

# A Learning Framework for Target Detection and Human Face Recognition in Real Time

Jiaxing Huang, University of Glasgow, Glasgow, UK

Zhengnan Yuan, University of Glasgow, Glasgow, UK

Xuan Zhou, University of Glasgow, Glasgow, UK

## ABSTRACT

Inspired by the function, mechanism and efficiency of the visual nerve system of human beings, a revolutionary detection and reorganization method named YOLO is present to provide an accurate, stable and fast arithmetic for a variety of targets, be it target detection for unmanned vehicle, car license recognition and optimization for surveillance. The traditional method for object detection is to reuse the classifier to implement detection, in contrast, the method named YOLOV2 process this problem by considering it in the mathematical area as a regression of spatially discrete bordered areas and relative class probability. However, as a cost of stable and fast response of this arithmetic, inaccurate detection maybe caused by YOLOV2 when the detected object is tiny (e.g., face recognition in surveillance). In this article, the authors provide a new method to further improve the performance of YOLOV2 by utilizing the accurate, stable and fast properties of YOLOV2 and editing the original code of YOLOV2 to eliminate the inaccuracy of tiny object detection, and implement this method on an embedded system.

## KEYWORDS

Deep Learning Frameworks, Increased Resolution, Real Time, Target Detection

## 1. INTRODUCTION

Target detection and face recognition, one of the most challenging issue in the field of computer vision, devotes at finding the correct position of the targets and deciding their subordinate categories (even recognize the face if detected). Artificial learning, a developing application which plays an important role in intelligent monitor, behavior recognition and flow detection has become a hot topic, especially its theory and application of deep learning. Comparing with traditional target detection method, utilizing deep learning in new target detection model has gradually became the main trend for research.

The traditional target detection method is generally divided into three stages. In first process, the features of some given graph are extracted from the candidate region that are selected on the image before. Then the classification is completed by the trained classifier. In the second process, manual acquisition is often required in extracting the expression information related to the target in the original input before the expression information doing classifier learning in the extracted feature information which is related to the target information (Carpenter, Grossberg, & Reynolds, 1991). For instance, SIFT (Scale-invariant feature transform) (Lowe, 2004), HOG (Histogram of gradient) (Dalal & Triggs, 2005). However, limitations also exit. On one aspect, artificial methods

DOI: 10.4018/IJTHI.2019070105

mainly depend on specific detection tasks. For different targets or the same target of different forms of objects, designers are expected to create desired methods of extracting the features and the final recognition performance of the model is also associated with their own experience (Ouyang & Wang, 2013). On other aspect, feature extraction and classification training are separated by the traditional test model (Sermanet, 2013). As a result, some useful information may not never be recovered from the train of classification once the features that are manually extracted in feature description are not minute enough to describe the target. These shortcomings prevent the traditional detection model from getting more consistent description with the target characteristics.

Before introducing convolutional neural networks, DPM (Deformable Parts Model) (Felzenszwalb, Girshick, Mcallester, & Ramanan, 2010) is a classical algorithm for target detection, whose core theory is achieving the combination with SVM (Zhu, Zheng, & Savvides, n.d.). In some testing tasks, DPM has desired results. CNN (Convolutional Neural Network) in deep learning combines artificial neural networks with convolution operations to recognize a wide variety of target model and is robust to a certain degree of distortion and deformation. Simultaneously, It uses sparse connections and weight sharing which greatly reducing the parameters of the traditional neural network parameters. That also the reason why many detection methods or models based on convolutional neural networks achieved significant results. In 2014, R-CNN (Girshick et al., 2014) algorithm, indicated by Girshick, introduced CNN in the field of target detection for the first time. Expectedly, the results are far better than the traditional approach. After R-CNN, other classical methods exist, be it SPPNet (He, Zhang, Ren, & Sun, 2015), Faster R-CNN (Ren et al., 2015), R-FCN (Dai, Li, He, & Sun, 2016), FPN (Lin et al., 2016), Mask RCNN (He et al., 2017), YOLO (Redmon, Divvala, Girshick, & Farhadi, 2016), SSD (Liu et al., 2016), YOLOv2 (Redmon & Farhadi, 2016). These algorithms can be divided into two categories: deep learning target detection algorithm based on candidate region and deep learning target detection algorithm based on regression method.

The deep learning target detection algorithm based on candidate region solves two problems about the traditional target detection. Firstly, lack of pertinence, time complexity and window redundancy for the region selection strategy based on sliding window. Secondly, lack of robustness to variations in diversity for manual design. Specifically, the method can obtain high recall rate in the case of fewer windows since the possible positions of the targets are pre-selected according to the texture and edge in the candidate region. This improvement reduces the time complexity of the subsequent operation, and obtains candidate window with higher quality than sliding window.

YOLO SSD and YOLOv2 are recommended in the depth learning target detection algorithm based on regression method. Nowadays, as one of the most important algorithm frame in the field of target detection, Faster R-CNN promoted a RPN structure for generating candidate regions and an anchor mapping mechanism for target frame position. These innovations directly stimulate and enlighten a series of follow-up research work, such as literature. This series of algorithm frames is more accurate than the YOLO in locating the target position, especially in detecting accuracy. However, there are three cons based on candidate region for the series of Faster R-CNN: Firstly, the algorithms use or refer the existing image classification models, such as VGG (Simonyan & Zisserman, 2014), ResNet (He, Zhang, Ren, & Sun, 2016) and so on. Comparing with the Darknet models of YOLO, those classification models would generate large amount of parameters and time consuming when training. As a result of this, Faster R-CNN algorithm requires higher for hardware devices, such as R-FCN, Mask R-CNN algorithm, the experimental part of the literature clearly proposes that network model requires 8 GPU trainings which is impossible for normal labs. Fortunately, only one laptop with graphic card whose memory is more than 4G is required for the Darknet model of YOLO, then the model training of PASCAL VOC data patterns can be achieved in the simple and low-cost environment. Apart from that, fast R-CNN algorithm series is far behind the real-time requirement for the checking speed. But YOLO can fix this problem since the GPU acceleration on PASCAL VOC can achieve 45FPS (YOLOv2 algorithm can achieve 67FPS). YOLO uses the entire image as the input of the network, directly detects the regression location of the object's borders and

12 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage:

[www.igi-global.com/article/a-learning-framework-for-target-detection-and-human-face-recognition-in-real-time/227402?camid=4v1](http://www.igi-global.com/article/a-learning-framework-for-target-detection-and-human-face-recognition-in-real-time/227402?camid=4v1)

This title is available in InfoSci-Journals, InfoSci-Journal Disciplines Communications and Social Science, InfoSci-Journal Disciplines Computer Science, Security, and Information Technology. Recommend this product to your librarian:

[www.igi-global.com/e-resources/library-recommendation/?id=2](http://www.igi-global.com/e-resources/library-recommendation/?id=2)

## Related Content

---

### Digital Transition in Alaska

Ping Lan (2012). *Social Development and High Technology Industries: Strategies and Applications* (pp. 136-150).

[www.igi-global.com/chapter/digital-transition-alaska/58718?camid=4v1a](http://www.igi-global.com/chapter/digital-transition-alaska/58718?camid=4v1a)

### Assessing a Spanish Translation of the End-User Computing Satisfaction Instrument

George E. Heilman and Jorge Brusa (2009). *Cross-Disciplinary Advances in Human Computer Interaction: User Modeling, Social Computing, and Adaptive Interfaces* (pp. 59-69).

[www.igi-global.com/chapter/assessing-spanish-translation-end-user/7278?camid=4v1a](http://www.igi-global.com/chapter/assessing-spanish-translation-end-user/7278?camid=4v1a)

### Addressing Data Collection Challenges in ICT for Development Projects

Devendra Dilip Potnis (2015). *International Journal of Information Communication Technologies and Human Development* (pp. 36-55).

[www.igi-global.com/article/addressing-data-collection-challenges-in-ict-for-development-projects/135532?camid=4v1a](http://www.igi-global.com/article/addressing-data-collection-challenges-in-ict-for-development-projects/135532?camid=4v1a)

## Examining the Varying Influence of Social and Technological Aspects on Adoption and Usage of Knowledge Management Systems

Andrea J. Hester (2011). *International Journal of Social and Organizational Dynamics in IT* (pp. 49-65).

[www.igi-global.com/article/examining-varying-influence-social-technological/50534?camid=4v1a](http://www.igi-global.com/article/examining-varying-influence-social-technological/50534?camid=4v1a)