The Application of Human Detection Based on YOLOv5

Chengcheng Hou *

CBIS Bilingual School, Foshan City, Guangdong Province, China

* Corresponding Author Email: jason_hou@cbis-stu.com.cn

Abstract. Human detection has several possible applications. A well-known example is autonomous driving technology, which incorporates gender recognition and fall detection for the elderly. This study will focus on the principles of these applications, which consist primarily of locating the positions of persons inside an image. It is a crucial piece of technology that protects individuals from danger. To fix this problem, this paper employed the deep learning algorithm, mainly supervised learning, which demands a significant quantity of labeled data. Fortunately, these data sets may be collected via security and in-vehicle cameras. YOLOv5 is a famous object detection algorithm, it can train the detector to operate faster and more precisely, enabling the system to identify any potential dangers instantly, which is considered in this study. The experimental results demonstrated that the employed YOLOv5 method can detect the location of human effectively, which proved the feasibility of the method used in the study.

Keywords: Machine Learning, Deep Learning, YOLO algorithm, Human Detection.

1. Introduction

Artificial Intelligence (AI) is an essential branch of computer science [1]. In the simplest terms, it refers to systems and programs that resemble human intellect in order to carry out tasks and have the ability to improve themselves repeatedly based on the data they gather [2]. Despite AI being a branch, it also has many branches below it. Machine Learning, Deep Learning, Robotics, and the list goes on [3]. Of all these branches, Machine Learning and Deep Learning are the most widely used ones, with countless applications. One of the most popular uses of machine learning is identifying things that meet the needs of particular persons. These technologies analyze still photographs, moving images, and audio to recognize the objects or phrases included, regardless of their language. In addition, it can finish the data integration process by generating linkages across diverse data sets as a result of evaluating and processing a large quantity of accessible data. However, there are still technologies based on these studies that are not fully developed, like autonomous piloting for vehicles.

However, people usually confuse object classification and object detection, which are different. Object detection is a bit more complex than object classification, where this paper tries to localize and recognize the object in images [4]. When people use to do object recognitions, they use the Convolutional Neural Network (CNN) algorithm to do so. The revelation that a convolutional neural network could be used to extract higher gradually- and higher-level representations of the visual content marked a significant advancement in constructing models for image categorization. A CNN uses the raw pixel data from the image as input to "learn" how to extract features like textures and forms and, in the end, determine what item they make up rather than preprocessing the data to create these features [5]. In May 2020, Evan et al. wrote an article about car classification implementing CNN [6]. It is a very successful model that can classify different types and brands of cars, but it is not enough if people hope to achieve autonomous driving. Object detection is more critical than object classification to locate cars or other obstacles.

On the other hand, YOLO is a much better algorithm to use when it comes to object detection. YOLO, standing for You Only Look Once, provides an extremely fast and accurate object detection algorithm that detects humans in real-time. It has many real-life applications, particularly for autonomous driving, vehicle detection, and intelligent video analytics [7]. There are many versions of YOLO, YOLOv2, YOLO9000, YOLOv4, and YOLOv5. Higher versions of YOLO provide quicker, more accurate detection, easier to operate the system, and can detect higher frames per second. An AI engineer, Bhavesh Singh Bisht, made a car detection system using YOLO [8].

However, the problem is that YOLO is the first version. It is way slower and less accurate than the newest versions.

The purpose is clearly to create a system that can detect objects using better algorithms (i.e. YOLOv5). The aim of this study is to create a system that could recognize humans. One reason is that another person already did that vehicle detection. Another reason is for humans to achieve autonomous driving, many functions such as learning the maps, remembering the local traffic system, and, most importantly, predicting and detecting potential accidents are essential. Nevertheless, one of the most significant "obstacles" that need to be considered is humans, which is why this study designed the system to be able to detect humans. The YOLOv5 algorithm was used to design the human detecting system, as YOLOv5 can not just identify the existence of humans but also locate the human location within the picture, and it is one of the latest official versions of YOLO. The result of the system had over 93% accuracy with the test dataset.

2. Methodology

2.1. Dataset preparation

The dataset used was found from various sites, mainly raw images taken from CCTV. The platform called Kaggle was used to collect data, which made the process much easier. Kaggle is a Google Limited Liability Company subsidiary that operates an online community for data scientists and machine learning practitioners. It enables users to discover and share data sets, study and construct models in a web-based data-science environment, collaborate with other data scientists and machine learning experts, and compete to solve data science problems. Figure 1 presents a part of data.

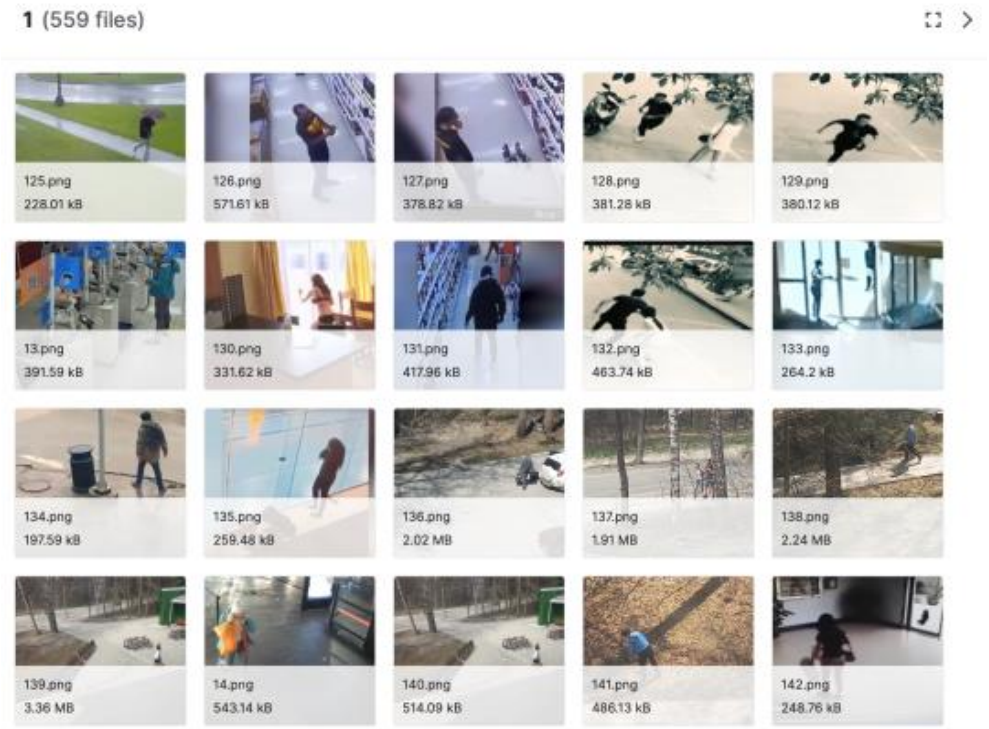

Figure 1. The sample data.

The majority of the images of people that this paperutilized originated from Kaggle, where a range of users uploaded them. At the start of 2022, Constantin Werner [9], who supplied nearly 2,000 images, was the first to submit it. Another one of our primary sources was a 2022 upload dataset by Meet Nagadia that had approximately 15,000 pictures [10]. Due to time restrictions, this paperwere limited to around one thousand photos, which this paperrandomly divided into eighty percent training data and twenty percent test data. This paperdefined the sorts of photographs this paperwanted, so

this paperwent through numerous datasets and filtered out the excellent images to create our data sets. The photos were then divided into two folders, the training dataset, and the test dataset, in a 4:1 ratio.

Makesense is the labeling website this paperutilized, and this papermade male and female labels. Figure 2 shows the example of annotated result. This papercategorized all the persons in the photo within a rectangle box based on these two labels. To avoid the machine from making erroneous conclusions, this paperhave opted only to preserve a record of persons who are marginally blocked. Using Makesense, these labeled data will be converted into txt files, which this papercan later customize to become our actual training data. In this dataset, class id 0 corresponds to the "man" class, whereas class id 1 corresponds to the "female" class. The following float values represent the bounding box's x, y, w, and h coordinates. As can be seen, these coordinates have been standardized between 0 and 1.

Figure 2. Example of annotated result from makesense

2.2. YOLO Algorithm

After annotating the data using Makesense, the next step is to build up the system using YOLOv5. YOLOv5 has numerous advantages. For starters, it is fast and accurate, making it one of the most well-known object identification algorithms. Furthermore, because YOLOv5 was first built in PyTorch, it benefits from the established PyTorch ecosystem with easy deployments and maintenance. However, pictures must often be tagged when employing deep learning for target recognition. Labeling is the image annotation tool this paperuse in general. Although the images tagged by this tool are saved in XML* format, Yolov5 only takes text files. The next step is to create the environment.

YOLOv5 is a convolutional neural network (CNN) that detects objects with high precision in real-time. This technique uses a single neural network to analyze the entire image, separate it into its component parts, and provide predictions regarding each component's bounding boxes and probabilities. A weight is assigned to these bounding boxes depending on the calculated probability. It is referred to as the "you only look once" strategy since it only provides predictions after a single forward propagation runs through the neural network. After non-max suppression, the found items are subsequently distributed (which ensures that the object detection algorithm only identifies each object once).

YOLOv4 debuted shortly after YOLOv5, which is based on the YOLOv4 algorithm. YOLOv5 experiences additional upgrades, as well as enhancements to its detection performance. It has not yet been established whether or not the YOLOv5 algorithm performs better than the YOLOv4 algorithm. It indicates that YOLOv5 has a positive influence on testing the COCO dataset. People have several doubts about how innovative the YOLOv5 algorithm will be.

Some people hold a favorable perception of it, while others hold a pessimistic attitude. The YOLOv5 detection system has tremendous opportunities for improvement, in our opinion. Despite their apparent simplicity and lack of originality, these enhancement recommendations can improve the performance of detection algorithms. The industry prefers to employ these tactics over a complex algorithm when attempting to achieve high detection accuracy. After reading this article, which will offer a comprehensive description of the new and improved ideas provided in the YOLOv5 detection method, this papermay attempt to apply these principles to other target identification methods.

3. Results and discussion

3.1. Raw results from the system

Figures 4, 5, and 6 illustrate the samples of training results. These numbers reflect a portion of the information was collected during the data collection stage. Overall, the following table depicts, for each of the three outcomes, the proportion of our test data attributable to that specific outcome.

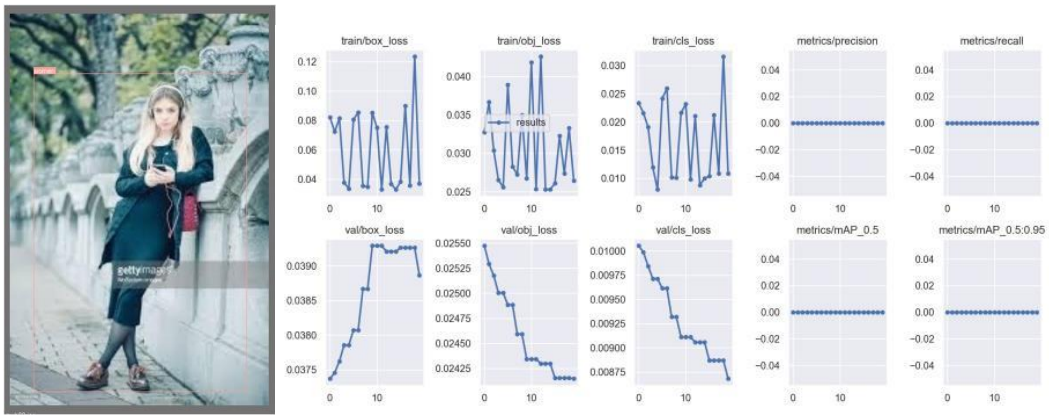

Figure 3. The result of example 1

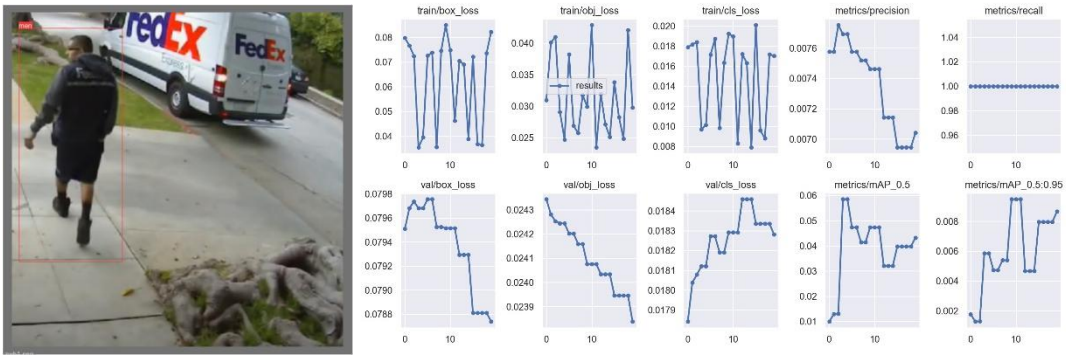

Figure 4. The result example 2

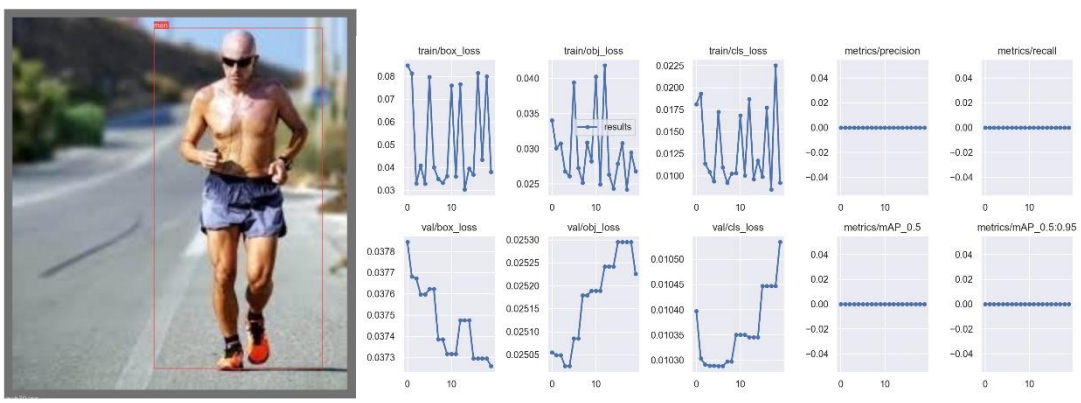

Figure 5. The result of example 3

The Successful tests in the following Table 1 signify that they were successful, indicating that the correct number of individuals were recognized and that the genders were distinguished correctly. The second one shows that the genders are incorrect. Figure 4-6 present the result of examples. However, the locations of the people in the photo are correct. The Misjudgments one suggests that non-human entities have been recognized as human, or more particularly, that the apparent presence of a particular individual has been disregarded.

Table 1. Testing result

Type	Amount	Percentage
Successful Tests	185	92.5
Faliure Tests	4	2
Misjudgements	11	5.5

The total success rate of our tests is relatively high, reaching 92.5 percent, with a failure rate of 5.5 percent and an error rate of 2 percent.

As a result of the failed test, this paperdetermined three fundamental causes for incorrect identifications. The first is anything with human-like characteristics, such as the same structure, color, appearance as humans, etc. For example, mannequins, scarecrows, or signboards are usually to fault for critical mistakes. The second case is when components are lacking, such as when humans lack limbs, legs, or heads, and the system fails to recognize them as humans. The third one is reserved only for gender-related mistakes. The bulk of mistakes, according to the results of our studies, come when males have long hair and females have short haircuts, causing the system to make a wrong decision.

4. Conclusions

The overall test result is good, it has a very high successful rate, yet there are still room for improvement. In order to keep these mistakes lower, there are things that this paperneed to do in the future to improve. The first thing is to provide more training. Despite that, more training can cause more biased examples, but it will lower the test errors. Secondly, this paperneed to use higher versions of YOLO. As this paperwere doing the project, this paperwere suggested by the professor to use YOLOv7, which is much quicker and more accurate, but this paperdid not use it as this paperwere almost done with the project. So using better versions of YOLO can increase the systems' accuracy. Last, this paperneed to provide more accurate training or add more variety to our training data. For example, this papercan outline the humans rather than label them using a rectangle inside the picture.

References

[1] Marketbusinessnews, How is Machine Learning Changing Digital Education? [R]. https://marketbusinessnews.com/how-is-machine-learning-changing-digital-education/303334/

[2] Oracle, What is Artificial Intelligence (AI)? [R], https://www.oracle.com/artificial-intelligence/what-is-ai/

[3] What are the types of artificial intelligence? [R]: Branches of ai. Edureka. (2021, July 29). Retrieved August 24, 2022

[4] Aktaş, Y. C. Object detection with Convolutional Neural Networks. Medium [R]. Retrieved August 25, 2022, from https://towardsdatascience.com/object-detection-with-convolutional-neural-networks-c9d729eedc18

[5] Google. (n.d.). ML Practicum: Image Classification | machine learning | google developers [R]. Google. Retrieved August 25, 2022, from https://developers.google.com/machine-learning/practica/image-classification/convolutional-neural-networks

[6] Evan E. and Henning K. et al. How to use a CNN to successfully classify car images [R]. Databricks. Retrieved August 25, 2022, from https://www.databricks.com/blog/2020/05/14/a-convolutional-neural-network-implementation-for-car-classification.html

[7] Yolo: Real-time object detection explained. V7. (n.d.) [R]. Retrieved August 25, 2022, from https://www.v7labs.com/blog/yolo-object-detection

[8] Bisht B. S. Object detection using Yolo and car detection implementation [R]. Medium. Retrieved August 25, 2022, from https://medium.com/analytics-vidhya/object-detection-using-yolo-and-car-detection-implementation-1ec79e882875

[9] Geoffrey H. Deep Learning [J]. In: Nature 521.7553 (2015), pp. 436–444.

[10] Meet Nagadia. Human Action Recognition (HAR) Dataset [R]. December 25, 2019. url: https ://www.kaggle.com/datasets/dasmehdixtr/human-action-recognition-dataset.