HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



Võ Minh Mạnh

TÓM TẮT ĐỀ ÁN TỐT NGHIỆP THẠC SĨ KỸ THUẬT

(Theo định hướng ứng dụng)

HÀ NỘI - 2025

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



Võ Minh Mạnh

ỨNG DỤNG MẠNG NƠ-RON NHÂN TẠO VÀ PHÉP BIẾN ĐỔI FOURIER ĐỂ PHÁT HIỆN HÌNH ẢNH THẺ CĂN CƯỚC GIẢ MẠO

CHUYÊN NGÀNH : KHOA HỌC MÁY TÍNH

MÃ SỐ: 8.48.01.01

TÓM TẮT ĐỀ ÁN TỐT NGHIỆP THẠC SĨ KỸ THUẬT

HÀ NỘI - 2025

Đề án tốt nghiệp được hoàn thành tại:

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Người hướng dẫn khoa học: TS. NGUYỄN TẤT THẮNG

Người phản biện 1: ……

Người phản biện 2: ……

Đề án tốt nghiệp sẽ được bảo vệ trước Hội đồng chấm đề án tốt nghiệp thạc sĩ tại Học viện Công nghệ Bưu chính Viễn thông

Vào lúc: ..... giờ ..... ngày ..... tháng .... năm .....

Có thể tìm hiểu đề án tốt nghiệp tại:

- Thư viện của Học viện Công nghệ Bưu chính Viễn thông

MỞ ĐẦU

1. Lý do chọn đề tài

Hiện nay việc sử dụng giấy tờ tùy thân (CCCD, hộ chiếu) để xác thực danh tính trên môi trường điện tử trở nên phổ biến, đặc biệt trong ngân hàng số. Hình thức này cho phép người dùng xác thực từ xa qua thiết bị kết nối Internet, nhanh chóng, tiện lợi, tiết kiệm chi phí, nâng cao trải nghiệm người dùng và hỗ trợ mở rộng dịch vụ tới vùng sâu, vùng xa, đồng thời tăng tính minh bạch và hiệu quả quản lý rủi ro.

Tuy nhiên, định danh điện tử cũng tiềm ẩn nhiều rủi ro gian lận. Do đó, nhu cầu tăng cường khả năng phát hiện các hình thức giả mạo trong hệ thống định danh điện tử là rất cấp thiết, với các giải pháp kết hợp thị giác máy tính, xử lý ảnh số và trí tuệ nhân tạo.

Trong bối cảnh đó, các mô hình mạng nơ-ron nhân tạo ngày càng được ứng dụng rộng rãi, trong khi phép biến đổi Fourier đã chứng minh hiệu quả trong phân tích cấu trúc tần số và phát hiện bất thường trong hình ảnh [1]. Việc kết hợp mạng nơ-ron nhân tạo với phép biến đổi Fourier để phát hiện hình ảnh thẻ căn cước giả mạo có tiềm năng giúp giảm rủi ro tài chính cho người dân. Vì vậy, đề tài áp dụng các mô hình học máy tiên tiến kết hợp với phép biến đổi Fourier trong xử lý ảnh, phát hiện dấu hiệu giả mạo trong ảnh thẻ căn cước, làm cơ sở cải tiến hệ thống định danh điện tử.

2. Tổng quan về vấn đề nghiên cứu

Đề tài tập trung vào bài toán phát hiện hình ảnh giấy tờ tùy thân không phải ảnh thật, cụ thể là ảnh thẻ căn cước được chụp gián tiếp qua màn hình thiết bị điện tử. Nghiên cứu gồm hai phần: (i) phân tích ưu, nhược điểm của VGG16, phép biến đổi Fourier và các bộ lọc tần số trong phát hiện giả mạo; (ii) xây dựng và đánh giá mô hình mạng nơ-ron kết hợp VGG16 và phép biến đổi Fourier để phát hiện dấu hiệu giả mạo trên ảnh thẻ căn cước. Trên cơ sở đó, đề án tốt nghiệp được trình bày theo ba chương như sau:

Chương 1: Tổng quan về vấn đề hình ảnh thẻ căn cước giả mạo.

Chương 2: Phương pháp phát hiện hình ảnh thẻ căn cước giả mạo

Chương 3: Xây dựng và đánh giá mô hình

3. Mục đích nghiên cứu

Mục đích của nghiên cứu là thử nghiệm và đánh giá mức độ hiệu quả của việc sử dụng mạng nơ-ron nhân tạo kết hợp với phép biến đổi Fourier trong phát hiện hình ảnh thẻ căn cước giả mạo. Để hoàn thành mục tiêu này, nghiên cứu sẽ tiến hành các nhiệm vụ sau:

- Nghiên cứu đặc trưng của hình ảnh thẻ căn cước và các dấu hiệu nhận biết giữa thẻ thật và thẻ giả.

- Xây dựng bộ dữ liệu hình ảnh thẻ căn cước thật và giả phục vụ cho quá trình huấn luyện và kiểm thử mô hình.

- Nghiên cứu chi tiết mô hình mạng nơ-ron nhân tạo VGG16 và cách thức nó có thể được áp dụng trong nhận dạng hình ảnh.

- Nghiên cứu phép biến đổi Fourier và ứng dụng của nó trong xử lý ảnh, đặc biệt là trong việc phát hiện những bất thường về tần số của hình ảnh.

- Tìm hiểu cách kết hợp giữa mạng nơ-ron nhân tạo VGG16 và phép biến đổi Fourier để tối ưu hóa quá trình phát hiện hình ảnh giả mạo.

- Xây dựng mô hình, tiến hành huấn luyện mạng nơ-ron, và đánh giá kết quả qua các tiêu chí như độ chính xác và hiệu suất của mô hình

Nghiên cứu này hướng đến việc cung cấp một giải pháp mới trong việc chống giả mạo giấy tờ tùy thân khi thực hiện định danh điện tử, góp phần nâng cao tính bảo mật cho các hệ thống này.

4. Đối tượng và phạm vi nghiên cứu

4.1. Đối tượng nghiên cứu

Đối tượng nghiên cứu của đề tài là các hình ảnh giấy tờ tùy thân, bao gồm: giấy chứng minh nhân dân, thẻ căn cước công dân, thẻ căn cước, hộ chiếu bao gồm cả hình ảnh thật và hình ảnh giả mạo.

4.2. Phạm vi nghiên cứu

a. Phạm vi không gian: Bộ dữ liệu mã nguồn mở hình ảnh thẻ căn cước công khai trên Internet khoảng 72.793 ảnh.

b. Phạm vi thời gian: Từ tháng 8 năm 2024 đến tháng 11 năm 2025.

5. Phương pháp nghiên cứu

Sử dụng phương pháp nghiên cứu định lượng, bao gồm thu thập các thông tin và dữ liệu dưới dạng số học, số liệu có tính chất thống kê để có được những thông tin cơ bản, tổng quát về đối tượng nghiên cứu nhằm phục vụ mục đích thống kê, phân tích.

CHƯƠNG 1 - TỔNG QUAN VỀ VẤN ĐỀ HÌNH ẢNH THẺ CĂN CƯỚC GIẢ MẠO

1.1. Giới thiệu vấn đề phát hiện hình ảnh thẻ căn cước giả mạo.

Một trong các hình thức lừa đảo trên môi trường điện tử là kẻ gian tạo giấy tờ tùy thân giả để mở tài khoản ngân hàng hoặc giả mạo danh tính. Với sự hỗ trợ của công nghệ, các giấy tờ giả ngày càng tinh vi, đặc biệt là kỹ thuật sử dụng hình ảnh thẻ căn cước hiển thị trên màn hình thiết bị (điện thoại, máy tính bảng, máy tính) để đánh lừa hệ thống rằng đó là thẻ thật.

1.2. Các nghiên cứu liên quan

Bài toán phát hiện giả mạo nói chung và phát hiện thẻ căn cước giả mạo nói riêng đã được nghiên cứu với nhiều hướng tiếp cận, từ trích xuất đặc trưng đơn giản đến các mô hình học sâu. Polevoy và cộng sự [2] sử dụng ResNet-50 [3] phân loại ảnh thẻ căn cước thật/giả trên các bộ dữ liệu công khai như DLC-2021 [2], MIDV-2020 [4]. Kunina và cộng sự [5] tập trung vào vùng biên thẻ căn cước, làm nổi bật các dải hẹp dọc mép ảnh và dùng biến đổi Hough nhanh để khai thác vân nhiễu moiré. Markham và cộng sự [6] tạo ảnh thẻ căn cước giả mạo bằng các mô hình sinh, sau đó huấn luyện MobileNetV2 [7] để phân loại thật – giả trên dữ liệu tổng hợp từ MIDV-2020 và DLC-2021. Al-Ghadi và cộng sự [8] đề xuất khung IDTrust với hai biến thể: DeepQD dùng bộ mã hoá CNN, GuidedDeepQD bổ sung bước lọc thông dải dựa trên FFT ở tiền xử lý. Zhang và cộng sự [9] tổng hợp, phân loại 43 nghiên cứu về phát hiện ảnh thẻ căn cước giả mạo, chỉ ra ưu, nhược điểm và các thách thức còn tồn tại.

Ở bài toán phát hiện khuôn mặt giả mạo, Zeng và cộng sự [10] sử dụng mô hình hai nhánh EfficientFormerV2 [11], trong đó một nhánh dùng ảnh gốc, nhánh còn lại dùng ảnh đã lọc Gaussian và biến đổi Fourier, kết hợp bằng chú ý chéo đa tầng. Yu và cộng sự [12] thay thế các lớp tích chập thông thường bằng tích chập chênh lệch tập trung (Central Difference Convolution), kết hợp cường độ và gradient

tại mỗi điểm ảnh, đạt kết quả tốt trên các bộ dữ liệu OULU-NPU [13] và CASIA-MFSD [14].

Nhìn chung, các phương pháp này ưu tiên các mô hình học sâu tiên tiến, quy mô lớn, đòi hỏi tài nguyên tính toán đáng kể, trong khi các kiến trúc mạng cơ bản, ổn định, dễ triển khai và tinh chỉnh cho bối cảnh thời gian thực, tài nguyên hạn chế vẫn chưa được khai thác đầy đủ.

1.3. Giải pháp giải quyết vấn đề của đề án

Đề án lựa chọn hướng tiếp cận dựa trên các mô hình cơ bản như VGG16. VGG16 được dùng để trích xuất và phân biệt đặc trưng hình ảnh, trong khi phép biến đổi Fourier và các bộ lọc tần số làm nổi bật các đặc trưng tần số tinh vi, khó nhận biết bằng không gian ảnh thông thường. Từ đó, đề án đề xuất kết hợp VGG16 với phép biến đổi Fourier nhằm tăng cường khả năng phát hiện dấu hiệu giả mạo trong ảnh thẻ căn cước và giấy tờ tùy thân.

Đề án triển khai qua ba giai đoạn: (i) xây dựng tập dữ liệu thẻ căn cước cho huấn luyện, kiểm tra và kiểm thử; (ii) xây dựng, huấn luyện, tinh chỉnh mô hình bằng Python, PyTorch, OpenCV và các công cụ liên quan; (iii) đánh giá mô hình về độ chính xác và hiệu năng, đồng thời so sánh với một số mạng khác.

1.4. Kết luận

Chương này khái quát bài toán giả mạo thẻ căn cước trong hệ thống định danh điện tử, làm rõ hình thức gian lận bằng ảnh chụp gián tiếp qua màn hình thiết bị khác, điểm qua các nghiên cứu liên quan và định hướng giải quyết của đề án. Các nghiên cứu trước chủ yếu tập trung vào mô hình mới, chưa khai thác đầy đủ việc tinh chỉnh các mô hình cơ bản và tận dụng ưu thế của các phương pháp xử lý ảnh như phép biến đổi Fourier rời rạc. Đề án lựa chọn tập trung vào mạng nơ-ron VGG16 và phép biến đổi Fourier, đồng thời nghiên cứu cách tích hợp hai hướng này để tối ưu phát hiện thẻ căn cước giả mạo, làm nền tảng lý thuyết cho các chương tiếp theo.

CHƯƠNG 2 - PHƯƠNG PHÁP PHÁT HIỆN HÌNH ẢNH THẺ CĂN CƯỚC GIẢ MẠO

2.1. Giới thiệu về phép biến đổi Fourier

Phép biến đổi Fourier là một phương pháp toán học quan trọng, được ứng dụng rộng rãi trong xử lý ảnh, nén và khôi phục hình ảnh, xử lý âm thanh, truyền thông số và nhiều lĩnh vực kỹ thuật – khoa học khác. Theo Nixon và Aguado [15], phép biến đổi này chuyển tín hiệu từ miền thời gian/không gian sang miền tần số, phân rã tín hiệu thành tổng các sóng sin và cos có tần số khác nhau.

2.1.1. Phép biến đổi Fourier rời rạc hai chiều.

Phép biến đổi Fourier rời rạc hai chiều là phép tính đưa tín hiệu như tín hiệu âm thanh hoặc tín hiệu hình ảnh từ miền thời gian hoặc không gian sang miền tần số, mục đích là phân giải tín hiệu đó thành tổng các thành phần sóng sin hoặc cos với các tần số khác nhau [15].

Thuật toán biến đổi Fourier rời rạc nhanh (FFT) do Cooley và Tukey (1965) đề xuất áp dụng nguyên tắc chia để trị, tách phép biến đổi Fourier rời rạc độ dài N (với $N = 2^i$) thành nhiều phép biến đổi nhỏ hơn cho đến $N = 2$, qua đó giảm số phép toán và hạ độ phức tạp từ $O(N^2)$ xuống $O(NlogN)$.

2.1.2. Nguyên lý và cách thực hiện lọc tần số bằng phép biến đổi Fourier trên ảnh.

2.1.2.1. Các đặc trưng tần số trong ảnh

Trong phân tích miền tần số, tần số thấp biểu diễn các biến đổi cường độ chậm, vùng tương đối đồng nhất và cấu trúc chung của ảnh (bóng, sắc độ tổng thể, hình dạng cơ bản) và thường tập trung quanh tâm phổ. Tần số cao biểu diễn các biến đổi nhanh, cạnh sắc nét, chi tiết kết cấu và nhiễu, đồng thời có thể chứa dấu vết bất thường do chỉnh sửa, ghép ảnh. Ảnh mờ thường giàu thành phần tần số thấp, trong khi ảnh

rõ có nhiều thành phần tần số cao, nên có thể dựa vào miền tần số để phân biệt ảnh rõ và ảnh mờ.

2.1.2.2. Khái niệm lọc tần số

Trong miền tần số, bộ lọc là mặt nạ áp lên phổ để giữ lại hoặc loại bỏ một dải tần số. Bộ lọc tần số cao loại bỏ thành phần tần số thấp và giữ lại chi tiết, biên, kết cấu; ngược lại, bộ lọc tần số thấp loại bỏ chi tiết tần số cao và giữ lại các thành phần tần số thấp.

2.1.2.3. Cách thực hiện bộ lọc tần số

Cho bức ảnh đầu vào là ảnh xám có kích thước $M \times N$.

Bước 1: Chuyển ảnh từ miền không gian sang miền tần số

Áp dụng biến đổi Fourier rời rạc hai chiều (2.2) để thu được phổ tần số, sau đó dùng phép dịch phổ (2.3) để đưa thành phần tần số thấp (tần số 0) về tâm, giúp dễ quan sát và thao tác với mặt nạ tròn đồng tâm (tần số thấp ở gần tâm, tần số cao ở biên).

$$Fp_{shift}(u,v) = Fp\left(\left(u + \frac{M}{2}\right) \bmod M, \left(v + \frac{N}{2}\right) \bmod N\right) \tag{2.3}$$

Bước 2: Thiết kế bộ lọc tần số

Xây dựng mặt nạ bộ lọc lý tưởng thông thấp (2.4) hoặc thông cao (2.5) dựa trên khoảng cách $D(u,v)$ đến tâm (2.6) và bán kính D_0.

$$H(u,v) = \begin{cases} 1, D(u,v) \le D_0, \\ 0, D(u,v) > D_0, \end{cases} \tag{2.4}$$

$$H(u,v) = \begin{cases} 0, D(u,v) \le D_0, \\ 1, D(u,v) > D_0, \end{cases} \tag{2.5}$$

$$D(u,v) = \sqrt{\left(u - \frac{M}{2}\right)^2 + \left(v - \frac{N}{2}\right)^2} \quad (2.6)$$

Chọn D_0 quá nhỏ hoặc quá lớn sẽ làm ảnh quá mờ hoặc lọc nhiễu kém (đối với thông thấp); giữ gần hết hoặc loại quá nhiều chi tiết (đối với thông cao). Bộ lọc lý tưởng có ưu điểm cắt "sạch" các thành phần ngoài ngưỡng, nhưng nhược điểm là gây ranh giới gắt giữa vùng giữ và vùng loại bỏ, tạo hiện tượng rung (ringing / hiệu ứng Gibbs).

Bước 3. Áp dụng bộ lọc vào ảnh tần số và biến đổi Fourier ngược

Sau khi thiết kế bộ lọc $H(u,v)$, (thông thấp hoặc thông cao), tiến hành nhân trực tiếp phổ tần số đã dịch $Fp_{shift}(u,v)$ với $H(u,v)$ để thu được phổ đã lọc $G(u,v)$, chỉ còn các thành phần tần số mong muốn. Tiếp đó, dịch phổ $G(u,v)$ ngược lại (unshift) để đưa tần số 0 về góc ma trận, rồi áp dụng biến đổi Fourier rời rạc ngược để thu được ảnh đã lọc $g(x,y)$ trong miền không gian.

2.2. Giới thiệu về mô hình mạng nơ-ron nhân tạo VGG16

2.2.1. Giới thiệu về học sâu, mạng nơ-ron nhân tạo.

Học sâu là một nhánh của học máy, sử dụng mạng nơ-ron nhiều lớp để biểu diễn dữ liệu ở các mức độ trừu tượng khác nhau; các lớp sau trích xuất đặc trưng ngày càng rõ nét từ đầu ra lớp trước. Mạng được xây dựng qua hai giai đoạn chính: huấn luyện và đánh giá.

2.2.1.1. Quá trình huấn luyện, hàm mất mát và quá trình tối ưu hóa mạng.

Trong quá trình huấn luyện, với mỗi lô gồm B mẫu, quá trình huấn luyện gồm: **(i)** lan truyền tiến để tính đầu ra $\hat{y}$ cho từng mẫu **(ii)** tính hàm mất mát $l(\hat{y}, y)$, **(iii)** tính véc-tơ lượng giảm (gradient) **(iv)** cập nhật tham số tối ưu W thông qua các thuật toán tối ưu véc-tơ lượng giảm (gradient) như SGD, Adam,…

Hàm mất mát tổng quát trên một batch được tính bằng trung bình các mất mát từng mẫu. Gradient $g = \frac{\partial L}{\partial W}$ cho biết hướng tăng nhanh nhất của hàm mất mát; tham số được cập nhật theo công thức dạng $W_{t+1} = W_t - \alpha \times \Delta(g_t)$, với α là tốc độ học. Quá trình này lặp lại qua nhiều chu kỳ huấn luyện (epoch) cho đến khi đạt số epoch đặt trước hoặc thỏa điều kiện dừng.

2.2.1.2. Quá trình đánh giá mạng

Quá trình đánh giá mạng dùng dữ liệu chưa xuất hiện trong huấn luyện để đo độ chính xác mô hình. Tập dữ liệu thường được chia thành ba phần: huấn luyện, kiểm tra và kiểm thử. Các kỹ thuật như dừng sớm có thể được dùng để nâng cao hiệu quả huấn luyện.

Với bài toán phân loại nhị phân, đầu ra gồm hai nhãn dương/âm; kết quả dự đoán được phân thành TP, FP, TN, FN. Từ đó lập ma trận nhầm lẫn (confusion matrix) để thống kê số trường hợp dự đoán đúng/sai theo từng loại. Sau đó đánh giá mạng bằng các thông số: độ chính xác tổng thể (Accuracy), độ chính xác của mẫu dương (Precision), độ phủ của mẫu dương (Recall), điểm trung bình điều hòa của mẫu dương (F1 Score), APCER (tỉ lệ mẫu giả bị phân loại thành thật), BPCER (tỉ lệ mẫu thật bị phân loại thành giả). Hiệu năng mô hình được đánh giá qua: số lượng tham số, độ phức tạp tính toán (FLOPs), tốc độ thực thi.

2.2.2. Giới thiệu về mạng nơ-ron tích chập

Mạng nơ-ron tích chập (CNN) được thiết kế để xử lý dữ liệu hình ảnh và tín hiệu dạng lưới. Khác với mạng truyền thống trải phẳng toàn bộ ảnh và nối đầy đủ vào từng nơ-ron, CNN xử lý cục bộ bằng phép tích chập giữa bộ lọc và các vùng nhỏ của ảnh, giúp học đặc trưng không gian (cạnh, góc, họa tiết) và giảm số tham số. CNN gồm các lớp tích chập, lớp kích hoạt, lớp gộp, lớp kết nối đầy đủ, lớp chuẩn hóa và lớp lược bỏ; các lớp xếp chồng, đầu ra lớp trước là đầu vào lớp sau.

2.2.2.1. Lớp tích chập.

Lớp tích chập (Conv) nhận đầu vào kích thước $H \times W \times C$ và sử dụng K bộ lọc $f \times f$ (mỗi bộ lọc có C kênh) trượt trên ảnh, tính tích vô hướng với vùng con, cộng bias rồi đưa qua hàm kích hoạt, tạo đầu ra kích thước $(H' \times W' \times K)$ phụ thuộc vào H, W, f, s, p. Mỗi bộ lọc học một đặc trưng riêng (cạnh, góc, họa tiết…); trọng số ban đầu khởi tạo ngẫu nhiên và được điều chỉnh trong quá trình huấn luyện để véc-tơ đầu ra biểu diễn đặc trưng phù hợp. Lớp tích chập có thể phát hiện đồng thời nhiều đường biên, kết cấu, cũng như các mẫu vệt phức tạp.

2.2.2.2. Lớp kích hoạt.

Lớp kích hoạt nhận đầu ra sau tích chập và áp dụng các hàm phi tuyến như ReLU, Leaky ReLU, Sigmoid, Softmax, Tanh… để giúp mạng học đặc trưng phức tạp hơn và hỗ trợ hội tụ.

2.2.2.3. Lớp gộp

Lớp gộp có nhiệm vụ giảm kích thước véc-tơ đặc trưng, từ đó giảm số tham số, tăng khả năng khái quát hóa và hạn chế quá khớp. Hai dạng phổ biến là Max Pooling (chọn giá trị lớn nhất trong từng cửa sổ) và Average Pooling (lấy giá trị trung bình trong cửa sổ); Max Pooling thường được ưa chuộng hơn vì giữ được các đặc trưng cục bộ nổi bật.

2.2.2.4. Lớp kết nối đầy đủ

Lớp kết nối đầy đủ nhận véc-tơ đặc trưng (đã được làm phẳng) từ khối tích chập – gộp và thực hiện nhiệm vụ phân loại/hồi quy cuối cùng. Mỗi nơ-ron trong lớp kết nối đầy đủ liên kết với toàn bộ đầu vào, tính tổng trọng số + bias rồi qua hàm kích hoạt; các lớp này nối tiếp nhau thành khối kết nối đầy đủ. Ở bài toán phân loại, lớp kết nối đầy đủ cuối thường có số nơ-ron bằng số nhãn và dùng Softmax để quyết định nhãn, giúp kết hợp các đặc trưng đã trích xuất để phân loại chính xác..

2.2.2.5. Lớp lược bỏ

Lớp lược bỏ (dropout) giúp ngăn quá khớp bằng cách ngẫu nhiên vô hiệu hóa một tỷ lệ p nơ-ron trong quá trình huấn luyện, buộc mạng phân tán đặc trưng trên nhiều nơ-ron. Khi kiểm thử, tất cả nơ-ron đều được sử dụng nhưng đầu ra được nhân với $(1-p)$ để bù cho việc lược bỏ trong huấn luyện; lớp này thường đặt sau lớp kết nối đầy đủ hoặc lớp kích hoạt. `

2.2.2.6. Lớp chuẩn hóa tích chập

Lớp chuẩn hóa tích chập chuẩn hóa đầu vào của từng lô nhỏvề trung bình 0, độ lệch chuẩn 1, rồi nhân với hệ số tỉ lệ γ và cộng hệ số dịch β. Mục tiêu là giảm sự thay đổi phân phối đầu vào giữa các lớp trong quá trình huấn luyện, cho phép dùng learning rate lớn hơn và tối ưu mô hình nhanh hơn.

2.2.3. Mô hình mạng nơ-ron VGG16

2.2.3.1. Lịch sử ra đời.

Mô hình mạng nơ-ron tích chập VGG16 sử dụng nhiều lớp tích chập với cửa sổ lọc nhỏ 3×3 thay vì các lọc lớn nhằm tăng khả năng phân biệt và giảm số tham số. VGG16 là mô hình hiệu quả cho nhận diện, phân loại ảnh, đạt độ chính xác Top-5 92,7% trên tập ImageNet [16], và thường được dùng làm mô hình tiền huấn luyện trong học chuyển giao.

2.2.3.2 Cấu trúc mạng

VGG16 có khoảng 138 triệu tham số. Ảnh đầu vào được chuẩn hóa về kích thước $224 \times 224 \times 3$ và trừ đi giá trị trung bình RGB của tập huấn luyện trước khi đưa vào mạng. Các lớp tích chập dùng stride = 1, padding = 1 để giữ nguyên kích thước không gian, còn lớp gộp dùng Max Pooling 2×2 với stride = 2 để giảm một nửa chiều rộng và chiều cao sau mỗi lần gộp..

2.2.3.3 Ưu điểm và nhược điểm của VGG16

VGG16 có ưu điểm là số lớp tích chập nhiều, giúp học đặc trưng trừu tượng và đạt độ chính xác cao trên ImageNet [16]. Cấu trúc đồng nhất với bộ lọc 3×3 và hai lớp kết nối đầy đủ 4096 nơ-ron giúp việc cài đặt, điều chỉnh dễ dàng. Tuy nhiên, mô hình có 138 triệu tham số (~533MB), đòi hỏi bộ nhớ và năng lực tính toán lớn, nên không phù hợp với hệ thống yêu cầu độ trễ thấp hoặc bộ nhớ hạn chế.

2.2.3.4. Ứng dụng của VGG16

VGG16 vẫn được dùng rộng rãi trong phân loại ảnh, phát hiện vật thể và trích xuất đặc trưng. VGG16 cũng hay được dùng như bộ trích xuất đặc trưng cho các bài toán khác (như zero-shot).

2.3. Giới thiệu về mô-đun tự chú ý

Mô-đun tự chú ý (self-attention) do Vaswani và cộng sự (2017) [17] đề xuất cho phép mô hình xác định mức độ liên quan giữa các phần tử trong cùng một chuỗi đầu vào, khắc phục hạn chế của mạng hồi quy truyền thống vốn phụ thuộc vào tính toán tuần tự, khó huấn luyện song song và dễ gặp vấn đề tiêu biến/bùng nổ gradient, đồng thời cải thiện khả năng học ngữ cảnh dài hạn.

Từ tập véc-tơ đầu vào, mô-đun tự chú ý xây dựng ba ma trận Q (Query), K (Key) và V (Value) bằng các phép chiếu tuyến tính, sau đó tính trọng số chú ý từ tích vô hướng tỉ lệ giữa Q và K, chuẩn hóa bằng hàm softmax để thu được tổ hợp tuyến tính các véc-tơ V. Các tầng tự chú ý có thể được nối hoặc cộng để tạo thành mô-đun tự chú ý đa tầng (multi-head attention). Bên cạnh đó, mô-đun chú ý chéo (cross-attention) có cấu trúc tương tự nhưng cho phép một chuỗi truy vấn tập trung vào các vị trí quan trọng của một chuỗi ngữ cảnh khác.

2.4. Giới thiệu về mô hình VGG16 kết hợp phép biến đổi Fourier.

Các ảnh thẻ căn cước giả mạo thường có những đặc trưng dễ nhận biết ở miền không gian như viền màn hình, mép khung thiết bị, được VGG16 trích xuất và phân loại. Tuy nhiên, chỉ dùng đặc trưng không gian có thể bỏ sót các dấu hiệu tinh vi như độ mờ, vân nhiễu. Do đó, mô hình kết hợp thêm miền tần số: phép biến đổi Fourier cùng bộ lọc tần số cao giúp khuếch đại các vân sọc do màn hình (nhiễu moiré) và sự khác biệt giữa ảnh mờ (chụp qua màn hình) và ảnh sắc nét (chụp trực tiếp), từ đó giúp phát hiện ảnh giả mạo hiệu quả hơn.

Về ưu điểm của mô hình hai nhánh VGG16 kết hợp phép biến đổi Fourier, việc kết hợp đặc trưng hình ảnh theo cả hai miền không gian và tần số giúp tăng khả năng phân biệt các mẫu khó, giảm thiểu số tham số cần huấn luyện, tăng tốc độ huấn luyện cho mô hình.

2.5. Kết luận

Chương này trình bày cơ sở lý thuyết cho mô hình phát hiện thẻ căn cước giả mạo, gồm: (i) phép biến đổi Fourier rời rạc và bộ lọc lý tưởng tần số cao cho ảnh số; (ii) mạng nơ-ron VGG16, các thành phần, cách huấn luyện và kiểm thử; (iii) mô-đun tự chú ý; và (iv) phương pháp kết hợp VGG16 với FFT và tự chú ý để vừa khai thác đặc trưng không gian, vừa tận dụng đặc trưng tần số cao nhằm phát hiện các dấu hiệu giả mạo tinh vi. Đây là nền tảng lý thuyết cho phần thực nghiệm ở chương sau.

CHƯƠNG 3 - XÂY DỰNG VÀ ĐÁNH GIÁ MÔ HÌNH

Chương này tập trung vào việc xây dựng và đánh giá mô hình phát hiện hình ảnh giả mạo từ dữ liệu thẻ căn cước. Kết quả cần đạt là một mạng nơ-ron nhân tạo kết hợp được ưu điểm của mạng VGG16 và phép biến đổi Fourier có kiến trúc đơn giản, tốc độ suy luận tốt và có hiệu quả tốt trong việc phát hiện ảnh thẻ giả mạo. Ngoài ra, mô hình cũng phải có khả năng mở rộng, dễ cải tiến.

3.1. Phân tích bài toán phát hiện thẻ căn cước giả mạo

Bài toán phát hiện thẻ căn cước giả mạo là một bài toán phân loại nhị phân trong đó: Ảnh thật là những bức ảnh được chụp trực tiếp và chân thực, tức là thẻ căn cước được đặt trước camera, không có bất kỳ bước trung gian nào gây suy giảm chất lượng. Ảnh giả là ảnh thẻ căn cước hiển thị trên thiết bị điện tử (điện thoại, máy tính bảng…) và chụp lại bằng thiết bị thu nhận ảnh. Đầu vào của bài toán này là một bức ảnh kỹ thuật số của thẻ căn cước định dạng RGB. Đầu ra là một nhãn phân loại rời rạc, trong đó: nhãn âm tương đương với “thật” (“genuine”); nhãn dương tương đương với nhãn “giả mạo” (“spoof”).

Để khẳng định tính đúng đắn và hiệu quả của mô hình đề xuất, quá trình thực nghiệm trong đề án được thiết kế với hai kịch bản thử nghiệm chính: (1) đánh giá so sánh với các nghiên cứu trước đây và (2) nghiên cứu loại bỏ thành phần.

3.2. Xây dựng tập dữ liệu

Đề án sử dụng hai bộ dữ liệu mở DLC-2021 [2] và MIDV-2020 [4] về giấy tờ tùy thân (thẻ căn cước, hộ chiếu) với cả ảnh thật và giả. Trong khuôn khổ đề án, chỉ dùng hai nhãn “or” (genuine/ảnh thật) và “re” (spoof/ảnh giả chụp từ màn hình). Tập huấn luyện + kiểm tra lấy 25.980 ảnh thật từ “MIDV-2020/clips” và 19.543 ảnh giả từ ba thư mục esp_id, lva_passport, rus_internalpassport, rồi chia ngẫu nhiên: 80% cho huấn luyện, 20% cho kiểm tra. Tập kiểm thử gồm 11.006 ảnh giả (nhãn dương)

và 16.264 ảnh thật (nhãn âm) liệt kê trong "screen_positive_test.lst" và "screen_negative_test.lst", trích từ các loại giấy tờ còn lại của DLC-2021 [2].

3.3. Xây dựng kiến trúc mạng học sâu.

Đề án sử dụng ngôn ngữ lập trình Python và các thư viện học sâu PyTorch, scikit-learn để xây dựng kiến trúc mạng học sâu cho bài toán phát hiện ảnh thẻ căn cước giả mạo.

3.3.1. Kiến trúc tổng quan của mạng

Kiến trúc đề xuất gồm hai nhánh song song: nhánh chính xử lý trực tiếp ảnh gốc và nhánh FFT xử lý ảnh sau khi qua biến đổi Fourier và lọc tần số cao. Ảnh RGB đầu vào kích thước 1080×1920 hoặc 2160×3840 được chuẩn hóa về [0;1].

Ở nhánh chính, ảnh được đưa qua khung VGG16 và chỉ giữ đến hết khối thứ 3, cho ra tensor đặc trưng $(B, 256, 28, 28)$. Ở nhánh FFT, ảnh được tiền xử lý bằng FFT + lọc tần số cao để làm nổi bật thành phần tần số cao, cũng được đưa qua một khung VGG16 tương tự, thu được tensor đặc trưng $(B, 256, 28, 28)$.

Hai tensor đặc trưng này được đưa vào mô-đun chú ý chéo đa tầng để học tương tác giữa đặc trưng miền không gian và đặc trưng tần số, tạo tensor hợp nhất $(B, 512)$. Tensor này sau đó đi qua ba lớp kết nối đầy đủ (có thể kèm chuẩn hóa) và lớp softmax để phân loại nhị phân hai lớp "genuine" và "spoof". Chi tiết xử lý trong từng mô-đun được trình bày trong các phần dưới.

3.3.2. Mô-đun tiền xử lý: phép biến đổi Fourier rời rạc nhanh kết hợp cơ chế lọc tần số cao

Để phục vụ cho nhánh FFT, cần chuẩn bị đầu vào cho nhánh này. Sau khi xử lý, sẽ thu được ảnh đã lọc tần số cao. Tiến hành áp dụng tuần tự các bước sau:

- Bước 1: Cắt và chuẩn hóa vùng cần quan tâm

Xác định tâm thẻ căn cước, cắt hai ô vuông: $L_1 = 896 \times 896$ (cho nhánh chính VGG16) và $L_2 = 448 \times 448$ (cho nhánh tần số FFT). Cả hai vùng đều được chỉnh kích thước về 224×224. L_1 được đưa vào nhánh chính, L_2 được đưa vào nhánh FFT. L_1 cung cấp vùng ngữ cảnh rộng cho các đặc trưng không gian, L_2 tập trung vào các chi tiết vi mô (ví dụ: vân nhiễu).

- Bước 2: Biến đổi Fourier rời rạc nhanh, dịch ảnh và áp dụng bộ lọc

Chuyển vùng L_2 sang ảnh xám, áp dụng FFT và FFTShift để đưa phổ tần số về tâm. Tạo mặt nạ high-pass hình tròn với bán kính $r = 8$, giữ lại thành phần tần số cao và loại bỏ tần số thấp, thu được phổ đã lọc.

- Bước 3: Dịch ngược và biến đổi ngược:

Thực hiện IFFTShift và IFFT để tái tạo ảnh thành phần tần số cao trong miền không gian. Lấy độ lớn (magnitude) của tensor phức và áp dụng biến đổi log để nén giá trị, làm nổi bật chi tiết nhỏ.

- Bước 4: Chuẩn hóa cực tiểu – cực đại

Chuẩn hóa cực tiểu – cực đại giúp nén giá trị ảnh về khoảng [0, 1], thu được tensor $(B, 1, 224, 224)$. Lặp kênh để thu được đầu vào hợp lệ có kích thước $(B, 3, 224, 224)$. Cuối cùng chuẩn hóa theo chuẩn ImageNet [16].

3.3.3. Mô-đun mạng nơ-ron tích chập khung sử dụng kiến trúc VGG16

Trong mô-đun này, sử dụng kiến trúc VGG16 và chỉ lấy 3 lớp đầu tiên trong mạng để giảm thiểu kích thước. Kích thước chi tiết các lớp được thể hiện trong bảng 3.2. Cả hai nhánh đều dùng phương pháp học chuyển giao với tham số được huấn luyện sẵn trên trên ImageNet [16].

3.3.4. Mô-đun chú ý chéo đa tầng

Mô-đun chú ý chéo đa tầng nhận hai tensor đặc trưng từ nhánh chính và nhánh FFT, cùng kích thước $(B, 256,28,28)$. Đầu tiên, hai tensor được làm phẳng theo chiều không gian và hoán vị để đưa về dạng chuỗi, rồi chiếu tuyến tính thành các bộ Q, K, V riêng cho mỗi nhánh. Tiếp theo, tính giá trị chú ý chéo đa tầng hai chiều:

a. Tính giá trị chú ý chéo của ảnh FFT đến ảnh RGB:

$$Attention_{RGB \leftarrow FFT} = softmax\left(\frac{Q_{RGB}K_{FFT}^{T}}{\sqrt{d_k}}\right)V_{FFT} \in \mathbb{R}^{(HW)\times B\times d_v} \quad (3.1)$$

b. Tính giá trị tự chú ý của ảnh RGB đến ảnh FFT:

$$Attention_{FFT \leftarrow RGB} = softmax\left(\frac{Q_{FFT}K_{RGB}^{T}}{\sqrt{d_k}}\right)V_{RGB} \in \mathbb{R}^{(HW)\times B\times d_v} \quad (3.2)$$

Kết quả từ các tầng được ghép lại, qua lớp tuyến tính để thu về kích thước $(HW) \times B \times C$ rồi cộng tắt (residual) với đặc trưng ban đầu để bảo toàn thông tin. Sau đó hoán vị lại, áp dụng global average pooling theo không gian, thu được hai véc-tơ $B \times C$ cho nhánh chính và nhánh FFT, rồi nối chúng thành véc-tơ chung Z $\in \mathbb{R}^{B\times 2C}$ làm đầu vào cho mô-đun phân loại.

3.3.5. Mô-đun phân loại

Mô-đun phân loại nhận tensor hợp nhất Z kích thước $(B, 2C)$ (với $C = 256$) sau khi kết hợp đặc trưng từ nhánh chính và nhánh FFT. Tensor này được đưa qua ba lớp kết nối đầy đủ liên tiếp: hai lớp đầu có cấu trúc tuyến tính, chuẩn hóa, ReLU và dropout, lớp cuối là tuyến tính cho ra 2 đầu ra tương ứng với hai nhãn "thật" và "giả". Các lớp này có nhiệm vụ trích lọc và tách biệt đặc trưng cuối cùng phục vụ phân loại.

3.3. Hàm mất mát và quá trình tối ưu hóa

Sau mô-đun phân loại, mô hình tạo ra hai giá trị đầu ra tương ứng với hai nhãn "thật" và "giả", được chuẩn hóa bằng hàm Softmax để thu được xác suất. Hàm mất mát sử dụng là Cross-Entropy nhị phân, từ đó tính lượng giảm (gradient) theo các tham số mạng và cập nhật bằng bộ tối ưu Adam với tốc độ học $\alpha = 10^{-4}$, các tham số $\beta_1 = 0.9, \beta_2 = 0.999$. Cách tối ưu này giúp mô hình học được đặc trưng phức tạp từ hai nhánh (nhánh chính và nhánh FFT), hội tụ nhanh mà vẫn hạn chế hiện tượng quá khớp.

3.4. Huấn luyện và đánh giá mô hình.

3.4.1. Cài đặt và huấn luyện mô hình.

Trong quá trình huấn luyện, mô hình sử dụng kích thước lô $B = 64$, dữ liệu được nạp song song với 7 luồng (workers) và được đảo ngẫu nhiên trước khi chia lô để tăng tính đa dạng và khả năng khái quát. Mô hình VGG16 kết hợp nhánh FFT được huấn luyện 5 chu kỳ với tổng 3644 bước trên GPU L4 của Google Colab [18] thông qua PyTorch Lightning và CUDA, giúp rút ngắn đáng kể thời gian huấn luyện; trong suốt quá trình, giá trị hàm mất mát được theo dõi liên tục. Hàm mất mát của tập huấn luyện và kiểm tra giảm ổn định theo thời gian, từ khoảng 0.2 xuống quanh 0.02 và không có dao động bất thường hay tăng trở lại, cho thấy mô hình hội tụ nhanh và học tốt các đặc trưng mà không bị mất ổn định trong quá trình huấn luyện.

3.4.2. Đánh giá mô hình.

Sau khi huấn luyện, mô hình được đánh giá trên tập kiểm thử chuẩn của DLC-2021 [2] với các chỉ số precision, recall, accuracy để so sánh với mô hình DLC-2021 cơ sở [2] và phương pháp dò nhiễu theo biên [5]. Đồng thời, các chỉ số FLOPs, số lượng tham số (Param) và tốc độ suy luận (Speed) cũng được đo lường; kết quả của ba mô hình được tổng hợp trong bảng 3.1.

Model	Precision (%)	Recall (%)	Accuracy (%)
DLC-2021 cơ sở	85.89	89.03	89.67
Bộ dò nhiễu theo biên	93.5	90.2	92.1
VGG16+FFT	90.62	92.49	91.71

Bảng 3.1: Kết quả đo trên mô hình DLC-2021 cơ sở, Bộ dò nhiễu theo biên và VGG16+FFT

Với cùng bộ dữ liệu và cách đánh giá, VGG16+FFT đạt accuracy 91.71%, precision 90.62% và recall 92.49%, cao hơn mô hình DLC-2021 cơ sở [2]. So với Bộ dò nhiễu theo biên [5], VGG16+FFT chỉ kém nhẹ về accuracy (91.71% so với 92.1%) và precision (90.62% so với 93.5%) nhưng có recall cao hơn (92.49% so với 90.2%), cho thấy mô hình ưu tiên phát hiện đầy đủ các mẫu giả mạo trong khi vẫn giữ độ chính xác tổng thể. Các kết quả này khẳng định VGG16+FFT là phương án mô hình hóa hiệu quả và ổn định cho bài toán phát hiện giả mạo trên DLC-2021.

Tốc độ thực thi được đánh giá trên CPU Intel Xeon @ 2.20 GHz (6 nhân/12 luồng) bằng cách đo thời gian suy luận trung bình trên hai thư mục grc_passport/01.or0004 (ảnh thật) và grc_passport/01.re0001 (ảnh giả). VGG16+FFT đạt thời gian 0.120 s/ảnh cho cả ảnh thật và giả, nhanh hơn DLC-2021 cơ sở (0.147–0.151 s/ảnh), trong khi thời gian của Bộ dò nhiễu theo biên không có sẵn. Mô hình VGG16+FFT chỉ dùng khoảng 4 triệu tham số (42 MB), có độ phức tạp khoảng 40.95 GFLOPs do kiến trúc hai nhánh và mô-đun chú ý chéo đa tầng, đồng thời đạt BPCER 9.03% và APCER 7.51%. Kết quả tổng hợp được trình bày trong Bảng 3.2. Nhìn chung, trong cùng giao thức đánh giá, VGG16+FFT cho độ chính xác xấp xỉ phương pháp dò nhiễu theo biên và tốt hơn rõ rệt DLC-2021 cơ sở, đồng thời đáp ứng yêu cầu vừa tin cậy vừa suy luận hiệu quả trên CPU.

Model	grc_passport/01.or0004	grc_passport/01.re0001
DLC-2021 cơ sở	0.147	0.151
Bộ dò nhiễu theo biên	-	-

VGG16+FFT	0.120	0.120

Bảng 3.2: Kết quả thời gian thực thực thi trung bình trên CPU Intel Xeon @ 2.20 GHz (x86-64), 6 nhân / 12 luồng

3.4.3. Nghiên cứu loại bỏ thành phần

Các khảo sát loại bỏ thành phần đều dùng chung bộ dữ liệu và giao thức đánh giá DLC-2021 [2] với các chỉ số precision, recall, accuracy; mô hình tham chiếu là VGG16+FFT (precision 90.62%, recall 92.49%, accuracy 91.71%). Ba biến thể được khảo sát và kết quả tương ứng như sau:

1. Thay mô-đun chú ý chéo đa tầng bằng ConcatGAP: precision giảm còn 91.29%, recall còn 83.44% và accuracy còn 88.10%. Mất cơ chế căn chỉnh linh hoạt giữa hai miền khiến nhiều mẫu giả mạo bị bỏ sót, cho thấy chú ý chéo đa tầng quan trọng trong việc hợp nhất tín hiệu không gian và tần số.

2. Thay VGG16+FFT bằng hai nhánh VGG16 không gian (không có nhánh FFT): đạt precision 91.38%, recall 84.80% và accuracy 88.73%. Thiếu đặc trưng miền tần số khiến mô hình bỏ lỡ các nhiễu, khẳng định đặc trưng FFT không dư thừa so với đặc trưng không gian.

3. Chỉ dùng một nhánh VGG16 không gian, bỏ nhánh FFT và mô-đun chú ý chéo: recall và accuracy đều giảm so với mốc tham chiếu, do thiếu cả gợi ý tần số bổ sung lẫn cơ chế căn chỉnh/tổng hợp thông tin; mô-đun chú ý chéo đa tầng không thể áp dụng khi kiến trúc chỉ có một nhánh.

3.5. Kết luận

Chương này trình bày cấu trúc mô hình VGG16 kết hợp phép biến đổi Fourier cho bài toán phát hiện thẻ căn cước giả mạo, cùng quy trình huấn luyện và đánh giá. Kết quả thực nghiệm cho thấy mô hình hội tụ ổn định, đạt F1 cao, APCER và BPCER thấp, thể hiện khả năng phân biệt tốt giữa mẫu thật và giả. Nhánh FFT bổ sung thông

tin tần số giúp phát hiện các dấu hiệu giả mạo tinh vi trên ảnh thẻ căn cước, trong khi quy trình huấn luyện và đánh giá đa chiều giúp đảm bảo tính khách quan và hạn chế quá khớp, khẳng định tính phù hợp của kiến trúc VGG16+FFT và tạo nền tảng cho các cải tiến tiếp theo.

KẾT LUẬN

Đề án đã giải quyết một bài toán có tính thực tiễn cao: phát hiện hình ảnh thẻ căn cước giả mạo trong bối cảnh định danh điện tử. Mô hình đề xuất kết hợp mạng nơ-ron VGG16 với phép biến đổi Fourier cho thấy khả năng nhận diện dấu hiệu giả mạo tốt, tốc độ thực thi nhanh, góp phần nâng cao độ chính xác, giảm rủi ro gian lận và cải thiện trải nghiệm người dùng trong các hệ thống định danh điện tử.

Kết quả thực nghiệm chứng minh rằng việc kết hợp một kiến trúc nền tảng như VGG16 với các kỹ thuật xử lý ảnh theo miền tần số có thể tạo ra mô hình vừa hiệu quả, vừa gọn nhẹ, dễ triển khai thực tế, đồng thời mở ra hướng cho các hệ thống định danh điện tử an toàn hơn.

Hướng nghiên cứu tiếp theo gồm: mở rộng sang nhiều kiểu giả mạo khác (ảnh in màu, đen trắng, có/không ép plastic), tăng độ đa dạng tập dữ liệu, tối ưu thêm kiến trúc VGG16+FFT và thử nghiệm các kiến trúc khung khác để tiếp tục cải thiện độ chính xác và tốc độ xử lý.

TÀI LIỆU THAM KHẢO

[1] Rafael C. Gonzalez, Richard E. Woods, Digital Image Processing, Pearson Education, 2018.

[2] Polevoy, Dmitry V. and Sigareva, Irina V. and Ershova, Daria M. and Arlazarov, Vladimir V. and Nikolaev, Dmitry P. and Ming, Zuheng and Luqman, Muhammad Muzzamil and Burie, Jean-Christophe, "Document Liveness Challenge Dataset (DLC-2021)," *Journal of Imaging,* tập 8, 2022.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, "Deep Residual Learning for Image Recognition," *Computer Vision and Pattern Recognition (cs.CV),* 2015.

[4] Konstantin Bulatov, Ekaterina Emelianova, Daniil Tropin, Natalya Skoryukina, Yulia Chernyshova, Alexander Sheshkus, Sergey Usilin, Zuheng Ming, Jean-Christophe Burie, Muhammad Muzzamil Luqman, Vladimir V. Arlazarov, "MIDV-2020: A Comprehensive Benchmark Dataset for Identity Document Analysis," *Computer Optics,* tập 46, số 2, pp. 252-270, 2022.

[5] I.A. Kunina, A.V. Sher, D.P. Nikolaev, "Screen recapture detection based on color-texture analysis of document boundary regions," *Computer Optics,* tập 47, số 4, pp. 650-657, 2023.

[6] Reuben Markham, Juan M. Espín, Mario Nieto-Hidalgo, Juan E. Tapia, "Open-Set: ID Card Presentation Attack Detection using Neural Transfer Style," 2023.

[7] "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 4510-4520, 2018.

[8] Musab Al-Ghadi; Joris Voerman, Mickaël Coustaty, Olivier Lessard, Nicolas Sidere, "IDTrust: Deep Identity Document Quality Detection with Bandpass Filtering," *2025 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW),* 2025.

[9] Chuo Jun Zhang, Asif Q. Gill, Bo Liu, Memoona J. Anwar, “AI-based Identity Fraud Detection: A Systematic Review,” *arXiv:2501.09239.*

[10] Dingheng Zeng, Liang Gao, Hao Fang, Guohui Xiang, Yue Feng, Quan Lu, "Bandpass Filter Based Dual-stream Network for Face Anti-spoofing," 2023.

[11] Yanyu Li, Ju Hu, Yang Wen, Georgios Evangelidis, Kamyar Salahi, Yanzhi Wang, Sergey Tulyakov, Jian Ren, “Rethinking Vision Transformers for MobileNet Size and Speed,” *Computer Vision and Pattern Recognition,* 2022.

[12] Zitong Yu, Chenxu Zhao, Zezheng Wang, Yunxiao Qin, “Searching Central Difference Convolutional Networks for Face Anti-Spoofing,” 2020.

[13] Boulkenafet, Zinelabinde and Komulainen, Jukka and Li, Lei and Feng, Xiaoyi and Hadid, Abdenour, “{OULU-NPU}: A Mobile Face Presentation Attack Database with Real-World Variations,” trong *12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017)*, 2017, pp. 612-618.

[14] Zhiwei Zhang; Junjie Yan; Sifei Liu; Zhen Lei; Dong Yi; Stan Z. Li, “A face antispoofing database with diverse attacks,” trong *2012 5th IAPR International Conference on Biometrics (ICB)*, IEEE.

[15] Mark S. Nixon & Alberto S. Aguado, Feature Extraction and Image Processing, Academic Press publications, 2008.

[16] J. Deng, W. Dong, R. Socher, L. -J. Li, Kai Li and Li Fei-Fei, “ImageNet: A large-scale hierarchical image database,” *2009 IEEE Conference on Computer Vision and Pattern Recognition,* pp. 248-255, 2009.

[17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin, “Attention Is All You Need,” *Computation and Language (cs.CL); Machine Learning (cs.LG),* 2017.

[18] G. C. Platform, “GPUs on Compute Engine – GPU machine types,” Google, 2024. [Trực tuyến]. Available: https://cloud.google.com/compute/docs/gpus.