

MetaSense: Few-Shot Adaptation to Untrained Conditions in Deep Mobile Sensing

Taesik Gong, Yeonsu Kim, Jinwoo Shin, and Sung-Ju Lee

摘要

最近深度学习和硬件支持的改进为移动传感提供了新的突破; 我们可以在人工智能驱动的设备上享受上下文服务和移动医疗。然而, 大多数相关研究只有在训练数据与目标数据分布具有一定程度的相似性时才能表现良好, 而在实践中, 特定用户的行为和设备会使传感器输入有所不同。因此, 在不同的用户和设备条件下, 这样的应用程序的性能可能会受到影响, 因为在如此无数的条件下训练深度模型是不可实现的。为了缓解这一问题, 我们提出了 **MetaSense**, 这是一种自适应深度移动传感系统, 仅利用来自目标用户的少数 (例如, 一个或两个) 数据实例。**MetaSense** 采用元学习来学习如何适应目标用户的情况, 通过在离线训练中排练由我们独特的任务生成策略生成的多个类似任务。训练后的模型能够在少量数据可用的情况下快速适应目标用户的情况。我们对运动和音频传感器的真实痕迹进行了评估, 结果表明, 在准确性方面, **MetaSense** 不仅比最先进的迁移学习高出 18%, 基于元学习的方法高出 15%, 而且对目标用户的适应时间也大大减少。

关键词: 人类活动识别; 移动传感; 深度学习; 元学习; 小样本学习

1 引言

智能设备 (智能手机和可穿戴设备) 通过不断改进的计算资源和感应功能丰富了我们的生活。在深度学习和移动 AI 处理器的支持下, 移动传感已针对多种应用开发, 包括人类活动识别^[1], 声音上下文识别^[2], 复杂的体育锻炼识别^[3], 无设备认证^[4], 手语识别^[5], 甚至可以预测一个人的情绪状态等。

尽管这些研究为智能设备提供了新颖的传感应用可能性, 但它们仍面临性能挑战。性能下降主要是由于存在大量个体条件所致; 我们将单个条件定义为在部署给用户时发生的多种依赖关系 (例如, 用户的行为和设备) 的组合, 通常与训练感知模型的条件不同。个体条件下的这些差异阻碍了移动感知系统无法为潜在的用户正常使用, 因此克服这一问题已成为一个重要的研究问题。

在繁琐的数据收集和标记过程中需要昂贵的成本和用户的努力, 这是不切实际的^[6]。是否有一个通用的解决方案, 以最小的用户数据克服移动感知中的个人差异? **MetaSense**, 它可以通过从目标用户提供少量的有标签数据, 来适应目标用户的深度感知模型。**MetaSense** 通过我们的任务生成策略提供多个类别任务来训练基本模型, 其中每个任务都是指适应特定条件。对于每个任务, 基本模型都会经历各种单独的条件, 并使它的参数适应新的/未知的条件。将基本模型部署给用户后, 只需少量有标签数据就可以适应目标用户的条件。由于 **MetaSense** 仅需非常少的目标用户有标签数据, 因此可以大大减少每个用户的繁琐数据收集过程, 同时仍可实现理想的性能。**MetaSense** 适用于任何深度学习模型 (即模型不可知), 并且不将其覆盖范围限制为特定类型的传感器或应用程序 (即条件不可知)。与传统的深度神经网络训练相比, **MetaSense** 还大大减少了适应时间, 这无疑对资源受限的移动设备有利。

2 相关工作

2.1 元学习

元学习 (也称为“学会学习”) 是一种模仿人类快速有效地适应新概念的能力的尝试。打个比方, 如果一个人已经知道如何骑旱冰鞋, 他就可以很容易地学会如何骑冰鞋。将这一概念引入机器学习, 元学习的目的是建立一个模型, 快速适应训练过程中从未遇到过的新情况。大多数现有的元学习方法都局限于用大量的图像数据 (例如, Omniglot^[7]) 来演示它们的性能。然而, 在移动传感中, 稀缺的数据集和前面提到的无数依赖组合使得情况有所不同, 因此需要一个独特而复杂的应用元学习的解决方案。

2.2 原型网络

原型网络是一种最先进的基于元学习的少镜头学习算法^[8]。在给定少量训练数据的情况下, PN 在嵌入空间中生成原型, 每个原型是每个类的代表。在推理中, PN 使用欧几里得距离度量对最近的原型进行分类。在移动感知领域内, 原型神经网络中, 它需要大量的数据, 通常超过数千个, 来训练大量的参数, 同时避免过度拟合到训练数据, 比较繁琐。

3 本文方法

3.1 本文方法概述

本文的方法, 如图 1 所示:

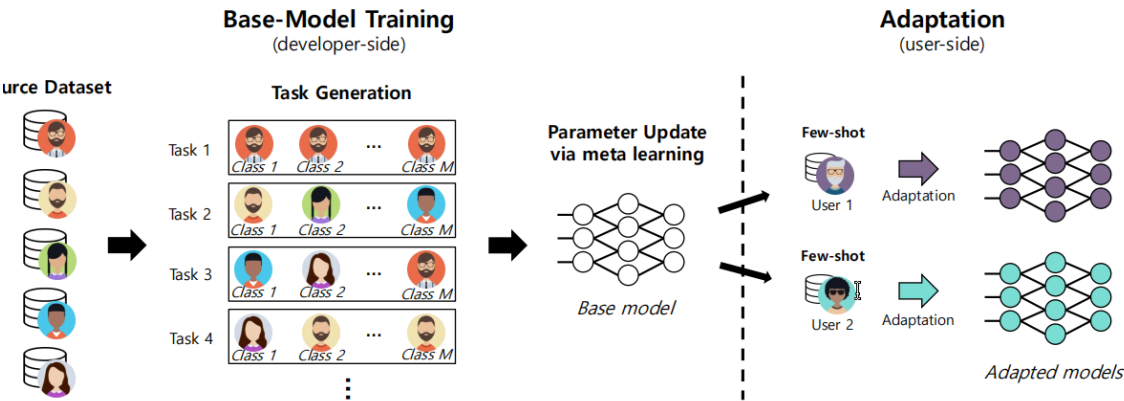


图 1: MetaSense 框架

MetaSense 通过两个步骤训练深度感知模型, 即基础模型训练和适应。具体来说, MetaSense 使基础模型只需要目标用户的少量有标签数据和梯度步骤就能学习如何适应新的条件。基本模型在一组任务上进行训练, 其中每个任务都是从源数据集生成的。个体任务模拟模型在一个新的未经训练的条件下执行的情况。经过训练, 基础模型通过几次拍摄就掌握了如何适应新环境的知识。在自适应步骤中, 目标用户向模型提供一些镜头, 模型用少量的梯度步骤对其参数进行自适应。在适应过程之后, 模型已经为目标用户的条件做好了准备。

3.2 任务生成

与现有的元学习方法不同, 在现有的元学习方法中, 任务是通过从大型可用数据集中随机采样生成的, 如何有效地利用有限的源数据集是将元学习应用于数据密集型移动感知的独特挑战。在移动感知中, 将每个任务视为每个单独的条件。因此 MetaSense 任务生成的目标是在给定源数据集的情况下

生成多样且真实的个人条件，以便 MetaSense 可以教会基础模型如何通过各种任务适应可能的条件变化。

3.3 参数更新

通过元学习来训练基模型的参数。具体来说，MetaSense 使用模型不可知元学习 (MAML)^[9]来更新参数。MAML 适用于任何使用梯度下降 (模型不可知) 的深度神经网络，只需要几个梯度步骤来更新模型。MAML 的假设是存在一些初始参数，这些参数可以转移到只需目标用户少量有标签样本的新任务中。MAML 对初始参数进行训练，使训练后的参数能够适应任务的变化。我们采用 MAML 的动机是，对于深度传感模型，存在有效的初始参数，这些参数可以在各个条件之间转换，因此参数可以在几个梯度步骤内适应目标条件。

由于该模型是基于原型网络衍生的，所以对每个任务分支撑集、查询集，对每个查询集算出交叉熵损失，在提供目标用户的少量有标签数据时，进行参数更新。

4 复现细节

4.1 与已有开源代码对比

该论文的源代码开源，对应的 GitHub 网站：https://github.com/TaesikGong/MetaSense_public，我在复现时借鉴此源代码，使用论文自带的两个数据集进行复现。使用数据增强策略，增加小样本学习中每个用户 shot 的数目，从而使得模型的准确率提升，能够更好的泛化新用户。

4.2 实验环境搭建

实验使用 AMAX 服务器，该服务器有 4 个 CPU，每个 CPU 有 12 核，CPU 型号为 Intel(R) Xeon(R) CPU E5-2650 v4 @ 2.20GHz，内存容量 96GB。数据集为 HAR（人类活动识别 IMU 数据集）、语音数据集。数据已经切分好训练集和测试集，将这些数据集加载到服务器中，选择相应训练网络的方法（原型网络、迁移学习等方式）。在数据未加载到模型里面时使用数据增强的方法，语音数据集时域频域的掩盖、时间扭曲，HAR 数据集加入随机噪声。然后将这些数据加载到模型中即可得到增强后的结果，在与未增强时的结果作比较。

4.3 创新点

在现有数据集的基础上，用了简单的数据增强，让目标用户的少量样本变多，从而在更新网络参数时能够更好的泛化那些未知的用户。

5 实验结果分析

原论文中使用的是 HAR 和 Speech 两个数据集，其 HAR 数据集的性能如图 2 所示。

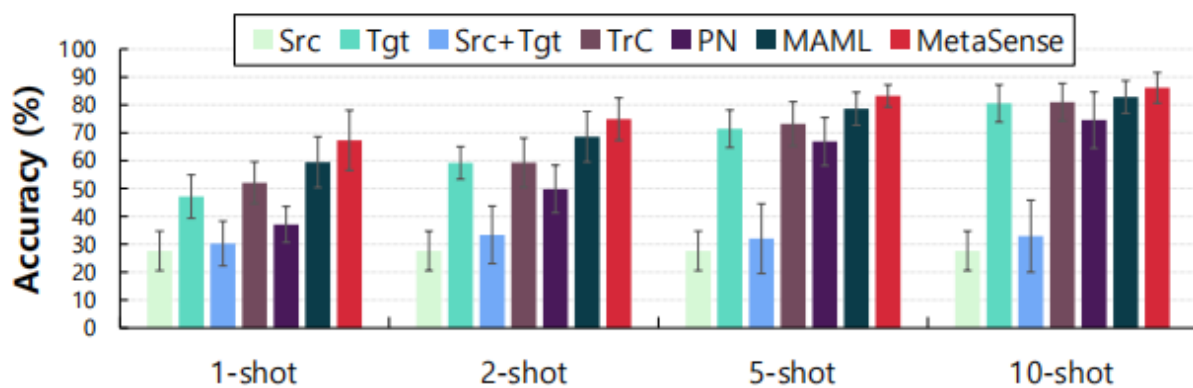


图 2: 原文 HAR 准确率

我在复现阶段由于没有达到原文中的准确率，所以我对该数据集做了一些简单的数据增强，与没做数据增强相比，变化不大，对比如图 3和图 4所示。

未增强

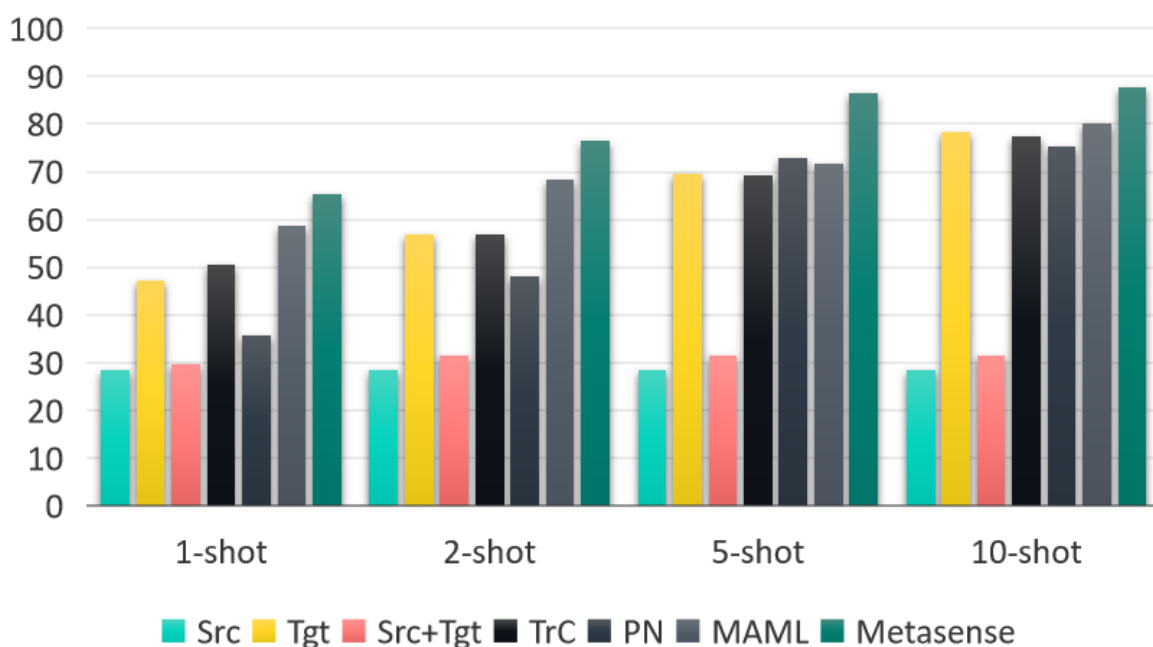


图 3: HAR 数据集未增强

增强

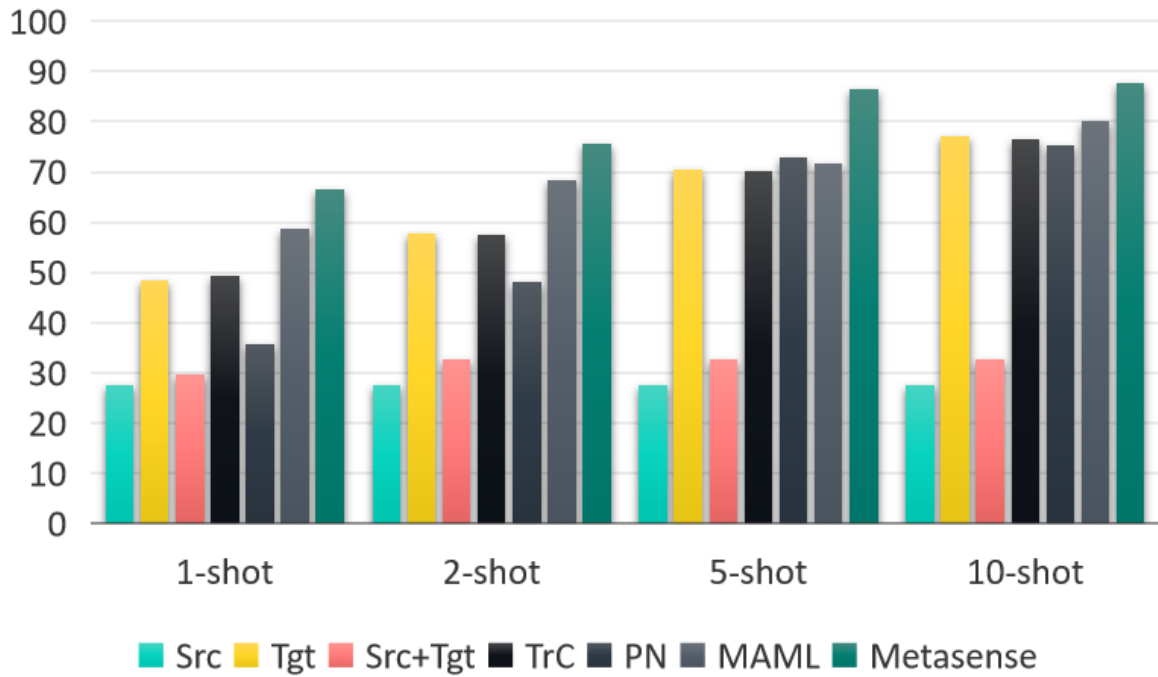


图 4: HAR 数据集增强

对于 Speech 数据集，其原文中的准确率如图 5 所示。

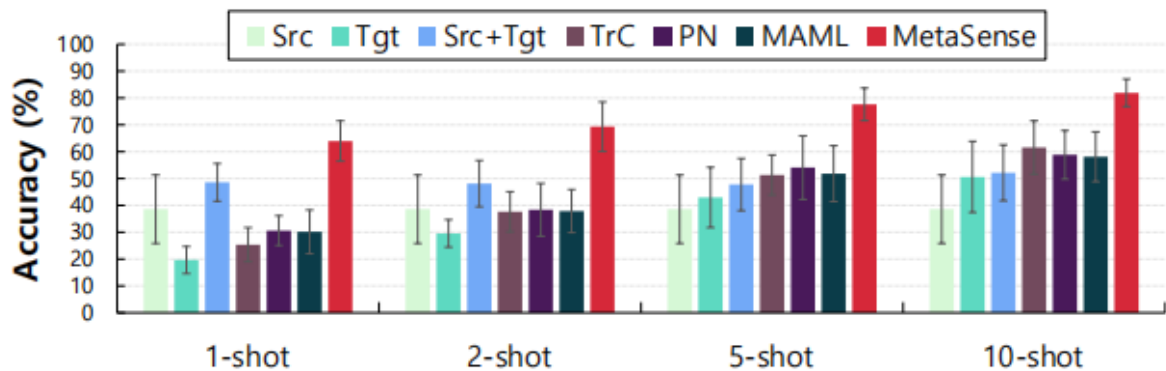


图 5: 原文 Speech 数据集准确率

我在使用这个数据集训练测试模型时，准确率与原文相差较大，一开始想到做简单的数据增强也是对这个数据集而言的。做完数据增强之后，准确率确实有所提升，但是与原文相比还是相差 2%-3%。其结果如图 6 和图 7 所示。

未增强

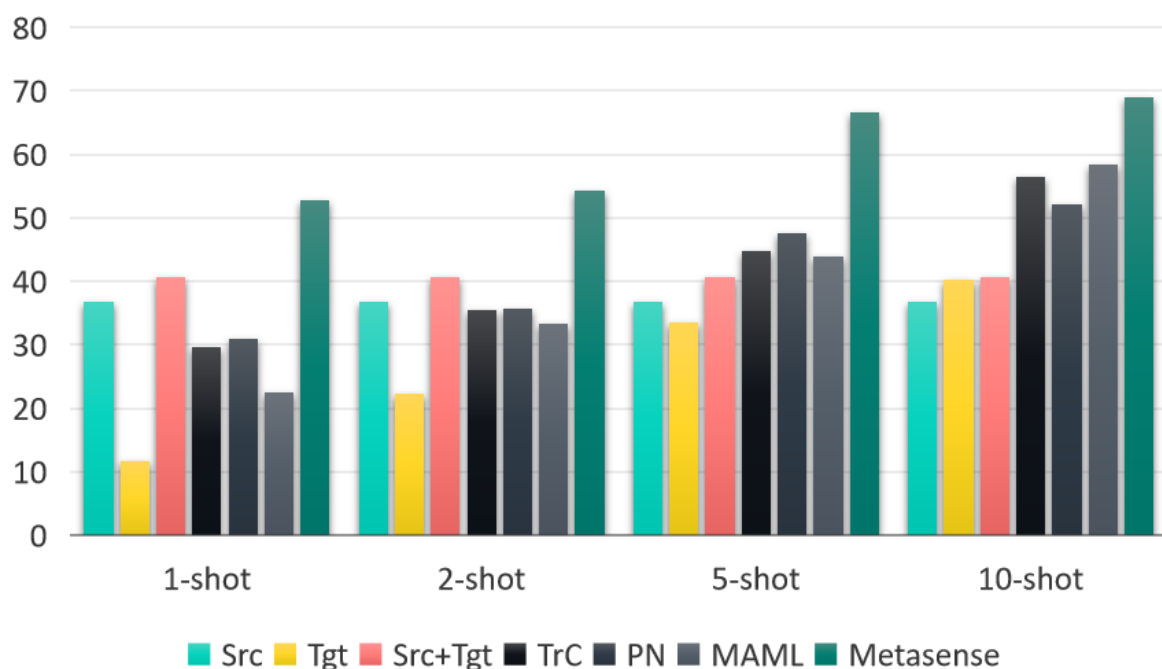


图 6: Speech 数据集未增强准确率

增强

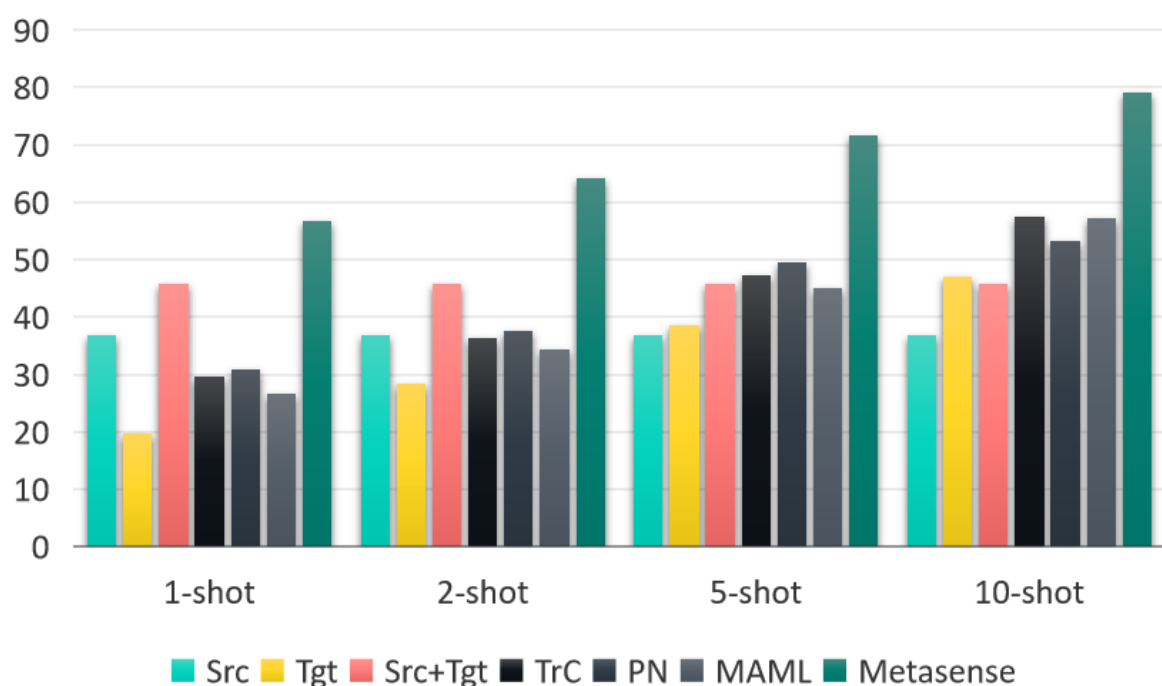


图 7: Speech 数据集增强准确率

6 总结与展望

小样本学习的方法确实能够提升跨人、跨设备的准确率，我用公开生理信号数据集去训练该模型时，准确率不高，可能是该系统所采用的小样本特征提取器不适合生理信号数据集，后续可以采用 MAML 的各种变体，来尝试着找到我这个领域的最合适的特征提取器，帮助我解决我现阶段遇到的

问题。

参考文献

- [1] RADU V, TONG C, BHATTACHARYA S, et al. Multimodal deep learning for activity and context recognition[J]. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2018, 1(04): 1-27.
- [2] LANE N D, GEORGIEV P, QENDRO L. DeepEar: robust smartphone audio sensing in unconstrained acoustic environments using deep learning[J]. Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, 2015: 283-294.
- [3] SORO A, BRUNNER G, TANNER S, et al. Recognition and Repetition Counting for Complex Physical Exercises with Deep Learning.[J]. Sensors, 2019, 19(3): 714.
- [4] CHAUHAN J, RAJASEGARAN J, SENEVIRATNE S, et al. Performance Characterization of Deep Learning Models for Breathing-based Authentication on Resource-Constrained Devices[J]. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2018, 2(4): 1-24.
- [5] DAI Q, HOU J, YANG P, et al. The sound of silence: end-to-end sign language recognition using smart-watch[J]. Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking, 2017: 462-464.
- [6] STISEN A, BLUNCK H, BHATTACHARYA S, et al. Smart devices are different: Assessing and mitigating mobile sensing heterogeneities for activity recognition.[J]. Proceedings of the 13th ACM conference on embedded networked sensor systems, 2015: 127-140.
- [7] Lake, M B, Salakhutdinov, et al. Human-level concept learning through probabilistic program induction [J]. Science, 2015, 350(6266): 1332-1338.
- [8] HUANG G, LIU Z, VAN L. Prototypical networks for few-shot learning[J]. In Proceedings of the IEEE conference on computer vision and pattern recognition., 2017: 4700-4708.
- [9] FINN C, ABBEEL P, LEVINE. S. Model-agnostic metalearning for fast adaptation of deep networks[J]. International conference on machine learning. PMLR, 2017: 1126-1135.