## RELATED WORK

## Video object detection

1. Kai Kang, Wanli Ouyang, Hongsheng Li, and Xiaogang Wang. 2016. Object detection from video tubelets with convolutional neural networks. In IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 817–825.

通过联合针对静态图像的物体检测方法和一般物体跟踪方法，从给定的视频中产生 许多小段的物体跟踪提议，来预测小段中含有物体的概率。

1. Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You only look once: Unified, real-time object detection. In IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 779–788.

将图像网格化，对每一个网格进行物体的·bounding box和类别的预测，最后进行 NMS处理，建立了一个端到端的模型。

1. Wenhan Luo, Junliang Xing, Xiaoqin Zhang, Xiaowei Zhao, and Tae-Kyun . 2014. Multiple object tracking: A literature review. *arXiv:1409.7618* (2014).

本文描述了视频中多目标跟踪（MOT）任务的相关方法和问题，提出了一个统一问题公式和一些现有方法的分类方式，介绍了state-of-the-art MOT算法的关键因素，并讨论了MOT算法的测评包括评价指标、公开数据集，开源代码的实现，和基准测试结果。

1. Long Ying, Tianzhu Zhang, and Changsheng Xu. 2015. Multi-object tracking via MHT with multiple information fusion in surveillance video. Multimedia Systems 21, 3 (2015), 313–326.

本文提出了一个基于多信息融合的多假设追踪算法，涉及到HSV-LBP表观特征、局 部动作模式以及排斥惯性模型。

1. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 770–778.

提出了一个残差网络框架，有利于训练更深的网络，使得图像的识别准确率得到提升。

1. Min Lin, Qiang Chen, and Shuicheng Yan. 2013. Network in network. arXiv:1312.4400 (2013).

提出MLP卷积层和全局平均池化，改进了传统的CNN网络，减少了需要训练的网络参 数。

1. Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in Neural Information Processing Systems. IEEE, 91–99.

Faster RCNN已经将特征抽取(feature extraction)，proposal提取，bounding box regression(rect refine)，classification都整合在了一个网络中，使得综合性能有较大 提高，在检测速度方面尤为明显。

1. Meng Wang, Changzhi Luo, Richang Hong, Jinhui Tang, and Jiashi Feng. 2016. Beyond object proposals: Random crop pooling for multi-label image recognition. IEEE Transactions on Image Processing 25, 12 (2016), 5678–5688.

针对多标签图像识别任务，提出一个随机剪切的池化方法，不需要预先产生大量的 object proposals，也不需要后处理步骤。

1. Wongun Choi. 2015. Near-online multi-target tracking with aggregated local flow descriptor. In IEEE International Conference on Computer Vision. IEEE, 3029–3037.

本文解决了MOT的两个关键挑战，一是如何准确地衡量两个检测之间的相似性， 二是如何有效地把全局跟踪的算法思想用至在线应用中。

1. Byungjae Lee, Enkhbayar Erdenee, Songguo Jin, Mi Young Nam, Young Giu Jung, and Phill Kyu Rhee. 2016. Multi-class Multi-object Tracking Using ChangingPoint Detection. In European Conference on Computer Vision. Springer, 68–83.

本文提出了一种新颖的多类别多物体跟踪框架，使用变点检测模型来检测突变和异 常。

1. Anton Milan, Stefan Roth, and Konrad Schindler. 2014. Continuous energy minimization for multitarget tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence 36, 1 (2014), 58–72.

该论文从运动的整体性出发，提出了一个整体的，比较贴合运动特征的能量函数，然后通过对该能量函数进行寻优得到较好的跟踪结果。

1. Dan Oneata, Jérôme Revaud, Jakob Verbeek, and Cordelia Schmid. 2014. Spatiotemporal object detection proposals. In European conference on computer vision. Springer, 737–752.

本文提出基于超体素单元的合并来获得视频的细管状提议，又提出一种新颖的产生超体素的方法，从对每帧所提取的超像素进行结构化聚类开始进行。

1. Xindi Shang, Tongwei Ren, Hanwang Zhang, Gangshan Wu, and Tat-Seng Chua.2017. Object trajectory proposal. In IEEE International Conference on Multimedia and Expo. IEEE.

本文提出一种产生物体轨迹提议的方法，也就是将移动物体与静止物体区分开来分 别产生轨迹提议，然后进行统一评价排序，合并最优结果。

## Visual relation detection

1. Bo Dai, Yuqi Zhang, and Dahua Lin. 2017. Detecting Visual Relationships With Deep Relational Networks. In IEEE Conference on Computer Vision and Pattern Recognition. IEEE.

本文针对视觉关系检测任务提出了一个新的框架，综合考虑表观特征、空间结构以及主宾与谓词间的统计依赖，分别预测主谓宾。

1. Yikang Li, Wanli Ouyang, Xiaogang Wang, and Xiao’ou Tang. 2017. ViP-CNN: Visual Phrase Guided Convolutional Neural Network. In IEEE Conference on Computer Vision and Pattern Recognition. IEEE.

本文提出了一种以视觉短语为指导的消息传递结构，更好地对各个预测分支模型之间的相互依赖关系进行建模。

1. Xiaodan Liang, Lisa Lee, and Eric P. Xing. 2017. Deep Variation-Structured Reinforcement Learning for Visual Relationship and Attribute Detection. In IEEE Conference on Computer Vision and Pattern Recognition. IEEE.

本文通过对图片中出现的物体、关系、属性构建全局语义图，运用一种变化结构化的遍历策略和增强学习来实现模型的运作。

1. Cewu Lu, Ranjay Krishna, Michael Bernstein, and Li Fei-Fei. 2016. Visual relationship detection with language priors. In European Conference on Computer Vision. Springer, 852–869.

本文通过利用关系中的主谓宾的语义信息和表观特征来进行建模，并引入了新的数据集Visual Relationship Dataset。

1. Hanwang Zhang, Zawlin Kyaw, Shih-Fu Chang, and Tat-Seng Chua. 2017. Visual Translation Embedding Network for Visual Relation Detection. In IEEE Conference on Computer Vision and Pattern Recognition. IEEE.

本文通过利用translation embedding的思想将主语物体的特征与宾语物体的特征之间的差与它们之间的谓语关系建立等式，进行关系检测的建模。

1. Hanwang Zhang, Zawlin Kyaw, Jinyang Yu, and Shih-Fu Chang. 2017. PPR-FCN:Weakly Supervised Visual Relation Detection via Parallel Pairwise R-FCN. In IEEE International Conference on Computer Vision. IEEE.

本文提出了一个基于成对的区域的并行全卷积神经网络，解决弱监督的视觉关系检测任务。

1. Michaela Regneri, Marcus Rohrbach, Dominikus Wetzel, Stefan Thater, Bernt Schiele, and Manfred Pinkal. 2013. Grounding action descriptions in videos. Transactions of the Association for Computational Linguistics 1 (2013), 25–36.

本文针对视频中发生的动作提出了一个具有多模态的语料库，为比较动作短语的相似性建立了标准的数据集并报告了有关视频中动作相似性的·实验结果。

1. C Lawrence Zitnick, Devi Parikh, and Lucy Vanderwende. 2013. Learning the visual interpretation of sentences. In IEEE International Conference on Computer Vision. IEEE, 1681–1688.

通过对句子进行关系三元组的提取，并对提取之后的这些元组进行特征学习，来解决场景生成和场景索引的任务。

## Action recognition

1. An-An Liu, Yu-Ting Su,Wei-Zhi Nie, and Mohan Kankanhalli. 2017. Hierarchical clustering multi-task learning for joint human action grouping and recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 39, 1 (2017), 102–114.

本文针对人类动作分组和识别任务，提出了一个分层群聚的多任务学习方法。

1. Li Niu, Xinxing Xu, Lin Chen, Lixin Duan, and Dong Xu. 2016. Action and event recognition in videos by learning from heterogeneous web sources. IEEE Transactions on Neural Networks and Learning Systems (2016).

本文通过对多种异质资源提取特征并进行融合，来训练出鲁棒的分类器。

1. Yan Yan, Elisa Ricci, Ramanathan Subramanian, Gaowen Liu, and Nicu Sebe. 2014. Multitask linear discriminant analysis for view invariant action recognition. IEEE Transactions on Image Processing 23, 12 (2014), 5599–5611.

本文通过对人类动作的多个视角的特征进行提取，然后用多任务的线性判别分析框架对提取到的特征进行学习，从而实现对人类动作的识别。

1. Yu-Gang Jiang, Qi Dai, Xiangyang Xue, Wei Liu, and Chong-Wah Ngo. 2012. Trajectory-based modeling of human actions with motion reference points. In European Conference on Computer Vision. Springer, 425–438.

本文采用全局和局部参考点来特征化移动信息，对相机移动有较好的鲁棒性，同时将物体间的关系考虑进去，产生了较好的行为特征表示。

1. Heng Wang and Cordelia Schmid. 2013. Action recognition with improved trajectories. In IEEE International Conference on Computer Vision. IEEE, 3551– 3558.

本文通过显式地估计相机移动来改善密集轨迹。

1. Karen Simonyan and Andrew Zisserman. 2014. Two-stream convolutional networks for action recognition in videos. In Advances in Neural Information Processing Systems. 568–576.

本文首次提出two stream网络，主要分为两个流，空间流处理静止图像帧，得到形状信息，时间流处理连续多帧稠密光流，得到运动信息。

1. LiminWang, Yu Qiao, and Xiaoou Tang. 2015. Action recognition with trajectory-pooled deep-convolutional descriptors. In IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 4305–4314.

本文在iDT和two-stream ConvNets的基础上，对提取出来的卷积特征图实施受限于轨迹的采样与池化来获得轨迹池化的深度卷积描述子。