

Bài báo nghiên cứu

PHÁT HIỆN KHẨU TRANG KHUÔN MẶT BẰNG PHƯƠNG PHÁP HỌC SÂU

Đàm Minh Linh^{1*}, Ngô Xuân Thoai¹, Hàn Minh Châu²

¹Học viện Công nghệ Bưu chính Viễn thông Cơ sở tại Thành phố Hồ Chí Minh, Việt Nam

²Trường Đại học Công nghệ Thành phố Hồ Chí Minh, Việt Nam

*Tác giả liên hệ: Đàm Minh Linh – Email: linhdm@ptit.edu.vn

Ngày nhận bài: 31-8-2023; ngày nhận bài sửa: 26-10-2023; ngày duyệt đăng: 30-10-2023

TÓM TẮT

Ô nhiễm môi trường và bệnh về đường hô hấp như là đại dịch COVID-19, có khả năng lây nhiễm qua không khí và ảnh hưởng đến sức khỏe của con người. Để bảo vệ sự an toàn cho bản thân và trong cộng đồng thì một trong các giải pháp được đề xuất là đeo khẩu trang. Vì thế, bài nghiên cứu này tập trung vào sự phát hiện khuôn mặt người có đeo khẩu trang hoặc không đeo khẩu trang từ dữ liệu của camera giám sát, dữ liệu video thu thập được kết hợp một thuật toán học sâu Convolutional Neural Network (CNN), học máy sẽ phân loại dữ liệu ra hai nhãn. Kết quả thực nghiệm đã đóng góp được hai vấn đề chính: một là phát hiện đeo khẩu trang và không đeo khẩu trang, hai là: đề xuất hai mô hình học sâu CNN được đánh giá và được so sánh cho sự hiệu quả cho các công thức: Accuracy, Precision, Recall và F1-Score với kết quả đã đạt được độ chính xác là 99.94%.

Từ khóa: mạng nơ ron (CNN); Deep learning; Face Mask; OpenCV; xử lý ảnh

1. Giới thiệu

Nhận dạng khẩu trang trên khuôn mặt người bằng phương pháp học sâu được triển khai cụ thể, các camera giám sát sẽ thu thập dữ liệu về khuôn mặt người, sau đó sẽ gửi dữ liệu thu thập được về máy chủ học máy, trên máy chủ sẽ so sánh giữa ảnh hoặc video gửi đến với chiết xuất đặc trưng từ bộ dữ liệu ảnh đã huấn luyện, cuối cùng là phân loại đưa ra kết quả có đeo khẩu trang hoặc không đeo khẩu trang.

Lĩnh vực thị giác máy tính đã ứng dụng các thuật toán học máy, học sâu vào nhận dạng khuôn mặt đeo khẩu trang, trong đó (Wang et al., 2020) đã sử dụng ba bộ dữ liệu khác nhau như là MFDD, RMFRD và SMFRD để đánh giá quá trình huấn luyện dữ liệu, đạt độ chính xác 95%. Ở nghiên cứu (Yadav et al., 2022) đã sử dụng mô hình CNN để nhận dạng mẫu dữ liệu đeo khẩu trang và không đeo khẩu trang với độ chính xác 97.25%. Để giảm sự lây nhiễm đại dịch COVID-19 (Li, 2020) đã đề xuất nhận dạng đeo khẩu trang theo thời gian thực có

Cite this article as: Dam Minh Linh, Ngo Xuan Thoai, & Han Minh Chau (2023). Detecting face mask using a deep learning method. *Ho Chi Minh City University of Education Journal of Science*, 20(11), 1931-1942.

thể áp dụng ở nơi công cộng, sử dụng công nghệ internet vạn vật (IoT), độ chính xác đạt 98.7%. Một phương pháp nhận dạng và phát hiện khuôn mặt sử dụng thư viện Open Computer Vision (OpenCV) bởi (Khan et al., 2019). Hệ thống hỗ trợ bằng công nghệ học sâu dùng phân loại khẩu trang chống lại COVID-19 nhằm bảo vệ an toàn cho con người được sử dụng kiến trúc YOLOv3 (Bhuiyan et al., 2020).

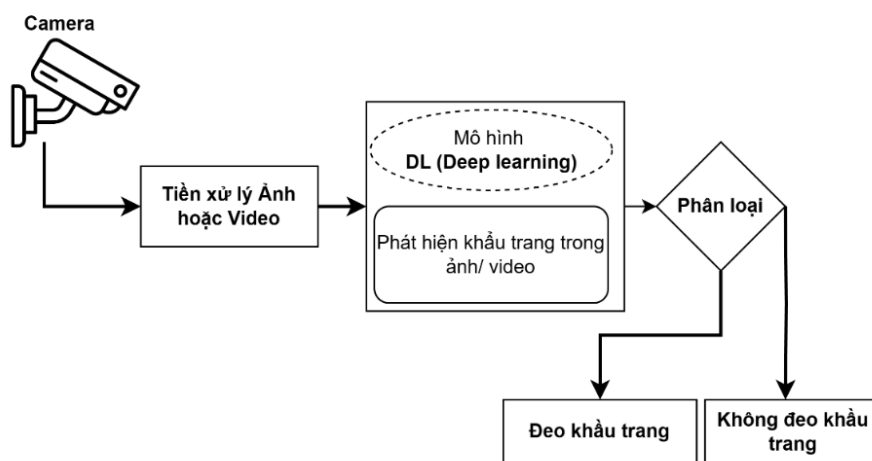
Trong nghiên cứu (Đam & Nguyen, 2023) sử dụng các thuật toán LDA, NB, KNN, DT, SVM và CNN đánh giá bộ dữ liệu hình thái khuôn mặt với mô hình đề xuất đạt độ chính xác 97.57%. Khảo sát thống kê của (Kore et al., 2023a) cho các nghiên cứu về các hệ thống để phân loại tính năng phát hiện khẩu trang cho dữ liệu cả hình ảnh và video bằng nhiều phương pháp khác nhau như Deep learning, machine learning, mobile Net, Res Net, YOLO, Google Net. Đề cập ở nghiên cứu này là liệt kê của các công trình nghiên cứu trước đó đã sử dụng nhiều phương pháp khác nhau cho hệ thống phát hiện khẩu trang bằng công nghệ máy học đã được phát triển.

Bài viết này gồm có 4 phần, các phần còn lại được trình bày như sau: Phần 2 là trình bày về đối tượng và phương pháp nghiên cứu: Đề xuất mô hình phát hiện khẩu trang, mô tả tập dữ liệu, phương pháp đánh giá, đề xuất 2 mô hình học sâu CNN; phần 3 là kết quả thực nghiệm và thảo luận của nghiên cứu; cuối cùng là phần 4 kết luận cho nghiên cứu này.

2. Đối tượng và phương pháp nghiên cứu

2.1. Đề xuất mô hình phát hiện khẩu trang

Ở Hình 1, phương pháp phát hiện khẩu trang bằng camera giám sát, dữ liệu sẽ được thu thập dưới dạng video hoặc ảnh, sau đó được chuẩn hóa dữ liệu được lưu trữ dạng mảng, kích thước ảnh..., tiếp theo là đưa dữ liệu vào mô hình dự đoán, so sánh chiết suất đặc trưng với bộ dữ liệu đã huấn luyện trước đó, cuối cùng là phân loại nhân, đưa ra kết quả.

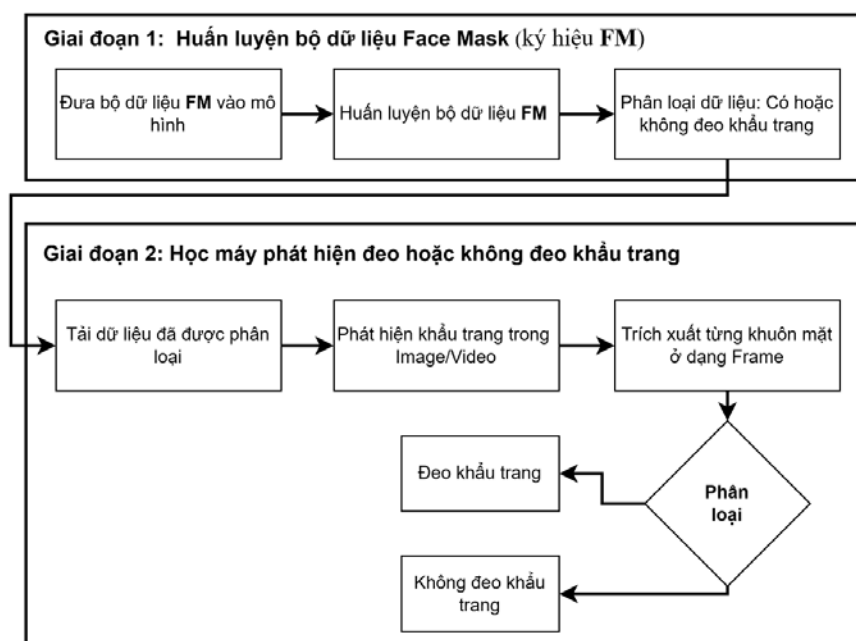


Hình 1. Phương pháp phát hiện khẩu trang

Trong Hình 2, cấu trúc của hệ thống học máy phát hiện khẩu trang đáp ứng thời gian thực được kết hợp với mô hình học máy, được phân chia làm hai giai đoạn:

Giai đoạn 1: Huấn luyện bộ dữ liệu Face Mask, tiếp theo là thực hiện phân loại nhãn dữ liệu sau khi huấn luyện.

Giai đoạn 2: Học máy phát hiện đeo khẩu trang hoặc không đeo khẩu trang, từ dữ liệu đã được phân loại nhãn sẽ được so sánh chiết suất đặc trưng với dữ liệu video từ camera, sau đó trích suất từng khuôn mặt ở dạng khung (frame), phân loại kết quả khuôn mặt có đeo khẩu trang hay không đeo khẩu trang.



Hình 2. Cấu trúc của hệ thống học máy, học sâu

2.2. Tập dữ liệu

Để đánh giá được nhận dạng khẩu trang với độ chính xác cao, cần có bộ dữ liệu chuẩn và tỉ lệ phân chia mẫu dữ liệu trong tập dữ liệu huấn luyện tương đối gần bằng nhau, số liệu được mô tả trong Bảng 1, được tham khảo nguồn từ (Jessica Li, 2021; MinhLinhEdu, 2023; Wang et al., 2020).

Bảng 1. Bộ dữ liệu Face Mask

Bộ dữ liệu có 3 loại	Mỗi loại dữ liệu gồm có 2 mẫu	
	Không đeo khẩu trang (WithoutMask)	Đeo khẩu trang (WithMask)
Dữ liệu train	5000	5000
Dữ liệu test	509	483
Dữ liệu validation	400	400
Tổng số	5909	5883

Bộ dữ liệu có tổng 11.792 ảnh, kích thước ảnh 256 * 256 pixel, có 3 loại dữ liệu: Dữ liệu huấn luyện (train) có 5000 mẫu ảnh đeo khẩu trang và 5000 mẫu ảnh không đeo khẩu trang, tương tự cho dữ liệu kiểm tra sau khi huấn luyện (test) và dữ liệu kiểm tra trong quá

trình huấn luyện (validation), mỗi loại có 2 mẫu dữ liệu là đeo khẩu trang và không đeo khẩu trang, được trình bày trong Bảng 1.



Hình 3. Trích xuất từ bộ dữ liệu mẫu đeo khẩu trang và không đeo khẩu trang

Ở Hình 3, mẫu dữ liệu ảnh đeo khẩu trang và không đeo khẩu trang được trích xuất từ bộ dữ liệu ảnh.

2.3. Phương pháp đánh giá

Để đánh giá hiệu quả chính xác các mô hình thực nghiệm thì sử dụng phương pháp đo bằng toán học cụ thể các công thức Accuracy (viết tắt Ac), Precision (viết tắt Pr) và F1-score (viết tắt F1S), Recall hoặc Detection Rate (viết tắt DR). Ở công thức 1, sử dụng ma trận nhầm lẫn các thuộc tính dương tính thật (TP), âm tính thật (TN), dương tính giả (FP) và âm tính giả (FN), được tham khảo (Nguyen & Ha, 2023; Huynh & Nguyen, 2022)

$$\text{Confusion Matrix} = \begin{bmatrix} TN & FP \\ FN & TP \end{bmatrix} \quad (1)$$

Ac là tiêu chí đánh giá độ chính xác trong quá trình học huấn luyện và kiểm tra dữ liệu cho mỗi vòng (epoch), ở công thức 2.

$$Ac = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

Pr là kết quả dự đoán giữa dữ liệu được kiểm tra trong quá trình huấn luyện, nếu giá trị gần bằng nhau thì kết quả dự đoán tốt, ở công thức 3.

$$Pr = \frac{TP}{TP + FP} \quad (3)$$

Công thức 4, F1S là trung bình của hai tiêu chí Pr và DR. F1S phù hợp hơn để đánh giá tỉ lệ nhận dạng các mẫu dữ liệu không cân bằng.

$$F1S = \frac{2 * Precision * DR}{Precision + DR} \quad (4)$$

Tỉ lệ phát hiện (Recall hoặc DR), giá trị DR càng gần với 1 thì kết quả cho một nhận dạng chính xác cao, ở công thức 5.

$$DR = \frac{TP}{TP + FN} \quad (5)$$

2.4. Đề xuất hai mô hình học sâu CNN

Mô hình học sâu (deep learning) thứ một: Được mô tả ở Hình 4, quá trình huấn luyện (train) và kiểm tra (test) bộ dữ liệu cho các tham số đầu vào ảnh thuộc tính **conv2d** = kích thước ảnh 256 * 256 pixel, đầu ra **dense_1** = 2 loại nhãn khác nhau, tổng số tham số được huấn luyện là 17.166.402, số vòng train epochs = 16, thời gian trung bình cho mỗi epoch =

25s, tuy nhiên thời gian thực cho mỗi epoch cao hơn từ 1s đến 3s. Mô hình này được huấn luyện và kiểm tra với ba kích thước batch_size khác nhau {32, 64, 128}.

```
model_32 = tf.keras.Sequential([
    tf.keras.layers.Rescaling(1./255),
    tf.keras.layers.Conv2D(32, 3, activation='relu'),
    tf.keras.layers.MaxPooling2D(pool_size=(2,2)),
    tf.keras.layers.Conv2D(64, 3, activation='relu'),
    tf.keras.layers.MaxPooling2D(pool_size=(2,2)),
    tf.keras.layers.Conv2D(128, 3, activation='relu'),
    tf.keras.layers.MaxPooling2D(pool_size=(2,2)),
    tf.keras.layers.Conv2D(256, 3, activation='relu'),
    tf.keras.layers.MaxPooling2D(pool_size=(2,2)),
    tf.keras.layers.Flatten(),
    tf.keras.layers.Dense(256, activation='relu'),
```

Hình 4. Mô hình CNN thứ một được thực nghiệm cho ba kích thước batch_size {32, 64, 128}

```
model = tf.keras.Sequential([
    tf.keras.layers.Rescaling(1./255),
    tf.keras.layers.Conv2D(64, 3, activation='relu'),
    tf.keras.layers.MaxPooling2D(pool_size=(2,2)),
    tf.keras.layers.Conv2D(128, 3, activation='relu'),
    tf.keras.layers.MaxPooling2D(pool_size=(2,2)),
    tf.keras.layers.Conv2D(256, 3, activation='relu'),
    tf.keras.layers.MaxPooling2D(pool_size=(2,2)),
    tf.keras.layers.Conv2D(512, 3, activation='relu'),
    tf.keras.layers.MaxPooling2D(pool_size=(2,2)),
    tf.keras.layers.Flatten(),
    tf.keras.layers.Dense(512, activation='relu'),
```

Hình 5. Mô hình CNN thứ 2 được thực nghiệm với kích thước batch_size = 128

Mô hình học sâu (deep learning) thứ hai: Ở Hình 5, các tham số đầu vào ảnh thuộc tính **conv2d** = kích thước ảnh 256 * 256 pixel, đầu ra **dense_1** = 2 loại nhãn khác nhau, tổng số tham số được huấn luyện là 68.661.378, số vòng train epochs = 16, thời gian trung bình cho mỗi epoch = 40s. Mô hình này được huấn luyện và kiểm tra ứng với kích thước batch_size = 128.

Các tham số ở các lớp cho mô hình CNN thứ hai được đề xuất gấp đôi tham số ở các lớp cho mô hình thứ nhất lần lượt là lớp tích chập hai chiều **conv2d** = {64, 32} và sau đó tăng dần, tương ứng cho các lớp còn lại. Nhược điểm của mô hình CNN thứ hai là tạo ra một ma trận trọng số tích chập có kích thước lớn, điều này dẫn đến tiêu tốn tài nguyên phần cứng nhiều hơn khi huấn luyện bộ dữ liệu và trong quá trình huấn luyện bộ dữ liệu sẽ bị trùng lặp thông tin dữ liệu cũ (nghĩa là sẽ huấn luyện lại dữ liệu trước đó). Bằng chứng là theo số liệu ở trên, kết quả tổng số tham số đã được huấn luyện bộ dữ liệu và thời gian huấn luyện trung bình cho mỗi epoch của mô hình CNN thứ hai lớn hơn rất nhiều so với mô hình CNN thứ một.

3. Kết quả và thảo luận

Trong bài thực nghiệm này, được đánh giá bằng máy server: Windows 10 Pro 20H2, CPU (2) 2.30GHz E5 2696 v3 (36 nhân, 72 luồng), RAM 64Gb DDR4, NVIDIA® GeForce RTX™ 3060 GPU 12Gb.

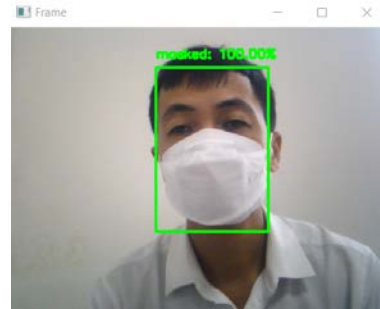
3.1. Kết quả thực nghiệm

Quá trình thực nghiệm, nghiên cứu đã sử dụng các thư viện như OpenCV hỗ trợ lĩnh vực thị giác máy tính về nhận dạng, phát hiện khuôn mặt, tìm kiếm ảnh; thư viện Keras hỗ trợ thuật toán học sâu CNN, xử lý dạng bài toán về dạng hình ảnh và TensorFlow hiệu quả cao tính toán linh hoạt nền tảng GPU hỗ trợ mạnh thuật toán học máy và học sâu.

Ở Hình 6, camera giám sát online nhận diện khuôn mặt không đeo khẩu trang, trong khi ở Hình 7 kết quả nhận diện khuôn mặt có đeo khẩu trang tỉ lệ chính xác 100%.

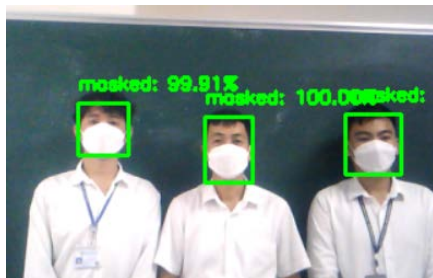


Hình 6. Kết quả không đeo khẩu trang



Hình 7. Kết quả nhận diện đeo khẩu trang

Tương tự ở Hình 8, camera giám sát online nhận diện tất cả ba khuôn mặt đeo khẩu trang, trong khi ở Hình 9 kết quả nhận diện khuôn mặt có đeo khẩu trang là một khuôn mặt và hai khuôn mặt không đeo khẩu trang, kết quả độ chính xác đạt 100%.



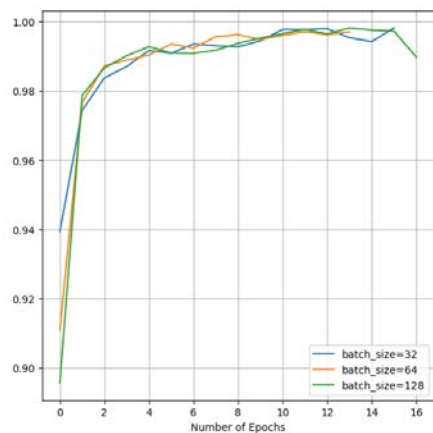
Hình 8. Kết quả nhận diện đeo khẩu trang



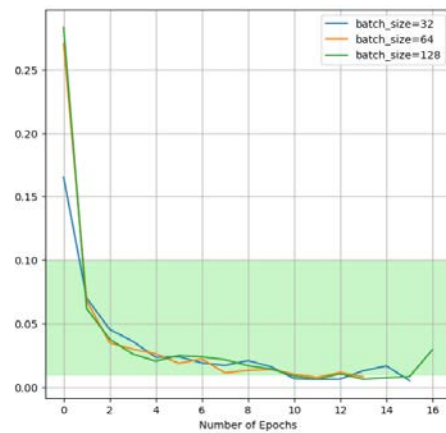
Hình 9. Kết quả nhận diện đeo khẩu trang

Ở biểu đồ Hình 10 tương ứng cho số liệu ở Bảng 2, sự so sánh độ chính xác (Accuracy) của dữ liệu huấn luyện của thuật toán CNN với nhiều batch size = {32, 64, 128}. Nhìn chung, ở vòng epoch = 1 thì độ chính xác của dữ liệu được huấn luyện với batch size = 32 đạt cao nhất là 93.68%, trong khi độ chính xác của hai batch size còn lại lần lượt là 91.7699%, 86.89%. Ở vòng epoch = 10, độ chính xác của các batch size dao động trong khoảng 1%. Ở vòng epoch = 16, độ chính xác của batch size = 32 là 99.97%, độ chính xác lần lượt các batch size còn lại 98.97% và 99.5%.

Hàm mất mát (loss) ở Hình 11, theo biểu đồ dữ liệu của các batch size càng tiệm cận với 0 thì mô hình huấn luyện dữ liệu đạt độ chính xác càng cao.



Hình 10. Sự so sánh độ chính xác



Hình 11. Sự so sánh hàm mất mát (loss)

Ở Bảng 2, mô tả về độ chính xác của mô hình học sâu CNN với các kích thước batch size {32, 64, 128} khác nhau, số vòng epoch = 16.

Bảng 2. Độ chính xác (accuracy) với kích thước batch size {32, 64, 128}

Số vòng (Epochs)	Số lượng mẫu dữ liệu (Batch size)		
	32	64	128
1	0.936800003	0.917699993	0.868900001
2	0.980400026	0.980799973	0.976300001
3	0.989700019	0.986400008	0.981899977
4	0.987699986	0.987200022	0.986599982
5	0.9903	0.989300013	0.989799976
6	0.9921	0.992999971	0.993399978
7	0.99150002	0.995599985	0.995800018
8	0.989799976	0.989700019	0.995000005
9	0.995500028	0.994099975	0.995700002
10	0.996699989	0.996299982	0.997500002
11	0.994799972	0.99119997	0.994400024
12	0.996399999	0.993700027	0.995199978
13	0.996900022	0.995599985	0.994899988
14	0.997300029	0.998300016	0.996399999
15	0.997699976	0.994199991	0.995999992
16	0.99970001	0.996399999	0.998199999

Ở Hình 12, với batch size = 32, bộ dữ liệu sau khi huấn luyện (train) và kiểm tra (test) kết quả thống kê độ chính xác của dữ liệu train đạt 99.94% và dữ liệu test là 98.99%. Kết quả cho thấy rất hiệu quả của mô hình này.

```
In [18]: e32 = model_32.evaluate(train_ds)
print(f'Train set accuracy: "{0:.2f}".format(e32[1]*100)}%')

313/313 [=====] - 12s 37ms/step - loss: 0.0031 - acc: 0.9994
call_m: 1.0004
Train set accuracy: 99.94%

In [19]: e32 = model_32.evaluate(test_ds)
print(f'Test set accuracy: "{0:.2f}".format(e32[1]*100)}%')

31/31 [=====] - 2s 65ms/step - loss: 0.0221 - acc: 0.9899 - f1_m: 0.5171
Test set accuracy: 98.99%
```

Hình 12. Độ chính xác của dữ liệu huấn luyện và dữ liệu kiểm tra với batch size = 32

Ở Hình 13, với batch size = 64, bộ dữ liệu sau khi huấn luyện (train) và kiểm tra (test) kết quả thống kê độ chính xác của dữ liệu train đạt 99.81% và dữ liệu test là 98.99%.

```
In [29]: e64 = model_64.evaluate(train_ds)
print(f'Train set accuracy: "{0:.2f}".format(e64[1]*100)}%')

157/157 [=====] - 10s 63ms/step - loss: 0.0052 - acc: 0.9981 - f1_m: 0.9981
Train set accuracy: 99.81%

In [31]: e64 = model_64.evaluate(test_ds)
print(f'Test set accuracy: "{0:.2f}".format(e64[1]*100)}%')

16/16 [=====] - 2s 95ms/step - loss: 0.0435 - acc: 0.9899 - f1_m: 0.5594
Test set accuracy: 98.99%
```

Hình 13. Độ chính xác của dữ liệu huấn luyện và dữ liệu kiểm tra với batch size = 64

Ở Hình 14, với batch size = 128, bộ dữ liệu sau khi huấn luyện (train) và kiểm tra (test) kết quả thống kê độ chính xác của dữ liệu train đạt 99.83% và dữ liệu test là 99.4%.

```
In [41]: e128 = model_128.evaluate(train_ds)
print(f'Train set accuracy: "{0:.2f}".format(e128[1]*100)}%')

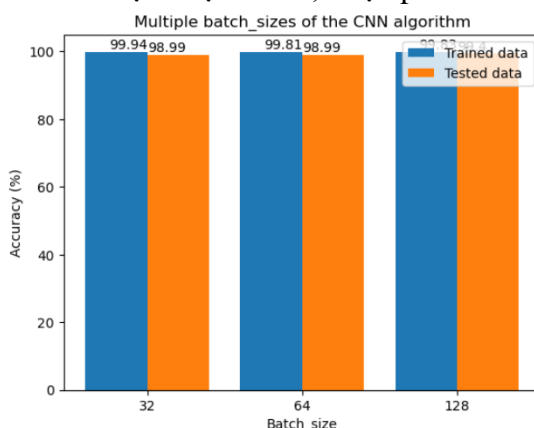
79/79 [=====] - 9s 110ms/step - loss: 0.0043 - acc: 0.9983 - f1_m
l1_m: 1.0000
Train set accuracy: 99.83%

In [42]: e128 = model_128.evaluate(test_ds)
print(f'Test set accuracy: "{0:.2f}".format(e128[1]*100)}%')

8/8 [=====] - 2s 291ms/step - loss: 0.0122 - acc: 0.9940 - f1_m: 0.
_m: 0.6250
Test set accuracy: 99.40%
```

Hình 14. Độ chính xác của dữ liệu huấn luyện và dữ liệu kiểm tra với batch size = 128

Với các kích thước với batch size {32, 64, 128} ở biểu đồ cột Hình 15, là sự so sánh độ chính xác của train và test dữ liệu được mô tả, được phân tích ở các Hình 12, 13, 14.



Hình 15. Độ chính xác của dữ liệu huấn luyện và dữ liệu kiểm tra với batch size {32, 64, 128}

Ở mô hình đề xuất thứ 2 thuật toán học sâu, thực nghiệm với kích thước batch size = 128, bộ dữ liệu sau khi huấn luyện (train) và kiểm tra (test) kết quả thống kê thì độ chính xác của dữ liệu train đạt 98.99% và dữ liệu test là 98.29%.

```
In [70]: e_train = model.evaluate(train_ds)
print(f'Train set accuracy: "{0:.2f}".format(e_train[1]*100)}%')

79/79 [=====] - 15s 188ms/step - loss: 0.0272 - acc: 0.9899
all_m: 1.0036
Train set accuracy: 98.99%

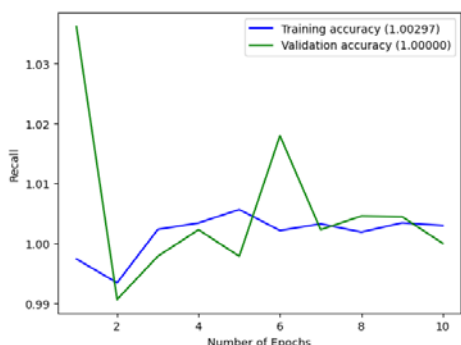
In [71]: e_test = model.evaluate(test_ds)
print(f'Test set accuracy: "{0:.2f}".format(e_test[1]*100)}%')

8/8 [=====] - 2s 252ms/step - loss: 0.0509 - acc: 0.9829 -
_m: 0.6326
Test set accuracy: 98.29%
```

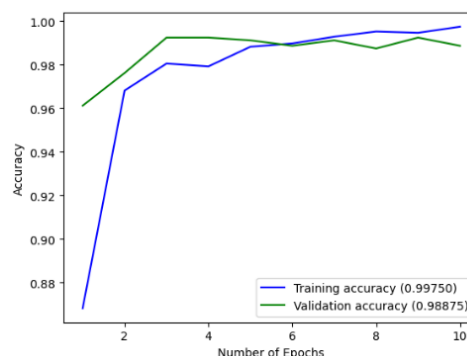
Hình 16. Độ chính xác của dữ liệu huấn luyện và dữ liệu kiểm tra với batch size {128}

Tỉ lệ phát hiện (Recall hoặc DR), giá trị DR càng gần với 1 thì kết quả cho một nhận dạng chính xác cao, kết quả ở Hình 17 cho thấy độ chính xác của dữ liệu kiểm tra (validation) trong quá trình huấn luyện tương ứng là 100%, 100.297%. Điều này chứng minh rằng mô hình học sâu đạt kết quả độ chính xác cao.

Độ chính xác ở Hình 18, là kết quả cho quá trình huấn luyện và kiểm tra dữ liệu lần lượt đạt 99.75% và 98.875%. Mỗi liên quan giữa học và kiểm tra dữ liệu là rất tốt.



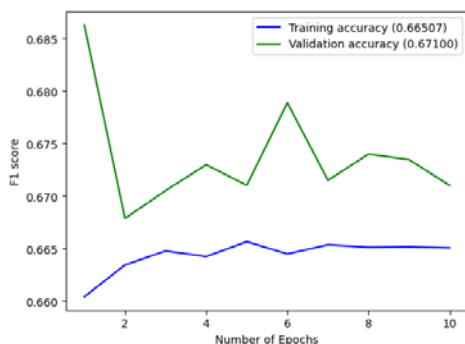
Hình 17. Tỷ lệ phát hiện



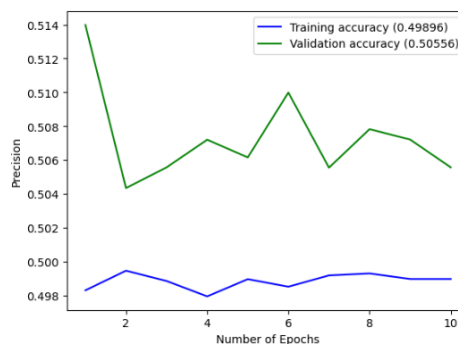
Hình 18. Độ chính xác

F1S là trung bình của Pr và DR. Ở Hình 19, tỉ lệ nhận dạng các mẫu dữ liệu cân bằng với nhau, kết quả lần lượt của dữ liệu huấn luyện và kiểm tra là 66.507%, 67.1%.

Kết quả dự đoán ở Hình 20, sự cân bằng giữa hai kết quả huấn luyện dữ liệu và kiểm tra dữ liệu đạt tỉ lệ gần bằng nhau.



Hình 19. Đánh giá cân bằng dữ liệu



Hình 20. Kết quả dự đoán

Theo bảng số liệu ở Bảng 3, là kết quả so sánh các công trình nghiên cứu về lĩnh vực thị giác máy tính đã ứng dụng học máy, học sâu vào nhận dạng khuôn mặt, khẩu trang. Kết quả mô hình đề xuất của nghiên cứu này đạt hiệu quả tốt hơn.

Bảng 3. Khảo sát kết quả công trình nghiên cứu các thuật toán liên quan

Tác giả	Accuracy %
(Wang et al., 2020)	95
(Yadav et al., 2022)	97.25
(Li, 2020)	98.7
(Kore et al., 2023b)	98.86
(Saiful Azian et al., 2022)	98.96 - 99.55
(Mare et al., 2021)	99.44
Mô hình đề xuất	99.94

3.2 Thảo luận

Trong nghiên cứu này, kết quả nhận dạng được khuôn mặt đeo khẩu trang và không đeo khẩu trang đáp ứng thời gian thực, phương pháp phát hiện đã ứng dụng công nghệ thuật toán học sâu (deep learning) đạt được kết quả chính xác cao.

Thực nghiệm đề xuất mô hình học sâu CNN thứ một, quá trình huấn luyện và kiểm tra bộ dữ liệu Face Mask với tổng số lượng ảnh 11.792, trong đó 5883 mẫu ảnh đeo khẩu trang và 5909 mẫu ảnh không đeo khẩu trang, kích thước 256 * 256 pixel, đầu ra dense_1 = 2 loại nhãn khác nhau, tổng số tham số được train là 17.166.402, số vòng train epochs = 16, thời gian trung bình cho mỗi epoch = 25s với ba kích thước batch size khác nhau {32, 64, 128}, kết quả độ chính xác của dữ liệu train lần lượt là 99.94%, 99.81%, 99.83% và dữ liệu test lần lượt là 98.99%, 98.99%, 99.4%. Kết quả kích thước batch size = 32 là hiệu quả nhất. Ở mô hình học sâu CNN thứ hai, tổng số tham số được train là 68.661.378, số vòng train epochs = 16, thời gian trung bình cho mỗi epoch = 40s, kích thước batch_size = 128. Kết quả độ chính xác của dữ liệu train đạt 98.99% và dữ liệu test là 98.29%. So sánh giữa hai mô hình học sâu CNN có cùng batch size = 128 thì mô hình thứ một đạt hiệu quả cao hơn cho dữ liệu train là 0.84% và dữ liệu test là 1.11%.

Qua khảo sát kết quả các công trình nghiên cứu, kết quả độ chính xác được đánh giá (Wang et al., 2020; Yadav et al., 2022; Li, 2020; Kore et al., 2023b; Saiful Azian et al., 2022; Mare et al., 2021) và mô hình đề xuất lần lượt là 95%, 97.25%, 98.7%, 98.86%, 98.96% - 99.55%, 99.44% và **99.94%**. Kết quả mô hình đề xuất tỉ lệ học chính xác, hiệu quả cao nhất.

4. Kết luận

Kết quả nghiên cứu này, nhóm chúng tôi đã nêu ra hai vấn đề chính: một là, từ camera giám sát trực tiếp sẽ phát hiện khuôn mặt người có đeo khẩu trang hoặc không đeo khẩu trang kết hợp thuật toán học sâu Convolutional Neural Network (CNN) đã được huấn luyện bộ dữ liệu Face Mask với tổng số lượng ảnh 11.792, trong đó 5883 mẫu ảnh đeo khẩu trang và 5909 mẫu ảnh không đeo khẩu trang, kích thước 256 * 256 pixel, kết quả đã phân loại chính xác; hai là, đề xuất hai mô hình học sâu CNN được đánh giá, so sánh sự hiệu quả bằng các tiêu chí Accuracy, Precision, Recall và F1-Score với kết quả huấn luyện bộ dữ liệu đạt độ chính xác là 99.94% cho mô hình học sâu (deep learning) CNN thứ một tương ứng với batch size = 32. Với kết quả này, đã ứng dụng được cho việc phát hiện khuôn mặt người có đeo khẩu trang hoặc không đeo khẩu trang đáp ứng thời gian thực.

❖ **Tuyên bố về quyền lợi:** Các tác giả xác nhận hoàn toàn không có xung đột về quyền lợi.

TÀI LIỆU THAM KHẢO

- Bhuiyan, M. R., Khushbu, S. A., & Islam, M. S. (2020). *A Deep Learning Based Assistive System to Classify COVID-19 Face Mask for Human Safety with YOLOv3*. In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (pp. 1-5). Kharagpur, India. <http://doi.org/10.1109/ICCCNT49239.2020.9225384>
- Dam, L. M., & Nguyen, T. H. (2023). An Approach to Human Face Recognition by Machine Learning Training. *Ho Chi Minh City University of Education Journal of Science*, 20(1), 165-179.
- Huynh, T. T., & Nguyen, T. H. (2022). On The Performance of Intrusion Detection Systems with Hidden Multilayer Neural Network using DSD Training. *International Journal of Computer Networks & Communications (IJCNC)*, 14(1), 117-137.
- Jessica Li. (2021, 6). *Kaggle*. (Product Manager at Kaggle) Retrieved from <https://www.kaggle.com/jessicali9530>
- Khan, M., Chakraborty, S., Astya, R., & Khepra, S. (2019). *Face Detection and Recognition Using OpenCV*. In *2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)* (pp. 116-119). Greater Noida, India. <http://doi.org/10.1109/ICCCIS48478.2019.8974493>
- Kore, S., Pise, P., Thakare, P., & Morey, J. (2023a). A Survey on Face Mask Detection. *Tijer - International Research Journal*, 10(2), 780-784.
- Kore, S., Pise, P., Thakare, P., & Morey, J. (2023b). Real Time Face Mask Detection System. *Journal For Basic Sciences*, 23(6), 452-461.
- Li, B. Q. (2020). Identifying Facemask-Wearing Condition Using Image Super-Resolution with Classification Network to Prevent COVID-19. *Sensors. MDPI*, 1-23.
- Mare, T., Duta, G., Georgescu, M.-I., Sandru, A., Alexe, B., Popescu, M., & Ionescu, R. T. (2021). *A realistic approach to generate masked faces applied on two novel masked face recognition data sets* (arXiv e-Print 2109.01745). <https://doi.org/10.48550/arXiv.2109.01745>
- MinhLinhEdu. (2023). *Face-Mask-Dataset*. <https://github.com/MinhLinhEdu/Face-Mask-Dataset-of-Jessica-Li-Product-Manager-at-Kaggle>
- Nguyen, S. H., & Ha, D. T. (2023). A Lightweight Method for Detecting Cyber Attacks in High-Traffic Large Networks Based on Clustering Techniques. *International Journal of Computer Networks & Communications (IJCNC)*, 15(1), 35-51.
- Saiful Azian, S. J. B., Ng, H., Yap, T. T. V., Tong, H. L., Goh, V. T., & Cher, D. T. (2022). *Face Mask Detection Using Deep Learning*. In *Proceedings of the International Conference on Computer, Information Technology and Intelligent Computing (CITIC 2022)* (pp. 279-288). Atlantis Press. https://doi.org/10.2991/978-94-6463-094-7_22
- Wang, Z., Wang, G., Huang, B., Xiong, Z., Hong, Q., Wu, H., Yi, P., Jiang, K., Wang, N., Pei, Y., Chen, H., Miao, Y., Huang, Z., & Liang, J. (2020). *Masked Face Recognition Dataset and Application* (arXiv preprint arXiv:2003.09093). <https://doi.org/10.48550/arXiv.2003.09093>
- Yadav, R., Gautam, S., & Das, R. R. (2022, September 6). *Covid Face Mask Detection Using Neural Networks (Version 1)*. Preprint available at Research Square. <https://doi.org/10.21203/rs.3.rs-1982995/v1>

DETECTING FACE MASK USING A DEEP LEARNING METHOD*Dam Minh Linh^{1*}, Ngo Xuan Thoai¹, Han Minh Chau²**¹Posts and Telecommunications Institute of Technology in Ho Chi Minh City, Vietnam**²Hutech University, Ho Chi Minh City, Vietnam***Corresponding author: Dam Minh Linh – Email: linhdm@ptit.edu.vn**Received: August 31, 2023; Revised: October 26, 2023; Accepted: October 30, 2023***ABSTRACT**

Environmental pollution and respiratory diseases such as the COVID-19 pandemic, are capable of being transmitted through the air and affecting human health. To protect the safety of oneself and the community, one of the proposed solutions is to wear a mask. Therefore, this study focuses on detecting the faces of people wearing or not wearing masks from surveillance camera data, collected video data combined with a Convolutional Neural Network (CNN) algorithm deep learning, machine learning will classify the data into two labels. The research results have two major contributions: (a) detecting wearing a mask and not wearing a mask, and (b) proposing two CNN deep learning models evaluated and compared for Accuracy, Precision, Recall, and F1-Score with an accuracy of 99.94%.

Keywords: Convolutional Neural Network (CNN); Deep learning; Face Mask; OpenCV; Image processing