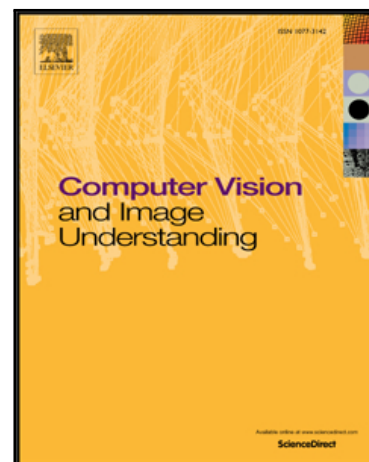


Accepted Manuscript

VideoLSTM Convolves, Attends and Flows for Action Recognition

Zhenyang Li, Kirill Gavriluk, Efstratios Gavves, Mihir Jain,
Cees G.M. Snoek

PII: S1077-3142(17)30174-1
DOI: [10.1016/j.cviu.2017.10.011](https://doi.org/10.1016/j.cviu.2017.10.011)
Reference: YCVIU 2629



To appear in: *Computer Vision and Image Understanding*

Received date: 12 May 2017
Revised date: 4 October 2017
Accepted date: 18 October 2017

Please cite this article as: Zhenyang Li, Kirill Gavriluk, Efstratios Gavves, Mihir Jain, Cees G.M. Snoek, VideoLSTM Convolves, Attends and Flows for Action Recognition, *Computer Vision and Image Understanding* (2017), doi: [10.1016/j.cviu.2017.10.011](https://doi.org/10.1016/j.cviu.2017.10.011)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Highlights

- To exploit both the spatial and temporal correlations in a video, we hardwire convolutions in the soft-Attention LSTM architecture.
- We introduce motion-based attention which guides better the attention towards the relevant spatial-temporal locations of the actions.
- We demonstrate how the attention generated from our VideoLSTM can be used for action localization by relying on the action class label only.
- We show the theoretical as well as practical merits of our VideoLSTM against other LSTM architectures for action classification and localization.

Download English Version:

<https://daneshyari.com/en/article/6937446>

Download Persian Version:

<https://daneshyari.com/article/6937446>

[Daneshyari.com](https://daneshyari.com)