Bryant University

Bryant Digital Repository

Computer Information Systems Conference Proceedings

Computer Information Systems Faculty Publications and Research

11-2019

Realtime Object Detection via Deep Learning-based Pipelines

James G. Shanahan
Bryant University, jshanahan@bryant.edu

Liang Dai
University of California - Berkeley

Follow this and additional works at: https://digitalcommons.bryant.edu/cis-conf-proceedings

Part of the Artificial Intelligence and Robotics Commons

Recommended Citation

Shanahan, James G. and Dai, Liang, "Realtime Object Detection via Deep Learning-based Pipelines" (2019). *Computer Information Systems Conference Proceedings.* Paper 1.
https://digitalcommons.bryant.edu/cis-conf-proceedings/1

This Conference Proceeding is brought to you for free and open access by the Computer Information Systems Faculty Publications and Research at Bryant Digital Repository. It has been accepted for inclusion in Computer Information Systems Conference Proceedings by an authorized administrator of Bryant Digital Repository. For more information, please contact dcommons@bryant.edu.

Realtime Object Detection via Deep Learning-based Pipelines

James G. Shanahan
Bryant University, Rhode Island, USA
& University of California, Berkeley
Berkeley, CA, US
James.Shanahan@gmail.com

Liang Dai
Facebook, Menlo Park, CA, USA
& University of California, Santa Cruz
Santa Cruz, CA, US
liangdai16@gmail.com

ABSTRACT

Ever wonder how the Tesla Autopilot system works (or why it fails)? In this tutorial we will look under the hood of self-driving cars and of other applications of computer vision and review state-of-the-art tech pipelines for object detection such as two-stage approaches (e.g., Faster R-CNN) or single-stage approaches (e.g., YOLO/SSD). This is accomplished via a series of Jupyter Notebooks that use Python, OpenCV, Keras, and Tensorflow. No prior knowledge of computer vision is assumed (although it will be help!). To this end we begin this tutorial with a review of computer vision and traditional approaches to object detection such as Histogram of oriented gradients (HOG).

CCS CONCEPTS

Artificial intelligence, machine learning, computer vision, object detection, tutorial

KEYWORDS: Computer vision, image classification, deep learning, object detection, region proposal networks, R-CNN, Fast R-CNN, Faster R-CNN, Single Shot MultiBox Detector (SSD), You Only Look Once (YOLO), RetinaNet, Mask R-CNN

1 INTRODUCTION

Computer vision (CV) has been revolutionized by deep learning in the past 7-8 years and vice versa. Exciting real world deployments of computer vision are appearing in the cloud and on the edge. For example, autonomous vehicles such as self-driving cars, face detection, checkout-less shopping, security systems, cancer detection, and more.

In this tutorial, we will briefly overview the basics of computer vision before focusing on object detection and other computer vision areas, from the following perspectives: state-of-the-art research, key algorithms, applications, and open challenges. We also present state-of-the-art pipelines that are being used in application areas, such as, advanced driver assistance systems (ADAS), driver monitoring systems (DMS), disease detection, such as lung cancer and heart disease, and security and surveillance systems.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.
CIKM '19, November 3–7, 2019, Beijing, China
© 2019 Copyright is held by the owner/author(s).
ACM ISBN 978-1-4503-6976-3/19/11. https://doi.org/10.1145/3357384.3360320

These pipelines are based on complex deep convolutional neural network (CNN) architectures (often 50-60 layers deep or more), multi-task loss functions, and are either two-stage (e.g., Faster R-CNN) or single-stage (e.g., YOLO/SSD) in nature. We will demonstrate in a Jupyter notebook how to build, train, and evaluate computer vision applications with a primary focus on building an object detection application from scratch to detect objects such as logos in images/video. Recent developments in object detection such as panoptic segmentation will also be reviewed.

Target audience: This tutorial provides a technical overview of modern computer vision systems for object detection for a general computing and engineering audience. This tutorial for beginners and practitioners in the field of object detection.

Prerequisite knowledge of audience: Programming background; comfort with mathematical and algorithmic reasoning; familiarity with basic machine learning concepts; exposure to deep learning; experience with Python (or the ability to learn it quickly). No prior knowledge of computer vision is assumed (although it will be help!).

2 Tutorial OUTLINE

Chapter	Topics
PART 1: Introduction and Infrastructure	Introduction Tutorial logistics Computer vision and object detection applications survey Infrastructure setup (Google CoLab, tutorial Docker container for local and on cloud CPU/GPU) OpenCV, PIL, Pillow *LAB: Transfer learning + Multi-task loss function*
PART 2: Traditional computer vision	Image processing and computer vision • Introduction and brief history • Convolutional kernels Traditional image processing and classification • Haar features, • Scale-invariant feature transform (SIFT), • and Histogram of oriented gradients (HOG) features Viola–Jones object detection framework OpenCV framework LAB: face detection using conventional image processing pipelines

PART 3: Convolutional neural networks	Convolutional neural networks (CNN) • Kernels, pooling, feature maps, ReLU • Biological inspiration for convolutional layers Loss functions for vision tasks Deep learning pipelines in Keras for image classification Deep NN Architectures for computer vision Skip connections (e.g., Resnet) Bottleneck layers (Inception Net) Reference architectures (AlexNet, VGGNet, Inception Net, ResNet) LAB: Image classification with AlexNet: Cifar10, Cifar100 Implement Alexnet from scratch Train network on GPU
PART 4: Object detection and two-stage approaches	Object detection overview Multitask loss functions Lab: finding Nemo • setting up the training data • detect the biggest object in the image Two-stage object detection explained • Region Proposal methods • Bounding box predictor and region classifier Two-stage approaches • R-CNN, Fast R-CNN, Faster R-CNN Object detection via segmentation approaches • Mask R-CNN
PART 5: Single stage object detection	Single step object detection • Pyramid architectures Single step approaches • Single Shot MultiBox Detector (SSD) • You Only Look Once (YOLO) • Retinanet and focal loss LAB: Logo detector using retinanet • Label training data • Set up logo training data • Train a logo detector • Evaluate
PART 6: Advanced applications	Autonomous vehicles Deploying vision applications on the edge MobileNet Disease detection
PART 7: Conclusions	Tutorial summary Where to go from here

3 PRESENTER BIOGRAPHY

Dr. James G. Shanahan, Bryant University and UC Berkeley

Dr. James G. Shanahan has spent the past 30 years developing and researching cutting-edge artificial intelligent systems splitting his time between industry and academia. As of the fall 2019, Jimi is the Rowe Professor of Data Science at Bryant University, Rhode Island. He also holds a visiting appointment at the University of California at Berkeley, and has (co) founded several companies that leverage AI/machine learning/deep learning/computer vision in verticals such as digital advertising, web search, and smart cameras. His latest venture focuses on making driving safer via advanced driver assistance systems (ADAS) and driver monitoring systems (DMS), along with rewarding safer driving with reduced insurance premiums. Previously he has held appointments at AT&T (Executive Director of Research), NativeX (SVP of data science), Xerox Research (staff research scientist), and Mitsubishi. He also advises several high-tech startups (including Aylien, ChartBoost, DigitalBank, LucidWorks, and others).

Dr. Shanahan received his PhD in engineering mathematics and computer vision from the University of Bristol, U. K., and holds a Bachelor of Science degree from the University of Limerick, Ireland. He is an EU Marie Curie fellow. In 2011 he was selected as a member of the Silicon Valley 50 (Top 50 Irish Americans in Technology).

Liang Dai, UC Santa Cruz and Facebook

Liang Dai is a data scientist at Facebook. At Facebook, he works on deep learning, large scale data mining projects in distributed platform, e.g. AWS, Hadoop, Spark, etc. He focuses on end to end data modeling pipeline: including preprocessing raw data, designing behavior and non-behavior features, selecting features based on experimental results, building predictive models and deploying models in production to handle large volume of ads placement requests. Liang is also pursuing Ph.D. in Technology Information and Management department, UC Santa Cruz. There he does research in data mining on digital marketing, including campaign evaluation, online experiment design, customer value improvement, etc. Previously he was a data scientist at NativeX, a leading ad technology company for mobile games. Liang received the B.S. and the M.S. from Information Science and Electronic Engineering department, Zhejiang University.

REFERENCES

[1] [Khan] -- Salman Khan, Hossein Rahmani, Syed Afaq Ali Shah (2018), A Guide to Convolutional Neural Networks for Computer Vision, Morgan and Claypool.

[2] [Rosebrock] -- Adrian Rosebrock (2018), Deep Learning for Computer Vision with Python, https://www.pyimagesearch.com/deep-learning-computer-vision-python-book/

[3] [Forsyth] -- Forsyth, David A., and Jean Ponce. "A modern approach." Computer vision: a modern approach (2003): 88-101.

[4] [Goodfellow] -- Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1). Cambridge: MIT press. PDF available online: https://www.deeplearningbook.org/

[5] [Szeliski] -- Szeliski, R. (2010). Computer vision: algorithms and applications. Springer Science & Business Media. PDF available online: http://szeliski.org/Book/drafts/SzeliskiBook_20100903_draft.pdf (Read Chapter 14 on object detection etc.)