HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

PHẠM THANH HÙNG

PHÁT HIỆN GIẢ MẠO KHUÔN MẶT SỬ DỤNG MẠNG HỌC SÂU

TÓM TẮT LUẬN VĂN THẠC SĨ KỸ THUẬT

(Theo định hướng ứng dụng)

HÀ NỘI - 2022

Luận văn được hoàn thành tại:

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Người hướng dẫn khoa học: GS. TS. TỪ MINH PHƯƠNG

Phản biện 1: PGS.TS Phạm Văn Cường……

Phản biện 2: TS. Nguyễn Xuân Thắng..……

Luận văn sẽ được bảo vệ trước Hội đồng chấm luận văn thạc sĩ tại Học viện Công nghệ Bưu chính Viễn thông

Vào lúc: 08 giờ 30 ngày 17 tháng 12 năm 2022

Có thể tìm hiểu luận văn tại:

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

MỞ ĐẦU

Ngày nay, các ứng dụng của trí tuệ nhân tạo ngày càng trở lên phổ biến, một trong các ứng dụng đó là nhận diện khuôn mặt. Tuy nhiên, sự phát triển của các ứng dụng này cũng kéo theo một vấn đề đó là phát hiện giả mạo khuôn mặt. Thuật ngữ 'giả mạo khuôn mặt' ở đây nhằm nói đến việc xây dựng các khuôn mặt giả mạo của một người thật bằng nhiều cách khác nhau như lấy ảnh chụp của người đó in ra giấy hay quay lại được một video có khuôn mặt của họ. Tất cả các hành động trên nhằm đánh lừa hệ thống nhận diện khuôn mặt rằng nạn nhân đang có mặt tại thời điểm xác thực khuôn mặt, từ đó đạt được các mục đích xấu như vượt qua các biện pháp bảo mật nhằm đánh cắp tiền hay đánh cắp thông tin cá nhân v.v. Thêm vào đó, sự thành công đáng kinh ngạc của mạng nơ-ron tích chập (convolution neural network - CNN) trong cuộc thi ImageNet [59] đã thu hút rất nhiều sự chú ý của các nhà nghiên cứu trong mảng thị giác máy tính nhằm khai thác các khả năng tiềm ẩn của phương pháp học sâu. Sự cải tiến ngày càng tăng của mạng CNN nói chung về phân loại hình ảnh và phát hiện đối tượng đã mở ra các nhánh và ứng dụng tiềm năng của CNN trong lĩnh vực chống giả mạo khuôn mặt. Với các lý do trên, em đã chọn đề tài luận văn là "Phát hiện giả mạo khuôn mặt sử dụng mạng học sâu".

Nội dung luận văn được chia thành 3 chương như sau:

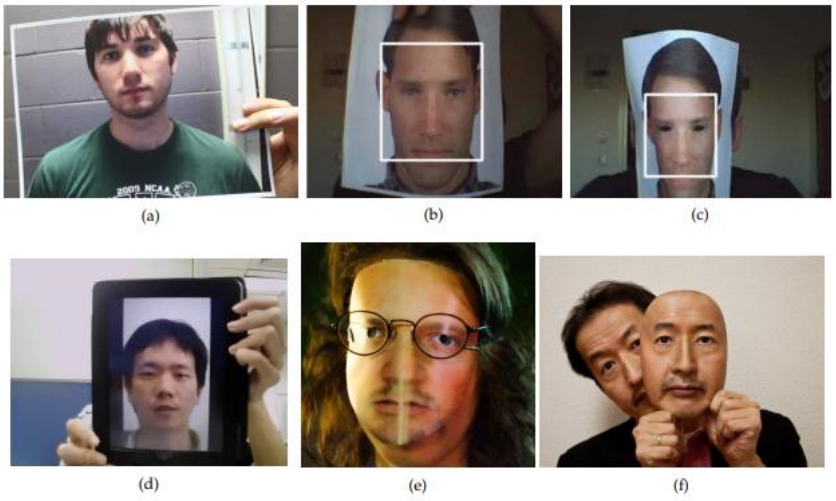
- CHƯƠNG 1: Bài toán phát hiện giả mạo khuôn mặt: Giới thiệu bài toán mà luận văn nghiên cứu và các nghiên cứu liên quan đã có.
- CHƯƠNG 2: Ứng dụng mạng học sâu vào bài toán phát hiện giả mạo khuôn mặt: Đưa ra một số lý thuyết về mạng học sâu, ý tưởng của việc đưa đặc trưng LBP vào mạng tích chập, cách tạo ảnh chiều sâu khuôn mặt từ mạng học sâu, giới thiệu mạng resnet-34, cách kết hợp các kỹ thuật trên. Bên cạnh đó, chương này sẽ nêu ra vấn đề thích ứng miền và ý tưởng sử dụng GAN để hạn chế vấn đề này.
- CHƯƠNG 3: Thử nghiệm và đánh giá: Trình bày về tập dữ liệu, các độ đo, các thử nghiệm, đưa ra các kết quả và rút ra kết luận.

CHƯƠNG 1: BÀI TOÁN PHÁT HIỆN GIẢ MẠO KHUÔN MẶT

Chương này sẽ trình bày định nghĩa của bài toán phát hiện giả mạo khuôn mặt cùng với các nghiên cứu liên quan tới bài toán này. Cụ thể, chương 1 sẽ giới thiệu các phương pháp dựa trên đặc trưng texture, các phương pháp dựa trên tương tác giữa người và máy, các thông tin về sự sống, chất lượng và chiều sâu của hình ảnh cũng được đề cập. Cuối cùng là các phương pháp dựa trên học sâu.

1.1. Giới thiệu bài toán phát hiện giả mạo khuôn mặt

Phát hiện giả mạo khuôn mặt là nhiệm vụ phát hiện hành vi xác minh khuôn mặt bằng cách sử dụng ảnh, video, mặt nạ hoặc một vật thay thế khác cho khuôn mặt của một người.



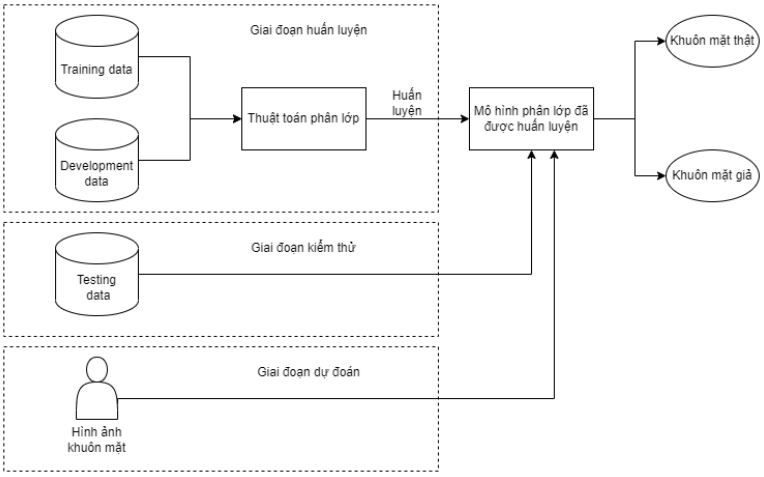
Hình 1-1: Các phương thức giả mạo khuôn mặt

CHƯƠNG 2: ỨNG DỤNG MẠNG HỌC SÂU VÀO BÀI TOÁN PHÁT HIỆN GIẢ MẠO KHUÔN MẶT

Chương này sẽ trình bày một số lý thuyết về mạng học sâu, ý tưởng đưa LBP vào mạng tích chập, sử dụng PRNet để tái tạo ảnh chiều sâu của khuôn mặt, giới thiệu về mạng resnet và cách kết hợp các kỹ thuật này lại với nhau để đưa ra được các phương pháp, kiến trúc mạng học sâu tương ứng cho bài toán phát hiện giả mạo. Đồng thời giới thiệu về vấn đề thích ứng miền và ý tưởng thử nghiệm nhằm khắc phục vấn đề này.

2.1. Ý tưởng giải quyết bài toán

Bài toán được đặt ra ban đầu trong luận văn đó là với một ảnh đầu vào, hệ thống phát hiện giả mạo cần trả về thông tin kết luận khuôn mặt xuất hiện trong bức ảnh đó là thật hay giả. Đây thực chất là một bài toán phân loại hai lớp. Vì vậy hướng tiếp cận của luận văn sẽ có 3 giai đoạn như hình 2-1.



Hình 2-1: Các giai đoạn trong quá trình xây dựng giải pháp phát hiện giả mạo khuôn mặt

Lấy cảm hứng từ các nghiên cứu được tìm hiểu ở chương 1, luận văn thấy rằng, có hai thông tin được sử dụng khá phổ biến để có thể phát hiện được khuôn mặt là thật hay giả mạo. Đó là đặc trưng LBP và thông tin về chiều sâu. Cụ thể, các mô hình mà luận văn thử nghiệm gồm: Resnet34, resnet34 kết hợp mạng tích chập khác biệt trung tâm (central difference convolution - CDC), resnet 34 kết hợp CDC và thông tin chiều sâu.

2.2. Ứng dụng học sâu vào bài toán phát hiện giả mạo khuôn mặt

2.2.1. Mạng tích chập khác biệt trung tâm (Central Difference Convolution - CDC)

Tương tự mạng tích chập thông thường, CDC cũng bao gồm hai bước lấy mẫu và tổng hợp. Việc lấy mẫu tương tự như trong tích chập thông thường trong khi bước tổng hợp có

khác biệt, CDC lấy tổng center-oriented gradient của các giá trị được lấy mẫu. Phương trình (2.6) trở thành

$$y(p_0) = \sum_{p_n \in R} w(p_n) \,.\, \left(x(p_0 + p_n) - x(p_0)\right) \quad (2.7)$$

Khi p_n = (0, 0), giá trị gradient luôn bằng 0 đối với chính vị trí trung tâm p_0.

Đối với bài toán phát hiện giả mạo khuôn mặt, cả thông tin ngữ nghĩa mức cường độ và thông tin chi tiết mức gradient đều rất quan trọng. Do đó CDC được tổng quát hóa thành:

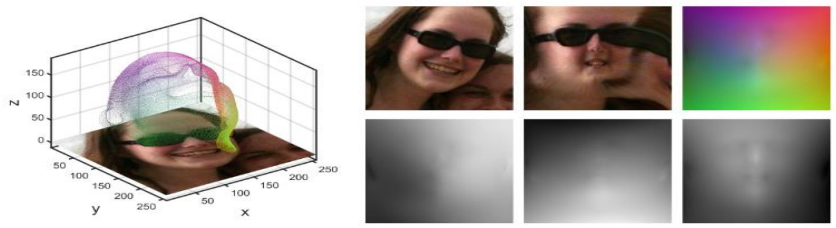
$$y(p_0) = (1 - \theta) \,.\, \sum_{p_n \in R} w(p_n) \,.\, x(p_0 + p_n) + \theta \,.\, \sum_{p_n \in R} w(p_n) \,.\, \left(x(p_0 + p_n) - x(p_0)\right) \quad (2.8)$$

Với thông số θ (theta) thuộc đoạn [0, 1] thể hiện tỷ lệ đóng góp giữa thông tin mức cường độ và mức gradient. Các giá trị của θ càng cao có nghĩa là thông tin về độ chênh lệch trung tâm càng có tầm quan trọng.

2.2.2. Tạo thông tin chiều sâu từ khuôn mặt

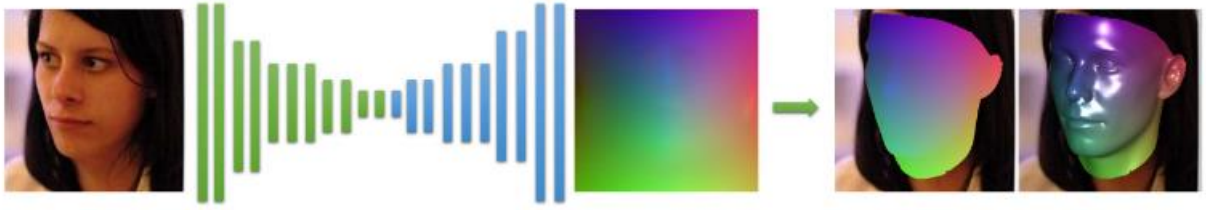
2.2.2.1. Biểu diễn khuôn mặt 3D

Để biểu diễn một khuôn mặt 3D trong máy tính, luận văn sử dụng một biểu đồ vị trí UV.



Hình 2-13: Hình minh họa của bản đồ vị trí UV. Bên trái: Biểu đồ 3D của hình ảnh đầu vào và đám mây điểm 3D của khuôn mặt. Bên phải: Hàng đầu tiên là hình ảnh 2D đầu vào, bản đồ kết cấu UV được trích xuất và bản đồ vị trí UV tương ứng. Hàng thứ hai là kênh x, y, z của bản đồ vị trí UV.

2.3.2.2. Kiến trúc mô hình và hàm mất mát



Hình 2-14: Kiến trúc của PRNet. Các hình chữ nhật màu xanh lá cây đại diện cho các khối Resnet và các hình chữ nhật màu xanh đại diện cho các lớp tích chập đã chuyển vị.

Mô hình học sâu dùng để cấu trúc lại thông tin chiều sâu của khuôn mặt tên là PRNet. Kiến trúc của PRNet được thể hiện trong hình 2-14.

Để học các tham số của mô hình, PRNet xây dựng một hàm mất mát mới để đo sự khác biệt giữa bản đồ vị trí và đầu ra của mô hình.

2.2.3. Mạng ResNet

ResNet, viết tắt của Residual Network là một loại mạng nơ-ron cụ thể đã được giới thiệu vào năm 2015 bởi Kaiming He, Xiangyu Zhang, Shaoqing Ren và Jian Sun.

2.2.4. Kết hợp CDC, thông tin chiều sâu và Resnet-34

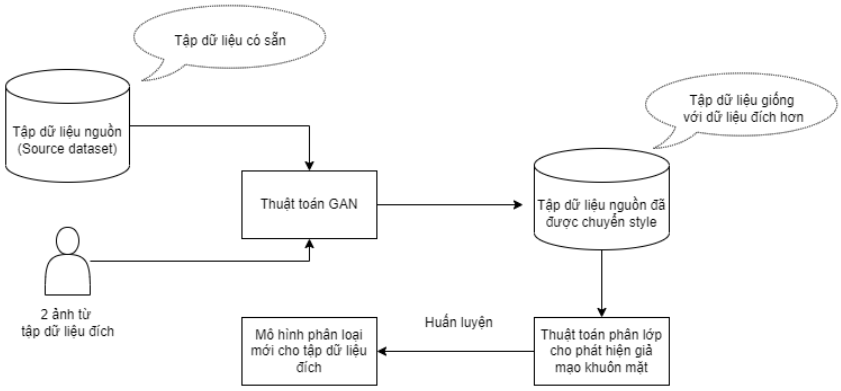
- Mô hình 1: Ban đầu mạng resnet-34 có lớp kết nối đầy đủ trả ra 1000 lớp do được huấn luyện với tập dữ liệu ImageNet. Tuy nhiên luận văn thay đổi để đầu ra chỉ trả về 2 giá trị để phù hợp với bài toán.
- Mô hình 2: Ở mô hình số 2, luận văn thực hiện thay thế toàn bộ các lớp tích chập ban đầu có trong mô hình 1 bằng lớp CDC.
- Mô hình 3: Ở mô hình số 3, luận văn tiếp tục thay đổi bắt nguồn từ mô hình số 2. Ở mô hình này luận vặn loại bỏ hẳn lớp kết nối đầy đủ cuối cùng. Sau đó, một lớp Upsample được đặt vào sau khối các lớp thứ 4 (gồm các lớp có số lượng kênh là 512) để thực hiện đưa feature map về kích thước 32x32. Bên cạnh đó, với mỗi một bức ảnh khuôn mặt đầu vào, luận văn sẽ đưa qua PRNet để nhận được một hình ảnh chiều sâu rồi tiếp tục đưa về kích thước 32x32.

2.3. Các vấn đề thích ứng miền

Nếu đầu vào ở giai đoạn kiểm thử khác đáng kể so với dữ liệu huấn luyện thì mô hình có thể không còn thực sự rất tốt nữa. Vấn đề này được gọi là vấn đề về thích ứng miền. Vấn đề này gặp nhiều trong bài toán phát hiện giả mạo khuôn mặt.

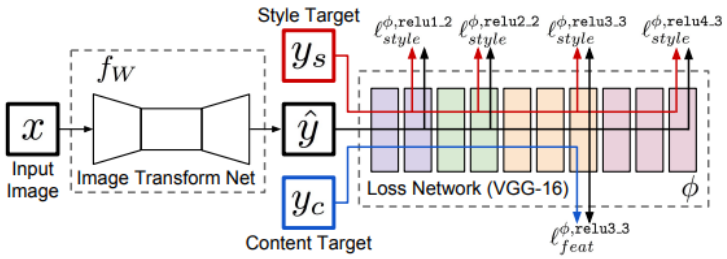
2.4. Ứng dụng GAN cho vấn đề thích ứng miền

Quá trình ứng dụng GAN của luận văn được mô tả qua hình 2-20.



Hình 2-20: Luồng thực hiện GAN trong luận văn

Để thực hiện ý tưởng này, luận văn có sử dụng một thuật toán chuyển kiểu được mô tả trong bài báo "Perceptual Losses for Real-Time Style Transfer and Super-Resolution" [55].



Hình 2-21: Kiến trúc tổng quan mạng chuyển đổi kiểu [55]

2.4.1. Mạng chuyển đổi hình ảnh

Layer	Activation size
Input	$3 \times 256 \times 256$
$32 \times 9 \times 9$ conv, stride 1	$32 \times 256 \times 256$
$64 \times 3 \times 3$ conv, stride 2	$64 \times 128 \times 128$
$128 \times 3 \times 3$ conv, stride 2	$128 \times 64 \times 64$
Residual block, 128 filters	$128 \times 64 \times 64$
Residual block, 128 filters	$128 \times 64 \times 64$
Residual block, 128 filters	$128 \times 64 \times 64$
Residual block, 128 filters	$128 \times 64 \times 64$
Residual block, 128 filters	$128 \times 64 \times 64$
$64 \times 3 \times 3$ conv, stride 1/2	$64 \times 128 \times 128$
$32 \times 3 \times 3$ conv, stride 1/2	$32 \times 256 \times 256$
$3 \times 9 \times 9$ conv, stride 1	$3 \times 256 \times 256$

Hình 2-22: Kiến trúc của mạng chuyển đổi kiểu

2.4.2. Hàm mất mát tri giác (Perceptual Loss function)

Phương pháp [55] định nghĩa hai hàm mất mát tri giác đo lường sự khác biệt về tri giác và ngữ nghĩa (semantic) của các hình ảnh ở mức cao.

Hàm mất mát tri giác đầu tiên đó là hàm mất mát tái cấu tạo đặc trưng (feature reconstruction loss). Phương pháp này sẽ thúc đẩy để các bức ảnh này tương tự nhau ở biểu diễn đặc trưng khi được tính bởi mạng mất mát Ø.

Hàm mất mát tri giác thứ hai đó là hàm mất mát tái cấu tạo kiểu (style reconstruction loss). Mục đích của hàm mất mát này sẽ là đo lường sự khác biệt giữa hai bức ảnh ở các đặc điểm như màu, cấu trúc, mẫu chung (common patterns), v.v.

2.5. Kết luận

Chương này đã trình bày các cơ sở lý thuyết của Central Difference Convolution, dựng thông tin chiều sâu cho khuôn mặt, residual network, và vấn đề thích ứng miền có thể gặp phải trong bài toán phát hiện giả mạo khuôn mặt và ý tưởng nhằm khắc phục vấn đề này dựa trên [55].

Chương tiếp theo sẽ mô tả quá trình chuyển bị dữ liệu, các độ đo nhằm đánh giá giải pháp cùng các thử nghiệm được thực hiện và kết quả cuối cùng.

CHƯƠNG 3: THỬ NGHIỆM VÀ ĐÁNH GIÁ

Chương 3 sẽ trình bày về các tập dữ liệu được sử dụng trong luận văn cũng như các bước tiền xử lý dữ liệu trong quá trình thực hiện, mô tả rõ hơn các thông số trong từng thực nghiệm như đã giới thiệu ở chương 2 cùng kết quả của các thực nghiệm đó, cuối cùng từ những kết quả này đưa ra một số nhận xét.

3.1. Dữ liệu thử nghiệm

3.1.1. Tập dữ liệu OULU

Tập dữ liệu phát hiện giả mạo khuôn mặt của Oulu-NPU bao gồm 4950 video về các cuộc tấn công, giả mạo và khuôn mặt thật. Sau khi rút gọn do dữ liệu gốc quá lớn và chia tập dữ liệu theo một cách chia sẵn có của OULU, luận văn sử dụng 1200 ảnh cho tập huấn luyện (training), 900 ảnh cho tập phát triển (development), 600 ảnh cho tập kiểm thử (testing).

3.1.2. Tập dữ liệu NUAA

Có tổng cộng 15 đối tượng (đánh số từ 1 đến 15) xuất hiện trong tập dữ liệu này. Trong mỗi khoảng thời gian, tập dữ liệu chứa cả ảnh chụp trực tiếp và ảnh chụp gián tiếp của các đối tượng.

Ba loại tấn công ảnh được mô phỏng trước webcam như trong hình 3-3.

Hình 3-3: Hình minh họa các cuộc tấn công ảnh khác nhau (từ trái sang phải): (1) di chuyển ảnh theo chiều ngang, theo chiều dọc, phía sau và phía trước; (2) xoay ảnh theo chiều sâu dọc theo trục dọc; (3) giống với (2) nhưng dọc theo trục hoành; (4) bẻ cong ảnh vào trong và ra ngoài theo trục tung; (5) giống như (4) nhưng dọc theo trục hoành.

3.2. Các độ đo

Để có thể đánh giá được các thuật toán, mô hình học sâu được thử nghiệm thì không thể thiếu đi được các độ đo phù hợp với bài toán. Cụ thể trong bài toán phát hiện giả mạo khuôn mặt, luận văn sử dụng 3 độ đo chính là APCER, BPCER và ACER.

3.3. Thực nghiệm

3.3.1. Thực nghiệm với riêng mạng resnet-34

Ở thí nghiệm này các thông số cho quá trình huấn luyện như số epoch, optimizer, learning rate, loss function lần lượt là 300, adam, 0.001, cross entropy. Pre-train model được huấn luyện từ tập dữ liệu ImageNet.

3.3.2. Thực nghiệm với mạng resnet-34 kết hợp CDC

Ở thực nghiệm này, các lớp tích chập trong mạng resnet-34 đã được thay thế bởi CDC, các thông số về số lượng epoch, optimizer, và loss function giống với ở mục 3.3.1.
Thông số theta cho CDC được chọn là 0.7. Các thông số về in channel, out channel, kernel size, stride, padding được giữ nguyên của các lớp tích chập tương ứng đã được thay thế.

3.3.3. Thực nghiệm với mạng resnet-34 kết hợp CDC và thông tin chiều sâu

Do quá trình sinh ra một ảnh chiều sâu cần nhiều thời gian, do vậy luận văn thực hiện sinh tất cả các hình ảnh chiều sâu cho từng khuôn mặt trước.

Trong quá trình huấn luyện, các hình ảnh chiều sâu này cũng sẽ được đọc cùng lúc với hình ảnh gốc. Tuy nhiên, đối với các hình ảnh giả mạo, luận văn sẽ thực hiện khởi tạo lại ma trận chiều sâu với toàn bộ các giá trị trong ma trận là 0.

Thông số cụ thể cho quá trình huấn luyện như sau: Số epoch là 300, optimizer là adam, learning rate là 0.0001, weight_decay là 5e-4, loss function là absolute loss và contrastive loss.

3.3.4. So sánh các kết quả thử nghiệm

Sau khi thực nghiệm với 3 phương án như trên, luận văn thu được kết quả sau:

Bảng 3-3: Kết quả thử nghiệm

	APCER	BPCER	ACER
Resnet-34	0.0958	0.175	0.1354
CDC + resnet-34	0.0313	0.266	0.1489
CDC + resnet-34 + thông tin chiều sâu	0.019	0.85	0.43

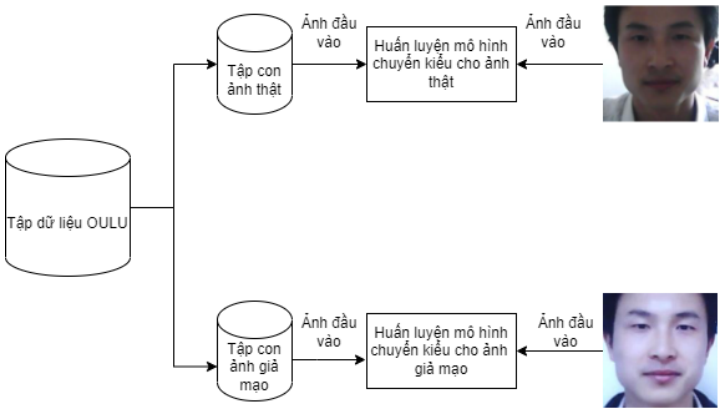
Qua kết quả trên, với độ đo APCER tỉ lệ lỗi giảm dần khi lần lượt kết hợp resnet-34 với CDC và ảnh chiều sâu, điều này cho thấy rằng CDC cùng với thông tin về chiều sâu có tác động tốt trong việc phát hiện được các trường hợp giả mạo, khi tỷ lệ phát hiện nhầm khuôn mặt giả mạo thành khuôn mặt thật là thấp.

Tuy nhiên điều ngược lại xuất hiện ở độ đo BPCER khi tỷ lệ lỗi lại tăng dần, đặc biệt CDC + resnet-34 + thông tin chiều sâu lại có tỉ lệ lỗi lớn hơn khá nhiều so với 2 thử nghiệm còn lại. Hiện tượng này có thể bắt nguồn từ cách khởi tạo ảnh chiều sâu cho ảnh giả mạo. Cụ thể hơn, hiện luận văn khởi tạo một ảnh chiều sâu màu đen cho ảnh giả mạo, tuy nhiên, ảnh giả mạo vẫn có thể có một chút thông tin chiều sâu tạo bởi bóng, màu sắc trên khuôn mặt nên khởi tạo hoàn toàn một màu đen có thể chưa biểu diễn được đặc trưng phân biệt giữ ảnh thật và ảnh giả mạo.

Với độ đo ACER là trung bình cộng của hai độ đo trên thì resnet-34 có nhỉnh hơn một chút so với CDC + resnet-34 tuy nhiên với mục đích là phát hiện khuôn mặt giả mạo mà vẫn cân bằng được 2 độ đo APCER và BPCER thì sự kết hợp giữa CDC và resnet-34 đang cho thấy những ưu thế của phương pháp này.

3.4. Thử nghiệm GAN trong vấn đề thích ứng miền

Do bài toán phát hiện giả mạo khuôn mặt có 2 lớp gồm khuôn mặt thật và khuôn mặt giả mạo nên sẽ cần 2 mô hình chuyển kiểu cho các lớp này. Quá trình phân chia dữ liệu được thể hiện tại hình 3-7.



Hình 3-7: Quá trình phân chia dữ liệu cho huấn luyện mô hình chuyển kiểu

Để huấn luyện cho mô hình chuyển kiểu luận văn sử dụng các thông số đầu vào như sau: Số epoch là 300, batch size là 4. Content weight là 1e5, style weight là 1e10, learning rate là 1e-3, optimizer là adam.

Bởi phương pháp [55] có mục tiêu là cực tiểu hóa hàm mất mát tối ưu nhất có thể nên luận văn sẽ đánh giá phương pháp này theo hai tiêu chí định tính và định lượng.

Về kết quả định tính, từ hình 3-8, luận văn thấy rằng dữ liệu ảnh được chuyển kiểu đang khác khá nhiều so với một hình ảnh chụp thông thường từ tập NUAA.

Hình 3-8: Từ trái qua phải là ảnh thật và ảnh giả mạo từ tập OULU đã chuyển kiểu

Về định lượng, tổng giá trị mất mát nằm trong khoảng 170000 và gần như không thay đổi quá nhiều từ epoch 150.

Từ các kết quả trên, luận văn thấy rằng việc ứng dụng [55] để tạo dữ liệu mới cho bài toán phát hiện giả mạo khuôn mặt hiện chưa đạt được như kỳ vọng của ý tưởng ban đầu.

3.5. Kết luận

Chương này đã trình bày về cách thu thập dữ liệu của các bộ dữ liệu được sử dụng trong luận văn, cùng với các độ đo dùng để đánh giá các mô hình. Tiếp theo đó là các thông số cụ thể đối với mỗi thử nghiệm. Qua đó luận văn thấy rằng, với mục đích phát hiện giả mạo khuôn mặt thì việc áp dụng CDC và thông tin chiều sâu đã giảm được tỷ lệ lỗi phát hiện sai các trường hợp giả mạo thành không giả mạo, từ đó đảm bảo được cho một hệ thống phát hiện khuôn mặt trở lên đáng tin cậy và an toàn hơn. Trong khi đó, phương pháp [55] hiện tại chưa thực sự phù hợp để giải quyết vấn đề thích ứng miền trong bài toán phát hiện giả mạo khuôn mặt.

KẾT LUẬN

Luận văn tập trung nghiên cứu các phương pháp nhằm phát hiện được các khuôn mặt giả mạo trong các hệ thống phát hiện khuôn mặt và đạt được các kết quả sau:

- Thực hiện khảo sát các phương pháp phát hiện giả mạo khuôn mặt
- Nghiên cứu và áp dụng CDC, sử dụng ảnh chiều sâu của khuôn mặt vào mạng resnet-34, so sánh về tác động của các kỹ thuật này trong việc phát hiện giả mạo khuôn mặt.
- Thử nghiệm sử dụng mô hình chuyển kiểu trong vấn đề thích ứng miền

Trong tương lai, luận văn có thể tiếp tục được nghiên cứu theo hướng giảm độ lỗi trong quá trình phát hiện khuôn mặt giả mạo, đồng thời nghiên cứu theo các phương pháp khắc phục vấn đề thích ứng miền trong bài toán này.