

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

NGUYỄN ĐÌNH TIÊM

**XÂY DỰNG HỆ THỐNG NHẬN
DIỆN CẢM XÚC KHUÔN MẶT
VỚI YOLOv5**

ĐỒ ÁN TỐT NGHIỆP ĐẠI HỌC
CHUYÊN NGÀNH CÔNG NGHỆ THÔNG TIN

THÁI NGUYÊN, NĂM 2022

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG



ĐỒ ÁN
TỐT NGHIỆP ĐẠI HỌC
Chuyên ngành Công nghệ thông tin

Đề tài:

**XÂY DỰNG HỆ THỐNG NHẬN DIỆN CẢM XÚC
KHUÔN MẶT VỚI YOLOv5**

Sinh viên thực hiện: NGUYỄN ĐÌNH TIÊM
Lớp CNTT K16I, hệ chính qui

Giáo viên hướng dẫn: TS. TRẦN QUANG QUÝ

Thái Nguyên, năm 2022

NHIỆM VỤ ĐỒ ÁN TỐT NGHIỆP

- Tìm hiểu bài toán nhận diện cảm xúc khuôn mặt, ý nghĩa và ứng dụng bài toán này, khảo sát các phương pháp giải quyết bài toán.
- Tìm hiểu về mạng Nơ-ron, mạng Nơ-ron tích chập
- Tìm hiểu, nghiên cứu về mô hình nhận diện đối tượng YOLO, các kiến thức liên quan như các phương pháp đánh giá mô hình, các phương pháp xử lý dữ liệu trong YOLO...
- Thu thập xử lý dữ liệu, thử nghiệm bài toán trên các bộ dữ liệu và đánh giá.

LỜI CẢM ƠN

Đi qua những năm tháng tại trường Đại học Công nghệ thông tin và truyền thông Thái Nguyên-ICTU tôi mới thấy được tuổi trẻ đáng giá như thế nào. Tuổi trẻ là những cơn mưa rào dù có làm tôi cảm lạnh nhưng tôi vẫn khát khao được đắm thân bút phá trong những trận mưa ấy.

Cảm ơn ICTU, 5 năm gắn bó với ICTU là một quãng thời gian thật sự quan trọng trong tuổi thanh xuân của tôi. Dù sau này có đi đến chân trời nào tôi sẽ mãi nhớ về những kỷ niệm với ICTU, tự hào mình là một người con của ICTU.

Cảm ơn những người bạn đã đồng hành cùng tôi suốt 5 năm học tại ICTU, cảm ơn các bạn đã cùng tôi cố gắng, thấu cảm và cùng nhau vượt qua những khó khăn. Cảm ơn bố mẹ đã vất vả làm việc để cho tôi có cơ hội bước trên con đường trí thức ngày hôm nay. Và tôi xin cảm ơn các thầy giáo, cô giáo, cán bộ nhân viên trường Đại học Công nghệ thông tin và truyền thông Thái Nguyên và thầy giáo Trần Quang Quý-một người thầy tận tâm đã hướng dẫn tôi hoàn thành đề án tốt nghiệp của mình.

Với trình độ và kiến thức còn hạn chế nên đề án còn nhiều thiếu sót, mong được thầy cô thông cảm và góp ý.

ICTU, tháng 01 năm 2022

Xin chân thành cảm ơn!

LỜI CAM ĐOAN

Đồ án tốt nghiệp này là sản phẩm tổng hợp toàn bộ các kiến thức mà tôi đã học và nghiên cứu trong thời gian học tập tại trường. Ý thức được điều đó, với tinh thần nghiêm túc, sự tự giác cùng sự cố gắng miệt mài của bản thân và sự hướng dẫn tận tình của thầy Trần Quang Quý. Tôi đã hoàn thành đồ án tốt nghiệp của mình.

Tôi xin cam đoan: nội dung bài làm của tôi là công sức tôi đã tìm hiểu, không sao chép từ nội dung của các đồ án khác, có sự tham khảo các mẫu đồ án trên thư viện nhà trường về nội dung và hình thức trình bày. Mọi thông tin sai lệch tôi xin hoàn toàn chịu trách nhiệm trước hội đồng bảo vệ.

MỤC LỤC	
NHIỆM VỤ ĐỒ ÁN TỐT NGHIỆP	3
LỜI CẢM ƠN	4
LỜI CAM ĐOAN	5
MỤC LỤC	6
MỤC LỤC HÌNH ẢNH	9
LỜI NÓI ĐẦU	12
CHƯƠNG 1. TỔNG QUAN VỀ BÀI TOÁN NHẬN DIỆN CẢM XÚC KHUÔN MẶT	14
1.1 Giới thiệu bài toán nhận diện cảm xúc trên khuôn mặt.	14
1.1.1 Phương pháp phân lớp	15
1.1.2 Phương pháp nhận diện đối tượng.....	15
CHƯƠNG 2. MẠNG NƠ-RON VÀ MẠNG NƠ-RON TÍCH CHẬP	17
2.1 Mạng Nơ-ron [3]	17
2.1.1 Mạng Nơ-ron sinh học	17
2.1.2 Mô hình mạng Nơ-ron	18
2.1.3 Hàm kích hoạt trong mạng Nơ-ron.....	18
2.2 Mạng Nơ-ron tích chập(Convolutional neural network)	20
2.2.1 Phép tính tích chập Convolutional.....	20
2.2.2 Tầng tích chập(Convolutional layer)	23
2.2.3 Tầng tổng hợp đặc trưng(Pooling layer).....	25
2.2.4 Tầng kết nối đầy đủ(Fully connected layer)	26
CHƯƠNG 3. GIẢI PHÁP NHẬN DIỆN ĐỐI TƯỢNG YOLO	28
3.1 Nội Dung	28
3.2 Một số khái niệm quan trọng trong một mô hình nhận diện đối tượng.	28
3.2.1 IOU-Giao trên hợp(Intersection over union).....	28
3.2.2 Phép đo độ chính xác(Accuracy) [4]	29
3.2.3 Ma trận nhầm lẫn(Confusion Matrix) [4]	30
3.2.4 Độ đo Precision và Recall.....	31
3.2.5 Precision Recall Curve, AP và mAP	32
3.2.6 Độ đo F1 score.....	33

3.3 Mô hình một giai đoạn(One stage) và mô hình hai giai đoạn (two stage detectors).....	34
3.3.1 Two stage detectors	34
3.3.2 One stage detectors	35
3.4 YOLOv1 [5] [6]	35
3.4.1 Cấu trúc.....	35
3.4.2 Hàm mất mát (Loss function)	38
3.4.3 Phương pháp loại bỏ bớt Boundingbox(Non-maximal suppression (NMS))	38
3.5 YOLOv2 [7] [6]	39
3.5.1 Cấu trúc.....	39
3.6 YOLOv3- Tốt hơn, không nhanh hơn, mạnh hơn(Better, not Faster, Stronger) [8] [6].	42
3.6.1 Cấu Trúc	42
3.6.2 Nhận dạng ở các tỉ lệ ảnh khác nhau.	44
3.6.3 Đầu ra của YOLOv3.....	45
3.6.4 Khung mặc định (Anchor box).....	46
3.7 YOLOv4- Tối ưu hóa tốc độ và độ chính xác(Optimal Speed and Accuracy of Object Detection) [9] [6].....	47
3.7.1 Backbone(Xương sống) – Trích xuất đặc trưng.	48
3.7.2 Neck (phần cổ) – Tổng hợp đặc trưng.	50
3.7.3 Head (phần đầu) – Bước nhận dạng.	55
3.8 YOLOv4 – Phương pháp cải thiện kết quả suy luận(Bag of Freebies).	56
3.8.1 Tăng sinh dữ liệu bằng cách: cắt và trộn- CutMix data augmentation.....	56
3.8.2 Tăng sinh dữ liệu bằng cách "khảm"- Mosaic data augmentation.	56
3.8.3 Phương pháp loại bỏ các điểm giống nhau(DropBlock regularization).	57
3.8.4 Làm mịn lớp nhãn học(Class label smoothing).	58
3.8.5 Túi đặc biệt cho phần xương sống(BoS cho backbone).	59
3.9 YOLOv5 [2]	61
CHƯƠNG 4. XỬ LÝ DỮ LIỆU, THỬ NGHIỆM MÔ HÌNH VÀ KẾT QUẢ	63
4.1 Các bước thu thập và xử lý dữ liệu.	63

4.1.1 Thu thập dữ liệu.....	63
4.1.2 Tiền xử lí dữ liệu	63
4.2 Phương pháp gán nhãn.....	64
4.2.1 Phương pháp gán nhãn thủ công.....	64
4.2.2 Phương pháp gán nhãn bán tự động.	65
4.2.3 Bộ dữ liệu.....	66
4.3 Kỹ thuật huấn luyện mô hình YOLOv5 trên Google Colab	67
4.3.1 Cấu trúc thư mục trước đào tạo cho YOLOv5	67
4.3.2 Sử dụng GoogleColab.....	68
4.4 Kết quả huấn luyện.....	69
4.4.1 Đánh giá mô hình YOLOv5s, YOLOv5x trên bộ dữ liệu Widerface.....	70
4.4.2 Đánh giá mô hình YOLOv5s, YOLOv5x trên bộ dữ liệu face_emoji	71
4.5 Demo	73
4.5.1 Kết quả nhận diện khuôn mặt trên mô hình YOLOv5 với bộ dữ liệu Widerface.....	73
4.5.2 Mô hình YOLOv5 Face moji.....	73
TÀI LIỆU THAM KHẢO	75
DANH MỤC THUẬT NGỮ.....	75
KẾT LUẬN	76
NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN.....	77
NHẬN XÉT CỦA GIẢNG VIÊN PHẢN BIỆN.....	78

MỤC LỤC HÌNH ẢNH

Hình 1.1.1 Mô hình phân lớp thông thường.....	15
Hình 1.1.2 Mô hình object detection Yolov5.....	16
Hình 2.1.1 Mạng Noron sinh học.....	17
Hình 2.1.2 Mô hình mạng Nơ-ron.....	18
Hình 2.1.3 Hàm biến đổi Hardlimit.....	19
Hình 2.1.4 Hàm biến đổi Linear.....	19
Hình 2.1.5 Hàm Sigmoid.....	20
Hình 2.2.1 Ý nghĩa phép tính convolution.....	22
Hình 2.2.2 Mô hình mạng nơ-ron.....	23
Hình 2.2.3 Input layer và Hidden Layer1.....	23
Hình 2.2.4 Phép tính convolution trên ảnh màu với $k=3$	24
Hình 2.2.5 Tensor X, W 3 chiều được viết dưới dạng 3 matrix.....	24
Hình 2.2.6 Thực hiện phép tính convolution trên ảnh màu.....	25
Hình 2.2.7 MaxPooling với Filter(2, 2) stride(2, 2).....	26
Hình 2.2.8 Max pooling và average pooling.....	26
Hình 2.2.9 Fully connected layer.....	26
Hình 2.2.10 Mô hình convolution neural network.....	27
Hình 3.2.1 Minh họa cách tính IOU.....	29
Hình 3.2.2 Minh họa IOU.....	29
Hình 3.2.3 Confusion matrix.....	30
Hình 3.2.4 Precison và Recall.....	31
Hình 3.2.5 Precision Recall Curve.....	32
Hình 3.2.6 Độ đo AP.....	33
Hình 3.3.1 Giai đoạn 1 của model tow stage.....	34
Hình 3.3.2 Giai đoạn 2 của model tow stage.....	34
Hình 3.3.3 Mô hình one stage(Yolo, SSD).....	35
Hình 3.4.1 Mô hình YOLOv1.....	35
Hình 3.4.2 Mô hình YOLOv1 chia hình thành SxS ô lưới.....	36
Hình 3.4.3 Trọng tâm của vật thể.....	37
Hình 3.4.4 Đầu ra YOLOv1.....	37
Hình 3.4.5 Hình ảnh trước và sau khi sử dụng NMS.....	39
Hình 3.5.1 Phát hiện nhiều đối tượng trong 1 ô lưới.....	40
Hình 3.5.2 Hình minh họa 1% số lượng Anchor Box trong mạng RetinaNet.....	41
Hình 3.5.3 Darknet 19.....	42
Hình 3.6.1 Darknet 53.....	43
Hình 3.6.2 Cấu trúc YOLOv3.....	44
Hình 3.6.3 Quá trình tìm kiếm vật thể.....	45
Hình 3.6.4 Kích thước của Tensor.....	46

Hình 3.6.5 Quá trình đào tạo để tinh chỉnh kích thước của anchor box sao cho giống với vật thể nhất.	46
Hình 3.7.1 So sánh mô hình Yolov3 với Yolov4	47
Hình 3.7.2 Cấu trúc YOLOv4	48
Hình 3.7.3 Các backbone được sử dụng trong YOLOv4.	48
Hình 3.7.4 EfficientNet	49
Hình 3.7.5 Cấu Trúc CSP	50
Hình 3.7.6 Cấu Trúc DenseNet	50
Hình 3.7.7 Một số cấu trúc sử dụng cho phần cổ (Neck).....	51
Hình 3.7.8 SPP YOLO	52
Hình 3.7.9 SPP YOLO(bỏ qua DC block)	52
Hình 3.7.10 Hoạt động của Path Aggregation Network.....	53
Hình 3.7.11 PANET	53
Hình 3.7.12 PAN thay đổi add thành concat.....	53
Hình 3.7.13 Mô hình Spatial Attention Module(SAM)	54
Hình 3.7.14 Mô hình Convolutional Block Attention Module	54
Hình 3.7.15 Mô hình SAM trong Yolov4	55
Hình 3.7.16 Quá trình đào tạo để tinh chỉnh kích thước của anchor box sao cho giống với vật thể nhất.	55
Hình 3.8.1 CutMix.....	56
Hình 3.8.2 Tăng dinh dữ liệu – Mosaic.....	57
Hình 3.8.3 Phương pháp DropBlock	57
Hình 3.8.4 Ảnh hưởng của Label Smoothing với mô hình	58
Hình 3.8.5 Mish activation function.....	59
Hình 3.8.6 Sự khác biệt khi sử dụng mish, reLu, Swish	60
Hình 3.8.7 Kết quả so sánh ReLU, Mish, Swish trên tập dữ liệu MNIST	61
Hình 3.9.1 Các mô hình Yolov5.....	61
Hình 3.9.2 Đánh giá mô hình YOLOv5 trên bộ dữ liệu COCO.....	62
Hình 4.1.1 Thu thập hình ảnh.....	63
Hình 4.1.2 Bộ dữ liệu Widerface [10].....	63
Hình 4.1.3 Chuyển đổi định dạng Widerface	64
Hình 4.2.1 Thiết lập các lớp	64
Hình 4.2.2 Tool labelImg gán nhãn dữ liệu.....	65
Hình 4.2.3 Mô hình gán nhãn tự động.....	65
Hình 4.2.4 Sử dụng labelImg chỉnh sửa nhãn	66
Hình 4.2.5 Dữ liệu sau khi xử lí	67
Hình 4.2.6 Chi tiết nhãn	67
Hình 4.3.1 Cấu trúc thư mục data Yolov5.....	68
Hình 4.3.2 Thiết lập GPU trên google colab	69
Hình 4.3.3 Cài đặt môi trường.....	69