

BÁO CÁO ĐỒ ÁN CUỐI KỲ

Môn học

**CS519 - PHƯƠNG PHÁP LUẬN
NGHIÊN CỨU KHOA HỌC**

Lớp học

CS519.O11

Giảng viên

PGS.TS. LÊ ĐÌNH DUY



Thời gian

09/2023 - 02/2024

----- Trang này cố tình để trống -----

THÔNG TIN CHUNG CỦA NHÓM

- Link YouTube video của báo cáo (tối đa 5 phút):
<https://www.youtube.com/watch?v=oUhNg9YhaqA>
- Link slides (dạng .pdf đặt trên Github của nhóm):
<https://github.com/Sunmapp/CS519.O11/blob/main/MoHinhFasterRCNNPhatHienDoiTuongThoiGianThucVoiMangDeXuatKhuVuc.pdf>
- Mỗi thành viên của nhóm điền thông tin vào một dòng theo mẫu bên dưới
- Sau đó điền vào Đề cương nghiên cứu (tối đa 5 trang), rồi chọn Turn in

| | |
|--|---|
| <ul style="list-style-type: none"> • Họ và Tên: Vương Vĩnh Thuận • MSSV: 20521997  | <ul style="list-style-type: none"> • Lớp: CS519.O11 • Tự đánh giá (điểm tổng kết môn): 8/10 • Số buổi vắng: 2 • Số câu hỏi QT cá nhân: 13 • Số câu hỏi QT của cả nhóm: 0 • Link Github: https://github.com/Sunmapp • Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm: <ul style="list-style-type: none"> ○ Làm slide ○ Tìm hiểu mô hình Fast R-CNN, RPN ○ Viết báo cáo phần nội dung |
| <ul style="list-style-type: none"> • Họ và Tên: Nguyễn Phương Duy • MSSV: 20521242  | <ul style="list-style-type: none"> • Lớp: CS519.O11 • Tự đánh giá (điểm tổng kết môn): 8/10 • Số buổi vắng: 3 • Số câu hỏi QT cá nhân: 3 • Số câu hỏi QT của cả nhóm: 12 • Link Github: https://github.com/duyduy2808 • Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm: <ul style="list-style-type: none"> ○ Làm Poster, nội dung |

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI

Mô hình Faster R-CNN: Phát hiện đối tượng thời gian thực với mạng đề xuất khu vực

TÊN ĐỀ TÀI TIẾNG ANH

Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks

TÓM TẮT

Những tiến bộ gần đây trong việc phát hiện đối tượng được thúc đẩy bởi sự thành công của các phương pháp đề xuất khu vực. Mặc dù CNN dựa trên khu vực ban đầu đáng kể nhờ chia sẻ các tích chập trên các đề xuất. Phiên bản mới nhất, Fast R-CNN, đạt được tốc độ gần thời gian thực bằng cách sử dụng mạng học sâu, khi bỏ qua thời gian tiêu tốn cho việc đề xuất khu vực. Hiện nay, việc đề xuất là rào cản tính toán ở thời gian kiểm tra trong các hệ thống phát hiện tiên tiến.

Các phương pháp đề xuất khu vực thường phụ thuộc vào các đặc trưng giá rẻ và các kế hoạch suy luận kinh tế. Selective Search, một trong những phương pháp phổ biến nhất, hợp nhất tham lam các siêu pixel dựa trên các đặc trưng cấp thấp được thiết kế. Tuy nhiên, so với các mạng phát hiện hiệu quả, Selective Search diễn ra chậm, mất 2 giây cho mỗi hình ảnh trong một triển khai CPU. EdgeBoxes hiện cung cấp sự cân đối tốt nhất giữa chất lượng và tốc độ đề xuất, mất 0.2 giây cho mỗi hình ảnh. Tuy nhiên, bước đề xuất khu vực vẫn tiêu tốn thời gian chạy như mạng phát hiện.

Do đó, trong bài báo này, chúng tôi chỉ ra rằng một thay đổi thuật toán - tính toán các đề xuất với một mạng nơ-ron tích chập sâu - dẫn đến một giải pháp hiệu quả, trong đó tính toán đề xuất gần như miễn phí so với tính toán của mạng phát hiện. Với mục tiêu này, chúng tôi giới thiệu một mạng học sâu mới bằng cách kết hợp mạng học sâu Fast R-CNN với RPN. Đầu được phát triển có chi phí tính toán cao, nhưng chi phí này đã được giảm

GIỚI THIỆU

Các mạng phát hiện đối tượng tiên tiến hiện nay phụ thuộc vào các thuật toán đề xuất khu vực để giả định vị trí của đối tượng. Những tiến bộ như SPPnet và Fast R-CNN đã làm giảm thời gian chạy của các mạng phát hiện này, làm nổi bật việc tính toán đề xuất khu vực như là một rào cản. Trong công việc này, chúng tôi giới thiệu Mạng Đề Xuất

Khu Vực (RPN) mà chia sẻ đầy đủ các tính năng tích chập của toàn bộ hình ảnh với mạng phát hiện, từ đó cho phép đề xuất khu vực gần như miễn phí. Một RPN là một mạng nơ-ron tích chập hoàn toàn, đồng thời dự đoán ranh giới của đối tượng và điểm số đối tượng tại mỗi vị trí. RPN được huấn luyện end-to-end để tạo ra các đề xuất khu vực chất lượng cao, được sử dụng bởi Fast R-CNN cho việc phát hiện. Chúng tôi tiếp tục hợp nhất RPN và Fast R-CNN thành một mạng duy nhất bằng cách chia sẻ các tính năng tích chập của họ—sử dụng thuật ngữ gần đây phổ biến trong mạng nơ-ron với cơ chế "chú ý", thành phần RPN cho biết mạng thống nhất nơi để tìm kiếm tạo nên mô hình Faster R-CNN.

MỤC TIÊU

- Nghiên cứu mạng học sâu CNN và thuật toán RPN hiện có và cải thiện hiệu suất của nó trong bài toán nhằm tạo ra phiên bản tích hợp của 2 phương pháp trên tạo ra mô hình Faster R-CNN.
- Thử nghiệm mô hình Faster R-CNN trên bộ dữ liệu aircraft tạo ra mô hình phân loại máy bay quân sự.

NỘI DUNG

Faster R-CNN

Hệ thống phát hiện đối tượng của chúng tôi, được gọi là Faster R-CNN, bao gồm hai mô-đun. Mô-đun đầu tiên là một mạng tích chập sâu hoàn toàn đề xuất các vùng và mô-đun thứ hai là bộ phát hiện Fast R-CNN sử dụng các vùng được đề xuất. Toàn bộ hệ thống là một mạng thống nhất để phát hiện đối tượng.

Sharing Features for RPN and Fast R-CNN

Tiếp theo, chúng tôi sẽ mô tả các thuật toán học một mạng hợp nhất bao gồm RPN và Fast R-CNN được chia sẻ với các lớp tích chập.

Cả mạng RPN và Fast R-CNN đều được huấn luyện riêng biệt sẽ thay đổi các lớp chập theo nhiều cách khác nhau. Vì thế, chúng tôi phát triển một kỹ thuật cho phép chia sẻ các lớp chập giữa mạng. Chúng tôi tìm được 3 cách để huấn luyện mạng với các đặc trưng:

(i): Alternating training. Chúng tôi sẽ huấn luyện mạng RPN trước, sử dụng các vùng đề xuất để huấn luyện thuật toán Fast R-CNN. Sau đó, lớp mạng được điều chỉnh bởi Fast R-CNN sẽ được dùng để khởi tạo mạng RPN và quá trình này được lặp đi lặp lại. Đây là biện pháp được dùng cho mọi thực nghiệm trong đề xuất này.

(ii): Approximate joint training. Các mạng RPN và Fast R-CNN được hợp nhất lại thành một mạng trong quá trình huấn luyện. Trong mỗi lần lặp SGD, cơ chế chuyển tiếp (forward pass) tạo sinh ra đề xuất vùng ở những điểm cố định, tính toán trước khi

huấn luyện bằng thuật toán Fast R-CNN. Cơ chế lan truyền ngược (backward propagation) xảy ra ở các lớp có dấu hiệu của hàm mất mát RPN và Fast R-CNN

(iii) Non-approximate joint training. Các khung bao đối tượng (bounding box) được dự đoán bằng mạng RPN sẽ là đầu vào cho lớp RoI pooling trong thuật toán Fast R-CNN. Chúng tôi cần một lớp RoI pooling có thể phân biệt được tọa độ của khung bao đối tượng.

Experiment-evaluation

Thực hiện một đánh giá toàn diện về phương pháp của chúng tôi trên bộ kiểm thử phát hiện PASCAL VOC 2007. Bộ dữ liệu này bao gồm khoảng 5.000 ảnh trainval và 5.000 ảnh thử nghiệm trên 20 loại đối tượng. Chúng tôi cũng cung cấp kết quả trên bộ kiểm thử PASCAL VOC 2012 cho mô hình.

Train in Aircraft dataset

- Sau khi thử nghiệm và hoàn thiện mô hình Faster R-CNN, chúng tôi thực hiện một mô hình phân loại các loại máy bay chiến đấu dựa trên bộ dữ liệu **Aircraft dataset**

KẾT QUẢ MONG ĐỢI

- Báo cáo mô hình Faster R-CNN mà nhóm phát triển được sử dụng trong bài toán phát hiện đối tượng.
- Kết quả thực nghiệm, đánh giá, so sánh các phương pháp với nhau và với mô hình Transformer truyền thống.
- Chương trình phát hiện máy bay quân sự với tốc độ thực tế.

KẾ HOẠCH THỰC HIỆN

- ❖ Tuần 1 – 4: Tìm hiểu các mô hình CNN đã có sẵn trước đó và kỹ thuật RPN.
 - Kết quả dự kiến:
 - Tài liệu chi tiết cấu trúc cơ bản các mô hình CNN.
 - .Tìm hiểu các mạng CNN đã có trước đó
 - Tài liệu chi tiết kỹ thuật RPN
- ❖ Tuần 5 – 8: Kết hợp mô hình Fast R-CNN với RPN tạo ra một mô hình mới gọi là Faster R-CNN
 - Kết quả dự kiến:
 - Thử nghiệm các mô hình CNN trước đó trên bộ dữ liệu **PASCAL VOC 2007**.
 - Kết hợp mô hình Fast R-CNN với RPN.
- ❖ Tuần 9 – 12: Huấn luyện mô hình Faster R-CNN trên tập dữ liệu Aircraft tạo ra mô hình phát hiện đối tượng máy bay quân sự.

➤ Kết quả dự kiến:

- Thu thập dữ liệu các loại máy bay chiến đấu tạo nên bộ dữ liệu Aircraft.
- Thử nghiệm trên bộ dữ liệu Aircraft

❖ Tuần 13 – 16: Xây dựng chương trình demo.

➤ Kết quả dự kiến:

- Chương trình minh họa.

❖ Phân công nhiệm vụ:

| Sinh viên thực hiện | Nhiệm vụ |
|---------------------|---|
| Vương Vĩnh Thuận | <ul style="list-style-type: none">● Tìm hiểu mô hình Fast R-CNN, RPN● Viết nội dung phần nội dung phương pháp● Làm slide |
| Nguyễn Phương Duy | <ul style="list-style-type: none">● Phụ trách tìm hiểu cách kết hợp Fast R-CNN với mạng RPN tạo nên Faster R-CNN.● Huấn luyện mô hình và đưa ra các bảng số liệu, đánh giá.● Xây dựng mô hình thử nghiệm trên bộ dữ liệu Aircraft● Viết báo cáo phần abstract, giới thiệu● Làm poster |

TÀI LIỆU THAM KHẢO

[1] K. He, X. Zhang, S. Ren, and J. Sun, “Spatial pyramid pooling in deep convolutional networks for visual recognition,” in European Conference on Computer Vision (ECCV), 2014.

[2] R. Girshick, “Fast R-CNN,” in IEEE International Conference on Computer Vision (ICCV), 2015.

[3] J. R. Uijlings, K. E. van de Sande, T. Gevers, and A. W. Smeulders, “Selective search for object recognition,” International Journal of Computer Vision (IJCV), 2013.

[4] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for

accurate object detection and semantic segmentation,” in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.

[5] C. L. Zitnick and P. Dollar, “Edge boxes: Locating object proposals from edges,” in European Conference on Computer Vision (ECCV), 2014.