YOLO-LITE: A Real-Time Object Detection Web Implementation

Rachel Huang

REU UNCW

July 13, 2018

Collaborators

Jonathan Pedoeem:

Summerlin Thompson:

Dr. Cuixian Chen:

Dr. Yishi Wang:

Overview

Why Object Detection?

Definition:

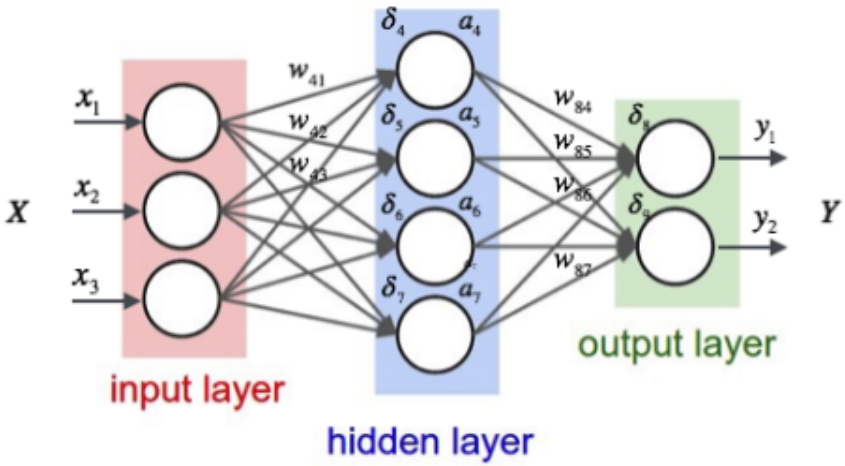
A field of computer vision to detect and classify objects.

Applications:

- Self-driving vehicles
- Security
- Face detection
- Sports (ball tracking)

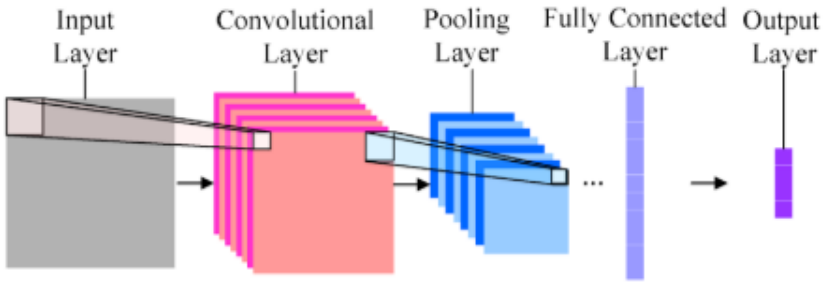
Neural Networks

Fully Connected Neural Network



https://medium.com/@curiousily/tensorflow-for-hackers-part-iv-neural-network-from-scratch-1a4f504dfa8

Convolutional Neural Network



http://www.mdpi.com/2078-2489/7/4/61

Introduction: YOLO

You-Only-Look-Once (YOLO)

1. Efficient object detection model.
2. Predicts bounding boxes and classification in one look.
3. mAP of 63.4% and 45 FPS.

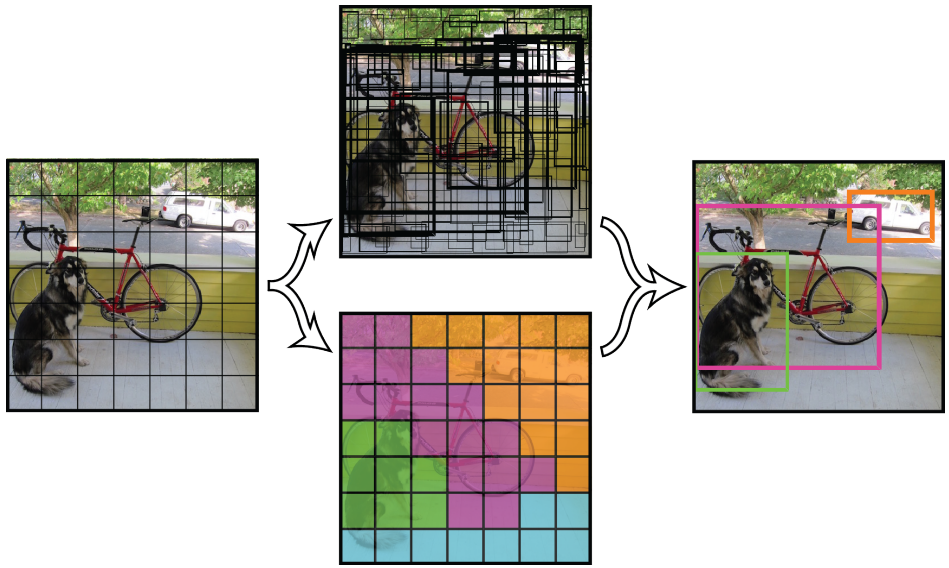


Figure:
[Redmon et al.(2016)Redmon, Divvala, Girshick, and Farhadi]

Data Sets	PASCAL VOC	MS COCO
Version	2007 + 2012	2014
Number of Classes	20	80
Number of Training Images	≈5,011	≈40,775

Introduction: Goals of YOLO-LITE

Goals:

- Implement an object detection algorithm that can work real time on a CPU
- Goal is around 10 FPS
- Maintain around 30% mAP
- Implement onto a website

Training on VOC

Model	mAP	FPS
tiny-YOLOv2-VOC	40.48%	2.4
trial 1	12.64%	1.56
trial 2	28.60%	6.94
trial 3	28.11%	9.5
trial 4	2.35%	5.2
trial 5	0.55%	3.5
trial 6	**30.80%**	**21**
trial 7	29.33%	9.7
trial 8	16.84%	5.7
trial 9	3.93%	—
trial 10	24.22%	—
trial 11	24.22%	21
trial 12	24.22%	8.2
trial 13	24.22%	21
trial 14	24.22%	3

Variables Tested:

- Number of Layers
- Number of filters
- Batch Normalization
- Size of input image

Training on COCO

Trial 6 Training:

- mAP: 11%
- 21 FPS
- Training time: 24 hours
- Loss: 6.5

Trial 6 Architecture:

- 6 layers
- No batch normalization
- Image size: 224 x 224

Comparison of Architectures

Tiny-YOLOv2-VOC

Layer	Filters	Size	Stride	Param
Conv1 (C1)	16	3	1	432
Max Pool (MP)		2	2	
C2	32	3	1	4,608
MP		2	2	
C3	64	3	1	18,432
MP		2	2	
C4	128	3	1	73,728
MP		2	2	
C5	256	3	1	294912
MP		2	2	
C6	512	3	1	1 Mil
MP		2	2	
C7	1024	3	1	4 Mil
C8	1024	3	1	9 Mil
C9	125	1	1	128,000
Regional	–	–	–	–

YOLO-LITE

Layer	Filters	Size	Stride	Param
C1	16	3	1	432
MP		2	2	
C2	32	3	1	4,608
MP		2	2	
C3	64	3	1	18,432
MP		2	2	
C4	128	3	1	73,728
MP		2	2	
C5	256	3	1	294,912
MP		2	2	
C6	125	1	1	32,000
Regional	–	–	–	–

Demo

Conclusion

Next Steps:

- Train our best model on COCO and VOC for a longer time
- Wrap up the website

Future Work:

- Continue increasing mAP
- Pretraining on ImageNet
- Pruning
- Combining R-CNN and YOLO

References

Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi.
You only look once: Unified, real-time object detection.
In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

Questions?