

ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC VIỆT NHẬT



BÁO CÁO NGHIÊN CỨU KHOA HỌC

**Đề tài: Giám sát chuyển động sinh học của cá mắt muỗi
và cá Medaka sử dụng thị giác máy tính**

Giảng viên hướng dẫn: Bùi Huy Kiên

Sinh viên thực hiện: Trần Tiến Toàn

Lê Minh Khôi

Chu Tuấn Việt

Bùi Duy Hưng

Hà Nội, 2023

I. TÓM TẮT

Trong bài nghiên cứu này, chúng tôi khảo sát ảnh hưởng của estrogen tổng hợp - một trong những chất gây ô nhiễm làm biến đổi tuyến nội tiết và hành vi của các loài cá nhỏ trong môi trường nước. Chất này thường được phát hiện trong môi trường với hàm lượng nhỏ. Tuy nhiên, các nghiên cứu trước đã cho thấy chúng có thể gây ảnh hưởng tới trao đổi chất và hành vi của nhiều động vật thủy sinh ngay cả ở hàm lượng rất nhỏ, ng/L. Với bộ dữ liệu khoảng 10 phút, chúng tôi sử dụng kỹ thuật thị giác máy tính để theo dõi và phân tích hành vi của cá dựa trên sự di chuyển của chúng trong ba khoang bể có nồng độ estrogen khác nhau. Chúng tôi áp dụng mô hình YOLOv8, một mô hình nhận diện đối tượng hiệu quả và nhanh chóng, để phát hiện và gắn nhãn các cá trong các video quay lại. Mô hình được chọn đạt được mAP@50 là 94,8%, cho thấy rằng mô hình có khả năng dự đoán chính xác vị trí các cá trong các video với một lượng lớn kết quả dương tính thật (true-positive) và ít kết quả bị dự đoán là âm tính thật (false-positive) và âm tính giả (false-negative). Sau đó, chúng tôi sử dụng các chỉ số thống kê để đánh giá sự thay đổi hành vi của cá theo thời gian và theo nồng độ estrogen. Kết quả cho thấy rằng estrogen tổng hợp có ảnh hưởng đáng kể lên hành vi của cá, làm giảm sự hoạt động và tương tác của chúng. Bài nghiên cứu này cung cấp một phương pháp mới và hiệu quả để nghiên cứu ảnh hưởng của các chất gây ô nhiễm lên hệ sinh thái thủy sinh, cũng như một ứng dụng tiềm năng của kỹ thuật thị giác máy tính trong lĩnh vực sinh học.

II. GIỚI THIỆU

Đông Nam Á và Đông Á là vùng đất có nền kinh tế sôi động và đông dân số, chiếm một phần ba của tổng dân số thế giới. Sự tăng trưởng kinh tế nhanh chóng, sự gia tăng đột ngột của dân số và sự phát triển mạnh mẽ của các hoạt động sản xuất, cùng với những thách thức từ biến đổi khí hậu, đã đưa môi trường trở thành một vấn đề quan trọng được các chính phủ quan tâm. Trong khu vực này, môi trường nước và sự tồn tại của các hệ thống sinh vật trong môi trường nước là một trong những chủ đề được thảo luận nhiều nhất. Trong cuộc chiến để bảo vệ và quản lý bền vững các nguồn tài nguyên biển cả và sông ngòi, việc hiểu sâu hơn về cách mà các loài cá phản ứng và thích nghi với biến đổi môi trường trở nên hết sức quan trọng. Một trong những yếu tố môi trường quan trọng mà chúng ta quan tâm là nồng độ estrogen trong nước - một hóa chất gây rối loạn nội tiết thường tồn tại ở nồng độ thấp hoặc ít trong môi trường, và có thể bắt chước chức năng của estrogen trong cơ thể động vật và con người [1] (Léon-Olea và cộng sự, 2014; Petrie và cộng sự, 2015). Các báo cáo về tác động của hóa chất gây rối loạn nội tiết đối với sức khỏe của con người và động vật đã xuất hiện khá nhiều trong những năm gần đây. Nhưng số lượng các án phẩm liên quan theo thống kê của PubMed cho từ khóa “estrogen trong

môi trường nước” không nhiều [2]. Phương pháp thường được sử dụng phổ biến trong việc xác định ảnh hưởng của các chất gây ô nhiễm có hoạt tính estrogen như estrogen tự nhiên và tổng hợp trong môi trường đó là sử dụng động vật nhỏ và cá nhỏ để nghiên cứu sự biến đổi về hành vi xã hội, sinh sản và chuyên hoá của chúng khi tiếp xúc với các chất ô nhiễm. Điều này đã được thực hiện bởi nhiều nghiên cứu trước đó.

Dường như, sự biến đổi trong hành vi của các động vật nhỏ, ví dụ như cá nhỏ và động vật gặm nhấm, đóng vai trò quan trọng trong việc đánh giá tác động của sự thay đổi trong môi trường, đặc biệt là tác động từ các chất gây ô nhiễm, đặc biệt là những chất gây biến đổi tuyển nội tiết và điều kiện vật lý của môi trường như nhiệt độ và độ ẩm. Tuy nhiên, điều đáng lưu ý là nhiều nghiên cứu trước đây chủ yếu dựa vào việc quan sát bằng mắt thường của các nghiên cứu viên và kỹ thuật viên trong môi trường phòng thí nghiệm. Công việc này thường mất thời gian và có tính nhảm chán. Hơn nữa, quá trình quan sát như vậy thường tạo ra sự không chắc chắn và không đồng nhất trong kết quả giữa các quan sát của các nghiên cứu viên và kỹ thuật viên khác nhau.

Trong dự án nghiên cứu này, bên cạnh việc quan sát sự biến đổi trong hành vi của cá nhỏ do tác động của chất gây ô nhiễm và sự thay đổi trong môi trường, chúng tôi đề xuất và phát triển một công cụ dựa trên thị giác máy tính để tự động quan sát và trích xuất thông tin liên quan từ các video và hình ảnh được ghi lại thông qua camera về sự thay đổi trong hành vi của cá trong môi trường. Công cụ này sẽ cung cấp cho người quan sát các thông tin quan trọng về hành vi của cá như mô hình di chuyển, sự tập hợp, tốc độ trung bình và vị trí của cá trong các khoảng thời gian khác nhau. Các thông tin thu thập sẽ đảm bảo tính nhất quán, khách quan và giảm thiểu sai lệch giữa các lần quan sát. Công cụ cũng cho phép tự động hóa việc quan sát, giúp thu thập dữ liệu một cách nhanh chóng. Dựa trên thông tin được trích xuất, chúng ta có thể thực hiện các phân tích thống kê để đánh giá những thay đổi trong hành vi của cá trong môi trường nước.

Ngoài ra, dự án nghiên cứu này dự kiến sẽ góp phần quan trọng vào việc thúc đẩy phát triển bền vững ở trong nước và trong khu vực, không chỉ thông qua việc công bố các kết quả nghiên cứu khoa học và kỹ thuật, mà còn qua việc cung cấp các giải pháp hiệu quả giúp các cơ quan quản lý môi trường trong nước và các quốc gia khác trong khu vực đánh giá tác động của các chất gây ô nhiễm trong môi trường nước và thiết lập các quy định liên quan đến sản xuất và sử dụng các hóa chất có khả năng gây ô nhiễm đó.

Trong dự án này chúng tôi lựa chọn cá bát muỗi (Tên khoa học: *Gambusia affinis*) và cá Medaka (Tên khoa học: *Oryzias latipes*) để tiến hành thí nghiệm. Cá hai loại cá này đều

có đặc điểm sinh học độc đáo. Cá băt muỗi (Gambusia) được biết đến với khả năng ưa nước ngọt và khả năng sinh sản nhanh chóng. Chúng thường được sử dụng trong nghiên cứu khoa học và quản lý môi trường để nghiên cứu sự tương tác của các loài cá với các yếu tố môi trường. Cá băt muỗi cũng có khả năng thích nghi với nhiều loại môi trường nước, từ ao nuôi cá đến các dòng sông và hồ nước lớn, làm cho chúng trở thành đối tượng nghiên cứu hữu ích cho việc khảo sát và hiểu biết về hệ sinh thái nước ngọt [3].

Cá medaka (*Oryzias latipes*) là một loài cá nước ngọt nhỏ có nhiều đặc điểm hữu ích trong nghiên cứu vì chu kỳ sinh sản ngắn. Cá medaka là một mô hình sinh học quan trọng, được sử dụng rộng rãi trong nghiên cứu gen và đã trở thành một hệ thống mẫu phổ biến trong nghiên cứu về sinh học phát triển, di truyền học, và đáng chú ý là nghiên cứu về tác động của các chất hóa học, bao gồm cả estrogen, lên sự phát triển và hành vi của cá. Sự linh hoạt và tính ứng dụng của cá Medaka đã giúp nó trở thành một công cụ quan trọng trong việc đánh giá tác động của các chất gây ô nhiễm môi trường lên sức kháng và sức kháng của cá, đặc biệt là trong việc nghiên cứu về tác động của estrogen trong nước.[4] Ngoài ra, cả hai loại cá này đều có những đặc điểm vật lý rõ ràng, giúp cho việc nhận diện bằng thị giác máy tính trở nên dễ dàng hơn.

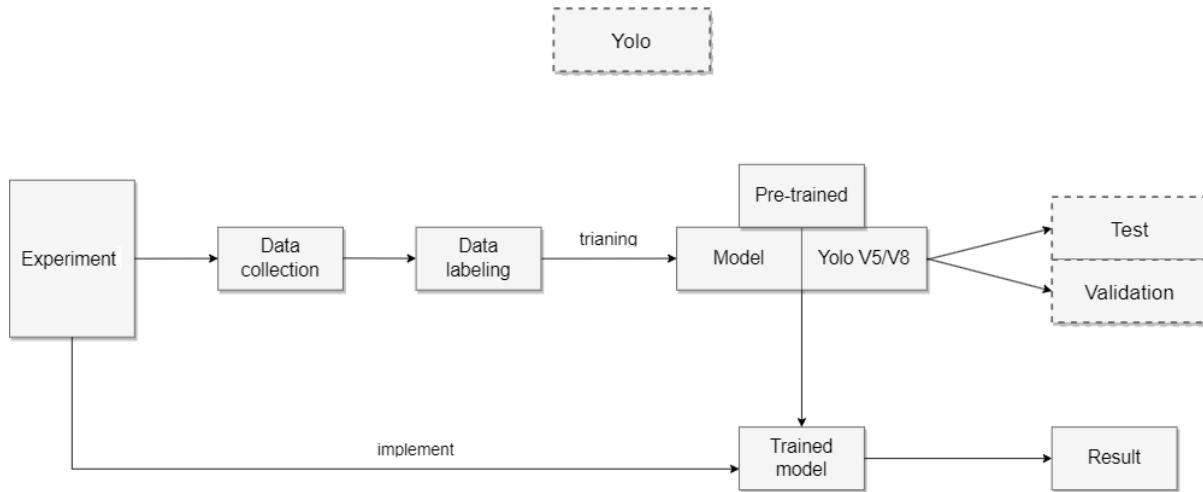
Trí tuệ nhân tạo đã và đang thay đổi, cách mạng hóa nhiều lĩnh vực khác nhau kể từ khi cách mạng Công nghiệp lần thứ tư ra đời [5]. Thị giác máy tính (CV) là một trong những lĩnh vực của trí tuệ nhân tạo được cách mạng hóa nhờ vào các kỹ thuật, phương pháp học sâu (Deep learning) giúp tạo ra được các mô hình thị giác máy tính hiệu quả và trở thành công nghệ quan trọng với mục tiêu đào tạo máy tính có thể thu nhận, nhận dạng, xử lý và phân tích các hình ảnh kỹ thuật số. [6] Với sự phát triển mạnh mẽ của việc xử lý dữ liệu hình ảnh, nó đã trở nên phổ biến trong nhiều lĩnh vực, bao gồm cả trong lĩnh vực thủy sản. Một trong số đó có thể kể tới việc sử dụng hệ thống thị giác máy tính và cảm biến quang học nhằm giám sát chất lượng tại chỗ và sau thu hoạch trong sản xuất thủy sản. Trong các nghiên cứu gần đây, kỹ thuật học sâu (Deep learning) còn được ứng dụng để xác định vị trí, theo dõi các hành vi bất thường của cá và từ đó có thể xác định và phát hiện các vấn đề như sức khỏe của cá, kiểm tra điều kiện môi trường sống, dự đoán và phòng ngừa khi có sự biến đổi trong môi trường sống. Với tính năng học mạnh mẽ của học sâu (Deep learning), thuật toán DL có một số lợi thế hơn so với các phương pháp học máy (Machine learning) thông thường. Hai loại kiến trúc mạng trong Deep learning Mạng thần kinh tích chập (CNN) và Mạng thần kinh hồi quy (RNN) có hiệu suất cao được sử dụng phổ biến rộng rãi. Các mô hình phát hiện đối tượng (object detection) dựa trên kiến trúc mạng CNN được chia thành hai loại: cách tiếp cận hai giai đoạn (2 stages) (e.g, Faster R-CNN) và cách tiếp cận một giai đoạn (single-stage) (v.d, YOLO). Cách tiếp cận một giai đoạn chỉ cần đi qua mạng lưới một lần để dự đoán các bounding boxes vì

vậy tốc độ xử lý thời gian thực nhanh hơn tuy nhiên về độ chính xác thấp hơn cách tiếp cận hai giai đoạn, do đó cách tiếp cận 1 giai đoạn phù hợp với ứng dụng, nghiên cứu đòi hỏi tốc độ xử lý thời gian thực và có thể chấp nhận độ chính xác ở mức trung bình.

Nhiều phương pháp khác nhau đã được thực hiện để phát hiện cá nhằm theo dõi hoạt động và hành vi của cá. Thế nhưng hầu hết các phương pháp được nghiên cứu với tốc độ khung hình thấp không phù hợp với nghiên cứu hành vi của cá bắt muỗi và cá medaka trong thời gian thực. Nghiên cứu này sẽ sử dụng YOLO để kiểm soát sự thay đổi của nồng độ estrogen với hành vi của cá. You Only Look Once (YOLO) là một thuật toán lan truyền và được sử dụng rộng rãi [7]. YOLO nổi tiếng với đặc tính phát hiện đối tượng của nó. Vào năm 2015, Redmond và cộng sự. đã giới thiệu phiên bản YOLO đầu tiên . Trong nhiều năm, các học giả đã xuất bản một số phiên bản tiếp theo của YOLO được mô tả là YOLOv2, YOLOv3, YOLOv4 và YOLOv5.Nghiên cứu này bài viết chỉ tập trung vào năm phiên bản YOLO V5 và YOLLO V8.Cốt lõi của thuật toán phát hiện mục tiêu YOLO nằm ở kích thước nhỏ và tốc độ tính toán nhanh của mô hình. Các cấu trúc của YOLO rất đơn giản. Nó có thể xuất trực tiếp vị trí và danh mục của hộp giới hạn thông qua mạng lưới thần kinh. Tốc độ của YOLO nhanh vì YOLO chỉ cần đưa hình ảnh vào mạng là nhận được kết quả phát hiện cuối cùng, do đó YOLO cũng có thể nhận ra việc phát hiện thời gian của video.Điều quan trọng là, các mô hình YOLO còn nhanh hơn trong quá trình đào tạo và có khả năng đạt được độ chính xác cao với kích thước mô hình nhỏ hơn. Chúng có thể được đào tạo (train) trên GPU đơn, điều này làm cho chúng trở nên dễ tiếp cận hơn với chúng tôi.

III. PHƯƠNG PHÁP

Sơ đồ quy trình



Hình 1. Sơ đồ quy trình thí nghiệm dự án giám sát cá sử dụng thị giác máy tính.

1. Chuẩn bị dữ liệu

1.1. Bố trí thí nghiệm

Chúng tôi tiến hành thử nghiệm trên 2 mô hình khác nhau. Trong các bức ảnh được trích xuất ra từ các video của hai mô hình có chứa một bể nước bao gồm 3 khoang bể gồm các loại cá khác nhau được ngăn cách nhau và có nồng độ estrogen thay đổi theo ngày.

Ở mô hình thứ nhất và mô hình thứ hai chúng tôi tiến hành 7 thí nghiệm với mỗi mô hình:

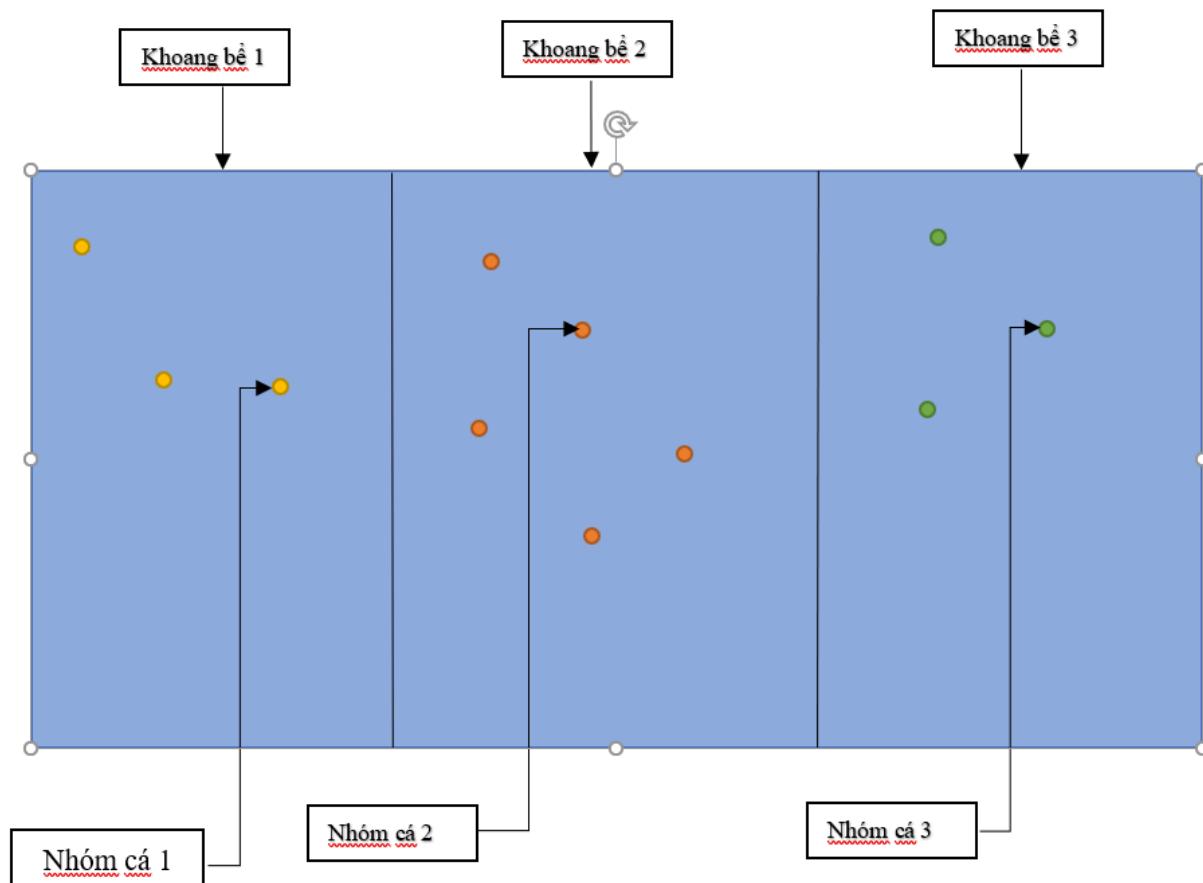
- Với mô hình 1, tương ứng với ba khoang bể từ trái qua phải lần lượt là cá Medaka cái (Female rice fish), Cá bắt muỗi đực (Male mosquito fish) và cá Medaka đực (Male rice fish).

- Với mô hình 2, ba khoang bể từ trái qua phải tương ứng với cá Medaka đực (Male rice fish), Cá bắt muỗi cái (Female mosquito fish) và cá Medaka cái (Female rice fish).

Với cả hai mô hình trên, chúng tôi thay đổi nồng độ estrogen ở khoang bể thứ 2 để theo dõi sự thay đổi hành vi của chúng qua từng ngày, từng nồng độ khác nhau:

- 0ng estrogen : ngày 0;
- 5ng estrogen: ngày 2, ngày 4, ngày 7, revert.
- 25ng estrogen: ngày 2, ngày 4.

Hình ảnh mô tả bể cá gồm cá Medaka cái, cá Mosquito đực, cá Medaka đực.

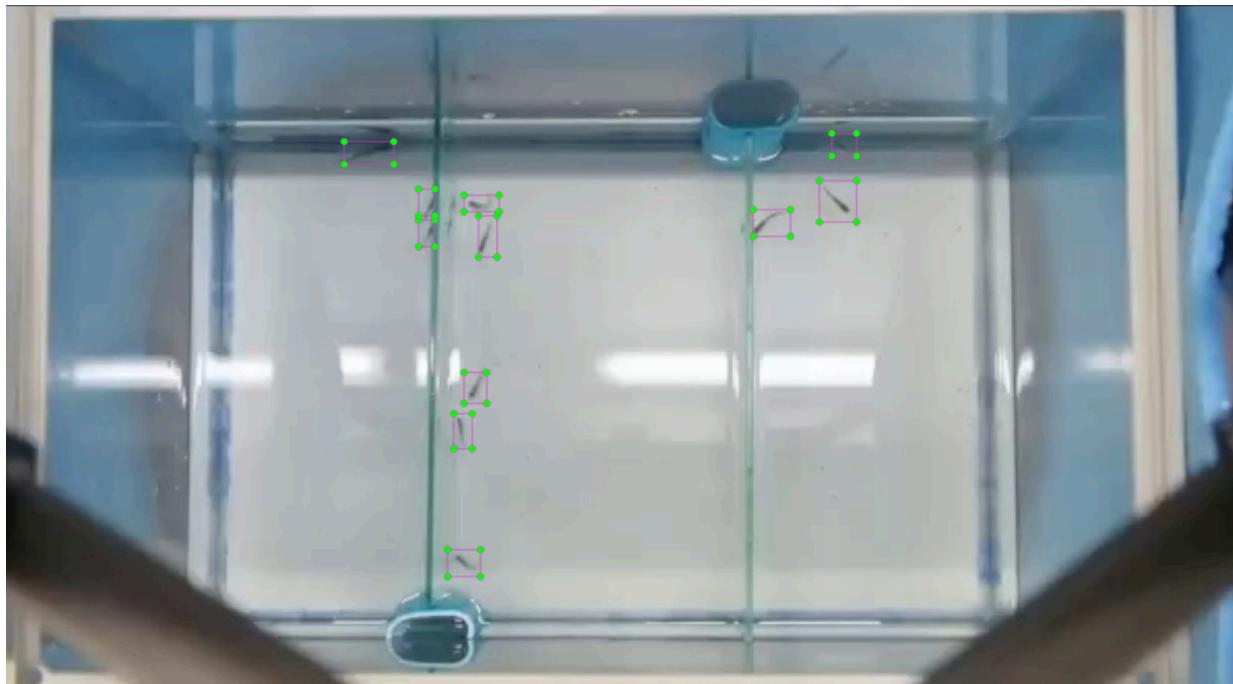


Hình 2: Ảnh bể cá

1.2 Thu thập và gán nhãn dữ liệu

Bộ dữ liệu được thu thập dưới dạng video 10 phút quay lại quá trình di chuyển của cá, và khoảng 600 khung ảnh được trích xuất từ các video được quay lại với camera 60fps. Độ phân giải của camera là 1080p.

Sau khi chuẩn bị dữ liệu, chúng tôi sẽ thực hiện công việc chia nhỏ video thành các frame ảnh sau đó gán nhãn cho dữ liệu bằng công cụ labelImg



Hình ảnh 3: Ảnh gán nhãn cá

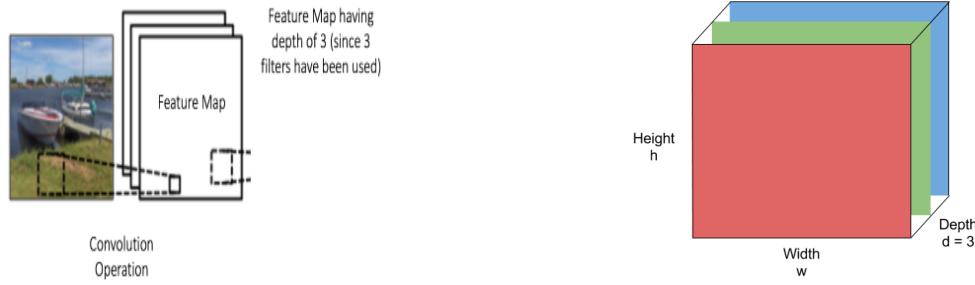
2. Mô hình

2.1. Kiến trúc mô hình

2.1.1. Mạng Nơ-ron Tích Chập (Convolutional Neural Networks - CNNs):

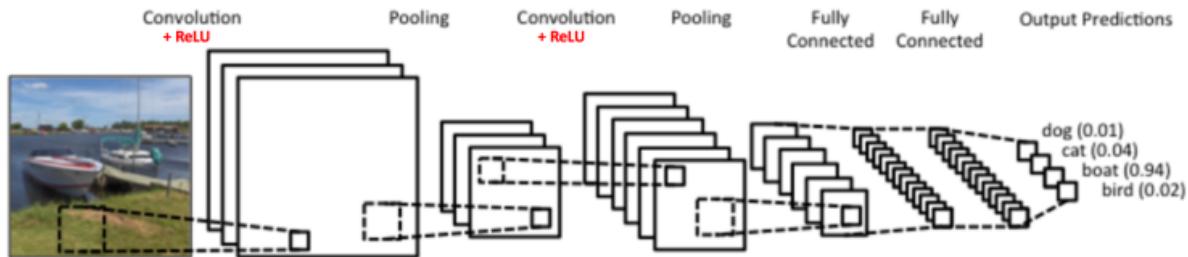
Mạng nơ-ron tích chập (CNN) là một trong những mô hình học sâu phổ biến và hiệu quả nhất dành cho xử lý ảnh và video. Đặc trưng của CNN là việc sử dụng các lớp tích chập, giúp trích xuất thông tin cục bộ từ dữ liệu đầu vào thông qua việc sử dụng các bộ lọc.

CNN sử dụng hình ảnh đầu vào dưới dạng một mảng pixel và phân loại nó thành các hạng mục nhất định (ví dụ: Chó, Mèo, Hổ,...). Độ phân giải của hình ảnh quyết định kích thước của mảng pixel, thường được biểu diễn dưới dạng $H \times W \times D$ (H : Chiều cao, W : Chiều rộng, D : Độ dày). Ví dụ: Một hình ảnh có độ phân giải 6×6 và là hình ảnh màu RGB sẽ được biểu diễn dưới dạng một mảng ma trận $6 \times 6 \times 3$ (trong đó, 3 là giá trị RGB).



Hình ảnh 4: Hình ảnh là mảng ma trận

Mỗi hình ảnh đầu vào sẽ chuyển nó qua 1 loạt các lớp tích chập với các bộ lọc (Kernels), tổng hợp lại các lớp được kết nối đầy đủ (Full Connected) và áp dụng hàm Softmax để phân loại đối tượng có giá trị xác suất giữa 0 và 1. [8]



Hình ảnh 5: Một convNet đơn giản

Các thành phần chính của CNN:

1. Lớp Tích Chập (Convolutional Layer): Dùng để trích xuất đặc trưng từ ảnh thông qua bộ lọc.

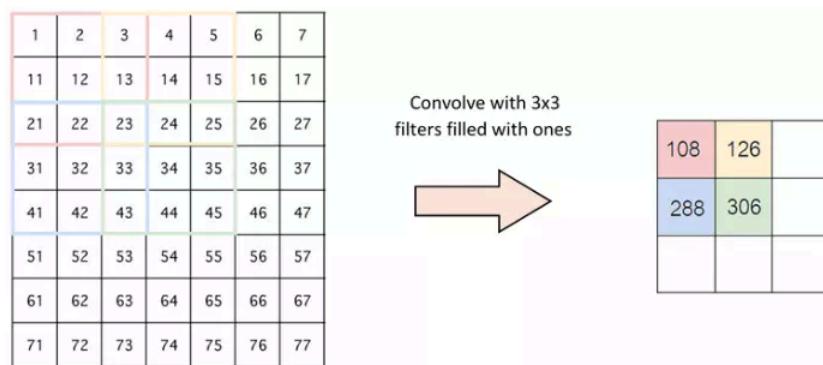
Lớp này thực hiện tích số chấm giữa hai ma trận, trong đó một ma trận là tập hợp các tham số có thể học được hay còn gọi là hạt nhân và ma trận còn lại là phần bị hạn chế của trường tiếp nhận. Hạt nhân nhỏ hơn về mặt không gian so với hình ảnh nhưng có chiều sâu hơn. Điều này có nghĩa là, nếu hình ảnh bao gồm ba kênh (RGB), chiều cao và chiều rộng của hạt nhân sẽ nhỏ về mặt không gian, nhưng độ sâu mở rộng đến cả ba kênh[22].

Một lớp tích chập bao gồm n_f bộ lọc có cùng kích thước và độ sâu. Đối với mỗi bộ lọc, chúng tôi kết hợp nó với âm lượng đầu vào để thu được n_f đầu ra. Tiếp theo, các đầu ra

được chuyển tới một số hàm kích hoạt, chẳng hạn như ReLU. Cuối cùng, nó được xếp chồng lên nhau thành một khối $(h - f_h + 1) \times (w - f_w + 1) \times n_f$.

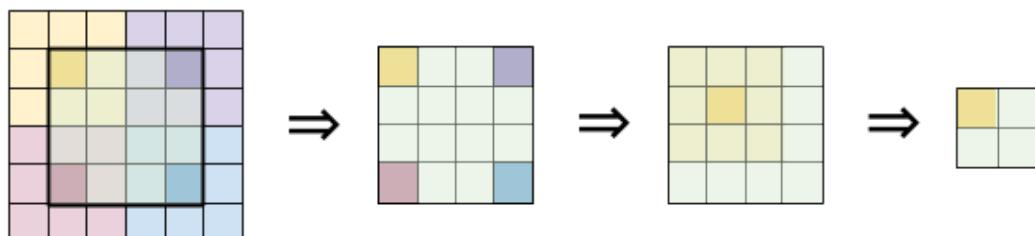
2. Bước nhảy (Stride) :

Stride là số pixel thay đổi trên ma trận đầu vào. Khi stride là 1 thì ta di chuyển các kernel 1 pixel. Khi stride là 2 thì ta di chuyển các kernel đi 2 pixel và tiếp tục như vậy. Hình dưới là lớp tích chập hoạt động với stride là 2.



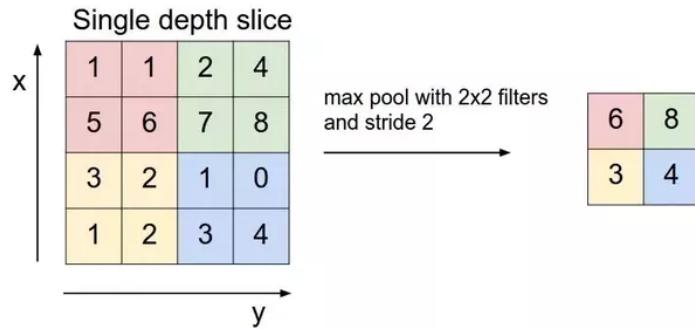
Hình ảnh 6: lớp tích chập hoạt động với stride là 2

3. Lớp Đệm (Padding): Thêm một lượng nhất định giá trị (thường là 0) vào viền của ảnh. Giúp giữ nguyên kích thước không gian của ảnh sau mỗi lớp tích chập.



Hình ảnh 7: lớp đệm

4. *Lớp Gộp (Pooling Layer)*: Lớp pooling sẽ giảm bớt số lượng tham số khi hình ảnh quá lớn. Không gian pooling còn được gọi là lấy mẫu con hoặc lấy mẫu xuống làm giảm kích thước của mỗi map nhưng vẫn giữ lại thông tin quan trọng



Hình ảnh 8: Lớp gộp

Nếu chúng ta có bản đồ kích hoạt có kích thước $W \times W \times D$, hạt nhân tổng hợp có kích thước không gian F và bước tiến S , thì kích thước của âm lượng đầu ra có thể được xác định theo công thức sau

$$W_{out} = \frac{W - F}{S} + 1$$

5. *Lớp Kết Nối Đầy Đủ (Fully Connected Layer)*: Sau một số lớp tích chập và gộp, chúng tôi đã làm phẳng dữ liệu của mình thành một vectơ và đưa dữ liệu đó vào một lớp được kết nối đầy đủ giống như trong mạng thần kinh thông thường.. Dùng để phân loại hoặc hồi quy dựa trên các đặc trưng đã được trích xuất. [9]

2.1.2. Phát hiện đối tượng (Object Detection):

Object Detection đề cập đến khả năng của hệ thống máy tính và phần mềm để định vị các đối tượng trong một hình ảnh và xác định từng đối tượng. Object Detection đã được sử dụng rộng rãi để phát hiện khuôn mặt, phát hiện xe, đếm số người đi bộ, hệ thống bảo mật và xe không người lái. Trong bài nghiên cứu này chúng tôi sử dụng để nhận diện hành vi của các loại cá.

Object Detection không chỉ nhận biết sự tồn tại của một đối tượng trong ảnh mà còn xác định vị trí của chúng thông qua việc vẽ một hộp bao quanh (bounding box).

Phần tử chính trong object detection:

- Bounding Box: Đây là hình chữ nhật được vẽ quanh một đối tượng, thường được biểu diễn bằng tọa độ góc trên bên trái và tọa độ góc dưới bên phải.



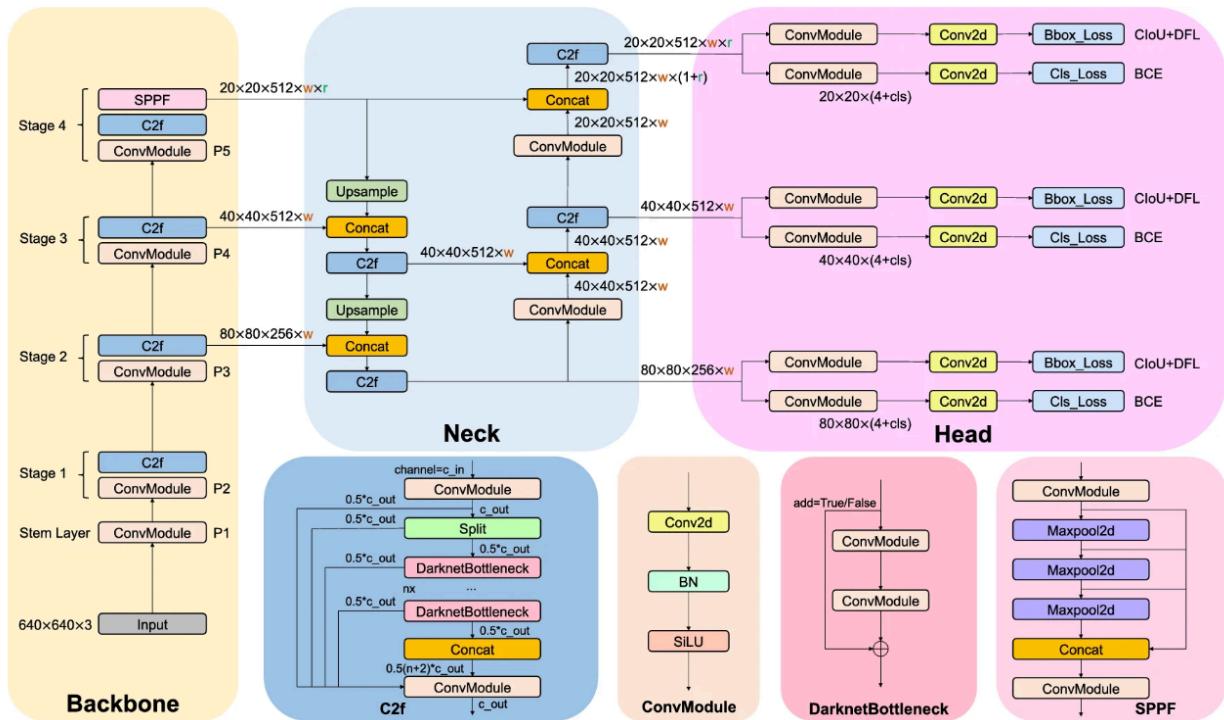
Hình ảnh 9: Phát hiện cá bằng mô hình YOLOv8

- Class Label: Nhãn cho biết loại đối tượng nào đang được nhận dạng, ở đây tài này, chúng tôi sử dụng nhãn “fish”
- Confidence Score: Điểm số cho biết mức độ tin cậy của mô hình trong việc dự đoán một đối tượng và vị trí của nó.

2.1.3. Kiến trúc YOLOv8

YOLOv8 chưa công bố bài báo nào về kiến trúc mạng, vì vậy chúng ta thiếu thông tin về phương pháp nghiên cứu trực tiếp và nghiên cứu cắt bỏ được thực hiện trong quá trình tạo ra nó. Nhưng qua sự phân tích tỉ mỉ file YOLOv8-p2.yaml ,repo github của RangeKing

YOLOv8 có cấu trúc chính tương tự với YOLOv5 bao gồm: mạng thần kinh, cổ, đầu dò, hàm loss. Ngoài ra YOLOv8 vẫn có một số nâng cấp về cấu trúc:



Hình ảnh 10: YOLOV8 Architecture từ Fracture detection in pediatric wrist trauma X-ray images using YOLOv8 algorithm

Kiến trúc của YOLOv8 bao gồm ba phần chính:

Backbone: Đây là phần chính của mạng. Đối với YOLOv8, đường trực được thiết kế bằng cấu trúc CSP-Darknet53 mới, một bản sửa đổi của kiến trúc Darknet được sử dụng trong các phiên bản trước. Kiến trúc CSP giúp cải thiện khả năng học của CNN và giảm chi phí tính toán của mô hình. YOLOv8 giới thiệu module C2f bao gồm 2 ConvModule và n lớp DarknetBottleNeck được kết nối với nhau thông qua Split và Concat. Khối DarknetBottleNeck được sử dụng để giảm số lượng kênh bằng cách sử dụng các bộ lọc kích thước nhỏ (1x1), tiếp theo sử dụng bộ lọc có kích thước lớn hơn để trích xuất thông tin không gian và cuối cùng nếu Bottleneck có kết nối Shortcut, đầu ra của lớp tích chập cuối cùng sẽ được cộng với đầu vào ban đầu của khối hoặc qua một lớp tích chập 1x1 để điều chỉnh kích thước phù hợp trước khi cộng. Bottleneck giúp giảm số lượng phép tính mà vẫn giữ được khả năng trích xuất đặc trưng. Lớp ConvModule bao gồm 3 lớp kết nối với nhau: Convolution2d-BatchNorm2d-SiLU. Khối cuối cùng là khối SPPF được sử dụng để giúp mô hình có khả năng xử lý các đầu vào có kích thước bất kỳ mà không làm mất thông tin đặc trưng. [10][11]

Cổ (Neck): Các mạng sâu hơn thường thu thập được nhiều thông tin đặc trưng hơn, dẫn đến kết quả dự đoán dày đặc tốt hơn. Tuy nhiên, các mạng quá sâu sẽ làm giảm thông tin vị trí của đối tượng, và quá nhiều phép toán tích chập sẽ dẫn đến mất thông tin đối với các đối tượng nhỏ. Do đó, việc sử dụng kiến trúc Mạng Hình Tháp Đặc Trưng (FPN) và Mạng Tổng Hợp Đường Dẫn (PAN) để kết hợp thông tin đặc trưng từ các tầng có độ phân giải khác nhau. Các lớp trên thu thập được nhiều thông tin hơn do có thêm các lớp mạng, trong khi các lớp dưới bảo toàn thông tin vị trí do có ít lớp tích chập hơn. Lấy cảm hứng từ YOLOv5, FPN thực hiện việc lấy mẫu lên từ trên xuống dưới để tăng lượng thông tin đặc trưng trong bản đồ đặc trưng dưới cùng; và PAN thực hiện lấy mẫu xuống từ dưới lên trên để thu thập thông tin bản đồ đặc trưng trên cùng. Hai đầu ra đặc trưng này được kết hợp lại để đảm bảo dự đoán chính xác cho các hình ảnh có kích thước khác nhau. FP-PAN (Hình Tháp Đặc Trưng - Tổng Hợp Đường Dẫn) trong mô hình được áp dụng và loại bỏ các phép toán tích chập trong quá trình lấy mẫu lên để giảm chi phí tính toán [12]. Cấu trúc này rất phù hợp với các đặc điểm nổi bật trên bộ ba tỷ lệ: nhỏ (20×20), trung bình (40×40) và mở rộng (80×80), đáp ứng nhu cầu chi tiết về đặc trưng trên các kích thước đối tượng khác nhau [13].

Head: Thành phần này tạo ra dự đoán dựa trên các đặc trưng được trích xuất từ 2 thành phần trước đó là Backbone và Neck. Khác với các phiên bản trước, mô hình YOLOv8 là mô hình Anchor-Free sẽ xác định tâm của đối tượng và ước tính khoảng cách giữa tâm và hộp giới hạn và dự đoán đối tượng. [14][15]

YOLOv8 là mô hình anchor-free . Điều này có nghĩa là nó dự đoán trực tiếp tâm của đối tượng thay vì phần bù từ anchor box .Tính năng phát hiện anchor-free giúp giảm số lượng dự đoán hộp, giúp tăng tốc độ Non-Maximum Suppression (NMS), một bước xử lý hậu kỳ phức tạp giúp sàng lọc các phát hiện đối tượng.

Trong quá trình training YOLOv8 thực hiện nhiều cải tiến cho hình ảnh training. Một sự gia tăng như vậy là tăng cường dữ liệu khám. Tăng cường dữ liệu khám là một kỹ thuật tăng cường đơn giản trong đó bốn hình ảnh khác nhau được ghép lại với nhau và đưa vào mô hình làm đầu vào. Điều này làm cho mô hình tìm hiểu các đối tượng thực tế từ các vị trí khác nhau và ở trạng thái che khuất một phần.

Tuy nhiên, việc tăng cường này được chứng minh bằng thực nghiệm là làm giảm hiệu suất nếu được thực hiện trong toàn bộ quá trình luyện tập. Sẽ rất có lợi nếu tắt nó trong mười giai đoạn training cuối. Loại thay đổi này là điển hình cho sự chú ý cẩn thận của mô hình YOLO đã được đưa ra trong thời gian làm thêm trong báo cáo YOLOv5 và trong nghiên cứu YOLOv8. [16]

2.2. Đánh giá

2.2.1. Ma trận nhầm lẫn (Confusion matrix)

Confusion matrix là một ma trận lấy các giá trị thực tế sau đó ánh xạ các giá trị này tới các giá trị dự đoán đánh giá kết quả của những bài toán phân loại với việc xem xét cả những chỉ số về độ chính xác và độ bao quát của các dự đoán cho từng lớp. Confusion matrix rất hữu ích trong các bài toán phân loại

- Một confusion matrix gồm 4 chỉ số đối với mỗi lớp phân loại:
 1. True Positive (TP): Khi dự đoán là dương tính và thực sự đúng
 2. True Negative (TN): Khi dự đoán là âm tính và thực sự đúng
 3. False Positive (FP): Khi dự đoán là dương tính và thực sự sai
 4. False Negative (FN): Khi dự đoán là âm tính và thực sự sai

[18]

2.2.2. Độ chính xác (Accuracy)

Độ chính xác (Accuracy) là một chỉ số đo lường tần số mà một mô hình học máy dự đoán đúng kết quả. Bạn có thể tính độ chính xác bằng cách chia số lượng dự đoán đúng cho tổng số dự đoán.

Cụ thể, độ chính xác (accuracy) đo lường tỷ lệ của các dự đoán đúng (true predictions) mà mô hình đã thực hiện so với tổng số dự đoán. Nó giúp bạn đánh giá mức độ chính xác của mô hình trong việc dự đoán kết quả.

Công thức tính:

$$\text{Accuracy} = \frac{\text{Số lượng dự đoán đúng}}{\text{Tổng số dự đoán}} = \frac{TP + TN}{TP + FP + TN + FN}$$

[17]

2.2.3. Precision

Precision là một chỉ số đo lường tần số mà một mô hình học máy dự đoán đúng lớp tích cực. Precision đo lường khả năng của mô hình trong việc đưa ra các dự đoán tích cực (positive predictions) và bao gồm tỷ lệ các dự đoán đúng là tích cực so với tổng số dự đoán tích cực mà mô hình thực hiện. Nó giúp đánh giá mức độ chính xác của mô hình trong việc đưa ra các dự đoán tích cực.

Hay nói cách khác, Precision trả lời cho câu hỏi: Trong tất cả các dự đoán tích cực, bao nhiêu trong số đó là thực sự đúng?

Công thức tính:

$$\text{Precision} = \frac{TP}{TP + FP}$$

[19]

2.2.4. Recall

Recall đo lường tỷ lệ của các trường hợp tích cực thực sự (true positives) mà mô hình đã xác định đúng so với tổng số trường hợp tích cực thực sự có trong dữ liệu. Nó giúp bạn đánh giá xem mô hình có xuất hiện bất kỳ trường hợp tích cực nào mà nó đã bỏ sót không

Hay nói cách khác, Recall trả lời cho câu hỏi: trong tất cả các trường hợp tích cực có trong dữ liệu, bao nhiêu trong số đó đã được mô hình phát hiện?

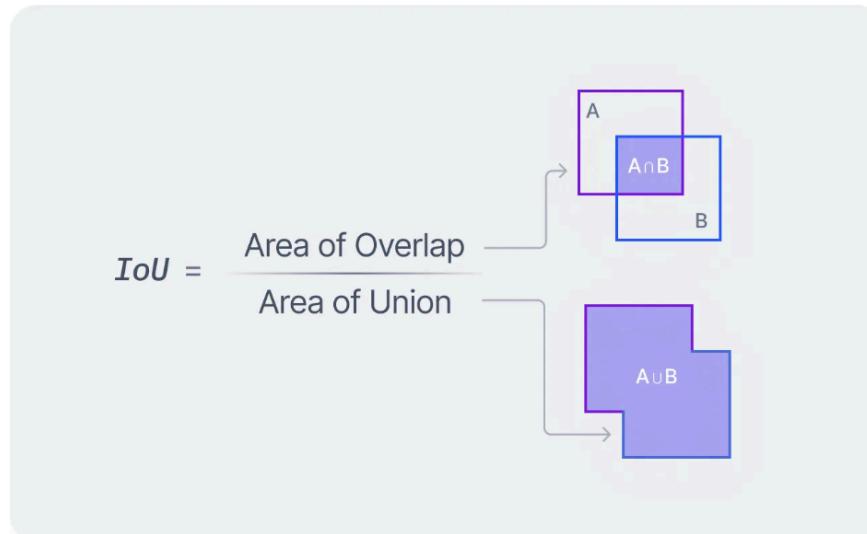
Công thức tính:

$$\text{Recall} = \frac{TP}{TP + FN}$$

[14]

2.2.5. IoU (Intersection over union)

- IoU là một chỉ số được sử dụng để đo độ chính xác trong việc định vị và tính toán sai lệch trong các mô hình phát hiện đối tượng.
- IoU (Intersection over Union) là giá trị đo lường sự chồng chéo giữa 2 ranh giới (boundary). Chúng ta sử dụng nó để đo lường mức độ mà ranh giới dự đoán của chúng ta trùng với ranh giới thực tế (ranh giới đối tượng thật). Trong một số bộ dữ liệu, chúng ta định trước một ngưỡng IoU để phân loại liệu dự đoán có phải là true positive (dự đoán đúng) hay false positive (dự đoán sai) hay không.
- Ngưỡng của IoU nằm trong ranh giới từ 0 đến 1. Giá trị ngưỡng của IoU cao nếu mức độ chồng chéo cao giữa các hộp sự thật và hộp dự đoán. Ngược lại nếu mức độ chồng chéo thấp thì giá trị IoU thấp



Hình ảnh 11: Ảnh công thức IoU [21]

hoặc có thể được tính bằng công thức dựa vào các chỉ số Confusion matrix:

$$IoU = \frac{TP}{TP + FN + FP}$$

- Ngưỡng IoU không có mức độ cố định mà sẽ phụ thuộc vào yêu cầu cụ thể của từng tác vụ và tập dữ liệu phát hiện đối tượng. Thông thường, ngưỡng IoU 0.5 được sử dụng phổ biến nhất, nghĩa là nếu IoU giữa ranh giới dự đoán và ranh giới thực tế lớn hơn hoặc bằng 0.5, thì dự đoán sẽ được coi là true positive.
- Các chỉ số Confusion matrix trong IoU:

TP: Khi IOU của hộp dự đoán so với hộp thực tế $>$ ngưỡng IoU

FP: Khi IOU của hộp dự đoán so với hộp thực tế $<$ ngưỡng IoU

TN: Thông số này ta có thể hiểu nó như là background.

FN: Bounding box của đối tượng không được phát hiện [21]

2.2.6. AP (Average Precision)

AP (Average Precision) là một thước đo đánh giá hiệu suất trong phát hiện đối tượng và tác vụ phân loại. Từ precision và recall ta có thể vẽ đường cong Precision và recall (PR

curve). Nó chỉ ra sự đánh đổi giữa precision và recall đối với nhiều giá trị confidence score khác nhau. AP chính là phần diện tích phía dưới đường cong PR curve, diện tích này đại diện cho mức độ chính xác trung bình của mô hình trong việc phát hiện các đối tượng tích cực (Positive). Nếu phần diện tích này lớn đồng nghĩa với việc precision và recall sẽ từ đó ta có thể suy ra rằng model có chất lượng tốt. [20]

2.2.7. mAP (mean Average Precision)

mAP được tính toán bằng trung bình cộng của các giá trị AP cho từng lớp hoặc vùng quan tâm riêng.

mAP@IoU có ý nghĩa trong việc đánh giá hiệu suất của mô hình phát hiện đối tượng dưới các ngưỡng IoU khác nhau. Nó giúp đánh giá khả năng của mô hình trong việc tìm ra và phân loại đối tượng dựa trên độ chính xác của phát hiện (IoU) và cách tính toán Precision và Recall ở mỗi ngưỡng IoU. Điều này có thể giúp bạn hiểu rõ hơn về hiệu suất của mô hình trong các tình huống khác nhau, đặc biệt khi bạn quan tâm đến độ chính xác của phát hiện dựa trên IoU.[12]

2.3. Huấn luyện và kiểm thử

Chúng tôi sẽ chia tập dữ liệu gồm ảnh và nhãn thành 2 tệp training và validate với tỷ lệ 7/3.Thuật toán YOLO sau khi được cấu hình, chỉ định số lượng lớp (loài cá) và kiến trúc của mạng lưới thần kinh. Tập dữ liệu có chủ thích sau đó được đưa vào model YOLO để training.Kết quả sau khi được training sẽ được sử dụng để predict các loài cá trong tệp testing và lưu lại kết quả.

2.4. Cài đặt và cấu hình

2.4.1. Cài đặt

Anaconda là một phần mềm phổ biến để quản lý khoa học dữ liệu và machine learning (học máy) trên Python và R. Nó bao gồm các thư viện và công cụ quan trọng để phân tích dữ liệu, xây dựng mô hình machine learning, và triển khai các ứng dụng khoa học dữ liệu.

Chúng ta cài đặt Anaconda Navigator 2.4 , ta cài đặt môi trường làm việc sử dụng Python 3.9.18 , sau đó chạy Jupiter Notebook. Sau khi cài đặt Anaconda Navigator cho máy , ta cài các thư viện cần thiết để nhận dạng cá : Có nhiều thư viện trong Python hỗ trợ cho

việc nhận dạng hình ảnh: OpenCV, TensorFlow, Keras, PyTorch, Numpy, Pandas, Matplotlib, Scikit-learn,..

Thư viện **ultralytics** là một bộ công cụ mạnh mẽ và linh hoạt để huấn luyện và sử dụng mô hình YOLO (You Only Look Once) cho nhiệm vụ phát hiện đối tượng trong hình ảnh và video.

Thư viện ultralytics cung cấp các công cụ và tài nguyên để huấn luyện mô hình YOLO từ đầu hoặc sử dụng các mô hình đã huấn luyện trước để thực hiện phát hiện đối tượng. Ngoài ra, thư viện này cung cấp các hàm và lớp hữu ích để đánh giá, kiểm tra và tối ưu hóa mô hình.

Thư viện **YOLO** (You Only Look Once) là một phương pháp phát hiện đối tượng dựa trên mạng nơ-ron sâu. YOLO được thiết kế để thực hiện dự đoán trực tiếp vị trí và lớp của các đối tượng trong một hình ảnh hoặc video. Phương pháp này nổi tiếng với tốc độ xử lý nhanh và khả năng phát hiện chính xác cao.

2.4.2. Phân bổ dữ liệu trên data.yaml

Đoạn mã cấu hình này xác định cách dữ liệu huấn luyện, kiểm thử và thử nghiệm được tổ chức, cũng như số lượng và tên của lớp mục tiêu cho việc huấn luyện mô hình nhận dạng hình ảnh. Cụ thể, đường dẫn đến thư mục chính của dữ liệu được chỉ định bởi path. Thư mục này chứa hình ảnh huấn luyện và kiểm thử. Mô hình được thiết kế để nhận dạng một loại cá duy nhất, được chỉ định bởi nc: 1 và tên của lớp là ['fish'].

2.4.3. Cấu hình

Cấu hình máy tính với CPU Intel i9 thế hệ thứ 13, GPU NVIDIA RTX A4000 16GB VRam và 32GB RAM là một hệ thống mạnh mẽ phù hợp cho việc nhận dạng hình ảnh cá. CPU Intel core i9 sẽ đảm bảo hiệu suất xử lý cao, giúp xử lý nhanh các thuật toán phức tạp. GPU RTX A4000 16GB là một lựa chọn xuất sắc cho việc huấn luyện và triển khai các mô hình học sâu như YOLO hay CNN, cung cấp khả năng tính toán song song mạnh mẽ và bộ nhớ đủ lớn cho việc xử lý các tập dữ liệu hình ảnh lớn. Cuối cùng, 32GB RAM sẽ giúp xử lý hiệu quả các tác vụ đa nhiệm và giảm thiểu thời gian chờ đợi khi xử lý dữ liệu lớn. Nhìn chung, cấu hình này sẽ cung cấp hiệu suất mạnh mẽ cho việc nhận dạng hình ảnh cá trong các ứng dụng thực tế.

2.5. Phân bố dữ liệu

Chúng tôi chia các dữ liệu cho các lần train lần lượt thành 25, 50, 100, 200, 400, 800 ảnh để đánh giá hiệu suất mô hình xem số lượng ảnh bao nhiêu là phù hợp

IV. THỰC THI

1 Dữ liệu và Tiền xử lý

Nghiên cứu sử dụng bộ dữ liệu bao gồm 14 video dữ liệu thí nghiệm của bể cá trong 2 thí nghiệm. Trước khi đào tạo mô hình, tập dữ liệu đã được tiền xử lý để đảm bảo chất lượng và tính nhất quán. Tập dữ liệu huấn luyện được sử dụng để huấn luyện mô hình, trong khi tập dữ liệu kiểm tra được sử dụng để đánh giá hiệu suất mô hình.

2 Lựa chọn mô hình

Mô hình YOLOv5 và YOLOv8 được lựa chọn để sử dụng cho quy trình huấn luyện.

3 Quy trình huấn luyện

Sử dụng tập dữ liệu huấn luyện để huấn luyện mô hình. Trong quá trình huấn luyện, mô hình sẽ cập nhật các trọng số dựa trên sai số giữa dự đoán của mô hình và nhãn thực tế. Quy trình huấn luyện được thực hiện qua 100 chu kỳ hoàn toàn qua toàn bộ tập dữ liệu huấn luyện. Trong mỗi chu kỳ, toàn bộ tập dữ liệu huấn luyện được sử dụng một lần để cập nhật trọng số của mô hình. Dừng sớm được sử dụng để ngừng đào tạo nếu mất mát xác thực không cải thiện trong mười kỳ liên tiếp.

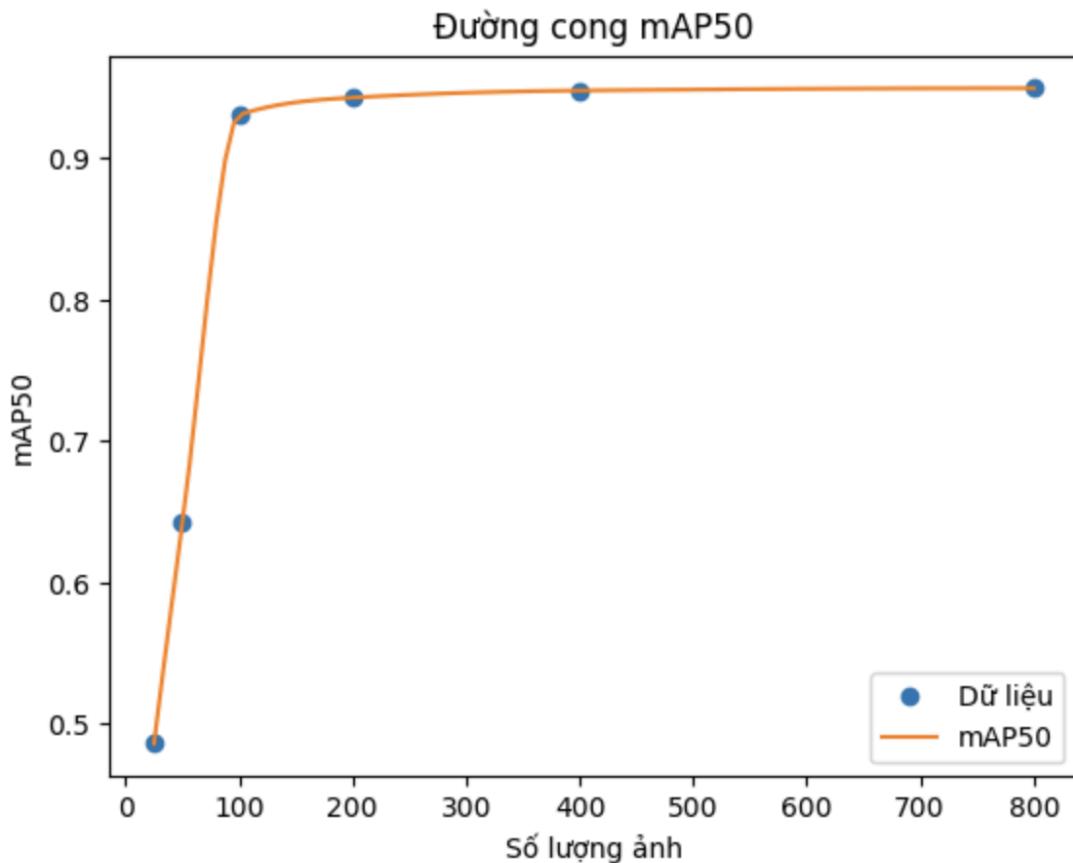
4.4 Hậu xử lý và Triển khai

Kiểm tra hiệu suất của mô hình huấn luyện và đánh giá độ chính xác, đánh giá mAP-50 và các thước đo hiệu suất khác để đảm bảo rằng mô hình hoạt động hiệu quả và so sánh với các mô hình khác để xác định mô hình hoạt động tốt nhất. Ở đây dự án này lựa chọn mô hình YOLOv8x. Sau khi quy trình huấn luyện hoàn thành, mô hình huấn luyện cung cấp hai trọng số: trọng số giai đoạn cuối cùng là trọng số của mô hình sau khi đã học từ toàn bộ tập dữ liệu huấn luyện và trọng số tốt nhất là trọng số được ghi lại tại thời điểm mà mô hình đạt được hiệu suất tốt nhất trên tập dữ liệu kiểm tra. Chúng tôi sử dụng trọng số tốt nhất để triển khai mô hình cho thí nghiệm của chúng tôi.

V. KẾT QUẢ

1. Đánh giá quá trình huấn luyện

Nhìn chung sau khoảng 100 bước huấn luyện, các mô hình đều hội tụ và đạt được điểm tối ưu. Các đại lượng accuracy đặc trưng cho độ chính xác, phân biệt đúng sai của dữ liệu và nhãn của nó, cụ thể là Precision acc (độ tin cậy), Recall acc (độ nhạy), đa phần các mô hình đều hội tụ đến trên 90% acc. mAP-50 đại diện cho độ chính xác của bounding box xác định được khá chính xác vật thể. Ngoài ra với dữ liệu mAP trong các tập training dataset từ 25,50, 100, 200, ..., 800 cho thấy rằng khoảng 200-400 ảnh là đủ cho việc huấn luyện mô hình. Các thông số cài đặt huấn luyện nếu không nhắc tới sẽ được lấy theo mặc định của YOLOv8 như learning rate hay là mô hình đánh giá sai số.



Hình 12 : So sánh mức độ mAP50 theo số lượng ảnh khác nhau

3.2 Đánh giá hiệu suất mô hình

Việc đánh giá hiệu suất mô hình sau khi huấn luyện nhằm đảm bảo rằng mô hình hoạt động đúng đắn, hiệu quả và phát hiện có đủ số lượng cá trong bể hay không do đó việc đánh giá cần được thực hiện một cách tỉ mỉ và chính xác.

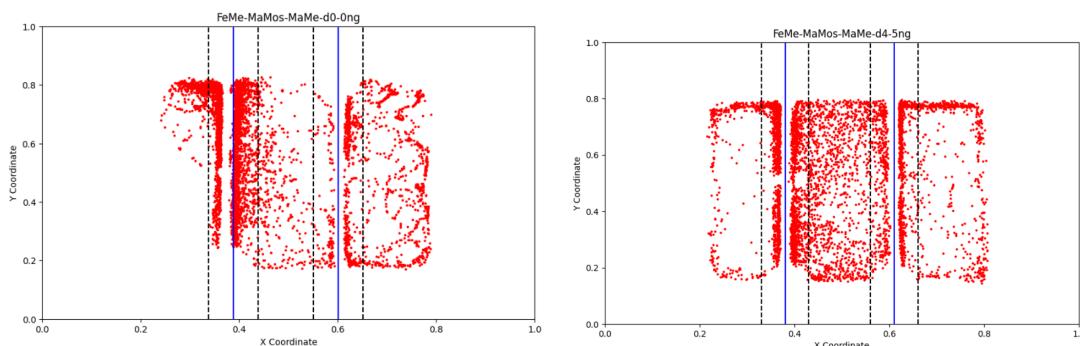
Việc đánh giá sẽ dựa trên số lượng cá bị phát hiện thiếu trong quá trình thí nghiệm, đó là một thước đo quan trọng để đánh giá được hiệu suất của mô hình. Nếu mô hình đạt độ chính xác tốt nhất, nó sẽ phát hiện được tất cả 11 con cá / 11 con cá trong toàn bộ quá trình thử nghiệm. Tuy nhiên, do một số ảnh hưởng như việc setup máy quay, góc độ quay, ánh sáng yếu hoặc các con cá bị che bởi ánh đèn dẫn đến việc các con cá bị khuất sẽ không được phát hiện bằng YOLO.

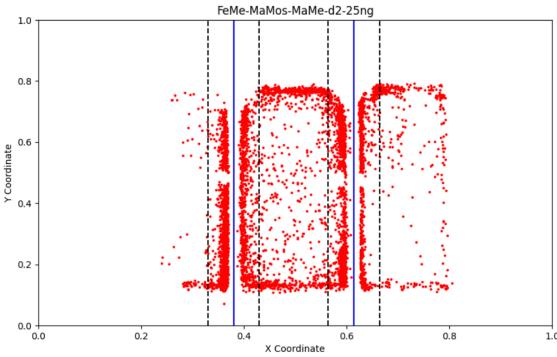
Kết quả sau khi sử dụng mô hình để phát hiện cá sẽ giúp chúng ta thu thập được toàn bộ vị trí của từng con cá trong quá trình thí nghiệm. Điều này cung cấp cho chúng ta các thông tin để phân tích hành vi của chúng bằng cách xác định vị trí của các con cá và theo dõi chúng theo thời gian, từ đó có thể xác định được các khu vực có mật độ cá lớn nhất từ đó có thể phân tích sự thay đổi trong hành vi của chúng.

3.3 Thu thập dữ liệu có ích để phân tích hành vi của cá

Theo như quan sát, với nồng độ estrogen khác nhau thì hành vi của cá có một vài sự thay đổi. Với nồng độ estrogen càng cao thì các con cá thường tập trung với nhau更多, mật độ cá xung quanh thành bể cao hơn. Ta chia vùng giới hạn đánh giá là các đường nét đứt màu đen vào khoảng 10mm, mỗi bên vách ngăn là 5mm. Đường màu xanh là 2 thành bể và các chấm màu đỏ là vị trí các con cá được phát hiện. Ta thấy rằng trong khoảng phạm vi 10mm, mật độ cá tập trung rất nhiều và các con cá có xu hướng lại gần nhau hơn. Với các thí nghiệm tương tự với nồng độ cao hơn thì thấy rằng các con cá có xu hướng tiến lại gần nhau khi nồng độ estrogen càng cao.

Ví dụ hình 14 dưới đây là Datavisualization của mô hình 1 trường hợp ngày 0- 0ng/l và ngày 2- 5ng/l , ngày 2- 25ng/l estrogen .

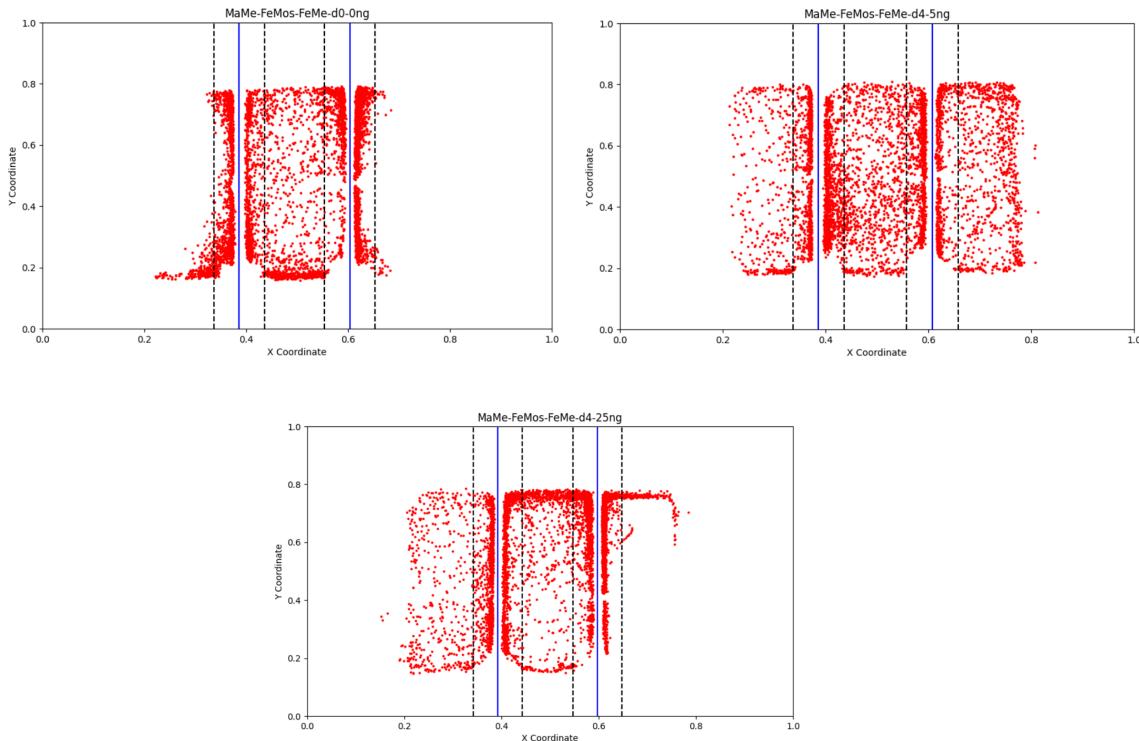




Hình 14: Datavisualization trong mô hình 1 với nồng độ estrogen lần lượt ngày 0 -0ng/l, ngày 2 -5ng/l, ngày 2 -25ng/l

Quan sát đồ thị hình 14 , chúng tôi nhận thấy rằng khi nồng độ estrogen ngày 0 -0ng/l thì mật độ chấm đỏ thường thưa và không dày như khi nồng độ estrogen tăng lên là ngày 2 -5ng/l, ngày 2 -25ng/l , mật độ cá tập trung cạnh 2 vách ngăn cũng nhiều hơn.

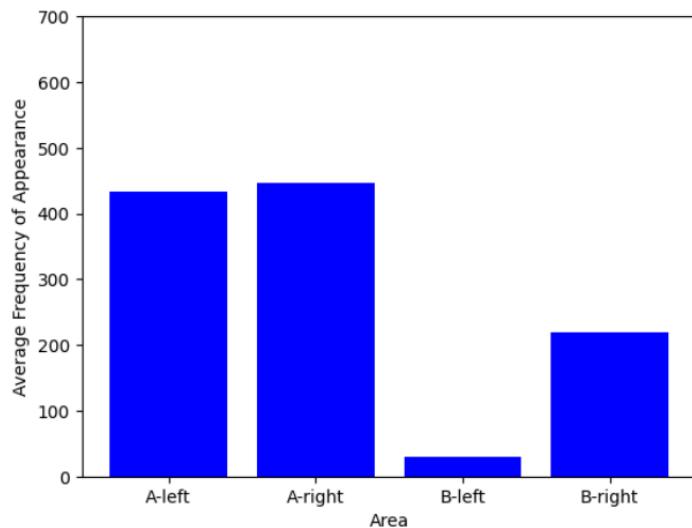
Ví dụ hình 15 dưới đây là Datavisualization của mô hình 2 trường hợp ngày 0 -0ng/l và ngày 4 -5ng/l ,ngày 4 -25ng/l estrogen .



Hình 15: Datavisualization trong mô hình 2 với nồng độ estrogen lần lượt ngày 0 -0ng/l, ngày 4 -5ng/l, ngày 4 -25ng/l

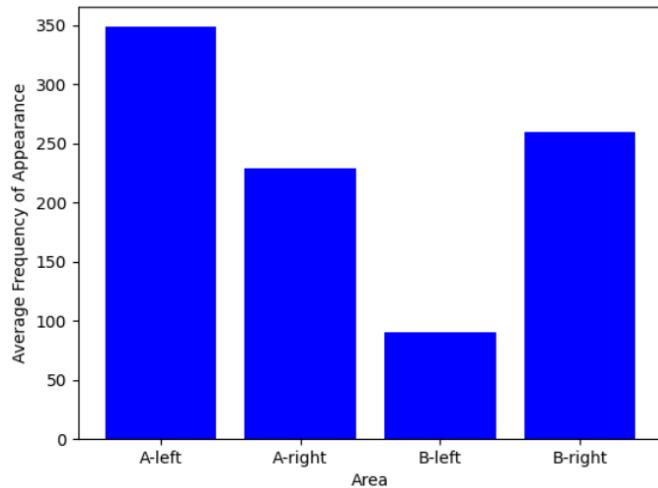
Quan sát đồ thị hình 15, chúng tôi nhận thấy rằng khi nồng độ estrogen ngày 0 -0ng/l thì mật độ chấm đỏ không dày như khi nồng độ estrogen tăng lên là ngày4 -5ng/l, ngày4 -25ng/l , mật độ cá tập trung cạnh 2 vách ngăn tăng nhiều hơn. Từ đó ta có thể dự đoán rằng với nồng độ estrogen càng cao và trong thời gian càng lâu thì những con cá sẽ bị thu hút lẫn nhau.

Để có thể đánh giá tốt hơn nữa về vị trí của cá , ta xây dựng một biểu đồ về tần suất xuất hiện của cá tập trung ở 2 bên vách ngăn . Đánh giá với mô hình 1 , ta có hình 16 là biểu đồ tần suất xuất hiện trung bình của cá ở 2 bên vách ngang A và vách ngăn B .



Hình 16: Biểu đồ tần suất xuất hiện trung bình của cá của mô hình 1 với nồng độ estrogen ngày 0 -0ng/l

Dựa vào đồ thị hình 16 ta thấy rằng tần suất xuất hiện của cá ở 2 bên vách ngăn A là khá lớn khoảng gần 450 lần . Còn lại ở vách ngăn B bên trái có tần suất xuất hiện cá rất ít khoảng 30 lần còn vách B bên phải xuất hiện khoảng 200 lần .

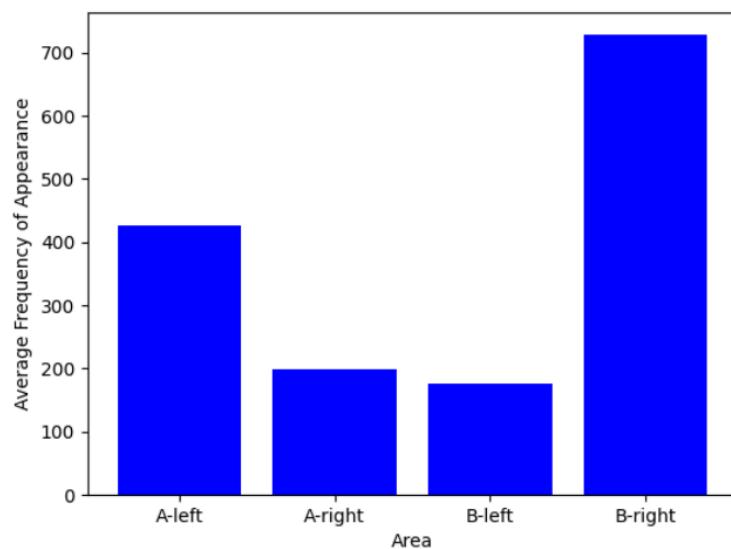


Hình 17: Biểu đồ tần suất xuất hiện trung bình của cá của mô hình 1 với nồng độ estrogen ngày 4 -5ng/l

Hình 17 ta thấy rằng với nồng độ estrogen là ngày 2 -5ng/l , tần suất xuất hiện của cá ở 2 bên vách ngăn A khá lệch nhau , khi ngăn trái A xuất hiện cá gần 350 lần thì ngăn phải A chỉ có khoảng gần 220 lần xuất hiện cá, bên vách B tần suất xuất hiện của cá ở ngăn trái vào khoảng gần 100 lần , nhưng ngăn bên phải có gần 250 lần xuất hiện cá .

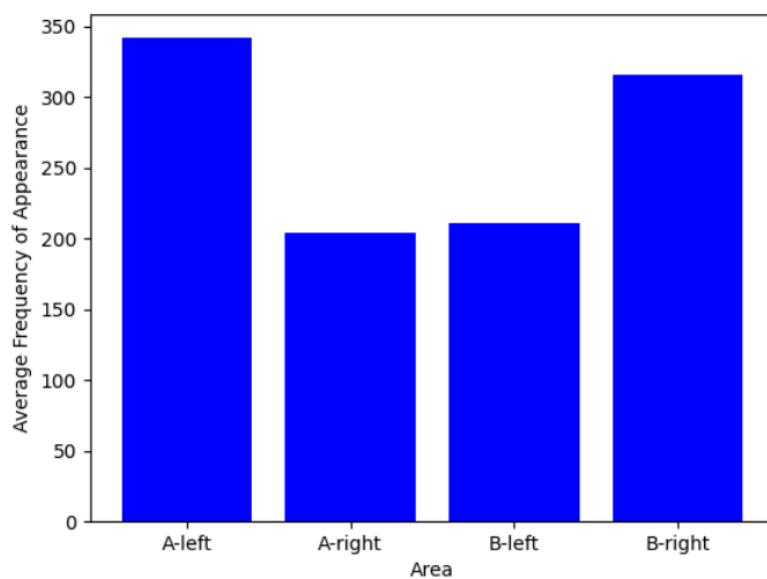
Điều đó cho ta thấy rằng hành vi của cá cũng có phần bị ảnh hưởng bởi nồng độ estrogen cao hay thấp .

Tương tự , ta thu được kết quả của mô hình 2 như sau :



Hình 18: Biểu đồ tần suất xuất hiện cá của mô hình 2 với nồng độ estrogen ngày 0 -0ng/l

Dựa vào đồ thị hình 18 với nồng độ estrogen là 0ng/l , mô hình 2 ta thấy rằng tần suất xuất hiện của cá ở 2 bên vách ngăn A lần lượt là vách ngăn A trái khoảng 400 lần , bên phải khoảng 200 lần. Còn lại ở vách ngăn B bên trái có tần suất xuất hiện cá khoảng 200 lần còn vách B bên phải xuất hiện cao khoảng 700 lần và 2 bên vách ngăn B lệch nhau khá nhiều .



Hình 19: Biểu đồ tần suất xuất hiện cá của mô hình 2 với nồng độ estrogen ngày 2 -5ng/l

Hình 19 ta thấy rằng với nồng độ estrogen là 5ng/l , tần suất xuất hiện của cá ở 2 bên vách ngăn A khá lệch nhau , khi ngăn trái A xuất hiện cá hơn 350 lần thì ngăn phải A chỉ có khoảng gần 200 lần tần suất xuất hiện cá, bên vách B tần suất xuất hiện của cá cũng khoảng như vậy .

VI. YOLOv5 vs YOLOv8

Đối với cả YOLOv5 và YOLOv8, ta thấy có một số biến động trong giá trị mAP50 ở các epoch đầu tiên, sau đó giá trị mAP50 dần ổn định hơn. Các biến động này có thể xuất phát từ việc mô hình đang trong quá trình học và cần thêm thời gian để hội tụ.

Giá trị mAP50 của YOLOv8x tăng mạnh ngay sau epoch đầu tiên nhưng sau đó có một số giá trị biến động trong các epoch tiếp theo.

Về cuối, hiệu suất của nó tăng dần và ổn định hơn. Giá trị mAP50 của nó gần sát với 0.95 ở các epoch cuối, cho thấy một hiệu suất khá tốt.

Cả YOLOv5 và YOLOv8 đều cho kết quả mAP50 tốt sau một số epoch, tuy nhiên YOLOv5m và YOLOv5n có xu hướng ổn định hơn so với các phiên bản khác.

Đối với YOLOv8, YOLOv8n ổn định hơn trong quá trình huấn luyện so với YOLOv8x, dù tất cả đều đạt được hiệu suất tương tự ở cuối quá trình huấn luyện.

Nếu muốn một mô hình có hiệu suất tốt ngay từ các epoch đầu, YOLOv8n có thể là lựa chọn tốt hơn. Trong khi đó, nếu muốn tối ưu hóa quá trình huấn luyện hơn, YOLOv8x có thể cung cấp cho ta nhiều thông tin hơn qua sự biến động của nó.

VII. KẾT LUẬN

Việc sử dụng thị giác máy tính để giám sát hành vi của cá có thể đóng góp cho xã hội ở nhiều mức độ và trong nhiều lĩnh vực khác nhau. Giám sát và phân tích hành vi của cá giúp nhà khoa học nắm bắt được những thay đổi về môi trường sống của chúng, như ô nhiễm nước, biến đổi khí hậu hay các yếu tố khác. Thông qua việc giám sát, các nhà nghiên cứu có thể hiểu rõ hơn về hành vi sinh sản, di cư, và tương tác xã hội giữa các loài cá. Điều này hỗ trợ việc phát triển các giải pháp bảo tồn cho các loài cá đang bị đe dọa. Việc theo dõi hành vi của cá có thể giúp nhận biết các dấu hiệu về sức khỏe, tăng trưởng và sự phát triển của chúng. Điều này giúp người nuôi cá tối ưu hóa điều kiện nuôi cấy và giảm thiểu tỷ lệ cá chết. Bằng cách minh họa và trình bày thông tin về hành vi của cá, dự án có thể giáo dục công chúng về tầm quan trọng của việc bảo tồn môi trường sống và tác động của con người lên thế giới tự nhiên.

Tóm lại, mô hình phát hiện cá là một công cụ quan trọng trong việc quản lý nguồn lợi cá và bảo vệ môi trường nước. Tuy nhiên, việc giảm thiểu nguồn nước bị ô nhiễm và quản lý tốt nguồn nước là yếu tố quan trọng để bảo vệ sức khỏe và hành vi tự nhiên của cá.

I. Tài liệu tham khảo

- [1] Léon-Olea M, et al. (2014) Current concepts in neuroendocrine disruption, General and Comparative Endocrinology 203: 158-173.
- [1] Petrie B, et al. (2015) A review on emerging contaminants in wastewaters and the environment: Current knowledge, understudied areas and recommendations for future monitoring. WaterResearch 72: 3-27.
- [2] Le, T. T. B., Huynh, U. D. H., Tuyet, H. T. D., & Ha, T. L. B. (2022). Study on the endometrium regeneration treatment using human injured uterine tissue model. *VNUHCM Journal of Natural Sciences*, 6(2), 2015-2022.
- [3] Pyke, G. H. (2005). A review of the biology of *Gambusia affinis* and *G. holbrooki*. *Reviews in Fish Biology and Fisheries*, 15, 339-365.
- [4] Iwamatsu, T. (2004). Stages of normal development in the medaka *Oryzias latipes*. *Mechanisms of development*, 121(7-8), 605-618.
- [5] Xu, M., David, J. M., & Kim, S. H. (2018). The fourth industrial revolution: Opportunities and challenges. *International journal of financial research*, 9(2), 90-95.
- [6] Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience*, 2018.

[7] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).

[8] Ujjwakan , An Intuitive Explanation of Convolutional Neural Networks (2016)

[9] MathWorks, What Is a Convolutional Neural Network? (2023)

[10] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 37(9), 1904-1916.

[11] Hidayatullah, P. (2024). YOLOV8 ARCHITECTURE . LinkedIn.

https://www.linkedin.com/posts/dr-priyanto-hidayatullah-38632715_hi-deep-learning-computer-vision-enthusiast-activity-7123935145916567552-9Ao_/ [2023]

[12] Ju, R. Y., & Cai, W. (2023). Fracture detection in pediatric wrist trauma X-ray images using YOLOv8 algorithm. *Scientific Reports*, 13(1), 20077.

[13] (Zhu, J., Hu, T., Zheng, L., Zhou, N., Ge, H., & Hong, Z. (2024). YOLOv8-C2f-Faster-EMA: An Improved Underwater Trash Detection Model Based on YOLOv8. *Sensors*, 24(8), 2483)

[14] Ultralytics. (n.d.). YOLOv8. Retrieved 2023, from
<https://docs.ultralytics.com/models/yolov8/#overview>

[15] YOLOv8. (2024). Neck Architecture. In YOLOv8 Architecture Explained. Retrieved from https://yolov8.org/yolov8-architecture-explained/#2_Neck_Architecture

[16] Glenn-jocher , Sergiuwaxmann: Ultralytics YOLOv5 Architecture (2023)

[17] Padilla, R., Netto, S. L., & Da Silva, E. A. (2020, July). A survey on performance metrics for object-detection algorithms. In *2020 international conference on systems, signals and image processing (IWSSIP)* (pp. 237-242). IEEE.

[18] Gad, A. F. (2020). Evaluating object detection models using mean average precision (mAP). *PaperspaceBlog*.

[19] Everingham, M., & Winn, J. (2012). The PASCAL visual object classes challenge 2012 (VOC2012) development kit. *Pattern Anal. Stat. Model. Comput. Learn., Tech. Rep.*, 2007(1-45), 5.

[20] Evidently AI Team: Accuracy vs. precision vs. recall in machine learning: what's the difference?

[21] Deva Shah: Intersection over Union (IoU): Definition, Calculation, Code

[22] Mayank Mishra. Published in [Towards Data Science](#). Convolutional Neural Networks, Explained

