

密集目标检测论文及代码:

传统任务排行榜:<https://paperswithcode.com/sota/object-detection-on-coco>

推荐看:

Cascade Mask R-CNN(Triple-ResNeXt152, multi-scale)

EfficientDet-D7(single-scale)

TridentNet (ResNet-101-Deformable, Image Pyramid)

PANet (ResNeXt-101, multi-scale)

一定要看: Mask R-CNN (HRNetV2p-W48 + cascade)

FreeAnchor (ResNeXt-101)

FCOS (ResNeXt-64x4d-101-FPN 4 + improvements)

FreeAnchor 的研究方向是现在很火的一个点

经典的: YOLOv3 @800 + ASFF* (Darknet-53)

Faster R-CNN (HRNetV2p-W48)

RetinaMask(ResNeXt-101-FPN-GN)

密集物体检测:Focal Loss for Dense Object Detection

ICCV 2017 • Tsung-Yi Lin • Priya Goyal • Ross Girshick • Kaiming He • Piotr Dollár

The highest accuracy object detectors to date are based on a two-stage approach popularized by R-CNN, where a classifier is applied to a sparse set of candidate object locations. In contrast, one-stage detectors that are applied over a regular, dense sampling of possible object locations have the potential to be faster and simpler, but have trailed the accuracy of two-stage detectors thus far. In this paper, we investigate why this is the case. We discover that the extreme foreground-background class imbalance encountered during training of dense detectors is the central cause. We propose to address this class imbalance by reshaping the standard cross entropy loss such that it down-weights the loss assigned to well-classified examples. Our novel Focal Loss focuses training on a sparse set of hard examples and prevents the vast number of easy negatives from overwhelming the detector during training. To evaluate the effectiveness of our loss, we design and train a simple dense detector we call RetinaNet. Our results show that when trained with the focal loss, RetinaNet is able to match the speed of previous one-stage detectors while surpassing the accuracy of all existing state-of-the-art two-stage detectors.

Code is at: <https://github.com/facebookresearch/Detectron>. (pytorch版本)

Keras: [fizyr/keras-retinanet](#)

TSM: Temporal Shift Module for Efficient Video Understanding

ICCV 2019 • Ji Lin • Chuang Gan • Song Han

The explosive growth in video streaming gives rise to challenges on performing video understanding at high accuracy and low computation cost. Conventional 2D CNNs are computationally cheap but cannot capture temporal relationships; 3D CNN based methods can achieve good performance but are computationally intensive, making it expensive to deploy. In this paper, we propose a generic and effective Temporal Shift Module (TSM) that enjoys both high efficiency and high performance. Specifically, it can achieve the performance of 3D CNN but maintain 2D CNN's complexity. TSM shifts part of the channels along the temporal dimension; thus facilitate information exchanged among neighboring frames. It can be inserted into 2D CNNs to achieve temporal modeling at zero computation and zero parameters. We also extended TSM to online setting, which enables real-time low-latency online video recognition and video object detection. TSM is accurate and efficient: it ranks the first place on the Something-Something leaderboard upon publication; on Jetson Nano and Galaxy Note8, it achieves a low latency of 13ms and 35ms for online video recognition. The code is available at: <https://github.com/mit-han-lab/temporal-shift-module>.

pytorch: [MIT-HAN-LAB/temporal-shift-module](#)

Tensorflow 目标检测代码库: [Tensorflow Object Detection API](#)