

(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
2020 年 12 月 3 日 (03.12.2020)



(10) 国际公布号
WO 2020/238902 A1

(51) 国际专利分类号:
G06K 9/32 (2006.01) **G06T 7/00** (2017.01)
G06K 9/34 (2006.01) **G06T 7/11** (2017.01)
G06K 9/62 (2006.01)

(21) 国际申请号: PCT/CN2020/092356

(22) 国际申请日: 2020 年 5 月 26 日 (26.05.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910455150.4 2019年5月29日 (29.05.2019) CN

(71) 申请人: 腾讯科技(深圳)有限公司 (TENCENT TECHNOLOGY (SHENZHEN) COMPANY LIMITED) [CN/CN]; 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。

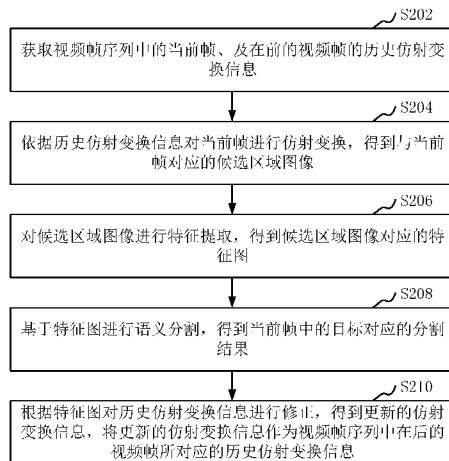
(72) 发明人: 陈思宏 (CHEN, Sihong); 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。

(74) 代理人: 北京三高永信知识产权代理有限公司 (BEIJING SAN GAO YONG XIN INTELLECTUAL PROPERTY AGENCY CO., LTD.); 中国北京市海淀区学院路蓟门里和景园 A 座 1 单元 102 室, Beijing 100088 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,

(54) **Title:** IMAGE SEGMENTATION METHOD, MODEL TRAINING METHOD, APPARATUSES, DEVICE AND STORAGE MEDIUM

(54) 发明名称: 图像分割方法、模型训练方法、装置、设备及存储介质



S202 ACQUIRE THE CURRENT FRAME IN A VIDEO FRAME SEQUENCE AND HISTORICAL AFFINE TRANSFORMATION INFORMATION OF A PRECEDING VIDEO FRAME
S204 PERFORM AFFINE TRANSFORMATION ON THE CURRENT FRAME ACCORDING TO THE HISTORICAL AFFINE TRANSFORMATION INFORMATION TO OBTAIN A CANDIDATE REGION IMAGE CORRESPONDING TO THE CURRENT FRAME
S206 PERFORM FEATURE EXTRACTION ON THE CANDIDATE REGION IMAGE TO OBTAIN A FEATURE MAP CORRESPONDING TO THE CANDIDATE REGION IMAGE
S208 PERFORM SEMANTIC SEGMENTATION ON THE BASIS OF THE FEATURE MAP TO OBTAIN A SEGMENTATION RESULT CORRESPONDING TO A TARGET IN THE CURRENT FRAME
S210 CORRECT THE HISTORICAL AFFINE TRANSFORMATION INFORMATION ACCORDING TO THE FEATURE MAP, TO OBTAIN UPDATED AFFINE TRANSFORMATION INFORMATION, AND TAKE THE UPDATED AFFINE TRANSFORMATION INFORMATION AS HISTORICAL AFFINE TRANSFORMATION INFORMATION CORRESPONDING TO THE SUBSEQUENT VIDEO FRAME IN THE VIDEO FRAME SEQUENCE

图 2

(57) **Abstract:** The present application relates to an image segmentation method, a model training method, apparatuses, a device and a storage medium. The image segmentation method comprises: acquiring the current frame in a video frame sequence and historical affine transformation information transferred by a preceding video frame; performing affine transformation on the current frame according to the historical affine transformation information to obtain a candidate region image corresponding to the current frame; performing feature extraction on the candidate region image to obtain a feature map corresponding to the candidate region image; performing semantic segmentation on the basis of the feature map to obtain a segmentation result corresponding to a target in the current frame; correcting the historical affine transformation information according to the feature map, to obtain updated affine transformation information, and taking the updated affine transformation information as historical affine transformation information corresponding to the subsequent video frame in the video frame sequence. The solutions provided in the present application improve the accuracy of image segmentation.

ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US,
UZ, VC, VN, WS, ZA, ZM, ZW。

- (84) 指定国(除另有指明, 要求每一种可提供的地区
保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG,
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,
IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告(条约第21条(3))。

(57) 摘要: 本申请涉及一种图像分割方法、模型训练方法、装置、设备及存储介质, 其中图像分割方法包括:
获取视频帧序列中的当前帧、及在前的视频帧所传递的历史仿射变换信息; 依据所述历史仿射变换信息对所述
当前帧进行仿射变换, 得到与所述当前帧对应的候选区域图像; 对所述候选区域图像进行特征提取, 得到所述
候选区域图像对应的特征图; 基于所述特征图进行语义分割, 得到所述当前帧中的目标对应的分割结果; 根据
所述特征图对所述历史仿射变换信息进行修正, 得到更新的仿射变换信息, 将所述更新的仿射变换信息作为所
述视频帧序列中在后的视频帧所对应的历史仿射变换信息。本申请提供的方案可以提高图像分割准确性。

图像分割方法、模型训练方法、装置、设备及存储介质

本申请要求于 2019 年 05 月 29 日提交的申请号为 201910455150.4、发明名称为“图像分割方法和装置、模型训练方法和装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及计算机技术领域，特别是涉及一种图像分割方法和装置、模型训练方法和装置。

背景技术

对图像或视频进行语义分割(semantic segmentation)是计算机视觉研究领域的热点之一，语义分割技术是指计算机设备将图片中属于一大类的区域都分割出来并给出其类别信息。

相关技术中对视频进行语义分割的方式，计算机设备需要对视频的每一帧进行关键点预测，得到每一帧的关键点。计算机设备通过模板，根据每一帧的关键点计算每一帧图像与模板的差异来获取变换参数，基于该变换参数进行仿射(Affine)变换得到 ROI(region of interest, 感兴趣区域)，随后在 ROI 上进行目标分割。

然而在上述语义分割方式中，在后的视频帧的关键点的预测依赖于在前的视频帧的目标分割结果，首帧的预测偏差会直接导致后续一系列视频帧的定位偏移，导致对目标对象的语义分割准确性低。

发明内容

本申请提供一种图像分割方法、模型训练方法、装置、设备及存储介质，能够提高语义分割的准确性。

根据本申请的一个方面，提供了一种图像分割方法，应用于计算机设备中，所述方法包括：

获取视频帧序列中的当前帧、及在前的视频帧所传递的历史仿射变换信息；

依据所述历史仿射变换信息对所述当前帧进行仿射变换，得到与所述当前帧对应的候选区域图像；

对所述候选区域图像进行特征提取，得到所述候选区域图像对应的特征图；

基于所述特征图进行语义分割，得到所述当前帧中的目标对应的分割结果；

根据所述特征图对所述历史仿射变换信息进行修正，得到更新的仿射变换信息，并将所述更新的仿射变换信息作为所述视频帧序列中在后的视频帧所对应的历史仿射变换信息。

根据本申请的一个方面，提供了一种图像分割装置，所述装置包括：

获取模块，用于获取视频帧序列中的当前帧、及在前的视频帧所传递的历史仿射变换信息；

仿射变换模块，用于依据所述历史仿射变换信息对所述当前帧进行仿射变换，得到与所述当前帧对应的候选区域图像；

特征提取模块，用于对所述候选区域图像进行特征提取，得到所述候选区域图像对应的特征图；

语义分割模块，用于基于所述特征图进行语义分割，得到所述当前帧中的目标对应的分

割结果；

参数修正模块，用于根据所述特征图对所述历史仿射变换信息进行修正，得到更新的仿射变换信息，并将所述更新的仿射变换信息作为所述视频帧序列中在后的视频帧所对应的历史仿射变换信息。

根据本申请的一个方面，提供了一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序被处理器执行时，使得所述处理器执行以下步骤：

获取视频帧序列中的当前帧、及在前的视频帧的历史仿射变换信息；

依据所述历史仿射变换信息对所述当前帧进行仿射变换，得到与所述当前帧对应的候选区域图像；

对所述候选区域图像进行特征提取，得到所述候选区域图像对应的特征图；

基于所述特征图进行语义分割，得到所述当前帧中的目标对应的分割结果；

根据所述特征图对所述历史仿射变换信息进行修正，得到更新的仿射变换信息，将所述更新的仿射变换信息作为所述视频帧序列中在后的视频帧所对应的历史仿射变换信息。

根据本申请的一个方面，提供了一种计算机设备，所述计算机设备包括存储器和处理器，所述存储器存储有计算机程序，所述计算机程序被所述处理器执行时，使得所述处理器执行以下步骤：

获取视频帧序列中的当前帧、及在前的视频帧的历史仿射变换信息；

依据所述历史仿射变换信息对所述当前帧进行仿射变换，得到与所述当前帧对应的候选区域图像；

对所述候选区域图像进行特征提取，得到所述候选区域图像对应的特征图；

基于所述特征图进行语义分割，得到所述当前帧中的目标对应的分割结果；

根据所述特征图对所述历史仿射变换信息进行修正，得到更新的仿射变换信息，将所述更新的仿射变换信息作为所述视频帧序列中在后的视频帧所对应的历史仿射变换信息。

上述图像分割方法、装置、计算机可读存储介质和计算机设备，依据在前的视频帧的历史仿射变换信息，对当前帧进行仿射变换，得到与当前帧对应的候选区域图像。在前的视频帧的历史仿射变换信息是经过修正后的参数，这样可大大提高候选区域图像获取的准确性。对与候选区域图像对应的特征图进行语义分割，可以准确得到当前帧中的目标对应的分割结果。并且，根据该特征图对历史仿射变换信息进行修正，将修正后的仿射变换信息传递至在后的视频帧，以供在后的视频帧使用。这样可对当前帧的定位起到纠正作用，减少了错误定位给后续的分割处理所带来误差，大大提高了对视频进行语义分割处理的准确性。

根据本申请的一个方面，提供了一种模型训练方法，应用于计算机设备中，所述方法包括：

获取视频帧样本、所述视频帧样本对应的样本标注信息、及所述视频帧样本对应的标准仿射变换信息；

将所述视频帧样本输入至目标分割模型中进行训练，通过所述目标分割模型，确定与所述视频帧样本对应的预测仿射变换信息；

依据所述预测仿射变换信息和所述标准仿射变换信息构建仿射损失函数；

通过所述目标分割模型，输出与所述视频帧样本对应的预测仿射变换差异信息、及所述视频帧样本中目标对应的预测分割结果；

根据所述预测仿射变换信息和所述标准仿射变换信息间的差异，确定标准仿射变换差异

信息；

依据所述标准仿射变换差异信息和所述预测仿射变换差异信息，构建仿射变换信息修正损失函数；

根据所述预测分割结果和所述样本标注信息，确定分割损失函数；

依据所述仿射损失函数、所述仿射变换信息修正损失函数、及所述分割损失函数，调整所述目标分割模型的模型参数继续训练，直至满足训练停止条件时停止训练。

根据本申请的一个方面，提供了一种模型训练装置，所述装置包括：

样本获取模块，用于获取视频帧样本、所述视频帧样本对应的样本标注信息、及所述视频帧样本对应的标准仿射变换信息；

确定模块，用于将所述视频帧样本输入至目标分割模型中进行训练，通过所述目标分割模型，确定与所述视频帧样本对应的预测仿射变换信息；

构建模块，用于依据所述预测仿射变换信息和所述标准仿射变换信息构建仿射损失函数；

输出模块，用于通过所述目标分割模型，输出与所述视频帧样本对应的预测仿射变换差异信息、及所述视频帧样本中目标对应的预测分割结果；

所述确定模块还用于根据所述预测仿射变换信息和所述标准仿射变换信息间的差异，确定标准仿射变换差异信息；

所述构建模块还用于依据所述标准仿射变换差异信息和所述预测仿射变换差异信息，构建仿射变换信息修正损失函数；

所述构建模块还用于根据所述预测分割结果和所述样本标注信息，确定分割损失函数；

模型参数调整模块，用于依据所述仿射损失函数、所述仿射变换信息修正损失函数、及所述分割损失函数，调整所述目标分割模型的模型参数继续训练，直至满足训练停止条件时停止训练。

根据本申请的一个方面，提供了一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序被处理器执行时，使得所述处理器执行以下步骤：

获取视频帧样本、所述视频帧样本对应的样本标注信息、及所述视频帧样本对应的标准仿射变换信息；

将所述视频帧样本输入至目标分割模型中进行训练，通过所述目标分割模型，确定与所述视频帧样本对应的预测仿射变换信息；

依据所述预测仿射变换信息和所述标准仿射变换信息构建仿射损失函数；

通过所述目标分割模型，输出与所述视频帧样本对应的预测仿射变换差异信息、及所述视频帧样本中目标对应的预测分割结果；

根据所述预测仿射变换信息和所述标准仿射变换信息间的差异，确定标准仿射变换差异信息；

依据所述标准仿射变换差异信息和所述预测仿射变换差异信息，构建仿射变换信息修正损失函数；

根据所述预测分割结果和所述样本标注信息，确定分割损失函数；

依据所述仿射损失函数、所述仿射变换信息修正损失函数、及所述分割损失函数，调整所述目标分割模型的模型参数并继续训练，直至满足训练停止条件时停止训练。

根据本申请的一个方面，提供了一种计算机设备，所述计算机设备包括存储器和处理器，所述存储器存储有计算机程序，所述计算机程序被所述处理器执行时，使得所述处理器执行

以下步骤：

获取视频帧样本、所述视频帧样本对应的样本标注信息、及所述视频帧样本对应的标准仿射变换信息；

将所述视频帧样本输入至目标分割模型中进行训练，通过所述目标分割模型，确定与所述视频帧样本对应的预测仿射变换信息；

依据所述预测仿射变换信息和所述标准仿射变换信息构建仿射损失函数；

通过所述目标分割模型，输出与所述视频帧样本对应的预测仿射变换差异信息、及所述视频帧样本中目标对应的预测分割结果；

根据所述预测仿射变换信息和所述标准仿射变换信息间的差异，确定标准仿射变换差异信息；

依据所述标准仿射变换差异信息和所述预测仿射变换差异信息，构建仿射变换信息修正损失函数；

根据所述预测分割结果和所述样本标注信息，确定分割损失函数；

依据所述仿射损失函数、所述仿射变换信息修正损失函数、及所述分割损失函数，调整所述目标分割模型的模型参数并继续训练，直至满足训练停止条件时停止训练。

上述模型训练方法、装置、计算机可读存储介质和计算机设备，在模型训练过程中一方面引入仿射变换监督信息，也就是标准仿射变换信息，以提高方位预测的准确性；另一方面可通过对预测仿射变换信息进行纠正训练，从而减少错误定位带来的分割误差。训练时将仿射损失函数、仿射变换信息修正损失函数、及分割损失函数叠加一起优化，使得各个部分在训练过程中相互影响，相互提升，这样训练得到的目标分割模型具有准确的视频语义分割性能。

附图说明

图 1 为一个实施例中目标分割方法和/或模型训练方法的应用环境图；

图 2 为一个实施例中图像分割方法的流程示意图；

图 3 为一个实施例中视频帧序列的结构示意图；

图 4 为一个实施例中获取视频帧序列中的当前帧、及在前的视频帧所传递的历史仿射变换信息步骤的流程示意图；

图 5 为一个实施例中目标分割模型的整体框架图；

图 6 为一个实施例中对心脏超声检测视频中的左心室进行目标分割的目标分割模型的架构示意图；

图 7 为一个实施例中目标分割模型的训练步骤的流程示意图；

图 8 为一个实施例中模板的获取流程图；

图 9 为一个实施例中模型训练方法的流程示意图；

图 10 为一个实施例中在模型训练过程中目标分割模型的架构示意图；

图 11 为一个具体实施例中图像分割方法的流程示意图；

图 12 为一个实施例中图像分割装置的结构框图；

图 13 为另一个实施例中图像分割装置的结构框图；

图 14 为一个实施例中模型训练装置的结构框图；

图 15 为一个实施例中计算机设备的结构框图。

具体实施方式

为了使本申请的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请进行进一步详细说明。应当理解，此处所描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。

图1为一个实施例中图像分割方法和/或模型训练方法的应用环境图。参照图1，该图像分割方法和/或模型训练方法应用于语义分割系统。该语义分割系统包括采集器110和计算机设备120。采集器110和计算机设备120可以通过网络连接，也可以通过传输线连接。计算机设备120可以是终端或服务器。其中，终端可以是台式终端或移动终端，移动终端具体可以手机、平板电脑、笔记本电脑等中的至少一种；服务器可以用独立的服务器或者是多个服务器组成的服务器集群来实现。

采集器110可以实时采集视频，将视频传输至计算机设备120，计算机设备120可以获取视频帧序列中的当前帧、及在前的视频帧所传递的历史仿射变换信息；依据历史仿射变换信息对当前帧进行仿射变换，得到与当前帧对应的候选区域图像；对候选区域图像进行特征提取，得到候选区域图像对应的特征图；基于特征图进行语义分割，得到当前帧中的目标对应的分割结果；根据特征图对历史仿射变换信息进行修正，得到更新的仿射变换信息，并将更新的仿射变换信息作为视频帧序列中在后的视频帧所对应的历史仿射变换信息。

需要说明的是，上述的应用环境只是一个示例，在一些实施例中，计算机设备120可以直接获取视频，对视频对应的视频帧序列中的各个视频帧按照上述步骤进行目标分割。

如图2所示，在一个实施例中，提供了一种图像分割方法。本实施例以该方法应用于上述图1中的计算机设备120来举例说明。参照图2，该图像分割方法包括如下步骤：

S202，获取视频帧序列中的当前帧、及在前的视频帧的历史仿射变换信息。

其中，视频帧序列是多于一帧的视频帧按照各视频帧所对应的生成时序而构成的序列。视频帧序列包括：按照生成时序排列的多个视频帧。视频帧是构成视频的基本单位，一段视频可以包括多个视频帧。视频帧序列可以是实时采集的视频帧所构成的序列，例如可以通过采集器的摄像头实时获取的视频帧序列，也可以是存储的视频对应的视频帧序列。

当前帧是当前处理的视频帧，比如第*i*帧；在前的视频帧是生成时间在当前帧之前的视频帧，可以是当前帧的前一帧或当前帧的第前几帧的视频帧，也可称作当前帧的历史视频帧。

历史仿射变换信息是在前的视频帧所传递的用于当前帧进行仿射变换的仿射变换信息。此处的“在前的视频帧所传递的”可以理解为：计算机设备根据在前的视频帧所传递的，或者，在前的视频帧所对应的。仿射变换，又称仿射映射，是指对一个空间向量矩阵进行线性变换后再进行平移变换得到另一个空间向量矩阵的过程，线性变换包括卷积运算。仿射变换信息是用于进行仿射变换所需的信息，可以是仿射变换参数、或用于指示如何进行仿射变换的指令。其中，仿射变换参数是指图像进行线性变换或平移变换所需的参考参数，比如旋转角度（angle）、横轴方向的平移像素（Shift_x），纵轴方向的平移像素（Shift_y）以及缩放系数（Scale）等信息。

计算机设备可以在对视频进行检测的过程中，获取当前帧以及在前的视频帧的历史仿射变换信息。其中，在前的视频帧的历史仿射变换信息，是指依据对在前的视频帧执行该图像分割方法时所得到的已经修正的、且用于当前帧进行仿射变换的参数。计算机设备可通过以下方式得到历史仿射变换信息：计算机设备在对在前的视频帧进行目标分割时，可根据在前的视频帧所对应的特征图，对在前的视频帧对应的仿射变换信息进行修正，得到更新的仿射

变换信息，该更新的仿射变换信息即可作为当前帧的历史仿射变换信息。

可以理解，在对整个视频帧序列进行目标分割的过程中，计算机设备对当前帧执行图像分割方法时，同样可根据当前帧的特征图对该历史仿射变换信息进行修正，得到更新的仿射变换信息，并将更新的仿射变换信息作为视频帧序列中在后的视频帧所对应的历史仿射变换信息。这样，在对视频帧序列进行目标分割时，可不断修正并传递仿射变换信息。这样可对当前帧的定位起到纠正作用，减少了错误定位给后续的分割处理所带来误差，以提高对视频进行语义分割处理的准确性。

可以理解，本申请所使用的“当前帧”用于描述当前本方法所处理的视频帧，“当前帧”是一个相对变化的视频帧，比如在处理当前帧的下一个视频帧时，则可以将该下一个视频帧作为新的“当前帧”。

在一个实施例中，计算机设备可将当前帧的前一帧所传递的历史仿射变换信息作为当前帧对应的仿射变换信息，以进行仿射变换。相应的，下一帧视频帧可将当前帧所传递的历史仿射变换信息作为下一帧对应的仿射变换信息。依次类推，每一帧视频帧均可将前一帧所传递的历史仿射变换信息作为该帧对应的仿射变换信息以进行仿射变换。

可以理解，在另一些实施例中，计算机设备还可将当前帧的前第 N (N 为正整数，且 N 大于 1) 帧所传递的历史仿射变换信息作为当前帧对应的仿射变换信息，以进行仿射变换。相应的，下一帧视频帧可将当前帧的前第 $N-1$ 帧所传递的历史仿射变换信息作为下一帧对应的仿射变换信息。依次类推，每一帧视频帧均可将前第 N 帧所传递的历史仿射变换信息作为该帧对应的仿射变换信息以进行仿射变换。

举例说明，参考图 3，对于视频帧序列[F1,F2,F3,F4,F5,F6]，计算机设备当前所处理的当前帧为 F4，那么当前帧 F4 可使用在前的视频帧 F1 所传递的历史仿射变换信息作为对应的仿射变换信息以进行仿射变换；那么当前帧 F4 可使用在前的视频帧 F1 所传递的历史仿射变换信息作为对应的仿射变换信息以进行仿射变换；视频帧 F5 可使用在前的视频帧 F2 所传递的历史仿射变换信息作为对应的仿射变换信息以进行仿射变换；视频帧 F6 可使用在前的视频帧 F3 所传递的历史仿射变换信息作为对应的仿射变换信息以进行仿射变换等等依次类推。

在一个实施例中，当当前帧为初始视频帧时，步骤 S202，也就是获取视频帧序列中的当前帧、及在前的视频帧的历史仿射变换信息的步骤包括以下步骤：

S402，获取视频帧序列中的初始视频帧。

其中，初始视频帧是视频帧序列中开始的视频帧。初始视频帧可以是视频帧序列中的第一帧视频帧，也可以是视频帧序列中的第 N 帧（比如第一个对焦稳定性达到预设条件的帧，第一个出现目标的帧），也可以是视频帧序列中最靠前的前 N (N 为正整数，且 N 大于 1) 帧视频帧。

可以理解，当计算机设备在执行该图像分割方法，每后一视频帧的仿射变换信息会参考前一帧视频帧的仿射变换信息时，则该初始视频帧为视频帧序列中最开始的视频帧。当计算机设备在执行该图像分割方法，计算机设备将当前帧的前第 N (N 为正整数，且 N 大于 1) 帧所传递的历史仿射变换信息作为当前帧对应的仿射变换信息时，则从该视频帧序列的第一帧开始至前第 N 帧视频帧均可称作初始视频帧。

S404，通过第一卷积神经网络提取初始视频帧的图像特征。

其中，卷积神经网络 (Convolutional Neural Network, 简称 CNN) 是一类包含卷积计算且具有深度结构的前馈神经网络 (Feedforward Neural Networks)。卷积神经网络中的隐含层

内的卷积核参数共享和层间连接的稀疏性的特定，使得卷积神经网络能够以较小的计算量对格点化特征（例如像素和音频）进行学习。卷积神经网络通常包括卷积层和池化层，可对输入的图像进行卷积和池化处理，以将原始数据映射到隐层特征空间。而图像特征是通过卷积神经网络处理后所得到的能够表示该初始视频帧的图像信息的空间向量矩阵。

可选地，该图像分割方法通过目标分割模型执行，计算机设备可将视频帧序列输入至目标分割模型中，通过目标分割模型中的第一卷积神经网络对初始视频帧进行处理，提取初始视频帧中的特征，得到相应的图像特征。S406，将图像特征输入至第一全连接网络，并通过第一全连接网络对图像特征进行处理，通过第一全连接网络的至少一个输出通道输出仿射变换信息。

全连接网络（Fully Connected Network）也可称作全连接层（fully connected layers, FC），全连接层在整个卷积神经网络中起到“分类器”的作用。全连接层可将卷积层和池化层所学到的图像特征映射到样本标记空间。

可选地，计算机设备可将图像特征输入至第一全连接网络，并通过该第一全连接网络对图像特征进行处理，通过第一全连接网络的至少一个输出通道输出仿射变换信息。

在一个实施例中，目标分割模型包括区域仿射网络（Region Affine Networks, 简称 RAN），该 RAN 网络包括卷积神经网络和全连接网络。可选地，计算机设备在 RAN 网络中输入视频帧序列中的初始视频帧，通过轻量级 MobileNet-V2 网络（轻量化网络）作为 Generator（生成器）抽取初始视频帧的图像特征，再通过一个输出通道（channel）为 4 的全连接网络回归出 4 个仿射变换参数，这 4 个参数分别为旋转角度、横轴方向的平移像素，纵轴方向的平移像素以及缩放系数。

S408，将输出的仿射变换信息作为初始视频帧对应的历史仿射变换信息。

可选地，计算机设备可将第一全连接网络所输出的仿射变换信息作为该初始视频帧所对应的仿射变换信息，并依据该仿射变换信息进行仿射变换，得到与初始视频帧对应的候选区域图像。

可以理解，对于初始视频帧而言，初始视频帧并不存在与之对应的可参考的在前的视频帧，因而也没有在前的视频帧的历史仿射变换信息供其使用。

在一个实施例中，目标分割模型在训练时可引入与仿射变换信息对应的监督信息作为训练样本进行模型训练。其中，与仿射变换对应的监督信息可以是与视频帧样本对应的标准仿射变换信息。该标准仿射变换信息是指将视频帧样本转换成模板（Template）所需要的仿射变换信息。该标准仿射变换信息可通过视频帧样本所包括的样本关键点位置信息和模板所包括的模板关键点位置信息进行反射相似度计算所得。其中，关于该模板是如何获得的、以及目标分割模型的训练过程，在后续的训练方法中会进行详细的介绍。

这样，通过引入与仿射变换信息对应的监督信息来训练目标分割模型，可使得该目标分割模型中的区域仿射网络学习到模板的信息，从而可准确地回归出初始视频帧相对于模板的仿射变换信息。

上述实施例中，通过卷积神经网络提取初始视频帧的图像特征，并通过第一全连接网络对图像特征进行处理，可预测出与初始视频帧对应的、且准确性更高的仿射变换信息，从而有助于提高后续处理中对目标进行分割的准确性。

在一个实施例中，当当前帧不为初始视频帧时，在缓存中读取在前的视频帧的历史仿射变换信息。

S204, 依据历史仿射变换信息对当前帧进行仿射变换, 得到与当前帧对应的候选区域图像。

可选地, 计算机设备依据历史仿射变换信息对当前帧进行仿射变换, 可以是依据仿射变换信息对当前帧中的目标所对应的位置、尺寸及方位等进行了纠正, 得到对应的候选区域图像。其中, 候选区域图像也可称作感兴趣区域 (ROI)。

在一个实施例中, 计算机设备可将视频帧序列输入至目标分割网络, 通过该目标分割模型执行该图像分割方法。其中, 目标分割模型是用于对视频中的目标对象进行语义分割的模型, 可以是机器学习模型。该目标分割模型可包括多个网络结构, 不同的网络结构包括各自网络所对应的模型参数, 不同的网络结构用于执行不同的动作。

在一个实施例中, 计算机设备可将视频帧序列输入至目标分割模型中, 通过目标分割模型所包括的 RAN 网络, 依据历史仿射变换信息对当前帧进行仿射变换, 得到与当前帧对应的候选区域图像。

S206, 对候选区域图像进行特征提取, 得到候选区域图像对应的特征图。

其中, 特征图又称 feature map, 是通过卷积神经网络对图像进行卷积和/或池化处理后所得到的空间向量矩阵, 可用于表示该图像的图像信息。可选地, 计算机设备可对候选区域图像进行特征提取, 得到候选区域图像对应的特征图。

在一个实施例中, 计算机设备可通过目标分割模型中的第二卷积神经网络, 对候选区域图像进行特征提取, 得到候选区域图像对应的特征图。该卷积神经网络可以是 MobileNet-V2、VGG (Visual Geometry Group, 视觉集合组) 网络、或 ResNet (deep Residual learning, 深度残差学习) 网络等。

其中, 第二卷积神经网络可以和第一卷积神经网络共享参数, 因而可以认为是相同的卷积神经网络, 此处用“第一”“第二”主要是用于区分处于目标分割模型中的不同位置处、且用于处理不同数据的卷积神经网络。

在一个实施例中, 对候选区域图像进行特征提取所得到的特征图, 融合了视频帧序列所包括的光流信息。

其中, 光流信息是图像的运动变化信息, 在本申请实施例中可用于表示视频帧序列中各像素点在视频帧中移动的信息, 包括视频画面中待检测目标的运动变化信息。在本申请实施例中, 前一帧视频帧所对应的光流信息可通过前一帧视频帧中的各像素所对应的位置、及当前帧中各像素所对应的位置来确定。

在一个实施例中, 可假定相邻两帧视频帧中目标对应的变化是较为微小的, 因而当前帧中目标对象所在的目标分割区域可以通过前一帧视频帧所对应的光流信息确定。比如, 当前帧中目标对象所在的目标分割区域, 可根据光流信息, 以及前一帧视频帧中目标对象所在的目标分割区域共同预测。

为使得目标分割模型中的卷积神经网络在对候选区域图像进行特征提取时, 可以融合对应的光流信息, 使得提取出的特征图融合光流信息, 那么在对目标分割模型的卷积神经网络进行训练时, 可引入判别器 (discriminator) 来共同训练。其中, 生成器和判别器共同构成生成式对抗网络 (Generative Adversarial Nets, GAN)。

在模型训练阶段, 对于当前帧所对应的特征图有两种特征形式: 一种是通过第二卷积神经网络基于当前帧所对应的候选区域图像而提取的特征图, 可称作 CNN 特征; 另一种是通过光流信息基于上一帧视频帧的特征图进行变换而得到的特征图, 可称作光流特征。为此, 可

设计判别器将这两种信息同时引入。也就是说，在模型训练过程中，可分别将 CNN 特征和光流特征中的任意一种输入至判别器中，判别器判断当前输入的特征是属于光流特征还是 CNN 特征。通过不断调整第二卷积神经网络的参数和判别器的参数，使得判别器无法分辨 CNN 特征和光流特征的区别，那么此时的第二卷积神经网络就可以生成融合了光流信息的特征图。其中，关于判别器和第二卷积神经网络之间更详细的训练过程，在后续模型训练阶段的实施例中将会有详细的描述。

上述实施例中，对候选区域图像进行特征提取所得到的特征图融合了视频帧序列所包括的光流信息，可避免分割结果出现误差，从而产生具有时序渐进性的合理分割结果。

S208，基于特征图进行语义分割，得到当前帧中的目标对应的分割结果。

其中，语义分割是指计算机设备将图片中属于一大类的区域都分割出来并给出其类别信息。分割结果可以是当前帧中属于目标对象的像素点构成的目标分割区域。

可选地，计算机设备可对特征图进行像素维度的检测，也就是基于候选区域图像所对应的特征图，对候选区域图像中每个像素进行检测，输出当前帧中的目标对应的检测结果。在一个实施例中，计算机设备可识别候选区域图像中各个像素各自所对应的类别，根据对应目标类别的各像素点构成目标区域。也就是将目标对象从候选区域图像中区分开来。

在一个实施例中，计算机设备可以通过目标分割模型中的全卷积神经网络对候选区域图像特征进行语义分割，输出当前帧中的目标对应的检测结果。

在一个实施例中，步骤 S208，也就是基于特征图进行语义分割，得到当前帧中的目标对应的分割结果的步骤包括：通过全卷积神经网络对特征图进行上采样处理，得到中间图像；通过全卷积神经网络对中间图像中的各像素分别进行像素级分类，得到各像素所对应的类别；依据各像素所对应的类别，输出对当前帧中的目标进行语义分割的分割结果。

其中，全卷积神经网络（Fully Convolutional Networks，简称 FCN）通常用于对输入图像进行逐像素分类。全卷积神经网络通常可采用反卷积层对最后一个卷积层的 feature map 进行上采样（Upsample），使它恢复到输入图像相同的尺寸，从而可以对每个像素都产生了一个预测，同时保留了原始输入图像中的空间信息，最后在上采样的特征图上进行逐像素分类。

像素级是指像素维度；像素级分类是指在像素维度上进行分类处理，是一种精细的分类方式。对中间图像中的各像素分别进行像素级分类，也可称作对中间图像进行像素级的分类，是对中间图像中的每个像素都产生一个预测，进而得到中间图像中每个像素各自所对应的类别。

可选地，计算机设备可通过目标分割模型中的全卷积神经网络对当前帧所对应的特征图进行上采样处理，得到中间图像，通过全卷积神经网络对中间图像中的各像素分别进行像素级分类，得到各像素所对应的类别。比如若候选区域图像中属于目标对象的像素点的类别为 1，不属于目标对象的像素点的类别为 0，则候选区域图像所有类别为 1 的像素点所构成的区域为目标分割区域，据此可将目标区域从候选区域图像中分割出来。比如通过红色或绿色突出显示目标分割区域。

在一个实施例中，依据各像素所对应的类别，输出对当前帧中的目标进行语义分割的分割结果的步骤包括：确定中间图像中对应目标类别的像素；从中间图像中，分割出由对应目标类别的各像素所组成的、且包括目标对象的目标分割区域。

在一个实施例中，在对该目标分割模型的全卷积神经网络进行训练时，可依据视频帧样本、及对视频帧样本中的目标对象进行标注的样本标注信息来训练该全卷积神经网络，训练

得到的该全卷积神经网络具备对像素进行分类的能力。其中，对视频帧样本中的目标对象进行标注的样本标注信息，可以是将对对应目标对象的像素标记为“1”，将其他的像素标记为“0”，以此来区分目标对象个非目标对象。

在一个实施例中，计算机设备可通过目标分割模型中的全卷积神经网络确定中间图像中对应目标类别的像素。并对属于目标类别的像素进行标注，比如将标注成红色或绿色等，以此从中间图像中，分割出由对应目标类别的各像素所组成的、且包括目标对象的目标分割区域。这样可实现在当前帧中准确地定位到目标对象，并可以准确地确定目标对象在当前帧中所占的面积大小。

在一个实施例中，计算机设备可以根据每一个视频帧的检测结果显示目标对象，以实现在连续的视频帧构成的视频中对目标进行自动分割的效果。

上述实施例中，通过全卷积神经网络对特征图进行像素级分类，可得到各像素各自所对应的类别，从而依据各像素所对应的类别，可从像素级别准确地确定出当前帧中的目标所在的目标分割区域，大大提高了对目标对象的分割能力。

S210，根据特征图对历史仿射变换信息进行修正，得到更新的仿射变换信息，将更新的仿射变换信息作为视频帧序列中在后的视频帧所对应的历史仿射变换信息。

其中，对历史仿射变换信息进行修正是指调整历史仿射变换参数，得到更新的仿射变换参数。可选地，计算机设备可根据特征图对历史仿射变换信息进行修正，得到更新的仿射变换信息，该更新的仿射变换信息可作为视频帧序列中在后的视频帧所对应的仿射变换信息。

在一个实施例中，计算机设备可通过目标分割模型所包括的第二全连接网络，对当前帧所对应的特征图进行处理，对该仿射变换信息进行修正，得到更新的仿射变换信息。

在一个实施例中，该目标分割模型所包括的第二全连接网络，可被训练成输出仿射变换差异结果，再依据仿射变换差异结果和在前的视频帧所传递的历史仿射变换信息，计算得到当前帧所传递的更新的仿射变换信息。计算机设备则可直接将该更新的仿射变换信息传递至在后的视频帧，供在后的视频帧进行仿射变换使用。

在一个实施例中，步骤 S210，也就是根据特征图对历史仿射变换信息进行修正，得到更新的仿射变换信息，并将更新的仿射变换信息作为视频帧序列中在后的视频帧所对应的历史仿射变换信息的步骤包括以下步骤：通过第二全连接网络，对特征图进行处理，通过第二全连接网络的至少一个输出通道输出仿射变换差异结果；依据仿射变换差异结果和在前的视频帧所传递的历史仿射变换信息，计算得到当前帧所传递的更新的仿射变换信息；将当前帧所传递的更新的仿射变换信息，作为视频帧序列中在后的视频帧所对应的历史仿射变换信息。

其中，第二全连接网络和第一全连接网络是相同的全连接网络，或者，是不同的全连接网络。其中，相同的全连接网络是指第一全连接网络和第二全连接网络的参数共享；不同的全连接网络是指第一全连接网络和第二全连接网络的具有各自的模型参数。

可选地，该目标分割模型所包括的第二全连接网络，可被训练成输出仿射变换差异结果。在这种情况下，可通过目标分割模型中的第二全连接网络对与当前帧对应的特征图进行处理，回归出仿射变换差异结果。可选地，该差异结果是进行归一化处理后的差异率。

进一步地，计算机设备可依据仿射变换差异结果和在前的视频帧所传递的历史仿射变换信息，计算得到当前帧所传递的更新的仿射变换信息。比如，当仿射变换信息为仿射变换参数时，计算机设备可通过以下公式计算得到更新的仿射变换信息：

$$\hat{\theta}'_{t+1} = f(\hat{\theta}'_t, \hat{\theta}_t) = \hat{\theta}'_t \times \hat{\theta}_t + \hat{\theta}_t;$$

其中, $\hat{\theta}'_{t+1}$ 表示当前帧所传递的更新的仿射变换参数; $\hat{\theta}'_t$ 表示仿射变换差异结果; $\hat{\theta}_t$ 表示当前帧所对应的仿射变换参数, 也就是在前的视频帧所传递的历史仿射变换参数。

进而, 计算机设备可将计算得到的更新的仿射变换信息, 作为当前帧所传递的历史仿射变换信息, 也就是将该更新的仿射变换信息传递至视频帧序列中在后的视频帧, 以供在后的视频帧依据该更新的仿射变换信息进行仿射变换。

可以理解, 当目标分割模型所包括的第二全连接网络被训练成输出仿射变换差异结果时, 该第二全连接网络在训练过程中的监督信息可以是标准仿射变换信息和当前帧所对应的仿射变换信息的差异信息。

上述实施例中, 通过第二全连接网络对特征图进行处理, 以纠正当前帧所使用的仿射变换信息, 得到更新的仿射变换信息。更新的仿射变换信息用于向后传递, 这样可对当前帧的定位起到纠正作用, 减少了错误定位带来的分割误差。

在一个实施例中, 该目标分割模型所包括的第二全连接网络可被训练成输出经纠正过的更新的仿射变换信息。计算机设备则可直接将该更新的仿射变换信息传递至在后的视频帧, 供在后的视频帧进行仿射变换使用。

可以理解, 当目标分割模型所包括的第二全连接网络被训练成输出经纠正过的更新的仿射变换信息时, 该第二全连接网络在训练过程中的监督信息可以是当前帧所对应的标准仿射变换信息。

上述图像分割方法, 依据在前的视频帧所传递的历史仿射变换信息, 对当前帧进行仿射变换, 得到与当前帧对应的候选区域图像。在前的视频帧所传递的历史仿射变换信息是经过修正后的参数, 这样可大大提高候选区域图像获取的准确性。对与候选区域图像对应的特征图进行语义分割, 可以准确得到当前帧中的目标对应的分割结果。并且, 根据该特征图对历史仿射变换信息进行修正, 将修正后的仿射变换信息传递至在后的视频帧, 以供在后的视频帧使用。这样可对当前帧的定位起到纠正作用, 减少了错误定位给后续的分割处理所带来误差, 大大提高了对视频进行语义分割处理的准确性。

在一个实施例中, 该图像分割方法通过目标分割模型执行, 该图像分割方法包括以下步骤: 获取视频帧序列中的当前帧、及在前的视频帧的历史仿射变换信息; 通过目标分割模型中的区域仿射网络, 依据历史仿射变换信息对当前帧进行仿射变换, 得到与当前帧对应的候选区域图像; 通过目标分割模型中的第二卷积神经网络, 对候选区域图像进行特征提取, 得到候选区域图像对应的特征图; 通过目标分割模型中的全卷积神经网络, 对特征图进行语义分割处理, 得到当前帧中的目标对应的分割结果; 通过目标分割模型中的第二全连接网络对历史仿射变换信息进行修正, 得到更新的仿射变换信息, 并将更新的仿射变换信息作为视频帧序列中在后的视频帧所对应的历史仿射变换信息。

这样, 通过已训练好的目标分割模型可自动化、且准确的分割出视频中的目标对象, 具有极强的实时性。并且端到端网络工程化程度高, 极易迁移到移动设备中, 自适应能力高。

如图 5 所示, 为一个实施例中目标分割模型的整体框架图。参照图 5, 整体框架图包括区域仿射网络 (RAN) 510、第二卷积神经网络 (generator) 520、全卷积神经网络 530 以及第二全连接网络 540。其中, 区域仿射网络 510 包括第一卷积神经网络 (generator) 512 和第一全连接网络 514。

在对视频中的目标对象进行目标分割时,按帧输入视频帧序列中的各个视频帧,若当前帧为初始视频帧,则通过第一卷积神经网络 512 对初始视频帧进行特征提取,得到图像特征,并将图像特征输入至第一全连接网络 514 中回归出当前的仿射变换信息。通过区域仿射网络 510 依据当前的仿射变换信息对初始视频帧进行仿射变换,得到对应的候选区域图像(ROI)。再通过第二卷积神经网络 520 对候选区域图像进行特征提取,得到候选区域图像对应的特征图。该特征图进入两个任务分支,在分割任务分支中,通过全卷积神经网络 530 进行上采样处理后得到分割预测图,输出分割结果;在定位任务分支中,通过第二全连接网络回归出仿射变换差异结果。再依据仿射变换差异结果纠正当前帧所对应的仿射变换信息,得到更新的仿射变换信息,将该更新的仿射变换信息传递至下一帧。

如图 5 所示,在下一帧视频帧中,RAN 网络依据更新的仿射变换信息对下一帧视频帧进行仿射变换,得到下一帧视频帧所对应的 ROI 区域,并通过第二卷积神经网络 520 对候选区域图像进行特征提取,得到候选区域图像对应的特征图。该特征图进入两个任务分支,在分割任务分支中,通过全卷积神经网络 530 进行上采样处理后得到分割预测图,输出分割结果;在定位任务分支中,通过第二全连接网络回归出仿射变换差异结果。再依据仿射变换差异结果纠正下一帧视频帧所对应的仿射变换信息,得到更新的仿射变换信息,将该更新的仿射变换信息传递至在后的视频帧。依次类推,最终实现对视频中的目标进行分割的效果。

在一个实施例中,视频帧序列属于对生物组织进行医学检测得到的检测视频,比如可以是心脏超声检测视频。视频帧中的目标为左心室,检测结果为分割出视频帧中的左心室。

如图 6 所示,图 6 为一个实施例中对心脏超声检测视频中的左心室进行目标分割的架构示意图。在该示意图中,前一帧为 $t-1$ 帧;当前帧为 t 帧。参照图 6,对心脏超声检测视频进行目标分割,对于前一帧视频帧,可通过 RAN 网络中的生成器和全连接网络生成预测的仿射变换信息 $\hat{\theta}_{t-1}$;再依据仿射变换信息 $\hat{\theta}_{t-1}$ 进行仿射变换 $R(\hat{\theta}_{t-1})$ 得到前一帧视频帧的候选区域图像 ROI。再通过生成器提取图像特征后分别进入分割任务分支和定位任务分支,得到分割结果 $t-1$ 、及仿射变换差异参数 $\hat{\theta}'_{t-1}$ 。

该仿射变换差异参数传递到当前帧,区域仿射网络依据仿射变换差异参数 $\hat{\theta}'_{t-1}$ 和预测的仿射变换信息 $\hat{\theta}_{t-1}$,来对当前帧进行仿射变换,如 $R(f(\hat{\theta}'_{t-1}, \hat{\theta}_{t-1}))$,得到候选区域图像 ROI。再通过生成器提取图像特征后分别进入分割任务分支和定位任务分支,得到分割结果 t 、及仿射变换差异参数 $\hat{\theta}'_t$ 。以此循环类推,从而实现了心脏超声检测视频中的左心室进行标注分割。

参考图 7,在一个实施例中,该图像分割方法通过目标分割模型执行,该目标分割模型的训练步骤包括:

S602,获取视频帧样本、视频帧样本对应的样本标注信息、及视频帧样本对应的标准仿射变换信息。

其中,视频帧样本、视频帧样本对应的样本标注信息、及视频帧样本对应的标准仿射变换信息为训练数据。视频帧样本对应的样本标注信息可以是对视频帧样本中的关键点进行标注的样本关键点位置信息、及对视频帧样本中的目标对象进行标注的样本区域位置信息。其中,视频帧样本中的关键点是用于确定目标对象的关键点,关键点的数量可以是 3 个、4 个或其他数量等。

以心脏超声检测视频为例,视频帧序列中的目标对象为左心室,那么相应的视频帧样本中的关键点可以是左心室尖端、及左心室二尖瓣膜两端,样本关键点位置信息可以是左心室

尖端、及左心室二尖瓣膜两端对位置信息；样本区域位置信息可以是视频帧样本中左心室所在的区域的位置信息。

标准仿射变换信息是视频帧样本相对于模版的仿射变换信息，也就是说视频帧样本依据该标准仿射变换信息，可进行仿射变换得到模板。其中，模板是依据多个视频帧样本统计出的可以代表标准视频帧的图像。

在一个实施例中，步骤 S602，也就是获取视频帧样本、视频帧样本对应的样本标注信息、及视频帧样本对应的标准仿射变换信息对步骤包括以下步骤：获取视频帧样本和相应的样本标注信息；样本标注信息包括样本关键点位置信息和样本区域位置信息；根据视频帧样本、样本关键点位置信息和样本区域位置信息，确定模板图像及模板图像对应的模板关键点位置信息；根据样本关键点位置信息和模板关键点位置信息，计算得到与视频帧样本对应的标准仿射变换信息。

可选地，计算机设备可从本地或其他计算机设备处获取多个视频帧样本。并对该视频帧样本采用人工标注或机器标注的方式标注出样本关键点和目标对象在视频帧样本中的位置区域。

进而计算机设备可根据多个包括样本标注信息的视频帧样本，确定模板、以及模板中的模板关键点位置信息。可选地，计算机设备可对多个视频帧样本中的关键点位置信息求平均后得到模板关键点位置信息。

比如，计算机设备可依据对每个视频帧样本中的关键点确定包括有目标对象的区域框，将该区域框外扩一定的范围，得到这个视频帧样本的 ROI。再计算所有视频帧样本对应的 ROI 的平均尺寸，并将所有视频帧样本对应的 ROI 调整到平均尺寸。对所有调整到平均尺寸的 ROI 图像求平均即可得到模板。各个 ROI 图像中的关键点的位置信息求平均即可得到模板的关键点位置信息。

下面以心脏超声检测视频为例详细说明书模板的获取步骤，参考图 8，图 8 为一个实施例中模板的获取流程图。如图 8 所示，计算机设备可预先通过采集器采集多种标准的心脏切面，比如 A2C (apical-2-chamber, A2C, 二腔切面)、A3C (apical-3-chamber, A3C, 三腔切面)、A4C (apical-4-chamber, A4C, 四腔切面)、A5C (apical-5-chamber, A5C, 五腔切面) 等作为原始图片，也就是作为视频帧样本，再将每张切面图中的 3 个关键点紧密外扩得到区域框，考虑到各种标准切面中的左心室都在右上方位置，为了获得更多心脏结构信息，可将区域框往左边、往下边各外扩一定比例，比如长宽的 50%。最后，区域框四周在这个框基础上外扩一定比例，比如长宽的 5%，得到这张切面图的 ROI。所有切面图的 ROI 调整尺寸到一个尺度（该尺寸为所有 ROI 的平均尺寸），求平均则得到模板。

进一步地，计算机设备可依据各个视频帧样本的尺寸、关键点位置信息，以及模板的尺寸、模板关键点位置信息，进行反射相似度计算，得到变换矩阵，该变换矩阵中包括仿射变换信息，通过该方法计算得到的仿射变换信息即为与该视频帧样本对应的标准仿射变换信息。

上述实施例中，根据视频帧样本、样本关键点位置信息和样本区域位置信息，可确定模板图像及模板图像对应的模板关键点位置信息。从而可将每张视频帧样本均与模板进行比较，以确定标准仿射变换信息，该标准仿射变换信息可作为后续模型训练的监督信息，用以使得目标分割模型可学习到模板的信息，从而大大提高仿射变换信息的预测准确性。

S604，将视频帧样本输入至目标分割模型中进行训练，通过目标分割模型，确定与视频帧样本对应的预测仿射变换信息。

可选地, 计算机设备可将视频帧样本输入到目标分割模型中, 根据目标分割模型执行前述的图像分割方法, 通过 RAN 网络获取与视频帧样本对应的预测仿射变换信息。

S606, 依据预测仿射变换信息和标准仿射变换信息构建仿射损失函数。

其中, 仿射损失函数用于评估预测仿射变换信息和标准仿射变换信息之间的差异程度。仿射损失函数承担了训练得到好的 RAN 网络的责任, 使目标分割模型中的 RAN 网络可以生成相对于模板来说准确的仿射变换信息, 这样引入仿射监督信息的使得仿射参数预测更加准确。

可选地, 计算机设备可依据预测仿射变换信息和标准仿射变换信息构建仿射损失函数。在一个实施例中, 计算机设备可通过距离函数, 比如 L1-Norm (L1-范数, 又称曼哈顿距离) 函数来计算预测仿射变换信息和标准仿射变换信息的损失, 也就是基于 L1-Norm 函数来构建预测仿射变换信息和标准仿射变换信息的仿射损失函数。可以理解, 在本申请实施例中, 也可采用其他的函数来构建损失函数, 只要该函数可以用来衡量预测仿射变换信息和标准仿射变换信息之间的差异程度即可, 比如 L2-Norm (又称欧几里德距离) 函数等。

S608, 通过目标分割模型, 输出与视频帧样本对应的预测仿射变换差异信息、及视频帧样本中目标对应的预测分割结果。

可选地, 计算机设备可将视频帧样本输入到目标分割模型中, 根据目标分割模型执行前述的图像分割方法, 输出与视频帧样本对应的预测仿射变换差异信息、及视频帧样本中目标对应的预测分割结果。

在一个实施例中, 计算机设备可通过目标分割模型中的 RAN 网络, 依据预测仿射变换信息对视频帧样本进行仿射变换, 得到对应的样本候选区域图像。通过目标分割模型中的第二卷积神经网络并对样本候选区域图像进行特征提取, 得到对应的样本特征图。通过目标分割模型中的全卷积神经网络, 对样本特征图进行语义分割, 得到视频帧样本中的目标对应的预测分割结果。通过目标分割模型中的第二全连接网络, 基于样本特征图对预测仿射变换信息进行修正, 得到与视频帧样本对应的预测仿射变换差异信息。

S610, 根据预测仿射变换信息和标准仿射变换信息间的差异, 确定标准仿射变换差异信息。

其中, 标准仿射变换差异信息是作为目标分割模型中仿射变换修正模块的监督信息, 也就是作为第二全连接网络在训练过程中的监督信息。可选地, 计算机设备可根据预测仿射变换信息和标准仿射变换信息间的差异, 确定标准仿射变换差异信息。比如, 当仿射变换信息为仿射变换参数时, 计算机设备可通过以下公式计算标准仿射变换差异信息:

$$\hat{\theta}_t^{\text{标准}} = \frac{\theta_t - \hat{\theta}_t}{\hat{\theta}_t}, \theta \in \{Angle, Shift_x, Shift_y, Scale\};$$

其中, $\hat{\theta}_t^{\text{标准}}$ 表示标准仿射变换差异参数; $\hat{\theta}_t$ 表示当前帧所对应的仿射变换参数, 也就是预测仿射变换参数; θ_t 表示标准仿射变换参数。

S612, 依据标准仿射变换差异信息和预测仿射变换差异信息, 构建仿射变换信息修正损失函数。

其中, 仿射变换信息修正损失函数用于评估预测仿射变换差异信息和标准仿射变换差异信息之间的差异程度。仿射变换信息修正损失函数承担了训练得到好的第二全连接网络的责任, 使目标分割模型中的第二全连接网络可以生成对预测仿射变换信息进行修正后的仿射变换差异信息。

可选地, 计算机设备可依据标准仿射变换差异信息和预测仿射变换差异信息, 构建仿射变换信息修正损失函数。在一个实施例中, 计算机设备可通过距离函数, 比如 L1-Norm 函数来计算标准仿射变换差异信息和预测仿射变换差异信息的损失, 也就是基于 L1-Norm 函数来构建仿射变换信息修正损失函数。可以理解, 在本申请实施例中, 也可采用其他的函数来构建仿射变换信息修正损失函数, 只要该函数可以用来衡量标准仿射变换差异信息和预测仿射变换差异信息之间的差异程度即可, 比如 L2-Norm 函数等。

可以理解, 该预测仿射变换差异信息用于确定更新的仿射变换信息, 并传递至视频帧序列中在后的视频帧。当仿射变换信息为仿射变换参数时, 可通过以下公式计算更新的仿射变换参数: $\hat{\theta}'_{t+1} = f(\hat{\theta}'_t, \hat{\theta}_t) = \hat{\theta}'_t \times \hat{\theta}_t + \hat{\theta}_t$; 其中, $\hat{\theta}'_{t+1}$ 表示当前帧所传递的更新的仿射变换参数; $\hat{\theta}'_t$ 表示预测仿射变换差异参数; $\hat{\theta}_t$ 表示预测仿射变换参数。

S614, 根据预测分割结果和样本标注信息, 确定分割损失函数。

其中, 分割损失函数用于评估预测分割结果和样本标注信息之间的差异程度。分割损失函数承担了训练得到好的全卷积神经网络的责任, 使目标分割模型中的全卷积神经网络可以准确地从输入的视频帧中分割出目标对象。可选地, 计算机设备可根据预测分割结果和样本标注信息, 确定分割损失函数。

S616, 依据仿射损失函数、仿射变换信息修正损失函数、及分割损失函数, 调整目标分割模型的模型参数并继续训练, 直至满足训练停止条件时停止训练。

其中, 训练停止条件是结束模型训练的条件。训练停止条件可以是达到预设的迭代次数, 或者是调整模型参数后的目标分割模型的性能指标达到预设指标。调整目标分割模型的模型参数, 是对目标分割模型的模型参数进行调整。

可选地, 计算机设备可依据仿射损失函数、仿射变换信息修正损失函数、及分割损失函数, 共同调整目标分割模型中各个网络结构的模型参数并继续训练, 直至满足训练停止条件时停止训练。

可以理解, 对于每个损失函数, 计算机设备可朝着减小相应的预测结果和参考参数之间的差异的方向, 调整模型参数。这样, 通过不断的输入视频帧样本, 得到预测仿射变换信息、预测仿射变换差异信息、及预测分割结果, 根据预测仿射变换信息与标准仿射变换信息之间的差异、预测仿射变换差异信息与标准仿射变换差异信息之间的差异、及预测分割结果和样本标注信息之间的差异调整模型参数, 以训练目标分割模型, 得到训练好的目标分割模型。

上述实施例中, 在模型训练过程中一方面引入仿射变换监督信息, 也就是标准仿射变换信息, 以提高方位预测的准确性; 另一方面可通过对预测仿射变换信息进行纠正训练, 从而减少错误定位带来的分割误差。训练时将仿射损失函数、仿射变换信息修正损失函数、及分割损失函数叠加一起优化, 使得各个部分在训练过程中相互影响, 相互提升, 这样训练得到的目标分割模型具有准确的视频语义分割性能。

参考图 9, 在一个实施例中, 该模型训练方法包括以下步骤:

S802, 获取第一视频帧样本和第二视频帧样本; 第一视频帧样本为第二视频帧样本在前的视频帧。

其中, 第一视频帧样本和第二视频帧样本是不同的视频帧样本。第一视频帧样本为第二视频帧样本在前的视频帧, 也就是说第一视频帧样本的生成时间在第二视频帧之前。在一个实施例中, 第一视频帧样本和第二视频帧样本可以是相邻的视频帧。

S804, 分别获取与第一视频帧样本及第二视频帧样本各自对应的样本标注信息、及与第一视频帧样本对应的标准仿射变换信息。

可选地, 计算机设备可分别获取与第一视频帧样本及第二视频帧样本各自对应的样本标注信息、及与第一视频帧样本对应的标准仿射变换信息。其中, 样本标注信息可包括样本关键点位置信息和样本区域位置信息。标准仿射变换信息的获取步骤可参考前述实施例中所描述的获取步骤。

S806, 将第一视频帧样本和第二视频帧样本作为样本对输入至目标分割模型中进行训练, 通过目标分割模型对第一视频帧样本进行处理, 得到与第一视频帧样本对应的预测仿射变换信息。

可选地, 参考图 10, 图 10 为一个实施例中在模型训练过程中目标分割模型的架构示意图。如图 10 所示, 计算机设备可将相邻的前后两帧视频帧样本作为样本对输入至目标分割模型中。通过目标分割模型对第一视频帧样本进行处理, 得到与第一视频帧样本对应的预测仿射变换信息 $\hat{\theta}_{t-1}$ 。

S808, 依据预测仿射变换信息和标准仿射变换信息构建仿射损失函数。

可选地, 计算机设备可依据预测仿射变换信息和标准仿射变换信息构建仿射损失函数。在一个实施例中, 计算机设备可通过距离函数, 比如 L1-Norm 函数来计算预测仿射变换信息和标准仿射变换信息的损失, 也就是基于 L1-Norm 函数来构建预测仿射变换信息和标准仿射变换信息的仿射损失函数。可以理解, 在本申请实施例中, 也可采用其他的函数来构建损失函数, 只要该函数可以用来衡量预测仿射变换信息和标准仿射变换信息之间的差异程度即可, 比如 L2-Norm 函数等。

S810, 依据预测仿射变换信息对第一视频帧样本进行仿射变换, 得到第一样本候选区域图像, 并对第一样本候选区域图像进行特征提取, 得到第一样本特征图。

可选地, 参考图 10 上半部分, 计算机设备可依据预测仿射变换信息对第一视频帧样本进行仿射变换, 得到第一样本候选区域图像, 并通过 Generator (生成器, 可通过卷积神经网络实现) 对第一样本候选区域图像进行特征提取, 得到与第一视频帧样本对应的第一样本特征图。

S812, 基于第一样本特征图进行语义分割, 得到第一视频帧样本中的目标对应的预测分割结果。

可选地, 参考图 10, 该第一样本特征图进行两个任务分支, 其中一个任务分支是分割任务分支。目标分割模型可通过全卷积神经网络对第一样本特征图进行语义分割处理, 通过全卷积神经网络进行两次上采样处理后, 基于各个像素预测, 得到第一视频帧样本中的目标对应的预测分割结果。

S814, 根据第一样本特征图对预测仿射变换信息进行修正, 得到与第一视频帧样本对应的预测仿射变换差异信息。

可选地, 参考图 10, 第二个任务分支就是定位任务分支, 在定位任务分支中, 第一样本特征图通过 channel 为 4 的全连接层回归出新的仿射变换差异参数, 也就是预测仿射变换差异信息。

S816, 根据预测仿射变换信息和标准仿射变换信息间的差异, 确定标准仿射变换差异信息。

可选地, 计算机设备可根据预测仿射变换信息和标准仿射变换信息间的差异, 确定标准

仿射变换差异信息。比如，当仿射变换信息为仿射变换参数时，计算机设备可通过以下公式计算标准仿射变换差异信息： $\hat{\theta}'_t \text{标准} = \frac{\theta_t - \hat{\theta}_t}{\hat{\theta}_t}$, $\theta \in \{Angle, Shift_x, Shift_y, Scale\}$; 其中， $\hat{\theta}'_t$ 标准表示标准仿射变换差异信息； $\hat{\theta}_t$ 表示当前帧所对应的仿射变换参数，也就是预测仿射变换参数； θ_t 表示标准仿射变换参数。

S818, 依据标准仿射变换差异信息和预测仿射变换差异信息，构建仿射变换信息修正损失函数。

可选地，计算机设备可依据标准仿射变换差异信息和预测仿射变换差异信息，构建仿射变换信息修正损失函数。在一个实施例中，计算机设备可通过距离函数，比如 L1-Norm 函数来计算标准仿射变换差异信息和预测仿射变换差异信息的损失，也就是基于 L1-Norm 函数来构建仿射变换信息修正损失函数。可以理解，在本申请实施例中，也可采用其他的函数来构建仿射变换信息修正损失函数，只要该函数可以用来衡量标准仿射变换差异信息和预测仿射变换差异信息之间的差异程度即可，比如 L2-Norm 函数等。

可以理解，该预测仿射变换差异信息用于确定更新的仿射变换信息，并传递至视频帧序列中在后的视频帧。当仿射变换信息为仿射变换参数时，可通过以下公式计算更新的仿射变换信息： $\hat{\theta}'_{t+1} = f(\hat{\theta}'_t, \hat{\theta}_t) = \hat{\theta}'_t \times \hat{\theta}_t + \hat{\theta}_t$; 其中， $\hat{\theta}'_{t+1}$ 表示当前帧所传递的更新的仿射变换信息； $\hat{\theta}'_t$ 表示预测仿射变换差异参数； $\hat{\theta}_t$ 表示预测仿射变换参数。

S820, 根据第一视频帧样本和第二视频帧样本，确定对应的光流信息，并依据光流信息和第一样本特征图，确定光流特征图。

可选地，计算机设备可根据第一视频帧样本和第二视频帧样本，确定对应的光流信息。比如，计算机设备可通过 Lucas-kanade (是一种两帧差分的光流计算方法) 光流方法计算第一视频帧样本所对应的光流信息。进而，计算机设备可依据光流信息和第一样本特征图，计算得到光流特征图。其中，该光流特征图可认为是融合了光流信息的、通过第一视频帧样本所预测的第二视频帧样本对应的特征图。

S822, 将光流特征图和第二样本特征图作为目标分割模型中判别器的样本输入，并通过判别器对样本输入进行分类处理，得到样本输入的预测类别。

可选地，该目标分割网络在模型训练阶段还包括判别器 (Discriminator)。计算机设备可将光流特征图和第二样本特征图作为目标分割模型中判别器的样本输入，输入两种中的任意一种，通过 Discriminator 判断输入的特征是光流特征图还是第二样本特征图。其中，第二样本特征图是第二视频帧样本所对应的样本特征图，也可称作 CNN 特征图。

S824, 依据预测类别及样本输入所对应的参考类别，构建对抗损失函数。

其中，样本输入所对应的参考类别可以是光流特征图和第二样本特征图分别对应的类别，比如光流类别和特征类别。Discriminator 本质是一个二分类网络，计算机设备可使用二分类交叉熵 (cross entropy) 作为 Discriminator 的损失函数，以判断样本输入是否为光流特征图。也就是，根据预测类别及样本输入所对应的参考类别，依据交叉熵函数构建目标分割模型的对抗损失函数。

S826, 依据光流特征图、第二样本特征图、及参考特征图，构建分割损失函数；参考特征图为对第二视频帧样本中的目标进行特征提取所得到的特征图。

可选地，计算机设备可对第二视频帧样本中的目标进行特征提取，得到参考特征图。进而计算机设备可依据光流特征图、第二样本特征图、及参考特征图，构建分割损失函数。

在一个实施例中，计算机设备可通过以下公式构建分割损失函数：

$$loss_{seg_t} = f_{dice}(F_{CNN}, F'_{CNN}) + f_{bce}(F_{CNN}, F'_{CNN}) + f_{mse}(F'_{OF}, F'_{CNN});$$

其中, F'_{CNN} , F'_{OF} 分别代表第二样本特征图和通过光流获取的光流特征图。 F_{CNN} 代表参考特征图。 f_{dice} , f_{bce} , f_{mse} 分别表示 Dice 计算公式, 二分类交叉熵计算公式, 均方差 (mean square error) 计算公式。其中, f_{mse} 越大, 表示第二样本特征图和光流特征图的差距越大, 从而加重惩罚 Generator 完成参数更新, 使得 Generator 产生更加符合光流特征的特征图。 f_{dice} 和 f_{bce} 则是促使 Generator 产生更加贴合人工标注信息的特征图。

S828, 依据仿射损失函数、仿射变换信息修正损失函数、对抗损失函数、及分割损失函数, 调整目标分割模型的模型参数并继续训练, 直至满足训练停止条件时停止训练。

可选地, 计算机设备可依据仿射损失函数、仿射变换信息修正损失函数、对抗损失函数、及分割损失函数, 共同调整目标分割模型中各个网络结构的模型参数并继续训练, 直至满足训练停止条件时停止训练。

在一个实施例中, 目标分割模型在训练时, 可采用交叉训练和共同训练相结合的方式进行训练。比如, 参考图 10, 计算机设备可先训练生成器一段时间后, 固定训练得到的参数, 暂时不再回传。再训练判别器, 之后再固定判别器的参数, 进而再训练生成器, 等训练结果稳定后再结合各个网络结构一起训练。那么此时的训练停止条件, 也可认为是收敛条件, 可以是, 判别器的损失函数不再下降, 判别器的输出稳定在 (0.5, 0.5) 左右, 判别器无法分辨出光流特征图和 CNN 特征图的区别。

可以理解, 当生成器和判别器两者抗衡之后, 整个网络达到收敛状态, 生成器最终将产生 CNN 特征和光流信息共有部分的特征, 而判别器将分不清光流特征和 CNN 特征的区别。在模型的使用阶段, 可移除判别器, 此时生成器将产生融合了光流信息的特征图。

在一个实施例中, 目标分割模型中的各个生成器可共享参数。也就是上, 上述图 9 中的三个生成器可认为是相同的生成器。

上述实施例中, 在模型训练过程中一方面引入仿射变换监督信息, 也就是标准仿射变换信息, 以提高方位预测的准确性; 另一方面可通过对预测仿射变换信息进行纠正训练, 从而减少错误定位带来的分割误差。再者, 采用了带有光流信息的对抗学习方式实现网络在时序上的一致性, 使得训练时针对性更强, 性能更佳。这样, 训练时依据仿射损失函数、仿射变换信息修正损失函数、对抗损失函数、及分割损失函数叠加一起优化, 使得各个部分在训练过程中相互影响, 相互提升, 这样训练得到的目标分割模型可以准确且平滑地从视频中分割出目标对象。

在一个实施例中, 提供了一种模型训练方法。本实施例主要以该方法应用于图 1 中的计算机设备来举例说明, 该模型训练方法包括以下步骤: 获取视频帧样本、视频帧样本对应的样本标注信息、及视频帧样本对应的标准仿射变换信息; 将视频帧样本输入至目标分割模型中进行训练, 通过目标分割模型, 确定与视频帧样本对应的预测仿射变换信息; 依据预测仿射变换信息和标准仿射变换信息构建仿射损失函数; 通过目标分割模型, 输出与视频帧样本对应的预测仿射变换差异信息、及视频帧样本中目标对应的预测分割结果; 根据预测仿射变换信息和标准仿射变换信息间的差异, 确定标准仿射变换差异信息; 依据标准仿射变换差异信息和预测仿射变换差异信息, 构建仿射变换信息修正损失函数; 根据预测分割结果和样本标注信息, 确定分割损失函数; 依据仿射损失函数、仿射变换信息修正损失函数、及分割损失函数, 调整目标分割模型的模型参数并继续训练, 直至满足训练停止条件时停止训练。

关于模型训练方法中各个步骤的详细说明可参考前述实施例中机器翻译模型的模型训练

步骤的说明, 训练方式是一致的, 在此不做重复说明。

在一个实施例中, 以心脏超声检测视频为例, 详细说明该目标分割模型的训练过程。参考图 9, 在训练时, 可将前后两帧视频帧样本作为样本对输入到 RAN 网络中。第一阶段中, 当前帧经过 RAN 网络的仿射变换对目标位置、尺寸以及方位进行了纠正, 得到与模板分布相似的 ROI 图像, 经过纠正的 ROI 图像减少了很多干扰, 如其他心腔与左心室的相似性, 图像标记以及伪影等带来的影响等。第二阶段中, 再次使用 Generator 对 ROI 图像进行特征抽取, 输出的特征进入两个任务分支, 在分割任务分支中, 输出的特征通过两次上采样后得到分割预测图, 输出分割结果; 在定位任务分支中, 特征通过 channel 为 4 的全连接层回归出新的仿射变换差异结果。第二阶段通过回归差值的方式对第一阶段产生的仿射变换信息起二次修正作用。

其中, 当仿射变换信息为仿射变换参数时, 第二阶段的仿射变换差异结果的监督信息可通过下列公式计算: $\hat{\theta}'_t \text{标准} = \frac{\theta_t - \hat{\theta}_t}{\hat{\theta}_t}$, $\theta \in \{Angle, Shift_x, Shift_y, Scale\}$; 其中, $\hat{\theta}'_t$ 标准表示标准仿射变换差异信息; $\hat{\theta}_t$ 表示当前帧所对应的仿射变换参数, 也就是预测仿射变换参数; θ_t 表示标准仿射变换参数。

由于该差值较小, 为了加速网络收敛, 可以使用 L1-Norm 函数算损失值。当前帧在第二阶段预测的仿射变换差异参数将用于计算更新的仿射变换信息并传播到下一帧视频帧中, 下一帧视频帧根据上述参数直接进行仿射变换得到 ROI, 同理, ROI 经过 Generator 提取特征, 再次预测出分割结果和仿射变换差异结果。以第一阶段为基础, 第二阶段进行二次仿射变换信息修正, 如上面公式所示。第二阶段预测出相对于第一阶段的仿射变换信息变化值, 这里同样以 L1-Norm 函数计算损失值。当前帧第二阶段预测的仿射变换差异参数将传播到下一帧中, 下一帧根据下述公式计算得到更新的仿射变换信息: $\hat{\theta}'_{t+1} = f(\hat{\theta}'_t, \hat{\theta}_t) = \hat{\theta}'_t \times \hat{\theta}_t + \hat{\theta}_t$; 其中, $\hat{\theta}'_{t+1}$ 表示当前帧所传递的更新的仿射变换信息; $\hat{\theta}'_t$ 表示预测仿射变换差异参数; $\hat{\theta}_t$ 表示预测仿射变换参数。同理, 下一帧的视频帧所对应的 ROI 经过生成器提取特征, 再次预测出分割结果和仿射变换差异结果。除此之外, 渐进式变化是视频中目标变化的重要特征。在心脏超声检测视频帧中, 左心室会随着时间逐渐扩大或者缩小, 基本不存在突然变化的情况。然而, 由于分割目标边界信息模糊以及伪影的干扰, 尽管加入了时序、方位以及结构等先验信息, 在某些视频帧上仍然会出现由于误分割引起的左心室容积突变。针对这种情况, 在模型训练时可引入光流信息。假定, 左心室中相邻两帧的变化是较为微小的, 下一帧视频帧可以通过上一帧视频帧的光流信息计算得到。在训练时, 对于当前帧应该有两种特征形式: 一种是通过 CNN 网络基于当前帧提取的特征, 另一种是通过光流信息基于上一帧的特征变换而来的特征。为此, 可设计判别器将这两种信息同时引入。如图 9 所示, 判别器 (Discriminator) 的输入有两种: 一种来源于生成器对下一帧 ROI 提取的特征, 一种来源于利用光流信息基于当前帧 ROI 特征变换而来的下一帧 ROI 特征, 输入两种中的任意一种, 判别器判断输入的特征属于光流变换的特征 (Flow Field) 还是 CNN 特征。这样, 引入判别器促使生成器产生具备光流信息和 CNN 本帧信息的分割特征。因此, 分割任务分支可采用如下损失函数: $loss_{seg_t} = f_{dice}(F_{CNN}, F'_{CNN}) + f_{bce}(F_{CNN}, F'_{CNN}) + f_{mse}(F'_{OF}, F'_{CNN})$;

其中, F'_{CNN} , F'_{OF} 分别代表第二样本特征图和通过光流获取的光流特征图。 F_{CNN} 代表参考特征图。 f_{dice} , f_{bce} , f_{mse} 分别表示 Dice 计算公式, 二分类交叉熵计算公式, 均方差计算公式。其中, f_{mse} 越大, 表示第二样本特征图和光流特征图的差距越大, 从而加重惩罚生成器完成参数更新, 使得生成器产生更加符合光流特征的特征图。 f_{dice} 和 f_{bce} 则是促使生成器产生更加贴合人工标注信息的特征图。

此外，对于判别器，使用二分类交叉熵作为损失函数用于判断输入是否为光流特征。两者抗衡之后，当整个网络达到收敛状态，生成器最终将产生 CNN 特征和光流信息共有部分的特征，而判别器将分不清光流特征和 CNN 特征的区别。模型使用时，判别器将被移除，生成器将产生融合了光流信息的特征图

下面结合应用场景，比如心脏早期筛查场景，对心脏超声检测视频中的左心室作为目标，通过该图像分割方法实现对左心室的分割来进行详细说明：

临床中，心脏早期筛查是预防以及诊断心脏疾病的重要措施。鉴于其筛查快速，价格低廉，信息丰富的优势，心脏 B 型超声是目前普遍性较高的早期筛查手段。在心脏超声检测中，临床上常以心动周期超声中左心室在四腔切面和二腔切面的面积，配合 Simpson 法（辛普森法）估量射血分数，作为诊断心功能的一个重要信息来源。而基于计算机辅助的左心室自动分割是计算心功能指标（如射血分数）的重要依据。然而，左心室物体边界模糊，且容易受伪影影像造成边缘缺失，严重影响了分割准确性。同时，左心室的变化和时间强烈相关，预测错误带来的左心室轮廓突变极容易导致临床指标的误计算。同时，超声视频筛查的落地对网络大小、实时性有很大的需求。

考虑到上述困难，本申请实施例中提出了基于 Region Affine Networks 的端到端视频目标分割模型，将在前的视频帧帧的目标结构信息（也就是在前的视频帧所传递的历史仿射变换信息）引入到当前帧，提升了分割性能；同时 Region Affine Networks 是有监督信息的可学习仿射变换信息的预测网络，仿射监督信息的引入使得仿射变化参数预测更加准确。并且，基于二阶段定位网络能够二次纠正在前的视频帧所传递的变换错误，增加网络鲁棒性，减少因为仿射变换信息错误带来的分割误差。同时，基于光流信息的对抗学习网络，在训练时可促使分割结果贴近时序变换渐进性，使得分割结果更加合理。整个网络端到端训练，各个部分相辅相成，相互提高。目标结构信息的引入减少噪声干扰，降低分割难度，使用轻量级的编码网络即可得到优异的分割结果。同时，视频的时序分析、时间平滑处理全部集中在训练阶段，减少了模型在使用过程中的操作处理，大大减少了目标分割的耗时，提高了效率。

本申请实施例所提供的图像分割方法可以用于临床中心脏超声检测配合 Simpson 法筛查心脏疾病，可以解放医师的双手，减少医师标注带来的重复劳动以及主观差异。由于实现该目标分割模型的各个网络结构小、实时性好，端到端网络工程化程度高，极易迁移到移动设备中。

本申请实施例中对心脏超声检测视频中的左心室进行分割所得到的分割结果，可作为临床上心脏 B 型超声结合 Simpson 法测量射血分数的自动化方案；专为视频单物体设计的端到端网络，引入了时序信息、目标的结构位置信息，能得到更加符合视频规律的分割结果；对抗学习网络自适应地增加了视频分割的平滑度，使得分割结果更加合理；该图像分割方法实现了高分割性能的轻量级网络，实时性极强，工程化程度高。

在一个实施例中，如图 11 所示，该图像分割方法包括以下步骤：

S1002，当当前帧为初始视频帧时，获取视频帧序列中的初始视频帧。

S1004，通过第一卷积神经网络提取初始视频帧的图像特征。

S1006，将图像特征输入至第一全连接网络，通过第一全连接网络对图像特征进行处理，通过第一全连接网络的至少一个输出通道输出仿射变换信息。

S1008, 将输出的仿射变换信息作为初始视频帧对应的历史仿射变换信息。

当当前帧不为初始视频帧时, 从缓存中读取在前的视频帧对应的历史仿射变换信息。

S1010, 依据历史仿射变换信息对当前帧进行仿射变换, 得到与当前帧对应的候选区域图像。

S1012, 通过目标分割模型中的第二卷积神经网络, 对候选区域图像进行特征提取, 得到候选区域图像对应的特征图; 特征图融合了视频帧序列所包括的光流信息。

S1014, 通过全卷积神经网络对特征图进行上采样处理, 得到中间图像。

S1016, 通过全卷积神经网络对中间图像中的各像素分别进行像素级分类, 得到各像素所对应的类别。

S1018, 确定中间图像中对应目标类别的像素。

S1020, 从中间图像中, 分割出由对应目标类别的各像素所组成的、且包括目标对象的目标分割区域。

S1022, 通过第二全连接网络对特征图进行处理, 通过第二全连接网络的至少一个输出通道输出仿射变换差异结果。

S1024, 依据仿射变换差异结果和在前的视频帧所传递的历史仿射变换信息, 计算得到当前帧所传递的更新的仿射变换信息。

S1026, 将当前帧所传递的更新的仿射变换信息, 作为视频帧序列中在后的视频帧所对应的历史仿射变换信息。

上述图像分割方法, 依据在前的视频帧所传递的历史仿射变换信息, 对当前帧进行仿射变换, 得到与当前帧对应的候选区域图像。在前的视频帧所传递的历史仿射变换信息是经过修正后的参数, 这样可大大提高候选区域图像获取的准确性。对与候选区域图像对应的特征图进行语义分割, 可以准确得到当前帧中的目标对应的分割结果。并且, 根据该特征图对历史仿射变换信息进行修正, 将修正后的仿射变换信息传递至在后的视频帧, 以供在后的视频帧使用。这样可对当前帧的定位起到纠正作用, 减少了错误定位给后续的分割处理带来误差, 大大提高了对视频进行语义分割处理的准确性。

图 11 为一个实施例中图像分割方法的流程示意图。应该理解的是, 虽然图 11 的流程图中的各个步骤按照箭头的指示依次显示, 但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明, 这些步骤的执行并没有严格的顺序限制, 这些步骤可以以其它的顺序执行。而且, 图 11 中的至少一部分步骤可以包括多个子步骤或者多个阶段, 这些子步骤或者阶段并不必然是在同一时刻执行完成, 而是可以在不同的时刻执行, 这些子步骤或者阶段的执行顺序也不必然是依次进行, 而是可以与其它步骤或者其它步骤的子步骤或者阶段的至少一部分轮流或者交替地执行。

如图 12 所示, 在一个实施例中, 提供了图像分割装置 1100, 包括获取模块 1101、仿射变换模块 1102、特征提取模块 1103、语义分割模块 1104 和参数修正模块 1105。

获取模块 1101, 用于获取视频帧序列中的当前帧、及在前的视频帧的历史仿射变换信息。

仿射变换模块 1102, 用于依据历史仿射变换信息对当前帧进行仿射变换, 得到与当前帧对应的候选区域图像。

特征提取模块 1103, 用于对候选区域图像进行特征提取, 得到候选区域图像对应的特征图。

语义分割模块 1104, 用于基于特征图进行语义分割, 得到当前帧中的目标对应的分割结果。

参数修正模块 1105, 用于根据特征图对历史仿射变换信息进行修正, 得到更新的仿射变换信息, 并将更新的仿射变换信息作为视频帧序列中在后的视频帧所对应的历史仿射变换信息。

在一个实施例中, 当当前帧为初始视频帧时, 获取模块 1101 还用于当当前帧为初始视频帧时, 获取视频帧序列中的初始视频帧; 通过第一卷积神经网络提取初始视频帧的图像特征; 将图像特征输入至包括第一全连接网络, 通过第一全连接网络对图像特征进行处理, 通过第一全连接网络的至少一个输出通道输出仿射变换信息; 将输出的仿射变换信息作为初始视频帧对应的历史仿射变换信息。

在一个实施例中, 对候选区域图像进行特征提取所得到的特征图融合有视频帧序列所包括的光流信息。

在一个实施例中, 语义分割模块 1104 还用于通过全卷积神经网络对特征图进行上采样处理, 得到中间图像; 通过全卷积神经网络对中间图像中的各像素分别进行像素级分类, 得到各像素所对应的类别; 依据各像素所对应的类别, 输出对当前帧中的目标进行语义分割的分割结果。

在一个实施例中, 语义分割模块 1104 还用于确定中间图像中对应目标类别的像素; 从中间图像中, 分割出由对应目标类别的各像素所组成的、且包括目标对象的目标分割区域。

在一个实施例中, 参数修正模块 1105 还用于通过第二全连接网络对特征图进行处理, 通过第二全连接网络的至少一个输出通道输出仿射变换差异结果; 依据仿射变换差异结果和在前的视频帧的历史仿射变换信息, 计算得到当前帧的更新的仿射变换信息; 将当前帧的更新的仿射变换信息, 作为视频帧序列中在后的视频帧所对应的历史仿射变换信息。

在一个实施例中, 特征提取模块 1103 还用于通过目标分割模型中的第二卷积神经网络, 对候选区域图像进行特征提取, 得到候选区域图像对应的特征图。语义分割模块 1104 还用于通过目标分割模型中的全卷积神经网络, 对特征图进行语义分割处理, 得到当前帧中的目标对应的分割结果。参数修正模块 1105 还用于通过目标分割模型中的第二全连接网络对历史仿射变换信息进行修正, 得到更新的仿射变换信息。

如图 13 所示, 在一个实施例中, 该图像分割装置还包括模型训练模块 1106, 用于获取视频帧样本、视频帧样本对应的样本标注信息、及视频帧样本对应的标准仿射变换信息; 将视频帧样本输入至目标分割模型中进行训练, 通过目标分割模型, 获取与视频帧样本对应的预测仿射变换信息; 依据预测仿射变换信息和标准仿射变换信息构建仿射损失函数; 通过目标分割模型, 输出与视频帧样本对应的预测仿射变换差异信息、及视频帧样本中目标对应的预测分割结果; 根据预测仿射变换信息和标准仿射变换信息间的差异, 确定标准仿射变换差异信息; 依据标准仿射变换差异信息和预测仿射变换差异信息, 构建仿射变换信息修正损失函数; 根据预测分割结果和样本标注信息, 确定分割损失函数; 依据仿射损失函数、仿射变换信息修正损失函数、及分割损失函数, 调整目标分割模型的模型参数并继续训练, 直至满足训练停止条件时停止训练。

上述图像分割装置, 依据在前的视频帧的历史仿射变换信息, 对当前帧进行仿射变换, 得到与当前帧对应的候选区域图像。在前的视频帧的历史仿射变换信息是经过修正后的参数, 这样可大大提高候选区域图像获取的准确性。对与候选区域图像对应的特征图进行语义分割,

可以准确得到当前帧中的目标对应的分割结果。并且，根据该特征图对历史仿射变换信息进行修正，将修正后的仿射变换信息传递至在后的视频帧，以供在后的视频帧使用。这样可对当前帧的定位起到纠正作用，减少了错误定位给后续的分割处理所带来误差，大大提高了对视频进行语义分割处理的准确性。

如图 14 所示，在一个实施例中，提供了模型训练装置 1300，包括样本获取模块 1301、确定模块 1302、构建模块 1303、输出模块 1304 和模型参数调整模块 1305。

样本获取模块 1301，用于获取视频帧样本、视频帧样本对应的样本标注信息、及视频帧样本对应的标准仿射变换信息。

确定模块 1302，用于将视频帧样本输入至目标分割模型中进行训练，通过目标分割模型，确定与视频帧样本对应的预测仿射变换信息。

构建模块 1303，用于依据预测仿射变换信息和标准仿射变换信息构建仿射损失函数。

输出模块 1304，用于通过目标分割模型，输出与视频帧样本对应的预测仿射变换差异信息、及视频帧样本中目标对应的预测分割结果。

确定模块 1302 还用于根据预测仿射变换信息和标准仿射变换信息间的差异，确定标准仿射变换差异信息。

构建模块 1303 还用于依据标准仿射变换差异信息和预测仿射变换差异信息，构建仿射变换信息修正损失函数。

构建模块 1303 还用于根据预测分割结果和样本标注信息，确定分割损失函数。

模型参数调整模块 1305，用于依据仿射损失函数、仿射变换信息修正损失函数、及分割损失函数，调整目标分割模型的模型参数并继续训练，直至满足训练停止条件时停止训练。

在一个实施例中，样本获取模块 1301 还用于获取视频帧样本和相应的样本标注信息；样本标注信息包括样本关键点位置信息和样本区域位置信息；根据视频帧样本、样本关键点位置信息和样本区域位置信息，确定模板图像及模板图像对应的模板关键点位置信息；根据样本关键点位置信息和模板关键点位置信息，计算得到与视频帧样本对应的标准仿射变换信息。

在一个实施例中，样本获取模块 1301 还用于获取第一视频帧样本和第二视频帧样本；第一视频帧样本为第二视频帧样本在前的视频帧；分别获取与第一视频帧样本及第二视频帧样本各自对应的样本标注信息、及与第一视频帧样本对应的标准仿射变换信息。确定模块 1302 还用于将第一视频帧样本和第二视频帧样本作为样本对输入至目标分割模型中进行训练，通过目标分割模型对第一视频帧样本进行处理，得到与第一视频帧样本对应的预测仿射变换信息。输出模块 1304 还用于依据预测仿射变换信息对第一视频帧样本进行仿射变换，得到第一样本候选区域图像，并对第一样本候选区域图像进行特征提取，得到第一样本特征图；基于第一样本特征图进行语义分割，得到第一视频帧样本中的目标对应的预测分割结果；根据第一样本特征图对预测仿射变换信息进行修正，得到与第一视频帧样本对应的预测仿射变换差异信息。该模型训练装置还包括对抗模块 1306，用于根据第一视频帧样本和第二视频帧样本，确定对应的光流信息，并依据光流信息和第一样本特征图，确定光流特征图；将光流特征图和第二样本特征图作为目标分割模型中判别器的样本输入，通过判别器对样本输入进行分类处理，得到样本输入的预测类别。构建模块 1303 还用于依据预测类别及样本输入所对应的参考类别，构建对抗损失函数；依据光流特征图、第二样本特征图、及参考特征图，构建分割

损失函数；参考特征图为对第二视频帧样本中的目标进行特征提取所得到的特征图。模型参数调整模块 1305 还用于依据仿射损失函数、仿射变换信息修正损失函数、对抗损失函数、及分割损失函数，调整目标分割模型的模型参数并继续训练，直至满足训练停止条件时停止训练。

上述模型训练装置，在模型训练过程中一方面引入仿射变换监督信息，也就是标准仿射变换信息，以提高方位预测的准确性；另一方面可通过对预测仿射变换信息进行纠正训练，从而减少错误定位带来的分割误差。训练时将仿射损失函数、仿射变换信息修正损失函数、及分割损失函数叠加一起优化，使得各个部分在训练过程中相互影响，相互提升，这样训练得到的目标分割模型具有准确的视频语义分割性能。

图 15 示出了一个实施例中计算机设备的内部结构图。该计算机设备可以是图 1 中的计算机设备。如图 15 所示，该计算机设备包括该计算机设备包括通过系统总线连接的处理器、存储器、和网络接口。其中，存储器包括非易失性存储介质和内存存储器。该计算机设备的非易失性存储介质存储有操作系统，还可存储有计算机程序，该计算机程序被处理器执行时，可使得处理器实现图像分割方法和/或模型训练方法。该内存存储器中也可储存有计算机程序，该计算机程序被处理器执行时，可使得处理器执行图像分割方法和/或模型训练方法。

本领域技术人员可以理解，图 15 中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的计算机设备的限定，的计算机设备可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

在一个实施例中，本申请提供的图像分割装置和或/模型训练装置可以实现为一种计算机程序的形式，计算机程序可在如图 15 所示的计算机设备上运行。计算机设备的存储器中可存储组成该图像分割装置的各个程序模块，比如，图 12 所示的获取模块、仿射变换模块、特征提取模块、语义分割模块和参数修正模块。各个程序模块构成的计算机程序使得处理器执行本说明书中描述的本申请各个实施例的图像分割方法中的步骤。还比如，图 14 所示的样本获取模块、确定模块、构建模块、输出模块和模型参数调整模块。各个程序模块构成的计算机程序使得处理器执行本说明书中描述的本申请各个实施例的模型训练方法中的步骤。

例如，图 15 所示的计算机设备可以通过如图 12 所示的图像分割装置中的获取模块执行步骤 S202。计算机设备可通过仿射变换模块执行步骤 S204。计算机设备可通过特征提取模块执行步骤 S206。计算机设备可通过语义分割模块执行步骤 S208。计算机设备可通过参数修正模块执行步骤 S210。

在一个实施例中，提供了一种计算机设备，包括存储器和处理器，存储器存储有计算机程序，计算机程序被处理器执行时，使得处理器执行上述图像分割方法和/或模型训练方法的步骤。此处图像分割方法和/或模型训练方法的步骤可以是上述各个实施例的图像分割方法和/或模型训练方法中的步骤。

在一个实施例中，提供了一种计算机可读存储介质，存储有计算机程序，计算机程序被处理器执行时，使得处理器执行上述图像分割方法和/或模型训练方法的步骤。此处图像分割方法和/或模型训练方法的步骤可以是上述各个实施例的图像分割方法和/或模型训练方法中的步骤。

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机程序来指令相关的硬件来完成，所述的程序可存储于一非易失性计算机可读取存储介质中，该程序在执行时，可包括如上述各方法的实施例的流程。其中，本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用，均可包括非易失性和/或易失性存储器。非易失性存储器可包括只读存储器（ROM）、可编程 ROM（PROM）、电可编程 ROM（EPROM）、电可擦除可编程 ROM（EEPROM）或闪存。易失性存储器可包括随机存取存储器（RAM）或者外部高速缓冲存储器。作为说明而非局限，RAM 以多种形式可得，诸如静态 RAM（SRAM）、动态 RAM（DRAM）、同步 DRAM（SDRAM）、双数据率 SDRAM（DDRSDRAM）、增强型 SDRAM（ESDRAM）、同步链路（Synchlink）DRAM（SLDRAM）、存储器总线（Rambus）直接 RAM（RDRAM）、直接存储器总线动态 RAM（DRDRAM）、以及存储器总线动态 RAM（RDRAM）等。

以上实施例的各技术特征可以进行任意的组合，为使描述简洁，未对上述实施例中的各个技术特征所有可能的组合都进行描述，然而，只要这些技术特征的组合不存在矛盾，都应当认为是本说明书记载的范围。

以上所述实施例仅表达了本申请的几种实施方式，其描述较为具体和详细，但并不能因此而理解为对本申请专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本申请构思的前提下，还可以做出若干变形和改进，这些都属于本申请的保护范围。因此，本申请专利的保护范围应以所附权利要求为准。

权利要求书

1、一种图像分割方法，应用于计算机设备中，所述方法包括：

获取视频帧序列中的当前帧、及在前的视频帧的历史仿射变换信息；

依据所述历史仿射变换信息对所述当前帧进行仿射变换，得到与所述当前帧对应的候选区域图像；

对所述候选区域图像进行特征提取，得到所述候选区域图像对应的特征图；

基于所述特征图进行语义分割，得到所述当前帧中的目标对应的分割结果；

根据所述特征图对所述历史仿射变换信息进行修正，得到更新的仿射变换信息，将所述更新的仿射变换信息作为所述视频帧序列中在后的视频帧所对应的历史仿射变换信息。

2、根据权利要求 1 所述的方法，其特征在于，所述根据所述特征图对所述历史仿射变换信息进行修正，得到更新的仿射变换信息，将所述更新的仿射变换信息作为所述视频帧序列中在后的视频帧所对应的历史仿射变换信息，包括：

通过第二全连接网络对所述特征图进行处理，通过所述第二全连接网络的至少一个输出通道输出仿射变换差异结果；

依据所述仿射变换差异结果和所述在前的视频帧的历史仿射变换信息，计算得到更新的仿射变换信息；

将所述更新的仿射变换信息，作为所述视频帧序列中在后的视频帧所对应的历史仿射变换信息。

3、根据权利要求 1 所述的方法，其特征在于，所述特征图融合有所述视频帧序列所包括的光流信息。

4、根据权利要求 1 所述的方法，其特征在于，所述获取视频帧序列中的当前帧、及在前的视频帧的历史仿射变换信息，包括：

当所述当前帧为初始视频帧时，获取所述视频帧序列中的所述初始视频帧；

通过第一卷积神经网络提取所述初始视频帧的图像特征；

将所述图像特征输入至第一全连接网络，通过所述第一全连接网络对所述图像特征进行处理，通过所述第一全连接网络的至少一个输出通道输出的仿射变换信息；

将输出的所述仿射变换信息作为所述初始视频帧对应的历史仿射变换信息。

5、根据权利要求 1 所述的方法，其特征在于，所述基于所述特征图进行语义分割，得到所述当前帧中的目标对应的分割结果，包括：

通过全卷积神经网络对所述特征图进行上采样处理，得到中间图像；

通过所述全卷积神经网络对所述中间图像中的各个像素分别进行像素级分类，得到所述各个像素所对应的类别；

依据所述各个像素所对应的类别，输出对所述当前帧中的目标进行语义分割的分割结果。

6、根据权利要求 5 所述的方法，其特征在于，所述依据所述各个像素所对应的类别，输出对所述当前帧中的目标进行语义分割的分割结果，包括：

确定所述中间图像中属于目标类别的像素；

从所述中间图像中，分割出由属于所述目标类别的各个所述像素所组成的、且包括所述目标的目标分割区域。

7、根据权利要求1所述的方法，其特征在于，所述方法通过目标分割模型执行；

所述对所述候选区域图像进行特征提取，得到所述候选区域图像对应的特征图，包括：

通过所述目标分割模型中的第二卷积神经网络，对所述候选区域图像进行特征提取，得到所述候选区域图像对应的特征图；

所述基于所述特征图进行语义分割，得到所述当前帧中的目标对应的分割结果，包括：

通过所述目标分割模型中的全卷积神经网络，对所述特征图进行语义分割处理，得到所述当前帧中的目标对应的分割结果；

所述根据所述特征图对所述历史仿射变换信息进行修正，得到更新的仿射变换信息，包括：

通过所述目标分割模型中的第二全连接网络对所述历史仿射变换信息进行修正，得到更新的仿射变换信息。

8、根据权利要求1至7中任一项所述的方法，其特征在于，所述方法通过目标分割模型执行，所述目标分割模型的训练步骤包括：

获取视频帧样本、所述视频帧样本对应的样本标注信息、及所述视频帧样本对应的标准仿射变换信息；

将所述视频帧样本输入至所述目标分割模型中进行训练，通过所述目标分割模型，确定与所述视频帧样本对应的预测仿射变换信息；

依据所述预测仿射变换信息和所述标准仿射变换信息构建仿射损失函数；

通过所述目标分割模型，输出与所述视频帧样本对应的预测仿射变换差异信息、及所述视频帧样本中目标对应的预测分割结果；

根据所述预测仿射变换信息和所述标准仿射变换信息间的差异，确定标准仿射变换差异信息；

依据所述标准仿射变换差异信息和所述预测仿射变换差异信息，构建仿射变换信息修正损失函数；

根据所述预测分割结果和所述样本标注信息，确定分割损失函数；

依据所述仿射损失函数、所述仿射变换信息修正损失函数、及所述分割损失函数，调整所述目标分割模型的模型参数并继续训练，直至满足训练停止条件时停止训练。

9、一种模型训练方法，应用于计算机设备中，所述方法包括：

获取视频帧样本、所述视频帧样本对应的样本标注信息、及所述视频帧样本对应的标准仿射变换信息；

将所述视频帧样本输入至目标分割模型中进行训练，通过所述目标分割模型，确定与所述视频帧样本对应的预测仿射变换信息；

依据所述预测仿射变换信息和所述标准仿射变换信息构建仿射损失函数；

通过所述目标分割模型，输出与所述视频帧样本对应的预测仿射变换差异信息、及所述视频帧样本中目标对应的预测分割结果；

根据所述预测仿射变换信息和所述标准仿射变换信息间的差异，确定标准仿射变换差异信息；

依据所述标准仿射变换差异信息和所述预测仿射变换差异信息，构建仿射变换信息修正损失函数；

根据所述预测分割结果和所述样本标注信息，确定分割损失函数；

依据所述仿射损失函数、所述仿射变换信息修正损失函数、及所述分割损失函数，调整所述目标分割模型的模型参数并继续训练，直至满足训练停止条件时停止训练。

10、根据权利要求9所述的方法，其特征在于，所述获取视频帧样本、所述视频帧样本对应的样本标注信息、及所述视频帧样本对应的标准仿射变换信息，包括：

获取视频帧样本和相应的样本标注信息；所述样本标注信息包括样本关键点位置信息和样本区域位置信息；

根据所述视频帧样本、所述样本关键点位置信息和所述样本区域位置信息，确定模板图像及所述模板图像对应的模板关键点位置信息；

根据所述样本关键点位置信息和所述模板关键点位置信息，计算得到与所述视频帧样本对应的标准仿射变换信息。

11、根据权利要求9或10所述的方法，其特征在于，所述获取视频帧样本、所述视频帧样本对应的样本标注信息、及所述视频帧样本对应的标准仿射变换信息，包括：

获取第一视频帧样本和第二视频帧样本；所述第一视频帧样本为所述第二视频帧样本在前的视频帧；

分别获取与所述第一视频帧样本及所述第二视频帧样本各自对应的样本标注信息、及与所述第一视频帧样本对应的标准仿射变换信息；

所述将所述视频帧样本输入至所述目标分割模型中进行训练，通过所述目标分割模型，确定与所述视频帧样本对应的预测仿射变换信息，包括：

将所述第一视频帧样本和所述第二视频帧样本作为样本对输入至目标分割模型中进行训练，通过所述目标分割模型对所述第一视频帧样本进行处理，得到与所述第一视频帧样本对应的预测仿射变换信息；

所述通过所述目标分割模型，输出与所述视频帧样本对应的预测仿射变换差异信息、及所述视频帧样本中目标对应的预测分割结果，包括：

依据所述预测仿射变换信息对所述第一视频帧样本进行仿射变换，得到第一样本候选区域图像，并对所述第一样本候选区域图像进行特征提取，得到第一样本特征图；

基于所述第一样本特征图进行语义分割，得到所述第一视频帧样本中的目标对应的预测分割结果；

根据所述第一样本特征图对所述预测仿射变换信息进行修正，得到与所述第一视频帧样本对应的预测仿射变换差异信息。

12、根据权利要求11所述的方法，其特征在于，所述方法还包括：

根据所述第一视频帧样本和所述第二视频帧样本，确定对应的光流信息，并依据所述光流信息和所述第一样本特征图，确定光流特征图；

将所述光流特征图和所述第二样本特征图作为所述目标分割模型中判别器的样本输入，并通过所述判别器对所述样本输入进行分类处理，得到所述样本输入的预测类别；

依据所述预测类别及所述样本输入所对应的参考类别，构建对抗损失函数；

所述根据所述预测分割结果和所述样本标注信息，确定分割损失函数包括：

依据所述光流特征图、所述第二样本特征图、及参考特征图，构建分割损失函数；所述参考特征图为对所述第二视频帧样本中的目标进行特征提取所得到的特征图；

所述依据所述仿射损失函数、所述仿射变换信息修正损失函数、及所述分割损失函数，调整所述目标分割模型的模型参数并继续训练，直至满足训练停止条件时停止训练，包括：

依据所述仿射损失函数、所述仿射变换信息修正损失函数、所述对抗损失函数、及所述分割损失函数，调整所述目标分割模型的模型参数并继续训练，直至满足训练停止条件时停止训练。

13、一种图像分割装置，所述装置包括：

获取模块，用于获取视频帧序列中的当前帧、及在前的视频帧的历史仿射变换信息；

仿射变换模块，用于依据所述历史仿射变换信息对所述当前帧进行仿射变换，得到与所述当前帧对应的候选区域图像；

特征提取模块，用于对所述候选区域图像进行特征提取，得到所述候选区域图像对应的特征图；

语义分割模块，用于基于所述特征图进行语义分割，得到所述当前帧中的目标对应的分割结果；

参数修正模块，用于根据所述特征图对所述历史仿射变换信息进行修正，得到更新的仿射变换信息，将所述更新的仿射变换信息作为所述视频帧序列中在后的视频帧所对应的历史仿射变换信息。

14、根据权利要求 13 所述的装置，其特征在于，所述参数修正模块，还用于通过第二全连接网络对特征图进行处理，通过所述第二全连接网络的至少一个输出通道输出仿射变换差异结果；依据仿射变换差异结果和所述在前的视频帧的历史仿射变换信息，计算得到所述当前帧的更新的仿射变换信息；将所述更新的仿射变换信息，作为所述视频帧序列中在后的视频帧所对应的历史仿射变换信息。

15、根据权利要求 13 所述的装置，其特征在于，所述特征图融合有所述视频帧序列所包括的光流信息。

16、一种模型训练装置，所述装置包括：

样本获取模块，用于获取视频帧样本、所述视频帧样本对应的样本标注信息、及所述视频帧样本对应的标准仿射变换信息；

确定模块，用于将所述视频帧样本输入至目标分割模型中进行训练，通过所述目标分割模型，确定与所述视频帧样本对应的预测仿射变换信息；

构建模块，用于依据所述预测仿射变换信息和所述标准仿射变换信息构建仿射损失函数；

输出模块，用于通过所述目标分割模型，输出与所述视频帧样本对应的预测仿射变换差异信息、及所述视频帧样本中目标对应的预测分割结果；

所述确定模块，还用于根据所述预测仿射变换信息和所述标准仿射变换信息间的差异，确定标准仿射变换差异信息；

所述构建模块，还用于依据所述标准仿射变换差异信息和所述预测仿射变换差异信息，构建仿射变换信息修正损失函数；

所述构建模块，还用于根据所述预测分割结果和所述样本标注信息，确定分割损失函数；

模型参数调整模块，用于依据所述仿射损失函数、所述仿射变换信息修正损失函数、及所述分割损失函数，调整所述目标分割模型的模型参数并继续训练，直至满足训练停止条件时停止训练。

17、根据权利要求 16 所述的装置，其特征在于，

所述样本获取模块，还用于获取第一视频帧样本和第二视频帧样本；所述第一视频帧样本为所述第二视频帧样本在前的视频帧；分别获取与所述第一视频帧样本及所述第二视频帧样本各自对应的样本标注信息、及与所述第一视频帧样本对应的标准仿射变换信息；

所述确定模块，还用于将所述第一视频帧样本和所述第二视频帧样本作为样本对输入至所述目标分割模型中进行训练，通过所述目标分割模型对第一视频帧样本进行处理，得到与所述第一视频帧样本对应的预测仿射变换信息；

所述输出模块，还用于依据所述预测仿射变换信息对所述第一视频帧样本进行仿射变换，得到第一样本候选区域图像，并对所述第一样本候选区域图像进行特征提取，得到第一样本特征图；基于所述第一样本特征图进行语义分割，得到所述第一视频帧样本中的目标对应的预测分割结果；根据所述第一样本特征图对预测仿射变换信息进行修正，得到与所述第一视频帧样本对应的预测仿射变换差异信息。

18、根据权利要求 16 所述的装置，其特征在于，所述装置还包括：

对抗模块，用于根据所述第一视频帧样本和所述第二视频帧样本，确定对应的光流信息，并依据所述光流信息和所述第一样本特征图，确定光流特征图；将所述光流特征图和所述第二样本特征图作为所述目标分割模型中判别器的样本输入，并通过所述判别器对样本输入进行分类处理，得到所述样本输入的预测类别；

所述构建模块，还用于依据所述预测类别及所述样本输入所对应的参考类别，构建对抗损失函数；依据所述光流特征图、所述第二样本特征图、及所述参考特征图，构建分割损失函数；所述参考特征图为对所述第二视频帧样本中的目标进行特征提取所得到的特征图；

所述模型参数调整模块还用于依据所述仿射损失函数、所述仿射变换信息修正损失函数、所述对抗损失函数、及所述分割损失函数，调整所述目标分割模型的模型参数并继续训练，直至满足训练停止条件时停止训练。

19、一种计算机可读存储介质，存储有计算机程序，所述计算机程序被处理器执行时，使得所述处理器执行如权利要求 1 至 12 中任一项所述方法的步骤。

20、一种计算机设备，包括存储器和处理器，所述存储器存储有计算机程序，所述计算机程序被所述处理器执行时，使得所述处理器执行如权利要求 1 至 12 中任一项所述方法的步骤。

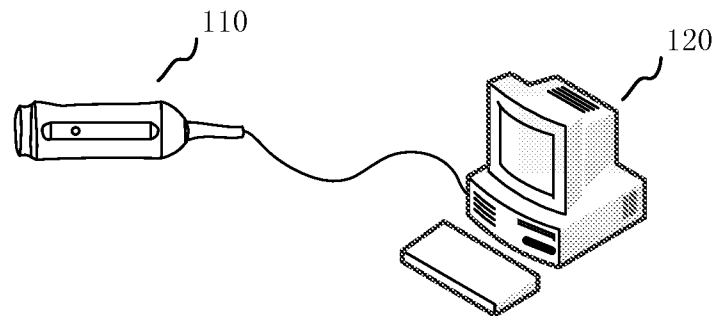


图 1

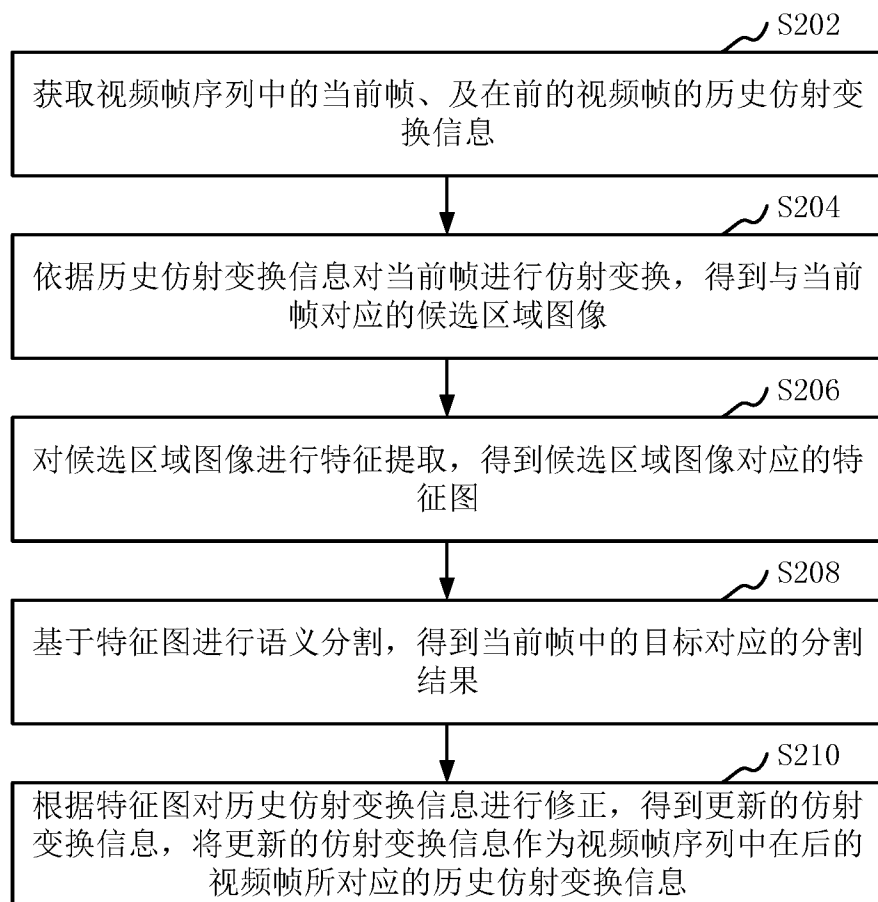


图 2

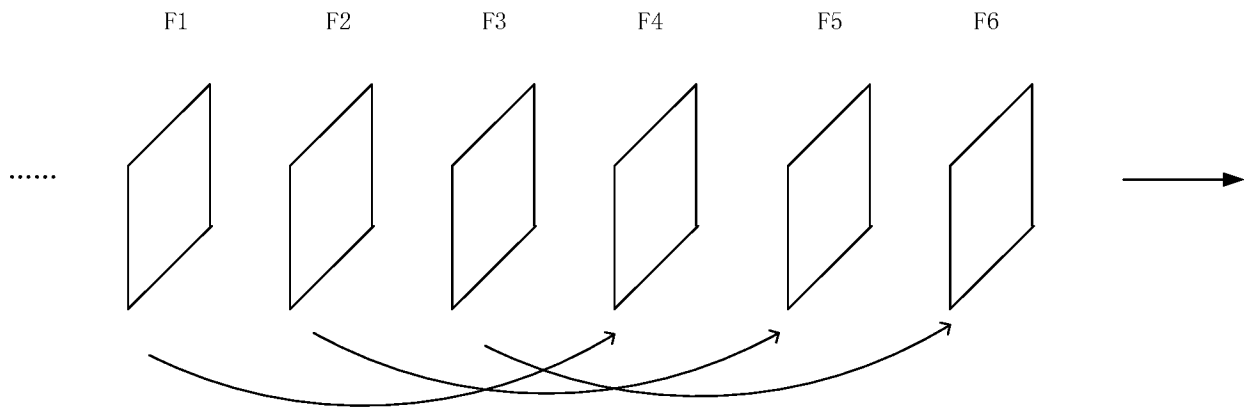


图 3

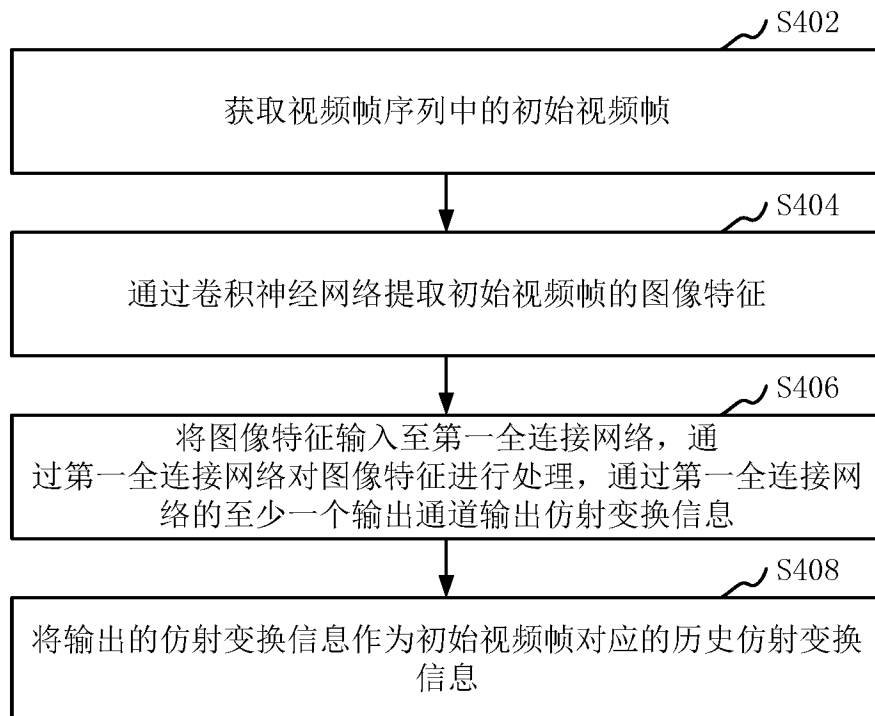


图 4

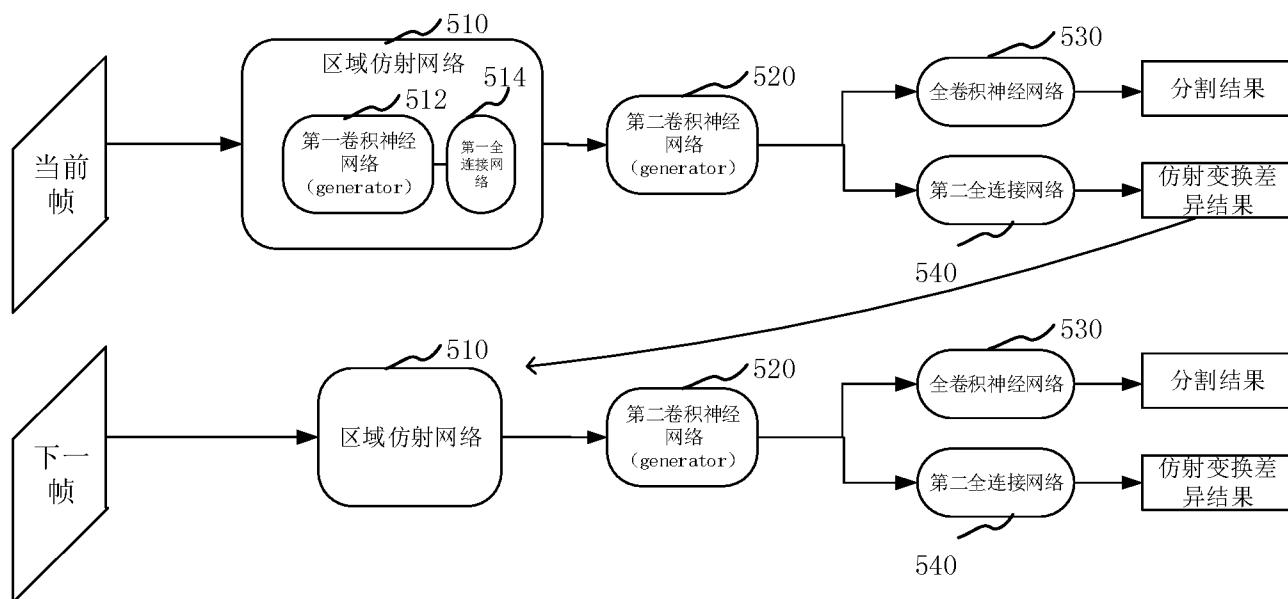


图 5

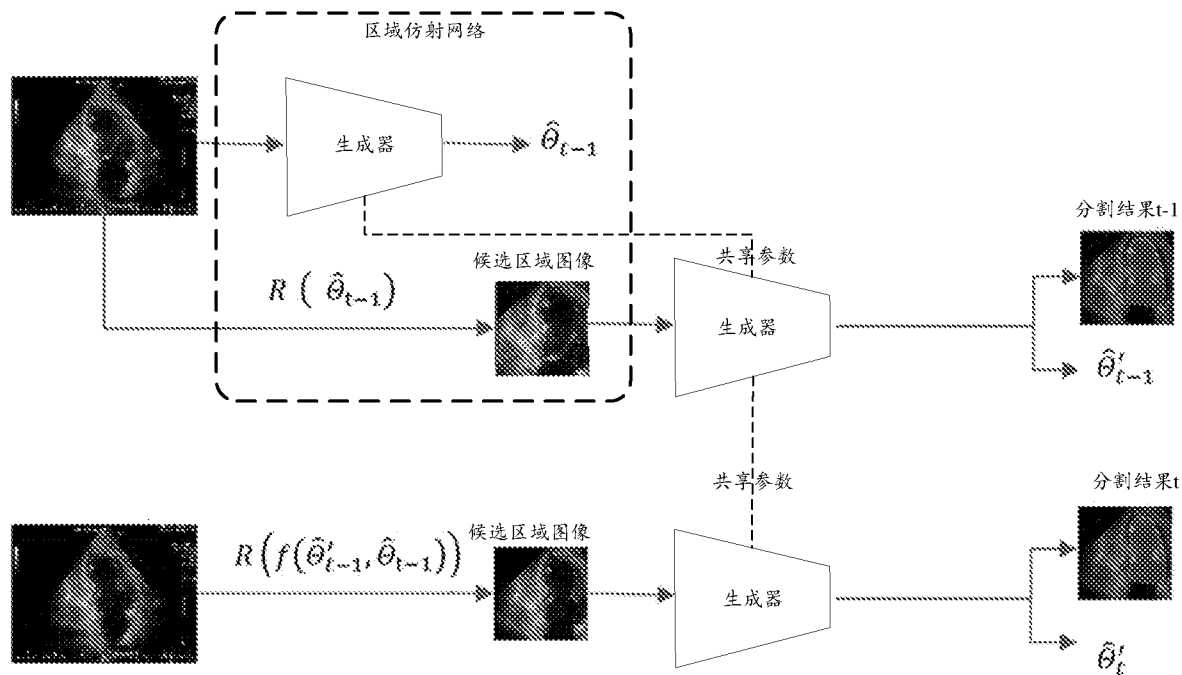


图 6

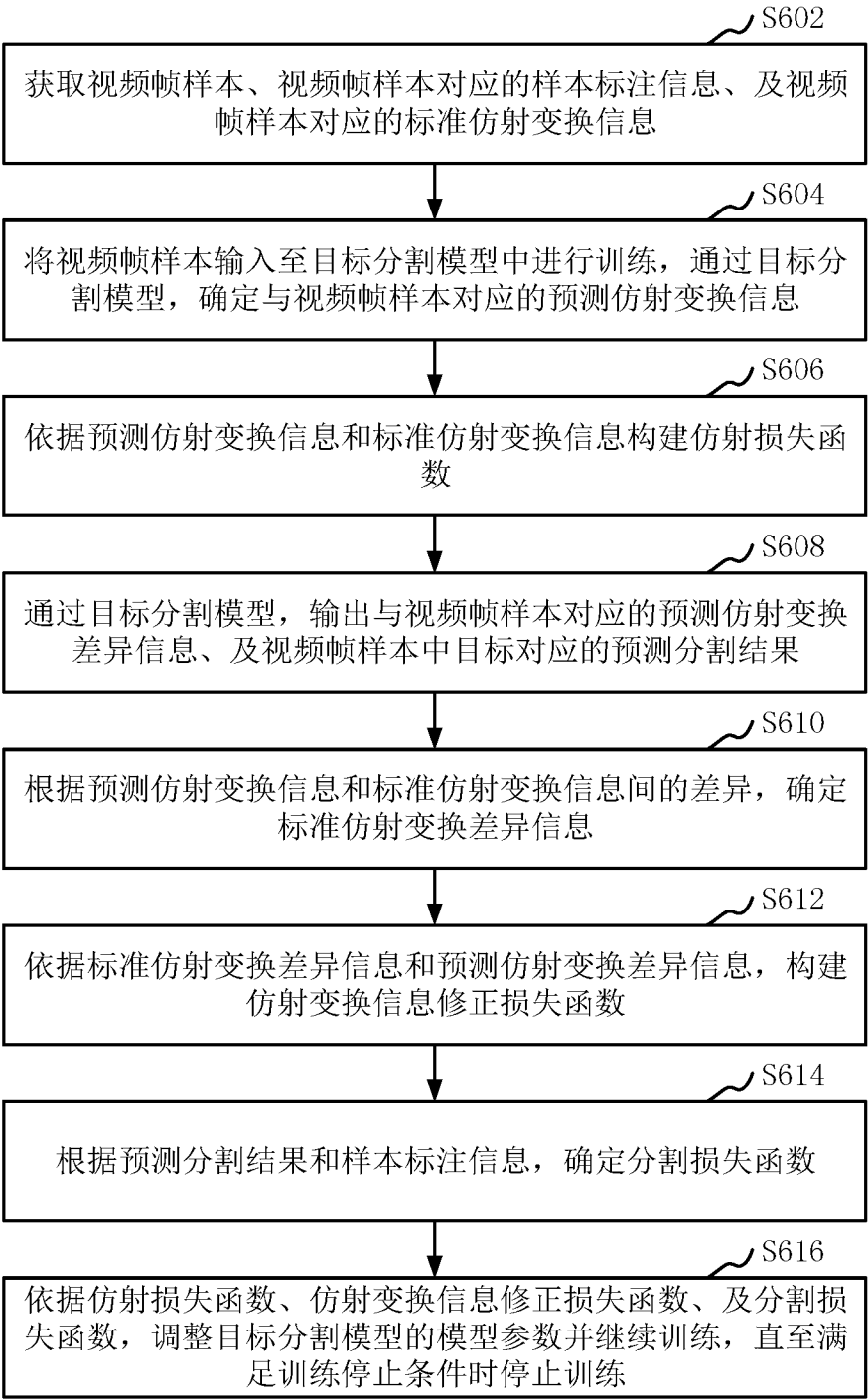


图 7

带有三个关键点标记的原始图片

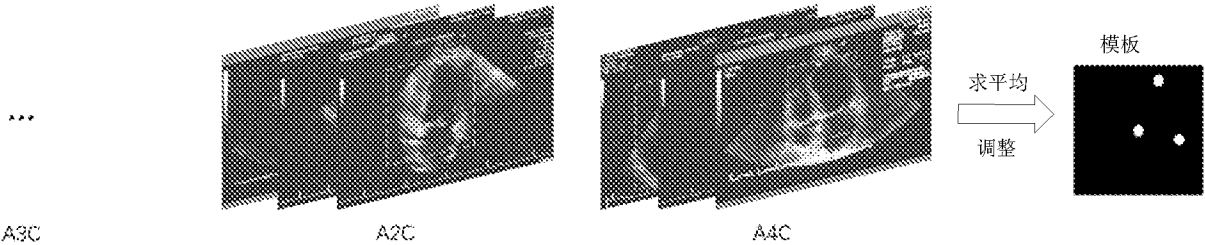


图 8

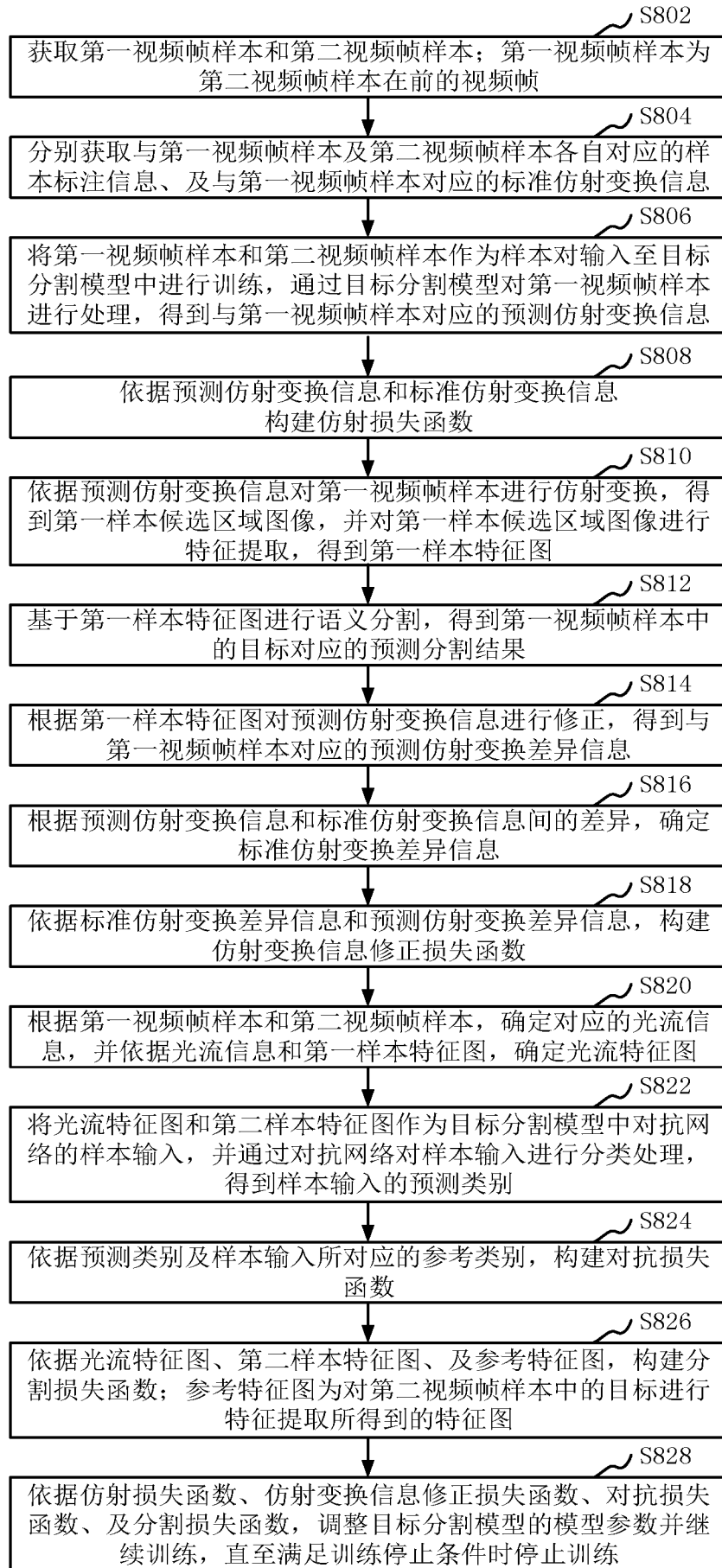


图 9

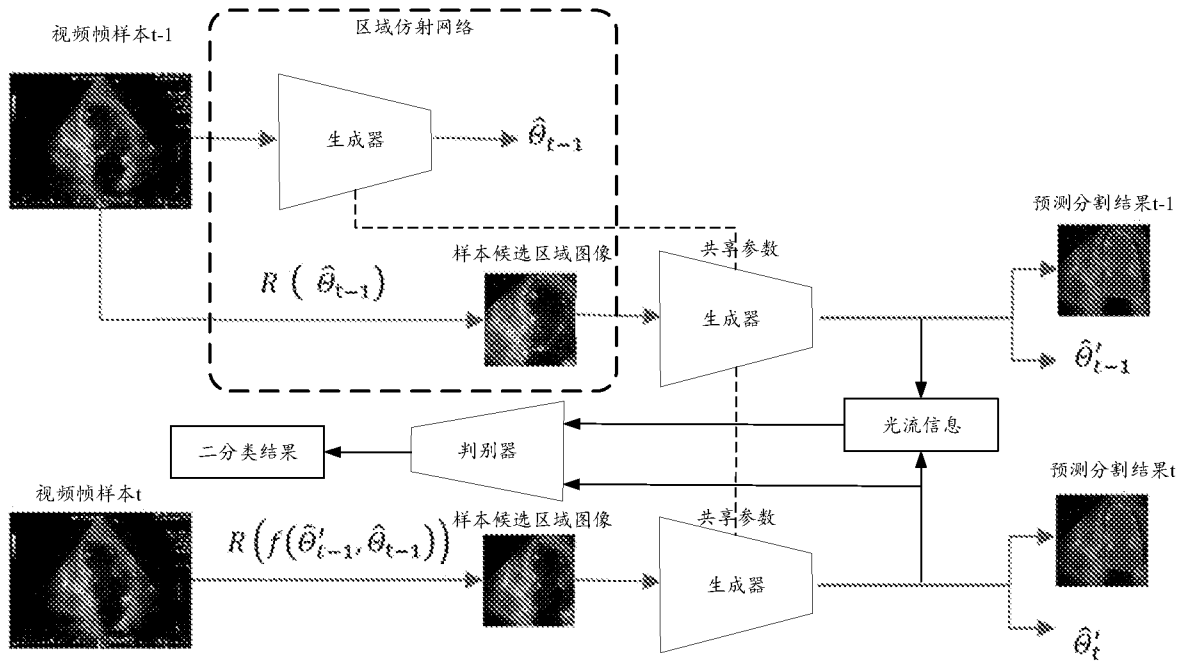


图 10

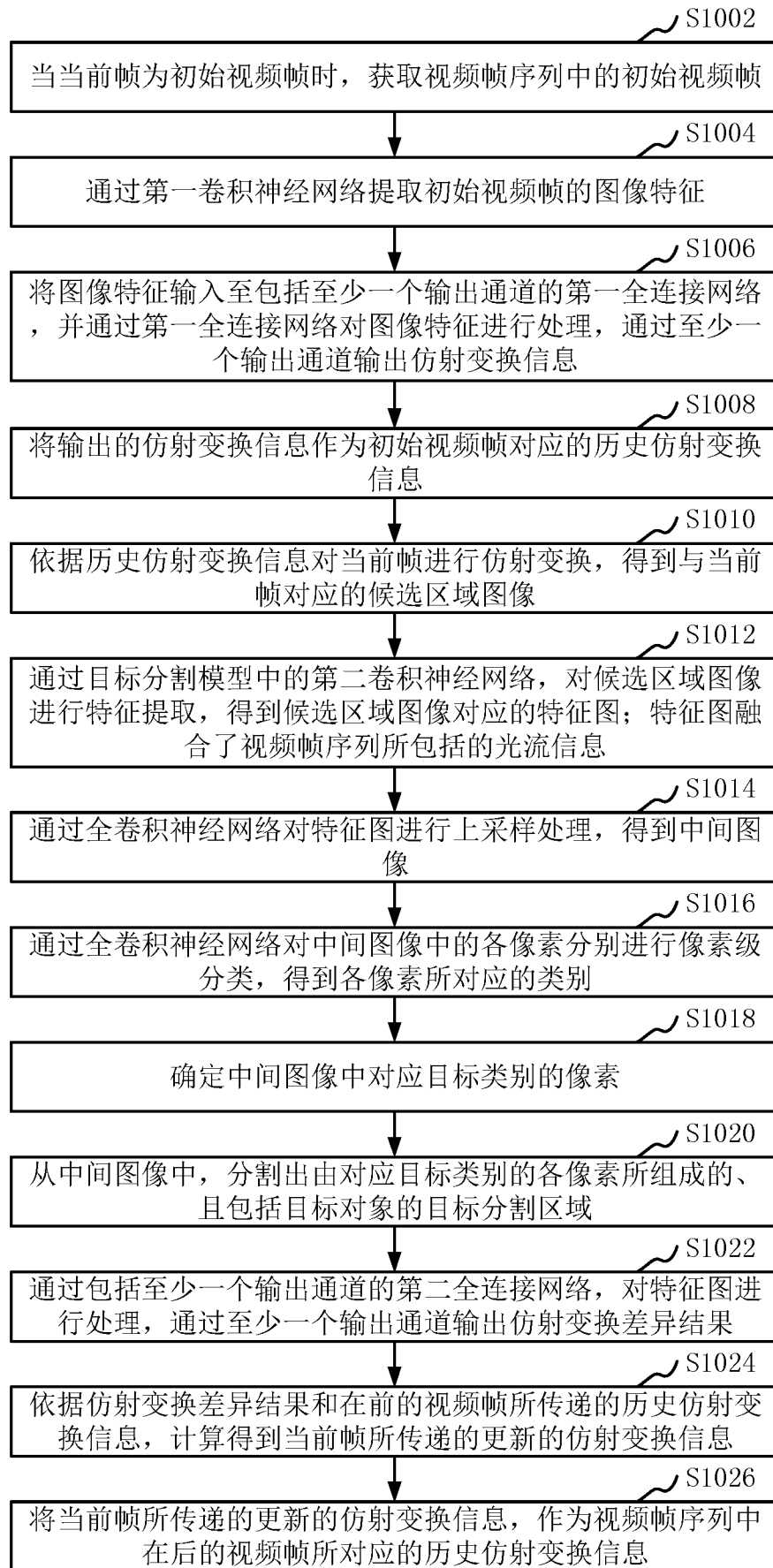


图 11

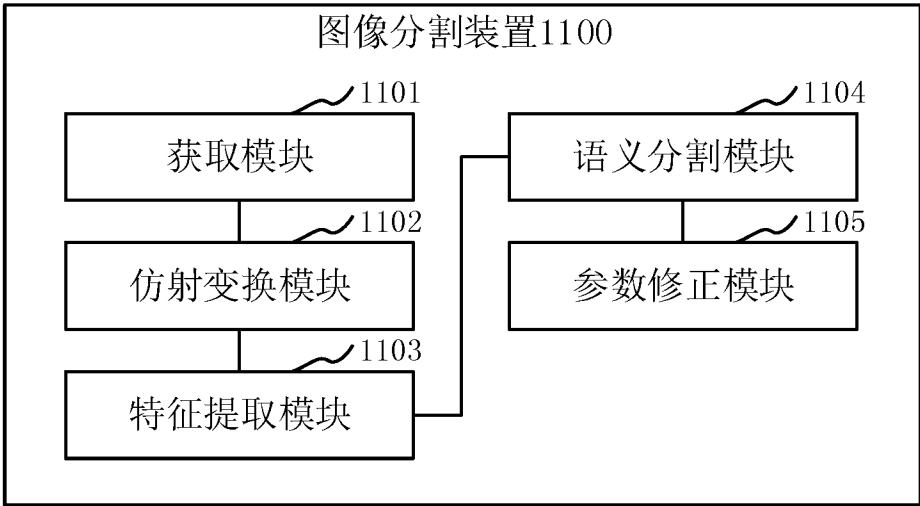


图 12

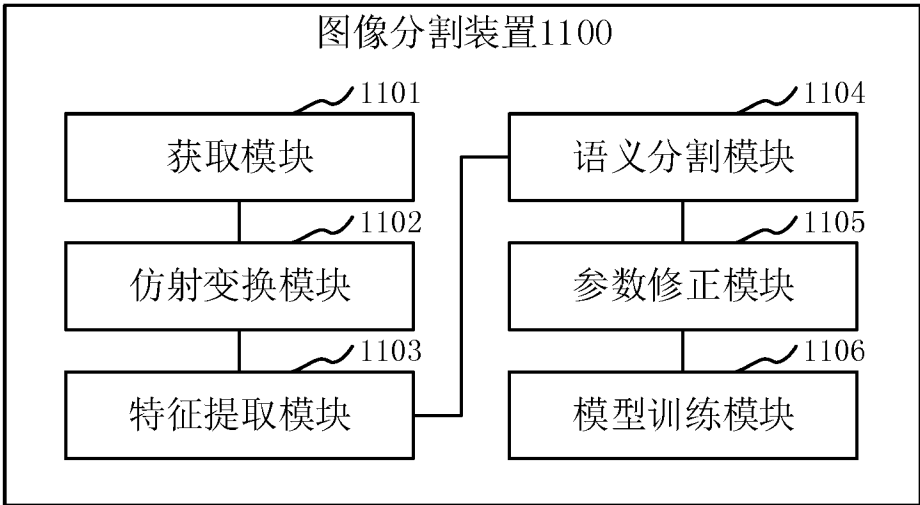


图 13

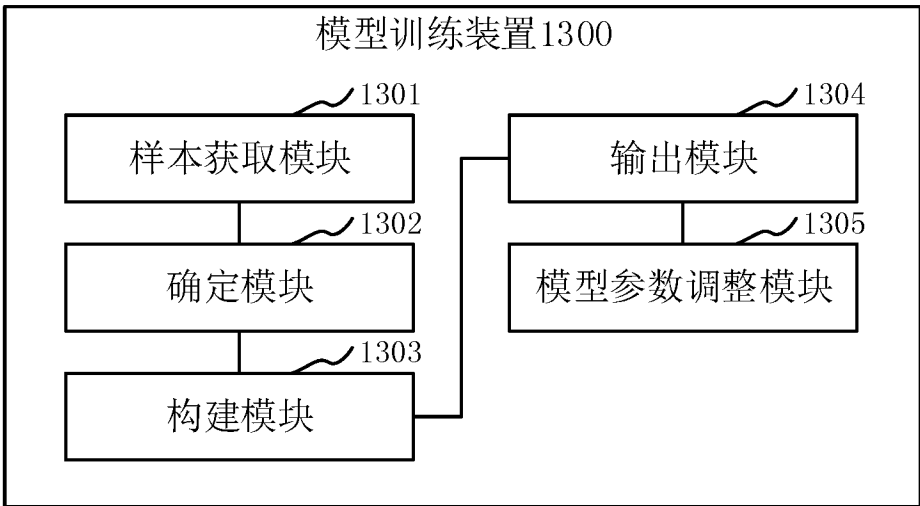


图 14

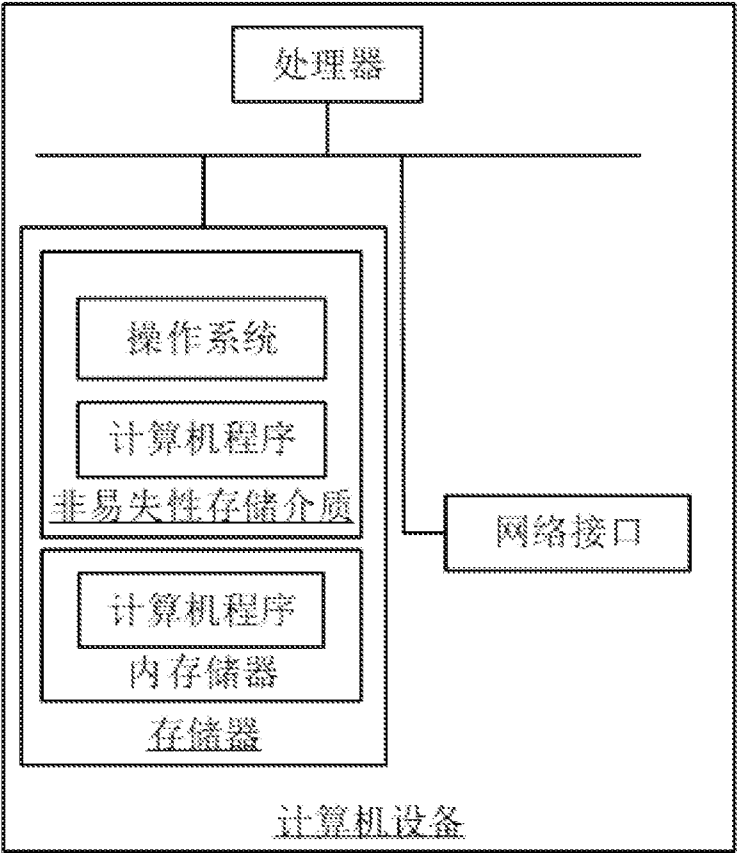


图 15

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/092356

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/32(2006.01)i; G06K 9/34(2006.01)i; G06K 9/62(2006.01)i; G06T 7/00(2017.01)i; G06T 7/11(2017.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06K G06T

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS; CNTXT; CNKI; SIPOABS; DWPI; USTXT; WOTXT; EPTXT: 视频, 图像, 帧, 分割, 划分, 仿射, 投影, 拉伸, 位移, 缩放, 动态, 更新, 校准, 校正, 背景, video, image, frame, dynamic, project, affine, segment, divide, partition, update, calibration, adjust, revise

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 110188754 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 30 August 2019 (2019-08-30) claims 1-15, description, paragraphs [0082]-[0239], figures 1-14	1-20
X	CN 101482923 A (LIU, Yun) 15 July 2009 (2009-07-15) description page 4 line 8 to page 16 line 3	1-8, 13-15, 19, 20
X	CN 102216957 A (ISIS INNOVATION LTD.) 12 October 2011 (2011-10-12) description, paragraphs [0072]-[0276]	1-8, 13-15, 19, 20
X	CN 108122234 A (BEIJING SENSETIME TECHNOLOGY DEVELOPMENT CO., LTD.) 05 June 2018 (2018-06-05) description, paragraphs [0051]-[0219]	9-12, 16-18, 19, 20
A	CN 107146239 A (WUHAN UNIVERSITY) 08 September 2017 (2017-09-08) entire document	1-20
A	CN 108492297 A (CHONGQING UNIVERSITY OF TECHNOLOGY) 04 September 2018 (2018-09-04) entire document	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

03 July 2020

Date of mailing of the international search report

18 August 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2020/092356

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	110188754	A	30 August 2019	None			
CN	101482923	A	15 July 2009	CN	101482923	B	23 May 2012
CN	102216957	A	12 October 2011	CN	102216957	B	16 July 2014
				US	2011254950	A1	20 October 2011
				EP	2359333	B1	19 December 2012
				WO	2010041034	A1	15 April 2010
				US	8810648	B2	19 August 2014
				GB	0818561	D0	19 November 2008
				EP	2359333	A1	24 August 2011
CN	108122234	A	05 June 2018	None			
CN	107146239	A	08 September 2017	CN	107146239	B	07 January 2020
CN	108492297	A	04 September 2018	None			

国际检索报告

国际申请号

PCT/CN2020/092356

A. 主题的分类 G06K 9/32(2006.01)i; G06K 9/34(2006.01)i; G06K 9/62(2006.01)i; G06T 7/00(2017.01)i; G06T 7/11(2017.01)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) G06K G06T 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用)) CNABS;CNTXT;CNKI;SIPOABS;DWPI;USTXT;WOTXT;EPTXT: 视频, 图像, 帧, 分割, 划分, 仿射, 投影, 拉伸, 位移, 缩放, 动态, 更新, 校准, 校正, 背景, video, image, frame, dynamic, project, affine, segment, divide, partition, update, calibration, adjust, revise		
C. 相关文件		
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 110188754 A (腾讯科技深圳有限公司) 2019年 8月 30日 (2019 - 08 - 30) 权利要求第1-15项, 说明书第[0082]-[0239]段, 附图1-14	1-20
X	CN 101482923 A (刘云) 2009年 7月 15日 (2009 - 07 - 15) 说明书第4页第8行至第16页第3行	1-8, 13-15, 19, 20
X	CN 102216957 A (埃西斯创新有限公司) 2011年 10月 12日 (2011 - 10 - 12) 说明书第[0072]-[0276]段	1-8, 13-15, 19, 20
X	CN 108122234 A (北京市商汤科技开发有限公司) 2018年 6月 5日 (2018 - 06 - 05) 说明书第[0051]-[0219]段	9-12, 16-18, 19, 20
A	CN 107146239 A (武汉大学) 2017年 9月 8日 (2017 - 09 - 08) 全文	1-20
A	CN 108492297 A (重庆理工大学) 2018年 9月 4日 (2018 - 09 - 04) 全文	1-20
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。		
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件		
国际检索实际完成的日期 2020年 7月 3日		国际检索报告邮寄日期 2020年 8月 18日
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		授权官员 任洪潮 电话号码 (86-512) 88995644

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/092356

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	110188754	A	2019年 8月 30日	无	
CN	101482923	A	2009年 7月 15日	CN	101482923 B 2012年 5月 23日
CN	102216957	A	2011年 10月 12日	CN	102216957 B 2014年 7月 16日
				US	2011254950 A1 2011年 10月 20日
				EP	2359333 B1 2012年 12月 19日
				WO	2010041034 A1 2010年 4月 15日
				US	8810648 B2 2014年 8月 19日
				GB	0818561 D0 2008年 11月 19日
				EP	2359333 A1 2011年 8月 24日
CN	108122234	A	2018年 6月 5日	无	
CN	107146239	A	2017年 9月 8日	CN	107146239 B 2020年 1月 7日
CN	108492297	A	2018年 9月 4日	无	