BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC SƯ PHẠM TP. HỒ CHÍ MINH

BÁO CÁO ĐỀ TÀI NGHIÊN CỨU KHOA HỌC

MỘT HỆ THỐNG ĐIỂM DANH BẰNG NHẬN DẠNG KHUÔN MẶT TRONG LỚP HỌC TRỰC TUYẾN

Giảng viên hướng dẫn: TS. NGUYỄN VIẾT HƯNG

Nhóm thực hiện:

Lê Tấn Lộc	45.01.104.135
Lê Ái Quốc Vinh	45.01.104.277
Chống Chí Dinh	46.01.104.029
Nguyễn Tô Thụy Anh	46.01.104.007

TP HỒ CHÍ MINH – 4/2022

Trường Đại học Sư Phạm Thành phố Hồ Chí Minh
Khoa Công Nghệ Thông Tin

BÁO CÁO ĐỀ TÀI NGHIÊN CỨU KHOA HỌC SINH VIÊN

Một hệ thống điểm danh bằng nhận dạng khuôn mặt trong lớp học trực tuyến

Nhóm sinh viên thực hiện

LÊ TẤN LỘC - 45.01.104.135

LÊ ÁI QUỐC VINH - 45.01.104.277

CHỐNG CHÍ DINH - 46.01.104.029

NGUYỄN TÔ THỤY ANH - 46.01.104.007

Giảng viên hướng dẫn

TS. NGUYỄN VIẾT HƯNG

MỤC LỤC

DANH MỤC HÌNH ẢNH

DANH MỤC BẢNG BIỂU

LỜI CẢM ƠN

Trong quá trình thực hiện bài nghiên cứu khoa học này, nhóm chúng em đã nhận được nhiều sự giúp đỡ từ các thầy cô trong trường Đại học Sư phạm Thành phố Hồ Chí Minh. Chúng em xin cảm ơn các thầy cô đã tận tình chỉ dẫn, giúp chúng em hiểu rõ hơn về đề tài và hướng chúng em đi đúng hướng.

Đặc biệt, chúng em xin cảm ơn thầy TS. Nguyễn Viết Hưng và anh Vương Lê Minh Nguyên đã truyền đạt vốn kiến thức quý báu cho chúng em, hướng dẫn, hỗ trợ về chuyên môn và theo dõi sát sao trong suốt quá trình thực hiện đề tài.

Một lần nữa chúng em xin chân thành cảm ơn.

Thay mặt nhóm thực hiện

Lê Tấn Lộc

CHƯƠNG 1. TỔNG QUAN

1.1. Giới thiệu đề tài

Trong tình hình đại dịch Covid-19 kéo dài, học trực tuyến đang đóng vai trò chủ đạo trong việc truyền đạt kiến thức của giáo viên và tự rèn luyện của học sinh, sinh viên nhưng vẫn giữ được an toàn cho bản thân. Tuy nhiên, việc đảm bảo hiệu quả trong giảng dạy vẫn gặp nhiều thách thức. Mạng xã hội là một môi trường thuận lợi để tuyên truyền các chiêu trò để gian lận, gây ra những tác động xấu đến chất lượng học tập và giảng dạy. Do đó, để hỗ trợ cho giáo viên, xây dựng một hệ thống điểm danh bằng nhận dạng khuôn mặt trong lớp học trực trực tuyến là cần thiết. Tuy nhiên, thực trạng gian lận trong việc điểm danh tồn tại khá phức tạp và tồn tại ở nhiều hình thức khác nhau, phổ biến nhất là việc sử dụng khuôn mặt giả mạo để điểm danh qua camera như: sử dụng video, hình ảnh bản thân đặt trước ống kính. Trong đó, thực trạng học sinh, sinh viên tìm cách để gian lận việc điểm danh diễn ra khá thường xuyên, làm giảm chất lượng học tập. Phổ biến nhất là việc sử dụng khuôn mặt giả mạo để điểm danh qua camera như: sử dụng video, hình ảnh bản thân đặt trước ống kính.

Vấn nạn hiện tại của học sinh, sinh viên là không chủ động trong việc học. Việc học tập trực tuyến dẫn đến các hệ lụy, tạo cơ hội để các bạn trẻ trở nên thụ động, không chuyên cần trong học tập. Từ đó, dẫn đến các học sinh, sinh viên tìm kiếm nhiều cách để gian lận trong việc điểm danh khi lên lớp.

Hơn nữa, cách mạng công nghiệp 4.0 nổ ra, làm cho công nghệ ngày càng phát triển, các bạn trẻ được tiếp xúc với công nghệ từ rất sớm, quá trình hình thành và phát triển của giới trẻ giúp cho việc mày mò các thiết bị công nghệ để sử dụng nhằm mục đích gian lận ngày càng tinh vi, nhất là trong giai đoạn học tập trực tuyến hiện nay.

Ngoài ra, điều kiện học tập trực tuyến của các bạn học sinh, sinh viên là không giống nhau. Chính vì vậy mà việc kiểm chứng giả mạo đối với các khuôn mặt từ camera khi học sinh điểm danh là rất khó - khó trong việc dựa vào chất lượng hình ảnh để xác định khuôn mặt.

Vì những lí do trên, nhóm nghiên cứu đã thực hiện bài nghiên cứu *một mô hình phát hiện gian lận điểm danh trong lớp học trực tuyến* nhằm giúp học sinh, sinh viên có tinh thần tự giác hơn trong việc học tập trực tuyến.

Mô hình này được xây dựng dựa trên thực trạng hiện nay của các lớp học trực tuyến. Đề tài với mục tiêu phân tích và phát hiện gian lận từ camera. Hệ thống sẽ sử dụng hình ảnh từ các camera sau buổi học, phân loại giả mạo và đưa ra kết quả.

Về mặt ý tưởng của một mô hình phát hiện gian lận điểm danh trong lớp học trực tuyến mà nhóm đề xuất dựa trên mục tiêu cơ bản của bài toán chống giả mạo khuôn mặt (Face Anti-Spoofing). Cụ thể mô hình sẽ dựa trên một số thông tin một người nào đó (học sinh, sinh viên) trong lúc điểm danh thông qua các thiết bị ghi hình, từ đó đưa ra quyết định xem người đó có phải giả mạo hay không?

Mô hình hướng tới các lớp học trực tuyến có nhu cầu điểm danh qua camera với mong muốn sẽ trở thành một công cụ phổ biến để tăng hiệu quả trong giảng dạy.

1.2. Mục tiêu cụ thể

Trong đề tài nghiên cứu này, nhóm nghiên cứu tập trung vào việc *phát hiện gian lận điểm danh trong lớp học trực tuyến*.

Những mục tiêu mà nhóm nghiên cứu muốn hướng đến và đã đạt được trong đề tài này là:

- Xây dựng một mô hình phát hiện người điểm danh (học sinh, sinh viên) có giả mạo hay không?
- Xây dựng một trang web, mô phỏng việc phát hiện giả mạo điểm danh trong lớp học trực tuyến.

Mục đích các nghiên cứu này nhằm nâng cao hiệu quả trong công tác giảng dạy của nhà trường.

1.3. Đối tượng và phạm vi nghiên cứu

1.3.1. Đối tượng nghiên cứu:

Đề tài mà nhóm đề xuất là một đề tài về xử lý ảnh số, với bài toán trọng tâm là Face Anti-Spoofing. Đề tài ban đầu nhằm vào các đối tượng là người học trong các lớp học trực tuyến.

1.3.2. Phạm vi nghiên cứu:

Phạm vi nghiên cứu đề tài là phát hiện được đối tượng giả mạo các hoạt động điểm danh bằng camera trong lớp học trực tuyến.

Phạm vi áp dụng có thể áp dụng trong các lớp học trực tuyến.

Giới hạn thông tin: Mô hình chỉ hỗ trợ cho biết người học có gian lận trong lúc điểm danh hay không, ngoài ra không cho biết bất kì thông tin nào khác như: tên, tuổi, giới tính, ...

Giới hạn chất lượng ảnh: Mô hình có thể hoạt động không tốt đối với các thiết bị ghi hình với chất lượng thấp.

Giới hạn về phương thức triển khai: Mô hình hiện tại không thể chạy trên thời gian thực, mà chỉ lấy vài khung hình của từng người học để phát hiện gian lận.

1.4. Phương pháp nghiên cứu

1.4.1. Phương pháp nghiên cứu lý thuyết

- Tìm hiểu các công trình nghiên cứu liên quan.
- Tìm hiểu về các bài toán và các phương pháp phát hiện giả mạo khuôn mặt.
- Tìm hiểu các phần mềm lớp học trực tuyến hiện có.

1.4.2. Phương pháp nghiên cứu thực nghiệm

- Phân tích yêu cầu của một mô hình phát hiện gian lận khuôn mặt.
- Xây dựng một mô hình phát hiện gian lận khuôn mặt thông qua một vài khung ảnh của người học.
- Xây dựng một trang web lớp học trực tuyến.
- Tích hợp mô hình phát hiện giả mạo khuôn mặt vào trang web lớp học trực tuyến đã xây dựng.
- Đánh giá kết quả đạt được dựa trên thực nghiệm.

1.5. Ý nghĩa khoa học và thực tiễn

1.5.1. Ý nghĩa khoa học

- Đóng góp vào quá trình nghiên cứu xử lý ảnh số trong bài toán phát hiện giả mạo khuôn mặt.
- Đóng góp vào quá trình nghiên cứu một mô hình điểm danh trong lớp học trực tuyến.

1.5.2. Ý nghĩa thực tiễn

- Xây dựng một công cụ hỗ trợ cho việc điểm danh các lớp học trực tuyến trong tình hình dịch bệnh Covid 19.
- Tạo tiền đề cho sự phát triển các ứng dụng lớp học trực tuyến.

CHƯƠNG 2. TỔNG QUAN VÀ CƠ SỞ LÝ THUYẾT

2.1. Tình hình nghiên cứu và phát triển

2.1.1. Tình hình nghiên cứu

Nhận dạng khuôn mặt cung cấp nhiều lợi thế so với các sinh trắc học có sẵn khác, nhưng nó đặc biệt dễ bị giả mạo. Năm 2012, De Marsico và cộng sự [31] đề xuất một giải pháp hữu hiệu và hiệu quả cho vấn đề giả mạo khuôn mặt. Cách tiếp cận được đề xuất có thể xác minh xem khuôn mặt có thực sự là 3D hay không vẫn duy trì chi phí tính toán thấp. Tương tác của người dùng cũng cho phép phát hiện các hành vi giả mạo phức tạp hơn, chẳng hạn như việc trình bày các video được quay trước. Tuy nhiên, một số người trong số họ chỉ phát hiện các cuộc tấn công rất đơn giản và hệ thống không thể phát hiện giả mạo thông qua mặt nạ chuyển động 3D.

Năm 2013, Bharadwaj và cộng sự [7] đã đề xuất một phương pháp chính xác và chi phí tính toán thấp hơn. Bài báo này trình bày một cách tiếp cận mới để phát hiện giả mạo trong video khuôn mặt bằng cách sử dụng tính năng phóng đại chuyển động. Phương pháp phóng đại chuyển động Eulerian được sử dụng để nâng cao các biểu cảm khuôn mặt thường thể hiện bởi các đối tượng trong một video đã quay. Tiếp theo, hai loại thuật toán trích xuất đặc trưng được đề xuất: (i) cấu hình LBP cung cấp hiệu suất được cải thiện so với các phương pháp dựa trên kết cấu tính toán đắt tiền khác và (ii) phương pháp ước lượng chuyển động sử dụng bộ mô tả HOOF. Nghiên cứu này trình bày một khuôn khổ mới để phát hiện giả mạo trong các hệ thống nhận dạng khuôn mặt. Sử dụng tính năng phóng đại chuyển động, video đầu vào của một chủ thể được cải tiến để phóng đại các nét mặt vi mô và vĩ mô tinh tế thường được thể hiện bởi một người thực.

Năm 2015, Zhenqi Xu và các cộng sự [48] đã giới thiệu phương pháp nhận dạng khuôn mặt giả mạo dựa theo temporal features (các đặc điểm nhận dạng theo thời gian) nhằm mô tả cấu trúc động của khuôn mặt, thay vì sử dụng ảnh đơn lẻ như thông thường, phương pháp này xem việc nhận dạng khuôn mặt thật giả như một bài toán phân loại video để trích xuất mối quan hệ về thời gian và các đặc điểm trên khuôn mặt từ các khung hình khác nhau trong video.

Năm 2017, nhận thấy các tính năng phát hiện và chống giả mạo khuôn mặt trên video vẫn chưa đủ, Gan và cộng sự [15] đã giới thiệu một tính năng chống giả mạo khuôn mặt dựa trên CNN (3D convolution neural network). Trong bài báo này, thay vì trích xuất các tính năng từ một hình ảnh, các tính năng được học từ các khung hình video. Để nhận ra tính năng chống giả mạo khuôn mặt, các tính năng không gian của khung

video liên tục được trích xuất bằng cách sử dụng mạng nơ-ron tích chập 3D (CNN) từ cấp khung hình video ngắn. Đến năm 2021, mặc dù các thiết bị phát hiện dựa trên Mạng lưới thần kinh phù hợp vani (CNN) có thể đạt được hiệu suất khả quan trong việc phát hiện khuôn mặt giả, Gan và các cộng sự [8] quan sát thấy rằng các thiết bị phát hiện có xu hướng tìm kiếm giả mạo trên một vùng hạn chế của khuôn mặt, điều này cho thấy rằng các thiết bị phát hiện thiếu hiểu biết về giả mạo. Do đó, họ đề xuất một khung tăng cường dữ liệu dựa trên sự chú ý để hướng dẫn việc tinh chỉnh máy dò và mở rộng sự chú ý của nó.

Cùng năm 2017, Atoum và cộng sự [45] giới thiệu một phương pháp chống giả mạo khuôn mặt mới dựa trên việc kết hợp hai luồng CNN. Không giống như các phương pháp trước đây trong việc chống giả mạo khuôn mặt sử dụng toàn khuôn mặt để phát hiện các cuộc tấn công bản trình bày, họ tận dụng cả hình ảnh khuôn mặt đầy đủ và các bản vá được trích xuất từ cùng một khuôn mặt để phân biệt giả mạo với khuôn mặt trực tiếp. Luồng CNN đầu tiên dựa trên diện mạo bản vá được trích xuất từ các vùng trên khuôn mặt. Luồng CNN thứ hai dựa trên ước tính độ sâu khuôn mặt bằng cách sử dụng hình ảnh toàn khuôn mặt. Việc kết hợp cả hai luồng CNN dẫn đến một cải tiến tổng thể được so sánh thuận lợi với SOTA *(state-of-the-art)*.

Năm 2019, Usman và các cộng sự [40] đã đề xuất mô hình nhận diện giả mạo khuôn mặt dựa trên phương pháp lọc thưa thớt (sparse filtering) được áp dụng để tạo ra thêm nhiều chi tiết học mẫu (features) bằng cách áp dụng ResNet. Các features được tạo ra được xây dựng ở dạng chuỗi tuần tự, đưa vào mô hình LSTM để xây dựng biểu diễn cuối cùng.

2.1.2. Thách thức

Giả mạo khuôn mặt là một nỗ lực để có được quyền truy cập vào hệ thống sinh trắc học khuôn mặt. Điều này xảy ra khi người xâm nhập cố gắng đánh lừa hệ thống bằng cách hiển thị một bản sao hình ảnh khuôn mặt chẳng hạn như: ảnh in, chuỗi video của hình ảnh, video cảnh động hoặc mô hình 3D của người cụ thể. Để đánh giá khuôn mặt giả hay thật, một số hình ảnh khuôn mặt được lấy mẫu ngẫu nhiên. Tuy nhiên, không có manh mối trực quan nào có thể hiểu được để chọn các mẫu giả và thật. Điều này cho thấy sự đa dạng của việc giả mạo các cuộc tấn công, khi kiểm tra trực quan hình ảnh khuôn mặt chứng tỏ rằng các cuộc tấn công mạo danh có thể rất giống nhau, ngay cả mắt người cũng có thể khó phân biệt mặt thật và mặt giả.

Hầu hết các biện pháp đối phó trước đây sử dụng mạng nơ ron tích chập (CNN) [48][25] phát hiện khuôn mặt trong hình ảnh và đặt lại để chứa hầu hết vùng có sự xuất hiện của khuôn mặt. Do đó, các biện pháp này phụ thuộc trực tiếp vào việc nhận diện khuôn mặt và đôi lúc điều này có thể xảy ra sai xót. Hơn nữa, bản thân mạng

CNN không thể học được các đặc điểm thời gian, để nắm bắt các dấu hiệu phân biệt giữa truy cập trong thời gian thực và tấn công mạo danh. Trong bài báo [31], Marsico và cộng sự đã sử dụng phương pháp dựa trên phép chiếu xạ ảnh 3D để chống giả mạo khuôn mặt chuyển động, nhưng không xử lý được với ảnh giả mạo 3D.

Hầu hết các công việc trước đây đều sử dụng các đặc trưng được tạo ra một cách thủ công và áp dụng các kỹ thuật học cạn (shallow learning) ví dụ như SVM và LDA để phát triển một hệ thống chống giả mạo khuôn mặt. Rất nhiều nghiên cứu chú ý đến sự khác biệt về kết cấu giữa mặt thật của người sống và mặt giả do các phương pháp nhân tạo làm ra. Nhiều nhóm nghiên cứu cũng đã tìm hiểu và phát triển hệ thống chống giả mạo khuôn mặt từ trước, và phần lớn các đặc điểm được trích xuất phải dựa vào các thuật toán xử lý ảnh như LBP [23], LBP-TOP [39], HOG [22], DoG [6][43] và còn rất nhiều. Tuy nhiên, các thuật toán xử lý ảnh chỉ có tỷ lệ nhận dạng tốt hơn cho một loại hình ảnh nhất định. Ngoài ra, các yếu tố ngoại cảnh như ánh sáng, độ chói hoặc chất lượng cảm biến camera cũng làm ảnh hưởng đến độ chính xác của mô hình. Thông thường, các phương pháp giả mạo bằng giấy, hình ảnh, video sẽ làm cho độ phân giải hình ảnh mà mô hình nhận được rất thấp, từ đó không tìm ra được đặc trưng giống với khuôn mặt thật và sẽ bị kết luận là giả mạo. Dựa theo nguyên lý này, các hệ thống giả mạo khuôn mặt hoạt động theo cơ chế này rất dễ bị đánh lừa bởi bức ảnh được in với các thiết bị in tiên tiến, có độ phân giải cao.

Biện pháp đối phó điển hình đối với các cuộc tấn công giả mạo là phát hiện khuôn mặt sống (hay còn gọi là phát hiện sự sống – liveness detection) nhằm mục đích phát hiện các dấu hiệu sinh lý của sự sống (chẳng hạn như chớp mắt, thay đổi biểu hiện trên khuôn mặt và cử động miệng). Ví dụ, G. Pan và đồng sự [14] đã đề xuất một phương pháp chống giả mạo dựa trên việc xác định sự chớp mắt bằng cách tích hợp một phương pháp dự đoán có cấu trúc trong khi Kollreider và đồng sự [26] đã đề xuất một phương pháp dựa trên luồng quang học để nắm bắt những chuyển động của hình ảnh khuôn mặt. Mặc dù các biện pháp đối phó như vậy có thể hiệu quả trong các trường hợp tấn công sử dụng ảnh, nhưng chúng thường không hiệu quả khi sử dụng video (hoặc đơn giản là lắc ảnh trước máy ảnh) như một phương tiện giả mạo. Một số nhà nghiên cứu đã cố gắng chống lại việc giả mạo video bằng cách sử dụng cấu trúc từ chuyển động để tính toán thông tin về độ sâu khuôn mặt. Một lần nữa, điều này có thể không hoạt động trong trường hợp tấn công giả mạo bằng cách sử dụng mặt nạ 3D.

Có rất nhiều dấu hiệu trực quan để phát hiện giả mạo đã được khám phá và có rất nhiều kết quả ấn tượng đã được báo cáo trên các cơ sở dữ liệu riêng lẻ như Replay Attack [9], CASIA [49], … Tuy nhiên, bản chất khác nhau của các cuộc tấn công giả mạo và các điều kiện không giống nhau khiến chúng ta không thể dự đoán được cách thức các kỹ thuật chống giả mạo đơn lẻ, ví dụ: phân tích kết cấu khuôn mặt, có thể

khái quát vấn đề trong các ứng dụng thực tế. Hơn nữa, chúng ta không thể lường trước tất cả các kịch bản tấn công có thể xảy ra và đưa chúng vào cơ sở dữ liệu vì trí tưởng tượng của con người luôn tìm ra những mánh khóe mới để đánh lừa các hệ thống hiện có.

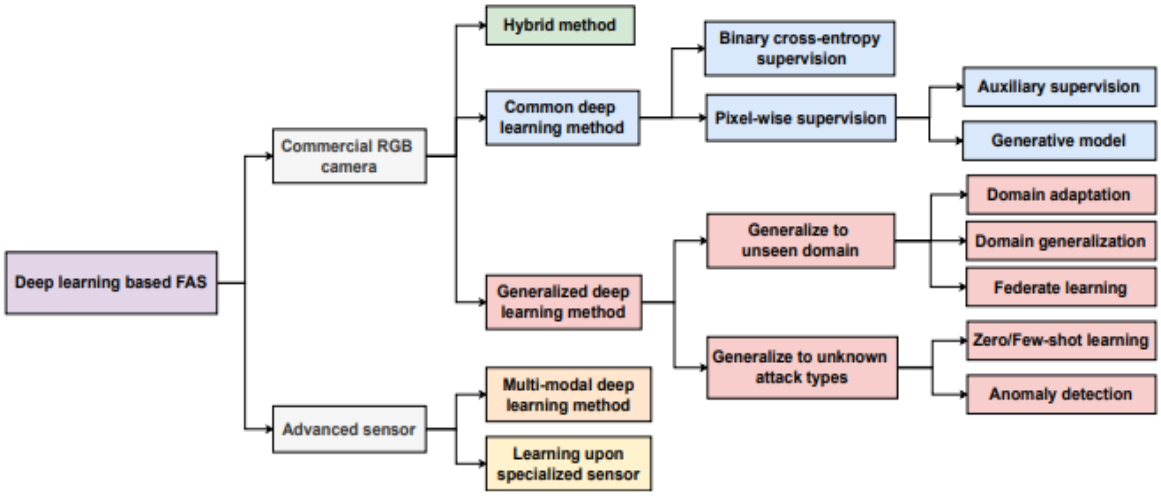
Thêm một điểm đáng chú ở các biện pháp chống giả mạo thường không được phát triển để hoạt động như một thủ tục độc lập mà là trong một chuỗi hoạt động chung với một hệ thống nhận dạng. Tuy nhiên, hầu hết các công trình nghiên cứu về chống giả mạo có xu hướng chỉ tập trung vào phần phát hiện giả mạo, do đó bỏ qua phần tích hợp biện pháp đối phó vào một hệ thống nhận dạng. Vấn đề là làm thế nào để kết hợp biện pháp chống giả mạo và nhận dạng sinh trắc học để hệ thống nhận dạng sinh trắc học kết hợp mạnh mẽ hơn đối với giả mạo và không bị giảm độ chính xác khi tiến hành nhận dạng [16][10].

Hình ảnh khuôn mặt được chụp từ các phương pháp giả mạo khuôn mặt có thể trông rất giống với hình ảnh được chụp từ khuôn mặt trực tiếp. Do đó, việc phát hiện giả mạo khuôn mặt có thể khó thực hiện nếu chỉ dựa trên một hình ảnh khuôn mặt đơn lẻ hoặc một chuỗi video tương đối ngắn. Tùy thuộc vào hình ảnh và chất lượng khuôn mặt giả, gần như không thể phân biệt được đâu là khuôn mặt thật và khuôn mặt giả mà không có bất kỳ thông tin cảnh nào hoặc chuyển động không tự nhiên hoặc các mẫu kết cấu khuôn mặt.

Việc phát hiện giả mạo dựa trên phân tích độ sống (liveness analysis) và chuyển động khá khó thực hiện bằng cách chỉ quan sát chuyển động khuôn mặt tự phát trong chuỗi hình ảnh hoặc video ngắn. Vấn đề này có thể được đơn giản hóa bằng cách nhắc người dùng thực hiện một số hành động theo yêu cầu ngẫu nhiên cụ thể (chẳng hạn như mỉm cười và di chuyển đầu sang bên phải). Phản hồi của người dùng (nếu có) sẽ cung cấp bằng chứng xác thực. Đây được gọi là phương pháp tiếp cận phản hồi thách thức để phát hiện giả mạo. Hạn chế của cách tiếp cận như vậy là nó yêu cầu sự hợp tác của người dùng, do đó làm cho quá trình xác thực tốn nhiều thời gian.

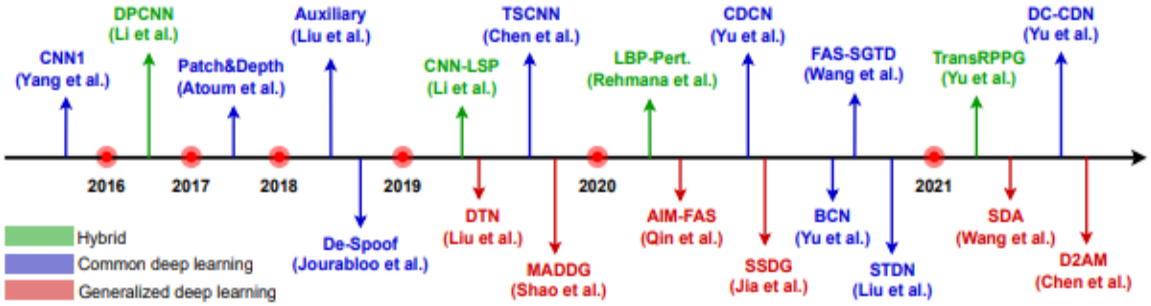
2.1.3. Các hướng tiếp cận phát triển

Gần đây, vào năm 2021 Zitong Yu và các cộng sự của mình [47] đã đề xuất một cấu trúc liên kết của các phương pháp tiếp cận *Face Anti-Spoofing - FAS* dựa trên học sâu (*Deep Learning*) được chia thành hai nhóm lớn: Commercial RGB camera và Advanced sensor.



Hình 2.1. Các phương pháp tiếp cận của FAS [47]

Commercial RGB camera



Hình 2.2. Tổng quan theo trình tự thời gian của các phương pháp FAS dựa trên học sâu dựa trên cột mốc quan trọng sử dụng commercial RGB camera [47]

Commercial RGB camera được sử dụng rộng rãi trong nhiều tình huống ứng dụng thực tế như: hệ thống kiểm soát truy cập, mở khóa thiết bị di động, ... Các phương pháp FAS dựa trên học sâu hiện hiện có sử dụng commercial RGB camera có ba loại chính:

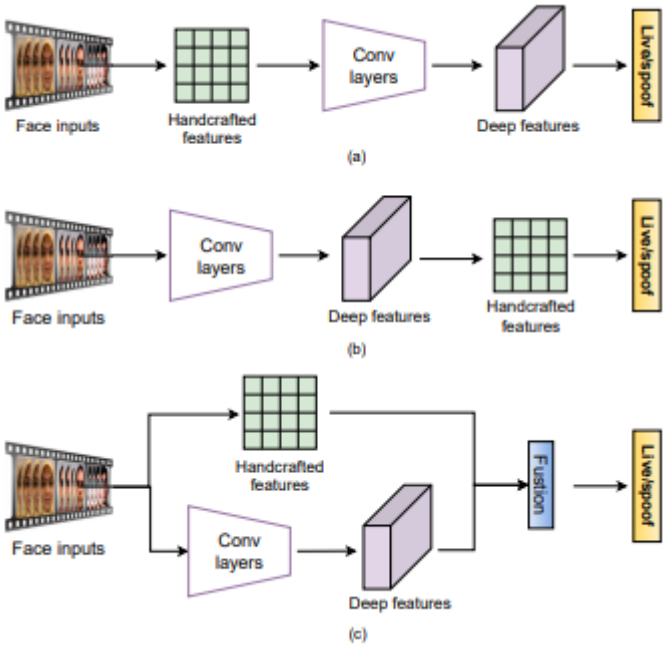
a. Hybrid Method: Một phương pháp từ sự kết hợp giữa handcrafted feature (đặc tính được sinh ra bởi các thuật toán) và deep learning features (đặc tính được trích xuất bởi các lớp học sâu)

b. Common deep learning method: Một phương pháp học sâu có giám sát (supervised deep learning) từ đâu đến cuối.

c. Generalized deep learning: Một cách học cao cấp hơn, mục đích là xác định các kiểu tấn công, giả dạng bằng khuôn mặt mà mắt người không thể xác định được.

Hybrid Method

Mặc dù học sâu và CNN đã đạt được nhiều thành tích thành công trong nhiều nhiệm vụ thị giác máy tính (ví dụ: phân loại hình ảnh, phân đoạn ngữ nghĩa và đối tượng), nhưng số lượng và tính đa dạng của dữ liệu đào tạo vẫn còn hạn chế. Vì thế một số thuật toán hiện đại được ra đời, cổ điển nhất là Local Binary Patterns (LBP)[37], tiếp đó là Histogram of Oriented Gradients (HOG) [32], Image Quality [20], Optical Flow Motion [20][38], ... Từ đây, nhiều bài báo đã có những đề xuất khác nhau về cấu trúc của mô hình.



Hình 2.3. Các khuôn mẫu kết hợp cho FAS [47]

(a) Các đặc tính được trích xuất bởi các lớp học sâu thông qua đặc tính được sinh ra bởi các thuật toán.

(b) Các đặc tính được sinh ra bởi các thuật toán thông qua đặc tính được trích xuất bởi các lớp học sâu.

(c) Sự kết hợp đồng thời giữa các đặc tính ở cả hai quá trình (thuật toán và học sâu).

Common Deep Learning Method

Kế thừa từ sự phát triển của convolutional neural network (CNN) cùng với sự đóng góp các bộ dữ liệu quy mô lớn như CelebA-Spoof [46], HiFiMask [4], ... thì phương pháp này ngày càng thu hút được nhiều sự chú ý. Các phương pháp common deep learning phổ biến cần dữ liệu đầu vào là khuôn mặt để phát hiện giả mạo. Các khuôn khổ học sâu phổ biến bao gồm:

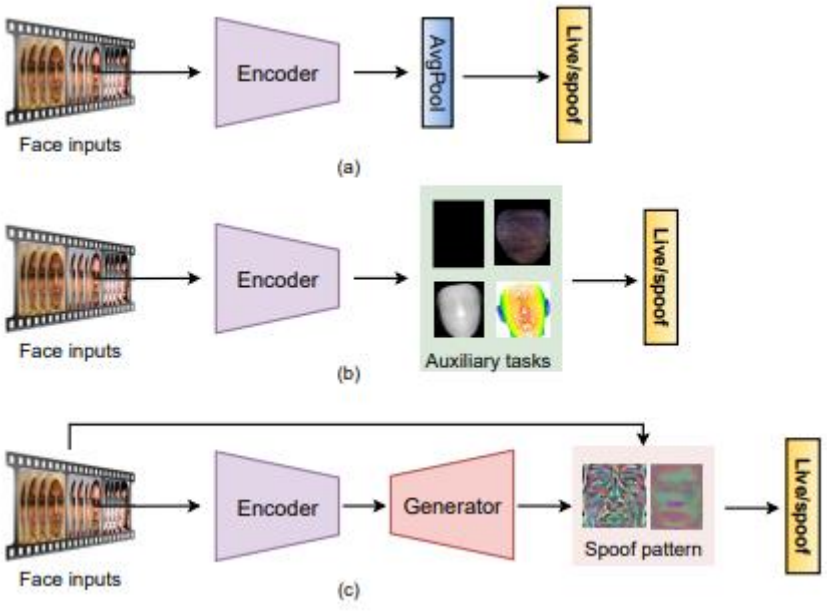
Direct Supervision with Binary Cross Entropy Loss: Giả sử mô hình chỉ trả về hai kết quả là có hoặc không giả mạo, thì có thể coi đây là bài toán phân lớp nhị phân, với làm kiểm tra mất mát là binary cross-entropy (CE):

$$\mathcal{L} = -y log\hat{y} - (1-y)\log(1-\hat{y})$$

Pixel-wise Supervision: Các mô hình học sâu dựa trên Direct Supervision with Binary Cross Entropy Loss sẽ dễ dàng học theo các đặt tính không có độ tin cậy cao (ví dụ: Đường viền). Ngược lại với phương pháp này có thể cung cấp các đặt tính liên quan đến ngữ cảnh một cách chi tiết góp phần nâng cao chất lượng của mô hình. Một số phương pháp đại diện cho khuôn khổ học sâu này:

Pixel-wise supervision with Auxiliary Task: Theo những kiến thức trước đây của con người về FAS thì hầu hết hình thức giả mạo bằng khuôn mặt (ví dụ: giấy in thường và màn hình điện tử) chỉ đơn thuần là không có thông tin chính xác về độ sâu của khuôn mặt, có thể được sử dụng làm tín hiệu để phát hiện có giả mạo hay không. Ngoài ra việc tạo ra và gán dãn cho các dữ liệu dạng 3D là rất tốn kém và không đủ chính xác, nhưng mặt nạ nhị phân (binary mask label)[2] có thể dễ dàng được tạo ra và tổng quát hơn nhiều, thông qua mặt nạ nhị phân, chúng ta có thể tìm thấy liệu cách giả mạo này có xuất hiện trong các trường hợp tương ứng hay không. Bên cạnh đánhgiá dựa trên độ sâu của khuôn mặt và mặt nạ nhị phân, có một số phương pháp phụ trợ thông tin khác như: pseudo reflection map, 3D points cloud map, ternary map và Fourier spectra.

Pixel-wise Supervision with Generative Model: Một xu hướng nổi bật là khai thác các mẫu giả mạo trực quan hiện có trong các mẫu giả mạo, nhằm cung cấp cách giải thích trực quan hơn về độ giả mạo của mẫu, từ đó phục vụ cho việc tìm hiểu các dấu hiệu giả mạo một cách đầy đủ.



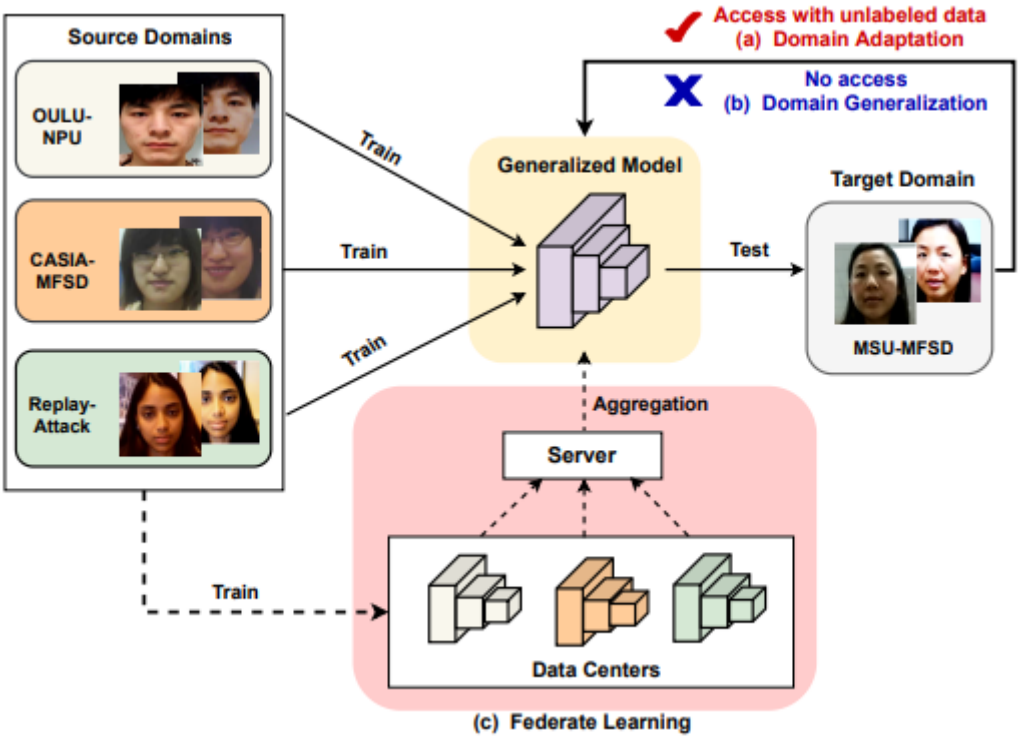
Hình 2.4. Các khuôn khổ học sâu phổ biến cho FAS [47]

(a) Direct supervision with binary cross entropy loss.
(b) Pixel-wise supervision with auxiliary tasks.
(c) Pixelwise supervision with generative model for implicit spoof.

Generalized Deep Learning Method

Các phương pháp FAS dựa trên học tập sâu từ đầu đến cuối phổ biến có thể khái quát hóa kém về các điều kiện ưu thế không nhìn thấy được (ví dụ: ánh sáng, hình dạng khuôn mặt và chất lượng máy ảnh) và các kiểu tấn công không xác định (ví dụ: mặt nạ làm bằng vật liệu mới). Do đó, các phương pháp này là không đáng tin cậy để áp dụng trong các ứng dụng thực tế với nhu cầu bảo mật mạnh mẽ. Do đó, ngày càng có nhiều nhà nghiên cứu tập trung vào việc nâng cao năng lực tổng quát hóa của các mô hình FAS. Mặt khác, các kỹ thuật miền thích ứng – domain adaptation (DA) và miền tổng quát hóa - domain generalization (DG) được tận dụng để phân loại giả mạo / không giả mạo mạnh mẽ dưới các biến thể miền không giới hạn.

Generalization to Unseen Domain



Hình 2.5. So sánh khung giữa các miền thích ứng - domain adaptation (DA), miền tổng quát hóa - domain generalization (DG) và học liên kết - federate learning (FL) [47].

Nhận thấy rằng:

(a) Các phương pháp DA cần mục tiêu (không được gắn nhãn) dữ liệu miền để tìm hiểu mô hình trong khi (b) Các phương pháp DG tìm hiểu mô hình tổng quát mà không có kiến thức từ miền đích. (c) FL coi mỗi miền nguồn như một trung tâm dữ liệu riêng và học mô hình tổng quát hóa trong máy chủ công cộng thông qua tổng hợp các mô hình từ các trung tâm dữ liệu cục bộ.

Sự thay đổi miền nghiêm trọng tồn tại giữa các miền nguồn và miền đích, điều này dễ dẫn đến hiệu suất kém trên tập dữ liệu mục tiêu thiên vị (ví dụ: MSU-MFSD) khi đào tạo mô hình sâu trực tiếp trên tập dữ liệu nguồn (ví dụ: OULU-NPU, CASIA-MFSD và Replay-Attack). Kỹ thuật *domain adaptation* tận dụng kiến thức từ miền đích để thu hẹp khoảng cách giữa miền nguồn và miền đích. Ngược lại, *domain generalization* giúp học trực tiếp biểu diễn tính năng tổng quát từ nhiều miền nguồn mà không cần bất kỳ quyền truy cập nào vào dữ liệu đích, điều này thực tế hơn cho việc triển khai trong thế giới thực. Để xem xét các vấn đề pháp lý và quyền riêng tư mà dữ liệu đào tạo thường không được chia sẻ trực tiếp giữa các chủ sở hữu dữ liệu (miền), *federate learning* được giới thiệu trong việc học các mô hình FAS tổng quát trong khi vẫn bảo toàn quyền riêng tư của dữ liệu.

Domain Adaptation: Kỹ thuật domain adaptation làm giảm bớt sự khác biệt giữa miền nguồn và miền đích. Trong hầu hết các phương pháp, việc phân phối các tính năng nguồn và đích được khớp trong một không gian đặc trưng đã học. Nếu các tính năng có sự phân bố tương tự, một bộ phân loại được đào tạo về các tính năng của mẫu nguồn cũng có thể được sử dụng để phân loại các mẫu không giả mạo/ giả mạo. Mặc dù, domain adaptation mạng lại lợi ích để giảm thiểu sự khác biệt về phân phối giữa nguồn và miền đích bằng cách sử dụng dữ liệu mục tiêu không được gắn nhãn, trong nhiều tình huống FAS trong thế giới thực, việc thu thập nhiều dữ liệu mục tiêu không được gắn nhãn (đặc biệt là các cuộc tấn công giả mạo) để đào tạo là rất khó và tốn kém.

Domain Generalization: Domain generalization giả định rằng tồn tại một không gian đặc trưng tổng quát bên dưới nhiều miền nguồn được nhìn thấy và miền đích không nhìn thấy nhưng lại được mô tả lại, trên đó mô hình đã học từ các miền nguồn được nhìn thấy có thể tổng quát hóa tốt cho miền đích không nhìn thấy. Domain generalization mang lại lợi ích cho các mô hình FAS hoạt động tốt trong miền không nhìn thấy, nhưng vẫn chưa biết liệu nó có làm giảm khả năng phân biệt để phát hiện giả mạo trong các tình huống đã thấy hay không.

Federate Learning: Một mô hình FAS tổng quát có thể được hoàn thiện khi được huấn luyện với hình ảnh khuôn mặt từ các vết thương khác nhau và các cách giả mạo khác nhau. Trên thực tế, dữ liệu về khuôn mặt đào tạo không được chia sẻ trực tiếp giữa các chủ sở hữu dữ liệu do các vấn đề pháp lý và quyền riêng tư. Để giải quyết thách thức này, *federate learning* một kỹ thuật học máy phân tán và bảo vệ quyền riêng tư, được giới thiệu trong FAS để đồng thời tận dụng nguồn thông tin không giả mạo / giả mạo phong phú có sẵn ở các chủ sở hữu dữ liệu khác nhau trong khi vẫn duy trì quyền riêng tư của dữ liệu. Cụ thể, mỗi trung tâm dữ liệu / chủ sở hữu tại địa phương đào tạo mô hình FAS của riêng mình. Sau đó, một máy chủ học một mô hình FAS toàn cầu bằng cách tổng hợp lặp đi lặp lại các bản cập nhật mô hình từ tất cả

các trung tâm dữ liệu mà không cần truy cập dữ liệu cá nhân ban đầu trong mỗi trung tâm. Mục đích của *federated learning* là để giải quyết vấn đề riêng tư của *các tập dữ liệu*. Tuy nhiên, nó bỏ qua các vấn đề về quyền riêng tư trong cấp độ *mô hình* cho FAS vì việc đào tạo mô hình toàn cầu cần nhiều nhóm để chia sẻ các mô hình địa phương của riêng họ, điều này có thể gây hại cho cạnh tranh thương mại.

Generalization to Unknown Attack Types: Bên cạnh các vấn đề chuyển đổi tên miền, các mô hình FAS dễ bị ảnh hưởng bởi các cách giả mạo mới xuất hiện trong các ứng dụng thực tế trong thế giới thực. Hầu hết các phương pháp học sâu trước đây đều hình thành FAS như một bài toán gần gũi để phát hiện các cách giả mạo được xác định trước khác nhau, các cách giả mạo này cần dữ liệu đào tạo quy mô lớn để bao gồm nhiều cuộc tấn công nhất có thể. Tuy nhiên, mô hình được đào tạo có thể dễ dàng trang bị một số cuộc tấn công thông thường (ví dụ: print - giả mạo bằng cách dùng tấm ảnh and replay – giả mạo bằng cách phát lại đoạn video giả mạo) và vẫn dễ bị tấn công bởi các loại tấn công không xác định (ví dụ: mask – dùng mặt nạ để giả mạo và makeup – trang điểm để giả mạo). Gần đây, nhiều nghiên cứu tập trung vào việc phát triển các mô hình FAS tổng quát để phát hiện tấn công giả mạo không xác định. Một mặt, *zero/few-shot learning* được sử dụng để cải thiện khả năng phát hiện giả mạo mới với rất ít hoặc thậm chí không có mẫu nào của các kiểu tấn công mục tiêu. Mặt khác, FAS cũng có thể được coi như một nhiệm vụ phân loại một lớp trong đó các mẫu khuôn mặt thật được tập hợp chặt chẽ và phát hiện bất thường được sử dụng để phát hiện các mẫu khuôn mặt giả mạo nằm ngoài phân bố.

Zero/Few-Shot Learning: Một cách đơn giản để phát hiện các cách giả mạo mới là kết nối mô hình FAS với đủ các mẫu của các cuộc tấn công mới. Tuy nhiên, việc thu thập dữ liệu được gắn nhãn cho mỗi cuộc tấn công mới là tốn kém và mất thời gian trong khi hành vi giả mạo tiếp tục phát triển. *Zero/Few-Shot Learning* nhằm mục đích tìm hiểu các đặc điểm tổng quát và phân biệt từ các cách giả mạo được xác định trước để phát hiện cách giả mạo mới chưa biết. *Few-Shot Learning* nhằm mục đích nhanh chóng thích ứng mô hình FAS với các cuộc tấn công mới bằng cách học từ cả các cách giả mạo xác định trước và rất ít mẫu được thu thập của cuộc tấn công mới. Mặc dù *Few-Shot Learning* có lợi nhưng các mô hình FAS cho phát hiện cuộc tấn công không xác định, hiệu suất giảm rõ ràng khi dữ liệu của các kiểu tấn công mục tiêu không có sẵn cho thích ứng (tức là trường hợp *Zero- Shot Learning*). Chúng tôi nhận thấy rằng việc phát hiện không thành công thường xảy ra trong các kiểu tấn công đầy thách thức (ví dụ: mặt nạ trong suốt, mắt vui nhộn và trang điểm), có chung phân bố ngoại hình tương tự với khuôn mặt thật.

Anomaly detection: Giả định rằng các mẫu trực tiếp thuộc loại bình thường vì chúng có chung biểu diễn tính năng tương tự và nhỏ gọn hơn trong khi các tính năng từ mẫu giả mạo có sự khác biệt lớn về phân bố trong không gian mẫu bất thường do phương

sai cao của các cách giả mạo. Dựa trên giả định, việc phát hiện bất thường trước hết thường đào tạo một bộ phân loại một lớp đáng tin cậy để phân cụm các mẫu trực tiếp một cách chính xác. Sau đó, bất kỳ mẫu nào (ví dụ: các cuộc tấn công không xác định) bên ngoài lề của cụm mẫu trực tiếp sẽ được phát hiện là giả mạo. Thay vì chỉ sử dụng các mẫu khuôn mặt thật, một số hoạt động nghiên cứu cũng đào tạo các hệ thống phát hiện bất thường tổng quát với cả hai mẫu khuôn mặt thật và khuôn mặt giả mạo. Mặc dù đáp ứng khả năng tổng quát hóa để phát hiện cuộc tấn công không xác định, các phương pháp FAS dựa trên phát hiện bất thường sẽ bị suy giảm khả năng phân biệt so với phân loại khuôn mặt thật / giả mạo thông thường trong các kịch bản thiết lập mở thế giới thực (tức là cả các cuộc tấn công đã biết và chưa biết).

Advanced Sensor

FAS dựa trên *Commercial RGB camera* là một giải pháp cân bằng tuyệt vời về mặt bảo mật và chi phí phần cứng trong các ứng dụng nhận dạng khuôn mặt hàng ngày (ví dụ: mở khóa di động và kiểm soát truy cập khu vực). Tuy nhiên, một số tình huống bảo mật cao (thanh toán bằng khuôn mặt và bảo vệ lối vào kho tiền) yêu cầu sai số chấp nhận sai rất thấp. Gần đây, các cảm biến tiên tiến với nhiều phương thức khác nhau được phát triển để tạo điều kiện thuận lợi cho FAS siêu an toàn. Hiệu suất và đánh giá cao của các cảm biến và mô-đun phần cứng khác nhau cho FAS về điều kiện môi trường (ánh sáng và khoảng cách) và các loại tấn công (in, phát lại, và mặt nạ 3D).

Uni-Modal Deep Learning upon Specialized Sensor

Dựa trên cảm biến / phần cứng chuyên dụng cho hình ảnh riêng biệt, các nhà nghiên cứu đã phát triển các phương pháp học sâu nhận biết cảm biến để có FAS hiệu quả. Một số nghiên cứu, Seo và Chung [24] đề xuất lightweight CNN Thermal Face để ước tính nhiệt độ khuôn mặt từ hình ảnh nhiệt và phát hiện sự giả mạo với nhiệt độ bất thường (ví dụ: ngoài phạm vi từ 36 đến 37 độ). Họ nhận thấy rằng hình ảnh nhiệt phù hợp hơn hình ảnh RGB để phát hiện tấn công kiểu *replay*. Rehman và cộng sự [44] giới thiệu một lớp chênh lệch trong CNN để trích xuất bản đồ chênh lệch động cho FAS dựa trên Stereo - âm thanh nổi, giúp cải thiện hiệu suất và độ mạnh của CNN đối với các loại mặt PA không xác định, ... Ngoài việc sử dụng phần cứng chuyên dụng như máy chiếu chấm hồng ngoại và máy ảnh chuyên dụng, một số phương pháp FAS được phát triển dựa trên các máy ảnh có thể nhìn thấy với đèn flash môi trường bổ sung. Ebihara và các cộng sự [11] thiết kế một bộ mô tả mới để đại diện cho các phản xạ đặc trưng và khuếch tán tận dụng các dấu hiệu khác biệt có và không có flash, hoạt động tốt hơn ResNet với đầu vào flash được ghép nối.

Multi-Modal Deep Learning

Với sự phát triển của công nghệ sản xuất và tích hợp phần cứng, các hệ thống FAS đa phương thức với chi phí chấp nhận được ngày càng được sử dụng nhiều hơn trong các ứng dụng thực tế. Trong khi đó, *multi-modal deep learning* trở nên phổ biến và tích cực trong cộng đồng nghiên cứu FAS.

Multi-Modal Fusion: Multi-Modal Fusion tập trung vào chiến lược tổng hợp cấp đặc tính. Zhang và cộng sự [36] đề xuất SD-Net sử dụng cơ chế tái trọng số tính năng để chọn các tính năng kênh thông tin trong số các phương thức RGB, độ sâu và NIR. Thay vì các phương pháp tiếp cận được đề cập ở trên bằng cách sử dụng hợp nhất cấp tính năng, có rất ít công trình xem xét hợp nhất cấp đầu vào và cấp quyết định. Hình ảnh tổng hợp được hợp nhất từ các phương thức tỷ lệ xám, độ sâu và NIR bằng cách xếp chồng các hình ảnh chuẩn hóa, và sau đó được đưa đến các máy phát hiện giả mạo. Liu và cộng sự [42] xây dựng khung phân tầng hai giai đoạn để thể hiện các tính năng dựa trên độ sâu và độ sâu từ các đầu vào độ sâu được xử lý trước nhiều lần (tức là chuẩn hóa, nhúng tỷ lệ và nhúng định hướng) và đầu vào VIS-NIR tổng hợp (tức là, ngăn xếp, tổng, và khác biệt).

Cross-Modal Translation: *Multi-Modal Deep Learning* cần các cảm biến bổ sung cho đầu vào hình ảnh khuôn mặt với các phương thức khác nhau. Tuy nhiên, trong một số trường hợp thông thường, chỉ có thể sử dụng các phương thức một phần (ví dụ: RGB). Để giải quyết vấn đề thiếu phương thức này ở giai đoạn suy luận, một số công trình áp dụng kỹ thuật dịch chéo phương thức để tạo ra dữ liệu phương thức còn thiếu cho FAS đa phương thức. Jiang và cộng sự [12] trước tiên đề xuất một chu trình dịch hình ảnh nhiều danh mục mới GAN tạo ra hình ảnh NIR tương ứng cho hình ảnh mặt RGB, sau đó tìm hiểu các tính năng hợp nhất từ các đầu vào xếp chồng lên nhau của RGB và hình ảnh NIR được tạo cho FAS.

2.2. Các phương pháp trích xuất đặc trưng hình ảnh

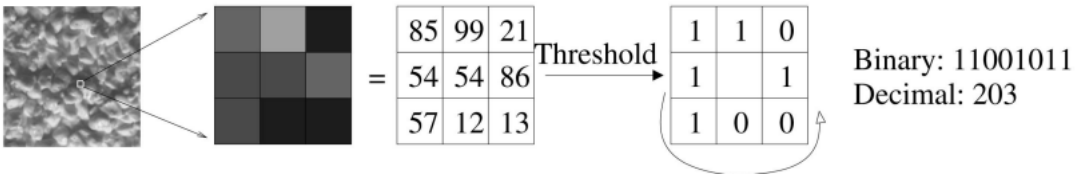
2.2.1. Giới thiệu

Trích xuất đặc trưng hình ảnh là quá trình ánh xạ một hình ảnh hình một vector đặc trưng. Trong xử lý ảnh số, trích xuất đặc trưng là rất quan trọng trong việc phân loại, các đặc trưng thường được trích xuất là các đặc điểm không gian, sự biến đổi, điểm biên và các đường viền.

2.2.2. Local Binary Pattern (LBP)

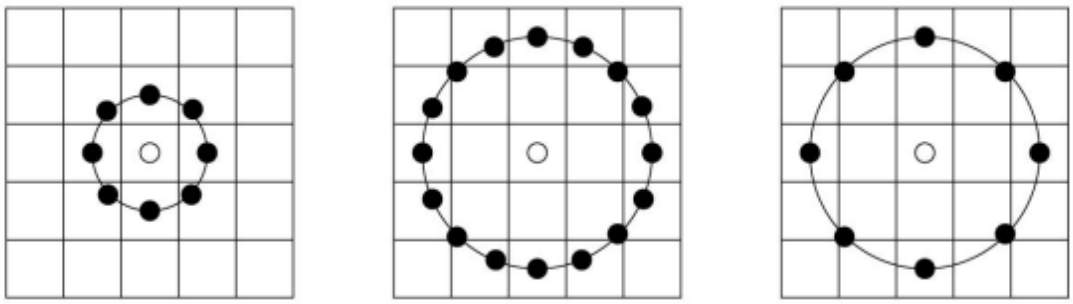
Thuật toán LBP ban đầu được thiết kế để mô tả kết cấu. Thuật toán sẽ gán nhãn cho mỗi pixel của hình ảnh với ngưỡng 3 vùng lân cận 3 của mỗi pixel với trung

tâm giá trị pixel và coi kết quả là số nhị phân. Sau đó, biểu đồ của nhãn có thể được sử dụng như một bộ mô tả kết cấu.



Hình 2.6. Ý tưởng cơ bản của thuật toán LBP

Việc xác định vùng lân cận cục bộ là một tập hợp các điểm lấy mẫu cách đều nhau trên một vòng tròn có tâm tại pixel được gắn nhãn cho phép bán kính và số lượng điểm lấy mẫu bất kỳ. Ký hiệu (P, R) sẽ được sử dụng cho vùng lân cận pixel có nghĩa là các điểm lấy mẫu P trên đường tròn bán kính R.



Hình 2.7. Các vùng lân cận hình tròn (8,1), (16,2) và (8,2).

Mẫu nhị phân cục bộ được gọi là đồng nhất nếu mẫu nhị phân chứa nhiều nhất hai chuyển đổi theo chiều dọc bit từ 0 sang 1 hoặc ngược lại khi mẫu bit được coi là hình tròn. Ví dụ: các mẫu 00000000 (0 chuyển tiếp), 01110000 (2 chuyển tiếp) và 11001111 (2 chuyển tiếp) là đồng nhất trong khi các mẫu 11001001 (4 chuyển tiếp) và 01010011 (6 chuyển tiếp) thì không. Trong tính toán biểu đồ LBP, các mẫu đồng nhất được gán một ngăn riêng cho mọi mẫu đồng nhất và tất cả các mẫu không đồng nhất được gán cho một ngăn duy nhất.

2.2.3. Histogram of oriented gradients (HOG)

HOG [32] là một phương pháp trừu tượng hóa đối tượng bằng cách trích xuất ra những đặc trưng của đối tượng đó và bỏ đi những thông tin không hữu ích, được sử dụng chủ yếu để mô tả hình dạng và sự xuất hiện của một đối tượng trong ảnh. Bản chất của phương pháp HOG là sử dụng thông tin về sự phân bố của các cường độ gradient (intensity gradient) hoặc của hướng biên (edge directins) để mô tả các đối tượng cục bộ trong ảnh. Các thuật toán HOG sẽ chia nhỏ một bức ảnh thành các vùng con, được gọi là cells, tính toán histogram về các hướng của gradients cho các

điểm nằm trong mỗi cell. Đặc trưng của bức ảnh tạo ra bằng cách ghép các histogram lại với nhau. Để tăng cường hiệu năng nhận dạng, các histogram cục bộ có thể được chuẩn hóa về độ tương phản bằng cách tính một ngưỡng cường độ trong một vùng lớn hơn cell, gọi là các blocks và sử dụng giá trị ngưỡng đó để chuẩn hóa tất cả các cell trong khối. Kết quả sau bước chuẩn hóa sẽ là một vector đặc trưng có tính bất biến cao hơn đối với các thay đổi về điều kiện ánh sáng.

2.2.4. Image Quality Assessment

Image Quality Assessment (IQA) như một phương pháp bảo vệ chống lại các cuộc tấn công giả mạo khuôn mặt. Trong tình trạng tiên tiến hiện nay, lý do đằng sau việc sử dụng các tính năng IQA để phát hiện sự sống được hỗ trợ bởi ba yếu tố:

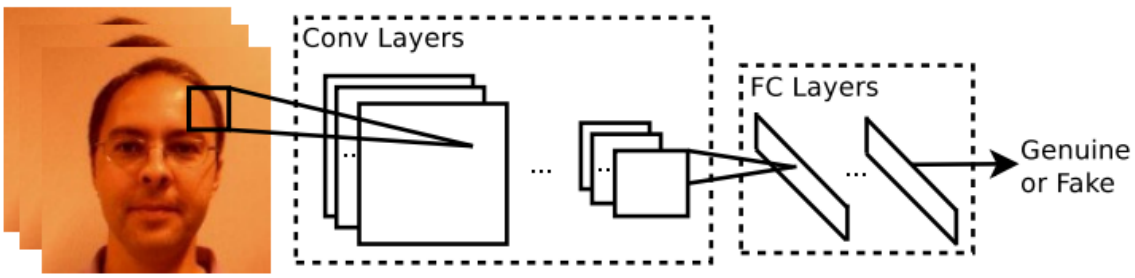
• Chất lượng hình ảnh đã được sử dụng thành công trong các công trình trước đây để phát hiện thao tác hình ảnh [34] và phân tích mật mã [18] trong lĩnh vực pháp y. Ở một mức độ nhất định, nhiều cuộc tấn công giả mạo khuôn mặt, đặc biệt là những cuộc tấn công liên quan đến việc chụp ảnh khuôn mặt hiển thị trong thiết bị 2D (ví dụ: các cuộc tấn công giả mạo với hình ảnh khuôn mặt in), có thể được coi là một loại thao tác hình ảnh có thể hiệu quả được phát hiện, như được thể hiện trong công trình nghiên cứu hiện tại, bằng cách sử dụng các tính năng chất lượng khác nhau.

• Ngoài các nghiên cứu trước đây trong lĩnh vực pháp y, các tính năng khác nhau đo lường các thuộc tính chất lượng đặc trưng của đặc điểm đã được sử dụng cho mục đích phát hiện độ sống trong các ứng dụng vân tay và mống mắt [21].

• Các nhà quan sát của con người thường đề cập đến "sự khác biệt về hình dáng bên ngoài" của các mẫu thật và giả để phân biệt giữa chúng. Các thước đo và phương pháp khác nhau được thiết kế cho IQA nhằm ước tính một cách khách quan và đáng tin cậy sự xuất hiện của hình ảnh mà con người cảm nhận được.

2.2.5. Convolutional Neural Network (CNNs)

Convolutional Neural Network (CNNs – Mạng nơ-ron tích chập) là một trong những mô hình Deep Learning tiên tiến, được sử dụng nhiều trong các bài toán nhận dạng các object trong ảnh. Jianwei Yang và các cộng sự [25] đã đề xuất một mô hình CNNs cho bài toán Face Anti-Spoofing, trên tập dữ liệu Replay Attacks.



Hình 2.8. Mô hình CNNs cho bài toán Face Anti-Spoofing của Jianwei Yang và các cộng sự [25]

2.2.6. Temporal-based Methods

Một trong những giải pháp sớm nhất để chống giả mạo khuôn mặt là dựa trên các dấu hiệu thái dương như nháy mắt [27]. Các phương pháp như [33] theo dõi chuyển động của miệng và môi để phát hiện nét mặt. Trong khi các phương pháp này có hiệu quả đối với các cuộc tấn công trên giấy điển hình, nhưng sẽ không hiệu quả khi những kẻ tấn công cố thực hiện một cuộc tấn công phát lại hoặc một cuộc tấn công bằng giấy với phần mắt / miệng bị cắt.

Cũng có những phương pháp dựa vào các đặc điểm tổng quát hơn, thay vì chuyển động khuôn mặt cụ thể. Cách tiếp cận phổ biến nhất là nối khung (frame concatenation). Ngoài ra, có một số công trình đề xuất các tính năng cụ thể theo thời gian, ví dụ, Haralick Features [1], Motion Mag [35], and Optical Flow [41].

2.2.7. Remote photoplethysmography

Remote photoplethysmography (rPPG) là kỹ thuật theo dõi các tín hiệu quan trọng, chẳng hạn như nhịp tim, mà không cần bất kỳ sự tiếp xúc nào với da người. Nghiên cứu bắt đầu với video khuôn mặt không có chuyển động hoặc ánh sáng thay đổi thành video có nhiều biến thể. Trong [13], Haan và các cộng sự ước tính tín hiệu rPPG từ video mặt RGB với các thay đổi về ánh sáng và chuyển động. Nó sử dụng sự khác biệt màu sắc để loại bỏ sự phản xạ đặc trưng và ước tính hai tín hiệu sắc độ trực giao. Sau khi áp dụng Band Pass Filter (BPM), tỷ lệ tín hiệu màu sắc được sử dụng để tính tín hiệu rPPG.

2.3. Các phương pháp phân loại

2.3.1. Đối với input bài toán là ảnh đơn lẻ

Mạng nơ-ron tích chập (CNN) được xem là một phương pháp vượt trội so với các mô hình khác trong lĩnh vực thị giác máy tính.

Trong bài báo [27] [28] CNN đóng vai trò như một công cụ trích xuất đặc trưng trên khuôn mặt (features extractor). Trong [25], Yang và các cộng sự đề xuất mô hình CNN như một công cụ phân loại để chống giả mạo khuôn mặt. Các dữ liệu khuôn mặt với các tỷ lệ trong thang đo không gian (spatial scales) khác nhau được xếp chồng lên nhau làm đầu vào và nhãn thật/giả được chỉ định làm đầu ra.

Javier Hernandez-Ortega và các cộng sự [17] đã sử dụng mô hình CNN kết hợp với các model đã được hoàn thiện sẵn (pretrained-models), cụ thể ở đây là ResNet-50, sau đó thông qua kỹ thuật fine tuning họ đã cho ra mô hình FaceQNet. Ngoài ra thì

Alotaibi cùng các cộng sự [5] đã giới thiệu một mô hình CNN để nhận dạng trên từng bức ảnh. Trong đó, giá trị của pixel hình ảnh đầu vào được chuẩn hóa trong đoạn [0, 1], đặt giá trị trung bình gần bằng 0 và phương sai gần bằng 1.

Tuy nhiên, so với các vấn đề liên quan đến khuôn mặt khác, chẳng hạn như nhận dạng khuôn mặt [29] và căn chỉnh khuôn mặt [3], về cơ bản vẫn còn ít nỗ lực và khám phá hơn về việc chống giả mạo khuôn mặt bằng kỹ thuật học sâu.

2.3.2. Đối với input bài toán là chuỗi ảnh tuần tự hoặc video

Mạng nơ-ron hồi quy (RNN) có thể xử lý thông tin dạng chuỗi (sequence/ time-series), có nghĩa là RNN có thể mang thông tin của frame (ảnh) từ state trước đến các state sau, rồi ở state cuối cùng là sự kết hợp của tất cả các ảnh để đưa ra dự đoán. Mô hình Long-short Term Memory (LSTM) là một biến thể của RNN, trong đó thông tin nào quan trọng sẽ được gửi vào và dùng bất kỳ đâu khi cần. Do đó mô hình LSTM có cả short term memory và long term memory.

Thay vì sử dụng bức ảnh đơn chiếc, các phương pháp sử dụng đầu vào ở dạng chuỗi ảnh/video như trong [48] được xem như bài toán phân loại video (video classification problem). Đầu vào của mô hình của là chuỗi các khung hình video (x_1, x_2, ..., x_n) và đầu ra là nhãn 0 hoặc 1 tương ứng với giả/thật.

Để trích xuất các đặc điểm thời gian từ chuỗi video, Baccouche và các cộng sự [30] áp dụng các lớp LSTM trên các đặc trưng biến đổi tính năng bất biến theo quy mô (Scale-invariant Feature Transform) và biểu diễn Bag-of words (BoW) để phân loại hành động thật hoặc giả. Ngoài ra, Gan và cộng sự [15] đã sử dụng mô hình 3D-CNN để xử lý đầu vào là chuỗi video liên tục và kết quả tích chập của nó là nhiều hơn một bản đồ đặc trưng (feature map), có thể duy trì tốt các đặc trưng của chuỗi video liên tục trong chiều thời gian.

CHƯƠNG 3. THỰC NGHIỆM CHƯƠNG TRÌNH

3.1. Cài đặt

Cấu hình máy thực hiện:

Bảng 3.1.1. Cấu hình máy để huấn luyện.

CPU	Intel Core I5 11400H
GPU	Nvidia GeForce RTX 3050 Laptop
RAM	16Gb
Ổ cứng	256GB SSD

Phiên bản: Python 3.9.7

IDE: Pycharm

3.2. Dữ liệu

3.2.1. Yêu cầu dữ liệu

Dữ liệu để kiểm tra là một tập các hình ảnh (gồm 10 hình ảnh) phải cho máy tính nhìn thấy được trên 2/3 khuôn mặt thật hay khuôn mặt giả mạo (hình ảnh, video được ghi lại trên thiết bị di động) mà không phải là một đối tượng nào khác và trong điều kiện không thiếu sáng, hoặc ngược sáng khi quay video trực tiếp trên không gian thực, thiết bị ghi hình phải bảo đảm có chất lượng hình ảnh trên 360p.

Hình 3.1.1. Ví dụ một số dữ liệu đã đạt yêu cầu

Hình 3.1.2. Ví dụ một số dữ liệu không đạt yêu cầu

3.2.2. Thống kê dữ liệu

Bộ dữ liệu: Replay-Attack [19]

Mô tả bộ dữ liệu:

1300 video về các cuộc tấn công và truy cập sau đó được chia theo cách sau:

Tập train: gồm 60 video khuôn mặt thật và 300 video khuôn mặt giả trong các điều kiện ánh sáng khác nhau.

Tập valid: chứa 60 video khuôn mặt thật và 300 video khuôn mặt giả trong các điều kiện ánh sáng khác nhau.

Tập test: gồm 80 video khuôn mặt thật và 400 video khuôn mặt giả trong các điều kiện ánh sáng khác nhau.

Tổ chức lưu trữ: một thư mục tên *dataset* lưu trữ hai thư mục là real (chứa các đoạn video có khuôn mặt thật, số lượng: 300 video); attack (chứa các đoạn video có khuôn mặt giả mạo, số lượng: 100 video). Mỗi video có khung hình 240x320, thời lượng là 9 giây với độ phân giải là 24 FPS.

Bảng 3.1.2. Thống kê số lượng video training và testing.

	Training (80%)	Testing (20%)
Real	80	20
Attack	320	80

3.2.3. Tiền xử lý dữ liệu

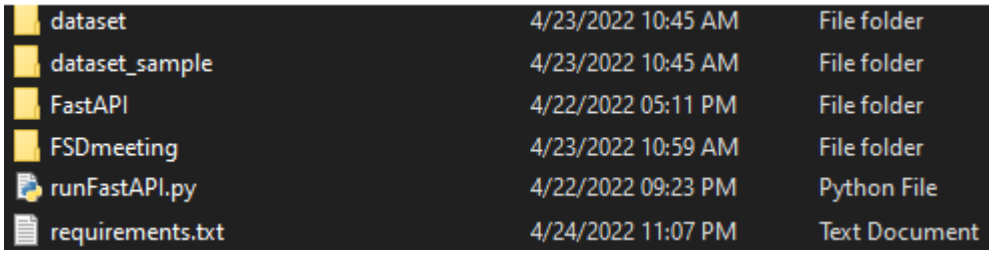
Tiền xử lý: Như đã trình bày ở trên mỗi video sẽ có thời lượng là 9 giây với độ phân giải là 24 FPS, , với mỗi khung hình là 240x320 nhưng vậy mỗi đoạn video sẽ luôn có 200 khung hình (do không lấy tối đa là 216, vì có một số đoạn video bị mất vài khung hình do ở giây cuối cùng), do vậy mỗi video sẽ được chia thành 20 mẫu, với một mẫu gồm tập hợp 10 khung hình và sẽ được một tập tin NumPy (.npy) với số chiều là [240,320,3,10]. Trong đó 240 và 320 là các thông số của kích thước của một khung hình, gồm 3 kênh màu (red, green, blue), và 10 khung hình.

Tổ chức dữ trữ: một thư mục *dataset_sample* lưu trữ hai thư mục là train (chứa các file.npy phục vụ cho việc huấn luyện, số lượng: 8000 file); valid (chứa các file.npy phục vụ cho việc kiểm tra, số lượng: 2000 file).

Bảng 3.1.3. Thống kê số lượng file.npy training và testing.

	Training (80%)	Testing (20%)
Real	1600	400
Attack	6400	1600

3.2.4. Cấu trúc thư mục



Hình 3.1.3. Cấu trúc thư mục của đề tài.

dataset lưu trữ hai thư mục là real (chứa các đoạn video có khuôn mặt thật, số lượng: 300 video); attack (chứa các đoạn video có khuôn mặt giả mạo, số lượng: 100 video). Mỗi video có khung hình 240x320, thời lượng là 9 giây với độ phân giải là 24 FPS.

Hình 3.1.4. Một số video trong thư mục attack.

Hình 3.1.5. Một số video trong thư mục real.

dataset_sample lưu trữ hai thư mục là train (chứa các file.npy phục vụ cho việc huấn luyện, số lượng: 8000 file); valid (chứa các file.npy phục vụ cho việc kiểm tra, số lượng: 2000 file). Mỗi file.npy sẽ được đặt tên theo dạng:

x_namefile_index_type

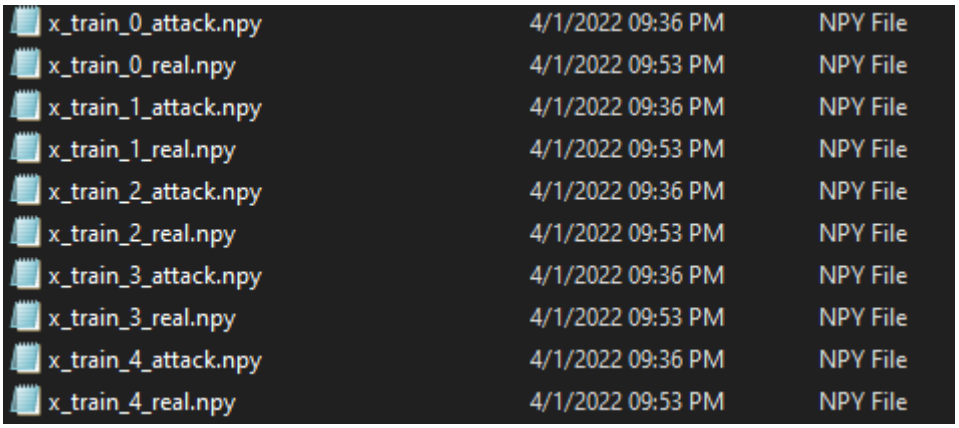
Trong đó:

x là kí tự đặc biệt để nhận biết đây là file dữ liệu.

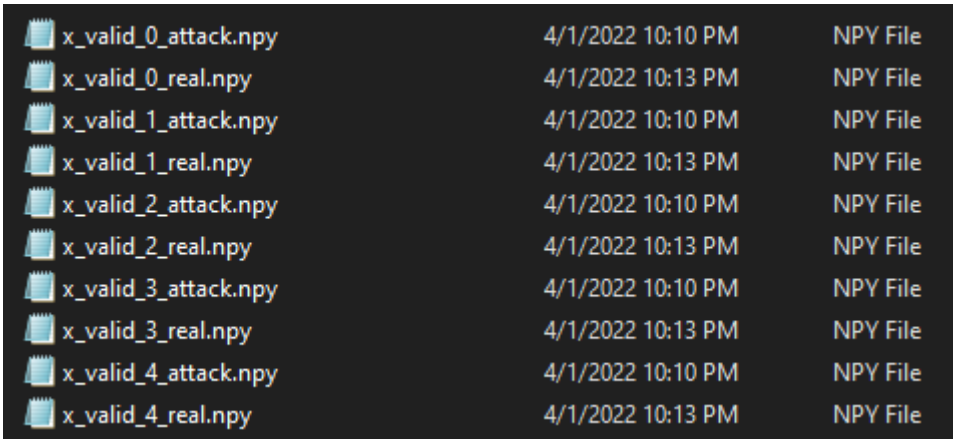
namefile chứa hai giá trị là “train”, “vaild” để nhận biết file đó thuộc tập nào.

index cho biết thông tin số thứ tự của của file đó.

type chứa hai giá trị “real”, “attack”.



Hình 3.1.6. Một số file.npy trong thư mục train



Hình 3.1.7. Một số file.npy trong thư mục vaild

FastAPI là thư mục cho phép việc truyền các khung hình từ Client (web) về mô hình để xử lý.

FSDmeeting là một chương trình ứng dụng web, nhằm mô phỏng quá trình điểm danh trên web.

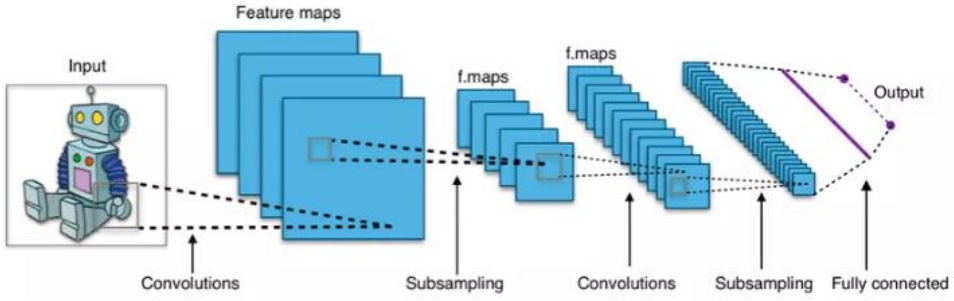
runFastAPI.py là chương trình để khởi động *FastAPI*.

requirements.txt là một file chứa các thư viện với phiên bản cần thiết cho python.

3.2.5. Phân tích mô hình

Thiết lập mô hình

Mô hình sử dụng mạng Convolutional Neural Network (CNN hoặc ConvNet) được tạm dịch là: Mạng nơ ron tích chập. Đây được xem là một trong những mô hình của Deep Learning – tập hợp các thuật toán để có mô hình dữ liệu trừu tượng hóa ở mức cao bằng cách sử dụng nhiều lớp xử lý cấu trúc phức tạp. Hiểu đơn giản, CNN là một mô hình của mạng nơ-ron học sâu, được áp dụng phổ biến nhất để phân tích hình ảnh trực quan.



Hình 3.1.8. Mô hình của một mạng CNN phổ biến

Bảng 3.1.4. Các tham số trong mô hình

Tham số	**Giá trị**
INIT_LR	1.e-4
EPOCHS	20
BS	1

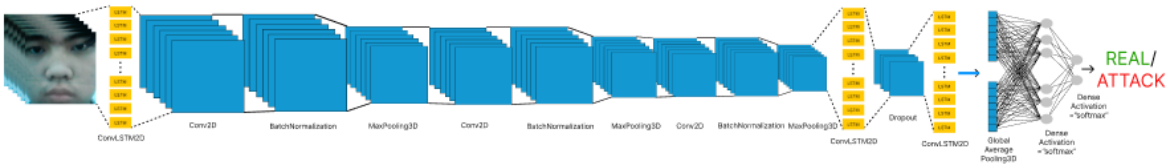
Trong đó:

Learning rate (INIT_LR) – Tốc độ học là một siêu tham số sử dụng trong việc huấn luyện các mạng nơ ron. Giá trị của nó là một số dương, thường nằm trong khoảng giữa 0 và 1. Tốc độ học kiểm soát tốc độ mô hình thay đổi các trọng số để phù hợp với bài toán. Tốc độ học lớn giúp mạng nơ ron được huấn luyện nhanh hơn nhưng cũng có thể làm giảm độ chính xác.

Một *Epoch* được tính là khi chúng ta đưa tất cả dữ liệu vào mạng neural network 1 lần. Khi dữ liệu quá lớn, chúng ta không thể đưa hết mỗi lần tất cả tập dữ liệu vào để huấn luyện được. Buộc lòng chúng ta phải chia nhỏ tập dữ liệu ra thành các batch (size nhỏ hơn).

Batch size (BS) là số lượng mẫu dữ liệu trong một batch. Ở đây, khái niệm batch size và số lượng batch (number of batch) là hoàn toàn khác nhau. Như đã nói ở trên, chúng ta không thể đưa hết toàn bộ dữ liệu vào huấn luyện trong 1 epoch, vì vậy chúng ta cần phải chia tập dữ liệu thành các phần (number of batch), mỗi phần có kích thước là batch size.

Mô hình được thiết lập như sau:



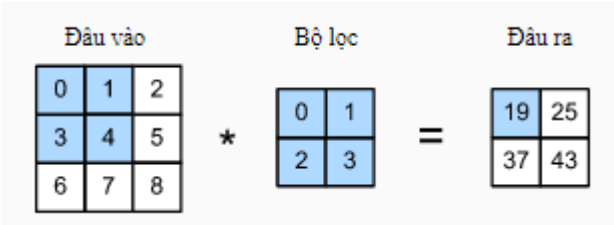
Hình 3.1.9. Mô hình training của đề tài

Trong đó:

ConvLSTM2D layer: LSTM (Mạng bộ nhớ dài-ngắn - Long Short-Term Memory networks) - là một dạng đặc biệt của RNN (Mạng nơ-ron hồi quy - Recurrent Neural Network), nó có khả năng học được các phụ thuộc xa. Ví dụ, dự đoán chữ cuối cùng trong đoạn: “I grew up in France… I speak fluent French.”. Rõ ràng là các thông tin gần (” I speak fluent”) chỉ có phép ta biết được đằng sau nó sẽ là tên của một ngôn ngữ nào đó, còn không thể nào biết được đó là tiếng gì. Muốn biết là tiếng gì, thì ta

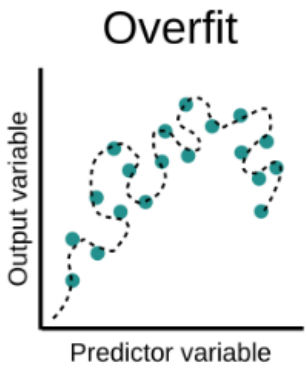
cần phải có thêm ngữ cảnh “I grew up in France” nữa mới có thể suy luận được. Rõ ràng là khoảng cách phj thuộc xa. Tương tự như một lớp LSTM, nhưng các phép biến đổi đầu vào và phép biến đổi lặp lại đều có tính chất tích chập.

Conv2D layer: Convolution 2D - tầng tích chập thực hiện phép toán tương quan chéo giữa đầu vào và hạt nhân, sau đó cộng thêm một hệ số điều chỉnh để có được đầu ra. Hai tham số của tầng tích chập là hạt nhân và hệ số điều chỉnh. Khi huấn luyện mô hình chứa các tầng tích chập, ta thường khởi tạo hạt nhân ngẫu nhiên.



Hình 3.1.10. Ví dụ về tính chập 2 chiều

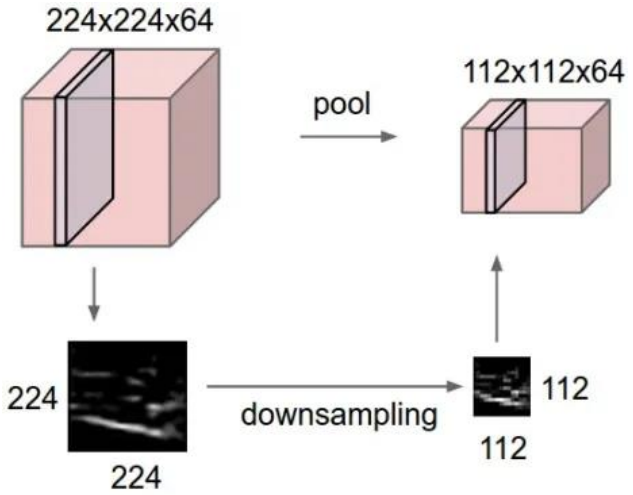
Dropout layer: là lớp sẽ bỏ qua một vài unit trong suốt quá trình train trong mô hình, những unit bị bổ qua được lựa chọn ngẫu nhiên, điều này sẽ chống lỗi over-fitting (quá khớp).



Hình 3.1.11. Một ví dụ minh họa cho over-fitting lỗi quá khớp.

BatchNormalization layer: Lớp chuẩn hóa đầu vào, chuẩn hóa hàng loạt áp dụng một phép biến đổi tuân theo xác suất phân phối chuẩn duy trì giá trị đầu ra trung bình gần bằng 0 và độ lệch chuẩn đầu ra gần bằng 1.

MaxPooling3D layer: Giảm kích thước dữ liệu nhưng vẫn giữ được các thuộc tính quan trọng.

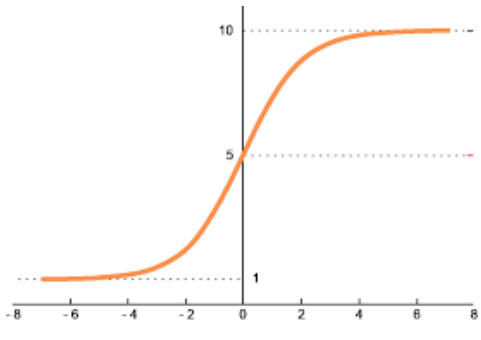


Hình 3.1.12. Ví dụ về lớp MaxPooling3D

GlobalAveragePooling3D: là lớp chuyển đổi dữ liệu ở lớp trước đó thành một vector một chiều, để nhập dữ liệu vào các lớp tiếp theo.

Desen layer: là lớp mạng nơ-ron nhân tạo trong đó mỗi nơ-ron nhận đầu vào từ tất cả nơ-ron của lớp trước đó. Activation là hàm kích hoạt (hàm tính toán) trong việc áp chức năng kích hoạt các phần tử trong vùng thỏa điều kiện tham số kích hoạt.

Bảng 3.1.5. Một số hàm kích hoạt đã được sử dụng trong mô hình

	Softmax
Công thức	$a_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$
Đồ thị	

Đánh giá mô hình

Bảng 3.1.6. Consfusion Matrix

		Predicted	
		Attack	Real
Groundtruth	Attack	368	132
	Real	137	363

Bảng trên là kết quả được thông kê trên tập kiểm thử 1000 file.npy, bao gồm 500 file.npy gán nhãn là real và 500 file.npy gán nhãn là attack.

Từ ma trận nhầm lẫn ta thấy:

Tỉ lệ mà mô hình dự đoán nhầm lẫn Attack (khuôn mặt giả mạo) là Real (khuôn mặt thật) là: 132/500 tức 26.4%

Tỉ lệ mà mô hình dự đoán nhầm lẫn Real (khuôn mặt thật) là Attack (khuôn mặt giả mạo) là: 137/500 tức 27.4%

Từ đó, nhận thấy rằng tỉ lệ phát hiện sai Attack (khuôn mặt giả mạo) gần giống như tỉ lệ phát hiện sai Real (khuôn mặt thật), nên mô hình này có tính khả thi cao, tuy nhiên cần tối ưu nhiều hơn để giảm độ sai số đến mức tối thiểu.

Bảng 3.1.7. Kết quả của mô hình

F1 Scores	0.73
Precision Scores	0.781
Recall Scores	0.726

Trong đó:

Tập dữ liệu kiểm thử có 1200 file.npy được gán nhãn là Real (khuôn mặt thật) hoặc là Attack (khuôn mặt giả mạo).

Giá trị của Precision Scores cho biết hệ thống dự đoán file.npy đó đúng nhãn đến 78.1%, hay nói cách khác là chỉ có 21.9% hệ thống dự đoán dự đoán nhầm nhãn khi ta cho tập dữ liệu kiểm thử gồm cả hai nhãn

Giá trị của Recall Scores cho biết hệ thống có thể phát hiện được tới 72.6% là đúng nhãn khi ta cho tập dữ liệu chỉ đúng một nhãn

Giá trị của Accuracy cho biết mức độ hiệu quả của toàn bộ hệ thống đối với việc phát hiện gian lận điểm danh trong lớp học trực tuyến.

3.2.6. Thực thi ngoài thực tế

Cách triển khai hệ thống

Cài đặt các môi trường cần thiết: nodejs, python.

Cài đặt các thư viện cần thiết:

- Nodejs: npm install
- Python: pip install -r requirements.txt

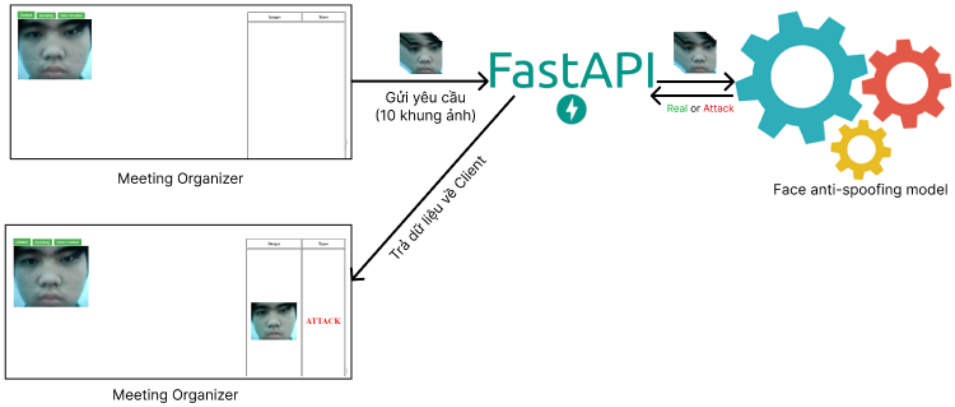
Khởi động file *runFastAPI.py* để khởi chạy FastAPI. Sau đó, vào thư mục *FSDmeeting*, tiếp tục chạy đồng thời hai câu lệnh:

npm start - Để khởi động ứng dụng.

ngrok.exe http 8000 - Để host trên localhost:8000

Hình 3.1.13. Giao diện ứng dụng web sau khi được khởi động

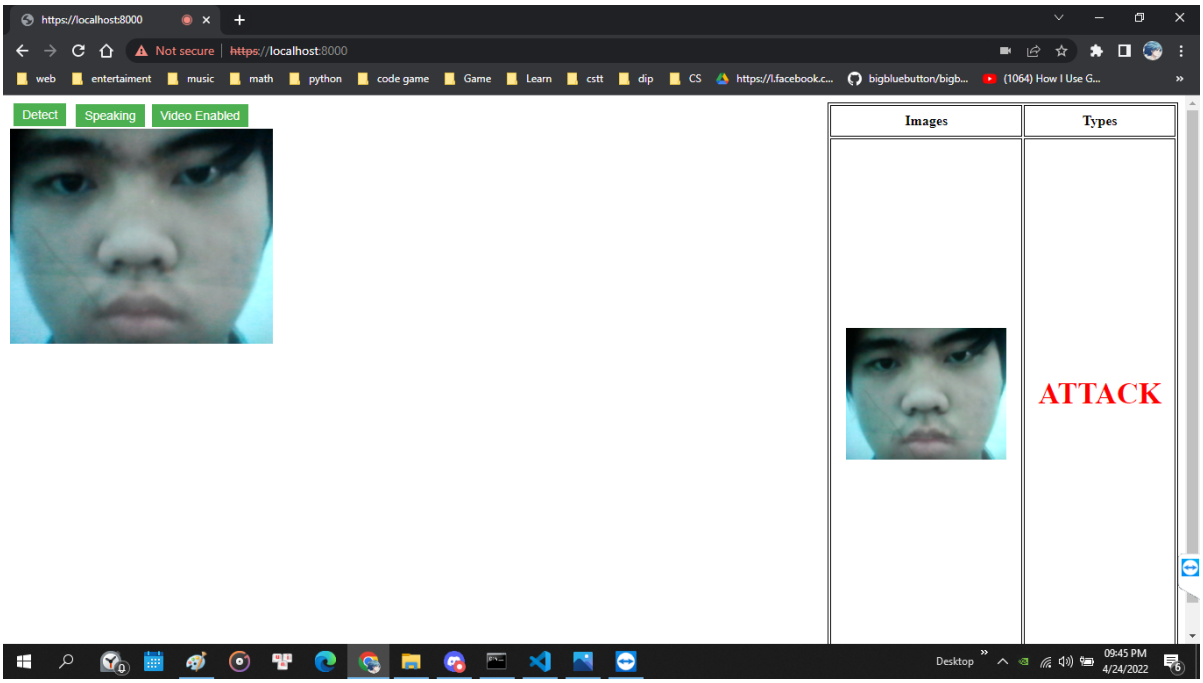
Cách hệ thống hoạt động



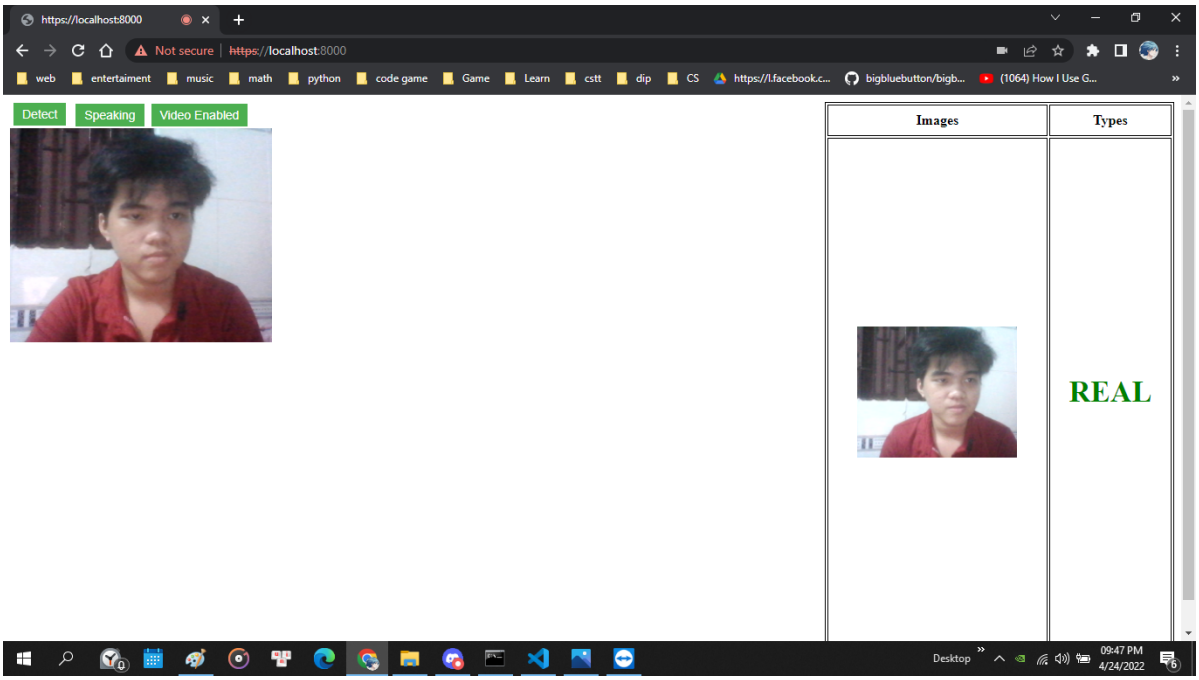
Hình 3.1.14. Sơ đồ hoạt động của hệ thống phát hiện giả mạo trong lớp học trực tuyến

Sau khi nhấn nút “Detect”, Meeting Organizer (Người tổ chức lớp học) sẽ gửi 10 khung hình đến Server (FastAPI). Sau đó mô hình Face anti-spoofing sẽ dự đoán kết quả là Real – khuôn mặt thật hoặc Attack – khuôn mặt giả mạo từ dữ liệu được nhận từ Server. Bước kế tiếp, hệ thống sẽ trả về kết quả về cho Server và sẽ hiện thị kết quả cho Meeting Organizer.

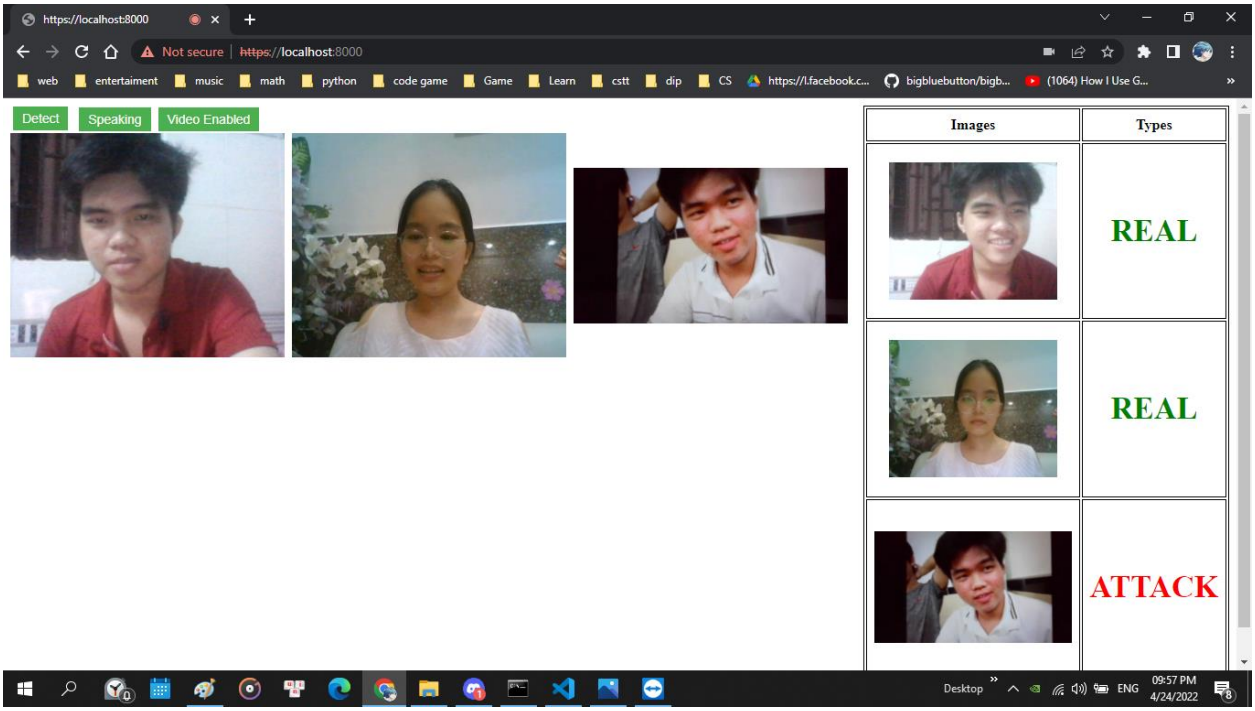
Kết quả thu được



Hình 3.1.15. Kết quả thu được đối với khuôn mặt giả mạo



Hình 3.1.16. Kết quả thu được đối với khuôn mặt thật



Hình 3.1.17. Kết quả thu được đối với nhiều khuôn mặt

CHƯƠNG 4. KẾT LUẬN

4.1. Đánh giá kết quả thực nghiệm

4.1.1. Đóng góp

Nghiên cứu đã trình bày về vấn đề phát hiện gian lận điểm danh trong lớp học trực tuyến, cụ thể là bài toán nhận dạng khuôn mặt thật giả, một bài toán quan trọng trong lĩnh vực thị giác máy tính. Đề tài tập trung nghiên cứu, phát triển về lý thuyết và ứng dụng đối với bài toán nhận dạng, đề xuất một số mô hình và giải pháp nhằm nâng cao hiệu quả đưa ra một số khung làm việc phục vụ cho quá trình nhận dạng. Nâng cao chất lượng nhận dạng đối tượng liên quan tới giả mạo khuôn mặt bằng cách thu nhập số lượng dữ liệu hình ảnh đủ lớn, tiến hành thực nghiệm ngoài thực tế để ghi chép những những sai sót, lập bảng thống kê về các tham số để chương trình được tối ưu hơn.

4.1.2. Hạn chế

Hệ thống trong đề tài mà chúng em đang nghiên cứu vẫn còn một số hạn chế như không thể nhận diện hoặc nhân diện sai trong nhiều điều kiện khác nhau, trong nhiều kiểu tấn công khác nhau. Mặc dù nghiên cứu còn những hạn chế và thiếu sót nhất định, nhưng chúng em sẽ tiếp tục cố gắng hoàn thiện trong thời gian tới. Mặc dù mô hình đã đạt được kết quả sơ bộ ban đầu, tuy nhiên vì nhiều lý do khác nhau, nhóm nghiên cứu vẫn chưa thể tìm ra được bộ thông số cho ra được kết quả tốt nhất có thể. Do lúc đầu mô hình đã không tích hợp với mô hình nhận dạng khuôn mặt trước khi đưa vào mô hình mà nhóm đang xây dựng, nên sẽ gặp vài kết quả không mong muốn. Ngoài ra, mô hình còn bị phụ thuộc vào chất lượng hình ảnh để có thể đưa ra quyết định, cũng như việc xử lý có phần sai lệch trong việc xác định độ rõ của bức ảnh/video.

4.2. Hướng phát triển

Hiện tại, do hạn chế về thời gian, nghiên cứu dừng lại thử nghiệm với tập dữ liệu có kích cỡ nhỏ cũng như chưa tìm ra được thông số huấn luyện cho ra kết quả tốt nhất. Nhóm nghiên cứu dự định phát triển đề tài với quy mô lớn hơn về mặt kỹ thuật, cũng như cố gắng tìm ra nhiều bộ thông số huấn luyện nhằm tìm ra kết quả tốt nhất có thể.

Mở rộng và thử nghiệm trên các kiểu thực thể và mối quan hệ thực thể khác. Cải tiến áp dụng các dạng biểu đồ trực quan khác nhau trong việc phân tích các thực thể.

Sau khi trải qua mùa dịch COVID thì hệ thống vẫn có thể được tiếp tục ứng dụng không chỉ riêng việc điểm danh, mà còn có thể ứng dụng vào các vấn đề liên quan tới bảo mật sinh trắc học với quy mô mô hình lớn hơn.

Ngoài ra hệ thống còn có thể được ứng dụng và phát triển trên các trang web học trực tuyến, các ứng dụng xác thực khuôn mặt cũng như kết hợp với hệ thống nhận dạng danh tính của con người trong không gian thực.

TÀI LIỆU THAM KHẢO

[1] A. Agarwal, R. Singh, and M. Vatsa. Face anti-spoofing using Haralick features. In BTAS, 2016.

[2] A. George and S. Marcel, "Deep pixel-wise binary supervision for face presentation attack detection," in ICB, no. CONF, 2019.

[3] A. Jourabloo and X. Liu. Large-pose face alignment via cnnbased dense 3d model fitting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4188–4196, 2016.

[4] A. Liu, C. Zhao, Z. Yu, J. Wan, A. Su, X. Liu, Z. Tan, S. Escalera, J. Xing, Y. Liang et al., "Contrastive context-aware learning for 3d high-fidelity mask face presentation attack detection," arXiv preprint arXiv:2104.06148, 2021.

[5] Alotaibi, Aziz, and Ausif Mahmood. "Deep face liveness detection based on nonlinear diffusion using convolution neural network." Signal, Image and Video Processing 11.4 (2017): 713-720.

[6] B. Peixoto, C. Michelassi, and A. Rocha, "Face liveness detection under bad illumination conditions," in Image Processing (ICIP), 2011, pp. 3557–3560.

[7] Bharadwaj, Samarth, et al. "Computationally efficient face spoofing detection with motion magnification." Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2013.

[8] Chengrui Wang, Weihong Deng. "Representative Forgery Mining for Fake Face Detection" This CVPR 2021 paper is the Open Access version, provided by the Computer Vision Foundation. Except for this watermark, it is identical to the accepted version; the final published version of the proceedings is available on IEEE Xplore.

[9] Chingovska, I., Anjos, A., & Marcel, S. (2012, September). On the effectiveness of local binary patterns in face anti-spoofing. In 2012 BIOSIG-proceedings of the international conference of biometrics special interest group (BIOSIG) (pp. 1-7). IEEE.

[10] Chingovska, I., Anjos, A., & Marcel, S. (2013). Anti-spoofing in action: joint operation with a verification system. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 98-104).

[11] F. Ebihara, K. Sakurai, and H. Imaoka, "Specular-and diffusereflection-based face spoofing detection for mobile devices," inIJCB. IEEE, 2020.

[12] F. Jiang, P. Liu, X. Shao, and X. Zhou, "Face anti-spoofing with generated near-infrared images," Multimedia Tools and Applications, vol. 79, no. 29, pp. 21 299–21 323, 2020.

[13] G. de Haan and V. Jeanne. Robust pulse rate from chrominance-based rPPG. IEEE Trans. Biomedical Engineering, 60(10):2878–2886, 2013.

[14] G. Pan, Z. Wu, and L. Sun. Liveness detection for face recognition. In K. Delac, M. Grgic, and M. S. Bartlett, editors, Recent Advances in Face Recognition, page Chapter 9. IN-TECH, 2008.

[15] Gan, Junying, et al. “3d convolutional neural network based on face anti-spoofing.” 2017 2nd international conference on multimedia and image processing (ICMIP). IEEE, 2017.

[16] Hadid, A. (2014). Face biometrics under spoofing attacks: Vulnerabilities, countermeasures, open issues, and research directions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 113-118).

[17] Hernandez-Ortega, Javier, et al. FaceQnet Quality assessment for face recognition based on deep learning. 2019 International Conference on Biometrics (ICB). IEEE, 2019.

[18] I. Avcibas et al., “Steganalysis using image quality metrics,” IEEE Trans. on Image Processing, vol. 12, pp. 221–229, 2003.

[19] I. Chingovska, A. Anjos, S. Marcel,"On the Effectiveness of Local Binary Patterns in Face Anti-spoofing"; IEEE BIOSIG, 2012.

[20] J. Galbally and S. Marcel, “Face anti-spoofing based on general image quality assessment,” in ICPR. IEEE, 2014.

[21] J. Galbally et al., “A high performance fingerprint liveness detection method based on quality related features,” Future Generation Computer Systems, vol. 28, pp. 311–321, 2012.

[22] J. Komulainen, A. Hadid, and M. Pietikainen, “Context based face anti-spoofing,” in Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on, IEEE, 2013, pp. 1–8.

[23] J. Maatta, A. Hadid, and M. Pietikainen, “Face spoofing detection from single images using micro-texture analysis,” in Biometrics (IJCB), 2011 International Joint Conference on, oct. 2011, pp. 1 –7.

[24] J. Seo and I.-J. Chung, “Face liveness detection using thermal face-cnn with external knowledge,” Symmetry, vol. 11, no. 3, p.360, 2019.

[25] J. Yang, Z. Lei, and S. S. Z. Li, “Learn Convolutional Neural Network for Face Anti-Spoofing,” ArXiv Prepr., no. 1408-5601, p. 8, 2014.

[26] K. Kollreider, H. Fronthaler, and J. Bign. Non-intrusive liveness detection by face images. Image and Vision Computing, 27:233–244, 2009.

[27] K. Patel, H. Han, and A. K. Jain. Cross-database face anti-spoofing with robust feature representation. In CCBR, pages 611–619, 2016.

[28] L. Li, X. Feng, Z. Boulkenafet, Z. Xia, M. Li, and A. Hadid.An original face anti-spoofing approach using partial convolutional neural network. In Image

Processing Theory Tools and Applications (IPTA), 2016 6th International Conferenceon, pages 1–6. IEEE, 2016.

[29] L. Tran, X. Yin, and X. Liu. Disentangled representation learning gan for pose-invariant face recognition. In Proceeding of IEEE Computer Vision and Pattern Recognition, Honolulu, HI, July 2017.

[30] M. Baccouche, F. Mamalet, C. Wolf, C. Garcia, and A. Baskurt. Action classification in soccer videos with long short-term memory recurrent neural networks. In Artificial Neural Networks–ICANN 2010, pages 154–159. Springer, 2010.

[31] M. De Marsico, M. Nappi, D. Riccio, and J. L. Dugelay, "Moving face spoofing detection via 3D projective invariants," Proc. - 2012 5th IAPR Int. Con! Biometrics, ICB 2012, pp. 73-78,2012.

[32] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in CVPR. IEEE, 2005.

[33] R. Shao, X. Lan, and P. C. Yuen. Deep convolutional dynamic texture learning with adaptive channeldiscriminability for 3D mask face anti-spoofing. In IJCB,2017.

[34] S. Bayram et al., "Image manipulation detection," Journal of Electronic Imaging, vol. 15, p. 041102, 2006.

[35] S. Bharadwaj, T. Dhamecha, M. Vatsa, and R. Singh. Face anti-spoofing via motion magnification and multifeature videolet aggregation. 2014.

[36] S. Zhang, A. Liu, J. Wan, Y. Liang, G. Guo, S. Escalera, H. J. Escalante, and S. Z. Li, "Casia-surf: A large-scale multi-modal benchmark for face anti-spoofing," TBIOM, vol. 2, no. 2, pp. 182–193, 2020.

[37] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," TPAMI, no. 12, pp. 2037–2041, 2006.

[38] T. Brox and J. Malik, "Large displacement optical flow: descriptor matching in variational motion estimation," TPAMI, vol. 33, no. 3, pp. 500–513, 2010.

[39] T. de Freitas Pereira, A. Anjos, J. M. De Martino, and S. Marcel, "Can face anti-spoofing countermeasures work in a real world scenario?" in ICB, June 2013.

[40] Usman Muhammad, Tuomas Holmberg, Wheidima Carneiro de Melo, Abdenour Hadid. "Face Anti-Spoofing via Sample Learning Based Recurrent Neural Network (RNN) ". Center for Machine Vision and Signal Analysis, University of Oulu Oulu, Finland, 2019.

[41] W. Bao, H. Li, N. Li, and W. Jiang. A liveness detection method for face recognition based on optical flow field. In IASP, pages 233–236, 2009.

[42] W. Liu, X. Wei, T. Lei, X. Wang, H. Meng, and A. K. Nandi, “Data fusion based two-stage cascade framework for multi-modality face anti-spoofing,” TCDS, 2021.

[43] X. Tan, Y. Li, J. Liu, and L. Jiang, “Face liveness detection from a single image with sparse low rank bilinear discriminative model,” in Computer Vision–ECCV 2010, Springer, 2010, pp. 504–517.

[44] Y. A. U. Rehman, L.-M. Po, and M. Liu, “Slnet: Stereo face liveness detection via dynamic disparity-maps and convolutional neural network,” Expert Systems with Applications, vol. 142, p. 113002, 2020.

[45] Y. Atoum, Y. Liu, A. Jourabloo, and X. Liu, “Face anti-spoofing using patch and depth-based cnns,” in IJCB, 2017.

[46] Y. Zhang, Z. Yin, Y. Li, G. Yin, J. Yan, J. Shao, and Z. Liu, “Celeba-spoof: Large-scale face anti-spoofing dataset with rich annotations,” in ECCV. Springer, 2020.

[47] Yu, Zitong et al. “Deep Learning for Face Anti-Spoofing: A Survey.” ArXiv abs/2106.14948 (2021): n. pag.

[48] Z. Xu, S. Li, and W. Deng, “Learning temporal features using LSTMCNN architecture for face anti-spoofing,” Proc. - 3rd IAPR Asian Conference of Pattern Recognition, ACPR 2015, pp. 141-145,2016.

[49] Zhiwei Zhang, Junjie Yan, Sifei Liu, Zhen Lei, Dong Yi, Stan Z. Li. A Face Antispoofing Database with Diverse Attacks. In proceedings of the 5th IAPR International Conference on Biometrics (ICB’12), New Delhi, India, 2012.