

GIẢI PHÁP NHẬN DẠNG KÝ TỰ TIẾNG TRUNG VIẾT TAY DỰA TRÊN MẠNG NƠON TÍCH CHẬP

HANDWRITTEN CHINESE CHARACTER RECOGNITION SOLUTION BASED ON CONVOLUTIONAL NEURAL NETWORK

Vũ Thị Duyên

Học viện Cảnh sát Nhân dân

Ngày nhận bài: 24/3/2023, Ngày chấp nhận đăng: 26/5/2023, Phản biện: PGS. TS Phạm Thanh Hiệp

Tóm tắt:

Gần đây, bài toán nhận dạng ký tự viết tay trở nên phổ biến hơn bởi những ứng dụng đa dạng của nó. So với bài toán nhận dạng các chữ số và chữ cái tiếng Anh viết tay, việc nhận dạng các ký tự tiếng Trung viết tay là một bài toán khó khăn hơn bởi nhiều lý do. Để giải quyết được vấn đề này, bài báo trình bày một giải pháp nhận dạng ký tự tiếng Trung viết tay dựa trên kiến trúc mạng nơon tích chập nổi tiếng LeNet-5. Kết quả thử nghiệm trên tập dữ liệu viết tay CASIA minh họa khả năng dự đoán chính xác ký tự tiếng Trung viết tay trong khoảng thời gian dưới 0,1 giây.

Từ khóa:

Nhận dạng ký tự, nhận dạng ký tự tiếng Trung, mạng nơon, mạng nơon tích chập.

Abstract:

Currently, because of its many applications, the handwritten character recognition problem is growing in popularity. The task of recognizing handwritten Chinese characters poses more of a challenge than that of recognizing handwritten numerals and English alphabets for a variety of reasons. To solve this problem, the paper presents a handwritten Chinese character recognition solution based on the famous convolutional neural network architecture Lenet-5. Experimental results on CASIA handwritten datasets demonstrate the ability to accurately predict handwritten Chinese characters for a period of fewer than 0.1 seconds.

Keywords:

Handwritten character recognition; handwritten chinese character recognition; neural networks; convolutional neural network.

1. GIỚI THIỆU CHUNG

Với sự phát triển nhanh chóng của các công nghệ học sâu (Deep Learning), nhiều bài toán liên quan đến nhận dạng mẫu (Pattern Recognition) đã đạt được những cải tiến đáng kể. Các bài toán thay

đổi từ phát hiện đối tượng và tạo hình ảnh đến tạo thơ. Trong đó, bài toán nhận dạng văn bản cũng là một ví dụ điển hình về việc học các đặc trưng của dữ liệu đầu vào được thực hiện bởi các thuật toán học sâu [1].

Để giải bài toán nhận dạng văn bản ví dụ như nhận dạng ký tự viết tay, mạng nơron tích chập (Convolutional Neural Network) đã trở thành kiến trúc điển hình được lựa chọn trong lĩnh vực học sâu và được sử dụng rộng rãi. Đã có rất nhiều nghiên cứu về việc sử dụng mạng nơron tích chập để nhận dạng các chữ số viết tay, bảng chữ cái tiếng Anh hoặc bảng chữ cái Latinh nói chung. Tuy nhiên, so với bài toán nhận dạng các chữ số viết tay và bảng chữ cái tiếng Anh, việc nhận dạng các ký tự tiếng Trung viết tay là một nhiệm vụ khó khăn hơn bởi nhiều lý do. Đầu tiên, có nhiều ký tự tiếng Trung hơn là các chữ số và ký tự tiếng Anh [2], [3]. Để so sánh, có 10 chữ số cho các bài toán nhận dạng chữ số thông thường và có 26 chữ cái cho tiếng Anh, trong khi có tổng cộng hơn 50.000 ký tự tiếng Trung và khoảng 3.000 ký tự trong số đó được sử dụng hàng ngày. Thứ hai, hầu hết các ký tự tiếng Trung có cấu trúc phức tạp hơn nhiều và bao gồm nhiều nét hơn so với các chữ số hoặc ký tự tiếng Anh. Thứ ba, phong cách viết tay cho ký tự tiếng Trung rất khác nhau, do sự truyền đạt từ thế hệ này sang thế hệ khác. Hơn nữa, sự tồn tại của chữ viết tay nối liền khiến cho việc nhận dạng thậm chí còn khó khăn hơn [4].

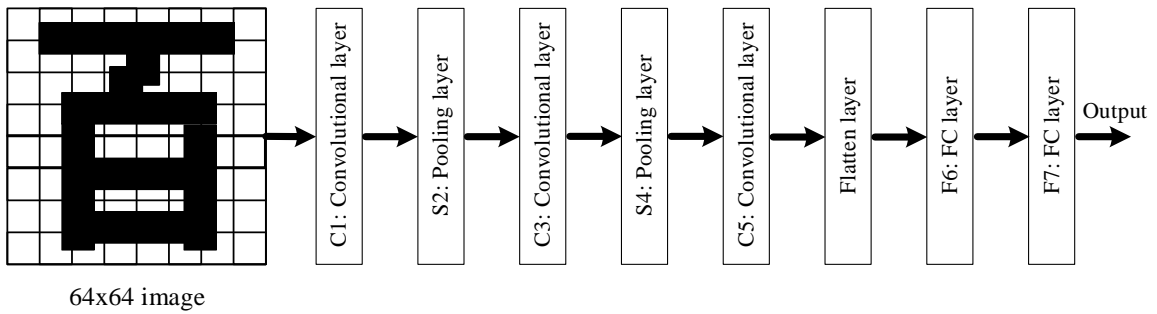
Nhận dạng ký tự tiếng Trung viết tay được tìm thấy trong nhiều ứng dụng, chẳng hạn như phân loại thư, đọc séc ngân hàng, ghi chú sách và ghi chú viết tay. Độ chính xác nhận dạng ký tự cao là điều cần thiết cho sự thành công của nhận

dạng văn bản viết tay [3]. Dựa vào sự cần thiết của việc giải quyết các bài toán trong thực tế này, bài báo trình bày giải pháp dựa trên mạng nơron tích chập LeNet-5 để nhận dạng ký tự tiếng Trung viết tay. Kết quả thử nghiệm đã cho thấy rằng giải pháp đưa ra đã nhận dạng chính xác các ký tự yêu cầu.

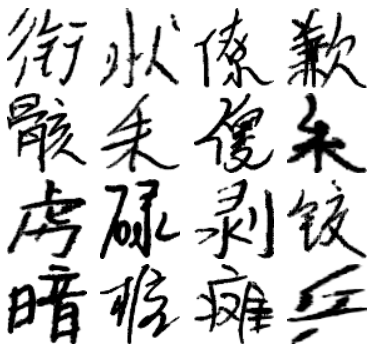
Phần còn lại của bài báo được tổ chức như sau. Mục 2 mô tả giải pháp nhận dạng ký tự tiếng Trung viết tay. Mục 3 kiểm chứng giải pháp và so sánh mô hình dựa trên kiến trúc của LeNet-5 trong giải pháp với mô hình khi sử dụng các siêu tham số (Hyperparameter) ngẫu nhiên. Mục 4 đưa ra kết luận bài báo.

2. GIẢI PHÁP NHẬN DẠNG KÝ TỰ TIẾNG TRUNG VIẾT TAY

Tập dữ liệu tiếng Trung viết tay được sử dụng trong bài báo này được xây dựng bởi Phòng nghiên cứu Quốc gia về nhận dạng mẫu thuộc Viện Tự động hóa của Viện Hàn lâm Khoa học Trung Quốc (CASIA). Các mẫu dữ liệu viết tay được tạo ra bởi 1.020 người viết sử dụng bút Anoto trên giấy, trong đó tập dữ liệu HWDB1.1, được viết bởi 300 người viết, chứa các ký tự viết tay độc lập khoảng 3.755 ký tự tiếng Trung cấp 1 GB2312-80 và 171 chữ số và ký hiệu [5]. Để đơn giản, 100 lớp ký tự tiếng Trung trong HWDB1.1 được sử dụng để kiểm chứng giải pháp. Một số ký tự tiếng Trung viết tay trong tập dữ liệu huấn luyện và thử nghiệm được minh họa như trong Hình .



Hình 1. Mô hình mạng nơron tích chập dựa trên kiến trúc LeNet-5



Hình 2. Ký tự tiếng Trung viết tay trong tập dữ liệu HWDB1.1

Để nhận dạng được ký tự tiếng Trung viết tay, giải pháp được dựa trên mô hình mạng nơron tích chập nổi tiếng LeNet-5. Mô hình này, được đề xuất bởi LeCun et al [6], đã được ứng dụng trong hàng loạt các bài toán nhận dạng [7] như chữ số viết tay trong tập dữ liệu MNIST, đối tượng trong tập dữ liệu CIFAR. Mô hình dựa trên kiến trúc của LeNet-5 để nhận dạng tiếng Trung được minh họa như hình 2. Trong mô hình này, đầu vào là các ảnh chữ viết tay kích thước 64x64 [4] và đầu ra là xác suất dự đoán ảnh đầu vào tương ứng với ký tự nào trong tập dữ liệu. Mô hình gồm 3 lớp tích chập (Convolutional layer), 2 lớp tổng hợp (Pooling layer), lớp Flatten để chuyển các ma trận nhiều chiều thành 1 chiều, và 2 lớp kết nối đầy đủ (FC

layer: Fully connected layer). Siêu tham số tại các lớp này được tổng hợp trong bảng 1. Bên cạnh đó, giải pháp được so sánh với mô hình sử dụng các siêu tham số ngẫu nhiên (gọi là Mô hình ngẫu nhiên), để minh chứng rằng việc lựa chọn siêu tham số đối với một mô hình mạng nơron là vô cùng quan trọng. Giả sử mô hình ngẫu nhiên và mô hình dựa trên LeNet-5 có số lượng lớp, dữ liệu đầu vào, kích thước lớp đầu ra là như nhau. Không gian tìm kiếm của các siêu tham số cho mô hình ngẫu nhiên được mô tả như trong bảng 1. Trong đó, một số siêu tham số phổ biến đối với từng lớp được chọn làm ứng cử trong không gian tìm kiếm như kích thước bộ lọc, lớp tổng hợp, hàm kích hoạt, thuật toán tối ưu như mô tả trong bảng 1.

Bảng 1. Siêu tham số cho mạng nơron tích chập

	Mô hình dựa trên LeNet-5	Mô hình ngẫu nhiên
Số bộ lọc tại lớp C1	6	Số nguyên ngẫu nhiên từ 5 đến 150 với bước nhảy 1
Số bộ lọc tại lớp C3	16	

Số bộ lọc tại lớp C5	120	
Số nơron tại lớp F6	84	
Kích thước bộ lọc tại C1, C3, C5	5	Số nguyên ngẫu nhiên từ 3 đến 5 với bước nhảy 1
Lớp tổng hợp tại S2 và S4	Lớp tổng hợp trung bình	Lớp tổng hợp trung bình hoặc lớn nhất
Số nơron tại lớp F7	100	
Hàm kích hoạt tại C1, C3, C5, F6	Hyperbolic tangent (Tanh)	ReLu hoặc Tanh hoặc Sigmoid hoặc Elu hoặc Linear hoặc Softplus hoặc Swish
Hàm kích hoạt tại F7	Softmax	
Hệ số học	0.01	Số thực ngẫu nhiên từ $1e-5$ đến $1e-2$ với bước nhảy $1e-5$
Hàm mất mát	Categorical cross-entropy	
Thuật toán tối ưu	SGD	Adamax hoặc Adam hoặc Nadam hoặc SGD

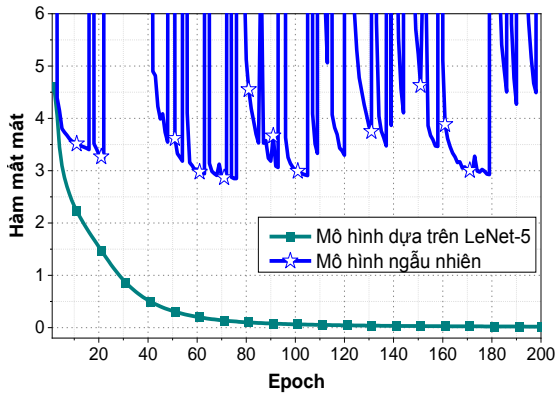
3. KẾT QUẢ THỬ NGHIỆM

Mô hình mạng nơron tích chập được xây dựng dựa trên giao diện lập trình ứng dụng Keras [8] và được thực thi trên

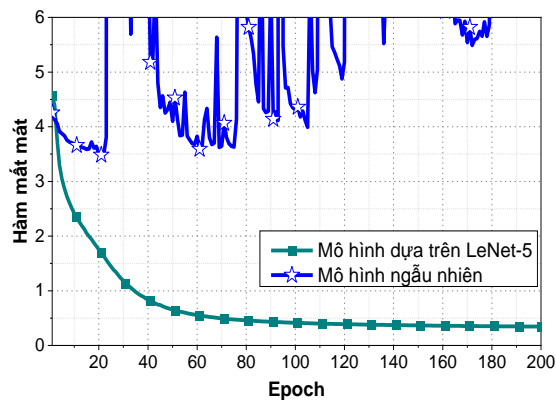
Google Colaboratory. Kết quả minh họa là giá trị trung bình của 100 lần chạy độc lập. Để mô hình có thể dự đoán được, đầu tiên, mô hình được huấn luyện với khoảng 25.000 hình ảnh qua 200 epochs, sau đó được thử nghiệm trên 5.000 hình ảnh (không thuộc tập dữ liệu được huấn luyện).

Hình 3 và hình 4 lần lượt minh họa giá trị hàm mất mát trên tập dữ liệu huấn luyện và thử nghiệm. Mô hình dựa trên kiến trúc của LeNet-5 đã cho thấy rằng chúng có thể gần như hội tụ kể từ epoch thứ 80 trở đi và không bị học thuộc (Overfitting) hay học chưa đủ (Underfitting). Trong khi đó, mô hình sử dụng siêu tham số ngẫu nhiên đã không thể hội tụ với giá trị hàm mất mát liên tục dao động. Điều này dẫn đến việc nhận dạng các ký tự tiếng Trung viết tay không được chính xác. Hình 5 minh họa độ chính xác nhận dạng các ký tự viết tay trên tập dữ liệu thử nghiệm qua từng epoch. Kết quả chỉ ra rằng, chỉ với khoảng 80 epoch, giải pháp nhận dạng ký tự tiếng Trung viết tay đã đạt độ chính xác lên tới 95%, trong khi đó, mô hình ngẫu nhiên chỉ đạt khoảng 32%. Đối với từng ký tự tiếng Trung viết tay, số lượng ảnh thử nghiệm và độ chính xác nhận dạng đối với từng mô hình được biểu diễn trong hình 6. Khi sử dụng mô hình dựa trên LeNet-5, hầu hết các ký tự đều đạt độ chính xác nhận dạng trên 80%, trong khi sử dụng mô hình ngẫu nhiên chỉ đạt khoảng 25% đến 40%. Một khi mô hình đã được huấn luyện, với mỗi ảnh đầu vào, mô hình dựa trên LeNet-5 chỉ cần khoảng 0,052 giây để có thể trả về kết quả (được

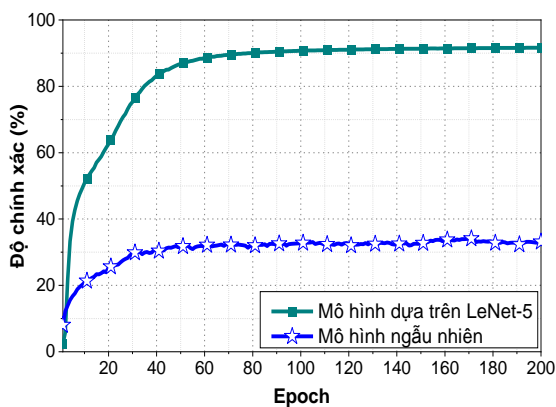
thực thi trên Google Colaboratory được trang bị NVIDIA Tesla T4 16GB GDDR6 PCIe 3.0).



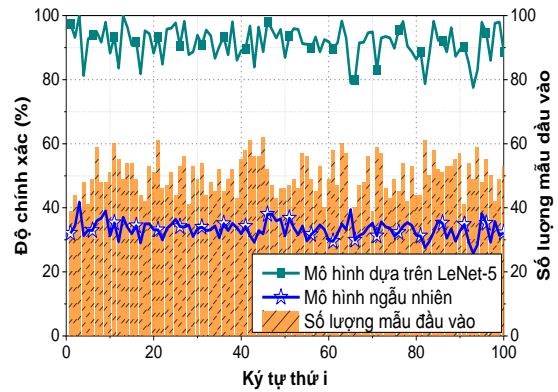
Hình 3. Giá trị hàm mất mát trên tập dữ liệu huấn luyện



Hình 4. Giá trị hàm mất mát trên tập dữ liệu thử nghiệm



Hình 5. Độ chính xác nhận dạng trên tập dữ liệu thử nghiệm



Hình 6. Độ chính xác nhận dạng trên từng ký tự

4. KẾT LUẬN

Bài báo này đã trình bày giải pháp nhận dạng ký tự tiếng Trung viết tay dựa mô hình mạng nơron tích chập. Mô hình trong giải pháp này dựa vào kiến trúc mạng của mô hình LeNet-5 đã cho thấy rằng chỉ với khoảng 80 epoch huấn luyện, mô hình đã có thể phân loại chính xác lên tới 95%. Mô hình dựa vào LeNet-5 cũng đã được so sánh với mô hình sử dụng các siêu tham số ngẫu nhiên trong cùng điều kiện như cùng tập dữ liệu, cùng số lượng lớp trong kiến trúc mạng. Kết quả của mô hình dựa vào LeNet-5 đã trả về kết quả vượt trội so với mô hình ngẫu nhiên. Tuy nhiên, để cải thiện độ chính xác phân loại, giải pháp tối ưu siêu tham số là chủ đề hấp dẫn nên được áp dụng vào các mô hình học sâu. Giải pháp này nhằm xác định tập siêu tham số tối ưu để cải thiện hiệu suất của mô hình như độ chính xác phân loại, thời gian tính toán. Trong tương lai, giải pháp tối ưu siêu tham số cũng như giải pháp nhận dạng tiếng Trung viết tay trực tuyến hay sử dụng mô

hình phức tạp trên tập dữ liệu lớn hơn hay triển khai mô hình sau khi huấn luyện trên các thiết bị di động Android, iOS, hay trên các thiết bị nhúng như Raspberry Pi và vi điều khiển sẽ là một chủ đề nghiên cứu đầy hứa hẹn.

LỜI CẢM ƠN

Tác giả chân thành cảm ơn sự góp ý và hỗ trợ về kỹ thuật của Thạc sĩ Nguyễn Văn Cường - Khoa Điện tử - Trường Đại học Công nghiệp Hà Nội để thực hiện nghiên cứu này.

TÀI LIỆU THAM KHẢO

- [1] P. Melnyk, Z. You, and K. Li, "A high-performance CNN method for offline handwritten Chinese character recognition and visualization," *Soft Computing*, vol. 24, no. 11, pp. 7977–7987, 2019.
- [2] N. Aleskerova and A. Zhuravlev, "Handwritten Chinese characters recognition using two-stage hierarchical convolutional neural network," 2020 17th International Conference on Frontiers in Handwriting Recognition, 2020.
- [3] X.-Y. Zhang, Y. Bengio, and C.-L. Liu, "Online and offline handwritten Chinese character recognition: A comprehensive study and new benchmark," *Pattern Recognition*, vol. 61. Elsevier BV, pp. 348–360, 2017.
- [4] Zhang, Y., "Deep convolutional network for handwritten Chinese character recognition," Computer Science Department, Stanford University, 2015.
- [5] C.-L. Liu, F. Yin, D.-H. Wang, Q.-F. Wang, "CASIA online and offline Chinese handwriting databases," *Proceeding of the 11th International Conference on Document Analysis and Recognition (ICDAR)*, Beijing, China, pp. 37-41, 2011.
- [6] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [7] Z.-H. Zhang, Z. Yang, Y. Sun, Y.-F. Wu, and Y.-D. Xing, "Lenet-5 convolution neural network with mish activation function and fixed memory step gradient descent method," 2019 16th International Computer Conference on Wavelet Active Media Technology and Information Processing, 2019.
- [8] François Chollet et al., Keras, 2015 [Online]. Available: <https://keras.io>.

Giới thiệu tác giả:



Tác giả Vũ Thị Duyên tốt nghiệp đại học và nhận bằng Thạc sĩ tại Trường Đại học Ngoại ngữ - Đại học Quốc gia Hà Nội năm 2002 và năm 2009.

Hướng nghiên cứu: tiếng Trung Quốc, giải pháp nâng cao hiệu quả giảng dạy và học tập tiếng Trung Quốc.