

HỌC VIỆN NÔNG NGHIỆP VIỆT NAM

LÝ THỊ PHƯƠNG

**ỨNG DỤNG MẠNG DEEP BOLTZMANN MACHINE
NHẬN BIẾT KHUÔN MẶT ĐEO KHẨU TRANG**

Ngành: Công nghệ thông tin
Mã số: 8 48 02 01
Người hướng dẫn: TS. Nguyễn Trọng Kương

NHÀ XUẤT BẢN HỌC VIỆN NÔNG NGHIỆP – 2024

LỜI CAM ĐOAN

Tôi xin cam đoan đây là công trình nghiên cứu của riêng tôi, các kết quả nghiên cứu được trình bày trong luận văn là trung thực, khách quan và chưa từng dùng để bảo vệ lấy bất kỳ học vị nào.

Tôi xin cam đoan rằng mọi sự giúp đỡ cho việc thực hiện luận văn đã được cảm ơn, các thông tin trích dẫn trong luận văn này đều được chỉ rõ nguồn gốc.

Hà Nội, ngày... tháng... năm...

Tác giả luận văn

Lý Thị Phượng

LỜI CẢM ƠN

Trong suốt quá trình học tập, nghiên cứu và thực hiện đề tài luận văn thạc sĩ, ngoài những cố gắng nỗ lực của bản thân, đề tài này còn có sự hướng dẫn nhiệt thành của quý Thầy cô, cùng với sự động viên khích lệ và ủng hộ của gia đình, bạn bè đồng nghiệp.

Lời đầu tiên Em xin cảm ơn chân thành và sâu sắc đến Thầy TS Nguyễn Trọng Kương đã dành rất nhiều thời gian hướng dẫn Em một cách tận tâm và sâu sát để giúp Em hoàn thành tốt luận văn này.

Em xin chân thành cảm ơn toàn thể thầy cô khoa công nghệ thông tin, Học viện Nông nghiệp Việt nam đã tận tình truyền đạt những bài học quý báu trong suốt quá trình học tập tại trường.

Cuối cùng, mặc dù đã có nhiều nỗ lực, song do thời gian và kinh nghiệm nghiên cứu khoa học còn hạn chế nên không thể tránh khỏi những thiếu sót. Em rất mong nhận được sự góp ý của các thầy cô và bạn bè đồng nghiệp để hiểu biết của mình ngày một hoàn thiện hơn.

Hà Nội, ngày... tháng... năm 20...

Tác giả luận văn

Lý Thị Phượng

MỤC LỤC

Lời cam đoan	i
Lời cảm ơn	ii
Mục lục	iii
Danh mục chữ viết tắt	v
Danh mục bảng	vi
Danh mục hình	vii
Danh mục sơ đồ	viii
Trích yếu luận văn	ix
Thesis abstract.....	xi
Phần 1. Mở đầu	1
1.1. Tính cấp thiết của đề tài.....	1
1.2. Mục tiêu nghiên cứu	2
1.3. Phương pháp nghiên cứu	2
1.4. Ý nghĩa khoa học và ý nghĩa thực tiễn	3
1.4.1. Ý nghĩa khoa học	3
1.4.2. Ý nghĩa thực tiễn	3
Phần 2. Bài toán và dữ liệu nguyên cứu	4
2.1. Nghiên cứu về nhận diện khuôn mặt	4
2.2. Bài toán nhận diện đeo khẩu trang	7
2.3. Tình hình nghiên cứu trong và ngoài nước.....	8
2.3.1. Tình hình nghiên cứu trong nước	8
2.3.2. Tình hình nguyên cứu ngoài nước	10
2.4. Nghiên cứu dữ liệu thu thập được	12
2.4.1. Số lượng ảnh, người tham gia lấy mẫu.....	12
2.4.2. Đặc trưng mẫu	14
Phần 3. Mạng boltzmann machine.....	17
3.1. Giới thiệu về học sâu	17
3.1.1. Khái niệm về học sâu.....	17
3.1.2. Mạng thần kinh nhân tạo	17

3.1.3.	Các thuật toán huấn luyện	22
3.2.	Giới thiệu về mạng restricted Boltzmann machine	25
3.2.1.	Mạng restricted Boltzmann machine	25
3.2.2.	Mạng classification Restricted Boltzmann machine	28
3.3.	Giới thiệu về mạng Deep Boltzmann Machine	29
Phần 4. Kết quả nghiên cứu.....		34
4.1.	Tiền xử lý dữ liệu nghiên cứu.....	34
4.1.1.	Thu thập dữ liệu.....	34
4.1.2.	Tiền xử lý dữ liệu	35
4.2.	Cấu hình mạng Deep Boltzmann Machine.....	37
4.2.1.	Biên dịch dữ liệu.....	38
4.2.2.	Huấn luyện, kiểm tra mô hình	39
4.3.	Kết quả nhận diện	41
4.3.1.	Ứng dụng nhận diện khuôn mặt đeo khẩu trang.....	43
4.3.2.	Cấu trúc mô hình nhận diện khuôn mặt đeo khẩu trang.....	43
Phần 5. Kết luận		48
5.1.	Kết luận.....	48
5.2.	Hướng phát triển.....	48
Tài liệu tham khảo		50

DANH MỤC CHỮ VIẾT TẮT

Chữ viết tắt	Nghĩa tiếng Việt
DL	Học sâu
AI	Trí tuệ nhân tạo.
CNN	Mạng lưới thần kinh tích chập.
DBM	Máy học Boltzmann sâu.
ML	Học máy.
NIST	Viện Tiêu chuẩn và Công nghệ.
NN	Mạng lưới thần kinh.
RBM	Máy học Boltzmann hạn chế.
WHO	Tổ chức Y tế Thế giới.
MTCNN	Mạng lưới thần kinh tích chập xếp tầng đa nhiệm.
RMFRD	Bộ dữ liệu nhận dạng khuôn mặt đeo khẩu trang trong thực tế.
PLA	Thuật toán học Perceptron.

DANH MỤC BẢNG

Bảng 3.1. Một số hàm kích hoạt.....	21
Bảng 4.1. Dữ liệu thu thập từ sinh viên tình nguyện tại Học viện Nông nghiệp Việt Nam.....	35
Bảng 4.2. Kết quả độ chính xác trong huấn luyện và kiểm tra của DBM.....	41
Bảng 4.3. Kết quả độ chính xác trong nhận diện khuôn mặt và khẩu trang của DBM , NN và SVM	47

DANH MỤC HÌNH

Hình 2.1.	Khuôn mặt với các các điểm đặc trưng.....	5
Hình 2.2.	Biểu đồ biểu diễn của bộ lọc Gaussian trong thời gian biến thiên.....	15
Hình 3.1.	Mạng perceptron đơn	18
Hình 3.2.	Mô hình mạng NN cơ bản.....	19
Hình 3.3.	Cấu trúc mạng restricted Boltzmann Machine.....	25
Hình 3.4.	Cấu trúc mạng classification Restricted Boltzmann Machine	29
Hình 3.5.	Mô hình DBM được sắp xếp thành đồ thị.....	31
Hình 4.1.	Hình ảnh đã qua tiền xử lý	36
Hình 4.2.	Sơ đồ mô hình Deep Boltzmann Machine	37
Hình 4.3.	Kết quả nhận diện khuôn mặt đeo khẩu trang trên màn hình của mô hình Deep Boltzmann Machine.....	46

DANH MỤC SƠ ĐỒ

Sơ đồ 4.1. Sơ đồ hệ thống nhận diện khuôn mặt đeo khẩu trang	42
--	----

TRÍCH YẾU LUẬN VĂN

Tên tác giả: Lý Thị Phượng

Tên Luận văn: Ứng dụng mạng Deep Boltzmann Machine nhận biết khuôn mặt đeo khẩu trang.

Ngành: Công nghệ thông tin (ứng dụng)

Mã số: 8 48 02 01

Tên cơ sở đào tạo: Học viện Nông nghiệp Việt Nam

Mục đích nghiên cứu

Xây dựng mô hình nhận diện khuôn mặt, đào tạo hệ thống và ứng dụng mạng DBM cho công nghệ phát hiện khẩu trang trên ảnh, hình thành cơ sở khoa học cho việc cải thiện hệ thống nhận diện khuôn mặt từ mô hình học máy đơn giản sang mô hình học sâu nhằm cải thiện tính năng phát hiện, mở rộng phạm vi nghiên cứu và nâng cao hiệu quả trong việc xây dựng ứng dụng nhận diện khuôn mặt kết hợp việc đeo khẩu trang.

Phương pháp nghiên cứu

Nội dung nghiên cứu trong luận văn gồm:

- Tổng quan lý thuyết chung nghiên cứu nhận diện khuôn mặt và thu thập dữ liệu về bài toán nhận biết đeo khẩu trang.
- Khái quát kiến thức mạng học sâu, xây dựng mô hình DBM cho nghiên cứu nhận biết khuôn mặt đeo khẩu trang.
- Nghiên cứu kết quả mô hình mạng DBM và cải thiện độ chính xác của nhận dạng cho bài toán thực tế.

Luận văn đã sử dụng các phương pháp nghiên cứu sau:

- Phương pháp phân tích, tổng hợp: phân tích, tổng hợp tài liệu liên quan đến nhận diện, mạng DBM, tác dụng của khẩu trang từ các bài viết, các tạp chí công nghệ như "Nhận diện người đeo khẩu trang bằng thời gian thực tế" của tác giả Bùi Xuân Quang, Các nhân tố ảnh hưởng đến ý định đeo khẩu trang của người dân tại nơi công cộng để phòng chống dịch Covid-19 - Nghiên cứu thực nghiệm tại Thành phố Hồ Chí Minh của tác giả Trần Phạm Khánh Toàn và một số bài báo khác.
- Phương pháp thu thập dữ liệu: dữ liệu được sự đồng ý và xử lý của 30 sinh viên Học viện Nông nghiệp Việt Nam để chọn ra 300 hình ảnh khuôn mặt và đeo khẩu trang trên từng góc độ khác nhau để ứng dụng vào mô hình nhận biết hình ảnh.
- Phương pháp mô hình hóa: Từ các kiến thức đã học, đã biết và sự hướng dẫn của Thầy Tiến sĩ Nguyễn Trọng Kương, cũng như các kiến thức chuyên môn của các chuyên gia trong lĩnh vực đã xây dựng thành công hệ thống mô hình DBM cho nhận biết khuôn mặt và khẩu trang, từ đó áp dụng vào công nghệ nhận diện khuôn mặt.

Kết quả chính và kết luận

Sau khi hoàn thiện bài luận văn, nghiên cứu đã đúc kết tổng quan về lý thuyết và ứng dụng của mạng học sâu DBM vào công nghệ nhận diện khuôn mặt đeo khẩu trang. Từ đó xây dựng và hoàn thiện bài toán nhận biết khẩu trang vào đời sống thực tế.

Nghiên cứu nêu nên sự quan trọng của việc đeo khẩu trang trong bối cảnh phòng chống dịch bệnh và thời tiết cực đoan đang xảy ra, đồng thời đóng góp vào công việc tăng cường an ninh an toàn tại các nơi đông đúc, cơ quan làm việc, nơi làm việc có tính ô nhiễm cao.

Tìm hiểu về lợi ích, ưu nhược điểm của mạng học sâu DBM khi triển khai trên bài toán nhận diện và đưa ra những thuận lợi, khó khăn mà mô hình đưa ra.

Thực hiện nghiên cứu thử nghiệm mô hình mạng DBM trên camera trên máy tính, từ đó đánh giá độ chính xác của mô hình dựa trên các kết quả mong muốn và hiệu chỉnh mô hình DBM phù hợp cho các tiến trình chạy thử ứng dụng. So sánh các kết quả độ chính xác của các mô hình mạng khác nhau như NN và SVM cho bài toán nhận biết khuôn mặt đeo khẩu trang, đưa ra kết luận rằng mô hình DBM là phù hợp với tiêu đề bài toán được đề ra.

Dữ liệu đầu vào của mô hình vẫn còn là một trong các hạn chế để xử lý và nhận diện chính xác, ngoài ra cần thay đổi và xây dựng các cấu trúc khác trong mô hình DBM nhằm đạt được kết quả tốt nhất trong nhận diện hình ảnh.

Tạo dựng chức năng chụp và lưu trữ dữ liệu mới khi thực hiện quá trình nhận diện, cải thiện số lượng dữ liệu và đạt kết quả chính xác tốt nhất cho mô hình học.

THESIS ABSTRACT

Master candidate: Lý Thị Phương

Thesis title: Application of Deep Boltzmann Machine for Recognizing Face Wearing Masks.

Major: Information Technology

Code: 8 48 02 01

Education organization: Vietnam National University of Agriculture

Research Objectives

Developing a facial recognition model, training the system, and applying the DBM network for mask detection technology in images, establishing a scientific basis for improving the facial recognition system from simple machine learning models to deep learning models to enhance detection features, expand research scope, and improve the effectiveness in building facial recognition applications that incorporate mask detection.

Materials and Methods

Research questions concern the following issues:

- The theoretical background on facial recognition and data collection for the problem of mask detection.
- Deep learning network knowledge, and the development of DBM model for researching masked face detection.
- DBM model results and recognition accuracy for practical application.

The thesis utilized the following research methods:

- Analytical and synthetic: analysis and synthesis documents related to facial recognition, DBM networks, and the effectiveness of mask from articles and technology journals such as “Real-time Recognition of Masked Individuals” by author Bui Xuan Quang, “Factors Affecting the Intention to Wear Masks among the Public Places to Prevent Covid-19 - An Empirical Study in Ho Chi Minh City” by author Tran Pham Khanh Toan and other articles.
- Data collection: data was collected with the consent and processing from 30 students of Vietnam National University of Agriculture to select 300 images of faces and faces with masks from different angles for application in the image recognition model.
- Modeling method: Base on learned knowledge, known, and under the guidance of DR Nguyen Trong Kuong, a successful DBM model recognizing faces and masks was developed, which was applied to facial detection technology.

Main findings and conclusions

After conducting the thesis, the research summarized an overview of the theory and

application of DBM networks in the technology of facial recognition with masks which developed and refined the problem of mask detection in real life.

The research mentions the importance of wearing masks in the context of disease prevention and extreme weather conditions, which enhances security and safety in crowded places, workplaces, and areas with high pollution levels.

The benefits and drawbacks of DBM networks were explored when deployed in facial recognition tasks, discussing the advantages and challenges presented by the model.

The DBM network model was experimented on a computer camera, then evaluated the accuracy of the model based on desired outcomes and adjusted it.

The accuracy results of different network models such as NN and SVM were compared for the problem of recognizing faces wearing masks, concluding that the DBM model is suitable for the problem.

Input data is one of the problems for accurate processing and detection, and the DBM model needs to change and construct different structures for the best results in image recognition.

A function should be developed to capture and store new data during the detection process, improving the amount of data and achieving the best possible accuracy for the learning model.

PHẦN 1. MỞ ĐẦU

1.1. TÍNH CẤP THIẾT CỦA ĐỀ TÀI

Trong bối cảnh của thời đại chuyển đổi số, việc áp dụng công nghệ vào đời sống xã hội, nhất là trong việc xác định những người đang đeo khẩu trang nơi công cộng, là công việc quan trọng để xác thực nhận dạng chính xác và duy trì khoảng cách an toàn ngăn chặn tiếp xúc gần với người khác. Trong đó, nhận diện khẩu trang là để xác định liệu một người có đang đeo khẩu trang hay không. Thực tế, vấn đề này là kỹ thuật đảo ngược của việc phát hiện khuôn mặt, nơi khuôn mặt được phát hiện bằng các thuật toán học máy khác nhau cho mục đích an ninh, xác thực và giám sát. Phát hiện khuôn mặt là một lĩnh vực then chốt trong lĩnh vực Thị giác máy tính và Nhận dạng mẫu.

Công nghệ nhận dạng có thể giúp xác định những người đeo khẩu trang trong không gian công cộng, hỗ trợ các nhà quản lý và nhân viên an ninh trong việc đảm bảo an ninh và tuân thủ quy định về y tế. Bên cạnh những lợi ích về sức khỏe, việc nhận diện người đeo khẩu trang còn đóng vai trò thiết yếu trong các ứng dụng liên quan đến an ninh, từ việc ngăn chặn những người khả nghi xuất hiện tại các cơ quan, ngân hàng cho đến việc giám sát các hoạt động đáng ngờ của đối tượng.

Mặc dù vậy tại những nơi tập trung nhiều người như cơ quan, trường học, khu chợ,... vẫn còn xuất hiện tình trạng có người không đeo khẩu trang, hoặc có mang theo bên người nhưng không đeo lên trên mặt. Nhiều người đeo khẩu trang nhưng không đúng quy định, không có tác dụng phòng chống các bệnh truyền nhiễm và bụi bẩn, tiềm ẩn nhiều nguy cơ mắc các bệnh truyền nhiễm.

Với sự phát triển nhanh chóng của mạng học sâu (một nhánh con của học máy), đặc biệt là mô hình DBM, tôi xin chọn đề tài “Ứng dụng mạng deep Boltzmann machine nhận biết khuôn mặt đeo khẩu trang” việc xây dựng mô hình học sâu để nhận dạng người đeo khẩu trang là cực kỳ quan trọng để giải quyết các vấn đề về nhận diện khẩu trang trên khuôn mặt, thúc đẩy sự tuân thủ quy định, đảm bảo an ninh và nâng cao hiệu suất tổng thể của giám sát và theo dõi trong không gian công cộng. Nó đóng góp quan trọng vào sự an toàn và phát triển của cộng đồng, đặc biệt là trong bối cảnh các tình huống về sức khỏe công cộng và an ninh.

1.2. MỤC TIÊU NGHIÊN CỨU

Xây dựng mô hình nhận diện khuôn mặt, đào tạo hệ thống và ứng dụng mạng DBM cho công nghệ phát hiện khẩu trang trên ảnh, hình thành cơ sở khoa học cho việc cải thiện hệ thống nhận diện khuôn mặt từ mô hình học máy đơn giản sang mô hình học sâu nhằm cải thiện tính năng phát hiện, mở rộng phạm vi nghiên cứu và nâng cao hiệu quả trong việc xây dựng ứng dụng nhận diện khuôn mặt kết hợp việc đeo khẩu trang.

1.3. PHƯƠNG PHÁP NGHIÊN CỨU

Nội dung nghiên cứu trong luận văn gồm:

Tổng quan lý thuyết chung nghiên cứu nhận diện khuôn mặt và thu thập dữ liệu về bài toán nhận biết đeo khẩu trang.

Khái quát kiến thức mạng học sâu, xây dựng mô hình DBM cho nghiên cứu nhận biết khuôn mặt đeo khẩu trang.

Nghiên cứu kết quả mô hình mạng DBM và cải thiện độ chính xác của nhận dạng cho bài toán thực tế.

Luận văn đã sử dụng các phương pháp nghiên cứu sau:

Phương pháp phân tích, tổng hợp: Phân tích, tổng hợp tài liệu liên quan đến nhận diện, mạng DBM, tác dụng của khẩu trang từ các bài viết, các tạp chí công nghệ như "Nhận diện người đeo khẩu trang bằng thời gian thực tế" của tác giả Bùi Xuân Quang, Các nhân tố ảnh hưởng đến ý định đeo khẩu trang của người dân tại nơi công cộng để phòng chống dịch Covid-19 - Nghiên cứu thực nghiệm tại Thành phố Hồ Chí Minh của tác giả Trần Phạm Khánh Toàn và một số bài báo khác.

Phương pháp thu thập dữ liệu: Dữ liệu được sự đồng ý và xử lý của 30 sinh viên Học viện Nông nghiệp Việt Nam để chọn ra 300 hình ảnh khuôn mặt và đeo khẩu trang trên từng góc độ khác nhau để ứng dụng vào mô hình nhận biết hình ảnh.

Phương pháp mô hình hóa: Từ các kiến thức đã học, đã biết và sự hướng dẫn của Thầy Tiến sĩ Nguyễn Trọng Kương, cũng như các kiến thức chuyên môn của các chuyên gia trong lĩnh vực đã xây dựng thành công hệ thống mô hình DBM cho nhận biết khuôn mặt và khẩu trang, từ đó áp dụng.

1.4. Ý NGHĨA KHOA HỌC VÀ Ý NGHĨA THỰC TIỄN

1.4.1. Ý nghĩa khoa học

Nghiên cứu về bài toán áp dụng mạng DBM để nhận diện khuôn mặt đeo khẩu trang không chỉ mang lại ứng dụng thực tế mà còn mở ra cánh cửa cho việc hiểu sâu hơn về tiềm năng và hạn chế của DL trong vấn đề giải quyết các thách thức đặc biệt như vấn đề che khuất một phần khuôn mặt và giảm đi khả năng nhận diện trên khuôn mặt con người.

Ngoài ra nghiên cứu không chỉ tập trung vào việc nhận diện khuôn mặt mà còn là tìm hiểu cách mà hệ thống học sâu có thể tự học và thích nghi với những biến đổi như đeo khẩu trang trên khuôn mặt. Điều này đặt ra những câu hỏi sâu rộng hơn về cách mà mô hình học sâu có thể hiểu và biểu diễn những đặc điểm tối ưu trong những tình huống cụ thể về nhận diện.

1.4.2. Ý nghĩa thực tiễn

Nâng cao hiệu suất hệ thống an ninh và quản lý an toàn: Bài toán nhận diện có thể tích hợp vào các hệ thống an ninh hiện đại, giúp nhận diện và theo dõi người dùng một cách chính xác ngay cả khi họ đeo khẩu trang. Điều này vô cùng quan trọng để duy trì an ninh tại các khu vực như trung tâm mua sắm, nhà ga, sân bay, hay các địa điểm công cộng khác.

Tăng cường quản lý an toàn: Công nghệ nhận diện khẩu trang còn có thể ứng dụng trong quản lý mức độ an toàn, đặc biệt là trong các lĩnh vực y tế, nơi mà tập trung nhiều loại bệnh truyền nhiễm và vi khuẩn từ người bệnh cần phải được giám sát khi đeo khẩu trang để phòng tránh các bệnh truyền nhiễm khi tiếp xúc gần.

Cải thiện trải nghiệm người dùng: Ở các lĩnh vực như công nghệ sinh trắc học, việc sử dụng công nghệ nhận diện khuôn mặt dưới lớp khẩu trang có thể cải thiện tính năng nhận diện của các đồ dùng công nghệ điện tử thông qua xác thực khuôn mặt để mở khóa, thanh toán nhanh từ khuôn mặt và xác định được đối tượng mà không cần tháo khẩu trang.

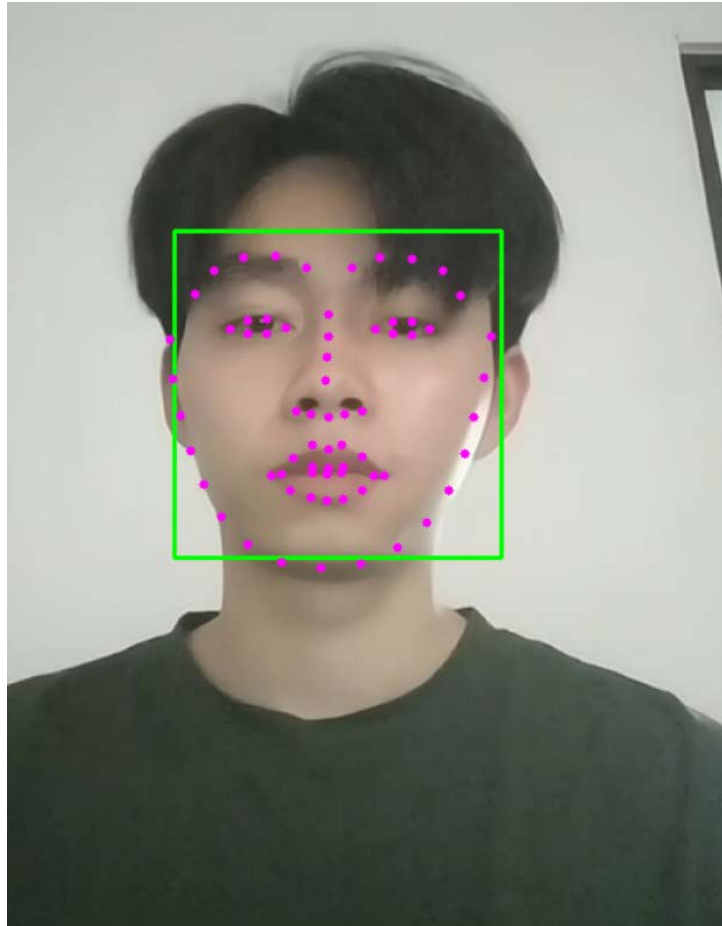
PHẦN 2. BÀI TOÁN VÀ DỮ LIỆU NGUYÊN CỨU

2.1. NGHIÊN CỨU VỀ NHẬN DIỆN KHUÔN MẶT

Nhận dạng khuôn mặt (Face Recognition) được nghiên cứu từ những năm 1980, là một lĩnh vực nghiên cứu của ngành thị giác máy tính, và cũng được xem là một lĩnh vực nghiên cứu của ngành sinh trắc học tương tự như nhận dạng vân tay, hay nhận dạng móng mắt. Trong khi nhận dạng vân tay và móng mắt có thể áp dụng trên thực tế một cách rộng rãi thì nhận dạng mặt người vẫn còn nhiều thách thức. So với nhận dạng vân tay và móng mắt, nhận dạng mặt có nguồn dữ liệu phong phú hơn và ít đòi hỏi sự tương tác có kiểm soát hơn.

Nhận dạng khuôn mặt là một phương pháp sinh trắc học để xác định và xác minh một cá nhân nào đó bằng cách so sánh một hình ảnh tĩnh hay hình ảnh kỹ thuật số với bản ghi dựa trên cơ sở dữ liệu được lưu trữ từ trước. Điểm khác biệt của nhận diện khuôn mặt so với nhận diện vân tay với móng mắt đó là trích chọn đặc trưng (feature extraction) của khu vực nhận diện.

Để làm rõ hơn trong vấn đề của đặc trưng giữa từng bộ phận, nghiên cứu cần hiểu rõ hơn về công nghệ nhận diện và đặc trưng làm mỗi người trở nên khác biệt. Trong nhận diện khuôn mặt, trích chọn đặc trưng dựa trên hình dạng, kích thước và vị trí tương đối của các đặc điểm trên khuôn mặt như mắt, mũi, miệng, hình dạng hàm và đường viền trên khuôn mặt. Có tổng cộng 68 điểm đánh dấu trên khuôn mặt để nhận dạng. Nhưng có một vấn đề cần làm rõ rằng vì sao mỗi người lại có đặc điểm riêng tạo nên sự khác nhau trên khuôn mặt. Các nhà nghiên cứu tại Đại học California đưa ra rằng sự đa dạng trên khuôn mặt chúng ta là kết quả của áp lực tiến hóa để đảm bảo rằng mọi người có thể dễ dàng nhận ra nhau [1]. Sau khi nghiên cứu chuyên sâu hơn về vấn đề này, họ đã đưa ra kết luận rằng gen di truyền có một số ảnh hưởng nhất định đối với đặc điểm thể chất và các hình thái khác trên cơ thể nhưng hoạt động tạo hình trên khuôn mặt không theo một quy luật chung nào. Các DNA chịu tác động quy định đặc điểm trên khuôn mặt có trình tự phức tạp hơn rất nhiều so với các đoạn DNA quy định chiều cao, cân nặng và các đặc điểm khác trên cơ thể ví dụ như chiều cao tai, độ rộng mũi, độ rộng hai mắt thay đổi nhiều hơn so với các bộ phận ngoài khuôn mặt. Chính vì các lý do như trên, con người dù mang các đặc điểm về gen di truyền vẫn mang các đặc điểm khác nhau phân biệt với những người khác.



Hình 2.1. Khuôn mặt với các các điểm đặc trưng

Tiếp đến là các đặc trưng trích chọn của vân tay tập trung vào các mẫu hình học như đường thẳng, vòng cung, vòng xoáy, cũng như các điểm đặc biệt như nơi các đường vân tay gặp nhau hay phân tách. Trong bào thai từ khi sinh ra có một lớp đệm mô tế bào phát triển dưới da trên mỗi đầu ngón tay, chúng có một mô hình chính của vân tay gồm vân xoáy, vân móc và vân cung. Khi các đường vân phát triển đôi khi chúng đan vào nhau tạo thành một hoặc tách nhau ra do nhiều yếu tố tự nhiên, và bởi vì sự phát triển bất định không theo hệ thống nên không có ai có thể đoán trước được các đường vân của các ngón tay theo hệ thống ngẫu nhiên mà trung bình có khoảng 50 điểm giao nhau giữa chúng. Khi đó một ngón tay của con người được coi là độc nhất trong hơn triệu triệu trường hợp có thể xảy ra của đường vân trên ngón tay.

Còn về nghiên cứu trên móng mắt, các đặc trưng phân biệt dựa trên các sợi mô trong móng mắt, chẳng hạn như kết cấu, màu sắc và các mẫu vòng trên con ngươi đôi mắt. Các đặc trưng của móng mắt trong từng người cũng rất đa dạng,

khi so sánh màu sắc và con người cho ra cấu tạo riêng không có ai trùng lặp và là một nghiên cứu rất phát triển được nghiên cứu rộng rãi tương tự như nhận diện khuôn mặt và nhận diện vân tay.

Sự khác biệt chính trong trích chọn đặc trưng giữa nhận diện khuôn mặt so với nhận diện vân tay và móng mắt nằm ở tính chất đầu vào và cách thức xác định và phân tích giữa chúng. Nhận diện khuôn mặt phụ thuộc vào các đặc điểm nhận dạng 2 chiều hoặc 3 chiều của từng bộ phận lớn và phức tạp, trong khi nhận diện của vân tay và móng mắt tập trung nhiều vào các đặc điểm nhỏ trên một khu vực nhận diện.

Một biểu diễn khuôn mặt được định nghĩa tương minh hướng tới mục tiêu xây dựng một không gian đặc trưng trực quan, trong thực tế các biểu diễn tương minh thường không chính xác. Để khắc phục tình trạng đó, các hướng tiếp cận sau này được dựa trên ý tưởng sử dụng các mô hình học máy thống kê có khả năng học để lựa chọn đặc trưng khuôn mặt của một tập mẫu cho trước.

Đối với nhận diện vân tay và móng mắt, cả hai đặc điểm trên đều có đường nét và hình dạng cụ thể trong khi trích xuất dữ liệu và có thể áp dụng được nó ngay lập tức thì nhận dạng khuôn mặt lại mang đến cách thức lớn hơn khi dữ liệu khuôn mặt thường phong phú và liên tục thay đổi. Với khuôn mặt con người, dữ liệu hình ảnh khó có thể chụp hoặc lưu trữ các góc nhìn của mỗi người một cách toàn diện mà bộ xử lý không cho phép [2]. Ví dụ như khuôn mặt của một người khi đi ngang hoặc đi về phía trước, việc phân biệt tổng thể khuôn mặt trên máy tính không thể khách quan khi nhận biết người đó là ai trong dữ liệu mà phải đặt người đó phải hướng chính diện đòi hỏi con người luôn có sự hợp tác nhất quán mới cho ra kết quả như mong muốn.

Hiện nay các công nghệ nhận dạng không chỉ dùng để xác thực nhân thân mà còn được dùng trong rất nhiều bài toán thực tiễn như kiểm soát vào/ra, kiểm soát truy cập mạng, đảm bảo mức độ an ninh cần thiết tại các khu vực quan trọng như nhà ga, sân bay, ngân hàng, hỗ trợ tự động chăm công, v.v. Tại Việt Nam, công nghệ nhận dạng cũng được ứng dụng rất phổ biến, điển hình là các hệ thống chăm công tự động dựa trên nhận dạng vân tay, mặt người, các hệ thống giám sát an ninh, phát hiện đối tượng, phát hiện đột nhập, phát hiện và cảnh báo sự cố, bất thường. Tuy nhiên, theo tìm hiểu của tôi, hầu hết các sản phẩm nhận dạng sinh trắc học hiện có tại Việt Nam đều được nhập khẩu từ nước ngoài. Một số công nghệ

được kể đến như DMIT.US là phần mềm sinh trắc dấu vân tay được chuyển giao công nghệ hoàn toàn từ Singapore, Morpho là công ty dẫn đầu về công nghệ nhận diện sinh trắc học từ khuôn mặt, vân tay và móng mắt được hợp tác và đầu tư với công ty của Việt Nam,...

2.2. BÀI TOÁN NHẬN DIỆN ĐEO KHẨU TRANG

Nhận diện đeo khẩu trang (Face Mask Detection) là một trong những bài toán con của nhận diện sinh trắc học nhằm nhận diện con người có đeo khẩu trang hay không. Công nghệ nhận diện này bắt đầu phát triển rầm rộ từ khi dịch bệnh Covid-19 hoành hành tại một số nước trên thế giới và các quốc gia yêu cầu người dân bắt buộc đeo khẩu trang khi đi ra ngoài và vấn đề ô nhiễm môi trường cũng làm cho nhiều người thích nghi và làm quen với khẩu trang mà thời kỳ đầu các nước phương Tây hay tỏ ra e ngại, thậm chí phản đối kịch liệt khi đeo chúng [3]. Ngoài ra các nước châu Âu, tỷ lệ đeo khẩu trang vô cùng hiếm gặp so với các nước châu Á, được coi là hiện tượng phổ biến trong thời gian dài để ngăn chặn sự lây lan của bệnh cảm lạnh và cúm [4]. Cùng với đó từ khi người dân bắt đầu đeo khẩu trang, các công nghệ nhận diện đã không còn đưa ra kết quả chính xác khi nhận diện người dùng, mở khóa trên thiết bị điện tử hay xác thực danh tính [5].

Có hai vấn đề đối với ý thức của con người trong việc đeo khẩu trang. Đầu tiên người ta xác định được sự khác biệt trong ý định đeo khẩu trang dựa trên các hoạt động giải trí thể chất và phi thể chất, như cảm giác khó chịu khi đeo khẩu trang do thiếu oxy lưu thông trong máu và nhiệt độ cơ thể tăng trong hoạt động thể dục thể thao tuy phải bắt buộc thực hiện để phòng tránh giọt bắn tại những nơi đông người hoặc sự tự do duy trì khoảng cách ở thư viện, rạp phim, triển lãm,... và ít hoạt động thể chất là hai hoạt động hoàn toàn khác nhau. Thứ hai là thái độ và hành vi nhận thức được xem xét trước mỗi nguy hiểm do dịch bệnh gây ra, cụ thể là các cảm xúc, quan điểm phòng ngừa y tế và hiểu biết về cách thức lây lan dịch bệnh, nhận thức các biện pháp phòng chống sẽ thúc đẩy các biện pháp phòng ngừa dịch bệnh hiệu quả hoặc có thể làm suy yếu các nỗ lực trong việc kiểm soát dịch bệnh gia tăng [5].

Mặc dù nhiều nhà nghiên cứu đã nỗ lực để thiết kế các thuật toán hiệu quả để phát hiện và nhận dạng khuôn mặt nhưng vẫn tồn tại một sự khác biệt cơ bản giữa “nhận diện khuôn mặt dưới lớp khẩu trang” và “nhận diện khẩu trang trên khuôn mặt”. Theo tài liệu trước đây có rất ít nghiên cứu đã cố gắng nhận diện khẩu trang

trên khuôn mặt. Vì có mục đích nhằm phát triển một kỹ thuật có thể phát hiện chính xác khẩu trang trên khuôn mặt, cụ thể là mạng DBM, ở các khu vực công cộng (chẳng hạn như trên đường, cửa chính tòa nhà, khu vui chơi,...v.v) để hạn chế người dân tiếp xúc với các nguồn lây bệnh phức tạp, phòng tránh các tác nhân ảnh hưởng đến cơ thể bởi vi khuẩn, khói bụi trong hệ thống hô hấp và từ đó nâng cao nhận thức chăm sóc sức khỏe cộng đồng [2]. Hơn nữa, việc nhận diện khẩu trang ở nơi công cộng không phải là điều dễ dàng vì tập dữ liệu có sẵn để nhận diện khẩu trang rất ít dẫn đến việc đào tạo mô hình rất khó khăn. Vì vậy bài toán nhận diện khuôn mặt đeo khẩu trang là đề tài quan trọng trong việc nâng cao khả năng nhận diện khuôn mặt và sinh trắc học, đồng thời áp dụng các mô hình AI khác nhau làm đa dạng hóa công nghệ nhận diện sau này [7].

2.3. TÌNH HÌNH NGHIÊN CỨU TRONG VÀ NGOÀI NƯỚC

2.3.1. Tình hình nghiên cứu trong nước

Tại Việt Nam, công nghệ nhận diện khuôn mặt trong thị giác máy tính nói riêng và trí tuệ nhân tạo nói chung đã được nghiên cứu từ lâu. Tuy nhiên việc phát triển nhận diện khuôn mặt khi đeo khẩu trang mới thực sự được quan tâm đến trong thời gian gần đây và đạt được kết quả hoàn toàn mới, tạo nên bước đột phá quan trọng khi áp dụng nhận diện khẩu trang đầu tiên vào trong đời sống. Trong đó, một số nhà khoa học và các trường đại học đã trình bày thành công một số báo cáo của mô hình nhận diện khuôn mặt đeo khẩu trang bằng các phương pháp sau.

Tại Viện Nghiên cứu Trí tuệ Nhân tạo VinAI Research thuộc Vingroup đã công bố đề tài về nhận diện khuôn mặt cho con người ngay cả khi đeo khẩu trang. Công nghệ sử dụng mô hình học sâu được nghiên cứu mới và ổn định nhất trong thời điểm đó đã cho ra kết quả đáng kinh ngạc khi giả lập trên camera trên điện thoại mà họ sản xuất để nhận diện khuôn mặt và có đeo khẩu trang thành công nhận diện và mở khóa thiết bị, đánh dấu bước tiến lớn so với công nghệ nhận diện khuôn mặt trên thế giới. Hệ thống nhận diện của VinAI đang phát triển có mô hình khá đơn giản, chỉ cần sử dụng các thiết bị thu thập thông tin từ các camera thông thường mà không cần các thiết bị cảm biến như tia hồng ngoại hoặc cảm biến chiều sâu (Time-Of-Flight) nên giá thành rẻ hơn các các thiết bị công nghệ khác và có thể tích hợp với các camera giám sát khác ngoài điện thoại. Sản phẩm được lên kế hoạch sẽ phát triển thêm các tính năng khác cho quét khuôn mặt, chấm công cho công nhân và hệ thống xác thực cho doanh nghiệp và xí nghiệp [8].

Tương tự, một đề tài khác tại trường Đại học Sư phạm Kỹ thuật, Đại học Đà Nẵng với chủ đề nhận diện người đeo khẩu trang trong thời gian thực đã có kết quả rất cao với 98,99% trong tập dữ liệu thử nghiệm với hệ thống xử lý ảnh và nhận diện hình ảnh [9]. Trong đó (Bùi Xuân Quang & Trần Duy Tân, 2022) đã khái quát được cấu trúc chương trình để xây dựng hệ thống nhận diện khẩu trang, qua đó thiết lập chương trình gồm hai quá trình tạo dựng mô hình cho việc phân loại hình ảnh và áp dụng mô hình đó để nhận diện khẩu trang trên khuôn mặt.

Trong một bước tiến mới khác của đề tài nhận diện khẩu trang, nhóm nghiên cứu của trường Đại học Công nghiệp Thực phẩm TP.HCM được đăng trên tạp chí Khoa học Công nghệ và Thực phẩm đã áp dụng các kỹ thuật AI tiên tiến như ArcFace, MTCNN, MobileNet và ResNet để giải quyết những thách thức trong việc nhận diện con người đeo khẩu trang hay không. Qua thử nghiệm trên cơ sở dữ liệu RMFRD, công trình đã đạt được độ chính xác ấn tượng, lên đến trên 72%. Đây là một tỉ lệ khả quan, phản ánh khả năng thích ứng và hiệu suất cao của mô hình trong điều kiện thực tế.

Trong nghiên cứu đăng trên tạp chí Khoa học Công nghệ Hàng hải, (Đỗ Thị Chang & cs., 2021) thuộc Trường Đại học Hàng hải Việt Nam đã công bố công trình nghiên cứu mang tên "Phát hiện người đeo khẩu trang trong thời gian thực". Công trình này đã áp dụng công nghệ học máy và xử lý ảnh nhằm giải quyết vấn đề giám sát đeo khẩu trang một cách tự động trong thời kỳ dịch bệnh. Đặc biệt hơn công trình đã phát triển thêm một mô hình nhắc nhở người đeo khẩu trang thông qua giọng nói, tạo nên giải pháp thông minh hỗ trợ phòng chống dịch bệnh. Đáng chú ý, kết quả nghiên cứu còn cho biết mô hình được đào tạo với độ chính xác lên tới 98,28% trên tập dữ liệu thử nghiệm, cho thấy sự hiệu quả và chính xác cao khi áp dụng trí tuệ nhân tạo trong việc phát hiện người đeo khẩu trang nơi công cộng. Mô hình hiện có nhiều điểm mới cho đề tài nhận diện khẩu trang đang được nghiên cứu, trong đó mô hình triển khai và chức năng nhắc nhở người dân đeo khẩu trang được chú ý để phát triển đề tài nghiên cứu.

Hay gần đây nhất, công an thành phố Hồ Chí Minh đã tổ chức hội nghị bàn về giải pháp phòng ngừa tội phạm cướp tài sản tại các ngân hàng. Đặc điểm chung của nhóm tội phạm này là ngụy trang kín đáo như đeo khẩu trang, đeo kính, mặc áo khoác,... để tránh bị nhận diện. Những nghi phạm trước khi tiến hành hành vi phạm tội thường đến trước để quan sát, nắm quy luật hoạt động để thực hiện kế hoạch. Do mục tiêu đề ra là để phòng tránh các nguy cơ xảy đến khi các đối tượng

xuất hiện, công an đã chủ trương kế hoạch lắp đặt hệ thống camera nhận diện khuôn mặt định dạng cá nhân tại các khu giao dịch kết hợp chuông báo hiệu tại địa điểm truyền đến trụ sở công an gần nhất để ngăn ngừa, phòng chống tội phạm trộm tài sản.

2.3.2. Tình hình nghiên cứu ngoài nước

Theo Viện Tiêu chuẩn và Công nghệ Quốc gia Mỹ (NIST) năm 2020, công nghệ nhận diện khuôn mặt tại Hoa Kỳ đã gặp nhiều khó khăn bởi người dân đã bắt đầu đeo khẩu trang thường xuyên hơn. Theo số liệu được thu thập được, công nghệ nhận diện khuôn mặt được thử nghiệm với 89 thuật toán khác nhau đã cho ra kết quả sai số chỉ khoảng 0,3% trên dữ liệu được thu thập được từ con người. Nhưng khi có người đeo khẩu trang, thuật toán lại đưa ra sai số rất lớn, từ 5% lên tới 50% khiến cho việc nhận diện gặp trục trặc hoặc có thể mất kết nối đối với đối tượng. NIST đã nghiên cứu phối hợp với Bộ An ninh nội địa và Cục Hải quan và Biên phòng Hoa Kỳ đã đưa ra kết quả nghiên cứu cho rằng công nghệ của họ khó có thể nhận diện được khuôn mặt của một người khi họ đang đeo khẩu trang.

(Mei Ngan & cs., 2020) của Viện NIST đã triển khai một đề tài nhằm khắc phục tình trạng nhận diện không chính xác khi người đeo khẩu trang đã làm cải thiện nhận diện khẩu trang với độ chính xác cao hơn. Các tình trạng này sẽ là tác nhân chính mà ta cần phải khắc phục khi sử dụng công nghệ phát hiện khuôn mặt và điều kiện để cải thiện chất lượng [10]. Cụ thể họ đã tìm ra được nguyên nhân khiến khẩu trang trên khuôn mặt là giảm tính nhận diện trong việc sử dụng AI để khắc phục tình trạng khẩu trang trên hình ảnh như sau: Hiệu suất từ chối kết quả hợp lệ, hiệu suất chấp nhận kết quả sai, độ bao phủ của khẩu trang, màu của khẩu trang, ánh sáng và điều kiện tự nhiên, không thể phát hiện và tạo mẫu khung không thể nhận diện từ các điểm trên khuôn mặt và tạo hộp giới hạn quanh nó.

Còn tại Trung Quốc việc đeo khẩu trang hiện nay hiện đang được giảm bớt so với các nước trong khu vực châu Á. Tại các thành phố lớn như Thượng Hải, Bắc Kinh,... mật độ ô nhiễm không khí do bụi mịn gây ra lớn hơn 40% so với Mỹ. Trung Quốc hiện vẫn đang ở vị trí cao trong danh sách các nước ô nhiễm nhất trên thế giới, vượt qua mức độ tiêu chuẩn của WHO mức cho phép. Chính tình hình nguy cấp đó họ đã cho ra nhiều công nghệ thông minh ngăn chặn bệnh tật tích hợp kiểm soát con người [11].

Với công nghệ Trí tuệ nhân tạo đang phát triển rầm rộ, các công ty công nghệ

đang phát triển nhiều công cụ cho công việc giám sát mà có thể nhận diện được danh tính người dân ngay cả khi họ đeo khẩu trang và xác minh danh tính. Không khó để nhận biết được trên các phố đi bộ, lối ra vào trong các trụ sở là những nơi thường xuyên có rất đông người ra vào và không gian hẹp không có khoảng cách tối thiểu khi có người bị ho hoặc mắc bệnh truyền nhiễm tiếp xúc với người khác. Để phòng tránh và tạo không gian thuận lợi cho mỗi người, Trung Quốc đã xây dựng một ứng dụng đặt trong hệ thống tính điểm tín dụng xã hội của người dân trên ứng dụng mạng xã hội WeChat, mỗi khi có người không đeo khẩu trang, đeo sai cách hoặc không giãn cách hợp lý trên đường hệ thống sẽ thông báo cho người dân qua ứng dụng và nếu như họ không chấp hành theo quy định thì sau khi bị cảnh cáo, người vi phạm sẽ bị trừ điểm xã hội dựa theo mức độ vi phạm đã gây ra.

Cùng với đó một công ty khác ở Trung Quốc có tên là SenseTime cũng đã phát triển công nghệ nhận diện khuôn mặt có sẵn để nhận dạng con người dù đang đeo khẩu trang [12]. Việc nhận diện khuôn mặt thường được thực hiện bằng cách xác định các điểm phân biệt trên khuôn mặt gồm xung quanh mắt, mũi, miệng rồi kết nối chúng với nhau để tạo thành một dấu hiệu đặc trưng của mỗi người nhưng công nghệ mới của công ty cho phép nhận dạng một người từ mắt và từ vùng phía trên mũi của họ. Ngoài ra sản phẩm cũng thiết lập máy quét đo nhiệt độ trong cơ thể nhận biết được ai đang bị sốt, nhiệt độ tăng cao và thông báo đến cơ quan quản lý, bảo vệ đề phòng bệnh lây lan và truyền nhiễm trong các đơn vị hoặc bệnh viện cũng như thử nghiệm trên một số địa điểm công cộng địa phương.

Ngoài ra một công trình nghiên cứu về nhận diện khuôn mặt đeo khẩu trang cũng đang được chú ý gần đây thực hiện bởi Chandrika Deb & cộng sự., 2020 tại Ấn Độ được rất nhiều người quan tâm. Nhờ sử dụng kiến trúc MobiNetV2, hệ thống nhận diện khẩu trang đạt hiệu quả về mặt tính toán trong đó tỷ lệ nhận diện khẩu trang và khuôn mặt lần lượt là 99% và 88%, đào tạo ra mô hình phân loại kết quả tương đối cao và triển khai tới các hệ thống nhúng như Google Coral, Raspberry Pi,...v.v dễ dàng. Do đó, hệ thống này có thể sử dụng trong các ứng dụng thời gian thực để phát hiện người đeo khẩu trang, hoặc có thể tích hợp với các hệ thống nhúng được áp dụng tại nhà ga, sân bay, chăn nuôi chuồng trại, trường học và nơi công cộng nhằm đảm bảo các nguyên tắc an toàn phòng chống các loại dịch bệnh và nhiễm trùng đường hô hấp.

Sau khi phân tích kỹ lưỡng các nghiên cứu liên quan đến nhận diện khuôn mặt và ứng dụng của các mạng Boltzmann trong và ngoài nước, luận văn xác định

một số vấn đề then chốt cần được nghiên cứu và phát triển, đồng thời làm nổi bật đóng góp của đề tài trong lĩnh vực này. Cụ thể rằng các nghiên cứu trước đây đã cho thấy tiềm năng lớn của mạng DBM trong việc học sâu không giám sát và giám sát, tuy nhiên, việc áp dụng chúng vào bài toán phát hiện khuôn mặt đeo khẩu trang vẫn còn là một lĩnh vực tương đối mới mẻ và chưa được khai thác triệt để. Điều này đặt ra nhu cầu thực hiện các nghiên cứu sâu hơn về công việc tích hợp DBM vào các hệ thống nhận diện khuôn mặt trong điều kiện bị che khuất bởi khẩu trang, đồng thời là công việc cấp thiết trong môi trường ngày càng ô nhiễm tại các thành phố lớn. Luận văn tập trung vào việc triển khai một mô hình DBM hiệu quả có khả năng nhận diện khuôn mặt khi đeo khẩu trang trong nghiên cứu cơ bản qua việc tối ưu cấu trúc mạng và tinh chỉnh các tham số xử lý trên bộ dữ liệu khuôn mặt của sinh viên. Qua đó bài toán không chỉ đóng góp vào phát triển công nghệ nhận diện khuôn mặt trong bối cảnh thực tiễn mà còn xây dựng từ các biện pháp phòng chống dịch bệnh hiệu quả, đảm bảo an toàn nơi công cộng và đảm bảo an ninh xã hội.

Để áp dụng các công trình nghiên cứu vào đề tài, trong mô hình nghiên cứu được chia làm 2 giai đoạn chính. Giai đoạn đầu tiên là thiết lập và xây dựng mô hình DBM cho việc nhận diện hình ảnh của người đeo khẩu trang và khuôn mặt con người, trong đó các lớp của mô hình được xếp bởi nhiều tầng mạng RBM cùng với kỹ thuật lan truyền ngược tối ưu hóa các tham số hàm mất mát đưa ra kết quả được biểu diễn qua các dãy nhị phân. Tiếp theo giai đoạn thứ hai sẽ sử dụng mô hình đã được đào tạo để phân loại hình ảnh của một hình ảnh, video hay trong thời gian thực với sự hỗ trợ của mô hình DNN để trích xuất đặc trưng, vẽ hộp giới hạn trên đối tượng là khuôn mặt con người và phân biệt giữa từng khuôn mặt đeo khẩu trang và khuôn mặt người bằng việc đổi màu giữa chúng.

2.4. NGHIÊN CỨU DỮ LIỆU THU THẬP ĐƯỢC

2.4.1. Số lượng ảnh, người tham gia lấy mẫu

Trong công tác nghiên cứu, dữ liệu được sử dụng dựa trên sự tham gia của nhóm gồm 30 sinh viên tình nguyện (nam, nữ) sinh viên Học viện Việt Nam, và được sự đồng ý tham gia được sự đồng ý của nhóm sinh viên tham gia tình nguyện để sử dụng dữ liệu hình ảnh khuôn mặt của họ. Ý tưởng ban đầu của thu thập dữ liệu cho nghiên cứu là quay video khuôn mặt của mỗi người trong từng góc độ khác nhau khi đeo khẩu trang và khuôn mặt thường không bị đồ vật nào che khuất, mà trong đó có vài bạn có đeo kính tạo thêm sự đa dạng trong dữ liệu sau đó chọn

ra những hình ảnh phù hợp nhất trong góc quay đó để xử lý, chuẩn hóa, trích xuất dữ liệu hình ảnh.

Thông thường, mỗi người tham gia có tất cả 5 góc quay khi di chuyển mặt bao gồm nhìn thẳng, quay sang trái, quay sang phải, nhìn lên trên, nhìn xuống dưới. Tuy nhiên thao tác trên chưa kể đến quá trình đeo khẩu trang nên khi đó mỗi người tham gia sẽ có 10 hình ảnh tương ứng cho hai điều kiện.

Việc lựa chọn người tham gia để thu thập dữ liệu bắt đầu với bạn bè, người quen của nhóm sinh viên khoa Công nghệ thông tin thuộc Học viện Nông nghiệp Việt Nam và những sinh viên khác được tư vấn và đồng ý cho thu thập dữ liệu trong khuôn viên trường. Công tác lựa chọn của nhóm sinh viên vô cùng đa dạng với kiểu tóc, khuôn mặt, độ tuổi khác nhau,... và được chuẩn bị đặt lịch hẹn trước. Những người quen, bạn bè của sinh viên ở xa Học viện khi được biết về quá trình thu thập dữ liệu đã nhiệt tình tham gia gửi video về quá trình quay khuôn mặt được nhóm sinh viên hướng dẫn và chữ ký xác nhận góp phần nâng cao công tác thu thập dữ liệu và lòng nhiệt tình của sinh viên Học viện Nông nghiệp Việt Nam đối với nghiên cứu này.

Về quy trình, nhóm sinh viên đã chuẩn bị địa điểm quay video vào buổi sáng tại khoa Công nghệ thông tin, Học viện Nông nghiệp Việt Nam. Công việc lấy mẫu do đại diện của nhóm sinh viên thực hiện sử dụng điện thoại ghi hình để quay video. Những người tham gia được ngồi trên ghế, nhìn thẳng sát đối diện với camera, hướng ánh sáng tỏa ra trên khuôn mặt không bị tối hay che lấp đi khi người thực hiện thay đổi góc quay mặt và được tập luyện trước khi quay video để thu thập dữ liệu không thể xảy ra sai sót để mất đi thời gian quay video lại. Trong đó về phần khẩu trang, tôi đã lựa chọn loại khẩu trang y tế màu trắng và xanh lục phổ biến trong công cộng, ngoài ra một số sinh viên khác đã chuẩn bị thêm các khẩu trang vải và màu khác nhau để đa dạng dữ liệu hơn.

Kết quả các mẫu dữ liệu thu được là 30 tệp video của 30 sinh viên tình nguyện, trong đó mỗi tệp cho ra 10 hình ảnh khuôn mặt bình thường và đeo khẩu trang theo từng góc độ khác nhau với tổng cộng 300 hình ảnh khuôn mặt của sinh viên tình nguyện và các nguồn hình ảnh khuôn mặt khác được thu thập và tổng hợp trên nghiên cứu được công bố, mạng xã hội và các nguồn khác.

Tuy nhiên điểm hạn chế lớn nhất của các hình ảnh đeo khuôn mặt được tổng hợp bên ngoài khác luôn có độ nghiêng không thỏa mãn với quy định của nghiên

cứu đề ra trong đề tài, trong đó khuôn mặt khi được đưa vào công đoạn xử lý phải căn chỉnh rất lâu và phải loại bỏ nhiều sự vật thừa không liên quan trong nhận diện, cùng với đó là tỷ lệ số lượng giữa khuôn mặt thường và đeo khẩu trang luôn có chênh lệch lớn. Điều này dẫn đến nhiều khả năng mô hình nhận diện của nghiên cứu không cao hoặc nhận diện khẩu trang xảy ra sai sót.

2.4.2. Đặc trưng mẫu

2.4.2.1. Biểu diễn hình ảnh trên máy tính

Hình ảnh được biểu diễn trong máy tính bằng các ma trận số. Có ba kiểu phổ biến là ảnh màu, ảnh xám và ảnh nhị phân. Để hiểu rõ hơn về từng loại ảnh, nghiên cứu sẽ tiến hành phân tích từng loại ảnh sau

Đầu tiên là ảnh màu, trong hình ảnh cơ bản có 3 bộ số gồm R, G, B tương ứng với đỏ, lục, lam có trọng số trong khoảng từ $[0, 255]$. Ba màu sắc của ánh sáng khi được tách trên lăng kính và trộn theo tỷ lệ nhất định sẽ tạo ra các dải màu khác nhau tổng cộng có 256^3 màu, tương ứng với hơn 16 triệu màu. Để tiện cho việc xử lý ảnh, mỗi hình ảnh được tách thành các ma trận pixel có các tầng màu (Red, Green, Blue) và mỗi bức ảnh được tạo lên từ bộ ba màu sắc được kể trên. Có hai loại không gian màu để biểu diễn trong ảnh màu là RGB và HSV, giữa hai không gian màu này có công dụng và áp dụng trong nhiều nghiên cứu khác nhau. Mỗi ô nhỏ bên trong một hình ảnh là đại diện cho một pixel. Trong hình ảnh thực, những pixel này rất nhỏ đến độ mắt người không thể phân biệt được.

Ngoài ra một loại ảnh khác được đề cập tiếp theo là ảnh xám. Trong ảnh xám, mỗi một pixel mang một giá trị trong khoảng $[0, 255]$. Ảnh xám có khả năng làm giảm khối lượng thông tin so với ảnh màu, giúp tăng tốc độ xử lý nhưng vẫn đảm bảo các tác vụ cần thiết.

Còn về ảnh nhị phân, mỗi một pixel trong nó có giá trị là 0 và 1. Ảnh nhị phân thể hiện rõ các cạnh và hình dạng vật thể. Ứng dụng của nhị phân có thể kể đến như lọc nhiễu và định dạng vật thể, giúp tăng cường xử lý trong ảnh mà không thay đổi kết cấu trong nó.

Bằng công dụng của ảnh xám và ảnh nhị phân, bài toán xử lý có thể chuyển đổi từ ảnh màu sang ảnh xám và từ ảnh xám sang ảnh nhị phân. Tuy nhiên nếu như chuyển đổi từ ảnh màu sang ảnh nhị phân, hình ảnh đó cần phải chuyển đổi sang ảnh xám và sử dụng các thuật toán chuyển đổi số để thay đổi sang ảnh nhị phân tương ứng.

2.4.2.2. Bộ lọc ảnh

Bộ lọc ảnh hay được hiểu là mượt ảnh, mịn ảnh là bước quan trọng trong xử lý ảnh, giúp hình ảnh đầu vào được xử lý tốt hơn, giảm nhiễu ảnh, làm chắc nét các cạnh viền và làm hình ảnh được mượt mà hơn.

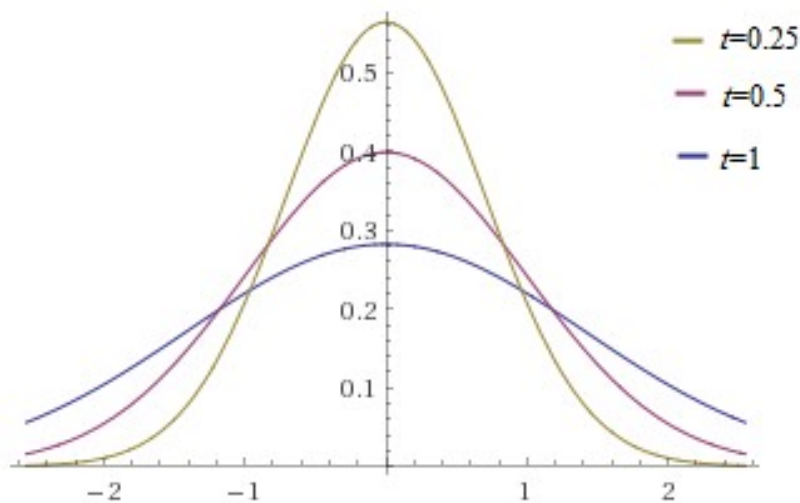
Nguyên tắc chung cho bộ lọc ảnh là cho một ma trận ảnh nhân với một ma trận lọc (Kernel). Ở đây nghiên cứu đã lựa chọn bộ lọc ảnh phù hợp là bộ lọc Gaussian hay còn được gọi là heat kernel (Gaussian heat kernel), nó thường được sử dụng để mô tả sự lan truyền của nhiệt trong không gian 2D/3D và cũng được sử dụng trong nhiều lĩnh vực khác nhau. Bộ lọc Gaussian có tính chất lan truyền nhiệt từ một điểm cố định ra xa theo thời gian và không gian.

Bộ lọc Gaussian thường được biểu diễn dưới dạng một hàm số, phụ thuộc vào biến thời gian t và không gian x và thường được ký hiệu là $H(t,x)$ hoặc $Kt(x)$. Hàm này thường được định nghĩa dựa trên phân phối Gaussian và có dạng

$$H(t, x) = \frac{1}{(4\pi t)^{\frac{d}{2}}} \cdot e\left(-\frac{\|x\|^2}{4t}\right) \quad (1)$$

Trong đó:

- d : Là số chiều của không gian 2D/3D
- t : Là thời gian
- x : Là vị trí trong không gian (vector x có d chiều)
- $\|x\|$: Là độ dài của vector x



Hình 2.2. Biểu đồ biểu diễn của bộ lọc Gaussian trong thời gian biến thiên

Bộ lọc Gaussian có các tính chất quan trọng bao gồm tính duy nhất của nó khi tất cả điểm bắt đầu và thời gian khác nhau và tính xấp xỉ khi t tiến gần đến 0 hoặc vô cùng. Nó thường được sử dụng trong lý thuyết xác suất để mô tả quá trình khuếch tán và trong xử lý ảnh để làm mịn hoặc làm nổi bật các đặc trưng của ảnh.

PHẦN 3. MẠNG BOLTZMANN MACHINE

3.1. GIỚI THIỆU VỀ HỌC SÂU

3.1.1. Khái niệm về học sâu

Học sâu (Deep Learning) là một phạm trù nhỏ của học máy (ML), và là một lĩnh vực nghiên cứu đầy hứa hẹn, đặt nền móng cho việc áp dụng mạng neural và các thuật toán tiên tiến để mô hình hoá dữ liệu ở các mức trừu tượng khác nhau. Điều này giúp giải quyết một loạt các bài toán khó khăn mà các mô hình học không sâu truyền thống thường gặp khó khăn, như thị giác máy tính, nhận diện giọng nói, xử lý ngôn ngữ tự nhiên, nhận dạng âm thanh ngôn ngữ và tin sinh học.

Trong học máy hiện đại, các mô hình học sâu không chỉ là công cụ mạnh mẽ mà còn là những người nghệ sĩ số liệu xuất sắc, đạt được độ chính xác đôi khi vượt xa hiệu suất của con người. Sự thành công này không chỉ đến từ việc huấn luyện trên bộ dữ liệu có nhãn mà còn từ cấu trúc mạng thần kinh với nhiều lớp, giúp mô hình hiểu biết và mô phỏng dữ liệu ở mức trừu tượng cao. Các mô hình học sâu không chỉ giỏi về khả năng mở rộng mạng neural, mà còn nổi bật với tính năng học tập tự động. Điều này đồng nghĩa với việc chúng có khả năng khai thác và học từ các đặc trưng tự động xuất phát từ dữ liệu thô, không cần sự can thiệp chi tiết từ con người.

Kiến trúc cơ bản của học sâu là mạng NN và có rất nhiều biến thể, hầu hết là các nhánh sinh ra từ kiến trúc ban đầu như: mạng neural sâu (Deep Neural Network), các máy Deep Boltzmann, Mạng neural tích chập (Convolutional neural networks - CNN), mạng niềm tin sâu (Deep Belief Network), mạng niềm tin sâu tích chập (Convolutional Deep Belief Network), mạng neural lưu trữ và truy xuất bộ nhớ lớn (Large Memory Storage And Retrieval Neural Network),...

Để giải thích chặt chẽ hơn, mạng DL có nhiều lớp ẩn với hàng triệu neural nhân tạo liên kết với nhau. Ở hai nút được liên kết với nhau tồn tại một trọng số tương ứng, số đó có thể dương nếu một nút kích thích nút còn lại và âm khi một nút ngăn cản quá trình đó. Các nút với trọng số cao hơn có ảnh hưởng lớn hơn lên các nút khác.

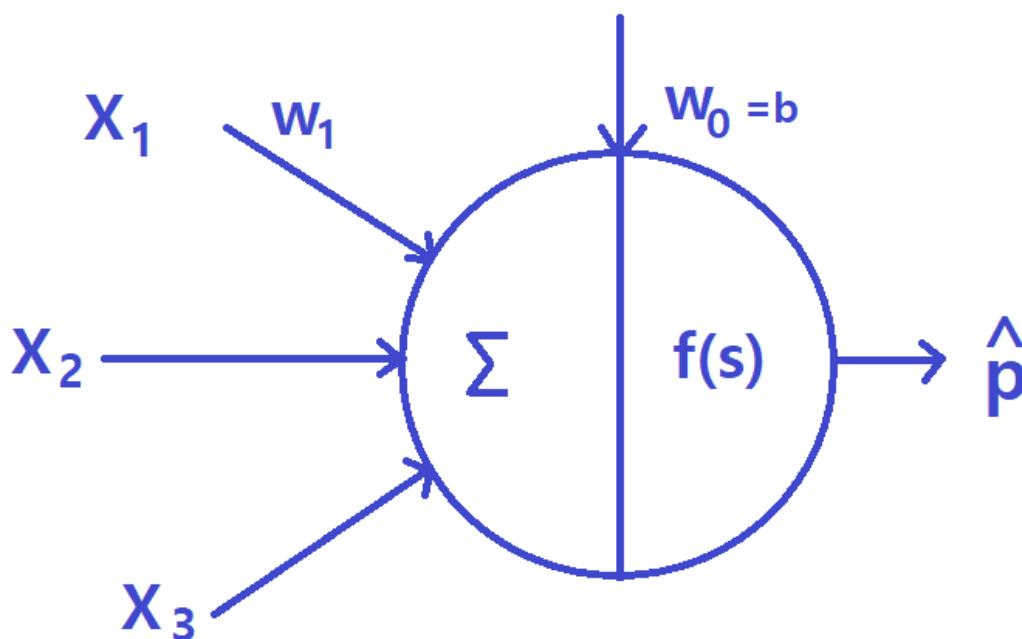
3.1.2. Mạng thần kinh nhân tạo

Mạng thần kinh nhân tạo (NN) là một mô hình toán học trong lĩnh vực trí tuệ nhân tạo được xây dựng để mô phỏng lại kiến trúc và hành vi của neural sinh học

trong não người. Nó là một hệ thống các neural nhân tạo nối với nhau thành các lớp và xử lý thông tin bằng cách truyền theo các kết nối giữa các neural, tạo ra một hệ thống thích ứng được máy tính sử dụng để học hỏi từ sai lầm khi học và liên tục cải thiện.

Bộ não con người là nguồn cảm hứng trong kiến trúc mạng NN. Các tế bào não của con người, hay còn được gọi là neural, tạo thành một mạng lưới có tính liên kết cao và cấu trúc phức tạp, gửi đi các tín hiệu điện đến nhau để giúp con người xử lý thông tin. Tương tự như vậy, mạng NN được tạo ra từ các tế bào neural nhân tạo cùng được phối hợp với nhau để giải quyết một vấn đề [13].

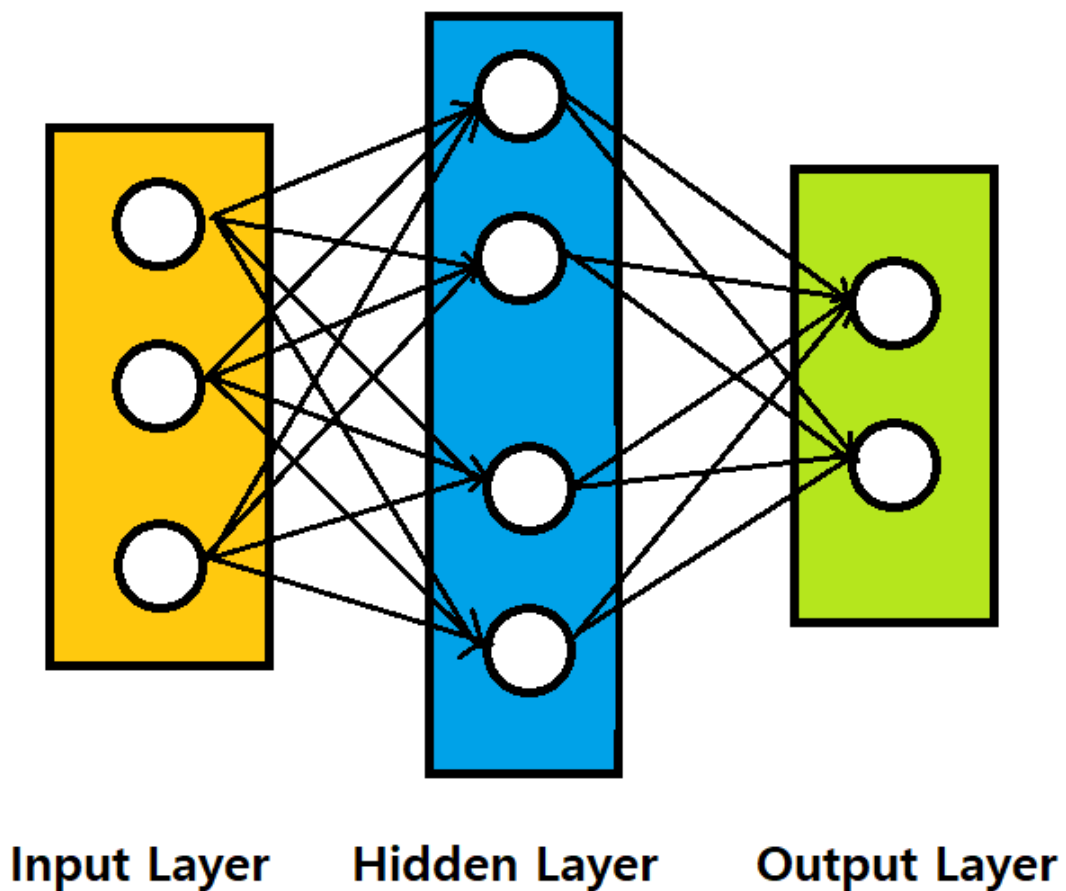
Để dễ dàng giải thích các thành phần của mạng neural, ta có một ví dụ về một mạng neural đơn giản là mạng perceptron đơn do Rosenblatt đưa ra vào năm 1957. Kiến trúc và hành vi của perceptron rất giống với neural sinh học và thường được coi là dạng cơ bản nhất của mạng neural. Các loại mạng neural khác nhau đã được phát triển dựa trên perceptron, và còn đang tiếp tục phát triển cho tới hiện nay.



Hình 3.1. Mạng perceptron đơn

Kiến trúc của mạng NN cơ bản liên kết với nhau theo ba lớp. Trong nghiên cứu, ba lớp dưới đây được xếp theo trình tự từ lớp đầu vào, lớp ẩn và lớp đầu ra. Trong đó lớp đầu vào mang thông tin từ thế giới bên ngoài như hình ảnh, âm thanh,

ký tự,... đi vào mạng NN. Các nút đầu vào xử lý dữ liệu, phân tích hoặc phân loại và chuyển các dữ liệu được xử lý sang các lớp tiếp theo. Ở giữa lớp đầu vào và lớp đầu ra tồn tại một hay nhiều lớp nhỏ được gọi là lớp ẩn, các lớp ẩn này nhận dữ liệu từ lớp đầu vào hay các lớp ẩn trước đó. Mỗi lớp ẩn phân tích dữ liệu đầu ra từ lớp trước, xử lý dữ liệu đó sâu hơn và rồi chuyển dữ liệu sang lớp tiếp theo.



Hình 3.2. Mô hình mạng NN cơ bản

Mạng NN được sử dụng trong nhiều bài toán trải dài khắp các lĩnh vực. Trong đó phổ biến nhất là bốn ứng dụng được nhiều nhà phát triển tìm hiểu và nghiên cứu là thị giác máy tính, nhận dạng giọng nói, xử lý ngôn ngữ tự nhiên và công cụ đề xuất thường hướng đến cá nhân hóa người dùng.

3.1.2.1. Neural

Neural là đơn vị cơ bản cấu tạo nên hệ thống thần kinh và là một phần quan trọng nhất trong bộ não. Bộ não của con người có khoảng 10 triệu neural và mỗi

neural liên kết với 10000 neural khác [13]. Ở mỗi neural có phần thân chứa các nhân , tín hiệu đầu vào đi qua các sợi nhánh và tín hiệu đầu ra đi qua các sợi trục kết nối với các neural khác. Mỗi neural nhận xung điện từ các neural khác qua sợi nhánh và khi các xung điện này đủ lớn để kích hoạt neural thì tín hiệu này đi qua sợi trục đến các sợi nhánh của các neural khác trong não, ở mỗi neural có cần quyết định kích hoạt các neural đó hay không, tương tự như cấu trúc hồi quy logic của học sâu.

Tương tự như kiến trúc và hành vi của neural sinh học, một neural và một perceptron có cấu trúc tương tự với đầu vào và đầu ra. Thông tin từ đầu vào đi qua neural sẽ được biến đổi, sau đó đi ra tại đầu ra. Nói cách khác, một neural là một tập hợp các hàm biến đổi toán học để biến đổi đầu vào thành đầu ra mong muốn. Trong ví dụ trên, mạng perceptron đơn được cấu tạo từ một perceptron duy nhất, sử dụng hàm tính tổng và một hàm phi tuyến f , hoạt động như một bộ phân lớp nhị phân với đầu vào là một vector đặc trưng $[x_1, x_2, x_3]$ và đầu ra là xác suất của một sự kiện nhất định.

3.1.2.2. Trọng số

Trong mạng perceptron đơn, mỗi đầu vào của vector đặc trưng không chỉ là một số liệu thuần túy, mà còn được gắn với một "lực lượng" đặc biệt, được biểu diễn qua trọng số tương đối (w). Trọng số này là một yếu tố quan trọng, làm nổi bật ảnh hưởng của từng đầu vào đối với quá trình hình thành giá trị tổng trong mạng. Nó như là một bức tranh động, nơi một số đường nét ẩn tượng hơn sẽ có đặc tính màu sắc và chiều sâu đặc biệt.

Mỗi trọng số không chỉ là một con số, mà là một mảng các quy tắc ẩn chứa dữ liệu của đầu vào tương ứng. Những đầu vào có các yếu tố quan trọng sẽ được ưu ái với trọng số lớn, trong khi những đầu vào ít quan trọng hơn sẽ có trọng số nhỏ. Đồng thời, độ lệch (bias) $w_0 = b$ cũng tham gia vào cuộc "hội tụ" này, như một trọng số đặc biệt, làm nổi bật tính chất chung của mạng .

$$s = w_0 + w_1x_1 + w_2x_1 + w_3x_3 = [w \ b][x \ 1]^T \quad (2)$$

Trong đó:

s : Là giá trị đầu ra của mạng perceptron đơn. Đây là tổng có trọng số của tất cả các đầu vào, kết hợp với độ lệch (bias).

w_0 : Là độ lệch (bias), là một trọng số đặc biệt trong mạng perceptron đơn,

đại diện cho tính chất chung của mạng. Nó cộng với tổng trọng số của tất cả các đầu vào.

w_1, w_2, w_3, \dots : Là các đầu vào của mạng, thường là các đặc trưng của dữ liệu. Mỗi đầu vào được nhân với trọng số tương ứng của nó trong quá trình tính tổng.

$[w \ b]$: Là một vector cột chứa tất cả các trọng số cùng với độ lệch được thêm vào cuối của vector.

3.1.2.3. Hàm kích hoạt

Hàm kích hoạt (activation function) được mô phỏng tỷ lệ truyền xung qua axon của một neural thần kinh. Trong một mạng NN, hàm kích hoạt đóng vai trò là thành phần phi tuyến tại đầu ra của các neural mà kết quả của hàm tính tổng được biến đổi thành một đầu ra mong muốn bằng cách sử dụng một hàm phi tuyến f (non-linear function). Các hàm kích hoạt thường gặp được liệt kê theo bảng 3.1.2.3 dưới đây:

Bảng 3.1. Một số hàm kích hoạt

Hàm kích hoạt	Công thức	Giá trị khoảng
Đồng nhất (Identity)	$f(x) = x$ (3)	$(-\infty, +\infty)$
Logistic (Sigmoid)	$f(x) = \sigma(x) = \frac{1}{1 + e^{-x}}$ (4)	$(0, 1)$
TanH	$f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ (5)	$(-1, 1)$
Rectified linear unit (ReLU)	$f(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} = \max(0, x)$ (6)	$[0, +\infty)$
Softmax	$f\left(\frac{\rightarrow}{x}\right) = \frac{e^{x_i}}{\sum_{j=1}^J e^{x_j}}, i \in [1, J]$ (7)	$(0, 1)$

Đồng nhất (Identity)

Trong đó:

- f : Là đại diện cho hàm số, nói cách khác, là cách chúng ta gọi hành động hoặc quy tắc ánh xạ một giá trị đầu vào sang một giá trị đầu ra.
- x : Là biến số, đại diện cho giá trị đầu vào của hàm số.

Logistic (Sigmoid) và TanH

Trong đó:

- x : Là giá trị đầu vào của hàm.
- e : Là số Euler, một hằng số xấp xỉ khoảng 2.71828.

Rectified linear unit (ReLU)

Trong đó:

- x : nhỏ hơn 0, hàm trả về 0 nếu lớn hơn hoặc bằng 0, hàm trả về chính giá trị x .

Softmax

Trong đó:

- $f(\rightarrow_x)$: Là đại diện cho phép biến đổi softmax được áp dụng cho điểm số ban đầu của lớp i .
- e : Là số Euler, một hằng số xấp xỉ khoảng 2.71828.
- x_j : Là điểm số ban đầu cho lớp i .
- $\sum_{j=1}^j e$: Là tổng của các điểm số được mũ của tất cả các lớp, từ $j = 1$ đến $j = J$.

Vì đầu ra mong muốn của bài toán là xác suất của một sự kiện, ta có thể sử dụng hàm sigmoid để giới hạn tổng giá trị trong khoảng (0,1).

3.1.3. Các thuật toán huấn luyện

Bên cạnh mạng neural, một mô hình học sâu cần có các thuật toán để huấn luyện mạng neural đó. Các thuật toán này dùng để tính toán mức độ sai sót của mô hình mạng, đồng thời tinh chỉnh và đưa ra kết quả đầu ra phù hợp với mô hình mạng trong công nghệ nhận diện, phân loại và nhiều bài toán khác nhau.

3.1.3.1. Sai số và hàm mất mát

Trong hầu hết mạng neural, sai số được tính bằng hiệu giữa đầu ra mong muốn và đầu ra dự đoán.

$$L(w) = \frac{\text{đầu ra mong muốn } (p)}{\text{đầu ra dự đoán } (p')} \quad (8)$$

Hàm mất mát (L) là hàm ánh xạ một biến cô trong xác suất hoặc các giá trị của một hoặc nhiều biến thành một số thực. Công thức thường gặp được viết thành.

$$\mathcal{L}_D(f_\omega) = \frac{1}{|D|} \sum_{(x,y) \in D} L(f_\omega(x), y) \quad (9)$$

Trong đó L là hàm mất mát.

Hàm mất mát trả về một giá trị thực không âm biểu diễn sự chênh lệch giữa hai đại lượng nhãn dự đoán $f_\omega(x)$ và nhãn giá trị thực y .

Trong đó:

- $\frac{1}{|D|}$: Là một hệ số chuẩn hóa, được sử dụng để tính trung bình của hàm mất mát trên toàn bộ tập dữ liệu D .
- $|D|$: Là số lượng mẫu trong tập dữ liệu.
- $\sum_{(x,y) \in D}$: Ký hiệu tổng, trong đó (x, y) là cặp dữ liệu trong tập D .
- $L(f_\omega(x), y)$: Đây là hàm mất mát cụ thể, tính toán sự chênh lệch giữa dự đoán của mô hình $f_\omega(x)$ và nhãn thực tế y .

Hàm mất mát có vai trò đánh giá độ chính xác của mô hình với một bộ trọng số tương ứng. Mục đích của quá trình huấn luyện là tìm bộ trọng số để độ lớn của hàm mất mát xuống mức nhỏ nhất (cực tiểu) hay có thể coi hàm mất mát làm hàm mục tiêu quan trọng trong quá trình huấn luyện mô hình.

3.1.3.2. Lan truyền ngược và hàm tối ưu

Trong quá trình huấn luyện mạng neural, việc giảm thiểu sai số là một bước quan trọng để đảm bảo dự đoán chính xác. Hàm mất mát $L(w)$, nơi w đại diện cho các tham số trong mô hình như các trọng số và độ lệch, chính là tiêu chí để đánh giá hiệu suất của mô hình. Để tối ưu hóa mô hình, chúng ta cần điều chỉnh các tham số này sao cho hàm mất mát đạt được giá trị cực tiểu.

Lan truyền ngược là một phương pháp quan trọng để giảm thiểu sai số trong mạng neural. Cụ thể, khi chúng ta đã có dự đoán tại lớp đầu ra, sai số tại lớp hiện tại sẽ được truyền ngược lại lớp trước đó. Quá trình giúp điều chỉnh các trọng số và độ lệch sao cho sai số giảm đi, từng bước một.

Để thực hiện quá trình tối ưu hóa, chúng ta sử dụng các hàm tối ưu hóa. Những hàm này thường tính độ dốc (gradient), tức là đạo hàm riêng của hàm mất mát đối với các trọng số. Các trọng số sau đó được cập nhật theo hướng ngược lại của độ dốc tính được. Quá trình được lặp lại nhiều lần cho đến khi chúng ta đạt được giá trị cực tiểu của hàm mất mát.

Điều này có thể được hiểu như việc tìm kiếm "địa điểm thấp nhất" trên bề mặt của hàm mất mát, và các hàm tối ưu giúp chúng ta dẫn dắt mô hình đến vị trí này để cải thiện hiệu suất dự đoán. Quá trình này đòi hỏi sự linh hoạt và khéo léo của các thuật toán tối ưu hóa để đảm bảo hội tụ đến giải pháp tối ưu.

3.1.3.3. Ước tính thời điểm thích ứng

Ước tính thời điểm thích ứng (Adam) là một thuật toán tối ưu hóa thường được các nhà nghiên cứu và nhà khoa học dữ liệu sử dụng trong mạng ML và mạng NN. Ý tưởng cơ bản của nó là sử dụng tốc độ học thích ứng trong từng tham số. Điều này có nghĩa là các nhà nghiên cứu có thể điều chỉnh tốc độ học trong quá trình huấn luyện trong từng tham số dựa trên độ dốc trước đó - đạo hàm của hàm có nhiều biến đầu vào. Điều này cho phép mô hình căn chỉnh hiệu quả các tốc độ học khác nhau cho các tham số khác nhau và giúp cải thiện tính hội tụ của quá trình đào tạo. Nói cách khác, nó giúp mô hình nhanh hơn mà không cần đào tạo thêm lần nữa.

Thuật toán Adam sử dụng hai sự kiện chính, quá trình đầu tiên (động lượng) và quá trình thứ hai (phương sai không tâm). Những sự kiện này được tính toán trong quá trình đào tạo và được sử dụng để điều chỉnh tốc độ học tập. Quá trình đầu tiên giúp tính ra tốc độ của tham số, trong khi quá trình thứ hai chứa thông tin về phương sai của độ dốc.

3.1.3.4. Dropout

Dropout là một thuật toán trong đó các neural được chọn ngẫu nhiên sẽ bị bỏ qua trong quá trình huấn luyện. Chúng bị "bỏ rơi" một cách ngẫu nhiên. Điều này có nghĩa là sự đóng góp của chúng vào việc kích hoạt các neural chuyển tiếp sẽ bị loại bỏ tạm thời ở lượt tiếp theo và bất kỳ quá trình cập nhật trọng số nào cũng không được áp dụng cho neural ở lượt di chuyển ngược lại.

Khi mạng neural học, trọng số neural sẽ phù hợp với bối cảnh của chúng trong mạng. Trọng số của các neural được điều chỉnh theo các tính năng cụ thể mang lại một số tính chuyên môn hóa. Các neural lân cận sẽ dựa vào tính chuyên môn hóa này, nếu đi ra quá xa có thể dẫn đến một mô hình mỏng manh quá chuyên biệt cho dữ liệu huấn luyện. Sự phụ thuộc vào bối cảnh của một neural trong quá trình huấn luyện được gọi là sự đồng thích ứng phức tạp.

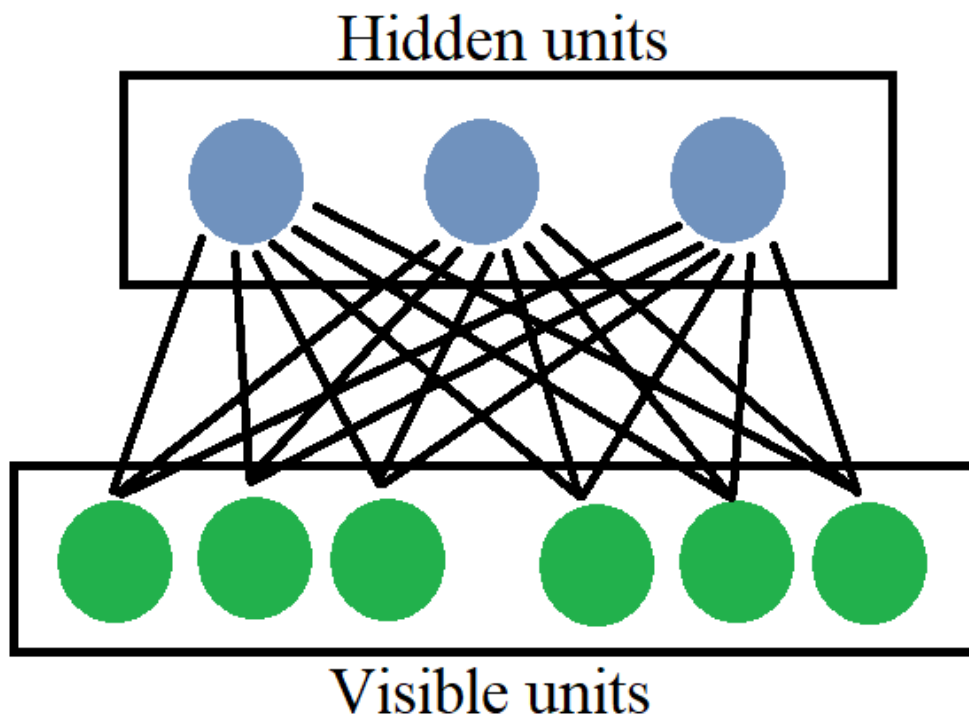
Có thể tưởng tượng được rằng các neural ngẫu nhiên bị loại khỏi mạng trong quá trình huấn luyện thì các neural khác sẽ phải can thiệp và xử lý việc biểu hiện

cần thiết để đưa ra dự đoán cho các neural bị thiếu. Điều này tin rằng có nhiều biểu diễn độc lập được học trong mạng.

3.2. GIỚI THIỆU VỀ MẠNG RESTRICTED BOLTZMANN MACHINE

3.2.1. Mạng restricted Boltzmann machine

Mạng Restricted Boltzmann Machine (RBM) là một mô hình xác suất sử dụng một lớp các biến nhị phân ẩn hoặc đơn vị để mô hình hóa sự phân bố của một lớp biến hiển thị. Được giới thiệu ban đầu bởi Smolensky vào năm 1986, RBM đã trở thành một công cụ quan trọng trong lĩnh vực máy học và học sâu.



Hình 3.3. Cấu trúc mạng restricted Boltzmann Machine

RBM bao gồm hai lớp chính: lớp biến hiển thị v (các đơn vị trong lớp "visible layer") và lớp biến ẩn h (các đơn vị trong lớp "hidden layer"). Lớp hiển thị biểu thị dữ liệu đầu vào, trong khi lớp ẩn biểu thị một tập hợp các tính năng được mạng học tập. Ngoài ra còn có thêm độ lệch trọng số (offset) với giá trị a_i tương ứng với lớp hiển thị v_i và giá trị b_j tương ứng với lớp ẩn h_j .

Mỗi đơn vị v được kết nối với mỗi đơn vị h thông qua một trọng số ω_{ij} có thể biểu diễn dưới dạng ma trận $m \times n$ và có giá trị nhị phân.

$$h_j = \sigma \left(\sum_i^m v_i \cdot \omega_{ij} + b_j \right) \quad (10)$$

Trong đó:

- h_j : Là đầu ra của neural thứ j trong một lớp ẩn.
- σ : Là hàm kích hoạt, thường là một hàm phi tuyến tính như hàm sigmoid.
- v_i : Là đơn vị thứ i trong lớp hiển thị.
- ω_{ij} : Là trọng số kết nối giữa đơn vị thứ i trong lớp hiển thị và đơn vị thứ j trong lớp ẩn.
- b_j : Là độ lệch trọng số (bias) cho neural thứ j trong lớp ẩn.

Với trọng số và độ lệch, mạng RBM được học với hàm năng lượng với công thức:

$$E(v, h) = -\sum_i a_{ij} - \sum_i b_j \cdot h_j - \sum_i - \sum_j v_i \cdot \omega_{ij} \cdot h_j \quad (11)$$

Trong đó:

- $E(v, h)$: Là năng lượng của mạng RBM với các giá trị v và h cho các lớp hiển thị và ẩn tương ứng.
- a_{ij} : Là độ lệch trọng số (bias) cho đơn vị thứ i trong lớp hiển thị v .
- b_j : Là độ lệch trọng số (bias) cho neural thứ j trong lớp ẩn h .
- ω_{ij} : Là trọng số kết nối giữa đơn vị thứ i trong lớp hiển thị và neural thứ j trong lớp ẩn.

RBM được đặc trưng bởi hàm năng lượng (energy function) dựa trên tham số trọng số và các giá trị hiện tại của các đơn vị ẩn và hiển thị. Hàm năng lượng này giúp xác định xác suất của một trạng thái cụ thể bằng công thức.

$$P(x, h) = \frac{Z}{2} = \frac{(\sum e^{-E(v,h)})}{2} \quad (12)$$

Trong đó:

- $P(x, h)$: Là xác suất của một trạng thái cụ thể (x, h) trong RBM, trong đó x là trạng thái của lớp hiển thị và h là trạng thái của lớp ẩn.
- Z : Là hằng số chuẩn hóa (partition function) để đảm bảo tổng xác suất của tất cả các trạng thái có tổng bằng 1.
- $E(v, h)$: Là năng lượng của mạng dựa trên trạng thái v và h .

Ở đây, Z là một hàm phân vùng được định nghĩa là tổng của $e^{-E(v,h)}$, hoặc được hiểu như là hằng số chuẩn hóa để đảm bảo các xác suất có tổng bằng 1.

Do cấu trúc cơ bản của RBM không có kết nối trong lớp nên các hàm kích hoạt của h độc lập lẫn nhau với các hàm kích hoạt của v . Xác suất kích hoạt riêng lẻ của mạng Boltzmann được đưa ra bởi công thức:

$$p((h_j|x)) = \text{sigm}(\sum_i w_{ij}x_j + c_j) \quad (13)$$

$$p((x_j|h)) = \text{sigm}\left(\sum_j w_{ij}h_j + b_j\right) \quad (14)$$

Trong đó:

- $p((h_j|x))$ và $p((x_j|h))$: Là xác suất rằng hidden unit j sẽ được kích hoạt (có giá trị là 1) khi biết giá trị của các visible units.
- $\text{sigm}()$: Là hàm sigmoid, một hàm kích hoạt phổ biến được sử dụng trong các mạng neural, biểu thị xác suất là một số nằm trong khoảng từ 0 đến 1.
- w_{ij} : Là trọng số kết nối giữa visible unit i và hidden unit j .
- c_j và b_j : Là độ lệch (bias) của hidden unit j .

Trong đó sigm được biểu thị là hàm đại số logic sigmoid với công thức:

$$\text{sigm}(x) = \frac{1}{1+e^{-x}}$$

RBM đã được áp dụng thành công trong nhiều lĩnh vực, bao gồm xử lý hình ảnh và văn bản. Trong quá trình huấn luyện RBM, thuật toán được thực hiện một cách không giám sát để lập mô hình phân phối đầu vào. Có thể sử dụng nhiều RBM để tạo thành các tầng xây dựng (layer) cho mô hình học sâu.

Trong việc giải quyết các vấn đề học giám sát, RBM thường được tích hợp với các thuật toán học khác như mạng neural (NN) hoặc bộ phân loại. Một trong những cách tiếp cận là sử dụng lớp ẩn của RBM để xử lý trước dữ liệu đầu vào, thay thế nó bằng đầu ra của lớp ẩn. Hoặc, các tham số của RBM có thể được sử dụng để khởi tạo quá trình chuyển tiếp trong mạng NN.

Tuy RBM mang lại nhiều ưu điểm tương tự mạng NN, nhưng ngoài ra nó cũng đặt ra nhiều thách thức khi triển khai mô hình. Vấn đề cơ bản là việc điều chỉnh đồng thời cả hai bộ siêu tham số của RBM và các thuật toán học khác là một nhiệm vụ phức tạp. Do RBM được đào tạo không giám sát, nó không nhìn thấy bản chất của nhiệm vụ giám sát cần giải quyết, và không đảm bảo rằng thông tin được trích xuất là đủ cho việc học giám sát trong không gian mô hình.

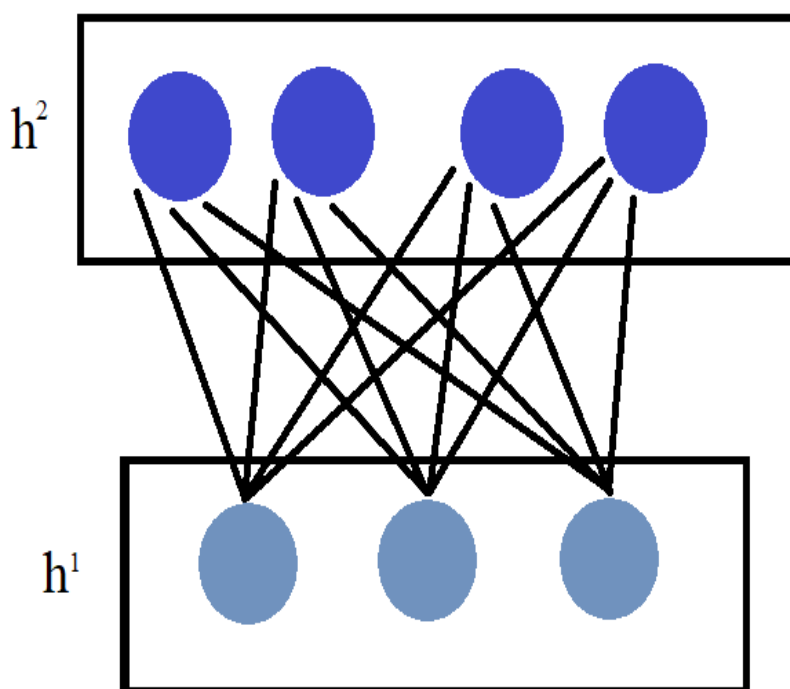
3.2.2. Mạng classification Restricted Boltzmann machine

Mạng classification Restricted Boltzmann Machine (classRBM) một biến thể linh hoạt của Restricted Boltzmann Machine (RBM) trong lĩnh vực máy học, được thiết kế đặc biệt để đối mặt với các nhiệm vụ phức tạp của bài toán phân loại. RBM, một loại mô hình máy học không giám sát, thường được áp dụng cho việc học biểu diễn và giảm chiều dữ liệu, nhưng classRBM nâng cao khả năng áp dụng của nó trong bối cảnh phân loại dữ liệu.

Trong bài toán phân loại lớp, mục tiêu chính của nó là gán nhãn (label) chính xác dựa trên từng dữ liệu đầu vào. Các mạng RBM ban đầu chỉ có thể xử lý dữ liệu không có nhãn, tức là không có đầu ra cụ thể được nhận diện từ trước. Tuy nhiên khi có một lớp nhãn được đưa vào, nó trở thành mạng classRBM có giám sát có thể dự đoán đầu ra chính xác cho dữ liệu mới dựa trên mối quan hệ học được từ dữ liệu đã gán nhãn trước. Kiến trúc của mô hình này thường gồm hai lớp quan trọng: lớp visible và lớp ẩn. Lớp visible đóng vai trò biểu diễn cho dữ liệu đầu vào, trong khi lớp ẩn chịu trách nhiệm học các đặc trưng biểu diễn dữ liệu và đồng thời có khả năng thực hiện nhiệm vụ phân loại.

Để đạt được hiệu suất tốt, classRBM thường được huấn luyện thông qua các phương pháp tiên tiến, ví dụ như Contrastive Divergence. Quá trình này giúp mô hình học các trọng số sao cho nó có khả năng tạo ra biểu diễn chất lượng cho dữ liệu đầu vào và đồng thời thực hiện nhiệm vụ phân loại một cách hiệu quả.

Một trong những ưu điểm lớn của classRBM nằm ở khả năng học đặc trưng không giám sát, cho phép mô hình tự động rút trích những đặc tính quan trọng từ dữ liệu. Điều này giúp nó trở thành một công cụ mạnh mẽ trong việc xử lý và phân loại dữ liệu đa dạng, đặc biệt là khi đối mặt với những bài toán phân loại có độ phức tạp cao. Tuy nhiên, như mọi mô hình, classRBM cũng đặt ra những thách thức và yêu cầu sự chú ý đặc biệt trong việc điều chỉnh các tham số và quá trình huấn luyện.



Hình 3.4. Cấu trúc mạng classification Restricted Boltzmann Machine

3.3. GIỚI THIỆU VỀ MẠNG DEEP BOLTZMANN MACHINE

Mạng Deep Boltzmann Machine (DBM) là một trong những mô hình học sâu không giám sát mà nói riêng và học sâu mà nói chung, đã đưa ra những đóng góp quan trọng trong việc hiểu và biểu diễn dữ liệu phức tạp. DBM được phát triển như một sự mở rộng của mô hình Restricted Boltzmann Machine (RBM), một loại mô hình đồng thuận có khả năng học biểu diễn từ dữ liệu [14].

Trong kiến trúc của DBM, nó có nhiều lớp ẩn, với mỗi lớp ẩn tạo ra một cấp biểu diễn tiếp theo cho dữ liệu. Điều quan trọng là, các lớp RBM trong DBM không được kết nối trực tiếp với nhau. Thay vào đó, chúng được kết nối thông qua các lớp ẩn trung gian. Mô hình này tạo ra một cấu trúc mạng phức tạp, mở ra khả năng học và biểu diễn các đặc trưng phức tạp hơn so với RBM thông thường.

Quá trình huấn luyện của DBM thường bắt đầu thông qua giai đoạn được gọi là "đào tạo theo từng tầng". Trong giai đoạn này, mỗi lớp ẩn được huấn luyện như là một RBM độc lập, sử dụng dữ liệu từ lớp trước đó. Mỗi lớp ẩn tìm cách học biểu diễn của dữ liệu một cách hiệu quả thông qua quá trình này. Trong trường hợp mạng DBM có 1 lớp hiển thị v và ba tầng ẩn lần lượt là h_1, h_2, h_3 , xác suất của chúng được thể hiện là:

$$P(v, h1, h2, h3) = \frac{1}{Z(0)} \exp(-E(v, h1, h2, h3; \theta)) \quad (15)$$

Trong đó:

- $P(v, h1, h2, h3)$: Là xác suất của các units trong mạng DBM, bao gồm visible units và hidden units.
- $Z(0)$: Là hằng số chuẩn hóa, cũng gọi là hàm phân partition, là tổng của toàn bộ xác suất trên tất cả các trạng thái có thể của mạng.
- $\exp(-E(v, h1, h2, h3; \theta))$: Là biểu thị sự "khả thi" của một cấu hình, với việc sử dụng hàm mũ để biểu thị mối quan hệ giữa năng lượng và xác suất.

Cùng với đó, hàm năng lượng của chúng được biểu thị thông qua:

$$E(v, h1, h2, h3, \theta) = -v^T \cdot w_1 h1 - h^T \cdot w_2 h2 \quad (16)$$

Trong đó:

- $E(v, h1, h2, h3, \theta)$: Là hàm năng lượng của mạng DBM.
- v : Là vector của visible units.
- $h1, h2, h3$: Lần lượt là vector của các hidden units tương ứng với ba tầng ẩn.
- w_1 và w_2 : Là các ma trận trọng số kết nối giữa các layers, với w_1 liên kết visible layer với hidden layer $h1$ và w_2 liên kết h với $h2$.
- θ : là vector tham số của mô hình, bao gồm tất cả các tham số cần thiết để xác định hàm năng lượng.

Trong đó $\theta = [W1, W2]$ là các tham số của mô hình, biểu thị tương tác giữa lớp hiển thị với tầng ẩn và giữa từng tầng ẩn khác.

Khi so sánh với hàm năng lượng của RBM, công thức sẽ là:

$$E(v, h) = -b_v^t - c^T \cdot h - v^T \cdot wh \quad (17)$$

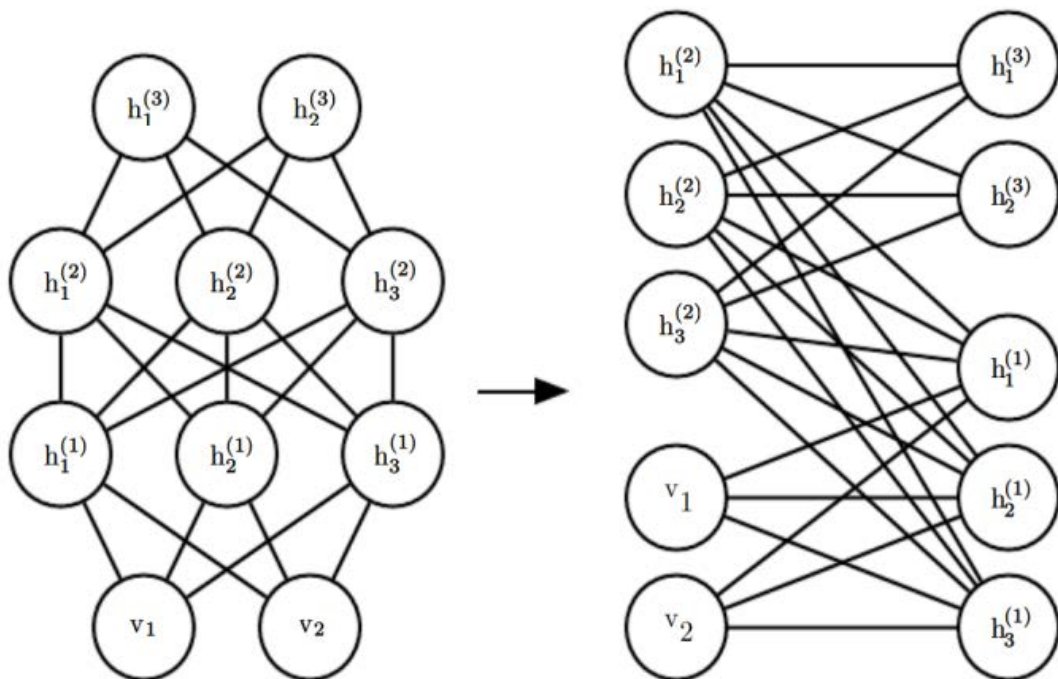
Trong đó:

- $E(v, h)$: Là hàm năng lượng của mạng DBM.
- v : Là vector của visible units.
- h : Là vector của hidden units.

- b_v^t và c^T : Là các vector độ lệch (bias) tương ứng của visible units và hidden units.
- w : Là ma trận trọng số kết nối giữa visible units và hidden units.

Sau giai đoạn pretraining, toàn bộ mô hình DBM được huấn luyện thông qua quá trình tối ưu hóa hàm năng lượng (energy function). Mục tiêu là điều chỉnh các trọng số và độ lệch trong mạng sao cho phân bố xác suất của dữ liệu được xấp xỉ tốt nhất với phân bố xác suất thực sự của dữ liệu. Quá trình này giúp tạo ra một biểu diễn sâu và phức tạp của dữ liệu, cho phép DBM nắm bắt những đặc trưng quan trọng và phức tạp trong dữ liệu đầu vào.

So với mạng BM được kết nối đầy đủ (các đơn vị được kết nối với các đơn vị khác), mạng DBM mang đến lợi ích tương tự được cấp bởi mạng RBM. Ngoài ra các lớp DBM có thể được tổ chức như một đồ thị.



Hình 3.5. Mô hình DBM được sắp xếp thành đồ thị

Trong đồ thị, mạng DBM có thể sử dụng các phương trình tương tự đối với các điều kiện của RBM, các đơn vị trong mỗi lớp có điều kiện độc lập với nhau dựa trên giá trị của các lớp lân cận từ đó suy ra sự phân phối trên các biến nhị phân có thể được mô tả bởi các tham số Bernoulli.

Khi tính toán phân phối $P(h/v)$, công thức đưa ra có tính xấp xỉ đơn giản hơn rất nhiều so với mạng DBN. Việc phân loại được thực hiện bằng phương pháp tự

suy luận gần đúng, hay được hiểu rằng trường trung bình của các đơn vị ẩn được cấp phát bởi các bước trong mạng neural truyền thẳng nhiều lớp (MLP) với hàm sigmoid và cùng có trọng số tương tự DBN. Ngoài ra vấn đề thiếu đi tương tác giữa các lớp có thể sử dụng các phương trình điểm cố định để tối ưu hóa giới hạn dưới biến đổi.

Khi sử dụng mô hình mạng DBM trong khoa học thần kinh, Sargur N. Srihari (2020) nêu lên việc sử dụng trường trung bình thích hợp cho phép tính suy luận gần đúng để nắm rõ tương tác phản hồi từ trên xuống dưới [15]. Điều này làm cấu trúc của mạng DBM trở nên thú vị theo quan điểm từ khoa học thần kinh bởi vì bộ não con người cũng được biết là sử dụng nhiều phương pháp kết nối phản hồi từ trên xuống dưới do đó mạng DBM được sử dụng làm công cụ mô hình tính toán cho các hiện tượng khoa học thần kinh thật sự.

Mạng Deep Boltzmann Machine (DBM) đã được chứng minh là một công cụ mạnh mẽ và linh hoạt trong việc học biểu diễn sâu và xử lý dữ liệu phức tạp, nhất là trong lĩnh vực xử lý ảnh, xử lý ngôn ngữ tự nhiên và các tác vụ phân loại phức tạp. Tính biểu diễn học được từ DBM đã cho phép nó đạt được những thành tựu đáng kể trong nhiều lĩnh vực khác nhau của học máy và trí tuệ nhân tạo, bao gồm cả các ứng dụng yêu cầu độ chính xác và tính tối ưu cao.

Trong bối cảnh của nghiên cứu nhận diện khuôn mặt đeo khẩu trang, việc lựa chọn sử dụng mạng DBM để thử nghiệm không chỉ dựa trên khả năng xử lý dữ liệu ảnh phức tạp mà còn bởi khả năng của nó trong việc trích xuất và học hỏi những đặc trưng sâu và phức tạp từ dữ liệu. Nhận diện khuôn mặt đeo khẩu trang là một thách thức đặc biệt vì khẩu trang che đi một phần khuôn mặt, làm giảm số lượng thông tin có sẵn mà các mô hình học máy truyền thống có thể sử dụng để nhận diện. Tuy nhiên, nhờ vào khả năng của DBM trong việc học biểu diễn các tín hiệu ẩn và giải mã các mối liên hệ phức tạp giữa các đặc điểm trên dữ liệu, mô hình này có thể phát hiện và hiểu các mẫu biến đổi phức tạp, bao gồm cả việc nhận diện các đặc điểm khuôn mặt khi bị che khuất bởi khẩu trang.

DBM, với cấu trúc đa tầng và khả năng mô hình hóa phân phối chung của dữ liệu, cung cấp một phương pháp tiếp cận mạnh mẽ để tổng hợp và phân tích dữ liệu từ nhiều nguồn khác nhau, giúp tăng cường khả năng nhận diện và phân biệt giữa khuôn mặt có đeo khẩu trang và không đeo khẩu trang. Việc áp dụng DBM trong bài toán này cũng cho phép tận dụng lợi thế của mô hình trong việc tự động

học hỏi các đặc trưng phức tạp mà không cần sự can thiệp nhiều từ con người, điều này rất quan trọng trong các ứng dụng thực tế nơi mà việc thu thập và gán nhãn dữ liệu thường xuyên là không khả thi. Kết quả là, DBM không chỉ cải thiện độ chính xác trong nhận diện khuôn mặt đeo khẩu trang mà còn góp phần vào sự phát triển của các hệ thống tự động có khả năng thích ứng và phản ứng hiệu quả với những thay đổi trong môi trường thực tế.

DBM là một mô hình deep learning mạnh mẽ có khả năng học cách biểu diễn dữ liệu phức tạp và học các đặc trưng phân cấp từ dữ liệu đầu vào. Sự linh hoạt của DBM trong việc mô hình hóa phân phối xác suất của dữ liệu cho phép nó xử lý các dạng dữ liệu đa dạng và không cần phải có giả định về phân phối dữ liệu. Với bài toán nhận dạng khuôn mặt đeo khẩu trang, việc xử lý và biểu diễn dữ liệu khuôn mặt có thể rất phức tạp vì sự biến đổi lớn giữa các khuôn mặt khác nhau và hiệu ứng của việc đeo khẩu trang. Một DBM có thể học được các biểu diễn cấp cao của khuôn mặt và đồng thời xử lý được sự biến đổi do việc đeo khẩu trang. Mạng DBM có khả năng học tự giám sát (unsupervised learning) mà không cần dữ liệu được gán nhãn, điều này có thể giúp tiết kiệm thời gian và chi phí trong quá trình huấn luyện.

PHẦN 4. KẾT QUẢ NGHIÊN CỨU

4.1. TIỀN XỬ LÝ DỮ LIỆU NGHIÊN CỨU

4.1.1. Thu thập dữ liệu

Dữ liệu trong nghiên cứu này dựa trên sự tham gia của nhóm và được sự cho phép của 30 sinh viên Học viện Nông nghiệp Việt Nam để sử dụng dữ liệu thông tin về khuôn mặt, ngoài ra còn có một số nguồn dữ liệu hình ảnh khác nhau trên mạng Internet như Maskk.com, Kaggle.com...v.v để góp phần đa dạng cơ sở dữ liệu hiện có. Trong khâu thu thập dữ liệu của các sinh viên Nông nghiệp, tôi đã lên các bước thực hiện sau:

Bước 1: Thiết kế bảng thu thập dữ liệu mẫu. Trong bảng thu thập này sẽ bao gồm thông tin về đề tài nghiên cứu, mục đích và cam kết cũng như tên và chữ ký của người được thu thập. Ở bảng này những người tham gia được khái quát về mục đích của việc lấy là mẫu hoàn toàn tự nguyện của bản thân và các dữ liệu cá nhân này sẽ chỉ được dùng trong công tác nghiên cứu thí nghiệm cũng như không được để lộ ra bên ngoài dưới bất kỳ hình thức nào.

Bước 2: Tìm hiểu về thiết bị ghi hình và chụp ảnh. Dựa trên điều kiện vật chất và các công cụ hiện có, các dữ liệu ban đầu sẽ được quay thành video cận cảnh khuôn mặt thường và khi đeo khẩu trang trong các góc quay khác nhau để sử dụng cho công việc nhận diện video. Việc quay video mang lại rất nhiều lợi ích để có thể thu được nhiều hình ảnh trong công đoạn xử lý dữ liệu sau này và có thể lựa chọn mẫu video có khuôn mặt ở bên trong khung hình và đường nét rõ ràng để sử dụng cho sinh viên đó.

Bước 3: Lên quy trình và thực hiện lấy mẫu. Theo kế hoạch đã phân công, các sinh viên sẽ lập lịch và thực hiện công việc quay phim dữ liệu tại khoa Công nghệ Thông tin, Học viện Nông nghiệp Việt Nam. Trong quá trình lấy dữ liệu, các sinh viên sẽ thử nghiệm các góc quay, khoảng cách lựa chọn vị trí phù hợp cho mục đích tổng thể khuôn mặt luôn xuất hiện trên đoạn dữ liệu. Trong đó, mỗi bạn sẽ sử dụng khẩu trang y tế với nhiều màu sắc từ trắng, xám, xanh lục, đen,... làm phong phú dữ liệu trong mô hình phân loại sau.

Bảng 4.1. Dữ liệu thu thập từ sinh viên tình nguyện tại Học viện Nông nghiệp Việt Nam

Trường hợp	Nam	Nữ	Kiểu dữ liệu	Độ tuổi	Số ảnh/người
Đeo khẩu trang (with_mask)	28	2	mp4	20 - 22	10
Không đeo khẩu trang (without_mask)	28	2	mp4	20 - 22	10

Sau khi thu được tập video cụ thể của mỗi người, dữ liệu hiện nay vẫn chưa được sử dụng bởi mô hình học sâu không thể xử lý được thông tin rất lớn của từng đoạn phim, thậm chí làm phức tạp khả năng đào tạo mô hình và gây ra kết quả sai sót lớn. Cách giải quyết đơn giản nhất là xử lý dữ liệu thô đó sang dạng dữ liệu phù hợp với đầu vào của mô hình, đó chính là hình ảnh. Hình ảnh khi được căn chỉnh mang lại nhiều lợi ích trong quá trình huấn luyện và kiểm tra mô hình, cũng như đảm bảo tính nhất quán giữa các hình ảnh trong quá trình nhận diện khuôn mặt đeo khẩu trang.

4.1.2. Tiền xử lý dữ liệu

Để đảm bảo khả năng nhận diện mạnh mẽ giữa khuôn mặt đeo khẩu trang và khuôn mặt thường, quá trình tiền xử lý dữ liệu không chỉ là một bước đơn thuần, mà là một quy trình chi tiết đặc biệt quan trọng trong chuẩn bị dữ liệu cho mô hình. Mục tiêu chính là tạo ra một tập dữ liệu có tính năng đặc biệt, giúp mô hình hiểu biết và trích xuất các điểm đặc trưng quan trọng trên khuôn mặt của nhiều người khác nhau, từ đó tạo nền tảng vững chắc cho khả năng nhận diện chính xác của mô hình.

Trong quá trình tiền xử lý, việc căn chỉnh hình ảnh đóng vai trò quan trọng, nhằm đảm bảo sự đồng đều về chất lượng và tính nhất quán trong dữ liệu. Điều này có thể bao gồm việc điều chỉnh cường độ sáng để làm nổi bật các đặc điểm trên khuôn mặt, cân bằng độ tương phản để làm cho hình ảnh rõ ràng, và thậm chí xoay hoặc phóng to/hẹp hình ảnh để có góc nhìn và kích thước chuẩn xác. Mặc dù vậy, việc thay đổi các đặc điểm của hình ảnh đó cũng mang đến nhiều rủi ro khi áp dụng vào mô hình nhận dạng vào bài toán thực tế khi các đặc điểm chính trên khuôn mặt làm thay đổi hiệu suất nhận diện của mô hình khi phát hiện khẩu trang trên khuôn mặt người đó và làm dữ liệu hình ảnh không phù hợp với vấn đề thực tế.

Bước tiếp theo của công đoạn căn chỉnh hình ảnh là xử lý kích thước hình ảnh để tất cả chúng có kích thước đồng nhất. Điều này đảm bảo mô hình được đào tạo trên dữ liệu có định dạng thống nhất, từ đó tránh được sự biến động không mong muốn trong quá trình học.

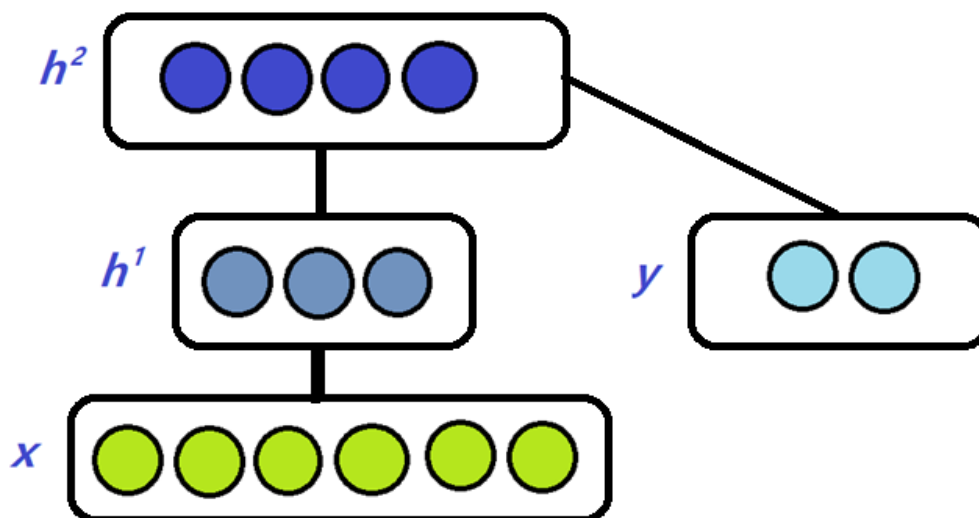
Ngoài ra, quá trình tiền xử lý có thể bao gồm việc loại bỏ nhiễu từ hình ảnh, chẳng hạn như ánh sáng môi trường không đồng đều hoặc đối tượng nền không mong muốn. Các kỹ thuật như lọc thông tin và cắt ghép có thể được áp dụng để tăng cường chất lượng của dữ liệu, giúp mô hình tập trung vào những đặc điểm quan trọng trong quá trình học.



Hình 4.1. Hình ảnh đã qua tiền xử lý

Quá trình tiền xử lý dữ liệu không chỉ là bước chuẩn bị đơn thuần, mà là một chiến lược tích hợp và chi tiết, giúp tối ưu hóa đặc trưng và đồng thời đảm bảo tính nhất quán và chất lượng của tập dữ liệu. Công đoạn này sẽ đóng vai trò quan trọng trong việc đạt được khả năng nhận diện chính xác và hiệu quả của mô hình trên đa dạng hình ảnh khuôn mặt.

4.2. CẤU HÌNH MẠNG DEEP BOLTZMANN MACHINE



Hình 4.2. Sơ đồ mô hình Deep Boltzmann Machine

Trong cấu trúc của ứng dụng trí tuệ nhân tạo, thứ quan trọng nhất là mô hình của nó. Việc lựa chọn một mô hình phù hợp sẽ là tiền đề để giúp bài toán đang được nghiên cứu giải quyết nhanh chóng.

Đối với đề tài nhận diện khẩu trang, việc xây dựng một mô hình phù hợp là bước quyết định không thể thiếu. Mô hình, được xem như "bộ não" của trí tuệ nhân tạo, được thiết kế với những ý tưởng và cấu trúc tương tự như bộ não của con người, với các đường neural để học và phân tích mọi khía cạnh của tập dữ liệu đa dạng.

Trong đề tài này, sự lựa chọn chủ yếu dựa vào việc sử dụng thư viện scikit-learn, một công cụ mạnh mẽ và linh hoạt trong việc phát triển mô hình máy học. Scikit-learn không chỉ giúp phân loại lớp, đào tạo, và kiểm tra mô hình một cách hiệu quả mà còn cung cấp khả năng đánh giá chi tiết và trực quan hóa thông qua biểu đồ tổng quát về độ chính xác và hàm mất mát trên từng bước đào tạo.

Quá trình này không chỉ là việc xây dựng một mô hình Deep Boltzmann Machine (DBM), mà còn là sự kết hợp và tận dụng các công cụ mạnh mẽ từ scikit-learn để tối ưu hóa quá trình nghiên cứu.

Nhờ vào sự tích hợp của DBM và scikit-learn, các công đoạn để triển khai, điều chỉnh, phân loại lớp, kiểm tra được giảm bớt gánh nặng trong quá trình thử nghiệm và điều chỉnh, tăng cường khả năng linh hoạt và dễ dàng trong quá trình thử nghiệm và đánh giá mô hình một cách tối ưu nhất.

Sự kết hợp này không chỉ tạo nên một môi trường làm việc thuận lợi mà còn giúp nâng cao chất lượng và hiệu suất của quá trình nghiên cứu. Kết quả là, người nghiên cứu có thể nhanh chóng đưa ra giải pháp đáng tin cậy cho bài toán nhận diện khẩu trang, đồng thời quá trình thực hiện không bị phức tạp trong các kỹ thuật không cần thiết.

4.2.1. Biên dịch dữ liệu

Để máy tính có thể học, điều đầu tiên là phải cho máy tính biết nó sẽ phải học cái gì. Tương tự như bộ não con người, thứ cần thiết để khi bắt đầu quá trình học chính là kiến thức, là dữ liệu liên quan đến công việc mình đang thực hiện.

Ở dữ liệu hình ảnh, máy tính không thể xử lý được hết từng ảnh một vì độ lớn của dữ liệu sẽ làm chậm khả năng vận hành của mô hình và có thể tiêu tốn tài nguyên dung lượng với những điểm dư thừa khi dữ liệu hình ảnh đó có nhiều điểm pixel không phải khuôn mặt, vấn đề đó được hiểu rằng trong một tấm ảnh màu với một điểm ảnh pixel bất kỳ sẽ có không gian màu khác nhau.

Khi đã có tập dữ liệu hình ảnh đã được xử lý để sử dụng cần phải duyệt thư mục chứa dữ liệu đó đầu tiên để lấy thông tin. Có 2 đường dẫn tương ứng là “FOLDER” dẫn đến thư mục chứa dữ liệu hình ảnh và “TYPE” là danh sách chứa các loại hình ảnh như trên, tập ảnh đeo khẩu trang và khuôn mặt thường.

Để đọc hình ảnh với mỗi tệp ảnh, một vòng lặp duyệt qua từng hình ảnh trong đường dẫn “TYPE”, cùng với sử dụng hàm `os.path` để lấy danh sách các tệp hình ảnh trong thư mục đó và thực hiện từng nhiệm vụ bên trong vòng lặp. Bằng hàm `cv2.imread`, hình ảnh được đọc và đưa vào máy tính trong khi đó hàm `cv2.resize` thực hiện thay đổi kích thước hình ảnh thành 128x128 pixel rồi truyền xuống công đoạn xử lý tiếp theo. Mục đích của việc giảm kích thước sẽ tối ưu chuyển hóa hình ảnh sang dạng mảng 1 chiều được lưu vào hàm `data` còn nhãn của tệp ảnh truyền vào hàm `labels` để sử dụng.

Hiểu thêm về chuyển hóa dạng 1 chiều, hình ảnh ban đầu có dạng ma trận 2 chiều khi được thu nhỏ thành 128x128 pixel vẫn còn nhiều tham số cần học trong mô hình do đó công việc tính toán của máy tính sẽ chậm đi khi xử lý thêm dữ liệu đầu vào. Trong khi đó hình ảnh được chuyển sang dạng véc tơ 1 chiều có kích thước 1x16384 (128x128), phù hợp với mô hình của học máy cũng như học sâu, giảm được chiều dữ liệu và làm mô hình sau này dễ quản lý hơn.

Sau cùng chuyển 2 hàm data và label thành dạng mảng numpy. Dữ liệu đã được chuyển đổi khi được đổi sang mảng numpy sẽ sử dụng hàm toán học và các phép toán đại số tuyến tính để tích hợp sử dụng các phương pháp và hàm của các thư viện học máy, học sâu như TensorFlow, Pytorch hay thậm chí cả thư viện Numpy sau này. Còn khi chuyển nhãn lớp label_indices, các dạng “with_mask” và “without_mask” sẽ đánh số là 0 và 1, được đặt làm giá trị đại diện cho việc xác định về sau.

4.2.2. Huấn luyện, kiểm tra mô hình

Sau khi chuyển đổi dữ liệu và nhãn để huấn luyện, bước tiếp theo để xây dựng mô hình là huấn luyện và kiểm tra mô hình đó.

Trong biến khai báo đầu tiên của tạo mô hình là chia dữ liệu của tập dữ liệu cho huấn luyện và kiểm tra. Biến khai báo bên ngoài được giải thích khái quát ý nghĩa để xây dựng phần dữ liệu, nhãn cho huấn luyện và kiểm tra như sau:

train_x: là tập dữ liệu hình ảnh dùng cho huấn luyện.

test_x: là tập dữ liệu hình ảnh để kiểm tra.

train_y: là nhãn tương ứng với tập dữ liệu huấn luyện.

test_y: là nhãn cho tập dữ liệu kiểm tra.

Mục đích chia dữ liệu dùng để kiểm soát dự đoán của mô hình trên dữ liệu độc lập, tối ưu hóa tham số và ngăn chặn quá trình học quá mức (Overfitting) mà không thể tổng quát hóa tốt nếu dùng toàn bộ dữ liệu cho tập huấn luyện.

Từ thư viện Scikit-learn, hàm train_test_split được sử dụng trong việc chia dữ liệu từ đầu vào. Bên trong hàm đó biến data và label_indices được xử lý ở trên và nhập tỷ lệ 20% dữ liệu cho việc kiểm tra còn lại 80% để huấn luyện mô hình mạng. Hàm stratify đảm bảo phân phối các lớp trong tập dữ liệu ban đầu và nhãn của chúng được chia như nhau, điều này có ích trong vai trò coi tập huấn luyện và kiểm tra đều có các lớp được đại diện đầy đủ.

Tiếp đó, tôi khởi tạo hàm StandardScaler cũng trong thư viện Scikit-learn, nó là một bộ giả định thống kê được sử dụng cho biến đổi dữ liệu sao cho giá trị trung bình bằng 0 và độ lệch chuẩn bằng 1. Trước khi chuẩn hóa, train_x và test_x được biến đổi thành một ma trận có số hàng bằng số lượng tương ứng của chúng và số cột bằng số lượng đặc trưng.

Mạng DBM là sự xếp chồng của nhiều lớp RBMs (Lecun & cộng sự., ,2015). Với thuật toán trong mạng RBM, nó cho phép tầng ẩn học phân phối của đầu ra, và đồng thời tham gia như một đầu ra của tầng ẩn khác. Trong đề tài tôi sử dụng mạng DBM với 3 tầng ẩn trong đó tầng ẩn thứ ba có sự tham gia của mạng classRBM để đào tạo học có giám sát kết hợp với nhãn để phân loại hình ảnh.

Cấu trúc cho mạng DBM trong nghiên cứu gồm có 3 tầng ẩn có sự tham gia của hàm BernoulliRBM của thư viện Scikit-learn, trong đó:

Ở tầng đầu vào x và tầng ẩn h_1 , lớp `rbm_layer1` đầu tiên có 200 neural tham gia. Các tham số W , U , b , c ban đầu được sinh ngẫu nhiên và nhỏ trong giới hạn $[10^0, 10^{-3}]$ bằng biến `learning_rate`, và từ biến `verbose` phản ánh mức độ chi tiết của quá trình huấn luyện, ở đây mô hình sẽ in thông tin về quá trình huấn luyện sau mỗi 10 lần lặp.

Tương tự như tầng ẩn trên, lớp `rbm_layer2` đều có các tham số đầu vào và thông số tương tự, ngoại trừ trong tầng ẩn này số neural giảm xuống còn 100.

Tầng ẩn classRBM cuối cùng có chức năng như 1 lớp phân loại (classifier) có 2 neural, được hiệu chỉnh như các tầng ẩn phía trên.

Sau khi thiết lập từng lớp RBM cho mạng DBM, tôi sẽ thực hiện công việc huấn luyện các mô hình trên dữ liệu huấn luyện. Cụ thể từng chuỗi của các tầng RBM đều giống nhau về cấu trúc và khác nhau về tên gọi của biến tầng ẩn.

Cấu trúc chung của tầng ẩn như sau:

Biến `rbm_layer` có phương thức `fit` sẽ tối ưu hóa các tham số của mô hình dựa trên `train_x` (là hình ảnh đã qua xử lý) còn biến `class_rbm_layer` của tầng classRBM có thêm `train_y` (nhãn tương ứng tập dữ liệu).

Chuyển đổi dữ liệu của `train_x` và `test_x` của tầng ẩn trên cho tầng ẩn tiếp theo bằng hàm `transform`.

Ở tầng ẩn classRBM tôi sẽ thêm 1 mô hình tuyến tính được gọi là Logistic Regression. Trong dạng này, đầu ra có thể được thể hiện dưới dạng xác suất. Mô hình logistic regression giống với linear regression ở khía cạnh đầu ra là số thực và giống với PLA ở việc đầu ra bị chặn ở khoảng $[0,1]$ tương tự như logistic function. Mô hình tuyến tính được thực hiện 1000 bước lặp cho dữ liệu bài toán được hội tụ tại 1 điểm ngẫu nhiên được coi là cân bằng nhiệt động lực học, để huấn luyện bộ phân loại của biến `train_x` của classRBM và nhãn `test_y`.

Sau cùng, tập dữ liệu kiểm tra `test_x` của tầng ẩn của `classRBM` được đánh giá thông qua đường vector trên mỗi đường lặp của mô hình mạng DBM, kết quả đặt tên là `predictions` sẽ đưa ra dự đoán của mô hình của tập kiểm tra so với tập huấn luyện đã được máy tính học.

4.3. KẾT QUẢ NHẬN DIỆN

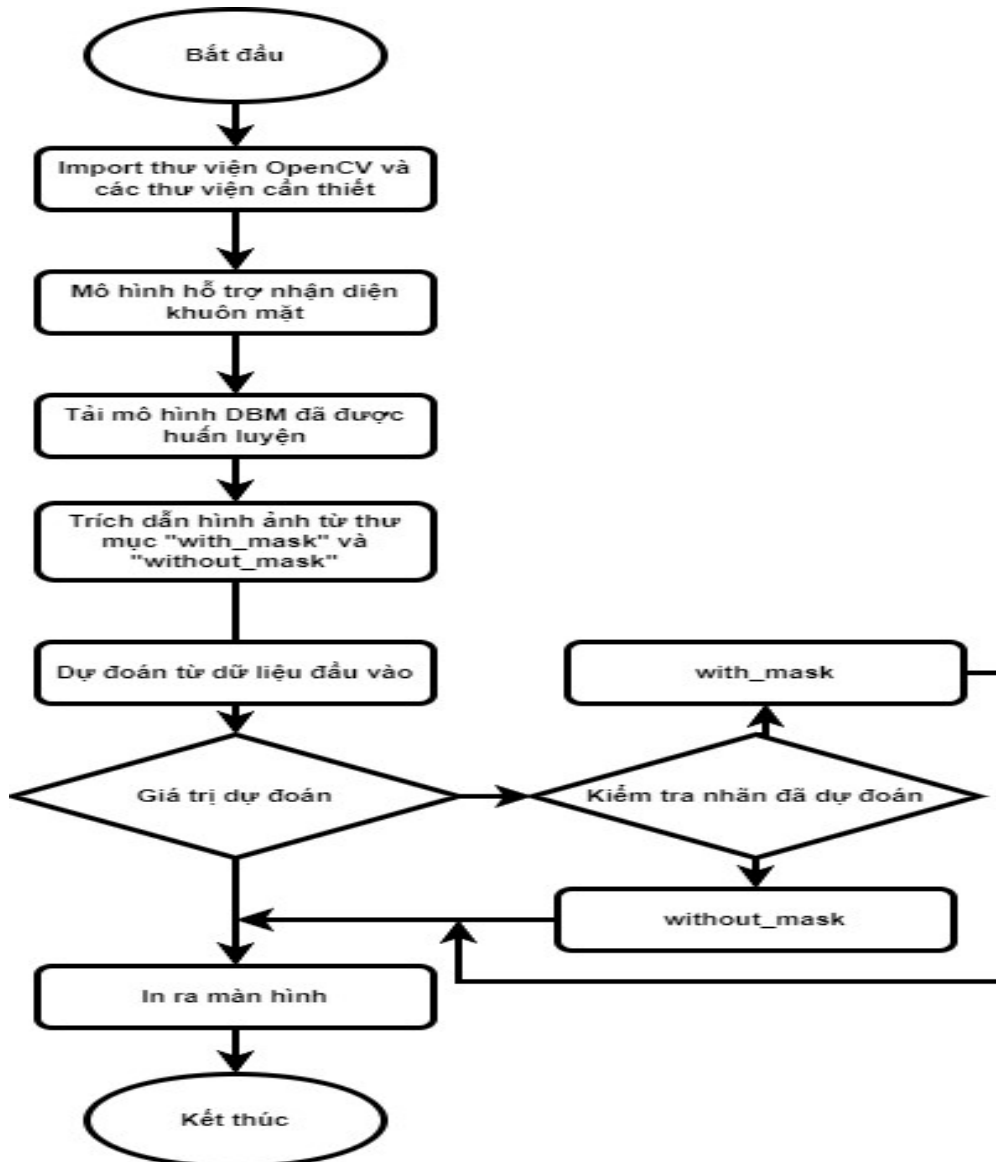
Sau khi đã có mô hình hoàn chỉnh, việc cần làm tiếp theo là dự đoán và đánh giá phân loại hình ảnh của mô hình đã kiểm tra để dự đoán các mẫu trong nhãn `test_y`. Kết quả sẽ cho ra xác suất dự đoán của mô hình.

Để tính độ chính xác một cách hiệu quả, trong thư Scikit-learn tôi dùng thư viện `classification_report` để máy tính tính toán độ chính xác giữa nhãn `test_y` và kết quả được đánh giá ở bước trên và in ra kết quả được thể hiện ở hình dưới đây.

Bảng 4.2. Kết quả độ chính xác trong huấn luyện và kiểm tra của DBM.

Nhãn	Precision	Recall	F1-score	Support
0	52%	65%	58%	60
1	53%	40%	46%	60
Accuracy			53%	120
Macro avg	53%	53%	52%	120
Weighted avg	53%	53%	52%	120

Sau khi thiết lập một mô hình hoàn chỉnh, kết quả của mô hình đó sẽ được lưu thành một tệp pickle để tái sử dụng các chỉ số trong mô hình nhận diện khuôn trang sau này. Tập tin được lưu có tên là “DBM.pkl” có kết quả được ghi lại dưới dạng số nhị phân.



Sơ đồ 4.1. Sơ đồ hệ thống nhận diện khuôn mặt đeo khẩu trang

Khi mô hình của bài toán phân loại hình ảnh được xây dựng và hoàn thiện, giai đoạn tiếp theo là nghiên cứu các mô hình nhận diện khuôn mặt để xác định khuôn mặt trên định dạng cụ thể và áp dụng mô hình cho bài toán thực tế.

Trong thời gian nghiên cứu các mô hình định dạng khuôn mặt khác nhau, tiêu chí lựa chọn đầu tiên là tối ưu hóa tài nguyên và khả năng nhận diện khuôn mặt hiệu quả tạo thuận lợi cho công việc xác định khẩu trang trên khuôn mặt, cùng với độ chính xác cao khi thay đổi yếu tố bên ngoài và ít tiêu tốn tài nguyên xử lý đồ họa. Trong đó Deep Neural Network(DNN) và Haar Cascade của thư viện mã nguồn mở OpenCV có đầy đủ các tiêu chí lựa chọn trên, và DNN là lựa chọn tốt nhất để nhận diện khuôn mặt trên hình ảnh.

4.3.1. Ứng dụng nhận diện khuôn mặt đeo khẩu trang

Sau khi đã lựa chọn xong mô hình nhận diện khuôn mặt là mô hình DNN, tôi tiến hành tạo chương trình nhận diện khuôn mặt đeo khẩu trang từ dữ liệu được xử lý bằng mô hình DBM. Ứng dụng dự định sẽ lựa chọn 3 chức năng để nhận diện bao gồm nhận diện bằng hình ảnh có trong dữ liệu, bằng video và cuối cùng thực hiện nhận diện bằng camera trên máy tính hay webcam được kết nối.

Bằng sự thuận tiện của giao diện người dùng, các phần chức năng đã được lồng vào trong 1 chương trình tổng mà không cần chia chúng thành các tệp nhỏ hơn để nhận diện. Khi đó trong cấu trúc của ứng dụng tôi có thể áp dụng thuật toán áp dụng mô hình nhận diện khuôn mặt và đặt vùng trung tâm (Region of interest) như là ô vuông trên các dữ liệu phân loại cần được kiểm tra khác nhau.

Trong nhận diện khẩu trang bằng hình ảnh và video, cấu trúc chung của chúng sẽ dùng dữ liệu được lưu tại cơ sở dữ liệu trên máy tính để sử dụng. Khi chọn một trong hai chức năng, hệ thống sẽ đưa ra một đường dẫn trong biểu tượng truy cập vào bộ nhớ và ổ đĩa (My Computer, This PC) để lựa chọn dữ liệu mà người dùng muốn kiểm tra. Còn với camera hệ thống sẽ cho ra kết quả màn hình với kỹ thuật phát trực tuyến trong thời gian thực như camera laptop, webcam kết nối với máy tính bàn hoặc camera an ninh.

Ở cấu trúc nhận diện chung tôi sẽ đặt các đường dẫn của mô hình của máy tính đã được đào tạo mạng DBM để so sánh giữa 2 khuôn mặt đeo khẩu trang với khuôn mặt thường và mô hình DNN để nhận diện khuôn mặt trên hình ảnh hoặc video cụ thể muốn kiểm tra.

4.3.2. Cấu trúc mô hình nhận diện khuôn mặt đeo khẩu trang

Mô hình đào tạo được lưu lại trên 1 thư mục riêng và tải lên trong chương trình cho việc gọi và sử dụng. Sau khi mô hình được lựa chọn đưa vào chương trình, nó sẽ được hàm joblib kết xuất và tải kết quả mà mô hình học sâu của mạng DBM được đào tạo có giá trị thuộc mảng numpy và lưu trữ trong biến của chương trình, khi đó chỉ cần gọi tên biến có chứa mô hình để sử dụng mà không cần đặt lại đường liên kết của tệp dữ liệu lần nữa.

Đối với mô hình DNN, đầu vào của nghiên cứu sử dụng 2 tệp mô hình lưu trữ thông tin của thư viện nhận diện trong đó 1 tệp dùng để lưu thông tin cấu trúc của mạng neural và tệp còn lại lưu thông tin về mức độ liên kết giữa các neural

(weight). Ngoài ra hàm `readNetFromTensorFlow` cho biết rằng hai tệp mô hình mạng nêu trên được lưu trữ ở định dạng của khung TensorFlow.

Khi đã thiết lập xong các mô hình để nhận diện khuôn mặt, tiếp theo đó sẽ là nhận diện đối tượng trên dữ liệu hình ảnh đầu vào. Điều này là vì mô hình của DNN chỉ hỗ trợ với hình ảnh thuộc màu RGB mà hình ảnh RGBA có mức độ rộng cao hơn để xử lý, với A là độ mờ của màu.

Sau khi đã chuyển đổi màu cho hình ảnh đầu vào, mạng DNN giờ sẽ tạo một blob từ hàm `blobFromImage`. Blob (Binary large object) thực hiện cho chuẩn hóa dữ liệu đầu vào hiện có và chuyển đổi hình ảnh phù hợp với mạng neural. Giống như kết quả đào tạo mô hình của mạng DBM, tập dữ liệu hình ảnh sẽ chuẩn hóa thành mã nhị phân. Các tham số của blob được liệt kê dưới đây:

`image`: là hình ảnh đầu vào được xử lý trước khi chuyển sang mạng DNN để phân loại.

`scalefactor`: là tham số dùng để thay đổi tỷ lệ hình ảnh đầu vào. Giá trị thường để mặc định là 1.0 là hình ảnh không chia tỷ lệ mà giữ thành dạng gốc.

`size`: là kích thước đầu ra của blob. Kích thước (300,300) tương ứng với `width`, `height` phù hợp với định dạng.

`mean`: là giá trị trừ trung bình của màu sắc từ hình ảnh đầu vào. Khi thực hiện phép trừ trung bình cân cung cấp 3 chỉ số của bộ màu RGB tương ứng với các số (104, 117, 123).

`swapRB`: theo mặc định của OpenCV hình ảnh đầu vào sẽ thực hiện việc hoán đổi từ màu RGB sang màu BGR. Trong bài toán này hình ảnh sẽ luôn sử dụng dải màu RGB nên giá trị đưa ra được đặt là “False”, nghĩa là không chuyển đổi giữa chúng.

Sau khi đã tạo được blob được định dạng với yêu cầu thông số trên, tôi thiết lập đầu vào cho mạng DNN bằng cách gán blob và thực hiện xác nhận bằng việc chạy mô hình mạng DNN trên đầu vào và đưa ra kết quả chứa khuôn mặt trong hình ảnh đó. Hình ảnh sau được chuyển đổi thành một đốm màu và truyền qua mạng bằng hàm `forward`.

Đối với mỗi đối tượng được phát hiện, điểm tin cậy được trích xuất từ mảng `detection`. Nếu điểm tin cậy lớn hơn 0.5, nó được coi là một phát hiện hợp lệ. Các kết quả phát hiện từ đầu ra là 1 ma trận 4 chiều, trong đó chiều thứ 3 lặp lại trên

các khuôn mặt được phát hiện với i là quy trình vòng lặp theo số lượng khuôn mặt và chiều thứ 4 chứa thông tin về hộp giới hạn và điểm cho mỗi khuôn mặt. Tọa độ đầu ra của hộp giới hạn được chuẩn hóa trong khoảng $[0,1]$. Do đó tọa độ phải được nhân với chiều cao và chiều rộng của ảnh gốc để có được hộp giới hạn chính xác trên ảnh.

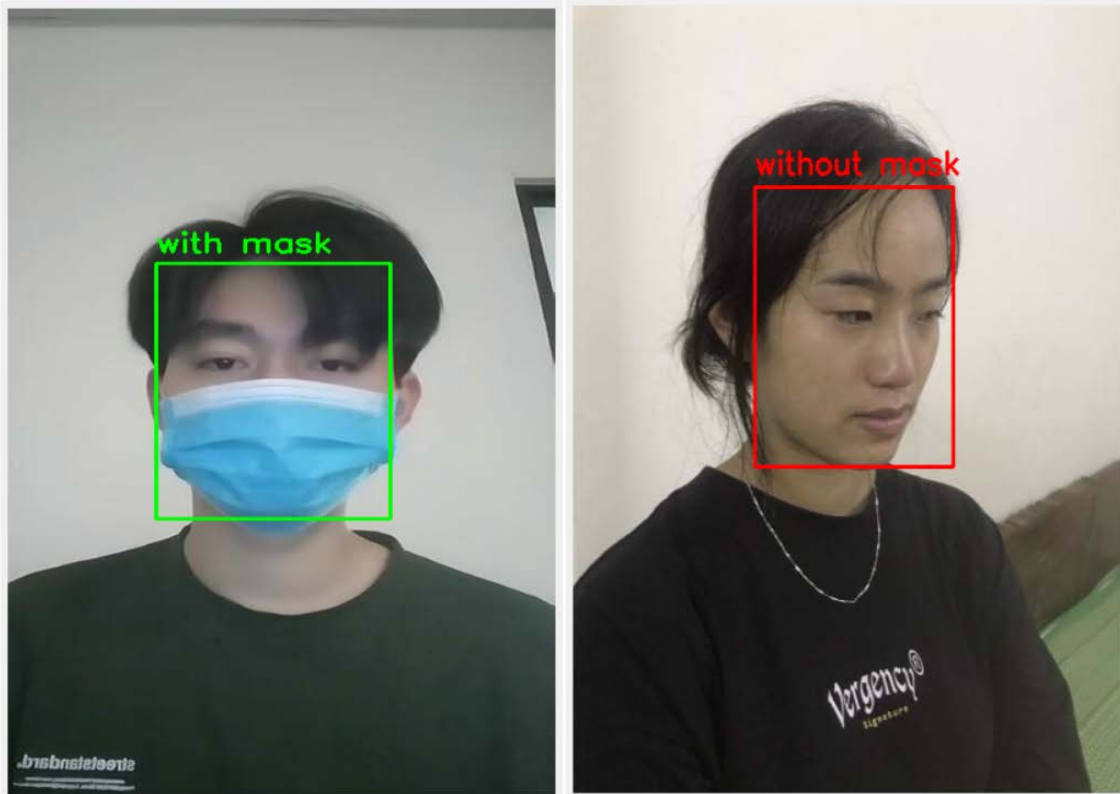
Sau khi tạo được hộp giới hạn để tạo vùng khuôn mặt, lúc này bên trong vùng hộp giới hạn đó chỉ còn hình ảnh khuôn mặt tại từng vị trí trên bức hình nhưng nhiệm vụ đặt ra vẫn chưa thể nhận biết được khuôn mặt đó có đeo khẩu trang hay không. Để trả lời cho bài toán đó, hình ảnh khuôn mặt được cắt trong hộp giới hạn cần phải được xử lý riêng trước khi đưa vào mô hình DBM để dự đoán.

Khi thực hiện tiền xử lý vùng khuôn mặt, trước tiên các hình ảnh khuôn mặt đặt trong biến face được thay đổi từ không gian màu BRG sang RGB để phù hợp với định dạng của nhiều mô hình thị giác máy tính. Cùng với đó kích thước của chúng được điều chỉnh thành 128×128 và khi chuyển đổi, một chiều mới được thêm vào và được chuẩn hóa giá trị pixel về khoảng từ 0 đến 1 bằng việc chia cho 255 tạo một định dạng mới đảm bảo đầu ra có kích thước cố định, chuẩn hóa và phù hợp với mô hình. Cuối cùng, hình ảnh được làm phẳng thành một vector 1 chiều. Tất cả các bước này nhằm mục đích tối ưu hóa đầu vào cho mô hình nhận dạng trong bước tiếp theo.

Sau khi hình ảnh khuôn mặt trong hộp giới hạn được cắt ra và xử lý, các kết quả của hình ảnh được lưu trong biến face sẽ chuyển đổi định dạng phù hợp với mô hình DBM. Với cấu trúc đã được định dạng cho mô hình được đào tạo, các kết quả của mạng DBM hiện tại là các vector đặc trưng một chiều. Để thay đổi định dạng cho mảng face nêu trên cần phải sử dụng phương thức reshape để biến đổi kết quả thành một mảng hai chiều với một hàng và tự động tính toán kích thước cột để giữ nguyên phần tử thành phần trong nó. Và khi đó các kết quả trong hộp giới hạn sẽ so sánh các thông số của mô hình mạng DBM đào tạo cho ra phương pháp phân biệt các hình ảnh người đeo khẩu trang và khuôn mặt người thường.

Tuy rằng hiện tại nghiên cứu đã có thể nhận diện được khẩu trang từ dữ liệu đưa vào nhưng máy tính vẫn chưa thể phân biệt được kết quả hiển thị rằng khuôn mặt đó đã đeo khẩu trang hay chưa. Bằng cấu trúc sẵn có của hộp giới hạn, kết quả so sánh mô hình nhận diện của mạng DBM được đối chiếu với giá trị $[0,1]$ rồi thay đổi màu cho nó. Khi kết quả lớn hơn 0.5, ở trên hộp giới hạn sẽ thông báo nhãn

với dòng chữ “with_mask” và màu của chữ cùng với khung hộp cho ra màu xanh lục còn ngược lại với khuôn mặt người thường, dòng chữ “without_mask” và hộp sẽ chuyển thành màu đỏ biểu thị cảnh báo vi phạm.



Hình 4.3. Kết quả nhận diện khuôn mặt đeo khẩu trang trên màn hình của mô hình Deep Boltzmann Machine

Việc xác thực kết quả mô hình ứng dụng nhận diện của mạng DBM trong nhận diện khuôn mặt đeo khẩu trang cần phải có kết quả tốt, chính xác nhất mới có thể được coi là mô hình ứng dụng. Tuy nhiên thực hiện ứng dụng mô hình cần được kiểm chứng rồi phân loại dựa theo điều kiện cụ thể đưa ra kết quả xác nhận, trong đó nghiên cứu thực nghiệm là nhiệm vụ đáng quan tâm nhất để kiểm thử, đánh giá chi tiết nhất.

Để so sánh khả năng học của DBM, tôi so sánh kết quả của DBM với NN trong cùng cấu trúc kích thước của mạng, có thể hiểu là NN được sử dụng có 3 tầng ẩn có kích thước lần lượt là, và các tham số của mạng NN cũng được thiết lập tương tự như DBM. Thống kê từ kết quả tốt nhất cho việc nhận diện khuôn mặt và đeo khẩu trang giữa DBM và NN được thể hiện ở bảng sau:

Bảng 4.3. Kết quả độ chính xác trong nhận diện khuôn mặt và khẩu trang của DBM , NN và SVM

Độ chính xác nhận diện khẩu trang và khuôn mặt	DBM	NN	SVM
Độ chính xác của nhận diện khuôn mặt	63%	60%	59%
Độ chính xác của nhận diện khẩu trang	65%	57%	60%

Kết quả so sánh cho thấy mô hình DBM cho kết quả vượt trội so với mô hình NN và SVM, vốn là cấu trúc mô hình nhận diện cơ bản cho bài toán nhận diện khuôn mặt và phân loại. Mô hình DBM còn cho thấy sự ổn định với các điều kiện khác nhau của bức ảnh khuôn mặt như điều kiện ánh sáng thay đổi, góc chụp thay đổi và độ che khuất khuôn mặt. Với các ưu điểm đó, mô hình phân loại vật lý trong học sâu, nhất là mô hình DBM, là mô hình tốt để ứng dụng cho việc nhận dạng khuôn mặt đeo khẩu trang.

PHẦN 5. KẾT LUẬN

5.1. KẾT LUẬN

Luận văn đã tiếp cận được những phương pháp học sâu và mạng DBM trong bài toán phân loại hình ảnh được nghiên cứu và công bố trên thế giới. Dựa vào đó luận văn đã tiến hành phân tích và xây dựng mô hình mạng DBM cho nhận diện khuôn mặt đeo khẩu trang.

Việc sử dụng mạng DBM trong nghiên cứu này cũng đã chứng minh khả năng nhận diện khuôn mặt đeo khẩu trang và cho thấy ứng dụng của học sâu trong nhiều lĩnh vực phân loại nói chung và nhận diện khẩu trang nói riêng. Dù vậy mô hình DBM cũng cần phải cải thiện đáng kể để nâng cao việc học sau này và điều chỉnh các tham số cũng như cấu trúc mạng cho việc đánh giá kết quả của mô hình.

5.2. HƯỚNG PHÁT TRIỂN

Tuy nhiên điểm khó khăn của mô hình DBM hiện tại là việc không thể phân loại hình ảnh dựa trên các lớp mô hình RBM đơn giản của thư viện Scikit-learn. Vì vậy, phương pháp đề xuất là sử dụng các thư viện hỗ trợ học máy khác như Keras, TensorFlow,... để xây dựng mạng DBM có độ chính xác cao, đồng thời tối ưu hóa dữ liệu lớn.

Nghiên cứu đã được thu thập dữ liệu video và hình ảnh khuôn mặt và đeo khẩu trang từ 30 sinh viên tình nguyện từ Học viện Nông nghiệp Việt Nam. Mặc dù dữ liệu thu thập được chưa đủ lớn và chưa thể đa dạng so với dữ liệu hình ảnh nói chung, nghiên cứu cũng góp phần bổ sung cơ sở và dữ liệu cần thiết cho các nghiên cứu sâu hơn sau này.

Để hệ thống nhận diện khẩu trang có độ chính xác cao hơn thì hướng phát triển dưới đây sẽ giúp mô hình tăng hiệu quả nhận diện, đồng thời phát triển mô hình hơn nữa.

Nhằm nâng cao khả năng tổng quát hóa và độ chính xác cao của mô hình, cần mở rộng tập dữ liệu đào tạo bằng cách thu thập các thu thập nhiều hình ảnh khuôn mặt người đeo khẩu trang từ các nguồn đa dạng khác nhau. Điều đó giúp mô hình học đa dạng các biến thể của khẩu trang và khuôn mặt trong nhiều tình huống cụ thể.

Ngoài việc áp dụng DBM để nhận diện khuôn mặt đeo khẩu trang. Trong thời

gian tới, tôi đề xuất việc kết hợp các kỹ thuật khác với DBM và so sánh đánh giá với kết quả hiện tại.

Hướng phát triển của nghiên cứu từ những đặc trưng của khuôn mặt có thể truy xuất được độ tuổi, chủng tộc, giới tính,...Ngoài ra nhận diện khẩu trang có thể cải tiến các thuật toán để nhận diện với các khẩu trang khác nhau từ khẩu trang vải, khẩu trang cho quảng cáo hay mặt nạ phòng độc đeo trên mặt.

Trong thời gian tới, luận văn sẽ cố gắng tinh chỉnh dữ liệu để cải thiện hiệu suất của mô hình. Theo hướng tiếp cận hiện tại, mô hình DBM sẽ xây dựng với tính năng của GPU để xử lý các vấn đề của hình ảnh đồ họa tạo dựng mô hình chính xác và thiết thực hơn.

TÀI LIỆU THAM KHẢO

- Viettimes - tin tức và phân tích chuyên sâu kinh tế, Q. T. <https://viettimes.vn/tai-sao-moi-nguoi-lai-co-mot-dac-diem-rieng-tren-khuon-mat-khong-ai-giong-ai-100-post142384.html>.
- Trần Phạm Khánh Toàn (2022) Các nhân tố ảnh hưởng đến ý định đeo khẩu trang của người dân tại nơi công cộng để phòng chống dịch Covid-19 - Nghiên cứu thực nghiệm tại Thành phố Hồ Chí Minh.
- Chu, D. K., Akl, E. A., Duda, S., Solo, K., Yaacoub, S., Schünemann, H. J. & Schünemann, H. J. (2020). Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and Covid-19: A systematic review and meta-analysis. *The Lancet*, 395(10242), 1973-1987.
- Rieger, M. O. (2020). To wear or not to wear? Factors influencing wearing face masks in Germany during the Covid-19 pandemic. *Social Health and Behavior*. 3(2): 50-52.
- Phil Nelson (15/11/2022) "OpenCV Face Detection: Cascade Classifier vs. YuNet", <https://opencv.org/blog/opencv-face-detection-cascade-classifier-vs-yunet/#:~:text=YuNet%20correctly%20detected%2010%20faces,similar%20to%20the%20previous%20test>.
- Kim, Y. J., Cho, J. H., & Kang, S. W. (2020). Study on the relationship between leisure activity participation and wearing a mask among koreans during Covid-19 crisis: Using TPB model. *International Journal of Environmental Research and Public Health*, 17(20), Article 7674.
- Vikas Gupta (22/10/2018) "Face Detection – Dlib, OpenCV, and Deep Learning (C++ / Python)".<https://learnopencv.com/face-detection-opencv-dlib-and-deep-learning-c-python>
- “Việt Nam có công nghệ nhận diện cả khi đeo khẩu trang”Sở khoa học công nghệ, <https://sokhcn.camau.gov.vn/wps/portal/?1dmy&page=trangchitiet&urile=wcm%3Apath%3A/sokhvcnlibrary/siteofkhohocvacongnghe/tintucsukien/khoahoccongnghe/vncocongnghehandangkhaustrang>.
- Bùi Xuân Quang và cộng sự (12/2022) "Nhận diện người đeo khẩu trang bằng thời gian thực tế". studocu, 2022, <https://www.studocu.com/vn/document/truong-dai-hoc-su-pham-ky-thuat-thanh-pho-ho-chi-minh/cung-cap-dien/do-an-2-xla-yes/61070500>.
- Mei Ngan và cộng sự (FRVT)(2020). "Face recognition accuracy with masks using pre-COVID-19 algorithms", <https://nvlpubs.nist.gov/nistpubs/ir/2020/NIST.IR.839911.pdf>.

Khuôn Mặt Đeo Khẩu Trang vì COVID-19 Sẽ Không Làm Khó Được Công Nghệ Mới.”
Viện Chiến Lược Thông Tin và Truyền Thông, niics.gov.vn/cong-nghe-thong-tin/khuon-mat-deo-khau-trang-vi-covid-19-se-khong-lam-kho-duoc-cong-nghe-moi.

Trung Quốc: Công nghệ nhận diện phát triển mạnh”Báo an ninh thế giới,
<https://antg.cand.com.vn/Khoa-hoc-Ky-thuat-hinh-su/Trung-Quoc-Cong-nghe-nhan-dien-phat-trien-manh-i575325/>

nttuan8 (2019) Bài 3: Neural network, Deep Learning cơ bản.<https://nttuan8.com/bai-3-neural-network/>

Ruslan Salakhutdinov và cộng sự (12/2022) "Deep Boltzmann Machines",
<https://www.utstat.toronto.edu/~rsalakhu/papers/dbm.pdf>

Sargur N. Srihari . Deep Boltzmann Machines
<https://cedar.buffalo.edu/~srihari/CSE676/20.4-DeepBoltzmann.pdf?fbclid=IwAR08ldg0QmyR-c3XT97k28SiiC8SC1XGdJNQpGCLsS4kOTUC1it4WjV4Yww>