PHÁT HIỆN GIẢ MẠO KHUÔN MẶT SỬ DỤNG CÔNG NGHỆ TRÍ TUỆ NHÂN TẠO

Lê Văn Hào[1], Trịnh Thị Anh Loan[1], Lê Việt Nam[1], Nguyễn Đức Toàn[2]

TÓM TẮT

Phát hiện giả mạo khuôn mặt là một bước quan trọng trong các hệ thống nhận dạng khuôn mặt. Gần đây, sự phát triển của các mạng nơ-ron tích chập (Convolution Neural Networks - CNNs) đang cho thấy kết quả vượt trội so với các phương pháp truyền thống sử dụng các thuật toán xử lý ảnh khác. Bên cạnh đó, xu hướng di động hóa đang đòi hỏi các phần mềm cần đáp ứng được khả năng thực thi trên các thiết bị có năng lực hạn chế như điện thoại, thiết bị nhúng. Trong bài báo này, chúng tôi đề xuất mạng nơ-ron tích chập hduNet được phát triển từ mạng MobilenetV2 của Google để phát hiện giả mạo khuôn mặt nhằm hướng tới mục tiêu chạy trên các thiết bị phần cứng yếu không sử dụng bộ xử lý đồ họa (GPU) mà vẫn đáp ứng độ chính xác. Ngoài ra, chúng tôi cũng bổ sung thêm 5000 dữ liệu ảnh mang đặc trưng của người châu Á để tăng cường hiệu quả và tránh việc mất cân bằng trong bộ dữ liệu chuẩn LCC_FASD [1] vốn chỉ thiên về ảnh giả mạo với 16885 ảnh giả mạo và chỉ 1942 ảnh thật. Cuối cùng, chúng tôi thực hiện đánh giá hiệu quả của mạng đề xuất trên tập dữ liệu mới thu thập và ứng dụng kết quả trong một ứng dụng thực tiễn cụ thể.

Từ khóa: *Giả mạo khuôn mặt, phương pháp học chuyển giao, phương pháp tinh chỉnh, mạng nơ-ron tích chập.*

1. ĐẶT VẤN ĐỀ

Các cuộc tấn công giả mạo đã trở thành mối đe dọa bảo mật nghiêm trọng cho các hệ thống xác thực, do chúng có thể được sử dụng để truy cập trái phép vào hệ thống bằng cách mạo danh người dùng được ủy quyền. Cụ thể, kẻ xấu có thể dễ dàng thực hiện các cuộc tấn công giả mạo đối với các hệ thống xác thực khuôn mặt bằng cách in ảnh của người được ủy quyền lên giấy hoặc bằng cách chụp ảnh và hiển thị trên thiết bị di động [2,3]. Nhằm đối phó với những thách thức này, một số kỹ thuật chống giả mạo đã được phát triển để phát hiện những hành vi giả mạo. Các hệ thống chống giả mạo dựa trên mạng nơ-ron tích chập gần đây đã thể hiện sự hiệu quả vượt trội của chúng so với các phương pháp truyền thống, vì thế chúng là giải pháp hứa hẹn để thay thế các kỹ thuật dựa trên đặc trưng và thuật toán học máy trước đây vốn dựa trên các đặc trưng cục bộ dễ nhạy cảm với nhiễu và kết quả kém chính xác.

[1] *Khoa Công nghệ Thông tin và Truyền thông, Trường Đại học Hồng Đức*
[2] *Sở Công Thương Thanh Hóa*

Tuy nhiên, có một xu hướng mới là nhận dạng khuôn mặt đang dần chuyển sang các thiết bị di động hoặc thiết bị nhúng. Điều này yêu cầu thuật toán chống giả mạo khuôn mặt cần được cải tiến để chạy với chi phí tính toán và lưu trữ ít hơn. Từ quan điểm này, việc thiết kế các thuật toán chống giả mạo dựa trên mạng nơ-ron tích chập trở nên thách thức hơn trong môi trường di động hoặc nhúng. Do đó, phát triển một thuật toán học sâu đủ tốt để có thể chạy được trên các thiết bị cấu hình thấp nhưng vẫn đáp ứng được độ chính xác của thuật toán vẫn đang cần nhiều đầu tư nghiên cứu.

Đóng góp chính của chúng tôi trong bài báo này là đề xuất một mạng nơ-ron học sâu hduNet phát triển từ mô hình MobileNetV2 được phát triển bởi Google. Bên cạnh đó, sau khi nghiên cứu những bộ dữ liệu về giả mạo khuôn mặt, chúng tôi nhận thấy điểm khó khăn và giới hạn về mức độ phong phú, đa dạng của các bộ dữ liệu hiện nay đều chưa đáp ứng. Bởi vì thế, chúng tôi đóng góp thêm vào 5000 dữ liệu ảnh trong bộ dữ liệu chuẩn LCC_FASD nhằm giảm tình trạng mất cân bằng và nâng cao hiệu quả của thuật toán để phù hợp với đặc trưng của người châu Á, cụ thể là người Việt Nam.

Bài báo được tổ chức như sau: Phần 2 trình bày các công việc liên quan đến những nghiên cứu về phát hiện giả mạo khuôn mặt. Phần 3 mô tả chi tiết về phương pháp đề xuất của chúng tôi. Phần 4 sẽ trình bày các quá trình thực nghiệm và kết quả của chúng tôi, bao gồm cả việc tiền xử lý dữ liệu, và hậu xử lý trong ngữ cảnh ứng dụng thực tiễn. Cuối cùng, các kết luận và những định hướng phát triển trong tương lai được trình bày ở phần 5.

2. CÁC KỸ THUẬT PHÁT HIỆN GIẢ MẠO KHUÔN MẶT

Nhìn chung, các nghiên cứu về phát hiện giả mạo có thể được chia thành 2 phương pháp chính gồm: phương pháp truyền thống và phương pháp sử dụng mạng nơ-ron tích chập CNNs.

Phương pháp truyền thống: Bài toán phát hiện giả mạo được quy về bài toán phân loại nhị phân bằng phương pháp sử dụng vector hỗ trợ (Support Vector Machine - SVM). Cụ thể, quá trình được thực hiện theo cách sau:

Bước 1. Trích chọn các đặc trưng bằng các bộ lọc khác nhau. Các đặc trưng được áp dụng chủ yếu bao gồm: Local Binary Patterns (LBP) [4,5,6], Scale Invariant Feature Transform (SIFT) [7], Speeded-Up Robust Features (SURF) [8], Histogram of Oriented Gradients (HOG) [9,10], Difference of Gaussian (DoG) [10].

Bước 2. Phân loại là giả hay thật bằng cách sử dụng thuật toán SVM hoặc Random Forest.

Tuy nhiên, các tác giả [11] chỉ ra rằng việc phát hiện đặc trưng bị ảnh hưởng rất nhiều bởi môi trường, ví dụ như điều kiện ánh sáng. Hơn nữa, phát hiện đặc trưng cho thấy các hạn chế của đặc trưng và các điểm đặc trưng không cung cấp nhiều thông tin như các phương thức CNN có thể mang lại với các tập dữ liệu khổng lồ.

Phương pháp CNNs: Về cơ bản, phương pháp sử dụng CNNs có thể được nhóm thành 3 nhóm.

Nhóm 1. Sử dụng duy nhất một khung hình màu RGB kết hợp với bộ phân loại. Hầu hết các phương pháp tiếp cận bằng cách sử dụng lớp cuối cùng trong mạng CNNs là tầng kết nối đầy đủ (Fully Connected Layer) để phân biệt khuôn mặt thật và giả. Bên cạnh đó, các tác giả [12] đã đề xuất một cách là không lấy đặc trưng ở tầng cuối cùng mà họ kết hợp sử dụng SVM và tầng gần cuối để phân biệt khuôn mặt thật và giả. Các tác giả σ13] tăng cường thêm bằng việc áp dụng mạng học sâu phát hiện chớp mắt để nâng cao kết quả. Và các nghiên cứu thấy rằng nhóm phương pháp sử dụng ảnh RGB kết hợp với mạng học sâu CNN vẫn có thể được cải thiện hiệu quả.

Nhóm 2. Sử dụng mạng CNN với nhiều khung ảnh RGB kết hợp với phương pháp đo áp lực tĩnh mạch Remote Photoplethysmography (rPPG) [14] để đưa ra quyết định. Phương pháp này cho kết quả tốt do những khuôn mặt giả sẽ không có các tín hiệu PPG này. Nhưng nhóm phương pháp này yêu cầu cần có máy ảnh chuyên dụng để có thể đo được PPG, đồng nghĩa với việc cần phát sinh thêm khoản chi phí vì cần mua thêm thiết bị ngoài.

Nhóm 3. Kết hợp nhiều loại ảnh RGB, ảnh hồng ngoại, ảnh 3D trên cùng một đối tượng để truyền vào mạng CNN nhằm trích chọn đặc trưng và đưa ra quyết định [15]. Nhóm phương pháp này mặc dù cho thấy độ chính xác cao nhất so với các nhóm khác nhưng yêu cầu nguồn dữ liệu và thiết bị phần cứng để đáp ứng. Bên cạnh đó, sử dụng nhiều loại ảnh cũng yêu cầu số lượng tính toán lớn, điều này làm cho thuật toán khó có thể đạt được tốc độ mong muốn.

Qua các phương pháp trên, việc áp dụng kiến trúc mạng trọng lượng nhẹ chưa có nhiều sự quan tâm. Trong phần tiếp theo, chúng tôi đề xuất mạng nơ-ron tích chập CNNs có tên hduNet dựa trên tinh chỉnh và tối ưu kiến trúc mạng CNNs nổi tiếng của Google là MobileNetV2 [16] để đáp ứng cả độ chính xác và thời gian xử lý.

3. PHƯƠNG PHÁP ĐỀ XUẤT

Trong phần này, chúng tôi sẽ giới thiệu chi tiết về mạng hduNet. Cách tiếp cận của chúng tôi là tinh chỉnh và tối ưu mạng nơ-ron nhân chập dựa trên một mô hình mạng đã được huấn luyện của Google là MobilenetV2. Phương pháp này thường được biết đến với tên gọi là học chuyển giao (transfer learning). Đây là phương pháp hiệu quả để cải thiện tốc độ và hiệu suất từ mô hình mạng được huấn luyện thực hiện một nhiệm vụ ban đầu chuyển sang thực hiện một nhiệm vụ thứ hai. Phương pháp này cũng giúp tránh được tình trạng học quá nhớ (overfitting) khi không có số lượng lớn dữ liệu huấn luyện từ đầu, và vì thế cũng đồng nghĩa với việc tiết kiệm được tài nguyên máy tính để phục vụ huấn luyện mô hình mạng nơ-ron.

3.1. Kiến trúc mạng đề xuất

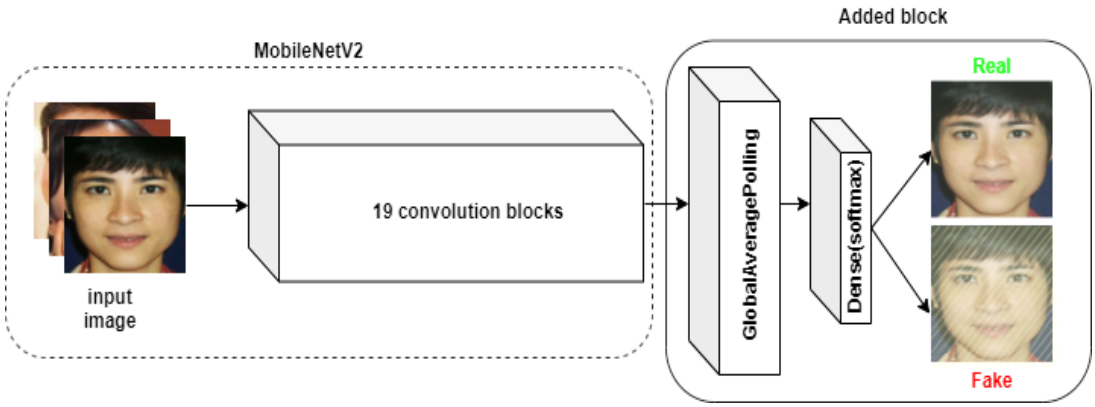
Đối với các mạng nơ-ron tích chập, thông thường có 2 hướng chính để thực hiện học chuyển giao: cách đơn giản là tách lấy bộ trích chọn đặc trưng (features extractor) hoặc kĩ thuật nâng cao mà cần đòi hỏi quá trình thực nghiệm đó là tinh chỉnh (fine-tunes) mô hình. Trong bài báo này, chúng tôi thực hiện theo hướng thứ 2 nhằm mục đích đạt được một mô hình mạng nơ-ron nhân chập tối ưu.

Bảng 1. So sánh mô hình mạng nổi tiếng đánh giá trên tập dữ liệu ImageNet

Mạng	Kích thước (MB)	Độ chính xác (%)	Số lượng tham số
Xception	88	79.0	22.910.480
VGG16	528	71.3	138.357.544
VGG19	549	71.3	143.667.240
ResNet50	99	74.9	25.636.712
InceptionV3	92	77.9	23.851.784
MobileNetV2	14	71.3	3.538.984

Mạng học sâu được chúng tôi đề xuất là hduNet được phát triển từ mạng MobileNetV2 [16], một trong những mạng học sâu tiên tiến được Google đề xuất năm 2018. Chúng tôi lựa chọn mạng MobilenetV2 nhằm kế thừa lại độ chính xác (đã được huấn luyện và kiểm thử trên bộ dữ liệu imagenet chứa 1,2 triệu ảnh [17]) và giải quyết được khó khăn (chi phí phần cứng, thời gian huấn luyện) mà hiện nay các thuật toán về mạng nơ-ron nhân tạo đang gặp phải. Ngoài ra, mạng MobileNetV2 có độ chính xác không thua kém các mô hình mạng phổ biến khác như VGG16, VGG19 trong khi lượng tham số chỉ gần 4 triệu, khoảng xấp xỉ 1/39 số lượng tham số của VGG16. Bảng 1 cho thấy thống kê so sánh độ chính xác, số lượng tham số mạng của một số kiến trúc mạng nổi tiếng khác.

Hình 1 mô tả kiến trúc tổng quan về hduNet. Khối nét đứt là phần kiến trúc của mạng MobileNetV2. Kiến trúc mạng MobileNetV2 nhận đầu vào là ảnh 2D với kích thước 224 x 224 pixel. Lớp nhân chập đầu tiên với 32 bộ lọc (filters), theo sau là 19 khối (gồm nhiều tầng nhân chập ghép cùng nhau). Hàm kích hoạt (activation function) được sử dụng là hàm rectifier linear unit (ReLU), tất cả kích thước của mặt nạ lọc là 3 x 3. Tầng kết nối đầy đủ (full connected layers) của MobilenetV2 được chúng tôi loại bỏ, thay vào đó chúng tôi bổ sung phần được bao bởi khối nét liền gồm. Tầng giảm chiều tham số, trong đó chúng tôi lựa chọn hàm giảm chiều là GlobalAveragePooling, theo sau là tầng kết nối và sử dụng hàm Softmax để phân loại ảnh thật hay ảnh giả mạo. Việc làm này nhằm điều chỉnh mục tiêu của kiến trúc mạng ban đầu để thực hiện mục tiêu của bài toán phát hiện giả mạo khuôn mặt. Trong phần kế tiếp, chúng tôi sẽ trình bày chi tiết việc huấn luyện mạng hduNet.



Hình 1. Kiến trúc tổng quan của mạng hduNet

3.2. Cơ sở dữ liệu

Ba cơ sở dữ liệu được sử dụng gồm LCC_FASD [1], NUAA [18] và hduDB là cơ sở dữ liệu đóng góp của chúng tôi.

(a) Minh họa hình ảnh trong tập dữ liệu LCC_FASD

(b) Minh họa hình ảnh trong tập dữ liệu NUAA

Hình 2. Một phần của 2 tập dữ liệu LCC_FASD và NUAA

Cơ sở dữ liệu Large Crowdcollected Facial Anti-Spoofing Dataset (LCC_FASD) chứa 3 tập con gồm trainning, development và evaluation. Tổng cộng gồm 243 đối tượng (người châu Úc) với 1942 ảnh thật và 16885 ảnh giả mạo. Hình ảnh được thu thập từ nhiều nguồn như Youtube, Amazon, Toloka,... với hình thức giả mạo trên 83 loại thiết bị (máy ảnh số, điện thoại,...) khác nhau. Hình 2 (a) mô tả một phần của tập dữ liệu LCC_FASD.

Cơ sở dữ liệu NUAA Photo Imposter Database (NUAA) chứa 15 đối tượng (người Trung Quốc), gồm 5105 ảnh thật và 7509 ảnh giả mạo tại nhiều vị trí (văn phòng, ngoài trời, …) và điều kiện ánh sáng khác nhau. Cơ sở dữ liệu chia làm 2 tập training và testing. Hình 2 (b) minh họa một phần ảnh thật và ảnh được giả mạo của bộ dữ liệu NUAA.

Sau khi phân tích 2 bộ cơ sở dữ liệu chuẩn, được sử dụng trong nhiều nghiên cứu về phát hiện giả mạo khuôn mặt [8,9,10,18], chúng tôi nhận thấy rằng. Cơ sở dữ liệu LCC_FASD tập trung phần lớn vào thu thập ảnh giả mạo với tỉ lệ ảnh thật/ảnh giả mạo chệnh lệch nhau rất lớn xấp xỉ 1/9 điều này có xu hướng dẫn đến kết quả quá trình huấn luyện mạng nơ-ron bị kém chính xác (underfitting). Trong khi đó, cơ sở dữ liệu NUAA với số lượng đối tượng không nhiều, chỉ 15 người khác nhau. Ngoài ra dạng tấn công của cơ sở dữ liệu NUAA chỉ là phương pháp chụp ảnh và in lại trên giấy A4 (2D print-attack) mà không có đa dạng hóa hình thức tấn công bằng cách quay chụp lại khuôn mặt từ thiết bị số như điện thoại, máy tính bảng,… (video replay attacks). Chính vì thế, với kỳ vọng có một mô hình mạng tốt, có tính tổng quát cao, phù hợp với đặc trưng của người châu Á trong cả quá trình huấn luyện và quá trình kiểm thử thực tế. Chúng tôi xây dựng cơ sở dữ liệu mới, hduNet, tổng hợp dựa trên 2 cơ sở dữ liệu trên và đóng góp thêm nhằm mục tiêu phù hợp với dữ liệu người châu Á, chi tiết được mô tả trong bảng 2.

Bảng 2. Thống kê số lượng ảnh trong tập dữ liệu hduDB

Phần	Ảnh giả mạo	Ảnh thật
Training	8000	4800
Valuation	2000	1200
Evaluation	4436	330
Tổng	14436	6330

4. THỰC NGHIỆM VÀ KẾT QUẢ

4.1. Môi trường thực nghiệm

Thuật toán được cài đặt bằng ngôn ngữ Python trên thư viện hỗ trợ phát triển thuật toán học sâu Keras[3]. Ngoài ra quá trình huấn luyện thực hiện trên máy tính với hệ điều hành Ubuntu 18.4-LTS được trang bị Intel(R) Xeon(R) W-2133 CPU @ 3.60GHz (16GB RAM), NVIDIA GeForce GTX GPU (11GB).

4.2. Tiền xử lý dữ liệu

Do sử dụng 3 nguồn dữ liệu khác nhau, để thực hiện huấn luyện mô hình mạng nơ-ron nhân chập chúng tôi cài đặt một số bước tiền xử lý dữ liệu gồm:

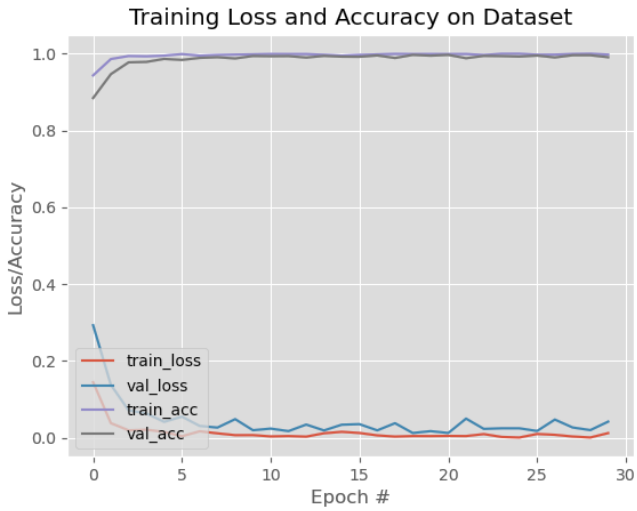
Cơ sở dữ liệu LCC_FASD và NUAA: Thực hiện co, giãn ảnh về kích thước chung là 128 x 128. Điều này đồng nghĩa với việc kích thước ảnh đầu vào trong mạng hduNet là 128 x128, không phải là 224 x 224 của mạng MobileNetV2. Điều này, cũng giúp giảm được chi phí tính toán trong mạng nơ-ron.

[3] *Thư viện lập trình mạng học sâu (deep learning): https://keras.io/*

Cơ sở dữ liệu hduNet: Vì dữ liệu do chúng tôi thực hiện đang ở mức thô gồm các video quay chụp từ camera an ninh, thiết bị di động,… nên chúng tôi thực hiện quá trình xử lý theo các bước như sau: Đầu tiên, video đầu vào được phân đoạn thành các khung hình tuần tự. Bước tiếp theo, chúng tôi thực hiện thuật toán phát hiện khuôn mặt trong khung hình. Thứ ba, với những khung hình có chứa khuôn mặt chúng tôi xác định kích thước của chúng và cắt vùng ảnh chứa khuôn mặt để lưu trữ. Cuối cùng, với mỗi ảnh mới chỉ chứa khuôn mặt được điều chỉnh về kích thước chuẩn 128 x 128.

4.3. Huấn luyện và đánh giá mô hình

Chúng tôi huấn luyện mạng hduNet với những tham số mạng gồm: learning rate là 0.0001, kích thước batch size là 32, thuật toán tối ưu Adam [19] và số lượng Epochs là 28. Hình 3 mô tả quá trình huấn luyện mạng hduNet.



Hình 3. Biểu đồ huấn luyện mạng hduNet với số lượng Epochs là 28

Qua hình 3 cho chúng ta thấy mô hình mạng nơ-ron hduNet học khá tốt, điều này được thể hiện qua 2 quá trình huấn luyện (training) và kiểm thử (validation). Biểu đồ hàm mục tiêu (loss) và độ chính xác (accuracy) của huấn luyện và kiểm thử đều bám sát nhau và đạt kết quả ấn tượng khi kết thúc huấn luyện với giá trị hàm mục tiêu giảm còn 0,02 và độ chính xác đạt 0.98 còn quá trình kiểm thử giá trị hàm mục tiêu và độ chính xác tương ứng đạt 0,08 và 0,97.

Để đánh giá chất lượng mạng nơ-ron và độ tổng quát dữ liệu của mô hình hduNet chúng tôi thực hiện chạy mô hình mạng trên một tập dữ liệu đánh giá (evaluation data) gồm 4436 ảnh giả mạo và 330 ảnh thật, không nằm trong dữ liệu dùng huấn luyện và kiểm thử. Độ đo mà chúng tôi sử dụng là 3 độ đo phổ biến được sử dụng đánh giá các hệ thống sinh trắc học gồm: Độ chính xác (Accuracy) công thức (3), Tỉ lệ phân loại ảnh giả mạo nhầm thành ảnh thật (False Acceptance Rate - FAR) công thức (1), Tỉ lệ phân loại ảnh thật bị nhầm thành giả mạo (False Rejection Rate - FRR) công thức (2). Bảng 3 diễn giải cách tính FAR, FRR và Accuracy.

Bảng 3. Ma trận nhầm lẫn (Confusion matrix)

Lớp		Phân lớp bởi hệ thống	
		Thật	Giả mạo
Phân lớp đúng (nhãn)	Thật	TP	FN
	Giả mạo	FP	TN

Trong đó:

TP: Số lượng các mẫu thuộc lớp thật được phân loại đúng là ảnh thật.

FP: Số lượng các mẫu thuộc lớp giả mạo được phân loại nhầm thành ảnh thật.

FN: Số lượng các mẫu thuộc lớp thật được phân loại thành giả mạo.

TN: Số lượng các mẫu thuộc lớp giả mạo được phân loại đúng là giả mạo.

N: Tổng số lượng các mẫu được đánh giá (N = 4436 + 330 = 4766).

$$FAR = \frac{FP}{N} (1) \qquad FRR = \frac{FN}{N} (2)$$

$$Accuracy = \frac{TP + TN}{N} (3)$$

Kết quả tương ứng chúng tôi thu được là FAR = 0.124, FRR =0.008 và Accuracy = 0.867. Kết quả của chúng tôi được đem so sánh với kết quả độ chính xác của nhóm tác giả [1] cùng sử dụng các mạng nơ-ron nhân chập CNNs khác.

Bảng 4 cho chúng ta thấy mặc dù kết quả độ chính xác của chúng tôi thấp hơn gần 10% so với kiến trúc mạng tốt nhất SeNet-154, nhưng theo bảng 1 chúng tôi đã chỉ ra rằng mục tiêu của kiến trúc mạng này cần đạt được là kích thước mạng cần đủ nhẹ với số lượng tham số ít. Chúng tôi đánh đổi một tỉ lệ để đạt được kiến trúc mạng nhẹ và độ chính xác vẫn ở mức cao là 86,7%. Ngoài ra, trong các kịch bản ứng dụng thực tế hệ thống nhận dạng khuôn mặt phải đối mặt với quá trình đối sánh và nhận dạng khuôn mặt, công đoạn này có thời gian tính toán tỉ lệ thuận với số lượng mẫu (người) trong cơ sở dữ liệu đối sánh. Hoặc là, trong các ứng dụng mà việc đọc dữ liệu từ luồng camera với tốc độ thông thường 24 hình/giây, tốc độ xử lý của các ứng dụng là điều cần được ưu tiên. Để đối phó với ràng buộc xử lý luồng video từ camera chúng tôi sẽ trình bày cụ thể trong phần tiếp theo.

Bảng 4. So sánh độ chính xác của hduNet và các CNNs

Mạng CNNs	Độ chính xác (%)
Xception	95.9
ResNext-50	94.0
SeNet-154	96.2
hduNet	86.7

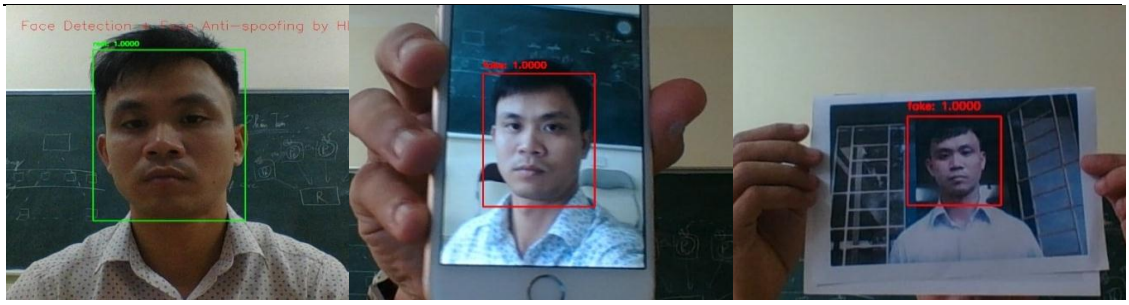
4.4. Ứng dụng thực tiễn

Trong các ứng dụng thực tiễn, việc phát hiện giả mạo khuôn mặt thường đi kèm cùng nhiều thuật toán khác như phát hiện khuôn mặt, đánh dấu điểm đặc trưng (mắt, mũi, miệng) trên khuôn mặt, trích chọn đặc trưng (embedding features) và phân loại. Ngoài ra, dữ liệu đầu vào thường sẽ là luồng video được đọc từ camera, webcam thay vì ảnh tĩnh như quá trình huấn luyện và kiểm thử. Chính vì thế, thuật toán trở nên thách thức hơn với việc xử lý luồng dữ liệu video (trung bình khoảng 24 khung hình/giây). Để giải quyết khó khăn này, chúng tôi cài đặt thêm một thuật toán hậu xử lý (post-processing) được trình bày dưới đây.

Thuật toán: phát hiện giả mạo khuôn mặt với dữ liệu video

Đầu vào: Ảnh thu nhận từ webcam, camera

01 : While True :
02 : Begin :
03 : count_real = 0
04 : Dò tìm và phát hiện khuôn mặt trong ảnh
05 : Cắt vùng khuôn mặt phát hiện
06 : Phát hiện khuôn mặt giả mạo
07 : count_real += 1
08 : If count_real > 5 :
09 : Begin :
10 : Hiển thị và vẽ hình bao quanh khuôn mặt với nhãn real
11 : count_real = 0
12 : End if :
13 : else
14 : Hiển thị và vẽ hình bao quanh khuôn mặt với nhãn fake
15 : End



(a) (b) (c)

Hình 4. Một số kết quả thực nghiệm phát hiện giả mạo khuôn mặt từ camera giám sát

(a) khuôn mặt thật trước camera, (b) khuôn mặt giả mạo được chụp bằng điện thoại, (c) khuôn mặt giả mạo được in từ ảnh

Để đánh giá về thời gian xử lý, chúng tôi thực nghiệm mô hình hduNet trên máy tính thông thường được trang bị CPU Intel® Core™ i5-5300U, RAM 4GB. Qua bảng 5 cho chúng ta thấy rằng, mặc dù hduNet có thời gian xử lý tối đa xấp xỉ đạt 15.4 FPS (Frames Per Second) nhưng với việc kết hợp vào bộ dò tìm khuôn mặt và hậu xử lý thì về tổng thể thời gian xử lý vẫn cho kết quả gần thời gian thực (real-time) với trung bình 22.5 FPS. Như vậy, kết hợp mô hình hduNet và áp dụng thêm kĩ thuật hậu xử lý, thuật toán của chúng tôi trở nên bền vững và có tiềm năng ứng dụng trong các giải pháp thực tế hơn. Hình 4 mô tả một số kết quả thực nghiệm khi chạy mô hình trong điều kiện thực tế.

Bảng 5. Thời gian xử lý

Mô hình	Max FPS
hduNet	15.4
Tích hợp hệ thống	22.5

5. KẾT LUẬN

Trong bài báo này, chúng tôi trình bày một hướng tiếp cận mới mà sử dụng phương pháp học chuyển giao trong các mạng nơ-ron tích chập để giải quyết vấn đề phát hiện giả mạo khuôn mặt. Ngoài ra, chúng tôi cũng đóng góp thêm để xây dựng một cơ sở dữ liệu dành cho việc phát hiện giả mạo khuôn mặt với đặc trưng khuôn mặt của người Châu Á. Cơ sở dữ liệu mới khắc phục được những hạn chế của 2 cơ sở dữ liệu LCC_ FASD và NUAA, trở thành một cơ sở dữ liệu mang tính đại diện tốt, mức độ tổng quát và đa dạng cao. Nhìn chung, giải pháp của chúng tôi đơn giản nhưng hiệu quả và dễ sử dụng trong các tình huống ứng dụng thực tế. Trong tương lai, chúng tôi hướng tới tích hợp mô hình phát hiện giả mạo khuôn mặt vào trong những ứng dụng như điểm danh, chấm công, khóa cửa nhận dạng khuôn mặt.

TÀI LIỆU THAM KHẢO

[1] D. Timoshenko, K. Simonchik, V. Shutov, P. Zhelezneva and V. Grishkin (2019), Large Crowdcollected Facial Anti-Spoofing Dataset, in *Computer Science and Information Technologies (CSIT)*, Yerevan, Armenia.

[2] N. Evans (2019), *Handbook of Biometric Anti-spoofing: Presentation Attack Detection*, Springer.

[3] Y. Li, K. Xu, Q. Yan, Y. Li, and R. H. Deng (2014), Understanding osn-based facial disclosure against face authentication systems, in *Proceedings of the 9th ACM symposium on Information, computer and communications security. ACM*.

[4] Tiago de Freitas Pereira, Andre Anjos, José Mario De Martino, and Sebastien Marcel (2013), Can face anti-spoofing countermeasures work in a real world scenario?, in *International Conference on Biometrics (ICB)*.

[5] Tiago de Freitas Pereira, Andre Anjos, José Mario De Martino, and Sebastien Marcel (2012), Lbp- top based countermeasuré against face spoofing attacks, in *Asian Conference on Computer Vision*.

[6] Jukka Maatta, Abdenour Hadid, and Matti Pietikainen (2011), Face spoofing detection from single images using micro-texture analysis, in *International joint conference on Biometrics (IJCB)*.

[7] Keyurkumar Patel, Hu Han, and Anil K Jain (2016), Secure face unlock: Spoof detection on smartphones, in *IEEE Transactions on Information Forensics and Security*.

[8] Z. Boulkenafet, J. Komulainen and A. Hadid (2017), Face Antispoofing Using Speeded-Up Robust Features and Fisher Vector Encoding, *IEEE Signal Processing Letters*, vol. 24, pp. 141-145,.

[9] J. Komulainen, A. Hadid and M. Pietikäinen (2013), Context based face anti-spoofing, in *IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, Arlington, VA.

[10] J. Yang, Z. Lei, S. Liao and S. Z. Li (2013), Face liveness detection with component dependent descriptor, in *International Conference on Biometrics (ICB)*, Madrid.

[11] Zezheng Wang, Chenxu Zhao, Yunxiao Qin, Qiusheng Zhou, and Zhen Lei (2018), Exploiting temporal and depth information for multi-frame face anti-spoofing, *CoRR*, vol. abs/1811.05118.

[12] Lei Li, Xiaoyi Feng, Zinelabidine Boulkenafet, Zhaoqiang Xia, Mingming Li, and Abdenour Hadid (2016), An original face anti-spoofing approach using partial convolutional neural network, in *Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*.

[13] Keyurkumar Patel, Hu Han, and Anil K Jain (2016), Cross-database face antispoofing with robust feature representation, in *Chinese Conference on Biometric Recognition*.

[14] Javier Hernandez-Ortega, Julian Fierrez, Aythami Morales, and Pedro Tome (2018), Time analysis of pulse-based face antispoofing in visible and nir, in *Conference on Computer Vision and Pattern Recognition Workshops*.

[15] Shifeng Zhang, Xiaobo Wang, Ajian Liu, Chenxu Zhao, Jun Wan, Sergio Escalera, Hailin Shi, Zezheng Wang, and Stan Z Li (2018), Casia-surf: A dataset and benchmark for large-scale multi-modal face anti-spoofing, *CoRR*, vol. abs/1812.00408.

[16] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. Chen (2018), MobileNetV2: Inverted Residuals and Linear Bottlenecks, in *Conference on Computer Vision and Pattern Recognition, Salt Lake City*, UT.

[17] J. Deng, W. Dong, R. Socher, L. Li, Kai Li and Li Fei-Fei (2009), ImageNet: A large-scale hierarchical image database, in *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL.

[18] Xiaoyang Tan, Yi Li, Jun Liu, and Lin Jiang (2010), Face liveness detection from a single image with sparse low rank bilinear discriminative model, in *the 11th European conference on Computer vision*, Berlin, Heidelberg.

[19] Kingma Diederik P, Ba Jimmy (2015), Adam: A Method for Stochastic Optimization, in *the 3rd International Conference for Learning Representations*, San Diego.

[20] Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H (2017), MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, in *ArXiv*.

A DEEP LEARNING TECHNIQUE FOR FRAUD FACE DETECTION

Le Van Hao, Trinh Thi Anh Loan, Le Viet Nam, Nguyen Duc Toan

ABSTRACT

Fraud face detection is a crucial procedure for many face recognition systems. In recent years, state-of-the-art approaches based on convolution neural networks (CNNs) show impressive results compared to traditional methods using hand-crafted features. In addition, the increasing trend of embedding the computer vision systems on mobile devices requires that the designed algorithms are capable of dealing with the time-critical constraint. In this paper, we first propose a CNN model, namely hduNet, developed from Google's MobilenetV2 that provides a flexible trade-off between latency and accuracy, to detect different face spoofing attacks. We then provide an addition dataset of roughly 5000 images capturing the characteristics of Vietnamse people. Combining with LCC_FASD [1] dataset (which is only 1942 real face images, while having 16855 fake face images), the proposed model is carefully fine-tuned to optimize the computational cost as well as the classification accuracy. To validate the model, different experiments have been conducted, demonstrating interesting performance in comparison with other methods.

Keywords: *Face anti-spoofing, transfer learning, fine-tunning, convolution neural network.*

** Ngày nộp bài: 27/7/2020; Ngày gửi phản biện: 3/8/2020; Ngày duyệt đăng: 28/10/2020*

** Bài báo này là kết quả nghiên cứu từ đề tài cấp cơ sở mã số ĐT-2019-26 của Trường Đại học Hồng Đức.*