



Center for Research in Computer Vision

UNIVERSITY OF CENTRAL FLORIDA

FINAL ORAL EXAMINATION

OF

Rui Hou

M.E., Arizona State University, 2012

B.S., Beijing University of Posts and Telecommunications, 2010

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

(COMPUTER SCIENCE)

26 June, 2019, 11:00 A.M.

HEC 103

DISSERTATION COMMITTEE

Professor Mubarak Shah, *Chair*, shah@crcv.ucf.edu

Professor Rahul Sukthankar, rahulsukthankar@gmail.com

Professor Kien Hua, kienhua@cs.ucf.edu

Professor Abhijit Mahalanobis, amahalan@cs.ucf.edu

DISSERTATION RESEARCH IMPACT

Automatic understanding of videos is one of the most active areas of computer vision research. It has applications in video surveillance, human computer interaction, video sports analysis, virtual and augmented reality, video retrieval etc. In this dissertation, we address four important tasks in video understanding, namely action recognition, temporal action localization, spatial-temporal action detection and video object/action segmentation.

This dissertation makes contributions to above tasks by proposing:

(1) A category level feature learning method for action recognition in well trimmed short videos. (2) A sub-action discovery approach for localizing actions in long untrimmed video, by automatically determining the number of sub-actions from the training set. (3) A 3D Tube Convolutional Neural Network (TCNN) based pipeline for action detection. It generalizes the popular faster R-CNN framework from images to videos. (4) An end-to-end encoder-decoder based 3D convolutional neural network pipeline, which is able to segment out the foreground objects from the background. Moreover, the action label can be obtained as well by passing the foreground object into an action classifier.

SELECTED PUBLICATIONS (Total Citations: 168)

1. **DaMN (Discriminative and Mutually Nearest): Exploiting Pairwise Category Proximity for Video Action Recognition**, Rui Hou, Amir R. Zamir, Rahul Sukthankar and Mubarak Shah, in *European Conference on Computer Vision (ECCV)*, 2014
2. **Real-Time Temporal Action Localization in Untrimmed Videos by Sub-Action Discovery**, Rui Hou, Rahul Sukthankar and Mubarak Shah, in *British Machine Vision Conference (BMVC)*, 2017.
3. **Tube Convolutional Neural Network (T-CNN) for Action Detection in Videos**, Rui Hou, Chen Chen and Mubarak Shah, in *IEEE International Conference on Computer Vision (ICCV)*, 2017.
4. **Efficient 3D CNN for Action/Object Segmentation in Video**, Rui Hou, Chen Chen, Rahul Sukthankar and Mubarak Shah, *under submission to British Machine Vision Conference (BMVC)*, 2019 (submitted).

DISSERTATION

ACTION RECOGNITION, TEMPORAL LOCALIZATION AND DETECTION IN TRIMMED AND UNTRIMMED VIDEO

This dissertation addresses the problem of action understanding in videos, which includes action recognition in trimmed video, temporal action localization in untrimmed video, spatial-temporal action detection and video object/action segmentation.

For action recognition, we propose a method for learning discriminative category-level features. The key observation is that one-vs-rest classifiers, which are ubiquitously employed for this task, face challenges in separating very similar categories (such as running vs. jogging). Our proposed method automatically identifies such pairs of categories using a criterion of mutual pairwise proximity in the (kernelized) feature space, and a category-level similarity matrix where each entry corresponds to the one-vs-one SVM margin for pairs of categories. We augment one-vs-rest classifiers with a judicious selection of “two-vs-rest” classifier outputs, formed from such discriminative and mutually nearest DaMN pairs.

Above video action recognition is designed for clip level classification in manually trimmed datasets. However, in the real world, the majority of videos are untrimmed, where an action of interest may occur only in a small part of a long video. Next, for temporal action localization, we propose to exploit the temporal structure of actions by modeling an action as a sequence of sub-actions and present a computationally efficient approach. A novel sub-action discovery algorithm is proposed, where the number of sub-actions for each action as well as their types are automatically determined from the training videos. To localize an action, an objective function combining appearance, duration and temporal structure of sub-actions is optimized as a shortest path problem in a network flow formulation.

In above method, the actions are only localized temporally. However, the goal of action detection is to detect every occurrence of a given action within a long video and to localize each detection both in space and time. Towards that end we propose an end-to-end deep network called Tube Convolutional Neural Network (T-CNN) for action detection in videos. The proposed architecture is a unified deep network that is able to recognize and localize action based on 3D convolution features. A video is first divided into equal length clips and next for each clip a set of tube proposals are generated based on 3D Convolutional Network (ConvNet) features. Finally, the tube proposals of different clips are linked together and spatio-temporal action detection is performed using these linked video proposals.

Localizing actions in videos at finer spatio-temporal level is continuously desired by computer vision researchers and developers. Finally, to pixel-wise localize actions, we propose an end-to-end encoder-decoder style 3D CNN for video object segmentation. The proposed approach leverages 3D separable convolutions and 3dpyramid pooling, which drastically reduces the number of trainable parameters. Additionally, the framework is extended to solve the video action segmentation problem by adding an extra classifier to predict the action label for actors in videos.



RUI HOU

1987	Born in Tonghua, China
2010	B.S., Beijing Univ. of Posts and Telecom., Beijing, China
2012	M.E., Arizona State University, Tempe, AZ
2014 & 2015	Software Engineer Intern, Google, Mountain View, CA
2012-2019	Ph.D., University of Central Florida, Orlando, FL
2018-	Software Engineer, Niantic Labs, Sunnyvale, CA