
视觉人体姿态估计研究取得进展

作者：writer 来源：中国科学院

本文原地址：<https://www.iikx.com/news/progress/36641.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

视觉人体姿态估计研究取得进展。

2D人体姿态估计是一项面向人体对象的基础视觉感知任务，旨在通过图像或视频精准定位人体姿态关键点位置，为理解人体运动和行为提供结构化数据。然而，现有方法在复杂场景下难以兼顾空间定位准确性与时间运动连贯性，尤其在视频中如何保持姿态估计的稳定性是一大挑战。同时，

在获得2D人体关键点基础上，3D人体姿态估计致力于恢复人体关键点在三维空间中的位置，这一技术在虚拟现实、运动分析、人机交互等领域具有广泛应用。然而，这一过程面临“深度模糊”等固有挑战，即同一组2D关键点或对应多个合理的3D姿态。目前，基于扩散生成式框架的方法一定程度上缓解了这一问题，但由于模型缺乏对人体外观轮廓、语义理解等方面的认识，导致在预测中生成不合理的姿态，这一问题在遮挡严重或人群密集等复杂场景下较为突出。

近日，中国科学院上海微系统与信息技术研究所研究团队，在视觉人体姿态估计研究中取得进展。团队围绕2D及3D人体姿态估计两项任务，分别提出了创新方法。相关研究成果发表在IEEE Transactions on Circuits and Systems for Video Technology和IEEE Transactions on Multimedia上。

团队提出了高效的层次化对比一致性约束（HICCON），其可灵活嵌入多种姿态估计模型中，提升2D姿态估计性能。该方法在空间域上

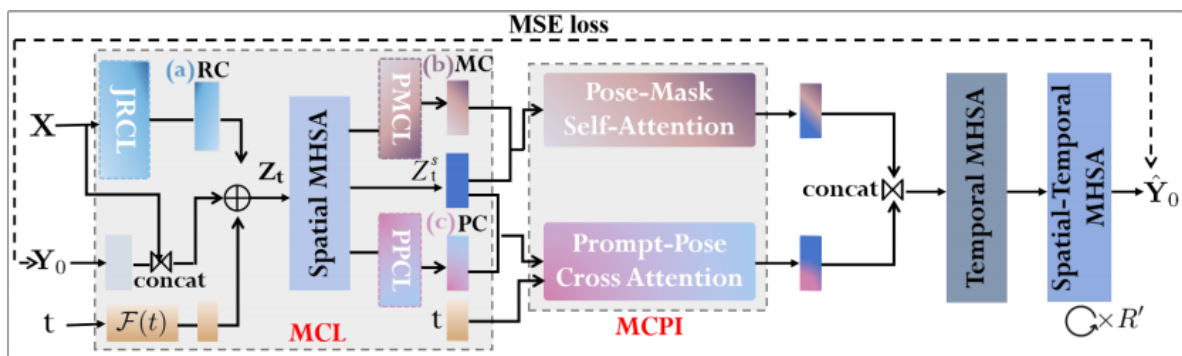
捕捉建模关键点与身体部件之间的关系，在时间域上捕捉帧与片段之间的运动模式。同时，HICCON可分别从空间与时间两个维度，提取

关键点点级、部件关系级、实例级及片段级等不同粒度多层次特征，并施加对比学习约束，增强模型对复杂姿态的判别能力。实验表明，在主流视频姿态数据集PoseTrack上，结合HICCON的多个模型均超过基准性能，且该方法表明了对比学习机制在人体视觉表征建模中的应用潜力和重要作用。

团队进一步提出了MMCPose模型，将多模态人体先验作为条件信号，引导扩散过程生成合理且准确的3D姿态。该模型融合了

人体关节拓扑关系、基于自然语言的部件描述、提升姿态关注度的人体掩码三类结构化信息。为更好地发挥多模态条件的引导作用，团队设计了多模态表征—姿态交互机制，实现引导信号与生成过程之间的深度交互，从而提升模型在姿态建模上的感知能力与生成质量。在Human3.6M和MPI-INF-3DHP等基准集上的测试表明，MMCPose取得了领先性能，特别是在Human3.6M上，将平均误差降至30.8毫米。该方法说明，多模态引导及人体先验知识对解决三维人体视觉感知任务具有关键作用。

研究工作得到科技创新2030重大项目、上海市自然科学基金等的支持。



在Human3.6M数据集上，MMCPose达到了最优性能

研究团队单位：上海微系统与信息技术研究所

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发