

视频行为识别中的时序保真度增强

许少武¹, 贾熹滨¹, 孙倩美², 常晶²

¹北京工业大学信息学部, 中国北京市, 100124

²首都医科大学附属北京朝阳医院, 中国北京市, 100020

摘要: 时序注意力机制对于视频行为识别至关重要, 它使模型能够聚焦于具有丰富语义信息的关键片段。然而, 这些模型常因训练多样性有限和缺乏细粒度时序监督而出现时序失真现象——即注意力权重与语义内容错位。尽管视频级标签提供了粗粒度的行为指引, 但细节约束的缺失导致注意力噪声持续存在, 尤其在包含干扰性空间元素的复杂场景中。针对这一问题, 本文提出时序保真度增强 (TFE) ——一种基于解耦信息瓶颈 (DisenIB) 理论的对抗性学习范式。TFE通过对抗性特征解耦将行为相关语义与虚假相关性分离, 从而缓解时序失真问题。该方法利用预训练表征进行初始化, 建立对抗学习流程, 即高时序注意力片段与行为相关性弱化的上下文相互竞争。该方法无需细粒度监督标签即可确保时序一致性, 并提升注意力权重的保真度。在UCF101、HMDB-51和Charades基准数据集上的大量实验验证了该方法的有效性, 结果表明TFE可令行为识别准确率显著提升。

关键词: 行为识别; 解耦信息瓶颈; 时序建模; 时序保真度

<https://doi.org/10.1631/FITEE.2500164>