

江俊广

(+86) 131-2683-0206 · JiangJunguang1123@outlook.com · [个人主页](#)

教育背景

清华大学, 软件工程, 硕士研究生 2020.8 - 2023.7

- 曾获得过综合优秀奖 (华为奖学金), 入选了 2022 腾讯犀牛鸟精英人才计划
- 导师为龙明盛, 研究方向为深度学习、迁移学习、多任务学习, 曾担任过 ICML、NIPS 等会议审稿人

清华大学, 软件工程, 工学学士 2015.8 - 2020.7

- GPA 3.9/4.0, 年级排名 2/83, 所有专业课程均在 95 分以上
- 曾获得国家奖学金, 综合优秀奖 (华为奖学金、恒大奖学金) 等荣誉

科研论文

Transferability in Deep Learning: A Survey, 第一作者

- 从深度学习的全生命周期 (预训练、自适应、评估) 介绍了如何获得、利用以及评估迁移性
- 被 *Foundations and Trends in Machine Learning* (Michael Jordan 主编, 收录机器学习领域的专著) 接收
- 负责撰写的部分包括: 针对预训练的生成学习和对比学习, 迁移学习中的灾难性遗忘问题与负迁移问题, 参数高效的迁移学习, 数据高效的迁移学习, 领域自适应理论部分, 基于统计距离的领域自适应方法, 基于对抗学习的领域自适应方法, 领域翻译, 半监督学习, 迁移性的实验评估

Debiased Self-Training for Semi-supervised Learning, 共同一作 NIPS 2022

- 系统地分析了自训练中存在的偏差问题, 并将其归类成数据偏差 (有限次数据采样导致的偏差) 和训练偏差 (使用错误的伪标签训练导致偏差的扩大)
- 为了消除训练偏差, 将伪标签的生成与使用进行解耦。模型的主塔只使用干净的标注数据进行训练, 并用主塔产生伪标签来训练辅助塔。为了减少数据偏差, 提出了最差情况估计。估计在当前的特征表示下, 模型在无标注数据上所能达到的最大偏差, 然后优化特征来降低最差估计下的偏差。
- 相比于此前的 SOTA 方法, 在包括标准半监督任务的 13 个数据集上获得了平均 6.3% 的提升

Decoupled Adaptation for Cross-Domain Object Detection, 第一作者 ICLR 2022

- 为了缓解分布偏移 (Domain Shift) 对于检测器性能的影响, 提出了解耦自适应框架。核心想法是分而治之: (1) 将原始的跨域物体检测问题分解为几个子问题; (2) 设计合适的算法来解决每个子问题; (3) 协调不同的算法/模块之间的关系, 从而解决最终的问题。这里的解耦包含两次含义: (1) 域自适应过程和检测器训练过程的解耦。(2) 分类域自适应任务和定位域自适应任务的解耦
- 在多个数据集上都取得了目前最好效果, 两个数据集上比现有方法的 mAP 高将近 10 个点。该方法也容易扩展到其他检测架构以及其他检测任务 (例如实例分割等)

Regressive Domain Adaptation for Unsupervised Keypoint Detection, 第一作者 CVPR 2021

- 从理论和实验的角度系统地分析了跨域回归和跨域分类问题的本质区别
- 基于差异散度 (Disparity Discrepancy) 学习理论, 针对以关键点检测为例的回归问题, 提出了首个有效的迁移学习算法。即通过引入了对抗回归器, 来刻画任务相关的分布距离。并通过输出空间的概率密度估计, 来降低有效假设空间的大小, 优化分布距离的估计
- 在多种关键点检测任务上进行了实验, 所有任务的 PCK 指标均提高了 10 个百分点以上

Resource Efficient Domain Adaptation, 第一作者 ACMMM 2020, Oral

- 探究了深度学习中迁移性与模型架构的关系。经过预训练后, 深度模型的允许假设空间 (allowed hypothesis space) 比浅层模型更小, 理论和实验上都获得了一个更紧的泛化误差界, 即更好的迁移性
- 为了将域自适应模型部署到低资源的场景, 提出了针对浅层模型的无监督迁移方法, 即先在深度模型上进行跨域迁移, 然后将其迁移性蒸馏到浅层模型, 并利用数据副本一致性过滤迁移性较差的实例

实习经历

腾讯 | TEG-精准推荐中心-社交广告组 2022.6-2022.12

- 入选腾讯犀牛鸟精英人才计划, 研究课题为多任务学习与多域学习。项目背景: 推荐系统中经常需要对用户的多种反馈进行建模 (多任务), 此外, 也需要对不同场景下的用户反馈行为, 例如信息流、新闻流和视频流等进行联合建模 (多域)

字节跳动 | data-抖音-推荐技术

2022.4-2022.6

- 参与搜推联动项目。为了克服推荐系统中的“信息茧房”问题，从用户的搜索记录中挖掘用户的长期兴趣并用于改进推荐系统的召回与排序。具体的，使用了 **Hive SQL / Spark** 清洗抖音的用户搜索数据，使用 **euler** 框架建立“用户-搜索”图，使用图嵌入的方式从“用户-搜索”图自动地学习用户以及搜索的表征，并将这些表征用于推荐系统的召回
- 该方法可以从用户的搜索行为中，自动地学出一个能够度量用户细粒度兴趣的特征空间

项目经历

开源项目 | 迁移学习算法库

2020.4 至今

- 以第一作者开源了迁移学习算法库 **TLib**，个人贡献了超过 **22k** 行代码
- 对现有的 **30** 多种跨域迁移、跨任务迁移算法进行封装、优化，支持了图片分类、物体检测、图像分割、关键点检测、行人重识别等多种任务
- 该项目在 **github** 上已获得超过 **1.7k** 个 **star**
- 作为工信部“2020 新兴平台软件”项目“人工智能开发框架和开放平台软件”清华单位的核心交付物

清华与快手 Y-tech 合作项目

2020.8-2021.4

- 负责 **CG** 生成数据在深度学习模型训练中的领域自适应项目的所有研发，目前已通过项目结题验收
- 提出基于回归域自适应方法的无监督关键点检测方法，将 **2D** 关键点的 **PCK** 指标从 **62%** 提高到了 **73%**，已提交专利申请，核心技术发表于 **CVPR (CCF A 类)**
- 提出基于注意力机制的 **3D** 关键点检测网络和基于自监督学习的 **3D** 关键点网络训练方法，将相对深度误差从 **0.126** 降低到了 **0.09** (相对提升 **28.5%**)。已提交专利申请

课程项目 | 弹幕 **biubiu** 小程序

2018.9-2019.1

- 本科软件工程课程作业，负责实现网站和小程序的后端逻辑，以及弹幕墙的桌面程序
- 曾被用于清华大学校园各类晚会的现场互动，例如 **2019** 年软件学院学生节
- 团队 (共 **4** 人) 获得腾讯 **5** 万元的资助

专业技能

- 有良好的的英语阅读与写作能力，**GRE: 325/340**，**TOFEL: 106/120**
- 熟练使用 **Python**，有 **Linux** 环境下的开发经验；熟练掌握 **Pytorch**，了解 **TensorFlow**