

计算机视觉（CV）模型偏差消除方法的研究进展

闫懿涛 赵剑波 莫运益 苏若虹^{通讯作者}

（南宁师范大学，广西 南宁 530100）

摘要：近年来，随着计算机视觉（CV）技术的快速发展，语义分割在自动驾驶、安防监控、虚拟试衣等实际场景中展现出广泛应用潜力，特别是人体解析任务在细粒度理解中扮演着关键角色。然而，当前主流预训练模型在处理复杂人体姿态、多样外观特征及遮挡等问题时，仍面临显著的偏差挑战，这些偏差严重影响了模型在公平性、鲁棒性与泛化能力方面的实际表现。传统的人体解析偏差缓解方法主要依赖数据增强、类别重加权、边界优化等手段，虽在一定程度上提升了模型性能，但难以彻底克服因数据不平衡、属性敏感性等因素导致的系统性误差。本文聚焦于语义分割中人体解析任务的偏差消除研究进展，系统梳理了近年来从数据构建、模型设计到训练机制等层面提出的多种缓解策略。文中还详细分析了在遮挡场景、边界处理和小样本泛化等方面的关键技术与挑战，旨在为后续构建更稳健、公平的人体解析模型提供理论参考与方法支持。

关键词：语义分割；人体解析；多尺度模型；CV 模型平衡性

1. 引言

1.1 研究背景与意义

随着深度学习技术的迅速发展，计算机视觉在图像理解任务中取得了广泛突破，尤其是语义分割在自动驾驶、虚拟现实、人体姿态估计等领域发挥着越来越重要的作用^[1]。作为语义分割中的一个核心子任务，人体解析致力于对人体不同部位进行像素级别的精确分割，是实现多种人机交互应用的基础。然而，近年来大量研究发现，受限于训练数据中个体属性分布不均、采集条件差异等因素，模型在不同性别、肤色、年龄或姿态上的识别效果存在显著偏差^[2]。这些偏差不仅削弱了模型的泛化能力和鲁棒性，还可能在特定应用中引发关于公平性与伦理性的担忧。例如，在虚拟试衣或公共安全系统中，对少数群体识别效果较差可能影响使用体验，甚至带来现实社会影响。因此，深入研究和有效缓解人体解析模型中的偏差问题，已成为推动相关技术可信性和实用性的重要课题。

1.2 现有方法与局限性

针对模型偏差问题，研究者提出了包括数据层改进、结构设计优化等多种方法^[3]。在数据层面，通过增强边缘样本、采样均衡或构建更加多样化的训练集以减少数据偏倚；在模型层面，设计关注局部区域或多尺度融合的结构，以提升对复杂人体姿态与遮挡的鲁棒性；在训练策略上，引入加权损失、对抗训练等机制，以降低模型对特定属性的过拟合。然而，这些方法往往依赖于对数据偏差的先验认知，难以适应未知属性分布的变化。同时，结构复杂度的提升也可能导致训练成本的显著增加，降低了方法在实际部署中的可扩展性。此外，在处理多任务融合和跨域泛化问题时，现有技术在保持公平性与性能之间仍存在明显权衡。因此，亟需发展更为稳健且高效的偏差缓解机制，以适应复杂多变的实际场景。

1.3 综述范围与贡献

本文旨在系统回顾当前计算机视觉领域中针对语义分割任务，特别是人体解析任务中模型偏差消除的最新研究进展。首先梳理了人体解析任务中常见的偏差来源，包括训练数据分布不均、遮挡与姿态变化、属性敏感性等因素；重点总结了从数据层、模型结构层以及训练策略层出发的各类消偏方法，包括数据增强与重采样、边界感知结构设计、多尺度融合机制、损失函数优化及正则化策略等，并对其在不同应用场景中的效果进行对比分析。接着，进一步讨论了当前方法在实际应用中所面临的挑战。通过本综述，期望为相关研究人员提供系统性参考，助力偏差消除技术在实际计算机视觉应用中的深入发展与落地。

2. 研究内容

2.1 语义分割

语义分割（Semantic Segmentation）是计算机视觉中的一项基础任务，旨在对图像中的每一个像素进行分类，使其对应于某一特定的语义类别^[4]。与图像分类或目标检测等任务相比，语义分割能够实现更加精细的视觉理解，对于自动驾驶、遥感影像分析、医学图像处理以及虚拟现实等应用场景具有重要意义。近年来，得益于深度卷积神经网络的发展，语义分割在精度和效率方面取得了显著突破，极大地提升了模型在不同尺度、复杂背景和细节结构中的识别能力^[5]。然而，语义分割仍面临诸如边界模糊、小目标识别困难、类间不平衡及跨域泛化能力不足等挑战，尤其在人类解析等细粒度分割任务中，对模型的结构设计与泛化能力提出了更高要求。在图 1 中展示了实例分割和语义分割的精度随着深度卷积神经网络的发展取得的进步。



图 1: 计算机视觉部分任务可视化展示。

(a) 原始图像输入，通常来源于数据集。(b) 经过实例分割产生的结果，识别图像中不同的对象。(c) 经过语义分割产生的结果，为每个像素分配一个类别标签，实现图像中不同对象的区域划分。

2.2 人体解析

人体解析（Human Parsing）是计算机视觉中一种细粒度语义分割任务，其目标是对人体图像进行像素级的语义标注，将人体划分为多个结构化部位，如头发、面部、上衣、裤子、手脚等^[6]。相较于传统语义分割任务，人体解析对分割精度和细节保留提出了更高要求，尤其在处理复杂背景、姿态变换、遮挡、衣物多样性以及小尺度部位时面临极大挑战。

在技术演进方面，人体解析经历了从传统浅层特征方法到深度学习驱动的阶段发展。早期方法多基于人工特征与图模型，准确率有限。随着 FCN 的提出，端到端的卷积网络成为主流，引入多尺度特征融合、空洞卷积与结构先验进一步提升了性能。近

年来,Transformer结构的引入显著增强了模型的全局建模能力,推动了人体解析精度的持续突破。

3. 计算机视觉中偏差消除方法

3.1 模型结构层方法

在模型结构设计层面,为缓解CV模型中的系统性偏差,研究者提出了多种嵌入式结构调整策略,旨在增强模型的公平性和泛化能力,降低对特定属性的敏感性。这些网络包含了多个卷积层和池化层,其中每一层的输出特征图会不断被处理和压缩,从而逐步提取出更为复杂和抽象的特征。在这些特征提取的层之后,网络进入了全连接层,这些全连接层的作用是将高维的特征向量转化为更为紧凑的表示,以便进行最终的分类判断。网络根据这些分类判断来决定最终的分类结果。图中右侧显示的分类结果是图像经过网络处理后,模型对每个像素或区域进行分类的输出,通常会用不同的颜色表示不同的类别,从而实现对图像的语义分割或实例分割。图2展示了深度学习模型分层设计的经典结构。

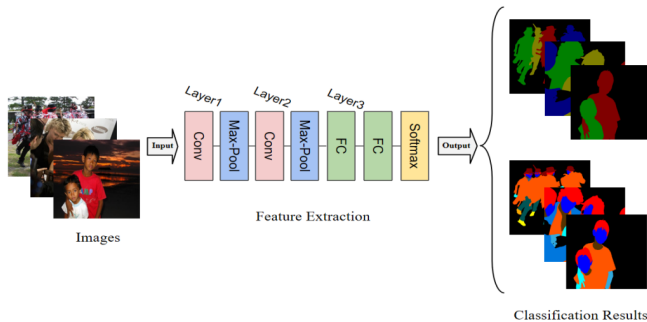


图2: 图像经过特征提取和分类的框架图,是深度学习中用于图像识别和分类的经典结构。

3.2 样本评估与反馈机制方法

有效的偏差检测和评估机制是推进偏差消除研究的基础。当前在CV领域,研究者逐渐意识到传统的全局精度指标(如mIoU、accuracy)无法真实反映模型在不同子群体或边缘样本上的表现差异,因此构建了一套更具解释力的评价体系。将测试集按分组性能(Group-wise Metrics)属性划分子集,分别计算每组的性能指标,进而比较组间差异。例如在人体解析中,研究者常用“少数族群IoU差值”“姿态复杂样本召回率”等来评估模型偏差程度。借鉴公平性指标(Fairness Metrics),NLP与算法公平性研究领域的方法,将其扩展应用于视觉任务。

3.3 国内外相关研究综述

随着计算机视觉技术的广泛应用,模型偏差问题逐渐成为研究热点。国外研究起步较早,重点关注算法公平性与系统性偏差控制,提出了如对抗消偏、公平性损失函数、特征解耦等方法。在语义分割和人体解析领域,学者们针对复杂姿态和遮挡问题提出了结构感知网络、类别敏感注意力机制等,提升了模型在长尾部位和多样人群中的表现,同时结合无监督领域自适应技术增强了跨域泛化能力。

国内研究虽起步稍晚,但近年来在结构创新与任务适配方面取得显著进展。例如,针对人体解析中的属性偏差与结构建模问题,研究者提出了属性解耦网络、图神经网络等方法,有效提升了模型的鲁棒性与公平性。

总体而言,国外更注重构建通用的偏差控制框架和评估体系,国内则强调任务导向与实际应用结合。尽管已有一定成果,偏差量化、交叉属性建模与跨域适应仍是未来需重点攻克的问题。

4. 挑战与未来展望

4.1 当前方法的局限性

4.1.1 数据质量与样本偏差问题

当前偏差消除方法在很大程度上依赖于高质量、覆盖全面的数据集。然而,在实际的人体解析任务中,训练数据常常存在语

义标注不准确、身体部位分割边界模糊、长尾类别样本稀缺等问题。这些问题不仅会限制模型对细粒度结构的学习,还可能导致模型在特定人群或边缘样本上的表现显著下降,进一步加剧性能偏差。

4.1.2 计算复杂度与模型泛化性问题

许多现有的偏差缓解方法通过引入复杂结构,如多尺度融合模块、辅助监督分支或对抗机制来提升模型的公平性与识别能力,但这也带来了额外的计算负担。在处理高分辨率图像或部署在边缘设备上时,这类方法可能难以满足实时性或资源约束的需求。

4.2 未来研究方向

4.2.1 多模态数据中的公平性建模方法

随着多模态视觉系统的快速发展,将图像、文本、深度图、姿态信息等融合用于人体解析已成为趋势。在这一背景下,如何在多模态融合过程中实现偏差控制与公平性建模成为新的研究挑战。未来的研究应聚焦于设计适用于多模态输入的消偏机制,如构建风格无关、属性鲁棒的统一特征空间,避免模型在融合不同模态特征时对某一模态或属性维度产生过度依赖。

4.2.2 偏差消除方法在更多CV任务中的泛化

尽管现有偏差缓解方法在图像分类、目标检测与人体解析等任务中已取得一定成果,但其在更广泛的视觉任务中的应用仍显不足。未来应进一步探索如何将偏差控制机制有效迁移至视频理解、行为识别、三维重建、跨域分割等复杂任务中。在这些任务中,数据分布、时序变化或场景复杂性更高,对模型的公平性提出了更大挑战。

总结:

随着计算机视觉技术在现实世界中的广泛部署,模型偏差问题正日益成为影响其公平性与可靠性的重要因素。本文聚焦于语义分割中的人体解析任务,系统梳理了当前模型偏差的主要表现形式及其潜在成因,综述了近年来在模型结构设计、评估机制构建和跨域自适应方法等方面的偏差消除技术进展。尽管现有方法在提升模型泛化性和公平性方面取得了一定成效,但仍面临数据质量受限、计算开销较大、跨模态信息融合困难等挑战。未来的研究应更加重视多模态融合背景下的公平建模、跨任务偏差控制策略的泛化能力,以及在实际部署中的效率与可解释性。我们期望本文的综述能为相关研究提供理论参考和实践指导,推动构建更加稳健、公平的人体解析系统在现实场景中的广泛应用。

参考文献:

- [1] 刘晨,张敏,王丽.计算机视觉中的图像分类算法及其应用研究[J].计算机工程与应用,2019,55(12):78-82.
- [2] 王伟,李强,张涛.基于深度学习的计算机视觉目标检测方法研究[J].计算机应用研究,2020,37(5):1234-1238.
- [3] 李明,张磊,王强.基于模型压缩的计算机视觉模型优化方法研究[J].计算机科学,2020,47(10):98-104.
- [4] 王浩,李敏,张云.基于深度学习的语义分割算法研究与应用[J].计算机应用研究,2020,37(12):3456-3460.
- [5] 张宇,李伟.(2020).计算机视觉的深度学习综述. IEEE 图像处理汇刊,29(5),1201-1215.
- [6] 孙晓,尚杰,张磊.(2020).基于深度学习的人体姿态估计综述. IEEE 模式分析与机器智能汇刊,42(5),1095-1110.

作者简介:

闫懿涛(2000—),男,硕士,主要从事计算机视觉(CV)、语义分割。赵剑波(1997—),男,硕士,主要从事自然语言处理(NLP)、因果推理。

莫运益(2000—),男,硕士,主要从事遥感图像语义分割。

通信作者:

苏若虹(2001—),女,本科,主要从事教育学。