ỨNG DỤNG HỌC CHUYỂN ĐỔI NHẬN DIỆN HÀNH VI GIAN LẬN TRONG PHÒNG THI

Phạm Văn Sự
Học Viện Công Nghệ Bưu Chính Viễn Thông

Tóm tắt: Nhận diện hành động và cử chỉ của con người đã và đang thu hút được sự quan tâm của rất nhiều nhà nghiên cứu trong những năm gần đây. Cùng với sự thành công của việc ứng dụng học sâu, rất nhiều bài toán về nhận diện hành động và cử chỉ của con người ở nhiều khía cạnh như thể thao, sinh hoạt, trợ giúp, y tế, … đã được xem xét và giải quyết. Trong bài báo này, nhóm nghiên cứu đề xuất một giải phát sử dụng học chuyển đổi để giải quyết bài toán nhận diện hành vi gian lận trong phòng thi. Bằng cách sử dụng một mạng học sâu đã được huấn luyện trên tập dữ liệu đủ lớn, giải pháp đề xuất sử dụng học chuyển đổi để cá thể hóa cho bài toán vốn không có nhiều dữ liệu để huấn luyện. Kết quả kiểm chứng trên bộ dữ liệu thu thập được cho thấy giải pháp đề xuất tận dụng được tính tối ưu của học sâu, nhờ học chuyển đổi giảm thời gian cần thiết huấn luyện lại mà vẫn đạt được kết quả nhận diện chính xác cao.

Từ khóa: Hành vi gian lận trong thi cử, học chuyển đổi, học sâu, mạng nơ-ron tích chập, nhận diện cử chỉ, nhận diện hành động.

I. GIỚI THIỆU

Nhận diện hành động và cử chỉ của con người là một trong những mảng được nghiên cứu sôi động nhất trong lĩnh vực thị giác máy tính. Rất nhiều nghiên cứu về mảng này đã được công bố trong những năm gần đây cho thấy được sự ứng dụng phong phú của nhận diện hành động và cử chỉ [1]-[5].

Lĩnh vực áp dụng của nhận diện hành động và cử chỉ của con người đầu tiên phải kể đến đó là nhận diện ngôn ngữ ký hiệu nhằm tạo sự thuận lợi trong giao tiếp với người điếc [6], [7]. Các ký hiệu tay được nhận diện, giải mã tự động nhờ các thuật toán được phát triển và cài đặt trên các ứng dụng giúp chúng ta có thể dễ dàng hiểu và tương tác với những người không có khả năng nói.

Một lĩnh vực áp dụng khác không kém phần quan trọng đó chính là lĩnh vực chăm sóc và theo dõi sức khỏe cho người già cô đơn [8], [9]. Nhờ sự trợ giúp của hệ thống camera cùng với các thuật toán nhận dạng các hành vi bất thường có thể giúp trung tâm chăm sóc hoặc người quản lý có thể hỗ trợ kịp thời.

Nhận diện hành động và cử chỉ cũng được áp dụng trong việc giám sát theo dõi sức khỏe người bệnh [10]. Video giám sát được phân tích và trích xuất các tham số động học để phát hiện các hành động và được phân loại nhằm đánh giá và trợ giúp việc chẩn đoán.

Bên cạnh đó, còn có rất nhiều các lĩnh vực ứng dụng khác mà nhận dạng hành động và cử chỉ con người đã tỏ ra là một giải pháp trợ giúp hữu hiệu trong các hệ thống giao tiếp người – máy dựa trên thị giác máy tính, chẳng hạn như phân tích ngữ cảnh ảnh qua các hành động thường nhật của cuộc sống [11]-[13], phân tích các hành động trong thể thao [14],[15], phân tích các hành động chủ thể để tạo các hoạt động chân thực cho các nhân vật hoạt hình 3D [16].

Sự thành công và thuận lợi cho phép nhận dạng hành động và cử chỉ được áp dụng rộng rãi trong thực tế có được là nhờ sự phát triển của thuật toán và công nghệ nhận diện hành động dựa trên thị giác máy tính, đặc biệt là học sâu.

Ở thế hệ công nghệ đầu tiên sử dụng giải quyết bài toàn nhận diện hành động và cử chỉ dựa trên ảnh thường tiếp cận theo cách trích chọn những đặc trưng thích hợp từ ảnh [17], [18]. Việc trích chọn đặc trưng thường dựa trên quan điểm chủ quan và kinh nghiệm. Điều này khiến cách tiếp cận này không khai thác được hết những thông tin có tính phân biệt mức trừu tượng cao từ dữ liệu ảnh vốn là những thông tin phức tạp. Và do đó, các phương pháp tiếp cận này thường chỉ tập trung vào một số hành động nhất nhưng độ chính xác cũng không cao [19].

Cùng với sự phát triển và hoàn thiện của kỹ thuật học sâu, hướng tiếp cận giải quyết các bài toán nhận diện hành vi đã được chuyển hướng sang sử dụng học sâu [5]-[7]. Với kỹ thuật học sâu, nhiều thông tin phức tạp dễ dàng được trích xuất – được học – trực tiếp từ dữ liệu thô. Đặc điểm này khiến cho học sâu được đánh giá là một phương pháp rất thành công trong việc học các đặc trưng trong dữ liệu phức tạp và cho kết quả chính xác cao. Tuy nhiên, để đảm bảo sự thành công của giải pháp tiếp cận sử dụng học sâu, một yêu cầu bắt buộc đó là cần một lượng dữ liệu đầu vào lớn và chứa đựng thông tin phong phú về vấn đề cần giải quyết [19], [20]. Một rào cản nữa của học sâu đó chính là

Tác giả liên lạc: Phạm Văn Sự,
Email: supv@ptit.edu.vn
Đến tòa soạn: 9/2020, chỉnh sửa: 11/2020, chấp nhận đăng: 12/2020.

thời gian cần thiết thực hiện huấn luyện cho mạng học sâu thường khá dài. Dù với sự hỗ trợ của phần cứng như GPU, thời gian cần thiết huấn luyện cho một bài toán mới với lượng dữ liệu lớn cũng phải kéo dài ít nhất vài ngày cho đến một tuần [19].

Ngoài việc cần đáp ứng nhu cầu rút ngắn thời gian đưa vào sử dụng của mạng, có rất nhiều bài toán ở một phạm vi cụ thể bó hẹp hơn việc có được lượng dữ liệu lớn để áp dụng một cách trực tiếp kỹ thuật học sâu là điều khó khăn. Vấn đề này có thể được khắc phục nhờ kỹ thuật học chuyển đổi [21], [22]. Học chuyển đổi là một dạng thức học máy trong đó thực hiện trích rút kiến thức đã học được từ một hoặc một số bài toán để rút ngắn thời gian và tăng hiệu quả giải quyết một bài toán khác có tính tương đồng.

Hành vi gian lận trong thi cử là một vấn đề nhạy cảm và phức tạp [23], [24]. Việc giám sát, tìm cách giảm nhỏ và tiến tới loại bỏ nhằm nâng cao chất lượng đào tạo trong các cơ sở giáo dục là việc làm hết sức cần thiết. Một số cơ sở giáo dục đã bước đầu lắp đặt các camera quan sát [25], [26]. Tuy nhiên, đây là một bài toán có sự thách thức lớn đòi hỏi nguồn nhân lực lớn và cần được đào tạo khi tiếp cận theo cách theo dõi thủ công. Trong bài báo này, nhóm nghiên cứu đề xuất giải pháp áp dụng học chuyển đổi nhằm phát hiện các hành vi gian lận trong phòng thi một cách tự động. Bằng cách sử dụng học chuyển đổi, giải pháp tận dụng tính ưu việt của các mạng học sâu đã được huấn luyện thuần thục áp dụng cho một lĩnh vực cụ thể vốn còn rất ít dữ liệu. Giải pháp đề xuất cho thấy tiết kiệm được thời gian huấn luyện, nhưng vẫn đảm bảo tính chính xác hứa hẹn là một giải pháp khả thi và có tính áp dụng cao.

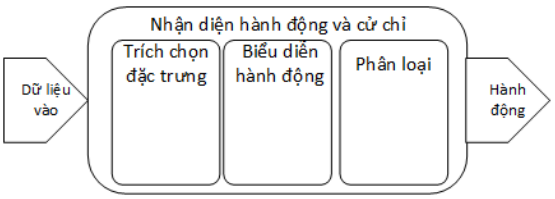
Phần còn lại của bài báo được tổ chức thành bốn phần. Phần II trình bày giải pháp đề xuất sử dụng học chuyển đổi để thực hiện nhận điện hành động và cử chỉ trong phòng thi. Phần III cung cấp các kết quả mô phỏng và các thảo luận. Cuối cùng, chúng tôi kết luận bài báo trong phần IV.

II. GIẢI PHÁP ĐỀ XUẤT

A. Cấu hình mạng học sâu cho bài toán nhận dạng hành vi gian lận trong phòng thi

Tương tự như một hệ thống nhận dạng ảnh, sơ đồ tổng quát của hệ thống nhận dạng hành động và cử chỉ được trình bày trong Hình 1. Một hệ thống nhận dạng hành động và cử chỉ về cơ bản gồm ba bước: trích xuất/học các đặc trưng; biểu diễn các hành động; và phân lớp các hành động. Mỗi một bước đều có một vai trò quan trọng trong việc nâng cao độ chính xác của việc nhận diện.

Học sâu có thể tăng khả năng mô tả dữ liệu phức tạp thông qua một số lớp biểu diễn. Thành công đầu tiên của học sau trong lĩnh vực thị giác máy tính được biết đến vào năm 2012, trong đó bài toán phân loại ảnh được giải quyết bằng cách xây dựng một mạng tích chập (CNN), thực hiện huấn luyện với 1,2 triệu bức ảnh độ phân giải cao và phân loại ảnh theo 1000 lớp [27]. Từ sau thành công đầu tiên, rất nhiều nghiên cứu trong lĩnh vực thị giác máy đã được đề xuất với cách tiếp cận học sâu [28]-[33].



Hình 1: Sơ đồ tổng quát hệ thống nhận dạng hành động và cử chỉ

Sơ đồ minh họa việc áp dụng học sâu vào bài toán nhận diện hành động và cử chỉ được trình bày trong Hình 2. Trong sơ đồ, một số lớp ẩn thực hiện mô hình hóa mối quan hệ phi tuyến, đầu ra của một lớp là đầu vào của lớp tiếp theo. Tại mỗi lớp, một mối quan hệ hàm phức tạp được học và hình thành một phân tầng biểu diễn thông tin về đối tượng, lớp sau trừu tượng/tổng quát hơn lớp trước [22].

Hình 2: Minh họa giải pháp học sâu giải quyết bài toán nhận diện hành động

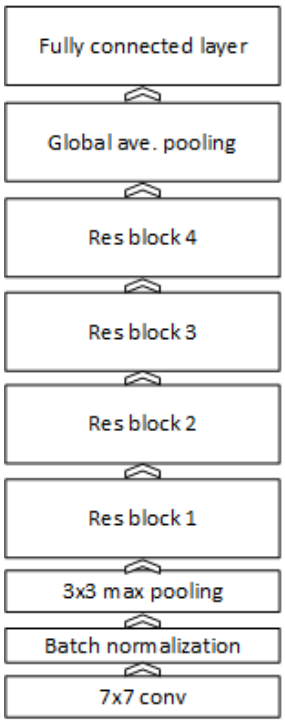
Các khối cấu thành trong mạng học sâu có thể được xây dựng từ nhiều phương thức khách nhau như: mạng tin sâu (DBN), máy Boltzman (BM), mạng nơ-ron sâu (DNN), mã hóa tự động (AE), mạng tính chập (CNN), mạng nơ-ron hồi quy (RNN), mạng với phần tử nhớ dài hạn – ngắn hạn (LSTM), ... Trong đó các nghiên cứu cho thấy các mạng CNN, RNN, và LSTM tỏ ra thích hợp hơn với bài toán nhận diện hành động.

Trong nghiên cứu này, chúng tôi sử dụng lớp mạng CNN làm cơ sở cho giải pháp, cụ thể sử dụng mạng ResNet-18 với sơ đồ trình bày trong Hình 3 [34]. Mạng ResNet được cấu thành bởi các khối hạt nhân chính có cấu trúc đặc biệt trong đó mỗi khối nội tại có liên kết rút ngắn (còn gọi là liên kết nội) được trình bày trong Hình 4 [34]. Với liên kết rút ngắn này, đầu vào của khối trước có thể truyền nhanh hơn sang các khối tiếp sau.

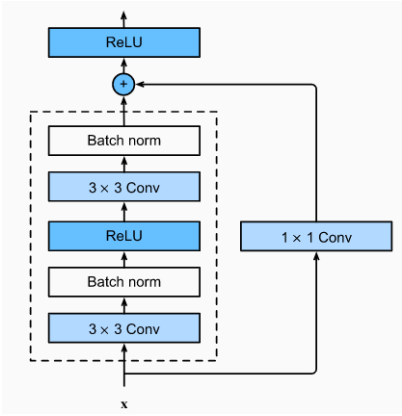
Hai lớp đầu tiên của ResNet tương tự với mạng GoogLeNet [34]: gồm một lớp tích chập 7x7 với bước dịch bằng 2 cho ra 64 kênh ra; theo sau là một lớp chọn phần tử lớn nhất (max pooling) 3x3 với bước dịch bằng 2. Tuy nhiên khác với GoogLeNet, sau mỗi lớp tích chập, một lớp chuẩn hóa theo nhóm được sử dụng.

Tiếp đến ResNet-18 sử dụng 4 mô-đun được tạo bởi các khối nội. Cuối cùng một lớp chọn trung bình toàn cục được thêm vào trước khi cho kết quả qua một lớp kết nối đầy đủ.

Các mạng ResNet khác nhau có thể dễ dàng đạt được bằng thay đổi số kênh đầu ra và số lớp khối nội. Với kiến trúc đơn giản, dễ dàng thay đổi khiến cho mạng ResNet được triển khai nhanh chóng và sử dụng rộng rãi. Đây cũng là lý do chính mà nhóm nghiên cứu xem xét và lựa chọn cấu hình mạng này.



Hình 3: Sơ đồ giản lược cấu hình mạng ResNet-18



Hình 4: Sơ đồ cấu trúc khối nội cấu thành của mạng ResNet

B. Chuẩn bị dữ liệu

Để thực hiện huấn luyện cho mạng học sâu, trong nghiên cứu này chúng tôi sử dụng bộ dữ liệu HMDB51 [37]. Trong nghiên cứu này, nhóm nghiên cứu tiếp cận bài toán theo hướng 2D. Tập dữ liệu video được thực hiện tiền xử lý bằng cách trích cắt khung chính với sự hỗ trợ của thư viện Yolov3 [38] thu được hơn 2,5 triệu ảnh tương ứng với 51 hành động. Tập ảnh được trộn ngẫu nhiên, phân chia thành 5 tập con và được sử dụng để thực hiện huấn luyện và đánh giá chéo.

Mặc dù các tập cơ sở dữ liệu hành động phong phú như KTH [36], UCF50 [37], … nhưng việc tìm tập dữ liệu cho các hành động vi phạm trong phòng thi hoàn toàn không dễ dàng. Thêm nữa, đây là tập dữ liệu có tính nhạy cảm. Theo hiểu biết của tác giả cho đến nay chưa có tập dữ liệu công khai thuộc chủ đề này. Ngoài ra, việc có được tập dữ liệu đủ lớn về chủ đề này hiện nay để có thể áp dụng trực tiếp mạng học sâu là điều rất khó.

Trong quá trình nghiên cứu tìm hiểu, nhóm nghiên cứu được sự cho phép của Trung tâm Khảo thí và Đảm bảo chất lượng tại Học viện Bưu chính Viễn thông đã thực hiện thu thập dữ liệu thụ động. Dữ liệu được thu thập một cách kín đáo và không có sự hợp tác của người học. Tập dữ liệu thô có tổng thời lượng khoảng 1,5 giờ đồng hồ được thu thập của nhiều sinh viên khác nhau với 8 nhóm hành động chính: sử dụng tài liệu trong lòng bàn tay để trên bàn (IPF), sử dụng tài liệu để trên tay để dưới gầm bàn (IPU), sử dụng tài liệu dưới giấy viết (IPO), quay trái sang nhìn/trao đổi (RL), quay phải sang nhìn/trao đổi (RR), quay sau phải để nhìn/trao đổi (RBR), quay sau trái để nhìn/trao đổi (RBL), nhổm người về trước nhìn/trao đổi (UF). Dữ liệu ảnh được trích xuất khung với sự hỗ trợ của thư viện Yolov3. Các khung hình ứng với các hành động thuộc nhóm hành động được chọn và đánh nhãn thủ công thu được khoảng 1640 khung hình tương ứng cho 8 nhóm hành động. Cụ thể, số lượng khung hình của mỗi nhóm hành động được trình bày trong Bảng 1. Một số hành động điển hình được minh họa trong Hình 5.

Bảng 1: Số lượng khung hình của mỗi nhóm hành động trong dữ liệu thu thập

Lớp hành động	Số lượng khung hình
IPF	229
IPU	171
IPO	190
RL	185
RR	236
RBL	174
RBR	252
UF	203

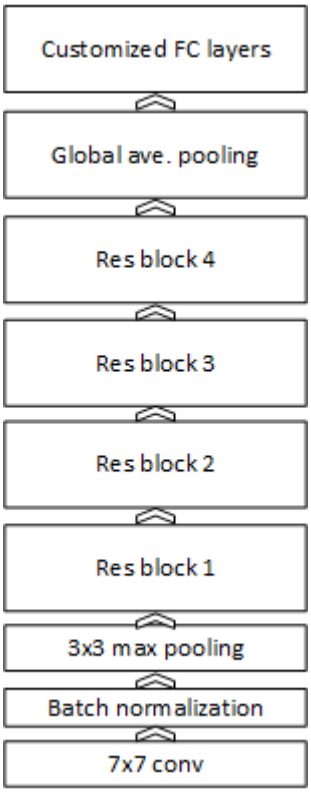
Hình 5: Minh họa một số hành động gian lận trong phòng thi

C. Giải pháp áp dụng học chuyển đổi

Để khắc phục việc thiếu dữ liệu cho mạng học sâu như đề cập ở trên, nhóm nghiên cứu xem xét việc áp dụng học chuyển đổi dựa trên đặc trưng nhằm chuyển đổi mạng ResNet sau khi đã được huấn luyện thuần thục để cá thể hóa cho bài toán nhận diện hành động gian lận trong phòng thi. Phương pháp học chuyển đổi dựa trên đặc trưng cho phép việc học chuyển đổi có thể thực hiện trên không gian đặc trưng được trừu tượng hóa thay vì phụ thuộc vào không gian ảnh thô đầu vào [22].

Ý tưởng cơ bản của học chuyển đổi dựa trên đặc trưng là coi các lớp phía trước của mạng, trừ một số lớp cuối cùng,

như các lớp biểu diễn đặc trưng. Với các bài toán có sự tương đồng, thay vì phải huấn luyện lại từ đầu thì chúng ta chỉ cần cá thể hóa thích hợp một số lớp cuối cùng [22]. Dựa trên ý tưởng đó, nhóm nghiên cứu thay đổi xây dựng lớp kết nối đầy đủ cuối cùng để phù hợp với tập các hành động quan tâm. Cụ thể, một lớp kết nối đầy đủ mới với số nút trong lớp phù hợp số lớp hành động được thêm vào. Sơ đồ minh họa mạng đề xuất trình bày trong Hình 6.



Hình 6: Sơ đồ giản lược kiến trúc đề xuất áp dụng học chuyển đổi

III. KẾT QUẢ THỰC NGHIỆM VÀ THẢO LUẬN

Để so sánh đánh giá kết quả, tập dữ liệu HMDB51 và tập dữ liệu thu thập được được sử dụng. Quá trình thực nghiệm và khảo sát sử dụng ngôn ngữ Python với thư viện Pytorch trên máy trạm với sự hỗ trợ của thiết bị GPU.

Đầu tiên, để đánh giá chất lượng của phương pháp đề xuất tập dữ liệu thu được từ bộ dữ liệu HMDB51 như mô tả trong phần II được sử dụng để huấn luyện và đánh giá với mô hình mạng ResNet-18. Sau khi mạng được huấn luyện thuần thục thể hiện thông qua các đánh giá mạng ổn định, lớp kết nối đầy đủ cuối cùng được cấu hình lại như đã trình bày. Tiếp đến bộ dữ liệu thu thập được được trộn ngẫu nhiên và chia thành ba phần với tỷ lệ 70%, 15% và 15% tương ứng cho phần tinh chỉnh, đánh lại và kiểm tra.

Để thực hiện đối sánh và đánh giá lợi ích của học chuyển đổi, toàn bộ dữ liệu thu thập được cũng được thực hiện trộn và chia như trên sau đó được đưa vào huấn luyện và đánh từ đầu cho mạng ResNet-18.

Kết quả đánh giá về độ chính xác cho thấy, với giải pháp đề xuất độ chính xác tính trung bình cho các lớp hành động đạt 88.35% trong khi việc thực hiện sử dụng dữ liệu huấn luyện từ đầu chỉ đạt khoảng 64.8%. Sở dĩ việc sử dụng dữ liệu huấn luyện lại từ đầu không đạt kết quả cao có thể bởi vì lượng dữ liệu quá nhỏ cho mỗi lớp hành động. Đặc biệt như quan sát trong Hình 5, có một số hành động có sự tương đồng cao. Do đó, khi dữ liệu không đủ lớn, việc học và biểu diễn chúng của mạng chưa đủ mạng để phân biệt được dẫn đến độ chính xác thấp.

Kết quả ma trận nhầm lẫn của phương pháp đề xuất được trình bày trong Bảng 2. Quan sát kết quả từ bảng chúng ta thấy rằng mặc dù phương pháp đề xuất có độ chính xác cao nhưng vẫn có môt số hành động có sự nhầm lẫn khá cao chẳng hạn như hành động sử dụng tài liệu trong tay để trước mặt dễ bị nhầm đến khoảng hơn 20% thành sử dụng tài liệu dưới giấy trước mặt. Bằng kiểm nghiệm quan sát trên minh họa Hình 5 thì thấy kết quả này hoàn toàn dễ hiểu vì hai hành động này có sự tương đồng đáng kể. Ngoài ra các hành động quay sang trái và quay về phía sau bên trái cũng có sự nhầm lẫn cao, tương tự cho hành động quay về phía bên phải. Điều này cũng là do những hành động này có sự tương đồng đáng kể.

Bảng 2: Kết quả đánh giá ma trận nhầm lẫn

		Dự đoán							
		IPF	IPU	IPO	RL	RR	RBL	RBR	UF
Thực tế	IPF	175	0	54	0	0	0	0	0
	IPU	0	171	0	0	0	0	0	0
	IPO	21	0	169	0	0	0	0	0
	RL	2	0	0	167	0	12	2	2
	RR	0	0	1	0	211	0	21	3
	RBL	0	0	1	17	3	147	0	6
	RBR	12	0	3	0	23	0	214	0
	UF	0	0	0	1	4	2	1	195

Giải pháp để giảm sự nhầm lẫn giữa các hành động này có thể được thực hiện bằng cách tiếp cận 3D trong đó tận dụng đặc tính chuỗi thời gian của hành động và áp dụng các kiến trúc mạng RNN hoặc LSTM thay vì CNN như hiện nay. Giải pháp này nhóm nghiên cứu xin trình bày trong nghiên cứu trong thời gian tới.

Kết quả quan sát về mặt thời gian cho thấy thời gian từ lúc bắt đầu thực hiện tinh chỉnh cho đến lúc kết quả đánh giá ổn định của giải pháp đề xuất là 2,23 phút trong khi thời gian để có kết quả đánh giá ổn định khi thực hiện huấn luyện mạng từ đầu là 27,51 phút. Các kết quả thời gian là giá trị trung bình của 150 lần thử nghiệm. Kết quả này cho thấy độ lợi rõ rệt về mặt thời gian khi áp dụng học chuyển đổi. Cũng cần nhấn mạnh rằng, nếu tính tổng thời gian huấn luyện dữ liệu cho bài toán gốc với bộ dữ liệu HMDB51 thì thời gian là 8,21 giờ. Tuy nhiên, khi quan tâm đến sự hạn chế về mặt dữ liệu cho bài toán áp dụng và khả năng về sự dịch chuyển kiến thức học được sẵn có sang một bài toán mới thì rõ ràng độ lợi về độ chính xác và thời gian giải quyết bài toán là rất đáng xem xét và có ý nghĩa hết sức thực tế.

IV. KẾT LUẬN

Trong bài báo này, chúng tôi đã đề xuất một phương pháp tiếp cận sử dụng học chuyển đổi để giải quyết bài toán nhận diện hành vi gian lận trong phòng thi. Giải pháp sử dụng học chuyển đổi dựa trên đặc trưng nhằm tận dụng tính ưu việt của mạng học sâu đã được huấn luyện thuần thục với một mục tiêu có nét tương đồng. Với giải pháp đề xuất, chất lượng theo khía cạnh độ chính xác được cải thiện đáng kể dù cơ sở dữ liệu nhỏ vốn dĩ không thích hợp cho việc áp

dụng mạng học sâu. Không những thế, thời gian đưa vào áp dụng mạng cho bài toán cũng được rút ngắn. Từ đó cho thấy, giải pháp đề xuất hứa hẹn có tính thực tiễn cao.

TÀI LIỆU THAM KHẢO

[1] Schuldt, Laptev and Caputo, "Recognizing Human Actions: A local SVM Approach," in Proc. ICPR'04, Cambridge, UK, 2004.

[2] C. Chen, B. Zhang, Z. Hou, J. Jiang, M. Liu, and Y. Yang. Action recognition from depth sequences using weighted fusion of 2d and 3d auto-correlation of gradients features. Multimedia Tools and Applications, pages 1–19, 2016

[3] T. Eleni. Gesture recognition with a convolutional long short term memory recurrent neural network. In ESANN, 2015.

[4] C. Feichtenhofer, A. Pinz, and A. Zisserman. Convolutional two-stream network fusion for video action recognition. In CVPR, 2016

[5] W. Ouyang, X. Chu, and X. Wang. Multi-source deep learning for human pose estimation. CVPR, pages 2337–2344, 2014

[6] Pigou L., Dieleman S., Kindermans PJ., Schrauwen B. (2015) Sign Language Recognition Using Convolutional Neural Networks. In: Agapito L., Bronstein M., Rother C. (eds) Computer Vision - ECCV 2014 Workshops. ECCV 2014. Lecture Notes in Computer Science, vol 8925. Springer, Cham

[7] Jie Huang, Wengang Zhou, Qilin Zhang, Houqiang Li, Weiping Li, Video-based Sign Language Recognition without Temporal Segmentation, arXiv:1801.10111 Medeley generated error.

[8] Crispim-Junior, C. F., Ma, Q., Fosty, B., Romdhane, R., Bremond, F., & Thonnat, M. (2015). Combining Multiple Sensors for Event Detection of Older People Health Monitoring and Personalized Feedback using Multimedia Data (pp. 179-194): Springer

[9] Foroughi, H., Yazdi, H. S., Pourreza, H., & Javidi, M. (2008). An eigenspace-based approach for human fall detection usingintegrated time motion image and multi-class support vector machine. Paper presented at the Intelligent Computer Communication and Processing, 2008. ICCP 2008. 4th International Conference on

[10] Kuo, Y.-M., Lee, J.-S., & Chung, P.-C. (2010). A visual context-awareness-based sleeping-respiration measurement system. Information Technology in Biomedicine, IEEE Transactions on, 14(2), 255-265

[11] Ahmad Jalal; Maria Mahmood; Abdul S. Hasan, Multi-features descriptors for Human Activity Tracking and Recognition in Indoor-Outdoor Environments, 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST)

[12] Y. Tang, Y. Tian, J. Lu, J. Feng and J. Zhou, "Action recognition in RGB-D egocentric videos," 2017 IEEE International Conference on Image Processing (ICIP), Beijing, 2017, pp. 3410-3414, doi: 10.1109/ICIP.2017.8296915.

[13] Jalal, A., Kamal, S. & Azurdia-Meza, C.A. Depth Maps-Based Human Segmentation and Action Recognition Using Full-Body Plus Body Color Cues Via Recognizer Engine. J. Electr. Eng. Technol. 14, 455–461 (2019). https://doi.org/10.1007/s42835-018-00012-w

[14] Q. V. Le, W. Y. Zou, S. Y. Yeung and A. Y. Ng, "Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis," CVPR 2011, Providence, RI, 2011, pp. 3361-3368, doi: 10.1109/CVPR.2011.5995496.

[15] P. Martin, J. Benois-Pineau, R. Péteri and J. Morlier, "Sport Action Recognition with Siamese Spatio-Temporal CNNs: Application to Table Tennis," 2018 International Conference on Content-Based Multimedia Indexing (CBMI), La Rochelle, 2018, pp. 1-6, doi: 10.1109/CBMI.2018.8516488.

[16] C. Ionescu, D. Papava, V. Olaru and C. Sminchisescu, "Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 7, pp. 1325-1339, July 2014, doi: 10.1109/TPAMI.2013.248.

[17] Wang, H., Kläser, A., Schmid, C., et al.: 'Dense trajectories and motion boundary descriptors for action recognition', Int. J. Comput. Vis., 2013, 103, pp. 60–79

[18] Wang, H., Schmid, C.: 'Action recognition with improved trajectories'. Proc.IEEE Int. Conf. on Computer Vision, 2013

[19] Maryam Koohzadi, Nasrollah Moghadam Charkari, Survey on deep learning methods in human action recognition, Special Section: Deep Learning in Computer Vision, IET Comput. Vis., 2017, Vol. 11 Iss. 8, pp. 623-632

[20] Zhu, F., Sha, L., Xie, J., and Fang, Y., From handcrafted to learned representations for human action recognition: A survey. Image and Vision Computing, 2016

[21] A. B. Sargano, X. Wang, P. Angelov and Z. Habib, "Human action recognition using transfer learning with deep representations," 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, 2017, pp. 463-469, doi: 10.1109/IJCNN.2017.7965890.

[22] Qiang Yang, Yu Zhang, Wenyuan Dai, and Sinno Jialin Pan, Transfer Learning, CUP. 2020

[23] Trần Đức Viên, Gian lận và thi cử: Lo âu về một ngày mai, Báo Tia sáng, Tháng 12, 2019

[24] Quỳnh Nguyễn, Cảnh giác gian lận trong thi cử, Báo nhân dân. Tháng 8,2020

[25] Hà Phượng, Chống gian lận thi cử: 100% các phòng thi đều được lắp camera, Pháp luật Online, Tháng 5,2019

[26] idp.com

[27] Krizhevsky, A., Sutskever, I., Hinton, G.E.: 'ImageNet classification with deep convolutional neural networks'. Advances in Neural Information Processing Systems, 2012

[28] Le, Q.V.: 'Building high-level features using large scale unsupervised learning'. 2013 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), 2013

[29] Peng, X., Zou, C., Qiao, Y., et al.: 'Action recognition with stacked fishervectors'. Computer Vision–ECCV 2014, 2014, pp. 581–595

[30] Rifai, S., Bengio, Y., Courville, , et al.: 'Disentangling factors of variation for facial expression recognition'. Computer Vision–ECCV 2012, 2012, pp. 808–822

[31] Ciresan, D., Meier, U., Schmidhuber, J.: 'Multi-column deep neural networks for image classification'. 2012 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2012

[32] Zeiler, M.D.: 'Hierarchical convolutional deep learning in computer vision' (New York University, 2013)

[33] Mnih, V., Kavukcuoglu, K., Silver, D., et al.: 'Human-level control through deep reinforcement learning', Nature, 2015, 518, (7540), pp. 529–533

[34] Aston Zhang and Zachary C. Lipton and Mu Li and Alexander J. Smola, Dive into Deep Learning, https://d2l.ai

[35] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre. HMDB: A Large Video Database for Human Motion Recognition. ICCV, 2011

[36] Christian Schuldt, Ivan Laptev and Barbara Caputo, "Recognizing Human Actions: A Local SVM Approach", in Proc. ICPR'04, Cambridge, UK

[37] http://vision.eecs.ucf.edu/data.html

[38] Redmon, Joseph and Farhadi, Ali, YOLOv3: An Incremental Improvement, https://arxiv.org/abs/1804.02767, 2018

APPLICATION OF TRANSFER LEARNING ON DETECTING EXAMINIATION CHEATING ACTION

Abstract: Human action and gesture recognition (HAR/HGR) has been an attractive research topic recently. By applying successfully deep learning to HAR, many aspects of daily life actions in sport, leisure, medical care, … have been recognized with significantly correctness. In this work, we propose a solution which combines transfer learning and deep learning to solve the case of recognizing the misbebaviour human actions in exams where the available data is limited. The evaluations on the collected data show that the proposed approach is a promising method. The solution can exploit the goodness of deep learning and leverage the short cut of transfer learning while still achieving the high performance.

Keywords: Examination cheating behavior, cheating action, transfer learning, deep learning (DL), convolutional neural network (CNN), human gesture recognition (HGR), human action recognition (HAR***)***

Phạm Văn Sự tốt nghiệp ngành Điện tử Viễn thông tại Đại học Bách Khoa Hà Nội (HUST) năm 1999, cao học ngành Kỹ thuật Điện – Điện tử tại Đại học Thông tin Liên lạc (ICU) Hàn Quốc năm 2004. Tác giả hiện là giảng viên Bộ môn Xử lý tín hiệu & Truyền thông, Khoa Kỹ thuật Điện tử I, Học viện Công nghệ Bưu chính Viễn thông. Các hướng nghiên cứu chính của tác giả gồm: Thiết kế mạch tích hợp số và tương tự, Xử lý ảnh, Xử lý tiếng nói, Thị giác máy tính, Thông tin số.