

# Thị giác máy tính cho Quản lý An toàn Nghiên cứu trong Ngành Xây dựng

Nguyen Quoc Khanh

Dept. of Vingroup AI Engineer Prgram Training

VinBigData Joint Stock Company

Ha Noi, Viet Nam

orcid.org/0009-0007-4278-149X

*Tóm tắt nội dung—Việc triển khai ngày càng nhiều công nghệ thị giác máy tính trong các quy trình công nghiệp góp phần đáng kể vào việc nâng cao lĩnh vực sản xuất về năng suất và an toàn cho người lao động. Công nhân sản xuất thường phải làm việc trong môi trường độc hại, xử lý các thiết bị nguy hiểm khác nhau, khiến tính mạng của họ bị đe dọa hàng ngày. Tai nạn lao động là lời nhắc nhở mà các công ty phải nỗ lực để giảm thiểu sự xuất hiện của nó và tác động bất lợi của chúng đối với cuộc sống của người lao động. Do sự phức tạp vốn có của các công trường xây dựng, quản lý an toàn vẫn là một thách thức thường trực và là trách nhiệm quan trọng. Mục tiêu chính của quản lý an toàn xây dựng là đảm bảo dự án hoàn thành thành công và an toàn, ưu tiên sức khỏe của người lao động và giảm thiểu tai nạn liên quan đến xây dựng và các chi phí liên quan. Sự ra đời của công nghệ thị giác máy tính (CV) đã cách mạng hóa các phương pháp tiếp cận truyền thống đối với quản lý an toàn xây dựng. Bài báo này xem xét tác động mang tính biến đổi của công nghệ CV đối với an toàn công trường và hiệu quả quản lý từ góc độ hướng ứng dụng. Ban đầu, bài báo cung cấp cái nhìn tổng quan về các nguyên tắc và phương pháp cơ bản làm nền tảng cho công nghệ CV, kèm theo mô tả chi tiết về phương pháp phân tích tài liệu được sử dụng. Sau đó, bài báo đi sâu vào các ứng dụng đa dạng của công nghệ CV, bao gồm giám sát công trường theo thời gian thực, quản lý an toàn cho người lao động, theo dõi hành vi của thiết bị và kiểm soát chất lượng vật liệu. Hơn nữa, bài báo nêu bật tiềm năng đáng kể của công nghệ CV trong việc giảm thiểu tai nạn, nâng cao hiệu suất an toàn và đưa ra những hiểu biết sâu sắc có giá trị cho các hướng nghiên cứu trong tương lai. Tóm lại, bài báo này trình bày một cái nhìn tổng quan toàn diện về việc ngành xây dựng theo đuổi việc tận dụng công nghệ CV để nâng cao thực hành quản lý an toàn, đóng vai trò như một nguồn tài nguyên thông tin và hướng dẫn.*

**Keywords—Computer vision, safety management, worker behavior.**

## I. GIỚI THIỆU

Ngành xây dựng đã trải qua quá trình mở rộng nhanh chóng do ảnh hưởng kinh tế toàn cầu đáng kể. Tuy nhiên, đây được coi là một trong những ngành nguy hiểm nhất, chiếm khoảng 20% số ca tử vong do tai nạn lao động trên toàn thế giới. Năm 2023, Bộ Lao động - Thương binh và Xã hội nước Cộng hòa Xã hội Chủ nghĩa Việt Nam đã ghi nhận 7.394 vụ tai nạn lao động, làm 7.553 người bị nạn. Tổng chi phí cho tai nạn lao động và thiệt hại tài sản lên tới gần 16.357 tỷ đồng và hơn 149.770 ngày công, chỉ tính riêng trong lĩnh vực quan hệ lao động. [1]. Theo thống kê của ILO, các công trường xây dựng trên toàn thế giới chứng kiến ít nhất 60.000 ca tử vong hàng năm, chiếm khoảng 1/6 tổng số ca tử vong liên quan đến công

việc. Ở các nước phát triển, 25%–40% số ca tử vong tại công trường xây dựng là do liên quan đến công việc. Điều đáng lo ngại là ở một số quốc gia, 30% công nhân xây dựng bị đau lưng hoặc các rối loạn cơ xương khớp khác [2]. Ngành xây dựng được công nhận rộng rãi là một trong những ngành có tỷ lệ mắc bệnh nghề nghiệp và tai nạn lao động cao nhất trên toàn cầu. Hơn nữa, những rủi ro cố hữu của ngành góp phần gây ra các vấn đề sức khỏe nghiêm trọng cho lực lượng lao động của ngành và làm tăng chi phí quản lý an toàn [3]. Do đó, cần phải ưu tiên và tăng cường các biện pháp quản lý an toàn trong suốt quá trình kỹ thuật và xây dựng.

Bản chất luôn thay đổi của các công trường xây dựng, bao gồm việc di chuyển của nhân sự và vị trí tạm thời của các thiết bị khác nhau, góp phần tạo nên vô số rủi ro có thể phát sinh [4]. Một yếu tố quan trọng khác là việc thường xuyên coi thường sự an toàn của người lao động. Cộng đồng kỹ thuật gần đây đã thừa nhận những hạn chế của các phương pháp quản lý an toàn truyền thống, chẳng hạn như chỉ dựa vào người giám sát an toàn, trong việc giải quyết hiệu quả thách thức quản lý này. Do đó, các nhà khoa học đã và đang tích cực phát triển một loạt các công nghệ thông minh tiên tiến, bao gồm thiết bị đeo, camera, robot và điện thoại thông minh, cùng với các kỹ thuật máy tính tinh vi như điện toán đám mây, Internet vạn vật và trí tuệ nhân tạo (AI). Những tiến bộ công nghệ này nhằm mục đích tự động hóa và tăng cường quản lý an toàn bằng cách thay thế các công cụ ra quyết định thông tin dựa trên con người thông thường [5].

Trong số các phương pháp tiên tiến này, Thị giác Máy tính (CV) đang nhanh chóng trở nên phổ biến do khả năng đặc biệt của nó trong việc xử lý dữ liệu hình ảnh phức tạp. Công nghệ CV cho phép giám sát thiết bị và an toàn của người lao động tại các công trường xây dựng, nhận dạng dữ liệu trực quan như số lượng và bề mặt vật liệu, đồng thời theo dõi chuyển động và thay đổi của đối tượng [6].

Sự gia tăng về tính khả dụng của camera độ phân giải cao, sự phát triển theo cấp số nhân của dung lượng cơ sở dữ liệu máy tính và sự tiến bộ của các công nghệ như học sâu, biểu đồ tri thức và mạng nơ-ron đã góp phần vào sự phổ biến ngày càng tăng của công nghệ CV [7]. (1) CV vượt trội trong việc tích hợp và xử lý hiệu quả một lượng lớn dữ liệu phức tạp; (2) CV hoạt động đáng tin cậy trong điều kiện nhiệt độ khác nghiệt, môi trường bụi bặm, trường điện từ, môi trường dễ nổ và các điều kiện làm việc nguy hiểm khác; (3) CV là một

công nghệ không xâm lấn, một khía cạnh quan trọng khi được áp dụng để giám sát an toàn xây dựng.

Bài báo này tập trung vào việc xem xét và phân tích tài liệu để khám phá các con đường chính cho nghiên cứu trong tương lai và đánh giá tiến bộ của công nghệ CV trong quản lý an toàn kỹ thuật. Nó tập trung vào các yếu tố khác nhau mà công nghệ CV sử dụng làm nhãn phân loại để phân tích thêm. Những yếu tố này bao gồm bốn lĩnh vực chính: giám sát công trường xây dựng, hướng dẫn và kiểm soát hành vi và hoạt động của người lao động, vận hành thiết bị an toàn và quản lý vật liệu. Nghiên cứu này nhằm mục đích cung cấp những hiểu biết sâu sắc và khuyến nghị có giá trị cho việc áp dụng công nghệ CV trong tương lai trong các lĩnh vực liên quan, cuối cùng góp phần vào sự phát triển hợp lý và bền vững của lĩnh vực kỹ thuật.

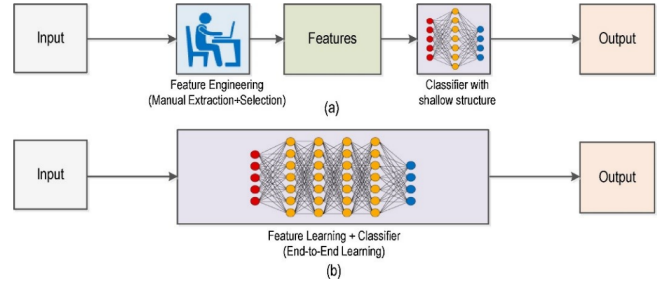
## II. TỔNG QUAN VỀ THỊ GIÁC MÁY TÍNH VÀ HỌC SÂU

Thị giác máy tính (CV), một nhánh của Trí tuệ nhân tạo (AI), có tiềm năng trích xuất thông tin chi tiết có giá trị từ dữ liệu hình ảnh kỹ thuật số, bao gồm hình ảnh và video. Là sự giao thoa nhiều ngành khác nhau như Khoa học máy tính, Toán học, Kỹ thuật, Tâm lý học và Vật lý, CV hướng đến mục tiêu sao chép các chức năng của hệ thống thị giác của con người. Lấy cảm hứng từ mắt người, Thị giác máy tính nỗ lực sao chép khả năng xử lý hình ảnh của nó. Sự kết hợp giữa Thị giác máy tính và Học sâu đã mang lại những tiến bộ đáng kể, đạt được độ chính xác và hiệu quả đặc biệt trong các tác vụ nhận dạng hình ảnh. Học sâu (DL), một lĩnh vực con của Học máy (ML), sử dụng các mạng tinh vi có khả năng học tự động từ dữ liệu không có nhãn. Các mạng này, đóng vai trò trung tâm trong DL, có khả năng học các mẫu và biểu diễn phức tạp từ dữ liệu mà không cần giám sát rõ ràng. Một quy trình cơ bản trong các tác vụ thị giác máy tính là trích xuất tính năng từ hình ảnh, bao gồm việc xác định ban đầu các cạnh, góc, màu sắc và đối tượng.

Bước đầu tiên trong việc trích xuất thông tin có ý nghĩa từ hình ảnh thô bao gồm việc phát hiện các yếu tố hình ảnh cơ bản như cạnh, góc, màu sắc và đối tượng.

Theo truyền thống, các thuật toán như SIFT (Biến đổi tính năng bất biến theo tỷ lệ), SURF (Tính năng mạnh mẽ được tăng tốc) và BRIEF (Tính năng cơ bản độc lập mạnh mẽ nhị phân) [78] đã được sử dụng để trích xuất tính năng từ hình ảnh thô. Tuy nhiên, các thuật toán thông thường này gặp khó khăn trong việc trích xuất hiệu quả các tính năng khi xử lý hình ảnh có chất lượng thấp hơn hoặc khi số lượng lớp để phân loại tăng lên. Các phương pháp Học sâu (DL) cung cấp các giải pháp để khắc phục những thách thức về trích xuất tính năng này. Hình 1 minh họa cách thức hoạt động của các thuật toán học sâu kết hợp với thị giác máy tính. Năm 2012, một bài báo nghiên cứu mang tính đột phá ImageNet Classification with Deep Convolutional Neural Networks đã giới thiệu tính ưu việt của các phương pháp học sâu so với các kỹ thuật tiên tiến hiện có (SOTA) trong nhận dạng hình ảnh tại cuộc thi CV ImageNet nổi tiếng. Sự kiện quan trọng này đánh dấu bước ngoặt trong lĩnh vực này, làm nổi bật những lợi ích đáng kể của việc tích hợp DL với CV. Bước đột phá này đã khơi dậy

sự quan tâm đến Mạng nơ-ron tích chập (CNN), dẫn đến sự phát triển của nhiều mô hình "ConvNet" khác nhau. Kiến trúc ConvNET đã chứng minh hiệu suất đặc biệt trong các tác vụ phân loại hình ảnh, đạt đến mức độ chính xác gần như hoàn hảo là 100%. Do đó, DL đã nổi lên như một công cụ không thể thiếu trong lĩnh vực Thị giác máy tính.



Hình 1. (a) Quy trình làm việc của Computer Vision truyền thống so với (b) Quy trình làm việc của Deep Learning. Hình từ [78]

Các phương pháp Deep Learning có khả năng tự động thu thập thông tin chi tiết có ý nghĩa từ dữ liệu đào tạo được chú thích. Thuộc tính này khiến chúng trở nên cực kỳ có giá trị trong Computer Vision, đặc biệt là trong việc sử dụng Mạng nơ-ron tích chập (CNN) cho các nhiệm vụ phân loại hình ảnh và phát hiện đối tượng.

### A. Các kỹ thuật Computer Vision cho Xây dựng ứng dụng trí tuệ nhân tạo

Sau đây là phân trình bày ngắn gọn về năm kỹ thuật Computer Vision chính.

1) *Phân loại hình ảnh*: Kỹ thuật này liên quan đến việc nhập một tập dữ liệu đào tạo hình ảnh được gắn nhãn vào hệ thống máy tính để xử lý. Sau đó, hệ thống phân tích các hình ảnh này để tìm hiểu và phân biệt các đặc điểm trực quan của chúng. Việc mở rộng các nhiệm vụ phân loại hình ảnh trên nhiều khung hình sẽ dẫn đến nhận dạng hành động bằng cách tích hợp các dự đoán từ mỗi khung hình.

2) *Phát hiện đối tượng*: Phát hiện đối tượng tập trung vào việc xác định và phân loại các đối tượng trong một hình ảnh, bao gồm việc tạo ra các hộp giới hạn (Bounding Box). Phát hiện đối tượng liên quan đến việc xác định và dán nhãn các đối tượng trong một hình ảnh. Áp dụng phát hiện đối tượng cho từng khung hình của chuỗi video sẽ tạo ra tác vụ theo dõi đối tượng [79].

3) *Theo dõi đối tượng*: Kỹ thuật này liên quan đến việc theo dõi một hoặc nhiều đối tượng chuyển động trong một cảnh được xác định. Theo dõi đối tượng là một kỹ thuật được sử dụng để theo dõi chuyển động của một hoặc nhiều đối tượng trong một cảnh cụ thể. Theo truyền thống, kỹ thuật này được sử dụng để quan sát và phân tích các tương tác trong thế giới thực, với các ứng dụng mở rộng sang các khung hình video.

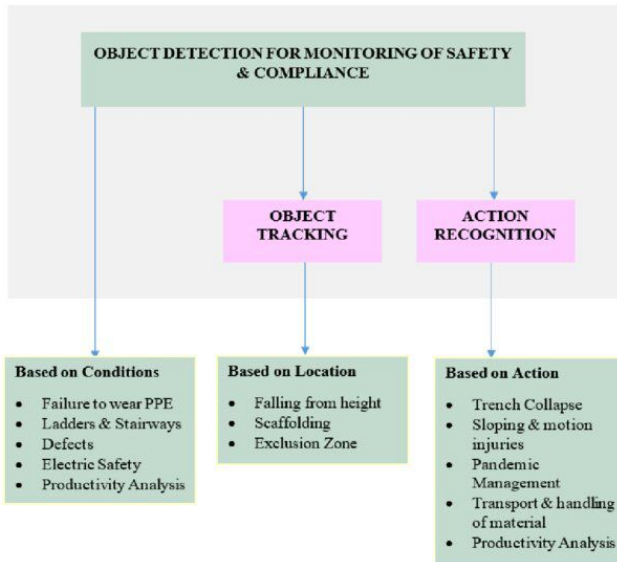
4) *Phân đoạn ngữ nghĩa*: Là một khía cạnh quan trọng của Thị giác máy tính, phân đoạn liên quan đến việc phân vùng một hình ảnh thành các nhóm pixel riêng biệt có thể được dán

nhân và phân loại. Phân đoạn ngữ nghĩa, một khía cạnh quan trọng của thị giác máy tính, chia một hình ảnh thành các phân đoạn có ý nghĩa. Cụ thể hơn, phân đoạn ngữ nghĩa nhằm mục đích giải mã vai trò của từng pixel trong một hình ảnh.

5) *Phân đoạn hình ảnh*: Kỹ thuật này phân loại tất cả các trường hợp của nhiều lớp khác nhau. Ví dụ, trong một cảnh phức tạp chứa nhiều đối tượng và nền chông chéo, kỹ thuật này cho phép phân loại tất cả các đối tượng, xác định sự khác biệt của chúng và hiểu được mối quan hệ giữa chúng. Những khía cạnh này đòi hỏi phải đánh giá toàn diện.

### III. KHUNG DỰA TRÊN THỊ GIÁC MÁY TÍNH ĐỂ GIÁM SÁT AN TOÀN TRONG XÂY DỰNG

Mặc dù có liên quan, an toàn và tuân thủ đại diện cho các khái niệm riêng biệt trong tin học. Cơ quan Quản lý An toàn và Sức khỏe Nghề nghiệp (OSHA), một bộ phận của Bộ Lao động Hoa Kỳ, nhấn mạnh việc tuân thủ là một yếu tố quan trọng trong giảm thiểu thương tích và tai nạn tại nơi làm việc, như đã nêu trong [80]. Các nỗ lực nghiên cứu trước đây đã đi sâu vào việc tìm hiểu các yếu tố nhân quả cơ bản gây ra tai nạn nghề nghiệp trong ngành xây dựng. Đáng chú ý là vào năm 2015, J. Seo và cộng sự. [81] đã giới thiệu một khuôn khổ được thiết kế để tăng cường giám sát sức khỏe và an toàn cho công nhân xây dựng. Việc xem xét các sự cố trước đó và các yếu tố góp phần của chúng nhấn mạnh nhu cầu nhấn mạnh nhiều hơn vào các giao thức an toàn và tuân thủ. Bằng cách thực hiện các biện pháp an toàn và tuân thủ mạnh mẽ, khả năng xảy ra các tai nạn như vậy có thể giảm đáng kể, như đã nêu trong [82] và [83].



Hình 2. Khung cho Hệ thống quản lý an toàn và tuân thủ

Hình 2 minh họa một khung cho Hệ thống quản lý an toàn và tuân thủ nhằm thúc đẩy môi trường xây dựng an toàn hơn. Bước đầu tiên, như được mô tả trong Hình 2, bao gồm việc biên soạn một tập dữ liệu hình ảnh thu được thông qua các thiết bị cảm biến hình ảnh. Tập dữ liệu này tận dụng ba kỹ thuật thị giác máy tính (CV) riêng biệt: (I) Phát hiện đối tượng;

(II) Theo dõi đối tượng; và (III) Nhận dạng hành động, tất cả đều được điều chỉnh để xác định các điều kiện không an toàn và không tuân thủ. Phát hiện đối tượng đóng vai trò là nền tảng để xác định các rủi ro dựa trên tình huống hoặc điều kiện, mở đường cho các phương pháp theo dõi đối tượng và nhận dạng hành động tiếp theo. Phát hiện đối tượng đóng vai trò là nền tảng để xác định các rủi ro dựa trên tình huống hoặc điều kiện, mở đường cho các phương pháp theo dõi đối tượng và nhận dạng hành động tiếp theo. Thuật toán theo dõi đối tượng giúp giám sát các hoạt động không an toàn và không tuân thủ dựa trên vị trí. Sử dụng hình ảnh tuần tự, các kỹ thuật nhận dạng hành động phân tích các hành động của các thực thể liên quan đến xây dựng (ví dụ: công nhân và thiết bị) để Sử dụng hình ảnh tuần tự, các kỹ thuật nhận dạng hành động phân tích các hành động của các thực thể liên quan đến xây dựng (ví dụ: công nhân và thiết bị) để xác định các hoạt động không an toàn, đi chệch khỏi các giao thức an toàn và tuân thủ đã thiết lập.

1) *Thiết bị bảo vệ cá nhân (PPE)*: Thiết bị bảo vệ cá nhân còn được gọi là "PPE", là thiết bị bảo vệ chống lại thương tích và bệnh tật cá nhân tại nơi làm việc. Những bệnh tật và thương tích này là kết quả của nhiều mối nguy hiểm tại nơi làm việc, ví dụ như mối nguy hiểm về phóng xạ, hóa chất, điện v.v. [84][85]. Trong Hình 3, nhiều loại PPE khác nhau được trình bày



Hình 3. Các loại PPE [84][85]

### IV. PHƯƠNG PHÁP NGHIÊN CỨU

Nghiên cứu này nhằm mục đích làm sáng tỏ trạng thái hiện tại của việc sử dụng công nghệ CV trong quản lý an toàn kỹ thuật thông qua phân tích toàn diện các tài liệu hiện có. Nghiên cứu sử dụng phân tích nội dung, một phương pháp được sử

dụng rộng rãi trong nghiên cứu khoa học xã hội. Phân tích nội dung là một kỹ thuật nghiên cứu chuyên biệt sử dụng phương pháp tiếp cận có hệ thống, khách quan và định lượng để phân tích nội dung của tài liệu. Mục tiêu chính của nó là khám phá các kết nối tiềm ẩn trong tài liệu và đưa ra dự đoán sáng suốt về các xu hướng trong tương lai. Phân cụm từ khóa, dựa trên phân tích thư mục, được sử dụng để lập bản đồ các điểm nóng nghiên cứu trong lĩnh vực an toàn xây dựng liên quan đến CV. Cách tiếp cận này đảm bảo tính trung lập trong việc phân loại các ứng dụng của công nghệ CV trong phần tiếp theo và đặt nền tảng cho cuộc thảo luận tiếp theo.

#### A. Tìm kiếm và lựa chọn tài liệu

Phần lớn tài liệu được xem xét trong bài báo này có nguồn gốc từ cơ sở dữ liệu Bộ sưu tập cốt lõi trong Web of Science (WOS). WOS lưu trữ một số hệ thống tài liệu học thuật uy tín nhất về khoa học và công nghệ, bao gồm SCI, SSCI, v.v. [8]. Nó bao gồm các tạp chí học thuật uy tín và có ảnh hưởng nhất trên toàn cầu và đóng vai trò như một nguồn tài nguyên quý giá để theo dõi sở thích nghiên cứu của các nhà khoa học quốc tế. WOS đặc biệt nổi tiếng với khả năng thư mục mạnh mẽ, cho phép đo lường tài liệu hiệu quả và hiệu quả cũng như xác định nghiên cứu phù hợp, chất lượng cao.

Nghiên cứu này sử dụng chiến lược tìm kiếm nâng cao sử dụng toán tử Boolean và từ đồng nghĩa để đảm bảo việc truy xuất tài liệu liên quan chính xác và toàn diện. Truy vấn tìm kiếm được sử dụng như sau: “TS=(CV OR based vision); TS=(construction OR architecture); TS=(safety management OR risk management); #3 AND #2 AND #1”. Tìm kiếm này đã mang lại 164 bài báo liên quan đến chủ đề nghiên cứu, bao gồm bài báo trên tạp chí, bài báo hội nghị, bài báo đánh giá và các ấn phẩm truy cập sớm.

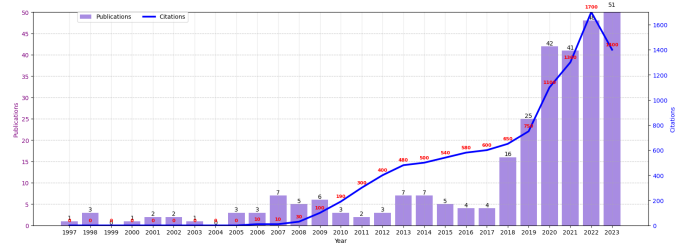
Các bài báo kéo dài từ năm 1997 đến năm 2023 và không áp dụng thêm bộ lọc thời gian nào do số lượng bài báo có hạn.

#### B. Phân tích tài liệu

1) *Phân tích dựa trên WOS*: WOS provides statistical analysis of the selected articles.

WOS cung cấp phân tích thống kê về các bài báo đã chọn. Hình 4 minh họa sự tăng trưởng hàng năm về số lượng ấn phẩm tập trung vào CV trong an toàn xây dựng. Biểu đồ cho thấy sự gia tăng đáng kể trong việc sử dụng công nghệ CV để quản lý an toàn trong ngành kỹ thuật sau năm 2018. Sự gia tăng đột biến này trùng với sự gia tăng phổ biến của các công nghệ như mạng nơ-ron và học sâu, cho thấy rằng những tiến bộ trong thuật toán máy tính đã mở đường cho việc áp dụng rộng rãi và tiến bộ của các ứng dụng CV.

Hình 5 cho thấy một số lượng đáng chú ý các ấn phẩm (71) có nguồn gốc từ Trung Quốc, nhấn mạnh trọng tâm nghiên cứu của quốc gia này về chủ đề này. Điều quan trọng cần lưu ý là con số được hiển thị trong Hình 5 cho Trung Quốc là 69 do loại trừ hai bài báo từ tỉnh Đài Loan. Tuy nhiên, con số điều chỉnh, bao gồm cả tỉnh Đài Loan, là 71. Tỷ lệ gần 50% ấn phẩm này cho thấy sự quan tâm ngày càng tăng của các nhà nghiên cứu Trung Quốc đối với an toàn của công nhân.



Hình 4. Số lần trích dẫn và bài báo theo thời gian

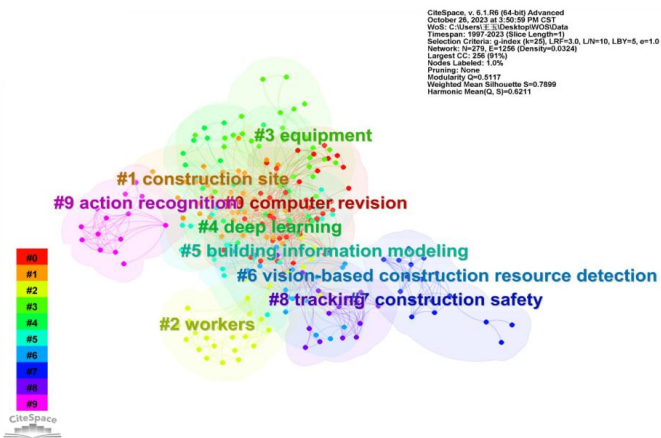
Năm 2018, Trung Quốc báo cáo 734 vụ tai nạn liên quan đến xây dựng và 840 trường hợp tử vong, trong khi năm 2020, có 689 vụ tai nạn và 794 trường hợp tử vong.



Hình 5. Minh họa mô tả phân bố địa lý của các ấn phẩm nghiên cứu

2) *Phân tích dựa trên CiteSpace*: Tận dụng CiteSpace, một công cụ chuyên dụng để phân tích trực quan các tài liệu nghiên cứu, nghiên cứu này đã thực hiện phân tích mạng đồng xuất hiện từ khóa trên tập dữ liệu gồm 164 bài báo. CiteSpace, dựa trên các thuật toán phân cụm phổ dựa trên lý thuyết đồ thị, cung cấp những lợi thế vốn có so với các mạng đồng tham chiếu truyền thống dựa vào việc liên kết các mối quan hệ để phân cụm nút thay vì các thuộc tính. Độ tin cậy của kết quả phân cụm có nguồn gốc từ CiteSpace thường được đánh giá bằng cách sử dụng hai số liệu chính: Silhouette (-S value), đại diện cho hồ sơ trung bình của các cụm và Modularity, cho biết giá trị của mô-đun phân cụm (-vQvalue). Nói chung, giá trị S lớn hơn 0,5 cho thấy phân cụm hợp lý, vượt quá 0,7 cho thấy phân cụm đạt yêu cầu, trong khi giá trị modularity là 0,7 được coi là đáng tin cậy. Trong phân tích này, như được mô tả trong Hình 6, kết quả phân cụm mang lại giá trị Q là 0,5117 và giá trị S là 0,7899, biểu thị sự tin cậy của kết quả. Kết quả, được hình dung trong Hình 6, cung cấp thông tin chi tiết về các điểm nóng nghiên cứu phổ biến trong lĩnh vực.

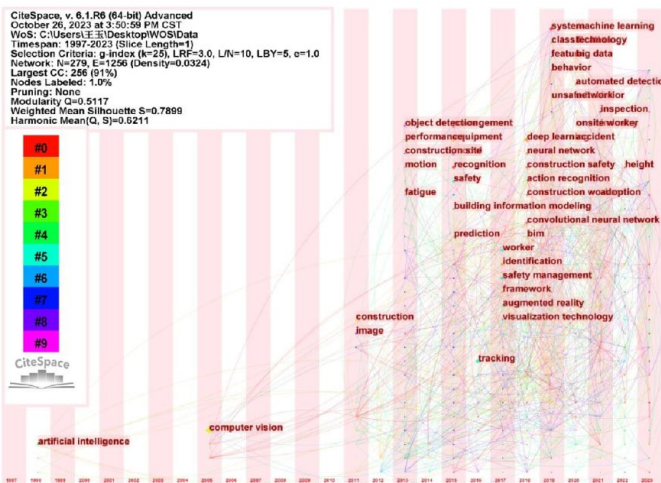
Hình 6 trình bày kết quả của phân tích cụm từ khóa bao gồm 10 cụm riêng biệt. Các từ khóa được đánh số tuần tự từ 0 đến 9, với giá trị số thấp hơn cho biết số lượng từ khóa nhiều hơn trong mỗi cụm. Mỗi cụm bao gồm một tập hợp các thuật ngữ có ngữ nghĩa liên quan. Thông qua phân tích toàn diện mười cụm từ khóa này, nghiên cứu này xác định



Hình 6. Biểu diễn theo thời gian của sự xuất hiện của từ khóa nổi bật dựa trên phân tích CiteSpace

bốn loại chính bao gồm các đối tượng của Thị giác Máy tính (CV) trong lĩnh vực quản lý an toàn kỹ thuật: công trường xây dựng, công nhân, thiết bị và vật liệu. Các danh mục này cung cấp một khuôn khổ nền tảng cho các cuộc thảo luận tiếp theo được trình bày trong nghiên cứu này.

Sự xuất hiện theo thời gian của các chủ đề nghiên cứu phù hợp chặt chẽ với chủ đề của nghiên cứu được minh họa thông qua mạng đồng xuất hiện được trình bày trong Hình 7. Đáng chú ý, các thuật ngữ như "mạng nơ-ron", "học sâu" và "mạng nơ-ron tích hợp" đã nổi lên nổi bật vào năm 2018. Giai đoạn này biểu thị một thời điểm quan trọng trong quá trình phát triển của lĩnh vực, được đánh dấu bằng sự xuất hiện và hội tụ tiếp theo của các từ khóa này,



Hình 7. Biểu diễn theo thời gian của sự xuất hiện của từ khóa nổi bật dựa trên phân tích CiteSpace

## V. ĐÁNH GIÁ TÀI LIỆU DỰA TRÊN PHÂN TÍCH NỘI DUNG

Từ góc độ sức khỏe và an toàn, có ba cấp độ phát triển và ứng dụng thị giác máy tính: Nhận dạng và theo dõi L1, đánh giá L2 và dự đoán L3, như được hiển thị trong Bảng 1.

Bảng 1  
KHUNG PHÁT TRIỂN CỦA THỊ GIÁC MÁY TÍNH

Mức độ phát triển	Chức năng	Câu hỏi nghiên cứu chính
<b>L1: Nhận dạng và theo dõi</b>	L1.1 Nhận dạng đối tượng	Đối tượng là gì?
	L1.2 Theo dõi đối tượng	Đối tượng ở đâu?
	L1.3 Nhận dạng hành động	Đối tượng đang làm gì?
<b>L2: Đánh giá</b>	L2.1 Đánh giá đối tượng	Đối tượng có phải là mối nguy hiểm không?
	L2.2 Đánh giá hành vi	Hành động có không an toàn không?
	L2.3 Đánh giá tình trạng	Điều kiện làm việc (kịch bản) có an toàn hay không an toàn?
<b>L3: Dự đoán</b>	L3.1 Dự đoán hành vi	Đối tượng sẽ hành xử như thế nào?
	L3.2 Dự đoán sự cố	Sự cố tiếp theo có xảy ra không?
	L3.3 Cảnh báo sớm	Các chỉ số hàng đầu của sự cố tiếp theo là gì?

### A. Giám sát công trường xây dựng

Trong bối cảnh quản lý công trường xây dựng, một ứng dụng nổi bật của CV nằm ở việc giám sát an toàn. Điều này bao gồm việc quan sát liên tục công trường bằng công nghệ CV để phát hiện các nguy cơ tiềm ẩn như hỏa hoạn, khói hoặc mất ổn định cấu trúc, cho phép kích hoạt kịp thời các giao thức ứng phó khẩn cấp. (2) Xác định các khu vực nguy hiểm: CV tạo điều kiện thuận lợi cho việc xác định và phân định các khu vực có nguy cơ cao hoặc các khu vực có chứa vật liệu nguy hiểm trong công trường xây dựng, từ đó nâng cao nhận thức của người lao động và thúc đẩy các biện pháp phòng ngừa. (3) Giám sát giao thông: Để đảm bảo an toàn giao thông, đặc biệt là trên đường bộ và cầu công trong và xung quanh khu vực dự án, các hệ thống hỗ trợ CV giám sát lưu lượng giao thông và mô hình di chuyển của phương tiện.

Trong một nghiên cứu gần đây tập trung vào giám sát an toàn, Yan và cộng đồng. [9] đã đề xuất một khuôn khổ khái niệm toàn diện để quản lý tai nạn xây dựng. Cách tiếp cận của họ đã tận dụng công nghệ CV và các phương pháp khác để quản lý hiệu quả hậu quả của một vụ sập nhà ở Thượng Hải bằng cách cung cấp nhận thức tình huống toàn diện. Tương tự, Ahn và cộng đồng. [10] đã giới thiệu một phương pháp luận mới để cải thiện khả năng phát hiện hỏa hoạn và giảm thiểu tai nạn và thiệt hại tài sản. Cách tiếp cận của họ liên quan đến việc phát triển một hệ thống CV được tích hợp với công nghệ truyền hình mạch kín trong nhà (CCTV), đạt được tỷ lệ chính xác đáng kể là 91% và thể hiện khả năng phát hiện hỏa hoạn trong môi trường trong nhà trong khung thời gian một giây.

Các mối nguy hiểm cụ thể hiện diện trên công trường xây dựng vốn đã thay đổi và phụ thuộc vào ngữ cảnh. Tuy nhiên, cách tiếp cận dựa trên công nghệ để xác định và đánh giá các khu vực có nguy cơ cao có thể giảm thiểu đáng kể sự xảy ra tai nạn. Theo dòng này, Pushkar và cộng đồng. [10] đề xuất

việc sử dụng tái tạo công trường 3D dựa trên hình ảnh như một phương tiện để xác định chính xác các vị trí nguy hiểm nhất, bao gồm các khu vực công nhân tiếp xúc với các rủi ro tiềm ẩn, chẳng hạn như khu vực vận hành máy móc và thiết bị, cũng như khu vực lưu trữ vật liệu. Các kỹ thuật xử lý hình ảnh thời gian thực, sử dụng nhiều camera và thuật toán CV, được sử dụng để giám sát sự an toàn của các khu vực được chỉ định này. Hơn nữa, Luo và cộng đồng. [12] đã phát triển một hệ thống giám sát video thông minh, tận dụng CV để phân biệt giữa nhân viên đứng yên và di chuyển trong các khu vực nguy hiểm khác nhau được đặc trưng bởi các mức độ rủi ro khác nhau.

Do việc sử dụng thường xuyên các phương tiện hạng nặng và thiết bị cơ khí tại các công trường xây dựng, việc quản lý tai nạn giao thông hiệu quả là điều tối quan trọng để tăng cường an toàn tổng thể. Zhu và cộng đồng. [13] đã phát triển một hệ thống cảnh báo sớm dựa trên CV được thiết kế để giám sát thời gian thực các điều kiện giao thông tại các ngã tư công trường xây dựng. Hệ thống hiệu quả về chi phí này xử lý hiệu quả các tình huống giao thông phức tạp liên quan đến xe buýt, phương tiện vận tải, công nhân và người đi bộ.

Tóm lại, việc tích hợp công nghệ CV cho các nhiệm vụ như giám sát an toàn công trường xây dựng, xác định khu vực nguy hiểm và quản lý giao thông có tiềm năng to lớn trong việc tăng cường an toàn xây dựng, giảm thiểu tai nạn và nâng cao chất lượng dự án. Do đó, những tiến bộ này đóng một vai trò trong việc đảm bảo việc thực hiện thành công các dự án xây dựng và bảo vệ sức khỏe cho người lao động.

1) *Nhận dạng và theo dõi người lao động*: Việc áp dụng công nghệ CV trong việc bảo vệ người lao động xây dựng có thể được phân loại thành ba lĩnh vực chính: (1) Các chương trình đào tạo và giáo dục được thiết kế để nâng cao nhận thức và kiến thức về an toàn cho người lao động. (2) Xác định hành vi của người lao động có khả năng gây nguy hiểm. (3) Theo dõi và xác định vị trí của người lao động trong thời gian thực.

Trong thập kỷ qua, các nhà nghiên cứu đã nỗ lực đáng kể để nhận ra các đối tượng liên quan đến dự án khác nhau bằng cách sử dụng thị giác máy tính. Như được hiển thị trong Bảng 2, trọng tâm nghiên cứu đã được đặt vào việc nhận dạng và theo dõi lực lượng lao động và thiết bị tại chỗ. Gần đây hơn, một số nỗ lực cụ thể đã được thực hiện để áp dụng nhận dạng đối tượng vào sức khỏe và an toàn xây dựng.

Kết quả nghiên cứu chỉ ra mối tương quan chặt chẽ giữa sự xuất hiện của tai nạn và hai yếu tố chính: đào tạo không đầy đủ và khả năng nhận thức hạn chế của công nhân xây dựng [14]. Nhiều nỗ lực nghiên cứu đã tập trung vào việc tận dụng những tiến bộ công nghệ để nâng cao hiệu quả đào tạo, trong đó thực tế tăng cường (AR) nổi lên như một công nghệ nổi bật trong lĩnh vực này. Khía cạnh cơ bản của những tiến bộ này liên quan đến việc triển khai CV [14]. Kivrak và cộng đồng. [15] đã phát triển một phương pháp giáo dục an toàn sáng tạo sử dụng kính thực tế tăng cường để mang đến trải nghiệm học tập hấp dẫn và nhập vai, cho phép người dùng hình dung các hoạt ảnh an toàn. Trong một nghiên cứu năm 2018, Eiris và cộng đồng đã phát triển một nền tảng hệ thống đào tạo an toàn kết hợp thực tế toàn cảnh 360 độ tăng cường

Bảng II  
CÁC ĐỐI TƯỢNG CÓ THỂ ĐƯỢC NHẬN DẠNG VÀ THEO DÕI

Các đối tượng có thể được nhận dạng và theo dõi	Tài liệu tham khảo
Cột bê tông	(Zhu et al. 2010)
Xe tải chở đất	(Rezazadeh Azar và McCabe 2011)
Máy xúc và xe tải chở đất	(Golparvar-Fard et al. 2012)
Máy xúc thủy lực	(Azar và McCabe 2012)
Cần trục tháp	(Yang et al. 2012)
Công nhân	(Park và Brilakis 2012; Zhu et al. 2016)
Máy xúc, máy xúc lật, máy ủi, máy lu và máy xúc lật bánh sau	(Tajeen và Zhu 2014)
Xe tải chở đất, máy xúc, máy xúc lật, xe trộn bê tông và máy lu đường	(Kim et al. 2017)
Nhận dạng nghề nghiệp	(Fang et al. 2018)
Xe trộn bê tông	(Kim và Kim 2018)
Dây an toàn	(Fang et al. 2018)
Lan can an toàn	(Kolar et al. 2018)
Mũ bảo hộ	(Fang et al. 2018)

(PARS) [16]. Thử nghiệm được thực hiện trên một nhóm 30 người tham gia cho thấy sự cải thiện đáng kể về khả năng xác định tất cả bốn mẫu nguy hiểm được chỉ định của họ. Kết quả của những nghiên cứu này cho thấy rằng công nghệ CV sẵn sàng tạo điều kiện thuận lợi cho việc phát triển các nền tảng đào tạo tiên tiến có khả năng nâng cao giáo dục và đào tạo về an toàn, từ đó cải thiện khả năng xác định và giải quyết các mối nguy hiểm tiềm ẩn của người lao động. Các sáng kiến đào tạo công nhân trong tương lai nhằm mục đích kết hợp các công nghệ tiên tiến để thúc đẩy sự tham gia thông qua trải nghiệm nhập vai.

Việc xác định và giám sát công nhân xây dựng tạo thành một khía cạnh quan trọng của việc giảm thiểu các sự cố an toàn. Tuy nhiên, việc thực hiện toàn diện nhiệm vụ này đã gặp nhiều thách thức do tính chất phức tạp và khắt khe của nó. Làm việc trên cao từ lâu đã được công nhận là một nghề nguy hiểm. Trên thực tế, một tỷ lệ đáng kể công nhân thể hiện sự thiếu nhận thức hoặc cố ý coi thường việc sử dụng các thiết bị an toàn thiết yếu, chẳng hạn như mũ bảo hiểm và dây đai an toàn. Để giải quyết những lo ngại này, Fang và cộng đồng. [17] đã đề xuất một phương pháp quản lý an toàn kết hợp mạng nơ-ron tích hợp và CV. Cách tiếp cận này cho phép thực hiện nhanh chóng các biện pháp can thiệp và phòng ngừa khi phát hiện các hành vi cụ thể, dẫn đến giảm đáng kể các sự cố rơi trong bối cảnh làm việc trên cao. Trong nghiên cứu của mình, Hayat và cộng đồng [18] đã trình bày một khuôn khổ mới sử dụng công nghệ thị giác máy tính để kiểm tra tự động mũ bảo hiểm an toàn, đạt tỷ lệ chính xác 92,44% ngay cả trong điều kiện đầy thách thức như ánh sáng yếu hoặc sự hiện diện của các vật thể nhỏ. Một nghiên cứu khác của Ding và cộng đồng. [19] đã giới thiệu một mô hình học sâu lai kết hợp Mạng nơ-ron tích hợp (CNN) và Bộ nhớ ngắn hạn dài hạn (LSTM) để tự động phân loại hành động của người lao động trên các tư thế khác nhau.

Trong công trình của mình, các nhà nghiên cứu Alateeq et al. [20] đã phát triển một phương pháp sử dụng cảnh quay

truyền hình mạch kín (CCTV) để nhận dạng công nhân, tích hợp các kỹ thuật học sâu và thị giác máy tính để cải thiện khả năng nhận dạng và phân loại các hoạt động của công nhân xây dựng và tuân thủ an toàn. Những tiến bộ này làm nổi bật sự tiến bộ nhanh chóng của thị giác máy tính trong việc giám sát hành vi của công nhân, đặc biệt là khi kết hợp với các thiết bị đeo được để tăng cường sự tiện lợi. Việc tích hợp các phương pháp thông minh như học sâu và các công nghệ camera tiên tiến như camera độ sâu, camera ống nhòm và camera quang học có thể góp phần vào phương pháp tiếp cận chính xác và tự động hơn để quản lý hành vi của công nhân [21].

Trong khi các hệ thống quản lý an toàn để theo dõi vị trí công nhân tại các công trường xây dựng thường dựa vào hệ thống Internet vạn vật (IoT) và các thiết bị cảm biến như Nhận dạng tần số vô tuyến (RFID) để biểu diễn thông tin vị trí công nhân một cách ảo, công nghệ thị giác máy tính cung cấp khả năng theo dõi thời gian thực, tiện lợi và ổn định hơn. Park và cộng sự [22] đã giới thiệu một hệ thống hình ảnh dựa trên máy tính để xác định và theo dõi công nhân xây dựng trong các khung hình video, cho phép phát hiện vị trí công nhân và xác định tọa độ không gian cho vật liệu, tòa nhà và thiết bị tại công trường xây dựng. Park và cộng sự [23] đã khám phá các chiến lược theo dõi công nhân trong môi trường ngoài trời đông đúc bằng cách kết hợp công nghệ camera với các phương pháp thị giác máy tính. Park và cộng sự [24] cũng đã đề xuất tích hợp thị giác máy tính với thực tế ảo và tương tác giữa người và máy tính để kết hợp các quy trình theo dõi và định vị công nhân. Một phương pháp theo dõi trực quan trực tuyến sử dụng Mạng nơ-ron tích chập đa miền (MD-CNN) đã chứng minh hiệu suất theo dõi thỏa đáng ngay cả trong các môi trường xây dựng phức tạp của Lin và cộng sự [25].

Theo truyền thống, việc đảm bảo an toàn cho công nhân tại các công trường xây dựng đòi hỏi nhiều nhân công, đòi hỏi các nhà quản lý và kỹ sư phải kiểm tra thực tế công trường để phát hiện các mối nguy tiềm ẩn. Công nghệ thị giác máy tính cung cấp giải pháp tự động theo dõi vị trí và hành vi của người lao động thông qua phân tích hình ảnh.

### B. Giám sát thiết bị

Tai nạn liên quan đến máy móc, đặc biệt là những tai nạn liên quan đến các phương tiện lớn như cần cẩu và máy xúc, là một mối lo ngại đáng kể tại các công trường xây dựng.

Azar và cộng sự [26] đã đề xuất một phương pháp thị giác máy tính mới sử dụng camera zoom chủ động để giám sát hoạt động của máy đào và xe ben, đạt độ chính xác nhận dạng từ 80

Nghiên cứu sâu rộng đã khám phá việc xác định thái độ trong máy móc xây dựng bằng nhiều phương pháp khác nhau, bao gồm cảm biến IMU, đặc biệt tập trung vào việc tránh va chạm máy móc như một khía cạnh quan trọng của quản lý an toàn. Zhang và cộng sự [28] đã đề xuất một khuôn khổ đánh giá để tăng cường an toàn tránh va chạm bằng cách sử dụng thị giác máy tính và các kỹ thuật lý luận mờ, xác định máy móc và sự gần gũi cũng như tình trạng chen chúc của công nhân là những tác nhân chính gây ra tai nạn va chạm. Để cải thiện hiệu quả của hệ thống, một thuật toán R-CNN nhanh hơn

đã được kết hợp để thiết lập mối quan hệ định lượng trong việc xác định các yếu tố này. Zhang et al [29] đã phát triển một khuôn khổ cho an toàn tránh va chạm bằng cách sử dụng các kỹ thuật thị giác máy tính (CV) và lý luận mờ. Họ phát hiện ra rằng sự gần gũi và đông đúc của máy móc và công nhân là nguyên nhân chính gây ra tai nạn va chạm. Các tác giả đã tích hợp một thuật toán R-CNN nhanh hơn để cải thiện hiệu quả của hệ thống bằng cách định lượng các yếu tố này.

Việc thực hiện thành công các công trình đất phụ thuộc vào việc vận hành máy móc lớn một cách khéo léo và quản lý các điều kiện địa chất phức tạp, có thể gây ra các mối nguy hiểm liên quan đến việc vận chuyển vật liệu. Một kỹ thuật thị giác mới sử dụng hai camera độ sâu đã được giới thiệu để đánh giá chính xác sự biến dạng không gian-thời gian của đất trong quá trình đào đất của Tsuchiya et al. [30]. Nghiên cứu này cũng đã sửa đổi mô hình hiện có để dự đoán sức cản của gầu trong quá trình đào, kết hợp các phép đo mới. Mặc dù những tiến bộ về thuật toán có thể nâng cao hơn nữa độ chính xác của phương pháp, nhưng nghiên cứu này có ý nghĩa quan trọng đối với việc sử dụng công nghệ thị giác trong các đánh giá chính xác. Naghshbandi et al. [31] đã xem xét ứng dụng công nghệ thị giác máy tính (CV) để tăng cường các biện pháp an toàn trong các hoạt động đào đất, tập trung vào an toàn tự động và giám sát mục tiêu. Việc sử dụng hiệu chuẩn đa camera, hợp nhất dữ liệu và các kỹ thuật ước tính thái độ đã góp phần cải thiện độ chính xác phát hiện và giảm độ phức tạp của quá trình xử lý.

Việc tích hợp công nghệ thị giác máy tính vào quy định về máy móc và phương tiện xây dựng có khả năng giảm thiểu các mối nguy hiểm liên quan đến các khu vực không an toàn và điểm mù, cải thiện quản lý hoạt động trong các môi trường đầy thách thức như công trình đất.

### C. Quản lý vật liệu và tài nguyên

Trong quản lý an toàn xây dựng, thị giác máy tính (CV) chủ yếu được sử dụng để xác định, phân loại và theo dõi vật liệu và tài nguyên như giàn giáo tạm thời và ván khuôn, tập trung vào việc giám sát chất lượng vật liệu để đảm bảo các tiêu chuẩn an toàn. Bản chất thâm dụng lao động của các công trường xây dựng đòi hỏi lực lượng lao động, máy móc, vật liệu và các điều khoản tạm thời đáng kể, và việc quản lý không đầy đủ các thành phần này có thể ảnh hưởng tiêu cực đến lịch trình xây dựng và các giao thức an toàn [32].

Trong cuộc điều tra của mình, Dimitrov và cộng sự [33] đã đề xuất một phương pháp phân loại vật liệu bằng cách sử dụng CV và bộ phân loại máy vectơ hỗ trợ. Các nhà nghiên cứu đã thu thập một tập dữ liệu bao gồm hơn 20 vật liệu xây dựng thường được sử dụng, với 150 hình ảnh thu được cho mỗi loại. Đáng chú ý là các hình ảnh đã được xử lý cẩn thận để đảm bảo điều kiện ánh sáng tối ưu tại công trường, giúp tăng độ chính xác của quá trình phân loại. Do đó, Mahami và cộng sự [34] đã phát triển và trình bày một phương pháp phân loại vật liệu trực quan dựa trên học sâu. Tổng cộng 1231 hình ảnh từ nhiều công trường xây dựng khác nhau đã được thu thập để đào tạo một tập dữ liệu đáng kể. Nghiên cứu đã đạt được độ chính xác phát hiện thành công là 97,35% đối với

nhiều loại vật liệu xây dựng, ngay cả trong môi trường phức tạp.

Lĩnh vực kiểm tra dựa trên thị giác để đánh giá chất lượng vật liệu xây dựng và tính toàn vẹn của cấu trúc, đặc biệt tập trung vào phát hiện vết nứt, đã có những tiến bộ đáng kể trong những năm gần đây. Trong nghiên cứu của mình, Dinh và cộng sự [35] đã tích hợp các robot kiểm tra và kỹ thuật CV để tạo điều kiện cho việc tự động trích xuất các vết nứt có thể nhìn thấy từ ảnh. Điều này đạt được bằng cách sử dụng ngưỡng nhị phân hóa ảnh, sử dụng các thuật toán phát hiện đỉnh không tham số. Theo Koch và cộng sự [36], CV đã có những tiến bộ đáng kể trong việc tự động phát hiện các vết nứt, hư hỏng mối nối và các lỗ hổng trong nhiều thành phần cơ sở hạ tầng khác nhau như đường hầm, vỉa hè nhựa đường và đường ống. Các tác giả nhấn mạnh tầm quan trọng của phương pháp định lượng trực quan này trong việc đảm bảo an toàn và chất lượng của các vật liệu này. Trong bài đánh giá gần đây của mình, Deng và cộng sự [37] đã giới thiệu ứng dụng của CV trong phát hiện vết nứt, cùng với việc tích hợp các kỹ thuật trí tuệ nhân tạo. Các kỹ thuật này đã chứng kiến những tiến bộ đáng kể, chuyển đổi từ các kỹ thuật xử lý hình ảnh thủ công (IPT) dựa trên các tính năng cấp thấp sang các phương pháp học tính năng tinh vi hơn. Do đó, các phương pháp tiếp cận hiện tại cung cấp kết quả định lượng toàn diện và chính xác.

## VI. NHỮNG THÁCH THỨC CỦA THỊ GIÁC MÁY TÍNH TRONG XÂY DỰNG

Mặc dù đã có nhiều nỗ lực tự động hóa việc xác định môi trường nguy hiểm tại các công trường xây dựng bằng thị giác máy tính, nhưng vẫn chưa có một hệ thống hoàn toàn tự động nào. Mặc dù chúng ta có thể xác định những cá nhân không đeo thiết bị bảo vệ cá nhân (PPE), nhưng việc xác định danh tính của họ và xác minh việc sử dụng PPE đúng cách, chẳng hạn như đảm bảo móc dây an toàn được cố định vào lan can, vẫn là một thách thức. Tuy nhiên, học sâu có tiềm năng xác định môi trường nguy hiểm theo thời gian thực thông qua phân tích dữ liệu. Hãy cùng tìm hiểu sâu hơn về các ứng dụng tiềm năng và rào cản kỹ thuật của các thuật toán học sâu, đặc biệt là mạng nơ-ron tích chập (CNN), trong quản lý an toàn. (1) xác định những cá nhân không tuân thủ các yêu cầu về PPE; và (2) xác định xem PPE có được sử dụng đúng cách hay không (ví dụ: nếu móc dây an toàn được gắn vào lan can). Tuy nhiên, học sâu có tiềm năng cung cấp phân tích dữ liệu để xác định môi trường nguy hiểm theo thời gian thực, tự động. Tiếp theo, chúng ta sẽ khám phá các ứng dụng tiềm năng và thách thức kỹ thuật của các thuật toán học sâu, đặc biệt là CNN, trong quản lý an toàn.

### A. Áp dụng thị giác máy tính vào thực tế

Điều kiện tiên quyết để xác định môi trường nguy hiểm hiệu quả (giảm thiểu phát hiện sai và đảm bảo độ chính xác) và ứng dụng thị giác máy tính thành công là quyền truy cập vào cơ sở dữ liệu hình ảnh toàn diện, chất lượng cao bao gồm nhiều loại khác nhau để đào tạo CNN. Tuy nhiên, việc thiếu cơ sở dữ liệu đủ lớn gây ra trở ngại đáng kể cho việc tận dụng thị giác máy tính để giám sát an toàn hiệu quả. Không giống như

các tập dữ liệu có sẵn công khai trong khoa học máy tính, chẳng hạn như ImageNet và Microsoft® Common Objects in Context (COCO), các tập dữ liệu xây dựng yêu cầu các đặc điểm riêng biệt. Những đặc điểm này bao gồm các cân nhắc về xung đột không gian, nền lộn xộn, che khuất, các thay đổi về tư thế và tỷ lệ, và bản chất năng động, luôn thay đổi của môi trường xây dựng. Do đó, nhiều mối nguy hiểm tiềm ẩn vẫn ẩn khỏi tầm nhìn thông thường (ví dụ: va chạm tiềm ẩn với nhà máy hoặc thiết bị) và việc đánh giá chất lượng kết cấu trở nên khó khăn. Khó khăn này phát sinh từ những hạn chế của các mô hình học sâu trong việc dự đoán các đối tượng trước đây không được quan sát và trích xuất thông tin ẩn thông qua thị giác máy tính.

Sự khan hiếm của các tập dữ liệu đã buộc các nhà nghiên cứu phải dựa vào các mẫu hình ảnh tương đối nhỏ cho công việc xác định môi trường nguy hiểm thử nghiệm của họ. Sự phụ thuộc vào dữ liệu hạn chế này đã khiến việc so sánh và đối chiếu các số liệu đánh giá được báo cáo như độ chính xác, khả năng thu hồi và độ chính xác với các phương pháp tiếp cận khác trở nên khó khăn. Sự thay đổi về chất lượng của các tập dữ liệu được sử dụng để đào tạo và thử nghiệm càng làm phức tạp thêm việc xác định tính hợp lệ và độ tin cậy của các kết quả được báo cáo trong các tài liệu hiện có. Do đó, có nhu cầu cấp thiết về các tiêu chí đánh giá mạnh mẽ và khách quan để so sánh và đối chiếu các phương pháp tiếp cận thị giác máy tính khác nhau được ủng hộ cho việc quản lý an toàn.

### B. Thách thức kỹ thuật

Như đã đề cập trước đó, các nghiên cứu về thị giác máy tính thường dựa vào các cơ sở dữ liệu nhỏ và các phương pháp học có giám sát để xác định hành vi không an toàn. Phương pháp này dẫn đến khái quát hóa yếu do hai yếu tố chính: (a) giả định rằng cơ sở dữ liệu đào tạo và thử nghiệm có cùng phân phối; và (b) những hạn chế của các mô hình học máy đào tạo trên các tập dữ liệu nhỏ, hạn chế sự thay đổi giữa các lớp và trong lớp. Do đó, điều này cản trở khả năng nhận dạng chính xác hành vi không an toàn và khái quát hóa thành các tập dữ liệu khác nhau [42]. Tuy nhiên, các kỹ thuật như học chuyển giao và tăng cường dữ liệu (ví dụ: cắt xén, lật và xoay ngẫu nhiên) có thể giảm thiểu các vấn đề về độ chính xác và độ tin cậy liên quan đến các tập dữ liệu đào tạo nhỏ.

Các mô hình học sâu rất giỏi trong việc học các mối tương quan giữa các tính năng đầu vào và đầu ra, nhưng chúng lại gặp khó khăn trong việc thiết lập quan hệ nhân quả. Việc hiểu được sự tương tác giữa hành vi của con người và môi trường làm việc tương ứng là rất quan trọng để ngăn ngừa thông tin xung quanh các mối nguy hiểm trên các công trường xây dựng. Ví dụ, an toàn dựa trên hành vi (BBS) quan sát và xác định các hành động không an toàn, cung cấp phản hồi trực tiếp cho những cá nhân thực hiện các hành động đó để khuyến khích hành vi an toàn hơn trong tương lai [43,44,45,46].

Mặc dù học sâu chứng minh được khả năng nhận dạng môi trường nguy hiểm, nhưng điều cần thiết là phải thừa nhận rằng các cách tiếp cận này thường mang tính nhiệm vụ cụ thể. Hạn chế này đặt ra một thách thức đáng kể vì không có cách tiếp cận đơn lẻ nào có thể xác định hiệu quả nhiều hành vi không an



toàn, khiến việc triển khai thực tế thị giác máy tính trở nên tốn kém và mất thời gian. Do đó, việc phát triển các thuật toán mới và đào tạo chúng để phát hiện nhiều hành vi và điều kiện không an toàn phổ biến gặp phải trên các công trường xây dựng là rất quan trọng.

Các mô hình học sâu thường hoạt động như 'hộp đen', thiếu tính minh bạch trong quá trình ra quyết định của chúng [47,48,49]. Mặc dù đã có những nỗ lực để giải quyết tình trạng mờ đục này bằng cách trực quan hóa các đóng góp của từng nút trong các mạng phức tạp bao gồm hàng triệu tham số, nhưng việc đạt được tính minh bạch hoàn toàn trong học sâu vẫn là một thách thức chưa được giải quyết [48]. Việc không thể xác định chính xác các tính năng được các nút này trích xuất và học được khiến việc hiểu quy trình phát hiện và xác định các tham số cần điều chỉnh để phát hiện mỗi nguy hiểm chính xác trở nên khó khăn, cản trở việc biện minh cho việc áp dụng học sâu [48].

## VII. CÁC LĨNH VỰC NGHIÊN CỨU TRONG TƯƠNG LAI VỀ HỌC SÂU VÀ THỊ GIÁC MÁY TÍNH

Để vượt qua những thách thức này và đảm bảo ứng dụng hiệu quả và hiệu suất của thị giác máy tính trong giám sát an toàn, chúng tôi đề xuất các lĩnh vực nghiên cứu tiềm năng phù hợp với khuôn khổ mà chúng tôi thiết kế và phát triển.

### A. Kết hợp học sâu và thị giác máy tính với công nghệ kỹ thuật số

Các phương pháp tiếp cận thị giác máy tính hiện tại trong xây dựng thường cho thấy mức độ sử dụng thông tin thấp, đòi hỏi mức độ chính xác cao hơn để phát hiện mỗi nguy hiểm. Khi các quy định về an toàn ngày càng trở nên phức tạp, phụ thuộc lẫn nhau và nghiêm ngặt do các yêu cầu theo luật định, các phương pháp tiếp cận thị giác máy tính hiện tại có thể gặp khó khăn trong việc theo kịp. Nếu những cách tiếp cận này không thích ứng được với nhu cầu pháp lý đang thay đổi và những sắc thái riêng của xây dựng, chúng có nguy cơ trở nên lỗi thời.

1) *Ontology và thị giác máy tính*: Để giải quyết những thách thức do các quy định về an toàn đang phát triển, chúng tôi đề xuất tích hợp ontology với các phương pháp thị giác máy tính. Ontology cung cấp một khuôn khổ kiến thức chính thức, được khái niệm hóa, cung cấp một biểu diễn đơn giản hóa của một miền bằng cách mô tả các đối tượng, khái niệm và mối quan hệ của chúng [50]. Mục đích của nó là cho phép các ứng dụng máy tính biểu diễn và lý luận về kiến thức một cách hiệu quả. Khi kết hợp với thị giác máy tính, các đối tượng có thể được tự động phát hiện và các thuộc tính như lớp và hình học có thể được trích xuất từ hình ảnh. Sự tích hợp này mở đường cho một khuôn khổ thị giác máy tính ngữ nghĩa bao gồm bốn quy trình chính: (1) mô hình hóa ontology về các mối nguy hiểm (ví dụ: hành vi không an toàn và trạng thái nhà máy); (2) phát hiện thực thể và thuộc tính bằng thị giác máy tính; (3) trích xuất các mối quan hệ ngữ nghĩa không gian và thời gian từ dữ liệu video; và (4) lý luận dữ liệu để xác định mỗi nguy hiểm.

Chúng tôi đề xuất rằng một mô hình ngữ nghĩa tích hợp ontology và thị giác máy tính, ngay cả với dữ liệu hạn chế, có thể hiệu quả để xác định mỗi nguy hiểm bằng cách sử dụng học sâu. Phương pháp này không chỉ tận dụng khả năng phát hiện đối tượng chính xác mà còn tận dụng các mối quan hệ không gian-thời gian giữa các đối tượng để lý luận về mỗi nguy hiểm.

Một số nghiên cứu đã chứng minh hiệu suất thỏa đáng của các phương pháp tiếp cận dựa trên thị giác máy tính hiện có trong việc phát hiện nhiều đối tượng khác nhau [51], cho thấy phương pháp tiếp cận ngữ nghĩa được đề xuất này, không yêu cầu cơ sở dữ liệu đào tạo cụ thể, có tiềm năng đáng kể.

2) *Nhóm với dữ liệu trực quan được xây dựng, mô hình theo kế hoạch và IoT*: Một hệ thống phân tích trực quan học sâu được đề xuất để đánh giá hiệu suất của dự án. Hệ thống này sẽ tận dụng mô hình tái tạo ngữ nghĩa ba chiều (3D) [52,53], được xây dựng từ dữ liệu theo kế hoạch và theo xây dựng, và sử dụng các kỹ thuật thị giác máy tính.

Hệ thống sẽ cho phép tạo tự động và có thể mở rộng các mô hình theo kế hoạch 3D chất lượng cao. Điều này sẽ đạt được bằng cách xử lý khối lượng lớn hình ảnh và video từ nhiều nguồn khác nhau và căn chỉnh chúng với mô hình theo kế hoạch.

Ngoài ra, nó sẽ nâng cao nhận thức về tình huống cho mục đích an toàn, hợp lý hóa phân tích khiếu nại và hỗ trợ điều tra tai nạn. Ảnh chụp nhanh và video tại hiện trường, căn chỉnh với mô hình theo kế hoạch, có thể được sử dụng để chú thích, báo cáo, lập tài liệu và truyền thông. Vai trò của thị giác máy tính trong hệ thống này sẽ mở rộng ra ngoài việc trích xuất đối tượng và xác định thuộc tính (ví dụ: khoảng cách, lớp).

Nó cũng sẽ bao gồm việc xây dựng mô hình ngữ nghĩa 3D theo kế hoạch theo thời gian thực. Đồng thời, dữ liệu cảm biến (ví dụ: vị trí) từ các thành phần được lắp đặt có thể được trích xuất và tích hợp với kết quả phát hiện thị giác máy tính thông qua IoT. Sự tích hợp này sẽ tạo điều kiện thuận lợi cho việc lưu trữ và cập nhật dữ liệu liên tục theo thời gian thực trong mô hình 3D của tài sản được xây dựng. Do đó, các mối nguy tiềm ẩn, bao gồm các khiếm khuyết hoặc hỏng hóc về mặt cấu trúc, có thể được xác định trong mô hình.

3) *Dữ liệu trực quan khi xây dựng, AR/VR và mô hình thông tin xây dựng*: Các ứng dụng AR đã được sử dụng rộng rãi trong xây dựng, tăng cường cả môi trường ảo và thực tế [54].

- Các ứng dụng này đã được phát triển để tạo điều kiện thuận lợi cho việc truy xuất thông tin trong quá trình xây dựng và quản lý cơ sở, đặc biệt là vì mục đích an toàn.
- Chúng cho phép trực quan hóa các tiện ích ngầm, tăng cường nhận thức trực quan để đảm bảo an toàn khi đào và kiểm tra tiện ích ngầm, đồng thời cung cấp hướng dẫn vận hành 3D theo thời gian thực được phủ lên trên công trường thực tế để hỗ trợ lắp ráp và các nhiệm vụ phức tạp khác.
- Có thể thu được hướng dẫn vận hành 3D theo thời gian thực được phủ lên trên công trường thực tế để hỗ trợ lắp ráp và các hoạt động phức tạp khác.

Các khả năng AR hiện có và các thiết bị di động (ví dụ: ARKit của Apple dành cho iOS, ARCore của Google dành cho

Android) đã đủ tiên tiến để hỗ trợ các ứng dụng trực quan đã xác định [55]. Tuy nhiên, cần phải cho phép các thiết bị di động có tài nguyên hạn chế hỗ trợ công cụ phân tích được hỗ trợ bởi học sâu.

### B. Thông tin chi tiết từ hợp nhất nhiều dữ liệu

Hợp nhất nhiều dữ liệu liên quan đến việc tích hợp dữ liệu từ nhiều nguồn khác nhau để tạo ra thông tin nhất quán, chính xác và hữu ích hơn so với thông tin mà một nguồn duy nhất có thể cung cấp[56]. Trong an toàn xây dựng, dữ liệu có thể bắt nguồn từ nhiều nguồn khác nhau, bao gồm báo cáo an toàn và cảm biến không trực quan.

1) *Sử dụng luồng video và hợp nhất nhiều mô hình*: Mặc dù học sâu theo truyền thống được sử dụng trong các cài đặt có giám sát với dữ liệu được gắn nhãn, nhưng đã có sự chuyển dịch đáng kể sang các mô hình không giám sát để nâng cao tốc độ phát hiện đối tượng, độ chính xác và độ tin cậy.

Dựa trên những tiến bộ trong học sâu và thị giác máy tính trong khoa học máy tính[57,58,59,60], một phương pháp học không giám sát dựa trên học sâu tự học được đề xuất để xác định mối nguy hiểm bằng cách sử dụng phát trực tuyến video. Trong phương pháp này, luồng video sẽ đóng vai trò là đầu vào cho mô hình học sâu được trang bị cơ chế học tự học. Cơ chế này, với các tham số có thể điều chỉnh, sẽ cho phép tự đào tạo liên tục bằng cách sử dụng luồng video, tạo ra các khung đầu ra để thử nghiệm.

Với sự phong phú của các nguồn dữ liệu hình ảnh cảm biến từ xa sẵn có, đã được sử dụng trong nhiều ứng dụng khác nhau[61-65], chúng tôi đề xuất hợp nhất chúng để nhận dạng mối nguy hiểm tự động trong xây dựng. Ví dụ, bằng cách hợp nhất hình ảnh nhiệt và 2D, có thể phát hiện ra các hành vi không an toàn như hút thuốc trong xây dựng. Điều này có thể thực hiện được vì nhiệt độ của điều thuốc cao hơn nhiệt độ của môi trường xung quanh, giúp dễ dàng phân biệt với các điểm không phải điểm nóng bằng hình ảnh nhiệt [66].

2) *Căn chỉnh giữa báo cáo thị giác máy tính và báo cáo văn bản*: Kết hợp dữ liệu văn bản và hình ảnh cho phép mô hình học sâu suy luận và hiểu bản chất của rủi ro. Sự kết hợp này cung cấp hai hướng nghiên cứu:

- Đầu tiên, tận dụng các báo cáo để nâng cao độ chính xác của việc xác định hành vi không an toàn. Điều này liên quan đến việc sử dụng học sâu và thị giác máy tính để trích xuất và mã hóa hình ảnh từ các biểu diễn đặc điểm (ví dụ: các vùng quan trọng).
- Thứ hai, tự động tạo báo cáo an toàn từ hình ảnh. Các cuộc kiểm tra an toàn tại chỗ hiện tại phụ thuộc rất nhiều vào việc lưu giữ hồ sơ thủ công về các mối nguy hiểm đã quan sát được, sau đó được chuyển vào hệ thống máy tính để tạo báo cáo[66,67].

3) *Căn chỉnh giữa dữ liệu thị giác máy tính và dữ liệu cảm biến không trực quan*: Dữ liệu cảm biến từ nhiều vị trí khác nhau trên công trường xây dựng có thể được sử dụng để phát hiện mối nguy hiểm. Các loại cảm biến khác nhau, chẳng hạn như cảm biến vị trí (ví dụ: Nhận dạng tần số vô tuyến, Hệ thống định vị toàn cầu), đã được triển khai để thu thập dữ liệu an toàn[69-71].

Bằng cách kết hợp hình ảnh với dữ liệu từ nhiều cảm biến không trực quan, phạm vi các mối nguy hiểm có thể phát hiện được bằng thị giác máy tính dựa trên học sâu có thể được mở rộng.

- **Hành vi không an toàn**: Điều này bao gồm việc xác định các hành vi không an toàn. Các cảm biến như cảm biến vị trí và cảm biến danh tính có thể nắm bắt tọa độ và danh tính của một cá nhân[74,75]. Sau đó, thị giác máy tính có thể trích xuất thông tin như lớp đối tượng, hoạt động, thuộc tính và tọa độ 3D của đối tượng[66,72,73].
- **Nhà máy không an toàn**: Cảm biến có thể được sử dụng để thu thập dữ liệu cơ học, chẳng hạn như mômen uốn và góc, từ cần cẩu trong quá trình nâng. Dữ liệu này, kết hợp với thị giác máy tính, cho phép xác định các yếu tố theo ngữ cảnh và điều kiện của nhà máy góp phần gây ra các tình huống không an toàn.
- **Khuyết tật về cấu trúc**: Các nhà nghiên cứu đã sử dụng dữ liệu cảm biến để theo dõi tình trạng cấu trúc nhằm xác định chính xác nguồn gốc khuyết tật[76, 77]. Ngoài ra, thị giác máy tính và học sâu đã được sử dụng để nhận dạng trực quan các khuyết tật. Tuy nhiên, các phương pháp này chưa xác định hiệu quả chất lượng thành phần bên trong của cơ sở hạ tầng dân dụng như đường hầm và cầu. Radar xuyên đất (GPR), một kỹ thuật địa vật lý sử dụng bức xạ điện từ ở tần số UHF/VHF, phát hiện tín hiệu phản xạ từ các cấu trúc ngầm. Chúng tôi đề xuất phát triển một hệ thống chụp cắt lớp vi tính (CT) tích hợp hình ảnh radar và thị giác máy tính dựa trên học sâu để trích xuất thông tin và chẩn đoán chất lượng cấu trúc bên trong.

## VIII. ỨNG DỤNG THỰC TIỄN THỊ GIÁC MÁY TÍNH TRONG AN TOÀN LAO ĐỘNG VÀO HỆ SINH THÁI VINBIGDATA

Công ty Cổ phần VinBigData được thành lập trên nền tảng một số thành quả nghiên cứu khoa học của Viện Nghiên cứu Dữ liệu lớn thuộc tập đoàn Vingroup (thành lập tháng 8/2018) trong lĩnh vực Khoa học Dữ liệu và Trí tuệ Nhân tạo (đặc biệt về xử lý hình ảnh và ngôn ngữ).

Với việc làm chủ công nghệ lõi cùng cơ sở dữ liệu đa ngành quy mô lớn, VinBigdata tập trung phát triển các sản phẩm, giải pháp có tính ứng dụng cao, tăng trải nghiệm người dùng, nâng cao chất lượng cuộc sống và tối ưu hoạt động sản xuất kinh doanh.

Trong bài báo này, hệ sinh thái Vizione là hệ sinh thái giải pháp phân tích hình ảnh thông minh sẽ được thảo luận.

Vizione là hệ sinh thái các giải pháp phân tích hình ảnh ứng dụng công nghệ Thị giác máy tính (CV) và Trí tuệ nhân tạo (AI), bao gồm: Nhận diện khuôn mặt, giám sát an ninh - an toàn, Phân tích hành vi khách hàng, Định danh khách hàng điện tử và Nhận diện giấy tờ tùy thân... Hệ sinh thái tổng hợp các giải pháp công nghệ toàn diện với mục tiêu mang lại cuộc sống an toàn - tiện nghi cho mọi người.

### A. Vizione Secure - Giải pháp camera giám sát thông minh

Vizione Secure là giải pháp camera giám sát thông minh toàn diện cho doanh nghiệp. Được phát triển từ công nghệ Thị giác



Hình 8. Tổng quan hệ sinh thái Vizione

máy tính (CV) và Trí tuệ nhân tạo (AI), Vizione Secure có thể hoạt động hiệu quả trong mọi môi trường với độ chính xác cao, đáp ứng đa dạng nhu cầu và quy mô, giúp tối ưu hóa quy trình vận hành.

#### Tính năng nổi bật

- 99% Nhận diện khuôn mặt chính xác
- 0,3s Tốc độ Nhận diện tức thì
- 90% Phát hiện đối tượng (người, phương tiện,...)

#### Nhận diện vật thể

- Phát hiện thú cưng
- Phát hiện đồ vật bỏ quên trên hành lang, trong thang máy
- Nhận diện vật thể lạ
- Phát hiện đỗ xe trái phép tại vùng cấm đỗ
- Nhận diện đồ vật quá tải so với quy định

#### Phân tích đặc điểm & hành vi

- Phân tích cảm xúc (vui/buồn/hài lòng)
- Nhận diện người đeo khẩu trang
- Phân tích thuộc tính khuôn mặt, giới tính, độ tuổi...
- Phát hiện mang dao/kiếm/có hành vi nguy hiểm

#### Tính năng thống kê

- Đếm người
- Bản đồ nhiệt
- Đếm lượt phương tiện
- Đếm đối tượng, đồ vật theo yêu cầu

#### Nhận diện khuôn mặt

- Nhận diện khách lạ, khách quen
- Nhận diện lượt khách lạ trong ngày
- Nhận diện đúng khách hàng (đã có ảnh trên hệ thống)
- Nhận diện và cảnh báo danh sách đen có sẵn trong dữ liệu

Các tính năng giám sát tự động của Vizione Secure hỗ trợ doanh nghiệp không chỉ trong việc phát hiện các hành vi nguy hiểm mà còn trong việc dự đoán và phòng ngừa tai nạn, tạo ra môi trường làm việc an toàn hơn:

### 1. Phát hiện hành vi nguy hiểm

- Hệ thống Vizione Secure có khả năng phân tích hành vi của công nhân trong thời gian thực, giúp phát hiện các hành vi nguy hiểm như mang vũ khí (dao, kiếm) hoặc thực hiện các thao tác sai quy trình.
- Thị giác máy tính được tích hợp với camera giám sát có thể phát hiện những hành vi không tuân thủ quy định an toàn lao động, chẳng hạn như không sử dụng trang bị bảo hộ hoặc đứng ở những khu vực nguy hiểm.
- Tự động nhận diện và cảnh báo khi phát hiện những hành vi bất thường, giúp người quản lý nhanh chóng can thiệp và phòng ngừa tai nạn.

### 2. Giám sát việc sử dụng thiết bị bảo hộ cá nhân (PPE)

- Thị giác máy tính giúp nhận diện và theo dõi xem công nhân có tuân thủ việc đeo các trang bị bảo hộ như mũ bảo hiểm, găng tay, áo phản quang hay khẩu trang trong khu vực làm việc không.
- Hệ thống sẽ phát hiện và đưa ra cảnh báo nếu phát hiện công nhân không sử dụng các thiết bị bảo hộ theo quy định, giúp doanh nghiệp nâng cao tính tuân thủ an toàn trong nhà máy hoặc công trường.

### 3. Giám sát khu vực nguy hiểm và hạn chế tiếp cận

- Vizione Secure giúp giám sát các khu vực nguy hiểm, khu vực hạn chế tiếp cận như vùng cấm xe đỗ, khu vực có máy móc hoạt động nguy hiểm.
- Hệ thống có khả năng phát hiện việc tiếp cận trái phép vào những khu vực này và ngay lập tức phát tín hiệu cảnh báo, giúp ngăn chặn các tai nạn không đáng có.
- Nhận diện và phân tích vị trí của nhân viên trong thời gian thực giúp quản lý việc ra vào khu vực an toàn, đảm bảo không có tình huống xâm nhập trái phép.

### 4. Phân tích cảm xúc và hành vi của nhân viên

- Công nghệ thị giác máy tính còn có thể được sử dụng để phân tích cảm xúc và hành vi của nhân viên trong môi trường làm việc, từ đó nhận biết được những dấu hiệu mệt mỏi, căng thẳng hay các biểu hiện cảm xúc khác có thể ảnh hưởng đến hiệu suất và sự an toàn.
- Thông qua các phân tích này, doanh nghiệp có thể đưa ra các giải pháp kịp thời để cải thiện điều kiện làm việc và giảm thiểu tai nạn lao động do yếu tố tâm lý hoặc sức khỏe gây ra.

### 5. Tối ưu hóa quy trình kiểm tra an toàn

- Hệ thống giám sát thông minh của Vizione Secure cho phép tự động hóa việc kiểm tra an toàn trong doanh nghiệp. Các hoạt động kiểm tra được ghi nhận và báo cáo tự động, giúp quản lý dễ dàng theo dõi và cải thiện các biện pháp an toàn lao động.
- Việc tích hợp với hệ thống phân tích dữ liệu giúp doanh nghiệp xác định các khu vực hoặc hành vi tiềm ẩn rủi ro, từ đó đưa ra các giải pháp cải tiến hiệu quả.

### 6. Đo lường và báo cáo số liệu an toàn

- Vizione Secure cung cấp các tính năng thống kê chi tiết về các vi phạm an toàn, sự cố xảy ra và những hành vi có nguy cơ cao. Dữ liệu này được thu thập và phân tích

tự động, hỗ trợ nhà quản lý trong việc đưa ra các quyết định cải thiện chính sách an toàn.

- Nhờ khả năng phân tích dữ liệu mạnh mẽ, hệ thống có thể báo cáo những thay đổi tích cực hoặc tiêu cực về mức độ an toàn, từ đó giúp doanh nghiệp điều chỉnh chiến lược an toàn lao động phù hợp.

### B. Vizione Access - Giải pháp kiểm soát ra vào bằng khuôn mặt

Vizione Access là giải pháp kiểm soát ra vào bằng khuôn mặt ứng dụng công nghệ Trí tuệ nhân tạo (AI) và Học sâu (Deep Learning), hỗ trợ các hoạt động chấm công, check-in/out, giám sát ra vào tại cơ sở, nhận ra khách hàng VIP, kiểm soát nhân viên và thực hiện các dịch vụ chuyên biệt.

#### Tính năng nổi bật

- **99,9%** Độ chính xác vượt trội
- **<1s** Tốc độ xử lý nghiệp vụ nhanh
- **24/7** Hoạt động ổn định
- **Nhận diện khuôn mặt**
  - Nhận diện khuôn mặt trong các trường hợp đặc biệt (đeo kính, đeo khẩu trang,...)
- **Phát hiện giả mạo**
- **Tích hợp dễ dàng vào hệ thống sẵn có**
- **Hỗ trợ và quản lý thiết lập báo cáo**

#### Ứng dụng của Vizione Access trong an toàn lao động:

Vizione Access không chỉ được sử dụng trong việc kiểm soát ra vào mà còn hỗ trợ tối ưu trong việc đảm bảo an toàn lao động tại nơi làm việc. Với khả năng nhận diện khuôn mặt trong thời gian thực, hệ thống có thể ngăn chặn những cá nhân không được phép vào khu vực nguy hiểm, đồng thời đảm bảo rằng nhân viên tuân thủ quy định an toàn, chẳng hạn như đeo khẩu trang hay kính bảo hộ. Tính năng phát hiện giả mạo giúp giảm nguy cơ xâm nhập trái phép, trong khi khả năng tích hợp dễ dàng vào các hệ thống giám sát hiện có đảm bảo việc triển khai hiệu quả trong môi trường sản xuất và công trường.

## IX. LỜI CẢM ƠN

Tôi cũng xin bày tỏ lòng biết ơn đến Viện Nghiên cứu Dữ liệu lớn VinBigdata vì những đóng góp quý báu thông qua các buổi thảo luận và sự hướng dẫn trong suốt quá trình thực hiện dự án.

Ngoài ra, tôi xin ghi nhận việc sử dụng ChatGPT, Axriv, ASCE Library, Science Direct, Taylor & Francis đã hỗ trợ rất nhiều trong việc tiến hành nghiên cứu này.

## TÀI LIỆU

- [1] The Ministry of Labor, War invalids and Social Affairs of the People's Republic of Viet Nam: Work accident situation 2023
- [2] International Labour Office 2005 Facts about safety at work.
- [3] Wang M, Wong P, Luo H, Kumar S, Delhi V and Cheng J 2019 Proc. Inter. Symposium on Automation and Robotics in Construction 36 399-406.
- [4] Guo B H, Zou Y, Fang Y, Goh Y M and Zou P X 2021 Safety Sci. 135 105130.
- [5] Arshad S, Akinade O, Bilal M and Bello S 2023 J. Build. Eng. 107049.
- [6] Xu S, Wang J, Shou W, Ngo T, Sadick A M and Wang X 2021 Archi. Comput. Methods Eng. 28 3383-3397.
- [7] Fang W, Love P E, Luo H and Ding L 2020 J. Advanced Engineering Informatics 43 100980.

- [8] Liu W, Meng Q, Li Z and Hu X 2021 Build. 11(9) 409.
- [9] Yan X and Kim Y C 2018 A conceptual framework of ITSMCA for a building collapse accident Eng. Constr. Archit. Ma. 25(6) 721-737.
- [10] Ahn Y, Choi H and Kim B.S 2023 Development of early fire detection model for buildings using computer vision-based CCTV J. Build. Eng. 65 105647.
- [11] Pushkar A, Senthilvel M and Varghese K 2018 Automated progress monitoring of masonry activity using photogrammetric point cloud ISARC 35 1-7.
- [12] Luo H, Wang M, Wong P K Y and Cheng J C 2020 Full body pose estimation of construction equipment using computer vision and deep learning techniques AUTCON 110 103016.
- [13] Zhu D, Wen H and Deng Y 2020 Pro-active warning system for the crossroads at construction sites based on computer vision Eng. Constr. Archit. Ma. 27(5) 1145-1168.
- [14] Wilkins J R 2011 Construction workers' perceptions of health and safety training programmes Constr. Ma. Econom. 29(10) 1017-1026.
- [15] Jeelani I, Albert A and Han K 2020 Improving Safety Performance in Construction Using Eye-Tracking, Visual Data Analytics, and Virtual Reality In Construction Research Congress 2020 American Society Civil Eng pp 395-404.
- [16] Eiris R, Gheisari M and Esmaceli B 2018 PARS: Using augmented 360-degree panoramas of reality for construction safety training Int. J. Environ. Health. Res. 15(11) 2452.
- [17] Fang W, Ding L, Luo H and Love P E 2018 AUTCON 91 53-61.
- [18] Hayat A and Morgado-Dias F 2022 J. Deep learning-based automatic safety helmet detection system for construction safety. Applied Sciences. 12(16) 8268.
- [19] Ding L, Fang W, Luo H, Love P E, Zhong B and Ouyang X 2018 J. A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory. AUTCON. 86 118-124.
- [20] Alateeq M M, Fathimathul Rajeena P P and Ali M A 2023 Construction Site Hazards Identification Using Deep Learning and Computer Vision Sustainability 15(3) 2358.
- [21] Yan X and Kim Y C 2018 A conceptual framework of ITSMCA for a building collapse accident Eng. Constr. Archit. Ma. 25(6) 721-737.
- [22] Park M W and Brilakis I 2012 AUTCON 28 15-25.
- [23] Park M.W, Palinginis E and Brilakis I 2012 Detection of construction workers in video frames for automatic initialization of vision trackers In Construction Research Congress 2012: Construction Challenges in a Flat World pp 940-949.
- [24] Park M W and Brilakis I 2016 AUTCON 72 129-142.
- [25] Liu W, Shao Y, Zhai S, Yang Z and Chen P 2023 IEICE Transactions on Information and Systems 106(5) 653-661.
- [26] Azar E.R 2016 Construction equipment identification using marker-based recognition and an active zoom camera J. Constr. Eng. Ma. 30(3) 04015033.
- [27] Xiao B and Kang S C 2021 Vision-based method integrating deep learning detection for tracking multiple construction machines J. Constr. Eng. Ma. 35(2) 04020071.
- [28] Zhang M, Cao Z, Yang Z and Zhao X 2020 J. Constr. Eng. Ma. 146(6) 04020051.
- [29] Zhang M and Ge S 2022 Vision and trajectory-Based dynamic collision prewarning mechanism for tower cranes J. Constr. Eng. Ma. 148(7) 04022057.
- [30] Tsuchiya K and Ishigami G 2020 Vision-based measurement of spatio-temporal deformation of excavated soil for the estimation of bucket resistive force J. Terramechanics 90 11-21.
- [31] Naghshbandi S N, Varga L and Hu Y 2021 J. Technologies for safe and resilient earthmoving operations: A systematic literature review AUTCON 125 103632.
- [32] Goodrum P M, Zhai D and Yasin M F 2009 J. Relationship between changes in material technology and construction productivity J. Constr. Eng. Ma. 135(4) 278-287.
- [33] Dimitrov A and Golparvar-Fard M 2014 J. Adv. Eng. Inform. 28(1) 37-49.
- [34] Mahami H, Ghassemi N, Darbandy M T, Shoeibi A, Hussain S, Nasirzadeh F, Alizadehsani R, Nahavandi D, Khosravi A and Nahavandi S 2020 J. arXiv preprint arXiv 16344.
- [35] Dinh T H, Ha Q P and La H M 2016 Computer vision based method for concrete crack detection. Int. Control Automation Robotics & Vision (ICARCV) 1-6.

- [36] Koch C, Georgieva K, Kasireddy V, Akinci B and Fieguth P 2015. *Adv. Eng. Inform.* 29(2) 196-210.
- [37] Deng J, Singh A, Zhou Y, Lu Y and Lee V C S 2022 *Constr Build Mater.* 356 129238.
- [38] S. Shetye, S. Shetty, S. Shinde, C. Madhu and A. Mathur, "Computer Vision for Industrial Safety and Productivity," 2023 International Conference on Communication System, Computing and IT Applications (CSCITA), Mumbai, India, 2023, pp. 117-120, doi: 10.1109/CSCITA55725.2023.10104764.
- [39] Wang, Yu. (2024). Research on computer vision application for safety management in construction. *Theoretical and Natural Science.* 30. 232-242. 10.54254/2753-8818/30/20241124.
- [40] Fang, Weili & Ding, Lieyun & Luo, Hanbin & Li, Heng & Peña-Mora, Feniosky & Zhong, Botao & Zhou, Cheng. (2020). Computer vision applications in construction safety assurance. *Automation in Construction.* 110. 103013. 10.1016/j.autcon.2019.103013.
- [41] Guo, B.; Zou, Y.; Chen, L. A Review of the Applications of Computer Vision to Construction Health and Safety. 2018.
- [42] J. Tompson, R. Goroshin, A. Jain, Y. LeCun, C. Bregler, Efficient object localization using convolutional networks, 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 648–656.
- [43] Y. Yu, H. Guo, Q. Ding, H. Li, M. Skitmore, An experimental study of real-time identification of construction workers' unsafe behaviors, *Autom. Constr.* 82 (2017) 193–206
- [44] R.M. Choudhry, Behavior-based safety on construction sites: a case study, *Accid. Anal. Prev.* 70 (2014) 14–23
- [45] H. Luo, W. Fang, P.E.D. Love, L. Ding, Computer vision for behaviour-based safety in construction: A review and future directions, *Adv. Eng. Inform.* 43 (2020) 100980.
- [46] S. Guo, P. Zhang, L. Ding, Time-statistical laws of workers' unsafe behavior in the construction industry: a case study, *Physica A: Statistical Mechanics and Its Applications* 515 (2019) 419–429
- [47] W. Samek, T. Wiegand, K.-R. Müller, Detection of construction workers under varying poses and changing the background in image sequences via very deep residual networks.
- [48] Marcus, G. Deep learning: A critical appraisal. (2018) arXiv preprint arXiv:1801.00631.
- [49] M.T. Ribeiro, S. Singh, C. Guestrin, "Why should I trust you?": Explaining the predictions of any classifier, (2016), <https://doi.org/10.18653/v1/n16-3020> Conference of the North American Chapter of the Associate for Computational Linguistics: Demonstrations
- [50] T. Gruber, A translation approach to portable ontology specifications, *Knowl. Acquis.* 5 (2) (1993) 199–220
- [51] W. Fang, L. Ding, B. Zhong, P.E.D. Love, H. Luo, Automated detection of workers and heavy equipment on construction sites: a convolutional neural network approach, *Adv. Eng. Inform.* 37 (2018) 139–149.
- [52] P. Tang, D. Huber, B. Akinci, R. Lipman, A. Lytle, Automatic reconstruction of asbuilt building information models from laser-scanned point clouds: a review of related techniques, *Autom. Constr.* 19 (2010) 829–843
- [53] R. Lu, I. Brilakis, Digital twinning of existing reinforced concrete bridges from labeled point clusters, *Autom. Constr.* 105 (2019).
- [54] M. Chu, J. Matthews, P.E.D. Love, Integrating mobile building information modeling and augmented reality systems: An experimental study, *Autom. Constr.* 85 (2018) 305–316
- [55] M. J. Kim, X. Wang, P.E.D. Love, H. Li, S. C. Kang, Virtual reality for the built environment: a critical review of recent advances, *Journal of Information Technology in Construction.* 18 (2), 279–305
- [56] J. Llinas, D.L. Hall, An introduction to multisensor data fusion, in *The Proceedings of the IEEE* (1997), pp. 6–23.
- [57] H. Hu, N. Phan, J. Geller, H. Vo, B. Manasi, X. Huang, S. Di Lorio, T. Dinh, S.A. Chun, Deep self-taught learning for detecting drug abuse risk behavior in tweets, *Computational Data and Social Networks*, Springer International Publishing, Cham, 2018, pp. 330–342
- [58] P. Kang, D. Kim, S. Cho, Semi-supervised support vector regression based on selftraining with label uncertainty: An application to virtual metrology in semiconductor manufacturing, *Expert Syst. Appl.* 51 (2016) 85–106.
- [59] X. Tang, F. Guo, J. Shen, T. Du, Facial landmark detection by semi-supervised deep learning, *Neurocomputing* 297 (2018) 22–32 <https://doi.org/10.1016/j.neucom.2018.01.080>
- [60] P.K. Rhee, E. Erdenee, S.D. Kyun, M.U. Ahmed, S. Jin, Active and semi-supervised learning for object detection with imperfect data, *Cogn. Syst. Res.* 45 (2017) 109–123 <https://doi.org/10.1016/j.cogsys.2017.05.006>
- [61] Aparna, Y. Bhatia, R. Rai, V. Gupta, N. Aggarwal, A. Akula, Convolutional neural networks based potholes detection using thermal imaging, *Journal of King Saud University - Computer and Information Sciences*, (2019).
- [62] E. Shi, Q. Li, D. Gu, Z. Zhao, A Method of Weather Radar Echo Extrapolation Based on Convolutional Neural Networks, *MultiMedia Modeling*, Springer International Publishing, Cham, 2018, pp. 16–28
- [63] P. Ding, Y. Zhang, W.-J. Deng, P. Jia, A. Kuijper, A light and faster regional convolutional neural network for object detection in optical remote sensing images, *ISPRS J. Photogramm. Remote Sens.* 1
- [64] N. Qin, X. Hu, H. Dai, Deep fusion of multi-view and multimodal representation of ALS point cloud for 3D terrain scene recognition, *ISPRS J. Photogramm. Remote Sens.* 143 (2018) 205–212
- [65] V.N. Nguyen, R. Jenssen, D. Roverso, Automatic autonomous vision-based power line inspection: a review of current status and the potential role of deep learning, *Int. J. Electr. Power Energy Syst.* 99 (2018) 107–120
- [66] H. Luo, C. Xiong, W. Fang, P.E.D. Love, B. Zhang, X. Ouyang, Convolutional neural networks: computer vision-based workforce activity assessment in construction, *Autom. Constr.* 94 (2018) 282–289
- [67] M.Z. Hossain, F. Sohel, M.F. Shiratuddin, H. Laga, A comprehensive survey of deep learning for image captioning, *ACM Comput. Surv.* 51 (2019) 1–36
- [68] U. Nadeem, S.A.A. Shah, F. Sohel, R. Togneri, M. Bennamoun, Deep learning for scene understanding, *Handbook of Deep Learning Applications*, Springer International Publishing, Cham, 2019, pp. 21–51
- [69] H. Li, M. Lu, S.-C. Hsu, M. Gray, T. Huang, Proactive behavior-based safety management for construction safety improvement, *Saf. Sci.* 75 (2015) 107–117
- [70] W. Lee, K.-Y. Lin, E. Seto, G.C. Migliaccio, Wearable sensors for monitoring on-duty and off-duty worker physiological status and activities in construction, *Automconstr.* 83 (2017) 341–353
- [71] S. Dong, H. Li, Q. Yin, Building information modeling in combination with real-time location systems and sensors for safety performance enhancement, *Saf. Sci.* 102 (2018) 226–237
- [72] W. Fang, L. Ding, B. Zhong, P.E.D. Love, H. Luo, Automated detection of workers and heavy equipment on construction sites: a convolutional neural network approach, *Adv. Eng. Inform.* 37 (2018) 139–149
- [73] X. Luo, H. Li, D. Cao, Y. Yu, X. Yang, T. Huang, Towards efficient and objective work sampling: recognizing workers' activities in site surveillance videos with twostream convolutional networks, *Autom. Constr.* 94 (2018) 360–370
- [74] T. Cheng, J. Teizer, G.C. Migliaccio, U.C. Gatti, Automated task-level activity analysis through fusion of real-time location sensors and worker's thoracic posture data, *Autom. Constr.* 29 (2013) 24–39
- [75] A. Kelm, L. Laußat, A. Meins-Becker, D. Platz, M.J. Khazaei, A.M. Costin, M. Helmus, J. Teizer, Mobile passive radio frequency identification (RFID) portal for automated and rapid control of personal protective equipment (PPE) on construction sites, *Autom. Constr.* 36 (2013) 38–52
- [76] D. Yang, J. Wang, D. Li, K.S.C. Kuang, Fatigue crack monitoring using plastic optical fibre sensor, *Procedia Structural Integrity* 5 (2017) 1168–1175
- [77] X. Jin, S. Yuan, J. Chen, On crack propagation monitoring by using reflection spectra of AFBG and UFBG sensors, *Sensors Actuators A Phys.* 285 (2019) 491–500
- [78] Mahony N, Campbell S, Carvalho A, Harapanahalli S, Hernandez G V, Krpalkova L, Riordan D, Walsh J 2010 Binary robust independent elementary features dblp, , deep learning vs. Traditional computer vision in advances in computer vision proceedings of the 2019 computer vision conference (cvc). Springer nature switzerland ag, pp. 128-144
- [79] Website Source: <https://www.analyticsvidhya.com/blog/2020/05/social-distancing-detection-tool-deeplearning> (last updated on 15-Oct-2020)
- [80] Website: <https://www.osha.gov> (last updated on 15-Oct-2020)
- [81] Seo J et al. 2015 Computer vision techniques for construction safety and health monitoring, *adv. Eng. Informat.* <http://dx.doi.org/10.1016/j.aei.2015.02.001>
- [82] Ringen K, Seegal J and England A 1995 Safety and health in the construction industry, *ann. Rev. Public health* 16 (1) pp 165–188
- [83] Reese C D and Eidson J V 2006 Handbook of osha construction safety and health, crc press

- [84] Website: <https://www.osha.gov/personal-protective-equipment> (last updated on 15-Oct-2020)
- [85] Osha, worker safety series: construction Website: <https://www.osha.gov/publications/ osha3252/3252.html> (last updated on 15-Oct-2020)