

图片检测方法、设备及计算机存储介质

5 本公开要求于 2023 年 10 月 25 日提交中国专利局、申请号为 202311396792.4、申请名称为“图片检测方法、设备及计算机存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本公开中。

技术领域

10 本说明书实施例涉及图片处理技术领域，尤其涉及一种图片检测方法、设备及计算机存储介质。

背景技术

15 在很多场景，均需要对图片进行检测，判定图片是否为某种特定类型的图片。目前，在对图片内容进行检测时，通常是通过一些目标检测模型对图片进行检测，目标检测模型可以输出图片中包括预设的多种类别的对象的概率。比如，将图片输入到目标检测模型后，目标检测模型可以输出图片中包括人、房子、车辆、树木等对象的概率，如果该概率大于预先设置的某个概率阈值，则认为图片中包括上述对象。目前在设置该概率阈值时，通常是针对所有的对象，统一设置一个概率阈值，比如，将概率阈值设置为 80%，如果目标检测模型输出图片中包括人的概率大于 80%，则认为图片中包括人，同理，针对其他类别的对象，比如，房子、车辆等，如果目标检测模型输出图片中包括房子、车辆的概率大于 80%，则认为图片中包括房子、车辆。

20 然而，针对不同的对象，由于目标检测模型在对该对象进行识别时，其难度不一样，如果采用统一的概率阈值作为评判标准，可能会使得目标检测模型对不同类别的对象的识别准确率不一致，导致某些类别的对象的识别准确率会过低，并且相关技术中概率阈值大多凭借人为经验去设置，比较难把控目标检测模型的判定标准，容易出现在检测某种特定类型的图片时，要么判定标准过于严格，将过多不属于特定类型的图片也筛选出来，要么判定标准过于宽松，导致过多属于该特定类型的图片无法识别到，造成过多的误检和漏检的问题。

30 发明内容

为克服相关技术中存在的问题，本说明书实施例提供了一种图片检测方法、设备及计算机存储介质。

根据本说明书实施例的第一方面，提供一种图片检测方法，所述方法包括：
获取待检测图片；

利用目标检测模型预测所述待检测图片中包括预设的类别标签库中的类别标签指示的对象的预测概率；

若所述预测概率大于所述类别标签对应的概率阈值，则确定所述待检测图片为目标类型的图片；

5 其中，所述类别标签库中的每个类别标签对应一个概率阈值，每个类别标签对应的概率阈值基于所述目标检测模型对多张样本图片的预测结果的准确率设置，针对每个类别标签，若所述目标检测模型预测任一样本图片中包括该类别标签指示的对象的概率大于该类别标签对应的概率阈值时，判定该样本图片中包括该类别标签指示的对象，则所述目标检测模型对所述多张样本图片的预测结果的准确率大于预设准确率阈值。

10 根据本说明书实施例的第二方面，提供一种图片检测方法，所述方法包括：

获取待检测图片；

利用图文转换模型得到所述待检测图片的文本描述；

根据所述文本描述判定所述待检测图片中是否包括目标对象；

15 如果是，则利用目标检测模型预测所述待检测图片中包括第一类别的对象的预测概率，在所述预测概率大于预设概率阈值的情况下，判定所述待检测图片为目标类型的图片；

其中，所述第一类别的对象在所述待检测图片中的占比小于预设的第一占比，所述待检测图片中包括所述目标对象为所述待检测图片中包括所述第一类别的对象的前提条件。

20 根据本说明书实施例的第三方面，提供一种电子设备，所述电子设备包括处理器、存储器、存储于所述存储器且可供所述处理器执行的计算机程序，所述计算机程序被执行时实现上述第一方面和/或第二方面提及的方法。

根据本说明书实施例的第四方面，提供一种计算机存储介质，所述计算机存储介质上存储有计算机程序，所述计算机程序被处理器执行时实现上述第一方面和/或第二方面提及的方法。

25 根据本说明书实施例的第五方面，提供一种计算机程序产品，包括计算机程序，所述计算机程序被处理器执行时实现如第一方面和/或第二方面提及的方法。

30 本说明书实施例的有益效果：在检测图片是否为目标类型的图片的场景，可以预先基于该目标类型的图片的特性构建一个类别标签库，类别标签库中可以包括一个或多个类别标签，即当图片中包括该类别标签库中的任意类别标签指示的对象时，则可以认定该图片为目标类型的图片。其中，为了保证目标检测模型在对每个类别标签下的对象进行识别时，均具有较准确的识别率，针对每个类别标签，可以预先确定该类别标签对应的概率阈值。每个类别标签的概率阈值可以基于目标检测模型对大量的样本图片的预测结果的准确率调整得到，比如，如果以目标检测模型预测的样本图片中包括该类别标签指示的对象的概率大于该类别标签对应的概率阈值时，则认为该样本图片中包括该类别标签指示的对象作为判定标准，则在该判定标准下，目标检测模型对这些样本图片的预测结果的准确率
35 大于预设准确率阈值。通过针对类别标签库中的每个类别标签，均预先设置与之适配的

概率阈值，使得标检测模型在对目标类型的图片进行检测时，准确度更高，可以减少误检或漏检的现象。且每个类别标签对应的概率阈值可以基于目标检测模型的预测结果的准确率调整，从而用户可以基于实际需求调整检测标准的严格程度，使得图片的检测更加灵活。

应当理解的是，以上的一般描述和后文的细节描述仅是示例性和解释性的，并不能限制本说明书实施例。

附图说明

此处的附图被并入说明书中并构成本说明书实施例的一部分，示出了符合本说明书实施例的实施例，并与说明书一起用于解释本说明书实施例的原理。

- 图 1 为利用目标检测对图像进行检测的示意图；
- 图 2 为本说明书一示例性实施例示出的图片检测方法的示意图；
- 图 3 为本说明书一示例性实施例示出的应用场景的示意图；
- 图 4 为本说明书一示例性实施例示出的图片检测方法的流程图；
- 图 5 为本说明书一示例性实施例示出的结合图文转换模型和目标检测模型对图片进行检测的示意图；
- 图 6 为本说明书一示例性实施例示出的构建标签类别库的示意图；
- 图 7 为本说明书一示例性实施例示出的电子设备的逻辑框图。

具体实施方式

这里将详细地对示例性实施例进行说明，其示例表示在附图中。下面的描述涉及附图时，除非另有表示，不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本说明书实施例相一致的所有实施方式。相反，它们仅是与如所附权利要求书中所详述的、本说明书实施例的一些方面相一致的装置和方法的例子。

在本说明书实施例使用的术语是仅仅出于描述特定实施例的目的，而非旨在限制本说明书实施例。在本说明书实施例和所附权利要求书中所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形式，除非上下文清楚地表示其他含义。还应当理解，本文中使用的术语“和/或”是指并包含一个或多个相关联的列出项目的任何或所有可能组合。

应当理解，尽管在本说明书实施例可能采用术语第一、第二、第三等来描述各种信息，但这些信息不应限于这些术语。这些术语仅用来将同一类型的信息彼此区分开。例如，在不脱离本说明书实施例范围的情况下，第一信息也可以被称为第二信息，类似地，第二信息也可以被称为第一信息。取决于语境，如在此所使用的词语“如果”可以被解释成为“在……时”或“当……时”或“响应于确定”。

在很多场景，均需要对图片进行检测，判定图片是否为某种特定类型的图片。比如，为了对产品或服务进行宣传，可以在一些应用或网站上展示广告图片，为了避免

广告图片中包括一些违规内容，比如，涉及暴力、不雅画面的内容，在展示这些广告图片之前，可以对这些广告图片进行检测，以判定其是否为包括上述违规内容的违规图片，如果是，则禁止展示。

一般而言，在判定图片是否为某种特定类型的图片时，一般是对图片中的内容进行检测，判定图片中是否包括某些类别的对象，进而判定其是否为特定类型的图片。以违规图片的检测为例，如果图片中包括枪支、粉末状物体、或者过度裸露的某些身体部位等，则可以将其认定为违规图片。

目前，在对图片内容进行检测时，通常是通过一些目标检测模型对图片进行检测，目标检测模型可以输出图片中包括预设的多种类别的对象的概率。比如，如图 1 所示，将图片输入到目标检测模型 11 后，目标检测模型 11 可以输出图片中包括人、房子、车辆、树木等对象的概率，如果该概率大于预先设置的某个概率阈值，则认为图片中包括上述对象。目前在设置该概率阈值时，通常是针对所有的对象，统一设置一个概率阈值，比如，将概率阈值设置为 80%，如果目标检测模型 11 输出图片中包括人的概率大于 80%，则认为图片中包括人，同理，针对其他类别的对象，比如，房子、车辆等，如果目标检测模型 11 输出图片中包括房子、车辆的概率大于 80%，则认为图片中包括房子、车辆。然而，针对不同的对象，由于目标检测模型 11 在对该对象进行识别时，其难度不一样，如果采用统一的概率阈值作为评判标准，可能会使得目标检测模型 11 对不同类别的对象的识别准确率不一致，导致某些类别的对象的识别准确率会过低，并且相关技术中概率阈值大多凭借人为经验去设置，比较难把控目标检测模型 11 的判定标准，容易出现在检测某种特定类型的图片时，要么判定标准过于严格，将过多不属于特定类型的图片也筛选出来，要么判定标准过于宽松，导致过多属于该特定类型的图片无法识别到，造成过多的误检和漏检的问题。

基于此，本公开实施例提供了一种图片检测方法，如图 2 所示，在检测图片是否为目标类型的图片的场景，可以预先基于该目标类型的图片的特性构建一个类别标签库，类别标签库中可以包括一个或多个类别标签，即当图片中包括该类别标签库中的任意类别标签指示的对象时，则可以认定该图片为目标类型的图片。其中，为了保证目标检测模型 11 在对每个类别标签下的对象进行识别时，均具有较准确的识别率，针对每个类别标签，可以预先确定该类别标签对应的概率阈值。每个类别标签的概率阈值可以基于目标检测模型 11 对大量的样本图片的预测结果的准确率调整得到，比如，如果以目标检测模型 11 预测的样本图片中包括该类别标签指示的对象的概率大于该类别标签对应的概率阈值时，则认为该样本图片中包括该类别标签指示的对象作为判定标准，则在该判定标准下，目标检测模型 11 对这些样本图片的预测结果的准确率大于预设准确率阈值。

通过针对类别标签库中的每个类别标签，均预先设置与之适配的概率阈值，使得目标检测模型 11 在对目标类型的图片进行检测时，准确度更高，可以减少误检或漏检的现象。

且每个类别标签对应的概率阈值可以基于目标检测模型 11 的预测结果的准确率调整，从

而用户可以基于实际需求调整检测标准的严格程度，使得图片的检测更加灵活。

本公开实施例的图片检测方法可以由各种手机、平板、电脑等终端设备执行，也可以由服务器或者服务器集群执行。比如，可以由终端设备在展示图片之前对图片进行检测，或者，也可以是终端设备在向服务器发送图片展示请求时，由服务器对待展示的图片进行检测，判定该图片不是目标类型的图片后，再发送给终端设备展示。比如，如图 3 所示，为本公开实施例的一种应用场景示意图，以电商产品的海外业务场景为例，商家通常会为其产品提供广告素材，这些素材中往往包括大量的图片，由于海外广告的投放对广告图片存在一定的要求，比如，禁止涉及暴力、枪支等内容的违规图片的投放。因此，第三方电商平台在投放这些商家的广告图片之前，可以对这些广告图片进行检测，过滤到一些违规图片。比如，商家可以通过第三方电商平台的客户端 21 上传这些广告图片，客户端 21 将广告图片发送至第三方电商平台的服务端 22，服务端 22 中可以设置目标检测模型 11，通过目标检测模型 11 检测这些广告图片是否为违规图片，如果是，则拒绝投放这类广告图片。当然，以上应用场景只是示例性的例子，本公开实施例提供的图片检测方案并不局限上述场景，针对所有图片检测的场景均适用。

如图 4 所示，本公开实施例的图片检测方法可以包括以下步骤：

S402、获取待检测图片；

在步骤 S402 中，可以获取待检测图片，其中，待检测图片可以是各种类型的需要检测的图片，本公开实施例不做限制。

S404、利用目标检测模型 11 预测待检测图片中包括预设的类别标签库中的类别标签指示的对象的预测概率；

在步骤 S404 中，在获取到待检测图片后，可以利用目标检测模型 11 对待检测图片进行检测，以预测图片中包括类别标签库中的类别标签指示的对象的预测概率。其中，考虑到通常图片中包括某些类别的对象时，则将其判定为某个目标类型的图片，因此，可以基于目标类型的图片的特性，即其包括哪些类别对象，预先构建类别标签库，类别标签库中可以包括一个或者多个类别标签。即当图片中包括该类别标签库中的类别标签指示的对象时，则该图片为目标类型的图片。

以目标类型的图片为违规图片为例，考虑到违规图片包括多种类别，比如，包括枪支、武器等暴力图片、包括毒品等违法的图片、包括过度裸露身体部位的不雅图片等等。因此，可以基于违规图片包括的内容构建类别标签库，该类别标签库可以包括以下类别标签：枪支、粉末、裸露的特定身体部位等。然后判定待检测图片中是否包括上述类别标签指示的对象，如果包括，则可以判定待检测图片为违规图片。

其中，目标检测模型 11 可以各种预训练好的开源的目标检测模型 11，比如，blip-2 (Bootstrapping Language-Image Pre-training, 引导式语言-图像预训练) 模型，Faster-CNN (Faster Region-based Convolutional Neural Network, 更快的区域卷积神经网络) 模型，也可以是自行利用样本图片训练得到的目标检测模型 11，只要具备目标检

测的功能即可，本公开实施例不做限制。

S406、若预测概率大于类别标签对应的概率阈值，则确定待检测图片为目标类型的图片；

其中，类别标签库中的每个类别标签对应一个概率阈值，每个类别标签对应的概率阈值基于目标检测模型 11 对多张样本图片的预测结果的准确率设置，针对每个类别标签，若目标检测模型 11 预测任一样本图片中包括该类别标签指示的对象的概率大于该类别标签对应的概率阈值时，判定该样本图片中包括该类别标签指示的对象，则目标检测模型 11 对多张样本图片的预测结果的准确率大于预设准确率。

在步骤 S406 中，为了保证目标检测模型 11 识别各种类别标签指示的对象的准确率，在基于目标类型的图片构建类别标签库后，针对类别标签库中的每个类别标签，可以预先确定其对应的概率阈值。在确定每个类别标签对应的概率阈值时，可以获取大量的样本图片，然后利用目标检测模型 11 对这些样本图片的预测结果的准确率设置该概率阈值。其中，如果样本图片是否包括该类别标签指示的对象的判定标准为：“目标检测模型 11 预测样本图片中包括该类别标签指示的对象的概率大于该类别标签对应的概率阈值时，则判定该样本图片中包括该类别标签指示的对象”，则按照该判定标准对这些样本图片进行检测时，目标检测模型 11 对这些样本图片的预测结果的准确率大于预设准确率阈值。其中，该准确率阈值可以基于用户对判定标准的严格程度灵活设置，比如，如果希望误检率低一些，则该准确率阈值可以设置得高一些。

比如，以类别标签为“人物”为例，可以获取多张包括人物和不包括人物的样本图片，利用目标检测模型 11 预测这些样本图片包括“人物”的概率，如果该概率大于概率阈值，则认为包括“人物”。其中，该概率阈值需使得目标检测模型 11 对这些样本图片的预测结果的准确率高于预设准确率阈值。比如，假设预设准确率阈值为 80%，样本图片为 100 张图片，则该概率阈值需使得目标检测模型 11 对该 100 张图片中的至少 80 张图片的预测结果是准确的。

在一些实施例中，在确定类别标签库中的每个类别标签对应的概率阈值时，可以获取多张样本图片，这些样本图片携带标签，该标签用于指示样本图片中是否包括该类别标签指示的对象。然后可以设置该概率阈值的初始阈值，利用目标检测模型 11 对样本图片进行检测，在目标检测模型 11 预测的样本图片中包括该类别标签指示的对象的概率大于预设的初始阈值时，则判定样本图片中包括该类别标签指示的对象，然后可以基于多张样本图片各自对应的判定结果统计该初始阈值下目标检测模型 11 预测结果的准确率。然后可以判定该准确率是否低于预设的准确率阈值，如果低于，则调整该初始阈值后再基于调整后的初始阈值重新对各张样本图片是否包括该类别标签指示的对象进行判定，重复上述过程，直至目标检测模型 11 的预测结果的准确率高于预设准确率阈值，将最终调整后的初始阈值作为该概率阈值。

比如，以类别标签为“人物”为例，可以先设置一个概率阈值的初始值（比如，0.6）

和调节幅度（比如，0.05），然后基于目标检测模型 11 预测的各样本图片中包括人物的概率，如果大于 0.6，则判定样本图片中包括人物，并且可以基于样本图片的标签确定对该样本图片的预测结果是否准确。在完成对所有的样本图片的预测后，可以统计预测结果的准确率，如果小于预设准确率阈值（比如，0.8），则可以增大该概率阈值，比如调整为 0.65，
5 然后继续重复上述步骤，直至最终的目标检测模型 11 针对这些样本图片的预测结果的准确率大于或等于 0.8，则将此时的调整后的概率阈值作为最终的概率阈值。

在一些实施例中，为了提高设置各类别标签对应的概率阈值的准确度，样本图片中可以涵盖正样本和负样本，即包括该类别标签指示的对象的样本图片和不包括该类别标签指示的对象的样本图片。并且，样本图片中可以包括一些易于混淆的正样本或负样本。比如，
10 以类别标签为“人物”为例，样本图片中可以包括一些含有雕塑、或者连衣裙等易于让目标检测模型 11 混淆，导致其判定图片中含有“人物”的样本图片。

考虑到大部分场景，待检测图片的数量可能较多，如果逐一利用目标检测模型 11 预测每张待检测图片中包括类别标签库中的类别标签指示的对象的概率，检测速度可能较慢。为了提高检测速度，可以同时结合图文转换模型对待检测图片进行检测。图文转换模型可以
15 直接将待检测图片中的内容转换为文本描述，然后可以直接判定该文本描述中是否包括类别标签库的类别标签，即可以判定待检测图片中是否包括该类别标签指示的对象。很明显，文字的匹配较为简单，这种方式相比于直接检测待检测图片中包括各类对象的概率效率要高一些，可以提升检测速度。但是，由于图文转换模型通常只能在较大的粒度上对图片的内容进行识别，即其描述的往往是图片中的大粒度的主体，无法对小粒度的对象进行
20 识别和描述。举个例子，以违规图片的检测为例，假设有一张裸露了特定身体部位的人物图片，图文转换模型可能往往只是识别到图片中的人，而无法识别到该裸露的特定身体部位，从而也无法检测出该违规图片。而目标检测模型 11 则刚好可以克服上述缺陷，即目标检测模型 11 可以对图片中更粒度的对象进行识别，但是其检测速度更慢。

为了既能提高检测速度，又可以提高检测结果的准确度，在一些实施例中，如图 5 所示，
25 可以结合上述两类模型进行图片的检测。比如，可以将类别标签库中的类别标签指示的对象划分为两类，第一类别的对象和第二类别的对象。其中，第一类别的对象在待检测图片中的占比通常较小，比如，小于预设的第一占比，这类对象往往可能是某些对象的局部区域，或者是一些细节部分。第二类别的对象在待检测图片中的占比通常较大，比如，大于预设的第二占比，即这类对象往往是图像的主体部分。其中，第一占比和第二占比可以
30 以相同，也可以不同。

针对第一类别的对象，由于图文转换模型 12 可能无法识别到，因此，可以通过图文转换模型 12 去判定图片中是否包括更大粒度、更易于识别的目标对象。其中，待检测图片中包括该目标对象是其包括该第一类别的对象的前提条件。在一些实施例中，第一类别的对象可以是目标对象的局部区域，比如，该第一类别的对象可以是裸露的某个身体部位，
35 目标对象则可以是人物，待检测图片包括裸露的某个身体部位的前提条件是该待检测图片

中包括人物。在利用目标检测模型 11 对待检测图片进行检测之前，可以先利用图文转换模型 12 得到该待检测图片的文本描述，然后根据该文本描述判定该待检测图片是否包括该目标对象，如果不包括，则说明待检测图片不可能包括该第一类别的对象，则无需再输入至该目标检测模型 11 进行下一步的更细粒度对象的检测。如果判定该待检测图片包括该目标对象，则可以进一步利用该目标检测模型 11 对该待检测图片进行检测，以进一步判定是否包括该第一类别的对象。

通过利用图文转换模型 12 进行预判，判定是否需要对待检测图片进行更细粒度的对象的检测，如果不需要，则无需再利用目标检测模型 11 进行检测，如果需要，则再进一步利用目标检测模型 11 进行检测，可以大大提高检测速度。此外，由于目标检测模型 11 的检测结果也无法保证 100%准确，存在一定的误检，而图文转换模型 12 对图像中的主体识别结果往往较为准确，因而基于图文转换模型的预判结果再执行进一步的检测，一定程度上也可以提高检测结果的准确性。

在一些实施例中，考虑到针对第二类别的对象，由于其在待检测图片中的占比较大，往往是图片中的主体，因而，直接利用图文转换模型 12 即可以识别到。因此，在基于图文转换模型 12 输出的文本描述判定待检测图片不包括目标对象时，可以进一步基于该文本描述判定待检测图片是否包括第二类别的对象，如果包括，则判定该待检测图片为目标类型的图片。对于大粒度的第二类别的对象，由于图文转换模型 12 即可以得到较准确的检测结果，可以直接通过图文转换模型 12 进行检测，进而提高检测速度。

其中，该图文转换模型 12 可以是任意能够将图片转换为文本的模型，该模型可以是预训练好的开源模型，也可以是自行训练的模型，本公开实施例不做限制。

在一些实施例中，考虑到如果仅基于目标检测模型 11 的预测结果，判定待检测图片是否为目标类型的图片，由于目标检测模型 11 的准确率有限，也可能存在误判的情况，为了进一步提高图片检测结果的准确率，可以同时结合图文转换模型 12 对待检测图片进行检测。比如，在判定待检测图片为目标类型的图片之前，还可以利用图文转换模型 12 得到待检测图片的文本描述，然后判定文本描述中是否包括类别标签库中的类别标签，如果待检测图片的文本描述中包括该类别标签库中的类别标签，且该目标检测模型 11 输出的待检测图片中包括该类别标签指示的对象的预测概率大于该类别标签对应的概率阈值，则判定该待检测图片为目标类型的图片。

通过同时结合图文转换模型 12 输出的文本描述，以及该目标检测模型 11 输出的预测概率，判定待检测图片是否为目标类型的图片，可以大大提升检测结果的准确率。

在一些实施例中，该图文转换模型 12 可以是 blip-2 模型，该目标检测模型 11 可以是 clip (Contrastive Language-Image Pretraining, 对比语言-图像预训练) 模型。blip-2 模型可以识别图片的主体内容，但无法下沉到更细粒度的维度 (比如，只能识别到图片中是否有人，而无法识别这个人的特定身体部位)，而 clip 模型可以输出当前图片中存在某些细粒度的对象的概率，比如，存在人的概率是 0.95、存在某个特定的身体部位的概率是

0.9, 但 clip 模型无法 100%识别出图片是否包括某个对象。以识别暴露特定身体部位的违规图片为例, blip-2 模型可以识别待检测图片中是否包括人, 而 clip 模型可以识别更细粒度的特定身体部位, 并输出包括特定身体部位的概率, 因而, 可以结合 blip-2 模型和 clip 模型进行图片识别容, 先通过 blip-2 识别出是否存在人, 如果存在则通过 clip 来判断是否存在某些不能暴露的身体部位, 以判定图片是否违规图片。

5 在一些实施例中, 类别标签库可以由用户手动构建, 也可以自动构建。比如, 考虑到针对同一种对象, 其命名或描述方式多种多样, 即其类别标签的表述方式也多种多样。比如, 针对“房屋”这一对象, 可以描述成“房子”、“楼房”、“建筑”等等。考虑到用户自行构建的类别标签, 其表述方式可能和模型输出的描述方式存在差异, 如果直接使用用户构建的类别标签, 可能存在因为模型输出的描述方式和用户构建的类别标签表述不一致的问题, 造成误识别。

举个例子, 以图文转换模型 12 为例, 假设将待检测图片输入至图文转换模型 12, 图文转换模型 12 输入出的文本描述为“房子前方停着一辆车”, 而用户针对“房屋”这一对象设置的类别标签为“建筑”, 因而, 在基于文本描述判定待检测图片中是否包括“房屋”时, 可能因为两者表述方式不一致, 导致最终判定图片中不包括“房屋”。同理, 针对目标检测模型 11 也是类似, 如果目标检测模型 11 输出的是“房子的概率为 0.9”, 而用户针对“房屋”这一对象设置的类别标签为“建筑”, 也可能会导致模型最终判定图片中不包括“房屋”, 从而造成误识别。

20 因此, 可以构建和模型的表述适配的类别标签库, 以避免模型因为表述不一致而出现误判的问题。比如, 如图 6 所示, 为了得到和上述图文转换模型 12 适配的类别标签, 在构建类别标签库时, 可以获取目标图片集, 该目标图片集中的图片为目标类型的图片, 然后将该目标图片集中的图片输入至该图文转换模型 12 中, 由图文转换模型 12 得到该图片对应的文本描述, 然后可以从该文本描述中提取指定类型的类别标签, 以构建类别标签库。通过从图文转换模型 12 输出的文本描述中提取符合需求的类别标签, 使得构建的类别标签和图文转换模型 12 的描述一致, 从而可以减少因为表述不一致而造成误检的现象。

30 在一些实施例中, 针对目标检测模型 11 也可以采用类似的方式, 了得到和上述目标检测模型 11 适配的类别标签, 在构建类别标签库时, 可以获取目标图片集, 该目标图片集中的图片为目标类型的图片, 然后将该目标图片集中的图片输入至该目标检测模型 11 中, 由目标检测模型 11 输出图片包括的各个类别标签的对象的概率, 然后可以从目标检测模型 11 输出的类别标签中提取指定类型的类别标签, 以构建类别标签库。

通过从目标检测模型 11 输出的类别标签中提取符合要求的类别标签, 使得构建的类别标签和目标检测模型 11 在预测过程中输出的类别标签一致, 从而可以减少因为表述不一致而造成误检的现象。

35 在一些实施例中, 为了可以自动提取出和目标类型的图片匹配的类别标签, 用户可以

先基于目标类型的图片的特性设置好基准类别标签，该指定类型的类别标签可以与用户设置的基准类别标签一致，或者也可以与用户设置的基准类别标签语义相同。

此外，考虑到大部分场景，待检测图片的数量可能较多，如果逐一利用目标检测模型预测每张待检测图片中包括类别标签库中的类别标签指示的对象的概率，检测速度可能较慢。为了提高检测速度，申请人想到的一种方式是可以结合图文转换模型对待检测图片进行检测。图文转换模型可以直接将待检测图片中的内容转换为文本描述，然后可以直接判定该文本描述中是否包括类别标签库的类别标签，即可以判定待检测图片中是否包括该类别标签指示的对象。很明显，文字的匹配较为简单，这种方式相比于直接检测待检测图片中包括各类对象的概率效率要高一些，可以提升检测速度。但是，由于图文转换模型通常只能在较大的粒度上对图片的内容进行识别，即其描述的往往是图片中的大粒度的主体，无法对小粒度的对象进行识别和描述。举个例子，以违规图片的检测为例，假设有一张裸露了特定身体部位的人物图片，图文转换模型可能往往只是识别到图片中的人，而无法识别到该裸露的特定身体部位，从而也无法检测出该违规图片。而目标检测模型则刚好可以克服上述缺陷，即目标检测模型可以对图片中更粒度的对象进行识别，但是其检测速度更慢。

基于此，本公开实施例提供了一种图片检测方法，可以结合目标检测模型和图文检测模型对待检测图片进行检测，并且设置相应的检测顺序和方式，使得既可以提高检测效率，又能保证检测结果具有较高的准确度。

该图片检测方法具体包括以下步骤：

获取待检测图片；

利用图文转换模型得到待检测图片的文本描述；

根据文本描述判定待检测图片中是否包括目标对象；

如果是，则利用目标检测模型预测待检测图片中包括第一类别的对象的预测概率，在预测概率大于预设概率阈值的情况下，判定待检测图片为目标类型的图片；其中，第一类别的对象在待检测图片中的占比小于预设的第一占比，待检测图片中包括目标对象为待检测图片中包括第一类别的对象的前提条件。

在一些实施例中，第一类别的对象为目标对象的局部区域。

在一些实施例中，方法还包括：

如果基于文本描述判定待检测图片中不包括目标对象，则基于文本描述判定待检测图片中是否包括第二类别的对象，其中，第二类别的对象在待检测图片中的占比大于预设的第二占比；

如果包括，则确定待检测图片为目标类型的图片。

其中，上述检测方法的具体实施细节可参考上述实施例中的描述，在此不再赘述。

与本说明书实施例提供的图片检测方法实施例相对应，本说明还提供了图片检测装置，装置包括：

获取模块，用于获取待检测图片；

预测模块，用于利用目标检测模型预测待检测图片中包括预设的类别标签库中的类别标签指示的对象的预测概率；

判定模块，用于若预测概率大于类别标签对应的概率阈值，则确定待检测图片为目标类型的图片；

其中，类别标签库中的类别标签对应的概率阈值基于目标检测模型对多张样本图片的预测结果的准确率设置，若目标检测模型预测任一样本图片中包括该类别标签指示的对象的概率大于该类别标签对应的概率阈值时，判定该样本图片中包括该类别标签指示的对象，则目标检测模型对多张样本图片的预测结果的准确率大于预设准确率阈值。

上述装置中各个单元的功能和作用的实现过程具体详见上述图片检测方法中对应步骤的实现过程，在此不再赘述。

对于装置实施例而言，由于其基本对应于方法实施例，所以相关之处参见方法实施例的部分说明即可。以上所描述的装置实施例仅仅是示意性的，其中作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本说明书实施例方案的目的。本领域普通技术人员在不付出创造性劳动的情况下，即可以理解并实施。

从硬件层面而言，如图 7 所示，为本说明书实施例的确定设备所属办公区域的装置所在设备的一种硬件结构图，除了图 7 所示的处理器 72、存储器 74 之外，该设备通常还可以包括其他硬件，如负责处理报文的转发芯片等；从硬件结构上来讲该设备还可能是分布式的设备，可能包括多个接口卡，以便在硬件层面进行报文处理的扩展。存储器 74 存储有计算机指令，处理器 72 执行计算机指令时实现上述任一实施例提及的方法。

本公开所涉及的用户信息（包括但不限于用户设备信息、用户个人信息等）和数据（包括但不限于用于分析的数据、存储的数据、展示的数据等），均为经用户授权或者经过各方充分授权的信息和数据，并且相关数据的收集、使用和处理需要遵守相关地区的相关法律法规和标准，并提供有相应的操作入口，供用户选择授权或者拒绝。

由于本说明书实施例对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，包括若干指令用以使得一台终端设备执行本说明书实施例各个实施例方法的全部或部分步骤。而前述的存储介质包括：U 盘、移动硬盘、只读存储器(Read-Only Memory, ROM)、随机存取存储器(Random Access Memory, RAM)、磁碟或者光盘等各种可以存储程序代码的介质。

本说明书实施例还提供一种计算机程序产品，包括计算机程序，该计算机程序被处理器执行时实现前述任一实施例的方法。计算机程序存储在可读存储介质中，电子设备的

至少一个处理器可以从可读存储介质读取计算机程序，至少一个处理器执行计算机程序使得电子设备执行上述任一方法实施例中图片检测方法的流程，具体功能和所能实现的技术效果此处不再赘述。

- 以上所述仅为本说明书实施例的较佳实施例而已，并不用以限制本说明书实施例，
- 5 凡在本说明书实施例的精神和原则之内，所做的任何修改、等同替换、改进等，均应包含在本说明书实施例保护的范围之内。

权 利 要 求 书

1、一种图片检测方法，其中，所述方法包括：
获取待检测图片；
利用目标检测模型预测所述待检测图片中包括预设的类别标签库中的类别标签指示
5 的对象预测概率；
若所述预测概率大于所述类别标签对应的概率阈值，则确定所述待检测图片为目标类
型的图片；

其中，所述类别标签库中的每个类别标签对应一个概率阈值，每个类别标签对应的概
率阈值基于所述目标检测模型对多张样本图片的预测结果的准确率设置，针对每个类别标
10 签，若所述目标检测模型预测任一样本图片中包括该类别标签指示的对象的概率大于该
类别标签对应的概率阈值时，判定该样本图片中包括该类别标签指示的对象，则所述目标
检测模型对所述多张样本图片的预测结果的准确率大于预设准确率阈值。

2、根据权利要求1所述的方法，其中，所述每个类别标签对应的概率阈值基于所述
目标检测模型对多张样本图片的预测结果的准确率设置，包括：
15 获取多张携带标签的样本图片，所述标签用于指示所述样本图片中是否包括该类别
标签指示的对象；

在所述目标检测模型预测的所述样本图片中包括该类别标签指示的对象的概率大于
预设的初始阈值的情况下，判定所述样本图片中包括该类别标签指示的对象，基于多张样
本图片各自对应的判定结果统计该初始阈值下所述目标检测模型预测结果的准确率；

20 在判定所述准确率低于预设准确率阈值的情况下，不断调整所述初始阈值，直至所述
目标检测模型的预测结果的准确率高于所述预设准确率阈值，并将调整后的初始阈值作为
所述概率阈值。

3、根据权利要求1所述的方法，其中，所述类别标签库中的类别标签指示的对象包
括第一类别的对象，所述第一类别的对象在所述待检测图片中的占比小于预设的第一占比；
25 所述利用目标检测模型预测所述待检测图片中包括类别标签库中的类别标签指示的对象的
预测概率，包括：

利用图文转换模型得到所述待检测图片的文本描述；
根据所述文本描述判定所述待检测图片中是否包括目标对象；
如果是，则利用目标检测模型预测所述待检测图片中包括所述第一类别的对象的预测
30 概率，其中，所述待检测图片中包括所述目标对象为所述待检测图片中包括所述第一类别
的对象的前提条件。

4、根据权利要求3所述的方法，其中，所述第一类别的对象为所述目标对象的局部
区域。

5、根据权利要求3所述的方法，其中，所述类别标签库中的类别标签指示的对象包
35 括第二类别的对象，所述第二类别的对象在所述待检测图片中的占比大于预设的第二占比；

所述方法还包括：

如果基于所述文本描述判定所述待检测图片中不包括所述目标对象，则基于所述文本描述判定所述待检测图片中是否包括所述第二类别的对象；

如果包括，则确定所述待检测图片为目标类型的图片。

- 5 6、根据权利要求 1 所述的方法，其中，在执行若所述预测概率大于所述类别标签对应的概率阈值，则确定所述待检测图片为目标类型的图片的步骤之前，所述方法还包括：

利用图文转换模型得到所述待检测图片的文本描述，判定所述文本描述中是否包括所述类别标签库中的类别标签；

- 10 若所述预测概率大于所述类别标签对应的概率阈值，则确定所述待检测图片为目标类型的图片，包括：

若所述文本描述中包括所述类别标签库中的类别标签，且所述预测概率大于所述类别标签对应的概率阈值，则确定所述待检测图片为目标类型的图片。

7、根据权利要求 1-6 任一项所述的方法，其中，所述类别标签库基于以下方式构建得到：

- 15 获取目标图片集，所述目标图片集中的图片为所述目标类型的图片；

将所述目标图片集中的图片输入至图文转换模型，从所述图文转换模型输出的所述图片的文本描述中提取指定类型的类别标签，以构建所述类别标签库；和/或将所述目标图片集中的图片输入至所述目标检测模型，从所述目标检测模型输出的类别标签中提取指定类型的类别标签，以构建所述类别标签库。

- 20 8、根据权利要求 7 所述的方法，其中，所述指定类型的类别标签与用户设置的基准类别标签语义相同或相近。

9、根据权利要求 3 或 6 所述的方法，其中，所述目标检测模型为 clip 模型，所述图文转换模型为 blip-2 模型。

10、一种图片检测方法，其中，所述方法包括：

- 25 获取待检测图片；

利用图文转换模型得到所述待检测图片的文本描述；

根据所述文本描述判定所述待检测图片中是否包括目标对象；

如果是，则利用目标检测模型预测所述待检测图片中包括第一类别的对象的预测概率，在所述预测概率大于预设概率阈值的情况下，判定所述待检测图片为目标类型的图片；

- 30 其中，所述第一类别的对象在所述待检测图片中的占比小于预设的第一占比，所述待检测图片中包括所述目标对象为所述待检测图片中包括所述第一类别的对象的前提条件。

11、一种电子设备，其中，所述电子设备包括处理器和存储器，所述存储器上存储有计算机程序，所述处理器执行所述计算机程序是可实现如权利要求 1-10 任一项所述的方法。

- 35 12、一种计算机存储介质，其中，所述计算机存储介质上存储有计算机程序，所述计

算机程序被处理器执行时实现上述权利要求 1-10 任一项所述的方法。

13、一种计算机程序产品，包括计算机程序，所述计算机程序被处理器执行时，实现如权利要求 1-10 中任一项所述的方法。