

# **BÀI THU HOẠCH**

## **CHỦ ĐỀ: PERSON RE-IDENTIFICATION**

### **A. Thông tin cá nhân**

MSSV: 19127723

Huỳnh Thị Mỹ Thanh

### **B. Bài thu hoạch**

#### **I. Problem statement**

- Input: Hình ảnh hoặc video chứa hình ảnh của một hoặc nhiều người.
- Output: các đặc trưng của người được tái định danh, có thể là một vector số hoặc mã hash để đại diện cho từng đối tượng. Các thông tin khác về đối tượng như tên, địa chỉ và các thuộc tính liên quan khác.

#### **II. Challenge**

- Nhiều ảnh ở nhiều góc độ khác nhau.
- Ảnh với chất lượng thấp.
- Ảnh chụp lúc ánh sáng khác nhau.
- Dáng đứng của người.
- Ảnh bị che khuất.
- Môi trường hoặc ảnh nền phức tạp và lộn xộn.
- Các bounding box được tạo sinh không đáng tin cậy.
- 1 hệ thống camera tự động cập nhật (thêm hoặc bớt camera trong hệ thống an ninh).
- Tập dữ liệu lớn dần nhưng lại cần 1 phương pháp truy xuất hiệu quả.
- Trường hợp không có tập dữ liệu học.
- Đối tượng thay đổi trang phục.

#### **III. Methods**

Một hệ thống closed world Re-ID cơ bản sẽ bao gồm 3 tác vụ chính là:

- Feature representation Learning
- Deep metric learning
- Ranking optimization

#### **A. Close world**

Feature representation – **global feature**: join learning of single image and cross-image representations

Single image tốt hơn cross image

Multiscale: phân biệt khi ở nhiều độ scale khác nhau, tăng cường khả năng phân biệt, tránh mất mát thông tin khi có nhiều layer

Feature representation – **local feature**:

Part aligned bilinear representations (chia cơ thể ra từng part): kết hợp việc học biểu diễn full body và phần đặc trưng local của body đó.

Pose attention: pose-driven matching, pose-guided part attention, semantically part alignment

Horizontal – divided body part region: với việc đặc trưng được chia ra làm

### **Auxiliary feature**

- Semantic attribute: sử dụng các thuộc tính ngữ nghĩa ví dụ "nam tóc ngắn" điều này sẽ tạo ra các descriptors hay characteristics tương ứng với mỗi identity.
- Viewpoint information: kết hợp nhiều viewpoint khác nhau để giải quyết vấn đề các vùng non-overlapping giữa các view khác nhau bằng cách tìm các vùng giống nhau giữa các view (shared view) và kết hợp chúng tại (view confusion).
- Domain information: sử dụng nhiều dataset khác nhau để tận dụng các target giống nhau.
- GAN: sử dụng mô hình tạo sinh ảnh để tạo ra tập ảnh có nhiều view pose, attribute của cùng một đối tượng kết hợp với mô hình phân biệt (discriminative model).
- Data augmentation: sử dụng các kỹ thuật như random resize, cropping, flip, random noise... để tăng độ đa dạng của tập dữ liệu học.

### **Video feature**

- RNN: sử dụng dữ liệu liên tục, với việc dữ liệu có thể được truyền giữa các time-steps thì RNN có thể giúp cải thiện hiệu suất học.
- Spatial and Temporal Attention Pooling Network: để tránh việc có các outlier, ta tích hợp RNN với 2 lớp pooling.
- Chia video dài ra làm các snippet ngắn.

### **Deep metric learning**

- Identity loss: Coi quá trình training như 1 bài toán phân lớp. Với đầu vào là ảnh xi được đánh nhãn yi. Xác suất dự đoán xi được phân lớp yi là  $p(xi|yi)$ . Hàm loss sẽ được tính bằng cross-entropy với n là số mẫu của mỗi batch trong tập train.

- Verification Loss: giúp phân biệt mẫu là positive hay negative của cặp ảnh đầu vào.
- Triplet Loss: coi quá trình training như xếp hạng ảnh truy vấn, khoảng cách giữa cặp ảnh positive và anchor nên nhỏ hơn cặp ảnh negative và anchor.

### **Ranking optimization:**

- Re-ranking: sử dụng độ tương đồng giữa các gallery để tối ưu hóa dãy ảnh được xếp hạng ban đầu.
- Rank fusion: khai thác nhiều ranking list để cải thiện performance của quá trình truy vấn.

### **B. Open-world Re-ID**

**Deep based Re-ID:** hình ảnh độ sâu lưu lại được hình dáng của người. “Recurrent attention models for depth-based person identification” - Kết hợp các convolutional network và recurrent network để xác định vùng cục bộ nhỏ của cơ thể.

**Text to Image Re-ID:** xử lý sự khớp giữa văn bản mô tả và hình ảnh RGB. Điều này là bắt buộc khi không thể có được hình ảnh trực quan của người truy vấn và chỉ có thể cung cấp văn bản mô tả thay thế.

“Deep cross model projection learning for image-text matching”: kiến trúc bao gồm 3 thành phần chính: visual CNN trích xuất đặc trưng, LSTM 2 chiều (bi-LSTM) để encode đặc trưng văn bản, module học chung để liên kết văn bản và hình ảnh.

**Visible infrared Re-ID:** xử lý sự khớp đa phương thức giữa hình ảnh nhìn thấy ban ngày và ảnh hồng ngoại ban đêm. “RGB infrared Cross Modality Person Re-identification” – có thể học các đặc trưng có ở ảnh thường và ảnh hồng ngoại.

**Cross resolution Re-ID:** tiến hành định danh giữa hình ảnh có độ phân giải thấp và độ phân giải cao. “Cascaded SR-GAN for scale-adaptive low resolution person re-identification” tạo hình ảnh người có độ phân giải cao theo cách xếp tầng, kết hợp thông tin nhận dạng.

**End to end Re-ID:** giảm bớt sự phụ thuộc vào bước bổ sung để tạo bounding box.

**Semi-supervised and Unsupervised Re-ID:** học các thành phần bất biến (chi tiết nổi bật) dẫn đến khả năng phân biệt hoặc khả năng mở rộng hạn chế. “Dynamic graph matching (DGM)” xây dựng ước tính nhãn dưới dạng bài toán khớp đồ thị hai phía. Học sâu có thể được áp dụng để tìm hiểu các mô hình Re-ID. Tuy nhiên, làm thế nào để xác định danh tính mới chưa biết vẫn là một vấn đề chưa được giải quyết.

## IV. Conclusion

Person re-identification là một lĩnh vực nghiên cứu đang phát triển nhanh chóng trong lĩnh vực nhận dạng hình ảnh. Hệ thống nhận dạng lại người có đầu vào là hình ảnh hoặc video chứa hình ảnh của người và đầu ra là kết quả nhận dạng lại người, đặc trưng của các người được nhận dạng lại, và các thông tin khác liên quan. Person re-identification có thể được ứng dụng trong nhiều lĩnh vực, từ an ninh, giám sát, quản lý đám đông, đến giao thông và nhiều lĩnh vực khác.

Tuy nhiên, person re-identification còn đối diện với nhiều thách thức, như khó khăn trong việc đối mặt với biến đổi về góc nhìn, ánh sáng, thay đổi trang phục, và che mặt của các đối tượng. Công nghệ person re-identification cũng đặt ra nhiều vấn đề liên quan đến quyền riêng tư và đạo đức trong việc sử dụng dữ liệu hình ảnh của con người.

Tuy nhiên, person re-identification vẫn có tiềm năng lớn để cải thiện tính năng của các ứng dụng nhận dạng hình ảnh hiện tại và đem lại nhiều lợi ích cho xã hội. Các nghiên cứu trong lĩnh vực này cần tiếp tục đổi mới, phát triển các phương pháp nhận dạng lại người mới, và đồng thời chú ý đến các vấn đề liên quan đến đạo đức và quyền riêng tư để đảm bảo sự phát triển bền vững và đúng đắn của công nghệ person re-identification trong tương lai.