

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM  
TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG  
KHOA CÔNG NGHỆ THÔNG TIN



**BÁO CÁO CUỐI KỲ: KHAI THÁC DỮ LIỆU  
VÀ KHAI PHÁ TRI THỨC**

**OBJECT RECOGNITION  
AND CLASSIFICATION**

*Người hướng dẫn:* **PhD Trần Thanh Phước**

*Người thực hiện:* **Huỳnh Văn Duy - 51703006**

**Trần Quốc Linh - 51703124**

**Khoá: 21**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2019**

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM  
TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG  
KHOA CÔNG NGHỆ THÔNG TIN



**BÁO CÁO CUỐI KỲ: KHAI THÁC DỮ LIỆU  
VÀ KHAI PHÁ TRI THỨC**

**OBJECT RECOGNITION  
AND CLASSIFICATION**

*Người hướng dẫn:* **PhD Trần Thanh Phước**

*Người thực hiện:* **Huỳnh Văn Duy - 51703006**

**Trần Quốc Linh - 51703124**

**Khoá: 21**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2019**

## LỜI CẢM ƠN

Cảm ơn thầy Trần Thanh Phước đã dạy môn Khai thác dữ liệu và khai phá tri thức hết sức nhiệt tình và tâm huyết, cảm ơn thầy đã định hướng và tạo cơ sở để chúng em có thể tự tìm hiểu và học thêm về ...

Còn đây là đồ án của chúng em, sản phẩm của quá trình học tập và rèn luyện dưới sự chỉ dẫn của thầy. Trong quá trình làm bài đồ án này, chúng em vẫn còn nhiều thiếu sót. Chúng em rất mong nhận được sự đánh giá và chỉ bảo từ thầy!

# **BÀI TẬP LỚN ĐƯỢC HOÀN THÀNH TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

Chúng em xin cam đoan đây là sản phẩm bài tập lớn của riêng chúng em. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu, hình ảnh được chính chúng em thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong bài tiểu luận còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

Nếu phát hiện có bất kỳ sự gian lận nào chúng em xin hoàn toàn chịu trách nhiệm về nội dung bài tập lớn của mình. Trường đại học Tôn Đức Thắng không liên quan đến những vi phạm tác quyền, bản quyền do chúng em gây ra trong quá trình thực hiện (nếu có).

*TP. Hồ Chí Minh, ngày 13 tháng 11 năm 2019*

*Tác giả 1*

*(Đã ký)*

*Huỳnh Văn Duy*

*Tác giả 2*

*(Đã ký)*

*Trần Quốc Lĩnh*

# PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN

Phần xác nhận của GV hướng dẫn

---

---

---

---

---

---

Tp. Hồ Chí Minh, ngày 13 tháng 11 năm 2019  
*(kí và ghi họ tên)*

Phần đánh giá của GV chấm bài

---

---

---

---

---

---

Tp. Hồ Chí Minh, ngày 13 tháng 11 năm 2019  
*(kí và ghi họ tên)*

## TÓM TẮT

Nhằm tổng kết lại toàn bộ kiến thức đã được học và vận dụng vào một bài toán thực tế, nhóm xin thực hiện bài tập cuối kỳ với đề tài: Nhận diện sinh viên trường Đại học Tôn Đức Thắng và đánh giá sinh viên có đang mặc trang trang phục đúng quy định hay không.

Phần trình bày của báo cáo này bao gồm các nội dung chính:

- Tìm hiểu và thu thập dữ liệu
- Mô tả giải thuật
- Kết quả
- Tổng kết

# MỤC LỤC

LỜI CẢM ƠN .....	3
CAM KẾT .....	4
ĐÁNH GIÁ CỦA GIÁO VIÊN .....	5
TÓM TẮT .....	6
MỤC LỤC .....	7
DANH MỤC CHÚ THÍCH CÁC HÌNH ẢNH .....	8
CHƯƠNG 1: TÌM HIỂU VÀ THU THẬP DỮ LIỆU .....	9
1.1 Tổng quan về phương pháp giải quyết .....	9
1.2 Thu thập dữ liệu .....	9
1.2.1. Cách thức thu thập dữ liệu .....	9
1.2.2. Xử lý dữ liệu đầu .....	10
CHƯƠNG 2: GIẢI THUẬT .....	11
2.1. Thế nào là Object recognition và Classification? .....	11
2.2. YOLO .....	11
2.3. YOLOv3 .....	14
CHƯƠNG 3: KẾT QUẢ .....	15
3.1. Kết quả thực nghiệm .....	15
3.2. Đánh giá kết quả .....	16
CHƯƠNG 4: TỔNG KẾT .....	17
4.1 Thành quả đạt được .....	17
4.2. Hạn chế .....	17
4.3. Hướng mở rộng .....	17
TÀI LIỆU THAM KHẢO .....	18

# DANH MỤC CHÚ THÍCH CÁC HÌNH ẢNH

## Hình ảnh

Hình 1: Một phần của tập dữ liệu thu thập được

Hình 2: Đóng khung và gắn label cho ảnh

Hình 3: Luồng công việc của model YOLO

Hình 4: Kiến trúc mạng (Network Architecture) của YOLO

Hình 5:  $B$  khung xung quanh ô  $i$

Hình 6: Kết quả thực nghiệm



# CHƯƠNG 1: TÌM HIỂU VÀ THU THẬP DỮ LIỆU

## 1.1 Tổng quan về phương pháp giải quyết.

Với đề tài *Nhận diện sinh viên trường Đại học Tôn Đức Thắng và đánh giá sinh viên có đang mặc trang phục đúng quy định hay không*, chúng ta dễ dàng nhận thấy được có hai tác vụ chính cần xử lý đó là: *Nhận diện* và *Đánh giá*.

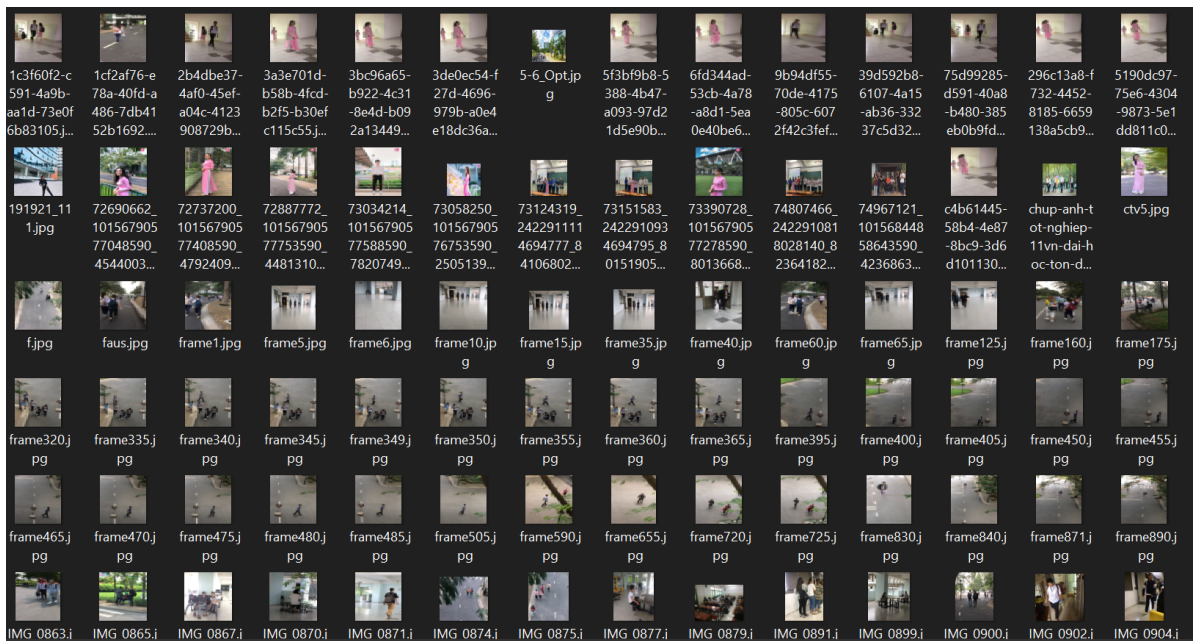
Đầu tiên, cũng như là phần quan trọng nhất trong tất cả các bài toán về xử lý dữ liệu, đó là thu thập và xử lý dữ liệu như thế nào. Còn về phương pháp, giải thuật sử dụng cho việc nhận diện và đánh giá sẽ được trình bày cụ thể ở chương 2.

## 1.2 Thu thập dữ liệu.

Để tăng tính chính xác và thực tế, toàn bộ dữ liệu (ảnh) đều được thu thập tại trường Đại học Tôn Đức Thắng. Tuy nhiên do còn nhiều hạn chế nên tổng số dữ liệu chỉ khoảng 150 mẫu (ảnh).

### 1.2.1. Cách thức thu thập dữ liệu.

Nhóm đã thu thập dữ liệu bằng cách tải từ internet, từ các trang thông tin, chụp bằng điện thoại, quay video và cắt ra thành từng frame ảnh. Ngoài trừ những bức ảnh tải về từ internet, đa phần dữ liệu mà nhóm chụp, quay đều đã được sự đồng ý của các bạn.



Hình 1: Một phần của tập dữ liệu thu thập được

### 1.2.2. Xử lý dữ liệu đầu.

Trước hết, những tấm ảnh bị mờ, ảnh có độ phân giải thấp, ảnh có quá nhiều đối tượng cần nhận diện ở một khung hình nhỏ sẽ được loại bỏ khỏi tập dữ liệu.

Kế đến nhóm dùng công cụ hỗ trợ VGG Image Annotator, để đóng khung cho đối tượng trong ảnh và gắn label (label này cho biết sinh viên có đang mặc trang phục đúng quy định hay không.)



Hình 2: Đóng khung và gắn label cho đối tượng

Sau khi đã đóng khung và gắn label cho toàn bộ ảnh trong tập (ngoại trừ một phần dữ liệu nhỏ dùng làm tập test cho model), từ công cụ trích xuất ra file .json lưu trữ thông tin về tên của ảnh và các tọa độ điểm cho từng khung trong từng ảnh. File .json và tập ảnh sẽ dùng để huấn luyện cho model của bài toán.

## CHƯƠNG 2: GIẢI THUẬT

### 2.1. Thế nào là Object recognition và Classification?

Về cơ bản, object recognition (hay object detection) là phương pháp nhận diện vị trí (tọa độ) của một đối tượng (object) trên một mặt phẳng (ở đây là ảnh). Classification là sự đánh giá (phân loại) đối tượng đó thuộc về lớp (label) nào.

Với sự phát triển không ngừng của kỹ thuật và công nghệ, ngày nay có rất nhiều giải thuật dùng để nhận diện một đối tượng (Selective Search, DPM, R-CNN, Fast R-CNN, Faster R-CNN, Mask R-CNN, SSD, RetinaNet,...) và phân lớp đối tượng đó (KNN, Decision Tree, Naive Bayes, CNN,...).

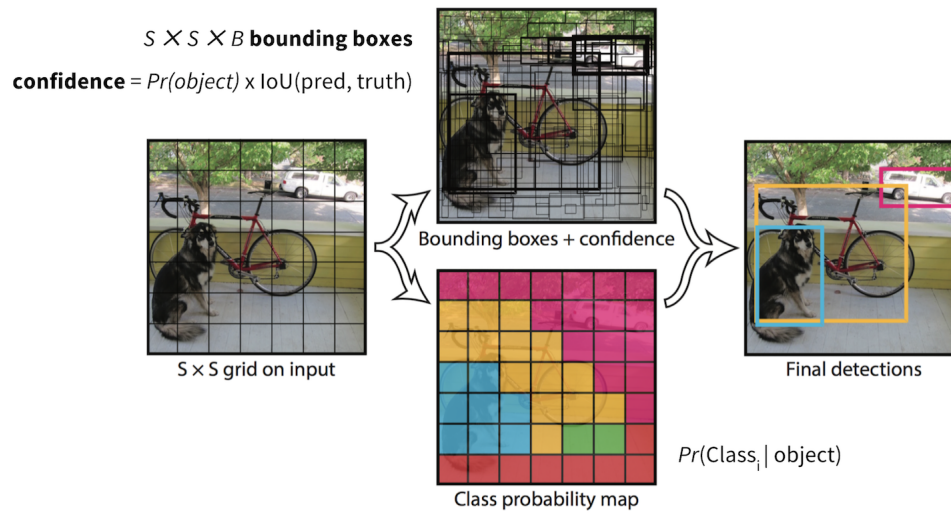
Vì có những ưu và nhược điểm khác nhau, nên các giải thuật kể trên được sử dụng cho các trường hợp khác nhau. Và giải thuật phù hợp nhất theo đánh giá của nhóm cho đề tài này là YOLOv3 và CNN.

### 2.2. YOLO.

YOLO là viết tắt của từ You Only Look Once, đây sự nỗ lực đầu tiên trong việc xây dựng một thiết bị nhận diện vật thể (object detector) chạy với thời gian thực. Bởi vì YOLO chỉ *Look Once* và không phải trải qua các bước trung gian như ở các giải thuật khác, nên tốc độ tính toán và cho ra kết quả của nó là cực kỳ nhanh.

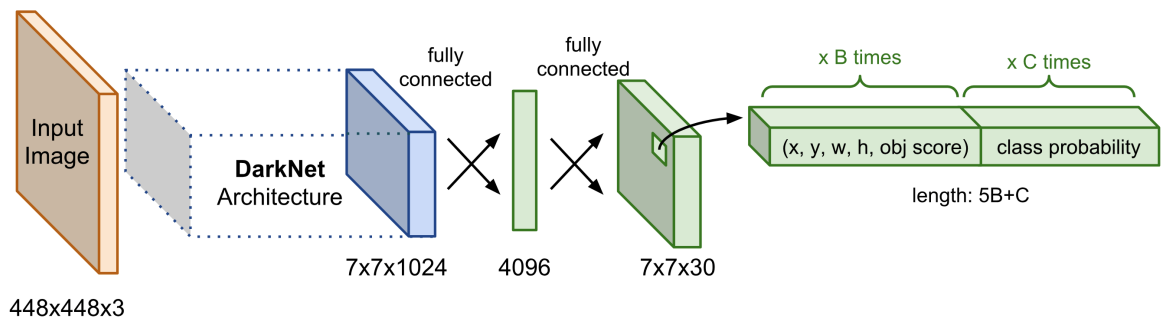
#### Luồng công việc (Workflow)

1. Chuẩn bị trước một model CNN đã được train cho việc phân lớp
2. Chia bức ảnh cần nhận diện và phân lớp đối tượng thành  $S \times S$  ô (cell). Nếu có phần nào của một đối tượng nằm vào một ô, thì ô này là "tín nhiệm" cho việc phát hiện sự tồn tại của đối tượng đó. Tại mỗi ô, tính:
  - Tọa độ dự đoán của B khung (bounding boxes), trong  $B$  khung này có khả năng có một hoặc nhiều khung bao quanh được đúng đối tượng cần tìm.
  - Độ tin cậy (confidence score), giá trị này thể hiện cho khả năng chứa một đối tượng của ô.
  - $K$  xác suất (probability), xác suất tồn tại của đối tượng  $C_i$  ( $i = 1, 2, \dots, K$ ) trong từng khung.
3. Tại lớp (layer) cuối của model CNN (đã chuẩn bị ở trên) điều chỉnh sao cho đầu ra (output) có dạng  $S \times S \times (5B + K)$ . Bởi một ảnh có  $S \times S \times B$  khung, tại mỗi khung có 4 giá trị thể hiện cho tọa độ dự đoán, 1 cho độ tin cậy và  $K$  xác suất tại mỗi ô.



Hình 3: Luồng công việc của model YOLO

Kết quả dự đoán cuối cùng có dạng  $S \times S \times (5B + K)$  được tạo bởi hai lớp liên kết đầy đủ (fully connected layer) cho toàn bộ bản đồ đặc tính tích chập (convolutional feature map).



Hình 4: Kiến trúc mạng (Network Architecture) của YOLO

### Mất mát (Loss Function)

Hàm tính mất mát dựa trên: tổng sai lệch giá trị giữa dự đoán với thực tế của tọa độ khung (localization loss) và phân lớp (classification loss). Ngoài ra còn có hai tham số được dùng cho việc điều chỉnh các giá trị sai lệch (loss) mà chúng ta muốn xảy ra là  $\lambda_{\text{noobj}}$  và  $\lambda_{\text{coord}}$ .

$$\mathcal{L}_{\text{loc}} = \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{\text{obj}} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2]$$

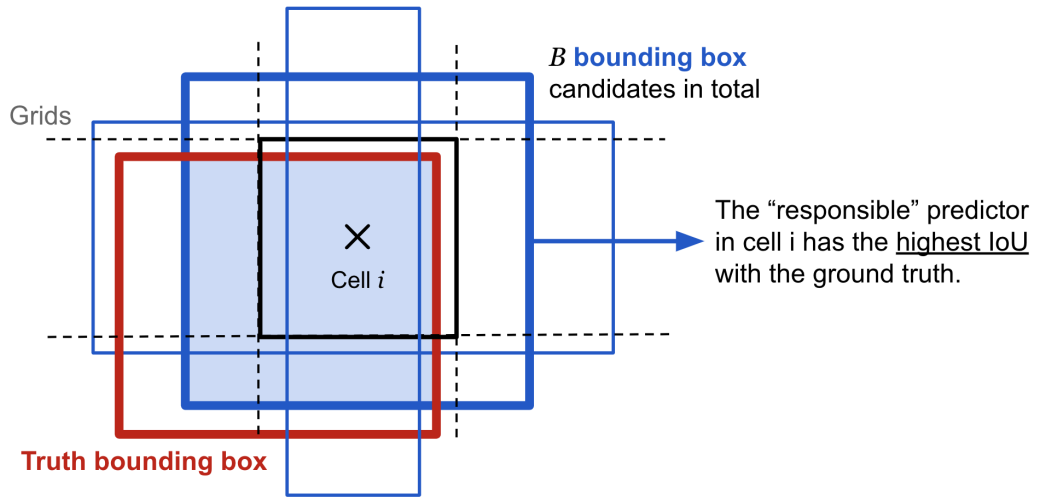
$$\mathcal{L}_{\text{cls}} = \sum_{i=0}^{S^2} \sum_{j=0}^B (1_{ij}^{\text{obj}} + \lambda_{\text{noobj}}(1 - 1_{ij}^{\text{obj}}))(C_{ij} - \hat{C}_{ij})^2 + \sum_{i=0}^{S^2} \sum_{c \in \mathcal{C}} 1_i^{\text{obj}} (p_i(c) - \hat{p}_i(c))^2$$

$$\mathcal{L} = \mathcal{L}_{\text{loc}} + \mathcal{L}_{\text{cls}}$$

Trong đó:

- $1_i^{\text{obj}}$  Một hàm biểu thị liệu ô  $i$  có chứa đối tượng nào hay không.
- $1_{ij}^{\text{obj}}$  Biểu thị khung thứ  $j$  của ô  $i$  có là khung tín nhiệm (xem hình 5) hay không.
- $C_{ij}$  Giá trị tin cậy của ô  $i$ .
- $\hat{C}_{ij}$  Giá trị tin cậy dự đoán của ô  $i$ .
- $\mathcal{C}$  Tập các lớp đối tượng có thể có.
- $p_i(c)$  Xác suất chứa đối tượng  $c$  của ô  $i$ ,  $c \in \mathcal{C}$ .
- $\hat{p}_i(c)$  Xác suất dự đoán đối tượng  $c$  tồn tại trong ô  $i$ .

Trong hình 5, ta thấy có  $B$  khung (bounding box) nằm xung quanh ô (cell)  $i$ , trong đó có một khung (khung được in đậm) là có số vùng trùng khớp lớn nhất so với khung thực tế. Khung này được xem là khung dự đoán "tín nhiệm".



Hình 5:  $B$  khung xung quanh ô  $i$

### 2.3. YOLOv3.

YOLOv3 hay You Only Look Once version 3, đúng như cái tên đây là phiên bản cải tiến của YOLO thế hệ đầu. Danh sách các thay đổi gồm:

1. **Dùng hồi quy tuyến tính (Logistic regression) để tính độ tinh cậy**, trong khi YOLO phiên bản đầu tính dựa trên tổng bình phương các sai số.
2. **Không còn softmax**, YOLOv3 phân lớp bằng nhiều hàm logistic tuyến tính một cách độc lập (multiple independent logistic classifier).
3. **Sử dụng Darknet + ResNet làm nền tảng.**
4. **Dự đoán trên nhiều tỉ lệ hơn (Multi-scale prediction).**
5. **Bỏ qua lớp nối kết (concatenation layer).**



## CHƯƠNG 3: KẾT QUẢ

### 3.1. Kết quả thực nghiệm.

Nhóm đã quay video và tách ra thành từng frame làm dữ liệu test cho model sau khi đã huấn luyện. Dưới đây là một số output.



Hình 6: Kết quả thực nghiệm.

### **3.2. Đánh giá kết quả.**

Đã xác định được tối tượng và đánh giá được các trường hợp sinh viên mặc trang phục không đúng quy định.

Tuy nhiên do dữ liệu dùng để huấn luyện còn ít nên vẫn còn tồn tại một vài hạn chế.

Đầu tiên về phần xác định tọa độ của đối tượng trong ảnh vẫn còn sai số đáng kể, sai đối tượng.

Và với phần xác định trang phục có đúng quy định hay không, đôi khi hệ thống vẫn còn có sự nhầm lẫn giữa những trang phục có cùng màu sắc và đánh giá sai.



## CHƯƠNG 4: TỔNG KẾT

### 4.1 Thành quả đạt được.

Nắm chắc các phương pháp trong môn Khai thác dữ liệu và khai phá tri thức, biết được cách thu thập và xử lý dữ liệu để có một model tốt. Các giải thuật đã được tìm hiểu rõ ràng hơn, đặc biệt là giải thuật về nhận diện và phân lớp đối tượng.

Khả năng phân công và hợp tác trong khi làm việc nhóm được cải thiện, công việc được hoàn thành hiệu quả và năng suất.

### 4.2. Hạn chế.

Dữ liệu thu thập còn ít, nên chưa tạo ra được sản phẩm đúng như mong đợi.

### 4.3. Hướng mở rộng.

Nếu hệ thống được huấn luyện với tập dữ liệu lớn sẽ có thể hoạt động chính xác và chi tiết hơn. Ví dụ sinh viên có đang đóng thùng hay không.

Kết hợp với camera an ninh của trường sẽ dễ dàng nhận diện được sinh viên ăn mặc không đúng quy định và xử phạt. Giúp nâng cao ý thức cho mỗi sinh viên, hình thành được tính kỷ luật giúp ích cho công việc sau này.

# TÀI LIỆU THAM KHẢO

## Tiếng Việt

[1]<https://deepmlml.com> › classic-object-detection

## Tiếng Anh

[1]<https://pjreddie.com/darknet/yolo/>

[2]<https://lilianweng.github.io/lil-log>