Learning disentangled behaviour patterns for wearable-based human activity recognition 论文复现

朱毓正

摘要

在基于可穿戴人类活动识别 (HAR) 的研究中，主要挑战之一是大类内的可变性问题，收集到的活动信号经常是由个人、环境或其他外部因素引起的噪声或偏差与具有行为特征的信号耦合而成的。在这种情况下很难学习到 HAR 任务的有效特征。为了解决这种问题，本复现论文提出了一种行为模式分离框架，可以将行为模式与无关的噪声 (个人行为风格、性别或环境噪音) 分离出来。本复现工作在论文的基础上添加 SEnet、ECAnet、CBAM block 三种不同的注意力机制，作用在不同的特征提取层之后，在相同迭代次数下能略微提升模型性能。

关键词：可穿戴传感；注意力机制；人类活动识别

1 引言

基于可穿戴设备的人类活动识别 (HAR) 是指利用穿戴放置在身体四肢、躯干、头部等的惯性单元 (IMU) 实时记录人活动时产生的加速度、速度、角度变化等信息，将获得的信号进行处理，获得对应活动的标签进行分类，它在健康监测、睡眠评估等方面都有着广泛的运用。

基于机器学习的方法利用信号的均值、方差、功率谱密度等手工特征，通过传统机器学习方法如 KNN、SVM 等来将特征映射到相应活动标签中。然而手工特征的设计往往需要反复试验，且泛化能力较弱，所需要的特征可能会因活动的不同而有所差异。利用深度学习进行特征提取能获得比手工特征更好的表现，因而在近几年的工作中被广泛的运用。

虽然深度学习可以从输入的传感器数据中获得适当的特征，但是在处理来自不同用户的多模态传感器数据时可能会面临挑战。不同用户的传感器数据包含个体差异性 (如性别、体重、习惯)，将原始传感器数据输入深度学习网络进行特征提取会同时获得这些冗余的特征表示，降低对特定活动的跨用户的识别精度。复现论文认为人行为活动的传感数据包含了特定活动的特征和个体间不同的冗余差异特征，旨在通过神经网络将这两种不同的特征分离，利用特定活动的特征来进行 HAR，提高分类的准确率。

原文工作利用了卷积神经网络 CNN 作为特征提取器，为了进一步提高模型提取特征的能力，本复现工作在原文的基础上添加了三种不同的注意力机制，以期望能提高模型提取特征的能力。

2 相关工作

本章将介绍复现论文涵盖的三个主要领域：人类活动识别、对抗学习、表征解耦学习。

2.1 人类活动识别 HAR

传统机器学习利用 K 临近算法 (KNN)，隐式马可夫模型 (HMM)，支持向量机 (SVM) 等方法来实现 HAR，这些模型的主要缺点是依赖手工特征或启发式信息。随着深度学习的发展，利用神经网

络提取原始信号的特征，能大大减少特征工程过程中的工作量。最常见的特征提取器是卷积神经网络(CNN)[1]，它通过堆叠多个卷积层来提取 HAR 的特征表示。DeepConvLSTM[2]通过为时间信息建模添加 LSTM 层来扩展 CNN。

2.2 对抗学习

Goodfellow 等人提出了对抗生成网络 (GAN)[3]，其拥有一个生成器和鉴别器，在这两者之间建立竞争。鉴别器的目标是鉴别出样本是真实数据还是由生成器生成的数据，而生成器的目标是生成足够真实的数据来欺骗鉴别器，使鉴别器无法判断该样本是真实数据还是生成数据。训练过程是固鉴别器，训练生成器，直到鉴别器无法区分数据后，再固定生成器，训练鉴别器。如此循环训练即可提升他们的性能。GAN 已经有许多应用，如对抗样本生成[4]，风格迁移[5]和自动驾驶[6]。

2.3 解耦表征学习

深度学习是数据驱动的，能自动的学习到输入数据中的协变因素 (如速度、角度，以及噪声信号中的特征)。在本复现论文中，输入数据是由噪声和特定活动特征纠缠的高维数据，噪声信号可以是个体间的差异如性别、体重、习惯等因素。在这里，噪声定义为“基于任务的不期望的变化因素”。如在活动分类中，性别可能是噪声因素，但在条件信号样本生成任务中 (生成具有性别条件的信号样本)，性别将会是关键因素。因此探索解耦这些不同因素对下游任务实现性能提升至关重要。

此外，解耦这些因素也有利于深度学习的可解释性研究。随着深度学习的发展，计算机视觉的最新工作开始利用 GAN 从图像或视频中学习可解释的表示。Liu 等人[7]引入了一个统一的特征解纠缠框架，以学习跨不同域的域不变特征。Hu 等人[8]提出了一种分离框架，可以将步态识别与相机视图分离，以实现视图不变的步态识别。DEAN[9]从语音信号中分离出与说话者相关的特征，以实现鲁棒的说话者自适应识别。

3 本文方法

3.1 本文方法概述

图 1为复现论文的方法示意图。复现原文中的方法由两部分构成：(1) 信号和冗余特征解耦网络，它将输入特征解耦为行为特征和冗余特征 (噪声)，在这里行为特征是指与人行为活动相关的特征，冗余特征是指如性别、身高等与人行为活动不相关的噪声。(2) 依赖减少网络，旨在减少活动信号和冗余特征之间的相关性。这两个模块与特征重构模块一起最大化了特征解耦的效果，保证最小的信息损失。整个网络以端到端的方式训练。

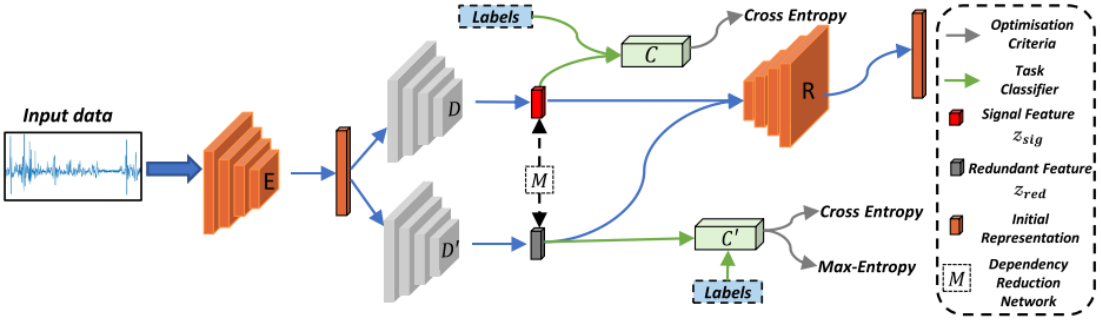

图 1: 方法示意图[10]，其中 (E),(D,D’),(C,C’),(R),(M) 分别表示编码器、解缠器，分类器，重建器和依赖减少网络

3.2 信号和冗余特征解耦网络

在这部分中，输入特征经过一个特征提取器 E 后，送入两个解耦器 D,D' 中。借助生成对抗网络的思想，D' 将生成冗余信号，与分类 C' 进行对抗。与此同时，D 将生成行为特征，并用监督学习的方式训练分类器 C。最后为了保证 D 与 D' 生成的特征完整性，生成的行为特征和冗余信号特征输入重建 R 进行重建，并与原始提取的特征信号做对比。

3.3 依赖减少网络

在这一步中，利用 D 和 D' 生成的行为特征和冗余特征的互信息来减少他们之间的相似度，进一步分离行为特征和冗余特征。

行为特征 z_{sig} 和冗余特征 z_{red} 之间的互相关可以表示为：

$$I(z_{sig}, z_{red}) = \int_{z_{sig}} \int_{z_{red}} log \frac{P(z_{sig}, z_{red})}{P(z_{sig}), P(z_{red})} \mathrm{d}z_{sig} \mathrm{d}z_{red} \tag{1}$$

其中，$P(z_{sig}, z_{red})$ 表示联合概率密度分布，$P(z_{sig}), P(z_{red})$ 是边缘概率密度分布。$I(z_{sig}, z_{red})$ 表示两种特征间的依赖性，最小化 $I(z_{sig}, z_{red})$ 便可使得两个特征分离更远。

3.4 算法与实现

复现原文模型中，使用 CNN 或 DeepConvLSTM 作为编码器 E；解耦器 D,D' 使用具有批量标准化层的单个全连接层；重建器 R 使用单个全连接层；依赖减少网络 M 使用两个全连接层；分类器 C,C' 使用两个全连接层。

4 复现细节

4.1 与已有开源代码对比

复现原文代码已开源 http://github.com/Jie-su/BPD。本次复现工作针对原文使用的两个数据集进行重现，同时为尝试提高模型提取特征的性能，在编码器 E 或解缠器 D,D' 后分别添加了三种不同的注意力机制 SEnet[11]、ECAnet[12]、CBAM[13]。

4.2 注意力机制模块

在复现工作中，为尝试提高模型提取特征的性能，如图 2所示, 在编码器 E 或解缠器 D,D' 后添加注意力机制。原文使用 CNN 和 DeepConvLSTM 作为编码器，输入数据的通道和空间维数多，可以使用通道注意力机制和空间注意力机制。通道注意力机制：通过神经网络的学习，给不同的通道赋不同的权重，强化重要特征。空间注意力机制：空间注意力机制旨在提升关键区域的特征表达，增强感兴趣的特定目标区域同时弱化不相关的背景区域。

4.2.1 Squeeze-and-Excitation Networks

如图 3所示为 SEnet 的示意图。SEnet[11]是通道注意力机制的一种，输入数据 X 经特征提取器提取出特征 U 后，使用全局池化顺着空间维度进行特征压缩，得到 $1 * 1 * C$ 的特征，再经过全连接层为每个通道生成权重，最后将权重与原始提取特征 U 结合，即得到带有通道注意力机制的输出。

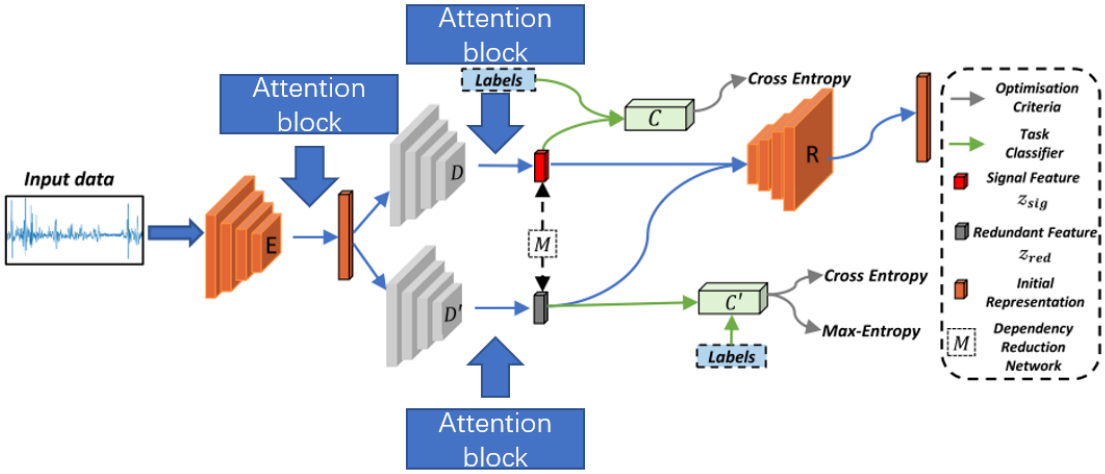

图 2: 添加注意力机制

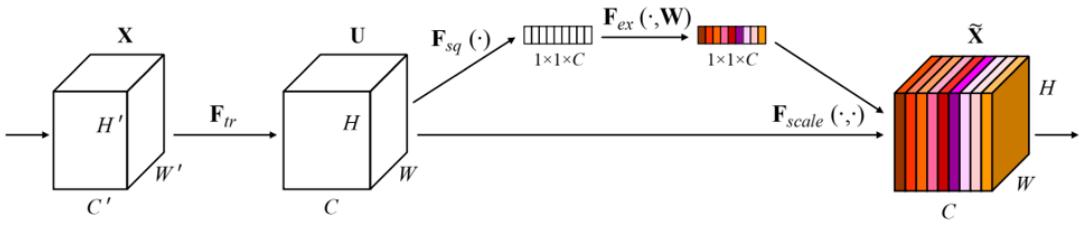

图 3: SEnet 示意图[11]

4.2.2 Efficient Channel Attention

如图 4所示为 ECAnet 的示意图。ECAnet[12]也是通道注意力机制的一种，它指出 SEnet 中对所有通道进行降维的操作效果并不好，同时也带来更多的参数，获得所有通道之间的依存关系效率不高也没有必要。于是该方法使用一维滑动卷积代替 SEnet 中的全连接层，减少参数量的同时获得局部的跨通道信息。

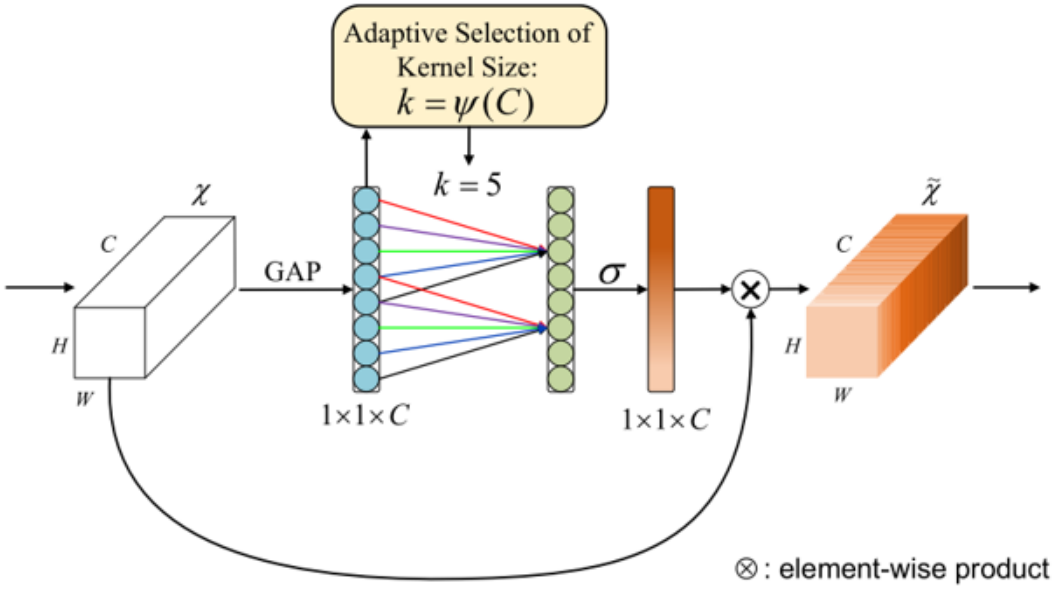

图 4: ECAnet[12]

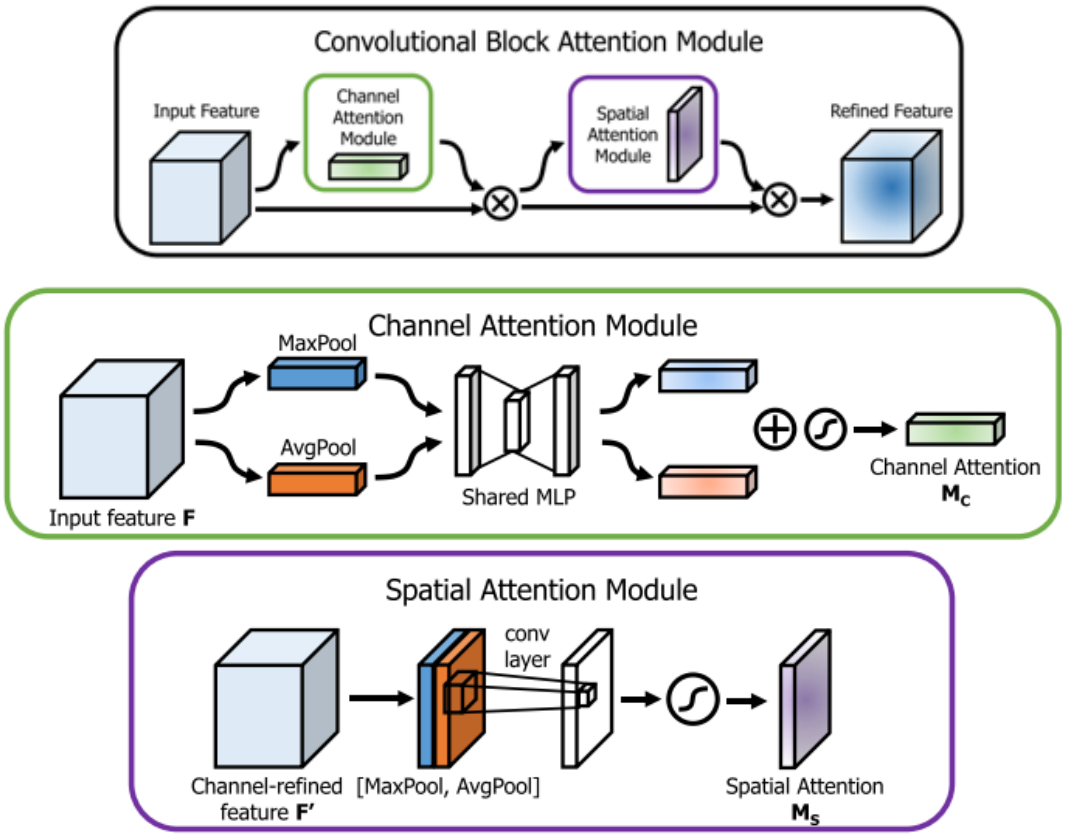

图 5: CBAM[13]

4.2.3 Convolutional Block Attention Module

如图 5所示为 CBAM 的示意图。CBAM[13]结合了通道注意力机制和空间注意力机制。输入特征首先经过通道注意力模块，获得通道权重后，再经过空间注意力模块，以获得空间权重。在通道注意力模块中，将空间特征分别按照最大池化和平均池化进行压缩，再通过同于一个多层感知机后得到各自权重，最后相加，即得到通道注意力权重。在空间注意力模块中，则将通道特征分别按照最大池化和平均池化进行压缩，再进行拼接，得到具有 2 通道的空间特征，利用卷积获得其空间注意力权重。

5 实验

5.1 数据集

本次复现实验使用 MHEALTH[14]、PAMAP2[15]数据集。

Mobile Health(MHEALTH) 数据集包含 10 名不同特征的受试者的身体运动和生命体征数据，包含 12 项非实验室环境下的活动。输入数据总维度为 23，惯性测量单元 (IMU) 分别部署在受试者胸部，右手腕和左脚腕，收集 3 轴加速度、3 轴陀螺仪、3 轴磁场的数据。此外胸部的 IMU 单元提供 2 个 ECG 心电测量，用于基本的心脏检测。

Physical Activity Monitoring(PAMAP2) 数据集包含 9 名受试者的 18 项不同活动的记录数据。IMU 单元部署再受试者胸部、手腕、脚踝。在复现原文中选择了其中的 12 项活动，使用 8 名受试者的所有 IMU 数据。值得注意的是，原文中使用了该数据集的所有 52 维通道，但经过检查发现，该数据集中有一维通道是无效的，需要经过数据清洗之后才能使用。因此复现工作使用了 8 名受试者的 12 项活动，通道数 51 维。

表 1: dataset

Dataset	Subject	Activity	Frequency	Sample	Dim	Wearing Position
MHEALTH	10	12	50Hz	0.34M	23	Chest,Ankle,Arm
PAMAP2	8	12	100Hz	2.84M	51	Wrist,Chest,Ankle

表 2: MHEALTH+CNN

Dataset BackBone	Subject	ECAnet	SEnet	CBAM	D-ECAnet	D-SEnet	原文复现	原文
MHEALTH CNN	1	0.9235	0.8900	0.8967	**0.9909**	0.8713	0.8748	0.9575
	2	0.8983	**0.9291**	0.9089	0.9263	0.9260	0.8936	0.9348
	3	0.8322	0.8490	0.8127	**0.9177**	0.8407	0.8692	0.8659
	4	0.9163	0.8972	0.8804	0.8524	0.9512	**0.9526**	0.9510
	5	0.9849	0.9842	**0.9885**	0.9539	0.9842	0.9815	0.9904
	6	**0.9913**	0.9809	0.9748	0.9868	0.9677	0.9841	0.9934
	7	0.9841	0.9876	0.9790	0.9850	0.9894	**0.9902**	0.9965
	8	0.9696	0.9758	0.9632	**0.9886**	0.9731	0.9615	0.9848
	9	0.9851	0.9878	0.9818	0.9782	**0.9887**	0.9878	0.9913
	10	0.9816	0.9874	0.9638	**0.9895**	0.9759	0.9891	0.9943
	Avg.	0.9467	0.9469	0.9350	**0.9569**	0.9468	0.9484	0.9660

表 3: MHEALTH+DeepConvLSTM

Dataset BackBone	Subject	ECAnet	SEnet	CBAM	D-ECAnet	D-SEnet	原文复现	原文
MHEALTH DeepConvLSTM	1	0.9225	**0.9829**	0.9223	0.9327	0.9036	0.9509	0.9554
	2	0.8537	**0.9577**	0.9291	0.8803	0.9382	0.8455	0.9085
	3	**0.8724**	0.8048	0.8522	0.8643	0.8538	0.8671	0.8711
	4	0.9350	0.8549	0.9027	0.9145	**0.9364**	0.8957	0.9573
	5	0.8751	0.7969	0.8624	0.9609	0.8706	**0.9663**	0.9804
	6	0.9686	**0.9795**	0.9037	0.9762	0.9757	0.9743	0.9766
	7	0.9687	0.9671	0.9653	**0.9770**	0.9650	0.9608	0.9760
	8	0.9729	**0.9751**	0.9659	0.9736	0.9391	0.9746	0.9775
	9	0.9798	0.9002	0.9780	0.9746	**0.9859**	0.9809	0.9869
	10	0.9855	0.9825	0.9829	0.9809	**0.9856**	0.9838	0.9865
	Avg.	0.9334	0.9202	0.9265	**0.9435**	0.9354	0.9400	0.9576

5.2 实验设置

数据预处理。将输入数据以 50% 的重叠率划分成 168 个固定大小的滑动窗口数据。由于数据集采样率各有不同，因此 MHEALTH、PAMAP2 的窗口长度分别是 3.36、1.68 秒，直接输入模型，没有额外的数据处理。

基线模型。复现工作的基线模型为复现原文的 BPD 模型，该模型在原文中有两个变体，分别以 CNN、DeepConvLSTM 作为特征提取器。复现工作在这两个变体模型的编码器 E，解缠器 D,D’ 前分别加入不同注意力机制，尝试提高模型提取特征的性能。

训练设置。使用 Xavier Normal 初始化网络参数，使用 Adam Optimier 优化，学习率设置 0.0001，batch size 为 64。由于资源有限，迭代次数为 50(原文 300)。所有算法由 PyTorch 实现，并且在 GeForce RTX 2080 Ti GPU 上运行。

评估方法。对每个数据集使用用户留一交叉验证来评估模型的性能，对每个数据集，评估总体表现和单一用户表现。使用平均 F1 分数作为评估指标，这一指标在 HAR 文献中广泛运用。

5.3 实验结果

如表 2、表 3所示，为 MHEALTH 数据集下的表现。其中，前三列数据 ECAnet、SEnet、CBAM 分别表示在图 2中特征提取器 E 后添加注意力模块，D-ECAnet、D-SEnet 则表示在解缠器 D 后添加注意力模块。在 50 迭代次数下，原文复现的结果并没有原文 (迭代次数 300 次) 给出的数据好，但是添加在解耦器后的注意力层发挥了作用，与相比原文复现的结果分别提高了 0.85% 和 0.35%。表 4、表

5为 PAMAP2 数据集下的表现。在 PAMAP2 数据集中，只有以 CNN 为 backbone 的 SEnet 表现超过了原文复现 1.88%。

本次复现的实验结果并没有原文中的数据结果好，最主要的原因是训练的 epoch 数目不同。原文指出其训练的最大 epoch 为 300，但由于在复现过程中消耗的资源太多，本次复现最大 epoch 设为 50。而在相同的训练条件下，注意力机制发挥了作用。

CBAM 在所有模型的表现都没有很突出，主要原因是输入数据是 IMU 数据，数据维度 $1 * t * c$，其中 t 指的是数据长度，c 指的是通道数。每个 IMU 数据仅是一维数据，而 CBAM 在处理这些数据时，空间注意力机制失去了作用，导致效果不佳。

表 4: PAMAP2+CNN

Dataset BackBone	Subject	ECAnet	SEnet	CBAM	D-ECAnet	D-SEnet	原文复现	原文
PAMAP2 CNN	1	**0.6743**	0.5913	0.6238	0.6405	0.6564	0.6480	0.6826
	2	0.7534	**0.8035**	0.6984	0.7297	0.7412	0.7395	0.8719
	3	0.8066	**0.8185**	0.7149	0.6802	0.7281	0.7927	0.8262
	4	0.7888	**0.8693**	0.8649	0.8298	0.8259	0.8515	0.8307
	5	0.8185	0.8475	**0.8589**	0.8541	0.8315	0.8222	0.8900
	6	**0.8769**	0.8625	0.7931	0.8486	0.8070	0.8287	0.8827
	7	**0.9375**	0.9188	0.9365	0.9317	0.9159	0.9301	0.9311
	8	0.3067	0.3407	0.3376	0.3847	**0.4213**	0.3705	0.4011
	Avg.	0.7453	**0.7565**	0.7285	0.7374	0.7409	0.7479	0.7895

表 5: PAMAP2+DeepConvLSTM

Dataset BackBone	Subject	ECAnet	SEnet	CBAM	D-ECAnet	D-SEnet	原文复现	原文
PAMAP2 DeepConvLSTM	1	0.6303	**0.6848**	0.6549	0.6471	0.6820	0.6840	0.6915
	2	0.6471	0.6510	0.7227	0.5555	0.5803	**0.7446**	0.8381
	3	0.6583	0.6419	0.6588	0.6558	**0.6956**	0.6402	0.8117
	4	0.8543	0.8278	0.8629	0.8367	**0.9077**	0.8455	0.8224
	5	**0.8802**	0.8309	0.8602	0.8418	0.8588	0.8304	0.8675
	6	0.7833	0.8128	0.7876	**0.8134**	0.8110	0.7966	0.8281
	7	0.9180	0.8364	0.9065	0.9155	**0.9184**	0.9112	0.9101
	8	0.3667	0.3894	0.3643	0.3704	0.3723	**0.4178**	0.4921
	Avg.	0.7173	0.7094	0.7272	0.7045	0.7283	**0.7338**	0.7827

6 总结与展望

本次复现的工作是解耦表征学习在 HAR 领域的运用，该论文对人活动的 IMU 数据进行解耦，分离出具有活动信号的特征和与任务无关的冗余特征，提高了 HAR 任务的性能。本次复现过程中出现了一些问题，在数据集处理方面，一开始没有仔细检查数据集的数据，直接按照原文给出的 52 维进行运算，总是报错，在重新进行数据清洗，去除其中一维之后，能成功运行。

在学习过注意力机制后，尝试将注意力机制添加到模型当中，但是添加的过程缺乏较为科学的指导，算是一种“尝试性”的工作，试试看哪一种方法添加进去好用。未来应更深入的学习这些方法背后的逻辑原理，以更有科学指导的方式结合到工作中。

本次复现效果提升不明显，主要的原因有：原文迭代次数 300 次，而本次复现由于资源有限最高只能迭代 50 次。其次，原文的模型已经相对比较复杂了，模型中的参数已经能够学习到非常多的信息了，再添加注意力机制效果提升就不明显。

参考文献

[1] YANG J, NGUYEN M N, SAN P P, et al. Deep convolutional neural networks on multichannel time series for human activity recognition[C]//Twenty-fourth international joint conference on artificial intelligence. 2015.

[2] ORDÓÑEZ F J, ROGGEN D. Deep convolutional and lstm recurrent neural networks for multimodal

wearable activity recognition[J]. Sensors, 2016, 16(1): 115.

[3] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.

[4] XIAO C, LI B, ZHU J Y, et al. Generating adversarial examples with adversarial networks[J]. arXiv preprint arXiv:1801.02610, 2018.

[5] JING Y, YANG Y, FENG Z, et al. Neural style transfer: A review[J]. IEEE transactions on visualization and computer graphics, 2019, 26(11): 3365-3385.

[6] ZHANG M, ZHANG Y, ZHANG L, et al. DeepRoad: GAN-based metamorphic testing and input validation framework for autonomous driving systems[C]//Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering. 2018: 132-142.

[7] LIU A H, LIU Y C, YEH Y Y, et al. A unified feature disentangler for multi-domain image translation and manipulation[J]. Advances in neural information processing systems, 2018, 31.

[8] HU B, GUAN Y, GAO Y, et al. Robust cross-view gait recognition with evidence: A discriminant gait GAN (DiGGAN) approach[J]. arXiv preprint arXiv:1811.10493, 2018.

[9] SANG M, XIA W, HANSEN J H. Deaan: Disentangled embedding and adversarial adaptation network for robust speaker representation learning[C]//ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2021: 6169-6173.

[10] SU J, WEN Z, LIN T, et al. Learning disentangled behaviour patterns for wearable-based human activity recognition[J]. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2022, 6(1): 1-19.

[11] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

[12] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 11534-11542.

[13] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.

[14] BANOS O, GARCIA R, HOLGADO-TERRIZA J A, et al. mHealthDroid: a novel framework for agile development of mobile health applications[C]//Ambient Assisted Living and Daily Activities: 6th International Work-Conference, IWAAL 2014, Belfast, UK, December 2-5, 2014. Proceedings 6. 2014: 91-98.

[15] REISS A, STRICKER D. Introducing a new benchmarked dataset for activity monitoring[C]//2012 16th international symposium on wearable computers. 2012: 108-109.