

I. Cải tiến mô hình

Để tăng độ chính xác cho mô hình, ta tiến hành các phương pháp cải tiến nhằm đem lại hiệu suất tốt, giảm sai số của mô hình.

Các Phương pháp cải tiến bao gồm:

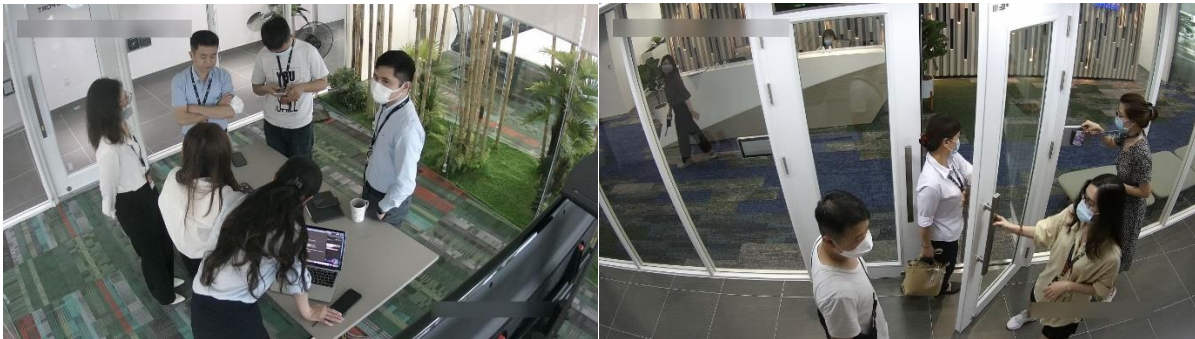
- Thêm data thật
- Thay đổi model mạnh hơn
- Augmentation
- ...

1. Phương pháp 1: Thêm data thật

Ta xem xét các bộ data trong thực tế và áp dụng chúng vào mô hình để “học” thực tế.

Bộ data 1:

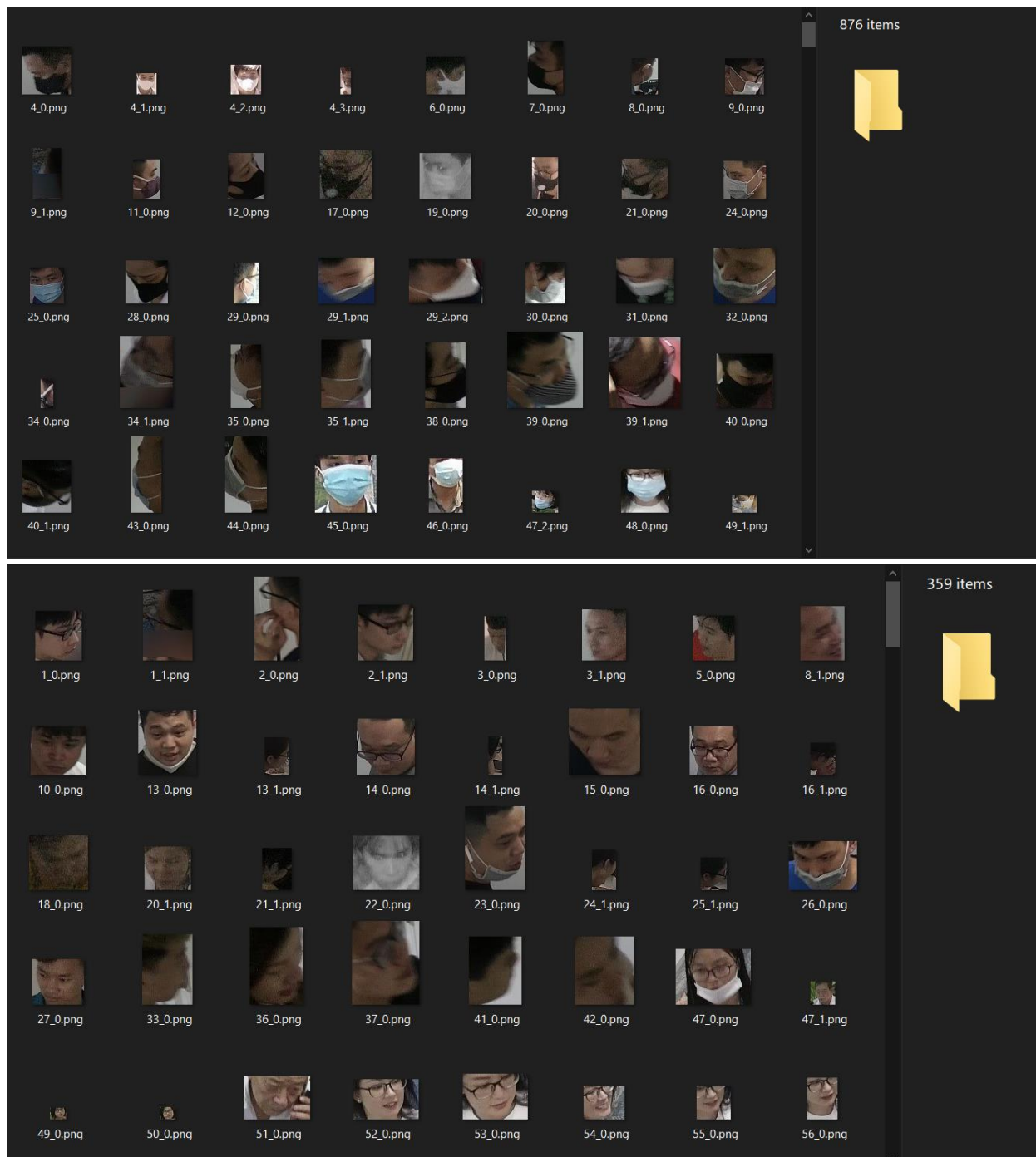
- Kích thước: 849 MB
- Gồm: file Images và Labels
 - Images: gồm 3 tập public_test, train và val với các ảnh (jpg) được thu thập từ camera giám sát tại các vị trí phòng ban ở 1 công ty.
 - Public_test: 88 ảnh
 - Train: 792 ảnh
 - Val: 184 ảnh
 - Labels: 3 tập public_test, train và val với các tập tin (txt) lưu trữ labels của từng ảnh tương ứng ở thư mục images.
- Ảnh minh họa



Đối với dataset này, ta tiến hành rút trích các khuôn mặt và gán label. Với các thông số box và label cho các ảnh có sẵn. Dùng một số thuật toán của OpenCV để cắt khuôn mặt theo box và tiến hành gán label cho các ảnh đó.

Kết quả:

- Tập Mask: 876 ảnh
- Tập No Mask: 359 ảnh



Test Mô hình cho bộ dữ liệu FPT:

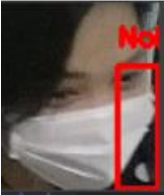

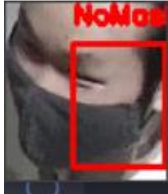


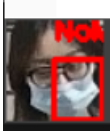
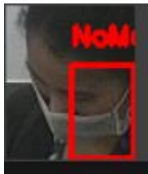
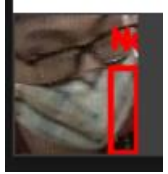
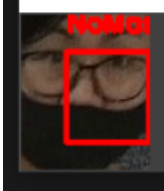
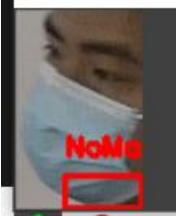
Tổng ảnh Input	876 (Mask) + 359 (NoMask) = 1235
Số lượng ảnh nhận diện được	125 (Mask) + 95 (NoMask) = 220

Số lượng ảnh không nhận diện được	352 (Mask) + 201 (NoMask) = 553
Số lượng ảnh đeo khẩu trang được nhận diện đúng label	1
Số lượng ảnh không đeo khẩu trang được nhận diện đúng label	95
Độ chính xác	$(1 + 95) / 220$ = 0.4364

Các giá trị không nhận diện được:

Mask	 ảnh bị che, góc từ phía sau	 Góc chụp từ phía sau	 Thiếu sáng	 Thiếu sáng, mất thông tin
	 Góc nghiêng	 Thiếu sáng	 Thiếu sáng, góc từ phía sau	
NoMask	 Góc cúi đầu	 Độ phân giải thấp, bị cắt hình	 Nghiêng, cúi đầu	 ảnh mờ, đen trắng
	 Nghiêng, cúi đầu	 ảnh mờ, đen trắng	 Nghiêng, cúi đầu	 ảnh mờ, đen trắng

Các giá trị nhận diện sai:

Mask				
				
				
No Mask	Nhận diện được 95/95 ảnh detect được			

Bộ data 2: [mirror] zalo ai challenge 2021 - 5k, tập data được lấy từ cuộc thi Zalo AI challenge 2021.

- Kích thước: 573 MB
- Gồm: public_test, sample_submission, train, mask.tflite
 - public_test: 542 ảnh các hoạt động được tổ chức trong đại dịch covid. Với file csv đánh số thứ tự ảnh.
 - sample_submission: 3 submission mẫu, các ảnh được phân loại theo 0 hoặc 1.
 - Train: gồm 4175 ảnh.
 - File mask.tflite
- Ảnh minh họa

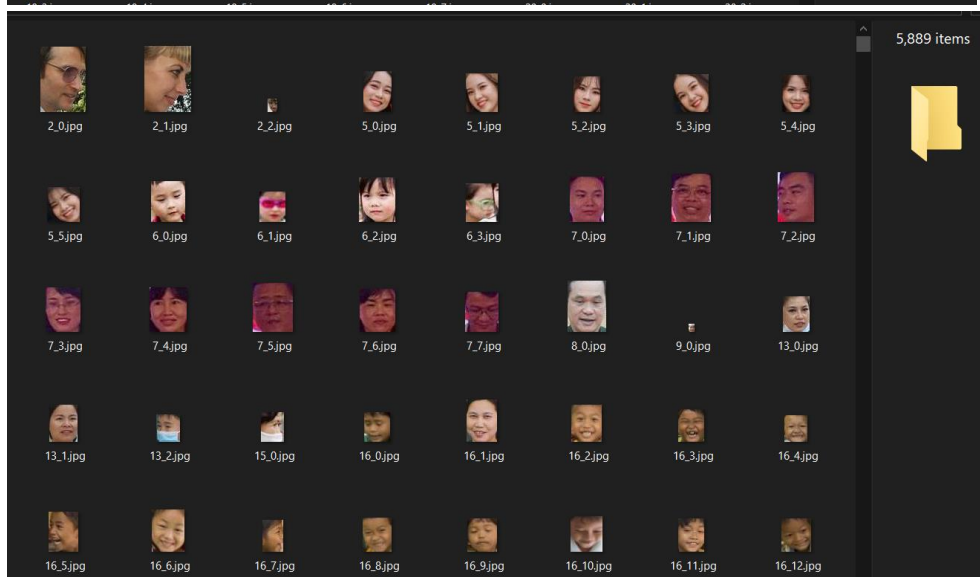
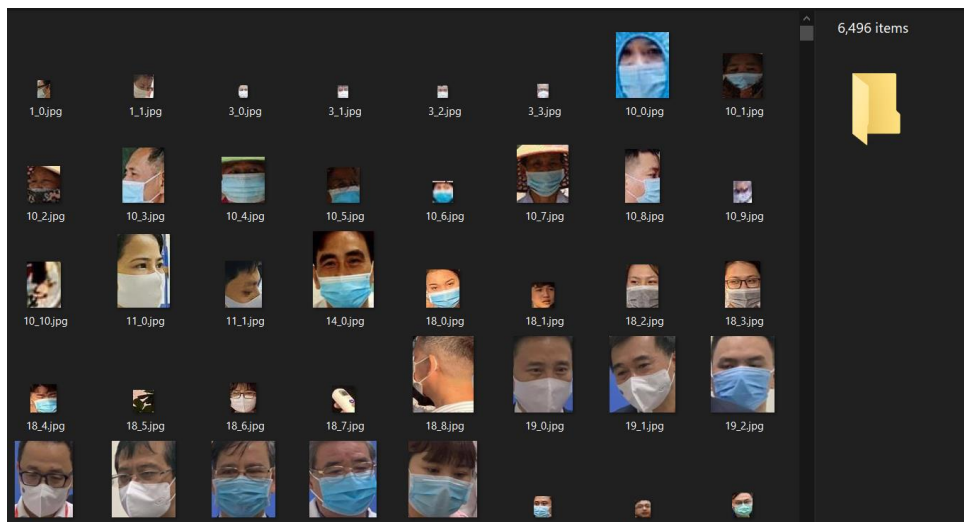


Đối với dataset này, ta tiến hành rút trích các khuôn mặt và gán label cho chúng.

Dùng một số thuật toán của OpenCV, MTCNN để xác định và rút trích khuôn mặt trong ảnh và gán label cho chúng.

Kết quả:

- Tập Mask: 6496 ảnh
- Tập No Mask: 5889 ảnh



Test Mô hình cho bộ dữ liệu ZALO:




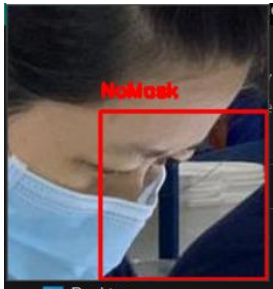
Tổng ảnh Input	6496 (Mask) + 5889 (NoMask) = 12385
Số lượng ảnh nhận diện được	472 (Mask) + 386 (NoMask) = 858
Số lượng ảnh không nhận diện được	4290 (Mask) + 3703 (NoMask) = 7993
Số lượng ảnh đeo khẩu trang được nhận diện đúng label	4
Số lượng ảnh không đeo khẩu trang được nhận diện đúng label	386
Độ chính xác	$(4 + 386) / 858$ = 0.4545


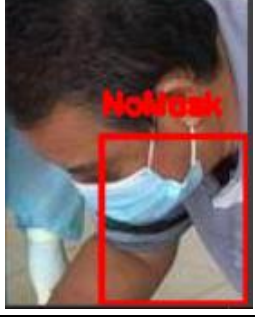
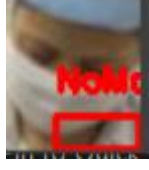

Các giá trị không nhận diện được:

Mask	 Góc nghiêng			
	 Ảnh mờ		 Bị vật cản che	
		 Độ phân giải thấp	 Góc nghiêng	
NoMask		 Góc nghiêng	 Độ phân giải thấp	

				 Độ phân giải thấp
	 Độ phân giải thấp, góc nghiêng		 góc nghiêng	 góc nghiêng

Các giá trị nhận diện sai:

Mask				
				
				

				
NoMask	Nhận diện đúng 386/386 ảnh			

Huấn luyện lại mô hình

Tiếp theo, ta tiến hành thêm các tập dữ liệu để tiến hành huấn luyện lại mô hình:

Cấu trúc của dataset như sau:

- **With_mask:** khoảng 24k ảnh
- **Without_mask:** khoảng 24k ảnh

Kết quả sau quá trình huấn luyện như sau:

```
Epoch 10/20
1275/1275 [=====] - 299s 234ms/step - loss: 0.1423 - accuracy: 0.9436 - val_loss: 0.1601 - val_accuracy: 0.9344
Epoch 11/20
1275/1275 [=====] - 296s 232ms/step - loss: 0.1368 - accuracy: 0.9461 - val_loss: 0.1588 - val_accuracy: 0.9348
Epoch 12/20
1275/1275 [=====] - 325s 255ms/step - loss: 0.1340 - accuracy: 0.9480 - val_loss: 0.1602 - val_accuracy: 0.9338
Epoch 13/20
1275/1275 [=====] - 316s 248ms/step - loss: 0.1288 - accuracy: 0.9494 - val_loss: 0.1546 - val_accuracy: 0.9369
Epoch 14/20
1275/1275 [=====] - 301s 236ms/step - loss: 0.1247 - accuracy: 0.9514 - val_loss: 0.1546 - val_accuracy: 0.9358
Epoch 15/20
1275/1275 [=====] - 311s 244ms/step - loss: 0.1204 - accuracy: 0.9534 - val_loss: 0.1544 - val_accuracy: 0.9380
Epoch 16/20
1275/1275 [=====] - 314s 246ms/step - loss: 0.1191 - accuracy: 0.9540 - val_loss: 0.1556 - val_accuracy: 0.9374
Epoch 17/20
1275/1275 [=====] - 282s 221ms/step - loss: 0.1150 - accuracy: 0.9557 - val_loss: 0.1528 - val_accuracy: 0.9377
Epoch 18/20
1275/1275 [=====] - 298s 234ms/step - loss: 0.1133 - accuracy: 0.9567 - val_loss: 0.1539 - val_accuracy: 0.9381
Epoch 19/20
1275/1275 [=====] - 302s 237ms/step - loss: 0.1097 - accuracy: 0.9573 - val_loss: 0.1479 - val_accuracy: 0.9418
Epoch 20/20
1275/1275 [=====] - 292s 229ms/step - loss: 0.1068 - accuracy: 0.9589 - val_loss: 0.1495 - val_accuracy: 0.9394
```


Test lại mô hình trên bộ fpt và zalo:





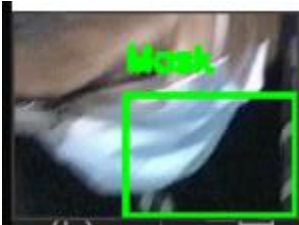


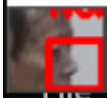




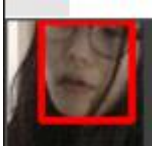

Test Mô hình cho bộ dữ liệu FPT:

Tổng ảnh Input	876 (Mask) + 359 (NoMask) = 1235
Số lượng ảnh nhận diện được	125 (Mask) + 95 (NoMask) = 220
Số lượng ảnh không nhận diện được	352 (Mask) + 201 (NoMask) = 553
Số lượng ảnh đeo khẩu trang được nhận diện đúng label	6
Số lượng ảnh không đeo khẩu trang được nhận diện đúng label	91
Độ chính xác	$(1 + 95) / 220$ = 0.4364

Các giá trị không nhận diện được:

Mask	 ảnh bị che, góc từ phía sau	 Góc chụp từ phía sau	 Thiếu sáng	 Thiếu sáng, mất thông tin
	 Góc nghiêng	 Thiếu sáng	 Thiếu sáng, góc từ phía sau	
NoMask	 Góc cúi đầu	 Độ phân giải thấp, bị cắt hình	 	
	 	 	 Nghiêng, cúi đầu	 ảnh mờ, đen trắng

Các giá trị nhận diện đúng:

Mask				
				
No Mask				
				





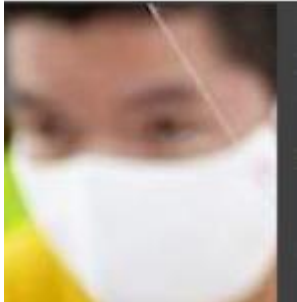















Bộ zalo:

Test Mô hình cho bộ dữ liệu ZALO:

Tổng ảnh Input	6496 (Mask) + 5889 (NoMask) = 12385
Số lượng ảnh nhận diện được	472 (Mask) + 386 (NoMask) = 858
Số lượng ảnh không nhận diện được	4290 (Mask) + 3703 (NoMask)

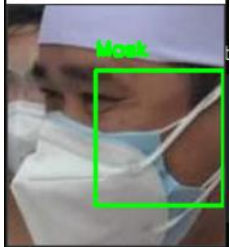
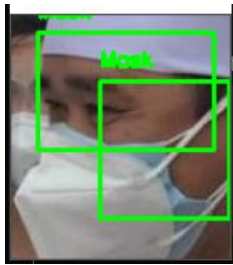


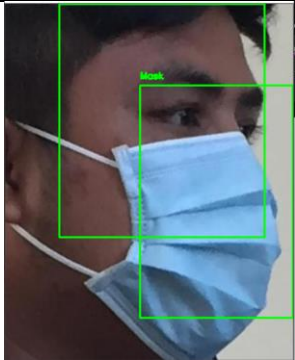



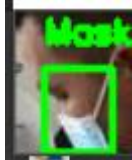

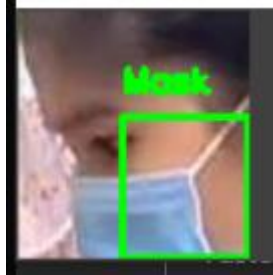
	= 7993
Số lượng ảnh đeo khẩu trang được nhận diện đúng label	26
Số lượng ảnh không đeo khẩu trang được nhận diện đúng label	383
Độ chính xác	$(26 + 383) / 858$ = 0.4767









Các giá trị không nhận diện được:

Mask	 Góc nghiêng			
	 Ảnh mờ		 Bị vật cản che	
		 Độ phân giải thấp	 Góc nghiêng	
NoMask		 Góc nghiêng	 Độ phân giải thấp	
				 Độ phân giải thấp

	 <p>Độ phân giải thấp, góc nghiêng</p>		 <p>góc nghiêng</p>	 <p>góc nghiêng</p>
--	-------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------

Các giá trị nhận diện đúng:

Mask				
				
				

No Mask				
				

TEST LẠI MÔ HÌNH:

Tính toán độ chính xác của mô hình thông qua confusion matrix:

Chỉnh sửa data: Thêm bộ data từ github của face_mask_detection và 2 tập validation của 2 bộ fpt và zalo

Thu được bộ data như sau:

- Mask: khoảng 26k ảnh
- No Mask: khoảng 26k ảnh

Thực hiện huấn luyện lại mô hình:

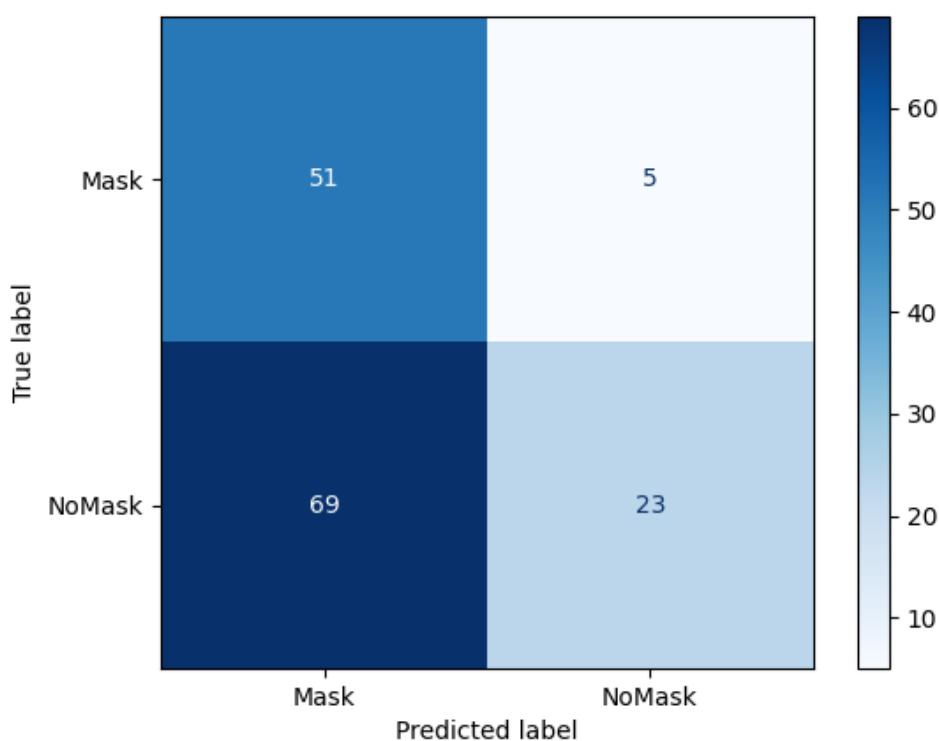

```

Epoch 17/20
582/1384 [=====>.....] - ETA: 2:14 - loss: 0.1137 - accuracy: 0.9543/usr/local/lib/python3.7/dist-packages/PIL/Image.py:
"Palette images with Transparency expressed in bytes should be "
1384/1384 [=====] - ETA: 0s - loss: 0.1147 - accuracy: 0.9557/usr/local/lib/python3.7/dist-packages/PIL/Image.py:96
"Palette images with Transparency expressed in bytes should be "
1384/1384 [=====] - 315s 227ms/step - loss: 0.1147 - accuracy: 0.9557 - val_loss: 0.1379 - val_accuracy: 0.9434
Epoch 18/20
573/1384 [=====>.....] - ETA: 2:15 - loss: 0.1105 - accuracy: 0.9583/usr/local/lib/python3.7/dist-packages/PIL/Image.py:
"Palette images with Transparency expressed in bytes should be "
1384/1384 [=====] - ETA: 0s - loss: 0.1111 - accuracy: 0.9574/usr/local/lib/python3.7/dist-packages/PIL/Image.py:96
"Palette images with Transparency expressed in bytes should be "
1384/1384 [=====] - 285s 206ms/step - loss: 0.1111 - accuracy: 0.9574 - val_loss: 0.1373 - val_accuracy: 0.9448
Epoch 19/20
518/1384 [=====>.....] - ETA: 2:26 - loss: 0.1082 - accuracy: 0.9593/usr/local/lib/python3.7/dist-packages/PIL/Image.py:
"Palette images with Transparency expressed in bytes should be "
1384/1384 [=====] - ETA: 0s - loss: 0.1089 - accuracy: 0.9585/usr/local/lib/python3.7/dist-packages/PIL/Image.py:96
"Palette images with Transparency expressed in bytes should be "
1384/1384 [=====] - 293s 211ms/step - loss: 0.1089 - accuracy: 0.9585 - val_loss: 0.1369 - val_accuracy: 0.9443
Epoch 20/20
116/1384 [=>.....] - ETA: 3:39 - loss: 0.1147 - accuracy: 0.9537/usr/local/lib/python3.7/dist-packages/PIL/Image.py:
"Palette images with Transparency expressed in bytes should be "
1384/1384 [=====] - ETA: 0s - loss: 0.1080 - accuracy: 0.9586/usr/local/lib/python3.7/dist-packages/PIL/Image.py:96
"Palette images with Transparency expressed in bytes should be "
1384/1384 [=====] - 318s 229ms/step - loss: 0.1080 - accuracy: 0.9586 - val_loss: 0.1362 - val_accuracy: 0.9454

```

Xử lý bộ dữ liệu test: Lấy dữ liệu từ bộ `fpt_public_test` để kiểm tra mô hình

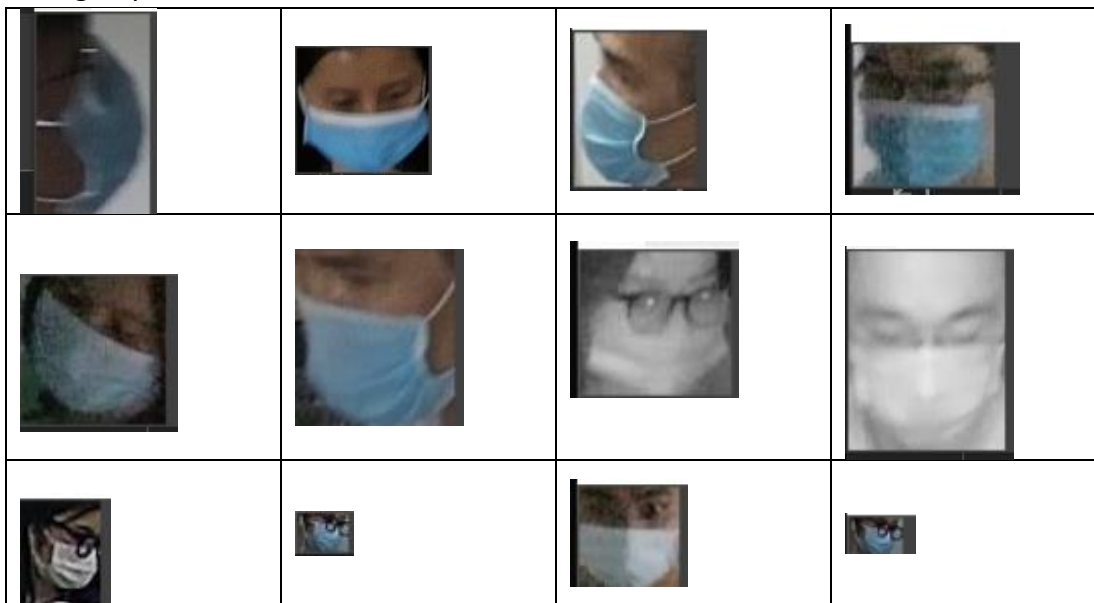
Kết quả:



Qua ảnh của confusion matrix ta có thể thấy:

- Tổng số lượng mẫu: $51 + 23 + 5 + 69 = 148$
- Độ chính xác của mô hình: $\frac{51+23}{51+23+5+69} = 50\%$
- Các trường hợp sai chủ yếu ở class Mask khi có 69 mẫu bị dự đoán sai thành No Mask

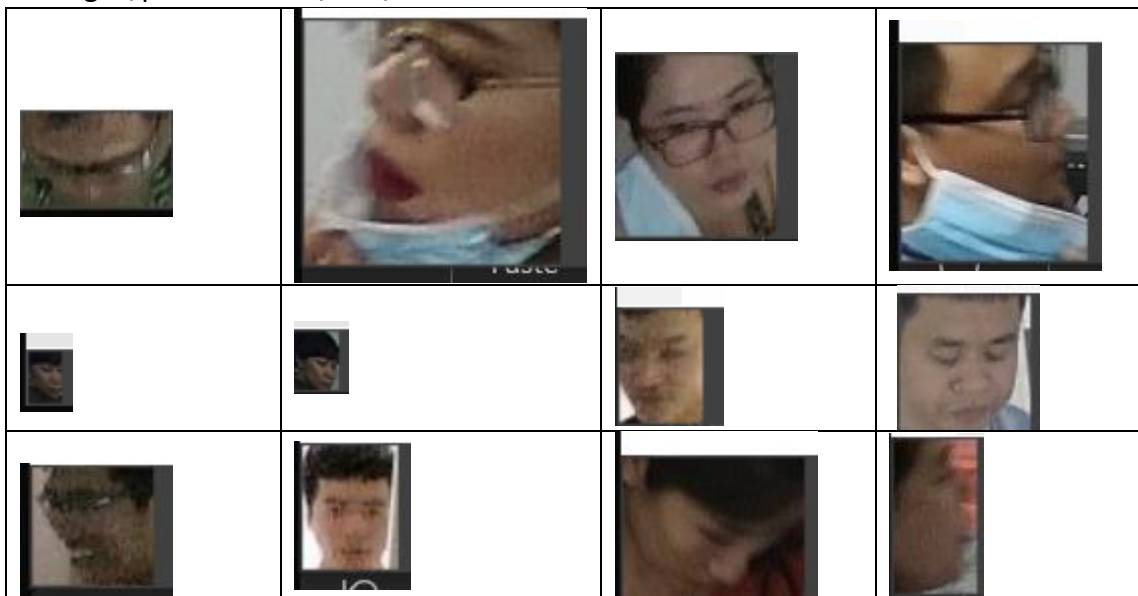
Các trường hợp Mask nhận diện sai thành No Mask:



Nhận xét: có thể do các nguyên nhân sau:

- Ảnh tối đen trắng
- Kích thước quá nhỏ
- Các góc bị mất mát chi tiết nhiều
- Góc nghiêng làm mất chi tiết

Các trường hợp NoMask nhận diện sai thành Mask:



Nhận xét: có thể do các nguyên nhân sau:

- Kích thước quá nhỏ
- Các góc bị mất mát chi tiết nhiều
- Góc nghiêng làm mất chi tiết