

TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI
VIỆN ĐIỆN TỬ-VIỄN THÔNG



BÁO CÁO BÀI TẬP LỚN

Đề tài:

**PHÂN LOẠI ẢNH UNG THƯ DẠ DÀY SỬ DỤNG
MACHINE LEARNING**

Giảng viên hướng dẫn: TS. NGUYỄN VIỆT DŨNG

Phó Thị Hà Anh	20140217	ĐT 06 K59
Trương Thanh Hiền	20141612	ĐT 05 K59
Ngô Bá Sơn	20143813	ĐT 03 K59

Hà nội, 06/201

LỜI MỞ ĐẦU

Ngày nay ung thư trong đường tiêu hóa là một trong những căn bệnh nguy hiểm trên toàn thế giới, tuy nhiên loại ung thư này có ít triệu chứng ở giai đoạn đầu. Chính vì vậy, việc cung cấp các công cụ tốt hơn để chẩn đoán có thể giúp việc phát hiện sớm và điều trị là vô cùng cần thiết. Các công nghệ xử lý hình ảnh y tế để chẩn đoán ung thư đã và đang được phát triển trong thế kỷ qua, đặc biệt là nội soi. Chúng cho phép thu được hình ảnh của các mô trên đường tiêu hóa với độ phân giải tốt. Các kỹ thuật để phân tích những hình ảnh này đã được phát triển và đã được sử dụng bởi hệ thống Chẩn đoán bằng máy tính (Computer Aided Diagnosis -CAD).

Trong hệ tiêu hóa, các kỹ thuật của hệ thống CAD đã cho phép các bác sĩ phân tích các hình ảnh nội soi. Tuy nhiên, sự nhận biết ung thư trong đường tiêu hóa là một vấn đề hết sức phức tạp, trong đó chỉ có các bác sĩ được đào tạo có tỷ lệ thành công cao. Một số giải pháp nhận dạng mẫu bao gồm vấn đề này đã được phát hiện và ứng dụng trong quá khứ. Trong đề tài này, chúng em đã sử dụng một mạng CNN (Convolutional neural network), từ mô hình mạng này và dữ liệu ảnh tạo nên một bộ phân loại Multiclass SVM.

Chúng em đã thực hiện qua bộ dữ liệu hình ảnh mà chúng em đã điều chỉnh tất cả các thông số có liên quan. Phương pháp của chúng em trình bày là sau khi áp dụng bước tiền xử lý biến đổi hình ảnh sang màu xám, giảm kích thước 50% thì sử dụng mạng neural này trích xuất ra đặc trưng (feature) của ảnh rồi đưa vào bộ phân loại Multiclass SVM. Chúng em đã đánh giá phương pháp và đưa ra được độ chính xác của bài tập này khoảng **85, 42%** khi chưa finetune, dựa trên chia tập ảnh random trong huấn luyện.

Cuối cùng, chúng em xin chân thành cảm ơn **TS. Nguyễn Việt Dũng** đã có những bài giảng bổ ích trong quá trình học tập và hướng dẫn để chúng em có thể hoàn thành đề tài, chúng em xin chúc thầy hạnh phúc và thành đạt trong cuộc sống.

MỤC LỤC

LỜI MỞ ĐẦU.....	2
MỤC LỤC.....	3
DANH MỤC HÌNH ẢNH.....	5
DANH MỤC BẢNG BIỂU.....	6
PHÂN CÔNG CÔNG VIỆC	7
Chương 1: Giới thiệu đề tài.....	8
1.1. Lý do chọn đề tài	8
1.2. Ảnh sử dụng và kỹ thuật nội soi (Endoscopy)	9
1.2.1 Khái niệm.....	9
1.2.2 Các kỹ thuật nội soi	10
1.2.3 Thị giác máy tính trong nội soi dạ dày	11
Chương 2: Tổng quan về đề tài nghiên cứu tương tự.....	13
2.1. Đề tài phân loại ảnh ung thư dạ dày sử dụng mô hình túi từ trực quan	13
2.1.1 Tổng quan đề tài	13
2.1.2 Thực hiện.....	13
2.1.3 Đánh giá	15
2.2. Xác định ung thư dạ dày sử dụng Random Forests trích xuất đặc trưng đưa vào mô hình túi từ trực quan	16
2.2.1 Tổng quan đề tài	16
2.2.2 Thực hiện.....	17
2.2.3 Đánh giá	18
Chương 3: Cơ sở lý thuyết của phương pháp.....	19
3.1. Giới thiệu về Machine Learning	19
3.2. Convolutional neural network (CNN)	20
3.2.1 Giới thiệu về mạng CNN	20
3.2.2 Mạng Resnet	26
3.3. Thuật toán SVM đa lớp (Multiclass SVM)	27
Chương 4: Phương pháp thực hiện	29
4.1. Hệ thống thực hiện.....	29
4.2. Dataset.....	29

4.3. Sơ đồ thực hiện	29
4.4. Các bước phân loại hình ảnh	31
4.4.1. Bước 1: Tiền xử lý.....	31
4.4.2. Bước 2: Sử dụng mạng Resnet-50 như một bộ trích xuất feature để huấn luyện bộ phân loại hình ảnh.....	33
4.4.3. Bước 3: Huấn luyện bộ phân loại SVM đa lớp sử dụng feature trích xuất từ mạng CNN 35	
4.4.4. Bước 4: Kiểm định bộ phân loại	35
KẾT QUẢ VÀ ĐÁNH GIÁ	36
TÀI LIỆU THAM KHẢO	41
PHỤ LỤC.....	42

DANH MỤC HÌNH ẢNH

Hình 1. 1 Minh họa về ung thư dạ dày.....	9
Hình 1. 2 Hình ảnh minh họa 7 công nghệ nội soi.	11
Hình 1. 3 Quy trình thu và xử lý ảnh nội soi trong hệ thống CAD.	12
Hình 2. 1 Hình ảnh mô tả các loại mô khác nhau từ dạ dày đại diện cho bộ dữ liệu.....	13
Hình 3. 1 Các ứng dụng sử dụng Machine learning.	19
Hình 3. 2 Mô hình mạng truyền ngược.....	20
Hình 3. 3 Mô hình mạng Convolutional neural network.	21
Hình 3. 4 Minh họa về từ các noron đầu vào và việc tạo ra noron ẩn.	22
Hình 3. 5 Tạo noron ẩn đầu tiên của lớp ẩn thứ nhất.	22
Hình 3. 6 Tạo noron thứ hai của lớp ẩn thứ nhất.	23
Hình 3. 7 Các feature map.....	24
Hình 3. 8 Minh họa việc lấy một đơn vị max-pooling.....	25
Hình 3. 9 Mô hình một mạng CNN đơn giản.....	25
Hình 3. 10 Minh họa Multiclass SVM.....	28
Hình 4. 1 Sơ đồ thực hiện.	30

DANH MỤC BẢNG BIỂU

Bảng 1. 1 Các kỹ thuật nội soi.....10

Bảng 2. 1 Kết quả kiểm định ở chế độ tối ưu.....15

PHÂN CÔNG CÔNG VIỆC

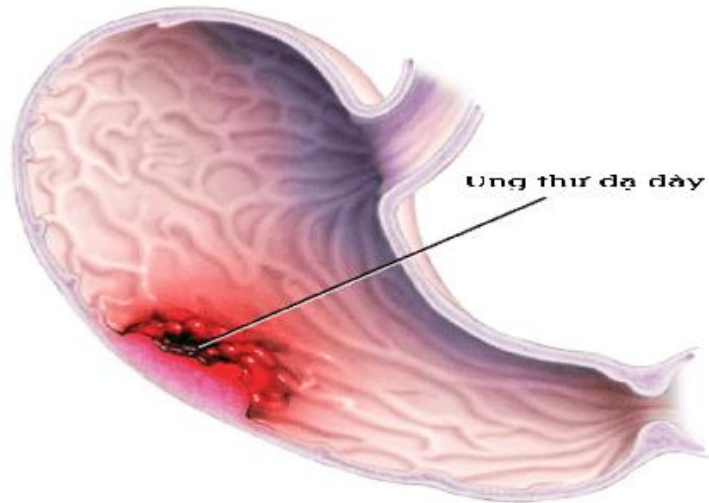
Thành viên	Công việc
Trương Thanh Hiền	Thực hiện phần code chính và thuyết trình
Phó Thị Hà Anh	Thực hiện phần code tiền xử lý
Ngô Bá Sơn	Làm báo cáo

Chương 1: Giới thiệu đề tài

1.1. Lý do chọn đề tài

Ung thư dạ dày là một trong những bệnh ung thư nghiêm trọng nhất, do tỷ lệ tử vong cao và sự phổ biến của nó. Ung thư có thể phát triển ở bất cứ phần nào của dạ dày, có thể lan ra khắp dạ dày và đến các cơ quan khác của cơ thể, đặc biệt là thực quản, phổi, hạch bạch huyết và gan. Ung thư dạ dày mỗi năm có thể gây ra 800.000 ca tử vong trên khắp thế giới.

Việc phát hiện ra ung thư dạ dày tương tự như với những loại ung thư khác, và nó không còn là một hoạt động y tế độc lập. Thật vậy, sự kết hợp giữa bí quyết chẩn đoán trong y tế với những tiến bộ có liên quan trong kỹ thuật ung thư, mới lạ và sáng tạo cho một đặc tính khách quan của mô. Kỹ thuật đó được tạo ra và phát triển bởi các kỹ sư y sinh. Mối quan hệ giữa y học và kỹ thuật đã trở nên gần gũi hơn trong vài năm qua, cho phép bác sĩ đi từ triệu chứng đơn giản (phân tích các dấu hiệu và triệu chứng bằng mắt thường), đến sử dụng các công cụ cung cấp sự đánh giá khách quan và tin cậy cho bệnh nhân, (ví dụ, nội soi, hoặc gần đây hơn là hệ thống chẩn đoán được hỗ trợ bởi máy tính CAD). Những giải pháp này mặc dù sáng tạo nhưng vẫn chưa đủ linh hoạt để giải quyết được các biến thể không xác định hoặc vô hình của hình thái khối u. Điểm bất lợi này thúc đẩy sự phát triển của một hệ thống phân tích hình ảnh ung thư dạ dày gắn với chiến lược nhận dạng ung thư, cho ra đánh giá đáng tin cậy trong trường hợp lâm sàng mới.



Hình 1. 1 Minh họa về ung thư dạ dày.

Mong muốn nghiên cứu và phát triển trong lĩnh vực Chẩn đoán hình ảnh đặc biệt là phương pháp học máy (Machine learning) và học sâu (Deep learning). Vì vậy, chúng em lựa chọn đề tài bài tập lớn: “Phân loại ảnh ung thư dạ dày sử dụng machine learning”.

1.2. Ảnh sử dụng và kỹ thuật nội soi (Endoscopy)

1.2.1 Khái niệm

Nội soi (Endoscopy) là một kỹ thuật y học hiện đại được ứng dụng trong việc khám bệnh, chẩn đoán bệnh bằng việc sử dụng các loại dụng cụ chuyên biệt để quan sát trực tiếp bên trong các cơ quan của cơ thể. Với kỹ thuật nội soi, người ta có thể quay phim, chụp hình bên trong các cơ quan, lấy dị vật, sinh thiết và thậm chí là thực hiện hiện phẫu thuật nội soi. Nội soi hiện nay được sử dụng trong hầu hết các chuyên khoa: tai mũi họng, tiêu hóa (thực quản, dạ dày, ruột già...) sản, ngoại, tiết niệu, xương khớp, thần kinh, thẩm mỹ...

Thiết bị nội soi là một thiết bị y tế có khả năng thu được hai hình ảnh đồng thời và thậm chí các mẫu mô từ đường tiêu hóa. Mặc dù có rất nhiều cấu hình nhưng các loại nội soi đều có cùng một số đặc điểm:

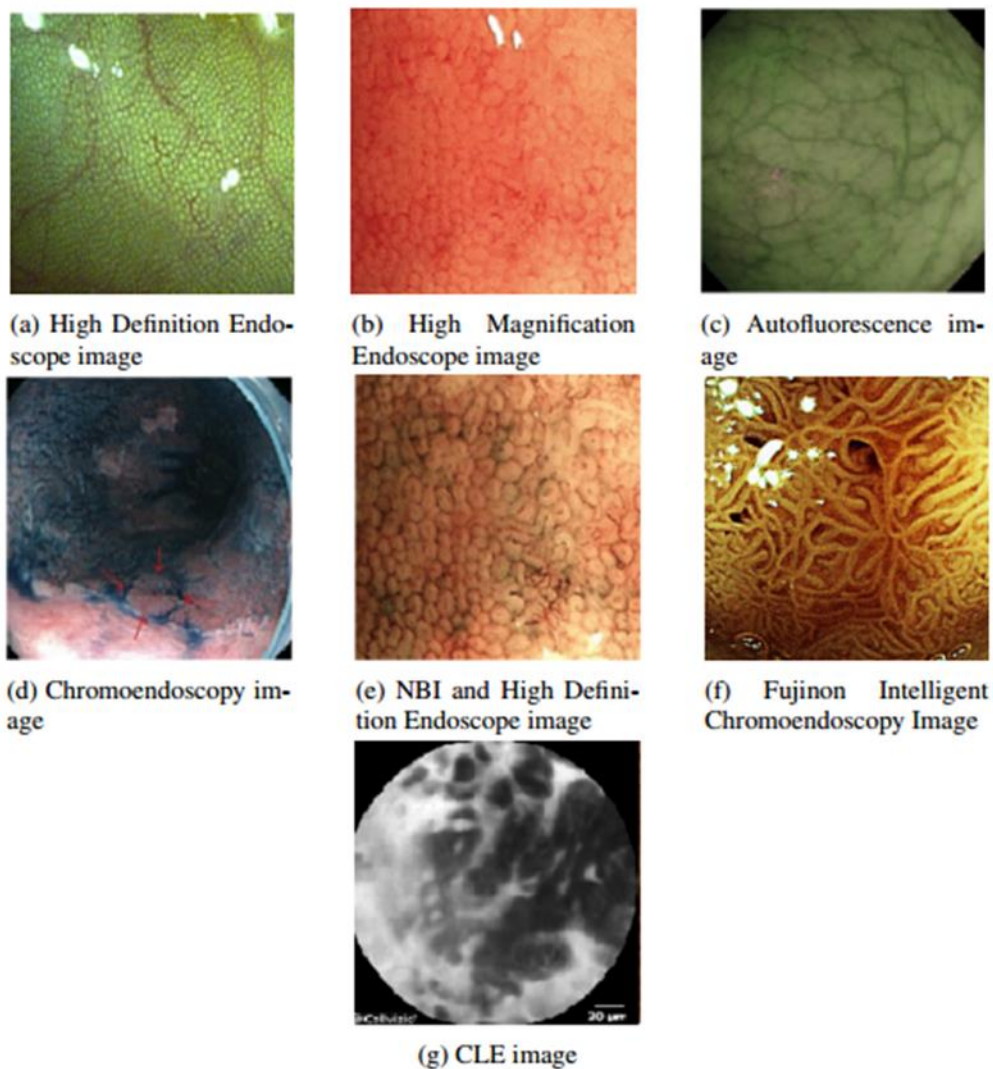
- Bao gồm một ống mềm dẻo.
- Có khả năng chiếu sáng và thu thập hình ảnh được kết hợp.

- Có một kênh phụ, có thể chụp lấy mẫu sinh thiết.

1.2.2 Các kỹ thuật nội soi

Bảng 1. 1 Các kỹ thuật nội soi.

CÔNG NGHỆ NỘI SOI	ỨNG DỤNG	MỤC ĐÍCH
Tiêu chuẩn và HD	Phát hiện những thay đổi tinh vi trong niêm mạc dễ dàng hơn	Tăng bề mặt nội soi và giảm các thành phần lạ trong ảnh
HME (Phóng đại cao)	Xác định hiện tượng tân mạch	Quan sát tốt chi tiết bề mặt nội soi và mạch máu.
AFI (Hình ảnh phát huỳnh quang tự động)	Xác định hoặc phát hiện sớm ung thư	Mô tả mô và phát hiện ung thư
NBI (Chụp hình với biên độ hẹp)	Xác định hiện tượng tân mạch kết đặc biệt kết hợp với nội soi phóng đại.	Tăng cường mô hình vi mô
I-scan và FICE	Xác định hiện tượng tân mạch kết đặc biệt kết hợp với nội soi phóng đại	Phát hiện, chẩn đoán và điều trị bệnh dạ dày ruột dễ dàng hơn
CLE (Chụp quang tuyến Laser)	Xác định hiện tượng tân mạch	Cung cấp hình ảnh cực nhỏ có độ phân giải cao



Hình 1. 2 Hình ảnh minh họa 7 công nghệ nội soi.

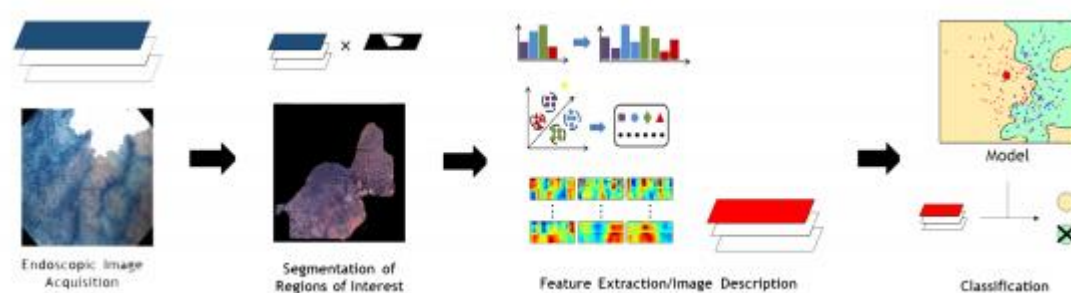
- a) Hình ảnh nội soi có độ phân giải cao của bề mặt niêm mạc bình thường
- b) Phóng đại cao của niêm mạc ruột bình thường
- c) AFI trong ruột kết bình thường
- d) Hình ảnh Chromoendoscope
- e) Hình ảnh NBI và HD
- f) Hình ảnh nội soi màu nâu nhạt của một tổn thương niêm mạc
- g) Hình ảnh mô bình thường sử dụng đầu dò CLE

1.2.3 Thị giác máy tính trong nội soi dạ dày

Việc thu thập lại hình ảnh của đường tiêu hóa bằng cách sử dụng đầu dò nội soi là một tiến trình dễ bị nhiễu. Bởi nó được thực hiện trong điều kiện không được kiểm soát, một số vấn đề liên quan đến chiếu sáng, xoay đầu dò, đổ bóng hoặc bị kẹt. Thị giác máy tính (CV) trong nội soi dạ dày phải đối phó với những bất ổn này cũng như với các mẫu hình ảnh yếu, cản trở khả năng nhận biết tổn thương.

Các công trình được phát triển trước đây trong lĩnh vực nghiên cứu này bao gồm một loạt các kỹ thuật mở rộng của CV. Trong lĩnh vực này nó sẽ sử dụng một thuật toán có khả năng trích xuất mạnh mẽ các đặc trưng liên quan. Mặc dù cần phải giải quyết một vấn đề y tế, nhưng nó cũng là một mô hình nhận dạng đầy thách thức. Một giải pháp khả thi có thể bao gồm các thuật toán xử lý ảnh mà chưa được thử trước đây trong lĩnh vực này.

Hình cho thấy tiến trình chung cho việc thu và xử lý hình ảnh nội soi trong một hệ thống CAD. Đặc biệt nhấn mạnh đến tính năng trích xuất và mô tả hình ảnh.



Hình 1. 3 Quy trình thu và xử lý ảnh nội soi trong hệ thống CAD.

Chương 2: Tổng quan về đề tài nghiên cứu tương tự

2.1. Đề tài phân loại ảnh ung thư dạ dày sử dụng mô hình túi từ trực quan

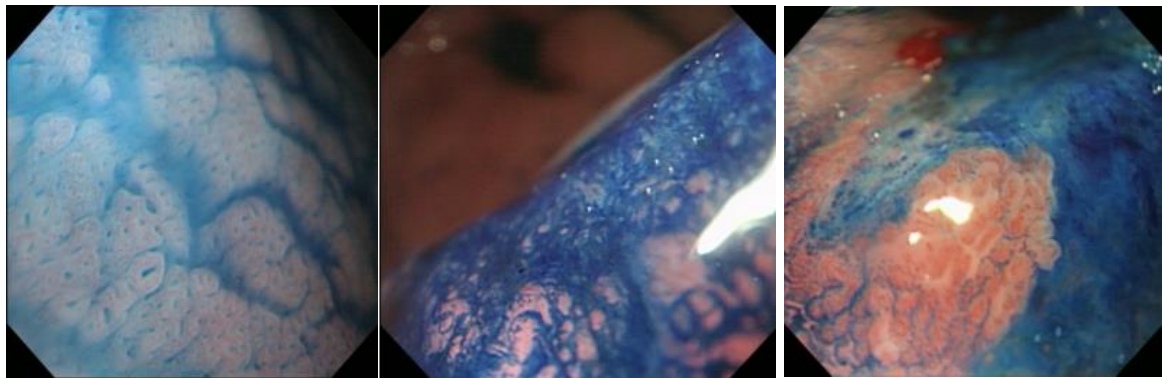
2.1.1 Tổng quan đề tài

Mô hình túi từ trực quan (Bag of words model) là một trong những phương pháp phổ biến nhất dành cho phân loại object (đặc biệt trong xử lý ảnh). Mô hình này học một bộ từ vựng từ của tất cả các văn bản, rồi mô hình hóa các văn bản bằng cách đếm số lần xuất hiện của mỗi từ trong văn bản đó. Từ đó mã hóa mỗi văn bản thành các vector.

2.1.2 Thực hiện

a) Dataset

Sử dụng tập dữ liệu hình ảnh nội soi “Chromoendoscopy” gồm 176 ảnh, mỗi ảnh có kích thước 518x481, trong đó có 56 ảnh mô bình thường, và 120 ảnh mô bất thường (96 ảnh mắc chứng metaplasia và 24 ảnh thuộc triệu chứng dysplasia) cùng với vùng ảnh quan tâm (Region Of Interest) tương ứng với mỗi ảnh, được xác định bởi bác sĩ.



a) Mô bình thường

b) Mô bị nhiễm dysplasi

c) Mô bị nhiễm metaplasia

Hình 2. 1 Hình ảnh mô tả các loại mô khác nhau từ dạ dày đại diện cho bộ dữ liệu.

b) Các bước thực hiện

Bước 1: Tiền xử lý

- Loại bỏ nhiễu và tách lấy vùng cần quan tâm ROI (ảnh xám) để thu được tập data mới chứa 120 ảnh.

- Tạo tập data mới ngẫu nhiên có 112 ảnh chia đều cho 2 lớp bình thường và bất thường.
- Chia tập data mới thành 2 tập ảnh trainingSets và testingSets với tỷ lệ 7:3.

Bước 2: Tạo túi từ trực quan (Túi từ tính năng – BoW)

- Trích xuất tính năng của từng ảnh trong trainingSets.
- Tối ưu hóa các tính năng.
- Tạo ra túi từ trực quan bằng phương pháp phân cụm K-means.

Bước 3: Đào tạo “Bộ phân loại” sử dụng hàm “trainImageCategoryClassifier” trong Matlab.

Bước 4: Kiểm định “Bộ phân loại” và tính tỷ lệ lỗi.

c) Các tham số

- Kích thước hình ảnh
- Tỷ lệ N = trainingSets: testingSets
- Kích thước Vocabulary (số cụm K trong thuật toán K-means)
- Hệ số StrongestFeatures
- Tham số PointSelection
- Tham số GridStep
- Độ lệch chuẩn của bộ lọc Gaussian

d) Kết quả

Bảng 2. 1 Kết quả kiểm định ở chế độ tối ưu.

<i>trainingSets</i>	<i>testingSets</i>
<p style="text-align: center;">PREDICTED</p> <p>KNOWN Abnormal_50 Normal_50</p> <p>-----</p> <p>Abnormal_50 0.92 0.08</p> <p>Normal_50 0.06 0.94</p> <p> * Độ chính xác trung bình: 93%</p>	<p style="text-align: center;">PREDICTED</p> <p>KNOWN Abnormal_50</p> <p>Normal_50</p> <p>-----</p> <p>Abnormal_50 0.94 0.06</p> <p>Normal_50 0.03 0.97</p> <p> * Độ chính xác trung bình: 95%</p>

2.1.3 Đánh giá

Phương pháp này đã đem lại hiệu suất cao với độ chính xác khoảng hơn 90% với tập dataset sử dụng, tương ứng với tỷ lệ dự đoán sai trong khoảng 6 ± 1 (dưới 10%), kết quả tương tự và có phần tốt hơn với kết quả của các bài báo đã được công bố như “Recognition of Cancer using Random Forests as a Bag-of-Words Approach for Gastroenterology” [2], “Superpixel-based classification of gastric chromoendoscopy images” [3].

Tuy nhiên phương pháp còn có hạn chế là việc xử lý ảnh và lấy các feature đặc trưng của ảnh còn là phương pháp thủ công.

2.2. Xác định ung thư dạ dày sử dụng Random Forests trích xuất đặc trưng đưa vào mô hình túi từ trực quan

2.2.1 Tổng quan đề tài

Mỗi năm, 2.8 triệu người được chẩn đoán mắc ung thư dạ dày và khoảng 1.8 triệu người chết. Các nghiên cứu gần đây đã chỉ ra rằng tỷ lệ sống sót là đặc biệt cao khi ung thư được chẩn đoán sớm. Điều này thúc đẩy việc tìm kiếm các phương pháp mới, khách quan và có tính định lượng để giúp đội ngũ y tế trong công việc này. Giải pháp thị giác máy tính (Computer Vision, CV) được áp dụng cho phân loại hình ảnh nội soi với tính hữu ích, thêm đánh giá một cách định lượng để cho các thiết lập tốt, chính xác và hạn chế tối đa việc xâm lấn. Tuy nhiên, việc thu thập hình ảnh nội soi bị ảnh hưởng bởi một số khó khăn mà sẽ khó nhận ra ung thư và làm giảm khả năng ứng dụng của các phương pháp CV.

Hiện nay, việc phân loại ảnh GE (ảnh dạ dày) đã được thực hiện bằng cách trích xuất các đặc trưng đặc biệt từ hình ảnh, hoặc trên miền tần số (ví dụ: bộ lọc Gabor) hoặc trong miền không gian (ví dụ: bộ mô tả SIFT), và cho nó vào mô hình phân loại sử dụng SVM hoặc kNN. Mặc dù trình bày kết quả đầy hứa hẹn, các phương pháp vẫn có thể được cải thiện thêm tính mô tả và độ chính xác. Ngoài ra, điều quan trọng là phải loại bỏ nhu cầu của người dùng điều chỉnh các thông số của thuật toán, làm định nghĩa tự động và thích ứng với các tập dữ liệu mới. Trong nghiên cứu này, họ sử dụng những tiến bộ gần đây trong CV để phân tích hình ảnh GE. Cách tiếp cận này thường được dùng để nhận ra đối tượng trong cảnh hỗn tạp bằng cách tạo biểu đồ số lần xuất hiện (từ vựng) của mỗi mẫu khác nhau (hình ảnh từ) trong ảnh. Một tập các hình ảnh từ này được gọi là BoW.

Các cách tiếp cận khác nhau có thể tạo ra BoW. Ở đây, họ đề xuất sử dụng RF để thực hiện phân cấp theo cường độ pixel đầu vào để tạo ra BoW, trong đó mỗi từ tương ứng với một mẫu riêng biệt và độc lập trên ảnh nội soi. RF là một tập các cây quyết định (DT). Họ lấy đầu vào dữ liệu và đưa nó vào một chuỗi các quyết định nhị phân. Chuỗi

này là cách tiếp cận lặp lại của các quyết định phụ phân vùng theo không gian đặc trưng. Với mỗi DT khác nhau các mẫu dữ liệu không được học một cách chính xác. Tuy nhiên, họ thêm ngẫu nhiên vào quá trình học và cải thiện xác suất để có mô hình đáng tin cậy hơn. RF được áp dụng để tạo BoW, tuy nhiên việc sử dụng chúng trong GE chưa được nghiên cứu. Cách tiếp cận tương tự đã được thử nghiệm trong GE, nhưng sử dụng SIFT và k-means để tạo BoW.

2.2.2 Thực hiện

Hình 2.2 mô tả quy trình phân tích và phân loại hình ảnh GE. Giai đoạn quan trọng của phương pháp bao gồm xây dựng vốn từ vựng của dữ liệu đầu vào, sử dụng BoW để tạo cường độ pixel đầu vào bằng RF.



Hình 2. 2 Phương pháp xây dựng mô hình phân loại cho ảnh GE.

a) Thu nhận và xử lý ảnh:

Thu hình ảnh nội soi dạ dày và vùng quan tâm ROI. Hình ảnh được thay đổi kích thước và làm mịn. Sau đó chia nhỏ ROI thành các phần nhỏ có chứa cường độ pixel của ảnh.

b) Giới thiệu trực quan

Tạo ra BoW sử dụng RF. Mỗi phần nhỏ đi xuống cây quyết định cho đến khi một nút lá được tạo ra. Lặp lại quá trình cho tất cả các cây, đếm số lần mỗi nút thuộc đường đi xuống, kết quả thu được từ vừng, từ đó xây dựng bộ phân loại.

c) Nhận dạng hình ảnh

Sử dụng SVM. Mô hình phân loại được tạo ra bởi các thông số học và các thông số đáng giá được định nghĩa bởi xác thực chéo 3-folds. Vì đang xử lý vấn đề đa lớp, với các giai đoạn ung thư khác nhau (mô thường, mô với metaplasia và mô với displasia), họ áp dụng chiến lược one-versus-all (OVA).

d) Kết quả

Họ nhận được lỗi trung bình $23.07 \pm 2.83\%$, so với $30.44 \pm 4.20\%$ khi sử dụng phương pháp chuẩn. Đặc biệt mỗi hình ảnh đầu vào mới cần khoảng 5 giây để được phân loại, cho phép gần với thời gian thực, phản hồi cho bác sĩ với độ tin cậy. Hơn nữa, họ loại bỏ sự khác biệt giữa metaplasia và displasia, đánh giá ung thư chỉ sử dụng mô bình thường và mô bệnh lý. Sai số trung bình thu được là $11.03 \pm 1.80\%$.

2.2.3 Đánh giá

Phương pháp có sự chính xác khá cao, đây là một hướng nghiên cứu có khả năng ứng dụng vào môi trường y tế: tạo chương trình giúp các bác sĩ giải quyết vấn đề phân loại và xử lý hình ảnh nội soi giúp cho việc chẩn đoán được nhanh và chính xác. Tuy nhiên độ chính xác lại không thể bằng phương pháp phía trước.

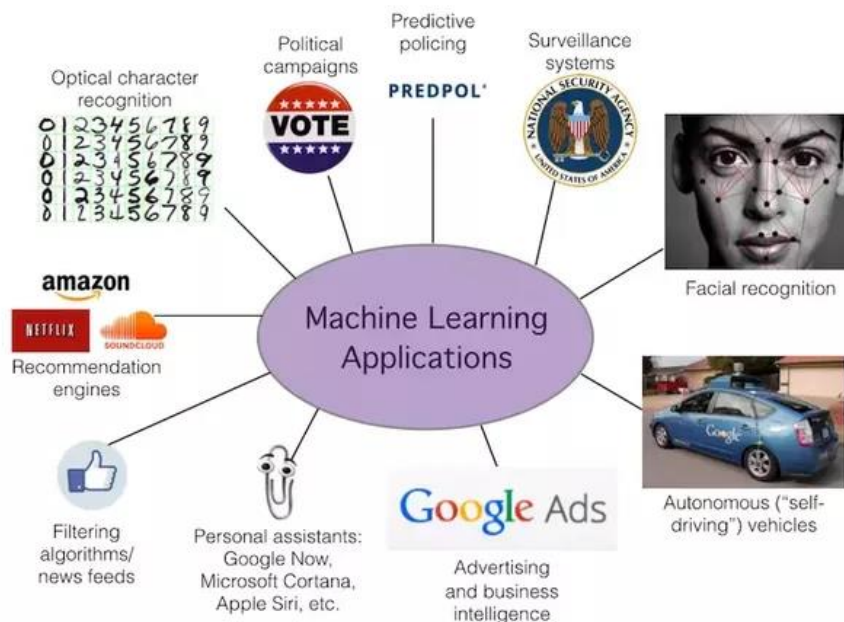
Chương 3: Cơ sở lý thuyết của phương pháp

3.1. Giới thiệu về Machine Learning

Machine Learning là một lĩnh vực của trí tuệ nhân tạo liên quan đến việc nghiên cứu và xây dựng các kỹ thuật cho phép các hệ thống "học" tự động từ dữ liệu để giải quyết những vấn đề cụ thể. Ví dụ như các máy có thể "học" cách phân loại thư điện tử xem có phải thư rác (spam) hay không và tự động xếp thư vào thư mục tương ứng. Học máy rất gần với suy diễn thống kê (statistical inference) tuy có khác nhau về thuật ngữ.

Học máy có liên quan lớn đến thống kê, vì cả hai lĩnh vực đều nghiên cứu việc phân tích dữ liệu, nhưng khác với thống kê, học máy tập trung vào sự phức tạp của các giải thuật trong việc thực thi tính toán. Nhiều bài toán suy luận được xếp vào loại bài toán NP-khó (nondeterministic polynomial time), vì thế một phần của học máy là nghiên cứu sự phát triển các giải thuật suy luận xấp xỉ mà có thể xử lý được.

Học máy hiện nay được áp dụng rộng rãi bao gồm máy truy tìm dữ liệu, chẩn đoán y khoa, phát hiện thẻ tín dụng giả, phân tích thị trường chứng khoán, phân loại các chuỗi DNA, nhận dạng tiếng nói và chữ viết, dịch tự động, chơi trò chơi và cử động robot (robot locomotion).



Hình 3. 1 Các ứng dụng sử dụng Machine learning.

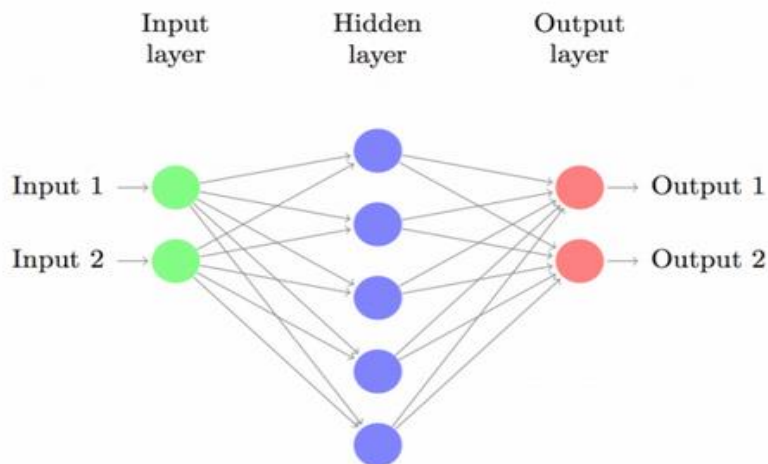
3.2. Convolutional neural network (CNN)

3.2.1 Giới thiệu về mạng CNN

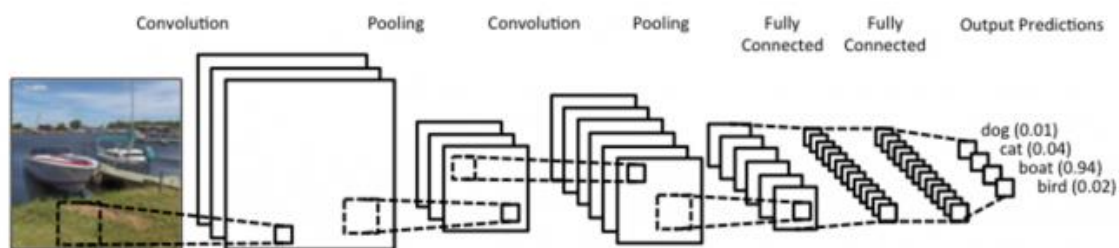
Mạng nơ-ron tích chập- CNN (convolutional neural network) là một trong những mô hình Deep Learning tiên tiến giúp cho chúng ta xây dựng được những hệ thống thông minh với độ chính xác cao như hiện nay như hệ thống xử lý ảnh lớn như Facebook, Google hay Amazon đã đưa vào sản phẩm của mình những chức năng thông minh như nhận diện khuôn mặt người dùng, phát triển xe hơi tự lái hay drone giao hàng tự động.

Mạng- là thuật ngữ để chỉ sự dùng một cấu trúc có các Node được liên kết với nhau bằng các cạnh, có một hoặc nhiều lớp kết nối với nhau, có đầu vào và đầu ra. Mạng CNN là một tập hợp các lớp Convolution chồng lên nhau và sử dụng các hàm nonlinear activation như ReLU và tanh để kích hoạt các trọng số trong các Node. Mỗi một lớp sau khi đi qua hàm kích hoạt sẽ tạo ra các thông tin trừu tượng hơn cho các lớp tiếp theo.

Trong mô hình mạng truyền ngược (feedforward neural network) thì mỗi nơ-ron đầu vào (input node) cho mỗi nơ-ron đầu ra trong các lớp tiếp theo. Mô hình này gọi là mạng kết nối đầy đủ (fully connected layer) hay mạng toàn vẹn (affine layer). Còn trong mô hình CNNs thì ngược lại. Các layer liên kết được với nhau thông qua cơ chế convolution. Layer tiếp theo là kết quả convolution từ layer trước đó, nhờ vậy mà ta có được các kết nối cục bộ. Như vậy mỗi nơ-ron ở lớp kế tiếp sinh ra từ kết quả của bộ lọc áp đặt lên một vùng ảnh cục bộ của nơ-ron trước đó.



Hình 3. 2 Mô hình mạng truyền ngược.



Hình 3. 3 Mô hình mạng Convolutional neural network.

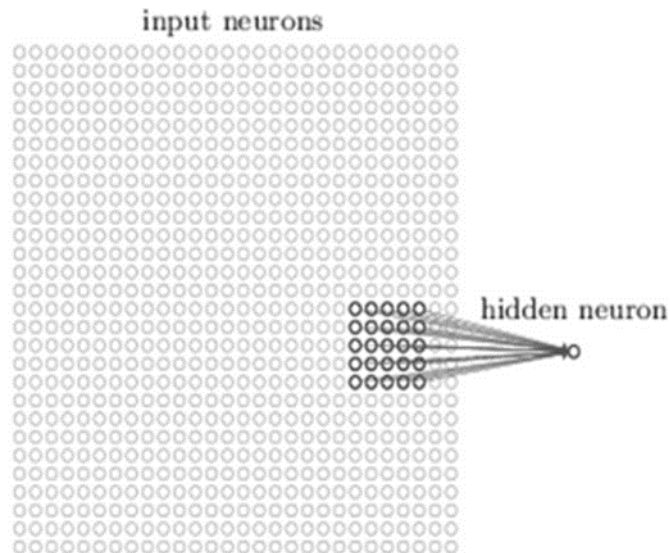
Mỗi một lớp được sử dụng các bộ lọc khác nhau. Thông thường có hàng trăm nghìn bộ lọc như vậy và kết hợp kết quả của chúng lại. Ngoài ra có một số layer khác như pooling/subsampling layer dùng để chắt lọc lại các thông tin hữu ích hơn (loại bỏ các thông tin nhiễu). Trong quá trình huấn luyện mạng (training) CNN tự động học các giá trị qua các lớp bộ lọc dựa vào cách thức mà bạn thực hiện. Ví dụ trong tác vụ phân lớp ảnh, CNNs sẽ cố gắng tìm ra thông số tối ưu cho các bộ lọc tương ứng theo thứ tự raw pixel > edges > shapes > facial > high-level features. Layer cuối cùng được dùng để phân lớp ảnh.

Trong mô hình CNN có 2 khía cạnh cần quan tâm là Location Invariance (tính bất biến) and Compositionality (tính kết hợp). Với cùng một đối tượng, nếu được chiếu theo các góc độ khác nhau (translation, rotation, scaling) thì độ chính xác của thuật toán sẽ bị ảnh hưởng đáng kể. Pooling layer sẽ cho bạn tính bất biến đối với phép dịch chuyển (translation), phép quay (rotation) và phép co giãn (scaling). Tính kết hợp cục bộ cho ta các cấp độ biểu diễn thông tin từ mức độ thấp đến mức độ cao và trừu tượng hơn thông qua convolution từ các bộ lọc. Đó là lý do tại sao CNNs cho ra mô hình với độ chính xác rất cao. Cũng giống như cách con người nhận biết các vật thể trong tự nhiên.

Mạng tích chập sử dụng 3 ý tưởng cơ bản: các trường tiếp nhận cục bộ (local receptive field), trọng số chia sẻ (shared weights) và tổng hợp (pooling).

Trường tiếp nhận cục bộ (local receptive field): đầu vào của mạng CNN là một ảnh. Ví dụ như 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. Trong mô hình mạng ANN truyền thống thì chúng ta sẽ kết nối các neuron đầu vào vào tầng ảnh. Tuy nhiên trong CNN chúng ta

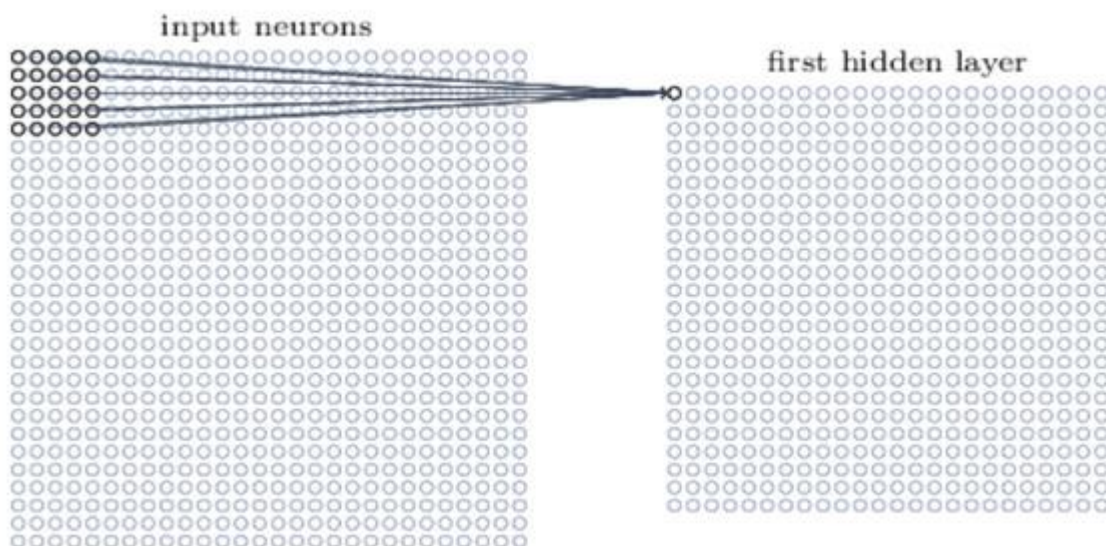
không làm như vậy mà chúng ta chỉ kết nối trong một vùng nhỏ của các noron đầu vào như một bộ lọc có kích thước 5x5 tương ứng 25 điểm ảnh đầu vào. Mỗi một kết nối sẽ học một trọng số và mỗi noron ẩn sẽ học một bias.



Hình 3. 4 Minh họa về từ các noron đầu vào và việc tạo ra noron ẩn.

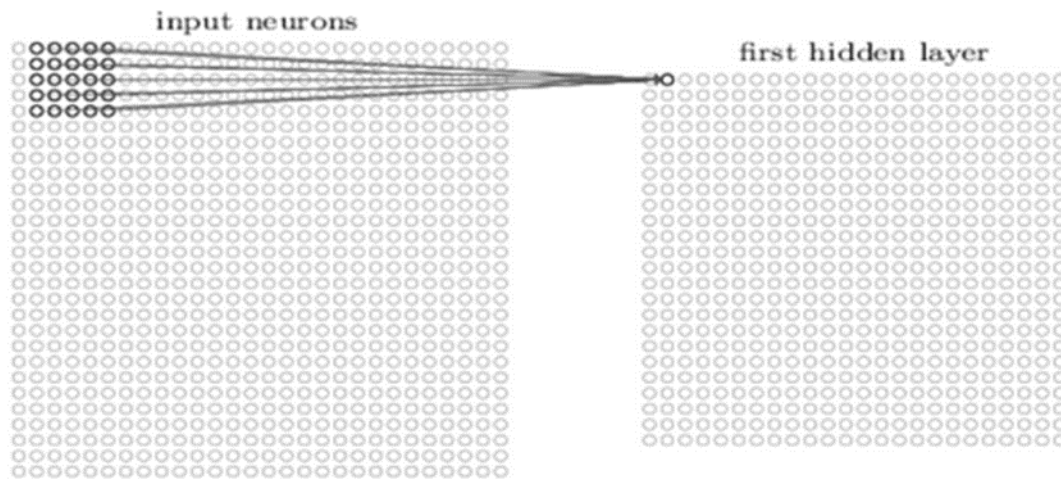
Tiếp tục bộ lọc sẽ trượt trên tám ảnh tương ứng với mỗi vùng trượt sẽ có một noron ẩn trong lớp ẩn đầu tiên.

- Bước 1: Tạo ra noron ẩn đầu tiên trong lớp ẩn thứ nhất



Hình 3. 5 Tạo noron ẩn đầu tiên của lớp ẩn thứ nhất.

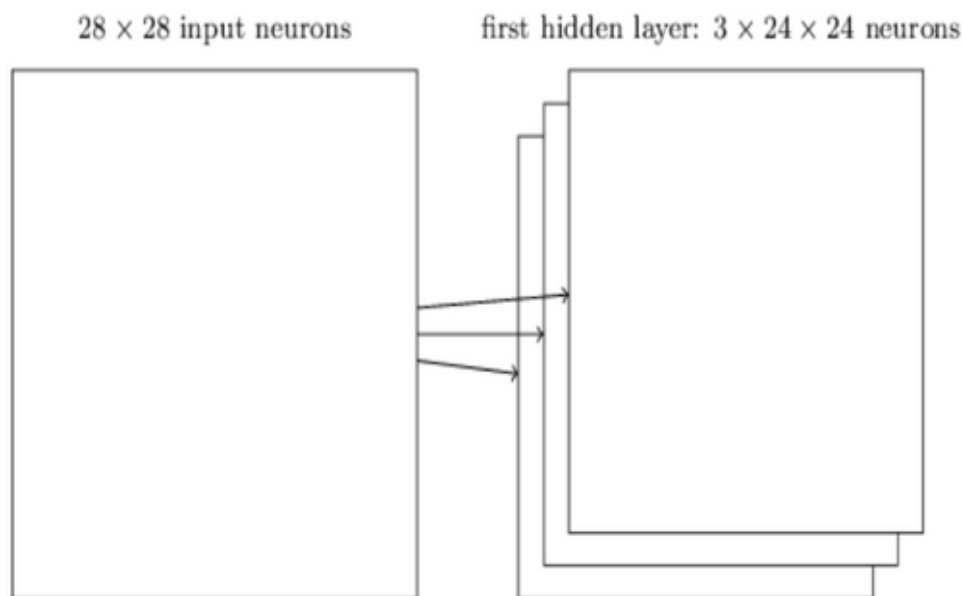
- Bước 2: Dịch bộ lọc qua bên phải một cột sẽ tạo được noron ẩn thứ hai.



Hình 3. 6 Tạo noron thứ hai của lớp ẩn thứ nhất.

Như vậy kết thúc nếu ta có ảnh đầu vào là 28x28 và filter có kích thước là 5x5 thì ta sẽ có bao nhiêu nút trong lớp ẩn. Kết quả là $28-5+1 = 24$ noron trong lớp ẩn đầu tiên do filter trượt đến dòng cuối chỉ có thể thực hiện 23 lần trước khi chạm đường biên. Đối với bài toán nhận dạng ảnh người 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.



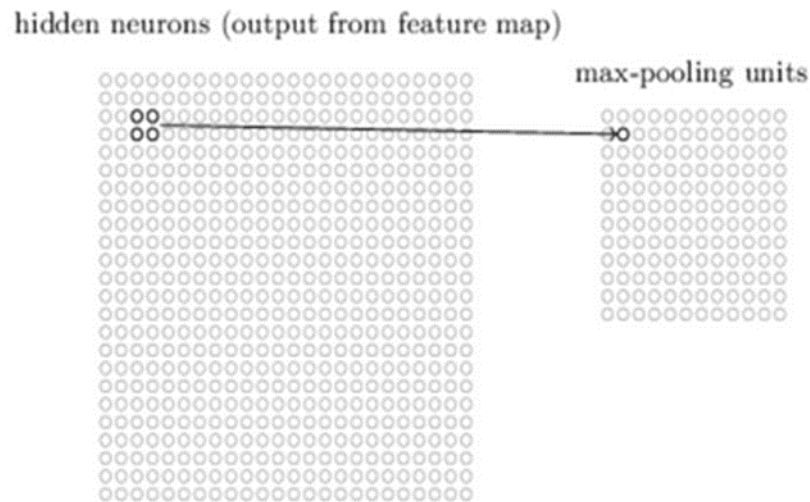
Hình 3. 7 Các feature map.

Như trong Hình 3.7 phía trên, một mạng có một ảnh đầu vào sẽ xây dựng 3 feature map. Như vậy để làm gì? Một shared weight và shared bias sẽ tạo một feature map. Như vậy ta tìm được 3 loại đặc trưng khác nhau với mỗi đặc trưng được phát hiện trên toàn bộ ảnh. Trong thực tế các bài toán sử dụng tìm được số lượng feature khác nhau như mạng LeNet sử dụng 6 bản đồ đặc trưng khác nhau.

Trọng số chia sẻ (shared weights): Mối quan hệ giữa số lượng Feature map với số lượng tham số là gì? Chúng ta thấy mỗi feature map cần $25 = 5 \times 5$ 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 lớp 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 nếu mô hình có số lượng tham số ít hơn thì nó sẽ chạy nhanh hơn.

Tổng hợp (pooling): lớp pooling thường được sử dụng ngay sau lớp convolutional để đơn giản hóa thông tin đầu ra để giảm bớt số lượng neuron. Thủ tục

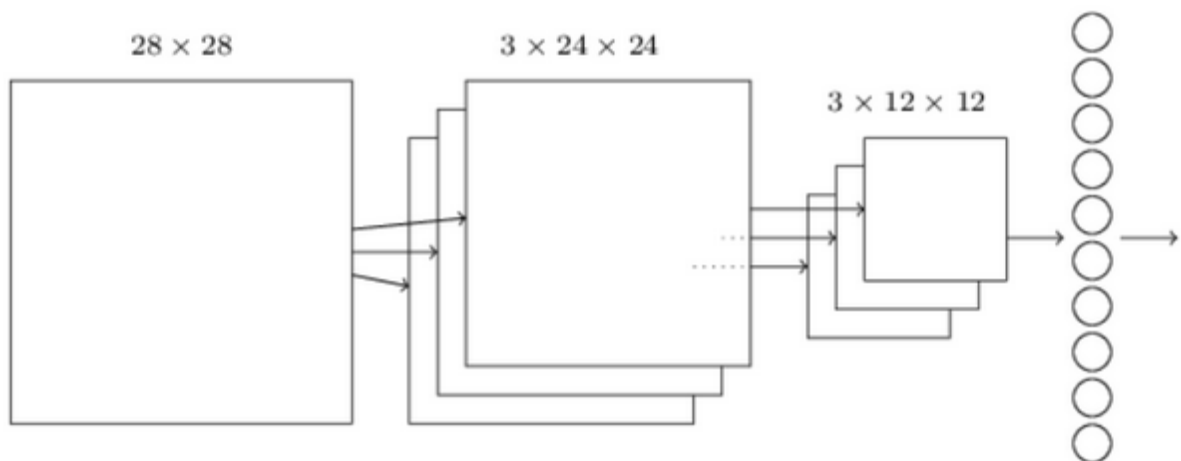
pooling phổ biến là Max pooling, thủ tục này chọn giá trị lớn nhất trong vùng đầu vào 2×2 .



Hình 3. 8 Minh họa việc lấy một đơn vị max-pooling.

Như vậy qua lớp Max Pooling thì số lượng noron giảm đi phân nửa. Trong một mạng CNN có nhiều Feature Map nên mỗi Feature Map chúng ta sẽ cho một Max Pooling khác nhau. Có thể thấy rằng Max Pooling là cách hỏi xem trong các đặc trưng này thì đặc trưng nào là đặc trưng nhất. Ngoài Max Pooling còn có L2 Pooling.

Cuối cùng ta đặt tất cả các lớp lại với nhau thành một CNN với đầu ra gồm các noron với số lượng tùy bài toán. Như bài toán nhận dạng chữ số viết tay thì đầu ra là 10.



Hình 3. 9 Mô hình một mạng CNN đơn giản.

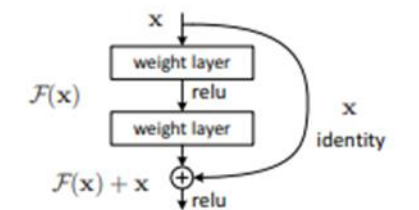
Lớp cuối cùng của các kết nối trong mạng là một lớp Fully- connected. Đó là, lớp này nối mọi neuron từ lớp max pooled tới mọi neuron của tầng đầu ra.

3.2.2 Mạng Resnet

Resnet là viết tắt của Residual Network- mạng học phần dư (residual learning).

Mạng CNN đã dẫn đến một loạt những đột phá cho phân loại hình ảnh. Nhiều tác vụ nhận dạng hình ảnh khác cũng được hưởng lợi từ các mô hình này. Vì vậy, trong những năm qua có một xu hướng đi sâu hơn, để giải quyết các nhiệm vụ phức tạp hơn và cũng tăng/cải thiện độ chính xác phân loại/nhận dạng. Nhưng, khi chúng ta đi sâu hơn, việc đào tạo mạng neuron trở nên khó khăn và độ chính xác cũng bắt đầu trở nên bão hòa và sau đó giảm xuống. Residual Learning cố gắng giải quyết cả hai vấn đề này.

Nói chung, trong một mạng CNN, một số lớp được xếp chồng lên nhau và được đào tạo cho nhiệm vụ trong tầm tay. Mạng học một số đặc trưng cấp thấp/trung bình/cao ở cuối các lớp của nó. Trong Residual Learning, thay vì cố gắng học các đặc trưng, chúng ta cố học các phần còn dư. Phần dư có thể đơn giản để hiểu như phép trừ các đặc trưng đã được học từ đầu vào của lớp đó. Resnet thực hiện điều này bằng cách sử dụng các kết nối phím tắt (kết nối trực tiếp đầu vào của lớp thứ n đến lớp thứ $n+x$). Điều này chứng minh rằng việc đào tạo dạng mạng này dễ hơn so với mạng CNN và vấn đề giảm độ chính xác cũng được giải quyết.



Hình 3. 10 Sơ đồ khối của Residual Network.

Resnet50 là một mạng Residual Network 50 lớp. Cũng có các biến thể khác như Resnet101 và Resnet152.

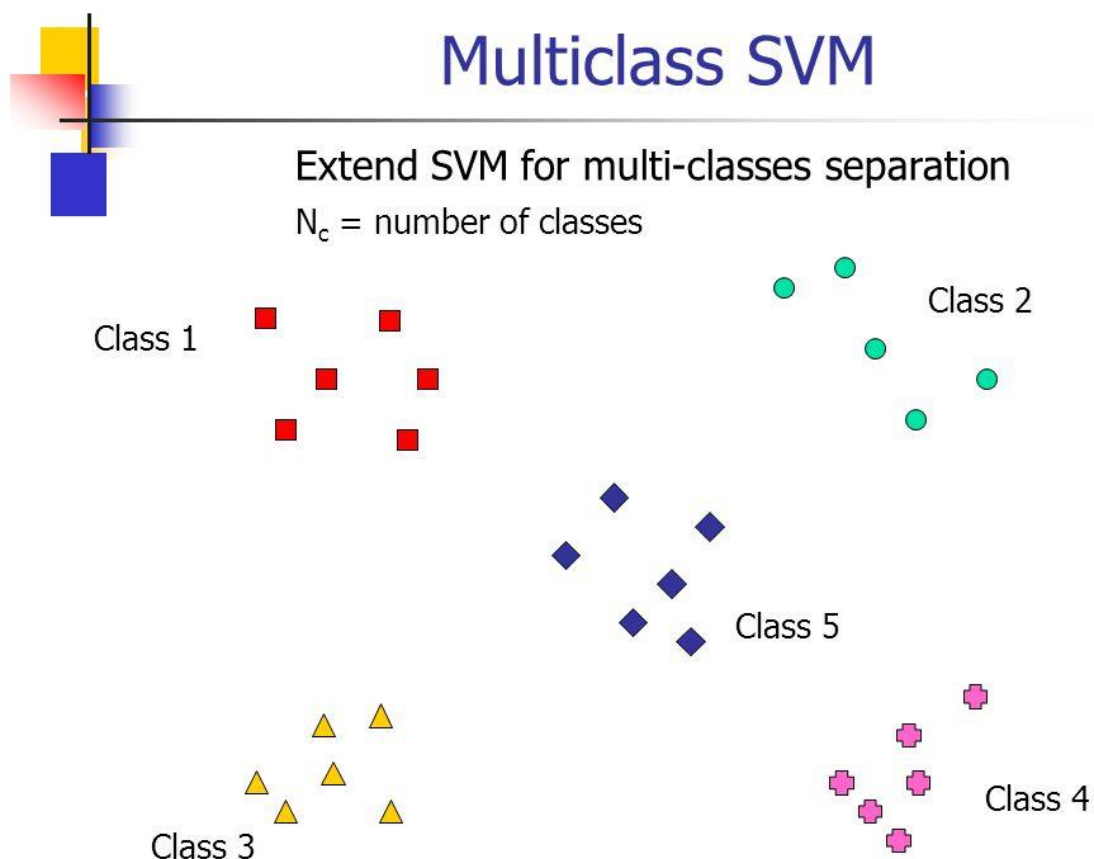
3.3. Thuật toán SVM đa lớp (Multiclass SVM)

Support vector machine (SVM) ban đầu được thiết kế cho phân loại nhị phân. Để mở rộng cho bài toán đa lớp, một số mô hình phân loại đã được đề xuất. Tuy nhiên, số lượng các biến là sản phẩm của số lượng mẫu theo số lớp, tạo ra độ phức tạp tính toán lớn.

SVM ban đầu tách các lớp nhị phân ($k=2$) với tiêu chí số dư được tối đa hóa. Tuy nhiên, vấn đề thực tế thường yêu cầu phân biệt cho hơn hai loại. Do đó, nhận dạng mẫu đa lớp có nhiều ứng dụng mở rộng bao gồm nhận dạng ký tự quang học (OCR), phát hiện xâm nhập, nhận dạng giọng nói và thông tin sinh học. Trong thực tế, vấn đề phân loại đa lớp ($k>2$) thường được phân tách thành một loạt các vấn đề nhị phân sao cho tiêu chuẩn SVM có thể được áp dụng trực tiếp. Hai sơ đồ đại diện là hướng tiếp cận một-nhiều (one-versus-rest, 1VR) và một-một (one-versus-one, 1V1). Cả hai hướng này đều là trường hợp đặc biệt của mã đầu ra sửa lỗi (Error Correcting Output Codes, ECOC) khi chia vấn đề đa lớp thành một tập các vấn đề nhị phân đã được định trước. Vấn đề chính của hướng tiếp cận này là xây dựng một ma trận ECOC tốt. Một đề xuất khác là giải quyết trực tiếp vấn đề đa lớp trong một quá trình tối ưu hóa duy nhất. Loại mô hình này kết hợp vấn đề tối ưu hóa nhiều lớp nhị phân vào một hàm khách quan duy nhất và đồng thời phân loại nhiều lớp. Tuy nhiên, độ phức tạp tính toán lớn hơn là cần thiết cho kích thước của vấn đề phương trình bậc hai (Quadratic Programming, QP).

Hơn nữa, Szedmak đã đề xuất một mô hình đa lớp cho SVM L1-norm. Trong công thức của họ, khung one-versus-rest được sử dụng. Một vấn đề tiềm năng là khi số lượng các lớp lớn, mỗi bộ phân loại nhị phân trở nên không cân bằng. Vấn đề phân loại không cân bằng xảy ra khi có nhiều mẫu của một số lớp hơn những lớp khác. Trong trường hợp này, bộ phân loại chuẩn có xu hướng bị áp đảo bởi các lớp có quy mô lớn và bỏ qua những lớp nhỏ. Thuật toán SVM đang xây dựng một phân tách siêu phẳng với lề tối đa. Từ chỉ có các vectơ hỗ trợ được dùng cho bộ phân loại và đa số mẫu xa ranh giới

quyết định có thể loại bỏ, SVM có thể chính xác hơn với dữ liệu không cân bằng vừa phải. Tuy nhiên, SVM nhạy cảm cao độ với bộ phân loại không cân bằng vì nó dễ tạo ra một bộ phân loại mà có sự đánh giá thiên vị mạnh hướng tới lớp lớn và sẽ cho độ chính xác thấp cho lớp nhỏ. Có một đề xuất phương pháp loại bỏ những khó khăn của one-versus-rest bằng cách xử lý nhiều lớp cùng nhau, đó là mở rộng phương pháp LS-SVM thành trường hợp đa lớp. Một nhược điểm của LS-SVM là giải pháp của nó được xây dựng từ hầu hết các ví dụ đào tạo. Ý tưởng PSVM (SVM gần đúng) được áp dụng vào vấn đề đa lớp có liên kết chặt chẽ với phương pháp one-versus-rest. Đối với mỗi vấn đề nhỏ được phân tách ra, giải pháp là tương tự như trường hợp nhị phân của nó: phân loại các mẫu mới bằng cách gán chúng vào gần hai mặt phẳng song song được đẩy cách nhau càng xa càng tốt.



Hình 3. 11 Minh họa Multiclass SVM.

Chương 4: Phương pháp thực hiện

4.1. Hệ thống thực hiện

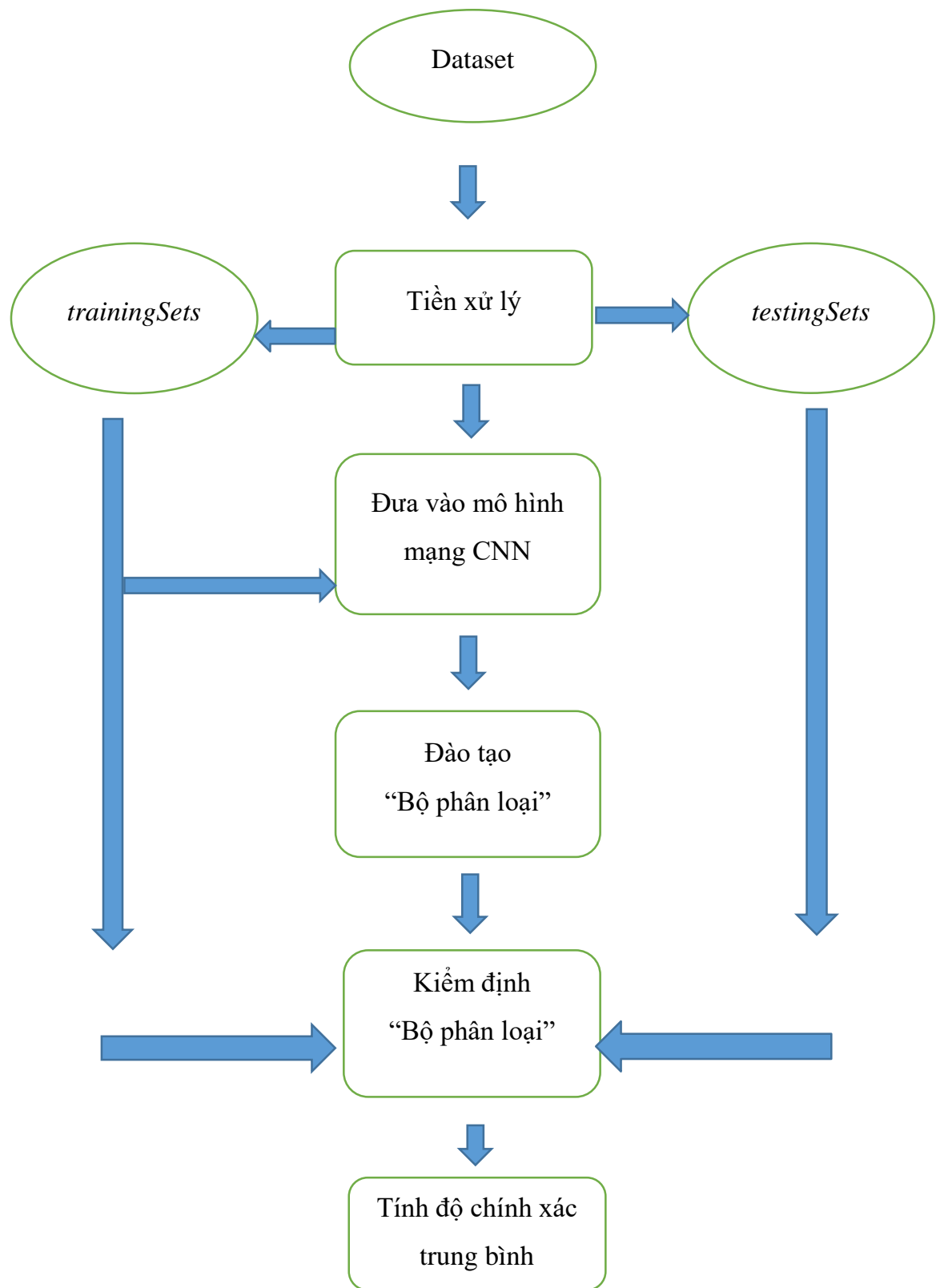
Quá trình được thực hiện trên phần mềm Matlab 2018a, cho phép áp dụng *Neural Network Toolbox*TM có các mạng CNN đã được đánh giá cao và sử dụng nhiều. Đồng thời, sử dụng hệ thống máy tính laptop với cấu hình CPU Intel Core i5-4210U và 4Gb Ram.

4.2. Dataset

Sử dụng tập dữ liệu hình ảnh nội soi “Chromoendoscopy” gồm 176 ảnh, mỗi ảnh có kích thước 518x481, trong đó có 56 ảnh mô bình thường, và 120 ảnh mô bất thường (96 ảnh mắc chứng metaplasia và 24 ảnh thuộc triệu chứng dysplasia) cùng với vùng ảnh quan tâm (Region Of Interest) tương ứng với mỗi ảnh, được xác định bởi bác sĩ. Minh họa về tập dataset nằm ở Hình 2.1 của Chương 2.

Hình ảnh Chromoendoscope thu được bằng cách sử dụng một máy nội soi Olympus GIF-H180 tại Viện Ung thư Bồ Đào Nha (IPO) Porto, Bồ Đào Nha trong các nghiên cứu lâm sàng thông thường. Ta chia tập dữ liệu thành 2 lớp ảnh là bình thường (Normal, gồm 56 ảnh) và bất bình thường (Abnormal, gồm 120 ảnh) để thực hiện quá trình phân loại.

4.3. Sơ đồ thực hiện

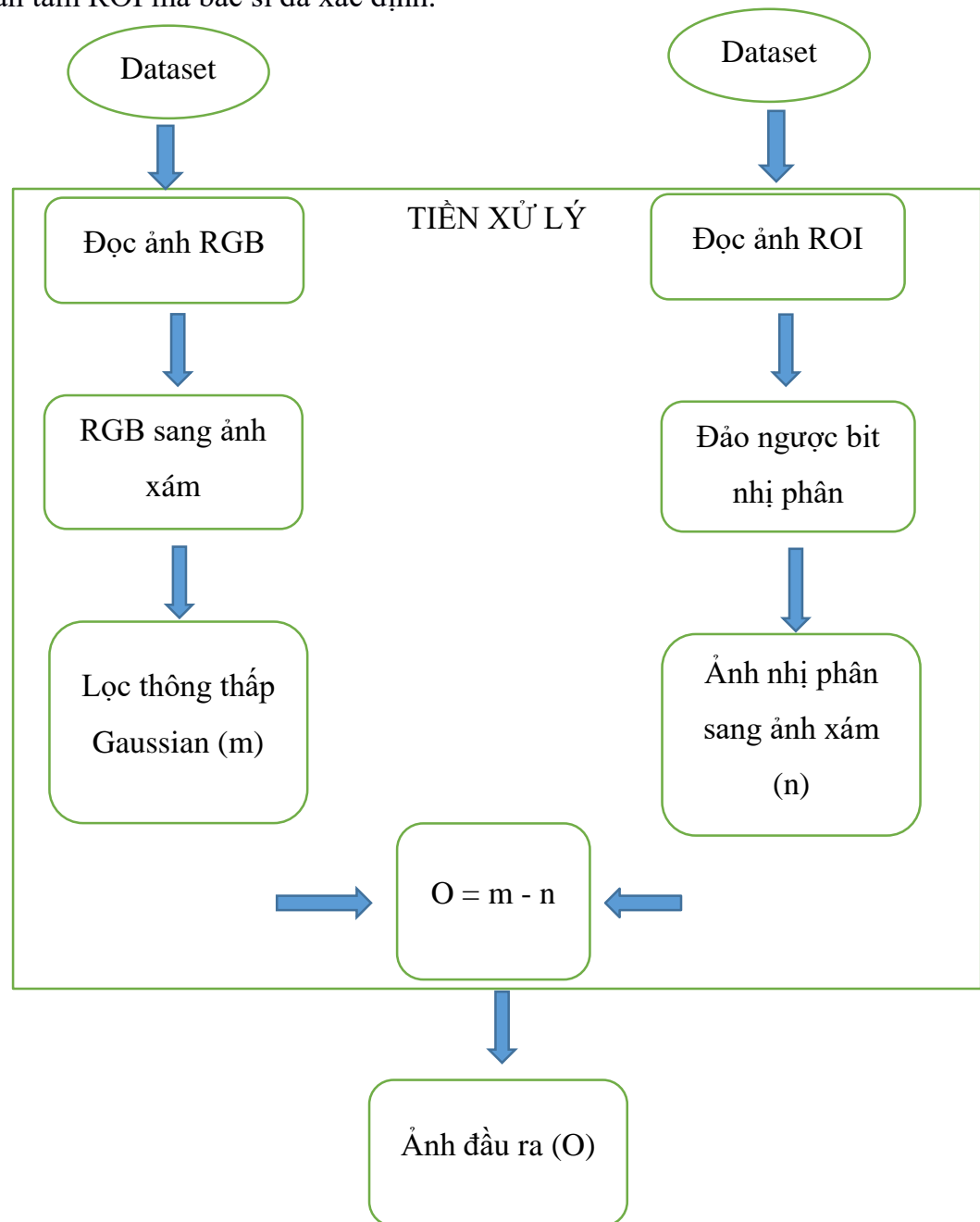


Hình 4. 1 Sơ đồ thực hiện.

4.4. Các bước phân loại hình ảnh

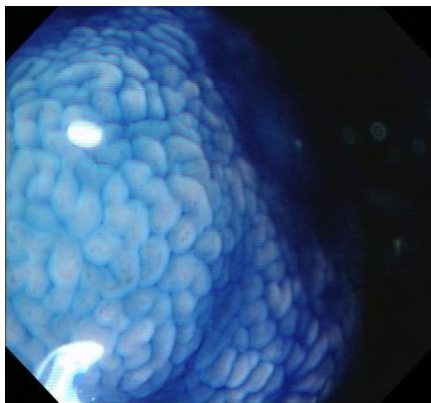
4.4.1. Bước 1: Tiền xử lý

Dữ liệu thô ban đầu (Dataset gốc) có chứa nhiễu do ánh sáng, sự chuyển động của đầu ống nội soi... cần phải loại bỏ và đồng thời cần phải phân vùng hình ảnh dựa vào vùng quan tâm ROI mà bác sĩ đã xác định.



Hình 4. 2 Sơ đồ thuật toán tiền xử lý.

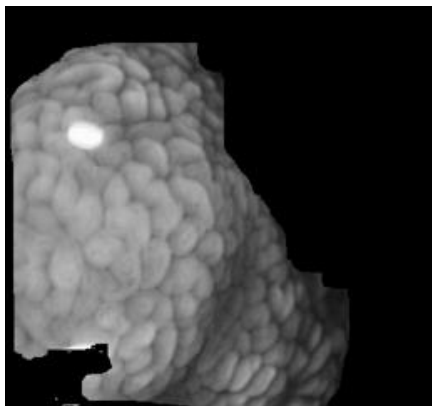
Sau khi thực hiện theo sơ đồ trên ta được ảnh xám O cùng kích thước với ảnh đầu vào và đã được loại bỏ nhiễu.



Ảnh RGB



Ảnh ROI



Ảnh O

Hình 4. 3 Ảnh O thu được sau bước tiền xử lý hai ảnh RGB và ảnh ROI.

Sau khi xử lý trên từng cặp ảnh (RBG, ROI) ta thu được tập data mới chứa 120 ảnh bất thường và 56 ảnh bình thường.

Cân bằng số lượng ảnh mỗi lớp, nghĩa là mỗi lớp bình thường và bất thường sẽ có 56 ảnh, 56 ảnh bất thường sẽ được chọn ngẫu nhiên từ 120 ảnh. Và tập data mới sẽ có 112 ảnh gồm 2 lớp bình thường và bất thường.

Chia tập data mới thành 2 tập ảnh trainingSets và testingSets với tỷ lệ chia là $N = 7:3$, tương ứng với 2 lớp.

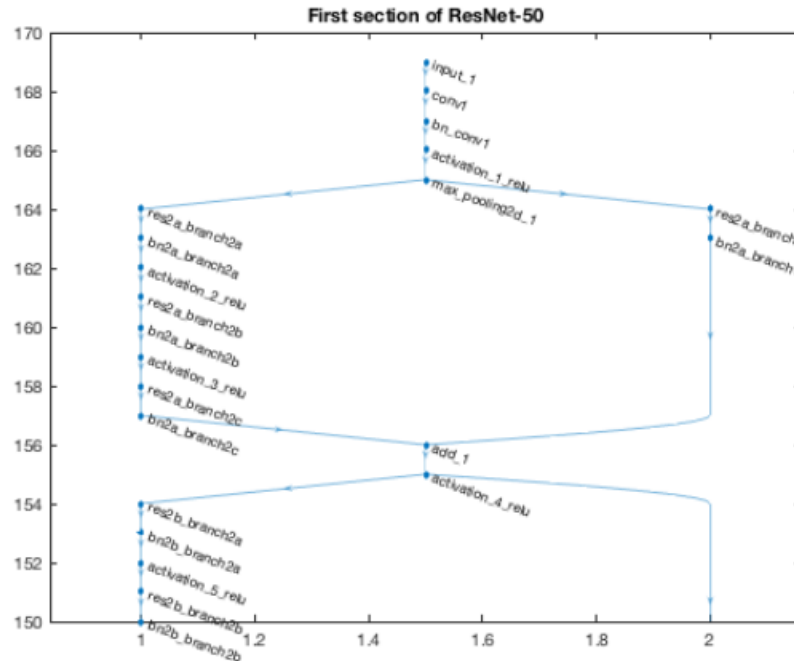
4.4.2. Bước 2: Sử dụng mạng Resnet-50 như một bộ trích xuất feature để huấn luyện bộ phân loại hình ảnh.

Một mạng nơon CNN là một kỹ thuật học máy mạnh mẽ từ lĩnh vực học sâu. CNN được đào tạo bằng cách sử dụng bộ dataset lớn với các hình ảnh đa dạng. Từ những bộ dataset lớn này, các CNN có thể tìm hiểu các biểu diễn feature phong phú cho một loạt các hình ảnh. Các biểu diễn feature này thường hoạt động tốt hơn các feature được tạo thủ công như HOG, LBP hoặc SURF. Một cách dễ dàng để tận dụng sức mạnh của CNN, mà không cần đầu tư thời gian và công sức vào đào tạo, là sử dụng một CNN giả định như một bộ trích xuất feature.

Các hình ảnh từ tập dataset “Chromoendoscopy” được phân loại bằng cách sử dụng một bộ phân loại SVM tuyến tính đa lớp được đào tạo với các tính năng CNN được trích xuất từ các hình ảnh. Ví dụ: Phân loại danh mục hình ảnh bằng cách sử dụng ví dụ về feature của túi tính năng (BoW) sử dụng các feature được tạo ra bởi SURF trong hàm BoW để đào tạo một bộ phân loại SVM đa lớp. Sự khác biệt ở đây là thay vì sử dụng các feature hình ảnh như HOG hoặc SURF, các feature được trích xuất bằng CNN.

Trong bài tập lớn này bọn em sẽ đưa ra một bộ phân loại được đào tạo bằng các feature của CNN cung cấp độ chính xác cao đặc biệt là chưa hề điều chỉnh các tham số của mạng. Nếu bài tập lớn này được phát triển và tối ưu hóa được mang sử dụng cũng như tìm được dataset hợp lý thì chúng em tin rằng độ chính xác sẽ cao hơn độ chính xác đạt được khi sử dụng BoW và SURF.

Đầu tiên, chúng em sử dụng mạng ResNet-50 đã được huấn luyện trước đó và được lấy từ *Neural Network Toolbox™ Model for ResNet-50 Network*.



Hình 4. 4 Minh họa về một đoạn ngắn của mô hình mạng Resnet-50.

Chuẩn bị tập huấn luyện (*trainingSet*) và kiểm định (*testSet*): Chia các tập dữ liệu thành hai tập là huấn luyện và kiểm định. Chọn 70% hình ảnh từ mỗi bộ cho dữ liệu huấn luyện và phần còn lại, 30% cho dữ liệu kiểm định. Ngẫu nhiên phân chia để tránh ảnh hưởng kết quả. Các tập huấn luyện và kiểm tra sẽ được xử lý bởi mô hình CNN.

Mạng Resnet-50 mà chúng em sử dụng chỉ có thể xử lý hình ảnh RGB là 224x224. Để tránh lưu lại tất cả các hình ảnh trong dataset sang định dạng này, bọn em sử dụng hàm *augmentedImageDatastore* để thay đổi kích thước và chuyển đổi bất kỳ hình ảnh thang độ xám thành RGB. Hàm *augmentedImageDatastore* có thể được sử dụng để tăng cường dữ liệu bổ sung cũng như khi được sử dụng để huấn luyện mạng.

Trích xuất các feature để huấn luyện bằng mạng CNN: Mỗi lớp của một mạng CNN tạo ra một phản ứng, hoặc kích hoạt, đến một hình ảnh đầu vào. Tuy nhiên, chỉ có một vài lớp trong một mạng CNN phù hợp cho việc trích xuất feature của ảnh. Các lớp ở đầu mạng nắm bắt các tính năng hình ảnh cơ bản, chẳng hạn như các cạnh và các đốm

màu. Điều này có thể giúp xây dựng hình ảnh trực quan tại sao các feature được trích xuất từ CNN hoạt động tốt cho các tác vụ nhận dạng hình ảnh.

Ta có thể dễ dàng trích xuất các feature từ một trong các lớp sâu hơn bằng cách sử dụng phương thức *activations*. Chọn lớp sâu như thế nào là phụ thuộc vào chúng ta, nhưng thường thì sẽ bắt đầu với lớp ngay trước khi lớp phân loại. Trong Resnet-50, lớp mà chúng em sử dụng này được đặt tên là 'fc1000', chúng em trích xuất các feature để huấn luyện bằng cách sử dụng lớp đó.

4.4.3. Bước 3: Huấn luyện bộ phân loại SVM đa lớp sử dụng feature trích xuất từ mạng CNN

Tiếp theo, sử dụng các feature của ảnh CNN để đào tạo một phân loại SVM đa lớp. Bộ giải mã Stochastic Gradient Descent nhanh được sử dụng để huấn luyện bằng cách đặt tham số '*Learners*' của hàm *fitcecoc* thành '*Linear*'. Điều này giúp đẩy nhanh quá trình đào tạo khi làm việc với các vector feature CNN nhiều chiều.

4.4.4. Bước 4: Kiểm định bộ phân loại

Lặp lại quy trình được sử dụng trước đó để trích xuất các feature của ảnh từ *testSet*. Các feature kiểm định sau đó có thể được chuyển đến bộ phân loại để đo độ chính xác của bộ phân loại được huấn luyện.

KẾT QUẢ VÀ ĐÁNH GIÁ

Kết quả của mô hình bộ phân loại trong nghiên cứu phân loại ảnh ung thư dạ dày khá là khả quan. Với mô hình bộ phân loại, bọn em tóm lược kết quả với tham số tỷ lệ $N = \text{trainingSet} : \text{testSet}$ như sau:

- $N = 7 : 3$

Tỷ lệ N	Lần	Độ chính xác		Độ chính xác trung bình	
		<i>trainingSet</i>	<i>testSet</i>	<i>trainingSet</i>	<i>testSet</i>
7 : 3	1	100	68.75	98.54	82.81
	2	97.5	81.25		
	3	97.5	81.25		
	4	100	100		
	5	95	93.75		
	6	97.5	87.5		
	7	100	81.25		
	8	100	75		
	9	97.5	75		
	10	100	87.5		
	11	97.5	75		
	12	100	87.5		

- $N = 5 : 5$

Tỷ lệ N	Lần	Độ chính xác		Độ chính xác trung bình	
		<i>trainingSet</i>	<i>testSet</i>	<i>trainingSet</i>	<i>testSet</i>
5 : 5	1	100	85.71	100	79.46
	2	100	71.43		
	3	100	67.86		

	4	100	82.14		
	5	100	85.71		
	6	100	82.14		
	7	100	89.29		
	8	100	89.29		
	9	100	75		
	10	100	75		
	11	100	82.14		
	12	100	67.86		

- N= 3: 7

Tỷ lệ N	Lần	Độ chính xác		Độ chính xác trung bình	
		<i>trainingSet</i>	<i>testSet</i>	<i>trainingSet</i>	<i>testSet</i>
3 : 7	1	93.7	60	97.91	72.92
	2	100	77.5		
	3	100	77.5		
	4	93.7	80		
	5	100	70		
	6	100	65		
	7	100	70		
	8	100	75		
	9	93.75	87.5		
	10	100	70		
	11	100	67.5		
	12	93.75	75		

- N= 6:4

Tỷ lệ N	Lần	Độ chính xác		Độ chính xác trung bình	
		<i>trainingSet</i>	<i>testSet</i>	<i>trainingSet</i>	<i>testSet</i>
6 : 4	1	97.06	81.42	99.27	79.51
	2	100	77.27		
	3	100	72.73		
	4	100	81.82		
	5	100	77.27		
	6	97.06	63.64		
	7	100	86.36		
	8	100	72.73		
	9	97.06	77.27		
	10	100	81.82		
	11	100	90.91		
	12	100	90.91		

- N= 8: 2

Tỷ lệ N	Lần	Độ chính xác		Độ chính xác trung bình	
		<i>trainingSet</i>	<i>testSet</i>	<i>trainingSet</i>	<i>testSet</i>
8 : 2	1	97.73	66.67	98.46	85.42
	2	97.73	83.33		
	3	97.73	83.33		
	4	97.73	100		
	5	97.73	91.67		
	6	97.73	75		
	7	100	83.33		
	8	97.73	83.33		

	9	99.73	91.67		
	10	100	91.67		
	11	100	83.33		
	12	97.73	91.67		

Kiểm định *trainingSet* nhằm mục đích kiểm tra tính đúng đắn của việc đào tạo bộ phân loại, nghĩa là chúng ta sử dụng *trainingSet* để tạo nên feature và các feature này ta đào tạo nên “Bộ phân loại” – Classifier. Như vậy khi sử dụng Classifier này kiểm tra lại tập *trainingSet* cho ra độ chính xác càng cao thì có nghĩa là quá trình đào tạo của ta là càng chuẩn xác.

Kiểm định *testSet* nhằm mục đích kiểm tra độ chính xác của Classifier với tập dữ liệu mới, tức là các ảnh cần phân loại. Trong nhiều trường hợp chúng ta thấy, với cùng một “Bộ phân loại” thì độ chính xác đối với tập *trainingSet* là cao, nhưng đối với tập *testSet* lại thấp và ngược lại. Chúng ta cần phải cân bằng 2 yếu tố này để vừa đảm bảo quá trình đào tạo là chuẩn xác và vừa đáp ứng được phân loại ảnh với độ chính xác cao.

Ngoài ra có nhiều yếu tố để lựa chọn phương án tối ưu. Như dựa vào số lượng ảnh mà chúng ta có và cần phân loại. Trong thực tế, luôn có giới hạn về tốc độ xử lý cũng như bộ nhớ của máy tính nên số lượng ảnh trong *trainingSets* sẽ có một giới hạn nhất định thỏa mãn điều kiện trên, đồng thời đảm bảo thời gian tính toán. Như vậy ta luôn biết trước được số lượng ảnh trong *trainingSet*.

Thông thường, khi có một vị trí nghi ngờ ung thư thì bác sĩ thường chụp nhiều ảnh về vị trí đó với các góc độ, khoảng cách chụp... khác nhau, tạo nên tập dữ liệu ảnh về vị trí nghi ngờ (*testSet*).

Ta luôn có thể tính toán được tỷ lệ $N = \text{trainingSet} : \text{testSet}$, và số lượng ảnh trong *testSet* luôn nhỏ hơn nhiều so với *trainingSet*. Dựa vào bảng kết quả độ chính xác khi thay đổi tỷ lệ N để chọn giá trị N sao cho phù hợp với bài toán thực tế.

Hơn nữa cũng cần phải cân bằng giữa tốc độ xử lý, và độ chính xác (tức là không làm mất đi quá nhiều chi tiết, tính năng đặc trưng trong ảnh).

Từ kết quả đạt được chúng em nhận ra rằng, chỉ với việc thay đổi tỷ lệ chia dataset kết quả chênh lệch khá nhiều, đồng thời chúng em vẫn chưa can thiệp vào mạng CNN để thay đổi một số tham số của một vài lớp trong mạng nên chưa thể có kết quả như mong muốn. Chúng em hy vọng rằng sẽ có thể nghiên cứu sâu hơn về lĩnh vực Deep-learning này để có thể hiểu biết thêm một lĩnh vực mạnh mẽ trong nền công nghiệp 4.0 hiện nay. Việc chúng em áp dụng Machine learning và Deep-learning trong nền y học hiện đại, đặc biệt trong công nghệ chẩn đoán hình ảnh và đặc biệt bài tập lớn này đã khiến cho chúng em thêm hiểu phần nào về việc quan trọng của xử lý ảnh trong y tế và phương hướng phát triển mạnh mẽ của ngành y sinh hiện đại.

TÀI LIỆU THAM KHẢO

- [1] S. Francisco, M. T. Coimbra, and R. Gamelas Sousa, “Recognition of Cancer using a Bag-of-Words Random Forest for Gastroenterology,” 2015
- [2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, “Deep Residual Learning for Image Recognition”, 2015
- [3]<https://www.quora.com/What-is-the-deep-neural-network-known-as-%E2%80%9CResNet-50%E2%80%9D>
- [4] https://vi.wikipedia.org/wiki/Ung_th%C6%B0_d%E1%BA%A1_d%C3%A0y
- [5] BMC Gastroenterology 11(1):135 · December 2011
- [6] <https://aidasub-chromogastro.grand-challenge.org>
- [7] https://vi.wikipedia.org/wiki/Học_máy
- [8] <https://techtalk.vn/tat-tan-tat-moi-kien-thuc-co-ban-ve-machine-learning.html>
- [9] Zhe Wang and Xiangyang Xue, Chapter 2: “Multi-Class Support Vector Machine”
- [10] Báo cáo xử lý ảnh y tế nhóm 10 K58, Phân loại ảnh ung thư dạ dày sử dụng mô hình túi từ trực quan, 06/2017.

PHỤ LỤC

Tiền xử lý

```
clc;
n=88; % So luong anh
for count=29:n
% Doc anh goc
input=imread(['C:\Users\Pho Ha Anh\Desktop\dataset\dataset\Group 2 -
Abnormal\CHGastro_Abnormal_',num2str(count),'.png']);
% Doc anh ROI
mask=imread(['C:\Users\Pho Ha Anh\Desktop\dataset\dataset\Group 2 -
Abnormal\CHGastro_Abnormal_',num2str(count),'_ROI.png']);
% Chuyen anh rgb sang anh xam
input = rgb2gray(input);
%-----
%          Lay Vung ROI
%-----
% Loc thong thap gaussian
k=fspecial('gaussian',[3 3],0.5);
input=imfilter(input,k);
%-----
IM2 = imcomplement(mask); % Xu ly anh nhi phan ROI: dao nguoc bit
gray = 255*uint8(IM2);    % Chuyen anh nhi phan thanh anh xam
out=input-gray;          % Lay vung anh can thiet
%-----
out = imresize(out,0.5); % Giam kich thuoc anh
filename =(['C:\Users\Pho Ha
Anh\Desktop\dataset\dataset\tien_xu_ly\anh_O_2\Anh_',num2str(count),'.png']
);
imwrite(out, filename);
end
```

Code chính

```
function CNN()
%load data
%outputFolder = fullfile('E:\BTL Y sinh\BTL XLAYT', 'data');
rootFolder = fullfile('E:\BTL_Y_sinh\BLT_XLAYT', 'data');
%path = char (rootFolder);
categories = {'anh_O_1', 'anh_O_2'};
% image and the category labels associated with each image
imds = imageDatastore(fullfile(rootFolder, categories), 'LabelSource',
'foldernames');
%summarize the number of images per category
tbl = countEachLabel(imds);
minSetCount = min(tbl(:,2)); % determine the smallest amount of images in a
category
%Use splitEachLabel method to trim the set.
imds = splitEachLabel(imds, minSetCount, 'randomize');
% Notice that each set now has exactly the same number of images.
countEachLabel(imds);
% Find the first instance of an image for each category
% normal = find(imds.Labels == 'anh_O_1', 1);
% abnormal = find(imds.Labels == 'anh_O_2', 1);
% figure
% subplot(1,2,1);
% imshow(readimage(imds,normal))
% subplot(1,2,2);
% imshow(readimage(imds,abnormal))
%70% training and 30% test
[trainingSet, testSet] = splitEachLabel(imds, 0.8, 'randomize');
% Load pretrained network
net = resnet50();
% profile = 'digitsnet.prototxt';
% datafile = 'digits_iter_10000.caffemodel';
%net = importCaffeNetwork(profile,datafile);
imageSize = net.Layers(1).InputSize;
% Get the network weights for the second convolutional layer
w1 = net.Layers(2).Weights;

% Scale and resize the weights for visualization
w1 = mat2gray(w1);
w1 = imresize(w1,5);

% Display a montage of network weights. There are 96 individual sets of
% weights in the first layer.
% figure
% montage(w1)
% title('First convolutional layer weights')
augmentedTrainingSet = augmentedImageDatastore(imageSize, trainingSet,
'ColorPreprocessing', 'gray2rgb');
augmentedTestSet = augmentedImageDatastore(imageSize, testSet,
'ColorPreprocessing', 'gray2rgb');

%extract trainfeature
featureLayer = 'fc1000';
trainingFeatures = activations(net, augmentedTrainingSet, featureLayer, ...
'MiniBatchSize', 32, 'OutputAs', 'columns');
```

```

% Get training labels from the trainingSet
trainingLabels = trainingSet.Labels;

% Train multiclass SVM classifier using a fast linear solver, and set
% 'ObservationsIn' to 'columns' to match the arrangement used for training
% features.
classifier = fitcecoc(trainingFeatures, trainingLabels, ...
    'Learners', 'Linear', 'Coding', 'onevsall', 'ObservationsIn',
    'columns');
% test on trainingSet
predictedtrainLabels = predict(classifier, trainingFeatures,
    'ObservationsIn', 'columns');
confMattrain = confusionmat(trainingLabels, predictedtrainLabels);
confMattrain = bsxfun(@rdivide, confMattrain, sum(confMattrain, 2));

% Extract test features using the CNN
testFeatures = activations(net, augmentedTestSet, featureLayer, ...
    'MiniBatchSize', 32, 'OutputAs', 'columns');

% Pass CNN image features to trained classifier
predictedLabels = predict(classifier, testFeatures, 'ObservationsIn',
    'columns');

% Get the known labels
testLabels = testSet.Labels;

% Tabulate the results using a confusion matrix.
confMat = confusionmat(testLabels, predictedLabels);

% Convert confusion matrix into percentage form
confMat = bsxfun(@rdivide, confMat, sum(confMat, 2));

% Display the mean accuracy
mean(diag(confMattrain))
mean(diag(confMat))
% newImage = imread(fullfile(rootFolder, 'anh_O_2', 'Anh_29.png'));
%
% % Create augmentedImageDatastore to automatically resize the image when
% % image features are extracted using activations.
% ds = augmentedImageDatastore(imageSize, newImage, 'ColorPreprocessing',
    'gray2rgb');
%
% % Extract image features using the CNN
% imageFeatures = activations(net, ds, featureLayer, 'OutputAs',
    'columns');
% label = predict(classifier, imageFeatures, 'ObservationsIn', 'columns')

```