CAN THO UNIVERSITY
COLLEGE OF INFORMATION AND COMMUNICATION TECHNOLOGY

BASIC-THESIS
(HIGH-QUALITY PROGRAM)

BUILDING OBJECT DETECTION MODEL WITH YOLOV5

Student: Nguyen Thi My Khanh
Student ID: B1910657
Class: 2019-2023 (K45)
Advisor: Dr. Tran Cong An

Can Tho, 04/2022

CAN THO UNIVERSITY
COLLEGE OF INFORMATION AND COMMUNICATION TECHNOLOGY
DEPARTMENT OF INFORMATION TECHNOLOGY

BASIC-THESIS
(HIGH-QUALITY PROGRAM)

BUILDING OBJECT DETECTION MODEL WITH YOLOV5

Student: Nguyen Thi My Khanh
Student ID: B1910657
Class: 2019-2023 (K45)
Advisor: Dr. Tran Cong An

Can Tho, 04/2022

ACKNOWLEDGMENTS

First of all, I would like to express my deep gratitude to Dr. Tran Cong An, who directly guided and oriented me to complete this graduation thesis. He spent time and created favorable conditions in all aspects so that I could complete the topic.

Can Tho, May …, 2022
Student

Nguyen Thi My Khanh

TABLE OF CONTENT

LIST OF FIGURES

LIST OF ABBREVIATIONS

No.	Abbreviation	Origin word
1	YOLO	You Only Look Once
2	AI	Artificial Intelligence
3	CV	Computer Vision
4	CNN	Convolutional Neural Network
5	CSP	Cross Stage Partial Networks

ABSTRACT

Object detect is a popular topic of deep learning because of its high applicability, easy-to-search data, and widely applied results. Object detection algorithms like Yolo have high speed and accuracy, making detection done in almost real time.

Besides that , the development of neural networks is gradually realizing the concept we often call *Computer Vision*. YOLOv5 has built an approach that makes Object detection really possible in life.

Therefore, this basic yearbook will build an object recognition model using YOLOv5 on google colaboratory, with a training dataset of COCO128. The goal is that upon completion, the model will successfully detect the data.

CHAPTER 1: INTRODUCTION

1.1. Statement of the problem

Nowadays, Artificial Intelligence (AI) is increasingly popular and contributes to profoundly changing many aspects of life. In particular, Computer Vision (CV) is an important area of AI that includes methods of collecting, processing images, analyzing and recognizing images.

Deep learning Network is the study of algorithms, computer programs for computers to learn and make predictions like humans. It is applied in many different fields such as science, engineering, other fields of life as well as the applications of classification and object detection. A typical example is CNN (Convolutional Neural Network) applied to automatic recognition, learn distinguishing patterns from images by successively stacking layers on top of each other, and in many applications CNN is considered a powerful image classifier and leverages technologies in the field of computer vision, driving the development of machine learning. But besides that, to classify an object, CNN technology consumes a lot of resources such as bandwidth, memory and processing capacity of hardware.

In order to reduce the above consumptive resources, more and more algorithms and models over time have emerged, and the typical model is the YOLOv5 model.

1.2. Related work

Currently, there are many YOLO models of object detect that have been conducted and succeeded, such as: YoloV4 model for vehicle identification and classification [1], YoloV5 model is also used to help automatic identification. The growth stages of cantaloupe plants during growth in greenhouses [2].

ứng coi như bị mất. Tệp CSV trung gian sẽ được hình thành trong đó chứa tọa độ của các hộp giới hạn được phân tích cú pháp để áp dụng Sort trên các đối tượng được phát hiện.

3.3. Phân loại và đếm các phương tiện lưu thông

Các phương tiện lúc này đã được gán ID riêng và cố định, nhóm nghiên cứu tiến hành kiểm tra phương tiện qua 5 khung hình liên tiếp. Quá trình được lặp lại cho tới khi kết quả trả về là đúng thì ID của phương tiện sẽ được thêm vào danh mục kết quả.

Vì thuật toán Sort chỉ định một ID duy nhất bằng cách lấy thông tin hộp giới hạn làm đầu vào, tức là các nhãn lớp được tạo ra bởi phát hiện đối tượng. Để khắc phục hạn chế này, một danh mục được tạo ra bằng cách duy trì ánh xạ các ID tạo bởi Sort tới các nhãn lớp tương ứng từ Yolov4. Một danh sách các ID duy nhất hiện đang được sử dụng và duy trì. Đối với mỗi khung sẽ tìm ra ID nào đã bị loại bỏ giữa các khung liên tiếp bằng cách so sánh danh sách ID duy nhất hiện tại và trước đó. Sau đó, các ID được tìm thấy ngoài danh sách sẽ được sử dụng để xóa các khóa tương ứng khỏi danh mục kết quả vì chúng không còn phù hợp nữa, từ đó xác định được số lượng từng loại xe trong khoảng thời gian cụ thể tương với với ID trong danh mục kết quả và nhãn lớp trong danh mục đối tượng. Quá trình được lặp lại cho tới khi chương trình kết thúc.

Hình 4.3: Nhận diện với mật độ lưu thông cao

Hình 4.4: Đếm các phương tiện lưu thông

Hình 4.4 mô tả thống kê các số lượng các loại phương tiện tương ứng. Với mỗi mật độ lưu thông khác nhau thì độ chính xác của thuật toán với các loại phương tiện là khác nhau. Tỉ lệ đô chính xác tương ứng được thể hiện trong *Bảng 4.1*. Kết quả

Figure 1. The YoloV4 model used to recognize and the type of the media media

Application of deep learning model in recognition of growth stages of *Cucumis melo* L. in greenhouse

Hoang Anh Tuan Dang[1*], Minh Thang Nguyen[2]

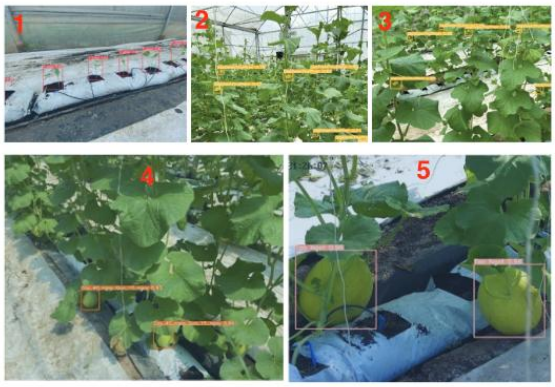
[1]*National Center for Technological Progress*
[2]*Vietnam - Korea Institute of Science and Technology*

Received 11 August 2021; accepted 16 September 2021

Abstract:

Despite the increasing application of deep learning (DL) models in various socioeconomics such as financial analysis and forecast, intelligent transport, self-driving, disease diagnosis, the effective use of this technology to support agricultural cultivation is still limited. This paper introduces the implementation of the lightest and state-of-the-art YOLOv5 architecture for automatic recognising of important growth stages of *Cucumis melo* L. from the camera images collected in the greenhouse. This image identification initiative achieved an average accuracy of 96% F1-score in the identification of the five growth stages of *Cucumis melo* L. using a limited set of training and testing data (total 2,818 images of *Cucumis melo* L.). These preliminary results lead to the conclusion that the YOLOv5 object detection and

ảnh trống trong tập dữ liệu này). Thông qua việc nghiên cứu tài liệu và kinh nghiệm canh tác dưa lưới của Trang trại Linh An (Giang Biên, Long Biên, Hà Nội), chúng tôi xác định 5 giai đoạn sinh trưởng, phát triển quan trọng của cây dưa lưới trồng trong nhà màng cần được nhận biết gồm: 7 ngày; thụ phấn; đã thụ phấn; cây được 40 ngày và quả được 15 ngày; tạo ngọt. Với việc nhận biết chính xác các giai đoạn phát triển quan trọng này sẽ giúp người trồng có những chế độ canh tác phù hợp với nhu cầu sinh trưởng của cây.



Hình 1. Các giai đoạn đã nhận diện thành công. 1: giai đoạn 7 ngày; 2: giai đoạn thụ phấn; 3: giai đoạn đã thụ phấn; 4: giai đoạn cây được 40 ngày và quả được 15 ngày; 5: giai đoạn tạo ngọt.

Tương đương với 5 giai đoạn sinh trưởng, phát triển chính nêu trên, chúng tôi có 5 lớp nhận diện trong tập dữ

Figure 2. The YoloV5 model identifies the growth stages of cantaloupe plants during growth in the greenhouse

1.3. The purpose of the topic

The goal of the project is to build an object recognition model, and allow to detect some objects in the data set.

1.4. Object and scope research

❖ Research object:

The object of the study is the Yolov5 model.

❖ Research scope:

Approximately 80 objects are identified in the dataset.

1.5. Research approach and methods

❖ General approach

Based on references. Then proceed to build and test the model on google colaboratory.

❖ Research methods

- Research theory from journals, articles, websites in related fields.
- Implement and evaluate the results achieved.

1.6. Topic outline

- Chapter 1: Introduction - overview of the year thesis.
- Chapter 2: Literature review - describes the theory used in the thesis.
- Chapter 3: Solution design and implementation.
- Chapter 4: Test Evaluation
- Chapter 5: Conclusion of achieved results and development direction.

CHAPTER 2: LITERATURE REVIEW

2.1. YoLo overview

YOLO (You Only Look Once) is an algorithm of the CNN network model for object detection and classification. Yolo is created from the combination of convolutional layers and connected layers. In which, the convolutional layers will extract the features of the image, while the connected layers will predict the probability and coordinates of the object.

In terms of accuracy, YOLO may not be the best algorithm but it is the fastest in the class of object detection models. It can achieve almost real-time speed, but the accuracy is no less than that of the top models.

Up to now, YOLO has 5 versions including v1, v2, v3, v4, v5. In which v5 is the latest version, improved in speed, accuracy and overcome the disadvantages of previous versions such as error in determining the position of objects, bounding box space constraints...[3]

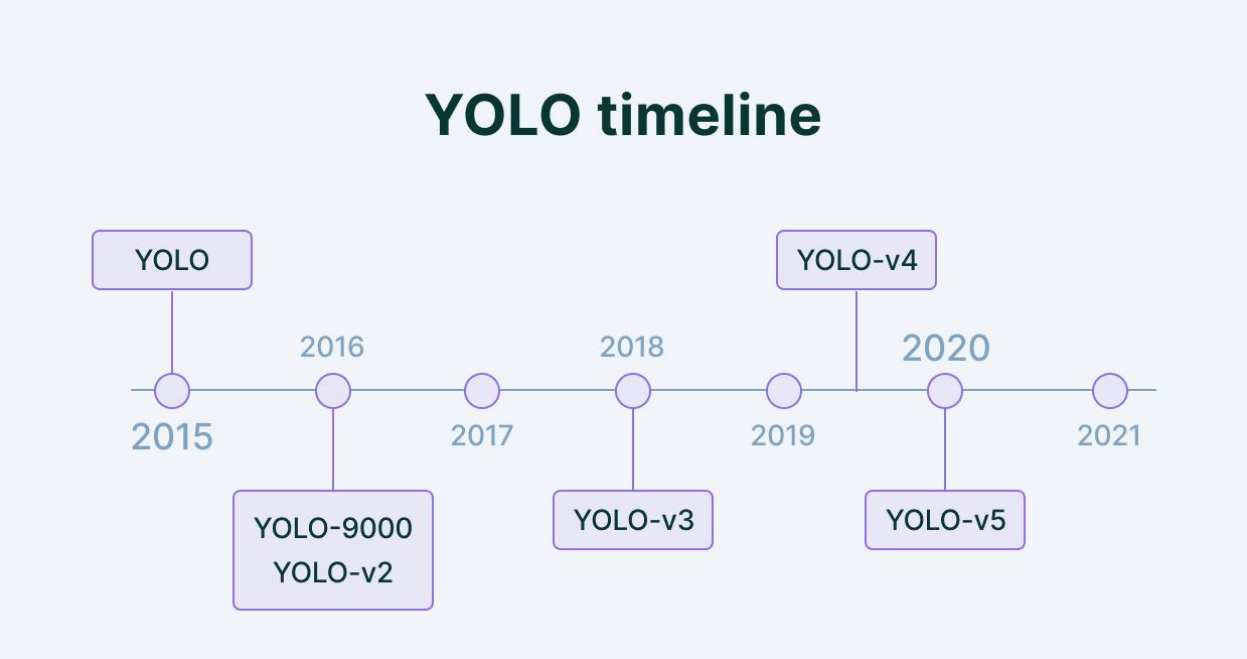


Figure 3. Yolo timeline

2.2. Yolov5 model

YOLOv5 is a recent release of the YOLO family of models, it detects the object by dividing the image into a grid system. Each cell in the grid is responsible for detecting objects in itself. YOLOv5 is the first of the YOLO models to be written in the PyTorch framework and it is much more lightweight and easy to use. Compared to other object detection frameworks, YOLOv5 is extremely easy to use for a developer implementing computer vision technologies into an application.

The advantages of YOLOv5 model: [4]

- Easy Install: YOLOv5 only requires the installation of torch and some lightweight python libraries.
- Fast Training: The YOLOv5 models train extremely quickly which helps cut down on experimentation costs as you build your model.
- Inference Ports that work: You can infer with YOLOv5 on individual images, batch images, video feeds, or webcam ports.
- Intuitive Layout: File folder layout is intuitive and easy to navigate while developing
- Easy Translation to Mobile: You can easily translate YOLOv5 from PyTorch weights to ONXX weights to CoreML to IOS.

YOLOv5 offers 5 versions with different network architectures:

+ YOLOv5n: nano version

+ YOLOv5s: small version

+ YOLOv5m: medium version

+ YOLOv5l: large version

+ YOLOv5x: extra-large version

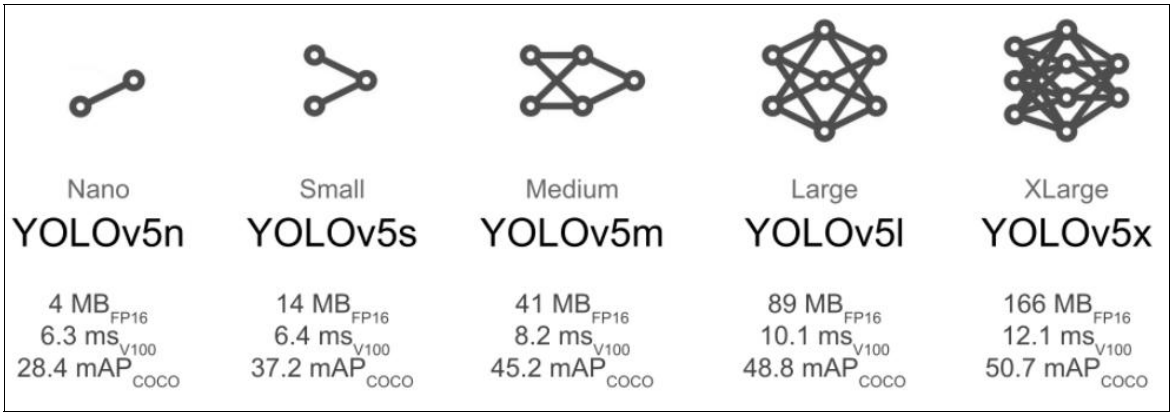


Figure 4. YoLov5 Network Architecture

2.3. Yolov5 architecture

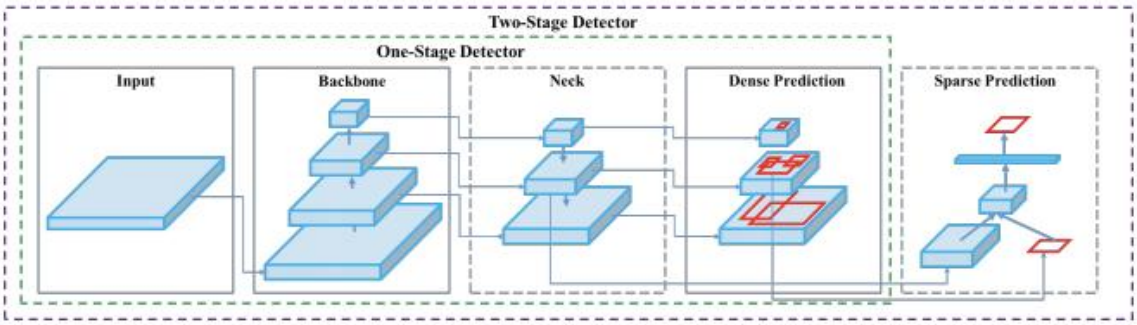


Figure 5. YoLov5 model architecture

The YOLO network consists of three main pieces:

+ Backbone: mainly used to extract important parts from given input image. In YOLOv5 is used as the backbone to extract feature information from input image.

+ Neck: used to create a typical pyramid grid. The pyramids help model generalize about object proportions, help identify the same object with different sizes and scales. YOLOv5 uses PANet to get feature pyramids. And pass it to Head for prediction.

+ Head: mainly used to perform the final detection part. It applies anchor boxes on the features and produces the final output vector. This Head part YOLOv5 is similar to the versions YOLOv3 and YOLOv4.[5]

CHAPTER 3 DESIGN AND INSTALLATION

3.1. Set up the environment and related libraries

Clone yolov5 from github and install related libraries

```
!git clone https://github.com/ultralytics/yolov5  # clone
%cd yolov5
%pip install -qr requirements.txt  # install

import torch
from yolov5 import utils
display = utils.notebook_init()  # checks
```

3.2. Validate

Download and unzip the dataset COCO128

```
# Download COCO
torch.hub.download_url_to_file('https://ultralytics.com/assets/coco128.zip', 'tmp.zip
')
!unzip -q tmp.zip -d ../datasets && rm tmp.zip
```

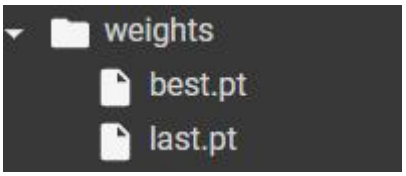
3.3. Train model

Trainning

```
# Train YOLOv5s on COCO128 for 3 epochs
!python train.py --img 640 --batch 16 --epochs 3 --data coco128.yaml --
weights yolov5s.pt --cache
```

After the training is over, there are 2 weights files saved

- Best model: best.pt
- Last model: last.pt



3.4. Detect

```
!python detect.py --weights best.pt --img 640 --conf 0.1 --source data/images
```

- weights: file weight used to predict
- img: image size (resolution)
- conf: reliability
- source: path to photo folder

Prediction results are saved in folder /yolov5/runs/detect/exp1/ (or exp2, exp3,… depend on the number of predictions and detect.py).


yolov5
data
models
runs
detect
exp
exp2
exp3
exp4
exp5
exp6
train

CHAPTER 4: TEST EVALUATION

Figure 6. Result obtained after dectect

➢ Model accuracy

STT	Tên đối tượng	Số lần kiểm tra	Nhận diện đúng	Nhận diện sai	Độ chính xác
1	Person	5	5	0	100%
2	Bicycle	5	5	0	100%
3	Car	5	4	1	80%
4	Motorcycle	5	5	0	100%
5	Airplane	5	5	0	100%
6	Bus	5	4	1	80%
7	Train	5	5	0	100%
8	Truck	5	4	1	80%
9	Boat	5	3	2	60%
10	Traffic light	5	5	0	100%
11	Fire hydrant	5	5	0	100%
12	Stop sign	5	5	0	100%
13	Parking meter	5	5	0	100%
14	Bench	5	4	1	80%
15	Bird	5	5	0	100%
16	Cat	5	5	0	100%
17	Dog	5	5	0	100%
18	Horse	5	3	2	60%
19	Sheep	5	4	1	80%
20	Cow	5	3	2	60%

21	Elephant	5	5	0	100%
22	Bear	5	4	1	80%
23	Zebra	5	5	0	100%
24	Giraffe	5	5	0	100%
25	Backpack	5	2	3	40%
26	Umbrella	5	5	0	100%
27	Handbag	5	0	5	0%
28	Tie	5	5	0	100%
29	Suitcase	5	5	0	100%
30	Frisbee	5	4	1	80%
31	Skis	5	3	2	60%
32	Snowboard	5	3	2	60%
33	Sports ball	5	3	2	60%
34	Kite	5	4	1	80%
35	Baseball bat	5	4	1	80%
36	Baseball glove	5	3	2	60%
37	Skateboard	5	4	1	80%
38	Surfboard	5	2	3	40%
39	Tennis racket	5	5	0	100%
40	Bottle	5	4	1	80%
41	Wine glass	5	5	0	100%
42	Cup	5	5	0	100%
43	Fork	5	2	3	40%
44	Knife	5	1	4	20%
45	Spoon	5	2	3	40%
46	Bowl	5	5	0	100%
47	Banana	5	5	0	100%
48	Apple	5	5	0	100%
49	Sandwich	5	5	0	100%
50	Orange	5	5	0	100%
51	Broccoli	5	5	0	100%
52	Carrot	5	5	0	100%
53	Hot dog	5	5	0	100%
54	Pizza	5	5	0	100%
55	Donut	5	5	0	100%
56	Cake	5	5	0	100%
57	Chair	5	5	0	100%
58	Couch	5	3	2	60%
59	Potted Plant	5	5	0	100%
60	Bed	5	4	1	80%
61	Dining table	5	3	2	60%
62	Toilet	5	4	1	80%
63	TV	5	3	2	60%

64	Laptop	5	4	1	80%
65	Mouse	5	4	1	80%
66	Remote	5	5	0	100%
67	Keyboard	5	5	0	100%
68	Cell phone	5	5	0	100%
69	Microwave	5	4	1	80%
70	Oven	5	3	2	60%
71	Toaster	5	0	5	0%
72	Sink	5	5	0	100%
73	Refrigerator	5	5	0	100%
74	Clock	5	5	0	100%
75	Book	5	5	0	100%
76	Vase	5	5	0	100%
77	Scissors	5	5	0	100%
78	Teddy bear	5	5	0	100%
79	Hair drier	5	1	4	20%
80	Toothbrush	5	4	1	80%

CHAPTER 5: CONCLUSION OF ACHIEVED RESULTS AND DEVELOPMENT DIRECTION

5.1. Achieved results

The basic model performed the recognition based on the training data set. However, there are some limitations:

- The number of objects in the dataset is limited.

- Accuracy is not high because the diversity of objects is still low

- Limited in recognizing many objects at the same time.

5.2. Development direction

The model is the basis for future related applications. The model when put into practice promises to bring certain benefits to users. Contribute to the development of machine learning.

REFERENCES

[1] Cường, TS Nguyễn Mạnh, et al. "Nghiên cứu thuật toán phân loại phương tiện giao thông dựa trên thị giác máy tính."

[2] Tuấn, Đ. H. A., & Thắng, N. M. (2021). Ứng dụng mô hình học sâu trong xác định các giai đoạn sinh trưởng của cây dưa lưới trồng trong nhà màng. *Bản B của Tạp chí Khoa học và Công nghệ Việt Nam*, *63*(11).

[3] https://www.v7labs.com/blog/yolo-object-detection

[4] https://blog.roboflow.com/yolov5-improvements-and-evaluation/

[5] https://machinelearningknowledge.ai/introduction-to-yolov5-object-detection-with-tutorial/

https://appsilon.com/object-detection-yolo-algorithm/

APPENDIX

Instructions for installation and use follow the following link:
https://github.com/mykhanhkhanh/yolov5model