

时序动作分割方法、装置及模型训练方法、装置和存储介质

相关申请的交叉引用

本申请是以 CN 申请号为 202410057649.0，申请日为 2024 年 1 月 15 日的申请为
5 基础，并主张其优先权，该 CN 申请的公开内容在此作为整体引入本申请中。

技术领域

本公开涉及人工智能技术领域，特别是一种时序动作分割方法、装置及模型训练
方法、装置和存储介质。

10

背景技术

时序动作分割技术是指，给定一段视频，预测出每一帧中的人体动作类别。相关
技术中，时序动作分割技术通常分为动作识别和时序推理两部分。在动作识别部分，
首先提取视频帧的视觉特征，然后通过机器学习技术（如深度神经网络）将视频帧特
15 征转换为度量空间中的特征向量，在度量空间中，不同动作类别的视频帧特征距离较
大，从而便于动作分类。在时序推理部分，将重点关注人体动作在时序上的连续性，
将一个连续的时间片段内的所有帧预测出相同的动作类别。

相关技术中，时序动作分割技术通常采用深度卷积神经网络作为视频特征提取器
和度量空间学习器，采用循环神经网络、时间卷积网络和 **Transformer** 架构对帧级的
20 动作特征进行时序建模，采用大规模视频标注数据训练深度神经网络，采用交叉熵损
失函数优化深度卷积神经网络。

发明内容

本公开的一个目的在于如何提高动作分割的准确度。

25 根据本公开的一些实施例的一个方面，提出一种时序动作分割模型的训练方法，
包括：对样本数据中每个模态的数据进行高维特征提取，获取每个模态的第一特征；
根据第一特征，通过降维处理，获取中间损失函数，并通过维度恢复获取每个模态的
第二特征；根据第二特征和补充特征，通过对不同模态的特征的融合，获取运动表示；
分别通过类别编码器网络、边界编码器网络处理运动表示，获取动作类别预测向量和
30 时序边界预测向量，确定动作类别预测结果，并确定预测损失函数，其中，时序动作

分割模型根据中间损失函数和预测损失函数调整参数，直至完成训练。

在一些实施例中，根据第二特征和补充特征，通过对不同模态的特征的融合，获取运动表示包括：根据第二特征和补充特征获取注意力增强的运动流特征和注意力增强的空间流特征；通过多头交叉注意力机制获取跨流注意力特征；根据跨流注意力特征、注意力增强的运动流特征和注意力增强的空间流特征，获取运动特征向量和空间特征向量；根据运动特征向量和空间特征向量，确定运动表示。

在一些实施例中，根据第一特征，通过降维处理，获取中间损失函数，并通过维度恢复获取每个模态的第二特征包括：对第一特征进行时序编码，经过全连接层确定中间类别特征，并确定中间类别损失函数；将中间类别特征恢复维度，并与第一特征相加，获取第二特征；将第二特征，通过降维处理和时序信息提取获取中间边界特征，并根据中间边界特征获取中间边界损失函数，其中，中间损失函数包括中间类别损失函数和中间边界损失函数。

在一些实施例中，分别通过类别编码器网络、边界编码器网络处理运动表示，获取动作类别预测向量和时序边界预测向量，确定动作类别预测结果包括：在编码器网络的起始处将类别编码器网络中的查询特征向量用于检索边界编码器网络中的键值对特征向量，获取边界特征；在编码器网络的末尾处将边界编码器网络中的查询特征向量用于检索边界编码器中的键值对特征向量，获取类别特征；根据边界特征和类别特征，通过维度调整获取动作类别预测向量和时序边界预测向量；在时序边界预测向量中根据置信度筛选出边界值，并根据边界值调整动作类别预测向量，获取动作类别预测结果。

在一些实施例中，确定预测损失函数包括：获取动作类别预测结果对应的动作分类损失函数；根据时序边界预测向量获取边界概率向量，并根据边界概率向量确定时序边界损失函数，其中，预测损失函数包括动作分类损失函数和时序边界损失函数。

在一些实施例中，根据第二特征和补充特征获取注意力增强的运动流特征和注意力增强的空间流特征包括：根据惯性模态的第二特征获取初始运动流特征，根据关键点模态的第二特征和边界框模态的第二特征获取初始空间流特征，其中，模态包括惯性、关键点和边界框；分别将初始运动流特征、初始空间流特征与补充特征相加，获取辅助增强运动流特征和辅助增强空间流特征；根据辅助增强运动流特征和补充特征，利用多头注意力机制，获取注意力增强的运动流特征和注意力增强的空间流特征。

在一些实施例中，根据跨流注意力特征、注意力增强的运动流特征和注意力增强

的空间流特征，获取运动特征向量和空间特征向量包括：通过逐帧卷积层处理第一跨流注意力特征和注意力增强的运动流特征，获取第三特征；通过逐帧卷积层处理第二跨流注意力特征和注意力增强的空间流特征，获取第四特征；将第三特征、第一跨流注意力特征和注意力增强的运动流特征相加，获取运动特征向量；将第四特征、第二跨流注意力特征和注意力增强的空间流特征相加，获取空间特征向量，其中，第一跨流注意力特征为通过将空间流的查询特征向量来匹配运动流的键值对特征向量所获取的跨流注意力特征，第二跨流注意力特征为通过将运动流的查询特征向量来匹配空间流的键值对特征向量所获取的跨流注意力特征。

10 在一些实施例中，根据运动特征向量和空间特征向量，确定运动表示包括：将运动特征向量与空间特征向量相加，作为运动表示。

15 在一些实施例中，将第二特征，通过降维处理和时序信息提取获取中间边界特征，并根据中间边界特征获取中间边界损失函数包括：通过一层卷积神经网络处理第二特征，获取第五特征；通过具有残差连接的多层卷积神经网络获取第五特征中的时序信息，并与第五特征相加，确定中间边界特征；将中间边界特征通过激活函数获得边界分类概率向量；根据边界分类概率向量确定中间边界损失函数。

20 在一些实施例中，该训练方法还包括：获取中间类别损失函数、中间边界损失函数、动作分类损失函数和时序边界损失函数的加权和，作为损失函数值，其中，时序动作分割模型根据损失函数值调整参数，中间损失函数包括中间类别损失函数和中间边界损失函数，预测损失函数包括动作分类损失函数和时序边界损失函数。

25 在一些实施例中，补充特征为根据物联网设备的工作日志确定。

根据本公开的一些实施例的一个方面，提出一种时序动作分割方法，包括：获取目标的每个模态的数据和补充数据；将每个模态的数据和补充维度数据输入时序动作分割模型，获取目标的动作类别预测结果，其中，时序动作分割模型为根据上文中提到的任意一种时序动作分割模型的训练方法生成。

25 在一些实施例中，该方法符合以下至少一项：模态包括惯性、关键点和边界框；或补充数据包括物联网设备的工作日志。

30 根据本公开的一些实施例的一个方面，提出一种时序动作分割模型的训练装置，包括：特征提取模块，被配置为对样本数据中每个模态的数据进行高维特征提取，获取每个模态的第一特征；特征约束模块，被配置为根据第一特征，通过降维处理，获取中间损失函数，并通过维度恢复获取每个模态的第二特征；融合模块，被配置为根

据第二特征和补充特征，通过对不同模态的特征的融合，获取运动表示；和交互双分支模块，被配置为分别通过类别编码器网络、边界编码器网络处理运动表示，获取动作类别预测向量和时序边界预测向量，确定动作类别预测结果，并确定预测损失函数，其中，时序动作分割模型根据中间损失函数和预测损失函数调整参数，直至完成训练。

5 在一些实施例中，该训练装置还包括：损失值确定单元，被配置为获取中间类别损失函数、中间边界损失函数、动作分类损失函数和时序边界损失函数的加权和，作为损失函数值，其中，时序动作分割模型根据损失函数值调整参数，中间损失函数包括中间类别损失函数和中间边界损失函数，预测损失函数包括动作分类损失函数和时序边界损失函数。

10 根据本公开的一些实施例的一个方面，提出一种时序动作分割装置，包括：数据获取模块，被配置为获取目标的每个模态的数据和补充数据；和预测模块，被配置为将每个模态的数据和补充维度数据输入时序动作分割模型，获取目标的动作类别预测结果，其中，时序动作分割模型为根据上文中提到的任意一种时序动作分割模型的训练方法生成。

15 根据本公开的一些实施例的一个方面，提出一种数据处理装置，包括：存储器；以及耦接至存储器的处理器，处理器被配置为基于存储在存储器的指令执行上文中提到的任意一种方法。

根据本公开的一些实施例的一个方面，提出一种计算机可读存储介质，其上存储有计算机程序指令，该指令被处理器执行时实现上文中提到的任意一种方法。

20 根据本公开的一些实施例的一个方面，提出一种计算机程序产品，包括计算机程序或指令，所述计算机程序或指令被处理器执行时实现上文中提到的任意一种方法。

根据本公开的一些实施例的一个方面，提出一种计算机程序，用于使处理器执行上文中提到的任意一种方法。

25 附图说明

此处所说明的附图用来提供对本公开的进一步理解，构成本公开的一部分，本公开的示意性实施例及其说明用于解释本公开，并不构成对本公开的不当限定。在附图中：

图 1 为本公开的时序动作分割模型的训练方法的一些实施例的流程图。

30 图 2 为本公开的时序动作分割模型的训练方法中模态信息约束的一些实施例的

流程图。

图 3 为本公开的时序动作分割模型的训练方法中特征融合过程的一些实施例的流程图。

图 4 为本公开的时序动作分割模型的训练方法中确定动作类别预测结果的一些
5 实施例的流程图。

图 5 为本公开的时序动作分割模型的一些实施例的示意图。

图 6 为本公开的时序动作分割方法的一些实施例的流程图。

图 7 为本公开的时序动作分割模型的训练装置的一些实施例的示意图。

图 8 为本公开的时序动作分割装置的一些实施例的示意图。

10 图 9 为本公开的数据处理装置的一些实施例的示意图。

图 10 为本公开的数据处理装置的另一一些实施例的示意图。

具体实施方式

下面通过附图和实施例，对本公开的技术方案做进一步的详细描述。

15 相关的时序动作分割技术通常将视频视觉颜色信息作为特征提取和模型训练的样本，从而提取动作特征用于帧级动作类别预测，采用多阶段调整或独立的边界分支来加强动作间的时序关系。这样的方式由于仅使用视觉颜色信息，表示和理解人体动作的角度受限。

多模态动作识别技术是指，给定多个种类的数据，对这些数据提供的互补信息进行融合，以获得全面的人体动作表示，进而对人体动作进行预测。相关多模态动作识别技术通常采用深度卷积神经网络和 **Transformer** 结构，通过大规模有标注数据训练网络模型，对不同模态的信息进行融合，得到动作识别结果。相关技术中多模态融合方法主要有：1) 数据级融合，指将多源数据集合，然后输入到特征学习的网络中；2) 特征级融合，指不同模态数据先转化为高维特征表达再进行融合；3) 决策级融合，指
20 每个模态数据分别训练好分类器，将每个分类器的预测分数进行融合。
25

相关的特征级融合的多模态融合方法，主要采用高维特征拼接、相加或注意力机制的方式，不能保证融合前多模态特征的质量，这可能会导致不相关信息的融合并引入干扰，且使用的特征融合方式较为简单，无法充分利用各个模态互补的信息。

针对上述问题，本公开提出一种时序动作分割方法、装置及模型训练方法、装置
30 和存储介质，以如何提高动作分割的准确度。

本公开的时序动作分割模型的训练方法的一些实施例的流程图如图 1 所示。

在步骤 S11 中，对样本数据中每个模态的数据进行高维特征提取，获取每个模态的第一特征。在一些实施例中，每条样本数据中包括针对处理对象（如同一个人）的，通过多种渠道、设备获取的多种模态数据。在一些实施例中，样本数据中还包括每帧
5 对应的动作类别标签。

在一些实施例中，上述模态可以包括惯性、关键点和边界框，从而避免了使用和依赖于视觉颜色，导致容易暴露人和周围环境的敏感信息、泄露隐私的风险，提高了安全性。

在一些实施例中，特征提取操作包括，对于惯性数据、关键点数据、边界框数据
10 分别进行高维特征提取，得到惯性的第一特征、关键点的第一特征，以及边界框的第一特征。在一些实施例中，可以针对每种模态的数据分别设置编码器进行高维特征提取。在一些实施例中，高维特征指的是维数大于等于预定数量的特征，在一些实施例中，预定数量为 1000。

在一些实施例中，第一特征为 $L \times D$ 维度的特征， L 为时间长度， D 为模态特征
15 维度。

在步骤 S13 中，根据第一特征，通过降维处理，获取中间损失函数，并通过维度恢复获取每个模态的第二特征。通过中间损失函数的约束，对每种模态的信息分别进行约束，提高后续融合步骤中所基于的特征的质量，降低干扰信息量。

在一些实施例中，中间损失函数包括中间类别损失函数和中间边界损失函数，从
20 而从类别、边界两个角度对各个模态的信息进行约束，提高特征优化质量。

在一些实施例中，模态信息约束的过程如图 2 所示。

在步骤 231 中，通过长短期记忆网络对第一特征进行时序编码，再经过全连接层处理得到中间类别特征，并确定中间类别损失函数。

在一些实施例中，中间类别特征的维数小于第一特征的维数。在一些实施例中，
25 中间类别特征为 $L \times C$ 维度的特征，其中 C 为预测的动作类别数。

在一些实施例中，通过 Softmax 操作得到 $L \times C$ 维度的分类概率向量，根据分类概率向量确定中间类别损失函数。在一些实施例中，使用随机梯度下降算法和反向传播算法，根据中间类别损失函数优化网络参数。

在步骤 232 中，将中间类别特征恢复维度，并与第一特征相加，获取第二特征。
30 在一些实施例中，可以使用全连接神经网络，将 $L \times C$ 维度的中间类别特征恢复至 L

×D 维度，进而与第一特征相加，获取第二特征。在一些实施例中，特征相加的方式为逐位相加。通过这样的方法，能够避免维度降低导致的高维信息丢失。

在步骤 233 中，将第二特征，通过降维处理和时序信息提取获取中间边界特征，并根据中间边界特征获取中间边界损失函数。

5 在一些实施例中，将第二特征作为输入信息，通过一层卷积神经网络处理第二特征，获取第五特征；进一步的，通过具有残差连接的多层卷积神经网络获取第五特征中的时序信息，并与第五特征相加，确定中间边界特征。

10 在一些实施例中，将中间边界特征通过激活函数获得边界分类概率向量，向量中的值在 0-1 之间，值越接近 0 表示对应帧的动作时序边界的概率越小，越接近 1 则该帧位于动作时序边界的概率越大。根据边界分类概率向量确定中间边界损失函数。在一些实施例中，使用随机梯度下降算法和反向传播算法，根据中间类别损失函数优化网络参数。

15 通过这样的方法，能够通过中间损失函数的约束，对每种模态的信息分别进行约束，提高后续融合步骤中所基于的特征的质量，降低干扰信息量；在降维处理后升维并与原信息相加，避免维度降低导致的高维信息丢失，提高信息全面性；从类别、边界两个角度对各个模态的信息进行约束，提高特征优化质量。

20 在步骤 S15 中，根据第二特征和补充特征，通过对不同模态的特征的融合，获取运动表示。在一些实施例中，补充特征可以为根据时序上稀疏的信息确定的特征，在一些实施例中，补充特征可以为只与特征动作相关的特征，例如根据物联网设备的工作日志确定的特征。在一些实施例中，可以获取补充数据作为一种额外的模态的数据，在上述步骤 S11 中，不执行高维特征提取，而是通过全连接层处理补充数据，获取补充特征。

在一些实施例中，特征融合过程如图 3 所示。

25 在步骤 351 中，根据第二特征和补充特征获取注意力增强的运动流特征和注意力增强的空间流特征。

在一些实施例中，可以先针对各个模态的特点，对不同模态的第二特征进行整合，例如，根据惯性模态的第二特征获取初始运动流特征，根据关键点模态的第二特征和边界框模态的第二特征获取初始空间流特征。在整合后的特征的基础上执行进一步操作。通过这样的方法，能够降低后续运算量，提高数据处理效率。

30 在一些实施例中，分别将初始运动流特征、初始空间流特征与补充特征相加，获

取辅助增强运动流特征和辅助增强空间流特征，通过这样的方法，能够在运动流、空间流的特征中分别增加补充特征信息，提高了辅助增强运动流特征和辅助增强空间流特征携带的有效信息量，有利于提高动作分割的准确度。

- 5 在一些实施例中，根据辅助增强运动流特征和补充特征，利用多头注意力机制，获取注意力增强的运动流特征和注意力增强的空间流特征。

在步骤 352 中，通过多头交叉注意力机制获取跨流注意力特征。

- 10 在一些实施例中，可以基于上文中得到的注意力增强的运动流特征和注意力增强的空间流特征，采用多头注意力机制对两种类型的特征相互进行跨流处理，例如通过将空间流的查询特征向量来匹配运动流的键值对特征向量获取第一跨流注意力特征，通过将运动流的查询特征向量来匹配空间流的键值对特征向量所获取的第二跨流注意力特征。在一些实施例中，跨流注意力特征包括第一跨流注意力特征和第二跨流注意力特征。

在步骤 353 中，根据跨流注意力特征、注意力增强的运动流特征和注意力增强的空间流特征，获取运动特征向量和空间特征向量。

- 15 在一些实施例中，可以根据跨流注意力特征、注意力增强的运动流特征和注意力增强的空间流特征，获取细粒度的动作特征向量，作为运动特征向量和空间特征向量。

- 20 在一些实施例中，通过逐帧卷积层处理第一跨流注意力特征和注意力增强的运动流特征，获取第三特征；通过逐帧卷积层处理第二跨流注意力特征和注意力增强的空间流特征，获取第四特征；将第三特征、第一跨流注意力特征和注意力增强的运动流特征相加，获取运动特征向量；将第四特征、第二跨流注意力特征和注意力增强的空间流特征相加，获取空间特征向量。通过这样的方法，运动特征向量中包括经过逐帧卷积层处理的第一跨流注意力特征和注意力增强的运动流特征，以及第一跨流注意力特征和注意力增强的运动流特征本身，空间特征向量中包括经过逐帧卷积层处理的第二跨流注意力特征和注意力增强的空间流特征，以及第二跨流注意力特征和注意力增强的空间流特征本身，提高了向量中信息的全面性。

在步骤 354 中，根据运动特征向量和空间特征向量，确定运动表示。在一些实施例中，将运动特征向量与空间特征向量相加，作为运动表示。通过这样的方式将细粒度的特征组合，提高运动表示的全面性和鲁棒性。

- 30 通过上文所示实施例中的方法，能够基于注意力机制和补充信息提高不同模态的信息的融合度，对动作的运动和空间信息的交互上下文进行编码，从而获得完整的动

作表示，提高运动流和空间流信息之间的融合度，提高对动作的表征能力。

在步骤 S17 中，分别通过类别编码器网络、边界编码器网络处理运动表示，获取动作类别预测向量和时序边界预测向量，确定动作类别预测结果，并确定预测损失函数。在一些实施例中，预测损失函数可以包括动作分类损失函数和时序边界损失函数，
5 从而从分类、边界两个角度对模型进行约束，提高模型的准确度。

在一些实施例中，可以在动作类别预测结果的基础上，通过 Softmax 操作得到分类概率向量，进而确定动作分类损失函数。在一些实施例中，使用随机梯度下降算法和反向传播算法，根据动作分类损失函数优化网络参数。

在一些实施例中，可以在时序边界测量向量的基础上，通过 Sigmoid 操作得到边
10 界概率向量，确定时序边界损失函数。在一些实施例中，使用随机梯度下降算法和反向传播算法，根据时序边界损失函数优化网络参数。

在一些实施例中，时序动作分割模型根据前述步骤 S13 中的中间损失函数，以及当前步骤的预测损失函数进行参数调整，直至完成训练。

在一些实施例中，可以根据需要设置不同损失函数的权重，以平衡不同损失函数
15 的重要性。在一些实施例中，可以设置中间损失函数与预测损失函数的权重不同，从而平衡中间损失和最终损失对模型的影响力度。

在一些实施例中，获取中间类别损失函数、中间边界损失函数、动作分类损失函数和时序边界损失函数的加权和，作为损失函数值，

其中，时序动作分割模型根据损失函数值调整参数，中间损失函数包括中间类别
20 损失函数和中间边界损失函数，预测损失函数包括动作分类损失函数和时序边界损失函数。

在一些实施例中，确定动作类别预测结果的流程图如图 4 所示。

在步骤 471 中，在编码器网络的起始处将类别编码器网络中的查询特征向量用于检索边界编码器网络中的键值对，获取边界特征。在一些实施例中，编码器网络包括
25 类别编码器网络和边界编码器网络，类别编码器网络和边界编码器网络分别、并行进行特征处理。编码器网络的起始处包括类别编码器网络的起始处和边界编码器网络的起始处。

在步骤 472 中，在编码器网络的末尾处将边界编码器网络中的查询特征向量用于检索边界编码器中的键值对特征向量，获取类别特征。在一些实施例中，编码器网络
30 的末尾处包括类别编码器网络的末尾处和边界编码器网络的末尾处。

在步骤 473 中, 根据边界特征和类别特征, 通过维度调整获取动作类别预测向量和时序边界预测向量。在一些实施例中, 可以通过全连接层将 $L \times D$ 维度的类别特征调整至 $L \times C$ 维度, 得到动作类别预测向量。在一些实施例中, 可以通过全连接层将 $L \times D$ 维度的边界特征调整至 $L \times 1$ 维度, 得到时序边界预测向量。

- 5 在步骤 474 中, 在时序边界预测向量中根据置信度筛选出边界值, 并根据边界值调整动作类别预测向量, 获取动作类别预测结果。

在一些实施例中, 可以设置置信度阈值, 在时序边界预测向量中筛选出置信度高于置信度阈值的边界值, 采用边界值对动作类别预测向量进行微调, 获得动作类别预测结果。

- 10 通过上文所示实施例中的方法, 能够通过多阶段的交互, 提高时序建模能力, 改善过度分割, 提高动作分割的准确度。

基于上文所示实施例中的方法, 能够提取多模态数据的特征, 提高动作表示的全面度; 对每种模态的信息分别进行约束, 提高后续融合步骤中所基于的特征的质量, 充分融合多模态特征, 提高了动作分割的准确度。

- 15 本公开的时序动作分割模型的一些实施例的示意图如图 5 所示。

时序动作分割模型包括特征提取模块、中间瓶颈模块、双流注意力融合模块和多阶段交互双分支模块。

1、特征提取模块

在一些实施例中, 特征提取模块的数量与数据来源或数据分类的数量相对应。

- 20 如图 5 (a) 所示, 针对不同模态的数据的特点, 采用不同的网络结构。

在一些实施例中, 惯性数据、人体关键点数据、人体边界框数据包含的信息丰富, 采用 **Transformer** 编码器结构 $G(\cdot)$ 进行高维特征提取, 得到每个模态的第一特征 $I_x = G(x)$, 其中, x 可表示为惯性数据、人体关键点数据、人体边界框数据。第一特征作为中间瓶颈模块的输入。

- 25 在一些实施例中, 来自物联网设备的工作日志在时间上稀疏, 并且只与特定动作有关, 因此作为补充数据, 通过简单的全连接层 $C(\cdot)$ 得到对应的补充特征 $I_l = C(l)$, 其中, l 表示来自物联网设备的工作日志信息。

2、中间瓶颈模块

在一些实施例中, 特征提取模块的数量与第一特征对应的模态的种类数量相对应。

- 30 以一个模态的第一特征 I 为例, $L \times D$ 维度的 I 作为输入, 其中 L 为时间长度, D 为

模态特征维度。通过降低特征维度,计算中间类别损失 $L_{c\ inner}$ 和中间边界损失 $L_{b\ inner}$,再通过升高特征维度得到增强的模态特征 I' ,该特征即为中间瓶颈模块的输出。中间瓶颈模块包括中间类别分支和中间边界分支。

在一些实施例中,中间瓶颈模块结构如图5(b)所示。

- 5 A、在中间类别分支中,给定一个 $L \times D$ 维度的模态第一特征 I ,将 I 输入中间类别分支。通过长短期记忆网络 $g(\cdot)$ 对 I 进行时序编码,再通过一个全连接层神经网络 $f(\cdot)$ 得到 $L \times C$ 的中间类别特征 $I_c = f(g(I))$,其中 C 为预测的动作类别数。然后通过 softmax 操作得到 $L \times C$ 维度的分类概率向量 $s(I_c)$ 。在一些实施例中,根据如下公式(1)计算中间类别损失函数。

$$10 \quad L_{c\ inner} = \frac{1}{T} \sum_t -\log(s(I_c)_{ct}) \quad (1)$$

在一些实施例中,使用随机梯度下降算法和反向传播算法优化网络参数。

进一步的,在中间类别分支中,在使用全连接层神经网络 $f(\cdot)$ 将特征维度从 $L \times C$ 重新调整至 $L \times D$,并将计算结果与原始特征输入 I 相加,如公式(2)所示,以避免维度降低导致的高维信息丢失。

$$15 \quad I' = f(I_c) + I \quad (2)$$

公式(2)中, I' 为中间类别分支的输出,同时也是中间边界分支的输入。

- B、在中间边界分支中,中间边界分支接收中间类别分支的输出 I' 作为输入,通过一层卷积神经网络 $p(\cdot)$ 将特征向量大小从 $L \times D$ 调整至 $L \times 1$,得到特征向量 $I_b = p(I')$ 。之后通过具有残差连接的多层卷积神经网络 $p'(\cdot)$ 建模特征中的时序信息,根据公式(2)
- 20 得到中间边界特征。

$$I_b' = p'(I_b) + I_b \quad (3)$$

- 对于中间边界特征 I_b' ,通过 sigmoid 操作得到 $L \times 1$ 维度的边界分类概率向量 $m(I_b')$,向量中的值都在 0-1 之间。给定一帧的边界分类概率向量,值越接近 0 表示该帧处于动作时序边界的概率越小,反之越靠近 1,该帧位于动作时序边界的概率越大。根据公式(4)计算中间边界损失函数。
- 25

$$L_{b\ inner} = \frac{1}{T} \sum_t (y_t \cdot \log(m(I_b')) + (1 - y_t) \cdot \log(1 - m(I_b'))) \quad (4)$$

上述公式中, y_t 是位于 t 时刻的边界真值。在一些实施例中,使用随机梯度下降算法和反向传播算法优化网络参数。

3、双流注意力融合模块

- 30 双流注意力融合模块以通过中间瓶颈模块得到的增强模态特征 I' 作为输入,得到

全面人体动作特征 I_F 。双流注意力融合模块由补充注意力融合模块和运动-空间注意力融合模块组成。

双流注意力融合模块结构如图 5 (c) 所示, 包括运动-空间注意力融合模块和运动-空间注意力融合模块。

5 A、补充注意力融合模块

通过上述的特征提取模块和中间瓶颈模块, 分别得到了惯性数据、人体关键点数据、人体边界框数据和来自物联网设备的工作日志分别对应的特征向量。将来自物联网设备的工作日志提供准确的表示作为补充特征 I_L , 将惯性模态的第二特征作为人体的运动流特征 I_M 。另外, 将关键点模态和边界框模态的第二特征相加, 作为空间流特征 I_S 。

在一些实施例中, 将补充特征 I_L 分别与运动流特征 I_M 和空间流特征 I_S 相加, 获得辅助增强的特征向量 I_{M+L} 和 I_{S+L} 。

辅助增强的特征向量与补充特征一起被送入多头注意力层, 使用如下公式 (5)

(6) 生成注意力增强的特征向量 I_{M+L}' 和 I_{S+L}' :

$$15 \quad I_{M+L}' = \text{softmax}\left(\frac{Q_L K_{M+L}^T}{\sqrt{d}}\right) V_{M+L} + I_{M+L} \quad (5)$$

$$I_{S+L}' = \text{softmax}\left(\frac{Q_L K_{S+L}^T}{\sqrt{d}}\right) V_{S+L} + I_{S+L} \quad (6)$$

上述公式中, d 为归一化常数, T 表示矩阵转置, 辅助增强的特征向量提供键 K 和值 V , 补充特征提供查询 Q 。

B、运动-空间注意力融合模块

20 通过多头交叉注意力机制建模运动特征和空间流特征。根据如下公式 (7) (8) 使用一个流的查询特征向量 Q 来匹配另一个流的键值对特征向量 K 、 V :

$$I_{M \rightarrow S} = \text{softmax}\left(\frac{Q_S K_M^T}{\sqrt{d}}\right) V_M \quad (7)$$

$$I_{S \rightarrow M} = \text{softmax}\left(\frac{Q_M K_S^T}{\sqrt{d}}\right) V_S \quad (8)$$

上述公式中, $I_{M \rightarrow S}$ 和 $I_{S \rightarrow M}$ 是编码运动流和空间流之间相关性的跨流注意力特征。

25 将跨流注意力特征通过逐帧卷积层 $c(\cdot)$ 转换成细粒度动作特征向量, 称为运动特征向量 $I_{M \rightarrow S}'$ 、空间特征向量 $I_{S \rightarrow M}'$ 。在一些实施例中, 可以根据公式 (9) (10) 确定运动特征向量 $I_{M \rightarrow S}'$ 、空间特征向量 $I_{S \rightarrow M}'$ 。

$$I_{M \rightarrow S}' = c(I_{M \rightarrow S} + I_{M+L}') + I_{M \rightarrow S} + I_{M+L}' \quad (9)$$

$$I_{S \rightarrow M}' = c(I_{S \rightarrow M} + I_{S+L}') + I_{S \rightarrow M} + I_{S+L}' \quad (10)$$

30 经过补充注意力融合模块和运动-空间注意力融合模块, 补充信息得以很好嵌入,

运动-空间两个流的信息得到充分交互。

在一些实施例中,通过逐元素加法运算将运动特征向量 $I_{M \rightarrow S}'$ 、空间特征向量 $I_{S \rightarrow M}'$ 组合,如公式(11)所示,得到全面且鲁棒的运动表示。

$$I_F = I_{M \rightarrow S}' + I_{S \rightarrow M}' \quad (11)$$

5 4、多阶段交互双分支模块

多阶段交互双分支模块包括类别分支和边界分支。将运动表示 I_F 分别送入两个分支。两个分支分别由多层 **Transformer** 编码器层组成。

第一阶段的信息交互位于动作分支的起始处,将从类别分支获得的查询特征向量 Q_C 来检索从边界分支获得的键值对特征向量 K_B 、 V_B 。经过多头交叉注意力获得第一次信息交互后的边界特征 $F_B = \text{softmax}(\frac{Q_C K_B^T}{\sqrt{d}}) V_B$ 。

第二阶段的信息交互位于两个分支的末尾,边界分支提供查询特征向量 Q'_B ,类别分支提供键和值的特征向量 K'_C 、 V'_C 。根据多头交叉注意力获得第二次信息交互后的类别特征 $F_C = \text{softmax}(\frac{Q'_B K'_C{}^T}{\sqrt{d}}) V'_C$ 。

在类别分支的末尾,由一个全连接层将 $L \times D$ 的特征调整至 $L \times C$,得到每一帧的动作类别预测向量 X_C 。在边界分支的末尾,由一个全连接层将 $L \times D$ 的特征调整至 $L \times 1$,得到时序边界预测向量 X_B 。首先设定阈值 α ,从 X_B 中筛选出置信度高的边界值 $\hat{X}_B = \max(X_B, \alpha)$ 。使用边界值 $\hat{X}_B$ 对类别预测向量 X_C 进行微调,如下公式(12)所示。

$$\hat{x}_C = \frac{X_C + \sum_{s \in \{-1,1\}} \sum_{\beta=1}^L \hat{x}_{B \ t+s:\beta} \exp\left(-k \sum_{j=1}^{\beta} \hat{x}_{B \ t+s:j}\right)}{1 + \sum_{s \in \{-1,1\}} \sum_{\beta} \exp\left(-k \sum_{j=1}^{\beta} \hat{x}_{B \ t+s:j}\right)} \quad (12)$$

上述公式中, k 控制权重的衰减率, L 为片段的时间长度, $s \in \{-1,1\}$ 表示在加重的过程中特征向量按照时间前向和后向两个方向进行聚合。得到的 $\hat{x}_C$ 即为网络最后的输出结果,为经过时序边界约束后的帧级动作类别预测结果。

在一些实施例中,通过 **softmax** 操作得到 $L \times C$ 维度的分类概率向量 $s(\hat{x}_C)$,根据公式(13)计算动作分类损失函数。

$$L_{c \ final} = \frac{1}{L} \sum_{t=1}^L -\log(s(\hat{x}_C)_{c,t}) \quad (13)$$

在一些实施例中,使用随机梯度下降算法和反向传播算法优化网络参数。

在一些实施例中,通过 **sigmoid** 操作得到 $L \times 1$ 维度的边界概率向量 $m(X_B)$,根据公式(14)计算时序边界损失函数:

$$L_{b \ final} = \frac{1}{L} \sum_{t=1}^L (y_t \cdot \log(m(X_B)) + (1 - y_t) \cdot \log(1 - m(X_B))) \quad (14)$$

在一些实施例中， y_t 是位于 t 时刻的边界真值。在一些实施例中，使用随机梯度下降算法和反向传播算法优化网络参数。

基于上文所示的时序动作分割模型的训练方法，基于上述时序动作分割模型，通过特征提取模块和中间瓶颈模块提取多模态数据的特征，通过双流注意力融合模块充分融合多模态特征，获得全面且鲁棒的人体动作表示，通过多阶段交互双分支模块建模动作类别和时序边界间的信息交互，加强网络的时序建模能力，从而实现准确的时序动作分割结果；为每个模态特征引入类别感知和时序感知能力的方法，通过计算中间类别损失和中间边界损失，为后续的特征融合提供最有效的信息；通过基于注意力机制的多阶段模态融合方法，结合了辅助信息并促进运动流和空间流信息之间的充分融合交互，使用注意力机制建模各个信息间的关系，提高了动作的表征能力；通过类别-边界的多阶段信息交互方法，表征动作类别和时序边界间的上下文关系，提高了时序建模能力，有效改善过度分割问题。

在一些实施例中，在训练过程中，收集样本数据，样本数据中包含一个时间段的多模态数据及每帧相对应的动作类别标签，顺序遍历多模态数据及标签。将样本数据中的多模态数据输入特征提取模块，在训练过程中得到中间损失函数和预测损失函数，根据损失函数，使用随机梯度下降算法和反向传播算法优化网络参数。

在一些实施例中，多模态融合的时序动作分割模型是一个端到端的网络，各个模块同时训练。在一些实施例中，训练过程中模型生成四个损失函数：中间类别损失函数 $L_{c\ inner}$ 、中间边界损失函数 $L_{b\ inner}$ 、动作分类损失函数 $L_{c\ final}$ 和时序边界损失函数 $L_{b\ final}$ ，按公式（15）将损失函数相加。

$$L = \lambda_1 L_{c\ inner} + (1 - \lambda_1) L_{c\ final} + \lambda_2 (L_{b\ inner} + L_{b\ final}) \quad (15)$$

在上述公式中， λ_1 和 λ_2 为权重系数， λ_1 平衡中间损失和最终损失， λ_2 用于平衡动作类别损失和时序边界损失，进而采用随机梯度下降算法优化总网络参数。

通过这样的方法，能够根据需要平衡中间损失和最终损失、动作类别损失和时序边界损失，提高模型的可控性。

本公开的时序动作分割方法的一些实施例的流程图如图6所示。

在步骤S61中，获取目标的每个模态的数据和补充数据。在一些实施例中，模态包括惯性、关键点和边界框。在一些实施例中，补充数据包括物联网设备的工作日志。

在步骤S62中，将每个模态的数据和补充维度数据输入时序动作分割模型，获取目标的动作类别预测结果。在一些实施例中，时序动作分割模型为根据上文中提到的

任意一种时序动作分割模型的训练方法生成。在一些实施例中，时序动作分割模型如图 5 所示。

基于上文所示实施例中的方法，能够提取多模态数据的特征，对每种模态的信息分别进行约束，提高后续融合步骤中所基于的特征的质量，充分融合多模态特征，提高了动作分割的准确度。

本公开的时序动作分割模型的训练装置的一些实施例的示意图如图 7 所示。

特征提取模块 701 能够对样本数据中每个模态的数据进行高维特征提取，获取每个模态的第一特征。在一些实施例中，特征提取模块 701 能够执行上文步骤 S11，以及图 5 对应的实施例中特征提取模块执行的任意一种方法。

特征约束模块 702 能够根据第一特征，通过降维处理，获取中间损失函数，并通过维度恢复获取每个模态的第二特征。在一些实施例中时序动作分割模型根据中间损失函数调整参数，直至完成训练。在一些实施例中，特征约束模块 702 能够执行上文步骤 S13、如图 2 所示实施例，以及图 5 对应的实施例中中间瓶颈模块执行的任意一种方法。

融合模块 703 能够根据第二特征和补充特征，通过对不同模态的特征的融合，获取运动表示。在一些实施例中，融合模块 703 能够执行上文步骤 S15、如图 3 所示实施例，以及图 5 对应的实施例中双流注意力融合模块执行的任意一种方法。

交互双分支模块 704 能够分别通过类别编码器网络、边界编码器网络处理运动表示，获取动作类别预测向量和时序边界预测向量，确定动作类别预测结果，并确定预测损失函数。在一些实施例中时序动作分割模型根据预测损失函数调整参数，直至完成训练。在一些实施例中，交互双分支模块 704 能够执行上文步骤 S17、如图 4 所示实施例，以及图 5 对应的实施例中多阶段交互双分支模块执行的任意一种方法。

这样的训练装置能够提取多模态数据的特征，对每种模态的信息分别进行约束，提高后续融合步骤中所基于的特征的质量，充分融合多模态特征，提高了动作分割的准确度。

在一些实施例中，该训练装置还包括损失值确定单元，能够获取中间类别损失函数、中间边界损失函数、动作分类损失函数和时序边界损失函数的加权和，作为损失函数值。在一些实施例中，时序动作分割模型根据损失函数值调整参数，中间损失函数包括中间类别损失函数和中间边界损失函数，预测损失函数包括动作分类损失函数和时序边界损失函数。在一些实施例中，可以根据需要进行权值调整。

这样的训练装置能够根据需要平衡中间损失和最终损失、动作类别损失和时序边界损失，提高模型的可控性。

本公开的时序动作分割装置的一些实施例的示意图如图 8 所示。

数据获取模块 801 能够获取目标的每个模态的数据和补充数据。

- 5 预测模块 802 将每个模态的数据和补充维度数据输入时序动作分割模型，获取目标的动作类别预测结果。时序动作分割模型可以为上文中提到的任意一种，为根据上文中提到的任意一种时序动作分割模型的训练方法，或采用上文中提到的任意一种时序动作分割模型的训练装置训练生成。

- 10 这样的装置能够提取多模态数据的特征，对每种模态的信息分别进行约束，提高后续融合步骤中所基于的特征的质量，充分融合多模态特征，提高了动作分割的准确度。

- 15 本公开数据处理装置的一个实施例的结构示意图如图 9 所示。数据处理装置包括存储器 901 和处理器 902。其中：存储器 901 可以是磁盘、闪存或其它任何非易失性存储介质。存储器用于存储上文中时序动作分割模型的训练方法或时序动作分割方法对应实施例中的指令。处理器 902 耦接至存储器 901，可以作为一个或多个集成电路来实施，例如微处理器或微控制器。该处理器 902 用于执行存储器中存储的指令，能够提高动作分割的准确度。

- 20 在一个实施例中，还可以如图 10 所示，数据处理装置 1000 包括存储器 1001 和处理器 1002。处理器 1002 通过 BUS 总线 1003 耦合至存储器 1001。该数据处理装置 1000 还可以通过存储接口 1004 连接至外部存储装置 1005 以便调用外部数据，还可以通过网络接口 1006 连接至网络或者另外一台计算机系统（未标出）。此处不再进行详细介绍。

在该实施例中，通过存储器存储数据指令，再通过处理器处理上述指令，能够提高动作分割的准确度。

- 25 在另一个实施例中，一种计算机可读存储介质，其上存储有计算机程序指令，该指令被处理器执行时实现时序动作分割模型的训练方法或时序动作分割方法对应实施例中的方法的步骤。本领域内的技术人员应明白，本公开的实施例可提供为方法、装置、或计算机程序产品。因此，本公开可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本公开可采用在一个或多个其中包含
30 有计算机可用程序代码的计算机可用非瞬时性存储介质（包括但不限于磁盘存储器、

CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

本公开是参照根据本公开实施例的方法、设备(系统)和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

至此,已经详细描述了本公开。为了避免遮蔽本公开的构思,没有描述本领域所公知的一些细节。本领域技术人员根据上面的描述,完全可以明白如何实施这里公开的技术方案。

可能以许多方式来实现本公开的方法以及装置。例如,可通过软件、硬件、固件或者软件、硬件、固件的任何组合来实现本公开的方法以及装置。用于所述方法的步骤的上述顺序仅是为了进行说明,本公开的方法的步骤不限于以上具体描述的顺序,除非以其它方式特别说明。此外,在一些实施例中,还可将本公开实施为记录在记录介质中的程序,这些程序包括用于实现根据本公开的方法的机器可读指令。因而,本公开还覆盖存储用于执行根据本公开的方法的程序的记录介质。

需要说明的是,本公开的说明书和权利要求书及附图中的术语“第一”、“第二”等是用于区别类似的对象,而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换,以便这里描述的本公开的实施例能够以除了在这里图示或描述的那些以外的顺序实施。此外,术语“包括”和“具有”以及他们的任何变形,意图在于覆盖不排他的包含,例如,包含了一系列步骤或单元的过程、方法、

系统、产品或设备不必限于清楚地列出的那些步骤或单元，而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

- 最后应当说明的是：以上实施例仅用以说明本公开的技术方案而非对其限制；尽管参照较佳实施例对本公开进行了详细的说明，所属领域的普通技术人员应当理解：依然可以对本公开的具体实施方式进行修改或者对部分技术特征进行等同替换；而不脱离本公开技术方案的精神，其均应涵盖在本公开请求保护的技术方案范围当中。
- 5

权 利 要 求

1. 一种时序动作分割模型的训练方法，包括：

对样本数据中每个模态的数据进行高维特征提取，获取每个模态的第一特征；

根据所述第一特征，通过降维处理，获取中间损失函数，并通过维度恢复获取每个模态的第二特征；

根据所述第二特征和补充特征，通过对不同模态的特征的融合，获取运动表示；

分别通过类别编码器网络、边界编码器网络处理所述运动表示，获取动作类别预测向量和时序边界预测向量，确定动作类别预测结果，并确定预测损失函数，

其中，所述时序动作分割模型根据所述中间损失函数和所述预测损失函数调整参数，直至完成训练。

2. 根据权利要求 1 所述的训练方法，其中，所述根据所述第二特征和补充特征，通过对不同模态的特征的融合，获取运动表示包括：

根据所述第二特征和补充特征获取注意力增强的运动流特征和注意力增强的空间流特征；

通过多头交叉注意力机制获取跨流注意力特征；

根据所述跨流注意力特征、所述注意力增强的运动流特征和所述注意力增强的空间流特征，获取运动特征向量和空间特征向量；

根据所述运动特征向量和所述空间特征向量，确定所述运动表示。

3. 根据权利要求 1 或 2 所述的训练方法，其中，所述根据所述第一特征，通过降维处理，获取中间损失函数，并通过维度恢复获取每个模态的第二特征包括：

对所述第一特征进行时序编码，经过全连接层确定中间类别特征，并确定中间类别损失函数；

将所述中间类别特征恢复维度，并与所述第一特征相加，获取所述第二特征；

将所述第二特征，通过降维处理和时序信息提取获取中间边界特征，并根据所述中间边界特征获取中间边界损失函数，

其中，所述中间损失函数包括所述中间类别损失函数和所述中间边界损失函数。

4. 根据权利要求 1~3 中任意一项所述的训练方法，其中，所述分别通过类别编码器网络、边界编码器网络处理所述运动表示，获取动作类别预测向量和时序边界预测

向量，确定动作类别预测结果包括：

在编码器网络的起始处将所述类别编码器网络中的查询特征向量用于检索所述边界编码器网络中的键值对特征向量，获取边界特征；

在编码器网络的末尾处将边界编码器网络中的查询特征向量用于检索所述边界编码器中的键值对特征向量，获取类别特征；

根据所述边界特征和所述类别特征，通过维度调整获取动作类别预测向量和时序边界预测向量；

在所述时序边界预测向量中根据置信度筛选出边界值，并根据所述边界值调整所述动作类别预测向量，获取所述动作类别预测结果。

5. 根据权利要求 1~4 中任意一项所述的训练方法，其中，所述确定预测损失函数包括：

获取所述动作类别预测结果对应的动作分类损失函数；

根据所述时序边界预测向量获取边界概率向量，并根据所述边界概率向量确定时序边界损失函数，

其中，所述预测损失函数包括所述动作分类损失函数和所述时序边界损失函数。

6. 根据权利要求 2 所述的训练方法，其中，所述根据所述第二特征和补充特征获取注意力增强的运动流特征和注意力增强的空间流特征包括：

根据惯性模态的第二特征获取初始运动流特征，根据关键点模态的第二特征和边界框模态的第二特征获取初始空间流特征，其中，所述模态包括惯性、关键点和边界框；

分别将所述初始运动流特征、所述初始空间流特征与所述补充特征相加，获取辅助增强运动流特征和辅助增强空间流特征；

根据所述辅助增强运动流特征和所述补充特征，利用多头注意力机制，获取所述注意力增强的运动流特征和所述注意力增强的空间流特征。

7. 根据权利要求 2 或 6 所述的训练方法，其中，所述根据所述跨流注意力特征、所述注意力增强的运动流特征和所述注意力增强的空间流特征，获取运动特征向量和空间特征向量包括：

通过逐帧卷积层处理第一跨流注意力特征和所述注意力增强的运动流特征，获取第三特征；

通过逐帧卷积层处理第二跨流注意力特征和所述注意力增强的空间流特征，获取第四特征；

将所述第三特征、所述第一跨流注意力特征和所述注意力增强的运动流特征相加，获取所述运动特征向量；

将所述第四特征、所述第二跨流注意力特征和所述注意力增强的空间流特征相加，获取所述空间特征向量，

其中，所述第一跨流注意力特征为通过将空间流的查询特征向量来匹配运动流的键值对特征向量所获取的所述跨流注意力特征，所述第二跨流注意力特征为通过将运动流的查询特征向量来匹配空间流的键值对特征向量所获取的所述跨流注意力特征。

8. 根据权利要求 2、6 或 7 所述的训练方法，其中，所述根据所述运动特征向量和所述空间特征向量，确定运动表示包括：

将所述运动特征向量与所述空间特征向量相加，作为所述运动表示。

9. 根据权利要求 3 所述的训练方法，其中，所述将所述第二特征，通过降维处理和时序信息提取获取中间边界特征，并根据所述中间边界特征获取中间边界损失函数包括：

通过一层卷积神经网络处理所述第二特征，获取第五特征；

通过具有残差连接的多层卷积神经网络获取所述第五特征中的时序信息，并与所述第五特征相加，确定所述中间边界特征；

将所述中间边界特征通过激活函数获得边界分类概率向量；

根据所述边界分类概率向量确定所述中间边界损失函数。

10. 根据权利要求 1~9 中任意一项所述的训练方法，还包括：

获取中间类别损失函数、中间边界损失函数、动作分类损失函数和时序边界损失函数的加权和，作为损失函数值，

其中，所述时序动作分割模型根据所述损失函数值调整参数，所述中间损失函数包括所述中间类别损失函数和所述中间边界损失函数，所述预测损失函数包括所述动作分类损失函数和所述时序边界损失函数。

11. 根据权利要求 1~9 中任意一项所述的训练方法，其中，所述补充特征为根据物联网设备的工作日志确定。

12. 一种时序动作分割方法，包括：

获取目标的每个模态的数据和补充数据；

将所述每个模态的数据和所述补充维度数据输入所述时序动作分割模型，获取所述目标的动作类别预测结果，其中，所述时序动作分割模型为根据权利要求 1~11 任意一项所述的时序动作分割模型的训练方法生成。

13. 根据权利要求 12 所述的时序动作分割方法，其中，所述方法符合以下至少一项：

所述模态包括惯性、关键点和边界框；或

所述补充数据包括物联网设备的工作日志。

14. 一种时序动作分割模型的训练装置，包括：

特征提取模块，被配置为对样本数据中每个模态的数据进行高维特征提取，获取每个模态的第一特征；

特征约束模块，被配置为根据所述第一特征，通过降维处理，获取中间损失函数，并通过维度恢复获取每个模态的第二特征；

融合模块，被配置为根据所述第二特征和补充特征，通过对不同模态的特征的融合，获取运动表示；和

交互双分支模块，被配置为分别通过类别编码器网络、边界编码器网络处理所述运动表示，获取动作类别预测向量和时序边界预测向量，确定动作类别预测结果，并确定预测损失函数，

其中，所述时序动作分割模型根据所述中间损失函数和所述预测损失函数调整参数，直至完成训练。

15. 一种时序动作分割装置，包括：

数据获取模块，被配置为获取目标的每个模态的数据和补充数据；和

预测模块，被配置为将所述每个模态的数据和所述补充维度数据输入所述时序动作分割模型，获取所述目标的动作类别预测结果，其中，所述时序动作分割模型为根据权利要求 1~11 任意一项所述的时序动作分割模型的训练方法生成。

16. 一种数据处理装置，包括：

存储器；以及

耦接至所述存储器的处理器，所述处理器被配置为基于存储在所述存储器的指令执行如权利要求 1 至 13 任一项所述的方法。

17. 一种计算机可读存储介质，其上存储有计算机程序指令，该指令被处理器执行时实现权利要求 1 至 13 任意一项所述的方法的步骤。

18. 一种计算机程序产品，包括计算机程序或指令，所述计算机程序或指令被处理器执行时实现权利要求 1 至 13 任意一项所述的方法。

19. 一种计算机程序，用于使处理器执行权利要求 1 至 13 任意一项所述的方法。