

東北大學



NORTHEASTERN
UNIVERSITY

硕士学位论文

THESIS FOR MASTER'S DEGREE

论文题目 基于对比学习的人体活动识别预训练模型
研究与应用

作 者	郭世哲
学 号	1971609
学 院(部)	计算机科学与工程
专 业	计算机技术
指导教师	焦明海 副教授

2022 年 7 月

分类号_____密级_____

UDC_____

学 位 论 文

基于对比学习的人体活动识别 预训练模型研究与实现

作 者 姓 名： 郭世哲

指 导 教 师： 焦明海 副教授

东北大学计算机科学与工程学院

申请学位级别： 硕士 学 科 类 别： 专业学位

学科专业名称： 计算机技术

论文提交日期： 2022 年 5 月 论文答辩日期： 2022 年 6 月

学位授予日期： 2022 年 7 月 答辩委员会主席： 易秀双 教授

评 阅 人： 盲审专家、徐彬 副教授

东 北 大 学

2022 年 7 月

A Thesis in Computer Technology

Research and Implementation of Pre-training Model for Human Activity Recognition Based on Contrastive Learning

By Guo Shizhe

Supervisor: Associate Professor Jiao Minghai

Northeastern University

May 2022

独创性声明

本人声明，所呈交的学位论文是在导师的指导下完成的。论文中取得的研究成果除加以标注和致谢的地方外，不包含其他人已经发表或撰写过的研究成果，也不包括本人为获得其他学位而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

学位论文作者签名：郭世哲

日 期： 2022 年 6 月 15 日

学位论文版权使用授权书

本学位论文作者和指导教师完全了解东北大学有关保留、使用学位论文的规定：即学校有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许论文被查阅和借阅。本人同意东北大学可以将学位论文的全部或部分内容编入有关数据库进行检索、交流。

作者和导师同意网上交流的时间为作者获得学位后：

半年 ☐ 一年 ☐ 一年半 ☐ 两年 ☒

学位论文作者签名：郭世哲

导师签名：焦明海

签字日期：2022 年 6 月 15 日

签字日期：2022 年 6 月 15 日

摘要

随着传感器技术与深度学习技术的发展,基于可穿戴设备的人体活动识别受到了越来越多的关注,在工业生产、健康监测、医疗保健和人机交互等多个领域有着广泛的应用。目前深度学习方法在活动识别任务上已经取得了不错的效果,但在实际的数据采集过程中,由于打标签等操作成本昂贵,高质量标签活动识别数据集规模普遍较小,一定程度上限制了有监督方法的效果。

本文尝试使用无监督方法,从无标签数据中学习数据分布。对比学习作为近两年提出的自监督表示学习方法,在图像识别和自然语言处理领域获得了巨大的成功。本文在将对比学习应用到人体活动识别领域时遇到以下问题:使用原始的无标签时间序列难以构造对比学习中正负样本对:(1)由于人体动作的连续性,无法直接选取给定样本的负样本;(2)由于人体动作的周期性,给定样本的正样本趋于同质化,不利于模型将样本均匀地映射到特征空间中。针对上述问题,本文开展了以下研究:

(1)针对人体动作的连续性问题,本文提出了基于对比学习的时间序列变化点检测模型 CL-CPD。针对变化点检测这个二分类问题,提出了简化的正负样本对构造方法,并提出适用于传感器数据的跨通道域注意力,最后使用空洞卷积和多尺度卷积解决不同动作周期长度不同的问题。实验结果表明,该方法相较基于统计的方法具有更好的普适性,且检测准确率高于主流的深度学习方法。

(2)针对人体动作的周期性问题,本文提出了基于 Transformer 的时间序列数据增强模型 TGAN。针对 Transformer 的解码问题,本文提出了一种适用于数据增强任务的对角 mask 方法,可以更有效的利用双向的序列信息。通过将 Transformer 与 GAN 结合,可以生成多样性的样本。与基于 DeepConvLSTM 构造的 RGAN 相比, TGAN 具有结构简单,训练速度快的优点。实验结果表明, TGAN 的数据增强效果不输于 RGAN,且具有更少的参数和更快的训练速度。与传统的基于统计的方法相比, TGAN 不需要手动设定参数,并可以通过调整隐向量生成多样化的增强样本。

(3)本文搭建了基于对比学习的活动识别预训练模型 CL-HARP,融合了上述研究内容,并提出“多对多”对比损失函数,使模型可以更高效的利用每一个样本。通过降维算法验证了预训练模型特征提取方法的有效性。实验结果表明,在将其应用于下游分类任务时,使用简单的全连接层获得了优于监督学习方法的准确率。

关键词: 人体活动识别; 对比学习; 变化点检测; 数据增强; 预训练模型

摘要

Abstract

With the development of sensor technology and deep learning technology, human activity recognition based on wearable devices has received more and more attention, and has a wide range of applications in many fields such as industrial production, health monitoring, healthcare and human-computer interaction. At present, deep learning methods have achieved good results on activity recognition tasks, but in the actual data collection process, the size of high-quality labeled activity recognition datasets is generally small due to expensive operations such as labeling, which limits the effectiveness of supervised methods to some extent.

In this study, we try to use an unsupervised method to learn data distribution from unlabeled data. Contrastive learning, a self-supervised representation learning method proposed in the last two years, has been a great success in the fields of image recognition and natural language processing. In this paper, we encounter the following problems when applying contrastive learning to the field of human activity recognition: it is difficult to construct positive and negative sample pairs in contrastive learning using the original unlabeled time series: (1) Due to the continuity of human actions, the negative samples of a given sample cannot be directly selected; (2) due to the periodicity of human actions, the positive samples of a given sample tend to be homogeneous, which is not conducive to the model to map the samples uniformly into the feature space. To address the above problems, the following studies are conducted in this paper.

(1) To address the continuity problem of human actions, this paper proposes CL-CPD, a time-series change-point detection algorithm based on contrastive learning. for change-point detection, a simplified positive and negative sample pair construction method is proposed for this binary classification problem, and cross-channel domain attention applicable to sensor data is proposed, and finally the problem of different action cycle lengths is solved using null convolution and multi-scale convolution. The experimental results show that the method has better generalizability compared with the statistical-based methods and the detection accuracy is higher than the mainstream deep learning methods.

(2) To address the periodicity of human actions, this paper proposes a Transformer-based time-series data enhancement model TGAN. for the decoding problem of Transformer, this paper proposes a diagonal mask method for data enhancement tasks, which can more effectively utilize the bi-directional sequence information. By combining Transformer with GAN, diverse samples can be generated. Compared with the RGAN constructed based on DeepConvLSTM, TGAN has the advantages of simple structure and fast training speed. Experimental results show that TGAN does not lose the data enhancement effect than RGAN, and has fewer parameters and faster training speed. Compared with traditional statistical-based methods, TGAN does not require manual parameter setting and can generate diverse enhancement samples by adjusting the hidden vector.

(3) In this study, we build a pre-training model CL-HARP for activity recognition based on contrastive learning, which incorporates the above research and proposes a “many-to-many” contrastive loss function so that the model can utilize each sample more efficiently. The effectiveness of the pre-trained model feature extraction method is verified by a dimensionality reduction algorithm. Experimental results show that the accuracy better than that of supervised learning methods is obtained using a simple fully connected layer when applying it to downstream classification tasks.

Keywords: human activity recognition; contrastive learning; change point detection; data augmentation; pre-training model

目 录

独创性声明	I
摘 要	II
Abstract	V
目 录	V
第 1 章 绪 论	1
1.1 研究背景	1
1.2 研究目的与意义	2
1.3 本文的主要工作	3
1.4 本文组织结构	4
第 2 章 相关工作概述	5
2.1 相关技术	5
2.1.1 激活函数	5
2.1.2 卷积神经网络	6
2.1.3 循环神经网络	8
2.1.4 注意力机制	9
2.1.5 预测结果评估标准	10
2.2 相关研究	11
2.2.1 可穿戴传感器数据	11
2.2.2 基于深度学习的人体活动识别方法	12
2.2.3 对比学习的相关研究	12
2.3 本章小结	18
第 3 章 基于对比学习的时间序列 CPD 模型	19
3.1 问题分析与提出	19
3.2 基于对比学习的时间序列 CPD 算法	20
3.2.1 使用对比学习解决 CPD 问题的可行性	20

3.2.2 简化的正负样本对构造方法	21
3.2.3 多尺度卷积网络.....	23
3.2.4 跨通道注意力模块.....	25
3.2.5 损失函数.....	26
3.2.6 模型结构.....	27
3.3 实验分析	27
3.3.1 数据集.....	27
3.3.2 数据预处理.....	28
3.3.3 数据分割.....	28
3.3.4 实验结果分析.....	30
3.4 本章小结	34
第 4 章 基于 Transformer 的人体活动数据增强模型	35
4.1 问题分析与提出	35
4.1.1 数据增强.....	35
4.1.2 对比学习正负样本对构建的正样本选取	35
4.2 基于 Transformer 的时间序列数据增强算法	37
4.2.1 技术选型.....	37
4.2.2 模型结构.....	37
4.2.3 对角 mask.....	40
4.2.4 损失函数.....	41
4.3 实验分析	41
4.3.1 数据集.....	41
4.3.2 实验环境配置.....	42
4.3.3 实验结果及分析.....	42
4.4 小结	46
第 5 章 基于对比学习的人体活动识别预训练模型	47
5.1 问题分析与提出	47
5.2 模型框架	49
5.2.1 变化点检测模块.....	50
5.2.2 数据增强模块.....	50
5.2.3 对比学习预训练模块	50

5.2.4 活动识别模块.....	51
5.3 系统功能与架构.....	51
5.3.1 需求分析与功能.....	51
5.3.2 系统架构.....	53
5.3.3 系统的模块划分.....	53
5.4 系统的设计实现.....	54
5.4.1 数据管理模块.....	54
5.4.2 预训练模块.....	56
5.4.3 活动识别模块.....	58
5.5 实验分析	58
5.5.1 数据集.....	58
5.5.2 实验结果.....	60
5.6 小结	63
第 6 章 总结与展望	65
6.1 总结	65
6.2 展望	66
参考文献	67
致 谢	73
攻读硕士学位期间研究成果	75

第一章 绪论	1
第二章 相关理论	10
第三章 系统架构设计	20
第四章 系统实现	30
第五章 实验结果与分析	40
第六章 结论与展望	50
参考文献	60
附录	70

第1章 绪论

1.1 研究背景

近年来,随着微处理器和物联网的发展,传感器和电子设备因其高性能、小尺寸、长时间续航等优点得到了越来越多的关注,在工业生产^[1]、健康监测^[2,10,11,54]、医疗保健^[3,9]和人机交互^[4]等领域有着广泛的应用。特别是随着近几年来智能家居,智能手机、手环等普及,在日常生活中对人体活动进行识别的需求也在逐渐加大,因此有越来越多的研究者投入到基于传感器的活动识别这一问题中。

基于传感器的活动识别与传感器的位置有很大关系。总体上可以分成基于周围环境传感器的活动识别和基于可穿戴传感器的活动识别。在基于周围环境传感器的方式中,传感器被放置在实验环境中的各个物体上。活动识别的效果完全由这些与物体决定。智能家居^[5-8]就是使用外部传感器进行活动识别的典型例子:通过将活动数据与具体的物体绑定,可以获得大量的先验知识,如当冰箱门上的传感器数据发生变化时,用户在开关冰箱门的概率极大,因而识别的准确率很高。

尽管基于外部传感器的活动识别准确率极高,但这种基于外部传感器的数据的活动识别是一种被动识别,完全依赖于用户与物体的交互(例如水杯、门把手和餐具等)。我们只能通过这些物品的传感器的数据来记录用户使用了该物体做了某些活动,进而推测用户可能在做什么样的动作。如果因为一些意外情况,用户无法与物体进行交互,那么就无法采集到任何数据。用户的动作被限制在“使用某物品”这一范围内,而很难对用户本身的动作做出识别,因此极大地限制了使用场景。除此之外,这种基于外部传感器的活动识别成本通常较高。

相比之下,基于可穿戴传感器的活动识别作为一种主动识别,可以在不受活动范围制约、便携且可以保护用户的隐私的前提下实时采集活动数据。但这也为活动识别带来了挑战,因为一个完整动作可能由多个基础动作组成,而不同动作的基础动作可能很相似,这种相似性使得我们难将一种动作与另一种动作区分开来。针对这一具有挑战性的问题,已经有很多研究者投入其中。本文的后续研究内容均为基于可穿戴传感器的人体活动识别。综述^[12]中提及了如何降低数据采集成本的问题,这里的采集成本不仅仅指采集,想要将采集到的数据应用于实际问题中,还需要额外的人工辅助来对这些数据进行标注。现阶段,想要将大规模地带标签活动数据应用于人体活动识

别的研究，主要面临以下两个问题：

(1) 大规模高质量标签数据难以获取

即使是在实验室条件下，我们也很难对采集的数据及时地进行标注，往往是在采集完毕后再参考视频等对原始数据进行标注。而且由于每个人的行为习惯不同，很难找到一种通用的标注方法。所以常用的采集方式是预设一组动作让受试者来做，这很大程度上限制了高质量标签数据集的规模^[13]。

(2) 缺少对基于传感器的人体活动数据的数据增强方法

数据增强是一种常用的扩充样本的方法。以图像识别问题为例，我们对原始图片进行裁剪，调整对比度或增加随机噪声等，并给新生成的图片打上和原始图片打上同样的标签，这样就扩充了训练样本集。在原始数据集较小时，数据增强可以扩充训练样本的分布，有效地解决过拟合问题。但由于人体活动数据的特殊性，传统的时间序列增强方法生成的样本并不能很好应用于活动识别领域，生成的样本相对单调，且针对不同的数据集需要手动设定参数。因此需要一种数据增强方法来生成多样化且真实的活动识别数据。

1.2 研究目的与意义

对比学习作为一种表示学习方法，可以用来学习数据分布并提取特征，在深度学习方法里，通常使用表示学习方法来搭建预训练模型。预训练模型作为一个重要的研究课题，在很多领域都有非常广泛的应用。通过事先在大量的无标签数据上完成模型预训练，往往可以更好的让模型学习到数据的分布^[14]。BERT^[17]作为 NLP 领域最成功的模型就是预训练模型的典型应用。当预训练模型足够好时，在预训练模型的基础上搭建新的分类器或者预测器，往往可以取得比直接在标签数据进行模型训练更好的效果^[18]。鉴于对比学习在 CV 和 NLP 领域取得了巨大成功^[16]，本文尝试搭建基于对比学习方法适用于活动识别领域的预训练模型。

本文的研究目的主要是解决搭建基于对比学习的活动识别预训练模型过程遇到两个问题，并最终搭建基于对比学习的预训练模型：

一是针对现有对比学习中负样本选取方法不适用于活动识别领域的问题，设计变化点检测算法来分割原始序列，协助后续对比学习框架中负样本的选取；二是现有的时间序列数据增强算法不适用于活动识别领域，在其生成的样本上构建的样本的难以应用到对比学习中。本文提出基于注意力机制的数据增强方法，协助后续对比学习框架中正样本的选取，而数据增强算法也可应用于其他时间序列。

在解决上述两个问题后，最终搭建完整的基于对比学习的活动识别预训练模型，以实现在无标签数据上完成特征提取，并将预训练模型应用于下游分类任务。

目前自监督学习方法和预训练模型并未在活动识别领域有深入的研究和应用。本文基于以上研究设计并实现了人体活动识别预训练模型，并在最终的识别任务上取得了优于监督方法的准确率，旨在更有效地利用无标签数据，并降低对高质量标签活动数据集的需求，探究将自监督学习方法应用于活动识别领域的可行性。

1.3 本文的主要工作

本文工作包含四个主要部分：

(1) 基于对比学习的时间序列变化点检测模型 CL-CPD

本文首先提出了基于对比学习的时间序列变化点检测的方法，将对比学习引入了时序分割问题，并提出了简化的正负样本对构造算法。考虑到不同位置传感器在不同活动中的对识别效果的影响不同的情况，提出适用于传感器数据的跨通道注意力，让网络可以自动学习到不同传感器、不同轴的特征对识别结果的影响权重。

在时序分割任务中，采用基于变化点检测的方法。对比学习的正负样本选择时，为了简化训练过程，对一次采样的结果进行多次复用。考虑到接下来的聚类任务，这一阶段更关注预测结果为假阴性的样本，它们对后续构建正负样本对影响较大。

在具体的分割模型中，引入了注意力机制，将空间域的注意力引入到时间序列中，提出了跨通道域注意力，学习不同传感器的位置信息；并提出一维空洞卷积，配合多尺度卷积搭建了多尺度的 CNN 的网络框架，解决不同动作周期不同的问题。

(2) 基于 Transformer 的时间序列数据增强模型 TGAN

本文提出了一种基于注意力机制的时间序列数据增强生成模型，基于该模型方法生成的数据，相较于传统的时间序列数据增强方法生成的数据，有更自然，更多样性的优点，而且避免了复杂的超参数的设定。其具有现有主流时间序列数据增强的效果，并且训练速度更快和模型更小的优点。针对 Transformer 的解码问题，本文提出了对角 mask，更适用于数据增强任务，且可以更好地利用样本的双向信息。本文设计的数据增强算法一定程度上解决了样本不足的问题，而且模型中使用原始序列进行增强样本生成的特性也利于后续对比学习框架中正负样本对的构造。

(3) 基于对比学习的活动识别预训练模型 CL-HARP

本文设计了完整的基于对比学习的活动识别预训练模型，结合上述研究内容解决了现有框架下活动数据正负样本对构建难的问题，并提出了多对多的损失函数，可

以更高效地利用每一个样本。在活动识别领域，有监督方法，尤其是深度学习方法常常受制于高质量训练数据不足的问题，本文基于自监督学习方法——对比学习，设计并实现了活动识别预训练模型。该预训练模型可以利用无监督数据有效地提取特征。在其基础上训练分类器进行活动识别时，分类准确率较有监督方法有 3%~5% 的提升。

（4）设计人体活动识别系统

本文基于上述研究内容设计了人体活动识别预训练系统。该系统主要有以下功能：对原始活动数据的变化点检测、对原始数据进行数据增强、搭建预训练模型以及活动识别。

1.4 本文组织结构

本章主要介绍了本文的研究背景，说明了本文的研究目的与意义，随后总结了本文的主要工作，并给出了本文的组织结构，本文后续章节的具体组织结构如下。

第 2 章介绍了本文的相关技术和研究。首先介绍了一些本文用到的神经网络模型和实验结果评价标准；其次介绍了活动识别领域的主流深度学习算法；最后对比学习的原理、主流应用框架和其中的技术细节进行阐述和总结。

第 3 章介绍了基于对比学习的时间序列变化点检测模型。首先介绍了本文提出的基于对比学习的时间序列变化点检测算法 CL-CPD，对其中正负样本对构建方法进行了改进，并在具体的模型上使用了空洞卷积、多尺度卷积和跨通道注意力；然后介绍了本文实验所使用的数据集和传感器数据预处理方法；最后设计了对比实验验证了模型的有效性。

第 4 章介绍了基于 Transformer 的人体活动数据增强模型。首先针对传统时间序列数据增强算法的局限性、基于 CNN 的生成模型出现模型坍塌和基于 RNN 的模型训练速度较慢的缺点，本文提出了基于 Transformer 的生成模型 TGAN，并对具体的 mask 方式和样本生成方式做了改进；最后设计了对比实验验证了算法的有效性，并对不同的数据增强方式和不同分类器准确率之间的关系进行简要的分析。

第 5 章介绍了基于对比学习的人体活动识别预训练模型 CL-HARP。首先利用第 3,4 章的算法解决了对比学习应用于活动识别领域中正负样本的选择难的问题，并提出“多对多”的损失函数，可以更高效率的利用每一个样本；然后搭建了完整的预训练模型，并对系统各模块进行了设计；最后通过对比实验验证了在预训练模型在下游分类任务上的效果。

第 6 章总结了全文的工作，并对工作中的一些不足和可以改进的部分进行展望。

第2章 相关工作概述

本章首先在 2.1 节介绍了常用的激活函数、深度学习模型和注意力机制；然后在 2.2 节介绍了深度学习方法在活动识别领域的研究，并对比学习的原理、主流框架及技术细节进行阐述。

2.1 相关技术

本节首先介绍了激活函数这一神经网络中的基本组成；然后由于多轴传感器数据同时具有规律的拓扑结构和时间序列的特征，本节介绍了卷积神经网络、循环神经网络和注意力机制；最后介绍分类结果的评价标准。

2.1.1 激活函数

全连接层本身只能实现线性变换，即使有多层全连接层，其本质与一层全连接层是一样的。在对实际问题的函数进行拟合时，仅靠线性变换往往难以满足要求。为此，我们需要引入非线性变换。在深度学习中，通常使用激活函数的方式实现非线性变换。引入激活函数后，理论上我们可以使用神经网络来逼近任意一个函数。

除了非线性变换，激活函数还应该具有可微属性或至少是部分可微的，在对神经网络中的参数进行求解时，我们需要求出每个环节的梯度来进行反向传播。当下最为常用的激活函数为 ReLU，它的主要优势是计算简单，如图 2.1 所示。

在加入激活函数后，由公式（2.1）就可以得到：

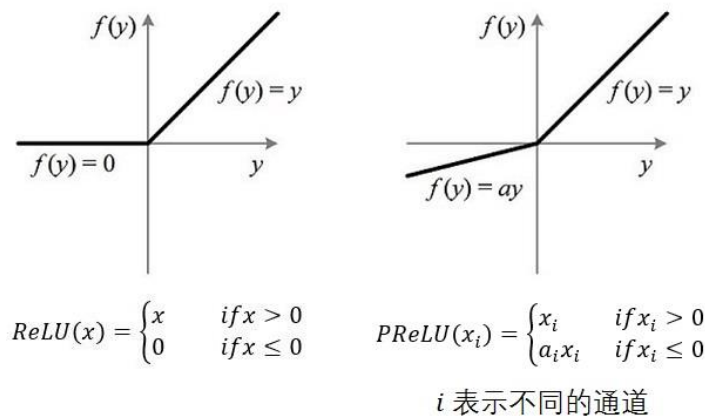


图 2.1 ReLU（左）和 PReLU（右）激活函数

Fig. 2.1 ReLU(left) and PReLU(right) activation function

$$a = f(z) = f(x \cdot w + b) \quad (2.1)$$

其中 x 为输入特征, w 为参数, b 为偏置项。

在 ReLU 的基础上, 何凯明^[19]等人提出了 PReLU, 即带参数的 ReLU。如图 2.1 所示, 当 $a_i = 0$, PReLU 退化为 ReLU。PReLU 在仅增加了一个参数 a_i 的情况下增加了激活函数的表达能力, 在深度学习模型中有广泛的应用。在更新参数时使用带动量的更新, 如公式 (2.2) 所示:

$$a_{i+1} = \mu a_i + \epsilon \frac{\partial \epsilon}{\partial a_i} \quad (2.2)$$

2.1.2 卷积神经网络

卷积神经网络 (CNN) 常用来处理图片数据, 这是因为图片本身具有非常规律的数据结构, 卷积网络可以很好的从中提取特征。以图像分类为例, 利用图像的平移不变性, CNN 通过引入卷积操作, 并复用卷积核, 大大减少了参数数量, 在提取了信息的同时, 并没有过多增加计算量, 可以很简单地实现降维等操作。而活动数据同样具有规律的拓扑结构, 同样可以用卷积神经网络来提取特征。

卷积神经网络的核心构成是卷积层和池化层, 这两层用来提取并过滤信息。最后使用全连接层来输出最终的结果, 本节重点介绍前两层。

2.1.2.1 卷积层

卷积层是卷积神经网络中最重要的组成部分, 其主要功能是从输入数据中提取特征, 在介绍卷积层之前先介绍以下几个参数:

(1) 卷积核的数量 K : 这个参数决定了下一层输入数据的维度, 可以通过增加卷积核的个数来对数据进行增维, 也可以通过减少卷积核的个数来对数据进行降维。

(2) 卷积核的尺寸 (C_i, F_w, F_h) : C_i 表示卷积核的输入通道, F_w 表示卷积核的宽, F_h 表示卷积核的高。每次卷积操作只关注当前局部的数据。

(3) 步长 (S_w, S_h) : 步长表示两次卷积操作之间的距离, S_w 和 S_h 分别表示在横向和纵向上的移动距离。

(4) 零填充: 零填充是指在原始输入数据和特征图的两端用 0 进行填充, 在经过卷积操作后, 特征图的尺寸通常会比原始数据小, 我们可以使用零填充来控制特征图的空间尺寸, 一般有 valid padding 和 same padding 两种。

这几个参数控制了每个卷积层输出的特征图大小, 假设输入数据的维度为 $W_i \times H_i \times C_i$, 输出数据的维度为 $W_j \times H_j \times C_j$, 维度的计算方式如公式 (2.3)、(2.4) 和 (2.5)

所示。卷积操作与步长的示意图如图 2.2 所示。

$$W_j = (W_i - F + 2P)/S + 1 \quad (2.3)$$

$$H_j = (H_i - F + 2P)/S + 1 \quad (2.4)$$

$$C_j = K \quad (2.5)$$

卷积层相对全连接层的最大优势就是参数少，我们在构建卷积层时，对于一个给定的卷积核，我们将它作用在该图像的所有通道上，而不是像全连接层那样给每个像素都设定独立的参数。如果我们想要更多维度的特征图，可以设定更多的卷积核。

卷积层的另一个优势就是可以并行计算。对于一个给定的卷积核，通道内各像素点的计算方式是一样的，彼此之间没有时间上的依赖关系，相互独立。这种特性使得我们可以实现 CNN 的并行训练，加快收敛速度。尤其随着 GPU 的发展，并行计算循环可以更简单地得到实现，我们可以搭建更庞大的，堆叠更多层的 CNN 模型。

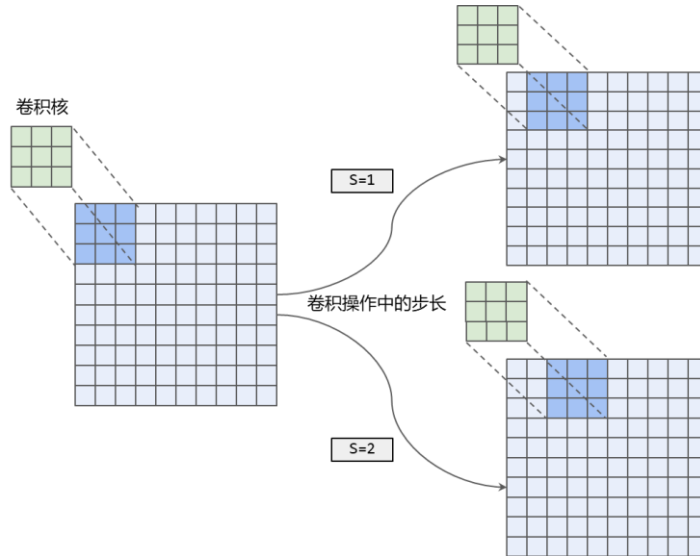


图 2.2 不同步长下的卷积操作

Fig. 2.2 Convolution operations at different stride

2.1.2.2 池化层

池化层（Pooling）是卷积神经网络中的一个重要的组成部分。如果没有池化层只有卷积层，那么在最后的全连接层会有过多的参数导致模型难以训练。类似降采样的作用，池化层可以减小特征图的大小，减少了计算量，同时避免不重要的特征对模型产生干扰。

除此之外池化层本身还有其他作用：增加感受野，池化层的步长越大，增加的感受野就越大，但同时丢失的信息就越多；带来特征的平移不变性和旋转不变性，以最大池化层为例，在每一次计算中，最终结果与最大值所处的位置无关，这是 CNN 网

络最重要的特性之一；池化操作与 **maxout** 激活函数相似，也是一种非线性操作，可以增加 CNN 的表达能力。如图 2.3 是最大池化层操作的示意图，我们按照池化的尺寸对特征图进行划分，并在每个子特征图中取最大值并组成新的特征图。可以看出特征图缩小的比例与步长大小有关。通常情况下，一次卷积操作与一次池化操作一起组成一个模块，这个模块可以堆叠很多层来提取高层次的抽象信息。

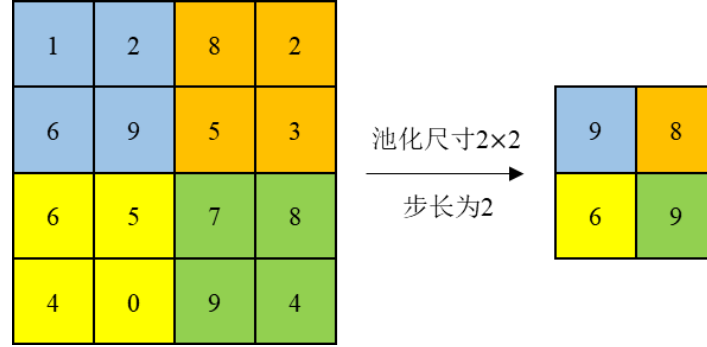


图 2.3 最大池化层运算

Fig. 2.3 Operation of maximum pooling layer

也有一些研究表明，并不是任务都适合池化层，例如在时间序列的异常检测问题中使用池化层会降低检测准确率^[20]，这可能是因为池化层为了减少计算量会忽略掉一些特征。

2.1.3 循环神经网络

循环神经网络 (RNN) 的结构除了包括输入层和输出层，还通过隐藏层对输入数据的时间特性进行建模。RNN 隐藏层的状态与当前所属的时刻有关，对于单向 RNN 模型， t 时刻的输出除了与 t 时刻的输入有关，还与 $t - 1$ 时刻的隐藏层状态有关^[21]。RNN 通过这种方式来构建不同时刻输入数据之间的关系，因此 RNN 模型特别适合用来对序列数据进行建模。RNN 网络的基本结构如图 2.4 所示：

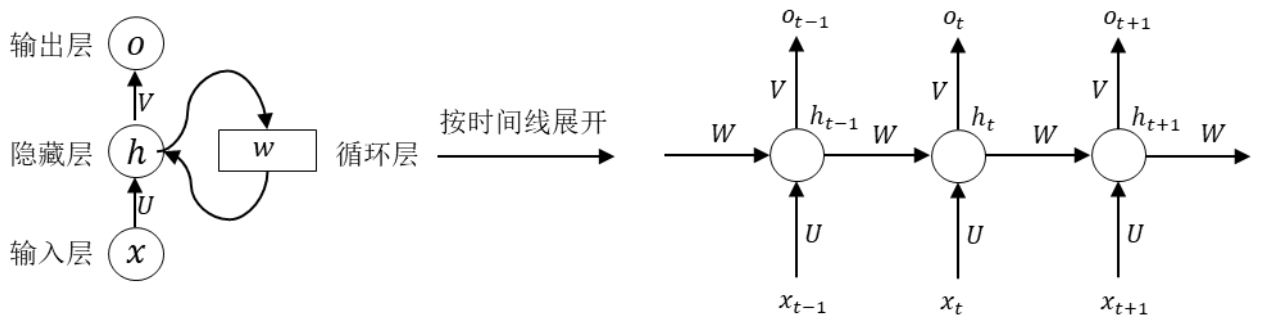


图 2.4 RNN 按时间线展开后的基本结构

Fig. 2.4 Basic structure of RNN expanded by timeline

2.1.3.1 LSTM

RNN 引入隐藏层的使得模型可以学习序列中的时间特征,但是随着序列的增长,RNN 模型容易出现梯度弥散的问题^[22],即 t 时刻的输出可以参考到 $t-1$ 、 $t-2$ 时刻隐藏层的状态,但是随着序列增长,离 t 时刻越远的隐藏层状态越难以将梯度传递过来。针对这个问题,Hochreiter 提出了 LSTM^[23],即长短时记忆网络。LSTM 对 RNN 的隐藏层进行了修改,引入了“门控”的概念,分别是输入门,输出门和遗忘门。其中起到决定作用的是遗忘门,它可以控制隐藏层在 t 时刻之前哪些状态需要被保留,而哪些状态需要被清除。

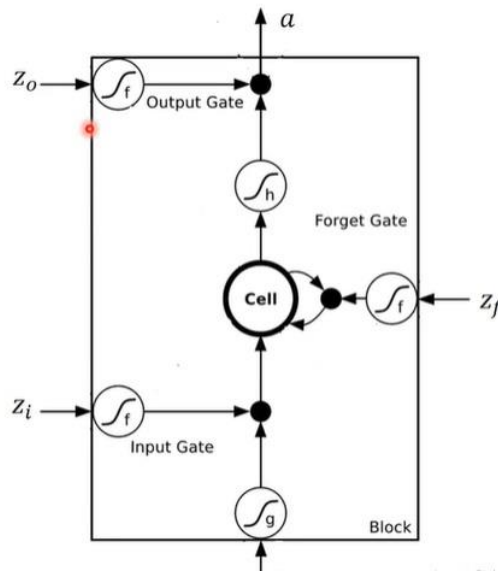


图 2.5 LSTM 结构示意图

Fig. 2.5 Schematic diagram of LSTM structure

尽管引入了遗忘门,但在处理超长序列时,仍然会存在梯度无法传递的问题,而且引入门控概念后,LSTM 的模型变得非常庞大,而且参数非常多难以训练。针对这个问题,Cho^[24]提出了 GRU,简化了门控的概念,精简了模型,但是并没有根本上解决这个问题。

2.1.4 注意力机制

注意力机制在上世纪 90 年代就有人已经提出了,但在 2014 年由 google mind 团队提出的基于注意力机制的 RNN 图像分类模型才让其受到广泛的关注^[25]。注意力模型旨在让模型自己学习所有数据中的注意力部分是哪些,并根据这些注意力部分进行预测或分类。最原始的注意力模型和普通的全连接网络在结构上很相似,但是两者有本质上的不同:在注意力模型中输入数据各维度之间的关系由输入本身决定,而在

普通的全连接层中，输入数据各维度之间的关系已经是确定的，我们能改变的只有输入数据的位置。最原始的注意力模型结构如图 2.6 所示。

与循环神经网络相比，注意力机制有以下优势：一是训练不受方向限制，也有人使用双向 Bi-LSTM^[26]来学习双向信息，但是这种人为引导的训练方式仍会遗漏信息；二是循环神经网络因为在 t 时刻的输出依赖于 $t - 1$ 时刻隐藏层的输出，所以并不能并行计算，这为大规模训练带来了困难，而注意力机制则没有这个限制，计算相对简单很多。针对 NLP 领域的 sequence-to-sequence 任务提出预训练模型 BERT，其中最重要的部分就是搭建了基于注意力机制的 Transformer 用来提取文字序列的前后关系以用于机器翻译问题。

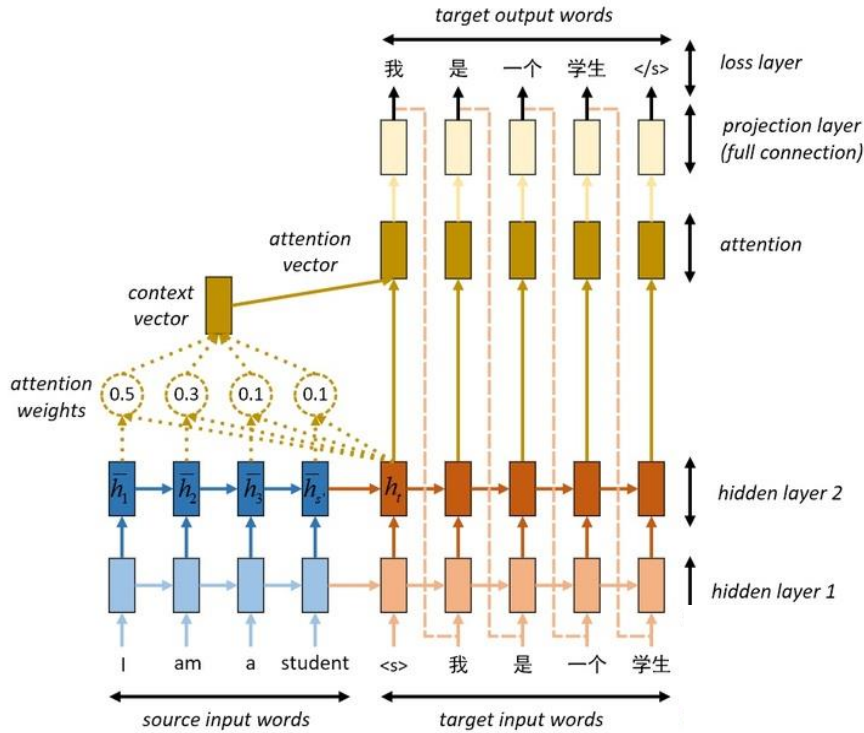


图 2.6 注意力模型结构

Fig. 2.6 Attention Model Structure

2.1.5 预测结果评估标准

2.1.5.1 准确率

变化点检测问题本质上是一个二分类问题，活动识别问题本质上是一个多分类问题。分类问题的评价标准中最经典的就是准确率。如果共有 n 个测试样本，其中正确分类的样本有 m 个，则准确率 a 为：

$$a = \frac{m}{n} \times 100\% \quad (2.6)$$

2.1.5.2 F1-score

F1-score 是衡量分类准确率的另一种指标,是在分类的准确率和召回率的基础上计算得到。相对于准确率, F1-score 更适用于评价各种类别的数量不均衡的样本,计算方式如公式 (2.7) 所示。

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \quad (2.7)$$

其中 *precision* 为准确率, *recall* 为召回率。

2.2 相关研究

本节主要介绍了基于可穿戴传感器活动识别领域和对比学习领域的相关研究。

2.2.1 可穿戴传感器数据

本文主要研究基于可穿戴传感器的活动识别,无论在实验室环境中还是实际的生产生活环境中,都可以使用智能手机手表等来采集数据,包括与动作相关的信息,例如动作的幅度,速度等信息,也包括用户本人的一些信息,例如体温,脉搏,和所处位置的磁场信息,在保护了用户隐私的前提下,对用户的活动影响也较小。对于不同类型的人体活动,使用不同的传感器数据进行识别,得到准确率也不同。目前主流的可穿戴传感器数据主要有三种:

(1) 三轴加速度

通过加速度计获取的三轴加速度是活动识别中最基础的数据,它负责测量目标物体在三维空间的三个坐标轴(x, y, z)上的线加速度,对于一些重复性比较强的动作如:跑步跳跃等的识别准确率较高^[32]。

(2) 三轴角速度

角速度记录的是目标物体在三维空间的三个坐标轴上偏向某一个轴的角速度,通过陀螺仪进行采集,同样分为三个轴。

(3) 磁场

磁力传感器可以用来测量设备所处位置磁场的强度和方向,以此确定目标的方位,它也是沿着三维空间的三个坐标轴进行收集的。在一些场景下,可以利用磁场数据来对设备进行校正,或者根据磁场数据来一定程度上修正设备的偏向。磁力传感器通常与加速度和角速度组合使用, Roggen 等人^[28]研究中公布的 opportunity 数据集就

是由三轴加速度、三轴角速度和磁场组成的。

2.2.2 基于深度学习的人体活动识别方法

深度学习模型因其强大的表达能力,被广泛应用到各个领域。针对活动识别这一分类问题,很多研究者利用深度学习方法提出了有效的方法,本小节对该领域比较经典的深度学习方法进行简要介绍。

Ronao 和 Cho^[29]提出将传感器信号转换成二维序列,并用处理图像的方式来处理这些序列的方法,使用 CNN 模型来对六种 Locomotion 动作进行识别,并通过对比实验证明了 CNN 模型在分类准确率优于 MLP, Naive Bayes 和 SVM 等传统方法。

Hughes 和 Correll^[30]通过构建分布式的 CNN 模型来对中层次的动作进行建模,如:开门,打开抽屉等,并对传感器在人体不同部位对最终结果产生的影响进行了论证。

FJ Ordóñez 等人^[31]在 CNN 和 LSTM 的基础上设计了 DeepConvLSTM 框架,他有如下优势:(1)适合多维可穿戴传感器(2)可以实现传感器数据融合(3)不需要专业知识来设计特征(4)可以显式地对激活特征的时间特性进行建模。因其优秀的建模性能,DeepConvLSTM 已经成为目前活动识别领域的主流基准测试方法。

Murad 和 Pyun^[32]设计了有多个 LSTM 组成的深度 RNN 模型,证明了无方向的 DRNN 模型优于双向 DRNN 模型和无向 DRNN 模型。

Gu 等人^[33]基于叠加深层去噪自编码器搭建了活动识别模型,可以结合来自多传感器数据一起进行建模,并且同样避免了手动设计特征。

尽管这些模型在各自的问题上取得了不错的成绩,但是这些方法都是基于有标签数据设计的深度学习模型,针对无标签数据的无监督方法的研究,如无监督嵌入学习^[34]则相对较少。

对比学习作为近两年提出的自监督学习方法,并没有在活动识别领域有深入的研究。本文基于这一现状,尝试将对比学习方法应用于活动识别领域。

2.2.3 对比学习的相关研究

由于有监督方法对于高质量标签数据的要求很高,我们考虑使用无监督方法处理活动识别任务。对比学习是在无监督学习的概念上提出的表示学习方法,并提出了“自监督学习”的概念,即不依赖标注数据,要从无标签数据中自己提取特征,从无监数据中构建出“伪标签”以实现“自监督”。表示学习的目标是为输入 x 学习一个表示 z ,最好的情况就是知道 z 就能知道 x 。从隐空间向量出发对原始数据进行生成,这

就是对比学习的第一大类：生成式自监督学习。比如还原文本中被掩盖的单词，或者还原图像中被掩盖的像素。变分自编码器^[53]与对抗生成网络^[34]是两种典型的生成式自监督学习方法，这类型的任务需要模型对原始数据的每个像素进行重构，具有一定的难度。但这种生成模型的前提需要假设被掩盖掉的元素是相互独立的，这在实际的训练时是无法实现的。

另一方面研究者从人的直觉出发，当我们在脑海重构某种图像时，并不关心细粒度的特征，而是通过更高维度的特征来给一张图片打标签，这就引出了无监督表示学习的另一种做法：判别式自监督学习。对比学习就属于判别式自监督学习的一种，判别式的输出通常是真/伪或者正/负，这种任务简单一些。其本质是：通过构造相似样本对（正样本）和不相似样本对（负样本），学习到一个表示学习模型，这个模型将正样本对映射到一起，而将负样本对的映射分开。

对比学习利用输入 x 到表示 z 过程中信息量的损失来衡量一种表示的好坏，以最大化互信息作为目标进行优化，就可以得到对比学习的损失函数。如图 2.7 所示，在特征空间里，我们希望正样本对之间的距离较近，负样本对之间的距离较远，如果可以定义一个合适的损失函数，那么这种映射就可以实现。

Chopra^[50]最早在 2005 年提出了有监督条件下的对比损失。给定输入样本集合 $\{x_i\}$ 和对应的标签 $y_i \in \{1, \dots, L\}$ ，映射函数 f_θ 的损失函数为：

$$L_{cont}(x_i, x_j, \theta) = 1[y_i = y_j] \|f_\theta(x_i) - f_\theta(x_j)\|_2^2 + 1[y_i \neq y_j] \max(0, \epsilon - \|f_\theta(x_i) - f_\theta(x_j)\|_2)^2 \quad (2.8)$$

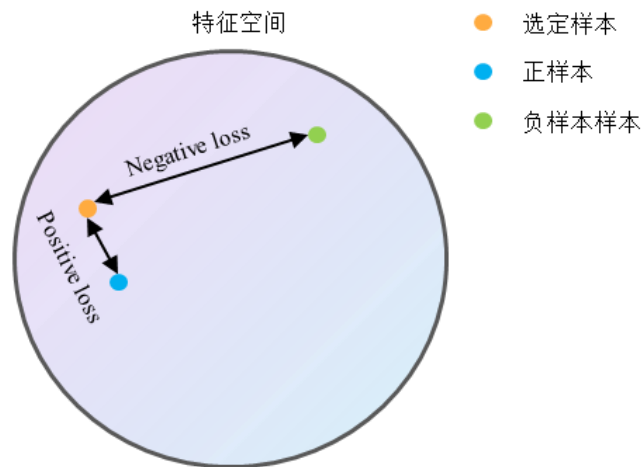


图 2.7 对比学习中正负样本损失

Fig. 2.7 Positive and negative sample loss in contrastive learning

在 Hinton 提出对比学习方法后，CV 和 NLP 领域均利用对比学习的获得了巨大的成功，在一些任务上已经超过了有监督模型。对比学习的核心任务是正负样本对的构建，针对这一问题，有很多研究提出了不同的优化方法。

2.2.3.1 MoCo

对比学习是在正负样本对集合中进行的，如果这个集合中的负例越多，想将正例与负例分开就越难，于是一种思路就是增加这个集合中的负例。由于 GPU 显存容量有限，通常情况下，我们不能将所有的负例映射都保存在 GPU 中。Memory bank 的思想可以解决这个问题，即将所有已经经过模型编码的样本存储在内存中，计算损失时将 memory bank 中存储的编码一起作为负例使用。但是 memory bank 中的编码是之前迭代过程中的编码，而编码器一直在更新，会出现同一批计算损失的样本编码方式不一致的问题，导致模型学习能力变差。有一种解决办法是把当前最新的编码加入 memory bank 中，但是仍然会存在不一致的问题。

He 在 2019 年提出了 MoCo (Momentum Contrast) 模型^[37]，利用 memory bank 的思想，从存储编码结果转向存储编码器，解决同一批样本的编码方式不同问题，首次得到了可以使用对比学习框架。核心思想是用动量编码器来代替不变的 memory bank，每次更新编码器时，一同更新动量编码器，更新比例可以由参数控制。其结构如图 2.8 所示，在方法在 ImageNet 的准确率超过了当时的所有无监督方法。

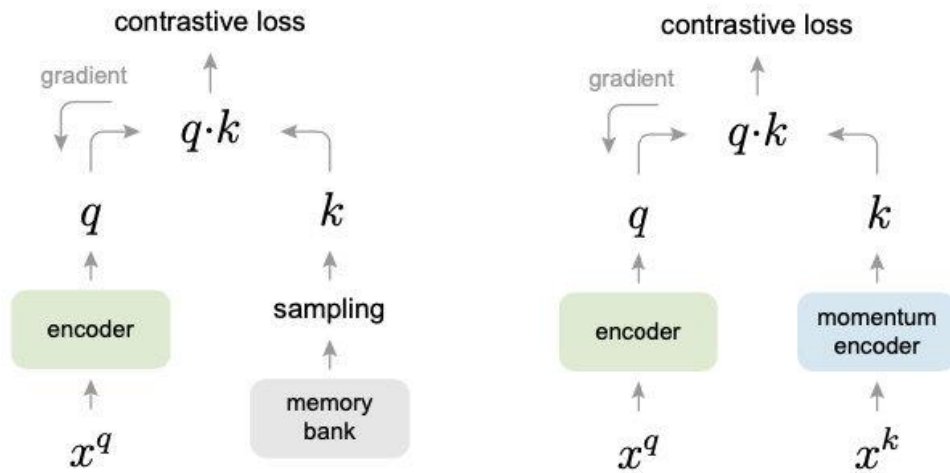


图 2.8 Memory bank (左) 和 MoCo (右) 结构

Fig. 2.8 The structure of memory bank(left) and MoCo(right)

2.2.3.2 SimCLR

SimCLR 是 Chen 在 2020 年提出的模型^[38]，其效果在分类任务上已与监督模型不分上下。MoCo 由于有两个编码层需要训练更新，导致模型非常庞大，对应硬件要求较高，SimCLR 则更关注对比学习的本质，即样本对的构造上。以图像为例，构造正例的方法如下：对于某张图片 x ，随机对其进行两种数据增强操作，如旋转，裁剪等，将得到的两个经过增强的新图像作为正样本对 $\langle x_1, x_2 \rangle$ 。训练时，同一批次内的任意其他样本都可以用来构建负样本对。对比学习模型将这些样本对映射到特征空间上，并在其中减少正样本对之间的距离，增加负样本对之间的距离，使模型过滤掉无用的噪声特征，去学习更通用的，更本质的特征，如：平移、旋转、亮度、剪切不变性等。SimCLR 通过实验验证了，通过融合多种数据增强操作，得到更为复杂的样本，可以增加对比学习模型任务难度，有利于增强对比学习模型的表达能力。

除了数据增强，SimCLR 还在编码层后加上了非线性映射头 $g(h_i) = W^2 \text{ReLU}(W^1 h_i)$ ，研究发现编码层编码后的特征包含了数据增强方式的信息，这些信息对最后的特征分类并没有帮助，通过一个非线性层将这些信息过滤掉，可以让模型学习到的特征表示集中于数据本身。这个非线性层只在搭建无监督预训练模型时使用，如果需要将其应用到后续的任务时，将编码层冻结并在其基础上搭建后续的模型。SimCLR 还改进了负例的选取，在同一 batch 内的所有其他数据均作为负例计算，在没有增加采样频率的同时，扩充了一次训练时的样本对大小。

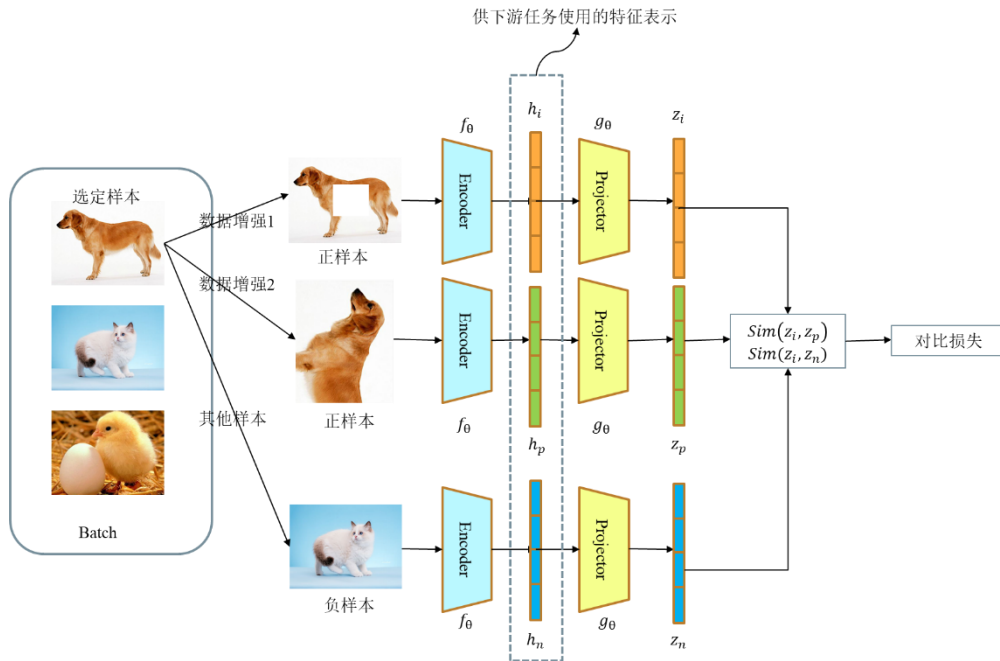


图 2.9 SimCLR 结构示意图

Fig. 2.9 Schematic diagram of SimCLR structure

以上这几点成为了后面其他对比学习模型的标准构成部分，在本文搭建的基于对比学习预训练模型中，也是从这三点出发提出针对于活动识别领域的改进。

2.2.3.3 BYOL

Grill 等人提出了的 BYOL 模型^[39]，提供了不同于前两个模型的思路。Tian 对图像中的 view 的含义做了阐述^[15]，比起视图，view 的概念更接近“观察一个物体的角度”，如果模型能通过一种角度去预测该物体在其他角度的成像，说明模型从不同数据中学习到了相同的本质。在表示学习中，现有的框架都是从样本的一个“视角”的表示去预测同一样本在其他“视角”下的表示。但是只关注这一点可能会出现模型坍塌的问题，即学到的所有表示都是相同的。对比学习的输出是对样本真伪的判别，因为每个的样本都是不同的，所以不能总判断样本是真实还是合成的。这使得模型不再输出单一的结果，从而避免模型坍塌。

而 BYOL 的作者则考虑在完全不使用负例的前提下学习到好的表示。该模型发现使用完全随机初始化的编码器也可达到 18.8% 的准确率，为了得到更好的编码器，该模型引入 MoCo 中的动量编码器，将两个编码器分为在线编码器（更新梯度）和目标编码器（不更新梯度）。该模型的优化目标是使用在线编码器来预测目标编码器的表示，即用一种表示去预测另一种表示。

尽管如此有些研究者中发现 BYOL 的本质还是使用“对比”，并在实验角度验证了 Batch Normalization 在对比学习中的重要意义，这也是一种隐式的对比，过滤掉该 batch 中样本之间相同的特征，留下不同的特征，说明即使在非对比学习框架中，也有对比的思想。

2.2.3.4 其他细节

1. 对比学习学习到的特征

Wang^[40]提出对特征的映射提出应满足要求：“对齐性（Alignment）”和“均匀性（Uniformly）”。“对齐性”指的是正样本对在单位超球面上的映射，应该有接近的特征，即两者在超球面上距离比较近；“均匀性”的是模型应该尽可能的保留原始数据的信息，这意味着模型需要将各个类别的映射均匀地分布在超球面上。如图 2.10 所示，各类别的特征应在不相互重叠前提下，尽量铺满整个超球面。

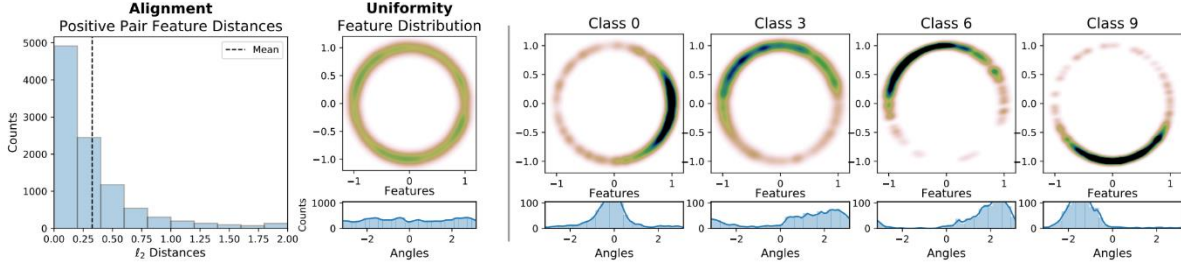


图 2.10 超球面上的特征映射应具有“均匀性”的特性

Fig. 2.10 The feature on the hypersphere should have the property of "uniformity"

2.模型坍塌

模型坍塌问题并不只存在于对比学习领域，无论是生成式的模型，还是判别式的模型，都会受到这个问题的困扰。以 GAN 模型为例，在模型训练过程中，目标映射不在神经网络可表示的泛函空间中，会导致收敛困难^[41]。以生成图像为例，模型可能倾向于生成一个最简单不易被误判的结果，尽管这个结果是正确的，但是失去了多样性，所有输入在特征空间的映射都集中到一个点上。这显然与利用 GAN 模型进行数据增强的初衷相悖。

3.对比聚类

Caron 等人在对比学习模型的训练过程中引入了聚类操作，提出 SwAV 模型^[42]。该模型也是上下分支对称结构，其中的图像增强、编码层以及映射头，与 SimCLR 基本保持一致。对于某个批次内某张图像 x 为例，将其增强后 $\langle x_i, x_j \rangle$ 映射到单位超球面中的对应 $\langle z_i, z_j \rangle$ ，之后对这一个批次内的样本进行对比聚类。假设其中一个分支的 x_j 聚类结果为 q_j ，则 SwAV 要求表示学习模型根据 x_i 预测 x_j 所在的类，即要将 z_i 也分到 q_j 类，具体损失函数为 z_i 和原型中每个类中心向量的交叉熵如公式(2.9)和(2.10)所示：

$$L_{aug1}(z_i, q_j) = -\sum_k q_j^k \cdot \log(p_i^k) \quad (2.9)$$

$$p_i^k = \frac{\exp(\frac{z_i^T c_k}{\tau})}{\sum_{k''} \exp(\frac{z_i^T c_{k''}}{\tau})} \quad (2.10)$$

其中 c_k 为第 k 个类的中心向量， τ 为温度超参数。综上，SwAV 的损失函数是两个分支损失之和，如公式(2.11)所示：

$$L_{SwAV} = L_{aug1}(z_i, q_j) + L_{aug2}(z_j, q_i) \quad (2.11)$$

在对比聚类模型中并没有明确地使用负例。但其核心思想与直接采用负例的对比学习模型是一样的，使用同一批的其他样本来约束模型起到负例的作用，也可以起

到防止模型坍塌的作用。在损失函数的分母中，不为同一类其他投影点 z_j 作为负例，使得不同类的投影在特征空间的距离变远，让其更“均匀的”铺满整个特征空间。如果这些样本的映射过于集中，模型的判定会趋于简单化，生成正确但单调的样本。

还有一些方法是在图片的特征矩阵上同时做实例级别的对比任务和聚类级别的对比任务。有研究者提出“标签即表示”的思想，将实例映射到维度与聚类数量相同的特征空间中，这样特征矩阵的行可以看作每一实例属于各个聚类的概率，而列则代表每一聚类在各个实例上的分布。这样做可以将聚类预测作为一个特殊的表示学习任务，从其他任务中独立出来。

总体上，对比学习作为一种较新的表示学习方法，目前只在 CV 和 NLP 领域广泛的应用，而在其在活动识别领域的应用尚未有深入研究。对比学习的几个关键问题在于如何构造相似实例，不相似实例，以及如何更好地构造对比学习损失函数。

2.3 本章小结

本章首先对后续算法中应用的主要技术进行了介绍，包括激活函数、CNN、RNN 和注意力机制等；然后介绍了人体活动识别领域的主流深度学习方法；最后着重介绍了对比学习的思想和主流对比学习框架，并对其中的一些技术细节进行了介绍。

第3章 基于对比学习的时间序列 CPD 模型

本章首先分析了活动数据连续性带来的负样本选取困难的问题，然后在分析了使用对比学习解决变化点检测问题(Change Point Detection, CPD)问题的可行性后，提出了基于对比学习的 CPD 模型用于分割原始活动数据以协助负样本选取，并对具体的改进和模型结构进行了阐述。最后通过实验和对比实验验证了模型的有效性。

3.1 问题分析与提出

在采集人体活动数据时，除了需要专业设备进行辅助，还需要志愿者的高度配合，而为了这些数据尤其是“长时间采集的未分割”的序列打标签，往往还需要专人进行监控，以上种种原因造成了高质量的带标签人体活动采集数据难的问题。以数据集 PAMAP2^[27]为例，该数据集采集了 9 个人的 7 种不同日常动作的传感器数据，该数据集因干扰数据少，标签精准而被广泛使用，但整体数据量并不大，想要在这些数据上通过有监督方法训练出精度高，具有泛化性的深度学习模型是很困难的。

由于受到标签数据少的限制，本文进而寻求无监督方法。对比学习作为近两年最新兴的无监督方法受到了广泛的关注，并在 CV 领域的得到了巨大的成功，具体表现为在很多任务上战胜了有监督方法。尽管图像数据和时间序列都是具有特定结构的数据，但将 CV 领域对比学习框架直接应用到人体活动识别会遇到很多问题。

对比学习中最重要的一环是正负样本对的构建，本章主要解决活动数据连续性带来的负样本选取困难问题：在图像领域，每张图片独立成帧，每个样本之间独立性强；在时间序列领域中，本文通过在原始序列上应用滑动窗口方法来构造样本，而对于人体活动识别问题，因为人体活动天生具有连续性，所以构造出来的样本独立性较差。我们在进行数据采集时，通常是让受试者做一段时间的动作 A，再做一段时间的动作 B，因此我们经常可以看到在长序列中构造出来的样本彼此很相像的情况。在选取正负样本对中的负样本时，由于没有标签作为参考，在随机选取的过程中，很容易将同一段内的样本作为负样本。在 CV 领域，图像天然可分，不同样本即使标签一样，也很难出现完全相同的两个样本，所以这种负样本可以作为 hard negative^[52]。但在活动识别问题中，这种连续性很容易让我们选取到完全一致的负样本，这种相似性有悖于对比学习的思想，所以我们有必要对原始时间序列进行分割，将其分割为彼此不同的几段，对于某个给定样本，在选取负样本时，本文从其它段中选取，而在选取正样本时，从本段中进行选取。

上面这个问题可以转化为时间序列的 CPD 问题。本章提出基于对比学习的时间序列模型 CL-CPD，旨在解决对比学习正负样本对构建中负样本选取难的问题。

3.2 基于对比学习的时间序列 CPD 算法

3.2.1 使用对比学习解决 CPD 问题的可行性

通过前面的分析，由于活动数据的连续性和周期性，现有的对比学习框架并不能直接用于解决活动识别问题。然而 CPD 问题与活动识别问题不同，虽然也是分类问题，但是 CPD 问题是个二分类任务，即我们只需要判别出某一个子序列是否是变化点即可，而不需要像活动识别问题的多分类任务那样识别出某一个子序列到底归属于哪一类别。在具体的序列中，变化点通常意味着波形的突变，对应着我们在日常活动中两种动作之间的过渡动作，即算法需要找到那些具有两个及以上周期或者不具有周期性动作的子序列。如图 3.1 所示，横坐标为 300 时波形发生了突变，若窗口选定为 200，那么 $[200,400]$ 就是包含变化点的子序列，我们就是要在原始序列中找到所有这样的子序列。

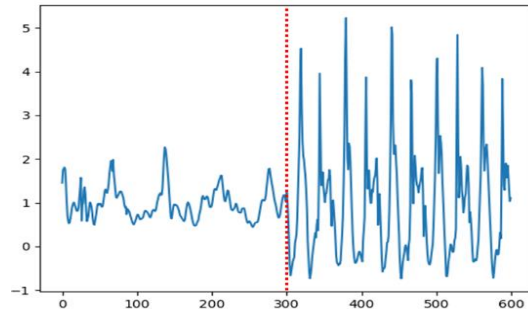


图 3.1 变化点示意图

Fig. 3.1 Schematic diagram of change points

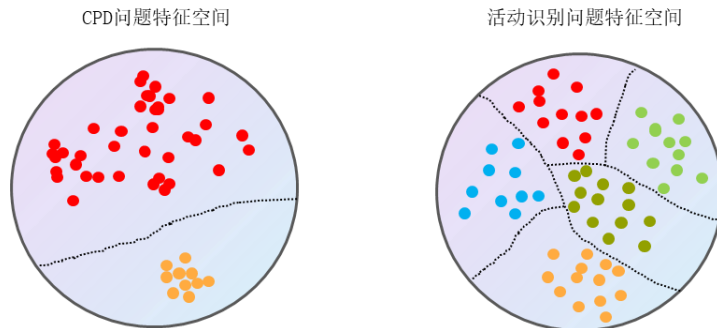


图 3.2 CPD 问题（左）与活动识别问题（右）的特征空间与决策边界

Fig. 3.2 Feature space and decision boundary of CPD problem(left) and activity recognition problem(right)

在特征空间，两个任务对于特征映射的要求也不同。如图 3.2（左）所示，对于

CPD 问题，只要可以将所有的样本大致分成两类即可；而不需要像活动识别问题那样将其均匀地按照类别映射到特征空间中，如图 3.2（右）所示。两个问题所需要的决策边界也不同，二分类问题的决策边界通常比较简单。

针对 CPD 这种二分类问题，为了不受活动数据的连续性和周期性的影响，构建正负样本对的方式也有所不同，我们直接从原始序列出发，按照对比学习的思想，对于某个确定的序列，我们将原始序列中距离最远的子序列选为负样本，将距离最近的子序列选为正样本。在对比学习框架下，我们根据上述特性来构建正负样本对。原始算法如算法 3.1 所示：

算法3.1 原始正负样本对构建算法

Algorithm 3.1 Original positive and negative sample pair construction

输入： 原始时间序列 T ，窗口长度 W

输出： 正负样本对集合 S

BEGIN

1. $[t] = \text{segment}(T, W)$ //按照窗口 W 对原始序列 T 进行分割，得到样本集 t
2. $S = \emptyset$ //初始化正负样本对集合 S 为空
3. **for** t_i **in** t **do**
4. $[distance] = \text{calculateDistance}(t_i, t)$ //计算序列 t_i 与样本集 t 中每个样本的距离
5. $Positive_i = \min Distance[distance]$ //选取距离最小的为正样本 $Positive_i$
6. $Negative_i = \max Distance[distance]$ //选取距离最大的为正样本 $Negative_i$
7. $\text{addToS}(\{t_i, Positive_i, Negative_i\})$ //将三元组 $\{t_i, Positive_i, Negative_i\}$ 加入到 S 中
8. **End for**
9. **Return** S

END

简单来说，对于某个确定的序列片段，我们要寻找到与其距离最大的序列片段作为其负样本，而与其距离最小的片段则作为其正样本。这种原始的构建方法的计算量是非常大的，我们要计算每个序列片段与整个序列中所有序列片段之间的距离。假设共有 N 个序列片段，若要为每个片段构造正负样本对，使用原始的算法，至少需要 N^2 次计算。

3.2.2 简化的正负样本对构造方法

由上一节的分析，时间序列的 CPD 问题可以理解为一个二分类问题，变化点为正样本，其他部分则为负样本。为了学习到序列片段是否包含到多种周期的动作或未包含一个完整周期这一特征，我们需要对样本的选取划分一定的范围，即我们从限定的连续的多个窗口，而不是从完整的序列进行正负样本对的构造。另外原始的构建算法时间复杂度过大，难以在实验中实现。本节提出一种简化的正负样本对构造方法，

减少了正负样本对构造所需的计算次数，如算法 3.2 所示。

算法3.2 简化的正负样本对构建算法

Algorithm 3.2: Simplified positive and negative sample pair construction

输入： 原始时间序列 T ，窗口长度 W ，连续片段个数 M

输出： 正负样本对集合 S

BEGIN

1. $[tm] = \text{segment}(T, W, M)$ //按照窗口 W 对序列 T 进行分割，将每 M 个样本组成得到本集 tm
2. $S = \phi$ //初始化正负样本对集合 S 为空
3. **for** tm_i **in** tm **do**
4. **for** t_j **in** tm_i **do**
5. $[distance] = \text{calculateDistance}(t_j, tm_x)$ //计算序列 t_i 与随机其他样本子集 tm_x 中每个样本的距离
6. $Positive_j = \minDistance[distance]$ //选取距离最小的位正样本 $Positive_j$
7. $Negative_j = \maxDistance[distance]$ //选取距离最大的位正样本 $Negative_j$
8. $\text{addToS}(\{t_j, Positive_j, Negative_j\})$ //将三元组 $\{t_j, Positive_j, Negative_j\}$ 加入到 S 中
9. **End for**
10. **End for**
11. **Return** S

END

简化的正负样本对构建算法如图 3.3 所示，核心思想是选取两个连续片段的窗口 w_1 和 w_2 ，每个窗口中都有若干个连续子序列。对于 w_1 中的一个确定的序列片段 T_i ，正负样本均在 w_2 中选取。对 w_1 中每个片段都进行同样的操作，通过复用片段来实现减少计算次数。假设共有 N 个序列片段，若要为每个片段构造正负样本对，使用改进的算法，需要 $M \times N$ 次计算即可。当 $M = N$ 时，改进的算法退化为原始算法，一般情况下， M 的取值越小（最小为 3），所需的计算量越小，但同时最终分类效果越差。在选取负样本时，可以将除了正样本的其他样本均作为负样本，即有多个负样本。以下图的 x_1 为例，最终加入集合的正负样本对为 $\langle x_1, y_3, (y_1, y_2, y_4, y_5, y_6) \rangle$ 。

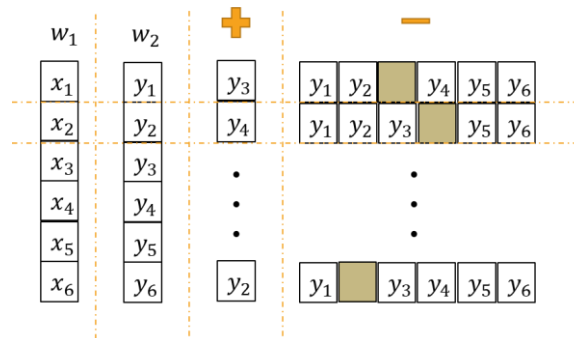


图 3.3 简化的正负样本对构建算法示意图

Fig. 3.3 Schematic diagram of the simplified positive and negative sample pair construction algorithm

3.2.3 多尺度卷积网络

3.2.3.1 多尺度卷积

CNN 结构在图像领域取得了巨大的成功,这与图像本身的结构特性等密不可分,而基于传感器的人体活动数据也有类似的结构特性。以一个完整的跳跃动作为例,受试者首先会有下蹲动作,此时 Z 轴线速度为负;然后受试者起跳此时 Z 轴线速度为正,受到重力的影响, Z 轴线速度逐渐减少变为负数;最后受试者落地 Z 轴线速度归 0。上述过程在实际序列上的体现则为一个特定的波形。如果一段序列中多次出现这种波形,那么这段序列对应跳跃动作的概率就会比较高。这与通过卷积操作来对图像中按照特定边缘进行语义分割的思想有很多的相似之处。时间序列领域的经典模型 DeepConvLSTM 也是通过卷积操作来初步提取特征,再使用提取到的特征输入到 LSTM 模型进行训练。

但是传统的基于 CNN 的活动识别模型没有考虑到各个动作的周期并不相同的问题,对同一个采集主体,每种动作的完整周期尚且不同,而对于不同的采集主体,动作之间的周期差异会更大。如果使用单一尺度的卷积核进行卷积操作,容易对某些特征过于关注,而忽略了其他特征。

为了解决这一问题本文提出多尺度卷积网络 MS-CNN,在时间维度上,解决窗口长度与动作周期不一致的问题。在每一个通道上使用不同尺寸的卷积核进行特征提取,并利用 padding 技术使得到的特征图尺寸相同。

3.2.3.2 空洞卷积

普通的卷积操作下,感受野的大小完全由卷积核的大小决定,想要增加感受野,就要增加卷积核的尺寸,但是过大的卷积核会在导致信息的丢失。上采样技术可以在一定程度上解决这个问题,但是上采样本身是一种“不可学习”的技术,无法针对不同的特征进行变化。针对这一问题, Fisher 提出了空洞卷积^[43]。

空洞卷积广泛应用于图像领域的语义分割^[44,45]任务,因为图像中每个要分割的物体大小是不同的,而普通的卷积操作的感受野是固定的,使用普通的卷积操作很难提高模型的准确率。尽管空洞卷积会有特征不连续的问题,不适合解决像素级别的预测问题,但是非常适合用来对不同尺寸的子图进行判别。空洞卷积通过扩张度参数来控制感受野的大小,扩张度越大,感受野就越大,但同时提取到的特征就越不连续,空洞卷积如图 3.4 所示。

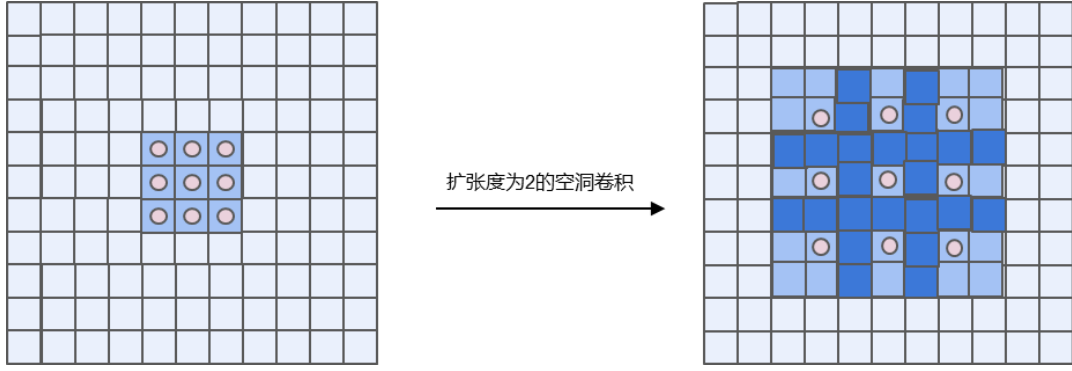


图 3.4 空洞卷积示意图

Fig 3.4 Schematic diagram of dilated convolution

本文将空洞卷积引入到时间序列，来增加一维卷积的感受野。与语义分割问题类似，人体活动也有周期不同的问题，例如上楼和跳跃两种动作的完整周期相差就较大，且在本章主要解决变化点检测问题，并不需要对序列进行预测，因此非常适合使用空洞卷积。一维空洞卷积如图 3.5 所示：

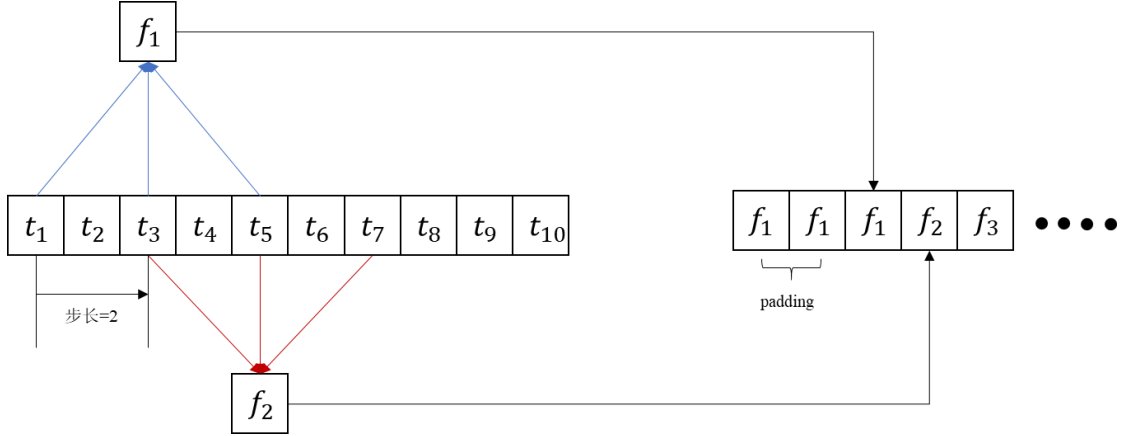


图 3.5 一维时间序列空洞卷积示意图

Fig. 3.5 Schematic diagram of dilated convolution for time-series

将多尺度卷积和一维空洞卷积结合后，MS-CNN 的结构如图 3.6 所示。我们这里使用 3 种不同尺度的卷积核进行卷积操作，并在最大的卷积核上应用了空洞卷积。在进行卷积操作后，我们将三个特征图拼接到一起，并借鉴 ResNet 的思想，将原始输入一同拼接到特征图上。

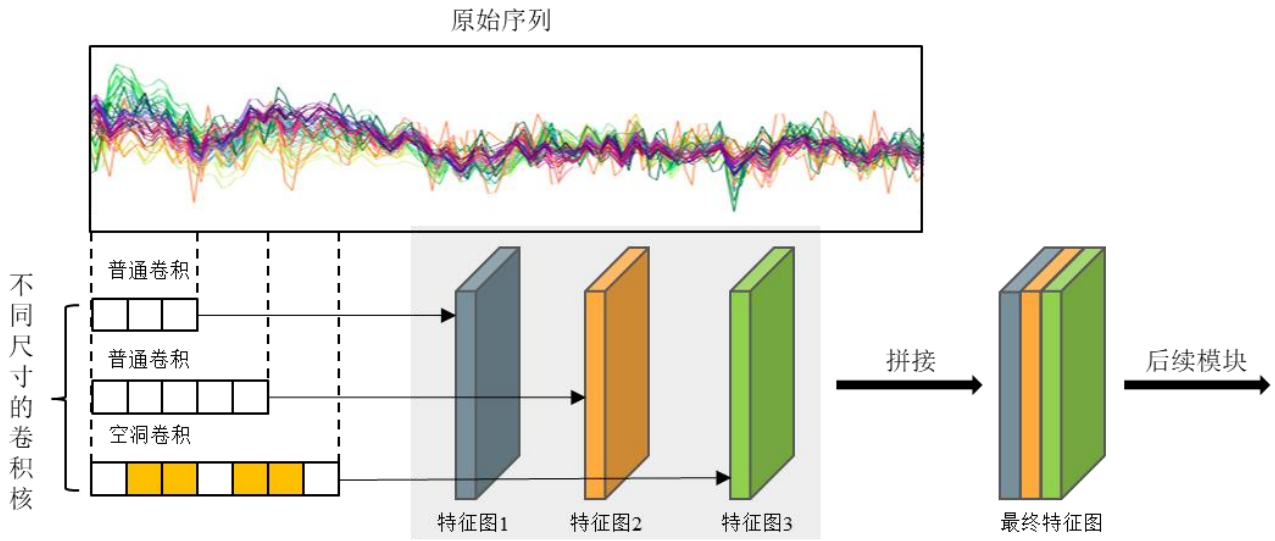


图 3.6 MS-CNN 结构示意图

Fig. 3.6 Schematic diagram of MS-CNN

3.2.4 跨通道注意力模块

活动识别领域的经典模型，往往只关注通道域的特征，而忽略了空间域的特征。以活动数据中最常用的三轴加速度为例，通常操作是将每个轴作为独立的通道，分别在上面进行卷积操作提取特征。但一个隐含的事实是，人能做的动作要同时受到人体机能和自然规律的束缚。这一特点在多传感器上体现的更为明显：例如当胸部的传感器三轴加速度同时为正值时，脚踝部的传感器三轴加速度不能同时为负。这说明通道间数据同样存在隐含的某种关系。本文受图像领域的注意力机制 CBAM 启发，提出适用于多维传感器数据的注意力模块，结构如图 3.7 所示：

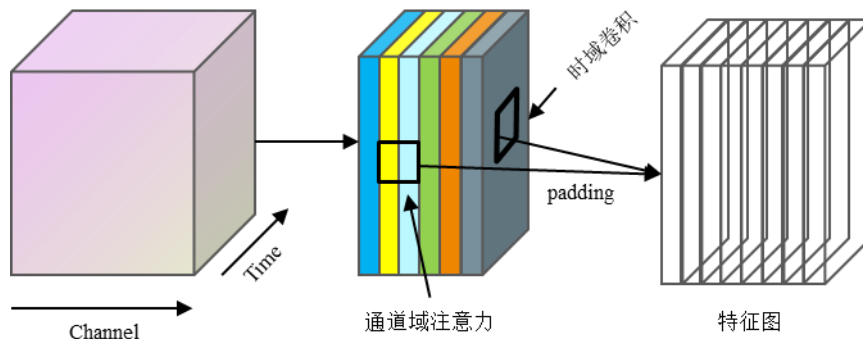


图 3.7 通道注意力层示意图

Fig 3.7 Schematic of channel attention

整个模块可以分为三层：第一层为卷积层，通过对跨通道的特征图做卷积，以此来提取跨通道的注意力，剩下两层为全连接层。在第三层中，为了与后面新的卷积模块拼接，为了调整输出维度，并借鉴 ResNet 的思想，将原始的特征图作为第三层输入得到最终的输出。

考虑到前面模块的复杂度，我们这里采用了如公式 (3.1) 的分段均值方式来代替卷积操作，并结合 padding 操作，在最终的实验里仍然取得了不错的效果，

$$Z_n^i = \frac{1}{L_f^i} \sum_{i=1}^{L_f^i} x_n(i) \quad (3.1)$$

整个跨通道注意力模块的结构如图 3.8 所示：

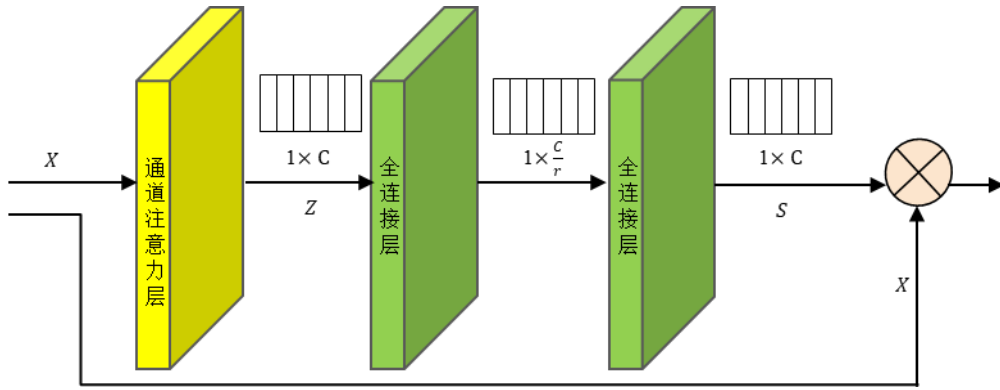


图 3.8 跨通道注意力模块示意图

Fig. 3.8 Schematic diagram of the cross-channel attention module

3.2.5 损失函数

为了度量正负样本对中各样本的相似性，首先需要指定损失函数中的相似性度量函数。本算法中使用余弦相似度作为度量方法：

$$Similarity(z_i, z_j) = \frac{z_i \cdot z_j}{|z_i| |z_j|} \quad (3.2)$$

由公式 (3.2) 可以看出，相似的样本的相似度接近 1，而不相似的样本相似度接近 0。这里的 z_i 和 z_j 是模型输出高维特征向量。

简化的正负样本对构建算法每次输出为一个给定样本，一个正样本和多个负样本的集合。对比学习的损失函数有很多种，其中 N-pair Loss^[51]非常适合处理这种一正多负的样本组合，如公式 (3.3) 所示：

$$L_{N-pair}(x, x^+, \{x_i^-\}_{i=1}^{N-1}) = -\log \left(\frac{\exp(f(x)^T f(x^+))}{\sum_{i=1}^{N-1} \exp(f(x)^T f(x_i^-)) + \exp(f(x)^T f(x^+))} \right) \quad (3.3)$$

其中 f 为公式 (3.2)。

3.2.6 模型结构

完整的模型结构如图 3.9 所示：

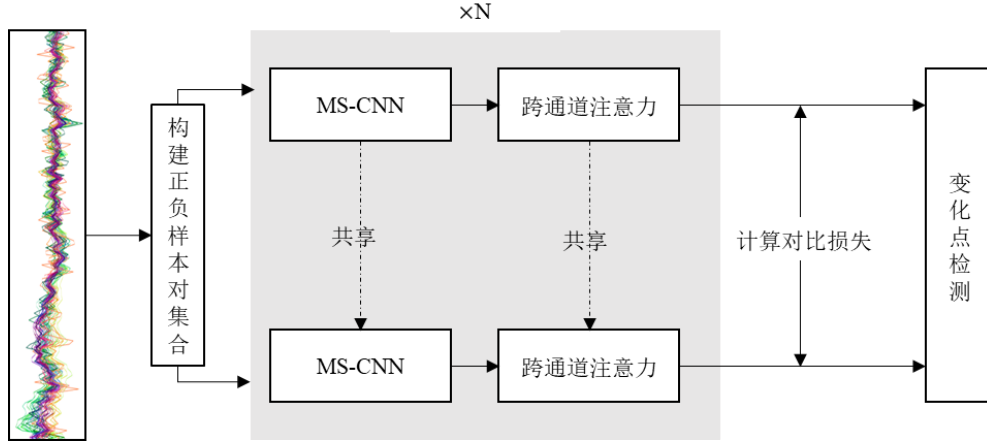


图 3.9 CL-CPD 模型结构图

Fig. 3.9 Structure of CL-CPD model

原始序列首先按照算法 3.2 构造正负样本对集合。集合中的每个元素都是一正多负的多个序列组合，将这些序列分别输入到由 MS-CNN 和跨通道注意力模块组成编码层，这个编码层可以堆叠多层。当对正负样本完成编码后，按照公式（3.3）计算对比损失，并根据对比损失优化编码器。当完成训练后，固定编码层参数，并在其基础上训练分类头，即进行变化点检测。

3.3 实验分析

3.3.1 数据集

在本章实验中，我们采用与提出 KL-CPD 的文献[49]同样的数据集 USC-HAD 和 HASL。

3.3.1.1 USC-HAD 数据集

USC-HAD 数据集采集自 14 个人的共计 12 种动作的日常活动场景。每个动作作用单独的时间序列进行记录。具体类别如下：{1: Walking Forward, 2: Walking Left, 3: Walking Right, 4: Walking Upstairs, 5: Walking Downstairs, 6: Running Forward, 7: Jumping Up, 8: Sitting, 9: Standing, 10: Sleeping, 11: Elevator Up, 12: Elevator Down}，数据集的基本信息见表 3.1。

我们从里面随机 30 条数据拼接到一起，并根据标签设定是否为分割点。在本节

实验里面我们只使用了三轴加速度作为输入数据。

3.3.1.2 HASC 数据集

HASC challenge 2011 数据集是由包含加速度和角速度多传感器采集得到的，本章试验中使用的数据集是 HASC 的子集，只包含三轴加速度。具体的类别如下：{1: stay, 2: walk, 3: jog, 4: skip, 5: stair up, 6: stair down}，数据集的基本信息见表 3.1。

表 3.1 本节实验中 USC-HAD 和 HASC 数据集的基本信息

Table 3.1 Basic information on USC-HAD and HASC datasets in this experiment

Dataset	Samples	Sequence	Channels	Change Points
HASC	39K	1	3	65
USC-HAD	97K	6	3	30

3.3.2 数据预处理

无论是机器学习方法还是深度学习方法，对属于数据进行预处理都是必不可少的。输入数据的不同属性的度量标准大概率不同，其中尺度较大的特征波动也较大，很容易将尺度较小的特征波动掩盖掉，造成信息的丢失。为了保证模型的学习稳定性，通常将特征调整到一个特定区间内。在人体活动识别问题中，可穿戴传感器的每一维数据都记录了相对独立的信息。为了保留这种独立性，最通常的方法是，我们在每一维上单独进行预处理。下面本文将按照这种设定方式，在每一条传感器信道上，对数据处理进行 z-score 标准化如公式(3.2)。

$$x' = \frac{x - \mu}{\sigma} \quad (3.2)$$

3.3.3 数据分割

原始的传感器数据长度往往从几千到几万不等，且包含各种不同的动作，这种原始序列并不能直接作为识别模型的输入。我们需要对原始序列进行划分，得到维度相同、长度相同的多个样本作为模型的输入。最常用的方法就是滑动窗口，本质上可以理解为值全为 1 的卷积操作。在使用滑动窗口方法时，要划定窗口的大小和每次滑动的步长。滑动窗口的分割过程如图 3.10 所示：

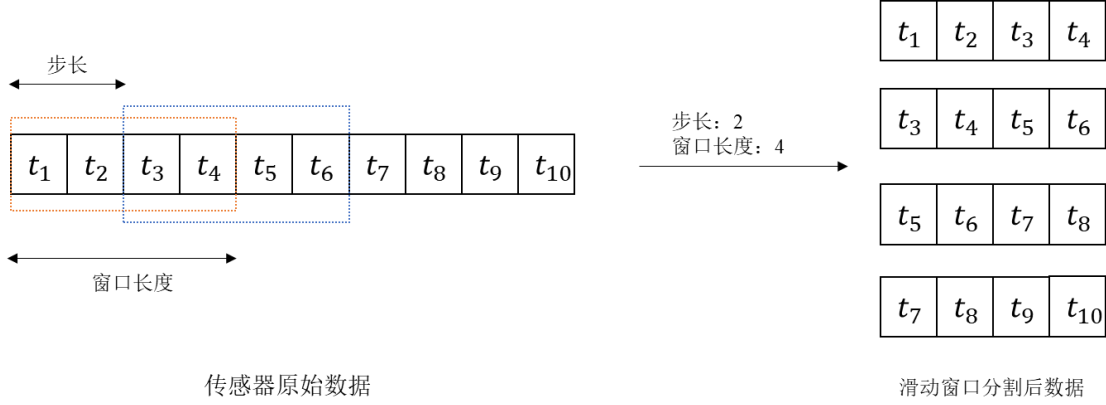


图 3.10 滑动窗口分割数据示意图

Fig. 3.10 Schematic diagram of sliding window segmentation data

(1) 窗口大小

窗口大小一般根据数据集提供的采样频率来设定。通常，一个窗口中应该至少包含一个完整的动作，但是过大的窗口也容易导致样本不足和数据混乱的问题。根据经验和实际情况，窗口应设定为 1s 到 2s 之间。考虑到本节实验采用的空洞卷积模型，窗口可以适当调大。在后续的实验中，本节分别尝试了长度为 100,200 和 400 的窗口，并在长度为 400 的窗口上取得了最高的准确率。

(2) 步长

滑动窗口的步长是指在两次划分过程中窗口之间的距离，也称为覆盖率，在窗口大小已经确定的情况下，可以通过步长来控制从原始序列中分割的样本个数。当步长过小时，会出现覆盖率过高的情况，样本之间过于相似造成了数据的冗余。当过大时，会出现覆盖率过低的情况，会造成信息的缺失，同时当原始序列长度较短时，过大的步长会造成样本不足的问题，难以利用深度学习训练出有效的模型。根据经验，步长一般选为窗口大小的一半，即覆盖率为 50%。

在 CPD 问题中，一般需要设定 Detection Margin 来辅助对实验结果的评价。由于输入样本受到划分样本时窗口的大小的影响，我们可以容忍一定程度的误差。具体算法如下：

算法3.3 变化点检测结果评价标准

Algorithm 3.3 Score Regimes for Change Point Detection

输入：实验分割结果集合 $estCP$ ，真实分割结果 $gtCP$ ，检测边界 DM

输出：F1-Score

BEGIN

1. $TP=0, FN=0, FP=0, TN=0$

2. $numCP = \text{length}(estCP)$

算法3.3（续） 变化点检测结果评价标准

Algorithm 3.3 (continued) Score Regimes for Change Point Detection

```

3.  visited = [0]×numCP
4.  for i = 1 to numCP do
5.      if estCP[i] == 1 and gtCP[i] == 1
6.          TP = TP+1; Visited[i] = 1
7.      else if estCP[i] == 0 and gtCP[i] == 0
8.          TN = TN+1; Visited[i]=1
9.      else if estCP[i] == 0 and gtCP[i] == 1
10.         If hasTrueIn(i,DP) //寻找i前后范围为DP的结果是否有尚未访问过的True
11.             TP = TP+1; Visited[i] = 1
12.         Else
13.             FN = FN+1
14.     Else
15.         FP += 1
16. End for
17. precision = TP/(TP+FP) //计算准确率
18. recall = TP/(TP+FN) //计算召回率
19. F1-Score = 2×precision×recall/(precision + recall) //计算F1-score
20. Return F1-Score
END

```

根据上述算法，一个 Detection Margin 中的正样本最多只会被计算一次，即使有多个偏离点出现，也只会计算一个真阳性。

3.3.4 实验结果分析

表 3.2 CL-CPD 与当前主流方法在 USC-HAD 上的 F1-score

Table 3.2 F1-score of CL-CPD and current mainstream methods on USC-HAD

Method	DM100		DM200		DM400	
	window	F1	window	F1	window	F1
FLOSS	100	0.2666	50	0.3666	400	0.4333
aHSIC	50	0.3333	50	0.3333	50	0.3999
RuLSIF	400	0.4666	400	0.4666	400	0.5333
ESPRESSO	200	0.6333	100	<u>0.8333</u>	100	<u>0.8333</u>
KL-CPD	100/4	<u>0.7426</u>	200/32	0.7180	400/16	0.6321
CL-CPD	400/128	0.8434	400/128	0.8434	400/128	0.8597

表 3.3 CL-CPD 与当前主流方法在 HASC 上的 F1-score
Table 3.3 F1-score of CL-CPD and current mainstream methods on HASC

Method	DM100		DM200		DM400	
	window	F1	window	F1	window	F1
FLOSS	60	0.3088	60	0.3913	100	0.5430
aHSIC	50	0.2308	40	0.3134	40	0.4167
RuLSIF	200	0.3433	200	<u>0.4999</u>	200	0.4999
ESPRESSO	100	0.2879	60	0.4233	100	<u>0.6933</u>
KL-CPD	60/4	<u>0.4785</u>	100/4	0.4726	200/64	0.4669
CL-CPD	60/64	0.5152	100/64	0.5807	200/64	0.7231

在上面两个表中，FLOSS^[55]，aHSIC^[56]，RuLSIF^[57]和 ESPRESSO^[58]是基于统计的非深度学习方法，KL-CPD 和 CL-CPD 是基于深度学习的方法。在后两种方法中除了要设定滑动窗口大小 window，还需要设定深度学习的 batch size，以 CL-CPD 在 USC-HAD 数据集上的实验为例，400/128 代表 window 设定为 400，同时 batch size 设定为 128。

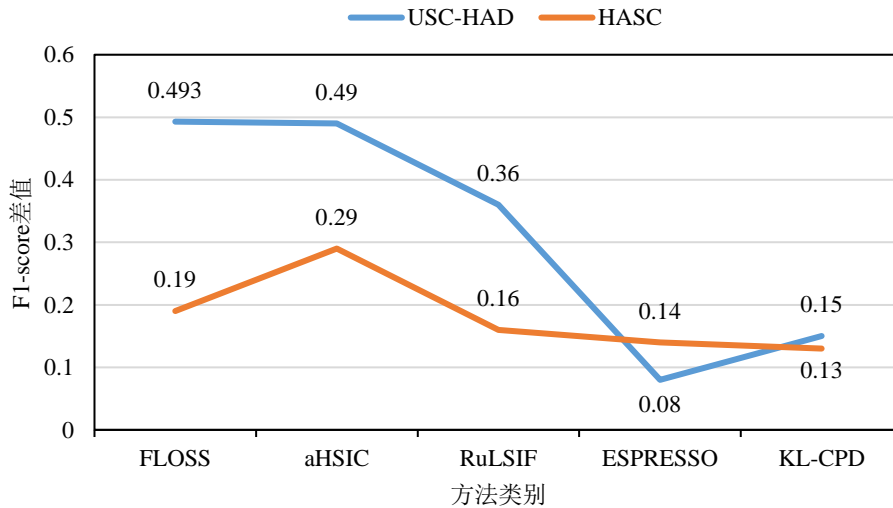


图 3.11 CL-CPD 与其他方法的平均 F1-score 提升比例

Fig. 3.11 Average F1-score improvement ratio of CL-CPD to other methods

CL-CPD 与其他方法的准确率比较结果如图 3.11 所示。可以看出 CL-CPD 大幅领先基于统计的传统方法，相较基于深度学习和基于熵分割的方法仍有一定优势。具体的，基准的 CPD 方法通常去寻找某种特定类型的时间序列中突然变化的点，所以这些方法并不能有效的泛化到不同类型的数据集上。例如 ESPRESSO 在 USC-HAD

数据集上取得了不错的效果，因为在这个数据集中，变化点在时序形状和统计特性上均有较大的变化，而在 HASC 数据集上，变化点有更稳定时序形状和统计特性，是一个缓慢变化的过程。无论是 ESPRESSO 还是其他的非深度学习方法，都不能同时在两个数据集上取得最优的效果，且严重依赖于窗口的划分和 Detection Margin 的设置，这也是传统方法最大的问题所在：超参数对变化点检测的准确率影响非常大，且在不同领域的数据集上表现波动会比较大。

为了探究跨通道注意力和空洞卷积对识别准确率的影响，本节进行了消融实验，对比了在原始算法、仅使用空洞卷积、仅使用跨通道注意力和同时使用空洞卷积和跨通道注意力的识别结果。循环测试 20 次的实验结果如图 3.9 和 3.10 所示，其中窗口长度为 400，batch size 设置为 128。

由图 3.12 可以看出在仅使用空洞卷积的情况下，F1-score 的波动较大，其最大值大于原始算法，最小值和平均值却小于原始算法，这是因为空洞卷积的稀疏性会导致提取到的特征较为离散，学习到的特征不够稳定；在仅使用跨通道注意力的情况下，F1-score 的波动相较仅使用空洞卷积的情况下更小，这是因为在跨通道注意力可以捕获范围更广的特征，甚至是全局的特征，这一特性使得学习到的特征更为稳定；而 CL-CPD 将前两者结合，得到的 F1-score 在绝大多数情况超过前两者，并且稳定性超过了仅使用跨通道注意力的方法，可以认为 CL-CPD 充分的结合两者的优势。图 3.13 更直观地展示了这一特性。

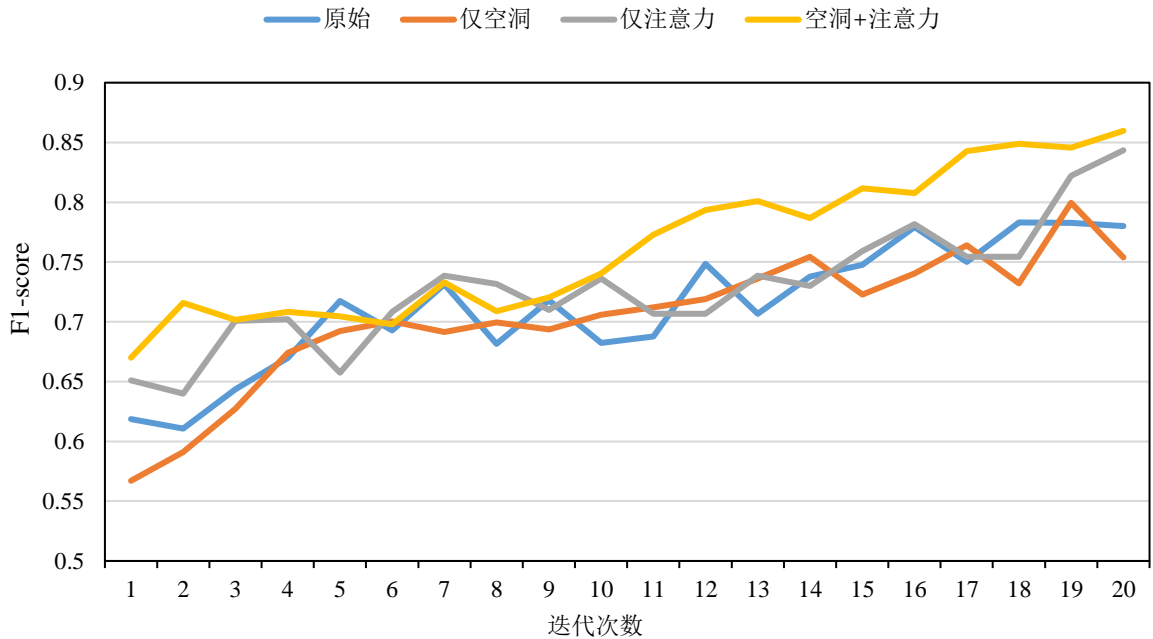


图 3.12 USC-HAD 数据集的消融实验结果

Fig. 3.12 Results of ablation experiments on the USC-HAD dataset

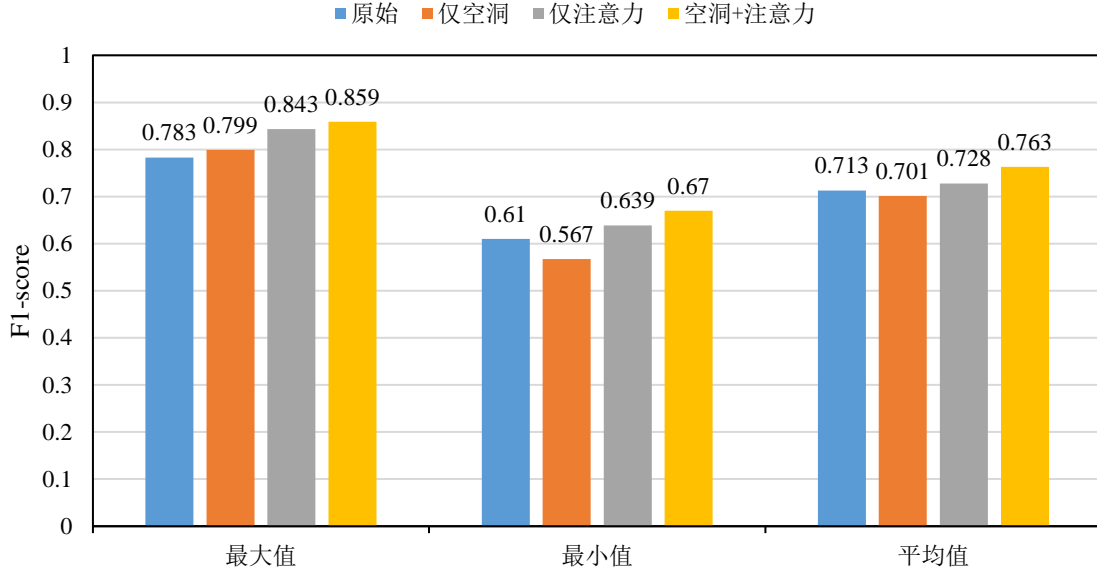


图 3.13 USC-HAD 数据集消融实验数据统计

Fig. 3.13 USC-HAD dataset ablation experiment data statistics

综上，相对基于统计的方法而言，基于深度学习的算法表现更稳定，受超参数和数据集特性的影响较小。根据前文提到的正负样本对构造方法，CL-CPD 方法实际上是一种小窗口内的双向时间序列表示学习方法：同时比较与选定片段距离最远和距离最近（较近）的片段，并将其映射到隐向量空间中合理的位置。回归到 CPD 是一个二分类问题的本质，这种小窗口内的表示学习足够稳定，且收敛快，尽管比较简单，但已经足够胜任这种二分类任务。如图 3.14 所示，模型在不超过 20 轮迭代的情况下就可以收敛，且正负样本之间的相似性得到了充分地建模，即正样本之间提高正样本对之间的相似性的同时，降低负样本对之间的相似性。

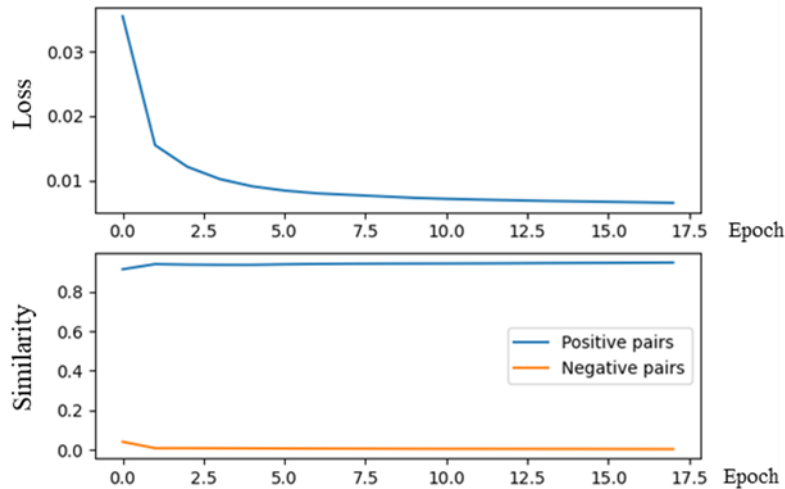
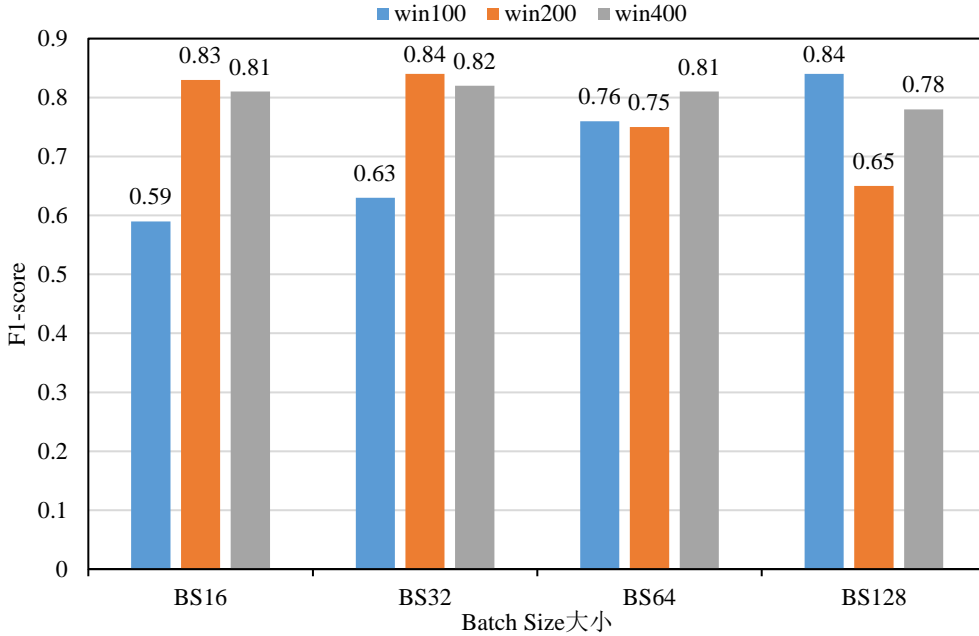


图 3.14 CL-CPD 训练损失与正负样本相似性

Fig. 3.14 CL-CPD training loss and similarity of positive and negative samples

上图中,可以看到在实验的开始阶段,正样本对的特征之间的相似度已经很高,这是因为我们在构建正样本时,选取的就是距离最小的样本,负样本对同理。

本章还额外进行了实验对 batch size 和窗口大小对准确率的影响进行探究。从图 3.15 可以看出来,总体上,在使用较小的窗口划分序列时,适合使用较小的 batch size,



而在使用较大的窗口划分序列时,适合使用较大的 batch size。这可能是因为较大的窗口样本因较多的特征需要更多的样本来决定模型优化的方向。

图 3.15 不同 batch size 和窗口长度对 F1-score 的影响

Fig. 3.15 Effects of different batch sizes and window lengths on F1-score

3.4 本章小结

本章针对活动数据的连续性导致的,正负样本对构建中的负样本选取困难的问题,先将其简化为时间序列的变化点检测问题,并提出了基于对比学习的时间序列变化点检测算法 CL-CPD。具体地,提出了简化的正负样本对构建方法,并对具体的模型结构进行了改进:引入了空洞卷积、多尺度卷积和跨通道注意力模块。通过与当前主流的非深度学习算法和深度学习算法进行对比,验证了 CL-CPD 算法的有效性。

第4章 基于 Transformer 的人体活动数据增强模型

本章首先分析了活动数据周期性带来的正样本构造困难的问题，然后通过对比现有技术优劣确定了基于 Transformer 的数据增强模型，并对模型结构和具体改进进行了阐述。最后通过实验和对比实验验证了模型的有效性。

4.1 问题分析与提出

4.1.1 数据增强

深度学习模型因其强大的表达能力广泛应用于各个领域，一般来说，模型结构越复杂，参数越多，其表达能力就越强，但是对训练样本数量的需求也就越大。当训练样本不足时，很容易发生过拟合现象。深度学习模型在计算机视觉领域取得了巨大的成功，这除了模型本身强大的表达能力之外，也与数量庞大，种类丰富的开源图像数据集密不可分。以最为常用的 ImageNet 为例，整个数据集 1420 多万张图片^[46]。庞大的数据集与模型相辅相成，共同促进了深度学习模型在图像领域的发展。尽管已经有如此庞大的数据集，如何更有效的利用数据，仍是人工智能领域非常重要的问题。数据增强是在不增加真实数据的前提下，在给定的数据上生成样本，使其等效于更多数据的一种数据处理方式，可以简单分为有监督的数据增强和无监督的数据增强。通常，使用正确的数据增强方法来扩充样本的分布，使得模型训练更加困难，避免在少量样本上过拟合，可以提升模型在测试时的性能。

在活动识别领域，无论是深度学习模型还是传统的统计模型或机器学习模型，往往受困于训练样本不足的问题^[47]。一般情况下，传感器数据抽象层级较高，人类难以对其有清晰的认识，很难直接参照原始数据进行分类标注，这导致高质量标签数据集规模小的问题更加严重。此外由于各个数据集采集数据的方式不尽相同，针对某一数据集的数据增强方法难以直接应用到另一数据集上。

4.1.2 对比学习正负样本对构建的正样本选取

对比学习的本质是构建正负样本对，并将它们映射到特征空间中，模型将正样本对映射到一起，而将负样本对的映射分开。理想情况下，负样本对可以理解为与选定样本差别最大的那个样本，在实际训练中，往往是选取不同类别的样本作为代替；而

正样本对可以理解为与选定样本最为接近的那个样本，在实际训练中，正样本往往根据选定样本自身来构建。

有监督的数据增强，即通过对原始数据的加工生成新的样本。以图像为例，裁剪、旋转、加噪声、改变亮度和改变对比度等操作，都是有监督的数据增强。在时间序列领域，常用的方法有增幅、震荡、前后翻转等，但并不是所有方法都适用于活动识别领域，因为人的所有活动都要受到人体的机能限制和自然规律的束缚，翻转这一数据增强方法并不适合应用于活动识别领域，另外通过这种方法生成的样本大多比较单调，和原样本差别不大，难以真正起到增加样本多样性的作用。

无监督的数据增强中最具代表性的方法是对抗生成网络（GAN）：通过学习输入样本的分布，模型从隐向量空间出发生成与输入数据分布一致的数据。这一方法在图像领域取得了巨大的成功，然而在时间序列尤其是活动识别领域，基于深度学习的数据增强方法尚未有深入的研究，也没有特别成熟的方法。本文在早期实验中首先尝试将传统的图像领域生成模型迁移到时间序列领域，但使用搭建基于 CNN 的 GAN 进行数据增强时，发生了模型坍塌的问题，即无论什么形式的输入，模型的输出求趋于一致，如图 4.1 所示。模型坍塌本身是一个非常复杂的问题，可以简单理解为梯度突然消失，而模型则趋于输出一个一定正确但是毫无变化的结果，这与我们想要进行数据增强的初衷相悖。

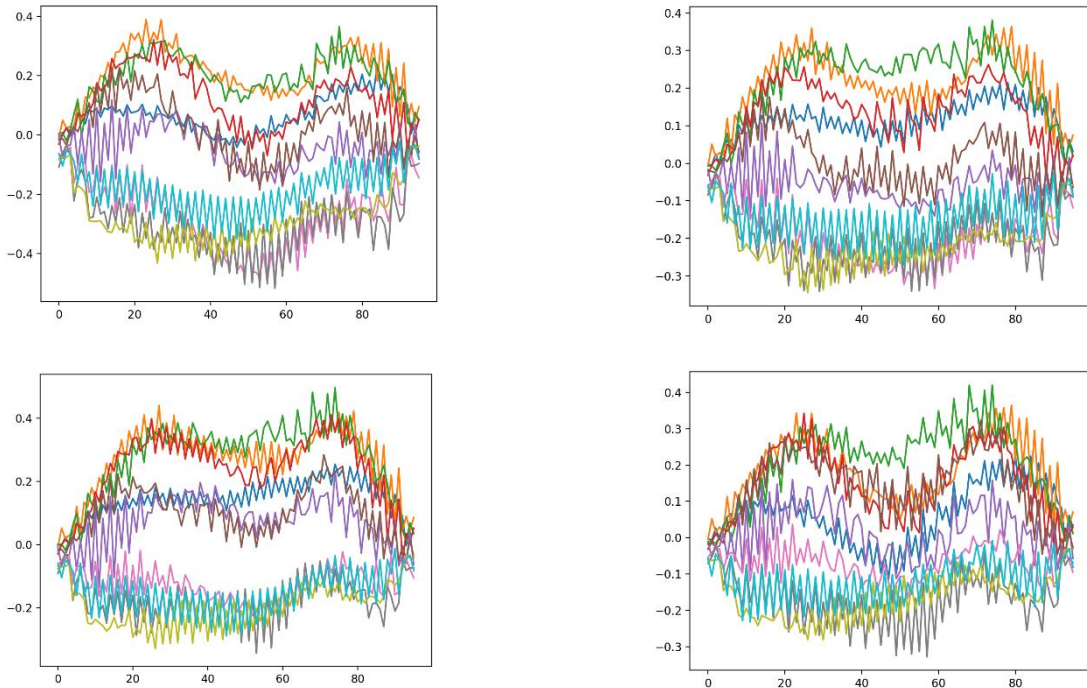


图 4.1 模型因为模型坍塌而输出非常相似的样本

Fig. 4.1 The model outputs very similar samples due to model collapse

本章从活动识别领域数据集规模小和对比学习框架正负样本对构造中正样本选取的需求出发, 基于时间序列本身的结构特性, 提出基于 Transformer 的生成模型。

4.2 基于 Transformer 的时间序列数据增强算法

4.2.1 技术选型

人体活动识别数据作为一种时间序列, t 时刻的动作与 $t-1$ 时刻及之前的动作的值有关, 也会对 t 时刻之后的取值产生影响, 这种特性非常适合用循环神经网络 RNN 来建模。但是 RNN 模型在 t 时刻的输入依赖于 $t-1$ 时刻模型的输出, 这导致了模型的并行训练能力非常差。本文在前期实验时, 受限于硬件条件, 即使使用基础的 RNN 模型进行预测, 迭代一个 epoch 也至少需要 5 分钟以上。

CNN 模型在时间序列分类任务上有出色的表现, 因为往往通过一段典型的、有的波形就可能判断出整个序列片段对应的类别。但是时间序列数据增强可以理解一种 Sequence-to-Sequence 任务, 我们不仅需要对局部特征进行学习, 还需要对序列中片段的前后关系建模, CNN 模型由于池化层的引入, 前后关联的特征很容易被忽略掉, 尽管有各种不同的池化技术, 但是这种 CNN 模型结构本身的特性难以从根本上解决这种问题。

自注意力机制 (Self-Attention) 最早在 2014 年由提出, 而 2018 年提出的 BERT 算法中最重要的部分便是注意力机制的应用 Transformer, 在 NLP 领域的机器翻译任务上大幅领先当时的算法, 并引领后续几年内的研究方向。在翻译任务中, 每个词的实际含义不仅取决于它自身, 也与前后文有较大的关系, 这与活动识别数据的数据增强有相同的思想。Transformer 的整个网络结构完全由注意力模块组成, 如图 2.6 所示。一个完整的注意力模块由多头注意力和前馈网络组成。这让 Transformer 可以进行大规模性并行计算, 并且基于 Transformer 的模型结构简单, 参数也相对较少, 模型训练起来更容易, 得到的最终模型也更小。

根据以上的分析, 本文最终选取了 Transformer 作为数据增强模型的基本结构。在目前结合了 CNN 与 RNN 技术的时间序列数据增强方法里, 最成功模型的是 RGAN^[48], 本文主要以 RGAN 作为参考, 对比两者在模型结构、增强效果、模型大小和训练时间上的差异。

4.2.2 模型结构

4.2.2.1 生成器

生成器的整体结构如图 4.2:

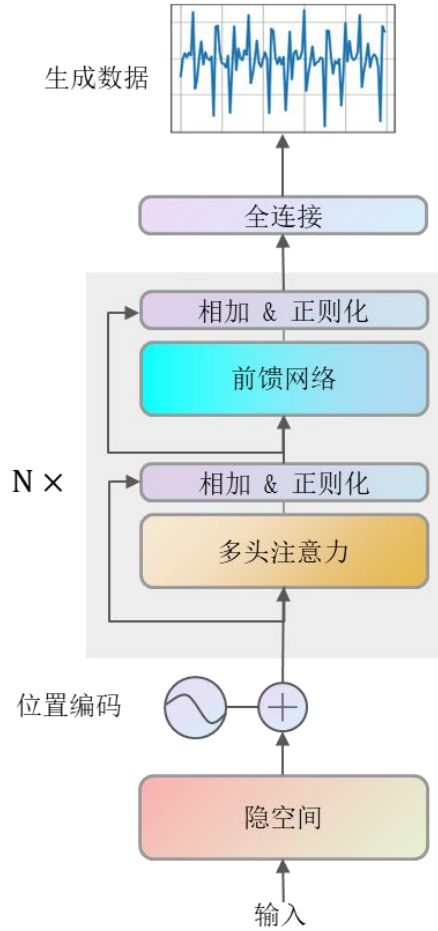


图 4.2 生成器结构图

Fig. 4.2 Structure of generator

生成器的任务是将隐空间随机选取的向量转化为合成数据，Transformer GAN (TGAN) 的生成器的整体结构与传统的 Transformer 编码层类似，可分为以下三部分：

首先是输入处理部分。从原始输入到隐空间向量的转换，TGAN 与 RGAN 是相通的。而在后续的处理上 TGAN 可以并行的处理一个 batch 的数据，而 RGAN 只能顺序地处理每个数据。此外为了让网络更好的学习到序列中前后位置的互相影响，我们对输入进行了 sin 位置编码。

接下来数据会进入注意力模块。注意力模块又可以分为多头注意力模块和前馈网络模块。

(1) 多头注意力模块：

多头注意力机制是通过缩放点积来实现的，计算方式如公式 (4.1)：

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (4.1)$$

其中, Q 代表 Query, 可以理解为完整的输入动作序列转化成的向量; K 代表 Key, 可以理解为一个完整动作里面的关键步骤, 或时间序列某种具有突出特征波形; V 代表 Value, 可以理解为所有可能的动作构成的样本矩阵。模型自主去学习哪些是关键向量、如何从查询序列和所有样本上抽取特征以及如何匹配他们之间的关系, 即注意力。在实际操作中, 我们通过在原始输入进行 1×1 的一维卷积得到这三个输入向量, d_k 是注意力的缩放因子, 一般取值为向量的维度。多头注意力网络后面是一个 skip 连接和层归一化, 目的是避免梯度消失并加速模型收敛。

(2) 前馈网络模块

前馈网络模块由两个全连接层组成, 第一层将特征向量映射到一个固定的维度, 第二层则将中间向量重新映射到原始的输入维度上, 以便于在后面堆叠多个注意力模块, 可以更好地提取特征。

最后是由一维卷积组成的, 带有 dropout 全连接层, 将编码层得到的中间向量转化为增强数据, 作为负样本输入给接下来的判别器。

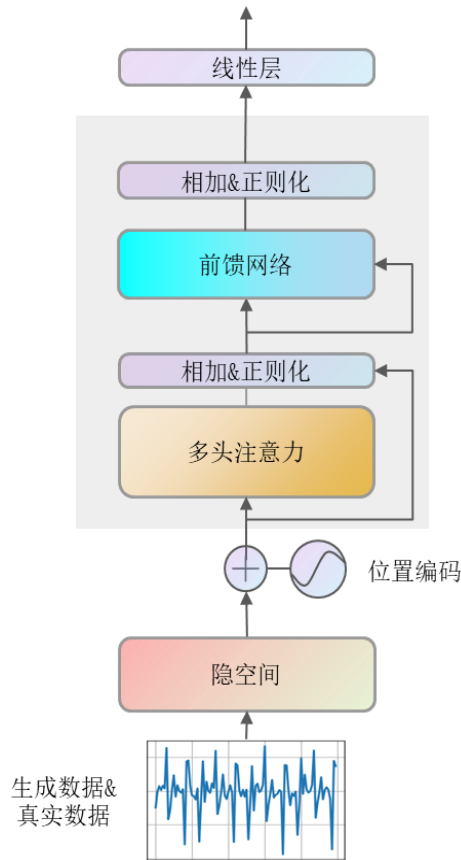


图 4.3 判别器结构图

Fig. 4.3 Structure of discriminator

4.2.2.2 判别器

判别器的整体结构如图 4.3 所示。判别器的任务是将真实数据和生成器生成的合成数据的真伪分别开来。整体上依然分为位置编码，注意力模块和最后的全连接层，与生成器的差别并不大。主要区别在于以下两点：

(1) 使用了不同的 **mask** 方式，详见 4.2.3 节。

(2) 在最后的全连接层上，没有使用生成器中的一维卷积，而是搭建了 Transformer 的分类器。在 RGAN 中，最终的识别使用的是 DeepConvLSTM。本文也尝试在 TGAN 中使用 DeepConvLSTM 作为最终的识别器，但是效果并不好。最终选定了 Transformer 的网络，在注意力模块中的前馈网络部分只保留一层即可，直接输出结果。具体的讨论见后续实验部分。

4.2.3 对角 mask

在输入的处理上，并没有使用与传统的注意力机制一样的 **mask** 方式，这是因为数据增强任务与机器翻译任务的目的不同，在机器翻译任务中，通常会使用“下三角 **mask**”，因为生成翻译序列时，我们实际上做的是对“下一个位置”的文本进行预测，通过 **mask** 技术，训练模型只根据当前的信息得出预测结果，而不对后续信息产生依赖；但在数据增强任务中，我们的任务是根据输入样本来生成一个同属性的新样本，并没有预测“下一步”的需求在里面，所以本文提出适用于时间序列数据增强的 **mask** 技术：对角 **mask**，如图 4.4 所示。

对于每个时刻 t_i ，我们关注的不是它的下一时刻，而是对这一时刻的“重构”。本质上我们并不是要对输入数据在 t_i 时刻的取值进行准确的还原，而是要捕捉到数据在 t_i 时刻的与前后两个方向的变化关系。我们将 t_i 时刻的信息 **mask** 掉，让模型利用前后的信息对这一时刻的信息进行推断。在实际使用模型时，我们也是使用原始样本作为输入，而将模型的输出作为增强结果，而不是像传统的 GAN 模型那样从直接隐向量空间出发生成增强数据。

在实际的实验中，我们通过三个一维卷积操作分别得注意力机制中的 Q 、 K 和 V ，利用它们进行多头注意力计算。多头注意力中的 **mask** 采用以 Q 的长度作为维度的单位矩阵。在计算完注意力向量后，与原始输入加和后输入到第一个前馈神经网络中，最后在将三个向量一起进行拼接，作为第二个前馈神经网络的输入，调整维度后，作为下一个多头注意力的模块的输入。

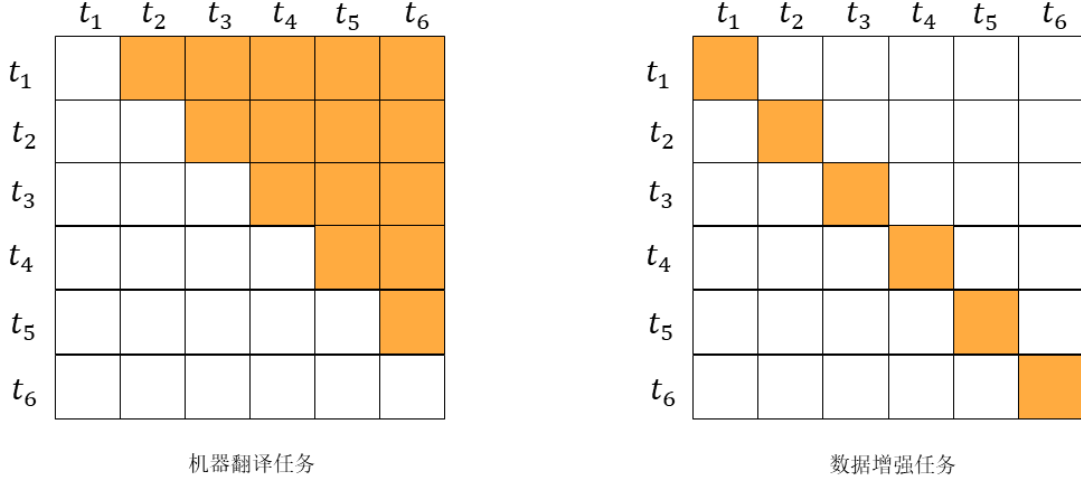


图 4.4 机器翻译（左）和数据增强（右）中使用的 mask 矩阵

Fig. 4.4 Mask matrix used in machine translation (left) and data augmentation (right)

4.2.4 损失函数

TGAN 的损失函数与原始 GAN 的损失函数一致，如公式 (4.2)：

$$L(x) = \min_G \max_D \mathbb{E}_{x \sim P_{data}} [\log D(x)] + \mathbb{E}_{z \sim p_z} [\log (1 - D(G(z)))] \quad (4.2)$$

其中 p_{data} 为原始数据分布， p_z 为生成数据分布。即生成器负责增强样本，而判别器负责将增强样本和原始样本分别出来。

4.3 实验分析

4.3.1 数据集

4.3.1.1 PAMAP2 数据集

PAMAP2 数据集采集自 9 个人的共计 7 种动作的日常活动场景。每个动作使用单独的时间序列进行记录。具体类别如下：{1: lying, 2: sitting, 3: standing, 4: walking, 5: running, 6: cycling, 7: Nordic walking }

4.3.1.2 RWHAAR 数据集

PAMAP2 数据集采集自 15 个人的共计 8 种动作的日常活动场景。每个动作使用单独的时间序列进行记录。具体类别如下：{1: climbing down, 2: climbing up, 3: jumping, 4: lying, 5: running, 6: sitting, 7: standing, 8: walking }

4.3.2 实验环境配置

本章实验的硬件配置如表 4.1:

表 4.1 硬件配置

Table 4.1 Configuration of hardware

设备名	详细属性	空间大小
CPU	Intel(R) Core(TM) i5-9300H CPU @3.20GHz	32.0 GB
GPU	NVIDIA GeForce GTX 1660Ti	6.0GB

实验所使用的深度学习框架, 软件配置等信息如表 4.2 所示。

表 4.2 软件配置

Table 4.2 Configuration of software

名称	详细属性
操作系统	Windows 10
开发语言	Python 3.7
工具包	Pytorch 1.7.0
	Pytorch-lightning 1.2.1

4.3.3 实验结果及分析

将本章提出的 TGAN 与当前主流的时间序列数据增强模型 RGAN 在识别准确率, 训练速度和型大小进行对比。训练流程如下: 首先训练数据增强模型; 其次将原始数据和生成数据一起输入给的识别器, 完成 GAN 的训练; 最后进行分类器的训练。对于 TGAN 使用的是 Transformer 识别器, 对于 RGAN 使用的是 DeepConvLSTM。

4.3.3.1 识别准确率与训练速度

表 4.3 TGAN 与 RGAN 在不同数据集上的性能对比

Table 4.3 Performance comparison of TGAN and RGAN on different datasets

DataSet	Model	Time(Per epoch)	F1-score
PAMAP2	RGAN	192sec	0.962
	TGAN	50sec	0.979
RWHAR	RGAN	476sec	0.918
	TGAN	109sec	0.912

根据实验结果可以看出 TGAN 和 RGAN 在两个数据集上的准确率互有胜负, 但是 TGAN 每轮迭代的速度快于 RGAN, 平均仅需约 30%的时间即可完成训练。

4.3.3.2 模型结构与参数

TGAN 和 RGAN 的模型结构和参数见表 4.4 和表 4.5。其中 TGAN 使用 PReLU 作为损失函数，PReLU 自身需要一个可学习参数。

表 4.4 TGAN 模型结构与参数数量

Table 4.5 TGAN model structure and number of parameters

Layer	Param	参数与结构
-BCELoss: 1-1	总参数: 9,397,202	输入 (MB): 0.01
-Generator: 1-2	训练参数: 8,473,199	前向/反向传播 (MB): 1.20
-Discriminator: 1-3	非训练参数: 924,003	参数 (MB): 37.59
-TransformerClassifier:1-4	加法操作 (M): 19.58	总计 (MB): 38.80
-F1: 1-5		
-Generator: 1-6		

表 4.5 RGAN 模型结构与参数数量

Table 4.5 RGAN model structure and number of parameters

Layer	Param	参数与结构
-BCELoss: 1-1	总参数: 21,270,164	输入(MB): 0.01
-Generator: 1-2	训练参数: 18,264,643	前向/反向传播 (MB): 17.17
-Discriminator: 1-3	非训练参数: 3,005,521	参数 (MB): 85.08
-DeepConvNet:1-4	加法操作 (G): 1.92	总计 (MB): 102.26
-F1: 1-5		

RGAN 因为编码层完全由卷积操作组成而与需要非常多的参数，transformer 需要的卷积操作相对少很多，从加法操作上来看 TGAN 和 RGAN 分别需要 19.58M 和 1.92G 次，另一方面 RGAN 中的 LSTM 层不能并行计算，也极大地增加模型训练所需要的时间。此外 TGAN 模型相对 RGAN 模型更小，在嵌入式设备如手环，智能手机上可以更方便地部署。

4.3.3.3 不同分类器对于识别结果的影响

识别结果的混淆矩阵如图 4.5 和图 4.6 所示：

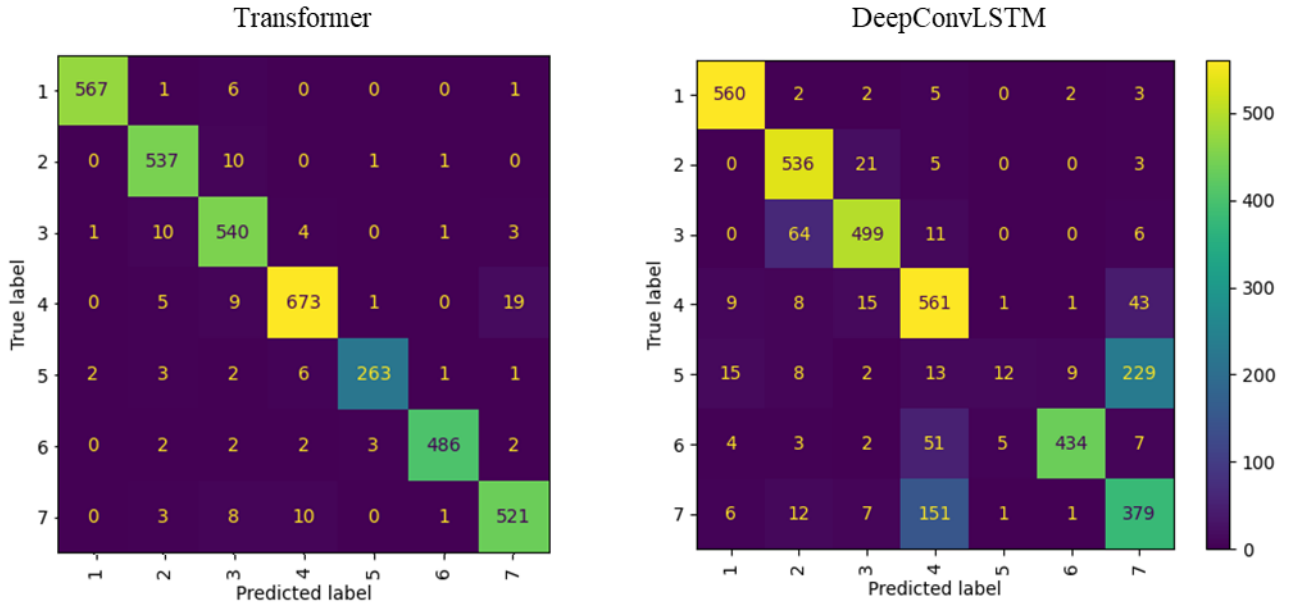


图 4.5 Transformer 和 DeepConvLSTM 在 PAMAP2 上结果的混淆矩阵

Fig. 4.5 Confusion matrix of Transformer and DeepConvLSTM results on PAMAP2

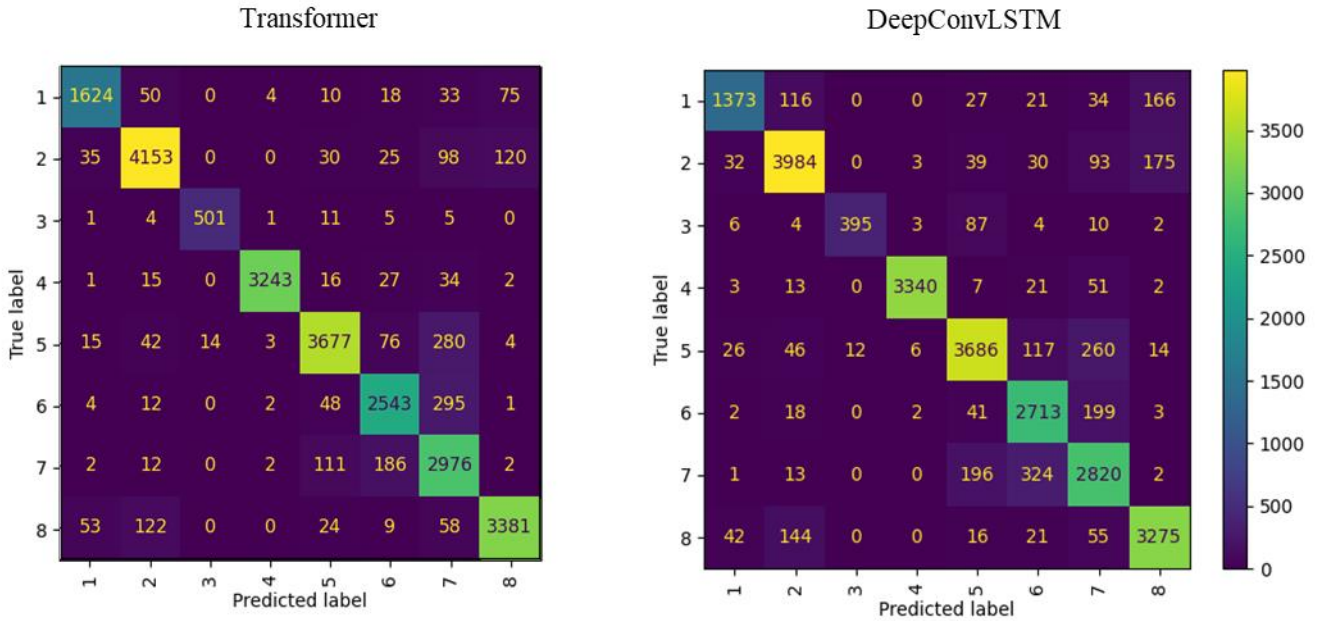


图 4.6 Transformer 和 DeepConvLSTM 在 RWTHAR 上结果的混淆矩阵

Fig. 4.6 Confusion matrix of Transformer and DeepConvLSTM results on RWTHAR

根据混淆矩阵可以看出 Transformer 在多元分类任务上优于 DeepConvLSTM。在实验的初期，我们认为 Transformer 在分类效果上会全面优于 DeepConvLSTM，但通过实验发现并不是这样。在尝试互换 TGAN 和 RGAN 所使用的最终分类器后，发现两者的准确率均有所下降。DeepConvLSTM 在 RGAN 上表现更好，而 Transformer 在

TGAN 上表现更好。这也许是因为 GAN 会去学习那些能够通过验证任务的数据分布，RGAN 的生成器和判别器部分整体结构与 DeepConvLSTM 非常相似，两者很容易相互匹配；TGAN 也是同理，基于 Transformer 的分类器，在依托于 Transformer 的编码解码层上更适用。

4.3.3.4 数据增强展示

最后在以 PAMAP2 为例展示一下数据增强的效果。

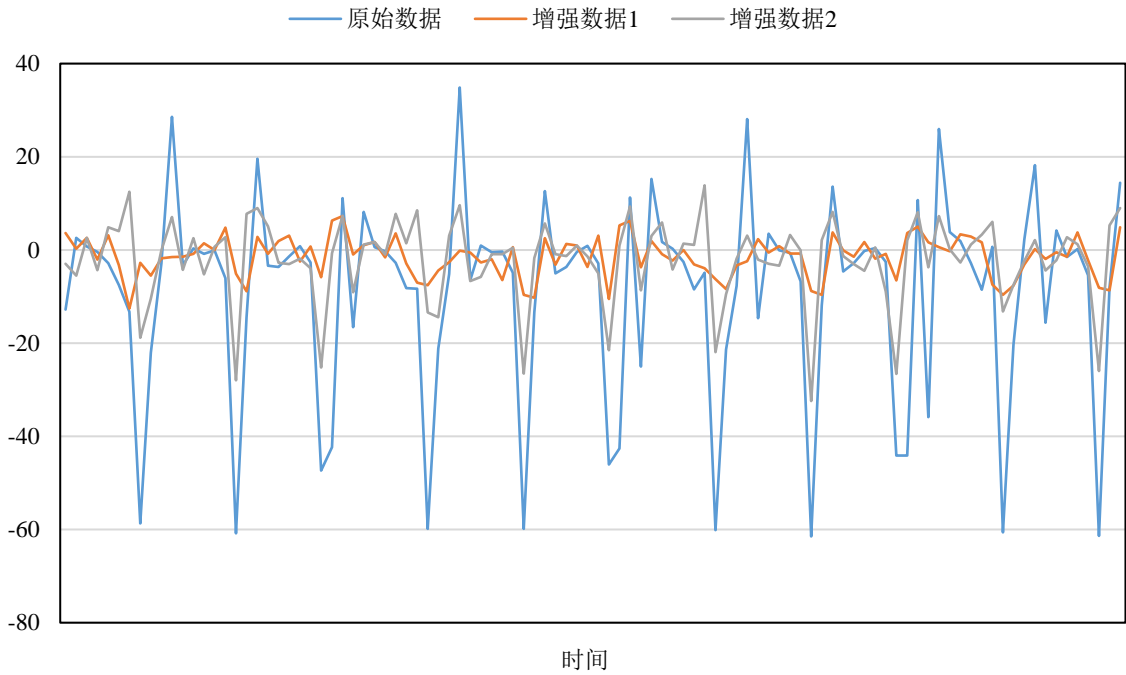


图 4.7 PAMAP2 中 cycling 的原始数据与增强数据

Fig. 4.7 Raw and enhanced data of cycling in PAMAP2

上图以 PAMAP2 中的 cycling 动作为例，展示了数据增强的效果，可以看出虽然在振幅上与原始数据相比有所减弱，但是模型可以很好地捕捉到原始数据的变化趋势，这也是数据增强的意义所在：生成与原始样本属性相同但不完全一致的样本来增加样本的多样性。与传统的数据增强方法如 TWP 相比，基于生成模型的方法生成的样本能保留每种序列的特性。如果想要多样化的数据，传统方法往往需要对每种类别单独设定参数，而使用生成模型，通过对隐空间向量的控制，可以更方便的，大量的生成增强样本供后续训练。图 4.7 就是通过在编码层得到的隐向量加上四倍高斯噪声得到的增强结果。由于噪声的随机性，每次生成的样本都不一样，如果想要不同的波动效果，可以控制高斯噪声的倍数来实现。这两个增强数据可以组成正样本对供对比学习模型训练。根据图 2.9，假设每个 Batch 中共 N 样本，我们可对每个样本生成两

个增强样本，共 $2N$ 个样本，每个样本的两个增强样本构成正样本对，而某个样本的增强与任意其他样本的增强样本构成负样本对。由此构建的正负样本对可供后续的对比学习框架使用。

4.4 小结

本章针对活动识别数据集规模小和活动数据的周期性导致的正样本对构造困难的问题，基于活动数据本身的时序特性和生成样本的多样性需求，提出了基于 Transformer 的生成模型 TGAN，并提出对角 mask 在解码过程中更好地利用双向信息。设计了对比实验与目前效果较好的深度学习方法 RGAN 进行了比较，用更小的模型和更快的训练速度，得到了不输于 RGAN 的效果，并对 GAN 学习的数据分布和不同分类器的准确率之间的关系进行了简要的讨论。最后展示了数据增强的效果并介绍了在后续对比学习中的样本生成方式。

第5章 基于对比学习的人体活动识别预训练模型

本章首先对活动数据的连续性问题 and 周期性问题带来的正负样本对构造问题进行了分析，然后结合前两章研究内容，提出完整的基于对比学习的人体活动识别预训练模型，并完成了系统的设计实现。最后通过对比实验验证了预训练模型的有效性。

5.1 问题分析与提出

在活动识别领域，带有高质量标签的数据非常难以采集，但随着传感器硬件和采集技术的进步，仅仅采集活动数据而不对其打标签是一个成本相对较低的任务。针对这个问题，本文提出一种适用于活动数据的无监督预训练方法：在无标签数据上训练出适用于各种下游任务（主要是分类任务）的特征提取器。

对比学习作为近两年新提出的无监督或称为自监督的表示学习方法，在 CV 和 NLP 领域取得了巨大的成功，但是将对比学习方法直接应用到时间序列领域上的效果并不好，问题主要出在正负样本对构造这个环节。这主要是由下面两个原因造成的：

（1）活动数据连续性过强不是自然可分的

在 CV 领域，每个图像独立成帧，在 NLP 领域，每个单词自然可分。在选取负样本时，一般选取另一张图片或者不同位置的单词即可。但是时间序列不是天然可分的，尽管有滑动窗口技术用来划分样本，但是划分出来的样本之间连续性过强，不利于负样本的选取。需要一种技术来将原始时间序列划分成若干段，在选取给定样本的负样本时，我们从该样本所在段的旁边两段中选取负样本。

（2）活动数据的周期性过强

在日常活动中，如走路，跑步等动作的周期性是非常强的，如图 5.1 所示，很可能出现几个甚至几十个动作的波形非常相似。当正样本的选择选定为两个非常相似的序列片段时，对比学习的“Triple-loss”会退化为“binary-loss”，如公式（5.1）所示，即 Positive sample 部分的损失已经非常小，对于一个样本对来说优化的难度变小了，但是对于整个样本空间来说，模型的表达能力却减弱了，尤其在测试时，由于每个受试者的行为模式不同，更难以将测试样本对映射到特征空间上恰当的位置上。根据“均匀性原则”，我们应该将特征尽量均匀的铺满特征空间，由于样本之

间的相似性过高，容易导致特征空间内的映射无法均匀地铺满整个特征空间，如图 5.2（左图），从而丢失了很多信息。如果不能在训练过程中尽可能地掌握各种动作的数据分布，模型就无法学习到正确的决策边界。由于受试者的行为习惯不同，很容易在测试的时候产生误判。如果特征可以均匀地分布在特征空间，学习到的决策边界也会更真实。综上需要一种数据增强技术来生成给定样本的正样本对，以增加正样本的多样性，使得模型可以更好的学习到数据分布。

$$L_i = -\log\left(\frac{\exp\left(\frac{1}{\tau}\right)}{\sum_{j=0}^K \exp\left(\frac{s(z_i, z_j)}{\tau}\right)}\right) \quad (5.1)$$

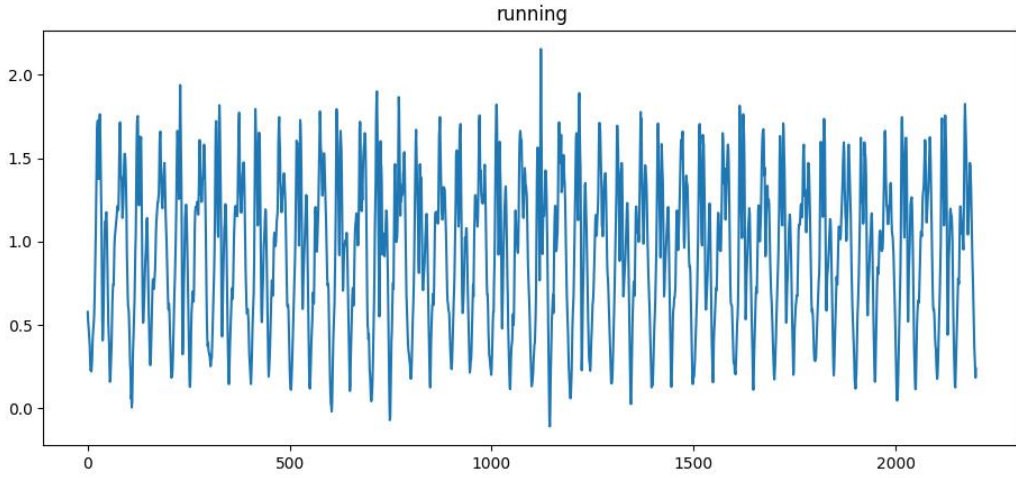


图 5.1 人体活动强周期性

Fig. 5.1 Strong periodicity of human activity

目前应用于活动识别的无监督预训练方法尚未有成熟的应用。针对上面的两个问题，本章以前两章的算法为基础，搭建完整的基于对比学习的活动识别预训练模型 CL-HARP。

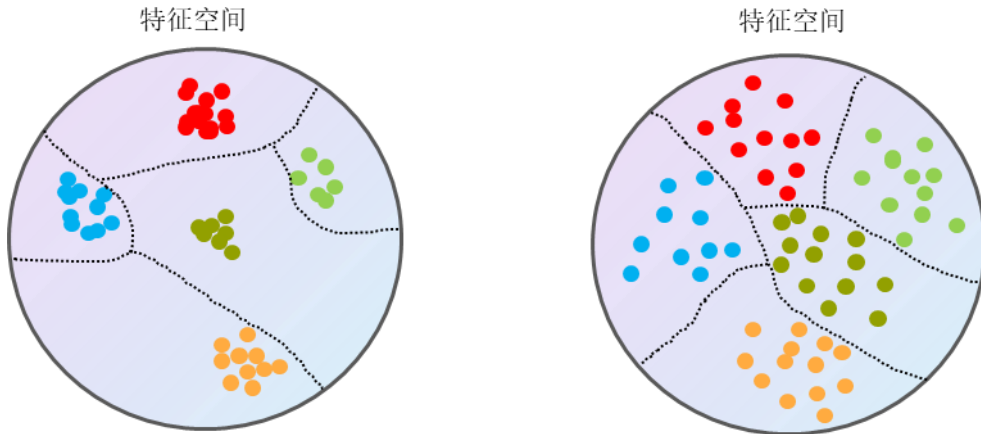


图 5.2 对比学习下的特征分布与决策边界

Fig. 5.2 Feature distribution and decision boundary under contrastive learning

5.2 模型框架

整体的模型框架如图 5.3 所示。

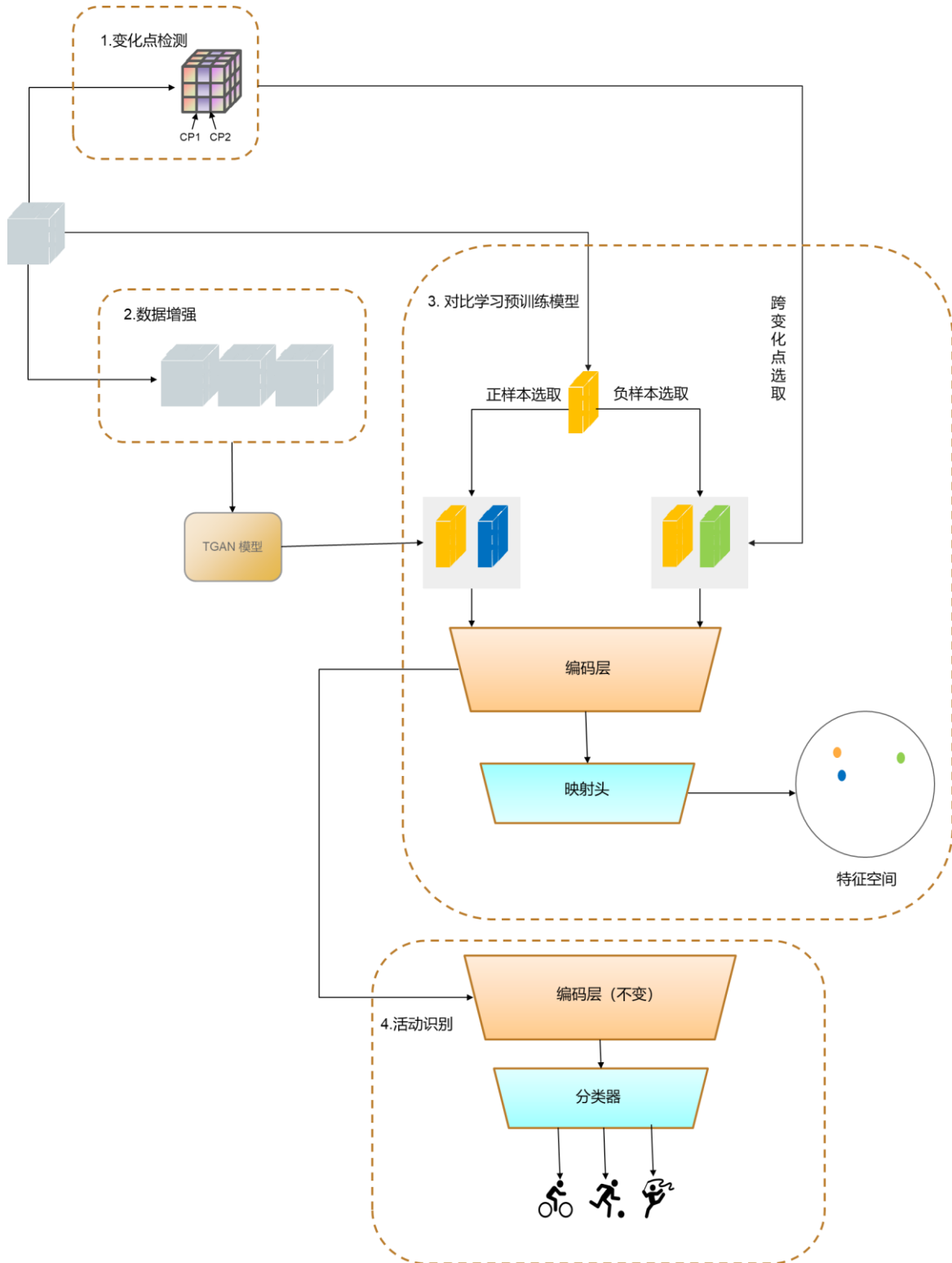


图 5.3 基于对比学习的人体活动识别预训练与识别模型框架

Fig. 5.3 Framework of Human Activity Recognition Pre-training and Recognition Model Based on Contrastive Learning

该框架由四部分组成，分别为变化点检测模块，数据增强模块，对比学习预训练模块，活动识别模块。

5.2.1 变化点检测模块

拿到无标签的原始数据后，首先需要将其分割为彼此不同的片段。如果原始序列有多条序列，则将他们拼接到一起，再一起进行分割。具体的算法是第三章中提出的 CL-CPD 算法。在完成分割后，相当于给每个动作打上了“弱标签”，尽管这种弱标签有误判的可能，但是注意在后续的负样本选取时，这种弱标签的作用主要是防止选取到同属性的样本，所以我们对“假阴性”的预测结果可以有较高的容忍度。

5.2.2 数据增强模块

数据增强模块的主要任务是学习到各种动作对应序列的分布。这个模块可以与变化点检测模块并行，使用的模型的第四章提出的 TGAN。在实际生成样本时，并不直接从隐空间向量出发，而是先将要增强的片段映射到隐空间，再进行还原，如果需要多个样本，可以在该向量上加上随机噪声。模型训练完成后，将效果最好的参数进行保存，以便于在预训练模型中使用。

5.2.3 对比学习预训练模块

在上面两个模块都完成训练后，可以对基于对比学习的预训练模型进行训练。

目前比较成熟的对比学习框架主要可以分为 MoCo、SimCLR 和 BYOL 三大类。其中 MoCo 和 BYOL 借鉴了 Memory bank 的思想，需要大量内存用来存储动量编码器，另一方面由于活动识别数据集普遍偏小，并不能提供太多的 batch 以供训练，batch 动量更新效果会受到影响，若 batch size 设置的过小，又会影响训练速度和 batch normalization 的效果。因其结构简单，SimCLR 框架的训练速度相较其他框架也更快一些。综上本节的预训练模型基于 SimCLR 搭建。

对于负样本的选取，我们根据变化点检测模块的结果，从该序列所在片段两边相邻片段中随机选取序列。对于正样本的选取，我们使用已经训练好的 TGAN 模型来生成该序列的增强数据。为了增强模型的学习能力，我们可以为一个样本选取多个正样本和多个样本负样本，让“1vs1”变为“多 vs 多”。而损失函数由原始的公式 (5.2) 变为公式 (5.3)。

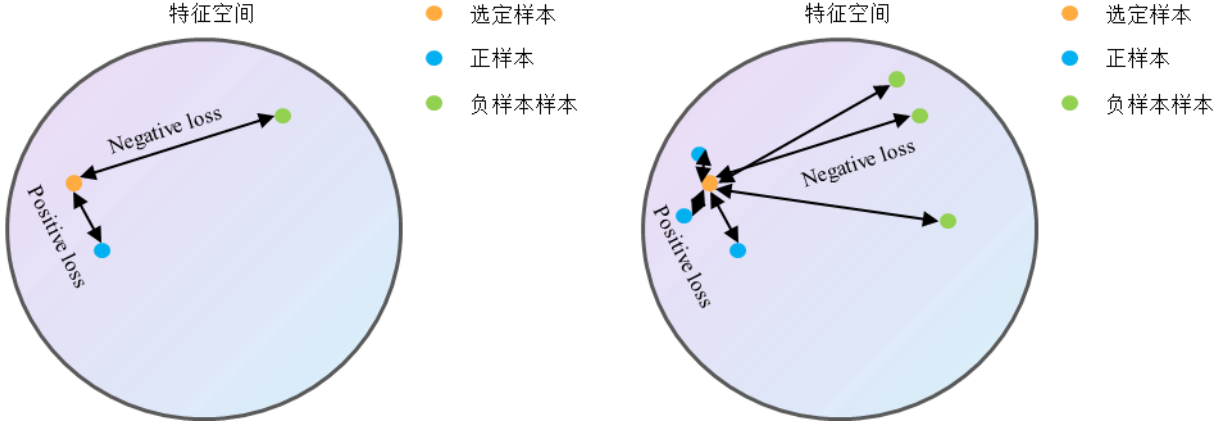


图 5.4 1vs1（左）与多 vs 多（右）在特征空间的示意图

Fig. 5.4 “1 vs 1”(left) and “multi vs multi”(right) in feature space

$$L_i = -\log \left(\frac{\exp\left(\frac{s(z_i, z_i^+)}{\tau}\right)}{\exp\left(\frac{s(z_i, z_j^-)}{\tau}\right) + \exp\left(\frac{s(z_i, z_i^+)}{\tau}\right)} \right) \quad (5.2)$$

$$L_i = -\log \left(\frac{\sum_{l=0}^{K^+} \exp\left(\frac{s(z_i, z_l^+)}{\tau}\right)}{\sum_{j=0}^{K^+} \exp\left(\frac{s(z_i, z_j^+)}{\tau}\right) + \sum_{j=0}^{K^-} \exp\left(\frac{s(z_i, z_j^-)}{\tau}\right)} \right) \quad (5.3)$$

构建完整负样本对后，将正负样本对输入给编码层，在本实验中使用的是 CNN 网络，整体结构与第 3 章相似，这里不再详细介绍。最后一层是映射层，用来将编码层的输出映射成指定维度的向量。

5.2.4 活动识别模块

在预训练完成后，我们可以将其应用在不同的下游任务上，在本实验中下游任务即分类任务。我们将保留与训练模型除最后映射头之外的所有模型结构与参数，在编码层后重新训练一个分类器。这个分类器结构不必太复杂，普通的全连接层即可胜任分类任务，因为在预训练过程中，已经对特征进行了提取，在这里我们只需要找决策边界即可。

5.3 系统功能与架构

5.3.1 需求分析与功能

本文需要设计并实现人体活动识别预训练与识别系统。第一，本系统提供不同用

户的个性服务，同时实现基本数据信息共享，因此用户的注册和基本信息管理功能是必须的；第二，模型需要数据作为基础才能进行训练，除了系统默认支持的几种数据集，还需要提供数据集的上传下载和预处理等基本管理功能；第三，用户需要对数据集的样本类别分布有所了解，因此需要提供数据查询与可视化功能；第四，模型中的两个核心算法变化点检测和数据增强在系统中要有对应的独立功能；第五，本系统的核心功能预训练模型；最后，作为一个最终任务是活动识别的系统，最重要的还是完成识别任务。整个系统的功能如下：

（1）用户注册

作为需要提供对外服务的系统，用户注册功能是必不可少的。为了给每个用户提供个性化的服务，用户在首次登录系统时，需要提供基础信息。在之后的使用中，可以通过用户名和密码登录自己的账户。

（2）数据管理

数据管理功能包含数据上传、下载和预处理的功能。用户可以将自己的数据上传至系统，并在上传时添加数据集的传感器个数、建议窗口长度、采样频率和动作类别名称等基本信息。服务器按照数据集的基本信息对数据进行格式化处理并存储，并生成 MD5 码作为该数据集的唯一标识。对于 MD5 码相同的数据集，服务器存储一份即可，不同用户不必重复上传，可以使用服务器提供的副本。

（3）数据查询与可视化

数据查询与可视化功能包括按照指定类别、指定时间段、指定传感器查看数据的功能，并对查询结果进行可视化，让用户对数据有更直观的认识。

（4）变化点检测

变化点检测是本系统的核心功能之一。该功能是对第 3 章提出的算法的封装，在训练完针对指定数据集变化点检测模型后，系统可以对该序列数据进行分割，并根据判定边界的不同对检测结果进行评价，同时通过可视化的方式与真实的变化点进行对比。该模型在无标签，即无监督下完成训练。系统将训练好的模型存储起来，用于后续的预训练模型搭建。

（5）数据增强

数据增强是本系统的另一个核心功能。该功能是对第 4 章提出的算法的封装。在训练完针对指定数据集的数据增强模型后，系统可以对该序列数据进行数据增强。用户可以指定要增强的类别，系统会将原始序列和增强序列进行可视化来检验增强效果。该模型在无标签，即无监督下完成训练。系统将训练好的模型存储起来，用于

后续的预训练模型搭建。

（6）预训练模型

搭建预训练模型是本系统的最重要的任务，该功能是基于前两个模块实现的。具体的结构如图 5.3 所示。在完成针对指定数据集预训练模型后，用户可以通过指定类别的方式来对预训练模型结果进行降维可视化。系统将训练好的预训练模型存储起来，用于后续的任务。

（7）活动识别

活动识别是本系统的最终任务。实际上预训练模型可以应用于各种不同的下游任务上，但在本系统中，我们主要关注活动识别任务。系统使用在该数据集上已经完成的预训练模型进行活动识别，并对识别结果进行分析。

5.3.2 系统架构

本小节首先对系统的整体架构进行了设计，整体系统功能的开发采用了前后端分离的模块化设计方法。该系统通过浏览器完成数据的上传和下载，然后将数据交给服务器进行处理并并存入数据库，其它功能模块对数据的操作均在服务器端完成。最后将各模块的结果发送给客户端/浏览器进行展示。图示系统的整体架构。

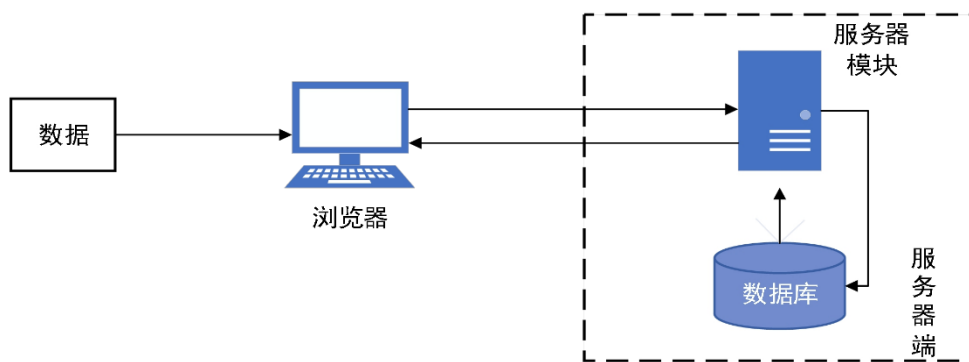


图 5.5 系统架构

Fig. 5.5 system structure

5.3.3 系统的模块划分

系统整体上划分为三个主要功能模块：

（1）数据管理模块。数据管理模块分为数据传输、数据预处理和数据查询三个功能子模块。主要用于使用浏览器将数据文件从客户端传输至服务器端，同时可以将其从服务器传输至客户端；并根据用户上传的数据元信息将数据格式化后存入数据库。同时为了便于用户对数据进行可视化查看，系统提供数据查询功能。用户通常难

以对原始的传感器数据有直观的认识，查询模块通过用户提供的动作类别和时间段对数据进行筛选，并以图像的形式进行展示。

(2) 预训练模块。预训练模块分为变化点检测、数据增强和预训练模型三个子功能模块。变化点检测模块主要使用 CL-CPD 算法处理数据，将检测结果展示给用户，同时提供不同判定边界下的结果，并将训练好的模型保存起来。数据增强模块主要使用 TGAN 模型处理数据，将增强结果展示给用户，并将训练好的模型保存起来。预训练模型 CL-HARP 基于前两个子模块已经训练好的模型进行搭建：利用 CL-CPD 的检测结果进行负样本选取，利用 TGAN 模型进行数据增强继续进行正样本选取，完成训练后，将最终的结果降维展示给用户，并将模型存储起来用于下游任务。

(3) 活动识别模块。活动识别模块针对选定的数据集划分训练集、测试集和验证集，利用已经训练好的 CL-HARP 模型构建分类器，并将最终的分类结果展示给用户。

5.4 系统的设计实现

本文提出的系统主要包含有数据管理、预训练和活动识别模块。本节将主要介绍数据管理、预训练和活动识别模块，数据查询作为系统的辅助模块，将不再进行详细介绍。

5.4.1 数据管理模块

系统的整个运行都以数据作为基础，本模块主要用于数据的上传下载和预处理，该模块的整体设计如图 5.6 所示。



图 5.6 数据管理模块设计图

Fig 5.6 Design diagram of data management module

目前开源的人体活动数据集的格式通常是以文件形式提供的。根据传感器的不同，采集的传感器数据常用的有三轴加速度和三轴角速度。受限于模型的输入，系统需要对用户提供的数据格式进行基本的校验，只有满足一定的条件才能存入数据库并进行后续的处理。表 5.1 说明了用户上传的数据格式。

表 5.1 数据格式

Table 5.1 Data structure

Key		Data						
User_id	Timestamp	Acc_x	Acc_y	Acc_z	Gyro_x	Gyro_y	Gyro_z	Label

其中 Timestamp 为非必须项，三轴加速度{Acc_x, Acc_y, Acc_z}和三轴角速度{Gyro_x, Gyro_y, Gyro_z}数据至少有一项不全为空，Label 为必须项。数据通过以上校验后会按照 User_id 和 Label 进行拆分：一级目录为数据集名的文件夹，二级目录以各动作的标签作为文件夹名，每个文件夹内按照 User_id 进行划分为若干个 csv 文件，每个文件中的数据格式同表 5.1。最终结构如图 5.7。

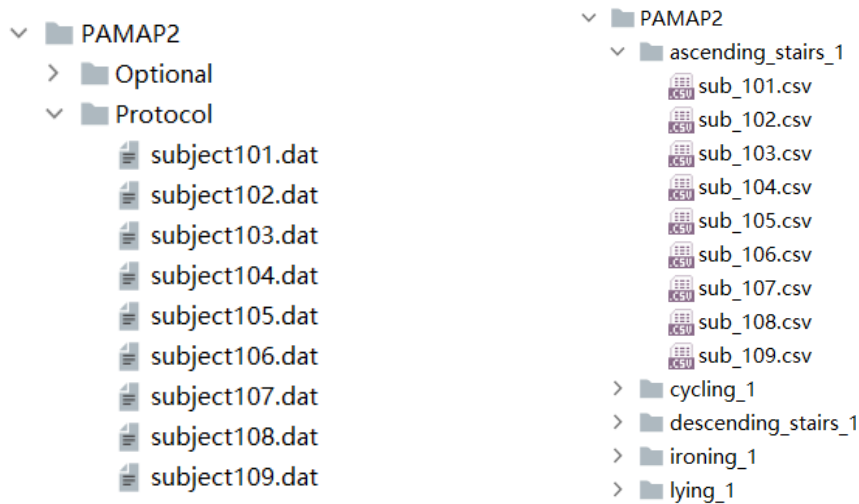


图 5.7 处理前（左）和处理后（右）的文件组织结构

Fig. 5.7 File organization structure before (left) and after (right) processing

此外，系统还会为每一个原始文件生成一个 MD5 码，当用户上传某个文件时，我们先在浏览器端对文件的 MD5 码进行校验。将 MD5 码上传后，系统扫描数据库中所有的 MD5 码，如果数据库中已经有相同 MD5 码的数据文件，则浏览器不再进行重复上传与存储，这样同时节省了通信和存储开销。该模块的整体流程如图 5.8 所示。

在经过上述处理后，将数据文件存储在本地磁盘上，并将文件名、文件的存储路径和 MD5 码存储在 MySQL 中。当用户需要对某个数据进行查看时，根据文件名在 MySQL 中查询对应的存储路径，并按照时间和动作标签进行可视化展示。

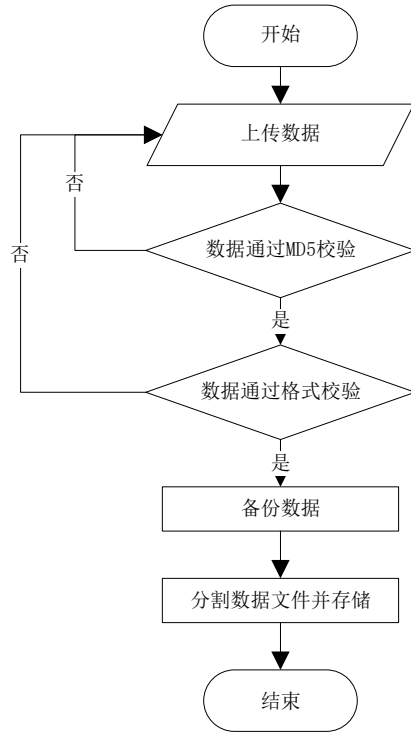


图 5.8 数据管理模块流程图

Fig. 5.8 Flow chart of data management module

5.4.2 预训练模块

预训练模块是系统中最重要功能模块，该模块又分为变化点检测、数据增强和预训练三个子功能模块，模块的整体流程如图 5.9 所示。其中变化点检测和数据增强子功能模块是对 CL-CPD 模型和 TGAN 模型的封装

首先按照文件名从本地磁盘加载数据集。然后并行训练该数据集的 CL-CPD 模型和 TGAN 模型。如图 5.3 所示，根据 CL-CPD 的判断结果和 TGAN 的增强结果构建对比学习模型 CL-HARP 需要的正负样本对。对于某个具体的样本，我们要在该样本的左右两个 CP 之间的区间内的样本进行数据增强，并将结果作为正样本使用。而负样本则从相邻的区间中样本数据增强结果进行选取。最后将选定的样本构建成正负样本对集作为预训练模型的输入。

上述准备步骤完成后，利用构建的正负样本对集进行预训练模型的训练，模型架构才用 SimCLR。在完成训练后，将模型的最终输出向量即为原始样本的特征向量。我们可以使用降维算法如 PCA、t-SNE 等来完成特征向量的可视化。

预训练模型训练完毕后，将最后一层的映射头移除，并将编码层的模型结构和参数进行保存，用于后续的分类任务使用。

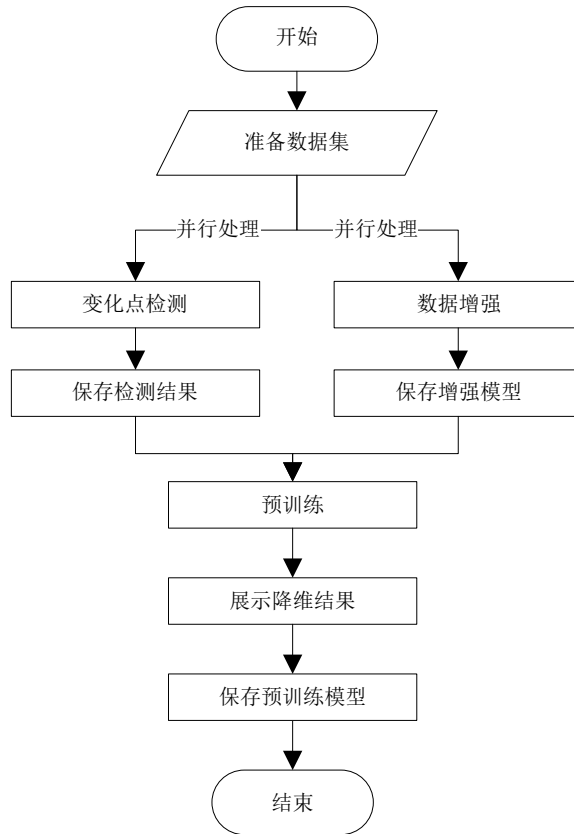


图 5.9 预训练模块流程图

Fig. 5.9 Flow chart of pre-trained module

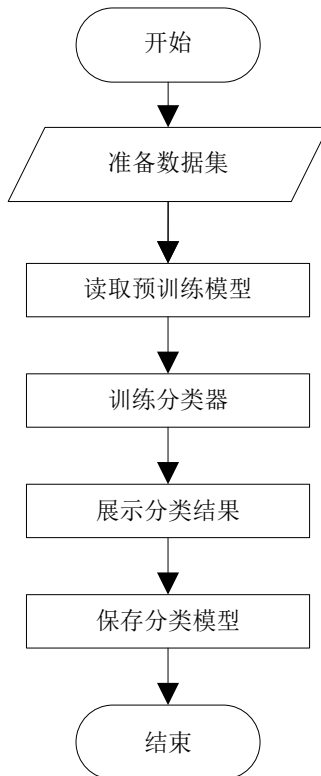


图 5.10 活动识别模块流程图

Fig. 5.10 Flow chart of activity recognition module

5.4.3 活动识别模块

在完成预训练模块后，我们可以在预训练模型的基础上搭建分类器，并完成活动识别任务。活动识别模块的流程如图 5.10 所示。

本文在预训练模块保存的模型的基础上，重新训练一个分类头，在训练的过程中，冻结预训练部分的模型参数，只对最终的分头中的参数进行训练。对于一个具体的样本，模型输出该动作属于每种动作的概率，并将最大概率的动作作为预测结果。

5.5 实验分析

5.5.1 数据集

5.5.1.1 Opportunity 数据集

Opportunity 数据集同时包含可穿戴传感器和周围环境的传感器。本章实验中仅使用可穿戴传感器中的 3 轴加速度传感器进行活动识别，其在人体的分布如图 5.11 所示。该数据集共采集自 4 个受试者 (S1, S2, S3, S4)，每个受试者采集 6 条活动数据，其中 5 次是受试者在自然状态下的一系列日常行为 (ADL)，1 次是受试者在受限条件下执行事先规划好的一系列活动 (Drill)。每个受试者的 ADL 活动数据之间差异比较大，而 Drill 活动数据之间的差异相对小很多。数据集共包含 {Locomotion, HL_Activity, LL_Left_Arm, LL_Left_Arm_Object, LL_Right_Arm, LL_Right_Arm_Object, ML_Both_Arms} 8 个数据集，其中 {Locomotion, HL_Activity, LL_Left_Arm, LL_Right_Arm, ML_Both_Arms} 这 6 个子集是基于可穿戴设备采集的数据，而 {LL_Right_Arm_Object, LL_Left_Arm_Object} 这两个子集则和在实验环境中可交互的环境传感器相关。在本实验中，仅采用了基于可穿戴设备的子集。在本实验中，仅使用了基于可穿戴设备的子集。Locomotion 包含 4 种基础运动模式 {走路，站立，静坐，平躺}，LL_Left_Arm 与 LL_Right_Arm 包含 13 种中层次运动模式 (开关抽屉、门等)，HL_Activity 包含 5 种抽象层次较高活动 (吃三明治，做早操等)。Locomotion 和 HL_Activity 子任务的所对应的数据集具体信息如表 5.2 和表 5.3。

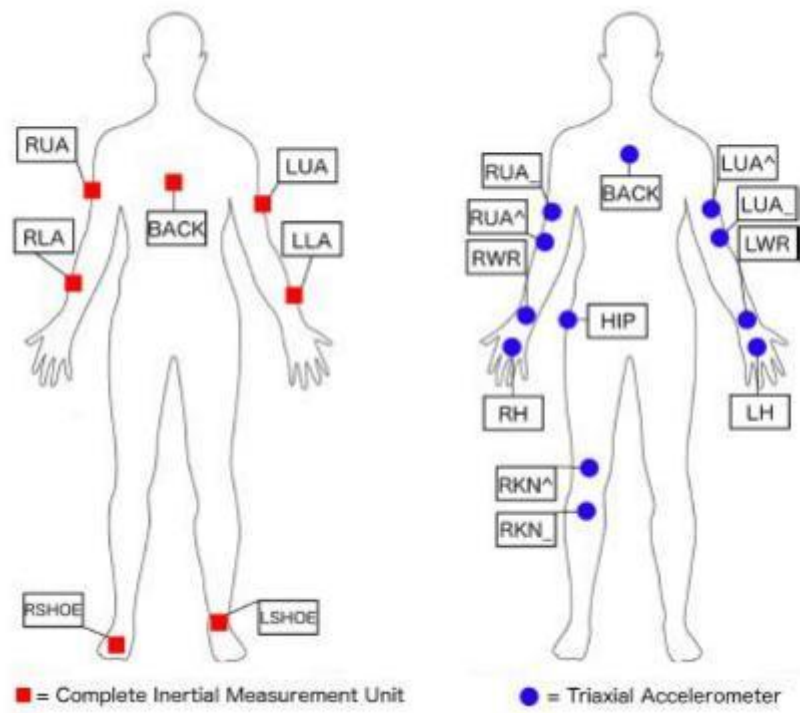


图 5.11 Opportunity 数据集传感器位置

Fig. 5.11 Sensor locations opportunity dataset

Locomotion 任务包括的 4 种活动状态，描述的是受试者整体所处的运动状态，包含的语义信息较少，抽象层级较低，在有足够的样本的情况下，模型训练难度较低，且最终的准确率较高。

HL_Activity 任务对应的 5 类活动，是抽象层次很高的动作，即在长时间内活动的一个总结，如喝咖啡是一个长达十几秒的，一系列动作的组合，包含很多的低层次信息，与 ML_Both_Arms 有较强的相关性。在不使用可交互物体上的传感器数据的情况下，想要实现较高的分类准确率是具有一定难度。

表 5.2 Locomotion 任务数据集样本分布

Table 5.2 Locomotion task dataset sample distribution

活动类别	样本数
Stand	19203
Walk	11282
Sit	8077
Lie	1435

表 5.3 HL_Activity 任务数据集样本分布
Table 5.3 HL_Activity task dataset sample distribution

活动类别	样本数
Relaxing	2584
Coffee time	5138
Early morning	7935
Clean up	4231
Sandwich time	9986

5.5.1.2 PAMAP2 数据集

本节中使用的 PAMAP2 数据集与第 4 章实验相同，详见 4.3.1.1。

5.5.2 实验结果

5.5.2.1 变化点检测

本节补充了 CL-CPD 算法在 PAMAP2 和 Opportunity 数据集的 Locomotion 子任务上的实验结果，如表 5.3 和表 5.4 所示。

表 5.3 CL-CPD 在 PAMAP2 上的 F1-score
Table 3.3 F1-score of CL-CPD on PAMAP2

Method	DM100		DM200		DM400	
	window	F1	window	F1	window	F1
CL-CPD	400/128	0.8303	400/128	0.8303	400/128	0.8834

表 5.4 CL-CPD 在 Opportunity 上的 F1-score
Table 3.3 F1-score of CL-CPD on Opportunity

Method	DM100		DM200		DM400	
	window	F1	window	F1	window	F1
CL-CPD	400/64	0.5226	400/64	0.5136	400/64	0.5740

PAMAP2 数据集因为预设动作占比高，所以采集到的数据有比较明显的分割点，CL-CPD 取得了较高的 F1-score；而 Opportunity 数据集受试者的自主程度更高，在不同种类动作之间的切换更随意，因此检测的准确率较低。

5.5.2.2 预训练模型

在完成预训练模型后，将映射头输出的向量利用 t-SNE 降维算法映射到二维空间上，效果如图 5.12 所示。

从 8 种动作中随机选取了 6 种动作进行降维映射，可以看出同种动作在二维特征空间的映射之间相对靠近，而不同动作的映射之间相对远离，这与对比学习的初衷相吻合。对比学习中的“对齐性”原则在上图中有较好的体现，而对于“均匀性”原则，6 种随机动作已经可以填满大部分特征空间，可以胜任后续的分类任务，可以认为本文的预训练模型可以提取处各个类别的高层次特征。

5.5.2.3 活动识别

在前面的预训练模型基础上进行活动识别，对于分类器的选择，因为预训练模型已经可以很好地提取特征了，本节实验全连接层来构建分类器，并将实验结果与 DeepConvLSTM, RAE 等经典活动识别模型进行比较，其他方法的结果来自综述[68]，具体方法对应的文献见图表。对于 Opportunity 数据集的 HL_Activity 任务，由于其抽象层级较高，对该任务的研究相对较少，本节只选取 DeepConvLSTM 和 LHNNet 进行对比实验。

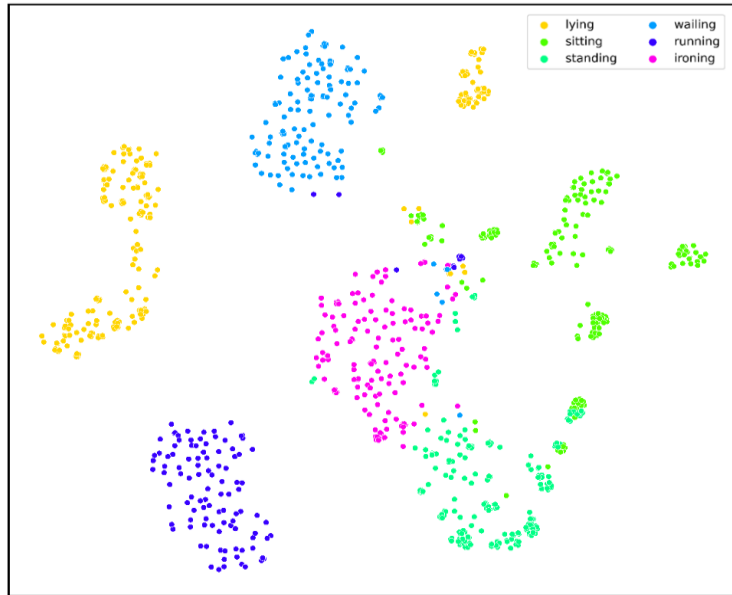


图 5.12 PAMAP2 预训练模型输出向量经 t-SNE 降维后的结果

Fig. 5.12 The result of t-SNE dimensionality reduction of the output vector of the PAMAP2 pre-training model

模型在 Opportunity 和 PAMAP2 数据集上的训练损失如图 5.13 所示。因为

PAMAP2 数据集采集的动作比较简单,所以模型收敛的较快,在设定的 200 个 epoch 之前已基本收敛,且损失的波动较少;而 Opportunity 数据集的 HL_Activity 任务抽象层级高,每一种动作由多种子动作组成,比较难以训练,模型收敛速度较慢,且训练损失有较大的波动。

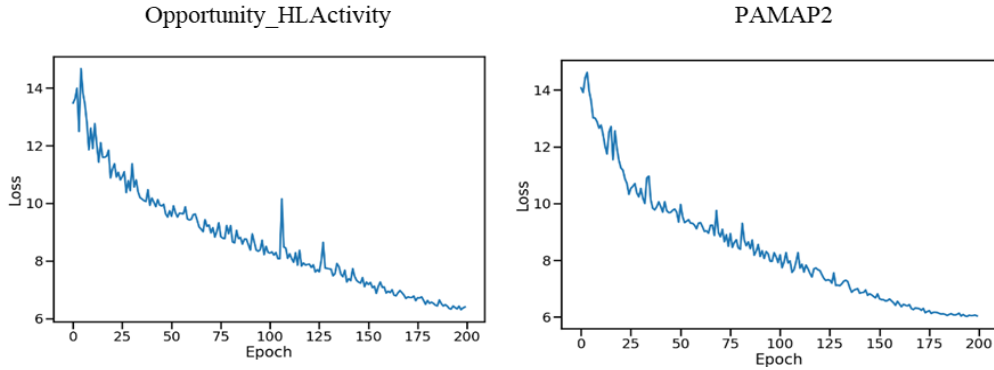


图 5.13 PAMAP2 (左) 和 Opportunity_HLActivity (右) 训练损失

Fig. 5.13 Training loss in PAMAP2(left) and Opportunity_HLActivity(right)

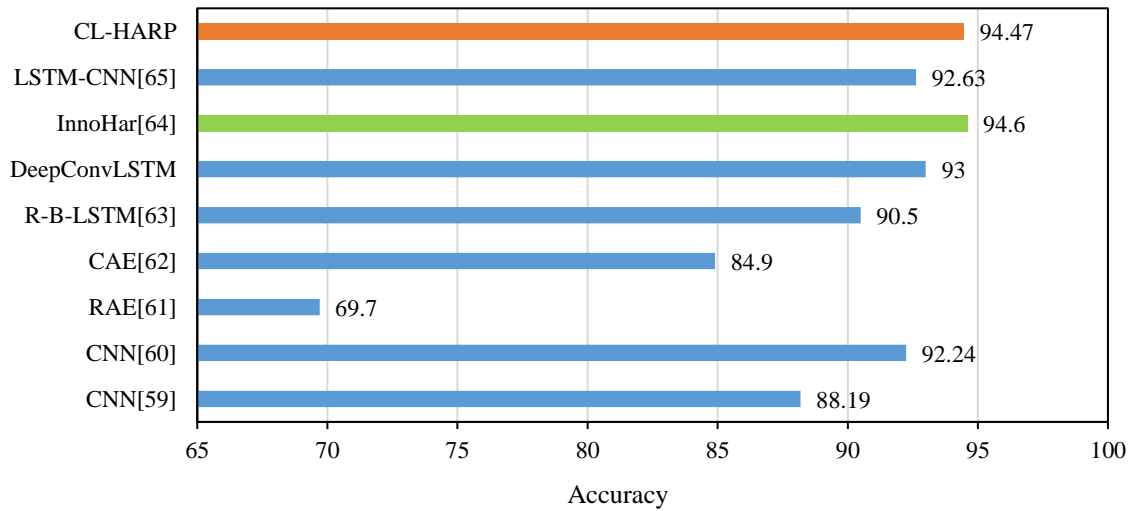


图 5.14 CL-HARP 与其他方法在 Opportunity Locomotion 任务的分类准确率

Fig. 5.14 The classification accuracy of CL-HARP and other methods in the Locomotion task

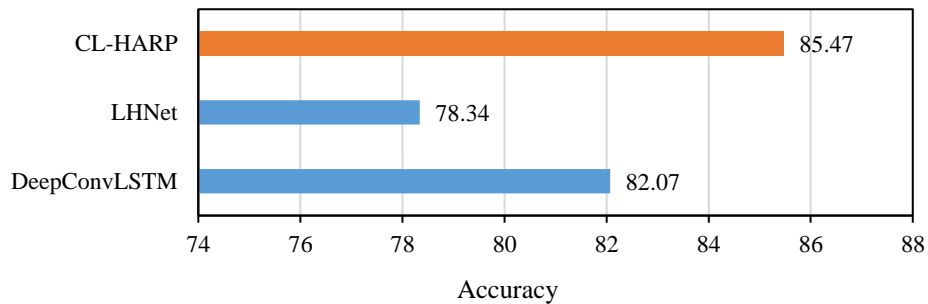


图 5.15 CL-HARP 与其他方法在 Opportunity HL_Activity 任务的分类准确率

Fig. 5.15 The classification accuracy of CL-HARP and other methods in the HL_Activity task

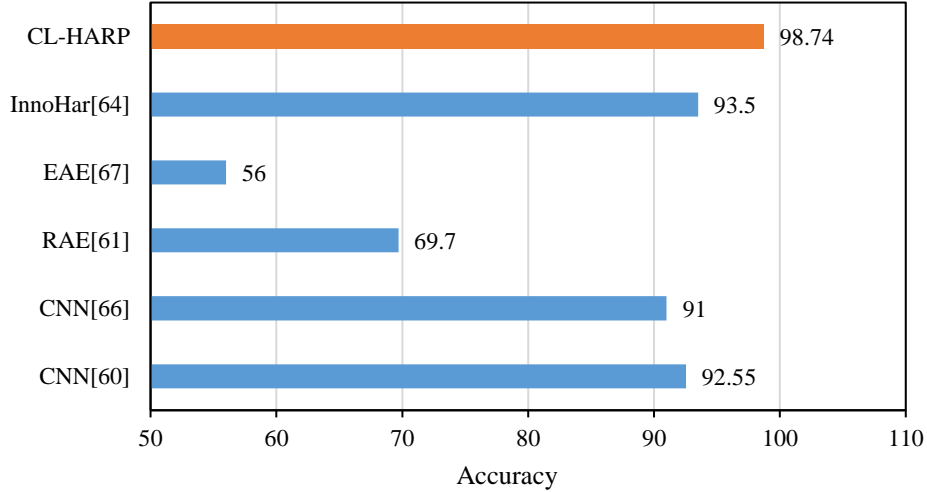


图 5.16 CL-HARP 与其他方法在 PAMAP2 的分类准确率

Fig. 5.16 The classification accuracy of CL-HARP and other methods in the PAMAP2

从图 5.14、5.15 和 5.16 可以得出，在 Opportunity 的 Locomotion 任务上，CL-HARP 均取得了次高的准确率，InnoHar 模型通过堆叠多层的 Inception Network 和 RNN 模块取得了最高的准确率，但参数庞大，且不能并行训练。CL-HARP 虽然在准确率上相较 InnoHar 低了 0.13%，但是其并行训练的优势可以更充分的利用硬件资源，在移动端部署上更具优势。

5.6 小结

本章在 SimCLR 的基础上提出基于对比学习的人体活动识别预训练模型 CL-HARP，结合前两章提出的 CPD 算法和数据增强的算法进行正负样本构建，并提出了多对多的损失函数，可以更高效地利用样本。通过预训练特征的降维结果可以看出，CL-HARP 可以通过预训练过程充分地从无监督数据中学习到数据分布。通过在 PAMAP2 和 Opportunity 的两个子任务上的对比实验表明，CL-HARP 在下游分类任务上的准确率不输于主流的监督学习方法。

第6章 总结与展望

6.1 总结

随着传感器技术与深度学习技术的发展, 基于传感器的人体活动识别问题受到了越来越多的关注。深度学习模型因其强大的表达能力, 逐渐成为分类任务中的主流方法, 目前已经有很多成熟的基于深度学习方法的人体活动识别模型, 如 DeepConvLSTM 已经可以取得较高的识别准确率。但是有监督的深度学习模型对训练数据的规模要求较高, 而在活动识别领域, 高质量标签数据集规模难以获取, 而无监督方法尚未在活动识别领域有大规模应用。近两年, 无监督表示学习方法中的对比学习在 CV 和 NLP 领域取得了巨大的成功, 如在 CV 领域, 基于对比学习搭建的预训练模型 SimCLR 在下游任务的准确率已经超过了监督模型。但是由于活动数据本身的连续性和周期性, 并不能直接将现有的对比学习框架应用于活动识别领域, 其问题主要出在正负样本对构造环节。

本文针对以上问题进行研究, 提出了基于对比学习的时间序列变化点检测模型 CL-CPD 来解决活动数据的连续性问题, 以协助负样本的选取; 提出了基于 Transformer 的活动数据增强模型来解决活动数据的周期性问题, 以协助正样本的选取, 并在一定程度上解决了活动数据样本不足的问题。最后在前两个模型的基础上, 搭建了完整的基于对比学习的活动识别预训练模型, 并通过对比实验验证了以上算法的有效性。

(1) 针对活动数据的连续性导致的负样本选取难的问题, 本文提出基于对比学习的时间序列变化点检测模型 CL-CPD。具体地, 提出一种简化的样本对构建方法, 对样本进行复用减少了计算量; 提出了跨通道的注意力机制, 用于捕捉不同传感器之间的关联; 并应用空洞卷积和多尺度卷积解决每种动作周期不同的问题。通过在开源数据集 USC-HAD 和 HASC 上进行了对比实验, 验证了所提出的算法的有效性。

(2) 针对活动数据的周期性导致的正样本选取难的问题, 本文提出基于 Transformer 的活动数据增强模型 TGAN。具体地, 利用注意力机制对序列数据建模并抽取特征, 一定程度上避免了传统 GAN 易发生模型坍塌的问题, 并提出了适用于数据增强任务的对角 mask 方法; 构建了基于 Transformer 的分类器, 并对不同结构生成模型与分类器之间的关系进行简要分析。通过在开源数据集 PAMAP2 和 RWHAR 上进行对比实验, 验证了 TGAN 在准确率不输于主流方法的前提下, 精简

了模型大小并加快了模型训练速度。

(3) 设计并搭建了完整的基于对比学习预训练模型 CL-HARP, 并融合了上述研究内容。对于预训练模型本身, 使用 t-SNE 算法对预训练结果进行降维和可视化, 展示了预训练模型学习数据分布的有效性。针对于下游的识别任务, 通过在公开数据集 PAMAP2 和 Opportunity 上进行对比实验, 验证了预训练模型在活动识别任务上的有效性。

6.2 展望

相对于目前自监督学习方法在活动识别领域的研究成果, 本文搭建的基于对比学习的活动识别预训练模型取得了一定的成果, 但是仍然有很多地方值得进一步研究:

(1) 本文提出的负样本选取方法适用于每种动作连续多次的情况, 对于持续时间过短的动作, 基于变化点检测负样本选取的方式可能会出现负样本过少的问题。之后的工作可以考虑探究具有更普适性的负样本选取方法。

(2) 本文提出的基于 TGAN 模型, 在训练数据分布过于集中地情况下, 即动作之间连续性和相似性过强的情况下, 仍可能出现模型坍塌的问题。之后的工作可以考虑使用非对称模型结构和基于重构损失的目标函数来进一步减少模型坍塌的可能。

(3) 本文搭建的预训练模型受限于不同数据集的采集方法不同, 需要针对不同的数据各自进行训练。之后的工作可以考虑探究更通用的数据预处理方式, 使得预训练模型可以应用在更广泛的数据集上, 从而取得更好的效果。

参考文献

- [1] Zheng X, Wang M, Ordieres-Meré J. Comparison of data preprocessing approaches for applying deep learning to human activity recognition in the context of industry 4.0[J]. Sensors, 2018, 18(7): 2146.
- [2] Baghezza R, Bouchard K, Bouzouane A, et al. From Offline to Real-Time Distributed Activity Recognition in Wireless Sensor Networks for Healthcare: A Review[J]. Sensors, 2021, 21(8): 2786.
- [3] Yim D, Lee W H, Kim J I, et al. Quantified activity measurement for medical use in movement disorders through IR-UWB radar sensor[J]. Sensors, 2019, 19(3): 688.
- [4] Mavropoulos T , Symeonidis S , Tsanousa A , et al. Smart integration of sensors, computer vision and knowledge representation for intelligent monitoring and verbal human-computer interaction[J]. Journal of Intelligent Information Systems, 2021, 57(2): 321-345.
- [5] Bermejo U , Almeida A , Bilbao A , et al. Embedding-based real-time change point detection with application to activity segmentation in smart home time series data[J]. Expert Systems with Applications, 2021, 185(2):115641.
- [6] Jarraya A , Bouzeghoub A , Borgi A , et al. DCR: A new distributed model for human activity recognition in smart homes[J]. Expert Systems with Application, 2020, 140(Feb.):112849.1-112849.19.
- [7] Lgf A , Sft B . Activity recognition and anomaly detection in smart homes - ScienceDirect[J]. Neurocomputing, 2020, 423: 362-372.
- [8] Civitarese G, Sztyler T, Riboni D, et al. POLARIS: Probabilistic and ontological activity recognition in smart-homes[J]. IEEE Transactions on Knowledge and Data Engineering, 2019, 33(1): 209-223.
- [9] Gm A , Be A , Rg B , et al. A machine learning approach for semi-automatic assessment of IADL dependence in older adults with wearable sensors[J]. Int. J. Medical Informatics.2022, 157: 104625.
- [10] Bi C , Huang J , Xing G , et al. SafeWatch: A Wearable Hand Motion Tracking System for Improving Driving Safety[C]. ACM Trans. Cyber Phys. Syst.2020, 4(1):

- 13:1-13:21.
- [11] Sridharan M , Bigham J , Campbell P M , et al. Inferring Micro-Activities through Wearable Sensing for ADL recognition of Home-Care Patients[J]. IEEE Journal of Biomedical and Health Informatics, 2019, PP(99):1-1.
 - [12] Nweke H F, Teh Y W, Al-Garadi M A, et al. Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges[J]. Expert Systems with Applications, 2018, 105: 233-261.
 - [13] Faridee A Z M , Khan M , Pathak N , et al. AugToAct: scaling complex human activity recognition with few labels[C] MobiQuitous: Computing, Networking and Services. 2019: 162-171.
 - [14] Jaiswal A, Babu A R, Zadeh M Z, et al. A survey on contrastive self-supervised learning[J]. Technologies, 2020, 9(1): 2.
 - [15] Tian Y, Sun C, Poole B, et al. What makes for good views for contrastive learning?[J]. Advances in Neural Information Processing Systems, 2020, 33: 6827-6839.
 - [16] Ni M, Huang H, Su L, et al. M3p: Learning universal representations via multitask multilingual multimodal pre-training[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 3977-3986.
 - [17] Devlin J , Chang M W , Lee K , et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[C]. NAACL-HLT (1) 2019: 4171-4186.
 - [18] Pfeiffer J, Rücklé A, Poth C, et al. AdapterHub: A Framework for Adapting Transformers[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. 2020: 46-54.
 - [19] He K, Zhang X, Ren S, et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1026-1034.
 - [20] Chadha G S, Islam I, Schwung A, et al. Deep Convolutional Clustering-Based Time Series Anomaly Detection[J]. Sensors, 2021, 21(16): 5488.
 - [21] Chen Z, Ma Q, Lin Z. Time-Aware Multi-Scale RNNs for Time Series Modeling[C]. IJCAI 2021: 2285-2291
 - [22] Kag A, Zhang Z, Saligrama V. Rnns incrementally evolving on an equilibrium manifold: A panacea for vanishing and exploding gradients?[C]//International

- Conference on Learning Representations. 2019.
- [23] Hochreiter S , Schmidhuber J . Long Short-Term Memory[J]. Neural Computation, 1997, 9(8):1735-1780.
- [24] Cho K, Van Merriënboer B, Gulcehre C, et al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation[C]. EMNLP 2014: 1724-1734
- [25] Mnih V, Heess N, Graves A. Recurrent models of visual attention[J]. Advances in neural information processing systems, 2014, 27.
- [26] Cross J, Huang L. Incremental Parsing with Minimal Features Using Bi-Directional LSTM[C]. ACL (2) 2016: 32–37.
- [27] Reiss A, Stricker D. Introducing a new benchmarked dataset for activity monitoring[C]//2012 16th international symposium on wearable computers. IEEE, 2012: 108-109.
- [28] Roggen D, Calatroni A, Rossi M, et al. Collecting complex activity datasets in highly rich networked sensor environments[C]//2010 Seventh international conference on networked sensing systems (INSS). IEEE, 2010: 233-240.
- [29] Ronao C A, Cho S B. Human activity recognition with smartphone sensors using deep learning neural networks[J]. Expert systems with applications, 2016, 59: 235-244.
- [30] Hughes D, Correll N. Distributed convolutional neural networks for human activity recognition in wearable robotics[M]//Distributed Autonomous Robotic Systems. Springer, Cham, 2018: 619-631.
- [31] Ordóñez F J, Roggen D. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition[J]. Sensors, 2016, 16(1): 115.
- [32] Murad A, Pyun J Y. Deep recurrent neural networks for human activity recognition[J]. Sensors, 2017, 17(11): 2556.
- [33] Gu F, Khoshelham K, Valae S, et al. Locomotion activity recognition using stacked denoising autoencoders[J]. IEEE Internet of Things Journal, 2018, 5(3): 2085-2093.
- [34] Sheng T, Huber M. Unsupervised Embedding Learning for Human Activity Recognition Using Wearable Sensor Data[C]. FLAIRS Conference 2020: 478-483
- [35] Diederik P. Kingma,Max Welling. Auto-Encoding Variational Bayes.[J]. CoRR,2013,abs/1312.6114.

- [36] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio. Generative Adversarial Networks[J]. Curran Associates, Inc, 2014 (27-2014): 2672.
- [37] He K, Fan H, Wu Y, et al. Momentum contrast for unsupervised visual representation learning[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 9729-9738.
- [38] Chen T, Kornblith S, Norouzi M, et al. A simple framework for contrastive learning of visual representations[C]//International conference on machine learning. PMLR, 2020: 1597-1607.
- [39] Grill J B, Strub F, Altché F, et al. Bootstrap your own latent-a new approach to self-supervised learning[J]. Advances in Neural Information Processing Systems, 2020, 33: 21271-21284.
- [40] Wang T, Isola P. Understanding contrastive representation learning through alignment and uniformity on the hypersphere[C]//International Conference on Machine Learning. PMLR, 2020: 9929-9939.
- [41] Lucas J, Tucker G, Grosse R, et al. Understanding Posterior Collapse in Generative Latent Variable Models[C]. DGS@ICLR 2019
- [42] Caron M, Misra I, Mairal J, et al. Unsupervised learning of visual features by contrasting cluster assignments[J]. Advances in Neural Information Processing Systems, 2020, 33: 9912-9924.
- [43] Yu F, Koltun V. Multi-Scale Context Aggregation by Dilated Convolutions[C]. ICLR (Poster) 2016
- [44] Wei Y, Xiao H, Shi H, et al. Revisiting dilated convolution: A simple approach for weakly-and semi-supervised semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7268-7277.
- [45] Li Y, Zhang X, Chen D. Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 1091-1100.
- [46] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database[C]//2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009: 248-255.

- [47] Yu Q, Lam W. Data augmentation based on adversarial autoencoder handling imbalance for learning to rank[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33(01): 411-418.
- [48] Hoelzemann A, Sorathiya N, Van Laerhoven K. Data Augmentation Strategies for Human Activity Data Using Generative Adversarial Neural Networks[C]//2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops). IEEE, 2021: 8-13.
- [49] Chang W C , Li C L , Yang Y , et al. Kernel Change-point Detection with Auxiliary Deep Generative Models[C]. ICLR(Poster) 2019
- [50] Chopra S, Hadsell R, LeCun Y. Learning a similarity metric discriminatively, with application to face verification[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). IEEE, 2005, 1: 539-546.
- [51] Sohn K. Improved deep metric learning with multi-class n-pair loss objective[J]. Advances in neural information processing systems, 2016, 29.
- [52] Gao T, Yao X, Chen D. SimCSE: Simple Contrastive Learning of Sentence Embeddings[C]. EMNLP (1) 2021: 6894-6910
- [53] Kieu T, Yang B, Guo C, et al. Outlier Detection for Time Series with Recurrent Autoencoder Ensembles[C]//IJCAI. 2019: 2725-2732.
- [54] Dai J, Bai X, Yang Z, et al. A pervasive fall detection system using mobile phones[C]//IEEE International Conference on Pervasive Computing and Communications Workshops. 2010:292-297
- [55] Gharghabi S, Yeh C C M, Ding Y, et al. Domain agnostic online semantic segmentation for multi-dimensional time series[J]. Data mining and knowledge discovery, 2019, 33(1): 96-130.
- [56] Yamada M, Kimura A, Naya F, et al. Change-point detection with feature selection in high-dimensional time-series data[C]// International Joint Conference on Artificial Intelligence. 2013: 1827-1833.
- [57] Liu S, Yamada M, Collier N, et al. Change-point detection in time-series data by relative density-ratio estimation[J]. Neural Networks, 2013, 43: 72-83.
- [58] Deldari S, Smith D V, Sadri A, et al. ESPRESSO: entropy and ShaPe aware time-Series SegmentatiOn for processing heterogeneous sensor data[J]. Proceedings of the

- ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2020, 4(3): 1-24.
- [59] Zeng M, Nguyen L T, Yu B, et al. Convolutional neural networks for human activity recognition using mobile sensors[C]//6th international conference on mobile computing, applications and services. IEEE, 2014: 197-205.
- [60] Moya Rueda F, Grzeszick R, Fink G A, et al. Convolutional neural networks for human activity recognition using body-worn sensors[C]//Informatics. Multidisciplinary Digital Publishing Institute, 2018, 5(2): 26.
- [61] Prabono A G, Yahya B N, Lee S L. Atypical sample regularizer autoencoder for cross-domain human activity recognition[J]. Information Systems Frontiers, 2021, 23(1): 71-80.
- [62] Varamin A A, Abbasnejad E, Shi Q, et al. Deep auto-set: A deep auto-encoder-set network for activity recognition using wearables[C]//Proceedings of the 15th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services. 2018: 246-253.
- [63] Zhao Y, Yang R, Chevalier G, et al. Deep residual bidir-LSTM for human activity recognition using wearable sensors[J]. Mathematical Problems in Engineering, 2018, 1-13.
- [64] Xu C, Chai D, He J, et al. InnoHAR: A deep neural network for complex human activity recognition[J]. Ieee Access, 2019, 7: 9893-9902.
- [65] Xia K, Huang J, Wang H. LSTM-CNN architecture for human activity recognition[J]. IEEE Access, 2020, 8: 56855-56866.
- [66] Wan S, Qi L, Xu X, et al. Deep learning models for real-time human activity recognition with smartphones[J]. Mobile Networks and Applications, 2020, 25(2): 743-755.
- [67] Garcia K D, de Sá C R, Poel M, et al. An ensemble of autonomous auto-encoders for human activity recognition[J]. Neurocomputing, 2021, 439: 271-280.
- [68] Thakur D, Biswas S, Ho E S L, et al. ConvAE-LSTM: Convolutional Autoencoder Long Short-Term Memory Network for Smartphone-Based Human Activity Recognition[J]. IEEE Access, 2022: 4137-4156.

致 谢

即将和自己的学生生涯告别，此刻思绪万千。相对每年如此多的本科毕业生，读研始终是少数人的选择，所以从本科毕业的那一刻起，我们便被附上了“社会人”的身份，有各种各样的压力等着我们去承担，而从研究生入学的那一刻，又重新拾起了学生的身份。学生身份总是一个避风港，它让我顺利地签一份工作，完成毕业论文的写作，也让我最后感受了作为学生可以心无旁骛学习的那份快乐，每每想到这里，心中都会有一丝窃喜。

硕士生和本科生的培养方式是有极大不同的。师者，传道受业解惑也。导师的身份没有变，而我们却需要发挥主观能动性从被动变主动。在这里我要感谢我的两位导师。他们督促我却不逼迫我，关心我又给我留有空间。就像冬夜的路灯，我和我的影子在灯下玩儿得不亦乐乎，而他们则默默守护着我，偶有雪花飘落，我抬头欣赏着美妙的夜景，仿佛已经忘记了他们的存在。希望他们注意身体，在白天好好休息，也希望他们的光可以指引更多的人走完这段夜路。

很多人的本科生活是命运的安排，而我相信硕士生活都是每个人自己的选择。依稀记得入学第一天的夜晚很安静，躺在床上还听得见虫鸣，我对自己说，别人做得到的，我也能做到。不过我是个容易懈怠的人，好在身边有无数优秀的人激励着我。他们强烈的求知欲，说到做到的执行力，坚持不懈的毅力虽让我难以望其项背，但跟在他们后面，我仿佛也走得更快更远了。努力的人值得回报，希望他们以后不要太累，也衷心祝福他们在以后的路上走得更远一些！

学习是快乐的，也是痛苦的，每个人的痛苦都各不相同。可志虽不同，道仍相合，同处于一个大环境中，我们虽不能对别人感同身受，却仍可以互相倾诉。感谢身边和不在身边的人的陪伴，我们偶尔聊上两句，便可从庞大的未知中短暂解脱，得到学习上的一些启示，又或是得到朴素的快乐。这些灵光一闪不可多得，我总是额外珍惜它们。祝我们之间的友谊长存，也希望你们在以后的路上走得不孤单。

我曾对老师说过“戛然而止才是常态”，也记得老师对我说“把握当下”。不再谈论过去的失败，也不去期盼未来的成功。二十一载书山路，上下求索不知途。告别学生身份后，还要面对人生中的未知，希望在以后的路上遇到困难时，能够想起此时此刻的自己，去面对，去尝试，去挑战，并赋予所有帮助过我的人一些人生意义。

攻读硕士学位期间研究成果

参加项目

发表文章

获得奖励

2019-2021 东北大学一等奖学金

2019 第二届辽宁省研究生数学建模竞赛二等奖

2020 华为 DIGIX 全球校园 AI 算法精英大赛 A 赛道决赛优胜奖