**ỦY BAN NHÂN DÂN THÀNH PHỐ HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC SÀI GÒN**

**Họ Tên Sinh Viên**

*Tống Đức Duy*

*Lê Trung Kiên*

**NGHIÊN CỨU ỨNG DỤNG YOLO CHO BÀI TOÁN GIÁM SÁT LỚP HỌC**

**ĐỀ CƯƠNG LUẬN VĂN cử nhân**

ngành: Công nghệ thông tin

Mã số NGÀNH: 7480201CLC

**Người hướng dẫn Chủ nhiệm chuyên ngành**

**GV TS.Đỗ Như Tài GV B**

**Thành phố Hồ Chí Minh, năm 2025**

# ĐỀ CƯƠNG LUẬN VĂN

## **1. Lý do chọn đề tài**

Trong thời đại công nghệ 4.0, việc ứng dụng trí tuệ nhân tạo (AI) vào các lĩnh vực giáo dục đã mở ra nhiều hướng phát triển mới. Đặc biệt, các mô hình học sâu (Deep Learning) và mạng nơ-ron tích chập (CNN) đã mang lại hiệu quả vượt trội trong các ứng dụng xử lý hình ảnh, bao gồm nhận diện khuôn mặt, phân tích hành vi, và giám sát thông minh.

YOLO (You Only Look Once), một phương pháp phát hiện đối tượng thời gian thực, nổi bật với tốc độ nhanh và độ chính xác cao, là công cụ phù hợp để giải quyết bài toán giám sát lớp học. Việc ứng dụng YOLO để nhận diện hành động, cảm xúc và trạng thái của học sinh từ video giám sát có ý nghĩa quan trọng trong việc đánh giá chất lượng giảng dạy, quản lý lớp học hiệu quả, và hỗ trợ giáo viên đưa ra phương pháp giảng dạy tối ưu.

Bài toán giám sát lớp học, bao gồm việc nhận diện khuôn mặt, trích xuất đặc điểm cảm xúc, và theo dõi hành động học sinh, không chỉ mang lại giá trị khoa học mà còn đóng góp thiết thực vào thực tiễn giáo dục.

## **2. Tổng quan vấn đề nghiên cứu**

## **2.1. Tình hình nghiên cứu hiện tại**

Phát hiện đối tượng và nhận diện khuôn mặt đã trở thành một chủ đề nghiên cứu quan trọng trong lĩnh vực thị giác máy tính. Các mô hình như R-CNN, SSD, và YOLO đã chứng minh hiệu quả trong việc phát hiện đối tượng. Trong đó, YOLO với ưu thế về thời gian xử lý nhanh, đã được áp dụng rộng rãi trong các bài toán yêu cầu xử lý thời gian thực.

Nhận diện cảm xúc từ khuôn mặt (Facial Emotion Recognition - FER) cũng là một nhánh quan trọng trong thị giác máy tính. Các nghiên cứu gần đây tập trung vào việc sử dụng các mạng CNN và cơ chế Attention để cải thiện độ chính xác khi nhận diện cảm xúc từ hình ảnh và video.

Tuy nhiên, việc tích hợp các kỹ thuật phát hiện đối tượng (YOLO), trích xuất đặc điểm khuôn mặt, và nhận diện cảm xúc vào một hệ thống đồng bộ để giải quyết bài toán giám sát lớp học vẫn còn là thách thức lớn.

## **2.2. Hướng tiếp cận của đề tài**

Luận văn tập trung nghiên cứu khả năng ứng dụng thuật toán YOLO trong việc phát hiện khuôn mặt từ video lớp học, kết hợp với các phương pháp phân tích cảm xúc khuôn mặt để nhận diện hành động và trạng thái cảm xúc của học sinh. Nghiên cứu sẽ tiến hành so sánh hiệu suất giữa mô hình tích hợp YOLO và các phương pháp truyền thống, từ đó đánh giá tính hiệu quả và khả năng ứng dụng trong bài toán giám sát lớp học.

Cơ chế nhận diện khuôn mặt được thực hiện qua thuật toán YOLO, một trong những kiến trúc hàng đầu về nhận diện đối tượng trong thời gian thực. Với khả năng phát hiện khuôn mặt nhanh chóng và chính xác, YOLO sẽ giúp hệ thống giám sát lớp học theo dõi đồng thời nhiều khuôn mặt, ngay cả trong môi trường phức tạp như lớp học đông người và điều kiện ánh sáng thay đổi.

Để phân tích cảm xúc khuôn mặt, luận văn tích hợp các module trích xuất đặc trưng khuôn mặt với các thuật toán nhận diện cảm xúc hiện đại. Quá trình này bao gồm việc sử dụng Face Feature Extraction để trích xuất đặc điểm khuôn mặt và áp dụng các mô hình nhận diện cảm xúc như CNN hoặc LSTM để xác định cảm xúc của từng học sinh.

Hệ thống giám sát lớp học sẽ được tối ưu hóa với sự hỗ trợ của Hungarian Matching, thông qua việc xây dựng cost matrix để theo dõi và cập nhật trạng thái khuôn mặt học sinh trong suốt quá trình phân tích. Kỹ thuật này giúp giảm thiểu lỗi nhận diện và đảm bảo hệ thống hoạt động hiệu quả trong thời gian thực.

Ngoài ra, nghiên cứu cũng xem xét các yếu tố ảnh hưởng đến hiệu suất nhận diện, bao gồm góc nhìn, biểu cảm khuôn mặt đa dạng, và sự khác biệt giữa các cá nhân. Thông qua việc xây dựng các kịch bản thực nghiệm và tối ưu hóa thuật toán, luận văn kỳ vọng sẽ cải thiện độ chính xác và khả năng áp dụng của hệ thống vào các lớp học thực tế.

## **3. Mục đích và nhiệm vụ nghiên cứu**

**Mục đích nghiên cứu:** Đánh giá khả năng ứng dụng YOLO trong việc giám sát lớp học thông qua nhận diện hành động và cảm xúc học sinh. Hệ thống sẽ giúp phát hiện các trạng thái như chú ý, mất tập trung, ngủ gật, hay các hành vi bất thường khác, từ đó hỗ trợ giáo viên quản lý lớp học hiệu quả hơn.

**Nhiệm vụ nghiên cứu:**

* **Xây dựng và huấn luyện mô hình**: Áp dụng YOLO để phát hiện khuôn mặt học sinh trong video giám sát lớp học. Tích hợp các phương pháp trích xuất đặc điểm khuôn mặt và nhận diện cảm xúc vào hệ thống.
* **Thiết lập kịch bản thực nghiệm**: Thiết kế các kịch bản thực nghiệm phù hợp và tiến hành thử nghiệm trên bộ dữ liệu tiêu chuẩn (Head Dataset và bộ dữ liệu cảm xúc FER+ hoặc tương tự).
* **Đánh giá hiệu suất mô hình**: Sử dụng các chỉ số đánh giá quan trọng như độ chính xác (Accuracy), Precision, Recall và F1-score để đo lường hiệu suất của các mô hình.
* **Phân tích hành vi của mô hình**: Nghiên cứu cách hệ thống giám sát lớp học phản hồi trong các trường hợp cụ thể như nhận diện cảm xúc, theo dõi hành động, và phân loại trạng thái chú ý/mất tập trung.
* **So sánh và kết luận**: So sánh hiệu suất giữa các mô hình có và không tích hợp các bước trích xuất đặc điểm và nhận diện cảm xúc. Đưa ra kết luận về hiệu quả của hệ thống trong việc giám sát lớp học.

## **4. Đối tượng và phạm vi nghiên cứu**

**Đối tượng nghiên cứu:** Luận văn tập trung nghiên cứu trên các phương pháp phát hiện đối tượng và nhận diện cảm xúc trong lĩnh vực thị giác máy tính. Với YOLO (You Only Look Once) được chọn làm mô hình nền tảng, luận văn sẽ áp dụng YOLO để phát hiện khuôn mặt học sinh trong video giám sát lớp học. Tiếp đó, các mạng nơ-ron tích chập (CNN) sẽ được tích hợp để trích xuất đặc điểm khuôn mặt và nhận diện cảm xúc. Nghiên cứu sẽ được triển khai trên tập dữ liệu Head Dataset và các bộ dữ liệu nhận diện cảm xúc phổ biến như FER+, nhằm kiểm chứng hiệu quả của các mô hình trong việc phát hiện hành vi và trạng thái cảm xúc của học sinh trong lớp học.

**Phạm vi nghiên cứu**

Phạm vi nghiên cứu của luận văn tập trung vào việc ứng dụng mô hình YOLO và các phương pháp nhận diện cảm xúc trong bài toán giám sát lớp học. Cụ thể, luận văn triển khai các mô hình phát hiện khuôn mặt và nhận diện cảm xúc từ video giám sát trong môi trường lớp học thực tế. Các nghiên cứu sẽ được thực hiện trên tập dữ liệu Head Dataset cho bài toán phát hiện khuôn mặt và các bộ dữ liệu cảm xúc như FER+ để nhận diện cảm xúc của học sinh.

Cụ thể, phạm vi nghiên cứu bao gồm:

* **Kiến trúc mạng**: Nghiên cứu và áp dụng mô hình YOLO để phát hiện khuôn mặt học sinh từ video giám sát lớp học. Các mạng nơ-ron tích chập (CNN) được sử dụng để trích xuất đặc điểm khuôn mặt và nhận diện cảm xúc.
* **Các kỹ thuật nhận diện cảm xúc:** Tích hợp các phương pháp học sâu, đặc biệt là CNN, để nhận diện chính xác cảm xúc của học sinh từ các khung hình video, bao gồm các trạng thái như chú ý, mất tập trung, hoặc ngủ gật.
* **Tập dữ liệu**: Sử dụng Head Dataset cho bài toán phát hiện khuôn mặt và các bộ dữ liệu nhận diện cảm xúc như FER+ để nhận diện cảm xúc học sinh. Tập dữ liệu được tiền xử lý để phù hợp với mục tiêu nghiên cứu và đảm bảo chất lượng đầu vào.
* **Đánh giá mô hình**: Sử dụng các chỉ số như độ chính xác (Accuracy), Precision, Recall, và F1-score để đo lường và so sánh hiệu quả của các mô hình có và không tích hợp các phương pháp trích xuất đặc điểm và nhận diện cảm xúc.

Phạm vi nghiên cứu được giới hạn ở việc giám sát trạng thái và cảm xúc học sinh từ các video giám sát có góc nhìn cố định, với mục tiêu đánh giá hiệu quả của hệ thống trong môi trường lớp học thực tế.

## **5. Phương pháp nghiên cứu**

Luận văn áp dụng phương pháp nghiên cứu kết hợp giữa phương pháp lý thuyết, phương pháp thực nghiệm và phương pháp chuyên gia, nhằm đảm bảo tính khoa học, hệ thống và hiệu quả trong quá trình nghiên cứu.

Phương pháp lý thuyết:

* Nghiên cứu các tài liệu liên quan đến mô hình YOLO và các kỹ thuật nhận diện cảm xúc từ hình ảnh và video.
* Phân tích và tổng hợp lý thuyết về các phương pháp học sâu, bao gồm các mô hình phát hiện đối tượng, trích xuất đặc điểm khuôn mặt và nhận diện cảm xúc.
* Tìm hiểu thuật toán Hungarian Matching để theo dõi trạng thái của học sinh qua từng khung hình video.
* Lựa chọn nguồn tài liệu uy tín từ các tạp chí khoa học hàng đầu như IEEE, Springer và các hội nghị chuyên ngành về thị giác máy tính và học sâu (CVPR, ICCV, NeurIPS).

Phương pháp thực nghiệm:

* Cài đặt mô hình: Sử dụng ngôn ngữ Python và các thư viện học sâu như TensorFlow, PyTorch để triển khai các mô hình YOLO và CNN.
* Tiền xử lý dữ liệu: Chuẩn bị tập dữ liệu Head Dataset và các bộ dữ liệu nhận diện cảm xúc (FER+). Tiến hành các bước tiền xử lý dữ liệu, như cắt khung hình, cân bằng dữ liệu và chuẩn hóa ảnh.
* Thực nghiệm: Huấn luyện và kiểm thử các mô hình phát hiện khuôn mặt và nhận diện cảm xúc trên tập dữ liệu đã chuẩn bị.
* Đánh giá: Sử dụng các chỉ số như độ chính xác (Accuracy), Precision, Recall và F1-score để đo lường hiệu suất của mô hình.
* So sánh hiệu quả: Đánh giá và so sánh hiệu suất giữa các mô hình có và không tích hợp trích xuất đặc điểm khuôn mặt và nhận diện cảm xúc.

Phương pháp chuyên gia:

* Tham khảo ý kiến của giảng viên hướng dẫn và các chuyên gia trong lĩnh vực thị giác máy tính để định hướng nghiên cứu, giải quyết các vấn đề kỹ thuật và đảm bảo tính chính xác của các mô hình.
* Dựa trên phản hồi từ chuyên gia, tối ưu hóa mô hình, cải thiện quy trình thực nghiệm và điều chỉnh các bước nghiên cứu để đạt kết quả tốt nhất.

Phương pháp nghiên cứu này kết hợp chặt chẽ giữa lý thuyết và thực nghiệm, cùng với sự hỗ trợ từ chuyên gia, nhằm đảm bảo rằng kết quả nghiên cứu mang tính khoa học, thực tiễn và có giá trị ứng dụng cao.

## **6. Giả thuyết khoa học**

Việc ứng dụng mô hình YOLO trong phát hiện khuôn mặt kết hợp với các phương pháp trích xuất đặc điểm khuôn mặt và nhận diện cảm xúc sẽ giúp cải thiện đáng kể hiệu suất giám sát lớp học, vượt trội so với các hệ thống truyền thống không sử dụng học sâu.

Hiệu suất của hệ thống giám sát lớp học phụ thuộc vào sự kết hợp giữa các bước phát hiện khuôn mặt, trích xuất đặc điểm và nhận diện cảm xúc. Các kỹ thuật tối ưu như Hungarian Matching và sử dụng bộ dữ liệu phù hợp sẽ đảm bảo hệ thống hoạt động chính xác, hiệu quả trong việc theo dõi trạng thái của học sinh theo thời gian.

## **7. Những đóng góp mới của đề tài**

Đề tài đã đề xuất một hệ thống tích hợp YOLO để phát hiện khuôn mặt kết hợp với các phương pháp trích xuất đặc điểm khuôn mặt và nhận diện cảm xúc nhằm giải quyết bài toán giám sát lớp học. Thông qua các thí nghiệm và phân tích chi tiết, nghiên cứu đã chứng minh rằng việc áp dụng YOLO giúp hệ thống phát hiện khuôn mặt nhanh và chính xác, ngay cả trong các điều kiện môi trường lớp học phức tạp. Đồng thời, các kỹ thuật trích xuất đặc điểm và nhận diện cảm xúc dựa trên mạng CNN đã hỗ trợ hiệu quả trong việc phân loại cảm xúc học sinh, từ đó phát hiện các trạng thái như chú ý, mất tập trung hoặc ngủ gật.

Ngoài ra, đề tài cũng đề xuất phương pháp sử dụng thuật toán Hungarian Matching để theo dõi trạng thái học sinh theo thời gian, giúp hệ thống duy trì sự liên tục trong việc giám sát và nâng cao hiệu quả quản lý lớp học. Cuối cùng, nghiên cứu đã cung cấp so sánh cụ thể về hiệu quả của từng bước trong hệ thống, làm sáng tỏ vai trò quan trọng của các thành phần trong việc tối ưu hóa hiệu suất giám sát lớp học. Kết quả nghiên cứu mang lại tiềm năng ứng dụng rộng rãi trong các hệ thống giáo dục thông minh và quản lý lớp học tự động.

## **8. Dự kiến kế hoạch nghiên cứu**

Bảng 8.1 Dự kiến kế hoạch nghiên cứu

|  |  |  |
| --- | --- | --- |
| **STT** | **Nội dung** | **Dự kiến thời gian thực hiện** |
| 1 | Nghiên cứu, chọn đề tài, xây dựng đề cương luận văn | 1 tháng |
| 3 | Nộp đề cương, sửa chữa hoàn thiện đề cương. | 0.5 tháng |
| 4 | Nghiên cứu, viết, hoàn thiện luận văn |  |
| Chương 1: Giới thiệu. | 0.5 tháng |
| Chương 2: Các nghiên cứu liên quan. | 0.5 tháng |
| Chương 3: Thực nghiệm | 2 tháng |
| Chương 4: Kết luận và hướng phát triển | 0.5 tháng |
| 5 | Chỉnh sửa, hoàn thiện luận văn và hồ sơ bảo vệ luận văn. | 1 tháng |

## **9. Dự kiến nội dung của luận văn**

**Chương 1:** Giới thiệu

* Lý do chọn đề tài.
* Mục tiêu và nhiệm vụ nghiên cứu.
* Đối tượng và phạm vi nghiên cứu.
* Phương pháp nghiên cứu.
* Giả thuyết khoa học.
* Những đóng góp mới của đề tài.
* Cấu trúc của luận văn.

**Chương 2**: Cơ sở lý thuyết

* Tổng quan về mô hình YOLO và mạng nơ-ron tích chập (CNN).
* Giới thiệu về thuật toán Hungarian Matching.
* Các phương pháp nhận diện cảm xúc khuôn mặt: từ trích xuất đặc điểm đến phân loại cảm xúc.
* Tổng quan về bài toán giám sát lớp học và các bộ dữ liệu tiêu biểu (Head Dataset, FER+).
* Các chỉ số đánh giá hiệu suất mô hình: Accuracy, Precision, Recall, F1-score.

**Chương 3**: Phương pháp nghiên cứu và mô hình đề xuất

* Quy trình nghiên cứu và các bước thực hiện.
* Tích hợp YOLO vào hệ thống phát hiện khuôn mặt:
* Mô tả cách sử dụng YOLO trên Head Dataset để phát hiện khuôn mặt.
* Xây dựng hệ thống tiền xử lý dữ liệu.
* Trích xuất đặc điểm và nhận diện cảm xúc:
* Sử dụng mạng CNN để phân tích cảm xúc khuôn mặt.
* Các kỹ thuật tối ưu hóa và tích hợp cảm xúc vào hệ thống giám sát.
* Xây dựng ma trận chi phí và thuật toán Hungarian Matching để theo dõi trạng thái học sinh.

**Chương 4**: Thực nghiệm và kết quả

* Thực hiện các thí nghiệm trên tập dữ liệu video lớp học:
* Huấn luyện và kiểm thử hệ thống phát hiện khuôn mặt và nhận diện cảm xúc.
* Đánh giá hiệu suất mô hình thông qua các chỉ số Accuracy, Precision, Recall, và F1-score.
* Phân tích và đánh giá hiệu quả hệ thống:
* So sánh giữa các mô hình có và không tích hợp nhận diện cảm xúc.
* Phân tích cách hệ thống theo dõi trạng thái và hành vi học sinh trong lớp học.
* Trình bày kết quả qua bảng biểu và hình ảnh trực quan.

**Chương 5**: Kết luận và hướng phát triển

* Tóm tắt các kết quả đạt được.
* Đánh giá hiệu quả của hệ thống giám sát lớp học dựa trên YOLO.
* Các hạn chế của nghiên cứu.
* Đề xuất các hướng nghiên cứu và ứng dụng trong tương lai:
* Mở rộng áp dụng hệ thống vào các bài toán khác như nhận diện hành vi trong môi trường công cộng.
* Khám phá các mô hình và thuật toán mới để nâng cao hiệu quả giám sát.

## **10. Danh mục tài liệu tham khảo**

[1] Tracking by Detection of Multiple Faces using SSD and CNN Features 2018