

ỨNG DỤNG HỌC SÂU TRONG NHẬN DẠNG CỬ CHỈ TAY

Nguyễn Trọng Khánh, Phạm Văn Cường

Khoa Công nghệ thông tin,
Học viện Công nghệ Bưu chính Viễn thông

Tóm tắt: Bài báo đề xuất một phương pháp nhận dạng cử chỉ tay dựa trên phương pháp học sâu. Dữ liệu cử chỉ tay được thu thập từ hai cảm biến: cảm biến gia tốc và con quay hồi chuyển. Dữ liệu này sau đó được tiền xử lý bằng cách lọc nhiễu và phân đoạn. Tiếp đến dữ liệu được đưa vào mạng học sâu để trích chọn đặc trưng và phân loại cử chỉ. Nhóm tác giả thử nghiệm với kiến trúc mạng học sâu đang được đánh giá cao trong nhận dạng cử chỉ tay động là mạng BaseLineCNN. Từ kết quả thực nghiệm có thể kết luận phương pháp học sâu nói chung và mạng BaselineCNN nói riêng cho kết quả tốt trong nhận dạng cử chỉ tay, với dữ liệu được thu thập từ cảm biến trên các thiết bị đeo tay phổ biến hiện nay.

Từ khóa: Nhận dạng cử chỉ tay, học sâu, BaselineCNN, hand gesture recognition.

I. MỞ ĐẦU

Bài toán nhận dạng cử chỉ tay đã được quan tâm và nghiên cứu từ những năm đầu của thế kỷ trước. Mục tiêu của bài toán là nhận dạng được hình dáng (cử chỉ tay tĩnh) hoặc hoạt động (cử chỉ tay động) của cử chỉ tay, bao gồm ngón tay, bàn tay, cánh tay để từ đó đưa ra các thông tin hữu ích trong tương tác người máy.

Trước đây, bài toán nhận dạng cử chỉ tay thường được tiếp cận theo hướng áp dụng thị giác máy tính. Theo đó các cử chỉ ngón tay, bàn tay sẽ được chụp lại và sử dụng các mô hình học máy để huấn luyện và nhận dạng. Ưu điểm phương pháp này là kết quả nhận dạng cao và ổn định. Tuy nhiên phương pháp này thường chỉ áp dụng hiệu quả trong bài toán nhận dạng cử chỉ tay tĩnh. Phạm vi thực hiện cử chỉ bị giới hạn trong vùng nhìn của camera nên chưa cho phép cử chỉ của người được nhận dạng mọi nơi, mọi lúc. Ngoài ra, cách tiếp cận thị giác máy tính thường đòi hỏi tài nguyên tính toán khá lớn.

Gần đây với sự phát triển mạnh mẽ của lĩnh vực vi điện tử, các cảm biến được tích hợp ngày càng nhiều trong các thiết bị IoT và di động. Các cảm biến như cảm biến gia tốc, cảm biến con quay hồi chuyển cho phép chúng ta thu thập được nhiều thông tin hữu ích phục vụ cho việc dự đoán các cử chỉ tay động, đặc biệt liên quan đến hoạt động của cổ tay, cánh tay. Phần lớn thiết bị đeo tay hiện nay đều được trang bị với những khối đo lường này, ví dụ như đồng hồ thông minh hoặc các vòng đeo tay theo dõi sức khỏe. Bằng cách kết hợp những thiết bị đeo tay này trong ngữ cảnh một hệ thống IoT như nhà thông minh, chúng ta có thể xây dựng

một hệ thống nhận dạng cử chỉ tay động phù hợp cho các bài toán khác nhau, ví dụ nâng cao trải nghiệm người dùng trong điều khiển nhà thông minh, hỗ trợ người khiếm thị trong giao tiếp, hoặc hỗ trợ trong tương tác người máy. Nhiều giải pháp được đưa ra với độ chính xác cao, ví dụ những nghiên cứu của Yang và đồng nghiệp [1], Monisha và đồng nghiệp [2], Lei và đồng nghiệp [3].

Đã có một số nghiên cứu ứng dụng thông tin thu thập từ các thiết bị đeo tay phổ biến [4] [5] [6] [7] [8] [9] [10], như sử dụng đồng hồ Apple Watch [4], Huawei Watch [10] để nhận dạng hoạt động của bàn tay, cổ tay. Tuy nhiên những hệ thống này chưa được tích hợp đầy đủ vào nền tảng IoT và việc triển khai trong thực tế còn gặp nhiều khó khăn. Một số nghiên cứu khác cho phép thực thi việc nhận dạng trực tiếp trên điện thoại thông minh, tuy nhiên lại gây khó khăn cho người dùng trong ghi nhớ cử chỉ.

Bài báo này đề xuất một phương pháp nhận dạng cử chỉ tay động liên quan đến hoạt động của cánh tay sử dụng các cảm biến được tích hợp sẵn trong thiết bị đeo (ví dụ như đồng hồ thông minh) phù hợp với cấu hình hạn chế của các nền tảng IoT như nhà thông minh. Mục tiêu của nghiên cứu hướng tới là nhận dạng hoạt động của tay (ví dụ gơ tay lên, xuống, sang trái, phải ...) thông qua dữ liệu được thu thập bởi các thiết bị đeo thông minh tại cổ tay.

Phần còn lại của bài báo được tổ chức như sau. Phần 2 giới thiệu một số nghiên cứu liên quan. Phần 3 giới thiệu phương pháp nhận dạng cử chỉ tay dựa trên học sâu. Phần 3 là thực nghiệm và phân tích kết quả. Phần 4 là kết luận và hướng phát triển trong tương lai.

II. NGHIÊN CỨU LIÊN QUAN

Nghiên cứu chủ yếu tập trung vào các giải pháp nhận dạng cử chỉ tay động, dựa trên thiết bị đeo tại cổ tay, do đó trong phần này, một số nghiên cứu liên quan sử dụng dữ liệu cảm biến để nhận dạng cử chỉ tay liên quan sẽ được đề cập.

Về cơ bản, nhận dạng cử chỉ tay động dựa trên cảm biến thường dựa trên dữ liệu trong không gian 3 chiều biến thiên theo thời gian. Các tính năng như góc uốn của ngón tay, hướng và vị trí tuyệt đối của bàn tay thường ở trong không gian 3 chiều, nên nó cần các thông tin về độ sâu, hướng, gia tốc ...

Gần đây, một trong những cảm biến thường được dùng để thu thập dữ liệu này là gia tốc kế. Gia tốc kế cung cấp dữ liệu trong không gian 3 chiều với 3 trục x, y, z; gia tốc của cảm biến theo các chiều được tính toán riêng biệt, do đó chúng ta có thể sử dụng số liệu này để xác định được hướng và chuyển động của cảm biến. Ví dụ, gắn cảm biến gia tốc vào dây đeo cổ tay, ta có thể xác định được khi nào và hướng cổ tay chuyển động hoặc quay. Đã có nhiều

Tác giả liên hệ: Nguyễn Trọng Khánh,

Email: khanht@ptit.edu.vn

Đến tòa soạn: 5/10/2021, chỉnh sửa: 12/11/2021, chấp nhận đăng: 12/12/2021.

nghiên cứu sử dụng cảm biến gia tốc để thu thập dữ liệu cho việc nhận dạng cử chỉ tay động, ví dụ như nghiên cứu của Liu và đồng nghiệp [12], Leung và đồng nghiệp [5], Kwon và đồng nghiệp [13], Hong và đồng nghiệp [7] ... Bên cạnh chỉ sử dụng cảm biến gia tốc, có một số nghiên cứu đề xuất kết hợp dữ liệu 3 chiều của cảm biến gia tốc với dữ liệu 3 chiều của cảm biến con quay hồi chuyển [13] [8] [14]. Sử dụng đồng thời cả hai cảm biến cho phép ta kết hợp dữ liệu gia tốc và hướng quay trên từng trục, từ đó cải thiện độ chính xác của việc nhận dạng cử chỉ tay động.

Ngoài ra, kết hợp 2 cảm biến trên với dữ liệu 3 chiều của từ kế (magnetometer) ta sẽ thu được dữ liệu cảm biến trong 9 trục khác nhau, hỗ trợ thông tin quay, hướng và độ nghiêng của cử chỉ. Cảm biến từ trường là một thiết bị hệ thống vì cơ điện từ quy mô nhỏ cho phép đo mức độ của từ trường. Một số nghiên cứu đã thử nghiệm kết hợp từ kế để thu thập dữ liệu cử chỉ tay động, ví dụ như Ordonez và đồng nghiệp [14], Senevirante và đồng nghiệp [15].

Có 3 phương pháp phân tích chính được áp dụng để phân loại hoặc nhận dạng cử chỉ tay: Phương pháp so khớp mẫu (Template-based methods); học máy cổ điển với các đặc trưng trích chọn bằng tay (Handcrafted features); và học sâu.

Hướng tiếp cận đầu tiên, phương pháp so khớp mẫu [7], dựa trên ý tưởng tính toán khoảng cách giữa tập cử chỉ mẫu và tập cần nhận dạng. Khoảng cách này sau đó được so sánh với một ngưỡng được đặt trước để xác định xem 2 bộ cử chỉ có phải là một hay không. Có nhiều loại khoảng cách được sử dụng, ví dụ như khoảng cách hình học [28], khoảng cách Euclid [29]. Một điển hình của hướng tiếp cận này là phương pháp Xoắn thời gian động - Dynamic Time Warping (DTW) [30]. DTW xác định cử chỉ tay động gần nhất bằng cách đo khoảng cách DTW giữa đặc trưng đầu vào và đặc trưng từng cử chỉ mẫu. Trước khi tính toán khoảng cách, DTW được sử dụng để căn chỉnh hai dữ liệu thông qua việc làm cong với phương pháp phi tuyến để làm cho dữ liệu khớp với nhau. Do quá trình căn chỉnh này, các thuật toán này có thể đối phó với các biến dạng trong dữ liệu phụ thuộc thời gian gây ra bởi các tốc độ khác nhau. Các thuật toán nhận dạng dựa trên DTW thể hiện hiệu suất tốt cho nhận dạng cử chỉ tay động. Tuy nhiên, hiệu suất nhận dạng phụ thuộc rất nhiều vào chất lượng của các mẫu được chọn. Do đó, vì các thuật toán này yêu cầu tìm các mẫu tối ưu từ bộ dữ liệu đã chuẩn bị, nên khó có thể thực hiện việc học theo thời gian thực với độ chính xác cao, đây là một trong những yêu cầu quan trọng đối với nhận dạng cử chỉ động.

Hướng tiếp cận thứ 2 có thể được xem như là sự phát triển từ hướng thứ nhất: nhận dạng cử chỉ tay động với học máy cổ điển bằng các bộ phân loại, sử dụng đặc trưng trích chọn bằng tay. Các kỹ thuật khác nhau được áp dụng để trích chọn đặc trưng trong miền thời gian, tần số hoặc cả 2 cho mỗi lớp cử chỉ cần nhận dạng. Sau đó, các kỹ thuật phân loại sẽ được áp dụng để xác định lớp của cử chỉ. Các kỹ thuật thường được sử dụng để phân loại ví dụ như máy véc tơ hỗ trợ [27, 30, 7] và mạng thần kinh (không có lớp ẩn) [29, 1, 28], K láng giềng gần nhất [32]. Các kỹ thuật này cũng có thể được kết hợp cùng sử dụng, ví dụ kết hợp cây quyết định với mô hình phân cụm k và các mô hình Markov ẩn được sử dụng trong [32]; kết hợp của các máy véc tơ hỗ trợ và các mô hình Markov được sử dụng trong [24]. Việc sử dụng các đặc trưng trích chọn bằng tay thường tốn thời gian trong xác định đặc trưng sử dụng. Ngoài ra, các nghiên cứu cũng gặp nhiều khó khăn khi muốn mở rộng

số lượng cử chỉ, hoặc nhận dạng cử chỉ phức tạp hơn, vì các đặc trưng này thường là kết quả của các biến đổi toán học, không phải xuất phát từ các hành vi [22].

Thế hệ thứ ba được xuất phát từ sự phát triển của các công nghệ học sâu, trong đó các đặc trưng cử chỉ tay động được tự động trích xuất từ dữ liệu thô. Phương pháp này sử dụng các mạng nơ ron nhiều lớp xử lý thông tin phi tuyến tính để trích xuất và phân loại đặc trưng, được sắp xếp theo thứ bậc; đầu ra lớp này là đầu vào của lớp kế. Chẳng hạn, Kwonet và đồng nghiệp [11] đã trình bày một công nghệ học sâu trong nhận dạng cử chỉ tay động bằng cách sử dụng mô hình phân loại mạng thần kinh tích chập (CNN) với 17 lớp ẩn. Ordonez và đồng nghiệp [22] năm 2016 cũng đề xuất mạng học sâu DeepConvLSTM bao gồm các lớp tích chập và lớp hồi quy LSTM, có khả năng tự học các đặc trưng của cử chỉ tay động. Mạng học sâu này đã được thử nghiệm trên bộ dữ liệu OPPORTUNITY [31], được thu thập bởi 5 thiết bị đeo với 15 cảm biến (5 cảm biến gia tốc, 5 cảm biến con quay hồi chuyển và 5 cảm biến từ kế). Kết quả nhận dạng cử chỉ đề xuất cao (91.5%) nhưng số thiết bị đeo không phù hợp với điều kiện thực tế trong cuộc sống. DeepConvLSTM là mạng được đánh giá cao trong nhận dạng cử chỉ tay vì nó đã xét đến yếu tố biến thiên theo thời gian của cử chỉ tay động (tích hợp mạng hồi quy bộ nhớ dài ngắn).

Các kỹ thuật học sâu là phương pháp đang hứa hẹn sẽ giải quyết triệt để yêu cầu của nhận dạng từ các thiết bị đeo tay thông thường. Đầu tiên, hiệu suất nhận dạng và thời gian cải thiện hơn so với các phương pháp trước đó. Thứ hai, nhận dạng cử chỉ tay động dựa trên học sâu có khả năng phát hiện ra các đặc trưng chưa được khám phá gắn liền với các hành vi của tay người, từ các chuyển động đơn giản ở các lớp thấp hơn đến các chuyển động phức tạp hơn ở các lớp trên.

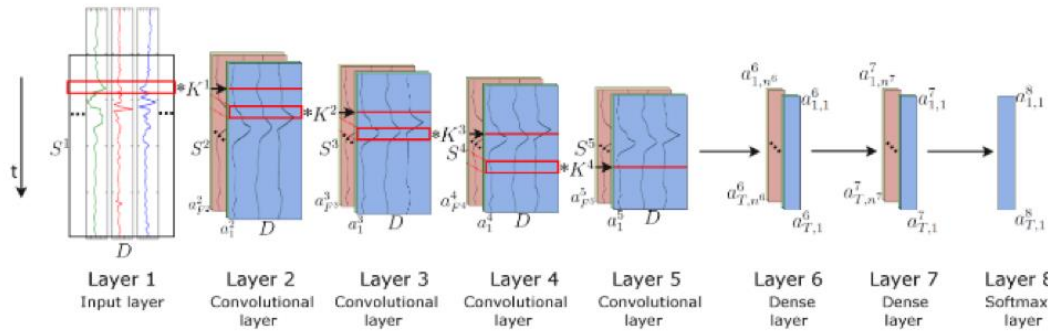
III. PHƯƠNG PHÁP ĐỀ XUẤT

Baseline CNN [14] là 1 mạng học sâu tích chập bao gồm các lớp tích chập 2 chiều, phi hồi quy và cuối cùng là lớp softmax dùng để phân loại hành động. Đầu tiên, dữ liệu cảm biến được tính toán qua 4 lớp tích chập để chiết xuất đặc trưng và trừu tượng hóa dữ liệu thành dạng bậc đặc trưng cao. Các lớp tích chập xử lý dọc theo trục thời gian, số lượng kênh cảm biến giống nhau ở các ảnh xạ đặc trưng đầu ra (feature map).

Hình 1 mô tả toán tử tích chập bằng ký hiệu '*', được tính dựa vào 1 nhân (kernel) có kích thước mô tả bằng hình chữ nhật màu vàng, với kích thước thực tế là 5×1 , bước nhảy bằng 1x1 và được đệm lề có hạn định (padding valid). Các đặc trưng ảnh xạ đầu ra ở mỗi bước tích chập này sau đó được tính qua hàm đơn vị tuyến tính chỉnh lưu (ReLU), hàm phi tuyến tính được định nghĩa là $\delta(x) = \max(0, x)$. Các lớp 6 và 7 là lớp phi tuyến tính dày đặc (dense layer), minh họa nhận thức đa lớp (multi-layer perception MLPs). Hàm kích hoạt cho mỗi đơn vị trong lớp dày đặc đầu tiên được tính dựa vào toàn bộ ảnh xạ đầu ra đặc trưng từ lớp tích chập cuối cùng. Mỗi đơn vị trong lớp dày đặc được thông qua hàm ReLU, với hàm phi tuyến tính được định nghĩa là $\delta(x) = \max(0, x)$ như trên. Đầu ra của mô hình là 1 lớp softmax, bản chất là 1 lớp dày đặc với hàm kích hoạt là softmax, biểu thị phân bố xác suất của các lớp phân loại.

Tóm lại, mô hình mạng được mô tả tóm tắt lại như sau:

$C(64) - C(64) - C(64) - C(64) - D(128) - D(128) - Sm$,
 với $C(x)$ minh họa 1 lớp tích chập với đầu vào là ảnh xạ đặc



Hình 1. Kiến trúc mạng BaselineCNN. Từ trái qua phải, dữ liệu bắt đầu từ cảm biến được thông qua 4 lớp tích chập để học các đặc trưng. 2 lớp dày đặc sau đó được tính qua 1 bước chuyển đổi phi tuyến tính, trong đó mang lại kết quả phân loại của lớp đầu ra với hàm hồi quy softmax ở bên phải. Độ dài của ảnh xạ đặc trưng ký hiệu là S_x và a_x minh họa hàm kích hoạt của lớp x .

trung, $D(y)$ là 1 lớp dày đặc với y đơn vị và S_m là lớp phân loại softmax. Chi tiết cấu trúc của mạng cho 2 mạng với kích thước đầu vào khác nhau được minh họa trong Bảng 1 và 2.

Trước khi bắt đầu quá trình huấn luyện, các đặc trưng được chuẩn hóa bởi Standardization, là một phương thức chuyển tỉ lệ đặc trưng về một miền để tính giảm gradient nhanh hơn, với công thức $z=(x-\mu)/\sigma$, trong đó z là đặc trưng sau khi chuẩn hóa, x là đặc trưng ban đầu, μ là giá trị trung bình, σ là độ lệch chuẩn của phân bố đặc trưng ban đầu. Sau đó, bắt đầu quá trình huấn luyện bằng phương pháp học có giám sát với dữ liệu và nhãn đã được gán, sau đó tính lan truyền ngược gradient từ lớp phân loại softmax đến lớp tích chập đầu tiên. Tham số của mạng được tối ưu hóa bởi giảm thiểu entropy chéo (cross-entropy) của hàm mất mát bằng sử dụng giảm gradient trên tập nhỏ (mini-batch) với cập nhật RMSProp. Để hiệu quả cho việc tính toán huấn luyện, dữ liệu được phân đoạn vào tập nhỏ với kích thước 128, các tham số của mạng được cập nhật sau mỗi bước lặp. Quá trình huấn luyện với tốc độ học (learning rate) là $10e-3$, yếu tố phân rã (decay-factor) là $p=0.9$. Trọng số của mạng được khởi tạo ngẫu nhiên. Ngoài ra, để tránh hiện tượng học quá vừa (overfitting), giá trị bỏ các đơn vị trong 2 lớp (dropout) cận cuối với xác suất của lớp đó là 0.5.

Bảng 1. Tham số cho mạng Baseline CNN với đầu vào 2 giây, để phân biệt cử chỉ Chọn và Unknown

Lớp	Kích thước tham số	Tổng kích thước	Kích thước đầu ra
1-Đầu vào	-	-	50x6x1
2	K: 5x1x64 b: 64	384	46x6x64
3	K: 5x1x64x64 b:64	20,544	42x6x64
4	K: 5x1x64x64 b:64	20,544	38x6x64
5	K: 5x1x64x64 b:64	20,544	34x6x64
6	W: 34x6x64x128 b: 128	1,671,296	128
7	W: 128x128 b:128	16,512	128
8	W: 128x2 b: 2	258	2
Tổng		1,750,082	

Bảng 2. Tham số cho mạng Baseline CNN với đầu vào 3 giây, để phân biệt 22 cử chỉ

Lớp	Kích thước tham số	Tổng kích thước	Kích thước đầu ra
1-Đầu vào	-	-	74x6x1
2	K: 5x1x64 b: 64	384	70x6x64
3	K: 5x1x64x64 b:64	20,544	66x6x64
4	K: 5x1x64x64 b:64	20,544	62x6x64
5	K: 5x1x64x64 b:64	20,544	58x6x64
6	W: 58x6x64x128 b: 128	2,850,944	128
7	W: 128x128 b:128	16,512	128
8	W: 128x22 b: 22	2838	22
Tổng		2,932,310	

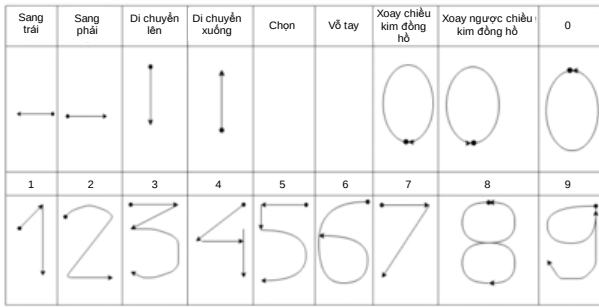
IV. THỰC NGHIỆM VÀ ĐÁNH GIÁ

A. Dữ liệu

Do phương pháp áp dụng theo hướng học sâu, dữ liệu là một trong những yếu tố quan trọng nhất. Bên cạnh phải có nhiều dữ liệu để huấn luyện học sâu được hiệu quả, các dữ liệu này cần phải có sự tổng quát để tránh vấn đề quá vừa dữ liệu (overfitting). Cho nên, nghiên cứu xây dựng tập cơ sở dữ liệu từ 20 tình nguyện viên bao gồm cả nam và nữ. Với mỗi tình nguyện viên, tổng số lần thực hiện cử chỉ là 900 lần cho 18 cử chỉ, trong đó mỗi loại cử chỉ được thực hiện 50 lần. Và kết quả là tập cơ sở dữ liệu sẽ bao gồm 18000 lần thực hiện cử chỉ từ các tình nguyện viên.

Hình 2 mô tả tập 18 các cử chỉ điều khiển, bao gồm di chuyển cánh tay theo 4 hướng: trái, phải, trên, dưới; rút tay lại rồi đưa tay về trước mặt; vỗ tay, vẽ vòng tròn theo chiều cùng, ngược kim đồng hồ, vẽ các chữ số từ 0 đến 9. Có thể nhận thấy, các cử chỉ này đều thông dụng, dễ hình dung, dễ ghi nhớ với mọi người, có tính ứng dụng cao và ý nghĩa khi điều khiển các thiết bị. Ví dụ, trên dưới tương ứng với việc tăng giảm nhiệt độ điều hòa, sang trái sang phải để chuyển đổi slide khi thuyết trình bằng máy chiếu hay các chữ số

tương ứng với điều khiển thay đổi kênh truyền hình trên vô tuyến ...



Hình 2. Tập cử chỉ sử dụng

B. Cài đặt thực nghiệm

Thực nghiệm được cài đặt trên môi trường như sau:

- Đồng hồ thông minh Sony Smartwatch 3: Bộ vi xử lý ARM A7 lõi tứ, 1.2 Ghz, 512 MB RAM, 4 GB, hỗ trợ kết nối Wifi và Bluetooth
- Điện thoại thông minh HTC One: bộ vi xử lý Qualcomm Snapdragon 801, 2GB RAM, Bộ nhớ trong 8GB
- Huấn luyện dữ liệu trên nền tảng Colab của Google

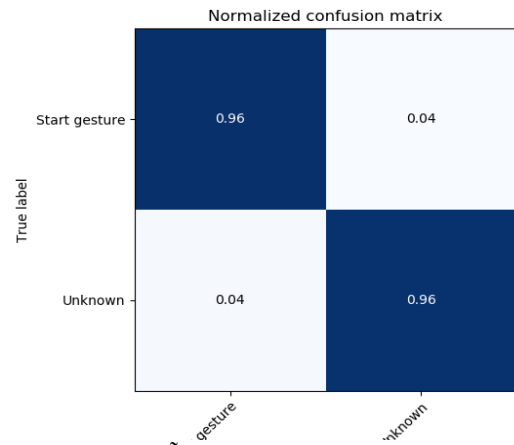
Dữ liệu được chia ngẫu nhiên dữ liệu thành 3 tập: 1 tập huấn luyện (training), 1 tập đánh giá mức độ chính xác trong quá trình huấn luyện (validation), 1 tập độc lập để kiểm thử (testing). Tập huấn luyện bao gồm 12 người, tập đánh giá mức độ chính xác trong quá trình huấn luyện gồm 3 người, 5 người còn lại dùng để kiểm thử độc lập.

Do quá trình huấn luyện sử dụng dịch cửa sổ cố định thời gian sinh ra các mẫu có số lượng không bằng nhau (chênh lệch lớn ở các nhãn như bắt đầu dơ tay xuống và đang dơ tay xuống, vv..), nên bên cạnh đánh giá theo độ chính xác, nghiên cứu áp dụng đánh giá theo mô hình F1-score bao gồm các bước: đánh giá độ chính xác (precision), độ bao phủ (recall), độ đo F1. Các giá trị được tính theo công thức:

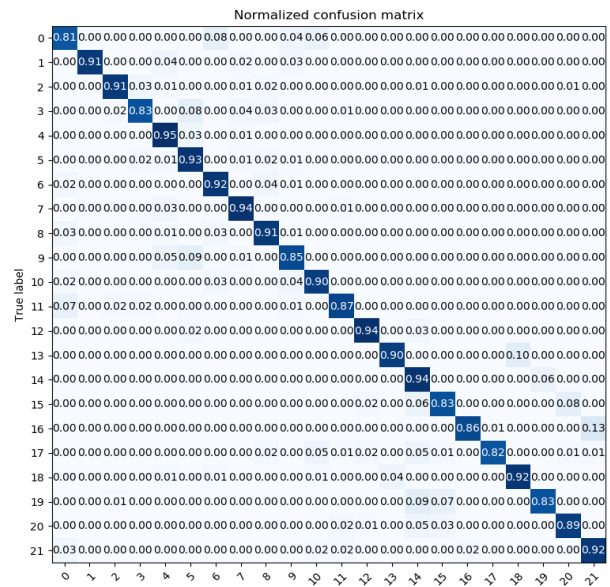
- $Precision = \frac{\text{Số lần đoán đúng nhãn}}{\text{Số lần đoán nhãn}}$
- $Recall = \frac{\text{Số lần đoán đúng nhãn}}{\text{Số nhãn thực tế}}$
- $F1 = \frac{2 * recall * precision}{recall + precision}$

C. Kết quả và thảo luận

Với 2 cử chỉ "Chọn" (Bắt đầu ra lệnh điều khiển) và Unknown, hệ thống áp dụng cửa sổ 2 giây, các cử chỉ còn lại có kích thước 3 giây, trong khi kích thước đầu vào của mạng học sâu trên cố định. Để giải quyết vấn đề này, thay vì sử dụng padding, khi triển khai hệ thống sử dụng 2 mạng tương ứng với 2 kích thước trên. Dưới đây là kết quả của 2 mạng đề cập, mỗi mạng đánh giá dựa vào loại cử chỉ tương ứng với kích thước cửa sổ 2 giây và 3 giây.



Hình 3. Ma trận hỗn loạn cho 2 cử chỉ Chọn và Unknown



Hình 4. Ma trận hỗn loạn cho 22 cử chỉ còn lại

Từ kết quả trên, ta có thể nhận xét mô hình BaselineCNN có khả năng phân loại tốt trên tập cử chỉ thu thập được. Đặc biệt với 2 cử chỉ đánh dấu bắt đầu hành động và cử chỉ Unknown, mạng cho kết quả rất tốt (đều 95.9% với F1-Score và độ chính xác). Với 22 cử chỉ còn lại, do số lượng mẫu ít hơn và số lượng lớp nhiều hơn, nên các chỉ số này có giảm nhưng vẫn trong ngưỡng chấp nhận được (88.91% với F1-score và 88.76% với độ chính xác).

Bảng 3. Kết quả thực nghiệm với mạng Baseline CNN

Loại	Precision	Recall	F1-score	Acc
2 giây	95.91	95.90	95.90	95.90
3giây	89.67	88.79	88.91	88.76

Việc sử dụng cửa sổ dịch chuyển cố định cho 22 cử chỉ điều khiển dẫn tới kết quả nhận dạng không chính xác ở 1 số mẫu. Do các cử chỉ như vẽ số 8 sẽ thường có thời gian lớn hơn so với vẽ số 1, vv... trong khi cửa sổ được cố định là 3s. Ngoài ra do không giới hạn góc mở vai, cổ tay, khuỷu tay, việc vẽ các số như số 0 và số 6, số 2 và số 3, dơ tay sang và bắt đầu dơ tay sang 4 hướng không có sự khác biệt

lớn dẫn tới khi nhận dạng dễ nhầm lẫn, từ bảng ma trận hỗn loạn ở trên ta có thể thấy cặp các cử chỉ này có độ chính xác thấp hơn so với đa số các cử chỉ khác.

V. KẾT LUẬN

Bài báo đã đề xuất một phương pháp nhận dạng cử chỉ tay động dựa trên phương pháp học sâu với các loại cảm biến phổ biến. Hai cảm biến chính đã được đề xuất, đó là cảm biến gia tốc và con quay hồi chuyển, 2 cảm biến thường được tích hợp trên các thiết bị đeo phổ biến. Giải pháp đề xuất bao gồm ba bước chính. Đầu tiên dữ liệu trong không gian 3 chiều, biến thiên theo thời gian của cử chỉ tay được thu thập. Dữ liệu này sau đó được tiền xử lý bằng cách lọc nhiễu và phân đoạn. Tiếp đến dữ liệu được đưa vào mạng học sâu để trích chọn đặc trưng và phân loại cử chỉ. Bài báo đã thử nghiệm với mạng học sâu đang được đánh giá cao trong nhận dạng cử chỉ tay động là mạng BaseLineCNN. Từ kết quả thực nghiệm có thể kết luận phương pháp học sâu nói chung và mạng BaselineCNN nói riêng cho kết quả tốt trong nhận dạng cử chỉ tay, với dữ liệu được thu thập từ cảm biến trên các thiết bị đeo tay phổ biến hiện nay.

TÀI LIỆU THAM KHẢO

- [1] A. Yang, S. M. Chun and J. -G. Kim, "Detection and recognition of hand gesture for wearable applications in IoMT," 2018 20th International Conference on Advanced Communication Technology (ICACT), 2018, pp. 1046-1053, doi: 10.23919/ICACT.2018.8323932.
- [2] M. Monisha and P. S. Mohan, "A novel IOT based approach to establish an ultra-low power self security system," 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), 2017, pp. 1-6, doi: 10.1109/ICIIECS.2017.8275874.
- [3] Y. Lei, W. Hongpeng, T. Dianxiong and W. Jue, "A real-time hand gesture recognition algorithm for an embedded system," 2014 IEEE International Conference on Mechatronics and Automation, 2014, pp. 901-905, doi: 10.1109/ICMA.2014.6885817.
- [4] Kwon, M. C., Park, G., and Choi, S. (2018). Smartwatch User Interface Implementation Using CNN-Based Gesture Pattern Recognition. *Sensors* (Basel, Switzerland), 18(9), 2997. <https://doi.org/10.3390/s18092997>.
- [5] Ho-Man Colman Leung, Chi-Wing Fu, and Pheng-Ann Heng. 2018. TwistIn: Tangible Authentication of Smart Devices via Motion Co-analysis with a Smartwatch. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 2, Article 72 (June 2018), 24 pages. <https://doi.org/10.1145/3214275>.
- [6] Lee, W., Liu, X., Shen, Y., Jin, H., and Lee, R.B. (2017). Secure Pick Up: Implicit Authentication When You Start Using the Smartphone. *Proceedings of the 22nd ACM on Symposium on Access Control Models and Technologies*.
- [7] Hong, F., You, S., Wei, M., Zhang, Y., and Guo, Z. (2016). MGRA: Motion Gesture Recognition via Accelerometer. *Sensors* (Basel, Switzerland), 16(4), 530. <https://doi.org/10.3390/s16040530>.
- [8] D. Iyer, F. Mohammad, Y. Guo, E. Al Safadi, B. J. Smiley, Z. Liang and N. K. Jain, "Generalized Hand Gesture Recognition for Wearable Devices in IoT: Application and Implementation Challenges," in *Machine Learning and Data Mining in Pattern Recognition*, Cham, 2016.
- [9] S. Kratz, M. Rohs and G. Essl, "Combining Acceleration and Gyroscope Data for Motion Gesture Recognition Using Classifiers with Dimensionality Constraints," in *Proceedings of the 2013 International Conference on Intelligent User Interfaces*, New York, NY, USA, 2013.
- [10] P. Zhu, H. Zhou, S. Cao, P. Yang and S. Xue, "Control with gestures: A hand gesture recognition system using off-the-shelf smartwatch," *Proceedings - 2018 4th International Conference on Big Data Computing and Communications, BIGCOM 2018*, pp. 72-77, 2018.
- [11] Ahmed, M. A., Zaidan, B. B., Zaidan, A. A., Salih, M. M., and Lakulu, M. (2018). A Review on Systems-Based Sensory Gloves for Sign Language Recognition State of the Art between 2007 and 2017. *Sensors* (Basel, Switzerland), 18(7), 2208. <https://doi.org/10.3390/s18072208>.
- [12] J. Liu, Z. Wang, L. Zhong, J. Wickramasuriya and V. Vasudevan, "uWave: Accelerometer-based personalized gesture recognition and its applications," in *2009 IEEE International Conference on Pervasive Computing and Communications*, 2009.
- [13] Kwon, M. C., Park, G., and Choi, S. (2018). Smartwatch User Interface Implementation Using CNN-Based Gesture Pattern Recognition. *Sensors* (Basel, Switzerland), 18(9), 2997. <https://doi.org/10.3390/s18092997>.
- [14] Yu Zhang, Tao Gu, Chu Luo, Vassilis Kostakos, and Aruna Seneviratne. 2018. FinDroidHR: Smartwatch Gesture Input with Optical HeartRate Monitor. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1, Article 56 (March 2018), 42 pages. <https://doi.org/10.1145/3191788>.
- [15] Gerald Bieber, Marian Haescher, and Matthias Vahl. 2013. Sensor requirements for activity recognition on smart watches. In *Proceedings of the 6th International Conference on Pervasive Technologies Related to Assistive Environments (PETRA '13)*. Association for Computing Machinery, New York, NY, USA, Article 67, 1-6. <https://doi.org/10.1145/2504335.2504407>.
- [16] Wile, D. J., Ranawaya, R., and Kiss, Z. H. (2014). Smart watch accelerometry for analysis and diagnosis of tremor. *Journal of neuroscience methods*, 230, 1-4. <https://doi.org/10.1016/j.jneumeth.2014.04.021>.
- [17] K. Liu, C. Chen, R. Jafari and N. Kehtarnavaz, "Multi-HMM classification for hand gesture recognition using two differing modality sensors," 2014 IEEE Dallas Circuits and Systems Conference (DCAS), 2014, pp. 1-4, doi: 10.1109/DCAS.2014.6965338.
- [18] Ahanathapillai, V., Amor, J. D., Goodwin, Z., and James, C. J. (2015). Preliminary study on activity monitoring using an android smart-watch. *Healthcare technology letters*, 2(1), 34-39. <https://doi.org/10.1049/htl.2014.0091>.
- [19] Figo, D., Diniz, P.C., Ferreira, D.R. et al. Preprocessing techniques for context recognition from accelerometer data. *Pers Ubiquit Comput* 14, 645-662 (2010). <https://doi.org/10.1007/s00779-010-0293-9>.
- [20] Zheng X, Wang M, Ordieres-Meré J. Comparison of Data Preprocessing Approaches for Applying Deep Learning to Human Activity Recognition in the Context of Industry 4.0. *Sensors*. 2018; 18(7):2146. <https://doi.org/10.3390/s18072146>.
- [21] Milošević, M., Van de Vel, A., Cuppens, K., Bonroy, B., Ceulemans, B., Lagae, L., Vanrumste, B., and Van Huffel, S. (2017). Feature selection methods for accelerometry-based seizure detection in children. *Medical and biological engineering and computing*, 55(1), 151-165. <https://doi.org/10.1007/s11517-016-1506-9>.
- [22] T. R. Bennett, J. Wu, N. Kehtarnavaz and R. Jafari, "Inertial Measurement Unit-Based Wearable Computers for Assisted Living Applications: A signal processing perspective," in *IEEE Signal Processing Magazine*, vol. 33, no. 2, pp. 28-35, March 2016, doi: 10.1109/MSP.2015.2499314.
- [23] O. D. Lara and M. A. Labrador, "A Survey on Human Activity Recognition using Wearable Sensors," in *IEEE Communications Surveys and Tutorials*, vol. 15, no. 3, pp. 1192-1209, Third Quarter 2013, doi: 10.1109/SURV.2012.110112.00192.
- [24] Shoaib M, Bosch S, Incel OD, Scholten H, Havinga PJM. Fusion of Smartphone Motion Sensors for Physical Activity Recognition. *Sensors*. 2014; 14(6):10146-10176. <https://doi.org/10.3390/s140610146>.
- [25] F. Attal, S. Mohammed, M. Dedabrishvili, F. Chamroukhi, L. Oukhellou and Y. Amirat, "Physical Human Activity Recognition Using Wearable Sensors," *Sensors*, vol. 15, no. 12, pp. 31314-31338, 2015.
- [26] S. J. Preece*, J. Y. Goulermas, L. P. J. Kenney and D. Howard, "A Comparison of Feature Extraction Methods for

- the Classification of Dynamic Activities From Accelerometer Data," IEEE Transactions on Biomedical Engineering, vol. 56, no. 3, pp. 871 - 879, 2009.
- [27] D. Rubine, "Specifying Gestures by Example," SIGGRAPH Comput. Graph., vol. 25, pp. 329-337, 7 1991.
- [28] J. O. Wobbrock, A. D. Wilson and Y. Li, "Gestures Without Libraries, Toolkits or Training: A \$1 Recognizer for User Interface Prototypes," in Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology, New York, NY, USA, 2007.
- [29] M. Müller, "Dynamic Time Warping,," In: Information Retrieval for Music and Motion. Springer, Berlin, Heidelberg, 2007.
- [30] Ricardo Chavarriaga, Hesam Sagha, Alberto Calatroni, Sundara Tejaswi Digumarti, Gerhard Tröster, José Del R. Millán, and Daniel Roggen. 2013. The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition. Pattern Recogn. Lett. 34, 15 (November, 2013), 2033–2042. <https://doi.org/10.1016/j.patrec.2012.12.014>.
- [31] Sen M Kuo and Bob H Lee, "Introduction to Real-Time Digital Signal Processing," in Real-Time Digital Signal Processing: Implementations and Applications, Second Edition, John Wiley and Sons, 2006.

APPLICATION OF DEEP LEARNING IN HAND GESTURE RECOGNITION

Abstract: The article proposes a deep learning-based method for hand gesture recognition. We used two sensors, including an accelerometer and gyroscope, to collect the input data. This data was then preprocessed to reduce noise and segment to different windows. Next, the data was fed into a deep neural network for feature extraction and gesture recognition. We have experimented with a common network architecture in hand gesture recognition, namely BaseLineCNN. From the experimental results, it can be concluded that the deep learning-based method in general and the BaselineCNN network in particular support good results in hand gesture recognition, with data collected from sensors, integrated into popular wearable devices.

Keywords: Hand gesture, Deep learning, CNN, Baseline



Nguyễn Trọng Khánh Nhận học vị Tiến sĩ năm 2013 tại Đại học Paris 6, Pháp. Hiện đang công tác tại Khoa Công nghệ Thông tin 1, Học viện Công nghệ Bưu chính Viễn thông. Lĩnh vực nghiên cứu: thị giác máy tính, IoT, Tính toán khắp nơi, học sâu, học máy, Mô hình hóa và mô phỏng các hệ thống phức tạp.

Email: khanhnt@ptit.edu.vn



Phạm Văn Cường Nhận học vị Tiến sĩ năm 2012 tại Đại học Newcastle, Vương quốc Anh. Hiện đang công tác tại Khoa Công nghệ Thông tin 1, Học viện Công nghệ Bưu chính Viễn thông. Lĩnh vực nghiên cứu: IoT, Tính toán khắp nơi, học sâu, học máy.

Email: cuongpv@ptit.edu.vn