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**  
https://digitalcommons.bryant.edu/cis-conf-proceedings/1
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).

CPS 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.

© 2019 Copyright is held by the owner/author(s).


2 Tutorial OUTLINE

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Topics</th>
</tr>
</thead>
</table>
| 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

<table>
<thead>
<tr>
<th>Topic</th>
<th>Details</th>
</tr>
</thead>
</table>
| 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

<table>
<thead>
<tr>
<th>Topic</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object detection overview</td>
<td></td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Topic</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single step object detection</td>
<td></td>
</tr>
</tbody>
</table>
- 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

<table>
<thead>
<tr>
<th>Topic</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autonomous vehicles</td>
<td></td>
</tr>
</tbody>
</table>
Deploying vision applications on the edge  
MobileNet  
Disease detection |

### PART 7: Conclusions

<table>
<thead>
<tr>
<th>Summary</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tutorial summary</td>
<td>Where to go from here</td>
</tr>
</tbody>
</table>

## 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).

### REFERENCES