MLKD 2024
The First International Conference on
Machine Learning and Knowledge Discovery
Amirkabir University of Technology, December 18-19, 2024

Enhancing Temporal Action Localization with 3D Input and Attention Mechanisms

Mozhgan Mokari[*] Khosrow Haj Sadeghi[†]

Abstract

Temporal action localization (TAL) in untrimmed videos poses a significant challenge due to the accurate determination of action timing and type within noisy or irrelevant content. In this paper, we introduce SeqAttNet, an innovative end-to-end network that aims to advance TAL performance through novel approaches. Our model combines attention mechanisms with a compact two-dimensional sequential network and utilizes 3D input aggregation to optimize accuracy and computational efficiency. SeqAttNet outperforms existing methods, achieving over 87% greater efficiency on the ActivityNet dataset while being over 70 times smaller compared to baseline. Despite its compact nature, SeqAttNet maintains competitive accuracy, surpassing larger models such as TriDet in overall efficiency and achieving more than twice the efficiency on the ActivityNet dataset. Our findings demonstrate that SeqAttNet effectively balances performance with computational cost, delivering high accuracy while notably reducing network complexity. This makes it a valuable tool for practical TAL applications, where both precision and efficiency are essential.

Keywords: Temporal action localization, efficient attention, and Action recognition

[*] Department of Electrical Engineering, Sharif University of Technology, `mozhgan.mokari@ee.sharif.edu`

[†] Department of Electrical Engineering, Sharif University of Technology, `ksadeghi@sharif.edu`