

# **BÁO CÁO ĐỒ ÁN CUỐI KỲ**

**Môn học**

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

**Lớp học**

**CS519.N11**

**Giảng viên**

**PGS.TS. LÊ ĐÌNH DUY**


**Thời gian**

**09/2022 - 02/2023**

----- *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=Jq76H0KSZtM>
- Link slides (dạng .pdf đặt trên Github của nhóm):  
[https://github.com/truong11062002/CS519.N11--PhuongPhapLuanNCKH/blob/main/AN\\_INTENSIVE\\_IMPROVEMENT\\_R-CNN\\_IN\\_VIETNAMESE\\_DOCUMENT\\_DETECTION.pdf](https://github.com/truong11062002/CS519.N11--PhuongPhapLuanNCKH/blob/main/AN_INTENSIVE_IMPROVEMENT_R-CNN_IN_VIETNAMESE_DOCUMENT_DETECTION.pdf)

<ul style="list-style-type: none"><li>● Họ và Tên: Nguyễn Nhật Trường</li><li>● MSSV: 20522087</li></ul> 	<ul style="list-style-type: none"><li>● Lớp: CS519.N11</li><li>● Tự đánh giá (điểm tổng kết môn): 9.5/10</li><li>● Số buổi vắng: 0</li><li>● Số câu hỏi QT cá nhân: 10</li><li>● Số câu hỏi QT của cả nhóm: 3</li><li>● Link Github: <a href="https://github.com/truong11062002">https://github.com/truong11062002</a></li><li>● Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:<ul style="list-style-type: none"><li>○ Lên ý tưởng và viết tóm tắt đề tài</li><li>○ Đưa ra mục tiêu của đề tài</li><li>○ Thiết kế poster và tìm tài liệu tham khảo</li><li>○ Thuyết trình</li></ul></li></ul>
<ul style="list-style-type: none"><li>● Họ và Tên: Lê Trương Ngọc Hải</li><li>● MSSV: 20520481</li></ul>	<ul style="list-style-type: none"><li>● Lớp: CS519.N11</li><li>● Tự đánh giá (điểm tổng kết môn): 9.5/10</li><li>● Số buổi vắng: 1</li><li>● Số câu hỏi QT cá nhân: 10</li><li>● Số câu hỏi QT của cả nhóm: 3</li><li>● Link Github:</li></ul>



<https://github.com/letruongngochai>

- Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:
  - Viết phần Nội dung và Phương pháp
  - Soạn slide
  - Tìm hiểu phần thực nghiệm và đánh giá
  - Phân công và quản lý tiến độ

- Họ và Tên: Lê Thị Phương Vy
- MSSV: 20520355



- Lớp: CS519.N11
- Tự đánh giá (điểm tổng kết môn): 9.5/10
- Số buổi vắng: 1
- Số câu hỏi QT cá nhân: 10
- Số câu hỏi QT của cả nhóm: 3
- Link Github: <https://github.com/Ceci-june>
- Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:
  - Viết phần Giới thiệu và Kết quả mong đợi
  - Thiết kế poster và làm slide
  - Tìm hiểu cách kết hợp mô hình
  - Làm video YouTube

# ĐỀ CƯƠNG NGHIÊN CỨU

## TÊN ĐỀ TÀI (IN HOA)

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

## TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

AN INTENSIVE IMPROVEMENT R-CNN IN VIETNAMESE DOCUMENT DETECTION

## TÓM TẮT (Tối đa 400 từ)

Việc tìm hiểu về bộ dữ liệu tài liệu tiếng Việt trước khi nghiên cứu là một việc cần thiết. Thông qua khảo sát, chúng tôi nhận thấy rằng bộ dữ liệu UIT-DODV là một bộ dữ liệu có thể rất tiềm năng bởi vì kết quả của những phương pháp trước đó đã thực nghiệm trên bộ dữ liệu này chưa được tối ưu hoàn toàn, hơn nữa điều kiện cũng rất phù hợp lượng tài nguyên mà nhóm đang có. Sau đó, nhóm tôi tìm hiểu các phương pháp cải thiện hiệu suất của R-CNN và giới thiệu phương pháp Guided Anchoring Cascade R-CNN - một sự đề xuất sẽ mang lại kết quả vượt trội so với phần còn lại. Cuối cùng là tiến hành thực hiện huấn luyện các phương pháp trên bộ dữ liệu UIT-DODV để cung cấp đánh giá đầy đủ chi tiết về bộ dữ liệu cũng như kết quả mô hình SOTA.

## GIỚI THIỆU (Tối đa 1 trang A4)

Quá trình số hóa tài liệu đã và đang diễn ra trong nhiều tổ chức và doanh nghiệp kể từ khi thời đại công nghệ 4.0 bắt đầu phát triển mạnh mẽ, các tài liệu truyền thống như (giấy, sổ, hóa đơn) đang dần chuyển hóa và thay bằng các tài liệu số (PDF, WORD, EXCEL) được lưu trữ trên các dịch vụ điện toán đám mây để thuận tiện cho truy cập, tìm kiếm, lưu trữ tài liệu. Với lượng lớn tài liệu như vậy việc tìm kiếm tài liệu trở nên khó khăn hơn bao giờ hết. Thế nên một mô hình tốt để nhận diện các thành phần, đối tượng có trong ảnh tài liệu là thật sự cần thiết. Từ đó, chúng tôi quyết định đặt vấn đề vào góc nhìn của bài toán phát hiện đối tượng. Document Object Detection (DOD) hướng đến nhiệm vụ phát hiện tự động các thành phần quan trọng (Caption, Table, Figure, ...) và cấu trúc của trang tài liệu. Bởi sự phổ biến và thông dụng của ngôn ngữ Anh, Trung, hầu hết những mô hình phát hiện đối tượng dành cho tài liệu hiện nay thường tập trung vào hai ngôn ngữ này. Tuy nhiên, các ngôn ngữ khác nhau cũng có các đặc trưng khác nhau từ cách trình bày khác nhau cho các đối tượng như chú thích, công thức cho đến ngữ nghĩa và ngữ pháp trong tài liệu. Do đó, nhóm chúng tôi quyết định chọn Tiếng Việt làm ngôn ngữ chủ đạo để thực hiện bài toán phát hiện đối tượng

trên tài liệu dạng ảnh Tiếng Việt.

**Input:** Một bức ảnh tài liệu, văn bản

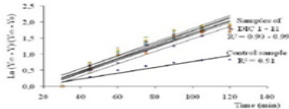
**Output:** Những hộp giới hạn bọc lấy đối tượng mà ta quan tâm kèm theo nhãn của nó được biểu diễn trên chính bức ảnh đầu vào.

Tạp chí Khoa học 2012-21a 43-51

Trường Đại học Cần Thơ

Với  $r$  là bán kính của hạt trích ly (m).  
Có thể giới hạn về phải ở đại lượng đầu tiên và bỏ qua các đại lượng kế tiếp,  
phương trình (8) trở thành:

$$\frac{Y_{m-1}-Y_m}{Y_m-Y_n} = A e^{-Bn} \quad (9)$$



Hình 1: Biểu diễn đồ thị của quá trình trích ly dầu từ Jatropha bằng dung môi hexane.  
Lấy biểu diễn dạng logarit, phương trình (8) sẽ chuyển thành đường thẳng với độ dốc k (như hình 1), từ đó các thông số  $D_{DH}$  được xác định:

$$D_{DH} = \frac{r^2}{2D} \quad (10)$$

Bảng 1: Kết quả thí nghiệm trích ly dầu bằng dung môi hexane từ hạt Jatropha xử lý bởi công nghệ DHC và không xử lý (control) sau 2 giờ trích ly

Mẫu	Giải pháp Crank R <sup>2</sup>	Sản lượng đầu Y <sub>m</sub> (g/g)	Sản lượng đầu Y <sub>n</sub> (g/g)	Hệ số khuếch tán D <sub>DH</sub> (10 <sup>-10</sup> m <sup>2</sup> /s)	Lượng dầu ban đầu Y <sub>n</sub> (g/g)	Y <sub>m</sub> (g/g)
DHC 1	0.915	0.515	112	5.44	225	0.470
DHC 2	0.908	0.506	110	5.62	232	0.459
DHC 3	0.919	0.504	110	5.52	228	0.459
DHC 4	0.903	0.506	110	5.32	220	0.465
DHC 5	0.908	0.511	111	5.13	212	0.473
DHC 6	0.900	0.505	110	5.36	221	0.462
DHC 7	0.939	0.501	109	5.29	218	0.459
DHC 8	0.901	0.493	107	5.90	244	0.444
DHC 9	0.966	0.495	108	5.21	215	0.448
DHC 10	0.935	0.491	107	5.36	221	0.448
DHC 11	0.952	0.492	107	5.50	227	0.444
Control	0.905	0.459	100	7.42	100	0.349

Số liệu thí nghiệm dùng cho mô hình toán học quá trình khuếch tán trên cân bằng qua các điểm thí nghiệm gần thời gian ban đầu  $t = 0$ . Chính vì thế, phép ngoại suy của mô hình toán thu được cho phép xác định được  $Y_n$  không bằng 0. Nghĩa là  $Y_n$

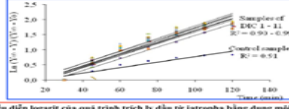
50

Tạp chí Khoa học 2012-21a 43-51

Trường Đại học Cần Thơ

Với  $r$  là bán kính của hạt trích ly (m).  
Có thể giới hạn về phải ở đại lượng đầu tiên và bỏ qua các đại lượng kế tiếp,  
phương trình (8) trở thành:

$$\frac{Y_{m-1}-Y_m}{Y_m-Y_n} = A e^{-Bn} \quad (9)$$



Hình 1: Biểu diễn đồ thị của quá trình trích ly dầu từ Jatropha bằng dung môi hexane.  
Lấy biểu diễn dạng logarit, phương trình (8) sẽ chuyển thành đường thẳng với độ dốc k (như hình 1), từ đó các thông số  $D_{DH}$  được xác định:

$$D_{DH} = \frac{r^2}{2D} \quad (10)$$

Bảng 1: Kết quả thí nghiệm trích ly dầu bằng dung môi hexane từ hạt Jatropha xử lý bởi công nghệ DHC và không xử lý (control) sau 2 giờ trích ly

Mẫu	Giải pháp Crank R <sup>2</sup>	Sản lượng đầu Y <sub>m</sub> (g/g)	Sản lượng đầu Y <sub>n</sub> (g/g)	Hệ số khuếch tán D <sub>DH</sub> (10 <sup>-10</sup> m <sup>2</sup> /s)	Lượng dầu ban đầu Y <sub>n</sub> (g/g)	Y <sub>m</sub> (g/g)
DHC 1	0.915	0.515	112	5.44	225	0.470
DHC 2	0.908	0.506	110	5.62	232	0.459
DHC 3	0.919	0.504	110	5.52	228	0.459
DHC 4	0.903	0.506	110	5.32	220	0.465
DHC 5	0.908	0.511	111	5.13	212	0.473
DHC 6	0.900	0.505	110	5.36	221	0.462
DHC 7	0.939	0.501	109	5.29	218	0.459
DHC 8	0.901	0.493	107	5.90	244	0.444
DHC 9	0.966	0.495	108	5.21	215	0.448
DHC 10	0.935	0.491	107	5.36	221	0.448
DHC 11	0.952	0.492	107	5.50	227	0.444
Control	0.905	0.459	100	7.42	100	0.349

Số liệu thí nghiệm dùng cho mô hình toán học quá trình khuếch tán trên cân bằng qua các điểm thí nghiệm gần thời gian ban đầu  $t = 0$ . Chính vì thế, phép ngoại suy của mô hình toán thu được cho phép xác định được  $Y_n$  không bằng 0. Nghĩa là  $Y_n$

50

a) Đầu vào

b) Đầu ra

## MỤC TIÊU (Viết trong vòng 3 mục tiêu)

- Khảo sát và phân tích bộ dữ liệu tài liệu dạng ảnh tiếng Việt UIT-DODV.
- Tìm hiểu và thực nghiệm đánh giá các phương pháp cải thiện hiệu suất của R-CNN: Double-Head R-CNN, Libra R-CNN và Guided Anchoring.
- Đề xuất phương pháp cải thiện họ RCNN từ sự kết hợp các module cải tiến với phương pháp thuộc họ R-CNN trên bài toán phát hiện đối tượng dạng ảnh tài liệu tiếng Việt.

## NỘI DUNG VÀ PHƯƠNG PHÁP

### 1. Nội dung

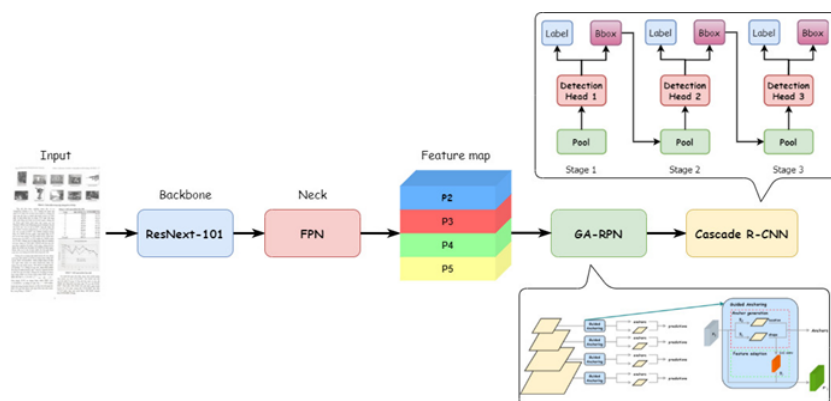
- Khảo sát và phân tích những đặc trưng của bộ dữ liệu dạng ảnh tiếng Việt UIT-DODV để tiếp cận gần hơn với bài toán POD (Page Object Detection). Đồng thời cũng phân tích và nhận xét về những thách thức có thể gặp khi tiếp cận đến bài toán bằng một ngôn ngữ không phổ biến trên thế giới nói chung và trong cộng đồng nghiên cứu nói riêng.
- Nghiên cứu và so sánh các loại hình phương pháp phát hiện đối tượng phổ biến như one-stage, two-stage để tìm ra điểm mạnh, điểm yếu nhằm định hướng kỹ cho hướng

nghiên cứu.

- Nghiên cứu về cơ sở lý thuyết của R-CNN để nắm được những đặc tính của phương pháp. Bên cạnh đó, chúng tôi còn tìm hiểu về các phương pháp gia tăng hiệu suất của họ R-CNN.
- Với giả thiết hiện có: “**Liệu có cách cải tiến hiệu suất phát hiện đối tượng trên trang tài liệu dạng ảnh của họ R-CNN?**”, chúng tôi sẽ nghiên cứu những module có thể kết hợp với những phương pháp họ R-CNN nhằm cải thiện chất lượng kết quả dự đoán và đề xuất ra một phương pháp mới.
- Huấn luyện và đánh giá mô hình để đưa ra những phân tích chuyên sâu. Từ đó không chỉ rút ra được kinh nghiệm trong quá trình huấn luyện, mà còn cải tiến thêm chất lượng mô hình.

## 2. Phương pháp

- Tìm hiểu và phân tích tổ chức của bộ dữ liệu UIT-DODV [1].
- Tìm hiểu và đánh giá các phương pháp cải thiện hiệu suất của R-CNN gồm: Double-Head R-CNN [2], Libra R-CNN [3] và Guided Anchoring [4].
- Đưa ra những nhận định về điểm yếu của những phương pháp trước. So sánh hiệu suất giữa Faster R-CNN và Cascade R-CNN và quyết định chọn phương pháp nào để làm detector chính cho phương pháp đề xuất.
- Cascade R-CNN [5] là một phần mở rộng nhiều giai đoạn của R-CNN trong đó 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.
- Chúng tôi tiến hành thử nghiệm kết hợp mô hình dựa trên kiến trúc Cascade R-CNN kết hợp với Guided Anchoring như tại hình bên dưới để nâng cao hiệu suất phát hiện đối tượng trên tài liệu dạng ảnh. Nếu kết quả tốt như mong đợi, chúng tôi sẽ tiến hành đánh giá và phân tích kết quả của phương pháp. Ngược lại, sẽ thử nghiệm kết hợp các mô hình còn lại của RCNN.



*Kiến trúc dự kiến Guided Anchoring Cascade R-CNN*

## KẾT QUẢ MONG ĐỢI

- Xây dựng bảng báo cáo tổng quan về bài toán phát hiện đối tượng trang.
- Xây dựng báo cáo và đánh giá các phương pháp phát hiện đối tượng phổ biến hiện nay.
- Xây dựng báo cáo về bộ dữ liệu tài liệu tiếng Việt UIT-DODV.
- Nếu kết quả thực nghiệm đạt kết quả như mong đợi tiến hành đề xuất một phương pháp mới cho bài toán phát hiện đối tượng trang.
- Xây dựng bảng đánh giá và phân tích kết quả thực nghiệm trên bộ dữ liệu UIT-DODV.

### **TÀI LIỆU THAM KHẢO** (*Định dạng DBLP*)

- [1]. Linh Truong Dieu, Thuan Trong Nguyen, Nguyen D. Vo, Tam V. Nguyen, Khang Nguyen: Parsing Digitized Vietnamese Paper Documents. CAIP (1) 2021: 382-392
- [2]. Yue Wu, Yinpeng Chen, Lu Yuan, Zicheng Liu, Lijuan Wang, Hongzhi Li, Yun Fu: Rethinking Classification and Localization for Object Detection. CVPR 2020: 10183-10192
- [3]. Jiangmiao Pang, Kai Chen, Jianping Shi, Huajun Feng, Wanli Ouyang, Dahua Lin: Libra R-CNN: Towards Balanced Learning for Object Detection. CVPR 2019: 821-830
- [4]. Jiaqi Wang, Kai Chen, Shuo Yang, Chen Change Loy, Dahua Lin: Region Proposal by Guided Anchoring. CVPR 2019: 2965-2974
- [5]. Zhaowei Cai, Nuno Vasconcelos: Cascade R-CNN: Delving Into High Quality Object Detection. CVPR 2018: 6154-6162



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