ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

ꕥ★ꕥ

Tên

TÊN ĐỀ TÀI

LUẬN VĂN THẠC SĨ

NGÀNH CÔNG NGHỆ THÔNG TIN

MÃ SỐ:

TP. HỒ CHÍ MINH, NĂM 2025

ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

ꕥ★ꕥ

Tên

TÊN ĐỀ TÀI

LUẬN VĂN THẠC SĨ

NGÀNH CÔNG NGHỆ THÔNG TIN

MÃ SỐ:

NGƯỜI HƯỚNG DẪN KHOA HỌC

TP. HỒ CHÍ MINH, NĂM 2024

LỜI CAM ĐOAN

Tôi xin cam kết luận này này do chính tôi thực hiện dưới sự hướng dẫn của thầy TS., TS.. Tôi không sao chép bất kỳ công trình nghiên cứu nào khác mà không trích dẫn rõ nguồn gốc.

Dữ liệu sử dụng trong luận văn có nguồn gốc rõ ràng. Các mã nguồn do tôi tự viết, các thư viện được sử dụng trong mã nguồn đều không vi phạm bản quyền.

Các số liệu trong kết quả nghiên cứu là trung thực và kết quả nghiên cũng chưa từng được công bố trên bất kỳ công trình nào khác.

Tôi xin cam đoan các điều bên trên là hoàn toàn đúng sự thật. Nếu có bất kỳ vi phạm nào ở các điều trên, tôi xin chịu hoàn toàn trách nhiệm và nhận hình thức kỷ luật từ Khoa và Nhà Trường.

Người thực hiện

LỜI CẢM ƠN

Lời đầu tiên tôi xin cảm ơn TS. Nguyễn Tấn Hoàng Phước đã tận tình và chu đáo hướng dẫn tôi trong suốt quãng thời gian thực hiện luận văn.

Tôi xin chân thành bày tỏ lòng biết ơn đến quý thầy cô trong Viện đào tạo Sau Đại học trường Đại học Công nghệ Thông tin đã cung cấp cho tôi những kiến thức, kinh nghiệm quý báu cho tôi trong suốt quá trình học tập và nghiên cứu tại trường.

Tôi cũng xin kính gởi lời cảm ơn đến gia đình và bạn bè cũng như những người thân đã luôn luôn quan tâm và giúp đỡ tôi trong suốt quá trình học tập và nghiên cứu để hoàn thành luận văn này.

Do thời gian và kiến thức có hạn cho nên luận văn sẽ không thể tránh khỏi những thiếu sót nhất định. Tôi rất cầu mong nhận được sự góp ý và lời chỉ dạy quý báu của quý thầy cô.

MỤC LỤC

DANH MỤC CÁC KÝ HIỆU VÀ CHỮ VIẾT TẮT

Viết tắt	Tên tiếng Anh	Tên tiếng Việt
CNN	Convolutional Neural Network	Mạng nơ-ron tích chập
ViT	Vision Transformer	Bộ biến đổi thị giác
YOLO	You Only Look Once	Chỉ cần nhìn một lần
RCNN	Region-Based Convolutional Neural Network	Mạng nơ-ron tích chập dựa trên vùng
FPS	Frames Per Second	Số khung hình mỗi giây
GPU	Graphics Processing Unit	Bộ xử lý đồ họa
AI	Artificial Intelligence	Trí tuệ nhân tạo
DL	Deep Learning	Học sâu
ML	Machine Learning	Học máy
IoU	Intersection over Union	Độ chồng lấp giữa các vùng
mAP	Mean Average Precision	Độ chính xác trung bình
FER	Facial Emotion Recognition	Nhận diện cảm xúc trên khuôn mặt
SERS	Student Emotion Recognition System	Hệ thống nhận diện cảm xúc học sinh

DANH MỤC HÌNH ẢNH

DANH MỤC BẢNG

LỜI MỞ ĐẦU

Trong thời đại công nghệ số hiện nay, trí tuệ nhân tạo (AI) đã trở thành một trong những công nghệ cốt lõi, được ứng dụng rộng rãi trong nhiều lĩnh vực như y tế, tài chính, sản xuất, và đặc biệt là giáo dục. Với sự phát triển nhanh chóng của các mô hình học sâu (Deep Learning), AI đang dần thay đổi cách chúng ta tiếp cận việc giảng dạy và học tập, mang lại những cải tiến vượt bậc trong việc cá nhân hóa trải nghiệm học tập, tự động hóa các quy trình quản lý, và nâng cao hiệu quả giảng dạy. Một trong những xu hướng nổi bật hiện nay là việc sử dụng AI để phân tích cảm xúc và nhận dạng danh tính học sinh trong lớp học, nhằm giúp giáo viên hiểu rõ hơn về trạng thái cảm xúc, mức độ tham gia và tình trạng học tập của học sinh.

Cảm xúc của học sinh đóng vai trò quan trọng trong quá trình học tập, không chỉ ảnh hưởng đến khả năng tiếp thu kiến thức mà còn tác động đến mối quan hệ giữa giáo viên và học sinh. Một học sinh vui vẻ, thoải mái sẽ dễ dàng tham gia vào các hoạt động học tập, trong khi những cảm xúc tiêu cực như lo lắng, buồn bã, hoặc tức giận có thể làm giảm khả năng tập trung và gây ra những trở ngại trong quá trình học tập. Tuy nhiên, việc nhận biết và theo dõi cảm xúc của từng học sinh trong lớp học đông đúc là một thách thức lớn đối với giáo viên. Đó là lý do tại sao các hệ thống phân tích cảm xúc tự động, sử dụng công nghệ AI, đang trở thành một giải pháp tiềm năng để hỗ trợ giáo viên trong việc nâng cao hiệu quả giảng dạy.

Đề tài "Hệ thống phân tích và đánh giá chất lượng trong môi trường giáo dục" tập trung vào việc nghiên cứu và phát triển một hệ thống tích hợp các mô hình học sâu để phân tích cảm xúc và nhận dạng danh tính học sinh. Hệ thống này không chỉ cung cấp thông tin về trạng thái cảm xúc của học sinh trong thời gian thực mà còn giúp giáo viên theo dõi danh tính của từng học sinh, từ đó đánh giá mức độ tham gia và hiệu quả của các phương pháp giảng dạy. Việc tích hợp khả năng nhận diện danh tính và cảm xúc trong một hệ thống duy nhất mở ra những cơ hội mới cho việc cá nhân hóa trải nghiệm học tập, đồng thời tạo điều kiện để giáo viên đưa ra các quyết định giảng dạy chính xác và kịp thời hơn.

Hệ thống được phát triển dựa trên các công nghệ tiên tiến như mạng nơ-ron tích chập (CNN) và các mô hình học sâu hiện đại như YOLOv8-Classify và ArcFace. Bên cạnh đó, nghiên cứu này cũng xây dựng bộ dữ liệu cảm xúc đặc thù cho trẻ em trong lớp học, đảm bảo rằng hệ thống có thể hoạt động hiệu quả trong các điều kiện thực tế. Với khả năng phân tích và xử lý dữ liệu trong thời gian thực, hệ thống không chỉ hỗ trợ giáo viên mà còn đóng vai trò quan trọng trong việc nâng cao chất lượng giáo dục, đặc biệt trong các bối cảnh như lớp học trực tuyến, trường học thông minh, và các chương trình đào tạo cá nhân hóa.

Bằng cách kết hợp giữa công nghệ AI và giáo dục, đề tài này không chỉ giải quyết các thách thức hiện tại mà còn góp phần định hình tương lai của ngành giáo dục. Việc áp dụng hệ thống phân tích cảm xúc và nhận dạng danh tính học sinh không chỉ giúp cải thiện chất lượng giảng dạy mà còn tạo ra một môi trường học tập tích cực, nơi mỗi học sinh được thấu hiểu và hỗ trợ một cách tối ưu. Với tiềm năng ứng dụng rộng lớn, nghiên cứu này hứa hẹn sẽ mang lại những đóng góp quan trọng cho việc xây dựng các lớp học thông minh và nâng cao trải nghiệm học tập cho thế hệ trẻ.

CHƯƠNG 1. TỔNG QUAN

1.1. Lý do chọn đề tài

Giáo dục chất lượng cao là yếu tố thiết yếu đối với sự phát triển của trẻ em, đặc biệt trong giai đoạn trung nhi đồng (từ 5 đến 12 tuổi), một giai đoạn quan trọng đối với sự phát triển cảm xúc, xã hội và nhận thức[1], [2]. Trong giai đoạn này, khả năng điều chỉnh và xử lý cảm xúc ảnh hưởng sâu sắc đến kết quả học tập, hành vi và mối quan hệ với giáo viên cũng như bạn bè, định hình các kết quả lâu dài đối với trẻ em [3]. Một môi trường cảm xúc tích cực khuyến khích sự tham gia và thành tích học tập, trong khi đó, những khó khăn trong việc điều chỉnh cảm xúc có thể cản trở quá trình học và nhấn mạnh sự cần thiết của việc quan tâm đến cảm xúc trong lớp học [4], [5].

Tuy nhiên, mặc dù vai trò quan trọng của cảm xúc trong động lực và hiệu quả lớp học đã được nhận thức rõ, vẫn còn thiếu các bộ dữ liệu và hệ thống mô hình toàn diện để đánh giá mối quan hệ giữa yếu tố cảm xúc và học tập trong các bối cảnh giáo dục thực tế [6]. Nhiều nghiên cứu hiện nay tập trung đánh giá hoặc chất lượng lớp học hoặc khả năng điều chỉnh cảm xúc, nhưng vẫn chưa tích hợp một cách toàn diện các khía cạnh này [7]. Ngoài ra, các bộ dữ liệu hiện tại thường bị hạn chế về tính đại diện cho nhóm tuổi trung nhi đồng hoặc các thống số cảm xúc đơn giản hóa [8].

Ví dụ, trong khi các bộ dữ liệu như “AffectNet” và “FER-2013” cung cấp các hình ảnh đã được gán nhãn cho nhận diện cảm xúc, chúng chủ yếu tập trung vào người lớn hoặc các danh mục cảm xúc nhất định, khiến chúng không thích hợp cho trẻ em trong bối cảnh lớp học [9], [10]. Tương tự, các nghiên cứu như “DEAP” và “GoEmotions” tuy mang lại những hiểu biết quý giá về nhận diện cảm xúc qua dữ liệu sinh lý hoặc văn bản, nhưng vẫn thiếu dữ liệu ngữ cảnh từ môi trường giáo dục thực tế [11], [12]. Khoảng cách này càng rõ rệt hơn đối với trẻ em từ 5 đến 12 tuổi, nơi các bộ dữ liệu chuyên biệt là cần thiết để hiểu rõ mối quan hệ phức tạp giữa cảm xúc và việc học [13].

Để khắc phục những hạn chế này, nghiên cứu này giới thiệu EmoLearn, một bộ

dữ liệu mới được thiết kế đặc biệt để đánh giá hiệu quả lớp học, động lực cảm xúc và nhận diện danh tính trong nhóm tuổi từ 5 đến 12. Những đóng góp chính của nghiên cứu bao gồm:

- Phát triển bộ dữ liệu EmoLearn nhằm đánh giá các động lực cảm xúc và tương tác trong lớp học đối với trẻ em từ 5 đến 12 tuổi.
- Huấn luyện và đánh giá ba mô hình trên bộ dữ liệu EmoLearn để thiết lập các chỉ số so sánh và kiểm chứng tính hữu ích của nó.
- Xây dựng khung chính sách đánh giá hiệu quả lớp học dựa trên động lực cảm xúc và mức độ tham gia của học sinh.
- Tạo ra một hệ thống phân loại cảm xúc để phân tích trạng thái cảm xúc trong các tình huống lớp học thực tế dựa trên chính sách hiệu quả lớp học.

Thông qua nghiên cứu này, chúng tôi kỳ vọng sẽ thu hẹp khoảng cách giữa lý thuyết và thực tiễn, đồng thời cung cấp một nguồn tài nguyên toàn diện hỗ trợ tích hợp các yếu tố cảm xúc và học tập trong đánh giá lớp học. Đóng góp này có tiềm năng thay đổi cách tiếp cận giáo dục, tạo ra các môi trường không chỉ hiệu quả về mặt học thuật mà còn hỗ trợ tích cực về mặt cảm xúc.

1.2. Mục tiêu nghiên cứu

1.2.1. Mục tiêu tổng quát

Mục tiêu tổng quát của luận văn này là phát triển một hệ thống dữ liệu và mô hình phân tích toàn diện nhằm đánh giá hiệu quả lớp học thông qua việc tích hợp các yếu tố động lực cảm xúc và mức độ tham gia của học sinh trong bối cảnh giáo dục thực tế. Trọng tâm là cung cấp một công cụ hỗ trợ mạnh mẽ để nâng cao hiệu quả giảng dạy và cải thiện chất lượng giáo dục cho trẻ em trong độ tuổi từ 5 đến 12, một giai đoạn phát triển quan trọng của đời sống học sinh.

Luận văn không chỉ tập trung vào việc thiết kế và triển khai các công cụ công nghệ mà còn hướng tới việc kết nối giữa lý thuyết và thực tiễn giáo dục, tạo ra một nguồn tài nguyên đáng tin cậy cho việc phân tích và quản lý cảm xúc học sinh trong

môi trường lớp học. Qua đó, nghiên cứu này kỳ vọng đóng góp vào việc xây dựng những lớp học không chỉ hiệu quả về mặt học thuật mà còn là nơi học sinh cảm thấy được hỗ trợ và khích lệ về mặt cảm xúc.

1.2.2. Mục tiêu cụ thể

Xây dựng bộ dữ liệu EmoLearn:

Một bộ dữ liệu chất lượng cao là nền tảng cho nghiên cứu về cảm xúc và hiệu quả lớp học. Do đó, mục tiêu đầu tiên của luận văn là xây dựng bộ dữ liệu EmoLearn, tập trung vào các đặc điểm cảm xúc và mức độ tham gia của học sinh trong độ tuổi 5-12. Bộ dữ liệu sẽ được thu thập từ nhiều nguồn đa dạng, phản ánh sự phong phú về ngữ cảnh giáo dục, văn hóa và hành vi học sinh.

Quá trình tăng cường dữ liệu sẽ được thực hiện để đảm bảo rằng bộ dữ liệu không chỉ toàn diện mà còn có khả năng áp dụng trong các mô hình học máy tiên tiến. Tăng cường dữ liệu bao gồm các phương pháp như tạo dữ liệu nhân tạo, xử lý ảnh và chuẩn hóa để cải thiện tính đại diện và giảm thiểu các sai lệch tiềm năng.

Huấn luyện và đánh giá bộ dữ liệu EmoLearn sử dụng các mô hình học sâu:

Hệ thống sẽ sử dụng các mô hình hiện đại như YOLOv8 Classify, Vision Transformer (ViT), và Faster RCNN để nhận diện và phân loại cảm xúc học sinh từ dữ liệu hình ảnh và video trong lớp học. Mỗi mô hình sẽ được đánh giá về độ chính xác, khả năng tổng quát hóa và hiệu quả thời gian thực.

Các mô hình sẽ được tích hợp để tạo thành một hệ thống liên kết từ nhận diện khuôn mặt, phân loại cảm xúc đến nhận diện danh tính của học sinh. Quá trình huấn luyện sẽ bao gồm tối ưu hóa các tham số và so sánh với các mô hình khác để đảm bảo hiệu quả cao nhất.

Đề xuất chính sách hiệu quả của lớp học dựa trên cảm xúc:

Một khung sách độ hiệu quả của lớp học sẽ được xây dựng, tập trung vào việc đánh giá hiệu quả lớp học dựa trên cảm xúc học sinh. Các chỉ số cảm xúc sẽ được

chuẩn hóa thành các thang đo định lượng, từ đó đưa ra độ hiệu quả của lớp học.

Chính sách này không chỉ hỗ trợ giáo viên nhận diện các vấn đề tiềm tàng trong lớp học mà còn cung cấp các gợi ý hành động cụ thể, chẳng hạn như điều chỉnh phương pháp giảng dạy hoặc thay đổi cấu trúc lớp học để cải thiện trải nghiệm học tập của học sinh.

Xây dựng hệ thống đánh giá mức độ hiệu quả của lớp học:

Hệ thống đánh giá được thiết kế để kết hợp dữ liệu từ các mô hình nhận diện khuôn mặt, phân loại cảm xúc và nhận diện danh tính, tạo thành một chuỗi phân tích toàn diện. Hệ thống sẽ cung cấp thông tin chi tiết về trạng thái cảm xúc chung của lớp học, sự tương tác giữa giáo viên và học sinh, cũng như động lực học tập.

Công cụ này sẽ tích hợp giao diện trực quan, dễ sử dụng, cho phép giáo viên theo dõi và điều chỉnh môi trường lớp học theo thời gian thực. Ngoài ra, việc sử dụng nhiều camera (multi-camera) trong hệ thống sẽ giúp loại bỏ các cảm xúc không đúng hoặc những trường hợp nhận diện sai, nâng cao tính chính xác và độ tin cậy của kết quả dự đoán.

1.3. Đối tượng nghiên cứu

Các đối tượng nghiên cứu chính của luận văn này bao gồm:

1.3.1. Dữ liệu cảm xúc và dữ liệu cảm xúc của trẻ em từ 5-12 tuổi

Dữ liệu cảm xúc đóng vai trò trung tâm trong nghiên cứu này, đặc biệt là dữ liệu từ nhóm trẻ em trong độ tuổi từ 5 đến 12. Đây là giai đoạn phát triển quan trọng, nơi cảm xúc của trẻ có ảnh hưởng sâu sắc đến khả năng học tập, các mối quan hệ xã hội và hành vi trong lớp học. Các bộ dữ liệu hiện có như AffectNet và FER-2013 tuy có giá trị nhưng chưa cung cấp đủ thông tin chi tiết về cảm xúc của trẻ em hoặc không đại diện đầy đủ cho nhóm tuổi này. Để khắc phục hạn chế, nghiên cứu sẽ phát triển bộ dữ liệu EmoLearn, tập trung vào cảm xúc trong các ngữ cảnh thực tế của lớp học, bao gồm các biểu hiện trên khuôn mặt, ngôn ngữ cơ thể và các yếu tố ngữ cảnh như tình trạng ánh

sáng, âm thanh, và tương tác giữa giáo viên và học sinh. Bộ dữ liệu này sẽ được thu thập từ nhiều nguồn, áp dụng các tiêu chuẩn kiểm tra chất lượng nghiêm ngặt để đảm bảo tính đại diện và tính ứng dụng cao trong phân tích cảm xúc học đường.

1.3.2. Các mô hình học sâu

Nghiên cứu sẽ triển khai một loạt các mô hình học sâu tiên tiến để phân tích dữ liệu cảm xúc, trong đó các mô hình như YOLOv8 Classify, Vision Transformer (ViT) và Faster RCNN đóng vai trò then chốt. YOLOv8 Classify sẽ được sử dụng cho các tác vụ nhận diện cảm xúc nhanh chóng và chính xác trong thời gian thực. Vision Transformer sẽ khai thác các đặc trưng phức tạp từ dữ liệu hình ảnh, giúp nhận diện các cảm xúc khó phân biệt và trích xuất thông tin từ các bối cảnh phức tạp. Faster RCNN sẽ đảm nhận việc nhận diện khuôn mặt và phân loại cảm xúc chi tiết, hỗ trợ việc phân tích đa lớp và cải thiện độ chính xác. Việc tích hợp các mô hình này trong một hệ thống thống nhất không chỉ nâng cao hiệu quả mà còn tạo điều kiện để xử lý các bài toán phân tích cảm xúc trong môi trường lớp học một cách toàn diện. Các bước huấn luyện và tối ưu hóa mô hình sẽ được tiến hành cẩn thận để đảm bảo khả năng hoạt động hiệu quả trên dữ liệu thực tế.

1.3.3. Các hệ thống đánh giá hiệu quả lớp học

Hệ thống đánh giá hiệu quả lớp học sẽ kết hợp nhiều thành phần để tạo nên một công cụ phân tích toàn diện, từ nhận diện khuôn mặt, phân loại cảm xúc, đến đánh giá tương tác giữa học sinh và giáo viên. Với sự hỗ trợ từ các camera đa góc, hệ thống sẽ loại bỏ các trường hợp nhận diện không chính xác do góc nhìn không phù hợp hoặc điều kiện ánh sáng không tốt. Đối tượng nghiên cứu bao gồm các phương pháp đo lường cảm xúc tổng thể của lớp học, mức độ tham gia của học sinh và động lực học tập. Hệ thống này không chỉ cung cấp dữ liệu theo thời gian thực mà còn tạo ra các báo cáo chi tiết về hiệu quả giảng dạy, giúp giáo viên hiểu rõ hơn về trạng thái cảm xúc của học sinh và đưa ra các điều chỉnh phù hợp. Những đánh giá này sẽ dựa trên các thang đo cảm xúc chuẩn hóa, cung cấp các chỉ số định lượng để so sánh và theo dõi hiệu quả lớp học trong thời gian dài.

1.3.4. Phương pháp tăng cường dữ liệu và xử lý dữ liệu

Một phần quan trọng trong nghiên cứu là các phương pháp tăng cường dữ liệu và xử lý dữ liệu để đảm bảo độ chính xác và tính đáng tin cậy của hệ thống. Tăng cường dữ liệu bao gồm các kỹ thuật như xoay ảnh, thay đổi độ sáng, thêm nhiễu, và chỉnh sửa màu sắc để làm giàu bộ dữ liệu và tăng khả năng tổng quát hóa của các mô hình học sâu. Các bước tiền xử lý, chẳng hạn như làm sạch dữ liệu, loại bỏ các mẫu không hợp lệ và chuẩn hóa dữ liệu, sẽ được thực hiện cẩn thận để giảm thiểu sai số. Ngoài ra, hệ thống sẽ sử dụng các phương pháp phân tích đa góc độ từ camera đa nguồn để loại bỏ những thông tin không nhất quán, tăng cường khả năng nhận diện cảm xúc chính xác trong các tình huống phức tạp. Phương pháp này không chỉ cải thiện hiệu quả của các mô hình mà còn góp phần xây dựng một cơ sở dữ liệu mạnh mẽ, hỗ trợ nghiên cứu tiếp theo.

1.4. Phạm vi nghiên cứu

Phạm vi đối tượng:

Nghiên cứu này tập trung vào nhóm học sinh trong độ tuổi từ 5 đến 12, một giai đoạn phát triển đặc biệt quan trọng đối với cảm xúc, hành vi và khả năng học tập. Đối tượng nghiên cứu không chỉ giới hạn ở học sinh mà còn mở rộng tới các yếu tố xung quanh như giáo viên, môi trường lớp học, và các công cụ hỗ trợ giảng dạy. Việc lựa chọn nhóm tuổi này xuất phát từ nhu cầu phân tích chuyên sâu về cách cảm xúc ảnh hưởng đến hiệu quả học tập trong giai đoạn hình thành các kỹ năng cơ bản và phát triển trí tuệ xã hội.

Phạm vi dữ liệu:

Phạm vi dữ liệu nghiên cứu bao gồm hình ảnh, video và các thông tin ngữ cảnh thu thập từ các lớp học thực tế. Dữ liệu sẽ được thu thập từ một lớp học tại Đồng Nai, Việt Nam, lớp học bao gồm các học sinh có độ tuổi từ 5-12 tuổi.

Phạm vi kỹ thuật:

Phạm vi kỹ thuật bao gồm việc áp dụng các mô hình học sâu hiện đại như YOLOv8 Classify, Vision Transformer (ViT), và Faster RCNN để phân tích dữ liệu cảm xúc. Ngoài ra, nghiên cứu sẽ triển khai các phương pháp tăng cường dữ liệu, xử lý dữ liệu đa góc từ camera và tích hợp các mô hình vào một hệ thống đánh giá toàn diện. Hệ thống sẽ hoạt động trong môi trường lớp học thực tế và cung cấp các kết quả đánh giá theo thời gian thực, hỗ trợ việc ra quyết định của giáo viên và nhà quản lý giáo dục.

1.5. Nội dung nghiên cứu

Nghiên cứu này cung cấp cái nhìn tổng quan về vai trò của cảm xúc trong môi trường giáo dục, đặc biệt đối với học sinh trong độ tuổi từ 5 đến 12. Các khái niệm cơ bản về nhận diện cảm xúc, tác động của cảm xúc đến hành vi học tập, và các yếu tố cảm xúc ảnh hưởng đến tương tác giữa học sinh và giáo viên sẽ được trình bày. Đồng thời, nghiên cứu phân tích các công trình liên quan, xác định khoảng trống trong tài liệu và lý do cần thiết phải phát triển một hệ thống đánh giá cảm xúc trong lớp học.

Nội dung tiếp theo tập trung vào việc phát triển bộ dữ liệu EmoLearn dành riêng cho việc phân tích cảm xúc học sinh trong môi trường lớp học. Bộ dữ liệu bao gồm hình ảnh, video và các dữ liệu ngữ cảnh thu thập từ các lớp học thực tế. Các kỹ thuật tăng cường dữ liệu như xoay ảnh, thay đổi ánh sáng, thêm nhiễu và chỉnh sửa màu sắc sẽ được áp dụng để làm giàu bộ dữ liệu. Ngoài ra, các phương pháp làm sạch và chuẩn hóa dữ liệu sẽ được thực hiện nhằm đảm bảo chất lượng và tính đại diện cao nhất của dữ liệu.

Nghiên cứu triển khai các mô hình học sâu tiên tiến để nhận diện cảm xúc và phân tích dữ liệu lớp học. Cụ thể, YOLOv8 Classify sẽ được sử dụng cho các tác vụ nhận diện cảm xúc nhanh chóng và chính xác trong thời gian thực. Vision Transformer (ViT) sẽ xử lý các đặc trưng phức tạp từ dữ liệu hình ảnh và video, trong khi Faster RCNN đảm nhận việc nhận diện khuôn mặt và phân loại cảm xúc chi tiết. Các mô hình sẽ được huấn luyện, kiểm thử và so sánh để tối ưu hóa hiệu quả hoạt động trên dữ liệu thực tế.

Hệ thống đánh giá hiệu quả lớp học được thiết kế và triển khai dựa trên các mô hình học sâu. Hệ thống sẽ sử dụng các camera đa góc để thu thập dữ liệu và giảm thiểu các lỗi nhận diện cảm xúc không chính xác. Hệ thống tích hợp các thành phần từ nhận diện khuôn mặt, phân loại cảm xúc, đến đánh giá tương tác giữa học sinh và giáo viên. Kết quả phân tích được trình bày thông qua giao diện trực quan, cung cấp thông tin chi tiết theo thời gian thực để hỗ trợ giáo viên điều chỉnh phương pháp giảng dạy.

Nghiên cứu đề xuất các chính sách và công cụ cụ thể nhằm hỗ trợ giáo viên cải thiện hiệu quả giảng dạy. Dựa trên các kết quả phân tích, các chính sách tập trung vào việc xây dựng môi trường học tập tích cực, nơi cảm xúc của học sinh được quan tâm đúng mức. Đồng thời, các công cụ kỹ thuật số sẽ được giới thiệu để hỗ trợ giáo viên theo dõi và quản lý cảm xúc học sinh một cách hiệu quả hơn.

Phần cuối cùng trình bày các kết quả đạt được từ hệ thống và mô hình được đề xuất, so sánh với các nghiên cứu hiện tại. Các chỉ số như độ chính xác của mô hình, hiệu quả của hệ thống đánh giá, và tác động của các chính sách giáo dục sẽ được phân tích chi tiết. Đồng thời, phần này cũng đề xuất các hướng nghiên cứu tiếp theo, mở rộng phạm vi ứng dụng của hệ thống sang các môi trường giáo dục khác nhau.

1.6. Cấu trúc luận văn

Luận văn được chia thành 6 chương như sau:

Chương 1: Trình bày bối cảnh nghiên cứu, lý do chọn đề tài, mục tiêu nghiên cứu, đối tượng, phạm vi, và ý nghĩa khoa học, thực tiễn của đề tài.

Chương 2: Đánh giá các nghiên cứu liên quan về dữ liệu, hệ thống phân loại cảm xúc, so sánh các phương pháp truyền thống và hiện đại, làm rõ ưu điểm của các hệ thống thông minh và mô hình học sâu.

Chương 3: Giới thiệu các khái niệm và công nghệ liên quan như xử lý hình ảnh, các thuật toán học sâu (YOLO, Faster-RCNN, ViT).

Chương 4: Trình bày quy trình thu thập dữ liệu, xây dựng bộ dữ liệu, áp dụng các phương pháp xử lý hình ảnh và triển khai mô hình học sâu.

Chương 5: Phân tích và đánh giá kết quả của các mô hình học sâu trên bộ dữ liệu, hiệu quả của hệ thống thông minh trong thực tế.

Chương 6: Tóm tắt những đóng góp của nghiên cứu, đánh giá các hạn chế, và đề xuất các hướng phát triển tiếp theo trong lĩnh vực nhận diện cảm xúc và đánh giá hiệu quả của lớp học dự vào cảm xúc.

CHƯƠNG 2. TỔNG QUAN TÌNH HÌNH NGHIÊN CỨU

2.1. Mối liên hệ giữa cảm xúc và hiệu quả của lớp học

Cảm xúc đóng vai trò trung tâm trong việc hình thành không khí và hiệu quả của lớp học, và điều này đã được nhiều nghiên cứu xác nhận. Một nghiên cứu toàn diện về khí hậu cảm xúc trong lớp học đã cho thấy rằng khi mối quan hệ cảm xúc trong lớp tích cực, học sinh không chỉ tăng cường sự gắn kết mà còn đạt được kết quả học tập tốt hơn [14]. Hiệu ứng này được trung gian bởi mức độ tham gia của học sinh và không bị ảnh hưởng bởi các đặc điểm cá nhân của giáo viên hay điều kiện tổ chức khác. Điều này cho thấy, để cải thiện hiệu quả giảng dạy, giáo viên cần quan tâm đến việc xây dựng một môi trường cảm xúc tích cực trong lớp học.

Một nghiên cứu khác tập trung vào các hoạt động trong lớp học và mức độ căng thẳng cảm xúc mà học sinh trải qua. Kết quả cho thấy rằng trả lời miệng trước bảng đen là một nguồn gây căng thẳng lớn, chiếm gần 40% thời gian trong lớp [15]. Những căng thẳng như vậy có thể làm giảm hiệu quả học tập và khả năng tương tác của học sinh, đặc biệt đối với những học sinh nhạy cảm. Nghiên cứu này đề xuất rằng việc áp dụng các chiến lược giảm căng thẳng, chẳng hạn như cung cấp thời gian chuẩn bị hoặc giảm áp lực từ các bài kiểm tra, có thể mang lại lợi ích lớn cho học sinh.

Ngoài ra, mối liên hệ giữa cảm xúc của giáo viên và học sinh cũng được nhấn mạnh qua nhiều nghiên cứu. Ví dụ, nghiên cứu của [16] đã chỉ ra rằng sự hứng thú của học sinh tăng lên đáng kể khi giáo viên thể hiện sự nhiệt tình và cảm xúc tích cực. Hiệu ứng này đặc biệt mạnh mẽ trong các môn học khó hoặc đòi hỏi sự tương tác cao, nơi mà mối quan hệ cảm xúc tích cực giữa giáo viên và học sinh trở thành một yếu tố không thể thiếu để đảm bảo thành công trong học tập.

Một nghiên cứu đáng chú ý khác về cảm xúc trong quản lý lớp học đã tập trung vào chương trình SECURe, nhấn mạnh tầm quan trọng của việc tích hợp các chiến lược quản lý cảm xúc vào các quy trình giảng dạy [7]. Chương trình này cho thấy rằng khi cảm xúc xã hội được quan tâm đúng mức, chất lượng giảng dạy không chỉ được cải

thiện mà còn tạo ra môi trường học tập tích cực hơn cho học sinh. Điều này đặc biệt quan trọng đối với những học sinh có nguy cơ thất bại trong học tập, khi sự hỗ trợ cảm xúc mạnh mẽ từ giáo viên có thể làm giảm xung đột và tăng cường hiệu quả giảng dạy.

2.2. Các nghiên cứu về bộ dữ liệu cảm xúc

Dữ liệu cảm xúc là nền tảng không thể thiếu cho các nghiên cứu và hệ thống phân tích cảm xúc. Bộ dữ liệu AffectNet đã trở thành tiêu chuẩn vàng cho việc nhận diện cảm xúc qua biểu cảm khuôn mặt, với dữ liệu đa dạng và quy mô lớn được thu thập từ các môi trường thực tế [9]. Bộ dữ liệu này không chỉ cung cấp thông tin chi tiết về các trạng thái cảm xúc mà còn hỗ trợ các mô hình học sâu trong việc nhận diện cảm xúc phức tạp, giúp các ứng dụng như theo dõi cảm xúc học sinh trong lớp học trở nên khả thi.

FER-2013 cũng là một bộ dữ liệu quan trọng trong việc phân loại cảm xúc, với tập trung vào biểu cảm khuôn mặt trong các tình huống thực tế [10]. Mặc dù quy mô nhỏ hơn so với AffectNet, FER-2013 đã chứng minh được tính hữu ích trong việc phát triển các hệ thống nhận diện cảm xúc thời gian thực. Điều này đặc biệt quan trọng trong các ứng dụng yêu cầu phân tích nhanh và chính xác, chẳng hạn như giám sát cảm xúc của học sinh để hỗ trợ giảng dạy.

Ngoài ra, các bộ dữ liệu tập trung vào tín hiệu sinh lý như DEAP và SEED đã mở ra những hướng đi mới trong nghiên cứu cảm xúc [11], [17]. DEAP sử dụng tín hiệu EEG để phân tích cảm xúc, trong khi SEED kết hợp EEG với các tín hiệu khác để cung cấp một cái nhìn toàn diện hơn về phản ứng cảm xúc. Tuy nhiên, do tập trung vào người lớn, khả năng áp dụng các bộ dữ liệu này trong môi trường lớp học vẫn còn hạn chế. Các nghiên cứu tiếp theo cần phải phát triển các bộ dữ liệu sinh lý dành riêng cho trẻ em để nâng cao độ chính xác trong việc phân tích cảm xúc.

Đối với trẻ em, các bộ dữ liệu như CAFE và KDEF-PT đã đóng góp đáng kể vào việc nghiên cứu cảm xúc khuôn mặt [18], [19]. Tuy nhiên, quy mô hạn chế và thiếu sự đa dạng văn hóa đã đặt ra những thách thức trong việc ứng dụng vào các lớp

học thực tế. GoEmotions, với hơn 58 nghìn nhận xét Reddit được gắn nhãn, đã cung cấp các phân loại cảm xúc chi tiết và phù hợp với nhiều ngữ cảnh, nhưng vẫn chưa tập trung vào trẻ em [20]. Những hạn chế này chỉ ra rằng cần có thêm các bộ dữ liệu đặc biệt được thiết kế cho trẻ em trong bối cảnh lớp học để cải thiện tính chính xác và ứng dụng của các hệ thống phân tích cảm xúc. Các bộ dữ liệu tương lai cần tập trung vào việc kết hợp hình ảnh, video và tín hiệu sinh lý để cung cấp một cái nhìn toàn diện hơn về cảm xúc của trẻ trong môi trường học tập.

2.3. Các nghiên cứu về hệ thống đánh giá độ hiệu quả của lớp học

Hệ thống đánh giá hiệu quả lớp học dựa trên cảm xúc đã có nhiều tiến bộ nhờ sự kết hợp của trí tuệ nhân tạo và học sâu. Một số nghiên cứu đã phát triển các hệ thống sử dụng camera đa góc để theo dõi cảm xúc và mức độ tham gia của học sinh. Ví dụ, hệ thống Smart Classroom không chỉ đo lường cảm xúc mà còn phân tích mức độ tập trung của học sinh, cung cấp thông tin phản hồi trực tiếp cho giáo viên [21]. Điều này không chỉ cải thiện chất lượng giảng dạy mà còn giúp giáo viên điều chỉnh chiến lược phù hợp với nhu cầu của từng học sinh.

Nghiên cứu của [22] đã áp dụng mô hình VGG16 tinh chỉnh để nhận diện cảm xúc khuôn mặt và đánh giá mức độ tham gia trong các lớp học trực tuyến. Kết quả cho thấy hệ thống không chỉ cung cấp thông tin chính xác mà còn giúp cải thiện trải nghiệm học tập tổng thể. Hơn nữa, hệ thống này có thể được mở rộng sang các lớp học thực tế để cung cấp đánh giá toàn diện hơn. Tuy nhiên, việc triển khai thực tế còn đòi hỏi sự cải tiến về khả năng hoạt động thời gian thực và tích hợp dữ liệu từ nhiều nguồn khác nhau.

Một nghiên cứu khác tập trung vào việc cải tiến thuật toán chú ý để tăng cường khả năng nhận diện cảm xúc trong các bối cảnh lớp học phức tạp [23]. Thuật toán này không chỉ giảm thiểu lỗi trong việc phân loại cảm xúc mà còn cung cấp các gợi ý hữu ích cho giáo viên trong việc điều chỉnh phương pháp giảng dạy. Các nghiên cứu tương tự đã phát triển hệ thống theo dõi cảm xúc thông qua các cảm biến phi tiếp xúc, giúp giảm thiểu sự gián đoạn trong lớp học.

Ngoài ra, các hệ thống như Student Emotion Recognition System (SERS) đã sử dụng chuyển động mắt và đầu để đánh giá sự tham gia của học sinh trong lớp học, cung cấp thông tin thời gian thực về mức độ tập trung và cảm xúc [24]. Điều này mở ra tiềm năng lớn trong việc kết hợp các cảm biến tiên tiến với hệ thống đánh giá lớp học, giúp giáo viên có thể hiểu rõ hơn về trạng thái cảm xúc và sự chú ý của học sinh. Mặc dù đạt được những tiến bộ đáng kể, phần lớn các hệ thống hiện tại vẫn tập trung vào môi trường trực tuyến hoặc người lớn, tạo ra khoảng trống nghiên cứu đối với trẻ em trong lớp học thực tế. Việc phát triển các hệ thống chuyên biệt cho trẻ em sẽ không chỉ cải thiện trải nghiệm học tập mà còn hỗ trợ giáo viên đưa ra các quyết định giảng dạy hiệu quả hơn.

CHƯƠNG 3. CƠ SỞ LÝ THUYẾT

3.1. Trí tuệ nhân tạo (AI) và Thị giác máy tính (CV)

3.1.1. Trí tuệ nhân tạo là gì?

Trí tuệ nhân tạo (Artificial Intelligence - AI) là một lĩnh vực nghiên cứu và ứng dụng trong khoa học máy tính, nhằm xây dựng các hệ thống có khả năng thực hiện các nhiệm vụ mà trước đây chỉ con người mới có thể đảm nhiệm. Những nhiệm vụ này bao gồm học hỏi, suy luận, lập luận logic, giải quyết vấn đề và thậm chí là ra quyết định một cách tự chủ [25]. AI không chỉ dừng lại ở việc mô phỏng hành vi thông minh của con người mà còn tìm cách tối ưu hóa và nâng cao hiệu quả của các hoạt động thông qua việc phân tích và xử lý dữ liệu ở quy mô lớn.

AI được hình thành dựa trên sự kết hợp của nhiều ngành khoa học, bao gồm toán học, thống kê, khoa học máy tính và thần kinh học. Trong thực tế, AI hiện đại chủ yếu dựa vào học máy (Machine Learning), đặc biệt là các thuật toán học sâu (Deep Learning), để nhận diện mẫu, phân tích dữ liệu và cải thiện hiệu năng của các hệ thống. Những hệ thống này không chỉ học hỏi từ dữ liệu mà còn có thể thích nghi với những tình huống mới mà không cần phải được lập trình trước.

Ứng dụng của AI đã trải rộng khắp các lĩnh vực, từ y tế (phân tích hình ảnh y khoa, dự đoán dịch bệnh) đến tài chính (phát hiện gian lận, tối ưu hóa danh mục đầu tư) và giao thông (phát triển xe tự lái, quản lý lưu lượng giao thông). Trong lĩnh vực công nghiệp, AI đã cách mạng hóa sản xuất thông qua tự động hóa, giảm thiểu sai sót và nâng cao năng suất. Ngoài ra, AI còn đóng vai trò quan trọng trong giáo dục, với các hệ thống học tập thông minh giúp cá nhân hóa trải nghiệm học tập cho từng học viên.

Tuy nhiên, sự phát triển của AI cũng đặt ra nhiều thách thức lớn. Về mặt đạo đức, AI đối mặt với các vấn đề liên quan đến quyền riêng tư, tính minh bạch và trách nhiệm đối với các quyết định tự động. Về mặt xã hội, việc AI thay thế con người trong một số ngành nghề có thể gây ra lo ngại về thất nghiệp và sự bất bình đẳng. Cuối cùng, các giới hạn kỹ thuật như độ tin cậy, khả năng giải thích và kiểm soát của AI vẫn đang

là những vấn đề cần được nghiên cứu sâu hơn.

Nhìn chung, trí tuệ nhân tạo không chỉ là công nghệ mang tính đột phá mà còn là một công cụ quan trọng định hình tương lai, thay đổi cách con người tương tác với máy móc và với thế giới xung quanh. Với tiềm năng không giới hạn, AI đang mở ra những cánh cửa mới cho sự sáng tạo, đổi mới và phát triển toàn cầu.

3.1.2. Các phương pháp trong trí tuệ nhân tạo

3.1.2.1. Học máy (Machine Learning -ML)

Học máy (Machine Learning - ML) là một lĩnh vực quan trọng trong trí tuệ nhân tạo (AI), tập trung vào việc phát triển các thuật toán và mô hình cho phép máy tính tự động học hỏi từ dữ liệu mà không cần lập trình chi tiết từng bước. Bằng cách phân tích các mẫu và mối quan hệ trong dữ liệu, ML giúp máy tính cải thiện hiệu suất và khả năng dự đoán trong nhiều nhiệm vụ khác nhau [26].

ML bao gồm ba phương pháp chính:

- Học có giám sát (Supervised Learning): Mô hình được huấn luyện bằng cách sử dụng tập dữ liệu có gán nhãn, trong đó mỗi đầu vào tương ứng với một đầu ra cụ thể. Mục tiêu là học cách ánh xạ từ đầu vào đến đầu ra để có thể dự đoán chính xác trên dữ liệu mới. Ví dụ, dự đoán giá nhà dựa trên diện tích và vị trí địa lý.
- Học không giám sát (Unsupervised Learning): Dữ liệu đầu vào không có nhãn, và mô hình tự động tìm kiếm các cấu trúc hoặc mẫu tiềm ẩn trong dữ liệu. Một ứng dụng phổ biến là phân cụm khách hàng dựa trên hành vi mua sắm để cải thiện chiến lược marketing.
- Học tăng cường (Reinforcement Learning): Mô hình học hỏi thông qua tương tác với môi trường và nhận phản hồi dưới dạng phần thưởng hoặc hình phạt. Phương pháp này được sử dụng rộng rãi trong đào tạo robot hoặc lập trình xe tự lái để tối ưu hóa hành vi qua nhiều lần thử nghiệm.

Ứng dụng của Học máy:

- Y tế: Dự đoán bệnh, phân tích hình ảnh y khoa và tối ưu hóa phác đồ điều trị.
- Tài chính: Phát hiện gian lận, dự đoán thị trường và quản lý rủi ro.
- Thương mại điện tử: Gợi ý sản phẩm dựa trên hành vi người dùng và phân tích xu hướng tiêu dùng.
- Giao thông: Phát triển xe tự lái, dự báo lưu lượng giao thông và tối ưu hóa chuỗi cung ứng.

Nhờ khả năng tự học và thích nghi, học máy đã trở thành công cụ quan trọng để giải quyết các vấn đề phức tạp và thúc đẩy đổi mới trong nhiều lĩnh vực, từ khoa học đến công nghiệp [27].

3.1.2.2. Học sâu (Deep Learning-DL)

Học sâu (Deep Learning - DL) là một nhánh của học máy (Machine Learning - ML), chuyên về việc sử dụng các mạng nơ-ron nhân tạo (Artificial Neural Networks - ANN) với nhiều tầng để học và biểu diễn dữ liệu phức tạp [28]. Khác với các thuật toán học máy truyền thống, DL có khả năng tự động trích xuất các đặc trưng từ dữ liệu, nâng cao độ chính xác trong các tác vụ như nhận dạng hình ảnh, xử lý ngôn ngữ tự nhiên và dự đoán chuỗi thời gian [29].

DL nổi bật nhờ các mô hình mạng nơ-ron sâu với nhiều tầng, nơi các tầng thấp hơn xử lý các đặc trưng cơ bản và các tầng cao hơn trừu tượng hóa các mẫu phức tạp hơn . Phương pháp này hoạt động hiệu quả trong cả môi trường học có giám sát và không giám sát, cho phép áp dụng trong nhiều lĩnh vực khác nhau. Tuy nhiên, DL thường yêu cầu một lượng dữ liệu lớn để huấn luyện mô hình và sử dụng các tập dữ liệu quy mô lớn như ImageNet [30]. Đặc biệt, DL vượt trội trong xử lý dữ liệu phi cấu trúc như hình ảnh, âm thanh và văn bản, điều mà các thuật toán học máy truyền thống thường khó xử lý .

Ứng dụng của DL đã mở rộng đáng kể, bao gồm thị giác máy tính với các tác vụ

như nhận diện khuôn mặt và phân loại hình ảnh, xử lý ngôn ngữ tự nhiên (NLP) trong chatbot hoặc dịch máy, y tế với phân tích hình ảnh chẩn đoán và thương mại điện tử với hệ thống gợi ý cá nhân hóa . Tuy nhiên, DL cũng gặp phải các thách thức lớn như chi phí tính toán cao, khó giải thích mô hình ("hộp đen"), và các rủi ro đạo đức liên quan đến quyền riêng tư

3.1.2.3. Xử lý ngôn ngữ tự nhiên (Natural Language Processing - NLP)

Xử lý ngôn ngữ tự nhiên (Natural Language Processing - NLP) là một lĩnh vực quan trọng trong trí tuệ nhân tạo (AI), tập trung vào việc phát triển các hệ thống giúp máy tính hiểu, phân tích và tạo ra ngôn ngữ tự nhiên của con người, bao gồm cả văn bản và lời nói. NLP kết hợp các kỹ thuật từ ngôn ngữ học, học máy và khoa học máy tính để xây dựng các mô hình có khả năng xử lý ngôn ngữ phức tạp một cách hiệu quả. Một trong những bước quan trọng đầu tiên trong NLP là tiền xử lý văn bản, nơi các kỹ thuật như phân tách từ (tokenization), chuẩn hóa văn bản và loại bỏ từ không cần thiết được áp dụng để chuẩn bị dữ liệu cho các bước phân tích tiếp theo [31].

Gần đây, với sự phát triển của các mô hình học sâu như Transformer, BERT và GPT, NLP đã đạt được những tiến bộ đáng kể. Những mô hình này không chỉ cải thiện độ chính xác trong các tác vụ như dịch ngôn ngữ và phân tích cảm xúc, mà còn giúp nâng cao khả năng hiểu ngữ cảnh của máy tính. Các ứng dụng như Google Dịch và trợ lý ảo Siri đã tận dụng những mô hình này để nâng cao hiệu quả và trải nghiệm người dùng [31]. Tuy nhiên, NLP vẫn đối mặt với những thách thức lớn, bao gồm việc xử lý các ngữ cảnh phức tạp, giảm thiểu tính thiên vị trong dữ liệu huấn luyện và đáp ứng sự đa dạng ngôn ngữ trên toàn cầu.

3.1.2.4. Các phương pháp khác

Ngoài các phương pháp chính như học máy (Machine Learning - ML), học sâu (Deep Learning - DL) và xử lý ngôn ngữ tự nhiên (Natural Language Processing - NLP), trí tuệ nhân tạo (AI) còn bao gồm nhiều phương pháp khác nhằm giải quyết các bài toán phức tạp và đa dạng trong thực tiễn. Một trong những phương pháp đáng chú ý là logic mờ (Fuzzy Logic), được sử dụng để xử lý các dữ liệu không chắc chắn và

không rõ ràng. Logic mờ cho phép máy tính mô phỏng cách con người đưa ra quyết định trong các môi trường không chắc chắn, từ đó được ứng dụng trong các hệ thống điều khiển và tối ưu hóa tự động .

Bên cạnh đó, hệ thống chuyên gia (Expert Systems) là một phương pháp quan trọng mô phỏng khả năng suy luận của con người để giải quyết các vấn đề trong các lĩnh vực cụ thể như y tế, tài chính và công nghiệp. Những hệ thống này dựa trên cơ sở tri thức được lập trình sẵn để đưa ra các quyết định và lời khuyên hiệu quả [32]. Ngoài ra, tính toán tiến hóa (Evolutionary Computation), bao gồm các thuật toán như thuật toán di truyền (Genetic Algorithms), được sử dụng để giải quyết các bài toán tìm kiếm và tối ưu hóa trong không gian tìm kiếm lớn, thường áp dụng trong các lĩnh vực như thiết kế kỹ thuật và công nghiệp [33].

Một phương pháp khác là mạng Bayes (Bayesian Networks), được sử dụng rộng rãi trong việc xử lý thông tin không chắc chắn và suy diễn xác suất. Mạng Bayes giúp xây dựng các hệ thống dự đoán và phát hiện gian lận, đặc biệt hiệu quả trong các bài toán yêu cầu phân tích dữ liệu lớn với độ tin cậy cao [34]. Những phương pháp này không chỉ mở rộng phạm vi ứng dụng của AI mà còn giúp tối ưu hóa hiệu suất trong các lĩnh vực từ công nghiệp đến nghiên cứu khoa học.

3.2. Tổng quan về các mô hình

3.2.1. Tổng quan về YOLO và YOLOv8

YOLO (You Only Look Once) là một thuật toán phát hiện vật thể (object detection) được phát triển bởi Joseph Redmon và cộng sự, lần đầu tiên được giới thiệu vào năm 2016. YOLO nổi bật với khả năng phát hiện vật thể trong thời gian thực bằng cách xử lý toàn bộ hình ảnh chỉ trong một bước duy nhất. Khác với các phương pháp trước đó như R-CNN, YOLO không cần phải chia hình ảnh thành nhiều vùng nhỏ để phân tích, nhờ đó đạt được tốc độ xử lý vượt trội mà vẫn đảm bảo độ chính xác [35].

YOLOv8 là phiên bản mới nhất trong họ thuật toán YOLO, được phát triển nhằm cải thiện cả về tốc độ và độ chính xác. Phiên bản này sử dụng kiến trúc mạng nơ-

ron hiện đại hơn với các cải tiến như Backbone mạnh mẽ, Head tiên tiến hơn, và các phương pháp tối ưu hóa huấn luyện. YOLOv8 không chỉ hoạt động hiệu quả trên các bộ dữ liệu lớn mà còn được tối ưu hóa để triển khai trên các thiết bị hạn chế tài nguyên như di động hoặc IoT. Bên cạnh đó, các chiến lược như augmentation dữ liệu và loss function mới đã giúp YOLOv8 nâng cao khả năng tổng quát hóa, đảm bảo độ chính xác trong các tác vụ phát hiện vật thể đa dạng [36].

Sự phát triển của YOLO và YOLOv8 đã mở rộng phạm vi ứng dụng trong nhiều lĩnh vực, từ xe tự lái, giám sát an ninh, đến y tế và thương mại điện tử. Khả năng phát hiện vật thể nhanh chóng và hiệu quả đã biến YOLO trở thành một công cụ tiêu chuẩn trong các hệ thống thị giác máy tính hiện đại.

3.2.2. Tổng quan về Vision Transfromer (ViT)

Vision Transformer (ViT) là một bước đột phá trong lĩnh vực thị giác máy tính, được giới thiệu bởi Dosovitskiy và cộng sự vào năm 2020. Kiến trúc ViT áp dụng cơ chế Transformer, vốn đã được chứng minh hiệu quả trong xử lý ngôn ngữ tự nhiên, để giải quyết các tác vụ thị giác như phân loại hình ảnh. Không giống như các mô hình dựa trên mạng nơ-ron tích chập (CNN), vốn phụ thuộc vào việc học các đặc trưng cục bộ từ hình ảnh, ViT xử lý hình ảnh bằng cách chia chúng thành các "patch" nhỏ có kích thước cố định (thường là 16x16 pixel), sau đó sử dụng các lớp Attention để nắm bắt mối quan hệ toàn cục giữa các vùng trong hình ảnh [37].

Kiến trúc ViT bao gồm ba thành phần chính:

- Patch Embedding: Hình ảnh được chia thành các patch, sau đó mỗi patch được ánh xạ thành một vector nhờ một lớp embedding. Điều này tương tự như việc mã hóa các token trong Transformer.
- Position Embedding: ViT sử dụng thông tin vị trí của các patch thông qua position embedding để bảo toàn thứ tự không gian, vì Transformer ban đầu được thiết kế cho dữ liệu tuần tự, không phải hình ảnh.
- Transformer Encoder: Các lớp Transformer encoder xử lý các patch

embedding và học các mối quan hệ giữa chúng, từ đó trích xuất các đặc trưng toàn cục.

ViT nổi bật nhờ khả năng học các đặc trưng toàn cục mạnh mẽ, điều mà CNN thường bị hạn chế do chỉ tập trung vào các đặc trưng cục bộ. Các nghiên cứu chỉ ra rằng ViT vượt trội hơn CNN khi làm việc với các tập dữ liệu lớn, chẳng hạn như ImageNet-21k và JFT-300M. Tuy nhiên, một hạn chế của ViT là nó hoạt động kém hiệu quả trên các tập dữ liệu nhỏ, vì cơ chế Attention yêu cầu một lượng lớn dữ liệu để huấn luyện. Để khắc phục, các biến thể như DeiT (Data-efficient Image Transformer) đã được phát triển nhằm giảm yêu cầu về dữ liệu, giúp ViT hoạt động tốt hơn trên các tập dữ liệu quy mô vừa và nhỏ [38].

Ứng dụng của ViT không chỉ dừng lại ở phân loại hình ảnh mà còn mở rộng sang các lĩnh vực như phát hiện vật thể, phân đoạn hình ảnh, và thậm chí cả phân tích y tế. Trong y học, ViT đã được sử dụng để phát hiện các bệnh từ hình ảnh chụp CT và MRI, với độ chính xác cao hơn so với các mô hình truyền thống. Trong giao thông, ViT đóng vai trò quan trọng trong các hệ thống xe tự lái, nơi việc nắm bắt toàn cục trong hình ảnh là yếu tố sống còn để phát hiện vật cản và điều hướng.

Việc tích hợp ViT vào thị giác máy tính đánh dấu một giai đoạn chuyển đổi quan trọng, từ việc phụ thuộc vào CNN sang khai thác sức mạnh của Attention. Mặc dù vẫn còn nhiều thách thức như chi phí tính toán cao và yêu cầu dữ liệu lớn, ViT đã chứng minh tiềm năng vượt trội, mở ra một hướng đi mới trong nghiên cứu và ứng dụng thị giác máy tính hiện đại.

3.2.3. Tổng quan về ResNet50

ResNet50, viết tắt của Residual Network 50, là một trong những kiến trúc mạng học sâu được sử dụng phổ biến nhất trong lĩnh vực thị giác máy tính. Đây là một phần mở rộng của mạng Residual Network do Kaiming He và các cộng sự giới thiệu vào năm 2015 [39], nhằm giải quyết vấn đề gradient biến mất (vanishing gradient) thường gặp ở các mạng rất sâu. Với 50 lớp, ResNet50 được thiết kế để học các đặc trưng phức

tạp và chi tiết từ dữ liệu lớn, đồng thời duy trì hiệu quả và độ chính xác cao trong quá trình huấn luyện.

Cấu trúc của ResNet50 đặc trưng bởi các khối dư (residual block), một sáng kiến đột phá giúp mạng học ánh xạ dư (residual mapping) thay vì ánh xạ trực tiếp giữa đầu vào và đầu ra. Mỗi khối dư bao gồm các lớp tích chập với bộ lọc nhỏ, được nối trực tiếp với đầu vào thông qua các đường dẫn tắt (shortcut connection). Thiết kế này cho phép thông tin và gradient truyền qua toàn bộ mạng một cách dễ dàng, giảm thiểu tình trạng gradient suy giảm trong các mạng sâu. Các khối dư trong ResNet50 được tổ chức thành nhiều nhóm, với số lượng bộ lọc tăng dần để học các đặc trưng từ cơ bản đến phức tạp.

ResNet50 mang lại nhiều ưu điểm vượt trội. Thứ nhất, nó có khả năng xử lý các mạng rất sâu mà không làm giảm hiệu suất, nhờ vào khả năng duy trì thông tin qua các khối dư. Thứ hai, nhờ vào cấu trúc linh hoạt, ResNet50 có thể áp dụng cho nhiều bài toán khác nhau trong thị giác máy tính, từ nhận diện khuôn mặt, phân loại cảm xúc, đến phát hiện đối tượng. Thứ ba, mô hình này đã đạt hiệu suất xuất sắc trên các tập dữ liệu chuẩn như ImageNet, trở thành một trong những kiến trúc phổ biến nhất cho các tác vụ nhận diện và phân loại.

Ứng dụng của ResNet50 rất đa dạng và phong phú. Trong lĩnh vực nhận diện khuôn mặt, ResNet50 thường được sử dụng như một phần của các hệ thống lớn, hỗ trợ xác định danh tính và phân loại cảm xúc. Trong phân loại đối tượng, ResNet50 có thể học và phân biệt hàng ngàn danh mục hình ảnh với độ chính xác cao. Ngoài ra, kiến trúc này thường được sử dụng làm nền tảng (backbone) trong các hệ thống phát hiện đối tượng như Faster R-CNN, phục vụ cho các bài toán như phát hiện vị trí và loại đối tượng trong ảnh. Không chỉ dừng lại ở lĩnh vực thị giác máy tính, ResNet50 còn được ứng dụng trong y học, hỗ trợ phân tích hình ảnh X-quang, MRI để phát hiện các bất thường.

Với thiết kế hiện đại, khả năng mở rộng, và hiệu năng vượt trội, ResNet50 là

một lựa chọn lý tưởng cho các bài toán học sâu phức tạp. Đây không chỉ là một công cụ mạnh mẽ cho các nhà nghiên cứu mà còn là nền tảng quan trọng cho các ứng dụng công nghiệp, từ nhận diện cảm xúc trong lớp học đến phát triển các hệ thống thông minh trong nhiều lĩnh vực khác nhau.

3.3. Các phương pháp xử lý và tăng cường hình ảnh

3.3.1. Các phương pháp xử lý ảnh

Trong thị giác máy tính, xử lý và tăng cường hình ảnh là những bước quan trọng giúp cải thiện chất lượng dữ liệu đầu vào và tăng hiệu quả của các mô hình học sâu. Quá trình xử lý hình ảnh bao gồm các kỹ thuật như giảm nhiễu, cân bằng sáng và tương phản, hoặc chuyển đổi không gian màu. Ví dụ, phương pháp giảm nhiễu bằng Gaussian Blur hoặc Median Filter được sử dụng để loại bỏ nhiễu mà không làm mất chi tiết quan trọng. Ngoài ra, cân bằng sáng và tương phản thông qua các thuật toán như histogram equalization giúp cải thiện khả năng nhận diện đặc trưng, đặc biệt trong các hình ảnh chụp ở điều kiện ánh sáng kém.

Tăng cường hình ảnh, mặt khác, tập trung vào việc tạo ra các biến thể của hình ảnh đầu vào để mở rộng tập dữ liệu huấn luyện mà không cần thu thập thêm dữ liệu thực tế. Các kỹ thuật phổ biến bao gồm xoay, phóng to/thu nhỏ, lật hình ảnh, hoặc thêm nhiễu để tăng khả năng chịu đựng của mô hình trước các biến đổi trong thế giới thực. Chẳng hạn, biến đổi màu sắc như thay đổi độ sáng hoặc độ tương phản không chỉ làm đa dạng dữ liệu mà còn cải thiện độ chính xác của mô hình khi xử lý dữ liệu từ các nguồn khác nhau [40]. Những phương pháp này đặc biệt quan trọng trong các ứng dụng như y tế, nơi chất lượng hình ảnh từ các thiết bị chụp X-quang hoặc MRI cần được nâng cao trước khi đưa vào phân tích.

Nhờ áp dụng các kỹ thuật xử lý và tăng cường hình ảnh, các hệ thống thị giác máy tính không chỉ cải thiện hiệu suất mà còn đạt được tính tổng quát cao hơn, hỗ trợ nhiều lĩnh vực từ giao thông, y tế đến thương mại điện tử.

3.3.2. Các phương pháp tăng cường ảnh

Tăng cường ảnh (Image Augmentation) là một kỹ thuật quan trọng trong học sâu, giúp mở rộng tập dữ liệu huấn luyện bằng cách tạo ra các biến thể mới từ hình ảnh ban đầu. Phương pháp này không chỉ làm tăng kích thước tập dữ liệu mà còn cải thiện khả năng tổng quát hóa của các mô hình, từ đó giảm thiểu tình trạng overfitting. Tăng cường ảnh được sử dụng phổ biến trong các lĩnh vực như nhận diện khuôn mặt, phát hiện vật thể, và phân loại hình ảnh.

Một trong những kỹ thuật tăng cường cơ bản là biến đổi hình học (Geometric Transformations), bao gồm các thao tác như xoay (rotation), phóng to hoặc thu nhỏ (scaling), lật ngang hoặc dọc (flipping). Các thao tác này giúp mô hình trở nên linh hoạt hơn trước các biến đổi về góc nhìn hoặc kích thước của hình ảnh trong thực tế [40]. Ngoài ra, thao tác cắt và ghép (Cropping and Padding) cũng là một phương pháp quan trọng, đặc biệt trong các bài toán như phát hiện vật thể, giúp mô hình học được các vùng khác nhau trong một hình ảnh.

Một nhóm phương pháp khác là biến đổi màu sắc (Color Transformations), như thay đổi độ sáng (brightness adjustment), độ tương phản (contrast adjustment), hoặc áp dụng hiệu ứng jitter màu sắc (color jittering). Những kỹ thuật này làm tăng khả năng của mô hình trong việc xử lý các hình ảnh được chụp ở các điều kiện ánh sáng và môi trường khác nhau. Ngoài ra, thêm nhiễu (Adding Noise) cũng được áp dụng để làm cho hình ảnh giống với các điều kiện thực tế hơn, giúp mô hình chịu đựng tốt hơn với dữ liệu có chất lượng thấp.

Các phương pháp tăng cường ảnh tiên tiến hơn, chẳng hạn như CutMix và MixUp, tạo ra các hình ảnh tổng hợp bằng cách kết hợp các hình ảnh gốc. Điều này không chỉ tăng sự đa dạng của tập dữ liệu mà còn giúp mô hình học được các đặc trưng quan trọng từ nhiều nguồn khác nhau cùng một lúc [41].

Tăng cường ảnh không chỉ cải thiện hiệu suất của các mô hình học sâu mà còn đóng vai trò thiết yếu trong các bài toán thực tế, đặc biệt khi dữ liệu đầu vào bị giới hạn về kích thước hoặc tính đa dạng. Đây là một bước quan trọng trong quy trình xử lý dữ liệu hình ảnh để đảm bảo các hệ thống thị giác máy tính hoạt động hiệu quả và đáng tin

cậy.

CHƯƠNG 4. PHƯƠNG PHÁP NGHIÊN CỨU

4.1. Phương pháp thu thập bộ dữ liệu

4.1.1. Môi trường thu thập dữ liệu

Dữ liệu được thu thập trong hai bố trí lớp học khác nhau nhằm đảm bảo sự đa dạng trong việc ghi nhận cảm xúc và tương tác của học sinh. Bố trí đầu tiên được gọi là "Bố trí bàn dọc", nơi bàn ghế được sắp xếp theo hình chữ U. Học sinh ngồi dọc theo ba cạnh của hình chữ U, tạo điều kiện thuận lợi cho việc giao tiếp và tương tác giữa các học sinh. Camera được đặt tại điểm mở của hình chữ U, mang lại góc nhìn toàn cảnh lớp học, giúp ghi nhận biểu cảm của đa số học sinh một cách rõ ràng và đồng đều.

Bố trí thứ hai là "Bố trí hướng về phía trước", trong đó tất cả bàn ghế và học sinh đều hướng về phía bảng. Cách bố trí này thường thấy trong các lớp học truyền thống, nơi học sinh tập trung vào giảng dạy trực tiếp từ giáo viên. Camera trong trường hợp này được đặt ở phía sau lớp học, cho phép thu thập dữ liệu biểu cảm của học sinh khi họ tham gia bài giảng hoặc hoạt động học tập. Cả hai bố trí này đều được lựa chọn nhằm phản ánh các tình huống lớp học thực tế, tạo điều kiện thuận lợi cho việc thu thập dữ liệu đa dạng và toàn diện. Hai bố trí lớp học được thể hiện ở Hình 4.1.

Hình 4.1. Hai bố trí lớp học thu dữ liệu.

4.1.2. Thiết bị thu thập dữ liệu

Thiết bị được sử dụng để thu thập dữ liệu là camera DJI Action 5 (Hình 4.2), một thiết bị ghi hình tiên tiến được thiết kế để ghi lại video chất lượng cao ngay cả

trong các điều kiện ánh sáng và không gian không đồng đều. DJI Action 5 được lựa chọn do tính linh hoạt, độ bền và khả năng chống rung, phù hợp để sử dụng trong môi trường lớp học năng động. Với khả năng ghi hình liên tục trong thời gian dài và chất lượng hình ảnh ổn định, thiết bị này đảm bảo rằng dữ liệu thu thập có độ rõ nét cao, hỗ trợ tốt cho các bước xử lý tiếp theo.

Hình 4.2. Thiết bị DJI Action 5.

4.1.3. Cách thức lắp đặt thiết bị thu thập dữ liệu

Camera được lắp đặt tại các vị trí chiến lược trong lớp học để đảm bảo ghi lại được đầy đủ các biểu cảm của học sinh. Cả hai bố trí lớp học đều có cùng chiều cao lắp đặt camera, được điều chỉnh ở mức 1,8 mét so với mặt sàn. Mức chiều cao này được lựa chọn để tránh bị che khuất bởi các vật dụng hoặc học sinh, đồng thời tạo góc nhìn tối ưu, giúp camera ghi lại khuôn mặt và biểu cảm của học sinh một cách rõ ràng nhất.

Trong "Bố trí bàn thẳng," camera được đặt tại điểm mở của hình chữ U, hướng thẳng vào lớp học, bao quát toàn bộ khu vực nơi học sinh ngồi. Cách lắp đặt này cho phép ghi lại biểu cảm của học sinh từ nhiều góc độ khác nhau. Đối với "Bố trí hướng về phía trước," camera được đặt ở phía trước lớp học và hướng về cuối lớp, đảm bảo

góc nhìn đồng nhất với tất cả học sinh, giúp phân tích biểu cảm và tương tác trực tiếp với bài giảng.

4.1.4. Quy trình thu thập dữ liệu

Dữ liệu được thu thập trong mỗi bố trí lớp học với tổng thời lượng là 45 phút cho mỗi bố trí. Video được ghi lại ở độ phân giải 4K và tốc độ 60 khung hình mỗi giây để đảm bảo chất lượng hình ảnh cao nhất. Thời gian ghi hình được chọn nhằm đảm bảo rằng dữ liệu thu thập đủ phản ánh các hoạt động học tập và tương tác trong lớp học.

Trong quá trình thu thập, các hoạt động học tập bao gồm cả bài giảng và các hoạt động thảo luận nhóm đã được tổ chức để đảm bảo tính đa dạng và đại diện của dữ liệu. Các camera được kiểm tra và hiệu chỉnh trước mỗi buổi thu thập để đảm bảo rằng tất cả thông tin cần thiết được ghi lại đầy đủ và chính xác. Sự đồng bộ giữa các bố trí lớp học cũng được thực hiện để đảm bảo dữ liệu từ cả hai cấu hình đều tương thích cho các bước phân tích tiếp theo.

4.2. Phương pháp tiền xử lý dữ liệu

Sau khi hoàn tất quá trình thu thập, dữ liệu video được đưa vào giai đoạn tiền xử lý nhằm chuẩn bị cho các bước phân tích tiếp theo. Giai đoạn tiền xử lý bao gồm hai bước chính: cắt khung hình và chọn lọc dữ liệu.

Đầu tiên, toàn bộ video được phân tách thành các khung hình với thời gian mỗi khung là 1 giây. Các khung hình này được phân tích bằng mô hình YOLOv8-Face để phát hiện và trích xuất khuôn mặt học sinh. Sau khi phát hiện, các khuôn mặt được chuẩn hóa về kích thước 160x160 pixel nhằm đảm bảo sự đồng nhất trong dữ liệu đầu vào. Quy trình này được minh họa rõ hơn trong Hình 4.3, mô tả cách các khung hình được cắt, trích xuất khuôn mặt và chuẩn hóa.

Tiếp theo, dữ liệu được chọn lọc để loại bỏ những khung hình không đạt yêu cầu. Các khung hình có thể bị loại bỏ nếu khuôn mặt bị che khuất, bị nhòe do chuyển động hoặc không đủ ánh sáng. Việc chọn lọc này đảm bảo rằng chỉ những dữ liệu chất

lượng cao được giữ lại để phân tích, giảm thiểu sai sót và tăng độ chính xác cho các mô hình nhận diện cảm xúc.

Hình 4.3. Quy trình xử lý dữ liệu.

Quy trình tiền xử lý này không chỉ cải thiện chất lượng dữ liệu mà còn tối ưu hóa hiệu suất của các bước phân tích tiếp theo, đảm bảo rằng mô hình học sâu có thể hoạt động tốt trên dữ liệu thực tế từ môi trường lớp học.

4.3. Gán nhãn hình ảnh

Quá trình gán nhãn hình ảnh đóng vai trò quan trọng trong việc xây dựng bộ dữ liệu chất lượng cao để huấn luyện và triển khai hệ thống nhận diện cảm xúc. Sau khi các khuôn mặt được trích xuất từ khung hình, mỗi hình ảnh được gắn nhãn cảm xúc dựa trên trạng thái biểu cảm của học sinh. Hệ thống phân loại cảm xúc chia các trạng thái này thành các nhóm tích cực, tiêu cực và trung tính. Mỗi hình ảnh được kiểm tra kỹ lưỡng để đảm bảo tính chính xác trong việc gán nhãn.

Các nhãn cảm xúc được định nghĩa như sau:

- Exciting (Tích cực): Đại diện cho trạng thái cảm xúc sôi nổi, vui mừng hoặc đầy hứng thú.
- Happy (Tích cực): Thể hiện trạng thái hạnh phúc, thoải mái và hài lòng.
- Sad (Tiêu cực): Biểu hiện cảm xúc buồn bã, thất vọng hoặc thiếu năng lượng.
- Neutral (Trung tính): Đại diện cho trạng thái cảm xúc trung lập, không thể hiện sự tích cực hay tiêu cực rõ rệt.

- Unknown (Trung tính): Sử dụng cho các hình ảnh mà cảm xúc không rõ ràng hoặc không thể phân loại.

Việc gán nhãn này không chỉ giúp chuẩn hóa bộ dữ liệu mà còn đảm bảo rằng các mô hình học sâu có thể học tập từ các biểu cảm được định nghĩa rõ ràng.

4.4. Huấn luyện mô hình

4.4.1. Môi trường huấn luyện

Quá trình huấn luyện mô hình được thực hiện trên nền tảng Google Colab. Đây là một môi trường điện toán đám mây mạnh mẽ với GPU miễn phí, giúp tăng tốc quá trình huấn luyện các mô hình học sâu. Colab tích hợp sẵn các thư viện phổ biến như TensorFlow, PyTorch và OpenCV, tạo điều kiện thuận lợi cho việc triển khai và kiểm thử mô hình. Hệ thống này đáp ứng yêu cầu xử lý dữ liệu lớn và đảm bảo hiệu quả cao cho các tác vụ như nhận diện khuôn mặt, phân loại cảm xúc và nhận dạng danh tính.

4.4.2. Các thông số huấn luyện

Quá trình huấn luyện các mô hình được thực hiện trong 100 epochs. Các thông số cơ bản như learning rate, batch size, và optimizer được cấu hình một cách cẩn thận để đảm bảo rằng mô hình học được một cách hiệu quả từ dữ liệu. Việc sử dụng Google Colab Pro với GPU Tesla V100-SXM2 giúp tăng tốc quá trình huấn luyện và đảm bảo rằng mô hình có thể xử lý khối lượng dữ liệu lớn một cách nhanh chóng và hiệu quả.

Component	Specification
CPU	Intel(R) Xeon(R) 2.30GHz
GPU	Tesla V100-SXM2
GPU memory size	16 GB
Memory	13 GB
Disk	167 GB
Deep learning architecture	Ultralytics YOLOv8.0.200 + Python-3.10 + torch-2.1.0+cu118

Bảng 4.1. Thông số môi trường huấn luyện.

Kết quả từ quá trình huấn luyện cho thấy mô hình đạt được độ chính xác và khả năng phát hiện đối tượng cao, với các chỉ số như mAP (mean Average Precision), độ chính xác (precision), và độ nhạy (recall) đều cho thấy sự cải thiện đáng kể qua từng epoch. Điều này chứng tỏ rằng việc xử lý và tăng cường dữ liệu đã giúp mô hình học được từ một tập hợp dữ liệu phong phú và đa dạng, từ đó cải thiện khả năng tổng quát hóa và độ chính xác trong việc nhận diện trang thiết bị bảo hộ lao động.

4.5. Triển khai hệ thống

Việc triển khai hệ thống phân tích hiệu quả lớp học là một quy trình phức tạp, được thiết kế với mục tiêu tối ưu hóa hiệu quả giảng dạy và quản lý lớp học thông qua việc ứng dụng các công nghệ học sâu hiện đại. Các bước trong quy trình triển khai không chỉ đảm bảo độ chính xác của các phân tích mà còn cung cấp thông tin chi tiết, hỗ trợ giáo viên trong việc ra quyết định. Quy trình này được chia thành bốn phần chính: đánh giá hiệu năng các mô hình, xây dựng chính sách đánh giá, thiết kế cấu trúc tổng thể của hệ thống, và xây dựng quy trình hoạt động chi tiết.

4.5.1. Đánh giá hiệu năng của các mô hình

Trước khi tích hợp vào hệ thống, các mô hình nhận diện danh tính và phân loại cảm xúc được đánh giá kỹ lưỡng dựa trên các tiêu chí về độ chính xác và tốc độ xử lý. Đối với phát hiện khuôn mặt, mô hình YOLOv8-Face đã được chọn trực tiếp nhờ hiệu suất vượt trội và khả năng đáp ứng yêu cầu thực tế.

Nhận diện danh tính:

Để nhận diện danh tính học sinh, các mô hình ArcFace, VGGFace, FaceNet và FaceNet512 được đưa vào thử nghiệm. Các tiêu chí đánh giá bao gồm:

- FPS (Frame per Second): Đo tốc độ xử lý khung hình trong môi trường thực tế.
- Độ chính xác (Accuracy): Xác định tỷ lệ nhận diện đúng danh tính học sinh.

- Khả năng hoạt động ổn định: Đánh giá hiệu năng của mô hình trong điều kiện ánh sáng yếu và góc nhìn phức tạp.

Sau khi phân tích, mô hình ArcFace được ưu tiên sử dụng nhờ độ chính xác cao trong khi FaceNet512 đảm bảo tốc độ xử lý nhanh chóng. Hệ thống được thiết kế để linh hoạt sử dụng một trong hai mô hình này tùy theo yêu cầu cụ thể.

Phân loại cảm xúc:

Để phân loại cảm xúc, hệ thống sử dụng các mô hình học sâu đã được huấn luyện trên bộ dữ liệu EmoLearn. Các trạng thái cảm xúc được gán nhãn gồm: Exciting, Happy, Sad, Neutral, và Unknown.

Các mô hình được đánh giá dựa trên khả năng phân biệt cảm xúc trong các điều kiện biểu cảm khác nhau. Các tiêu chí chính gồm:

- Độ chính xác: Tỷ lệ dự đoán đúng cảm xúc so với nhãn thực tế.
- Tốc độ xử lý: Thời gian trung bình cần để phân loại cảm xúc cho một khuôn mặt.

Mô hình có hiệu năng tốt nhất được chọn để tích hợp vào hệ thống, đảm bảo khả năng hoạt động ổn định và chính xác trong thời gian thực.

4.5.2. Chính sách đánh giá mức độ hiệu quả

Chính sách đánh giá hiệu quả lớp học là nền tảng để phân tích và đưa ra các nhận định về môi trường học tập. Chính sách này được xây dựng với các tiêu chí cụ thể để đảm bảo tính toàn diện và minh bạch trong đánh giá.

Tỷ lệ cảm xúc tích cực: Một lớp học được coi là hiệu quả nếu ít nhất 70% cảm xúc được phân loại là tích cực (Exciting và Happy).

Tỷ lệ cảm xúc trung tính và không rõ ràng: Cảm xúc trung tính (Neutral) và không rõ ràng (Unknown) không được vượt quá 20% tổng số cảm xúc ghi nhận.

Tỷ lệ cảm xúc tiêu cực: Cảm xúc tiêu cực (Sad) cần được giữ ở mức tối thiểu, không vượt quá 10%.

Theo dõi học sinh đặc biệt: Các học sinh có cảm xúc tiêu cực hoặc biểu hiện bất thường sẽ được ghi nhận lại trong hệ thống để giáo viên theo dõi và can thiệp kịp thời.

4.5.3. Cấu trúc tổng thể hệ thống

Hệ thống được thiết kế với các thành phần chính sau đây để đảm bảo hoạt động hiệu quả và tự động:

Camera lớp học:

- Camera được lắp đặt ở 4 vị trí xung quanh lớp học để ghi lại toàn cảnh lớp học trong cả hai bố trí: "Bố trí bàn dọc" và "Bố trí hướng về phía trước".
- Camera ghi hình với độ phân giải 4K và tốc độ 60 khung hình/giây, đảm bảo dữ liệu đầu vào có chất lượng cao.

Máy chủ xử lý:

- Máy chủ chạy các mô hình YOLOv8-Face, ArcFace hoặc FaceNet512, và mô hình phân loại cảm xúc.
- Đảm nhận việc phân tích dữ liệu và tổng hợp thông tin để đánh giá hiệu quả lớp học.

Giao diện người dùng: Giao diện trực quan hiển thị các kết quả phân tích, bao gồm bảng điều khiển, danh sách học sinh cần theo dõi, và các báo cáo chi tiết về hiệu quả lớp học.

4.5.4. Quy trình hoạt động của hệ thống

Hệ thống hoạt động theo quy trình chi tiết, với các bước được sắp xếp tuần tự để đảm bảo tính chính xác và hiệu quả trong phân tích dữ liệu:

Bước 1: Phát hiện khuôn mặt

Hệ thống bắt đầu bằng việc phát hiện khuôn mặt của học sinh trong các khung hình video. Bốn camera được lắp đặt ở bốn góc của lớp học nhằm đảm bảo quan sát toàn bộ khu vực, hạn chế tình trạng khuôn mặt học sinh bị che khuất do góc quay hoặc chuyển động. Mỗi camera ghi lại video với độ phân giải 4K và tốc độ 60 khung hình/giây, cung cấp dữ liệu chất lượng cao cho quá trình phân tích. Mô hình YOLOv8-Face được sử dụng để phát hiện và định vị khuôn mặt một cách nhanh chóng và chính xác.

Bước 2: Cắt và chuẩn hóa khuôn mặt

Các khuôn mặt được trích xuất từ khung hình video và chuẩn hóa về kích thước 160x160 pixel để đảm bảo đồng nhất đầu vào cho các mô hình phân tích tiếp theo. Nếu một khuôn mặt xuất hiện đồng thời trên nhiều camera, hệ thống sẽ chọn khuôn mặt có độ rõ nét và độ tự tin cao nhất, loại bỏ các khung hình mờ hoặc không đạt yêu cầu. Việc này giúp giảm thiểu sai số trong các bước phân tích tiếp theo.

Bước 3: Phân loại cảm xúc

Khuôn mặt sau khi được chuẩn hóa được đưa vào mô hình phân loại cảm xúc để xác định trạng thái cảm xúc của học sinh. Các trạng thái cảm xúc bao gồm: Exciting, Happy, Sad, Neutral, và Unknown. Hệ thống sử dụng mô hình học sâu đã được huấn luyện trên bộ dữ liệu EmoLearn để đảm bảo độ chính xác cao trong việc phân loại. Kết quả phân loại cảm xúc từ các camera được tổng hợp để tăng cường độ tin cậy, đặc biệt trong các trường hợp cảm xúc không rõ ràng hoặc bị ảnh hưởng bởi góc quay.

Bước 4: Nhận diện danh tính

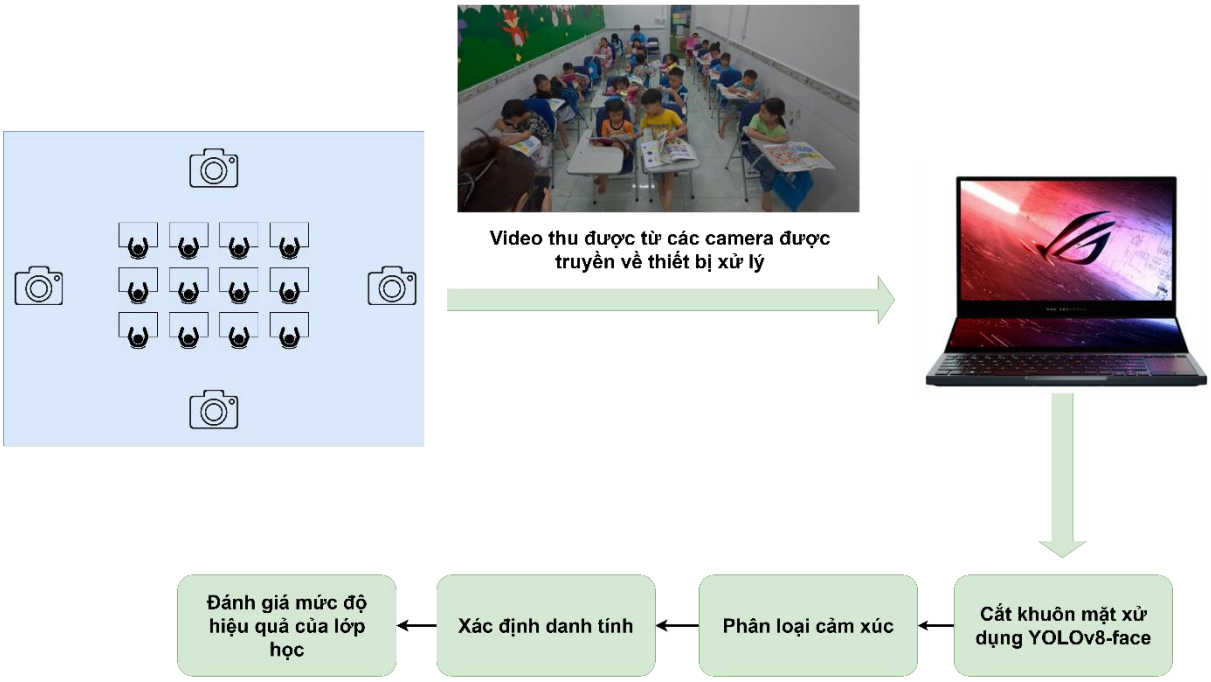
Sau khi phân loại cảm xúc, các khuôn mặt được đối chiếu với cơ sở dữ liệu danh tính học sinh để xác định danh tính cụ thể của từng người. Hệ thống sử dụng các mô hình như ArcFace và FaceNet512 để thực hiện nhận diện danh tính. Mô hình ArcFace đảm bảo độ chính xác cao, trong khi FaceNet512 cung cấp tốc độ xử lý nhanh, giúp cân bằng giữa hiệu suất và thời gian xử lý.

Bước 5: Tổng hợp và đánh giá

Hệ thống tổng hợp dữ liệu cảm xúc và danh tính của tất cả học sinh để phân tích hiệu quả lớp học. Dữ liệu từ các camera được kết hợp, sử dụng thuật toán đa camera để loại bỏ các khung hình trùng lặp hoặc không đạt tiêu chuẩn. Hiệu quả lớp học được đánh giá dựa trên các tiêu chí đã định trong chính sách, bao gồm tỷ lệ cảm xúc tích cực, trung tính và tiêu cực. Các học sinh có biểu hiện bất thường được ghi nhận lại để giáo viên có thể theo dõi và đưa ra các can thiệp phù hợp.

Bước 6: Lưu trữ và hiển thị dữ liệu

Kết quả phân tích được lưu trữ trong cơ sở dữ liệu và hiển thị trên giao diện người dùng. Giao diện cung cấp thông tin chi tiết về từng học sinh, bao gồm trạng thái cảm xúc, danh tính, và các khuyến nghị dành cho giáo viên. Hệ thống cũng cho phép xuất báo cáo tổng quan về hiệu quả lớp học, hỗ trợ giáo viên trong việc đánh giá và cải thiện chất lượng giảng dạy.



Hình 4.4. Quy trình hoạt động của hệ thống.

Hệ thống với quy trình hoạt động chặt chẽ này không chỉ cung cấp thông tin phân tích chính xác mà còn đảm bảo khả năng mở rộng và thích nghi với nhiều môi trường học tập khác nhau. Việc sử dụng cấu hình đa camera giúp tăng cường độ tin cậy và hiệu quả, đáp ứng tốt các yêu cầu thực tế trong quản lý lớp học.

4.6. Các chỉ số đánh giá

Một tập hợp các chỉ số đánh giá được sử dụng để đánh giá khả năng phát hiện của mô hình. Các chỉ số này bao gồm độ chính xác (precision), độ nhớ (recall), độ chính xác trung bình trung bình (mean average precision - mAP) tại các ngưỡng IoU là 0.5 và 0.5:0.95, số lượng tham số mô hình, kích thước mô hình và tốc độ phát hiện.

Độ Chính xác (Percision – P), một chỉ số quan trọng, định lượng tỷ lệ các mẫu tích cực được dự đoán chính xác bởi mô hình so với tất cả các mẫu được phát hiện. Nó được tính như sau:

$$P = \frac{TP}{TP + FP} \#(1)$$

Độ Nhớ (Recall – R), chỉ số thiết yếu khác, biểu thị tỷ lệ các mẫu tích cực được dự đoán chính xác bởi mô hình so với tổng số mẫu tích cực hiện có:

$$R = \frac{TP}{TP + FN} \#(2)$$

Độ Chính xác Trung bình (Average Percision – AP) là một chỉ số được suy ra từ diện tích dưới đường cong độ chính xác-độ nhớ, biểu thị hiệu suất độ chính xác-độ nhớ của mô hình:

$$AP = \int_0^1 P(R)d(R) \#(4)$$

Độ Chính xác Trung bình có nghĩa (mean Average Percision – mAP) là một chỉ số tổng hợp lấy trung bình có trọng số của các giá trị AP trên tất cả các danh mục mẫu.

Nó cung cấp một đánh giá toàn diện về hiệu suất của mô hình trên tất cả các danh mục, và được tính như sau:

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \#(5)$$

Ở đây, AP_i đề cập đến giá trị AP cho chỉ số danh mục i, và N đại diện cho tổng số danh mục trong bộ dữ liệu huấn luyện (trong trư ờng hợp này, N là 10).

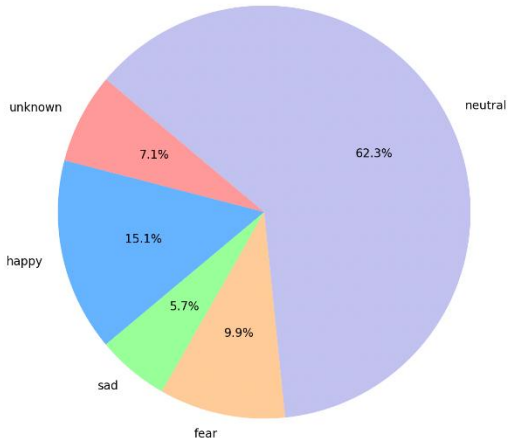
CHƯƠNG 5. KẾT QUẢ NGHIÊN CỨU

5.1. Kết quả thu thập dữ liệu

Trong quá trình triển khai và thu thập dữ liệu, hệ thống đã ghi nhận tổng cộng 15,000 hình ảnh từ các video lớp học. Sau khi hoàn thành bước cắt khung hình và trích xuất khuôn mặt, các hình ảnh được gán nhãn dựa trên các trạng thái cảm xúc, bao gồm: Unknown, Happy, Sad, Fear, và Neutral. Tỉ lệ phân bố của các nhãn được thể hiện ở Hình 5.1.

Sự phân bố ban đầu của các nhãn trong bộ dữ liệu như sau:

- Unknown: 1,065 hình ảnh, biểu thị trạng thái cảm xúc không rõ ràng hoặc không thể xác định.
- Happy: 2,272 hình ảnh, đại diện cho cảm xúc vui vẻ và hạnh phúc.
- Sad: 852 hình ảnh, mô tả trạng thái buồn bã hoặc thất vọng.
- Fear: 1,486 hình ảnh, biểu thị trạng thái lo lắng hoặc sợ hãi.
- Neutral: 9,365 hình ảnh, phản ánh cảm xúc trung tính, không thể hiện rõ nét sự tích cực hay tiêu cực.

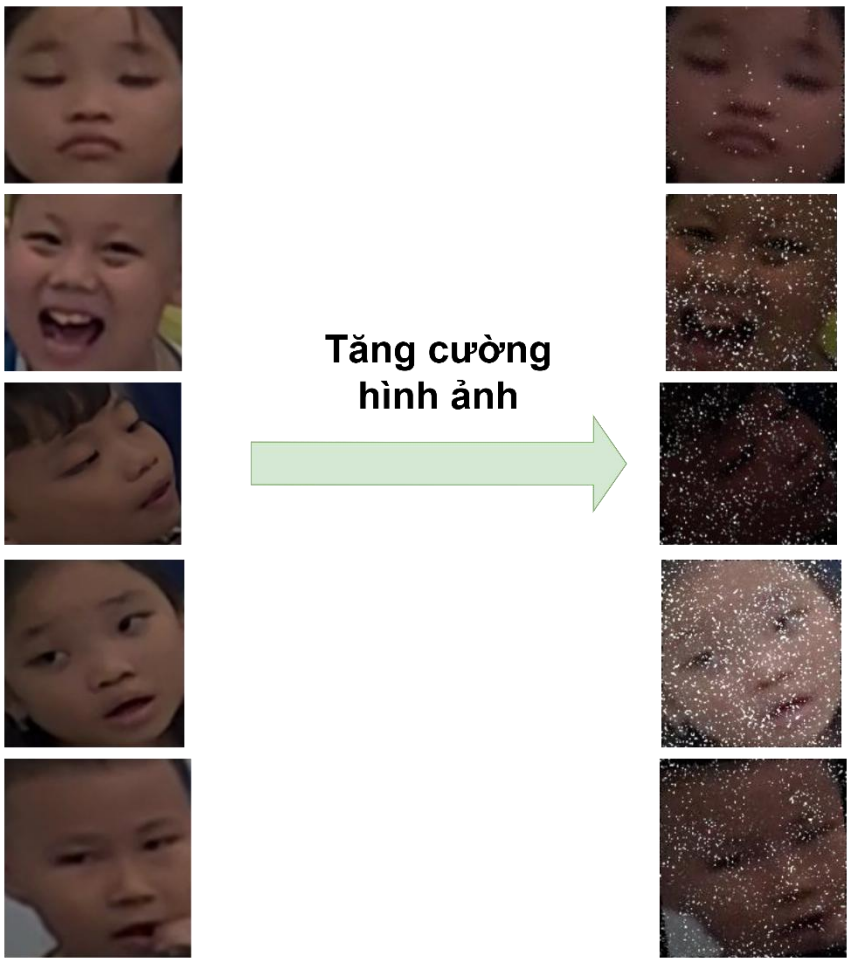


Hình 5.1. Tỉ lệ phân bố của các cảm xúc.

Sau quá trình tăng cường, bộ dữ liệu mở rộng lên tổng cộng 75,000 hình ảnh, giữ nguyên tỷ lệ phân bố giữa các nhãn như ban đầu. Điều này giúp đảm bảo rằng hệ

thống có thể xử lý hiệu quả các biểu cảm cảm xúc trong nhiều điều kiện khác nhau, đồng thời nâng cao tính ổn định và khả năng tổng quát hóa của các mô hình học sâu được huấn luyện.

Bộ dữ liệu cuối cùng không chỉ hỗ trợ huấn luyện mô hình mà còn đóng vai trò quan trọng trong việc đánh giá hiệu suất của hệ thống trong môi trường thực tế, đảm bảo rằng hệ thống có thể hoạt động hiệu quả và đáp ứng nhu cầu phân tích cảm xúc và quản lý lớp học. Một số hình ảnh từ tập dữ liệu gốc và sau khi tăng cường được thể hiện ở Hình 5.2.



Hình 5.2. Một số hình ảnh sau khi tăng cường.

5.2. Kết quả huấn luyện

Quá trình huấn luyện các mô hình được thực hiện trên hai phiên bản của bộ dữ

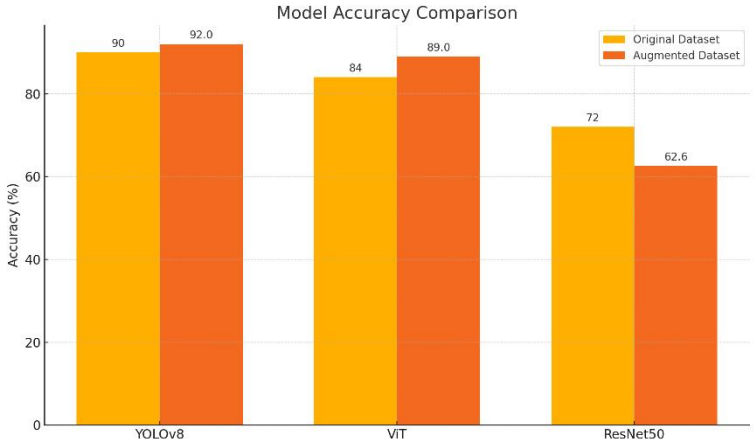
liệu: bộ dữ liệu gốc và bộ dữ liệu đã được tăng cường. Các mô hình sử dụng bao gồm YOLOv8, ViT (Vision Transformer), và ResNet50. Kết quả huấn luyện được đánh giá dựa trên hai chỉ số chính là độ chính xác (accuracy) và mất mát (loss), cung cấp cái nhìn tổng quan về hiệu suất của từng mô hình trong các điều kiện dữ liệu khác nhau. Kết quả này được thể hiện trực quan qua sơ đồ Hình 5.1 (độ chính xác) và Hình 5.2 (mất mát).

Trên bộ dữ liệu gốc, YOLOv8 đạt độ chính xác cao nhất với 90% và giá trị mất mát là 0.2. Điều này chứng tỏ YOLOv8 có khả năng mạnh mẽ trong việc học và nhận diện các đặc trưng từ dữ liệu đầu vào, phù hợp cho các tác vụ yêu cầu phát hiện chính xác biểu cảm khuôn mặt. ViT cũng thể hiện hiệu suất tốt với độ chính xác 84% và giá trị mất mát 0.22. Mô hình này cho thấy tiềm năng trong việc xử lý dữ liệu hình ảnh phức tạp, mặc dù chưa đạt được mức hiệu suất của YOLOv8. ResNet50, với độ chính xác 72% và giá trị mất mát 0.38, có hiệu suất thấp hơn so với hai mô hình còn lại. Điều này phản ánh giới hạn của ResNet50 trong việc học các đặc trưng biểu cảm từ bộ dữ liệu gốc so với các kiến trúc hiện đại hơn.

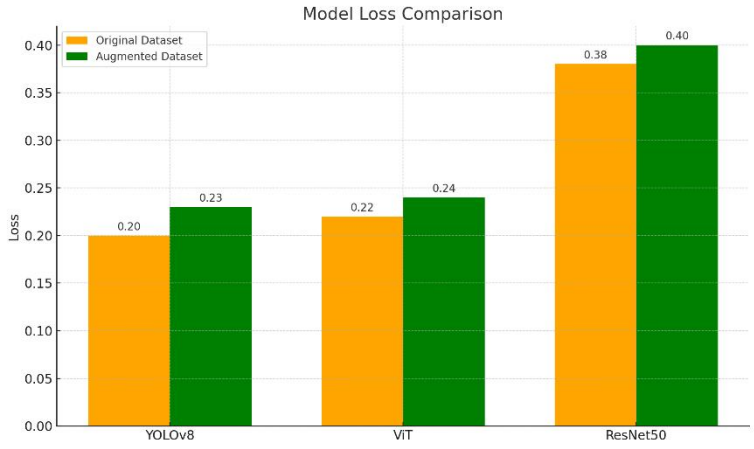
Khi áp dụng các kỹ thuật tăng cường dữ liệu, bộ dữ liệu được mở rộng từ 15,000 hình ảnh lên 75,000 hình ảnh, giúp tăng tính đa dạng và khả năng tổng quát hóa cho các mô hình. Trên bộ dữ liệu tăng cường, YOLOv8 tiếp tục dẫn đầu với độ chính xác 92% và giá trị mất mát 0.23. Mặc dù giá trị mất mát tăng nhẹ, độ chính xác được cải thiện, khẳng định khả năng tổng quát hóa tốt của mô hình này khi được huấn luyện trên dữ liệu phong phú hơn. ViT cũng có sự cải thiện đáng kể với độ chính xác đạt 89% và giá trị mất mát 0.24. Những kết quả này cho thấy ViT đã tận dụng hiệu quả sự đa dạng trong dữ liệu tăng cường để nâng cao hiệu suất nhận diện biểu cảm. Tuy nhiên, ResNet50 lại có hiệu suất giảm sút trên bộ dữ liệu tăng cường, chỉ đạt độ chính xác 62.6% và giá trị mất mát 0.4. Hiệu suất giảm này có thể do ResNet50 gặp khó khăn trong việc học từ dữ liệu tăng cường với độ phức tạp cao hơn, đặc biệt khi so sánh với các mô hình hiện đại như YOLOv8 và ViT.

Hình 5.3 minh họa sự so sánh độ chính xác giữa các mô hình trên cả hai phiên

bản của bộ dữ liệu. Có thể thấy rõ rằng YOLOv8 luôn giữ vị trí dẫn đầu, tiếp theo là ViT, trong khi ResNet50 có hiệu suất thấp nhất. Hình 5.4 trình bày biểu đồ mất mát của các mô hình, với xu hướng cho thấy rằng YOLOv8 và ViT duy trì mức mất mát thấp và ổn định hơn so với ResNet50. Các sơ đồ này không chỉ cung cấp một cái nhìn trực quan về hiệu suất của các mô hình mà còn làm nổi bật vai trò của dữ liệu tăng cường trong việc cải thiện khả năng học của các mô hình học sâu.



Hình 5.3.Một số hình ảnh sau khi tăng cường.



Hình 5.4.Một số hình ảnh sau khi tăng cường.

Nhìn chung, kết quả huấn luyện cho thấy YOLOv8 là mô hình mạnh mẽ nhất, với khả năng thích ứng tốt trên cả bộ dữ liệu gốc và tăng cường. ViT cũng thể hiện sự cải thiện đáng kể khi áp dụng tăng cường dữ liệu, làm nổi bật tiềm năng của kiến trúc này trong các ứng dụng nhận diện biểu cảm. Trong khi đó, ResNet50 có những hạn chế

rõ rệt khi làm việc với dữ liệu phức tạp, đặt ra nhu cầu xem xét các phương pháp tối ưu hóa hoặc thay thế mô hình cho các bài toán yêu cầu độ chính xác cao hơn. Các kết quả này không chỉ cung cấp cơ sở để lựa chọn mô hình phù hợp nhất cho các ứng dụng thực tế mà còn khẳng định tầm quan trọng của việc xử lý và tăng cường dữ liệu để cải thiện hiệu suất của các mô hình học sâu.

5.3. Kết quả triển khai hệ thống

Hệ thống EmoLearn được triển khai với sự tích hợp của các mô-đun phát hiện khuôn mặt, phân loại cảm xúc và nhận diện danh tính, nhằm phân tích chi tiết và toàn diện các động lực học lớp học. Kết quả triển khai được đánh giá thông qua hai cấu hình: hệ thống không sử dụng multicamera và hệ thống có multicamera. Cả hai cấu hình được trình bày và so sánh chi tiết để làm rõ hiệu quả và hạn chế của từng phương pháp.

Mô hình YOLOv8-Face đã được chọn làm công cụ phát hiện khuôn mặt chính của hệ thống nhờ vào độ chính xác vượt trội trên cả hai bộ dữ liệu. Với độ chính xác đạt 90.0% trên bộ dữ liệu gốc và cải thiện lên 92.8% trên bộ dữ liệu tăng cường, YOLOv8-Face chứng minh khả năng xử lý ổn định trong các môi trường lớp học đa dạng. Những điều kiện này bao gồm ánh sáng thay đổi và khuôn mặt bị che khuất, đảm bảo rằng dữ liệu đầu vào đạt chất lượng cao nhất cho các bước phân tích tiếp theo. Điều này thể hiện vai trò quan trọng của mô-đun phát hiện khuôn mặt trong toàn bộ hệ thống.

Về nhận diện danh tính, mô hình ArcFace thể hiện sự vượt trội với độ chính xác đạt 88.0%, cao hơn đáng kể so với các mô hình khác như VGGFace (83.0%), FaceNet (62.0%) và FaceNet512 (58.0%). ArcFace cung cấp sự ổn định trong việc nhận diện danh tính học sinh ngay cả trong các tình huống khó khăn như ánh sáng yếu hoặc góc nhìn không thuận lợi. Bảng 5.1 minh họa chi tiết kết quả của các mô hình, cho thấy khả năng của ArcFace trong việc đảm bảo tính toàn vẹn dữ liệu danh tính, là yếu tố quan trọng trong phân tích cảm xúc và hành vi học sinh.

Mô hình	**Accuracy**

ArcFace	88.0
VGGFace	83.0
FaceNet	62.0
FaceNet512	58.0

Bảng 5.1. Kết quả nhận diện danh tính.

Sự khác biệt về hiệu suất hệ thống giữa cấu hình không sử dụng multicamera và có multicamera được thể hiện rõ ràng qua các chỉ số hiệu suất. Trong cấu hình không sử dụng multicamera, hệ thống đạt tốc độ xử lý cao với trung bình 21.60 FPS và thời gian xử lý mỗi khung hình là 0.047 giây. Tuy nhiên, điểm tin cậy trung bình chỉ đạt 0.88, cho thấy hạn chế về độ chính xác khi phải xử lý các khung hình từ một góc camera duy nhất. Tình trạng này dẫn đến việc bỏ sót hoặc sai lệch trong nhận diện khuôn mặt, đặc biệt khi học sinh quay đi hướng khác hoặc khuất mặt.

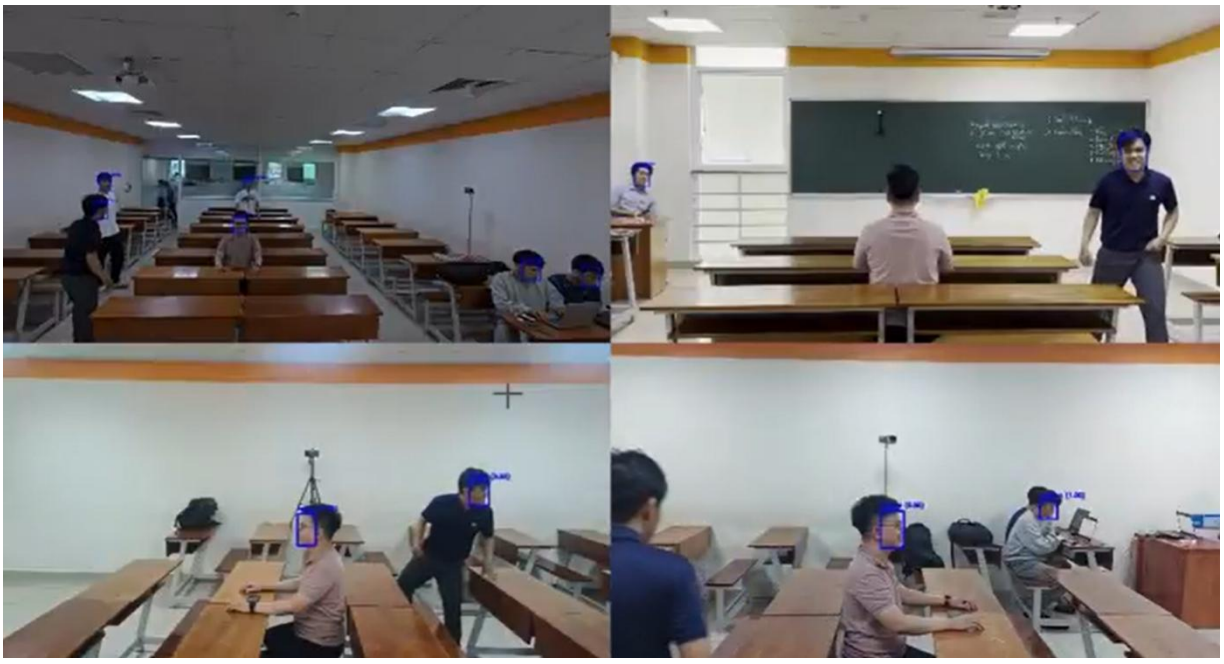
Cấu hình sử dụng multicamera, với bốn camera được lắp đặt tại các góc lớp học, cải thiện đáng kể độ tin cậy của hệ thống. Điểm tin cậy trung bình tăng lên 0.93, đảm bảo dự đoán cảm xúc chính xác hơn và toàn diện hơn. Tuy nhiên, tốc độ xử lý giảm xuống còn 7.20 FPS, với thời gian xử lý trung bình là 0.14 giây mỗi khung hình. Mặc dù sử dụng bộ nhớ tăng lên (73.90% so với 58.50% trong cấu hình không multicamera), lợi ích từ việc tăng cường khả năng quan sát và độ chính xác rõ ràng vượt trội. Bảng 5.2 so sánh các chỉ số hiệu suất giữa hai cấu hình hệ thống, làm nổi bật vai trò của multicamera trong việc nâng cao chất lượng phân tích.

Chỉ số	**Không có multicam**	**Có multicam**
FPS trung bình	21.60	7.20
Thời gian xử lý mỗi khung hình (giây)	0.047	0.14
Điểm tin cậy trung bình	0.88	0.93
Sử dụng bộ nhớ (%)	58.50	73.90

Bảng 5.2. Hiệu suất giữa hai hệ thống.

Hình 5.5 minh họa hoạt động của hệ thống không sử dụng multicamera, trong khi Hình 5.6 thể hiện cấu hình hệ thống có multicamera. Sự khác biệt giữa hai hình ảnh này làm rõ lợi ích của multicamera trong việc cung cấp góc nhìn toàn diện và hạn chế các sai số do góc quay hẹp hoặc che khuất. Các minh họa này không chỉ làm nổi bật sự khác biệt về cấu hình mà còn cho thấy hiệu quả của việc áp dụng công nghệ multicamera trong môi trường lớp học.

Hình 5.5. Hệ thống không multicam.



Hình 5.5. Hệ thống multicam.

Kết quả triển khai hệ thống cho thấy rằng việc tích hợp các mô hình học sâu hiện đại và cấu hình multicamera mang lại hiệu suất cao và độ tin cậy vượt trội. Hệ thống không chỉ đáp ứng yêu cầu về thời gian thực mà còn cung cấp thông tin phân

tích chi tiết, hỗ trợ giáo viên trong việc quản lý lớp học và nâng cao chất lượng giảng dạy. Điều này khẳng định rằng EmoLearn là một giải pháp mạnh mẽ, đáp ứng tốt các nhu cầu thực tế của giáo dục hiện đại.

CHƯƠNG 6. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

6.1. Kết luận

Hệ thống EmoLearn đã được triển khai thành công với sự tích hợp các mô hình học sâu tiên tiến nhằm phân tích hiệu quả lớp học dựa trên cảm xúc và danh tính học sinh. Hệ thống đã chứng minh khả năng vượt trội trong việc cung cấp thông tin chi tiết và đáng tin cậy về động lực học lớp học. Việc sử dụng mô hình YOLOv8-Face để phát hiện khuôn mặt, cùng với ArcFace để nhận diện danh tính, đã giúp đảm bảo độ chính xác cao trong các điều kiện lớp học phức tạp. Cấu hình multicamera mang lại những cải thiện đáng kể về độ tin cậy và khả năng quan sát toàn diện, cho phép hệ thống xử lý hiệu quả ngay cả trong các tình huống khó khăn như góc quay hẹp hoặc ánh sáng thay đổi. Những kết quả này không chỉ khẳng định hiệu quả của hệ thống mà còn mở ra tiềm năng ứng dụng rộng rãi trong lĩnh vực giáo dục.

Tuy nhiên, hệ thống vẫn tồn tại một số hạn chế cần được khắc phục trong tương lai. Việc sử dụng cấu hình multicamera yêu cầu tài nguyên phần cứng cao, đồng thời làm giảm tốc độ xử lý do khối lượng dữ liệu lớn hơn. Hơn nữa, hệ thống hiện tại chỉ tập trung vào các trạng thái cảm xúc cơ bản, chưa đáp ứng được yêu cầu phân tích các trạng thái cảm xúc phức tạp hơn. Điều này đặt ra yêu cầu tiếp tục nghiên cứu và phát triển để tối ưu hóa hiệu suất và mở rộng phạm vi ứng dụng của hệ thống.

6.2. Định hướng phát triển

Trong tương lai, hệ thống EmoLearn sẽ được cải tiến để khắc phục những hạn chế hiện tại và mở rộng khả năng ứng dụng. Một trong những hướng đi quan trọng là tối ưu hóa hiệu suất hệ thống thông qua việc sử dụng các mô hình học sâu hiện đại hơn với khả năng xử lý nhanh và chính xác hơn. Việc áp dụng các thuật toán giảm độ phức tạp và tối ưu hóa quy trình huấn luyện sẽ giúp cải thiện tốc độ và hiệu quả sử dụng tài nguyên.

Ngoài ra, việc mở rộng phân tích cảm xúc để bao gồm các trạng thái phức tạp như căng thẳng, hứng thú, hoặc mức độ chú ý sẽ làm tăng tính chính xác và giá trị của

hệ thống trong các ứng dụng thực tế. Hệ thống cũng có thể được áp dụng trong các môi trường học tập đa dạng, bao gồm lớp học trực tuyến hoặc các không gian học tập với sự tham gia của các nhóm văn hóa khác nhau.

Hợp tác với các tổ chức giáo dục và tiến hành các thử nghiệm thực tế sẽ là một phần quan trọng trong chiến lược phát triển. Điều này không chỉ giúp kiểm nghiệm hiệu quả của hệ thống trong các bối cảnh thực tế mà còn cung cấp thông tin quý giá để điều chỉnh và hoàn thiện các tính năng. Cuối cùng, việc tích hợp hệ thống với các công nghệ tiên tiến như phân tích video thời gian thực hoặc ứng dụng trí tuệ nhân tạo trong quản lý lớp học sẽ mở ra những cơ hội mới để cải thiện chất lượng giáo dục và trải nghiệm học tập của học sinh.

TÀI LIỆU THAM KHẢO

[1] J. Allen, A. Gregory, A. Mikami, J. Lun, B. Hamre, and R. Pianta, "Observations of Effective Teacher–Student Interactions in Secondary School Classrooms: Predicting Student Achievement With the Classroom Assessment Scoring System—Secondary," *School Psychology Review*, vol. 42, no. 1, pp. 76–98, Mar. 2013, doi: 10.1080/02796015.2013.12087492.

[2] D. J. Pope, H. Butler, and P. Qualter, "Emotional Understanding and Color-Emotion Associations in Children Aged 7-8 Years," *Child Development Research*, vol. 2012, no. 1, p. 975670, 2012, doi: 10.1155/2012/975670.

[3] M. E. S. Loevaas *et al.*, "Emotion regulation and its relation to symptoms of anxiety and depression in children aged 8–12 years: does parental gender play a differentiating role?," *BMC Psychol*, vol. 6, no. 1, p. 42, Aug. 2018, doi: 10.1186/s40359-018-0255-y.

[4] G. Hagenauer, T. Hascher, and S. E. Volet, "Teacher emotions in the classroom: associations with students' engagement, classroom discipline and the interpersonal teacher-student relationship," *Eur J Psychol Educ*, vol. 30, no. 4, pp. 385–403, Dec. 2015, doi: 10.1007/s10212-015-0250-0.

[5] C. Braet *et al.*, "Emotion Regulation in Children with Emotional Problems," *Cogn Ther Res*, vol. 38, no. 5, pp. 493–504, Oct. 2014, doi: 10.1007/s10608-014-9616-x.

[6] A. N. Veraksa, M. N. Gavrilova, and F. Pons, "The impact of classroom quality on young children's emotion understanding," *European Early Childhood Education Research Journal*, vol. 28, no. 5, pp. 690–700, Sep. 2020, doi: 10.1080/1350293X.2020.1817240.

[7] S. M. Jones, R. Bailey, and R. Jacob, “Social-emotional learning is essential to classroom management,” *Phi Delta Kappan*, vol. 96, no. 2, pp. 19–24, Oct. 2014, doi: 10.1177/0031721714553405.

[8] F. L. Brown *et al.*, “Psychological interventions for children with emotional and behavioral difficulties aged 5–12 years: An evidence review,” *Cambridge Prisms: Global Mental Health*, vol. 11, p. e75, Jan. 2024, doi: 10.1017/gmh.2024.57.

[9] A. Mollahosseini, B. Hasani, and M. H. Mahoor, “AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild,” *IEEE Trans. Affective Comput.*, vol. 10, no. 1, pp. 18–31, Jan. 2019, doi: 10.1109/TAFFC.2017.2740923.

[10] L. Zahara, P. Musa, E. Prasetyo Wibowo, I. Karim, and S. Bahri Musa, “The Facial Emotion Recognition (FER-2013) Dataset for Prediction System of Micro-Expressions Face Using the Convolutional Neural Network (CNN) Algorithm based Raspberry Pi,” in *2020 Fifth International Conference on Informatics and Computing (ICIC)*, Nov. 2020, pp. 1–9. doi: 10.1109/ICIC50835.2020.9288560.

[11] S. Koelstra *et al.*, “DEAP: A Database for Emotion Analysis ;Using Physiological Signals,” *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, Jan. 2012, doi: 10.1109/T-AFFC.2011.15.

[12] T. R. S. De Silva, K. Y. Dayananda, R. C. Galagama Arachchi, M. K. S. B. Amerasekara, S. Silva, and N. Gamage, “Solution to Measure Employee Productivity with Employee Emotion Detection,” in *2022 4th International Conference on Advancements in Computing (ICAC)*, Dec. 2022, pp. 210–215. doi: 10.1109/ICAC57685.2022.10025132.

[13] I. A. M. Verpaalen, G. Bijsterbosch, L. Mobach, G. Bijlstra, M. Rinck, and A. M. Klein, “Validating the Radboud faces database from a child’s perspective,” *Cognition and Emotion*, vol. 33, no. 8, pp. 1531–1547, Nov. 2019, doi: 10.1080/02699931.2019.1577220.

[14] M. R. Reyes, M. A. Brackett, S. E. Rivers, M. White, and P. Salovey, “Classroom emotional climate, student engagement, and academic achievement,” *Journal of Educational Psychology*, vol. 104, no. 3, pp. 700–712, 2012, doi: 10.1037/a0027268.

[15] V. Prokofieva, S. Kostromina, S. Polevaia, and F. Fenouillet, “Understanding Emotion-Related Processes in Classroom Activities Through Functional Measurements,” *Front. Psychol.*, vol. 10, Oct. 2019, doi: 10.3389/fpsyg.2019.02263.

[16] A. C. Frenzel, T. Goetz, O. Lüdtke, R. Pekrun, and R. E. Sutton, “Emotional transmission in the classroom: Exploring the relationship between teacher and student enjoyment,” *Journal of Educational Psychology*, vol. 101, no. 3, pp. 705–716, 2009, doi: 10.1037/a0014695.

[17] W.-L. Zheng and B.-L. Lu, “Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks,” *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, Sep. 2015, doi: 10.1109/TAMD.2015.2431497.

[18] V. LoBue and C. Thrasher, “The Child Affective Facial Expression (CAFE) set: validity and reliability from untrained adults,” *Front. Psychol.*, vol. 5, Jan. 2015, doi: 10.3389/fpsyg.2014.01532.

[19] M. V. Garrido and M. Prada, “KDEF-PT: Valence, Emotional Intensity, Familiarity and Attractiveness Ratings of Angry, Neutral, and Happy Faces,” *Front. Psychol.*, vol. 8, Dec. 2017, doi: 10.3389/fpsyg.2017.02181.

[20] D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, and S. Ravi, “GoEmotions: A Dataset of Fine-Grained Emotions,” in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault, Eds., Online: Association for Computational Linguistics, Jul. 2020, pp. 4040–4054. doi: 10.18653/v1/2020.acl-main.372.

[21] M. M. A. Parambil, L. Ali, F. Alnajjar, and M. Gochoo, “Smart Classroom: A Deep Learning Approach towards Attention Assessment through Class Behavior Detection,” in *2022 Advances in Science and Engineering Technology International Conferences (ASET)*, Feb. 2022, pp. 1–6. doi: 10.1109/ASET53988.2022.9735018.

[22] K. V. Karan, V. Bahel, R. Ranjana, and T. Subha, “Transfer Learning Approach for Analyzing Attentiveness of Students in an Online Classroom Environment with Emotion Detection,” in *Innovations in Computational Intelligence and Computer Vision*, S. Roy, D. Sinwar, T. Perumal, A. Slowik, and J. M. R. S. Tavares, Eds., Singapore: Springer Nature, 2022, pp. 253–261. doi: 10.1007/978-981-19-0475-2_23.

[23] L. Li and D. Yao, “Emotion Recognition in Complex Classroom Scenes Based on Improved Convolutional Block Attention Module Algorithm,” *IEEE Access*, vol. 11, pp. 143050–143059, 2023, doi: 10.1109/ACCESS.2023.3340510.

[24] Krithika L.B and Lakshmi Priya GG, “Student Emotion Recognition System (SERS) for e-learning Improvement Based on Learner Concentration Metric,” *Procedia Computer Science*, vol. 85, pp. 767–776, Jan. 2016, doi: 10.1016/j.procs.2016.05.264.

[25] H. Abbass, “Editorial: What is Artificial Intelligence?,” *IEEE Transactions on Artificial Intelligence*, vol. 2, no. 2, pp. 94–95, Apr. 2021, doi: 10.1109/TAI.2021.3096243.

[26] S. Angra and S. Ahuja, “Machine learning and its applications: A review,” in *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*, Mar. 2017, pp. 57–60. doi: 10.1109/ICBDACI.2017.8070809.

[27] H. Wang, C. Ma, and L. Zhou, “A Brief Review of Machine Learning and Its Application,” in *2009 International Conference on Information Engineering and Computer Science*, Oct. 2009, pp. 1–4. doi: 10.1109/ICIECS.2009.5362936.

[28] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[29] J. Schmidhuber, “Deep Learning in Neural Networks: An Overview,” *Neural Networks*, vol. 61, pp. 85–117, Jan. 2015, doi: 10.1016/j.neunet.2014.09.003.

[30] Soniya, S. Paul, and L. Singh, “A review on advances in deep learning,” in *2015 IEEE Workshop on Computational Intelligence: Theories, Applications and Future Directions (WCI)*, Oct. 2015, pp. 1–6. doi: 10.1109/WCI.2015.7495514.

[31] J. Chai and A. Li, “Deep Learning in Natural Language Processing: A State-of-the-Art Survey,” in *2019 International Conference on Machine Learning and Cybernetics (ICMLC)*, Jul. 2019, pp. 1–6. doi: 10.1109/ICMLC48188.2019.8949185.

[32] M. Eremia, C.-C. Liu, and A.-A. Edris, “Expert Systems,” in *Advanced Solutions in Power Systems: HVDC, FACTS, and Artificial Intelligence*, IEEE, 2016, pp. 731–754. doi: 10.1002/9781119175391.ch15.

[33] S. Yang, “Evolutionary Computation for Dynamic Optimization Problems,” in *Proceedings of the Companion Publication of the 2015 Annual Conference on Genetic and Evolutionary Computation*, in GECCO Companion ’15. New York, NY, USA: Association for Computing Machinery, Jul. 2015, pp. 629–649. doi: 10.1145/2739482.2756589.

[34] M. Ashcroft, “An introduction to Bayesian networks in systems and control,” in *18th International Conference on Automation and Computing (ICAC)*, Sep. 2012, pp. 1–6. Accessed: Jan. 12, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/6330539

[35] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 779–788. doi: 10.1109/CVPR.2016.91.

[36] J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, “A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS,” *Machine Learning and Knowledge Extraction*, vol. 5, no. 4, Art. no. 4, Dec. 2023, doi: 10.3390/make5040083.

[37] A. Dosovitskiy *et al.*, “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale.” arXiv, Jun. 2021. doi: 10.48550/arXiv.2010.11929.

[38] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jegou, “Training data-efficient image transformers & distillation through attention,” in *Proceedings of the 38th International Conference on Machine Learning*, PMLR, Jul. 2021, pp. 10347–10357. Accessed: Jan. 12, 2025. [Online]. Available: https://proceedings.mlr.press/v139/touvron21a.html

[39] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” *CoRR*, vol. abs/1512.03385, 2015, Accessed: Nov. 01, 2023. [Online]. Available: http://arxiv.org/abs/1512.03385

[40] C. Shorten and T. M. Khoshgoftaar, “A survey on Image Data Augmentation for Deep Learning,” *Journal of Big Data*, vol. 6, no. 1, p. 60, Jul. 2019, doi: 10.1186/s40537-019-0197-0.

[41] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, “mixup: Beyond Empirical Risk Minimization,” Apr. 27, 2018, *arXiv*: arXiv:1710.09412. Accessed: Jun. 24, 2024. [Online]. Available: http://arxiv.org/abs/1710.09412