HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



LƯƠNG ĐỨC THUẬN

NGHIÊN CỨU ÁP DỤNG HỌC TỰ GIÁM SÁT CẢI THIỆN PHÂN LOẠI BỆNH THỰC VẬT TRONG ĐIỀU KIỆN THIẾU DỮ LIỆU GÁN NHÃN

ĐỀ ÁN TỐT NGHIỆP THẠC SĨ KỸ THUẬT

(Theo định hướng ứng dụng)

HÀ NỘI - 2025

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

LƯƠNG ĐỨC THUẬN

NGHIÊN CỨU ÁP DỤNG HỌC TỰ GIÁM SÁT CẢI THIỆN PHÂN LOẠI BỆNH THỰC VẬT TRONG ĐIỀU KIỆN THIẾU DỮ LIỆU GÁN NHÃN

CHUYÊN NGÀNH : KHOA HỌC MÁY TÍNH

MÃ SỐ : 8.48.01.01

ĐỀ ÁN TỐT NGHIỆP THẠC SĨ KỸ THUẬT

(Theo định hướng ứng dụng)

NGƯỜI HƯỚNG DẪN KHOA HỌC

TS. Hoàng Văn Thành

GS.TS. Từ Minh Phương

HÀ NỘI - 2025

BỘ THÔNG TIN VÀ TRUYỀN THÔNG
HỌC VIỆN CÔNG NGHỆ
BƯU CHÍNH VIỄN THÔNG

CỘNG HOÀ XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập - Tự do - Hạnh phúc

BÁO CÁO GIẢI TRÌNH
SỬA CHỮA, HOÀN THIỆN ĐỀ ÁN TỐT NGHIỆP THẠC SĨ

Họ và tên học viên: Lương Đức Thuận.....

Chuyên ngành: Khoa học máy tính .....

Khóa: 2024 đợt I.....

Tên đề tài: Nghiên cứu áp dụng học tự giám sát cải thiện phân loại bệnh thực vật trong điều kiện thiếu dữ liệu gán nhãn .....

.....

Người hướng dẫn khoa học: TS. Hoàng Văn Thành – GS.TS. Từ Minh Phương .....

Ngày bảo vệ: 24/01/2026.....

Các nội dung học viên đã sửa chữa, bổ sung trong đề án tốt nghiệp theo ý kiến đóng góp của Hội đồng chấm đề án tốt nghiệp:

TT	Ý kiến hội đồng	Sửa chữa của học viên
1	Ngoài các phương pháp tự giám sát còn phương pháp nào để so sánh.	Đã bổ sung các phương pháp hiện nay vào mục 1.3.2
2	Phân tích và giải thích kỹ hơn các thành phần đưa vào phương pháp đề xuất.	Đã chỉnh sửa tại mục 3.1, phân tích chi tiết tại các mục 3.4, 3.5, 3.6
3	Phân tích nguyên nhân của kết quả thực nghiệm.	Đã phân tích chi tiết các kết quả tại mục 4.3
v	Làm rõ tốc độ suy diễn, độ trễ triển khai thực tế. Làm trơn nhãn và độ chính xác hỗn hợp có ảnh hưởng như thế nào đến độ trễ	Đã bổ sung giải thích trong mục 3.4.1
5	Các phương pháp khác đánh giá kết quả thực nghiệm	Đã bổ sung thêm vào mục 4.2
6	Hiện tượng MAE hoạt động không ổn định. Liệu có phải đặc thù kiến trúc CNN (thiếu cơ chế Global Attention) khiến việc tái tạo các vùng ảnh bị che khuất không đóng góp nhiều vào trong việc học đặc trưng ngữ nghĩa cho các mạng nơ-ron tích chập	Đã bổ sung thêm vào mục 4.3

	nhỏ không.	
7	Giải thích về ngưỡng tin cậy, cách xác định ngưỡng tự động, phân tích trường hợp lỗi điển hình khi mô hình đối mặt với với dữ liệu chứa nhiễu trong tập large để thấy rõ giới hạn hiện tại của học tương phản	Giải thích và bổ sung trong mục 3.2.2

Hà Nội, ngày 05 tháng 02 năm 2026

Ký xác nhận của

CHỦ TỊCH HỘI ĐỒNG CHẤM ĐỀ ÁN	THƯ KÝ HỘI ĐỒNG	NGƯỜI HƯỚNG DẪN KHOA HỌC	HỌC VIÊN
TS. Nguyễn Duy Phương	TS. Đỗ Thị Liên	TS. Hoàng Văn Thành	Lương Đức Thuận

CNTT 3

BỘ KHOA HỌC VÀ CÔNG NGHỆ
HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

CỘNG HOÀ XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập - Tự do - Hạnh phúc

BIÊN BẢN
HỌP HỘI ĐỒNG CHẤM ĐỀ ÁN TỐT NGHIỆP THẠC SĨ

Căn cứ quyết định số Quyết định số 3299/QĐ-HV ngày 26 tháng 12 năm 2025 của Giám đốc Học viện Công nghệ Bưu chính Viễn thông về việc thành lập Hội đồng chấm đề án tốt nghiệp thạc sĩ. Hội đồng đã họp vào hồi ..8..... giờ ..15.. phút, ngày 24 tháng 01 năm 2025 tại Học viện Công nghệ Bưu chính Viễn thông để chấm đề án tốt nghiệp thạc sĩ cho:

Học viên: **Lương Đức Thuận**

Tên đề án tốt nghiệp: **Nghiên cứu áp dụng học tự giám sát cải thiện phân loại bệnh thực vật trong điều kiện thiếu dữ liệu gán nhãn**

Ngành: **Khoa học máy tính** Mã số: **8.48.01.01**

Các thành viên của Hội đồng chấm đề án tốt nghiệp có mặt: ..5...../ 05

TT	HỌ VÀ TÊN	TRÁCH NHIỆM TRONG HĐ	GHI CHÚ
1	**TS. Nguyễn Duy Phương**	Chủ tịch	
2	**TS. Đỗ Thị Liên**	Thư ký	
3	**PGS.TS. Đỗ Trung Tuấn**	Phản biện 1	
4	**PGS.TS. Nguyễn Hữu Quỳnh**	Phản biện 2	
5	**PGS.TS. Nguyễn Long Giang**	Uỷ viên	

Các nội dung thực hiện:

1. Chủ tịch Hội đồng điều khiển buổi họp. Công bố quyết định của Giám đốc Học viện Công nghệ Bưu chính Viễn thông về việc thành lập Hội đồng chấm đề án tốt nghiệp thạc sĩ.
2. Người hướng dẫn khoa học hoặc thư ký đọc lý lịch khoa học và các điều kiện bảo vệ đề án tốt nghiệp của học viên (Có bản lý lịch khoa học và kết quả các môn học cao học của học viên kèm theo).
3. Học viên trình bày tóm tắt đề án tốt nghiệp.
4. Phản biện 1 đọc nhận xét (có văn bản kèm theo)
5. Phản biện 2 đọc nhận xét (có văn bản kèm theo)
6. Các câu hỏi của thành viên Hội đồng:

Ngoài phương pháp học tự giám sát thì có phương pháp khác để đối sánh không.
Phân tích kỹ hơn và giải thích tính liên quan giữa các thành phần đầu vào và phương pháp đề xuất
Phân tích nguyên nhân của kết quả thực nghiệm
Làm rõ hơi để suy diễn

7. Trả lời của học viên:

Học viên tiếp thu và làm rõ lại.

8. Thư ký đọc nhận xét về quá trình thực hiện đề án tốt nghiệp của học viên *(có văn bản kèm theo).*

9. Hội đồng họp riêng:

- Bầu Ban kiểm phiếu:
  1. Trưởng Ban kiểm phiếu: Đỗ Trung Tuấn
  2. Uỷ viên Ban kiểm phiếu: Đỗ Thị Liên
  3. Uỷ viên Ban kiểm phiếu: Nguyễn Long Giang
- Hội đồng chấm đề án tốt nghiệp bằng bỏ phiếu kín.
- Ban kiểm phiếu làm việc:
- Trưởng Ban kiểm phiếu báo cáo kết quả kiểm phiếu *(có Biên bản họp Ban kiểm phiếu kèm theo)*
- **Điểm trung bình của đề án tốt nghiệp:** 9,5

Kết luận:

1. Các nội dung cần chỉnh sửa, hoàn thiện sau bảo vệ đề án tốt nghiệp:

- Đề nghị học viên rà soát lại nội dung giải pháp đề xuất và giải thích rõ kết quả thực nghiệm theo góp ý của hội đồng.

2. Đề nghị Học viện công nhận (hoặc ~~không~~) và cấp bằng (hoặc ~~không~~) thạc sĩ cho học viên: Vũng Đức Thuận

3. Đề án tốt nghiệp có thể phát triển thành đề tài nghiên cứu cho NCS.....Vũng Đức Thuận.....

Buổi làm việc kết thúc vào.....9h..... cùng ngày.

Chủ tịch	**Thư ký**
TS. Nguyễn Duy Phương	**TS. Đỗ Thị Liên**

BỘ KHOA HỌC VÀ CÔNG NGHỆ
HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập Tự do Hạnh phúc

BẢN NHẬN XÉT
Đề án tốt nghiệp Thạc sỹ kĩ thuật

Tên đề tài đề án : **Nghiên cứu áp dụng học tự giám sát cải thiện phân loại bệnh thực vật trong điều kiện thiếu dữ liệu gán nhãn**

Chuyên ngành: Khoa học máy tính

Mã số: 8.48.01.01.

Tên học viên : **Lương Đức Thuận**

Họ và tên người nhận xét: Đỗ Trung Tuấn,

Đơn vị, Cơ quan công tác: Học viện Công nghệ Bưu chính Viễn thông

Ý KIẾN NHẬN XÉT

1. Tính thời sự, cấp thiết, ý nghĩa khoa học và thực tiễn của đề tài

Dữ liệu gán nhãn là thành phần cốt lõi của Học có giám sát. Nó đóng vai trò để mô hình học tập, từ đó có thể tự đưa ra dự đoán chính xác khi gặp dữ liệu mới chưa được gán nhãn.

Với học tự giám sát, có xu hướng chủ đạo trong năm 2026, cho phép mô hình tự học từ dữ liệu thô không cần nhãn. Mô hình tự tạo ra các "nhiệm vụ giả" như dự đoán phần bị thiếu của hình ảnh hoặc từ tiếp theo trong văn bản.

Đề tài có ứng dụng về phân loại bệnh thực vật, sử dụng học tự giám sát.

Đề án tốt nghiệp đáp ứng mã ngành đào tạo sau đại học, chuyên ngành khoa học máy tính, mã số 8.48.01.01. Học viên có tập thể hướng dẫn là TS. Hoàng văn Thành và GS. TS. Từ Minh Phương.

2. Về nội dung, chất lượng của đề án, các kết quả đã đạt được

Học viên đã có các tài liệu liên quan đến việc bảo vệ kết quả nghiên cứu và ứng

dụng, gồm bản viết đề án tốt nghiệp và bản tóm tắt đề án, đề cương viết đề án đã được bộ môn chuyên ngành thẩm định và góp ý.

Nội dung bản tóm tắt đề án tốt nghiệp thể hiện các ý chính của bản viết đề án tốt nghiệp. Đề án tốt nghiệp của học viên chia thành các chương, liên quan đến các vấn đề sau :

- Bài toán phân loại bệnh thực vật khi thiếu dữ liệu gán nhãn cho bài toán học;
- Khái niệm liên quan đến kết quả của đề án : mạng nơ ron tích chập, nguyên lí học tự giám sát;
- Mô hình học tự giám sát, gồm (i) đề xuất mô hình; (ii) chiến lược tang dữ liệu; (iii) thiết kế mạng nơ ron; (iv) xây dựng hàm tổn thất;
- Thử nghiệm bằng chương trình.

3. Những vấn đề cần giải thích thêm

Liên quan đến kết quả của học viên, qua bản viết đề án tốt nghiệp, người nhận xét thấy :

- Có thể kiêm tra kết quả thu được bằng phương pháp khác không ?

4. Kết luận

Theo các kết quả đạt được của đề án tốt nghiệp về nghiên cứu và thực nghiệm, căn cứ vào yêu cầu của đề án tốt nghiệp thạc sĩ chuyên ngành, tôi đồng ý để học viên Lương Đức Thuận được bảo vệ luận văn trước Hội đồng chấm đề án tốt nghiệp thạc sĩ.

Hà nội, 22 tháng 01 năm 2026

Người nhận xét kí tên

Đỗ Trung Tuấn

Đỗ Trung Tuấn

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM

Độc lập – Tự do – Hạnh phúc

BẢN NHẬN XÉT ĐỀ ÁN TỐT NGHIỆP THẠC SĨ

(Dùng cho người phản biện)

Tên đề tài đề án tốt nghiệp: NGHIÊN CỨU ÁP DỤNG HỌC TỰ GIÁM SÁT CẢI THIỆN PHÂN LOẠI BỆNH THỰC VẬT TRONG ĐIỀU KIỆN THIẾU DỮ LIỆU GÁN NHÃN

Chuyên ngành: KHOA HỌC MÁY TÍNH

Mã chuyên ngành: 8.48.01.01

Họ và tên học viên: Lương Đức Thuận

Họ và tên người nhận xét: Nguyễn Hữu Quỳnh

Học hàm, học vị: PGS.TS

Chuyên ngành: KHMT

Cơ quan công tác: Trường Đại học CMC

Số điện thoại: 0912174793.....E-mail: nhquynhml@gmail.com.....

NỘI DUNG NHẬN XÉT

I/ Cơ sở khoa học và thực tiễn, tính cấp thiết của đề tài:

- Đề tài tập trung giải quyết vấn đề trong ứng dụng AI vào nông nghiệp là sự khan hiếm dữ liệu hình ảnh bệnh thực vật được gán nhãn bởi chuyên gia.

-Đề tài dựa trên các nghiên cứu hiện đại về thị giác máy tính và học sâu (Deep Learning), đặc biệt là hướng đi mới về học tự giám sát (Self-supervised Learning - SSL) để khai thác nguồn dữ liệu không nhãn khổng lồ.

II/ Nội dung của đề án tốt nghiệp, các kết quả đã đạt được:

-Học viên đã nghiên cứu và áp dụng kỹ thuật Học tương phản (Contrastive Learning) với thuật toán MoCo v2.

- Cải tiến mô hình tích hợp song song luồng học tự giám sát và luồng học giám sát, sử dụng mạng xương sống (backbone) ResNet10t gọn nhẹ phù hợp cho thiết bị biên.

- Mô hình đạt độ chính xác 72.02% trên tập PlantDoc với chỉ 5.4 triệu tham số, tăng hơn các mô hình nặng như VGG16 hay ResNet152 (từ 58-139 triệu tham số)

III/ Những vấn đề cần giải thích thêm:

- Kết quả thực nghiệm cho thấy MAE hoạt động không ổn định và cho kết quả kém hơn MoCo v2 trên kiến trúc ResNet10t. Học viên hãy phân tích sâu hơn nguyên nhân kỹ thuật của hiện tượng này. Liệu có phải do đặc thù của kiến trúc CNN (thiếu cơ chế Global Attention

như trong Vision Transformer) khiến việc tái tạo các vùng ảnh bị che khuất (masked pixels) không đóng góp nhiều vào việc học đặc trưng ngữ nghĩa cho các mạng nơ-ron tích chập nhỏ?.

- Đề án chỉ ra một phát hiện quan trọng: tập dữ liệu "Small" (ít ảnh nhưng độ tinh khiết cao) cho kết quả tốt hơn tập "Large" (nhiều ảnh nhưng nhiều nhiễu). Học viên cần giải thích rõ hơn về "ngưỡng tin cậy" 0.44 khi lọc dữ liệu. Làm thế nào để xác định ngưỡng này một cách tự động cho các loại cây trồng khác mà không cần thử sai nhiều lần? Đồng thời, hãy phân tích một vài trường hợp lỗi (Failure Analysis) điển hình khi mô hình đối mặt với dữ liệu nhiễu trong tập "Large" để thấy rõ giới hạn hiện tại của phương pháp học tương phản.

- Mặc dù mô hình đạt được sự tối ưu về số lượng tham số (5.4 triệu tham số), học viên cần cho biết thêm về tốc độ suy diễn thực tế (Inference Time) hoặc mức tiêu thụ bộ nhớ RAM khi chạy trên các phần cứng hạn chế như Raspberry Pi hay Jetson Nano. Việc sử dụng kỹ thuật "Làm trơn nhãn" (Label Smoothing) và "Độ chính xác hỗn hợp" (Mixed Precision) có ảnh hưởng như thế nào đến độ trễ này trong thực tế triển khai?.

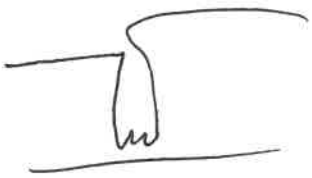
IV/ Kết luận:

Đề án có cơ sở khoa học rõ ràng, phương pháp tiếp cận phù hợp và đạt được các kết quả thực nghiệm có ý nghĩa cả về mặt học thuật lẫn ứng dụng. Nội dung đề án đáp ứng tốt các yêu cầu của một luận văn Thạc sĩ kỹ thuật, học viên đủ điều kiện để bảo vệ trước Hội đồng.

Đồng ý cho phép học viên bảo vệ đề án tốt nghiệp.

Ngày 19 tháng 1 năm 2025

NGƯỜI NHẬN XÉT



19% Tính tương đồng nói chung

Tổng cộng của tất cả các kết quả trùng khớp, bao gồm cả các nguồn trùng lặp, cho mỗi c...

Đã lọc khỏi Báo cáo

- Mục lục tham khảo
- Văn bản được trích dẫn
- Văn bản được trích dẫn
- Kết quả trùng khớp nhỏ (ít hơn 8 từ)

Nguồn hàng đầu

13% Nguồn Internet
7% Ấn bản
8% Bài tập được nộp (bài của học sinh)

Người hướng dẫn khoa học: TS. Hoàng Văn Thành

Học viên:
Thuần

Lương Đức Thuần

Được quét bằng CamScanner

LỜI CAM ĐOAN

Tôi cam đoan đây là công trình nghiên cứu của tôi.

Các số liệu, kết quả nêu trong đề án là trung thực và chưa từng được ai công bố trong bất kỳ công trình nào khác.

Tác giả đề án

Lương Đức Thuận

LỜI CẢM ƠN

Em xin được bày tỏ sự tri ân chân thành đến toàn thể quý thầy cô và cán bộ của Học viện Công nghệ Bưu chính Viễn thông Hà Nội, đặc biệt là các giảng viên thuộc Khoa Công nghệ Thông tin 1 và Khoa Sau Đại học.

Đặc biệt, em xin gửi lời cảm ơn sâu sắc đến thầy TS. Hoàng Văn Thành, người đã hướng dẫn em trong suốt quá trình thực hiện đề tài. Những góp ý sâu sắc, những chia sẻ về kinh nghiệm nghiên cứu, và sự tận tình chỉ bảo của thầy không chỉ giúp em hoàn thiện luận văn mà còn góp phần định hình phương pháp nghiên cứu khoa học của bản thân.

Bên cạnh đó, em xin gửi lời cảm ơn đến gia đình, người thân và bạn bè - những người đã luôn đồng hành, ủng hộ và động viên em trong suốt quá trình học tập và nghiên cứu. Sự hiện diện và những lời động viên kịp thời của họ là nguồn động lực to lớn giúp em vượt qua những thách thức và hoàn thành luận văn một cách tốt nhất.

Đề án này không thể tránh khỏi những hạn chế và thiếu sót, em chân thành mong nhận được những ý kiến đóng góp quý báu từ quý thầy cô, các nhà nghiên cứu và đồng nghiệp. Những góp ý này sẽ là cơ sở quan trọng để em tiếp tục hoàn thiện và nâng cao chất lượng nghiên cứu của mình, đồng thời mở rộng hiểu biết trong lĩnh vực chuyên môn.

Hà Nội, tháng 12 năm 2025

Lương Đức Thuận

MỤC LỤC

LỜI CAM ĐOAN i
LỜI CẢM ƠN ii
MỤC LỤC iii
DANH MỤC VIẾT TẮT v
DANH MỤC BẢNG vi
DANH MỤC HÌNH ẢNH vii
MỞ ĐẦU 1
CHƯƠNG 1: TỔNG QUAN VỀ BÀI TOÁN PHÂN LOẠI BỆNH THỰC VẬT TRONG ĐIỀU KIỆN THIẾU DỮ LIỆU GÁN NHÃN 4
1.1. Thị giác máy tính và vai trò trong nông nghiệp chính xác 4
1.1.1. Tổng quan về giám sát và chẩn đoán bệnh cây trồng tự động 4
1.1.2. Ứng dụng của thị giác máy tính và học sâu trong nhận diện bệnh thực vật 4
1.2. Tổng quan bài toán phân loại bệnh thực vật trong điều kiện thiếu dữ liệu gán nhãn . 6
1.2.1. Thực trạng khan hiếm dữ liệu gán nhãn chất lượng cao trong thực tế 6
1.2.2. Tác động của thiếu dữ liệu đến hiệu suất của mô hình 7
1.2.3. Bài toán phân loại bệnh thực vật trong điều kiện thiếu dữ liệu gán nhãn 9
1.3. Các nghiên cứu liên quan gần đây 10
1.3.1. Các phương pháp nghiên cứu truyền thống trước đây 10
1.3.2. Phương pháp hiện nay 11
1.4. Kết luận chương 1 13
CHƯƠNG 2: CƠ SỞ LÝ THUYẾT 15
2.1. Các thành phần chính của mạng nơ-ron tích chập 15
2.2. Các mô hình CNN nổi tiếng phù hợp với bài toán phân loại bệnh thực vật. 20
2.2.1. Quá trình huấn luyện mô hình học sâu 20
2.2.2. Các mô hình CNN nổi tiếng phù hợp 21
2.3. Học tự giám sát: nguyên lý và kiến trúc chung và các kĩ thuật liên quan 23
2.3.1. Định nghĩa 23
2.3.2. Phân loại 25
2.4. Kết luận chương 2 27

CHƯƠNG 3: ĐỀ XUẤT MÔ HÌNH ÁP DỤNG SSL CHO BÀI TOÁN PHÂN LOẠI BỆNH THỰC VẬT TRONG ĐIỀU KIỆN THIẾU DỮ LIỆU GÁN NHÃN 28
3.1. Đề xuất kiến trúc mô hình 28
3.2. Xây dựng dữ liệu 29
3.2.1. Tập dữ liệu PlantDoc 29
3.2.2. Xây dựng tập dữ liệu SSL 31
3.3. Chiến lược tăng cường dữ liệu 35
3.3.1. Tăng cường dữ liệu cho Plandoc dataset 35
3.3.2. Tăng cường dữ liệu cho SSL dataset 36
3.4. Thiết kế thành phần mạng nơ-ron 37
3.4.1. Mạng xương sống 37
3.4.2. Chiếu phi tuyến MLP 38
3.4.3. Bộ phân loại CLS 39
3.5. Cơ chế huấn luyện và cập nhập trọng số 39
3.6. Xây dựng hàm mất mát 40
3.6.1. Hàm mất mát phân loại với kỹ thuật làm trơn nhãn 41
3.6.2. Hàm mất mát tương phản InfoNCE cho học tự giám sát 41
3.6.3. Hàm mất mát tổng hợp và chiến lược trọng số động 42
3.7. Các kỹ thuật liên quan 42
3.7.1. Gradient Accumulation 42
3.7.2. Automatic Mixed Precision 43
3.7.3. Điều chuẩn Suy giảm trọng số 43
3.8. Kết luận chương 3 43
CHƯƠNG 4: THỰC NGHIỆM VÀ ĐÁNH GIÁ 45
4.1. Cấu hình phần cứng 45
4.2. Các tiêu chí đánh giá 45
4.2.1. Độ chính xác tổng thể (Accuracy) 46
4.2.2. Độ che phủ (Recall) 47
4.2.3. Độ chính xác (Precision) 47
4.2.4. F1-score 47
4.3. Kết quả thực nghiệm 48
4.4. Kết luận chương 4 53
KẾT LUẬN 55

TÀI LIỆU THAM KHẢO 57

DANH MỤC VIẾT TẮT

Viết tắt	Tiếng Anh	Tiếng Việt
ReLU	Rectified linear unit	Đơn vị tuyến tính được chỉnh lưu
SGD	Stochastic gradient descent	Giảm gradient ngẫu nhiên
AI	Artificial intelligence	Trí tuệ nhân tạo
CNN	Convolutional neural network	Mạng Nơ-ron tích chập
SVM	Support vector machine	Máy hỗ trợ vector
SSL	Self-supervised learning	Học tự giám sát
	Supervised learning	Học giám sát
	Unsupervised learning	Học không giám sát
k-NN	K near neighbor	K hàng xóm gần nhất
SimCLR	Simple Framework for Contrastive Learning of Visual Representations	Khung đơn giản để học tương phản các biểu diễn trực quan
MoCo	Momentum Contrast	Sự tương phản về động lượng
MAE	Masked Autoencoders	Bộ mã hóa tự động có mặt nạ
MLP	Multi-layer Perceptron	Mạng nơ-ron nhiều lớp
AMP	Automatic Mixed Precision	Độ chính xác hỗn hợp tự động

DANH MỤC BẢNG

Bảng 1: Bộ dữ liệu Plantdoc 29
Bảng 2: Tập dữ liệu cho quá trình học tự giám sát 35
Bảng 3: Tham số tăng cường cho quá trình huấn luyện 36
Bảng 4: Tham số tăng cường cho quá trình kiểm thử 36
Bảng 5: Tham số tăng cường dữ liệu cho dữ liệu học tự giám sát 36
Bảng 6: Kết quả các phương pháp học giám sát và học tự giám sát trên bộ dữ liệu Plantdoc 49
Bảng 7: Thực nghiệm hệ số alpha, beta của mô hình đề xuất trên tập dữ liệu plantdoc 51
Bảng 8: Kết quả với các tập dữ liệu không giám sát trên kiến trúc mới 51
Bảng 9: Hiệu suất các mô hình trên tập dữ liệu plantdoc 52

DANH MỤC HÌNH ẢNH

Hình 1: Minh hoạ ứng dụng thị giác máy tính trong nông nghiệp 4
Hình 3: Số mẫu tập dữ liệu plantdoc cho thấy sự mất cân bằng dữ liệu..... 6
Hình 4: Underfit, Optimal, Overfit 8
Hình 5: Các thành phần cơ bản của mạng nơ-ron tích chập 15
Hình 6: Lớp tích chập 16
Hình 8: Đồ thị hàm sigmoid và ReLU 17
Hình 9: Cách ReLU hoạt động 17
Hình 10: Max pooling và average pooling 18
Hình 11: Mô phỏng mạng kết nối đầy đủ 19
Hình 12: Mô hình đề xuất 29

MỞ ĐẦU

Trong bối cảnh cuộc Cách mạng công nghiệp lần thứ tư đang diễn ra mạnh mẽ, việc ứng dụng công nghệ thông tin và trí tuệ nhân tạo đã tạo nên những bước đột phá sâu rộng trong mọi mặt của đời sống kinh tế xã hội. Đặc biệt trong lĩnh vực nông nghiệp, sự tích hợp của các công nghệ tiên tiến không chỉ dừng lại ở việc số hóa quy trình quản lý mà còn thâm nhập sâu vào các khâu canh tác cốt lõi, đặt nền móng vững chắc cho nền nông nghiệp chính xác và thông minh. Nền tảng này được xây dựng dựa trên các trụ cột then chốt nhằm tối ưu hóa năng suất và đảm bảo an ninh lương thực bền vững. Trụ cột đầu tiên là giám sát môi trường và sinh trưởng cây trồng, cho phép thu thập dữ liệu liên tục theo thời gian thực. Trụ cột thứ hai là quản lý canh tác thông minh, hỗ trợ ra quyết định dựa trên dữ liệu để sử dụng hiệu quả tài nguyên đất và nước. Trụ cột cuối cùng, và cũng là yếu tố quan trọng nhất, là chẩn đoán và bảo vệ thực vật tự động, ứng dụng thị giác máy tính để phát hiện sớm dịch bệnh, giúp giảm thiểu rủi ro và nâng cao chất lượng nông sản thương phẩm.

Việc phát hiện sớm và chính xác các loại bệnh hại trên cây trồng đóng vai trò quyết định trong việc bảo vệ mùa màng và hiệu quả kinh tế của người nông dân. Các nghiên cứu nông học đã chỉ ra rằng việc can thiệp xử lý dịch bệnh ngay từ giai đoạn khởi phát có thể giảm thiểu thiệt hại năng suất lên tới mức tối đa và hạn chế đáng kể lượng thuốc bảo vệ thực vật cần sử dụng. Trước đây, quy trình này phụ thuộc hoàn toàn vào kinh nghiệm quan sát thủ công, vốn tốn nhiều thời gian, nhân lực và thường mang tính chủ quan. Với sự hỗ trợ của công nghệ xử lý ảnh hiện đại, các hệ thống chẩn đoán tự động có khả năng nhận diện các dấu hiệu bất thường như vết đốm, sự biến đổi màu sắc lá hay các tổn thương mô học với tốc độ và độ chính xác cao. Sự chuyển dịch từ canh tác dựa trên kinh nghiệm sang canh tác dựa trên dữ liệu hình ảnh không chỉ giúp khoanh vùng dịch bệnh nhanh chóng mà còn hỗ trợ các chuyên gia đưa ra phác đồ xử lý kịp thời, góp phần xây dựng một nền nông nghiệp xanh và bền vững hơn.

Tuy nhiên, quá trình triển khai các mô hình trí tuệ nhân tạo vào thực tiễn sản xuất nông nghiệp đang đối mặt với những rào cản kỹ thuật to lớn. Thách thức lớn nhất nằm ở khía cạnh dữ liệu huấn luyện. Để đạt được độ chính xác cao, các mô hình học sâu truyền thống đòi hỏi một lượng lớn dữ liệu hình ảnh chất lượng cao và phải được gán nhãn chính xác bởi các chuyên gia bệnh học cây trồng. Công việc gán nhãn

này tiêu tốn chi phí và thời gian rất lớn, dẫn đến tình trạng khan hiếm dữ liệu định danh nghiêm trọng, đặc biệt đối với các loại bệnh mới hoặc hiếm gặp. Bên cạnh đó, dữ liệu ảnh thu thập ngoài thực địa thường chứa nhiều nhiễu nền, điều kiện ánh sáng thay đổi và không đồng nhất về định dạng, gây khó khăn cho quá trình trích xuất đặc trưng. Vấn đề mất cân bằng dữ liệu cũng là một trở ngại lớn khi số lượng mẫu cây khỏe mạnh thường áp đảo so với mẫu bệnh, khiến mô hình có xu hướng thiên vị và bỏ sót các ca bệnh quan trọng. Những hạn chế này đặt ra yêu cầu cấp thiết về việc tìm kiếm các phương pháp học tập mới có khả năng giảm thiểu sự phụ thuộc vào dữ liệu gán nhãn thủ công.

Nhận thức được tầm quan trọng của việc tháo gỡ nút thắt về dữ liệu trong chẩn đoán bệnh cây trồng, luận văn tập trung nghiên cứu và đề xuất giải pháp ứng dụng kỹ thuật học tự giám sát nhằm cải thiện hiệu năng phân loại bệnh thực vật trong điều kiện thiếu dữ liệu gán nhãn. Về phương pháp luận, nghiên cứu khai thác sức mạnh của các thuật toán học tương phản tiên tiến để tận dụng nguồn dữ liệu hình ảnh khổng lồ chưa được gán nhãn, từ đó huấn luyện mô hình trích xuất được các đặc trưng bệnh học mạnh mẽ và tổng quát. Luận văn sử dụng các bộ dữ liệu chuẩn như PlantDoc kết hợp với các nguồn dữ liệu mở rộng để thực nghiệm và đánh giá. Các mô hình mạng nơ-ron tích chập được lựa chọn và tối ưu hóa nhằm đảm bảo khả năng vận hành hiệu quả ngay cả khi số lượng mẫu huấn luyện có nhãn là cực kỳ hạn chế.

Trên cơ sở phân tích thực trạng và ý nghĩa khoa học của bài toán, luận văn thực hiện đề tài: “Nghiên cứu áp dụng học tự giám sát cải thiện phân loại bệnh thực vật trong điều kiện thiếu dữ liệu gán nhãn”.

Nội dung đề án sẽ bao gồm phần mở đầu, phần kết luận và các chương nội dung được tổ chức như sau:

CHƯƠNG 1: TỔNG QUAN VỀ BÀI TOÁN PHÂN LOẠI BỆNH THỰC VẬT TRONG ĐIỀU KIỆN THIẾU DỮ LIỆU GÁN NHÃN

- Trình bày về tầm quan trọng và các ứng dụng của thị giác máy tính trong nông nghiệp nói chung và bài toán phân loại bệnh thực vật nói riêng.
- Phân tích sâu về sự khan hiếm dữ liệu được gán nhãn. Nêu bật tác động của chúng đến hiệu suất mô hình.

- Giới thiệu tổng quan về các giải pháp tiềm năng, dẫn dắt đến sự cần thiết của áp dụng học Tự Giám Sát.
- Tóm tắt các vấn đề và xác định mục tiêu cần giải quyết trong các chương tiếp theo.

CHƯƠNG 2: CƠ SỞ LÝ THUYẾT

- Các thành phần chính của mạng nơ-ron tích chập
- Các mô hình CNN nổi tiếng phù hợp với bài toán phân loại bệnh thực vật.
- Học Tự Giám Sát (SSL) nguyên lý và kiến trúc chung và các kĩ thuật liên quan.

CHƯƠNG 3: ĐỀ XUẤT MÔ HÌNH ÁP DỤNG SSL CHO BÀI TOÁN TOÁN PHÂN LOẠI BỆNH THỰC VẬT TRONG ĐIỀU KIỆN THIẾU DỮ LIỆU GÁN NHÃN

- Dựa trên backbone có ít tham số, phù hợp với bài toán như mobilenet v2.
- Thực nghiệm các kĩ thuật học tự giám sát
- Xây dựng mô hình đề xuất
- Luận giải tại sao việc áp dụng phương pháp này lại mang lại lợi ích vượt trội. học tự giám sát cung cấp điểm khởi đầu tốt cho các đặc trưng
- Mô tả chi tiết các thành phần, luồng dữ liệu và hàm mất mát của mô hình.

CHƯƠNG 4: THỰC NGHIỆM VÀ ĐÁNH GIÁ

- Mục tiêu thực nghiệm để kiểm chứng hiệu quả của mô hình.
- Trình bày các kịch bản thực nghiệm
- Phân tích, so sánh, trực quan hoá và bình luận về kết quả, làm rõ ưu thế của mô hình đề xuất.

CHƯƠNG 1: TỔNG QUAN VỀ BÀI TOÁN PHÂN LOẠI BỆNH THỰC VẬT TRONG ĐIỀU KIỆN THIẾU DỮ LIỆU GÁN NHÃN

1.1. Thị giác máy tính và vai trò trong nông nghiệp chính xác

1.1.1. Tổng quan về giám sát và chẩn đoán bệnh cây trồng tự động

Trong những năm gần đây, việc triển khai các hệ thống giám sát và chẩn đoán bệnh tự động trở thành một yêu cầu cấp thiết. Các hệ thống này cho phép theo dõi tình trạng của cây trồng theo thời gian thực trên quy mô lớn, đảm bảo khả năng phát hiện sớm các bệnh trên cây trồng ngay từ giai đoạn khởi phát, khi các dấu hiệu còn mờ nhạt đối với mắt thường. Khả năng vận hành liên tục và độ tin cậy cao của máy móc giúp loại bỏ các sai số do yếu tố tâm lý hay sự mệt mỏi của con người, cung cấp cơ sở dữ liệu chính xác để kích hoạt các quy trình can thiệp kịp thời.



Hình 1: Minh hoạ ứng dụng thị giác máy tính trong nông nghiệp

1.1.2. Ứng dụng của thị giác máy tính và học sâu trong nhận diện bệnh thực vật

Sự tích hợp của thị giác máy tính vào lĩnh vực bệnh học cây trồng đã trải qua một bước tiến hóa mang tính nền tảng, chuyển dịch từ các kỹ thuật xử lý ảnh sang

các mô hình học sâu mới. Trong giai đoạn đầu, các phương pháp nhận diện chủ yếu sử dụng việc trích xuất đặc trưng thủ công, nơi các kỹ sư phải thiết kế các thuật toán để định lượng các thuộc tính vật lý như màu sắc, kết cấu bề mặt hay hình dạng hình học của vết bệnh. Mặc dù các phương pháp này đạt được nhiều thành công trong điều kiện phòng thí nghiệm với phông nền đơn giản, chúng thường bộc lộ sự kém hiệu quả khi đối mặt với sự phức tạp của môi trường trong tự nhiên, nơi các yếu tố như nhiễu nền, ánh sáng thay đổi hay sự chồng chéo của tán lá làm sai lệch các đặc trưng đã được thiết kế sẵn. Sự ra đời của mạng nơ-ron tích chập sâu đã khắc phục phần nào hạn chế này bằng cơ chế học biểu diễn phân cấp, cho phép hệ thống tự động khám phá và tối ưu hóa các đặc trưng được trích xuất từ dữ liệu thô mà không cần sự can thiệp của con người trong giai đoạn trích xuất đặc trưng.

Sức mạnh của các mô hình học sâu nằm ở khả năng mô phỏng cơ chế thị giác sinh học thông qua hàng loạt các lớp tính toán phi tuyến. Tại các tầng thấp, mạng nơ-ron học cách nhận biết các cấu trúc cơ bản như cạnh, góc và sự chuyển tiếp màu sắc. Khi thông tin đi qua các tầng sâu hơn, các cấu trúc này được tổng hợp thành các đặc trưng ngữ nghĩa phức tạp hơn như đốm bệnh, vết hoại tử hay cấu trúc gân lá biến dị. Khả năng trừu tượng hóa thông tin này cho phép các kiến trúc mạng hiện đại như Residual Networks hay EfficientNet phân biệt chính xác các loại bệnh có triệu chứng hình thái tương đồng, thậm chí phân biệt được các giai khác nhau của cùng một bệnh. Việc áp dụng các kiến trúc tiên tiến này đã nâng độ chính xác của các hệ thống chẩn đoán lên mức tương đương hoặc thậm chí vượt trội hơn so với đánh giá của chuyên gia trong một số tác vụ cụ thể.

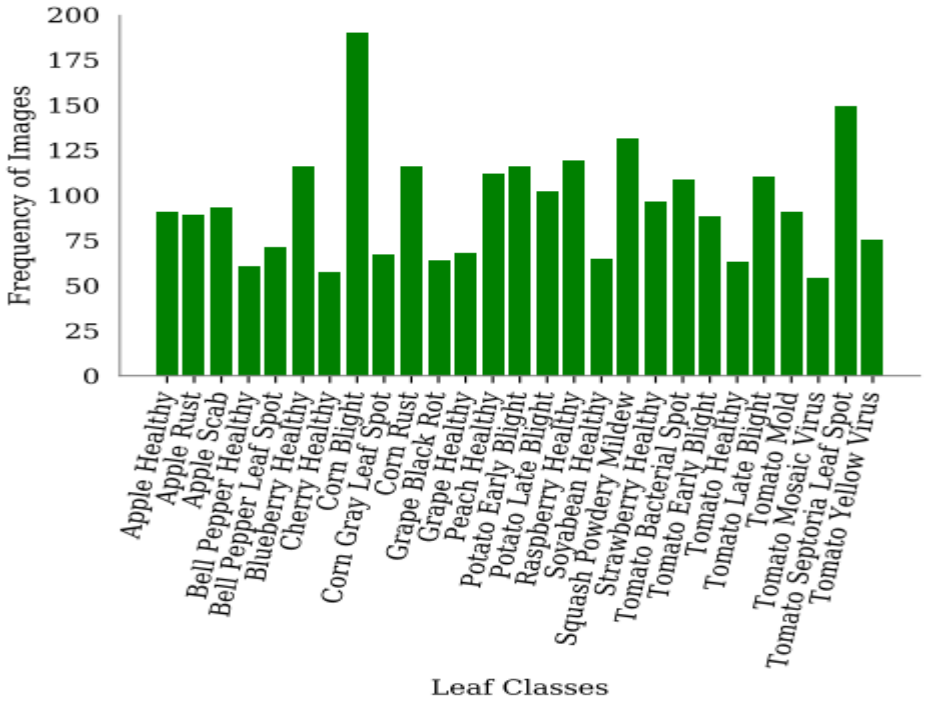
Ngày nay, sự thành công của phương pháp học giám sát truyền thống hiện nay đang phụ thuộc chặt chẽ vào số lượng và chất lượng của dữ liệu huấn luyện có nhãn. Để đạt được độ chính xác cao và khả năng tổng quát hóa tốt, các mô hình học sâu đòi hỏi hàng nghìn, thậm chí hàng triệu hình ảnh mẫu đã được các chuyên gia nông học định danh chính xác. Yêu cầu khắt khe này tạo ra một rào cản vô hình trong việc triển khai công nghệ đối với các loại cây trồng mới hoặc các bệnh dịch hiếm gặp, nơi dữ liệu thu thập được thường rất hạn chế và rời rạc. Chính thách thức về sự phụ thuộc vào dữ liệu nhãn đã thúc đẩy cộng đồng nghiên cứu chuyển hướng sang khai thác các

phương pháp học tập tiên tiến hơn, nhằm tận dụng tốt các nguồn hình ảnh khổng lồ chưa được gán nhãn đang tồn tại trong thực tế sản xuất.

1.2. Tổng quan bài toán phân loại bệnh thực vật trong điều kiện thiếu dữ liệu gán nhãn

1.2.1. Thực trạng khan hiếm dữ liệu gán nhãn chất lượng cao trong thực tế

Trong bài toán phân loại bệnh thực vật, khái niệm dữ liệu chất lượng cao không chỉ dừng lại ở độ phân giải hình ảnh mà còn nằm ở độ chính xác của nhãn dữ liệu. Quy trình gán nhãn cho các tập dữ liệu bệnh cây trồng phức tạp hơn rất nhiều so với các tác vụ nhận diện vật thể thông thường. Công việc này đòi hỏi sự tham gia trực tiếp từ các chuyên gia nông học hoặc những người có nhiều năm kinh nghiệm và kiến thức chuyên môn sâu để có thể phân biệt chính xác các triệu chứng bệnh lý, vốn thường có sự tương đồng cao về hình thái giữa các chủng bệnh khác nhau hoặc biến thiên mạnh mẽ theo từng thời kì, từng giai đoạn phát triển của cây. Sự phụ thuộc vào nhân lực chuyên gia khiến cho chi phí gán nhãn tăng cao, tốn kém nhiều thời gian và khó thực hiện, dẫn đến tình trạng khan hiếm dữ liệu định danh nghiêm trọng đối với các loại cây trồng bản địa hoặc các loại bệnh hiếm gặp.



Hình 2: Số mẫu tập dữ liệu plantdoc cho thấy sự mất cân bằng dữ liệu

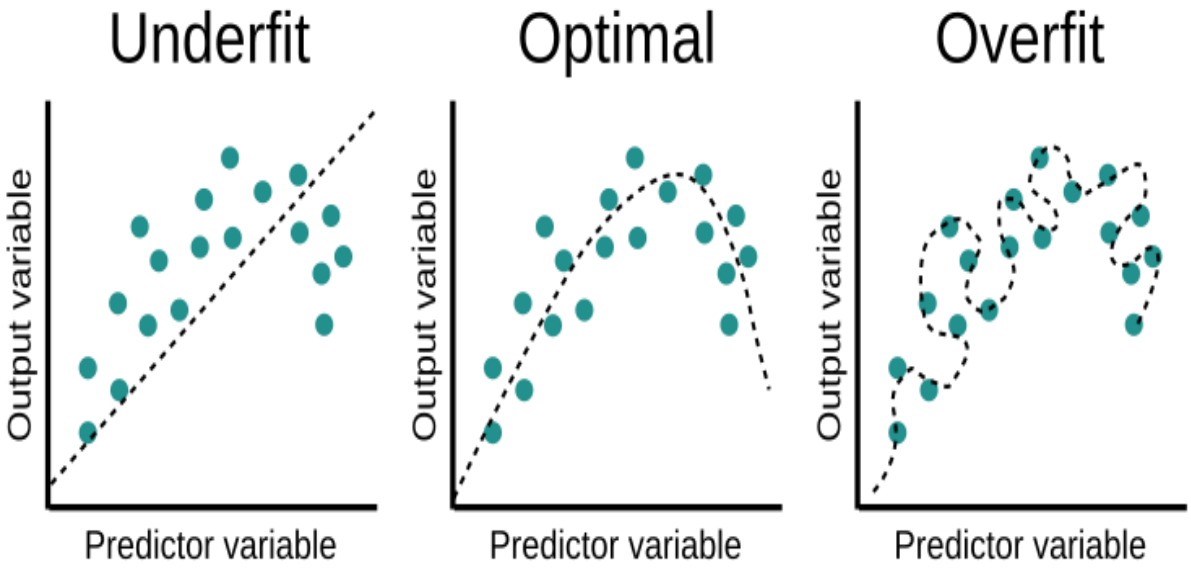
Bên cạnh vấn đề về nhân lực, tính chất phức tạp và nhiễu loạn của môi trường thực tế cũng làm trầm trọng thêm sự thiếu hụt dữ liệu đạt chuẩn. Các hình ảnh thu thập ngoài tự nhiên thường chịu tác động bởi ngoại cảnh như phông nền phức tạp chứa nhiều đất đá, cỏ dại, điều kiện ánh sáng thay đổi hoặc sự che khuất giữa các tầng lá. Trong khi đó, phần lớn các bộ dữ liệu công khai hiện có thường được thu thập trong môi trường tiêu chuẩn có sự kiểm soát với phông nền đơn giản. Sự chênh lệch về phân phối dữ liệu này khiến cho nguồn dữ liệu gán nhãn sẵn có không phản ánh đúng thực tế sản xuất. Hơn nữa, hiện tượng mất cân bằng dữ liệu tự nhiên, nơi số lượng mẫu của các cây khỏe mạnh hoặc các bệnh phổ biến áp đảo hoàn toàn so với các bệnh nguy hiểm nhưng ít gặp, càng khiến cho việc xây dựng một bộ dữ liệu huấn luyện có tính cân bằng và đại diện trở thành một trong những thách thức lớn đối với cộng đồng nghiên cứu.

1.2.2. Tác động của thiếu dữ liệu đến hiệu suất của mô hình

Vấn đề thiếu hụt dữ liệu gán nhãn tác động trực tiếp đến cơ chế học tập của các mạng nơ-ron tích chập sâu, dẫn đến hệ quả kỹ thuật điển hình là hiện tượng quá khớp. Về mặt lý thuyết, các mô hình học sâu hiện đại thường sở hữu hàng triệu tham số tự do, tạo nên một không gian giả thuyết vô cùng lớn để xấp xỉ các hàm phi tuyến phức tạp. Để mô hình có thể hội tụ về một giải pháp tối ưu đại diện cho bản chất của dữ liệu, lượng mẫu huấn luyện cần thiết phải tương xứng với độ phức tạp của mạng. Khi số lượng mẫu huấn luyện quá ít so với năng lực biểu diễn của mô hình nó có xu hướng học thuộc lòng các đặc điểm nhiễu ngẫu nhiên hoặc các chi tiết nền không liên quan thay vì trích xuất các quy luật bệnh học cốt lõi. Trong trường hợp này, mô hình hoạt động như một bộ nhớ máy móc, đạt độ chính xác cao trên dữ liệu huấn luyện nhưng lại thất bại trong việc đưa ra dự đoán đúng đối với các dữ liệu mới chưa nhìn thấy.

Sự hạn chế về dữ liệu ảnh hưởng nghiêm trọng đến biên quyết định giữa các lớp bệnh. Trong không gian đặc trưng đa chiều, khi số lượng mẫu hỗ trợ thưa thớt, mô hình sẽ thiết lập các ranh giới phân loại dựa trên những đặc điểm cục bộ thiếu tính đại diện. Đối với bài toán phân loại bệnh thực vật mô hình có thể nhận diện sai lệch dựa trên các yếu tố ngoại cảnh như màu sắc đất, bóng râm hoặc góc chụp thay

vì dựa trên hình thái vết bệnh thực sự. Ví dụ, nếu tập dữ liệu huấn luyện chỉ chứa hình ảnh lá bệnh đốm đen được chụp trong điều kiện trời nắng, mô hình có nguy cơ cao sẽ gán nhãn sai cho các lá khỏe mạnh được chụp trong điều kiện tương tự, hoặc bỏ sót các lá bệnh được chụp trong điều kiện trời râm. Sự phụ thuộc vào các tương quan giả này làm giảm chất lượng của hệ thống chẩn đoán khi triển khai trong môi trường canh tác thực tế đa dạng.



Hình 3: Underfit, Optimal, Overfit

Khả năng tổng quát hóa kém là hệ quả tất yếu của việc mô hình học trên các dữ liệu hạn chế, khiến mô hình trở nên mong manh trước sự thay đổi của miền dữ liệu. Một hệ thống chẩn đoán lý tưởng cần phải có khả năng trích xuất được các đặc trưng bất biến, nghĩa là các đặc điểm nhận dạng bệnh phải không thay đổi bất chấp các biến thiến về điều kiện thu thập ảnh. Tuy nhiên, khi dữ liệu gán nhãn khan hiếm, mô hình không có đủ cơ hội tiếp xúc với các biến thể đa dạng của cùng một loại bệnh để học được tính bất biến này. Kết quả là hiệu suất của mô hình sẽ sụt giảm mạnh khi có sự chênh lệch phân phối giữa dữ liệu huấn luyện và dữ liệu kiểm thử thực tế. Đây là thách thức lớn nhất ngăn cản việc chuyển giao các mô hình từ phòng thí nghiệm ra tự nhiên, đặt ra yêu cầu cấp thiết về việc tìm kiếm các phương pháp học tập mới có khả năng khai thác tri thức từ nguồn dữ liệu không nhãn để bù đắp cho sự thiếu hụt dữ liệu giám sát.

1.2.3. Bài toán phân loại bệnh thực vật trong điều kiện thiếu dữ liệu gán nhãn

Bài toán phân loại bệnh thực vật trong điều kiện thiếu dữ liệu gán nhãn không chỉ đơn giản là một thách thức về kỹ thuật, mà là một vấn đề nan giải liên quan đến việc tối ưu hóa không gian biểu diễn đặc trưng trong bối cảnh thông tin giám sát thưa thớt. Về bản chất, đây là việc tìm kiếm một hàm ánh xạ phi tuyến từ không gian hình ảnh nhiều chiều sang không gian nhãn bệnh học, sao cho hàm số này vẫn giữ được độ chính xác cao trên các mẫu chưa từng gặp dù chỉ được huấn luyện trên một tập mẫu hỗ trợ cực kỳ hạn chế. Trong học máy thống kê, vấn đề này vi phạm giả định cơ, nơi mà sự phân bố của tập huấn luyện không đủ để đại diện đầy đủ cho phân bố thực tế của quần thể dữ liệu bệnh ngoài tự nhiên. Khi số lượng mẫu gán nhãn không đủ lớn để bao phủ hết các biến thể của không gian đặc trưng, các mô hình học sâu có xu hướng thiết lập các đường bao quyết định dựa trên các đặc điểm nhiễu cục bộ thay vì các đặc trưng bản chất, dẫn đến sự suy giảm về độ tin cậy của hệ thống dự đoán.

Đặc thù sinh học của đối tượng cây trồng làm gia tăng đáng kể độ phức tạp của bài toán này so với các tác vụ nhận diện vật thể thông thường. Các triệu chứng bệnh lý trên lá cây thường biểu hiện sự biến thiên nội lớp rất lớn và sự tương đồng ngoại lớp cao. Cụ thể, cùng một loại bệnh có thể biểu hiện các hình thái tổn thương hoàn toàn khác nhau tùy thuộc vào giai đoạn của bệnh, giống cây hoặc điều kiện dinh dưỡng. Ngược lại, các loại bệnh khác nhau do nấm, vi khuẩn hay virus gây ra lại có thể tạo ra các vết hoại tử có hình dạng và màu sắc rất giống nhau. Trong điều kiện thiếu dữ liệu gán nhãn để cung cấp đủ các ví dụ phản biện cho mô hình, mạng nơ-ron rất khó phân tách rạch ròi các lớp bệnh này trong không gian véc-tơ, dẫn đến hiện tượng nhầm lẫn phổ biến giữa các bệnh có triệu chứng tương đương. Khả năng phân biệt tinh tế này đòi hỏi mô hình phải học được các đặc trưng ngữ nghĩa mức cao, điều mà các phương pháp học giám sát khó đạt được khi thiếu hụt số lượng mẫu cần thiết để hội tụ.

Một khía cạnh khác làm nên tính đặc thù và thách thức của bài toán này là sự phân bố dữ liệu mất cân bằng tự nhiên, dữ liệu được thường tập trung vào một số loại bệnh phổ biến hoặc cây khỏe mạnh, trong khi các mẫu bệnh hiếm hoặc mới phát sinh

lại chiếm tỷ trọng rất nhỏ. Khi áp dụng các thuật toán học giám sát tiêu chuẩn trên các tập dữ liệu mất cân bằng và thiếu nhãn này, mô hình sẽ nảy sinh xu hướng thiên vị đối với các lớp đa số để tối thiểu hóa hàm mất mát tổng thể, dẫn đến việc bỏ sót hoặc phân loại sai các lớp thiểu số. Điều này đặc biệt nguy hiểm trong bảo vệ thực vật vì các bệnh hiếm thường là các bệnh mới lây lan cần được phát hiện sớm. Do đó, bài toán đặt ra không chỉ là nâng cao độ chính xác trung bình mà còn là đảm bảo khả năng phát hiện các lớp dữ liệu hiếm thông qua việc khai thác tối đa thông tin từ các mẫu ít ỏi sẵn có.

1.3. Các nghiên cứu liên quan gần đây

1.3.1. Các phương pháp nghiên cứu truyền thống trước đây

Trong giai đoạn khởi đầu của kỷ nguyên thị giác máy tính nông nghiệp, các thuật toán học máy đóng vai trò chủ đạo định hình quy trình chẩn đoán bệnh tự động. Các phương pháp kinh điển như Máy véc-tơ hỗ trợ, K-láng giềng gần nhất hay Rừng ngẫu nhiên hoạt động dựa trên quy trình hai bước tách biệt: trích xuất đặc trưng thủ công và phân loại. Các nghiên cứu trong giai đoạn này tập trung vào việc thiết kế các bộ lọc để định lượng các thuộc tính thị giác như màu sắc, kết cấu và hình dạng của vết bệnh. Mặc dù đạt được độ chính xác nhất định trong môi trường phòng thí nghiệm được kiểm soát, nhóm phương pháp này bộc lộ hạn chế nghiêm trọng về khả năng tổng quát hóa do phụ thuộc hoàn toàn vào tri thức chuyên gia trong khâu thiết kế đặc trưng. Sự cứng nhắc này khiến mô hình dễ dàng thất bại khi đối mặt với các biến thể hình thái phức tạp hoặc điều kiện nhiễu loạn của môi trường tự nhiên, tạo ra một rào cản khi triển khai thực tế.

Sự ra đời của Mạng nơ-ron tích chập đã đánh dấu bước ngoặt thay đổi hoàn toàn mô thức tiếp cận, chuyển từ trích xuất đặc trưng thủ công sang cơ chế học biểu diễn đầu-cuối. Các kiến trúc mạng sâu như AlexNet, VGG hay ResNet có khả năng “tự động học các đặc trưng phân cấp từ dữ liệu thô, từ các cạnh nét đơn giản” đến các cấu trúc bệnh học phức tạp". Khả năng này đã giúp CNN vượt xa các phương pháp truyền thống trước đây về độ chính xác nhận diện. Tuy nhiên, sức mạnh của CNN đi kèm với một "cơn khát" dữ liệu khổng lồ. Để tối ưu hóa hàng triệu tham số trọng số,

các mô hình này yêu cầu lượng dữ liệu gán nhãn cực lớn. Khi học tập dữ liệu nhỏ hẹp về bệnh thực vật, CNN thường rơi vào trạng thái quá khớp, học thuộc lòng các đặc điểm nhiễu của tập huấn luyện thay vì nắm bắt bản chất của bệnh, dẫn đến hiệu suất suy giảm đáng kể trên tập kiểm thử.

Để khắc phục vấn đề thiếu hụt dữ liệu của CNN thuần túy, phương pháp Học chuyển giao đã trở thành chiến lược tiêu chuẩn trong hầu hết các nghiên cứu hiện đại. Bằng cách khởi tạo trọng số từ các mô hình đã được tiền huấn luyện trước trên các bộ dữ liệu lớn như ImageNet, học chuyển giao cho phép mô hình hội tụ nhanh hơn và đạt hiệu suất cao hơn ngay cả với dữ liệu gán nhãn hạn chế. Mặc dù vậy, phương pháp này không phải là cách giải quyết triệt để vấn đề. Sự chênh lệch về miền dữ liệu giữa các vật thể thông thường trong ImageNet và các đặc điểm sinh học tinh tế của lá cây có thể dẫn đến hiện tượng chuyển giao tiêu cực. Hơn nữa, học chuyển giao vẫn là một phương pháp học giám sát, nghĩa là hiệu quả tinh chỉnh cuối cùng vẫn bị giới hạn bởi chất lượng và số lượng của tập dữ liệu đích có nhãn.

Nhằm nỗ lực đẩy cao giới hạn độ chính xác, các phương pháp Học kết hợp và Học đa phương thức cũng đã được khai thác triệt để. Các nghiên cứu theo hướng này thường kết hợp dự đoán từ nhiều kiến trúc mạng khác nhau hoặc lai ghép giữa trích xuất đặc trưng CNN với các bộ phân loại truyền thống như SVM để tận dụng ưu điểm của từng thành phần. Dù cải thiện được tính ổn định và độ chính xác tổng thể, nhưng cái giá phải trả là sự gia tăng phức tạp trong tính toán và chi phí tài nguyên, làm giảm khả năng triển khai trên các thiết bị biên. Quan trọng hơn, các phương pháp này vẫn chưa giải quyết được gốc rễ là: “làm thế nào để mô hình học được các biểu diễn đặc trưng tốt mà không phụ thuộc quá nhiều vào nhãn dữ liệu con người, vốn là tài nguyên đắt đỏ và khan hiếm nhất” trong bài toán này.

1.3.2. Phương pháp hiện nay

Mặc dù các phương pháp supervised visual plant disease identification đạt hiệu quả cao khi dữ liệu gắn nhãn dồi dào, hiệu năng của chúng suy giảm đáng kể trong các kịch bản thiếu dữ liệu gán nhãn. Để khắc phục những hạn chế này, các nghiên cứu gần đây chuyển sang Contrastive Learning, nhằm học discriminative

representations thông qua việc kéo gần các mẫu cùng ngữ nghĩa và đẩy xa các mẫu khác ngữ nghĩa trong không gian embedding, từ đó cải thiện khả năng tổng quát hóa cho bài toán phân loại bệnh thực vật. Mu et al. [14] đề xuất một framework supervised contrastive few-shot learning (SC-FSL) cho phân loại bệnh thực vật đạt độ chính xác 79.51% chỉ với 30 ảnh huấn luyện trên tập dữ liệu PlantVillage. Pham et al. [15] xuất một kỹ thuật semi-supervised learning kết hợp ambiguity rejection để loại bỏ những pseudo-labels mơ hồ trước khi dùng chúng trong huấn luyện tiếp, giúp tăng tính chính xác của mẫu gán nhãn tự động. Thử nghiệm trên hai bộ dữ liệu coffee và banana leaf disease cho thấy phương pháp vẫn đạt độ chính xác 99.46% với Banana dataset và 97.27% với BRACOL dataset chỉ với 50% labeled data. Trong nghiên cứu [1] đã đề xuất một khung tự giám sát nhẹ sử dụng phương pháp đối chiếu SimCLR để tiền huấn luyện backbone ResNet-50 trên tập dữ liệu ảnh thực địa nông nghiệp chưa gán nhãn nhằm loại bỏ sự phụ thuộc vào dữ liệu có gán nhãn lớn và ứng dụng cho nhiều nhiệm vụ thị giác trong nông nghiệp. Trong một nghiên cứu khác [2] đã áp dụng học đối chiếu tự giám sát lên ảnh nông nghiệp, sử dụng nhiều ảnh không gán nhãn để tiền huấn luyện mạng neural. Kết quả cho thấy phương pháp này vượt trội hơn so với các phương pháp học sâu truyền thống trong các tác vụ phân loại downstream khi lượng ảnh gán nhãn ít, với mức tăng tới 14 % điểm top-1 accuracy trung bình. Kim và cộng sự [3] đã phát triển một mô hình phát hiện bệnh thực vật-aware theo đối tượng bằng cách kết hợp bản đồ saliency và tiền huấn luyện tự giám sát. Họ sử dụng normalizing flows và positional encodings để giảm nhiễu nền và tập trung vào vùng có dấu hiệu bệnh. Kết quả thử nghiệm trên các dataset BRACOL và PlantVillage đạt hiệu suất theo image-level và pixel-level lần lượt là 99.96 % và 99.99 %, cải thiện đáng kể so với không dùng tiền huấn luyện tự giám sát. Al Mamun và cộng sự [4] thực hiện tổng quan hệ thống các nghiên cứu về phát hiện bệnh cây sử dụng học tự giám sát, bao gồm các phương pháp, kiến trúc mạng, dataset và kết quả thực nghiệm. Bài review cung cấp một cái nhìn tổng hợp và cập nhật toàn diện về các ứng dụng SSL trong phát hiện bệnh thực vật, nhằm hỗ trợ cải tiến và phát triển các mô hình hiệu quả khi dữ liệu gán nhãn hạn chế. Fang và cộng sự [5] giới thiệu phương pháp SEED, một học

tự giám sát kết hợp chưng cất kiến thức từ một mạng lớn sang mạng nhỏ. Thay vì học trực tiếp từ dữ liệu không gán nhãn, student encoder được huấn luyện để mô phỏng phân phối điểm tương đồng do teacher suy ra trên một tập thể instances. Phương pháp này mang lại cải thiện lớn về độ chính xác top-1 cho các mạng nhỏ như EfficientNet-B0 (từ 42.2 % lên 67.6 %) và MobileNetV3-Large (từ 36.3 % lên 68.2 %) trên ImageNet-1k.

Việc áp dụng học tự giám sát trên các mô hình mạng nơ-ron tích chập có kích thước lớn, điển hình như ResNet-50, đã chứng minh được hiệu quả vượt trội trong việc trích xuất đặc trưng bệnh học thực vật, đặc biệt trong bối cảnh dữ liệu gán nhãn khan hiếm. Tuy nhiên, khi chuyển sang các kiến trúc mạng nhỏ gọn như EfficientNet hay MobileNet để phục vụ nhu cầu triển khai trên các thiết bị biên, hiệu suất của học tự giám sát thường bị suy giảm nếu không có các kỹ thuật hỗ trợ chuyên biệt. Phương pháp SEED đã giải quyết vấn đề này bằng cơ chế chưng cất kiến thức từ một mạng giáo viên lớn sang mạng học sinh nhỏ, mang lại bước nhảy vọt về độ chính xác cho các mô hình nhẹ. Mặc dù vậy, hạn chế chí tử của hướng tiếp cận này là gánh nặng tài nguyên tính toán khổng lồ phát sinh từ quá trình huấn luyện cặp mô hình giáo viên - học sinh, đi ngược lại với tiêu chí tối ưu hóa tài nguyên cho các ứng dụng nông nghiệp thông minh chi phí thấp. Từ thực tiễn đó, luận văn nhận định rằng đang tồn tại một khoảng trống nghiên cứu quan trọng trong việc tối ưu hóa quy trình học tự giám sát trực tiếp cho các mô hình hạng nhẹ mà không cần thông qua các cơ chế trung gian tốn kém tài nguyên. Việc khai thác hiệu quả học tự giám sát trên các kiến trúc khiêm tốn này không chỉ giúp giải quyết bài toán thiếu hụt dữ liệu gán nhãn mà còn mở ra cơ hội triển khai rộng rãi các hệ thống chẩn đoán bệnh tự động chi phí thấp, phù hợp với điều kiện canh tác thực tế.

1.4. Kết luận chương 1

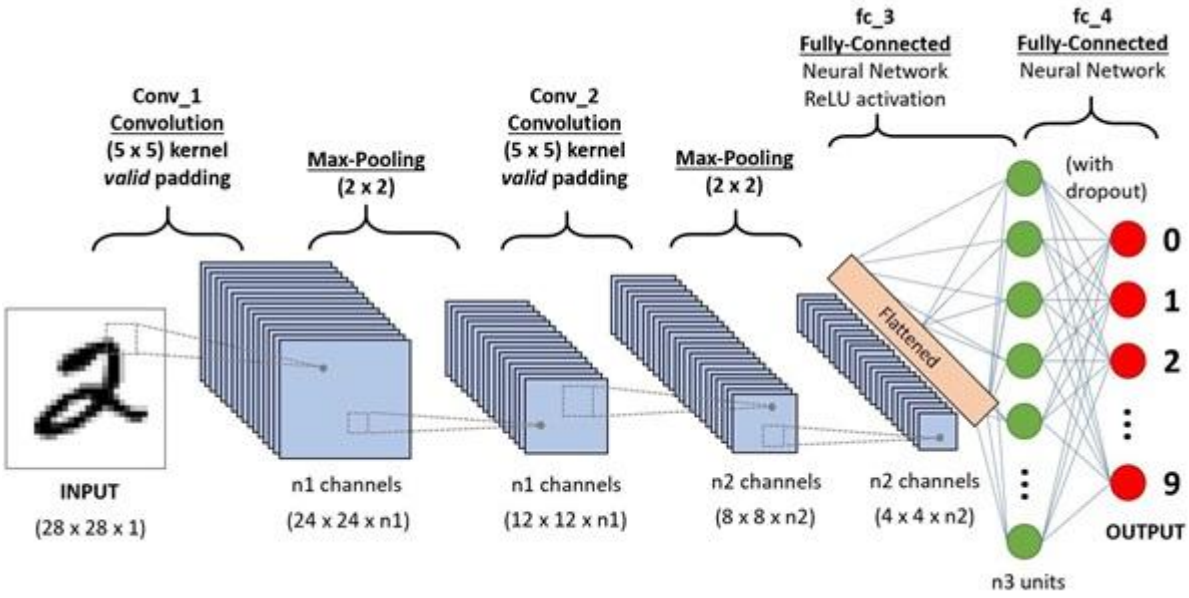
Chương 1 đã cung cấp cái nhìn toàn cảnh về việc ứng dụng thị giác máy tính trong nông nghiệp, đồng thời chỉ ra rào cản lớn nhất hiện nay là sự phụ thuộc vào nguồn dữ liệu gán nhãn thủ công vốn tốn kém và khan hiếm. Các phân tích cho thấy dù học sâu đạt thành tựu lớn, nhưng khi thiếu dữ liệu chất lượng cao, các mô hình

truyền thống thường gặp hiện tượng quá khớp và suy giảm hiệu suất nghiêm trọng trong môi trường thực tế nhiễu loạn. Để khắc phục điều này, học tự giám sát nổi lên như một giải pháp đột phá giúp khai thác nguồn dữ liệu không nhãn khổng lồ, tuy nhiên việc áp dụng hiệu quả kỹ thuật này lên các mô hình hạng nhẹ dành cho thiết bị biên vẫn là một thách thức kỹ thuật chưa được giải quyết triệt để. Từ khoảng trống nghiên cứu đó, luận văn xác định mục tiêu xây dựng một quy trình học tự giám sát tối ưu cho các kiến trúc mạng nhỏ gọn, giúp cân bằng giữa độ chính xác và chi phí tài nguyên. Các nền tảng lý thuyết cần thiết để hiện thực hóa mục tiêu này sẽ được tiếp tục trình bày chi tiết trong Chương 2 kế tiếp.

CHƯƠNG 2: CƠ SỞ LÝ THUYẾT

2.1. Các thành phần chính của mạng nơ-ron tích chập

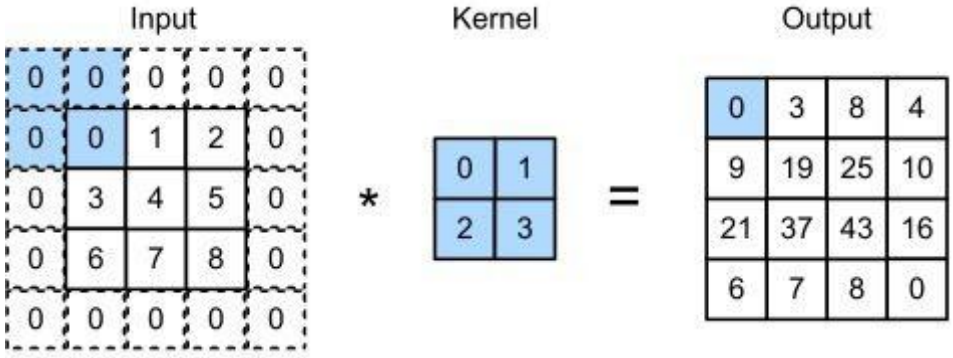
Một mạng nơ-ron tích chập bao gồm lần lượt một lớp đầu vào, nhiều lớp ẩn, và một lớp đầu ra. Lớp đầu vào sẽ khác nhau tùy theo các ứng dụng cụ thể. Các lớp ẩn, được coi là khối lõi của kiến trúc CNN, bao gồm một chuỗi các lớp tích chập, lớp gộp và cuối cùng xuất đầu ra thông qua lớp kết nối đầy đủ.



Hình 4: Các thành phần cơ bản của mạng nơ-ron tích chập

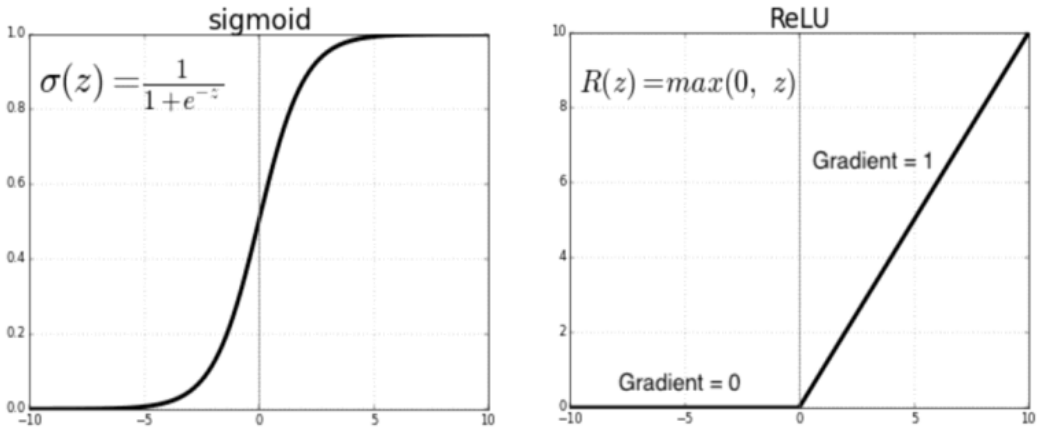
Lớp tích chập được xem là khối nền tảng quan trọng nhất của cấu trúc của mạng nơ-ron tích chập, giữ vai trò trung tâm trong việc chuyển đổi hình ảnh đầu vào thành các biểu diễn đặc trưng có ý nghĩa. Cụ thể, một ảnh đầu vào với kích thước xác định được ánh xạ thành các bản đồ đặc trưng thông qua quá trình tích chập giữa ảnh và một tập các bộ lọc có thể học được, trong đó mỗi bộ lọc là một ma trận nhỏ đại diện cho một mẫu đặc trưng cục bộ của ảnh. Các bộ lọc này lần lượt trượt trên toàn bộ không gian ảnh với một bước trượt cố định, tại mỗi vị trí thực hiện phép nhân tương ứng từng phần tử và phép cộng tổng để tạo ra một giá trị đầu ra duy nhất, từ đó hình thành nên bản đồ đặc trưng. Toàn bộ các phần tử trong bộ lọc, bao gồm trọng số và độ chệch, đều là các tham số của mạng và không được thiết lập thủ công, mà được tự động học thông qua quá trình lan truyền ngược dựa trên dữ liệu huấn luyện. Nhờ

cơ chế “kết nối cục bộ và chia sẻ trọng số”, lớp tích chập cho phép mạng nơ-ron học được các đặc trưng hình ảnh hiệu quả với số lượng tham số hạn chế, đồng thời nâng cao khả năng khái quát hóa so với các mạng kết nối đầy đủ truyền thống, như đã được chứng minh bởi các nghiên cứu nền tảng về mạng CNN của LeCun và cộng sự [6] cũng như các tổng quan hiện đại về học sâu trong xử lý ảnh.



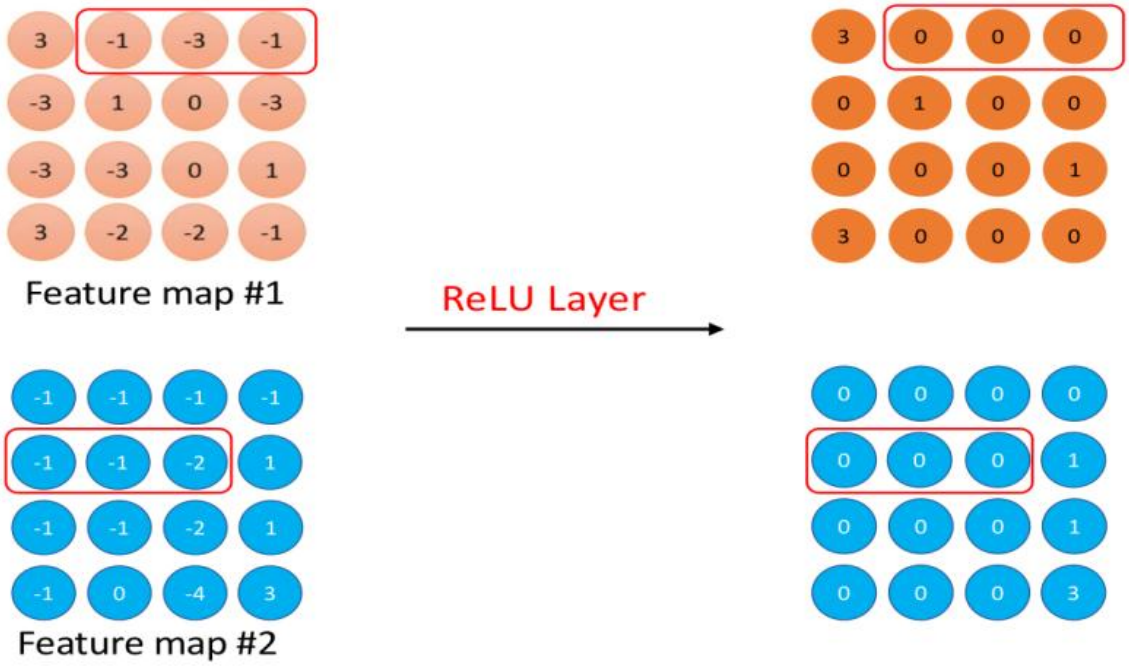
Hình 5: Lớp tích chập

Trong mạng nơ-ron tích chập, việc chèn lớp kích hoạt phi tuyến như ReLU sau mỗi lớp tích chập là cần thiết nhằm khắc phục hạn chế tuyến tính của phép tích chập. Bản chất của lớp tích chập chỉ bao gồm các phép nhân và cộng nên vẫn là một phép biến đổi tuyến tính; do đó, nếu xếp chồng nhiều lớp tích chập liên tiếp mà không có phi tuyến hóa, toàn bộ mạng về mặt toán học vẫn tương đương với một phép biến đổi tuyến tính duy nhất, dẫn đến khả năng biểu diễn bị suy giảm nghiêm trọng. Lớp ReLU giúp phá vỡ tính tuyến tính này bằng cách loại bỏ các giá trị âm và giữ lại các kích hoạt dương, qua đó cho phép mạng học được các quan hệ phi tuyến phức tạp giữa các đặc trưng hình ảnh. Ngoài ra, ReLU còn góp phần cải thiện quá trình huấn luyện mạng sâu nhờ đặc tính gradient đơn giản, giúp giảm hiện tượng gradient biến mất và hỗ trợ mạng hội tụ ổn định hơn, đặc biệt quan trọng trong các bài toán phân loại ảnh có độ phức tạp cao như nhận diện bệnh thực vật trong điều kiện thiếu dữ liệu gán nhãn. Hàm kích hoạt, thực hiện một phép biến đổi phi tuyến tính, đóng vai trò quan trọng trong CNN để quyết định xem một nơ-ron có nên được kích hoạt hay bỏ qua. Có nhiều loại hàm kích hoạt được sử dụng sau lớp tích chập, chẳng hạn như hàm hyperbolic và hàm sigmoid, trong đó ReLU là hàm kích hoạt được sử dụng phổ biến nhất trong mạng nơ-ron, đặc biệt là trong CNN.



Hình 6: Đồ thị hàm sigmoid và ReLU

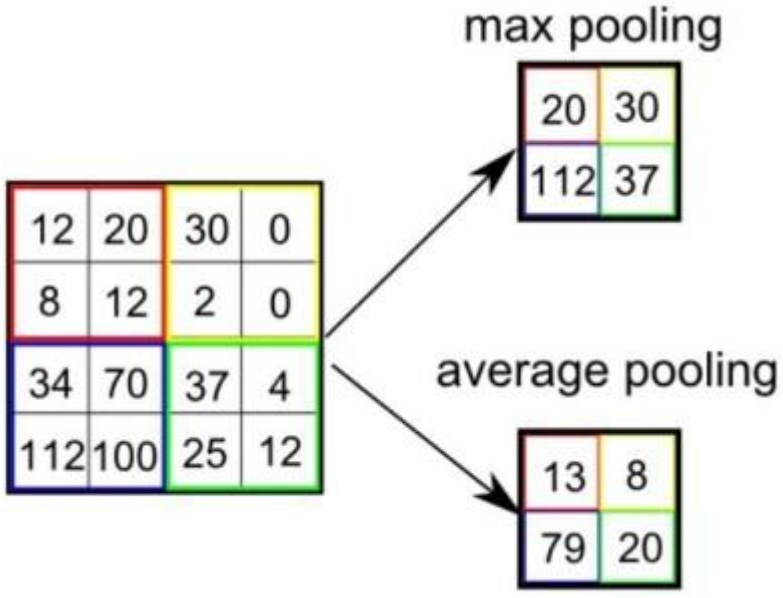
Hình 6 minh họa một phiên bản đơn giản hóa của lớp ReLU. Mỗi phần tử đơn lẻ trong các bản đồ đặc trưng, được xác định từ lớp tích chập trước đó, sẽ được tính toán thêm bởi hàm kích hoạt ReLU trong lớp này. Cụ thể, tất cả các giá trị dương sẽ được giữ nguyên, trong khi các giá trị âm được thay thế bằng 0. Đầu ra sau lớp ReLU, có cấu trúc mạng giống hệt với bản đồ đặc trưng từ lớp tích chập trước, sẽ được sử dụng làm đầu vào cho lớp tích chập tiếp theo.



Hình 7: Cách ReLU hoạt động

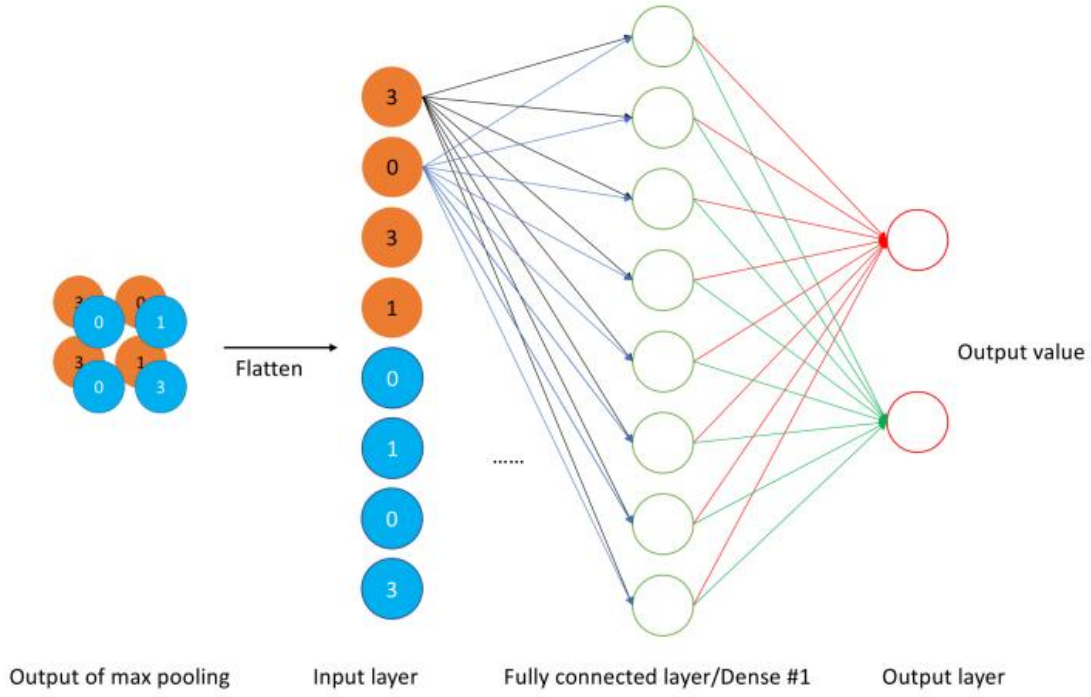
Lớp gộp là một thành phần vô cùng quan trọng trong việc xây dựng mạng nơ-ron tích chập, được sử dụng nhằm giảm dần kích thước không gian của các bản đồ đặc trưng sinh ra từ các lớp tích chập trước đó, đồng thời làm nổi bật những đặc trưng quan trọng nhất của ảnh. Bằng cách tổng hợp thông tin cục bộ trong từng vùng nhỏ, lớp gộp giúp mạng trở nên ít nhạy cảm hơn với các biến đổi nhỏ về vị trí của đặc trưng trong không gian ảnh. Trong thực tế, nhiều phép gộp khác nhau đã được đề xuất, bao gồm gộp trung bình, gộp chuẩn L2 và gộp giá trị lớn nhất, trong đó gộp giá trị lớn nhất được sử dụng phổ biến hơn cả là do khả năng giữ lại các phản hồi đặc trưng mạnh mẽ nhất trong mỗi vùng con của bản đồ đặc trưng.

Nguyên lý hoạt động của gộp giá trị lớn nhất dựa trên giả định rằng vị trí chính xác tuyệt đối của một đặc trưng không quan trọng bằng sự tồn tại tương đối của đặc trưng đó trong một vùng lân cận. Cụ thể, mỗi bản đồ đặc trưng đầu vào được chia thành các vùng con không chồng lấn với kích thước cố định, và trong mỗi vùng con, chỉ giá trị lớn nhất được giữ lại để tạo thành đầu ra của lớp gộp. Quá trình này dẫn đến việc giảm đáng kể kích thước không gian của bản đồ đặc trưng, ví dụ từ kích thước 4×4 xuống còn 2×2, trong khi vẫn bảo toàn các phản hồi đặc trưng quan trọng nhất. Nhờ đó, mạng có khả năng xây dựng các biểu diễn bất biến hơn đối với các dịch chuyển nhỏ trong ảnh đầu vào.



Hình 8: Max pooling và average pooling

Việc đưa lớp gộp giá trị lớn nhất vào mạng nơ-ron tích chập mang lại nhiều lợi ích thực tiễn. Trước hết, do kích thước không gian của bản đồ đặc trưng bị thu nhỏ, số lượng phép tính và tham số trong các lớp phía sau được giảm xuống, từ đó làm giảm độ phức tạp tính toán và cải thiện hiệu quả huấn luyện. Bên cạnh đó, bằng cách chỉ giữ lại các phản hồi nổi trội nhất trong từng vùng, lớp gộp góp phần hạn chế hiện tượng quá khớp, khi mô hình quá phụ thuộc vào các chi tiết cục bộ hoặc nhiễu trong dữ liệu huấn luyện. Đặc tính này đặc biệt quan trọng trong các bài toán phân loại ảnh bệnh thực vật, nơi dữ liệu gán nhãn thường hạn chế và mô hình cần có khả năng khái quát tốt để nhận diện các biểu hiện bệnh trong những điều kiện hình ảnh đa dạng.



Hình 9: Mô phỏng mạng kết nối đầy đủ

Lớp "kết nối đầy đủ" thường được đặt ở các tầng cuối của mạng nơ-ron tích chập, sau các lớp tích chập và lớp gộp, với mục tiêu tổng hợp và chuyển đổi các đặc trưng đã được trích xuất thành thông tin có tính phân biệt phục vụ cho bài toán phân loại. Trong lớp này, mỗi nơ-ron được kết nối trực tiếp với toàn bộ các nơ-ron của lớp trước đó, cho phép mạng tích hợp đồng thời thông tin từ tất cả các bản đồ đặc trưng

đã học được ở các tầng thấp hơn. Thông qua cơ chế kết nối toàn phần, lớp "kết nối đầy đủ" đóng vai trò như một "bộ phân loại", ánh xạ các biểu diễn đặc trưng mức cao sang không gian nhãn đầu ra. Để nâng cao khả năng học và hiệu quả hội tụ của mạng, các nơ-ron trong lớp "kết nối đầy đủ" thường sử dụng hàm kích hoạt ReLU, giúp đưa tính phi tuyến vào mô hình và hỗ trợ mạng học được các ranh giới phân lớp phức tạp, từ đó cải thiện hiệu suất phân loại trong các ứng dụng xử lý ảnh, bao gồm cả bài toán nhận dạng bệnh thực vật trong điều kiện dữ liệu gán nhãn hạn chế.

2.2. Các mô hình CNN nổi tiếng phù hợp với bài toán phân loại bệnh thực vật.

2.2.1. Quá trình huấn luyện mô hình học sâu

Quy trình phát triển và tối ưu hóa các thuật toán học sâu cho bài toán phân loại bệnh thực vật đòi hỏi một chuỗi các bước kỹ thuật chặt chẽ nhằm đảm bảo mô hình có khả năng chẩn đoán chính xác các loại bệnh từ dữ liệu hình ảnh. Quá trình này khởi đầu bằng việc thu thập và chuẩn bị dữ liệu từ đa dạng các nguồn khác nhau, đảm bảo các mẫu bệnh được gán nhãn chính xác bởi các chuyên gia nông nghiệp hoặc bệnh học thực vật. Dữ liệu thô sau đó được chuẩn hóa về kích thước, độ sáng và độ tương phản để đồng nhất đầu vào cho mạng nơ-ron. Đặc biệt trong bối cảnh thiếu hụt dữ liệu gán nhãn, kỹ thuật tăng cường dữ liệu đóng vai trò then chốt thông qua các thao tác như xoay ảnh, lật ảnh, thay đổi độ sáng hoặc thêm nhiễu nhằm làm phong phú tập huấn luyện và giúp mô hình học được đa dạng các biến thể của vết bệnh trong điều kiện thực tế.

Giai đoạn thiết kế và huấn luyện mô hình tập trung vào việc lựa chọn các kiến trúc tiên tiến như ResNet hoặc Vision Transformer kết hợp với các lớp kết nối đầy đủ cho tác vụ phân loại. Quá trình này đòi hỏi sự tinh chỉnh tỉ mỉ các siêu tham số như tốc độ học, kích thước lô, đồng thời áp dụng các kỹ thuật chuẩn hóa theo lô và loại bỏ ngẫu nhiên để ngăn ngừa hiện tượng quá khớp. Hàm mất mát được sử dụng thường là entropy chéo đa lớp phù hợp với bài toán phân loại nhiều loại bệnh thực vật khác nhau. Trong suốt quá trình huấn luyện, các chỉ số hiệu năng như độ chính xác, độ nhạy, F1-score cần được theo dõi liên tục để đánh giá khả năng hội tụ của mô hình.

Các chiến lược như dừng sớm và lưu trữ mô hình tốt nhất được áp dụng để đảm bảo kết quả tối ưu trên tập dữ liệu kiểm định.

Sau quá trình huấn luyện, mô hình phải được đánh giá một cách toàn diện và tối ưu hóa trước khi xem xét khả năng triển khai thực tế. Việc phân tích chi tiết các chỉ số đánh giá và ma trận nhầm lẫn giúp nhận diện các trường hợp dự đoán sai, từ đó thực hiện tinh chỉnh các siêu tham số hoặc thử nghiệm các kiến trúc mạng khác nhau nhằm cải thiện độ chính xác. Bước xác thực trên dữ liệu thực địa là giai đoạn quan trọng để kiểm chứng độ tin cậy của hệ thống trong các tình huống canh tác thực tế, so sánh kết quả dự đoán với đánh giá của chuyên gia. Cuối cùng, mô hình được đóng gói và triển khai thông qua các giao diện người dùng, đi kèm với cơ chế theo dõi hiệu suất liên tục để cập nhật và cải thiện hệ thống.

2.2.2. Các mô hình CNN nổi tiếng phù hợp

Trong bối cảnh ứng dụng thực tiễn của nông nghiệp thông minh, các hệ thống chẩn đoán bệnh cây trồng thường xuyên phải vận hành tại những khu vực canh tác xa xôi, nơi hạ tầng mạng thiếu ổn định và nguồn năng lượng bị giới hạn. Do đó, yêu cầu đặt ra đối với mô hình học sâu không chỉ dừng lại ở độ chính xác cao mà còn phải đảm bảo tính hiệu quả về mặt tính toán và kích thước lưu trữ để có thể triển khai trơn tru trên các thiết bị biên như điện thoại, máy tính nhúng hay các thiết bị bay không người lái. Xuất phát từ thực tiễn này, luận văn tập trung nghiên cứu và lựa chọn các kiến trúc mạng tích chập được thiết kế cho việc tối ưu hóa tài nguyên, tiêu biểu là các dòng mô hình MobileNet, EfficientNet và các biến thể gọn nhẹ của kiến trúc ResNet. Đây là các "ứng viên sáng giá" để đảm nhận vai trò mạng xương sống trong khung làm việc học tự giám sát, giúp cân bằng giữa việc trích xuất đặc trưng mạnh mẽ và tốc độ suy diễn thời gian thực.

Kiến trúc MobileNet được xem là tiêu chuẩn vàng cho các ứng dụng trên thiết bị di động nhờ việc giới thiệu khái niệm tích chập tách biệt theo chiều sâu. Thay vì thực hiện phép tính tích chập tiêu chuẩn tốn kém tài nguyên, MobileNet chia nhỏ quá trình này thành hai bước riêng biệt là tích chập theo chiều sâu và tích chập điểm, giúp giảm thiểu đáng kể số lượng tham số và khối lượng tính toán mà không làm suy giảm

quá nhiều độ chính xác. Phiên bản MobileNetV2 còn cải tiến hơn nữa với cấu trúc dư thừa đảo ngược và các khối nút thắt cổ chai tuyến tính, cho phép bảo toàn thông tin phong phú trong các không gian chiều thấp. Khi được sử dụng làm bộ mã hóa cho học tự giám sát, MobileNet thể hiện ưu thế vượt trội trong việc tạo ra các véc-tơ đặc trưng gọn nhẹ, cho phép quá trình huấn luyện và tinh chỉnh diễn ra nhanh chóng ngay cả khi tài nguyên phần cứng hạn chế, đáp ứng hoàn hảo nhu cầu chẩn đoán bệnh tức thời tại tự nhiên.

Bên cạnh dòng mô hình tối ưu tốc độ, kiến trúc EfficientNet mang lại một cách tiếp cận đột phá trong việc tìm kiếm sự cân bằng giữa độ chính xác và hiệu suất thông qua cơ chế mở rộng kích thước mạng hỗn hợp. Thay vì mở rộng mạng một cách thủ công và rời rạc theo chiều sâu, chiều rộng hay độ phân giải, EfficientNet sử dụng một hệ số tổng hợp để điều chỉnh đồng bộ cả ba yếu tố này dựa trên năng lực tính toán sẵn có. Đặc biệt, các phiên bản nhỏ như EfficientNet-B0 hay B1 cung cấp khả năng trích xuất đặc trưng râts tốt so với các mô hình cùng kích thước nhờ vào việc tối ưu hóa kiến trúc bằng tìm kiếm kiến trúc mạng nơ-ron. Trong bài toán phân loại bệnh thực vật với dữ liệu thiếu nhãn, khả năng biểu diễn đặc trưng mạnh mẽ của EfficientNet giúp mô hình nắm bắt được các chi tiết bệnh lý tinh tế và đa dạng về tỷ lệ, đồng thời kiến trúc này cũng tương thích tốt với các thuật toán học tương phản để tạo ra không gian biểu diễn chất lượng cao.

Mặc dù xu hướng tối ưu hóa trọng lượng đang chiếm ưu thế, kiến trúc Mạng dư thừa ResNet vẫn đóng vai trò là xương sống tiêu chuẩn và ổn định nhất trong các nghiên cứu về học tự giám sát hiện đại. Tuy nhiên, thay vì sử dụng các phiên bản quá sâu và nặng nề, luận văn hướng sự quan tâm đến các biến thể gọn nhẹ như ResNet-10 hoặc ResNet-18. Cơ chế kết nối tắt đặc trưng của kiến trúc này giúp giải quyết triệt để vấn đề tiêu biến đạo hàm, cho phép tín hiệu huấn luyện lan truyền xuyên suốt qua các tầng mạng. Sự ổn định này là yếu tố then chốt giúp các thuật toán học tự giám sát như MoCo hay SimCLR hội tụ tốt, trích xuất được các đặc trưng phân cấp từ các chi tiết gân lá vi mô đến cấu trúc hình thái vĩ mô. Việc lựa chọn ResNet làm cơ sở đối chiếu không chỉ đảm bảo tính tin cậy khoa học nhờ sự phổ biến rộng rãi của nó trong

cộng đồng nghiên cứu, mà còn cung cấp một nền tảng vững chắc để đánh giá hiệu quả cải thiện của phương pháp học tự giám sát đề xuất so với các phương pháp học giám sát truyền thống trên cùng một kiến trúc mạng.

Tuy nhiên, việc lựa chọn được một kiến trúc mạng nơ-ron tối ưu về tài nguyên và hiệu suất mới chỉ giải quyết được một phần bài toán. Sức mạnh thực sự của các mô hình CNN kể trên chỉ có thể được khai phóng khi chúng được dẫn dắt bởi một cơ chế huấn luyện phù hợp, đặc biệt trong bối cảnh khắc nghiệt về sự thiếu hụt thông tin định danh. Khi các phương pháp học giám sát truyền thống trở nên kém hiệu quả do sự phụ thuộc vào nhãn dữ liệu thủ công, yêu cầu cấp thiết đặt ra là phải tìm kiếm một mô thức học tập mới cho phép các mạng backbone này tự chủ động khai thác tri thức từ nguồn dữ liệu thô khổng lồ. Đây chính là tiền đề quan trọng dẫn dắt đến nội dung nghiên cứu về học tự giám sát, một hướng tiếp cận đột phá giúp biến đổi cách thức các mô hình thị giác máy tính hình thành không gian biểu diễn đặc trưng, sẽ được trình bày chi tiết về nguyên lý và các kỹ thuật cốt lõi trong phần tiếp theo.

2.3. Học tự giám sát: nguyên lý và kiến trúc chung và các kĩ thuật liên quan

2.3.1. Định nghĩa

Học tự giám sát được định nghĩa là một mô thức học máy tiên tiến nằm ở giao điểm giữa học có giám sát và học không giám sát, trong đó hệ thống tự động sinh ra các tín hiệu giám sát từ chính cấu trúc nội tại của dữ liệu đầu vào mà không cần sự can thiệp gán nhãn của con người. Khác với học không giám sát thuần túy vốn tập trung vào việc tìm kiếm ra các cấu trúc ẩn hoặc mật độ phân phối của dữ liệu như gom cụm hay giảm chiều dữ liệu, học tự giám sát hướng tới mục tiêu trích xuất các biểu diễn đặc trưng giàu ngữ nghĩa và có khả năng chuyển giao. Bản chất cốt lõi của phương pháp này là việc chuyển đổi từ bài toán không giám sát thành một bài toán có giám sát thực nghiệm bằng cách che giấu hoặc biến đổi một phần thông tin của dữ liệu gốc và yêu cầu mạng nơ-ron phải tái tạo hoặc dự đoán lại phần thông tin đó dựa trên phần còn lại.

Trong khuôn khổ lý thuyết thông tin, học tự giám sát có thể được hiểu là quá trình tối đa hóa lượng thông tin tương hỗ giữa đầu vào và biểu diễn nội bộ của mô

hình, đồng thời loại bỏ các yếu tố nhiễu không quan trọng. Để thực hiện điều này, quy trình học tập thường được chia thành hai giai đoạn riêng biệt nhưng "liên kết chặt chẽ" với nhau. Giai đoạn đầu tiên là tiền huấn luyện, nơi mô hình giải quyết một nhiệm vụ được thiết kế đặc biệt gọi là tác vụ tiền đề. Trong tác vụ này, các nhãn giả được sinh ra hoàn toàn tự động dựa trên các thuộc tính của dữ liệu như quan hệ không gian, thời gian hoặc màu sắc. Giai đoạn thứ hai là tinh chỉnh, trong đó các đặc trưng đã học được từ tác vụ tiền đề sẽ được chuyển giao và áp dụng vào nhiệm vụ mục tiêu thực tế hay còn gọi là tác vụ hạ nguồn, thường là các bài toán phân loại hoặc phát hiện đối tượng với số lượng dữ liệu gán nhãn hạn chế.

Sự khác biệt căn bản giữa học tự giám sát và học có giám sát nằm ở nguồn gốc của tín hiệu lỗi dùng để cập nhật trọng số mạng. Trong học có giám sát, tín hiệu lỗi được tính toán dựa trên sự sai lệch dự đoán của mô hình và nhãn thủ công do con người cung cấp, do đó hiệu suất bị giới hạn bởi tri thức và số lượng nhãn của chuyên gia. Ngược lại, học tự giám sát tính toán tín hiệu lỗi dựa trên khả năng của mô hình trong việc khôi phục hoặc phân biệt các phần của dữ liệu gốc. Ví dụ, trong bài toán phân loại bệnh thực vật, thay vì dạy mô hình phân biệt ngay lập tức giữa lá bệnh và lá lành, học tự giám sát có thể yêu cầu mô hình học cách ghép các mảnh cắt của một chiếc lá lại với nhau hoặc phân biệt hai khung hình khác nhau của cùng một chiếc lá so với các lá khác. Thông qua quá trình này, mạng nơ-ron buộc phải học được các đặc trưng cấp cao về hình thái, kết cấu và cấu trúc của lá cây, tạo tiền đề vững chắc cho việc phân loại chính xác sau này.

Nhìn nhận một cách tổng quát, học tự giám sát đại diện cho "một bước tiến hóa quan trọng" trong khả năng nhận thức của máy tính, chuyển từ việc học vẹt dựa trên đáp án có sẵn sang việc học hiểu thông qua quan sát và suy luận về mối tương quan giữa các dữ liệu. Đây là "chìa khóa" để giải quyết bài toán nút thắt cổ chai về dữ liệu gán nhãn, cho phép các hệ thống trí tuệ nhân tạo tận dụng nguồn tài nguyên dữ liệu số khổng lồ đang tồn tại trong thực tế để xây dựng các mô hình bền vững và hiệu quả hơn.

2.3.2. Phân loại

Dựa trên "bản chất" của hàm mục tiêu và cơ chế tương tác với dữ liệu, các phương pháp học tự giám sát trong thị giác máy tính hiện đại có thể được chia thành ba nhóm chính là phương pháp tạo sinh, phương pháp tương phản và phương pháp điều chỉnh mô hình hay còn gọi là học không tương phản. Mỗi nhóm phương pháp tiếp cận bài toán trích xuất đặc trưng theo một triết lý riêng biệt, từ việc tái tạo lại các chi tiết mức điểm ảnh cho đến việc phân biệt các thực thể ở mức ngữ nghĩa cao trong không gian tiềm ẩn.

Nhóm phương pháp đầu tiên và mang tính nền tảng là các mô hình tạo sinh (Generative Methods). Tư tưởng chủ đạo của hướng tiếp cận này là yêu cầu mạng nơ-ron tái tạo lại các phần dữ liệu bị che khuất hoặc bị biến đổi từ dữ liệu đầu vào. Các kỹ thuật điển hình trong nhóm này bao gồm Autoencoders và các biến thể như Variational Autoencoders hay GANs, trong đó mô hình học cách nén dữ liệu vào một không gian biểu diễn nhỏ gọn và sau đó giải mã ngược lại thành hình ảnh gốc. Một hướng phát triển khác của nhóm này là sử dụng các tác vụ tiền đề dựa trên ngữ cảnh như tô màu ảnh xám (Colorization), khôi phục phần ảnh bị khuyết (In-painting) hoặc sắp xếp lại các mảnh ghép xáo trộn (Jigsaw Puzzles). Mặc dù các phương pháp tạo sinh giúp mô hình học được phân phối của dữ liệu rất tốt, nhưng nhược điểm lớn của chúng là chi phí tính toán cao do phải xử lý từng điểm ảnh và sự lãng phí tài nguyên vào việc mô phỏng các chi tiết nhiễu tần số cao không cần thiết cho tác vụ phân loại. Gần đây, phương pháp Masked Image Modeling với đại diện là Masked Autoencoders (MAE) đã làm sống lại hướng tiếp cận này bằng cách che đi tỷ lệ lớn các vùng ảnh và buộc mô hình tái tạo lại, cho thấy hiệu quả vượt trội khi kết hợp với kiến trúc Vision Transformers, tuy nhiên độ hiệu quả khi áp dụng với các mô hình CNN chưa hiệu quả.

Nhóm phương pháp thứ hai và cũng là trọng tâm mà luận văn nghiên cứu là các phương pháp học tương phản (Contrastive Learning). Thay vì cố gắng tái tạo lại hình ảnh đầu vào, học tương phản tập trung vào việc học một không gian biểu diễn đặc trưng nơi các biến thể khác nhau của cùng một hình ảnh (cặp mẫu dương) được

kéo lại gần nhau, trong khi các hình ảnh khác nhau (cặp mẫu âm) bị đẩy ra xa. Kỹ thuật SimCLR là một ví dụ điển hình cho hướng đi này khi sử dụng độ lớn lô dữ liệu cực đại cùng các phép tăng cường dữ liệu mạnh mẽ để tối ưu hóa hàm mất mát InfoNCE; tuy nhiên, sự phụ thuộc vào kích thước lô lớn của nó tạo ra rào cản đáng kể về tài nguyên phần cứng. Để giải quyết vấn đề đó, kiến trúc Momentum Contrast (MoCo) đã được đề xuất với việc sử dụng một hàng đợi từ điển động và cơ chế cập nhật trọng số động lượng cho bộ mã hóa khóa, cho phép tách biệt kích thước lô khỏi số lượng mẫu âm. Đặc biệt, nghiên cứu này tập trung khai thác phiên bản cải tiến MoCo v2, một kiến trúc lai ghép tối ưu kế thừa cơ chế bộ nhớ của MoCo đồng thời tích hợp các cải tiến của SimCLR bao gồm việc thay thế lớp tuyến tính bằng đầu chiếu phi tuyến (MLP Projection Head) và áp dụng chiến lược tăng cường dữ liệu bất đối xứng với kỹ thuật làm mờ Gaussian (Gaussian Blur). Sự kết hợp này không chỉ giúp mô hình ngăn chặn hiện tượng học vẹt dựa trên phân phối màu sắc cục bộ mà còn cho phép trích xuất các đặc trưng phân biệt mạnh mẽ và bền vững ngay cả trên các hệ thống hạn chế tài nguyên, tạo nên nền tảng kỹ thuật vững chắc để giải quyết bài toán thiếu dữ liệu gán nhãn trong luận văn này.

Nhóm phương pháp cuối cùng là học không tương phản hay các phương pháp điều chỉnh mô hình (Distillation/Non-Contrastive Methods), ra đời nhằm khắc phục sự phức tạp của việc quản lý các mẫu âm của học tương phản. Các kỹ thuật tiêu biểu như BYOL (Bootstrap Your Own Latent) hay SimSiam đã chứng minh rằng việc sử dụng các cặp mẫu âm là không bắt buộc để tránh hiện tượng sụp đổ mô hình (model collapse). Thay vào đó, các phương pháp này sử dụng kiến trúc mạng bất đối xứng, trong đó một mạng giáo viên cung cấp mục tiêu hồi quy cho mạng học sinh. BYOL sử dụng cơ chế cập nhật trung bình trượt cho mạng giáo viên tương tự như MoCo nhưng loại bỏ hoàn toàn hàng đợi mẫu âm, trong khi SimSiam đơn giản hóa hơn nữa bằng cách sử dụng thao tác dừng gradient (stop-gradient) ở một nhánh để ngăn chặn sự đồng nhất hóa trọng số. Mặc dù đạt hiệu suất cao và đơn giản hóa quy trình huấn luyện, các phương pháp này thường đòi hỏi sự tinh chỉnh kỹ lưỡng về các siêu tham

số tối ưu hóa và đôi khi kém ổn định hơn so với các phương pháp tương phản chuẩn mực khi áp dụng trên các miền dữ liệu chuyên biệt như bệnh học thực vật.

2.4. Kết luận chương 2

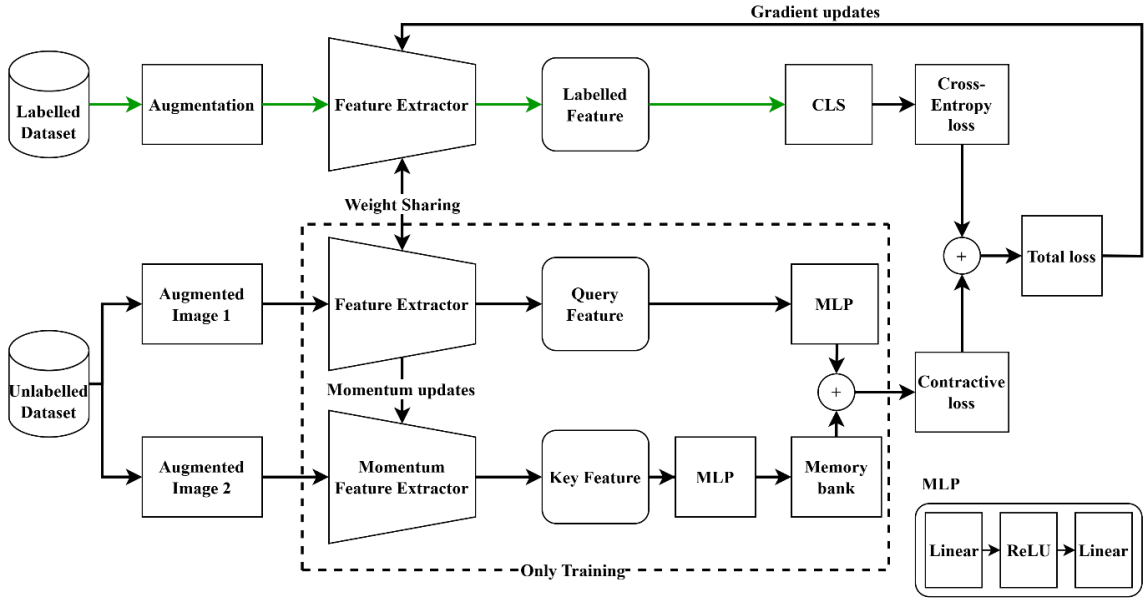
Chương 2 đã hệ thống hóa toàn diện cơ sở lý thuyết về mạng nơ-ron tích chập và phương pháp học tự giám sát, thiết lập nền tảng kỹ thuật vững chắc để giải quyết bài toán phân loại bệnh thực vật trong bối cảnh khan hiếm dữ liệu gán nhãn. Nội dung chương đi sâu phân tích các thành phần cốt lõi của CNN và khẳng định vai trò thiết yếu của các kiến trúc hạng nhẹ như MobileNet, EfficientNet hay ResNet-10t trong việc đáp ứng nhu cầu triển khai thực tế trên các thiết bị biên hạn chế tài nguyên. Bên cạnh đó, chương này cũng làm rõ cơ chế hoạt động của học tự giám sát, đồng thời chứng minh ưu thế của phương pháp học tương phản, tiêu biểu là MoCo v2, so với các phương pháp tạo sinh trong việc trích xuất đặc trưng ngữ nghĩa mạnh mẽ mà không đòi hỏi chi phí tính toán quá lớn. Tuy nhiên, tổng quan nghiên cứu cũng chỉ ra một khoảng trống quan trọng khi việc áp dụng trực tiếp SSL lên các mạng nhỏ thường dẫn đến suy giảm hiệu suất nếu thiếu các chiến lược tối ưu hóa chuyên biệt như chưng cất kiến thức hoặc cơ chế động lượng. Từ những phân tích đó, luận văn xác định hướng tiếp cận là xây dựng một kiến trúc lai ghép, kết hợp mạng xương sống hạng nhẹ với cơ chế cập nhật động lượng và hàng đợi bộ nhớ để cân bằng giữa độ chính xác và hiệu quả tài nguyên. Những nhận định lý thuyết này đóng vai trò là tiền đề then chốt để dẫn dắt sang Chương 3.

CHƯƠNG 3: ĐỀ XUẤT MÔ HÌNH ÁP DỤNG SSL CHO BÀI TOÁN PHÂN LOẠI BỆNH THỰC VẬT TRONG ĐIỀU KIỆN THIẾU DỮ LIỆU GÁN NHÃN

3.1. Đề xuất kiến trúc mô hình

Trong bối cảnh bài toán phân loại bệnh thực vật, thách thức lớn nhất đối với các mô hình học sâu hiện đại không nằm ở khả năng thiết kế kiến trúc mạng phức tạp, mà là vấn đề khan hiếm dữ liệu gán nhãn chất lượng cao. Các phương pháp học giám sát truyền thống khi đối mặt với việc hạn chế dữ liệu thường dễ rơi vào trạng thái quá khớp, dẫn đến khả năng tổng quát hóa kém trên thực tế. Để giải quyết nút thắt này, luận văn đề xuất một kiến trúc lai ghép, tích hợp song song hai cơ chế luồng học tự giám sát và luồng học giám sát. Mục tiêu cốt lõi của thiết kế này là tận dụng tối đa thông tin từ các mẫu dữ liệu, buộc mô hình phải học được các đặc trưng bất biến và mạnh mẽ thông qua việc giải quyết bài toán học tương phản thay vì chỉ phụ thuộc vào nhãn phân loại. Cơ chế này đóng vai trò như một bộ điều hướng, giúp Backbone trích xuất đặc trưng tốt hơn, từ đó cải thiện đáng kể độ chính xác phân loại ngay cả khi số lượng mẫu gán nhãn đầu vào là rất khiêm tốn.

Luồng xử lý của mô hình như hình Hình 10 bắt đầu bằng việc tiếp nhận hai nguồn dữ liệu đầu vào gồm tập dữ liệu có nhãn Plantdoc và tập SSL Dataset, cả hai đều trải qua các kỹ thuật tăng cường dữ liệu để tạo ra các biến thể đa dạng trước khi đi vào mạng nơ-ron tích chập Backbone. Tại đây, quá trình xử lý phân tách thành hai nhánh song song: nhánh giám sát đưa các véc-tơ đặc trưng qua lớp phân loại (CLS) để tính toán CLS loss dựa trên nhãn bệnh, trong khi đó, nhánh tự giám sát dẫn truyền đặc trưng qua mạng chiếu (MLP) và sử dụng cơ chế cập nhật động lượng (Momentum update) kết hợp với bộ nhớ đệm (Memory bank) để tính toán SSL loss. Trọng số của toàn bộ hệ thống cuối cùng được tối ưu hóa dựa trên hàm mất mát tổng hợp (Total loss), đảm bảo sự cân bằng giữa khả năng định danh chính xác của học giám sát và khả năng biểu diễn đặc trưng phong phú từ học tự giám sát.



Hình 10: Mô hình đề xuất

3.2. Xây dựng dữ liệu

3.2.1. Tập dữ liệu PlantDoc

Để đánh giá hiệu quả trong điều kiện thực tế của mô hình, luận văn sử dụng tập dữ liệu PlantDoc, được giới thiệu bởi Singh và cộng sự vào năm 2019 [7]. Khác với các bộ dữ liệu phổ biến trước đó như PlantVillage, vốn chủ yếu bao gồm các hình ảnh được chụp trong điều kiện tiêu chuẩn với sự kiểm soát chặt chẽ với nền đơn sắc, PlantDoc được xây dựng để đại diện cho các kịch bản canh tác thực tế in-the-wild scenarios [7].

Về mặt thống kê, tập dữ liệu bao gồm tổng cộng 2552 hình ảnh, bao phủ 13 loài cây trồng thương mại quan trọng (như táo, ngô, nho, khoai tây, cà chua, v.v.) và được phân chia thành 27 lớp (bao gồm 17 lớp bệnh và 10 lớp khỏe mạnh) [7]. Các hình ảnh trong tập dữ liệu được thu thập thông qua nhiều nguồn trên Internet, đảm bảo sự phong phú, đa dạng về nguồn gốc và đặc điểm hình ảnh [7].

Bảng 1: Bộ dữ liệu Plantdoc

Class	Train quantity	Test quantity
Apple Scab Leaf	77	10
Apple leaf	82	9
Apple rust leaf	79	10
Bell_pepper leaf	53	8
Bell_pepper leaf spot	62	9
Blueberry leaf	105	11
Cherry leaf	47	10
Corn Gray leaf spot	64	4
Corn leaf blight	180	12
Corn rust leaf	106	10
Peach leaf	103	9
Potato leaf early blight	109	8
Potato leaf late blight	97	8
Raspberry leaf	109	7
Soyabean leaf	57	8
Squash Powdery mildew leaf	123	6
Strawberry leaf	88	8
Tomato Early blight leaf	74	9
Tomato Septoria leaf spot	137	11
Tomato leaf	54	8
Tomato leaf bacterial spot	98	9
Tomato leaf late blight	101	10
Tomato leaf mosaic virus	44	10
Tomato leaf yellow virus	69	6
Tomato mold leaf	85	6
grape leaf	57	12
grape leaf black rot	56	8
Total	**2,316**	**236**

Đặc điểm nổi bật nhất khiến PlantDoc trở thành một thách thức lớn đối với các thuật toán học máy truyền thống là tính chất phi cấu trúc của nó. Các mẫu ảnh trong tập dữ liệu không chỉ chứa mẫu lá bệnh mà còn bao gồm nhiều yếu tố nhiễu nền phức tạp như đất, tay người nông dân, các bộ phận khác của cây, hoặc các loài thực vật khác mọc xen kẽ [7]. Hơn nữa, điều kiện nhiễu, ánh sáng, góc chụp và độ

phân giải của ảnh cũng biến thiên rất lớn. Chính vì số lượng mẫu gán nhãn tương đối hạn chế (trung bình chưa đến 100 ảnh mỗi lớp) cùng với độ nhiễu cao, PlantDoc là một ứng viên lý tưởng để kiểm chứng giả thuyết của luận văn: sử dụng học tự giám sát để trích xuất đặc trưng tốt hơn khi dữ liệu gán nhãn vừa thiếu, vừa khó.

3.2.2. Xây dựng tập dữ liệu SSL

Để đảm bảo việc hiệu quả tối ưu cho quá trình học tự giám sát, dữ liệu không nhãn đầu vào mô phỏng cần thỏa mãn đồng thời hai yếu tố là sự đa dạng về hình thái và tính tương đồng về mặt ngữ nghĩa với miền dữ liệu đích. Xuất phát từ yêu cầu này, luận văn đề xuất và hiện thực hóa quy trình xây dựng tập dữ liệu SSL thông qua cơ chế lọc dựa trên độ tương đồng đặc trưng. Bước đầu tiên của quy trình là khởi tạo một bể chứa dữ liệu ứng viên quy mô lớn bằng cách tổng hợp các nguồn dữ liệu thực vật mở đa dạng bao gồm PlantDoc [7], PlantWild [8], PlantSeg [9] và FieldPlant [10]. Việc tích hợp đa nguồn này giúp mở rộng không gian mẫu bao quát nhiều điều kiện môi trường, ánh sáng và góc chụp khác nhau, tạo tiền đề vật liệu phong phú cho mô hình học các đặc trưng bất biến mạnh mẽ.

Bộ dữ liệu PlantWild được xây dựng nhằm phục vụ đánh giá mô hình phân loại bệnh thực vật với điều kiện thực tế, với ảnh được thu thập ngoài môi trường tự nhiên, có phông nền phức tạp và nhiều biến thiên về ánh sáng, góc chụp cũng như chủng loại cây trồng. Dataset này bao gồm tổng cộng 18542 ảnh, được phân chia thành 89 lớp khác nhau, trong đó có 33 lớp cây khỏe và 56 lớp cây bị bệnh. Phân bố số lượng ảnh giữa các lớp thể hiện sự mất cân bằng dữ liệu, khi lớp có nhiều ảnh nhất đạt 589 ảnh, trong khi lớp ít nhất chỉ có 44 ảnh. Ngoài dữ liệu hình ảnh, mỗi lớp còn được gán kèm nhiều mô tả văn bản thu thập từ các nguồn tri thức mở và sinh tự động, giúp bổ sung thông tin ngữ nghĩa cho từng loại bệnh. Quy trình xây dựng dữ liệu được thực hiện chặt chẽ với nhiều vòng đánh giá thủ công nhằm đảm bảo độ chính xác của nhãn. Với quy mô lớn, số lượng lớp đa dạng và đặc trưng in-the-wild rõ rệt, PlantWild đóng vai trò như một bộ benchmark quan trọng cho các nghiên cứu phân loại bệnh thực vật trong điều kiện thiếu dữ liệu gán nhãn. PlantWild cũng đặc biệt

phù hợp để đánh giá hiệu quả của phương pháp học tự giám sát trong khi dữ liệu không đồng nhất.

Bên cạnh PlantWild, bộ dữ liệu PlantSeg được phát triển nhằm mở rộng từ bài toán phân loại sang bài toán phân vùng bệnh thực vật, cung cấp thông tin chi tiết hơn về vị trí và hình dạng của vùng bệnh trên lá. PlantSeg bao gồm khoảng 11400 ảnh có nhãn phân vùng bệnh và khoảng 8000 ảnh lá khỏe được phân theo từng loại cây. Toàn bộ dữ liệu đều thu thập kiểu in-the-wild, phản ánh sát các tình huống thực tế ngoài tự nhiên với phông nền tự nhiên và nhiều yếu tố gây nhiễu. Mỗi ảnh được gán nhãn phân vùng thủ công và trải qua quá trình kiểm tra bởi chuyên gia nhằm đảm bảo chất lượng. Với quy mô lớn và nhãn phân vùng chi tiết, PlantSeg hiện là một trong những bộ dữ liệu phân đoạn bệnh thực vật ngoài thực địa lớn nhất được công bố, đóng vai trò quan trọng trong việc đánh giá các mô hình học sâu, đặc biệt là các hướng tiếp cận kết hợp học tự giám sát và bán giám sát khi chỉ có một lượng nhỏ dữ liệu gán nhãn chính xác.

FieldPlant là bộ dữ liệu tập trung vào bối cảnh triển khai thực tế trong ruộng và vườn, khắc phục hạn chế của nhiều bộ dữ liệu trước đây vốn được thu thập trong môi trường phòng thí nghiệm với phông nền đơn giản. Dataset này bao gồm 5170 ảnh được chụp trực tiếp ngoài thực địa, trong đó có tổng cộng 8629 lá riêng lẻ được gán nhãn thủ công, phản ánh tình huống mỗi ảnh có thể chứa nhiều lá khác nhau. FieldPlant bao phủ 27 lớp bệnh, với quá trình gán nhãn được thực hiện bởi chuyên gia bệnh thực vật nhằm đảm bảo tính chính xác. Mặc dù không cung cấp nhãn phân vùng chi tiết cho vùng bệnh như PlantSeg, FieldPlant vẫn có giá trị lớn đối với các bài toán phân loại bệnh ở mức lá trong môi trường nền phức tạp, góp phần kiểm thử khả năng thích nghi của mô hình khi chuyển từ dữ liệu phòng thí nghiệm sang dữ liệu thực địa.

Mục tiêu cốt lõi là ứng dụng học tự giám sát nhằm nâng cao hiệu quả phân loại bệnh thực vật trong bối cảnh thiếu dữ liệu gán nhãn, ba bộ dữ liệu PlantWild, PlantSeg và FieldPlant có tính bổ trợ rõ rệt, phù hợp với nhiều khía cạnh của bài toán nghiên cứu. PlantWild cung cấp một tập dữ liệu quy mô lớn, đa dạng về lớp và có mức độ mất cân bằng cao, tạo điều kiện thuận lợi để học các biểu diễn đặc trưng giàu

tính khái quát từ dữ liệu chưa gán nhãn. PlantSeg đóng vai trò như một nguồn dữ liệu gán nhãn chất lượng cao, cho phép đánh giá và tinh chỉnh mô hình trong các bài toán yêu cầu thông tin không gian chi tiết về vùng bệnh. Trong khi đó, FieldPlant phản ánh trực tiếp những thách thức trong điều kiện thực tế ngoài môi trường tự nhiên, với phông nền phức tạp và nhiều đối tượng xuất hiện cùng lúc. Các kết quả thực nghiệm trước đây cho thấy, mặc dù mô hình học sâu có thể đạt độ chính xác rất cao trên dữ liệu thu thập trong phòng thí nghiệm, hiệu năng thường suy giảm đáng kể khi áp dụng lên dữ liệu thực địa, chỉ ra sự tồn tại rõ rệt của khoảng cách miền giữa hai bối cảnh. Trên cơ sở đó, bộ dữ liệu phục vụ huấn luyện học tự giám sát trong luận văn được tạo ra từ các tập dữ liệu nêu trên, đồng thời áp dụng khoảng cách cosin để đo lường độ tương đồng đặc trưng và lựa chọn tập dữ liệu phù hợp cho quá trình đánh giá.

Trước hết, một mô hình đặc trưng cơ sở được sử dụng để ánh xạ các mẫu dữ liệu có nhãn đầu vào thành các feature embeddings chuẩn hóa trong không gian biểu diễn liên tục. Các dữ liệu cùng nhãn ban đầu sẽ được xem là cùng cụm để tiến hành xây dựng prototypes. Trong các cụm riêng lẻ ban đầu, thuật toán tiến hành gom nhóm các embedding đại diện bằng K-Means clustering trong không gian đặc trưng. Khác với việc ấn định số cụm thủ công, số lượng cụm tối ưu được xác định tự động thông qua Elbow Method, giúp cân bằng giữa khả năng mô hình hóa đa dạng ngữ nghĩa và tránh over-partitioning. Tâm của mỗi cụm sau khi hội tụ được xem như một prototype, đại diện cho một semantic mode trong phân phối dữ liệu.

Song song với đó, pipeline áp dụng một cơ chế two-stage deduplication nhằm giảm thiểu ảnh hưởng của dữ liệu dư thừa. Ở pha thứ nhất, các bản sao chính xác được loại bỏ thông qua cryptographic hashing, đảm bảo mỗi mẫu vật lý chỉ xuất hiện một lần. Ở pha thứ hai, các mẫu gần trùng lặp về mặt thị giác được phát hiện bằng perceptual hashing kết hợp với Hamming distance, cho phép loại bỏ các biến thể gần giống nhau. Cơ chế này giúp tăng tính đa dạng hiệu quả của tập dữ liệu mà không làm tăng quy mô.

Sau bước tiền xử lý, mỗi mẫu dữ liệu không nhãn được so sánh với toàn bộ tập semantic prototypes bằng Cosine similarity. Với mỗi mẫu, thuật toán xác định

prototype gần nhất và ghi nhận giá trị tương đồng cực đại tương ứng. Dựa trên các giá trị này, một similarity threshold được áp dụng để loại bỏ các mẫu không đủ liên quan về mặt ngữ nghĩa. Đồng thời, để giải quyết vấn đề semantic imbalance – khi một số cụm ngữ nghĩa chiếm ưu thế về số lượng – thuật toán tích hợp thêm một chiến lược cluster-aware balancing. Cụ thể, trong mỗi nhóm, các mẫu được sắp xếp theo độ tương đồng giảm dần và chỉ giữ lại một số lượng giới hạn các mẫu có độ tin cậy cao hơn threshold và vượt ngưỡng nhất định nhằm đảm bảo sự cân bằng tương đối giữa các cụm.

Dựa trên phân phối của các điểm số tương đồng này, luận văn tạo ba tập thực nghiệm với quy mô và độ tin cậy khác nhau thông qua chiến lược lọc theo ngưỡng kết hợp chiến lược cân bằng dữ liệu. Cụ thể, ba ngưỡng tin cậy được thiết lập thực nghiệm là 0.44, 0.36 và 0.31 ứng với ba phiên bản dữ liệu: Nhỏ (Small), Trung bình (Medium) và Lớn (Large). Tập dữ liệu Nhỏ được hình thành bằng cách chỉ chọn những ảnh có độ tương đồng vượt qua 0.44. Đây là tập dữ liệu có độ tinh khiết cao nhất, bao gồm những mẫu sát nhất với miền dữ liệu gốc, giúp mô hình học các đặc trưng bệnh học chính xác nhưng hạn chế về sự đa dạng biến thể. Mở rộng không gian mẫu, tập dữ liệu Trung bình sử dụng ngưỡng 0.36 và tập dữ liệu Lớn sử dụng ngưỡng 0.31. Việc hạ thấp ngưỡng lọc cho phép chấp nhận thêm các mẫu dữ liệu có độ biến thiên lớn hơn về góc chụp, điều kiện ánh sáng hoặc các biểu hiện bệnh lý không điển hình. Chiến lược này giúp gia tăng đáng kể kích thước bộ dữ liệu, cung cấp cho mô hình khả năng tiếp cận với các biến thể phong phú của môi trường thực tế, nâng cao khả năng tổng quát hóa. Tuy nhiên, việc nới lỏng ngưỡng cũng đồng nghĩa với việc có một tỷ lệ nhiễu nhất định trong dữ liệu huấn luyện. Việc phân cấp dữ liệu thành ba mức độ dựa trên ngưỡng tương đồng như trên cho phép luận văn thực hiện các đánh giá toàn diện về sự đánh đổi giữa chất lượng và số lượng dữ liệu đối với hiệu suất cuối cùng của mô hình học tự giám sát.

Ngưỡng tin cậy trong nghiên cứu này được xác định dựa trên giá trị độ tương đồng Cosine giữa véc-tơ đặc trưng của ảnh và trung bình các nguyên mẫu lớp, đóng vai trò là bộ lọc để đảm bảo tính nhất quán về mặt ngữ nghĩa cho dữ liệu đầu vào. Mặc

dù nghiên cứu đã đề xuất chiến lược chọn lọc dữ liệu hiệu quả, tuy nhiên việc xác định ngưỡng tối ưu hiện vẫn dựa trên kết quả thực nghiệm khảo sát thay vì một cơ chế tự động hóa hoàn toàn. Khi mở rộng quy mô sang tập dữ liệu "Large", việc hạ thấp ngưỡng tin cậy để gia tăng số lượng mẫu đã vô tình đưa vào các biến số nhiễu đặc trưng của môi trường thực địa như sự chồng chéo của tán lá, nhiễu nền hoặc các đối tượng không liên quan về mặt bệnh lý.

Chi tiết dữ liệu được thể hiện như trong

Bảng 2: Tập dữ liệu cho quá trình học tự giám sát

3.3. Chiến lược tăng cường dữ liệu

Trong bài toán phân loại bệnh thực vật với điều kiện thiếu hụt dữ liệu gán nhãn, chiến lược tăng cường dữ liệu đóng vai trò quyết định đến khả năng hội tụ và tính tổng quát hóa của mô hình. Luận văn đề xuất một cơ chế tăng cường dữ liệu kép, được thiết kế riêng biệt cho hai luồng xử lý: luồng học giám sát và luồng học tự giám sát. Việc tách biệt này nhằm đảm bảo rằng dữ liệu có nhãn giữ được các đặc trưng ngữ nghĩa quan trọng để phân loại, trong khi dữ liệu tự giám sát phải trải qua các biến đổi nhiễu loạn mạnh mẽ để phục vụ việc học các đặc trưng bất biến.

3.3.1. Tăng cường dữ liệu cho Plandoc dataset

Đối với tập dữ liệu có nhãn Plantdoc, mục tiêu của quá trình tăng cường là gia tăng số lượng mẫu huấn luyện ảo và giảm thiểu hiện tượng quá khớp mà không làm sai lệch các đặc trưng bệnh học trên lá cây. Dựa trên đặc thù hình ảnh thực vật, luận văn áp dụng chuỗi các phép biến đổi hình học và quang học ở mức độ vừa phải. Cụ thể, các ảnh đầu vào được cắt ngẫu nhiên và thay đổi kích thước (Random Resized Crop) với tỷ lệ diện tích từ 60% đến 100% so với ảnh gốc, đảm bảo chủ thể chính là lá bệnh vẫn nằm trong vùng quan sát. Bên cạnh đó, kỹ thuật lật ngang (Random Horizontal Flip) và xoay ngẫu nhiên một góc 15 độ (Random Rotation) được sử dụng để giúp mô hình bất biến với hướng của lá cây trong không gian. Về mặt quang học, phép biến đổi màu sắc (Color Jitter) được áp dụng với các tham số độ sáng, độ tương

phản và độ bão hòa đều được điều chỉnh ở mức 0.2, giúp mô hình thích nghi với các điều kiện ánh sáng khác nhau khi chụp ảnh thực tế. Cuối cùng, tất cả các ảnh đều được chuẩn hóa theo phân phối chuẩn của tập ImageNet để tăng tốc độ hội tụ của thuật toán tối ưu.

Bảng 3: Tham số tăng cường cho quá trình huấn luyện

STT	Phép biến đổi	Tham số chính
1	RandomResizedCrop	size = 224, scale = (0.6, 1.0)
2	RandomHorizontalFlip	mặc định
3	RandomRotation	góc = 15
4	ColorJitter	brightness = 0.2, contrast = 0.2, saturation = 0.2
5	Normalize	mean = [0.485, 0.456, 0.406], std = [0.229, 0.224, 0.225]

Bảng 4: Tham số tăng cường cho quá trình kiểm thử

STT	Phép biến đổi	Tham số chính
1	Resize	size = 256
2	CenterCrop	size = 224
3	Normalize	mean = [0.485, 0.456, 0.406], std = [0.229, 0.224, 0.225]

3.3.2. Tăng cường dữ liệu cho SSL dataset

Đối với tập dữ liệu không gán nhãn dùng cho học tự giám sát, mục tiêu của quá trình tăng cường là tạo ra các cặp mẫu huấn luyện có tính chất tương phản mạnh mẽ từ cùng một ảnh gốc, buộc mô hình phải học được các đặc trưng ngữ nghĩa bất biến thay vì ghi nhớ các chi tiết nhiễu. Dựa trên cơ chế của thuật toán MoCo v2, luận văn áp dụng một chuỗi các phép biến đổi ngẫu nhiên đồng nhất nhưng được thực thi độc lập hai lần cho mỗi ảnh đầu vào. Cụ thể, các ảnh được cắt ngẫu nhiên và thay đổi kích thước (Random Resized Crop) với tỷ lệ diện tích dao động lớn từ 20% đến 100% so với ảnh gốc, giúp mô hình học được mối quan hệ giữa các phần cục bộ và toàn thể của lá bệnh. Để ngăn chặn việc mô hình dựa vào biểu đồ màu sắc để phân loại, phép biến đổi màu (Color Jitter) được áp dụng mạnh tay với các tham số độ sáng, độ tương phản, độ bão hòa đều là 0.4 và độ lệch màu

là 0.1 với xác suất 80%, kết hợp với việc chuyển đổi sang ảnh xám (Random Grayscale) với xác suất 20%. Đặc biệt, kỹ thuật làm mờ Gaussian (Gaussian Blur) nhằm loại bỏ các chi tiết tần số cao, buộc mạng nơ-ron phải tập trung vào hình dáng và kết cấu tổng thể của vết bệnh. Cuối cùng, kỹ thuật lật ngang ngẫu nhiên (Random Horizontal Flip) và chuẩn hóa theo phân phối chuẩn ImageNet được áp dụng để hoàn thiện dữ liệu đầu vào cho quá trình huấn luyện.

Bảng 5: Tham số tăng cường dữ liệu cho dữ liệu học tự giám sát

STT	Phép biến đổi	Tham số chính
1	RandomResizedCrop	size, scale = (0.2, 1.0)
2	RandomApply (ColorJitter)	brightness = 0.4, contrast = 0.4, saturation = 0.4, hue = 0.1, p = 0.8
3	RandomGrayscale	p = 0.2
4	RandomApply (GaussianBlur)	kernel size = 23, sigma = (0.1, 2.0), p = 0.5
5	RandomHorizontalFlip	mặc định
6	Normalize	mean = [0.485, 0.456, 0.406], std = [0.229, 0.224, 0.225]

3.4. Thiết kế thành phần mạng nơ-ron

3.4.1. Mạng xương sống

Sự phát triển mạnh mẽ của Điện toán biên và Internet vạn vật đang đặt ra nhu cầu xử lý dữ liệu trực tiếp bằng di động và các hệ thống nhúng. Tại các khu vực canh tác nông nghiệp, nơi hạ tầng mạng thiếu ổn định và băng thông hạn chế, việc phụ thuộc vào các mô hình học sâu quy mô lớn vận hành trên máy chủ đám mây bộc lộ rõ những rào cản về độ trễ truyền tải, chi phí duy trì và tính khả dụng. Vì vậy, luận văn này lựa chọn hướng tiếp cận dựa trên các kiến trúc mô hình hạng nhẹ, thay vì trông chờ vào những mạng nơ-ron khổng lồ vốn khó triển khai trong điều kiện thực địa.

ResNet10T (resnet10t.c3_in1k) là một biến thể rút gọn của họ ResNet được tối ưu trong thư viện timm nhằm đạt hiệu năng tốt trên những hệ thống giới hạn tài

nguyên. Mô hình sử dụng kiến trúc stem ba lớp convolution kết hợp với cơ chế downsampling bằng average-pool và các nhánh tắt 1×1, giữ lại ưu điểm ổn định của kiến trúc residual nhưng giảm đáng kể độ sâu và chi phí tính toán. Nhờ số tham số thấp và độ phức tạp nhỏ, ResNet10T phù hợp với các bài toán mô hình nhẹ, tốc độ suy luận cao và vẫn duy trì khả năng trích xuất đặc trưng ở mức chấp nhận được trên các thiết bị biên.

Trong triển khai thực tế, ResNet10T cho thấy độ trễ suy diễn thấp và tốc độ xử lý cao nhờ kiến trúc gọn nhẹ, số lượng tham số ít và chi phí tính toán nhỏ, giúp giảm đáng kể thời gian truy cập bộ nhớ và overhead điều phối toán tử, vốn là các yếu tố chi phối latency trong môi trường tài nguyên hạn chế. Label Smoothing chỉ tác động trong giai đoạn huấn luyện thông qua việc làm mềm phân phối nhãn, không làm thay đổi đồ thị suy diễn, do đó hầu như không ảnh hưởng trực tiếp đến độ trễ triển khai. Ngược lại, Mixed Precision (ví dụ FP16 hoặc BF16) có ảnh hưởng trực tiếp đến latency, cho phép giảm băng thông bộ nhớ và khai thác hiệu quả các đơn vị tính toán vector hoặc Tensor Cores, từ đó rút ngắn thời gian suy diễn và cải thiện hiệu năng triển khai, đặc biệt trong các kịch bản suy luận thời gian thực và trên thiết bị biên.

3.4.2. Chiếu phi tuyến MLP

Khối MLP projection head được triển khai dưới dạng một mạng hai tầng tuyến tính nhằm biến đổi không gian đặc trưng thô từ encoder sang một không gian biểu diễn phù hợp hơn cho bài toán tự giám sát. Ở tầng đầu, đầu vào có kích thước encoder_dim được mở rộng gấp đôi giúp mô hình học được các quan hệ phi tuyến phức tạp hơn, đi kèm hàm kích hoạt ReLU để tăng khả năng biểu diễn. Tầng tuyến tính thứ hai nén lại không gian này về feature_dim, tạo ra vector đặc trưng cuối cùng dùng cho mục tiêu so sánh và tối ưu hóa trong học tương phản. Thiết kế mở rộng rồi thu hẹp này giúp mô hình tách biệt được không gian biểu diễn cho nhiệm vụ và không gian biểu diễn cho encoder, qua đó ổn định việc học. Khối MLP này với vai trò là "cầu nối" giữa backbone và hàm mục tiêu tự giám sát, đảm bảo các đặc trưng học được mang tính phân tách và hữu ích cho giai đoạn fine-tune sau đó.

3.4.3. Bộ phân loại CLS

Khối CLS được thiết kế theo dạng tuyến tính đơn giản, nhận đầu vào là vector đặc trưng cuối cùng do backbone trích xuất và ánh xạ trực tiếp sang không gian nhãn. Cấu trúc tối giản này phù hợp với mục tiêu tách biệt rõ ràng giữa phần học biểu diễn và phần ra quyết định, đảm bảo rằng toàn bộ sức mạnh mô hình tập trung vào việc học đặc trưng thay vì phụ thuộc vào một kiến trúc phân loại phức tạp. Việc sử dụng một lớp tuyến tính duy nhất cũng giúp giảm thiểu số lượng tham số, từ đó tránh overfitting trong bối cảnh dữ liệu hạn chế. Đây chính là tầng chịu trách nhiệm chuyển đổi biểu diễn trừu tượng sang dự đoán cụ thể, đóng vai trò là cầu nối cuối cùng giữa không gian đặc trưng và nhiệm vụ phân loại bệnh thực vật.

3.5. Cơ chế huấn luyện và cập nhập trọng số

Trong kiến trúc mạng lai ghép mà trong luận văn đề xuất, quy trình huấn luyện được thiết kế không chỉ dựa trên phương thức học giám sát truyền thống mà là sự phối hợp đồng bộ giữa hai cơ chế cập nhật trọng số riêng biệt: cập nhật dựa trên đạo hàm cho nhánh mạng chính và cập nhật động lượng cho nhánh mạng phụ trợ. Chiến lược này được trong luận văn xây dựng nhằm giải quyết bài toán cốt lõi là duy trì sự ổn định của không gian biểu diễn đặc trưng khi học từ dữ liệu không nhãn, đồng thời vẫn đảm bảo tính linh hoạt cần thiết để tiếp thu các đặc trưng phân loại từ nguồn dữ liệu có nhãn vốn hạn chế.

Đối với nhánh mạng xương sống chính và các đầu ra dự đoán bao gồm đầu phân loại và đầu chiếu, cơ chế cập nhật trọng số tuân theo nguyên tắc tối ưu hóa hàm mất mát tổng hợp thông qua thuật toán lan truyền ngược. Như được minh họa trong sơ đồ kiến trúc, tại mỗi bước huấn luyện, mạng xương sống chính đóng vai trò là bộ mã hóa truy vấn, tiếp nhận dữ liệu ảnh để trích xuất véc-tơ đặc trưng. Các tham số của nhánh này được hệ thống cập nhật tức thời dựa trên gradient của tổng hàm mất mát, là sự kết hợp giữa mất mát phân loại từ nhãn dữ liệu Plantdoc và mất mát tương phản từ dữ liệu tự giám sát. Việc cập nhật trực tiếp này cho phép mô hình nhanh chóng điều chỉnh các trọng số để tối ưu hóa khả năng phân biệt giữa các lớp bệnh, đồng thời học được các đặc trưng bất biến từ lượng lớn dữ liệu không nhãn được đưa vào.

Song song với quá trình trên, nhánh mạng phụ trợ được trong luận văn thiết kế hoạt động với vai trò là bộ mã hóa khóa và chỉ tham gia vào quá trình huấn luyện mà không thực hiện nhiệm vụ suy diễn. Điểm đặc biệt trong kiến trúc đề xuất là trọng số của mạng phụ trợ không được cập nhật thông qua lan truyền ngược gradient, nhằm tránh hiện tượng dao động mạnh của các biểu diễn đặc trưng trong từ điển mẫu âm. Thay vào đó, trong luận văn áp dụng cơ chế cập nhật động lượng, trong đó các tham số của mạng phụ trợ ở bước hiện tại được tính toán theo công thức dưới đây. Với hệ số động lượng m được thiết lập ở mức cao, k đại diện cho tham số bộ mã hoá khoá và q đại diện cho tham số bộ mã hoá chính, điều này giúp bộ mã hoá khoá mượt mà hơn so với bộ mã hoá chính. Cơ chế này đảm bảo tính nhất quán của các đặc trưng trong bộ nhớ đệm, giúp quá trình so khớp tương phản giữa truy vấn và khóa diễn ra ổn định, ngăn chặn hiệu quả hiện tượng mô hình hội tụ về các giải pháp tầm thường.

$$k_{new} = k \cdot m + q \cdot (1 - m)$$

Bên cạnh cơ chế cập nhật trọng số, hệ thống được tích hợp một bộ nhớ đệm dạng hàng đợi để lưu trữ các véc-tơ đặc trưng từ các bước huấn luyện trước đó. Trong mỗi chu kỳ huấn luyện, sau khi véc-tơ đặc trưng của lô hiện tại sẽ bị mã hóa bởi mạng động lượng, chúng sẽ được đẩy vào hàng đợi và các mẫu cũ nhất sẽ bị loại bỏ theo nguyên tắc vào trước ra trước. Việc kết hợp mạng mã hóa động lượng với hàng đợi cho phép mô hình duy trì một tập hợp lớn các mẫu phủ định mà không phụ thuộc vào kích thước lô giới hạn của phần cứng. Điều này đặc biệt có ý nghĩa trong bối cảnh tài nguyên tính toán hạn chế của luận văn, giúp mô hình so sánh và phân biệt được các đặc trưng tinh tế của bệnh thực vật trên một không gian mẫu rộng lớn, từ đó nâng cao độ chính xác tổng thể của hệ thống ngay cả khi dữ liệu gán nhãn đầu vào là khan hiếm.

3.6. Xây dựng hàm mất mát

Để giải quyết đồng thời việc phân loại bệnh thực vật chính xác và bài toán học biểu diễn đặc trưng mạnh mẽ trong điều kiện thiếu dữ liệu, luận văn xây dựng một hàm mục tiêu tổng hợp. Hàm mục tiêu này là sự kết hợp tuyến tính giữa hai thành phần: hàm mất mát giám sát và hàm mất mát tự giám sát. Việc thiết kế hàm mất mát

đa nhiệm này đóng vai trò như một cơ chế điều hướng, buộc mô hình vừa phải tối ưu hóa độ chính xác phân loại, vừa phải duy trì cấu trúc tô pô của không gian đặc trưng sao cho các biến thể của cùng một loại bệnh luôn nằm gần nhau.

3.6.1. Hàm mất mát phân loại với kỹ thuật làm trơn nhãn

Đối với nhánh giám sát, mục tiêu là cực tiểu hóa sai số giữa phân phối xác suất dự đoán và nhãn thực tế của dữ liệu PlantDoc. Thay vì sử dụng hàm mất mát Entropy chéo tiêu chuẩn, luận văn áp dụng kỹ thuật Entropy chéo có làm trơn nhãn. Trong bối cảnh dữ liệu bệnh thực vật thường chứa nhiễu, việc sử dụng véc-tơ nhãn cứng dễ dẫn đến tình trạng mô hình quá tự tin vào dự đoán, gây ra hiện tượng quá khớp. Do đó, luận văn thiết lập tham số làm trơn $\epsilon = 0.1$, phân phối lại một phần xác suất nhỏ cho các lớp còn lại thay vì tập trung toàn bộ vào lớp đúng. Hàm mất mát này khuyến khích mô hình học các đặc trưng mang tính tổng quát hơn thay vì ghi nhớ máy móc các mẫu huấn luyện, qua đó cải thiện khả năng thích nghi trên tập dữ liệu kiểm thử.

$$\mathcal{L}_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c} \log p_{i,c}$$

Với N là số mẫu trong 1 batch, C là số lượng lớp, $y_{i,c}$ nhãn của mẫu i theo encode, $p_{i,c}$ là xác suất của mẫu i cho các lớp.

3.6.2. Hàm mất mát tương phản InfoNCE cho học tự giám sát

Đối với nhánh tự giám sát, luận văn sử dụng hàm mất mát InfoNCE. Mục tiêu của hàm này là tối đa hóa độ tương đồng giữa véc-tơ truy vấn q và véc-tơ khóa dương $k+$, đồng thời cực tiểu hóa độ tương đồng giữa q và tất cả các véc-tơ khóa âm $k-$ được lưu trữ trong bộ nhớ đệm. Trong không gian đặc trưng, phép đo độ tương đồng được thực hiện thông qua tích vô hướng giữa hai véc-tơ đã chuẩn hóa. Để kiểm soát độ sắc nhọn của phân phối xác suất, luận văn sử dụng một tham số nhiệt độ $\tau = 0.07$, mô hình phải tìm ra đúng mẫu dương duy nhất trong số hàng nghìn mẫu âm tính lấy từ hàng đợi, buộc mạng Backbone phải trích xuất được các đặc trưng phân biệt cốt lõi của lá bệnh bất chấp các biến đổi về màu sắc hay hình học.

$$\mathcal{L}_{MoCov2} = -\log \frac{\exp\left((q \cdot k^{+})/\tau\right)}{\exp\left((q \cdot k^{+})/\tau\right) + \sum_{j=1}^{K} \exp\left((q \cdot k_j^{-})/\tau\right)}$$

3.6.3. Hàm mất mát tổng hợp và chiến lược trọng số động

Hàm mất mát toàn cục giúp cập nhật trọng số cho mạng là tổng có trọng số của hai thành phần nêu trên, được biểu diễn qua công thức $\mathcal{L}_{total} = \alpha \cdot \mathcal{L}_{SSL} + \beta \cdot \mathcal{L}_{CLS}$. Trong đó, $\mathcal{L}_{CLS}$ là hàm mất mát cho nhánh phân loại, $\mathcal{L}_{SSL}$ là hàm mất mát cho nhánh học tự giám sát. β là trọng số của thành phần phân loại, thường được giữ cố định ở giá trị 1.0 nhằm đảm bảo mục tiêu chính của bài toán luôn được ưu tiên. Ngược lại, hệ số α điều khiển mức độ đóng góp của quá trình tự giám sát để mô hình tận dụng mạnh mẽ các đặc trưng từ học tự giám sát nhằm định hình không gian biểu diễn. Thiết kế này cho phép mô hình đồng thời khai thác thông tin nhãn để tăng cường khả năng phân biệt lớp, đồng thời tận dụng dữ liệu không nhãn để học các biểu diễn giàu ngữ nghĩa và có khả năng tổng quát hóa tốt hơn.

3.7. Các kỹ thuật liên quan

3.7.1. Gradient Accumulation

Trong các phương pháp học tương phản điển hình như MoCo, kích thước lô huấn luyện có ảnh hưởng trực tiếp đến chất lượng của biểu diễn đặc trưng học được, bởi số lượng mẫu âm tính tỷ lệ thuận với batch size, từ đó giúp mô hình phân biệt tốt hơn giữa các biểu diễn khác nhau trong không gian đặc trưng. Tuy nhiên, batch size lớn trong thực tế thường bị giới hạn bởi GPU, đặc biệt khi huấn luyện các mô hình sâu trên dữ liệu ảnh có độ phân giải cao. Để khắc phục hạn chế này, luận văn áp dụng kỹ thuật tích lũy gradient, trong đó các giá trị gradient không được sử dụng để cập nhật trọng số, mà được cộng dồn qua nhiều bước lặp liên tiếp trước khi thực hiện một lần cập nhật tham số. Cách tiếp cận này cho phép mô phỏng hiệu quả một batch size lớn hơn so với khả năng phần cứng thực tế, qua đó duy trì số lượng mẫu âm tính cần thiết cho quá trình học tương phản, đồng thời đảm bảo sự ổn định quá trình huấn luyện học tự giám sát mà không gây ra hiện tượng tràn bộ nhớ.

3.7.2. Automatic Mixed Precision

Nhằm nâng cao hiệu quả huấn luyện và giảm mức tiêu thụ bộ nhớ trong quá trình học tự giám sát, luận văn áp dụng cơ chế Automatic Mixed Precision như một giải pháp tối ưu hóa tính toán. Phương pháp này cho phép thực hiện các phép toán có chi phí tính toán lớn, đặc biệt là nhân ma trận và tích chập, dưới định dạng số thực dấu phẩy động 16 bit thay cho định dạng 32 bit truyền thống, qua đó giúp giảm đáng kể dung lượng bộ nhớ sử dụng và tăng thông lượng xử lý trên các bộ xử lý đồ họa hiện đại. Đồng thời, để đảm bảo độ ổn định và khả năng hội tụ của mô hình trong quá trình huấn luyện, một bản sao trọng số có độ chính xác cao vẫn được duy trì và sử dụng trong bước cập nhật tham số. Việc kết hợp cơ chế này với kỹ thuật cân bằng gradient giúp hạn chế các vấn đề về mất mát độ chính xác số học và hiện tượng tiêu biến gradient, từ đó duy trì chất lượng học của mô hình trong khi vẫn đạt được tốc độ huấn luyện cao hơn so với việc sử dụng hoàn toàn độ chính xác đơn lẻ.

3.7.3. Điều chuẩn Suy giảm trọng số

Nhằm giảm thiểu hiện tượng quá khớp trong quá trình huấn luyện các mạng nơ-ron sâu, luận văn áp dụng kỹ thuật điều chuẩn suy giảm trọng số như một thành phần quan trọng trong chiến lược tối ưu hóa mô hình. Phương pháp này hoạt động bằng cách bổ sung một hạng tử phạt tỷ lệ với chuẩn L2 của các tham số mạng vào hàm mất mát, qua đó hạn chế sự gia tăng quá mức của các trọng số trong quá trình học. Khi kết hợp với bộ tối ưu hóa AdamW, suy giảm trọng số được tách biệt rõ ràng khỏi bước cập nhật gradient, giúp bộ tối ưu hướng quá trình học về các nghiệm có biên độ trọng số nhỏ và phân bố ổn định hơn. Cách tiếp cận này góp phần làm mượt đường biên quyết định của mô hình, giảm mức độ nhạy cảm đối với nhiễu trong dữ liệu huấn luyện và từ đó cải thiện khả năng tổng quát hóa trên tập dữ liệu kiểm thử, đặc biệt trong bối cảnh dữ liệu gán nhãn hạn chế của bài toán phân loại bệnh thực vật.

3.8. Kết luận chương 3

Chương 3 đã trình bày toàn diện thiết kế kiến trúc mô hình lai ghép, tích hợp song song luồng học giám sát và luồng học tự giám sát theo cơ chế tương phản MoCo v2 trên nền tảng mạng xương sống hạng nhẹ ResNet10t nhằm giải quyết triệt để bài toán phân loại bệnh thực vật trong điều kiện khan hiếm dữ liệu gán nhãn. Điểm nhấn

quan trọng của chương là quy trình xây dựng dữ liệu tự giám sát thông minh, trong đó nguồn dữ liệu thô từ các tập PlantWild, PlantSeg và FieldPlant được tinh lọc nghiêm ngặt thông qua thuật toán so khớp độ tương đồng Cosine với các nguyên mẫu lớp của tập PlantDoc để loại bỏ nhiễu và đảm bảo tính nhất quán ngữ nghĩa. Song hành với đó, chiến lược tăng cường dữ liệu kép được thiết kế chuyên biệt, áp dụng các biến đổi nhiễu loạn mạnh mẽ như làm mờ Gaussian cho nhánh tự giám sát để học cấu trúc nội tại, trong khi nhánh giám sát chỉ chịu các biến đổi nhẹ để bảo toàn đặc trưng bệnh học. Về cơ chế vận hành, mô hình sử dụng cập nhật trọng số động lượng cho bộ mã hóa khóa kết hợp với hàng đợi bộ nhớ để duy trì sự ổn định của từ điển mẫu âm, giúp tách biệt kích thước lô khỏi giới hạn phần cứng. Hàm mục tiêu tổng hợp là sự kết hợp tuyến tính giữa hàm mất mát InfoNCE và Entropy chéo có làm trơn nhãn, được điều tiết bởi trọng số động nhằm cân bằng giữa khả năng phân biệt lớp và chất lượng biểu diễn đặc trưng. Để tối ưu hóa hiệu suất trên tài nguyên hạn chế, các kỹ thuật tiên tiến như tích lũy gradient và tính toán độ chính xác hỗn hợp tự động đã được tích hợp sâu vào quy trình huấn luyện, cho phép mô phỏng kích thước lô lớn mà không gây tràn bộ nhớ. Tổng thể thiết kế này không chỉ khắc phục được hạn chế quá khớp của các mô hình CNN truyền thống mà còn đảm bảo tính khả thi khi triển khai trên các thiết bị biên nhờ kiến trúc gọn nhẹ và tối ưu. Những đề xuất chi tiết về kiến trúc, phương pháp xây dựng dữ liệu và chiến lược tối ưu hóa này đã thiết lập một khung làm việc vững chắc về mặt lý thuyết và kỹ thuật. Đây chính là tiền đề then chốt để tiến hành các kịch bản thực nghiệm định lượng, đánh giá tính hiệu quả và so sánh hiệu năng của mô hình được trình bày chi tiết trong Chương 4 kế tiếp.

CHƯƠNG 4: THỰC NGHIỆM VÀ ĐÁNH GIÁ

4.1. Cấu hình phần cứng

Giai đoạn đầu tiên bao gồm việc xây dựng tập dữ liệu tự giám sát và thực hiện các thuật toán tuyển chọn mẫu dựa trên nguyên mẫu được thực hiện trên thiết bị với cấu hình phần cứng cơ bản. Cụ thể, hệ thống sử dụng Intel Core i5 1135G7 2.4 GHz, kết hợp cùng bộ nhớ truy cập ngẫu nhiên 8 GB. Việc thuật toán lọc và chuẩn hóa dữ liệu có thể vận hành trơn tru trên cấu hình này cho thấy chi phí tính toán cho bước chuẩn bị dữ liệu là hoàn toàn chấp nhận được, cho phép người dùng dễ dàng mở rộng hoặc cập nhật nguồn dữ liệu đầu vào mà không gặp rào cản về phần cứng.

Ngược lại, giai đoạn huấn luyện mô hình lai ghép đòi hỏi khả năng xử lý song song mạnh mẽ để thực hiện các phép nhân ma trận và tích chập trên quy mô lớn, đặc biệt là việc duy trì hàng đợi bộ nhớ và tính toán gradient cho cả hai nhánh giám sát và tự giám sát. Do đó, các kịch bản thử nghiệm chính thức được triển khai trên nền tảng cloud hỗ trợ tăng tốc phần cứng. Hệ thống sử dụng GPU P100 với dung lượng 16 GB. Việc trang bị tài nguyên phần cứng mạnh mẽ này là yếu tố then chốt, cho phép thiết lập kích thước lô dữ liệu đủ lớn để đảm bảo sự hội tụ ổn định của hàm mất mát tương phản, đồng thời rút ngắn thời gian huấn luyện trên các tập dữ liệu phức tạp.

4.2. Các tiêu chí đánh giá

Việc đánh giá hiệu năng đóng vai trò nền tảng trong việc xác định tính khả thi và độ tin cậy của mô hình học sâu đối với bài toán phân loại bệnh thực vật, đặc biệt khi triển khai trong điều kiện thực tế với nguồn dữ liệu gán nhãn hạn chế và phân phối lớp không đồng đều. Phần này trình bày hệ thống các tiêu chí định lượng được sử dụng để thẩm định mô hình bao gồm Accuracy, Precision, Recall, F1-score, đồng thời mở rộng sang một số chỉ số bổ trợ như Confusion Matrix, Macro-averaged Metrics và AUC-ROC nhằm cung cấp cái nhìn toàn diện hơn. Cụ thể, Accuracy phản ánh tỷ lệ phân loại đúng trên toàn bộ tập dữ liệu và đóng vai trò là thước đo tổng quát về hiệu suất của mạng nơ-ron; tuy nhiên, chỉ số này có thể gây sai lệch trong bối cảnh dữ liệu mất cân bằng. Trong bài toán chẩn đoán bệnh cây trồng, Recall giữ vai trò then chốt trong việc đánh giá khả năng phát hiện đầy đủ các mẫu bệnh, qua đó hạn

chế nguy cơ bỏ sót các trường hợp nhiễm bệnh ngoài thực địa. Song hành với đó, Precision đo lường mức độ chính xác của các dự đoán dương tính, giúp giảm thiểu các cảnh báo sai có thể dẫn đến lãng phí nguồn lực can thiệp nông nghiệp. Để dung hòa giữa Precision và Recall, F1-score được sử dụng như một chỉ số tổng hợp, đặc biệt hiệu quả trong các kịch bản dữ liệu mất cân bằng như tập PlantDoc. Bên cạnh đó, việc phân tích Confusion Matrix cho phép nhận diện chi tiết các dạng lỗi phân loại giữa các nhóm bệnh, trong khi Macro-F1 đảm bảo mỗi lớp được đánh giá công bằng bất kể quy mô mẫu, và AUC-ROC hỗ trợ đánh giá khả năng phân tách lớp của mô hình trong điều kiện ngưỡng quyết định thay đổi, góp phần nâng cao độ tin cậy của kết luận thực nghiệm.

4.2.1. Độ chính xác tổng thể (Accuracy)

Độ chính xác tổng thể (Accuracy) là thước đo định lượng cơ bản nhất phản ánh năng lực phân loại chung của mô hình trên toàn bộ tập dữ liệu kiểm thử. Về mặt toán học, chỉ số này được định nghĩa là tỷ lệ giữa tổng số mẫu được dự đoán chính xác (bao gồm cả các mẫu dương tính - True Positives và mẫu âm tính - True Negatives) trên tổng không gian mẫu. Trong bối cảnh phân loại bệnh thực vật, Accuracy cho biết tỷ lệ phần trăm các lá cây (bao gồm cả lá bệnh và lá khỏe mạnh) được mô hình định danh đúng.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Trong đó:

- True Positive (TP) đại diện cho số lượng ảnh lá bệnh được mô hình nhận diện chính xác là có bệnh
- True Negative (TN) là số lượng ảnh lá khỏe mạnh (hoặc không thuộc loại bệnh đang xét) được phân loại đúng là âm tính
- False Positive (FP - Lỗi loại I) xảy ra khi mô hình báo động giả, tức là phân loại nhầm một lá khỏe mạnh (hoặc bệnh khác) thành loại bệnh đang quan tâm
- False Negative (FN - Lỗi loại II) là trường hợp bỏ sót, khi lá thực sự có bệnh lại bị mô hình phân loại là khỏe mạnh hoặc thuộc nhóm khác.

4.2.2. Độ che phủ (Recall)

Recall là thước đo khả năng phát hiện các mẫu dương tính của mô hình. Nó được định nghĩa là tỷ lệ giữa số lượng mẫu dương tính được phân loại chính xác (TP) trên tổng số mẫu thực sự thuộc lớp dương tính (TP + FN).

$$Recall = \frac{TP}{TP + FN}$$

Trong bài toán chẩn đoán bệnh cây trồng, Recall mang ý nghĩa thực tiễn vô cùng quan trọng, thậm chí được ưu tiên hơn trong các giai đoạn sàng lọc ban đầu. Một giá trị Recall cao đồng nghĩa với việc mô hình có khả năng nhận diện được hầu hết các trường hợp nhiễm bệnh, giảm thiểu tối đa tỷ lệ bỏ sót (False Negative). Việc bỏ sót một cây bị bệnh (FN) có thể dẫn đến hậu quả nghiêm trọng hơn nhiều so với việc cảnh báo nhầm, bởi mầm bệnh không được phát hiện kịp thời sẽ lây lan ra cả cánh đồng, gây thiệt hại kinh tế lớn. Do đó, mục tiêu của mô hình đề xuất là tối ưu hóa Recall để đảm bảo tính an toàn sinh học và khả năng kiểm soát dịch hại.

4.2.3. Độ chính xác (Precision)

Precision là thước đo độ tin cậy của các dự đoán dương tính, được định nghĩa là tỷ lệ giữa số lượng mẫu dương tính được phân loại chính xác (TP) trên tổng số mẫu được mô hình phân loại là dương tính (TP + FP).

$$Precision = \frac{TP}{TP + FP}$$

Precision cho biết trong số các trường hợp mô hình cảnh báo là "có bệnh", bao nhiêu phần trăm là chính xác. Trong ứng dụng thực tế, Precision cao giúp giảm thiểu tỷ lệ báo động giả (False Positive). Nếu Precision thấp, hệ thống sẽ đưa ra nhiều cảnh báo sai, dẫn đến việc người nông dân phun thuốc bảo vệ thực vật không cần thiết lên các cây khỏe mạnh. Điều này không chỉ gây lãng phí chi phí, nhân lực mà còn tác động tiêu cực đến môi trường và để lại dư lượng hóa chất trên nông sản. Vì vậy, Precision là tiêu chí then chốt để đảm bảo tính kinh tế và hiệu quả sử dụng tài nguyên trong quy trình canh tác thông minh.

4.2.4. F1-score

F1-score là trung bình điều hòa (harmonic mean) giữa Precision và Recall, cung cấp một thước đo tổng hợp để đánh giá hiệu năng mô hình một cách toàn diện, đặc biệt trong các trường hợp dữ liệu phân bố không đồng đều như tập PlantDoc. Công thức tính F1-score giúp dung hòa sự đánh đổi tự nhiên giữa độ phủ và độ chính xác dự báo.

$$F1 - score = \frac{2 \cdot (Precision \cdot Recall)}{Precision + Recall}$$

Chỉ số F1 đạt giá trị cao nhất khi và chỉ khi cả Precision và Recall đều cao, đồng nghĩa với việc mô hình vừa có khả năng phát hiện tốt các ca bệnh (không bỏ sót), vừa đảm bảo độ tin cậy của các dự đoán (không báo nhầm). Trong nghiên cứu này, F1-score được sử dụng như tiêu chí chính để so sánh hiệu quả giữa các kiến trúc mạng khác nhau, bởi nó phản ánh đúng năng lực thực tế của hệ thống trong việc cân bằng giữa mục tiêu kiểm soát dịch bệnh (thiên về Recall) và mục tiêu tối ưu hóa chi phí xử lý (thiên về Precision), đảm bảo mô hình hoạt động ổn định và hiệu quả trên cả hai phương diện.

4.3. Kết quả thực nghiệm

Trong giai đoạn đầu thử nghiệm các phương pháp, luận văn tiến hành đánh giá mô hình theo nhiều phương pháp khác nhau của học tự giám sát như SimCLR, MoCo v2 và MAE mô hình học giám sát thuần tuý trên cùng bộ dữ liệu plantdoc, các pretraining của học tự giám sát được huấn luyện trên cùng bộ dữ liệu plantdoc làm đầu vào cho tác vụ hạ nguồn. Cụ thể, trong luận văn thử nghiệm với với mỗi phương pháp tiền huấn luyện, mô hình được thử nghiệm dưới hai giao thức chính. Thứ nhất, backbone được giữ cố định và chỉ huấn luyện bộ phân loại trên tập dữ liệu gán nhãn, nhằm đánh giá chất lượng biểu diễn học được từ giai đoạn tiền huấn luyện. Thứ hai, toàn bộ mô hình được fine-tune end-to-end trên tập dữ liệu downstream để khảo sát khả năng thích nghi của biểu diễn khi có cập nhật tham số đầy đủ.

Số liệu thực nghiệm tại *Bảng 6* minh chứng rằng hiệu năng phân loại phụ thuộc mật thiết vào sự tương thích giữa phương pháp tiền huấn luyện và chiến lược tinh chỉnh hạ nguồn, trong đó mô hình học có giám sát truyền thống đạt độ chính xác tham

chiếu là 65.25%. Trong nhóm các phương pháp học tự giám sát, MoCo v2 thể hiện tính ổn định và hiệu quả vượt trội nhất với độ chính xác khi tinh chỉnh toàn bộ đạt 64.87%, tiệm cận rất sát với mức tham chiếu, khẳng định rằng cơ chế cập nhật động lượng giúp mô hình duy trì được các đặc trưng chất lượng cao ngay cả khi dữ liệu gán nhãn hạn chế. Trái lại, phương pháp MAE bộc lộ sự phân hóa rõ rệt về hành vi học tập khi chỉ đạt độ chính xác rất thấp 28.81% ở chế độ đóng băng tham số do bản chất tập trung vào tái tạo điểm ảnh, nhưng lại có bước nhảy vọt lên 64.41% khi được phép tinh chỉnh toàn bộ trọng số mạng. Kết quả này, cùng với nghiên cứu [5], khẳng định rõ ràng rằng việc áp dụng học tự giám sát một cách rập khuôn lên các mô hình có kích thước nhỏ không chỉ không mang lại cải thiện mà thậm chí còn khiến hiệu suất suy giảm so với việc sử dụng mô hình được huấn luyện từ đầu hoặc chỉ tiền huấn luyện sơ bộ. Điều này củng cố thêm tính thách thức của bài toán và là cơ sở thực tiễn vững chắc cho việc lựa chọn MoCo v2 làm giải pháp nền tảng tối ưu cho nghiên cứu này.

Sự thiếu ổn định và hiệu quả hạn chế của MAE khi áp dụng lên các mạng CNN có thể được giải thích thông qua sự xung đột giữa chiến lược che khuất và định kiến quy nạp về tính cục bộ. Khác với Vision Transformers sử dụng cơ chế Self-Attention để thiết lập các mối tương quan tầm xa ngay từ các tầng đầu tiên, CNN vận hành dựa trên các cửa sổ trượt với trường thụ cảm hữu hạn. Đối với các mô hình CNN có kích thước nhỏ, giới hạn về chiều sâu và khả năng mở rộng trường thụ cảm khiến mạng khó có thể tổng hợp thông tin từ các vùng ảnh chưa bị che khuất để tái tạo các vùng bị trống ở khoảng cách xa. Do đó, thay vì học được các đặc trưng ngữ nghĩa cấp cao mang tính cấu trúc, mạng có xu hướng rơi vào trạng thái nội suy pixel dựa trên các cấu trúc texture cục bộ lân cận. Hệ quả là nhiệm vụ tái thiết trở nên quá khó khăn hoặc không đóng góp vào việc hình thành các biểu diễn trừu tượng, dẫn đến sự mất ổn định trong quá trình hội tụ và làm suy giảm khả năng tổng quát hóa của bộ trích xuất đặc trưng.

Bảng 6: Kết quả các phương pháp học giám sát và học tự giám sát trên bộ dữ liệu Plantdoc

Pretraining paradigm	Downtream task	Accuracy	Precision	Recall	F1-Score
Pretrained	Full fine-tuning	**65.25**	**64.98**	**62.4**	**63.66**
SimCLR	Frozen backbone	60.59	58.49	58.77	58.63
	Full fine-tuning	59.75	58.07	59.82	58.93
MoCoV2	Frozen backbone	62.71	59.35	61.53	60.42
	Full fine-tuning	64.87	61.7	65.79	63.68
MAE	Frozen backbone	28.81	39.72	28.62	33.27
	Full fine-tuning	64.41	67.29	64.92	66.08

Xuất phát từ đặc thù của kiến trúc mô hình đề xuất là sự tích hợp đồng thời hai cơ chế tối ưu hóa riêng biệt, bao gồm nhánh phân loại có giám sát và nhánh tương phản tự giám sát, việc xác định tỷ trọng đóng góp tối ưu của từng luồng tín hiệu là yếu tố then chốt quyết định sự hội tụ của mạng nơ-ron. Để làm rõ mức độ tác động và sự tương hỗ giữa hai hàm mục tiêu này, luận văn đã thiết kế một kịch bản thực nghiệm định lượng được tổng hợp chi tiết tại *Bảng 7*, tập trung khảo sát sự biến thiên hiệu năng của hệ thống khi thay đổi trọng số điều tiết trong hàm mất mát tổng hợp. Cụ thể, thí nghiệm giữ cố định trọng số beta của hàm mất mát phân loại ở giá trị một để duy trì tính ổn định của nhiệm vụ chính, đồng thời điều chỉnh trọng số alpha của hàm mất mát tương phản thay đổi tịnh tiến trong dải giá trị từ 0.3 đến 2.0. Mục tiêu của khảo sát này là tìm kiếm điểm cân bằng lý tưởng, nơi các đặc trưng bất biến từ học tự giám sát bổ trợ hiệu quả nhất cho biên quyết định phân loại mà không gây nhiễu loạn tín hiệu giám sát.

Kết quả thực nghiệm cho thấy một xu hướng biến thiên hình parabol rõ rệt của các chỉ số hiệu năng, với đỉnh cao nhất hội tụ tại giá trị alpha bằng 0.5. Tại thiết lập này, mô hình đạt được sự cộng hưởng tối ưu giữa hai nhánh với độ chính xác tổng thể là 69.49% và điểm F1 đạt 70.73%. Ngược lại, khi giá trị alpha được đẩy lên quá cao (từ 0.6 đến 2.0), hiệu năng hệ thống ghi nhận sự suy giảm dần đều, chạm đáy 64.83% tại alpha bằng 2.0. Điều này minh chứng rằng khi tín hiệu tự giám sát chiếm ưu thế quá lớn, nó có xu hướng lấn át tín hiệu định danh, buộc mô hình tập trung quá mức

vào việc phân biệt các biến thể ảnh cục bộ thay vì học các đặc trưng ngữ nghĩa tổng quát cần thiết cho việc phân loại bệnh. Do đó, việc lựa chọn alpha bằng 0.5 được xác định là cấu hình tiêu chuẩn cho các thử nghiệm so sánh tiếp theo của luận văn.

Bảng 7: Thực nghiệm hệ số alpha, beta của mô hình đề xuất trên tập dữ liệu plantdoc

alpha	Accuracy	Precision	Recall	F1-Score
alpha=0.3	66.53	69.62	66.2	67.87
alpha=0.4	66.95	69.24	66.83	68.01
alpha=0.5	**69.49**	**72.67**	**68.89**	**70.73**
alpha=0.6	66.53	68.75	66.41	67.56
alpha=0.7	66.53	69.53	66.54	68
alpha=0.8	65.68	67.22	65.19	66.19
alpha=0.9	65.25	66.98	64.68	65.81
alpha=1.0	65.68	68.21	65.3	66.72
alpha=1.5	66.1	71.3	66	68.55
alpha=2.0	64.83	67.97	64.72	66.31

Tiếp nối quy trình xây dựng dữ liệu dựa trên độ tương đồng ngữ nghĩa đã được đề xuất và hiện thực hóa tại Chương 3, luận văn tiến hành thực nghiệm đánh giá tác động thực tế của ba tập dữ liệu tự giám sát được phân cấp gồm tập Small, Medium và Large lên hiệu năng của mô hình phân loại. Các tập dữ liệu này được trích xuất từ nguồn dữ liệu thô hỗn hợp thông qua các ngưỡng lọc khác nhau, đại diện cho sự đánh đổi trực tiếp giữa độ tinh khiết của dữ liệu và sự đa dạng của các mẫu. Mục tiêu cốt lõi của kịch bản thử nghiệm này không chỉ dừng lại ở việc kiểm chứng khả năng tận dụng tài nguyên dữ liệu mở, mà quan trọng hơn là làm sáng tỏ mối quan hệ phi tuyến giữa số lượng mẫu và hiệu quả huấn luyện trong cơ chế học tương phản. Thông qua việc so sánh đối chiếu kết quả giữa các tập dữ liệu, luận văn nhằm khẳng định giả thuyết rằng trong bài toán học tự giám sát, việc kiểm soát nghiêm ngặt các tín hiệu nhiễu và đảm bảo độ tương đồng cao với tác vụ đích đóng vai trò quyết định hơn so với việc đơn thuần gia tăng số lượng mẫu đầu vào một cách ồ ạt và thiếu chọn lọc.

Bảng 8: Kết quả với các tập dữ liệu không giám sát trên kiến trúc mới

Dataset	Accuracy	Precision	Recall	F1-Score
PlantDoc	69.49	72.67	68.89	70.73
SSL small (3088)	**72.02**	**73.79**	**71.24**	**72.49**
SSL medium (3989)	70.34	71.94	69.96	70.94
SSL large (5621)	69.49	71.87	69.73	70.78

Kết quả định lượng được tổng hợp tại *Bảng 8* cho thấy một bức tranh rõ nét về sự tác động của chất lượng dữ liệu không nhãn đối với hiệu suất phân loại cuối cùng. Cụ thể, khi tích hợp tập dữ liệu SSL Nhỏ bao gồm 3.088 ảnh có độ tương đồng cao nhất với các nguyên mẫu bệnh, mô hình đạt được hiệu năng vượt trội trên mọi thước đo với độ chính xác tăng lên mức 72.02% và F1-Score đạt 72.49%, cải thiện đáng kể so với mức cơ sở 69.49% khi chỉ sử dụng dữ liệu PlantDoc gốc. Tuy nhiên, một xu hướng đáng chú ý là khi tiếp tục mở rộng không gian mẫu sang các tập Trung bình (3.989 ảnh) và Lớn (5.621 ảnh) bằng cách hạ thấp ngưỡng lọc, hiệu năng của mô hình bắt đầu có xu hướng suy giảm dần và quay trở lại mức ngang bằng với ban đầu. Hiện tượng này minh chứng rằng lợi ích từ việc gia tăng sự đa dạng biến thể đã bị triệt tiêu bởi sự xâm nhập của các tín hiệu nhiễu và các mẫu ngoại lai không sát với miền dữ liệu bệnh học mục tiêu. Điều này khẳng định tầm quan trọng của chiến lược tuyển chọn dựa trên nguyên mẫu, đồng thời chỉ ra rằng một tập dữ liệu có kích thước vừa phải nhưng đảm bảo độ "sạch" và sát ngữ nghĩa sẽ mang lại giá trị huấn luyện cao hơn nhiều so với một tập dữ liệu lớn nhưng chứa nhiều tạp nhiễu.

Ngoài ra, luận văn cũng tiến hành so sánh phương pháp đề xuất với các nghiên cứu liên quan được công bố trên cùng tập dữ liệu PlantDoc trên các thước đo hiệu năng truyền thống như Độ chính xác (Accuracy), Precision, Recall và F1-Score, luận văn đặc biệt chú trọng phân tích chỉ số về số lượng tham số (Parameters) để đánh giá tính hiệu quả về mặt tài nguyên – một yếu tố then chốt khi triển khai mô hình trên các thiết bị biên trong nông nghiệp thông minh. Các số liệu thống kê chi tiết về hiệu năng của các mô hình được tổng hợp và trình bày tại *Bảng 9*.

Bảng 9: Hiệu suất các mô hình trên tập dữ liệu plantdoc

Model	Accuracy	Precision	Recall	F1-Score	Params (M)
DenseNet121 [11]	61.01	-	-	-	8
ResNet152 [11]	60.16	-	-	-	60.2
VGG16 [11]	58.05	-	-	-	**138.4**
Ensemble (CNNs) [12]	68	69	69	68	-
ViT-Base [12]	69	61	59	60	-
InceptionResNet V2 [13]	70.53	-	-	-	55.8
Mô hình đề xuất	**72.02**	**73.79**	**71.24**	**72.49**	5.4

Phân tích số liệu từ bảng so sánh cho thấy mô hình đề xuất đã xác lập được một ưu thế vượt trội về sự cân bằng giữa hiệu suất nhận diện và chi phí tài nguyên khi so sánh với các nghiên cứu liên quan. Với độ chính xác đạt đỉnh 72.02% và chỉ số F1 đạt 72.49%, giải pháp tích hợp học tự giám sát MoCo v2 trên nền tảng mạng nhẹ không chỉ vượt qua các kiến trúc CNN kinh điển có độ sâu lớn như VGG16 hay ResNet152 mà còn chiến thắng cả những mô hình phức tạp hơn như InceptionResNet V2 hay Vision Transformer với khoảng cách đáng kể. Đáng chú ý nhất là sự chênh lệch khổng lồ về số lượng tham số; trong khi các mô hình đối chứng đòi hỏi từ 55 triệu đến gần 140 triệu tham số để đạt độ chính xác dao động quanh mức 58-70%, mô hình đề xuất chỉ cần vỏn vẹn 5.4 triệu tham số để đạt hiệu năng cao nhất. Điều này minh chứng sắc bén rằng trong bối cảnh dữ liệu gán nhãn khan hiếm và nhiễu tạp như PlantDoc, chiến lược sử dụng một mô hình gọn nhẹ kết hợp với không gian biểu diễn đặc trưng chất lượng cao từ dữ liệu không nhãn mang lại hiệu quả thực tế tốt hơn việc gia tăng kích thước mạng một cách cơ học, đồng thời mở ra khả năng ứng dụng khả thi trên các thiết bị biên hạn chế tài nguyên.

4.4. Kết luận chương 4

Chương 4 đã hoàn tất việc kiểm chứng thực nghiệm mô hình lai ghép trên nền tảng phần cứng GPU P100, sử dụng bộ tiêu chí đánh giá toàn diện gồm Accuracy, Precision, Recall và F1-score để đảm bảo tính khách quan khoa học trong bối cảnh dữ liệu PlantDoc đầy thách thức. Các kết quả so sánh chiến lược tiền huấn luyện đã chỉ ra rằng MoCo v2 là phương pháp tối ưu nhất cho các mạng xương sống hạng nhẹ, đạt độ chính xác tinh chỉnh 64.87% và khắc phục hoàn toàn sự bất ổn định của phương

pháp MAE vốn thường thất bại trên các mô hình nhỏ. Thông qua quá trình tinh chỉnh siêu tham số, thực nghiệm đã xác định được điểm cân bằng lý tưởng giữa nhánh giám sát và tự giám sát tại trọng số alpha=0.5, giúp hệ thống đạt độ chính xác cơ sở là 69.49%. Một trong những phát hiện quan trọng nhất của chương là việc chứng minh bằng số liệu rằng chất lượng dữ liệu không nhãn đóng vai trò quyết định hơn số lượng; cụ thể, tập dữ liệu Small được lọc sạch nhiễu dựa trên độ tương đồng ngữ nghĩa đã giúp mô hình đạt hiệu suất đỉnh với độ chính xác 72.02% và F1-score 72.49%. Ngược lại, việc nới lỏng ngưỡng lọc để gia tăng kích thước dữ liệu trong các tập Medium và Large lại dẫn đến sự suy giảm hiệu năng, khẳng định rằng việc kiểm soát nhiễu trong học tự giám sát là yếu tố then chốt. Khi đặt lên bàn cân so sánh với các nghiên cứu tiên tiến khác, giải pháp đề xuất đã thể hiện sự vượt trội ấn tượng trước các mô hình hạng nặng như VGG16, ResNet152 hay thậm chí là Vision Transformer. Điểm đột phá nằm ở chỗ hiệu suất vượt trội này đạt được chỉ với 5.4 triệu tham số, thấp hơn hàng chục lần so với con số 58 đến 139 triệu tham số của các mô hình đối chứng, qua đó chứng minh tính hiệu quả về mặt tài nguyên. Điều này minh chứng sắc bén cho luận điểm rằng việc tối ưu hóa không gian biểu diễn đặc trưng thông qua học tương phản hiệu quả hơn nhiều so với việc gia tăng độ sâu của mạng một cách cơ học trong điều kiện thiếu dữ liệu.

KẾT LUẬN

Trên cơ sở nghiên cứu bài toán phân loại bệnh thực vật và nền tảng lý thuyết về thị giác máy tính hiện đại, luận văn đã hoàn thành các mục tiêu đề ra ban đầu và đạt được những kết quả khả quan. Cụ thể, nghiên cứu đã khái quát hóa được bức tranh toàn cảnh về tầm quan trọng của nông nghiệp chính xác, đồng thời chỉ ra điểm nghẽn cốt lõi trong các mô hình học sâu hiện nay là sự phụ thuộc quá lớn vào dữ liệu gán nhãn thủ công. Thông qua việc tổng hợp và phân tích các phương pháp tiên tiến, luận văn đã đề xuất và hiện thực hóa thành công một kiến trúc lai ghép, tích hợp cơ chế học tự giám sát MoCov2 với các mạng xương sống gọn nhẹ. Điểm đóng góp nổi bật của đề tài là việc xây dựng quy trình tuyển chọn dữ liệu không nhãn dựa trên độ tương đồng ngữ nghĩa với các nguyên mẫu lớp, cho phép tận dụng nguồn dữ liệu mở khổng lồ để cải thiện không gian biểu diễn đặc trưng mà không tốn kém chi phí gán nhãn. Kết quả thực nghiệm đã chứng minh giải pháp đề xuất đạt hiệu năng vượt trội so với các phương pháp học giám sát truyền thống và các kiến trúc phức tạp khác trên tập dữ liệu PlantDoc, đặc biệt trong bối cảnh hạn chế về dữ liệu mẫu.

Bên cạnh những thành tựu đã đạt được, quá trình nghiên cứu vẫn tồn tại một số hạn chế nhất định cần được khắc phục trong tương lai. Mặc dù chiến lược tuyển chọn dữ liệu tự giám sát đã phát huy hiệu quả với các tập dữ liệu nhỏ và trung bình, nhưng việc mở rộng quy mô lên các tập dữ liệu lớn với ngưỡng lọc thấp đã dẫn đến sự xuất hiện của nhiễu tín hiệu, gây ảnh hưởng tiêu cực đến hiệu suất hội tụ của mô hình. Thêm vào đó, quá trình tinh chỉnh các siêu tham số trọng số trong hàm mất mát tổng hợp mới chỉ dừng lại ở các khảo sát thực nghiệm rời rạc, chưa áp dụng các thuật toán tối ưu hóa tự động để tìm ra cấu hình lý tưởng nhất cho từng kịch bản cụ thể. Ngoài ra, dù mô hình đề xuất có kiến trúc gọn nhẹ, quy trình huấn luyện giai đoạn tiền tự giám sát vẫn đòi hỏi tài nguyên tính toán đáng kể, tạo ra rào cản nhất định khi muốn cập nhật tri thức liên tục cho hệ thống ngay tại biên.

Từ những kết quả và hạn chế đã phân tích, có thể thấy việc phát triển một hệ thống chẩn đoán bệnh cây trồng tự động, bền vững và chi phí thấp là một quá trình nghiên cứu lâu dài, đòi hỏi sự cân bằng tinh tế giữa độ chính xác thuật toán và hiệu

quả tài nguyên. Trong phạm vi của luận văn, nghiên cứu đã chứng minh được tính khả thi của việc thay thế sự giám sát của con người bằng tri thức trích xuất từ cấu trúc dữ liệu, mở ra một hướng tiếp cận đầy tiềm năng cho các bài toán nông nghiệp thiếu dữ liệu. Hướng phát triển tiếp theo của đề tài sẽ tập trung vào việc cải tiến thuật toán lọc nhiễu cho dữ liệu không nhãn nhằm mở rộng quy mô huấn luyện mà không làm suy giảm chất lượng đặc trưng. Đồng thời, nghiên cứu cũng sẽ hướng tới việc tối ưu hóa quy trình triển khai thực tế thông qua các kỹ thuật nén mô hình như chưng cất kiến thức hoặc lượng tử hóa, giúp hệ thống có thể vận hành mượt mà trên các thiết bị di động hoặc thiết bị bay không người lái với năng lực tính toán hạn chế.

Thông qua đề tài này, luận văn mong muốn đóng góp một giải pháp thực tiễn cho bài toán cấp thiết về bảo vệ thực vật trong kỷ nguyên số. Các kỹ thuật xử lý ảnh và học tự giám sát được trình bày không chỉ có ý nghĩa trong phạm vi hẹp của bài toán phân loại bệnh cây trồng mà còn có giá trị tham khảo cho cộng đồng nghiên cứu khi đối mặt với các vấn đề khan hiếm dữ liệu trong các lĩnh vực khác. Những kinh nghiệm thực nghiệm và cơ sở lý luận được đúc kết từ nghiên cứu này sẽ là tiền đề vững chắc để phát triển các hệ thống nông nghiệp thông minh toàn diện hơn, hỗ trợ đắc lực cho người nông dân trong việc giám sát mùa màng và nâng cao năng suất canh tác.

TÀI LIỆU THAM KHẢO

[1] Sornapudi, S., & Singh, R. (2024). Self-supervised backbone framework for diverse agricultural vision tasks. *arXiv preprint arXiv:2403.15248*.

[2] Güldenring, R., & Nalpantidis, L. (2021). Self-supervised contrastive learning on agricultural images. *Computers and Electronics in Agriculture*, *191*, 106510.

[3] Kim, T., Kim, H., Baik, K., & Choi, Y. (2022). Instance-aware plant disease detection by utilizing saliency map and self-supervised pre-training. *Agriculture*, *12*(8), 1084.

[4] Al Mamun, A., Ahmedt-Aristizabal, D., Zhang, M., Hossen, M. I., Hayder, Z., & Awrangjeb, M. (2024). Plant Disease Detection Using Self-supervised Learning: A Systematic Review. *IEEE Access*.

[5] Fang, Z., Wang, J., Wang, L., Zhang, L., Yang, Y., & Liu, Z. (2021). Seed: Self-supervised distillation for visual representation. *arXiv preprint arXiv:2101.04731*.

[6] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (2002). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278-2324.

[7] Singh, D., Jain, N., Jain, P., Kayal, P., Kumawat, S., & Batra, N. (2020). PlantDoc: A dataset for visual plant disease detection. In Proceedings of the 7th ACM IKDD CoDS and 25th COMAD (pp. 249-253).

[8] Wei, T., Chen, Z., Huang, Z., & Yu, X. (2024, October). Benchmarking in-the-wild multimodal disease recognition and a versatile baseline. In Proceedings of the 32nd ACM International Conference on Multimedia (pp. 1593-1601).

[9] Wei, T., Chen, Z., Yu, X., Chapman, S., Melloy, P., & Huang, Z. (2024). Plantseg: A large-scale in-the-wild dataset for plant disease segmentation. arXiv preprint arXiv:2409.04038.

[10] Moupojou, E., Tagne, A., Retraint, F., Tadonkemwa, A., Wilfried, D., Tapamo, H., & Nkenlifack, M. (2023). FieldPlant: A dataset of field plant images for plant disease detection and classification with deep learning. IEEE Access, 11, 35398-35410.

[11] Balafas, V., Karantoumanis, E., Louta, M., & Ploskas, N. (2023). Machine learning and deep learning for plant disease classification and detection. *IEEE Access*, *11*, 114352-114377.

[12] Salman, Z., Muhammad, A., & Han, D. (2025). Plant disease classification in the wild using vision transformers and mixture of experts. *Frontiers in Plant Science*, *16*, 1522985.

[13] Sama, N., David, E., Rossetti, S., Antona, A., Franchetti, B., & Pirri, F. (2023). A new large dataset and a transfer learning methodology for plant phenotyping in Vertical Farms. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 540-551).

[14] Mu, J., Feng, Q., Yang, J., Zhang, J., & Yang, S. (2024). Few-shot disease recognition algorithm based on supervised contrastive learning. *Frontiers in plant science*, *15*, 1341831.

[15] Pham, T. C., Nguyen, T. N., & Nguyen, V. D. (2025). Ambiguity-aware semi-supervised learning for leaf disease classification. *Scientific Reports*, *15*(1), 14070.