



XÂY DỰNG CÔNG CỤ PHÁT HIỆN BẢN ĐỒ VIỆT NAM VI PHẠM CHỦ QUYỀN QUỐC GIA TRONG TÀI LIỆU ẢNH

Nguyễn Trần Linh Tâm¹, Lê Việt Phương²

¹ VNPT An Giang

² Trường Cao đẳng nghề An Giang

Thông tin chung:

Ngày nhận bài: 23/07/2024

Ngày nhận kết quả bình

duyet:

21/11/2024

Ngày chấp nhận đăng:

01/2025

Title: Building a tool to detect Viet Nam map in violation of national sovereignty in document images

Keywords:

Document Image Classification, CNN, VGG16, InceptionV3, ResNet50, Transfer learning

Từ khóa:

Phân loại ảnh, Mạng nơ-ron tích chập, VGG16, InceptionV3, ResNet50, Học chuyển giao

ABSTRACT

Recently, the country has continuously discovered images of a map of Vietnam with a cow-tongue line that falsely represents national sovereignty is embedded in documents and publications that violate Vietnam's law on border sovereignty. This causes many difficulties in licensing the publication of documents and publications of the Department of Information and Communications of An Giang province. Therefore, it is necessary to there is a need form a tool to automatically detect Vietnam map images containing the image of the cow- tongue line in the image document applying for a publishing license.

In this paper, we study and use a convolutional neural network (Convolutional Neraul Network - CNN) with models VGG16, InceptionV3 and ResNet50 models combined with transfer learning to detect Vietnam map images containing cow's tongue line images in image documents. Experimental results have shown that these three models are feasible, however, ResNet50 model achieved accuracy. Experimenting on a set of 2,000 image documents, ResNet50 model gives accurate results over 95%. The high detection probability of the ResNet50 model demonstrates the model's suitability and is also the basis for building tools for practical implementation at the Department of Information and Communications of An Giang province.

TÓM TẮT

Gần đây, trong nước liên tục phát hiện hình ảnh bản đồ Việt Nam có đường lưỡi bò thể hiện không đúng chủ quyền quốc gia được cài cắm vào trong các tài liệu, ấn phẩm vi phạm pháp luật Việt Nam về chủ quyền biên giới. Điều này gây rất nhiều khó khăn trong công tác cấp phép xuất bản tài liệu ấn phẩm của Sở Thông tin và Truyền thông tỉnh An Giang. Do đó rất cần có một công cụ tự động phát hiện các hình ảnh bản đồ Việt Nam có chứa ảnh đường lưỡi bò trong tài liệu ảnh xin cấp phép xuất bản.

Trong bài báo này chúng tôi nghiên cứu sử dụng mạng nơ-ron tích chập (Convolutional Neraul Network - CNN) với các mô hình VGG16, InceptionV3 và ResNet50 kết hợp học chuyển giao để phát hiện ảnh bản đồ Việt Nam có chứa ảnh đường lưỡi bò trong tài liệu ảnh. Kết quả thực nghiệm

đã cho thấy ba mô hình này đều khả thi, tuy nhiên mô hình ResNet50 cho kết quả tốt nhất, thực nghiệm trên tập 2.000 tài liệu ảnh, thì mô hình ResNet50 cho kết quả chính xác trên 95%. Xác suất phát hiện cao của mô hình ResNet50 thể hiện sự phù hợp của mô hình và cũng là cơ sở để xây dựng công cụ để triển khai thực tế tại Sở Thông tin và Truyền thông tỉnh An Giang.

1. GIỚI THIỆU

Tài liệu ảnh (Document Image) bao gồm các thành phần: ký tự và phần không phải ký tự. Thành phần không phải ký tự có thể là biểu bảng, ảnh,... và được gọi chung là đối tượng đồ họa (Graphical Object). Đối tượng đồ họa đóng vai trò quan trọng trong việc hiểu nội dung hay trích xuất thông tin của tài liệu. Vì vậy, việc phát hiện đối tượng đồ họa trở thành một đề tài được nhiều nhà nghiên cứu quan tâm do tầm quan trọng của bài toán và sự thách thức về độ khó mà chúng mang lại. Do chúng không có một cấu trúc, hình dạng và đặc trưng cụ thể và dường như việc đề xuất một phương pháp cho phát hiện tất cả đối tượng đồ họa là vô cùng khó khăn.

Bài toán phân loại tài liệu ảnh (Document Image Classification) đã xuất hiện từ lâu và đã có rất nhiều phương pháp, bài báo, công trình nghiên cứu được đưa ra nhằm cải thiện thuật toán, nâng cao tính hiệu quả, độ chính xác cho kết quả phân loại. Bắt đầu với các thuật toán sử dụng học máy (machine learning), chủ yếu phân tích nghiên cứu và thử nghiệm các phương pháp trích đặc trưng nhằm chọn ra phương pháp tối ưu nhất để đưa vào bộ phân loại. Kết quả thu được tương đối khả quan. Những năm gần đây cùng với sự phát triển mạnh mẽ của phần cứng máy tính cũng như sự bùng nổ dữ liệu trên Internet, một nhánh nghiên cứu của học máy là học sâu (deep learning) đã đạt được nhiều thành tựu đáng kể, đặc biệt trong lĩnh vực xử lý ảnh và ngôn ngữ tự nhiên.

Các hệ thống xử lý ảnh lớn đã khai thác mô hình mạng nơ-ron tích chập và đưa vào sản phẩm của

mình những chức năng thông minh như nhận diện khuôn mặt người dùng (Facebook), phát triển xe hơi tự lái (Google), giao hàng tự động (Amazon), phân loại ảnh y tế (Nima Tajbakhsh, Jae Y. Shin, Suryakanth R. Gurudu, & Jianming Liang, 2017; ZhiFei Lai and HuiFang Deng, 2017; AhmedT. Sahlol, DaliaYousri, AhmedA. Ewees, & MohamedAbd Elaziz, 2020). Vào hai ngày 14 và 15/12/2019 trong khuôn khổ Zalo AI Summit 2019 đã có 70 đội với hơn 150 kỹ sư tham dự cuộc thi Zalo AI Hackathon với chủ đề phát hiện đường lười bò phi pháp bằng AI. Ngoài ra, trong nước cũng có nhiều công trình nghiên cứu được ứng dụng trong nhiều lĩnh vực phân loại, xử lý ảnh tại Việt Nam có thể kể đến như: tác giả Lê Việt Phương tiếp cận bài toán đối sánh ảnh về đặc trưng cục bộ (Le Viet Phuong, Nayef, Visani, Ogier, & Tran Cao De, 2014; Le Viet Phuong, & cs, 2015) và nghiên cứu “Logo detection, recognition and spotting in context by matching local visual features” (Lê Việt Phương, 2015, tr.1).

Hiện trạng việc kiểm duyệt các tài liệu ấn phẩm xin cấp phép xuất bản tại Sở Thông tin và Truyền thông tỉnh An Giang đang thực hiện thủ công; thời hạn giải quyết 15 ngày là một thách thức không nhỏ cho đơn vị cấp phép về áp lực thời gian và độ chính xác trong kiểm tra, thẩm định. Bởi nhiệm vụ cán bộ kiểm tra là phải duyệt về nội dung tiêu đề, mục lục, các đoạn tóm tắt,... tiếp đó cần kiểm duyệt nội dung từng mục, bố cục, hình ảnh trong tài liệu và thực hiện một công việc quan trọng là rà soát các hình ảnh liên quan

đến các hình ảnh cảm lưu hành theo quy định của pháp luật như hình ảnh bản đồ Việt Nam có chứa đường lưỡi bò vi phạm chủ quyền quốc gia (sau đây gọi tắt là bản đồ có đường lưỡi bò) (liên quan đến lĩnh vực cấp phép xuất bản, năm 2019 các tác giả Phạm Thị Minh Thư và Lê Việt Phương cũng đã có công trình nghiên cứu liên quan, đó là “Xây dựng công cụ hỗ trợ đọc xuất bản phẩm lưu chiếu dựa trên phương pháp đối sánh ảnh tài liệu”).

Xuất phát từ nhu cầu thực tế cùng với các thành công trong các nghiên cứu về thị giác máy tính trong phân loại ảnh, bài báo đề xuất nghiên cứu ứng dụng các mô hình học sâu hiện đại để xây dựng công cụ phát hiện bản đồ có đường lưỡi bò trong tài liệu ảnh. Công cụ này khi hoàn thành sẽ được triển khai tại Bộ phận cấp phép xuất bản Sở Thông tin và Truyền thông tỉnh An Giang, góp phần giảm bớt phần nào về áp lực thời gian và tăng độ chính xác trong khi thực thi nhiệm vụ của các cán bộ tại đây.

Nội dung tiếp theo sẽ giới thiệu chi tiết nội dung giải thuật xử lý và phân loại ảnh để phát hiện bản đồ có đường lưỡi bò trong tài liệu ảnh. Các kết quả thực nghiệm và thảo luận được trình bày ở Mục 3, Kết luận và hướng phát triển sẽ được nêu ở Mục 4 của bài báo.

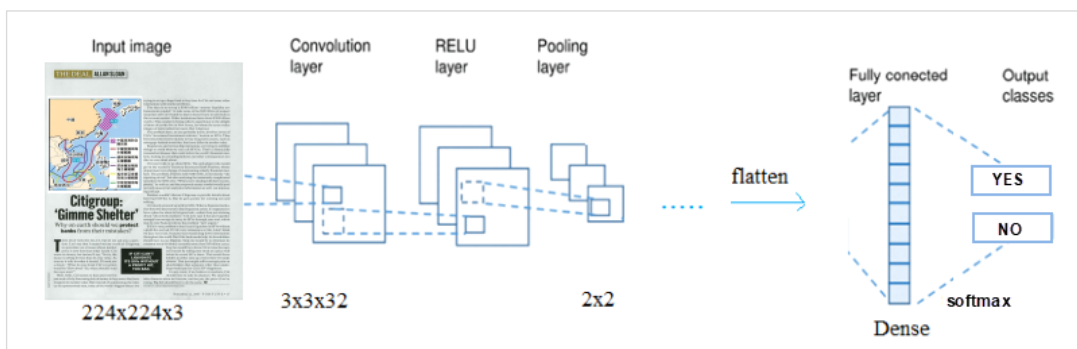
2. NỘI DUNG NGHIÊN CỨU

Trong bài báo này tập trung nghiên cứu nội dung xây dựng mô hình học sâu kết hợp học chuyển giao để phân loại tài liệu ảnh thành 2 lớp: lớp các tài liệu ảnh chứa bản đồ có đường lưỡi bò và lớp các tài liệu ảnh không có bản đồ có đường lưỡi bò. Trên cơ sở mô hình nhận được, thực hiện xây dựng công cụ phát hiện tài liệu ảnh có bản đồ có đường lưỡi bò hay không.

2.1 Xây dựng mô hình

Có nhiều phương pháp để phân loại ảnh. Trong bài báo, phương pháp được chọn là mạng nơ-ron tích chập bởi các ưu điểm vượt trội của nó đã được đề cập ở trên.

2.1.1 Sơ đồ mô hình mạng nơ-ron tích chập



Hình 1. Mô hình mạng nơ-ron tích chập

Một mô hình mạng nơ-ron tích chập có đầu vào là một tài liệu ảnh với các pixel thô và biến đổi nó thông qua các lớp tích chập (convolution), lớp kích hoạt phi tuyến ReLU, lớp gộp Pooling và lớp liên kết đầy đủ (fully connected) để cho đầu ra là tài liệu ảnh thuộc 2 lớp là: lớp các tài liệu ảnh có bản đồ có đường lưỡi bò và lớp các tài liệu ảnh không có bản đồ có đường lưỡi bò.

2.1.2 Hàm kích hoạt sử dụng (Activation function)

Tại các lớp tích chập, sử dụng ReLU là hàm kích hoạt. Hàm này có công thức dễ thực hiện tính toán và hiệu quả với nhiều loại bài toán, với tốc độ thực hiện nhanh dẫn đến thời gian huấn luyện mô hình tương đối nhanh so với hàm kích hoạt khác.

Tại tầng liên kết đầy đủ cuối cùng, sử dụng hàm

Softmax. Hàm Softmax thường được sử dụng ở tầng đầu ra, nhằm đánh giá xác suất nhận phân loại của dữ liệu đầu vào của tầng đầu ra.

2.1.3 Lớp làm phẳng (Flatten)

Lớp này có nhiệm vụ chuyển đổi kết quả đầu ra từ lớp tích chập là mảng nhiều chiều và chuyển đổi thành vec-tơ một chiều trước khi được vào tầng kết nối đầy đủ để thực hiện quá trình phân loại.

2.1.4 Xây dựng mô hình huấn luyện kết hợp học chuyển giao

Do dữ liệu đầu vào dùng để huấn luyện mô hình tương đối ít (5.000 ảnh cho mỗi lớp) khó có thể tìm kiếm và mở rộng thêm nên các mô hình huấn luyện từ chúng sẽ khó có thể dự báo tốt. Để tận dụng lại tri thức từ các pretrained-model với cùng tác vụ phân loại sẽ giúp các mô hình được huấn luyện dự báo tốt hơn với dữ liệu mới, nghiên cứu đề xuất xây dựng mô hình huấn luyện bằng các mạng CNN kết hợp với học chuyển giao (transfer learning) trên tập dữ liệu ImageNet (tập dữ liệu chứa 1,2 triệu ảnh với 1.000 thể loại khác nhau).

Với yêu cầu bài toán đặt ra, chúng tôi sử dụng phương pháp Fine tuning để huấn luyện lại, chỉ giữ lại phần ConvNet trong CNN và bỏ đi FCs. Sau đó thêm các Fully Connected layer mới vào output của ConvNet. Khi train model được chia làm 2 giai đoạn:

Giai đoạn 1: Vì các fully connected layer mới thêm vào có các hệ số được khởi tạo ngẫu nhiên tuy nhiên các layer trong ConvNet của pre-trained model đã được train với ImageNet dataset nên sẽ không train (đóng băng/freeze) trên các layer trong ConvNet.

Giai đoạn 2: unfreeze các layer trên ConvNet của pre-trained model và train trên các layer của ConvNet của pre-trained model và các layer mới. Có thể unfreeze tất cả các layer trong ConvNet của model hoặc chỉ unfreeze một vài layer cuối (tùy vào thời gian và GPU sử dụng).

Tóm lại, quá trình huấn luyện sử dụng học chuyển giao bằng Fine tuning được thực hiện: sau khi lấy ra các đặc điểm của ảnh bằng việc sử dụng ConvNet của pre-trained model, thì ta sẽ coi đây là input của 01 CNN mới bằng cách thêm các ConvNet và Fully Connected layer. Lý do là ConvNet của ImageNet model có thể lấy ra được các thuộc tính chung, nhưng đặc tính khác nên cần thêm 1 số ConvNet mới để học thêm các thuộc tính của hình ảnh cần huấn luyện.

Xử lý dữ liệu đầu vào: ảnh đầu vào là ảnh chụp có nhiều kích thước khác nhau vì vậy cần phải chuyển sang dữ liệu dưới dạng một mảng (array), thứ bậc của quy trình này bao gồm nhận diện địa chỉ ảnh (folder chứa ảnh), thu nhỏ ảnh về một kích thước chung là 224×224 (Vì pre-trained model đã được train với kích thước ảnh cố định, nên khi dùng pre-trained model cần resize lại ảnh có kích ảnh bằng kích thước mà ConvNet của pre-trained model yêu cầu), tạo batch với kích thước 64 (qua thực nghiệm, nhận thấy giá trị này cho kết quả ổn định nhất), xác định loại mảng nhận được trả về được mã hóa “binary”.

Để có thể đối sánh và chọn ra được mô hình huấn luyện tốt nhất, chọn nghiên cứu kết hợp học chuyển giao với 3 mô hình mạng CNN phổ biến hiện nay, đó là VGG16, InceptionV3 và ResNet50.

➤ Mạng VGG16:

Xây dựng mô hình huấn luyện dùng mạng VGG16 có sử dụng pre-trained weights của ImageNet. Thực hiện tinh chỉnh:

- Thêm hai lớp kết nối đầy đủ với 4.096 đơn vị ẩn và hàm kích hoạt ReLU;
- Sử dụng dropout với tỷ lệ là 0,5;
- Thêm một lớp kết nối đầy đủ cuối cùng, sử dụng hàm kích hoạt “softmax”.

➤ Mạng InceptionV3:

Xây dựng mô hình dùng mạng InceptionV3 sử dụng pre-trained weights của ImageNet. Thực

hiện tinh chỉnh:

- Thêm hai lớp kết nối đầy đủ với 4.096 đơn vị ẩn và hàm kích hoạt ReLU;
- Sử dụng dropout với tỷ lệ là 0,5;
- Thêm một lớp kết nối đầy đủ cuối cùng, sử dụng hàm kích hoạt “softmax”.

➤ **Mạng ResNet50:**

Xây dựng mô hình dùng mạng ResNet50 có sử dụng pre-trained weights của ImageNet. Thực hiện tinh chỉnh:

- Thêm hai lớp kết nối đầy đủ với 4.096 đơn vị ẩn và hàm kích hoạt ReLU;
- Sử dụng dropout với tỷ lệ là 0,5;
- Thêm một lớp kết nối đầy đủ cuối cùng, sử dụng hàm kích hoạt “softmax”.

➤ **Huấn luyện mô hình:**

• **Thuật toán tối ưu (Optimizer)**

Nghiên cứu này sử dụng thuật toán tối ưu Adam (Adaptive moment Estimation). Adam là một thuật toán tối ưu có thể biến đổi thích nghi tỷ lệ học (learning rate) và được thiết kế đặc trưng cho các mô hình học sâu. Thuật toán Adam có thể được xem như là sự kết hợp của RMSprop và SGD (Stochastic Gradient Descent). Nó sử dụng giá trị bình phương của gradient để chia tỷ lệ học tập như thuật toán RMSprop và nó tận dụng động lượng (momentum) bằng cách sử dụng trung bình di chuyển của giá trị gradient thay vì gradient như SGD.

• **Hàm lỗi (Loss function)**

Hàm lỗi (hay còn gọi là hàm tổn hao) dùng để đo lường mức độ chính xác của mô hình trong quá trình huấn luyện. Cần giảm thiểu giá trị của hàm này để điều khiển mô hình đi đúng hướng (chỉ số lỗi càng ít, chính xác càng cao). Hàm lỗi sử dụng trong nghiên cứu này là Cross Entropy tính toán độ chênh lệch giữa 2 phân phối xác suất của dự đoán và của nhãn. Nhãn được biến đổi thành một vec-tơ one-hot với tất cả phần tử trong vec-tơ

nhận giá trị 0 trừ một vị trí biểu diễn nhãn nhận giá trị 1.

• **Kỹ thuật Drop-out để giảm Over-fitting**

Khi sử dụng full connected layer, các nơ ron sẽ phụ thuộc lẫn nhau trong suốt quá trình huấn luyện, điều này làm giảm sức mạnh cho mỗi nơ ron và dẫn đến trường hợp over-fitting trên tập dữ liệu huấn luyện, và để cải thiện được vấn đề này sử dụng kỹ thuật dropout. Kỹ thuật drop-out là cách thức giả định một phần các unit (nơ-ron) bị ẩn đi trong quá trình huấn luyện, qua đó làm giảm tích hòa trộn mà dẫn đến hiện tượng over-fitting. Tại mỗi bước trong quá trình huấn luyện, khi thực hiện quá trình lan truyền xuôi (forward propagation) đến tầng sử dụng drop-out, thay vì tính toán tất cả unit có trên tầng đó, tại mỗi unit thực hiện lựa chọn ngẫu nhiên một số unit được thực hiện tính toán dựa trên xác suất p.

2.2 Xây dựng công cụ phát hiện

Chúng tôi sử dụng thư viện mã nguồn mở OpenCV, Tensorflow, Keras framework hỗ trợ xây dựng, huấn luyện mô hình, cài đặt bằng ngôn ngữ Python. Sử dụng Framework Flask xây dựng API để deploy model.

Công cụ phát hiện ảnh bản đồ có đường lưới bỏ trong tài liệu ảnh được thiết kế dưới dạng website theo kiến trúc Client/Server, có 3 tầng như ở Hình 2. Trong đó:

➤ **Tầng thứ nhất** là tầng giao diện người dùng, là ứng dụng website chạy trên internet, cho phép người dùng upload các tài liệu ảnh, gửi lên server và hiển thị kết quả nhận dạng do server gửi về;

➤ **Tầng thứ hai** là tầng server quản lý cấu hình hệ thống, ví dụ cấu hình giao thức gửi/nhận dữ liệu với client, giao thức sử dụng trong hệ thống là HTTP;

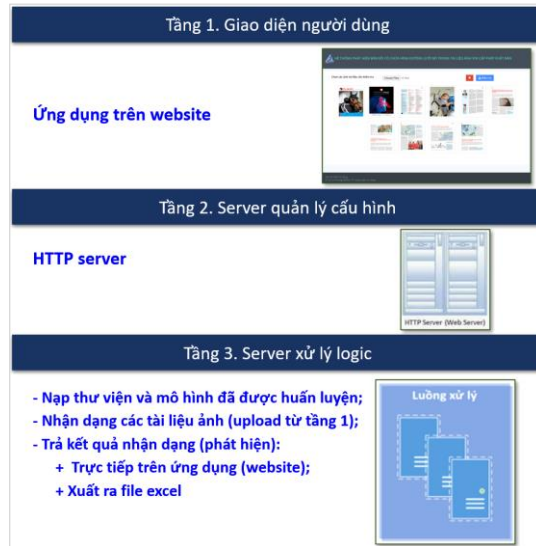
➤ **Tầng thứ ba** là tầng server thực hiện logic xử lý các yêu cầu từ client, như:

- Nạp thư viện và mô hình nhận dạng đã được

huấn luyện;

- Nhận dạng các tài liệu ảnh được upload từ tầng thứ nhất bằng API giao tiếp với mô hình nhận dạng;

- Trả kết quả nhận dạng (phát hiện) bằng cách: hiển thị kết quả lên website và xuất kết quả nhận dạng ra file excel.



Hình 2. Kiến trúc Hệ thống phát hiện ảnh bản đồ Việt Nam

có hình đường lưới bỏ trong tài liệu ảnh

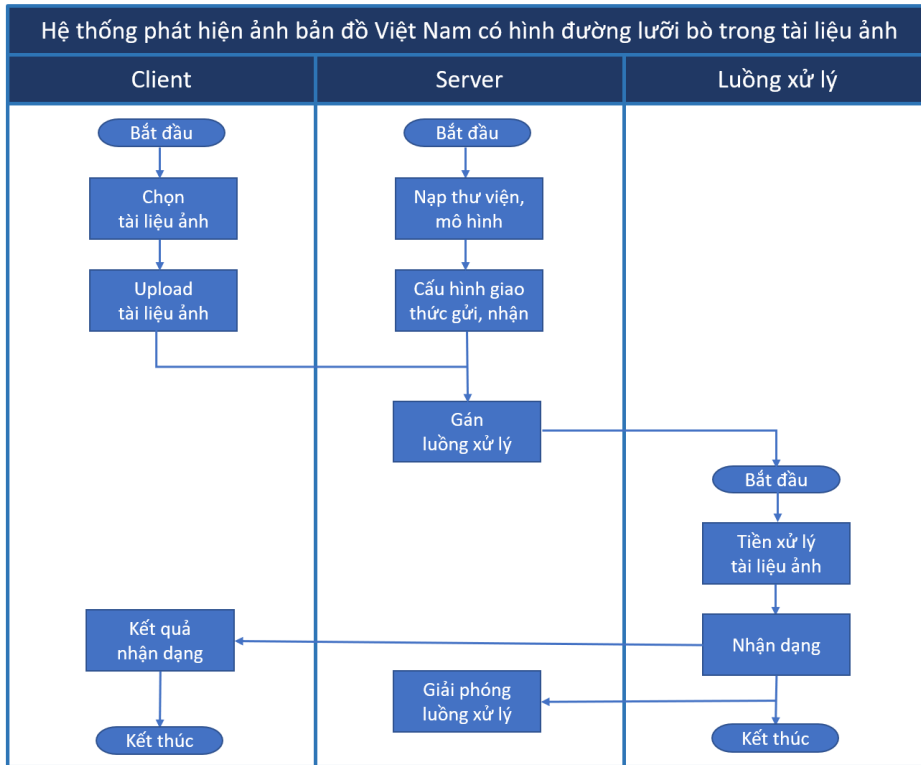
Luồng hoạt động chính của hệ thống thể hiện trong Hình 3, trong đó các bước thực hiện của server và client từ lúc khởi động ban đầu tới lúc kết thúc như sau:

➤ **Client (ứng dụng trên website):**

- 1) Người dùng khởi động ứng dụng;
- 2) Người dùng upload các tài liệu ảnh cần kiểm tra;
- 3) Ứng dụng đợi nhận kết quả nhận dạng (phát hiện đối tượng) từ máy chủ gửi về và hiển thị cho người dùng.

➤ **Chương trình Server:**

- 1) Chương trình được khởi động và nạp các thư viện cần thiết;
- 2) Chương trình nạp mô hình nhận dạng đã được huấn luyện trước đó;
- 3) Bắt đầu quá trình tiền xử lý tài liệu ảnh, tính toán nhận dạng bằng mô hình trên;
- 4) Trả kết quả nhận dạng về cho ứng dụng client.



Hình 3. Luồng hoạt động chính của hệ thống

3. KẾT QUẢ THỰC NGHIỆM, THẢO LUẬN

3.1 Xây dựng tập dữ liệu

Cần xây dựng tập dữ liệu huấn luyện gồm 2 lớp, được đặt tên như sau:

- **Lớp YES:** bao gồm các tài liệu ảnh có chứa bản đồ đường lưới bò;
- **Lớp NO:** bao gồm các tài liệu ảnh không chứa bản đồ có đường lưới bò.

Tuy nhiên, qua tìm hiểu thực tế, việc thu thập dữ liệu từ các nguồn sách, báo, tài liệu, hình ảnh trên Internet thì thường tìm được dưới dạng ảnh bản đồ Việt Nam có đường lưới bò, nhưng ít tìm thấy các tài liệu ảnh có chứa bản đồ Việt Nam có đường lưới bò (thuộc lớp YES). Nên chúng tôi đề xuất phương án xây dựng dữ liệu giả lập để khởi tạo các tài liệu ảnh theo 2 lớp YES và lớp

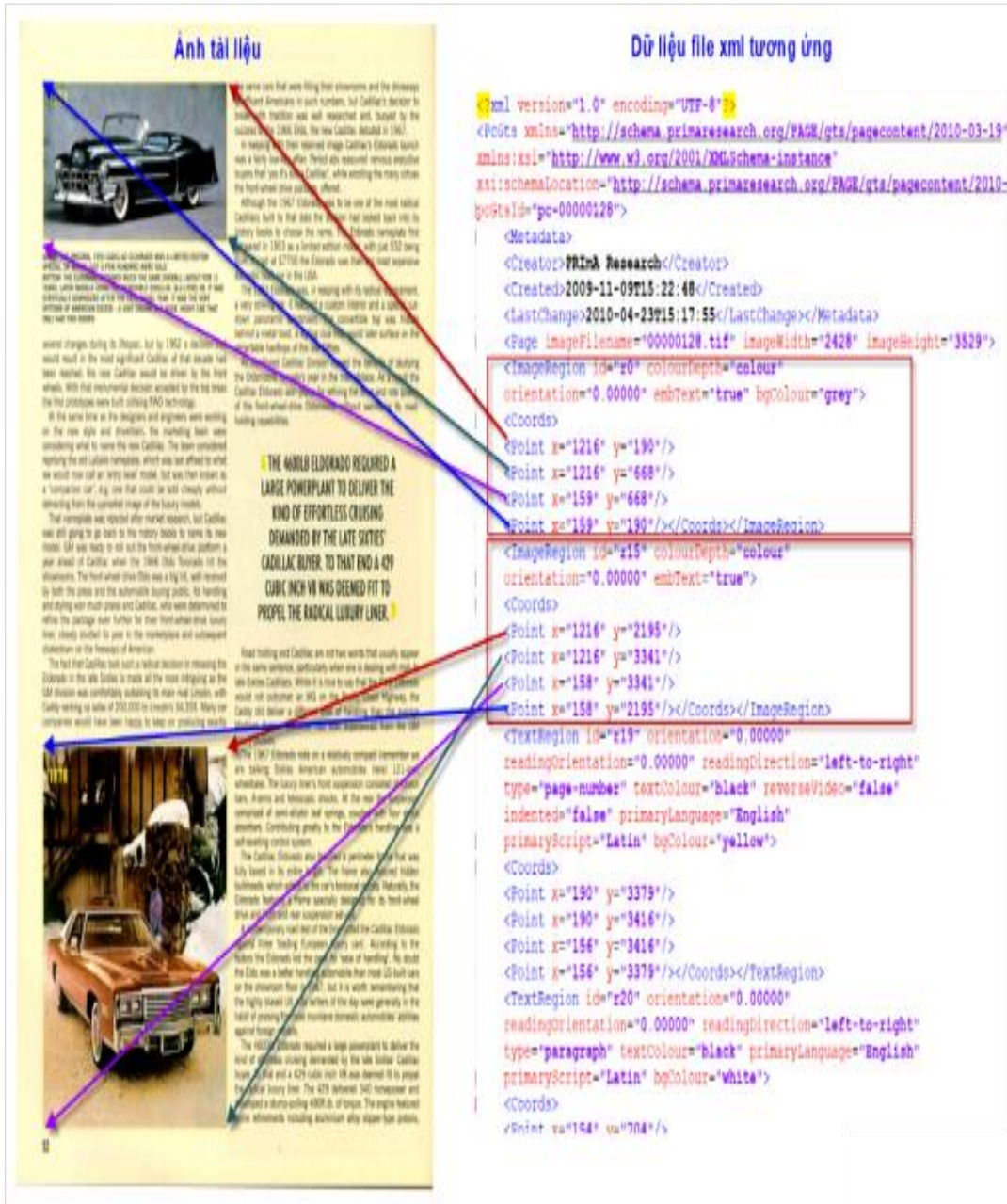
NO. Phương án đề xuất như sau:

✓ **Bước 1:** Sử dụng các tài liệu ảnh trong tập dữ liệu PRIMa của Đại học Salford (Manchester), tập dữ liệu này chứa các tài liệu ảnh được scan từ các tài liệu thực tế, từ các tạp chí, bài báo,... Trong các tài liệu ảnh có nhiều bố cục khác nhau, gồm: các vùng ảnh văn bản và các vùng chứa hình ảnh; các tài liệu ảnh có ground-truth để dễ dàng tách và chèn các đối tượng. Trong giới hạn của đề tài này, thì nguồn tài liệu ảnh tìm được là phù hợp để xây dựng bộ dữ liệu.

- Nguồn:

<https://www.primaresearch.org/dataset>.

- Dữ liệu gồm: 293 file tài liệu ảnh và 293 file xml tương ứng trong file lưu vị trí tọa độ 4 đỉnh của đối tượng của đối tượng ảnh trong tài liệu như Hình 4.

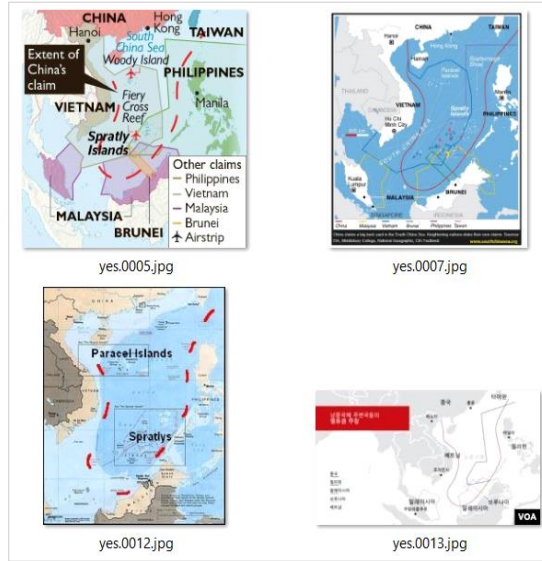


Hình 4. Mô tả 1 tài liệu ảnh và tọa độ các đối tượng ảnh trong file xml

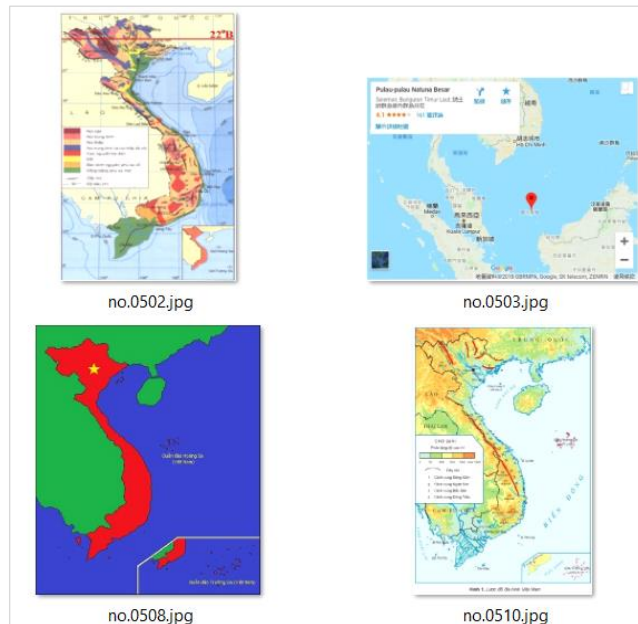
✓ **Bước 2:** tìm kiếm và thu thập các hình ảnh từ nguồn internet, hình ảnh thuộc 2 lớp:

- Tập thứ nhất gồm **1.000** ảnh bản đồ Việt Nam có chứa đường lưới bò (để ghép vào tài liệu ảnh thu được ở bước 1, hình thành lớp YES) (Hình 5);

- Tập thứ hai gồm **1.000** ảnh theo các nội dung: bản đồ không có đường lưới bò, và các hình ảnh khác (phong cảnh, đồ vật, con người, con vật, ...) (để ghép vào tài liệu ảnh thu được ở bước 1, hình thành lớp NO) (Hình 6).



Hình 5. Một số ảnh trong danh sách 1.000 ảnh bản đồ Việt Nam có chứa đường lưỡi bò (lớp YES)



Hình 6. Một số ảnh trong danh sách 1.000 ảnh bản đồ không chứa đường lưỡi bò (lớp NO)

✓ **Bước 3:** phân chia tập dữ liệu thu thập được ở Bước 2 thành 3 tập dữ liệu.

Bằng cách lấy ngẫu nhiên các ảnh, thực hiện phân chia tập dữ liệu thành 3 tập: tập dữ liệu

huấn luyện (train) chiếm 64%, tập dữ liệu dùng đánh giá mô hình (validation) chiếm 16% và tập dữ liệu dùng kiểm tra (test) chiếm 20%. Xem kết quả lưu ở Bảng 1.

Bảng 1. Chia dữ liệu thành 3 tập train, validation và test

Tập dữ liệu	Lớp	Thư mục	Tỷ lệ	Số lượng (ảnh)
Huấn luyện (train)	YES	train\YES	64%	640
	NO	train\NO		640
Đánh giá (validation)	YES	val\YES	16%	160
	NO	val\NO		160
Kiểm tra (test)	YES	test\YES	20%	200
	NO	test\NO		200
Tổng	1.000 ảnh lớp YES			
	1.000 ảnh lớp NO			

✓ **Bước 4:** do mục tiêu bài toán muốn phân loại, phát hiện các tài liệu ảnh có chứa đối tượng ảnh là các bản đồ có đường lười bò hay không, đề xuất phương án chọn ghép ảnh đối tượng ảnh vào các tài liệu ảnh để mô hình huấn luyện có thể nhận dạng được các đối tượng ảnh. Sau một số lần thử nghiệm với kết quả tương đối ổn định, nghiên cứu đề xuất chọn ghép lần lượt từng ảnh đối tượng từ tập dữ liệu có được ở Bảng 1 lần lượt với 5 tài liệu ảnh có được ở Bước 1, theo quy tắc sau:

- Chọn 1 ảnh đối tượng từ các thư mục ở Bảng 1.
- Tiếp theo, tìm ngẫu nhiên trong thư mục tài liệu ảnh, chọn file tài liệu ảnh thỏa điều kiện về kích thước (căn cứ theo tọa độ đối tượng ảnh lưu ở

file xml tương ứng) như sau: (Chiều rộng đối tượng/Chiều rộng ảnh trong tài liệu ảnh) × 100 nằm trong khoảng từ 50% đến 120%. Thực hiện:

- Tăng/giảm kích thước đối tượng ảnh ghép vào theo tỷ lệ trên;
- Ghép đối tượng vào vị trí trên tài liệu ảnh, theo tọa độ trên file xml tương ứng;
- Canh giữa đối tượng ghép vào vùng ảnh trên tài liệu ảnh, phần dư sau khi ghép sẽ được tô trắng.

- Thực hiện lặp lại các bước trên, đến khi tìm đủ 5 tài liệu ảnh để ghép ảnh đối tượng vào.

Kết quả tập dữ liệu dùng huấn luyện mô hình:

Bảng 2. Dữ liệu của 3 tập train, validation và test để huấn luyện mô hình

Tập dữ liệu	Lớp	Thư mục	Tỷ lệ	Số lượng (ảnh)
Huấn luyện (train)	YES	train\YES	64%	3.200
	NO	train\NO		3.200
Đánh giá (validation)	YES	val\YES	16%	800
	NO	val\NO		800
Kiểm tra (test)	YES	test\YES	20%	1.000
	NO	test\NO		1.000
Tổng				1.000 ảnh lớp YES
				1.000 ảnh lớp NO



Hình 7. Tài liệu ảnh sau khi ghép các ảnh đối tượng vào

3.2 Kết quả thực nghiệm

Với tập dữ liệu thu thập được, sau khi chuẩn hóa về kích thước chung 224×224, được dùng làm dữ liệu đầu vào cho các mô hình huấn luyện đã trình bày. Để có cơ sở khách quan nhất trong việc so sánh, đánh giá, lựa chọn mô hình huấn luyện sử dụng cho hệ thống phát hiện đối tượng ảnh bản đồ có đường lười bò trong tài liệu ảnh theo yêu cầu bài toán đặt ra. Thực hiện 10 lần huấn luyện, đối với 03 mô hình mạng VGG16, InceptionV3, ResNet50, cụ thể như sau:

- Mỗi lần lặp chọn tập dữ liệu đầu vào bằng cách chọn ngẫu nhiên theo Bước 3 và Bước 4 ở Mục 3.1.
- Lần lượt thực hiện huấn luyện trên tập dữ

liệu đó với 03 mô hình mạng VGG16, InceptionV3, ResNet50.

Khai báo giá trị các thông số trong quá trình huấn luyện trên 03 mô hình mạng nói trên, như sau:

- Khai báo hàm loss là “binary_crossentropy”, sử dụng thuật toán tối ưu “Adam”;
- Tỷ lệ học (learning rate) bằng 0,01 và độ đo khai báo “Accuracy” là độ chính xác của mô hình;
- Batch size = 64;
- Num epochs = 30.

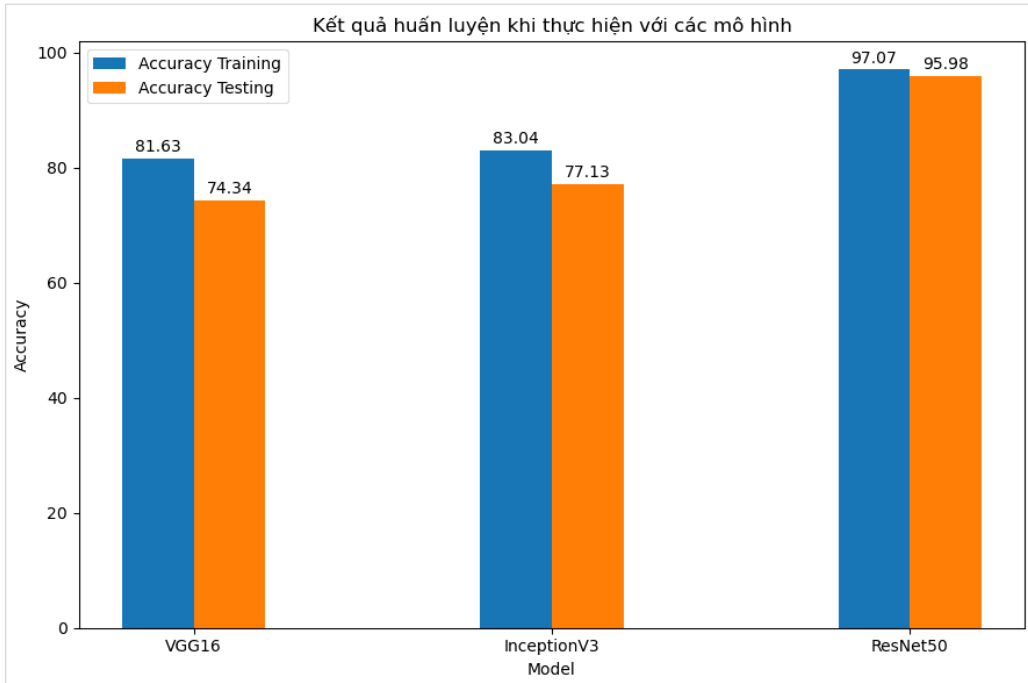
Kết quả huấn luyện các mô hình lần lượt đối với 10 lần chọn tập dữ liệu đầu vào ngẫu nhiên được thể hiện ở Bảng 3.

Bảng 3. Kết quả 10 lần huấn luyện của 3 mô hình VGG16, InceptionV3 và ResNet50

Lần thực hiện	Mô hình huấn luyện	Training accuracy	Testing accuracy
1	VGG16	83,06%	76,30%
	InceptionV3	84,25%	78,18%
	ResNet50	97,69%	96,63%
2	VGG16	81,94%	73,60%
	InceptionV3	83,06%	75,44%
	ResNet50	98,00%	95,17%

Lần thực hiện	Mô hình huấn luyện	Training accuracy	Testing accuracy
3	VGG16	81,81%	72,35%
	InceptionV3	82,00%	78,86%
	ResNet50	97,75%	96,21%
4	VGG16	81,19%	73,50%
	InceptionV3	83,75%	76,16%
	ResNet50	97,50%	96,21%
5	VGG16	82,56%	73,00%
	InceptionV3	83,44%	77,54%
	ResNet50	96,31%	96,91%
6	VGG16	82,38%	75,40%
	InceptionV3	83,81%	78,29%
	ResNet50	97,19%	95,93%
7	VGG16	80,75%	72,45%
	InceptionV3	84,44%	79,47%
	ResNet50	96,31%	96,78%
8	VGG16	79,69%	73,90%
	InceptionV3	82,00%	76,85%
	ResNet50	97,00%	94,86%
9	VGG16	82,69%	76,40%
	InceptionV3	81,13%	74,55%
	ResNet50	96,88%	95,37%
10	VGG16	81,19%	76,45%
	InceptionV3	82,56%	75,96%
	ResNet50	96,06%	95,77%
Trung bình	VGG16	81,63%	74,34%
	InceptionV3	83,04%	77,13%
	ResNet50	97,07%	95,98%

Bảng 3 có thể hiện giá trị trung bình accuracy train và test của 10 lần huấn luyện đối với các mô hình huấn luyện tương ứng; được thể hiện ở biểu đồ như Hình 8.



Hình 8. Kết quả (trung bình) huấn luyện khi thực hiện với các mô hình

Qua đồ thị biểu diễn trung bình kết quả huấn luyện mô hình ở Hình 13 và kết quả huấn luyện hình ở Bảng 3, có thể thấy rằng mô hình ResNet50 cho kết quả huấn luyện tốt nhất và ổn định nhất. Nghiên cứu sẽ chọn ResNet50 làm mô hình huấn luyện để xây dựng công cụ theo yêu cầu bài toán đặt ra.

Để khẳng định việc chọn mô hình huấn luyện ResNet50 là tốt hơn so với các mô hình đã được đề cập. Chúng tôi sử dụng các độ đo như: Accuracy, Precision, Recall, và bảng Ma trận nhầm lẫn (Confusion matrix) để đánh giá hiệu suất của bài toán phân loại ảnh. Để ước lượng các độ đo này, có thể dựa vào bảng 4:

3.3 Tiêu chí đánh giá

Bảng 4. Bảng ma trận nhầm lẫn (Confusion matrix)

Lớp dự đoán \ Lớp nhãn	Lớp dự đoán	
	NO	YES
NO	TN	FP
YES	FN	TP

Trong lớp dự đoán, ứng với mỗi tài liệu ảnh được ánh xạ vào một tập hợp {NO,YES}, với YES là tích cực (positive) ứng với tài liệu ảnh có chứa bản đồ có đường lưỡi bò, và NO là tiêu cực (negative) ứng với tài liệu ảnh không có bản đồ có đường lưỡi bò. Đối với lớp nhãn (phân loại),

sử dụng nhãn {NO,YES} để mô tả, với YES là phân loại tài liệu ảnh có chứa bản đồ có đường lưỡi bò và NO là phân loại tài liệu ảnh không có bản đồ có đường lưỡi bò.

3.4 Kết quả và đánh giá

Qua thực nghiệm 10 lần huấn luyện mô hình

ResNet50 với tập dữ liệu đầu vào ngẫu nhiên (của từng lần) và thực hiện kiểm chứng mô hình đã huấn luyện trên với dữ liệu kiểm chứng (từ tập dữ liệu test theo cách chọn ở 2 Bước 3 và Bước 4 của Mục 3.1). Kết quả huấn luyện trung bình với độ đo khác như Accuracy, Precision và Recall như sau:

Accuracy: 95,98 (lấy test accuracy trung bình

của mô hình ResNet50).

Từ kết quả kiểm chứng với chỉ số đánh giá là độ chính xác phân lớp của mô hình, nhận được ở mức trung bình tương đương 95,98% với bộ dữ liệu kiểm chứng. Xem xét với độ đo khác như Precision (trung bình) và Recall (trung bình) từ kết quả ở Bảng 5:

Bảng 5. Kết quả kiểm chứng của mô hình ResNet50

Accuracy score: 0,9598				
Report:				
	precision	recall	f1-score	support
NO	0,96	0,96	0,96	1.000
YES	0,96	0,96	0,96	1.000
accuracy			0,96	
macro avg	0,96	0,96	0,96	2.000
weighted avg	0,96	0,96	0,96	2.000

Kết quả nhận được từ 3 độ đo trên thể hiện ở Bảng 6.

Bảng 6. Kết quả phân loại của mô hình ResNet50

Nhân	Precision	Recall
NO	0,96	0,96
YES	0,96	0,96
Trung bình	0,96	0,96

Với số trung bình đo độ Precision cho thấy trong dữ liệu kiểm chứng thì 96% dữ liệu được phân loại đúng nhãn. Trung bình độ đo Recall cho thấy là mẫu dữ liệu có nhãn phân loại được phân loại đúng nhãn của nó với xác suất là 96%.

Để đánh giá được mô hình cụ thể hơn cần xem kết quả phân loại trên từng nhãn. Kết quả phân loại đúng, sai của từng nhãn của 10 lần thực hiện huấn luyện được trình bày ở Bảng 7.

Bảng 7. Kết quả phân loại theo từng nhãn ở 10 lần thực hiện huấn luyện mô hình ResNet50

Lần thực hiện	Nhãn phân loại	Phân loại đúng	Phân loại sai	Tổng số mẫu
1	NO	966	34	1.000
	YES	967	33	1.000
2	NO	943	57	1.000

Lần thực hiện	Nhãn phân loại	Phân loại đúng	Phân loại sai	Tổng số mẫu
3	YES	961	39	1.000
	NO	964	36	1.000
	YES	960	40	1.000
4	NO	963	37	1.000
	YES	961	39	1.000
5	NO	982	18	1.000
	YES	955	45	1.000
6	NO	944	56	1.000
	YES	975	25	1.000
7	NO	977	23	1.000
	YES	959	41	1.000
8	NO	948	52	1.000
	YES	946	54	1.000
9	NO	953	47	1.000
	YES	955	45	1.000
10	NO	945	55	1.000
	YES	971	29	1.000
Trung bình	NO	958	42	1.000
	YES	961	39	1.000

Tại lớp có nhãn YES thì mô hình phân loại đúng 961 mẫu trên tất cả 1.000 mẫu, tương đương với 96,1%. Với lớp có nhãn phân loại là NO, mô hình phân loại đúng 958 mẫu trên tổng số 1.000 mẫu, chiếm 95,8%.

Nhận xét: Với các kết quả đạt được trên cho thấy độ chính xác phân loại của mô hình tương đối ổn định.

3.5 Ảnh hưởng của tỷ lệ học (learning rate) đến mô hình

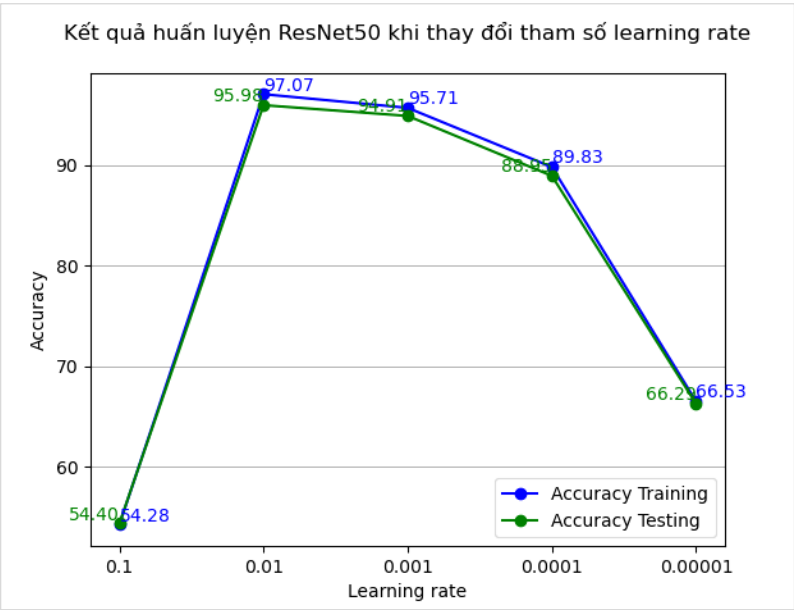
Bảng 8. Bảng kết quả tổng hợp huấn luyện mô hình ResNet50 khi thay đổi tỷ lệ học

Trong thử nghiệm này, tiến hành thay đổi tỷ lệ học để xem sự ảnh hưởng của nó đến độ chính xác của mô hình ResNet50. Thực hiện 10 lần quá trình huấn luyện với tập dữ liệu đầu vào ngẫu nhiên (của từng lần) được chọn (theo cách chọn ở 2 Bước 3 và Bước 4 của Mục 3.1).

Kết quả 10 lần huấn luyện mô hình ResNet50, với các tỷ lệ học khác nhau thể hiện trong Bảng 8.

Lần lập	lr = 0,1	lr = 0,01		lr = 0,001			lr = 0,0001		lr = 0,00001	
	Accuracy (%)	Accuracy (%)		Accuracy (%)			Accuracy (%)		Accuracy (%)	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
1	50,75	49,66	97,69	96,63	96,56	95,65	89,81	90,01	68,00	69,77
2	53,94	54,19	98,00	95,17	95,94	93,75	91,06	88,83	65,63	65,43
3	51,81	51,31	97,75	96,21	96,06	94,84	88,94	89,24	64,63	66,50
4	50,31	50,71	97,50	96,21	95,81	95,61	87,94	88,86	63,94	61,20
5	55,88	57,11	96,31	96,91	95,31	95,57	90,50	87,80	71,31	68,16
6	65,31	66,68	97,19	95,93	95,75	94,66	91,12	88,97	60,94	64,84
7	53,06	51,11	96,31	96,78	95,31	95,95	91,69	88,67	68,38	65,30
8	56,00	60,04	97,00	94,86	95,31	93,95	87,81	88,30	71,25	70,82
9	54,75	53,33	96,88	95,37	95,69	94,72	89,50	89,30	66,63	66,11
10	50,94	48,84	96,06	95,77	95,31	94,39	89,88	89,47	64,63	64,80
Trung bình	54,28	54,40	97,07	95,98	95,71	94,91	89,83	88,95	66,53	66,29

Lấy trung bình 10 lần thực nghiệm khi huấn luyện mô hình ResNet50 với các tham số tỷ lệ học lần lượt là 0,1; 0,01; 0,001; 0,0001 và 0,00001 cho kết quả như biểu đồ ở Hình 9.



Hình 9. Kết quả huấn luyện mô hình ResNet50 khi thay đổi tham số tỷ lệ học

Từ biểu đồ ở Hình 9 cho thấy rằng tỷ lệ học ảnh hưởng rất lớn đến độ chính xác của mô hình. Với các thử nghiệm tỷ lệ học khác nhau độ chính xác cao nhất khi tỷ lệ học bằng 0,01. Độ chính xác giảm nhanh khi tỷ lệ học quá lớn hoặc quá nhỏ. Điều này cho thấy khi chọn tỷ lệ học cần phải

chọn hợp lý. Khi chọn quá lớn sẽ dẫn đến mô hình không hội tụ còn khi quá nhỏ dẫn đến mô hình hội tụ chậm.

3.6 Thảo luận, đề xuất

3.6.1. Thảo luận

Bằng thực nghiệm, có thể đánh giá mô hình ResNet50 cho kết quả phát hiện tương đối tốt trong đa số các trường hợp. Tuy nhiên, bên cạnh đó đôi khi vẫn còn một số trường hợp phát hiện cho kết quả chưa chính xác, đó là các trường hợp phát hiện sai (tài liệu ảnh không có bản đồ có

đường lưỡi bò, nhưng hệ thống lại phát hiện có) và không phát hiện đối tượng (tài liệu ảnh có bản đồ có đường lưỡi bò, nhưng hệ thống không phát hiện được). Các trường hợp phát hiện chưa chính xác, bao gồm:

✓ Chất lượng ảnh của các đối tượng ghép vào tài liệu ảnh (không rõ ràng, nhiều chi tiết phức tạp,...): đôi khi hệ thống không phát hiện được nếu đối tượng không rõ ràng; trường hợp ảnh có nhiều chi tiết phức tạp thì hệ thống phát hiện sai hoặc không phát hiện được.



Hình 10. Không phát hiện khi đối tượng không rõ nét



Hình 11. Phát hiện sai (ảnh trái) và không phát hiện được (ảnh phải)

Khi đối tượng trong tài liệu ảnh có nhiều chi tiết phức tạp

- ✓ Ảnh đối tượng có kích thước nhỏ: thỉnh thoảng hệ thống phát hiện sai nếu đối tượng có kích thước nhỏ.



Hình 12. Phát hiện sai khi đối tượng ảnh có kích thước nhỏ

- ✓ Các tài liệu ảnh có chứa ảnh đối tượng có các chi tiết gần tương đồng ảnh đường lười bò (như đường hàng hải, đường hàng không, hướng

gió, hướng bão, có hình dạng gần giống đường lười bò,...): đôi khi hệ thống sẽ phát hiện sai.



Hình 13. Phát hiện sai khi ảnh đối tượng có các chi tiết như: đường bay (ảnh trái), đường hàng hải (ảnh phải)



Hình 14. Phát hiện sai khi ảnh đối tượng có các chi tiết như: hướng gió (ảnh trái), hướng bão (ảnh phải)



Hình 15. Phát hiện sai khi đối tượng có các chi tiết giống như đường lưới bò

✓ Kích thước tài liệu ảnh đưa vào nhận dạng có khổ giấy ngang (Landscape): một số trường

hợp hệ thống không thể phát hiện được hoặc phát hiện sai.



Hình 16. Không phát hiện được đối tượng khi tài liệu ảnh có khổ giấy ngang (Landscape)



Hình 17. Phát hiện sai khi tài liệu ảnh có khổ ngang (Landscape)

* **Tóm lại**, có thể thấy rằng các nguyên nhân làm cho mô hình ResNet50 cho kết quả phát hiện đối tượng sai và không phát hiện như bảng 9:

Bảng 9. Nguyên nhân mô hình phát hiện đối tượng sai và không phát hiện

Tài liệu ảnh	Phát hiện sai	Không phát hiện được
Chứa đối tượng không rõ nét		×
Chứa đối tượng có nhiều chi tiết phức tạp	×	×
Chứa đối tượng có kích thước nhỏ	×	
Chứa đối tượng gần giống đường lười bò	×	
Khô&ng ảnh tài liệu xoay ngang (Landscape)	×	

Mục tiêu bài toán là phát hiện đối tượng (detect object), nên quan tâm nhiều đến vấn đề làm thế nào để giảm tỷ lệ nhầm lẫn khi không phát hiện được đối tượng (tài liệu ảnh cần kiểm tra có ảnh bản đồ có đường lưỡi bò, nhưng hệ thống không phát hiện được), riêng đối với các trường hợp phát hiện sai (tài liệu ảnh cần kiểm tra không có đường lưỡi bò, nhưng hệ thống phát hiện có) sẽ ít ảnh hưởng đến nhu cầu thực tế của bài toán đặt ra (do phát hiện nhầm vẫn tốt hơn bỏ sót trường hợp).

Nhìn vào Bảng 9 có thể thấy mô hình hoạt động khá hiệu quả, đa số các trường hợp hệ thống nhầm lẫn là do phát hiện sai đối tượng.

3.6.2. Đề xuất

Trong quá trình 10 lần thực nghiệm, bằng quan sát nhận thấy kết quả phát hiện sai hoặc không phát hiện thường là do chất lượng đối tượng ghép, kích thước tài liệu ảnh. Do đó, để cải thiện độ, nâng cao chính xác cho mô hình huấn luyện, đề xuất các phương pháp xử lý sau:

Thứ nhất, cải thiện chất lượng ảnh các đối tượng ghép vào tài liệu ảnh.

Thứ hai, tìm nguồn các tài liệu ảnh có khổ giấy ngang (Landscape), bổ sung vào tập dữ liệu huấn luyện.

Thứ ba, thu thập các đối tượng ảnh có chi tiết phức tạp, và các đối tượng ảnh gần tương đồng nhau để làm phong phú thêm cho tập dữ liệu huấn luyện thuộc 2 lớp YES và NO.

4. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Bài báo đã đề xuất sử dụng mô hình huấn luyện ResNet50 kết hợp framework Flask để xây dựng một công cụ tự động phát hiện các hình ảnh bản đồ Việt Nam có đường lưỡi bò trong các tài liệu xin cấp phép xuất bản; mô hình này có độ chính xác trên 95%. Công cụ này hỗ trợ công tác cấp phép xuất bản, giúp tự động hóa trong việc tìm phát hiện các trang tài liệu ảnh có chứa hình ảnh không được phép xuất bản đó là các bản đồ Việt

Nam có đường lưỡi bò tại Bộ phận cấp phép xuất bản Sở Thông tin và Truyền thông tỉnh An Giang. Qua đó phần giảm bớt phần nào về áp lực thời gian và tăng độ chính xác trong khi thực thi nhiệm vụ của các cán bộ tại đây. Thời gian tới, nhóm tác giả sẽ tiếp tục nghiên cứu cải thiện, nâng cao độ chính xác của mô hình; thử nghiệm với một số cơ sở dữ liệu khác, phát hiện: ảnh bản đồ Việt Nam không có các quần đảo Trường Sa và Hoàng Sa; ảnh bản đồ Việt Nam có tên các đảo và quần đảo không đúng.

TÀI LIỆU THAM KHẢO

- AhmedT. Sahlol, DaliaYousri, AhmedA. Ewees, MohammedA.A.Al-qaness, Robertas Damasevicius & MohamedAbd Elaziz., (2020). COVID-19 image classification using deep features and fractional-order marine predators algorithm. *Scientific Reports*. Retrieved from <https://www.nature.com/articles/s41598-020-71294-2?proof=t>.
- Sahlol, A.T., Yousri, D., Ewees, A.A., Al-qaness, M.A.A, Damasevicius, R., & Elaziz . M.(2020). COVID-19 image classification using deep features and fractional-order marine predators algorithm. *Scientific Reports*,10 (1), 15364. <https://doi.org/10.1038/s41598-020-71294-2>.
- Le Viet Phuong. (2015). Logo Detection, Recognition and Spotting in Context by Matching Local Visual Features. Retrieved from <https://tel.archives-ouvertes.fr/tel-01373417>.
- Le Viet Phuong, Nayef, N., Visani, M., Ogier, J. M., & Trần Cao Đê., (2014). Document retrieval based on logo spotting using key-point matching. In *Proceedings - International Conference on Pattern Recognition* (pp. 3056– 3061). IEEE. Retrieved from <https://doi.org/10.1109/ICPR.2014.527>.

- Le Viet Phuong, & Tran Cao De., (2015). Key-point matching with post-filter using SIFT and BRIEF in logo spotting. In *The 2015 IEEE RIVF International Conference on Computing & Communication Technologies - Research, Innovation, and Vision for Future (RIVF)* (pp. 89–93). IEEE. Retrieved from <https://doi.org/10.1109/RIVF.2015.7049880>
- Confusion Matrix / Ma trận nhầm lẫn / Ma trận lỗi (19/08/2021).
Retrieved from <https://viblo.asia/p/confusion-matrix-ma-tran-nham-lan-ma-tran-loi-V3m5WQB7ZO7>.
- Nima Tajbakhsh., Jae Y. Shin., Suryakanth R., Gurudu, R., Todd Hurst., Christopher B. Kendall., Michael B. Gotway., & Jianming Liang. (2017). Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Trans. Med. Image*,35(5),1299-1312.
Retrieved from <https://ieeexplore.ieee.org/abstract/document/7426826>.
- Phạm Thị Minh Thư. & Lê Việt Phương. (2019). Xây dựng công cụ hỗ trợ đọc xuất bản phẩm lưu chiều dựa trên phương pháp đối sánh ảnh tài liệu. *AGU International Journal of Sciences – 2019*, 23 (2), 96 – 108.
- ZhiFei Lai.& HuiFang Deng. (2018). Medical Image Classification Based on Deep Features Extracted by Deep Model and Statistic Feature Fusion with Multilayer Perceptron. *Computational Intelligence and Neuroscience*. Retrieved from <https://www.hindawi.com/journals/cin/2018/2061516>.