

FaSRnet: 用于人体姿态估计的特征和语义修正网络

仲元红¹, 徐乾锋¹, 钟代笛², 杨勋³, 王姗姗⁴

¹重庆大学微电子与通信工程学院, 中国重庆市, 400044

²重庆大学生物工程学院, 中国重庆市, 400044

³中国科学技术大学信息科学与技术学院, 中国合肥市, 230039

⁴安徽大学物质科学与信息技术研究院, 中国合肥市, 230039

摘要: 由于运动模糊、视频失焦和遮挡等因素, 多帧人体姿态估计是一项有挑战性的任务。利用连续帧之间的时间一致性是解决这一问题的有效方法。目前, 大多数方法通过修正最终热图来利用时间一致性。热图包含了关键点的语义信息, 可在一定程度上提高检测质量。它们由特征生成, 但这些方法很少考虑特征级别的修正。本文提出一种人体姿态估计框架, 该框架在特征和语义层面进行了改进。将辅助特征与当前帧的特征对齐, 以减少不同特征分布带来的损失。使用注意力机制将辅助特征与当前特征融合。在语义方面, 使用相邻热图之间的差异作为辅助特征来修正当前热图。在大型基准数据集PoseTrack2017和PoseTrack2018上验证了该方法的有效性。

关键词: 人体姿态估计; 多帧修正; 热图和偏移估计; 特征对齐; 多人

<https://doi.org/10.1631/FITEE.2200639>