NHẬN DIỆN KHUÔN MẶT GIẢ MẠO SỬ DỤNG KỸ THUẬT HỌC SÂU
FACE LIVENESS DETECTION USING DEEP LEARNING

PHẠM MINH NHỰT[1a], PHAN ANH CANG[1], NGUYỄN THÁI NGHE[2]
[1]Trường Đại học Sư phạm Kỹ thuật Vĩnh Long
[2]Đại học Cần Thơ
[a]Tác giả liên hệ: nhutpm9@fe.edu.vn

Nhận bài (Received): 24/03/2023; Phản biện (Reviewed): 28/03/2023; Chấp nhận (Accepted): 24/04/2023

TÓM TẮT

Trong thời đại kỹ thuật số hiện nay, việc sử dụng nhận diện khuôn mặt đã trở nên phổ biến hơn bao giờ hết trong nhiều lĩnh vực khoa học công nghệ. Điều này đã làm giảm bớt nhiều vấn đề liên quan và hiệu quả hơn về nhận dạng người. Tuy nhiên, vẫn còn một số lỗ hổng liên quan đến vấn đề về an ninh thông tin, bảo mật trong hệ thống nhận diện khuôn mặt. Bài báo này đề xuất phương pháp chống giả mạo khuôn mặt và phát hiện sự sống dựa trên việc phát triển mô hình CNN (Mạng nơ-ron tích chập) sử dụng tập dữ liệu CelebA Spoof. Kết quả thực nghiệm của mô hình đề xuất đạt độ chính xác trung bình là 87%. Điều này có thể góp phần vào việc cải thiện hiệu suất của công nghệ nhận dạng khuôn mặt.

Từ khóa: Học sâu, Nhận diện khuôn mặt giả mạo, Phát hiện sự sống khuôn mặt, Mạng nơ-ron tích chập

ABSTRACT

In the current modern digital era, facial recognition is used more commonly than ever in numerous sectors of technology and science. Compared to other biometric identification techniques such as fingerprints, iris scans, etc., facial recognition systems have reduced many related issues and are more effective for human identification. However, some gaps related to information security and confidentiality in the human face recognition system still exist. This paper proposes an anti-forgery and liveness detection method based on the development of the CNN (Convolutional Neural Network) model using the CelebA Spoof dataset. The experimental results of the proposed model achieve an average accuracy of 87%. This can contribute to improving the performance of facial recognition technology.

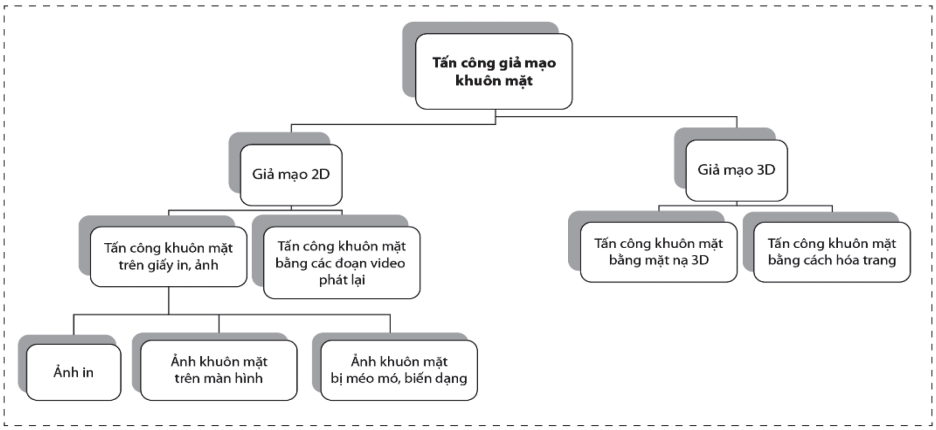
Keywords: *deep learning, face spoofing, liveness detection, convolution neural network*

1. MỞ ĐẦU

1.1. Giới thiệu bài toán

Hệ thống nhận dạng khuôn mặt có lợi thế so với các công nghệ sinh trắc học khác như mống mắt và dấu vân tay vì nó rất thuận tiện, không tiếp xúc và không xâm lấn[1].

Một hệ thống nhận dạng khuôn mặt thông thường không có các biện pháp chống giả mạo, nó rất dễ bị lỗi nếu có nỗ lực tấn công giả mạo vào hệ thống. Trường hợp còn tồi tệ hơn với các hệ thống nhận dạng khuôn mặt, vì ảnh/video của người dùng đã đăng ký có thể dễ dàng lấy được qua internet, mạng xã hội hoặc chỉ cần chụp khuôn mặt của họ bằng máy ảnh, ngay cả khi không có sự đồng ý. Điều này dẫn đến một trong những vấn đề thách thức: cho một Hình ảnh hoặc Video, và Mặt nạ 3D được chụp từ máy ảnh, chứa các khuôn mặt người. Vậy làm cách nào để trích xuất các đặc trưng để có thể phân biệt được khuôn mặt thật và khuôn mặt giả mạo một cách hiệu quả? Hình 1. Trong bài báo này, chúng tôi giải quyết vấn đề này và đề xuất việc sử dụng các mạng nơ-ron sâu để thực hiện việc phát hiện khuôn mặt và chống giả mạo.



Hình 1. Các kiểu giả mạo khuôn mặt. [2]

Cách tiếp cận được thực hiện bằng cách sử dụng bộ dữ liệu có tên là bộ dữ liệu chống giả mạo CelebA. Đây là bộ dữ liệu mới được xuất bản vào năm 2020. Nó bao gồm 625.537 bức ảnh về 10.177 chủ đề. Vì vậy, Nó đủ lớn để tập luyện và có kết quả rất tốt. Thử nghiệm của chúng tôi được thực hiện bằng cách sử dụng CNN tuần tự bao gồm nhiều giai đoạn. Đầu tiên là phần tiền xử lý chịu trách nhiệm về phần trích xuất tính năng. Sau đó, việc tăng cường sẽ được thực hiện. Sau đó, dữ liệu sẽ được chuyển đến các lớp Mạng nơ-ron tích chập để huấn luyện. Sau đó, mô hình sẽ có thể phân loại xem hình ảnh là trực tiếp hay ảnh giả mạo. Xác thực chéo (CV) cũng được sử dụng để kiểm tra khả năng dự đoán dữ liệu mới của mô hình đối với dữ liệu mới mà không được sử dụng để ước tính dữ liệu đó, nhằm đánh dấu các vấn đề có thể xảy ra trong mô hình như quá khớp hoặc sai lệch lựa chọn.

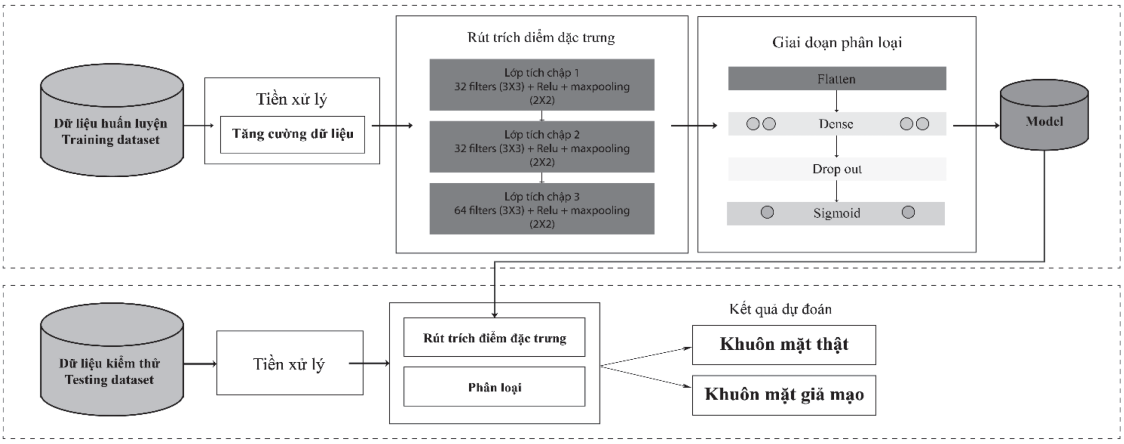
1.2. Những nghiên cứu liên quan

Sengur et al.[3], trình bày cách tiếp cận mạng nơ ron tích chập (CNN) để phát hiện sự sống động của khuôn mặt và nó đã được áp dụng trên bộ dữ liệu NUAA được chia thành 5761 ảnh để thử nghiệm và 3491 ảnh để huấn luyện, mô hình này đạt độ chính xác 83,38%. Akbulut et al.[4] được trình bày như trong hình 3 bên dưới, triển khai CNN cũng trên tập dữ liệu NUAA nhưng họ đã chia tập dữ liệu thành 5761 ảnh để thử nghiệm và 1748 ảnh để huấn luyện và mô hình đã đạt được một độ chính xác là 84,04% nhưng khi họ sử dụng các trường tiếp nhận cục bộ (LRF)-ELM, họ đạt được độ chính xác là 76,31% trên cùng một bộ dữ liệu. Komulainen và Pietikanen. [5] đã sử dụng cùng một bộ dữ liệu NUAA và chia thành 3362 mẫu thử nghiệm và 5761 mẫu đào tạo nhưng mô hình này sử dụng

mẫu nhị phân cục bộ (LBP) dựa trên phân tích kết cấu vi mô và điều này phụ thuộc vào một số tính năng như tính năng và biểu đồ sóng con Gabor của Gradient. Biểu diễn này giúp họ dễ dàng sử dụng các bộ phân loại máy vectơ hỗ trợ tuyến tính (SVM) trở nên dễ dàng đối với họ, kết quả của hệ thống được đề xuất được tính bằng cách sử dụng tỷ lệ lỗi bằng nhau (EER) và đường cong (AUC). Vì vậy, kết quả cho các công việc của họ đã có sự cải thiện về EER từ 2,8% lên 1,1% và AUC từ 0,995 lên 0,999. Wen et al.[6] đã đề xuất một thuật toán hệ thống hiệu quả dựa trên phân tích biến dạng hình ảnh, hệ thống trích xuất bốn tính năng quan trọng từ vectơ đặc trưng IDA là độ mờ, khoảnh khắc màu, phản xạ gương và đa dạng màu sắc, bộ dữ liệu họ sử dụng là bộ dữ liệu tấn công phát lại bao gồm trong số 1.300 bản ghi video về cả nỗ lực truy cập thực và tấn công của 50 đối tượng khác nhau, mô hình này đạt được độ chính xác là (TPR trung bình=90,5% & FAR=0,01). Liu et al.[7] đã sử dụng cùng bộ dữ liệu Replay-attack nhưng họ sử dụng thuật toán Mạng cây sâu (DTN), vì nó cho kết quả chính xác tổng thể là 95,5%. Bộ dữ liệu Replay-attack được sử dụng bởi Ito et al.[8] khi họ sử dụng thuật toán CNN để trích xuất các đặc trưng từ hình ảnh và phân loại chúng thành thật và giả bằng Máy vectơ hỗ trợ (SVM), mô hình chạy trên bộ dữ liệu Replay-attack bao gồm một tập hợp các chuỗi video được lấy từ 50 đối tượng của cả kịch bản thực và giả. Do đó, tỷ lệ lỗi (EER) khi sử dụng các phương pháp này lần lượt là 2,3% và 0,75&. Pinto et al.[9] đã trình bày một cách tiếp cận mới được sử dụng bằng cách sử dụng một kỹ thuật tận dụng các dấu hiệu nhiễu được tạo bởi video được thu lại. Phương pháp này cũng sử dụng phổ Fourier, sau đó là tính toán nhịp điệu thị giác thu được nhiễu được tạo ra từ các video đã ghi. Tập dữ liệu được sử dụng trong bài báo này là Print Attack Database. Nó bao gồm 200 video của 50 người dùng khác nhau và 200 video về các cuộc tấn công giả mạo.

1.3. Phương pháp đề xuất

Mô hình được đề xuất sử dụng kiến trúc học sâu tuần tự, trong đó các lớp tích chập và tổng hợp được xếp chồng lên nhau từ đầu vào đến đầu ra. Hệ thống đề xuất gồm 2 giai đoạn là đào tạo và kiểm thử. Chi tiết hệ thống được biểu diễn trong hình 2.



Hình 2. Kiến trúc hệ thống được đề xuất

1.3.1 Giai đoạn đào tạo

a. Tiền xử lý: Trong giai đoạn này chúng tôi tiến hành tăng cường dữ liệu bằng các biện pháp như lật ảnh, xoay ảnh. Điều này làm dữ liệu sẽ trở nên nhiều hơn và tăng cường độ chính xác khi đào tạo.

b. Rút trích đặc trưng và đào tạo: Dựa vào dữ liệu đã được tiền xử lý, chúng tôi tiến hành rút trích đặc trưng của các khuôn mặt thật và giả và đưa qua mô hình học sâu để thực hiện đào tạo.

1.3.2 Giai đoạn kiểm thử

Trong giai đoạn này, chúng tôi thực hiện dự đoán và phát hiện khuôn mặt giả mạo sử dụng tập dữ liệu kiểm thử dựa trên mô hình đào tạo ở giai đoạn 1. Hình ảnh đầu vào được trích xuất từ camera sau đó thực hiện việc phát hiện khuôn mặt và tiến hành đưa qua mô hình đào tạo để đưa ra kết quả dự đoán.

2. KẾT QUẢ NGHIÊN CỨU

2.1. Môi trường cài đặt

Các thử nghiệm được thực hiện trên máy được trang bị Intel Core i7-4790 @ 3.6GHZ, bộ nhớ 16GB và thẻ GPU NVIDIA GTX1060. Máy chạy trên Windows 11.

2.2. Tập dữ liệu thực nghiệm và kịch bản áp dụng

CelebA-Spoof chứa 625.537 ảnh của 10.177 đối tượng, lớn hơn bất kỳ bộ dữ liệu hiện có nào liên quan đến lĩnh vực này. Hình ảnh giả mạo được lấy từ 8 bối cảnh với hơn 10 cảm biến. CelebA-Spoof chứa 10 chú thích loại giả mạo, cũng như 40 chú thích thuộc tính được kế thừa từ tập dữ liệu CelebA gốc [10]

2.3. Kết quả đạt được sau khi huấn luyện

2.3.1. Kết quả đào tạo

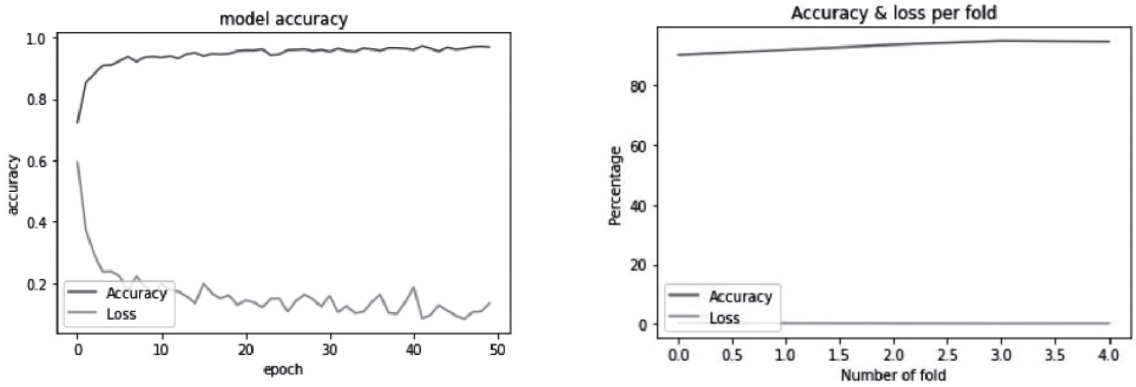
Bảng 1 cho thấy accracy, precision, recall và F1- score của quá trình đào tạo, kiểm thử và giá trị cross-validation. Kết quả cho thấy độ chính xác là 87% trong quá trình thử nghiệm, 96,9% trong quá trình đào tạo và 94,7% trong quá trình xác thực chéo. Sự khác biệt giữa độ chính xác kiểm tra và đào tạo là không lớn. Tuy nhiên, các thước đo khác (chẳng hạn như độ chính xác, thu hồi và thước đo F1) trong thử nghiệm cao hơn so với xác thực chéo. Lý do là xác thực chéo kiểm tra hệ thống trên một phần dữ liệu được nhìn thấy trước đó. Các giá trị thấp của độ chính xác, thu hồi và thước đo F1 cho thấy bộ phân loại cần nhiều dữ liệu hơn để được đào tạo tốt trong quá trình xác thực chéo. Hơn nữa, xác thực chéo holdout nhằm mục đích sử dụng thay vì xác thực chéo 5 lần.

Bảng 1. Kết quả thực nghiệm các phương pháp

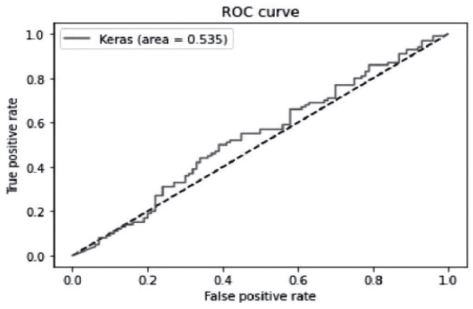
Model	Type	Đo lường			
		Độ chính xác (Accuracy)	Dự đoán (Precision)	Thu hồi (Recall)	Thước đo F (F measure)
Mạng neural tích chập	Đào tạo	96.9%	96.9%	96.4%	88.2%
Mạng neural tích chập	Thử nghiệm	87.0%	93.6%	78.2%	86.8%
Mạng neural tích chập	Xác thực chéo	94.7%	50.9%	81.9%	59.9%

2.3.2. Độ đo Loss và Accuracy

Bộ dữ liệu được chia thành 2000 hình ảnh để đào tạo và 200 hình ảnh để thử nghiệm. Số lượng kỷ nguyên là 50. Chúng tôi đã sử dụng 50 kỷ nguyên để huấn luyện mô hình. Hình 4 cho thấy độ chính xác của thử nghiệm và tổn thất trên mỗi kỷ nguyên. Người ta nhận thấy rằng độ chính xác tăng lên và tổn thất giảm đi. Đường cong ROC [11] cũng được tạo ra như thể hiện trong Hình 5. Diện tích dưới đường cong là 0,535.



Hình 4. Kiểm tra độ chính xác và hàm mất mát của mô hình trên tập huấn luyện (training set) và trên tập kiểm thử (validation set)

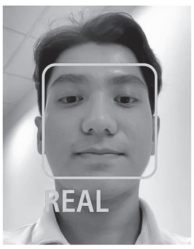


Hình 5. Biểu đồ ROC curve cho thử nghiệm

2.3.3. Một số kết quả thực nghiệm

Để kiểm tra độ chính xác hệ thống, chúng tôi tiến hành thực nghiệm hệ thống trên một số trường hợp: real time và sử dụng hình ảnh của người được chụp. Kết quả thực nghiệm được thể hiện trong Bảng 2

Bảng 2. Một số kết quả thực nghiệm

L1	L2	L3	L4
			
N1	N2	N3	N4
			

Kết quả thực nghiệm ta thấy được, xét 4 trường hợp trong 2 ngữ cảnh, hệ thống đề xuất cho kết quả khả thi trong việc phát hiện khuôn mặt thật và giả cụ thể trường hợp real-time hệ thống cho kết quả chính xác trong việc đưa ra kết quả tuy nhiên trong trường hợp giả mạo hệ thống lại có sự sai lệch khi có một dự đoán cho kết quả

sai. Điều này cho thấy độ chính xác của hệ thống còn hạn chế và cần được cải thiện để có thể góp phần xây dựng hệ thống chống giả mạo.

3. KẾT LUẬN VÀ HƯỚNG TƯƠNG LAI

Nhiều hệ thống bảo mật đã quan tâm đến tính năng Phát hiện sự sống của khuôn mặt để ngăn chặn giả mạo khuôn mặt. Phương pháp CNN được đề xuất đã đạt được độ chính xác tương đối chấp nhận được đối với thử nghiệm là 87% và đối với xác thực chéo là 94,7%. Hệ thống bao gồm hai giai đoạn là giai đoạn rút trích đặc trưng và giai đoạn phân loại. Tập dữ liệu được sử dụng là CelebA-Spoof xuất hiện vào năm 2020. Tập huấn luyện chứa 2000 hình ảnh được chia thành 1000 hình ảnh trực tiếp và 1000 hình ảnh giả mạo để có được tập hợp cân bằng và tập hợp thử nghiệm chứa 200 hình ảnh.

Nhiều kỹ thuật mới dự định sẽ được sử dụng trong công việc trong tương lai chẳng hạn như mạng lưới thần kinh viên nang. Các mô hình CNN khác như mô hình được đào tạo trước sẽ được so sánh với CNN thông thường để đạt được hiệu suất cao nhất. Một loại giả mạo khuôn mặt khác sẽ được sử dụng như video sử dụng thuật toán học sâu để phát hiện giả mạo khuôn mặt trong các video, chẳng hạn như RNN, một chủ đề quan trọng về giả mạo khuôn mặt.

TÀI LIỆU THAM KHẢO

[1] D. Wen, H. Han, and A. K. Jain, "Face spoof detection with image distortion analysis," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 746–761, Apr. 2015, doi: 10.1109/TIFS.2015.2400395.

[2] S. Kumar, S. Singh, and J. Kumar, "A comparative study on face spoofing attacks," in *Proceeding - IEEE International Conference on Computing, Communication and Automation, ICCCA 2017*, Institute of Electrical and Electronics Engineers Inc., Dec. 2017, pp. 1104–1108. doi: 10.1109/CCAA.2017.8229961.

[3] Y. Chen, H. Jiang, C. Li, X. Jia, S. Member, and P. Ghamisi, "IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING 1 Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks."

[4] Y. Akbulut, A. Şengür, Ü. Budak, and S. Ekici, "Deep learning based face liveness detection in videos," in *IDAP 2017 - International Artificial Intelligence and Data Processing Symposium*, Institute of Electrical and Electronics Engineers Inc., Oct. 2017. doi: 10.1109/IDAP.2017.8090202.

[5] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face Spoofing Detection Using Colour Texture Analysis," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 8, pp. 1818–1830, Aug. 2016, doi: 10.1109/TIFS.2016.2555286.

[6] D. Wen, H. Han, and A. K. Jain, "Face spoof detection with image distortion analysis," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 746–761, Apr. 2015, doi: 10.1109/TIFS.2015.2400395.

[7] Y. Liu, J. Stehouwer, A. Jourabloo, and X. Liu, "Deep Tree Learning for Zero-shot Face Anti-Spoofing."

[8] M. Asia-Pacific Signal and Information Processing Association. Annual Summit and Conference (9th : 2017 : Kuala Lumpur, Asia-Pacific Signal and Information Processing Association, and Institute of Electrical and Electronics Engineers., *Ninth Asia-Pacific Signal and Information Processing Association Annual Summit and Conference : APSIPA ASC 2017 : proceedings : 12-15 December 2017, Kuala Lumpur, Malaysia.*

[9] A. Da Silva Pinto, H. Pedrini, W. Schwartz, and A. Rocha, "Video-based face spoofing detection through visual rhythm analysis," in *Brazilian Symposium of Computer Graphic and Image Processing*, 2012, pp. 221–228. doi: 10.1109/SIBGRAPI.2012.38.

[10] Y. Zhang *et al.*, "CelebA-Spoof: Large-Scale Face Anti-Spoofing Dataset with Rich Annotations," Jul. 2020, [Online]. Available: http://arxiv.org/abs/2007.12342

[11]"The meaning and use of the area under a receiver operating characteristic (roc) curve".