

NGHIÊN CỨU MẠNG RESNET VÀ ỨNG DỤNG CHO BÀI TOÁN PHÂN LOẠI ẢNH

RESEARCH RESNET NETWORK AND APPLICATION OF SELECTION FOR IMAGE CLASSIFICATION PROBLEM

Nguyễn Hồng Phong^{1,*}, Hạ Ngọc Nam¹, Vi Văn Oanh¹,
Đặng Đình Phúc¹, Vũ Thị Bình Yên², Trần Hùng Cường³

TÓM TẮT

Trong thời đại khoa học công nghệ ngày nay, kỹ thuật học sâu hay kỹ thuật học nhiều tầng (deep learning) đã được ứng dụng rộng rãi. Kỹ thuật này giúp hiện thực hóa một hệ thống machine learning với hiệu quả vượt trội. Các thuật toán trong kỹ thuật học sâu dựa vào các mạng lưới thần kinh, trong đó các lớp của các nút giống như nơ-ron bắt chước cách não bộ của con người phân tích thông tin tìm các mối quan hệ có ý nghĩa, các lớp trong mạng nơ-ron lọc và sắp xếp thông tin, mỗi lớp nơ-ron giao tiếp và tinh chỉnh đầu ra từ lớp trước. Sử dụng mạng ResNet giúp bài toán đơn giản và tập trung vào cải thiện thông tin thông qua độ dốc của mạng, kiến trúc này có thể được huấn luyện mạng nơ-ron với độ sâu hàng nghìn lớp và nó nhanh chóng trở thành kiến trúc phổ biến nhất trong thị giác máy tính. Các thiết bị thông minh có khả năng nhận diện cảm xúc, cử chỉ con người trở nên cần thiết.

Từ khóa: Resnet, nhận diện cảm xúc, cử chỉ.

ABSTRACT

In today's science and technology era, deep learning or deep learning techniques have been widely applied. This technique helps to realize a machine learning system with outstanding efficiency. Deep learning algorithms rely on neural networks, in which layers of neural-like nodes mimic the way the human brain analyzes information to find meaningful relationships, Layers in a neural network filter and organize information, each layer of neurons communicating and refining the output from the previous layer. Using ResNet network to simplify the problem and focus on improving information through the gradient of the network, this architecture can be trained neural network with a depth of thousands of layers and it quickly becomes a popular architecture. most variable in computer vision. Smart devices capable of recognizing human emotions and gestures are necessary.

Keywords: Resnet, emotion recognition, gestures.

¹Lớp CNTT 05 - K14, Khoa CNTT, Trường Đại học Công nghiệp Hà Nội

²Lớp Kỹ thuật phần mềm 02 - K14, Khoa CNTT, Trường Đại học Công nghiệp Hà Nội

³Khoa CNTT, Trường Đại học Công nghiệp Hà Nội

*Email: phongnguyen2001k38gmail.com

1. GIỚI THIỆU

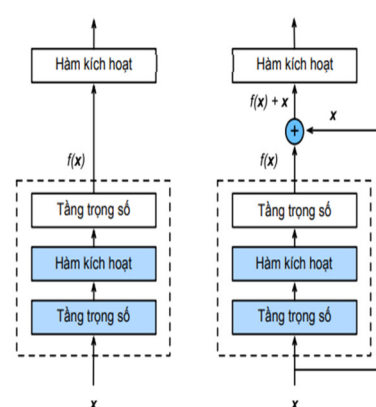
Cử chỉ là một trong những phương thức giao tiếp hữu hiệu của con người. Việc nhận diện được cử chỉ sẽ giúp các thiết bị có thể thực hiện yêu cầu hoặc đưa ra những quyết

định tự động. Bên cạnh đó việc phân tích và nhận diện cảm xúc cũng rất quan trọng, nó giúp các hệ thống được tích hợp có thể điều chỉnh hành vi cho phù hợp với khách hàng. Đặc biệt là đối với các doanh nghiệp trong thời đại số, khi mà dữ liệu người dùng trở thành tài nguyên quý giá nhất.

ResNet không phải là kiến trúc đầu tiên sử dụng các kết nối tắt, Highway Network là một ví dụ. Trong thử nghiệm cho thấy Highway Network hoạt động không tốt hơn ResNet. Giải pháp ResNet đưa ra đơn giản hơn và tập trung vào cải thiện thông tin thông qua độ dốc của mạng. Sau ResNet hàng loạt biến thể của kiến trúc này được giới thiệu. Thực nghiệm cho thấy những kiến trúc này có thể được huấn luyện mạng nơ-ron với độ sâu hàng nghìn lớp và nó nhanh chóng trở thành kiến trúc phổ biến nhất trong thị giác máy tính (computer vision).

Trong bài báo này, các tác giả trình bày về trình bày về nghiên cứu, áp dụng được mạng ResNet cho hai bài toán phân loại cảm xúc và nhận diện cử chỉ tay. Trong đó, mô hình đã có thể sử dụng webcam để nhận diện cảm xúc theo thời gian thực.

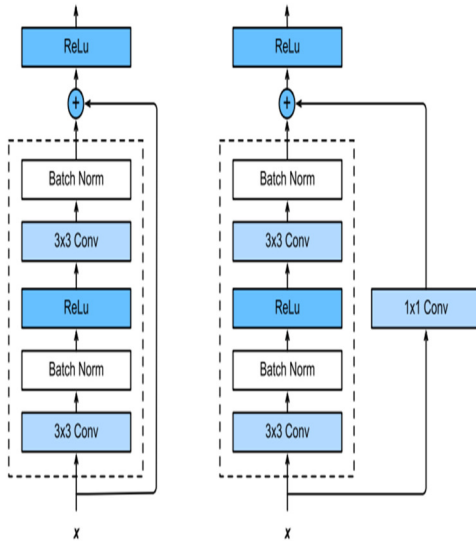
2. CƠ SỞ LÝ THUYẾT VÀ CÁC PHƯƠNG PHÁP THỰC HIỆN



Hình 1. Cấu trúc khối phần dư Resnet

Ký hiệu đầu vào là x . Giả sử ánh xạ lý tưởng muốn học được là $f(x)$, và được dùng làm đầu vào của hàm kích hoạt. Phần nằm trong viền nét đứt bên trái phải khớp trực tiếp với ánh xạ $f(x)$. Điều này có thể không đơn giản nếu chúng ta không cần khối đó và muốn giữ lại đầu vào x . Khi đó,

phần nằm trong viên nét đứt bên phải chỉ cần tham số hoá độ lệch khỏi giá trị x , bởi vì ta đã trả về $x + f(x)$. Trên thực tế, ánh xạ phần dư thường dễ tối ưu hơn, vì chỉ cần đặt $f(x) = 0$. Nửa bên phải mô tả khối phần dư cơ bản của ResNet. Về sau, những kiến trúc tương tự đã được đề xuất cho các mô hình chuỗi (sequence model), sẽ đề cập ở phần sau.



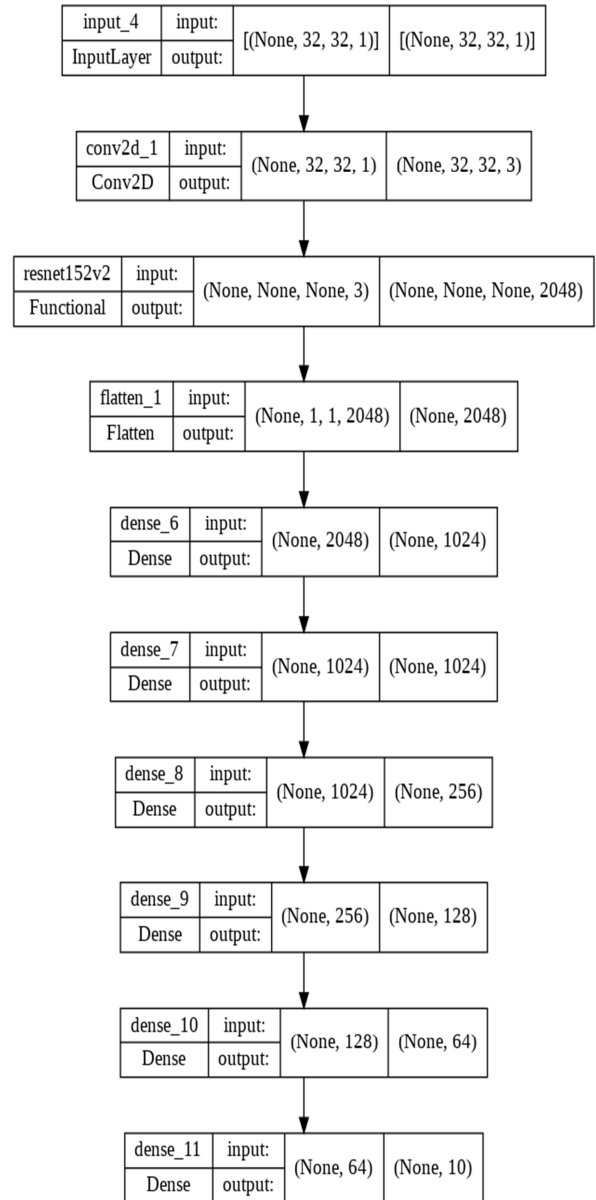
Hình 2. So sánh khi không có và có tầng tích chập 1×1



Hình 3. Mô hình mạng ResNet 152

Nếu muốn thay đổi số lượng kênh hoặc sai bước trong khối phần dư, cần thêm một tầng tích chập 1×1 để thay đổi kích thước đầu vào tương ứng ở nhánh ngoài.

Có thể thay đổi số kênh và các khối phần dư trong mô đun để tạo ra các mô hình ResNet khác nhau, ví dụ mô hình 152 tầng của ResNet152.



Hình 4. Cấu trúc mạng ResNet trong nhận diện cử chỉ tay

3. KẾT QUẢ ĐẠT ĐƯỢC

Bài toán nhận diện cảm xúc

Dữ liệu nghiên cứu được lấy từ Kaggle : FER2013

Fer2013 chứa khoảng 30.000 hình ảnh RGB trên khuôn mặt với các biểu cảm khác nhau với kích thước giới hạn ở 48×48 và các nhãn chính của nó có thể được chia thành 7 loại:

0 = Tức giận, 1 = Chán nản, 2 = Sợ hãi, 3 = Hạnh phúc, 4 = Buồn, 5 = Bất ngờ, 6 = Tự nhiên.

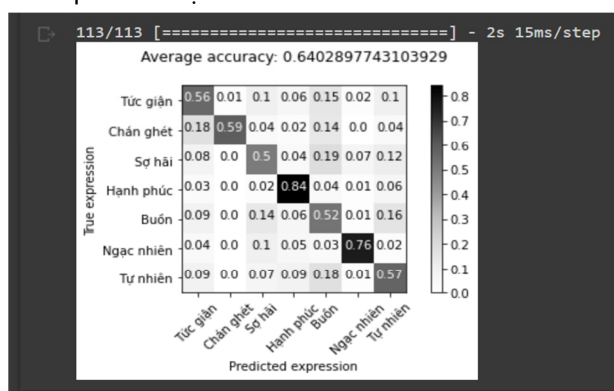
Trạng thái “chán nản” có số lượng hình ảnh tối ít nhất với 600 mẫu, trong khi các nhãn khác có gần 5.000 mẫu mỗi nhãn. Các hình ảnh trong thư mục train được chia thành tập huấn luyện và tập xác nhận.

Các hình ảnh trong thư mục train được chia thành tập huấn luyện và tập xác nhận:

emotion		pixels	Usage
0	0	70 80 82 72 58 58 60 63 54 58 60 48 89 115 121...	Training
1	0	151 150 147 155 148 133 111 140 170 174 182 15...	Training
2	2	231 212 156 164 174 138 161 173 182 200 106 38...	Training
3	4	24 32 36 30 32 23 19 20 30 41 21 22 32 34 21 1...	Training
4	6	4 0 0 0 0 0 0 0 0 0 3 15 23 28 48 50 58 84...	Training

Hình 5. Nguồn dữ liệu được sử dụng

Kết quả thu được:



Hình 6. Kết quả nhận dạng biểu cảm theo tỷ lệ

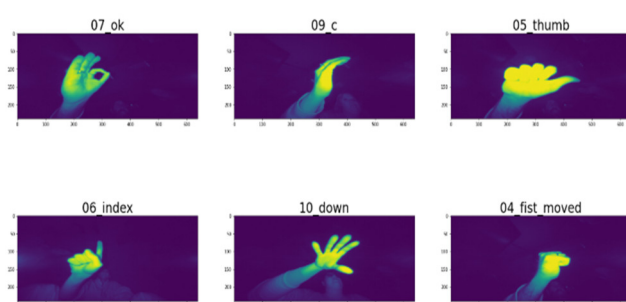
Qua các kết quả mô phỏng số, có thể thấy rằng:

- Độ chính xác trung bình khoảng 64% đặc biệt với 2 cảm xúc hạnh phúc và ngạc nhiên với độ chính xác cao.
- Có sự cải thiện về độ chính xác so với những mạng CNN trước đây, có thể ứng dụng trong thực tế.

Bài toán nhận diện cử chỉ tay

Dữ liệu nghiên cứu được lấy từ Kaggle (<https://www.kaggle.com/datasets/gti-upm/leapgestrecog>).

Hình ảnh cử chỉ tay



Hình 7. Dữ liệu cử chỉ tay

Cơ sở dữ liệu bao gồm 10 cử chỉ tay khác nhau được thực hiện bởi 10 đối tượng khác nhau (5 nam và 5 nữ). Mỗi thư mục gốc đều chứa các ảnh hồng ngoại của một đối tượng. Tên thư mục là định danh của từng chủ đề khác nhau.

Đánh giá mô hình sau khi huấn luyện như hình 8.

Dựa vào bảng kết quả, thấy được: Mô hình có sự chính xác cao và hoạt động tốt với ảnh tĩnh.

	precision	recall	f1-score	support
01_palm	1.00	1.00	1.00	204
02_l	1.00	1.00	1.00	215
03_fist	1.00	1.00	1.00	217
04_fist_moved	1.00	1.00	1.00	182
05_thumb	1.00	1.00	1.00	207
06_index	1.00	1.00	1.00	198
07_ok	0.53	1.00	0.69	201
08_palm_moved	1.00	1.00	1.00	204
09_c	1.00	1.00	1.00	194
10_down	0.00	0.00	0.00	178
accuracy			0.91	2000
macro avg	0.85	0.90	0.87	2000
weighted avg	0.86	0.91	0.88	2000

Hình 8. Đánh giá model nhận diện cử chỉ tay

4. KẾT LUẬN

Bài báo trình bày nghiên cứu, áp dụng được mạng Resnet cho hai bài toán phân loại cảm xúc và nhận diện cử chỉ tay. Trong đó, mô hình đã có thể sử dụng webcam để nhận diện cảm xúc theo thời gian thực.

Có thể mở rộng bài toán phân loại cảm xúc và áp dụng cho hệ thống cảnh báo lái xe an toàn, giao diện người - máy,...; mở rộng bài toán nhận diện cử chỉ tay và áp dụng cho hệ thống điều khiển từ xa, hệ thống hỗ trợ người khiếm thính,...

TÀI LIỆU THAM KHẢO

- [1]. Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam. *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications* (arXiv:1704.04861).
- [2]. Asifullah Khan, Anabia Sohail, Umme Zahoor, Aqsa Saeed Qureshi, 2020. *A Survey of the Recent Architectures of Deep Convolutional Neural Networks*. Artificial Intelligence Review, 53.
- [3]. Aston Zhang, Zachary C. Lipton, Mu Li, Alexander J. Smola, 2021. *Dive into Deep Learning: Dive into Deep Learning Release 0.16.6*. <https://arxiv.org/abs/2106.11342>
- [4]. *Deep to Learning*. <https://d2l.aivivn.com>.
- [5]. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, 2015. *Deep Residual Learning for Image Recognition*. <https://arxiv.org/pdf/1512.03385.pdf>.
- [6]. Nguyễn Tuấn, 2019. *Transfer learning và data augmentation*. Ngày truy cập 01/2022. <https://nttuan8.com/bai-9-transfer-learning-va-data-augmentation>.
- [7]. Vũ Hữu Tiệp, 2018. *Machine Learning cơ bản*. Ngày tháng truy cập 02/2022. <https://machinelearningcoban.com>.
- [8]. *Transfer Learning và bài toán Face Recognition*. <https://viblo.asia/p/transfer-learning-va-bai-toan-face-recognition-3Q75w7xD5Wb>.
- [9]. Zhang X, Zhou X, Lin M, Sun J, 2018. *ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices*. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.