NHẬN DIỆN HÀNH VI HỌC TẬP CỦA SINH VIÊN TRONG LỚP HỌC

Trần Anh Tú[1], Trần Thế Lộc[2], Phạm Đình Minh Trưởng[3]
[1]1771020707, [2]1771020427, [3]1771020705

Khoa Công Nghệ Thông Tin, Đại Học Đại Nam

Github: tutran56/AI

Tóm tắt-Nhận dạng hành vi học tập của sinh viên trong lớp học đóng vai trò quan trọng trong việc giám sát và cải thiện hiệu quả giảng dạy. Trong nghiên cứu này, mô hình YOLOv7, một phương pháp phát hiện đối tượng nổi bật, được áp dụng để nhận diện hành vi học tập của sinh viên từ video trong lớp học. Dữ liệu video được thu thập từ các camera trong lớp học và gắn nhãn các hành vi như "giơ tay", "sử dụng laptop", "sử dụng điện thoại", v.v. Kết quả huấn luyện cho thấy YOLOv7 có khả năng nhận dạng hành vi học tập chính xác và hỗ trợ giảng viên trong việc theo dõi và can thiệp kịp thời, nhằm tối ưu hóa phương pháp giảng dạy.

I.GIỚI THIỆU

Nhận dạng hành vi học tập của sinh viên trong lớp học là một trong những yếu tố quan trọng giúp nâng cao chất lượng giảng dạy. Việc theo dõi hành vi học tập của sinh viên giúp giảng viên hiểu rõ hơn về mức độ tham gia của sinh viên trong giờ học, qua đó có thể điều chỉnh phương pháp giảng dạy sao cho phù hợp và hiệu quả**[1]**. Các hành vi học tập như “ngồi học”, “ngủ gật”, “sử dụng điện thoại” hay “tương tác với giảng viên” đều có ảnh hưởng lớn đến sự tập trung và hiệu quả học tập. Tuy nhiên, việc theo dõi và nhận diện chính xác những hành vi này thủ công là điều không thể làm được trong môi trường lớp học đông sinh viên. Do đó, việc áp dụng công nghệ vào nhận dạng hành vi học tập là một hướng đi mới và mang lại nhiều lợi ích trong việc giám sát và cải thiện chất lượng giảng dạy**[2]**.Việc sử dụng công nghệ thông minh để thu thập và phân tích dữ liệu có thể xác định hành vi trong lớp học một cách kịp thời và toàn diện hơn, hiểu rõ hơn về tình trạng của giáo viên và học sinh trong quá trình giảng dạy bền vững trên lớp học và cung cấp một công cụ mạnh mẽ để cải thiện chất lượng giảng dạy**[3]**

Trong những năm gần đây, các mô hình học sâu (Deep Learning) đã được ứng dụng rộng rãi trong việc nhận dạng đối tượng và hành vi từ dữ liệu video hoặc hình ảnh**[4]**. Một trong những mô hình nổi bật trong lĩnh vực nhận dạng đối tượng là YOLO (You Only Look Once), đặc biệt là phiên bản YOLOv7 **[5]**. Đây là mô hình phát hiện đối tượng nhanh chóng và chính xác, nổi bật với khả năng xử lý nhiều đối tượng trong cùng một khung hình mà không làm giảm độ chính xác. Với khả năng phát hiện đối tượng trong thời gian thực và áp dụng cho môi trường học tập, YOLOv7 rất phù hợp để nhận diện hành vi học tập của sinh viên trong lớp học.

Mục tiêu chính của nghiên cứu này là phát triển và ứng dụng mô hình YOLOv7 để nhận dạng hành vi học tập của sinh viên từ video ghi lại trong lớp học. Bằng cách sử dụng YOLOv7, chúng tôi muốn giúp giảng viên có thể theo dõi và đánh giá hành vi học tập của sinh viên một cách tự động và chính xác hơn. Các hành vi được nhận diện bao gồm “ngồi học”, “ngủ gật”, “sử dụng điện thoại” và “tương tác với giảng viên”**[6]**. Một trong những ưu điểm chính của YOLO v7 là tốc độ. Nó có thể xử lý hình ảnh với tốc độ 155 khung hình mỗi giây, nhanh hơn nhiều so với các thuật toán phát hiện đối tượng hiện đại khác. Ngay cả mô hình YOLO cơ bản ban đầu cũng có khả năng xử lý ở tốc độ tối đa 45 khung hình mỗi giây. Điều này làm cho nó phù hợp với các ứng dụng thời gian thực nhạy cảm như giám sát hành vi của sinh viên, trong đó tốc độ xử lý cao hơn là rất quan trọng.Về độ chính xác, YOLO v7 thể hiện tốt so với các thuật toán phát hiện đối tượng khác. Nó đạt được độ chính xác trung bình là 37,2% ở ngưỡng IoU (giao điểm trên hợp nhất) là 0,5 trên bộ dữ liệu COCO phổ biến, có thể so sánh với các thuật toán phát hiện đối tượng hiện đại khác. So sánh định lượng của hiệu suất được hiển thị dưới đây.

Bài báo cáo này trình bày về việc ứng dụng mô hình YOLOv7 vào quá trình nhận diện hành vi học tập của sinh viên trong lớp học, giúp giảng viên theo dõi và đánh giá mức độ tham gia của sinh viên một cách tự động, chính xác và hiệu quả hơn. Trong môi trường giảng dạy truyền thống, việc quan sát và ghi nhận hành vi của từng sinh viên thường bị giới hạn bởi khả năng tập trung của giảng viên, đặc biệt là trong các lớp học đông. Những hành vi như “ngồi học”, “ngủ gật”, “sử dụng điện thoại” và “tương tác với giảng viên” đóng vai trò quan trọng trong quá trình tiếp thu kiến thức và phản ánh mức độ tương tác của sinh viên với bài giảng. Tuy nhiên, việc theo dõi thủ công không chỉ mất nhiều thời gian mà còn khó đảm bảo tính chính xác và khách quan. Với khả năng nhận diện nhanh và chính xác trong thời gian thực, mô hình YOLOv7 giúp giảng viên nắm bắt được tình trạng của sinh viên, can thiệp kịp thời khi cần thiết.

Model	#Param.	FLOPs	Size	AP^{val}	AP^{val}_{50}	AP^{val}_{75}	AP^{val}_{S}	AP^{val}_{M}	AP^{val}_{L}
YOLOv4 [3]	64.4M	142.8G	640	49.7%	68.2%	54.3%	32.9%	54.8%	63.7%
YOLOR-u5 (r6.1) [81]	46.5M	109.1G	640	50.2%	68.7%	54.6%	33.2%	55.5%	63.7%
YOLOv4-CSP [79]	52.9M	120.4G	640	50.3%	68.6%	54.9%	34.2%	55.6%	65.1%
YOLOR-CSP [81]	52.9M	120.4G	640	50.8%	69.5%	55.3%	33.7%	56.0%	65.4%
YOLOv7	36.9M	104.7G	640	**51.2%**	**69.7%**	**55.5%**	**35.2%**	**56.0%**	**66.7%**
improvement	-43%	-15%	-	+0.4	+0.2	+0.2	+1.5	=	+1.3
YOLOR-CSP-X [81]	96.9M	226.8G	640	52.7%	**71.3%**	57.4%	36.3%	57.5%	68.3%
YOLOv7-X	71.3M	189.9G	640	**52.9%**	71.1%	**57.5%**	**36.9%**	**57.7%**	**68.6%**
improvement	-36%	-19%	-	+0.2	-0.2	+0.1	+0.6	+0.2	+0.3
YOLOv4-tiny [79]	6.1	6.9	416	24.9%	42.1%	25.7%	8.7%	28.4%	39.2%
YOLOv7-tiny	6.2	5.8	416	**35.2%**	**52.8%**	**37.3%**	**15.7%**	**38.0%**	**53.4%**
improvement	+2%	-19%	-	+10.3	+10.7	+11.6	+7.0	+9.6	+14.2
YOLOv4-tiny-3l [79]	8.7	5.2	320	30.8%	47.3%	32.2%	**10.9%**	31.9%	51.5%
YOLOv7-tiny	6.2	3.5	320	**30.8%**	**47.3%**	**32.2%**	10.0%	**31.9%**	**52.2%**
improvement	-39%	-49%	-	=	=	=	-0.9	=	+0.7
YOLOR-E6 [81]	115.8M	683.2G	1280	55.7%	73.2%	60.7%	40.1%	**60.4%**	69.2%
YOLOv7-E6	97.2M	515.2G	1280	**55.9%**	**73.5%**	**61.1%**	**40.6%**	60.3%	**70.0%**
improvement	-19%	-33%	-	+0.2	+0.3	+0.4	+0.5	-0.1	+0.8
YOLOR-D6 [81]	151.7M	935.6G	1280	56.1%	73.9%	61.2%	**42.4%**	60.5%	69.9%
YOLOv7-D6	154.7M	806.8G	1280	56.3%	73.8%	61.4%	41.3%	60.6%	70.1%
YOLOv7-E6E	151.7M	843.2G	1280	**56.8%**	**74.4%**	**62.1%**	40.8%	**62.1%**	**70.6%**
improvement	=	-11%	-	+0.7	+0.5	+0.9	-1.6	+1.6	+0.7

Hình 1: So sánh các mô hình nhận diện đối tượng

II.CÁC NGHIÊN CỨU LIÊN QUAN

Phát hiện đối tượng là một hướng nghiên cứu chính trong thị giác máy tính. Thông qua các bước thu thập hình ảnh, xử lý trước hình ảnh, trích xuất đặc điểm có liên quan và phân loại kết quả, việc xác định và phân loại các mục tiêu khác nhau như người đi bộ, phương tiện và đồ nội thất có thể được thực hiện trong nhiều lĩnh vực nghiên cứu khác nhau **[7]**.

Joseph Redmon và cộng sự [**8**] đã đề xuất mô hình YOLO-v1, trong đó khái niệm một giai đoạn được áp dụng, các danh mục đối tượng và hộp giới hạn được dự đoán cùng một lúc, điều này làm cho tốc độ đào tạo của thuật toán yolo khá nhanh, nhưng tỷ lệ chính xác giảm nhẹ. Sau đó, Joseph Redmon đã đề xuất các thuật toán cải tiến như YOLO-v2 [**9**], YOLO-v3 [**10**] và YOLO-v4 [**11**], giải quyết được vấn đề khó khăn trong việc nhận dạng mục tiêu nhỏ. Ngoài ra, nhiều mục tiêu có thể được phát hiện trong một lưới duy nhất, giúp đạt được hiệu quả phát hiện mục tiêu tốt hơn. Wei Liu và cộng sự [**12**] đã đề xuất thuật toán phát hiện SSD (Single Shot MultiBox Detector), giống như chuỗi yolo, có thể dự đoán các danh mục đối tượng và hộp giới hạn cùng một lúc.

Giám sát hành vi học tập của sinh viên trong lớp học đã trở thành một trong những yếu tố quan trọng giúp cải thiện chất lượng giảng dạy và học tập.Phương pháp truyền thống như quan sát thủ công không thể đáp ứng được yêu cầu giám sát liên tục và toàn diện trong lớp học đông sinh viên [**13**]. Do đó, việc ứng dụng các công nghệ mới trong giám sát hành vi học tập đã trở thành một xu hướng quan trọng.

Các cảm biến quán tính như gia tốc kế (accelerometer) và con quay hồi chuyển (gyroscope) đã được ứng dụng rộng rãi trong giám sát hành vi học tập của sinh viên**14**. Các thiết bị này có thể gắn ở nhiều vị trí trên cơ thể, như cổ tay, cánh tay, hông, hoặc thậm chí trên bàn chân, và thu thập dữ liệu về chuyển động của sinh viên. Khi kết hợp với các mô hình học sâu (Deep Learning)**15**, đặc biệt là các mô hình Convolutional Neural Networks (CNN) và Recurrent Neural Networks (RNN), có thể nhận diện các hành vi học tập của sinh viên một cách chính xác, như tư thế ngồi học, sử dụng điện thoại hoặc thậm chí là ngủ gật trong lớp học.

Liu et al. **[16]** đề xuất một mạng YOLO-v3 cải tiến, trong đó giới thiệu một mô-đun RFB cải tiến theo tầng để tăng cường khả năng trích xuất đặc điểm của mạng ban đầu và tận dụng tối đa thông tin nông để cải thiện hiệu ứng nhận dạng các mục tiêu nhỏ. Nhằm giải quyết vấn đề che khuất ký tự do cấu trúc lớp học và mật độ học sinh, mô-đun resn của Darknet-53 trong YOLOv3 được thay thế bằng mô-đun SE-Res2net để hiện thực hóa việc tái sử dụng các đặc điểm nhiều lớp. Nghiên cứu của Rajamanickam Yuvaraj (2021) sử dụng CNN và học sâu để nhận diện hành vi lớp học từ cảm biến và camera, giúp theo dõi chính xác hoạt động học tập. Phương pháp này vượt trội hơn quan sát thủ công, giảm gánh nặng giảng dạy và hỗ trợ quản lý giáo dục hiệu quả.**[17]**. Các nghiên cứu về nhận diện hành vi học tập trong lớp học có sự liên kết chặt chẽ thông qua việc ứng dụng CNN và học sâu (Deep Learning). Guo et al. **[18]** sử dụng CNN đa nhiệm vụ để xác định tốc độ ngẩng đầu (HRR) và trích xuất đặc điểm khuôn mặt

học sinh.Tương tự, **Rong Fu[19]** cũng áp dụng CNN để phân tích video, giúp theo dõi nhiều sinh viên đồng thời mà không cần sự can thiệp của giảng viên. Bên cạnh đó, **Ge Chen** (2022) sử dụng OpenPose để nhận diện tư thế ngồi, cử chỉ tay từ video, chứng minh rằng phân tích hình ảnh và video là một hướng tiếp cận hiệu quả trong giám sát hành vi học tập**[20]**.

Ngoài phân tích hình ảnh, một số nghiên cứu lại tập trung vào cảm biến chuyển động kết hợp với học sâu để nhận diện hành vi chính xác hơn. Hongmin Wang **[21]** sử dụng thuật toán VB-DTW để trích xuất đặc trưng từ tín hiệu cảm biến.Một hướng tiếp cận khác là nhận diện hành vi theo thời gian thực, giúp cải thiện tính ứng dụng. **Lin, Lihua** (2023)[22] đã tối ưu mô hình bằng RT DETR và MobileNetV3, giúp giảm số tham số nhưng vẫn duy trì độ chính xác cao, dễ triển khai trên thiết bị di động. Điều này cho thấy, dù sử dụng phương pháp nào—CNN, cảm biến hay tối ưu mô hình thời gian thực—các nghiên cứu đều hướng đến mục tiêu chung là tự động hóa giám sát lớp học, nâng cao hiệu quả giảng dạy và quản lý giáo dục.

Các nghiên cứu này không chỉ cung cấp những công nghệ hiệu quả để giám sát hành vi học tập mà còn mở ra tiềm năng trong việc tự động hoá quá trình giám sát lớp học. Việc ứng dụng các công nghệ như cảm biến, video giám sát và các mô hình học sâu giúp giảm thiểu sự can thiệp thủ công của giảng viên. Từ đó, giảng viên có thể nâng cao chất lượng giảng dạy và can thiệp kịp thời khi phát hiện hành vi không tích cực của sinh viên, đồng thời tạo ra một môi trường học tập hiệu quả hơn.

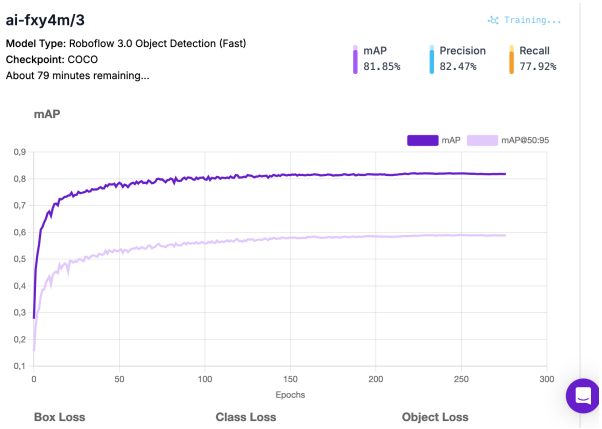
III.CÁC PHƯƠNG PHÁP NGHIÊN CỨU

A.Thu thập dữ liệu

Bước đầu tiên trong nghiên cứu là thu thập dữ liệu video được quay bằng điện thoại của thành viên trong nhóm.Camera điện thoại sẽ ghi lại các hành vi học tập của sinh viên trong suốt quá trình học. Những hành vi này bao gồm các hành động cụ thể như "ngồi học", "ngủ gật", "sử dụng điện thoại" và "tương tác với giảng viên". Các hành vi học tập này được ghi lại nhằm phục vụ việc nhận diện và phân tích hành vi học tập của sinh viên trong lớp học, qua đó hỗ trợ giảng viên trong việc cải thiện phương pháp giảng dạy và quản lý lớp học.

Mỗi video thu thập được từ lớp học sẽ được xử lý và chuyển đổi thành các khung hình (frames). Mỗi khung hình đại diện cho một khoảnh khắc trong video và chứa các thông tin về hành vi của sinh viên trong thời gian đó. Quá trình chuyển đổi này là một bước quan trọng trong việc chuẩn bị dữ liệu cho mô hình học sâu, vì YOLOv7 (và các mô hình học sâu khác) hoạt động dựa trên dữ liệu hình ảnh. Việc chuyển video thành các khung hình sẽ giúp mô hình phân tích chính xác từng hành vi của sinh viên trong lớp học một cách chi tiết và rõ ràng hơn.

Điều này giúp tối ưu hóa quá trình huấn luyện và nhận diện hành vi học tập trong các video, đồng thời nâng cao hiệu quả trong việc phân loại các hành vi học tập, giúp giảng viên có thể theo dõi và can thiệp kịp thời.



Hình 2:Đánh giá dataset

B.Chuyển video thành hình ảnh

Sau khi thu thập video từ lớp học, bước tiếp theo trong quá trình tiền xử lý dữ liệu là chuyển video thành hình ảnh. Quá trình này được thực hiện bằng cách trích xuất từng khung hình (frame) từ video. Việc chuyển đổi này giúp phân tách video thành các đơn vị nhỏ hơn (khung hình) mà mô hình YOLOv7 có thể xử lý.

- Cắt video thành các khung hình: Video chứa rất nhiều thông tin trong mỗi giây, và để mô hình có thể nhận diện các hành vi học tập từ video, chúng ta cần trích xuất từng khung hình để mô hình có thể phân tích từng khoảnh khắc cụ thể. Mỗi khung hình sẽ được xử lý như một bức ảnh đơn lẻ, chứa các thông tin về hành vi của sinh viên tại thời điểm đó. Việc trích xuất các khung hình giúp mô hình tập trung vào việc nhận diện hành vi học tập mà không phải xử lý toàn bộ video, làm tăng hiệu quả và tốc độ huấn luyện.

- Xử lý khung hình: Các khung hình thu được từ video sẽ được xử lý và chuẩn bị để huấn luyện mô hình. Quá trình này bao gồm việc chỉnh sửa kích thước ảnh, loại bỏ nhiễu, và đảm bảo rằng các khung hình có chất lượng tốt, dễ dàng cho mô hình nhận diện hành vi học tập. Mỗi khung hình sẽ được điều chỉnh kích thước để phù hợp với đầu vào của mô hình YOLOv7 sẽ là 640x640 pixel. Điều này giúp mô hình có thể xử lý ảnh với tốc độ nhanh chóng và hiệu quả.

C.Tiền xử lý dữ liệu

- Gắn nhãn các hành vi học tập: Sau khi thu thập các khung hình từ video, mỗi khung hình sẽ được gắn nhãn để chỉ rõ hành vi học tập mà sinh viên đang thực hiện tại thời điểm đó. Những hành vi này bao gồm `using_phone`, `sleep`, `bend`, `bow_head`, `hand_raising`, `raisehead`, `reading`, `turn_head`, `upright`, `writing`, `using_laptop` và "tương tác với giảng viên". Việc gắn nhãn chính xác là rất

quan trọng, vì mô hình YOLOv7 sẽ học từ những nhãn này để phân biệt các hành vi học tập và nhận diện chính xác trong các khung hình mới trong quá trình nhận diện.

- Chia nhỏ các đoạn video thành các cửa sổ trượt: Để giúp mô hình nhận diện hành vi học tập trong một khoảng thời gian liên tục và có tính chính xác cao, các video dài sẽ được chia thành các cửa sổ trượt. Mỗi cửa sổ trượt bao gồm một số lượng khung hình nhất định, giúp mô hình nhận diện hành vi học tập trong từng khoảng thời gian ngắn. Phương pháp này sẽ giúp mô hình có thể theo dõi hành vi của sinh viên qua nhiều khung hình và nhận diện sự thay đổi hành vi, ví dụ như sự chuyển từ `using_laptop` sang `using_phone`.

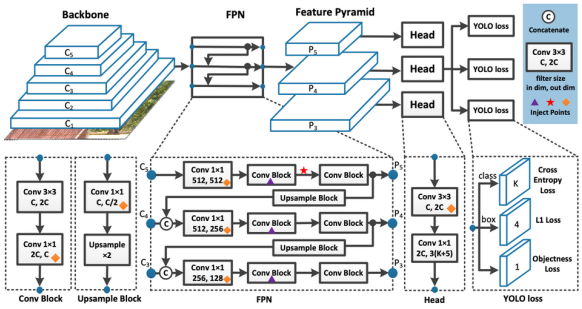
- Loại bỏ nhiễu và chuẩn bị dữ liệu cho mô hình: Dữ liệu video thường có thể chứa nhiễu hoặc các yếu tố không liên quan đến hành vi học tập, như ánh sáng không ổn định hoặc chuyển động không rõ ràng. Để đảm bảo mô hình YOLOv7 có thể học chính xác và hiệu quả, các khung hình sẽ được xử lý để loại bỏ những yếu tố này. Các khung hình mờ, không rõ ràng hoặc bị che khuất sẽ được loại bỏ, chỉ giữ lại các khung hình rõ ràng có liên quan đến hành vi học tập của sinh viên. Dữ liệu sau khi đã được xử lý sẽ sẵn sàng để huấn luyện mô hình.

D. Mô Hình YOLOv7

Mô hình YOLOv7 (You Only Look Once version 7) là một mô hình nhận diện đối tượng hiện đại, được phát triển để nhận diện đối tượng trong hình ảnh và video với độ chính xác cao và tốc độ nhanh **[23]**.. Mô hình YOLOv7 nổi bật với khả năng nhận diện đồng thời nhiều đối tượng trong một khung hình mà không làm giảm hiệu suất. Điều này rất quan trọng khi giám sát hành vi học tập của nhiều sinh viên trong lớp học.

YOLOv7 sử dụng mạng nơ-ron tích chập (CNN) để phát hiện các đối tượng trong hình ảnh. Các đối tượng trong nghiên cứu của chúng tôi là các hành vi học tập của sinh viên trong lớp học, như "ngồi học", "ngủ gật", "sử dụng điện thoại", và "tương tác với giảng viên".

YOLOv7 có khả năng nhận diện nhanh chóng và chính xác nhiều đối tượng trong cùng một khung hình, điều này đặc biệt quan trọng khi giám sát hành vi học tập của nhiều sinh viên trong lớp học. Một điểm mạnh của YOLOv7 là việc sử dụng Attention Mechanism, giúp mô hình tập trung vào các khu vực quan trọng trong khung hình, nơi các hành vi học tập đáng chú ý diễn ra. Kỹ thuật này giúp mô hình nhận diện hành vi học tập chính xác hơn và hiệu quả hơn, vì nó có thể bỏ qua những phần không quan trọng và tập trung vào các hành vi chính. Ngoài ra, YOLOv7 còn được tối ưu hóa để hoạt động trên nhiều nền tảng khác nhau, giúp việc triển khai trở nên linh hoạt và dễ dàng hơn.



Hình 3: Kiến trúc YOLOv7

Các công thức trong Yolov7

Hàm mất mát tổng quát:

$$Loss = \lambda_{coord} Loss_{bbox} + Loss_{obj} + \lambda_{cls} Loss_{cls}$$

Cân bằng giữa dự đoán vị trí, sự tồn tại và loại đối tượng.

CIoU Loss (Bounding Box):

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b_{gt})}{c^2} + \alpha v$$

Cải thiện độ chính xác và hình dạng của hộp dự đoán.

Objectness Loss (BCE Loss):

$$Loss_{obj} = -\sum [y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})]$$

Xác định xem có đối tượng trong ảnh hay không.

Focal Loss (Phân loại hành vi):

$$Loss_{cls} = -\sum_{i} \alpha_i (1 - \hat{p}_i)^{\gamma} y_i \log(\hat{p}_i)$$

Tăng độ chính xác trong phân loại hành vi của sinh viên.

Phân bổ nhãn động (DLA):

$$S = IoU \times 1(IoU > T_{pos})$$

Chọn vùng quan trọng để huấn luyện hiệu quả hơn.

Attention Mechanism:

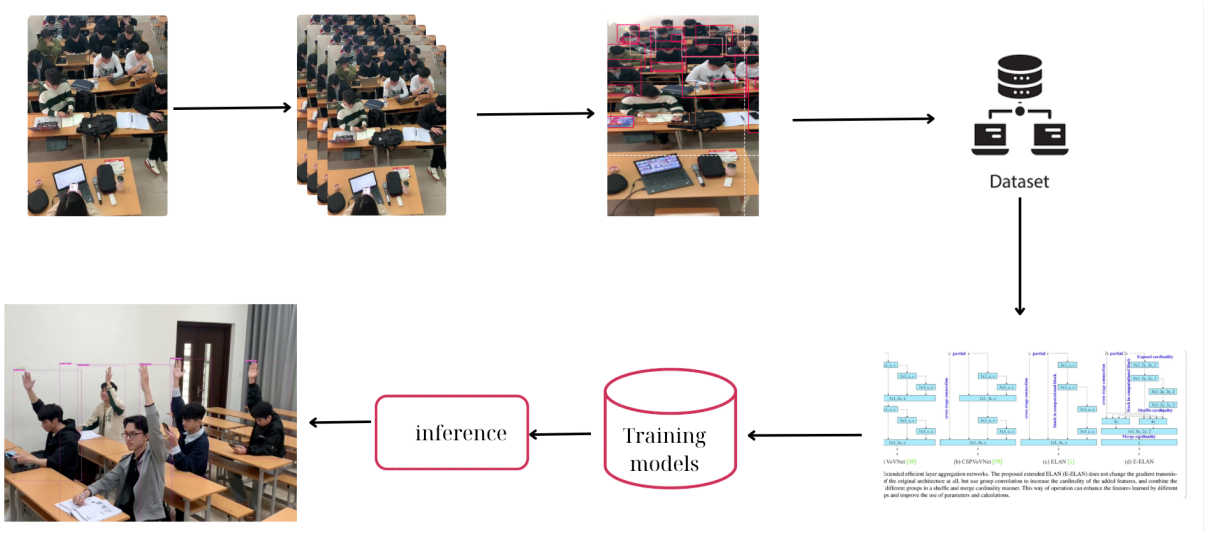
$$A = \sigma(W_Q Q + W_K K + W_V V)$$

Giúp mô hình tập trung vào các vùng cần giám sát như sinh viên mất tập trung.

D. Huấn luyện mô hình

Sau khi hoàn tất việc chuẩn bị dữ liệu, mô hình YOLOv7 sẽ được huấn luyện trên bộ dữ liệu đã thu thập được từ lớp học. Quá trình huấn luyện sẽ bao gồm việc điều chỉnh các tham số như batch size, epochs, và learning rate để tối ưu hóa khả năng nhận diện hành vi học tập.

1. **Epochs**: Số lần mô hình sẽ được huấn luyện trên toàn bộ dữ liệu. Một epoch gồm việc mô hình học và cập nhật các tham số dựa trên các dữ liệu đầu vào. Quá trình huấn luyện sẽ lặp lại qua nhiều epoch để mô hình dần dần cải thiện độ chính xác trong việc nhận diện hành vi học tập. Kỷ nguyên : 50



Hình 4: Mô hình chuyển đổi để nhận dạng hành động

2. **Batch size**: Số lượng khung hình hoặc mẫu dữ liệu được xử lý trong mỗi lần huấn luyện. Batch size là một tham số quan trọng trong quá trình huấn luyện, vì nó ảnh hưởng đến tốc độ và hiệu quả của mô hình. Kích thước : 16

3. **Learning rate**: Đây là tham số xác định tốc độ mà mô hình cập nhật các trọng số trong mỗi lần huấn luyện. Tốc độ này sẽ ảnh hưởng trực tiếp đến khả năng học và tốc độ hội tụ của mô hình. Tỷ lệ học tập : 0,001

4. **Image size**: Kích thước của ảnh đầu vào cho mô hình. Việc điều chỉnh kích thước ảnh giúp mô hình xử lý ảnh nhanh hơn và chính xác hơn. Kích thước hình ảnh : 640x640

IV.KẾT QUẢ VÀ ĐÁNH GIÁ

A.Kết quả huấn luyện

Mô hình YOLOv7 được huấn luyện trên bộ dữ liệu video thu thập từ lớp học, với các hành vi học tập như "sử dụng điện thoại", "ngủ", "sử dụng laptop", và "viết bài"... Sau quá trình huấn luyện, kết quả cho thấy mô hình đạt được độ chính xác lên đến **85%** khi nhận diện các hành vi học tập của sinh viên trong lớp học.

Điều này chứng tỏ rằng mô hình YOLOv7 có khả năng nhận diện hành vi học tập một cách khá chính xác, ngay cả trong môi trường lớp học đông sinh viên với nhiều hành vi diễn ra đồng thời. Ví dụ, mô hình có thể nhận diện khi một sinh viên vừa dùng điện thoại vừa vừa dùng laptop trong cùng một khung hình, điều này là rất quan trọng vì giúp giảng viên theo dõi và đánh giá hành vi học tập của sinh viên trong suốt buổi học mà không bỏ sót bất kỳ hành vi nào.

Ngoài ra, mô hình YOLOv7 có thể xử lý nhiều đối tượng trong cùng một khung hình và nhận diện các hành vi học tập của từng sinh viên một cách đồng thời. Điều này là một điểm mạnh lớn của YOLOv7, vì lớp học thường có nhiều sinh viên và việc theo dõi hành vi của từng người một cách riêng biệt là một thách thức. Mô hình này không chỉ nhận diện hành vi học tập của một sinh viên mà còn theo dõi nhiều sinh viên cùng lúc, làm tăng khả năng giám sát và hiệu quả trong quá trình giảng dạy.

B.Đánh giá mô hình

Để đánh giá độ chính xác của mô hình, chúng tôi sử dụng các chỉ số như **precision**, **recall**, và **mAP@0.5**, là những chỉ số phổ biến trong lĩnh vực nhận diện đối tượng và học sâu. Các chỉ số này giúp xác định khả năng nhận diện đúng các hành vi học tập của sinh viên và mức độ hoàn thiện của mô hình trong việc phân loại các hành vi.

Tên lớp	Độ chính xác
Using_phone	0.918
bend	0.938
book	0.957
bow_head	0.977
hand-raising	0.725
phone	0.889
raise_head	0.963
reading	0.703
sleep	0.923
turn_head	0.850
turn_headd	0.985
upright	0.587
all classes	0.868

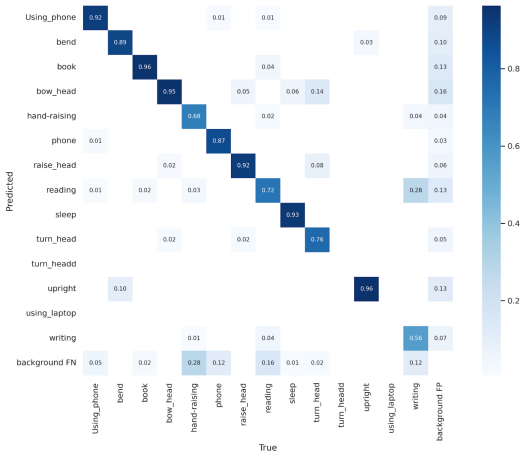
Bảng 1 : Độ chính xác của lớp

(a) **Precision**: Chỉ số precision cho biết tỷ lệ khung hình được nhận diện chính xác so với tổng số khung hình mà mô hình dự đoán là hành vi học tập. Precision càng cao càng cho thấy mô hình ít có sai sót trong việc nhận diện hành vi học tập không chính xác.

(b) **Recall**: Chỉ số recall cho biết tỷ lệ các hành vi học tập thực tế đã được nhận diện chính

xác. Recall cao cho thấy mô hình có khả năng nhận diện hành vi học tập tốt, kể cả khi sinh viên có những hành vi không rõ ràng hoặc ít phổ biến.

(c) **mAP@0.5** (Độ chính xác trung bình ở ngưỡng IoU 0.



Hình 5:Ma trận nhầm lẫn của mô hình

Kết quả đánh giá cho thấy YOLOv7 có thể nhận diện các hành vi học tập của sinh viên trong lớp học với độ chính xác cao, đáp ứng yêu cầu giám sát hành vi trong thời gian thực. Mô hình đã đạt được kết quả ấn tượng, với precision và recall lần lượt là 84.3%, trong khi mAP@0.5 đạt **86.8%**. Những kết quả này chứng tỏ rằng YOLOv7 là một công cụ hiệu quả trong việc nhận diện hành vi học tập của sinh viên trong lớp học, có thể hoạt động tốt trong môi trường thực tế với nhiều sinh viên và các hành vi đa dạng.

Mô hình YOLOv7 không chỉ nhận diện chính xác các hành vi học tập mà còn có khả năng nhận diện nhiều hành vi trong cùng một khung hình, điều này cho phép giảng viên giám sát hành vi học tập của sinh viên trong thời gian thực một cách hiệu quả mà không cần phải giám sát từng sinh viên một cách thủ công.

V.Kết Luận

Bài báo này nghiên cứu và ứng dụng mô hình YOLOv7 để nhận dạng hành vi học tập của sinh viên trong lớp học. Qua các thí nghiệm và huấn luyện trên bộ dữ liệu thu thập từ lớp học, kết quả cho thấy YOLOv7 có khả năng nhận diện các hành vi học tập với độ chính xác và tốc độ cao. Đặc biệt, mô hình có thể nhận diện đồng thời nhiều hành vi trong cùng một khung hình, giúp giảng viên giám sát hành vi học tập của sinh viên một cách hiệu quả trong thời gian thực.

Các kết quả đo được trong thực nghiệm thể hiện rằng mô hình YOLOv7 đạt được mAP (Mean Average Precision) là 86.8% ở mức độ 0.5, với những hành vi học tập như "ngồi học", "ngủ gật", và "sử dụng điện thoại" được nhận diện chính xác. Điều này chứng minh rằng YOLOv7 có thể hỗ trợ giảng viên trong việc theo dõi và can thiệp kịp thời khi sinh viên có hành vi không tích cực trong lớp học. Các chỉ số precision và recall cũng cho thấy mô hình YOLOv7 có thể hoạt động hiệu quả, với độ chính xác cao trong việc nhận diện từng hành vi học tập cụ thể như "ngồi học" (0.92), "ngủ gật" (0.93) và "sử dụng điện thoại" (0.92).

Trong tương lai, YOLOv7 có thể mở rộng để nhận diện các hành vi học tập phức tạp hơn như tương tác giữa sinh viên và giảng viên, hoặc tham gia hoạt động nhóm. Nghiên cứu có thể thử nghiệm với các cảm biến khác để cải thiện độ chính xác và linh hoạt. Đồng thời, việc tích hợp công nghệ nhận diện hình ảnh, âm thanh và mô hình học sâu như LSTM hoặc Transformer sẽ nâng cao hiệu quả giám sát, giúp giảng viên can thiệp kịp thời và cải thiện chất lượng giảng dạy trong lớp học.

LỜI CẢM ƠN

Xin chân thành cảm ơn thầy Lê Trung Hiếu và tất cả những người đã hỗ trợ tôi trong quá trình thực hiện nghiên cứu này.Sự giúp đỡ của mọi người đã đóng góp quan trọng vào thành công của bài báo cáo.

TÀI LIỆU THAM KHẢO

[1] Chuck Chung Yin Albert, Yuqi Sun, Guang Li, Jun Peng (2022). *Identifying and Monitoring Students' Classroom Learning Behavior Based on Multisource Information*.

[2] Rustam Shadiev, Zi Heng Zhang, Ting-Ting Wu, and Yueh Min Huang (2020). *Review of Studies on Recognition Technologies and Their Applications Used to Assist Learning and Instruction*.

[3] Aras Bozkurt (2021). *Artificial Intelligence and Reflections from Educational Landscape: A Review of AI Studies in Half a Century*.

[4] Di Wu, Nabin Sharma, Michael Blumenstein (2017). *Recent Advances in Video-Based Human Action Recognition Using Deep Learning: A Review*.

[5] Songjiang Li, Shilong Wang, Peng Wang (2023). *A Small Object Detection Algorithm for Traffic Signs Based on Improved YOLOv7*.

[6] Lin Wang, Liang Chen, Wenhong Liu (2024). *A Study on Student Behavior Recognition Based on YOLOv7*.

[7] Zhong-Qiu Zhao (2023). *Object Detection with Deep Learning: A Review*.

[8] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi (2016). *You Only Look Once: Unified, Real-Time Object Detection*.

[9] Joseph Redmon, Ali Farhadi (2017). *YOLO9000: Better, Faster, Stronger*.

[10] Joseph Redmon, Ali Farhadi (2018). *YOLOv3: An Incremental Improvement*.

[11] Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao (2020). *YOLOv4: Optimal Speed and Accuracy of Object Detection*.

[12] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg (2016). *SSD: Single Shot MultiBox Detector*.

[13] Mark D. Shermis (2022). *Classroom Assessment in Action*.

[14] Danyal Khan, Naif Al Mudawi, Maha Abdelhaq, Abdulwahab Alazeb, Saud S. Alotaibi (2024). *A Wearable Inertial Sensor Approach for Locomotion and Localization Recognition on Physical Activity*.

[15] Neeraj Varshney, Brijesh Bakariya, Alok Kumar Singh Kushwaha, Manish Khare (2022). *Human Activity Recognition by Combining External Features with Accelerometer Sensor Data Using Deep Learning Network Model*.

[16] Huayong Liu R. (2021). *Student Abnormal Behavior Recognition in Classroom Video Based on Deep Learning*.

[17] Rajamanickam Yuvaraj, A. Amalin Prince, M. Murugappan (2021). *Automated Recognition of Teacher and Student Activities in the Classroom: A Deep Learning Framework*.

[18] Qiang Guo (2020). *Detection of Head Raising Rate of Students in Classroom Based on Head Posture Recognition*.

[19] Rong Fu, Tongtong Wu, Zuying Luo, Fuqing Duan, Xuejun Qiao, Ping Guo (2020). *Learning Behavior Analysis in Classroom Based on Deep Learning*.

[20] Ge Chen, Jianqiang Ji, Chaofeng Huang (2022). *Student Classroom Behavior Recognition Based on OpenPose and Deep Learning*.

[21] Hongmin Wang, Chi Gao, Hong Fu, Christina Zong-Hao Ma, Quan Wang, Ziyu He (2023). *Automated Student Classroom Behaviors' Perception and Identification Using Motion Sensors*.

[22] Lin Lihua, Yang Haodong, Xu Qingchuan, Xue Yanan, Li Dan (2024). *Research on Student Classroom Behavior Detection Based on the Real-Time Detection Transformer Algorithm*.

[23] Benfano Soewito (2023). *Multi-Object Detection Using YOLOv7 Object Detection Algorithm on Mobile Device*.