

THÔNG TIN LUẬN ÁN

Tên luận án: **Phương pháp phát hiện đối tượng khó trong ảnh.**

Chuyên ngành: Khoa học máy tính

Mã số: 62.48.01.01

Họ tên NCS : Nguyễn Duy Khánh

Hướng dẫn khoa học: PGS.TS. Dương Anh Đức, PGS.TS. Lê Đình Duy

Cơ sở đào tạo : Trường Đại học Công nghệ thông tin – ĐHQG TP. HCM

1. TÓM TẮT

Phát hiện đối tượng là một trong những bài toán quan trọng của thị giác máy tính với các ứng dụng trải rộng trong nhiều lĩnh vực khác nhau như: công nghệ robot, xử lý ảnh y khoa, các hệ thống giám sát, hệ thống tương tác người-máy, giao thông thông minh. Trong những năm gần đây, các phương pháp phát hiện đối tượng đã phát triển mạnh mẽ, đặc biệt đạt được bước cải tiến lớn về cả độ chính xác và tốc độ xử lý, điển hình như các phương pháp dựa trên mạng học sâu như R-CNN, Fast R-CNN, Faster R-CNN, Mask R-CNN, YOLO, SSD, Retinanet.

Bằng việc xem xét kết quả từ các mô hình phát hiện đối tượng hiện có, chúng tôi nhận thấy có nhiều đối tượng khó có thể bị bỏ qua hoặc dự đoán sai dưới các tình huống phức tạp trong thực tế. Việc phát hiện thành công các đối tượng khó này sẽ hứa hẹn nâng cao hiệu suất cho các mô hình phát hiện đối tượng. Do vậy, trong luận án này chúng tôi tập trung vào việc đề xuất các phương pháp để phát hiện các đối tượng khó, nhằm cải tiến các mô hình phát hiện đối tượng hiện có.

Trong quá trình thực hiện luận án, NCS đã công bố 4 bài tạp chí: JVCI-2019 (ISI), MTAP-2019 (ISI), IEEE TIP-2019 (ISI), Electronics-2020 (ISI) và 4 bài hội thảo quốc tế: ICONIP-2013 (Rank A), IC3INA-2016, FAIR-2017, ACM IMCOM-2018.

2. CÁC ĐÓNG GÓP CHÍNH CỦA LUẬN ÁN

Nội dung của luận án bao gồm các vấn đề nghiên cứu sau đây:

i. Đề xuất phương pháp phát hiện lại đối tượng khó trên sử dụng mạng học sâu ở hai giai đoạn.

Lấy ý tưởng từ các cách tiếp cận khai thác đối tượng khó, trong luận án này chúng tôi đề xuất một phương pháp tập trung vào việc khai thác các đối tượng khó mà các thuật

toán phát hiện đối tượng hiện đại dựa trên CNN chưa giải quyết tốt. Chúng tôi đề xuất kết hợp bộ phát hiện đối tượng dễ và bộ phát hiện đối tượng khó, trong đó bộ phát hiện đối tượng khó được huấn luyện nhằm phát hiện các đối tượng bị bỏ sót của bộ phát hiện đối tượng dễ.

ii. Đề xuất phương pháp phát sinh dữ liệu nhân tạo nhằm tăng cường hiệu quả của các bộ phát hiện đối tượng trên tập các đối tượng khó.

Xuất phát từ ý tưởng phát sinh dữ liệu nhân tạo để tăng cường cho việc huấn luyện bộ phát hiện đối tượng, chúng tôi đề xuất một phương pháp phát sinh dữ liệu nhân tạo có định hướng. Việc phát sinh dữ liệu nhân tạo cần tập trung vào các đối tượng khó và thông thường ít xuất hiện trong dữ liệu. Chúng tôi đề xuất phát sinh dữ liệu nhân tạo phải được thực hiện kết hợp với quá trình khai thác đối tượng khó trong tập dữ liệu.

iii. Luận án cũng áp dụng kết quả của mô hình phát hiện đối tượng vào bài toán liên quan: phát hiện đối tượng chính trong ảnh (*Salient Object Detection*).

Chúng tôi trích xuất các mặt nạ phân vùng ảnh chạy trên các khung bao từ mô hình phát hiện đối tượng để tính toán thông tin ngữ nghĩa. Khung bao đối tượng có thể thu được từ các bộ phát hiện đối tượng khó đã đề xuất hay bất kỳ phương pháp phát hiện đối tượng tân tiến nào khác. Sau đó chúng tôi đề xuất các ảnh xạ tường minh và không tường minh sử dụng thông tin ngữ nghĩa để phát hiện đối tượng chính với độ chính xác cao.

3. NHỮNG VẤN ĐỀ CÒN BỎ NGỎ CẦN TIẾP TỤC NGHIÊN CỨU

Điểm yếu của các phương pháp đề xuất là chi phí tính toán lớn và quá trình huấn luyện trải qua nhiều giai đoạn. Do đó, các phương pháp này có thể được cải tiến theo hướng tích hợp các giai đoạn vào một kiến trúc mạng thống nhất, hỗ trợ quá trình huấn luyện theo cơ chế liên mạch. Đồng thời với đó là việc rút trích đặc trưng (bản đồ đặc trưng) có thể được chia sẻ ở các bộ phát hiện để giảm thiểu chi phí tính toán.

CÁN BỘ HƯỚNG DẪN

NGHIÊN CỨU SINH

DƯƠNG ANH ĐỨC

LÊ ĐÌNH DUY

NGUYỄN DUY KHÁNH

THESIS INFORMATION

Title:	Hard object detection methods in images
Major:	Computer Science
Major Code:	62.48.01.01
PhD student :	NGUYEN Duy Khanh
Advisors:	Assoc.Prof. Duong Anh Duc, Assoc.Prof. Le Dinh Duy
University :	University of Information Technology, VNU-HCMC

1. ABSTRACT

Object detection is one of the most critical problems in the computer vision field. It has a wide range of application such as robotics, medical image analysis, surveillance systems, human-computer interaction, smart traffic control. Recently, deep learning-based object detection methods have significantly improved accuracy and running time performance such as R-CNN, Fast R-CNN, Faster R-CNN, Mask R-CNN, YOLO, SSD, Retinanet.

However, these methods still fail to detect some hard objects correctly in complicated scenes. In this thesis, we propose hard-object detection methods to improve the performance of existing object detection methods.

We published 4 journal articles: JVCI-2019 (ISI), MTAP-2019 (ISI), IEEE TIP-2019 (ISI), Electronics-2020 (ISI) and 4 conference papers: ICONIP-2013 (Rank A), IC3INA-2016, FAIR-2017, ACM IMCOM-2018.

2. MAIN CONTRIBUTIONS:

- i. **We propose a 2-staged deep learning-based hard-object detection method for residual set:**

We design an easy-object detection module and a hard-object detection module. We train the hard-object detection module to detect residual objects which are ignored by the easy-object detection module.

ii. We propose a data generation method to improve the performance of existing object detection methods on the hard-object set:

If conventionally augmented data improve the overall performance of detectors, then well-intended generated data boost their performance on hard objects. Our data generation module is, therefore, trained in conjunction with the hard-object mining process.

iii. We apply our methods on the salient object detection problem:

We extract the semantic priors from object masks which can be collected from our hard object detector or other state-of-the-art detectors. After that, two individual maps, namely, the explicit saliency map and the implicit saliency map, are computed from the semantic priors. These two maps are adaptively fused together incorporating with post-processing to yield a highly accurate saliency map.

3. OPEN PROBLEMS

Our method still suffers from the high computational cost and multiple-stages of training. Therefore, future works should focus on integrating them into one end-to-end system. Feature maps should also be shared across different detectors to reduce the computational cost.

ADVISORS

PHD STUDENT

DUONG ANH DUC

LE DINH DUY

NGUYEN DUY KHANH