

TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI

# ĐỒ ÁN TỐT NGHIỆP

Xây dựng phương pháp học liên tục sử dụng  
Diffusion Model trong bài toán phát hiện vật thể

NGÔ VĂN TÂN

tan.nv210769@sis.hust.edu.vn

Ngành: Khoa học máy tính

Giảng viên hướng dẫn: PGS.TS. Huỳnh Quyết Thắng

Chữ kí GVHD

Khoa: Khoa học máy tính

Trường: Công nghệ Thông tin và Truyền thông

HÀ NỘI, 01/2025

# LỜI CẢM ƠN

Giây phút này thật tuyệt vời, tôi cảm giác mọi mệt mỏi trong 4 năm qua dần tan biến. Thời gian trôi thật nhanh, đã đến lúc tôi nói lời tạm biệt với Bách Khoa.

Tôi xin gửi lời cảm ơn chân thành đến thầy Huỳnh Quyết Thắng và thầy Ngô Văn Linh. Hai thầy đã chia sẻ, định hướng giúp tôi hoàn thành đồ án này. Gặp được các thầy thật sự là một sự kiện đặc biệt trong cuộc đời sinh viên của tôi.

Tôi cũng xin cảm ơn gia đình tôi đã luôn bền bỉ, luôn cố gắng, luôn nỗ lực, luôn động viên tôi trước những khó khăn, thách thức. Gia đình chính là nguồn động lực to lớn kiến tạo nên thành công của tôi.

Bốn năm học khó khăn nhưng nó đã nhẹ nhàng hơn vì đã có những người bạn luôn tin yêu và giúp đỡ. Cảm ơn những người đã ở lại và biết ơn những người đã ra đi trong quãng đường sinh viên của tôi.

Có một chút tiếc nuối trong tôi vì quá trình học tập không như mục tiêu tôi đề ra, nhưng ít nhất tôi đã vượt qua nó. Bách khoa đã mang lại cho tôi những trải nghiệm đầy thú vị. Và khi đã trải nghiệm đủ, tôi phải ra đi và bước tiếp trên con đường mới. Chào tạm biệt Bách Khoa thân yêu! Giây phút này thật quá trọn vẹn.

# LỜI CAM KẾT

Họ và tên sinh viên: Ngô Văn Tân

Điện thoại liên lạc: 0364143275

Email: tanngo06022002@gmail.com

Lớp: Khoa học máy tính 06

Hệ đào tạo: Cử nhân

Tôi – *Ngô Văn Tân* – cam kết Đồ án Tốt nghiệp (ĐATN) là công trình nghiên cứu của bản thân tôi dưới sự hướng dẫn của *PGS.TS. Huỳnh Quyết Thắng*. Các kết quả nêu trong ĐATN là trung thực, là thành quả của riêng tôi, không sao chép theo bất kỳ công trình nào khác. Tất cả những tham khảo trong ĐATN – bao gồm hình ảnh, bảng biểu, số liệu, và các câu từ trích dẫn – đều được ghi rõ ràng và đầy đủ nguồn gốc trong danh mục tài liệu tham khảo. Tôi xin hoàn toàn chịu trách nhiệm với dù chỉ một sao chép vi phạm quy chế của nhà trường.

*Hà Nội, ngày      tháng      năm*

Tác giả ĐATN

*Ngô Văn Tân*

# TÓM TẮT NỘI DUNG ĐỒ ÁN

Phát hiện vật thể là một trong những nhiệm vụ quan trọng trong lĩnh vực học máy và thị giác máy tính, với nhiều ứng dụng thực tiễn như giám sát an ninh, tự động lái xe và y tế. Tuy nhiên, hầu hết các phương pháp hiện tại chỉ tập trung vào việc huấn luyện mô hình trên một tập dữ liệu cố định, gây khó khăn khi mô hình phải thích ứng với dữ liệu mới mà không quên đi kiến thức đã học trước đó.

Trong đồ án này, tôi đề xuất đề "Xây dựng phương pháp học liên tục sử dụng Diffusion Model trong bài toán phát hiện vật thể" với sự kết hợp backbone RT-DETR với phương pháp phát lại dữ liệu bằng mô hình phân tán Stable Diffusion và ControlNet. Lý do chọn hướng tiếp cận này là RT-DETR cung cấp hiệu suất và tốc độ phát hiện cao, trong khi Stable Diffusion cùng ControlNet đảm bảo tạo ra các exemplar chất lượng cao, giữ nguyên các đặc trưng quan trọng như bounding box và vị trí đối tượng mà không cần lưu trữ dữ liệu gốc.

Giải pháp đề xuất bao gồm việc huấn luyện mô hình RT-DETR trên Task 0, sử dụng Stable Diffusion và ControlNet để phát lại dữ liệu exemplar cho Task 1, và áp dụng kỹ thuật knowledge distillation kết hợp với pseudo labels để duy trì kiến thức cũ khi học từ dữ liệu mới.

Đóng góp chính của đồ án bao gồm việc tích hợp RT-DETR làm backbone cho mô hình học liên tục, thiết kế mô hình Stable Diffusion kết hợp ControlNet để phát lại dữ liệu exemplar, và đánh giá hiệu quả của phương pháp đề xuất thông qua các thí nghiệm thực tế. Kết quả nghiên cứu khẳng định tính khả thi và hiệu quả của phương pháp trong việc nâng cao khả năng học liên tục trong phát hiện vật thể, mở ra hướng đi mới cho các nghiên cứu tiếp theo trong lĩnh vực này.

Sinh viên thực hiện

(Ký và ghi rõ họ tên)

## MỤC LỤC

<b>CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI.....</b>	<b>1</b>
1.1 Bài toán học liên tục trong phát hiện vật thể.....	1
1.2 Các giải pháp hiện tại và hạn chế .....	1
1.3 Mục tiêu và định hướng giải pháp .....	2
1.4 Đóng góp của đề án .....	3
1.5 Bố cục đề án .....	4
<b>CHƯƠNG 2. NỀN TẢNG LÝ THUYẾT .....</b>	<b>5</b>
2.1 Ngữ cảnh của bài toán.....	5
2.2 Học sâu .....	5
2.2.1 Tổng quan về học sâu .....	5
2.2.2 Mạng nơ-ron nhân tạo .....	5
2.2.3 Gradient Descent .....	6
2.2.4 Mạng nơ-ron tích chập.....	7
2.3 Vision Transformer.....	8
2.3.1 Tổng quan .....	8
2.3.2 Cơ chế tự chú ý đa đầu.....	8
2.4 Hungarian Matching và Hungarian Loss .....	9
2.4.1 Hungarian Matching.....	9
2.4.2 Hungarian Loss.....	9
2.5 Học liên tục trong phát hiện vật thể .....	10
2.5.1 Sử dụng mô hình phân tán trong phát hiện vật thể .....	11
2.6 RT-DETR: Real-Time Detection Transformer.....	11
2.6.1 Giới thiệu về DETR và RT-DETR.....	11
2.6.2 Kiến trúc và cơ chế hoạt động .....	11

2.6.3 Ứng dụng trong học liên tục .....	12
2.7 Stable Diffusion và ControlNet trong Phát lại dữ liệu .....	12
2.7.1 Giới thiệu về Stable Diffusion .....	12
2.7.2 ControlNet-Canny: Mô hình điều khiển quá trình sinh dữ liệu .....	13
2.7.3 Quá trình phát lại dữ liệu bằng Stable Diffusion và ControlNet .....	13
2.7.4 Lợi ích và thách thức .....	13
2.8 Tổng kết chương 2 .....	14
<b>CHƯƠNG 3. PHƯƠNG PHÁP ĐỀ XUẤT.....</b>	<b>15</b>
3.1 Tổng quan giải pháp.....	15
3.2 Phân chia và Xử lý Nhiệm vụ (Tasks) .....	15
3.2.1 Phân chia dữ liệu thành các nhiệm vụ.....	15
3.2.2 Xử lý dữ liệu cho từng nhiệm vụ.....	16
3.3 Phát triển và Tối ưu RT-DETR trong Học Liên tục.....	16
3.3.1 Sự hiệu quả của mô hình nền tảng RT-DETR.....	16
3.3.2 Tích hợp RT-DETR vào học liên tục .....	17
3.4 Sử dụng Stable Diffusion và ControlNet để Phát lại Dữ liệu .....	17
3.4.1 Trích xuất Canny-image.....	18
3.4.2 Tạo prompt cho mô hình Stable Diffusion - ControlNet .....	18
3.5 Pseudo Labels trong Quá Trình Huấn Luyện .....	19
3.5.1 Phương Pháp Pseudo Labels .....	20
3.5.2 Kết Quả và Đánh Giá .....	21
3.5.3 Kết Luận .....	21
3.6 Phương pháp Huấn luyện và Tối ưu Mô hình .....	22
3.6.1 Quy trình huấn luyện.....	22
3.6.2 Hàm Loss.....	22
3.6.3 Pseudocode cho quá trình huấn luyện.....	22

3.7 Hạn chế của giải pháp đề xuất.....	23
3.7.1 Chất lượng ảnh Canny được tạo ra .....	23
3.7.2 Chất lượng prompt .....	24
3.7.3 Chất lượng của Stable Diffusion - ControlNet.....	24
3.8 Tổng kết chương 3 .....	25
<b>CHƯƠNG 4. ĐÁNH GIÁ THỰC NGHIỆM.....</b>	<b>26</b>
4.1 Tổng quan .....	26
4.2 Bộ dữ liệu COCO2017 .....	26
4.3 Các tham số đánh giá .....	28
4.4 Tiền xử lý và tăng cường dữ liệu .....	29
4.5 Môi trường thử nghiệm .....	31
4.6 Hiệu suất mô hình trên task 0.....	31
4.7 Tác dụng của Pseudo Labels .....	32
4.7.1 Hiệu quả của Pseudo Labels trong Học Liên tục .....	32
4.7.2 Phân tích Thử nghiệm về Tác dụng của Pseudo Labels.....	32
4.7.3 Kết luận về Tác dụng của Pseudo Labels .....	33
4.8 Tác dụng của ControlNet Model.....	33
4.8.1 Phân tích Kết quả Thử nghiệm.....	34
4.8.2 Đánh Giá Tác dụng của ControlNet .....	35
4.8.3 Kết luận về Tác dụng của ControlNet .....	35
4.9 Hiệu suất trên Task 1 và Phát sinh Quên lãng .....	35
4.9.1 Hiện tượng quên lãng tri thức cũ ở task 0 khi huấn luyện task 1 .....	35
4.9.2 Kết quả sau khi điều chỉnh hàm Loss .....	35
4.10 Kết quả đánh giá.....	36
4.11 Tổng kết chương 4.....	36

<b>CHƯƠNG 5. KẾT LUẬN .....</b>	<b>38</b>
5.1 Kết luận.....	38
5.2 Hướng phát triển trong tương lai .....	38
<b>TÀI LIỆU THAM KHẢO.....</b>	<b>41</b>
<b>PHỤ LỤC.....</b>	<b>43</b>
<b>A. Một số kết quả của mô hình ControlNet-Canny .....</b>	<b>43</b>



## DANH MỤC HÌNH VẼ

Hình 2.1	Kiến trúc đơn giản của mạng nơ-ron nhân tạo. . . . .	6
Hình 2.2	Cách hoạt động của bộ lọc trong mạng nơ-ron tích chập. . . .	7
Hình 2.3	Kiến trúc Vision Transformer . . . . .	8
Hình 2.4	Kiến trúc RT-DETR . . . . .	11
Hình 2.5	Kiến trúc Stable Diffusion . . . . .	12
Hình 2.6	Kiến trúc Controlnet . . . . .	13
Hình 3.1	Sơ đồ luồng hoạt động của giải pháp đề xuất. . . . .	16
Hình 3.2	Cách tạo prompt cho mỗi ảnh từ caption và chú thích . . . . .	20
Hình 3.3	Sơ đồ luồng hoạt động của giải pháp đề xuất. . . . .	20
Hình 3.4	Pseudo code cho quá trình huấn luyện . . . . .	23
Hình 4.1	Phân phối số lượng annotations của 80 class trong COCO2017.	28
Hình 4.2	Chỉ số AP được tính bằng phần diện tích màu xanh bên dưới đường Precision-Recall. . . . .	29
Hình 4.3	So sánh giữa hàm mất mát và giá trị mAP của backbone RT- DETR khi huấn luyện task 0 . . . . .	31
Hình 4.4	So sánh sự hiệu quả của model khi sử dụng pseudo_labels . .	33
Hình 4.5	Các hình ảnh sinh ra từ Controlnet-Canny Model . . . . .	34
Hình 4.6	Thực nghiệm về mức độ quên lãng kiến thức khi học task mới	36
Hình 4.7	AP score của model đã cải thiện khi chọn được hàm Loss phù hợp . . . . .	36
Hình A.1	Một số kết quả khác của Controlnet-Canny Model . . . . .	44

## DANH MỤC BẢNG BIỂU

Bảng 4.1	Kết quả độ đo AP score với tỉ lệ phân chia dữ liệu 70-10. Các kết quả được <b>in đậm</b> thể hiện kết quả tốt nhất.(các số liệu khác lấy trong nghiên cứu liên quan) . . . . .	37
Bảng 4.2	Kết quả độ đo AP score với tỉ lệ phân chia dữ liệu 40-40. Các kết quả được <b>in đậm</b> thể hiện kết quả tốt nhất.(các số liệu khác lấy trong nghiên cứu liên quan) . . . . .	37