International Core Journal of Engineering Volume 8 Issue 7, 2022
ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0070

# Research on Object Detection based on HoloLens2

Hao Su[a], Zilong Guo[b], and Haotian Lu[c]

School of Automation, Wuhan University of Technology, Wuhan 430070, China

[a]303065@whut.edu.cn, [b]302987@whut.edu.cn, [c]303036@whut.edu.cn

## Abstract

**In recent years, with the continuous development of society, emerging science and technology have emerged in an endless stream. HoloLens2 came into being, it is widely used in military, medical and educational fields, through the combination of deep learning object detection technology and mixed reality technology, to achieve users and the surrounding environment of the holographic projection interaction. With the superior capabilities of the HoloLens2 device, the hybrid world around the user becomes interactive and actionable, connecting the virtual and real worlds in a more natural way. In this paper, the YOLO-v4 deep learning object detection algorithm is used, and the detection of target objects is developed in conjunction with HoloLens2. Experimental results show that the system can more accurately detect the target object and achieve the purpose of virtual and real interaction.**

## Keywords

**HoloLens2; Object Detection; YOLO-v4.**

## 1. Introduction

There are some basic tasks in the field of computer vision: image classification, object detection, example division and meaning division, of which object detection has received widespread attention as the most basic task in computer vision in recent years. Object detection, also known as object extraction, is an image segmentation based on the geometric and statistical features of the target. The segmentation and identification of targets is one, and its accuracy and real-time are important capabilities of the entire system. In recent years, with the rapid development of deep learning technology, new blood has been injected into the detection target and significant breakthroughs have been made. At present, object detection is widely used in automatic operation, robot vision, video surveillance and other fields.

Similarly, deep learning algorithms have formed a relatively complete system through continuous evolution. YOLO was proposed by Redmon et al. in 2015 and is the first single-segment detector in the field of deep learning. The biggest advantage of the YOLO algorithm is the fast processing speed, compared with the previous version, YOLOv2 proposed a joint training algorithm, the basic idea is to use two data sets to train the detector at the same time, the detection detection data set and the classification data set to locate the position of the object on the detection data set, the classification data set to increase the type of object recognized by the detector. YOLOv3 features the introduction of FPNs to achieve multi-scale prediction, and the use of a better underlying network Darknet-53 and binary cross-entropy loss function to achieve a balance of speed and accuracy by changing the network structure of the model. YOLOv4 is a big milestone in the YOLO series, with mAP on the MS-CCOCO dataset reaching 43.5% and a staggering 65FPS speed. Through all-round improvement, performance is so improved. [1].

ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0070

In this paper, we first developed an object detection system that trains network models by executing the Yolo-v4 deep learning object detection algorithm, creating unique data sets, and detecting common objects.After the training, the performance of the model is evaluated. The host can then establish TCP communication with the HoloLens2 device, which can process the objects that need to be detected and return the processed data to the host. This paper is mainly divided into the following parts: the second section introduces the deep learning object detection algorithm, the third section introduces the composition of the target detection system, the fourth section mainly analyzes the target detection results, and the fifth section summarizes the work and puts forward the shortcomings and prospects.

## 2. Object Detection based on the YOLO-v4 Algorithm

Object detection algorithms can be roughly divided into one-step algorithms and two-step algorithms, but the Yolo algorithm is a typical single-step algorithm, and the Yolo algorithm is known to be fast, and compared to other algorithms, Yolo's framework is simple and versatile. The network structure of YOLOv4 mainly includes the following parts: input, backbone , algorithm neck.

### 2.1 The Network Structure of YOLO-v4

#### 2.1.1 Data Entry

At the input, YOLOv4 uses self-adversarial training (SAT) and Mosaic's data augmentation approach to enhance the robustness of the network. With the SAT strategy, the neural network is updated in reverse, trained on perturbed images, and data expansion is achieved. Using Mosaic method to randomly calibrate, crop 4 images, and synthesize 1 image, increasing the small target in the sample and alleviating GPU pressure. [2].

#### 2.1.2 Backbone Network

Based on darknet 53 combined with CSPnet ideas, the CSPDarknet 53 network structure is proposed. The feature map of the basic layer is divided into two parts using the CSP module: one part obtains the residual result by the volume calculation; the other part is not calculated but is hierarchically merged with the residual result obtained in the previous part, which can maintain high accuracy while reducing the amount of computation. Using a tiny Mish activation function, the function image is shown in Figure 1. A slight tolerance for negative values of the Mish function can provide better gradient flow to avoid hard zero boundaries in ReLU, and smooth activation functions allow better information to penetrate deeper into the neural network, resulting in better precision and generalization. [3].

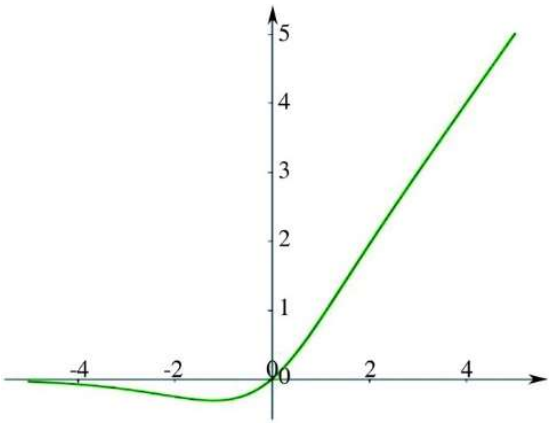


**Figure 1.** Mish activation function

#### 2.1.3 Algorithm Neck

SPP modules already exist in YOLOv3 and continue to be used in YOLOv4. The first sizes used in the SPP module are 1× 1, 5×5, 9×9, 13× maximizing 13 pooled cores and incorporating feature maps

ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0070

of different scales. Rather than simply using the k×k maximum pooling scheme by using the SPP module scheme, the receiving range of trunk features can be increased more effectively, and the most important context features can be significantly separated. The researchers at YOLOv4 used 608 × in the 608 size image test, the module was used in YOLOv4 because the SPP module increased the AP 50 by 2.7% at an additional computational cost of 0.5%. [4].

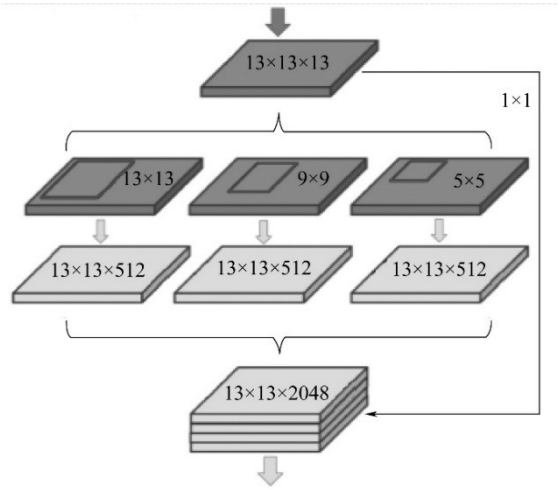


**Figure 2.** SPP module structure

## 2.2 YOLO-v4 Loss Function

The loss function of Yolo-v4 consists of three parts: the confidence loss function, the categorical loss function, and the bounding box regression loss function. [5].

1) Confidence loss function:

When the number of grids is expressed in S, the feature map on each scale is divided into s*s grids, B represents the number of anchor boxes produced by each grid, W indicates whether it is a positive and negative sample, if the sample is a negative sample, W is 0, and vice versa is 1. ^C represents the true reliability of the sample, taking the value of 0 or 1, and C represents the predicted reliability of the sample, whose value is 1 or 0.

$$loss1 = \sum_{i=0}^{s^2} \sum_{j=0}^{B} W_{ij}^{obj} [\hat{C}_i^j \log(C_i^j) + (1 - \hat{C}_i^j) \log(1 - C_i^j)] - \sum_{i=0}^{s^2} \sum_{j=0}^{B} (1 - W_{ij}^{obj}) [\hat{C}_i^j \log(C_i^j) + (1 - \hat{C}_i^j) \log(1 - C_i^j)] \quad (1)$$

2) Classification loss function:

$$loss2 = -\sum_{i=0}^{s^2} \sum_{j=0}^{B} W_{ij}^{obj} \sum_{c=1}^{C} \left[\hat{p}_i^j(c) \log\left(p_i^j(c)\right) - \left(1 - \hat{p}_i^j(c)\right) \log\left(1 - p_i^j(c)\right)\right] \quad (2)$$

Here, ^p^p represents the predicted probability that the object belongs to class c in the jth bounding box of the ith grid, and p represents the true probability.

3) Bounding box regression loss function:

ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0070

Iou is commonly used in object detection to reflect the detection effect of prediction boxes and target boxes. It is defined as:

$$Iou = \frac{|A \cap B|}{|A \cup B|} \tag{3}$$

The regression loss function of the original BBox is replaced by CIou in Yolo-v4. CIou is an improvement over Iou, the formula of which is as follows:

$$CIou = Iou - \frac{\rho^2(b,b^{gt})}{c^2} - \alpha v \tag{4}$$

$$\alpha = \frac{v}{(1-Iou)+v} \tag{5}$$

$$v = \frac{4}{\pi^2}(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h})^2 \tag{6}$$

where b, the center point ρ representing the prediction box and the true box represents the European distance between the center point of the prediction box and the center point of the actual box. c represents the distance diagonal of the minimum bounding $b^{gt}$box covering both the prediction box and the real box, w and h are the width and height of the prediction box, and the width and height of the real box, respectively. α the box height related to the real box and the prediction box, the closer the two are, the smaller the value of αv. Thus, the bounding box regression loss function of Yolo-v4 is $w^{gt}h^{gt}$:

$$loss3 = 1 - Iou + \frac{\rho^2(b,b^{gt})}{c^2} + \alpha v \tag{7}$$

## 3. System Development Environment

### 3.1 PC Environment

The environment required for the design of the YOLO algorithm is completed on the Windows 10 system, including: YOLO-v4, Visual Studio2019, OpenCV, CUDA, CUDNN, etc. Start YOLO-v4 from the computer to process the detected target and return the results to the HoloLens2 device, as shown in Table 1.

**Table 1.** Computer experimental environment

| | |
|---|---|
| Operating System | Windows10 |
| CPU | AMDRyzen74800UwithRadeonGraphics1.80GHz |
| CUDA<br>GPU | V11.2<br>AMDRadeon(TM)Graphics |
| VisualStudio<br>Algorithm | AMDRadeon(TM)Graphics VisualStudio2019<br>YOLO-v4 |

### 3.2 HoloLens2

Develop applications on the HoloLens platform with Unity 3D, use Visual Studio 2019 to equip Unity Project on HoloLens, and analyze the experimental surroundings using HoloLens2. The user can issue commands through gestures and sounds, collect image information in the environment, send it to the

ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0070

computer, and after the computer processes the image information, the various information of the detected object is returned to the device.

**Figure 3.** HoloLens2 device

## 4. Analysis of Target Detection Results

In this paper, the HoloLens2 device obtains a relatively good detection effect on the target object. In order to better judge the detection effect, in the experimental results, the recall rate and the accuracy rate (Precision) are used as the benchmark to determine the quality of the test results, where the recall rate indicates how many of the positive cases in the sample are predicted correctly; the accuracy rate indicates how many of the samples predicted as positive are true positive samples. Detection accuracy and recall are defined as follows: [6][9].

$$P = \frac{Z_{TP}}{Z_{TP} + Z_{FP}} \tag{8}$$

$$R = \frac{Z_{TP}}{Z_{TP} + Z_{FN}} \tag{9}$$

where ,$Z_{TP}$ is the correct dectection target,$Z_{FN}$ is the detection target of the missed test,$Z_{FP}$ is the wrong dectection target.

The targets of this training are boxes, cups, mice, chairs, table lamps and other commonly used items in life, and the recall rate and accuracy rate of each target object are shown in Table 2.

**Table 2.** The recall and precision of each target object

| The target object | Recall | Precision |
|---|---|---|
| box | 74.2% | 97.2% |
| cup | 72.1% | 98.7% |
| mouse | 63.5% | 98.2% |
| chair | 73.3% | 96.7% |
| table lamp | 75.6% | 92.8% |

According to the analysis of the recall rate and accuracy rate of each target object obtained above, it can be seen that the detection effect of the box, cup and chair of the target detection is better, followed by the lamp and mouse.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0070

## 5. Conclusion

Although this paper implements HoloLens-based object detection, there are still many problems:

(1) HoloLens2 camera stabilization function is poor, the user wears on the head to take photos is easy to blur, making the photo unusable.

(2) HoloLens2's current computing power is limited, but uploading photos accounts for the computation, which can easily lead to cartoons on the device and affect the user's sense of experience.

(3) The model trained using the Yolo-v4 algorithm is not effective enough for the detection of occluded or overlapping objects and small target objects, and sometimes missed detection.

Compared to the HoloLens generation, the HoloLens 2 improves wearing comfort, enhances immersion, eye tracking capabilities, multiple gestures, integrates advanced AI technology, and accelerates the realization of business value. The future development prospects of HoloLens are very broad.

## References

[1] XIE Fu,ZHU Dingju. A Review of Deep Learning Object Detection Methods[J].Computer System Applications,2022,31(2):1-12.

[2] LI Weigang,YANG Chao,JIANG Lin,ZHAO Yuntao. Indoor Scene Object Detection Based on Improved YOLOv4 Algorithm[J/OL].Advances in Laser and Optoelectronics:1-19[2022-04-23].

[3] YU Peidong,WANG Xin,JIANG Gangwu,LIU Jianhui,XU Baiqi. A Typical Object Detection Algorithm for Remote Sensing Images With Improved YOLOv4[J].Journal of Surveying and Mapping Science and Technology,2021,38(03):280-286.

[4] SHI Ruipeng,JIANG Dani,FANG Qing. Aircraft Target Detection based on YOLOv4 remote sensing image[J].Bulletin of Surveying and Mapping,2021(S1):134-138.

[5] CHEN Qian. Real-time object detection in the workshop based on YOLOv4[J].Modern Computer, 2021 (16):164-168.

[6] LIU Yangfan,CAO Lihua,LI Ning,ZHANG Yunfeng. Spatial infrared weak object detection based on YOLOv4[J].Liquid Crystals and Displays,2021,36(4):615-623.

[7] ZHOU Hang. Object detection and localization based on HoloLens[D].Huazhong University of Science and Technology, 2019.

[8] GONG Chibing. Exploring the New Characteristics of HoloLens 2 Mixed Reality[J].Modern Information Technology,2020,4(02):121-123.

[9] Zeng laughed. Research on Object Detection Technology Based on HoloLens2[J].Modern Computer, 2021(14):92-95.