ISSN: 1859-2171
e-ISSN: 2615-9562
TNU Journal of Science and Technology 225(06): 451 - 458

PHÂN LOẠI CHỮ SỐ CHO CÁC CAMERA NHẬN DIỆN BIỂN SỐ GIAO THÔNG TẠI VIỆT NAM

Lê Hữu Tôn*, Nguyễn Hoàng Hà
Trường Đại học Khoa học và Công nghệ Hà Nội

TÓM TẮT

Nhận dạng ký tự là một bài toán nghiên cứu quan trọng và được áp dụng trong nhiều bài toán thực tế, trong đó có bài toán nhận dạng các biển số xe cho các camera giám sát giao thông. Các bài toán nhận dạng thường xây dựng một mô hình phân loại cho tất cả các lớp. Tuy nhiên, độ khó để phân loại các lớp ký tự là không đồng đều, một số ký tự dễ bị phân loại nhầm hơn các kí tự khác. Việc xây dựng một mô hình phân loại duy nhất cho tất cả các lớp ký tự dẫn đến việc dự đoán các lớp ký tự có độ chính xác rất khác nhau. Trong bài báo này, chúng tôi trình bày một phương pháp giúp cải thiện độ chính xác trong việc nhận dạng các kí tự khó bằng cách xây dựng một bộ phân loại 2 lớp. Trong đó, bộ phân loại thứ nhất được áp dụng cho tất cả các loại ký tự, bộ phân loại thứ 2 có tác dụng phân loại lại các ký tự khó, nhằm sửa lại những lỗi phân loại của bộ phân loại thứ nhất. Thực nghiệm trên 2 tập dữ liệu SHVN và tập dữ liệu các chữ số trích xuất từ camera nhận dạng biển số tại Việt Nam cho thấy phương pháp được đề xuất giúp cải thiện độ chính xác của 1 số ký tự đến 1,4%.
Từ khóa: *Xử lý hình ảnh; nhận dạng ký tự; mạng nơron tích chập; học sâu; phân loại hình ảnh*

Ngày nhận bài: 18/5/2020; Ngày hoàn thiện: 28/5/2020; Ngày đăng: 31/5/2020

CHARACTER RECOGNITION FOR LICENSE PLATE RECOGNITION TRAFFIC CAMERA IN VIETNAM

Le Huu Ton*, Nguyen Hoang Ha
University of Science and Technology of Hanoi

ABSTRACT

Optical Character Recognition (OCR) is an active research direction with many practical applications, including digital character classification for license plate recognition on traffic cameras. The OCR models usually deploy a single classifier for all the categories in the dataset. However, the classification difficulties among all the classes in the dataset are different, some characters are easier to be misclassified compared to the others. Due to this reason, the classification performances across the classes are not equal. In this paper, we deploy a 2-stage classifier in order to improve the classification accuracy for difficult classes. The first classifier is used to classify all the classes while the second one is used only for difficult classes, in order to refine the predictions made by the first classifier. The experiment results on two datasets SVHN and license plate characters demonstrate that the proposed method helps to enhance the classification accuracy of some difficult classes by 1.4%.
Keywords: *Image processing; optical character recognition; convolutional neural network; deep learning; image classification.*

Received: 18/5/2020; Revised: 28/5/2020; Published: 31/5/2020

* Corresponding author. *Email: le-huu.ton@usth.edu.vn*

1. Giới thiệu

Phân loại ký tự là một bài toán nghiên cứu quan trọng và được áp dụng trong nhiều ứng dụng thực tế, trong đó có bài toán nhận dạng các biển số xe cho các camera giám sát giao thông. Trong bài toán phân loại, các mô hình phân loại thường nhận đầu vào là một ảnh ký tự và dự đoán xem ký tự chứa trong ảnh là ký tự nào. Trong những năm qua, đã có nhiều phương pháp được công bố để giải quyết các bài toán này. Các phương pháp này thường được chia làm hai hướng chính. Hướng thứ nhất sử dụng các đặc trưng thủ công như các đoạn ký tự [1], nét chữ [2] hoặc các điểm đặc trưng [3] để trích xuất các vector đặc trưng của từng ký tự và dùng các vector đặc trưng này để nhận dạng ký tự. Ở hướng nghiên cứu thứ hai, các công bố thường sử dụng thuật toán học sâu với mạng nơron tích chập (convolutional neural network - CNN) để nhận dạng ký tự. Một số công bố tiêu biểu có thể kể đến như mạng AlexNet [4], MobileNet [5] hay Population Based Augmentation [6].

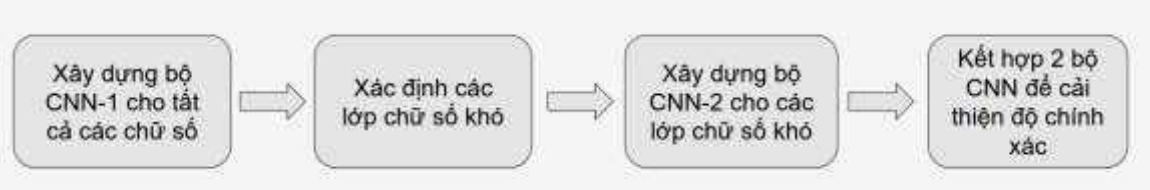
Nhận dạng chữ số là bài toán riêng trong nhận dạng ký tự, trong đó mỗi ảnh đầu vào sẽ thuộc vào một trong 10 ký tự từ 0 đến 9. Tính chất của 10 lớp chữ số này là khác nhau: ví dụ số lượng mẫu trong tập dữ liệu, mối quan hệ giữa các lớp cũng như sự chồng chéo giữa các lớp dữ liệu là khác nhau. Do vậy, một số lớp chữ số trở nên khó nhận dạng hơn các lớp chữ số khác. Vấn đề này là một trong những đặc tính về mất cân bằng dữ liệu được mô tả trong [7]. Một số phương pháp được giới thiệu để giải quyết vấn đề này như AdaBoost [8] hay ensemble-based classifiers [9]. Tuy nhiên, các phương pháp này chưa được áp dụng thử nghiệm trên các tập dữ liệu nhận dạng chữ số.

Để nâng cao độ chính xác của các lớp chữ số khó trong bài toán phân loại chữ số, chúng tôi đề xuất mô hình sử dụng kết hợp 2 bộ phân loại CNN. Bộ thứ nhất được sử dụng để phân loại tất cả các lớp chữ số. Bộ thứ hai được dùng riêng để phân loại các lớp chữ số khó. Tùy vào kết quả phân loại của bộ CNN thứ nhất, mô hình sẽ lựa chọn một số ảnh thuộc các lớp chữ số khó nhận dạng để nhận dạng lại. Thực nghiệm trên hai tập dữ liệu "The Street View House Number" [10] và bộ dữ liệu các chữ số trong biển số xe cho thấy phương pháp đề xuất giúp tăng khả năng nhận diện các ký tự khó lên đến 1,4%, là một tỷ lệ đáng kể khi trong thực tế việc cải thiện từng phần trăm khi độ chính xác vượt 90% thường rất khó khăn. Theo tìm hiểu của chúng tôi, đây là lần đầu tiên phương pháp kể trên được áp dụng cho hai tập dữ liệu này.

Phần còn lại của bài báo được trình bài như sau. Phương pháp phân loại ảnh chữ số sử dụng hai bộ CNN được trình bày trong phần 2. Phần 3 giới thiệu và phân tích các kết quả thực nghiệm. Cuối cùng, phần 4 đưa ra kết luận về phương pháp được đề xuất.

2. Phương pháp nghiên cứu

Trong phần này chúng tôi trình bày phương pháp đề xuất nhằm cải thiện độ chính xác của một số lớp ký tự khó nhận diện. Hầu hết các phương pháp nhận diện ký tự hiện nay đều sử dụng thuật toán học sâu với mô hình mạng nơron tích chập CNN. Đặc điểm của phương pháp này là các mô hình có khả năng tự học các đặc trưng tốt nhất của tập dữ liệu huấn luyện và tiến hành phân loại. Khi huấn luyện một mô hình nhằm phân loại tất cả các chữ số (0-9), mạng CNN sẽ cố học những đặc trưng để phân loại tất cả các kí tự này dựa trên một hàm tối ưu nhằm thu được tỉ lệ nhận diện cao nhất trên toàn tập dữ liệu. Tuy nhiên, tùy vào các tập dữ liệu khác nhau, độ khó trong việc phân loại các lớp ký tự là khác nhau. Việc huấn luyện mô hình để học các đặc trưng sử dụng chung cho tất cả các lớp ký tự dẫn đến độ chính xác của các lớp ký tự cũng khác nhau. Một số lớp ký tự dễ phân biệt sẽ cho độ chính xác cao, trong khi đó, các ký tự khó phân biệt sẽ có độ chính xác thấp hơn. Việc gia tăng độ phức tạp của mạng CNN (gia tăng số lớp, số bộ lọc ở mỗi lớp) thường làm tăng đáng kể tốc



Hình 1. *Các bước xử lý chính của phương pháp*

độ tính toán và nhiều khi không giải quyết triệt để vấn đề (do mô hình vẫn học các đặc trưng tốt nhất cho cả bộ dữ liệu chứ không tập trung giải quyết các lớp dữ liệu khó). Xuất phát từ nguyên nhân trên, chúng tôi đề xuất sử dụng thêm một mạng CNN được huấn luyện riêng cho các ký tự khó. Sau khi sử dụng một mạng CNN chung (sau đây gọi là CNN-1) để phân loại các ký tự, chúng tôi lọc ra những trường hợp các kí tự khó, có độ chính xác thấp và đưa vào bộ CNN thứ hai (sau đây gọi là CNN-2). Do mạng CNN-2 được huấn luyện chỉ để nhận dạng các ký tự khó, nó sẽ cố học các đặc trưng tốt nhất để nhận dạng các ký tự khó này (thay vì các đặc trưng để nhận dạng tất cả các ký tự), độ chính xác trong việc nhận diện các ký tự khó được cải thiện đáng kể. Các bước thực hiện chính của phương pháp được mô tả như trong hình 1.

- **Huấn luyện một mạng CNN chung để nhận diện các lớp ký tự**: Đối với các bài toán phân loại ký tự có nhiều mô hình CNN đã được giới thiệu với độ phức tạp và hiệu năng khác nhau. Trong khuôn khổ bài báo này, chúng tôi sử dụng 2 mạng CNN là MobileNet [5] và AlexNet [4], đây đều là các mạng CNN đơn giản hoặc có tốc độ tính toán nhanh, giúp các bộ nhận dạng ký tự có khả năng đáp ứng yêu cầu chạy trong thời gian thực. Chúng tôi hiểu rõ, đây không phải là những mạng CNN có độ chính xác cao nhất hay chạy nhanh nhất cho tập dữ liệu nhận dạng chữ số. Các mạng này được lựa chọn với mục đích kiểm tra tính hiệu quả của mô hình CNN 2 lớp so với việc sử dụng 1 mô hình CNN duy nhất.

- **Xác định các ký tự khó nhận diện**: Việc xác định các ký tự khó nhận diện phụ thuộc vào từng tập dữ liệu và ứng dụng. Trong nghiên cứu của chúng tôi, việc xác định này được dựa vào ma trận nhầm lẫn (confusion matrix). Với mỗi bài toán, dữ liệu thường được chia làm 3 phần, tập huấn luyện (training data), tập dữ liệu xác thực (validation data) và tập dữ liệu kiểm thử (testing data). Mô hình được huấn luyện dựa trên dữ liệu huấn luyện và được lựa chọn dựa trên độ chính xác trên tập dữ liệu xác thực. Mô hình có độ chính xác cao nhất trên bộ dữ liệu xác thực được chọn làm mô hình cuối cùng. Mô hình này sẽ được kiểm nghiệm trên tập dữ liệu kiểm thử để đưa ra độ chính xác cuối cùng của mô hình. Thông thường, các tập dữ liệu huấn luyện, xác thực được lựa chọn sao cho chúng có cùng tính chất, độ phức tạp, độ phân bố so với bộ dữ liệu kiểm thử. Sau khi mô hình được huấn luyện, chúng tôi tính toán ma trận nhầm lẫn của mô hình trên tập dữ liệu xác thực và từ đó tìm ra các lớp chữ số hay bị nhầm lẫn với nhau. Chúng tôi sẽ huấn luyện thêm bộ phân loại CNN-2 để kiểm tra lại kết quả nhận diện các lớp chữ số khó này.

- **Huấn luyện mạng CNN để nhận diện các ký tự khó**: Sau khi đã xác định được các lớp chữ số khó nhận diện hoặc hay bị nhầm lẫn với nhau, chúng tôi sử dụng các chữ số thuộc lớp này để xây dựng bộ phân loại CNN-2 dành cho các lớp chữ số khó. Chúng tôi vẫn sử dụng lại các tập dữ liệu huấn luyện, xác thực và kiểm thử như bước trước để xây dựng bộ phân loại CNN này. Tuy nhiên, thay vì sử dụng toàn bộ các lớp dữ liệu, chúng tôi chỉ sử dụng các lớp chữ số khó, có độ chính xác thấp để huấn luyện bộ nhận diện này. Ở bước này, chúng ta có thể sử dụng một mạng CNN có

cấu trúc giống hoặc khác so với bộ CNN-1 sử dụng ở bước một. Tuy nhiên, để kết quả thực nghiệm được đánh giá một cách khách quan và không phụ thuộc vào việc lựa chọn cấu trúc mạng CNN, chúng tôi sử dụng chung cấu trúc mạng CNN đã sử dụng ở bước một với thay đổi duy nhất là thay đổi số lượng đầu ra (output) của mạng CNN.

- Kết hợp hai bộ CNN để tăng độ chính xác: ở bước này, chúng tôi giả thiết bộ CNN-2 sẽ có độ chính xác cao hơn so với bộ CNN-1 trong việc phân loại các lớp chữ số khó. Giả thiết này là khả thi do bộ CNN-2 được huấn luyện để tối ưu hóa khả năng nhận diện cho các ảnh thuộc các lớp chữ số này, thay vì tối ưu khả năng nhận diện cho tất cả các lớp chữ số. Vấn đề đặt ra là làm thế nào để xác định những ảnh nào cần phân loại lại bằng bộ CNN-2. Với mỗi ảnh đầu vào, bộ phân loại CNN-1 sẽ trả về xác suất mà ảnh đầu vào thuộc từng lớp chữ số. Kết quả nhận diện cuối cùng sẽ được chọn là lớp chữ số có xác suất cao nhất. Mục đích của bộ CNN-2 là phân loại lại các ảnh thuộc các lớp chữ số khó phân loại. Nhằm tránh khỏi việc bộ CNN-2 phân loại lại các ảnh thuộc các lớp chữ số khác, chúng tôi lựa chọn chỉ phân loại lại những ảnh mà hai xác suất cao nhất được đề xuất với bộ CNN-1 đều thuộc vào các lớp chữ số khó và hay bị nhận dạng nhầm.

3. Kết quả và bàn luận

Tập dữ liệu: Để minh họa cho mô hình phân loại của mình, chúng tôi sử dụng 2 tập dữ liệu sau đây:

- Tập dữ liệu "The Street View House Number (SVHN)" [10]. Đây là tập dữ liệu chứa các ảnh chữ số trong thế giới thực. Tập dữ liệu này được thu thập từ các ảnh số nhà của thuộc Google Street View. Mỗi ảnh trong tập SVHN có kích thước 32 x 32. SVHN bao gồm 73.257 ảnh trong tập huấn luyện và 26.032 ảnh trong tập kiểm thử. Chúng tôi sử dụng một phần dữ liệu trong tập huấn luyện để xây dựng bộ xác thực với tỷ lệ 8-2, 58.606 ảnh được sử dụng để huấn luyện mô hình, 14.651 ảnh được sử dụng để xác thực mô hình.

- Tập dữ liệu chữ số trong biển số xe tại Việt Nam: đây là tập dữ liệu được thu thập từ các camera nhận diện biển số tại Việt Nam, bao gồm cả biển số xe ô tô và biển số xe máy. Các ảnh trong bộ dữ liệu này có kích thước 30 x 50. Để thống nhất với bộ dữ liệu SVHN, chúng tôi chỉ tiến hành nhận diện các chữ số và tạm thời bỏ qua các kí tự. Bộ dữ liệu bao gồm 21.900 ảnh, với 10 lớp chữ số (0-9). Trong thực nghiệm của mình, chúng tôi sử dụng 14.235 ảnh để huấn luyện mô hình, 3.285 ảnh để xác thực mô hình và 4.380 ảnh để kiểm thử mô hình. Các ảnh trong bộ dữ liệu được thu bởi nhiều camera, trong các điều kiện ánh sáng khác nhau, các ảnh có nền ảnh (background) khá đồng nhất do đều được cắt ra từ biển số xe. Tuy nhiên, do đặc điểm các ảnh được trích xuất từ các xe đang chuyển động, nên các ảnh đều bị ảnh hưởng ít nhiều bởi nhiễu chuyển động. Do ảnh hưởng của nhiễu chuyển động, một số chữ số dễ bị nhầm sang chữ số khác, đặc biệt là lớp chữ số 5 và 6 (ví dụ ảnh số 5 trong hàng 2, hình 2b). Đặc điểm này hiếm khi xảy ra với các bộ dữ liệu ảnh tĩnh khác.

Một số mẫu ảnh trong 2 tập dữ liệu này được minh họa như trong hình 2.

- Xây dựng bộ CNN-1 cho tất cả các lớp chữ số: Ở bước này chúng tôi dùng dữ liệu huấn luyện để xây dựng các mô hình nhận diện có thể nhận diện tất cả lớp chữ số. Mô hình nhận diện được lựa chọn là mô hình có độ chính xác nhận diện cao nhất trong tập xác thực. Độ chính xác của các bộ nhận diện này được thể hiện trong bảng 1. Chúng ta có thể thấy hầu hết các phương pháp đều cho độ chính xác trên 90%. Độ chính xác trên tập dữ liệu biển số xe có độ chính xác cao hơn so với tập SVHN do tính chất của tập dữ liệu. Các chữ số trong bộ dữ liệu biển số xe thường dễ phân biệt hơn, sử dụng cùng một phông chữ và nền cũng ít nhiễu hơn.

Bảng 1. *Độ chính xác của bộ nhận dạng chữ số của hai mạng AlexNet, MobileNet với hai tập dữ liệu SVHN và biển số xe*

Tập dữ liệu	AlexNet	MobileNet
SVHN	93,67 (%)	91,53 (%)
Biển số xe	97,26 (%)	98,52 (%)

- Xác định các lớp chữ số khó nhận diện. Để xác định các lớp chữ số khó nhận diện, chúng ta dựa vào ma trận nhầm lẫn của bộ CNN-1 trên bộ dữ liệu xác thực. Hình 3 thể hiện ma trận nhầm lẫn của các bộ nhận dạng trên 2 tập dữ liệu khác nhau.

Đối với tập dữ liệu SVHN, chúng ta có thể thấy 1 số lớp dữ liệu có độ nhận diện chính xác chưa cao như số 3 (88,1%), số 8 (88,6%). Trong đó có 3,1% số 3 bị nhầm thành số 5, 2,2% số 5 bị nhầm sang số 3, 4,4% số 8 bị nhầm thành số 6, 2,2% số 6 bị nhầm thành số 8. Như vậy, chúng ta có thể xây dựng thêm bộ CNN-2 để phân biệt các lớp chữ số 3 và 5, 6 và 8. Chúng tôi lựa chọn xây dựng bộ phân loại cho chữ số 6 và 8. Bộ phân loại các lớp chữ số 3 và 5 có thể làm tương tự. Những ảnh nào mà bộ phân loại đầu tiên CNN-1 trả về hai xác suất cao nhất là 6 và 8 sẽ được đưa vào kiểm tra lại sử dụng bộ CNN-2. Trong thực nghiệm của chúng tôi, có 846/1155 trên tổng số 26032 ảnh trong bộ kiểm thử phải chạy qua bộ CNN-2 với lần lượt các mạng AlexNet và MobileNet. Điều này dẫn đến thời gian tính toán tăng thêm 3,2% và 4,4%.

Đối với tập dữ liệu biển số, việc xác định các lớp chữ số khó phân loại là phức tạp hơn. Thay vì có 2 lớp kí tự hay bị phân loại nhầm cho nhau như tập SVHN, tập dữ liệu biển số có 3 lớp kí tự hay bị nhầm lẫn chéo với nhau. Ví dụ, có 4,2% số 5 bị nhầm sang số 6, 3,2% số 6 bị nhầm sang số 5, 2,2% số 6 bị nhầm sang số 8, 1,5% số 8 bị nhầm sang số 6. Do đó, với tập dữ liệu này chúng tôi xác định 3 lớp kí tự khó phân biệt là lớp chữ số 5, lớp chữ số 6 và lớp chữ số 8. Bộ phân loại thứ 2 sẽ được dùng để phân loại 3 lớp chữ số ở trên. Những ảnh nào mà bộ phân loại đầu tiên CNN-1 trả về hai xác suất cao nhất trùng với các tập sau sẽ được đưa vào phân loại lại bởi bộ phân loại thứ hai: (5,6), (6,8) và (5,8). Trong thực nghiệm của chúng tôi, có 183/256 trên tổng số 4380 ảnh trong bộ kiểm thử phải chạy qua bộ CNN-2 với lần lượt các mạng AlexNet và MobileNet. Điều này tương đương với việc thời gian tính toán tăng thêm 4,2% và 5,8%.

Độ chính xác của các bộ phân loại thứ 2 này được thể hiện trong bảng 2. Có thể thấy độ chính xác của bộ phân loại này thường cao hơn độ chính xác của bộ phân loại đầu tiên do chỉ phải tập trung phân loại một số ít các lớp dữ liệu.

Kết hợp hai bộ CNN để tăng độ chính xác. Do bộ CNN-2 thường có độ chính xác cao hơn bộ CNN-1, kết hợp 2 bộ nhận dạng nhìn chung giúp gia tăng độ chính xác của toàn hệ thống. Hình 4 thể hiện ma trận nhầm lẫn của 2 tập dữ liệu trước và sau khi kết hợp với bộ phân loại CNN-2. Độ chính xác trong việc phân loại các lớp chữ số khó tăng lên đáng kể. Bảng 3 thể hiện độ chính xác trong việc phân loại các lớp chữ số khó cũng như tổng hợp thời gian tính toán cần tăng thêm do việc sử dụng 2 bộ CNN. Có thể thấy, với phương pháp được đề xuất, độ chính xác cho các lớp chữ số khó tăng lên khoảng 1,4% với yêu cầu tăng thêm 4,4% thời gian tính toán.

Hình 2. *Bộ dữ liệu sử dụng trong thực nghiệm: (a) SVHN, (b) biển số xe*

Bảng 2. *Độ chính xác của bộ nhận dạng các lớp chữ số khó của hai mạng AlexNet, MobileNet với hai tập dữ liệu SVHN và biển số xe*

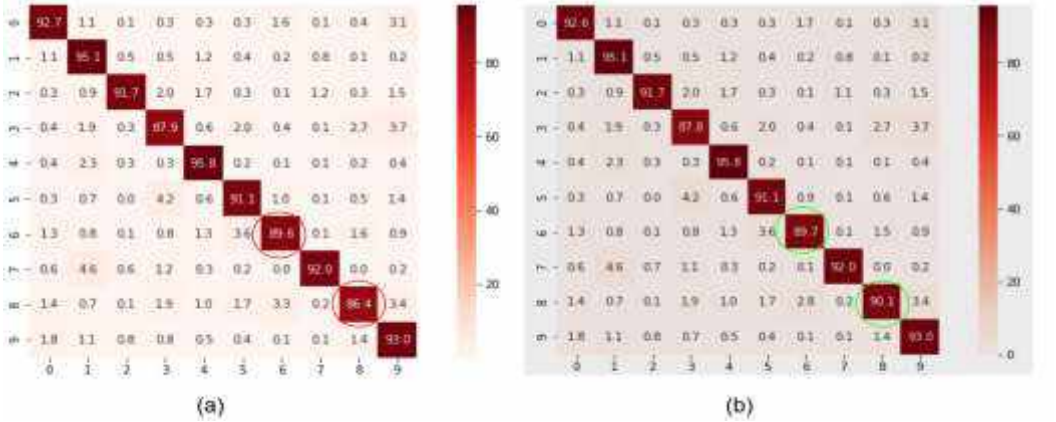
Tập dữ liệu	AlexNet	MobileNet
SVHN (6-8)	96,1 (%)	95,7 (%)
Biển số xe (5,6,8)	98,86 (%)	99,12 (%)

Như vậy, liệu tăng thêm số lớp các mạng CNN có tiếp tục cải thiện độ chính xác của mô hình? Chúng tôi đã thử nghiệm sử dụng thêm 1 lớp CNN thứ 3 để phân loại lại các chữ số khó này nhưng độ chính xác của mô hình gần như không cải thiện. Với cả 2 bộ dữ liệu, nghiên cứu ma trận nhầm lẫn sau khi sử dụng mạng CNN 2 lớp cho thấy, ngoài các lớp chữ số khó đã được chọn để phân loại ở mạng CNN-2 không còn lớp chữ số nào có độ chính xác thấp hơn hẳn các lớp khác. Việc huấn luyện thêm một lớp CNN thứ 3 lại cho các lớp chữ số khó không cải thiện được độ chính xác so với lớp CNN-2. Thêm nữa, phương pháp được đề xuất chỉ thích hợp nếu như tìm được một số lớp chữ số hay bị nhầm với nhau. Nếu một lớp chữ số có độ chính xác thấp nhưng ma trận nhầm lẫn phân bố đều cho các lớp chữ số khác thì sẽ khó ra quyết định huấn luyện lớp CNN-2 cho những lớp chữ số nào. Do vậy, chúng tôi đề xuất chỉ dừng lại ở việc sử dụng 2 lớp CNN, thay vì tiếp tục gia tăng số lớp của CNN của mô hình.

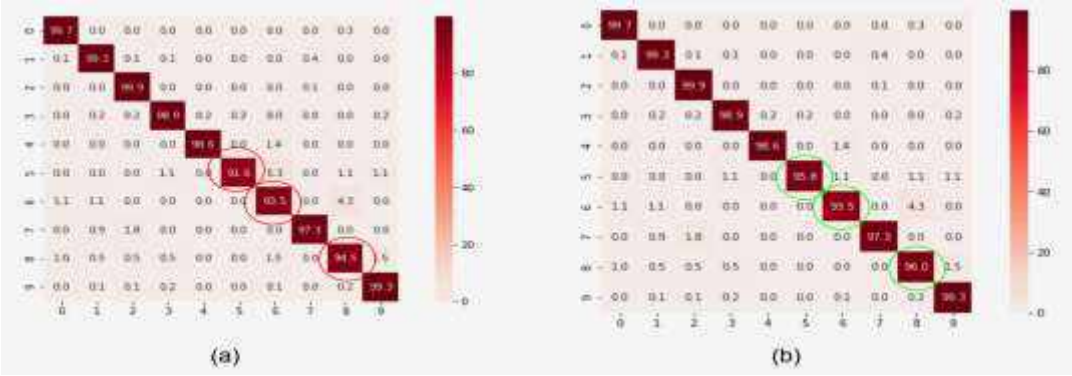
4. Kết luận

Trong bài báo này, chúng tôi đã trình bày phương pháp sử dụng mạng CNN 2 lớp để cải thiện độ chính xác của một số lớp dữ liệu khó phân loại.

Hình 3. *Ma trận nhầm lẫn của bộ nhận dạng CNN-1 với tập dữ liệu xác thực: (a) SHVN và MobileNet, (b) biển số xe và AlexNet*



Hình 4. *Ma trận nhầm lẫn của mô hình MobileNet trên tập dữ liệu SVHN: (a) sử dụng 1 mạng CNN-1, (b) kết hợp 2 mạng CNN*



Hình 5. *Ma trận nhầm lẫn của mô hình AlexNet trên tập dữ liệu biển số xe: (a) sử dụng 1 mạng CNN-1, (b) kết hợp 2 mạng CNN*

Bảng 3. *Độ chính xác của bộ phân loại các kí tự khó và thời gian tính toán cần tăng thêm*

	AlexNet [4]			**MobileNet [5]**		
	1 bộ CNN	**2 bộ CNN**	**Thời gian**	**1 bộ CNN**	**2 bộ CNN**	**Thời gian**
SVHN (6-8)	91,9 (%)	**92,75 (%)**	+3,2 (%)	88 (%)	**89,9(%)**	+4,4 (%)
Biển số xe (5,6,8)	93,2 (%)	**95,1 (%)**	+4,2 (%)	96,8 (%)	**97,5 (%)**	+5,8 (%)

Kết quả thực nghiệm cho thấy phương pháp này giúp gia tăng độ chính xác của các lớp dữ liệu này lên khoảng 1,4% với thêm 4,4% thời gian tính toán. Nên nhớ rằng, khi độ chính xác vượt ngưỡng 95%, việc cải thiện dù chỉ 1% độ chính xác là vô cùng khó khăn. Tính hiệu quả của phương pháp được chứng minh trên tập dữ liệu nhận dạng các chữ số, tuy nhiên, phương pháp có thể áp dụng cho nhiều bộ dữ liệu khác nhau khi mà độ khó trong việc nhận dạng các lớp dữ liệu không đồng đều.

Lời cảm ơn

Bài báo này được hỗ trợ bởi Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam thông qua nhiệm vụ khoa học mã số GUST.STS.NV2017-TT01.

TÀI LIỆU THAM KHẢO/ REFERENCES

[1]. C. Yao, X. Bai, B. Shi, and W. Liu, "Strokelets: A learned multi-scale representation for scene text recognition," *In Proceedings of the IEEE Conference on Computer Vision and PatternRecognition (CVPR)*, 2014, pp. 4042-4049.

[2]. M. Buta, L. Neumann and J. Matas, "FASText: Efficient Unconstrained Scene Text Detector," *2015 IEEE International Conference on Computer Vision (ICCV), Santiago*, 2015, pp. 1206-1214, doi: 10.1109/ICCV.2015.143.

[3]. T. Q. Phan, P. Shivakumara, S. Tian, and C. L. Tan, "Recognizing text with perspective distortion in natural scenes," *In Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2013, pp. 569-576.

[4]. A. Krizhevsky, I. Sutskever, and G Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *In Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1 (NIPS), 2012*.

[5]. A. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, 2017. [Online]. Available: http://arxiv.org/abs/1704.04861 [Accessed May 12, 2018].

[6]. D. Ho, E. Liang, I. Stoica, P. Abbeel, and X. Chen, Population Based Augmentation: Efficient Learning of Augmentation Policy Schedules, 2019, [Online]. Available: https://arxiv.org/abs/1905.05393 [Accessed May 12, 2019].

[7]. M. Galar, A. Fernández, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem:

bagging-, boosting-, and hybrid-based approaches," *IEEE Trans. Syst., Man, Cybernet., Part C: Appl. Rev.*, vol. 42, no. 4, pp. 463-484, 2012.

[8]. Freund, and R. E. Schapire, "A decision-theoretic generalization ofon-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119-139, 1997.

[9]. L. Rokach, "Ensemble-based classifiers,"*Artif. Intell. Rev.*, vol. 33, pp. 1-39, 2010.

[10]. Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and Y. Andrew, "Reading Digits in Natural Images with Unsupervised Feature Learning NIPS," *Workshop on Deep Learning and Unsupervised Feature Learning*, 2011.