

# ĐẠI HỌC QUỐC GIA HÀ NỘI

## TRƯỜNG ĐẠI HỌC CÔNG NGHỆ

### Segmentation Cải Bó Xôi

Học phần: Học máy

Lý Trường Phước

Mã sinh viên: 21020934

## I. GIỚI THIỆU

Cải bó xôi (Spinach) là một loại rau có giá trị dinh dưỡng cao, thường được trồng trong nhà kính hoặc trang trại nông nghiệp công nghệ cao. Trong các ứng dụng nông nghiệp chính xác (precision agriculture), việc phân đoạn (segmentation) cây cải bó xôi từ hình ảnh là bước quan trọng giúp tự động hóa các tác vụ như theo dõi tăng trưởng, ước lượng sinh khối, và phát hiện sớm sâu bệnh.

Trong đề tài này, nhóm thực hiện bài toán phân đoạn cây cải bó xôi từ hình ảnh đầu vào, sử dụng tập dữ liệu đã được gán nhãn thủ công (manual labelling). Mục tiêu là sử dụng một mô hình học sâu có thể tách chính xác vùng lá cây ra khỏi nền ảnh, phục vụ cho các ứng dụng phân tích và đánh giá tăng trưởng cây trồng trong thực tế.

## II. DỮ LIỆU

Với bộ dữ liệu được cung cấp sẵn bao gồm 180 ảnh với kích thước 2560x1920 pixel. Bộ dữ liệu được đưa lên CVAT đánh label thủ công:

- Chọn công cụ vẽ Polygon hoặc Brush để phân vùng từng lá.

- Gán nhãn cho mỗi lá là `'leaf'`.

- Mỗi vùng được vẽ riêng biệt sẽ là một instance khác nhau.

Quá trình đánh nhãn với các cây, lá độc lập quá trình gán nhãn rất dễ dàng. Với các cây trưởng thành có các phân vùng lá chèn lên nhau, quá trình đánh nhãn quy ước theo đặc trưng, 1 lá được xác định khi có màu sắc tương đồng lớn, có gân lá chính hoặc đủ cọng lá. Với các lá bị chèn vào nhau có thể gộp lại thành 1 lá chính đảm bảo nhãn phủ kín viền lá.

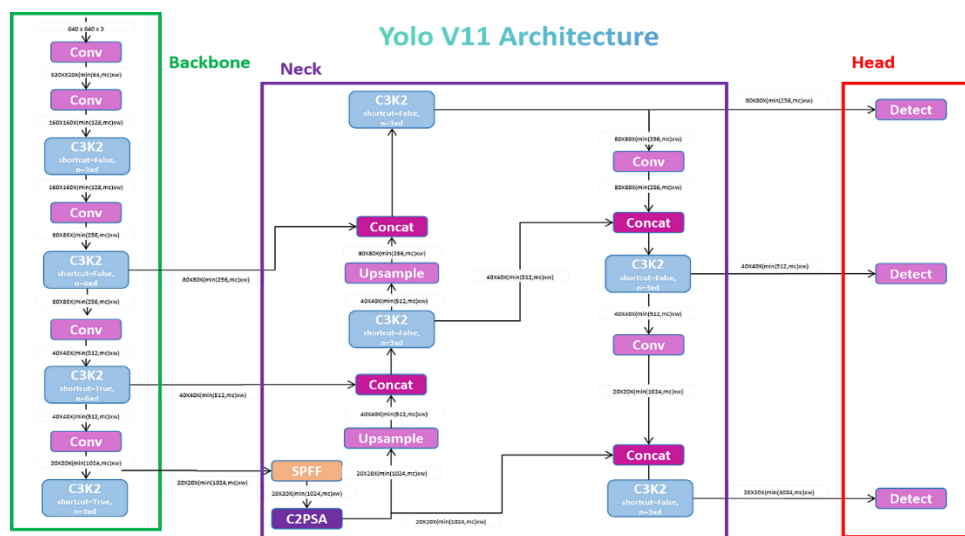
Hoàn thành quá trình đánh nhãn thực hiện xuất dữ liệu với định dạng Ultralytics YOLO Segmentation 1.0. Dữ liệu có được đảm bảo có thể thực hiện được bài toán segmentation với

mô hình YOLO. Tuy nhiên do giới hạn free của web chỉ có thể tải dc nhãn dán file xml và yaml về phần ảnh cần tự lấy từ bộ dữ liệu đã được cung cấp. Gộp ảnh có sẵn và file tải về được sau quá trình đánh nhãn sau đó đưa lên roboflow để thực hiện tăng cường ảnh và resize ảnh về 640x640( phù hợp với đa số các mô hình, nhỏ, nhẹ nhưng vẫn đảm bảo chất lượng).

## II.MÔ HÌNH

YOLOv11 là phiên bản mới thuộc dòng mô hình You Only Look Once (YOLO), nổi bật với tốc độ xử lý nhanh và khả năng nhận diện chính xác cao. Khác với các phiên bản YOLO trước đó chủ yếu được thiết kế cho bài toán detection (phát hiện vật thể), YOLOv11 hỗ trợ trực tiếp cho các bài toán segmentation (phân đoạn ảnh) thông qua nhánh xuất mask trong cấu trúc mạng.

YOLOv11 kế thừa thiết kế module Backbone–Neck–Head, đồng thời cải tiến mạnh ở phần Neck bằng cách tích hợp các kiến trúc như BiFPN (Bidirectional Feature Pyramid Network) hoặc SPPF++, giúp tăng khả năng trích xuất đặc trưng ở nhiều mức độ (multi-scale features). Bên cạnh đó, YOLOv11 áp dụng loss function tối ưu hơn cho segmentation như Binary Cross Entropy Loss (BCE) kết hợp Dice Loss, giúp mô hình học tốt hơn trong bài toán phân đoạn nhị phân như nhận diện vùng lá cải bó xôi.



Hình 1. Cấu trúc YOLOv11

### 1. Backbone (Khung xương) – màu xanh lá cây

- Dùng để trích xuất đặc trưng cơ bản từ ảnh đầu vào.
- Các lớp Conv và C3K2 (một biến thể của C3 với kernel size nhỏ hơn) giúp phát hiện đặc trưng cấp thấp.
- Kết hợp shortcut (residual connections) giúp truyền gradient tốt hơn trong quá trình huấn luyện.

### 2. Neck – màu tím

- Gồm nhiều tầng Concat và Upsample, kết hợp với các module đặc biệt như:
    - SPPF (Spatial Pyramid Pooling - Fast): giúp trích xuất đặc trưng ở nhiều tỉ lệ vùng khác nhau.
    - C2PSA (nếu đúng là Channel-wise Position Self Attention): tăng khả năng học phụ thuộc không gian và kênh.
  - Đảm nhiệm việc kết hợp đặc trưng từ nhiều tầng để phục vụ cho nhận diện và phân đoạn.
- 

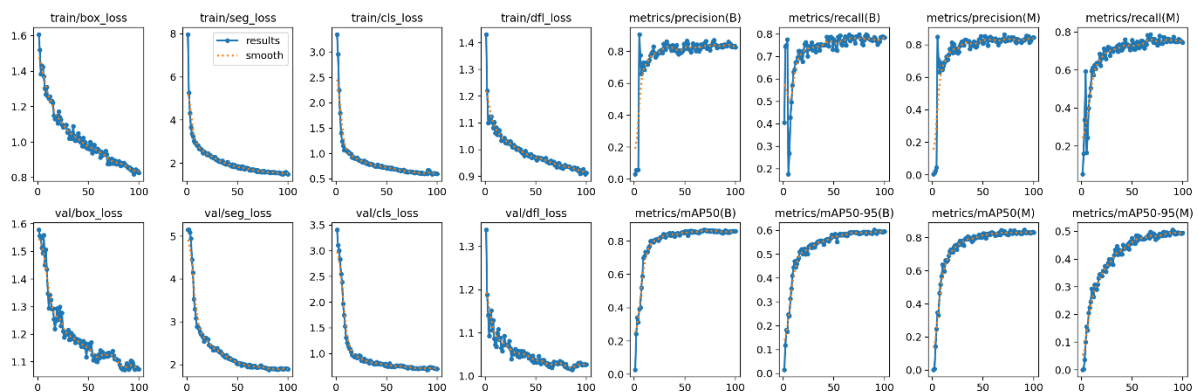
### 3. Head – màu đỏ

- Phần đầu ra chứa các lớp Detect cho các kích thước ảnh khác nhau (multi-scale output).
- Trong phiên bản segmentation, ngoài các box head còn có thể có mask head (không thể hiện rõ trong ảnh, có thể do bản thiết kế tùy biến).

Việc lựa chọn YOLOv11 cho bài toán segmentation cây cải bó xôi dựa trên các ưu điểm sau:

- Tốc độ xử lý nhanh, phù hợp với yêu cầu triển khai thời gian thực trong các hệ thống nông nghiệp thông minh.
- Hiệu suất cao trong bài toán segmentation, nhờ tích hợp nhánh sinh mặt nạ (mask head) và sử dụng các kỹ thuật tăng cường đặc trưng sâu.
- Kiến trúc gọn nhẹ, dễ dàng huấn luyện trên các bộ dữ liệu nhỏ như 180 ảnh đã được gán nhãn thủ công.
- Hỗ trợ tốt định dạng nhãn segmentation từ CVAT, cho phép tích hợp liền mạch vào pipeline huấn luyện mà không cần chuyển đổi phức tạp.

## IV. KẾT QUẢ



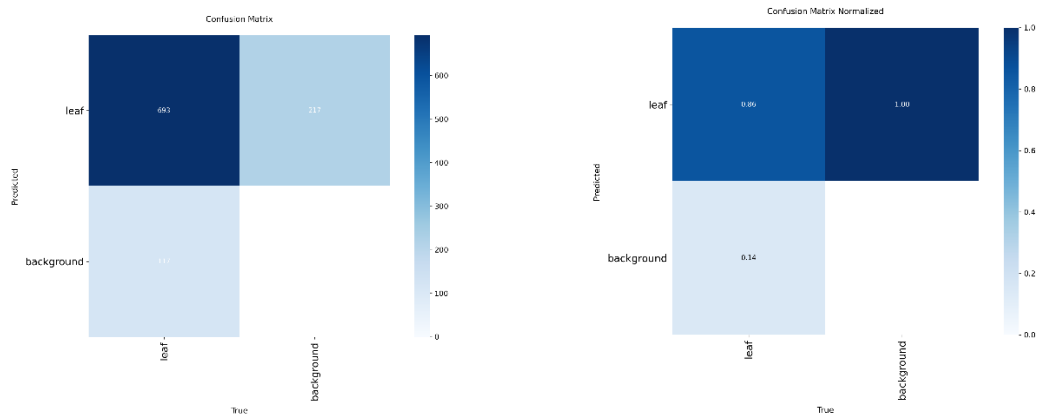
Hình 1 Đồ thị Loss Function

- train/box\_loss & val/box\_loss:
    - Giảm đều qua từng epoch, từ  $\sim 1.6$  về  $\sim 0.8$  (train) và  $\sim 1.1$  (val), cho thấy mô hình học được định vị tốt bounding box quanh vùng lá.
  - train/seg\_loss & val/seg\_loss:
    - Giảm mạnh từ  $\sim 8$  xuống dưới 2 (train), và từ  $\sim 5$  xuống  $\sim 2$  (val), chứng tỏ mô hình học tốt phân vùng mask.
  - train/cls\_loss & val/cls\_loss:
    - Giảm từ  $\sim 3.2$  xuống  $\sim 0.6$ , phản ánh khả năng phân biệt giữa “leaf” và “background” cải thiện rõ rệt.
  - train/df\_l\_loss & val/df\_l\_loss:
    - DFL (Distribution Focal Loss) cũng giảm rõ rệt, khẳng định mô hình học tốt phân bố các điểm neo (anchor) trong phát hiện vật thể.
- ⇒ Nhận xét tổng quan loss: Mô hình không gặp tình trạng overfitting rõ rệt, loss train và val đều giảm ổn định.

Các chỉ số đánh giá hiệu suất (Precision, Recall, mAP)

- Precision(B & M) đạt  $> 0.85$  sau 100 epochs  $\rightarrow$  mô hình dự đoán “leaf” chính xác, ít dương tính giả (false positive).
- Recall(B & M) đạt  $0.75\text{--}0.8$   $\rightarrow$  mô hình nhận diện được hầu hết các vùng thực sự là “leaf”, tuy vẫn bỏ sót một số ít vùng.
- mAP50(B & M) đạt  $\sim 0.85\text{--}0.88$   $\rightarrow$  hiệu quả mô hình cao khi chấp nhận sai số nhỏ ( $\text{IoU} \geq 0.5$ ).

- mAP50-95(B & M) đạt 0.5–0.6 → hiệu quả vẫn khá tốt ngay cả khi yêu cầu dự đoán khắt khe hơn về IoU.
- ⇒ Tổng kết: Các chỉ số mAP và Recall/Precision đều cho thấy mô hình huấn luyện thành công và đạt độ chính xác cao trong phân đoạn lá cải bó xôi.



Hình 2 Confusion Matrix

Ma trận thường:

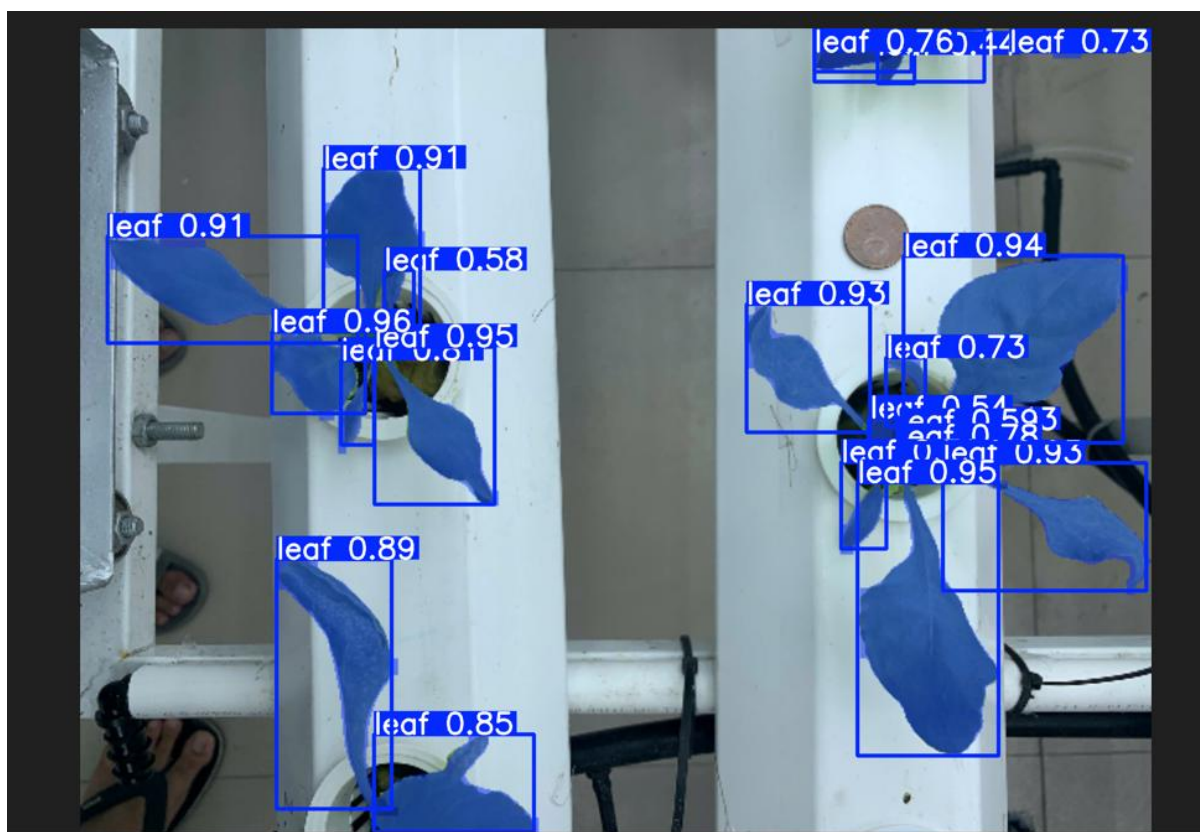
- Dự đoán đúng “leaf” (TP): 693 mẫu.
- Dự đoán nhầm “background” là “leaf” (FP): 217 mẫu.
- Dự đoán nhầm “leaf” là “background” (FN): 117 mẫu.

Ma trận chuẩn hóa:

- Precision cho lớp "leaf": 86%
- Recall cho lớp "leaf": 86%
- Lớp "background": độ chính xác 100%, nhầm lẫn chủ yếu đến từ vùng biên.

⇒ Nhận xét:

- Mô hình phân biệt tốt giữa lá và nền, song cần cải thiện thêm với các vùng biên hoặc lá mờ, chồng lên nhau.
- Việc dự đoán sai vùng “background” là “leaf” phần lớn đến từ nhiều nền hoặc viền mép không rõ nét.



Hình 3 Kết quả test sau khi train mô hình

Kết quả test cho thấy mô hình YOLOv11 nhận diện và phân đoạn lá cải bó xôi rất chính xác. Tất cả các lá đều được phát hiện đầy đủ với bounding box rõ ràng và confidence cao (đa số  $> 0.9$ ). Các mask phân đoạn (vùng màu xanh) bám sát hình dạng thực tế của lá, kể cả với lá cong hoặc chồng nhau.

Mô hình vẫn giữ độ chính xác tốt với những lá nhỏ hoặc che khuất (vẫn được phát hiện với confidence  $\sim 0.58$ ). Tuy nhiên, một vài vùng có box trùng nhẹ, có thể tinh chỉnh thêm ngưỡng confidence hoặc thuật toán NMS để loại bỏ dự đoán dư thừa.

Tổng thể, mô hình hoạt động hiệu quả, đủ đáp ứng yêu cầu trong ứng dụng phân đoạn lá cải bó xôi tự động. Nhưng do giới hạn về dữ liệu xuất ra không có thông tin về group hoặc id cây đã được xác định trong quá trình gán nhãn. Kết quả cuối hoàn thành về bài toán phân đoạn nhưng chưa thể phân loại riêng các lá thuộc các cây khác nhau.