖所列出的项目中的任何一个以及全部可能的组合方式。

相关技术中，单纯使用图像特征的检索技术，无法充分利用图片本身可附加的文本信息，所以无法更好的学习图像中各模块的相关性。文本信息本身是一种很全面、立体的监督信息。缺失了文本信息的参与，神经网络模型对于复杂角度、形状多变的物体（例如服饰）等目标的表征能力相对较弱。

为解决上述问题，在神经网络模型训练过程中，结合了图像识别神经网络和文本识别神经网络，分别输入样本图像以及与样本图像对应的语义文本信息，目的是为了让图像识别神经网络能够更好的学习图像的语义特征。在实际应用的时候，只使用训练好的图像识别神经网络，再与数据库存储的特征向量进行相似度计算。这样能更好地学习图像的语义特征，输出更准确地结果。

下面将结合附图详细描述本公开的实施例。

根据本公开的一方面，提供了一种神经网络模型的训练方法。如图 1 所示，训练方法 100 可以包括：步骤 101、获取样本图像以及样本图像相对应的语义文本信息；步骤 102、将样本图像输入图像识别神经网络模型，获取图像识别神经网络模型输出的样本图像相对应的第一特征向量；步骤 103、将语义文本信息输入文本识别神经网络模型，获取文本识别神经网络模型输出的语义文本信息相对应的第二特征向量；步骤 104、基于第一特征向量和第二特征向量，计算第一损失值；步骤 105、至少基于第一损失值，调整图像识别神经网络模型和文本识别神经网络模型的参数。由此，通过利用样本图像以及与样本图像对应的语义文本信息对图像识别神经网络和文本识别神经网络进行训练，能够使得图像识别神经网络学习到图像的语义信息。

在一个示例中，样本图像可以是物品的图像，语义文本信息是能够反映图片内容的并且更丰富的信息。例如，样本图像中包括鼠标、键盘和显示器，其对应的语义文本信息可以包括无线鼠标键盘、液晶显示器以及品牌名称、颜色等。再例如，样本图像可以为上衣图像，此时样本图像对应的语义文本信息可以包括品牌名称、上衣、短袖、运动装等等。

根据一些实施例，在将样本图像输入图像神经网络模型之前，可以对样本图像进行预处理，然后将预处理后的样本图像输入图像识别神经网络模型，获得与样本图像相对应的第一特征向量。预处理可以包括对样本图像进行尺寸调整、角度矫正等。

根据一些实施例，在将语义文本信息输入文本识别神经网络模型之前，可以获取语义文本信息中的至少一个关键字，并将语义文本信息相对应的至少一个关键字输入文本识别神经网络模型，获得与语义文本信息相对应的第一特征向量。

示例性的，可以但不限于计算第一特征向量和第二特征向量之间的欧式距离，并将该欧式距离确定为第一损失值，以基于第一损失值调整模型参数，使得第一损失值尽可能得小。

根据一些实施例，样本图像包括样本对象。如图 2 所示，神经网络模型的训练方法 200 可以包括：

步骤 201、获取样本图像以及样本图像对应的语义文本信息。

步骤 202、标记样本图像中包围样本对象的真实边界框以及样本对象的真实标签。

根据一些实施例，可以获取语义文本信息的至少一个关键字，将至少一个关键字中的一个或多个关键字作为真实标签。

在一个示例中，获取猫的样本图像以及与该样本图像的对应的语义文本信息，其中，该语义文本信息可以包括猫、猫的品种、猫的颜色、猫的特征（例如短腿、短耳）等等关键字。在这种场景下，真实标签可以为语义文本信息所包括的关键字“猫”

步骤 203、将样本图像输入图像识别神经网络模型，获取图像识别神经网络模型输出的样本图像对应的第一特征向量、预测边界框和样本对象的预测标签。

步骤 204、基于预测边界框、预测标签、真实边界框和真实标签，计算第二损失值。

根据一些实施例，可以基于真实边界框和预测边界框的交并比计算损失值，也可以基于真实边界框和预测边界框的中心距离计算损失值，还可以通过其他方式计算损失值，在此不做限定。

步骤 205、将语义文本信息输入文本识别神经网络模型，获取文本识别神经网络模型输出的语义文本信息相对应的第二特征向量。

步骤 206、基于第一特征向量和第二特征向量，计算第一损失值。

步骤 207、基于第一损失值和第二损失值，调整图像识别神经网络模型的参数。

步骤 208、基于第一损失值，调整文本识别神经网络模型的参数。

上述训练方法中，图像神经网络模型还具有目标检测的功能，使得模型能够更好得提取样本对象的图像特征。

图 2 中的步骤 201、步骤 205 和步骤 206 分别与图 1 中的步骤 101、步骤 103 和步骤 104 类似，在此不做赘述。

根据一些实施例，如图 3 所示，神经网络模型的训练方法 300 可以包括：

步骤 301、获取样本图像以及样本图像相对应的语义文本信息。

步骤 302、标记样本图像中样本对象的真实标签。

步骤 303、确定样本图像中样本对象所处的前景区域。

步骤 304、对样本图像进行裁剪，以得到前景图像。

步骤 305、将前景图像输入图像识别神经网络模型，获得图像识别神经网络模型输出的样本图像对应的第一特征向量和样本对象的预测标签。

步骤 306、基于预测标签和真实标签，计算第三损失值。

根据一些实施例，将预测标签和真实标签在一个空间进行转化，分别得到各自的特征向量，并由此计算两个特征向量的欧氏距离，最后得到第三损失值。

步骤 307、将语义文本信息输入文本识别神经网络模型，获取文本识别神经网络模型输出的语义文本信息相对应的第二特征向量。

步骤 308、基于第一特征向量和第二特征向量，计算第一损失值。

步骤 309、基于第一损失值和第三损失值，调整图像识别神经网络模型的参数。

根据一些实施例，基于步骤 308 中得到的第一损失值与步骤 306 中表示预测标签和真实标签的损失值的第三损失值，其中损失值是用来表示预测值与真实值的差距。在训练神经网络时，可以通过不断的改变神经网络中所有的参数，使损失值不断减小，从而训练出更准确的神经网络模型。

步骤 310、基于第一损失值，调整文本神经网络识别神经网络模型的参数。

图 3 中步骤 301、步骤 307 和步骤 308 与图 1 中的步骤 101、步骤 103 和步骤 104 类似，在此不做赘述。

根据本公开的另一方面，还提供了一种基于神经网络模型的图像检索方法，所述神经网络模型为上述的训练方法进行训练来得到的，所述神经网络模型包括图像识别神经网络模型和文本识别神经网络模型。如图 4 所示，图像检索方法 400 可以包括：

步骤 401、将待检测图像输入图像识别神经网络模型，获取图像识别神经网络模型输出的图像特征向量。

上述方案中，利用训练好的图像识别神经网络来获取图像特征，由于该图像特征包含了图像的语义信息，从而能够提高图像检索结果的准确性。

在一个示例中，图像识别神经网络模型可以通过引入卷积神经网络中常用的层次化构建方式来构建的层次化 transformer 模型。transformer 模型结合 CNN 与 self-attention 结构，神经网络中前层使用具有滑动窗口机制卷积神经网络提取 low level 特征，深层采用具有自注意力机制的 transformer 模型提取 high level 特征，在图像检索任务上效果提升非常明显。

步骤 402、基于待检测图像的图像特征向量，从数据库中确定待检测图像的第一匹配图像集。

根据一些实施例，数据库中存储有多个图像各自相应的多个图像特征向量，分别计算待检测图像的图像特征向量与在数据库中存储的多个图像特征向量中的每一个图像特征向量之间的欧式距离。可以于相应的欧式距离，确定数据库中与待检测图像匹配的多个图像。

根据一些实施例，其中数据库中存储的多个图像特征向量也可以通过图 1-图 3 展示的方法 100、方法 200 和方法 300 中任一方法训练得到的图像识别神经网络模型来获得。

根据一些实施例，待检测图像包括目标对象，如图 5 所示，图像检索方法 500 可以包括：

步骤 501、将待检测图像输入图像识别神经网络模型，获取图像识别神经网络模型输出的图像特征向量、目标对象的目标边界框和目标对象的目标标签。

步骤 502、基于待检测图像的图像特征向量，从数据库中确定待检测图像的第一匹配图像集。

步骤 503、将目标标签输入文本识别神经网络模型，获取文本识别神经网络模型输出的文本特征向量。

步骤 504、基于文本特征向量，从第一匹配图像集中确定待检测图像的至少一个匹配图像。

图 5 中的步骤 502 与图 4 中的步骤 402 类似，在此不作赘述。

根据一些实施例，分别计算文本识别神经网络模型输出的文本特征向量与步骤 502 中确定的第一匹配图像集所包括的各图像相对应的文本特征向量之间的欧式距离，从第一匹配图像集中确定输入的待检测图像的至少一个匹配图像。

由此，待检测图像通过图像识别神经网络模型确定第一匹配图像集后，再利用与待检测图像相关的语音文本信息通过与图像识别神经网络模型一起训练的文本神经网络模型进行进一步的确定，最终确定的图像与输入的待检测图像的匹配程度更高。

根据一些实施例，如图 6 所示，图像检索方法 600 包括：

步骤 601、将待检测图像输入图像识别神经网络模型，获取图像识别神经网络模型输出的图像特征向量、目标对象的目标边界框和目标对象的目标标签。

步骤 602、基于待检测图像的图像特征向量，从数据库中确定待检测图像的第一匹配图像集。

步骤 603、将目标标签输入文本识别神经网络模型，获取文本识别神经网络模型输出的文本特征向量。

步骤 604、基于文本特征向量，从数据库中确定待检测图像的第二匹配图像集。

根据一些实施例，计算文本特征向量与数据库中存储的文本特征向量欧氏距离，筛选出符合要求的图像，这些图像共同组成与待检测图像的第二匹配图像集。

根据一些实施例，其中数据库中存储的多个文本特征向量也可以为通过图 1-图 3 展示的方法 100、方法 200 和方法 300 中任一方法训练得到的神经网络模型中的文本识别神经网络模型来获得。

步骤 605、基于第一匹配图像集和第二匹配图像集，确定待检测图像的至少一个匹配图像。

根据一些实施例，基于图像特征向量和文本特征向量分别与数据库中的数据进行比对，计算图像特征向量和数据库中的图像特征向量的欧式距离以及文本特征向量和数据库中的文本特征向量的欧式距离，最终通过两次比对结果来确定最终的匹配图像。

在一个示例中，将两次比对结果都包括的图像作为匹配图像或者将两次比对结果按照相似度进行排序，取分数最高的若干图像作为最终的匹配图像。

图 6 中的步骤 601-步骤 603 与图 5 中的步骤 501-步骤 503 相似，在此不做赘述。

根据一些实施例，如图 7 所示，图像检索方法 700 可以包括：

步骤 701、确定待检测图像中目标对象所处的前景区域。

步骤 702、对待检测图像进行裁剪，以得到前景图像。

步骤 703、将前景图像作为图像识别神经网络模型的输入，获取图像识别神经网络模型输出的图像特征向量和目标对象的目标标签。

步骤 704、基于待检测图像的图像特征向量，从数据库中确定待检测图像的第一匹配图像集。

步骤 705、将目标标签输入文本识别神经网络模型，获取文本识别神经网络模型输出的文本特征向量。

步骤 706、基于文本特征向量，从第一匹配图像集中确定待检测图像的至少一个匹配图像。

图 7 中的步骤 704-步骤 706 与图 5 中的步骤 502-步骤 504 相似，在此不做赘述。

根据一些实施例，如图 8 所示，图像检索方法 800 可以包括：

步骤 801、确定待检测图像中目标对象所处的前景区域。

步骤 802、对检测图像进行裁剪，以得到前景图像。

步骤 803、将前景图像作为图像识别神经网络模型的输入，获取图像识别神经网络模型输出的图像特征向量和目标对象的目标标签。

步骤 804、基于待检测图像的图像特征向量，从数据库中确定待检测图像的第一匹配图像集。

步骤 805、将目标标签输入文本识别神经网络模型，获取文本识别神经网络模型输出的文本特征向量。

步骤 806、基于文本特征向量，从数据库中确定待检测图像的第二匹配图像集。

步骤 807、基于第一匹配图像及和第二匹配图像集，确定待检测图像的至少一个匹配图像。

利用本公开实施例中的图像检索方法能够充分利用图像的语义文本信息，提升图像检索结果的准确性，提升用户体验。

例如，目标图像为上衣图像，该上衣图像的语义信息可以包括品牌名称、上衣、短袖、运动装等等。利用现有的检索方法检索到的结果包括相同品牌图像标志的下装，而利用本公开实施例中的图像检索方法能够充分利用图像对应的语义文本信息，检索结果可以仅包括相同品牌图像标志的运动款短袖上衣，大大提升图像检索结果的准确性，提升用户体验。

图 8 中的步骤 804-步骤 807 与图 6 中的步骤 602-步骤 605 相似，在此不做赘述。

根据本公开的另一方面，还提供了一种神经网络模型的训练装置 900。如图 9 所示，神经网络模型训练装置 900 包括：第一获取单元 901，被配置用于获取样本图像以及样本图像相对应的语义文本信息；第二获取单元 902，被配置用于将样本图像输入图像识别神经网络模型，获取图像识别神经网络模型输出的样本图像相对应的第一特征向量；第三获取单元 903，被配置用于将语义文本信息输入文本识别神经网络模型，获取文本识别神经网络模型输出的语义文本信息相对应的第二特征向量；计算单元 904，被配置为基于第一特征向量和第二特征向量，计算第一损失值；以及调参单元 905，被配置为至少基于第一损失值，调整图像识别神经网络模型的参数。

根据一些实施例，其中训练装置 900 还包括：第一标记单元，被配置为标记样本图像中包围样本对象的真实边界框以及样本对象的真实标签。

根据一些实施例，其中计算单元 904 被进一步配置为基于预测边界框、预测标签、真实边界框和真实标签，计算第二损失值。

根据一些实施例，其中调参单元 905 包括：第一调参子单元，被配置为基于第一损失值和第二损失值，调整图像识别神经网络模型的参数；以及第二调参子单元，被配置为基于第一损失值，调整文本识别神经网络模型的参数。

根据一些实施例，其中第一标记单元被进一步配置为获取语义文本信息的至少一个关键字，将至少一个关键字中的一个或多个关键字作为真实标签。

根据一些实施例，其中训练装置 900 还包括：确定单元被配置为在将样本图像输入图像识别神经网络模型之前，确定样本图像中样本对象所处的前景区域；以及裁剪单元，被配置为对样本图像进行裁剪，以得到前景图像，并将前景图像作为图像识别神经网络模型的输入。

根据一些实施例，其中训练装置 900 还包括：第二标记单元，被配置为标记样本图像中样本对象的真实标签；

根据一些实施例，其中计算单元 904 被进一步配置为基于预测标签和真实标签，计算第三损失值。

根据一些实施例，其中调参单元 905 包括：第三调参子单元，被配置为基于第一损失值和第三损失值，调整图像识别神经网络模型的参数；以及第四调参子单元，被配置为基于第一损失值，调整文本识别神经网络模型的参数。

训练装置 900 的单元 901 至单元 905 的操作和前面描述的步骤 101 至步骤 105 的操作类似，在此不做赘述。

根据本公开的另一方面，还提供了一种基于神经网络模型的图像检索装置 1000，该神经网络模型为通过上述的训练方法进行训练来得到的，该神经网络模型包括图像识别神经网络模型和文本识别神经网络模型。如图 10 所示，图像检索装置 1000 包括：第一获取单元 1001，被配置用于将待检测图像输入图像识别神经网络模型，获取图像识别神经网络模型输出的图像特征向量；以及第一确定单元 1002，被配置用于基于待检测图像的图像特征向量，从数据库中确定待检测图像的第一匹配图像集。

根据一些实施例，其中待检测图像包括目标对象，图像识别神经网络模型的输出还包括包围目标对象的目标边界框和目标对象的目标标签。

根据一些实施例，其中图像检索装置 1000 还包括第二获取单元，被配置用于将目标标签输入文本识别神经网络模型，获取文本识别神经网络模型输出的文本特征向量；以及第二确定单元，被配置用于基于文本特征向量，从第一匹配图像集中确定待检测图像的至少一个匹配图像。

根据一些实施例，其中图像检索装置 1000 还包括第三获取单元，被配置用于将目标标签输入文本识别神经网络模型，获取文本识别神经网络模型输出的文本特征向量；第三确定单元，被配置用于基于文本特征向量，从数据库中确定待检测图像的第二匹配图像集；以及第四确定单元，被配置用于基于第一匹配图像集和第二匹配图像集，确定待检测图像的至少一个匹配图像。

根据一些实施例，其中待检测图像包括目标对象，并且图像检索装置 1000 还包括：第五确定单元，被配置用于在将待检测图像输入图像识别神经网络模型之前，确定待检测图像中目标对象所处的前景区域；裁剪单元，被配置用于对待检测图像进行裁剪，以得到前景图像，并将前景图像作为图像识别神经网络模型的输入。

根据一些实施例，其中图像识别神经网络模型的输出还包括目标对象的目标标签。

根据一些实施例，其中图像检索装置 1000 还包括第四获取单元，被配置用于将目标标签输入文本识别神经网络模型，获取文本识别神经网络模型输出的文本特征向量；以及第六确定单元，被配置用于基于文本特征向量，从第一匹配图像集中确定待检测图像的至少一个匹配图像。

根据一些实施例，其中，图像识别神经网络模型的输出还包括目标对象的目标标签，并且图像检索装置 1000 还包括第五获取单元，被配置用于将目标标签输入文本识别神经网络模型，获取文本识别神经网络模型输出的文本特征向量；第七确定单元，被配置用于基于文本特征向量，从数据库中确定待检测图像的第二匹配图像集；以及第八确定单元，被配置用于基于第一匹配图像集和第二匹配图像集，确定待检测图像的至少一个匹配图像。

图像检索装置 1000 的单元 1001 和单元 1002 的操作和前面描述的步骤 401 和步骤 405 的操作类似，在此不做赘述。

本公开的技术方案中，所涉及的用户个人信息的获取，存储和应用等，均符合相关法律法规的规定，且不违背公序良俗。

根据本公开的实施例，还提供了一种电子设备、一种可读存储介质和一种计算机程序产品。

参考图 11，现将描述可以作为本公开的服务器或客户端的电子设备的结构框图，其是可以应用于本公开的各方面的硬件设备的示例。电子设备旨在表示各种形式的数字电子的计算机设备，诸如，膝上型计算机、台式计算机、工作台、个人数字助理、服务器、刀片式服务器、大型计算机、和其它适合的计算机。电子设备还可以表示各种形式的移动装置，诸如，个人数字处理、蜂窝电话、智能电话、可穿戴设备和其它类似的计算装置。本文所示的部件、它们的连接和关系、以及它们的功能仅仅作示例，并且不意在限制本文中描述的和/或者要求的本公开的实现。

如图 11 所示，设备 1100 包括计算单元 1101，其可以根据存储在只读存储器 (ROM) 1102 中的计算机程序或者从存储单元 1108 加载到随机访问存储器 (RAM) 1103 中的计算机程序，来执行各种适当的动作和处理。在 RAM 1103 中，还可存储设备 1100 操作所需的各种程序和数据。计算单元 1101、ROM 1102 以及 RAM 1103 通过总线 1104 彼此相连。输入/输出 (I/O) 接口 1105 也连接至总线 1104。

设备 1100 中的多个部件连接至 I/O 接口 1105，包括：输入单元 1106、输出单元 1107、存储单元 1108 以及通信单元 1109。输入单元 1106 可以是能向设备 1100 输入信息的任何类型的设备，输入单元 1106 可以接收输入的数字或字符信息，以及产生与电子设备的用户设置和/或功能控制有关的键信号输入，并且可以包括但不限于鼠标、键盘、触摸屏、轨迹板、轨迹球、操作杆、麦克风和/或遥控器。输出单元 1107 可以是能呈现信息的任何类型的设备，并且可以包括但不限于显示器、扬声器、视频/音频输出终端、振动器和/或打印机。存储单元 1108 可以包括但不限于磁盘、光盘。通信单元 1109 允许设备 1100 通过诸如因特网的计算机网络和/或各种电信网络与其他设备交换信息/数据，并且可以包括但不限于调制解调器、网卡、红外通信设备、

无线通信收发机和/或芯片组，例如蓝牙™设备、1302.11设备、WiFi设备、WiMax设备、蜂窝通信设备和/或类似物。

计算单元 1101 可以是各种具有处理和计算能力的通用和/或专用处理组件。计算单元 1101 的一些示例包括但不限于中央处理单元（CPU）、图形处理单元（GPU）、各种专用的人工智能（AI）计算芯片、各种运行机器学习模型算法的计算单元、数字信号处理器（DSP）、以及任何适当的处理器、控制器、微控制器等。计算单元 1101 执行上文所描述的各个方法和处理，例如神经网络模型训练方法和图像检索方法。例如，在一些实施例中，神经网络模型训练方法和图像检索方法可被实现为计算机软件程序，其被有形地包含于机器可读介质，例如存储单元 1108。在一些实施例中，计算机程序的部分或者全部可以经由 ROM 1102 和/或通信单元 1109 而被载入和/或安装到设备 1100 上。当计算机程序加载到 RAM 1103 并由计算单元 1101 执行时，可以执行上文描述的方法神经网络模型训练方法和图像检索方法的一个或多个步骤。备选地，在其他实施例中，计算单元 1101 可以通过其他任何适当的方式（例如，借助于固件）而被配置为执行方法神经网络模型训练方法和图像检索方法。

本文中以上描述的系统和技术和各种实施方式可以在数字电子电路系统、集成电路系统、场可编程门阵列（FPGA）、专用集成电路（ASIC）、专用标准产品（ASSP）、芯片上系统的系统（SOC）、负载可编程逻辑设备（CPLD）、计算机硬件、固件、软件、和/或它们的组合中实现。这些各种实施方式可以包括：实施在一个或者多个计算机程序中，该一个或者多个计算机程序可在包括至少一个可编程处理器的可编程系统上执行和/或解释，该可编程处理器可以是专用或者通用可编程处理器，可以从存储系统、至少一个输入装置、和至少一个输出装置接收数据和指令，并且将数据和指令传输至该存储系统、该至少一个输入装置、和该至少一个输出装置。

用于实施本公开的方法的程序代码可以采用一个或多个编程语言的任何组合来编写。这些程序代码可以提供给通用计算机、专用计算机或其他可编程数据处理装置的处理器或控制器，使得程序代码当由处理器或控制器执行时使流程图和/或框图中所规定的功能/操作被实施。程序代码可以完全在

机器上执行、部分地在机器上执行，作为独立软件包部分地在机器上执行且部分地在远程机器上执行或完全在远程机器或服务器上执行。

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更具体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、随机存取存储器（RAM）、只读存储器（ROM）、可擦除可编程只读存储器（EPROM 或快闪存储器）、光纤、便捷式紧凑盘只读存储器（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。

为了提供与用户的交互，可以在计算机上实施此处描述的系统和技术，该计算机具有：用于向用户显示信息的显示装置（例如，CRT（阴极射线管）或者 LCD（液晶显示器）监视器）；以及键盘和指向装置（例如，鼠标或者轨迹球），用户可以通过该键盘和该指向装置来将输入提供给计算机。其它种类的装置还可以用于提供与用户的交互；例如，提供给用户的反馈可以是任何形式的传感反馈（例如，视觉反馈、听觉反馈、或者触觉反馈）；并且可以用任何形式（包括声输入、语音输入或者、触觉输入）来接收来自用户的输入。

可以将此处描述的系统和技术实施在包括后台部件的计算系统（例如，作为数据服务器）、或者包括中间件部件的计算系统（例如，应用服务器）、或者包括前端部件的计算系统（例如，具有图形用户界面或者网络浏览器的用户计算机，用户可以通过该图形用户界面或者该网络浏览器来与此处描述的系统和技术实施方式交互）、或者包括这种后台部件、中间件部件、或者前端部件的任何组合的计算系统中。可以通过任何形式或者介质的数字数据通信（例如，通信网络）来将系统的部件相互连接。通信网络的示例包括：局域网（LAN）、广域网（WAN）和互联网。

计算机系统可以包括客户端和服务端。客户端和服务端一般远离彼此并且通常通过通信网络进行交互。通过在相应的计算机上运行并且彼此具有客

户端-服务器关系的计算机程序来产生客户端和服务器的关系。服务器可以是云服务器，也可以为分布式系统的服务器，或者是结合了区块链的服务器。

应该理解，可以使用上面所示的各种形式的流程，重新排序、增加或删除步骤。例如，本公开中记载的各步骤可以并行地执行、也可以顺序地或以不同的次序执行，只要能够实现本公开公开的技术方案所期望的结果，本文在此不进行限制。

虽然已经参照附图描述了本公开的实施例或示例，但应理解，上述的方法、系统和设备仅仅是示例性的实施例或示例，本发明的范围并不由这些实施例或示例限制，而是仅由授权后的权利要求书及其等同范围来限定。实施例或示例中的各种要素可以被省略或者可由其等同要素替代。此外，可以通过不同于本公开中描述的次序来执行各步骤。进一步地，可以以各种方式组合实施例或示例中的各种要素。重要的是随着技术的演进，在此描述的很多要素可以由本公开之后出现的等同要素进行替换。

权 利 要 求 书

1. 一种神经网络模型的训练方法,所述神经网络模型包括图像识别神经网络模型和文本识别神经网络模型,所述方法包括:

获取样本图像以及样本图像相对应的语义文本信息;

将所述样本图像输入图像识别神经网络模型,获取所述图像识别神经网络模型输出的所述样本图像相对应的第一特征向量;

将语义文本信息输入文本识别神经网络模型,获取所述文本识别神经网络模型输出的所述语义文本信息相对应的第二特征向量;

基于所述第一特征向量和第二特征向量,计算第一损失值; 以及

至少基于所述第一损失值,调整所述图像识别神经网络模型和文本识别神经网络模型的参数。

2. 如权利要求 1 所述的方法,其中,所述样本图像包括样本对象,所述图像识别神经网络模型的输出还包括预测边界框和所述样本对象的预测标签,

并且所述方法还包括:

标记所述样本图像中包围所述样本对象的真实边界框以及所述样本对象的真实标签;

基于所述预测边界框、所述预测标签、所述真实边界框和真实标签,计算第二损失值,

其中,至少基于所述第一损失值,调整所述图像识别神经网络模型和文本识别神经网络模型的参数包括:

基于所述第一损失值和第二损失值,调整所述图像识别神经网络模型的参数; 以及

基于所述第一损失值,调整所述文本识别神经网络模型的参数。

3. 如权利要求 2 所述的方法,还包括:

获取所述语义文本信息的至少一个关键字,将所述至少一个关键字中的一个或多个关键字作为所述真实标签。

4. 如权利要求 1 所述的方法，其中，所述样本图像包括样本对象，并且所述方法还包括：

在将所述样本图像输入所述图像识别神经网络模型之前，确定所述样本图像中所述样本对象所处的前景区域；

对所述样本图像进行裁剪，以得到前景图像，并将所述前景图像作为图像识别神经网络模型的输入。

5. 如权利要求 4 所述的方法，其中，所述图像识别神经网络模型的输出还包括所述样本对象的预测标签，

并且所述方法还包括：

标记所述样本图像中所述样本对象的真实标签；

基于所述预测标签和真实标签，计算第三损失值，

其中，至少基于所述第一损失值，调整所述图像识别神经网络模型和文本识别神经网络模型的参数包括：

基于所述第一损失值和第三损失值，调整所述图像识别神经网络模型的参数；以及

基于所述第一损失值，调整所述文本识别神经网络模型的参数。

6. 一种基于神经网络模型的图像检索方法，所述神经网络模型为通过权利要求 1-5 中任一项所述的训练方法进行训练来得到的，所述神经网络模型包括图像识别神经网络模型和文本识别神经网络模型，所述方法包括：

将待检测图像输入图像识别神经网络模型，获取所述图像识别神经网络模型输出的图像特征向量；以及

基于所述待检测图像的图像特征向量，从数据库中确定所述待检测图像的第一匹配图像集。

7. 如权利要求 6 所述的方法，其中，所述待检测图像包括目标对象，所述图像识别神经网络模型的输出还包括包围所述目标对象的目标边界框和所述目标对象的目标标签，

并且所述方法还包括：

将所述目标标签输入文本识别神经网络模型，获取所述文本识别神经网络模型输出的文本特征向量；以及

基于所述文本特征向量，从所述第一匹配图像集中确定所述待检测图像的至少一个匹配图像。

8. 如权利要求 6 所述的方法，其中，所述待检测图像包括目标对象，所述图像识别神经网络模型的输出还包括包围所述目标对象的目标边界框和所述目标对象的目标标签，

并且所述方法还包括：

将所述目标标签输入文本识别神经网络模型，获取所述文本识别神经网络模型输出的文本特征向量；

基于所述文本特征向量，从数据库中确定所述待检测图像的第二匹配图像集；以及

基于所述第一匹配图像集和第二匹配图像集，确定所述待检测图像的至少一个匹配图像。

9. 如权利要求 6 所述的方法，其中，所述待检测图像包括目标对象，并且所述方法还包括：

在将所述待检测图像输入所述图像识别神经网络模型之前，确定所述待检测图像中所述目标对象所处的前景区域；

对所述待检测图像进行裁剪，以得到前景图像，并将所述前景图像作为图像识别神经网络模型的输入。

10. 如权利要求 9 所述的方法，其中，所述图像识别神经网络模型的输出还包括所述目标对象的目标标签，

并且所述方法还包括：

将所述目标标签输入文本识别神经网络模型，获取所述文本识别神经网络模型输出的文本特征向量；以及

基于所述文本特征向量，从所述第一匹配图像集中确定所述待检测图像的至少一个匹配图像。

11. 如权利要求 9 所述的方法，其中，所述图像识别神经网络模型的输出还包括所述目标对象的目标标签，

并且所述方法还包括：

将所述目标标签输入文本识别神经网络模型，获取所述文本识别神经网络模型输出的文本特征向量；

基于所述文本特征向量，从数据库中确定所述待检测图像的第二匹配图像集；以及

基于所述第一匹配图像集和第二匹配图像集，确定所述待检测图像的至少一个匹配图像。

12. 一种神经网络模型的训练装置，所述神经网络模型包括图像识别神经网络模型和文本识别神经网络模型，所述装置包括：

第一获取单元，被配置用于获取样本图像以及样本图像相对应的语义文本信息；

第二获取单元，被配置用于将所述样本图像输入图像识别神经网络模型，获取所述图像识别神经网络模型输出的所述样本图像相对应的第一特征向量；

第三获取单元，被配置用于将语义文本信息输入文本识别神经网络模型，获取所述文本识别神经网络模型输出的所述语义文本信息相对应的第二特征向量；

计算单元，被配置为基于所述第一特征向量和第二特征向量，计算第一损失值；以及

调参单元，被配置为至少基于所述第一损失值，调整图像识别神经网络模型的参数。

13. 如权利要求 12 所述的装置，其中，所述样本图像包括样本对象，所述图像识别神经网络模型的输出还包括预测边界框和所述样本对象的预测标签，

并且所述装置还包括：

第一标记单元，被配置为标记所述样本图像中包围所述样本对象的真实边界框以及所述样本对象的真实标签，

其中，所述计算单元被进一步配置为基于所述预测边界框、所述预测标签、所述真实边界框和真实标签，计算第二损失值，

其中，所述调参单元包括：

第一调参子单元，被配置为基于所述第一损失值和第二损失值，调整所述图像识别神经网络模型的参数；以及

第二调参子单元，被配置为基于所述第一损失值，调整所述文本识别神经网络模型的参数。

14. 如权利要求 12 所述的装置，其中，所述第一标记单元被进一步配置为获取所述语义文本信息的至少一个关键字，将所述至少一个关键字中的一个或多个关键字作为所述真实标签。

15. 如权利要求 12 所述的装置，其中，所述样本图像包括样本对象，并且所述装置还包括：

确定单元，被配置为在将所述样本图像输入所述图像识别神经网络模型之前，确定所述样本图像中所述样本对象所处的前景区域；以及

裁剪单元，被配置为对所述样本图像进行裁剪，以得到前景图像，并将所述前景图像作为图像识别神经网络模型的输入。

16. 如权利要求 12 所述的装置，其中，所述图像识别神经网络模型的输出还包括所述样本对象的预测标签，

并且所述装置还包括：

第二标记单元，被配置为标记所述样本图像中所述样本对象的真实标签；

其中，所述计算单元被进一步配置为基于所述预测标签和所述真实标签，计算第三损失值，

其中，所述调参单元包括：

第三调参子单元，被配置为基于所述第一损失值和第三损失值，调整所述图像识别神经网络模型的参数；以及

第四调参子单元，被配置为基于所述第一损失值，调整所述文本识别神经网络模型的参数。

17. 一种基于神经网络模型的图像检索装置，所述神经网络模型为通过权利要求 1-5 中任一项所述的训练方法进行训练来得到的，所述神经网络模型包括图像识别神经网络模型和文本识别神经网络模型，包括：

第一获取单元，被配置用于将待检测图像输入图像识别神经网络模型，获取所述图像识别神经网络模型输出的图像特征向量；以及

第一确定单元，被配置用于基于所述待检测图像的图像特征向量，从数据库中确定所述待检测图像的第一匹配图像集。

18. 根据权利要求 17 所述的装置，其中，所述待检测图像包括目标对象，所述图像识别神经网络模型的输出还包括包围所述目标对象的目标边界框和所述目标对象的目标标签，

并且所述装置还包括：

第二获取单元，被配置用于将所述目标标签输入文本识别神经网络模型，获取所述文本识别神经网络模型输出的文本特征向量；以及

第二确定单元，被配置用于基于所述文本特征向量，从所述第一匹配图像集中确定所述待检测图像的至少一个匹配图像。

19. 根据权利要求 17 所述的装置，其中，所述待检测图像包括目标对象，所述图像识别神经网络模型的输出还包括包围所述目标对象的目标边界框和所述目标对象的目标标签，

并且所述装置还包括：

第三获取单元，被配置用于将所述目标标签输入文本识别神经网络模型，获取所述文本识别神经网络模型输出的文本特征向量；

第三确定单元，被配置用于基于所述文本特征向量，从数据库中确定所述待检测图像的第二匹配图像集；以及

第四确定单元，被配置用于基于所述第一匹配图像集和第二匹配图像集，确定所述待检测图像的至少一个匹配图像。

20. 根据权利要求 17 所述的装置，其中，所述待检测图像包括目标对象，并且所述装置还包括：

第五确定单元，被配置用于在将所述待检测图像输入所述图像识别神经网络模型之前，确定所述待检测图像中所述目标对象所处的前景区域；

裁剪单元，被配置用于对所述待检测图像进行裁剪，以得到前景图像，并将所述前景图像作为图像识别神经网络模型的输入。

21. 根据权利要求 20 所述的装置，其中，所述图像识别神经网络模型的输出还包括所述目标对象的目标标签，

并且所述装置还包括：

第四获取单元，被配置用于将所述目标标签输入文本识别神经网络模型，获取所述文本识别神经网络模型输出的文本特征向量；以及

第六确定单元，被配置用于基于所述文本特征向量，从所述第一匹配图像集中确定所述待检测图像的至少一个匹配图像。

22. 根据权利要求 20 所述的装置，其中，所述图像识别神经网络模型的输出还包括所述目标对象的目标标签，

并且所述装置还包括：

第五获取单元，被配置用于将所述目标标签输入文本识别神经网络模型，获取所述文本识别神经网络模型输出的文本特征向量；

第七确定单元，被配置用于基于所述文本特征向量，从数据库中确定所述待检测图像的第二匹配图像集；以及

第八确定单元，被配置用于基于所述第一匹配图像集和第二匹配图像集，确定所述待检测图像的至少一个匹配图像。

23. 一种电子设备，包括：

至少一个处理器；以及

与所述至少一个处理器通信连接的存储器；其中

所述存储器存储有可被所述至少一个处理器执行的指令，所述指令被所述至少一个处理器执行，以使所述至少一个处理器能够执行权利要求 1-11 中任一项所述的方法。

24. 一种存储有计算机指令的非瞬时计算机可读存储介质，其中，所述计算机指令用于使所述计算机执行根据权利要求 1-11 中任一项所述的方法。

25. 一种计算机程序产品，包括计算机程序，其中，所述计算机程序在被处理器执行时实现权利要求 1-11 中任一项所述的方法。

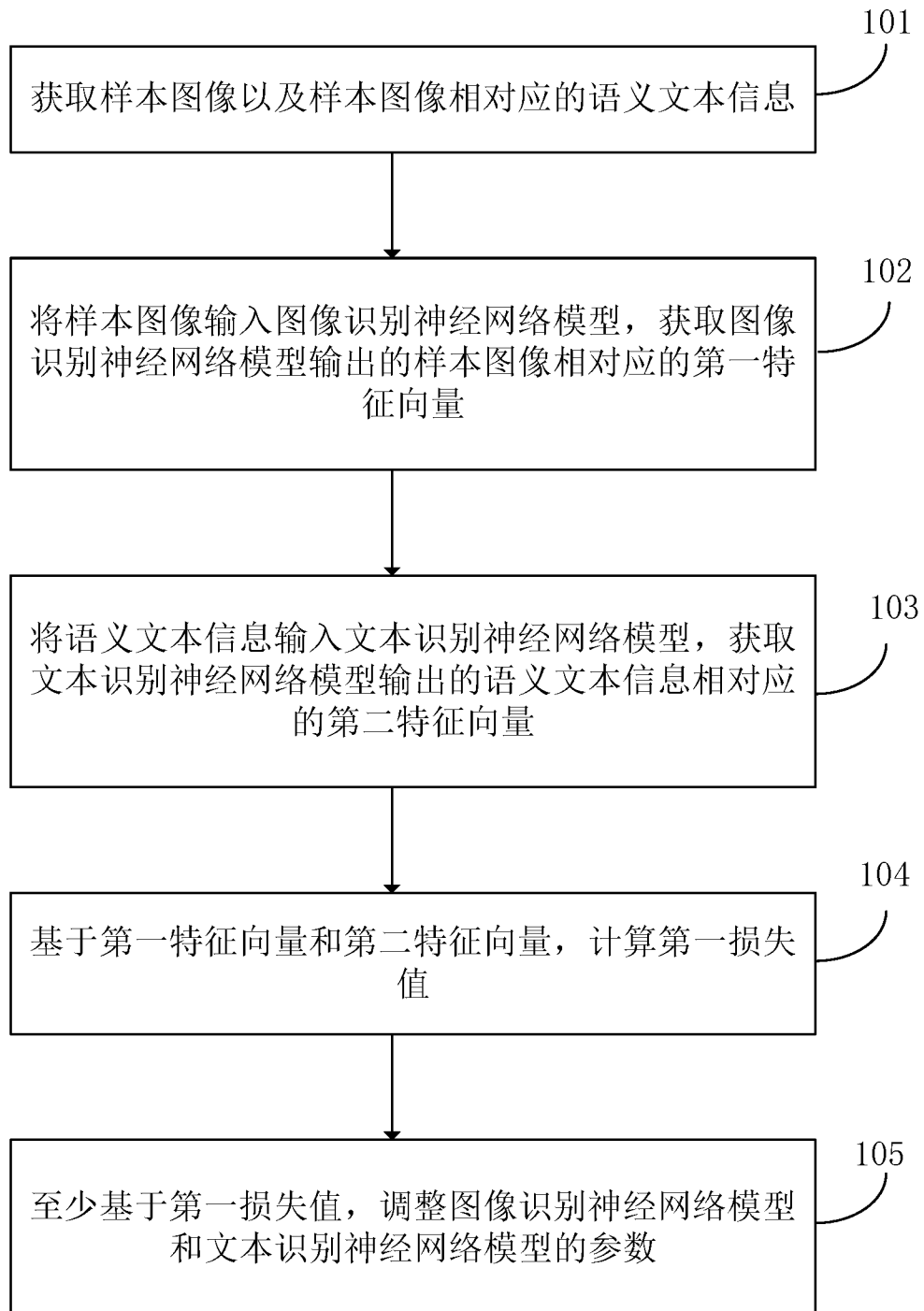
100

图 1

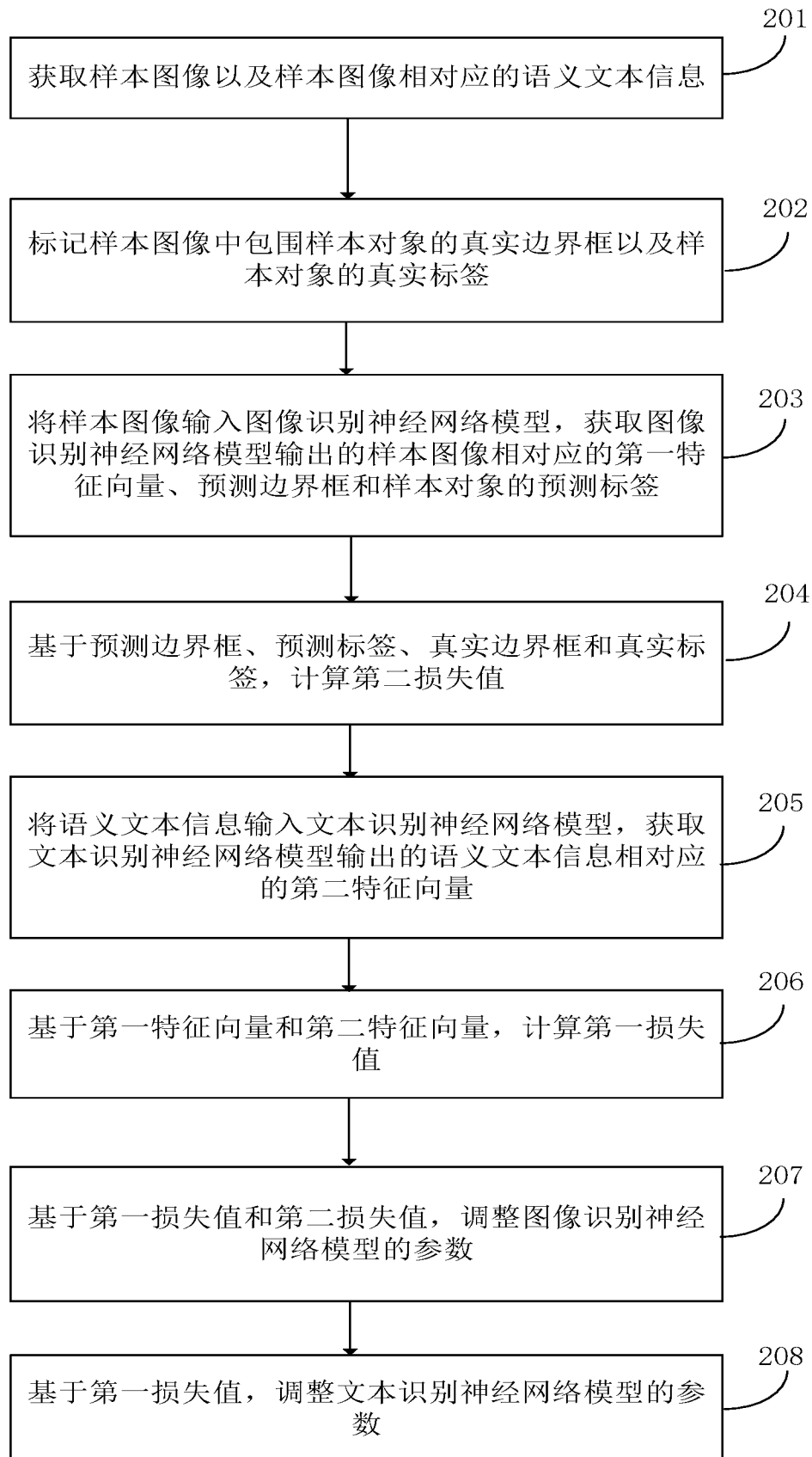
200

图 2

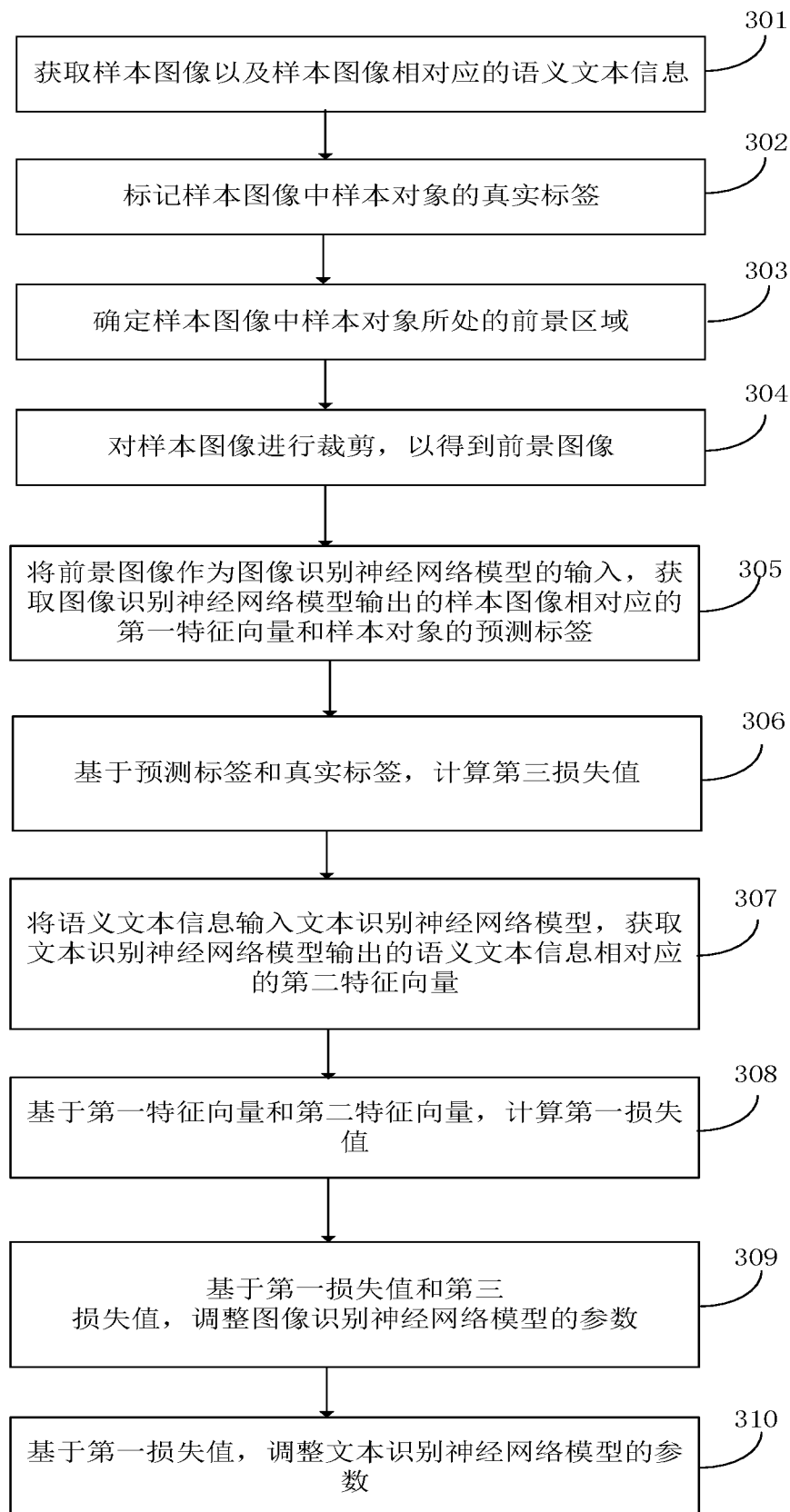
300

图 3

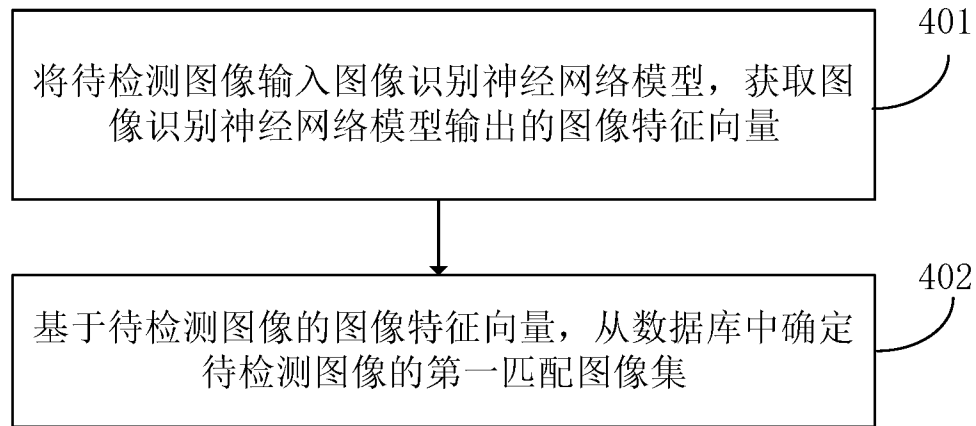
400

图 4

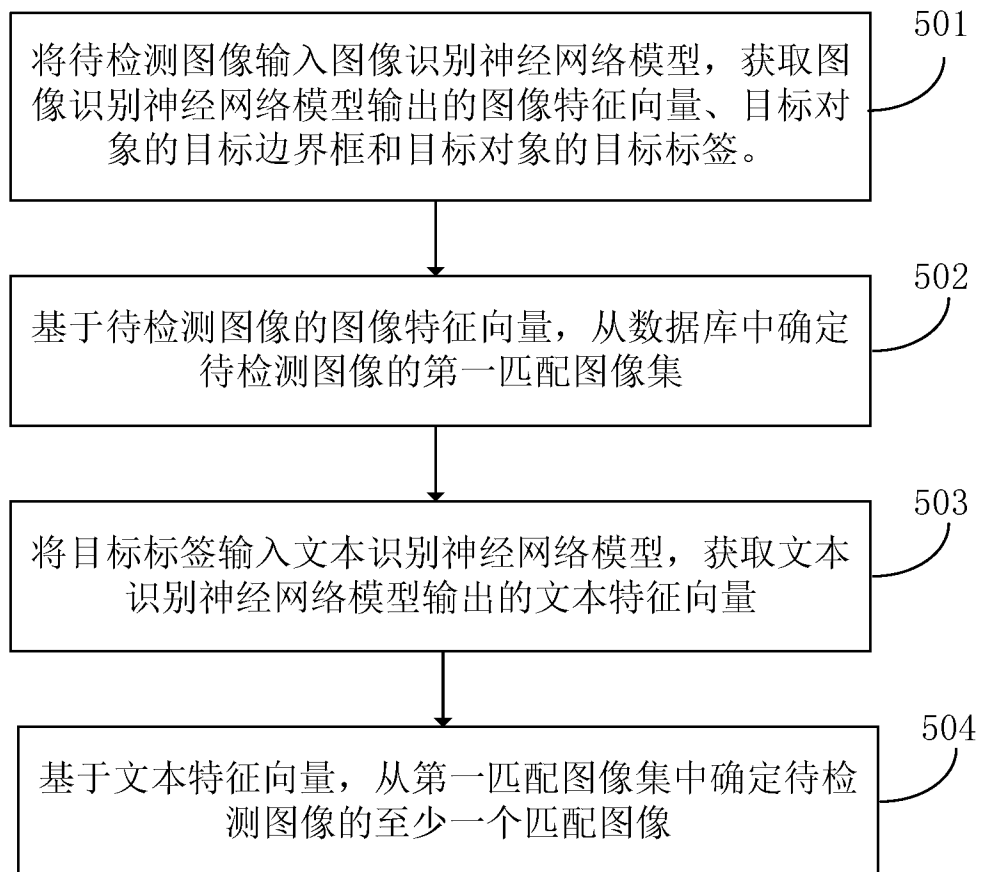
500

图 5

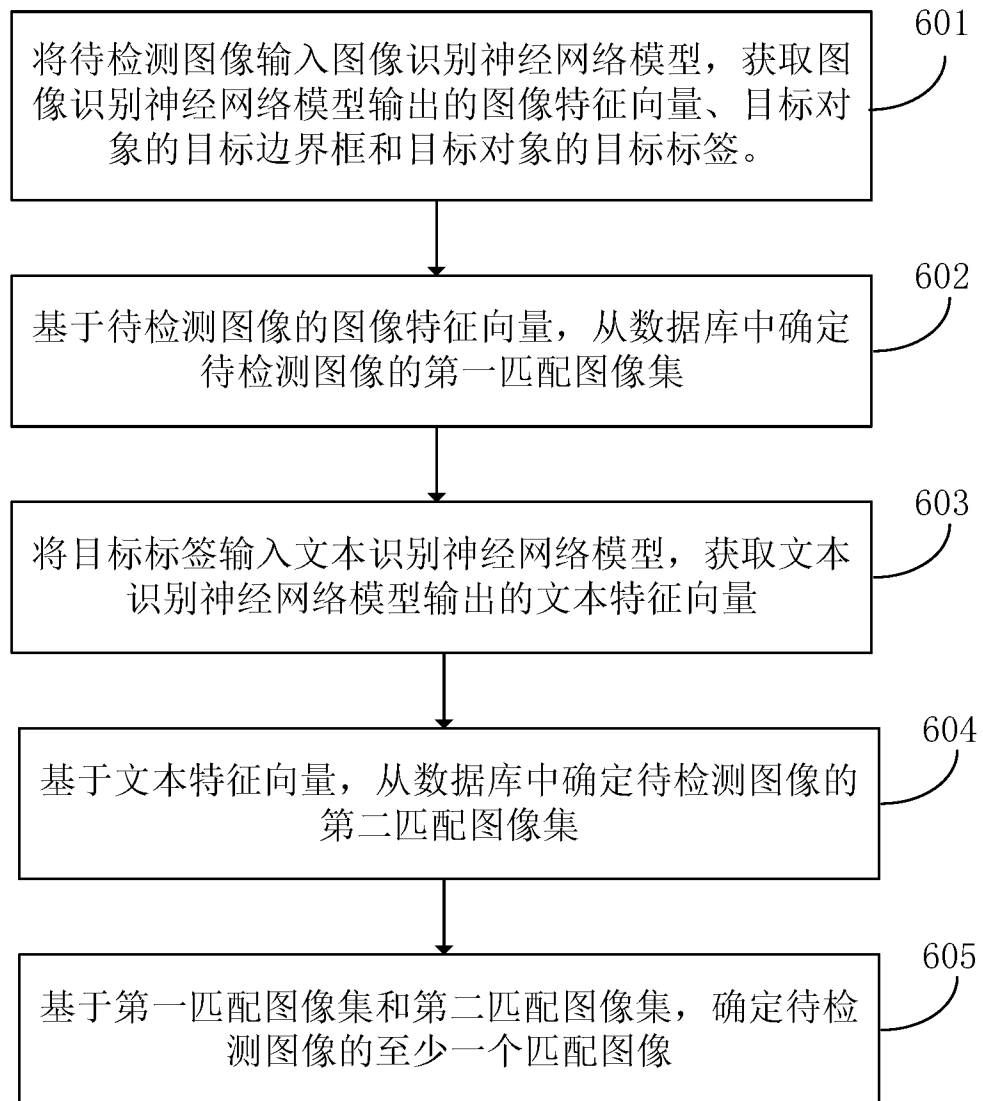
600

图 6

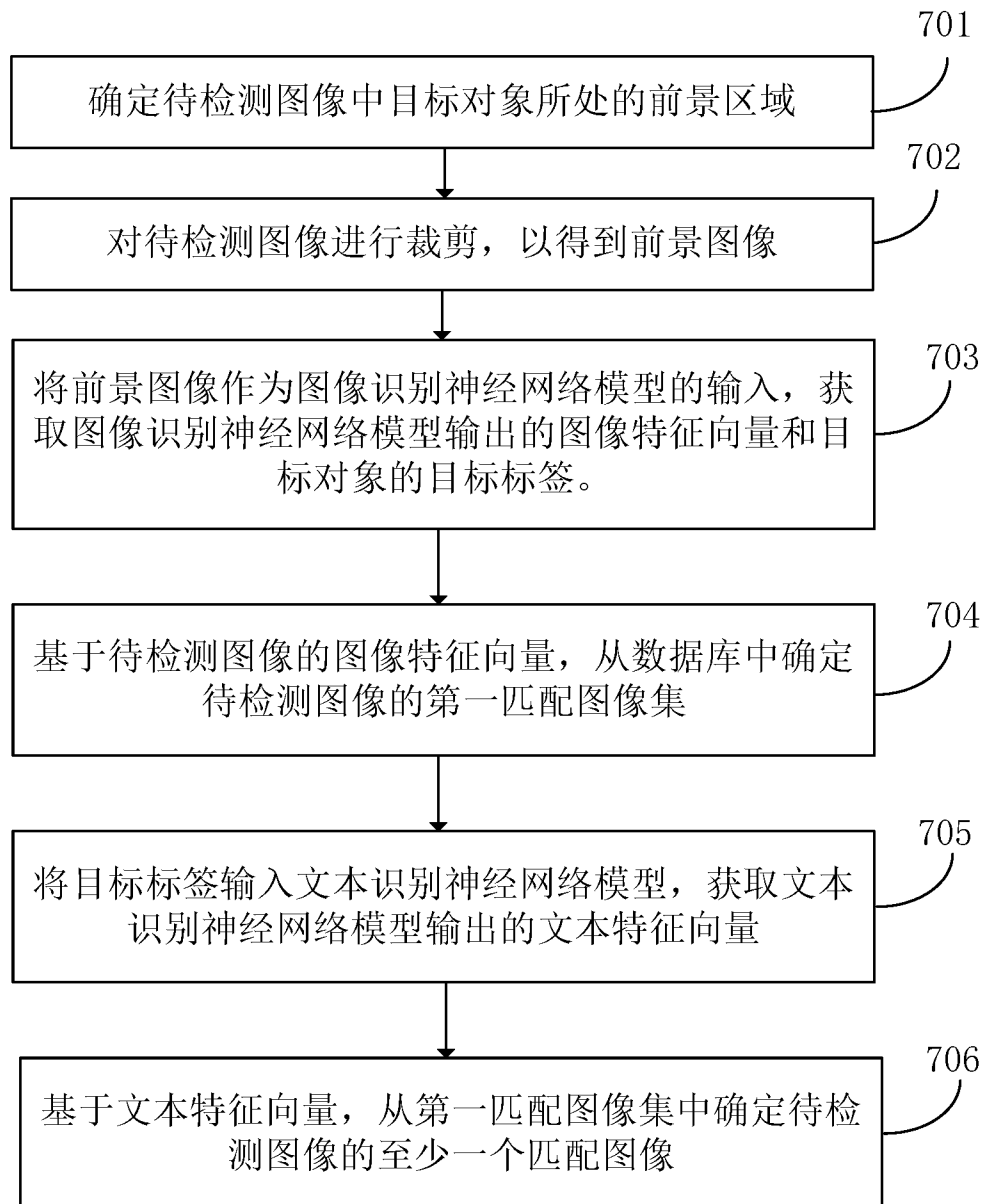
700

图 7

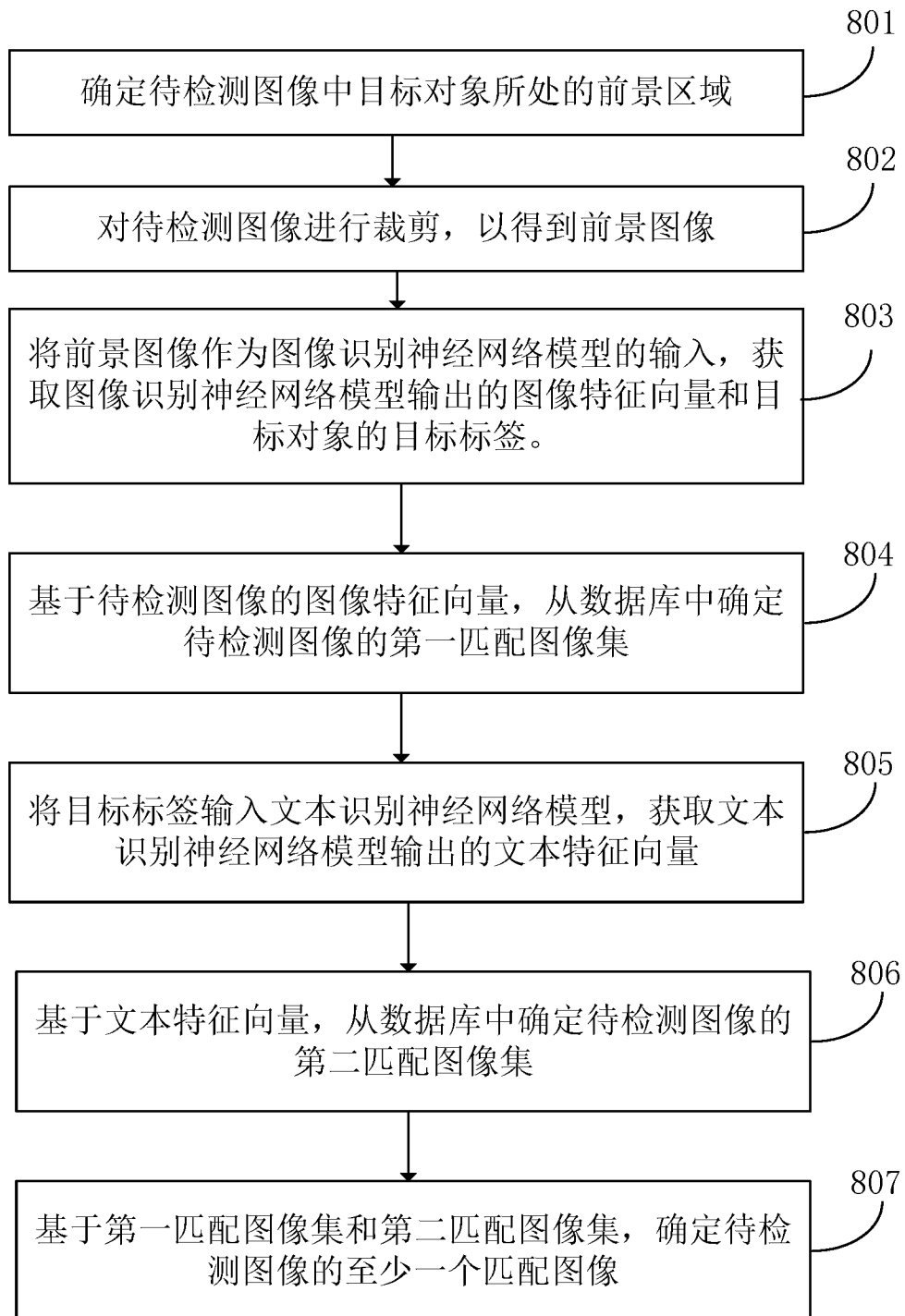
800

图 8

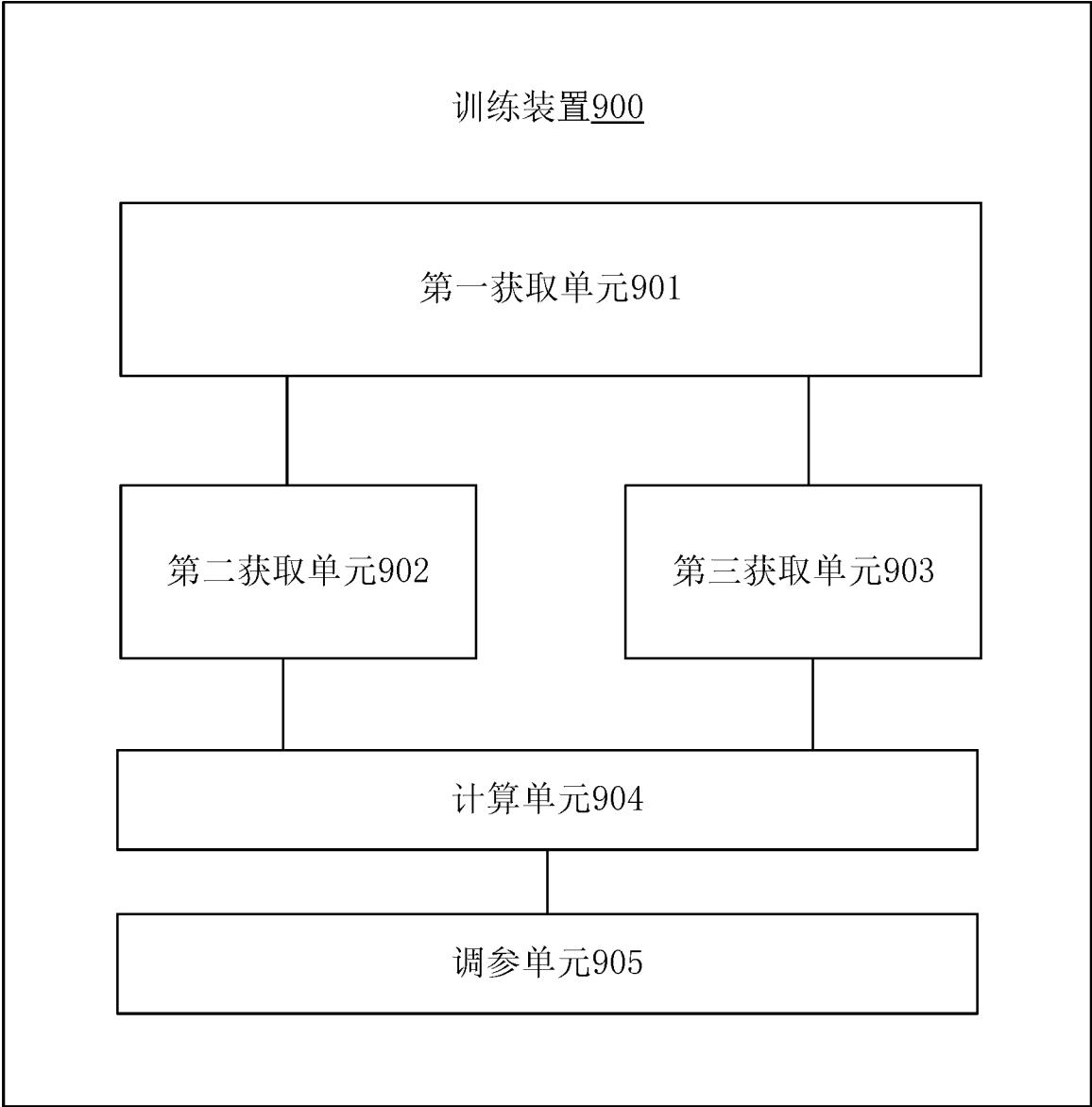


图 9

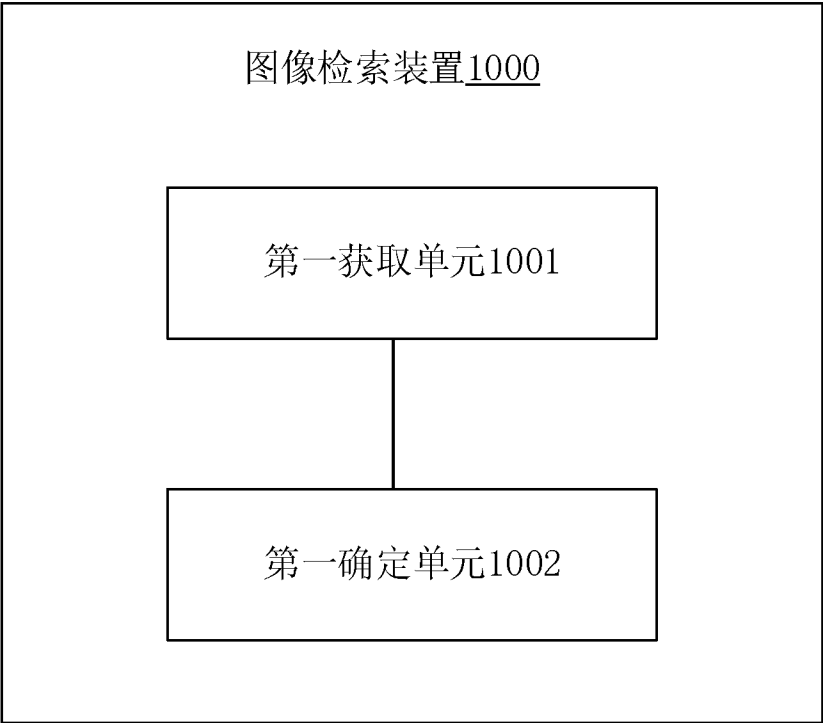


图 10

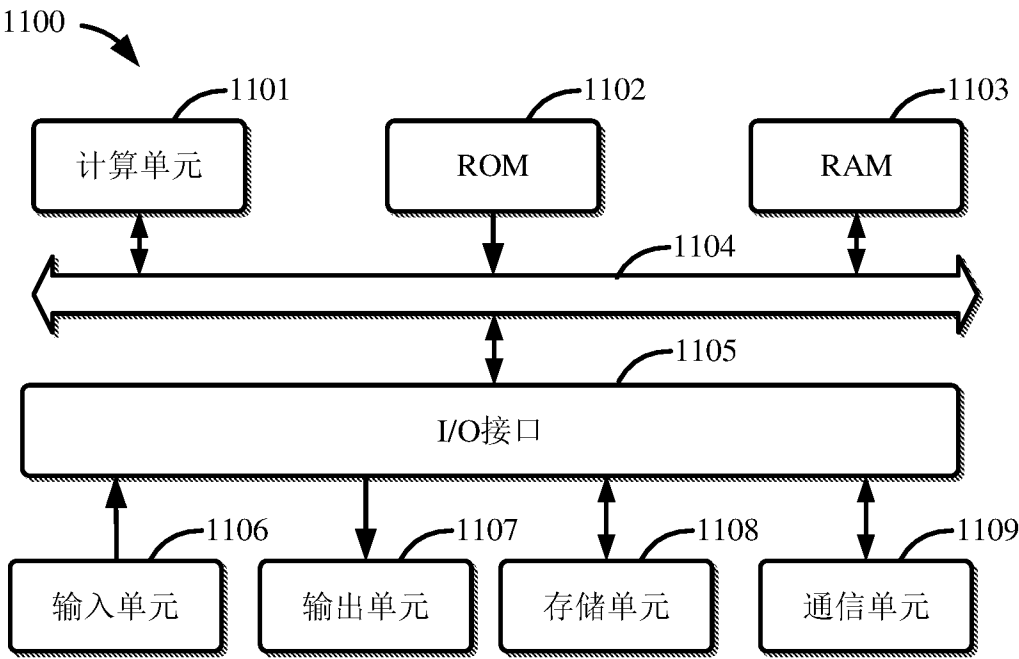


图 11

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/089626

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/583(2019.01)i; G06K 9/62(2022.01)i; G06V 10/40(2022.01)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)

G06F; 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)

CNABS, CNTXT, CNKI, DWPI, ENTXT, ENTXTC: 图像, 检索, 搜索, 查询, 文本, 语义, 训练, 特征, 损失, 识别, 区域推荐网络; image, search, retrieve, query, text, semantic, training, feature, loss, recognition, RPN.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 113656582 A (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.) 16 November 2021 (2021-11-16) claims 1-25	1-25
Y	CN 112364195 A (TIANJIN UNIVERSITY) 12 February 2021 (2021-02-12) claims 1-6, description paragraphs 0034-0051, and figure 2	1-25
Y	CN 109522967 A (GUANGZHOU COMMA SMART RETAIL CO., LTD.) 26 March 2019 (2019-03-26) description, paragraphs 0035-0057	1-25
Y	CN 104298749 A (HANGZHOU TAOTAOSOU TECHNOLOGY CO., LTD.) 21 January 2015 (2015-01-21) description, paragraphs 0031-0055	4-11, 15, 17-22
A	CN 110866140 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 06 March 2020 (2020-03-06) entire document	1-25
A	CN 112612913 A (XIAMEN MEIYA PICO INFORMATION CO., LTD.) 06 April 2021 (2021-04-06) entire document	1-25

☒ 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

22 June 2022

Date of mailing of the international search report

29 June 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/CN2022/089626

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 11017019 B1 (SHUTTERSTOCK, INC.) 25 May 2021 (2021-05-25) entire document	1-25

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2022/089626

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	113656582	A	16 November 2021	None	
CN	112364195	A	12 February 2021	None	
CN	109522967	A	26 March 2019	None	
CN	104298749	A	21 January 2015	None	
CN	110866140	A	06 March 2020	None	
CN	112612913	A	06 April 2021	None	
US	11017019	B1	25 May 2021	None	

国际检索报告

国际申请号

PCT/CN2022/089626

A. 主题的分类

G06F 16/583(2019.01)i; G06K 9/62(2022.01)i; G06V 10/40(2022.01)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)

G06F; 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))

CNABS, CNTXT, CNKI, DWPI, ENTXT, ENTXTC: 图像, 检索, 搜索, 查询, 文本, 语义, 训练, 特征, 损失, 识别, 区域推荐网络; image, search, retrieve, query, text, semantic, training, feature, loss, recognition, RPN.

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 113656582 A (北京百度网讯科技有限公司) 2021年11月16日 (2021 - 11 - 16) 权利要求1-25	1-25
Y	CN 112364195 A (天津大学) 2021年2月12日 (2021 - 02 - 12) 权利要求1-6, 说明书0034-0051段, 图2	1-25
Y	CN 109522967 A (广州逗号智能零售有限公司) 2019年3月26日 (2019 - 03 - 26) 说明书0035-0057段	1-25
Y	CN 104298749 A (杭州淘淘搜科技有限公司) 2015年1月21日 (2015 - 01 - 21) 说明书0031-0055段	4-11, 15, 17-22
A	CN 110866140 A (腾讯科技深圳有限公司) 2020年3月6日 (2020 - 03 - 06) 全文	1-25
A	CN 112612913 A (厦门市美亚柏科信息股份有限公司) 2021年4月6日 (2021 - 04 - 06) 全文	1-25
A	US 11017019 B1 (SHUTTERSTOCK, INC.) 2021年5月25日 (2021 - 05 - 25) 全文	1-25

☐ 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

2022年6月22日

国际检索报告邮寄日期

2022年6月29日

ISA/CN的名称和邮寄地址

中国国家知识产权局(ISA/CN)

中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

张会

电话号码 62412328

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2022/089626

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	113656582	A	2021年11月16日	无	
CN	112364195	A	2021年2月12日	无	
CN	109522967	A	2019年3月26日	无	
CN	104298749	A	2015年1月21日	无	
CN	110866140	A	2020年3月6日	无	
CN	112612913	A	2021年4月6日	无	
US	11017019	B1	2021年5月25日	无	