ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ

Phạm Hải Đăng

NGHIÊN CỨU, PHÁT TRIỂN PHƯƠNG PHÁP HỌC MÁY HIỆN ĐẠI ĐỂ NÂNG CAO HIỆU NĂNG TÁI ĐỊNH DANH NGƯỜI

Ngành đào tạo: Hệ thống thông tin
Mã: 9480104

TÓM TẮT LUẬN ÁN TIẾN SỸ
HỆ THỐNG THÔNG TIN

Hà Nội - 2025

Công trình được hoàn thành tại: Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội.

Người hướng dẫn khoa học:

- Phó giáo sư, Tiến sĩ Nguyễn Ngọc Hóa
- Phó giáo sư, Tiến sĩ Nguyễn Ngọc Tú

Phản biện: PGS.TS Phạm Văn Cường
Phản biện: PGS.TS Nguyễn Long Giang
Phản biện: PGS.TS Vũ Hải

Luận án sẽ được bảo vệ trước Hội đồng cấp Đại học Quốc gia chấm luận án tiến sĩ họp tại.....
vào hồi ngày tháng năm

Có thể tìm hiểu luận án tại:

- Thư viện quốc gia Việt Nam.
- Trung tâm Thông tin - Thư viện, Đại học Quốc gia Hà Nội.

MỞ ĐẦU

Bối cảnh và động lực nghiên cứu

Trong bối cảnh đô thị hóa và chuyển đổi số, hệ thống camera giám sát ngày càng quan trọng trong đảm bảo an ninh và quản lý trật tự. Tuy nhiên, để theo dõi liên tục một cá nhân khi họ di chuyển qua nhiều camera, cần đến bài toán tái định danh người (Person Re-Identification). Khác với nhận dạng khuôn mặt, tái định danh người khai thác đặc trưng toàn thân như trang phục, dáng đi hay phụ kiện, nhưng gặp thách thức lớn do thay đổi ánh sáng, tư thế, góc nhìn hoặc trang phục. Ban đầu, tái định danh người dựa trên đặc trưng thủ công (color histogram, HOG) và độ đo khoảng cách, song hiệu quả hạn chế. Từ năm 2014, học sâu trở thành hướng chủ đạo, với các mạng CNN, ViT tự động học đặc trưng giàu thông tin và phân biệt tốt hơn. Hiện có ba hướng tiếp cận chính: học có giám sát (SL) với dữ liệu gán nhãn đầy đủ, thích ứng miền không giám sát (UDA) khi mô hình cần chuyển sang miền mới không nhãn, và học không giám sát hoàn toàn (USL) trực tiếp từ dữ liệu thô. Luận án tập trung nghiên, cứu phát triển phương pháp nâng cao hiệu năng tái định danh người trong cả ba thiết lập SL, UDA và USL nhằm tối ưu hóa không gian đặc trưng, nâng cao độ chính xác và hướng tới ứng dụng trong giám sát an ninh, đô thị thông minh.

Thách thức và vấn đề nghiên cứu

Sự biến đổi ngoại hình, Hạn chế về nhãn dữ liệu, Khả năng biểu diễn đặc trưng và lựa chọn mạng cốt lõi phù hợp, Chất lượng nhãn giả và tính ổn định trong huấn luyện không giám sát, Khả năng mở rộng.

Mục tiêu và nội dung nghiên cứu

Mục tiêu: Nghiên cứu, phát triển các phương pháp mới nhằm nâng cao hiệu năng tái định danh người, tiếp cận theo cả ba hướng học có giám sát, thích ứng miền không giám sát và không giám sát.

Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu: Đối tượng là ảnh và tập ảnh người thu được từ các camera giám sát. Các bộ dữ liệu Market-1501[1], DukeMTMC-reID[2], CUHK03[3], MSMT17[4]. Hệ thống, mô hình, và phương pháp trích xuất, so sánh đặc trưng hình ảnh.

Phạm vi nghiên cứu: Đề xuất các giải pháp học máy hiện đại, thực nghiệm trên các bộ dữ liệu chuẩn. Nghiên cứu tập trung vào các

vấn đề cốt lõi ảnh hưởng trực tiếp đến hiệu năng không bao gồm các lớp bài toán chuyên biệt khác như thay đổi điều kiện ánh sáng (Cross-modality person ReID), thay đổi quần áo và ngoại hình (Long-term person ReID), che khuất (Occluded person ReID), hay khả năng mở rộng trong so khớp danh tính.

Các phương pháp nghiên cứu

Các phương pháp được sử dụng là lý thuyết, phân tích, mô hình hóa, thực nghiệm.

Đóng góp chính

Với mục tiêu nghiên cứu và phạm vi nghiên cứu được xác định rõ ràng, luận án có những đóng góp sau:

- Đóng góp 1: Theo hướng tiếp cận có giám sát, luận án đề xuất phương pháp "SCM-ReID" sử dụng học tương phản có giám sát kết hợp với bốn hàm mất mát khác để nâng cao hiệu năng tái định danh người. Công bố [CT2].

- Đóng góp 2: Theo hướng tiếp cân học thích ứng miền không giám sát, luận án đề xuất phương pháp hai phương án "IAQGA" và "DAPRH" để nâng cao hiệu năng theo cả hai giai đoạn là học có giám sát trên miền nguồn và học không giám sát trên miền đích. Công bố [CT4] với "IQAGA" và [CT1], [CT5] đối với "DAPRH".

- Đóng góp 3: Theo hướng tiếp cận học không giám sát, luận án đề xuất phương pháp "ViTC-UReID" sử dụng mạng cốt lõi ViT và tích hợp thông tin camera để nâng cao hiệu năng tái định danh người. Công bố [CT3].

Cấu trúc luận án

Luận án có bốn chương chính, đi kèm với phần mở đầu và kết luận. Các phần có nội dung cụ thể như sau. Chương 1 trình bày tổng quan và nền tảng lý thuyết. Chương 2 trình bày phương pháp đề xuất "SCM-ReID" theo hướng tiếp cận học có giám sát. Chương 3 trình bày hai phương pháp đề xuất "IQAGA" và "DAPRH" theo hướng tiếp cận học thích ứng miền không giám sát. Chương 4 trình bày phương pháp đề xuất "ViTC-UReID" theo hướng tiếp cận học không giám sát. Kết luận và hướng phát triển: trình bày tóm tắt các đóng góp chính, bàn luận, hạn chế và hướng phát triển.

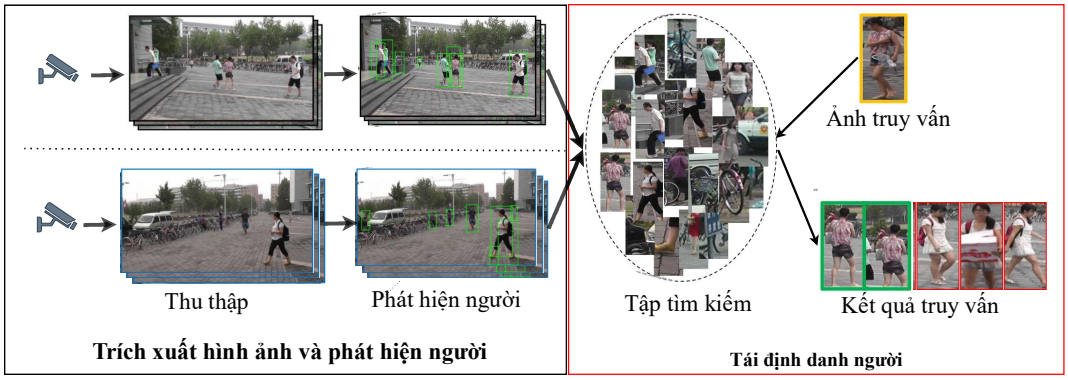
Chương 1
TỔNG QUAN NGHIÊN CỨU VÀ KIẾN THỨC NỀN TẢNG

Trong chương này, luận án trình bày tổng quan về bài toán tái định danh người, kiến thức nền tảng và tổng quan nghiên cứu liên quan, để chỉ ra khoảng trống nghiên cứu và hướng tiếp cận.

1.1 Giới thiệu bài toán tái định danh người

1.1.1 Phát biểu bài toán

Tái định danh người là bài toán nhận diện lại danh tính của một định danh qua tập hình ảnh thu từ nhiều camera trong điều kiện thay đổi về góc nhìn, ánh sáng và bối cảnh. Đây là một mô-đun quan trọng trong hệ thống tái định danh (Hình 1.1).



Hình 1.1: Hệ thống tái định danh người đầy đủ

Hàm mục tiêu của bài toán được biểu diễn như sau:

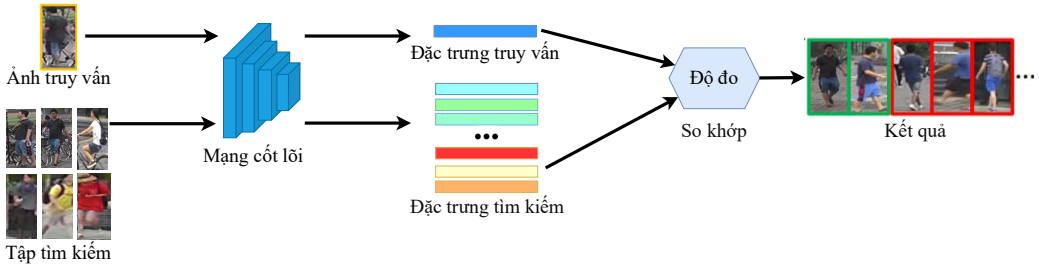
$$\mathcal{R}(q_i, G) = \text{Sort}\left(\{g_j\}_{j=1}^{N} \mid \text{sim}(q_i, g_j)\right), \tag{1.1}$$

Khi chúng ta chỉ quan tâm đến k kết quả có độ tương đồng cao nhất (truy xuất top-k). Khi đó, hàm mục tiêu được điều chỉnh như sau:

$$\mathcal{R}_k(q_i, G) = \text{Top}_k\left(\mathcal{R}(q_i, G)\right) \tag{1.2}$$

1.1.2 Phương án so khớp hình ảnh

Quy trình này được minh họa trong Hình 1.2.



Hình 1.2: Quá trình trích xuất đặc trưng và so khớp

1.1.3 Học có giám sát

Học có giám sát là hướng tiếp cận truyền thống và hiệu quả trong bài toán tái định danh người, nơi mô hình được huấn luyện trên dữ liệu đã gán nhãn danh tính.

Mô hình hóa: Mục tiêu của hướng tiếp cận này là cải thiện hiệu năng của mô hình tái định danh người f_θ được huấn luyện và đánh giá trên một tập dữ liệu đã gán nhãn với hàm mục tiêu:

$$score = \mathcal{F}_{eval}\left(\mathcal{F}_{SL}(f_\theta, \mathcal{D}), G\right) \tag{1.3}$$

1.1.4 Học thích ứng miền không giám sát

Phương pháp này tận dụng mô hình huấn luyện sẵn để thích nghi hiệu quả với dữ liệu mới, phản ánh nhu cầu thực tế trong triển khai hệ thống tái định danh.

Mô hình hóa: Mục tiêu của hướng tiếp cận này là sử dụng mô hình f_θ được huấn luyện trên một bộ dữ liệu có nhãn (miền nguồn) và học các đặc trưng mới trên một bộ dữ liệu khác không có nhãn (miền đích) để đánh giá kết quả trên tập dữ liệu kiểm tra ở miền đích.

$$score = \mathcal{F}_{eval}(\mathcal{F}_{USL}\left(\mathcal{F}_{SL}(f_\theta, \mathcal{D}_s), \mathcal{D}_t\right), G_t), \tag{1.4}$$

1.1.5 Học không giám sát

Học không giám sát cho phép huấn luyện tái định danh người mà không cần nhãn, giảm chi phí và tăng khả năng ứng dụng.

Mô hình hóa: Mục tiêu của hướng tiếp cận này là sử dụng mô hình f_θ được huấn luyện trên một bộ dữ liệu không có nhãn.

$$score = \mathcal{F}_{eval}\left(\mathcal{F}_{USL}(f_\theta, \mathcal{D}), G\right) \tag{1.5}$$

1.1.6 Dữ liệu đánh giá

Đánh giá trên các tập dữ liệu chuẩn như CUHK03, Market-1501, DukeMTMC-ReID và MSMT17. Trong luận án, các tên bộ dữ liệu được giản lược: "CUHK" chỉ CUHK03, "Market" cho Market-1501, "DukeMTMC" cho DukeMTMC-reID, và "MSMT" cho MSMT17. Với thiết lập học thích ứng miền không giám sát, ký hiệu "Market-to-Duke", "Duke-to-Market", "Market-to-MSMT" và "Duke-to-MSMT".

1.1.7 Độ đo đánh giá

Hiệu năng của mô hình thường được đánh giá dựa trên hai chỉ số chính: Đặc trưng tích lũy theo hạng (CMC - Rank-n) và bình quân độ chính xác trung bình (mAP).

1.2 Kiến thức nền tảng

1.2.1 Mạng cốt lõi trích xuất đặc trưng

Mạng cốt lõi CNN

Nhiệm vụ của mạng cốt lõi là ánh xạ dữ liệu thô thành một đặc trưng đại diện có ý nghĩa để dùng cho các tác vụ như phân loại, phát hiện, nhận diện. Một số mạng cốt lõi nổi tiếng hay được sử dụng trong các bài toán thị giác máy tính như Lenet-5[5], VGGNet[6], Resnet[7]...

Mạng cốt lõi ViT

Mạng CNN thường gặp khó khăn trong việc nắm bắt các đặc điểm có mối liên kết sâu và thông tin ngữ cảnh toàn cục, vốn rất quan trọng trong việc xử lý các thách thức về hình dáng, tư thế, độ sáng và độ che khuất. Nhằm khắc phục, Vision Transformer (ViT) đã được Dosovitskiy và các cộng sự[8] đề xuất tận dụng sức mạnh của cơ chế tự chú ý trong kiến trúc Transformer cho phép mô hình hóa quan hệ toàn cục giữa các vùng ảnh một cách trực tiếp, đồng thời bỏ qua hoàn toàn việc sử dụng các tầng tích chập.

1.2.2 Học độ đo sâu

DML giúp mô hình học biểu diễn đặc trưng có tính phân biệt cao bằng cách kéo gần đặc trưng các ảnh cùng danh tính và đẩy xa các ảnh khác danh tính trong không gian đặc trưng. Việc tối ưu này thường dựa trên các hàm mất mát nhằm tăng độ chính xác nhận dạng dưới nhiều điều kiện khác nhau. Các hàm mất mát được giới thiệu là phân

loại, trung tâm, bộ ba, bộ ba trọng tâm, InfoNCE, tương phản tự giám sát, tương phản có giám sát.

1.2.3 Kỹ thuật xếp hạng lại

Sử dụng K-reciprocal để xếp hạng lại truy vấn.

1.2.4 Thuật toán phân cụm

Sử dụng DBSCAN[9] hoặc k-mean[10].

1.3 Tổng quan nghiên cứu liên quan

1.3.1 Tái định danh người dựa trên học có giám sát

Tổng quan các nghiên cứu liên quan về tái định danh người theo hướng học có giám sát.

1.3.2 Tái định danh người dựa trên học thích ứng miền không giám sát

Tổng quan các nghiên cứu liên quan về tái định danh người theo hướng học thích ứng miền không giám sát.

1.3.3 Tái định danh người dựa trên học không giám sát

Tổng quan các nghiên cứu liên quan về tái định danh người theo hướng học không giám sát.

1.4 Các nghiên cứu trong nước

Nghiên cứu về tái định danh người tại Việt Nam đã có bước tiến từ đặc trưng thủ công đến các phương pháp học sâu, tiệm cận xu hướng quốc tế như học biến đổi, mô hình bất định, xếp hạng lại hay triển khai trên thiết bị biên. Những kết quả này khẳng định năng lực nghiên cứu trong nước và tạo nền tảng cho các hệ thống giám sát thông minh

1.5 Khoảng trống nghiên cứu

Các khoảng trống chính trong tái định danh người gồm:

1. **Học có giám sát:** Các hàm mất mát phổ biến (phân loại, trung tâm, bộ ba, tương phản) còn hạn chế về khả năng tổng quát, phân biệt liên lớp, chọn mẫu và duy trì tâm cụm; việc kết hợp chúng chưa được khai thác đầy đủ.

2. **Thích ứng miền không giám sát:** Khác biệt phân phối giữa miền nguồn–đích làm suy giảm hiệu năng; sử dụng đặc trưng toàn

cục bỏ qua thông tin cục bộ; nhãn giả thiếu tin cậy. Cần khai thác đặc trưng đa mức và cải thiện nhãn giả.

3. **Học không giám sát:** ResNet-50 ổn định nhưng kém mô hình hóa quan hệ không gian; ViT mạnh về tự chú ý nhưng chưa tận dụng đặc trưng cục bộ và thông tin camera ID để tăng phân biệt liên miền.

1.6 Hướng tiếp cận nghiên cứu

Với bối cảnh, tổng quan và khoảng trống nghiên cứu đã được xác định, luận án tập trung đề xuất các phương pháp nâng cao hiệu năng tái định danh người theo ba hướng tiếp cận chính: học có giám sát, học thích ứng miền không giám sát, và học không giám sát.

- Hướng tiếp cận 1: Nghiên cứu, phân tích ưu và nhược điểm của các hàm mất mát phổ biến được sử dụng trong bài toán tái định danh người, qua đó đề xuất chiến lược kết hợp các hàm mất mát một cách hiệu quả nhằm tận dụng toàn diện các đặc tính bổ sung lẫn nhau giữa chúng.

- Hướng tiếp cận 2: Nghiên cứu, phát triển phương pháp nâng cao hiệu năng của bài toán tái định danh người trong thiết lập học thích ứng miền không giám sát thông qua các hướng tiếp cận chính: (i) giảm sự sai lệch giữa phân phối đặc trưng của miền nguồn và miền đích, (ii) đánh giá chất lượng ảnh, (iii) tăng cường khả năng biểu diễn đặc trưng bằng cách kết hợp thông tin từ cả đặc trưng toàn cục và đặc trưng cục bộ, và (iv) tinh chỉnh nhãn giả nhằm nâng cao độ chính xác và độ tin cậy của nhãn huấn luyện trong miền mục tiêu.

- Hướng tiếp cận 3: Nghiên cứu, phân tích ưu và nhược điểm của kiến trúc CNN và ViT làm cơ sở cho đề xuất sử dụng ViT, đồng thời tích hợp thông tin camera vào quá trình huấn luyện giúp tăng hiệu năng tái định danh người.

1.7 Kết chương

Chương 1 đã giới thiệu cụ thể về bài toán tái định danh người, trình bày các kiến thức nền tảng cũng như đưa ra tổng quan các nghiên cứu liên quan. Trên cơ sở đó, luận án đã xác định được những khoảng trống nghiên cứu, từ đó đề xuất các hướng tiếp cận phù hợp nhằm nâng cao hiệu năng tái định danh người theo ba hướng chính là học có giám sát, học thích ứng miền không giám sát và học không giám sát.

Chương 2
NÂNG CAO HIỆU NĂNG TÁI ĐỊNH DANH NGƯỜI CÓ GIÁM SÁT DỰA TRÊN KẾT HỢP CÁC HÀM MẤT MÁT

Trong chương này, luận án trình bày phương pháp đề xuất "SCM-ReID" để nâng cao hiệu năng tái định danh người theo hướng tiếp cận có giám sát bằng cách sử dụng hàm mất mát tương phản có giám sát kết hợp với bốn hàm mất mát phổ biến gồm: phân loại, trung tâm, bộ ba và bộ ba trọng tâm. Kết quả nghiên cứu được công bố tại [CT2].

2.1 Đặt vấn đề

Trong bối cảnh của bài toán tái định danh người, nhiều hàm mất mát khác nhau đã được để xuất để tối ưu hiệu suất của mô hình, mỗi hàm cung cấp những ưu điểm riêng biệt, cũng như phải đối mặt với những hạn chế cụ thể[11]. Các phương pháp hiện tại thường dựa vào các hàm mất mát riêng lẻ hoặc kết hợp nhiều hàm để cải thiện hiệu suất tái định danh người. Xuất phát từ nhận định đó, chúng tôi đề xuất phương pháp kết hợp học tập tương phản có giám sát với học độ đo sâu, có tên là "SCM-ReID", nhằm nâng cao hiệu năng tái định danh người. Cụ thể, chúng tôi tích hợp SCL cùng với bốn hàm mất mát phổ biến: phân loại, bộ ba, trung tâm và bộ ba trọng tâm từ mô hình cơ sở được đề xuất bởi Wieczorek và cộng sự.

2.2 Phương pháp đề xuất

2.2.1 Hướng tiếp cận

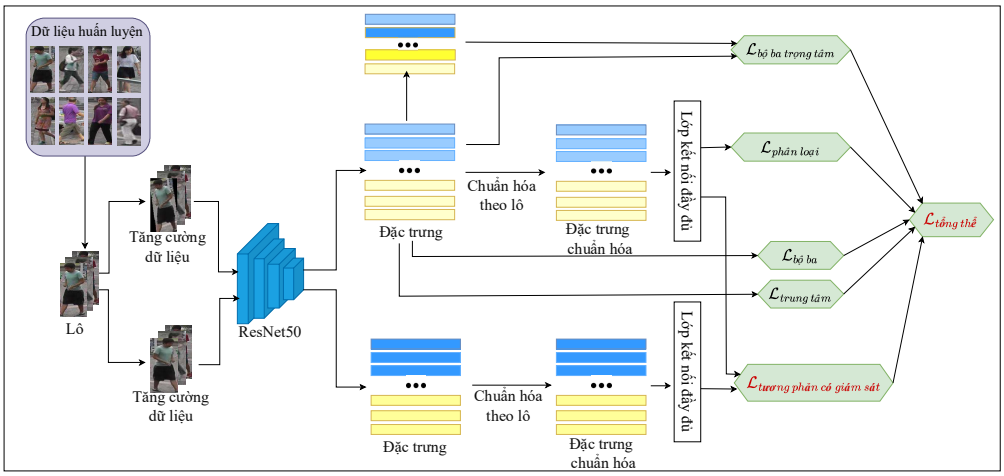
Phương pháp đề xuất tổng thể được minh họa trong Hình. 2.1.

2.2.2 Mô hình cơ sở

Dựa trên các hàm mất mát phân loại, trung tâm, bộ ba, bộ ba trọng tâm, hàm mất mát tổng quát của mô hình cơ sở được trình bày trong Công thức 2.1 như sau:

$$\mathcal{L}_{bl} = \mathcal{L}_{ce} + \lambda_{tri}\mathcal{L}_{tri} + \lambda_{ct}\mathcal{L}_{ct} + \lambda_{ctl}\mathcal{L}_{ctl}, \tag{2.1}$$

trong đó, λ_{tri}, λ_{ct}, λ_{ctl} là các siêu tham số điều chỉnh ảnh hưởng của từng hàm mất mát với hàm mất mát tổng.



Hình 2.1: Mô hình kiến trúc của phương pháp đề xuất - SCM-ReID.

2.2.3 Hàm mất mát tương phản có giám sát

Hàm mất mát tương phản có giám sát so sánh mỗi mẫu với toàn bộ mẫu cùng và khác danh tính trong lô, thay vì chỉ một cặp như bộ ba, nhờ đó xử lý tốt biến thiên nội lớp và nâng cao hiệu quả nhận dạng.

$$\mathcal{L}_{\text{sup}} = \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp\left(\text{sim}(\mathbf{z}_i, \mathbf{z}_p)/\tau\right)}{\sum_{a \in N(j)} \exp\left(\text{sim}\left(\mathbf{z}_i, \mathbf{z}_a\right)/\tau\right)}, \tag{2.2}$$

2.2.4 Hàm mất mát tổng thể

Hàm mục tiêu tối ưu của chúng tôi khi kết hợp với hàm mất mát tương phản có giám sát được biểu diễn trong Công thức 2.3 như sau:

$$\mathcal{L}_{overall} = \mathcal{L}_{bl} + \lambda_{sup}\mathcal{L}_{sup}, \tag{2.3}$$

2.3 Thực nghiệm nghiệm và đánh giá kết quả

Thực nghiệm toàn diện trên các bộ dữ liệu Market, CUHK03(L) và CUHK(D).

2.3.1 Chi tiết cài đặt

Thực nghiệm trên máy tính trang bị GPU NVIDIA A100 80GB VRAM.

2.3.2 Phân tích và đánh giá kết quả

Ảnh Hưởng của trọng số λ_{sup}: Giá λ là 0.2 cho kết quả tốt nhất.

Bảng 2.1: So sánh SCM-ReID với các phương pháp tiên tiến.

Phương pháp	Market				CUHK03(L)				CUHK03(D)			
	mAP	**R1**	**R5**	**R10**	**mAP**	**R1**	**R5**	**R10**	**mAP**	**R1**	**R5**	**R10**
BoT (CVPR 19)	94.2	95.4	-	-	-	-	-	-	-	-	-	-
CTL (ICONIP 21)	98.3	98	98.6	99.5	-	-	-	-	-	-	-	-
MCTL (JoBD 23)	98.63	98.4	99.1	**99.6**	-	-	-	-	-	-	-	-
CTupletL (3PGCIC 23)	98.57	98.3	-	-	94.38	92.3	-	-	93.39	91.10	-	-
PRONET++ (arXiv 23)	95.3	96.4	-	-	91.9	90.6	-	-	89.2	87.8	-	-
CA-Jaccard (CVPR 24)	94.5	96.2	-	-	-	-	-	-	-	-	-	-
SCM-ReID (Ours)	**98.84**	**98.66**	**99.14**	99.5	**96.92**	**95.57**	**98.57**	**99.14**	**96.53**	**95.14**	**98.35**	**99.00**

Ảnh hưởng của kích thước hình ảnh: Kích thước ảnh 256×128 đạt hiệu suất cao nhất.

Ảnh hưởng của kích thước lô: Kích thước lô 16x4 nổi bật là có hiệu suất ổn định cao nhất trên tất cả các bộ dữ liệu.

Hiệu suất tính toán: Đánh giá hiệu suất tính toán của SCM-ReID so với mô hình cơ sở trên bộ dữ liệu Market như một ví dụ đại diện.

Trực quan hóa: Trình bày kết quả trực quan trên bộ dữ liệu Market.

Trường hợp khó định danh: Chỉ ra các trường khó định danh.

So sánh với các phương pháp tiên tiến: So sánh SCM-ReID với các phương pháp tiên tiến để khẳng định hiệu quả của phương pháp đề xuất.

2.4 Kết luận và hướng phát triển

Chương này tập trung vào việc giải quyết thách thức cải thiện hiệu suất mô hình tái định danh người người có giám sát. Bằng cách tận dụng mạng cốt lõi Resnet-50 sử dụng học tương phản có giám sát và học độ đo sâu, chúng tôi đề xuất phương pháp SCM-ReID, kết hợp hàm mất mát tương phản có giám sát với bốn hàm mất mát khác: phân loại, trung tâm, bộ ba, bộ ba trọng tâm để đạt được kết quả vượt trội.

Chương 3

NÂNG CAO HIỆU NĂNG TÁI ĐỊNH DANH NGƯỜI THÍCH ỨNG MIỀN KHÔNG GIÁM SÁT DỰA TRÊN TĂNG CƯỜNG DỮ LIỆU, KẾT HỢP ĐẶC TRƯNG VÀ TINH CHỈNH NHÃN GIẢ

Chương này trình bày hai phương pháp học thích ứng miền không giám sát là IQAGA và DAPRH. Đầu tiên, IQAGA sử dụng GAN để chuyển phong cách ảnh từ miền nguồn sang miền đích, kết hợp đánh giá chất lượng ảnh nhằm giảm ảnh nhiễu. Thứ hai, DAPRH bổ sung DIM để thu hẹp chênh lệch đặc trưng giữa các miền, đồng thời kết hợp đặc trưng toàn thể (toàn cục và cục bộ) cùng kỹ thuật tinh chỉnh nhãn giả để nâng cao độ phân biệt của mô hình.

Kết quả nghiên cứu công bố tại [CT1], [CT4] và [CT5].

3.1 Đặt vấn đề

Các kết quả ở chương trước cho thấy mô hình tái định danh đạt hiệu năng cao khi huấn luyện và đánh giá trên cùng một bộ dữ liệu. Tuy nhiên, hiệu năng giảm đáng kể khi áp dụng sang tập dữ liệu khác do chênh lệch phân phối giữa các miền[12]. Sự khác biệt về góc nhìn, ánh sáng, nền cảnh và phong cách camera giữa các bộ dữ liệu ảnh hưởng mạnh đến khả năng tổng quát hóa. Ví dụ, mô hình huấn luyện trên Market đạt mAP 80.1% và Rank-1 92.8%, nhưng khi đánh giá trên Duke, chỉ còn 25.8% mAP và 43.7% Rank-1 (Bảng 3.1), cho thấy rõ tác động tiêu cực của sự khác biệt miền.

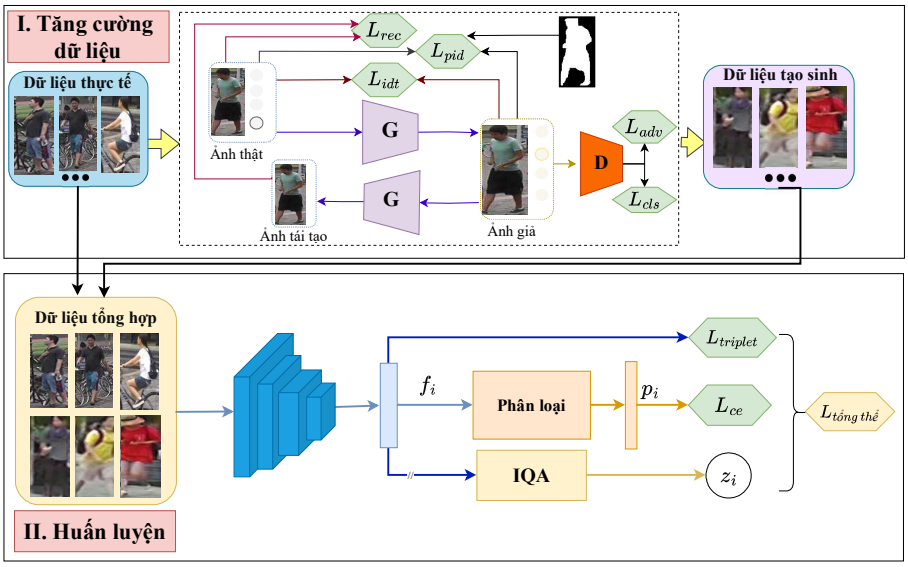
Để giảm thiểu sự suy giảm hiệu năng do lệch phân phối giữa các miền, chúng tôi đề xuất hai phương pháp theo hướng học thích ứng miền không giám sát.

Phương pháp đầu tiên, IQAGA, sử dụng StarGAN để sinh ảnh phong cách miền đích, kết hợp mô-đun đánh giá chất lượng ảnh dựa trên chuẩn hóa đặc trưng nhằm điều chỉnh trọng số huấn luyện, từ đó giảm tác động của ảnh nhiễu và nâng cao độ ổn định mô hình.

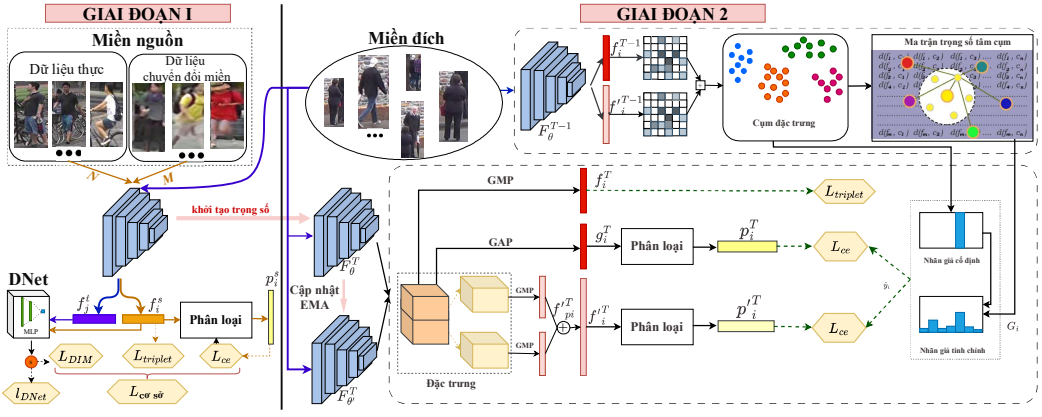
Phương pháp thứ hai, DAPRH, tiếp tục tăng cường dữ liệu bằng GAN nhưng chọn lọc ảnh sinh theo tỷ lệ tối ưu thay vì sử dụng toàn bộ. Đồng thời, phương pháp khai thác đặc trưng toàn thể và tinh chỉnh nhãn giả, giúp mô hình học tốt hơn từ dữ liệu không nhãn ở miền đích.

Bảng 3.1: Sự suy giảm hiệu năng khi đánh giá trên các tập dữ liệu khác nhau.

Tập huấn luyện	**Tập đánh giá**			
	Market		Duke	
	mAP	R1	mAP	R1
Market	80.1	92.8	25.8	43.7
Duke	26.2	55.3	70.2	84.1



Hình 3.1: Kiến trúc tổng thể của phương pháp đề xuất - IQAGA.



Hình 3.2: Kiến trúc tổng thể của phương pháp DAPRH.

DAPRH còn tích hợp ánh xạ bất biến miền (DIM) để thu hẹp chênh lệch đặc trưng giữa hai miền trong giai đoạn huấn luyện. Đồng thời, phương pháp khai thác dữ liệu không nhãn thông qua nhãn giả mềm theo khoảng cách cụm, kết hợp đặc trưng toàn–cục và cục–bộ, cùng kiến trúc Teacher-Student nhằm tăng độ ổn định. Hình 3.2 minh họa tổng quan phương pháp.

Phương pháp DAPRH được xây dựng dựa trên ba thành phần chính: (i) sử dụng GAN (StarGAN) để sinh ảnh theo phong cách miền đích và DIM để thu hẹp khoảng cách đặc trưng giữa hai miền; (ii) kết hợp đặc trưng toàn cục và cục bộ để tạo biểu diễn tổng thể giàu thông tin; (iii) áp dụng nhãn giả mềm thay vì nhãn cứng và sử dụng kiến trúc Teacher-Student nhằm cải thiện hiệu quả học không giám sát.

3.2 Phương pháp nâng cao học có giám sát trên miền nguồn

Đây là hướng tiếp cận đầu tiên nhằm cải thiện hiệu năng của mô hình trong giai đoạn học có giám sát trên miền nguồn..

3.2.1 Thích ứng miền mức hình ảnh

Sử dụng StarGAN[13] để chuyển đổi phong cách miền nguồn sang miền đích mà vẫn giữ nguyên đặc trưng danh tính.

3.2.2 Thích ứng miền mức đặc trưng

Sử dụng Domain-Invariant Mapping (DIM)[14] để giảm thiểu sự khác biệt phân phối đặc trưng giữa miền nguồn và miền đích.

3.2.3 Đánh giá chất lượng ảnh và điều chỉnh trọng số huấn luyện

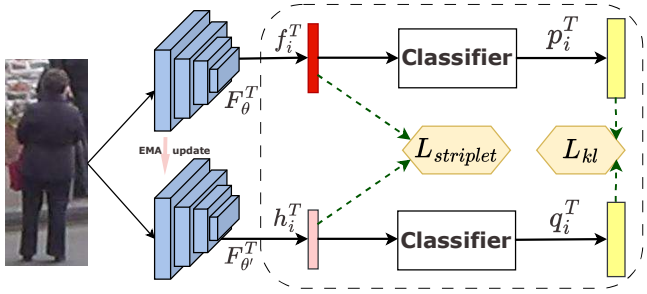
Sử dụng chuẩn hóa đặc trưng để ước lượng giá trị chất lượng z_i của ảnh x_i để đánh giá[15].

$$z_i = \left\lfloor \frac{\|f_i\| - \mu_f}{\sigma_f / h} \right\rceil_{-1}^{1} \in \quad [-1, 1], \tag{3.1}$$

Dựa trên chỉ số chất lượng này, hàm mất mát tổng thể được biểu diễn như sau:

$$\mathcal{L}(x_i, y_i) = (z_i * \lambda + 1) * (\mathcal{L}_{ce}(x_i, y_i) + \mathcal{L}_{tri}(x_i)) \tag{3.2}$$

ở đây, λ là tham số điều khiển ảnh hưởng của giá trị z_i.



Hình 3.3: Tổng quát về kiến trúc Mean-Teacher.

3.3 Phương pháp nâng cao học thích ứng miền không giám sát trên miền đích

3.3.1 Đặc trưng tổng thể dựa trên đặc trưng toàn cục và cục bộ

Nhánh đặc trưng toàn cục: Để lấy đặc trưng toàn cục của một hình ảnh x_i tại epoch T, trước tiên, chúng tôi cần trích xuất bản đồ đặc trưng tương ứng $F_\theta^T(x_i) \in \mathcal{R}^{C\times H\times W}$, trong đó C, H và W lần lượt là thông tin của kênh, chiều cao và chiều rộng của hình ảnh.

Nhánh đặc trưng phần cục bộ: Bên cạnh thông tin toàn cục, chúng tôi cũng xem xét các lợi ích đến từ thông tin cục bộ được trích xuất từ các vùng của hình ảnh. Do đó, giống như[16, 17, 18], chúng tôi có thể thu được các đặc trưng cục bộ bằng cách chia ngang bản đồ đặc trưng thành K vùng phân chia đều nhau và sau đó áp dụng lớp GMP.

3.3.2 Tinh chỉnh nhãn giả dựa trên tâm cụm

Tỉnh chỉnh nhãn giả cứng (one-hot) bằng nhãn giả mềm (one-soft) dựa vào khoảng cách đến tâm cụm và lọc bỏ các đặc trưng không đủ tốt để cải thiện đặc trưng tâm cụm dựa vào độ đo Silhouette[19].

3.3.3 Huấn luyện dựa trên kiến trúc Mean-Teacher

Đề xuất sử dụng kiến trúc Mean-Teacher để xử lý vấn đề sự biến động và không thể đoán trước trong giai đoạn học không giám sát.

3.3.4 Hàm mất mát tổng thể

Cuối cùng, để tối ưu hóa mô hình tái định danh trong giai đoạn II, chúng tôi sử dụng hàm mất mát nhận dạng, hàm mất mát bộ ba để học thông tin mới từ bộ dữ liệu, và hai hàm mất mát trong Mục 3.3.3 để học từ mô hình Teacher. Với nhãn giả đã tinh chỉnh, công thức của hàm mất mát mục tiêu có thể được viết như sau:

Bảng 3.2: Hiệu năng cơ sở mô hình "IQAGA" (%).

Phương pháp	Duke → Market				Market → Duke			
	mAP	R1	R5	R10	mAP	R1	R5	R10
Baseline	26.2	55.3	67.8	75.2	25.8	43.7	57.9	66.2
+GAN	35.1	68.6	83.1	87.8	31.5	54.2	67.6	72.5
+GAN+ASSESSMENT	36.3	70.2	84.4	88.4	32.1	55.5	68.2	73.2

$$\begin{aligned}\mathcal{L}_{\text{USL}} = (1-w_1)\mathcal{L}_{id}(p_i, p'_i, \hat{\mathbf{y}}_i) + w_1\mathcal{L}_{kl}(p_i, q_i) \\ + (1-w_2)\mathcal{L}_{tri}(f_i, y_i) + w_2\mathcal{L}_{stri}(f_i, h_i)\end{aligned} \tag{3.3}$$

Trong hàm mất mát mục tiêu, hai trọng số w_1, w_2, được điều chỉnh là 0.4 và 0.8, được sử dụng để quản lý việc học từ Teacher và học thông tin mới. Sau đó, p_i, p'_i, và q_i đại diện cho đầu ra lớp phân loại từ các nhánh toàn cục, cục bộ và toàn cục của mô hình giáo viên, tương ứng. Cuối cùng, f_i và h_i là các đặc trưng toàn cục được trích xuất từ mô hình Student và Teacher, tương ứng.

3.4 Thực nghiệm và đánh giá kết quả

Đặt ra 9 RQ để thực nghiệm giải quyết.

3.4.1 Chi tiết triển khai

Hạ tầng tính toán: CPU Xeon (Platinum 8160), 256GB RAM và một GPU NVIDIA Tesla T4 với 16GB bộ nhớ.

Bộ dữ liệu đánh giá: "Market-to-Duke", "Duke-to-Market", "Market-to-MSMT", "Duke-to-MSMT".

Tham số thích ứng miền ở mức hình ảnh: Tham số kiến trúc StarGAN[13].

Tham số huấn luyện có giám sát trên miền nguồn: Tham số huấn luyện mô hình có giám sát trên miền nguồn.

Tham số huấn luyện không giám sát trên miền đích: Tham số huấn luyện không giám sát trên miền đích.

3.4.2 Kết quả học có giám sát trên miền nguồn

Thích ứng miền mức hình ảnh: Hiệu năng mAP và Rank-1 tăng lên khoảng 10% tùy từng trường hợp.

Thích ứng miền mức đặc trưng: Hiệu năng mAP và Rank-1 tăng lên khoảng 5% tùy từng trường hợp.

Đánh giá chất lượng để điều chỉnh trọng số huấn luyện: λ = 0.8

Bảng 3.3: Hiệu năng cơ sở của mô hình "DAPRH" (%).

Phương pháp	Duke → Market				Market → Duke				Market → MSMT				Duke → MSMT			
	mAP	R1	R5	R10	mAP	R1	R5	R10	mAP	R1	R5	R10	mAP	R1	R5	R10
Base.	26.2	55.3	67.8	75.2	25.8	43.7	57.9	66.2	4.7	14.4	22.1	26.5	6.3	17.8	29.3	36.2
+GAN	36.2	66.4	81.1	86.1	34.4	57.6	70.3	76.0	9.6	28.8	40.3	52.9	12.0	35.8	48.1	53.4
+DIM	30.1	60.4	75.1	80.8	28.8	51.1	66.0	71.2	8.6	25.3	35.9	41.6	10.7	31.2	43.0	48.8
Our base.	36.4	68.9	83.2	87.9	35.4	58.2	72.4	77.5	12.4	36.0	49.2	54.4	16.0	43.8	56.8	62.6

Bảng 3.4: So sánh với các phương pháp tiên tiến.

Phương pháp	Nơi công bố	Duke → Market				Market → Duke			
		mAP	R1	R5	R10	mAP	R1	R5	R10
PTGAN[4]	CVPR18	-	38.6	-	66.1	-	27.4	-	50.7
SPGAN+LMP[20]	CVPR18	26.9	58.1	76.0	82.7	26.4	46.9	62.6	68.5
PGPPM[21]	IEEE.TMM20	33.9	63.9	81.1	86.4	17.9	36.3	54.0	61.6
CDCSA[18]	ACM.MM22	34.0	66.8	82.1	87.9	31.4	53.8	67.8	72.2
IQAGA		**36.3**	**70.2**	**84.4**	**88.4**	**32.1**	**55.5**	**68.2**	**73.2**

Kết hợp thích ứng miền mức hình ảnh và đánh giá chất lượng để điều chỉnh trọng số huấn luyện: Nâng cao hiệu năng khoảng 2%.

Kết hợp thích ứng miền ở mức hình ảnh và thích ứng miền ở mức đặc trưng:

Kết quả nâng cao hiệu năng

3.4.3 Kết quả thích ứng không giám sát trên miền đích

Đặc trưng tổng thể với GMP: Cải thiện 1–2% với mAP và 0.5–1% Rank-1 so với mô hình chỉ dùng đặc trưng toàn cục.

Tinh chỉnh nhãn giả: Cải thiện mAP và Rank-1 từ 0.3–2%.

Phân tích giá trị của các siêu tham số:

Phân tích tỷ lệ N:M. Tỷ lệ $N : M = 4 : 1$ là tối ưu.

Bảng 3.5: Ảnh hưởng của lớp pooling và học theo đặc trưng cục bộ (%).

Phương pháp	Duke → Market				Market → Duke				Market → MSMT				Duke → MSMT			
	mAP	R1	R5	R10	mAP	R1	R5	R10	mAP	R1	R5	R10	mAP	R1	R5	R10
Base.	81.5	92.0	96.9	98.0	58.8	74.5	82.7	85.5	29.8	55.6	66.5	73.9	32.6	60.5	73.2	79.0
+GMP	84.5	93.0	97.2	98.0	69.0	81.6	90.4	92.5	32.7	60.2	72.5	77.0	34.1	62.8	74.8	79.3
+GMP+PL	85.6	93.9	98.0	98.7	70.3	82.8	91.7	94.2	34.2	63.2	75.1	79.5	35.6	65.6	77.2	81.2

Phân tích hệ số α. Đối với trường hợp Market-to-Duke và Market-to-MSMT, mô hình đạt hiệu năng cao nhất khi $\alpha = 0.5$, trong khi với Duke-to-Market, giá trị tối ưu là $\alpha = 0.7$.

Ảnh hưởng của các trọng số điều khiển. Giá trị của w_1 và w_2 trong Phương trình 3.3 được đặt là 0.4 và 0.8.

So sánh với các phương pháp tiên tiến

Chúng tôi so sánh DAPRH với các phương pháp UDA gần đây (Bảng 3.6) để trả lời RQ9. Kết quả cho thấy DAPRH đạt hiệu năng vượt trội: 85.9% mAP, 94.4% Rank-1 (Duke-Market); 72.0% mAP, 83.7% Rank-1 (Market-Duke); và kết quả ấn tượng trên Market-MSMT (35.8% / 64.8%) và Duke-MSMT (36.0% / 65.5%).

3.5 Kết chương

Nghiên cứu này tập trung vào bốn thách thức chính trong tái định danh người thích ứng miền không giám sát: (i) chênh lệch phân phối giữa các miền, (ii) ảnh chất lượng thấp từ GAN, (iii) đặc trưng toàn cục chưa đủ thông tin, và (iv) nhãn giả thiếu chính xác. Để giải quyết, chúng tôi đề xuất hai phương pháp: IQAGA, sử dụng StarGAN kết hợp đánh giá chất lượng ảnh; và DAPRH, mở rộng thêm DIM để thu hẹp khoảng cách đặc trưng, kết hợp đặc trưng toàn cục–cục bộ, cùng cơ chế nhãn giả mềm trong kiến trúc Teacher-Student.

Các thực nghiệm toàn diện trên các trường hợp Market-to-Duke, Duke-to-Market, Market-to-MSMT, Duke-to-MSMT đã chứng minh cho hiệu quả của hai phương pháp đề xuất IQAGA và DAPRH trong tác bài toán định danh người thích ứng miền không giám sát. Tuy nhiên, GAN có thể sinh ảnh kém chất lượng, và độ tin cậy của nhãn giả vẫn phụ thuộc vào phân cụm. Trong tương lai, chúng tôi hướng đến cải tiến cơ chế nhãn giả và tối ưu quy trình huấn luyện để tăng tính ổn định và khả năng thích ứng mô hình.

Bảng 3.6: So sánh các phương pháp tiên tiến trong tái định danh người thích ứng miền không giám sát (%).

Phương án	Nơi công bố	Duke → Market				Market → Duke				Market → MSMT				Duke → MSMT			
		mAP	R1	R5	R10	mAP	R1	R5	R10	mAP	R1	R5	R10	mAP	R1	R5	R10
DIMGLO[14]	ACM.MM20	65.1	88.3	94.7	96.3	58.3	76.2	85.7	88.5	20.7	49.7	–	66.1	24.4	56.5	–	70.0
MMT[22]	ICLR20	71.2	87.7	94.9	96.9	65.1	78.0	88.8	92.5	22.9	49.2	63.1	68.8	23.3	50.1	63.9	69.8
SpCL[23]	NIPS20	76.7	90.3	96.2	97.7	68.8	82.9	90.1	92.5	26.8	53.7	65.0	69.8	26.5	53.1	65.8	70.5
UNRN[24]	AAAI21	78.1	91.9	96.1	97.8	69.1	82.0	90.7	93.5	25.3	52.4	64.7	69.7	26.2	54.9	67.3	70.6
CCTSE[25]	IEEE.TIFS21	78.4	92.9	96.9	97.8	72.6	**85.0**	92.1	93.9	33.2	62.3	74.1	78.5	34.5	63.9	75.3	79.6
GLT[26]	CVPR21	79.5	92.2	96.5	97.8	69.2	82.0	90.2	92.8	26.5	56.6	67.5	72.0	27.7	59.5	70.1	74.2
AWB[27]	IEEE.TIP22	81.0	93.5	97.4	98.3	70.9	83.8	**92.3**	94.0	29.0	57.3	70.7	75.9	29.0	57.9	71.5	76.6
CMFC[18]	ACM.MM22	81.0	94.0	97.1	98.3	71.2	83.2	91.6	94.0	–	–	–	–	–	–	–	–
CDCL[28]	KBS23	81.5	92.8	97.6	98.7	70.2	82.7	91.3	93.9	29.5	59.7	71.4	75.9	30.3	61.4	74.8	77.1
SECRET[29]	AAAI22	83.0	93.3	–	–	69.2	82.0	–	–	31.7	60.0	–	–	–	–	–	–
RESL[30]	AAAI22	83.1	93.2	96.8	98.0	72.3	83.9	91.7	93.6	33.6	**64.8**	74.6	79.6	34.2	65.2	74.6	80.1
LF_2[17]	ICPR22	83.2	92.8	**97.8**	98.4	**73.5**	83.7	91.9	94.3	–	–	–	–	–	–	–	–
MCRN[31]	AAAI22	83.8	93.8	97.5	98.5	71.5	84.5	91.7	93.8	32.8	64.4	75.1	79.2	35.7	**67.5**	**77.9**	**81.6**
DAPRH		**85.9**	**94.4**	**97.8**	**98.8**	72.0	83.7	92.1	**94.5**	**35.8**	**64.8**	**77.0**	**81.1**	**36.0**	65.5	77.0	80.9

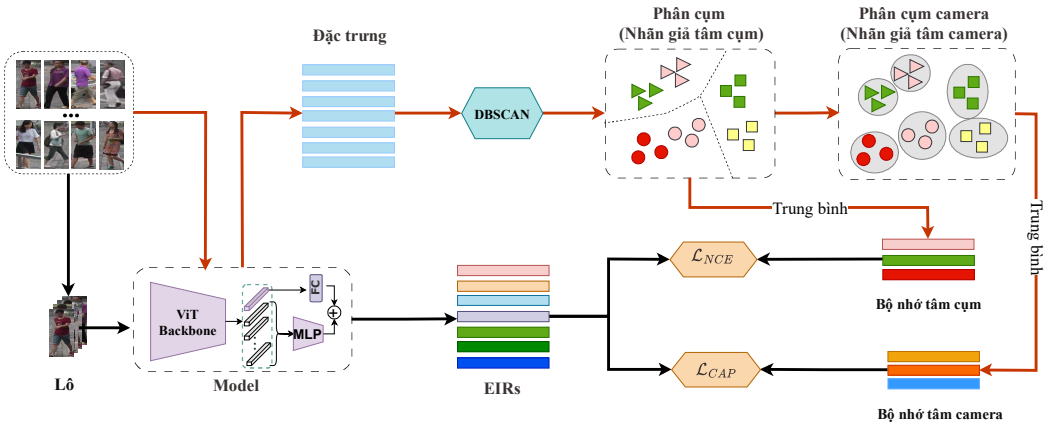
Chương 4
NÂNG CAO HIỆU NĂNG TÁI ĐỊNH DANH NGƯỜI KHÔNG GIÁM SÁT DỰA TRÊN KIẾN TRÚC ViT VÀ THÔNG TIN CAMERA

Trong chương này, luận án trình bày phương pháp "ViTC-UReID" nhằm nâng cao hiệu năng tái định danh người trong hướng tiếp cận học không giám sát sử dụng mạng cốt lõi ViT và thông tin camera. Các kết quả nghiên cứu trong chương này được công bố tại [CT3].

4.1 Đặt vấn đề

Các phương pháp học thích ứng miền (UDA) tuy hiệu quả nhưng phụ thuộc vào dữ liệu có nhãn và giả định phân phối tương đồng giữa các miền, gây hạn chế trong thực tế. Ngược lại, học không giám sát hoàn toàn (USL) linh hoạt hơn nhưng vẫn đối mặt với nhãn giả kém chính xác và biểu diễn đặc trưng chưa đầy đủ.

Để khắc phục, chúng tôi đề xuất ViTC-UReID kết hợp ViT và thông tin camera để khai thác hiệu quả cả đặc trưng toàn cục, cục bộ và giảm sai lệch liên miền. Phương pháp đề xuất đạt kết quả vượt trội trên Market, MSMT và CUHK, cho thấy tính hiệu quả và khả năng ứng dụng cao trong hệ thống giám sát thông minh.



Hình 4.1: Kiến trúc phương pháp đề xuất ViTC-UReID.

4.2 Phương pháp đề xuất

Phương pháp ViTC-UReID được phát triển dựa trên hai ý tưởng chính: i) tăng cường biểu diễn ảnh trong không gian đặc trưng dựa trên

mạng cốt lõi ViT; ii) tích hợp thông tin định danh camera. Hình 4.1 minh họa trực quan các thành phần chính của phương pháp đề xuất.

4.2.1 Mô hình cơ sở

Sử dụng DBSCAN[9] để phân cụm và gán nhã giả:

$$\mathcal{C} = DBSCAN(\{\Omega(x_i) \mid x_i \in \mathcal{D}\}), \tag{4.1}$$

Huấn luyện với ClusterNCE[32].

$$\mathcal{L}_{NCE} = -\log \frac{\exp(q \cdot c_+/\tau)}{\sum_{k=1}^{K} \exp(q \cdot c_k/\tau)}, \tag{4.2}$$

4.2.2 Tăng cường biểu diễn đặc trưng hình ảnh

Ban đầu, kiến trúc cơ sở được đề cập được minh họa trong Hình 4.1, tuân theo mạng cốt lõi ViT để trích xuất đặc trưng cho ảnh. Như đã trình bày trong phần đặt vấn đề, bộ mã hóa ảnh được khởi tạo với mô hình đã huấn luyện trước LUPersonVIT[33]. Cụ thể, với một ảnh đầu vào I có kích thước $R^{(H\times W\times C)}$, ảnh sẽ được chia thành các mảnh không chồng lắp nhau với số lượng $M = H \times W/S^2$, trong đó S là kích thước của mỗi mảnh. Các mảnh này sau đó thêm thông tin nhúng tuyến tính và trở thành các tham số có thể học được. Một token $[CLS]$ được thêm vào đầu chuỗi để đại diện cho cấp độ ảnh. Tiếp theo, tập các mảnh được ký hiệu là $P = p_{cls}, p_1, p_2, ..., p_{M-1}, p_M$ được đưa vào các khối transformer của bộ mã hóa ảnh. Quá trình này tạo ra một chuỗi biểu diễn ảnh có D chiều.

4.2.3 Kết hợp thông tin Camera

Thách thức cốt lõi trong bài toán tái định danh người là làm sao để nhận dạng chính xác cùng một người qua các camera khác nhau. Lấy cảm hứng từ nghiên cứu của Wang và cộng sự [34], trong phần này, chúng tôi tập trung vào học các thông tin định danh camera liên miền, nhằm mục tiêu học các đặc trưng của người giữa các camera khác nhau.

Hàm mất mát tương phản giữa các camera:

$$\mathcal{L}_{CAP} = -\frac{1}{|\mathcal{P}(i)|} \sum_{j \in \mathcal{P}(i)} \log \frac{\exp(q \cdot c_a^j/\tau_c)}{\exp(q \cdot c_a^j/\tau_c) + \sum_{k \in \mathcal{N}(i)} \exp(q \cdot c_k/\tau_c)}, \tag{4.3}$$

trong đó $\mathcal{P}$ và $\mathcal{N}$ lần lượt là tập cùng định danh và khác định danh thông tin camera khó xác định, và τ_c là siêu tham số điều chỉnh.

Bảng 4.1: Nghiên cứu lược bỏ trên bộ dữ liệu Market and MSMT.

Phương pháp	Market				MSMT			
	mAP	R1	R5	R10	mAP	R1	R5	R10
Pretrained	11.3	34.2	49.7	57.3	4.7	20.5	29.5	34.1
Baseline	90.7	96.4	98.8	99.4	53.3	78.9	88.0	90.8
Baseline + *EIR*	92.5	96.5	98.8	99.5	55.7	80.6	88.7	91.3
Baseline + *CAP*	92.0	96.6	98.9	99.5	62.9	85.4	92.1	94.0
Baseline + *EIR* + *CAP*	92.8	97.1	99.1	99.5	63.6	85.8	92.3	94.1

4.2.4 Hàm mục tiêu huấn luyện

Để tối ưu hóa mô hình, chúng tôi sử dụng hai hàm mất mát ClusterNCE và CAP trong quá trình huấn luyện. Hàm mục tiêu cuối cùng trong quá trình huấn luyện được viết như sau:

$$\mathcal{L}_{OBJ} = \mathcal{L}_{NCE} + \lambda * \mathcal{L}_{CAP} \tag{4.4}$$

Trong đó, λ là hệ số dùng để điều chỉnh mức độ ảnh hưởng của thành phần CAP trong quá trình huấn luyện.

4.3 Thực nghiệm và đánh giá

Để chứng minh hiệu quả của ViTC-UReID, chúng tôi thực hiện một loạt thực nghiệm toàn diện nhằm phân tích tính hiệu quả của đề xuất.

4.3.1 Tham số thực nghiệm

Sử dụng ViT-B/16 làm bộ mã hóa ảnh và được khởi tạo bằng trọng số từ mô hình LUPerson đã huấn luyện trước[33].

4.3.2 Thực nghiệm lược bỏ

Phân tích mô hình cơ sở: Việc khởi tạo mạng cốt lõi ViT từ các tham số đã được huấn luyện từ tập dữ liệu không gán nhãn quy mô lớn LUPerson[33] đã được chứng minh là một hướng tiếp cận hiệu quả trong các bài toán tái định danh người.
Kết quả các cách kết hợp thành phần: Như thể hiện trong Bảng 4.1, việc tích hợp EIR giúp cải thiện hiệu suất mô hình lên đến 2% ở cả chỉ số R1 và mAP. Kết quả cho thấy EIR đạt độ ổn định khi top-K vượt quá 0.2. Tương tự, chúng tôi tinh chỉnh các mô hình có bổ sung hàm mất mát CAP, và trình bày kết quả ở hàng thứ tư của Bảng 4.1.

4.3.3 Hiệu suất tính toán

Mức sử dụng bộ nhớ GPU gần như tương đương. ViTC-UReID cần 180 giây cho mỗi epoch tăng 28.5% so với 140 giây của mô hình cơ sở.

4.3.4 Phân tích trực quan

Cụ thể, 10 ảnh đầu tiên của quá trình truy xuất đã được thể hiện, bao gồm cả các trường hợp thành công và thất bại, được sử dụng để so sánh giữa mô hình cơ sở và ViTC-UReID.

4.3.5 So sánh với các phương pháp tiên tiến

Kết quả thí nghiệm cuối cùng được trình bày trong Bảng 4.2, cho thấy phương pháp đề xuất đạt hiệu suất vượt trội trên nhiều chỉ số đánh giá khi so sánh với các phương pháp nổi tiếng hiện nay.

Bảng 4.2: So sánh với các phương pháp tiên tiến. † phương pháp sử dụng thông tin camera.

Mạng cốt lõi	Phương pháp	Nơi công bố	Market				MSMT				CUHK			
			mAP	R1	R5	R10	mAP	R1	R5	R10	mAP	R1	R5	R10
Phương pháp có giám sát - Fully supervised methods														
CNN	ABNET+	CVPR22	93.0	95.7	–	–	62.2	80.8	–	–	79.1	80.6	–	–
	NFormer[35]													
	ProNet++[36]	CoRR23	90.2	96.0	–	–	65.5	85.4	–	–	82.7	85.2	–	–
VIT	TransReID[37]	ICCV21	89.5	95.2	–	–	69.4	86.2	–	–	–	–	–	–
	CLIP-ReID	AAAI23	89.6	95.5	–	–	73.4	88.7	–	–	–	–	–	–
	TMGF†[38]	WACV23	91.9	96.3	98.9	99.3	70.3	88.2	94.1	95.4	–	–	–	–
Phương pháp không giám sát - Fully unsupervised methods														
CNN	CC[32]	ACCV22	83.0	92.9	97.2	98.0	33.0	62.0	71.8	76.7	–	–	–	–
	PPLR†[16]	CVPR22	84.4	94.3	97.8	98.6	42.2	73.3	83.5	86.5	–	–	–	–
	ISE[39]	CVPR22	85.3	94.3	98.0	98.8	37.0	67.6	77.5	81.0	–	–	–	–
VIT	PASS[40]	ECCV22	88.5	94.9	–	–	41.0	67.0	–	–	–	–	–	–
	TMGF†[38]	WACV23	89.5	95.5	98.0	98.7	58.2	83.3	90.2	92.1	–	–	–	–
	PCL-CLIP†[41]	Arxiv23	88.4	94.8	98.0	98.7	**65.5**	84.9	92.0	94.0	–	–	–	–
	ACFL-VIT[42]	PR24	89.1	95.1	–	–	45.7	70.1	–	–	–	–	–	–
	TCMM[43]	Arxiv25	90.5	96.0	–	–	52.0	78.4	–	–	–	–	–	–
VIT	ViTC-UReID	Our	**92.8**	**97.1**	**99.1**	**99.3**	63.6	**85.8**	**92.3**	**94.1**	**89.8**	**91.1**	**95.1**	**97.1**

4.4 Kết luận và hướng phát triển

Trong chương này, luận án trình bày phương pháp ViTC-UReID sử dụng mạng cốt lõi ViT kết hợp thông tin camera nhằm đồng thời thu nhận đặc trưng toàn cục và cục bộ, tăng tính nhất quán liên miền. Phương pháp cho kết quả vượt trội trên các bộ dữ liệu chuẩn như Market, MSMT và CUHK. Tuy nhiên, vẫn còn tồn tại một số hạn chế như độ tin cậy của nhãn giả trong môi trường nhiễu và yêu cầu tính toán cao của ViT. Trong tương lai, luận án hướng đến cải thiện nhãn giả, nghiên cứu kiến trúc Transformer nhẹ và mở rộng học đặc trưng theo thông tin camera để tăng khả năng thích ứng trong các điều kiện thực tế phức tạp.

KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Đóng góp chính

Trong bối cảnh đô thị thông minh với hệ thống camera giám sát ngày càng phổ biến, tái định danh người giữ vai trò quan trọng trong hệ thống giám sát thông minh. Khác với nhận dạng khuôn mặt, tái định danh dựa vào đặc trưng toàn thân (trang phục, dáng đi, phụ kiện) nhưng vẫn đối mặt với nhiều thách thức về sự biến đổi ngoại hình, hạn chế về dữ liệu, khả năng biểu diễn đặc trưng, chất lượng nhãn giả và khả năng mở rộng.. Để khắc phục, luận án đóng góp ba hướng chính.

- Đóng góp 1: Theo hướng tiếp cận học có giám sát, đề xuất phương pháp SCM-ReID sử dụng hàm mất mát tương phản có giám sát kết hợp với bốn hàm mất mát phổ biến (phân loại, bộ ba, trung tâm, bộ ba trọng tâm) giúp mô hình học được đặc trưng có tính phân biệt và khả năng khái quát tốt hơn.

- Đóng góp 2: Theo hướng tiếp cận học thích ứng miền không giám sát, đề xuất phương pháp IQAGA và DAPRH bằng cách kết hợp các chiến lược: thu hẹp khoảng cách phân phối giữa hai miền thông qua GAN và ánh xạ bất biến miền (DIM), đánh giá chất lượng ảnh để điều chỉnh trọng số huấn luyện, tích hợp đặc trưng cục bộ và toàn cục để tăng tính biểu diễn, và tinh chỉnh nhãn giả mềm dựa trên khoảng cách tâm cụm nhằm nâng cao độ ổn định và độ chính xác..

- Đóng góp 3: Theo hướng tiếp cận học không giám sát đề xuất phương pháp ViTC-UReID tận dụng kiến trúc ViT thay thế CNN để tăng cường biểu diễn đặc trưng toàn cục, đồng thời tích hợp thông tin camera nhằm học đặc trưng phù hợp với từng góc nhìn, từ đó cải thiện hiệu năng tái định danh người không giám sát.

Từ những kết quả đạt được qua ba đóng góp, có thể khẳng định rằng luận án đã hoàn thành mục tiêu đề ra ban đầu là đề xuất phương pháp mới để nâng cao hiệu năng tái định danh người theo cả ba hướng tiếp cận học có giám sát, thích ứng miền không giám sát và không giám sát.

Bàn luận

Kết quả thực nghiệm cho thấy ba hướng tiếp cận chính trong luận án đều mang lại hiệu quả nhất định. Với học có giám sát, phương pháp SCM-ReID đạt mAP 98.84% và Rank-1 98.66% trên Market, gần 97%

trên CUHK03, chứng minh hiệu năng cao nhưng phụ thuộc hoàn toàn vào dữ liệu gán nhãn nên khó triển khai rộng rãi. Với thích ứng miền không giám sát, phương pháp IQAGA còn hạn chế (mAP 36.3%, Rank-1 70.2% Duke-to-Market), trong khi DAPRH cải thiện đáng kể (mAP 85.9%, Rank-1 94.4%), cho thấy lợi ích của việc khai thác dữ liệu không nhãn tại miền đích, song vẫn bị giới hạn khi khoảng cách miền lớn. Cuối cùng, với học không giám sát, ViTC-UReID đạt mAP 92.8%/Rank-1 97.1% trên Market, 63.6%/85.8% trên MSMT và 89.8%/91.1% trên CUHK, vượt trội hơn các mô hình CNN không giám sát và tiệm cận học có giám sát.

Trong triển khai thực tế, học có giám sát phù hợp môi trường kiểm soát tốt nhưng tốn chi phí gán nhãn; thích ứng miền giảm chi phí nhưng phức tạp và kém ổn định; học không giám sát hứa hẹn cho hệ thống quy mô lớn song còn hạn chế về độ chính xác và tính ổn định. Như vậy, mỗi hướng tiếp cận đều có ưu – nhược điểm riêng, và việc lựa chọn phụ thuộc vào yêu cầu của từng hệ thống giám sát thông minh.

Hạn chế

Mặc dù các phương pháp đề xuất đã cải thiện hiệu năng ReID, vẫn còn một số hạn chế. Với học có giám sát, mô hình nhạy với kích thước lô, tốn bộ nhớ, chưa xử lý tốt che khuất, thay đổi góc nhìn lớn hay ngoại hình tương đồng, và phụ thuộc vào ResNet-50 nên khó mở rộng. Với thích ứng miền không giám sát, GAN có thể tạo ảnh kém chất lượng và nhãn giả dễ sai lệch, làm giảm hiệu quả. Với học không giám sát, việc dùng ViT đòi hỏi tài nguyên tính toán lớn và phụ thuộc vào nhãn giả, dễ bị nhiễu khi dữ liệu chất lượng thấp hoặc cá thể quá giống nhau.

Hướng phát triển

Dựa trên các hạn chế đã phân tích, một số hướng nghiên cứu tiềm năng gồm: (i) áp dụng hoặc kết hợp linh hoạt nhiều hàm mất mát để tối ưu không gian đặc trưng; (ii) cải thiện chất lượng ảnh sinh bằng GAN hoặc thay thế bằng các kỹ thuật mới như mô hình khuếch tán; (iii) mở rộng đánh giá trên nhiều kiến trúc mạng và bộ dữ liệu đa dạng để kiểm chứng khả năng tổng quát; (iv) phát triển chiến lược tinh chỉnh nhãn giả hiệu quả hơn cho huấn luyện không giám sát.

DANH MỤC CÔNG TRÌNH KHOA HỌC

Tạp chí

[CT1] **Dang H. Pham**, Anh D. Nguyen and Hoa N. Nguyen, "GAN-based Data Augmentation and Pseudo-Label Refinement with Holistic Features for Unsupervised Domain Adaptation Person Re-Identification", Journal of Knowledge-Based Systems, Elsevier.
https://doi.org/10.1016/j.knosys.2024.111471. (SCI-E, Q1-Scopus).

[CT2] **Dang Hai Pham** and Hoa Ngoc Nguyen "SCM-ReID: Enhancing Person Re-Identification by Supervised Contrastive-Metric Learning and Hybrid Loss Optimization", Journal of Electronic Imaging.
https://doi.org/10.1117/1.JEI.34.4.043001. (SCI-E, Q3-Scopus).

[CT3] **Hai Dang Pham**, Ngoc Tu Nguyen, Ngoc Hoa Nguyen "ViTC-UReID: Enhancing Unsupervised Person ReID with Vision Transformer Image Encoder and Camera-Aware Proxy Learning", Journal of Computer Science and Cybernetics (Chấp nhận đăng).

Hội nghị

[CT4] **Dang H. Pham**, Anh D. Nguyen, Long V. Vu and Hoa N. Nguyen, "IQAGA: Image Quality Assessment-Driven Learning with GAN-Based Dataset Augmentation for Cross-Domain Person Re-Identifi cation", SOICT 2023 - 12th International Symposium on Information and Communication Technology.
https://doi.org/10.1145/3628797.3628961. (WoS, Scopus).

[CT5] Anh D. Nguyen, **Dang H. Pham** and Hoa N. Nguyen, "GAN-based Data Augmentation and Pseudo-Label Refinement for Unsupervised Domain Adaptation Person Re-Identification", ICCCI 2023 - 15th International Conference on Computational Collective Intelligence.
https://doi.org/10.1007/978-3-031-41456-5_45. (WoS, Scopus).