Deep Learning in IoT and Edge Computing for Safety Applications

Wilbur Kai Heng LUA
Department of Computing Science
University of Glasgow
United Kingdom
2508444L@glasgow.ac.uk

Peter ChunYu YAU
Department of Computing Science
University of Glasgow
United Kingdom
PeterCY.Yau@glasgow.ac.uk
0000-0001-8431-2770

Chee Kiat SEOW
Department of Computing Science
University of Glasgow
United Kingdom
CheeKiat.Seow@glasgow.ac.uk
0000-0002-6499-9410

Dennis WONG
Faculty of Applied Sciences
Macao Polytechnic University
Macau, China
cwong@mpu.edu.mo
0000-0002-6242-164X

***Abstract*— The COVID-19 pandemic has resulted in the mandate of public safety measures such as face mask-wearing. This paper provides an overview of using artificial intelligence (AI), computer vision and the Internet of Things (IoT) to implement face-mask detection systems as a public health safety solution. This paper reviews how edge computing can overcome traditional cloud computing issues. The report also examines the current state of computer vision, convolutional neural networks and their potential in the public health and safety domain. Finally, we will recommend a methodology that leverages Artificial Intelligence of Things (IoT) and Computer vision to build an intelligent edge system for the public health domain.**

Index Terms — Internet-of-Things, IoT, Edge Computing, Computer Vision, Deep Learning, Low Cost, Medical Facility, Hospital, Mask, Public Health, Public Safety, COVID-19

I. Introduction

According to World Health Organization (WHO), in 2022, there will be 520 million confirmed cases, causing 6.2 million deaths worldwide [1]. Research studies by the Infectious diseases society of America have shown that countries with stringent regulations have seen drastically reduced numbers of local community transmissions [2]. Governments have mandated mandatory facial mask-wearing rules to combat and contain COVID-19 (SARS-CoV-2) respiratory virus transmission. In this paper, we will discuss a framework for how Artificial Intelligence (AI) and deep learning techniques such as convolutional neural networks and Edge computing techniques can be used to deploy an IoT solution for face-mask detection solutions, such as unlocking the mobile devices. We aim to evaluate the performances of different algorithms and the architecture design with various hardware. This paper serves as a foundation overview of the project execution.

II. Literature Review

A. Core Technology

At the core of the solution, deep learning algorithms such as Convolutional Neural Network (CNN) and deep learning frameworks such as Keras, PyTorch or TensorFlow remain the popular implementation tools and cornerstones for computer vision and object detection projects. The adoption of CNNs proliferated when AlexNet [3]won the ImageNet Challenge (LSVRC-2012), outperforming other detection methods in terms of both speed and accuracy. Since then, other popular neural networks such as VGG-16 [4], VGG-19, Inception[5] and ResNet [6] have surfaced, and deep neural networks have achieved better accuracy on image classification and object detection tasks on large-scale datasets such as MNIST, MS-COCO and CIFAR-10. CNN remains the most popular deep learning method used to extract features from unstructured data such as images, video, audio, and textual documents. These CNN models also serve as backbone feature extractors for complex computer vision tasks such as object detection models.

B. Applications of Computer Vision for Public Safety

a) Third-generation surveillance system in Smart City: Computer vision and video analytics of surveillance systems for Smart city applications have been ongoing for the past few decades, and it remains one of the key areas of research. Tomi D. Raty has termed current surveillance technology systems as the third-generation surveillance system (3GSS) [7]To improve from previous generations, existing surveillance systems are more location-aware, scalable, and distributive. Importantly, with AI/ML technologies advancing, some 3GSS systems also have video analytics capabilities and are context-aware.

Computer vision and artificial intelligence techniques such as object detection continue to be active research fields to bring intelligence to current traditional passive surveillance. A typical architecture of current 3GSS systems comprises of a machine or server which uses data collected from data sources such as surveillance cameras or sensors to perform object detection, behavioural, or scene-based analysis to detect any safety or health risk in the location. With the increased deployments of Close-circuit television (CCTV) in the community, there will continue to be a substantial increase in the adoption of computer vision-based solutions for smart-city applications. Public safety contains a broad spectrum of plausible applications.

Zhang et al. listed some public safety applications that have benefitted from video surveillance and analytics, such as policing, where video surveillance captures suspicious

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

activities or people in the community, to transportation and emergency medical services (EMS) applications [8]

b) Applications in Intelligent traffic systems: Buch et al. reviewed the current computer vision techniques for Urban surveillance such as smart traffic and pedestrian monitoring and noted that these are fast-emerging areas of research to develop safety-related solutions as part of an intelligent traffic system (ITS). [9] The ITS covers a wide area of applications such as detection of traffic violations, pedestrian safety and visual inspection for traffic or crowd control. Computer vision techniques such as Optical Character Recognition (OCR) are used to detect and analyze vehicle license registration plates either for parking, access control or traffic violation purposes. Real-time detection systems have also been researched for accident detection for traffic surveillance purposes.[10]

c) Application in Emergency Medical Centers and Hospitals: The COVID-19 pandemic has resulted in shortages of resources and manpower. With the increased number of patients requiring medical attention and close observations, a closed-loop monitoring system can be deployed to assist and alleviate monitoring duties and improve the overall efficiency of caretakers and medical staff. The patient monitoring system is also used to detect any accidents and abnormal movements of the patient and alert the medical staff to check on the patient. Kittipanya-Ngam et al. researched the implementation of a fall detection system that uses computer vision and deep learning capabilities which can be deployed at CCTV blind spots or stairs to detect if a patient or elderly has fallen and alert the medical staff to render immediate medical attention [11]

d) Applications for Public safety COVID-19 measures: Since the start of the COVID-19 pandemic, research and literature on computer vision and deep learning solutions for public safety saw a sudden spike, notably in facial mask detector systems and social distancing surveillances. The solutions proposed were mostly built using CV, Deep Neural Networks (DNN) and Object detector models. Singh et al. proposed a Face mask detection system using YOLOv3 and a region-based proposal network, R-CNN.[12] Jignesh et al. proposed using transfer learning on the InceptionV3 model to develop a facial mask detection (FMD) system. [13] Hou et al. proposed a social distancing detection system that identifies pedestrians and calculates if safe distancing measures were complied with. The solution was built using the YOLOv3 object detection model. [14]

C. Trends in Video Surveillance and Analytics

a) Cloud Computing Challenges: The proliferation of the Internet of Things (IoT) and Cyber-Physical Systems has caused a massive increase in data and information produced. By 2025, it is forecasted that 30.9 billion IoT devices will be connected to the internet, and these devices will contribute up to 79.4 Zettabytes (ZB) of data.[15] Cloud computing infrastructure is insufficient to deal with such a massive volume of data, and such usage causes bottlenecks in the network bandwidth. The system would require a constant high-speed connection to the cloud.[16]

Furthermore, it is inadvisable to use cloud computing for video surveillance and analytics of public safety as it deals with private data such as people's physical features. Pushing these sensitive data through the network to the cloud exposes it to potential cybersecurity attacks such as identity theft and data breaches. Additionally, public safety applications have certain level timing constraints. Using a cloud architecture, the system would be affected by network connectivity issues or low-network bandwidth.

b) Edge Computing: The Edge computing paradigm introduces means a new means of data processing; In an edge computing architecture, edge nodes are placed in close proximity to the sensor device to reduce inefficient network bandwidth utilisation. It improves the system's response time by bringing down the overall latency, and fulfils privacy preservation requirements.

Deep learning for Edge Computing and IoT architectures has shown to be effective in reducing data processing requirements and relieving the pressure of the cloud. [18], [19] However, DNNs in IoT devices have their own sets of challenges as these devices are resource-constrained, and traditional CNN models will not be able to run on such devices. It is important to manage the trade-off of lightweight models, accuracy and real-time requirement constraints. Techniques such as network pruning and quantization reduce the overall size of the model by shrinking and removing connections. Using embedded GPU devices such as NVIDIA Jetson TX1 and Nano has also been shown to improve model performance on the edge. [20]

III. Study Approach

This study aims to experiment with a practical and effective approach for Face Mask Detection by using Computer vision and DNNs. Evaluation will be done on different state of the arts single-shot object detectors such as YOLOv5 and EfficientDet.

We will also leverage techniques to optimise networks so that the trained models can be deployed onto resource-constrained IoT devices. We have also chosen the NVIDIA Jetson Nano as the intended edge device for deployment. Convolutional operations and matrix arithmetic are expensive operations that demand extensive computational powers; as such, the NVIDIA Jetson Nano ecosystem is preferred over similar single-board embedded devices like Raspberry Pi 3B+/4 (RPI) as the Graphics Processor (GPU) uses NVIDIA CUDA cores as compared to a CPU centric device like RPI, which enable the device to perform simultaneous parallel computations and high-performance GPUs are preferred for DNNs as they allow a device to achieve better performance overall. [21]

Convolutional neural networks (CNNs) function as backbone feature extractors of object detector models. Traditional CNNs such as ResNet and VGG-16 have relatively high layers resulting in high computational parameters and model complexity. Furthermore, the model size is often too large to be able to run on edge devices. With the proliferation of IoT as well as deep learning in IoT devices, there is currently a rising trend in research towards the construction of computationally efficient lightweight models with significantly fewer parameters whilst still achieving comparable accuracies. Models such as MobileNet[22] and EfficientNet [23]family developed by

Fig 1: samples of correctly masked faces (dataset CMFD), and samples of incorrectly masked faces (dataset IMFD)
Source: Masked Face Net [28]

Google utilizes novel techniques such as depth-wise convolutions and compound scaling to produce state-of-the-art lightweight models that are leaner and have lower power consumption which is extremely good for resourced-constrained embedded devices. These models will be heavily explored in the research as suitable candidates for feature extractors in the backbone of object detectors.

The choice of DNN models also impacts the performance of the overall solution. Object detection models can be split into two different categories: 1) *Two-stage detectors*, such as R-CNN[24] and Faster R-CNN[25] are Region Proposal Networks (RPN) that will first run a region proposal stage to extract regions of interest (ROI) before performing classification and detection on the ROI. 2) *Single-shot detectors,* such as the YOLO object detection family and Single Shot Detector (SSD) [26], [27] treat the object detection process as a regression problem, and the model will perform localisation, classification and detection of multiple objects in the frame within one forward propagation of the neural network.

Benchmark studies have revealed that the Single-shot detectors of the YOLOv4 variant obtained similar accuracies and average precision (AP) compared to Faster R-CNN while achieving 2-3 times higher performance in terms of FPS. A survey on existing COVID-19 Face Mask detection systems showed that Single-stage detectors like YOLOv2, YOLOv3, and Single-shot detector (SSD) architectures were predominantly used as the chosen architecture. In the case of the study approach, we will focus on developing solutions using the Single-shot framework as it has been proven to have a higher performance, which is an important area of consideration for edge deployment scenarios.

Prior to the emergence of COVID-19, there literature and data set volume of masked individuals were low as most of the images captured were meant for human face detection such as VGG Face and Microsoft Celebrity Face dataset. Currently, there are several open-source datasets publicly available for deep learning training such as Kaggle's Face mask detection dataset, the Real-world masked face dataset (RMFD) as well as the MaskedFaceNet. [28] These datasets will allow the model to be trained and exposed to sufficiently diverse data.

A. *Measurement metrics used to evaluate model accuracy*

The detection model will be evaluated using metrics such as precision, model recall, F1 score, mean average precision score (mAP) as well as the intersection over union (IoU) score.

Metrics of model accuracy measurement

$$Precision = \frac{TP}{(TP+FP)}$$

1) *Precision*

The precision of the model will calculate the probability that the predicted bounding box outputted by the model matches the actual ground truth bounding box. It refers to the positive predictive value.

$$Recall = \frac{TP}{(TP+FN)}$$

2) *Recall*

The recall metric will assess the sensitivity or truth positive rates of the model, the higher the recall means that the model manages to successfully predict correct values.

$$F1\ score = 2 * \frac{(Precision * Recall)}{Precision + Recall}$$

3) *F1 Score*

(harmonic mean between precision and recall)

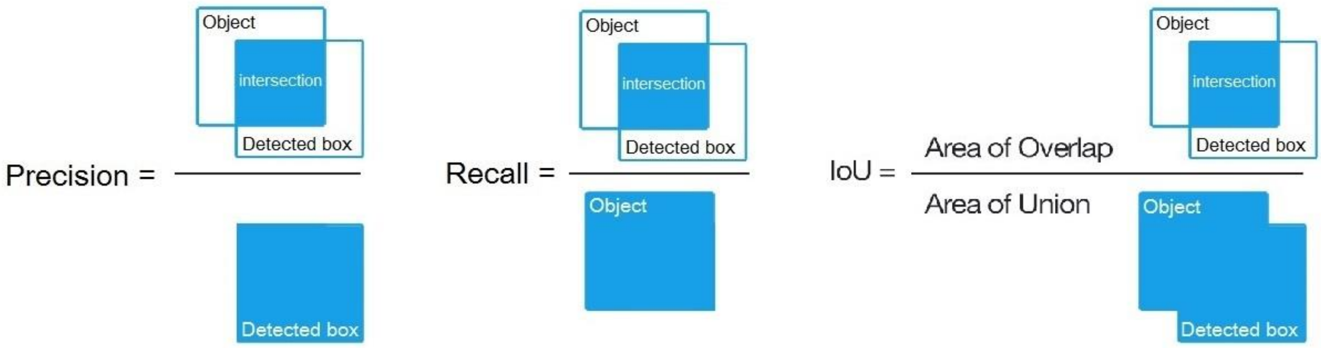

Fig 2: Precision, Recall and Intersection over Union (IoU).
Source [29]

The F1 score is an important function that factors in both precision and recall. F1 analyses the balance of both precision and recall and the possibility of uneven class distribution. Additionally, the function considers both precision and recall and prevents instances where the model accuracy is high but only for certain classes such as true negatives (TN). For example, the model could have a high number of true negatives but is unable to detect actual positive cases.

4) Mean Average Precision (mAP) =

$$\sum_{k=0}^{k=n-1} [Recalls(k) - Recalls(k+1)] * Precisions(k)$$

AP metric allows us to summarize the precision and recall curve into one singular value representing the average precision of all precisions.

5) Intersection over Union

Intersection over union (IoU) calculates how well the predicted bounding box matches the ground truth bounding box. Typically, the IoU, is done by calculating the intersect sections over the total union of both bounding boxes. For the study, the IoU threshold will be set to be 0.5.

IV. DISCUSSION

Computer vision and deep learning applications continue to be an emerging area of research for public safety applications. Following the discussion, a COVID-19 Face Mask detection will be built and deployed on the edge with deep learning. Embedded low-cost and small single-board devices such as the NVIDIA Jetson ecosystem are now equipped with powerful GPU that can be utilised to do parallel computing to speed up AI and deep learning on edge devices. Furthermore, a rising trend of deep learning applications and being research to be deployed on edge for computational offloading to bring down the overall latency of the system. Several state-of-the-art, efficient models such as SqueezeNet, EfficientNet and MobileNet are lightweight enough to run on these small resource-constrained devices. This study aims to leverage existing open-source facial mask datasets such as Kaggle FMD, and MaskedFaceNet to train the neural network.

The study will also utilise lightweight CNNs and Single-shot detectors such as YOLOv5 variants and SSD to train and deploy a COVID-19 Face mask detection system capable of being deployed on an NVIDIA Jetson Nano kit. The paper will also present the methodology and performance evaluation of the trained models to contribute to the current literature on future public safety applications.

ACKNOWLEDGMENT

This research is supported by the Macao Polytechnic University research grant (Project code: RP/FCA-02/2022). The research of the fourth author is also supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Science and ICT (MSIT), Korea (No. 2020R1F1A1A01070666).

REFERENCES

[1] “WHO Coronavirus (COVID-19) Dashboard.”

[2] R. v. Tso and B. J. Cowling, “Importance of Face Masks for COVID-19: A Call for Effective Public Education,” *Clinical Infectious Diseases*, vol. 71, no. 16. Oxford University Press, pp. 2195–2198, Oct. 15, 2020. doi: 10.1093/cid/ciaa593.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks.” [Online]. Available: http://code.google.com/p/cuda-convnet/

[4] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” Sep. 2014, [Online]. Available: http://arxiv.org/abs/1409.1556

[5] C. Szegedy, V. Vanhoucke, S. Ioffe, and J. Shlens, “Rethinking the Inception Architecture for Computer Vision.”

[6] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” Dec. 2015, [Online]. Available: http://arxiv.org/abs/1512.03385

[7] T. D. Räty, “Survey on contemporary remote surveillance systems for public safety,” *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 40, no. 5, pp. 493–515, Sep. 2010, doi: 10.1109/TSMCC.2010.2042446.

[8] Q. Zhang, H. Sun, X. Wu, and H. Zhong, “Edge video analytics for public safety: A review,”

Proceedings of the IEEE, vol. 107, no. 8, pp. 1675–1696, Aug. 2019, doi: 10.1109/JPROC.2019.2925910.

[9] N. Buch, S. A. Velastin, and J. Orwell, “A review of computer vision techniques for the analysis of urban traffic,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 3. pp. 920–939, Sep. 2011. doi: 10.1109/TITS.2011.2119372.

[10] E. P. Ijjina, D. Chand, S. Gupta, K. Goutham, and B. Tech, “Computer Vision-based Accident Detection in Traffic Surveillance; Computer Vision-based Accident Detection in Traffic Surveillance,” 2019.

[11] P. Kittipanya-Ngam and O. Soh Guat, “Computer Vision Applications for Patients Monitoring System,” 2012.

[12] S. Singh, U. Ahuja, M. Kumar, K. Kumar, and M. Sachdeva, “Face mask detection using YOLOv3 and faster R-CNN models: COVID-19 environment,” *Multimedia Tools and Applications*, vol. 80, no. 13, pp. 19753–19768, May 2021, doi: 10.1007/s11042-021-10711-8.

[13] G. J. Chowdary, N. S. Punn, S. K. Sonbhadra, and S. Agarwal, “Face Mask Detection using Transfer Learning of InceptionV3,” Sep. 2020, doi: 10.1007/978-3-030-66665-1_6.

[14] Y. C. Hou, M. Z. Baharuddin, S. Yussof, and S. Dzulkifly, “Social Distancing Detection with Deep Learning Model,” in *2020 8th International Conference on Information Technology and Multimedia, ICIMU 2020*, Aug. 2020, pp. 334–338. doi: 10.1109/ICIMU49871.2020.9243478.

[15] Lionel Sujay Vailshery, “Internet of Things (IoT) - statistics & facts,” *statista*, May 11, 2021.

[16] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, “Edge Computing: Vision and Challenges,” *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 637–646, Oct. 2016, doi: 10.1109/JIOT.2016.2579198.

[17] Chiradeep BasuMallick, “Top 10 Edge Computing Platforms in 2022,” Feb. 08, 2022.

[18] J. Chen and X. Ran, “Deep Learning With Edge Computing: A Review,” *Proceedings of the IEEE*, 2019, doi: 10.1109/JPROC.2019.2921977.

[19] H. Li, K. Ota, and M. Dong, “Learning IoT in Edge: Deep Learning for the Internet of Things with Edge Computing,” *IEEE Network*, vol. 32, no. 1, pp. 96–101, Jan. 2018, doi: 10.1109/MNET.2018.1700202.

[20] A. Marchisio *et al.*, “Deep Learning for Edge Computing: Current Trends, Cross-Layer Optimizations, and Open Research Challenges,” in *Proceedings of IEEE Computer Society Annual Symposium on VLSI, ISVLSI*, Jul. 2019, vol. 2019-July, pp. 553–559. doi: 10.1109/ISVLSI.2019.00105.

[21] Institute of Electrical and Electronics Engineers. Turkey Section. and Institute of Electrical and Electronics Engineers, *HORA 2020 : 2nd International Congress on Human-Computer Interaction, Optimization and Robotic Applications : proceedings : June 26-27, 2020, Turkey*.

[22] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” Jan. 2018, [Online]. Available: http://arxiv.org/abs/1801.04381

[23] M. Tan and Q. v. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” May 2019, [Online]. Available: http://arxiv.org/abs/1905.11946

[24] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” Nov. 2013, [Online]. Available: http://arxiv.org/abs/1311.2524

[25] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” Jun. 2015, [Online]. Available: http://arxiv.org/abs/1506.01497

[26] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” Jun. 2015, [Online]. Available: http://arxiv.org/abs/1506.02640

[27] W. Liu *et al.*, “SSD: Single Shot MultiBox Detector,” Dec. 2015, doi: 10.1007/978-3-319-46448-0_2.

[28] A. Cabani, K. Hammoudi, H. Benhabiles, and M. Melkemi, “MaskedFace-Net – A dataset of correctly/incorrectly masked face images in the context of COVID-19,” *Smart Health*, vol. 19, Mar. 2021, doi: 10.1016/j.smhl.2020.100144.

[29] R. Padilla, S. L. Netto, E. A. B. da Silva, and S. L. Netto, “A Survey on Performance Metrics for Object-Detection Algorithms Energy Conservation in Wireless Sensor Networks for IoT Applications View project Light Fields Compression View project A Survey on Performance Metrics for Object-Detection Algorithms”, doi: 10.1109/IWSSIP48289.2020.