

GROUP 2 DIGITAL IMAGE PROCESSING ABSTRACT REPORT

Ngày 29 tháng 5 năm 2024

1 Summary of approaches

1.1 Tìm thuật toán sử dụng cho bài toán detect bounding box

- **Thành công:** config được thành công mô hình YOLO từ thư viện Ultralytics và tận dụng tối đa các mô hình thuộc họ YOLO này. Dựa vào forum của Ultralytics trên Github, nhóm đã đọc và biết được rằng 2 mô hình YOLOv5 và YOLOv8 là 2 mô hình mạnh nhất trong họ YOLO nên đã quyết định thử nghiệm và tune các tham số quanh 2 mô hình này.
- **Thất bại:**
 - Các mô hình deep learning: so với proposal report, nhóm đã không config được theo thư viện detectron2 để thử nghiệm các kiến trúc đã đề cập trong modelzoo của detectron2. Trong quá trình tìm kiếm, nhóm cũng đã thử nghiệm với Resnet config cho bài toán bounding box nhưng rồi nhận ra là Resnet không phải kiến trúc SOTA cho bài toán này
 - Các mô hình không phải deep learning: Vì nhóm khởi điểm ban đầu hầu hết các thành viên chưa có kiến thức deep learning, nên nhóm đã đi tìm những thuật toán cơ bản không phải mô hình học sâu để nhóm về implement lại nhưng kết quả không khả quan lắm.

1.2 Phân tích bộ dữ liệu

Những phân tích được thực hiện trên bộ dữ liệu:

- **Phân tích có ích:** bức ảnh lệch hay không: một bức ảnh lệch là ảnh có thể có số chữ trang bên trái nhiều hơn hẳn so với số chữ trang bên phải, hoặc số chữ nửa trên bức ảnh nhiều hơn hẳn nửa dưới (nói chung là 2 bên không đều nhau). Nhóm đã phân tích và tìm ra những bức ảnh như vậy bởi vì trong quá trình train, trong 1 lần train với số epoch thấp (tầm 30) với mô hình YOLOv8, nhóm đã nhận ra là lúc đó mô hình có xu hướng làm tệ trên các bức ảnh bị "lệch" nên đã tìm những bức ảnh đó và tập trung augment vào chúng.
- **Phân tích không có ích:** nhóm đã phân tích các tham số x, y, w, h của các bounding box trong tập train và val, cũng như so sánh distribution giữa các tham số này của 2 tập và nhận ra không có gì bất thường cả. Chi tiết hơn về các loại phân tích được đề cập trong full report của nhóm.

1.3 Sinh thêm dữ liệu

Dựa trên bộ dataset NomNaOCR được public trước đó, nhóm đã đi tìm hiểu các tools như Tesseact, PaddleOCR nhưng không thành công vì không config được hoặc tool đó chỉ dành cho dải chữ.

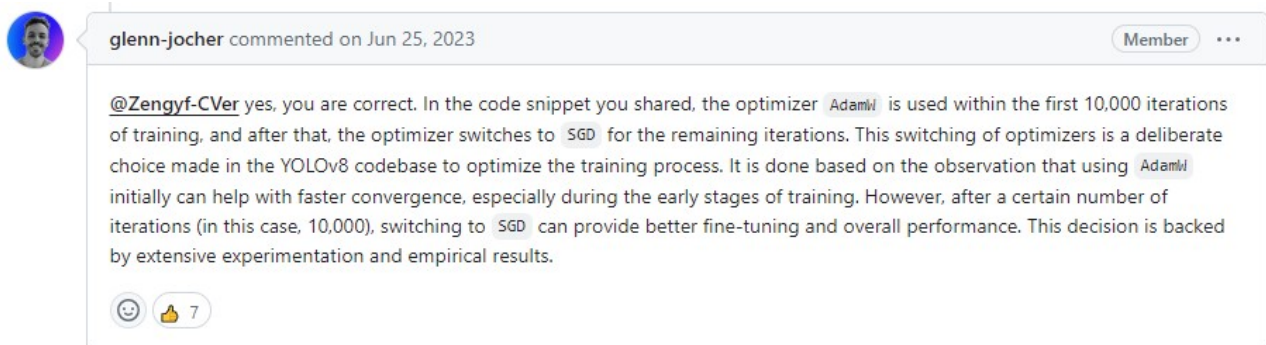
1.4 Tăng cường bộ dữ liệu

Nhóm dùng 4 cách augment: scale, rotate, dịch ảnh và thêm nhiễu. Chi tiết các nhiễu được đề cập trong full report của nhóm. Sau khi có data augment thì nhóm đã chia nhau ra tự tạo bộ dữ liệu mới của riêng mình và train thêm trên bộ dữ liệu đó.

1.5 Fine tune tham số

Sau khi đọc và tìm hiểu thì nhóm quyết định chia nhau ra thử nghiệm và thay đổi các siêu tham số sau so với các siêu tham số mặc định ban đầu:

- Batch size: 3 - 5 - 8 - 16 với mục đích là với số batch to thì để xem model train có ổn định và đỡ stochastic hơn không.
- Imgsz: 640 - 640, 1280x1280 vì ở phần phân tích dữ liệu, nhóm cũng có phân tích xem có các ảnh có kích cỡ ban đầu là bao nhiêu pixel. Và kết quả là có 3 nhóm kích cỡ chính dựa theo tên của các ảnh (Ví dụ tên ảnh là "nlvnpf-0137-01-001" thì ảnh thuộc nhóm kích cỡ "0137"). Nếu resize 640 x 640 pixel thì ảnh sẽ nhỏ hơn nên nhóm cũng quyết định thử cả size 1280 x 1280 nữa vì nhóm nghĩ việc resize như vậy cũng sẽ làm cho model có performance tốt hơn trên các ảnh lớn hơn và data được augment cũng có dữ liệu được scale random.
- Optimizer: auto, SGD, AdamW. Khi nhóm research về các siêu tham số, nhóm đã đọc được bình luận bên dưới của tác giả framework Ultralytics



Hình 1: Nhận xét của tác giả framework Ultralytics về các optimizer

- Ngoài ra nhóm còn thử nghiệm với các version khác nhau của YOLOv5, YOLOv8 nữa khi đọc về performance của các version small, medium, big,...

1.6 Semi-supervised

Tại đầu tuần này, nhóm đã lấy mô hình có số điểm tốt nhất của nhóm là 88.9% trên tập val gốc ban đầu và đem dự đoán với 52 ảnh random từ tập dữ liệu NomNaOCR, tuy nhiên chỉ có 3 ảnh trong số đó được dự đoán đúng gần như hoàn toàn nên nhóm đã pick 3 ảnh đó cho vào tập dữ liệu mới để học. Tuy nhiên mô hình cũng không tăng mAP lên nhiều cho lắm.

2 Description of tasks

Tăng Vĩnh Hà, MSV: 22028129. Self - evaluation: 25%. Những việc đã làm: Đề ra pipeline cho nhóm và lead, Tổ chức meeting hàng tuần để kiểm tra tiến độ và hỗ trợ mọi người, Hướng dẫn mọi người làm quen với deep learning, Đi research toàn bộ 1.1 và config YOLO, hướng dẫn mọi người chạy YOLO, Chuẩn bị các hàm cơ bản cho việc train, viết script cuối cùng để test, Đi research về siêu tham số để thiết kế các bộ siêu tham số cho mọi người thử nghiệm, Viết abstract report, full report, Thử nghiệm và thiết kế các bộ dữ liệu mới với các ảnh semi - supervised để mọi người train thêm. Thử nghiệm config detectron2 phần 1.1 và sinh thêm bộ dữ liệu phần 1.3

Lê Thị Hải Anh, MSV:22028162. Self-evaluation: 25%. Những việc đã làm: Đi research về các thư viện sử dụng cho việc augment data, Làm 70% các phương pháp tăng cường bộ dữ liệu, Train model và thử nghiệm với 2 bộ tham số để so sánh, Thiết kế các bộ dữ liệu mới để tự mình train thêm, Viết full report

Vũ Nguyệt Hằng, MSV:22028079. Self-evaluation: 25%. Những việc đã làm: Viết proposal report, full report, Làm 30 % các phương pháp tăng cường bộ dữ liệu, Train model và thử nghiệm với 2 bộ tham số để so sánh, Thiết kế các bộ dữ liệu mới để tự mình train thêm, Làm slides, Thử nghiệm các tool và tìm cách cho phần 1.3

sinh thêm dữ liệu tuy thất bại

Lê Xuân Hùng, MSV:22028172. Self-evaluation: 25%. Những việc đã làm: Di research các cách phân tích dữ liệu với dạng ảnh, Phân tích dữ liệu và đưa ra kết luận cho các chiến lược augment của nhóm, Train model và thử nghiệm với 2 bộ tham số để so sánh, Thiết kế các bộ dữ liệu mới để tự mình train thêm, Viết full report, Thử nghiệm config mô hình detectron2 cho phần 1.1