1. Zheng Shou, Dongang Wang, Shih-Fu Chang:Temporal Action Localization in Untrimmed Videos via Multi-stage CNNs. CVPR 2016: 1049-1058
2. Kensho Hara, Hirokatsu Kataoka, Yutaka Satoh:Can Spatiotemporal 3D CNNs Retrace the History of 2D CNNs and ImageNet? CoRR abs/1711.09577 (2017)
3. Colin Lea, Michael D. Flynn, René Vidal, Austin Reiter, Gregory D. Hager:Temporal Convolutional Networks for Action Segmentation and Detection. CVPR 2017: 1003-1012
4. Du Tran, Lubomir D. Bourdev, Rob Fergus, Lorenzo Torresani, Manohar Paluri:Learning Spatiotemporal Features with 3D Convolutional Networks. ICCV 2015: 4489-4497
5. Zheng Shou, Jonathan Chan, Alireza Zareian, Kazuyuki Miyazawa, Shih-Fu Chang:CDC: Convolutional-De-Convolutional Networks for Precise Temporal Action Localization in Untrimmed Videos. CVPR 2017: 1417-1426
6. Kai Kang, Hongsheng Li, Tong Xiao, Wanli Ouyang, Junjie Yan, Xihui Liu, Xiaogang Wang:Object Detection in Videos with Tubelet Proposal Networks. CVPR 2017: 889-897
7. Yue Zhao, Yuanjun Xiong, Limin Wang, Zhirong Wu, Xiaoou Tang, Dahua Lin:Temporal Action Detection with Structured Segment Networks. ICCV 2017: 2933-2942
8. Xizhou Zhu, Yujie Wang, Jifeng Dai, Lu Yuan, Yichen Wei:Flow-Guided Feature Aggregation for Video Object Detection. ICCV 2017: 408-417
9. Yuan, Xiaodan Liang, Xiaolong Wang, Dit-Yan Yeung, Abhinav Gupta:Temporal Dynamic Graph LSTM for Action-Driven Video Object Detection. ICCV 2017: 1819-1828
10. Xizhou Zhu, Yuwen Xiong, Jifeng Dai, Lu Yuan, Yichen Wei:Deep Feature Flow for Video Recognition. CVPR 2017: 4141-4150
11. Khurram Soomro, Mubarak Shah:Unsupervised Action Discovery and Localization in Videos. ICCV 2017: 696-705
12. Limin Wang, Yuanjun Xiong, Zhe Wang, Yu Qiao, Dahua Lin, Xiaoou Tang, Luc Van Gool:Temporal Segment Networks: Towards Good Practices for Deep Action Recognition. ECCV (8) 2016: 20-36
13. Ke Yang, Peng Qiao, Dongsheng Li, Shaohe Lv, Yong Dou:Exploring Temporal Preservation Networks for Precise Temporal Action Localization. AAAI 2018
14. Victor Escorcia, Fabian Caba Heilbron, Juan Carlos Niebles, Bernard Ghanem:DAPs: Deep Action Proposals for Action Understanding. ECCV (3) 2016: 768-784
15. Rui Hou, Chen, Mubarak Shah:Tube Convolutional Neural Network (T-CNN) for Action Detection in Videos. ICCV 2017: 5823-5832
16. Yicong Tian, Rahul Sukthankar, Mubarak Shah:Spatiotemporal Deformable Part Models for Action Detection. CVPR 2013: 2642-2649
17. Xiaojiang Peng, Cordelia Schmid:Multi-region Two-Stream R-CNN for Action Detection. ECCV (4) 2016: 744-759
18. Serena Yeung, Olga Russakovsky, Greg Mori, Li Fei-Fei:End-to-End Learning of Action Detection from Frame Glimpses in Videos. CVPR 2016: 2678-2687
19. Shugao Ma, Jianming Zhang, Nazli Ikizler-Cinbis, Stan Sclaroff:Action Recognition and Localization by Hierarchical Space-Time Segments. ICCV 2013: 2744-2751
20. Dan Oneata, Jakob J. Verbeek, Cordelia Schmid:Efficient Action Localization with Approximately Normalized Fisher Vectors. CVPR 2014: 2545-2552

[1]提出了一种用于时序动作定位的多阶段3D卷积网络，包括多尺度视频片段的生成、多阶段segment-CNN以及后处理**（temporal action localization）**。[2]观察到深度2D卷积神经网络经过ImageNet数据集预训练后迁移到detection，segmentation，captioning等其他任务上取得的显著成就后，尝试在Kinetics数据集上对深度3D卷积神经网络预训练，观察迁移到detection、summarization以及Optical flow的提取等任务上的效果**（action recognition）**。[3]提出了一个基于时序卷积的卷积-逆卷积网络，来实现对视频帧的类别分类**（action segmentation and detection）**。[4]提出了一种用于提取视频时空特征的3D卷积神经网络**（action recognition）**。[5]基于C3D提出了一种卷积-逆卷积的神经网络，输入一小段视频，输出frame-level的动作类别概率和segment-level的时序动作定位**（temporal action localization）**。[6]将传统针对静止图像的FasterRCNN框架进行了扩展，将视频物体检测框架扩展为候选时空管道生成和候选时空管道识别两个模块，基于这些高质量的候选时空管道应用编码-解码LSTM网络进行时空管道的识别能够有效的提升检测整体的正确率**（video object detection）**。[7]提出了一种结构化分段的时序金字塔网络，既产生高质量的动作提议又提升了对动作的准确识别率和定位**（temporal action localization）**。[8]提出使用时间域的聚合提升单帧特征学习，使得对视频中每一帧的物体识别率得到提高**（video object detection）**。[9]提出了一种以动作描述驱动的时序动态图网络**（video object detection）**。[10]利用视频中连续帧内容高度相关这一特点只对关键帧计算CNN特征提取，然后通过一个flow field将关键帧的CNN特征propagte到其他帧去，减少了计算量**（video object detection）**。[11]提出了一种对视频动作进行聚类、标记以及定位的无监督方法**（temporal action localization）**。[12]提出一种时序分段网络，并基于two-stream法进行改进**（action recognition）**。[13]基于[5]提出TPN（temporal preservation networks）模型，将temporal convolution替换为temporal preservation convolution能够在不进行时序池化操作的情况下获得同样大小的感受野而不缩短时序长度，从而更好地保留时序信息**（temporal action localization）**。[14]提出了一种产生动作提议的高效算法**（temporal action localization）**。[15]提出一种管道卷积神经网络，先将一个视频分成clips，分别从clips中产生候选动作管道，然后这些clips中的候选管道进行合理地衔接，最后对这些衔接好的候选管道进行动作检测**（temporal action detection）**。[16]提出了一种时空可变部分模型（SDPM），对每类动作学习出一种时空模式，也就是将这种时空模式与目标动作进行尽可能地匹配**（temporal action localization）**。[17]提出一种多区域的two-stream Faster R-CNN网络，在appearance R-CNN和motion R-CNN中选择多个身体区域（即上半身、下半身和边界区域）来提高帧级动作检测的性能**（temporal action localization）**。[18]提出了一种通过预测待处理的下一帧位置来逐步完成对视频的动作检测**（temporal action localization）**。[19]提出一种对视频帧的层级式的分割表示方法，第一层表示出该视频帧的最基本的分割（root），第二层则更细粒度地将root进行分割表示**（temporal action localization）**。[20]基于对图片的费舍向量表示提出了对视频的近似的正则化费舍向量表示方法**（temporal action localization）**。