

OPEN WORLD INSTANCE SEGMENTATION USING PROBABILISTIC OBJECTNESS ON INSTANCE ACTIVATION MAP

Trần Hoàng Bảo Ly^{1,2, 3}

¹ Trường Đại học Công nghệ Thông tin

² Đại học Quốc gia Thành phố Hồ Chí Minh

³ 21521109@gm.uit.edu.vn

What ?

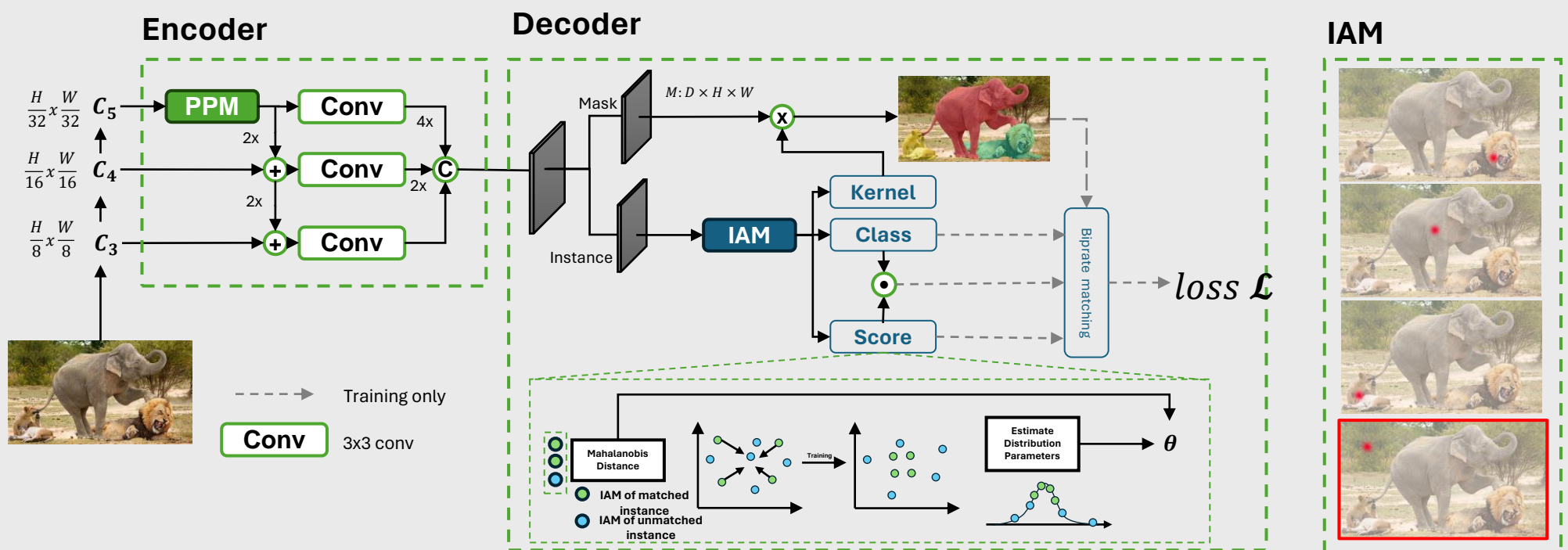
Chúng tôi giới thiệu phương pháp để giải quyết và đánh giá bài toán open world instance segmentation (OWIS).

- Đề xuất mô hình end-to-end open world instance segmentation.
- Xây dựng tập dữ liệu mới để đánh giá mô hình open world instance segmentation.
- Huấn luyện, đánh giá và so sánh mô hình trên các tập dữ liệu khác nhau.

Why ?

- Open world instance segmentation (OWIS) giúp giảm sự phụ thuộc vào số lượng hạn chế của các tập huấn luyện, giúp mô hình ứng dụng rộng rãi hơn trong thực tế từ xe tự hành, tự động hóa, y tế, giáo dục,... OWIS gồm 2 mục tiêu: 1) phát hiện lớp thực thể không có trong tập huấn luyện, 2) Học tăng cường các lớp thực thể mới.
- Các nghiên cứu hiện nay chỉ phát hiện thực thể, không quan tâm đến lớp, hoặc phụ thuộc vào các mô hình open world object detection.

Overview



Description

1. Encoder

Chúng tôi xây dựng encoder dựa trên ý tưởng của Feature Pyramid Network (FPN), kết hợp với Pyramid Pooling module (PPM) để trích xuất được các đặc trưng ở các scale khác nhau, đầu ra của encoder $\frac{1}{8}$ x độ phân giải của ảnh đầu vào.

2. Decoder

Decoder dựa trên instance activation map (IAM) bao gồm hai nhánh:

- Nhánh instance: thực hiện nhiệm vụ đưa ra các IAM của các thực thể, cũng như các đặc trưng của thực thể cho việc phân đoạn và phân loại.
- Nhánh mask: để encoder các đặc trưng mask của thực thể. Các mask dùng để phân đoạn cho mỗi thực thể có thể thu được bằng cách nhân nhánh mask với kernel tương ứng ở nhánh instance.

Nhánh instance còn bao gồm hai đầu ra là score và class.

- Đầu ra score được đổi thành module xác suất về đối tượng. Phân phối xác suất cho tập IAM ký hiệu là θ , ban đầu được ước lượng thông qua đường trung bình động hàm mũ, và huấn luyện thông qua hàm loss dựa trên khoảng cách Mahalanobis. Với thiết kế trên, ta có thể tách rời việc học thực thể và lớp thực thể. Cụ thể xác suất IAM_t thuộc thực thể lớp l sẽ bằng xác suất IAM_t là thực thể, nhân với xác suất IAM_t thuộc lớp l .

$$p(l|q) = f_{cls}^t(q) \cdot f_{obj}^t(q)$$

- Đầu ra class dùng để dự đoán lớp của đối tượng được huấn bằng hàm sigmoid focal loss.

3. IAM

Tập IAM sẽ được sinh ra bởi module IAM tương ứng với N dự đoán, và với K ground-truth thực thể, ta có thể huấn luyện module IAM bằng thuộc toán Hungarian matching.

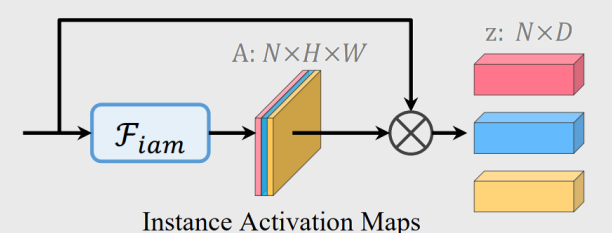


Figure 1. IAM module.

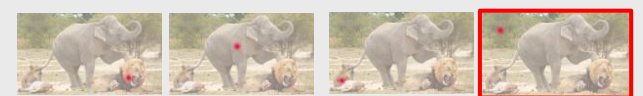


Figure 2. Tập IAM dự đoán. (IAM khoanh đỏ tương ứng cho IAM dự đoán sai, tức thuộc về background.)

4. Tập dữ liệu đánh giá mô hình OWIS

Dựa trên nguồn tổng hợp 21K ảnh từ tập Pascal VOC và hơn 130K ảnh từ tập MS-COCO, với 30 lớp thực thể, bao gồm toàn bộ 20 lớp từ tập Pascal VOC, Outdoor, Accessories, Appliance, Truck, Sports, Food, Electronic, Indoor, Kitchen, Furniture.