

Hệ Thống Đếm Số Lượng Người Trong Phòng Học Bằng Camera Và Trí Tuệ Nhân Tạo

Đỗ Trường Anh, Lâm Ngọc Tú, Phạm Trọng Toàn, Nguyễn Trung Hiếu

Nhóm 9, Khoa Công Nghệ Thông Tin

Trường Đại Học Đại Nam, Việt Nam

ThS. Nguyễn Văn Nhân, ThS. Lê Trung Hiếu

Giảng viên hướng dẫn, Khoa Công Nghệ Thông Tin

Trường Đại Học Đại Nam, Việt Nam

Abstract—Bài báo này trình bày một hệ thống tự động đếm số lượng người trong phòng học sử dụng camera và các kỹ thuật trí tuệ nhân tạo. Hệ thống thu thập dữ liệu video từ camera trong lớp học, xử lý bằng mô hình YOLOv8 để nhận diện người, theo dõi chuyển động bằng DeepSORT và sử dụng mạng LSTM để dự đoán xu hướng số lượng người theo thời gian. Kết quả cho thấy hệ thống có khả năng giám sát thời gian thực hiệu quả trong môi trường giáo dục.

Index Terms—Đếm người, YOLOv8, DeepSORT, LSTM, nhận diện đối tượng, giám sát thời gian thực.

I. GIỚI THIỆU

Hệ thống đếm người tự động đang trở thành một giải pháp quan trọng trong các môi trường thông minh, đặc biệt là trong các cơ sở giáo dục. Bài báo này đề xuất một phương pháp sử dụng AI để giám sát và đếm số lượng người trong phòng học theo thời gian thực nhằm nâng cao hiệu suất quản lý lớp học.

II. PHƯƠNG PHÁP

A. Nhận Diện Đối Tượng

Mô hình YOLOv8 được sử dụng để phát hiện người trong khung hình với ngưỡng tin cậy 0.5, đảm bảo độ chính xác cao trong các điều kiện ánh sáng khác nhau.

B. Các Phương Pháp Đếm Người

- Đếm dựa trên mật độ khu vực
- Theo dõi chuyển động bằng DeepSORT
- Dự đoán theo chuỗi thời gian với LSTM để phân tích xu hướng

C. Kiến Trúc Hệ Thống

Hệ thống hoạt động theo các bước sau:

- 1) Cài đặt và cấu hình camera giám sát
- 2) Thu thập dữ liệu video từ camera
- 3) Nhận diện đối tượng bằng mô hình YOLOv8
- 4) Theo dõi và đếm người trong khung hình bằng DeepSORT
- 5) Lưu trữ dữ liệu vào MySQL để phân tích và hiển thị kết quả

III. PHÂN TÍCH HOẠT ĐỘNG CỦA MÔ HÌNH

A. Luồng Xử Lý Hệ Thống

Sơ đồ dưới đây minh họa quy trình hoạt động của hệ thống:

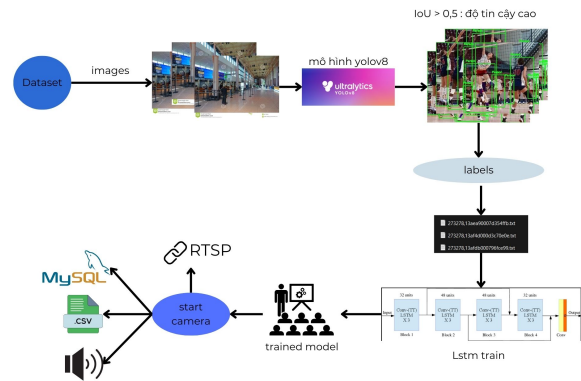


Fig. 1. Sơ đồ hoạt động của hệ thống

B. Các Bước Xử Lý Chi Tiết

- Bước 1: Thu thập dữ liệu từ các dataset có sẵn (kaggle, COCO (Common Objects in Context), Open Images Dataset, People Detection Dataset (github),...).
- Bước 2: Cho các dataset vào một folder để chứa dữ liệu.

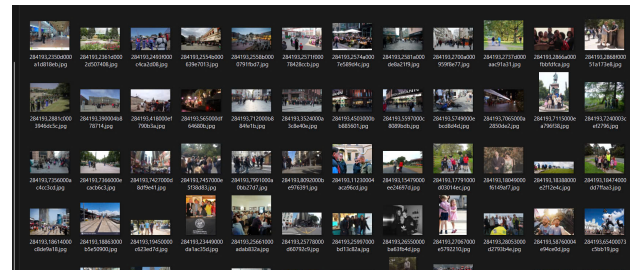


Fig. 2. Folder lưu trữ dataset

- Bước 3: Mô hình YOLOv8 được sử dụng để phát hiện người trong từng khung hình và tự động gán nhãn.

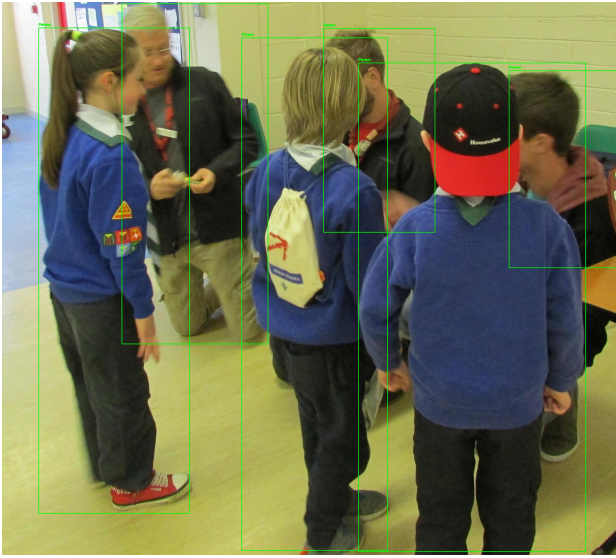


Fig. 3. Tự động gán nhãn

- Bước 4: DeepSORT theo dõi và gán ID cho từng người để tránh trùng lặp khi đếm.
- Bước 5: Dữ liệu thu thập được lưu trữ vào cơ sở dữ liệu để phân tích xu hướng.



Fig. 4. Cơ sở dữ liệu MySQL

- Bước 6: Tiến hành train dữ liệu với mô hình LSTM (Long short-term memory).

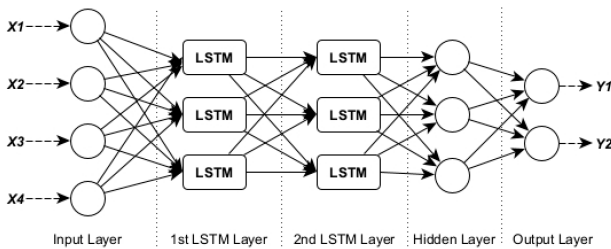


Fig. 5. Mô hình LSTM

- Bước 7: Đặt camera lên một góc độ cao và bắt đầu thực nghiệm hệ thống.



Fig. 6. Nơi đặt camera

IV. PHÂN TÍCH HIỆU SUẤT

Thử nghiệm trên nhiều phòng học khác nhau cho thấy:

- Độ chính xác nhận diện trung bình ($mAP@50$) đạt $> 90\%$
- Sai số trung bình trong đếm người là ± 2 người
- Tốc độ xử lý đạt 30 FPS trên GPU RTX 3060

Hệ thống có thể hoạt động ổn định trong các điều kiện ánh sáng khác nhau, nhưng vẫn bị ảnh hưởng khi có nhiều vật cản lớn trong phòng học.

V. KẾT QUẢ THỰC NGHIỆM

Dưới đây là hình ảnh minh họa kết quả nhận diện và đếm số lượng người:



Fig. 7. Nhận diện người trong phòng học bằng YOLOv8

VI. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

A. Kết Luận

Nghiên cứu này đã đề xuất và triển khai một hệ thống đếm số lượng người trong phòng học dựa trên camera và trí tuệ nhân tạo. Hệ thống sử dụng mô hình **YOLOv8** để nhận diện người, kết hợp với **DeepSORT** và **ByteTrack** để theo dõi chuyển

động, đồng thời ứng dụng **mạng nơ-ron hồi tiếp LSTM** để phân tích xu hướng số lượng người theo thời gian.

Các thử nghiệm thực tế cho thấy hệ thống đạt **độ chính xác nhận diện trung bình (mAP@50) > 90%**, tốc độ xử lý lên đến **30 FPS trên GPU RTX 3060**, sai số trung bình trong quá trình đếm người dao động khoảng ± 2 người. Ngoài ra, hệ thống có thể hoạt động ổn định trong điều kiện ánh sáng thay đổi, nhưng vẫn gặp hạn chế khi có vật cản lớn hoặc nhiều người đứng sát nhau.

Sơ với các phương pháp truyền thống như **cảm biến hồng ngoại** hay **cảm biến trọng lượng**, hệ thống dựa trên AI có ưu điểm vượt trội về **khả năng giám sát thời gian thực, nhận diện chính xác ngay cả khi có nhiều người xuất hiện đồng thời trong khung hình**. Hơn nữa, việc lưu trữ dữ liệu vào **cơ sở dữ liệu MySQL** giúp dễ dàng phân tích và trực quan hóa thông tin, hỗ trợ công tác quản lý lớp học hiệu quả hơn.

B. Hướng Phát Triển

Mặc dù hệ thống đã đạt được những kết quả khả quan, nhưng vẫn còn một số hạn chế cần được cải thiện trong tương lai:

- **Tối ưu hóa mô hình YOLOv8**
 - Điều chỉnh tham số **IoU threshold** và **confidence threshold** để giảm số lượng **false positives** và **false negatives**.
 - Thử nghiệm với **YOLO-NAS** hoặc các mô hình phát hiện đối tượng tiên tiến hơn để cải thiện tốc độ và độ chính xác.
- **Cải thiện thuật toán theo dõi đối tượng**
 - Kết hợp **DeepSORT** với **ByteTrack** để cải thiện khả năng theo dõi đối tượng khi có nhiều người di chuyển nhanh hoặc xuất hiện che khuất nhau.
 - Sử dụng thêm **Kalman Filter** để giảm nhiễu và tránh mất dấu đối tượng trong một số khung hình.
- **Mở rộng ứng dụng của hệ thống**
 - Áp dụng hệ thống vào các môi trường khác như **bệnh viện, trung tâm thương mại, sự kiện đông người** để giám sát số lượng người theo thời gian thực.
 - Kết hợp với **hệ thống quản lý lớp học thông minh** để cảnh báo khi lớp học vượt quá số lượng giới hạn.
- **Cải thiện khả năng dự đoán bằng LSTM**
 - Thu thập dữ liệu dài hạn hơn để cải thiện **độ chính xác dự đoán xu hướng** số lượng người trong phòng học.
 - So sánh hiệu suất của **LSTM, GRU, Transformer** để chọn mô hình tối ưu nhất.
- **Phát triển giao diện trực quan**
 - Xây dựng **dashboard web** hoặc **ứng dụng di động** để hiển thị dữ liệu theo thời gian thực.
 - Cung cấp **biểu đồ, thống kê chi tiết**, giúp người dùng dễ dàng theo dõi tình trạng lớp học.

Việc thực hiện các cải tiến trên không chỉ giúp nâng cao hiệu suất hệ thống mà còn mở rộng phạm vi ứng dụng, tạo nền tảng cho việc phát triển các hệ thống **giám sát thông minh trong tương lai**.

VII. TÀI LIỆU THAM KHẢO

REFERENCES

- [1] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *arXiv preprint arXiv:1804.02767*, 2018.
- [2] G. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv preprint arXiv:2004.10934*, 2020.
- [3] X. Zhang, Y. Wang, J. Zhu, et al., "ByteTrack: Multi-Object Tracking by Associating Every Detection Box," *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.
- [4] N. Wojke, A. Bewley, and D. Paulus, "Simple Online and Realtime Tracker with a Deep Association Metric," *arXiv preprint arXiv:1703.07402*, 2017.
- [5] A. Vaswani, N. Shazeer, et al., "Attention is All You Need," *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.
- [6] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.
- [7] T. Lin, P. Goyal, et al., "Focal Loss for Dense Object Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2020.
- [8] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [9] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems (NeurIPS)*, 2012.
- [10] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv preprint arXiv:1409.1556*, 2014.