

具备合适教师的多出口自蒸馏

孙武杰, 陈德仿, 王灿, 叶德仕, 冯雁, 陈纯
浙江大学计算机科学与技术学院, 中国杭州市, 310000

摘要: 多出口架构允许早停推理以减少计算成本, 这使其可以在资源受限的情况下使用。最近的研究将多出口架构与自蒸馏相结合, 以在不同网络深度上同时实现高效率和卓越性能。然而, 现有方法主要从深层出口或单一集成中传递知识, 以指导所有出口, 而没有考虑学生和教师之间不适当的学习差距可能会降低模型性能, 特别是对于浅层出口而言。为解决这个问题, 提出具备合适教师的多出口自蒸馏方法, 为每个出口提供多样化且适当的教师知识。在我们的方法中, 根据不同可训练的集成权重, 从所有出口获得多个集成教师。每个出口从所有教师那里接收知识, 并重点关注其所对应的主教师, 以保持适当的学习差距并实现高效的知识传递。通过这种方式, 我们的方法在保证学习效率的同时实现了多样化的知识蒸馏。在CIFAR-100、TinyImageNet以及3个细粒度数据集上的实验结果表明, 我们的方法在各种网络架构中始终优于最先进的多出口自蒸馏方法。

关键词: 多出口架构; 知识蒸馏; 学习差距

<https://doi.org/10.1631/FITEE.2200644>