



# THÔNG TIN CHUNG CỦA NHÓM

- Link YouTube video của báo cáo: <https://www.youtube.com/watch?v=6rFDN0k93BI>
- Link slides: [https://github.com/Truong99zvc/CS519.P11/blob/main/CS519\\_Proposal.pdf](https://github.com/Truong99zvc/CS519.P11/blob/main/CS519_Proposal.pdf)

<ul style="list-style-type: none"><li>● Họ và Tên: Trương Phúc Trường</li><li>● MSSV: 22521587</li></ul> 	<ul style="list-style-type: none"><li>● Lớp: CS519.P11</li><li>● Tự đánh giá (điểm tổng kết môn): 8.5/10</li><li>● Số buổi vắng: 1</li><li>● Số câu hỏi QT cá nhân: 6</li><li>● Link Github: <a href="https://github.com/Truong99zvc/CS519.P11">https://github.com/Truong99zvc/CS519.P11</a></li><li>● Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:<ul style="list-style-type: none"><li>○ Lên ý tưởng đề tài nghiên cứu</li><li>○ Chỉnh sửa phần tóm tắt, giới thiệu. Viết phần nội dung và phương pháp, mục tiêu, kết quả mong đợi. Làm Poster. Chỉnh sửa Slide.</li><li>○ Kiểm tra và chỉnh sửa video YouTube</li></ul></li></ul>
<ul style="list-style-type: none"><li>● Họ và Tên: Võ Đình Trung</li><li>● MSSV: 22521571</li></ul> 	<ul style="list-style-type: none"><li>● Lớp: CS519.P11</li><li>● Tự đánh giá (điểm tổng kết môn): 8.5/10</li><li>● Số buổi vắng: 1</li><li>● Số câu hỏi QT cá nhân: 6</li><li>● Link Github: <a href="https://github.com/votrung654/CS519.P11">https://github.com/votrung654/CS519.P11</a></li><li>● Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:<ul style="list-style-type: none"><li>○ Lên ý tưởng đề tài nghiên cứu</li><li>○ Chỉnh sửa phần nội dung và phương pháp. Viết phần tóm tắt, giới thiệu. Làm Slide. Làm Poster</li><li>○ Quay video YouTube</li></ul></li></ul>

# ĐỀ CƯƠNG NGHIÊN CỨU

## TÊN ĐỀ TÀI

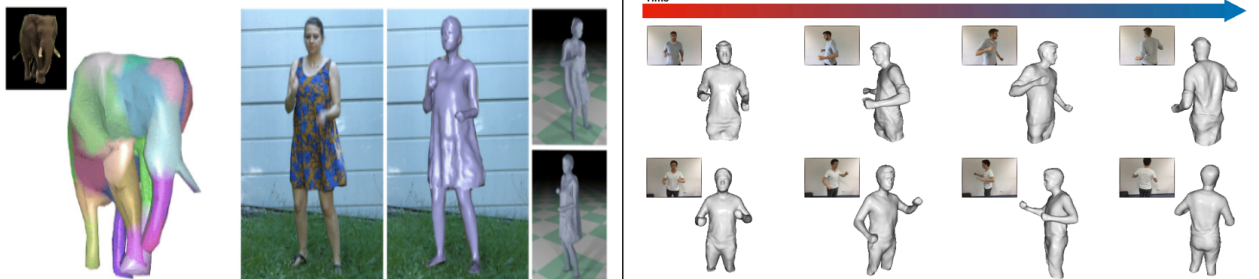
MỞ RỘNG KHẢ NĂNG TÁI TẠO 3D THỜI GIAN THỰC CỦA NEURALRECON TRONG MÔI TRƯỜNG ĐỘNG

## TÊN ĐỀ TÀI TIẾNG ANH

EXTENDING NEURALRECON FOR REAL-TIME 3D RECONSTRUCTION IN DYNAMIC ENVIRONMENTS

## TÓM TẮT

Nghiên cứu này đề xuất một phương pháp tái tạo 3D thời gian thực từ video đơn ảnh với khả năng xử lý cả môi trường tĩnh và các đối tượng động. Phương pháp này mở rộng “NeuralRecon” [1], vốn tập trung chủ yếu vào môi trường tĩnh, bằng cách tích hợp module phát hiện và theo dõi đối tượng (object tracking) [2], kết hợp với tái tạo 3D chuyên biệt cho các đối tượng động. Các kết quả tái tạo này được hòa trộn vào mô hình môi trường tĩnh, đảm bảo sự chính xác và nhất quán trong cả môi trường động. Phương pháp sử dụng các mô hình học sâu được huấn luyện trước, sau đó tinh chỉnh để phù hợp với kiến trúc của “NeuralRecon”. Hiệu quả của giải pháp sẽ được đánh giá trên các bộ dữ liệu chứa đối tượng động, so sánh trực tiếp với “NeuralRecon” gốc. Kỳ vọng rằng phương pháp đề xuất sẽ cải thiện đáng kể khả năng tái tạo 3D trong môi trường động, mở ra tiềm năng ứng dụng trong các lĩnh vực AR/VR, robot và giám sát.



**Hình 1.** Tái tạo 3D từ video đơn ảnh ở môi trường tĩnh và môi trường động

## GIỚI THIỆU

Công nghệ tái tạo 3D từ video đơn ảnh đã đạt những bước tiến đáng kể, mở ra tiềm năng ứng dụng trong AR, VR, robot tự hành, thiết kế công nghiệp và giám sát. NeuralRecon nổi lên như một giải pháp tiên tiến, cho phép tái tạo 3D chính xác và thời gian thực nhờ mạng học sâu để học và hợp nhất các đặc trưng hình học. Tuy nhiên, NeuralRecon chủ yếu hoạt động trong môi trường tĩnh và gặp khó khăn với các đối tượng động, hạn chế khả năng ứng dụng trong bối cảnh thực tế.

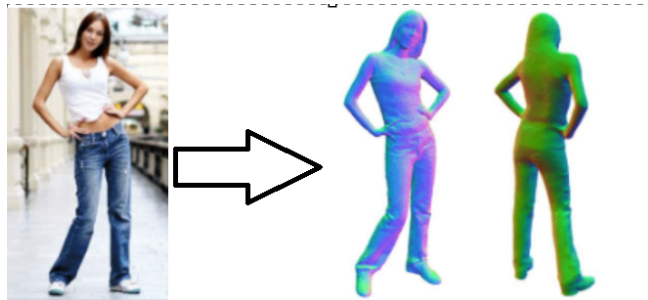
Thách thức chính khi tái tạo 3D với đối tượng động là tính phức tạp của chúng: các đối tượng như con người hay phương tiện thường gây nhiễu, che khuất, và thay đổi hình dạng, vị trí liên tục. Phương pháp tái tạo truyền thống không xử lý tốt các thách thức này, dẫn đến lỗi artifacts (các lỗi do quá trình xử lý gây ra), mô hình thiếu chính xác, và kết quả không nhất quán.

Đề tài này nhằm mở rộng NeuralRecon để tái tạo 3D chính xác và nhất quán trong môi trường động. Chúng tôi tích hợp các module phát hiện và theo dõi đối tượng (object detection và tracking) vào kiến trúc NeuralRecon để nhận diện, theo dõi và tái tạo riêng biệt các đối tượng động bằng kỹ thuật chuyên biệt. Kết quả tái tạo sẽ được hòa trộn mượt mà với mô hình môi trường tĩnh.

Phương pháp đề xuất sử dụng các mô hình học sâu đã được huấn luyện trước, sau đó tinh chỉnh để phù hợp với framework NeuralRecon. Hiệu quả của phương pháp sẽ được đánh giá trên các bộ dữ liệu chứa đối tượng động và so sánh với phiên bản gốc NeuralRecon.

### **Phát biểu bài toán:**

- **Input:** Video đơn ảnh của một bối cảnh (scene) động, chứa cả các đối tượng tĩnh và động (ví dụ: con người, phương tiện giao thông).
- **Output:** Mô hình 3D chính xác và nhất quán của toàn bộ scene, bao gồm cả các phần tĩnh và động.



**Hình 2.** Ví dụ về kết quả của bài toán tái tạo 3D

## MỤC TIÊU

- ❖ **Phát triển một phương pháp mở rộng NeuralRecon:** Mục tiêu chính là xây dựng một hệ thống có khả năng tái tạo 3D thời gian thực trong môi trường có các đối tượng chuyển động, dựa trên nền tảng của NeuralRecon.
- ❖ **Tích hợp module object detection và tracking:** Mục tiêu thứ hai là tích hợp module object detection và tracking vào NeuralRecon, cho phép hệ thống phát hiện và theo dõi các đối tượng chuyển động trong scene.
- ❖ **Đánh giá phương pháp trên dataset có đối tượng động:** Mục tiêu cuối cùng là đánh giá hiệu quả của phương pháp đề xuất trên các dataset có chứa đối tượng động, so sánh kết quả với NeuralRecon gốc và các phương pháp khác.

## NỘI DUNG VÀ PHƯƠNG PHÁP

Trong đề tài này, chúng tôi sẽ thực hiện nghiên cứu các nội dung chính sau:

- ❖ **Bước 1:** Nghiên cứu tổng quan
  - Phương pháp: Tìm kiếm và tổng hợp các bài báo khoa học, tài liệu kỹ thuật về tái tạo 3D từ video đơn ảnh trong môi trường động [3]. Nghiên cứu các phương pháp object detection và tracking hiện đại, đánh giá ưu nhược điểm của từng phương pháp (Siamese Networks [2], DETR [4]). Phân tích kiến trúc và cơ chế hoạt động của NeuralRecon để xác định các điểm có thể cải tiến và mở rộng.
- ❖ **Bước 2:** Thiết kế và xây dựng module tích hợp
  - Phương pháp:
    - Lựa chọn mô hình object detection và mô hình tracking phù hợp.

- Tinh chỉnh các mô hình này trên dataset phù hợp, có thể sử dụng các dataset có chứa cả cảnh tĩnh và đối tượng động.
- Xây dựng giao diện kết nối giữa các module với pipeline của NeuralRecon. Thiết kế cơ chế truyền tải thông tin từ module phát hiện và theo dõi đối tượng đến các module tái tạo 3D tiếp theo.

❖ **Bước 3:** Tái tạo 3D cho đối tượng động

➤ Phương pháp:

- Nghiên cứu các phương pháp tái tạo 3D object-level có thể áp dụng cho các đối tượng động (shape completion [5], NeRF [6]).
- Xem xét khả năng sử dụng biến thể của NeuralRecon hoặc các phương pháp khác để tái tạo 3D cho đối tượng động.

❖ **Bước 4:** Hòa trộn đối tượng động và tĩnh

➤ Phương pháp:

- Nghiên cứu các phương pháp blending và ghép ảnh tạo ra sự chuyển tiếp tự nhiên (Poisson blending [7], làm mờ biên).
- Tối ưu hóa để đạt được độ chính xác và hiệu suất thời gian thực.

❖ **Bước 5:** Đánh giá và so sánh

- Phương pháp: Sử dụng các dataset benchmark có chứa cả cảnh tĩnh và đối tượng động (KITTI dataset [8],...). So sánh kết quả của phương pháp đề xuất với NeuralRecon gốc và các phương pháp khác bằng các metric đánh giá về độ chính xác, tính nhất quán, và hiệu suất thời gian thực.

## KẾT QUẢ MONG ĐỢI

- ❖ Một hệ thống mở rộng của NeuralRecon có khả năng tái tạo 3D thời gian thực trong môi trường động.
- ❖ Module object detection và tracking tích hợp vào NeuralRecon, có thể phát hiện và theo dõi các đối tượng động một cách chính xác.
- ❖ Phương pháp tái tạo 3D riêng biệt cho các đối tượng động, và các kỹ thuật hòa trộn đối tượng động và tĩnh một cách mượt mà.
- ❖ Báo cáo chi tiết về kết quả thực nghiệm, bao gồm cả các so sánh và đánh giá về

hiệu quả của phương pháp.

**TÀI LIỆU THAM KHẢO** (*Định dạng DBLP*)

- [1]. Jiaming Sun, Yiming Xie, Linghao Chen, Xiaowei Zhou, Hujun Bao: NeuralRecon: Real-Time Coherent 3D Reconstruction from Monocular Video. CVPR 2021: 15598-15607
- [2]. Luca Bertinetto, Jack Valmadre, João F. Henriques, Andrea Vedaldi, Philip H.S. Torr: Fully-Convolutional Siamese Networks for Object Tracking. ECCV 2016: 850-865
- [3]. Richard A. Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J. Davison, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Andrew Fitzgibbon: KinectFusion: Real-Time 3D Reconstruction and Interaction Using a Moving Depth Camera. ISMAR 2011: 127-136
- [4]. Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, Sergey Zagoruyko: End-to-End Object Detection with Transformers. ECCV 2020: 213-229
- [5]. Christian B. Choy, Danfei Xu, JunYoung Gwak, Kevin Chen, Silvio Savarese: 3D-R2N2: A Unified Approach for Single and Multi-view 3D Object Reconstruction. ECCV 2016: 628-644
- [6]. Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, Ren Ng: NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. ECCV 2020: 151-174
- [7]. Patrick Pérez, Michel Gangnet, Andrew Blake: Poisson Image Editing. SIGGRAPH 2003: 313-322
- [8]. Andreas Geiger, Philip Lenz, Raquel Urtasun: Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. CVPR 2012: 3354-3361