

(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  
2025 年 1 月 2 日 (02.01.2025)



(10) 国际公布号  
**WO 2025/001898 A1**

(51) 国际专利分类号:

G06V 10/40 (2022.01)

(21) 国际申请号:

PCT/CN2024/099597

(22) 国际申请日:

2024 年 6 月 17 日 (17.06.2024)

(25) 申请语言:

中文

(26) 公布语言:

中文

(30) 优先权:

202310787598.2 2023年6月29日 (29.06.2023) CN

(71) 申请人: 北京字跳网络技术有限公司

(BEIJING ZITIAO NETWORK TECHNOLOGY CO., LTD.) [CN/CN]; 中国北京市海淀区紫金数码园4号楼2层0207, Beijing 100190 (CN)。

(72) 发明人: 刘一衡 (LIU, Yiheng); 中国北京市朝阳区七圣中街12号院融中心B1小邮局, Beijing 100028 (CN)。

吴俊塔 (WU, Junta); 中国北京市朝阳区七圣中街12号院融中心B1小邮局, Beijing 100028 (CN)。

(74) 代理人: 北京市柳沈律师事务所 (LIU, SHEN & ASSOCIATES); 中国北京市海淀区彩和坊路10号1号楼10层, Beijing 100080 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF,

(54) Title: IMAGE PROCESSING METHOD AND APPARATUS, ELECTRONIC DEVICE, AND COMPUTER-READABLE MEDIUM

(54) 发明名称: 一种图像处理方法、装置、电子设备、计算机可读介质

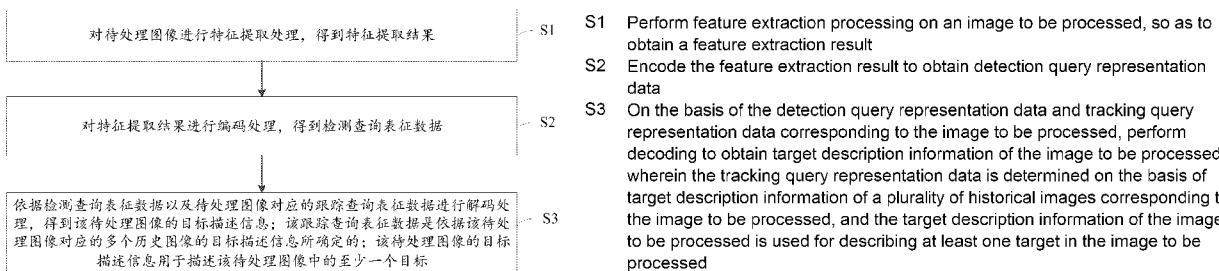


图 1

(57) Abstract: Embodiments of the present invention provide an image processing method and apparatus, an electronic device, and a computer-readable medium. The method comprises: first performing feature extraction processing on an image to be processed, so as to obtain a feature extraction result; then, encoding the feature extraction result to obtain image feature representation data and detection query representation data; and finally, on the basis of the image feature representation data, the detection query representation data, and tracking query representation data corresponding to said image, performing decoding to obtain target description information of said image, wherein the target description information can represent some targets in said image.

(57) 摘要: 本公开的实施例提供一种图像处理方法、装置、电子设备、计算机可读介质, 该方法包括: 先对该待处理图像进行特征提取处理, 得到特征提取结果; 再对该特征提取结果进行编码处理, 得到图像特征表征数据以及检测查询表征数据; 最后, 依据该图像特征表征数据、该检测查询表征数据以及该待处理图像对应的跟踪查询表征数据进行解码处理, 得到该待处理图像的目标描述信息, 以使该目标描述信息能够表示出该待处理图像中一些目标。

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))。

## 一种图像处理方法、装置、电子设备、计算机可读介质

5 本申请要求于 2023 年 6 月 29 日递交的中国专利申请第 202310787598.2 号的优先权，在此全文引用上述中国专利申请公开的内容以作为本申请的一部分。

### 技术领域

10 本公开的实施例涉及一种图像处理方法、装置、电子设备、计算机可读介质。

### 背景技术

目标跟踪用于从视频数据中分析出一些目标的运动状态信息(比如,在每帧视频图像中所处位置等)。

15

### 发明内容

本公开提供了一种图像处理方法、装置、电子设备、计算机可读介质。

为了实现上述目的,本公开提供的技术方案如下:

本公开提供一种图像处理方法,所述方法包括:

20 对待处理图像进行特征提取处理,得到特征提取结果;

对所述特征提取结果进行编码处理,得到检测查询表征数据;

依据所述检测查询表征数据以及所述待处理图像对应的跟踪查询表征数据进行解码处理,得到所述待处理图像的目标描述信息;所述跟踪查询表征数据是依据所述待处理图像对应的多个历史图像的目标描述信息所确定的;所述待处理图像的目标描述信息用于描述所述待处理图像中的至少一个目标。

25

在一种可能的实施方式下,所述对所述特征提取结果进行编码处理,得到检测查询表征数据,包括:

对所述特征提取结果进行编码处理,得到图像特征表征数据以及检测查询表征数据;

30 所述依据所述检测查询表征数据以及所述待处理图像对应的跟踪查询表

征数据进行解码处理，得到所述待处理图像的目标描述信息，包括：

依据所述图像特征表征数据、所述检测查询表征数据以及所述待处理图像对应的跟踪查询表征数据进行解码处理，得到所述待处理图像的目标描述信息。

- 5        在一种可能的实施方式下，所述多个历史图像包括第一类图像和第二类图像；

所述跟踪查询表征数据的确定过程，包括：

从所述第一类图像的目标描述信息中提取第一跟踪目标描述信息；

依据所述第一跟踪目标描述信息以及所述第二类图像的目标描述信息，

- 10      确定所述跟踪查询表征数据。

在一种可能的实施方式下，所述依据所述第一跟踪目标描述信息以及所述第二类图像的目标描述信息，确定所述跟踪查询表征数据，包括：

依据所述第一跟踪目标描述信息中的目标位置表征数据以及所述第二类图像的目标描述信息中的目标内容特征表征数据，确定第二跟踪目标描述信息；

- 15      根据所述第二跟踪目标描述信息和所述第一跟踪目标描述信息，确定所述跟踪查询表征数据。

在一种可能的实施方式下，所述待处理图像与所述待处理图像对应的多个历史图像属于同一个视频数据；

- 20      所述待处理图像的时序晚于所述第一类图像的时序；  
所述第一类图像的时序晚于所述第二类图像的时序。

在一种可能的实施方式下，所述依据所述检测查询表征数据以及所述待处理图像对应的跟踪查询表征数据进行解码处理，得到所述待处理图像的目标描述信息，包括：

- 25      依据所述检测查询表征数据和所述跟踪查询表征数据，确定待处理查询表征数据；

依据所述待处理查询表征数据进行解码处理，得到所述待处理图像的目标描述信息。

- 30      在一种可能的实施方式下，所述依据所述图像特征表征数据、所述检测查询表征数据以及所述待处理图像对应的跟踪查询表征数据进行解码处理，得

到所述待处理图像的目标描述信息，包括：

依据所述检测查询表征数据和所述跟踪查询表征数据，确定待处理查询表征数据；

依据所述图像特征表征数据和所述待处理查询表征数据进行解码处理，

5 得到所述待处理图像的目标描述信息。

在一种可能的实施方式下，所述待处理图像的目标描述信息是利用预先训练好的解码器所确定的；

所述解码器包括至少一个解码层，所述解码层中的自注意力模块是利用第一注意力掩模进行实现的，所述第一注意力掩模用于屏蔽属于同一个目标的在不同历史图像下对应的查询表征数据之间的交互。

在一种可能的实施方式下，所述至少一个目标包括一个或者多个待跟踪目标；

所述至少一个解码层包括第一解码层和第二解码层；

所述解码器还包括特征融合模块；

15 所述特征融合模块用于针对由所述第一解码层输出的跟踪查询处理数据进行融合处理，得到跟踪查询融合数据；所述跟踪查询处理数据是依据所述第一解码层以及所述跟踪查询表征数据所确定的；所述跟踪查询处理数据包括由所述第一解码层针对各所述待跟踪目标分别输出的处理后的查询表征数据；所述跟踪查询融合数据包括由所述特征融合模块针对各所述处理后的查询表征数据分别输出的融合后的查询表征数据；

所述第二解码层用于针对所述跟踪查询融合数据以及由所述第一解码层输出的检测查询处理数据进行处理；所述检测查询处理数据是依据所述第一解码层以及所述检测查询表征数据所确定的。

25 在一种可能的实施方式下，所述特征融合模块包括信息移除分支网络和信息添加分支网络；

所述跟踪查询融合数据的确定过程，包括：

利用所述信息移除分支网络对所述跟踪查询处理数据进行处理，得到保留信息特征；

30 利用所述信息添加分支网络对所述跟踪查询处理数据进行处理，得到待添加信息特征；

根据所述保留信息特征和所述待添加信息特征，确定所述跟踪查询融合数据。

在一种可能的实施方式下，所述保留信息特征的确定过程，包括：

对所述跟踪查询处理数据进行时序信息提取处理，得到第一时序信息；

- 5 对所述第一时序信息与所述跟踪查询处理数据进行自注意力处理，得到自注意力处理结果；

对所述自注意力处理结果进行全连接处理，得到全连接处理结果；

依据所述全连接处理结果，确定待移除信息表征数据；

- 10 依据所述待移除信息表征数据，对所述跟踪查询处理数据进行信息移除处理，得到所述保留信息特征。

在一种可能的实施方式下，所述待添加信息特征的确定过程，包括：

对所述跟踪查询处理数据进行时序信息提取处理，得到第二时序信息；

对所述第二时序信息与所述跟踪查询处理数据进行自注意力处理，得到所述待添加信息特征。

- 15 在一种可能的实施方式下，所述特征融合模块中的自注意力层是利用第二注意力掩模进行实现的，所述第二注意力掩模用于屏蔽属于不同目标的查询表征数据之间的交互。

在一种可能的实施方式下，所述待处理图像的目标描述信息是利用预先训练好的目标跟踪模型所确定的；

- 20 所述目标跟踪模型的训练损失的确定过程，包括：

依据第一图像数据、所述第一图像数据对应的多个历史图像的目标描述信息、以及所述目标跟踪模型，确定所述第一图像数据的目标描述信息；

从所述第一图像数据的目标描述信息中提取第一组信息和第二组信息；

- 25 依据所述第一组信息和所述第一图像数据对应的目标标签数据，得到第一损失；

依据所述第二组信息和所述目标标签数据，得到第二损失；

依据所述第一损失和所述第二损失，确定所述目标跟踪模型的训练损失。

在一种可能的实施方式下，所述第一组信息对应的时序晚于所述第二组信息对应的时序。

- 30 在一种可能的实施方式下，所述依据所述第二组信息和所述目标标签数

据，得到第二损失，包括：

依据标签匹配结果、所述第二组信息、以及所述目标标签数据，确定所述第二损失；所述标签匹配结果是根据所述第一组信息与所述目标标签数据之间的匹配结果所确定的。

- 5        在一种可能的实施方式下，所述第二组信息包括至少一个目标预测结果；  
所述第二损失的确定过程，包括：

依据所述标签匹配结果、所述第二组信息、以及所述目标标签数据，确定各所述目标预测结果对应的损失；

- 10       依据所述至少一个目标预测结果对应的损失的平均值，确定所述第二损失。

在一种可能的实施方式下，所述待处理图像的目标描述信息是利用预先训练好的目标跟踪模型所确定的；

所述目标跟踪模型的训练过程，包括：

利用至少一个第二图像数据以及各所述第二图像数据的目标检测标签，

- 15       对初始模型进行训练，得到待优化模型；

利用至少一个图像序列以及各所述图像序列的目标跟踪标签，对所述待优化模型中的解码模块进行训练，得到所述目标跟踪模型。

- 20       在一种可能的实施方式下，所述至少一个目标包括一个或者多个待跟踪目标；所述跟踪查询表征数据包括各所述待跟踪目标在多个历史图像下对应的查询表征数据。

在一种可能的实施方式下，所述待处理图像是指从图像序列中抽取的一个图像；所述目标描述信息包括至少一个目标预测结果；

所述得到所述待处理图像的目标描述信息之后，所述方法还包括：

从所述目标描述信息中提取满足预设参考条件的目标预测结果；

- 25       利用所述图像序列中的下一帧图像更新所述待处理图像，利用所述满足预设参考条件的目标预测结果中的查询表征数据，更新所述跟踪查询表征数据，并继续执行所述对待处理图像进行特征提取处理的步骤。

在一种可能的实施方式下，所述图像序列为视频数据。

本公开提供了一种图像处理装置，包括：

- 30       提取单元，被配置为对待处理图像进行特征提取处理，得到特征提取结

果;

编码单元,被配置为对所述特征提取结果进行编码处理,得到检测查询表征数据;

5 解码单元,被配置为依据所述检测查询表征数据以及所述待处理图像对应的跟踪查询表征数据进行解码处理,得到所述待处理图像的目标描述信息;所述跟踪查询表征数据是依据所述待处理图像对应的多个历史图像的目标描述信息所确定的;所述待处理图像的目标描述信息用于描述所述待处理图像中的至少一个目标。

本公开提供了一种电子设备,所述设备包括:处理器和存储器;

10 所述存储器,被配置为存储指令或计算机程序;

所述处理器,被配置为执行所述存储器中的所述指令或计算机程序,以使得所述电子设备执行本公开提供的图像处理方法。

本公开提供了一种计算机可读介质,所述计算机可读介质中存储有指令或计算机程序,当所述指令或计算机程序在设备上运行时,使得所述设备执行  
15 本公开提供的图像处理方法。

本公开提供了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本公开提供的图像处理方法的程序代码。

## 20 附图说明

为了更清楚地说明本公开实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本公开中记载的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其它的附图。

25 图1为本公开实施例提供的一种图像处理方法的流程图;

图2为本公开实施例提供的一种目标跟踪模型的示意图;

图3为本公开实施例提供的一种特征融合模块的示意图;

图4为本公开实施例提供的一种图像处理装置的结构示意图;

图5为本公开实施例提供的一种电子设备的结构示意图。

### 具体实施方式

经研究发现,对于上文基于卡尔曼滤波的两阶段跟踪方案来说,因该方案依赖于卡尔曼滤波器对于运动的建模,其要求目标的运动速度和方向不能变化太快。但是,因低帧率视频中的目标在帧间变化更加剧烈(比如,会出现快速的速度、方向、外观的变化等),如此导致该方案在低帧率视频下的目标跟踪效果严重下降,从而导致该方案不适于处理低帧率视频。

基于上述研究,为了更好地提高目标跟踪效果,本公开提供了一种图像处理方法,该方法包括:对于待处理图像(比如,一个视频数据中的某一帧视频图像)来说,先对该待处理图像进行特征提取处理,得到特征提取结果,以使该特征提取结果用于表征该待处理图像所携带的图像信息;再对该特征提取结果进行编码处理,得到检测查询表征数据(比如,多个检测 Query);最后,依据该检测查询表征数据以及该待处理图像对应的跟踪查询表征数据(比如,基于多个历史图像所确定的跟踪 Query)进行解码处理,得到该待处理图像的目标描述信息(比如,目标内容特征表征数据、目标位置表征数据、以及类别置信度等),以使该目标描述信息能够表示出该待处理图像中一些目标。其中,因该跟踪查询表征数据是依据该待处理图像对应的多个历史图像的目标描述信息所确定的,以使该跟踪查询表征数据能够更好地表示出该待处理图像中一些目标的历史状态,从而使得基于该跟踪查询表征数据所确定的目标描述信息能够更好地表示出该待处理图像中一些目标,如此有利于提高目标跟踪效果。

另外,当本公开提供的图像处理方法用于实现针对视频数据的目标跟踪处理时,因该图像处理方法需要参考多个历史图像的目标描述信息(比如,Query),以确定某一个视频图像中的目标,以使在针对该视频图像的目标确定过程中能够从这些历史图像的目标描述信息中获取到比较多的可参考内容(比如,目标在不同历史图像中所处状态等),如此能够有效地克服因视频数据的帧率较低而导致的问题,从而能够有效地提高针对低帧率视频的目标跟踪效果,进而使得本公开提供的图像处理方法适用于针对具有任何帧率的视频数据进行目标跟踪处理,如此有利于提高该图像处理方法在目标跟踪领域中的普适性。

此外,本公开不限定图像处理方法的执行主体,例如,本公开实施例提供

的图像处理方法的任一实施方式可以应用于终端设备或服务器等具有数据处理功能的设备。又如，本公开实施例提供的图像处理方法的任一实施方式也可以借助不同设备（例如，终端设备与服务器、两个终端设备、或者两个服务器）之间的数据通信过程进行实现。其中，终端设备可以为智能手机、计算机、个人数字助理（Personal Digital Assistant, PDA）或平板电脑等。服务器可以为独立服务器、集群服务器或云服务器。

为了使本技术领域的人员更好地理解本公开方案，下面将结合本公开实施例中的附图，对本公开实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅是本公开一部分实施例，而不是全部的实施例。基于本公开中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本公开保护的范围。

为了更好地理解本公开所提供的技术方案，下面先结合一些附图对本公开提供的图像处理方法进行说明。如图 1 所示，本公开实施例提供的图像处理方法，包括下文 S1-S3。其中，该图 1 为本公开实施例提供的一种图像处理方法的流程图。

S1：对待处理图像进行特征提取处理，得到特征提取结果。

其中，待处理图像是指需要进行目标确定处理的图像数据（比如，图 2 所示的第  $t$  帧视频图像  $I^t$ ）。 $t$  为正整数。需要说明的是，本公开不限定目标的实施方式，例如，该目标可以是指任意一种可检测的对象（比如，人、动物、植物、物体等）。

另外，本公开不限定上文待处理图像，比如，在一些应用场景（比如，针对某个图像序列进行目标跟踪处理等场景）下，该待处理图像是指从该图像序列中抽取的一个图像（比如，位于第  $t$  个排列位置的图像数据）。其中，该图像序列用于记录按照某种顺序（比如，时序）进行排列的多个图像数据；而且本公开不限定该图像序列的实施方式，例如，其可以采用现有的或者未来出现的任意一种图像序列（比如，视频数据）进行实施。

特征提取结果用于表示上文待处理图像所携带的图像信息。

另外，本公开不限定上文特征提取结果的确定过程（也就是，S1 的实施方式），比如，其可以采用现有的或者未来出现的任意一种能够针对一个图像数据进行特征提取处理的方法（比如，借助特征提取器等方式）进行实施。

可见，在一些应用场景下，上文 S1 具体可以为：将待处理图像输入预先构建的特征提取器，得到该特征提取器输出的特征提取结果，以使该特征提取结果能够表示出该待处理图像所携带的图像信息。其中，该特征提取器用于针对该特征提取器的输入数据进行特征提取处理；而且本公开不限定该特征提取器的实施方式，例如，该特征提取器可以采用现有的或者未来出现的任意一种具有特征提取功能的设备进行实施。又如，该特征提取器可以采用卷积神经网络（Convolutional Neural Networks, CNN）进行实施。还如，该特征提取器可以采用下文目标跟踪模型中的特征提取模块进行实施。

需要说明的是，本公开不限定上段中 CNN 的实施方式，例如，其可以采用基于 CoCo 数据集预训练的 ResNet50 进行实施。

基于上述两段内容可知，在一种可能的实施方式下，当本公开提供的图像处理方法借助预先训练好的目标跟踪模型（比如，图 2 所示的多目标跟踪模块）进行实现时，上文 S1 具体可以为：利用该目标跟踪模型中的特征提取模块（比如，图 2 所示的 CNN），对上文待处理图像进行特征提取处理，得到该待处理图像对应的特征提取结果。

基于上文 S1 的相关内容可知，对于待处理图像（比如，图 2 所示的第  $t$  帧视频图像  $I$ ）来说，在获取到该待处理图像之后，对该待处理图像进行特征提取处理，得到特征提取结果（比如，由图 2 所示的 CNN 所输出的数据），以使该特征提取结果能够表示出该待处理图像所携带的图像信息，以便后续能够基于该特征提取结果，确定该待处理图像中的一些目标。

S2：对特征提取结果进行编码处理，得到检测查询表征数据。

其中，检测查询表征数据用于表征上文待处理图像中一些目标；而且本公开不限定该检测查询表征数据，比如，该检测查询表征数据可以包括该待处理图像中多个目标对应的查询表征数据（例如，Query）。其中，该目标对应的查询表征数据用于描述该待处理图像中一个目标。可见，在一种可能的实施方式下，该检测查询表征数据可以包括一个或者多个目标在该待处理图像下对应的查询表征数据。其中，该目标在该待处理图像下对应的查询表征数据用于表征该目标在该待处理图像下所处状态。

另外，本公开不限定上段中目标对应的查询表征数据的实施方式，例如，其可以包括该目标对应的目标内容特征表征数据以及该目标对应的目标位置

表征数据。其中，该目标内容特征表征数据用于表征该目标在该待处理图像中所呈现的特征（比如，在姿态、衣服颜色等方面所呈现的特征）；而且本公开不限定该目标内容特征表征数据的实施方式，例如，该目标内容特征表征数据可以采用特征向量进行实施。该目标位置表征数据用于表征该目标在该待处理图像中所处位置；而且本公开不限定该目标内容特征表征数据的实施方式，比如，其可以采用边界框进行实施。

此外，本公开不限定上文 S2 中编码处理的实施方式，例如，其可以采用现有的或者未来出现的任意一种能够进行编码处理的方法（比如，借助编码器等方式）进行实施。

可见，在一种可能的实施方式下，上文 S2 具体可以为：将上文特征提取结果输入预先构建的编码器，得到该编码器输出的检测查询表征数据。其中，该编码器用于针对该编码器的输入数据进行编码处理；而且本公开不限定该编码器的实施方式，例如，该编码器可以采用现有的或者未来出现的任意一种具有编码功能的设备进行实施。又如，该编码器可以采用 Transformer Encoder 进行实施。还如，该编码器可以采用下文目标跟踪模型中的编码模块进行实施。

需要说明的是，本公开不限定上段中 Transformer Encoder 的实施方式，例如，其可以采用 DINO（DETR with Improved DeNoising Anchor Boxes）检测算法中的 Transformer Encoder 模块进行实施。

基于上述两段内容可知，在一种可能的实施方式下，当本公开提供的图像处理方法借助预先训练好的目标跟踪模型（比如，图 2 所示的多目标跟踪模块）进行实现时，上文 S2 具体可以为：在利用该目标跟踪模型中的特征提取模块从上文待处理图像中确定出特征提取结果之后，由该目标跟踪模型中的编码模块针对该特征提取结果进行编码处理，得到该待处理图像对应的检测查询表征数据（比如，图 2 所示的  $D^t$ ）。

需要说明的是，对于图 2 所示的  $D^t$  来说， $D^t$  用于表示图 2 所示的图像数据  $I$  中一些目标；而且该  $D^t$  可以采用下文公式（1）进行表示。

$$D^t = \{d_1^t, d_2^t, \dots, d_{N_{det}}^t\} \quad (1)$$

式中， $D^t$  表示通过针对上文图像数据  $I$  进行处理（比如，特征提取处理+编码处理）所确定的 Query； $d_1^t$  表示该图像数据  $I$  中第 1 个目标对应的 Query，

而且该 $d_1^t$ 包括该第 1 个目标对应的目标内容特征表征数据以及该第 1 个目标对应的目标位置表征数据； $d_2^t$ 表示该图像数据  $I$  中第 2 个目标对应的 Query，而且该 $d_2^t$ 包括该第 2 个目标对应的目标内容特征表征数据以及该第 2 个目标对应的目标位置表征数据；...（以此类推）； $d_{N_{det}}^t$ 表示该图像数据  $I$  中第  $N_{det}$  个目标对应的 Query，而且该 $d_{N_{det}}^t$ 包括该第 $N_{det}$ 个目标对应的目标内容特征表征数据以及该第 $N_{det}$ 个目标对应的目标位置表征数据； $N_{det}$ 表示预先设定的 Query 数目，以使该 $N_{det}$ 用于表示该 $D^t$ 所涉及的 Query 个数。

实际上，在一些应用场景下，为了更好地提高目标跟踪效果，本公开还提供了上文 S2 中编码处理的一种可能的实施方式，在该实施方式中，该 S2 具体可以为：对特征提取结果进行编码处理，得到图像特征表征数据以及检测查询表征数据。其中，该图像特征表征数据用于表征上文待处理图像所携带的信息；而且本公开不限定该图像特征表征数据的实施方式，例如，其可以采用现有的或者未来出现的任意一种针对图像数据所确定的特征编码向量（比如，图 2 所示的图像特征图）进行实施。

另外，本公开不限定上段中编码处理的实施方式，例如，该编码处理具体可以为：将上文特征提取结果输入预先构建的编码器，得到该编码器输出的图像特征表征数据以及检测查询表征数据。又如，当本公开提供的图像处理方法借助预先训练好的目标跟踪模型（比如，图 2 所示的多目标跟踪模块）进行实现时，该编码处理具体可以为：在利用该目标跟踪模型中的特征提取模块从上文待处理图像中确定出特征提取结果之后，由该目标跟踪模型中的编码模块针对该特征提取结果进行编码处理，得到该待处理图像对应的图像特征表征数据（比如，图 2 所示的图像特征图）以及该待处理图像对应的检测查询表征数据（比如，图 2 所示的 $D^t$ ）。

基于上段内容可知，在一种可能的实施方式下，对于待处理图像（比如，图 2 所示的第  $t$  帧视频图像  $I$ ）来说，在获取到该待处理图像对应的特征提取结果（比如，由图 2 所示的 CNN 所输出的数据）之后，对该特征提取结果进行编码处理，得到图像特征表征数据（比如，图 2 所示的图像特征图）以及检测查询表征数据（比如，图 2 所示的 $D^t$ ），以便后续能够基于该图像特征表征数据以及该检测查询表征数据，确定该待处理图像中的一些目标。

基于上文 S2 的相关内容可知，对于待处理图像（比如，图 2 所示的第  $t$

帧视频图像  $I$ ) 来说, 在获取到该待处理图像对应的特征提取结果 (比如, 由图 2 所示的 CNN 所输出的数据) 之后, 对该特征提取结果进行编码处理, 至少得到检测查询表征数据 (比如, 图 2 所示的  $D^t$ ), 以便后续能够基于该检测查询表征数据, 确定该待处理图像中的一些目标。

- 5        S3: 依据检测查询表征数据以及待处理图像对应的跟踪查询表征数据进行解码处理, 得到该待处理图像的目标描述信息; 该跟踪查询表征数据是依据该待处理图像对应的多个历史图像的目标描述信息所确定的; 该待处理图像的目标描述信息用于描述该待处理图像中的至少一个目标。

10       其中, 待处理图像对应的跟踪查询表征数据是指在针对该待处理图像进行目标确定处理时所需依据的跟踪 Query, 以使该跟踪查询表征数据能够表示出该待处理图像中一些已跟踪到的目标 (比如, 下文“待跟踪目标”) 在历史阶段所呈现的状态。

15       另外, 待处理图像对应的跟踪查询表征数据是依据该待处理图像对应的多个历史图像的目标描述信息所确定的。其中, 该历史图像的时序早于该待处理图像的时序; 而且本公开不限定该待处理图像对应的多个历史图像, 比如, 当该待处理图像为图 2 所示的第  $t$  帧视频图像  $I$  时, 该待处理图像对应的多个历史图像可以包括图 2 所示的第  $t-1$  帧视频图像  $I^{t-1}$ 、第  $t-2$  帧视频图像  $I^{t-2}$ 、第  $t-3$  帧视频图像  $I^{t-3}$  等。该历史图像的目标描述信息是指在针对该历史图像进行目标确定处理时所得到的结果, 以使该历史图像的目标描述信息能够描述  
20       出该历史图像中一些目标。

需要说明的是, 本公开不限定目标描述信息的实施方式, 例如, 该目标描述信息可以包括一些目标对应的查询表征数据 (比如, 该目标对应的特征向量+该目标对应的边界框)。又如, 该目标描述信息可以还包括这些目标对应的类别置信度。该类别置信度用于表示该目标以多大概率属于某一类别。

- 25       还需要说明的是, 本公开不限定上文“待处理图像对应的多个历史图像”中“多个”的实施方式, 比如, 其可以为 2 个、3 个或者其他大于 1 的正整数。

此外, 本公开不限定上文待处理图像对应的跟踪查询表征数据的确定过程的实施方式, 例如, 当该待处理图像对应的多个历史图像包括  $R$  个图像数据时, 该跟踪查询表征数据的确定过程具体可以包括: 将第  $r$  个图像数据的目标描述信息中存在的部分或者全部目标对应的查询表征数据 (Query), 确定  
30       出该历史图像中一些目标。

为该第  $r$  个图像数据对应的跟踪 Query,  $r$  为正整数,  $r \leq R$ ,  $R$  为正整数; 再将  $R$  个图像数据对应的跟踪 Query 进行集合, 得到该跟踪查询表征数据, 以使该跟踪查询表征数据包括该待处理图像对应的多个历史图像中一些已跟踪到的目标对应的跟踪 Query。

- 5 实际上, 为了更好地提高目标确定效果, 本公开还提供了上文待处理图像对应的跟踪查询表征数据的确定过程的一种可能的实施方式, 在该实施方式中, 当该待处理图像对应的多个历史图像包括第一类图像 (比如, 图 2 所示的第  $t-1$  帧视频图像  $I^{t-1}$ ) 和第二类图像 (比如, 图 2 所示的第  $t-2$  帧视频图像  $I^{t-2}$ 、第  $t-3$  帧视频图像  $I^{t-3}$  等) 时, 该待处理图像对应的跟踪查询表征数据的
- 10 确定过程可以包括下文步骤 11-步骤 12。

步骤 11: 从第一类图像的目标描述信息中提取第一跟踪目标描述信息。

- 其中, 第一类图像用于提供上文待处理图像中已跟踪到的目标的最近一次历史所处位置; 而且本公开不限定该第一类图像, 比如, 当该待处理图像与该待处理图像对应的多个历史图像属于同一个视频数据时, 该待处理图像的
- 15 时序晚于该第一类图像的时序, 而且该第一类图像的时序晚于该多个历史图像中除了该第一类图像以外其他图像 (也就是, 上文第二类图像) 的时序。

- 可见, 在一种可能的实施方式下, 当上文待处理图像与该待处理图像对应的多个历史图像属于同一个视频数据时, 上文第一类图像是指该多个历史图像中存在的、时序最接近于该待处理图像的时序的历史图像 (比如, 图 2 所示的第  $t-1$  帧视频图像  $I^{t-1}$ )。
- 20

第一跟踪目标描述信息用于描述从上文第一类图像提取到的、在确定上文待处理图像中一些目标时所需参考的最近一次历史信息, 以使该第一跟踪目标描述信息能够表示出在时序最接近于该待处理图像的时序的历史图像中出现的一些已跟踪到的目标。

- 25 另外, 本公开不限定上文第一跟踪目标描述信息的实施方式, 例如, 其可以包括至少一个已跟踪到的目标的在上文第一类图像中所呈现的目标内容特征表征数据以及目标位置表征数据, 以使该目标内容特征表征数据能够表示出该已跟踪到的目标在时序最接近于该待处理图像的时序的历史图像中所处状态, 并使得该目标位置表征数据能够表示出该已跟踪到的目标在时序最接近于该待处理图像的时序的历史图像中所处位置。
- 30

此外，本公开不限定上文第一跟踪目标描述信息的确定过程，例如，其具体可以为：在获取到上文第一类图像的目标描述信息之后，先从该目标描述信息中抽取至少一个满足预设筛选条件的 Query；再依据这些 Query，确定该第一跟踪目标描述信息，以使该第一跟踪目标描述信息包括这些 Query。其中，

5 预设筛选条件可以预先设定，例如，该预设筛选条件可以为：该 Query 被分配了跟踪对象标识（比如，ID），而且该 Query 的 IOU(Intersection over Union) 大于预设阈值。该跟踪对象标识用于唯一标识针对某个已跟踪到的目标的跟踪过程。

基于上文步骤 11 的相关内容可知，对于上文待处理图像（比如，图 2 所示的第  $t$  帧视频图像  $I^t$ ）来说，在获取到该待处理图像对应的、距离该待处理图像最近的一个历史图像（比如，图 2 所示的第  $t-1$  帧视频图像  $I^{t-1}$ ）的目标描述信息之后，可以从该目标描述信息抽取一些已跟踪到的目标的目标内容特征表征数据及其目标位置表征数据，确定为第一跟踪目标描述信息（比如，图 2 所示的  $X^t$  中  $t-1$  对应的跟踪 Query），以使该第一跟踪目标描述信息能够表示出在这些已跟踪到的目标在该历史图像中所处状态。

10 15

步骤 12：依据上文第一跟踪目标描述信息以及上文第二类图像的目标描述信息，确定上文待处理图像对应的跟踪查询表征数据。

其中，第二类图像用于代表上文待处理图像对应的多个历史图像中除了上文第一类图像以外的其他图像；而且本公开不限定该第二类图像，比如，当

20 该待处理图像与该待处理图像对应的多个历史图像属于同一个视频数据时，该待处理图像的时序晚于该第一类图像的时序，而且该第一类图像的时序晚于该第二类图像的时序。

可见，在一种可能的实施方式下，当上文待处理图像与该待处理图像对应的多个历史图像属于同一个视频数据时，上文第二类图像是指该多个历史图像中存在的时序不是最接近于该待处理图像的时序的历史图像（比如，图 2 所示的第  $t-2$  帧视频图像  $I^{t-2}$ 、第  $t-3$  帧视频图像  $I^{t-3}$  等）。

25

另外，本公开不限定上文步骤 12 的实施方式，例如，其具体可以为：先从上文第二类图像的目标描述信息中提取一些已跟踪到的目标的目标内容特征表征数据；再将该目标内容特征表征数据与上文第一跟踪目标描述信息中

30 相应目标的目标内容特征表征数据进行预设处理（比如，整合处理、融合处理

或者拼接处理等), 得到上文待处理图像对应的跟踪查询表征数据, 以使该跟踪查询表征数据不仅能够表示出这些已跟踪到的目标的在最近一帧历史图像中所处位置, 还能够比较全面的表示出这些已跟踪到的目标的在最近多帧历史图像中所呈现的特征。

- 5           实际上, 为了更好地提高目标确定效果, 本公开还提供了上文步骤 12 的另一种可能的实施方式, 其具体可以包括下文步骤 121-步骤 122。

步骤 121: 依据上文第一跟踪目标描述信息中的目标位置表征数据以及上文第二类图像的目标描述信息中的目标内容特征表征数据, 确定第二跟踪目标描述信息。

- 10           其中, 第二跟踪目标描述信息用于描述依据上文第二类图像所确定的、在确定上文待处理图像中一些目标时所需参考的跟踪 Query。

- 另外, 本公开不限定上文步骤 121 的实施方式, 例如, 其具体可以为: 将上文第一跟踪目标描述信息中的目标位置表征数据以及上文第二类图像的目标描述信息中相应目标的目标内容特征表征数据进行组合, 得到至少一个已跟踪到的目标的组合 Query, 以使该组合 Query 能够表示出该已跟踪到的目标在该第二类图像中所呈现的特征、以及该已跟踪到的目标在上文第一类图像中所处位置; 再依据这些组合 Query, 确定为第二跟踪目标描述信息 (比如, 图 2 所示的 $X^t$ 中 t-2 对应的跟踪 Query 或者 t-3 对应的跟踪 Query), 以使该第二跟踪目标描述信息包括这些组合 Query。
- 15

- 20           基于上文步骤 121 的相关内容可知, 对于上文待处理图像 (比如, 图 2 所示的第 t 帧视频图像  $I^t$ ) 来说, 在获取到该待处理图像对应的、距离该待处理图像不是最近的一个历史图像 (比如, 图 2 所示的第 t-2 帧视频图像  $I^{t-2}$ 、或者第 t-3 帧视频图像  $I^{t-3}$  等) 的目标描述信息之后, 可以先从该目标描述信息中抽取一些已跟踪到的目标的目标内容特征表征数据, 以使该目标内容特征表征数据能够表示出该已跟踪到的目标在该历史图像中所呈现的特征; 再将该目标内容特征表征数据与上文第一跟踪目标描述信息中相应目标的目标位置表征数据进行组合, 得到该已跟踪到的目标的组合 Query; 最后, 依据该组合 Query, 确定上文第二跟踪目标描述信息, 以使该第二跟踪目标描述信息能够表示出依据该历史图像所确定的、在确定上文待处理图像中一些目标时所需
- 25
- 30           参考的跟踪 Query。

步骤 122: 根据上文第二跟踪目标描述信息和上文第一跟踪目标描述信息, 确定上文待处理图像对应的跟踪查询表征数据。

需要说明的是, 本公开不限定上文步骤 122 的实施方式, 例如, 其具体可以  
以为: 针对上文第二跟踪目标描述信息与上文第一跟踪目标描述信息进行信  
5 息整合处理, 得到上文待处理图像对应的跟踪查询表征数据 (比如, 图 2 所  
示的 $X^t$ ), 以使该跟踪查询表征数据包括一些已跟踪到的目标对应的跟踪  
Query 序列 (比如, 图 2 所示的序列 $X_i^t$ 、序列 $X_{i+1}^t$ 、以及序列 $X_{i+2}^t$ 等)。

还需要说明的是, 对于图 2 中所示的  $i$ 、 $i+1$ 、以及  $i+2$  分别表示针对一个  
已跟踪到的目标所分配的跟踪对象标识,  $i$  为正整数。

10 基于上文步骤 121 至步骤 122 的相关内容可知, 在一些应用场景下, 对  
于上文待处理图像 (比如, 图 2 所示的第  $t$  帧视频图像  $I$ ) 来说, 如果该待处  
理图像对应的多个历史图像包括第一类图像 (比如, 图 2 所示的第  $t-1$  帧视频  
图像  $I^{-1}$ ) 和第二类图像 (比如, 图 2 所示的第  $t-2$  帧视频图像  $I^{-2}$ 、第  $t-3$  帧视  
频图像  $I^{-3}$  等), 则可以利用该第一类图像的目标描述信息中的目标位置表征  
15 数据以及该第二类图像的目标描述信息中的目标内容特征表征数据, 构建该  
第二类图像对应的跟踪 Query (也就是, 上文第二跟踪目标描述信息); 再结  
合该第二类图像对应的跟踪 Query 以及该第一类图像对应的跟踪 Query (也就  
是, 上文第一跟踪目标描述信息), 确定上文待处理图像对应的跟踪查询表征  
数据, 以使该跟踪查询表征数据能够更好地表示出该待处理图像中一些已跟  
20 踪到的目标在历史阶段所呈现的状态。

基于上文步骤 11 至步骤 12 的相关内容可知, 在一些应用场景下, 对于  
上文待处理图像 (比如, 图 2 所示的第  $t$  帧视频图像  $I$ ) 来说, 可以利用在历  
史时间段内针对距离该待处理图像最近的  $N_{max}$  帧历史图像 (比如, 图 2 所示  
的第  $t-1$  帧视频图像  $I^{-1}$ 、第  $t-2$  帧视频图像  $I^{-2}$ 、第  $t-3$  帧视频图像  $I^{-3}$  等) 所确  
25 定的目标内容特征表征数据、以及针对距离该待处理图像最近的一帧历史图  
像 (比如, 图 2 所示的第  $t-1$  帧视频图像  $I^{-1}$  等) 所确定的目标位置表征数据  
(比如, 边界框), 构建该待处理图像对应的跟踪查询表征数据 (比如, 图 2  
所示的 $X^t$ ), 以便后续能够利用该跟踪查询表征数据中所记录的所有跟踪  
Query 协同跟踪该待处理图像中的一些目标, 以使该跟踪查询表征数据可以被  
30 用于确定该待处理图像中出现的一些已经跟踪到的目标, 并使得上文检测查

询表征数据（比如，由图 2 所示的 $D^t$ ）被用于检测该待处理图像中新出现的目标，以便后续能够基于该跟踪查询表征数据以及该检测查询表征数据，确定该待处理图像的目标描述信息。

基于上文待处理图像对应的跟踪查询表征数据的相关内容可知，在一种可能的实施方式下，当上文待处理图像中存在至少一个目标时，如果该至少一个目标包括  $Y$  个待跟踪目标，则该跟踪查询表征数据包括第 1 个待跟踪目标在多个历史图像下分别对应的查询表征数据、第 2 个待跟踪目标在多个历史图像下分别对应的查询表征数据、……（以此类推）、以及第  $Y$  个待跟踪目标在多个历史图像下分别对应的查询表征数据。其中，第  $y$  个待跟踪目标是指在该待处理图像中存在的、而且在历史阶段已经跟踪到的目标；而且该第  $y$  个待跟踪目标在一个历史图像下对应的查询表征数据用于表征该第  $y$  个待跟踪目标在该历史图像中所呈现的状态， $y$  为正整数， $y \leq Y$ ， $Y$  为正整数。

基于上段内容可知，对于上文待处理图像对应的跟踪查询表征数据来说，如果该跟踪查询表征数据用于描述一个或者多个已跟踪到的目标的历史状态，则该跟踪查询表征数据包括每个已跟踪到的目标的至少两个跟踪 Query，以便在后续处理过程（比如，解码过程等）中，对于每个已跟踪到的目标来说，该目标对应的各个跟踪 query 分别作为一个独立的 Query 完成针对该目标的跟踪预测处理，且各 Query 在此过程中会充分利用该目标对应的其他 Query 中所携带的信息（比如，目标内容特征表征数据+目标位置表征数据等）完成针对其自身的迭代更新处理，以实现协同跟踪，以使针对各 Query 所执行的协同跟踪过程均用于共同跟踪同一个目标，如此能够更好地提高针对该目标预测所得的特征的可靠性，从而使得针对该目标预测所得的特征能够更好地表达该目标在该待处理图像中所处状态，如此有利于提高目标跟踪效果。

上文“待处理图像的目标描述信息”用于描述该待处理图像中的至少一个目标；而且本公开不限定该“待处理图像的目标描述信息”的实施方式，比如，该“待处理图像的目标描述信息”可以包括一些目标的查询表征数据（Query）以及这些目标的类别置信度。其中，该目标的查询表征数据可以包括该目标的目标内容特征表征数据以及该目标的目标位置表征数据（比如，边界框等）。该目标的类别置信度用于表示该目标分别以多大概率属于某些类别（比如，人、动物、汽车等）。

又如，当上文待处理图像对应的检测查询表征数据包括多个检测 Query，而且上文待处理图像对应的跟踪查询表征数据包括多个跟踪 Query 时，该待处理图像的目标描述信息可以包括各个检测 Query 对应的目标预测结果以及各个跟踪 Query 对应的目标预测结果。其中，第  $j$  个检测 Query 对应的目标预测结果是指依据该第  $j$  个检测 Query 推导出一个目标在该待处理图像中所处状态； $j$  为正整数， $j \leq J$ ， $J$  为正整数， $J$  表示该检测 Query 的个数。第  $m$  个跟踪 Query 对应的目标预测结果是指依据该第  $m$  个跟踪 Query 推导出一个目标在该待处理图像中所处状态； $m$  为正整数， $m \leq M$ ， $M$  为正整数， $M$  表示该跟踪 Query 的个数。

另外，本公开不限定上文“待处理图像的目标描述信息”的确定过程（也就是，上文 S3 的实施方式），例如，其可以借助预先构建的解码器进行实施。其中，该解码器用于针对该解码器的输入数据进行解码处理；而且本公开不限定该解码器，比如，该解码器可以采用图 2 所示的多目标跟踪模型中的解码模块进行实施。

此外，在一些应用场景下，当上文 S2 采用“对特征提取结果进行编码处理，得到图像特征表征数据以及检测查询表征数据”这一步骤进行实施时，上文 S3 具体可以为：依据该图像特征表征数据、该检测查询表征数据以及上文待处理图像对应的跟踪查询表征数据进行解码处理，得到该待处理图像的目标描述信息。需要说明的是，该段中所示的 S3 的实施方式也可以借助预先构建的解码器进行实施。

还有，为了更好地提高目标确定效果，本公开还提供了上文 S3 的一种可能的实施方式，其具体可以包括下文步骤 21-步骤 22。

步骤 21：依据上文检测查询表征数据和上文跟踪查询表征数据，确定待处理查询表征数据。

其中，待处理查询表征数据用于描述在针对上文待处理图像进行目标确定处理时所需参考的所有 Query（比如，多个检测 Query+多个跟踪 Query），以使该待处理查询表征数据能够更好地表示出该待处理图像中一些目标所具有的特点。

另外，本公开不限定上文待处理查询表征数据的确定过程，例如，其具体可以为：将上文检测查询表征数据和上文跟踪查询表征数据进行级联处理（也

就是，拼接处理），得到待处理查询表征数据，以使该待处理查询表征数据包括该检测查询表征数据和该跟踪查询表征数据，从而使得该待处理查询表征数据能够更好地表示出该待处理图像中一些目标所具有的特点。

5 步骤 22：依据上文待处理查询表征数据进行解码处理，得到待处理图像的目标描述信息。

需要说明的是，本公开不限定上文步骤 22 的实施方式，例如，其具体可以为：将上文待处理查询表征数据输入预先构建的解码器（比如，图 2 所示的多目标跟踪模型中的解码模块），以使该解码器能够依据该待处理查询表征数据进行解码处理，得到并输出上文待处理图像的目标描述信息（比如，由图 10 2 所示的最后一个解码层所输出的数据），以使该目标描述信息能够描述出该待处理图像中的一些目标。

又如，在一种可能的实施方式下，当上文 S2 采用“对特征提取结果进行编码处理，得到图像特征表征数据以及检测查询表征数据”这一步骤进行实施时，该步骤 22 具体可以为：依据该图像特征表征数据和上文待处理查询表征数据 15 数据进行解码处理，得到待处理图像的目标描述信息。

需要说明的是，本公开不限定上段中步骤“依据该图像特征表征数据和上文待处理查询表征数据进行解码处理，得到待处理图像的目标描述信息”的实施方式，例如，其具体可以为：将该图像特征表征数据和该待处理查询表征数据输入预先构建的解码器（比如，图 2 所示的多目标跟踪模型中的解码模块）， 20 以使该解码器能够依据该图像特征表征数据和该待处理查询表征数据进行解码处理，得到并输出上文待处理图像的目标描述信息（比如，由图 2 所示的最后一个解码层所输出的数据），以使该目标描述信息能够描述出该待处理图像中的一些目标。

基于上文步骤 21 至步骤 22 的相关内容可知，在一些应用场景下，对于 25 待处理图像来说，在获取到该待处理图像对应的检测查询表征数据以及该待处理图像对应的跟踪查询表征数据之后，可以先将这两种 Query 进行级联处理；再将级联结果直接送入预先构建的解码器，以使该解码器能够输出该待处理图像的目标描述信息。

基于上段内容可知，在一种可能的实施方式下，当本公开提供的图像处理 30 方法借助预先训练好的目标跟踪模型（比如，图 2 所示的多目标跟踪模块）

进行实现时，上文 S3 具体可以为：对于上文待处理图像（比如，图 2 所示的第  $t$  帧视频图像  $I$ ）来说，在获取到该待处理图像对应的跟踪查询表征数据（比如，图 2 所示的  $X^t$ ）、由该目标跟踪模型中的编码模块输出的该待处理图像对应的检测查询表征数据（比如，图 2 所示的  $D^t$ ）以及该待处理图像对应的图像特征表征数据（比如，图 2 所示的图像特征图）之后，可以先将该跟踪查询表征数据与该检测查询表征数据进行级联处理；再由该目标跟踪模型中的解码模块依据该级联结果以及该图像特征表征数据进行解码处理，以使该解码模块能够通过自回归的方式得到并输出该待处理图像的目标描述信息（比如，由图 2 所示的最后一个解码层所输出的数据），以使该目标描述信息包括一些目标的 Query 以及这些目标的类别置信度。

基于上文内容可知，在一种可能的实施方式下，上文待处理图像的目标描述信息是利用预先训练好的解码器所确定的。其中，该解码器可以用于依据上文检测查询表征数据以及上文待处理图像对应的跟踪查询表征数据进行解码处理，得到该待处理图像的目标描述信息；或者，该解码器也可以用于依据上文图像特征表征数据、该检测查询表征数据以及该待处理图像对应的跟踪查询表征数据进行解码处理，得到该待处理图像的目标描述信息。

另外，为了更好地提高目标确定效果，本公开还提供了上文解码器的一种可能的实施方式，在该实施方式下，该解码器可以包括至少一个解码层，而且该解码层中的自注意力模块是利用第一注意力掩模进行实现的，该第一注意力掩模用于屏蔽属于同一个目标的在不同历史图像下对应的查询表征数据之间的交互，以使该解码层达到时序阻隔的效果。需要说明的是，本公开不限定该解码层的个数，比如，其可以为 6 个。

需要说明的是，本公开不限定上段中“同一个目标的在不同历史图像下对应的查询表征数据”的实施方式，例如，当上段所示的解码器是指第  $h$  个解码器时，该“同一个目标的在不同历史图像下对应的查询表征数据”是指在该第  $h$  个解码器中的自注意力模块的输入数据中存在的对应于同一个目标的不同跟踪 Query； $h$  为正整数， $h \leq$  解码器的个数（比如，6）。

基于上段内容可知，在一种可能的实施方式下，上文解码器可以是由多个（比如，6 个）时序阻隔的解码层(temporal blocking decoder)堆叠而成的，而且每个时序阻隔的解码层中的自注意力模块均是利用第一注意力掩模进行实

现的，以使各个自注意力模块均能够借助该第一注意力掩模实现屏蔽属于同一个目标的在不同历史图像下对应的查询表征数据之间的交互，以避免同一个目标的不同跟踪 Query 之间相互抑制。

- 上文第一注意力掩模是指针对上文解码层中的自注意力模块所配置的注意力掩模，以使该自注意力模块能够借助该注意力掩模实现屏蔽属于同一个目标的在不同历史图像下对应的查询表征数据之间的交互；而且该第一注意力掩模被配置为屏蔽属于同一个目标的在不同历史图像下对应的查询表征数据之间的交互。另外，本公开不限定该第一注意力掩模的实施方式，例如，图 2 所示的解码层中的自注意力模块中的第一注意力掩模可以采用下文表 1 所示的注意力掩模进行实施。

|                 | $d_1^t$  | $d_2^t$  | $\dots$  | $d_{N_{det}}^t$ | $X_i^t(1)$ | $X_{i+1}^t(1)$ | $X_{i+1}^t(2)$ | $X_{i+1}^t(3)$ | $X_{i+2}^t(1)$ | $X_{i+2}^t(2)$ |
|-----------------|----------|----------|----------|-----------------|------------|----------------|----------------|----------------|----------------|----------------|
| $d_1^t$         | 1        | 1        | $\dots$  | 1               | 1          | 1              | 1              | 1              | 1              | 1              |
| $d_2^t$         | 1        | 1        | $\dots$  | 1               | 1          | 1              | 1              | 1              | 1              | 1              |
| $\vdots$        | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$        | $\vdots$   | $\vdots$       | $\vdots$       | $\vdots$       | $\vdots$       | $\vdots$       |
| $d_{N_{det}}^t$ | 1        | 1        | $\dots$  | 1               | 1          | 1              | 1              | 1              | 1              | 1              |
| $X_i^t(1)$      | 1        | 1        | $\dots$  | 1               | 1          | 1              | 1              | 1              | 1              | 1              |
| $X_{i+1}^t(1)$  | 1        | 1        | $\dots$  | 1               | 1          | 1              | 0              | 0              | 1              | 1              |
| $X_{i+1}^t(2)$  | 1        | 1        | $\dots$  | 1               | 1          | 0              | 1              | 0              | 1              | 1              |
| $X_{i+1}^t(3)$  | 1        | 1        | $\dots$  | 1               | 1          | 0              | 0              | 1              | 1              | 1              |
| $X_{i+2}^t(1)$  | 1        | 1        | $\dots$  | 1               | 1          | 1              | 1              | 1              | 1              | 0              |
| $X_{i+2}^t(2)$  | 1        | 1        | $\dots$  | 1               | 1          | 1              | 1              | 1              | 0              | 1              |

表 1 解码层中的自注意力模块中的注意力掩模的一种可能的实施方式

- 需要说明的是，对于上文表 1 来说， $X_i^t(1)$ 表示以  $i$  作为跟踪对象标识的目标在图像数据  $I^{-1}$  下所对应的跟踪 Query； $X_{i+1}^t(1)$ 表示以  $i+1$  作为跟踪对象标识的目标在图像数据  $I^{-1}$  下所对应的跟踪 Query； $X_{i+1}^t(2)$ 表示以  $i+1$  作为跟踪对象标识的目标在图像数据  $I^{-2}$  下所对应的跟踪 Query； $X_{i+1}^t(3)$ 表示以  $i+1$  作为跟踪对象标识的目标在图像数据  $I^{-3}$  下所对应的跟踪 Query；...（以此类推）。

- 基于上述解码器的相关内容可知，在一种可能的实施方式下，上文解码器可以是由多个时序阻隔的解码层堆叠而成的，而且可以通过将每个时序阻隔的解码层内自注意力模块中的注意力掩模被配置为屏蔽属于同一个目标的不同跟踪 Query 之间做交互的方式，以避免同一个目标的不同跟踪 Query 之间

相互抑制。例如，对于该解码层中的自注意力模块来说，该自注意力模块可以不仅用于进行不同目标的 Query 之间的交互处理，还用于进行同一个 Query 之间的交互处理；但是该自注意力模块会屏蔽同一个目标下的不同跟踪 Query 之间的交互处理。

- 5       可见，由于在传统的自注意力模块中可以让该自注意力模块各项输入信息进行平等的交互，故为了有效地避免同一个目标的多个跟踪 query 彼此抑制，可以向该自注意力模块中添加掩模（比如，上文第一注意力掩模或者上文表 1 所示的掩模），以得到上文解码层中的自注意力模块，以使该解码层中的自注意力模块就是添加了掩模的自注意力模块。其中，因在添加了掩模的自
- 10   注意力模块中能够屏蔽属于同一个目标的不同跟踪 Query 之间做交互，如此能够有效地避免在添加了掩模的自注意力模块中出现同一个目标的多个跟踪 query 彼此抑制这一现象，从而有利于提高目标跟踪效果。

- 实际上，为了更好地提高解码效果，本公开还提供了上文解码器的另一种可能的实施方式，在该实施方式中，当该解码器包括至少一个解码层，而且该
- 15   至少一个解码层包括第一解码层和第二解码层时，该解码器可以还包括特征融合模块（比如，图 2 所示的特征融合模块）。其中，该第一解码层用于代表该至少一个解码层中存在的任意两个相邻解码层中位置靠前的一个解码层。该第二解码层用于代表该至少一个解码层中存在的任意两个相邻解码层中位置靠后的一个解码层。该特征融合模块位于第一解码层与第二解码层之间。为
- 20   了便于理解，下面结合示例进行说明。

- 作为示例，当上文解码器包括依次排列的 6 个解码层时，如果第一解码层为该解码器中的第 1 个解码层，则第二解码层为该解码器中的第 2 个解码层，而且在该第 1 个解码层与该第 2 个解码层之间存在一个特征融合模块；如果第一解码层为该解码器中的第 2 个解码层，则第二解码层为该解码器中的
- 25   第 3 个解码层，而且在该第 2 个解码层与该第 3 个解码层之间存在一个特征融合模块；……（以此类推）。

- 基于上述两段内容可知，对于上文解码器中存在的任意一个三元组（第一解码层，特征融合模块，第二解码层）来说，该三元组中不同元素之间的关联关系具体可以为：该特征融合模块用于针对由该第一解码层输出的跟踪查询
- 30   处理数据进行融合处理，得到跟踪查询融合数据；而且该第二解码层用于针对

由该特征融合模块输出的跟踪查询融合数据以及由该第一解码层输出的检测查询处理数据进行处理。其中，该跟踪查询处理数据是依据所述第一解码层以及上文跟踪查询表征数据所确定的；该跟踪查询处理数据包括由该第一解码层针对各待跟踪目标分别输出的处理后的查询表征数据；该跟踪查询融合数据包括由该特征融合模块针对各处理后的查询表征数据分别输出的融合后的查询表征数据；该检测查询处理数据是依据该第一解码层以及上文检测查询表征数据所确定的。

上文跟踪查询处理数据是指由上文第一解码层针对已跟踪到的目标所输出的 Query；而且该跟踪查询处理数据是依据该第一解码层以及上文待处理图像对应的跟踪查询表征数据所确定的。例如，如果该第一解码层为上文解码器中的第 1 个解码层，则由该第一解码层所输出的跟踪查询处理数据是指由该第 1 个解码层针对上文跟踪查询表征数据进行处理所得到的 Query，以使由该第一解码层所输出的跟踪查询处理数据包括由该第 1 个解码层针对各个已跟踪到的目标分别输出的处理后的查询表征数据；如果该第一解码层为上文解码器中的第 2 个解码层，则由该第一解码层所输出的跟踪查询处理数据是指由该第 2 个解码层针对该第 2 个解码层的输入数据中存在的用于描述已跟踪到的目标的 Query 进行处理所得到的 Query，以使由该第一解码层所输出的跟踪查询处理数据包括由该第 2 个解码层针对各个已跟踪到的目标分别输出的处理后的查询表征数据；如果该第一解码层为上文解码器中的第 3 个解码层，则由该第一解码层所输出的跟踪查询处理数据是指由该第 3 个解码层针对该第 3 个解码层的输入数据中存在的用于描述已跟踪到的目标的 Query 进行处理所得到的 Query，以使由该第一解码层所输出的跟踪查询处理数据包括由该第 3 个解码层针对各个已跟踪到的目标分别输出的处理后的查询表征数据；……（以此类推）。其中，该处理后的查询表征数据是指由该第一解码层针对一个已跟踪到的目标迭代所得的 Query。

上文检测查询处理数据是指由上文第一解码层针对待检测的目标所输出的 Query；而且该检测查询处理数据是依据该第一解码层以及上文检测查询表征数据所确定的。例如，如果该第一解码层为上文解码器中的第 1 个解码层，则由该第一解码层所输出的检测查询处理数据是指由该第 1 个解码层针对上文检测查询表征数据进行处理所得到的 Query；如果该第一解码层为上文解码

器中的第 2 个解码层, 则由该第一解码层所输出的检测查询处理数据是指由该第 2 个解码层针对该第 2 个解码层的输入数据中存在的用于描述待检测到的目标的 Query 进行处理所得到的 Query; 如果该第一解码层为上文解码器中的第 3 个解码层, 则由该第一解码层所输出的检测查询处理数据是指由该第 3 个解码层针对该第 3 个解码层的输入数据中存在的用于描述待检测到的目标的 Query 进行处理所得到的 Query; ... (以此类推)。

上文跟踪查询融合数据是指由上文特征融合模块针对由上文第一解码层所输出的与各个已跟踪到的目标相关联的 Query 进行融合所得到的融合结果, 以使该跟踪查询融合数据包括由该特征融合模块针对各已跟踪到的目标分别输出的融合后的查询表征数据, 从而使得该跟踪查询融合数据能够更好地表示出各个已跟踪到的目标所呈现的特点。其中, 该融合后的查询表征数据是指将该已跟踪到的目标以及与该已跟踪到的目标相关联的 Query 进行融合所得到的融合结果。比如, 对于已跟踪到的目标 A 来说, 如果该目标 A 对应于 3 个跟踪查询处理数据  $Query_{A-1}$ ,  $Query_{A-2}$ ,  $Query_{A-3}$ , 则在该  $Query_{A-1}$  对应融合过程中, 需要借助融合模块将该  $Query_{A-1}$  与  $Query_{A-2}$ ,  $Query_{A-3}$  所携带的信息进行交互, 得到该  $Query_{A-1}$  对应的融合后的查询表征数据  $Query_{A-1}'$ ; 在该  $Query_{A-2}$  对应融合过程中, 需要借助融合模块将该  $Query_{A-2}$  与  $Query_{A-1}$ ,  $Query_{A-3}$  所携带的信息进行交互, 得到该  $Query_{A-2}$  对应的融合后的查询表征数据  $Query_{A-2}'$ ; 在该  $Query_{A-3}$  对应融合过程中, 需要借助融合模块将该  $Query_{A-3}$  与  $Query_{A-1}$ ,  $Query_{A-2}$  所携带的信息进行交互, 得到该  $Query_{A-3}$  对应的融合后的查询表征数据  $Query_{A-3}'$ 。

另外, 本公开不限定上文特征融合模块的实施方式, 例如, 在一些应用场景下, 该特征融合模块可以包括多个融合子模块, 不同融合子模块分别用于针对不同已跟踪到的目标所关联的跟踪 Query 进行融合处理。可见, 该融合子模块的个数可以依据已跟踪到的目标的个数进行确定。又如, 在另一种可能的实施方式下, 该特征融合模块可以用于执行以下步骤: 将第 1 个已跟踪到的目标所关联的跟踪 Query 进行融合处理, 得到并输出该第 1 个已跟踪到的目标对应的融合后的查询表征数据, 以使该第 1 个已跟踪到的目标对应的融合后的查询表征数据包括该第 1 个已跟踪到的目标所关联的各个跟踪 Query 分别对应的融合结果, 从而使得该第 1 个已跟踪到的目标对应的融合后的查询

表征数据中的 Query 个数与该第 1 个已跟踪到的目标所关联的跟踪 Query 个数保持一致;再将第 2 个已跟踪到的目标所关联的跟踪 Query 进行融合处理,得到并输出该第 2 个已跟踪到的目标对应的融合后的查询表征数据,以使该第 2 个已跟踪到的目标对应的融合后的查询表征数据包括该第 2 个已跟踪到的目标所关联的各个跟踪 Query 分别对应的融合结果,从而使得该第 2 个已跟踪到的目标对应的融合后的查询表征数据中的 Query 个数与该第 2 个已跟踪到的目标所关联的跟踪 Query 个数保持一致;.....(以此类推)。

实际上,为了更好地提高融合效率,本公开还提供了上文特征融合模块的一种可能的实施方式,在该实施方式下,该特征融合模块包括信息移除分支网络和信息添加分支网络;而且利用该特征融合模块确定上文跟踪查询融合数据的过程,具体可以包括下文步骤 31-步骤 33。

步骤 31: 利用上文信息移除分支网络对上文跟踪查询处理数据进行处理,得到保留信息特征。

其中,信息移除分支网络用于针对上文特征融合模块的输入数据进行信息移除处理;而且本公开不限定该信息移除分支网络,比如,该信息移除分支网络可以包括两个自注意力层、一个全连接层以及一个 sigmoid 层。

保留信息特征用于表示通过针对上文特征融合模块的输入数据进行信息移除处理之后所保留下来信息。

另外,本公开不限定上文保留信息特征的确定过程,例如,其具体可以包括下文步骤 311-步骤 315。

步骤 311: 对上文跟踪查询处理数据进行时序信息提取处理,得到第一时序信息,以使该第一时序信息能够表示出该跟踪查询处理数据中所携带的时序信息。

需要说明的是,本公开不限定步骤 311 的实施方式,例如,该步骤 311 可以借助一个自注意力层进行实施。

基于上文步骤 311 的相关内容可知,对于上文信息移除分支网络来(比如,图 3 所示的信息移除分支网络)说,在该信息移除分支网络接收到由上文第一解码层所输出的跟踪查询处理数据(比如,图 3 所示的 $\bar{X}^t$ )之后,可以由该信息移除分支网络中的第一个自注意力层针对该跟踪查询处理数据进行时序信息提取处理,得到第一时序信息(比如,图 3 所示的时序信息 1),

以使该第一时序信息能够表示出该跟踪查询处理数据中所携带的时序信息。

需要说明的是，对于图 3 所示的 $\bar{X}^t$ 来说，该 $\bar{X}^t$ 可以表示由图 2 中第 1 个解码层针对图 2 所示的 $X^t$ 进行处理所得到的结果；而且 $\bar{X}^t = (F^t, \hat{B}^t)$ ， $F^t$ 用于描述由该第 1 个解码层针对一些已跟踪到的目标预测所得的特征信息； $\hat{B}^t$ 表示由该第 1 个解码层针对一些已跟踪到的目标预测所得的位置信息。

步骤 312: 对上文第一时序信息与上文跟踪查询处理数据进行自注意力处理，得到自注意力处理结果。

需要说明的是，本公开不限定步骤 312 的实施方式，例如，该步骤 312 可以借助一个自注意力层进行实施。

10 基于上文步骤 312 的相关内容可知，对于上文信息移除分支网络（比如，图 3 所示的信息移除分支网络）来说，在由该信息移除分支网络中的第一个自注意力层输出上文第一时序信息（比如，图 3 所示的时序信息 1）之后，可以由该信息移除分支网络中的第二个自注意力层依据该第一时序信息与上文跟踪查询处理数据进行自注意力处理，得到自注意力处理结果，以使该自注意力处理结果能够表示出需要将哪些信息进行剔除处理。

15 步骤 313: 对上文自注意力处理结果进行全连接处理，得到全连接处理结果。

需要说明的是，本公开不限定步骤 313 的实施方式，例如，该步骤 313 可以借助一个全连接层进行实施。

20 基于上文步骤 313 的相关内容可知，对于上文信息移除分支网络（比如，图 3 所示的信息移除分支网络）来说，在由该信息移除分支网络中的第二个自注意力层输出上文自注意力处理结果之后，可以由该信息移除分支网络中的全连接层针对该自注意力处理结果进行全连接处理，得到全连接处理结果。

25 步骤 314: 依据上文全连接处理结果，确定待移除信息表征数据，以使该待移除信息表征数据能够表示出上文跟踪查询处理数据中哪些信息需要被移除。

需要说明的是，本公开不限定步骤 314 的实施方式，例如，该步骤 314 可以借助一个 sigmoid 层进行实施。

30 基于上文步骤 314 的相关内容可知，对于上文信息移除分支网络（比如，图 3 所示的信息移除分支网络）来说，在由该信息移除分支网络中的全连接

层输出上文全连接处理结果之后,可以由该信息移除分支网络中的 sigmoid 层针对该全连接处理结果进行处理,得到待移除信息表征数据(比如,图 3 所示的  $Z$ ),以使该待移除信息表征数据能够表示出上文跟踪查询处理数据中哪些信息需要被移除。

- 5        步骤 315: 依据上文待移除信息表征数据,对上文跟踪查询处理数据进行信息移除处理,得到保留信息特征。

需要说明的是,本公开不限定步骤 315 的实施方式,例如,该步骤 315 具体可以为:先依据上文待移除信息表征数据,确定待保留信息表征数据(比如,图 3 所示的  $1-Z$ );再将上文跟踪查询处理数据中的目标内容特征表征数据与  
10    该待保留信息表征数据相乘,得到保留信息特征(例如,图 3 所示的  $F^t \times (1-Z)$ )。

基于上文步骤 311 至步骤 315 的相关内容可知,在一种可能的实施方式下,上文信息移除分支网络可以包括两个自注意力层、一个全连接层以及一个 sigmoid 层;其中,第 1 个自注意力层用于搜集时序信息,第 2 个自注意力层  
15    用于融合多个 Query 的特征,该全连接层用于针对该第 2 个自注意力层的输出数据进行全连接处理;该 sigmoid 层用于针对该全连接层的输出数据进行处理,以得到门控值(也就是,上文待移除信息表征数据),以便后续能够基于该门控值(比如,图 3 所示的  $Z$ ),针对该信息移除分支网络的输入数据进行信息移除处理。

- 20        基于上文步骤 31 的相关内容可知,对于上文特征融合模块来说,如果该特征融合模块包括信息移除分支网络,则在该特征融合模块接收到由上文第一解码层所输出的跟踪查询处理数据(比如,图 3 所示的  $\bar{X}^t$ )之后,由该信息移除分支网络针对该跟踪查询处理数据进行信息移除处理,得到保留信息特征(例如,图 3 所示的  $F^t \times (1-Z)$ ),以便后续基于该保留信息特征,确定  
25    出针对该跟踪查询处理数据的 Query 融合结果。

步骤 32: 利用上文信息添加分支网络对上文跟踪查询处理数据进行处理,得到待添加信息特征。

- 其中,信息添加分支网络用于从上文特征融合模块的输入数据中提取出所需添加的信息;而且本公开不限定该信息添加分支网络,比如,该信息添加  
30    分支网络可以包括两个自注意力层。

待添加信息特征是指在进行信息添加处理时所需依据的信息。

另外，本公开不限定上文待添加信息特征的确定过程，例如，其具体可以包括下文步骤 321-步骤 322。

5 步骤 321: 对上文跟踪查询处理数据进行时序信息提取处理，得到第二时序信息，以使该第二时序信息能够表示出该跟踪查询处理数据中所携带的时序信息。

需要说明的是，本公开不限定步骤 321 的实施方式，例如，该步骤 321 可以借助一个自注意力层进行实施。

10 基于上文步骤 321 的相关内容可知，对于上文信息添加分支网络（比如，图 3 所示的信息添加分支网络）来说，在该信息添加分支网络接收到由上文第一解码层所输出的跟踪查询处理数据（比如，图 3 所示的 $\bar{X}^t$ ）之后，可以由该信息添加分支网络中的第一个自注意力层针对该跟踪查询处理数据进行时序信息提取处理，得到第二时序信息（比如，图 3 所示的时序信息 2），以使该第二时序信息能够表示出该跟踪查询处理数据中所携带的时序信息。

15 步骤 322: 对上文第二时序信息与上文跟踪查询处理数据进行自注意力处理，得到待添加信息特征。

需要说明的是，本公开不限定步骤 322 的实施方式，例如，该步骤 322 可以借助一个自注意力层进行实施。

20 基于上文步骤 322 的相关内容可知，对于上文信息移除分支网络（比如，图 3 所示的信息添加分支网络）来说，在由该信息添加分支网络中的第一个自注意力层输出上文第二时序信息（比如，图 3 所示的时序信息 2）之后，可以由该信息移除分支网络中的第二个自注意力层依据该第二时序信息与上文跟踪查询处理数据进行自注意力处理，得到待添加信息特征（比如，图 3 所示的 $F_{add}^t$ ），以使该待添加信息特征能够表示出在进行信息添加处理时所需依据的信息。

25 基于上文步骤 321 至步骤 322 的相关内容可知，在一种可能的实施方式下，上文信息添加分支网络可以包括两个自注意力层；其中，第 1 个自注意力层用于搜集时序信息，第 2 个自注意力层用于融合多个 Query 的特征。

30 基于上文步骤 32 的相关内容可知，对于上文特征融合模块来说，如果该特征融合模块包括信息添加分支网络，则在该特征融合模块接收到由上文第

一解码层所输出的跟踪查询处理数据（比如，图 3 所示的 $\bar{X}^t$ ）之后，由该信息添加分支网络针对该跟踪查询处理数据进行信息提取处理，得到待添加信息特征（比如，图 3 所示的 $F_{add}^t$ ），以便后续基于该待添加信息特征，确定出针对该跟踪查询处理数据的 Query 融合结果。

- 5        步骤 33：根据上文保留信息特征和上文待添加信息特征，确定上文跟踪查询融合数据。

需要说明的是，本公开不限定上文步骤 33 的实施方式，例如，其可以采用下文公式（2）进行实施。

$$F_r^t = LN(2F^t \times (1 - Z) + F_{add}^t) \quad (2)$$

- 10        式中， $F_r^t$ 表示上文特征融合模块的输出结果（比如，上文跟踪查询融合数据）； $LN()$ 表示层级标准化处理； $F^t$ 表示上文特征融合模块的输入数据中的目标内容特征表征数据； $Z$ 表示由上文特征融合模块中信息移除分支网络内 sigmoid 层所输出的门控值； $F^t \times (1 - Z)$ 表示由上文特征融合模块中信息移除分支网络所输出的数据（比如，上文保留信息特征）； $F_{add}^t$ 表示由上文特征融合模块中信息添加分支网络内第二个自注意力层所输出的数据（比如，上文待添加信息特征）。
- 15

- 基于上文步骤 31 至步骤 33 的相关内容可知，对于上文特征融合模块来说，该特征融合模块可以包括两个分支网络，即信息移除分支网络和信息添加分支网络。每个分支网络都包含两个叠加的自注意力层，第一个自注意力层用于搜集时序信息，第二个自注意力层用于根据该时序信息融合多个 Query 的特征。另外，对于信息移除分支网络来说，该信息移除分支网络中的第二个自注意力层的输出数据会进一步经过一个全连接层和一个 sigmoid 层得到门控值，该门控值与该信息添加分支网络的输出数据共同决定了融合后的 Query 信息（如上文公式（2）所示）。可见，因该特征融合模块是通过两个分支来模拟信息交互过程的，以使对于任一目标的某个跟踪 Query 来说，该特征融合模块能够更高效地将该目标除了该某个跟踪 Query 以外的其他跟踪 Query 所携带的有效信息整合至该某个跟踪 Query 内，以完成针对该某个跟踪 Query 的自我迭代更新，如此有利于提高同一个目标的不同跟踪 Query 之间的交互效率及效果，从而有利于提高目标跟踪效果。
- 25

- 30        另外，对于上文特征融合模块来说，为了确保该特征融合模块可以同时执

行针对所有已跟踪到的目标的 Query 融合处理，本公开还提供了该特征融合模块的一种可能的实施方式，在该实施方式中，该特征融合模块中的自注意力层是利用第二注意力掩模进行实现的，该第二注意力掩模用于屏蔽属于不同目标的查询表征数据之间的交互，如此能够确保在该特征融合模块中只进行

5 同一个目标的所有跟踪 Query 之间的交互。

其中，第二注意力掩模是指针对上文特征融合模块中的自注意力层所配置的注意力掩模；而且该第二注意力掩模被配置为屏蔽属于不同目标的查询表征数据之间的交互。另外，本公开不限定该第二注意力掩模的实施方式，例如，图 2 所示的特征融合模块中的自注意力层中的第二注意力掩模可以采用

10 下文表 2 所示的注意力掩模进行实施。

需要说明的是，对于上文特征融合模块中的每个分支网络（比如，信息移除分支网络或者信息添加分支网络）来说，因该分支网络包括两个自注意力层，故在一种可能的实施方式下，该分支网络中每个自注意力层所使用的第二注意力掩模分别可以采用类似于下文表 2 所示的注意力掩模进行实施。

15 还需要说明的是，对于下文表 2 来说，该表 2 所涉及的 $X_i^t(1)$ 、 $X_{i+1}^t(1)$ 、 $X_{i+1}^t(2)$ 、 $X_{i+1}^t(3)$ 、 $X_{i+2}^t(1)$ 、以及 $X_{i+2}^t(2)$ 的相关内容请参见上文。

|                | $X_i^t(1)$ | $X_{i+1}^t(1)$ | $X_{i+1}^t(2)$ | $X_{i+1}^t(3)$ | $X_{i+2}^t(1)$ | $X_{i+2}^t(2)$ |
|----------------|------------|----------------|----------------|----------------|----------------|----------------|
| $X_i^t(1)$     | 1          | 0              | 0              | 0              | 0              | 0              |
| $X_{i+1}^t(1)$ | 0          | 1              | 1              | 1              | 0              | 0              |
| $X_{i+1}^t(2)$ | 0          | 1              | 1              | 1              | 0              | 0              |
| $X_{i+1}^t(3)$ | 0          | 1              | 1              | 1              | 0              | 0              |
| $X_{i+2}^t(1)$ | 0          | 0              | 0              | 0              | 1              | 1              |
| $X_{i+2}^t(2)$ | 0          | 0              | 0              | 0              | 1              | 1              |

表 2 特征融合模块中的自注意力层中的注意力掩模的一种可能的实施方式

基于上文特征融合模块的相关内容可知，该特征融合模块可以用于同时

20 执行针对所有已跟踪到的目标的 Query 融合处理；而且可以通过将该特征融合模块内自注意力层中的注意力掩模配置为避免不同目标的 Query 之间进行交互的方式，来确保在该特征融合模块内只进行同一个目标的 Query 之间交互。

另外，对于上文特征融合模块中的每个分支网络（比如，信息移除分支网

25 络或者信息添加分支网络）来说，因该分支网络包括两个自注意力层，故该分

支网络中第  $g$  个自注意力层所使用的第二注意力掩模用于屏蔽该第  $g$  个自注意力层的输入数据中出现的不同目标的查询表征数据之间的交互,  $g$  为正整数,  $g \leq 2$ 。

5 基于上文解码器的相关内容可知, 在一种可能的实施方式下, 该解码器可以由多个时序阻隔的解码层堆叠而成, 而且在任意两个相邻解码层之间会插入一个特征融合模块, 如此能够确保该解码器具有更好地解码效果。

10 基于上文 S1 至 S3 的相关内容可知, 对于本公开实施例提供的图像处理方法来说, 先对待处理图像进行特征提取处理, 得到特征提取结果, 以使该特征提取结果用于表征该待处理图像所携带的图像信息; 再对该特征提取结果进行编码处理, 至少得到检测查询表征数据(比如, 多个检测 Query); 最后, 至少依据该检测查询表征数据以及该待处理图像对应的跟踪查询表征数据(比如, 基于多个历史图像所确定的跟踪 Query)进行解码处理, 得到该待处理图像的目标描述信息(比如, 目标内容特征表征数据、目标位置表征数据、以及类别置信度等), 以使该目标描述信息能够表示出该待处理图像中一些目标。其中, 因该跟踪查询表征数据是依据该待处理图像对应的多个历史图像的目标描述信息所确定的, 以使该跟踪查询表征数据能够更好地表示出该待处理图像中一些目标的历史状态, 从而使得基于该跟踪查询表征数据所确定的目标描述信息能够更好地表示出该待处理图像中一些目标, 如此有利于提高目标跟踪效果。

20 实际上, 在一些应用场景下, 在获取到一个图像数据的目标描述信息之后, 可以从该目标描述信息中挑选出一些高质量的目标预测结果作为该图像数据的目标确定结果。基于此, 本公开还提供了上文图像处理方法的一种可能的实施方式, 在该实施方式下, 当上文待处理图像的目标描述信息包括至少一个目标预测结果(比如, 多个检测 Query 对应的目标预测结果+多个跟踪 Query 对应的目标预测结果)时, 该图像处理方法除了包括上文 S1-S3 以外, 可以还包括下文步骤 41。其中, 该步骤 41 的执行时间晚于上文 S3 的执行时间。

步骤 41: 在得到待处理图像的目标描述信息之后, 从该目标描述信息中提取满足预设参考条件的目标预测结果, 得到该待处理图像的目标确定结果。

30 其中, 预设参考条件是指在从上文待处理图像的目标描述信息中筛选高质量的目标预测结果时所需依据的条件; 而且本公开不限定该预设参考条件,

例如，其可以为：属于检测 Query 对应的目标预测结果或者属于最近一个历史图像的跟踪 Query 对应的目标预测结果，而且类别置信度达到预设阈值。

待处理图像的目标确定结果用于描述该待处理图像中一些预测比较准确的目标。

- 5       另外，本公开不限定上文待处理图像的目标确定结果的确定过程，例如，其具体可以为：在得到待处理图像的目标描述信息之后，从该目标描述信息中提取出检测 Query 对应的目标预测结果和最近一个历史图像的跟踪 Query 对应的目标预测结果，均作为待处理预测结果；再判断每个待处理预测结果中的类别置信度是否达到预设阈值，若达到，则可以确定该待处理预测结果所描述
- 10       的目标比较准确，故可以依据该待处理预测结果，确定该待处理图像的目标确定结果，以使该目标确定结果包括该待处理预测结果。

- 基于上文步骤 41 的相关内容可知，在一些应用场景下，在确定出一个图像数据的目标描述信息之后，从该目标描述信息中提取一些高质量的目标预测结果，得到该图像数据的目标确定结果，以使该目标确定结果能够更好地表示出该图像数据中存在哪些目标，如此有利于提高目标确定效果。
- 15

- 实际上，当本公开实施例提供的图像处理方法应用于目标跟踪领域时，在获取到某一帧图像数据的目标描述信息之后，可以进一步利用该目标描述信息，确定下一帧图像对应的跟踪查询表征数据，如此有利于提高针对下一帧图像的目标确定效果。基于此，本公开还提供了上文图像处理方法的一种可能的
- 20       实施方式，在该实施方式下，当上文待处理图像是指从图像序列（比如，视频数据）中抽取的一个图像，而且该待处理图像的目标描述信息包括至少一个目标预测结果（比如，多个检测 Query 对应的目标预测结果+多个跟踪 Query 对应的目标预测结果）时，该图像处理方法可以至少包括上文 S1-S3 以及下文步骤 51-步骤 52。其中，该步骤 51 的执行时间晚于该 S3 的执行时间。

- 25       步骤 51：在得到待处理图像的目标描述信息之后，从该目标描述信息中提取满足预设参考条件的目标预测结果。

      需要说明的是，步骤 51 的相关内容类似于上文步骤 41 中部分内容，为了简要起见，在此不再赘述。

- 步骤 52：利用上文图像序列中的下一帧图像更新待处理图像，利用上文
- 30       满足预设参考条件的目标预测结果中的查询表征数据，更新该待处理图像对

应的跟踪查询表征数据，并返回继续执行上文步骤 S1 及其后续步骤，直至该图像序列中所有图像数据均被遍历。

上文步骤 52 中的图像序列是指需要进行目标跟踪处理的序列；而且本公开不限定该图像序列的实施方式，例如，其可以采用视频数据进行实施。

- 5       上文步骤 52 中的下一帧图像是指在上文图像序列中存在的、与上文步骤 51 所涉及的待处理图像位置相邻、而且位置比该步骤 51 所涉及的待处理图像位置靠后的图像数据。

- 10       另外，本公开不限定上文步骤 52 的实施方式，例如，当上文图像序列为图 2 所涉及的视频数据，而且上文 S1-S3 以及步骤 51 均用于针对图 2 所示的图像数据  $I$  进行处理时，该步骤 52 具体可以为：在从该图像数据  $I$  的目标描述信息中提取满足预设参考条件的目标预测结果（比如，由图 2 所示的多目标跟踪模型所输出的数据）之后，可以将图像数据  $I^{+1}$  作为新的待处理图像，并利用该目标预测结果中的查询表征数据（Query）更新该待处理图像对应的跟踪查询表征数据，以便后续能够基于更新后的待处理图像及其对应的跟踪
- 15       查询表征数据，执行针对该图像数据  $I^{+1}$  的目标确定过程。

- 基于上文步骤 51 至步骤 52 的相关内容可知，在一些应用场景下，在获取到当前帧视频图像的目标描述信息之后，可以从该目标描述信息中提取满足预设参考条件的目标预测结果；再利用该目标预测结果中的 Query，确定下一帧视频图像对应的跟踪查询表征数据，以使该跟踪查询表征数据包括该
- 20       Query，如此有利于提高目标跟踪效果。

      需要说明的是，在一些应用场景下，对于一个跟踪 Query 对应的目标预测结果来说，如果该目标预测结果中的类别置信度较低，则可以判定该目标预测结果中的 Query 失效。另外，如果失效持续帧数超过预定的数量时视为此目标消失并不再继续跟踪，相应的跟踪 Query 会被删除。

- 25       还需要说明的是，当上文待处理图像是指从图像序列（比如，视频数据）中抽取的一个图像时，如果该待处理图像为该图像序列中的首帧图像（也就是，第 1 帧图像），则因不存在该待处理图像对应的多个历史图像，以使该待处理图像对应的跟踪查询表征数据也是不存在的，故针对该待处理图像的图像处理过程具体可以为：先对该待处理图像进行特征提取处理，得到特征提取
- 30       结果；再对该特征提取结果进行编码处理，得到检测查询表征数据；然后，依

据该检测查询表征数据进行解码处理,得到该待处理图像的目标描述信息;但是,如果该待处理图像为该图像序列中的非首帧图像(比如,第2帧图像、第3帧图像等),则可以利用上文S1-S3实现针对该待处理图像的图像处理过程。

基于上文图像处理方法的相关内容可知,该图像处理方法可以借助预先  
5 训练好的目标跟踪模型进行实施。也就是,在一种可能的实施方式下,当该目标跟踪模型包括特征提取模块、编码模块以及解码模块时,该图像处理方法具体可以为:先由该特征提取模块针对上文待处理图像(比如,图2所示的图像数据 $I^t$ )进行特征提取处理,得到特征提取结果;再由该编码模块针对该特征提取结果进行编码处理,得到图像特征表征数据(比如,图2所示的图像  
10 特征图)以及检测查询表征数据(比如,图2所示的 $D^t$ );最后,由该解码模块依据该图像特征表征数据、该检测查询表征数据以及该待处理图像对应的跟踪查询表征数据(例如,图2所示的 $X^t$ )进行解码处理,得到该待处理图像的目标描述信息(比如,由图2所示的最后一个解码层所输出的数据)。

实际上,为了更好地提高上文目标跟踪模型的性能,本公开还提供了该目  
15 标跟踪模型的训练过程的一种可能的实施方式,其具体可以包括下文步骤61-步骤62。

步骤61:利用至少一个第二图像数据以及各第二图像数据的目标检测标签,对初始模型进行训练,得到待优化模型。

其中,第二图像数据是指在针对目标跟踪模型的第一训练阶段中所需使  
20 用的图像数据;而且本公开不限定该第二图像数据的获取方式,例如,该第二图像数据可以是指从任意一个目标检测训练数据集中所抽取的图像数据。

第二图像数据的目标检测标签用于描述该第二图像数据中每个目标实际  
所处位置;而且本公开不限定该第二图像数据的目标检测标签的获取方式,比如,可以借助人工标注的方式进行实施。又如,在一些应用场景下,如果该第  
25 二图像数据是指从某个目标检测训练数据集中所抽取的图像数据,则该第二图像数据的目标检测标签可以为该目标检测训练数据集中存在的、与该第二图像数据具有对应关系的标签信息。

初始模型用于代表在第一训练阶段内所需训练的目标跟踪模型;而且本  
公开不限定该初始模型,比如,该初始模型可以包括特征提取模块、编码模块  
30 以及解码模型;而且该特征提取模块是指基于CoCo数据集预训练的ResNet50,

该编码模块是指 DINO 检测算法中的 Transformer Encoder 模块。

待优化模型是指在针对上文初始模型进行目标检测方面训练所得到的模型，以使该待优化模型具有较好的目标检测功能。可见，该待优化模型可以代表经过第一训练阶段所得到的目标跟踪模型。

- 5        步骤 62：利用至少一个图像序列以及各图像序列的目标跟踪标签，对上文待优化模型中的解码器进行训练，得到目标跟踪模型。

上文步骤 62 中的图像序列是指在针对目标跟踪模型的第二训练阶段中所需使用的图像数据序列；而且本公开不限定该图像序列，比如，其可以是指视频数据。另外，本公开不限定该图像序列的获取方式，例如，该图像序列可以是指从任意一个目标跟踪训练数据集中所抽取的视频数据。

10        图像序列的目标跟踪标签用于描述该图像序列中每个图像数据内各个目标实际所处状态；而且本公开不限定该图像序列的目标跟踪标签的获取方式，比如，可以借助人工标注的方式进行实施。又如，在一些应用场景下，如果该图像序列是指从某个目标跟踪训练数据集中所抽取的视频数据，则该图像序列的目标跟踪标签可以为该目标跟踪训练数据集中存在的、与该视频数据具有对应关系的标签信息。

15        基于上文步骤 61 至步骤 62 的相关内容可知，在一些应用场景下，可以采用两阶段训练方式完成针对上文目标跟踪模型的训练过程，而且该两阶段训练方式具体可以为：先将基于图片的目标检测数据视为包括帧长为 1 的视频数据的多目标跟踪训练数据，并利用该多目标跟踪训练数据，针对该目标跟踪模型进行第一训练阶段；再将训练好的模型中的特征提取模块以及编码模块进行固定，并利用包括多帧视频数据的多目标跟踪训练数据，针对该模型中的解码模块进行第二阶段的模型训练，以得到最终训练好的目标跟踪模型，以使该目标跟踪模型具有较好的性能。

25        另外，为了更好地提高上文目标跟踪模型的训练效果，可以在针对上文目标跟踪模型的训练过程中采用一些数据增强手段（比如，Mosaic 和 mixup 等数据增强手段）。基于此可知，该目标跟踪模型的训练过程具体可以为：在获取到至少一个第二图像数据之后，先针对这些第二图像数据进行数据增强处理，得到增强后的图像数据；再利用增强后的图像数据以及这些增强后的图像数据的目标检测标签，对初始模型进行训练，得到待优化模型，以便在获取到

针对上文至少一个图像序列进行数据增强所得到的增强后的图像序列之后，利用该增强后的图像序列以及该增强后的图像序列的目标跟踪标签，对上文待优化模型中的解码器进行训练，得到目标跟踪模型，如此有利于提高该目标跟踪模型的训练效果。

5       此外，为了更好地提高上文目标跟踪模型的训练效果，可以采用多尺度训练方式进行模型训练。基于此可知，对于上文目标跟踪模型来说，在针对该目标跟踪模型的训练过程中可以随机地调整被训练模型中的尺度参数，以确保最终训练好的目标跟踪模型可以适用于处理具有任一种尺度的图像数据，如此有利于提高该目标跟踪模型的普适性。

10       还有，为了更好地提高模型训练效果，本公开还提供了一种上文目标跟踪模型的训练损失的确定过程，为了便于理解，下面结合示例进行说明。

作为示例，当上文目标跟踪模型的训练过程中使用了某个图像序列，而且该图像序列包括第一图像数据以及该第一图像数据对应的多个历史图像时，该目标跟踪模型的训练损失的确定过程，具体可以包括下文步骤 71-步骤 75。

15       步骤 71：依据第一图像数据、该第一图像数据对应的多个历史图像的目标描述信息、以及上文目标跟踪模型，确定该第一图像数据的目标描述信息。

其中，第一图像数据用于代表在一轮训练过程中所使用的图像数据。例如，该第一图像数据可以是图 2 所示的图像数据  $I$ 。

20       第一图像数据对应的多个历史图像是指在利用该第一图像数据进行模型训练时所需参考的图像数据。例如，当该第一图像数据是图 2 所示的图像数据  $I$  时，该第一图像数据对应的多个历史图像可以包括图 2 所示的图像数据  $I^{-1}$ 、图像数据  $I^{-2}$ 、图像数据  $I^{-3}$  等。

25       第一图像数据的目标描述信息用于描述该第一图像数据的一些目标；而且该“第一图像数据的目标描述信息”的确定过程类似于上文“待处理图像的目标描述信息”的确定过程，为了简要起见，在此不再赘述。

30       基于上文步骤 71 的相关内容可知，对于一轮训练过程来说，在获取到第一图像数据以及该第一图像数据对应的多个历史图像的目标描述信息之后，可以基于这些信息以及需要训练的目标跟踪模型，确定出该第一图像数据的目标描述信息，以便后续能够基于该目标描述信息，分析该目标跟踪模型所具有的性能。

步骤 72: 从第一图像数据的目标描述信息中提取第一组信息和第二组信息。

其中, 第一组信息是指上文第一图像数据的目标描述信息中存在的、满足一定条件的目标预测结果; 而且该条件可以预先设定, 例如, 该条件具体可以为: 属于检测 Query 对应的目标预测结果或者属于最近一个历史图像的跟踪 Query 对应的目标预测结果 (比如, 图 2 所示的由长方形虚线框所框定的 Query)。

第二组信息是指上文第一图像数据的目标描述信息中除了第一组信息以外的其他目标预测结果 (比如, 图 2 所示的由正方形虚线框所框定的 Query)。

另外, 本公开不限定上文第一组信息和第二组信息之间的关联关系, 例如, 该第一组信息对应的时序晚于该第二组信息对应的时序。其中, 该第一组信息对应的时序是指在预测该第一组信息时所需依据的图像数据的时序。比如, 如果该第一组信息包括图 2 所示的由长方形虚线框所框定的 Query, 则该第一组信息对应的时序包括图 2 所示的图像数据  $I$  的时序以及图 2 所示的图像数据  $I^{-1}$  的时序。该第二组信息对应的时序是指在预测该第二组信息时所需依据的图像数据的时序。比如, 如果该第二组信息包括图 2 所示的由正方形虚线框所框定的 Query, 则该第二组信息对应的时序包括图 2 所示的图像数据  $I^{-2}$  的时序、图 2 所示的图像数据  $I^{-3}$  的时序等。

基于上文步骤 72 的线管内容可知, 对于一轮训练过程来说, 在获取到第一图像数据的目标描述信息之后, 可以依据该目标描述信息中各个目标预测结果对应的时序, 将该目标描述信息拆分成两组信息进行损失计算。其中, 第一组信息包括所有检测 Query 对应的目标预测结果和最近一个历史图像的跟踪 Query 对应的目标预测结果组成, 而且该第一组信息中不存在多个 Query 跟踪同一个目标的情况。第二组信息是指去掉了最近一个历史图像的跟踪 Query 之后的其他跟踪 Query 的目标预测结果, 而且该第二组信息中存在多个 Query 跟踪同一个目标的情况。

步骤 73: 依据上文第一组信息和上文第一图像数据对应的目标标签数据, 得到第一损失。

其中, 第一图像数据对应的目标标签数据是指在利用该第一图像数据进行模型训练时所需依据的标签信息 (比如, 目标对应的跟踪对象标识标签、该

目标对应的边界框标签、该目标对应的类别标签等),以使该目标标签数据能够表示出该第一图像数据中实际存在哪些目标。

另外,本公开不限定上文步骤 73 的实施方式,例如,其可以采用现有的或者未来出现的任意一种二分匹配损失(bipartite matching loss)进行实施。

- 5       基于上文步骤 73 的相关内容可知,对于一轮训练过程来说,在获取到第一组信息(比如,所有检测 Query 对应的目标预测结果以及最近一个历史图像的跟踪 Query 对应的目标预测结果)之后,因此组中不存在多个 Query 跟踪同一个目标的情况,故可以直接利用二分匹配损失,针对该第一组信息进行损失计算;而且该计算过程大致可以为:对于已跟踪到的目标的跟踪 Query 对应的目标预测结果来说,直接依据该跟踪 Query 对应的跟踪对象标识,将该 Query 对应的目标预测结果与上文第一图像数据对应的目标标签数据中具有该跟踪对象标识的目标进行关联,如此能够实现针对已跟踪到的目标的跟踪 Query 对应的目标预测结果的匹配过程;然后,将该目标标签数据中剩余的未关联任何目标预测结果的目标视为新出现的目标,并基于边界框以及类别等方面的匹配过程,从所有检测 Query 对应的目标预测结果中确定出与该新出现的目标相关联的目标预测结果,如此能够实现针对一些检测 Query 对应的目标预测结果的匹配过程;还有,可以针对那些已经关联到新出现的目标的检测 Query 对应的目标预测结果配置一个新的跟踪对象标识,以便后续能够基于被关联至该目标标签数据中某个目标的目标预测结果进行二分匹配损失计算,得到这些被关联至该目标标签数据中某个目标的目标预测结果对应的损失值。另外,对于未关联到该目标标签数据中任何目标的目标预测结果需要视为背景这一类别,故可以计算这些未关联到该目标标签数据中任何目标的目标预测结果对应的类别预测损失。
- 10
- 15
- 20

- 需要说明的是,对于上段中“被关联至该目标标签数据中某个目标的目标预测结果”来说,可以依据该目标预测结果中的目标位置表征数据与该被关联的目标的边界框标签之间的相似度、以及该目标预测结果中的类别置信度与该被关联的目标的类别标签之间的相似度,确定该目标预测结果对应的损失值。
- 25

- 还需要说明的是,对于上文“未关联到该目标标签数据中任何目标的目标预测结果”来说,可以依据该目标预测结果中的类别置信度与该背景对应的类
- 30

别标签之间的相似度，确定该目标预测结果对应的损失值。

步骤 74: 依据上文第二组信息和上文第一图像数据对应的目标标签数据，得到第二损失。

需要说明的是，本公开不限定上文步骤 74 的实施方式，例如，在一些可能的实施方式下，上文步骤 74 具体可以包括：依据标签匹配结果、上文第二组信息、以及上文第一图像数据对应的目标标签数据，确定第二损失。其中，该标签匹配结果用于描述该第二组信息中每个目标预测结果与该第一图像数据对应的目标标签数据中哪个目标关联。

另外，本公开不限定上文标签匹配结果的确定过程，例如，其可以采用现有的或者未来出现的任意一种将一个已跟踪到的目标的跟踪 Query 对应的目标预测结果与某个标签数据中的目标进行关联的方法（比如，二分匹配损失所涉及的关联方式）进行实施。

实际上，在一些应用场景下，上文第二组信息所描述的已跟踪到的目标与上文第一组信息所描述的已跟踪到的目标大致相同，故为了更好地节省计算资源，上文标签匹配结果可以根据上文第一组信息与该第一图像数据对应的目标标签数据之间的匹配结果进行确定。为了便于理解，下面结合示例进行说明。

作为示例，假设上文第二组信息包括跟踪 Query1 对应的目标预测结果，上文第一组信息包括跟踪 Query2 对应的目标预测结果，而且该跟踪 Query1 对应的跟踪对象标识与该跟踪 Query2 对应的跟踪对象标识相同。基于此该假设可知，该跟踪 Query1 与该跟踪 Query2 用于描述同一个已跟踪到的目标，故可以依据上文第一组信息与该第一图像数据对应的目标标签数据之间的匹配结果，确定该跟踪 Query2 对应的目标预测结果与该第一图像数据对应的目标标签数据中哪个目标相关联；并将确定出的与该跟踪 Query2 对应的目标预测结果相关联的目标，直接确定为与该跟踪 Query1 对应的目标预测结果相关联的目标，如此能够实现基于该第一组信息与该第一图像数据对应的目标标签数据之间的匹配结果，确定上文标签匹配结果的目的。

另外，本公开不限定上文第二损失的计算过程，例如，当上文第二组信息包括至少一个目标预测结果（比如，图 2 所示的由正方形虚线框所框定的 Query）时，该第二损失的计算过程具体可以包括下文步骤 741-步骤 742。

步骤 741: 依据上文标签匹配结果、上文第二组信息、以及上文第一图像数据对应的目标标签数据, 确定各目标预测结果对应的损失。

需要说明的是, 本公开不限定上文步骤 741 的实施方式, 例如, 其具体可以  
以为: 先依据上文标签匹配结果、上文第二组信息、以及上文第一图像数据对  
应的目标标签数据, 确定该第二组信息中各目标预测结果对应的标签信息(比  
如, 边界框标签+类别标签); 再依据该第二组信息中的第  $n$  个目标预测结果  
与该第  $n$  个目标预测结果对应的标签信息, 确定该第  $n$  个目标预测结果对应  
的损失(比如, 依据该第  $n$  个目标预测结果中的目标位置表征数据与该第  $n$  个  
目标预测结果对应的标签信息中的边界框标签之间的相似度、以及依据第  $n$   
10 个目标预测结果中的类别置信度与该第  $n$  个目标预测结果对应的标签信息中  
的类别标签之间的相似度, 确定该第  $n$  个目标预测结果对应的损失),  $n$  为正  
整数,  $n \leq N$ ,  $N$  表示上文第二组信息中目标预测结果的个数。

步骤 742: 依据上文第二组信息中至少一个目标预测结果对应的损失的平  
均值, 确定第二损失。

15 基于上文步骤 74 的相关内容可知, 对于上文第二组信息来说, 因此组中  
存在多个 Query 跟踪同一个目标的情况, 故本公开可以借助跟踪一致性损失  
(tracking object consistency loss), 确定该第二组信息对应的第二损失。其中,  
该跟踪一致性损失保持了与上文二分匹配损失一样的损失计算方式, 但是在  
该跟踪一致性损失的计算过程中可以直接根据上文第一组信息与该第一图像  
20 数据对应的目标标签数据之间的匹配结果, 确定与该第二组信息中各个目标  
预测结果相关联的标签信息。另外, 基于第二组信息计算所得的第二损失是  
在此组内取平均, 而不是与第一组中所有目标预测结果对应的损失合并之后才  
取平均的, 如此能够有效地避免因第二组信息中的目标预测结果个数过多而  
导致的不良影响(比如, 导致基于该第一组中检测 Query 对应的目标预测结  
25 果所确定的损失被稀释等问题), 如此有利于提高模型训练效果。

步骤 75: 依据第一损失和第二损失, 确定目标跟踪模型的训练损失。

需要说明的是, 本公开不限定步骤 75 的实施方式, 例如, 该步骤 75 具  
体可以为: 直接将第一损失和第二损失之间的和值, 确定为目标跟踪模型的训  
练损失。又如, 该步骤 75 具体可以为: 将第一损失和第二损失进行加权求和,  
30 得到该目标跟踪模型的训练损失。还如, 该步骤 75 具体可以为: 将第一损失

和第二损失进行集合，得到该目标跟踪模型的训练损失。

基于上文步骤 71 至步骤 75 的相关内容可知，如图 2 所示，在针对上文目标跟踪模型的训练过程中，该目标跟踪模型的训练损失可以根据二分匹配损失与跟踪一致性损失之间的和值进行确定，如此使得该训练损失能够更好地表示出该模型的性能，从而有利于提高模型训练效果。

需要说明的是，本公开不限定上文步骤 71-步骤 75 所示的训练损失的应用场景，例如，当该步骤 71-步骤 75 所示的训练损失应用于上文步骤 62（也就是，第二训练阶段）时，可以将该步骤 71-步骤 75 中出现的“目标跟踪模型”替换为上文待优化模型即可。

另外，在针对上文目标跟踪模型的训练过程中，当前帧（比如，上文第一图像数据）所涉及的各个跟踪 Query 对应的目标预测结果和关联到新出现的目标的检测 Query 对应的目标预测结果都被分配了相应的跟踪对象标识，故对于任意一个被分配了跟踪对象标识的目标预测结果来说，如果确定被分配了跟踪对象标识的目标预测结果与其关联的目标之间的 IOU 大于一个预定阈值，则可以确定已较好地跟踪到该目标，故可以将与该目标相关联的目标预测结果中的 Query，按照一个预设的概率送入到后续帧中作为历史跟踪 Query（也就是，正样本）参与跟踪。同时与跟踪 Query 相关联的目标位置较近的未分配到跟踪对象标识的检测 Query 对应的目标预测结果中的 Query，会以一定概率作为噪声跟踪 Query（也就是，负样本），关联一个不存在的目标（比如，虚拟目标）并送入后续帧中作为历史跟踪 Query 参与训练，增加模型对噪声的鲁棒性。

基于上文目标跟踪模型的相关内容可知，该目标跟踪模型属于一种基于多 Query 的端到端多目标跟踪算法，其使用每个目标的多个历史 Query 构建当前帧中该目标的协同跟踪 Query，共同跟踪目标以提升特征的可靠性，以使该目标跟踪模型在高帧率视频中和低帧率视频中都获得更好的性能，且处理速度比较快。

基于本公开实施例提供的图像处理方法，本公开实施例还提供了一种图像处理装置，下面结合图 4 进行解释和说明。其中，图 4 为本公开实施例提供的一种图像处理装置的结构示意图。需要说明的是，本公开实施例提供的图像处理装置的技术详情，请参照上文图像处理方法的的相关内容。

如图 4 所示, 本公开实施例提供的图像处理装置 400, 包括:

提取单元 401, 用于对待处理图像进行特征提取处理, 得到特征提取结果;

编码单元 402, 用于对所述特征提取结果进行编码处理, 得到检测查询表征数据;

- 5        解码单元 403, 用于依据所述检测查询表征数据以及所述待处理图像对应的跟踪查询表征数据进行解码处理, 得到所述待处理图像的目标描述信息; 所述跟踪查询表征数据是依据所述待处理图像对应的多个历史图像的目标描述信息所确定的; 所述待处理图像的目标描述信息用于描述所述待处理图像中的至少一个目标。

- 10       在一种可能的实施方式下, 所述编码单元 402, 具体用于: 对所述特征提取结果进行编码处理, 得到图像特征表征数据以及检测查询表征数据;

所述解码单元 403, 具体用于: 依据所述图像特征表征数据、所述检测查询表征数据以及所述待处理图像对应的跟踪查询表征数据进行解码处理, 得到所述待处理图像的目标描述信息。

- 15       在一种可能的实施方式下, 所述多个历史图像包括第一类图像和第二类图像; 所述跟踪查询表征数据的确定过程, 包括: 从所述第一类图像的目标描述信息中提取第一跟踪目标描述信息; 依据所述第一跟踪目标描述信息以及所述第二类图像的目标描述信息, 确定所述跟踪查询表征数据。

- 20       在一种可能的实施方式下, 所述跟踪查询表征数据的确定过程, 包括: 依据所述第一跟踪目标描述信息中的目标位置表征数据以及所述第二类图像的目标描述信息中的目标内容特征表征数据, 确定第二跟踪目标描述信息; 根据所述第二跟踪目标描述信息和所述第一跟踪目标描述信息, 确定所述跟踪查询表征数据。

- 25       在一种可能的实施方式下, 所述待处理图像与所述待处理图像对应的多个历史图像属于同一个视频数据; 所述待处理图像的时序晚于所述第一类图像的时序; 所述第一类图像的时序晚于所述第二类图像的时序。

在一种可能的实施方式下, 所述解码单元 403, 具体用于: 依据所述检测查询表征数据和所述跟踪查询表征数据, 确定待处理查询表征数据; 依据所述待处理查询表征数据进行解码处理, 得到所述待处理图像的目标描述信息。

- 30       在一种可能的实施方式下, 所述解码单元 403, 具体用于: 依据所述检测

查询表征数据和所述跟踪查询表征数据,确定待处理查询表征数据;依据所述图像特征表征数据和所述待处理查询表征数据进行解码处理,得到所述待处理图像的目标描述信息。

5 在一种可能的实施方式下,所述待处理图像的目标描述信息是利用预先训练好的解码器所确定的;所述解码器包括至少一个解码层,所述解码层中的自注意力模块是利用第一注意力掩模进行实现的,所述第一注意力掩模用于屏蔽属于同一个目标的在不同历史图像下对应的查询表征数据之间的交互。

10 在一种可能的实施方式下,所述至少一个目标包括一个或者多个待跟踪目标;所述至少一个解码层包括第一解码层和第二解码层;所述解码器还包括特征融合模块;所述特征融合模块用于针对由所述第一解码层输出的跟踪查询处理数据进行融合处理,得到跟踪查询融合数据;所述跟踪查询处理数据是依据所述第一解码层以及所述跟踪查询表征数据所确定的;所述跟踪查询处理数据包括由所述第一解码层针对各所述待跟踪目标分别输出的处理后的查询表征数据;所述跟踪查询融合数据包括由所述特征融合模块针对各所述处  
15 理后的查询表征数据分别输出的融合后的查询表征数据;所述第二解码层用于针对所述跟踪查询融合数据以及由所述第一解码层输出的检测查询处理数据进行处理;所述检测查询处理数据是依据所述第一解码层以及所述检测查询表征数据所确定的。

20 在一种可能的实施方式下,所述特征融合模块包括信息移除分支网络和信息添加分支网络;

所述跟踪查询融合数据的确定过程,包括:利用所述信息移除分支网络对所述跟踪查询处理数据进行处理,得到保留信息特征;利用所述信息添加分支网络对所述跟踪查询处理数据进行处理,得到待添加信息特征;根据所述保留信息特征和所述待添加信息特征,确定所述跟踪查询融合数据。

25 在一种可能的实施方式下,所述保留信息特征的确定过程,包括:对所述跟踪查询处理数据进行时序信息提取处理,得到第一时序信息;对所述第一时序信息与所述跟踪查询处理数据进行自注意力处理,得到自注意力处理结果;对所述自注意力处理结果进行全连接处理,得到全连接处理结果;依据所述全连接处理结果,确定待移除信息表征数据;依据所述待移除信息表征数据,对  
30 所述跟踪查询处理数据进行信息移除处理,得到所述保留信息特征。

在一种可能的实施方式下，所述待添加信息特征的确定过程，包括：对所述跟踪查询处理数据进行时序信息提取处理，得到第二时序信息；对所述第二时序信息与所述跟踪查询处理数据进行自注意力处理，得到所述待添加信息特征。

- 5       在一种可能的实施方式下，所述特征融合模块中的自注意力层是利用第二注意力掩模进行实现的，所述第二注意力掩模用于屏蔽属于不同目标的查询表征数据之间的交互。

在一种可能的实施方式下，所述待处理图像的目标描述信息是利用预先训练好的目标跟踪模型所确定的；

- 10       所述目标跟踪模型的训练损失的确定过程，包括：依据第一图像数据、所述第一图像数据对应的多个历史图像的目标描述信息、以及所述目标跟踪模型，确定所述第一图像数据的目标描述信息；从所述第一图像数据的目标描述信息中提取第一组信息和第二组信息；依据所述第一组信息和所述第一图像数据对应的目标标签数据，得到第一损失；依据所述第二组信息和所述目标标签数据，得到第二损失；依据所述第一损失和所述第二损失，确定所述目标跟踪模型的训练损失。
- 15

在一种可能的实施方式下，所述第一组信息对应的时序晚于所述第二组信息对应的时序。

- 20       在一种可能的实施方式下，所述第二损失的确定过程，包括：依据标签匹配结果、所述第二组信息、以及所述目标标签数据，确定所述第二损失；所述标签匹配结果是根据所述第一组信息与所述目标标签数据之间的匹配结果所确定的。

在一种可能的实施方式下，所述第二组信息包括至少一个目标预测结果；

- 25       所述第二损失的确定过程，包括：依据所述标签匹配结果、所述第二组信息、以及所述目标标签数据，确定各所述目标预测结果对应的损失；依据所述至少一个目标预测结果对应的损失的平均值，确定所述第二损失。

在一种可能的实施方式下，所述待处理图像的目标描述信息是利用预先训练好的目标跟踪模型所确定的；

- 30       所述目标跟踪模型的训练过程，包括：利用至少一个第二图像数据以及各所述第二图像数据的目标检测标签，对初始模型进行训练，得到待优化模型；

利用至少一个图像序列以及各所述图像序列的目标跟踪标签，对所述待优化模型中的解码模块进行训练，得到所述目标跟踪模型。

在一种可能的实施方式下，所述至少一个目标包括一个或者多个待跟踪目标；所述跟踪查询表征数据包括各所述待跟踪目标在多个历史图像下对应的查询表征数据。

在一种可能的实施方式下，所述待处理图像是指从图像序列中抽取的一个图像；所述目标描述信息包括至少一个目标预测结果；

所述图像处理装置 400 还包括：

筛选单元，用于从所述目标描述信息中提取满足预设参考条件的目标预测结果；

更新单元，用于利用所述图像序列中的下一帧图像更新所述待处理图像，利用所述满足预设参考条件的目标预测结果中的查询表征数据，更新所述跟踪查询表征数据，并继续执行所述对待处理图像进行特征提取处理的步骤。

在一种可能的实施方式下，所述图像序列为视频数据。

基于上述图像处理装置 400 的相关内容可知，对于本公开实施例提供的图像处理装置 400 来说，先对待处理图像进行特征提取处理，得到特征提取结果，以使该特征提取结果用于表征该待处理图像所携带的图像信息；再对该特征提取结果进行编码处理，得到检测查询表征数据（比如，多个检测 Query）；最后，依据该检测查询表征数据以及该待处理图像对应的跟踪查询表征数据（比如，基于多个历史图像所确定的跟踪 Query）进行解码处理，得到该待处理图像的目标描述信息（比如，目标内容特征表征数据、目标位置表征数据、以及类别置信度等），以使该目标描述信息能够表示出该待处理图像中一些目标。其中，因该跟踪查询表征数据是依据该待处理图像对应的多个历史图像的目标描述信息所确定的，以使该跟踪查询表征数据能够更好地表示出该待处理图像中一些目标的历史状态，从而使得基于该跟踪查询表征数据所确定的目标描述信息能够更好地表示出该待处理图像中一些目标，如此有利于提高目标跟踪效果。

另外，本公开实施例还提供了一种电子设备，所述设备包括处理器以及存储器：所述存储器，用于存储指令或计算机程序；所述处理器，用于执行所述存储器中的所述指令或计算机程序，以使得所述电子设备执行本公开实施例

提供的图像处理方法的任一实施方式。

参见图 5，其示出了适于用来实现本公开实施例的电子设备 500 的结构示意图。本公开实施例中的终端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、PDA（个人数字助理）、PAD（平板电脑）、PMP（便携式多媒体播放器）、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 TV、台式计算机等等的固定终端。图 5 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 5 所示，电子设备 500 可以包括处理装置（例如中央处理器、图形处理器等）501，其可以根据存储在只读存储器（ROM）502 中的程序或者从存储装置 508 加载到随机访问存储器（RAM）503 中的程序而执行各种适当的动作和处理。在 RAM503 中，还存储有电子设备 500 操作所需的各种程序和数据。处理装置 501、ROM 502 以及 RAM 503 通过总线 504 彼此相连。输入/输出（I/O）接口 505 也连接至总线 504。

通常，以下装置可以连接至 I/O 接口 505：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 506；包括例如液晶显示器（LCD）、扬声器、振动器等的输出装置 507；包括例如磁带、硬盘等的存储装置 508；以及通信装置 509。通信装置 509 可以允许电子设备 500 与其他设备进行无线或有线通信以交换数据。虽然图 5 示出了具有各种装置的电子设备 500，但是应理解的是，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。

特别地，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 509 从网络上被下载和安装，或者从存储装置 508 被安装，或者从 ROM502 被安装。在该计算机程序被处理装置 501 执行时，执行本公开实施例的方法中限定的上述功能。

本公开实施例提供的电子设备与上述实施例提供的方法属于同一发明构思，未在本实施例中详尽描述的技术细节可参见上述实施例，并且本实施例与上述实施例具有相同的有益效果。

本公开实施例还提供了一种计算机可读介质，所述计算机可读介质中存储有指令或计算机程序，当所述指令或计算机程序在设备上运行时，使得所述设备执行本公开实施例提供的图像处理方法的任一实施方式。

需要说明的是，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更具体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、随机访问存储器（RAM）、只读存储器（ROM）、可擦式可编程只读存储器（EPROM 或闪存）、光纤、便携式紧凑磁盘只读存储器（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、RF（射频）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 HTTP（Hyper Text Transfer Protocol，超文本传输协议）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（“LAN”），广域网（“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，使得该电子设备可以执行上述方法。

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——包括局域网(LAN)或广域网(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，图示了按照本公开各种实施例的系统、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，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单元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，单元/模块的名称在某种情况下并不构成对该单元本身的限定。

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(FPGA)、专用集成电路(ASIC)、专用标准产品(ASSP)、片上系统(SOC)、复杂可编程逻辑设备(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。

需要说明的是，本公开中各个实施例采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似部分互相参见即可。对于实施例公开的系统或装置而言，由于其与实施例公开的方法相对应，所以描述的比较简单，相关之处参见方法部分说明即可。

应当理解，在本公开中，“至少一个（项）”是指一个或者多个，“多个”是指两个或两个以上。“和/或”，用于描述关联对象的关联关系，表示可以存在三种关系，例如，“A 和/或 B”可以表示：只存在 A，只存在 B 以及同时存在 A 和 B 三种情况，其中 A，B 可以是单数或者复数。字符“/”一般表示前后关联对象是一种“或”的关系。“以下至少一项（个）”或其类似表达，是指这些项中的任意组合，包括单项（个）或复数项（个）的任意组合。例如，a，b 或 c 中的至少一项（个），可以表示：a，b，c，“a 和 b”，“a 和 c”，“b 和 c”，或“a 和 b 和 c”，其中 a，b，c 可以是单个，也可以是多个。

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

结合本文中所公开的实施例描述的方法或算法的步骤可以直接用硬件、处理器执行的软件模块，或者二者的结合来实施。软件模块可以置于随机存储器（RAM）、内存、只读存储器（ROM）、电可编程 ROM、电可擦除可编程 ROM、寄存器、硬盘、可需要说明的是，本公开不限定上文步骤 33 的实施方式，移动磁盘、CD-ROM、或技术领域内所公知的任意其它形式的存储介质中。

- 对所公开的实施例的上述说明，使本领域专业技术人员能够实现或使用本公开。对这些实施例的多种修改对本领域的专业技术人员来说将是显而易见的，本文中所定义的一般原理可以在不脱离本公开的精神或范围的情况下，在其它实施例中实现。因此，本公开将不会被限制于本文所示的这些实施例，
- 5 而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。

## 权利要求书

1. 一种图像处理方法，包括：

对待处理图像进行特征提取处理，得到特征提取结果；

5 对所述特征提取结果进行编码处理，得到检测查询表征数据；

依据所述检测查询表征数据以及所述待处理图像对应的跟踪查询表征数据进行解码处理，得到所述待处理图像的目标描述信息；所述跟踪查询表征数据是依据所述待处理图像对应的多个历史图像的目标描述信息所确定的；所述待处理图像的目标描述信息用于描述所述待处理图像中的至少一个目标。

10 2. 根据权利要求 1 所述的方法，其中，所述多个历史图像包括第一类图像和第二类图像；

所述跟踪查询表征数据的确定过程，包括：

从所述第一类图像的目标描述信息中提取第一跟踪目标描述信息；

依据所述第一跟踪目标描述信息以及所述第二类图像的目标描述信息，

15 确定所述跟踪查询表征数据。

3. 根据权利要求 2 所述的方法，其中，所述依据所述第一跟踪目标描述信息以及所述第二类图像的目标描述信息，确定所述跟踪查询表征数据，包括：

20 依据所述第一跟踪目标描述信息中的目标位置表征数据以及所述第二类图像的目标描述信息中的目标内容特征表征数据，确定第二跟踪目标描述信息；

根据所述第二跟踪目标描述信息和所述第一跟踪目标描述信息，确定所述跟踪查询表征数据。

25 4. 根据权利要求 2 所述的方法，其中，所述待处理图像与所述待处理图像对应的多个历史图像属于同一个视频数据；

所述待处理图像的时序晚于所述第一类图像的时序；

所述第一类图像的时序晚于所述第二类图像的时序。

5. 根据权利要求 1 所述的方法，其中，所述依据所述检测查询表征数据以及所述待处理图像对应的跟踪查询表征数据进行解码处理，得到所述待处理图像的目标描述信息，包括：

30

依据所述检测查询表征数据和所述跟踪查询表征数据，确定待处理查询表征数据；

依据所述待处理查询表征数据进行解码处理，得到所述待处理图像的目标描述信息。

- 5        6. 根据权利要求 1 所述的方法，其中，所述待处理图像的目标描述信息是利用预先训练好的解码器所确定的；

所述解码器包括至少一个解码层，所述解码层中的自注意力模块是利用第一注意力掩模进行实现的，所述第一注意力掩模用于屏蔽属于同一个目标的在不同历史图像下对应的查询表征数据之间的交互。

- 10       7. 根据权利要求 6 所述的方法，其中，所述至少一个目标包括一个或者多个待跟踪目标；

所述至少一个解码层包括第一解码层和第二解码层；

所述解码器还包括特征融合模块；

- 15       所述特征融合模块用于针对由所述第一解码层输出的跟踪查询处理数据进行融合处理，得到跟踪查询融合数据；所述跟踪查询处理数据是依据所述第一解码层以及所述跟踪查询表征数据所确定的；所述跟踪查询处理数据包括由所述第一解码层针对各所述待跟踪目标分别输出的处理后的查询表征数据；所述跟踪查询融合数据包括由所述特征融合模块针对各所述处理后的查询表征数据分别输出的融合后的查询表征数据；

- 20       所述第二解码层用于针对所述跟踪查询融合数据以及由所述第一解码层输出的检测查询处理数据进行处理；所述检测查询处理数据是依据所述第一解码层以及所述检测查询表征数据所确定的。

8. 根据权利要求 7 所述的方法，其中，所述特征融合模块包括信息移除分支网络和信息添加分支网络；

- 25       所述跟踪查询融合数据的确定过程，包括：

利用所述信息移除分支网络对所述跟踪查询处理数据进行处理，得到保留信息特征；

利用所述信息添加分支网络对所述跟踪查询处理数据进行处理，得到待添加信息特征；

- 30       根据所述保留信息特征和所述待添加信息特征，确定所述跟踪查询融合

数据。

9. 根据权利要求 8 所述的方法，其中，所述保留信息特征的确定过程，包括：

对所述跟踪查询处理数据进行时序信息提取处理，得到第一时序信息；

5 对所述第一时序信息与所述跟踪查询处理数据进行自注意力处理，得到自注意力处理结果；

对所述自注意力处理结果进行全连接处理，得到全连接处理结果；

依据所述全连接处理结果，确定待移除信息表征数据；

10 依据所述待移除信息表征数据，对所述跟踪查询处理数据进行信息移除处理，得到所述保留信息特征。

10. 根据权利要求 8 所述的方法，其中，所述待添加信息特征的确定过程，包括：

对所述跟踪查询处理数据进行时序信息提取处理，得到第二时序信息；

15 对所述第二时序信息与所述跟踪查询处理数据进行自注意力处理，得到所述待添加信息特征。

11. 根据权利要求 7 所述的方法，其中，所述特征融合模块中的自注意力层是利用第二注意力掩模进行实现的，所述第二注意力掩模用于屏蔽属于不同目标的查询表征数据之间的交互。

20 12. 根据权利要求 1 所述的方法，其中，所述待处理图像的目标描述信息是利用预先训练好的目标跟踪模型所确定的；

所述目标跟踪模型的训练损失的确定过程，包括：

依据第一图像数据、所述第一图像数据对应的多个历史图像的目标描述信息、以及所述目标跟踪模型，确定所述第一图像数据的目标描述信息；

从所述第一图像数据的目标描述信息中提取第一组信息和第二组信息；

25 依据所述第一组信息和所述第一图像数据对应的目标标签数据，得到第一损失；

依据所述第二组信息和所述目标标签数据，得到第二损失；

依据所述第一损失和所述第二损失，确定所述目标跟踪模型的训练损失。

30 13. 根据权利要求 12 所述的方法，其中，所述第一组信息对应的时序晚于所述第二组信息对应的时序。

14. 根据权利要求 12 所述的方法，其中，所述依据所述第二组信息和所述目标标签数据，得到第二损失，包括：

依据标签匹配结果、所述第二组信息、以及所述目标标签数据，确定所述第二损失；所述标签匹配结果是根据所述第一组信息与所述目标标签数据之间的匹配结果所确定的。

15. 根据权利要求 14 所述的方法，其中，所述第二组信息包括至少一个目标预测结果；

所述第二损失的确定过程，包括：

依据所述标签匹配结果、所述第二组信息、以及所述目标标签数据，确定各所述目标预测结果对应的损失；

依据所述至少一个目标预测结果对应的损失的平均值，确定所述第二损失。

16. 根据权利要求 1 所述的方法，其中，所述待处理图像的目标描述信息是利用预先训练好的目标跟踪模型所确定的；

15 所述目标跟踪模型的训练过程，包括：

利用至少一个第二图像数据以及各所述第二图像数据的目标检测标签，对初始模型进行训练，得到待优化模型；

利用至少一个图像序列以及各所述图像序列的目标跟踪标签，对所述待优化模型中的解码模块进行训练，得到所述目标跟踪模型。

20 17. 根据权利要求 1 所述的方法，其中，所述至少一个目标包括一个或者多个待跟踪目标；所述跟踪查询表征数据包括各所述待跟踪目标在多个历史图像下对应的查询表征数据。

25 18. 根据权利要求 1-17 任一项所述的方法，其中，所述待处理图像是指从图像序列中抽取的一个图像；所述目标描述信息包括至少一个目标预测结果；

所述得到所述待处理图像的目标描述信息之后，所述方法还包括：

从所述目标描述信息中提取满足预设参考条件的目标预测结果；

30 利用所述图像序列中的下一帧图像更新所述待处理图像，利用所述满足预设参考条件的目标预测结果中的查询表征数据，更新所述跟踪查询表征数据，并继续执行所述对待处理图像进行特征提取处理的步骤。

19. 根据权利要求 18 所述的方法，其中，所述图像序列为视频数据。

20. 一种图像处理装置，包括：

提取单元，被配置为对待处理图像进行特征提取处理，得到特征提取结果；

5       编码单元，被配置为对所述特征提取结果进行编码处理，得到检测查询表征数据；

      解码单元，被配置为依据所述检测查询表征数据以及所述待处理图像对应的跟踪查询表征数据进行解码处理，得到所述待处理图像的目标描述信息；所述跟踪查询表征数据是依据所述待处理图像对应的多个历史图像的目标描述信息所确定的；所述待处理图像的目标描述信息用于描述所述待处理图像中的至少一个目标。

21. 一种电子设备，包括：处理器和存储器；其中，

      所述存储器，被配置为存储指令或计算机程序；

15       所述处理器，被配置为执行所述存储器中的所述指令或计算机程序，以使  
      得所述电子设备执行权利要求 1-19 任一项所述的方法。

22. 一种计算机可读介质，其中，所述计算机可读介质中存储有指令或计算机程序，当所述指令或计算机程序在设备上运行时，使得所述设备执行权利要求 1-19 任一项所述的方法。

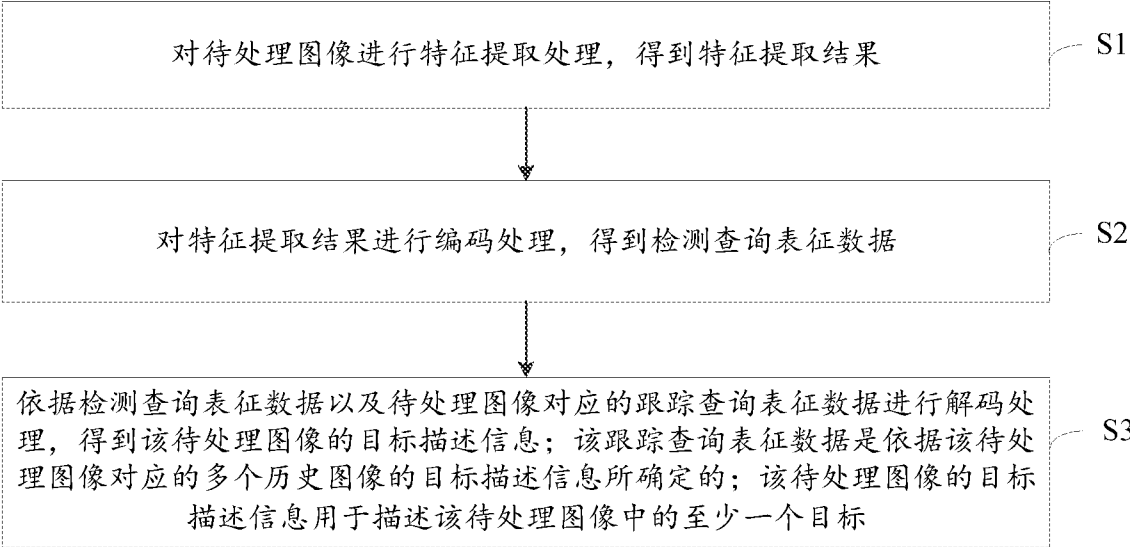


图 1

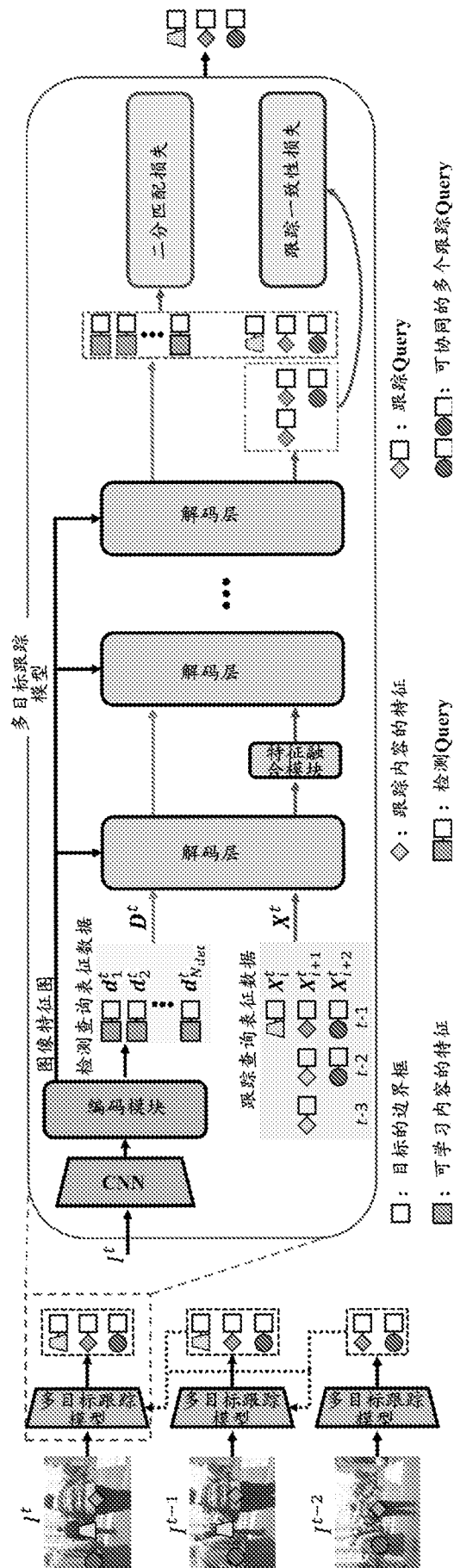


图 2

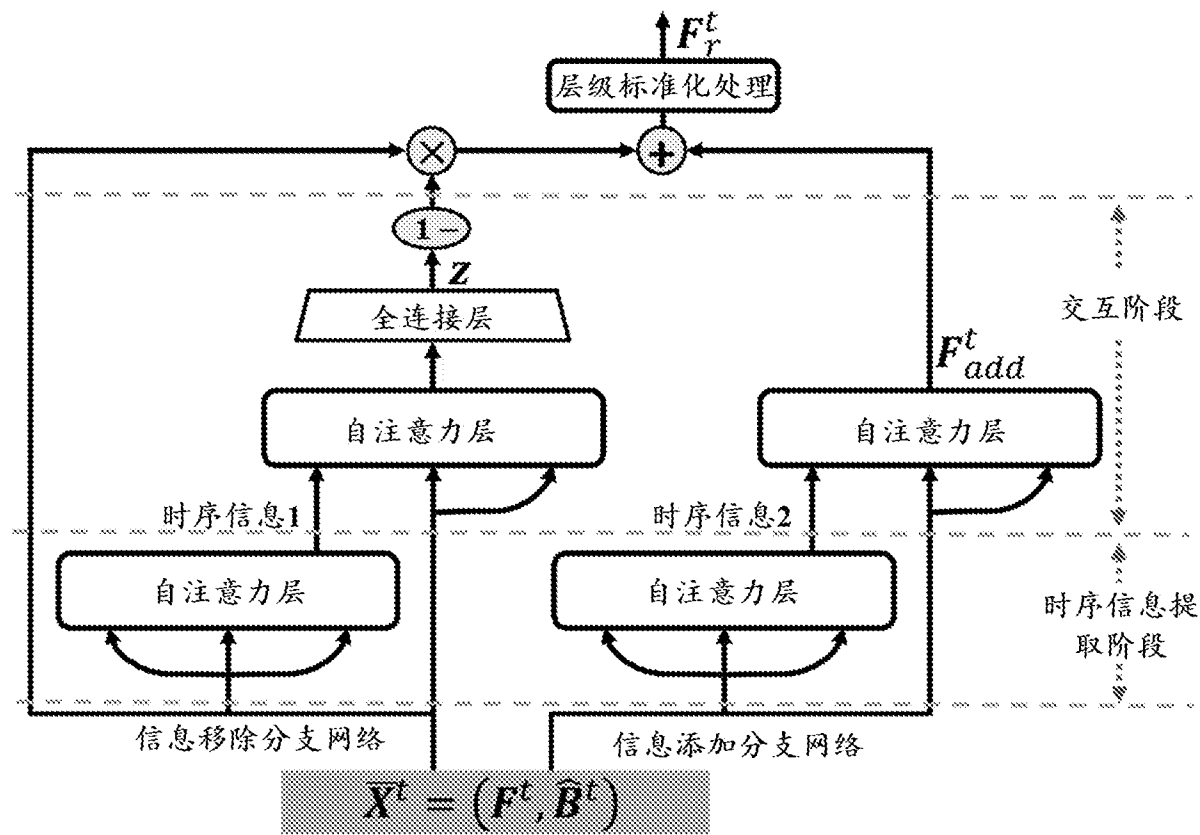


图 3

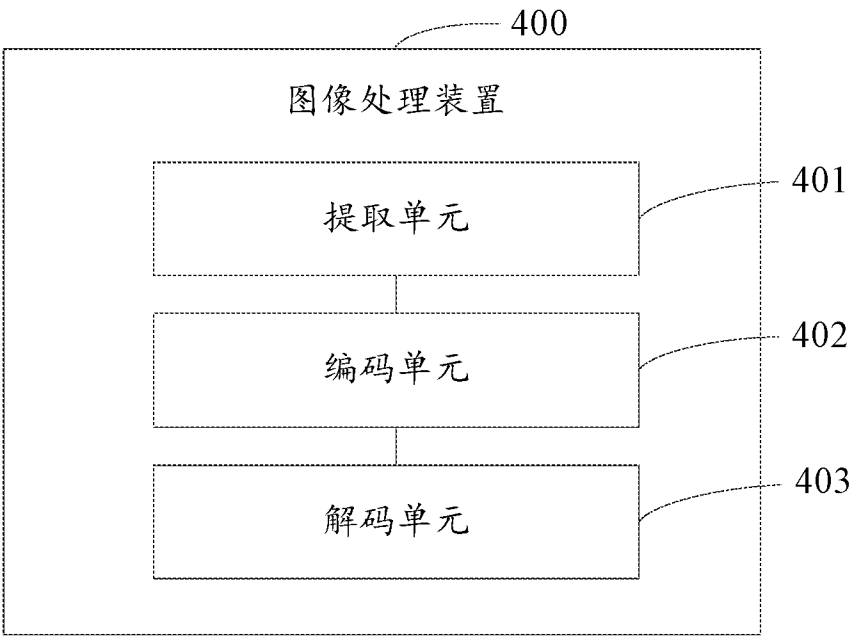


图 4

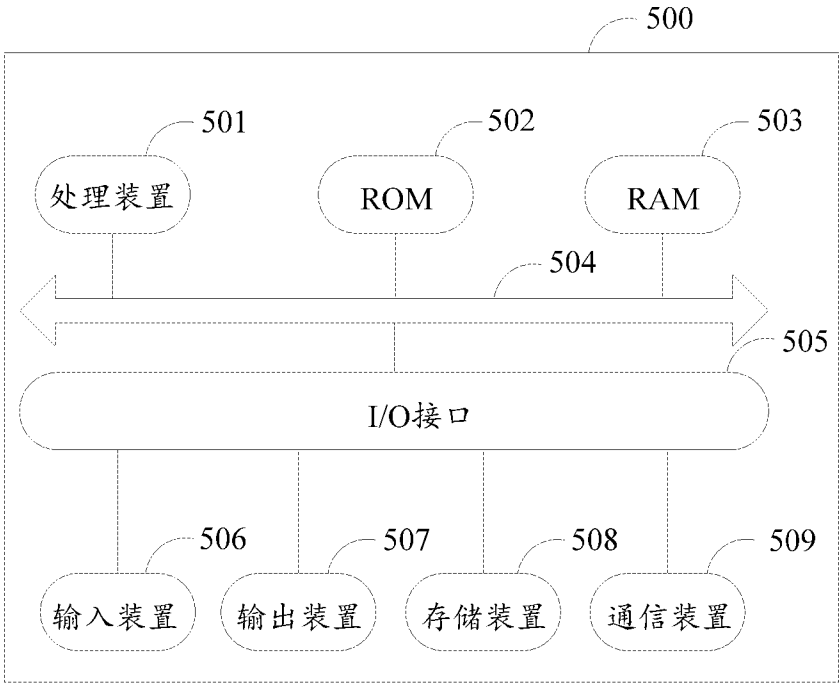


图 5

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2024/099597

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                                                                                                                                                                                                |                                                                             |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|
| <b>A. CLASSIFICATION OF SUBJECT MATTER</b><br>G06V10/40(2022.01)i<br><br>According to International Patent Classification (IPC) or to both national classification and IPC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                |                                                                             |
| <b>B. FIELDS SEARCHED</b><br><br>Minimum documentation searched (classification system followed by classification symbols)<br>IPC: G06V<br><br>Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched<br><br>Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)<br>VEN, CNABS, CNTXT, WOTXT, EPTXT, USTXT, CNKI, IEEE: 图像, 特征, 提取, 跟踪, 查询, 历史, 帧, 目标, 描述, 编码, 解码, 视频, image, feature, extraction, track, query, history, frame, target, description, encode, decode, video                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                                                                                                                                                                                                                |                                                                             |
| <b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                                                                                                                                                                                                                |                                                                             |
| Category*                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                             | Relevant to claim No.                                                       |
| X                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | ZENG, Fangao et al. "MOTR: End-to-End Multiple-Object Tracking with Transformer" <i>ECCV 2022</i> , 30 October 2022 (2022-10-30), page 660, paragraphs 2 to 3, page 664, 3rd-to-last paragraph to last paragraph, and figure 3 | 1, 5, 16-22                                                                 |
| A                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | CN 112347852 A (SHANGHAI JIAO TONG UNIVERSITY) 09 February 2021 (2021-02-09) entire document                                                                                                                                   | 1-22                                                                        |
| A                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | CN 115861893 A (HARBIN INSTITUTE OF TECHNOLOGY) 28 March 2023 (2023-03-28) entire document                                                                                                                                     | 1-22                                                                        |
| A                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | CN 113963032 A (ZHEJIANG UNIVERSITY OF TECHNOLOGY) 21 January 2022 (2022-01-21) entire document                                                                                                                                | 1-22                                                                        |
| A                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | US 2022301183 A1 (BEIJING Baidu NETCOM SCIENCE & TECHNOLOGY CO., LTD.) 22 September 2022 (2022-09-22) entire document                                                                                                          | 1-22                                                                        |
| <input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                                                                                                                                                                                |                                                                             |
| <p>* Special categories of cited documents:</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"D" document cited by the applicant in the international application</p> <p>"E" earlier application or patent but published on or after the international filing date</p> <p>"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)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"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</p> <p>"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</p> <p>"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</p> <p>"&amp;" document member of the same patent family</p> |                                                                                                                                                                                                                                |                                                                             |
| Date of the actual completion of the international search<br><b>27 August 2024</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                                                                                                                                                                                                                | Date of mailing of the international search report<br><b>30 August 2024</b> |
| Name and mailing address of the ISA/CN<br><b>China National Intellectual Property Administration (ISA/CN)<br/>China No. 6, Xitucheng Road, Jimenqiao, Haidian District,<br/>Beijing 100088</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                                                                                | Authorized officer<br><br><br>Telephone No.                                 |

INTERNATIONAL SEARCH REPORT  
Information on patent family members

International application No.  
**PCT/CN2024/099597**

| Patent document cited in search report |            |    | Publication date (day/month/year) | Patent family member(s) | Publication date (day/month/year) |
|----------------------------------------|------------|----|-----------------------------------|-------------------------|-----------------------------------|
| CN                                     | 112347852  | A  | 09 February 2021                  | None                    |                                   |
| CN                                     | 115861893  | A  | 28 March 2023                     | None                    |                                   |
| CN                                     | 113963032  | A  | 21 January 2022                   | None                    |                                   |
| US                                     | 2022301183 | A1 | 22 September 2022                 | None                    |                                   |

A. 主题的分类

G06V10/40(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)

IPC: G06V

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))

VEN, CNABS, CNTXT, WOTXT, EPTXT, USTXT, CNKI, IEEE: 图像, 特征, 提取, 跟踪, 查询, 历史, 帧, 目标, 描述, 编码, 解码, 视频, image, feature, extraction, track, query, history, frame, target, description, encode, decode, video

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   |
|-----|---------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| X   | ZENG, Fangao 等. "MOTR: End-to-End Multiple-Object Tracking with Transformer" ECCV 2022, 2022年10月30日 (2022 - 10 - 30), 第660页第2-3段、第664页倒数第1-3段, 图3 | 1,5,16-22 |
| A   | CN 112347852 A (上海交通大学) 2021年2月9日 (2021 - 02 - 09) 全文                                                                                             | 1-22      |
| A   | CN 115861893 A (哈尔滨工业大学) 2023年3月28日 (2023 - 03 - 28) 全文                                                                                           | 1-22      |
| A   | CN 113963032 A (浙江工业大学) 2022年1月21日 (2022 - 01 - 21) 全文                                                                                            | 1-22      |
| A   | US 2022301183 A1 (BEIJING BAIDU NETCOM SCIENCE & TECHNOLOGY CO., LTD.) 2022年9月22日 (2022 - 09 - 22) 全文                                             | 1-22      |

☐ 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

“D” 申请人在国际申请中引证的文件

“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

2024年8月27日

国际检索报告邮寄日期

2024年8月30日

ISA/CN的名称和邮寄地址

中国国家知识产权局  
中国北京市海淀区蓟门桥西土城路6号 100088

授权官员

马晋涛

电话号码 (+86) 010-53961410

PCT/ISA/210 表(第2页) (2022年7月)

国际检索报告  
关于同族专利的信息

国际申请号  
PCT/CN2024/099597

| 检索报告引用的专利文件 |            |    | 公布日<br>(年/月/日) | 同族专利 | 公布日<br>(年/月/日) |
|-------------|------------|----|----------------|------|----------------|
| CN          | 112347852  | A  | 2021年2月9日      | 无    |                |
| CN          | 115861893  | A  | 2023年3月28日     | 无    |                |
| CN          | 113963032  | A  | 2022年1月21日     | 无    |                |
| US          | 2022301183 | A1 | 2022年9月22日     | 无    |                |