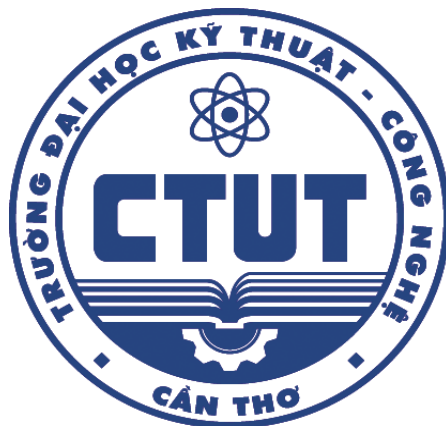


TRƯỜNG ĐẠI HỌC KỸ THUẬT – CÔNG NGHỆ CẦN THƠ
KHOA CÔNG NGHỆ THÔNG TIN



ĐỒ ÁN HỌC PHẦN 2
NGHIÊN CỨU CÁC KỸ THUẬT PHÂN
ĐOẠN ẢNH VÀ TRÍCH XUẤT ĐẶC TRƯNG

GVHD: Th.S Hà Lê Ngọc Dung

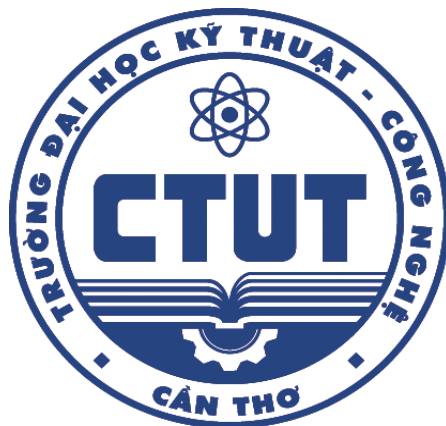
GVPB:.....

Sinh viên thực hiện:

- | | |
|----------------------------|----------------------|
| 1: Từ Thái Bảo | MSSV: 1900222 |
| 2: Lâm Thiện Nhân | MSSV: 1900558 |
| 3: Nguyễn Duy Khánh | MSSV: 1900540 |

Ngành: Khoa học máy tính - 2019

Cần Thơ, năm 2022



NGHIÊN CỨU CÁC KỸ THUẬT PHÂN ĐOẠN ẢNH VÀ TRÍCH XUẤT ĐẶC TRƯNG

GVPB:.....

1: Từ Thái Bảo **MSSV: 1900222**

2: Lâm Thiên Nhân **MSSV: 1900558**

3: Nguyễn Duy Khánh **MSSV: 1900540**

Ngành: Khoa học máy tính - 2019

Cần Thơ, năm 2022

NHẬN XÉT, ĐÁNH GIÁ CỦA GIẢNG VIÊN HƯỚNG DẪN

Đồ án 2

Đề tài: *NGHIÊN CỨU CÁC KỸ THUẬT PHÂN ĐOẠN ẢNH VÀ TRÍCH XUẤT ĐẶC TRƯNG*

Sinh viên thực hiện:

1. Từ Thái Bảo

MSSV: 1900222

2. Lâm Thiện Nhân

MSSV: 1900558

3. Nguyễn Duy Khánh

MSSV: 1900540

Ngành: Khoa học máy tính - 2019

Giảng viên hướng dẫn: Th.S Hà Lê Ngọc Dung

Nhận xét, đánh giá:

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

....., ngày ... tháng ... năm 2022

Giảng viên hướng dẫn

NHẬN XÉT, ĐÁNH GIÁ CỦA GIẢNG VIÊN PHẢN BIỆN

Đồ án 2

Đề tài: *NGHIÊN CỨU CÁC KỸ THUẬT PHÂN ĐOẠN ẢNH VÀ TRÍCH XUẤT ĐẶC TRƯNG*

Sinh viên thực hiện:

- | | |
|---------------------|---------------|
| 1. Từ Thái Bảo | MSSV: 1900222 |
| 2. Lâm Thiện Nhân | MSSV: 1900558 |
| 3. Nguyễn Duy Khánh | MSSV: 1900540 |

Ngành: Khoa học máy tính - 2019

Giảng viên hướng dẫn: Th.S Hà Lê Ngọc Dung

Nhận xét, đánh giá:

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

....., ngày ... tháng ... năm 2022

Giảng viên phản biện

LỜI CAM ĐOAN

Nhóm chúng em xin cam đoan rằng đồ án này là công trình nghiên cứu của nhóm thực hiện dựa trên những kiến thức đã được học, nghiên cứu và tìm hiểu một số đề tài và các phương án đi trước, không sao chép từ bất cứ công trình đã có trước đó. Mọi thứ được dựa trên sự cố gắng cũng như sự nỗ lực của bản thân.

Những phần có sử dụng tài liệu tham khảo có trong đồ án đã được liệt kê và nêu rõ ra tại phần tài liệu tham khảo.

Cần Thơ, ngày 20 tháng 6 năm 2022

Nhóm sinh viên thực hiện

Sinh viên thực hiện 1



Từ Thái Bảo

Sinh viên thực hiện 2



Lâm Thiện Nhân

Sinh viên thực hiện 3



Nguyễn Duy Khánh

LỜI CẢM ƠN

Được sự giúp đỡ nhiệt tình của quý thầy cô bộ môn Khoa học máy tính, Khoa Công nghệ thông tin, Trường Đại học Kỹ thuật - Công nghệ Cần Thơ, những người đã dìu dắt chúng em tận tình, đã truyền đạt cho em những kiến thức và những bài học quý giá trong suốt thời gian chúng em theo học tại trường. Sau gần ba tháng nghiên cứu chúng em đã hoàn thành đề tài: ***“NGHIÊN CỨU CÁC KỸ THUẬT PHÂN ĐOẠN ẢNH VÀ TRÍCH XUẤT ĐẶC TRƯNG”***.

Chúng em chân thành gửi lời cảm ơn đến tất cả các thầy cô trong khoa Công Nghệ Thông Tin đã hỗ trợ chúng em hoàn thành tốt đồ án lần này. Đặc biệt nhất là chúng em muốn dành lời cảm ơn đến cô Hà Lê Ngọc Dung, người đã tận tâm hướng dẫn và trực tiếp giúp đỡ chúng em trong suốt quá trình nghiên cứu đồ án. Với sự chỉ bảo nhiệt tình của cô, chúng em đã có định hướng tốt trong việc triển khai và thực hiện các yêu cầu trong quá trình làm đồ án.

Để có thể hoàn thành tốt được đề tài, ngoài này sự nỗ lực học hỏi của bản thân, còn có sự hướng dẫn tận tình từ quý thầy cô và sự giúp đỡ tận tình từ gia đình và bạn bè đã tạo cho chúng em những điều kiện tốt nhất trong suốt quá trình học tập và nghiên cứu.

Xin chân thành cảm ơn!

TÓM TẮT ĐỒ ÁN

Đây là đồ án phục vụ cho việc nghiên cứu của sinh viên tìm hiểu về các kỹ thuật phân đoạn và trích xuất đặc trưng của hình ảnh. Với tham vọng tìm hiểu, học hỏi những điểm mới mẽ của các kỹ thuật này nên chúng em mạnh dạn đăng ký đề tài này mong rằng sẽ một phần nào đó giúp cho mọi người có thể nắm bắt và tìm hiểu được các kỹ thuật của phân đoạn và trích xuất đặc trưng của hình ảnh. Ngoài ra, Đồ án còn tìm hiểu sơ lược và trình bày về bệnh viêm phổi và xây dựng một hệ thống nhận dạng bệnh viêm phổi.

Bố cục của đề tài: *Nghiên cứu các kỹ thuật phân đoạn ảnh và trích xuất đặc trưng* bao gồm 6 phần:

- Chương I: Tổng quan về Xử lý hình ảnh
- Chương II: Cơ sở lý thuyết
- Chương III: Tìm hiểu về bệnh viêm phổi Pneumonia
- Chương IV: Xây dựng hệ thống nhận dạng bệnh viêm phổi
- Chương V: Kết luận
- Chương VI: Tài liệu tham khảo

MỤC LỤC

NHẬN XÉT, ĐÁNH GIÁ CỦA GIẢNG VIÊN HƯỚNG DẪN	i
NHẬN XÉT, ĐÁNH GIÁ CỦA GIẢNG VIÊN PHẢN BIỆN	ii
LỜI CAM ĐOAN	i
LỜI CẢM ƠN	ii
TÓM TẮT ĐỒ ÁN	iii
MỤC LỤC	iv
DANH MỤC BẢNG BIỂU, HÌNH ẢNH	vi
DANH MỤC TỪ VIẾT TẮT, THUẬT NGỮ	viii
CHƯƠNG I: TỔNG QUAN VỀ XỬ LÝ ẢNH VÀ THỊ GIÁC MÁY TÍNH	1
1.1. Lý do chọn đề tài	1
1.2. Xử lý ảnh	2
1.3. Thị giác máy tính	2
CHƯƠNG II: CƠ SỞ LÝ THUYẾT	4
2.1. Các phương pháp phân đoạn ảnh	4
2.1.1. Giới thiệu	4
2.1.2. Phân loại dựa trên ngưỡng cố định	4
2.1.3. Phân loại dựa trên lược đồ (Histogram)	4
2.1.4. Phân đoạn dựa theo đường biên	7
2.2. Các phương pháp trích xuất đặc trưng	10
2.2.1. Phương pháp trích xuất đặc trưng SIFT	11
2.2.2. Phương pháp trích xuất đặc trưng HoG	17
2.2.3. Phương pháp trích chọn đặc trưng LBP	23
2.3. Tìm hiểu về Mạng thần kinh tích chập CNN	29
2.3.1. Giới thiệu	29
2.3.2. Kiến trúc mạng CNN	30
2.3.3. Xây dựng mô hình CNN	33
2.3.4. Cách chọn tham số cho CNN	37
CHƯƠNG III: TÌM HIỂU VỀ BỆNH VIÊM PHỔI PNEUMONIA	38
3.1. Giới thiệu	38
3.2. Các nguyên nhân gây nên bệnh viêm phổi	38
3.2.1. Phân loại dựa trên nguyên nhân gây viêm	38
3.2.2. Phân loại dựa trên nguyên nhân lây nhiễm bệnh	39

3.3. Các triệu chứng của bệnh viêm phổi.....	40
3.4. Các phương pháp chuẩn đoán bệnh viêm phổi	40
3.5. Các phương pháp điều trị bệnh viêm phổi	41
3.6. Các phương pháp phòng bệnh.....	41
CHƯƠNG IV: XÂY DỰNG HỆ THỐNG NHẬN DẠNG BỆNH VIÊM PHỔI	43
4.1. Giới thiệu	43
4.2. Thực nghiệm	43
4.2.1. Cài đặt các thư viện cần thiết.....	43
4.2.2. Thu thập bộ dữ liệu.....	44
4.2.3. Tiền xử lý dữ liệu	45
4.2.4. Tăng cường dữ liệu.....	46
4.2.5. Cấu hình cho mạng CNN	47
4.2.6. Phân tích mô hình	48
CHƯƠNG V: KẾT LUẬN.....	50
5.1. Kết quả đạt được	50
5.2. Hạn chế	50
5.3. Hướng phát triển	50
CHƯƠNG VI: TÀI LIỆU THAM KHẢO	51

DANH MỤC BẢNG BIỂU, HÌNH ẢNH

Bảng 2.1: Bảng thể hiện biến thể LBP, LBP đồng dạng	25
Bảng 2.2: Bảng thể hiện ví dụ về tính toán đặc trưng LBP	29
Bảng 2.3: Bảng mô tả phương pháp kích hoạt ReLU	32
Bảng 2.4: Max Pooling và Average Pooling kích thước 2x2.....	32
Bảng 4.1: Bảng mô tả cấu hình cho tăng cường dữ liệu.....	46
Hình 2. 1: Hình ảnh mô tả thuật toán đối xứng nền	6
Hình 2. 2: Hình ảnh mô tả thuật toán tam giác.....	7
Hình 2. 3: Hình ảnh mô tả Bimodal Histogram.....	7
Hình 2. 4: Hình ảnh mô tả các loại đường biên.....	8
Hình 2. 5: Hình ảnh mô tả việc tính toán các ảnh DoG từ các ảnh kẻ mờ	12
Hình 2. 6: Hình ảnh mô tả quá trình tìm điểm cực trị trong các hàm sai khác DoG.....	13
Hình 2. 7: Hình ảnh mô tả quá trình lựa chọn các điểm hấp dẫn	15
Hình 2. 8: Tính độ lớn và hướng của Gradient.....	16
Hình 2. 9: Hình ảnh mô tả mạng lưu lược đồ định hướng	17
Hình 2. 10: Hình ảnh mô tả biến thể R-HoG và C-HoG	18
Hình 2. 11: Hình ảnh mô tả quá trình tiền xử lý ảnh.....	18
Hình 2. 12: Hình ảnh mô tả gradient theo hướng X (Trái), Y (Giữ), XY (Phải)	19
Hình 2. 13: Hình ảnh mô tả quá trình chia khối trích đặc trưng HOG	20
Hình 2. 14: Hình ảnh mô tả quá trình chia ảnh thành các ô để tính các biểu đồ.....	21
Hình 2. 15: Hình ảnh mô tả quá trình tính biểu đồ mức xám.....	21
Hình 2. 16: Hình ảnh mô tả biểu đồ kết quả của tính lược đồ mức xám.....	22
Hình 2. 17: Hình ảnh mô tả về trích chọn đặc trưng bởi HoG	23
Hình 2. 18: Hình ảnh mô tả tập hợp các điểm xung quanh Ptt.....	24
Hình 2. 19: Hình ảnh mô tả các mẫu LBP đồng dạng	25
Hình 2. 20: Hình ảnh mô tả về các trường hợp của LBP sau khi quay với góc 15 độ	27
Hình 2. 21: Phương pháp LBP không phụ thuộc vào độ sáng cho cùng một đối tượng giống nhau	29
Hình 2. 22: Hình ảnh mô tả kiến trúc mạng CNN.....	30
Hình 2. 23: Hình ảnh mô tả phương pháp tích chập.....	31
Hình 2. 24: Hình ảnh mô tả lớp kết nối đầy đủ	33
Hình 2. 25: Vị trí bắt đầu của trường tiếp nhận cục bộ	34
Hình 2. 26: Trường tiếp nhận cục bộ với ba bản đồ đặc trưng.....	35
Hình 2. 27: Ví dụ Max Pooling 2x2.....	36

Hình 2. 28: Max pooling với ba bản đồ đặc trưng.....	36
Hình 3.1 : Ảnh phổi bình thường (phải) và viêm phổi (trái)	39
Hình 4.1: Ảnh phổi bình thường (phải) và ảnh viêm phổi (trái)trong tập dữ liệu.....	43

DANH MỤC TỪ VIẾT TẮT, THUẬT NGỮ

STT	Từ viết tắt	Tên đầy đủ	Dịch nghĩa
1	SIFT	Scale-invariant feature transform	Quy mô biến đổi tính năng bất biến
2	HoG	Histogram of Oriented Gradients	Biểu đồ các đường dốc
3	LBP	Local binary patterns	Mẫu nhị phân địa phương
4	CNN	Convolutional Neural Network	Mạng thần kinh tích chập

CHƯƠNG I: TỔNG QUAN VỀ XỬ LÝ ẢNH VÀ THỊ GIÁC MÁY TÍNH

1.1. Lý do chọn đề tài

Trong bối cảnh ứng dụng công nghệ thông tin ngày phát triển, ứng dụng vào phân loại hiện nay đang phát triển rất mạnh ở rất nhiều lĩnh vực như: học thuật, kinh doanh, bảo mật, y tế... và các ở các đối tượng như: nhà nghiên cứu xã hội, chính phủ và các tổ chức phi lợi nhuận khác dữ liệu có vai trò quan trọng. Trong đó lượng thông tin mà con người thu thập dưới dạng hình ảnh rất lớn. Các thông tin hình ảnh được tổng hợp phân tích xử lý bằng các kỹ thuật phân đoạn ảnh và trích xuất đặc trưng trong các lĩnh vực khác nhau. Các bài toán nhận dạng và phân loại ảnh là một trong những bài toán quan trọng trong lĩnh vực thị giác máy tính do có nhiều ứng dụng như phát hiện đối tượng, phân lớp tự động, tổ chức quản lý và truy xuất ảnh-video, gán chú thích và mô tả nội dung ảnh-video, phát hiện và nhận dạng bệnh trong lĩnh vực y tế... Vì các tổ chức này sở hữu một lượng lớn dữ liệu không có cấu trúc và việc xử lý dữ liệu sẽ trở nên dễ dàng hơn rất nhiều nếu như các dữ liệu này được chuẩn hóa bởi các chủ đề/nhãn. Nền tảng công nghệ để thực hiện bài toán phân loại chính là trí tuệ nhân tạo (Artificial Intelligence) và học sâu (Deep Learning).

Trong lĩnh vực Thị giác máy tính (Computer Vision), nhờ những thành tựu của lĩnh vực học sâu mà trong những năm gần đây, ta đã chứng kiến được nhiều thành tựu vượt bậc. Các hệ thống xử lý ảnh lớn như Facebook, Google hay Amazon đã đưa vào sản phẩm của mình những chức năng thông minh như nhận diện khuôn mặt người dùng, phát triển xe hơi tự lái hay drone giao hàng tự động.

Từ lâu các nhà khoa học đã nhận thấy những ưu điểm của bộ óc con người và tìm cách bắt chước để thực hiện trên những máy tính, tạo cho nó có khả năng học tập, nhận dạng và phân loại. Vì vậy các nhà khoa học đã nghiên cứu và sáng tạo ra mạng nơron nhân tạo. Nó thực sự được chú ý và nhanh chóng trở thành một hướng nghiên cứu mới triển vọng đặc biệt là lĩnh vực nhận dạng, dự đoán và phân loại.

Trong lĩnh vực Y tế, việc phát hiện và chuẩn đoán bệnh một cách chính xác sẽ giúp tiết kiệm được thời gian cho các y bác sĩ cũng như đem lại hiệu quả cao trong việc chữa bệnh. Ứng dụng của xử lý ảnh và thị giác máy tính đã giúp giải quyết những vấn đề đó bằng việc xử lý ảnh để phát hiện và nhận dạng khối u, cải thiện ảnh X quang, nhận dạng đường biên mạch máu từ những ảnh chụp bằng tia X,...

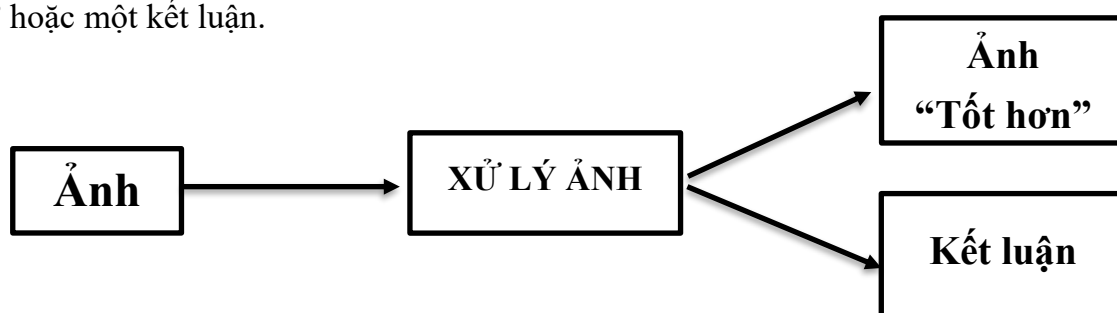
Với sự hấp dẫn của đề tài và những thách thức còn đang ở phía trước, với niềm đam mê công nghệ hiện đại và những ứng dụng thực tế tuyệt vời của nó, với khát khao khám phá và chinh phục những chi thức mới mẻ, Chúng em đã chọn đề tài nghiên cứu: ***“Nghiên cứu các kỹ thuật phân đoạn ảnh và trích xuất đặc trưng”*** trong đồ án lần này của mình. Mục tiêu chính của đồ án 2 này là tìm hiểu về nghiên cứu các kỹ thuật phân

đoạn ảnh và trích xuất đặc trưng, tìm hiểu về bệnh viêm phổi và xây dựng một hệ thống nhận dạng bệnh viêm phổi dựa trên mạng thần kinh tích chập CNN.

1.2. Xử lý ảnh

Con người thu nhận thông tin qua các giác quan, trong đó thị giác đóng vai trò quan trọng nhất. Những năm trở lại đây với sự phát triển của phần cứng máy tính, xử lý ảnh và đồ hoạ đã phát triển một cách mạnh mẽ và có nhiều ứng dụng trong cuộc sống. Xử lý ảnh và đồ hoạ đóng một vai trò quan trọng trong tương tác người máy.

Quá trình xử lý ảnh được xem như là quá trình thao tác ảnh đầu vào nhằm cho ra kết quả mong muốn. Kết quả đầu ra của một quá trình xử lý ảnh có thể là một ảnh “tốt hơn” hoặc một kết luận.



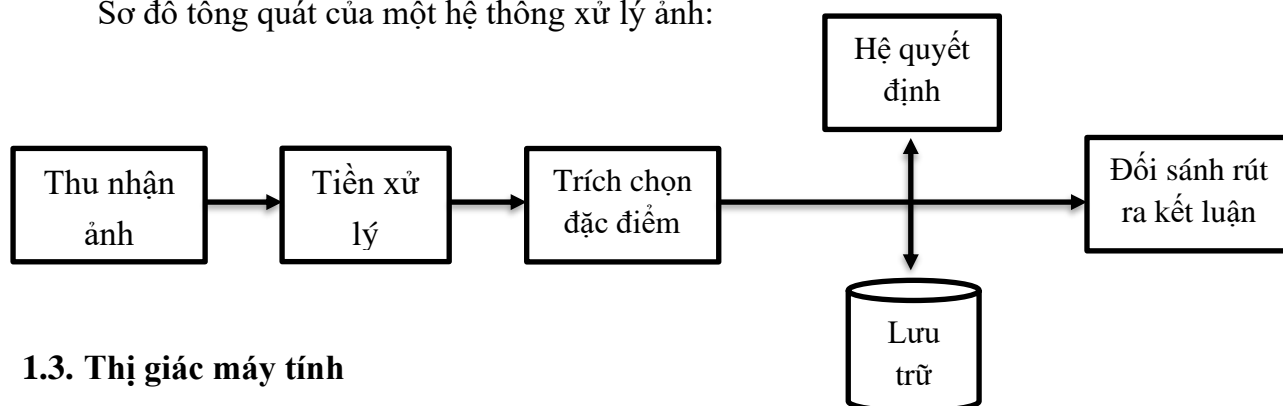
Xử lý ảnh là đối tượng nghiên cứu của lĩnh vực thị giác máy, là quá trình biến đổi từ một ảnh ban đầu sang một ảnh mới với các đặc tính và tuân theo ý muốn của người sử dụng. Xử lý ảnh có thể gồm quá trình phân tích, phân lớp các đối tượng, làm tăng chất lượng, phân đoạn và tách cạnh, gán nhãn cho vùng hay quá trình biên dịch các thông tin hình ảnh của ảnh.

“Thị giác máy tính” và “xử lý ảnh” thường là các cụm từ đi chung với nhau. Các công việc của chúng cũng tương đồng với nhau, chúng cùng xử lý một đối tượng là ảnh để cho ra các bộ dữ liệu phù hợp với nhu cầu phân tích của các cá nhân hay tổ chức nhằm phục vụ cho nhu cầu phát triển kinh tế hay phát triển một mô hình dự đoán, đánh giá.

Các lĩnh vực xử lý ảnh điển hình như: trích chọn đặc trưng, nhận dạng, nén ảnh,...

Xử lý ảnh bao gồm tập hợp của một số bước như sau: thu thập và tiền xử lý ảnh, biểu diễn dữ liệu và nhận dạng, đưa ra quyết định.

Sơ đồ tổng quát của một hệ thống xử lý ảnh:



1.3. Thị giác máy tính

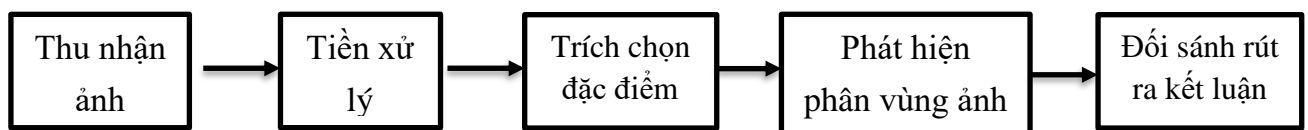
Thị giác máy tính (Computer Vision) được định nghĩa là một lĩnh vực bao gồm các phương pháp thu nhận, xử lý ảnh kỹ thuật số, phân tích và nhận dạng các hình ảnh và nói chung là dữ liệu đa chiều từ thế giới thực để cho ra các thông tin số hoặc biểu tượng. Thị giác máy tính cũng được mô tả là sự tổng thể của một dải rộng các quá trình tự động và tích hợp và các thể hiện cho các nhận thức thị giác

Thị giác máy là một lĩnh vực đã và đang phát triển mạnh mẽ trên thế giới hiện nay. Kể từ những năm 70 của thế kỷ 20 khi mà khả năng tính toán của các máy tính trở nên mạnh mẽ, nó có thể giải quyết được những công việc cần tới tốc độ cao như xử lý các tập ảnh hay các đoạn video thì thị giác máy được nhắc đến, nghiên cứu và phát triển cho đến ngày nay.

Nhìn chung là dữ liệu đa chiều từ thế giới thực để cho ra các thông tin số hoặc biểu tượng. Vì vậy dễ dàng thấy được một điều là rất nhiều lĩnh vực khác nhau đều liên quan đến Computer Vision vì nó có thể thay thế cho cặp mắt của con người để làm việc trong thời gian dài thậm chí hiệu suất có thể vượt trội hơn so với con người.

Ngoài ra nó còn có thể trích xuất được rất nhiều dữ liệu để dùng cho việc nghiên cứu và đánh giá một vấn đề nào đó theo một cách trực quan và rất thực tế.

Sơ đồ tổng quát của một hệ thống thị giác máy tính:



CHƯƠNG II: CƠ SỞ LÝ THUYẾT

2.1. Các phương pháp phân đoạn ảnh

2.1.1. Giới thiệu

Biên độ của các thuộc tính vật lý của ảnh (độ phản xạ, độ truyền sang, màu sắc,...) là một đặc tính đơn giản và rất hữu ích. Nếu biên độ đủ lớn đặc trưng cho phân đoạn ảnh. Biên độ trong bộ cảm biến hồng ngoại có thể phản ánh vùng có nhiệt độ thấp hay nhiệt độ cao.

Việc chọn ngưỡng trong kỹ thuật này là bước vô cùng quan trọng, thông thường người ta tiến hành theo các bước chung như sau:

- Xem xét lược đồ xám của ảnh để xác định đỉnh và khe, nếu ảnh có nhiều đỉnh và khe thì các khe có thể sử dụng để chọn ngưỡng.
- Chọn ngưỡng T sao cho xác định trước η của toàn bộ số mẫu thấp hơn T .
- Điều chỉnh ngưỡng dựa trên xét lược đồ của điểm lân cận
- Chọn ngưỡng bằng cách xem xét lược đồ xám của những điểm tiêu chuẩn đã chọn.

Một thuật toán đơn giản trong kỹ thuật này là :giả sử rằng mình chúng ta đang quan tâm tới các đối tượng sáng (object) trên nền tối (background), một tham số T gọi là ngưỡng độ sáng sẽ được chọn 1 ảnh $f[x,y]$ theo cách:

$$\begin{array}{ll} \text{If } f[x,y] \geq T & f[x,y] = \text{object} = 1 \\ \text{Else} & f[x,y] = \text{Background} = 0 \end{array}$$

Ngược lại, với các đối tượng tối trên nền sáng chúng ta có thuật toán sau:

$$\begin{array}{ll} \text{If } f[x,y] < T & f[x,y] = \text{object} = 1 \\ \text{Else} & f[x,y] = \text{Background} = 0 \end{array}$$

Có rất nhiều thuật toán chọn ngưỡng : ngưỡng cố định, dựa trên lược đồ, sử dụng Entropy, sử dụng tập mờ, chọn ngưỡng thông qua sự không ổn định của lớp và tính thuần nhất của vùng,...

2.1.2. Phân loại dựa trên ngưỡng cố định

Đây là một phương pháp chọn ngưỡng độc lập với dữ liệu ảnh. Chương trình ứng dụng sẽ làm việc với các ảnh có độ tương phản khá cao, trong đó các đối tượng quan tâm rất tối còn nền gần như là đồng nhất và rất sáng thì việc chọn ngưỡng $T=128$ (xét trên thang độ sáng từ 0 tới 255) là một giá trị chọn khá chính xác. Chính xác ở đây hiểu theo nghĩa là số các điểm ảnh bị phân lớp sai là cực tiểu.

2.1.3. Phân loại dựa trên lược đồ (Histogram)

Trong hầu hết các trường hợp, ngưỡng được chọn từ lược đồ sáng của vùng hay ảnh cần phân đoạn. Có rất nhiều kỹ thuật chọn ngưỡng tự động xuất phát từ lược đồ xám $\{h[b] \mid b=0,1,2 \dots 2^B - 1\}$ đã được đưa ra. Nhưng kỹ thuật này có thể tận dụng những lợi

thể do sự làm trơn dữ liệu lược đồ ban đầu mang lại nhằm loại bỏ những giao động nhỏ về độ sáng. Tuy nhiên các thuật toán làm trơn cần phải cẩn thận, không được làm dịch chuyển các vị trí đỉnh của lược đồ. Nhận xét này dẫn tới thuật toán làm trơn dưới đây :

$$h_{smooth}[b] = \frac{1}{W} \sum_{w=-(W-1)/2}^{(W-1)/2} h_{raw}[b-w]$$

Trong đó:

$W / 2 \neq 1$ và thường được chọn là 3 hoặc 5

b có giá trị từ 0 đến 255

2.1.3.1. Thuật toán đẳng hiệu – Isodata

Đây là kỹ thuật chọn ngưỡng theo kiểu lặp do Ridler và Calvard đưa ra.

Thuật toán được mô tả như sau:

- Bước 1 : Chọn giá trị ngưỡng khởi động $\theta_0 = 2^{B-1}$
- Bước 2 : Tính các trung bình mẫu ($m_{f,0}$) của những điểm ảnh thuộc đối tượng và ($m_{b,0}$) của những điểm ảnh nền.
- Bước 3 : Tính ngưỡng trung gian theo công thức :

$$\theta_k = \frac{m_{f,k-1} + m_{b,k-1}}{2} \text{ với } k = 1, 2, \dots$$

- Bước 4 : nếu $\theta_k = \theta_{k-1}$: kết thúc và dùng thuật toán. Nếu không bằng nhau gán $\theta_k = \theta_{k-1}$ ta được một ngưỡng mới, quay về bước 1

Ưu điểm và nhược điểm của thuật toán đẳng hiệu

Ưu điểm: Đây là phương pháp:

- Có khả năng tổ chức.
- Mềm dẻo trong xử lý khử bỏ những cụm có kích thước quá nhỏ
- Có khả năng tách biệt được các cụm có tính chất hoàn toàn khác nhau
- Có khả năng kết hợp những cụm gần giống nhau thành 1 cụm

Nhược điểm:

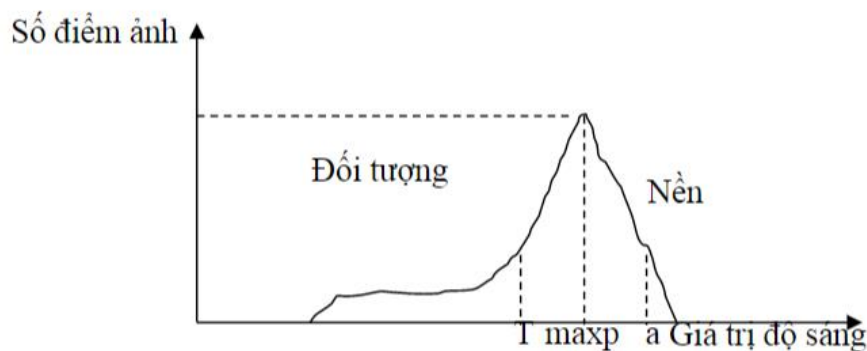
- Quá nhiều tham số cần phải cung cấp bởi người dùng, mặc dù chúng không phải là đại lượng cần phải biết.
- Các cụm được là hình cầu được xác định bởi hàm khoảng cách
- Giá trị K phụ thuộc vào những tham số do người sử dụng qui định và nó cũng không phải là những giá trị tốt nhất
- Cụm trung bình thường không phải là mẫu tốt nhất cho một cụm.

2.1.3.2. Thuật toán đối xứng nền

Kỹ thuật này dựa trên sự giả định là tồn tại hai đỉnh phân biệt trong lược đồ nằm đối xứng qua đỉnh có giá trị lớn nhất trong phần lược đồ thuộc về các điểm ảnh nền. Kỹ thuật này có thể tận dụng ưu điểm của việc làm trơn được mô tả trong phương trình.

Đỉnh cực đại $\max p$ tìm được nhờ tiến hành tìm giá trị cực đại trong lược đồ. Sau đó thuật toán sẽ áp dụng ở phía *không phải là điểm ảnh thuộc đối tượng ứng* với giá trị cực đại đó nhằm tìm ra giá trị độ sáng a ứng với giá trị phần trăm $p\%$ mà: $P(a) = p\%$, trong đó $P(a)$ là hàm phân phối xác suất về độ sáng được định nghĩa như sau:

Hàm phân phối xác suất $P(a)$ thể hiện xác suất chọn được một giá trị độ sáng từ một vùng ánh sáng cho trước, sao cho giá trị này không vượt qua một giá trị này cho trước, sao cho giá trị này không vượt quá giá trị sáng cho trước a . Khi a biến thiên từ $-\infty$ đến $+\infty$, $P(a)$ sẽ nhận các giá trị từ 0 đến 1, $P(a)$ là hàm đơn điệu không giảm theo a , do vậy $dP/da \geq 0$.



Hình 2. 1: Hình ảnh mô tả thuật toán đối xứng nền

Giả thiết ở đây là ảnh có các đối tượng tối trên nền sáng. Giả sử mức là 5% thì có nghĩa là ta phải ở bên phải đỉnh $\max p$ một giá trị a sao cho $P(a) = 95\%$. Do tính đối xứng đã giả định ở trên, chúng ta sử dụng độ dịch chuyển về phía trái của điểm cực đại tìm giá trị ngưỡng T :

$$T = \max p - (a - \max p)$$

Kỹ thuật này dễ dàng điều chỉnh được cho phù hợp với tình huống ảnh có các đối tượng sáng trên một nền tối.

2.1.3.3. Thuật toán tam giác

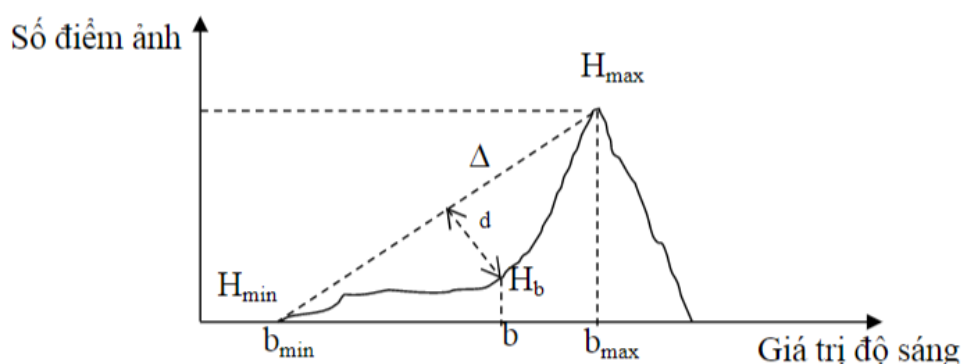
Khi một ảnh có các điểm ảnh thuộc đối tượng tạo nên một đỉnh yếu trong lược đồ ảnh thì thuật toán tam giác hoạt động rất hiệu quả. Thuật toán này do Zack đề xuất và được mô tả như sau:

Bước 1: Xây dựng đường thẳng Δ là đường nối 2 điểm là (H_{\max}, b_{\max}) và (H_{\min}, b_{\min}) trong đó H_{\max} là điểm Histogram ứng với độ sáng nhỏ nhất b_{\min} .

Bước 2: Tính khoảng cách d từ H_b của lược đồ (ứng với điểm sáng b) đến Δ Trong đó, $b \in [b_{\max}, b_{\min}]$.

Bước 3: Chọn ngưỡng $T = \max \{H_b\}$

Minh họa thuật toán tam giác bởi hình vẽ như sau :



Hình 2. 2: Hình ảnh mô tả thuật toán tam giác

Ngưỡng T được chọn ở tại vị trí cực tiểu địa phương của Histogram nằm giữa hai đỉnh của Histogram. Điểm cực đại địa phương của Histogram có thể dễ dàng được phát hiện bằng cách sử dụng biến đổi chóp mũ (Top Hat). Thuộc vào tình huống chúng ta đang phải làm việc là với nhưng đối tượng sáng trên nền tối hay đối tượng tối trên nền sáng mà phép biến đổi top hat sẽ có một trong hai dạng sau:

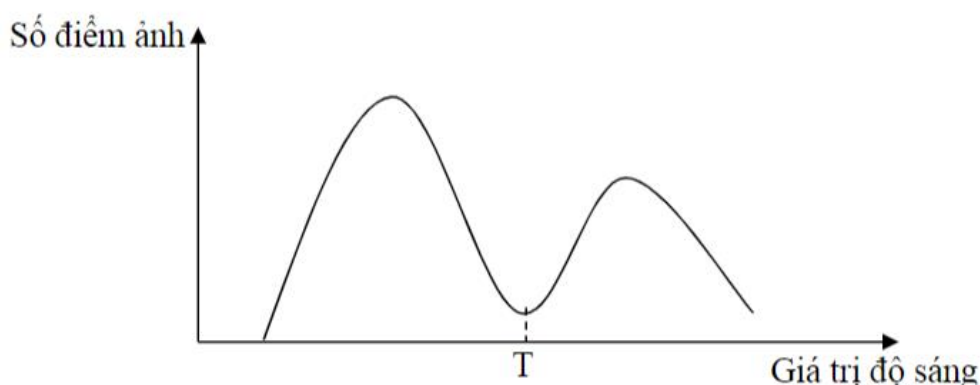
Các đối tượng sáng:

$$TopHat(A, B) = A - \max_B(\min_A))$$

Các đối tượng tối:

$$TopHat(A, B) = A - \min_B(\max_A))$$

Việc tính toán giá trị cực tiểu địa phương của Histogram thì khó nếu Histogram nhiều. Do đó, trong trường hợp này nên làm trơn Histogram.



Hình 2. 3: Hình ảnh mô tả Bimodal Histogram

Trong một số ứng dụng nhất định, cường độ của đối tượng hay nền thay đổi khá chậm. Trong trường hợp này, Histogram ảnh có thể không chứa hai thùy phân biệt rõ ràng, vì vậy có thể phải dùng ngưỡng thay đổi theo không gian. Hình ảnh được chia thành những khối hình vuông, Histogram và ngưỡng được tính cho mỗi khối tương ứng. Nếu Histogram cục bộ không phải là Bimodal Histogram thì ngưỡng được tính bằng cách nội suy ngưỡng của các khối láng giềng.

2.1.4. Phân đoạn dựa theo đường biên

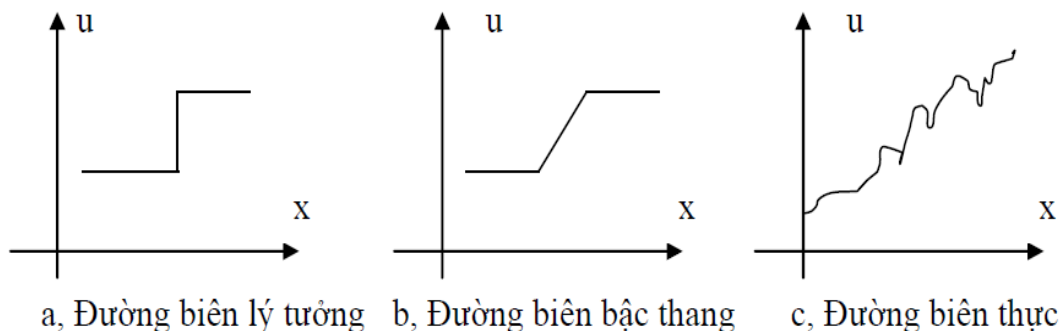
2.1.4.1. Giới thiệu

Biên là một đặc tính rất quan trọng để phân vùng các đối tượng. Một điểm ảnh được coi là điểm biên khi có sự thay đổi nhanh hoặc đột ngột về mức xám hoặc màu sắc. Chính vì vậy mà trong nhiều ứng dụng người ta sử dụng cách phân đoạn dựa theo biên.

Phát hiện biên trực tiếp: Phương pháp này làm nổi biên dựa vào sự biến thiên mức xám của ảnh. Kỹ thuật chủ yếu dùng để phát hiện biên ở đây là dựa vào sự biến đổi cấp xám theo hướng. Cách tiếp cận theo đạo hàm bậc nhất của ảnh dựa trên kỹ thuật Gradient, nếu lấy đạo hàm bậc hai của ảnh dựa trên biến đổi gia ta có kỹ thuật Laplace.

Phát hiện biên gián tiếp: Nếu bằng cách nào đó ta phân được ảnh thành các vùng thì ranh giới giữa các vùng đó gọi là biên. Kỹ thuật dò biên và phân vùng ảnh là hai bài toán đối ngẫu nhau vì dò biên để thực hiện phân lớp đối tượng mà khi đã phân lớp xong nghĩa là đã phân vùng được ảnh và ngược lại, khi đã phân vùng ảnh đã được phân lớp thành các đối tượng, do đó có thể phát hiện được biên.

Phương pháp phát hiện biên trực tiếp tỏ ra khá hiệu quả và ít chịu ảnh hưởng của nhiễu, song nếu sự biến thiên độ sáng không đột ngột, phương pháp tỏ ra kém hiệu quả, phương pháp phát hiện biên gián tiếp tuy khó cài đặt, song lại áp dụng khá tốt trong trường hợp này.



Hình 2. 4: Hình ảnh mô tả các loại đường biên

2.1.4.2. Các phương pháp phát hiện biên

Kỹ thuật phát hiện biên Gradient

Phương pháp Gradient là phương pháp dò biên cục bộ dựa vào cực đại của đạo hàm. Theo định nghĩa, Gradient là một vectơ có các thành phần biểu thị tốc độ thay đổi giá trị của điểm ảnh theo hai hướng x và y. Các thành phần của gradient được tính theo công thức:

$$\begin{cases} \frac{\partial f(x,y)}{\partial x} \approx \frac{f(x+dx,y)-f(x,y)}{dx} \\ \frac{\partial f(x,y)}{\partial y} \approx \frac{f(x,y+dy)-f(x,y)}{dy} \end{cases}$$

Trong đó:

dx, dy là khoảng cách (tính bằng số điểm) theo hướng x và y.

Tuy ta nói là lấy đạo hàm nhưng thực chất chỉ là mô phỏng và xấp xỉ đạo hàm bằng các kỹ thuật nhân chập vì ảnh số là tín hiệu rời rạc nên đạo hàm không tồn tại.

a) Kỹ thuật Prewitt

Đây là phương pháp lâu đời nhất, cổ điển nhất. Kỹ thuật Prewitt được mô tả sử dụng 2 mặt nạ nhân chập xấp xỉ đạo hàm theo 2 hướng x và y là:

$$H_x = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}$$

$$H_y = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

Các bước tính toán của kỹ thuật Prewitt:

Bước 1: Tính $I \otimes H_x$ và $I \otimes H_y$

Bước 2: Tính $I \otimes H_x + I \otimes H_y$

b) Kỹ thuật Sobel

Tương tự như kỹ thuật Prewitt, kỹ thuật Sobel sử dụng 2 mặt nạ nhân chập theo 2 hướng x, y là:

$$H_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

$$H_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

Các bước tính toán của kỹ thuật Sobel:

Bước 1: Tính $I \otimes H_x$ và $I \otimes H_y$

Bước 2: Tính $I \otimes H_x + I \otimes H_y$

c) Kỹ thuật la bàn

Kỹ thuật la bàn đo gradient theo 8 hướng ngược chiều kim đồng hồ cách nhau 45° và sử dụng 8 mặt nạ chập theo 8 hướng $0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ, 315^\circ$.

$$H_1 = \begin{bmatrix} -5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}$$

$$H_3 = \begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix}$$

$$H_5 = \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix}$$

$$H_7 = \begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix}$$

$$H_2 = \begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}$$

$$H_4 = \begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix}$$

$$H_6 = \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix}$$

$$H_8 = \begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix}$$

Các bước tính toán thuật toán La bàn:

Bước 1: Tính $I \otimes H_i$; $i = 1, 8$

Bước 2: $\sum_{i=1}^8 I \otimes H_i$

Kỹ thuật phát hiện biên Laplace

Các Phương pháp gradient chỉ thích hợp cho các vùng ảnh độ tương phản thay đổi có tính nhảy bậc, điều này gây khó khăn cho phát hiện các đường thẳng. Khi mức xám thay đổi chậm, miền chuyển tiếp trải rộng, phương pháp cho hiệu quả hơn đó là phương pháp sử dụng đạo hàm bậc hai Laplace. Phương pháp Laplacian cho phép xác định đường biên dựa vào giá trị 0 của đạo hàm bậc hai của ảnh. Laplacian của một ảnh tại điểm $I(x,y)$ được tính như sau:

$$L(x, y) = \frac{\partial^2 y}{\partial x^2} + \frac{\partial^2 y}{\partial y^2}$$

Laplacian được kết hợp với bộ lọc làm mịn ảnh để tìm biên. Xét công thức sau:

$$h(r) = -e^{-\frac{r^2}{2\sigma^2}}$$

Trong đó:

$$r^2 = x^2 + y^2$$

σ : là độ lệch chuẩn

Nếu thực hiện phép tích chập của hàm này với ảnh cần tìm biên, kết quả là ảnh sẽ bị mờ đi, mức độ mờ phụ thuộc vào giá trị của σ . Laplacian của h tức đạo hàm bậc hai của h theo r là:

$$\nabla^2 h(r) = -\left[\frac{r^2 - \sigma^2}{\sigma^4}\right] e^{-\frac{r^2}{2\sigma^2}}$$

Hàm này thường được gọi là Laplacian of a Gaussian (LoG) do có dạng Gaussian.

Trong phương pháp này, bộ lọc Gaussian được kết hợp với Laplacian cho phép hiển thị những vùng ảnh có cường độ thay đổi nhanh do đó làm tăng hiệu quả phát hiện biên. Nó cho phép làm việc với một diện tích rộng hơn xung quanh điểm ảnh đang được nghiên cứu nhằm phát hiện chính xác hơn vị trí của đường biên. Nhược điểm của phương pháp này là không xác định được hướng của biên do sử dụng ba bộ lọc Laplacian quá khác nhau có dạng như sau.

$$H_1 = \begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix} \quad H_2 = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \quad H_3 = \begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix}$$

2.2. Các phương pháp trích xuất đặc trưng

Mỗi đối tượng nói chung đều có những đặc trưng riêng, đặc trưng chính là yếu tố giúp phân biệt đối tượng này với đối tượng khác. Ví dụ đặc trưng của xe ô tô là có 4 bánh, có vô lăng,... còn đặc trưng của xe máy là có hai bánh; đặc trưng của mỗi con người thể hiện là chiều cao, cân nặng, màu tóc, màu da,... Tóm lại đặc trưng là các yếu

tổ xác định nên đối tượng. Một đối tượng chỉ được xác định khi có đủ số đặc trưng xác định nên nó.

Trích xuất đặc trưng là tìm ra điểm đặc trưng của đối tượng so với đối tượng khác tùy theo mục đích nhận dạng trong quá trình xử lý và nhận dạng ảnh ảnh. Là quá trình chọn lọc một tập con chứa các thuộc tính liên quan để sử dụng trong quá trình xây dựng mô hình. Các kỹ thuật trích chọn đặc trưng được dùng cho một số lý do:

- Đơn giản hóa các mô hình để giúp các nhà nghiên cứu diễn dịch dễ dàng hơn.
- Giảm thời gian huấn luyện của các mô hình.
- Tăng cường tổng quát hóa bằng cách giảm sự quá khớp.

2.2.1. Phương pháp trích xuất đặc trưng SIFT

2.2.1.1. Giới thiệu

SIFT là viết tắt của Scale-invariant feature transform. Là một thuật toán tiêu biểu và có hiệu quả khá cao dựa theo các đặc trưng cục bộ bất biến trong ảnh. SIFT là một feature descriptor được sử dụng trong lĩnh vực thị giác máy tính và xử lý hình ảnh được dùng để nhận dạng đối tượng, matching image, hay áp dụng cho các bài toán phân loại. Giải thuật lần đầu được giới thiệu bởi David Lowe vào năm 1999.

SIFT sẽ đưa ra các kết quả ổn định với những tỉ lệ của ảnh khác nhau, bên cạnh đó cũng có thể nói giải thuật này có tính rotation-invariant.

Để đối sánh và nhận dạng hai ảnh thì cần tìm tập đặc trưng giống nhau trong hai ảnh, dựa vào hướng và tỉ lệ để có thể biết đối tượng trong ảnh gốc đã xoay, thu phóng bao nhiêu so với ảnh đem đối sánh. Cách tiếp cận của thuật toán này dựa vào điểm bất biến cục bộ của ảnh, chúng được trích xuất ra, được định hướng và mô tả sao cho hai đặc trưng ở hai vùng khác nhau thì khác nhau. Số lượng các điểm đặc trưng có một tầm quan trọng trong vấn đề nhận dạng đối tượng, để nhận dạng một đối tượng nhỏ trong một ảnh chứa tập hợp các đối tượng hỗn độn thì cần ít nhất 3 điểm đặc trưng giống nhau để phát hiện và bóc tách đối tượng. Đối với vấn đề thực hiện nhận dạng đối tượng bất kì thì ban đầu thường sử dụng SIFT để tạo một hệ dữ liệu các đặc trưng được trích xuất từ dữ liệu ảnh gốc. Sau đó với mỗi ảnh đối tượng đem nhận dạng ta dùng giải thuật SIFT trích xuất tập đặc trưng từ ảnh và đem đối sánh với hệ dữ liệu đặc trưng để tìm ra tập đặc trưng giống nhau, từ đó nhận dạng đối tượng trong cơ sở dữ liệu ảnh ban đầu.

2.2.1.2. Các bước trích chọn đặc trưng SIFT

Phương pháp trích rút các đặc trưng bất biến SIFT được tiếp cận theo phương pháp lọc kim tự tháp, theo đó phương pháp được thực hiện lần lượt theo các bước sau:

Bước 1: Phát hiện các điểm cực trị

Bước đầu tiên sẽ tìm các điểm tiềm năng có thể trở thành điểm đặc trưng bằng phương pháp lọc kim tự tháp dựa vào việc thay đổi tham số bộ lọc Gaussian, cần dò tìm các vị trí và các độ đo mà chúng bất biến trong các khung nhìn khác nhau của cùng một đối tượng. Các vị trí đó bất biến về độ đo có thể được dò tìm bằng cách tìm kiếm các đặc trưng ổn định trên toàn bộ các độ đo có thể, sử dụng một hàm liên tục về số đo vốn rất nổi tiếng có tên là hàm độ đo không gian Witkin 1983.

Theo các công bố của Koenderink (1984) và Lindeberg (1994) thì hàm Gaussian là hàm tốt nhất để biểu diễn độ đo không gian của ảnh 2 chiều. Vì vậy, độ đo không gian của một ảnh sẽ được định nghĩa như công thức sau:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

Trong đó:

$L(x, y, \sigma)$: là hàm không gian tỷ lệ của ảnh I

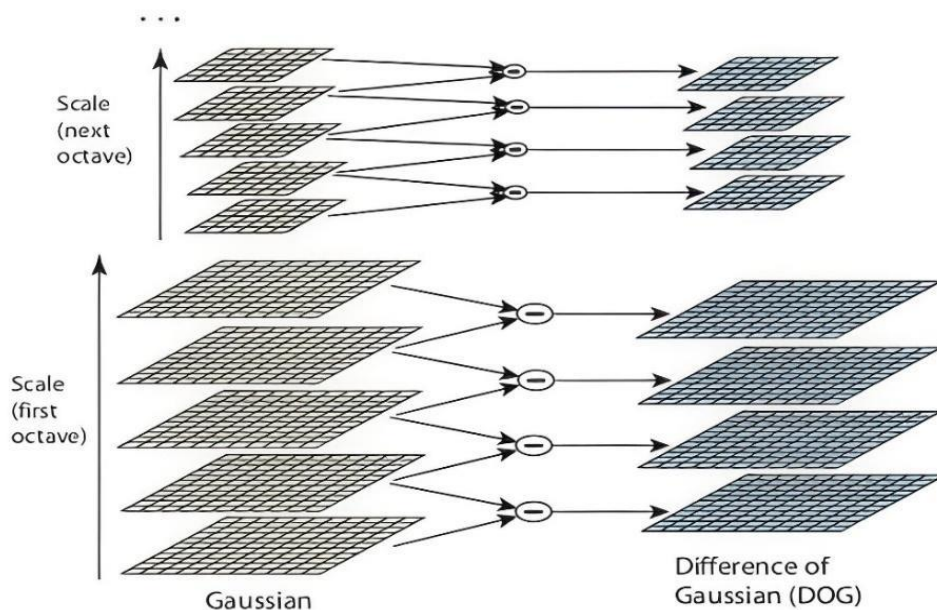
$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$ là hàm Gaussian

$I(x, y)$: là ảnh đầu vào

Để tìm những điểm đặc trưng có tính bất biến cao, thuật toán được sử dụng là tìm cực trị cục bộ của đạo hàm của hàm Gaussian kí hiệu là $D(x, y, \sigma)$. Hàm này được tính toán từ sự sai khác giữa 2 độ đo không gian cạnh nhau của một ảnh với tham số đo lệch nhau một hằng số k.

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y)$$

Việc lựa chọn hàm Gaussian là vì nó là kỹ thuật rất hiệu quả để tính toán L cũng như làm tăng độ mịn của ảnh, mà L thì luôn phải được tính rất nhiều để mô tả đặc trưng trong không gian đo, và sau đó D sẽ được tính một cách đơn giản chỉ với phép trừ ma trận điểm ảnh với chi phí thực hiện thấp.



Hình 2. 5: Hình ảnh mô việc tính toán các ảnh DoG từ các ảnh kề mờ

Hơn nữa, đạo hàm của hàm Gaussian (DoG) có thể được sử dụng để tạo ra một sự xấp xỉ có kích thước chuẩn của hàm Gaussian ($\sigma^2 \nabla^2 G$). Việc chuẩn hóa đạo hàm bậc hai với hệ số σ^2 là cần thiết cho bất biến đo trở nên đúng. Các giá trị cực đại và cực tiểu của $\sigma^2 \nabla^2 G$ chính là những giá trị có tính ổn định nhất so với một loạt các hàm đánh giá khác như : Gradient, Hessian hay Harris.

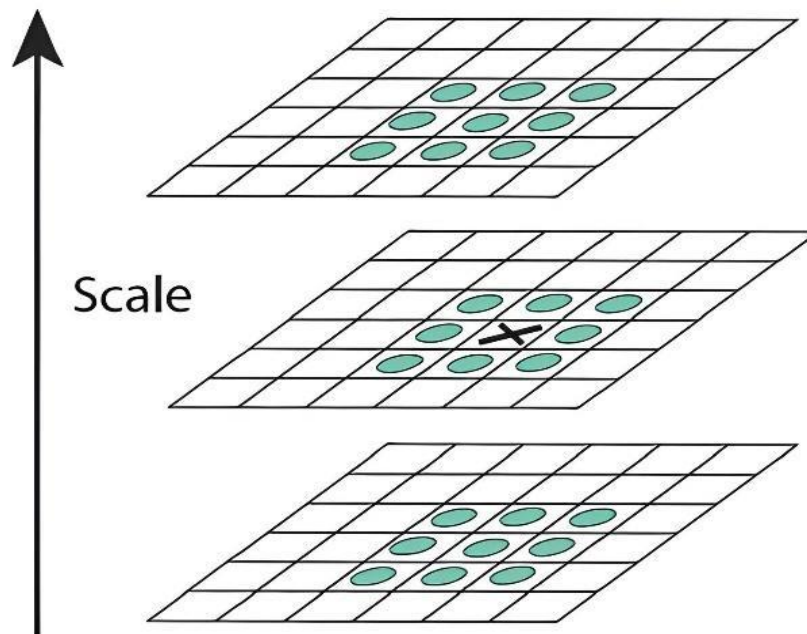
Như vậy, $\nabla^2 G$ có thể được tính thông qua việc xấp xỉ đạo hàm riêng tại các tham số đo gần nhau $k\sigma$ và σ :

$$\sigma^2 \nabla^2 G \approx \frac{G(x,y,k\sigma) - G(x,y,\sigma)}{k\sigma - \sigma}$$

Như vậy, bước đầu tiên của giải thuật SIFT là phát hiện các điểm hấp dẫn với bộ lọc Gaussian ở các tỉ lệ khác nhau và các ảnh DoG từ sự khác nhau của các ảnh kề mờ.

Hai hình ảnh liên tục trong một nhóm tám được chọn để tính toán. Sau đó, cặp đôi tiếp theo được thực hiện, và lặp đi lặp lại quá trình. Điều này được thực hiện cho tất cả các nhóm tám. Các hình ảnh kết quả là một xấp xỉ bất biến tỷ lệ của Laplacian of Gaussian.

Tìm các cực trị trong các lớp kết quả theo từng miền cục bộ. Cụ thể là tại mỗi điểm trên các lớp kết quả sẽ được so sánh với 8 điểm lân cận trên cùng lớp và 9 điểm lân cận trên mỗi lớp khác.



Hình 2. 6: Hình ảnh mô tả quá trình tìm điểm cực trị trong các hàm sai khác DoG

(X là điểm hiện tại, các vòng tròn màu xanh là các láng giềng của nó)

X sẽ được đánh dấu là điểm hấp dẫn nếu nó là cực đại hoặc cực tiểu của tất cả 26 láng giềng. Thông thường, một vị trí không cực đại hoặc không cực tiểu sẽ không phải đi qua tất cả 26 kiểm tra. Một vài kiểm tra ban đầu thường là đủ để loại bỏ nó.

Để tăng sự hiệu quả khi dò tìm các điểm cực trị. Các tần số này sẽ được xác định thông qua phương pháp thử nghiệm. Sau khi thử nghiệm với nhiều nguồn dữ liệu ảnh khác nhau tần số lấy mẫu trong không gian đo tốt nhất là 3, và tần số lấy mẫu $\sigma = 1.6$.

Bước 2: Định vị các điểm đặc trưng

Bước tiếp theo sẽ loại bỏ các điểm có độ tương phản hoặc tính đặc trưng cục bộ ít hơn các điểm khác hoặc có xu hướng là đường biên đối tượng. Bước thực hiện này gồm 3 giai đoạn :

a) Phép nội suy lân cận cho vị trí đúng của điểm tiềm năng:

Phép nội suy lân cận sử dụng mở rộng Taylor cho hàm Difference of Gaussian $D(x,y,\sigma)$:

$$D(X) = D + \frac{\partial D^t}{\partial X} X + \frac{1}{2} X \frac{\partial^2 D}{\partial^2 X} X$$

Trong đó:

D và đạo hàm của nó được tính tại một điểm tiềm năng.

$X = (x,y,\sigma)$: là khoảng cách từ điểm đó.

\hat{X} : lấy đạo hàm của hàm trên với đối số X và tiến dần đến 0:

Vùng chứa điểm hấp dẫn được xác định qua \hat{X} :

$$\hat{X} = - \frac{\partial^2 D^{-1}}{\partial X^2} \frac{\partial D}{\partial X}$$

Nếu $\hat{X} > 0.5$ theo một chiều nào đó thì nó có chỉ số cực trị không gần với các điểm tiềm năng khác, nó sẽ bị thay đổi và phép nội suy sẽ thay thế vai trò của nó bằng điểm khác gần nó.

b) Loại trừ các điểm có tính tương phản kém

Những điểm có \hat{X} thỏa mãn (< 0.5) được thêm vào tập hợp mẫu tốt nhất, tiếp tục phân tích tiếp.

Dùng \hat{X} để loại những điểm cực trị không ổn định.

$$D(\hat{X}) = D + \frac{1}{2} \frac{\partial D^t}{\partial X} \hat{X}$$

Nếu $D(\hat{X})$ nhỏ hơn 0.03 thì sẽ bị loại bỏ.

c) Loại bỏ các điểm dư thừa theo biên

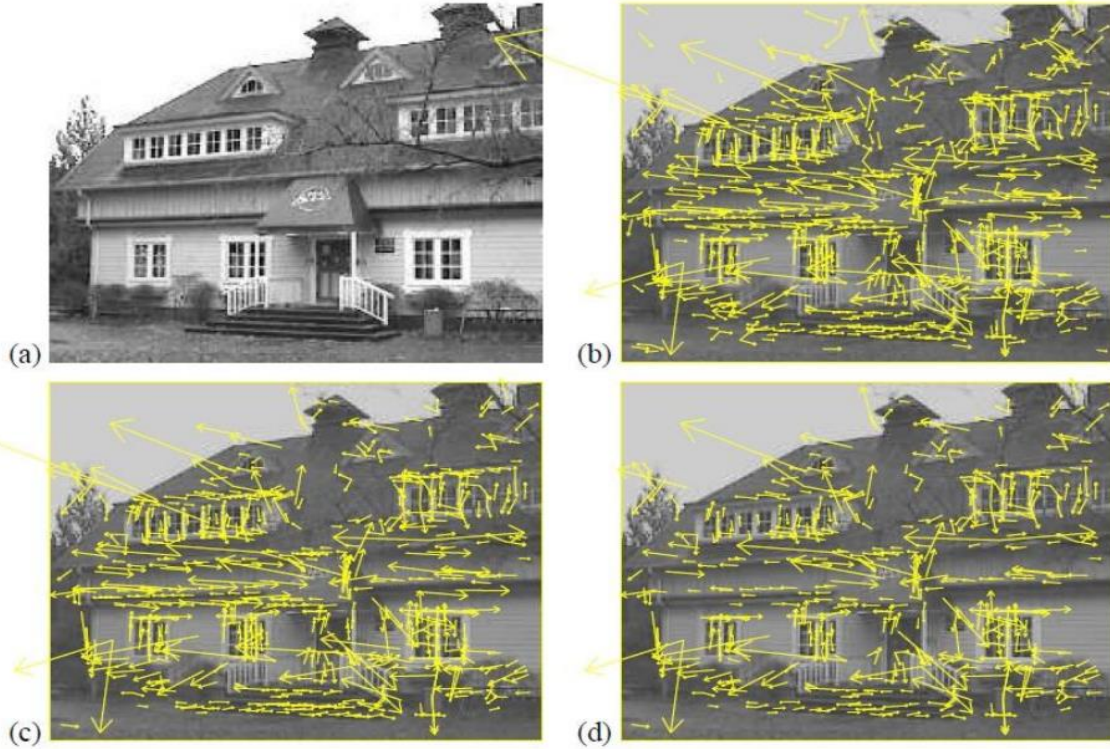
Sử dụng hàm DoG sẽ cho tác động mạnh đến biên khi vị trí của biên là khó xác định và vì vậy các điểm tiềm năng trên biên sẽ không bất biến và bị nhiễu. Và để tăng sự ổn định cho các điểm sẽ được chọn làm điểm đặc biệt ta sẽ loại trừ các điểm tiềm năng khó định vị.

Sau khi áp dụng hàm DoG sẽ làm đường biên ảnh không rõ ràng và độ cong chính sẽ có giá trị lớn hơn nhiều so với độ cong dọc theo biên vì vậy cần loại bỏ bớt các điểm

đặc biệt dọc theo cùng một biên. Giải pháp cho việc này là sử dụng giá trị của ma trận Hessian cấp 2 :

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}$$

Các giá trị riêng của H tỉ lệ thuận với độ cong của D, các phần tử của H là D_{xx} và D_{yy} .



Hình 2. 7: Hình ảnh mô tả quá trình lựa chọn các điểm hấp dẫn

Trong bước này số điểm hấp dẫn giảm, điều này giúp tăng hiệu quả và sự mạnh mẽ của thuật toán. Điểm hấp dẫn bị loại bỏ nếu nó có độ tương phản thấp hoặc nếu nó nằm trên một cạnh.

Bước 3: Xác định hướng cho các điểm đặc trưng

Xác định một định hướng cho mỗi điểm hấp dẫn bằng cách chỉ định một hướng phù hợp cho mỗi điểm hấp dẫn dựa trên các thuộc tính hình ảnh cục bộ, dựa vào hướng của điểm hấp dẫn này ta có thể biết được điểm hấp dẫn bất biến với phép quay ảnh.

Sau khi thử nghiệm với một số phương pháp tiếp cận để xác định hướng, phương pháp sau đây đã được tìm thấy cho kết quả ổn định nhất. Tại mỗi điểm hấp dẫn người ta tính toán biểu đồ hướng Gradient trong vùng láng giềng của điểm hấp dẫn. Độ lớn và hướng của các điểm hấp dẫn được xác định lần lượt theo công thức 2.22 và 2.23:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + ((L(x, y+1) - L(x, y-1)))^2}$$

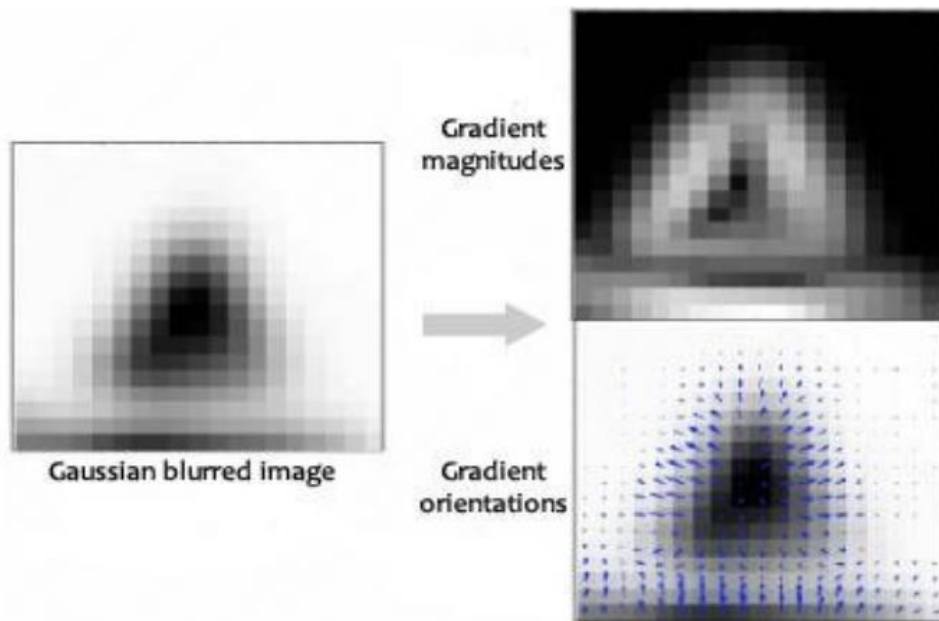
$$\theta(x, y) = \tan^{-1} \left(\frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \right)$$

Trong đó:

$m(x, y)$: là độ lớn của vector định hướng

$\theta(x, y)$: là hướng của vector định hướng

Một biểu đồ hướng được hình thành từ định hướng gradient của các điểm lấy mẫu trong một khu vực xung quanh điểm hấp dẫn. Đỉnh trong biểu đồ hướng tương ứng với hướng chủ đạo của gradient. Đỉnh cao nhất trong biểu đồ được phát hiện, và sau đó bất kỳ điểm nào khác có cao điểm là 80% so với đỉnh cao nhất cũng được sử dụng cũng tạo ra một điểm hấp dẫn với định hướng đó. Đối với các địa điểm có nhiều đỉnh cường độ tương tự sẽ có nhiều điểm hấp dẫn tạo ra tại cùng một vị trí và tỷ lệ, nhưng có hướng khác nhau.

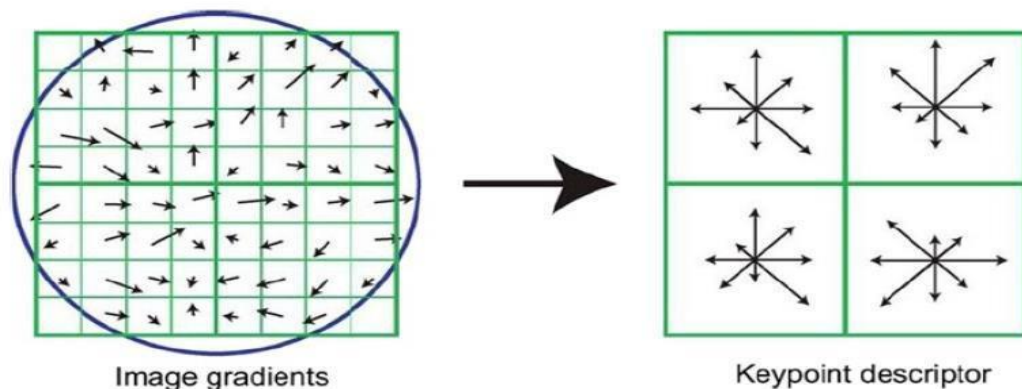


Hình 2. 8: Tính độ lớn và hướng của Gradient

Bước 4: Mô tả các điểm đặc trưng

Bước tiếp theo đây sẽ tính toán một bộ mô tả cho một vùng ảnh cục bộ mà có tính đặc trưng cao. Một cách tiếp cận đơn giản đó là lấy mẫu mật độ ảnh cục bộ lân cận điểm đặc trưng ở một độ đo thích hợp, và đối sánh các mật độ này sử dụng độ đo tương quan chuẩn. Tuy nhiên, hệ số tương quan đơn giản thì lại rất nhạy cảm với sự thay đổi mà gây ra sự đăng ký nhầm các mẫu, chẳng hạn như các biến đổi Affine, phối cảnh 3 chiều, hoặc bóp méo mềm.

Cách tiếp cận này dựa trên một mô hình thị giác sinh học, cụ thể là mô hình noron phức tạp trong hệ thống não bộ. Các noron sẽ tương ứng với một gradient tại một hướng và tần số không gian cụ thể, nhưng vị trí của gradient trên võng mạc được phép trượt trên một phạm vi nhỏ của khung nhìn.



Hình 2. 9: Hình ảnh mô tả mạng lưu lược đồ định hướng

Điểm hấp dẫn sau khi được xác định hướng sẽ được biểu diễn dưới dạng các vector $4 \times 4 \times 8 = 128$ bằng cách tổng hợp các vector định hướng của các điểm trong khu vực, các vector này có đặc điểm :

- Chung gốc.
- Độ dài mỗi vector tương ứng độ lớn gradient m của nó

2.2.2. Phương pháp trích xuất đặc trưng HoG

2.2.2.1. Giới thiệu

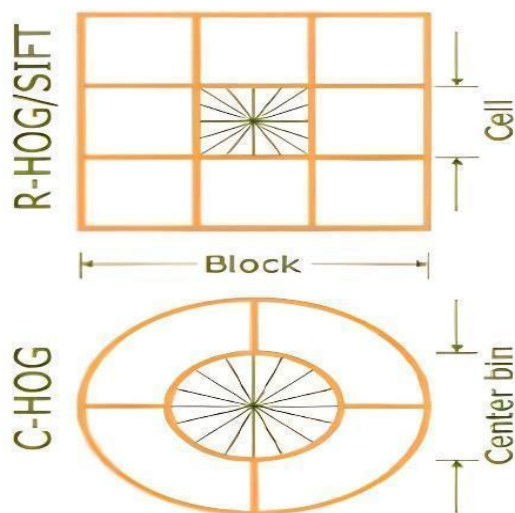
HoG là viết tắt của Histogram of Gradient là đặc trưng được dùng nhiều trong lĩnh vực phát hiện đối tượng HoG là một loại “feature descriptor”. Mục đích của HoG là trừu tượng hóa đối tượng bằng cách trích xuất ra những đặc trưng của đối tượng đó và bỏ đi những thông tin không hữu ích. Vì vậy, HoG được sử dụng chủ yếu để mô tả hình dạng và sự xuất hiện của một đối tượng trong ảnh.

Ý tưởng chính trong đặc trưng HoG xuất phát từ hình dạng và trạng thái của vật có thể được đặc trưng bằng sự phân bố về cường độ và hướng của cạnh. sử dụng thông tin về sự phân bố của các cường độ gradient hoặc của hướng biên để mô tả các đối tượng cục bộ trong ảnh. sử dụng thông tin về sự phân bố của các cường độ gradient hoặc của hướng biên để mô tả các đối tượng cục bộ trong ảnh.

HoG được cài đặt bằng cách chia nhỏ một bức ảnh thành các vùng con, được gọi là “tế bào” và với mỗi cell, ta sẽ tính toán một Histogram về các hướng của gradients cho các điểm nằm trong cell. Ghép các Histogram lại với nhau ta sẽ có một biểu diễn cho bức ảnh ban đầu. Để tăng cường hiệu năng nhận dạng, các Histogram cục bộ có thể được chuẩn hóa về độ tương phản bằng cách tính một ngưỡng cường độ trong một vùng lớn hơn cell, gọi là các khối (blocks) và sử dụng giá trị ngưỡng đó để chuẩn hóa tất cả các cell trong khối.

Kết quả sau bước chuẩn hóa sẽ là một vector đặc trưng có tính bất biến cao hơn đối với các thay đổi về điều kiện ánh sáng. Đặc trưng HoG được tính trên cả một vùng. Vì vậy, để có được đặc trưng của toàn bộ cửa sổ ta phải kết hợp nhiều vùng liên tiếp lại với nhau.

Đặc trưng HoG có một vài biến thể thường gặp như: R-HoG, R2-HoG, C-HoG. Các đặc trưng này khác nhau ở các phân bố và hình dạng của các ô như trong hình:



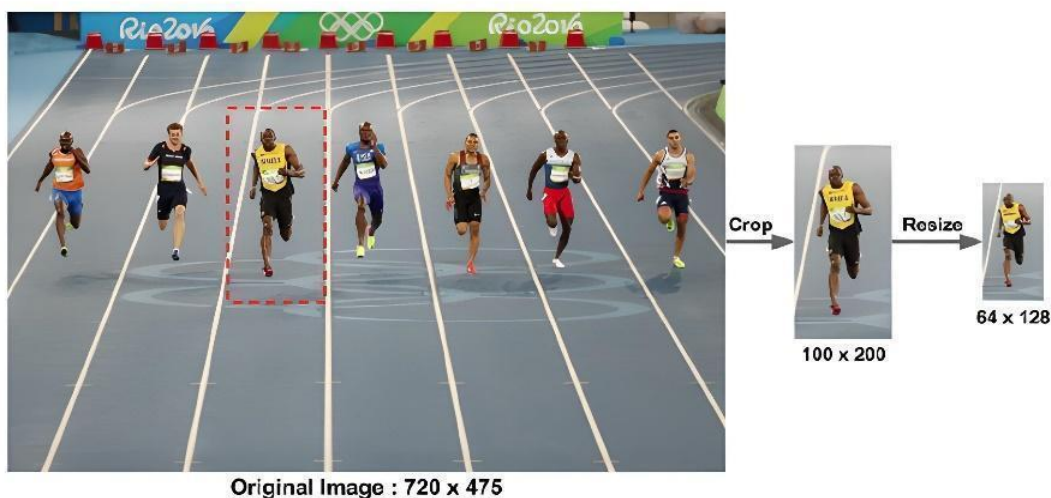
Hình 2. 10: Hình ảnh mô tả biến thể R-HoG và C-HoG

Từ một ảnh cho trước, qua việc rút trích đặc trưng trên ảnh đó cho ta một vector đại diện cho đối tượng đó. Việc rút ra được các vector đặc trưng như vậy giúp chuyển không gian ảnh sang không gian véc tơ để có thể xử lý và tính toán đơn giản hơn trên máy tính.

2.2.2.2. Các bước trích chọn đặc trưng HoG trên ảnh

Bước 1: Tiền xử lý

Trong hầu hết các thuật toán xử lý ảnh, bước đầu tiên là tiền xử lý dữ liệu ảnh. Để thuận tiện cho việc chia đều hình ảnh thành các khối, ô và tính toán đặc trưng ở các bước tiếp theo, cần thực hiện resize kích thước tất cả các hình ảnh trong dữ liệu về một kích thước chung.



Hình 2. 11: Hình ảnh mô tả quá trình tiền xử lý ảnh

Bước 2: Tính toán Gradient

Đây là bước trong phương pháp HoG và thường được thực hiện bằng hai phép nhân chập ảnh gốc với 2 nhân và 1 chiều tương ứng với các toán tử lấy đạo hàm ảnh theo hai hướng Ox và Oy, cụ thể là

Đây là bước được thực hiện bằng hai phép nhân chập ảnh gốc với 2 chiều, tương ứng với các toán tử lấy đạo hàm theo hai hướng Ox và Oy. Trong đó, 2 hướng tương ứng đó là:

$$D_x = [-1 \ 0 \ 1]$$

$$D_y = [1 \ 0 \ -1]^T$$

Trong đó:

T là phép chuyển vị của ma trận

Với một ảnh đầu vào là I , thì sẽ có 2 ảnh đạo hàm riêng theo 2 hướng được tính theo công thức:

$$I_x = I * D_x$$

$$I_y = I * D_y$$

Khi đó, có thể tính được Gradient bao gồm hai thành phần là cường độ(Gradient Magnitude) và hướng biến thiên(Gradient Derection) theo công thức lần lượt như sau:

$$G = \sqrt{I_x^2 + I_y^2}$$

$$\theta = \arctan \left(\frac{I_x}{I_y} \right)$$



Hình 2. 12: Hình ảnh mô tả gradient theo hướng X (Trái), Y (Giữa), XY (Phải)

Đối với hình ảnh màu, gradient của ba kênh(red, green và blue) được đánh giá. Độ lớn của gradient tại một điểm ảnh là giá trị lớn nhất của cường độ gradient của ba kênh, và góc là góc tương ứng với gradient tối đa.

Bước 3: Tính vector đặc trưng cho từng ô

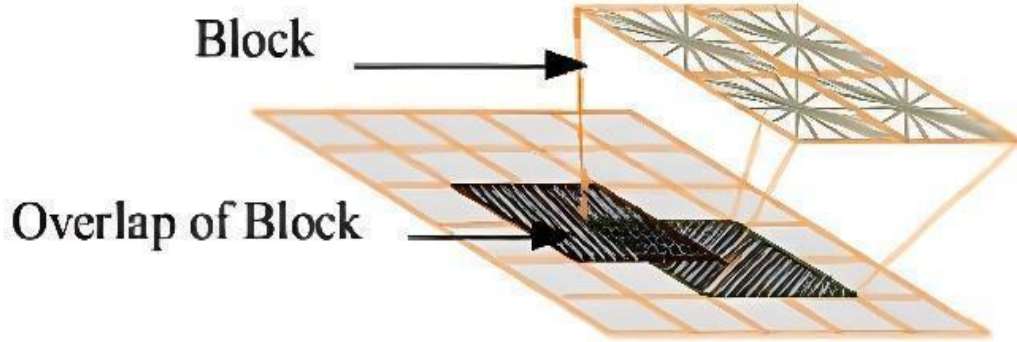
Để tính toán vector đặc trưng cho từng ô (cell). Chia hình ảnh đầu ra thành nhiều khối (block), mỗi khối lại chia đều thành các ô bằng nhau, mỗi ô có số pixel bằng nhau. Các khối được xếp chồng lên nhau. Để xác định được số khối sử dụng công thức sau:

$$n_{block/image} = \left(\frac{W_{image} - W_{block} - W_{cell}}{W_{cell}} + 1 \right) * \left(\frac{H_{image} - H_{block} - H_{cell}}{H_{cell}} + 1 \right)$$

Trong đó:

$W_{image} - W_{block} - W_{cell}$ lần lượt là chiều rộng của ảnh, khối và ô.

$H_{image} - H_{block} - H_{cell}$ lần lượt là chiều dài của ảnh, khối và ô.



Hình 2. 13: Hình ảnh mô tả quá trình chia khối trích đặc trưng HOG

Bước 4: Tính vector đặc trưng cho từng khối

Chia không gian hướng thành p bin (số chiều vector đặc trưng của ô).

Việc lưu trữ chính xác từng giá trị góc (orientation) của từng vị trí (x,y) không mang lại nhiều kết quả, do vậy sẽ chia không gian góc ra thành các bin. Việc phân chia bin càng nhỏ sẽ càng làm tăng độ chính xác, thực nghiệm cho thấy kích thước bin khoảng 20° là cho kết quả tốt nhất. Hướng Gradient sẽ chạy trong khoảng 0 độ đến 180 độ, trung bình 20° mỗi bin chia thành 9 bin như sau:

0 - 20; 21 - 40; 41 - 60; 61- 80; 81 - 100; 101 - 120; 121- 140; 141 -160; 161-180.

Rời rạc hóa góc hướng nghiêng tại mỗi điểm ảnh vào trong các bin.

Góc hướng nghiêng tại pixel (x,y) có độ lớn $\alpha(x,y)$ được rời rạc hóa vào một trong p bin.

Trường hợp rời rạc hóa unsigned-HOG với p=9:

$$B_{(x,y)} = \text{round} \left(\frac{p * \alpha(x,y)}{\pi} \right) \text{mod } p$$

Trường hợp rời rạc hóa signed-HOG với p=18:

$$B_{(x,y)} = \text{round} \left(\frac{p * \alpha(x,y)}{2\pi} \right) \text{mod } p$$

Giá trị bin được định lượng bởi tổng cường độ biến thiên của các pixel thuộc về bin đó. Để tính toán vector đặc trưng cho khối ta sẽ nối các vector đặc trưng của từng

cell trong block lại với nhau để được vector đặc trưng của block này. Nối các vector đặc trưng ô để được vector đặc trưng khối. Số chiều vector đặc trưng của khối tính theo công thức:

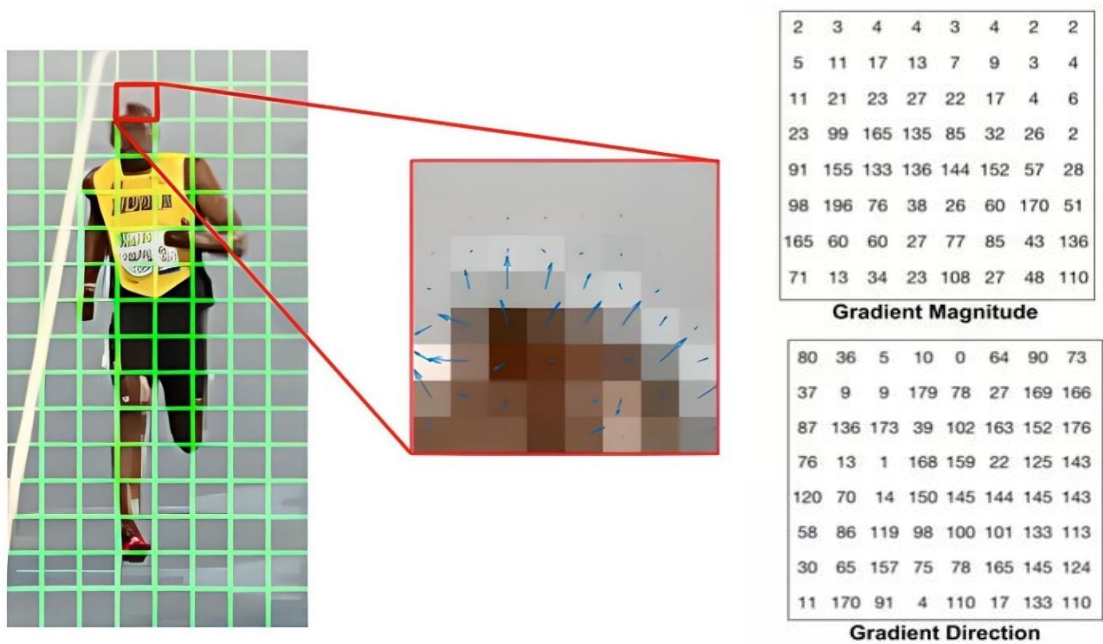
$$size_{block} = n_{cell} * size_{cell}$$

Trong đó:

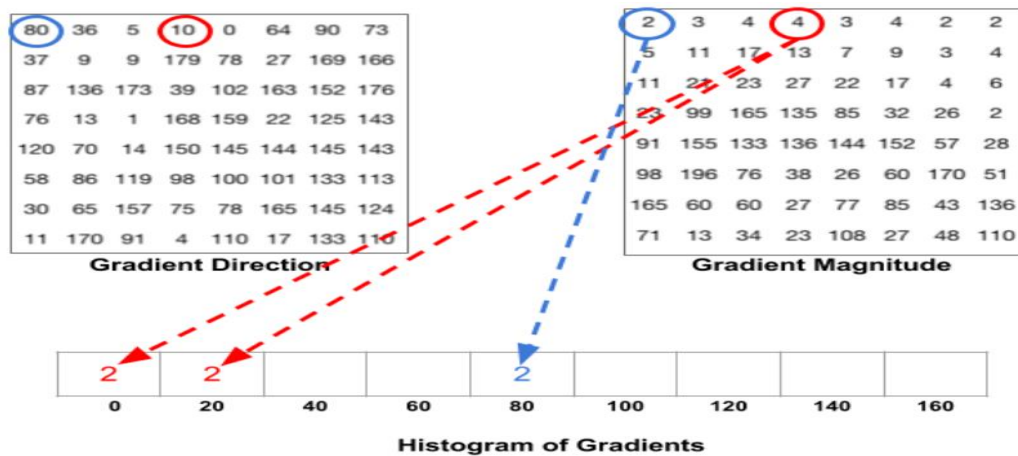
n_{cell} là số ô trong khối

$size_{cell}$ là số chiều của vector đặc trưng của ô. $size_{cell} = 9$ khi (unsigned-HoG) hoặc $size_{cell} = 18$ khi (signed-HoG).

Trong trường hợp này, hình ảnh của có kích thước là 64x128, ta sẽ chia mỗi hình ảnh thành các block có kích thước 16x16. Mỗi block sẽ bao gồm 4 cell, mỗi cell có kích thước là 8x8.



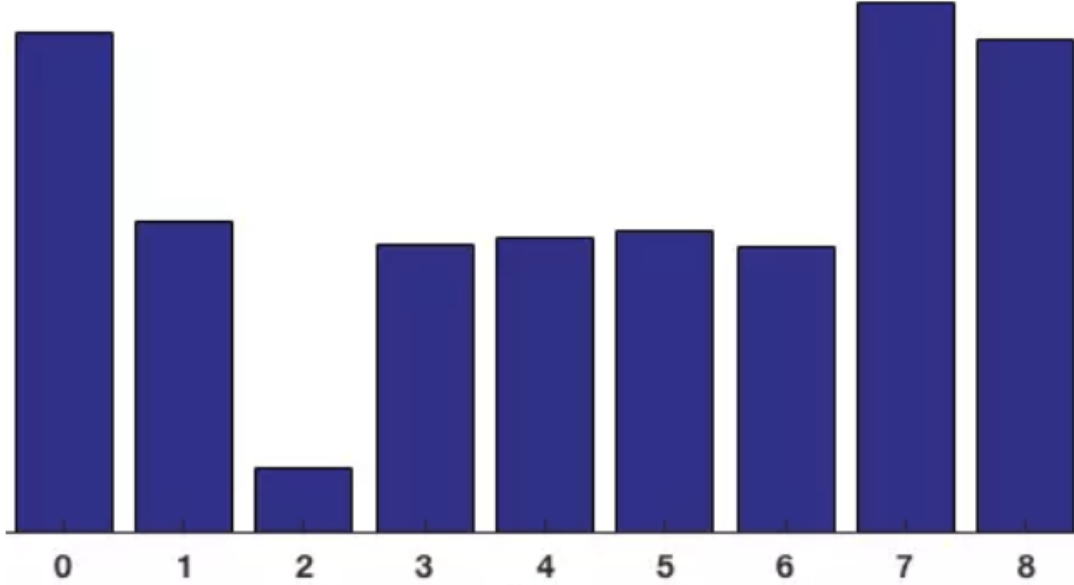
Hình 2. 14: Hình ảnh mô tả quá trình chia ảnh thành các ô để tính các biểu đồ



Hình 2. 15: Hình ảnh mô tả quá trình tính biểu đồ mức xám

Quá trình tính lược đồ mức xám được minh họa trong hình 2.15. Để tính được giá trị cho các ô trong bảng mức xám ta sẽ căn cứ vào bảng hướng gradient và giá trị tương

ứng. Chẳng hạn bin 80 sẽ chứa số 2 khi đối chiếu hai ô tương ứng của các bảng trên. Ta thêm 2 vào bin thứ 5 (hướng 80 độ) Tiếp theo là pixel có bao quanh màu đỏ. Nó có hướng 10 độ và cường độ 4. Vì không có bin 10 độ, nên ta vote cho bin 0 độ và 20 độ, mỗi bin thêm 2 đơn vị. Sau khi vote hết các pixel trong một cell kích thước 8x8 vào 9 bin, ta có thể thu được kết quả như sau:



Hình 2. 16: Hình ảnh mô tả biểu đồ kết quả của tính lược đồ mức xám

Bước 5: Tính vector đặc trưng cho từng ảnh.

Trước khi thực hiện tính véc tơ đặc trưng các khối trước khi tính véc tơ đặc trưng cho ảnh. Để chuẩn hóa véc tơ đặc trưng cho các khối ta có thể thực hiện bằng một trong các công thức như sau:

$$\text{L2-norm: } f = \frac{v}{\sqrt{v_2^2 + e^2}}$$

$$\text{L1-norm: } f = \frac{v}{\sqrt{v_1 + e}}$$

$$\text{L1-sqrt: } f = \sqrt{\frac{v}{v_1 + e}}$$

Trong đó:

v : là vectơ đặc trưng ban đầu của khối.

v_k : là k- norm của v ($k = 1, 2$).

e : là hằng số nhỏ nhất.

Để tính toán véc tơ đặc trưng cho ảnh, ta chỉ cần ghép các véc tơ đặc trưng của từng block lại với nhau để được véc tơ đặc trưng của cả window. Lưu ý rằng các block không đặt tách biệt nhau mà gối lên nhau, cách nhau một khoảng bằng kích thước của cell theo mỗi chiều.

Ghép các vector đặc trưng khối tạo nên ảnh để được đặc trưng R-HOG cho ảnh. Số chiều vector đặc trưng của ảnh tính theo công thức:

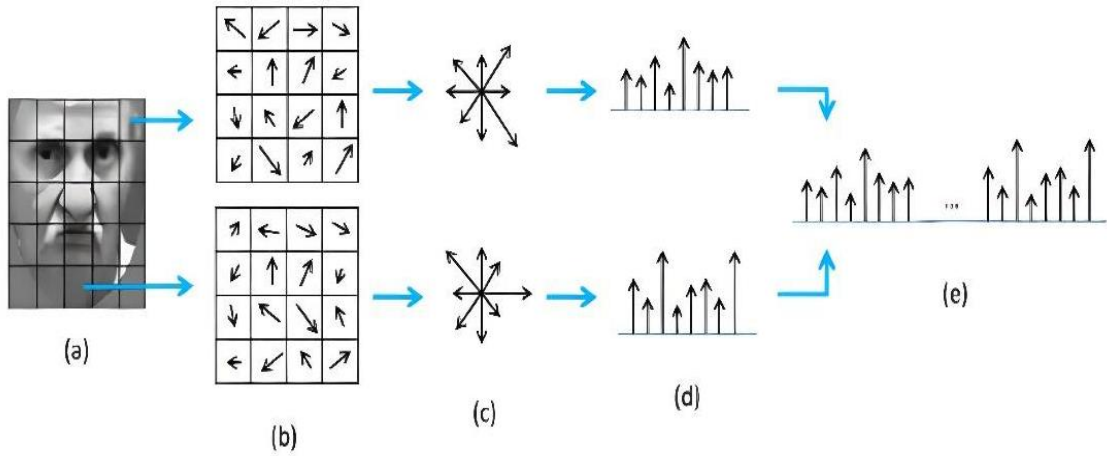
$$size_{future/image} = n_{block/image} * size_{future/block}$$

Trong đó:

$n_{block/image}$: Là số khối trên ảnh.

$size_{future/block}$: Là số chiều vecto đặc trưng của mỗi khối

Tổ hợp của tất cả các đồ thị sẽ là các đặc trưng HoG của ảnh.



Hình 2. 17: Hình ảnh mô tả về trích chọn đặc trưng bởi HoG

2.2.3. Phương pháp trích chọn đặc trưng LBP

2.2.3.1. Giới thiệu

LBP là viết tắt của Local Binary Pattern hay là mẫu nhị phân địa phương được Ojala trình bày vào năm 1996 như là một cách đo độ tương phản cục bộ của ảnh. Phiên bản đầu tiên của LBP được dùng với 8 điểm ảnh xung quanh và sử dụng giá trị của điểm ảnh ở trung tâm làm ngưỡng. Giá trị LBP được xác định bằng cách nhân các giá trị ngưỡng với trọng số ứng với mỗi điểm ảnh sau đó cộng tổng lại. Hình dưới minh họa cách tính độ tương phản trực giao là hiệu cấp độ xám trung bình của các điểm ảnh lớn hơn hoặc bằng ngưỡng với các điểm ảnh thấp hơn ngưỡng.

Kể từ khi được đưa ra, theo định nghĩa là bất biến với những thay đổi đơn điệu trong ảnh đen trắng. Để cải tiến phương pháp, bổ sung thêm phương pháp tương phản trực giao địa phương. Hình dưới minh họa cách tính độ tương phản trực giao là ký hiệu cấp độ xám trung bình của các điểm ảnh lớn hơn hoặc bằng ngưỡng với các điểm ảnh thấp hơn ngưỡng. Phân phối hai chiều của mã LBP và độ tương phản cục bộ được lấy làm đặc trưng gọi là LBP/C.

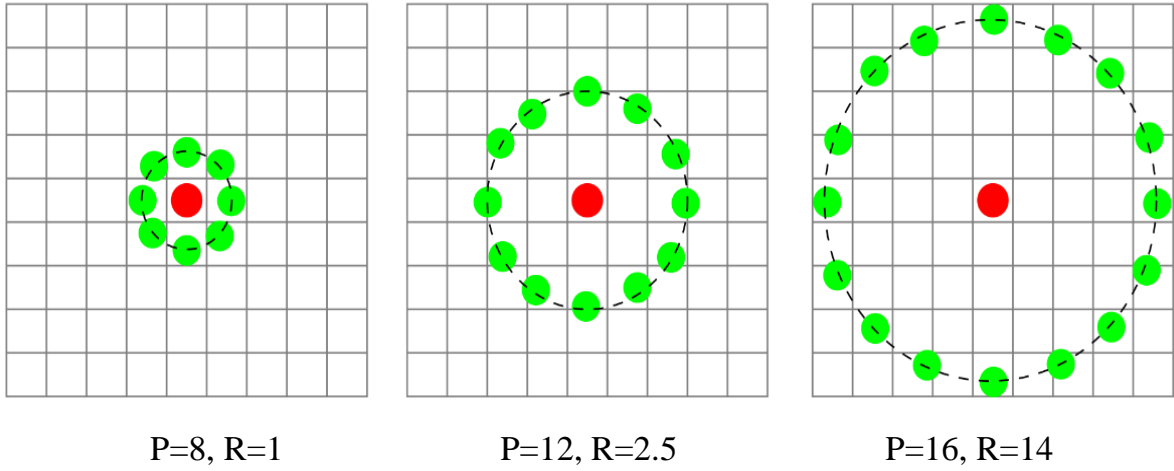
65	21	43	Ngưỡng	1	0	0	Trọng số	$1 * 2^0$	$0 * 2^1$	$0 * 2^2$
24	45	79		0		1		$0 * 2^7$		$1 * 2^3$
88	57	32		1	1	0		$1 * 2^6$	$1 * 2^5$	$0 * 2^4$

Bảng 2.1: Bảng thể hiện về phương pháp LBP

Dãy LBP được Ojala trình bày vào năm 2002. Định nghĩa một cấu trúc điểm ảnh T là một phân phối đại số của cấp độ xám của P + 1 (P > 0) điểm ảnh.

$$T = t(gc, g_0, \dots, g_{p-1})$$

Với g_c ứng với cấp độ xám của điểm ảnh trung tâm P_{tt} , g_p ($p = 0, \dots, 1$) tương ứng với P điểm ảnh xung quanh, P điểm ảnh này nằm trên đường tròn bán kính R và tâm là P_{tt}



Hình 2. 18: Hình ảnh mô tả tập hợp các điểm xung quanh P_{tt}

Không mất thông tin, có thể trừ g_c đi một lượng là g_c

$$T = t(g_c, g_0 - g_c, \dots, g_{p-1} - g_c)$$

Giả sử sự sai số giữa g_p và g_c là độc lập với g_c , ta có thể nhân tử hóa g_c như sau:

$$T = t(g_c) t(g_0 - g_c, \dots, g_{p-1} - g_c)$$

$t(g_c)$ biểu thị xu hướng độ sáng tối của cả bức ảnh nên không liên quan đến kết cấu của ảnh cục bộ do đó có thể bỏ qua

$$T \sim t((g_0 - g_c), \dots, (g_{p-1} - g_c))$$

Mặc dù tính bất biến ngược với độ thay đổi tỷ lệ xám của điểm ảnh, sự khác biệt ảnh hưởng bởi tỷ lệ. Để thu được đặc điểm bất biến với bất kỳ một sự thay đổi nào của ảnh đen trắng chỉ quan tâm đến dấu của độ lệch:

$$T \sim t(s(g_0 - g_c), \dots, s(g_{p-1} - g_c))$$

Với s là hàm dấu: $s(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$

Trọng số 2_p được dùng cho các hàm dấu, $s(g_p - g_c)$ để chuyển sự khác biệt giữa các điểm ảnh bên cạnh về một giá trị duy nhất.

$$LBP_{P,R} = \sum_{p=0}^{p-1} s(g_p - g_c) * 2^p$$

Với P pixel thì có 2_p giá trị $LBP_{P,R}$ trong khoảng $[0, 2_{p-1}]$ nhưng để đơn giản ta có thể chọn một số giá trị trong 2_p giá trị ký hiệu là $LBP_{P,R}^u$.

2.2.3.2. Các bước trích chọn đặc trưng LBP trên ảnh

Thông tin LBP của pixel tại trung tâm của mỗi khối ảnh sẽ được tính dựa trên thông tin của các pixel lân cận. Có thể tóm tắt các bước tiến hành như sau:

Bước 1: Xác định bán kính làm việc

Bước 2: Tính giá trị LBP cho pixel ở trung tâm (xc, yc) khối ảnh dựa trên thông tin của các pixel lân cận.

$$LBP_{P,R}(x_c, y_c) = \sum_{p=1}^P s(g_p - g_c) 2^p$$

Trong đó, (g_p) là giá trị grayscale của các pixel lân cận, (g_c) là giá trị grayscale của các trung tâm và (s) là hàm nhị phân được xác định như sau: $s(z) = 1$ nếu giá trị $z \geq 0$.

47	51	65
62	70	70
80	83	78

-23	-19	-5
-8	*	0
10	13	8

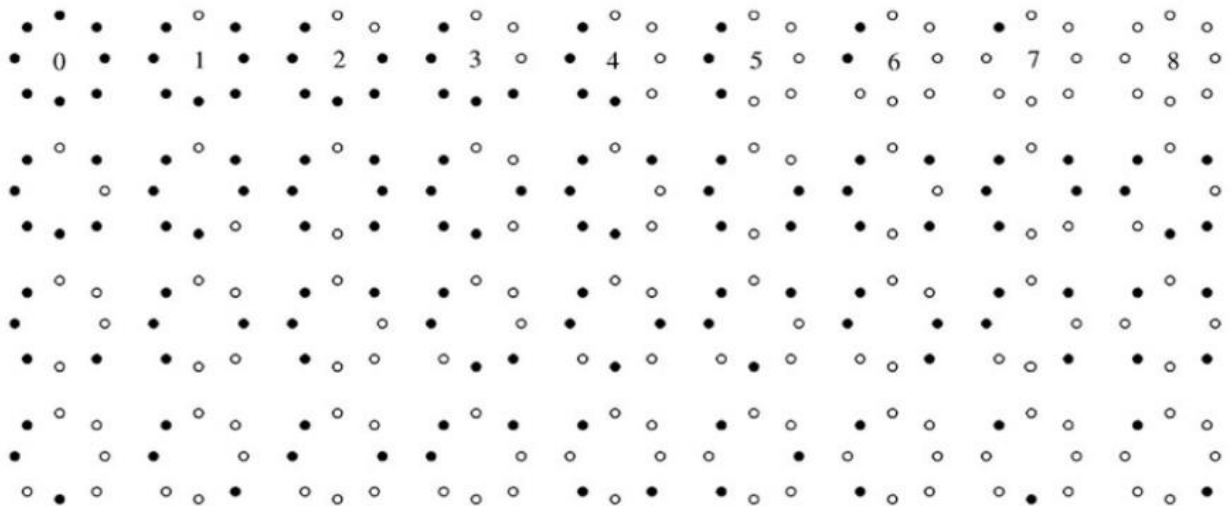
0	0	0
0	*	1
1	1	1

Bảng 2.1: Bảng thể hiện biến thể LBP, LBP đồng dạng

$$1*2^0 + 1*2^1 + 1*2^2 + 1*2^3 + 0*2^4 + 0*2^5 + 0*2^6 + 0*2^7 = 15.$$

Một mẫu nhị phân được gọi là đồng dạng khi xét chuỗi bit xoay vòng thì có nhiều nhất là 2 lần thay đổi từ giá trị bit 0 sang 1 hoặc từ giá trị bit 1 sang 0. Ví dụ: 00000000 có 0 sự chuyển tiếp, 01110000 có 2 sự chuyển tiếp, 11001111 có 2 sự chuyển tiếp nên đây là LBP đồng dạng. 11001001 có 4 sự chuyển tiếp, 01010011 có 6 sự chuyển tiếp nên không phải là LBP đồng dạng.

Dựa trên định nghĩa này, bảng ánh xạ cho thấy 36 mẫu nhị phân bất biến xoay duy nhất có thể xảy ra trong vùng lân cận $P = 8$ điểm, trong đó vòng tròn đen có giá trị bit là 0 và vòng tròn trắng có giá trị bit là 1. Hình vẽ sau đây thể hiện 59 nhãn (mẫu) và minh họa về biểu đồ của đặc trưng LBP đồng dạng.



Hình 2. 19: Hình ảnh mô tả các mẫu LBP đồng dạng

Nguyên lý phân lớp không tham biến:

Trong phân lớp, sự khác biệt giữa mẫu và mô hình phân phối LBP được đánh giá bởi kiểm tra thống kê không tham biến. Phương pháp tiếp cận này có ưu điểm là không

cần phải có những giả thiết về phân phối của các đặc trưng. Thông thường, những kiểm tra thống kê được chọn cho mục đích là nguyên lý crossentropy được giới thiệu bởi Kullback (1968). Sau đó, Sokal và Rohlf (1969) gọi cách đo này là thống kê G.

$$G(S,M) = 2 * \sum_{b=1}^B [S_b * \log S - S_b * \log M_b]$$

Trong đó:

S, M là: kí hiệu phân phối mẫu và mô hình mong muốn.

S_b và M_b là xác suất để b thuộc vào phân phối mẫu hoặc mô hình.

B là số phần tử trong phân phối.

Thống kê G sử dụng trong phân lớp có thể viết lại như sau:

$$L(S,M) = - \sum_{b=1}^B S_b \log M_b$$

Kiến trúc mô hình có thể xem như xử lý ngẫu nhiên có đặc tính có thể xác định bởi phân phối LBP. Trong một phân lớp đơn giản, mỗi lớp được biểu diễn bởi một mô hình phân phối đơn giản M_i . Tương tự, một kiến trúc mẫu không xác định có thể miêu tả bởi phân phối S. L là một giả ma trận đo khả năng mẫu S có thể thuộc lớp i.

Lớp C của một mẫu không xác định có thể được xác định bởi luật “hàng xóm gần nhất”: $C = \operatorname{argmin}_i L(S, M_i)$

Bên cạnh đó, một thống kê log-likelihood có thể xem như đơn vị đo sự khác biệt và có thể sử dụng để liên kết nhiều bộ phân lớp giống như bộ phân lớp k-NN hoặc self-organizing map (SOM). Log-likelihood đúng trong một số trường hợp nhưng không ổn định khi mà cỡ mẫu nhỏ. Trong trường hợp này Chi-square- distance thường cho kết quả tốt hơn:

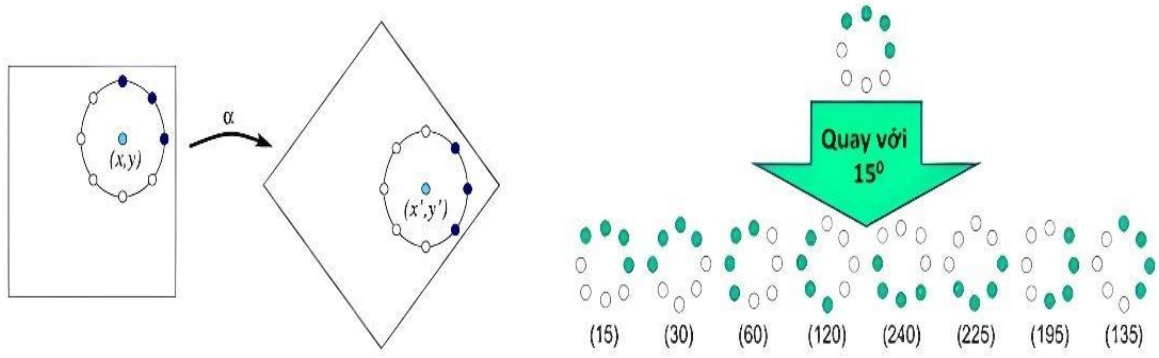
$$X^2(S,M) = \sum_{b=1}^B \frac{(S_b - M_b)^2}{S_b + M_b}$$

Để đạt được độ chính xác cao sử dụng giao Histogram

$$H(S,M) = \sum_{b=1}^B \min(S_b, M_b)$$

LBP bất biến với phép quay

Giả sử $I_\alpha(x, y)$ là ảnh quay góc (α) của ảnh $I(x, y)$. Với phép quay này điểm ảnh (x, y) sẽ nằm tại vị trí (x', y') như hình vẽ sau đây (hình trái). Trong ví dụ này (hình phải): tất cả 8 mẫu LBP bên dưới được ánh xạ về mẫu LBP đầu tiên vì mẫu đầu tiên cho giá trị nhỏ nhất.



Hình 2. 20: Hình ảnh mô tả về các trường hợp của LBP sau khi quay với góc 15 độ

LBP đồng dạng có khả năng bất biến với phép quay kết hợp của mẫu LBP đồng dạng và LBP bất biến với phép quay có thể tạo nên một dạng biến thể khác của LBP (uniform rotation – invariant LBP). Dựa trên định nghĩa này, bảng ánh xạ cho bán kính làm việc P-neighbors sẽ có P + 2 nhãn (label). Có nghĩa là có 10 nhãn trong trường hợp làm việc với 8-neighbour.

Để không bị ảnh hưởng bởi sự quay, mỗi giá trị LBP cần quay ngược lại về vị trí ban đầu, cách tốt nhất là tạo ra tất cả các trường hợp quay của một mẫu, sự quay có thể định nghĩa như sau:

$$LBP_{R,I}^i = \min\{ROR(LBP_{P,R}, i) \mid i = 0, 1, \dots, P - 1\}$$

Trong đó:

ri là viết tắt của rotation invariant (quay bất biến),

ROR(x,i) dịch vòng tròn số nhị phân P - bit (x) i lần

Độ tương phản và kết cấu mẫu:

Kết cấu có thể được coi là một hiện tượng hai chiều được đặc trưng bởi hai đặc tính trực giao: cấu trúc không gian (mô hình) và độ tương phản (độ mạnh của mô hình). Quay bất biến tương phản địa phương có thể được đo trong một hình tròn đối xứng xung quanh giống như LBP:

Kết cấu có thể được coi là một hiện tượng hai chiều được đặc trưng bởi hai đặc tính trực giao: cấu trúc không gian (mô hình) và độ tương phản (độ mạnh của mô hình). Quay bất biến tương phản địa phương có thể được đo trong một hình tròn đối xứng xung quanh giống như LBP:

$$VAR_{P,R} = \frac{1}{p} \sum_{p=0}^{p=1} (g_p - \mu)^2$$

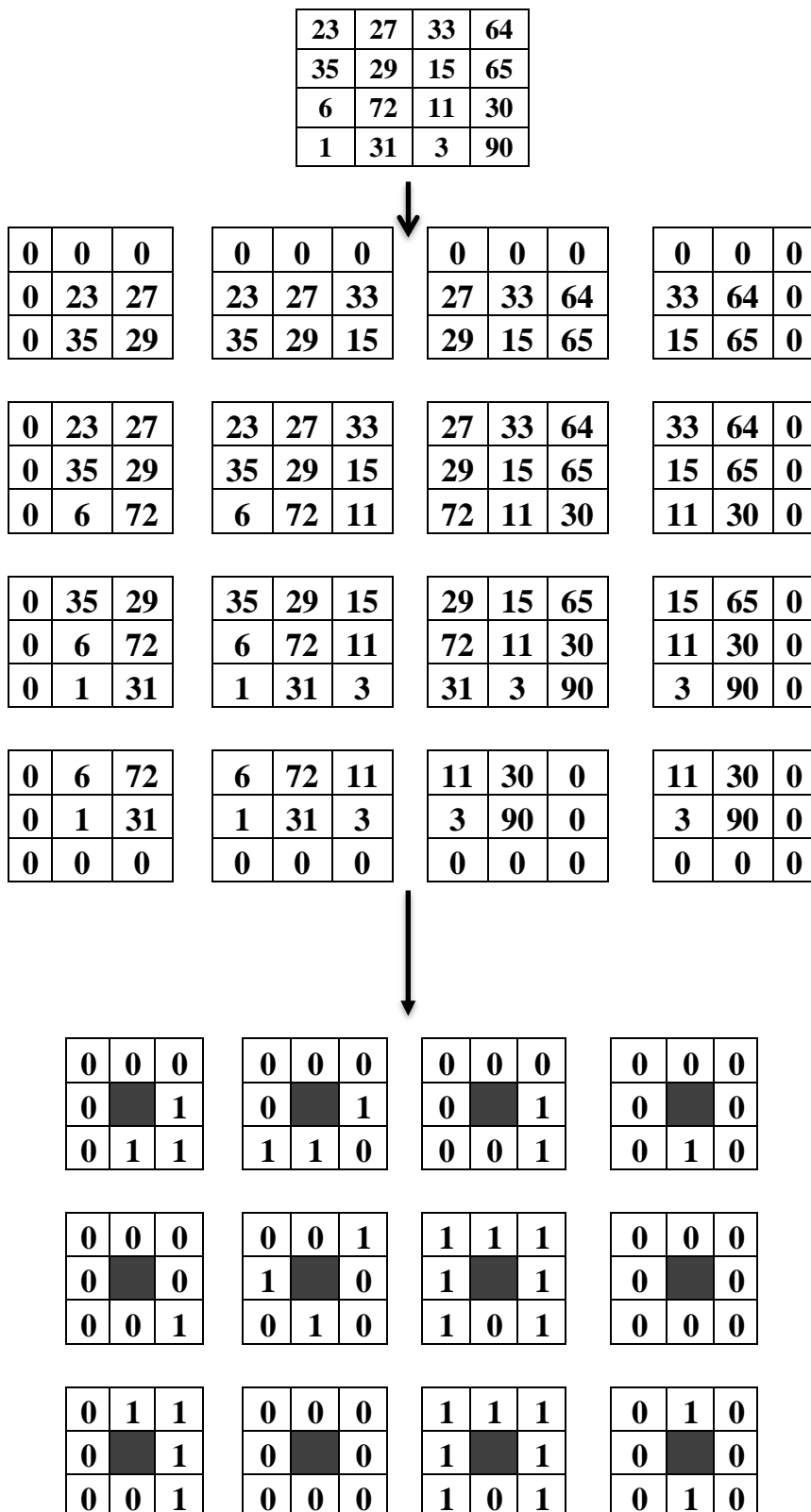
Trong đó:

$$\mu = \frac{1}{p} \sum_{p=0}^{p-1} g_p$$

Tổng hợp lại ta có: $LBP_{P1,R1}^{ri} / VAR_{P2,R2}$

Ví dụ về trích rút đặc trưng LBP trên ảnh số:

Ví dụ với ảnh có kích thước 4x4, sẽ tính được giá trị đặc trưng LBP như sau:

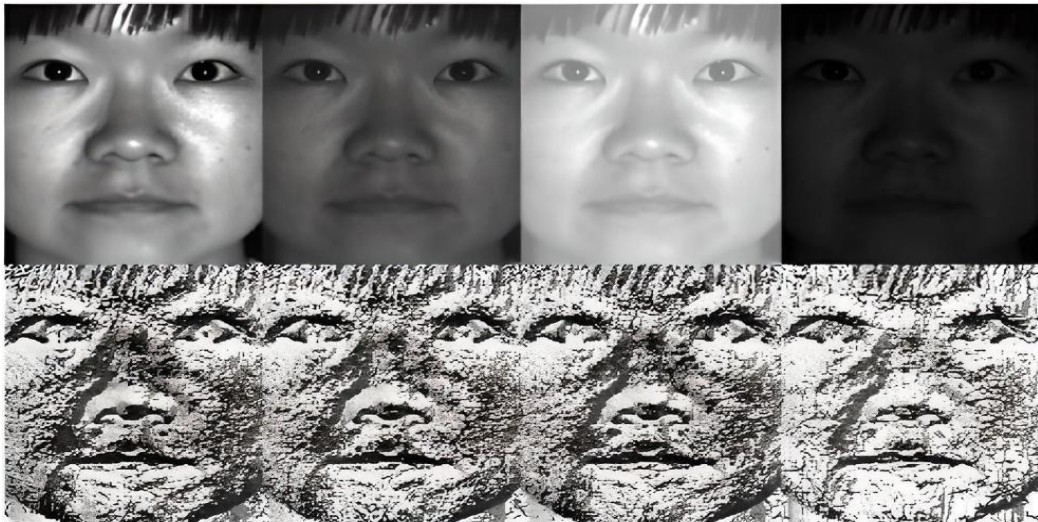


0	1	1		0	1	0		1	1	1		0	0	0
0		1		0		3		1		1		0		0
0	0	0		0	0	0		0	0	0		0	0	0

56	104	24	32
16	164	223	2
30	0	223	34
14	2	143	0

Bảng 2.2: Bảng thể hiện ví dụ về tính toán đặc trưng LBP

Nhận xét: Thuật toán trích rút đặc trưng LBP cài đặt đơn giản, thời gian tính toán giá trị đặc trưng nhanh vì nó làm việc với giá trị nguyên. Thuật toán LBP do sử dụng các giá trị sai khác về đánh giá giữa các điểm lân cận nên thuật toán LBP không bị lỗi khi ảnh có độ sáng khác nhau.



Hình 2.21: Phương pháp LBP không phụ thuộc vào độ sáng cho cùng một đối tượng giống nhau

2.3. Tìm hiểu về Mạng thần kinh tích chập CNN

2.3.1. Giới thiệu

CNN là viết tắt Convolutional Neural Network hay mạng thần kinh tích chập là một trong những mô hình Deep Learning tiên tiến. Được lấy cảm hứng từ não của con người.

CNN có lịch sử khá lâu đời. Kiến trúc gốc của mô hình CNN được giới thiệu bởi một nhà khoa học máy tính người Nhật vào năm 1980. Sau đó, năm 1998, Yan LeCun lần đầu huấn luyện mô hình CNN với thuật toán backpropagation cho bài toán nhận dạng chữ viết tay.

Có thể lấy hình ảnh đầu vào, gán độ quan trọng cho các đặc trưng hoặc đối tượng khác nhau trong hình ảnh và có thể phân biệt được từng đặc trưng hoặc đối tượng này với nhau. Công việc tiền xử lý được yêu cầu cho mạng nơron tích chập thì ít hơn nhiều so với các thuật toán phân loại khác. Trong các phương thức sơ khai, các bộ lọc được thiết kế bằng tay, với một quá trình huấn luyện để chọn ra các bộ lọc hoặc các đặc trưng phù hợp thì mạng nơron tích chập lại có khả năng tự học để chọn ra bộ lọc hoặc các đặc trưng tối ưu nhất.

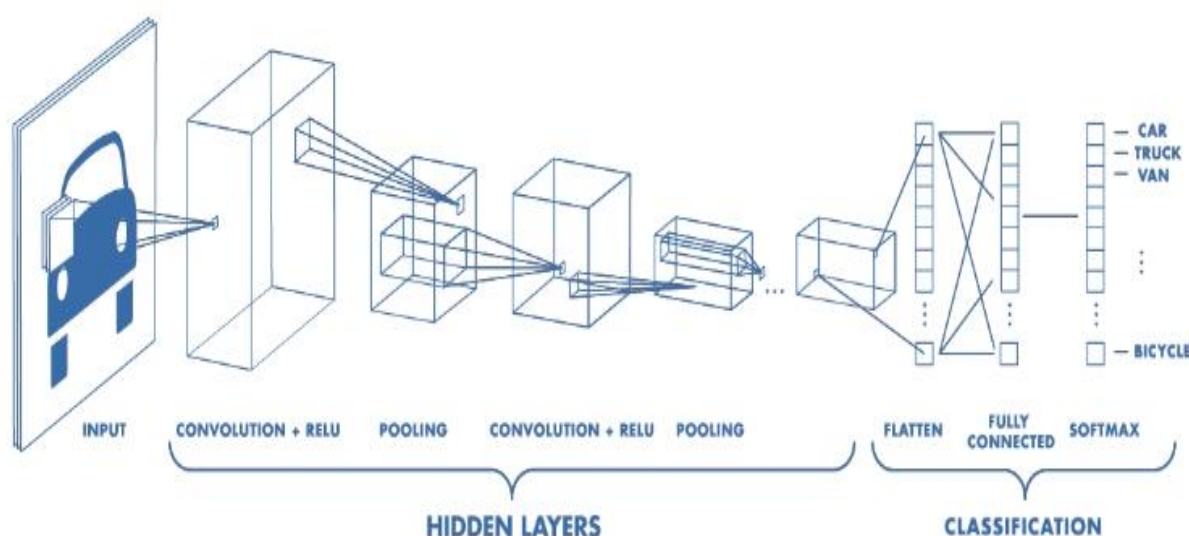
CNN có mối liên kết chặt chẽ với sinh học, cụ thể là của võ não thị giác, nơi xử lý thông tin liên quan đến hình ảnh từ các tế bào cảm thụ ánh sáng nằm ở mắt người. Năm 1962, 2 nhà thần kinh học người Mỹ là Hubel and Wiesel đã thực hiện thí nghiệm khám phá cách tổ chức của các tế bào não để xử lý thông tin thị giác và các tổ chức này đảm nhận nhiệm vụ nào.

2.3.2. Kiến trúc mạng CNN

Mạng Nơ-ron Tích Chập có kiến trúc khác với Mạng Nơ-ron thông thường. Mạng Neural Tích Chập được chia thành 3 chiều: rộng, cao, và sâu. Kể đến, các Neural trong mạng không liên kết hoàn toàn với toàn bộ Neural kế đến nhưng chỉ liên kết tới một vùng nhỏ. Cuối cùng, một tầng đầu ra được tối giản thành véc-tơ của giá trị xác suất.

Mạng CNN gồm hai thành phần:

- Phần trích xuất đặc trưng: trong phần này, mạng sẽ tiến hành tính toán hàng loạt phép tích chập và phép hợp nhất (pooling) để phát hiện các đặc trưng.
- Phần phân lớp: tại phần này, một lớp với các liên kết đầy đủ sẽ đóng vai trò như một bộ phân lớp các đặc trưng đã rút trích được trước đó. Tầng này sẽ đưa ra xác suất của một đối tượng trong hình.



Hình 2. 22: Hình ảnh mô tả kiến trúc mạng CNN

2.3.2.1. Rút trích đặc trưng

Lớp tích chập - Convolution Layer

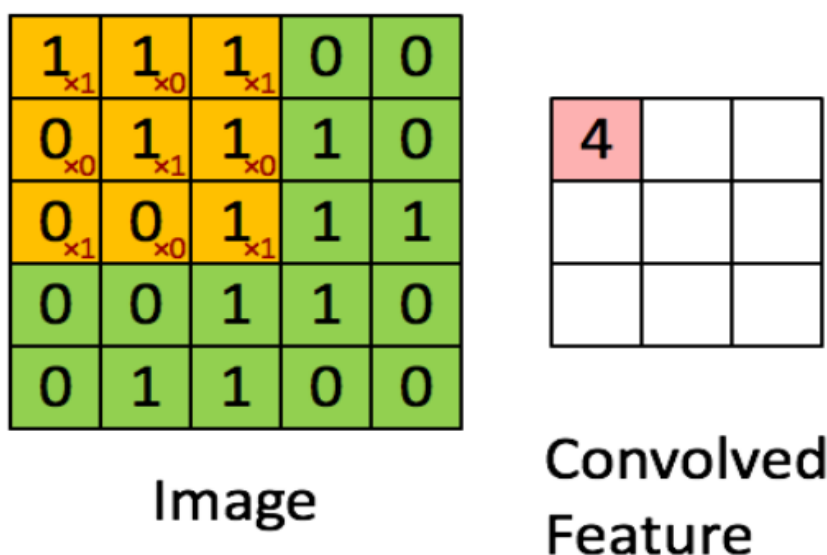
Tích chập là một khối quan trọng trong CNN. Là lớp đầu tiên để trích xuất các tính năng từ hình ảnh đầu vào. Tích chập duy trì mối quan hệ giữa các điểm ảnh bằng cách tìm hiểu các tính năng hình ảnh bằng cách sử dụng các ô vuông nhỏ của dữ liệu đầu vào. Nó là 1 phép toán có 2 đầu vào như ma trận hình ảnh và 1 bộ lọc hoặc hạt nhân.

Trong trường hợp CNN, tích chập được thực hiện trên giá trị đầu vào của dữ liệu và kernel/filter để tạo ra một bản đồ đặc trưng. Thực hiện phép tích chập bằng cách trượt kernel/filter theo dữ liệu đầu vào. Tại mỗi vị trí, ta tiến hành phép nhân ma trận và tính tổng các giá trị để đưa vào bản đồ đặc trưng.

Trong thực tế, tích chập được thực hiện trên không gian 3 chiều. Vì mỗi hình ảnh được biểu diễn dưới dạng 3 chiều: rộng, cao, và sâu. Chiều sâu ở đây chính là giá trị màu sắc của hình (RGB).

Ta thực hiện phép tích chập trên đầu vào nhiều lần khác nhau. Mỗi lần sử dụng một kernel/filter khác nhau. Kết quả ta sẽ thu được những bản đồ đặc trưng khác nhau. Cuối cùng, kết hợp toàn bộ bản đồ đặc trưng này thành kết quả cuối cùng của tầng tích chập.

Trong hình ảnh minh họa như **hình 2.23**, ma trận bên trái là một hình ảnh trắng đen được số hóa. Ma trận có kích thước 5×5 và mỗi điểm ảnh có giá trị 1 hoặc 0 là giao điểm của dòng và cột. Convolution. Sliding Window hay còn gọi là kernel, filter hoặc feature detect là một ma trận có kích thước nhỏ như trong ví dụ trên là 3×3 . Convolution hay tích chập là nhân từng phần tử bên trong ma trận 3×3 với ma trận bên trái. Kết quả được một ma trận gọi là Convolved feature được sinh ra từ việc nhân ma trận Filter với ma trận ảnh 5×5 bên trái.



Hình 2. 23: Hình ảnh mô tả phương pháp tích chập

Hàm phi tuyến - ReLU

Đây chính là một hàm kích hoạt trong Neural Network. Có thể biết đến hàm kích hoạt này với một tên gọi khác là Activation Function. Nhiệm vụ chính của hàm kích hoạt là mô phỏng lại các Neuron có tỷ lệ truyền xung qua Axon.

Với đầu ra là: $f(x) = \max(0, x)$

Trong quá trình trượt kernel/filter trên dữ liệu đầu vào, ta sẽ quy định một bước nhảy với mỗi lần di chuyển. Thông thường ta lựa chọn thường chọn bước nhảy là 1. Nếu kích thước bước nhảy tăng, kernel/filter sẽ có ít ô trùng lặp.

Bởi vì kích thước đầu ra luôn nhỏ hơn đầu vào nên ta cần một phép xử lý đầu vào để đầu ra không bị co giãn. Đơn giản ta chỉ cần thêm một lề nhỏ vào đầu vào. Một lề với giá trị 0 sẽ được thêm vào xung quanh đầu vào trước khi thực hiện phép tích chập.

15	20	-10	35	Hàm ReLU	15	20	0	35
18	-110	25	100		18	0	25	100
20	-15	25	-10		20	0	25	0
101	75	18	23		101	75	18	23

Bảng 2.3: Bảng mô tả phương pháp kích hoạt ReLU



Lớp gộp- Pooling Layer

Khi đầu vào quá lớn, các lớp Pooling Layer sẽ được dịch chuyển vào giữa những lớp Convolutional Layer nhằm giảm các Parameter. Mục đích của tầng này là để nhanh chóng giảm số chiều. Việc này giúp giảm thời gian học và hạn chế việc overfitting.

Pooling Layer được biết đến với ba loại phổ biến là: Max Pooling, Average Pooling và Sum Pooling.

Tại Pooling Layer, khi sử dụng lớp Max Pooling thì số lượng Parameter có thể sẽ giảm đi. Vì vậy, Convolutional Neural Network sẽ xuất hiện nhiều lớp Filter Map, mỗi Filter Map đó sẽ cho ra một Max Pooling khác nhau.

Lớp pooling sẽ giảm bớt số lượng tham số khi hình ảnh quá lớn. Không gian pooling còn được gọi là lấy mẫu con hoặc lấy mẫu xuống làm giảm kích thước của mỗi map nhưng vẫn giữ lại thông tin quan trọng.

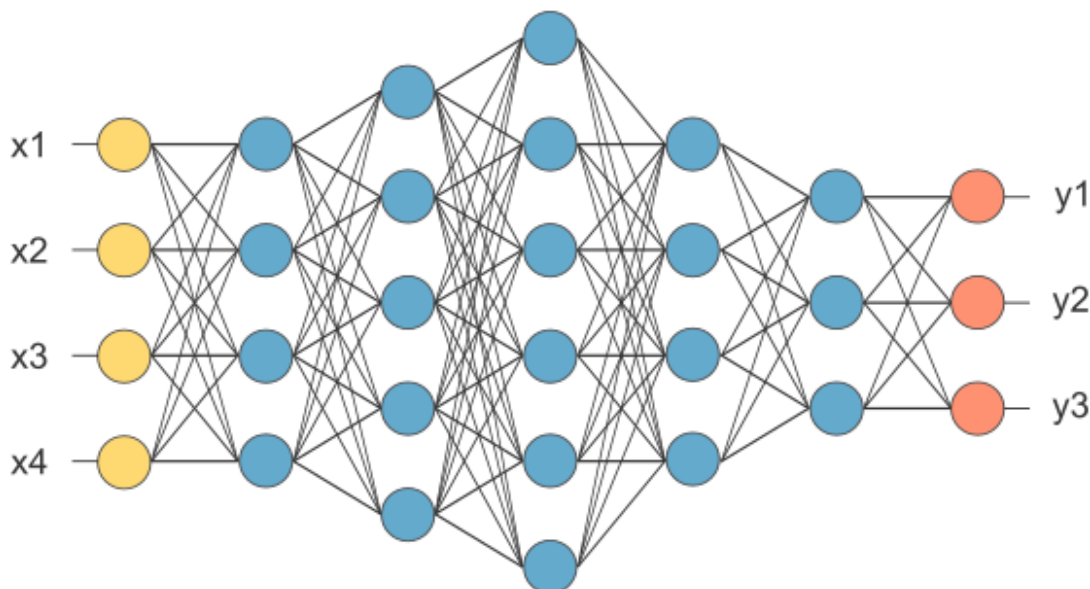
				Max Pooling			
12	20	30	0		20	30	
8	12	2	0		112	37	
34	70	37	4				
112	100	25	12				
				Average Pooling			
					13	8	
					79	20	

Bảng 2.4: Max Pooling và Average Pooling kích thước 2x2

2.3.2.2. Phân lớp

Trong phần phân lớp, sử dụng một vài tầng với kết nối đầy đủ (**Fully Connected Layer**) để xử lý kết quả của phần tích chập. Vì đầu vào của mạng liên kết đầy đủ là 1 chiều, ta cần làm phẳng đầu vào trước khi phân lớp. Tầng cuối cùng trong mạng CNN là một tầng liên kết đầy đủ, phần này hoạt động tương tự như mạng nơ-ron thông thường.

Tầng này có chức năng chuyển ma trận đặc trưng ở tầng trước thành vector chứa xác suất của các đối tượng cần được dự đoán. Kết quả thu được cuối cùng cũng sẽ là một véc-tơ với các giá trị xác suất cho việc dự đoán như mạng nơ-ron thông thường.



Hình 2. 24: Hình ảnh mô tả lớp kết nối đầy đủ

2.3.3. Xây dựng mô hình CNN

Mạng CNN sử dụng 3 ý tưởng cơ bản:

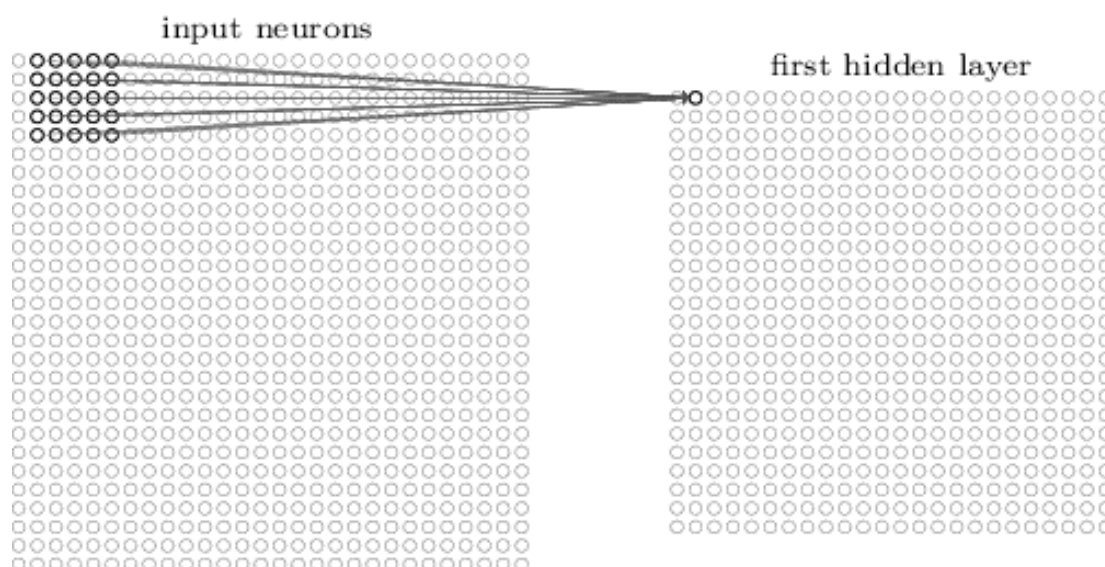
- Trường tiếp nhận cục bộ (Local receptive field)
- Trọng số chia sẻ (Shared weights)
- Tổng hợp (Pooling)

2.3.3.1. Trường tiếp nhận cục bộ (Local receptive field)

CNN sử dụng các mối tương quan không gian tồn tại trong dữ liệu đầu vào. Mỗi trong các lớp đồng thời của mạng nơ-ron kết nối một số nơ-ron đầu vào. Vùng cụ thể này được gọi là Trường tiếp nhận cục bộ (Local Receptive Field). Nó chỉ tập trung vào các tế bào thần kinh ẩn (hidden neurons). Nơ-ron ẩn (hidden neuron) sẽ xử lý dữ liệu đầu vào bên trong trường được đề cập mà không nhận ra những thay đổi bên ngoài ranh giới cụ thể.

Trong các tầng kết nối đầy đủ được chỉ ra, đầu vào của mạng CNN là một ảnh đã được mô tả là một đường thẳng đứng chứa các nơ-ron. Giả sử trong mô hình mạng tích chập, ảnh có kích thước 28x28 thì tương ứng đầu vào là một ma trận có 28x28 và giá trị mỗi điểm ảnh là một ô trong ma trận cường độ điểm ảnh.

Chúng ta sẽ kết nối các điểm ảnh đầu vào cho các nơron ở tầng ẩn. Để được chính xác hơn, mỗi nơron trong lớp ẩn đầu tiên sẽ được kết nối với một vùng nhỏ của các nơron đầu vào như một filter có kích thước 5×5 tương ứng $(28 - 5 + 1) = 24$ điểm ảnh đầu vào. Mỗi một kết nối sẽ học một trọng số và mỗi neuron ẩn sẽ học một bias. Mỗi một vùng 5×5 gọi là một trường tiếp nhận cục bộ, đối với một nơron ẩn cụ thể, chúng ta có thể có các kết nối như **hình 2.25**.



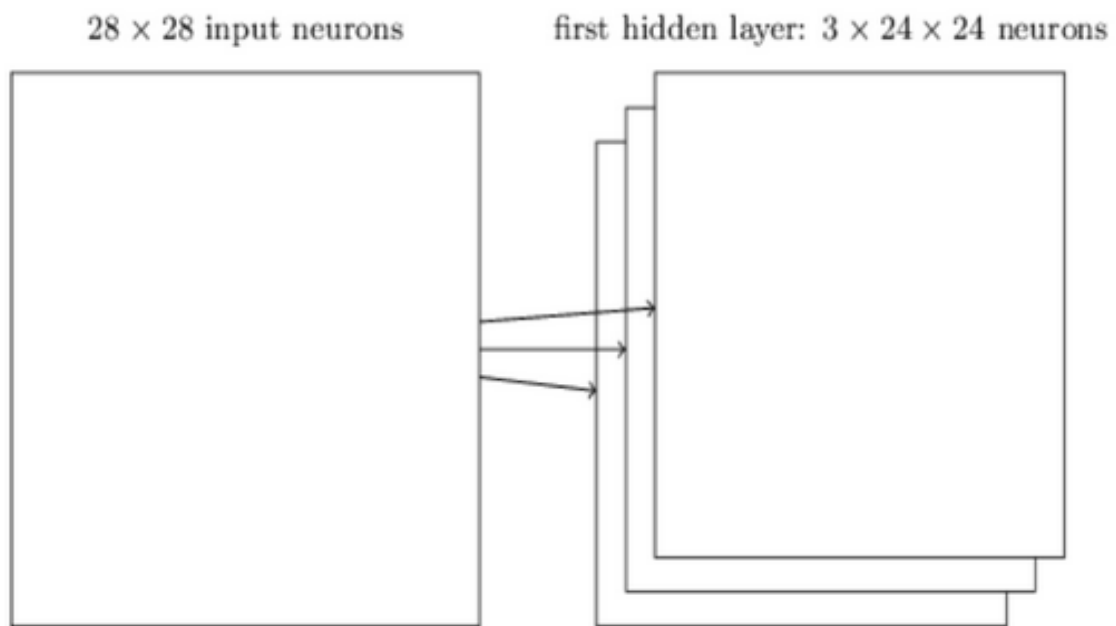
Hình 2. 25: Vị trí bắt đầu của trường tiếp nhận cục bộ

Vùng đó trong bức ảnh đầu vào được gọi là vùng tiếp nhận cục bộ cho nơron ẩn. Đó là một cửa sổ nhỏ trên các điểm ảnh đầu vào. Mỗi kết nối sẽ học một trọng số. Và nơron ẩn cũng sẽ học một độ lệch (overall bias). Ta có thể hiểu rằng nơron lớp ẩn cụ thể là học để phân tích trường tiếp nhận cục bộ cụ thể của nó.

Sau đó chúng ta trượt trường tiếp nhận cục bộ trên toàn bộ bức ảnh. Đối với mỗi trường tiếp nhận cục bộ, có một nơron ẩn khác trong tầng ẩn đầu tiên.

Sau đó chúng ta trượt trường tiếp nhận cục bộ một điểm ảnh bên phải (tức là bằng một nơron), để kết nối với một nơron ẩn thứ hai.

Với bài toán nhận dạng ảnh ta thường gọi ma trận lớp đầu vào là feature map, trọng số xác định các đặc trưng là shared weight và độ lệch xác định một feature map là shared bias. Như vậy đơn giản nhất là qua các bước trên chúng ta chỉ có 1 feature map. Tuy nhiên trong nhận dạng ảnh chúng ta cần nhiều hơn một feature map.



Hình 2. 26: Trường tiếp nhận cục bộ với ba bản đồ đặc trưng

Như vậy, local receptive field thích hợp cho việc phân tách dữ liệu ảnh, giúp chọn ra những vùng ảnh có giá trị nhất cho việc đánh giá phân lớp.

2.3.3.2. Trọng số chia sẻ (Shared weights)

Các trọng số cho mỗi kernel/filter phải giống nhau. Tất cả các nơ-ron trong lớp ẩn đầu sẽ phát hiện chính xác feature tương tự chỉ ở các vị trí khác nhau trong hình ảnh đầu vào. Chúng ta gọi việc map từ input layer sang hidden layer là một Feature map. Ta cần tìm ra mối quan hệ giữa số lượng Feature map với số lượng tham số.

Mỗi Feature map cần $5 \times 5 = 25$ shared weight và 1 shared bias. Như vậy mỗi feature map cần $5 \times 5 + 1 = 26$ tham số.

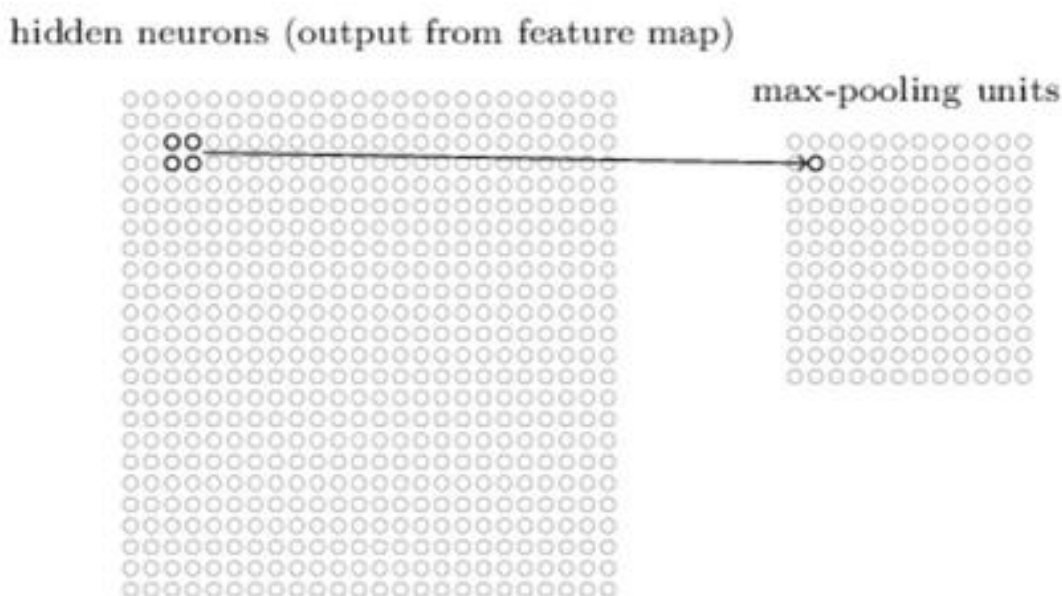
Như vậy nếu có 10 feature map thì có $10 \times 26 = 260$ tham số. Chúng ta xét lại nếu layer đầu tiên có kết nối đầy đủ nghĩa là chúng ta có $28 \times 28 = 784$ neuron đầu vào như vậy ta chỉ có 30 neuron ẩn. Như vậy ta cần $28 \times 28 \times 30$ shared weight và 30 shared bias. Tổng số tham số là $28 \times 28 \times 30 + 30$ tham số lớn hơn nhiều so với CNN. Ví dụ vừa rồi chỉ mô tả để thấy được sự ước lượng số lượng tham số chứ chúng ta không so sánh được trực tiếp vì 2 mô hình khác nhau. Nhưng điều chắc chắn là nếu mô hình có số lượng tham số ít hơn thì nó sẽ chạy nhanh hơn.

2.3.3.3. Tổng hợp (Pooling)

Lớp pooling thường được sử dụng ngay sau lớp Convolutional. Những gì các lớp pooling làm là để đơn giản hóa thông tin đầu ra để giảm bớt số lượng neuron.

Ví dụ, mỗi đơn vị trong lớp pooling có thể thu gọn một vùng 2×2 nơ-ron trong lớp trước. Một thủ tục pooling phổ biến là Max Pooling. Trong Max Pooling, một đơn vị

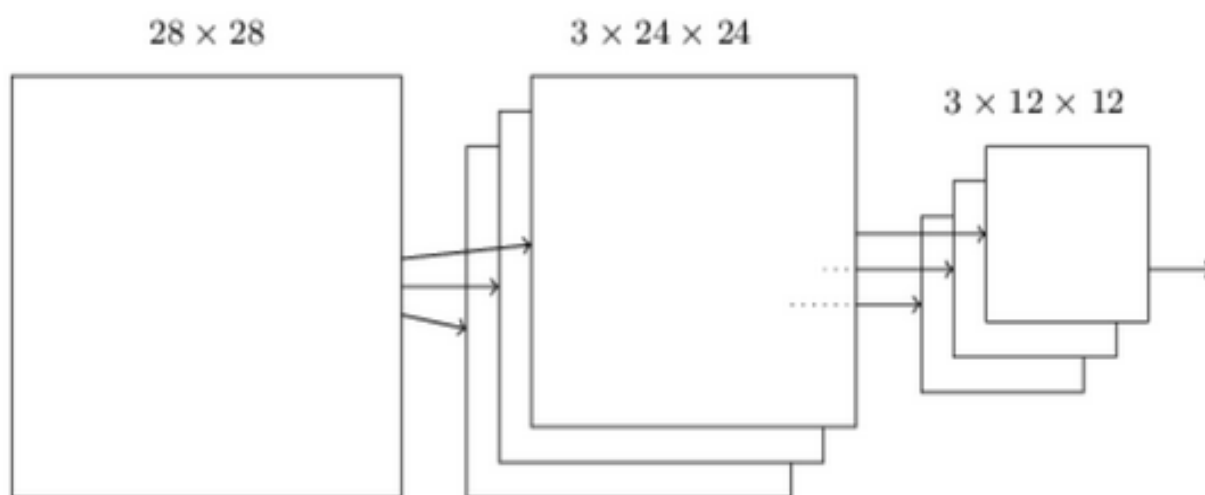
Pooling chỉ đơn giản là kết quả đầu ra kích hoạt giá trị lớn nhất trong vùng đầu vào 2×2 , như minh họa trong sơ đồ sau:



Hình 2. 27: Ví dụ Max Pooling 2x2

Lưu ý rằng bởi vì chúng ta có 24×24 nơron đầu ra từ các lớp tích chập, sau khi pooling chúng ta có 12×12 nơron.

Như đã đề cập ở trên, lớp tích chập thường có nhiều hơn một bản đồ đặc trưng. Chúng ta áp dụng Max Pooling cho mỗi bản đồ đặc trưng riêng biệt. Vì vậy, nếu có ba bản đồ đặc trưng, các lớp tích chập và Max Pooling sẽ kết hợp như sau:



Hình 2. 28: Max pooling với ba bản đồ đặc trưng

Có thể hiểu Max Pooling như là một cách cho mạng để hỏi xem một đặc trưng nhất được tìm thấy ở bất cứ đâu trong một khu vực của ảnh. Sau đó nó bỏ đi những thông tin định vị chính xác. Trực giác là một khi một đặc trưng đã được tìm thấy, vị trí chính xác của nó là không quan trọng như vị trí thô của nó so với các đặc trưng khác. Một lợi ích

lớn là có rất nhiều tính năng gộp ít hơn (fewer pooled features), và vì vậy điều này sẽ giúp giảm số lượng các tham số cần thiết trong các lớp sau.

Một Convolutional layer bao gồm các feature map khác nhau. Mỗi một feature map giúp phát hiện một vài đặc trưng trong bức ảnh. Lợi ích lớn nhất của trọng số chia sẻ là giảm tối đa số lượng tham số trong mạng CNN.

Như vậy qua lớp Max Pooling thì số lượng neuron giảm đi phân nửa. Trong một mạng CNN có nhiều Feature Map nên mỗi Feature Map chúng ta sẽ cho mỗi Max Pooling khác nhau. Chúng ta có thể thấy rằng Max Pooling là cách hỏi xem trong các đặc trưng này thì đặc trưng nào là đặc trưng nhất.

Cuối cùng ta đặt tất cả các lớp lại với nhau thành một CNN với đầu ra gồm các neuron với số lượng tùy bài toán.

2.3.4. Cánh chọn tham số cho CNN

Hiệu quả hoạt động của mạng CNN phụ thuộc rất nhiều vào việc lựa chọn các tham số sau:

- Số các convolution layer: càng nhiều các convolution layer thì performance càng được cải thiện. Sau khoảng 3 hoặc 4 layer, các tác động được giảm một cách đáng kể
- Filter size: thường filter theo size 5×5 hoặc 3×3
- Pooling size: thường là 2×2 hoặc 4×4 cho ảnh đầu vào lớn

Trong thực tế, tùy vào ứng dụng cụ thể mà ta chọn các tham số khác nhau. Thông thường ta sẽ thực hiện nhiều lần việc train test để chọn ra được param tốt nhất.

(Phương pháp thử sai.)

CHƯƠNG III: TÌM HIỂU VỀ BỆNH VIÊM PHỔI PNEUMONIA

3.1. Giới thiệu

Viêm phổi (Pneumonia) là hiện tượng viêm nhiễm của nhu mô phổi bao gồm viêm phế nang, túi phế nang, ống phế nang, tổ chức liên kết khe kẽ và viêm tiểu phế quản tận cùng. Các phế nang, đường dẫn khí trong phổi có thể chứa đầy chất lỏng hoặc mủ, gây ho có đờm hoặc mủ, sốt, ớn lạnh và khó thở. Thông thường, có nhiều tác nhân gây ra tình trạng viêm phổi, nhưng thường do vi khuẩn, virus và nấm. Bệnh viêm phổi có nhiều mức độ khác nhau từ viêm nhẹ, đến viêm phổi nặng. Bệnh đặc biệt nguy hiểm đối với trẻ sơ sinh, trẻ nhỏ và người trên 65 tuổi có nhiều bệnh nền, hoặc hệ miễn dịch suy yếu.

Viêm phổi có thể có mức độ nghiêm trọng từ nhẹ đến đe dọa đến tính mạng. Viêm phổi nghiêm trọng nhất đối với trẻ sơ sinh và trẻ nhỏ, những người trên 65 tuổi, và những người có vấn đề về sức khỏe hoặc hệ miễn dịch suy yếu.

3.2. Các nguyên nhân gây nên bệnh viêm phổi

Viêm phổi có thể được phân loại theo nhiều cách khác nhau tùy theo nguyên nhân và mục đích phân loại, nhưng nhìn chung hiện nay viêm phổi chủ yếu được phân loại dựa trên nguyên nhân gây viêm phổi và nguồn lây nhiễm bệnh.

3.2.1. Phân loại dựa trên nguyên nhân gây viêm

Viêm phổi do vi khuẩn

Vi khuẩn là nguyên nhân của hầu hết các trường hợp viêm phổi ở người trưởng thành. Viêm phổi do vi khuẩn thường lây truyền qua đường giọt bắn khi người khỏe mạnh hít, nuốt phải các giọt chứa đầy vi khuẩn sau khi người mắc bệnh ho hoặc hắt hơi. Những người có hệ miễn dịch suy yếu, những người có các bệnh lý nền mạn tính sẽ dễ bị viêm phổi do vi khuẩn hơn người bình thường.

Viêm phổi do virus

Viêm phổi do virus thường không nghiêm trọng và kéo dài trong một thời gian ngắn. Tuy nhiên, virus cúm có thể khiến viêm phổi trở nên nghiêm trọng và gây tử vong. Phụ nữ mang thai hoặc bệnh nhân mắc bệnh tim hay phổi nên cẩn thận với bệnh này. Diễn hình là viêm phổi do virus SARS-CoV-2.

Viêm phổi do nấm

Loại viêm phổi này do hít phải các bào tử của nấm, hay gặp ở những người có vấn đề sức khỏe mạn tính hoặc hệ thống miễn dịch suy yếu. Viêm phổi do nấm phát triển rất nhanh, các bào tử nấm khi hít phải bám vào phổi. Người hút thuốc lá, sinh sống ở môi trường bụi bẩn, ẩm mốc, dễ mắc bệnh viêm phổi do nấm.

Viêm phổi do hoá chất

Viêm phổi do hóa chất hay còn được gọi là viêm phổi hít. Đây là bệnh viêm phổi rất ít gặp, nhưng mức độ nguy hiểm cao. Viêm phổi do hóa chất xảy ra với nhiều mức

độ khác nhau. Mức độ nặng của tình trạng sẽ phụ thuộc vào: loại hóa chất, thời gian phơi nhiễm, thể trạng người bệnh, các biện pháp sơ cứu đã thực hiện... Ngoài ra, các hóa chất gây viêm phổi còn có thể gây hại cho nhiều cơ quan khác.

3.2.2. Phân loại dựa trên nguyên nhân lây nhiễm bệnh

Viêm phổi bệnh viện

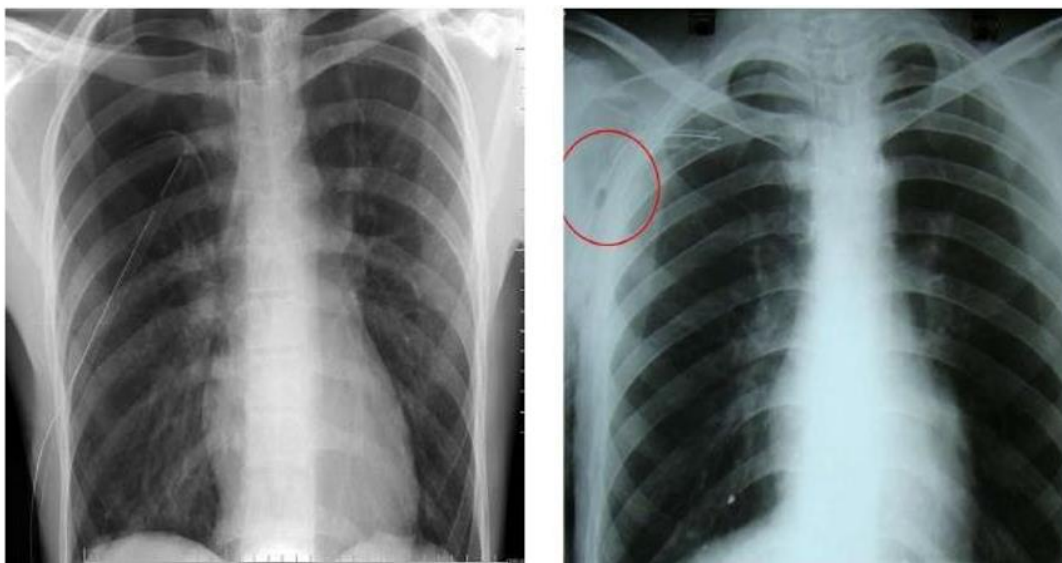
Là loại viêm phổi xuất hiện sau khi bệnh nhân nhập viện 48 giờ, mà trước đó bệnh nhân hoàn toàn không có các biểu hiện của viêm phổi. Viêm phổi bệnh viện là một vấn đề nghiêm trọng, bởi tác nhân gây ra viêm phổi bệnh viện có thể kháng với nhiều loại kháng sinh, khiến việc điều trị gặp nhiều khó khăn.

Các loại vi khuẩn gây ra căn bệnh này thường xâm nhập thông qua nhu phổi rồi dẫn vào trong máu. Đây chính là nguyên nhân gây ra tình trạng sốc nhiễm trùng, nhiễm trùng huyết, viêm màng não...

Viêm phổi cộng đồng

Viêm phổi mắc phải ở cộng đồng là tình trạng nhiễm khuẩn của nhu mô phổi xảy ra ở ngoài bệnh viện bao gồm viêm phế nang, tiểu phế quản tận hoặc viêm tổ chức kẽ của phổi. Viêm phổi được biểu hiện bằng viêm phổi thùy, viêm phổi đám hoặc viêm phổi không điển hình. Đặc điểm chung là có hội chứng đông đặc phổi và bóng mờ phế nang hoặc mô kẽ trên phim X-quang phổi. Bệnh do vi khuẩn, vi rút, nấm và một số tác nhân khác, nhưng không do trực khuẩn lao.

Viêm phổi có thể ở mức độ nghiêm trọng từ nhẹ đến đe dọa tính mạng. Nguyên nhân gây ra viêm phổi cộng đồng rất đa dạng, thông thường nhất là viêm phổi do vi khuẩn, virus. Sử dụng vắc - xin có thể giúp phòng tránh virus cúm và một số loại vi khuẩn nhất định có khả năng gây ra viêm phổi. Viêm phổi cộng đồng ở trẻ nhỏ còn có một loại đặc biệt hay gặp, đó là viêm phổi hít, xảy ra sau khi trẻ hít phải thức ăn, chất lỏng hoặc chất nôn vào trong phổi.



Hình 3.1 : Ảnh phổi bình thường (phải) và viêm phổi (trái)

3.3. Các triệu chứng của bệnh viêm phổi

Để kịp thời phát hiện vấn đề liên quan tới hệ hô hấp, cần nắm được một số triệu chứng đặc trưng của bệnh viêm phổi. Dựa vào tình trạng sức khỏe, mức độ nghiêm trọng của bệnh, các triệu chứng tương đối khác nhau.

Triệu chứng thường gặp

Đa số bệnh nhân đều trải qua một số triệu chứng, đó là sốt cao, những cơn ho kéo dài liên tục và kèm theo đờm. Họ rơi vào cảm giác mệt mỏi, cơ thể uể oải và cực kỳ khó chịu vì cơn ho liên tục làm phiền. Khi mắc bệnh, các túi khí ở phổi bị tắc nghẽn khiến người bệnh thiếu oxy và rất khó thở. Đặc biệt, ở bên phổi bị tổn thương bạn sẽ cảm thấy đau tức ngực.

Triệu chứng của người bệnh nặng

Khá nhiều bệnh nhân không may mắc phải bệnh viêm phổi ở mức độ nặng, tình trạng sức khỏe của bệnh nhân suy giảm cực kỳ nghiêm trọng. Với người bệnh nặng, triệu chứng sẽ là:

- Xuất hiện triệu chứng sốt cao, kèm theo đó là tình trạng nhức đầu tức ngực.
- Ho kéo dài và liên tục: Khi căn bệnh viêm phổi ngày càng chuyển biến nặng, tần suất các cơn ho sẽ xảy ra ngày càng thường xuyên hơn.
- Mệt mỏi, đau cơ: Những triệu chứng mà căn bệnh này sẽ khiến cho cơ thể trở nên mệt mỏi, uể oải, đau vùng cơ, đau khớp.
- Khó thở: Tình trạng khó thở thường xảy ra vào mỗi khi người bệnh hoạt động mạnh. Người bệnh luôn có cảm giác không thể hít đủ không khí.
- Đỏ nhiều mồ hôi, cơ thể bị ớn lạnh. Người bệnh cảm thấy cơ thể bị ớn lạnh dù cho đã được giữ ấm. Bên cạnh đó, rằng có thể bị đánh lộn cập vào với nhau.

3.4. Các phương pháp chuẩn đoán bệnh viêm phổi

Theo Trung tâm Kiểm soát và Phòng ngừa dịch bệnh Hoa Kỳ. Các xét nghiệm chẩn đoán viêm phổi sẽ được thực hiện sau khi qua những bước kiểm tra sơ bộ qua những biểu hiện bên ngoài. Tiếp theo đó, bác sĩ sẽ theo dõi tình trạng bệnh lý của người bệnh bằng những thiết bị y tế và chỉ định bệnh nhân thực hiện các xét nghiệm như:

X quang phổi

X quang ngực là xét nghiệm tốt nhất để chẩn đoán viêm phổi, giúp xác định mức độ và vị trí của nhiễm trùng, mức độ tổn thương.

Chụp cắt lớp vi tính (CT)

Xét nghiệm này cho thấy hình ảnh chi tiết hơn về phổi hoặc kiểm tra các biến chứng như áp xe phổi hoặc tràn dịch màng phổi. Đối tượng phù hợp với phương pháp này đó là người bị viêm phổi ở mức độ nặng, điều trị bệnh không đem lại hiệu quả cao.

Xét nghiệm máu

Xét nghiệm công thức máu, để kiểm tra hệ thống miễn dịch có đang chống lại nhiễm trùng hay không và cố gắng xác định loại sinh vật gây ra nhiễm trùng.

Xét nghiệm đờm

Bác sĩ có thể thu thập mẫu đờm từ việc ho hoặc dịch nhầy từ sâu trong phổi, gửi mẫu đến phòng thí nghiệm để kiểm tra.

Nuôi cấy dịch màng phổi

Đối với xét nghiệm này, bác sĩ sẽ lấy mẫu chất lỏng từ khoang màng phổi để kiểm tra vi khuẩn có thể gây viêm phổi.

Đo oxy xung

Xét nghiệm này sử dụng một cảm biến nhỏ được gắn vào ngón tay hoặc tai để ước tính lượng oxy trong máu. Viêm phổi có thể ngăn phổi di chuyển đủ oxy vào máu.

Nội soi phế quản

Đây là một thủ thuật xét nghiệm để nhìn vào bên trong đường thở của phổi. Đồng thời, thu thập các mẫu chất lỏng từ vị trí viêm phổi hoặc lấy sinh thiết mô phổi nhỏ để giúp tìm ra nguyên nhân gây viêm phổi.

3.5. Các phương pháp điều trị bệnh viêm phổi

Thuốc kháng sinh

Liệu pháp kháng sinh là phương pháp điều trị chính cho bệnh viêm phổi mắc phải tại cộng đồng. Điều trị thích hợp bao gồm bắt đầu dùng kháng sinh theo kinh nghiệm càng sớm càng tốt, thời gian tốt nhất là 8 giờ sau khi xuất hiện. Bởi vì sinh vật khó xác định, chế độ kháng sinh theo kinh nghiệm được lựa chọn dựa trên các mầm bệnh có thể xảy ra và mức độ nghiêm trọng của bệnh tật.

Điều trị bằng thuốc kháng vi-rút

Có thể được chỉ định cho một số trường hợp viêm phổi do virus. Nhắm vào các phần cụ thể của vi-rút để ngăn chặn nó sinh sôi trong cơ thể, giúp ngăn ngừa bệnh nghiêm trọng.

Chăm sóc hỗ trợ

Chăm sóc hỗ trợ bao gồm truyền dịch, thuốc hạ sốt, thuốc giảm đau, và cho những bệnh nhân bị giảm oxy máu, oxy. Dự phòng bệnh huyết khối tắc mạch và huyết động sớm cải thiện kết quả cho bệnh nhân viêm phổi nhập viện.

3.6. Các phương pháp phòng bệnh

Để phòng ngừa chứng viêm phổi, có thể thực hiện theo những cách sau:

- Giữ ấm cơ thể, tránh bị nhiễm lạnh và không đến gần người đang bị cảm cúm, đặc biệt là với trẻ em, người lớn tuổi, người có sức đề kháng yếu.
- Vệ sinh răng miệng và cổ họng sạch sẽ bằng nước muối sinh lý.

- Thường xuyên vệ sinh sạch sẽ nhà cửa, giữ gìn không khí trong lành.
- Tránh xa nơi có không khí ô nhiễm, khói thuốc lá hay các chất độc hại có thể gây kích thích phổi.
- Đeo khẩu trang khi đi ra ngoài và rửa tay thường xuyên bằng xà phòng.
- Tiêm phòng vacxin ngừa cúm: đặc biệt là các nhóm đối tượng: trẻ nhỏ, người già, người có bệnh lý đường hô hấp mạn tính: COPD, hen, giãn phế quản,....
- Duy trì chế độ ăn uống lành mạnh, cân bằng với đầy đủ các chất dinh dưỡng.
- Tập thể dục thường xuyên.

CHƯƠNG IV: XÂY DỰNG HỆ THỐNG NHẬN DẠNG BỆNH VIÊM PHỔI

4.1. Giới thiệu

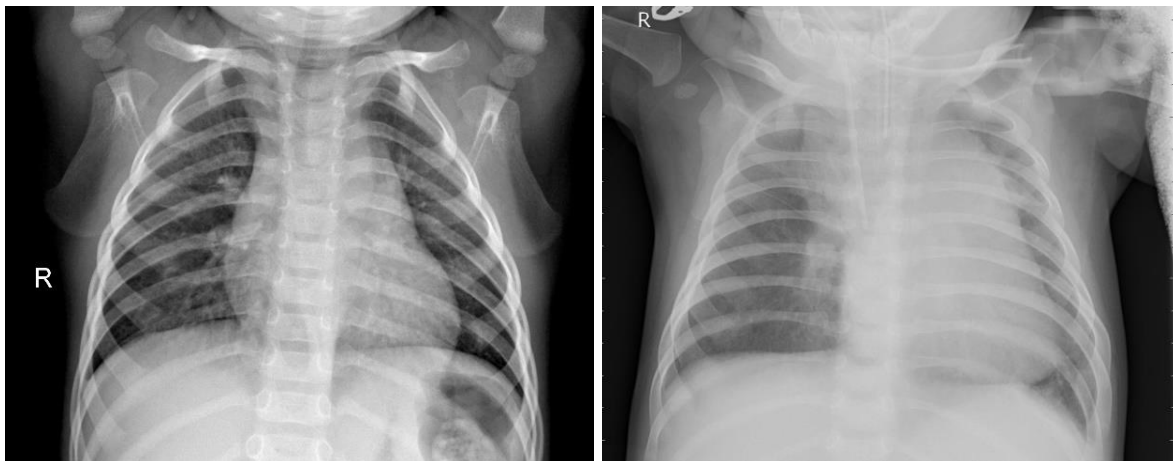
Hệ thống nhận dạng bệnh viêm phổi này sử dụng bộ dữ liệu gồm những hình ảnh X-quang về phổi được lấy trực tiếp từ Cơ sở dữ liệu Kaggle.

<https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>

Bộ dữ liệu được tổ chức thành 3 thư mục (train, test, val) và mỗi thư mục chứa các thư mục con cho từng loại hình ảnh (Pneumonia / Normal). Có 5.863 hình ảnh x-quang dưới định dạng ảnh JPEG của 2 loại hình ảnh (Pneumonia / Normal).

Hình ảnh X-quang ngực (trước-sau) được chọn từ nhóm hồi cứu bệnh nhi từ một đến năm tuổi từ Trung tâm Y tế Phụ nữ và Trẻ em Quảng Châu, Quảng Châu. Tất cả hình ảnh X-quang ngực được thực hiện như một phần của chăm sóc lâm sàng định kỳ cho bệnh nhân.

Đối với việc phân tích các hình ảnh X-quang phổi, tất cả các hình chụp X-quang phổi ban đầu đều được sàng lọc để kiểm soát chất lượng bằng cách loại bỏ tất cả các hình quét chất lượng thấp hoặc không thể đọc được. Các chẩn đoán hình ảnh sau đó được phân loại bởi hai bác sĩ chuyên môn trước khi được xóa để đào tạo hệ thống học máy. Để giải thích cho bất kỳ lỗi chấm điểm nào, bộ đánh giá cũng đã được kiểm tra bởi một chuyên gia thứ ba.



Hình 4.1: Ảnh phổi bình thường (phải) và ảnh viêm phổi (trái) trong tập dữ liệu

4.2. Thực nghiệm

4.2.1. Cài đặt các thư viện cần thiết

```
import cv2
import os
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import keras
```

```

from keras.models import Sequential
from keras.layers import Dense, Conv2D, MaxPool2D, Flatten,
    Dropout, BatchNormalization
from keras.preprocessing.image import ImageDataGenerator
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report,
    confusion_matrix
from keras.callbacks import ReduceLROnPlateau
from tensorflow.keras.utils import to_categorical, plot_model

```

4.2.2. Thu thập bộ dữ liệu

```

! pip install -q kaggle
# Chọn file kaggle.json
from google.colab import files
files.upload()
!rm -r ~/.kaggle
!mkdir ~/.kaggle
!mv ./kaggle.json ~/.kaggle/
!chmod 600 ~/.kaggle/kaggle.json
!kaggle datasets list
!kaggle datasets download -d paultimothymooney/chest-xray-
pneumonia

#Giải nén dữ liệu
!unzip -u "/content/chest-xray-pneumonia.zip" -d "/content/"

#Xây dựng hàm gán nhãn cho từng bộ dữ liệu
labels = ['PNEUMONIA', 'NORMAL']
img_size = 150
def get_training_data(data_dir):
    data = []
    for label in labels:
        path = os.path.join(data_dir, label)
        class_num = labels.index(label)
        for img in os.listdir(path):
            try:
                img_arr = cv2.imread(os.path.join(path, img),
                    cv2.IMREAD_GRAYSCALE)
                resized_arr = cv2.resize(img_arr, (img_size, i
mg_size))

                data.append([resized_arr, class_num])
            except Exception as e:
                print(e)
    return np.array(data)

#Tải từng bộ dữ liệu, kết hợp với hàm get_training_data
train = get_training_data('/content/chest_xray/train')
test = get_training_data('/content/chest_xray/test')
val = get_training_data('/content/chest_xray/val')

# Thống kê số lượng của từng bộ dữ liệu

```



```

print("The number of Train data: ",len(train))
print("The number of Test data: ",len(test))
print("The number of Val data: ",len(val))
print("The number of DATASET", len(train)+len(test)+len(val))

```

4.2.3. Tiền xử lý dữ liệu

Chuẩn hoá dữ liệu về thang độ xám có giá trị pixel nằm trong khoảng từ 0 đến 255. Ngoài ra, những hình ảnh này có kích thước không được đồng nhất. Do đó cần xử lý trước dữ liệu trước khi đưa nó vào mô hình.

Bước đầu tiên, thực hiện chuẩn hóa dữ liệu về thang độ xám để giảm ảnh hưởng của sự khác biệt về độ chiếu sáng. Hơn nữa, các mô hình máy học xử lý dữ liệu trên $[0...1]$ nhanh hơn so với trên $[0...255]$.

Tiếp đến, chuyển đổi tất cả hình ảnh của bộ dữ liệu về kích thước 150×150 . Và sao đó định hình lại thành 1 ma trận 1 chiều có dạng $150 \times 150 \times 1$ cho bộ dữ liệu.

```

# Xem biểu đồ thống kê số lượng của tập dữ liệu train

```

```

l = []
for i in train:
    if(i[1] == 0):
        l.append("Pneumonia")
    else:
        l.append("Normal")
sns.set_style('darkgrid')
sns.countplot(l)

```

```

# Xem trước hình ảnh của cả hai lớp.

```

```

plt.figure(figsize = (5,5))
plt.imshow(train[0][0], cmap='gray')
plt.title(labels[train[0][1]])

plt.figure(figsize = (5,5))
plt.imshow(train[-1][0], cmap='gray')
plt.title(labels[train[-1][1]])

```

```

x_train = []
y_train = []
x_val = []
y_val = []
x_test = []
y_test = []

```

```

for feature, label in train:
    x_train.append(feature)
    y_train.append(label)

```

```

for feature, label in test:
    x_test.append(feature)
    y_test.append(label)

```

```

for feature, label in val:
    x_val.append(feature)
    y_val.append(label)

# Chuẩn hoá dữ liệu
x_train = np.array(x_train) / 255
x_val = np.array(x_val) / 255
x_test = np.array(x_test) / 255

# Thay đổi kích thước dữ liệu
x_train = x_train.reshape(-1, img_size, img_size, 1)
y_train = np.array(y_train)

x_val = x_val.reshape(-1, img_size, img_size, 1)
y_val = np.array(y_val)

x_test = x_test.reshape(-1, img_size, img_size, 1)
y_test = np.array(y_test)

```

4.2.4. Tăng cường dữ liệu

Như đã trình bày trước đó, bộ dữ liệu được phân phối thành hai phần, tức là huấn luyện và kiểm tra, trong đó dữ liệu được sử dụng để huấn luyện sẽ không bao giờ được sử dụng để kiểm tra và ngược lại.

Từ bộ dữ liệu, nhận thấy rằng dữ liệu không được cân đối, tức là gần 75% hình ảnh minh họa bệnh viêm phổi trong khi 25% còn lại thể hiện hình ảnh bình thường. Đây được coi là một vấn đề nghiêm trọng được gọi là sự mất cân bằng dữ liệu, để khắc phục điều này, ta sử dụng dữ liệu nâng cao cho việc lấy mẫu lại tập dữ liệu. Tăng cường dữ liệu đã giúp tạo ra gần như hai hình ảnh từ một hình ảnh thông qua các phương pháp tăng cường khác nhau và do đó, hỗ trợ cân bằng tập dữ liệu và tránh vấn đề quá khớp.

Hơn nữa, các kỹ thuật tăng dữ liệu được sử dụng bao gồm phép biến đổi hình học được thể hiện trong bảng dưới đây.

Cấu hình cài đặt cho tăng cường dữ liệu	Giá trị tham số
Rescale	1/255
Rotation range	30°
Width shift range	0.1
Height shift range	0.1
Zoom_range	0.2
Horizontal flip	True
Vertical flip	False

Bảng 4.1: Bảng mô tả cấu hình cho tăng cường dữ liệu

Do bộ dữ liệu mất cân bằng nên ta áp dụng các kỹ thuật tăng cường dữ liệu

```
datagen = ImageDataGenerator(  
    featurewise_center=False,  
    samplewise_center=False,  
    featurewise_std_normalization=False,  
    samplewise_std_normalization=False,  
    zca_whitening=False,  
    rotation_range = 30,  
    zoom_range = 0.2,  
    width_shift_range=0.1,  
    height_shift_range=0.1,  
    horizontal_flip = True,  
    vertical_flip=False)  
datagen.fit(x_train)
```

4.2.5. Cấu hình cho mạng CNN

```
model = Sequential()  
model.add(Conv2D(32 , (3,3) , strides = 1 , padding = 'same' ,  
    activation = 'relu' ,  
    input_shape = (150,150,1)))  
model.add(BatchNormalization())  
model.add(MaxPool2D((2,2) , strides = 2 , padding = 'same'))  
model.add(Conv2D(64 , (3,3) , strides = 1 , padding = 'same' ,  
    activation = 'relu'))  
model.add(Dropout(0.1))  
model.add(BatchNormalization())  
model.add(MaxPool2D((2,2) , strides = 2 , padding = 'same'))  
model.add(Conv2D(64 , (3,3) , strides = 1 , padding = 'same' ,  
    activation = 'relu'))  
model.add(BatchNormalization())  
model.add(MaxPool2D((2,2) , strides = 2 , padding = 'same'))  
model.add(Conv2D(128 , (3,3) , strides = 1 , padding = 'same' ,  
    activation = 'relu'))  
model.add(Dropout(0.2))  
model.add(BatchNormalization())  
model.add(MaxPool2D((2,2) , strides = 2 , padding = 'same'))  
model.add(Conv2D(256 , (3,3) , strides = 1 , padding = 'same' ,  
    activation = 'relu'))  
model.add(Dropout(0.2))  
model.add(BatchNormalization())  
model.add(MaxPool2D((2,2) , strides = 2 , padding = 'same'))  
model.add(Flatten())  
model.add(Dense(units = 128 , activation = 'relu'))  
model.add(Dropout(0.2))  
model.add(Dense(units = 1 , activation = 'sigmoid'))  
model.compile(optimizer = "rmsprop" , loss = 'binary_crossentropy' , metrics = ['accuracy'])
```

Tóm tắt, hiển thị một số tham số (trọng số và độ lệch) trong mỗi lớp

```
model.summary()
```

```

#Điều chỉnh tỉ lệ học
learning_rate_reduction = ReduceLRonPlateau(monitor='val_accuracy', patience = 2, verbose=1, factor=0.3, min_lr=0.000001)

#Đưa dữ liệu vào mô hình học máy
history = model.fit(datagen.flow(x_train,y_train, batch_size = 32) ,epochs = 12, validation_data = datagen.flow(x_val, y_val) ,callbacks = [learning_rate_reduction])

#Xem, độ chính xác, độ mất mát của việc huấn luyện mô hình
print("Loss of the model is - " ,
      model.evaluate(x_test,y_test)[0])
print("Accuracy of the model is - " ,
      model.evaluate(x_test,y_test)[1]*100 , "%")
      Xem tóm tắt mô hình
plot_model(model, to_file='cnn.png', show_shapes=True)

# Lưu mô hình
model.save('model.h5')
#model = load_model('model.h5')

```

4.2.6. Phân tích mô hình

```

epochs = [i for i in range(12)]
fig , ax = plt.subplots(1,2)
train_acc = history.history['accuracy']
train_loss = history.history['loss']
val_acc = history.history['val_accuracy']
val_loss = history.history['val_loss']
fig.set_size_inches(20,10)

ax[0].plot(epochs , train_acc , 'go-
' , label = 'Training Accuracy')
ax[0].plot(epochs , val_acc , 'ro-
' , label = 'Validation Accuracy')
ax[0].set_title('Training & Validation Accuracy')
ax[0].legend()
ax[0].set_xlabel("Epochs")
ax[0].set_ylabel("Accuracy")

ax[1].plot(epochs , train_loss , 'g-
o' , label = 'Training Loss')
ax[1].plot(epochs , val_loss , 'r-
o' , label = 'Validation Loss')
ax[1].set_title('Testing Accuracy & Loss')
ax[1].legend()
ax[1].set_xlabel("Epochs")
ax[1].set_ylabel("Training & Validation Loss")
plt.show()

# Dự đoán mô hình
predictions = (model.predict(x_test) > 0.5).astype("int32")
predictions = predictions.reshape(1,-1)[0]

```

```

predictions[95:115]
arr_predictions = predictions[95:115]

for i in range(len(arr_predictions)):
    if arr_predictions[i] == 0:
        print("Image-{}: Pneumonia".format(i))
    else:
        print("Image-{}: Normal".format(i))

# Đánh giá mô hình

print(classification_report(y_test, predictions, target_names
= ['Pneumonia (Class 0)', 'Normal (Class 1)']))

# Xây dựng ma trận hỗn loạn
cm = confusion_matrix(y_test, predictions)
cm = pd.DataFrame(cm , index = ['0', '1'] , columns = ['0', '1']
)

# Trực quan hoá ma trận hỗn loạn
plt.figure(figsize = (10,10))
sns.heatmap(cm, cmap= "Blues", linecolor = 'black' ,
            linewidth = 1 , annot = True,
            fmt='', xticklabels = labels, yticklabels = labels)

```

CHƯƠNG V: KẾT LUẬN

Qua quá trình tìm hiểu, nghiên cứu về các kỹ thuật phân đoạn ảnh và trích suất đặc trưng, tìm hiểu sơ lược về bệnh viêm phổi. Đồ án đã trình bày được các vấn đề như sau:

5.1. Kết quả đạt được

Đồ án nghiên cứu đã trình bày được các kỹ thuật phân đoạn và trích suất đặc trưng của hình ảnh.

Nghiên cứu lý thuyết chung về mô hình mạng thần kinh tích chập CNN.

Tìm hiểu lý thuyết về bệnh viêm phổi, các nguyên nhân và phương pháp phòng bệnh. được những công nghệ có thể áp dụng cho y khoa, cụ thể trong bài này là những công nghệ áp dụng cho nhận dạng bệnh viêm phổi.

Xây dựng được một hệ thống minh hoạ ứng dụng mạng CNN nhận dạng bệnh viêm phổi với độ chính xác khá cao.

5.2. Hạn chế

Do thời gian thực hiện đồ án tương đối hạn chế nên không thể tránh được những thiếu sót nhất định.

Hệ thống nhận dạng chưa được tối ưu hóa về mặt giao diện, vì vậy chưa phù hợp để triển khai cho người dùng.

Ngoài ra độ chính xác nhận dạng là 92.6% là khá cao, nhưng để áp dụng vào thực tế thì cần cải thiện mô hình nhận dạng để nâng cao tỉ lệ hơn nữa.

Chưa thể tích hợp các phương pháp trích suất đặc trưng đã được trình bày vào hệ thống nhận dạng bệnh viêm phổi.

Chưa nghiên cứu sâu về đề tài, khả năng thu thập, lọc thông tin còn kém.

Hạn chế về vốn hiểu biết Tiếng Anh, dẫn đến hạn chế về việc sử dụng các tài liệu nước ngoài.

5.3. Hướng phát triển

Tiếp tục nghiên cứu và tìm hiểu về lĩnh vực xử lý ảnh, các phương pháp trích chọn đặc trưng khác, phân tích ưu nhược điểm của chúng.

Xây dựng giao diện cho mô hình để người dùng dễ dàng sử dụng.

Xây dựng một thuật toán mạnh mẽ để nhận dạng được chính xác hơn.

Nghiên cứu ứng dụng cho bài toán nhận dạng, phân loại các bệnh về y học. Các tập dữ liệu lớn hơn, các thuật toán khác của học máy cần được thử nghiệm và đánh giá để có cái nhìn toàn diện hơn về các kết quả đạt được.

CHƯƠNG VI: TÀI LIỆU THAM KHẢO

- [1]. TS Đỗ Năng Toàn, TS Phạm Việt Bình (2007). Giáo trình Xử lý ảnh. Khoa Công nghệ Thông Tin, Đại học Thái Nguyên.
- [2]. Nguyễn Thị Hoàn (2010). Phương pháp trích chọn đặc trưng ảnh trong thuật toán học máy tìm kiếm ảnh áp dụng vào bài toán tìm kiếm sản phẩm. Đồ án tốt nghiệp, Đại Học Quốc Gia Hà Nội, Trường Đại Học Công Nghệ.
- [3]. Bùi Thị Thúy Nga (2011). Tìm hiểu một số phương pháp trích chọn đặc trưng và ứng dụng tra cứu ảnh dựa trên nội dung. Đồ án tốt nghiệp, Trường Đại học Dân lập Hải Phòng.
- [4]. Ngô Huy Chương (2012). Tìm hiểu phương pháp phân đoạn ảnh y học. Đồ án tốt nghiệp, Trường Đại học Dân lập Hải Phòng.
- [5]. Dương Anh Hùng (2013). "Nghiên cứu ứng dụng Eigenfaces và đặc trưng cục bộ LBP cho bài toán nhận dạng mặt người. Luận văn Thạc sĩ, Đại học Đà Nẵng
- [6]. Dương Văn Cường (2018). Nghiên cứu một số phương pháp trích chọn đặc trưng cho ảnh và ứng dụng trong bài toán phân loại trạng thái cảm xúc khuôn mặt. Luận văn Thạc sĩ, Đại học Thái Nguyên.
- [7]. Dalal, Navneet, and Bill Triggs (2005). "Histograms of oriented gradients for human detection." Computer Vision and Pattern Recognition (CVPR).