极客大学

极客大学机器学习训练营
人工智能应用简介

王然

众微科技 AI Lab 负责人

二〇二一年九月二十三日

大纲

大纲

极客大学

图像识别

极客大学

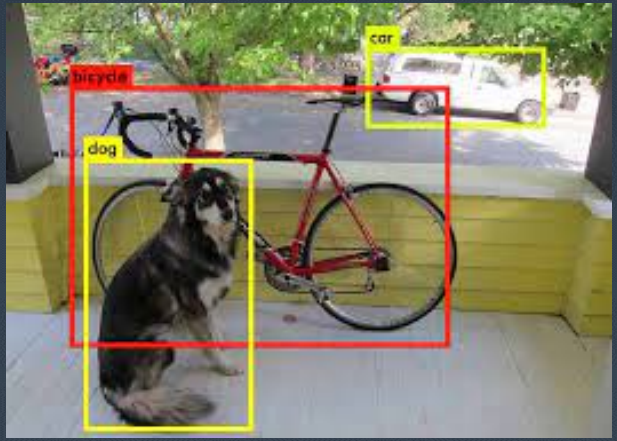
- 本质与其他深度学习应用类似；
- 来源于 ImageNet；
- 经典模型见Torch Vision Model Zoo。
- 请注意，并不是越深的模型效果越理想；经常需要组合不同的模型；
- ResNet50 一般是很好的开始的 baseline。

图像识别实现

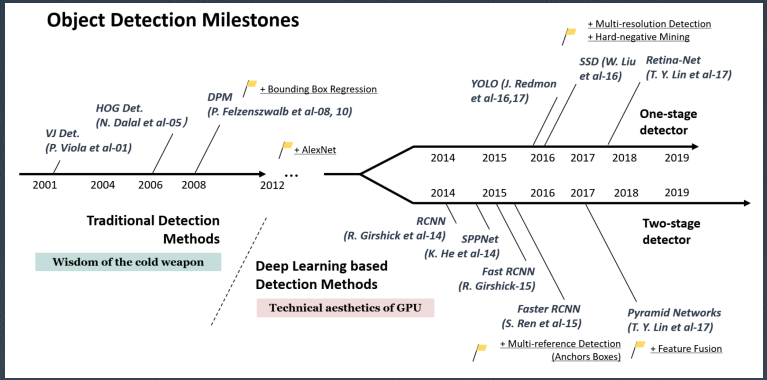
极客大学

- 可以直接使用 PyTorch Lightening 进行实现；
- 具体实现见 Colab。

图像检测任务



经典模型发展脉络


Object Detection Milestones
+ Multi-resolution Detection
+ Hard-negative Mining
SSD (W. Liu et al-16)
Retina-Net (T. Y. Lin et al-17)
YOLO (J. Redmon et al-16,17)
One-stage detector
+ Bounding Box Regression
DPM (P. Felzenszwalb et al-08, 10)
HOG Det. (N. Dalal et al-05)
VI Det. (P. Viola et al-01)
+ AlexNet
2001
2004
2006
2008
2012
...
2014
2015
2016
2017
2018
2019
2014
2015
2016
2017
2018
2019
Two-stage detector
RCNN (R. Girshick et al-14)
SPPNet (K. He et al-14)
Fast RCNN (R. Girshick-15)
Faster RCNN (S. Ren et al-15)
Pyramid Networks (T. Y. Lin et al-17)
Traditional Detection Methods
Wisdom of the cold weapon
Deep Learning based Detection Methods
Technical aesthetics of GPU
+ Multi-reference Detection (Anchors Boxes)
+ Feature Fusion

重要经典模型

极客大学

- RCNN 系列 (Girshick et al. 2014; Girshick 2015; Ren et al. 2015)；
- YOLO 系列 (Redmon et al. 2016; Redmon and Farhadi 2017, 2018; Bochkovskiy, Wang, and Liao 2020)；
- 很多网络设计目的不仅仅是为了准确率，同样也是为了提高训练和推断效率；
- 最新的模型引入了 transformer 概念，如 Swin Transformer(Liu et al. 2021)。

实现

- 主要采用Detectron2；
- 实现见 Colab Notebook。

大纲

传统 NLP 模型

极客大学

- 传统 NLP 模型一般建立在 Word Embedding + 上层网络基础上；
- 上层网络一般指 LSTM 或 CNN；
- 虽然传统 NLP 也称之为深度学习，但是网络深度一般有限；
- LSTM 难以在平衡计算效率的前提下提升网络深度。

Transformer 及 BERT

极客大学

- 突破口之一：Transformer 使得网络的深度可以在充分利用计算资源的基础上进行增加；
- 突破口之二：BERT 的预训练模型；
- 当前手段：预训练语言模型 + 任务相关的上层网络。

BERT 当中的特殊 Token

- [CLS]
- [SEP]
- [PAD]

一些其他语言模型

极客大学

- Roberta(Liu et al. 2019);
- XLNet(Yang et al. 2019);
- Albert(Lan et al. 2019);
- 一些巨大的语言模型 (Brown et al. 2020; Fedus, Zoph, and Shazeer 2021)
- 一些很小的语言模型：蒸馏 (Sanh et al. 2019) 或改变 Transformer 结构 (Choromanski et al. 2020)
- 大部分时候不要自己进行预训练;
- 英文和中文预训练效果有明显差异。

实践

见官方 Colab

大纲

见 Model-based RL 部分

大纲

- Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao (2020). "Yolov4: Optimal speed and accuracy of object detection". In: *arXiv preprint arXiv:2004.10934*.
- Brown, Tom B et al. (2020). "Language models are few-shot learners". In: *arXiv preprint arXiv:2005.14165*.
- Choromanski, Krzysztof et al. (2020). "Rethinking attention with performers". In: *arXiv preprint arXiv:2009.14794*.
- Fedus, William, Barret Zoph, and Noam Shazeer (2021). "Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity". In: *arXiv preprint arXiv:2101.03961*.
- Girshick, Ross (2015). "Fast r-cnn". In: *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448.
- Girshick, Ross et al. (2014). "Rich feature hierarchies for accurate object detection and semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580–587.

- Lan, Zhenzhong et al. (2019). "Albert: A lite bert for self-supervised learning of language representations". In: *arXiv preprint arXiv:1909.11942*.
- Liu, Yinhan et al. (2019). "Roberta: A robustly optimized bert pretraining approach". In: *arXiv preprint arXiv:1907.11692*.
- Liu, Ze et al. (2021). "Swin transformer: Hierarchical vision transformer using shifted windows". In: *arXiv preprint arXiv:2103.14030*.
- Redmon, Joseph and Ali Farhadi (2017). "YOLO9000: better, faster, stronger". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7263–7271.
- – (2018). "Yolov3: An incremental improvement". In: *arXiv preprint arXiv:1804.02767*.
- Redmon, Joseph et al. (2016). "You only look once: Unified, real-time object detection". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788.

- Ren, Shaoqing et al. (2015). "Faster r-cnn: Towards real-time object detection with region proposal networks". In: *Advances in neural information processing systems* 28, pp. 91–99.
- Sanh, Victor et al. (2019). "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter". In: *arXiv preprint arXiv:1910.01108*.
- Yang, Zhilin et al. (2019). "Xlnet: Generalized autoregressive pretraining for language understanding". In: *Advances in neural information processing systems* 32.