

CẢI THIẾN CHUYÊN SÂU HỌ R-CNN TRONG BÀI TOÁN PHÁT HIỆN ĐỐI TƯỢNG DỰA TRÊN TÀI LIỆU TIẾNG VIỆT DẠNG ẢNH

Lê Thị Phương Vy^{1,4} Lê Trương Ngọc Hải^{2,4} và Nguyễn Nhật Trường^{3,4}

{¹20520355, ²20520481, ³20522087}@gm.uit.edu.vn ⁴Trường Đại học Công nghệ Thông tin ĐHQG TP.HCM

What ?

- Phát hiện đối tượng** là quá trình xác định vị trí và định danh các đối tượng trong hình ảnh.
- Phát hiện đối tượng trong văn bản dạng ảnh** là quá trình xác định vị trí và định danh các đối tượng từ trong hình ảnh có chứa văn bản.

Why ?

- Kết quả của những phương pháp trước đó đã thực nghiệm trên bộ dữ liệu UIT-DODV chưa được tối ưu hoàn toàn.
- Điều kiện rất phù hợp lượng tài nguyên mà nhóm đang có.
- Phát hiện đối tượng trong văn bản dạng ảnh** là một bài toán mới bắt đầu được nghiên cứu kể từ năm 2010.
- Bài toán này vẫn còn nhiều **thách thức** vì cách trình bày các đối tượng ngữ nghĩa trong tiếng Việt tạo ra nhiều khó khăn trong việc rút trích đặc trưng các thông tin.
- Hơn nữa, các đặc trưng rất nhỏ trong văn và khó phân biệt gây ảnh hưởng tới **hiệu quả** của mô hình
- Vì thế, việc làm sao để **kết hợp** mô-đun mới vào mô hình nhằm tăng độ chính xác của RCNN một cách hiệu quả là rất quan trọng.

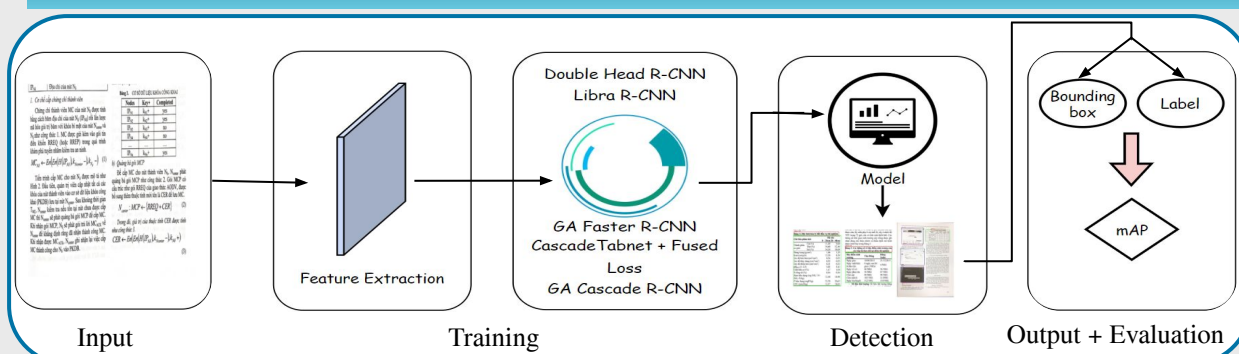
Hypothesis

"Liệu có cách cải tiến hiệu suất phát hiện đối tượng trên trang tài liệu dạng ảnh của họ R-CNN?"

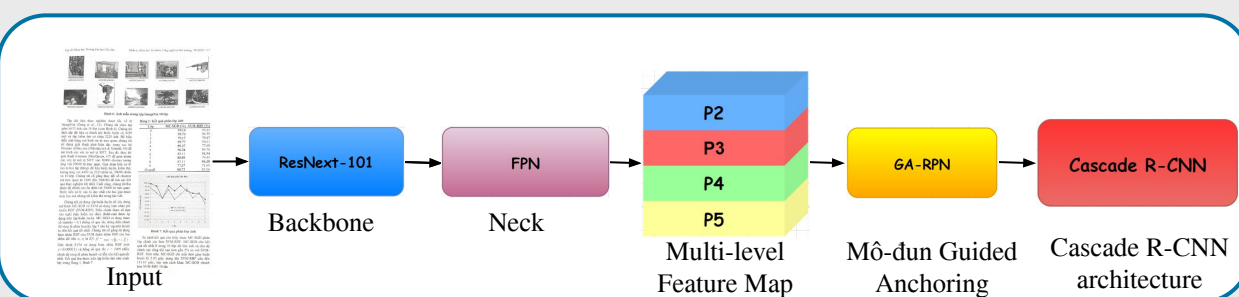
How ?

- Khảo sát các hướng tiếp cận gần đây cho bài toán phát hiện đối tượng trên tài liệu dạng ảnh.
- Khảo sát các phương pháp cải tiến hiệu suất R-CNN.
- Đề xuất kết hợp Cascade R-CNN và Guided Anchoring.
- Nếu phương pháp đạt kết quả tốt, chúng tôi đặt tên là Guided Anchoring Cascade R-CNN.

Description



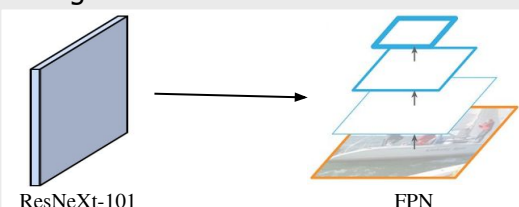
Hình 1. Mô tả quy trình thực nghiệm cho bài toán phát hiện đối tượng trang



Hình 2. Mô tả kiến trúc Guided Anchoring Cascade RCNN

1. Trích xuất đặc trưng:

- ResNeXt-101 làm backbone cho mô hình
- Sau đó FPN được sử dụng để trích xuất các đặc trưng thể hiện dưới các bộ feature map với các kích thước tương ứng với các tầng đã được xây dựng trong kiến trúc từ trước

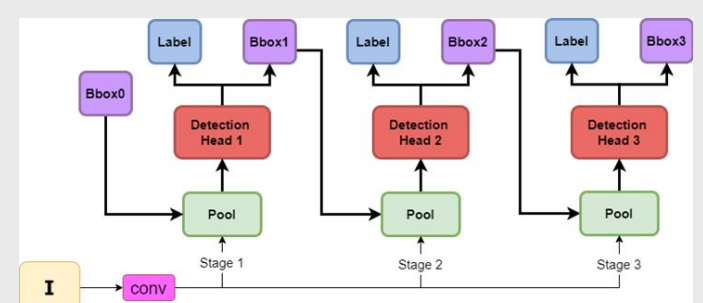


2. Mô-đun Guided Anchoring:

- Guided Anchoring có baseline là **Region Proposal Network (RPN)** cải thiện quá trình tạo ra các region proposal hiệu quả và linh hoạt nhờ vào việc dự đoán vị trí một điểm trung tâm của vật thể mà có khả năng tồn tại cũng như các scales và tỷ lệ khung hình tại các vị trí khác nhau.
- Ngoài ra **Guided Anchoring** loại bỏ cơ chế khởi tạo các hộp neo (anchor box) với các thông số mặc định, kích thước và tỉ lệ của một anchor có thể thay đổi linh hoạt giúp đa dạng quá trình học và phát hiện vật thể có kích thước đặc biệt.

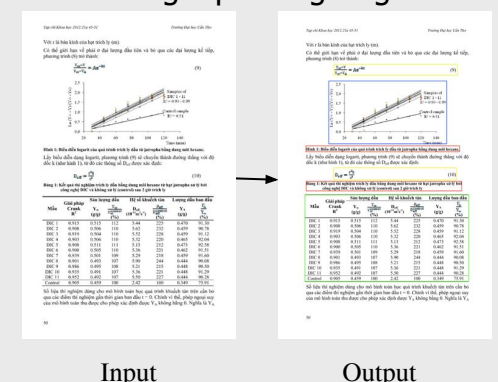
3. Mô hình Cascade RCNN

- Phần mở rộng **nhiều giai đoạn** của R-CNN.
- Các giai đoạn của máy dò được chọn lọc **tuần tự** hơn. Qua đó đã cải thiện được những hạn chế của Faster R-CNN



Mô tả bài toán:

- Input:** tập hợp các ảnh dạng tài liệu tiếng Việt.
- Output:** vị trí của các đối tượng trong ảnh được thể hiện bằng các hộp giới hạn tối thiểu với các màu sắc khác nhau cho từng lớp tương ứng.



Input

Output