



Ngày nhận bài: 13/4/2026; Ngày thẩm định: 19/6/2026; Ngày duyệt đăng: 29/6/2026.

# PHÁT HIỆN CHÁY SỚM BẰNG HÌNH ẢNH SỬ DỤNG TRÍ TUỆ NHÂN TẠO: TỔNG QUAN PHƯƠNG PHÁP, THÁCH THỨC VÀ XU HƯỚNG ỨNG DỤNG

Thượng tá, TS PHẠM THANH TÙNG

Phó trưởng Khoa Khoa học cơ bản và Ngoại ngữ, Trường Đại học PCCC

\*Tác giả liên hệ: Phạm Thanh Tùng (Email: tungpt@daihocpccc.edu.vn)

**Tóm tắt:** Phát hiện cháy sớm có ý nghĩa quan trọng trong công tác phòng cháy, chữa cháy (PCCC), góp phần rút ngắn thời gian phản ứng và hạn chế thiệt hại. Cùng với sự phát triển của camera giám sát và trí tuệ nhân tạo, phát hiện cháy bằng hình ảnh trở thành một hướng nghiên cứu nhiều tiềm năng. Bài báo tổng quan các phương pháp phát hiện cháy sớm bằng hình ảnh sử dụng trí tuệ nhân tạo, gồm phương pháp truyền thống dựa trên đặc trưng, phương pháp học máy cổ điển, phương pháp học sâu trên ảnh và phương pháp xử lý video có xét thông tin theo thời gian. Trên cơ sở phân tích các nghiên cứu tiêu biểu, bài báo làm rõ ưu điểm, hạn chế và khả năng ứng dụng của từng nhóm phương pháp. Kết quả cho thấy các mô hình học sâu, đặc biệt là mô hình phát hiện đối tượng, đang giữ vai trò chủ đạo. Tuy nhiên, phát hiện cháy ở giai đoạn sớm vẫn là thách thức lớn đối với đám cháy nhỏ, khói mỏng và bối cảnh giám sát phức tạp. Bài báo cũng nhấn mạnh vai trò của xử lý liên khung hình trong nâng cao độ tin cậy cảnh báo.

**Từ khóa:** phát hiện cháy sớm, trí tuệ nhân tạo, thị giác máy, học sâu, phát hiện khói, phát hiện lửa, camera giám sát.

**Abstract:** Early fire detection plays an important role in fire prevention and firefighting, contributing to shortened response times and mitigated damages. With the development of surveillance cameras and artificial intelligence, image-based fire detection has become a promising research direction. This paper provides a comprehensive review of early image-based fire detection methods using AI, encompassing conventional feature-based methods, classical machine learning approaches, deep learning-based image analysis, and video processing techniques incorporating temporal information. Based on the review of representative studies, the paper analyzes the strengths, limitations, and applicability of each group of methods. The results show that deep learning models, especially object detection models, currently play a dominant role. However, early-stage fire detection remains a major challenge for small fires, with thin smoke and complex surveillance scenes. The paper also emphasizes the role of inter-frame processing in improving the reliability of fire alarms.

**Keywords:** early fire detection, artificial intelligence, computer vision, deep learning, smoke detection, flame detection, surveillance camera.

## 1. Đặt vấn đề

Phát hiện cháy sớm có ý nghĩa đặc biệt quan trọng trong công tác PCCC, vì giúp rút ngắn thời gian phản ứng, hạn chế cháy lan và giảm thiểu thiệt hại về người và tài sản. Trong thực tế, các hệ thống báo cháy truyền thống dựa trên đầu báo khói, đầu báo nhiệt và đầu báo ngọn lửa đã được sử dụng rộng rãi và giữ vai trò nền tảng trong nhiều công trình [1], [2]. Tuy nhiên, trong một số bối cảnh như: không gian lớn, không gian mở, khu vực giám sát diện rộng hoặc các môi trường có điều kiện vận hành phức tạp, các phương pháp này vẫn bộc lộ những hạn chế nhất định về phạm vi giám sát, khả năng cung cấp thông tin trực quan và mức độ đáp ứng đối với yêu cầu cảnh báo sớm [1], [2].

Sự phát triển của camera giám sát và trí tuệ nhân tạo đã mở ra một hướng tiếp cận mới cho bài toán phát hiện cháy, trong đó các dấu hiệu của lửa và khói được nhận diện trực tiếp từ ảnh hoặc video. So với các phương pháp dựa trên cảm biến điểm, cách tiếp cận này có ưu thế ở khả năng quan sát diện rộng, giám sát liên tục và cung cấp bằng chứng trực quan về hiện trường [2], [3]. Trong những năm gần đây, các kỹ thuật học máy và học sâu đã được áp dụng ngày càng rộng rãi trong phát hiện cháy bằng hình ảnh, cho phép mô hình tự động học đặc trưng từ dữ liệu thay vì phụ thuộc hoàn toàn vào các quy tắc hoặc đặc trưng thủ công [4] – [7].

Mặc dù đạt được nhiều kết quả tích cực, phát hiện cháy ở giai đoạn sớm vẫn là một bài toán khó. Ở thời điểm này, đám cháy thường xuất hiện dưới dạng vùng lửa nhỏ hoặc khói mỏng, chiếm rất ít diện tích trong ảnh và dễ bị ảnh hưởng bởi điều kiện chiếu sáng, khoảng cách quan sát, nền cảnh phức tạp hoặc các nguồn gây nhiễu như: phản xạ, đèn, bụi và hơi nước [5], [7]. Do đó, nhiều mô hình có thể đạt kết quả tốt trên các bộ dữ liệu tổng quát nhưng hiệu năng lại suy giảm khi áp dụng cho các tình huống cháy nhỏ hoặc cháy mới phát sinh [7], [8]. Đồng thời, trong hệ thống camera giám sát, kết quả phát hiện trên từng khung hình riêng lẻ chưa đủ để tạo ra cảnh báo tin cậy, do dễ phát sinh báo giả hoặc bỏ sót ngắn hạn [9], [10].

Từ thực tế đó, việc tổng quan có hệ thống các phương pháp phát hiện cháy sớm bằng hình ảnh sử dụng trí tuệ nhân tạo là cần thiết. Bài báo này tập trung phân loại các hướng tiếp cận chính, phân tích ưu điểm, hạn chế và khả năng ứng dụng của từng nhóm phương pháp, đồng thời nhấn mạnh hai vấn đề trung tâm của lĩnh vực hiện nay, gồm phát hiện các đám cháy nhỏ ở giai đoạn đầu và nâng cao độ tin cậy cảnh báo trong hệ thống camera giám sát. Trên cơ sở đó, bài báo làm rõ các thách thức nổi bật và định hướng các nghiên cứu tiếp theo đối với bài toán phát hiện cháy sớm.

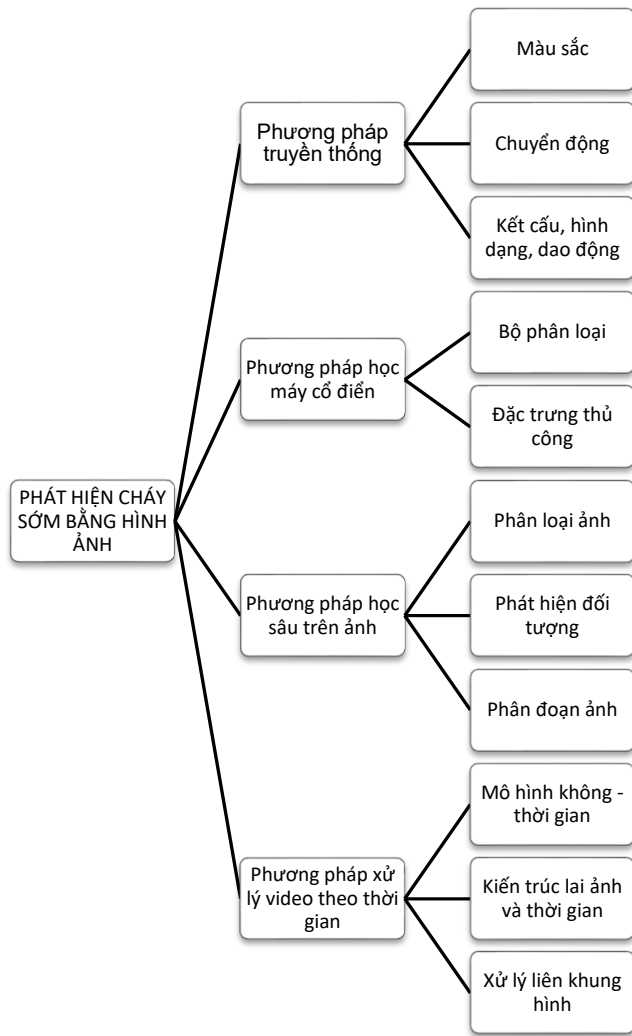
## 2. Tổng quan các hướng tiếp cận trong phát hiện cháy sớm bằng hình ảnh

Các phương pháp phát hiện cháy bằng hình ảnh có thể được chia thành bốn nhóm chính: phương pháp truyền thống dựa trên đặc trưng, phương pháp học máy cổ điển, phương pháp học sâu trên ảnh và phương pháp xử lý video có xét thông tin theo thời gian [5], [9]. Sự phân loại này không chỉ phản ánh tiến trình phát triển kỹ thuật, mà còn cho thấy mức độ phù hợp khác nhau của từng nhóm phương pháp đối với bài toán phát hiện cháy sớm trong các kịch bản giám sát khác nhau. Cấu trúc tổng quát của các nhóm phương pháp được minh họa trong Hình 1.

### 2.1. Phương pháp truyền thống dựa trên đặc trưng

Đây là nhóm phương pháp xuất hiện sớm nhất, chủ yếu khai thác các đặc trưng trực quan của lửa và khói như: màu sắc, chuyển động, kết cấu và dao động cường độ sáng [3], [11]. Một số nghiên cứu sử dụng các mô hình màu trong không gian RGB, HSV hoặc YCbCr để phân đoạn vùng lửa [3], [11], trong khi các nghiên cứu khác khai thác đặc trưng chuyển động và hình dạng của ngọn lửa hoặc khói trong chuỗi video [12], [13].

Ưu điểm của các phương pháp này là cấu trúc đơn giản, chi phí tính toán thấp và có thể triển khai trên các hệ thống có tài nguyên hạn chế. Tuy nhiên, nhược điểm nổi bật là phụ thuộc mạnh vào đặc trưng thủ công, do đó dễ bị ảnh hưởng bởi điều kiện chiếu sáng, phản xạ, nguồn sáng tương tự lửa hoặc các hiện tượng môi trường như sương mù, bụi và hơi nước [5], [9]. Vì vậy, khi bối cảnh giám sát phức tạp, hiệu năng của nhóm phương pháp này thường không ổn định.



Hình 1: Sơ đồ phân loại các hướng tiếp cận phát hiện cháy sớm bằng hình ảnh.

### 2.2. Phương pháp học máy cổ điển

Nhóm phương pháp này vẫn dựa trên đặc trưng thủ công, nhưng thay vì dùng các luật quyết định cố định, việc phân loại được thực hiện bởi các mô hình học máy như: bộ phân loại vector (Support Vector Machine – SVM), Random Forest, thuật toán phân loại dựa trên các mẫu gần nhất trong không gian đặc trưng (K-Nearest Neighbors - KNN) hoặc mô hình học máy một tầng ẩn (Extreme Learning Machine - ELM) [4]. Cách tiếp cận này giúp cải thiện khả năng phân biệt giữa vùng cháy và vùng không cháy, đặc biệt trong các bài toán có đặc trưng thị giác tương đối rõ.

Tuy nhiên, hiệu quả của các phương pháp học máy cổ điển vẫn phụ thuộc nhiều vào chất lượng của đặc trưng đầu vào. Khi dữ liệu có sự biến thiên lớn về góc quan sát, điều kiện môi trường hoặc kích thước đối tượng cháy, các đặc trưng thủ công thường không

còn đủ khả năng mô tả, dẫn đến suy giảm hiệu năng [5], [6]. Do đó, nhóm phương pháp này phù hợp hơn với các bài toán có bối cảnh tương đối ổn định hơn là các hệ thống giám sát cháy sớm trong điều kiện thực tế phức tạp.

### 2.3. Phương pháp học sâu trên ảnh

Sự phát triển của học sâu đã tạo ra bước tiến lớn trong phát hiện cháy bằng hình ảnh. Các mô hình dùng mạng nơ-ron tích chập (Convolutional Neural Network – CNN) có khả năng tự động học đặc trưng từ dữ liệu, nhờ đó cải thiện đáng kể độ chính xác so với các phương pháp trước [6], [14]. Trong thực tế, các hướng triển khai chính gồm phân loại ảnh, phát hiện đối tượng và phân đoạn ảnh.

Trong số đó, các mô hình phát hiện đối tượng như mô hình phát hiện đối tượng hai giai đoạn (Faster Region-based CNN - Faster R-CNN), mô hình phát hiện đối tượng một giai đoạn (Single Shot MultiBox Detector - SSD) và đặc biệt là mô hình phát hiện đối tượng (You Only Look Once – YOLO) đang được sử dụng rộng rãi nhất, do có thể đồng thời xác định vị trí và loại đối tượng cháy hoặc khói với tốc độ xử lý đủ cho thời gian thực [15], [16]. Ngoài ra, các mô hình phân đoạn ảnh cũng được sử dụng để xác định chi tiết vùng cháy ở mức pixel, nhất là trong các ứng dụng cần mô tả hình dạng và phạm vi của đám cháy [17]. Tuy nhiên, trong bối cảnh phát hiện cháy sớm, khó khăn lớn nhất vẫn là khả năng nhận diện các đám cháy nhỏ hoặc khói mỏng khi chúng chỉ chiếm một phần rất nhỏ trong khung hình [8], [18].

### 2.4. Phương pháp phát hiện cháy từ video và xử lý theo thời gian

Đối với hệ thống giám sát thực tế, dữ liệu đầu vào thường là video, do đó việc khai thác thông tin theo thời gian giữa các khung hình có vai trò quan trọng. Nhóm này có thể chia thành hai hướng chính.

Hướng thứ nhất là các mô hình học sâu khai thác trực tiếp đặc trưng không gian - thời gian, như CNN kết hợp mạng bộ nhớ dài – ngắn hạn (Long Short-Term Memory - LSTM), 3D CNN hoặc biến thể LSTM có tích chập (ConvLSTM) [10], [18], [19]. Các mô hình này học trực tiếp sự biến đổi của lửa và khói theo chuỗi khung hình, nhờ đó có thể tăng độ

nhạy đối với các dấu hiệu cháy còn yếu và giảm tác động của nhiễu nền.

Hướng thứ hai là các kiến trúc lai, trong đó mô hình phát hiện theo ảnh hoặc theo từng khung hình được kết hợp với cơ chế xử lý liên khung hình như xác nhận theo số khung liên tiếp, phân tích tính bền vững của hộp bao hoặc lọc theo thời gian [9], [27]. Đây là hướng tiếp cận đặc biệt phù hợp với hệ thống camera giám sát vì vừa tận dụng được hiệu quả của mô hình phát hiện đối tượng trên ảnh, vừa cải thiện độ tin cậy cảnh báo ở mức video.

So với phát hiện trên ảnh đơn, các phương pháp có xét thông tin theo thời gian có ưu thế ở khả năng nâng cao độ tin cậy cảnh báo, giảm báo giả và phù hợp hơn với hệ thống camera giám sát. Điều này đặc biệt quan trọng trong các bối cảnh có hoạt động nền cao, nơi việc chỉ dựa vào từng khung hình riêng lẻ thường không đủ để đưa ra quyết định ổn định [9], [10].

### 3. So sánh và đánh giá các phương pháp

Nhìn chung, các nhóm phương pháp phát hiện cháy bằng hình ảnh khác nhau rõ rệt về độ phức tạp, yêu cầu dữ liệu và khả năng thích ứng với bối cảnh triển khai. Một số kết quả tiêu biểu được tổng hợp trong Bảng 1. Các phương pháp truyền thống và học máy cổ điển có ưu điểm là nhẹ, dễ triển khai, nhưng hiệu quả thường suy giảm khi nền cảnh phức tạp hoặc có nhiều đối tượng gây nhầm lẫn. Trong khi đó, các mô hình học sâu cho thấy ưu thế rõ rệt về khả năng học đặc trưng và độ chính xác, đặc biệt trong các bài toán phát hiện khói, lửa từ ảnh và video giám sát [14], [20].

Nếu xét theo yêu cầu của bài toán phát hiện cháy sớm, các phương pháp không nên chỉ được đánh giá bằng độ chính xác tổng thể mà cần xem xét khả năng nhận diện khi dấu hiệu cháy còn yếu. Ở góc độ này, các mô hình phát hiện đối tượng thuộc nhóm học sâu trên ảnh có ưu thế hơn các mô hình phân loại ảnh do vừa xác định được vị trí, vừa phù hợp với hệ thống camera giám sát. Tuy nhiên, hiệu quả của chúng phụ thuộc mạnh vào kích thước đám cháy trong ảnh và độ phân giải làm việc. Trên D-Fire, một bộ dữ liệu nổi bật về cháy nhỏ và khoảng cách xa, cấu hình YOLOv5l chỉ đạt độ chính xác trung bình khi  $IoU \geq 0,5$  (mAP@0.5) bằng 79,10%; khi chuyển sang đánh

giá ở mức video với phân tích theo thời gian, cấu hình YOLOv5s kết hợp kiểm tra tính bền vững của hộp bao đạt độ chính xác 72,86% và chỉ số trung bình điều hòa ( $F_1$ -score) bằng 0,78 [21]. Kết quả này cho thấy ngay cả với các mô hình phát hiện đối tượng hiện đại, phát hiện cháy sớm trong điều kiện đám cháy nhỏ vẫn còn là một thách thức đáng kể.

Hiệu năng của mô hình cũng thay đổi mạnh theo đặc điểm bộ dữ liệu. Trên RISE, một bộ dữ liệu về khói công nghiệp với bối cảnh phức tạp hơn, *Timeception* đạt  $F_1$ -score trung bình 0,823 trên 6 cách chia dữ liệu, trong khi EFFNet cải thiện lên 0,837 nhờ tăng cường biểu diễn tiền cảnh và thông tin theo thời gian [22], [23]. Hai phương pháp này đều thuộc nhóm xử lý video có xét thông tin theo thời gian, và kết quả của chúng cho thấy ưu thế của việc khai thác đặc trưng không gian - thời gian trong các bối cảnh khó. Trên FiSmo, KutralNet chỉ đạt độ chính xác kiểm tra 77,43%, sau đó KutralNet+ cải thiện lên 80,62% với  $F_1$ -score tăng từ 0,8094 lên 0,8407 [24], [25]. Đây là ví dụ cho thấy các mô hình học sâu trên ảnh, dù có cải tiến, vẫn chịu ảnh hưởng rõ rệt bởi độ khó của dữ liệu và chưa thể bảo đảm khả năng tổng quát hóa cao trong mọi tình huống.

Từ Bảng 1 có thể thấy, mỗi nhóm phương pháp có phạm vi áp dụng khác nhau. Các phương pháp truyền thống dựa trên đặc trưng có ưu điểm là đơn giản, tốc độ nhanh và chi phí tính toán thấp. Do đó, chúng có thể phù hợp với các bối cảnh giám sát ổn định, nơi camera cố định, nền cảnh ít thay đổi và dấu hiệu cháy tương đối rõ. Tuy nhiên, nhóm phương pháp này không còn giữ vai trò chủ đạo trong các nghiên cứu hiện nay, vì hiệu năng dễ suy giảm khi gặp điều kiện ánh sáng phức tạp, phản xạ, vật thể có màu tương tự lửa hoặc các hiện tượng giống khói.

Nhóm học máy cổ điển là bước phát triển tiếp theo, khi các đặc trưng thủ công được kết hợp với các bộ phân loại như SVM, Random Forest hoặc KNN. So với các luật ngưỡng cố định, nhóm phương pháp này có khả năng học ranh giới phân loại từ dữ liệu, nhờ đó cải thiện hiệu quả trong một số bối cảnh. Tuy nhiên, bản chất của phương pháp vẫn phụ thuộc vào đặc trưng do con người thiết kế. Vì vậy, khi đối

tượng cháy nhỏ, khói mỏng hoặc nền cảnh thay đổi mạnh, các đặc trưng thủ công thường không đủ khả năng mô tả.

Các mô hình học sâu trên ảnh, đặc biệt là các mô hình phát hiện đối tượng như YOLO, SSD và Faster R-CNN, hiện đang là hướng nghiên cứu chủ đạo. Ưu điểm nổi bật của nhóm này là khả năng tự động học đặc trưng từ dữ liệu và đồng thời xác định vị trí vùng cháy hoặc khói trong ảnh. Trong đó, YOLO thường được đánh giá cao về tốc độ, phù hợp với yêu cầu xử lý thời gian thực; Faster R-CNN thường có độ chính xác tốt hơn trong một số bối cảnh nhưng chi phí tính toán cao hơn; SSD có ưu điểm về tốc độ nhưng có thể kém hiệu quả hơn với mục tiêu nhỏ. Điều này cho thấy lựa chọn mô hình cần căn cứ vào mục tiêu triển khai cụ thể, thay vì chỉ dựa trên một chỉ số độ chính xác.

Đối với dữ liệu video, việc xử lý từng khung hình riêng lẻ chưa đủ để tạo ra cảnh báo tin cậy. Các mô hình khai thác thông tin theo thời gian như CNN-LSTM, 3D CNN, ConvLSTM hoặc các kiến trúc lai giữa mô hình phát hiện đối tượng và cơ chế xác nhận liên khung hình có ưu thế trong việc giảm báo giả và tăng độ ổn định cảnh báo. Tuy nhiên, nhóm phương pháp này thường có chi phí tính toán cao hơn và cần dữ liệu video được gán nhãn phù hợp. Vì vậy, trong các hệ thống camera giám sát thực tế, hướng tiếp cận

có triển vọng là kết hợp mô hình phát hiện đủ nhanh với cơ chế xử lý liên khung hình để xác nhận sự tồn tại bền vững của dấu hiệu cháy.

Ngoài các chỉ số về độ chính xác, việc đánh giá phương pháp phát hiện cháy sớm bằng hình ảnh cần xem xét thêm chi phí tính toán, tốc độ xử lý, yêu cầu phần cứng và khả năng triển khai thực tế. Các phương pháp truyền thống và học máy cổ điển có ưu điểm là nhẹ, dễ triển khai trên hệ thống tài nguyên hạn chế, nhưng kém bền vững khi điều kiện chiếu sáng, nền cảnh hoặc môi trường giám sát thay đổi.

Trong số đó, các mô hình phát hiện đối tượng như: Faster R-CNN, SSD và YOLO được sử dụng phổ biến do có khả năng xác định vị trí vùng cháy hoặc khói trong ảnh. Faster R-CNN là mô hình phát hiện đối tượng hai giai đoạn, thường có độ chính xác tốt nhưng chi phí tính toán cao hơn; SSD và YOLO là các mô hình phát hiện đối tượng một giai đoạn, có tốc độ xử lý nhanh hơn và phù hợp hơn với yêu cầu thời gian thực. Đối với video, các mô hình như 3D CNN, CNN-LSTM hoặc Transformer khai thác được thông tin theo thời gian, giúp tăng độ ổn định cảnh báo và giảm báo giả, nhưng cũng đòi hỏi tài nguyên tính toán lớn hơn. Vì vậy, lựa chọn phương pháp cần cân bằng giữa độ chính xác, tốc độ xử lý, độ trễ cảnh báo và điều kiện triển khai của từng hệ thống PCCC cụ thể.

*Bảng 1: So sánh một số nghiên cứu tiêu biểu theo bộ dữ liệu và kết quả đánh giá.*

Nhóm phương pháp	Nghiên cứu tiêu biểu	Kết quả/chỉ số chính	Chi phí tính toán và khả năng triển khai	Nhận xét
Truyền thống dựa trên đặc trưng	Töreyn et al. [12]; Celik và Demirel [3]	Chủ yếu đánh giá bằng độ chính xác thực nghiệm, tốc độ xử lý và tỷ lệ báo giả	Thấp; có thể chạy trên phần cứng hạn chế; dễ tích hợp như bộ lọc sơ bộ	Phù hợp bối cảnh đơn giản, nhưng nhạy với ánh sáng, phản xạ, sương, bụi và nguồn sáng giống lửa
Học sâu trên ảnh - phân loại	KutralNet [24]; KutralNet+ [25]	Acc = 77,43%; KutralNet+ Acc = 80,62%, F1-score = 0,8407	Trung bình; có thể tối ưu để chạy trên thiết bị hạn chế	Phù hợp nhận biết có/không có cháy trong ảnh, nhưng không cung cấp chính xác vị trí lửa/khói
Học sâu trên ảnh - phát hiện đối tượng	Faster R-CNN, SSD, YOLO; de Venâncio et al. [21]	YOLOv5l: mAP@0.5 = 79,10%; YOLOv5s + thời gian: Acc = 72,86%, F1-score = 0,78	Trung bình đến cao; SSD và YOLO là mô hình một giai đoạn nên có tốc độ nhanh hơn Faster R-CNN; YOLO thường được ưu tiên trong yêu cầu thời gian thực	Là hướng chủ đạo hiện nay. Faster R-CNN có ưu thế về độ chính xác nhưng tính toán nặng hơn; SSD có tốc độ nhanh nhưng có thể hạn chế với mục tiêu nhỏ; YOLO cân bằng tốt giữa tốc độ và độ chính xác, nhưng vẫn khó với

				đám cháy nhỏ, khói mỏng và đối tượng ở xa
Học sâu phân đoạn ảnh	Frizzi et al. [17]	Đánh giá theo độ chính xác phân đoạn, IoU/Dice trong các nghiên cứu liên quan	Cao hơn mô hình phát hiện đối tượng; cần nhân pixel, khó triển khai thời gian thực nếu tài nguyên hạn chế	Có giá trị khi cần xác định phạm vi cháy/khói chi tiết, nhưng chi phí gán nhãn và tính toán lớn
Xử lý video theo thời gian	Hsu et al. [22]; Cao et al. [23]	Timeception F1-score = 0,823; EFFNet F1-score = 0,837	Cao hơn xử lý ảnh đơn; cần chuỗi khung hình và tài nguyên lưu trữ/tính toán lớn hơn	Hiệu quả trong bối cảnh nền phức tạp nhờ khai thác biến thiên theo thời gian
Kiến trúc lai ảnh-thời gian	Nguyen et al. [10]; Xie et al. [27]	Tăng độ ổn định cảnh báo so với xử lý từng khung hình	Trung bình đến cao; có thể triển khai theo mô hình phát hiện đối tượng + bộ lọc liên khung hình	Phù hợp thực tế PCCC vì giảm báo giả và hạn chế cảnh báo do nhiễu ngắn hạn

#### 4. Những thách thức nổi bật trong phát hiện cháy sớm bằng hình ảnh

Mặc dù các phương pháp phát hiện cháy bằng hình ảnh đã đạt được nhiều tiến bộ, phát hiện cháy ở giai đoạn sớm vẫn là một bài toán khó. Các thách thức

chủ yếu của bài toán này không chỉ đến từ bản chất thị giác của lửa và khói ở thời điểm mới phát sinh (Hình 2), mà còn liên quan đến chất lượng dữ liệu, bối cảnh giám sát và cách thức đánh giá hệ thống trong thực tế [5], [7], [9].



a. Đám cháy nhỏ



b. Khói mỏng

Hình 2: Một số thách thức điển hình trong phát hiện cháy sớm bằng hình ảnh.

##### 4.1. Hạn chế của dữ liệu hiện có

Một trong những thách thức lớn nhất của lĩnh vực này là dữ liệu huấn luyện và đánh giá còn hạn chế cả về quy mô lẫn tính đại diện. Nhiều bộ dữ liệu công khai có số lượng mẫu còn ít, điều kiện thu thập không đồng nhất, hoặc tập trung chủ yếu vào các tình huống cháy đã phát triển rõ rệt [7], [9]. Trong không ít trường hợp, video dương tính bắt đầu khi lửa hoặc khói đã hiện rõ trong khung hình, nên khó đánh giá đúng năng lực phát hiện sớm của mô hình.

Ngoài ra, dữ liệu âm tính trong nhiều bộ dữ liệu chưa thật sự phản ánh đầy đủ các yếu tố gây nhầm lẫn trong môi trường giám sát thực tế, như đèn pha, phản xạ ánh sáng, hơi nước, bụi, khói không do cháy hoặc các

vật thể có màu sắc gần với lửa [9]. Điều này làm cho một số mô hình có thể đạt kết quả tốt trên tập kiểm tra nhưng khi triển khai thực tế lại phát sinh nhiều báo giả.

##### 4.2. Thách thức đối với đám cháy nhỏ và khói mỏng

Ở giai đoạn đầu, dấu hiệu cháy thường biểu hiện dưới dạng đám lửa nhỏ hoặc khói mỏng, chỉ chiếm một vùng rất nhỏ trong ảnh. Đây là nguyên nhân khiến phát hiện cháy sớm có liên hệ chặt với bài toán phát hiện mục tiêu nhỏ [8]. Khi kích thước đối tượng quá nhỏ so với toàn bộ khung hình, thông tin đặc trưng dễ bị mất trong quá trình thay đổi kích thước ảnh, trích xuất đặc trưng hoặc qua các tầng giảm kích thước của mạng nơ-ron.

Các nghiên cứu gần đây cũng cho thấy nhiều hệ thống đạt hiệu quả tốt trên các bộ dữ liệu tổng quát, nhưng hiệu năng giảm rõ khi đối tượng cháy nhỏ hoặc ở xa [9], [21]. Trong các tình huống như vậy, việc chỉ tăng độ sâu mạng hoặc dùng kiến trúc mạnh hơn chưa chắc đã giải quyết triệt để vấn đề, nếu bộ dữ liệu huấn luyện chưa đủ mẫu cháy nhỏ và quy trình huấn luyện chưa hướng tới đặc điểm của bài toán phát hiện sớm.

### **4.3. Ảnh hưởng của nền cảnh phức tạp và báo giả**

Trong môi trường giám sát thực tế, nền cảnh thường không ổn định và có thể chứa nhiều yếu tố dễ gây nhầm lẫn với lửa hoặc khói. Các nguồn sáng mạnh, phản xạ trên bề mặt kim loại hoặc kính, khí thải xe cộ, bụi, hơi nước, sương mù hay các chuyển động bất thường trong cảnh đều có thể làm tăng xác suất báo giả [5], [9]. Đây là vấn đề đặc biệt đáng chú ý trong các khu vực như đường hầm, nhà xưởng, bãi xe, kho hàng hoặc dây chuyền sản xuất.

Theo phân tích của Gragnaniello và cộng sự, mức độ hoạt động nền là một trong hai yếu tố có ảnh hưởng quyết định đến độ khó của bài toán phát hiện cháy từ video [9]. Khi hoạt động nền tăng, cả nguy cơ báo giả lẫn nguy cơ bỏ sót đều tăng theo. Vì vậy, một phương pháp cho kết quả tốt trong bối cảnh tĩnh chưa chắc còn duy trì được hiệu quả trong môi trường giám sát động và phức tạp.

### **4.4. Khó khăn khi chuyển từ phát hiện trên ảnh sang cảnh báo trên video**

Một thách thức quan trọng khác là khoảng cách giữa phát hiện trên ảnh và cảnh báo cháy trong hệ thống giám sát video. Trong bài toán ảnh, mô hình thường được đánh giá dựa trên độ chính xác phân loại hoặc phát hiện ở từng khung hình. Tuy nhiên, trong hệ thống thực tế, yêu cầu quan trọng hơn là khả năng duy trì phát hiện ổn định theo thời gian và tạo ra cảnh báo đủ tin cậy [9], [10].

Nếu chỉ dựa trên kết quả của từng khung hình riêng lẻ, hệ thống rất dễ tạo ra cảnh báo sai do nhiễu ngắn hạn hoặc bỏ sót khi lửa, khói chỉ xuất hiện mờ nhạt trong một số khung hình đầu tiên. Vì vậy, với video giám sát, phát hiện cháy sớm không nên dừng ở

mức nhận diện đối tượng trên ảnh, mà cần gắn với các cơ chế xử lý liên khung hình để xác nhận sự tồn tại bền vững của dấu hiệu cháy. Đây là hướng đi ngày càng được nhấn mạnh trong các nghiên cứu gần đây về phát hiện cháy và khói từ video [9], [10], [18], [19].

### **4.5. Hạn chế trong đánh giá khả năng phát hiện sớm**

Một khó khăn nữa là các chỉ số đánh giá thông dụng như Accuracy, Precision, Recall hay mAP chưa phản ánh đầy đủ yêu cầu của hệ thống cảnh báo cháy sớm. Trong thực tế, cần quan tâm thêm đến khả năng phát hiện ở giai đoạn đầu, độ trễ cảnh báo, mức độ ổn định của cảnh báo và tần suất báo giả theo thời gian vận hành [9]. Tuy nhiên, do thiếu các bộ dữ liệu được chú thích rõ thời điểm xuất hiện cháy và thiếu các giao thức đánh giá thống nhất cho bài toán cảnh báo sớm, việc so sánh giữa các nghiên cứu vẫn còn gặp nhiều khó khăn.

Từ các phân tích trên có thể thấy rằng thách thức của phát hiện cháy sớm không chỉ nằm ở lựa chọn mô hình, mà còn gắn chặt với dữ liệu, bối cảnh triển khai và cách đánh giá hệ thống. Đây cũng là cơ sở cho thấy các nghiên cứu tiếp theo cần tập trung đồng thời vào xây dựng bộ dữ liệu chuyên biệt, nâng cao khả năng phát hiện mục tiêu nhỏ và phát triển các cơ chế xử lý theo thời gian nhằm nâng cao độ tin cậy cảnh báo trong hệ thống camera giám sát.

## **5. Xu hướng phát triển và khả năng ứng dụng**

Từ các nghiên cứu hiện nay có thể thấy rằng phát hiện cháy sớm bằng hình ảnh đang chuyển dần từ các phương pháp dựa trên đặc trưng thủ công sang các mô hình học sâu và xử lý video theo thời gian. Các phương pháp truyền thống và học máy cổ điển không còn là hướng chủ đạo, nhưng vẫn có giá trị trong hệ thống đơn giản, tài nguyên hạn chế hoặc làm lớp tiền xử lý. Trong khi đó, các mô hình học sâu trên ảnh, đặc biệt là mô hình phát hiện đối tượng, đang giữ vai trò trung tâm; các phương pháp xử lý video theo thời gian ngày càng quan trọng trong việc giảm báo giả và nâng cao độ ổn định cảnh báo [9], [10].

### **5.1. Xu hướng phát triển mô hình theo hướng phát hiện cháy sớm**

Xu hướng nổi bật hiện nay là phát triển các mô hình phù hợp hơn với bài toán phát hiện mục tiêu nhỏ.

Do lửa và khói ở giai đoạn đầu thường chiếm diện tích rất nhỏ trong ảnh, các nghiên cứu tập trung cải thiện trích xuất đặc trưng đa tỉ lệ, tăng cường cơ chế attention và tối ưu mô hình nhẹ để triển khai trên camera hoặc thiết bị biên [8], [18].

Đối với các phương pháp truyền thống và học máy cổ điển, xu hướng hiện nay chủ yếu là sử dụng ở vai trò hỗ trợ, như sàng lọc vùng nghi ngờ, tiền xử lý hoặc kết hợp trong các kiến trúc lai. Đối với học sâu trên ảnh, trọng tâm là nâng cao khả năng phát hiện cháy nhỏ, khói mỏng, đối tượng ở xa và bảo đảm tốc độ thời gian thực. Với dữ liệu video, nhiều nghiên cứu chuyển sang kết hợp mô hình phát hiện đối tượng theo khung hình với xử lý theo thời gian để tăng độ nhạy và giảm nhiễu nền [10], [18], [19].

### **5.2. Xu hướng phát triển dữ liệu và giao thức đánh giá**

Dữ liệu đang trở thành yếu tố quyết định trong phát hiện cháy sớm. Các nghiên cứu cần xây dựng bộ dữ liệu phản ánh sát hơn điều kiện thực tế, đặc biệt là các tình huống cháy nhỏ, khói mỏng, nền cảnh phức tạp và mẫu âm tính khó như ánh đèn, phản xạ, bụi, hơi nước hoặc khói không do cháy [7], [9].

Bên cạnh đó, giao thức đánh giá cũng cần gắn với yêu cầu vận hành thực tế. Ngoài Accuracy, Precision, Recall hay mAP, cần quan tâm thêm đến độ trễ cảnh báo, độ ổn định theo thời gian, tần suất báo giả và khả năng xử lý trên thiết bị biên. Điều này giúp đánh giá đúng hơn khả năng ứng dụng của mô hình trong hệ thống camera giám sát.

### **5.3. Khả năng ứng dụng trong hệ thống camera giám sát PCCC**

Phát hiện cháy bằng hình ảnh có thể ứng dụng trong nhà xưởng, kho hàng, bãi xe, đường hầm, khu dân cư, công trình công cộng và một số không gian mở có nguy cơ cháy [2], [9]. Ưu điểm của hướng tiếp cận này là tận dụng được hạ tầng camera sẵn có, cung cấp thông tin trực quan về vị trí và diễn biến ban đầu của sự cố.

Tuy nhiên, việc lựa chọn phương pháp cần phụ thuộc vào từng bối cảnh. Với môi trường ít nhiễu, phạm vi quan sát ngắn, các mô hình nhẹ hoặc phương pháp truyền thống có thể đáp ứng ở mức nhất định.

Với khu vực phức tạp, nhiều nguồn gây nhiễu lẫn hoặc yêu cầu phát hiện rất sớm, cần ưu tiên mô hình học sâu kết hợp xử lý liên khung hình để bảo đảm độ tin cậy [9], [27].

### **5.4. Định hướng nghiên cứu tiếp theo**

Hướng nghiên cứu tiếp theo cần tập trung vào ba nội dung chính: xây dựng bộ dữ liệu chuyên biệt cho phát hiện cháy sớm; phát triển mô hình phát hiện tốt hơn đối với cháy nhỏ, khói mỏng và mục tiêu ở xa; kết hợp phát hiện trên ảnh với xử lý theo thời gian để nâng cao độ ổn định cảnh báo trong video giám sát.

Như vậy, xu hướng phát triển của lĩnh vực không phải là thay thế hoàn toàn các phương pháp cũ, mà là tổ chức lại vai trò của từng nhóm phương pháp. Các phương pháp truyền thống và học máy cổ điển chủ yếu giữ vai trò hỗ trợ; các mô hình học sâu trên ảnh giữ vai trò trung tâm trong phát hiện và định vị cháy/khói; còn các phương pháp xử lý video theo thời gian là hướng quan trọng để nâng cao độ tin cậy cảnh báo trong thực tế.

## **6. Kết luận**

Phát hiện cháy sớm bằng hình ảnh sử dụng trí tuệ nhân tạo là hướng nghiên cứu có nhiều tiềm năng trong công tác PCCC, đặc biệt khi hệ thống camera giám sát ngày càng được triển khai rộng rãi. Qua tổng quan các nhóm phương pháp chính, có thể thấy lĩnh vực này đã phát triển từ các phương pháp truyền thống dựa trên đặc trưng, sang học máy cổ điển, học sâu trên ảnh và xử lý video có xét thông tin theo thời gian.

Các phương pháp truyền thống và học máy cổ điển hiện không còn là hướng chủ đạo do phụ thuộc nhiều vào đặc trưng thủ công và kém bền vững trong môi trường phức tạp. Tuy nhiên, chúng vẫn có giá trị trong các hệ thống đơn giản, tài nguyên hạn chế hoặc vai trò tiền xử lý. Ngược lại, các mô hình học sâu trên ảnh, đặc biệt là mô hình phát hiện đối tượng, đang giữ vai trò trung tâm nhờ khả năng tự động học đặc trưng và xác định vị trí vùng cháy/khói. Bên cạnh đó, các phương pháp xử lý video theo thời gian ngày càng quan trọng vì giúp giảm báo giả và nâng cao độ ổn định cảnh báo.

Tuy vậy, bài toán phát hiện cháy sớm vẫn còn nhiều thách thức, nhất là thiếu dữ liệu về cháy nhỏ,

khói mỏng, nền cảnh phức tạp và các mẫu âm tính khó. Vì vậy, hướng phát triển tiếp theo cần tập trung vào xây dựng bộ dữ liệu chuyên biệt, phát triển mô hình phù hợp với mục tiêu nhỏ, tối ưu mô hình cho thiết bị biên và kết hợp phát hiện trên ảnh với xử lý liên khung hình. Đây là cơ sở để phát triển các hệ thống cảnh báo cháy sớm bằng hình ảnh có khả năng ứng dụng thực tế trong công tác PCCC. ■

### TÀI LIỆU THAM KHẢO

1. National Fire Protection Association, 2017, *Fire Detection Systems Handbook*, NFPA, Quincy, MA.
2. A. E. Çetin, K. Dimitropoulos, B. Gouverneur, N. Grammalidis, O. Günay, Y. H. Habiboğlu, *et al.*, 2013, *Video fire detection – review*, *Digital Signal Processing*, vol. 23, no. 6, tr. 1827–1843.
3. T. Celik, H. Demirel, 2009, *Fire detection in video sequences using a generic color model*, *Fire Safety Journal*, vol. 44, no. 2, tr. 147–158.
4. B. Ko, K.-H. Cheong, J.-Y. Nam, 2009, *Fire detection based on vision sensor and support vector machines*, *Fire Safety Journal*, vol. 44, no. 3, pp. 322–329, doi: 10.1016/j.firesaf.2008.07.006.
5. A. Gaur, A. Singh, A. Kumar, A. Kumar, K. Kapoor, 2020, *Video flame and smoke based fire detection algorithms: A literature review*, *Fire Technology*, vol. 56, no. 5, tr. 1943–1980.
6. S. Geetha, C. Abhishek, C. Akshayanat, 2021, *Machine vision based fire detection techniques: A survey*, *Fire Technology*, vol. 57, no. 2, tr. 591–623.
7. J. Jin, T. Wang, N. Alhusaini, S. Zhao, H. Liu, K. Xu, *et al.*, 2023, *Video fire detection methods based on deep learning: Datasets, methods, and future directions*, *Fire*, vol. 6, no. 8, tr. 315.
8. Z. Li, Y. Liu, W. Jin, 2020, *Deep learning for small object detection: A survey*, *Pattern Recognition*.
9. D. Gragnaniello, A. Greco, C. Sansone, B. Vento, 2024, *Fire and smoke detection from videos: A literature review under a novel taxonomy*, *Expert Systems with Applications*, vol. 255, tr. 124783.
10. M. D. Nguyen, H. N. Vu, D. C. Pham, B. Choi, S. Ro, 2021, *Multistage real-time fire detection using convolutional neural networks and long short-term memory networks*, *IEEE Access*, vol. 9, tr. 146667–146679.
11. T.-H. Chen, P.-H. Wu, Y.-C. Chiou, 2004, *An early fire-detection method based on image processing*, *Proceedings of the International Conference on Image Processing*, tr. 1707–1710.
12. B. U. Töreyn, Y. Dedeoğlu, U. Güdükbay, A. E. Çetin, 2006, *Computer vision based method for real-time fire and flame detection*, *Pattern Recognition Letters*, vol. 27, no. 1, tr. 49–58.
13. K. Dimitropoulos, P. Barmpoutis, N. Grammalidis, 2014, *Spatio-temporal flame modeling and dynamic texture analysis for automatic video-based fire detection*, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 2, tr. 339–351.
14. A. Krizhevsky, I. Sutskever, G. E. Hinton, 2012, *ImageNet classification with deep convolutional neural networks*, *Advances in Neural Information Processing Systems*, tr. 1097–1105.
15. J. Redmon, S. Divvala, R. Girshick, A. Farhadi, 2016, *You only look once: Unified, real-time object detection*, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, tr. 779–788.
16. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, *et al.*, 2016, *SSD: Single shot multibox detector*, *Proceedings of the European Conference on Computer Vision*, tr. 21–37.
17. S. Frizzi, M. Bouchouicha, J.-M. Ginoux, E. Moreau, M. Sayadi, 2021, *Convolutional neural network for smoke and fire semantic segmentation*, *IET Image Processing*, vol. 15, no. 3, tr. 634–647.
18. H. Lin, G. Zhang, Y. Xu, Q. Zhang, 2019, *Smoke detection on video sequences using 3D convolutional neural networks*, *Fire Technology*, vol. 55, tr. 1827–1847.
19. H. Tao, M. Lu, Z. Hu, Z. Xin, J. Wang, 2022, *Attention-aggregated attribute-aware network with redundancy reduction convolution for video-based industrial smoke emission recognition*, *IEEE Transactions on Industrial Informatics*, vol. 18, no. 11, tr. 7653–7664.

20. K. Mardani, N. Vretos, P. Daras, 2023, *Transformer-based fire detection in videos*, *Sensors*, vol. 23, no. 6, tr. 3035.

21. P. V. A. de Venâncio, R. J. Campos, T. M. Rezende, A. C. Lisboa, A. V. Barbosa, 2023, *A hybrid method for fire detection based on spatial and temporal patterns*, *Neural Computing and Applications*, vol. 35, no. 13, tr. 9349–9361.

22. Y.-C. Hsu, T.-H. K. Huang, T.-Y. Hu, P. Dille, S. Prendi, R. Hoffman, *et al.*, 2021, *Project RISE: Recognizing industrial smoke emissions*, *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, tr. 14813–14821.

23. Y. Cao, Q. Tang, X. Wu, X. Lu, 2021, *EFFNet: Enhanced feature foreground network for video smoke source prediction and detection*, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 4, pp. 1820–1833, doi: 10.1109/TCSVT.2021.3083112.

24. A. Ayala, B. J. T. Fernandes, F. Cruz, D. Macêdo, A. L. I. Oliveira, C. Zanchettin, 2020, *KutralNet: A portable deep learning model for fire recognition*, *Proceedings of the International Joint Conference on Neural Networks*, tr. 1–8.

25. A. Ayala, B. J. T. Fernandes, F. Cruz, D. Macêdo, C. Zanchettin, 2022, *Convolution optimization in fire classification*, *IEEE Access*, vol. 10, tr. 23642–23658.

26. Y. Zhao, J. Zhang, K. L. Man, 2020, *LSTM-based model for unforeseeable event detection from video data*, *International Journal of Design, Analysis & Tools for Integrated Circuits & Systems*, 2020, Vol 9, Issue 1, p41.

27. Y. Xie, J. Zhu, Y. Cao, Y. Zhang, D. Feng, Y. Zhang, *et al.*, 2020, *Efficient video fire detection exploiting motion-flicker-based dynamic features and deep static features*, *IEEE Access*, vol. 8, tr. 81904–81917.