

MÔ HÌNH HỌC CHUYỂN TIẾP CHO CÁC DỊCH VỤ DỰA TRÊN ĐỊNH DANH

A TRANSFER LEARNING MODEL FOR IDENTIFIER-BASED SERVICES

Nguyễn Mạnh Cường^{1,*}, Nguyễn Lương Bằng¹,
Phạm Ngọc Huân¹, Phí Trung Hiếu¹

TÓM TẮT

Các dịch vụ dựa trên định danh ngày càng phổ biến và mang lại nhiều tiện ích cho người dùng. Định danh tự động giúp mang lại các trải nghiệm dịch vụ cao cấp cho người thụ hưởng trong rất nhiều lĩnh vực như giáo dục, nghỉ dưỡng, chăm sóc sức khỏe, chăm sóc khách hàng. Nhiều mô hình và phương pháp đã được đề xuất để giải quyết bài toán định danh người dùng, trong đó kỹ thuật dựa trên hình ảnh khuôn mặt được sử dụng rộng rãi do có nhiều ưu điểm về khả năng thu thập dữ liệu, khả năng cá biệt hóa. Tuy nhiên, một hệ thống định danh với độ chính xác cao và tốc độ theo thời gian thực vẫn là mục tiêu của nhiều nghiên cứu trong thời gian gần đây. Trong bài báo này, chúng tôi giới thiệu một phương pháp học chuyển tiếp kết hợp giữa mô hình mạng nơ ron nhân tạo CNN và mô hình máy véc tơ hỗ trợ SVM cho bài toán này. Một kiến trúc CNN được đề xuất và được sử dụng như là bộ trích rút thuộc tính cho mô hình SVM làm nhiệm vụ phân lớp các đối tượng. Các kết quả thu được cho thấy sự cải thiện đáng kể về độ chính xác trong định danh cũng như thời gian huấn luyện so với các mô hình riêng lẻ.

Từ khóa: Học chuyển tiếp, mạng nơ ron, trích chọn thuộc tính, SVM.

ABSTRACT

Identity-based services are becoming more and more popular and bring many benefits to users. Particularly, automatic identification helps bring high-class service experiences to beneficiaries in many fields such as education, resort travel, health care, customer care. Many models and methods have been proposed to solve the problem of user identification, in which face image-based techniques are widely used due to many advantages in terms of data collection ability, personalization. However, an identification system with high accuracy and real-time speed is still the goal of many studies in recent times. In this paper, we introduce a transfer learning based method that combines CNN and SVM models for the face identification problem. A CNN architecture is proposed and used as a feature extractor and then, the SVM model for object classification. The obtained results show a significant improvement in the accuracy of the image classification as well as the training time.

Keywords: Transfer learning, neural network, feature extraction, SVM.

¹Khoa Công nghệ thông tin, Trường Đại học Công nghiệp Hà Nội

*Email: manhcuong.nguyen@hau.edu.vn

Ngày nhận bài: 10/3/2022

Ngày nhận bài sửa sau phản biện: 05/4/2022

Ngày chấp nhận đăng: 25/4/2022

1. GIỚI THIỆU

Dịch vụ dựa trên định danh hiện rất phổ biến trong nhiều lĩnh vực và có thể bắt gặp ở nhiều nơi như: Các khu

nghỉ dưỡng cao cấp, nhà hàng, khách sạn, trung tâm ngoại ngữ, thư viện, phòng học, phương tiện giao thông... Định danh giúp trả lời cho câu hỏi "ai đang sử dụng dịch vụ?" và ràng buộc trách nhiệm của người thụ hưởng cũng như giúp quản lý quá trình sử dụng dịch vụ. Từ rất lâu, người ta đã có nhiều phương pháp giải quyết bài toán định danh như điểm danh, bán vé, sử dụng thẻ. Tuy nhiên, những dịch vụ cao cấp thường tìm cách định danh mà không ảnh hưởng tới trải nghiệm của khách hàng. Hiện nay, nhiều hệ thống đã áp dụng thành công các công nghệ nhận dạng khuôn mặt tự động giúp khách hàng có các trải nghiệm tốt hơn khi sử dụng dịch vụ.

Trong lĩnh vực nhận dạng dựa trên sinh trắc học, định danh dựa trên khuôn mặt hiện vẫn là chủ đề nghiên cứu sôi động trong những năm gần đây. Rất nhiều nghiên cứu đã đề xuất các phương pháp khá hiệu quả và được ứng dụng rộng rãi trong nhiều lĩnh vực. Trong số đó, các nghiên cứu ứng dụng mô hình SVM (Support Vector Machine) và mô hình mạng nơ ron tích chập CNN (Convolutional Neural Network) đã thu được nhiều kết quả khả quan [1, 6, 9-11, 15, 20, 21].

Mô hình SVM được sử dụng từ khá sớm và được hỗ trợ bởi các lập luận toán học chặt chẽ. Mô hình này tỏ ra khá hiệu quả cho bài toán phân lớp dữ liệu vector nhờ độ chính xác cao, tốc độ phân lớp nhanh [7]. Tuy nhiên, để đạt hiệu quả cao hơn, các mô hình SVM thường phải kết hợp với các kỹ thuật tiền xử lý dữ liệu, trích chọn đặc trưng khá phức tạp để giảm nhiễu, giảm số chiều của dữ liệu đầu vào. Với vấn đề xử lý dữ liệu có số chiều lớn, nhiều kỹ thuật đã được đề xuất với mục tiêu là loại bỏ các thông tin dư thừa, giảm không gian lưu trữ và thời gian thực hiện, đồng thời giúp cải thiện hiệu suất của mô hình [7, 12, 13].

Một trong những ứng cử viên sáng giá để giải quyết bài toán định danh từ ảnh khuôn mặt là mạng nơ ron nhân tạo (Artificial Neural Network) [1, 9-11, 14]. Chúng có khả năng xử lý mạnh mẽ trên các ảnh nhiễu và cơ chế học dựa trên kinh nghiệm tỏ ra khá hiệu quả [15]. Các mạng nơ ron truyền thống này cũng đòi hỏi các kỹ thuật tiền xử lý dữ liệu khá phức tạp như là phân đoạn ảnh, trích chọn đặc trưng, loại bỏ thông tin dư thừa để đảm bảo một kết quả phân lớp tốt. Trong bối cảnh đó, mạng nơ ron tích chập (CNN) được coi là một biến thể của mạng nơ ron nhân tạo có khả năng khắc

phục được một số nhược điểm kể trên. Các lớp tích chập đóng vai trò như là các bộ lọc nhiễu và dò biên ảnh trong khi các lớp lấy mẫu tiếp theo tính toán các giá trị trung bình cục bộ đóng vai trò cắt giảm số chiều cho ảnh. Các thao tác này cũng giúp cho mạng CNN có khả năng xử lý các ảnh đầu vào bị biến dạng, xoay hay co giãn [15]. Các lớp kết nối đầy đủ (Fully Connected Layers) thông thường sử dụng hàm Softmax cho mục đích phân lớp. Mặc dù vậy, vẫn còn một số thách thức lớn cho bài toán nhận dạng khuôn mặt từ ảnh mà điển hình là việc xử lý một khối lượng dữ liệu có số chiều lớn và nâng cao độ chính xác nhận dạng.

Với mục đích tăng độ chính xác phân lớp, một số nghiên cứu đề xuất xem xét một giải pháp thay thế hàm Softmax trong mô hình CNN bằng SVM cho bài toán phân lớp hoặc kết hợp hai mô hình này [2, 3, 5, 16-18]. Các kết quả nghiên cứu cho thấy việc sử dụng SVM trong mạng nơ ron nhân tạo cho một kết quả khả quan hơn so với việc sử dụng hàm Softmax thông thường trên một số bộ dữ liệu. Năm 2019, A. F. Agarap [4] thực hiện việc thay thế này trên một kiến trúc mạng CNN đơn giản với hai lớp tích chập và sử dụng Max Pooling. Thay thế cho hàm Softmax là một mô hình SVM tuyến tính. Các kết quả thu được trên một số bộ dữ liệu cho thấy độ chính xác phân lớp của CNN-Softmax và CNN-SVM gần như giống nhau và cần phải tiến hành nhiều nghiên cứu hơn nữa.

Trong bài báo này, chúng tôi đề xuất một mô hình học chuyển tiếp (Transfer Learning) [8, 14, 19], là sự kết hợp của hai mô hình CNN và SVM, gọi là DSVM. Khác với các mô hình thay thế CNN-SVM được giới thiệu trong [4] khi mà CNN và SVM được huấn luyện đồng thời trong cùng một mô hình, DSVM tách rời quá trình huấn luyện của CNN và SVM. Trong đó, một kiến trúc CNN được thiết kế làm nhiệm vụ lọc nhiễu, trích rút đặc trưng và giảm số chiều trên ảnh. Véc tơ đặc trưng thu được từ mô hình này được xem như đầu vào của một mô hình SVM đa lớp. Sự kết hợp này được kỳ vọng sẽ tận dụng được ưu điểm của cả hai mô hình khi mà CNN thực hiện việc trích chọn đặc trưng rất hiệu quả trên ảnh, trong khi SVM lại có độ chính xác phân lớp tốt nếu dữ liệu đầu vào được tiền xử lý hiệu quả.

Phần tiếp theo của bài báo có cấu trúc như sau: Phần 2 sẽ giới thiệu mô hình SVM và CNN. Trong phần 3, chúng tôi giới thiệu mô hình học chuyển tiếp DSVM. Phần 4 được dành để trình bày một số kết quả thực nghiệm và cuối cùng là một số kết luận.

2. MÔ HÌNH SVM VÀ CNN

2.1. Mô hình máy véc tơ hỗ trợ SVM

Trong phần này, chúng tôi giới thiệu sơ lược về mô hình SVM. Để có một mô tả chi tiết hơn, xin xem trong [12, 13]. Xét bài toán phân lớp nhị phân. Cho trước một tập dữ liệu huấn luyện gồm n mẫu:

$$X = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \in \mathbb{R}^{n \times (d+1)}$$

trong đó, x_i là một véc tơ trong không gian \mathbb{R}^d và $y_i \in \{-1, 1\}$ là tập các nhãn lớp. Một siêu phẳng phân tách tập X thành hai miền có dạng:

$$f(x) = \langle w, x \rangle + b, \tag{1}$$

với $w \in \mathbb{R}^d$ và $b \in \mathbb{R}$. Mục tiêu của bài toán huấn luyện SVM là tìm ra một siêu phẳng phân tách "tốt nhất" tập X theo nghĩa là lề của siêu phẳng (margin) đạt cực đại. Để tìm được bộ (w, b) như vậy, ta giải bài toán tối ưu sau:

$$\min_{w, b, \xi} \left\{ \frac{1}{2} \|w\|_2^2 + C \sum_{i=1}^n \xi_i \right\} \tag{2}$$

sao cho thỏa mãn:

$$\Omega: \begin{cases} (w, b, \xi) \in \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}_+^n \\ y_i (\langle w, x_i \rangle + b) \geq 1 - \xi_i, \forall 1 \leq i \leq n \end{cases}$$

trong đó, $\langle \cdot, \cdot \rangle$ là một tích vô hướng được định nghĩa trong không gian \mathbb{R}^n , ξ_i là các biến slack được thêm vào để nới lỏng điều kiện phân lớp và C là tham số điều chỉnh. Thay vì giải bài toán (2), ta thường xem xét bài toán đối ngẫu của nó như sau:

$$\min_{\alpha} \frac{1}{2} \alpha^T H \alpha - \vec{1} \alpha \tag{3}$$

sao cho thỏa mãn: $\Delta: \begin{cases} y^T \alpha = 0 \\ 0 \leq \alpha_i \leq C, i = 1, \dots, n \end{cases}$. Trong đó,

$y = (y_1, y_2, \dots, y_n)$, $\vec{1}$ là véc tơ với toàn bộ các thành phần đều bằng 1 và H là một ma trận đối xứng được xác định bởi:

$$H_{i,j} = y_i y_j \langle \phi(x_i) \phi(x_j) \rangle = y_i y_j K(x_i, x_j) \tag{4}$$

Ở đây, $\phi(\cdot)$ là một ánh xạ từ không gian ban đầu (input space) sang không gian đặc trưng (feature space) có số chiều cao hơn nhằm xử lý trường hợp dữ liệu không phân tách tuyến tính. Hàm $K(\cdot)$ được gọi là hàm nhân (kernel function) được định nghĩa:

$$K(x, y) = \langle \phi(x) \phi(y) \rangle \tag{5}$$

2.2. Mô hình mạng nơ ron tích chập

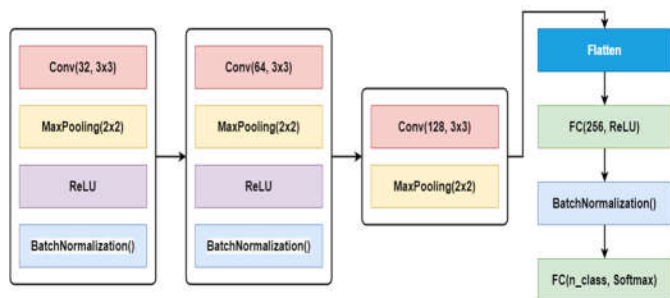
Hiệu quả của mô hình CNN phụ thuộc rất nhiều vào kiến trúc của mạng. Trong phần này, chúng tôi giới thiệu một kiến trúc mạng nơ ron tích chập được dùng cho DSVM. Mục tiêu là sử dụng một mạng không quá phức tạp nhưng vừa đủ cho các mục đích trích chọn đặc trưng và cắt giảm số chiều của ảnh. Mô hình được đề xuất (hình 1) bao gồm 3 khối chính với 3 lớp tích chập (Convolutional layer). Lớp tích chập có số lượng các bộ lọc (Filter) được đề xuất lần lượt là 32, 64 và 128. Sau mỗi lớp tích chập là một lớp gộp (Max pooling layer) với kích thước 2x2.

Tại hai khối đầu tiên, hàm kích hoạt phi tuyến ReLU có dạng $f(x) = \max(0, x)$ sẽ được thực hiện ngay sau lớp gộp để chuyển các giá trị âm thành 0. Tiếp theo, lớp Batch Normalization (BN) có vai trò chuẩn hóa các đặc trưng về trạng thái zero-mean với độ lệch chuẩn bằng 1. Lớp BN có thể giúp tránh được hiện tượng các giá trị rơi vào khoảng bão hòa sau khi đi qua các hàm kích hoạt phi tuyến, tức là đảm bảo rằng không có giá trị nào quá cao hoặc quá thấp. Điều này có thể giúp giảm thiểu hiện tượng quá khớp (overfitting).

Khối thứ 3 cũng là khối cuối cùng trước khi dữ liệu được làm phẳng để đi vào phân lớp. Khối này không bao gồm các lớp ReLU và BN. Hai lớp kết nối đầy đủ (Fully Connected layer - FC) được triển khai. Lớp FC đầu tiên có kích thước

256 *node*. Dữ liệu đầu ra của lớp này cũng được chuyển tới các lớp ReLU và BN rồi chuyển tới lớp FC cuối cùng. Lớp FC này sử dụng hàm phân loại *Softmax* với n_class là số lượng các lớp tùy thuộc vào bộ dữ liệu.

Hàm tổn thất (*loss function*) được sử dụng trong huấn luyện mô hình là *Sparse categorical cross entropy*. Thuật toán tối ưu trong quá trình này là *Adam*. Quá trình huấn luyện mô hình kéo dài 42 *epoch* và *batch size* là 16.

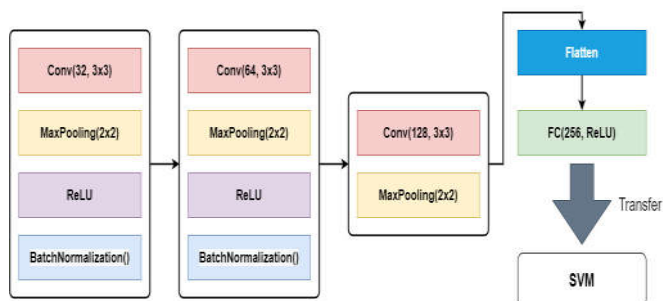


Hình 1. Mô hình mạng nơ ron tích chập được đề xuất

3. MÔ HÌNH HỌC CHUYỂN TIẾP DSVM

Mạng nơ ron tích chập được giới thiệu trong hình 1 có thể tóm lược lại thành các bước như sau:

- (1) CONV2D: 3×3, 32 filters, 1 stride, padding: same
- (2) MaxPool: 2×2
- (3) ReLU: $\max(0, x)$
- (4) BatchNormalization
- (5) CONV2D: 3×3, 64 filters, 1 stride, padding: same
- (6) MaxPool: 2×2
- (7) ReLU: $\max(0, x)$
- (8) BatchNormalization
- (9) CONV2D: 3×3 size, 128 filters, 1 stride, padding: same
- (10) MaxPool: 2×2
- (11) Flatten
- (12) FC: 256 hidden neurons, ReLU
- (13) BatchNormalization
- (14) FC: n_class , Softmax



Hình 2. Mô hình học chuyển tiếp DSVM

Để sử dụng mạng nơ ron này cho mô hình học chuyển tiếp DSVM, trước tiên mạng được huấn luyện với đầy đủ các bước. Các quá trình tiếp theo được thực hiện như sau: Mỗi ảnh đầu vào được đưa qua mạng CNN đã được huấn luyện ở trên nhưng bỏ đi các bước (13) và (14). Đầu ra của

bước (12) được lưu lại dưới dạng một véc tơ đặc trưng 256 chiều. Trong pha tiếp theo, véc tơ đặc trưng này được chuyển tiếp tới một mô hình SVM sử dụng chiến lược phân đa lớp *one vs one* và đóng vai trò như đầu vào của mô hình DSVM.

4. MỘT SỐ KẾT QUẢ THỰC NGHIỆM

Chúng tôi tiến hành thực nghiệm các mô hình trên môi trường *Google Colab* với ngôn ngữ lập trình *Python*. Trước tiên, mạng nơ ron tích chập với các bước như mô tả ở trên được huấn luyện và phân lớp trên các bộ dữ liệu. Các kết quả nhận dạng của mạng này được coi là kết quả của mô hình CNN riêng lẻ và được dùng để so sánh với mô hình học chuyển tiếp DSVM. Tương tự như vậy, một mô hình SVM riêng lẻ cũng được thực thi.

Các tham số cho mô hình SVM riêng lẻ và trong DSVM được xác định như sau. Tham số điều chỉnh C trong (2) được chọn cố định là 1. Hàm nhân được sử dụng là *Gaussian Kernel* với tham số σ được chọn lần lượt trong tập $\{10^{-9}, 10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}\}$. Với mỗi bộ dữ liệu và mỗi bộ tham số, chúng tôi sử dụng thủ tục *5-fold cross validation* để tính ra kết quả trung bình. Cuối cùng, kết quả thực nghiệm của các mô hình (bao gồm độ chính xác phân lớp - *Accuracy*, *Precision*, *Recall*, *F1-score*, thời gian huấn luyện, thời gian phân lớp) trên mỗi bộ dữ liệu là kết quả trung bình tốt nhất theo *Accuracy* của các bộ tham số trên tập dữ liệu đó.

4.1. Dữ liệu thực nghiệm

Quá trình thực nghiệm được tiến hành trên 6 bộ dữ liệu ảnh khuôn mặt người. Trong đó, bộ *TLFace* do chúng tôi tự thu thập, các bộ còn lại là các bộ dữ liệu phổ biến có thể dễ dàng thu thập trên internet. Chi tiết của các bộ dữ liệu xem trong bảng 1.

Bảng 1. Dữ liệu thực nghiệm

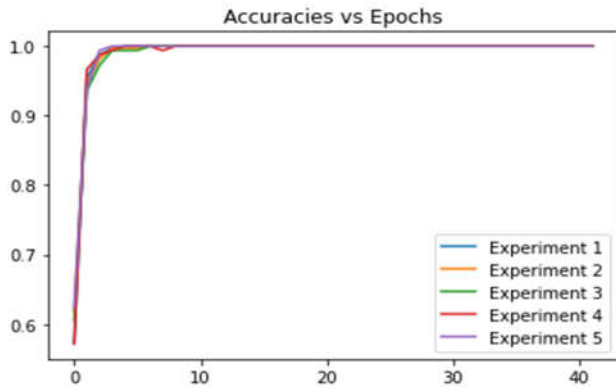
Dữ liệu	Số mẫu	Số lớp	Kích thước (pixel)	Số kênh màu
TLFace	411	20	120×80	3
AT&T	400	40	112×92	1
Georgia Tech Face	750	50	211×151	3
AR Face	2392	92	165×120	3
The extended Yale Face	2432	38	192×168	1
FEI Face	2800	200	240×320	3

4.2. Kết quả thực nghiệm

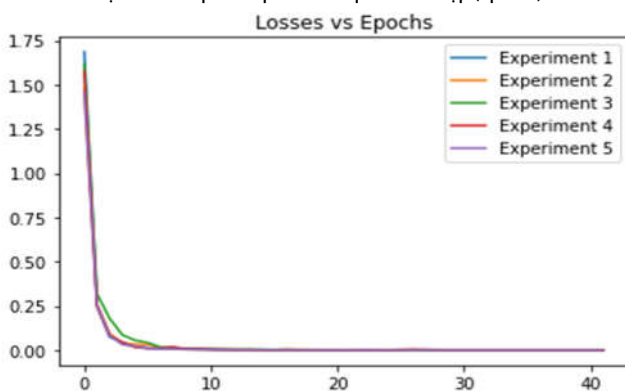
Để kiểm tra tính ổn định, tốc độ hội tụ của mô hình CNN đề xuất ở trên. Chúng tôi tiến hành thử nghiệm mô hình này trên bộ dữ liệu *TLFace*. Hình 3 và 4 biểu thị độ chính xác phân lớp và giá trị của hàm tổn thất qua các lần lặp (*Epochs*) trong quá trình huấn luyện.

Các kết quả cho thấy tốc độ hội tụ khá nhanh của mô hình này trên *TLFace* (sau khoảng 5 bước lặp). Quá trình huấn luyện cũng diễn ra ổn định, không có hiện tượng giảm đột ngột (*drop*) của *Accuracy*. Các kết quả thử nghiệm tiếp theo trên ba mô hình CNN, SVM và DSVM được trình

bày trong bảng 2. Trong đó, chỉ số về độ chính xác phân lớp trên các bộ dữ liệu huấn luyện (*Training set*) và dữ liệu kiểm tra (*Test set*) được báo cáo.



Hình 3. Độ chính xác phân lớp của CNN qua các lần lặp (Epochs)



Hình 4. Giá trị hàm tổn thất của CNN qua các lần lặp (Epochs)

Bảng 2. Độ chính xác phân lớp trên bộ dữ liệu TLFace

Dữ liệu	CNN	SVM	DSVM
Training set	100%	98,97%	99,57%
Test set	95,13%	89,29%	99,51%

Các kết quả này cho thấy mô hình CNN được đề xuất đảm bảo một độ chính xác phân lớp tốt hơn mô hình SVM trên dữ liệu TLFace và với các tập tham số được chọn. Trong trường hợp này, CNN cũng cho kết quả tốt nhất trong ba mô hình trên tập dữ liệu huấn luyện. Tuy nhiên, với dữ liệu kiểm tra (test set), DSVM có độ chính xác đạt hơn 99%, tốt hơn so với hơn 95% của CNN và hơn 89% của SVM.

Với kết quả này, chúng tôi tiến hành thực nghiệm chi tiết hơn với các bộ dữ liệu còn lại. Thay vì chỉ đánh giá trên *Accuracy*, các mô hình còn được đánh giá trên nhiều độ đo khác nhau. Các kết quả được thực hiện chỉ trên tập dữ liệu kiểm tra (*test set*).

Bảng 3. Độ chính xác phân lớp và thời gian thực thi của các mô hình

Dữ liệu	Accuracy (%)			Training time (s)			Testing time (s)		
	CNN	SVM	DSVM	CNN	SVM	DSVM	CNN	SVM	DSVM
AT&T	98,25	97,75	99,25	6,18	3,69	0,10	0,11	0,47	0,01
Georgia Tech Face	95,47	85,33	99,60	35,96	126,97	0,28	0,15	15,78	0,05

AR Face	99,96	65,34	100	70,95	820,16	3,39	0,21	102,84	0,57
The extended Yale Face	99,18	83,76	99,84	95,79	338,68	0,70	0,3	53,54	0,34
FEI Face	93,68	-	99,21	540,23	-	3,98	1,26	-	1,06

Bảng 4. Kết quả thử nghiệm các mô hình trên một số độ đo khác

Dữ liệu	Precision (%)			Recall (%)			F1-score (%)		
	CNN	SVM	DSVM	CNN	SVM	DSVM	CNN	SVM	DSVM
AT&T	98,25	97,75	99,25	98,25	97,75	99,25	98,25	97,75	99,25
Georgia Tech Face	95,47	85,33	99,60	95,47	85,33	99,60	95,47	85,33	99,60
AR Face	99,96	65,34	100	99,96	65,34	100	99,96	65,34	100
The extended Yale Face	99,18	83,76	99,84	99,18	83,76	99,84	99,18	83,76	99,84
FEI Face	93,68	-	99,21	93,68	-	99,21	93,68	-	99,21

Bảng 3 trình bày các kết quả trên các độ đo *Accuracy*, thời gian huấn luyện và thời gian phân lớp của các mô hình. Các kết quả thử nghiệm cho thấy:

- Mô hình SVM cho kết quả khá tốt trên các bộ dữ liệu thử nghiệm. Độ chính xác phân lớp trên *test sets* đạt từ 65,34 tới 97,75%, trung bình đạt 83%. Tuy nhiên, với dữ liệu đầu vào là ảnh ban đầu và không có các bước tiền xử lý hiệu quả cũng như không được hỗ trợ tính toán bởi GPU, mô hình SVM không thực thi được trên FEI Face.

- Mô hình CNN cho kết quả tốt hơn SVM về cả độ chính xác phân lớp lẫn thời gian thực thi. Kết quả thử nghiệm trên các bộ dữ liệu kể trên cho thấy: CNN có độ chính xác phân lớp lớn hơn 93,68%, trung bình đạt 97,31%. Kết quả này tốt hơn SVM trên tất cả các bộ dữ liệu thử nghiệm. Hơn nữa, CNN còn cho thấy một độ ổn định về độ chính xác phân lớp. Thời gian huấn luyện và thời gian phân lớp của CNN cũng nhanh hơn SVM trên tất cả các bộ dữ liệu mà SVM có thể thực thi. Điều này một phần là nhờ CNN được hỗ trợ tính toán bởi GPU, trong khi SVM thì không.

- DSVM cho kết quả tốt hơn CNN về độ chính xác phân lớp trên tất cả các bộ dữ liệu thử nghiệm (giao động từ 99,21 tới 100%, trung bình đạt 99,58%). Các kết quả này cũng cho thấy một sự ổn định của DSVM trên các bộ dữ liệu khác nhau. Hơn nữa, thời gian huấn luyện và thời gian phân lớp cũng nhanh hơn hoặc xấp xỉ với mô hình CNN.

Các kết quả thử nghiệm với các độ đo khác như *Precision*, *Recall*, *F1-score* (bảng 4) cũng cho kết quả tương tự: CNN cung cấp các kết quả tốt hơn SVM trên tất cả các bộ dữ liệu mà SVM có thể thực thi và DSVM cho kết quả ổn định, tốt hơn CNN trên tất cả các trường hợp thử nghiệm.

5. KẾT LUẬN

Trong bài báo này, chúng tôi giới thiệu một kỹ thuật học chuyển tiếp trên mạng nơ ron tích chập CNN và mô hình SVM, gọi là DSVM cho bài toán định danh dựa trên khuôn mặt từ ảnh. Trước tiên, chúng tôi đề xuất một mô hình CNN đơn giản cho mục đích này. Sau khi huấn luyện

mô hình CNN trên các tập dữ liệu, chúng tôi tiến hành loại bỏ lớp kết nối đầy đủ cuối cùng và đầu ra của mạng nơ ron lúc này là một véc tơ đặc trưng. Dữ liệu đặc trưng này được chuyển tiếp tới một mô hình SVM phân đa lớp theo chiến lược *one vs one* để cho ra kết quả phân lớp cuối cùng. Các kết quả thu được trên một số bộ dữ liệu thử nghiệm cho thấy sự cải thiện đáng kể về hiệu suất của mô hình DSVM bao gồm độ chính xác phân loại cũng như thời gian huấn luyện cùng một số độ đo khác so với các mô hình riêng lẻ ban đầu.

LỜI CẢM ƠN

Các nghiên cứu trong bài báo này được hỗ trợ bởi công ty Cổ phần Kinh doanh và Công nghệ MEGASOFT Việt Nam trong khuôn khổ dự án “Ứng dụng trí tuệ nhân tạo trong hệ sinh thái phần mềm Megakids”.

TÀI LIỆU THAM KHẢO

- [1]. H. Rondik, M. A. Adnan, 2021. *Deep Learning Convolutional Neural Network for Face Recognition, A Review*. International Journal of Science and Business (IJSAB), 2021.
- [2]. S. Ahlawat, C. Amit, 2020. *Hybrid CNN-SVM Classifier for Handwritten Digit Recognition*. Procedia Computer Science.
- [3]. H. Basly, et al, 2020. *CNN-SVM Learning Approach based Human Activity Recognition*. International Conference on Image and Signal Processing (ICISP).
- [4]. A. F. Agarap, 2019. *An architecture combining convolutional neural network (cnn) and support vector machine (SVM) for imageclassification*. arXiv:1712.03541v2.
- [5]. Abien Fred M. Agarap, 2018. *A Neural Network Architecture Combining Gated Recurrent Unit (GRU) and Support Vector Machine (SVM) for Intrusion Detection in Network Traffic Data*. ICMLC 2018, Macau, China.
- [6]. Shepley Andrew, 2019. *Deep Learning For Face Recognition, A Critical Analysis*. arXiv:1907.12739 [cs.CV].
- [7]. Nguyen Manh Cuong, Nguyen Van Thien, 2018. *Using the SCAD function for the problem of cutting the number of support vectors in the SVM model*. VNICT, Vietnam.
- [8]. H. Mahub, B. Jordan, F. Diego, 2018. *A Study on CNN Transfer Learning for Image Classification*. UK Workshop on Computational Intelligence (UKCI).
- [9]. A. Mohamad, et al., 2018. *Convolutional Neural Networks for Electrocardiogram Classification*. Journal of Medical and Biological Engineering.
- [10]. W. Jie, L. Zihao, 2018. *Research on Face Recognition Based on CNN*. IOP Conference Series, Earth and Environmental Science.
- [11]. M. Coşkun et al., 2017. *Face recognition based on convolutional neural network*. International Conference on Modern Electrical and Energy Systems (MEES).
- [12]. Le Thi Hoai An, Nguyen Manh Cuong, 2016. *Efficient Algorithms for Feature Selection in Multi-class Support Vector Machine*. Annals of Operations Research.
- [13]. Nguyen Manh Cuong, Nguyen Van Thien, 2016. *A Method for Reducing the Number of Support Vectors in Fuzzy Support Vector Machine*. Advanced Computational Methods for Knowledge Engineering, Volume 453 of the series Advances in Intelligent Systems and Computing, pp. 17-27, 2016.

- [14]. Y. Jason, et al., 2014. *How transferable are features in deep neural networks?*. Advances in Neural Information Processing Systems (NIPS).
- [15]. A. R. Syafeeza, et al., 2014. *Convolutional Neural Network for Face Recognition with Pose and Illumination Variation*. International Journal of Engineering and Technology (IJET), Vol 6 No 1.
- [16]. Yichuan Tang, 2013. *Deep Learning using Linear Support Vector Machines*. International Conference on Machine Learning 2013, Challenges in Representation Learning Workshop, Atlanta, Georgia, USA. arXiv preprint arXiv:1306.0239.
- [17]. Abdulrahman Alalshekmubarak, Leslie S Smith, 2013. *A novel approach combining recurrent neural network and support vector machines for time series classification*. in Innovations in Information Technology (IIT), 9th International Conference on. IEEE, pp. 42–47.
- [18]. N. Xiao-Xiao, S. Ching, 2012. *A novel hybrid CNN-SVM classifier for recognizing handwritten digits*. Pattern Recognition.
- [19]. S. J. Pan, Q. Yang, 2010. *A Survey on Transfer Learning*. IEEE Transaction on Knowledge and Data Engineering.
- [20]. F. Omar, H. Md. Al, 2009. *Face recognition using PCA and SVM*. 3rd International Conference on Anti-counterfeiting, Security, and Identification in Communication.
- [21]. Zhao Lihong, et al., 2009. *Face recognition based on multi-class SVM*. Chinese Control and Decision Conference.

AUTHORS INFORMATION

**Nguyen Manh Cuong, Nguyen Luong Bang,
Pham Ngoc Huan, Phi Trung Hieu**

Faculty of Information Technology, Hanoi University of Industry