

DETECTING AND MEASURING ENVIRONMENTAL DISASTERS BASED ON IMAGE SEGMENTATION DEEP LEARNING TECHNIQUE

Nguyen Quang Thi*, Nguyen Quang Uy, Phung Kim Phuong, Nguyen Minh Tri, Nguyen Manh Son
Le Quy Don Technical University

ARTICLE INFO	ABSTRACT
Received: 23/9/2022	It's essential for any environmental monitoring system to detect and measure any sources of smoke or aerosol clouds that can pollute the air or predict environmental disasters like massive forest fires. To detect and monitor smoke in the air, aerial imagery is a very effective method, especially for large areas. The presence of aerosol and particles can affect transparency and spectral properties of the atmosphere, for this reason we can detect smoke in aerial images visually. Based on these properties, in this paper, we evaluate the possibility of using automated models for aerial image analysis and detection, fragmentation, and measurement of smoke clouds in image data collected using deep learning neural networks. We have built pixel-by-pixel labeled datasets in large numbers and improve and train our segmentation models that derived from Unet neural network architecture. The test models were evaluated by IoU (Intersection Over Union) measurement and false alarm rate. The test results demonstrate deep learning models that enable reliable and efficient detection of smoke in environmental and security applications.
Revised: 03/11/2022	
Published: 03/11/2022	
KEYWORDS	
Smoke detection	
Segmentation	
Semantic segmentation	
Model	
Environmental disasters	

PHÁT HIỆN VÀ QUAN TRẮC THẢM HỌA MÔI TRƯỜNG SỬ DỤNG KỸ THUẬT HỌC SÂU PHÂN ĐOẠN ẢNH

Nguyễn Quang Thi*, Nguyễn Quang Uy, Phùng Kim Phương, Nguyễn Minh Trí, Nguyễn Mạnh Sơn
Trường Đại học Kỹ thuật Lê Quý Đôn

THÔNG TIN BÀI BÁO	TÓM TẮT
Ngày nhận bài: 23/9/2022	Việc phát hiện và đo lường đám khói cũng như mây sol khí gây ô nhiễm luôn là hoạt động cần thiết đối với bất kỳ hệ thống giám sát môi trường nào trong việc dự đoán sự cố môi trường hoặc cháy rừng. Để phát hiện và giám sát khói tồn tại trong không khí, việc sử dụng hình ảnh từ trên không là phương pháp rất hiệu quả, đặc biệt là đối với các khu vực rộng lớn. Sự hiện diện của sol khí và các hạt có thể ảnh hưởng đến độ trong suốt và các đặc tính quang phổ của khí quyển, vì lý do này mà chúng ta có thể phát hiện khói trong các hình ảnh trên không một cách trực quan. Dựa vào các đặc tính trên, trong bài báo này, chúng tôi đánh giá khả năng sử dụng các mô hình tự động để phân tích không ảnh, phát hiện, phân mảnh và đo lường các đám khói trong dữ liệu hình ảnh được thu thập bằng mạng nơ-ron học sâu Unet. Chúng tôi đã xây dựng các bộ dữ liệu được gán nhãn pixel-by-pixel với số lượng lớn nhằm cải thiện và huấn luyện các mô hình phân đoạn có nguồn gốc từ kiến trúc mạng nơ-ron Unet. Các mô hình thử nghiệm được đánh giá bằng phép đo IoU (Intersection Over Union) và tỷ lệ chấp nhận sai. Kết quả thử nghiệm cho thấy các mô hình học sâu cho phép phát hiện khói một cách đáng tin cậy và hiệu quả trong các ứng dụng an ninh và môi trường.
Ngày hoàn thiện: 03/11/2022	
Ngày đăng: 03/11/2022	
TỪ KHÓA	
Phát hiện khói	
Phân đoạn	
Phân đoạn ngữ nghĩa	
Mô hình	
Thảm họa môi trường	

DOI: <https://doi.org/10.34238/tnu-jst.6551>

* Corresponding author. Email: thinq.isi@lqdtu.edu.vn

1. Giới thiệu

Sự cố môi trường là sự cố xảy ra trong quá trình hoạt động của con người hoặc biến đổi của tự nhiên, gây ô nhiễm, suy thoái hoặc biến đổi môi trường nghiêm trọng. Ô nhiễm không khí cũng là một trong những tác động gây ô nhiễm môi trường, trong không khí có chứa các chất gây ô nhiễm gây nguy hiểm cho sức khỏe con người, hoặc tạo ra những tác động xấu đối với môi trường.

Ô nhiễm không khí có thể xảy ra do các hiện tượng tự nhiên (từ bụi, cháy rừng, núi lửa, chất phóng xạ trong tự nhiên, hoặc từ hoạt động của con người: do hút thuốc lá, quá trình đốt rom rạ thải ra khói, trong quá trình sản xuất công nghiệp, do đốt các nhiên liệu hóa thạch: than, dầu, khí đốt tạo ra các chất hữu cơ chưa cháy hết, muội than, bụi, quá trình thất thoát, rò rỉ trên dây chuyền công nghệ, do quá trình vận chuyển các hóa chất bay hơi, bụi,...). Ô nhiễm môi trường không khí có thể xảy ra trên quy mô một vùng, một quốc gia, một khu vực hoặc ô nhiễm môi trường không khí toàn cầu.

Hoạt động quan trắc đóng vai trò quan trọng trong việc đánh giá sự hiện diện của các tạp chất dạng aerosol do các chất này làm ảnh hưởng đến độ trong suốt và các tính chất khúc xạ của không khí. Từ đó, có thể giảm thiệt hại về tài sản trong cảnh báo cháy sớm và giải quyết được các vấn đề về mức độ phát tán của khói gây ảnh hưởng đến sức khỏe của con người. Hơn nữa, việc phát hiện và đo đạc lượng khói này là cần thiết để đánh giá ảnh hưởng của chúng đối với khí hậu và là căn cứ để ban hành các quy định bảo vệ môi trường trên quy mô lớn hơn.

Giải pháp phát hiện và đo đạc trình sát các đám khói đã được nghiên cứu và triển khai trên các thuật toán học máy và các mô hình học sâu. Phương pháp nghiên cứu phổ biến là sử dụng các thuật toán SVM (Support Vector Machines) [1], mô hình Fast R-CNN (Fast - Region-based Convolutional Neural Networks) [2], mô hình Smoke-Unet [1], mô hình Unet và Yuan [3]. Support Vector Machines (SVM) được sử dụng rộng rãi trong việc phát hiện khói của bộ ảnh viễn thám. Bộ phân loại SVM có thể tận dụng sự kết hợp của kết cấu, màu sắc và các đặc điểm khác của cảnh viễn thám và phân biệt thành công các điểm ảnh có khói với các điểm ảnh không khói [4] – [6]. Các kỹ thuật học máy khác, chẳng hạn như phân cụm K-mean, phân loại tuyến tính [7] và thuật toán BPNN [8], đã được sử dụng để phân biệt các pixel khói. Tuy nhiên, vẫn còn là một thách thức để trích xuất các vùng khói vì sự đa dạng về hình dạng, màu sắc, kết cấu, độ sáng và thành phần của aerosol là các sol khí. Nhìn chung, phương pháp xử lý ảnh, thị giác máy tính chưa giải quyết được bài toán phát hiện khói với các yêu cầu đặt ra; vì vậy cần phải nghiên cứu các mô hình học sâu để có thể giải quyết được nhược điểm này.

Phương pháp đề xuất của chúng tôi là xây dựng bộ dữ liệu lớn kết hợp huấn luyện và kiểm thử với nhiều mô hình học sâu để tăng hiệu quả phát hiện của các mô hình. Trong bài báo này, chúng tôi tiến hành thử nghiệm trên 03 mô hình Unet [1], Mask R-CNN [9], DeepLabv3 [10] và đo lường đám khói để đánh giá lượng tạp chất phát tán ra trong môi trường. Từ đó sớm có biện pháp khắc phục nhằm giải quyết hiệu quả vấn đề ô nhiễm môi trường. Trong quá trình nghiên cứu chúng tôi sử dụng đa dạng các mô hình học sâu và thị giác máy tính. Các mô hình thử nghiệm được đánh giá bằng độ đo IoU và tỷ lệ “chấp nhận sai”. Điều đó, giúp cho việc phát hiện khói thông qua hình ảnh và video trở nên chính xác hơn.

Bài báo được bố cục theo 5 mục chính: Phần 1 - Giới thiệu; Phần 2 đưa ra cơ sở lý thuyết liên quan được sử dụng trong bài báo; Phần 3 trình bày chi tiết về bộ dữ liệu và các phương pháp đo đạc được sử dụng; Phần 4 là kết quả thực nghiệm và đóng góp của bài báo. Cuối cùng là phần kết luận và tài liệu tham khảo được tham chiếu trong bài viết.

2. Cơ sở lý thuyết liên quan

2.1. Các phương pháp phát hiện khói bằng xử lý hình ảnh, thị giác máy tính và các mô hình học sâu

Giải pháp sử dụng các mô hình Unet, ResUnet, Smoke-Unet để phân đoạn và phân tích các tham số về khói lửa rừng có thể phân đoạn hiệu quả pixel khói trong bộ dữ liệu ảnh viễn thám

Landsat-8 Multispectral [1]. Để thực hiện điều này, nghiên cứu [1] đã phân tích các tham số của dữ liệu viễn thám và chỉ số viễn thám được sử dụng để phát hiện cháy rừng. Đầu tiên, một bộ dữ liệu ảnh đa góc độ viễn thám có độ phân giải cao về khói lửa rừng, bao gồm các năm, mùa, khu vực và độ che phủ đất khác nhau, được thành lập. Sau đó Smoke-Unet, một mô hình mạng phân đoạn khói dựa trên một Unet cải tiến kết hợp với các cơ chế và các khối dư được đề xuất. Các phương pháp dựa trên Unet gần đây [11] cũng đã có nhiều tiến bộ trong lĩnh vực viễn thám [12], [13]. Smoke-Unet, dựa trên kiến trúc Unet kết hợp các khối dư [14] và cơ chế chú ý [15], đã được đề xuất. Kết quả thử nghiệm cho thấy tỷ lệ chính xác của pixel khói khi sử dụng Smoke-Unet được đề xuất cao hơn 3,1% so với Unet, có thể phân đoạn hiệu quả pixel khói trong ảnh viễn thám. Tuy nhiên, chúng tôi nhận thấy mô hình cho kết quả phát hiện chưa cao vì bộ dữ liệu ít. Trong quá trình thử nghiệm, mô hình Unet-11 do chúng tôi đề xuất có độ chính xác khi phát hiện khói là 77,57%.

Bài toán sử dụng mạng nơ ron Convolutions để phát hiện và xác định vị trí khói và lửa [3]. Mạng này tạo ra nhãn lửa và khói trong một hình ảnh RGB (Red, Green, Blue) theo phân đoạn. Mục đích của công việc này là giúp lực lượng cứu hộ đánh giá mức độ đám cháy hoặc theo dõi đám cháy mới xảy ra trong thời gian thực với một máy ảnh được nhúng trong phương tiện nghiệp vụ. Để huấn luyện mạng này, một cơ sở dữ liệu với các hình ảnh và nhãn tương ứng có đã được tạo. So sánh mạng này với các mạng phân khúc tốt nhất như U-Net và mạng Yuan đã làm nổi bật hiệu quả của nó về độ chính xác của vị trí, giảm phân loại dương tính giả như mây hoặc sương mù trong hình ảnh. Kết quả mô hình Yuan cho kết quả phát hiện với IoU cao hơn 5,5% so với mô hình mạng Unet với cùng bộ dữ liệu.

Bên cạnh đó, việc phát hiện lửa và khói trong hình ảnh bằng cách sử dụng kỹ thuật xử lý hình ảnh và học sâu ngày càng được quan tâm nhiều hơn, một số tác giả đã trình bày phương pháp sử dụng CNN (Convolutional neural network) để phát hiện lửa hoặc khói [16]. Vẫn theo ý tưởng sử dụng camera giám sát, K. Muhammad [17] đã huấn luyện ra mô hình SqueezeNet để phát hiện đám cháy và xác định vị trí cũng như hiểu ngữ nghĩa của hiện trường vụ cháy. Q. Zhang và cộng sự [18] đã huấn luyện Faster R-CNN cách phát hiện khói trong các đám cháy rừng trên đất hoang bằng cách tạo ra các hình ảnh tổng hợp với việc bổ sung khói tổng hợp vào các hình ảnh rừng bình thường.

Bảng 1. Bảng so sánh kết quả các mô hình học sâu của các công trình nghiên cứu đã tham khảo

Mô hình	Bộ dữ liệu	F1	Accuracy	IoU
Smoke_Unet	Landsat-8 Multispectral	0,725	0,773	-
Unet	Landsat-8 Multispectral	0,735	0,699	-
Res-Unet	Landsat-8 Multispectral	0,701	0,805	-
Fast R-CNN	Firefront	-	-	0,575
Unet	SmokeFire	-	0,977	0,663
Yuan	SmokeFire	-	0,981	0,718

Kết quả phát hiện của mô hình Faster R-CNN với bộ dữ liệu Firefront là $IoU = 57,5\%$. Chúng tôi nhận thấy bộ dữ liệu còn nhiều ảnh nhiễu dẫn tới kết quả phát hiện còn thấp. Quá trình huấn luyện và thử nghiệm cho thấy mô hình Mask R-CNN với bộ dữ liệu tự xây dựng có độ chính xác khi phát hiện là $IoU = 80,41\%$. Kết quả của các mô hình học sâu được đề cập trên đây thể hiện ở Bảng 1. So sánh F1 cho thấy mô hình Unet cho kết quả cao hơn các mô hình còn lại trên cùng bộ dữ liệu huấn luyện. Kết quả phát hiện của mô hình Unet thấp hơn mô hình Yuan 5% trên cùng bộ dữ liệu SmokeFire. Bên cạnh đó, mô hình Fast R-CNN còn cho kết quả phát hiện thấp với $IoU = 57,5\%$. Nguyên nhân có thể là do bộ dữ liệu huấn luyện còn khá ít, số lượng ảnh nhiễu còn nhiều làm ảnh hưởng đến kết quả kiểm thử của các mô hình.

2.2. Các phương pháp đo đạc được sử dụng

Bài toán phát hiện hình ảnh phân đoạn khói dựa trên Convolutional Neural Network (CNN) cho cả môi trường rõ ràng và mờ ảo [6]. Không giống như các phương pháp hiện có, chúng tôi sử dụng kiến trúc CNN hiệu quả, được gọi là EfficientNet, để phát hiện khói với độ chính xác cao hơn. Chúng tôi cũng phân đoạn vùng khói bằng DeepLabv3, được hỗ trợ bởi bộ mã hóa và giải mã hiệu quả cùng với bộ phân loại pixel thông minh để phân đoạn khói một cách tối ưu. Kết quả phát hiện khói của chúng tôi cho thấy mức tăng đáng kể lên đến 3% về độ chính xác và giảm 0,46% về tỷ lệ chấp nhận sai, trong khi phân đoạn đưa ra mức tăng đáng kể 2% và 1% về độ chính xác và ngưỡng IoU. Điều đó cho thấy phương pháp này [6] trở nên phù hợp để phát hiện và phân đoạn khói trong thế giới thực. Trong quá trình huấn luyện và thử nghiệm, chúng tôi nhận thấy mô hình Mask R-CNN trên tập dữ liệu được gán nhãn cho kết quả cao nhất với chỉ số IoU là 80,41% và tỷ lệ phát hiện nhầm là 4,89%.

3. Chi tiết về bộ dữ liệu và phương pháp đo đạc được sử dụng

3.1. Chi tiết về bộ dữ liệu



Hình 1. Một số hình ảnh của bộ dữ liệu và nhãn đã xây dựng

Thu thập và xử lý dữ liệu là một bước rất quan trọng khi thực hiện một bài toán ứng dụng trí tuệ nhân tạo. Các mô hình học sâu không thể hoạt động nếu không có dữ liệu. Khi tập dữ liệu quá nhỏ, hiện tượng phổ biến có thể dễ dẫn đến *overfitting* khi mô hình không thể học đầy đủ các thuộc tính để tổng quát hóa. Trong nghiên cứu này, chúng tôi đã xây dựng bộ dữ liệu lớn thu thập hình ảnh đám khói của các đám cháy, khói thải từ nhà máy, khói từ phương tiện giao thông, khói

cháy rừng, ảnh khói tự tạo quay bằng flycam, điện thoại và hình ảnh trên Internet, Kaggle, Youtube. Sau đó, chúng tôi tiến hành gán nhãn thủ công bằng công cụ VGG Image Annotator (VIA) để chuẩn hóa dưới dạng tệp JSON.

Hình 1 là kết quả gán nhãn trên các bộ dữ liệu do chúng tôi tự xây dựng, ở đây, các hình ảnh và nhãn của bộ dữ liệu được gán pixel-by-pixel với số lượng lớn gồm các hình ảnh về đám khói trên thực tế và cả hình ảnh khói màu.

Từ các tập dữ liệu công khai, chúng tôi thu thập hình ảnh các đám khói với các kích thước đầu vào khác nhau. Kết quả thu được một bộ gồm 10.000 hình ảnh về khói được sử dụng để huấn luyện và đánh giá thuật toán được đề xuất. Với số lượng dữ liệu này đủ để đảm bảo các mô hình huấn luyện, thử nghiệm toàn diện và đa dạng. Bộ dữ liệu được công để công khai trên máy chủ theo đường dẫn: http://aiwave.vn/data/smoke_dataset.rar

3.2. Phương pháp đo đạc được sử dụng

Đối với bước phân đoạn hình ảnh, IoU là chỉ số chính để đánh giá độ chính xác của mô hình. Trong trường hợp này, vùng quan tâm không nhất thiết phải là hình chữ nhật. Nó có thể có bất kỳ hình dạng đều đặn hoặc bất thường. Điều đó có nghĩa là các dự đoán các nhãn phân đoạn chứ không phải hộp giới hạn. Do đó, quá trình phân tích từng pixel được thực hiện ở đây. Vì các giá trị của TP, FP và FN không là gì khác ngoài diện tích hoặc số pixel; chúng ta có thể viết IoU như sau.

$$IoU = \frac{TP}{TP+FP+FN} \quad (1)$$

Tỷ lệ chấp nhận sai (FAR) là một đơn vị được sử dụng để đo lường số lượng chấp nhận sai trung bình trong một hệ thống bảo mật sinh trắc học. Nó đo lường và đánh giá hiệu quả và độ chính xác của hệ thống sinh trắc học. FAR chủ yếu giúp xác định mức độ chính xác của hệ thống bảo mật sinh trắc học. FAR được tính bằng cách chia số lần chấp nhận sai cho số lần thử nhận dạng.

$$FAR = \frac{FP}{FP+TN} \quad (2)$$

Trong đó:

- Positive (P): Điểm có giá trị vượt (hoặc bằng) ngưỡng.
- Negative (N): Điểm có giá trị dưới ngưỡng.
- False Positive (FP): Tổng số trường hợp dự báo các quan sát thuộc nhãn Negative thành Positive
- True Positive (TP): Tổng số trường hợp dự báo khớp Positive.
- False Positive (FP): Tổng số trường hợp dự báo các quan sát thuộc nhãn Negative thành Positive
- False Negative (FN): Tổng số trường hợp dự báo các quan sát thuộc nhãn Positive thành Negative
- True Negative (TN): Tổng số trường hợp dự báo khớp Negative.

4. Kết quả đánh giá và đóng góp của bài báo

4.1. Kết quả đánh giá

Trong phần này, phương pháp đề xuất được đánh giá trong 02 bước. Đầu tiên, chúng tôi đánh giá các mô hình phát hiện khói trên bộ thử nghiệm bao gồm 2000 hình ảnh về khói do nhóm nghiên cứu xây dựng, độc lập với 8000 hình ảnh được đào tạo. Sau đó, dựa trên việc lựa chọn mô hình phát hiện, các thuật toán phát hiện đối tượng sẽ được đánh giá trên từng mô hình.

Bảng 2. Bảng so sánh các thông số huấn luyện và kết quả phát hiện của 03 mô hình Unet tự xây dựng

Mô hình	Kích thước ảnh đầu vào	Bath size	Epochs	Framework	IoU
Unet-7	512x512	2	65	Pytorch	62,25%
Unet-9	512x512	2	65	Pytorch	70,81%
Unet-11	512x512	2	65	Pytorch	77,57%

Kết quả ở Bảng 2 cho ta thấy với cùng kích thước đầu vào là 512x512, cùng số bath size là 2 và số Epochs là 65 thì mô hình Unet-11 cho kết quả IoU = 77,57%, cao hơn các mô hình còn lại được huấn luyện trên cùng bộ dữ liệu.

Ta thấy rằng, nếu tăng lớp thêm trong huấn luyện thì khả năng phát hiện của mô hình càng cao. Nhưng không thể tăng quá nhiều lớp vì nếu vậy sẽ tạo ra nhiều params làm cho mô hình không thể chạy được với bộ dữ liệu 10.000 hình ảnh. Ở mô hình Unet-7 trainable params là 592 ngàn tham số, ở mô hình Unet-9 trainable params là 1,9 triệu tham số. Ở mô hình Unet-11 trainable params là 124 triệu tham số.

Chi tiết huấn luyện: Có 03 mô hình được huấn luyện, bao gồm: Unet-11, DeepLabv3 và Mask R-CNN trên bộ dữ liệu huấn luyện gồm 9000 ảnh có khói và 1000 ảnh không có khói. Tập dữ liệu huấn luyện được chia theo tỷ lệ 80/20 để huấn luyện và thử nghiệm.

Thông tin chi tiết của các mô hình huấn luyện (bao gồm kích thước ảnh đầu vào, Bath size, Epochs,...) được mô tả ở Bảng 3.

Bảng 3. Thông tin chi tiết của 03 mô hình huấn luyện

Mô hình	Kích thước ảnh đầu vào	Bath size	Epochs/ Iterations	Framework
Mask R-CNN	512x512	128	20000	Pytorch
DeepLabv3	512x512	2	20	Pytorch
Unet-11	512x512	2	65	Pytorch

Chi tiết về độ phức tạp của các mô hình sử dụng với các thông tin về số lớp mạng được tiến hành huấn luyện, số lượng tham số của từng mô hình, dung lượng bộ nhớ sau khi huấn luyện, thời gian huấn luyện các mô hình, số lượng hình ảnh xử lý được trên giây với tập dữ liệu huấn luyện được chia theo tỷ lệ 80/20 để huấn luyện và thử nghiệm.

Bảng 4. Thông tin chi tiết về độ phức tạp của 03 mô hình huấn luyện

Mô hình	Số lớp mạng	Độ phân giải mỗi hình ảnh	Số lượng tham số	Dung lượng bộ nhớ sử dụng	Thời gian huấn luyện	Số hình ảnh xử lý/ giây
Mask R-CNN	15	512*512	62 triệu	4 GB GPU	6 giờ 14 phút	15
DeepLabv3	11	512*512	60 triệu	4 GB GPU	23 giờ 57 phút	16
Unet-11	11	512*512	124 triệu	4,5 GB GPU	25 giờ 35 phút	21

Nhận xét về độ phức tạp của 03 mô hình ở Bảng 4 với thời gian huấn luyện là nhiều nhất, dung lượng bộ nhớ lớn nhất, số lượng tham số cũng lớn nhất và số lượng hình ảnh xử lý trên giây cũng là nhiều nhất trên mô hình Unet-11.

So sánh kết quả: Trong phần này, nhóm nghiên cứu đánh giá kết quả thử nghiệm trên 03 mô hình được trình bày ở trên về bộ dữ liệu thử nghiệm gồm 2000 hình ảnh chứa các đám khói đã được xử lý gán nhãn và tách biệt khỏi bộ dữ liệu huấn luyện.

Bảng 5. So sánh kết quả kiểm thử của 03 mô hình

Mô hình	Kích thước ảnh đầu vào	IoU	FAR
Mask R-CNN	512x512	80,41%	4,89%
DeepLabv3	512x512	78,92%	5,78%
Unet-11	512x512	77,57%	6,32%

Bảng 5 so sánh kết quả kiểm thử của 03 mô hình với cùng bộ dữ liệu, chúng ta thấy ở mô hình Mask R-CNN cho kết quả phát hiện với IoU = 80,41% cao hơn khoảng 3% so với mô hình Unet-11. Đặc biệt, mô hình Mask R-CNN còn cho tỷ lệ phát hiện nhầm FAR = 4,89% thấp nhất so với các mô hình còn lại khi được huấn luyện trên cùng bộ dữ liệu.

Các hình ảnh kiểm thử bên dưới được lấy ngẫu nhiên trên Internet không nằm trong bộ dữ liệu huấn luyện và kiểm thử. Kết quả thể hiện ở Hình 2 có thể thấy được mô hình Mask R-CNN cho kết quả phát hiện chính xác hơn các mô hình còn lại với các hình ảnh ban ngày lẫn ban đêm.



Hình 2. Kết quả dự đoán của 03 mô hình Unet, Deeplabv3 và Mask R-CNN

Chúng tôi đã thực hiện thử nghiệm trên một video bất kì trên Internet ở Hình 3 gồm: 01 video có nền đơn giản với khoảng cách gần nền ít đối tượng, 01 video có nền phức tạp với nhiều đối tượng (nhà, đường, cánh đồng, ao hồ) ở khoảng cách xa.

Phương pháp đánh giá: đánh giá dựa trên tỷ lệ phát hiện. Vì ảnh sau khi qua mô hình được đưa tiếp qua hàm *sigmoid* (có giá trị giữa phạm vi 0 và 1) nên chúng tôi chọn ngưỡng là 0,5 để giảm thiểu tỷ lệ FP và FN trong quá trình dự đoán. Bên cạnh đó, chúng tôi cũng đánh giá FPS, số lượng vị trí dự đoán đúng trên tổng số nhãn được đánh nhãn thủ công của tất cả khung hình trong quá trình thử nghiệm.



Hình 3. Điều kiện thử nghiệm. Nền đơn giản (trái) với 1565 khung và nền phức tạp (bên phải) với 1230 khung hình

Kết quả đánh giá: Đối với mỗi video, để đánh giá chúng tôi thực nghiệm trên hệ điều hành Windows với cấu hình CPU Core i7-7500U-2.7GHz, GPU GTX Nvidia 3060 Ti.

Trong video đầu tiên với nền đơn giản ít đối tượng, mô hình Mask R-CNN có thể dễ dàng phát hiện được đám khói trên video với tỷ lệ dự đoán đúng là 1354/1565 bằng 86,51% với kết quả 8,87 FPS. Trong video thứ hai với nhiều đối tượng hơn tỷ lệ phát hiện đám khói trên video với tỷ lệ dự đoán đúng là 943/1230 bằng 76,67% với kết quả 10,83 FPS.

4.2. Đóng góp của bài báo

i) Xây dựng bộ dữ liệu lớn bao gồm 10000 hình ảnh lấy từ (ảnh khói các cuộc thi trên Kaggle, ảnh khói màu từ Youtube, video khói tự tạo, ...), gồm có:

- Bộ dữ liệu tự gán nhãn: 9000 hình ảnh được gán bằng tay bởi công cụ VGG Image Annotator (VIA) ở dạng điểm trong đa giác. Sau khi gán nhãn, nó sẽ lưu với tệp JSON. Sau đó chuyển đổi từ định dạng tệp JSON sang định dạng nhãn. Để tạo đầu vào cho quá trình huấn luyện và thử nghiệm.

- Bộ dữ liệu không có khói: 1000 hình ảnh không chứa nhãn.

ii) Xây dựng kiến trúc mạng Mask R-CNN và DeepLabv3 để so sánh kết quả đánh giá hiệu suất phát hiện trên cùng một tập dữ liệu.

iii) Cải tiến kiến trúc mạng UNET-11 từ UNET-9 để tăng khả năng phát hiện các đối tượng của kiến trúc. Trong mô hình UNET-9, chúng tôi đã xây dựng 5 lớp Convolution đơn giản trong phần bộ mã hóa. Trong tất cả 5 lớp, chúng tôi sử dụng bộ lọc 3x3. Lớp đầu tiên là 64 bộ lọc, lớp thứ hai là 128 bộ lọc, lớp thứ ba là 256 bộ lọc, lớp thứ tư là 512 bộ lọc và lớp cuối cùng là 1024 bộ lọc. Sau đó, bộ giải mã ở đây có 4 lớp tiếp theo là U -shown đầu tiên với các lớp thứ tự là 512, 256, 128 và 64 bộ lọc. Xây dựng mô hình UNET-11. Chúng tôi xây dựng 06 lớp Convolution trong phần Encoder. Trong tất cả 06 lớp, chúng tôi sử dụng bộ lọc 3x3. Trong lớp đầu tiên là 64 bộ lọc, trong lớp thứ hai là 128 bộ lọc, lớp thứ ba là 256 bộ lọc, lớp thứ tư là 512 bộ lọc, lớp thứ năm là 1024 bộ lọc, ở lớp cuối cùng là bộ lọc 2048. Sau đó, bộ giải mã ở đây có 5 lớp tiếp theo là U -shown đầu tiên với các lớp thứ tự là 1024, 512, 256, 128 và 64 bộ lọc. Tương tự, chúng ta có thể xây dựng mô hình Unet-11. Trong mỗi lớp, chúng tôi sử dụng maxpooling2d với kích thước là 2x2. Vì vậy, sau mỗi lớp, kích thước của hình ảnh sẽ giảm đi 2 lần, và độ sâu của hình ảnh tăng lên 2 lần.

5. Kết luận

Phương pháp phát hiện và đo đạc khói bằng cách sử dụng các mô hình phát hiện mạng nơ-ron phổ biến gần đây đã được đề xuất trong bài báo. Các mô hình này được huấn luyện và đánh giá để ước tính hiệu quả và áp dụng học sâu trong phát hiện khói. Nhóm nghiên cứu cũng chuẩn bị một số video khói tự tạo để đánh giá kết quả. Kết quả nghiên cứu cho thấy tính khả thi của phương pháp đề xuất trong hoạt động phát hiện và đo đạc các thông số ô nhiễm môi trường. Tuy nhiên kết quả phát hiện còn chưa được cao do bộ dữ liệu quá đa dạng và không tập trung vào được đối tượng đám khói cụ thể. Kết quả phát hiện với video có nhiều đối tượng ở khoảng cách xa cho kết quả phát hiện tốt hơn các mô hình nghiên cứu khác đã công bố.

Trong tương lai, chúng tôi dự kiến thực hiện xây dựng hệ thống kết hợp với flycam để có thể lấy thông tin hình ảnh, video về máy chủ để thực hiện phát hiện và xử lý đo đạc. Xây dựng mô hình với khả năng phát hiện cao hơn, xây dựng thêm bộ dữ liệu đa dạng hơn để có thể tăng hiệu suất phát hiện và đo đạc của hệ thống.

TÀI LIỆU THAM KHẢO/ REFERENCES

- [1] Z. Wang, P. Yang, H. Liang, C. Zheng, J. Yin, Y. Tian, and W. Cui, "Semantic Segmentation and Analysis on Sensitive Parameters of Forest Fire Smoke Using Smoke-Unet and Landsat-8 Imagery," *Remote Sens.*, vol. 14, no. 1, 2022, doi: 10.3390/rs14010045.
- [2] B. Amaral, A. Bernardino, and C. Barata, "Fire and Smoke Detection in Aerial Images," *Portuguese Conference on Pattern Recognition (RecPad)*, October 2020, pp. 1-2.
- [3] S. Frizzi, M. Bouchouicha, G. Jean-Marc, E. Moreau, and M. Sayadi, "Convolutional neural network for smoke and fire semantic segmentation," *IET Image Process*, vol. 15, pp. 634–647, 2021, doi:

- 10.1049/ipr2.12046.
- [4] M. J. Garay, D. M. Mazzoni, R. Davies, and D. Diner, "The application of support vector machines to the analysis of global datasets from MISR," in *Proceedings of the Fourth Conference on Artificial Intelligence Applications to Environmental Science*, San Diego, CA, USA, January 2005, pp. 1-19.
- [5] D. Mazzoni, J. A. Logan, D. Diner, R. Kahn, L. Tong, and Q. Li, "A data-mining approach to associating MISR smoke plume heights with MODIS fire measurements," *Remote Sensing of Environment*, vol. 107, pp. 138–148, 2007.
- [6] D. Mazzoni, M. J. Garay, R. Davies, and D. Nelson, "An operational MISR pixel classifier using support vector machines," *Remote Sensing of Environment*, vol. 107, pp. 149–158, 2007.
- [7] X. L. Li, J. Wang, W. G. Song, J. Ma, L. Telesca, and Y. M. Zhang, "Automatic Smoke Detection in MODIS Satellites Data based on K-means Clustering and Fisher Linear Discrimination," *Photogramm. Eng. Remote Sens.*, vol. 80, pp. 971–982, 2014.
- [8] X. L. Li, W. G. Song, L. Lian, and X. Wei, "Forest Fire Smoke Detection Using Back-Propagation Neural Network Based on MODIS Data," *Remote Sens.*, vol. 7, pp. 4473–4498, 2015.
- [9] W. Su, J. Zhang, C. Yang, R. Page, T. Szinyei, C. D. Hirsch, and B. J. Steffenson, "Automatic Evaluation of Wheat Resistance to Fusarium Head Blight Using Dual Mask R-CNN Deep Learning Frameworks in Computer Vision," *Remote Sens.*, vol. 13, no. 1, 2021, 10.3390/rs13010026.
- [10] S. Khan, K. Muhammad, T. Hussain, J. D. Ser, F. Cuzzolin, S. Bhattacharyya, Z. Akhtar, and V. H. C. de Albuquerque, "DeepSmoke: Deep learning model for smoke detection and segmentation in outdoor environments," *Expert Systems with Applications*, vol. 182, no. 15, 2021, doi: 10.1016/j.eswa.2021.115125.
- [11] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," 2015. [Online]. Available: <https://arxiv.org/pdf/1505.04597.pdf>. [Accessed September 29, 2021].
- [12] Y. Bao, W. Liu, O. Gao, Z. Lin, and Q. Hu, "E-Unet++: A Semantic Segmentation Method for Remote Sensing Images," in *Proceedings of the 2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, Chongqing, China, 18–20 June 2021, pp. 1858–1862.
- [13] X. Li, Z. Du, Y. Huang, and Z. Tan, "A deep translation (GAN) based change detection network for optical and SAR remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 179, pp. 14–34, 2021.
- [14] A. Markham, I. Ilyassov, M. Aitzhanov, M. F. Demirci, and A. M. Ozbayoglu, "Deep learning based investment strategy: Technical indicator clustering and residual blocks," *Soft Comput.*, vol. 25, pp. 5151–5161, 2021.
- [15] S. Kastner and L. G. Ungerleider, "Mechanisms of visual attention in the human cortex," *Annu. Rev. Neurosci.*, vol. 23, pp. 315–341, 2000.
- [16] P. Barmoutis, K. Dimitropoulos, K. Kaza, and N. Grammalidis, "Fire detection from images using faster CNN and multidimensional texture analysis," *2019 International Conference IEEE on Acoustics, Speech, and Signal Processing (ICASSP)*, 2019, pp. 8301–8305.
- [17] K. Muhammad, J. Ahmad, Z. Lv, and P. Bellavista, "Efficient CNN-based deep fire detection and locating in video surveillance applications," *IEEE Transactions on Systems, People, and Cybernetics: Systems*, 2018, pp. 1419–1434.
- [18] Q. X. Zhang, G. H. Lin, Y. M. Zhang, G. Xu, and J. J. Wang, "Wildland forest fire smoke detection based on faster r-cnn using synthetic smoke images," *Procedia Engineering*, vol. 211, pp. 441–446, 2018.