XÁC MINH CHỮ KÝ DỰA TRÊN KỸ THUẬT HỌC SÂU

SIGNAGTURE VERIFICATION USING DEEP LEARNING

Trần Minh Nhân, Trần Đại Gia Khánh, Hồ Phước Tiến*

Trường Đại học Bách khoa - Đại học Đà Nẵng[1]

*Tác giả liên hệ: hptien@yahoo.com

(Nhận bài: 18/4/2022; Chấp nhận đăng: 10/6/2022)

Tóm tắt - Xác minh chữ ký viết tay có vai trò rất quan trọng trong việc bảo mật và xác định danh tính của người dùng khi liên quan đến các hoạt động hành chính, công ty hay ngân hàng. Sau giai đoạn đầu với những phương pháp xác minh chữ ký theo cách tiếp cận truyền thống, gần đây một số giải thuật dựa trên học sâu đã cho thấy nhiều kết quả hứa hẹn đối với bài toán này. Tuy nhiên, vẫn có ít nghiên cứu nhằm tổng hợp và so sánh các mô hình học sâu để từ đó có thể giúp cải thiện xác minh chữ ký một cách hiệu quả hơn. Bài báo này sẽ xây dựng và so sánh các mô hình học sâu gần đây – thông qua nhiều kiến trúc khác nhau – đối với bài toán xác minh chữ ký. Kết quả cho thấy, việc tách biệt quá trình học thuộc tính của ảnh chữ ký với bộ phân loại mang lại hiệu quả xác minh cao nhất. Ngoài ra, bài báo còn đề xuất sử dụng bộ phân loại mới – XgBoost – nhằm cải thiện kết quả xác minh so với phương pháp trước đây.

Từ khóa - Xác minh chữ ký; học sâu; mạng nơron tích chập; mạng Capsule; mạng Transformer

Abstract - Verification of handwritten signatures plays a very important role in securing and determining user information concerning activities in administration, companies or banks. Following early methods based on traditional approach, recent deep learning based algorithms have shown promising results for signature verification. Yet, there are few studies which have been carried out to review and compare these models, and consequently help improve signature verification effectively. This paper will build and compare several deep learning models – with various architectures – for signature verification. The results shows that separating feature learning from classification can bring the highest verification efficiency. Besides, the paper also proposes to use a new classifier – XgBoost – to improve the signature verification consequence compared with the previous method.

Key words - Signature verification; deep learning; Convolution Neural Network; Capsule Network; Transformer Network

1. Giới thiệu bài toán xác minh chữ kí

Chữ kí là một trong những dấu hiệu phổ biến nhất và thường được dùng để xác nhận danh tính của một cá nhân. Chữ kí con người có vai trò quan trọng trong các hoạt động đời sống, nhất là khi liên quan đến tính xác thực của văn bản, biểu mẫu hay giấy tờ ngân hàng. Chính vì thế, việc xác minh chữ kí – nhằm xác định xem một chữ kí nào đó có khớp với chữ kí mà ta đã biết hay không – thật sự có ý nghĩa lớn. Thông thường, việc xác minh này được thực hiện bằng mắt người, tức làm thủ công. Tuy nhiên, đây là công việc khá phức tạp và tốn nhiều thời gian. Từ đó đặt ra bài toán là làm thế nào để có thể tự động xác minh chữ kí một cách nhanh chóng và hiệu quả [1, 2, 3, 4]. Dù vậy, cho đến nay, xác minh chữ kí vẫn chưa được nghiên cứu một cách rộng rãi, khi so với xác minh, nhận dạng các đặc điểm sinh trắc học khác (như khuôn mặt, vân tay).

Chữ kí con người có những đặc điểm làm cho việc xác minh thực sự khá thách thức. Chữ kí được đặc trưng bởi nhiều yếu tố tinh tế như nét nhỏ, độ cong, hướng [1]. Chữ kí của cùng một người, nhưng tại hai thời điểm khác nhau, có thể không giống nhau. Ta có thể hình dung chữ kí của một người có thể phụ thuộc vào trạng thái tâm lý của người đó khi kí tên. Một hệ thống xác minh chữ kí hiệu quả phải có khả năng rút ra được những thuộc tính đặc trưng của chữ kí của một người nào đó, và phân biệt được với chữ kí của người khác hay chữ kí giả mạo.

Nhìn chung, có hai hệ thống xác minh chữ kí: Trực tuyến (online) và ngoại tuyến (offline) [2]. Hệ thống xác minh chữ kí trực tuyến có ưu điểm khi có thể khai thác các yếu tố thời gian của chữ kí, hay lực tác động khi kí. Trong khi đó, hệ thống xác minh chữ kí ngoại tuyến không khai thác được các thông tin này, mà chỉ dựa trên hình ảnh của chữ kí. Tuy nhiên, hệ thống ngoại tuyến lại phổ biến và thực tế hơn, ví dụ như ta chụp hay scan chữ kí để kiểm tra. Bài báo này sẽ quan tâm đến hệ thống xác minh chữ kí ngoại tuyến.

Một số nghiên cứu trước đây đã cố gắng giải quyết bài toán xác minh chữ kí ngoại tuyến, và có thể chia thành hai hướng chính như sau: Cách tiếp cận truyền thống và cách tiếp cận theo học sâu hay mạng neuron. Với cách tiếp cận truyền thống, ảnh chữ kí được trích thuộc tính thông qua các công cụ như biến đổi Wavelet, Fourier, histogram [3, 4]. Mục tiêu của bước trích thuộc tính là rút ra được những đặc điểm đặc thù của chữ kí như độ cong, góc, hướng. Sau đó, ta sẽ dùng một khoảng cách, ví dụ khoảng cách Euclidean, để so sánh hai vector thuộc tính, một từ chữ kí thật và một từ chữ kí cần xác minh. Nếu khoảng cách này đủ nhỏ thì ta xem hai chữ kí này là của cùng một người, ngược lại thì ta xem đó là chữ kí giả mạo. Ngoài ra, ta còn có những công cụ khác để xác định sự tương tự giữa hai vector, và có thể áp dụng cho xác minh chữ kí, như cosine similarity hay DTW (Dynamic Time Warping) [5]. Bên cạnh đó, bước tiền xử lý cũng thường được thêm vào trước khi trích thuộc tính để việc xác minh đạt hiệu quả cao hơn [2].

Gần đây, với sự phát triển của kĩ thuật học sâu, một số nghiên cứu cũng đã áp dụng mạng neuron tích chập (CNN) cho bài toán xác minh chữ kí [1, 2, 6]. Nhìn chung, với cách tiếp cận này, một mô hình học sâu sẽ cố gắng học được một phép đo khoảng cách phù hợp với việc xác minh chữ kí [1, 2]. Tức là, với hai chữ kí giống nhau thì mô hình cho sẽ

[1] The University of Danang - Univeristy of Science and Technology (Tran Minh Nhan, Tran Dai Gia Khanh, Ho Phuoc Tien)

cho ra khoảng cách tương đối nhỏ; Ngược lại, với hai chữ kí khác nhau thì mô hình sẽ cho ra khoảng cách lớn. Bên cạnh đó, cũng như các mô hình CNN khác (ví dụ cho bài toán nhận dạng), mô hình CNN cho xác minh chữ kí cũng khai thác ưu điểm về trích thuộc tính một cách hiệu quả hơn (so với cách tiếp cận truyền thống như histogram hay biến đổi Wavelet) [1, 6]. Sau đó, vector thuộc tính này sẽ được đưa vào một bộ phân loại hay so khớp cổ điển để xác minh chữ kí thật hay giả mạo.

Bài báo này sẽ xây dựng và so sánh một số phương pháp học sâu đối với bài toán xác minh chữ kí. Trong đó, một số được dựa trên những kĩ thuật mới được đề xuất gần đây và cho kết quả tích cực trong lĩnh vực thị giác máy tính [7, 8]. Mục đích của việc so sánh này nhằm đưa ra một bức tranh tương đối tổng thể về các kĩ thuật học sâu dùng cho xác minh chữ kí, và từ đó chỉ ra những yếu tố cần thiết để xây dựng một mô hình xác minh chữ kí hiệu quả.

Bên cạnh đó, dựa trên kết quả so sánh từ các mô hình học sâu khác nhau, bài báo cũng sẽ đề xuất cách cải thiện phương pháp xác minh chữ kí. Một cách tiếp cận hiệu quả để phân biệt chữ kí là tách riêng phần trích thuộc tính – thông qua việc chiếu ảnh chữ kí vào một không gian có số chiều tương đối lớn, mà ở đó các chữ kí khác nhau có thể được phân biệt một cách dễ dàng – và phần phân loại. Bằng cách giữ lại khối trích thuộc tính đã được huấn luyện hiệu quả, bài báo đề xuất một cách phân loại mới (XgBoost) và cho kết quả tốt hơn phương pháp thường dùng trước đây.

2. Xây dựng mô hình xác minh chữ ký

Phần này sẽ trình bày cụ thể năm mô hình học sâu dùng để xác minh chữ kí, mà sẽ được thực hiện và so sánh trong phần thực nghiệm sau này. Những mô hình này được tổng hợp từ những phương pháp nổi bật gần đây; Một số xuất phát từ bài toán xác minh chữ kí, nhưng cũng có mô hình đến từ bài toán khác và đang cho kết quả ấn tượng hiện nay. Việc bổ sung những phương pháp mới này nhằm đánh giá khả năng của chúng khi áp dụng vào bài toán xác minh chữ kí. Ngoài ra, những đề xuất khác của bài báo này cũng sẽ được chỉ rõ.

2.1. Mô hình 1: Mạng song song

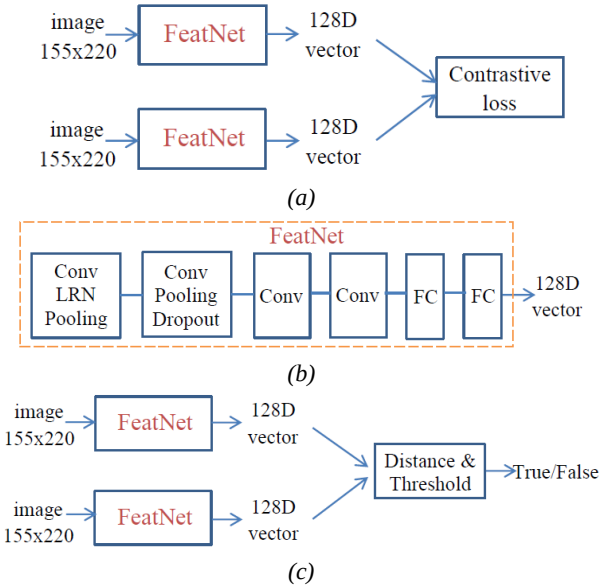
Mô hình mạng song song chứa hai mạng con giống hệt nhau, xuất phát từ mô hình SigNet [2]. Hai mạng con này có kiến trúc giống nhau và có trọng số giống nhau (Hình 1). Trong quá trình huấn luyện, việc cập nhập trọng số được sao chép cùng lúc cho cả hai mạng con. Mỗi mạng con bao gồm các lớp tích chập (với kernel có kích thước khác nhau), max pooling, và FC (Fully Connected). Ngoài ra, mạng còn sử dụng lớp Local Response Normalization (LRN) và Dropout để tăng tính tổng quát hóa. Hàm kích hoạt Rectified Linear Units (ReLU) được sử dụng trong toàn bộ mô hình. Đầu ra của mỗi mạng con là một vector 128 chiều.

Hai mạng con này được kết nối với một hàm tổn hao (contrastive loss), dựa trên hàm tính khoảng cách Euclidean giữa hai vector đầu vào (Hình 1a). Trong quá trình huấn luyện, mô hình sẽ tìm cách tối thiểu hóa khoảng cách giữa hai vector ứng với cặp chữ kí "thật-thật" và tối đa hóa khoảng cách giữa hai vector ứng với cặp chữ kí "thật-giả". Hàm tổn hao được cho như sau [2]:

$$L(s_1, s_2, y) = \alpha(1-y)D_w^2 + \beta y \max(0, m - D_w)^2 \quad (1)$$

Trong đó, s_1 và s_2 là hai ảnh chữ kí ở đầu vào. y là nhãn của cặp ảnh chữ kí đầu vào, $y = 0$ với hai ảnh "thật-thật" và $y = 1$ với hai ảnh "thật-giả". D_w là khoảng cách Euclidean giữa hai vector đầu ra của hai mạng con. α và β là hai hệ số điều chỉnh, m là ngưỡng (margin) để đảm bảo khoảng cách giữa hai ảnh "thật-giả" phải đủ lớn.

Trong quá trình kiểm tra (testing), ta sẽ dùng một ngưỡng (được chọn thông qua tập validation) để xác định chữ kí thật hay chữ kí giả, tùy theo khoảng cách giữa chúng. Hình 1c mô tả quá trình kiểm tra: Quá trình này cũng tương tự như quá trình huấn luyện (Hình 1a), điểm khác biệt duy nhất là hàm tổn hao (khi huấn luyện) được thay thế bởi hàm tính khoảng cách và phép lấy ngưỡng.

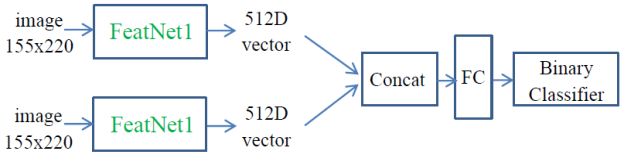


Hình 1. *(a) Kiến trúc mạng song song khi huấn luyện, (b) Chi tiết về khối FeatNet, (c) Kiến trúc mạng song song khi kiểm tra ("True" ứng với ảnh thật, "False" ứng với ảnh giả)*

2.2. Mô hình2: Mạng song song - phân loại nhị phân

Mô hình này cũng tương tự như mô hình mạng song song trong Mục 2.1. Tuy nhiên, để tránh phải chọn ngưỡng, ta sẽ thêm vào một bộ phân loại để mạng tự động phát hiện hai chữ kí là "thật-thật" hay "thật-giả". Cụ thể, hai vector đầu ra ở hai mạng con sẽ được ghép với nhau, và tiếp tục đi qua một số lớp FC, trước khi được phân loại nhị phân. Do đó, hàm tổn hao được sử dụng ở đây là Binary Cross Entropy (Hình 2).

Quá trình kiểm tra cũng được thực hiện tương tự như quá trình huấn luyện. Đầu ra của bộ phân loại nhị phân sẽ cho biết hai chữ kí đầu vào là giống hay khác nhau (thật/giả, ứng với True/False ở Hình 1c).



Hình 2. *Kiến trúc mạng song song-phân loại nhị phân*

2.3. Mô hình3: CNN-Capsule

Mô hình này cũng có kiến trúc tổng thể giống mô hình mạng song song trong Mục 2.1, tức có hai mạng con chia sẻ trọng số chung, và một hàm tổn hao tính khoảng cách. Tuy nhiên, khác biệt ở đây liên quan đến cấu trúc bên trong

của mạng con ("FeatNet" trong Hình 1). Thay vì chỉ sử dụng các lớp của mạng CNN truyền thống, ta sẽ sử dụng thêm cấu trúc Capsule [9]. Capsule có khả năng biểu diễn mối quan hệ cấu trúc tốt hơn mạng CNN truyền thống, do đó có thể nhận dạng đối tượng một cách ổn định hơn, nhất là khi có sự biến thiên ở đầu vào (ví dụ, góc nhìn thay đổi).

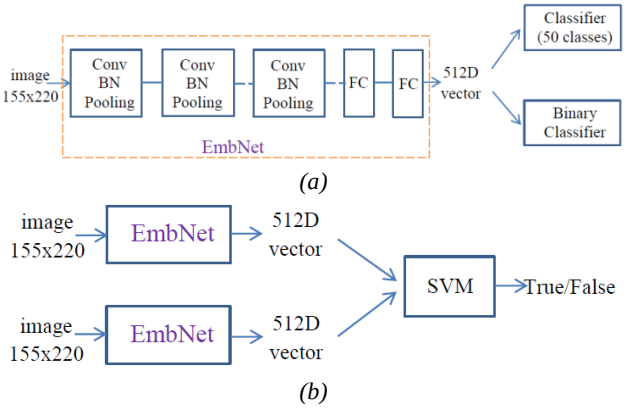
Mô hình CNN-Capsule trong mục này được mong đợi sẽ kết hợp các ưu điểm của CNN và Capsule. Các vector thuộc tính được trích xuất từ các lớp của CNN (3 lớp tích chập) sẽ được đưa vào mạng Capsule (cơ bản gồm các lớp tích chập và định tuyến - routing). Điều này giúp làm giảm kích thước mạng và tăng tốc độ tính toán. Do giới hạn về số trang, bài báo này không đi sâu vào chi tiết của mạng Capsule; Phần này có thể tìm thấy dễ dàng trong các tài liệu tham khảo liên quan.

Hàm tổn hao và quá trình kiểm tra của mô hình CNN-Capsule cũng tương tự như mô hình mạng song song (Mục 2.1).

2.4. Mô hình 4: Mạng Embedding

Mục tiêu của mô hình này là cố gắng biểu diễn một cách tốt nhất một ảnh chữ kí cho trước [1]. Do đó, đầu vào của mô hình này chỉ cần một ảnh chữ kí (Hình 3).

Để huấn luyện mô hình Embedding cho việc xác minh chữ kí, bên cạnh các lớp CNN truyền thống, ta sẽ kết hợp hai hàm tổn hao đồng thời. Hàm thứ nhất dùng để phân loại người kí tên và hàm thứ hai để phân loại chữ kí đưa vào là thật hay giả [10, 1]. Như vậy, dữ liệu trong quá trình huấn luyện là các bộ (X, y, f). Trong đó, X là ảnh chữ kí, y là định danh tác giả của chữ kí (target user), và f là một biến nhị phân cho biết X là chữ kí thật hay giả mạo. Trong mô hình này, hàm Categorical Cross Entropy được dùng để phân loại người kí tên và Binary Cross Entropy để phân loại chữ kí thật – giả.



Hình 3. *(a)Mô hình mạng Embedding khi huấn luyện, (b) dùng thuộc tính được trích từ mạng Embedding để xác minh chữ kí*

Trong quá trình kiểm tra, để phân biệt chữ kí thật-giả, ta sẽ lần lượt tính embedding vector của hai ảnh cho trước (một tham chiếu và một ảnh cần xác minh), rồi áp dụng một bộ phân loại nhị phân. Trong bài báo này, ta sẽ sử dụng lại bộ phân loại SVM như ở [1], đồng thời đề xuất sử dụng bộ phân loại mới là XgBoost [11]. XgBoost khai thác một tập hợp cây quyết định (*decision tree*) và *boosting*, và đã cho kết quả ấn tượng với các bài toán phân loại, hồi quy, và xếp hạng. Nhìn chung, XgBoost có độ phức tạp cao hơn các phương pháp phân loại truyền thống khác (ví dụ SVM), nhưng thường cho kết quả tốt hơn. Lý thuyết về cây quyết định và *boosting* có thể được tìm thấy trong nhiều tài liệu liên quan. Trong phần thực nghiệm, ta sẽ xem xét cụ thể hiệu quả của bộ phân loại XgBoost, và so sánh với SVM, đối với bài toán xác minh chữ ký.

2.5. Mô hình5: Transformer

Tương tự như mô hình mạng Embedding ở Mục 2.4, ta cũng sẽ xây dựng một mô hình để biểu diễn ảnh chữ kí. Tuy nhiên, khác với mô hình Embedding vốn sử dụng các lớp CNN truyền thống, mô hình Transformer sẽ sử dụng cấu trúc Transformer [8] để trích thuộc tính từ ảnh chữ kí ở đầu vào.

Thời gian vừa qua, mạng Transformer đã gây tiếng vang lớn với bài toán xử lý ngôn ngữ tự nhiên và đã trở thành công cụ ưu tiên trong lĩnh vực này [8]. Đặc điểm của Transformer là cho phép tập trung vào những phần quan trọng của đầu vào, để từ đó quá trình học thuộc tính có thể sẽ hiệu quả hơn. Từ thành công vượt bậc đó, Transformer bắt đầu được áp dụng vào các bài toán thị giác máy tính [12]. Bài báo này tiếp tục xem xét những ưu điểm của Transformer trong việc xác minh chữ kí [13].

Khi áp dụng cho ảnh, Transformer thường được xử lý như sau. Ảnh đầu vào được chia thành ảnh con (patch), rồi đưa qua Patch Encoder để mã hóa các ảnh con thành *word embedding*. Đồng thời, vị trí của ảnh con cũng được mã hóa thông qua *position embedding*. Hai embedding này được kết hợp lại, và đi qua 8 transformer layers, gồm các lớp con Normalization, Muti-head Attention, FC. Đầu ra của transformer layer tiếp tục qua một số lớp FC để tạo thành embedding vector kích thước 512, biểu diễn ảnh chữ kí đầu vào.

Tương tự như mô hình Embedding, trong quá trình huấn luyện, ta sẽ sử dụng hai hàm tổn hao: Phân loại người kí tên và phân loại thật-giả.

Trong quá trình kiểm tra, embedding vector của hai ảnh chữ kí (ảnh tham chiếu và ảnh cần xác minh) được đưa qua bộ phân loại SVM để xác định chữ kí thật hay giả (xem Hình 3b).

Để dễ dàng theo dõi năm mô hình học sâu đã đề cập, Bảng 1 tóm tắt các đặc điểm chính của những mô hình này. Có thể hình dung các mô hình 1, 2, và 3 thuộc nhóm mạng song song (gồm hai nhánh giống nhau, so sánh hai ảnh đầu vào); Còn các mô hình 4 và 5 tập trung vào việc học thuộc tính của ảnh và chỉ có một nhánh.

Bảng 1. *Tóm tắt các mô hình học sâu được thực hiện*

Mô hình	Kiến trúc mạng khi huấn luyện	Hàm tổn hao (huấn luyện)	Kiểm tra (testing)
Mô hình 1: Mạng song song	2 nhánh song song chia sẻ trọng số, 2 ảnh đầu vào	Contrastive Loss	Tính khoảng cách & lấy ngưỡng
Mô hình 2: Mạng song song-phân loại nhị phân	2 nhánh song song chia sẻ trọng số, 2 ảnh đầu vào	Binary Cross Entropy	Phân loại nhị phân
Mô hình 3: CNN-Capsule	2 nhánh song song chia sẻ trọng số, 2 ảnh đầu vào	Contrastive Loss	Tính khoảng cách & lấy ngưỡng
Mô hình 4: Mạng Embedding (SVM)	1 nhánh, 1 ảnh đầu vào	Categorical Cross Entropy & Binary Cross Entropy	SVM
Mô hình 5: Mạng Transformer	1 nhánh, 1 ảnh đầu vào	Categorical Cross Entropy & Binary Cross Entropy	SVM

3. Thực nghiệm và kết quả

3.1. Dữ liệu

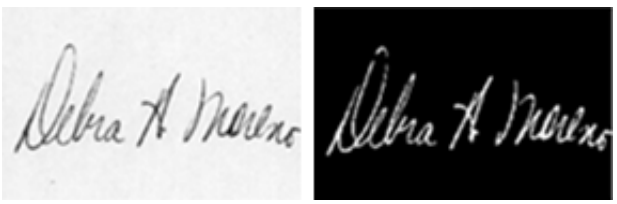
Tập dữ liệu chữ kí CEDAR (có thể được tải từ địa chỉ http://www.cedar.buffalo.edu/NIJ/data/signatures.rar) bao gồm chữ kí ảnh xám của 55 người dùng thuộc nhiều quốc gia và nghề nghiệp khác nhau [4]. Mỗi người kí 24 chữ kí của mình. Đồng thời, mỗi người cũng sẽ được giao nhiệm vụ thực hiện giả mạo chữ kí của 3 người trong tập dữ liệu, 8 bản giả mạo cho mỗi chữ kí; Tổng cộng có 24 chữ kí giả mạo. Do đó, bộ dữ liệu sẽ chứa 1320 chữ kí thật và 1320 chữ kí giả mạo. Bộ dữ liệu này được chia thành ba tập con: training-validation-testing lần lượt ứng với 45-5-5 người. Tuy nhiên, cách sử dụng tập training và validation của năm mô hình có chút khác biệt như sau.

Đối với các mô hình mạng song song (mô hình 1, 2 và 3), ta sẽ dùng hai tập training (45 người) và validation (5 người) một cách tách biệt như ở trên. Với mô hình trích thuộc tính (mô hình 4 và 5), để phù hợp với bộ phân loại người kí tên, ta gộp cả hai tập con training và validation ban đầu để tạo thành tập mới gồm 50 người, rồi sau đó chia lại theo tỉ lệ 8:2 giữa training và validation.

Trong quá trình kiểm tra, ta luôn sử dụng tập testing (5 người) để đảm bảo năm mô hình được đánh giá một cách khách quan.

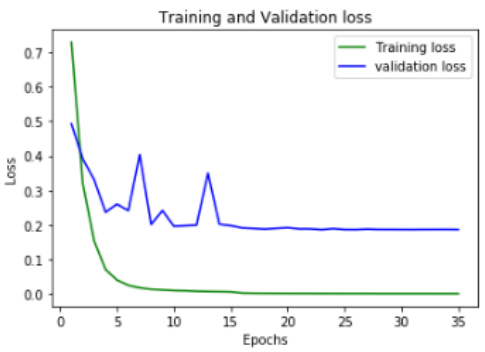
3.2. Tiền xử lý dữ liệu

Đầu tiên, ảnh của bộ dữ liệu gốc được đưa về kích thước cố định, phù hợp với đầu vào của mô hình, bằng cách sử dụng nội suy song tuyến tính. Sau đó, ảnh được khử nhiễu bằng bộ lọc thông thấp Gaussian và nhị phân hóa bằng phương pháp lấy ngưỡng Otsu [14]. Chữ kí được dịch về trung tâm của ảnh, và đảo giá trị pixel để nền có giá trị bằng 0 và chữ kí có giá trị 1 (Hình 4).



Hình 4. *Minh họa tiền xử lý ảnh: Bên trái là ảnh gốc, bên phải là ảnh sau tiền xử lý, từ bộ dữ liệu CEDAR [4]*

3.3. Huấn luyện



Hình 5. *Minh họa hàm tổn hao trong quá trình huấn luyện mạng song song*

Các mô hình được huấn luyện dựa trên phương pháp Gradient Descent để cập nhật trọng số cho đến khi hội tụ. Tốc độ học (learning rate) thay đổi theo thời gian (hay số epoch). Ví dụ, tốc độ học giảm theo tỉ lệ 0.1 tùy theo kết quả của hàm tổn hao dựa trên tập validation (validation loss). Các mô hình được huấn luyện cho đến khi hội tụ. Hình 5 minh họa các hàm tổn hao trong quá trình huấn luyện của mạng song song (mô hình 1).

3.4. Kết quả

Mục này sẽ trình bày kết quả xác minh chữ kí trên tập testing của bộ dữ liệu CEDAR. Lưu ý rằng, với mỗi lần xác minh ta có hai ảnh chữ kí: Ảnh tham chiếu (mà ta đã biết người kí) và ảnh cần xác minh; và cho kết quả là "thật"/"giả".

Chất lượng của mô hình được đánh giá thông qua tỉ lệ xác minh đúng:

$$\text{Tỉ lệ xác minh đúng} = \frac{\text{Số lần xác minh đúng}}{\text{Tổng số lần xác minh}} \quad (2)$$

Một lần xác minh được gọi là đúng nếu ảnh cần xác minh là chữ kí giả và mô hình cho ra kết quả là "giả"; Hoặc ảnh cần xác minh là chữ kí thật và mô hình cho ra kết quả là "thật".

3.4.1. So sánh năm mô hình xác minh

Trong mục này, ta so sánh tỉ lệ xác minh đúng của năm mô hình đã mô tả ở Mục 2. Với mỗi mô hình, ta thay đổi các thông số để có được kết quả tốt nhất. Chú ý rằng các mô hình này có sự khác biệt liên quan đến kiến trúc, cũng như độ phức tạp tính toán. Ở đây, bài báo chỉ tập trung vào tỉ lệ xác minh chữ kí của các mô hình. Kết quả được thể hiện Bảng 2.

Bảng 2. *Tỉ lệ xác minh đúng của năm mô hình*

Mô hình	Tỉ lệ xác minh đúng
Mô hình 1: Mạng song song	83,31%
Mô hình 2: Mạng song song-phân loại nhị phân	84,16%
Mô hình 3: CNN-Capsule	73,18%
Mô hình 4: Mạng Embedding (SVM)	94,09%
Mô hình 5: Mạng Transformer	89,60%

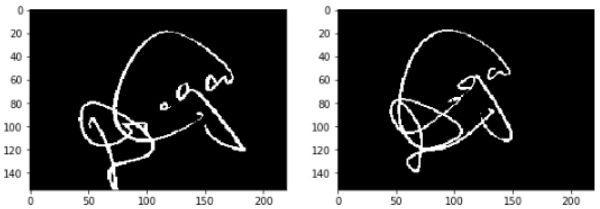
Theo kết quả ở Bảng 2 ta thấy, phương pháp trích xuất thuộc tính từ ảnh chữ kí (mô hình 4 và 5) kết hợp với một bộ phân loại riêng biệt (ở đây là SVM) cho kết quả tốt hơn so với các mô hình mạng song song (mô hình 1, 2, và 3). Trong đó, mô hình 4 có tỉ lệ xác minh đúng cao nhất, với 94,09%. Có thể bằng cách ép mô hình học chữ kí thật và giả từ các người khác nhau đã giúp cho việc biểu diễn chữ kí một cách hiệu quả hơn. Hay nói cách khác, các chữ kí sau khi đi qua mô hình kiểu này, sẽ được chiếu lên một không gian, mà ở đó các chữ kí khác nhau sẽ tách rời nhau hơn. Từ đó, việc sử dụng một bộ phân loại truyền thống (như SVM) để phân loại các chữ kí này sẽ cho kết quả xác minh tốt.

Tuy nhiên, ta cũng lưu ý rằng, mô hình trích thuộc tính (ví dụ mô hình 4) phụ thuộc vào số lượng người, liên quan đến bộ phân loại người kí tên trong quá trình huấn luyện. Khi số lượng người thay đổi thì ta phải thay đổi kiến trúc và huấn luyện lại từ đầu. Trong khi đó, các mô hình mạng song song lại không gặp vấn đề này, bởi chúng không phụ thuộc vào số lượng người. Thực tế, đầu vào của các mạng song song chỉ là cặp ảnh thật-thật hoặc thật-giả.

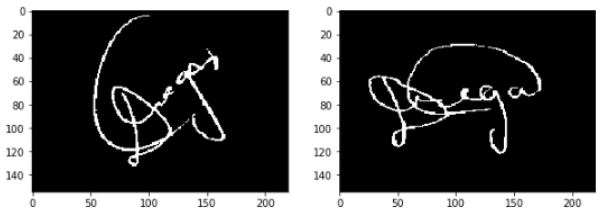
Một mục tiêu của bài báo này là đánh giá khả năng xác minh chữ kí của các mạng Capsule và Transformer, vốn đã thành công trong các bài toán khác. Kết quả thực nghiệm ở

đây cho thấy, các mô hình CNN vẫn tốt hơn Capsule hay Transformer (Transformer ở mô hình 5 cho kết quả tương đối cao, nhưng có thể do đóng góp của kiến trúc trích thuộc tính tách biệt). Có thể ảnh chữ kí không có những đặc điểm (ví dụ tính cấu trúc) phù hợp với thế mạnh của mạng Capsule hay Transformer. Tuy nhiên, ta nên xem đây là những thử nghiệm bước đầu, và có thể cần có những nghiên cứu chi tiết hơn để đánh giá thêm về khả năng của hai loại mạng này.

Bên cạnh đó, ta cũng lưu ý rằng, việc xác nhận chữ kí thông qua hình ảnh cũng có nhiều vấn đề phức tạp. Ví dụ, cùng một người kí tên của mình, nhưng tại những thời điểm khác nhau cũng có thể cho ra những chữ kí khác nhau. Điều này gây khó khăn cho việc kiểm tra bằng mắt người, và tất nhiên đối với cả máy tính. Hình 6 minh họa trường hợp chữ kí thật bị mô hình nhận nhầm thành giả. Ngược lại, có chữ kí giả, nếu xét từng đường nét riêng lẻ thì lại rất giống chữ kí thật (Hình 7). Do đó, một mô hình xác minh tốt cần có khả năng nhận ra các đường nét chi tiết cũng như hình dáng tổng thể của chữ kí. Thật sự, điều này không phải lúc nào cũng dễ dàng.



Hình 6. *Hai chữ kí thật của cùng một người nhưng mô hình cho là hai chữ kí khác nhau*



Hình 7. *Chữ kí giả (bên phải) được mô hình cho là giống chữ kí thật (bên trái)*

3.4.2. So sánh SVM và XgBoost

Kết quả thực nghiệm từ Bảng 2 cho thấy, mô hình Embedding cho kết quả xác minh tốt nhất. Ở đây, ta sẽ dựa trên mô hình này để cải thiện tỉ lệ xác minh đúng. Lưu ý rằng xác minh chữ kí theo mạng Embedding gồm hai bước: Trích thuộc tính và phân loại. Do phần trích thuộc tính phức tạp hơn, cần thời gian huấn luyện lâu hơn, và đã tạo ra vector thuộc tính tương đối hiệu quả (ứng với bộ dữ liệu cho trước), nên ta sẽ tập trung cải thiện bộ phân loại. Cụ thể, ta sẽ so sánh SVM (đã dùng ở [1]) và XgBoost (dựa trên tập hợp cây quyết định), mà gần đây đã trở thành một công cụ hiệu quả cho các bài toán phân loại hay hồi quy.

Bảng 3. *Tỉ lệ xác minh đúng của mô hình mạng Embedding*

	SVM	**XgBoost**
Mô hình 4-Mạng Embedding	94,09%	94,92%

Bảng 3 cho thấy, ưu điểm của bộ phân loại XgBoost so với SVM: Khi kết hợp với thuộc tính được trích từ mạng Embedding, XgBoost đã làm tăng tỉ lệ xác minh đúng lên 94,92%. Kết quả này cho thấy, hiệu quả của XgBoost trong việc phân loại, và việc kết hợp giữa bộ trích thuộc tính (sử dụng mạng CNN) và bộ phân loại XgBoost có thể là một giải pháp tốt cho bài toán xác minh chữ kí.

4. Kết luận

Bài báo này đã trình bày bài toán xác minh chữ kí, đây là vấn đề có ý nghĩa quan trọng đối với chữ kí điện tử hay hoạt động ngân hàng. Ta đã xem xét năm mô hình học sâu khác nhau, từ mạng CNN truyền thống cho đến các mạng mới được đề xuất gần đây như Capsule và Transformer, từ kiến trúc mạng song song đến mạng trích thuộc tính, cũng như phân tích ưu, nhược điểm của chúng. Kết quả thực nghiệm cho thấy, mạng trích thuộc tính kết hợp với một bộ phân loại riêng biệt cho tỉ lệ xác minh đúng cao nhất.

Đồng thời, bài báo còn đề xuất sử dụng XgBoost cho việc phân loại. Khi kết hợp với thuộc tính được trích từ mạng CNN, XgBoost cho phép cải thiện rõ rệt khả năng xác minh chữ kí.

Bên cạnh đó, những mạng Capsule và Transformer có thể được tiếp tục phân tích và cải thiện để đánh giá khả năng của chúng trong bài toán xác minh chữ kí. Một hướng khác là sử dụng Graph Neural Network (GNN), đây có thể là một cách tiếp cận đầy hứa hẹn với bài toán này khi đặc điểm của chữ kí khá phù hợp với dạng đồ thị (graph).

TÀI LIỆU THAM KHẢO

[1] L. G. Hafemann, R. Sabourin, and L. S. Oliveira, "Learning Features for Offline Handwritten Signature Verification using Deep Convolutional Neural Networks", *Pattern Recognition*, 70, 2017, 163-176.

[2] S. Dey, A. Dutta, J. I. Toledo, S. K. Ghosh, J. Llados, and U. Pal, "SigNet: Convolutional Siamese Network for Writer Independent Offline Signature Verification", *arXiv*:1707.02131, 2017.

[3] M. B. Yilmaz and B. Yanikoglu, "Score level fusion of classifiers in offline signature verification", *Information Fusion*, 32 (Part B), 2016, 109–119.

[4] M. K. Kalera, S. N. Srihari, and A. Xu, "Offline signature verification and identification using distance statistics", *International Journal of Pattern Recognition and Artificial Intelligence*, 18 (7), 2004, 1339–1360.

[5] G. Omer and S. Micha, "Dynamic Time Warping and Geometric Edit Distance: Breaking the Quadratic Barrier", *Association for Computing Machinery*, 14 (4), 2018, 1-17.

[6] L. G. Hafemann, L. S. Oliveira, and R. Sabourin, "Analyzing features learned for offline signature verification using Deep CNNs", *23rd International Conference on Pattern Recognition*, 2016, 2989-2994.

[7] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules", *Neural Information Processing Systems*, 2017, 3859–3869.

[8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is All you Need", *Neural Information Processing Systems*, 2017, 6000–6010.

[9] E. Parcham, M. Ilbeygi, and M. Amini, "CBCapsNet: A novel writer-independent offline signature verification model using a CNN-based architecture and capsule neural networks", *Expert Systems with Applications*, 185, 2021, 115649.

[10] O. Sener and V. Koltun, "Multi-task learning as multi-objective optimization", *Neural Information Processing Systems*, 2018, 525–536

[11] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System", *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, 785–794.

[12] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale", *arXiv*:2010.11929, 2020.

[13] X. Lu, L. Huang, and F. Yin, "Cut and Compare: End-to-end Offline Signature Verification Network", *25th International Conference on Pattern Recognition*, 2021, 3589-3596.

[14] N. Otsu, "A threshold selection method from gray-level histograms", *IEEE Transactions on Systems, Man, and Cybernetics*, 9 (1), 1979, 62–66.