THÔNG TIN CHUNG CỦA NHÓM

- Link YouTube video của báo cáo (tối đa 5 phút): https://www.youtube.com/watch?v=UETgBJtTgFc
- Link slides (dạng .pdf đặt trên Github của nhóm): Thach Nguyễn Ngoc/CS2205.FEB2025/DeCuong.FinalReport.Template.Slide.pdf

- Họ và Tên: Nguyễn Ngọc Thạch
- MSSV: 240101072

- Lớp: CS2205.FEB2025
- Tự đánh giá (điểm tổng kết môn): 8/10
- Số buổi vắng: 1
- Số câu hỏi QT cá nhân: 3
- Số câu hỏi QT của cả nhóm: 3
- Link Github: https://github.com/NguyenNgocThach2001/CS2205.FEB2025

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI (IN HOA)

MÔ PHỎNG ĐÁM ĐÔNG VỚI HỌC TĂNG CƯỜNG CÓ HƯỚNG DẪN

TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

CROWD SIMULATION WITH GUIDED REINFORCEMENT LEARNING

TÓM TẮT *(Tối đa 400 từ)*

Đề tài nghiên cứu ứng dụng học tăng cường có hướng dẫn trong mô phỏng hành vi đám đông, sử dụng dữ liệu trajectory thu thập từ video thực tế. Bằng cách huấn luyện agent thông qua thuật toán PPO với shaping reward được thiết kế từ hành vi người thật, mô hình có thể sinh ra các hành vi di chuyển tự nhiên như đi theo nhóm, tránh va chạm, hoặc thay đổi hướng linh hoạt. Mục tiêu của nghiên cứu là xây dựng pipeline xử lý dữ liệu, thiết lập môi trường mô phỏng, đánh giá hiệu quả hành vi học được, và thử nghiệm trên các kịch bản thực tế như phố đi bộ hoặc sự kiện ngoài trời. Mô hình có tiềm năng được triển khai trong lĩnh vực như trò chơi điện tử, phim hoạt hình hoặc diễn hoạt kỹ thuật số, nơi cần các hành vi đám đông chân thực và sống động. Kết quả mong đợi là mô hình có thể tổng quát hóa tốt, tạo tiền đề cho các ứng dụng mô phỏng đám đông trong đô thị thông minh, an ninh, hoặc truyền thông số.

GIỚI THIỆU *(Tối đa 1 trang A4)*

Mô phỏng hành vi đám đông là một bài toán quan trọng trong các lĩnh vực như phim ảnh, trò chơi, kiến trúc, an toàn công cộng và giao thông. Tuy nhiên, việc tái hiện các hành vi đa dạng của con người như đi theo nhóm, đổi hướng bất ngờ, do dự, hoặc tránh va chạm là một thách thức khi chỉ dựa vào các quy tắc thủ công.

Hướng tiếp cận sử dụng học tăng cường có hướng dẫn (guided reinforcement learning) cho phép mô hình học chính sách hành vi từ dữ liệu trajectory người thật. Thay vì học hoàn toàn bằng thử sai, agent được định hướng bởi phần thưởng được thiết kế từ dữ liệu thực tế, giúp hành vi trở nên tự nhiên và ổn định hơn trong môi trường mô phỏng. Mô hình được kỳ vọng có thể thích ứng tốt với các tình huống chưa

từng thấy trong quá trình huấn luyện.

Các hướng tiếp cận mô phỏng đám đông trước đây thường dựa vào quy tắc hoặc dữ liệu. Ví dụ, PAG (Parametric Action Graph) tổ chức hành vi thành biểu đồ hành động có tham số để mô phỏng theo kịch bản cụ thể, phù hợp cho hoạt cảnh trong phim và game [1]. Trong khi đó, mô hình học sâu như Social GAN lại sử dụng mạng đối kháng để sinh trajectory mang tính xã hội, học được hành vi tự nhiên nhưng thiếu tính tương tác thời gian thực [2].

Đề tài này tập trung xây dựng pipeline mô phỏng hành vi đám đông từ dữ liệu video, trích xuất trajectory và huấn luyện agent bằng thuật toán PPO [3] kết hợp với tín hiệu hướng dẫn từ dữ liệu. Việc shaping reward dựa trên phân phối trạng thái người thật giúp quá trình học có định hướng rõ ràng, là đặc trưng trong hướng tiếp cận guided reinforcement learning [4]. Mô hình được kỳ vọng có thể sinh hành vi tự nhiên cho lĩnh vực như trò chơi điện tử, phim hoạt hình hoặc diễn hoạt kỹ thuật số, nơi cần hành vi đám đông sống động, không lặp lại. Qua đó, mong muốn đánh giá được tính khả thi, hiệu quả và khả năng áp dụng mô hình để tái hiện lại đặc trưng con người. Có thể là mô phỏng lại đường phố, khu du lịch ở Việt Nam (nếu có đủ dữ liệu).

Input và Output của bài toán:

- Input: tập hợp video ngắn có môi trường đa dạng (chuỗi vị trí theo thời gian).
- Output: chính sách (policy) sinh hành vi di chuyển tự nhiên cho agent, thể hiện qua vector hành động tại mỗi bước thời gian.

MỤC TIÊU *(Viết trong vòng 3 mục tiêu)*

- Tìm hiểu cơ sở lý thuyết và nguyên lý hoạt động của học tăng cường có hướng dẫn (guided reinforcement learning), đặc biệt là thuật toán PPO và shaping reward từ dữ liệu thật.
- Xây dựng hoàn chỉnh bao gồm: nhận video thực tế, tracking người, trích xuất trajectory, thiết kế môi trường mô phỏng, và huấn luyện agent.
- Đánh giá hiệu quả của mô hình dựa trên các chỉ số định lượng (ADE, FDE) và định tính (tính tự nhiên, khả năng tương tác xã hội, phản ứng linh hoạt).

NỘI DUNG VÀ PHƯƠNG PHÁP

1. Nội dung:

- Tìm hiểu khái niệm học tăng cường có hướng dẫn (Guided Reinforcement Learning), trong đó agent học thông qua tương tác với môi trường nhưng được định hướng bởi phần thưởng thiết kế dựa trên dữ liệu hành vi thật. Sử dụng thuật toán PPO để học chính sách hành vi. Phân tích đặc điểm hành vi đám đông như giữ nhóm, tránh va chạm, thay đổi hướng linh hoạt.
- Bài toán mô phỏng hành vi đám đông:
    + Input: trạng thái môi trường (vị trí, vận tốc, mục tiêu, vai trò, khoảng cách đến người khác).
    + Output: hành động tiếp theo (vector hướng và quãng đường).
    + Mạng học: Multi-layer Perceptron (MLP).
    + Hành vi học được: điều hướng tự nhiên, giữ nhóm, tránh va chạm, thay đổi hướng linh hoạt.
- Dữ liệu huấn luyện:
    + ETH: một tập dữ liệu công khai gồm các video được quay từ góc nhìn bird's eye view tại ETH Zurich, Thụy Sĩ. Được chú thích thủ công với tất cả video bao gồm người đi bộ một mình cũng như người đi bộ di chuyển theo nhóm.
    + UCY: một tập dữ liệu công khai gồm các video được quay từ góc nhìn trên cao tại University of Cyprus. Được chú thích thủ công với tất cả video bao gồm người đi bộ một mình cũng như người đi bộ di chuyển theo nhóm.
    + Dù 2 bộ dữ liệu trên đã có sẵn trajectory nhưng nghiên cứu bao gồm học hành vi đám đông từ video input nên cần xây dựng pipeline trích xuất trajectory.
- Pipeline trích xuất trajectory:
    + Sử dụng mô hình YOLO để gán bounding box cho người đi đường.
    + Sử dụng DeepSORT để gán ID và theo dõi người qua chuỗi thời gian.
    + Làm sạch dữ liệu, loại bỏ track ngắn, mờ, mất ID…
    + Tạo trajectory phục vụ reward shaping.

2. Phương pháp:

- Thiết kế môi trường mô phỏng:
  + Môi trường 3D đơn giản, mỗi agent di chuyển trong không gian có các agent khác. Sử dụng Unity Engine cho environment, và OpenAI Gym cho cầu nối.
  + Trạng thái: gồm vị trí, vận tốc, mục tiêu, các cá thể lân cận.
  + Hành động: di chuyển theo vector chỉ hướng và độ lớn mỗi bước thời gian.
  + Mục tiêu agent: di chuyển đến điểm đích.
- Huấn luyện chính sách với PPO:
  + Chính sách nhận trạng thái (state) gồm vị trí bản thân, người xung quanh, mục tiêu.
  + Hành động là vector hướng và độ lớn (continuous action).
  + Thư viện sử dụng để huấn luyện là ml-agents, được triển khai cho Unity Engine sử dụng Pytorch.
- Reward Shaping:
  + Thay vì phần thưởng cố định, agent được thưởng càng lớn nếu trajectory sinh ra từ chính sách càng giống người.
  + Mô hình không cần biết mục tiêu thật, chỉ học từ reward được tạo bởi trajectory từ video đám đông thực tế.
- Đánh giá mô hình:
  + Định lượng: sử dụng các chỉ số sai số như ADE (Average Displacement Error), FDE (Final Displacement Error) để đo mức độ gần đúng với hành vi thật.
  + Định tính: quan sát khả năng giữ nhóm, tránh va chạm, di chuyển tự nhiên và phản ứng khi thay đổi môi trường.
  + So sánh với baseline: mô hình rule-based, Social GAN.

KẾT QUẢ MONG ĐỢI

- Mô hình học tăng cường có hướng dẫn sẽ học được chính sách điều hướng tự nhiên, phản ánh được các đặc trưng hành vi người thật như tránh va chạm, giữ nhóm, thay đổi hướng linh hoạt.

- Agent có thể sinh hành vi tự nhiên trong các môi trường đông đúc, không cần gán cứng hành vi hay luật điều khiển thủ công.
- Mô hình hoạt động ổn định khi tăng số lượng agent, và có thể tổng quát hóa hành vi với các tình huống chưa từng thấy trong huấn luyện.
- Pipeline có thể được sử dụng để tái hiện đám đông từ video.

TÀI LIỆU THAM KHẢO *(Định dạng DBLP)*

[1]. Panayiotis Charalambous and Yiorgos Chrysanthou. The PAG Crowd: A Graph Based Approach for Efficient Data-Driven Crowd Simulation. In Computer Graphics Forum, Vol. 33, No. 2, pp. 95–108, 2014.

[2]. Agrim Gupta, Justin Johnson, Li Fei-Fei, Silvio Savarese, and Alexandre Alahi. Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks. arXiv preprint arXiv:1803.10892, 2018.

[3]. J. Eßer, N. Bach, C. Jestel, O. Urbann, and S. Kerner. Guided Reinforcement Learning: A Review and Evaluation for Efficient and Effective Real-World Robotics. In IEEE Robotics and Automation Magazine, Vol. 30, No. 1, pp. 74–85, 2023.

[4]. John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. "Proximal Policy Optimization Algorithms." arXiv preprint arXiv:1707.06347, 2017.

[5]. Donghun Lee, Hyunseok Lee, Jaesik Park, Young Min Choi, and Seungjin Oh. Crowd Simulation by Deep Reinforcement Learning. Technical report, MRL, Seoul National University, 2018.

[6]. Panayiotis Charalambous. GREIL-Crowds: Crowd Simulation with Deep Reinforcement Learning and Examples. In ACM Transactions on Graphics (Proc. SIGGRAPH), Vol. 42, No. 4, Article 142, 2023.