TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN

BÁO CÁO TIẾN ĐỘ GIỮA KỲ

SINH TRẮC HỌC

CHỦ ĐỀ: NHẬN DIỆN DÁNG ĐI

1 THÔNG TIN CHUNG

Giảng viên hướng dẫn:

- PGS. TS. Lê Hoàng Thái (Khoa Công nghệ thông tin)
- Thầy Dương Thái Bảo (Khoa Công nghệ thông tin) - Tìm kiếm học vị của Thầy (ThS? TS?,...) và bổ sung trước khi nộp

Nhóm sinh viên thực hiện:

1. Phạm Thái Huy (MSSV: 21120081)
2. Nguyễn Đức Mạnh (MSSV: 22120204)
3. Lê Quang Vĩnh Quyền (MSSV: 22120307)

2 NỘI DUNG BÁO CÁO

2.1 Tổng hợp nội dung chương sách được chọn

Quyền phụ trách phần 1,4,5,P3D

Mạnh phụ trách phần 2,3

Tổng hợp đầy đủ nội dung, với tổ chức thư mục **BẮT BUỘC** giống hoàn toàn của chương. **Lưu ý:**

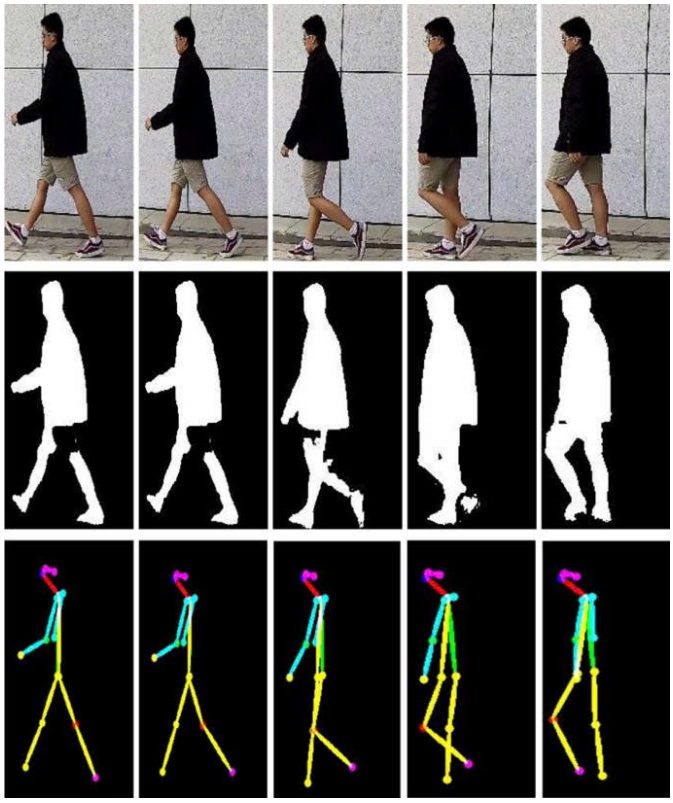
- Đề cập đầy đủ thông tin cốt lõi. Bỏ qua các *kể chuyện dài dòng* (tóm tắt lại nếu cần).
- Các biểu thức toán phải được viết bằng môi trường toán của LaTeX.
- Đề cập đầy đủ các số liệu.
- Chèn hình ảnh nếu cần thiết.
- Không sao chép nội dung trong sách.

2.1.1 Giới thiệu

Với mục tiêu định danh con người từ khoảng cách xa và trong môi trường không bị ràng buộc, nhận diện dáng đi tập trung phân tích và xác định danh tính của cá nhân dựa trên các kiểu đi bộ đặc trưng hoặc cách thức vận động của họ qua các chuỗi hình ảnh được thu thập từ camera. Khác với các đặc điểm sinh trắc học vật lý truyền thống như dấu vân tay, khuôn mặt hay mống mắt, dáng đi được xếp vào nhóm sinh trắc học hành vi. Đặc điểm này được hình thành từ cấu trúc cơ xương độc nhất và thói quen vận động của mỗi người, tạo nên một dấu ấn riêng biệt có khả năng phân biệt cao. Ngay từ những năm 1970, các nghiên cứu tâm lý học nền tảng của Cutting và Kozlowski đã chứng minh rằng con người có khả năng nhận diện người quen chỉ thông qua sự chuyển động của các điểm sáng mô phỏng dáng

đi mà không cần bất kỳ thông tin chi tiết nào về ngoại hình hay khuôn mặt.

Một trong những lợi thế quan trọng nhất giúp nhận diện dáng đi trở nên nổi bật là khả năng hoạt động hiệu quả trong các điều kiện khó khăn. Trong khi nhận diện khuôn mặt hay mống mắt yêu cầu đối tượng phải hợp tác, đứng gần thiết bị thu nhận và cần hình ảnh độ phân giải cao, nhận diện dáng đi có thể thực hiện từ khoảng cách xa, có thể lên tới hàng trăm mét và chấp nhận dữ liệu có độ phân giải thấp. Đặc biệt, đây là một phương thức nhận diện không xâm lấn, nghĩa là hệ thống có thể thu thập dữ liệu thụ động qua camera giám sát mà không cần sự tương tác trực tiếp hay sự chủ động hợp tác của đối tượng. Hơn nữa, do dáng đi là một hành vi vô thức bắt nguồn từ cơ chế sinh học, do đó việc ngụy trang hay giả mạo dáng đi trong thời gian dài là vô cùng khó khăn đối với các đối tượng muốn che giấu danh tính.



Hình 1: Nhận diện dáng đi.

Nhìn chung, việc ứng dụng hệ thống nhận diện dáng đi hiện thực phải đối mặt với rất nhiều thử thách lớn đến từ các yếu tố ngoại cảnh. Độ chính xác của hệ thống thường sẽ rất dễ bị ảnh hưởng nghiêm trọng bởi sự thay đổi của góc nhìn của camera, đây được xem là yếu tố gây nhiễu lớn nhất làm thay đổi hình dạng hình học của đối tượng trên khung hình. Bên cạnh đó, các điều kiện về trang phục như áo khoác dày che khuất cơ thể, hoặc trạng thái mang vác vật dụng như ba lô, túi xách cũng làm thay đổi trọng tâm và biên độ dao động của dáng đi. Các yếu tố môi trường khác như bề mặt đường đi, điều kiện ánh sáng hay sự che khuất bởi vật cản cũng góp phần làm giảm hiệu suất nhận diện khi áp dụng vào các tình huống đời sống.

2.1.2 Vấn đề

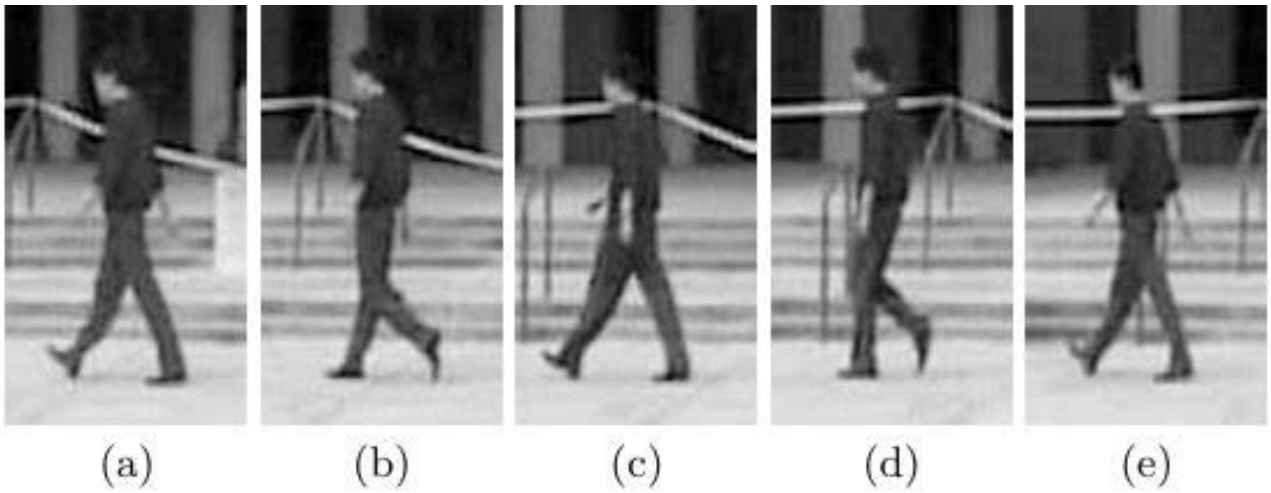
2.1.3 Phương pháp

2.1.4 Thảo luận và Hướng nghiên cứu tiếp theo

Hình dáng và động lực học của dáng đi

Trong nghiên cứu về dáng đi, sự tranh luận giữa vai trò của hình dáng và động lực học luôn là tiêu điểm. Các nghiên cứu thực nghiệm ban đầu chỉ ra rằng con người có khả năng nhận diện danh tính dựa trên các đặc điểm động lực học ngay cả khi hình dáng bị che khuất, tuy nhiên các phương pháp phân tích dựa trên hình dáng hình bóng lại mang lại hiệu quả vượt trội. Nhiều thuật toán tiên tiến gần đây đã chứng minh rằng việc tập trung vào hình dáng của từng giai đoạn trong chu kỳ bước đi giúp hệ thống đạt được độ chính xác cao hơn, đặc biệt là khi phải đối mặt với các biến số khó như thay đổi bề mặt đi bộ. Mặc dù vậy, động lực học vẫn đóng vai trò không thể thiếu vì nó chứa đựng các thông tin về tốc độ và sự chuyển tiếp giữa các pha vận động vốn mang tính đặc trưng cho từng cá nhân. Tuy nhiên, các nghiên cứu mới gần đây chỉ ra rằng việc phụ thuộc quá nhiều vào hình dáng sẽ khiến hệ thống dễ bị sai lệch khi đối tượng thay đổi trang phục. Do

đó, xu hướng hiện tại là phát triển các mô hình học biểu diễn không gian - thời gian phân cấp, cho phép tách biệt các đặc trưng chuyển động cốt lõi ra khỏi các đặc điểm hình dáng bề ngoài dễ thay đổi, từ đó tận dụng sức mạnh của cả hai yếu tố này.



Hình 2: Minh họa chu kỳ dáng đi được chia thành bốn giai đoạn: (i) tựa chân phải; (ii) lăng chân trái; (iii) tựa chân trái; và (iv) lăng chân phải, tương ứng với các trạng thái từ (a) đến (e). Khoảng thời gian cả hai chân cùng tiếp xúc mặt sàn được gọi là giai đoạn hỗ trợ kép.

Chất lượng hình bóng và nhận dạng dáng đi

Chất lượng của các hình bóng phụ thuộc vào khả năng phân biệt giữa nền và đối đượng. Trong môi trường ngoài trời, các yếu tố nhiễu như bóng đổ, thay đổi ánh sáng và sự chuyển động của hậu cảnh khiến việc tách hình bóng trở nên cực kỳ khó khăn. Tuy nhiên, một phát hiện chỉ ra rằng sự sụt giảm hiệu suất khi thay đổi bề mặt hoặc thời gian không hoàn toàn do lỗi xử lý hình bóng ở cấp độ thấp gây ra. Ngay cả khi sử dụng các hình bóng đã được tiền xử lý thủ công, kết quả vẫn cho thấy sự suy giảm đáng kể, nghĩa là bản thân dáng đi của con người đã có những thay đổi cơ bản khi điều kiện môi trường thay đổi. Điều này cho thấy các công trình tương tai thay vì tìm kiếm các phương pháp tốt hơn để phát hiện hình bóng nhằm cải thiện nhận dạng, việc nghiên cứu và tách biệt các thành phần của dáng đi không thay đổi theo giày dép, bề mặt hoặc thời gian sẽ hiệu quả hơn.

Các biến số

Thách thức của nhận diện dáng đi nằm ở việc duy trì độ ổn định trước các tác động của các yếu tố ngoại cảnh. Trong khi các yếu tố như loại giày dép hay việc mang theo túi xách có tác động tương đối nhỏ, thì sự thay đổi về bề mặt đi bộ và khoảng cách thời gian giữa các lần thu thập dữ liệu lại gây ra những ảnh hưởng tiêu cực nghiêm trọng. Đặc biệt, nhận dạng dáng đi theo thời gian là một bài toán khó do sự thay đổi về trang phục theo mùa và những biến đổi tự nhiên trong cơ thể người theo thời gian. Ngoài ra, việc di chuyển trong các môi trường thực tế với các góc nhìn camera đa dạng, vật cản che khuất và độ phân giải thấp vẫn là những trở ngại cần được giải quyết. Do đó các nghiên cứu trong tương lai cần tập trung vào việc mô hình hóa các thành phần dáng đi không thay đổi hoặc tìm cách dự đoán sự biến đổi của dáng đi khi chuyển từ bề mặt này sang bề mặt khác.

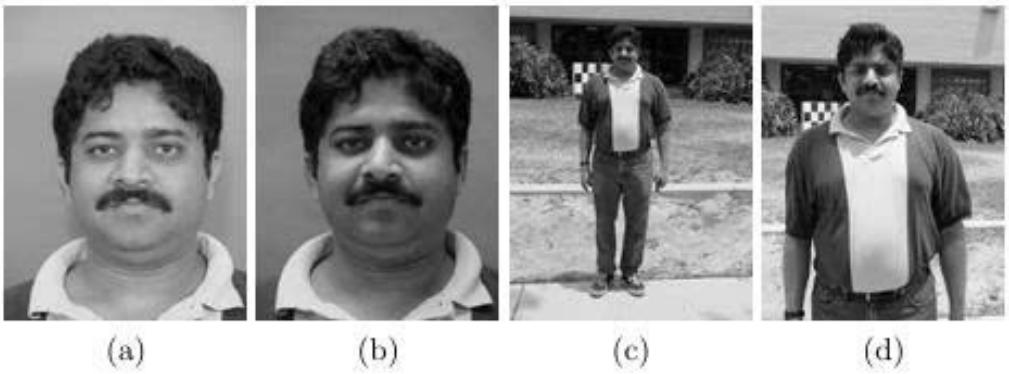
Các bộ dữ liệu trong tương lai

Sự phát triển của nhận dạng dáng đi phụ thuộc rất lớn vào quy mô và độ đa dạng của các bộ dữ liệu. Các chuyên gia nhận định rằng cần có những bộ dữ liệu khổng lồ với quy mô lên tới hàng nghìn đối tượng để hỗ trợ việc hiểu sâu hơn về sự biến thiên của dáng đi trong điều kiện ngoài trời và theo thời gian. Các bộ dữ liệu chuẩn như CASIA-B hay OU-MVLP đang dần trở nên bão hòa khi các mô hình Học sâu đã đạt độ chính xác rất cao trên đó. Hiện nay, các bộ dữ liệu như GREW hay Gait3D đã bắt đầu chuyển hướng từ môi trường phòng thí nghiệm sang môi trường thực tế với hàng chục nghìn danh tính và hàng triệu chuỗi hình ảnh. Một hướng đi mới đầy triển vọng là sử dụng dữ liệu tổng hợp được tạo ra từ các mô hình cơ thể người ảo, giúp giải quyết vấn đề thiếu hụt dữ liệu được dán nhãn và giảm bớt các rào cản về chi phí thu thập dữ liệu thực tế. Đồng thời, việc khai thác dữ liệu video không nhãn trên quy mô lớn thông qua các phương pháp học tự giám sát đang trở thành một lĩnh vực nghiên cứu đầy tiềm năng.

Kết hợp khuôn mặt và dáng đi

Mặc dù dáng đi có ưu thế ở khoảng cách xa, thì việc kết hợp giữa nhận dạng dáng đi và khuôn mặt là một giải pháp tối ưu để nâng cao độ tin cậy của hệ thống.

Dáng đi có thể được sử dụng khi đối tượng ở khoảng cách xa, trong khi khuôn mặt sẽ phát huy tác dụng khi đối tượng tiến lại gần camera hơn. Việc kết hợp đa phương thức dáng đi và mặt người mang lại hiệu suất vượt trội so với việc chỉ sử dụng một loại sinh trắc học đơn lẻ, đồng thời giúp hệ thống bền bỉ hơn trước các nhiễu động của từng loại dữ liệu. Trong tương lai, việc tích hợp này không chỉ dừng lại ở khuôn mặt mà còn có thể mở rộng sang các phương thức khác như thông tin về chiều cao, kích thước các bộ phận cơ thể hoặc dữ liệu từ nhiều góc nhìn camera cùng lúc để tạo ra một hồ sơ định danh toàn diện và chính xác hơn



Hình 3: Các mẫu khuôn mặt dưới nhiều điều kiện khác nhau: (a) và (b) là ảnh tập mẫu trong điều kiện ánh sáng khác nhau; (c) và (d) là ảnh tập kiểm tra chụp ngoài trời ở khoảng cách xa và gần.

Quyền riêng tư và bảo mật sinh trắc học

Một khía cạnh mới đang ngày càng được chú trọng là tính bảo mật và quyền riêng tư của dữ liệu dáng đi. Do dáng đi có thể được thu thập từ xa mà không cần sự hợp tác hay nhận biết của đối tượng, công nghệ này làm dấy lên những lo ngại nghiêm trọng về quyền riêng tư cá nhân và sự giám sát đại chúng. Bên cạnh đó, vấn đề an ninh của chính hệ thống AI cũng đáng báo động với sự xuất hiện của các cuộc tấn công đối kháng, nơi kẻ tấn công có thể thay đổi một chút dáng đi hoặc thêm nhiễu vào video để đánh lừa hệ thống. Các luật lệ như quy định chung về Bảo vệ Dữ liệu Châu Âu (GDPR) đã đặt ra những hạn chế chặt chẽ đối với việc sử dụng dữ liệu sinh trắc học, thúc đẩy cộng đồng nghiên cứu phải tìm kiếm các giải pháp bảo vệ quyền riêng tư. Các hướng nghiên cứu mới bao gồm việc phát

triển các phương pháp ẩn danh hóa, mã hóa video dáng đi sao cho hệ thống nhận dạng vẫn hoạt động nhưng danh tính con người không thể bị quan sát bằng mắt thường, cũng như tăng cường khả năng chống lại các cuộc tấn công giả mạo hoặc tấn công đối nghịch nhằm đánh lừa hệ thống nhận dạng.

2.1.5 Kết luận

Nhìn lại toàn bộ quá trình phát triển, nhận dạng dáng đi đã khẳng định vị thế là một trong những công nghệ sinh trắc học tiềm năng nhất, với điểm mạnh về khả năng định danh tầm xa và không xâm lấn mà các phương pháp truyền thống như khuôn mặt hay vân tay không thể thay thế. Sự chuyển dịch mạnh mẽ từ các kỹ thuật thị giác máy tính cổ điển sang các kiến trúc Học sâu tiên tiến đã nâng tầm lĩnh vực này, giúp các hệ thống hiện đại đạt được độ chính xác ấn tượng trên các bộ dữ liệu tiêu chuẩn. Các nghiên cứu đột phá gần đây đã chứng minh rằng việc khai thác sâu các biểu diễn không gian - thời gian phân cấp là chìa khóa để tách biệt các đặc trưng vận động cốt lõi khỏi những yếu tố nhiễu loạn của bề mặt, mở ra triển vọng to lớn trong việc giải mã hành vi vận động của con người.

Tuy nhiên, thực tế triển khai đã chỉ ra một khoảng cách đáng kể về tính thực tiễn giữa môi trường phòng thí nghiệm lý tưởng và thế giới thực hỗn loạn. Như một vài phân tích thực nghiệm đã làm rõ hiệu suất của các thuật toán hàng đầu vẫn sụt giảm nghiêm trọng khi đối mặt với sự đa dạng không giới hạn của các biến số ngoại cảnh như góc quay camera an ninh phức tạp, sự thay đổi trang phục theo mùa, hay điều kiện ánh sáng và vật che khuất trong môi trường tự nhiên. Điều này khẳng định rằng, mặc dù chúng ta đã giải quyết tốt bài toán so khớp mẫu trong điều kiện kiểm soát, nhưng bài toán nhận dạng ở ngoài thực tế vẫn là thách thức lớn cần tiếp tục giải quyết.

Trong tương lai, sự phát triển của nhận dạng dáng đi sẽ không còn đơn thuần là cuộc đua về các chỉ số độ chính xác trên tập dữ liệu cũ, mà sẽ là sự chuyển mình sang các hệ thống thông minh, bền vững và an toàn hơn. Xu hướng tất yếu

sẽ là sự kết hợp đa phương thức giữa dữ liệu để vượt qua giới hạn của camera quang học, cùng với việc áp dụng các kỹ thuật học không giám sát để tận dụng nguồn dữ liệu khổng lồ chưa gán nhãn. Đồng thời, khi công nghệ này đi sâu vào đời sống, các vấn đề về bảo mật chống giả mạo và bảo tồn quyền riêng tư sẽ trở thành những trụ cột quan trọng ngang hàng với hiệu năng kỹ thuật, đảm bảo rằng nhận dạng dáng đi không chỉ là một công cụ giám sát hiệu quả mà còn là một công nghệ có trách nhiệm và đáng tin cậy.

2.2 Phương pháp trình bày

Người phụ trách: Huy.

- Trình bày chi tiết về phương pháp gốc đã chọn. (đặt vấn đề, đề xuất phương pháp, tiến hành thực nghiệm, phân tích kết quả, bàn luận, tổng kết)
- Phân tích của nhóm về hạn chế tiềm ẩn của phương pháp.

2.3 Hướng nghiên cứu và thực nghiệm

Hướng nghiên cứu: Dựa trên phân tích về hạn chế tiềm ẩn, nhóm đều xuất các giải pháp thay thế, tái thực nghiệm, so sánh với phương pháp gốc, phân tích và bàn luận.

2.3.1 Thay thế kỹ thuật ARME thành P3D

Cơ sơ sở lý thuyết

Xét một thao tác tích chập trên một vùng cơ thể thứ j tại cấp độ l. Giả sử đầu vào là tensor $X \in \mathbb{R}^{C_{in} \times T \times H \times W}$. Bộ lọc có kích thước không gian $k \times k$ và thời gian d.

Đối với kĩ thuật ARME, thì ARME sử dụng tích chập 3D tiêu chuẩn để trích xuất đặc trưng không gian và thời gian đồng thời. Cụ thể là dùng một kernel 3D kích thước $d \times k \times k$ trượt trên cả ba chiều (thời gian, chiều cao, chiều rộng).
Với mỗi vùng j, đầu ra Y_j được tính bằng:

$$Y_j = f_j(X_j) = W_{3D} * X_j + b$$

Trong đó $W_{3D} \in \mathbb{R}^{C_{out} \times C_{in} \times d \times k \times k}$.

Đối với kĩ thuật P3D, P3D tách một kernel 3D kích thước $d \times k \times k$ thành hai kernel riêng biệt: một kernel không gian $(1 \times k \times k)$ và một kernel thời gian $(d \times 1 \times 1)$. Theo cơ chế như sau:

- Kernel không gian (S): Sử dụng bộ lọc kích thước $1 \times k \times k$, tương đương với 2D CNN để học các đặc trưng hình ảnh.
- Kernel thời gian (T): Sử dụng các bộ lọc $d \times 1 \times 1$ để xây dựng các kết nối thời gian giữa các bản đồ đặc trưng liền kề.

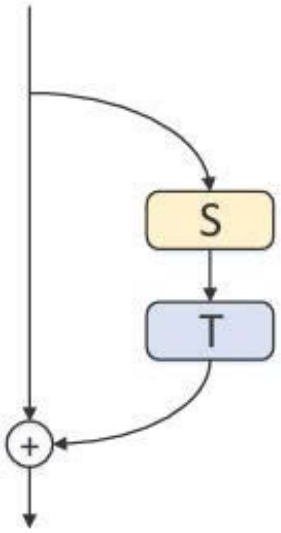
Thay vì tính trực tiếp, ta thực hiện tuần tự:

$$Y_{Spatial} = S(X_j) = W_S * X_j \quad (\text{kernel } 1 \times k \times k)$$

$$Y_j = T(Y_{Spatial}) = W_T * Y_{Spatial} \quad (\text{kernel } d \times 1 \times 1)$$

Trong đó $W_S \in \mathbb{R}^{C_{out} \times C_{in} \times 1 \times k \times k}$ và $W_T \in \mathbb{R}^{C_{out} \times C_{out} \times d \times 1 \times 1}$.
Trong nghiên cứu này, tôi áp dụng cấu trúc Residual của P3D-A. Cụ thể là thành phần thời gian (T) đi trực tiếp sau thành phần không gian (S) trên cùng một đường dẫn. Đầu ra của không gian là đầu vào của thời gian.

$$x_{t+1} = (I + T \cdot S) \cdot x_t = x_t + T(S(x_t))$$



P3D-A

Hình 4: Kiến trúc của khối P3D-A: Các bộ lọc không gian (S) và thời gian (T) được sắp xếp nối tiếp.

Ưu nhược điểm và độ phức tạp tính toán

Về mặt tính toán, module ARME trong HSTL sử dụng tích chập 3D tiêu chuẩn $(3 \times 3 \times 3)$ nên tốn kém tài nguyên với số lượng tham số tỷ lệ thuận với $3 \times 3 \times 3 = 27$. Trong khi đó, kỹ thuật P3D tách kernel này thành hai phần riêng biệt: không gian $(1 \times 3 \times 3)$ và thời gian $(3 \times 1 \times 1)$, giúp giảm số lượng tham số xuống chỉ còn tỷ lệ với $1 \times 3 \times 3 + 3 \times 1 \times 1 = 12$. Như vậy, việc chuyển sang P3D giúp giảm khối lượng tính toán và tham số khoảng 2.25 lần. Sự tối ưu này cực kỳ quan trọng đối với kiến trúc chia vùng của HSTL, cho phép bạn xây dựng mô hình sâu hơn hoặc xử lý dữ liệu lớn hơn mà không bị quá tải bộ nhớ.

Mặc dù P3D đã giảm đáng kể số lượng tham số và chi phí tính toán, tuy nhiên việc tách biệt không gian và thời gian có thể làm giảm khả năng học các mối tương quan chặt chẽ tức thời giữa hai miền này so với ARME.

Mức độ cải thiện kỳ vọng

Việc thay thế ARME bằng P3D được kỳ vọng sẽ mang lại các lợi ích sau:

- Giảm đáng kể chi phí tính toán và bộ nhớ, cho phép mô hình xử lý các chuỗi video dài hơn hoặc nhiều vùng cơ thể hơn mà không bị quá tải.
- Vì P3D sẽ có khả năng khái quát hóa cực tốt trên nhiều tác vụ video và bộ dữ liệu khác nhau như Sports-1M, UCF101. Khi đưa vào HSTL, nó có thể giúp mô hình hoạt động ổn định hơn trên các tập dữ liệu ngoài thực tế như GREW hoặc Gait3D.
- Trong nhận diện dáng người, việc sử dụng tích chập P3D trong tập CASIA-B hy vọng sẽ cải thiện độ chính xác như trong DeepGaitV2-P3D đã cải thiện độ chính xác tới 9.1% so với phiên bản 2D thuần túy trên một số bộ dữ liệu.

Lưu ý:

- Phân tích chi tiết giữa kĩ thuật gốc và kĩ thuật đề xuất (ưu nhược điểm, độ phức tạp, biểu thức toán, mức độ cải thiện kỳ vọng đóng góp vào hiệu suất của toàn mô hình).
- Không chèn mã nguồn vào báo cáo.

Hướng đề xuất	**Người phụ trách**
Thay **Conv3D** bằng **P3D**	Mạnh, Quyền
Thay **Triplet Loss** bằng **Circle Loss**	Huy
Bổ sung kĩ thuật **TSM** vào module chứa **Conv3D**	Huy

3 LỜI KẾT

DEADLINE: 22h - 25/12/2025

4 TÀI LIỆU THAM KHẢO