

DanHAR: Dual Attention Network For Multimodal Human Activity Recognition Using Wearable Sensors

作者: Wenbin Gao, Lei Zhang, Qi Teng, Jun He, Hao Wu
发表刊物/会议: Applied Soft Computing
时间: 2021

符合实验室研究方向: 开放环境中的自监督活动识别

论文选择:

1. 出版物: CCF推荐 A/B 类 (A为主, 少量B), IEEE/ACM Transactions, 顶级大学最新arXiv论文, 朱/胡整理的论文列表
2. 时间: 近两年优先
3. 主题: 开放环境 (新类发现、迁移学习、连续/终身学习等), 自监督技术 (对比学习, 生成方法), 活动识别 (传感器/设备/系统, 活动识别应用, 时序分析技术, 知识数据混合的活动识别), 论文周一前发在群里, 供其他人提前了解

研究概述（每个听众都要掌握）

- 简述要解决的问题/动机

- 1) 近年来，在深度神经网络中，提出了两种结合门控循环单元(GRU)和长短期记忆(LSTM)网络的注意方法，可以同时捕获感知信号在空间和时域的依赖关系，然而与卷积神经网络(cnn)相比，递归网络通常具有较弱的表征能力。
- 2) 通过与CNN的结合，将硬注意力和软注意力两种注意力应用于时域，更多地关注长序列的目标活动。然而，只能分辨出关注的地方，而忽略了频道信息，这在决定关注什么方面起着重要作用。与门控循环单元(GRU)和长短期记忆(LSTM)网络相比，它们无法解决多模态传感信号的时空依赖性。

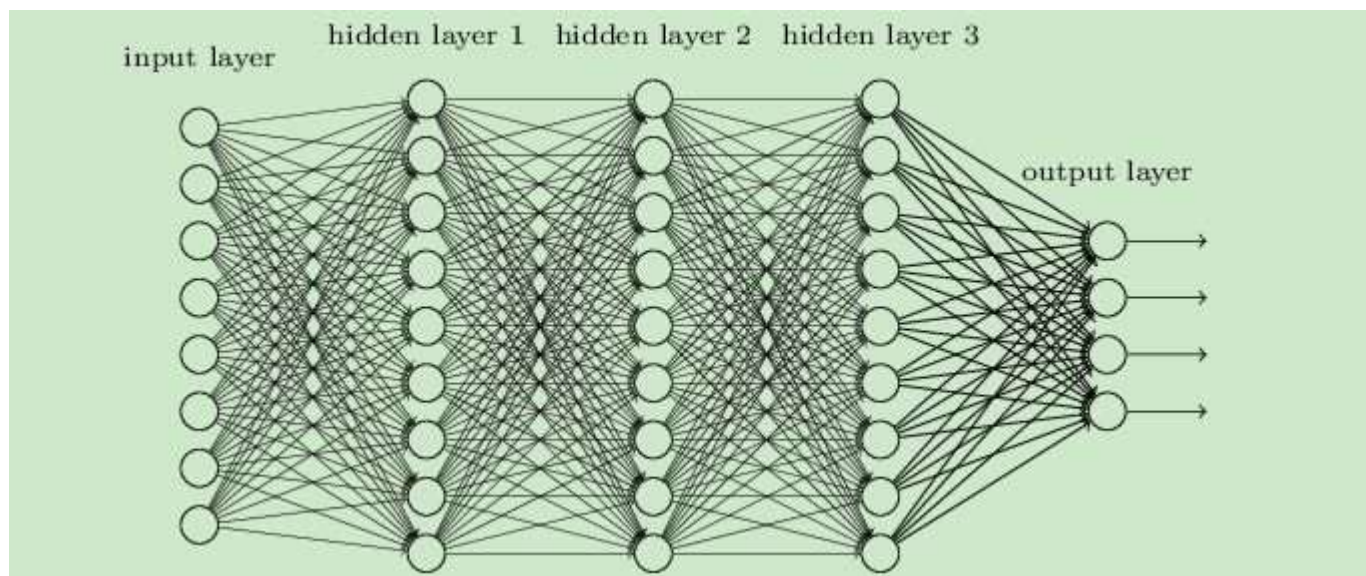
- 简述如何解决该问题

- 1) 提出了一种新的双注意方法DanHAR，在CNN上引入了信道注意和时间注意混合的框架，提高多模态HAR的可理解性方面具有优势。

背景

- **DNN (Deep Neural Networks 深度神经网络)**

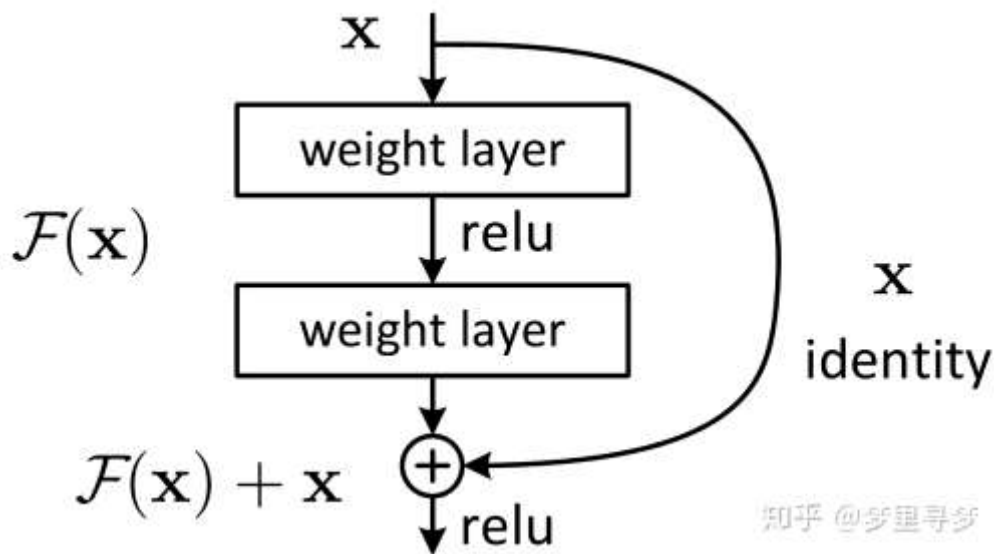
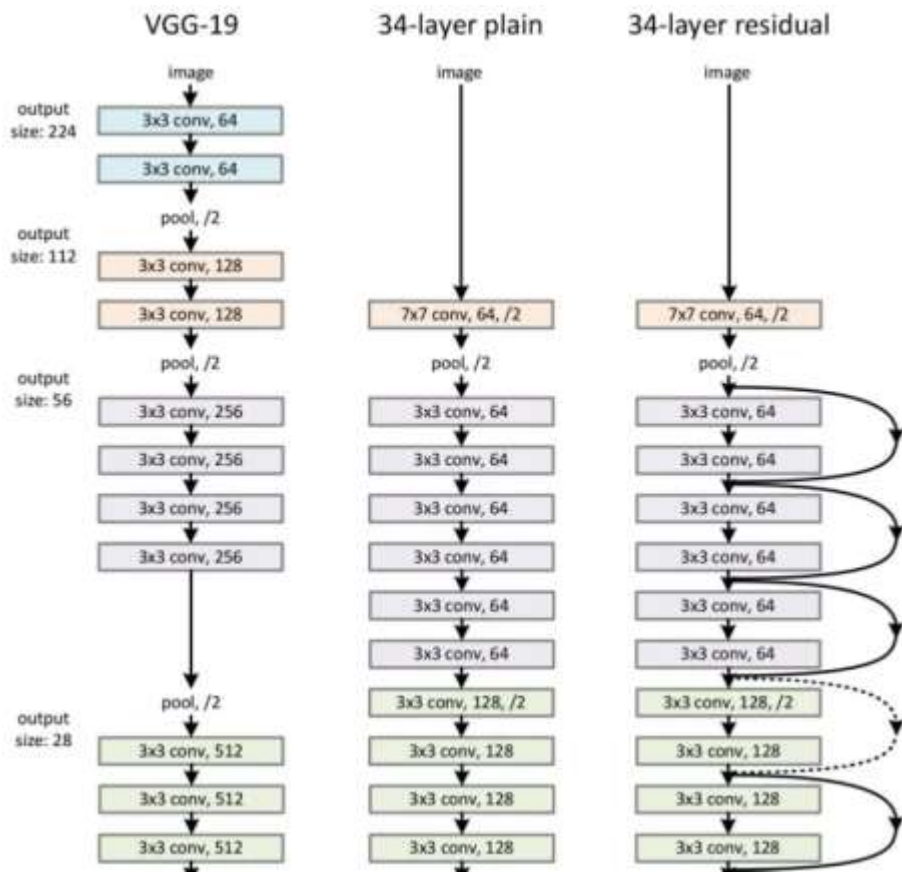
用是一种机器学习模型，其灵感来源于人工神经网络的结构和功能。它由多个神经网络层（也称为隐藏层）组成，每一层包含多个神经元（也称为节点或单元），这些层之间通过权重连接进行信息传递和变换。从而表达能力强、适用于多模态数据、表达能力强。然后却有计算复杂度高、数据需求大、过拟合风险高等缺点。



背景

● ResNet (Residual Network 残差网络)

是深度学习中的一个重要卷积神经网络架构，它在解决深层神经网络训练过程中出现的梯度消失和梯度爆炸问题方面具有创新性的思想。ResNet的核心思想是引入了残差块（residual block），允许网络在学习过程中学习恒等映射（identity mapping）的残差，从而使网络更易于训练和优化。



背景

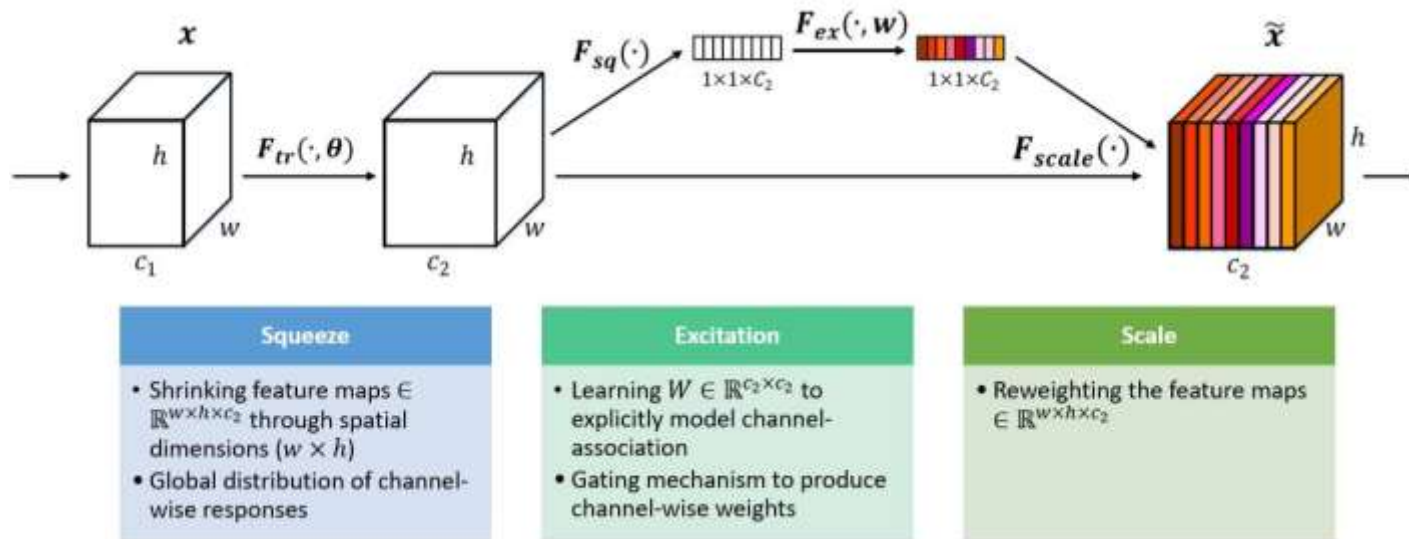
● SENet (Squeeze-and-Excitation Networks)

是一种用于卷积神经网络（CNN）的注意力机制，旨在增强网络对重要特征的感知能力。SENet通过学习通道之间的关系，动态地调整通道特征的权重，从而提高了网络的表示能力和性能。

SENet的核心思想是在每个卷积层之后引入一个Squeeze-and-Excitation块，包括两个步骤：Squeeze和Excitation。

1.Squeeze: 在Squeeze阶段，SENet通过全局平均池化操作对每个通道的特征进行压缩，将通道特征转换为一个小的全局描述符（向量）。这有助于捕获每个通道中的全局信息。

2.Excitation: 在Excitation阶段，SENet通过引入两个全连接层，对Squeeze阶段得到的全局描述符进行学习。这两个全连接层分别负责学习通道间的权重关系，以及将学习到的权重应用于每个通道特征。这样，每个通道的特征都可以得到自适应的调整，使得网络更加关注重要的特征通道



背景

- SENet (Squeeze-and-Excitation Networks)

1.Squeeze: 将 $H*W*C$ 的输入转换成 $1*1*C$ 的输出, 相当于表明该层 C 个feature map的数值分布情况, 或者叫全局信息

$$z_c = \mathbf{F}_{sq}(\mathbf{u}_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j). \quad (2)$$

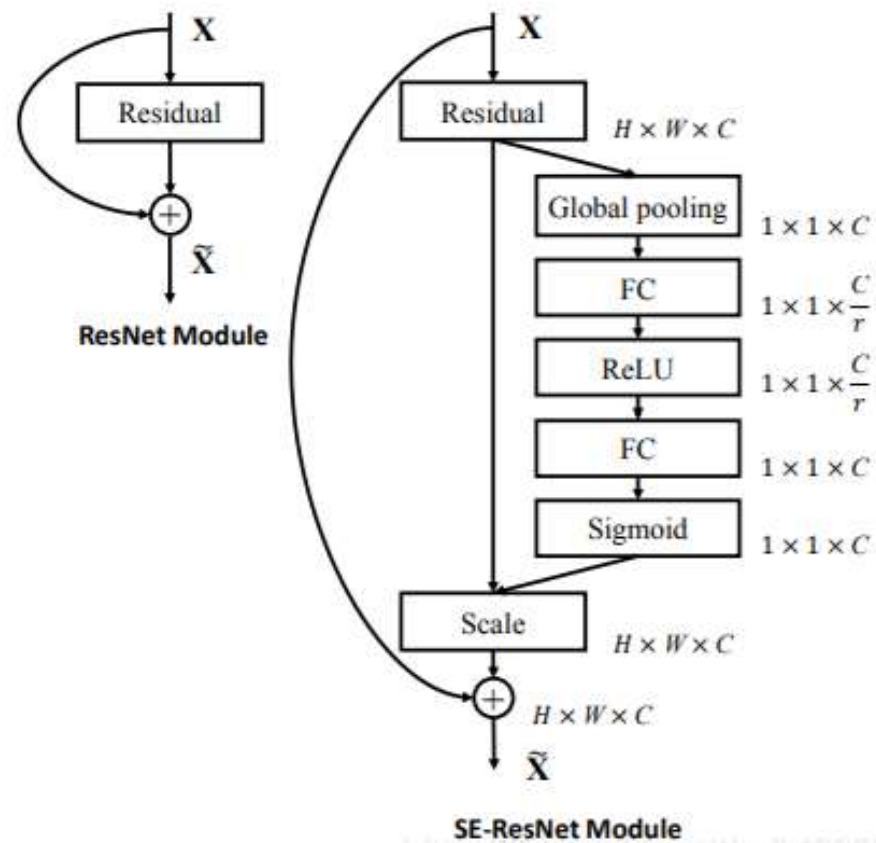
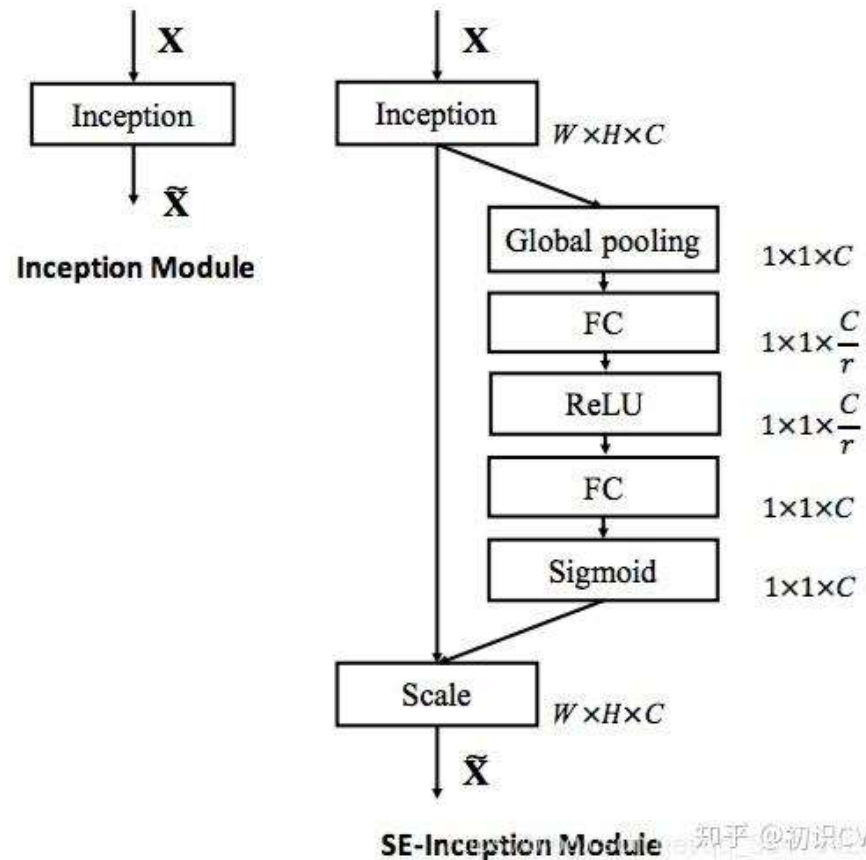
2.Excitation: 前面squeeze得到的结果是 z , 先用 W_1 乘以 z , 就是一个全连接层操作, W_1 的维度是 $C/r * C$, 这个 r 是一个缩放参数, 这个参数的目的是为了减少channel个数从而降低计算量。又因为 z 的维度是 $1*1*C$, 所以 W_1z 的结果就是 $1*1*C/r$; 然后再经过一个ReLU层, 输出的维度不变; 然后再和 W_2 相乘, 和 W_2 相乘也是一个全连接层的过程, W_2 的维度是 $C*C/r$, 因此输出的维度就是 $1*1*C$; 最后再经过sigmoid函数, 得到 s 。

$$\mathbf{s} = \mathbf{F}_{ex}(\mathbf{z}, \mathbf{W}) = \sigma(g(\mathbf{z}, \mathbf{W})) = \sigma(\mathbf{W}_2 \delta(\mathbf{W}_1 \mathbf{z})), \quad (3)$$

背景

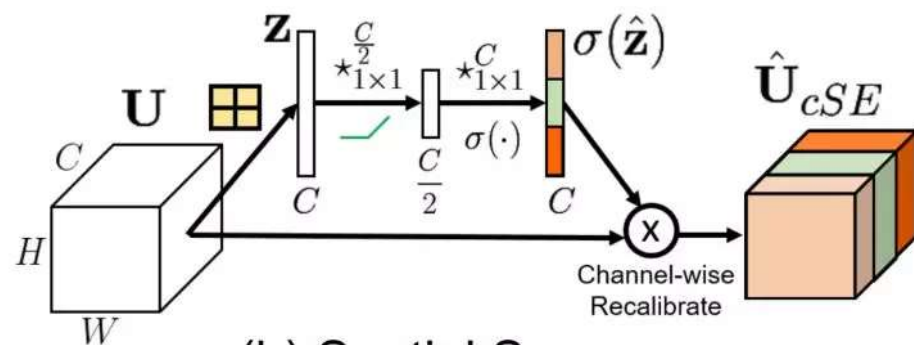
- SENet (Squeeze-and-Excitation Networks)

SENet在具体网络中的应用:



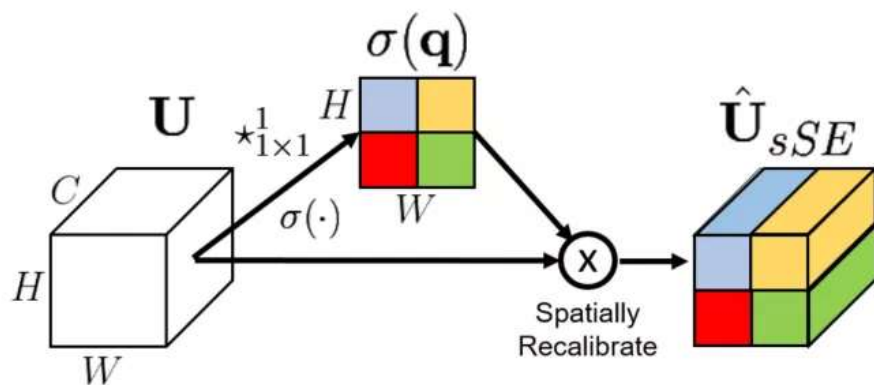
背景

- cSE :



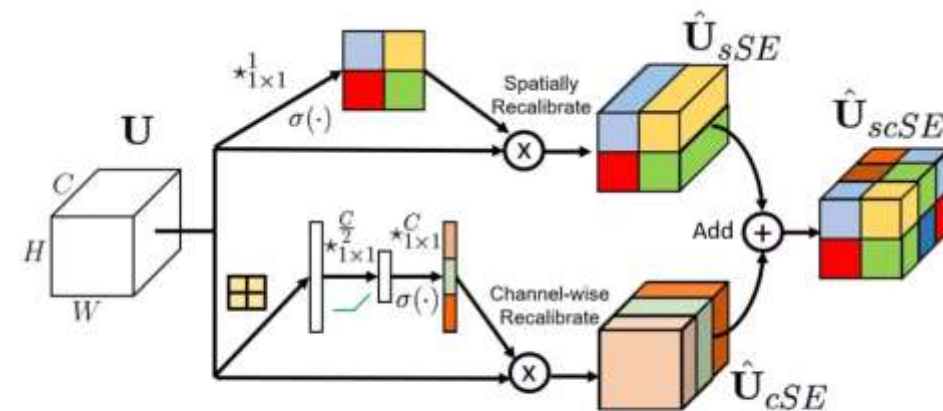
(b) Spatial Squeeze and Channel Excitation (cSE)

- sSE :



(c) Channel Squeeze and Spatial Excitation (sSE)

- scSE :



(d) Concurrent Spatial and Channel Squeeze and Excitation (scSE)

- **CBAM (Convolutional Block Attention Module 卷积注意力模块)**

是一种用于前馈卷积神经网络的简单而有效的注意力模块。给定一个中间特征图，CBAM模块会沿着两个独立的维度（通道和空间）依次推断注意力图，然后将注意力图与输入特征图相乘以进行自适应特征优化。由于CBAM是轻量级的通用模块，因此可以忽略的该模块的开销而将其无缝集成到任何CNN架构中，并且可以与基础CNN一起进行端到端训练。

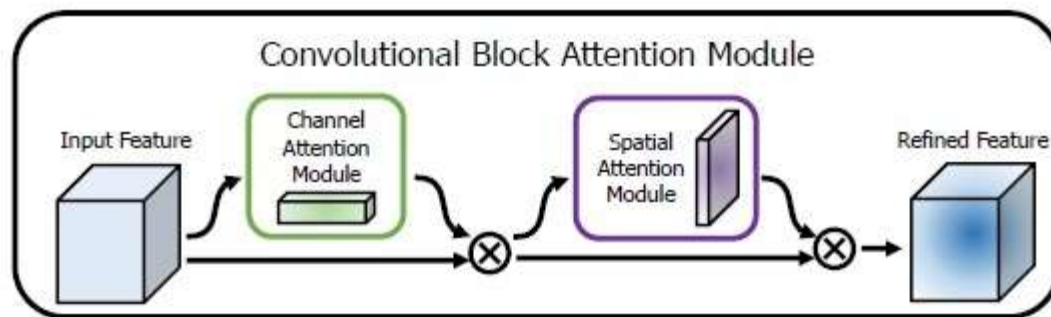


Fig. 1: The overview of CBAM. The module has two sequential sub-modules: *channel* and *spatial*. The intermediate feature map is adaptively refined through our module (CBAM) at every convolutional block of deep networks.

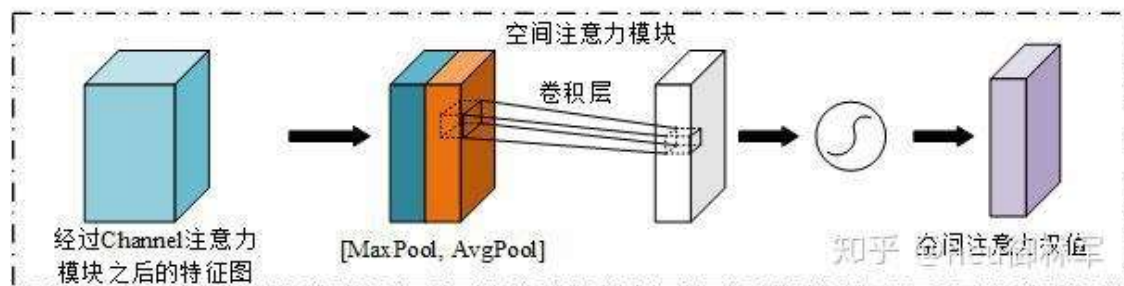
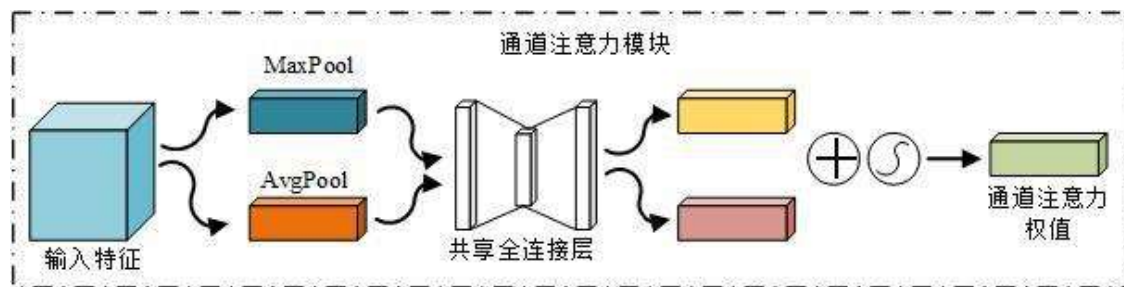
背景

● 通道注意力模块

别经过基于width和height的global max pooling 和global average pooling，然后分别经过MLP。将MLP输出的特征进行基于element-wise的加和操作，再经过sigmoid激活操作，生成最终的channel attention featuremap。将该channel attention featuremap和input featuremap做elementwise乘法操作，生成Spatial attention模块需要的输入特征。

● 空间注意力模块

将Channel attention模块输出的特征图作为本模块的输入特征图。首先做一个基于channel的global max pooling 和global average pooling，然后将这2个结果基于channel 做concat操作。然后经过一个卷积操作，降维为1个channel。再经过sigmoid生成spatial attention feature。最后将该feature和该模块的输入feature做乘法，得到最终生成的特征。



- BCBAM与ResNet网络结构组合

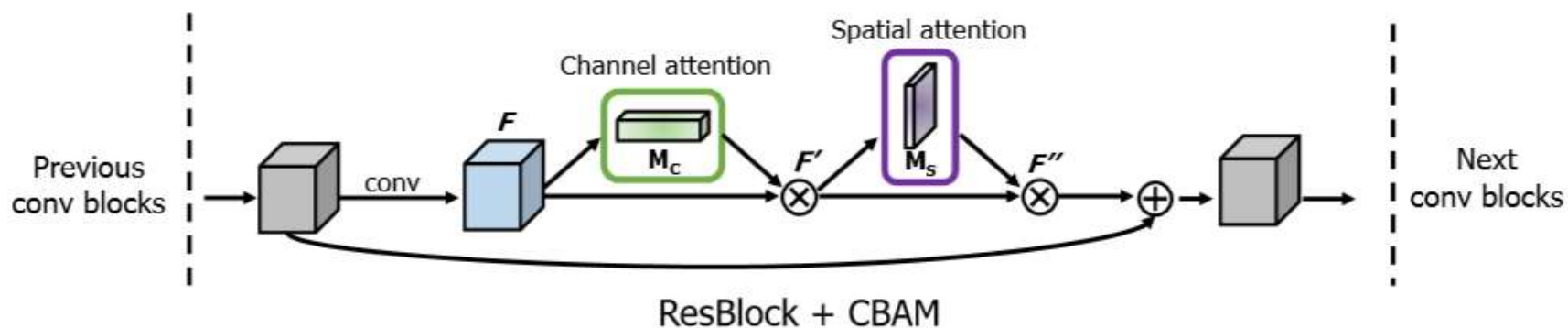


Fig. 3: **CBAM integrated with a ResBlock in ResNet[5]**. This figure shows the exact position of our module when integrated within a ResBlock. We apply CBAM on the convolution outputs in each block.

背景

- **BN (Batch Normalization 批归一化)**

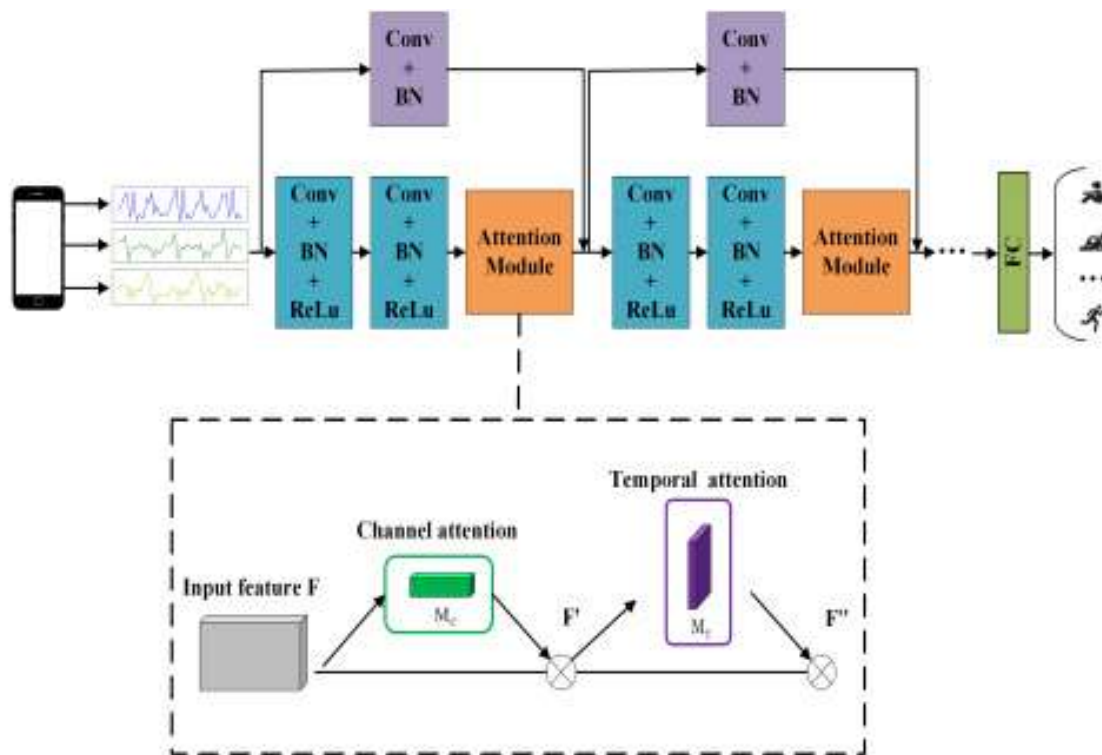
用于加速神经网络的训练并提高其稳定性和收敛速度。它的思想是对每个神经网络层的输入进行归一化处理，使得输入的均值接近0，标准差接近1。这可以帮助网络在训练过程中更快地收敛，并减少训练过程中出现的梯度消失和梯度爆炸等问题

给定两个概率分布 X 和 Y ,

研究方法

- 方法的整体结构

设计了包含卷积、通道注意力模块和时间注意力模块的DanHAR网络。首先，沿时间轴移动一个固定长度的滑动窗口，将时间序列数据分割成一组短信号，整个网络将每个时间窗口作为输入，然后将其送入卷积层以提取特征表示。其次，网络通过最大池化和平均池化两种方法沿时间轴聚合特征，生成信道注意图，然后通过Sigmoid函数生成概率。第三，利用最大池化和平均池化技术对特征图的通道信息进行跨通道轴聚合，生成后续的时间注意力图。



研究方法

- 通道注意力模块 (Channel Attention Submodule)

参照SENet，并在其基础上同时使用平均池化和最大池化特征，使更有效地进行特征聚合。

CNN相应的特征图 $A \in R^{C \times H \times W}$ ，其中C、W、H分别是通道（即滤波器的数量）、宽度（时间轴）和高度（传感器轴）维度。

$$W_C = \sigma \left(f_{\{w_1, w_2\}} (g_1(A)) + f_{\{w_1, w_2\}} (g_2(A)) \right)$$

由于两个FC层的设计目的是捕捉非线性跨通道相互作用，因此Eq.(1)可以进一步表示为：

$$W_C = \sigma \left(w_2 ReLU(w_1 g_1(A)) + w_2 ReLU(w_1 g_2(A)) \right)$$

$$g_1 = \frac{1}{WH} \sum_{i=1, j=1}^{W, H} A_{ij} \quad g_2 = \max_{i=1, j=1}^{W, H} A_{ij}$$

为了减少参数: w_1 和 w_2 分别设置为 $C \times (C/r)$ 和 $(C/r) \times C$ ，其中 r 为简化比。

研究方法

- 时间注意力模块 (Temporal Attention Submodule)

使用加法策略来推断频道注意力，这对两个汇集的特征具有同等的重要性。参考scSE和CBAM模块，同样的通过应用平均池化和最大池化操作来聚合特征映射的通道信息来计算时间注意力。将两个池化特征连接起来，沿着通道索引聚合信息。然后通过标准卷积层对连接的池化特征进行卷积，

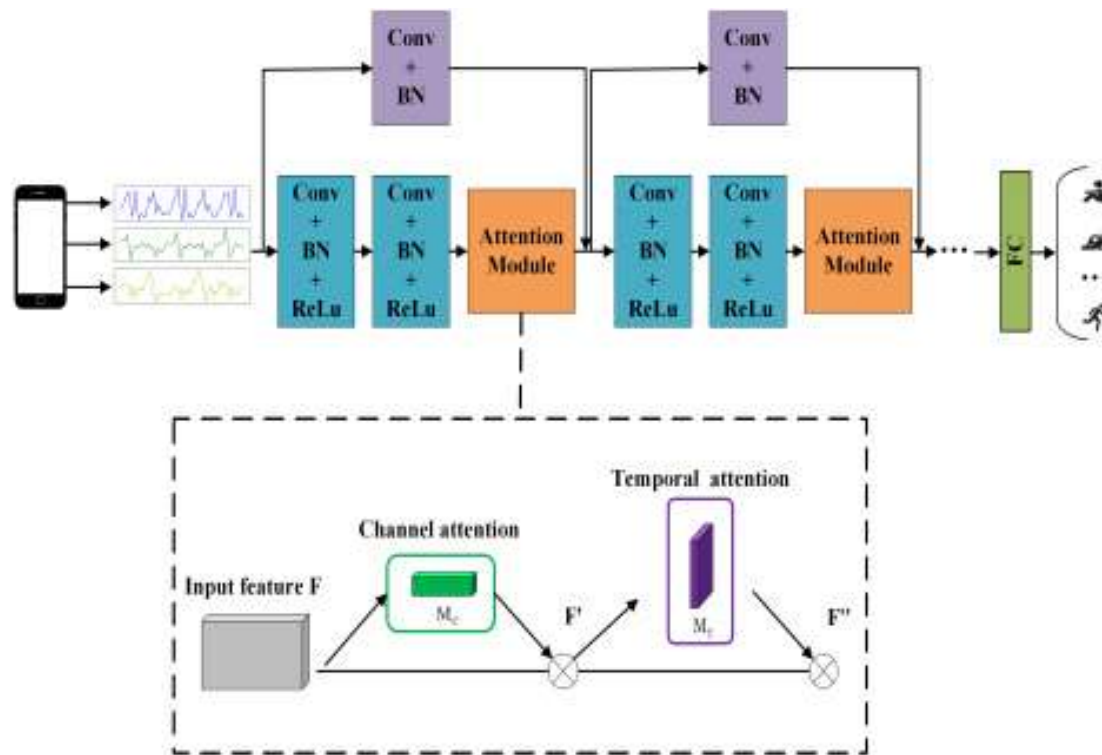
$$W_T = \sigma \left(f^{n \times 1} ([g_1(A); g_2(A)]) \right)$$

$n \times 1$ 表示卷积滤波器的大小，
 g_1, g_2 表示平均池化和最大池化。
过滤器大小被调整并设置为 7×1 。

研究方法

- 残差网络 (Residual Network)

如果没有足够的训练数据，构建大规模DNN容易出现过拟合。因此，残差网络只包含3组卷积层，每组卷积层包含一个2个卷积层组成的构建块。每个构建块的核大小相同。总共有6个卷积层，每个构建块都有一个捷径连接，它的作用是跳过该块进行恒等映射，并添加该块的残差映射，成为最终的底层映射。残差网络由双层跳过实现，其中包含ReLU和批归一化处理。

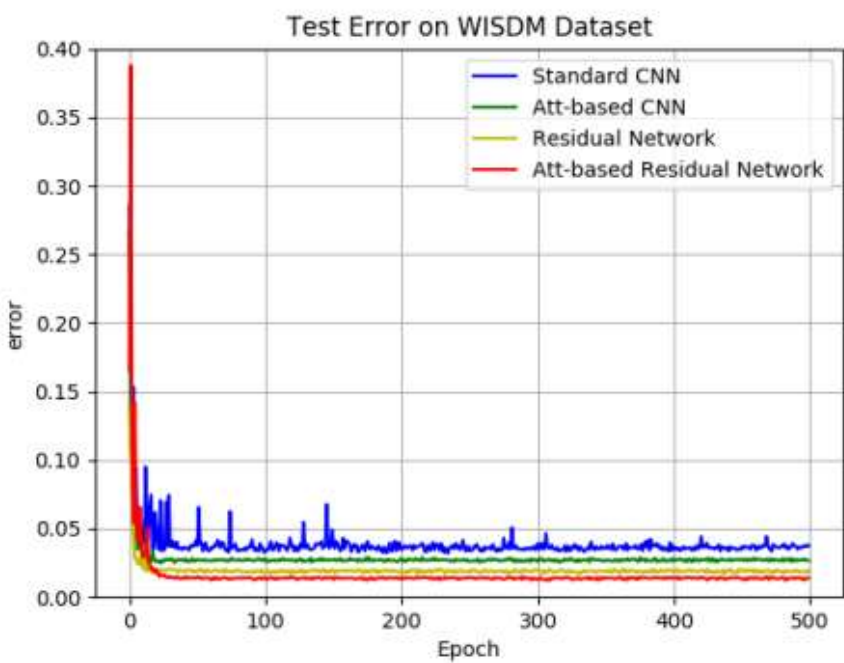


实验

- 实验结果及性能比较

WISDM数据集:

特性	说明
数据集名称	WISDM (Wireless Sensor Data Mining)
数据采集设备	智能手机
传感器	加速度传感器 (X轴、Y轴、Z轴)
数据采集方式	志愿者佩戴智能手机进行日常生活活动记录
活动类型标签	Walking、Jogging、Upstairs、Downstairs等
数据集版本	原始版本和预处理版本
用途	人体活动识别、移动设备运动分析研究
数据集来源	WISDM项目或相关研究论文
数据格式	时间戳、加速度数据、活动标签等
注意事项	遵循数据集使用规范，了解数据清洗和预处理过程



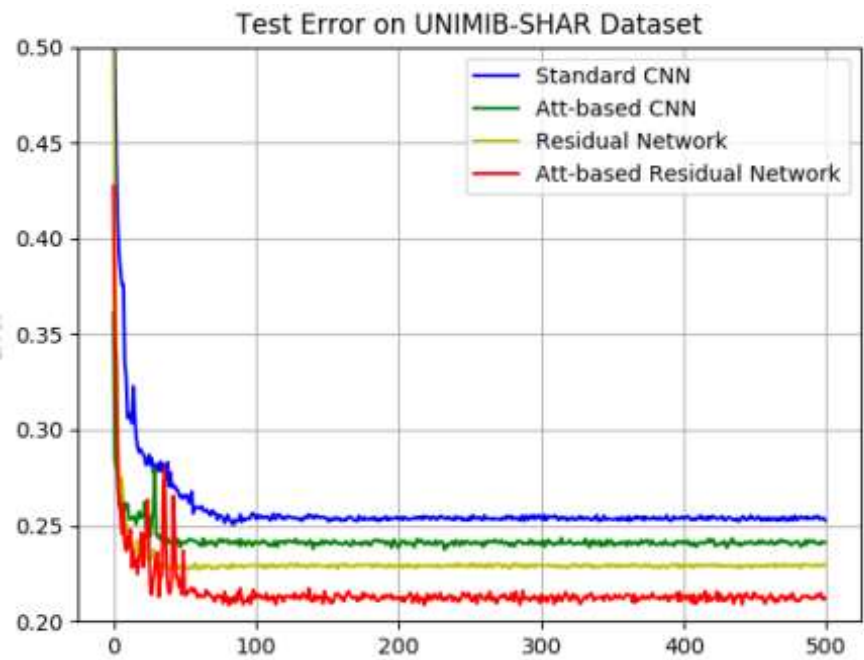
Model	Test Acc	Params
Standard CNN	74.83%	1.62M
Att-based CNN	76.07%	1.63M
Residual Network	77.12%	2.38M
Att-based Residual Network	79.03%	2.40M
Li <i>et al.</i> 2018 [35]	74.97%	-
Long <i>et al.</i> 2019 [34]	76.04%	-
Teng <i>et al.</i> 2020 [33]	78.07%	-

实验

- 实验结果及性能比较

UNIMIB SHAR数据集:

特性	说明
数据集名称	UNIMIB SHAR (University of Milano-Bicocca Smartphone Human Activity Recognition)
数据采集设备	智能手机
传感器	加速度传感器 (X轴、Y轴、Z轴)
数据采集方式	志愿者进行日常生活活动记录
活动类型标签	Walking、Jogging、Stairs Up、Stairs Down等
数据集来源	University of Milano-Bicocca研究项目
数据格式	时间戳、加速度数据、活动标签等
数据量	约3000个样本 (来自17名志愿者的多个实验)
用途	人体活动识别、姿势估计、移动设备运动分析等
特点	包含日常生活中的多种活动, 适用于多类别分类任务
注意事项	遵循数据集使用规范, 了解数据收集过程和标签含义



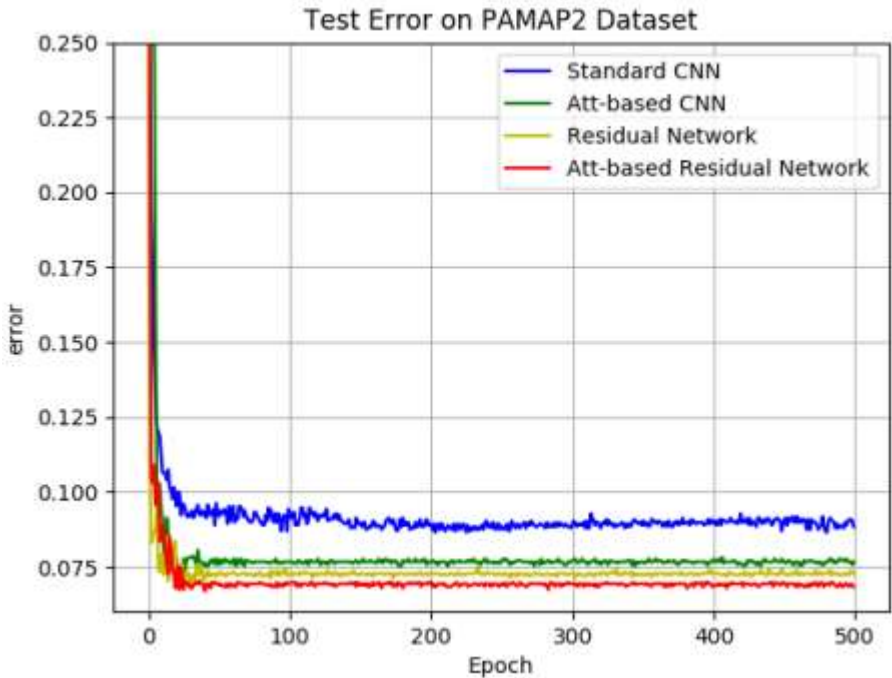
Model	Test Acc	Params
Standard CNN	74.83%	1.62M
Att-based CNN	76.07%	1.63M
Residual Network	77.12%	2.38M
Att-based Residual Network	79.03%	2.40M
Li <i>et al.</i> 2018 [35]	74.97%	-
Long <i>et al.</i> 2019 [34]	76.04%	-
Teng <i>et al.</i> 2020 [33]	78.07%	-

实验

- 实验结果及性能比较

PAMAP2数据集:

特性	说明
数据集名称	PAMAP2 (Physical Activity Monitoring Data Set 2)
数据采集设备	多个传感器, 包括心率监测、加速度计等
数据采集方式	志愿者进行不同的运动和活动, 同时监测多个生理信号
活动类型标签	Walking、Running、Cycling、Sitting等
数据集来源	Swiss Federal Institute of Sport Magglingen (SFISM)
数据格式	时间戳、加速度数据、心率等多维度数据
数据量	约2055个样本 (来自9名志愿者的多个实验)
用途	人体活动识别、生理信号分析、运动健康监测等
特点	多传感器数据, 包含多种活动和生理信号
注意事项	遵循数据集使用规范, 了解数据收集过程和标签含义



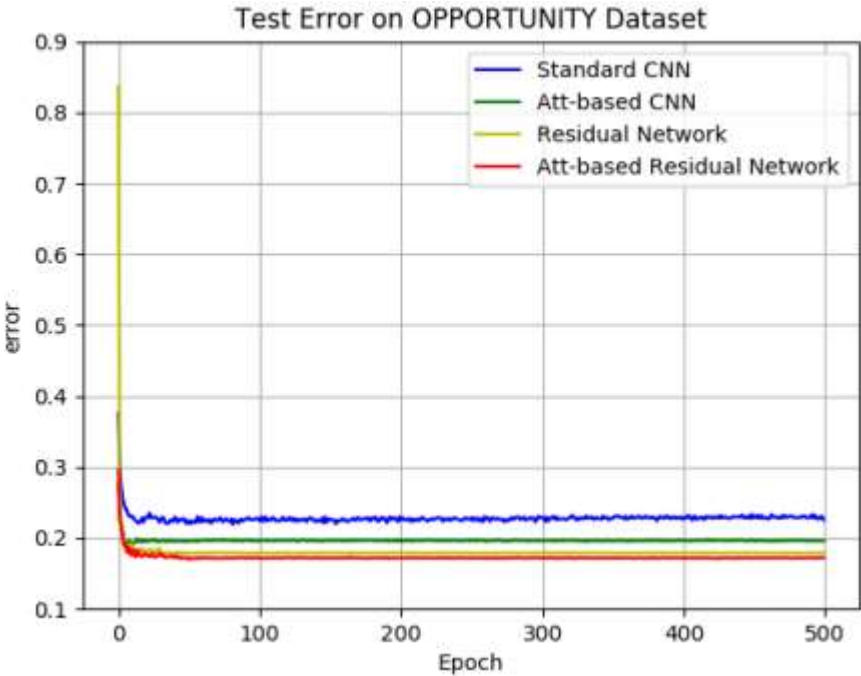
Model	Test Acc	Params
Standard CNN	91.21%	2.73M
Att-based CNN	92.45%	2.75M
Residual Network	92.72%	3.48M
Att-based Residual Network	93.16%	3.51M
Zeng <i>et al.</i> 2018 [13]	89.96%	-
Ma <i>et al.</i> 2018 [12]	89.30%	-
Teng <i>et al.</i> 2020 [33]	92.97%	-

实验

- 实验结果及性能比较

OPPORTUNITY数据集:

特性	说明
数据集名称	OPPORTUNITY Activity Recognition Data Set
数据采集设备	多个传感器，包括加速度计、陀螺仪、磁强计等
数据采集方式	志愿者进行不同的运动和活动，同时监测多个传感器数据
活动类型标签	Walking、Sitting、Standing、Lying Down等
数据集来源	European project HUMABIO
数据格式	时间戳、多种传感器数据、活动标签等
数据量	约427,444个样本（来自12名志愿者的多个实验）
用途	人体活动识别、行为分析、环境感知等
特点	包含多种传感器数据，适用于多模态传感器融合研究
注意事项	遵循数据集使用规范，了解数据收集过程和标签含义



Model	Test Acc	Params
Standard CNN	77.53%	1.15M
Att-based CNN	80.42%	1.17M
Residual Network	82.05%	1.55M
Att-based Residual Network	82.75%	1.57M
Zeng <i>et al.</i> 2014 [37]	76.83%	-
Ordóñez <i>et al.</i> 2016 [38]	78.90%	-
Hammerla <i>et al.</i> 2016 [36]	74.50%	-
Teng <i>et al.</i> 2020 [33]	81.00%	-

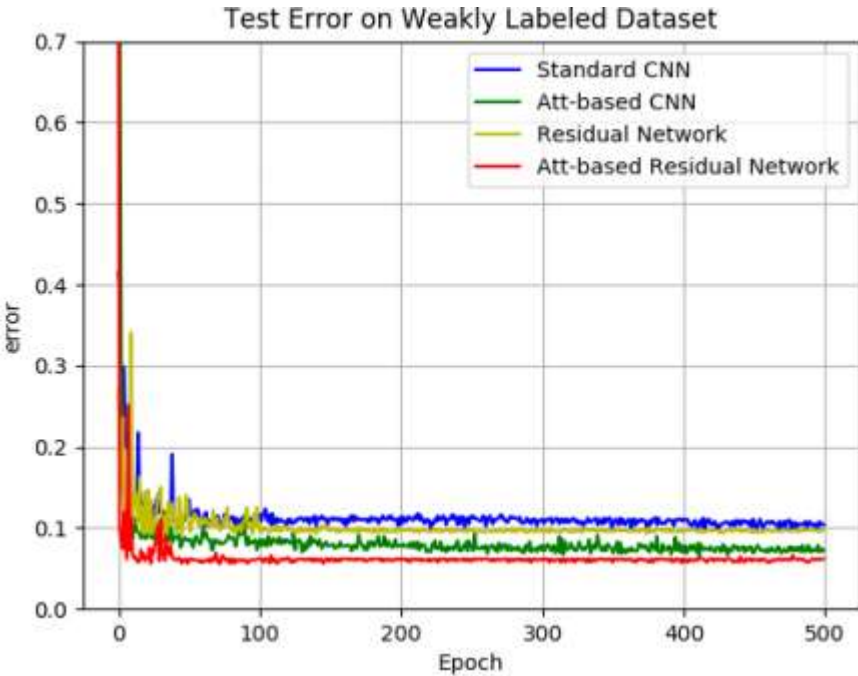
实验

- 实验结果及性能比较

弱标记数据集:

Activity	Label	Number
going upstairs	0	20194
going downstairs	1	18611
jumping	2	14088
jogging	3	23264
total	-	76157

特性	说明
数据采集设备	iphone7中的三轴加速度传感器
数据采集方式	10名参与者将智能手机放在他们的裤袋中，以进行人类日常活动
活动类型标签	志愿者进行不同的运动和活动，同时监测多个传感器数据
活动类型标签	Walking、jogging、jumping、going up、down stairs等
数据格式	时间戳、加速度数据、活动标签等
数据量	时间戳、多种传感器数据、活动标签等
数据量	约76157个样本



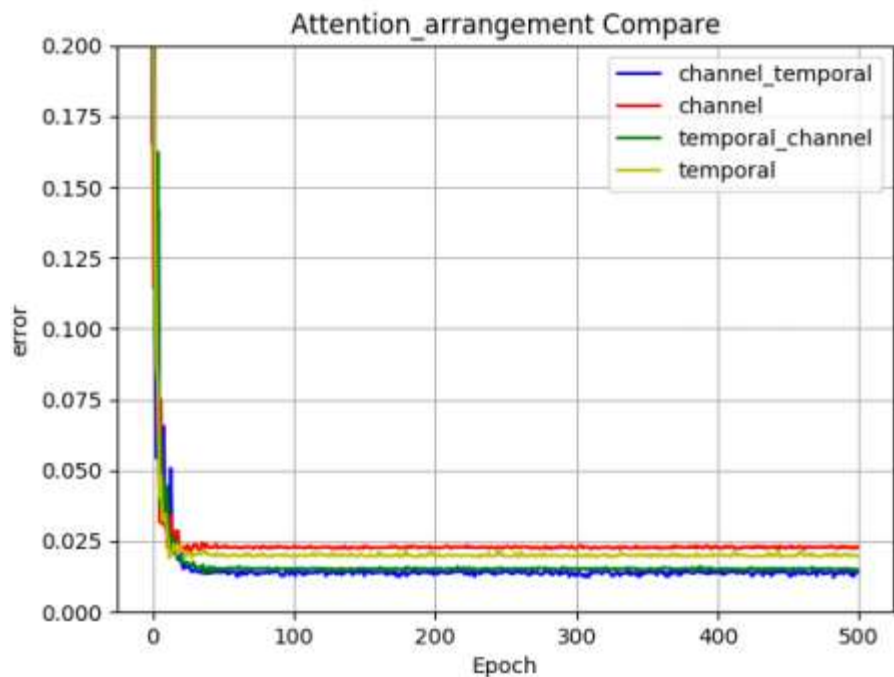
Model	Test Acc	Params
Standard CNN	89.86%	0.70M
Att-based CNN	93.62%	0.71M
Residual Network	90.93%	0.94M
Att-based Residual Network	94.86%	0.95M
DeepConvLSTM [14]	90.04%	-
Attention-based CNN [14]	93.83%	-

实验

- 消融实验

研究了以不同方式排列通道和时间注意子模块的影响：顺序通道-时间注意、顺序时间-通道注意、单独通道注意和单独时间注意。

通道注意倾向于告诉我们应该注意什么，而时间注意倾向于告诉我们应该注意哪里。



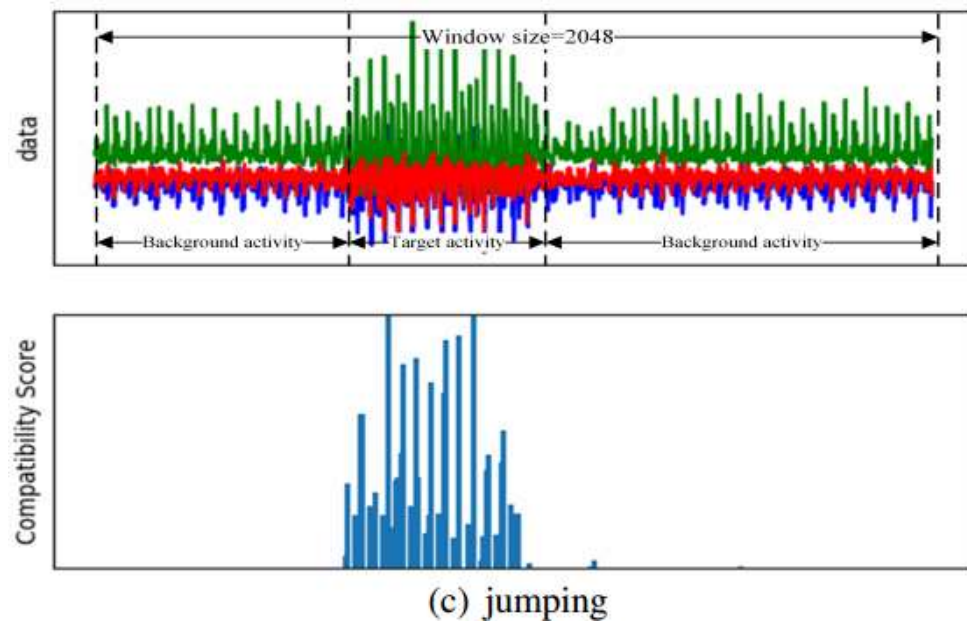
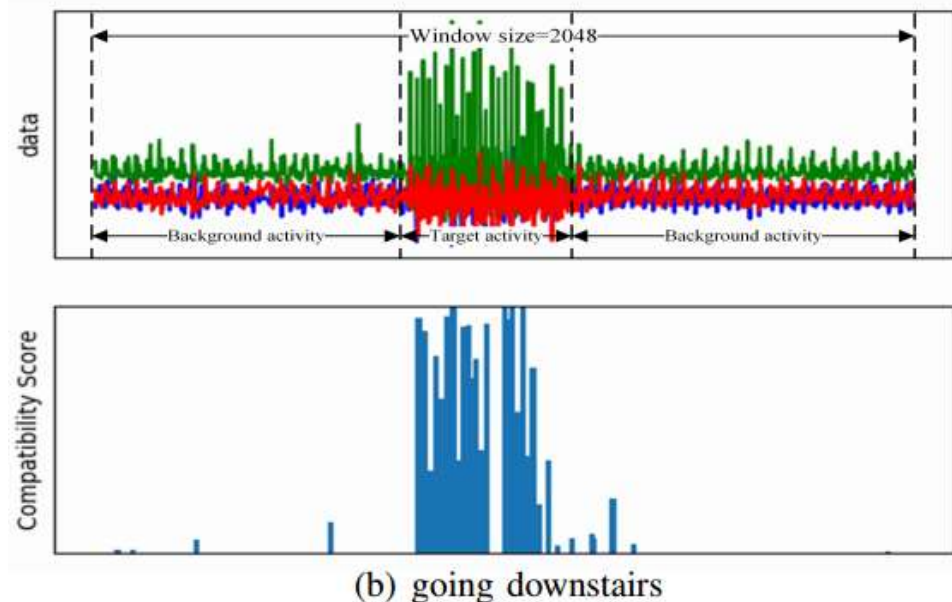
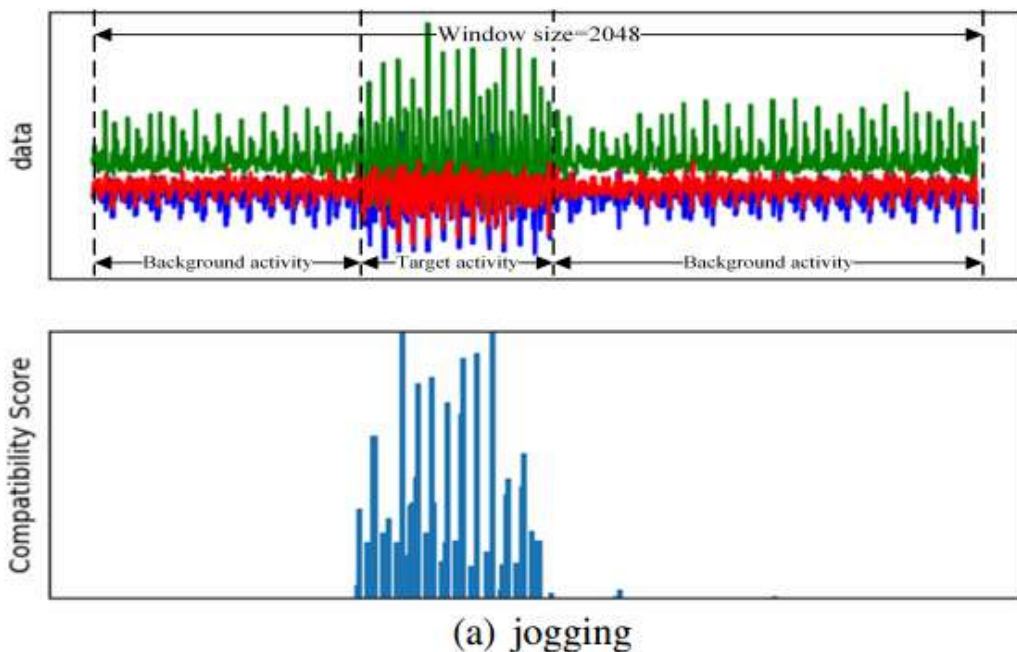
Module	Test Acc	Params
Channel	98.09%	2.32M
Temporal	98.12%	2.30M
Temporal+Channel	98.62%	2.32M
Channel+Temporal	98.85%	2.32M

实验

- 可视化分析

为了更好地理解信道和时间注意是如何提高准确性的，我们提供了可视化分析。

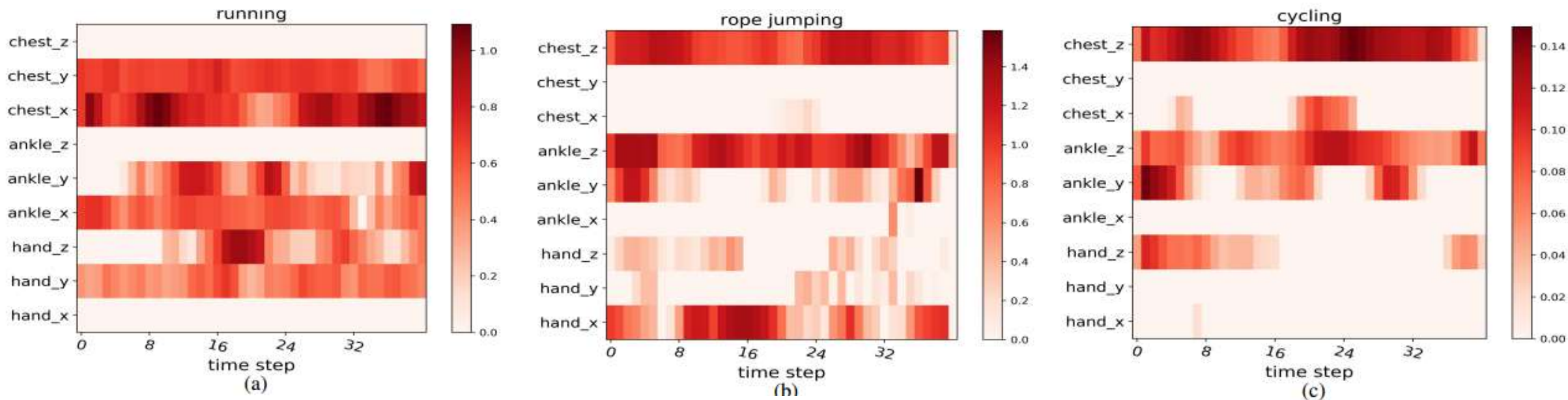
与严格标记的基准数据集不同，整个信号段包含目标活动，以及大量背景活动噪声(即行走)。我们的注意力方法倾向于高度重视感兴趣活动的一小部分，而忽略这些背景活动，这表明DanHAR可以从长序列中捕获重要部分，以提高性能以及传感器信号的可理解性。



实验

- 可视化分析

双注意机制可以自动学习不同传感器的优先级，在特征融合方面表现更好，在处理多模态HAR任务时取得了更好的性能。



a、对于跑步活动，通道注意力高度重视手部传感器(Hand_y, Hand_z)、胸部传感器(Chest_x, Chest_y)和脚踝传感器(Ankle_x, Ankle_y)。

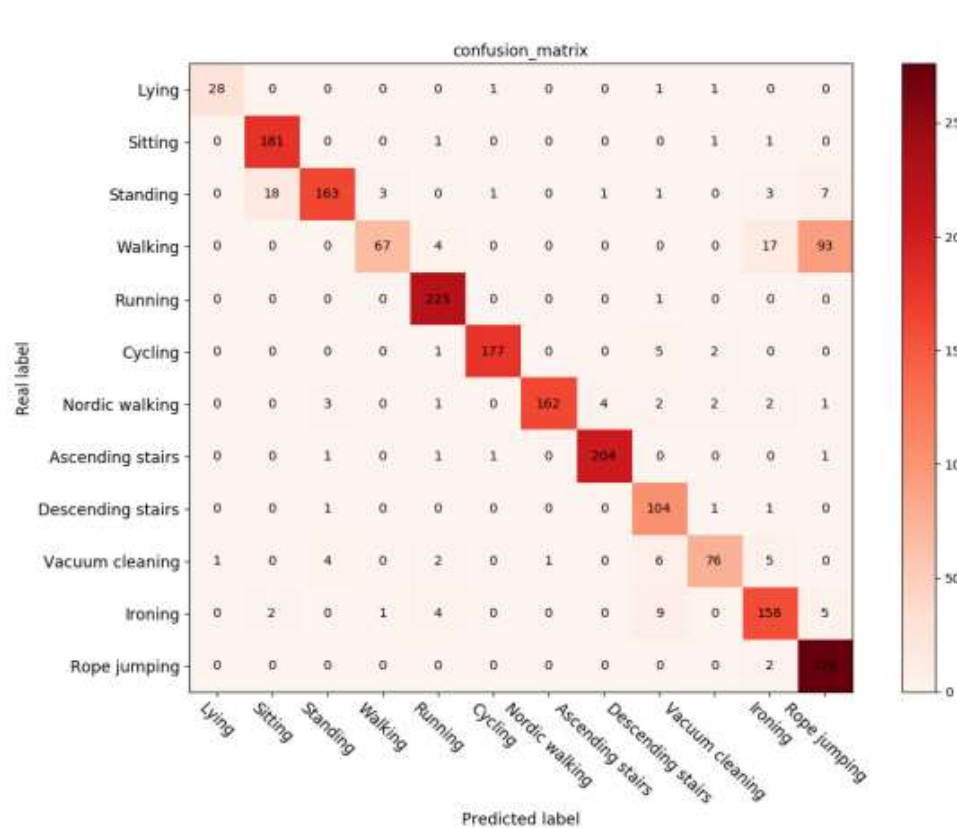
b、对于跳绳活动，通道注意力更侧重于手部传感器(Hand_x)、胸部传感器(Chest_z)和脚踝传感器(Ankle_z)。

c、对于自行车运动，我们的方法更关注胸部传感器(Chest_z)和脚踝传感器(Ankle_z)。结果是可以解释的，并且与我们的直觉一致。

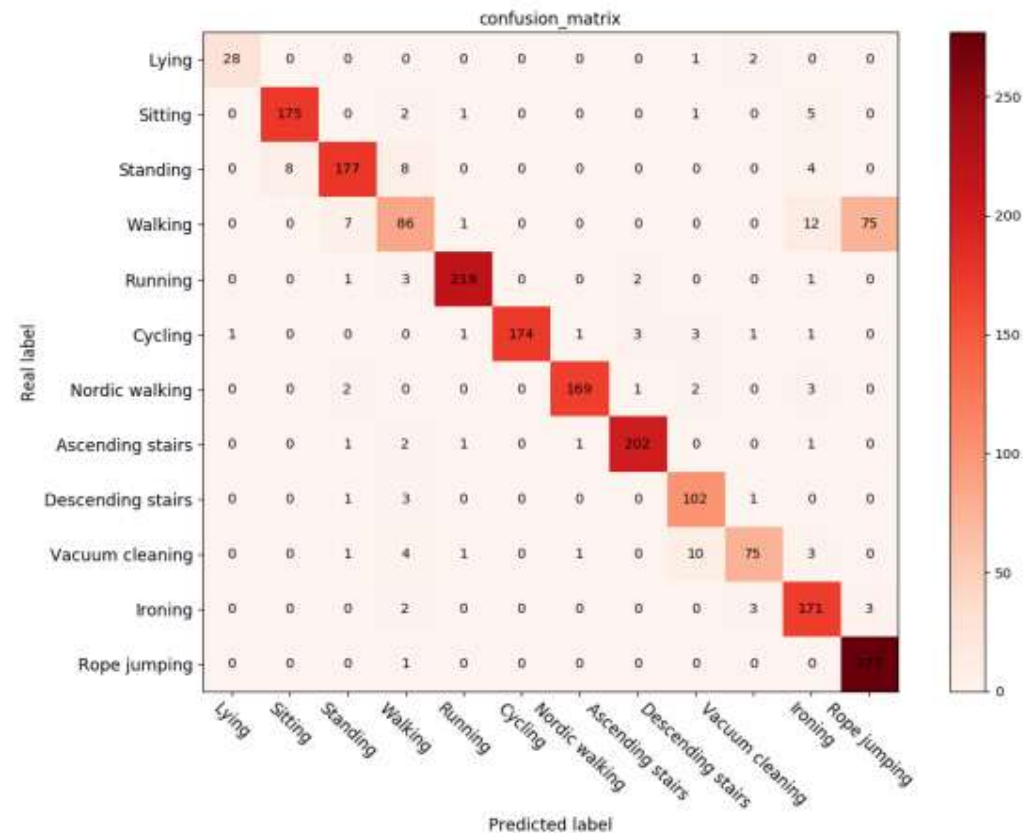
实验

- 可视化分析

最后在PAMAP2数据集上比较了DanHAR和残差网络的混淆矩阵，许多错误分类是由于两个非常相似的活动类别(如跳绳和步行)之间的混淆，这在以前被认为是非常难以区分的，而从结果来看，残差网络有93个错误，而DanHAR只有75个错误分类，这证实了本文方法在识别精度上的优越性。



Residual network



DanHAR

结论

- 简述本文结论

从多模态传感器数据中识别人类活动是一项具有挑战性的任务。在本文中，我们首次提出了称为DanHAR的双注意方法，该方法同时使用[通道注意](#)和[时间注意](#)来更好地理解和改进用于各种多模态HAR任务的深度网络。DanHAR采用混合框架，将[双注意机制与CNN或残差网络](#)相结合，融合多模态传感器信息，具有更好地捕捉多模态感知数据时空格局的能力。

提问思路

- 动机合理吗？要解决问题有价值吗？这个问题背后的本质是什么（关键技术问题/难点）
- 方法背后的假设是什么？假设什么情况下不成立？不成立的时候方法还有用吗，如何改进？
- 提出的方法是从本质上解决问题吗？
- 实验能支持论文的结论吗？数据的量/分布/变化足够支持得到可靠的结论吗？
- 实验设置符合真实世界的情况吗？和真实世界有什么不同？
- 实验结果之所以好，主要是方法的那个部分在起作用？
- 从这篇论文中还能得出哪些结论？这篇论文如何加深了你对开放环境/自监督/活动识别的理解
- 要解决的问题与“开放环境中的自监督活动识别”有什么联系/相同的地方，对我们工作的帮助或启发
- 这篇论文的写作/结构/组织好在哪里？他是如何包装自己的贡献的，如何使得工作看起来又新颖又有价值？