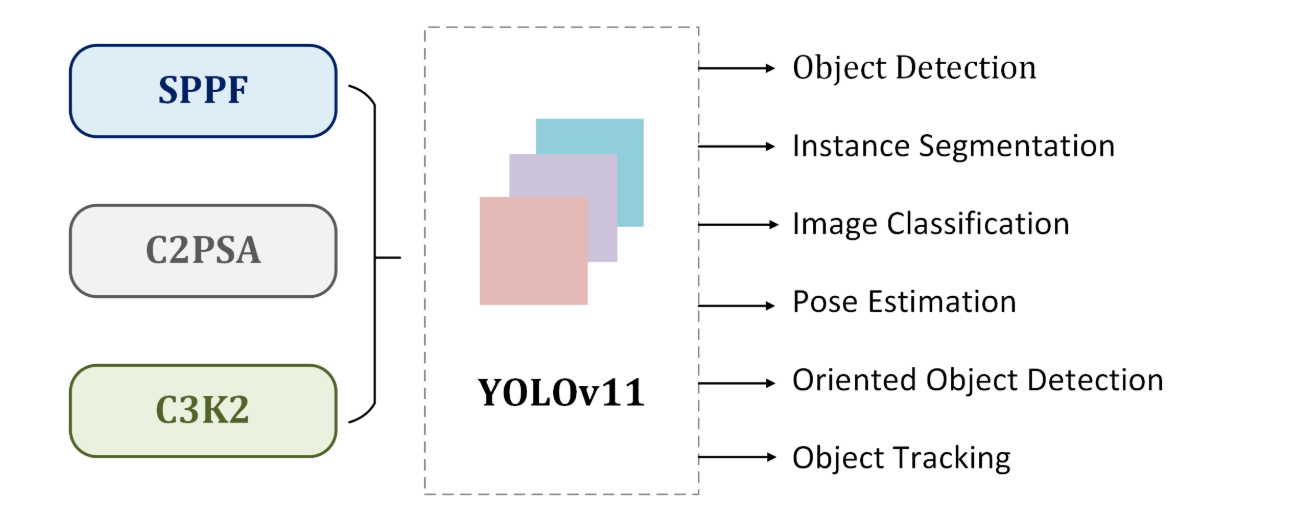
**YoLov11**

Cải thiện 3 model là: SPPF, C2PSA, C3K2

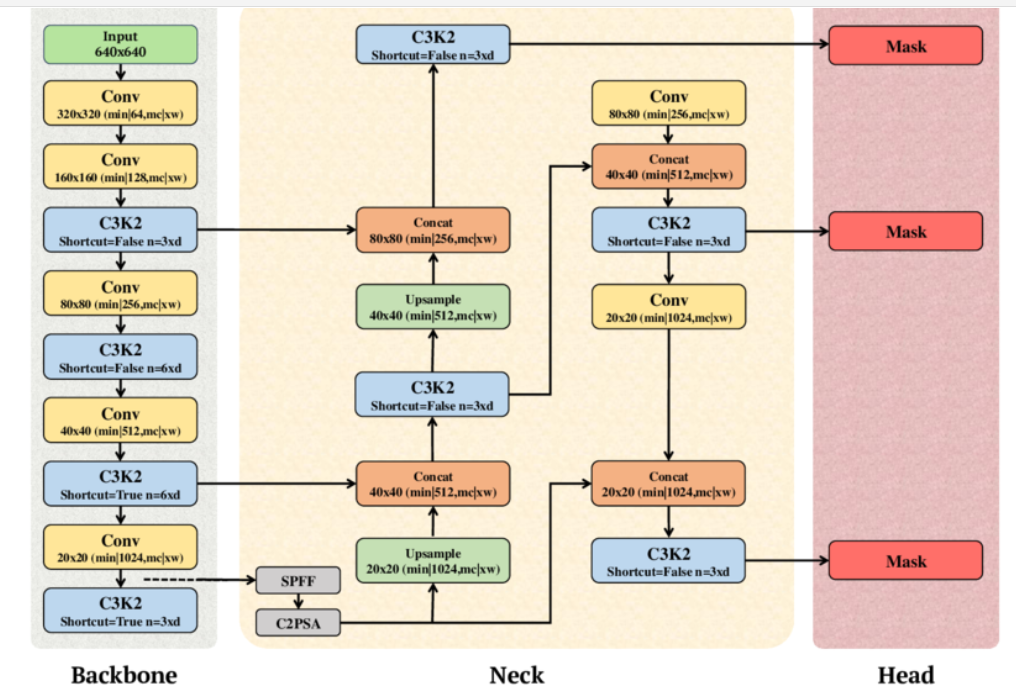


* Yolov11 cải tiến thêm phát hiện đối tượng hơn các mô hình truyền hình thống trước như khả năng ước lượng tư thế và nó có thể dùng cho việc mà phân đoạn ảnh nghĩa segmentation nó giúp mở rộng các ứng dụng trong nhiều lĩnh vực thực tiễn nhiều ngành nghề khác nhau
* SPPF: là được cải thiện từ yolov5 để tối ưu trong yolov11
* C2PSA: là giúp YoLov11 tăng cường cơ chế chú ý không gian (spatial attention) trong các bản đồ đặc trưng.
* C3K2: được để nhanh hơn và hiệu quả hơn, giúp nâng cao hiệu suất tổng hợp đặc trưng.

Ưu điểm:

* Tăng độ chính xác, tốc độ nhanh hơn, có số lượng tham số ít hơn có khả năng trích xuất đặc trưng được cải thiện ngoài ra nó có khả năng thích nghi các bối cảnh khác nhau và hỗ trợ nhiều trong nhiều nhiệm vụ.

Yolov11 có 3 phần: Backbone, Neek, Head

****

* Backbone: xử lý đầu vào và trích xuất đặc trưng quan trọng thông qua các lớp và các khối C3K2

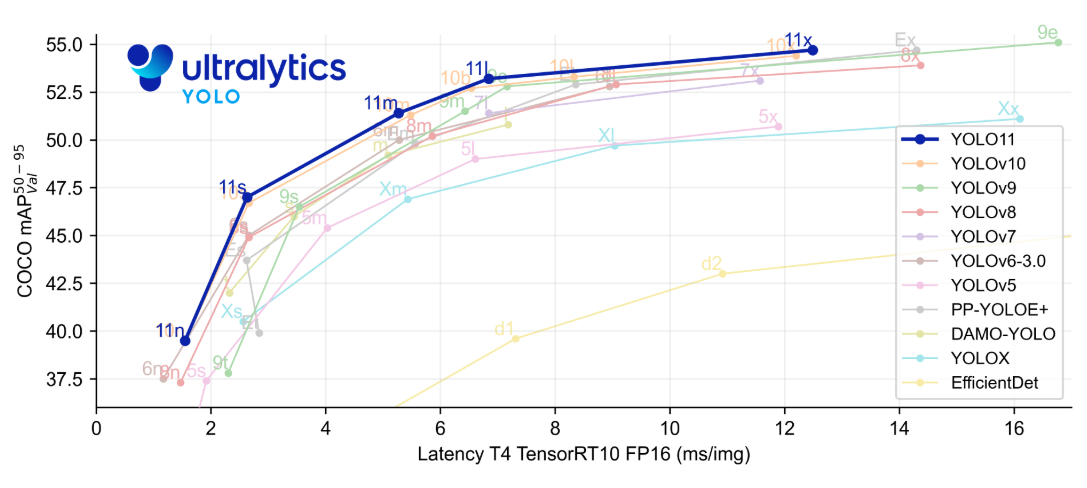
*Cấu trúc Backbone:*

* Conv Layers: Lần lượt giảm kích thước xuống, các lớp convolution này giúp lọc và trích xuất đặc trưng sơ cấp.
* C3K2 Blocks: là cải tiến của C2f, giúp tổng hợp đặc trưng hiệu quả hơn.
* SPPF: Kết hợp thông tin từ nhiều vùng không gian khác nhau, giúp mô hình học đặc trưng tốt hơn. Cải thiện khả năng phát hiện đối tượng với kích thước đa dạng.
* C2PSA:Một cơ chế chú ý không gian (Spatial Attention) giúp mô hình tập trung vào vùng quan trọng của ảnh.

*Cấu trúc Neck:* Giúp mô hình phát hiện đối tượng với nhiều kích thước khác nhau. Sử dụng các kỹ thuật upsampling, concatenation, C3K2 blocks tối ưu nhất.

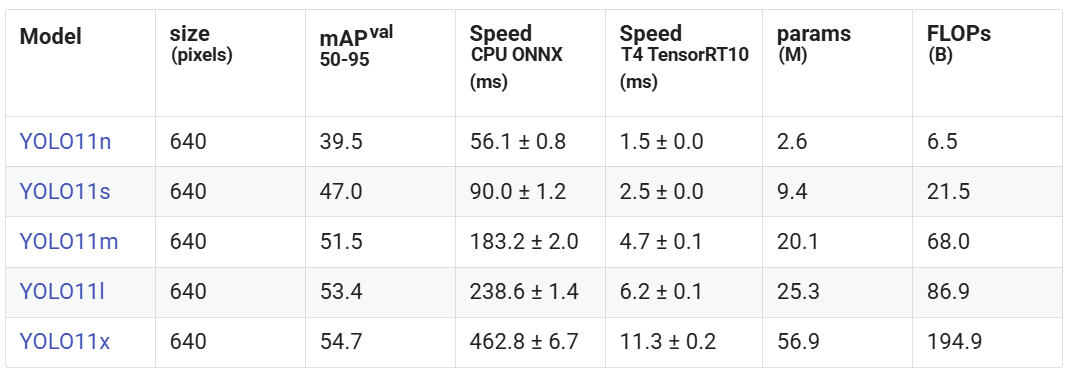
* Upsampling: Giúp truyền tải đặc trưng từ tầng sâu lên các tầng trên.
* Concat: Kết hợp thông tin từ các tầng khác nhau để tận dụng cả đặc trưng chi tiết và đặc trưng trừu tượng.
* C3K2 Block: Tăng cường khả năng tổng hợp đặc trưng từ nhiều nguồn dữ liệu, để cải thiện nhận diện.

Cấu trúc head*:* Kết Hợp thông tin các đặc trưng để dự đoán, điều chỉnh thông tin đầu ra trước khi tạo Mask, cuối cùng là cho là ra kết quả cho từng đối tượng trong ảnh và hỗ trợ phân đoạn ảnh.



* COOP AP (%): đo độ chính xác: càng cao thì mô hình càng chính xác.
* Latency (ms): thời gian xử lý: càng thấp thì mô hình càng nhanh.
* YOLOv11 là mô hình có hiệu suất tốt nhất trong tất cả các phiên bản YOLO tính đến hiện tại. Nó vượt trội cả về độ chính xác (mAP) lẫn khả năng mở rộng giữa các phiên bản nhẹ và mạnh.

Các phiên bản của yolov11:



N, S: Phù hợp với thiết bị nhúng, điện thoại, camera real-time, tốc độ rất nhanh, tiêu tốn ít tài nguyên.

M, B: Cân bằng giữa tốc độ & độ chính xác, phù hợp cho các ứng dụng thời gian thực.

L, X: Dùng cho cloud/server, phù hợp với các bài toán AI nặng, đòi hỏi độ chính xác cao.

**Độ chính xác:** Độ chính xác cao hơn với ít tham số hơn: Với những cải tiến trong thiết kế mô hình, YOLO11m đạt được chỉ số mAP (mean Average Precision) cao hơn trên bộ dữ liệu COCO trong khi sử dụng ít hơn 22% tham số so với YOLOv8m, giúp mô hình hiệu quả về mặt tính toán mà không làm giảm độ chính xác.

**FBS**:

* **1000 FPS (YOLOv11n):** Benchmark tối ưu (A100, TensorRT, không hiển thị ảnh)
* **140 FPS (YOLOv11x)**:Sử dụng trên máy tính cá nhân hoặc khi không dùng TensorRT

**Ảnh Hưởng:**

* Trong yolov11 vẫn chưa được cải thiện về Trường hợp sương mù không nhận biết và phân biệt rõ nhưng cũng đã cải thiện hơn là yolov8.
* Những trường hợp mất điện và ngược sáng thì khó bắt chi tiết được vật.