OXFORD BROOKES UNIVERSITY

Mask Wearing Detection Based on Faster R-CNN

Department of Computing Science,
Chengdu University of Technology, Oxford Brookes University

Student: Lucy 202018010103

Supervisor: Happy Nkanta Monday

Contents

INTRODUCTION

OXFORD BROOKES UNIVERSITY

COVID-19

- Novel coronavirus pneumonia outbreak (COVID-19) in Wuhan, December 2019.
- Caused by SARS-CoV-2 virus.
- Transmission through person-to-person contact and contaminated surfaces.

Measures

- Vaccination
- Wearing a mask (is most efficient)
- ....

AIM

OXFORD BROOKES UNIVERSITY

Aim:

Development of an efficient mask detection system using Faster-RCNN algorithm with ResNet-50 and FPN backbone network to detect the wearing of masks in public places and promote public health.

Audience:

- Public Health Departments and Government Agencies
- Hospitals and Health Facilities
- Transportation Sector
- General Public

BACKGROUND REVIEW

OXFORD BROOKES UNIVERSITY

Researcher	Method	Results	Disadvantage
Qin et al. [1]	SRC-Net: Super Resolution (SR) network with MobileNetV2	Accuracy 98.70%	Small dataset, few features, slow detection speed
Xu et al. [2]	SSD-based algorithm combined with channel attention mechanism	Accuracy 90.2%, recall 86.5%, F1 score 88.2%	/
Amit et al. [3]	Two-stage detector: RetinaFace detection with NASNetMobile classifier	Accuracy 98.28%, recall 100%, F1 score 99.13%	High complexity, low video frame rate
Madhura et al. [4]	Facemasknet	Accuracy 98.6%	Small dataset, contains only 35 images, may have regional bias
Loey et al. [5]	YOLOv2 + ResNet-50	Average accuracy 81%	ADAM optimizer outperforms SGDM Unable to recognize occluded faces in videos

What is Faster-RCNN and Why i choose it

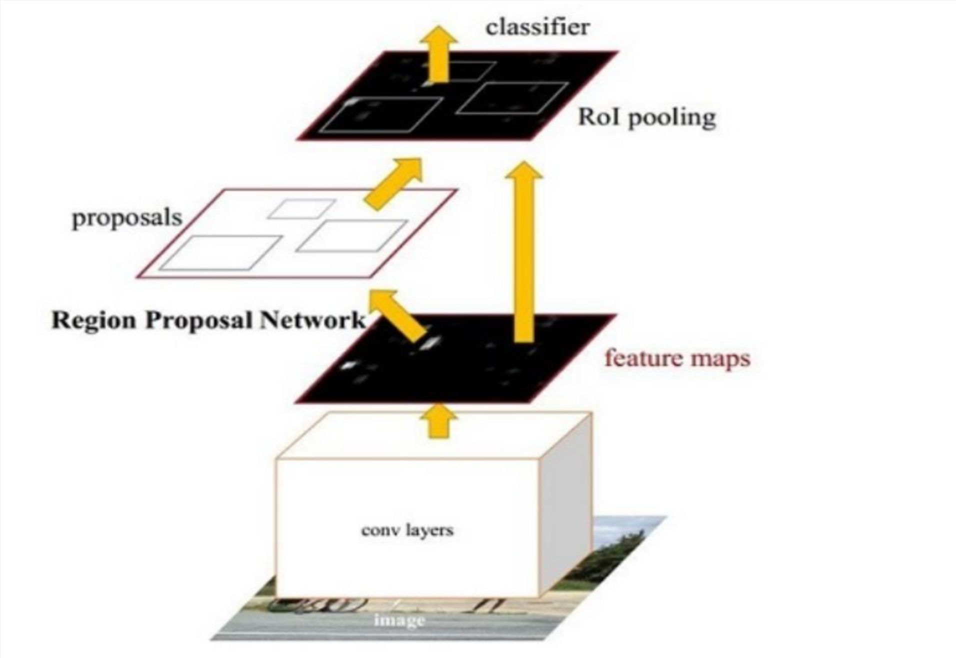
OXFORD BROOKES UNIVERSITY

- R-CNN and Fast R-CNN use selective search to generate region suggestions.

- Faster R-CNN introduces RPN to embed region suggestions into the network.

- Faster R-CNN combines RPN and Fast R-CNN through a shared feature layer.

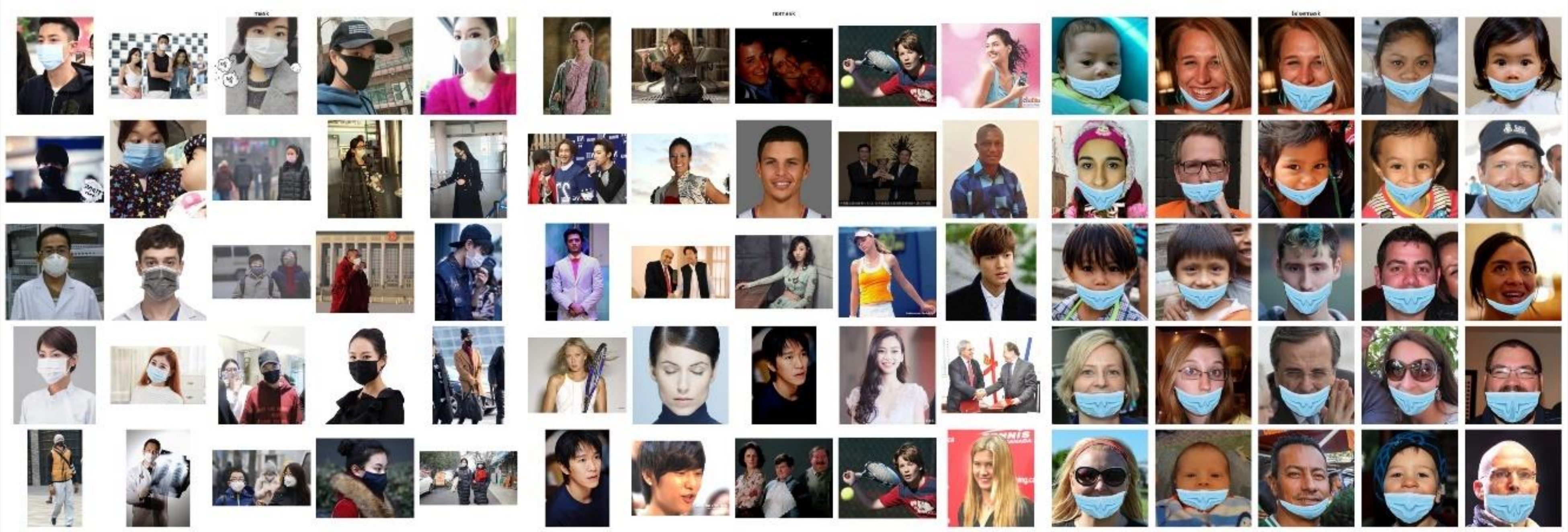
- RPN is efficient, making Faster R-CNN an order of magnitude faster than Fast R-CNN.



DATASETS

Overview

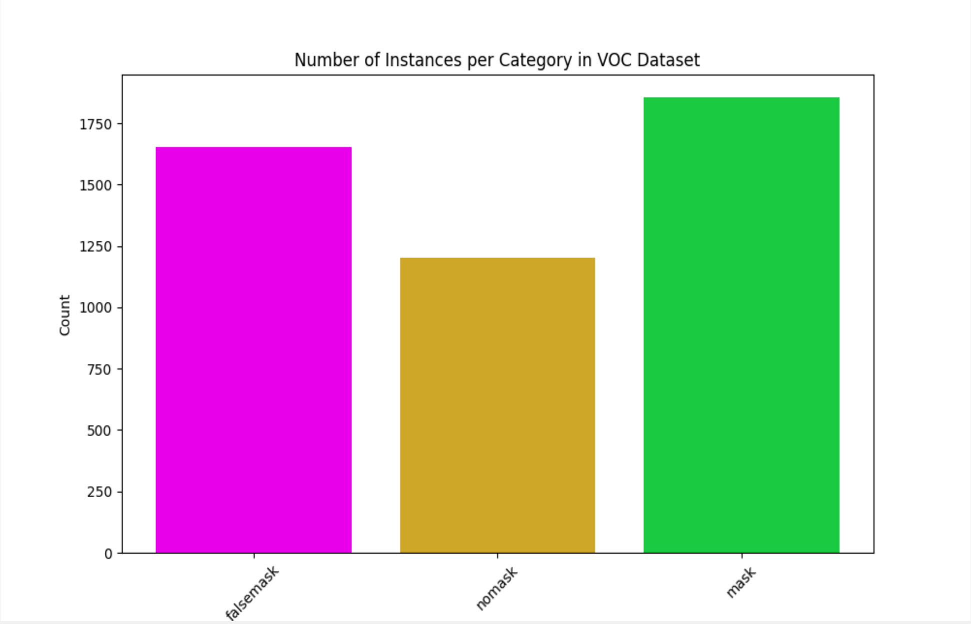
- A dataset containing three categories (mask, no mask, false mask) was constructed, totaling 3954 images.
- The images were mainly derived from the Masked Face-Net dataset [6].
- The dataset is diverse, covering different races, ages, genders, and lighting conditions.

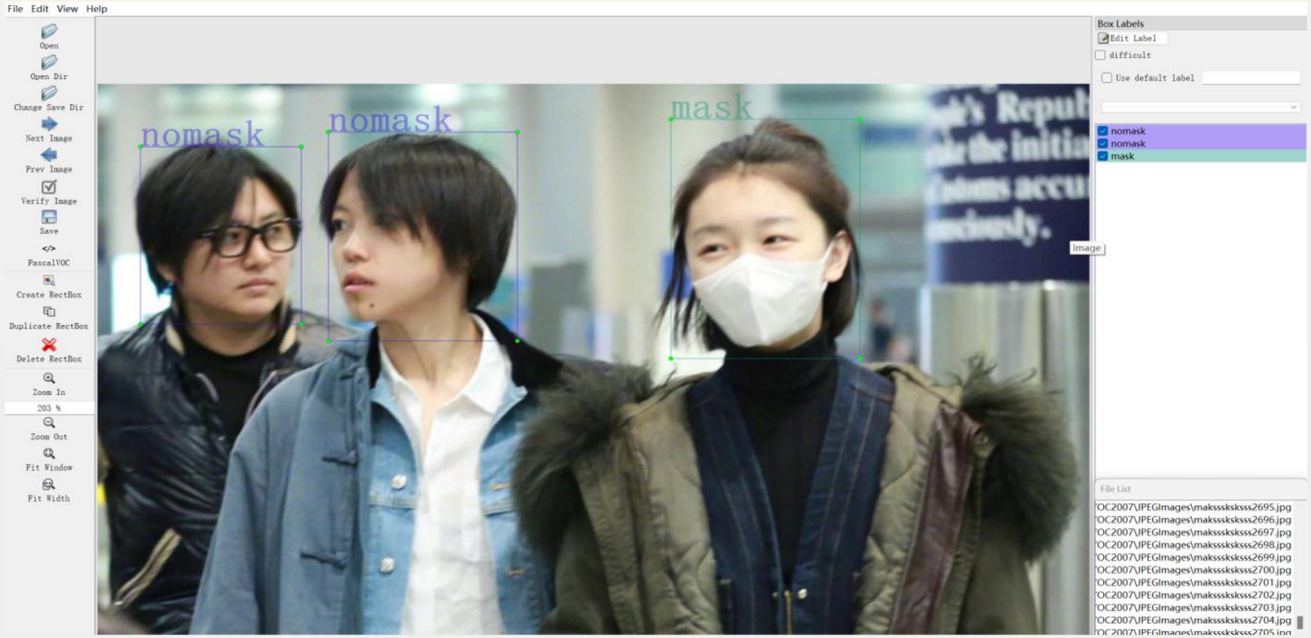


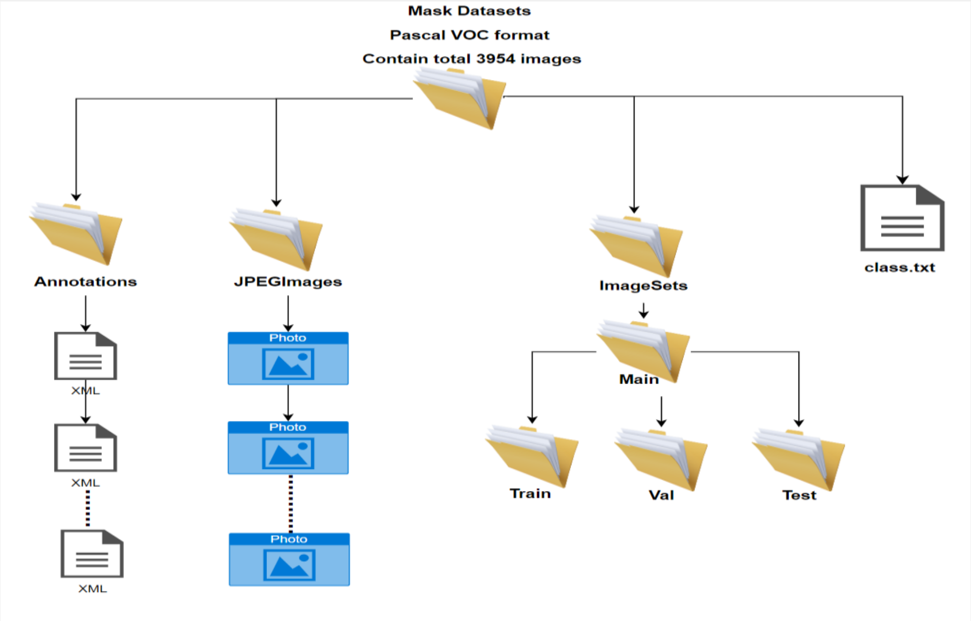
DATASETS

OXFORD BROOKES UNIVERSITY

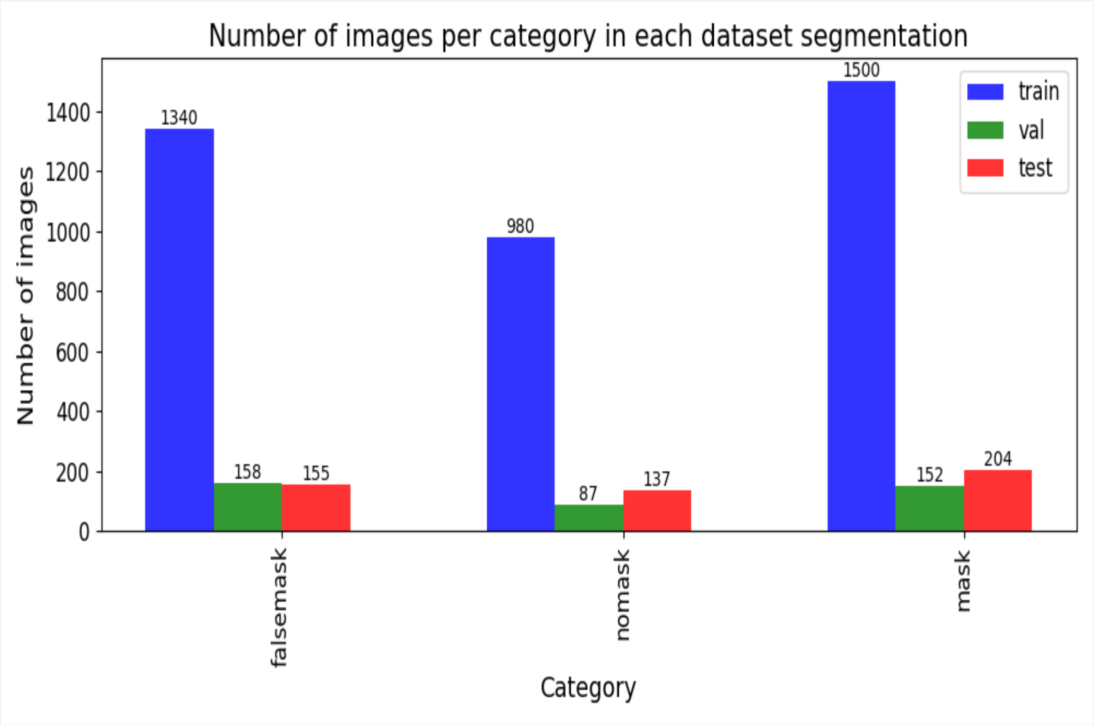
- Pascal VOC format was used
- Labeled image data using labelImg
- mask contains 1856 samples, no mask contains 1204 samples, false mask contains 1653 samples.



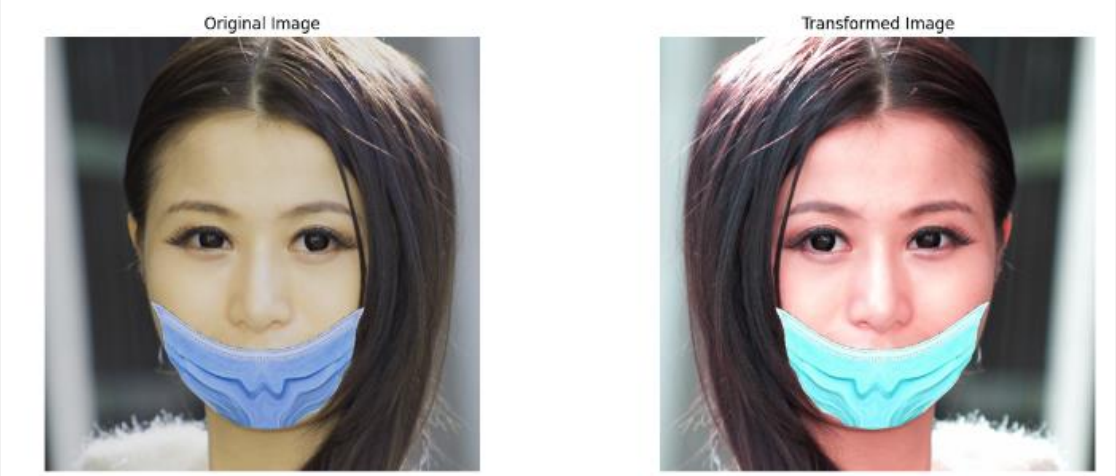




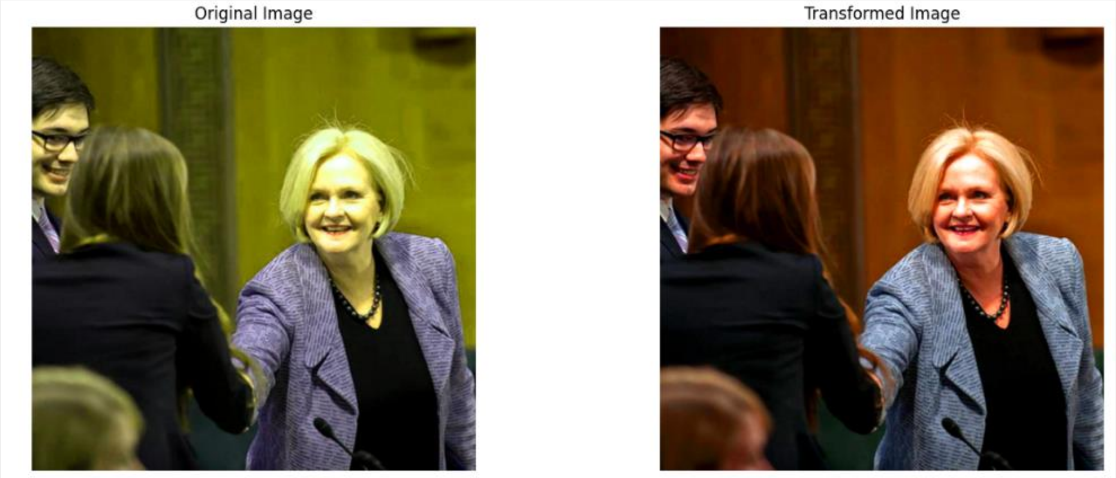
DATASETS



I divided the datasets by the 8:1:1 ratio

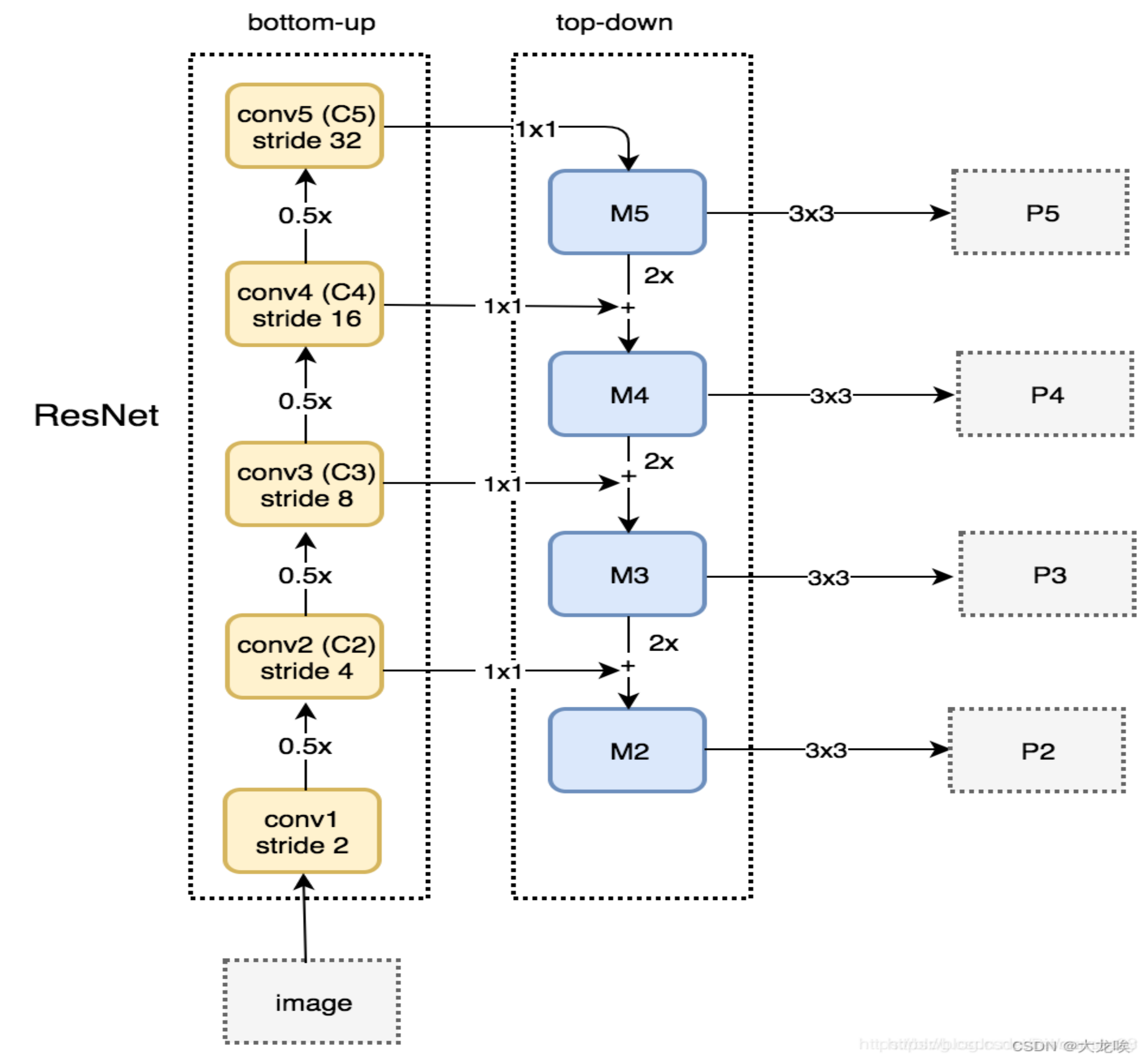


- Flipped the image horizontally with a probability of 50%.
- Randomize the color properties of an image



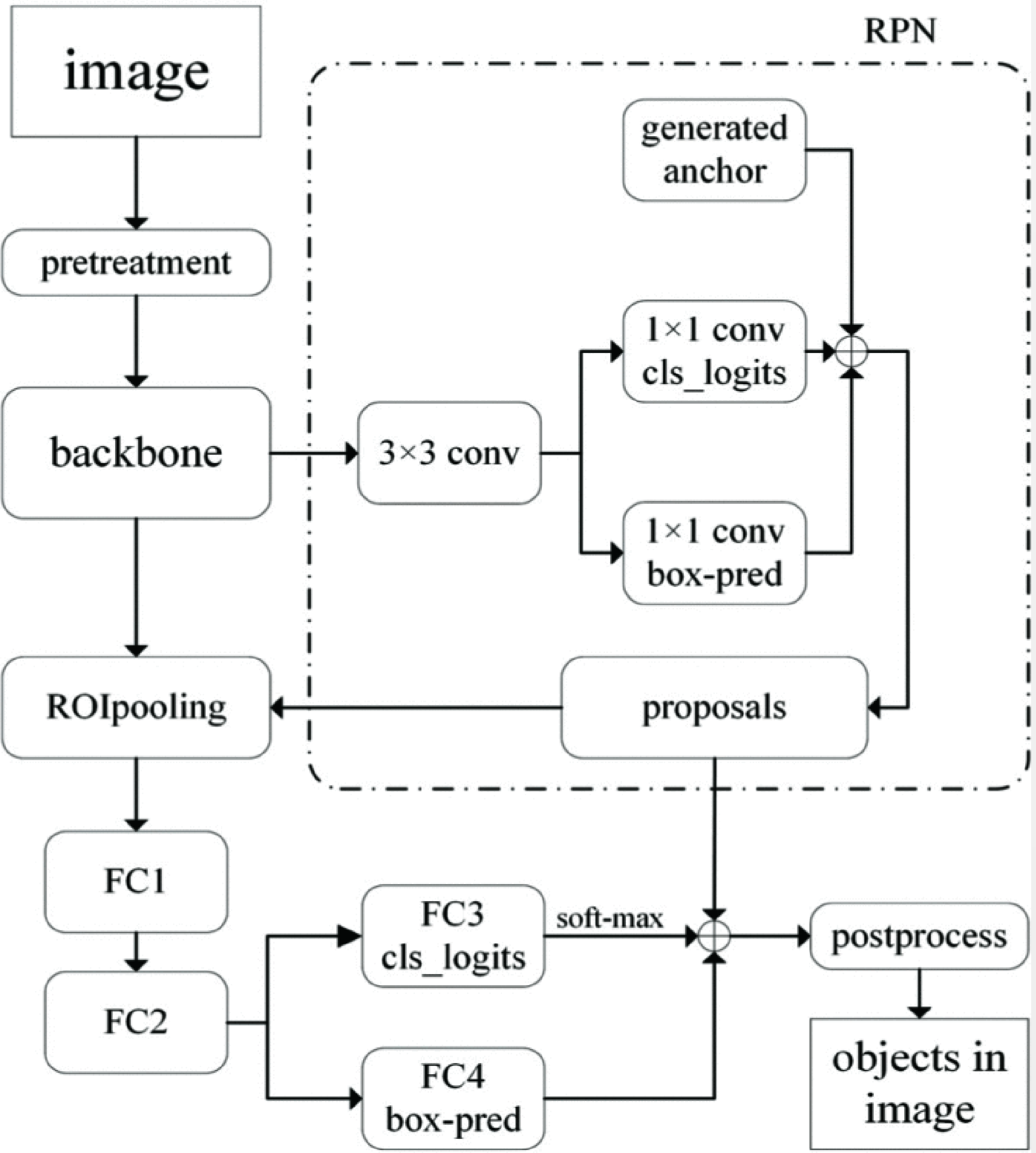
MODEL ARCHITECTURE

OXFORD BROOKES UNIVERSITY



In this project, FPN and ResNet50 are used as the backbone network of Faster R-CNN.

MODEL ARCHITECTURE



① Faster R-CNN extracts features using ResNet50+FPN

② RPN generates anchors, computes target probability and positional adjustment (NMS's)

③ Filtered region proposals and feature maps enter ROI pooling layer

④ ROI pooling layer through FCL classification and edge adjustment

⑤ Final labeling of object location

EXPERIEMNT

OXFORD BROOKES UNIVERSITY

Table 2. Basic information about experimental software and hardware		
Experimental environment		Configuration instructions
Hardware	CPU	12th Gen Intel(R) Core (TM) i9-12900H (20 CPUs), ~2.5GHz
	GPU	NVIDIA GeForce RTX 3060
	Memory	32768MB RAM
Software	Operating system	Windows 11, 64 bits
	Programming environment	Python 3.8.8; Pytorch 1.13.0; TensorFlow 2.13.0; NumPy 1.24.3

Hyperparameter	Freeze Training Phase	Unfreeze Training Phase
learn	0.005	0.005
batch size	8	8
epoch	5	20
optimizer	SGD	SGD
weight decay	0.0005	0.0005
learning rate scheduler	/	StepLR, Step Size=3, Gamma=0.33

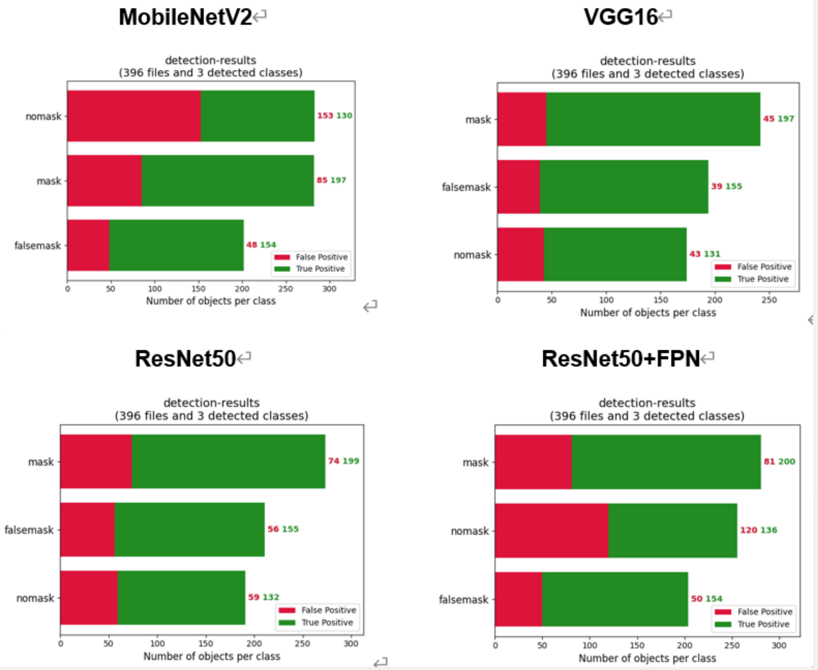
Initial Parameter Settings

RESULTS

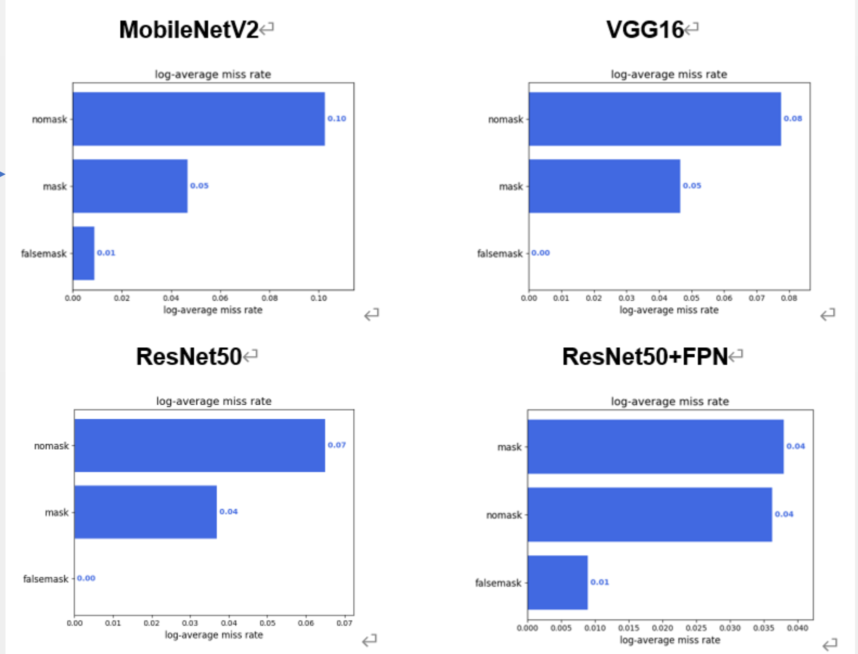
OXFORD BROOKES UNIVERSITY

Metrics	MobileNetV2	VGG16	ResNet50	ResNet50+FPN
F1	95.00%	96.67%	96.00%	96.00%
Recall	95.59%	96.45%	96.81%	96.45%
Precision	94.34%	97.00%	95.31%	94.64%
mAP	96.48%	96.99%	97.59%	98.40%

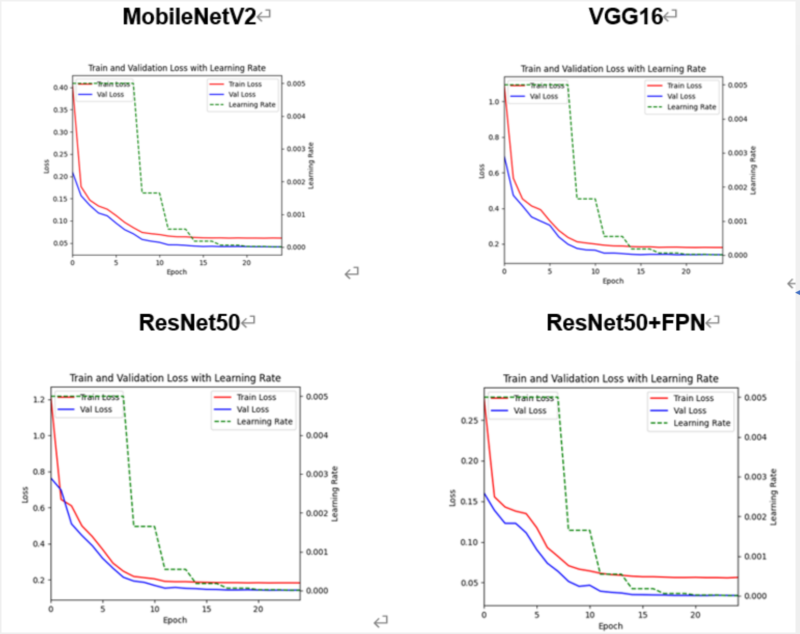
TP and FP



LAMR

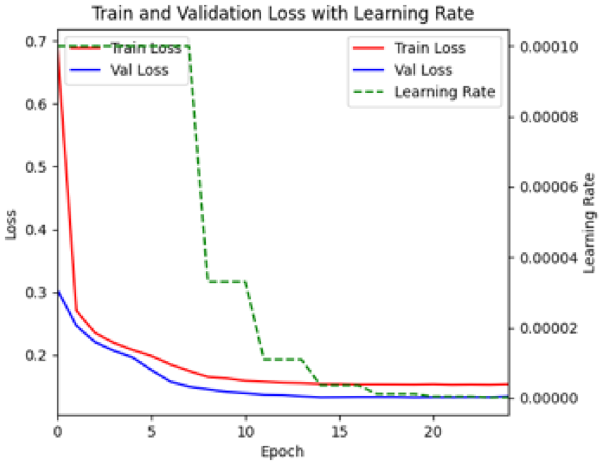


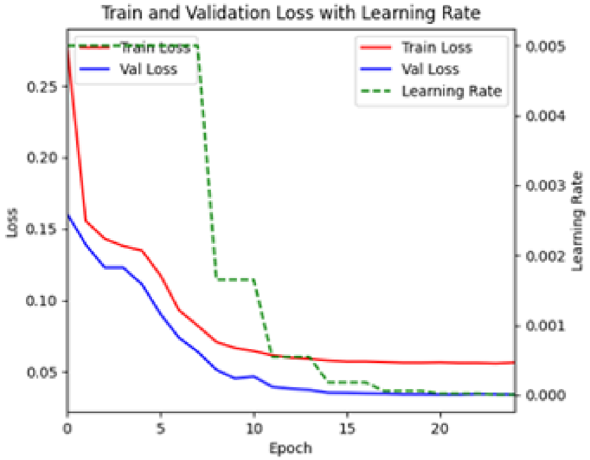
Training Results

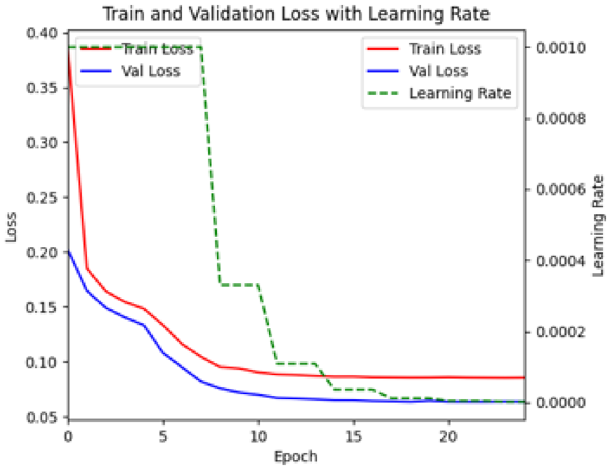


RESULTS

0.0001	0.005	0.001







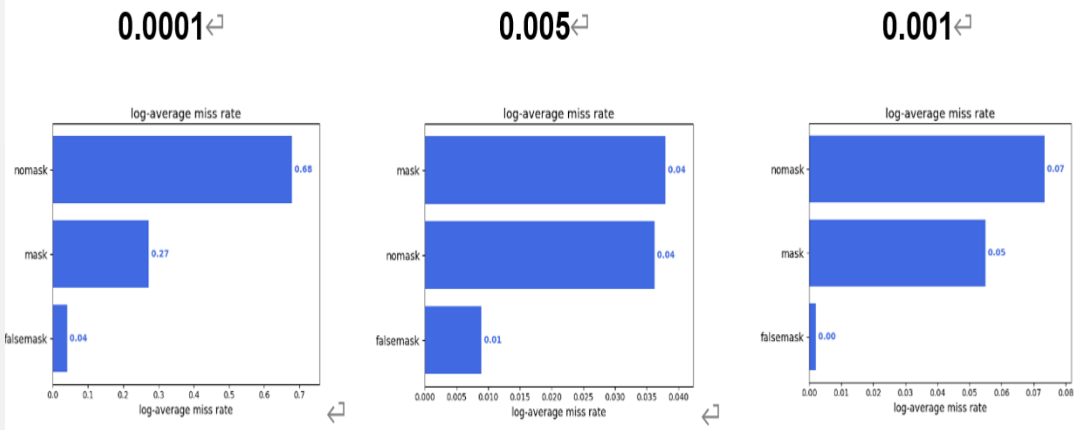
Training Results

RESULTS

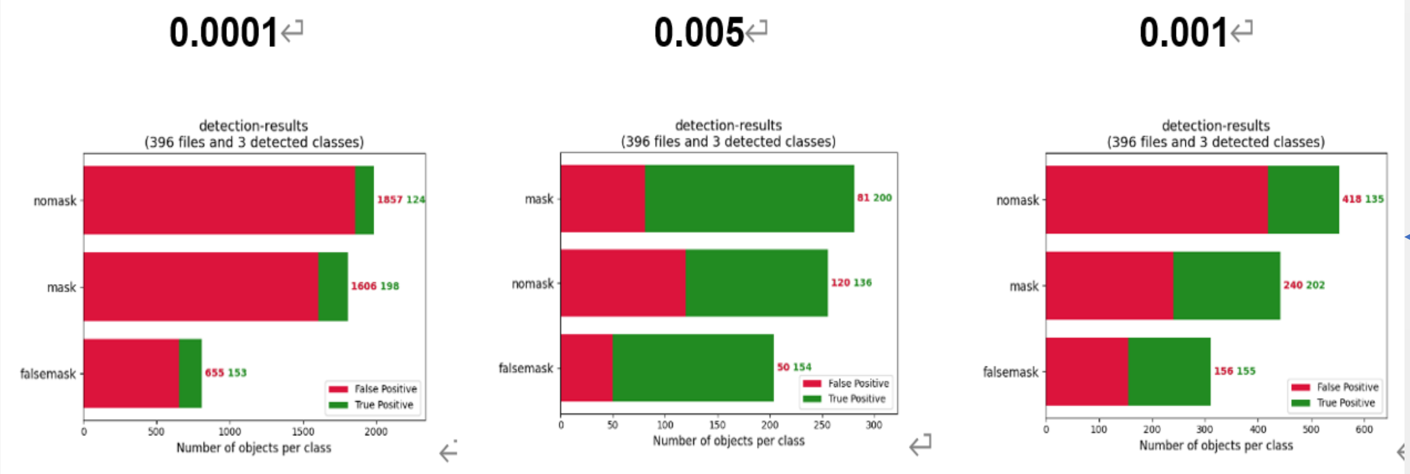
OXFORD BROOKES UNIVERSITY

Metrics	0.0001	0.001	0.005
F1	69.67%	92.33%	96.00%
Recall	78.13%	96.62%	96.45%
Precision	63.53%	88.33%	94.64%
mAP	79.03%	97.89%	98.40%

LAMR



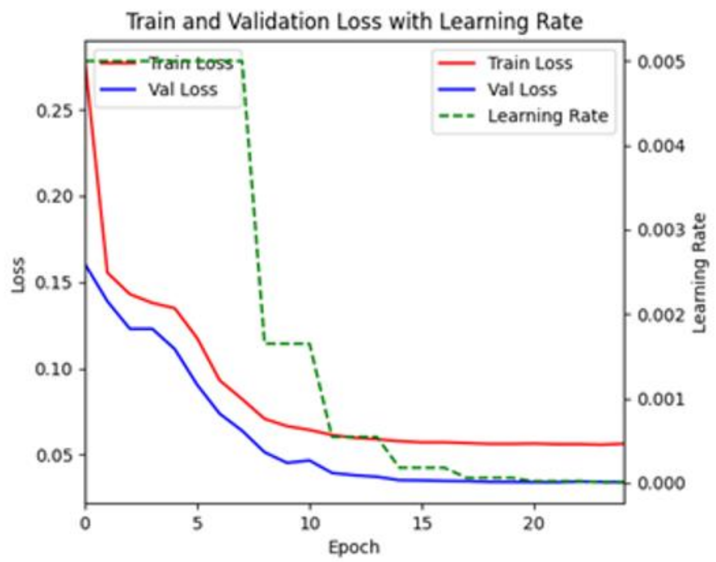
TP and FP



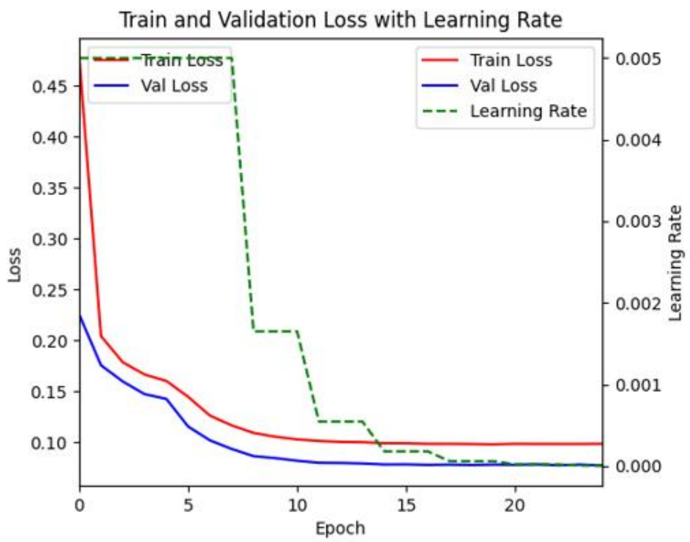
RESULTS

OXFORD BROOKES UNIVERSITY

SGD



ASGD

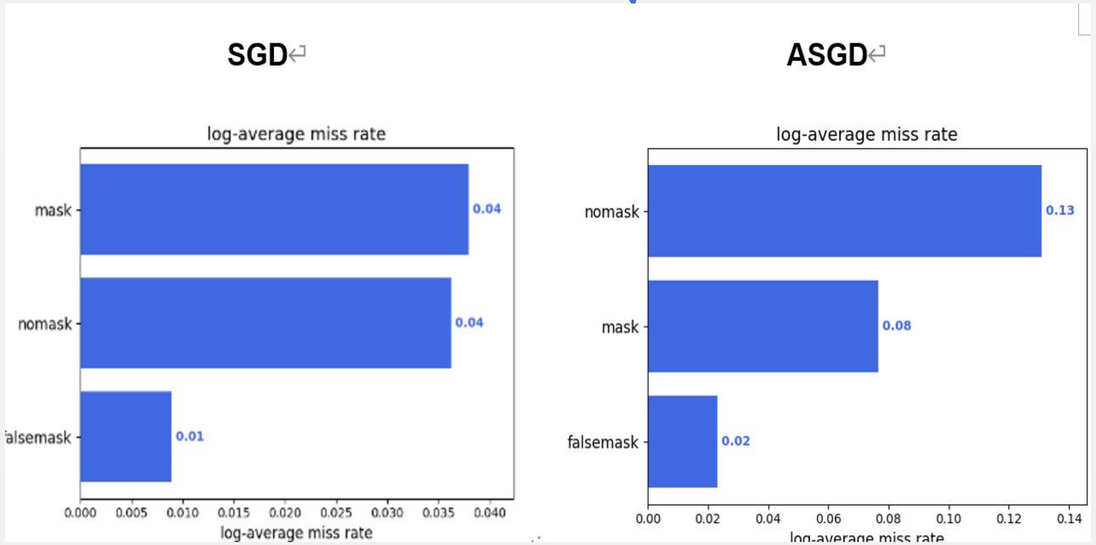


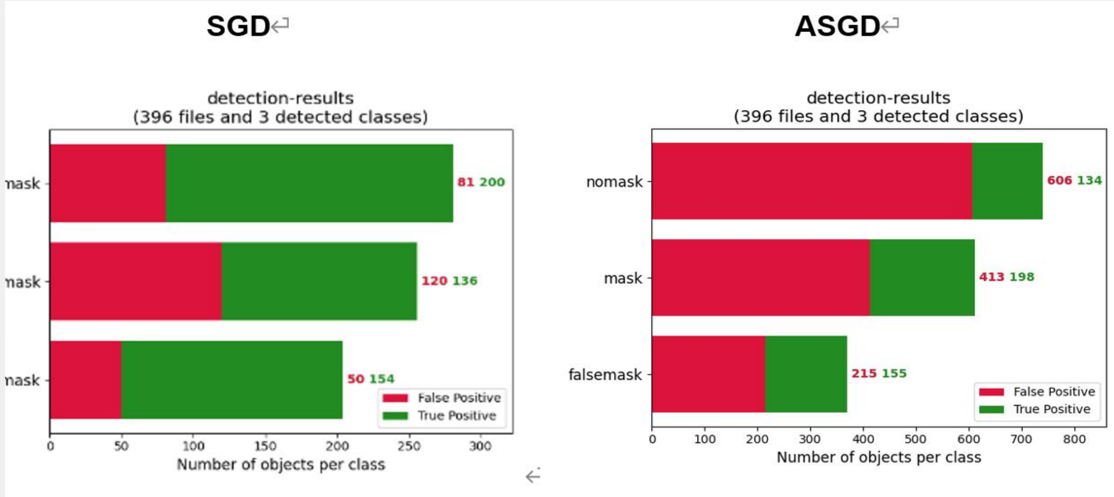
Training Results

RESULTS

Metrics	SGD	ASGD
F1	96.00%	88.67
Recall	96.45%	94.37
Precision	94.64%	83.67
mAP	98.40%	96.01%

LAMR





TP and FP

RESULTS

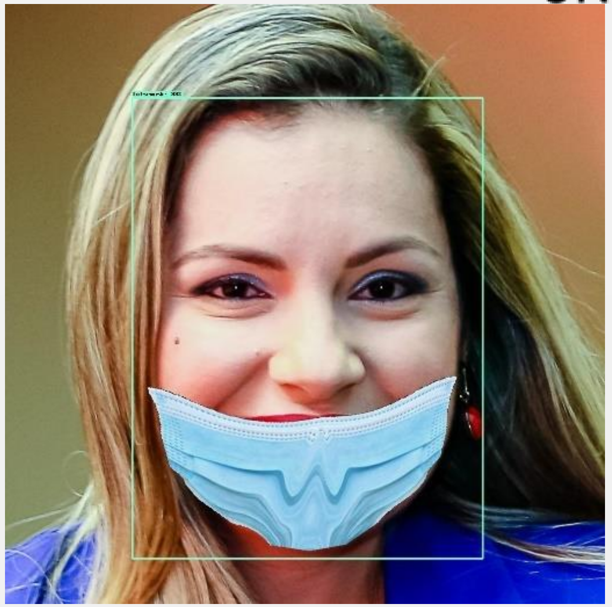
OXFORD BROOKES UNIVERSITY

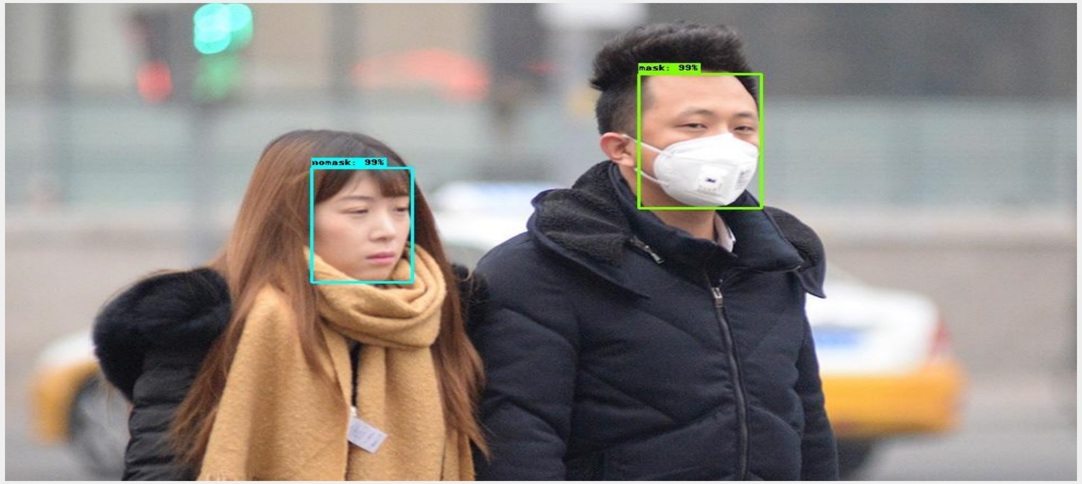
Hyperparameter	Freeze Training Phase	Unfreeze Training Phase
Backbone	ResNet50+FPN	ResNet50+FPN
Learning Rate	0.005	0.005
Batch Size	8	8
Epoch	5	20
Optimizer	SGD	SGD
Weight Decay	0.0005	0.0005
Learning Rate Scheduler	/	StepLR, Step Size=3 Gamma=0.33

Final Parameters Setting

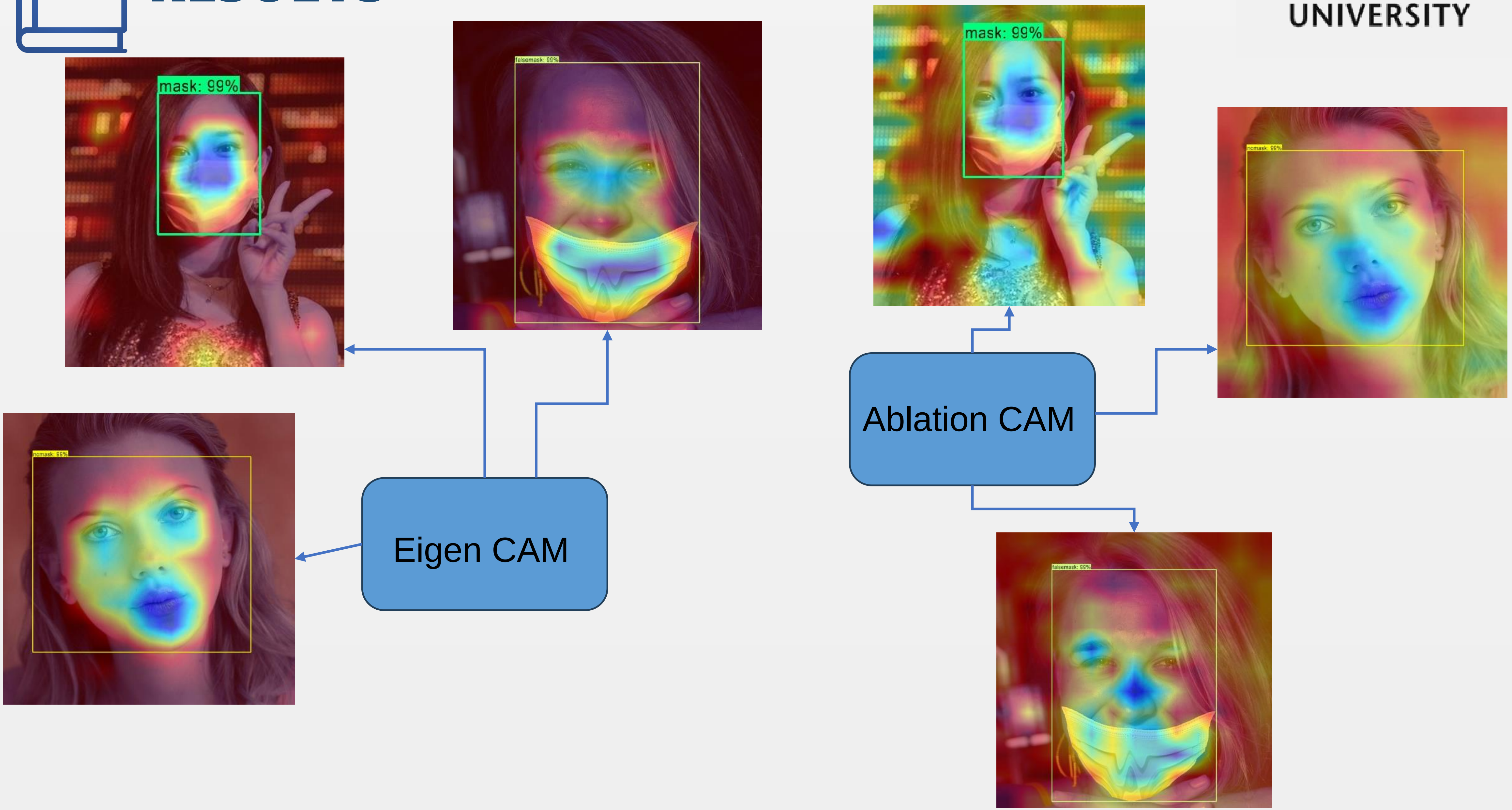
RESULTS

OXFORD BROOKES UNIVERSITY

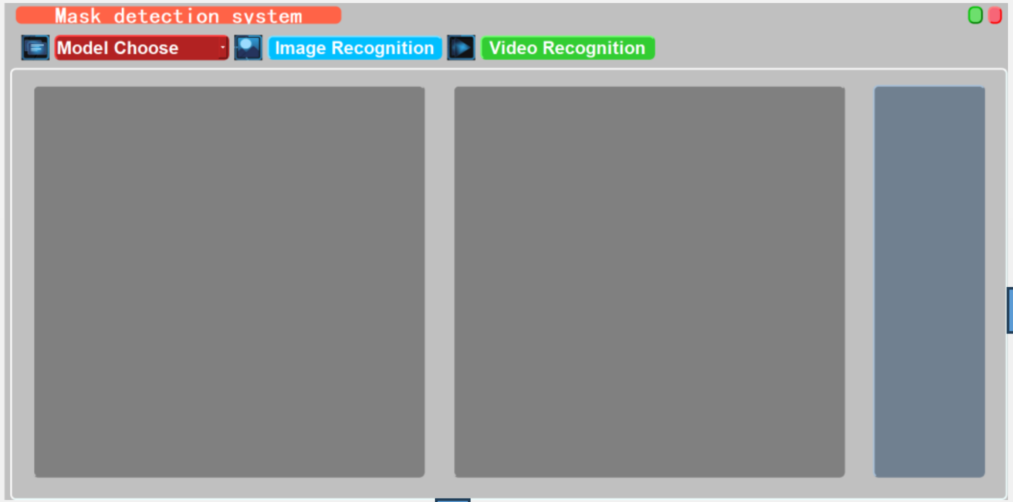


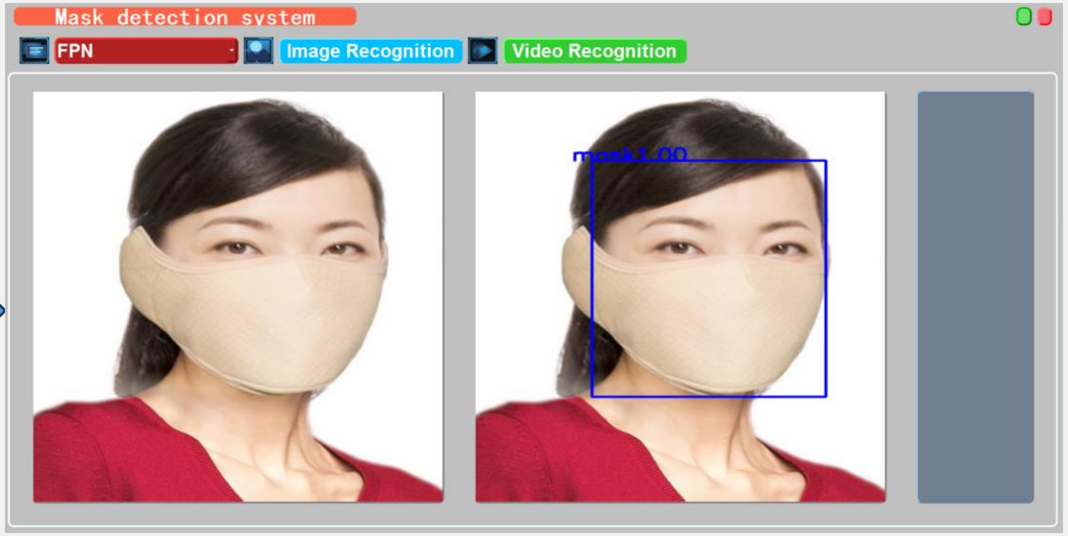


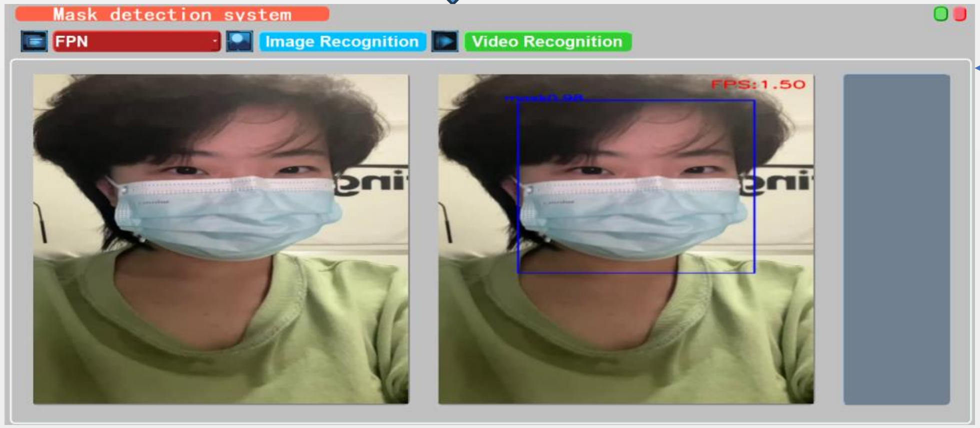
RESULTS



RESULTS







Low video frame

CONCLUSION

OXFORD BROOKES UNIVERSITY

Limitation

- Model training is resource intensive.
- The dataset categories are not balanced.
- Mis-wearing mask images are mostly PS processed.

Future Work

- Utilizing GANs to generate diverse training data
- Collect more mask images from real scenes
- Introduce attention mechanism

REFERENCE

OXFORD BROOKES UNIVERSITY

[1] B. Qin and D. Li, ‘Identifying Facemask-Wearing Condition Using Image Super-Resolution with Classification Network to Prevent COVID-19’, Sensors, vol. 20, no. 18, p. 5236, Sep. 2020, doi: 10.3390/s20185236.

[2] M. Xu, H. Wang, S. Yang, and R. Li, ‘Mask wearing detection method based on SSD-Mask algorithm’, in 2020 International Conference on Computer Science and Management Technology (ICCSMT), Shanghai, China: IEEE, Nov. 2020, pp. 138–143. doi: 10.1109/ICCSMT51754.2020.00034.

[3] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, ‘RetinaFace: Single-Shot Multi-Level Face Localisation in the Wild’, in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA: IEEE, Jun. 2020, pp. 5202–5211. doi: 10.1109/CVPR42600.2020.00525.

[4] M. Inamdar and N. Mehendale, ‘Real-Time Face Mask Identification Using Facemasknet Deep Learning Network’, SSRN Journal, 2020, doi: 10.2139/ssrn.3663305.

[5] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, ‘Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection’, Sustainable Cities and Society, vol. 65, p. 102600, Feb. 2021, doi: 10.1016/j.scs.2020.102600.

[6] MaskedFace-Net, March. 2024, [online] Available: https://github.com/cabani/MaskedFace-Net.