

ĐỀ CƯƠNG CHI TIẾT**TÊN ĐỀ TÀI: PHÁT HIỆN SỚM HÀNH VI BẤT THƯỜNG CỦA CON NGƯỜI
TỪ CHUỖI CHUYỂN ĐỘNG KHUNG XƯƠNG TRONG VIDEO GIÁM SÁT****TÊN TIẾNG ANH: EARLY DETECTION OF ABNORMAL HUMAN
BEHAVIOR FROM SKELETON MOTION SEQUENCES IN SURVEILLANCE
VIDEOS****Cán bộ hướng dẫn:** PGS.TS Nguyễn Đình Thuân**Thời gian thực hiện:** Từ ngày 26/01/2026 đến ngày 30/05/2026**Sinh viên thực hiện:**

Nguyễn Thị Hải Hà – 22520372

Nguyễn Thị Thanh Hiếu – 22520443

Nội dung đề tài:**• Tình hình chung**

Trong bối cảnh đô thị hóa hiện nay, giám sát an ninh công cộng là một trong những lĩnh vực trọng yếu ứng dụng công nghệ thị giác máy tính, đặc biệt tại các không gian tập trung đông người như trung tâm thương mại, nhà ga, trường học và khu dân cư. Trong đó, nhiệm vụ phát hiện bất thường trong video đóng vai trò then chốt nhằm nhận diện kịp thời các sự kiện hoặc hành vi của con người có khả năng gây nguy hiểm cho an toàn xã hội [1]. Dù là hành vi bạo lực hay phản ứng trước sự cố thì các hành vi này đều

có sự thay đổi bất ngờ về dáng điệu hoặc quỹ đạo di chuyển trong một khoảng thời gian ngắn [1, 4]. Do đó, việc xây dựng hệ thống có khả năng phát hiện sớm các hành vi bất thường nhằm hỗ trợ giám sát chủ động, giảm tải cho nhân viên an ninh và tăng khả năng phản ứng kịp thời là yêu cầu cấp thiết trong bối cảnh hiện nay.

Tuy nhiên, các phương pháp giám sát truyền thống dựa trên phân tích hình ảnh tĩnh hoặc xử lý toàn bộ dữ liệu pixel đang bộc lộ nhiều hạn chế. Trong thực tế, các hệ thống phân tích video hiện hành thường phụ thuộc vào mô hình học sâu dựa trên dữ liệu RGB đòi hỏi tài nguyên tính toán lớn và không ổn định khi điều kiện môi trường thay đổi. Về mặt lý thuyết, việc cố gắng nhận diện hành vi chỉ dựa trên ngoại hình được xem là một bài toán khó xác định rõ ràng do mô hình chịu ảnh hưởng lớn từ các yếu tố nhiễu ngoại cảnh như thay đổi góc quay, điều kiện ánh sáng và sự phức tạp của phong nền [2]. Ngoài ra, việc gán nhãn hành vi trong video cũng tốn kém và không khả thi ở quy mô lớn, khiến mô hình khó học được ranh giới chính xác giữa hành vi bình thường và bất thường. Bên cạnh đó, các đặc trưng dựa trên pixel thường chứa nhiều thông tin dư thừa và nhiễu, gây gánh nặng cho quá trình tính toán và làm giảm khả năng phân biệt giữa tín hiệu của hành vi thực sự và nhiễu của môi trường dễ dẫn đến hiện tượng cảnh báo sai trong môi trường thực tế.

Trước những hạn chế đó, hướng nghiên cứu phân tích chuyển động của con người thông qua biểu diễn khung xương đang trở thành xu hướng nổi bật trong lĩnh vực phát hiện hành vi bất thường. Thay vì phụ thuộc vào chi tiết hình ảnh, phương pháp này sử dụng các điểm khớp (keypoints) [5] của cơ thể người để mô hình hóa chuỗi chuyển động khung xương thông qua việc phân tích chuyển động của con người theo thời gian, giúp giảm thiểu ảnh hưởng của nhiễu thị giác và đảm bảo tính ẩn danh. Các nghiên cứu gần đây đã chứng minh rằng các mô hình skeleton-based không chỉ đạt độ chính xác cao hơn trong môi trường phức tạp mà còn có thể huấn luyện nhanh, yêu cầu ít tài nguyên và hoạt động ổn định ở thời gian thực, cho phép mô hình nắm bắt được ngữ nghĩa hành vi một cách trực quan và rõ ràng hơn [1, 2, 4].

Mặc dù đạt được nhiều tiến bộ, hướng tiếp cận dựa trên chuyển động vẫn đối mặt

với những thách thức đáng kể. Cụ thể, hiệu quả của mô hình phụ thuộc chặt chẽ vào độ chính xác của bộ trích xuất khung xương và khả năng mô hình hóa ngữ cảnh theo thời gian dài hạn. Trong các cảnh đông người hoặc có nhiều đối tượng di chuyển chồng lấn, việc nhận dạng và theo dõi khung xương thường bị lỗi, dẫn đến chuỗi chuyển động bị ngắt quãng hoặc sai lệch [1, 4]. Bên cạnh đó, ranh giới giữa hành vi bình thường và bất thường đôi khi rất mong manh và phụ thuộc chặt chẽ vào ngữ cảnh diễn ra trong một chuỗi thời gian dài, cùng một hành vi có thể được xem là bình thường trong hoàn cảnh này nhưng bất thường trong hoàn cảnh khác [3]. Do đó, việc phát triển mô hình có khả năng học thích nghi và duy trì độ tin cậy khi dữ liệu đầu vào không hoàn hảo, đồng thời phân biệt chính xác các biến thể chuyển động tự nhiên của con người vẫn là một bài toán mở cần tiếp tục được nghiên cứu.

Xuất phát từ thực tế trên, nghiên cứu này đề xuất một phương pháp phát hiện hành vi bất thường trong video giám sát dựa trên chuỗi chuyển động khung xương của con người. Phương pháp tập trung vào việc mô hình hóa quy luật chuyển động của cơ thể theo thời gian để phát hiện các sai lệch so với hành vi bình thường đã học, hướng tới khả năng phát hiện sớm, chính xác và tiết kiệm tài nguyên tính toán. Cách tiếp cận này không chỉ giảm sự phụ thuộc vào ngoại hình và điều kiện môi trường mà còn phù hợp để triển khai trong các hệ thống giám sát an ninh thông minh quy mô lớn.

- **Phát biểu bài toán**

Đầu vào của bài toán là chuỗi dữ liệu video giám sát trong không gian công cộng, trong đó ghi nhận sự xuất hiện và chuyển động của con người theo thời gian. Dữ liệu video ban đầu được tiền xử lý và trích xuất các điểm khớp (keypoints) của cơ thể con người, qua đó biểu diễn tư thế (pose) của con người tại từng thời điểm quan sát.

Đầu ra của bài toán là mức độ bất thường của hành vi theo thời gian, cho phép hệ thống phát hiện sớm các hành vi có nguy cơ và hỗ trợ đưa ra cảnh báo kịp thời trong môi trường giám sát an ninh công cộng.

- **Mục tiêu:**

- Phát triển mô hình và pipeline phát hiện sớm hành vi bất thường của con người trong không gian công cộng dựa trên dữ liệu video giám sát.
- Nghiên cứu, phân tích và đánh giá hiệu quả của một số mô hình học máy và học sâu tiêu biểu trong bài toán phát hiện hành vi bất thường thông qua biểu diễn chuyển động của con người theo thời gian.
- Đề xuất và thực nghiệm cải tiến trong mô hình nhằm nâng cao độ chính xác, khả năng phát hiện sớm và tính ổn định của hệ thống.

• **Đối tượng và phạm vi nghiên cứu:**

- **Đối tượng:** Hành vi của con người trong video giám sát tại các khu vực công cộng, tập trung vào việc nhận diện và phát hiện sớm các hành vi có dấu hiệu bất thường so với hành vi bình thường.
- **Dữ liệu:** tập trung vào dữ liệu video giám sát trong không gian công cộng có chứa hoạt động của con người (hành vi bình thường và bất thường).
 - + Các bộ dữ liệu video có chứa hành vi bất thường công khai như UBnormal [6].
 - + Các bộ dữ liệu video về hành vi bình thường công khai như ShanghaiTech [7] và CUHK Avenue [8].
- **Phạm vi nghiên cứu:**
 - + Khảo sát các công trình nghiên cứu liên quan đến nhận dạng hành động và phát hiện hành vi bất thường trong video, ưu tiên các hướng tiếp cận dựa trên phân tích chuyển động của con người theo thời gian, trong đó có các phương pháp dựa trên khung xương người.
 - + Nghiên cứu và áp dụng các phương pháp tiền xử lý và trích xuất đặc trưng từ dữ liệu video, bao gồm các kỹ thuật biểu diễn chuyển động của con người như khung xương.
 - + Nghiên cứu, phát triển và huấn luyện các mô hình học sâu cho phát hiện hành vi bất thường, tiêu biểu như SeeKer – Sequential Keypoint Density Estimator [1], LSTM hoặc Autoencoder.

- + Thực nghiệm và đánh giá mô hình bằng độ đo chuyên biệt như AUROC nhằm đo lường hiệu quả phát hiện hành vi bất thường của mô hình.

- **Thách thức:**

- **Về dữ liệu:**

- + Dữ liệu video giám sát trong môi trường công cộng thường có chất lượng không đồng đều, chịu ảnh hưởng của nhiều yếu tố như điều kiện ánh sáng, góc quay camera, độ phân giải và hiện tượng che khuất, gây khó khăn cho việc trích xuất và biểu diễn chính xác hành vi của con người.
 - + Các hành vi bất thường thường xuất hiện với tần suất thấp so với hành vi bình thường, dẫn đến sự mất cân bằng dữ liệu, gây ảnh hưởng đến khả năng học và đánh giá của mô hình.
 - + Dữ liệu video giám sát có thể chứa nhiều hoặc các hành vi nằm gần ranh giới giữa hành vi bình thường và bất thường, gây khó khăn cho quá trình huấn luyện và đánh giá mô hình
 - + Việc gán nhãn dữ liệu hành vi bất thường trong video đòi hỏi nhiều công sức và độ chính xác cao.

- **Về phương pháp:**

- + Trong bối cảnh các mô hình phát hiện hành vi bất thường chủ yếu dựa trên việc học các mẫu hành vi bình thường từ dữ liệu huấn luyện, hệ thống cần có khả năng nhận diện hiệu quả các sai lệch so với hành vi đã học.
 - + Các mô hình phân tích chuyển động của con người theo thời gian cần xử lý tốt dữ liệu chuỗi, đồng thời đảm bảo tính ổn định và khả năng tổng quát khi đối mặt với sự đa dạng về hình dáng cơ thể, tốc độ chuyển động và kiểu hành vi.
 - + Việc khai thác và kết hợp thông tin không gian – thời gian một cách hiệu quả trong video là vấn đề phức tạp, đòi hỏi thiết kế mô hình và pipeline xử lý phù hợp.

- **Về tính ứng dụng thực tiễn:**

- + Hệ thống phát hiện hành vi bất thường cần đạt hiệu năng xử lý đủ nhanh để đáp ứng yêu cầu giám sát gần thời gian thực trong các môi trường triển khai thực tế.
- + Dữ liệu video giám sát liên quan trực tiếp đến con người, do đó hệ thống cần xem xét và đảm bảo các yêu cầu về quyền riêng tư trong quá trình thu thập và xử lý dữ liệu.
- + Việc triển khai thực tế đòi hỏi hạ tầng tính toán có khả năng tích hợp với hệ thống giám sát hiện hữu, cũng như đảm bảo độ tin cậy và ổn định khi hoạt động trong các điều kiện đa dạng của môi trường công cộng.

- **Kết quả mong đợi:**

- Xây dựng và đánh giá được mô hình học sâu có khả năng phát hiện sớm hành vi bất thường của con người trong video giám sát, dựa trên các biểu diễn chuyển động theo thời gian phù hợp với dữ liệu video trong không gian công cộng.
- Hoàn thiện pipeline xử lý dữ liệu video cho bài toán phát hiện hành vi bất thường, bao gồm các bước tiền xử lý, trích xuất biểu diễn chuyển động, mô hình hóa chuỗi thời gian và tính toán điểm bất thường cho từng đoạn video hoặc khung hình.
- Thực hiện đánh giá hiệu năng phát hiện hành vi bất thường trên các bộ dữ liệu tiêu biểu như UBnormal, ShanghaiTech và CUHK Avenue thông qua các độ đo phù hợp.
- Phân tích và so sánh kết quả thực nghiệm giữa các hướng tiếp cận biểu diễn chuyển động khác nhau, qua đó làm rõ đặc điểm, ưu điểm và hạn chế của từng hướng tiếp cận trong bối cảnh phát hiện hành vi bất thường từ dữ liệu video giám sát.
- Rút ra các nhận xét và đánh giá tổng quan từ kết quả thực nghiệm, làm cơ sở cho việc định hướng lựa chọn mô hình và pipeline phù hợp cho bài toán phát hiện hành vi bất thường trong phạm vi nghiên cứu của đề tài.

Tài liệu tham khảo:

- [1] Delić, A., Grcic, M., & Šegvić, S. (2025). Sequential keypoint density estimator: an overlooked baseline of skeleton-based video anomaly detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 11579-11589).
- [2] Hirschorn, O., & Avidan, S. (2023). Normalizing flows for human pose anomaly detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 13545-13554).
- [3] Shen, G., Ouyang, Y., Lu, J., Yang, Y., & Sanchez, V. (2024). Advancing video anomaly detection: a bi-directional hybrid framework for enhanced single-and multi-task approaches. IEEE Transactions on Image Processing.
- [4] Morais, R., Le, V., Tran, T., Saha, B., Mansour, M., & Venkatesh, S. (2019). Learning regularity in skeleton trajectories for anomaly detection in videos. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 11996-12004).
- [5] Tahir, R., & Cai, Y. (2025). Multi-human pose estimation by deep learning-based sequential approach for human keypoint position and human body detection. Journal of Shanghai Jiaotong University (Science), 30(6), 1103-1113.
- [6] Acsintoae, A., Florescu, A., Georgescu, M. I., Mare, T., Sumedrea, P., Ionescu, R. T., ... & Shah, M. (2022). Ubnormal: New benchmark for supervised open-set video anomaly detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 20143-20153).
- [7] Yu, S., Zhao, Z., Fang, H., Deng, A., Su, H., Wang, D., ... & Wu, W. (2023). Regularity learning via explicit distribution modeling for skeletal

video anomaly detection. IEEE Transactions on Circuits and Systems for Video Technology.

- [8] Chen, W., Liu, Z., Zhao, H., Zhou, S., Li, H., & Liu, Y. H. (2020, October). CUHK-AHU dataset: Promoting practical self-driving applications in the complex airport logistics, Hill and urban environments. In 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 4283-4288). IEEE.

Kế hoạch thực hiện:

- **Kế hoạch làm việc**

STT	Thời gian	Mô tả công việc
1	Tuần 1 (26/1/2026 – 1/2/2026)	<ul style="list-style-type: none">– Tìm hiểu cơ sở lý thuyết về học máy và học sâu liên quan đến phân tích chuỗi dữ liệu và video.– Khảo sát các kỹ thuật tiền xử lý dữ liệu và các dạng biểu diễn chuyển động của con người.
2	Tuần 2 – 4 (2/2/2026 – 15/3/2026)	<ul style="list-style-type: none">– Nghiên cứu và phân tích các bộ dữ liệu tiêu biểu như ShanghaiTech, UBnormal và CUHK Avenue.– Thực hiện tiền xử lý và chuẩn hóa dữ liệu cho bài toán học hành vi bình thường.
3	Tuần 5 (16/3/2026 – 22/3/2026)	<ul style="list-style-type: none">– Khảo sát một số phương pháp phát hiện hành vi bất thường hiện có.– Lựa chọn hướng tiếp cận và mô hình phù hợp để triển khai trong phạm vi đề tài.
4	Tuần 6 – 8	<ul style="list-style-type: none">– Thiết kế pipeline tổng thể cho bài toán phát

	(23/3/2026 – 5/4/2026)	<p>hiện hành vi bất thường dựa trên dữ liệu video.</p> <ul style="list-style-type: none"> – Cài đặt và xây dựng kiến trúc mô hình học sâu ban đầu.
5	<p>Tuần 9 – 11</p> <p>(06/4/2026 – 19/4/2026)</p>	<ul style="list-style-type: none"> – Huấn luyện mô hình trên dữ liệu hành vi bình thường hoặc dữ liệu huấn luyện phù hợp theo hướng tiếp cận đã chọn. – Điều chỉnh siêu tham số và tối ưu hiệu suất mô hình.
6	<p>Tuần 12</p> <p>(20/4/2026 – 26/4/2026)</p>	<ul style="list-style-type: none"> – Thử nghiệm phát hiện hành vi bất thường trên các tập dữ liệu kiểm thử. – Tính toán các chỉ số đánh giá.
7	<p>Tuần 13 – 15</p> <p>(27/4/2026 – 10/5/2026)</p>	<ul style="list-style-type: none"> – So sánh và phân tích kết quả với các mô hình tham khảo. – Đề xuất các cải tiến nhằm nâng cao độ chính xác hoặc khả năng phát hiện sớm.
8	<p>Tuần 16</p> <p>(11/5/2026 – 17/5/2026)</p>	<ul style="list-style-type: none"> – Tổng hợp kết quả thực nghiệm, rút ra nhận xét và đánh giá. – Viết báo cáo kết quả và tài liệu kỹ thuật cho mô hình. – Kiểm tra, hoàn thiện toàn bộ khóa luận và tài liệu kèm theo.

• **Phân công công việc:**

Công việc	Hải Hà	Thanh Hiếu
Tìm hiểu cơ sở lý thuyết về học máy, học sâu và các hướng tiếp cận phát hiện hành vi bất thường trong	50%	50%

video giám sát		
Khảo sát và tiền xử lý các bộ dữ liệu công khai (ShanghaiTech, UBnormal, CUHK Avenue)	50%	50%
Thiết kế pipeline tổng thể và lựa chọn hướng tiếp cận, mô hình học sâu phù hợp cho bài toán phát hiện hành vi bất thường	50%	50%
Cài đặt và huấn luyện mô hình trên dữ liệu đã xử lý	50%	50%
Thử nghiệm, đánh giá mô hình	50%	50%
So sánh, phân tích kết quả thực nghiệm và đề xuất các điều chỉnh hoặc cải tiến trong phạm vi đề tài	50%	50%
Viết báo cáo khóa luận, trình bày kết quả và chuẩn bị slide bảo vệ	50%	50%
Xác nhận của CBHD (Ký tên và ghi rõ họ tên) PGS. TS Nguyễn Đình Thuần	TP. HCM, ngày 06 tháng 02 năm 2026 Sinh viên (Ký tên và ghi rõ họ tên) Nguyễn Thị Hải Hà Nguyễn Thị Thanh Hiếu	