

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

KHOA KHOA HỌC MÁY TÍNH



BÁO CÁO ĐỒ ÁN CUỐI KỲ

MÔN HỌC: XỬ LÝ ẢNH VÀ ỨNG DỤNG

ĐỀ TÀI: FOREST FIRE AND SMOKE DETECTION

GVHD: TS. CÁP PHẠM ĐÌNH THẮNG

SINH VIÊN THỰC HIỆN:

HỒ HỒNG HÀ – 20520480

TRƯƠNG VĂN CHINH – 20521137

ĐINH QUANG ĐÔNG – 20521189

Thành phố Hồ Chí Minh, tháng 1 năm 2024

Mục lục

CHƯƠNG 1: GIỚI THIỆU ĐỀ TÀI	3
1.1. Giới thiệu	3
1.2. Input, output của bài toán	3
CHƯƠNG 2: PHƯƠNG PHÁP THỰC HIỆN	4
2.1. YOLOv8	4
2.2. Faster-RCNN	7
2.3. Dataset	8
CHƯƠNG 3: HUẤN LUYỆN MÔ HÌNH VÀ KẾT QUẢ.....	8
3.1. Huấn luyện mô hình	8
3.1.1. YOLOv8	8
3.1.2. Faster-RCNN	9
3.2. Kết quả thực nghiệm.....	9
CHƯƠNG 4: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN	9
4.1. Kết luận.....	9
4.2. Khó khăn.....	10
4.3. Hướng phát triển.....	11

CHƯƠNG 1: GIỚI THIỆU ĐỀ TÀI

1.1. Giới thiệu

Cháy rừng là một trong những thảm họa thiên nhiên nguy hiểm và tốn kém nhất, gây nguy hiểm nghiêm trọng đến tài sản, tính mạng con người và ảnh hưởng xấu đến cân bằng sinh thái tự nhiên của rừng. Việc phát hiện chính xác và kịp thời các đám cháy rừng là rất quan trọng để thực hiện các biện pháp chữa cháy rừng, đồng thời giảm thiểu thiệt hại về vật chất, con người và thiên nhiên do cháy rừng gây ra.

Rừng không chỉ đóng vai trò quan trọng trong việc cung cấp hàng hóa vật chất và dịch vụ môi trường, mà còn góp phần quan trọng vào sự phát triển công nghiệp và kinh tế nông thôn. Rừng cung cấp nguồn gỗ, nhựa, cao su, thực phẩm, thuốc men và nhiều sản phẩm khác. Đồng thời, rừng cung cấp các dịch vụ môi trường quan trọng cho toàn cầu, như tạo ra oxy qua quá trình quang hợp và giảm hiệu ứng nóng lên toàn cầu bằng cách hấp thụ khí CO₂.

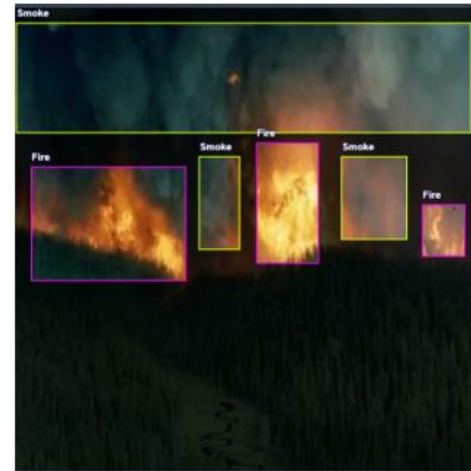
Khi rừng bị cháy, nó không chỉ gây ra thiệt hại về tài sản và tính mạng con người, mà còn tạo ra một loạt vấn đề môi trường và khí hậu. Khói từ đám cháy rừng có thể thải ra khí CO₂ và các chất gây ô nhiễm khác, góp phần vào biến đổi khí hậu và hiện tượng nóng lên toàn cầu. Vì vậy, việc phát hiện sớm cháy rừng là một nhiệm vụ cấp bách. Tuy nhiên, phát hiện cháy rừng ở giai đoạn sớm đôi khi gặp khó khăn do nhiều yếu tố như sự cô lập của khu vực, khả năng tiếp cận hạn chế, điều kiện thời tiết khắc nghiệt và thiếu nhân lực giám sát. Để vượt qua những thách thức này, một mô hình phát hiện cháy rừng tự động được tích hợp vào các thiết bị drone và máy bay không người lái (UAV) có thể đóng vai trò quan trọng trong việc phát hiện sớm và kịp thời các đám cháy và khói.

1.2. Input, output của bài toán

- Input: Một tấm ảnh số hoặc frame trong video chứa cảnh quan sát về rừng.
- Output: Tấm ảnh số đã được vẽ lên bounding box để đánh dấu vị trí của lửa hoặc khói (nếu xuất hiện trong ảnh), cùng với nhãn tương ứng cho mỗi bounding box.
- Ví dụ:



Input

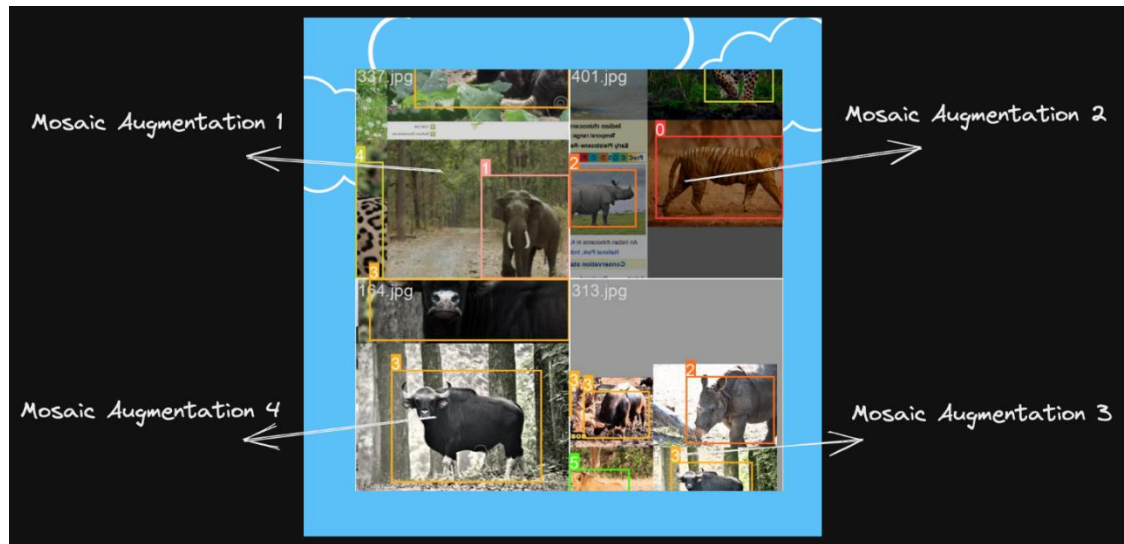


Output

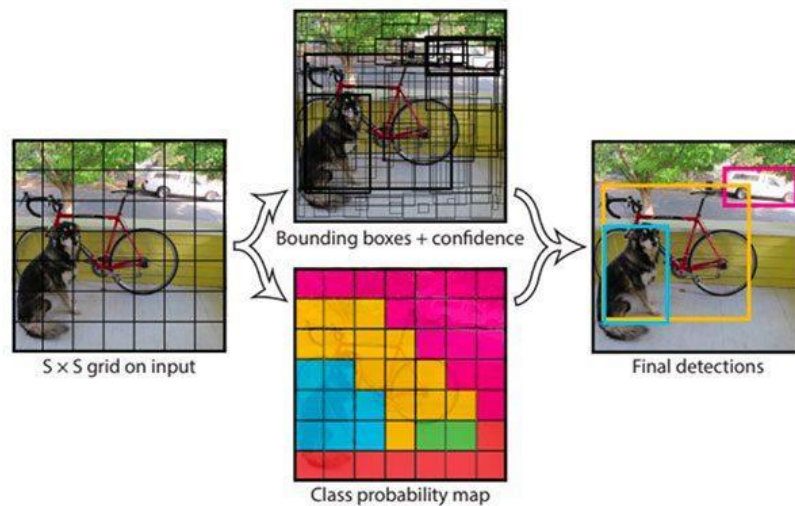
CHƯƠNG 2: PHƯƠNG PHÁP THỰC HIỆN

2.1. YOLOv8

- YOLOv8 là phiên bản mới nhất trong dòng mô hình YOLO. YOLO là viết tắt của cụm từ “You Only Look Once”, tức chỉ khả năng dự đoán tất cả các đối tượng có mặt trong một hình ảnh chỉ trong một lần truyền dữ liệu.
- Điểm khác biệt chính mà YOLO mang lại là cách đặt vấn đề. Tác giả đã đặt lại vấn đề phát hiện đối tượng dưới dạng một bài toán hồi quy (dự đoán tọa độ hộp giới hạn) thay vì phân loại.
- Các mô hình YOLO được huấn luyện trước trên các bộ dữ liệu lớn như COCO và ImageNet. Điều này cho phép chúng vừa có khả năng cung cấp dự đoán cực kỳ chính xác với các lớp đã được huấn luyện, vừa có thể học các lớp mới một cách tương đối dễ dàng.
- Các mô hình YOLO cũng có tốc độ huấn luyện nhanh hơn, với độ chính xác cao và kích thước mô hình nhỏ. Chúng có thể được huấn luyện trên các GPU đơn, do đó dễ tiếp cận hơn đối với các nhà phát triển.
- Được công bố vào đầu năm 2023, YOLOv8 đã mang lại nhiều điểm tích cực so với phiên bản tiền nhiệm, như phát hiện không dùng anchor, giới thiệu lớp tích chập C3 và tăng cường mosaic.



- YOLOv8 là mô hình SOTA mã nguồn mở được xây dựng và duy trì bởi nhóm Ultralytics. Nó được phân phối dưới Giấy phép Công cộng GNU, cho phép người dùng chia sẻ, sửa đổi và phân phối phần mềm một cách tự do. Cộng đồng YOLOv8 đang phát triển và rất sôi động.
- Hoạt động của YOLOv8

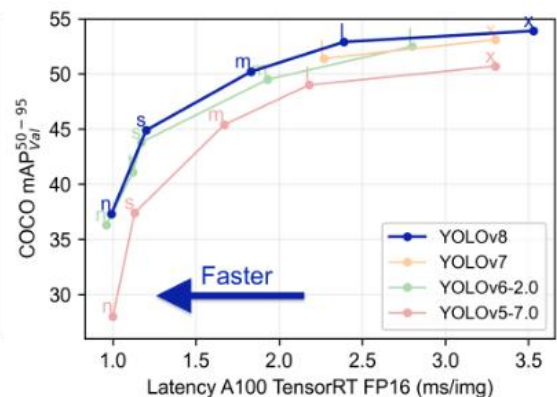
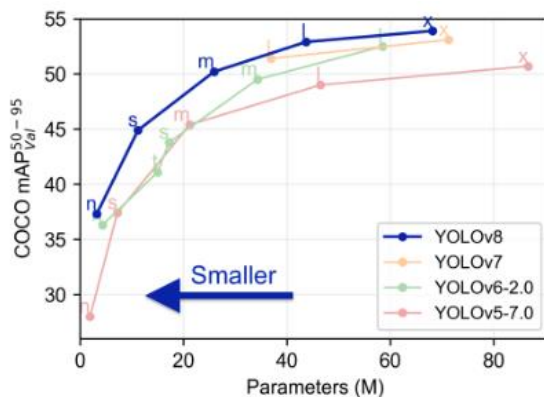


- Input Image: Thuật toán lấy hình ảnh đầu vào và thay đổi kích thước của nó thành kích thước cố định phù hợp để xử lý. Kích thước này thường được xác định dựa trên kiến trúc mạng được sử dụng.
- Grid Division: Hình ảnh được chia thành một lưới S x S, trong đó mỗi ô lưới chịu trách nhiệm dự đoán các đối tượng có trong ô đó.
- Bounding Box Prediction: Trong mỗi ô lưới, YOLO dự đoán nhiều Bounding Box. Mỗi Bounding Box được xác định bởi năm thuộc tính: (x, y, w, h, confidence). Tọa độ (x, y) biểu thị tâm của Bounding Box so với ô lưới, trong khi w và h biểu thị chiều rộng và chiều cao của box. Confidence biết mức độ tin cậy của thuật toán đối với hộp chứa đối tượng.

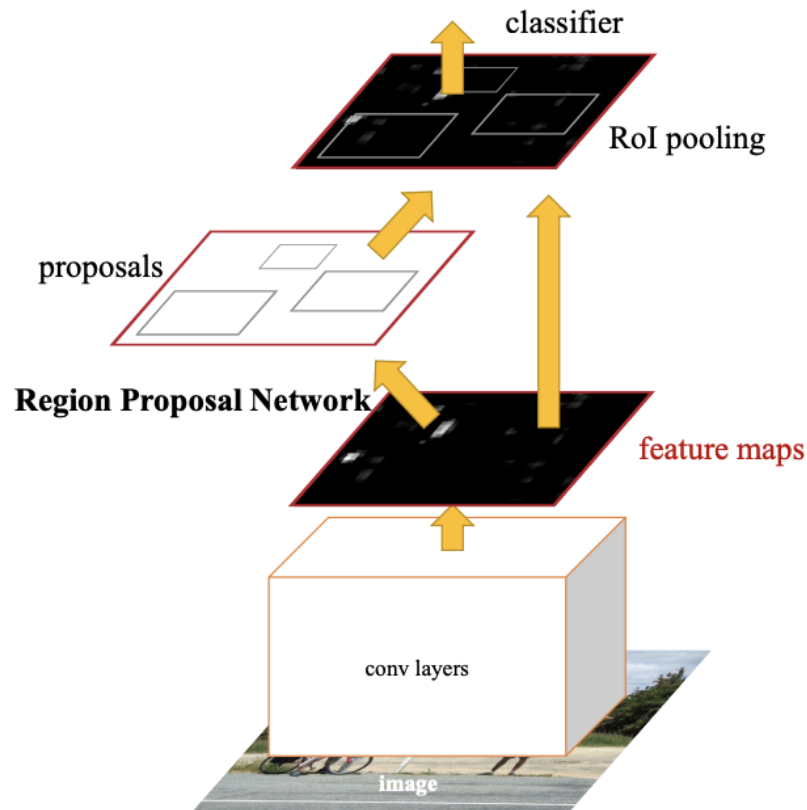
- Class Prediction: cùng với các dự đoán về Bounding Box, YOLO cũng dự đoán xác suất của các lớp khác nhau cho mỗi box. Số lượng xác suất của lớp phụ thuộc vào tập dữ liệu được sử dụng. Các xác suất của lớp này thể hiện khả năng mỗi lớp có mặt trong Bounding Box.
 - Confidence Thresholding: YOLO áp dụng ngưỡng tin cậy để lọc ra những dự đoán có độ tin cậy thấp.
 - Final Output: đầu ra của thuật toán YOLO là một tập hợp các Bounding Box, mỗi Bounding Box được liên kết với nhãn lớp và Confidence. Các Bounding Box này đại diện cho các đối tượng được phát hiện trong ảnh đầu vào.
- YOLOv8 được cung cấp với nhiều phiên bản với kích thước mô hình khác nhau phù hợp với mục đích sử dụng.

Model	size (pixels)	mAP _{val} ⁵⁰⁻⁹⁵	Speed CPU ONNX (ms)	Speed A100 TensorRT (ms)	params (M)	FLOPs (B)
YOLOv8n	640	37.3	80.4	0.99	3.2	8.7
YOLOv8s	640	44.9	128.4	1.20	11.2	28.6
YOLOv8m	640	50.2	234.7	1.83	25.9	78.9
YOLOv8l	640	52.9	375.2	2.39	43.7	165.2
YOLOv8x	640	53.9	479.1	3.53	68.2	257.8

- So với các phiên bản trước thì YOLOv8 có độ chính xác cao hơn tuy nhiên tốc độ chậm hơn.



2.2. Faster-RCNN



- Faster R-CNN (Region-based Convolutional Neural Network) là một mô hình phát hiện đối tượng được sử dụng rộng rãi trong lĩnh vực xử lý ảnh và thị giác máy tính. Mô hình này được giới thiệu bởi Shaoqing Ren, Kaiming He, Ross Girshick và Jian Sun vào năm 2015 nhằm cải thiện hiệu suất và tốc độ của việc phát hiện đối tượng.
- Faster R-CNN là một cải tiến của R-CNN và Fast R-CNN, kết hợp các ưu điểm của cả hai mô hình trước đó. Nó sử dụng sự kết hợp giữa mạng CNN để trích xuất đặc trưng và một mạng RPN (Region Proposal Network) để tạo ra các vùng đề xuất (region proposals) tiềm năng chứa đối tượng.
- Quy trình hoạt động của Faster R-CNN bao gồm ba giai đoạn chính:
 - Quá trình trích xuất đặc trưng: Faster R-CNN sử dụng một mạng CNN như VGG16 hoặc ResNet để trích xuất đặc trưng từ hình ảnh đầu vào. Quá trình này giúp chuyển đổi hình ảnh thành một biểu diễn đặc trưng có khả năng bắt chước các đặc điểm quan trọng của đối tượng.
 - Quá trình tạo ra các vùng đề xuất: Mạng Region Proposal Network (RPN) được sử dụng với đầu vào là các anchor box được định nghĩa trước, mỗi anchor box được định nghĩa trước và được gắn kết với một vị trí trên feature map của ảnh đầu vào. Mạng RPN sẽ phân loại xem anchor box

nào có chứa đối tượng hay không, đồng thời RPN có một bộ hồi quy để dự đoán thay đổi vị trí và kích thước của anchor box sao cho phù hợp với vị trí và kích thước của vật thể. Cuối cùng, RPN sử dụng kết quả từ bộ phân loại và bộ hồi quy để tạo ra các đề xuất vùng quan tâm (region proposals). Các đề xuất này là các hộp giới hạn ứng viên có xác suất cao chứa vật thể và đã được điều chỉnh vị trí. Số lượng đề xuất thường lớn hơn rất nhiều so với số lượng thực tế của các vật thể trong ảnh.

- Sau RPN chúng ta thu được region proposals, rất nhiều trong số chúng bị overlap với nhau nên thuật toán Non-max suppression (NMS) được sử dụng để loại bỏ đi các region proposals bị dư thừa.
- Quá trình phân loại và dự đoán: Sau khi có được các vùng đề xuất, Faster R-CNN sử dụng một bộ phân loại đa lớp (multi-class classifier) để xác định lớp và dự đoán độ chính xác của các đối tượng trong các vùng đề xuất đó. Đồng thời, một bộ hồi quy (regressor) được áp dụng để dự đoán vị trí chính xác của các đối tượng trong vùng đề xuất.
- Faster R-CNN đã đạt được sự thành công lớn trong việc phát hiện và nhận dạng đối tượng trong ảnh với độ chính xác cao và tốc độ nhanh hơn so với các mô hình trước đó. Nó đã trở thành một trong những mô hình phát hiện đối tượng tiêu chuẩn trong cộng đồng nghiên cứu và ứng dụng.

2.3. Dataset

- Dữ liệu nhóm sử dụng là dữ liệu từ <https://universe.roboflow.com/kirzone/fire-iejcs/dataset/2>
- Bộ dữ liệu gồm có 1706 hình ảnh đã được gán nhãn. Được chia thành 3 tập:
 - train - 1194 ảnh
 - val - 340 ảnh
 - test - 172 ảnh

CHƯƠNG 3: HUẤN LUYỆN MÔ HÌNH VÀ KẾT QUẢ

3.1. Huấn luyện mô hình

- Các mô hình đều được huấn luyện trên kaggle với GPU P100

3.1.1. YOLOv8

- YOLOv8n:
 - Backbone: CSPDarknet53
 - Tổng tham số: ~30,000,000
 - Số epoch: 100
- YOLOv8x:
 - Backbone: CSPDarknet53
 - Tổng tham số: ~68,100,000
 - Số epoch: 100

3.1.2. Faster-RCNN

- Backbone: MobileNetv3-Large FPN
- Tổng tham số: ~18,800,000.
- Số epoch: 100

3.2. Kết quả thực nghiệm

	mAP@50	mAP
Faster-RCNN	0.367	0.115
YOLOv8x	0.349	0.118
YOLOv8n	0.361	0.123

CHƯƠNG 4: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

4.1. Kết luận

- Trong đồ án này, nhóm em đã nghiên cứu và thực hiện việc phát hiện cháy rừng bằng hai mô hình phổ biến là YOLOv8 và Faster R-CNN. Mục tiêu của nhóm là xây dựng một hệ thống có khả năng phát hiện và định vị chính xác các vùng khói/lửa trong hình ảnh, từ đó đóng góp vào việc giám sát và dự báo nguy cơ cháy rừng.
- Kết quả thực nghiệm cho thấy cả YOLOv8 và Faster R-CNN đều cho kết quả khá tin cậy. Tuy nhiên, mỗi mô hình có những ưu điểm và hạn chế riêng.
- Với việc áp dụng YOLOv8 vào bài toán phát hiện cháy rừng đã đạt được kết quả phát hiện nhanh chóng và chính xác. Mô hình này có khả năng xử lý ảnh nhanh hơn so với Faster R-CNN, giúp giảm thời gian tính toán và tăng hiệu suất.
- Ý tưởng của Faster R-CNN là sử dụng một mạng nơ-ron tích chập để rút trích đặc trưng và một mạng phân loại riêng để định vị đối tượng. Mô hình này cung cấp kết quả phát hiện chính xác và độ tin cậy cao. Tuy nhiên, Faster R-CNN có tốc độ xử lý chậm hơn so với YOLOv8, đặc biệt khi áp dụng cho các ảnh có độ phân giải cao hoặc trong quá trình thời gian thực.
- Tùy thuộc vào yêu cầu và ứng dụng cụ thể, chúng ta có thể lựa chọn mô hình phù hợp. Nếu yêu cầu tốc độ xử lý nhanh và phát hiện thời gian thực là ưu tiên hàng đầu, YOLOv8 là một lựa chọn tốt. Trong khi đó, nếu độ chính xác và độ tin cậy cao là yếu tố quan trọng hơn, Faster R-CNN có thể được ưu tiên.
- Đồ án này đã đánh giá và so sánh hiệu năng của YOLOv8 và Faster R-CNN trong việc phát hiện cháy rừng trong ảnh chụp trong rừng. Kết quả cho thấy cả hai mô hình đều mang lại kết quả đáng tin cậy, và chúng có thể được áp dụng trong các ứng dụng thực tế khác nhau. Sự lựa chọn giữa YOLOv8 và Faster R-

CNN phụ thuộc vào yêu cầu cụ thể của bài toán và các yếu tố như tốc độ xử lý, độ chính xác và độ tin cậy.

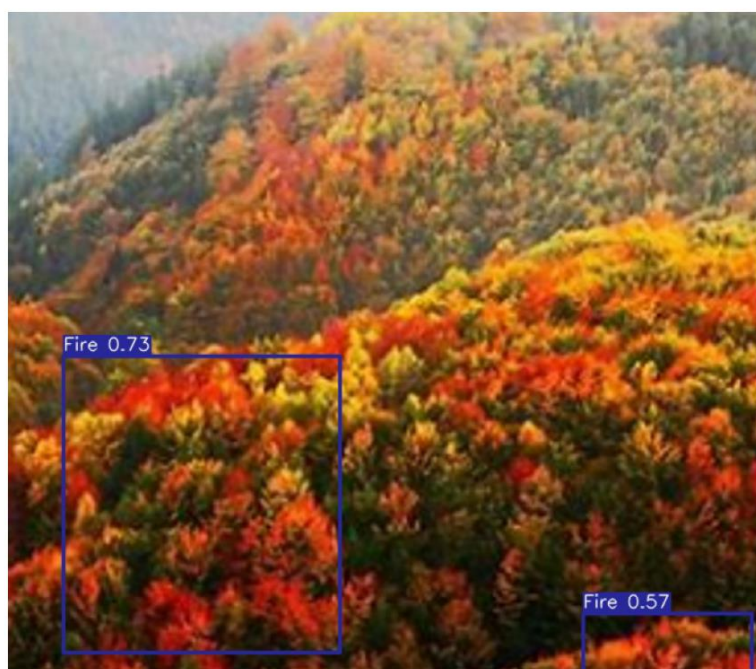
4.2. Khó khăn

Trong quá trình dự đoán, mô hình có thể gặp một số khó khăn:

- Dự đoán nhầm lẫn sương mù, mây thành khói.



- Bị ảnh hưởng bởi các mùa, đặc biệt là mùa thu và có thể bị nhiễu bởi điều kiện chiếu sáng.



- Các đám cháy nhỏ rất khó để được phát hiện vì bị che khuất bởi cây cối rậm rạp.



4.3. Hướng phát triển

- Thực hiện gán nhãn lại dữ liệu một cách tốt hơn.
- Thu thập, tổng hợp tạo ra bộ dữ liệu đa dạng và lớn hơn.
- Phát triển thành hệ thống chạy theo thời gian thực có thể tích hợp vào những thiết bị bay không người lái từ đó giúp hỗ trợ phát hiện kịp thời.
- Nghiên cứu sử dụng các kỹ thuật hiệu chỉnh (fine-tuning) để cải thiện độ chính xác của các mô hình.