

## [论文概览] AAAI 2018 行为识别论文概览



流浪者

计算机视觉爱好者，微信公众号“faiculy”

已关注

Evan 等 28 人赞了该文章

<个人网页blog已经上线，一大波干货即将来袭：[faiculy.com/](http://faiculy.com/)>

/\* 版权声明：公开学习资源，只供线上学习，不可转载，如需转载请联系本人。\*/

QQ交流群：451429116

### Action Detection

[1] ++Action Recognition from Skeleton Data via Analogical Generalization over Qualitative Representations Kezhen Chen\*, Kenneth Forbus++

- **思路：**从骨架图中学习人的行为

[2] ++Action Recognition with Coarse-to-Fine Deep Feature Integration and Asynchronous Fusion Weiyao Lin\*, Yang Mi, Jianxin Wu, Ke Lu, Hongkai Xiong++

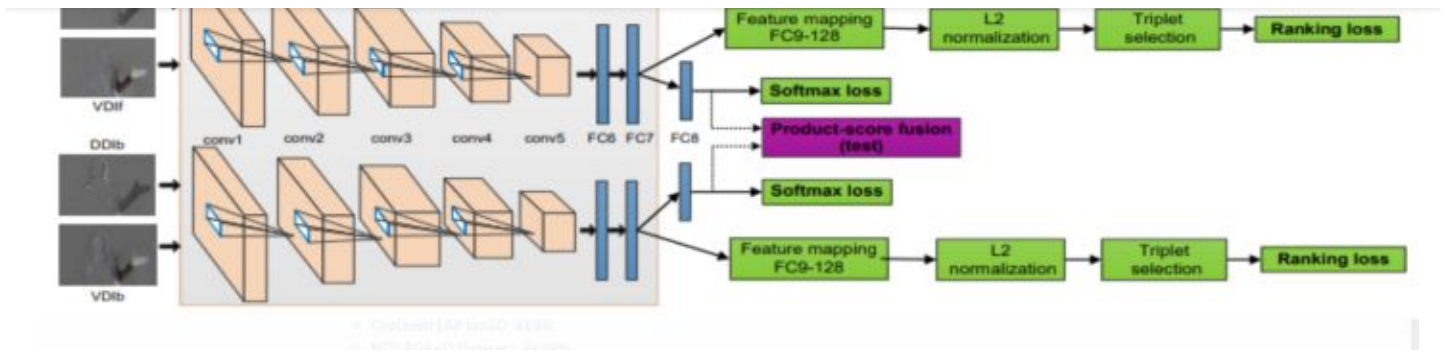
- **提高精度的方法：**
  - 生成更加具有针对性的动作特征，来更好的代表某个动作
  - 减少不同信息流的异步性
- **思路：**
  - 由粗到细的网络提取共享的深层特征，然后逐步融合获得更好的表征特征
  - 异步融合网络，在不同时间融合来自不同流的信息
- **结果：**
  - 无IDT的。UCF101上是94.3%，HMDB51是69.0%
  - 有IDT的。UCF101上是95.2%，HMDB51上是72.6%

[3] ++Cooperative Training of Deep Aggregation Networks for RGB-D Action Recognition Pichao Wang\*, Wanqing Li, Jun Wan, Philip Ogunbona, Xinwang Liu++

- **网络结构：**该篇文章针对的是RGB-D图像，所以并没有细看~



知乎

首发于  
计算机视觉

### 思路：

- 在RGB-D的视觉特征和深度特征上训练c-ConvNet卷积网络
- 通过联合ranking loss和softmax loss能增强深度可分离特征的学习，也就是可以学到更加具有区分性的深度特征

### 实验结果

- ChaLearn LAP IsoGD: 44.8%
- NTU RGB+D Dataset: 89.08%
- SYSU 3D HOI dataset: 98.33%

[4] ++Hierarchical Nonlinear Orthogonal Adaptive-Subspace Self-Organizing Map based Feature Extraction for Human Action Recognition Yang Du, Chunfeng Yuan\*, Weiming Hu, Hao Yang++

- 简介：**这篇文章是中科院自动化所提出来的，一种针对行为识别的特征生成的方法。传统的手写特征要求规则苛刻，而深度学习提取特征的方法需要大量的标记数据。文章提出的 Nonlinear Orthogonal Adaptive-Subspace Self-Organizing Map(NOASSOM)是一种折中的考虑。

### 思路：论文的主要贡献点

- 添加一个非线性正交图层使得NOASSOM能处理非线性的数据，而且通过核技巧可以避免定义具体非线性正交图。
- 修改ASSOM的损失函数，使得每个输入样本都被用来单独的训练模型
- 提出一个层次化的NOASSOM，能提取更具代表性（区分性，独特性）的特征

### 实验结果：

- HMDB-51上：NOASSOM+iDT, 69.3%
- UCF-101上：NOASSOM+iDT, 93.8%
- KTH上：NOASSOM+FV, 98.2%

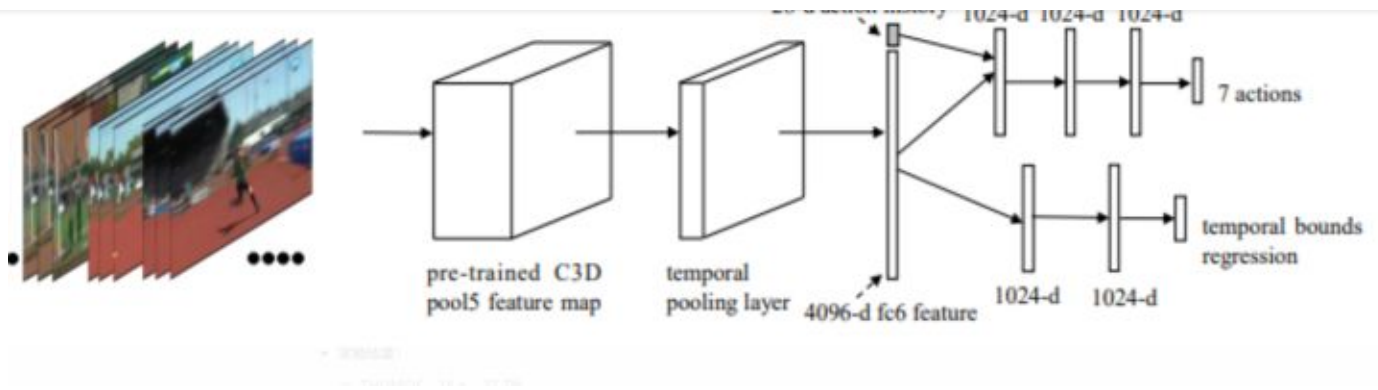
[5] ++SAP: Self-Adaptive Proposal Model for Temporal Action Detection based on Reinforcement Learning Jinjia Huang, Nannan Li, Ge Li\*, Ronggang Wang, Wenmin Wang++

- 简介：**北京大学深圳研究生院，行为检测文章。作者认为从人类认知来看，行为检测应该是分为两个部分，第一部分是粗定位，第二部分是精修。所以作者提出SAP，自适应的行为检测方法

### 网络结构



知乎

首发于  
计算机视觉

- **思路：**先遍历整个视频，发现一些行为记录（label），来学习一个代理。利用强化学习，特别是Deep Q-Learning 算法来学习代理的决策策略。
- **实验结果：**
  - THUMOS '14上, 27.7%
- **开源代码：**[github.com/hjjpku/Action](https://github.com/hjjpku/Action)

[6] ++Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition Sijie YAN\*, Yuanjun XIONG, Dahua LIN, xiaou Tang++

- **简介：**港中文汤晓鸥实验室，从论文题目可以知道，这篇文章设计一种基于骨架图做行为识别的空间时间卷积网络。传统的方法是通过手工制作或者遍历规则来建模骨架，这样得到的结果不仅代表性有限，而且泛化能力比较差。作者提出的ST-GCN能自动从数据中学习时间和空间的模型。
- **网络结构**

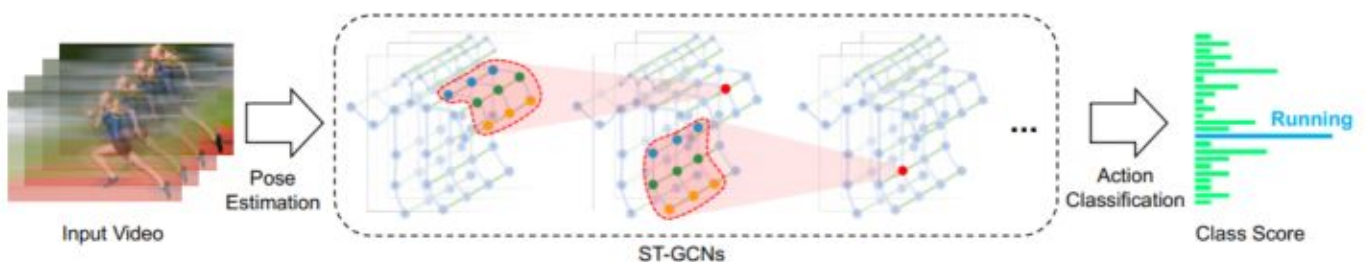


Figure 2: We perform pose estimation on videos and construct spatial temporal graph on skeleton sequences. Multiple layers of spatial-temporal graph convolution (ST-GCN) will be applied and gradually generate higher-level feature maps on the graph. It will then be classified by the standard Softmax classifier to the corresponding action category.

- **思路：**
  - 在视频上先对每一帧做姿态估计（Kinetics 数据集上文章用的是OpenPose），然后可以构建出一个空间上的骨架时序图。
  - 然后应用ST-GCN网络提取高层特征
  - 最后用softmax分类器进行分类
- **实验结果：**
  - Kinetics dataset: 30.7%
  - NTU-RGB+D : 在cross-subject(X-Sub)和cross-View(X-View)上表现是81.5%, 88.3%

[7] ++Spatio-Temporal Graph Convolution for Skeleton Based Action Recognition

Chaolong Li\*, Zhen Cui, Wenming Zheng, Chunyan Xu, Jian Yang++

[8] ++T-C3D: Temporal Convolutional 3D Network for Real-time Action Recognition LIU

KUN, Wu Liu\*, Chuang Gan, Mingkui Tan, Huadong Ma++

[9] ++Unsupervised Deep Learning of Mid-Level Video Representation for Action

Recognition Jingyi Hou\*, Xinxiao Wu, Jin Chen, Jiebo Luo, yunde Jia++

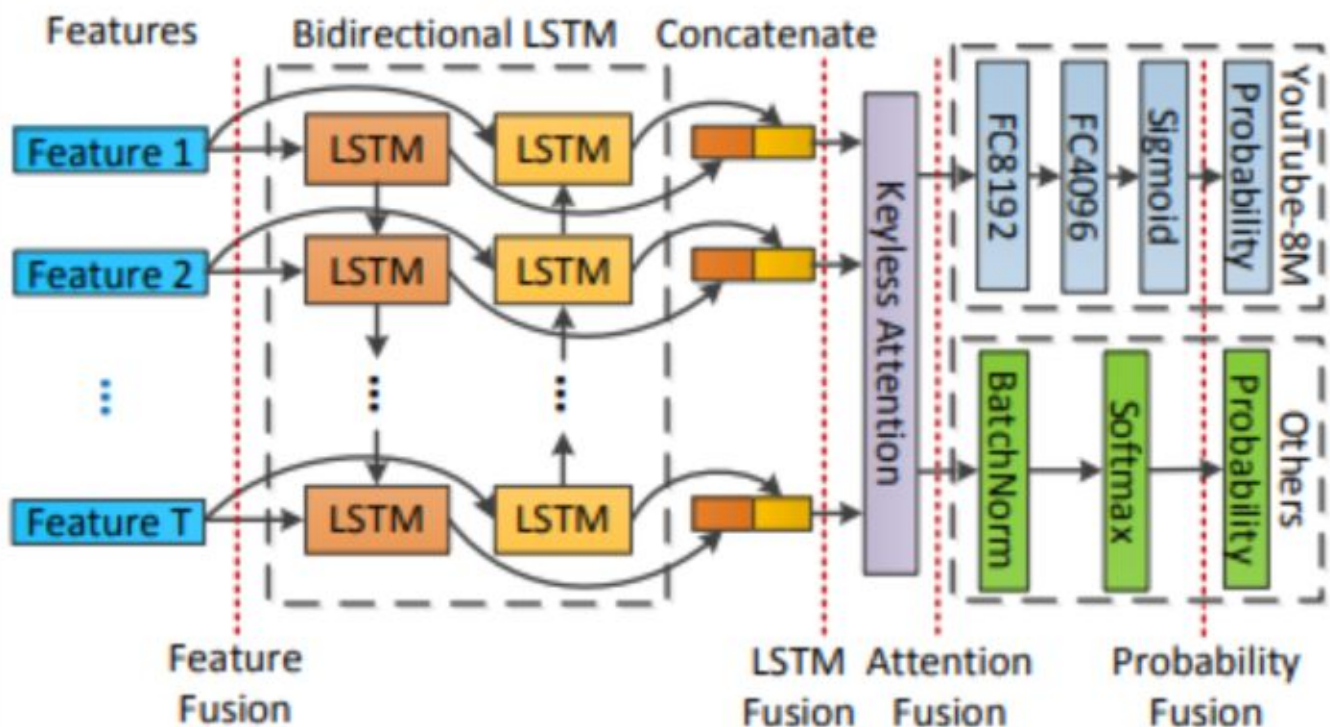
[10] ++Unsupervised Representation Learning with Long-Term Dynamics for Skeleton

Based Action Recognition Nenggan Zheng, Jun Wen, Risheng Liu\*, liangqu Long, Jianhua Dai, Zhefeng Gong++

[11] ++Multimodal Keyless Attention Fusion for Video Classification Xiang Long\*, Chuang

Gan, Gerard De melo, Xiao Liu, Yandong Li, Fu Li, Shilei Wen++

- **简介：**清华大学论文，根据题目，Multimodal Keyless 可以知道，这篇文章采用了多模态的方式。而且走的是RNN (LSTM) 的路线。
- **思路：**Multimodal Representation意思是多模式表示，在行为识别任务上，文章采用了视觉特征 (Visual Features, 包含RGB特征 和 flow features) ；声学特征 (Acoustic Feature) ；前面两个特征都是针对时序，但是时序太长并不适合直接喂到LSTM，所以作者采用了分割的方法 (Segment-Level Features) ，将得到的等长的Segment喂到LSTM。
- **网络结构**



#### • 实验结果：

- **特点：**该文章实验在多个数据集上，文章称鲁棒性比较好。



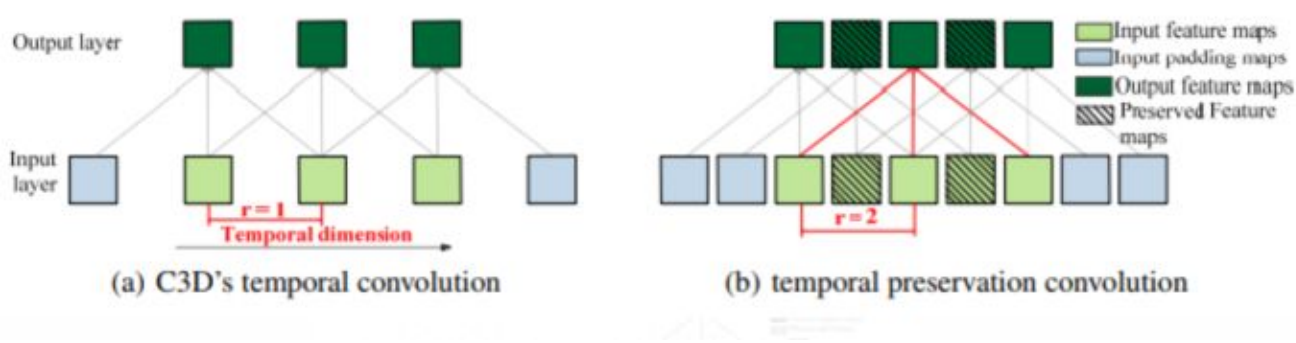


- Kinetics上, Top-1:77.0%, Top-5: 93.2%
- YouTube-8M GAP@20,60K Valid: 80.9%, Test: 82.2%

Action Localization

[12] ++Exploring Temporal Preservation Networks for Precise Temporal Action Localization Ke Yang\*, Peng Qiao, Dongsheng Li, Shaohe Lv, Yong Dou++

- **简介:** 这篇文章是 杨科大佬的文章。Temporal Preservation Network, TPC, 时序保留网络。
- **思路:** 这篇文章是在CDC网络的基础进行改进的，CDC最后是采用了时间上上采样，空间下采样的方法做到了 per-frame action predictions，而且取得了可信的行为定位的结果。但是在 CDC filter之前时间上的下采样存在一定时序信息的丢失。作者提出的TPC网络，采用时序保留卷积操作，这样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。
- **TPC 时序保留卷积:**



- **实验结果:** THUMOS'14上, 28.2%

编辑于 2018-03-07

计算机视觉

已赞同 28

▼

● 添加评论

➤ 分享

★ 收藏

...

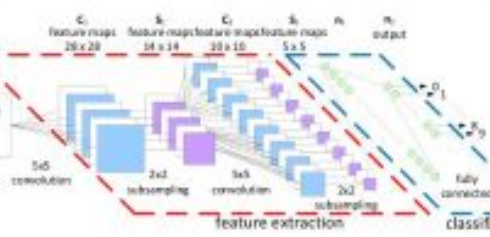
文章被以下专栏收录

FAICULTY

计算机视觉  
计算机视觉、机器学习算法学习

已关注





feature maps: 28 x 28, 24 x 24, 20 x 16, 16 x 16, 12 x 12, 10 x 10

5x5 convolution, 2x2 subsampling, 3x3 convolution, 2x2 subsampling, fully connected, classification

论文笔记——基于深度学习的  
视频行为识别/动作识别（一）

星号then



[CVPR 2018论文笔记] 光  
行为识别的结合研究

林天威

还没有评论

写下你的评论...

