

THÔNG TIN CHUNG CỦA NHÓM

- Link YouTube video của báo cáo (tối đa 5 phút):
https://www.youtube.com/watch?v=QC4JuZk_Xoc
- Link slides (dạng .pdf đặt trên Github của nhóm):
<https://github.com/tronghau112/CS2205.CH201/blob/main/H%E1%BA%ADu%20%C4%90inh%20Tr%E1%BB%8Dng%20-%20CS2205.SEP2025.DeCuong.FinalReport.Template.Slide.pdf>
- Họ và Tên: Đinh Trọng Hậu
- MSSV: 250101018



- Lớp: CS2205.CH201
- Tự đánh giá (điểm tổng kết môn): 9.5/10
- Số buổi vắng: 0
- Số câu hỏi cá nhân QT: 4
- Link Github:
<https://github.com/tronghau112/CS2205.CH201>

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI (IN HOA)

HỌC TỔ HỢP SỬ DỤNG TRANSFORMER VÀ MẠNG NO-RON TÍCH CHẬP
CHO NHẬN DẠNG KHUÔN MẶT KHI ĐEO KHẨU TRANG

TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

ENSEMBLE LEARNING USING TRANSFORMERS AND CONVOLUTIONAL
NETWORKS FOR MASKED FACE RECOGNITION

TÓM TẮT

Nhận dạng khuôn mặt là một trong những bài toán quan trọng của thị giác máy tính và đã được ứng dụng rộng rãi trong các hệ thống an ninh, xác thực danh tính và quản lý truy cập. Tuy nhiên, trong bối cảnh thực tế, đặc biệt sau đại dịch COVID-19, việc đeo khẩu trang đã trở nên phổ biến, dẫn đến việc che khuất phần lớn các đặc trưng khuôn mặt quan trọng như mũi và miệng. Hiện tượng này làm suy giảm đáng kể hiệu suất của các mô hình nhận dạng khuôn mặt dựa trên học sâu truyền thống.

Bài báo “Ensemble Learning using Transformers and Convolutional Networks for Masked Face Recognition” đề xuất một phương pháp nhận dạng khuôn mặt khi đeo khẩu trang dựa trên học tổ hợp (ensemble learning), kết hợp các mô hình mạng nơ-ron tích chập (CNN) và Transformer. Phương pháp này tận dụng khả năng trích xuất đặc trưng cục bộ hiệu quả của CNN cùng với khả năng mô hình hóa mối quan hệ toàn cục của Transformer nhằm tăng cường độ bền vững của hệ thống trước hiện tượng che khuất.

Nghiên cứu sử dụng bộ dữ liệu Masked-LFW, được tạo ra từ bộ dữ liệu công khai LFW thông qua công cụ MaskTheFace, để huấn luyện và đánh giá mô hình. Kết quả thực nghiệm cho thấy mô hình học tổ hợp đạt hiệu suất cao hơn so với các mô hình đơn lẻ, cải thiện đáng kể khả năng nhận dạng trong điều kiện khuôn mặt bị che khuất

bởi khẩu trang.

Trên cơ sở đó, đề tài tập trung phân tích phương pháp học tổ hợp trong bài báo, đánh giá khả năng áp dụng của phương pháp trên dữ liệu công khai và mã nguồn mở, đồng thời làm rõ tiềm năng triển khai trong các hệ thống nhận dạng khuôn mặt hoạt động trong môi trường thực tế.

GIỚI THIỆU

Nhận dạng khuôn mặt là một lĩnh vực nghiên cứu trọng tâm của thị giác máy tính và trí tuệ nhân tạo, với nhiều ứng dụng quan trọng trong thực tiễn như kiểm soát an ninh, xác thực người dùng và quản lý truy cập. Các hệ thống nhận dạng khuôn mặt hiện đại đạt độ chính xác cao khi khuôn mặt được quan sát đầy đủ trong điều kiện lý tưởng. Tuy nhiên, trong môi trường thực tế, khuôn mặt thường bị che khuất bởi nhiều yếu tố khác nhau, trong đó việc đeo khẩu trang là thách thức phổ biến nhất trong những năm gần đây.

Việc khẩu trang che khuất phần dưới khuôn mặt khiến các đặc trưng nhận dạng quan trọng không còn đầy đủ, làm giảm hiệu quả của các mô hình nhận dạng khuôn mặt truyền thống. Do đó, bài toán nhận dạng khuôn mặt khi đeo khẩu trang (Masked Face Recognition) đã trở thành một hướng nghiên cứu quan trọng, thu hút nhiều sự quan tâm trong cộng đồng nghiên cứu học sâu.

Bài báo được lựa chọn trong đề tài này để xuất một phương pháp học sâu dựa trên ensemble learning, kết hợp các mô hình CNN và Transformer nhằm tận dụng ưu điểm của từng kiến trúc. CNN cho phép trích xuất đặc trưng cục bộ từ các vùng khuôn mặt không bị che khuất, trong khi Transformer giúp học các mối quan hệ toàn cục giữa các đặc trưng này. Sự kết hợp này góp phần nâng cao tính ổn định và độ chính xác của hệ thống trong điều kiện che khuất.

Nghiên cứu có ý nghĩa cả về mặt học thuật và thực tiễn do sử dụng dữ liệu công khai,

mã nguồn mở và có khả năng triển khai thử nghiệm, phù hợp với yêu cầu phát triển các hệ thống nhận dạng khuôn mặt trong môi trường thực tế phức tạp.

MỤC TIÊU

Nghiên cứu và phân tích bài toán nhận dạng khuôn mặt khi đối tượng đeo khẩu trang.

1. Xây dựng và đánh giá mô hình học sâu kết hợp CNN và Transformer cho nhận dạng khuôn mặt bị che khuất.
2. So sánh hiệu quả của phương pháp học tổ hợp với các mô hình đơn lẻ trên bộ dữ liệu công khai.

NỘI DUNG VÀ PHƯƠNG PHÁP

1. Nội dung nghiên cứu

Nghiên cứu tập trung vào việc phân tích và đánh giá bài toán nhận dạng khuôn mặt khi đeo khẩu trang, trong đó các đặc trưng quan trọng của khuôn mặt bị che khuất, dẫn đến suy giảm hiệu suất của các hệ thống nhận dạng khuôn mặt truyền thống. Nội dung nghiên cứu được triển khai theo các hướng chính sau:

- **Phân tích bài toán Masked Face Recognition (MFR):** Trình bày đặc điểm và các thách thức chính của bài toán MFR, đồng thời so sánh sự khác biệt giữa nhận dạng khuôn mặt thông thường và nhận dạng khuôn mặt trong điều kiện che khuất. Phân tích các nguyên nhân gây suy giảm độ chính xác khi các vùng đặc trưng như mũi, miệng và cằm không còn được quan sát đầy đủ.
- **Khảo sát các phương pháp liên quan:** Tổng hợp và phân tích các hướng tiếp cận tiêu biểu trong lĩnh vực nhận dạng khuôn mặt khi đeo khẩu trang, bao gồm tăng cường dữ liệu, khai thác các vùng khuôn mặt không bị che khuất, cơ chế attention và các mô hình học sâu hiện đại. Qua đó làm rõ ưu điểm và hạn chế của từng phương pháp, cũng như bối cảnh áp dụng phù hợp.

- **Phân tích phương pháp học tổ hợp:** Làm rõ vai trò của phương pháp học tổ hợp trong việc kết hợp nhiều mô hình học sâu nhằm tăng tính ổn định và giảm sự phụ thuộc vào một kiến trúc đơn lẻ. Phân tích cách kết hợp CNN và Transformer để tận dụng đồng thời đặc trưng cục bộ và quan hệ toàn cục của khuôn mặt, đồng thời chỉ ra các pattern thường gặp trong bài toán MFR, như việc CNN hoạt động hiệu quả trên các vùng không bị che khuất trong khi Transformer gặp khó khăn hơn khi dữ liệu huấn luyện có quy mô hạn chế.
- **Định hướng đóng góp của nghiên cứu:** Đề tài tập trung tái hiện và phân tích phương pháp học tổ hợp được đề xuất trong bài báo, đồng thời đánh giá khả năng áp dụng phương pháp trên dữ liệu công khai và mã nguồn mở, từ đó rút ra các nhận xét và định hướng cải tiến cho các nghiên cứu tiếp theo.

2. Phương pháp nghiên cứu

2.1. Chuẩn bị và xử lý dữ liệu

- Sử dụng bộ dữ liệu công khai Labeled Faces in the Wild (LFW) làm dữ liệu gốc.
- Áp dụng công cụ MaskTheFace để tạo bộ dữ liệu Masked-LFW, mô phỏng các trường hợp khuôn mặt đeo khẩu trang với nhiều kiểu che khuất khác nhau.
- Thực hiện tiền xử lý dữ liệu bao gồm phát hiện và căn chỉnh khuôn mặt, chuẩn hóa kích thước ảnh và chia dữ liệu thành các tập huấn luyện, kiểm tra và đánh giá.

2.2. Xây dựng mô hình học sâu

- **Mô hình CNN:** Huấn luyện và tinh chỉnh các mô hình CNN phổ biến như VGG, EfficientNet và FaceNet nhằm trích xuất đặc trưng cục bộ từ các vùng khuôn mặt không bị che khuất.
- **Mô hình Transformer:** Sử dụng Transformer để học các mối quan hệ toàn cục giữa các đặc trưng còn lại trên khuôn mặt, giúp mô hình khai thác hiệu quả

thông tin từ vùng mắt và trán.

2.3. Học tổ hợp (Ensemble Learning)

- Kết hợp đầu ra của các mô hình CNN và Transformer thông qua các chiến lược như bỏ phiếu đa số và trung bình xác suất dự đoán.
- Phương pháp học tổ hợp giúp giảm sự phụ thuộc vào từng mô hình riêng lẻ và tăng tính bền vững của hệ thống.

2.4. Đánh giá và so sánh

- Đánh giá hiệu suất mô hình thông qua các chỉ số Accuracy, Precision và Recall, đồng thời xem xét khả năng mở rộng đánh giá theo các tiêu chí xác thực khuôn mặt trong các nghiên cứu liên quan.
- So sánh kết quả của mô hình học tổ hợp với các mô hình đơn lẻ và phân tích các trường hợp nhận dạng sai nhằm đề xuất hướng cải thiện.

KẾT QUẢ MONG ĐỢI

Mô hình học tổ hợp đạt hiệu suất nhận dạng cao hơn so với các mô hình truyền thống trong điều kiện khuôn mặt bị che khuất.

- Chứng minh tính hiệu quả và tính ổn định của phương pháp ensemble learning đối với bài toán Masked Face Recognition.
- Cung cấp cơ sở cho việc triển khai và mở rộng các hệ thống nhận dạng khuôn mặt trong môi trường thực tế.

TÀI LIỆU THAM KHẢO (*Định dạng DBLP*)

[1]. Mohammed R. Al-Sinan, Aseel F. Haneef, and Hamzah Luqman.
Ensemble Learning using Transformers and Convolutional Networks for Masked Face Recognition.

WACV Workshops, 2023.

[2]. Mei Ngan, Patrick Grother, and Kayee Hanaoka.

Ongoing face recognition vendor test (FRVT).

NIST Report, 2020.

[3]. Dale Fitousi, Noga Rotschild, Chaya Pnini, and Omri Azizi.

Understanding the impact of face masks on the processing of facial identity, emotion, age, and gender.

Frontiers in Psychology, 2021.

[4]. Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller.

Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments.

University of Massachusetts Amherst, 2007.

[5]. Jia Deng, Jiankang Guo, Niannan Xue, and Stefanos Zafeiriou.

ArcFace: Additive angular margin loss for deep face recognition.

CVPR 2019.

[6]. Mingxing Tan and Quoc V. Le.

EfficientNet: Rethinking model scaling for convolutional neural networks.

ICML 2019.

[7]. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones,

Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin.

Attention Is All You Need.

NeurIPS 2017.

[8]. Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, et al.

An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale.

ICLR 2021.

