A View Khung phân loại độc lập cho các tư thế Yoga

Mustafa Chasmai Khoa
học và Kỹ thuật Máy tính Học viện
Công nghệ Ấn Độ Delhi
cs1190341@cse.iitd.ac.in

Kỹ thuật điện
Nirjhar Das Học viện
Công nghệ Ấn Độ Delhi ee3190585@iitd.ac.in

Trường Công nghệ
Thông tin Aman Bhardwaj Học viện
Công nghệ Ấn Độ Delhi
aman.bhardwaj@cse.iitd.ac.in

Rahul Garg
Khoa học và Kỹ thuật Máy tính Học
viện Công nghệ Ấn Độ Delhi
rahulgarg@cse.iitd.ac.in

TRỪU TƯỢNG

Yoga là một môn tập luyện được hoan nghênh trên toàn cầu và được khuyến khích rộng rãi để có một cuộc sống khỏe mạnh. Duy trì tư thế đúng trong khi thực hiện Yogasana là vô cùng quan trọng. Trong công việc này, chúng tôi sử dụng phương pháp học chuyển đổi từ các mô hình Ước tính tư thế con người để trích xuất 136 điểm chính trải khắp cơ thể nhằm huấn luyện bộ phân loại Rừng ngẫu nhiên được sử dụng để ước tính các Yogasana. Kết quả được đánh giá trên cơ sở dữ liệu video yoga mở rộng được thu thập trong nhà gồm 51 đối tượng được ghi lại từ 4 góc camera khác nhau. Chúng tôi đề xuất một sơ đồ 3 bước để đánh giá khả năng khái quát hóa của một lớp học Yoga bằng cách kiểm tra nó trên 1) khung hình không nhìn thấy, 2) đối tượng không nhìn thấy và 3) góc máy ảnh không nhìn thấy. Chúng tôi lập luận rằng đối với hầu hết các ứng dụng, độ chính xác của việc xác thực trên các đối tượng không nhìn thấy và các góc máy ảnh không nhìn thấy sẽ là quan trọng nhất. Chúng tôi phân tích thực nghiệm trên ba bộ dữ liệu công khai, lợi thế của việc học chuyển đổi và khả năng rò rỉ mục tiêu. Chúng tôi tiếp tục chứng minh rằng độ chính xác của phân loại phụ thuộc rất nhiều vào phương pháp xác thực chéo được sử dụng và thường có thể gây hiểu nhầm. Để thúc đẩy nghiên cứu sâu hơn, chúng tôi đã cung cấp công khai bộ dữ liệu điểm chính và mã.

KHÁI NIỆM CCS

• Phương pháp tính toán Nhận biết và hiểu hoạt động.

TỪ KHÓA

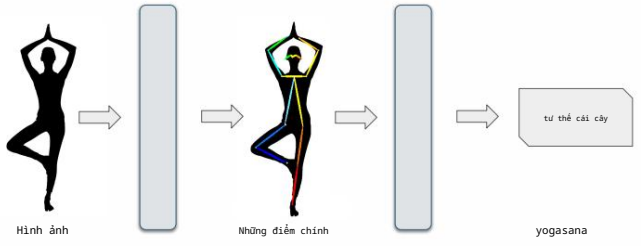
Ước tính tư thế, Yogasana, Học chuyển giao

1. GIỚI THIỆU

Human Activity Recognition (HAR) là một trong những vấn đề quan trọng trong Computer Vision. Nó bao gồm nhận dạng chính xác hoạt động đang được thực hiện bởi một người hoặc một nhóm với sự trợ giúp của hình ảnh, video hoặc dữ liệu thô được thu thập từ Cảm biến. Rất nhiều ứng dụng của HAR bao gồm nhưng không giới hạn ở cuộc sống tích cực và được hỗ trợ (AAL), giám sát chăm sóc sức khỏe, an ninh và giám sát, viễn thông và tự động hóa nhà thông minh [36].

Những tiến bộ này trong HAR đã được hỗ trợ bởi các bộ dữ liệu mở quy mô lớn đã được phân loại rộng rãi theo cấp độ hành động, cấp độ hành vi, cấp độ tương tác và cấp độ hoạt động nhóm. Một danh sách đầy đủ các bộ dữ liệu cùng với thông tin chi tiết của chúng có thể được tham khảo trong cuộc khảo sát [3].

Yoga là một môn tập luyện các bài tập về thể chất và tinh thần có nguồn gốc từ Ấn Độ cổ đại. Mục đích mang lại [illegible] cho cơ thể, tâm trí và linh hồn. Trong những thập kỷ qua, Yoga đã trở nên phổ biến rộng rãi trên toàn cầu như một [illegible] nghệ thuật và khoa học về [illegible] lành mạnh. Liên Hợp Quốc đã tuyên bố [illegible] 1 tháng 6 là 'Ngày [illegible]'. Một số nhà nghiên cứu đã nghiên cứu [illegible] giá các lợi ích y tế của tập Yoga [illegible] 4, 37, 45, 46]. Ba [illegible] nghiên cứu về thời gian diễn ra đại dịch COVID-19 gần đây [38] chỉ [illegible] người tập Yoga trong thời gian cách ly xã hội đã trải qua mức độ căng thẳng, lo lắng và trầm cảm thấp hơn.



Hình 1: Một cái nhìn đơn giản về kiến trúc 2 giai đoạn của Bộ phân loại Yogasana với bộ trích xuất đặc trưng và bộ phân loại.

'Asana' là bản dịch tiếng Phạn của tư thế. Để nhận được lợi ích tối đa từ Yogasana và để ngăn chặn bất kỳ tác dụng phụ và chấn thương nào trong quá trình luyện tập, người ta cần thực hiện chúng một cách chính xác. Điều này làm cho vấn đề nhận biết và điều chỉnh tư thế Yoga trở thành một vấn đề quan trọng. Điều này yêu cầu quyền truy cập vào một huấn luyện viên Yoga, có thể không phổ biến rộng rãi. Do đó, việc sử dụng huấn luyện viên Yoga dựa trên ML được đề xuất. Tuy nhiên, Yogasana liên quan đến các tư thế cơ thể phức tạp không được thấy trong các hoạt động chung. Do đó, các bộ dữ liệu huấn luyện hiện có dành cho HAR có khả năng thất bại trong việc ước tính chính xác các tư thế Yoga.

Một số nghiên cứu đã tạo bộ dữ liệu tùy chỉnh để ước tính tư thế yoga [7, 8, 15, 20, 33, 35, 43, 44, 48]. YogaNet [33] dựa trên bản đồ dịch chuyển góc khớp 3D và thu thập thông tin 3D bằng cách sử dụng các hệ thống chụp chuyển động 3D phức tạp, trong khi phần còn lại sử dụng bộ dữ liệu 2D ở dạng video [15, 20, 43 , 48] hoặc hình ảnh [7, 8 , 35, 44].

Chỉ một số bộ dữ liệu được liệt kê đã được cung cấp công khai [20, 35, 44, 48]. Yoga-82 [44] là một bộ dữ liệu quy mô lớn được tạo ra từ 28,4 hình ảnh của những người thực hiện Yogasana trong tự nhiên, được thu thập từ nhiều nguồn internet khác nhau cùng với các hình ảnh đó. Những asana này đã được phân loại thành 82 tư thế. Yoga-82 là cơ sở dữ liệu thách thức nhất

, ,

Bảng 1: Tóm tắt các công việc liên quan theo trình tự thời gian

Công việc	Asana Đối tượng Góc			Chìa khóa điểm	Mã số Mở	tập dữ liệu Mở	phương thức [1]	Khả năng của Rò rỉ mục tiêu	Chuyển khoản Học hỏi
Hồi giáo và cộng sự. [19]	-	5	1	20	KHÔNG	KHÔNG	khung	Đúng	KHÔNG
Maddala và cộng sự. [33]	42	10	1	25	KHÔNG	KHÔNG	khung	Đúng	KHÔNG
Yadav và cộng sự. [48]	6	12	1	18	Đúng	Đúng	khung	Đúng	Đúng
Verma et al. [44]	82	-	-	-	KHÔNG	Đúng	hình ảnh	KHÔNG	KHÔNG
Jain et al. [20]	10	27	1	-	KHÔNG	Đúng	video	KHÔNG	KHÔNG
Gupta và cộng sự. [15]	1	20	1	-	KHÔNG	KHÔNG	khung	Đúng	KHÔNG
Của chúng tôi	20	51	4	136	Đúng	Đúng	khung	KHÔNG	Đúng

[1] Phương thức có thể được hiểu là (1) khung hình - trích xuất khung hình đầu vào từ video, (2) hình ảnh - hình ảnh tĩnh không được lấy từ bất kỳ video, (3) video - sử dụng đoạn video hoàn chỉnh làm đầu vào. Lưu ý rằng chế độ khung với sự phân tách ngẫu nhiên sẽ dẫn đến rò rỉ mục tiêu.

có sẵn cho đến nay để phân loại Yogasana. Các bộ dữ liệu khác đã được mở có kích thước nhỏ hơn, do đó không thể được sử dụng để đào tạo mô hình Deep Learning. Đối với những trường hợp như vậy Chuyển Học tập (TL) chứng tỏ là một kỹ thuật hiệu quả và có khả năng để cho hiệu suất tốt hơn. Các mô hình sử dụng TL dự kiến sẽ khái quát hóa tốt so với những cái đang sử dụng hạn chế bộ dữ liệu.

Ước tính tư thế con người là một điểm thu hút chính trong thị giác máy tính và đã là một lĩnh vực nghiên cứu tích cực trong vài thập kỷ qua. Gần đây, một số phương pháp mới đã được đề xuất có thể được phân loại rộng rãi thành hai loại - (i) cách tiếp cận từ trên xuống và (ii) cách tiếp cận từ dưới lên. Trong cách tiếp cận đầu tiên, các kiến trúc đầu tiên dự đoán một hộp giới hạn xung quanh con người và những giới hạn này các hộp được xử lý riêng để dự đoán tọa độ điểm chốt chung xuất hiện trên con người. Các công trình nổi bật theo cách tiếp cận này là AlphaPose [10], Simple Baseline [47], HRNet [42], DARK [50] và Tư thế DC [31].

Trong cách tiếp cận thứ hai, các điểm chính được dự đoán đầu tiên mà không cần được giao cho một người cụ thể (trong số những người khác) trong hình ảnh. Thay vào đó, sau khi dự đoán các điểm chính, chúng được nhóm lại cùng nhau và được gán cho những người khác nhau trong hình ảnh. Các công trình nổi bật theo phương pháp này bao gồm OpenPose [6], MultiPoseNet [23], HigherHRNet [9], PifPaf [24], ĐƠN GIẢN [51]. Vì dữ liệu yoga là hạn chế về kích thước, học tập chuyển giao từ các mô hình này được đào tạo trên quy mô lớn bộ dữ liệu ước tính tư thế công khai quy mô có thể được sử dụng để cải thiện hiệu suất ước tính tư thế yoga.

Trong bài báo này, chúng tôi sử dụng một ước tính tư thế mạnh mẽ và hiệu quả mô hình, AlphaPose [10], để ước tính tư thế của người biểu diễn Yoga, theo sau là một khu rừng ngẫu nhiên đơn giản để phân loại yoga được hình thành. Chúng tôi cũng thử nghiệm các phương pháp gần đây hơn DCPose [31] và KAPAO [34] trên bộ dữ liệu Yoga-82 [44] để so sánh tác động của các phương pháp ước tính tư thế khác nhau đối với tổng thể của chúng tôi. đường ống dẫn.

Lựa chọn mô hình ước tính tư thế ảnh hưởng đến giai đoạn đầu tiên của đường ống và hiệu suất tốt ở đây lan truyền đến tổng thể tốt hiệu suất của đường ống. Thiết kế đường ống của chúng tôi cho phép lựa chọn của các mô hình ước tính tư thế và có thể thúc đẩy sự phát triển trong miền này.

Chúng tôi cũng thu thập một bộ dữ liệu nội bộ mở rộng, một tập hợp con trong số đó đã được sử dụng để đánh giá hiệu suất của mô hình của chúng tôi bằng cách sử dụng một chiến lược đánh giá ba giai đoạn bao gồm đánh giá trên 1) khung không nhìn thấy của Yogasana, 2) đối tượng không nhìn thấy, và 3) không nhìn thấy góc độ. Tóm lại, những đóng góp chính của công việc này là ba nếp gấp:-

- Một bộ dữ liệu mở rộng nắm bắt rõ ràng các tư thế yoga từ bốn góc máy ảnh cho mỗi trong số 51 đối tượng thực hiện 20 asana. Để thúc đẩy nghiên cứu trong tương lai, chúng tôi tạo tập dữ liệu và mã các điểm chính của cơ thể được suy luận Alpha Pose [10] một cách công khai có sẵn.
- Khung phân loại khung nhìn độc lập, trực giao để lựa chọn các thuật toán ước tính hoặc phân loại tư thế được sử dụng, sử dụng một chiến lược đánh giá ba giai đoạn để cung cấp một ước tính chính xác hơn về khả năng khái quát hóa của một người mẫu.
- Một quy trình đơn giản nhưng hiệu quả liên quan đến tính toán ít hơn phức tạp và suy luận thời gian thực cùng với tính cạnh tranh hiệu suất được đánh giá toàn diện trên các bộ dữ liệu hiện có.

Trong phần còn lại của bài báo, trước tiên chúng ta thảo luận về các công việc liên quan trong Phần 2. Trong Phần 3, chúng tôi thảo luận về phương pháp thu thập dữ liệu của chúng tôi, hai kiến trúc sân khấu để ước tính yogasana, khả năng của mục tiêu rò rỉ và các kỹ thuật đánh giá mới của chúng tôi được thiết kế để loại bỏ mục tiêu rò rỉ. Kết quả thử nghiệm của chúng tôi và ba bộ dữ liệu[2] có sẵn công khai có thể được tìm thấy trong Phần 4, sau đó là phần kết luận nhận xét và phạm vi tương lai trong Phần 6.

2 CÔNG VIỆC LIÊN QUAN

Ước tính tư thế con người, là một trong những vấn đề phổ biến nhất trong thị giác máy tính, có nhiều bộ dữ liệu điểm chuẩn quy mô lớn và có sẵn công khai như COCO [29], Halpe [10, 28], MPII [2], CrowdPose [27] và HiEve [30]. Bộ dữ liệu Microsoft COCO [29] là một trong những bộ dữ liệu quan trọng nhất bộ dữ liệu phổ biến trong ước tính tư thế con người. Nó bao gồm 200.000 hình ảnh với chú thích tư thế của 17 điểm chính chung. bộ dữ liệu MPII [2] chứa khoảng 29.000 hình ảnh con người thực hiện các hoạt động khác nhau được chụp từ các góc độ khác nhau. Tập dữ liệu này được chú thích với 15 điểm chốt liên kết cơ thể cùng với cờ hiển thị của chúng. TRONG CrowdPose [27], bộ dữ liệu chứa khoảng 20.000 hình ảnh với khoảng tổng cộng 80.000 người trải dài trên tất cả những hình ảnh này.

[2] Hình ảnh trong tập dữ liệu [35] rất giống với hình ảnh trong Yoga-82 [44], nhưng [35] tương đối kích thước nhỏ hơn. Vì vậy, chúng tôi thử nghiệm trên Yoga-82 [44] và không khám phá [35]

Hình ảnh của bộ dữ liệu này được lấy mẫu từ ba bộ dữ liệu hiện có dựa trên một số liệu có tên là Chỉ số đám đông. Bộ dữ liệu HiEve [30] là bộ dữ liệu tư thế con người lớn nhất với tổng số >1 triệu tư thế trải rộng trên 31 video. Tập dữ liệu này đặc biệt tập trung vào các sự kiện phức tạp và đông đúc như lối vào/lối ra ở tàu điện ngầm, va chạm, đánh nhau, v.v. Trong trường hợp không có chuẩn tư thế Yoga, hiệu suất trên các bộ dữ liệu này có thể được coi là đại diện cho hiệu suất của các phương pháp ước tính tư thế khác nhau về các tư thế yoga.

Theo cách tiếp cận từ trên xuống như được mô tả trong Phần 1, các tác giả của Đường cơ sở đơn giản [47] đề xuất một quy trình đơn giản nhưng hiệu quả để cung cấp đường cơ sở vững chắc cho các phương pháp ước lượng đặt ra. Kiến trúc của họ dựa trên ResNet [16] theo sau là một vài lớp tích chập. Với kiến trúc đơn giản này, họ đạt được 73,7 mAP trên bộ dữ liệu COCO và 74,6 mAP và 57,8 điểm MOTA3 trên bộ dữ liệu PoseTrack [1]. Trong HRNet [42], các tác giả đề xuất một kiến trúc sâu với các mạng con có độ phân giải cao đến thấp song song với việc trao đổi thông tin lặp đi lặp lại qua các mạng con đa độ phân giải. Họ đạt được 75,5 mAP trên bộ dữ liệu COCO và 74,9 mAP và 57,9 MOTA trên bộ dữ liệu PoseTrack. Hơi trực giao với các phương pháp này, DARK [50] nghiên cứu biểu diễn tọa độ điểm chính trong kiến trúc ước lượng tư thế con người. Họ đề xuất một quy trình giải mã dựa trên mở rộng Taylor hiệu quả của sơ đồ nhiệt chung được dự đoán tới tọa độ điểm chính trong không gian hình ảnh gốc và sơ đồ mã hóa tọa độ trung tâm pixel phụ không thiên vị. Với đường trục HRNet-W48 [42] , DARK [50] đạt được 76,2 mAP trên bộ dữ liệu COCO. Theo một hướng khác, Lite-HRNet [49] đề xuất một mạng có độ phân giải cao nhẹ, tập trung vào việc giảm chi phí tính toán trong khi không làm giảm đáng kể hiệu suất . Họ áp dụng khối xáo trộn từ ShuffleNet [32] đến HRNet [42] để hiển thị mức tăng hiệu suất. Sau đó, họ thay thế các tổ hợp điểm (1 × 1) chuyên sâu về tính toán trong các khối phát ngẫu nhiên bằng trọng số kênh có điều kiện, trong đó trọng số được học trên tất cả các kênh qua nhiều độ phân giải. Họ đạt được 69,7 mAP trên bộ dữ liệu COCO và 87,0 PCKh4 trên bộ dữ liệu MPII.

Có một số phương pháp dựa trên cách tiếp cận từ dưới lên vì nó giảm chi phí tính toán trong ước tính tư thế nhiều người vì mô hình không cần xử lý riêng từng người trong ảnh. OpenPose [6] đề xuất các trường ái lực bộ phận (PAF) là phương pháp biểu thị mối quan hệ theo cặp phi cấu trúc giữa các bộ phận cơ thể của những người khác nhau trong ảnh. Họ đạt được 61,8 mAP trên tập dữ liệu COCO bằng phương pháp vanilla trong khi mô hình chân+cơ thể đạt được 65,3 mAP trên tập dữ liệu COCO. Trong MultiPoseNet [23], các tác giả thiết kế một kiến trúc sâu bao gồm một xương sống dùng chung của trình trích xuất tính năng, sau đó được cung cấp thành hai mạng con song song—một là mạng con phát hiện/phân đoạn người và một là mạng con phát hiện điểm chính. Đầu ra của hai mạng con này sau đó được đưa vào một mạng gọi là Pose Residual

Mạng gán các điểm chính cho những người được phát hiện. Kiến trúc này đạt được 69,6 mAP trên bộ dữ liệu COCO. HigherHRNet [9] xử lý sự thay đổi tỷ lệ của con người trong ảnh. Nó tạo ra kim tự tháp tính năng có độ phân giải cao với sự giám sát đa độ phân giải trong quá trình đào tạo và tổng hợp bản đồ nhiệt đa độ phân giải trong quá trình suy luận. Đường ống đặc biệt nhắm đến con người nhỏ bé trong hình ảnh và cảnh đông đúc. Nó đạt được 70,5 mAP trên bộ dữ liệu COCO. PifPaf [24] dựa vào các trường cường độ bộ phận (pif) và trường liên kết bộ phận (paf) để định vị các bộ phận cơ thể và liên kết các bộ phận cơ thể với nhau để tạo thành đầy đủ các tư thế của con người. Họ cũng sử dụng mất mát Laplace cho hồi quy để mã hóa sự không chắc chắn. Phương pháp này đạt được 66,7 mAP trên tập dữ liệu COCO. Trong SIMPLE [51], các tác giả nhằm mục đích thu hẹp khoảng cách về hiệu suất xét về độ chính xác giữa phương pháp tiếp cận từ trên xuống và phương pháp tiếp cận từ dưới lên. Quy trình này sử dụng mô phỏng các bản đồ nhiệt ước tính của phương pháp tiếp cận hiệu suất cao từ trên xuống để chuyển kiến thức về các tính năng cấp cao từ mô hình từ trên xuống sang mô hình từ dưới lên. Các mô-đun ước tính tư thế và phát hiện con người chia sẻ cùng một xương sống và được hợp nhất bằng cách coi các vấn đề là các vấn đề học điểm để cả hai nhiệm vụ có thể mang lại lợi ích cho nhau. Nó đạt được 71,1 mAP trên bộ dữ liệu COCO và 69,5 mAP và 55,7 MOTA trên bộ dữ liệu PoseTrack. DEKR [12] là một quy trình đơn giản nhưng hiệu quả, sử dụng các cấu trúc tích chập thích ứng và cấu trúc nhiều nhánh để tham gia vào các vùng pixel khác nhau có liên quan cho các điểm chính khác nhau.

Các tác giả lập luận rằng để tìm hiểu các điểm chính bằng hồi quy, mô hình cần tập trung vào các vùng điểm chính. Điều này đạt được nhờ phần mở rộng pixel-khôn ngoan của mạng biến áp không gian kích hoạt các pixel gần một điểm chính cho phép mô hình tìm hiểu các biểu hiện đại diện phong phú từ các pixel được kích hoạt này. Hơn nữa, cấu trúc đa nhánh giúp mô hình tập trung vào các pixel có liên quan đến từng điểm chính một cách riêng biệt, do đó học cách biểu diễn không bị rối.
Phương pháp này đạt được 71,0 mAP trên bộ dữ liệu COCO.

Một mô hình khác trong ước tính tư thế là mô hình của các phương pháp dựa trên hồi quy, trong đó tọa độ điểm chính được xử lý trực tiếp dưới dạng tar get và mô hình được tạo để tìm hiểu ánh xạ dựa trên hồi quy tới tọa độ pixel. Mặc dù các phương pháp này ít tốn kém hơn về mặt tính toán so với các phương pháp dựa trên bản đồ nhiệt, nhưng hiệu suất của chúng lại thấp hơn. Điều này là do các mô hình này không kết hợp được thông tin theo ngữ cảnh xung quanh điểm chính và cũng không thể nắm bắt được sự không chắc chắn cố hữu trong chú thích điểm chính, đặc biệt là trong các trường hợp tắc và mờ chuyển động. Một phương pháp đáng chú ý trong lĩnh vực này là Ước tính khả năng ghi nhật ký còn lại [26] thực sự đạt được hiệu suất cao hơn so với SOTA trên bộ dữ liệu COCO. Trong [26], các tác giả đề xuất một mô hình hồi quy mới và hiệu quả với thiết kế tham số hóa lại và Ước tính khả năng ghi nhật ký dư (RLE), thay vì tìm hiểu phân phối thực tế của tọa độ điểm chính, cố gắng tìm hiểu sự thay đổi của phân phối từ phân phối giả định. Với đường trục HRNet-w48 và RLE, các tác giả đạt được 75,7 mAP trên tập dữ liệu COCO. Theo cách tương tự, các tác giả của [34] đề xuất một phương pháp không có bản đồ nhiệt mà họ đặt tên là KAPAO (Điểm chính và đặt dưới dạng đối tượng), trong đó các điểm chính riêng lẻ và tập hợp các điểm chính (tư thế) liên quan được mô hình hóa dưới dạng các đối tượng trong một giai đoạn dày đặc. khung phát hiện dựa trên neo. KAPAO giải quyết vấn đề ước tính tư thế con người nhiều người trong một giai đoạn bằng cách phát hiện đồng thời tư thế con người và các đối tượng điểm chính và kết hợp các phát hiện để khai thác điểm mạnh của cả hai biểu hiện đại diện đối tượng. KAPAO đạt được 70,3 mAP trên bộ dữ liệu COCO với đường ống nhanh hơn 1-2 bậc độ lớn. Chúng tôi sử dụng KAPAO như một phương pháp ước tính tư thế trong thử nghiệm trên bộ dữ liệu Yoga-82 [44] để quan sát tác động của các mô hình ước tính tư thế khác nhau trên hệ thống của chúng tôi.

Độ chính xác trung bình và độ chính xác theo dõi nhiều đối tượng của 3mean
4Tỷ lệ điểm chính xác dựa trên phần đầu

, ,

Khung AlphaPose [10] bao gồm ba mô-đun–(i) Mạng biến áp đối xứng không gian (SSTN) (ii) Loại bỏ tham số Pose Non Max Suppression (NMP) và (iii) Trình tạo đề xuất có hướng dẫn Pose (PGPG). SSTN được sử dụng để trích xuất khu vực một người chất lượng cao trong ảnh từ hộp giới hạn không chính xác có thể đến từ bộ phát hiện đối tượng phụ tối ưu. Parametric Pose NMS loại bỏ các tư thế dư thừa bằng cách sử dụng thước đo khoảng cách tư thế mới. PGPG được sử dụng để tăng cường dữ liệu đào tạo nhằm tạo các hộp giới hạn (dưới mức tối ưu) dựa trên tư thế đã cho được sử dụng để đào tạo SSTN. Nó đạt được kết quả cao trên điểm chuẩn MPII [2] (76,7 mAP) và bộ dữ liệu COCO (72,3 mAP) và có thể cung cấp tốc độ khung hình 23 khung hình/giây khi được cung cấp dữ liệu video. Do đó, chúng tôi chọn AlphaPose [10] làm phương pháp ước tính tư thế của mình.

DCPose [31] nhằm mục đích giải quyết vấn đề ước tính tư thế nhiều người trong dữ liệu video. Khung mã hóa bối cảnh điểm chính không gian-thời gian thành các phạm vi tìm kiếm được bản địa hóa, tính toán dư lượng tư thế và sau đó tinh chỉnh các ước tính bản đồ nhiệt điểm chính. Cụ thể, quy trình bao gồm ba mô-đun dành riêng cho nhiệm vụ–(i)
Mạng Pose Temporal Merger (PTM), mạng này thực hiện tổng hợp điểm chính trên ba khung hình liên tiếp bằng cách gộp nhóm , do đó bản địa hóa phạm vi tìm kiếm cho mạng điểm chính (ii) Pose Residual Fusion (PRF), mạng này thu được hiệu quả phần dư tư thế giữa khung hình hiện tại và các khung liền kề và (iii)

Pose Correction Network (PCN) bao gồm năm lớp chập song song với các tốc độ giãn nở khác nhau để lấy mẫu lại các bản đồ nhiệt điểm chính trong phạm vi tìm kiếm được bản địa hóa. Phương pháp này đạt được 79,2 mAP trên tập dữ liệu PoseTrack. Chúng tôi cũng thử nghiệm với DCPose trên Yoga-82 [44] để quan sát tác động của một phương pháp ước tính tư thế khác đối với quy trình của chúng tôi.

Nghiên cứu sâu rộng đã được thực hiện trong việc áp dụng ước tính và phân loại tư thế trong Yoga. Hồi giáo và cộng sự. [19] đã sử dụng các điểm chính để có được các góc khớp đã chọn. Họ đã sử dụng độ lệch của các góc này so với một tập hợp các góc tham chiếu làm độ chính xác của asana.
Mặc dù họ không tham gia vào việc phân loại asana, nhưng các thí nghiệm và kết quả của họ đã chứng minh rằng các điểm chính được phát hiện từ ước tính tư thế thực sự là các đặc điểm phù hợp với asana.

YogaNet [33] đã mở rộng công việc này bằng cách sử dụng JADM thay vì chọn các góc đặc biệt. Việc sử dụng các góc thay vì các điểm chính cho phép họ cải thiện sự bất biến về vị trí và kích thước của phương pháp ở một mức độ nào đó. Tuy nhiên, cả công việc này và công việc trước đó đều phụ thuộc rất nhiều vào các điểm chính được phát hiện bởi Microsoft Kinect [52], mà chúng tôi nhận thấy là hoạt động kém hơn so với các khung ước tính tư thế gần đây hơn như AlphaPose [10] và OpenPose [6]. Phương pháp học sâu này để trích xuất các điểm chính là một giải pháp thay thế tương đối rẻ tiền, chỉ yêu cầu hình ảnh RGB so với phương pháp dựa trên độ sâu và tia hồng ngoại mà Kinect sử dụng.

Yadav và cộng sự. [48] đã sử dụng OpenPose [6] để trích xuất điểm chính và LSTM [17] để khai thác thông tin tạm thời, xây dựng một đường dẫn đầu cuối hoàn chỉnh để phân loại từ các video yoga.
Thay vì sử dụng tọa độ điểm chính, họ đã sử dụng các tính năng trung gian do OpenPose học được và xử lý chúng bằng CNN [25] trước khi đưa chúng vào LSTM [17]. Chúng tôi cải thiện công việc này bằng cách sử dụng công cụ ước tính tư thế hoạt động tốt hơn, tập dữ liệu lớn hơn và đánh giá cụ thể hơn.

[5]Bản đồ dịch chuyển góc chung

Yoga-82[44] là một trong những bộ dữ liệu phân loại Yoga quy mô lớn đầu tiên được cung cấp công khai. Nó bao gồm 28,4K hình ảnh của những người thực hiện một trong 82 yogasana cụ thể được thu thập từ các công cụ tìm kiếm trên web. Cấu trúc phân cấp ba cấp cho nhãn, với 20 và 6 siêu lớp ở các cấp tiếp theo đã được cung cấp.
Tuy nhiên, họ đã xây dựng vấn đề của mình dưới dạng phân loại tư thế thay vì ước tính và do đó, không chứa các chú thích điểm chính.
Ngoài ra, một phần đáng kể dữ liệu của họ bao gồm các hình ảnh và biểu đồ clip art, thay vì hình ảnh thực của con người, điều này có thể dẫn đến hiệu suất kém khi sử dụng TL từ các phương pháp được đào tạo trên các ví dụ trong thế giới thực.

Jain et al. [20] đã sử dụng một kiến trúc hoàn chỉnh từ đầu đến cuối cho vấn đề này. Để tận dụng mối quan hệ không gian-thời gian một cách hiệu quả, họ đã đề xuất sử dụng CNN 3D [21]. Tuy nhiên, họ đã sử dụng một bộ dữ liệu nhỏ chỉ có 261 video.

YogaHelp [15] đã thực hiện một cách tiếp cận hơi khác và sử dụng các cảm biến chuyển động khác nhau để cung cấp phản hồi và hướng dẫn cải thiện sự cải thiện của một người thực hiện yoga. Họ đã khám phá và thử nghiệm rộng rãi một asana duy nhất cho các chủ đề khác nhau về chuyên môn khác nhau. Họ đã thiết kế các thông số khác nhau để xác định tính chính xác của asana và sử dụng chúng để đánh giá. Họ phát hiện ra rằng bằng cách sử dụng hệ thống phản hồi, những người mới bắt đầu tập yoga đã cho thấy sự cải thiện đáng kể trong khoảng thời gian ngắn bốn tuần. Phát hiện của họ chứng minh những lợi ích tổng thể của công việc này và cho thấy rằng một huấn luyện viên yoga thực sự sẽ rất hữu ích, đặc biệt là đối với những người mới bắt đầu trong lĩnh vực này.

Yadav và cộng sự. [48] đã sử dụng mô hình OpenPose [6] được đào tạo trước để trích xuất các điểm chính của chúng. Đây là tác phẩm đầu tiên sử dụng Học chuyển giao để có hiệu suất tốt hơn trong phân loại yoga. Việc sử dụng mô hình được đào tạo trước cho phép họ làm việc với khung học sâu hoạt động ngay cả với tập dữ liệu có kích thước tương đối nhỏ. Họ đã sử dụng một bộ dữ liệu yoga tùy chỉnh chỉ có 6 asana, 15 đối tượng và một góc máy quay thống nhất. Tuy nhiên, để đánh giá từng khung hình, họ đã sử dụng dữ liệu được phân tách ngẫu nhiên để xác thực, có khả năng dẫn đến rò rỉ mục tiêu.
Điều này cho phép họ đạt được độ chính xác 100% ở ba trong số sáu asana mà họ xem xét.

Một trong những mối quan tâm chính mà chúng tôi tìm thấy trong các công trình hiện có là rò rỉ mục tiêu đã được xác định chính thức trong phần 3.4.1. Nói chung, rò rỉ mục tiêu dẫn đến độ chính xác cao hơn trong quá trình thử nghiệm mặc dù khả năng khái quát hóa mô hình cơ bản có thể không so sánh được với khả năng quan sát được trên bộ thử nghiệm.

Trong công việc của mình, chúng tôi cố gắng giải quyết một số hạn chế mà chúng tôi tìm thấy trong các tác phẩm hiện có. Chúng tôi thu thập dữ liệu Yoga một cách có hệ thống hơn . Số lượng đối tượng lớn hơn và các biến thể ở các góc máy ảnh khác nhau cho phép chúng tôi khái quát hóa các mô hình của mình tốt hơn và có được những đánh giá thực tế hơn. Sử dụng chiến lược đánh giá ba cấp độ, chúng tôi giải quyết rò rỉ mục tiêu đang phổ biến ở hầu hết các công trình hiện có. Cuối cùng, chúng tôi đã khám phá kiến trúc hai giai đoạn thay vì cách tiếp cận từ đầu đến cuối. Điều này cho phép chúng tôi sử dụng Transfer Learning và tận dụng dữ liệu quy mô lớn hiện có và nghiên cứu sâu rộng dưới dạng các mô hình được đào tạo trước. Trong miền này, nơi không có bộ dữ liệu quy mô lớn , việc sử dụng TL cho phép chúng tôi cải thiện hiệu suất của mình và mở đường cho các công việc trong tương lai.

3. PHƯƠNG PHÁP LUẬN

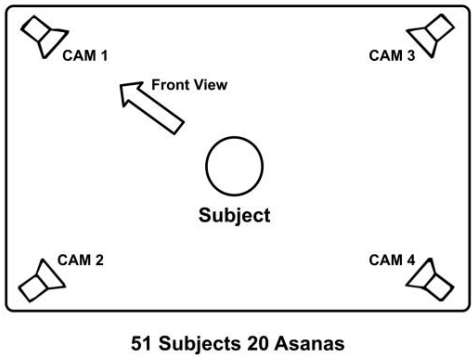
Phương pháp của chúng tôi có thể được chia thành 3 phần. Đầu tiên, chúng tôi thu thập tập dữ liệu về tư thế yoga phong phú đồng thời ghi nhớ những hạn chế và thiếu sót của tập dữ liệu mở hiện có. Thứ hai,

chúng tôi sử dụng kiến trúc 2 giai đoạn với ước tính tư thế con người để trích xuất tính năng và phân loại dựa trên cây quyết định để nhận dự đoán asana cho từng khung hình được cung cấp làm đầu vào hình ảnh. Cuối cùng, chúng tôi sử dụng chiến lược đánh giá ba cấp độ để đánh giá hiệu suất của mô hình.

3.1 Thu thập dữ liệu

Sử dụng Microsoft Kinect [52], chúng tôi thu thập dữ liệu dưới dạng video của các đối tượng thực hiện yogasana. Để khái quát hóa tốt hơn các mô hình của chúng tôi giữa các cá nhân, 51 tình nguyện viên đã được liên hệ làm đối tượng cho dữ liệu của chúng tôi. Mỗi người được yêu cầu thực hiện 20 tư thế yoga và 1 tư thế tĩnh (để biểu thị không có asana nào được thực hiện) trong một căn phòng có ánh sáng ổn định và cơ sở vật chất phù hợp. Mỗi video kéo dài khoảng 2-5 phút. Một số asana là hai bên và các đối tượng thực hiện các tư thế này hai lần, mỗi bên một lần. Hai bên được dán nhãn khác nhau[6] . Đối với các asana như vậy, các đối tượng được yêu cầu lần lượt thực hiện asana theo cả hai hướng và các dấu thời gian video cho thời gian bắt đầu và kết thúc của mỗi hướng cũng được ghi lại.

Một số asana có thể khó phân loại hơn từ hướng quay mặt về phía trước, trong khi chúng có thể dễ dàng phân loại từ một số góc độ khác. Ngoài ra, trong khi hoạt động với người dùng trong một kịch bản trong thế giới thực, bộ phân loại có thể được cung cấp hình ảnh từ nhiều hướng khác nhau. Để làm cho các mô hình của chúng tôi mạnh mẽ ở các góc nhìn khác nhau, mọi asana được thực hiện bởi từng đối tượng được ghi lại từ 4 camera khác nhau đặt ở các góc của căn phòng, như trong Hình 2.



Hình 2: Thiết lập được sử dụng để thu thập dữ liệu

Sau khi tất cả các video yoga được thu thập, công cụ ước tính tư thế Al phaPose [10] lần đầu tiên được chạy trên tất cả chúng cùng nhau, lưu trữ các điểm chính cho từng khung hình trong mỗi video. Vì chúng tôi không có chú thích thực tế cơ bản cho các điểm chính, nên chúng tôi đã sử dụng trọng số mô hình được đào tạo trước trên bộ dữ liệu Halpe Full Body [10, 28] trực tiếp cho giai đoạn đầu tiên này . Khi tất cả các video đã được suy luận, chúng tôi bắt đầu xử lý video cho giai đoạn phân loại thứ hai.

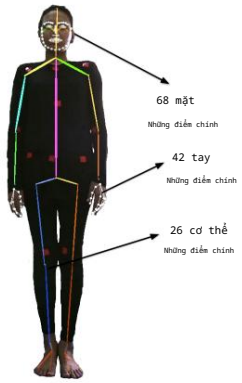
Đối với công việc này, chúng tôi xem xét suy luận theo từng khung và do đó, cần dữ liệu theo từng khung cho giai đoạn phân loại thứ hai. Trong khi thu thập dữ liệu, chúng tôi cũng ghi lại các dấu thời gian tương ứng với thời gian bắt đầu và kết thúc của asana. Từ mỗi asana, chúng tôi lấy mẫu thống nhất một số khung hình giữa thời điểm bắt đầu và kết thúc, sao cho các khung hình cách đều nhau. Chúng tôi quan sát thấy rằng các đối tượng đã một chút thời gian để vào tư thế cuối cùng của asana. Vì vậy, chúng tôi đã sử dụng các khung hình từ đầu video cho đến khi bắt đầu một asana làm khung hình 'Tĩnh', thêm một lớp bổ sung vào bộ phân loại của chúng tôi. Lớp bổ sung này đã giúp cải thiện hiệu suất tổng thể, vì nếu không thì những khung hình này sẽ bị phân loại sai thành bất kỳ asana nào. Tuy nhiên, chúng tôi ước tính khoảng 2-3% điểm dữ liệu bị dán nhãn sai khiến đối tượng bị mất thăng bằng khi thực hiện yogasana. Các điểm chính được AlphaPose phát hiện trong tất cả các khung và nhãn asana tương ứng đã tạo thành tập dữ liệu huấn luyện cho bộ phân loại của chúng tôi.

Tóm lại, video của 51 đối tượng được ghi lại một cách có hệ thống từ 4 góc camera khác nhau. Từ những video này, các khung hình trong đó asana thực tế đang được thực hiện được lấy mẫu thống nhất, với tối đa 200 khung hình trên mỗi phân đoạn video. Bạn có thể tìm thấy mô tả chi tiết hơn về số lượng đối tượng được ghi lại và các khung hình được trích xuất cho từng tư thế yoga và máy ảnh, như được sử dụng trong các phần tiếp theo, trong tài liệu bổ sung ở Phụ lục A.
Katichakrasana được thực hiện bởi số lượng đối tượng tối đa, tức là 47, và các asana khác nhau được ghi lại cho số lượng đối tượng khác nhau. Theo hiểu biết tốt nhất của chúng tôi, đây là tập dữ liệu yoga đầu tiên xem xét rõ ràng các asana từ các góc máy ảnh khác nhau và có số lượng đối tượng được ghi lại nhiều nhất.

3.2 Ước lượng tư thế

Ước tính tư thế là giai đoạn đầu tiên và khó khăn nhất trong hệ thống mô hình của chúng tôi. Một người muốn phân biệt các yogasana khác nhau sẽ chủ yếu sử dụng tư thế và tư thế của đối tượng để so sánh . Do đó, việc sử dụng ước lượng tư thế để trích xuất các đặc điểm phong phú đặc trưng cho tư thế của đối tượng là điều hiển nhiên . Đã có nghiên cứu sâu rộng trong lĩnh vực này và chúng tôi tận dụng các kiến trúc và bộ dữ liệu hiện có trong công việc của mình. Chuyển giao kiến thức từ các bộ dữ liệu quy mô lớn có sẵn công khai cho phép chúng tôi khắc phục những hạn chế về sự khan hiếm dữ liệu chú thích ở một mức độ nào đó.

AlphaPose [10] là một phương pháp phổ biến để Ước tính Pose. Nó tuân theo một khuôn khổ hai bước trong đó lần đầu tiên nó phát hiện các hộp giới hạn của con người và sau đó ước tính tư thế trong mỗi hộp một cách độc lập. AlphaPose [10] tương đối nhanh, khiến nó trở nên lý tưởng cho các tác vụ thời gian thực. Đây là hệ thống nguồn mở đầu tiên đạt được hơn 70 mAP (75 mAP) trên tập dữ liệu COCO [29] và 80+ mAP (82,1 mAP) trên tập dữ liệu MPII [2].



Hình 3: Các điểm chính được phát hiện bởi AlphPose được đào tạo trên Halpe.

[6]Các asana song phương được dán nhãn riêng là asana_left và asana_right. Asana đơn phương được dán nhãn bình thường là asana

, ,

Hình 4: Hình ảnh mẫu hiển thị các điểm chính do AlphaPose phát hiện. (a) có cùng một tư thế và cùng một chủ thể từ 4 góc máy ảnh khác nhau, (b) có cùng một chủ thể thực hiện các tư thế khác nhau trong khi (c) có các chủ thể khác nhau ở tư thế 'tĩnh'.

AlphaPose được đào tạo trước trên bộ dữ liệu Halpe Full Body [10, 28] , phát hiện 136 điểm chính trên một người. Như có thể thấy trong Hình 3, nó phát hiện 68 điểm chính trên khuôn mặt của người đó, 42 điểm chính trên cả hai tay và 26 điểm chính nằm rải rác trên phần còn lại của cơ thể. Ban đầu, chúng tôi sử dụng tọa độ của tất cả 136 điểm chính này làm vectơ đặc trưng của mình. Tuy nhiên, vì một số lượng lớn các điểm chính ở mặt và tay không liên quan lắm đến việc phân loại các tư thế yoga, nên chúng tôi nhận thấy rằng việc thay thế các bộ[7] này bằng các giá trị trung bình, tối thiểu và tối đa của 10 điểm chính có độ tin cậy cao nhất của chúng cho phép chúng tôi đạt được kết quả tốt hơn hiệu suất. Ngoài các điểm chính, AlphaPose [10] cũng phát hiện các hộp giới hạn xung quanh đối tượng. Chúng tôi nhận thấy rằng việc bao gồm tỷ lệ khung hình của các hộp giới hạn này làm tính năng cũng giúp tăng hiệu suất. Chúng tôi cũng chuẩn hóa các tọa độ điểm chính đối với hộp giới hạn, để có kích thước bất biến tốt hơn. Do đó, chúng tôi thu được một vectơ đặc trưng 71 chiều (35 điểm chính x 2 tọa độ + tỷ lệ khung hình) từ giai đoạn đầu tiên của chúng tôi và chuyển nó sang giai đoạn phân loại thứ hai.

3.3 Phân loại

Giai đoạn thứ hai này của đường ống là một nhiệm vụ phân loại chung. Các điểm chính được phát hiện bởi công cụ ước tính tư thế được sử dụng làm tính năng bởi trình phân loại. Để kiểm tra tính hợp lệ của phương pháp của chúng tôi, chúng tôi huấn luyện bộ phân loại Random Forest [4] trên dữ liệu điểm chính được suy luận. Sau đó , chúng tôi tiếp tục khám phá các phương pháp tăng cường khác như bộ phân loại ADAboost [40], Gradient Boosting [22] và Bagging [41] và cũng là tập hợp các phương pháp hoạt động tốt nhất. Sau khi chuẩn bị tập dữ liệu của các khung và các lớp tương ứng, như đã giải thích trong Phần 3.1, trước tiên chúng tôi huấn luyện bộ phân loại Rừng ngẫu nhiên cho giai đoạn phân loại thứ hai.

[7] 21 điểm quan trọng bên trái, 21 điểm quan trọng bên phải và 68 điểm quan trọng trên mặt là bộ ba điểm quan trọng được xem xét

3.4 Đánh giá

3.4.1 Rò rỉ mục tiêu. Khi chia ngẫu nhiên tập dữ liệu thành các nếp gấp huấn luyện và kiểm tra, nếu các điểm dữ liệu thay đổi liên tục, thì các điểm dữ liệu huấn luyện và kiểm tra sẽ có rất ít thay đổi giữa chúng. Về mặt hình thức, nếu chúng ta chia dữ liệu theo tỷ lệ, các điểm dữ liệu trong bất kỳ : ,một dự kiến tập hợp dấu (+) nào sẽ nằm trong tập huấn luyện, trong khi phần còn lại sẽ nằm trong tập kiểm tra. Bây giờ, nếu tập hợp con các điểm dữ liệu (+) này rất giống nhau, thì thử nghiệm không gây khó khăn gì cho mô hình vì nó đã ghi nhớ các điểm dữ liệu tương tự như các điểm dữ liệu trong tập hợp thử nghiệm. Do đó, mô hình được quan sát là hoạt động rất tốt, đôi khi thậm chí với độ chính xác 100%. Điều này đặc biệt phổ biến trong dữ liệu chuỗi thời gian và được gọi chính thức là Rò rỉ mục tiêu.

Trong trường hợp phân loại tư thế Yoga, có thể rò rỉ mục tiêu nếu các khung hình từ một video được chia thành tập luyện và thử nghiệm. Trong hầu hết các tác phẩm trước đây sử dụng dữ liệu video, quá trình phân chia thử nghiệm và đào tạo là ngẫu nhiên và do đó, việc rò rỉ mục tiêu là không thể tránh khỏi. Chúng tôi cũng sử dụng dữ liệu video và do đó, rò rỉ mục tiêu là mối quan tâm chính đối với chúng tôi. Để loại bỏ khả năng rò rỉ mục tiêu, chúng tôi phải đảm bảo rằng dữ liệu đào tạo và kiểm tra đến từ các video khác nhau. Chúng tôi thử nghiệm hai phương pháp như vậy ngoài việc đánh giá theo khung như được mô tả bên dưới.

3.4.2 Đánh giá ba giai đoạn. Đầu tiên, chúng tôi xem xét chiến lược đánh giá thông thường của thử nghiệm theo khung, trong đó dữ liệu được chạy được chia thành các tập huấn luyện và thử nghiệm.

Thứ hai, chúng tôi xem xét việc phân chia dữ liệu theo chủ đề. Vì mỗi video trong dữ liệu của chúng tôi chỉ ghi lại một đối tượng duy nhất nên chiến lược này sẽ đảm bảo rằng các mẫu đào tạo và thử nghiệm không bao giờ liền kề nhau và do đó, loại bỏ khả năng rò rỉ mục tiêu. Ngoài ra, nó cũng sẽ cho phép chúng tôi kiểm tra tính khái quát của các mô hình của chúng tôi trên các đối tượng không nhìn thấy được.

Cuối cùng, chúng tôi xem xét việc phân chia máy ảnh dữ liệu một cách khôn ngoan. Một lần nữa, vì mỗi video chỉ có dữ liệu từ một camera, nên việc rò rỉ mục tiêu sẽ không xảy ra nếu dữ liệu được chia nhỏ theo camera. Hơn nữa, chiến lược này

Bảng 2: Kết quả theo khung trong xác thực chéo gấp 10 lần. Cái này tương đương với việc huấn luyện & kiểm tra trên cả 4 camera trong Bảng 4

Tư thế ID Yoga	Thu hồi chính xác Điểm F1
0 Garudasana còn lại	98,19% 99,22% 98,70%
1 Garudasana đúng	99,57% 99,92% 99,74%
2 Gorakshasana	99,88% 99,98% 99,93%
3 Katichkrasana	100,00% 99,87% 99,93%
4 Natavasana còn lại	99,93% 99,88% 99,91%
5 Natavarasana đúng	99,75% 99,92% 99,83%
6 Pranamasana còn lại	99,87% 99,83% 99,85%
7 Pranamasana đúng	99,83% 99,98% 99,91%
8 Tadasana	99,82% 99,88% 99,85%
9 Vrikshasana còn lại	100,00% 99,92% 99,96%
10 Vrikshasana đúng	99,83% 98,32% 99,07%
11 Vẫn	99,72% 99,65% 99,68%
Trung bình	99,70% 99,70% 99,70%

sẽ cho phép chúng tôi kiểm tra tính khái quát của mô hình của chúng tôi trên định hướng máy ảnh không nhìn thấy. Vị trí của các điểm chính sẽ rất khác nhau đối với dữ liệu được chụp từ các góc độ khác nhau và tốt hiệu suất trên các máy ảnh sẽ là thước đo cụ thể hơn về hiệu suất thực tế của mô hình cho các mẫu trong tự nhiên.

3.5 Chi tiết triển khai

Chúng tôi sử dụng một GPU RTX 5000 duy nhất với bộ nhớ 16 GB cho hầu hết thí nghiệm của chúng tôi. Chúng tôi sử dụng tạ được tập trên toàn thân Halpe [28], PoseTrack[1] và COCO [29] cho Alphapose, DCPose và KAPAO tương ứng, với các mô hình phát hiện 1368 , 17 và 17 điểm chính tương ứng. Chúng tôi đã sử dụng các siêu đường kính được đề xuất trong chính bài báo phản hồi và các chi tiết khác về chúng phương pháp có thể được tìm thấy trong công việc của họ. Đối với trình xác định lớp giai đoạn hai của chúng tôi , chúng tôi sử dụng các phần mềm tự triển khai do thư viện cung cấp sklearning. Đối với các thử nghiệm trên bộ dữ liệu Yoga-82, chúng tôi sử dụng một nhóm gồm Tăng cường độ dốc biểu đồ, LightGBM và Rừng ngẫu nhiên, trong khi đối với các thử nghiệm còn lại, chúng tôi sử dụng ngẫu nhiên tiêu chuẩn Rừng. Chúng tôi sử dụng tiêu chí Gini với tổng số 500 cây và cho phép cây để phát triển cho đến khi tất cả các lá đều sạch. Thông tin chi tiết khác về thiết lập và chạy các thử nghiệm có thể được tìm thấy trong kho lưu trữ mã của chúng tôi.

4 KẾT QUẢ THỰC NGHIỆM

Mặc dù chúng tôi đã ghi lại dữ liệu cho 20 asana, chúng tôi chỉ sử dụng một tập hợp con của những điều này để đánh giá của chúng tôi. Chúng tôi sử dụng 72k khung hình được trích xuất kéo dài 11 asana và một lớp tĩnh, tổng cộng là 12 lớp. 12 này các lớp bao gồm asana trái và phải cho song phương. Mỗi tư thế có tổng cộng 6000 mẫu, dẫn đến một bộ dữ liệu rất cân bằng. Đối với lớp học tĩnh, chúng tôi lấy tất cả các khung hình trong video trước khi bắt đầu của asana thực tế, với thời gian đệm là 1 giây. Chúng tôi nhận ra rằng điều này bộ đệm có thể không đủ và do đó, một phần dữ liệu của chúng tôi có thể bị dán nhãn sai. Thêm chi tiết về kích thước của tập dữ liệu của chúng tôi và bạn có thể tìm thấy các bản phân phối thông minh của máy ảnh và chủ thể trong tài liệu bổ sung Phụ lục A.

[¨] 136 điểm chính sau đó giảm xuống còn 35, như được mô tả trong Phương pháp luận

Bảng 3: Kết quả theo chủ đề trong xác thực chéo gấp 10 lần

Tư thế ID Yoga	Thu hồi chính xác Điểm F1
0 Garudasana còn lại	88,32% 99,02% 93,36 %
1 Garudasana đúng	96,32% 99,58% 97,93%
2 Gorakshasana	99,85% 98,08% 98,96%
3 Katichkrasana	99,97% 99,53% 99,75%
4 Natavasana còn lại	99,97% 97,47% 98,70%
5 Natavarasana đúng	98,41% 99,22% 98,81%
6 Pranamasana còn lại	99,61% 94,38% 96,93%
7 Pranamasana đúng	98,93% 98,40% 98,66%
8 Tadasana	98,16% 98,78% 98,47%
9 Vrikshasana còn lại	100,0% 99,87% 99,93%
10 Vrikshasana đúng	99,00% 93,70% 96,28%
11 Vẫn	98,69% 97,72 % 98,20%
Trung bình	98,10% 97,97% 97,99%

4.1 Đánh giá theo khung

Đánh giá theo khung là chiến lược đánh giá tiêu chuẩn được sử dụng bởi hầu hết các công trình hiện có. Chúng tôi đánh giá mô hình của mình bằng vanilla 10-gấp xác nhận chéo. Các kết quả theo lớp có thể được nhìn thấy trong Bảng 2. Trình phân loại rừng ngẫu nhiên thu được độ chính xác trung bình, thu hồi và F1 điểm 99,70% mỗi. Hiệu suất cao như vậy là do giải thích vấn đề rò rỉ mục tiêu (Phần 3.4.1) có thể gây hiểu lầm và có thể cũng được chứng kiến trong các nghiên cứu khác báo cáo độ chính xác gần như hoàn hảo cho phân loại của họ.

4.2 Đánh giá theo chủ đề

Bộ dữ liệu của chúng tôi có 51 đối tượng thực hiện các tư thế yoga khác nhau. Các các đối tượng có thể hình, lứa tuổi và giới tính khác nhau. bất kể đặc điểm của đối tượng, tư thế họ thực hiện trong một asana nên giống nhau. Do đó, bộ phân loại sẽ có thể khái quát hóa xuyên suốt các môn học. Để kiểm tra điều này, chúng tôi tạo các nếp gấp trong đó tập dữ liệu được chia đối tượng khôn ngoan thành tỷ lệ 9: 1. Sau khi đào tạo bộ phân loại của chúng tôi trên bộ môn đầu tiên, chúng tôi kiểm tra các môn còn lại. Các môn thi này sẽ không được mô hình nhìn thấy, và do đó, chiến lược này ngăn chặn mô hình để chỉ cần ghi nhớ các đối tượng đào tạo, và loại bỏ khả năng rò rỉ mục tiêu (Mục 3.4.1).

Chúng tôi tạo 10 nếp gấp như vậy, với các đối tượng được chọn ngẫu nhiên, trong luân phiên, sao cho mỗi đối tượng được kiểm tra chính xác một lần. Lớp kết quả khôn ngoan của những thí nghiệm này có thể được tìm thấy trong Bảng 3. Như có thể được nhìn thấy, khu rừng ngẫu nhiên hoạt động tốt ngay cả trên các đối tượng, với độ chính xác và thu hồi trung bình là 98,10% và 97,97%, dẫn đến điểm F1 là 97,99%. So với kết quả thông minh về khung hình, những kết quả này kết quả luôn thấp hơn khoảng 1-2% cho cả ba chỉ số mà chúng tôi sử dụng để đánh giá.

4.3 Đánh giá thông minh về máy ảnh

Chúng tôi hy vọng việc đánh giá thông minh bằng máy ảnh sẽ khó khăn hơn so với hai phương pháp trước đó. Thứ nhất, nhiều asana có thể rất khó thực hiện. xác định từ quan điểm bên hoặc phía sau. Thứ hai, cùng một asana, được xem từ các máy ảnh khác nhau sẽ có tọa độ điểm chính rất khác nhau và việc khái quát hóa phạm vi biến thể rộng như vậy là thách thức. Cuối cùng, một số góc độ sẽ có độ che phủ cao hơn nhiều

, ,

Bảng 4: Kết quả thông minh về máy ảnh với số lượng máy ảnh được đào tạo khác nhau

Tư thế ID Yoga	Đào tạo về 3 Camera Thu hồi chính xác Điểm F1	Đào tạo trên 2 Camera Thu hồi chính xác Điểm F1	Đào tạo trên 1 Camera Thu hồi chính xác Điểm F1
0 Garudasana còn lại	78,48% 81,10% 79,77%	75,88% 78,64% 77,23%	59,22% 76,63% 64,37%
1 Garudasana đúng	83,46% 87,79% 85,53%	71,04% 82,88% 76,28%	67,18% 79,78% 72,36%
2 Gorakshasana	77,97% 68,84% 72,62%	73,86% 77,03% 75,30%	74,59% 57,68% 64,90%
3 Katichkrasana	80,22% 71,15% 75,39%	61,16% 77,56% 67,11%	73,18% 64,48% 64,45%
4 Natavasana còn lại	99,94% 99,48% 99,71%	99,66% 99,47% 99,57%	96,80% 99,42% 98,05%
5 Natavarasana phải	68,86% 47,00% 55,18%	6 Pranamasana trái 64,49% 46,01% 53,25%	54,65% 37,16% 43,73%
	60,10% 82,26% 69,36%	52,30% 63,28% 57,15%	37,31% 50,98% 42,95%
7 Pranamasana đúng	77,76% 89,18% 83,05%	74,94% 73,57% 74,03%	55,02% 55,78% 55,29%
8 Tadasana	98,07% 96,51% 97,28%	96,67% 90,37% 93,34%	78,39% 85,73% 79,82%
9 Vrikshasana còn lại	95,02% 93,03% 93,98%	96,16% 90,29% 93,08%	94,26% 78,43% 84,61%
10 Vrikshasana đúng	76,66% 74,42% 75,44%	69,92% 59,92% 64,17%	68,84% 44,72% 51,14%
11 Vẫn	86,70% 85,66% 85,98%	92,39% 70,50% 78,95%	76,08% 65,67% 69,53%
Trung bình	81,94% 81,37% 81,11%	77,37% 75,79% 75,79%	69,63% 66,37% 65,93%

hơn những cái khác, dẫn đến hiệu suất ước tính tư thế cũng kém hơn.
Hiệu suất kém trong giai đoạn đầu tiên được truyền sang giai đoạn thứ hai
giai đoạn phân loại là tốt.

Các kết quả có thể được nhìn thấy trong Bảng 4. Các kết quả nhất quán thấp hơn so với những phương pháp thu được từ các phương pháp trước đó, với độ chính xác trung bình và khả năng thu hồi chỉ lần lượt là 81,94% và 81,37% trong trường hợp đào tạo trên 3 góc máy ảnh. Chúng tôi cũng quan sát thấy rằng thay đổi số góc được sử dụng trong đào tạo ảnh hưởng trực tiếp đến hiệu suất. Một xu hướng giảm rõ ràng có thể được nhìn thấy trong các màn trình diễn của đào tạo trên 3, 2 và 1 máy ảnh trong Bảng 4. Mẫu này cho thấy rằng bao gồm nhiều góc máy ảnh hơn trong khi đào tạo có xu hướng để cho kết quả tốt hơn. Bao gồm tất cả bốn góc máy ảnh sẽ là tương đương với hai chiến lược đánh giá trước đó, cả hai đều đã cho kết quả tốt hơn đáng kể.

Chúng tôi thực hiện thử nghiệm ba cấp độ tương tự với các điểm chính được phát hiện bởi Microsoft Kinect [52] cũng vậy. Sử dụng các điểm chính này, giá trị trung bình của F1 điểm số cho khung hình, chủ đề và máy ảnh khôn ngoan là 69,49%, lần lượt là 57,91% và 35,21%. Đây là thấp hơn đáng kể so với điểm số tương ứng mà chúng tôi thu được bằng cách sử dụng các điểm chính của AlphaPose [10] , với cùng bộ phân loại giai đoạn thứ hai. Vì vậy, một tư thế tốt công cụ ước tính là điều cần thiết để có hiệu suất tốt trong phương pháp của chúng tôi.

4.4 Thí nghiệm trên Yadav et al. tập dữ liệu

Yadav và cộng sự. [48] đã thu thập một bộ dữ liệu nội bộ bao gồm 88 video, với 15 đối tượng được ghi lại trong 6 asana. Chúng tôi nghi ngờ rằng đánh giá theo khung của họ có thể có rò rỉ mục tiêu và đánh giá mô hình của chúng tôi trên dữ liệu của họ cả về khung và chủ đề.

Sử dụng tất cả các khung được trích xuất từ các video cùng nhau, các tác giả đã có thể đạt được độ chính xác 99,04% trên bộ thử nghiệm của họ. BẰNG có thể được nhìn thấy trong Bảng 5, chúng tôi thu được kết quả tương tự khi đào tạo trên chỉ 200 khung hình được trích xuất thống nhất từ mỗi video. Tuy nhiên, đi qua các khung được trích xuất, chúng tôi quan sát thấy rằng một số khung đang có chủ đề trong quá trình chuyển đổi, và do đó rất có thể sẽ là dán nhãn sai. Chúng tôi ước tính rằng khoảng 3,5% khung hình được trích xuất của chúng tôi đã bị dán nhãn sai. Với suy nghĩ này, độ chính xác lớn hơn 96,5% chỉ ra rõ ràng rằng mô hình quá phù hợp và có thể có được mục tiêu rò rỉ.

Bảng 5: Kết quả theo khung của Yadav et al. [48] bộ dữ liệu

Tư thế ID Yoga	Thu hồi chính xác Điểm F1
0 Bhujangasana	98,85% 99,44% 99,14%
1 Padamasana	98,38% 99,79% 99,08%
2 Shavasana	99,59% 98,29% 98,93%
3 tư thế	99,87% 99,13% 99,50%
4 Trikonasana	99,81% 99,58% 99,69%
5 Vrikshasana	99,37% 99,40% 99,38%
Trung bình (của chúng tôi)	99,31% 99,27% 99,29%
Trung bình (Yadav và cộng sự [48])	98,97% 99,11% 99,03%

Bảng 6: Kết quả theo chủ đề của Yadav et al. [48] bộ dữ liệu

Tư thế ID Yoga	Thu hồi chính xác Điểm F1
0 Bhujangasana	94,41% 95,83% 95,12%
1 Tư thế	92,33% 94,96% 93,63%
2 Shavasana 3	98,73% 95,51% 97,09%
Tadasana 4	92,24% 94,30% 93,26%
Trikonasana 5	98,73% 98,80% 98,77%
Vrikshasana	94,76% 91,07% 92,88%
Trung bình	95,20% 95,07% 95,12%

Tuy nhiên, các kết quả theo chủ đề khôn ngoan được chấp nhận hơn. Các kết quả chính xác có thể được nhìn thấy trong Bảng 6. Những kết quả này tiếp tục chiến lược sự mạnh mẽ của chiến lược đánh giá của chúng tôi. Vài ví dụ nơi mô hình hoạt động kém có thể xuất hiện trong Hình 5. Lưu ý rằng trong phần lớn các trường hợp phân loại sai, các điểm chính được phát hiện bởi AlphaPose [10] có lỗi. Trong một số, có một con người thứ hai trong thị giác, trong khi những cái khác bị dán nhãn sai. Hình thức tương đối kém này của AlphaPose [10] trên dữ liệu yogasana chỉ ra nhu cầu về bộ dữ liệu chú thích điểm chính mới có chứa một số khó khăn tư thế phổ biến ở đây.

Bảng 7: Kết quả theo khung của Jain et al. [20] bộ dữ liệu

Tư thế ID Yoga	Thu hồi chính xác Điểm F1
0 Tư Thế Vòng Hoa	99,27% 98,27% 98,76%
1 tư thế em bé hạnh phúc	99,39% 98,65% 99,02%
Tư thế 2 đầu gối	98,78% 98,59% 98,69%
3 tư thế Lunge	96,70% 96,83% 96,76%
4 Tư Thế Ngọn Núi	97,78% 99,10% 98,44%
5 tư thế tấm ván	97,88% 96,41% 97,14%
6 Tư Thế Giơ Tay	94,02% 97,21% 95,59%
Gập người về phía trước 7 chỗ ngồi	98,63% 98,76% 98,69%
8 tư thế nhân viên	98,85% 98,47% 98,66%
9 Cúi gập người về phía trước	94,57% 93,83% 94,20%
Trung bình (của chúng tôi)	97,59% 97,61% 97,59%
Trung bình (Jain et al. [20])	91% 91% 91%

Bảng 8: Độ chính xác cho tất cả các cấp bậc trong Yoga-82 [44]. Hiệu suất của ba biến thể của họ, MobileNet-V2 và DenseNet-201 được lấy trực tiếp từ bài báo Yoga-82.

Phương pháp	Độ chính xác Top-1 Cấp 1	Cấp 2	Cấp 3
MobileNet-V2 [39]	-	-	71,11%
DenseNet-201 [18]	-	-	74,91%
Yoga-82 Biến thể 1 [44]	83,84%	85,10%	79,35%
Yoga-82 Biến thể 2 [44]	89,81%	84,59%	79,08%
Yoga-82 Biến thể 3 [44]	87,20%	84,42%	78,88%
Của chúng ta (DCPose[31])	86,47%	85,05%	55,07%
Của chúng ta (ĐÓNG[34])	86,53%	83,75%	78,01%
Của chúng ta	91,21%	87,91%	80,14%

4.5 Thí nghiệm trên Jain et al. tập dữ liệu

Jain et al. [20] bộ dữ liệu bao gồm 27 đối tượng thực hiện 10 tư thế khác nhau. Các tác giả đề xuất việc sử dụng CNN 3D trên phân đoạn video của mỗi 16 khung hình. Tuy nhiên, chúng tôi phân tích các hiệu suất phân loại dựa trên hình ảnh của chúng tôi trên tập dữ liệu của họ. Một lần nữa, chúng tôi lấy mẫu 100 khung hình từ mỗi video và sử dụng các khung hình này cho các phân tích tiếp theo. Mặc dù bộ dữ liệu này bao gồm các đối tượng được ghi lại, chúng tôi không thể xác định đối tượng nào đã tham gia vào video nào và do đó, sẽ không sử dụng chiến lược đánh giá khôn ngoan theo chủ đề của chúng tôi ở đây. Những video này được ghi lại từ một góc máy ảnh thống nhất duy nhất và do đó, chúng tôi không thể thực hiện máy ảnh-khôn ngoan đánh giá một trong hai.

Để đánh giá theo khung hình, chúng tôi đã trích xuất tổng cộng 23090 điểm cao khung hình có độ phân giải và thu nhỏ lại chúng thành độ phân giải thấp hơn vì của các ràng buộc tài nguyên. Các kết quả khung khôn ngoan của phương pháp của chúng tôi trên tập dữ liệu này có thể được tìm thấy trong Bảng 7. Ngay cả với dự đoán dựa trên hình ảnh thay vì phân đoạn video, phương pháp của chúng tôi vẫn vượt trội so với phương pháp 3D của họ Phương pháp dựa trên CNN cho độ chính xác 91%. Đây là một điều rõ ràng trình diễn lợi thế của việc sử dụng học chuyển tiếp qua đầu để kết thúc đào tạo trong miền khan hiếm dữ liệu này.

4.6 Thử nghiệm trên tập dữ liệu Yoga-82

Yoga-82 [44] là một trong số ít các điểm chuẩn có sẵn công khai cho phân loại yoga Mặc dù các tác giả đã cung cấp các liên kết để tải xuống từng hình ảnh, một số liên kết không thể truy cập được. Chúng tôi đã có thể trích xuất 18k hình ảnh trong tổng số 28k trong điểm chuẩn. Chúng tôi tiếp tục quản lý dữ liệu để xóa hình ảnh giống như clipart và cuối cùng đã có 12k hình ảnh. Chúng tôi sử dụng tập hợp con 12k hình ảnh này cho tất cả các mục đích tiếp theo của chúng tôi Phân tích. Bộ dữ liệu điểm chính AlphaPose [10] cho những hình ảnh này sẽ được cung cấp công khai để tái sản xuất.

Để chứng minh rõ hơn những lợi ích của việc sử dụng phương pháp học chuyển giao, chúng tôi đã giữ cho giai đoạn phân loại càng đơn giản càng tốt. Đối với tất cả các phân tích ở trên, chúng tôi đã sử dụng một bộ phân loại rừng ngẫu nhiên đơn giản. Tuy nhiên, các khu rừng ngẫu nhiên hoạt động khá kém trên bộ dữ liệu Yoga-82 [44] đầy thách thức hơn . Sử dụng một nhóm Random Forests [4], Gradient Boosting [11] và LightGBM [22], chúng tôi thu được kết quả tương đương với báo cáo của các tác giả. Một mô tả ngắn gọn về một số mẫu được phân loại sai có thể được nhìn thấy trong Hình 5.

Phân tích của chúng tôi dựa trên tập hợp con của tập dữ liệu gốc. Vì vậy, hình thức bên ngoài chỉ mang tính chất đại diện, không dùng để so sánh chính xác.

Chúng tôi cũng so sánh phương pháp này với các phương pháp được đề xuất gần đây DCPose[31] và KAPAO[34], nhưng quan sát thấy kết quả kém hơn một chút. Chúng tôi tin rằng điều này có thể là do số lượng điểm chính ít hơn (17 trên cơ thể) và không có thông tin thời gian cần thiết cho DCPose. Các điểm chính được suy luận bởi tất cả các phương pháp này cho hình ảnh Yoga 82 sẽ được cung cấp công khai để tái tạo. Của chúng tôi các kết quả cùng với những kết quả thu được bởi các tác giả Yoga-82 [44] và một số đường cơ sở được sử dụng bởi họ có thể được tìm thấy trong Bảng 8.

Trái ngược với bộ dữ liệu của chúng tôi và bộ dữ liệu Yadav et al [48] , Yoga-82 bao gồm các hình ảnh trong tự nhiên. Cùng với nhiều chủ đề phong phú hơn và tạo dáng, hình ảnh của họ cũng từ nhiều góc máy khác nhau. Ngoài ra, hình ảnh không ở trong môi trường được kiểm soát, và điều kiện nền không đồng nhất. Tất cả những yếu tố này làm cho đây là một điểm chuẩn thách thức hơn và do đó, mô hình của chúng tôi có mức thấp hơn hiệu suất ở đây so với các bộ dữ liệu khác. Người mẫu khác trong Bảng 8 là những kiến trúc sâu với mức độ phức tạp cao. Tận dụng học tập chuyển giao, chúng tôi đã có thể vượt trội hơn họ bằng cách đào tạo các bộ phân loại tương đối đơn giản hơn. Đây là xác nhận rõ ràng của tính hữu ích của việc học chuyển giao trong lĩnh vực yoga khan hiếm dữ liệu này phân loại.

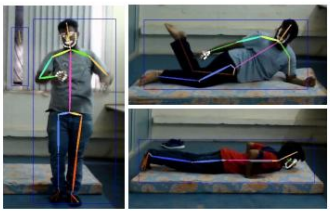
5 THẢO LUẬN

Mặc dù trong các thử nghiệm của mình, chúng tôi sử dụng AlphaPose[10], một cách tiếp cận tương tự có thể được áp dụng bằng các phương pháp ước tính tư thế khác, gần đây hơn. Chúng tôi thử nghiệm KAPAO[34] và DCPose[31] trên Yoga-82 để quan sát hiệu quả của các phương pháp ước lượng tư thế khác nhau trong giai đoạn đầu của đường ống của chúng tôi. Để chứng minh lợi ích của việc học chuyển giao trên miền khan hiếm dữ liệu này, chúng tôi chỉ ra rằng ngay cả với một bộ phân loại đơn giản, các điểm chính mà AlphaPose học được dẫn đến kết quả khá hứa hẹn. Hiệu suất nhất quán trên 4 bộ dữ liệu được xem xét chứng tỏ hiệu quả của phương pháp này. Trong khi phân loại Yogasana trước đây các phương pháp sử dụng các thuật toán phức tạp với đào tạo từ đầu đến cuối, phương pháp đơn giản hơn của chúng tôi có thể đạt được kết quả tương đương.

Trong số ba chiến lược đánh giá mà chúng tôi khám phá, chiến lược theo khung đánh giá được sử dụng phổ biến nhất. Hầu hết các công việc hiện có trong lĩnh vực này sử dụng chính chiến lược này. Tuy nhiên, khi sử dụng khung

, ,

(a) Một số hình ảnh ví dụ từ Yoga-82 đã bị phân loại sai. Trên cùng bên trái là một ví dụ về đảo ngược chân. Trên cùng bên phải là hình ảnh có độ phân giải rất thấp, dẫn đến khả năng dự đoán kém hơn. Các ví dụ còn lại có khả năng dự đoán kém do các bộ phận cơ thể bị che khuất.



(b) Các khung chuyển tiếp bị dán nhãn sai. Tư thế của đối tượng không phù hợp với asana được dán nhãn. Ở bên trái, đối tượng chưa bắt đầu tạo dáng. Ở trên cùng bên phải, anh ấy đang thay đồ, trong khi ở dưới cùng bên phải, anh ấy đã hoàn thành.

(c) AlphaPose dự đoán kém. Trong trái, tay lỡ. Ở trên cùng bên phải, toàn bộ con người bị bỏ sót và phát hiện dương tính giả. Ở dưới cùng bên phải, hộp giới hạn chính xác, ảnh hưởng đến quá trình chuẩn hóa.

Hình 5: Một số ví dụ được phân loại sai trong bộ dữ liệu chuẩn Yoga-82 [44] và Yadav et al [48].

đã được trích xuất từ video, chiến lược này dẫn đến vấn đề nghiêm trọng là rò rỉ mục tiêu. Yadav và cộng sự. [48] đã sử dụng chiến lược này và báo cáo độ chính xác 100% trong một nửa số asana mà họ đã thử nghiệm . Kết quả của chúng tôi trong Bảng 2 cũng rất cao, với độ chính xác 100% ở hai tư thế. Một xu hướng tương tự có thể được nhìn thấy với bộ dữ liệu của Jain et al [20] . Do rò rỉ mục tiêu, chúng tôi tin rằng đánh giá theo khung không phải là một chiến lược thử nghiệm lý tưởng.

Với các đánh giá thông minh về chủ đề, mô hình của chúng tôi được thử nghiệm trên các chủ đề mà nó chưa từng thấy trong quá trình đào tạo. Chúng tôi cũng thu được kết quả tốt với chiến lược này, như có thể thấy trong Bảng 3. Điều này cho thấy rằng mô hình của chúng tôi mạnh mẽ so với các biến thể của đối tượng thực hiện asana. Đây là ảnh hưởng trực tiếp của việc sử dụng Transfer Learning trong công cụ ước tính tư thế giai đoạn đầu của chúng tôi. Mặc dù trình phân loại chưa được đào tạo về một chủ đề cụ thể, nhưng công cụ ước tính tư thế đã được đào tạo trên nhiều loại người hơn, mặc dù từ một tập dữ liệu khác. Điều này cho phép nó mạnh mẽ đối với các biến thể của đối tượng và vì bộ phân loại chỉ yêu cầu các điểm chính từ công cụ ước tính tư thế này, hiệu suất tốt này được truyền trực tiếp đến kết quả cuối cùng. Tương tự, các thuộc tính mong muốn khác của công cụ ước tính tư thế sẽ được truyền đến kết quả cuối cùng và do đó, những tiến bộ trong lĩnh vực phân loại tư thế yogasana cũng khuyến khích sự phát triển và nghiên cứu trong lĩnh vực ước tính tư thế.

Chiến lược đánh giá thứ ba mà chúng tôi sử dụng là đánh giá bằng máy ảnh. Trong Bảng 4, có thể thấy rằng khi mô hình được đào tạo trên số góc máy ảnh ít hơn, nó có xu hướng hoạt động kém hơn. Điều này cho thấy rằng việc bao gồm nhiều góc máy ảnh hơn trong quá trình đào tạo có tác động tích cực đến hiệu suất và khả năng khái quát hóa của mô hình. Ngay cả đối với cùng một tư thế đang được xem xét, các góc máy ảnh khác nhau sẽ tạo ra các tọa độ điểm chính rất khác nhau. Chúng tôi tin rằng việc bao gồm tất cả các tọa độ khác nhau này sẽ cho phép một mô hình tìm hiểu các mô hình độc lập với chế độ xem chất lượng tốt hơn, do đó cải thiện hiệu suất của nó.

Kết quả của việc sử dụng KAPAO[34] và DCPose[31] trên Yoga-82 cho thấy hiệu suất không tăng so với các thử nghiệm của chúng tôi với AlphaPose[10], mặc dù hai phương pháp này vượt trội đáng kể so với AlphaPose trên bộ dữ liệu tiêu chuẩn. Chúng tôi tin rằng điều này là do mô hình của AlphaPose mà chúng tôi sử dụng được đào tạo trước trên bộ dữ liệu Halpe Full Body [28] có 136 điểm chính được chú thích trên cơ thể con người trong khi mô hình của DCPose được đào tạo trước trên bộ dữ liệu PoseTrack [1] chứa 17 điểm chính và KAPAO được đào tạo trước -Huân luyê n về bộ dữ liệu COCO[29] bao gồm 17 điểm chính. Vì các tư thế Yoga liên quan đến cơ thể con người thực hiện các tư thế cực kỳ phi tuyến tính, nên cần có nhiều điểm chính hơn trên cơ thể để phân loại đầy đủ thành công một Yogasana. Ngoài ra, hiệu suất của DCPose có thể thấp hơn vì DCPose chủ yếu là mô hình theo dõi tư thế phù hợp với dữ liệu video, yêu cầu khung hình trước đó và khung hình tiếp theo của khung hình hiện tại để cải thiện khả năng ước tính tư thế của khung hình hiện tại. Vì Yoga-82 không phải là tập dữ liệu video nên không có sự liên tục về thời gian từ hình ảnh này sang hình ảnh tiếp theo, điều này có thể làm giảm hiệu suất của DCPose trên tập dữ liệu, sau đó dẫn đến hiệu suất kém của bộ phân loại.

Bên cạnh việc bản thân mô hình có khả năng thay đổi góc máy ảnh mạnh mẽ, một hướng khác có thể là học trực tiếp cách thể hiện tư thế độc lập với chế độ xem. Nghiên cứu đang thực hiện về ước tính tư thế 3D chỉ sử dụng dữ liệu video cung cấp một hướng mới trong đó các tư thế được phát hiện từ các góc máy ảnh rất khác nhau có thể có các biểu diễn tư thế giống hệt nhau, do đó tăng khả năng khái quát hóa đối với các góc máy ảnh . Mặc dù đặc biệt nổi bật trong các tư thế Yogasana, nhưng nhu cầu khái quát hóa qua các góc máy ảnh khác nhau cũng thể hiện rõ trong nhiều ứng dụng ước lượng tư thế con người khác. Một tình huống đặc biệt khó khăn trong đó nhiều thuật toán ước tính tư thế không thành công là tắc nghẽn, trong đó một phần của bộ phận cơ thể bị ẩn hoặc không hiển thị trực tiếp. Đáng chú ý là một bộ phận cơ thể bị che khuất khi nhìn từ một góc máy ảnh này có thể thực sự được nhìn thấy rõ ràng từ một góc máy ảnh khác. Nếu một mô hình thực sự độc lập với chế độ xem, thì nó sẽ có thể lấy tư thế từ chế độ xem không bị che và sử dụng nó cho tư thế trong chế độ xem bị che. Do đó, việc nghiên cứu các phương pháp phân loại tư thế độc lập xem có thể giúp giảm thiểu vấn đề khớp cắn. Một mô hình phân loại tư thế khái quát hóa các góc máy ảnh không nhìn thấy sẽ khá có lợi cho cộng đồng.

Yoga-82 [44] có hình ảnh tự nhiên, với sự thay đổi đáng kể về góc máy ảnh, đối tượng, điều kiện ánh sáng nền và nhiều yếu tố khác. Hiệu suất trên tập dữ liệu này đối với các lớp cấp độ 3 (82 asana) rất giống với các đánh giá thông minh về máy ảnh của chúng tôi, trong đó các biến thể khác với góc máy ảnh là tối thiểu. Điều này cho thấy rằng các biến thể góc máy ảnh có tác động nhiều nhất đến hiệu suất của một phương pháp, so với các biến thể thông minh về khung hình và chủ thể. Do đó, một khung phân loại độc lập về chế độ xem có thể đánh giá hiệu quả khả năng khái quát hóa của mô hình đối với các góc máy ảnh khác nhau sẽ mang lại kết quả tốt hơn nhiều

ước tính hiệu suất của mô hình cho một tập dữ liệu trong tự nhiên. Chúng tôi hy vọng công việc này có thể đóng vai trò là cơ sở vững chắc cho các công trình tiếp theo của chúng tôi và các nhà nghiên cứu khác.

6 KẾT LUẬN VÀ CÔNG VIỆC TƯƠNG LAI

Trong bài báo này, chúng tôi đề xuất một cấu trúc hai giai đoạn để phân loại các tư thế yoga. Chúng tôi sử dụng AlphaPose [10], một phương pháp ước tính tư thế được thiết lập tốt làm công cụ trích xuất đặc trưng trong giai đoạn đầu tiên và phân loại rừng ngẫu nhiên trong giai đoạn thứ hai. Trong trường hợp không có bộ dữ liệu quy mô lớn để ước tính tư thế yoga, chúng tôi sử dụng các mô hình được đào tạo trước trên bộ dữ liệu quy mô lớn hiện có Halpe Full Body [28]. Việc học chuyển giao từ các mô hình ước tính tư thế được kỳ vọng sẽ mang lại hiệu suất tốt hơn cho việc phân loại Yogasana. Chúng tôi tạo một tập dữ liệu mới để phân loại Tư thế Yoga tập trung vào các quan điểm khác nhau về một chủ đề. Theo hiểu biết tốt nhất của chúng tôi, bộ dữ liệu yoga của chúng tôi là bộ dữ liệu yoga đầu tiên xem xét rõ ràng các tư thế asana từ 4 góc máy ảnh khác nhau và có số lượng đối tượng lớn nhất được ghi lại một cách có hệ thống. Dữ liệu về điểm chính cũng như mã mô hình được sử dụng sẽ được tạo thành nguồn mở để đẩy nhanh hơn nữa nghiên cứu trong lĩnh vực này.

Chúng tôi đưa ra sơ đồ đánh giá gồm 3 bước để đánh giá mức độ mạnh mẽ của mô hình đối với các nguồn biến thể khác nhau. Do những thiếu sót cụ thể với các phương pháp đánh giá riêng lẻ, chúng tôi ủng hộ việc sử dụng cả ba để kiểm tra hiệu quả khả năng khái quát hóa của một mô hình.
Chúng tôi chứng minh rằng mô hình của chúng tôi hoạt động đáng ngưỡng mộ ở hai phần đầu tiên và có khả năng cạnh tranh với các phương pháp phân loại Yogasana đã xuất bản trước đó trên 3 bộ dữ liệu có sẵn công khai. Quan sát thấy hiệu suất thấp hơn đáng kể khi được đánh giá qua các góc máy ảnh, chúng tôi thảo luận về những thách thức cụ thể liên quan đến cài đặt này và lập luận rằng đánh giá thông minh về máy ảnh này thường có thể đáng tin cậy hơn như một thước đo.

Có một vài hạn chế liên quan đến cách tiếp cận của chúng tôi. Vì chúng tôi dựa vào ước tính tư thế và phát hiện chính xác các điểm chính trong giai đoạn đầu tiên, nên hầu hết các thách thức xuất hiện trong ước tính tư thế cũng phản ánh một cách tự nhiên trong cách tiếp cận của chúng tôi. Loại trừ và đảo ngược là hai trường hợp thất bại thường được nghiên cứu. Những trường hợp này đặc biệt khó khăn trong trường hợp Yogasana vì các tư thế phức tạp liên quan ở đây, trong đó một số bộ phận cơ thể thường khó phân biệt hoặc bị che khuất khỏi tầm nhìn do các bộ phận cơ thể khác. Hầu hết các bộ dữ liệu ước tính tư thế quy mô lớn bao gồm các tư thế bình thường hàng ngày có thể không cho phép một mô hình thực hiện tốt các tư thế phức tạp như vậy. Ngay cả trong trường hợp tắc nghẽn xảy ra, các bộ phận cơ thể ẩn thường không được dán nhãn. Bộ dữ liệu của chúng tôi cũng chỉ chứa các nhãn phân loại và sẽ cần nhiều nỗ lực và thời gian hơn để thu thập các chú thích điểm chính xác cho nó. Bên cạnh sự khan hiếm dữ liệu, khả năng khái quát hóa kém đối với các góc máy ảnh cũng là một hạn chế đối với phương pháp của chúng tôi. Chúng tôi tin rằng công việc trong tương lai có thể cố gắng giảm bớt nhiều hạn chế này. Những hạn chế cho thấy phạm vi đáng kể để nghiên cứu thêm về cách phân loại tư thế độc lập và nhu cầu về bộ dữ liệu ước tính tư thế quy mô lớn với các tư thế phức tạp tương tự như Yogasana. Các phương pháp ước tính tư thế 3D và sử dụng thông tin chuyên sâu để gắn nhãn các bộ phận cơ thể bị che khuất hoặc bị che khuất cũng có thể giúp giảm thiểu một số thách thức này.

Trong các ứng dụng chẳng hạn như ứng dụng của huấn luyện viên Yoga tự động, sẽ mong muốn có các mô hình mạnh mẽ trên các máy ảnh khác nhau góc độ. Để đạt được điều này, một trong hai cần đào tạo các mô hình trên dữ liệu lớn hơn bằng cách sử dụng nhiều góc quay khác nhau (kết quả của chúng tôi cho thấy kết hợp nhiều máy ảnh hơn trong tập huấn luyện sẽ cải thiện bài kiểm tra theo hình thức), hoặc người ta cần kết hợp các phương pháp thị giác máy tính để làm cho các mô hình này không phụ thuộc vào góc máy ảnh. Kết quả của chúng tôi về các góc quay không nhìn thấy được cho thấy vẫn còn phạm vi đáng kể để cải thiện việc phân loại Yogasana từ các góc quay khác nhau. Đây là một lĩnh vực thú vị cho nghiên cứu trong tương lai. Chúng tôi tin rằng những nghiên cứu tiếp theo về phân loại Yogasana độc lập với chế độ xem camera cũng sẽ tìm thấy nhiều ứng dụng khác. Nó cũng có khả năng nâng cao trình độ tiên tiến nhất trong lĩnh vực ước tính tư thế con người.

NGƯỜI GIỚI THIỆU

[1] Mykhaylo Andriluka, Umar Iqbal, Eldar Insafutdinov, Leonid Pishchulin, Anton Milan, Juergen Gall, và Bernt Schiele. 2018. Posetrack: Điểm chuẩn để ước tính và theo dõi tư thế con người. Trong Kỷ yếu hội nghị IEEE về thị giác máy tính và nhận dạng mẫu. 5167–5176.

[2] Mykhaylo Andriluka, Leonid Pishchulin, Peter Gehler, và Bernt Schiele. 2014. Ước tính tư thế con người 2d: Tiêu chuẩn mới và phân tích hiện đại. Trong Kỷ yếu của Hội nghị IEEE về Tầm nhìn Máy tính và Nhận dạng Mẫu. 3686–3693.

[3] Djamila Romaissa Beddiar, Brahim Nini, Mohammad Sabokrou, và Abdenour Hadid. 2020. Nhận dạng hoạt động của con người dựa trên tầm nhìn: một cuộc khảo sát. Công cụ và ứng dụng đa phương tiện 79, 41 (2020), 30509–30555.

[4] Leo Breiman. 2001. Rừng ngẫu nhiên. Học máy 45, 1 (2001), 5–32.

[5] Arndt Büssing, Andreas Michalsen, Sat Bir S Khalsa, Shirley Telles, và Karen J Sherman. 2012. Tác dụng của yoga đối với sức khỏe thể chất và tinh thần: một bản tóm tắt ngắn các bài đánh giá. Y học thay thế và bổ sung dựa trên bằng chứng 2012 (2012).

[6] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, và Yaser Sheikh. 2019. OpenPose: ước tính tư thế 2D nhiều người trong thời gian thực bằng cách sử dụng Part Affinity Fields. Các giao dịch của IEEE về phân tích mẫu và trí thông minh của máy 43, 1 (2019), 172–186.

[7] Hua-Tsung Chen, Yu-Zhen He, và Chun-Chieh Hsu. 2018. Hệ thống đào tạo yoga có sự trợ giúp của máy tính. Công cụ và ứng dụng đa phương tiện 77, 18 (2018), 23969–23991.

[8] Hua-Tsung Chen, Yu-Zhen He, Chun-Chieh Hsu, Chien-Li Chou, Suh-Yin Lee, và Bao-Shuh P Lin. 2014. Nhận biết tư thế yoga để tự rèn luyện. Trong Hội nghị quốc tế về mô hình hóa đa phương tiện. Springer, 496–505.

[9] Bowen Cheng, Bin Xiao, Jingdong Wang, Honghui Shi, Thomas S Huang, và Lei Zhang. 2020. Higherhrnet: Học biểu diễn theo tỷ lệ để ước tính tư thế con người từ dưới lên . Trong Kỷ yếu hội nghị IEEE/CVF về thị giác máy tính và nhận dạng mẫu. 5386–5395.

[10] Hao-Shu Fang, Shuqin Xie, Yu-Wing Tai, và Cewu Lu. 2017. Rmpe: Ước tính tư thế nhiều người theo khu vực. Trong Kỷ yếu của hội nghị quốc tế IEEE về thị giác máy tính. 2334–2343.

[11] Jerome H Friedman. 2002. Tăng cường độ dốc ngẫu nhiên. Thống kê tính toán & phân tích dữ liệu 38 (2002), 367–378.

[12] Zigang Geng, Ke Sun, Bin Xiao, Zhaoxiang Zhang, và Jingdong Wang. 2021. Ước tính tư thế con người từ dưới lên thông qua hồi quy điểm chính không đồng nhất. Trong Kỷ yếu của Hội nghị IEEE/CVF về Thị giác máy tính và Nhận dạng mẫu. 14676–14686.

[13] Neha P Gothe, Imadh Khan, Jessica Hayes, Emily Erlenbach, và Jessica S Damoi seaux. 2019. Tác dụng của yoga đối với sức khỏe não bộ: tổng quan hệ thống các tài liệu hiện tại. Độ dẻo của não 5, 1 (2019), 105–122.

[14] Erik J Groessl, Deepak Chopra, và Paul J Mills. 2015. Tổng quan về nghiên cứu yoga đối với sức khỏe và hạnh phúc. Tạp chí Yoga & Vật lý trị liệu 5, 4 (2015), 1.

[15] Ashish Gupta và Hari Prabhat Gupta. 2021. Trợ giúp Yoga: Tận dụng các cảm biến chuyển động để học cách thực hiện đúng bài tập Yoga với phản hồi. Giao dịch của IEEE về Trí tuệ nhân tạo (2021).

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, và Jian Sun. 2016. Học phần dư sâu để nhận dạng hình ảnh. Trong Kỷ yếu hội nghị IEEE về thị giác máy tính và nhận dạng mẫu. 770–778.

[17] Sepp Hochreiter và Jürgen Schmidhuber. 1997. Trí nhớ ngắn hạn dài. Tính toán thần kinh 9, 8 (1997), 1735-1780.

[18] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, và Kilian Q Weinberger. 2017. Mạng tích chập được kết nối dày đặc. Trong Kỷ yếu hội nghị IEEE về thị giác máy tính và nhận dạng mẫu. 4700–4708.

[19] Muhammad Usama Islam, Hasan Mahmud, Faisal Bin Ashraf, Iqbal Hossain, và Md Kamrul Hasan. 2017. Nhận dạng tư thế yoga bằng cách phát hiện các điểm khớp của con người trong thời gian thực bằng cách sử dụng microsoft kinect. Năm 2017 hội nghị công nghệ nhân đạo IEEE Khu vực 10 (R10-HTC). IEEE, 668–673.

[20] Shrajal Jain, Aditya Rustagi, Sumeet Saurav, Ravi Saini, và Sanjay Singh. 2021. Kiến trúc học sâu lấy cảm hứng từ CNN ba chiều để nhận dạng tư thế Yoga trong môi trường thế giới thực. Điện toán thần kinh và ứng dụng 33, 12 (2021), 6427–6441.

, ,

[21] Shuiwang Ji, Wei Xu, Ming Yang, và Kai Yu. 2012. Mạng thần kinh tích chập 3D để nhận dạng hành động của con người. Các giao dịch của IEEE về phân tích mẫu và trí thông minh của máy 35, 1 (2012), 221–231.

[22] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, và Tie-Yan Liu. 2017. Lightgbm: Cây quyết định tăng độ dốc hiệu quả cao . Những tiến bộ trong hệ thống xử lý thông tin thần kinh 30 (2017), 3146–3154.

[23] Muhammed Kocabas, Salih Karagoz, và Emre Akbas. 2018. Multiposenet: Ước tính tư thế nhiều người nhanh chóng bằng cách sử dụng mạng dư tư thế. Trong Kỷ yếu của hội nghị châu Âu về tầm nhìn máy tính (ECCV). 417–433.

[24] Sven Kreiss, Lorenzo Bertoni, và Alexandre Alahi. 2019. Pifpaf: Các trường tổng hợp để ước tính tư thế con người. Trong Kỷ yếu hội nghị IEEE/CVF về thị giác máy tính và nhận dạng mẫu. 11977–11986.

[25] Yann LeCun, Léon Bottou, Yoshua Bengio, và Patrick Haffner. 1998. Học tập dựa trên độ dốc được áp dụng để nhận dạng tài liệu. Proc. IEEE 86, 11 (1998), 2278–2324 .

[26] Jiefeng Li, Siyuan Bian, Ailing Zeng, Can Wang, Bo Pang, Wentao Liu và Cewu Lu, 2021. Hồi quy tư thế con người với ước tính khả năng log còn lại. Trong Kỷ yếu của Hội nghị Quốc tế IEEE/CVF về Tầm nhìn Máy tính. 11025–11034 .

[27] Jiefeng Li, Can Wang, Hao Zhu, Yihuan Mao, Hao-Shu Fang, và Cewu Lu. 2019. Tư thế đám đông: Ước tính tư thế cảnh đông đúc hiệu quả và một tiêu chuẩn mới. Trong Kỷ yếu hội nghị IEEE/CVF về thị giác máy tính và nhận dạng mẫu. 10863–10872.

[28] Yong-Lu Li, Liang Xu, Xinpeng Liu, Xijie Huang, Yue Xu, Shiyi Wang, Hao-Shu Fang, Ze Ma, Mingyang Chen, và Cewu Lu. 2020. Pastanet: Hướng tới động cơ tri thức hoạt động của con người. Trong Kỷ yếu của Hội nghị IEEE/CVF về Thị giác máy tính và Nhận dạng mẫu 382–391.

[29] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, và C. Lawrence Zitnick. 2014. Microsoft coco: Các đối tượng phổ biến trong ngữ cảnh. Trong hội nghị châu Âu về thị giác máy tính. Springer, 740–755.

[30] Weiyao Lin, Huabin Liu, Shizhan Liu, Yuxi Li, Rui Qian, Tao Wang, Ning Xu, Hongkai Xiong, Guo-Jun Qi, và Nicu Sebe. phân tích video trọng tâm trong các sự kiện phức tạp.arXiv preprint arXiv:2005.04490 (2020).

[31] Zhenguang Liu, Haoming Chen, Runyang Feng, Shuang Wu, Shouling Ji, Bailin Yang và Xun Wang. 2021. Mạng kép liên tiếp sâu để ước tính tư thế con người. Trong Kỷ yếu của Hội nghị IEEE/CVF về Thị giác máy tính và Nhận dạng mẫu .525–534.

[32] Ningning Ma, Xiangyu Zhang, Hai-Tao Zheng, và Jian Sun. 2018. Shufflenet v2: Hướng dẫn thiết thực để thiết kế kiến trúc cnn hiệu quả. Trong Kỷ yếu của hội nghị châu Âu về tầm nhìn máy tính (ECCV). 116–131.

[33] Teja Kiran Kumar Maddala, PVV Kishore, Kiran Kumar Eepuri, và Anil Kumar Dande. 2019. YogaNet: Nhận dạng Yoga Asana 3-D bằng cách sử dụng Bản đồ vị trí góc nhìn chung với ConvNets. IEEE Giao dịch trên Đa phương tiện 21, 10 (2019), 2492–2503.

[34] William McNally, Kanav Vats, Alexander Wong, và John McPhee. Năm 2021. Suy nghĩ lại về cách biểu diễn điểm chính: Lập mô hình điểm chính và tư thế làm đối tượng để ước tính tư thế con người nhiều người. bản in trước arXiv arXiv:2111.08557 (2021).

[35] Học giả Niharika. 2020. Bộ dữ liệu tư thế Yoga mới. https://www.kaggle.com/General/192938

[36] Suneth Ranasinghe, Fadi Al Machot, và Heinrich C Mayr. 2016. Đánh giá về các ứng dụng của hệ thống công nhận hoạt động liên quan đến hiệu suất và đánh giá. Tạp chí quốc tế về mạng cảm biến phân tán 12, 8 (2016), 1550147716665520.

[37] Alyson Ross và Sue Thomas. 2010. Lợi ích sức khỏe của yoga và tập thể dục: đánh giá các nghiên cứu so sánh. Tạp chí y học thay thế và bổ sung 16, 1 (2010), 3–12.

[38] Pooja Swami Sahni, Kamlesh Singh, Nitesh Sharma, và Rahul Garg. 2021. Yoga , một chiến lược hiệu quả để tự quản lý các vấn đề liên quan đến căng thẳng và giữ gìn sức khỏe trong thời gian phong tỏa vì COVID19: Một nghiên cứu cắt ngang. PloS một 16, 2 (2021), d0245214.

[39] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, và Liang Chieh Chen. 2018. Mobilenetv2: Đảo ngược phần dư và tắc nghẽn tuyến tính. Trong Kỷ yếu hội nghị IEEE về thị giác máy tính và nhận dạng mẫu. 4510–4520.

[40] Robert E Schapire. 2013. Giải thích về adaboost. Trong suy luận thực nghiệm. mùa xuân, 37–52.

[41] Marina Skurichina và Robert PW Duin. 1998. Đóng bao cho các bộ phân loại tuyến tính. Nhận dạng mẫu 31, 7 (1998), 909–930.

[42] Ke Sun, Bin Xiao, Dong Liu, và Jingdong Wang. 2019. Học biểu diễn chuyên sâu với độ phân giải cao để ước tính tư thế con người. Trong Kỷ yếu hội nghị IEEE/CVF về thị giác máy tính và nhận dạng mẫu. 5693–5703.

[43] Edwin W Trejo và Peijiang Yuan. 2018. Nhận dạng các tư thế Yoga thông qua hệ thống tương tác với thiết bị Kinect. Năm 2018, Hội nghị Quốc tế lần thứ 2 về Khoa học Robot và Tự động hóa (ICRAS). IEEE, 1–5.

[44] Manisha Verma, Sudhakar Kumawat, Yuta Nakashima, và Shanmuganathan Raman. 2020. Yoga-82: bộ dữ liệu mới để phân loại chi tiết con người tư thế. Trong Kỷ yếu của Hội thảo IEEE/CVF về Thị giác Máy tính và Hội thảo Nhận dạng Mẫu. 1038–1039.

[45] Donna Vương. 2009. Việc sử dụng yoga cho sức khỏe thể chất và tinh thần ở người lớn tuổi: tổng quan tài liệu. Tạp chí Yoga trị liệu quốc tế 19, 1 (2009), 91–96.

[46] Catherine Woodyard. 2011. Khám phá tác dụng trị liệu của yoga và khả năng nâng cao chất lượng cuộc sống của nó. Tạp chí quốc tế về yoga 4, 2 (2011), 49.

[47] Bin Xiao, Haiping Wu, và Yichen Wei. 2018. Đường cơ sở đơn giản để ước tính và theo dõi tư thế con người. Trong Kỷ yếu của hội nghị châu Âu về tầm nhìn máy tính (ECCV). 466–481.

[48] Santosh Kumar Yadav, Amitojdeep Singh, Abhishek Gupta, và Jagdish Lal Raheja. 2019. Nhận dạng Yoga trong thời gian thực bằng cách sử dụng học sâu. Điện toán thần kinh và ứng dụng 31, 12 (2019), 9349–9361.

[49] Changqian Yu, Bin Xiao, Changxin Gao, Lu Yuan, Lei Zhang, Nong Sang, và Jingdong Wang. 2021. Lite-hrnet: Mạng nhẹ độ phân giải cao. Trong Kỷ yếu của Hội nghị IEEE/CVF về Tầm nhìn và Thị giác Máy tính Nhận dạng Mẫu 10440–10450.

[50] Feng Zhang, Xiatian Zhu, Hanbin Dai, Mao Ye, và Ce Zhu. 2020. Biểu diễn tọa độ nhận biết phân phối để ước tính tư thế con người. Trong Kỷ yếu hội nghị IEEE/CVF về thị giác máy tính và nhận dạng mẫu. 7093–7102.

[51] Jiabin Zhang, Zheng Zhu, Jiwen Lu, Junjie Huang, Guan Huang, và Jie Zhou. Năm 2021. Đơn giản: Mạng đơn với tính năng bắt chước và học điểm để ước tính tư thế con người từ dưới lên. Trong Kỷ yếu Hội nghị AAAI về Trí tuệ nhân tạo, Tập. 35. 3342–3350.

[52] Trịnh Hữu Trương. 2012. Cảm biến kinect của Microsoft và tác dụng của nó. đa phương tiện IEEE 19, 2 (2012), 4–10.

VẬT LIỆU BỔ SUNG

MÔ TẢ BỘ DỮ LIỆU

A.1 Quay Video Yoga Các video đối tượng

thực hiện các tư thế yoga khác nhau được ghi lại một cách có hệ thống. Chúng tôi có 51 đối tượng thực hiện 20 asana và họ được ghi lại từ 4 góc camera khác nhau. Do một số hạn chế, một số asana không thể được thực hiện bởi một số đối tượng và một số video không thể quay được ở một số góc máy. Nói chung, có tổng cộng 3532 video đã được ghi lại. Cùng với các video thô ở định dạng avi, dữ liệu IR và độ sâu cũng được thu thập bằng Microsoft Kinect. Ngoài những thứ này, bản thân Kinect cũng cung cấp các keypoint do nó phát hiện và những keypoint này cũng được lưu lại để sử dụng trong tương lai.

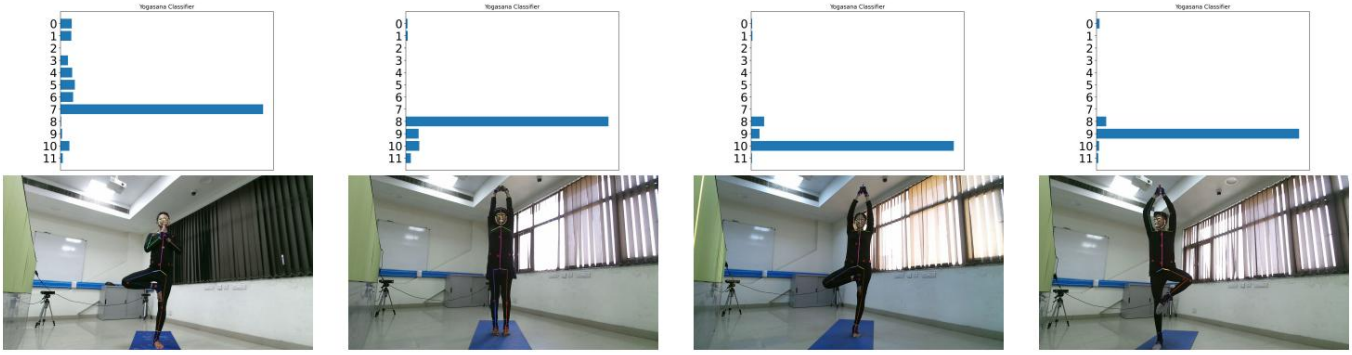
Các video được quay trong môi trường trong nhà được kiểm soát với đủ ánh sáng. Điều này cho phép chúng tôi theo dõi chính xác hiệu quả của từng thành phần của bộ dữ liệu. Ví dụ: trong khi phân tích hiệu suất của mô hình của chúng tôi với các đối tượng không nhìn thấy, các góc máy ảnh được phân bổ tương đối đồng đều hơn và trong khi phân tích tương tự với các góc máy ảnh không nhìn thấy, các đối tượng được phân bổ ở cả hai. Điều này cho phép chúng tôi tách riêng ảnh hưởng của sự thay đổi đối tượng và sự thay đổi góc máy ảnh. Ngược lại, một bộ dữ liệu được ghi lại trong tự nhiên có thể có biến thể lớn hơn nhưng sẽ thiếu sự kiểm soát. Ví dụ: nếu hiệu suất trên hai tập hợp con khác nhau, sẽ rất khó xác định nguyên nhân chính xác của sự khác biệt. Sự khác biệt có thể là do đối tượng, góc máy ảnh, điều kiện ánh sáng hoặc thậm chí là hậu cảnh.

A.2 Chạy AlphaPose Trong trình phân

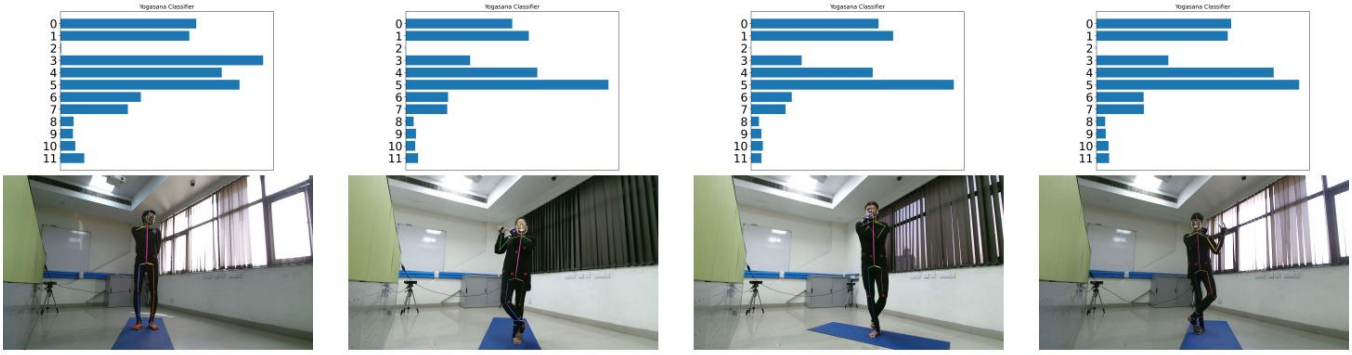
loại Tư thế Yoga đã hình dung, các video đã ghi sẽ được xử lý ngoại tuyến, trong khi để hoạt động trong thời gian thực, nó sẽ là một phần của quy trình hoàn chỉnh. Đối với tập dữ liệu video đã ghi của chúng tôi, chúng tôi đã chạy suy luận AlphaPose trên từng video và lưu trữ chúng riêng biệt. Mô hình AlphaPose nguồn mở cung cấp dưới dạng đầu ra là json chứa các điểm chính, điểm số của chúng và hộp giới hạn được phát hiện của con người tương ứng cho mỗi lần phát hiện trong từng khung hình của mỗi video. Th

Bảng 9: Mô tả tập dữ liệu của chúng tôi

Tư thế ID Yoga	Số môn học				Số khung hình				
	Máy ảnh 1	Máy ảnh 2	Máy ảnh 3	Máy ảnh 4	Máy ảnh 1	Máy ảnh 2	Máy ảnh 3	Máy ảnh 4	Tổng cộng
0 Garudasana còn lại	35	33	32	12	1916	1738	1724	622	6000
1 Garudasana đúng	40	35	33	9	2129	1807	1613	451	6000
2 Gorakshasana	36	33	32	7	2040	1852	1706	402	6000
3 Katichkrasana	47	43	39	15	1962	1798	1592	648	6000
4 Natavasana còn lại	37	26	27	5	2369	1651	1677	303	6000
5 Natavarasana đúng	32	25	26	1	2316	1792	1826	66	6000
6 Pranamasana còn lại	36	32	29	4	2148	1879	1739	234	6000
7 Pranamasana đúng	34	31	26	2	2143	2026	1710	121	6000
8 Tadasana	41	36	38	18	1807	1596	1746	851	6000
9 Vrikshasana còn lại	46	42	42	11	1913	1787	1791	509	6000
10 Vrikshasana đúng	38	32	32	4	2211	1789	1784	216	6000
11 Vẫn	36	31	27	2	2231	1972	1693	104	6000
Tổng cộng	-	-	-	-	25185	21687	20601	4527	72000



Hình 6: Một số mẫu có dự đoán tốt. Mô hình có thể phân biệt giữa trái và phải cho các tư thế song phương



Hình 7: Một số mẫu có dự đoán kém. Các tư thế có tính đối xứng cao và bộ phân loại có khá nhiều nhầm lẫn.

ngoài các tọa độ điểm chính, nó cũng cung cấp các hình ảnh trực quan để xem kết quả một cách định tính. Sau khi dự đoán AlphaPose được lấy cho tất cả các video, chúng tôi chuyển sang giai đoạn tiếp theo của khai thác khung.

A.3 Trích xuất khung hình

Chúng tôi đã lên kế hoạch làm việc trên một bộ phân loại dựa trên hình ảnh thay vì dựa trên video để đạt được hiệu suất gần thời gian thực với đường ống hoàn chỉnh. Do đó, trình phân loại của chúng tôi cần dữ liệu dựa trên hình ảnh để đào tạo và kiểm tra. Chúng tôi có được dữ liệu này bằng cách trích xuất các khung hình riêng lẻ từ các video đã ghi và sử dụng chúng được phát hiện tương ứng những điểm chính. Mỗi video của chúng tôi dài khoảng 2-5 phút và đang

, ,

(a) Một số ví dụ với các đối tượng chuyển tiếp giữa các asana

(b) Một số ví dụ với dự đoán AlphaPose kém.

Hình 8: Một số ví dụ về khung phân loại sai.

được ghi lại ở mức 30,,tổng số khung hình quá lớn so với mục đích của chúng tôi. Thay vì lấy tất cả khung hình trong một video, chúng tôi đã lấy mẫu thống nhất khoảng 200 khung hình trong mỗi phần của video. Một lần nữa, vì các đối tượng sẽ thực hiện cả tư thế bên trái và bên phải cho các tư thế song phương, nên chúng tôi xử lý từng người trong số họ một cách riêng biệt và trích xuất 200 khung hình cho cả tư thế bên trái và bên phải.

Chúng tôi đã ghi lại các dấu thời gian tương ứng với từng phần của tư thế. Chúng tôi quan sát thấy rằng dấu thời gian giữa các máy ảnh hơi khác nhau và do đó có khả năng ghi nhãn sai. Ngoài ra, trong các video thường có một giai đoạn chuyển tiếp, trong đó đối tượng đang vào tư thế, thay đổi từ trái sang phải hoặc đi ra khỏi tư thế đó. Chúng tôi đã lấy bộ đệm 1 giây trước và sau mỗi dấu thời gian cho các khoảng thời gian chuyển tiếp này và chỉ lấy mẫu các khung của chúng tôi từ phần còn lại.

Bên cạnh những điều này, chúng tôi quan sát thấy rằng trong một số trường hợp, đối tượng tạm thời mất thăng bằng và không ở tư thế thực trong một thời gian. Chúng tôi không thể ghi lại các dấu thời gian tương ứng với từng trường hợp này và do đó, một số khung có thể bị gắn nhãn sai vì điều đó.

Sau khi trích xuất các khung hình từ tất cả các video, chúng tôi tiếp tục lấy mẫu phụ của tập dữ liệu như đã thảo luận trong Phần 3.1 của toàn văn. Có thể thấy sự phân bổ khung hình giữa các đối tượng và góc máy ảnh đối với từng loại được lấy mẫu phụ trong Bảng 9. Sự phân bổ này hơi lệch một chút so với máy ảnh 4 và điều này được phản ánh trong hình thức đánh giá thông minh về máy ảnh, trong đó bao gồm cả máy ảnh nhỏ hơn máy ảnh 4 và thử nghiệm trên máy ảnh 3 lớn hơn có định hướng tương tự dẫn đến kết quả kém.

Theo hiểu biết tốt nhất của chúng tôi, bộ dữ liệu yoga của chúng tôi là bộ dữ liệu yoga đầu tiên xem xét rõ ràng các tư thế asana từ 4 góc máy ảnh khác nhau và có số lượng đối tượng được ghi lại nhiều nhất. Yoga-82, một bộ dữ liệu có sẵn công khai, có những hình ảnh dường như được chụp từ các góc máy ảnh khác nhau. Tuy nhiên, những hình ảnh này là trong tự nhiên và không thể tách biệt hiệu ứng của sự thay đổi góc máy ảnh với các biến thể khác. Chúng tôi tin rằng việc tách rời biến thể góc camera sẽ cho phép chúng tôi tập trung tốt hơn vào việc khái quát hóa nó và hy vọng bộ dữ liệu của chúng tôi cũng sẽ giúp các nhà nghiên cứu khác cải thiện phương pháp của họ .

B PHÂN TÍCH ĐỊNH TÍNH Ngoài đánh giá Định lượng trong toàn văn, chúng tôi cũng cố gắng đánh giá định tính hiệu suất của phương pháp trên bộ dữ liệu của chúng tôi . Có thể xem trực quan các dự đoán phân loại đối với một số mẫu trong Hình 6 và Hình 7. Thay vì chỉ hiển thị giá trị dự đoán cuối cùng, chúng tôi trực quan hóa xác suất do mô hình đưa ra cho từng lớp hoặc nhãn 'mềm'. Các xác suất được hiển thị dưới dạng biểu đồ thanh, với trục y có id asana, phù hợp với quy ước trong Bảng 9 và trục x là các dự đoán cho mỗi lớp.

Trong Hình 6, chúng tôi trình bày một số ví dụ hoạt động tốt. Trong các ví dụ này, lớp được dự đoán là chính xác, đồng thời, xác suất cho các lớp khác tương đối thấp, nghĩa là bộ phân loại ít nhầm lẫn hơn. Hình phụ*s 3 và 4 ở đây thực ra là phiên bản bên trái và bên phải của cùng một tư thế yoga và bộ phân loại cũng có thể phân biệt giữa chúng một cách dễ dàng. Những ví dụ này bao gồm các asana có một số đặc điểm giúp chúng dễ thực hiện hơn.

phân biệt, và do đó, bộ phân loại dễ dàng dự đoán chính xác.

Mặt khác, các ví dụ trong Hình 7 có nhiều sự nhầm lẫn hơn. Mặc dù các dự đoán thực sự chính xác, bộ phân loại cũng dự đoán xác suất cao cho các lớp sai. Như vậy, dự đoán có thể sai hoàn toàn đối với một ví dụ tương tự khác.

C. VÍ DỤ ĐƯỢC PHÂN LOẠI MIS

Trong phần này, chúng tôi thảo luận về một số ví dụ trong đó việc đánh giá thông minh bằng máy ảnh đã dự đoán sai. Trong Hình 8a, đối tượng vẫn chưa vào tư thế chính của asana, đã hoàn thành asana và đang rời khỏi nó hoặc bị mất thăng bằng ở giữa. Chúng tôi không thể xác định chính xác tất cả các ví dụ này và do đó, khoảng 2-3% khung hình trong tập dữ liệu của chúng tôi có thể bị gắn nhãn sai. Trong nghiên cứu tiếp theo, chúng tôi dự định quản lý tập dữ liệu của mình tốt hơn bằng cách xác định tất cả các khung bị phân loại sai này và xóa chúng khỏi tập dữ liệu.

Trong Hình 8a, từ trái sang phải, một đối tượng đang vào tư thế chính của động tác Vrikshasana, một đối tượng đang thay đổi từ trái sang phải đối với động tác Vriskshasana, một đối tượng mất thăng bằng khi thực hiện Garudasana và một đối tượng mất thăng bằng khi thực hiện động tác Vriskshasana.

Trong Hình 8b, chúng tôi hiển thị một số ví dụ trong đó AlphaPose hoạt động kém. Các tư thế phổ biến trong Yoga không có trong hầu hết các bộ dữ liệu hiện có để ước tính tư thế và HAR, do đó, mô hình AlphaPose được đào tạo trên các hình ảnh tương đối dễ dàng hơn. Điều này chỉ ra nhu cầu về các bộ dữ liệu ước tính tư thế có các tư thế phức tạp hơn.

Trong Hình 8b, từ trái sang phải, Al phaPose không phát hiện ra điểm mấu chốt nào đối với Katichakrasana, điểm mấu chốt của cánh tay và bàn tay bị bỏ sót đối với Katichakrasana, cánh tay lại bị bỏ sót đối với Tadasana, Vrikshasana và Katichakrasana.