


THÔNG TIN CHUNG CỦA NHÓM DDD

- Link YouTube video của báo cáo (tối đa 5 phút):
(<https://www.youtube.com/watch?v=e9PhturUY6A>)
- Link slides (dạng .pdf đặt trên Github của nhóm):
(<https://github.com/anhdungbmt2001/CS519.M11.KHCL/blob/main/Final%20Report/CS519.DeCuong.FinalReport.Slides.pdf>)
- Mỗi thành viên của nhóm điền thông tin vào một dòng theo mẫu bên dưới
- Sau đó điền vào Đề cương nghiên cứu (tối đa 5 trang), rồi chọn Turn in

<ul style="list-style-type: none">• Họ và Tên: Nguyễn Trọng Doanh• MSSV: 19521368 	<ul style="list-style-type: none">• Lớp: CS519.M11.KHCL• Tự đánh giá (điểm tổng kết môn): 8/10• Số buổi vắng: 0• Số câu hỏi QT cá nhân: 14• Số câu hỏi QT của cả nhóm: 5• Link Github: NguyenTrongDoanh (Nguyễn Trọng Doanh) (github.com)• Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:<ul style="list-style-type: none">○ Tìm hiểu về đề tài○ Làm slide thuyết trình○ Làm video youtube○ Góp ý chỉnh sửa nội dung○ Góp ý làm poster
<ul style="list-style-type: none">• Họ và Tên: Lưu Anh Dũng• MSSV: 19521392	<ul style="list-style-type: none">• Lớp: CS519.M11.KHCL• Tự đánh giá (điểm tổng kết môn): 8/10• Số buổi vắng: 0• Số câu hỏi QT cá nhân: 14• Số câu hỏi QT của cả nhóm: 7



- Link Github:
<https://github.com/anhdungbmt2001/CS519.M11.KHCL>
- Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:
 - Chọn đề tài
 - Làm file nội dung
 - Làm poster
 - Góp ý làm file thuyết trình
 - Góp ý làm video

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI (IN HOA)

TÁI TẠO BÀN TAY 3D DỰA TRÊN MÔ HÌNH THÔNG QUA HỌC TỰ GIÁM SÁT

TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

MODEL-BASED 3D HAND RECONSTRUCTION VIA SELF-SUPERVISED LEARNING

TÓM TẮT *(Tối đa 400 từ)*

Việc tạo lại bàn tay 3D từ hình ảnh RGB là một thách thức do cấu hình bàn tay khác nhau và độ sâu không rõ ràng. Để có thể tái tạo bàn tay 3D từ hình ảnh một cách đáng tin cậy, hầu hết các phương pháp hiện đại đều dựa vào gắn nhãn 3D ở giai đoạn huấn luyện, nhưng việc gắn nhãn và thu thập nhãn 3D rất tốn kém. Vì thế nên chúng tôi nghiên cứu và dự định sẽ đề xuất một mô hình mạng tái tạo bàn tay 3D tự giám sát có thể ước tính tư thế, hình dạng, kết cấu của bàn tay và góc nhìn của máy ảnh, qua đó giảm bớt sự phụ thuộc vào dữ liệu được gắn nhãn.

Ý tưởng ban đầu của mô hình là thu thập các đặc điểm hình học từ hình ảnh đầu vào thông qua các điểm khóa 2D mà có thể dễ dàng phát hiện và truy cập, sau đó sử dụng tính nhất quán giữa các phương pháp biểu diễn 2D và 3D để hợp lý hóa đầu ra của mạng neural. Mô hình học tự giám sát này được mong đợi rằng có thể đạt hiệu suất tương đương hoặc cao hơn so với các phương pháp và mô hình học có giám sát hoàn toàn gần đây.

GIỚI THIỆU *(Tối đa 1 trang A4)*

Việc tái tạo 3D bàn tay người từ một hình ảnh duy nhất rất quan trọng đối với các tác vụ thị giác máy tính như nhận dạng hành động liên quan đến bàn tay, thực tế tăng cường (AR), dịch ngôn ngữ ký hiệu và tương tác giữa con người với máy tính. Tuy nhiên, do sự đa dạng về mặt hình dạng, kết cấu, tư thế của bàn tay và sự mơ hồ về chiều sâu trong việc tái tạo 3D, việc tái tạo bàn tay 3D dựa trên hình ảnh vẫn là một

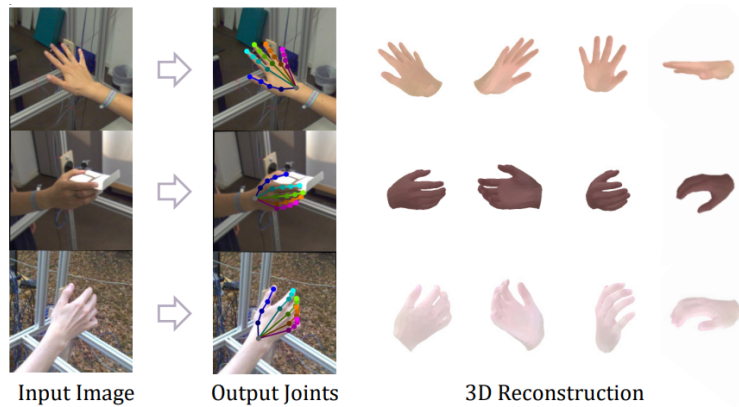
vấn đề khó khăn.

Trong những năm gần đây đã xuất hiện sự tiến bộ nhanh chóng trong việc khôi phục các biểu diễn 3D của bàn tay con người từ hình ảnh. Hầu hết các phương pháp được đề xuất để dự đoán tư thế tay 3D từ hình ảnh chiều sâu hoặc hình ảnh RGB. Tuy nhiên, thông tin bề mặt cần thiết trong một số ứng dụng chẳng hạn như nắm một vật thể bằng tay ảo là không đủ. Để hiển thị tốt hơn thông tin bề mặt của bàn tay, các nghiên cứu trước đây dự đoán lưới tam giác bằng cách hồi quy tọa độ mỗi đỉnh hoặc bằng cách làm biến dạng mô hình bàn tay tham số. Việc xuất ra biểu diễn chiều cao như vậy từ đầu vào 2D là một thách thức đối với các mạng neural, dẫn đến quá trình huấn luyện phụ thuộc nhiều vào các phương pháp gắn nhãn 3D như quét tay dày đặc, lưới tay tham số được trang bị mô hình hoặc các đầu nối 3D được gắn nhãn bởi con người.

Chúng tôi quan sát và nhận thấy rằng các tín hiệu 2D trong không gian hình ảnh có liên quan chặt chẽ với mô hình bàn tay 3D trong thế giới thực. Các điểm khoá (keypoint) 2D của bàn tay chứa thông tin cấu trúc phong phú, qua đó có thể sử dụng trực tiếp các nhãn 2D và hình ảnh đầu vào để học các biểu diễn cấu trúc và kết cấu mà không cần sử dụng nhãn 3D. Tuy nhiên, việc gắn nhãn các điểm khoá 2D của bàn tay vẫn còn tốn nhiều công sức. Để tiết kiệm chi phí gắn nhãn thủ công, một số biểu diễn hình học có thể được trích từ hình ảnh không được gắn nhãn để giúp tái tạo hình dạng và sử dụng thông tin kết cấu có trong hình ảnh đầu vào để giúp lập mô hình kết cấu.

Dựa trên những quan sát và tìm hiểu ở trên, chúng tôi đề xuất một ý tưởng về một mô hình tái tạo bàn tay 3D dựa trên những thông tin giám sát thu được từ các hình ảnh đầu vào và loại bỏ các nhãn được gắn thủ công.

- Input: 1 hình ảnh chụp một bàn tay ở tư thế tự do.
- Output: Hình biểu diễn các khớp của bàn tay và mô hình 3D của bàn tay ở tư thế giống như trong input.



Hình ảnh mô tả input và output của bài toán

MỤC TIÊU (Viết trong vòng 3 mục tiêu)

- Nghiên cứu và xây dựng thành công mô hình mạng neural tái tạo bàn tay 3D thông qua học tự giám sát, cho ra kết quả chính xác các khớp và kết cấu 3D của bàn tay từ một hình ảnh duy nhất mà không sử dụng bất kỳ dữ liệu đã gắn nhãn nào.
- Tìm hiểu các mô hình tái tạo bàn tay 3D hiện có và phát triển mô hình của chúng tôi để đạt độ chính xác và tốc độ xử lý cao hơn các mô hình hiện có.
- Thử nghiệm mô hình trên một số bộ dữ liệu khó, qua đó cải thiện độ chính xác của mô hình.

NỘI DUNG VÀ PHƯƠNG PHÁP

Phương pháp mà chúng tôi nghiên cứu sẽ sử dụng bộ mã hoá tự động để nhận hình ảnh bàn tay làm đầu vào, và trả về hình dạng, kết cấu của tay và góc nhìn máy ảnh, sau đó tạo ra nhiều biểu diễn 2D trong không gian hình ảnh và thiết kế các hàm mất mát cho quá trình huấn luyện. Để hiện thực hoá phương pháp này, có một số vấn đề cần giải quyết. Đầu tiên, làm thế nào để sử dụng hiệu quả các điểm khoá 2D để giám sát việc tái tạo bàn tay 3D có hình dáng lạ? Thứ hai, vì không sử dụng bất kỳ dữ liệu có gắn nhãn nào, làm cách nào để xử lý nhiễu?

- Để giải quyết vấn đề đầu tiên, một bộ mã hoá tự động sẽ được sử dụng để dự đoán các khớp và hình dạng 3D, trong đó các khớp 3D đầu ra sẽ được chiếu vào không gian hình ảnh và trong quá trình huấn luyện sẽ được căn chỉnh với

các điểm khoá đã được phát hiện. Tuy nhiên, nếu chỉ căn chỉnh các điểm khoá trong không gian hình ảnh, có thể sẽ xảy ra nhiều tư thế tay không hợp lệ.

Ngoài ra, các điểm khoá 2D không thể làm giảm mơ hồ về tỷ lệ của bàn tay 3D được dự đoán. Do đó, cần nghiên cứu giải pháp để mạng neural trả về bàn tay 3D với tư thế và kích thước hợp lý.

- Để giải quyết vấn đề thứ hai, cần có một công cụ dự đoán điểm khoá 2D và một sự mất mát mới về tính nhất quán 2D-3D. Công cụ dự đoán điểm khoá 2D sẽ trả về các điểm khoá 2D và sự mất mát nhất quán 2D-3D sẽ liên kết công cụ dự đoán điểm khoá 2D và mạng tái tạo 3D để làm cho cả hai cùng giúp đỡ nhau trong quá trình huấn luyện.

KẾT QUẢ MONG ĐỢI

- Mô hình đạt độ chính xác hơn 80% khi thử nghiệm trên một số bộ dữ liệu lớn như FreiHAND, HO-3D,...
- Mô hình đạt độ chính xác cao hơn so với ít nhất 3 mô hình state-of-the-art.
- Mô hình được ứng dụng trong các ứng dụng thực tế ảo (VR), thực tế tăng cường (AR).

TÀI LIỆU THAM KHẢO (*Định dạng DBLP*)

[1]. Anil Armagan, Guillermo Garcia-Hernando, Seungryul Baek, Shreyas Hampali, Mahdi Rad, Zhaohui Zhang, Shipeng Xie, MingXiu Chen, Boshen Zhang, Fu Xiong, et al:

Measuring generalisation to unseen viewpoints, articulations, shapes and objects for 3d hand pose estimation under hand-object interaction.

European Conference on Computer Vision, 2020.

[2]. Vassilis Athitsos, Stan Sclaroff:

Estimating 3d hand pose from a cluttered image.

CVPR 2003.

[3]. Liuhao Ge, Zhou Ren, Yuncheng Li, Zehao Xue, Yingying Wang, Jianfei Cai, Junsong Yuan:

3d hand shape and pose estimation from a single rgb image.

CVPR 2019.

[4]. Yana Hasson, Bugra Tekin, Federica Bogo, Ivan Laptev, Marc Pollefeys, Cordelia Schmid:

Leveraging photometric consistency over time for sparsely supervised hand-object reconstruction.

CVPR 2020.

[5]. Markus Holl, Markus Oberweger, Clemens Arth, Vincent Lepetit:

Efficient physics-based implementation for realistic hand-object interaction in virtual reality.

IEEE Conference on Virtual Reality and 3D User Interfaces, 2018.

[6]. Maria Parelli, Katerina Papadimitriou, Gerasimos Potamianos, Georgios Pavlakos, Petros Maragos:

Exploiting 3d hand pose estimation in deep learning-based sign language recognition from rgb videos.

European Conference on Computer Vision. Springer, 2020.