生物智能与算法 Report-1

赵昱 21821275

1　简介

目标检测问题分为两个阶段，即目标定位（确定目标所在位置，回归问题）和目标分类（确定候选框内的目标所属类别），常规方法通常采用滑动窗口加人工特征分类器进行检测。

YOLO 系列 [1–3] 是首个使用单个 CNN 模型同时处理回归和分类两个问题的神经网络模型。

本项目选择 YOLO 系列作为多类别目标检测（通常叫做通用目标检测）所采用的模型，尝试复现每个版本并进行比较，同时针对一些特殊应用（如人脸检测等单类别问题），根据自己的需求训练模型。

2　选择论文

（1）：[YOLO v1-CVPR2016] You Only Look Once: Unified, Real-Time Object Detection

（2）：[YOLO v2-CVPR2017] YOLO9000: Better, Faster, Stronger

（3）：[YOLO v3-(arxiv)2018] YOLOv3: An Incremental Improvement

3　目标

（1）：精读论文，理解模型，分析方法的优缺点和创新之处

（2）：复现论文

（3）：和其他方法进行对比实验

3.1　论文摘要与翻译

3.1.1　[YOLO v1-CVPR2016] You Only Look Once: Unified, Real-Time Object Detection

Abstract:

We present YOLO, a new approach to object detection. Prior work on object detection repurposes classifiers to perform detection. Instead, we frame object detection as a regression problem to spatially separated bounding boxes and associated class probabilities. A single neural network predicts bounding boxes and class probabilities directly from full images in one evaluation. Since the whole detection pipeline is a single network, it can be optimized end-to-end directly on detection performance.

Our unified architecture is extremely fast. Our base YOLO model processes images in real-time at 45 frames per second. A smaller version of the network, Fast YOLO, processes an astounding 155 frames per second while still achieving double the mAP of other real-time detectors. Compared to state-of-the-art detection systems, YOLO makes more localization errors but is less likely to predict false positives on background. Finally, YOLO learns very general representations of objects. It outperforms other detection methods, including DPM and R-CNN, when generalizing from natural images to other domains like artwork.

摘要翻译：

我们提出 YOLO——一种新的目标检测方法。以前的目标检测工作重新利用分类器来执行检测。相反，我们将目标检测框架看作回归问题从空间上分割边界框和相关的类别概率。单个神经网络在一次评估中直接从完整图像上预测边界框和类别概率。由于整个检测流程是单一网络，因此可以直接对检测性能进行端到端的优化。

我们的单一网络架构非常快。我们的基础 YOLO 模型以 45 帧/秒的速度实时处理图像。网络的一个较小版本，快速 YOLO，每秒能处理惊人的 155 帧，同时实现其它实时检测器两倍的 mAP。与最先进的检测系统相比，YOLO 产生了更多的定位误差，但不太可能在背景上的预测假阳性。最后，YOLO 学习目标非常通用的表示。当从自然图像到艺术品等其它领域泛化时，它都优于其它检测方法，包括 DPM 和 R-CNN。

3.1.2 [YOLO v2-CVPR2017] YOLO9000: Better, Faster, Stronger

Abstract:

We introduce YOLO9000, a state-of-the-art, real-time object detection system that can detect over 9000 object categories. First we propose various improvements to the YOLO detection method, both novel and drawn from prior work. The improved model, YOLOv2, is state-of-the-art on standard detection tasks like PASCAL VOC and COCO. Us ing a novel, multi-scale training method the same YOLOv2 model can run at varying sizes, offering an easy tradeoff between speed

and accuracy. At 67 FPS, YOLOv2 gets 76.8 mAP on VOC 2007. At 40 FPS, YOLOv2 gets 78.6 mAP, outperforming state-of-the-art methods like Faster R-CNN with ResNet and SSD while still running significantly faster. Finally we propose a method to jointly train on object detection and classification. Using this method we train YOLO9000 simultaneously on the COCO detection dataset and the ImageNet classification dataset. Our joint training allows YOLO9000 to predict detections for object classes that don't have labelled detection data. We validate our approach on the ImageNet detection task. YOLO9000 gets 19.7 mAP on the ImageNet detection validation set despite only having detection data for 44 of the 200 classes. On the 156 classes not in COCO, YOLO9000 gets 16.0 mAP. YOLO9000 predicts detections for more than 9000 different object categories, all in real-time.

摘要翻译:

我们提出 YOLO9000, 这是一种先进的、实时的目标检测系统, 可以检测超过 9000 个类别。首先, 我们对 YOLO 做了一些改进, 得到了提升后的模型 YOLO v2, 这是一种在如 Pascal VOC 以及 COCO 等任务上表现出色, 使用一种新奇的、多尺度的训练方法, YOLO v2 可以适应不同的尺寸, 可以在速度与精度之间提供一个很好的平衡。在 67 帧率的时候, YOLO V2 可以获得 76.8 的 mAP 在 VOC2007 上, 40 帧率的时候, 可以获得 78.6 的 mAP 值, 这比 Faster R-cnn 要好。最后, 我们提出了一种联合训练的方法来进行目标检测和分类。使用这种方法, 我们训练了 YOLO 9000, 它可以在 ImageNet 和 COCO 上同时进行训练, 我们的这种联合训练方法允许 YOLO9000 可以进行无标签目标的预测。我们在 ImageNet 检测任务中验证了我们的方法, 在 44 类已有数据集中 YOLO9000 获得了 19.7 的 mA P 值, 在其余的 156 类没有的样本中, 获得了 16.0 的 mAP 值。但是 YOLO9000 可不是只能预测 200 类数据, 它可以同时检测出 9000 种目标, 并且仍然可以实时的进行。

3.1.3 [YOLO v3-(arxiv)2018] YOLOv3: An Incremental Improvement

Abstract:

We present some updates to YOLO! We made a bunch of little design changes to make it better. We also trained this new network that's pretty swell. It's a little bigger than last time but more accurate. It's still fast though, don't worry. At 320 x 320 YOLOv3 runs in 22 ms at 28.2 mAP, as accurate as SSD but three times faster. When we look at the old .5 IOU mAP detection metric YOLOv3 is quite good. It achieves 57:9 AP50 in 51 ms on a Titan X, compared to 57:5 AP50 in 198 ms by RetinaNet, similar performance but 3.8x faster. As always, all the

code is online at https://pjreddie.com/yolo/.

摘要翻译:

本文为 YOLO 提供了一系列更新！它包含一堆小设计, 可以使系统的性能得到更新; 也包含一个新训练的、非常棒的神经网络, 虽然比上一版更大一些, 但精度也提高了。不用担心, 虽然体量大了点, 它的速度还是有保障的。在输入 320×320 的图片后, YOLOv3 能在 22 毫秒内完成处理, 并取得 28.2mAP 的成绩。它的精度和 SSD 相当, 但速度要快上 3 倍。和旧版数据相比, v3 版进步明显。在 Titan X 环境下, YOLOv3 的检测精度为 57.9AP, 用时 51ms; 而 RetinaNet 的精度只有 57.5AP, 但却需要 198ms, 相当于 YOLOv3 的 3.8 倍。

参考文献

[1] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[2] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.

[3] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.