THIẾT KẾ MÔ HÌNH MẠNG NƠ RON NHÂN CHẬP CHO BÀI TOÁN NHẬN DẠNG GIỚI TÍNH TỪ ẢNH MẶT NGƯỜI

DESIGNING A CONVOLUTIONAL NEURAL NETWORK FOR GENDER IDENTIFICATION FROM FACIAL IMAGES

NGUYỄN HỮU TUÂN*, NGUYỄN VĂN THỦY

Khoa Công nghệ Thông tin, Trường Đại học Hàng hải Việt Nam

**Email liên hệ: huu-tuan.nguyen@vimaru.edu.vn*

Tóm tắt

Giới tính là một trong những thông tin quan trọng và có ích có thể xác định từ ảnh mặt người. Các kỹ thuật áp dụng cho bài toán nhận dạng giới tính được công bố gần đây đều dựa trên các phương pháp học sâu và cho các kết quả cao hơn so với cách tiếp cận truyền thống dựa trên các đặc trưng cục bộ được trích chọn từ các thuật toán trích chọn đặc trưng từ ảnh. Trong bài báo này, nhóm tác giả tập trung vào việc thiết kế một mô hình mạng nơ ron nhân chập và kết hợp với việc áp dụng các kỹ thuật tăng cường dữ liệu để đưa ra một hệ thống giải quyết bài toán. Kết quả thực nghiệm thu được trên tập dữ liệu ảnh mặt người công cộng LFW cho thấy hệ thống đề xuất đạt được tỉ lệ chính xác cao (97,5%) và tương đương với các hệ thống đã được công bố.

***Từ khóa**: Nhận dạng giới tính, học sâu, mạng nơ ron nhân chập, tăng cường dữ liệu, LFW.*

Abstract

Gender is among the most important and useful information that can be identified from human facial images. Recent techniques for gender classification problem have been mostly based on deep learning methods and have gained higher results than conventional approaches which are relied on local features extracted from input pictures. In this paper, we focus on building a convolutional neural network and combine several data augmentation methods to build up a gender classification system. Obtained experimental results upon public face image database LFW show that our system achieves high accuracies (97.5%) and is compared with published works in the literature.

***Keywords**: Gender classification, deep learning, convolutional neural network, data augmentation, LFW.*

1. Giới thiệu

Trong số các dữ liệu sinh trắc học của một con người, hình ảnh khuôn mặt là nguồn dữ liệu hữu ích nhất vì từ đó có thể xác định được nhiều thông tin quan trọng liên quan như danh tính, giới tính, độ tuổi, cảm xúc, dân tộc, độ hấp dẫn. Thông tin giới tình từ ảnh mặt người là một thông tin được quan tâm nhiều bởi các nhà khoa học và các công ty công nghiệp vì từ đó có thể xác định được xu hướng tiêu dùng, loại hình dịch vụ cần cung cấp cho khách hàng và xây dựng các hệ thống tương tác người máy. Một hệ thống nhận dạng giới tính dựa trên ảnh mặt người thường gồm các bước sau: 1 - phát hiện vùng ảnh mặt người trong ảnh input, 2 - tiền xử lý, 3 - trích chọn đặc trưng, 4 - học để giảm số chiều và loại bỏ các thông tin dư thừa, 5 - phân lớp, trong đó kết quả ouput của bước trước là dữ liệu input của bước sau. Các cách tiếp cận cũ thường dựa chủ yếu vào thuật toán trích chọn đặc trưng được dùng ở bước 3. Gần đây, các hệ thống nhận dạng giới tính mới công bố có sự dịch chuyển sang sử dụng các kỹ thuật học sâu dựa trên các mạng nơ ron phức tạp, sử dụng số lượng ảnh huấn luyện lớn, đòi hỏi thời gian tính toán huấn luyện lâu nhưng cho kết quả tốt hơn so với các hệ thống dựa trên kỹ thuật trích chọn đặc trưng cục bộ.

Học sâu (Deep learning) là một lĩnh vực con của lĩnh vực học máy (machine learning) với các mô hình toán học gọi là mạng nơ ron (neural network) có cấu trúc được xây dựng dựa trên sự mô phỏng cấu trúc và chức năng của bộ não con người. Khái niệm mạng nơ ron không phải là một khái niệm mới mà đã được đề xuất từ năm 1959 [1]. Tuy nhiên các mạng nơ ron thời kỳ đầu (còn được gọi là các mạng nơ ron truyền thống hoặc nông - swallow) có cấu trúc đơn giản với 2-3 lớp ẩn nằm giữa lớp input và output có những hạn chế cố hữu: do cấu trúc đơn giản nên sức mạnh của mạng không lớn, độ chính xác khi áp dụng vào các bài toán nhận dạng với dữ liệu có tính đa dạng không cao, không tận dụng được nguồn dữ liệu lớn để cải tiến sức mạnh của mạng. Mạng học sâu (deep neural network) là một mở rộng của mạng nơ ron truyền thống với nhiều lớp ẩn phức tạp ở giữa lớp input và output, sử dụng các hàm biến đổi phi tuyến cho việc trích chọn đặc trưng và biến đổi các đặc trưng, trong đó kết quả output của lớp trước sẽ là dữ liệu input cho lớp sau. Do cấu trúc phức tạp nên mạng học sâu không có cấu trúc kết nối đầy đủ cho tất cả các lớp ẩn. Các mạng học sâu

sử dụng các kỹ thuật học có giám sát và không có giám sát cho các bài toán phân lớp. Sức mạnh của mạng học sâu so với mạng nơ ron truyền thống có được là do cấu trúc của mạng cho phép chúng có thể học và trích chọn được các biểu diễn đặc trưng của dữ liệu đầu vào ở nhiều mức khác nhau. Có nhiều mô hình mạng học sâu khác nhau và chúng thích hợp cho các bài toán khác nhau nhưng trong nội dung của bài báo này, chúng tôi tập trung vào mạng nơ ron nhân chập (Convolutional neural network) vì đây là mô hình mạng phù hợp nhất với các bài toán nhận dạng hình ảnh nói chung và bài toán nhận dạng giới tính nói riêng.

Cấu trúc các phần tiếp theo của bài toán bao gồm: phần 2 trình bày về một số nghiên cứu liên quan, phần 3 trình bày về hệ thống đề xuất và các kỹ thuật tăng cường dữ liệu được sử dụng, phần 4 trình bày về kết quả thực nghiệm trên cơ sở dữ liệu ảnh LFW, cuối cùng là phần kết luận.

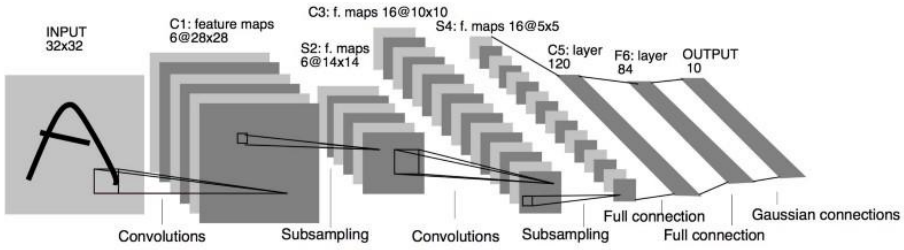
2. Các nghiên cứu liên quan

Các hệ thống nhận dạng giới tính truyền thống thường dựa trên các phương pháp trích chọn đặc trưng cục bộ, chẳng hạn như các công bố [2], [3], [4]. Tuy nhiên gần đây các đề xuất mới dần chuyển sang sử dụng các tiếp cận dựa trên các mạng học sâu. Trong [5], các tác giả đã sử dụng một mạng nhân chập cho cả hai bài toán nhận dạng tuổi và giới tính với kết quả tốt hơn so với các tiếp cận truyền thống. Eidinger và các cộng sự [6] đã xây dựng một mô hình dựa trên mạng tin sâu (deep bilief network) để nhận dạng các ảnh ở các điều kiện phức tạp. Cũng dựa trên mạng tin sâu, một công trình khác phải kể đến do Zhang và các cộng sự đề xuất trong [7]. Lao và các cộng sự [8] đã dựa trên mạng học sâu với các kỹ thuật và đặc trưng cục bộ để nhận dạng giới tính. Nói chung so với cách tiếp cận truyền thống cách tiếp cận dựa trên mạng học sâu có các ưu điểm sau: độ chính xác cao hơn đặc biệt là khi làm việc với các ảnh có điều kiện phức tạp, đa dạng về ánh sáng, độ mờ, cảm xúc, che khuất. Bên cạnh đó, các mạng học sâu cũng sử dụng được hết nguồn dữ liệu ảnh có số lượng và kích thước lớn. Ngược lại các hệ thống dựa trên học sâu cũng có một số yêu cầu: cần dữ liệu lớn để huấn luyện mô hình mạng, cần có phần cứng đặc thù đủ mạnh để thực hiện việc huấn luyện (các hệ thống có chip GPU chuyên dụng với bộ nhớ lớn) và thời gian huấn luyện lâu (từ vài giờ, cho tới vài ngày hoặc thậm chí vài tuần).

3. Thiết kế mô hình mạng nơ ron nhân chập và áp dụng các kỹ thuật tăng cường dữ liệu cho bài toán nhận dạng giới tính

3.1. Mô hình mạng nơ ron nhân chập đề xuất

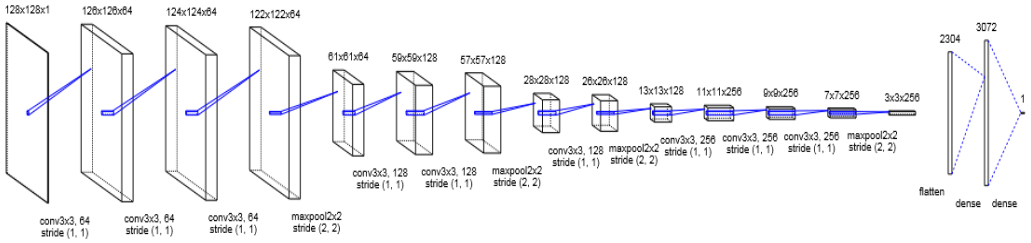
Mạng nơ ron học sâu khi sử dụng cho bài toán nhận dạng giới tính dựa trên ảnh mặt người sẽ thực hiện 3 công việc tương ứng với 3 bước trong hệ thống nhận dạng như đã đề cập ở phần 1: trích chọn đặc trưng, học và phân lớp. Do đó, độ chính xác của hệ thống hoàn toàn phụ thuộc vào sức mạnh của mạng nơ ron sử dụng. Các thành phần của một mạng nơ ron nhân chập (Hình 1) gồm có: lớp input, lớp nhân chập (convolution), lớp tổng hợp đặc trưng (pooling), lớp dropout, lớp output. Lớp input gồm các dữ liệu từ ảnh input đưa vào mạng, lớp output là lớp có số nút tương ứng với số nhãn đầu ra, trong bài toán nhận dạng giới tính là 2 nút. Lớp input được kết nối đầy đủ với lớp ẩn đầu tiên của mạng, còn lớp ẩn cuối cùng của mạng (nằm trước lớp output) sẽ kết nối với lớp output.



***Hình 1. Mô hình mạng nhân chập đề xuất bởi LeCun và các cộng sự** [9]*

Các thành phần của mạng nơ ron nhân chập có vai trò khác nhau. Lớp nhân chập đóng vai trò là các bộ lọc với mục đích sinh ra các ma trận đặc trưng (feature map) từ dữ liệu input nhận được từ lớp trước. Lớp tổng hợp (pooling) thực hiện việc chọn lọc và giữ lại các đặc trưng quan trọng nhất. Việc chọn lọc là cần thiết vì mỗi mạng sử dụng nhiều nhân khác nhau và sẽ sinh ra số ma trận đặc trưng rất lớn, nếu giữ nguyên sẽ nâng chi phí tính toán lên rất lớn một cách không cần thiết. Riêng lớp dropout mới được bổ sung trong các mô hình gần đây [10] và thường được dùng sau mỗi khối (block) gồm lớp nhân chập và lớp tổng hợp dựa trên một quan sát là mạng sẽ mạnh hơn nếu một phần ngẫu nhiên của mỗi block được bỏ ra ngoài quá trình học (giống như một giám đốc có n trợ lý nhưng luôn chỉ dùng khoảng 80% số trợ lý để nếu có ai đó nghỉ thì hệ thống vẫn hoạt động tốt).

Mô hình chúng tôi đề xuất cho bài toán nhận dạng giới tính từ ảnh mặt người có cấu trúc như trong Hình 2 bên dưới.



Hình 2. Mô hình mạng nơ ron nhân chập đề xuất cho bài toán nhận dạng giới tính

Có thể thấy trên Hình 2 mô hình đề xuất của chúng tôi gồm 8 khối chính, mỗi khối gồm lần lượt 64, 64, 128, 128, 128, 256, 256, 256 bộ nhân chập. Các lớp tổng hợp sử dụng hàm max (maxpooling) lần lượt được sử dụng sau mỗi khối nhân chập. Các lớp dropout với tỉ lệ 0,5 được sử dụng sau hàm tổng hợp của mỗi khối. Tổng số tham số của mô hình mạng là hơn 40 triệu.

3.2. Kỹ thuật tăng cường dữ liệu

Bên cạnh vai trò chủ đạo của mô hình mạng sử dụng cho mỗi hệ thống nhận dạng giới tính dựa trên ảnh mặt đối với độ chính xác của hệ thống, một yếu tố cũng rất quan trọng nữa là dữ liệu. Thông thường khi chưa đạt tới ngưỡng, các hệ thống sẽ càng chính xác hơn nếu dữ liệu học của nó càng nhiều. Tuy nhiên đối với bài toán nhận dạng hình ảnh, số dữ liệu ảnh cho hệ thống học thường quá ít (ví dụ cơ sở dữ liệu LFW [11] chỉ có 13.233 bức ảnh) do đó cần các kỹ thuật tăng cường số lượng ảnh để tránh hiện tượng quá khớp (overfitting) và cải thiện hiệu năng của mạng. Trong bài báo này chúng tôi sử dụng 5 kỹ thuật xử lý ảnh (Hình 3) để sinh 5 ảnh từ 1 ảnh input và do đó tổng số ảnh huấn luyện sẽ là 6*N với N là số ảnh huấn luyện. Các kỹ thuật cụ thể gồm có: cân bằng histogram, xoay, dịch, cắt xén (shear), lấy đối xứng.

Hình 3. Ảnh mặt người và một số kỹ thuật tăng cường dữ liệu

4. Kết quả thực nghiệm và phân tích

Cơ sở dữ liệu LFW

Để đánh giá độ chính xác của mô hình đề xuất chúng tôi sử dụng sơ sở dữ liệu ảnh công cộng LFW theo giao thức chuẩn được đề xuất bởi Dago và các cộng sự [12]. Tập ảnh LFW gồm 13.233 ảnh được chia thành 5 tập con với số lượng ảnh xấp xỉ nhau (xem chi tiết trong [12]) để thực hiện 5 lần thử nghiệm, mỗi lần thử nghiệm 4 tập con được dùng làm tập huấn luyện cho mô hình, tập con còn lại được dùng làm tập test và kết quả được lấy là trung bình cộng của 5 lần chạy.

Kết quả thực nghiệm

Bảng 1. Kết quả thực nghiệm trên cơ sở dữ liệu LFW so sánh với một số công bố khác

Phương pháp	Độ chính xác (%)	Số ảnh sử dụng để thử nghiệm
[8]	95,6	13.233
[3]	95,6	13.233
Phương pháp đề xuất - không có tăng cường dữ liệu	**95,7**	**13.233**
[13]	96,9	13.233
[14]	97,3	13.233
Phương pháp đề xuất - có tăng cường dữ liệu	**97,5**	**13.233**

Kết quả thực nghiệm của hệ thống đề xuất được trình bày trong Bảng 1 sau 100 epoch cho mỗi lần huấn luyện. Dựa vào Bảng 1 chúng ta thấy rõ việc sử dụng kết hợp các kỹ thuật tăng cường dữ liệu cho kết quả tốt hơn hẳn (độ chính xác tăng từ 95,7% lên 97,5%), điều này là do các kỹ thuật tăng cường dữ liệu một mặt tăng số lượng ảnh huấn luyện, mặt khác tăng độ da dạng của các ảnh

được huấn luyện nên mạng sẽ mạnh hơn và cho kết quả cao hơn. Kết luận quan trọng thứ 2 là kết quả của hệ thống do chúng tôi đề xuất cao hơn một số phương pháp được công bố gần đây. Điều này chứng tỏ mô hình mạng nơ ron nhân chập do chúng tôi đề xuất hiệu quả cho bài toán nhận dạng giới tình từ ảnh mặt người.

5. Kết luận

Trong bài báo này, nhóm tác giả đã nghiên cứu và đề xuất một mô hình mạng nơ ron học sâu sử dụng các bộ lọc nhân chập áp dụng cho bài toán nhận dạng giới tính từ ảnh mặt người. Các kỹ thuật xử lý ảnh khác nhau cũng được áp dụng cho việc tăng cường dữ liệu huấn luyện cho mô hình. Hệ thống đề xuất đã được thử nghiệm với cơ sở dữ liệu ảnh mặt người công cộng LFW theo giao thức chuẩn. Việc so sánh với các cách tiếp cận khác cho thấy hệ thống đề xuất khá hiệu quả và kỹ thuật tăng cường dữ liệu đóng vai trò quan trọng trong việc tăng cường sức mạnh cho mạng nơ ron nhân chập. Trong tương lai các tác giả mong muốn thử nghiệm hệ thống đề xuất cho các bài toán khác có liên quan tới nhận dạng mặt như nhận dạng cảm xúc, độ tuổi. Một hướng khác mà nhóm tác giả rất quan tâm là sử dụng thêm các kỹ thuật tăng cường dữ liệu để nâng cao độ chính xác của hệ thống.

TÀI LIỆU THAM KHẢO

[1] X. Liu, B. V. K. V. Kumar, Y. Ge, C. Yang, J. You, and P. Jia, “Normalized face image generation with perceptron generative adversarial networks,” in *2018 IEEE 4th International Conference on Identity, Security, and Behavior Analysis (ISBA)*, pp. 1-8, 2018.

[2] A. R. Ardakany and A. M. Jula, “Gender Recognition Based On Edge Histogram,” *Int. J. Comput. Theory Eng.*, vol. 4, no. 2, pp. 127-130, 2012.

[3] A. M. Mirza, M. Hussain, H. Almuzaini, G. Muhammad, H. Aboalsamh, and G. Bebis, “Gender Recognition Using Fusion of Local and Global Facial Features,” in *Advances in Visual Computing*, Springer, pp. 493-502, 2013.

[4] H. Moeini, K. Faez, and A. Moeini, “Real-world gender classification via local Gabor binary pattern and three-dimensional face reconstruction by generic elastic model,” *IET Image Process.*, vol. 9, no. 8, pp. 690-698, Aug. 2015.

[5] G. Levi and T. Hassner, “Age and gender classification using convolutional neural networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 34-42, 2015.

[6] E. Eidinger, R. Enbar, and T. Hassner, “Age and Gender Estimation of Unfiltered Faces,” *IEEE Trans. Inf. Forensics Secur.*, vol. 9, no. 12, pp. 2170-2179, Dec. 2014.

[7] K. Zhang, L. Tan, Z. Li, and Y. Qiao, “Gender and Smile Classification Using Deep Convolutional Neural Networks,” pp. 739-743, 2016.

[8] Z. Liao, S. Petridis, and M. Pantic, “Local Deep Neural Networks for Age and Gender Classification,” *ArXiv Prepr. ArXiv170308497*, 2017.

[9] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.

[10] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” *J. Mach. Learn. Res.*, vol. 15, pp. 1929-1958, 2014.

[11] G. B. Huang, M. Mattar, T. Berg, E. Learned-Miller, and others, “Labeled Faces in the Wild: A Database forStudying Face Recognition in Unconstrained Environments,” 2008.

[12] P. Dago-Casas, D. González-Jiménez, L. L. Yu, and J. L. Alba-Castro, “Single-and cross-database benchmarks for gender classification under unconstrained settings,” in *Computer vision workshops (ICCV Workshops), 2011 IEEE international conference on*, pp. 2152-2159, 2011.

[13] S. Jia and N. Cristianini, “Learning to classify gender from four million images,” *Pattern Recognit. Lett.*, vol. 58, pp. 35-41, Jun. 2015.

[14] J. Mansanet, A. Albiol, and R. Paredes, “Local Deep Neural Networks for gender recognition,” *Pattern Recognit. Lett.*, vol. 70, pp. 80-86, Jan. 2016.

Ngày nhận bài: 24/4/2019
Ngày nhận bản sửa: 09/5/2019
Ngày duyệt đăng: 13/5/2019