TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP. HCM

Khoa Điện – Điện Tử

HCMUTE

KHÓA LUẬN TỐT NGHIỆP

SỬ DỤNG THUẬT TOÁN YOLO NHẬN DIỆN KÍ HIỆU TAY HỖ TRỢ GIAO TIẾP CHO NGƯỜI KHIẾM THÍNH - KHIẾM THỊ

GVHD: PGS.TS Trương Ngọc Sơn

SVTH: Phạm Hữu Nghĩa - 20119256

Nguyễn Xuân Hải - 20119221

TP Hồ Chí Minh, tháng 5 năm 2024

MỤC LỤC

DANH MỤC TỪ VIẾT TẮT

YOLO	You Only Look Once
ML	Machine Learning
CLI	Command Line Interface
JN	Jetson Nano
ReLU	Rectified Linear Unit
COCO	Common Objects in Context
RP Pi	Rasberry Pi
SOTA	State-of-the-art
AP	Average Precision
mAP	Mean Average Precision
IoU	Intersection over Union
NMS	Non-Maximum Suppression
FPN	Feature Pyramid Network
PAN	Path Aggregation Network
CNN	Convolution Neural Network

DANH MỤC HÌNH ẢNH

LỜI CẢM ƠN

Để hoàn thành đồ án tốt nghiệp, sinh viên thực hiện đề tài xin chân thành cảm ơn:

Thầy Trương Ngọc Sơn – Giảng viên Bộ môn Kỹ thuật máy tính – Viễn thông, Trường Đại Học Sư Phạm Kỹ Thuật TP.HCM đã theo sát, tận tình hướng dẫn, giúp đỡ cũng như tạo những điều kiện thuận lợi trong suốt quá trình thực hiện để em có thể hoàn thành đề tài một cách tốt nhất.

Nhóm cũng xin chân thành cảm ơn các thầy cô giáo trong trường Đại học Sư Phạm Kỹ Thuật Tp.HCM nói chung, các thầy cô trong Bộ môn Kỹ Thuật Máy Tính nói riêng đã cho nhóm em kiến thức về các môn đại cương cũng như các môn học chuyên ngành, giúp nhóm em có được cơ sở lý thuyết vững vàng và tạo điều kiện giúp đỡ nhóm em trong suốt quá trình học tập. Các bạn sinh viên của tập thể lớp 20119CL2 đã có những giúp đỡ thiết thực, cung cấp tài liệu liên quan, cũng như động viên trong quá trình thực hiện đề tài.

Tp. Hồ Chí Minh, Tháng 5 năm 2024
Nhóm sinh viên thực hiện

Phạm Hữu Nghĩa- Nguyễn Xuân Hải

CHƯƠNG I: GIỚI THIỆU TỔNG QUÁT VỀ ĐỀ TÀI

1.1 Tính cấp thiết của đề tài

Người khiếm thính - khiếm thị gặp nhiều khó khăn trong giao tiếp do không thể nghe hoặc nhìn thấy. Khả năng giao tiếp hiệu quả đóng vai trò quan trọng trong việc giúp họ hòa nhập vào cộng đồng và nâng cao chất lượng cuộc sống.

Ký hiệu tay là ngôn ngữ phi ngôn ngữ quan trọng giúp người khiếm thính - khiếm thị giao tiếp với nhau và với người bình thường. Tuy nhiên, việc sử dụng ngôn ngữ ký hiệu thủ công đòi hỏi người sử dụng phải có kỹ năng cao và tốn nhiều thời gian học tập.

Các phương pháp giao tiếp hiện có cho người khiếm thính - khiếm thị như bảng chữ cái nổi, máy trợ thính, máy cấy ốc tai còn nhiều hạn chế.

1.2 Lý do chọn đề tài

Hiện tại, người khiếm thính - khiếm thị đang gặp phải nhiều khó khăn trong việc giao tiếp, điều này ảnh hưởng đến cuộc sống của họ và khả năng hòa nhập vào cộng đồng. Các phương pháp giao tiếp hiện có như ngôn ngữ ký hiệu, chữ nổi Braille và thiết bị hỗ trợ đều đặn ra những thách thức, bao gồm việc tốn thời gian học tập và khó sử dụng trong một số tình huống.

Để giải quyết vấn đề này, một giải pháp được đề xuất là sử dụng thuật toán YOLO để nhận diện ký hiệu tay và hỗ trợ giao tiếp cho người khiếm thính - mù. Hệ thống nhận diện này dựa trên thuật toán YOLO, với những ưu điểm như chính xác, hiệu quả và dễ sử dụng.

Nó tập trung vào việc nhận diện các ký hiệu tay cơ bản trong ngôn ngữ ký hiệu Việt Nam và có thể tích hợp vào thiết bị di động. Bằng cách này, hệ thống có thể giúp nâng cao chất lượng cuộc sống của người khiếm thính - khiếm thị, giúp họ giao tiếp một cách hiệu quả hơn với mọi người và hòa nhập tốt hơn vào cộng đồng.

1.3 Đối tượng và phương pháp nghiên cứu

1.3.1 Đối tượng nghiên cứu

Người khiến thính- khiếm thị : Nhóm đối tượng này khó khăn trong giao tiếp do không thể nghe hoặc nhìn thấy. Thuật toán nhận diện ký hiệu tay có thể giúp họ giao tiếp hiệu quả hơn so với người khác.

Ký hiệu tay: Hệ thống ngôn ngữ phi ngôn ngữ được sử dụng bởi người khiếm thính - khiếm thị để giao tiếp. Thuật toán YOLO có thể được sử dụng để nhận diện các ký hiệu tay này một cách chính xác và hiệu quả.

Thuật toán YOLO (You Only Look Once): Là một thuật toán học máy được sử dụng để phát hiện vật thể trong hình ảnh và video. Thuật toán YOLO có thể được sử dụng để nhận diện các ký hiệu tay trong thời gian thực.

1.3.2 Mục tiêu nghiên cứu

Đề tài được thực hiện với mục đích đưa ra được giải pháp và triển khai được đề tài để giúp những người khiếm thính - khiếm thị có thể giao tiếp với mọi người một cách dễ dàng hơn. Đề tài thực hiện bằng cách áp dụng thuật toán YOLO vào việc nhận dạng kí hiệu tay để giúp mọi người có thể nhận biết được thông điệp mà không cần phải mất thời gian để học kí hiệu tay.

Một cách tổng quát, đề tài dự kiến đạt được những mục tiêu sau: - Hệ thống có thể nhận diện chính xác các kí hiệu tay trong ngôn ngữ kí hiệu , bao gồm cả các kí hiệu đơn lẻ và các cụm từ.

- Hệ thống có thể hoạt động hiệu quả trong môi trường thực tế với nhiều điều kiện ánh sáng khác nhau.

- Hệ thống có thể chuyển đổi các kí hiệu tay được nhận diện thành văn bản hoặc âm thanh, giúp người khiếm thính - khiếm thị có thể giao tiếp với người khác.

- Hệ thống này cũng có thể được sử dụng trong các ứng dụng khác như giáo dục, y tế và dịch vụ khách hàng.

1.3.3 Phương pháp nghiên cứu

Đề tài thực hiện bằng cách áp dụng thuật toán YOLO (YOLOv8) để thực hiện việc nhận dạng kí hiệu tay, lấy ngõ vào với camera, ngõ ra là màn hình hiển thị văn bản được dịch, loa để đọc văn bản cho người khiếm thị. Một cách tổng quát, nhóm đã có hướng làm đề tài như sau:

Xác định bài toán: Mục đích của đề tài là dịch được ký hiệu tay nên nhóm cần phải có một model dạng detection/segmentation. Vì thế nhóm đã chọn YOLOv8 để thực hiện đề tài (Vì tính mới, nhận dạng nhanh có thể đến mức Real-Time, dễ ứng dụng và phổ biến)

Kiến thức cần có: Nhóm cần phải tìm hiểu cách hoạt động của CNN, các hàm kích hoạt như ReLU, Sigmoid, Softmax. Các vấn đề của hàm kích hoạt. Các thuật ngữ: Anchor box, bounding box, feature map, non-max suppression. Xem lại các thuật toán Forward Propagation, Back Propagation, Gradient Descent.

Kiến trúc mạng YOLO: Tìm hiểu về kiến trúc mạng YOLO (Lớp, inputs, outputs, loss function).

Chuẩn bị Dataset*:* Lấy dataset từ COCO, Kaggle hoặc tự thu thập dữ liệu và gán label thủ công.

Thực hiện train dữ liệu và đánh giá mô hình*:* Huấn luyện mô hình YOLO trên Google Collab hoặc trên Local.

Đánh giá mô hình và dự đoán dữ liệu mới.

Tìm hiểu về máy tính nhúng Jetson Nano/RP Pi*:* Cấu trúc phần cứng ARM, GPIO, cổng giao tiếp.

Thực thi mô hình lên phần cứng: Áp dụng kiến thức về lập trình nhúng, thực thi mô hình lên Jetson Nano/ RP Pi, giao tiếp ngoại vi: Camera, Loa, Màn hình.

1.4 Giới hạn đề tài

1.4.1 Về mặt kỹ thuật

Độ chính xác của hệ thống nhận diện ký hiệu tay: Hiệu suất của hệ thống nhận diện ký hiệu tay phụ thuộc vào nhiều yếu tố như chất lượng hình ảnh, tốc độ cử chỉ, độ phức tạp của ký hiệu tay, v.v. Do đó, độ chính xác của hệ thống có thể không đạt được 100% trong mọi trường hợp.

Khả năng nhận diện các ký hiệu tay phức tạp: Một số ký hiệu tay có thể rất phức tạp và khó nhận diện bằng thuật toán YOLO. Hệ thống có thể gặp khó khăn trong việc nhận diện chính xác các ký hiệu tay này.

Khả năng nhận diện ký hiệu tay trong môi trường nhiễu: Hệ thống có thể gặp khó khăn trong việc nhận diện ký hiệu tay trong môi trường nhiễu như tiếng ồn, ánh sáng yếu, v.v

1.4.2 Về mặt ứng dụng

Yêu cầu người dùng có kiến thức về ngôn ngữ ký hiệu*:* Để sử dụng hệ thống hiệu quả, người dùng cần có kiến thức cơ bản về ngôn ngữ ký hiệu.

Chưa thể thay thế hoàn toàn giao tiếp bằng lời nói: Hệ thống nhận diện ký hiệu tay chưa thể thay thế hoàn toàn giao tiếp bằng lời nói. Trong một số trường hợp, người dùng vẫn cần sử dụng các phương thức giao tiếp khác như viết hoặc sử dụng bảng chữ cái.

CHƯƠNG II: THIẾT KẾ MÔ HÌNH

2.1 Đặc tả kỹ thuật

2.1.1 Giới thiệu

Mô hình nhận diện và dịch ngôn ngữ ký hiệu sang văn bản và giọng nói để giúp người khiếm thính và khiếm thị giao tiếp hiệu quả hơn.

2.1.2 Yêu cầu chức năng

Nhận diện các cử chỉ ngôn ngữ ký hiệu trong thời gian thực với độ chính xác cao. Chuyển đổi cử chỉ ngôn ngữ ký hiệu sang văn bản. Chuyển đổi văn bản sang giọng nói. Phân tích biểu cảm khuôn mặt của người đối diện. Hiển thị thông tin (văn bản) cho người dùng.

2.1.3 Yêu cầu phi chức năng

Hiệu suất*:* Hệ thống có thể xử lý video đầu vào với tốc độ tối thiểu 30 khung hình/giây.

Độ chính xác: Tỷ lệ nhận diện cử chỉ chính xác phải đạt tối thiểu 80%.

Khả năng bảo trì: Hệ thống dễ dàng sửa lỗi và nâng cấp.

Khả năng sử dụng: Giao diện người dùng đơn giản và dễ sử dụng.

2.1.4 Giao diện

- Hệ thống sử dụng camera để thu thập video đầu vào.

- Giao diện người dùng hiển thị video đầu vào, kết quả nhận diện cử chỉ, văn bản tương ứng và biểu cảm khuôn mặt của người đối diện.

- Hệ thống có thể sử dụng loa hoặc tai nghe để phát ra giọng nói.

2.1.5 Ràng buộc

- Hệ thống được triển khai trên máy tính nhúng Raspberry Pi hoặc Jetson Nano.

- Hệ thống sử dụng mô hình YOLOv8 để nhận diện cử chỉ.

- Hệ thống sử dụng dịch vụ Google Text-to-Speech để chuyển đổi văn bản sang giọng nói.

2.1.6 Kiểm tra và xác minh

- Hệ thống được kiểm tra với tập dữ liệu cử chỉ ngôn ngữ ký hiệu.

- Hệ thống được đánh giá về hiệu suất, độ chính xác và khả năng sử dụng

2.2 Thiết kế hệ thống

2.2.1 Lựa chọn thuật toán

Nhận diện cử chỉ: Sử dụng thuật toán YOLOv8 hoặc các biến thể khác (YOLOv7, YOLOv5,...) để nhận diện cử chỉ ngôn ngữ kí hiệu trong thời gian thực với độ chính xác cao.

Chuyển đổi cử chỉ thành văn bản:Lấy output khi nhận diện kí hiệu xuất ra văn bản.

Chuyển văn bản thành giọng nói:Sử dụng các dịch vụ chuyển văn bản thành giọng nói sử dụng thư viện có sẵn.

Chuyển giọng nói sang văn bản và từ văn bản sang hình ảnh: Sử dụng Google Speech to Text Recognition để dịch từ âm thanh sang văn bản, từ văn bản ta xử lý chuỗi để lấy hình ảnh nhận diện cơ thể cho người khuyết tật có thể hiểu thông điệp từ tệp các hình ảnh nhóm tự thu thập.

Phân tích biểu cảm khuôn mặt: Train tập dữ liệu về biểu cảm khuôn mặt với YOLO, cùng với tập dữ liệu hand gesture.

2.2.2 Thiết kế kiến trúc hệ thống

Module thu thập dữ liệu*:* Sử dụng camera để thu thập hình ảnh hoặc video người dùng thực hiện ngôn ngữ ký hiệu.

Module tiền xử lý: Xử lý hình ảnh đầu vào (cắt, xoay, thay đổi kích thước) và chuẩn hóa dữ liệu.

Module nhận diện cử chỉ: Sử dụng thuật toán YOLO để nhận diện cử chỉ trong hình ảnh/video.

Module ngôn ngữ: Xử lý output sau nhận diện để chuyển đổi cử chỉ thành văn bản.

Module tổng hợp giọng nói: Sử dụng dịch vụ TTS để chuyển đổi văn bản thành giọng nói.

Module phân tích biểu cảm: Sử dụng thuật toán nhận diện khuôn mặt và mô hình học sâu để phân tích biểu cảm.

Module giao diện người dùng: Hiển thị thông tin (văn bản) và tương tác với người dùng.

2.2.3 Triển khai và thử nghiệm

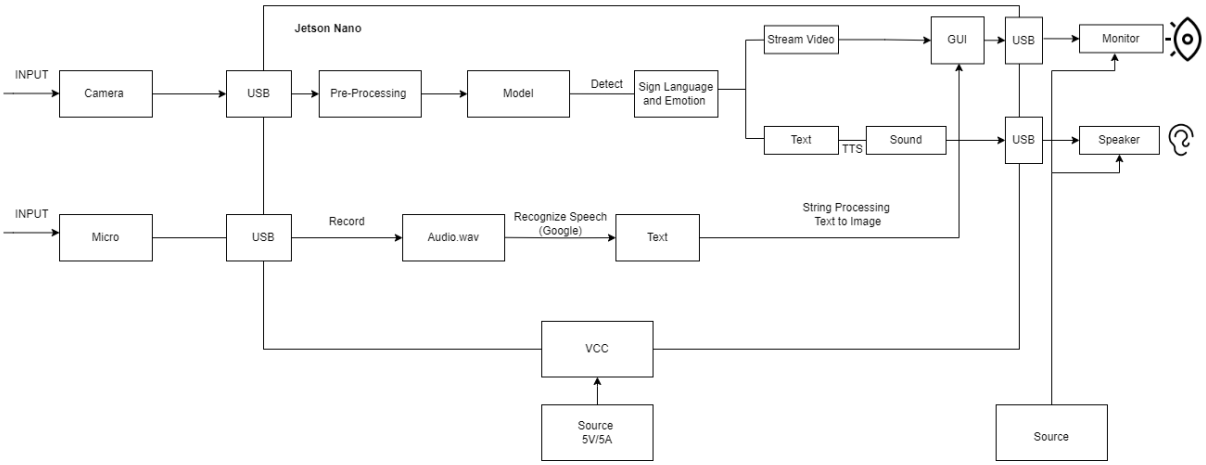
Lựa chọn phần cứng: Sử dụng máy tính nhúng Raspberry Pi hoặc Jetson Nano để triển khai hệ thống.

Phát triển phần mềm: Viết mã cho các mô-đun trong hệ thống sử dụng Python, C++, hoặc ngôn ngữ lập trình phù hợp.

Huấn luyện mô hình*:* Huấn luyện mô hình ngôn ngữ và mô hình phân tích biểu cảm trên tập dữ liệu phù hợp.

Thử nghiệm và đánh giá: Thử nghiệm hệ thống với các trường hợp sử dụng khác nhau và đánh giá hiệu năng, độ chính xác, và khả năng sử dụng.

2.2.4 Sơ đồ khối hệ thống



Hình 2. 1: Sơ đồ khối hệ thống

CHƯƠNG III: SƠ LƯỢC VỀ NGÔN NGỮ CƠ THỂ

3.1 Khái niệm ngôn ngữ cơ thể trong giao tiếp

Chúng ta không thể phủ nhận tầm quan trọng của giao tiếp bằng lời nhưng những buổi diễn thuyết, bữa tiệc hay chỉ đơn thuần là các buổi nói chuyện sẽ trở nên kém hấp dẫn nếu như không có giao tiếp bằng cử chỉ lời nói.

Không phải lúc nào con người ta cũng có thể dùng lời nói để diễn đạt suy nghĩ của mình. Chỉ cần tinh tế một chút trong giao tiếp chúng ta sẽ nhận ra ngay chúng ta không chỉ giao tiếp bằng lời nói mà cả bằng ngôn ngữ cơ thể.

Ngôn ngữ cơ thể có thể giúp đỡ chúng ta rất nhiều trong cuộc sống hàng ngày. Những cử chỉ, hành động, nét mặt, thậm chí cách đi đứng của bạn cũng có thể bộc lộ khá nhiều suy nghĩ, giúp đồng nghiệp và bạn bè hiểu rõ thông điệp bạn muốn truyền đến họ hơn. Đặc biệt, nếu bạn là 1 nhà lãnh đạo hoặc người nổi tiếng, mỗi cử chỉ hành động của bạn sẽ được rất nhiều người chú ý. Có 1 số đã dùng những cử chỉ, hành động độc đáo trở thành đặc trưng cho chính họ. Ngôn ngữ cơ thể là các cử chỉ hành động của cơ thể như nét mặt, cách nhìn, điệu bộ, và khoảng cách giao tiếp.

3.2 Các tín hiệu ngôn ngữ cơ thể

Trong những nét đặc trưng góp phần vào thế mạnh của ngôn ngữ như một công cụ để truyền đạt thông tin là đặc tính tượng trưng, sử dụng những từ ngữ để thay thế cho một điều gì đó vượt xa hơn so với ý nghĩa thực của chúng. Để đánh giá đúng đặc tính này của ngôn ngữ, chúng ta cần nghiên cứu dấu hiệu và biểu tượng và phân biệt chúng.

Một dấu hiệu là bất cứ những gì chúng ta sử dụng để ám chỉ hay nhắc đến như một dấu hiệu, một điều gì đó.Với mỗi một dấu hiệu ngôn ngữ, theo nguyên lí chung của việc thành lập, một dấu hiệu có hai mặt: Mặt biểu hiện (hình thức tín hiệu) và Mặt được biểu hiện (nội dung tín hiệu).

Mặt hình thức của dấu hiệu là những dạng âm thanh khác nhau mà trong quá trình nói năng con người đã thiết lập nên mã cụ thể cho mình, đó chính là đặc trưng âm thanh cụ thể của từng ngôn ngữ.

Còn mặt nội dung (cái được biểu hiện) là những thông tin, những thông điệp về những mảnh khác nhau của thế giới hiện tại mà con người đang sống, hoặc những dấu hiệu hình thức để phân cắt tư duy, phân cắt thực tại.

Mối liên hệ giữa cái biểu hiện và cái được biểu hiện là mối liên hệ rất đặc trưng của ngôn ngữ. Đặc trưng này thể hiện ở chỗ: mỗi một cái biểu hiện luôn chỉ có một cái được biểu hiện tương ứng. Khi mối liên hệ 1 - 1 này bị cắt đứt thì các quá trình giao tiếp sẽ bị ảnh hưởng hoặc không thể thực hiện được...

3.3 Vai trò của ngôn ngữ cơ thể trong giao tiếp

Phương tiện ngôn ngữ đóng vai trò quan trọng trọng giao tiếp nhưng theo kết quả điều tra gần đây, ngôn ngữ được truyền đạt bằng lời nói hay chữ viết chỉ chiếm 20%, 80% còn lại được biểu đạt bằng ngôn ngữ cơ thể.

Nó phản ánh chân thật và đầy đủ các mối quan hệ do đó không chỉ giúp con người hiểu được nhau mà còn giúp hoàn thiện các mối quan hệ đó.

Chúng ta không thể phủ nhận tầm quan trọng của giao tiếp bằng lời nhưng những buổi diễn thuyết, bữa tiệc hay chỉ đơn thuần là các buổi nói chuyện sẽ trở nên kém hấp dẫn nếu như không có giao tiếp cử chỉ.

3.4 Sử dụng ngôn ngữ bằng tay trong giao tiếp của người Khuyết tật

Người khiếm thính cũng muốn được đóng góp cống hiến cho xã hội, ngôn ngữ cử chỉ chính là một phương tiện hiệu quả cho họ.

Hình 3. 1: Ngôn ngữ bằng tay trong giao tiếp của người Khuyết tật

Đây thực sự là loại ngôn ngữ được sử dụng rất phổ biến hiện nay. Không như trước đây chỉ có những người bị khuyết tật mới sử dụng mà bây giờ rất nhiều người bình thường cũng đang học loại hình ngôn ngữ này để có thể giao tiếp với thế giới người khuyết tật.

Điếc" có nghĩa là gì? Nếu bạn hét lớn hết mức có thể, âm thanh đo được khoảng 80 đêxiben. Chỉ những người không thể nghe tiếng hét như vậy mới thực sự được xem là người Điếc. Người bị mất thính lực ít hơn được xem như "nghe kém".

Làm thế nào để giao tiếp với người khiếm thính?

Cách người Khiếm thính giao tiếp thường phụ thuộc vào thời gian bị mất thính lực của họ. Những người sinh ra là người Điếc hoặc mất thính lực trước khi bắt đầu học nói thường sử dụng ngôn ngữ ký hiệu. Những người bị mất thính lực sau khi đã học nói thường sẽ giao tiếp bằng lời nói và đọc tỉnh hiệu môi.

Không nên cho rằng vì một người Điếc có đeo máy trợ thính, anh ta có thể nghe được điều bạn đang nói. Anh ta chỉ có thể nghe được những âm thanh đặc biệt hay tiếng động nên.

Làm thế nào để có thể nhận biết người tôi đang giao tiếp là người Khiếm thính?

Mất thính lực thường được coi như là "khuyết tật ẩn" vì thế có thể không có cách nào biết một người bị mất thính lực nặng. Những người bị điếc sâu có thể không đeo máy trợ thính.

Một vài người Khiếm thính có mang thẻ ghi thông tin vắn tắt về cách giao tiếp với người khiếm thính. Nếu có ai đó đưa cho bạn một trong những cái thẻ như vậy, bạn nên biết rằng người mang thẻ bị mất thính lực và có thể gặp khó khăn khi giao tiếp với bạn.

Lời nói của người Khiếm thính có thể nghe hơi lạ. Âm lượng của giọng nói có thể không thích hợp hay họ phát âm một vài từ nghe rất lạ. Cần nhớ rằng người Khiếm thính không thể nghe giọng nói của chính họ và vài người Khiêm thính đã học nói chưa bao giờ nghe được một từ đơn giản nào cả.

Một cách khác cho thấy một người có thể là người Khiếm thính nếu người đó dùng tay để viết ra những yêu cầu. Những người sử dụng ngôn ngữ ký hiệu không nói chuyện thì thường hay chuẩn bị viết và giấy.

Làm thế nào để giao tiếp với người Khiếm thính?

Trước hết, hãy xem người Khiếm thính đó giao tiếp như thế nào. Nếu họ hỏi bạn bằng lời nói,chắc chắn rằng họ sẽ cần nghe băng đọc tín hiệu môi khi bạn trả lờiHãy nhìn thẳng vào người khiêm thính, nếu nhìn sang chỗ khác người khiếm thính sẽ không thấy môi của bạn. "Nói rõ ràng chậm rãi, Đừng hét to"`. Bảo đảm rằng phía sau lưng bạn không có ánh đèn sáng chói có thể làm cho người khiếm thính khó nhìn thấy khuôn mặt của bạn. Nên nói cả câu hơn là trả lời từng từ một – 70% việc đọc tín hiệu môi là đoán và nhiều từ trông rất giống nhau. Nói cả câu giúp đoán được nội dung. Hãy kiên nhẫn, nếu được yêu cầu lặp lại, hãy cố gắng chuyển giọng một cách nhẹ nhàng, điều này giúp người khiếm thính hiểu dễ dàng hơn.

Nếu người khiếm thính vẫn chưa hiểu, đừng bỏ cuộc, hãy viết ra giấy. Với người Điếc sử dụng ngôn ngữ ký hiệu, họ vẫn có thể muốn nghe bằng đọc tín hiệu môi. Đáng buồn là có rất ít người nghe biết ngôn ngữ ký hiệu và người Điếc lại quen với cách cổ gắng giao tiếp với người nghe.

Ngoài những vấn đề trên, cần lưu ý thêm: Hãy cố gắng sử dụng bảng chữ cái ngôn ngữ ký hiệu đánh vần bằng tay bất cứ tên gọi hay những từ không thông thường nào. (Xem bảng chữ cái). Sử dụng ngôn ngữ cử chỉ giải thích điều bạn muốn nói. Ví dụ, dùng bàn tay thể hiện kích thước và hình dạng hoặc thể hiện chiều hướng bằng cách chỉ, có thể rất hữu dụng. Sử dụng nét mặt để diễn tả nội dung.

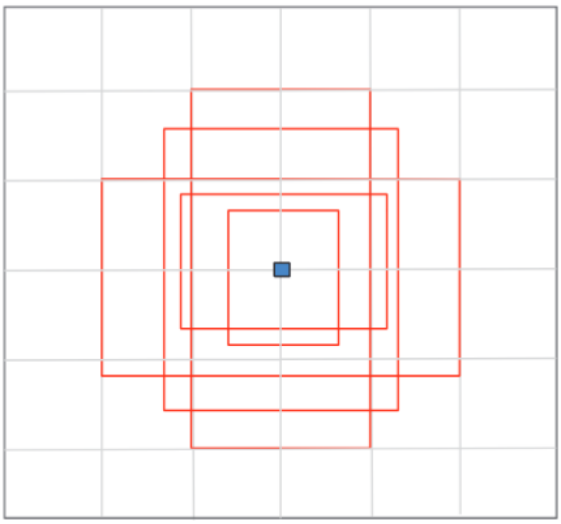
CHƯƠNG IV: SƠ LƯỢC VỀ MÔ HÌNH YOLO

4.1 Mạng YOLO

YOLO là một thuật toán phát hiện đa đối tượng nhanh chóng, theo thời gian thực. YOLO bao gồm một mạng lưới thần kinh tích chập duy nhất dự đoán đồng thời các hộp giới hạn và xác suất lớp của các đối tượng bên trong chúng. YOLO đào tạo trên hình ảnh đầy đủ và mạng được thiết lập để giải quyết các vấn đề hồi quy nhằm phát hiện các đối tượng. Do đó, YOLO không cần một quy trình xử lý phức tạp nên tốc độ xử lý cực kỳ nhanh.

4.1.1 Anchors

Anchors là một hoặc nhiều hình chữ nhật được đặt tại mỗi điểm chập của feature map. Trong Hình 4.1, có năm anchors hình chữ nhật (được hiển thị bằng đường viền màu đỏ) được đặt tại một điểm (được hiển thị bằng màu xanh lam).

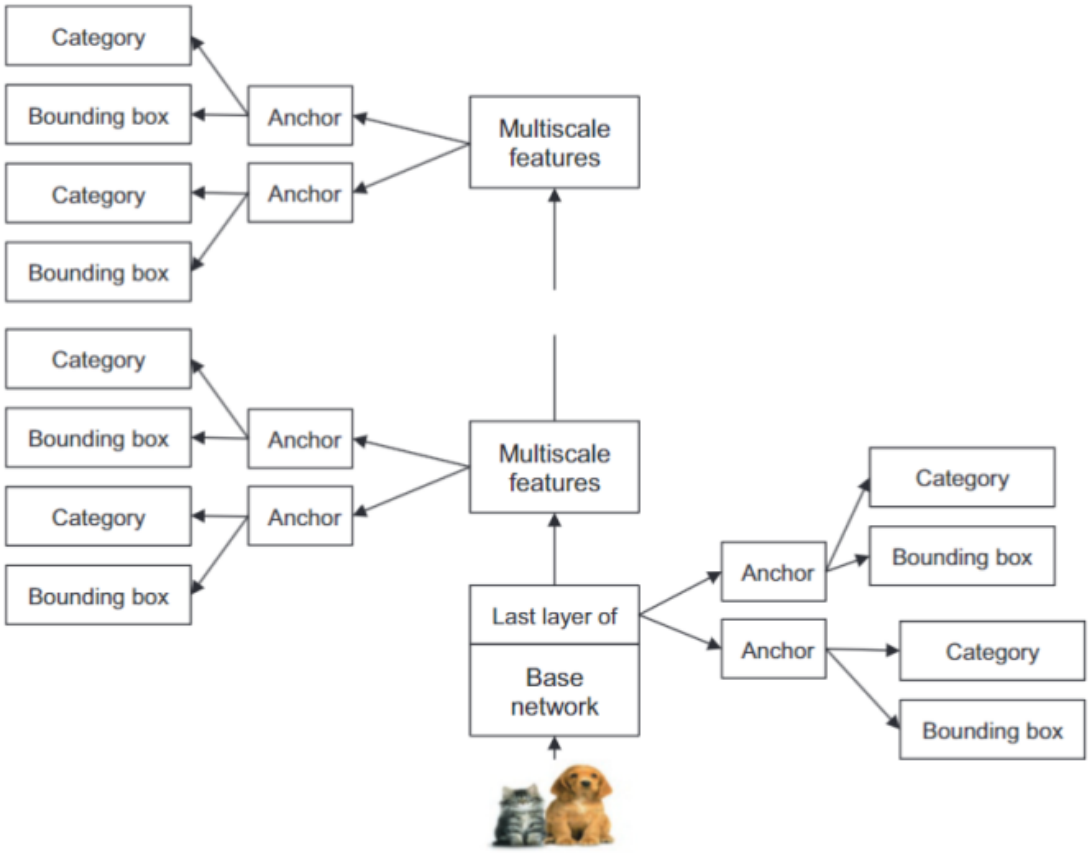


Hình 4. 1: Anchors Boxs

Khi thuật toán YOLO xử lý một hình ảnh, nó sẽ áp dụng các anchor box lên hình ảnh. Sau đó, thuật toán sẽ dự đoán xác suất mỗi anchor box chứa một đối tượng và loại đối tượng đó là gì. Nếu xác suất dự đoán đủ cao, thuật toán sẽ coi đó là một đối tượng và hiển thị nó trong hình ảnh.

4.1.2 Nhận diện đa tỷ lệ Feature Map

Các lớp chập gắn vào cuối mạng cơ sở được thiết kế sao cho các lớp này giảm kích thước dần dần. Điều này cho phép nó dự đoán các đối tượng ở nhiều tỷ lệ. Ta có thể hình dung chúng như Hình 4.2 bên dưới:

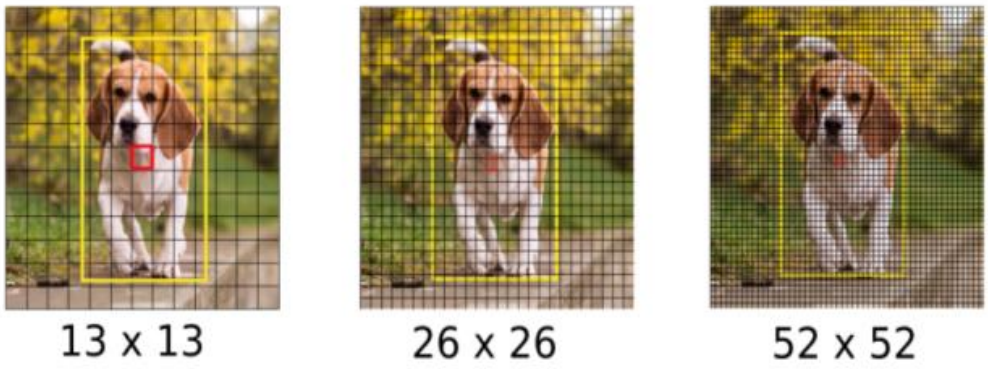


Hình 4. 2: Các lớp tích chập có kích thước giảm dần để dự đoán các đối tượng và các bounding boxes theo tỉ lệ.

Như hình dưới, mỗi lớp nhận diện, lớp cuối cùng của lớp cơ sở dự đoán độ lệch của bốn toạ độ của bounding boxes và object class categories. Các bounding boxes và vật thể được dự đoán qua anchor boxes.

Hình 4.3 mô phỏng việc Feature map chia nhỏ dần qua các output, điều này giúp mô hình nhận diện được các object có kích thước lớn, các output có

feature map chia nhỏ hơn giúp nhận diện các object nhỏ hơn trong khi anchor box vẫn giữ nguyên.



Hình 4. 3: Feature map chia nhỏ dần

Feature map được tạo ra từ các lớp tích chập của mạng nơ-ron, trong đó mỗi lớp sẽ tạo ra một feature map khác nhau. Các feature map có kích thước khác nhau tùy thuộc vào cấu trúc của mạng nơ-ron và các tham số của quá trình huấn luyện. Feature map cuối cùng trong mạng YOLO thường chứa thông tin chi tiết về các đối tượng trong hình ảnh cùng với vị trí và độ tin cậy của chúng. Được sử dụng để dự đoán các bounding box và các lớp của đối tượng trong hình ảnh.

4.1.3 Tiêu chí đánh giá dữ liệu và Non-Maximum Suppression (NMS)

NMS là một kỹ thuật quan trọng quan trọng được sử dụng trong YOLO để cải thiện độ chính xác và hiệu quả của việc phát hiện đối tượng.

Avarage Precision (AP), hay còn được thường gọi là Mean Average Precision (mAP), là những thông số thường được dùng để đánh giá hiệu xuất của mô hình. Nó đo độ chính xác trung bình trên tất cả các cụm, cung cấp số liệu để có thể so sánh với các model.

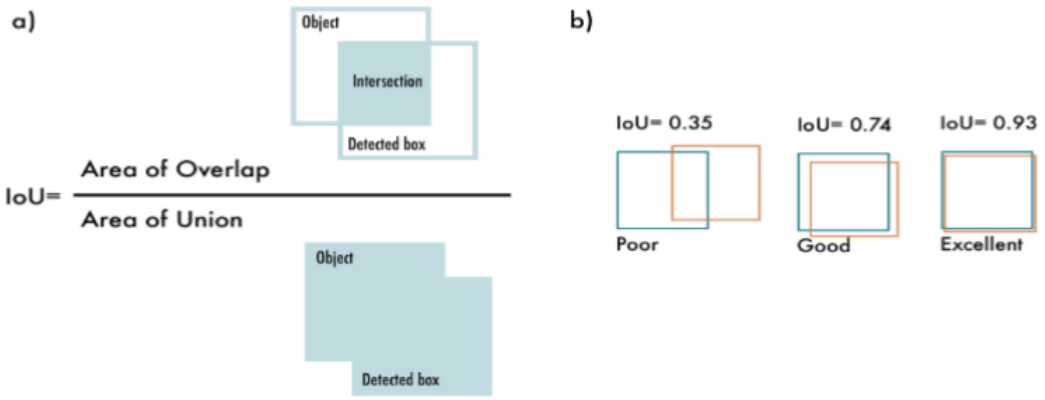
Cách AP/mAP hoạt động:

AP/mAP hoạt động dựa trên chỉ số precision-recall, xử lý nhiều loại đối tượng và dự đoán dựa trên IoU.

Precision và Recall: Độ chính xác (Precision) đo mức độ chính xác của các dự đoán dương tính của mô hình, trong khi độ nhớ lại (Recall) đo tỷ lệ các trường hợp dương tính thực tế mà mô hình xác định đúng. Thường có sự đánh đổi giữa độ chính xác và độ nhớ lại; ví dụ, việc tăng số đối tượng được phát hiện (độ nhớ lại cao hơn) có thể dẫn đến nhiều dương tính giả (độ chính xác thấp hơn).

Xử lý nhiều dạng vật thể: Mô hình nhận dạng vật thể cần xác định và định vị nhiều loại đối tượng khác nhau trong một bức ảnh. Chỉ số AP (Độ chính xác trung bình) giải quyết vấn đề này bằng cách tính toán riêng biệt độ chính xác trung bình (AP) của từng loại, sau đó lấy giá trị trung bình của các AP này trên tất cả các loại (đó là lý do tại sao nó còn được gọi là độ chính xác trung bình của trung bình). Cách tiếp cận này đảm bảo hiệu suất của mô hình được đánh giá cho từng loại riêng lẻ, cung cấp một đánh giá toàn diện hơn về hiệu suất tổng thể của mô hình.

IoU:Nhận dạng đối tượng có mục đích định vị chính xác các đối tượng trong ảnh bằng cách đặt các bounding boxes. AP kết hợp với IoU để đánh giá mức độ chính xác của bounding boxes, IoU là tỉ lệ giữa phần diện tích bị chồng lắp giữa object và bounding box so với phần giới hạn thực tế.



Hình 4. 4: Intersection of Union. a) IoU tính tỷ lệ giữa phần chồng lắp giữa phần object boxes và phần nhận diện so với toàn bộ b) Đánh giá các vị trí vùng nhận diện khác nhau với các IoU khác nhau

Hình ảnh a: Hộp giới hạn dự đoán (xanh lá) chỉ che một phần nhỏ hộp giới hạn thực tế (đỏ). IoU thấp, gần bằng 0, thể hiện dự đoán sai vị trí và kích thước đối tượng.

Hình ảnh b: Hộp giới hạn dự đoán (xanh lá) bao phủ phần lớn hộp giới hạn thực tế (đỏ). IoU cao, gần bằng 1, thể hiện dự đoán vị trí và kích thước đối tượng tương đối chính xác.

IoU là một chỉ số quan trọng để đánh giá độ chính xác của mô hình phát hiện đối tượng. Nó cung cấp một cách đơn giản và trực quan để đo lường mức độ trùng lặp giữa hộp giới hạn dự đoán và hộp giới hạn thực tế. IoU được sử dụng rộng rãi trong các ứng dụng phát hiện đối tượng, bao gồm đánh giá hiệu suất mô hình, lựa chọn hộp giới hạn tốt nhất và hệ thống theo dõi đối tượng.

4.2 Các thành phần cơ bản của một mô hình

4.2.1 Backbone

Backbone là phần "xương sống" của mô hình học máy, đóng vai trò trích xuất đặc trưng từ dữ liệu đầu vào. Nó thường sử dụng các mạng nơ-ron tích chập (CNN) được đào tạo sẵn như ResNet, VGGNet, MobileNet. Quá trình trích xuất đặc trưng của backbone sẽ tạo ra các bản đồ đặc trưng với kích thước và độ phân giải khác nhau

4.2.2 Neck

Neck là phần "cổ" của mô hình, có nhiệm vụ kết hợp các bản đồ đặc trưng từ backbone. Nó thường sử dụng các mô-đun như FPN (Feature Pyramid Network) hoặc PAN (Path Aggregation Network) để kết hợp thông tin từ các bản đồ đặc trưng ở nhiều độ phân giải khác nhau. Việc kết hợp này giúp cải thiện hiệu quả của mô hình trong việc nhận diện các đối tượng có kích thước khác nhau.

4.2.3 Head

Head là phần "đầu" của mô hình, thực hiện các dự đoán cuối cùng. Nó thường sử dụng các mô-đun phân loại và hồi quy để dự đoán vị trí và loại của các

đối tượng trong ảnh. Có thể có nhiều head khác nhau cho các nhiệm vụ khác nhau, ví dụ như head cho phân loại đối tượng, head cho phát hiện điểm ảnh, v.v.

4.3 Kiến trúc mô hình

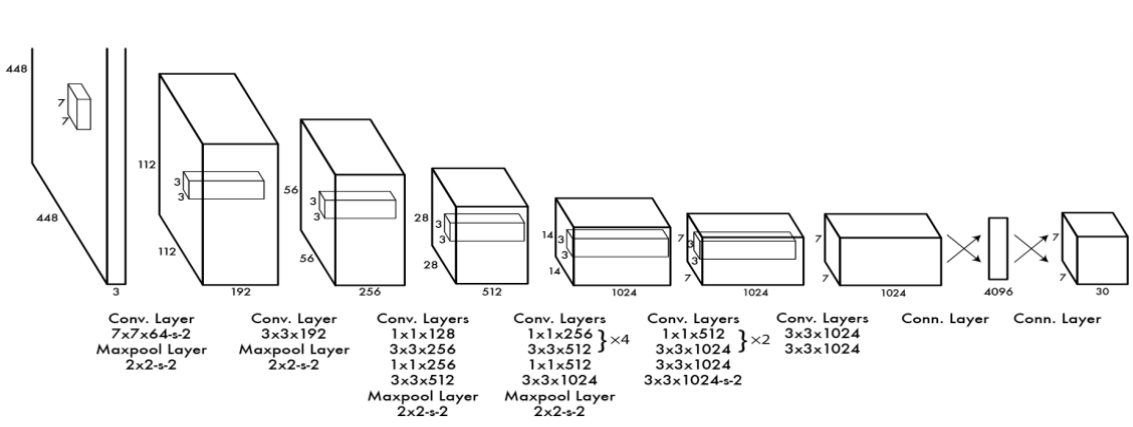
Việc xuất bản: Phát hiện đối tượng theo thời gian thực" thống nhất được đề xuất lần đầu tiên bởi Redmon và cộng sự vào năm 2015, một trong những thuật toán phát hiện đối tượng phổ biến nhất, YOLOv1, lần đầu tiên được mô tả là có cách tiếp cận " hoàn toàn đơn giản".

Khi mới bắt đầu, YOLOv1 có thể xử lí hình ảnh ở tốc độ 4 khung hình/giây, trong khi một biến thể YOLO nhanh, có thể đạt tốc độ lên tới 155 khung hình/giây. Nó cũng đạt được mAP cao so với các thuật toán phát hiện đối tượng vào thời điểm đó.

Đề xuất chính của YOLO là coi việc phát hiện đối tượng là một vấn đề hồi quy một lần. YOLOv1 bao gồm một neural network duy nhất, dự đoán các bounding boxes và xác suất lớp liên quan trong một đánh giá duy nhất. Mô hình cơ bản của YOLO hoạt động bằng cách trước tiên chia hình ảnh đầu vào thành lưới SxS trong đó mỗi ô (i,j) bounding boxes B, điểm tin cậy cho mỗi box và xác suất của lớp C. Đầu ra cuối cùng sẽ là một tensor có hình dạng SxSx(Bx5+C).

4.3.1 YOLOv1

Kiến trúc YOLOv1 bao gồm 24 convolutional layer theo sau là 2 fully connected layer .Tại đây lấy 20 convolutional layer từ backbone của mạng và cùng với việc bổ sung average pooling layer và fully connected layer duy nhất, nơi nó được huấn luyện trước và xác thực trên bộ dữ liệu ImageNet. Trong quá trình suy luận, 4 lớp cuối cùng và 2 lớp fully connected layer được thêm vào mạng; tất cả đều khởi tạo ngẫu nhiên.



Hình 4. 5: Kiến trúc YOLOv1

YOLOv1 sử dụng phương pháp Gradient Descent làm công cụ tối ưu hóa; Hàm loss được hiển thị ở đây. Hàm Loss bao gồm 2 phần, localization loss (Hàm loss cục bộ) và classification loss (loss phân loại). Sự mất mát nội địa hóa đo lường lỗi giữa tọa độ bounding boxes được dự đoán và hộp giới hạn thực tế. Sự mất mát phân loại đo lường sai số của lớp được dự đoán và sự thật cơ bản.

4.3.2 YOLOv5

YOLOv5 là mô hình phát hiện đối tượng được Ultralytic phát triển , người sáng tạo ra YOLOv1 và YOLOv3 ban đầu , giới thiệu vào năm 2020. YOLOv5 đặt được hiệu suất SOTA trên tập dữ liệu chuẩn COCO. Đồng thời huấn luyện và triển khai nhanh chóng và hiệu quả. YOLOv5 đũa thực hiện một số thay đổi về mặt kiến trúc, đáng chú ý nhất là phương pháp tiêu chuẩn hóa mô hình 3 thành phần,Backbone, neck ,head.

Backbone của YOLOv5 là Darknet53, một kiến trúc mạng tập trung vào việc trích xuất các tính năng được đặc trưng bởi các cửa sổ lọc nhỏ(small filter windows) và các kết nối còn lại(residual connections). Kết nối một phần qua từng giai đoạn cho phép kiến trúc đạt được luồng Gradient phong phú hơn đồng thời giảm tính toán như mô tả do Wang và cộng sự đề xuất.

Neck được mô tả bởi Teven và cộng sự, của YOLOV5 kết nối backbone với head, mục đích là tổng hợp và tinh chỉnh các đặc điểm được trích xuất bởi backbone, tập trung vào việc nâng cao thông tin không gian và ngữ nghĩa trên các quy mô khác nhau. Module nhóm kim tự tháp không gian loại bỏ rằng buộc kích thước cố định của mạng, giúp loại bỏ nhu cầu làm cong, tăng cường hoặc cắt xén hình ảnh. Tiếp đến là module mạng tổng hợp đường dẫn CSP, kết hợp các tính năng đã học trong Backbone và rút ngắn đường thông tin giữa các lớp thấp hơn và cao hơn.

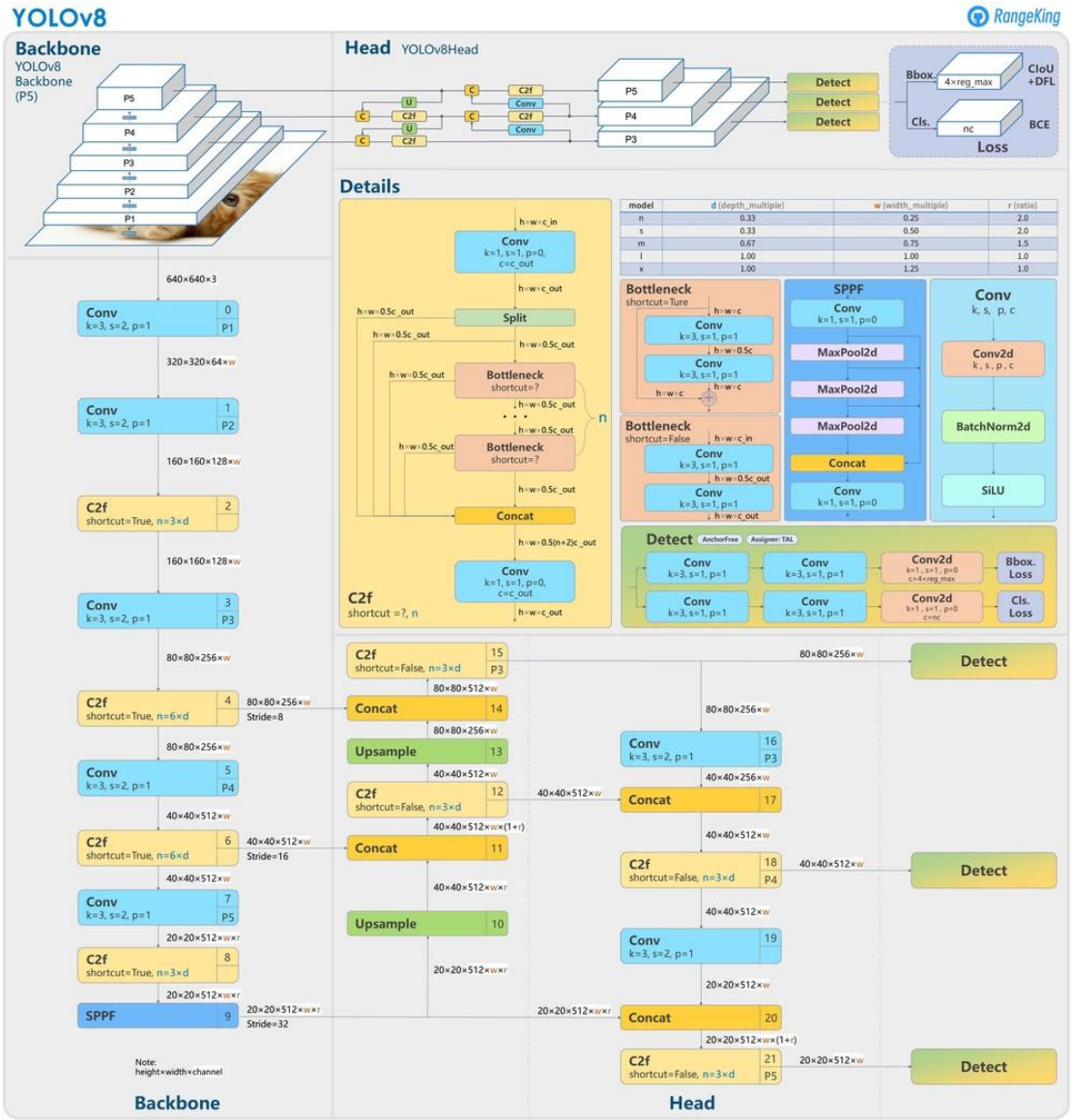
Head của YOLOv5 bao gồm 3 nhánh, mỗi nhánh dự đoán một thang đo tính năng khác nhau. Trong ấn phẩm ban đầu của mô hình, người sáng tạo đã sử dụng kích thước ô lưới 13x13, 26x26 và 52x52, mỗi ô cell dự đoán B=3 hộp giới hạn. Mỗi điểm đầu tạo ra các hộp giới hạn, xác suất của lớp và điểm tin cậy. Cuối cùng sử dung Non-maximum Suppression (NMS)(mạng sử dụng Ngăn chặn không tối đa) để học các hộp chồng chéo.

YOLOv5 kết hợp các hộp anchor box, các hộp đóng khung có kích thước cố định để dự đoán vị trí và kích thước của vật thể trong hình ảnh. Thay vì dự đoán giới hạn tùy ý các hộp cho từng phiên bản đối tượng, mô hình dự đoán tọa độ của các hộp anchor box với tỷ lệ khung hình được xác định trước và chia tỉ lệ và điều chỉnh chúng để phù hợp với thể hiện của đối tượng.

4.3.3 YOLOv8

YOLOv8 là phiên bản mới nhất của mô hình phát hiện đối tượng YOLO. Phiên bản mới này có kiến trúc tương tự như phiên bản trước đó, nhưng nó có thêm nhiều cải tiến so với các phiên bản trước của YOLO chẳng hạn như phiên bản mới kiến trúc mạng neural Network, sử dụng cả mạng Feature Pyramid Network (FPN) và Path Aggregation Network (PAN) và công cụ ghi nhãn mới giúp đơn giản hóa quá trình chú thích . Công cụ ghi nhãn này chứa một số tính năng hữu ích như tự động ghi nhãn , các phím tắt ghi nhãn và các phím nóng có thể tùy chỉnh. Sự kết hợp các tính năng này giúp việc chú thích hình ảnh phục vụ

cho việc huấn luyện mô hình trở nên dễ dàng hơn. FPN hoạt động bằng cách giảm dần độ phân giải không gian của hình ảnh đầu vào và đồng thời tăng số lượng các kênh chức năng. Điều này dẫn đến việc tạo ra các bản đồ đặc trưng có khả năng phát hiện các vật thể ở khoảng cách khác nhau, quy mô và độ phân giải. Mặt khác kiến trúc PAN, tổng hợp các tính năng từ các cấp độ khác nhau của mạng thông qua việc bỏ qua kết nối. Bằng cách làm như vậy, mạng có thể nắm bắt các đặc điểm tốt hơn ở nhiều tỉ lệ và độ phân giải, điều này rất quan trọng để phát hiện chính xác các đối tượng kích cỡ và hình dạng khác nhau.



Hình 4. 6: Kiến trúc YOLOv8

Cấu trúc mạng nơ-ron YOLOv8 (Hình 4.6), một mô hình phát hiện đối tượng hiệu quả được phát triển bởi Ultralytics. YOLOv8 là phiên bản nâng cấp của YOLOv5, với nhiều cải tiến về hiệu suất và tốc độ.

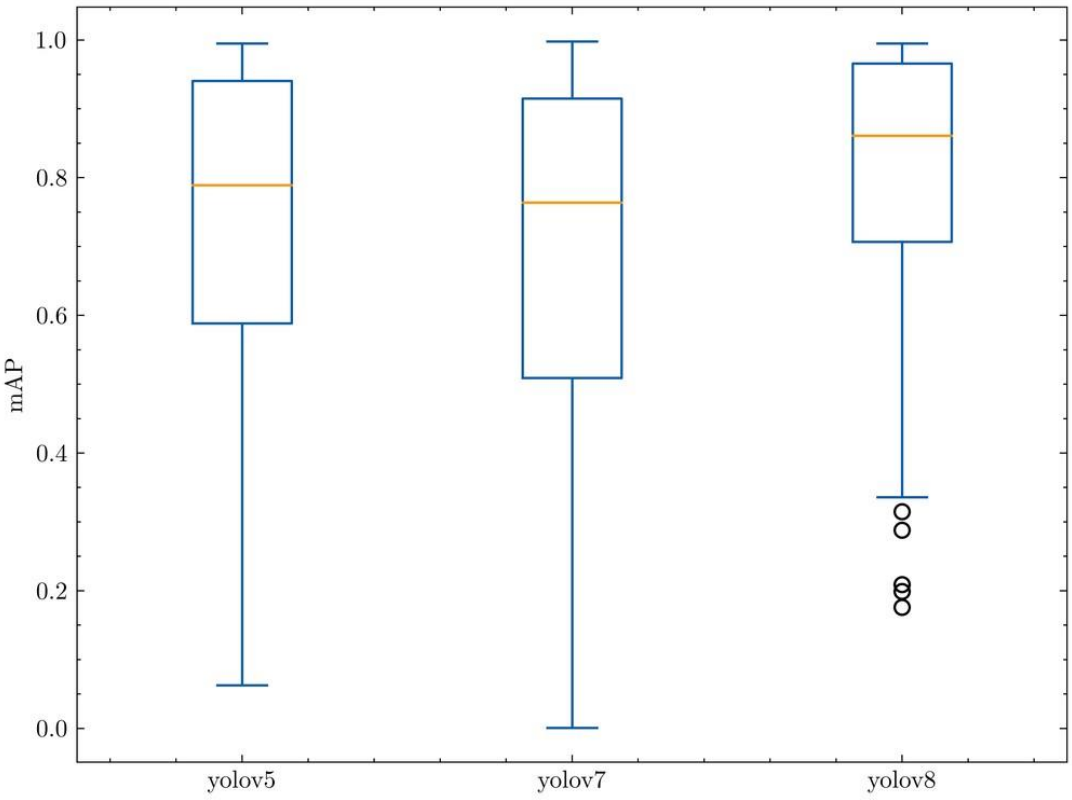
Cấu trúc tổng thể của YOLOv8 bao gồm hai phần chính: **Backbone** và **Head.**

Backbone: Phần này dùng để trích xuất các đặc trưng từ hình ảnh đầu vào. Trong YOLOv8, Backbone sử dụng kiến trúc CSPNet (Cross Stage Partial connections) với các cải tiến như Cấu trúc Bottleneck để giảm số lượng tham số và tăng hiệu quả tính toán, cùng với SPPF (Spatial Pyramid Pooling Feature) để tăng cường khả năng trích xuất đặc trưng đa kích thước.

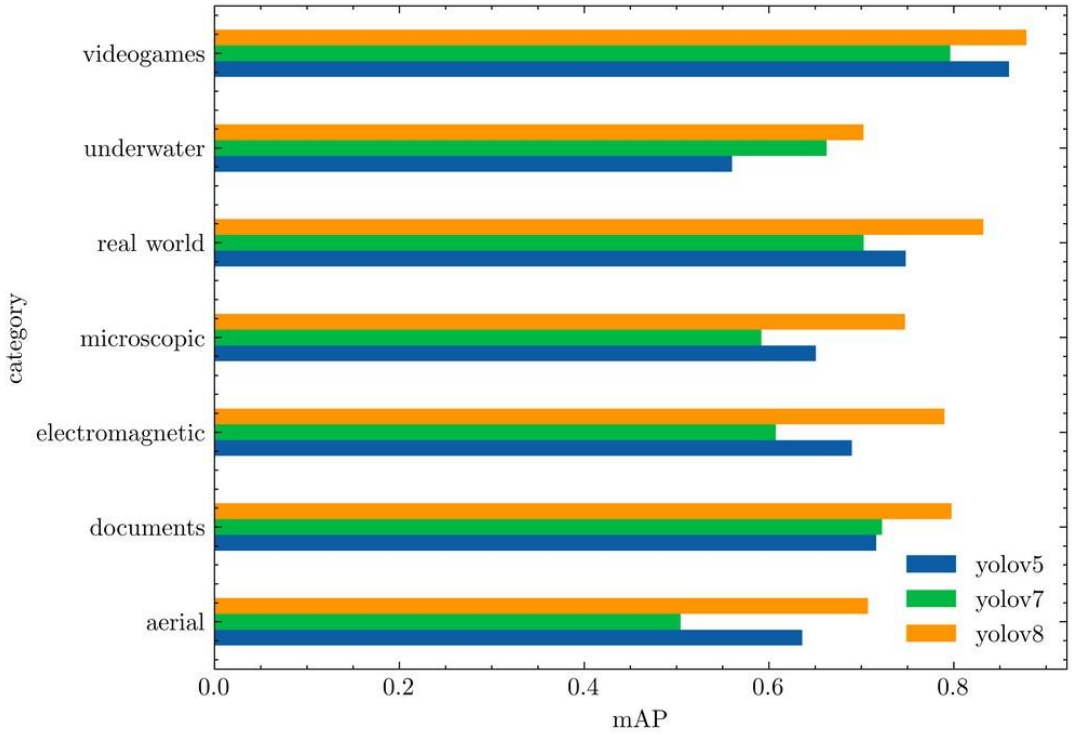
Head: Phần này có trách nhiệm dự đoán vị trí và lớp đối tượng trong hình ảnh. Head bao gồm YOLOv8Head để dự đoán hộp giới hạn và lớp đối tượng, cùng với quá trình Detect để xử lý kết quả dự đoán và tạo ra hộp giới hạn cuối cùng cho các đối tượng được phát hiện.

4.3.4 So sánh giữa hai mô hình tương đồng YOLOv5 và YOLOv8

Lí do mà YOLOv8 bị so sánh với YOLOv5 và không phải bất kì phiên bản YOLO nào khác mà là YOLOv5, hiệu suất và số liệu gần với YOLOv8 hơn. Tuy nhiên YOLOv8 vượt trội hơn YOLOv5 hơn khi chúng ta nói về mAP (hình 4.7), điều này cho thấy YOLOv8 có outlier hơn khi được đo dựa trên RF100 là 100 mẫu tập dữ liệu từ Robotflow là một kho lưu trữ dữ liệu của 100000 bộ dữ liệu. ta cũng thấy được rằng YOLOv8 vượt trội hơn YOLOv5 cho từng loại RF100. Từ hình chúng ta có thể thấy rằng YOLOv8 tạo ra kết quả tương tự hoặc tốt hơn YOLOv5 (hình 4.8 .).



Hình 4. 7: Biểu đồ dạng Box Plot trên thể hiện sự so sánh giá trị mAP của YOLOv8 cao hơn so với các phiên bản tiền nhiệm là YOLOv7 và YOLOv5



Hình 4. 8 :Hiệu suất của các mô hình phát hiện đối tượng khác nhau trên tập dữ liệu Roboflow 100

Tập dữ liệu này bao gồm 100 hình ảnh về các đối tượng khác nhau, và các mô hình được đánh giá dựa trên khả năng xác định và định vị chính xác các đối tượng này. Biểu đồ hiển thị độ chính xác trung bình (mAP) của mỗi mô hình, đây là thước đo hiệu suất tổng thể của mô hình. mAP càng cao, mô hình càng hoạt động tốt.

Như bạn có thể thấy, mô hình YOLOv8 vượt trội so với các mô hình khác trên tập dữ liệu Roboflow 100. Điều này có nghĩa là YOLOv8 có thể xác định và định vị chính xác các đối tượng trong hình ảnh với độ chính xác cao hơn so với các mô hình khác.

Một điểm khác biệt nữa của hai mô hình là quá trình huấn luyện dữ liệu. YOLOv8 đã được huấn luyện trên phạm vi rộng hơn và đa dạng hơn tập dữ liệu so với YOLOv5. YOLOv8 đã được huấn luyện trên một sự kết hợp giữa tập dữ liệu COCO và một số tập dữ liệu khác, trong khi YOLOv5 được huấn luyện chủ yếu trên bộ dữ liệu COCO. Vì lí do đó, YOLOv8 có hiệu suất tốt hơn trên phạm vi rộng hơn trên các loạt hình ảnh.

YOLOv8 bao gồm công cụ ghi nhãn mới có tên Roboflow Annotate được sử dụng để chú thích hình ảnh với đói tượng nhiệm vụ phát hiện hình ảnh để huấn luyện mô hình dễ dàng hơn và bao gồm một số tính năng như ghi nhãn tự động, ghi nhãn phím tắt và phím nóng có thể tùy chỉnh, Ngược lại YOLOv5 sử dụng một công cụ ghi nhãn khác có tên là LabelImg. LabelImg là một công cụ chú thích hình ảnh đồ họa mã nguồn mở cho phép người dùng của nó vẽ các hộp giới hạn xung quanh đối tượng quan tâm trong một hình ảnh, sau đó xuất các chú thích trong YOLO dạng để huấn luyện mô hình.

YOLOv8 bao gồm các kỹ thuật xử lí hậu kì tiên tiến hơn YOLOv5, đây là một tập hợp các thuật toán được áp dung cho các hộp giới hạn được dự đoán và tính khách quan điểm số được tạo ra bởi mạng lưới thần kinh.Những Kỹ thuật này giúp tinh chỉnh các kết quả phát hiện, loại bỏ các phát hiện dư thừa và cải thiện độ chính xác tổng thể của các dự đoán. YOLOv8 sử dụng Soft-NMS, một

biến thể của kỹ thuật NMS được sử dụng trong YOLOv5. Soft-NMS áp dụng phần mềm ngưỡng cho các hộp giới hạn chồng chéo thay vì loại bỏ chúng hoàn toàn. Trong khi đó NMS loại bỏ các hộp giới hạn chồng chéo và chỉ giữ lại những hộp có điểm khách quan cao nhất.

Trong kiến trúc YOLO thường có một số đầu ra, đứng đầu chịu trách nhiệm dự đoán các khía cạnh khác nhau của đối tượng được phát hiện, chẳng hạn như tọa độ hộp giới hạn, xác suất của lớp và điểm số khách quan. Những đầu ra này thường được kết nối với một vài lớp cuối cùng của mạng Neural và được huấn luyện để đưa ra một tập hợp các giá trí có thể dự dụng để phân loại các đối tượng trong một hình ảnh. Các số lượng và loại đầu ra được sử dụng khác nhau tùy theo về thuật toán phát hiện đối tượng cụ thể và các yêu cầu của nhiệm vụ hiện tại. YOLOv5 có 3 đầu ra trong khi YOLOv8 có 1 đầu ra . YOLOv8 không có các anchors cell, vừa và lớn thay vì sử dụng cơ chế phát hiện không có neo dự đoán trực tiếp tâm của một đối tượng thay vì phần bù từ hộp anchor đã biết, điều này làm giảm số lượng hộp dự đoán và giúp tăng tốc quá trình xử lí hậu kì.

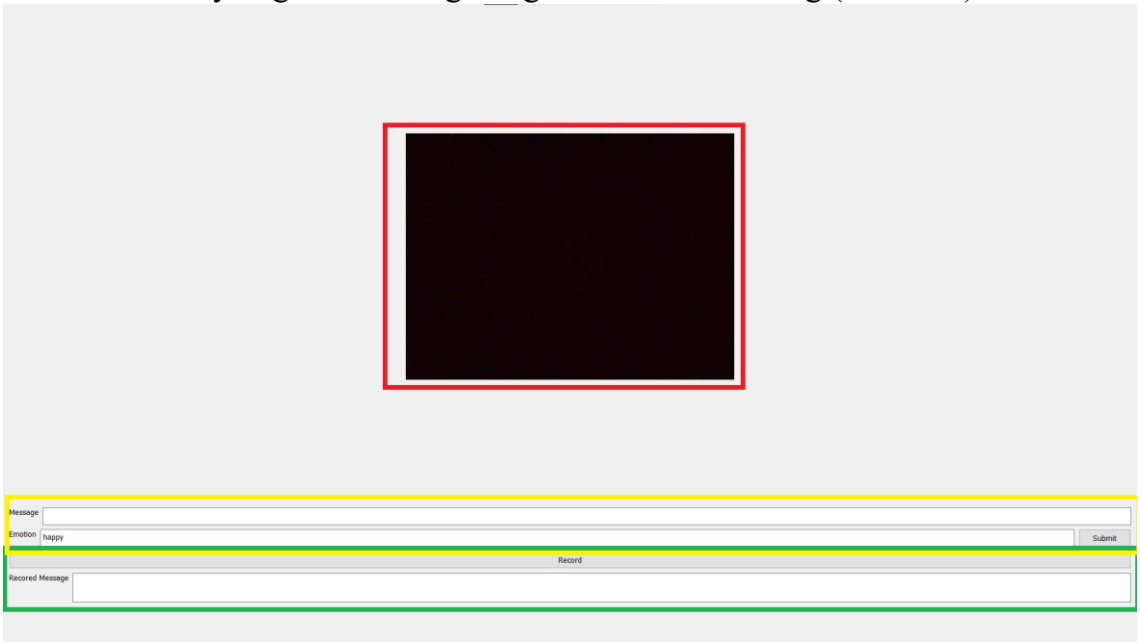
Công bằng mà nói thì YOLOv8 chậm hơn một chút so với YOLOv5 liên quan tới tốc độ phát hiện đối tượng. Tuy nhiên YOLOv8 vẫn có thể xử lí hình ảnh theo thời gian thực trên các GPU hiện đại.

Cả YOLOv5 và YOLOv8 đều sử dụng khả năng tăng cường khảm trên tập huấn luyện. Tăng cường khảm là một kỹ thuật tăng cường dữ liệu lấy bốn hình ảnh ngẫu nhiên từ tập huấn luyện và kết hợp chúng thành một hình ảnh khảm duy nhất. Hình ảnh này, trong đó mỗi góc phần tư chứa một phần cắt ngẫu nhiên từ một trong bốn hình ảnh đầu vào, sau đó được sử dụng làm đầu vào cho mô hình.

CHƯƠNG V: KẾT QUẢ VÀ THỰC NGHIỆM

5.1 Giao diện

Dưới đây là giao diện ứng dụng khi bắt đầu sử dụng (Hình 5.1)



Hình 5. 1: Giao diện phần mềm giao tiếp giữa người binh thường và người khuyết tật

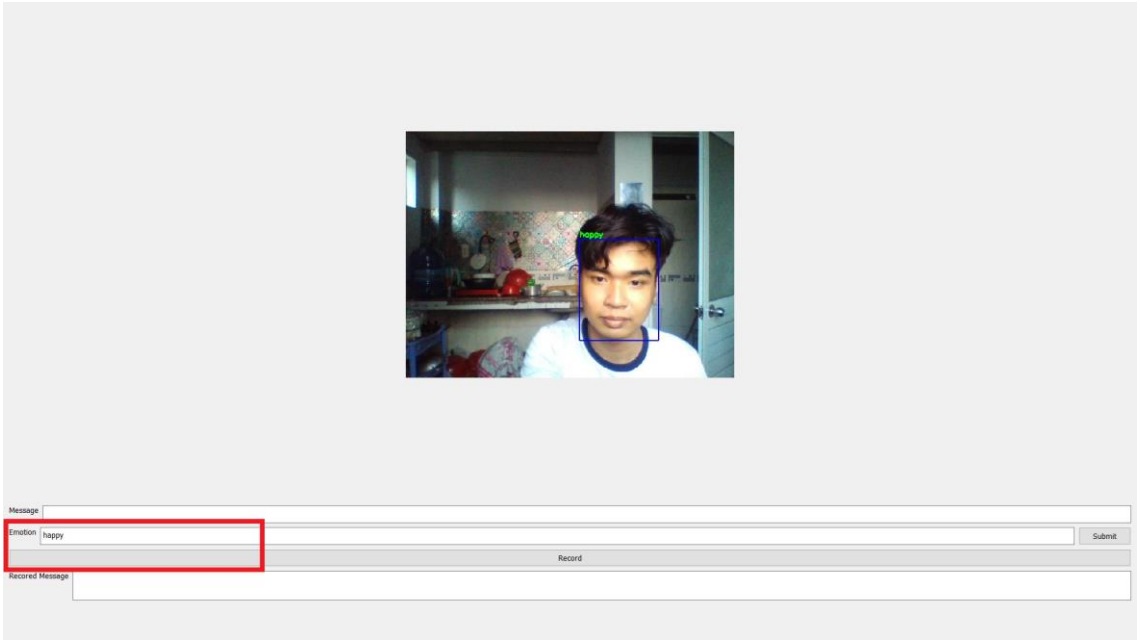
Khung màu đỏ: Thể hiện cho camera thực hiện việc nhận dạng các ngôn ngữ bằng tay và cảm xúc.

Khung màu vàng: Thể hiện cho việc nhận dạng các ký hiệu tay của người khuyết tật để chuyển sang dạng Text cho người bình thường có thể hiểu.

Khung màu xanh: Thể hiện cho chức năng ghi lại âm thanh của người bình thường. Sau đó sẽ chuyển thành các ký hiệu tay để giao tiếp với người khuyết tật. Các ký hiệu này sau đó sẽ xuất hiện lên màn hình.

5.2 Kết quả và thực nghiệm

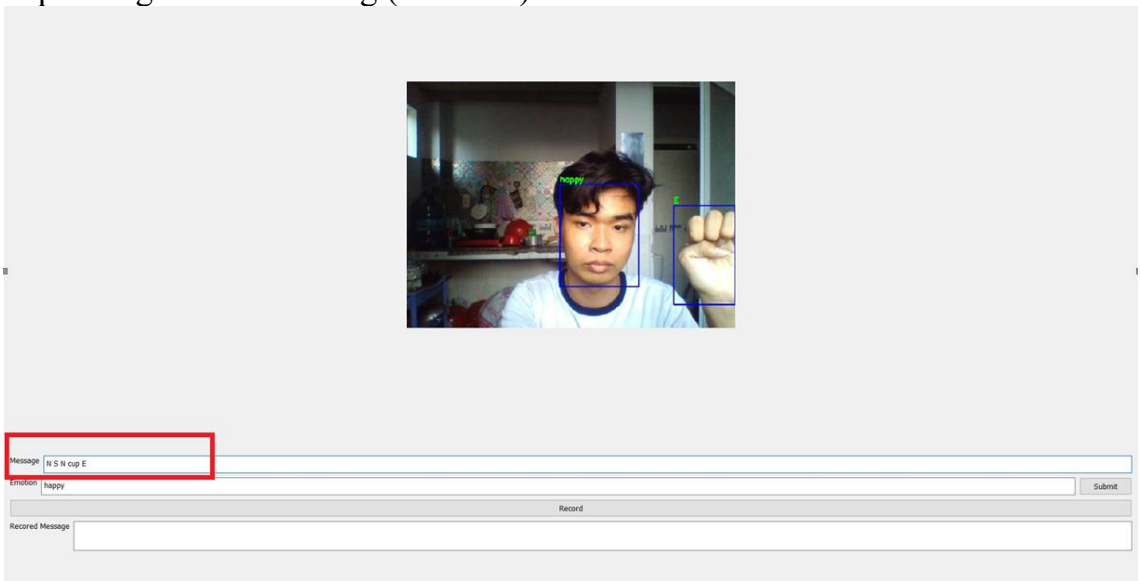
Dưới dây là hình ảnh nhận diện Emotion của đối tượng (Hình 5.2)



Hình 5. 2: Nhận diện cảm xúc của đối tượng

Trong quá trình nhận diện, Camera sẽ liên tục cập nhật hình ảnh và nhận diện cảm xúc của đối tượng trong khung hình. Tại đây đối tượng đang có cảm xúc là "Happy" nên sẽ nhận diện Emotion là "Happy"

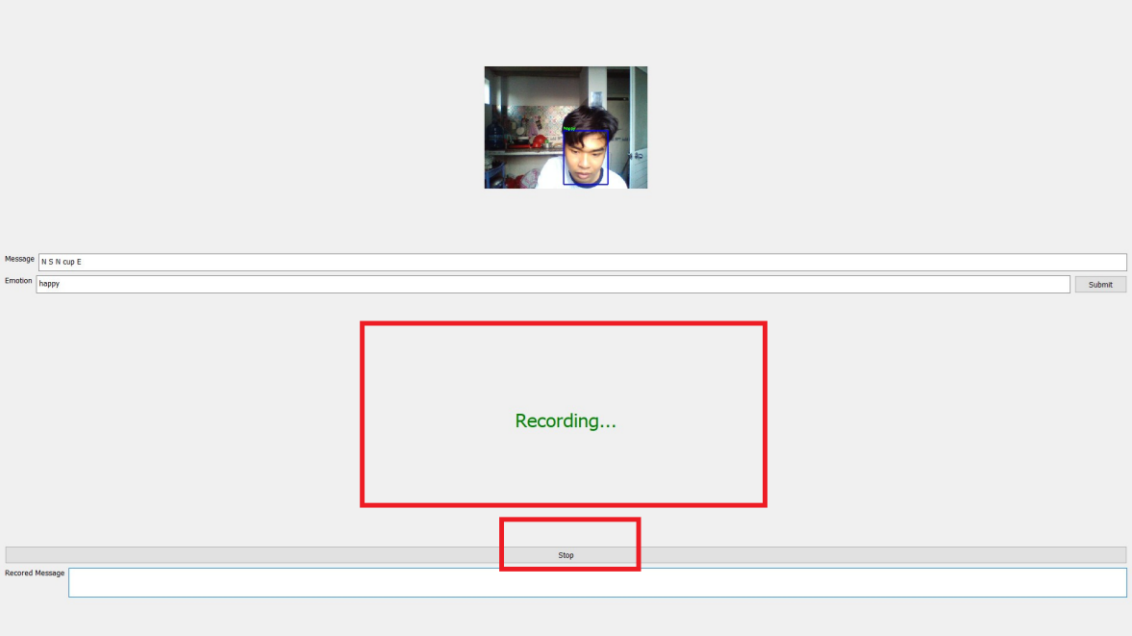
Kế đến, Nhận diện các ký hiệu tay để chuyển về dạng Văn bản để giao tiếp với người bình thường (Hình 5.3)



Hình 5. 3: Nhận diện các ký hiệu tay để chuyển sang dạng Văn bản để giao tiếp với người bình thường

Khi camera nhận diện được các ký hiệu tay thì đồng thời trong khung "Message" (ô màu đỏ được tô đậm) sẽ nhận diện và liên tục xuất hiện các dữ liệu dạng Văn bản khi đã nhận diện được.

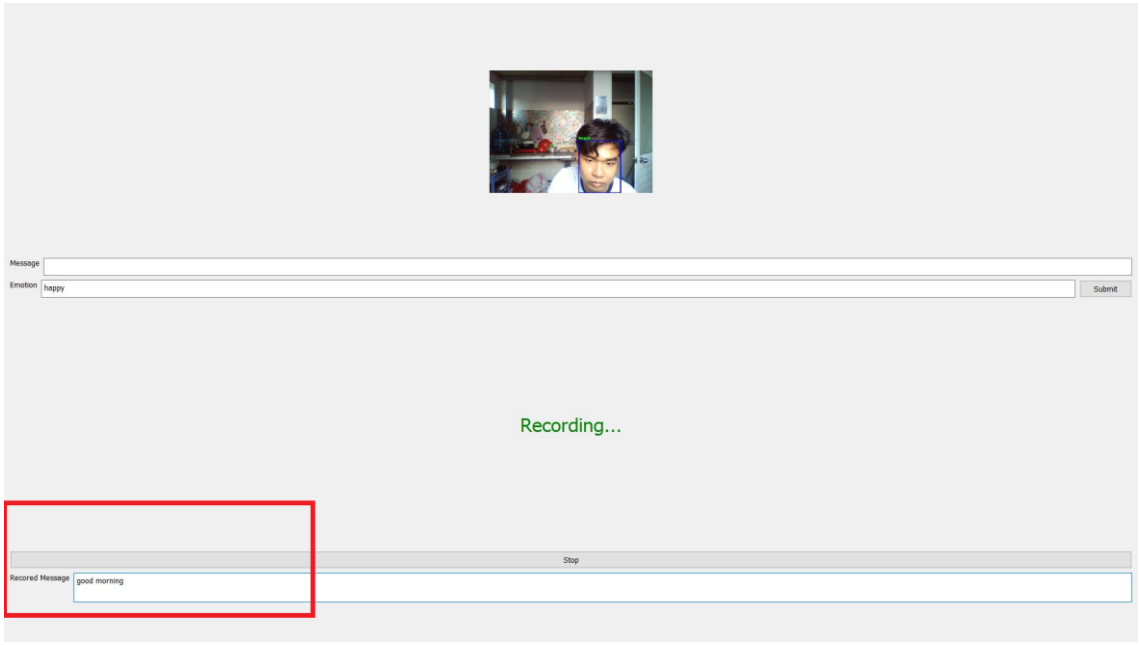
Giao tiếp giữa người bình thường với người khiếm thính- khiếm thị (Hình 5.4, 5.5, 5.6)



Hình 5. 4: Giao diện giao tiếp khi bắt đầu sử dụng Record

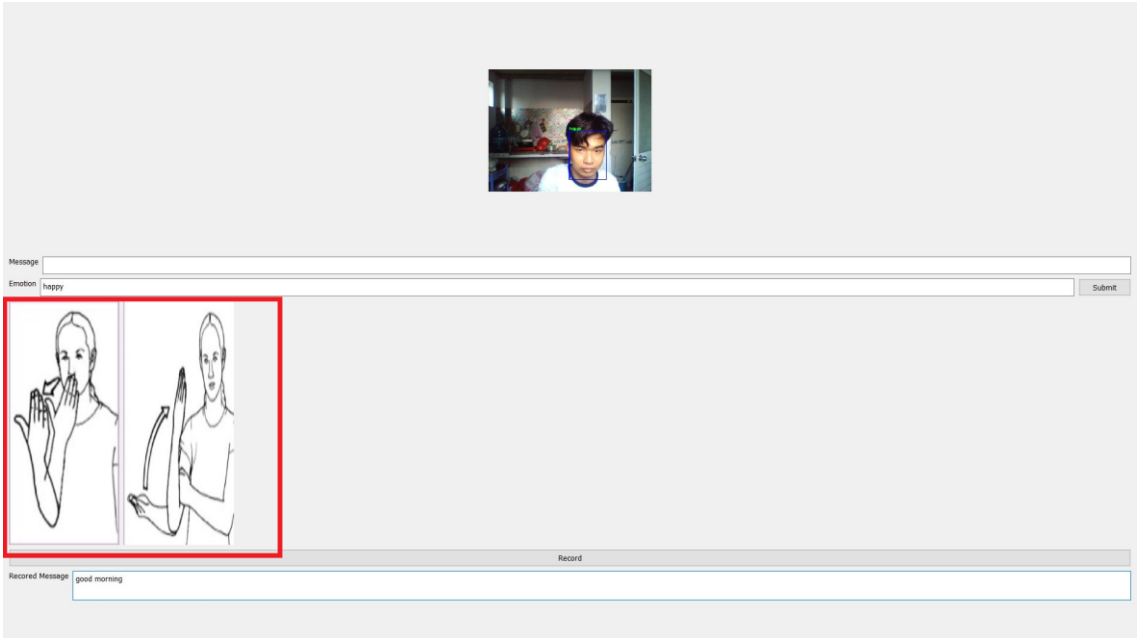
Sau khi nhấn nút "Record" trên giao diện, Giao diện sẽ chuyển sang chế độ Ghi âm lại tiếng nói để nhận diện tiếng nói sau đó chuyển sang dạng Văn bản. Màn hình sẽ xuất hiện chữ "Recording" màu xanh trên màn hình, để thể hiện đang ghi âm lại giọng nói.

Khi muốn dừng chế độ Record thì bấm nút "Stop" trên giao diện. Khi đó, chúng ta bấm dừng thì quá trình Record của chúng ta kết thúc và chương trình sẽ xuất hiện dữ liệu kiểu Văn bản trong khung "Record Message".



Hình 5. 5 : Giao diện sau khi Record và nhận diện được giọng nói thành dạng văn bản

Đây là hình thể hiện cho quá trình Ghi âm để chuyển thành dạng Văn bản, khi chúng ta nói thì chương trình sẽ xử lí và nhận diện lời nói và chuyển đổi thành dạng Văn bản liên tục. Đưa các dữ liệu nhận diện được đưa vào khung "Record Message" (Khung màu đỏ).



Hình 5. 6: Giao diện chương trình sau khi nhận diện được giọng nói và xuất hình ảnh lên màn hình

Sau khi hoàn thành quá trình Record, chương trình sẽ xuất ra màn hình các khung hình sử dụng Ký hiệu tay trong giao tiếp với người khiếm thính-Khiếm thị trên màn hình. Từ đó người Khiếm tính- Khiếm thị có thể quan sát và giao tiếp lại với người bình thường thông qua các bước trên. Ở khung hình này, quá trình ghi âm lại được dữ liệu là ‘good morning” thì chương trình sẽ xử lí và xuất ra màn hình những khung hình có dữ liệu đúng với các dữ liệu thu thập ở đầu vào dạng Text.

CHƯƠNG VI: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN