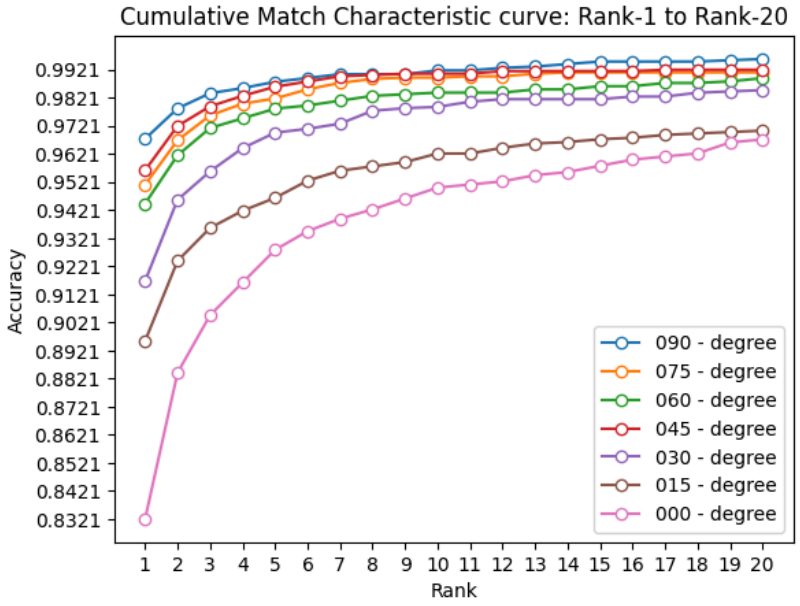
GIẢI TRÌNH VÀ TIẾP THU Ý KIẾN PHẢN BIỆN LUẬN VĂN

Kính gửi các thầy trong hội đồng,
Em xin được bổ sung thông tin cho luận văn dựa vào nội dung phản biện từ các thầy như sau

1. **Ý kiến của TS. Trần Hải Anh**

Ý Kiến 01:
Luận văn dùng Rank-1 Accuracy làm chỉ số chính nhưng thiếu các đánh giá bổ trợ thường gặp (Rank-5, ROC/CMC curve, phân tích theo điều kiện khó như thay đổi trang phục, vật phẩm mang theo, v.v.)

Học viên giải trình:
Bổ sung chỉ số: Em đã cập nhật bảng kết quả thực nghiệm chi tiết từ Rank-1 đến Rank-20 tại trang 42 của luận văn và sử dụng chúng cho biểu đồ CMC.
Đồ thị CMC: Đồ thị đường cong đặc tính khớp tích lũy (CMC) cho tất cả các góc quan sát (từ 0◦ đến 90◦) đã được bổ sung tại trang 43.
Về điều kiện môi trường: Các yếu tố biến đổi như trang phục hay vật phẩm mang theo hiện nằm ngoài phạm vi cung cấp của bộ dữ liệu OU-MVLP và mục tiêu nghiên cứu cụ thể của đề tài này



Ý Kiến 02: Baseline lựa chọn hạn chế và lý do lựa chọn còn sơ sài (chủ yếu vì "dùng OU-MVLP" và "công bố gần đây"). Chưa làm rõ: có tái triển khai cùng pipeline tiền xử lý hay trích số "best accuracy" từ bài báo

Học viên giải trình:
Lý do chọn OU-MVLP: Bộ dữ liệu này được ưu tiên vì đã cung cấp sẵn các hình ảnh năng lượng dáng đi (GEI) được tiền xử lý chuẩn hóa. Việc này đảm bảo tính tương đồng về quy trình xử lý dữ liệu (pipeline) với các nghiên cứu cơ sở được trích dẫn. Ngoài ra, đây là bộ dữ liệu lớn nhất thế giới và thường dùng chuẩn mực trong lĩnh vực vì đã được xử lý cẩn thận.
Tiêu chí chọn Baseline: Các bài báo được chọn làm baseline đều là các công bố gần đây có độ chính xác Rank-1 tối ưu. Em trích dẫn kết quả tốt nhất được công bố trực tiếp trong các bài báo này để đảm bảo tính khách quan.

Ý Kiến 03: Quy tắc chia train/validation theo ngưỡng ID("<08000" và ">=08000") được nêu ngắn gọn nhưng chưa thấy mô tả rõ test protocol, seed, số lần chạy hay sai số/ độ lệch chuẩn
Chính tác giả cũng thừa nhận trong hướng tương lai rằng cần nhấn mạnh cross dataset generation và thử trên các bộ khác như CASIA-B

Học viên giải trình:
Hạn chế thực nghiệm: Do hạn chế về hạ tầng tính toán (sử dụng tài nguyên miễn phí trên Kaggle), việc thực hiện kiểm định chéo K-fold hiện chưa khả thi. Thời gian huấn luyện mô hình theo phân tách dữ liệu hiện tại đã kéo dài hơn 02 tháng.
Thử nghiệm trên CASIA-B: Việc mở rộng sang bộ dữ liệu CASIA-B gặp khó khăn do dữ liệu gốc chỉ cung cấp video dáng đi, đòi hỏi quy trình tiền xử lý phức tạp để chuyển đổi sang GEI. Hơn nữa, hầu hết các công bố trước đó không mô tả chi tiết quy trình chuyển đổi này, gây khó khăn cho việc tái lập kết quả tương đương.

Ý Kiến 04: Mâu thuẫn/ thiếu nhất quán trong mô tả hệ thống
Trong phần gợi ý tương lai, luận văn nói hệ thống hiện tại tổng hợp từ một tập cố định 4 góc camera, trong khi phần thực nghiệm nhấn mạnh two-view (và trình bày bằng ma trận 2 góc) . Cần làm rõ "cấu hình mặc định " và cơ chế chọn số/ góc view trong huấn luyện và suy luận

Học viên giải trình:

Tính linh hoạt của mô hình: Mô hình đề xuất có khả năng suy luận linh hoạt với đầu vào từ một, hai hoặc nhiều góc nhìn khác nhau tùy thuộc vào nguồn lực hệ thống thực tế. Trong đó em tập trung nghiên cứu tìm hiểu khả năng kết hợp nhiều view để nâng cao độ chính xác nhận dạng so với một view. Tuy nhiên để có thể so sánh hiệu năng của mô hình trong nghiên cứu với các mô hình đã được xuất bản, em đã sử dụng kết quả suy luận từ 1 view để đưa vào báo cáo so sánh.
Mục đích thực nghiệm: Sử dụng kết quả từ một góc nhìn để so sánh trực tiếp hiệu năng với các nghiên cứu (SOTA) chỉ dùng một góc nhìn hiện đang có. Hiện tại chỉ với một góc nhìn, phương pháp đề xuất đã cho kết quả tốt hơn các phương pháp SOTA hiện tại cũng dùng đặc trưng dáng đi GEI. Bổ sung thực nghiệm hai góc nhìn để chứng minh rằng độ chính xác có thể cải thiện rõ rệt mà không cần tái huấn luyện mô hình.
Hướng phát triển: Em định hướng phát triển cơ chế tự động lựa chọn các góc nhìn tối ưu (adaptive view selection) dựa trên điều kiện môi trường trong tương lai.

Ý Kiến 05: Luận văn kết luận phương pháp "computation efficient", nhưng phần phương pháp lại thừa nhận một lựa chọn kiến trúc "ổn định hơn" không áp dụng được vì thiếu tài nguyên tinh toán . Chưa có đo đạc định lượng thời gian train/infer, FLOPs, VRAM, throughput(số mẫu / giây) nên nhận định "hiệu quả chưa đủ căn cứ"

Học viên giải trình:
Em rất cám ơn thầy đã chỉ ra điểm chưa chính xác, em xin đính chính lại như sau:
Kiến trúc mô hình: Trong nội dung báo cáo, kiến trúc mà em sử dụng có dùng hàm Dropout nhằm tối ưu hóa chi phí huấn luyện.Tuy nhiên phương pháp này dựa khá nhiều vào việc lựa chọn ngẫu nhiên các giá trị bị "không hóa", "zeroed out". Điều đó khiến cho kiến trúc không ổn định trong quá trình huấn luyện. Kiến trúc "ổn định hơn" (không dùng Dropout) là một phương án đang trong quá trình nghiên cứu thực nghiệm và chứng minh toán học, kiến trúc này không nhằm mục tiêu tối ưu hóa tài nguyên như mô hình chính được đề xuất.
Bổ sung định lượng: Em đã bổ sung các thông số về thời gian huấn luyện/suy luận và FLOP tại mục "Other measurement" (trang 42) để làm căn cứ cho nhận định về tính hiệu quả của phương pháp.

2. **Ý kiến của PGS.TS Phạm Tuấn Minh**

Ý Kiến 01: The author states that the first contribution is the proposal of a "GNN inspired knowledge propagation mechanism" for the problem, in which embeddings from one camera angle exchange information with those from neighboring angles. This

contribution should be presented more clearly. Specifically, the author should clarify the definition of node and edges in the GNN, identify which parameters are learnable at this stage, and explain how this component is trained.

Học viên giải trình:
Em rất xin lỗi về phần trình bày chưa được tường minh khiến cho người đọc thấy nội dung chưa đủ tính thuyết phục và tính nhất quán. Em xin được bổ sung thông tin như sau:
Về cơ chế lan truyền tri thức lấy cảm hứng từ GNN:
Định nghĩa: Đề xuất này kế thừa tư tưởng về cơ chế lan truyền thông điệp (message passing) của mạng đồ thị (GNN) chứ không xây dựng cấu trúc nút (node) và cạnh (edge) tường minh.
Cơ chế hoạt động: Dựa trên giả thuyết về tính tương đồng của thông tin thực thể giữa các góc nhìn, em thực hiện chuyển giao tham số (transfer learning) từ mô hình đã huấn luyện ở góc nhìn này sang góc nhìn khác. Quá trình này giúp mô hình mới kế thừa tri thức và tăng khả năng phân biệt đối tượng. Nội dung chi tiết đã được bổ sung tại trang 36.

Ý Kiến 02: The author further claims that the second contribution is the use of a Transformer-based model to synthesize embeddings from multiple camera angles into 128-dimensional vectors, which then is used for identity recognition via cosine similarity comparison. This contribution also requires clearer presentation. For example, the author should clearly explain the representation of each token fed into the Transformer, the number of encoded layers, the number of attention heads per encoder layer, the meaning of these model parameters, and how their values are selected.

The author should also discuss the rationale for choosing a Transformer in the proposed model. Using the output of a Transformer as a stored representation for identity recognition may not be fully appropriate given the basic role of the model, as the resulting embedding strongly depends on input set at the time of generation. Fundamentally, transformers are designed not for feature extraction but for modeling dependency relationships among elements in a sequence.

Học viên giải trình:
Em hoàn toàn nhất trí với phát biểu ở trên về bản chất của Transformer.
Về việc sử dụng mô hình Transformer:

Trong nghiên cứu này, em sử dụng CNN đóng vai trò là bộ trích xuất đặc trưng (descriptor) và Transformer đóng vai trò là bộ suy luận (reasoner) để khai thác các mối quan hệ tiềm ẩn giữa các đặc trưng.

Cách tiếp cận này tương tự như việc ứng dụng mô hình BERT trong phân loại văn bản, tập trung vào việc mô hình hóa sự phụ thuộc giữa các yếu tố trong chuỗi đặc trưng để đưa ra kết quả nhận dạng cuối cùng.