Detection of the autonomous car robot using Yolo

F.I. Abd-AL Sahib, H. B. Taher, R F. Ghani

1University of Thi-Qar – College of Education for Pure Sciences, Thi-Qar, Iraq.

edpcompfatima@utq.edu.iq  Hazeem.sci@utq.edu.iq  110016@uotechnology.edu.iq

Abstract. One of the important object detection applications in smart transportation systems is vehicle detection. Working on self-driving car robots has become an important experiment in recent years to take advantage of innovations and ideas in real self-driving cars, and the detection of robots by multiple algorithms is the most important phase in this work. To solve the problems of self-driving car robot detection. Such as not recognizing shape. In this paper, via the Yolov2 algorithm, we trained a new model for robots. It was proven with the comparison experiments that the proposed method is successful for robot detection. In addition, the proposed model demonstrated excellent feature extraction ability with network visualization.

1. Introduction

Computer vision is a tool for making machines "look". This allows computers and other devices to take the place of human eyes to recognize objects within the scene and further the processing of images. The new research trend for using artificial intelligence systems to extract data from images is image processing and computer vision. A significant branch of computer vision is object detection. Object detection is efficient when it is fast and accurate. Good object detection algorithms should be convenient for the lives of individuals. The purpose of this paper is to classify and locate self-driving car robots. It is important in the autonomous driving system to detect self-driving cars and their location. A self-driving car, also known as an autonomous vehicle (AV or auto), driverless car, or robot-car is a vehicle that is capable of sensing its environment and moving safely with little or no human input. Object detection is a very important part of artificial intelligence. Which is a forward-feedback neural network. It has a unique superiority in object recognition with its special structure of local weights sharing. The interest in having deeper hidden layers has begun to surpass the performance of classical methods in various fields, particularly in pattern recognition. The Convolutional Neural Network is one of the most popular deep neural networks (CNN). It takes its name from the mathematical linear operation called convolution between matrices. CNN has several layers [1], including the convolution layer, pooling layer and fully-connected layer. In machine learning issues, CNN has outstanding results. Especially those applications dealing with image data. The traditional detectors [3] first extract suggestions and extract features, then do the classification [2]. It is proposed that location and classification in a single CNN be implemented. We want a real-time object detection algorithm in the self-driving system, so we chose you only look once (YOLO) as a guide to constructing a better method where More than twice the average accuracy of other real-time systems is reached by YOLO. To construct our own training set and testing set, we select some samples and label some samples manually. YOLO algorithm is an algorithm based on regression, instead of selecting the interesting part of an Image, it predicts classes and bounding boxes for the whole image in one run of the Algorithm [4]. The work discussed in this paper forms part of an automated monitoring system for vehicles. The main objective is to train a system to classify a new class. Which self-driving cars robot define and monitor the behavior of autonomous robots. In this way, it is possible to track important events that took place on the way. The yolo detection system. Figure 1 shows how YOLO is refreshingly basic.
2. The proposed method

In Figure 2 Block diagram of the proposed system.

2.1. Data collection and prepare data

The images were collected by the iPhone 8 Plus camera, and the total number of images was 1235 the self-driving car robot, and since the system is under supervision depend on human work, Manual labeling for supervised learning, we use LabelImg to label images. LabelImg is available on mac, windows, and UNIX we do the labeling process to determine the Center X, y in the image, width, height and it belongs to any class. Within the process of pre-processing the data, missing labels for some image we removed them.

2.2. Yolo algorithm

YOLO divides the image into SxS grids and predicts B bounding box and C class probability for each grid cell. Each bounding box consists of five predictions: w, h, x, y, and object confidence. The values of w and h represent the width and height of the box relative to the whole image. The values of (x, y) represent the center coordinates of the box relative to the bounds of the grid cell. The object confidence represents the reliability of existing object in the box to improve the YOLO prediction accuracy, Redmon et al. proposed a new version YOLOv2 in 2017. A new network structure Darknet-19 was designed by removing the full connection layers of the network, and batch normalization was applied to each layer. Referring to the anchor mechanism of Faster R-CNN, k-means clustering was used to obtain the anchor boxes. In addition, the predicted boxes were retrained with direct prediction. Compared with YOLO, YOLOv2 greatly improves the accuracy and speed of object detection. The Self-driving car robots dataset was divided into a training dataset and validation dataset with the ratio of 8:2, where the numbers of images in the training dataset 988 and numbers of images in the testing dataset 247. in this paper we used tensor flow lite keras [5]. TensorFlow [6] Lite is designed to run machine learning models with just a few kilobytes of memory on microcontrollers and other devices. The core runtime is only 16 KB and several basic models can be run. Keras is an API designed not for machines but for human beings. Keras follows good cognitive load reduction practices: it provides consistent and easy APIs, minimizes the number of user actions needed for common use cases, and provides simple and actionable error messages. An algorithm is a type of Supervised machine learning.
2.2.1. Loss function

Different object sizes had different effects on the entire model when training YOLOv2, which resulted in greater errors for larger objects than for smaller objects. The loss estimate for the width and height of the bounding boxes was improved using normalization in order to reduce this effect. The enhanced loss feature is seen in Equation 1 [9].
2.2.2. Model Architecture

The propose system detection network has 23 convolutional layers as shown in 3. The image size is 512 * 512. The images enter these layers and the features are extracted from them through the filters in each layer, where each layer has specific filters to extract the features starting from the lower level, middle level, and higher level. The edge might be detected in the first layers and then the simpler shapes in the second layers, and then the higher-level features such as Self-driving robot. The total of the filters in the first layer is (32) filters with dimensions of (3 * 3) and strides (1x1), and in order to preserve the shape of the image, the process padding is done by adding zeros around the image. After that cancel the bias and adding BatchNormalization to speed up the work. The type of activation function is Relu. Sigmoid and Tanh have been the most common non-linearity for several years. However, for the following reasons, the Rectified Linear Unit (ReLU) [1] has been used more commonly recently

1 - For both function and gradient, ReLU has simpler definition.
2 - The saturated function, like sigmoid and tanh, causes backpropagation problems. The gradient signal starts to disappear, which is called the “vanishing gradient” as the neural network architecture is deeper. This occurs because virtually anywhere but the middle, the gradient of those functions is very close to zero. The ReLU has a constant gradient for the positive input, however. Although the feature is not distinguishable, it can be overlooked in the actual implementation phase.

3 - A sparser representation produces the ReLU. The zero in the gradient leads to a complete zero being obtained. Sigmoid and tanh, however, still have non-zero gradient outcomes, which may not be in favor of training for training.

Figure 3. Layers of Yolov2 – self – driving car robot
Pooling: The main idea of pooling is to reduce samples to reduce complexity in layers. One of the most popular types of assembly is max-pooling, where you take an m* m patch and pass a filter over it that takes the largest value between the values. One of the most common sizes used in max-pooling is 2×2 [7-10]. As can be seen in Fig 4.

2.2.3 Non-max suppression

The objects in the image can be of various sizes and shapes, and to perfectly capture each of these, several border boxes are created by object detection algorithms. (Image on the left). Ideally, we must have a single bounding box for each object in the image. Something like that image to the right. To choose the best bounding box, from the multiple bounding boxes, predicted Non-max suppression is used in these object detection algorithms. This technique is used to "suppress" and hold only the best boxes that are less likely to be bound.

3. Implementation Results

In this paper, we have presented an algorithm for self-driving car robot detection and tracking and were score 82%, and a new class was called a self-driving car as in figure 5. This class can be used in autonomous automobile robotics projects for autonomous vehicle path control or collision control. The first table shows the result after detection the class. The execution time of the object detection is real time. To calculate (F measure) we need Precision and Recall:
i. Precision measures how accurate your predictions are. i.e. the percentage of your predictions are correct.

ii. Recall measures how well you find all the positives

\[ F \text{ measure} = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \]

\[ F \text{ measure} = \frac{\text{TP}}{\text{TP} + 1/2(\text{FP} + \text{FN})} \]

\[ F \text{ measure} = \frac{202.54}{202.54 + 1/2(44.46)} = 0.90 \]

Where TP Means number of true positives, FN mean number of false negatives. FP mean number of false positives.

The values taken in the equation are the set of test. In measuring object detection FP is sometimes canceled.

Figure 5. Detection results of YOLOv2_Self-driving car
4. Conclusion and Future Work

In this paper, a model called the YOLOv2_Self-driving car was proposed for vehicle detection by YOLOv2. The resulting system is both efficient and accurate based on the experimental results, the map of YOLOv2_Self-driving car could reach 82%. Therefore, the proposed network is effective for Self-driving car robot detection. Although the model proposed in this paper achieved ideal experimental results, the number of images amounts of data is relatively low. In future work, we will collect more actual data to further study how to improve the accuracy and speed of Self-driving car detection.

5. References

[1] Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017, August). Understanding of a convolutional neural network. International Conference on Engineering and Technology (ICET) (pp. 1-6). IEEE, 2017

[2] Felzenszwalb, P., McAllester, D., & Ramanan, D. (2008, June). A discriminatively trained, multiscale, deformable part model. In 2008 IEEE conference on computer vision and pattern recognition (pp. 1-8). IEEE.

[3] Felzenszwalb, P. F., Girshick, R. B., McAllester, D., & Ramanan, D. (2009). Object detection with discriminatively trained part-based models. IEEE transactions on pattern analysis and machine intelligence, 32(9), 1627-1645.

[4] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

[5] https://keras.io/

[6] https://www.tensorflow.org/install

[7] Scherer, D., Müller, A., & Behnke, S. (2010, September). Evaluation of pooling operations in convolutional architectures for object recognition. In International conference on artificial neural networks (pp. 92-101). Springer, Berlin, Heidelberg.

[8] Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering, 160(1), 3-24.

[9] Sang, J., Wu, Z., Guo, P., Hu, H., Xiang, H., Zhang, Q., & Cai, B. (2018). An improved YOLOv2 for vehicle detection. Sensors, 18(12), 4272.

[10] T. A. Al-asadi and A. J. Obaid, "Object detection and recognition by using enhanced speeded up robust feature," International Journal of Computer Science and Network Security (IJCSNS), vol. 16, no. 4, pp. 66-71, 2016.