

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



NGUYỄN MẠNH HIẾU

**PHÁT HIỆN VÀ NHẬN DẠNG HÌNH DÁNG LOẠI VIÊN
THUỐC**

Chuyên ngành: HỆ THỐNG THÔNG TIN

Mã số: 8.48.01.04

TÓM TẮT LUẬN VĂN THẠC SĨ

HÀ NỘI - NĂM 2022

Luận văn được hoàn thành tại:

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Người hướng dẫn khoa học: PGS. TSKH. HOÀNG ĐĂNG HẢI

Phản biện 1: TS. Nguyễn Xuân Thắng

Phản biện 2: PGS.TS. Đỗ Trung Tuấn

Luận văn sẽ được bảo vệ trước Hội đồng chấm luận văn thạc sĩ tại Học viện Công nghệ Bưu chính Viễn thông

Vào lúc: 10 giờ 00 ngày 17 tháng 12 năm 2022

Có thể tìm hiểu luận văn tại:

- Thư viện của Học viện Công nghệ Bưu chính Viễn thông

MỞ ĐẦU

1. Lý do chọn đề tài

Cùng với sự tiến bộ, phát triển của kinh tế - khoa học - xã hội, thế giới đang ngày càng tập trung hơn vào vấn đề chăm sóc sức khỏe cho con người; trong đó, việc nâng cao khả năng tự chăm sóc của người dân trở thành trọng tâm ở nhiều quốc gia. Việc tự chăm sóc sức khỏe tốt có thể giảm thiểu đáng kể các trường hợp phải nhập viện, phòng tránh các trường hợp có thể gây ra bệnh tật, nguy hiểm và trong một số trường hợp có thể dẫn tới tử vong. Tuy nhiên, với sự đa dạng của các loại thuốc viên trên thị trường, người bình thường thường, đặc biệt là những người lớn tuổi, thường gặp nhiều khó khăn trong việc phân biệt hoặc nhận dạng các viên thuốc không có nhãn mác, do đó tỉ lệ các trường hợp phải nhập viện hoặc gặp các vấn đề về sử dụng sai thuốc hiện nay tương đối cao. Theo [1], có 1/5 tác dụng phụ do thuốc gây ra có liên quan đến việc bệnh nhân sử dụng thuốc tại nhà; khoảng 18 triệu người từ 12 tuổi trở lên có thể sử dụng sai thuốc trị liệu tâm lý theo toa mỗi năm; 10% thuốc trên toàn cầu là hàng giả. Hậu quả sai sót về thuốc có thể ảnh hưởng nghiêm trọng đến sức khỏe và chất lượng cuộc sống của con người, dùng sai thuốc có thể dẫn đến các tác dụng phụ nghiêm trọng, nhập viện và thậm chí tử vong. Mỗi năm, tại Hoa Kỳ có khoảng hơn 7 triệu bệnh nhân bị ảnh hưởng, 7.000 đến 9.000 người chết do lỗi sử dụng thuốc [2], tại Ấn Độ có 5,2 triệu người chết do các sai sót y tế [3]... Chính vì vậy, việc phát triển các hệ thống phát hiện và nhận dạng hình dáng loại viên thuốc là một nhu cầu thực tiễn cấp thiết và đang là một chủ đề nghiên cứu được quan tâm nhiều (ví dụ [9-14]).

Mặc dù các viên thuốc có thể được phân biệt dựa trên đặc trưng cơ bản của các viên thuốc như: hình dạng, kích thước, màu sắc và dấu ấn riêng... nhưng vẫn cần những kiến thức chuyên môn nhất định và không tránh khỏi những nhầm lẫn chủ quan. Do đó, một hệ thống nhận dạng tự động có thể hỗ trợ hiệu quả cho cả bệnh nhân và nhân viên y tế trong việc phát hiện và nhận dạng các viên thuốc một cách chính xác, nhanh chóng và khách quan.

Đã có nhiều công trình nghiên cứu về nhận dạng và phân loại viên thuốc trong những năm qua, điển hình như [9-14]; trong đó, một số phương pháp phổ biến được áp dụng để phát hiện và phân loại viên thuốc là so khớp ảnh, hình dáng dựa vào trích chọn các đặc trưng cơ bản của viên thuốc. Dựa trên nền tảng sử dụng công nghệ mạng nơ ron nhân tạo (Artificial Neuron Network - ANN) cho hiệu quả rất tốt đối với các bài toán

nhận dạng đối tượng, các phương pháp học sâu (Deep Learning) với các mô hình mạng nơ ron tích chập (Convolutional Neural Network – CNN) đã được đề xuất nhằm tăng khả năng phát hiện, nhận dạng chính xác viên thuốc. Đã có nhiều mô hình CNN được đề xuất [4], điển hình là mô hình CNN kết hợp vùng (Region-CNN hay R-CNN), mô hình R-CNN nhanh (Fast R-CNN) [5], mô hình R-CNN nhanh hơn (Faster R-CNN) [6]. Một cải tiến mới của Faster R-CNN là Mask R-CNN được đề xuất năm 2017 [7]. Cho tới nay, Mask R-CNN là mô hình có nhiều ưu điểm vượt trội nhất trong số các mô hình R-CNN và cũng đã được đề xuất áp dụng cho mô hình nhận dạng viên thuốc, ví dụ [8]. Tuy nhiên, như đã chỉ ra trong [8], độ chính xác của Mask R-CNN phụ thuộc vào nhiều yếu tố như kỹ thuật phân đoạn ảnh, xác định vùng kết hợp, tập dữ liệu huấn luyện, môi trường kiểm thử.

Từ tình hình trên, việc nghiên cứu đề tài về ***“Phát hiện và nhận dạng hình dáng loại viên thuốc”*** là hết sức cần thiết, đáp ứng nhu cầu thực tiễn, góp phần xây dựng hệ thống phần mềm có khả năng hỗ trợ người dùng, nhân viên y tế, viện dưỡng lão nhằm nâng cao chất lượng chăm sóc sức khỏe người bệnh, phòng ngừa những rủi ro từ việc sử dụng sai loại thuốc.

2. Mục đích và nhiệm vụ nghiên cứu

- **Mục đích nghiên cứu:** Nghiên cứu, đề xuất giải pháp nhằm phát hiện và nhận dạng hình dáng loại viên thuốc một cách tự động bằng học sâu, thực nghiệm đánh giá giải pháp đã đề xuất.
- **Nhiệm vụ nghiên cứu:**
 - Xác định mục đích, yêu cầu phát hiện và nhận dạng hình dáng loại viên thuốc.
 - Nghiên cứu các phương pháp học sâu và khả năng áp dụng cho bài toán phát hiện và nhận dạng hình dáng loại viên thuốc.
 - Nghiên cứu xây dựng hệ thống (giải pháp) phát hiện và nhận dạng hình dáng loại viên thuốc bằng phương pháp học sâu.
 - Thực hiện thử nghiệm, đánh giá kết quả

3. Đối tượng và phạm vi nghiên cứu

* **Đối tượng:** Phương pháp phát hiện và nhận dạng hình dáng loại viên thuốc trong ảnh bằng học sâu.

*** Phạm vi nghiên cứu:**

- Các phương pháp phân đoạn ảnh và nhận dạng đối tượng.
- Phương pháp phân đoạn ảnh và nhận dạng hình dáng loại viên thuốc bằng học sâu.
- Xây dựng hệ thống phát hiện và nhận dạng hình dáng loại viên thuốc bằng học sâu.
- Thử nghiệm huấn luyện và đánh giá hệ thống trên bộ dữ liệu ảnh thuốc được lựa chọn và gán nhãn từ bộ dữ liệu ảnh CURE (chứa 8.973 hình ảnh của 196 lớp viên thuốc) [9].

4. Phương pháp nghiên cứu

*** Phương pháp nghiên cứu lý thuyết:**

- Lý thuyết liên quan vấn đề nghiên cứu.
- Tìm hiểu về một số phương pháp, mô hình học sâu hiện đại trong giải quyết bài toán nhận dạng đối tượng.
- Ứng dụng phát hiện và nhận dạng hình dáng loại viên thuốc.

*** Phương pháp nghiên cứu thực nghiệm:**

- Khảo sát, thu thập, xây dựng bộ dữ liệu ảnh thuốc mẫu.
- Lập trình ứng dụng thực nghiệm phương pháp đề xuất.
- Đánh giá kết quả thu được.

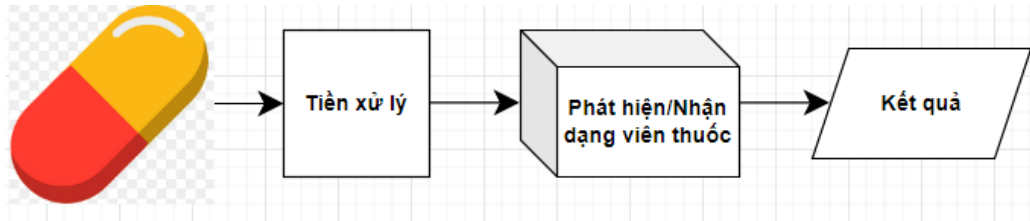
5. Cấu trúc của luận văn

- Chương 1 trình bày cơ sở lý thuyết về phát hiện và nhận dạng hình dáng loại viên thuốc.
- Chương 2 trình bày giải pháp hệ thống phát hiện và nhận dạng hình dáng loại viên thuốc sử dụng Mask R-CNN
- Chương 3 trình bày kết quả thực nghiệm và đánh giá.

CHƯƠNG 1: CƠ SỞ LÝ THUYẾT

1.1 Bài toán phát hiện và nhận dạng loại viên thuốc

Bài toán phát hiện và nhận dạng loại viên thuốc bao gồm các phần chính như sau:
 Image → Tiền xử lý (Pre-processing) → Phát hiện/nhận dạng (Detection/Recognition)
 → Output Hình Error! No text of specified style in document..1.



Hình Error! No text of specified style in document..1. Thành phần chính của bài toán phát hiện, nhận dạng hình dáng viên thuốc

1.2 Tiền xử lý dữ liệu

Phát hiện và nhận dạng hình dáng loại viên thuốc dựa trên việc xử lý dữ liệu đầu vào là các hình ảnh, do đó việc tiền xử lý dữ liệu - ở đây là tiền xử lý ảnh, trích xuất đặc trưng mặt không gian là quá trình tất yếu của mô hình định hướng xây dựng.

Ứng dụng tiền xử lý dữ liệu sẽ giúp ta tăng độ đa dạng của dữ liệu mẫu, giải quyết được nhiều hơn các trường hợp trên thực tế và giúp nâng cao độ chính xác, cải thiện khả năng phát hiện, phân loại, nhận dạng viên thuốc. Cũng nhờ tiền xử lý ảnh mà tiết kiệm được chi phí thời gian cho quá trình chuẩn bị dữ liệu cho huấn luyện.

1.3 Phân đoạn ảnh và nhận dạng đối tượng

1.3.1 Phân đoạn hình ảnh

Phân đoạn hình ảnh là một lĩnh vực cơ bản của thị giác máy tính được hỗ trợ bởi một lượng lớn nghiên cứu liên quan đến cả các thuật toán dựa trên xử lý hình ảnh và các kỹ thuật dựa trên học máy.

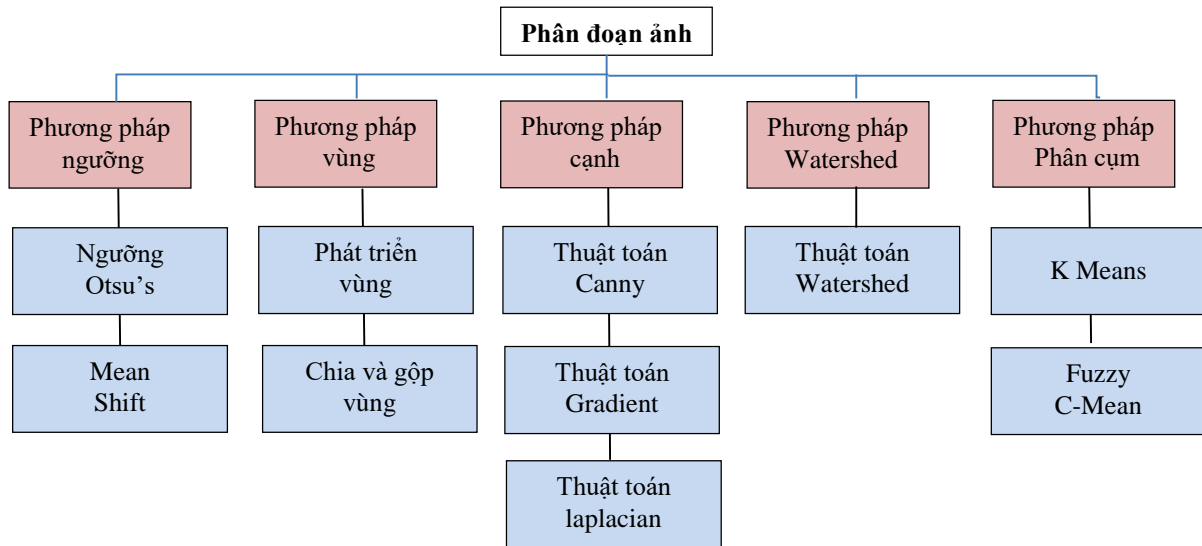
Phân đoạn hình ảnh là một miền phụ của thị giác máy tính và xử lý hình ảnh kỹ thuật số nhằm mục đích nhóm các vùng hoặc phân đoạn tương tự của hình ảnh dưới nhãn lớp tương ứng của chúng.

1.3.2 Các cách tiếp cận trong phân đoạn hình ảnh

- Phân đoạn ngữ nghĩa - Phân đoạn theo loại đối tượng (Semantic Segmentation).
- Phân đoạn phiên bản - Phân đoạn từng đối tượng (Instance Segmentation).

- Phân đoạn khái quát.

1.4. Các kỹ thuật phân đoạn ảnh



Hình Error! No text of specified style in document..2. Một số phương pháp phân đoạn ảnh truyền thống [20]

1.5. Phân loại đối tượng

Phân loại nhãn đơn: Phân loại nhãn đơn là nhiệm vụ phân loại phổ biến nhất trong phân loại ảnh có giám sát.

Phân loại đa nhãn: Phân loại đa nhãn là một nhiệm vụ phân loại trong đó mỗi hình ảnh có thể chứa nhiều hơn một nhãn và một số hình ảnh có thể đồng thời chứa tất cả các nhãn.

1.6. Mạng nơ ron nhân tạo

1.6.1. Mạng nơ-ron tích chập

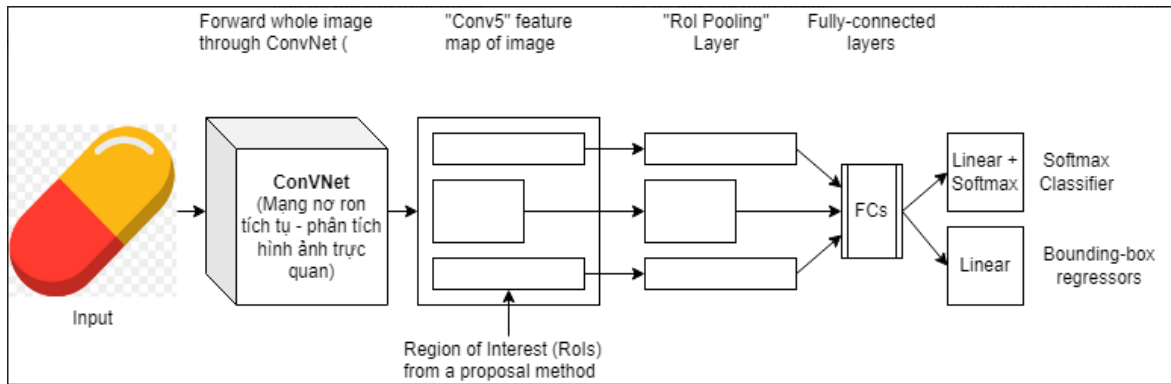
Một mạng nơ-ron tích chập bao gồm một lớp đầu vào, các lớp ẩn và một lớp đầu ra. Trong mọi mạng nơron lan truyền thẳng (Feed-forward Neural Network - FNN), các lớp giữa được gọi là lớp ẩn vì các đầu vào và đầu ra của chúng bị che bởi hàm kích hoạt và tích chập cuối cùng, các lớp ẩn bao gồm các lớp thực hiện các phép chập. Sản phẩm đầu ra là sản phẩm bên trong Frobenius và hàm kích hoạt của nó có thể là ReLU, Sigmoid...

1.6.2. Mô hình R-CNN

Mô hình R-CNN đề xuất một loạt các hộp trong hình ảnh và kiểm tra xem có bất kỳ hộp nào trong số này chứa bất kỳ đối tượng nào hay không. R-CNN sử dụng tìm kiếm

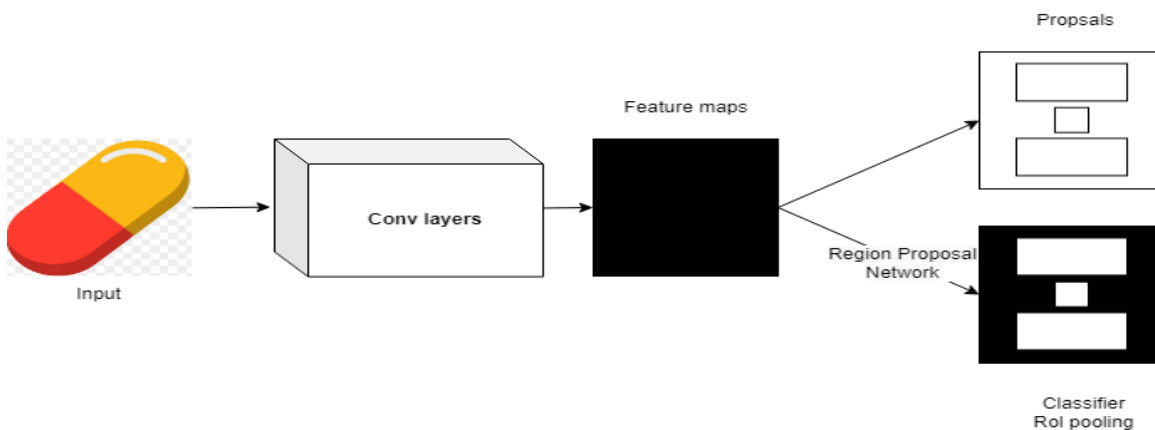
có chọn lọc để trích xuất các hộp từ một hình ảnh (các hộp này được gọi là vùng). Về cơ bản, có bốn vùng hình thành một đối tượng: tỷ lệ, màu sắc, kết cấu và vùng bao quanh khác nhau. Tìm kiếm có chọn lọc xác định các mẫu này trong hình ảnh và dựa trên đó, đề xuất các vùng khác nhau.

1.6.3. Mô hình Fast R-CNN



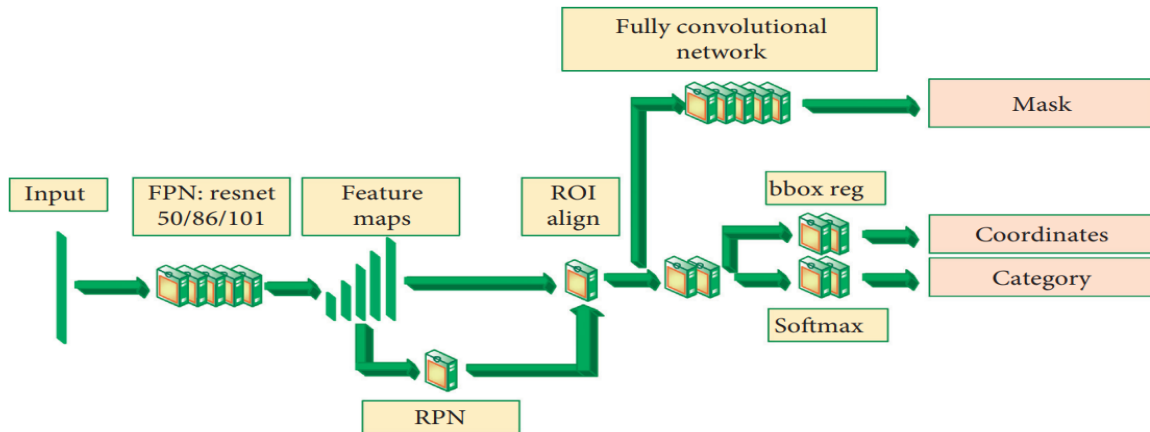
Hình *Error! No text of specified style in document.*3. Kiến trúc tổng quan của mô hình Fast R-CNN

1.6.4. Mô hình Faster R-CNN



Hình *Error! No text of specified style in document.*4. Mô hình luồng Faster R-CNN

1.6.5. Mạng Mask R-CNN



Hình *Error! No text of specified style in document..5*. Mô hình luồng của Mask R-CNN [7]

1.7. Một số nghiên cứu liên quan

- (1) Tháng 9/2020, Naoto Usuyama và cộng sự [24]
- (2) Tháng 5/2020, Alphonso Woodbury [11]
- (3) Năm 2020, Suwat Tangwattananuwat và cộng sự [13]
- (4) Năm 2020, Suiyi Ling và cộng sự [9]
- (5) Yu và Chen [25]

1.8. Kết chương

Hiện nay các nghiên cứu liên quan, nhất là các phương pháp dựa trên kỹ thuật tiền xử lý ảnh, tìm cạnh, tìm đường biên... chưa thu được độ chính xác cao trên bộ dữ liệu ảnh thuốc do người dùng cung cấp (vì dữ liệu này có độ tương phản viên thuốc và nền thấp, viên thuốc bị xoay, bị chồng đè, che lấp...), do đó, việc ứng dụng các mô hình học sâu sử dụng kỹ thuật phân đoạn lớp đối tượng (instance segmentation), Unet, Linknet... là một trong những định hướng nghiên cứu tiềm năng hiện nay để giải quyết bài toán.

Qua quá trình khảo sát một số phương pháp, các lý thuyết, vấn đề liên quan bài toán nhận dạng hình dáng loại viên thuốc; luận văn đề xuất phương pháp giải quyết bao gồm các bước: bắt đầu bằng việc tăng cường dữ liệu để làm đa dạng hơn các ảnh viên thuốc mẫu sau đó tiến hành phân tích tập dữ liệu hình ảnh viên thuốc trong bộ dữ liệu CURE; kích thước, hình dạng, màu sắc, dấu ấn... đều là các thông tin tạo nên đặc trưng của viên thuốc được trích xuất thông qua một mạng nơ-ron nhân tạo khung (backbone) học chuyển giao từ mô hình mạng học sâu (như: Resnet50, Resnet101) với trọng số được

huấn luyện trước trên bộ dữ liệu COCO kết hợp với kiến trúc mạng Mask R-CNN để huấn luyện mô hình học sâu (model); sau khi tạo cơ sở dữ liệu và mô hình, hệ thống đề xuất có khả năng thực hiện trích xuất đặc trưng tự động của hình ảnh các viên thuốc (theo định dạng và quy định chung trong việc thu thập tập dữ liệu hình ảnh viên thuốc) và sau đó trả về kết quả dự đoán và xác định viên thuốc.

Chương 2: XÂY DỰNG HỆ THỐNG PHÁT HIỆN VÀ NHẬN DẠNG HÌNH DÁNG LOẠI VIÊN THUỐC

2.1. Mô hình hệ thống

Luận văn đã tập trung vào 02 hướng tiếp cận chính để giải quyết bài toán nhận dạng hình dạng viên thuốc là: (1) Phân đoạn và phát hiện viên thuốc bằng kỹ thuật truyền thống, trong đó, sẽ tiến hành xử lý ảnh, phát hiện cạnh của các viên thuốc qua đó nhận dạng hình dáng viên thuốc; và (2) Phát hiện viên thuốc dựa trên kỹ thuật học máy, học sâu.

2.2. Các tiêu chí đánh giá

- **Precision:** Là một số liệu được xác định cho mỗi lớp mà mô hình dự đoán.
- **Recall:** Recall tương tự như precision được xác định cho mỗi lớp, cho biết tỷ lệ dữ liệu trong tập kiểm tra thuộc về lớp được xác định chính xác.
- **F1 Score** giúp đạt được sự cân bằng giữa độ precision và recall để có được ý tưởng trung bình về cách mô hình hoạt động.
- **IoU** (Intersection over Union) là chỉ số đánh giá được sử dụng để đo độ chính xác của mô hình phát hiện đối tượng (trong đó có Mask R-CNN).
- **mAP** (mean Average Precision) là giá trị trung bình của thông số Độ chuẩn trung bình (Average Precision - AP) trên tất cả các lớp với tất cả các giá trị ngưỡng giao hộp giới hạn đối tượng IoU (để xác định các giá trị dự đoán là TP, TN, FP hay FN; IoU càng cao thì dự đoán càng tốt).

2.3. Thu thập dữ liệu

Luận văn sử dụng bộ dữ liệu CURE do nhóm tác giả Suiyi Ling xây dựng và cung cấp công khai là một trong những bộ dữ liệu đa dạng, chất lượng ảnh sát với thực tế do người dùng cung cấp nhất.

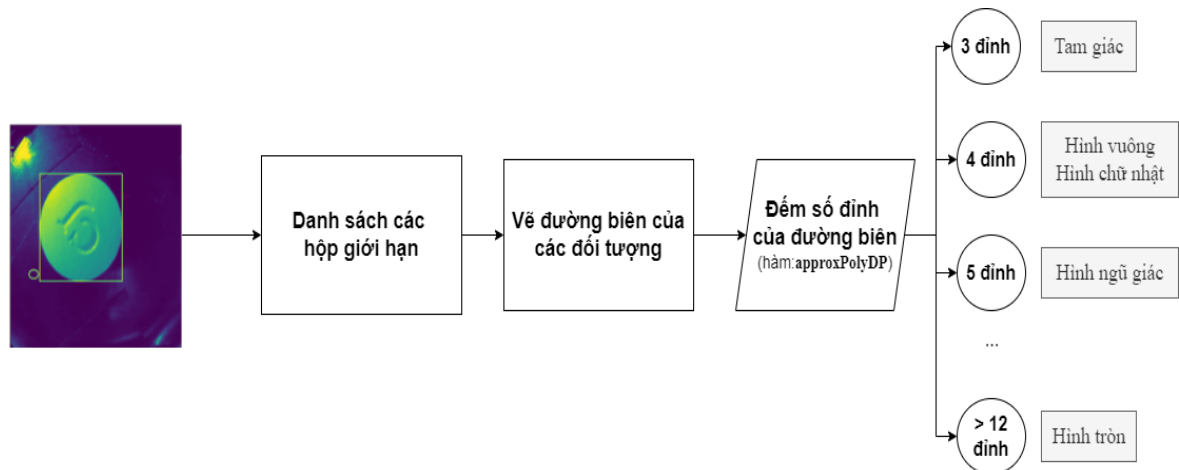
2.4. Một số thuật toán phân đoạn và nhận dạng bằng học máy truyền thống

Bảng Error! No text of specified style in document..1. So sánh một số kỹ thuật phân đoạn hình ảnh truyền thống

Kỹ thuật	Miêu tả	Ưu điểm	Nhược điểm
Phương pháp ngưỡng	Tập trung vào việc tìm các giá trị đỉnh dựa trên biểu đồ của hình ảnh để tìm các pixel tương tự	Đơn giản; không yêu cầu tiền xử lý ảnh phức tạp	Nhiều chi tiết có thể bị bỏ qua, thường gặp lỗi ngưỡng
Phương pháp dựa trên cạnh	Dựa trên phát hiện sự gián đoạn không giống như phát hiện tương tự	Tốt cho hình ảnh có độ tương phản cao giữa các đối tượng.	Không thích hợp cho hình ảnh nhiễu
Phương pháp dựa trên khu vực	Dựa trên việc phân vùng hình ảnh thành các vùng đồng nhất	Hoạt động thực sự tốt đối với hình ảnh có lượng nhiễu đáng kể, có thể lấy điểm đánh giá của người dùng để đánh giá nhanh	Tiêu tốn thời gian và bộ nhớ
Các thuật toán phân đoạn truyền thống	Chia hình ảnh thành k cụm đồng nhất, loại trừ lẫn nhau - thu được các đối tượng	Các phương pháp đã được chứng minh, được củng cố bằng logic mờ và hữu ích hơn cho ứng dụng thời gian thực.	Việc xác định hàm chi phí để tối giản có thể khó khăn.
Phương pháp watershed	Dựa trên giải thích cấu trúc liên kết của ranh giới hình ảnh	Các phân đoạn thu được ổn định hơn, các ranh giới được phát hiện có sự khác biệt	Tính toán độ dốc cho các đường biên rất phức tạp.

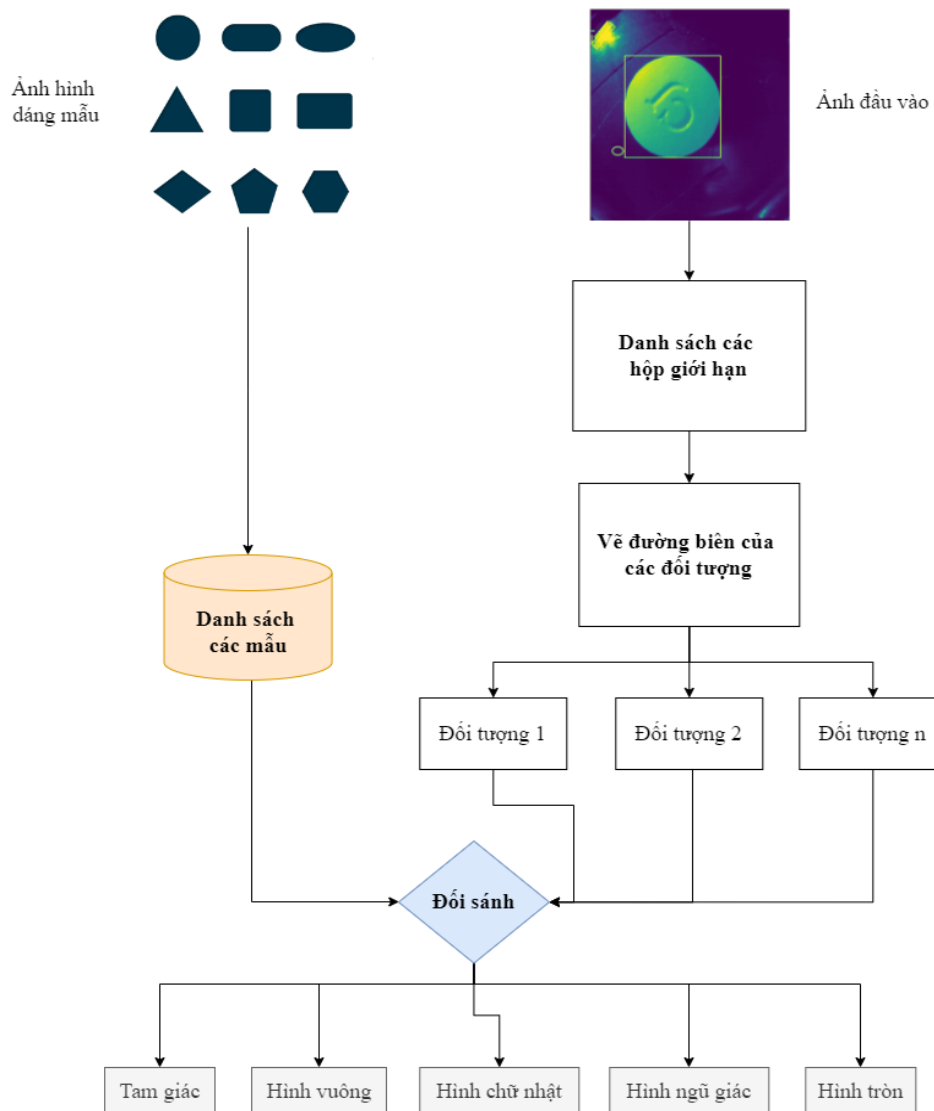
2.5. Nhận dạng hình dáng loại viên thuốc bằng phương pháp truyền thống

2.5.1. Phương pháp hình học



Hình *Error! No text of specified style in document.*6. Nhận dạng hình dáng bằng phương pháp đếm số đỉnh

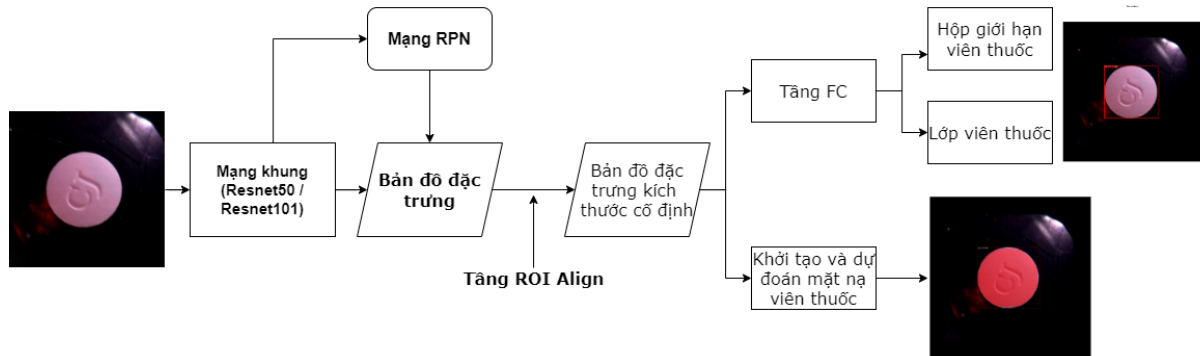
2.5.2. Phương pháp đối sánh mẫu



Hình *Error! No text of specified style in document.*7. Nhận dạng hình dáng viên thuốc bằng kỹ thuật đối sánh mẫu

2.6. Giải pháp phát hiện và nhận dạng hình dáng loại viên thuốc bằng mô hình học sâu Mask R-CNN

2.6.1. Mô hình hệ thống



Hình Error! No text of specified style in document..8. Mô hình đề xuất

2.6.2. Tiền xử lý ảnh

Áp dụng trong xây dựng hệ thống giải quyết bài toán, luận văn đã sử dụng thư viện *imgaug* để tăng cường sự đa dạng cho dữ liệu mẫu thông qua các kỹ thuật tiền xử lý ảnh như: lật ảnh (*fliplr*), áp dụng bộ lọc mờ ngẫu nhiên (*GaussianBlur*), cụ thể:

```
import imgaug

augmentation = imgaug.augmenters.Sometimes(0.5, [
    imgaug.augmenters.Fliplr(0.5),
    imgaug.augmenters.GaussianBlur(sigma=(0.0, 5.0))
])
```

2.6.3. Phát hiện và nhận dạng bằng Mask R-CNN

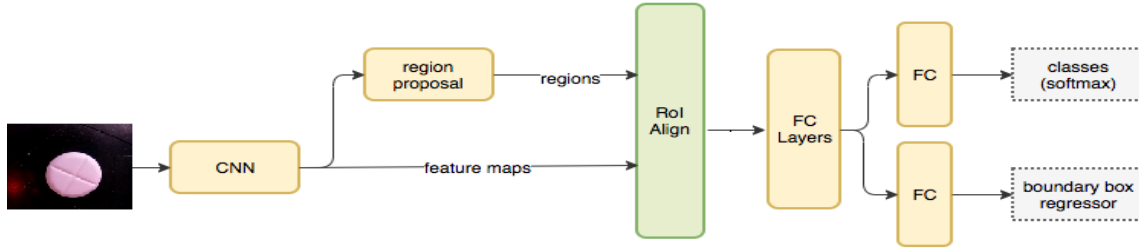
2.6.3.1. Mạng đề xuất vùng khởi tạo các hộp giới hạn viên thuốc

Với dữ liệu đầu vào là các bản đồ đặc trưng của hình ảnh thu được sau khi đi qua mạng nơ-ron tích chập (CNN), cụ thể là mạng **Resnet50** hoặc **Resnet101**,

Bước 1: Mạng đề xuất vùng (RPN) chạy một bộ phân loại nhị phân nhỏ trên nhiều hộp giới hạn (anchor) trên hình ảnh để xác định các Vùng quan tâm (Region of Interest – RoI) và trả về xác suất đánh giá việc có đối tượng hoặc không có đối tượng. Các anchor có điểm đối tượng cao (anchor dương) được chuyển sang giai đoạn hai để được phân loại. Thông thường, ngay cả các anchor dương cũng không bao phủ hoàn toàn các đối tượng. Vì vậy, RPN cũng hội quy một cách tinh chỉnh (tìm sự cân bằng về vị trí và kích thước) được áp dụng cho các anchor để dịch chuyển nó và thay đổi kích thước từng chút một cho đúng ranh giới của đối tượng;

Bước 2: ROI xuất ra nhiều hộp giới hạn viên thuốc thay vì một hộp xác định duy nhất và biến đổi chúng thành một chiều cố định;

Bước 3: Các đặc trưng đã bị biến đổi sau đó được đưa vào các lớp được kết nối đầy đủ (FC) để phân loại bằng cách sử dụng hàm *softmax* và dự đoán hộp ranh giới viên thuốc được tinh chỉnh thêm với thông số IoU bằng cách sử dụng mô hình hồi quy.

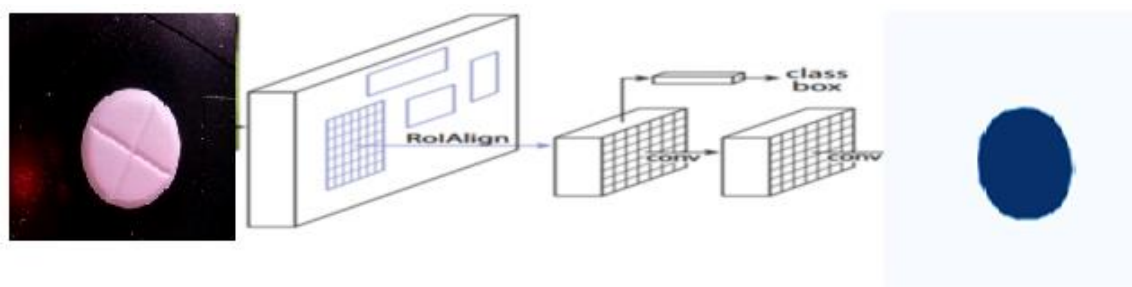


Hình 9. Mạng RPN đề xuất các ROI và dự đoán các lớp, hộp giới hạn viên thuốc

Mục tiêu của RPN: Mục tiêu của RPN (RPN Target) là các giá trị huấn luyện cho mạng RPN. Để tạo các mục tiêu, Mask R-CNN bắt đầu với một lưới các anchor bao gồm toàn bộ hình ảnh ở các tỷ lệ khác nhau, sau đó tính giá trị giao nhau của các hộp giới hạn viên thuốc (IoU) của các anchor với viên thuốc thực.

2.6.3.2. Bộ phận lớp mặt nạ viên thuốc và tăng cường dữ liệu

Trong giai đoạn thứ hai, song song với việc đề dự đoán độ lệch lớp và hộp giới hạn viên thuốc, Mask R-CNN cũng xuất ra một mặt nạ nhị phân cho mỗi RoI. Nhánh mặt nạ có một đầu ra K_{m^2} chiều cho mỗi RoI, trong đó, mã hóa K mặt nạ nhị phân độ phân giải $m \times m$ cho mỗi K lớp. Để áp dụng kỹ thuật này, cần áp dụng hàm sigmoid trên mỗi pixel và xác định LMask là tổn thất entropy chéo nhị phân trung bình (cross-entropy). Đối với một RoI được liên kết với lớp ảnh nền (growth-truth) k, LMask chỉ được xác định trên k-mask tương ứng (đầu ra mặt nạ khác không đóng góp vào giá trị hàm lỗi (loss)). Từ công thức trên, hệ thống tính toán được giá trị hàm lỗi của mặt nạ khi dự đoán viên thuốc, là kết quả rất quan trọng trong phân đoạn viên thuốc ở mức pixel-to-pixel.



Hình *Error! No text of specified style in document..10*. Mask R-CNN dự đoán mặt nạ viên thuốc

2.6.4. Huấn luyện mô hình nhận dạng hình dáng viên thuốc

Hiện nay, Mask R-CNN đã được các nhà nghiên cứu huấn luyện trên nhiều bộ dữ liệu lớn, trong đó có bộ dữ liệu COCO (là bộ dữ liệu phục vụ các bài toán phát hiện đối tượng, phân đoạn ảnh, chú thích ảnh với hơn 1.5 triệu đối tượng thuộc về 80 lớp khác nhau); do đó, việc sử dụng kỹ thuật học chuyển giao với bộ trọng số được huấn luyện trước trên bộ dữ liệu COCO trên sẽ đem lại hiệu quả cho hệ thống xây dựng, giúp khởi tạo bộ trọng số hợp lý hơn, tăng hiệu quả trích xuất các đặc trưng viên thuốc. Và áp dụng dữ liệu đã được tăng cường, ta có đầu vào là 02 bộ dữ liệu: *dataset_train* và *dataset_val*, sử dụng tỉ lệ học máy *LEARNING_RATE* khởi tạo không quá nhỏ cũng không quá lớn sẽ giúp mô hình huấn luyện hiệu quả hơn (nếu tỉ lệ quá nhỏ thì mô hình lâu hội tụ, nếu tỉ lệ quá lớn thì mô hình huấn luyện thiếu sự chính xác); đồng thời, cần một số lần lặp *epochs* đủ lớn để huấn luyện trên toàn bộ bộ dữ liệu thì mô hình mới thu được độ chính xác cao. Tuy nhiên, việc áp dụng dừng huấn luyện (*callback*) khi độ chính xác đạt đến giá trị nhất định là không thể thiếu, điều này làm giảm khả năng mô hình bị *overfitting* trên bộ dữ liệu huấn luyện. Các thông số tối ưu như Adam, SGD, RmsProp... được áp dụng sẽ giúp cập nhật các bộ trọng số hợp lý hơn cho mô hình.

Kết quả của Mask R-CNN khi dự đoán một ảnh trả về theo dạng từ điển với các khóa cho *hộp giới hạn*, *mặt nạ*, *lớp viên thuốc* và *độ chính xác*. Cụ thể:

- '*rois*': Vùng quan tâm (ROI) cho các viên thuốc được phát hiện.
- '*mask*': Mặt nạ cho các viên thuốc được phát hiện.
- '*class_ids*': Số định danh tương ứng với lớp viên thuốc được phát hiện.
- '*score*': Xác suất tin cậy cho mỗi lớp viên thuốc dự đoán.

2.6.5. Khởi tạo cấu hình mô hình và bộ dữ liệu ảnh thuốc

- Số lớp ảnh thuốc (*number_class* = 9).
- Giá trị ngưỡng tối thiểu để xác định một đối tượng là loại hình dạng viên thuốc hay không (*min-confidence*).
- Tọa độ và nhãn các ảnh viên thuốc đã được gán (*annotation_file.json*).

2.7 Kết chương

Qua nghiên cứu và thực nghiệm xây dựng chương trình, chúng tôi đã tiến hành tiếp cận giải quyết bài toán bằng 02 phương pháp (một là kỹ thuật xử lý ảnh truyền thống và một là kỹ thuật học sâu sử dụng mô hình Mask R-CNN).

Mặc dù phương pháp truyền thống đơn giản, áp dụng nhiều thuật toán lọc ảnh, đối sánh ảnh... phổ biến nhưng việc xử lý các thách thức liên quan bài toán nhiều khả năng sẽ kém hiệu quả, nhất là khi ảnh viên thuốc trùng màu với nền. Bên cạnh đó, phương pháp sử dụng mô hình Mask R-CNN là một trong những mô hình mạng nơ-ron nhân tạo tiên tiến, hiệu quả trong giải quyết bài toán phân đoạn ảnh đối tượng, hứa hẹn sẽ đem lại hiệu quả cao hơn trong giải quyết bài toán đặt ra.

Chương 3: KẾT QUẢ THỰC NGHIỆM VÀ HƯỚNG PHÁT TRIỂN

3.1. Môi trường thực nghiệm và bộ dữ liệu

Chương trình thực nghiệm trên nền tảng Google Colab, với hỗ trợ GPU – Testla 4 (40GB), sử dụng ngôn ngữ Python; chạy huấn luyện và kiểm thử trên bộ dữ liệu CURE chia thành **09 loại hình dạng** khác nhau gồm: '*Capsule*', '*Double_round*', '*Heart*', '*Modified_rectangle*', '*Octagon*', '*Oval*', '*Pentagon*', '*Round*', '*Triangle*'. luận văn đã gom nhóm và gán nhãn pixel gần **1.700 ảnh** của **09 lớp** hình dạng trên (những viên thuốc chưa gán nhãn chủ yếu là 02 loại hình dạng *Round* và *Capsule* đã được gán nhãn lượng lớn 02 loại thuốc này nên việc gán nhãn thêm sẽ không làm tăng hiệu quả của mô hình).

3.2. Kết quả thực nghiệm

3.2.1. Nhận dạng hình dáng viên thuốc bằng phương pháp phát hiện cạnh

Luận văn đã thực nghiệm một số kỹ thuật áp dụng các bộ lọc khác nhau, qua đó các cạnh của viên thuốc có thể được thể hiện rõ hơn, có sự tương phản nhất định so với

nền. Tuy nhiên, do ảnh hưởng của các điều kiện tự nhiên (ánh sáng, độ chói, bóng, tương phản...) trên ảnh của bộ dữ liệu mẫu CURE khiến phương pháp truyền thống không thể phát hiện hoàn chỉnh đủ cạnh, viền tại một số viên thuốc hoặc phát hiện sai nhiều.

Ngoài ra, luận văn đã thực nghiệm trên một số ảnh thực tế người dùng cung cấp, khi ảnh có nhiều viên thuốc, nhất là có những viên thuốc có sự chồng, lấp thì phương pháp truyền thống hoạt động không hiệu quả ngay từ bước phân đoạn viên thuốc; không thể phát hiện chính xác cạnh của các viên thuốc dẫn đến không thể phân loại hình dạng theo yêu cầu bài toán đặt ra.

3.2.2. Thực nghiệm phát hiện và nhận dạng hình dáng viên thuốc bằng mô hình Mask R-CNN

3.2.2.1. Gán nhãn dữ liệu

Đây là công việc sử dụng thời gian và công sức đáng kể. Kết quả gán nhãn thu được các tệp JSON lưu trữ thông tin về lớp hình dạng, tọa độ, vị trí của viên thuốc cũng như mặt nạ của chúng ('Capsule', 'Double_round', 'Heart', 'Modified_rectangle', 'Octagon', 'Oval', 'Pentagon', 'Round', 'Triangle').

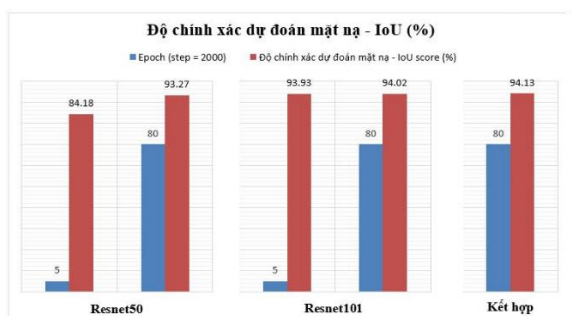
3.2.2.2. Các giá trị tham số huấn luyện mô hình

Bảng Error! No text of specified style in document..2. Giá trị tham số huấn luyện mô hình Mask R-CNN

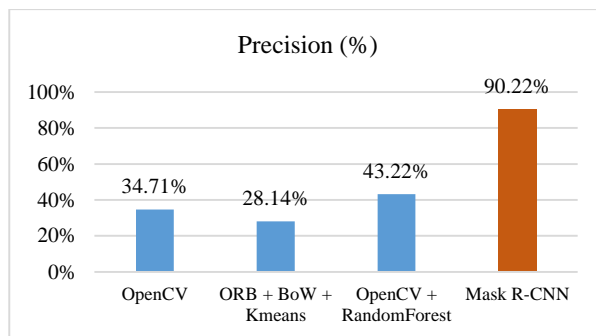
STT	Tham số	Giá trị
1	<i>Optimizer</i>	SGD
2	<i>learning_rate</i>	0.001
3	<i>batch_size</i>	1
4	<i>epoch</i>	80
5	<i>loss_function</i>	rpn_class_loss = 1 rpn_bbox_loss = 1 mrcnn_class_loss = 1 mrcnn_bbox_loss = 1 mrcnn_mask_loss = 1

3.2.2.3. Độ chính xác dự đoán của mô hình

Qua kết quả thực nghiệm trên bộ dữ liệu đánh giá, chúng tôi thu được kết quả độ chính xác khi dự đoán mặt nạ viên thuốc IoU score đạt **94,13%**, độ chính xác phân lớp đạt **90.22%** cao hơn khi áp dụng các phương pháp khác như: OpenCV, ORB + BoW + KMeans, OpenCV + RandomForest; được thể hiện dưới dạng đồ thị ở các đồ thị Hình Error! No text of specified style in document..1 và Hình Error! No text of specified style in document..2. Trong đó, để đánh giá kết quả, chương trình thực nghiệm trên 02 kiến trúc là Resnet50, Resnet101 với số lần lặp khác nhau là 5 lần và 80 lần.



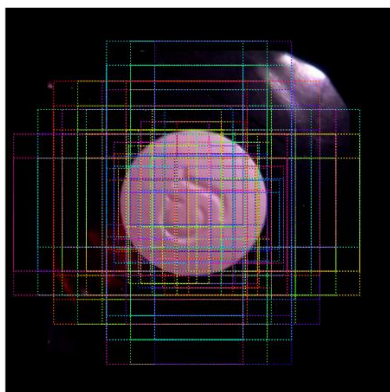
Hình Error! No text of specified style in document..1. Kết quả độ chính xác theo tỉ lệ chồng lấp IoU của mô hình thực nghiệm với các kiến trúc và thông số khác nhau



Hình Error! No text of specified style in document..2. Kết quả so sánh độ chính xác của phương pháp đề xuất với phương pháp truyền thống

* **Kết quả dự đoán trên 01 mẫu ngẫu nhiên cụ thể như sau:**

(1) Đề xuất vùng chứa viên thuốc

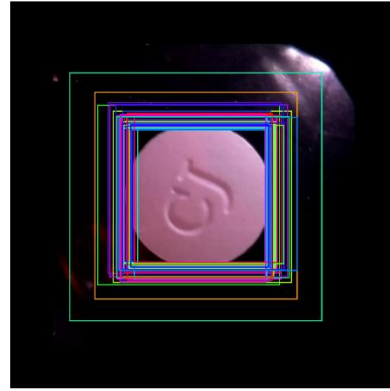
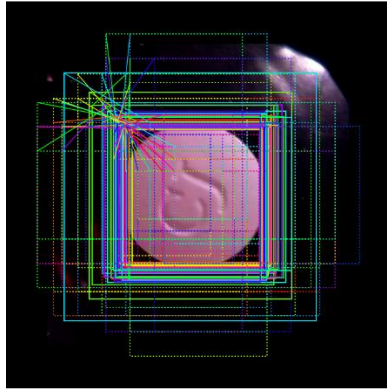


Hình Error! No text of specified style in document..3. Các anchor dương trước khi sàng lọc (chấm) và sau khi sàng lọc (liền)

(2) Phân loại vùng đề xuất



Hình Error! No text of specified style in document..4. Các anchor box được tinh chỉnh sau khi loại bỏ những hộp độ chính xác thấp



Hình Error! No text of specified style in document..5. *Hiển thị các vùng đề xuất cuối cùng*



Hình Error! No text of specified style in document..6. *Phân loại các vùng đề xuất hình dạng viên thuốc*

(3) *Tạo ra các mặt nạ viên thuốc và dự đoán lớp hình dáng*

Bước này lấy các vùng phát hiện được (tính chỉnh hộp giới hạn và định danh lớp) từ bước trước và chạy tầng đầu của Mask để tạo mặt nạ phân đoạn cho các trường hợp.



Hình Error! No text of specified style in document..7. *Tạo ra mặt nạ phân đoạn cho các viên thuốc*

Hình Error! No text of specified style in document..8. *Kết quả phát hiện và nhận dạng viên thuốc*

3.2.3. Thời gian xử lý

Thời gian huấn luyện mô hình Mask R-CNN sắp xỉ 8 giờ.

Thời gian trung bình để phát hiện và nhận dạng hình dáng loại viên thuốc trong 1 ảnh là khoảng 1.75 giây.

3.3. Kết chương

Từ kết quả thực nghiệm trên cho thấy, các phương pháp phân đoạn dựa trên việc phát hiện cạnh (phương pháp truyền thống) chưa đưa được kết quả có độ chính xác cao, khó có thể áp dụng được cho thực tiễn do ảnh các viên thuốc có thể có màu trắng và màu trùng với màu nền, làm giảm độ tương phản của đối tượng với nền. Đồng thời, các viên thuốc có khả năng bị đổ bóng, chói sáng, chồng, lấp hoặc che nhau, do đó cần có phương pháp tương tự watershed nhưng hiệu quả cao hơn để phân đoạn. Do đó, việc ứng dụng mô hình mạng nơ-ron nhân tạo Mask R-CNN để giải quyết bài toán là hết sức cần thiết, kết quả thực nghiệm đã cho thấy tính khả thi của phương pháp đề xuất trong giải quyết bài toán phát hiện và nhận dạng hình dáng loại viên thuốc. Cụ thể:

3.3.1. Ưu điểm

Phương pháp nhận dạng hình dáng loại viên thuốc dựa trên mô hình học sâu Mask R-CNN với bộ dữ liệu viên thuốc CURE đã thu được kết quả tương đối cao, tỉ lệ dự đoán mặt nạ viên thuốc IoU đạt 94,13%, cho thấy hướng phát triển khả quan khi áp dụng vào bài toán nhận dạng viên thuốc trên các ảnh thực tế do người dùng cung cấp. Một số ưu điểm nổi bật mà luận văn đã đóng góp như sau:

- Trên bộ dữ liệu này, phương pháp đề xuất đã thu được kết quả thực nghiệm với độ chính xác đạt **94.13%** cao hơn phương pháp của *Suiyi Ling và cộng sự năm 2020* [9] thực nghiệm mô hình đa luồng và một số phương pháp khác.
- Đã chú thích mức độ pixel cho bộ dữ liệu hình ảnh viên thuốc với hơn **1.700** viên thuốc, là tiền đề rất lớn cho các bài toán phân đoạn và nhận dạng viên thuốc chính xác cao trong tương lai.
- Qua thực nghiệm cho thấy, phương pháp dựa trên hình học trong nhận dạng hình dáng viên thuốc chưa thực sự hiệu quả, chịu ảnh hưởng lớn từ kết quả tìm đường bao quanh đối tượng, khó liệt kê hết và chính xác các hình, đặc biệt, không chính xác khi các viên thuốc có hình đặc biệt như trái tim, hoa... hay cạnh viên không được làm nét, mịn; đồng thời, phương pháp nhận dạng hình dáng loại viên thuốc bằng đối sánh mẫu cũng chịu

ảnh hưởng lớn từ việc phân đoạn ảnh bằng các bộ lọc tiền xử lý truyền thống nên việc áp dụng cho bài toán thực tế trong đời sống sẽ kém hiệu quả.

3.3.2. Một số hạn chế

Bên cạnh những kết quả đã đạt được, phương pháp đề xuất còn tồn tại một số khó khăn, hạn chế, như:

- Mặc dù chương trình đã được huấn luyện trên những lớp hình dạng viên thuốc phổ biến nhất hiện nay, tuy nhiên, độ đa dạng về hình dáng loại viên thuốc chưa cao (hiện có 09 lớp hình dáng, thiếu nhiều lớp viên thuốc hình dáng đặc biệt), khi áp dụng vào thực tiễn cần bổ sung thêm nhiều dữ liệu mẫu và chú thích điểm ảnh.
- Một số lớp hình dạng loại viên thuốc có số lượng hình ảnh hạn chế, do đó hiệu quả của mô hình khi nhận dạng những viên thuốc có hình dạng này chưa cao.
- Nền tảng áp dụng mô hình Mask R-CNN cần được cải tiến, ví dụ như sử dụng các thư viện Pytorch, Tensorflow... mới hơn để tăng hiệu năng, hiệu quả của chương trình.

KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN TIẾP

Trong phạm vi nghiên cứu, luận văn đã trình bày bài toán nhận dạng hình dáng loại viên thuốc; trong đó, tập trung vào một số kỹ thuật nhằm giải quyết bài toán, như: phân đoạn viên thuốc bằng phương pháp tìm đường biên truyền thống; phân đoạn và nhận dạng hình dáng viên thuốc bằng mô hình học sâu. Từ đó đề xuất xây dựng chương trình thực nghiệm thông qua phương pháp sử dụng mô hình Mask R-CNN, xây dựng mô hình huấn luyện của hơn **1.700** hình ảnh viên thuốc thuộc **09** loại hình dáng khác nhau, từ đó thu được độ chính xác nhận dạng đạt xấp xỉ **94,13%**.

Bám sát mục tiêu, nhiệm vụ, sử dụng đúng đắn các phương pháp nghiên cứu khoa học, luận văn đã thu được một số thành công và về cơ bản đã đạt được mục tiêu, nhiệm vụ đặt ra. Trong tương lai, để cải thiện hiệu suất của hệ thống, luận văn định hướng mở rộng nghiên cứu tăng cường bộ dữ liệu mẫu, giảm nhiễu; đồng thời, cần xem xét, nghiên cứu ứng dụng những thành công của các phương pháp tiên tiến hiện đại bắt nguồn từ thị giác máy tính, khai thác dữ liệu lớn và khoa học máy tính, ứng dụng các kỹ thuật mới để phát hiện, nhận dạng đối tượng.

Hướng phát triển tiếp: Việc phát hiện và nhận dạng hình dáng loại viên thuốc là nền tảng hết sức quan trọng cho lĩnh vực Tin Sinh học, đặc biệt là vấn đề nghiên cứu phân loại viên thuốc. Do đó, để cải tiến hiệu quả, trong tương lai có thể nghiên cứu các kỹ thuật học sâu mới hơn, nâng cấp, cải thiện thuật toán Mask R-CNN; đồng thời, nghiên cứu, phát triển việc phân loại viên thuốc dựa trên màu sắc, ký hiệu viên thuốc, quá đó giúp đưa việc nhận dạng viên thuốc trở thành hệ thống hoàn chỉnh hơn, không chỉ dừng lại ở bài toán nhận dạng hình dáng loại viên thuốc.

Luận văn là công trình nghiên cứu công phu, nghiêm túc, song do đây là vấn đề khó và phức tạp, phạm vi nghiên cứu rộng, cộng thêm những khó khăn khách quan, cũng như kiến thức còn hạn chế nên chắc chắn còn nhiều khiếm khuyết. Rất mong nhận được sự quan tâm, góp ý của các nhà khoa học, nhà hoạt động thực tiễn và đồng nghiệp. Cuối cùng, xin chân thành cảm ơn các đơn vị liên quan, các đồng chí, đồng nghiệp, đặc biệt là thầy hướng dẫn khoa học đã tận tình giúp đỡ để hoàn thành luận văn này./.

TÀI LIỆU THAM KHẢO

- [1] A. Jacobson, “Medication errors statistics 2022,” *SingleCare Team*, 01/2022.
- [2] Tariq, Rayhan A.; Vashisht, Rishik; Sinha, Ankur; Scherbak, Yevgeniya, “Medication Dispensing Errors And Prevention,” *NCBI*, 01/2021.
- [3] Kapil G Zirpe, Bhavika Seta, Sharvari Gholap, “Incidence of Medication Error in Critical Care Unit of a Tertiary Care Hospital: Where Do We Stand?,” *PMC*, 2020.
- [4] Mandal, Manav, “Introduction to Convolutional Neural Networks (CNN),” *Data Science Blogathon.*, 2021.
- [5] Girshick, Ross, “Fast r-cnn.,” volume 10.1109/ICCV.2015.169. , 2015.
- [6] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” *arXiv*, 2016.
- [7] R. Girshick, “Mask R-CNN,” *Facebook AI Research (FAIR)*, 2018.
- [8] Hyuk-Ju Kwon, Hwi-Gang Kim, Sung-Hak Lee, “Pill Detection Model for Medicine Inspection Based on Deep Learning,” *chemosensors - MDPI*, 2021.
- [9] Suiyi Ling et al, “Few-Shot Pill Recognition,” *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* , volume doi: 10.1109/CVPR42600.2020.00981, pp. 9786-9795, 2020.
- [10] World Health Organization, Medication Errors, ISBN 978-92-4-151164-3, 2016.
- [11] A. Woodbury, “Increasing Medication Safety with Deep Learning Image Recognition,” *RxVision*, 2020.
- [12] Wong, Y. F. et al., “Development of fine-grained pill identification algorithm using deep convolutional network,” *J. Biomed. Inform.* 74, 2017.
- [13] S. Tangwattananuwat, “The Identification of Pill Images Using Convolutional,” 2020.
- [14] M.A.V. Neto, J.W.M. de Souza, P.P. Reboucas Filho and W.D.O. Antonio, “CoforDes: An invariant feature extractor for the drug pill identification,” *IEEE 31st Int. Symp. on Computer-Based Medical Systems (CBMS), Karlstad*, 2018.
- [15] N. L. o. Medicine, “nlm.nih.gov,” 2016. [Trực tuyến]. Available: https://www.nlm.nih.gov/pubs/techbull/ma16/brief/ma16_pill_challenge.html.

- [16] J.S. Wang, A. Ambikapathi, Y. Han, S.L. Chung, H.W. Ting and C.F. Chen, "Highlighted deep learning based identification of pharmaceutical blister packages," *IEEE 23rd Int. Conf. on Emerging Technologies and Factory Automation (ETFA)*, Turin, 2018.
- [17] Maier, Andreas & Syben, Christopher & Lasser, Tobias & Riess, Christian, "A Gentle Introduction to Image Segmentation for Machine Learning," *A gentle introduction to deep learning in medical image processing. Zeitschrift für Medizinische Physik.*, volume 29, 10.1016/j.zemedi.2018.12.003. , 2021.
- [18] N. Barla, "The Complete Guide to Panoptic Segmentation," PerceptronAI, <https://www.v7labs.com/blog/panoptic-segmentation-guide>, 2022.
- [19] Arpan Kumar, Anamika Tiwari , "A Comparative Study of Otsu Thresholding and K-means Algorithm of Image Segmentation," *International Journal of Engineering and Technical Research (IJETR)*, volume 9, number 5, p. 1, 2019.
- [20] H. Bandyopadhyay, "An Introduction to Image Segmentation: Deep Learning vs. Traditional," V7, 2022.
- [21] Michielan, Lisa & Terfloth, Lothar & Gasteiger, Johann & Moro, Stefano, "Comparison of Multilabel and Single-Label Classification Applied to the Prediction of the Isoform Specificity of Cytochrome P450 Substrates.," *Journal of chemical information and modeling*, volume 49. 10.1021/ci900299a., pp. 2588-605, 2009.
- [22] Gershenson, Carlos, "Artificial Neural Networks for Beginners," 2003.
- [23] Radhamadhab Dalai, Kishore Kumar Senapati, "Comparison of Various RCNN techniques for Classification of Object from Image," *International Research Journal of Engineering and Technology (IRJET)*, volume 04, 2017.
- [24] N. Usuyama, L. Naoto, "ePillID Dataset: A Low-Shot Fine-Grained Benchmark for Pill Identification," *arXiv*, 2020.
- [25] Chen, J. Yu and Z., "Accurate system for automatic pill recognition using imprint information," *IET Image Process.*, volume 9, 2015.
- [26] S. Prasad, "analytixlabs," 2022. [Trực tuyến]. Available: <https://www.analytixlabs.co.in/blog/what-is-image-segmentation/>. [access 2022].
- [27] K V, Lalitha & .R, Amrutha & Michahial, Stafford , "Implementation of Watershed Segmentation," *IJARCCCE*, volume 5, number 10.17148/IJARCCCE.2016.51243. , pp. 196-199, 2016.

- [28] Olaf Ronneberger, Philipp Fischer, Thomas Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” *MICCAI*, 2015.
- [29] Sachin Mehta, Mohammad Rastegari, Linda Shapiro, and Hannaneh Hajishirzi, “Espnetv2: A light-weight, power efficient, and general purpose convolutional neural network,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, p. pages 9190–9200, 2019.