Đề tài: NHẬN DIỆN CÁC LOÀI ĐỘNG VẬT SỬ DỤNG MÔ HÌNH YOLOv7

Thực hiện: Nguyễn Quang Tính

Business Understanding

- Hệ động vật vô cùng đa dạng với vô vàn những loài sống trên cạn, loài bay trên trời cho đến loài sống dưới nước,... Chúng ẩn chứa rất nhiều điều kỳ diệu và là một mắt xích quan trọng góp phần gia tăng sự phong phú cũng như tạo nên một hệ sinh thái tự cân bằng tự nhiên.
- Dạy cho trẻ em biết các con vật là cơ hội để trẻ làm quen với thế giới động vật nói riêng và môi trường tự nhiên nói chung. Trong quá trình đó, trẻ học hỏi được những bài học mới về muông thú giúp trẻ nuôi dưỡng tâm hồn, tình yêu thương và ý thức bảo vệ thế giới tự nhiên.

Business Understanding

Mục tiêu:

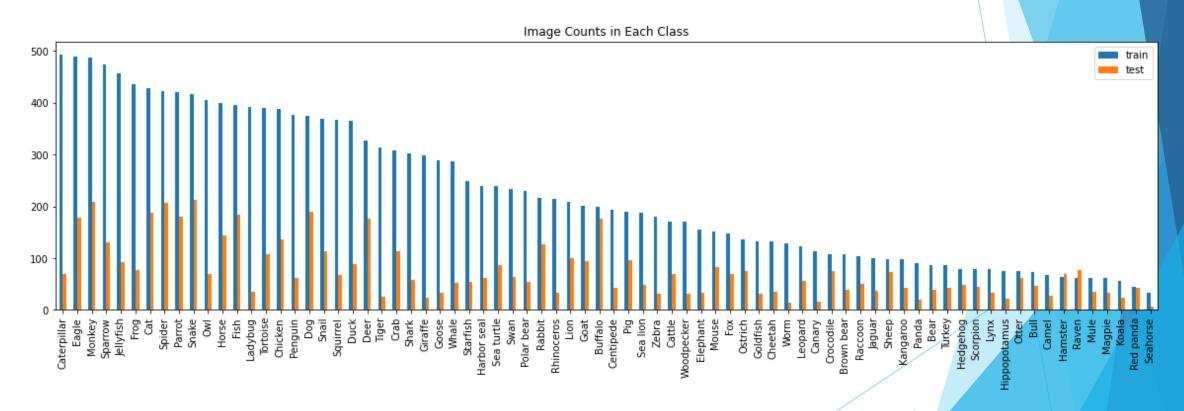
Nhận diện được các loài động vật có trong ảnh và video giúp bé biết được tên loài vật.

Tiêu chí đánh giá:

MAP@0.5 (Mean Average Precision ứng với ngưỡng IoU = 0.5) > 0.6

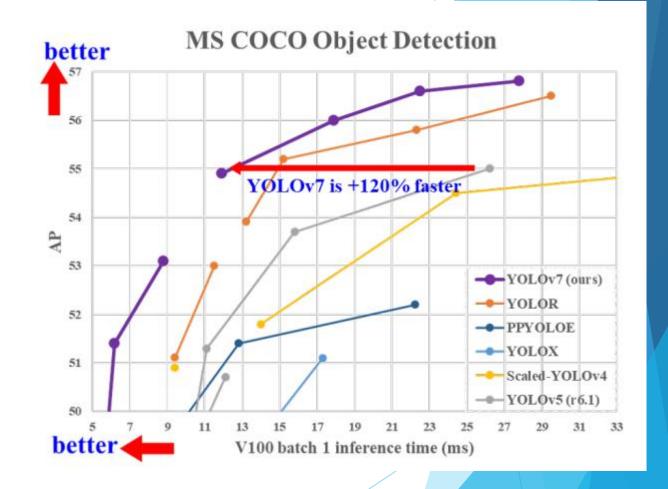
Data Understanding

• Số lượng dữ liệu: có 21758 ảnh của 76 loài động vật chia thành 2 tập Train và Test. Trong đó tập Train chứa 16363 ảnh và tập Test chứa 5395 ảnh. Mỗi file ảnh đều có file text tương ứng chứa tên loài vật và tham số bouding box của chúng.



Đề xuất phương pháp

- Với bài toán Object Detection thì YOLO là thuật toán thường được sử dụng vì tốc độ nhanh mà vẫn có độ chính xác khá cao. Trong đồ án này đế xuất sử dụng YOLOv7 vì các ưu điểm của nó.
- YOLOv7 vượt qua mọi model Object Detection trong cả tốc độ và độ chính xác từ 5 FPS tới 160 FPS và đạt độ chính xác cao nhất với 56.8% AP trong số toàn bộ các model Object Detection real-time, có tốc độ 30 FPS hoặc hơn trên GPU V100. YOLOv7 cũng vươt qua cả: YOLOR, YOLOX, Scaled-YOLOv4, YOLOv5, DETR, Deformable DETR, DINO-5scale-R50, ViT-Adapter-B cũng như rất nhiều mạng Object Detection khác cả về mặt tốc độ cũng như độ chính xác. Hơn nữa, YOLOv7 được train trên COCO từ đầu mà không sử dụng bất kì pretrained nào.

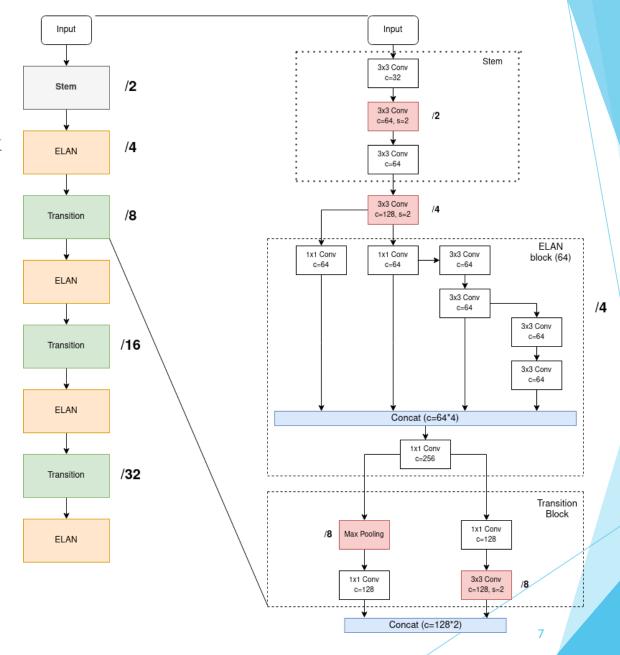


Kiến trúc YOLOv7

- Backbone: ELAN
- Neck: SPPCSPC + (CSP-OSA)PANet + RepConv
- Head: YOLOR + Auxiliary head

Backbone

- Trước khi tiến vào ELAN Block đầu tiên trong backbone, ảnh đầu vào sẽ đi qua Stem Block
- Các ELAN Block được kết nối với nhau thông qua các Transition Block. Mỗi Transition Block là một lần giảm kích cỡ của feature maps đi 2 lần.
- Backbone hoàn chỉnh của YOLOv7 là tập hợp của các ELAN Block và các Transition block



Neck

- Thực hiện phát hiện vật thể ở trên những scale khác nhau.
- Trong Top Down pathway của YOLOv7, feature maps từ scale /16 sau khi được kết hợp với feature maps từ scale /32 sẽ được xử lý thêm thông qua CSP-OSA rồi mới kết hợp với feature maps từ scale /8.
- Với Bottom Up pathway của YOLOv7, feature maps từ scale dưới sẽ được đi qua CSP-OSA để xử lý, và downsample thông qua Transition Block trước khi kết hợp với feature maps của scale phía trên.

1x1 Conv

c=1024

3x3 Conv

c=1024

1x1 Conv c=1024

1x1 Conv

5x5 MP

9x9 MP

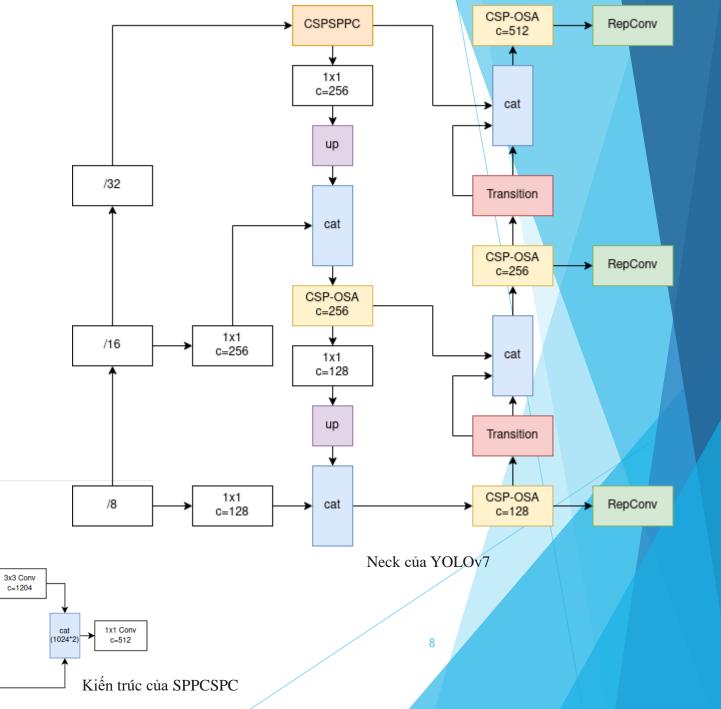
same

13x13 MP

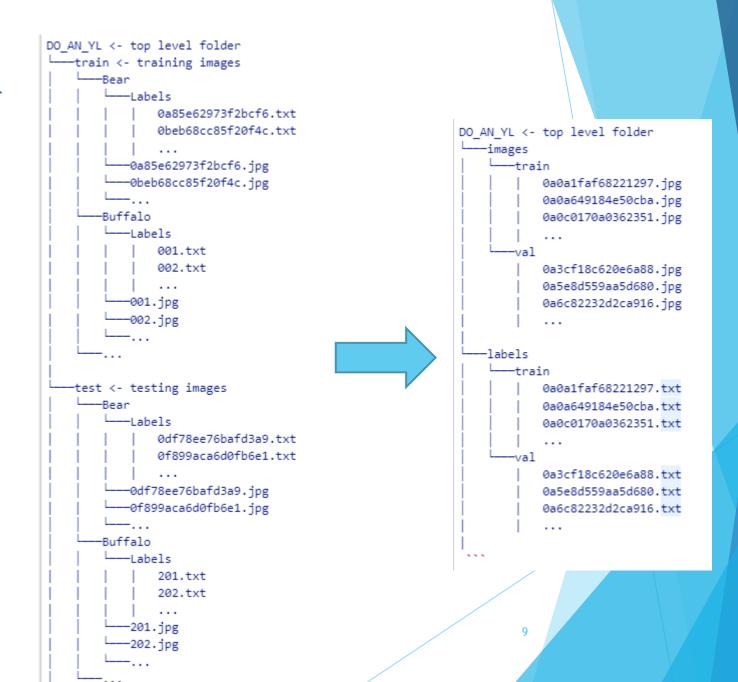
1x1 Conv

c=1024

(1024*4)



Data Preparation



Huấn luyện YOLOv7 với dữ liệu

Chỉnh sửa file yolov7.yaml và tạo file mydataset.yaml để phù hợp với number of classes của dữ liệu.

```
yolov7.yaml X
1 # parameters
2 nc: 76 # number of classes
3 depth multiple: 1.0 # model depth multiple
4 width multiple: 1.0 # layer channel multiple
# Tạo file data yaml
!rm data/mydataset.yaml
!echo 'train: ../images/train' >> data/mydataset.yaml
!echo 'val: ../images/val' >> data/mydataset.yaml
!echo 'nc: 76' >> data/mydataset.yaml
with open("data/mydataset.yaml", "a") as yaml file:
    yaml file.write("names: "+str(animals to detect))
 mydataset.yaml X
 1 train: ../images/train
 2 val: ../images/val
 3 nc: 76
 4 names: ['Bear', 'Brown bear', 'Buffalo', 'Bull', 'Camel', 'Canary', 'Cat',
```

Huấn luyện YOLOv7 với dữ liệu

obi

Labels

6222

0.01767 0.006085 0.008709

box

Images

5395

Epoch

149/149

gpu mem

12.6G

Class

all

cls

total

0.03247

0.618

```
# Clone yolov7
 !git clone https://github.com/WongKinYiu/yolov7.git
 cd yolov7
 !pip install -r requirements.txt
 # Tải weight yolov7
 !wget https://github.com/WongKinYiu/yolov7/releases/download/v0.1/yolov7.pt
 # Train model YOLOv7 với dữ liêu
 !python train.py --epochs 150 --batch-size 16 --data data/mydataset.yaml --img 640 640 --cfg cfg/training/yolov7.yaml
 --weights yolov7.pt --hyp data/hyp.scratch.custom.yaml
 !python train.py --epochs 150 --batch-size 16 --data data/mydataset.yaml --img 640 640 --cfg cfg/training/yolov7.yaml
 --weights runs/train/exp9/weights/last.pt --hyp data/hyp.scratch.custom.yaml
  Epoch
          gpu mem
                        box
                                  obi
                                            cls
                                                    total
                                                            labels img size
                   0.01748 0.006175 0.009019
                                                                          640: 100% 1023/1023 [16:17<00:00, 1.05it/s]
147/149
           12.6G
                                                 0.03268
            Class
                                   Labels
                                                                      mAP@.5 mAP@.5:.95: 100% 169/169 [01:49<00:00, 1.55it/s]
                       Images
              all
                         5395
                                     6222
                                                                       0.623
                                                0.638
                                                            0.592
                                                                                    0.488
                                  obi
                                            cls
                                                    total
                                                            labels img size
  Epoch
          gpu mem
                       box
                                                                          640: 100% 1023/1023 [16:41<00:00, 1.02it/s]
148/149
           12.6G
                   0.01752 0.006203 0.00893
                                                 0.03265
                                                                 21
                                                                      mAP@.5 mAP@.5:.95: 100% 169/169 [01:49<00:00, 1.54it/s]
            Class
                                   Labels
                                                    Ρ
                       Images
              all
                         5395
                                     6222
                                                 0.64
                                                            0.592
                                                                       0.623
                                                                                    0.489
```

labels img size

0.624

22

0.609

640: 100% 1023/1023 [16:43<00:00, 1.02it/s]

0.49

mAP@.5 mAP@.5:.95: 100% 169/169 [01:56<00:00, 1.45it/s]

Los Function

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

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

cell contain object

probability distribution classes

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

- $\mathbb{1}_i^{\text{obj}}$: Hàm indicator có giá trị 0,1 nhằm xác định xem cell i có chứa vật thể hay không. Bằng 1 nếu chứa vật thể và 0 nếu không chứa.
- $\mathbb{1}_{ij}^{\text{obj}}$: Cho biết bounding box thứ j của cell i có phải là bouding box của vật thể được dự đoán hay không?
- ullet C_{ij} : Điểm tin cậy của ô i, P(contain object) * IoU (predict bbox, ground truth bbox)
- \hat{C}_{ij} : Điểm tự tin dự đoán.
- C: Tập hợp tất cả các lớp.
- $p_i(c)$: Xác suất có điều kiện, có hay không ô i có chứa một đối tượng của lớp $c \in \mathcal{C}$.
- $\hat{p}_i(c)$: Xác suất có điều kiện dự đoán.

Metrics

IoU (Intersection over Union) là độ tương đồng giữa ground truth bounding box với predicted bounding box của mô hình

$$IoU_{pred}^{truth} = \frac{truth \cap pred}{truth \cup pred}$$

Precision và Recall

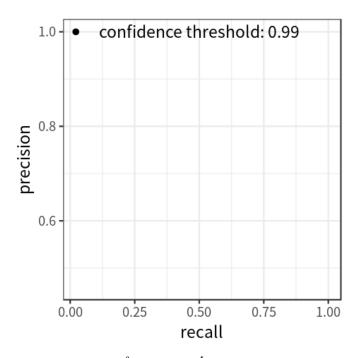
$$Precision = \frac{TP}{TP + FP} \qquad \qquad Recall = \frac{TP}{TP + FN}$$

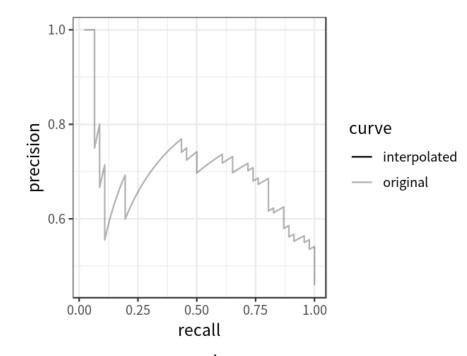
Với bài toán Object Detection:

- •TP (True Positive): là các predicted box với IoU lớn hơn hoặc bằng giá trị IoU cố định
- •FP (False Positive): là các predicted box với IoU nhỏ hơn giá trị IoU cố định
- •FN (False Negative): mô hình không bắt được đối tượng trong ảnh (ứng với ground truth tương ứng) Vậy ta có Precision và Recall cho bài toán Object Detection:

$$Precision = \frac{true_object_detection}{all_detected_boxes} \qquad \qquad Recall = \frac{true_object_detection}{all_ground_truth_boxes}$$

Metrics

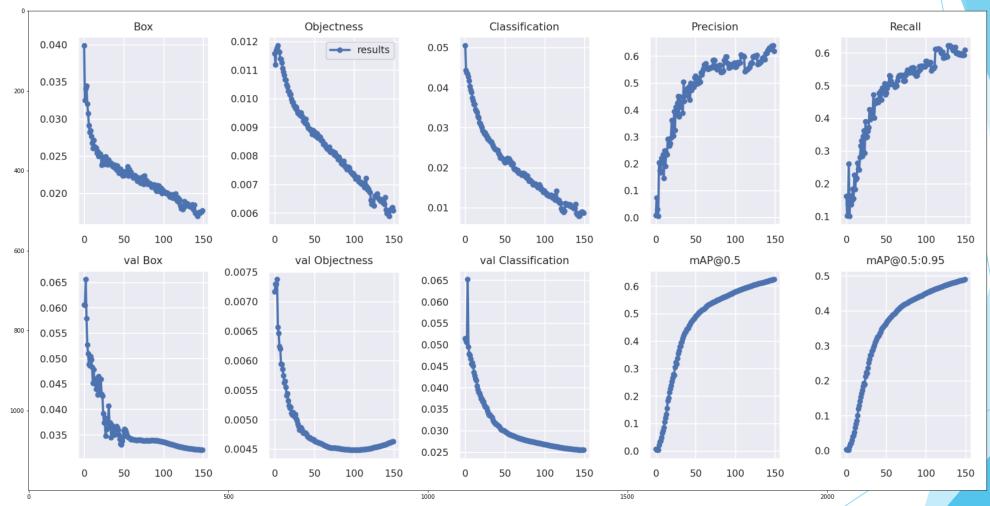




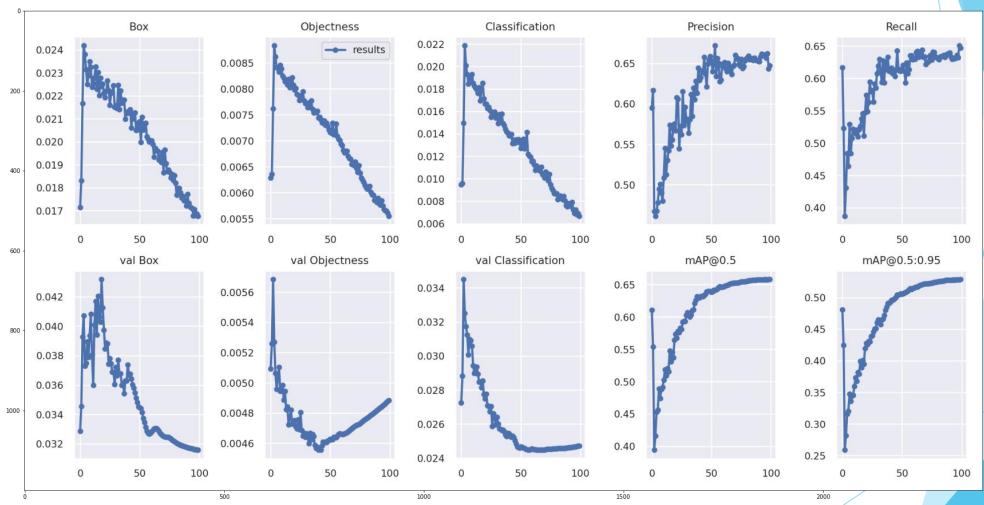
PR-curve thể hiện mối liên hệ giữa 2 giá trị Precision và Recall khi điều chỉnh ngưỡng confidence score

AP (Average Precision) là trung bình precision ứng với các mốc recall tương ứng trong đoạn từ 0 đến 1, tính toán bằng phần diện tích bao phủ dưới PR-curve theo công thức:

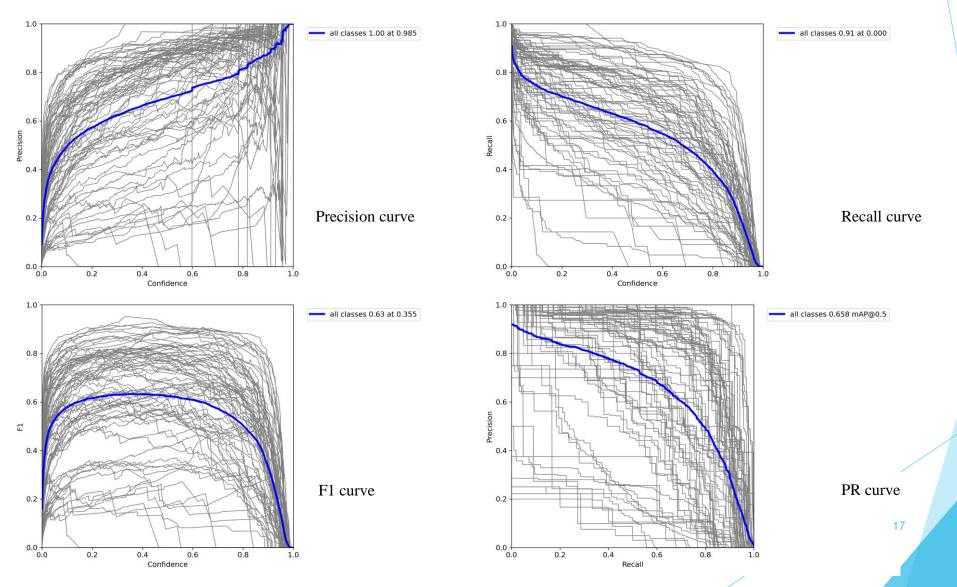
bang phan diện tich bao phù dươi PR-curve theo công thức:
$$AP = \sum_{i=1}^{N} (r_n - r_{n-1}) p_{interpolated} (r_n)$$
 mAP là trung bình AP score của N class, được định nghĩa với công thức:
$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i$$

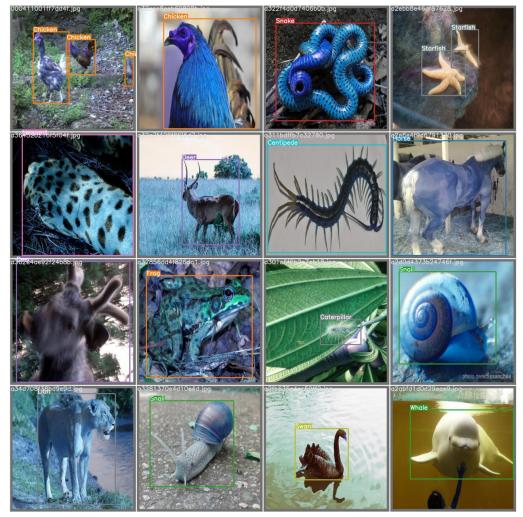


Nhận xét: Loss giảm dần, mAP có chiều hướng tăng cho thấy quá trình học của mô hình khá tốt.



Nhận xét: mAP chỉ cải thiện ít ở những epoch cuối cùng.



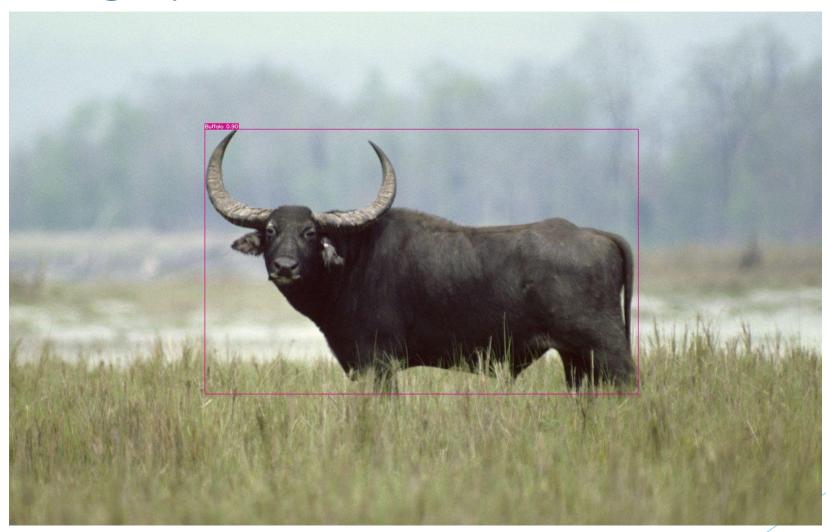




Label

Nhận xét

- Mô hình nhận diện tốt đối với ảnh chụp đầy đủ các bộ phận của động vật.
- Có thể không nhận diện được hoặc nhận diện sai nếu ảnh chỉ có một bộ phận nhỏ của động vật.
- Còn nhầm lẫn giữa các loài có nhiều nét tương đồng nhau như: Whale và Shark









Đề Xuất Phát Triển

- Cần tinh tế trong việc chọn dữ liệu ngay từ đầu để giảm bớt sự tương đồng các đặc điểm của các loài động vật khác nhau.
- ► Hoàn thiện tốt có thể ứng dụng để dạy cho trẻ em nhận biết các loài động vật khác nhau.