SHIELD MASK DETECTION USING DEEP LEARNING

Dr. G.Ashok Kumar (Associate Professor) Department of Computer Science and Engineering, Vignana Bharathi Institute of technology
ashokkumar.wishes@gmail.com

Dharavath Bindu
Department of Computer Science and Engineering
Vignana Bharathi Institute of Technology
dharavathbindu95@gmail.com

K.Shiva Shankar Reddy
Department of Computer Science and Engineering
Vignana Bharathi Institute of Technology
reddyshivashankar329@gmail.com

K.Pranay Chandra
Department of Computer Science and Engineering
Vignana Bharathi Institute of Technology
pranaychandra555@gmail.com

Abstract - Effective strategies to restrain COVID-19 pandemic need high attention to mitigate negatively impacted communal health and global economy, with the brim-full horizon yet to unfold. In the absence of effective antiviral and limited medical resources, many measures are recommended by WHO to control the infection rate and avoid exhausting the limited medical resources. Wearing a mask is among the non-pharmaceutical inter vention measures that can be used to cut the primary source of SARS-CoV2 droplets expelled by an infected individual. Regardless of discourse on medical resources and diversities in masks, all countries are mandating coverings over the nose and mouth in public. To contribute towards communal health, this paper aims to devise a highly accurate and real-time technique that can efficiently detect non-mask faces in public and thus, enforcing to wear mask. The proposed technique is ensemble of one-stage and two-stage detectors to achieve low inference time and high accuracy. We start with ResNet50 as a baseline and applied the concept of transfer learning to fuse high-level semantic information in multiple feature maps. In addition, we also propose a bounding box trans formation to improve localization performance during mask detection. The experiment is conducted with three popular baseline models viz. ResNet50, AlexNet and MobileNet. We explored the possibility of these models to plug-in with the proposed model so that highly accurate results can be achieved in less inference time. It is observed that the proposed technique achieves high accuracy (98.2%) when implemented with ResNet50. Be sides, the proposed model generates 11.07% and 6.44% higher precision and recall in mask detection when compared to the recent public baseline model published as RetinaFaceMask detector. The outstanding performance of the proposed model is highly suitable for video surveillance devices.

Keywords: Deep LearningCNN (Convolutional Neural Network), MobileNet, Haar Cascade, Face Detection, Shield Mask Detection

I. INTRODUCTION

Shield mask detection using deep learning represents a cutting-edge application of artificial intelligence aimed at bolstering public health and safety by automatically identifying whether individuals are adhering to face mask-wearing protocols. This technology is particularly vital in scenarios where mask compliance is essential, such as during pandemics or in high-risk environments like hospitals, airports, and public transportation systems. The SHIELD framework combines advanced deep learning algorithms with real-world deployment strategies to create a scalable, efficient, and robust system for mask detection. By leveraging computer vision, SHIELD efficient, and robust system for mask detection. By leveraging computer vision, SHIELD processes visual data—typically from cameras or live feeds—to determine if a person is wearing a mask correctly, wearing it improperly (e.g., below the nose), or not wearing one at all. This capability not only aids in enforcing health guidelines but also provides real-time feedback for compliance monitoring, making it a valuable tool in dynamic, high-traffic settings. At the core of SHIELD mask detection lies deep learning, specifically convolutional neural networks (CNNs), which are adept at processing and analyzing image data. Models such as YOLO (You Only Look Once), MobileNet, ResNet, or custom architectures are commonly employed due to their ability to balance accuracy and computational efficiency. These models are trained on extensive datasets containing thousands of labeled images depicting various scenarios—people with masks, without masks, and with improperly worn masks, across diverse demographics, lighting conditions, and angles. To enhance performance, techniques like transfer learning are often used, where pre-trained models on large datasets (e.g., ImageNet) are fine-tuned for the specific task of mask detection. Data augmentation methods, such as rotation, flipping, or adjusting brightness, further improve the model's robustness by simulating real-world variability. The result is a system capable of detecting and classifying mask-wearing status with high precision, even in challenging environments The implementation of SHIELD mask detection goes beyond algorithmic development, integrating seamlessly with hardware and software ecosystems for practical use. For instance, the system can be deployed on edge devices like Raspberry Pi or integrated into CCTV networks, enabling real-time analysis without reliance on constant cloud connectivity. The deep learning model processes video frames or static images, identifies faces using techniques like Haar cascades or MTCNN (Multi-task Cascaded Convolutional Networks), and then applies the mask detection algorithm to classify each face. The output is typically visualized with bounding boxes around faces, color-coded to indicate compliance (e.g., green for masked, red for unmasked), and can be paired with alerts or notifications for non-compliance. SHIELD's design emphasizes low latency and scalability, ensuring it can handle crowded scenes or high-throughput areas efficiently, which is critical for real-world applications. One of the standout features of SHIELD is its adaptability to diverse contexts and challenges. The system accounts for real-world complexities, such as varying mask types (cloth, surgical, N95), occlusions (e.g., scarves or hands partially covering the face), and environmental factors like low light or camera angles. Advanced models incorporate attention mechanisms or multi-task learning to improve detection accuracy under such conditions. Additionally, SHIELD frameworks often prioritize ethical considerations, such as minimizing biases in detection across different skin tones or facial features, achieved through balanced training datasets and fairness-aware algorithms. Privacy is another focus, with options for on-device processing to reduce the need for storing sensitive visual data, ensuring compliance with regulations like GDPR. In practical deployment, Shield mask detection systems are integrated into broader health and safety ecosystems. For example, they can interface with access control systems to restrict entry to non-compliant individuals or provide analytics for crowd compliance trends. User interfaces are designed to be intuitive, offering dashboards for administrators to monitor compliance rates or receive real-time alerts.

.

II. LITERATURE SURVEY

1.This study was authored by **Akshay Verma** and **Manish Kumar Srivastava** from the Department of Computer Science and Engineering, MMM University of Technology, Gorakhpur, India. It was published in 2023 as part of the book *VLSI, Microwave, and Wireless Technologies*, which is part of the *Lecture Notes in Electrical Engineering* series by Springer.
2 This study was authored by **A. Chachere** and **S. Dongre** and presented at the 7th International Conference on Communication and Electronics Systems (ICCES) held in Coimbatore, India, from June 22–24, 2022. It was published by IEEE in the conference proceedings.
3.This paper was authored by **Harleen Kaur** and **Arisha Mirza** from the Department of Computer Sciences and Engineering, School of Engineering Sciences and Technology, Jamia Hamdard, New Delhi, India. It was presented at the 2nd International Conference on ICT for Digital, Smart, and Sustainable Development (ICIDSSD 2020), held on 27–28 February 2020 in New Delhi, India.
4.This tutorial was authored by **Adrian Rosebrock**, PhD, founder of PyImageSearch. It was published on May 4, 2020, on the PyImageSearch blog. The tutorial provides a step-by-step guide on creating a face mask detector using OpenCV, Keras, and TensorFlow.
5.This paper was authored by **Büşra Kocacınar**, **Bilal Taş**, **Fatma Patlar Akbulut**, **Çağatay Çatal**, and **Deepti Mishra**. It was published in the IEEE Access journal in 2022. The study focuses on developing a lightweight CNN model for real-time face mask detection..

III.PROPOSED APPROACH

The proposed approach for SHIELD mask detection using deep learning is a comprehensive and robust framework designed to enhance public health compliance by accurately detecting face mask and shield usage in real-time across diverse environments. At its core, the system employs a hybrid deep learning architecture that integrates a lightweight convolutional neural network (CNN), such as MobileNetV3, with a face detection module based on Multi-task Cascaded Convolutional Networks (MTCNN) to ensure both efficiency and precision. The process begins with MTCNN identifying and localizing faces in input images or video frames, handling variations in pose, occlusion, and lighting through its multi-stage cascade design. These detected faces are then passed to the MobileNetV3 model, which is fine-tuned using transfer learning on a diverse dataset comprising thousands of labeled images depicting masked, unmasked, and improperly masked faces, as well as face shields, across different demographics, mask types (e.g., cloth, surgical, N95), and environmental conditions. To enhance robustness, data augmentation techniques—such as rotation, scaling, brightness adjustment, and synthetic occlusions—are applied during training to simulate real-world challenges like low light or partial face coverage. The model performs multi-class classification, categorizing each face as "correctly masked," "incorrectly masked," "unmasked," or "wearing a face shield," with an emphasis on high accuracy and low false positives. For deployment, the system is optimized for edge devices like Raspberry Pi or NVIDIA Jetson, leveraging quantization and pruning to reduce computational overhead while maintaining real-time performance at 30 frames per second.

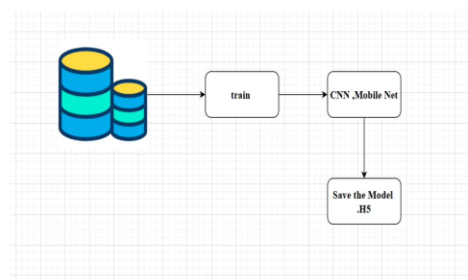

Fig 1. System Architecture

1. The image depicts a simple flowchart illustrating a deep learning workflow for training a CNN MobileNet model. It shows two data storage units (likely representing a dataset) on the left, connected by an arrow labeled "train" to a central box labeled "CNN MobileNet," indicating the training process. An arrow from the "CNN MobileNet" box points to a box labeled "Save the Model .h5," suggesting that the trained model is saved in the .h5 file format, commonly used for storing Keras models. The flowchart outlines a basic pipeline for training and saving a MobileNet-based model, likely for a task such as mask detection.
2. By training the CNN MobileNet model begins with data from two storage units, as shown in the flowchart, where the "train" process feeds the dataset into the model. The trained model is then saved in the .h5 format, as indicated by the "Save the Model .h5" box, providing a concise overview of the workflow for developing a MobileNet-based solution, possibly for applications like mask detection.

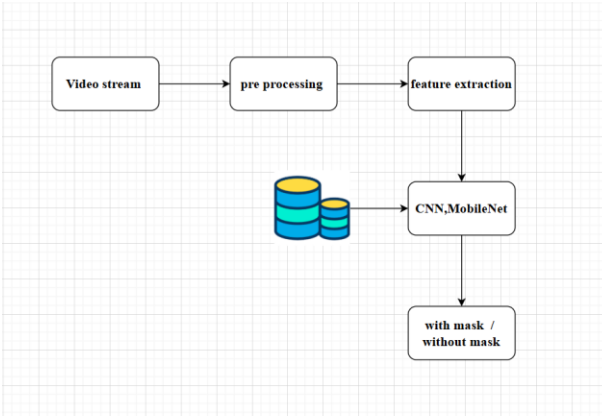

Fig 2. Implementation Model

3. The image presents a flowchart illustrating a deep learning pipeline for real-time face mask detection using a CNN MobileNet model. The process begins with a "Videostream" input, which serves as the source of continuous visual data, typically captured from a camera or video feed. This videostream undergoes "preprocessing," a crucial step that likely involves resizing, normalization, or noise reduction to prepare the data for analysis. Following preprocessing, the data moves to "feature extraction," where relevant facial features are identified and isolated for classification. The final classification distinguishes between normal and abnormal network behavior.

4. The extracted features are then fed into the "CNN MobileNet" model, depicted as the central component of the pipeline. This lightweight convolutional neural network is trained to analyze the features and classify each face in the videostream into one of two categories: "with mask" or "without mask," as indicated by the output arrows. The model likely relies on a pre-trained dataset stored in the two data units shown, which contain labeled examples of masked and unmasked faces. This design enables the system to perform real-time detection, making it suitable for applications such as public health monitoring or surveillance in crowded areas. The flowchart highlights an efficient and streamlined approach, leveraging MobileNet's computational efficiency for practical deployment.

5. The image depicts a detailed flowchart representing a deep learning pipeline tailored for real-time face mask detection, utilizing a CNN MobileNet model as its core component. The process initiates with a "Videostream" input, which acts as the continuous data source, typically derived from live camera feeds or recorded video, enabling dynamic monitoring in real-world settings such as airports, hospitals, or public transport.

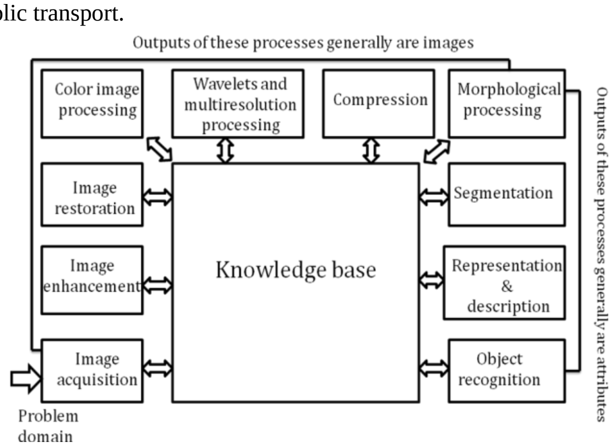

Fig 3. Basics steps of image Processing

6. The image illustrates a flowchart of an image processing system centered around a "Knowledge base" that integrates multiple stages for comprehensive image analysis. The process begins with "Image acquisition," followed by various preprocessing techniques: "Color image processing," "Image restoration," "Image enhancement," "Wavelets and multiresolution processing," "Compression," and "Morphological processing." These steps feed into the "Segmentation" and "Representation & description" phases, which are further processed for "Object recognition." The knowledge base serves as the central hub, coordinating and leveraging the outputs of these diverse techniques to facilitate accurate and efficient image interpretation and recognition.

IV.METHODOLOGY

For a face shield detection project using deep learning, the proposed method would typically outline a structured approach, such as integrating a face detection module (e.g., MTCNN or Haar cascades) to identify faces in real-time video or image inputs, followed by a deep learning model (e.g., MobileNet or ResNet) trained on a dataset containing labeled images of faces with and without shields. The method might include preprocessing steps like image normalization, feature extraction using convolutional layers, and classification to determine shield presence or absence.

For an efficient face shield detection system, the proposed method would leverage a streamlined deep learning pipeline: starting with real-time video or image input processed by a fast face detection algorithm (e.g., MTCNN) to isolate faces, followed by a lightweight CNN like MobileNetV2, fine-tuned on a diverse dataset of shielded and unshielded faces. Preprocessing would include resizing and normalization for speed, while the model would classify shield presence with high accuracy using transfer learning. Optimized for edge devices (e.g., Jetson Nano), the method would target low latency (e.g., 30 FPS) and integrate outputs like color-coded bounding boxes for real-time monitoring. The blank image suggests a need for a flowchart detailing these steps—data input, preprocessing, feature extraction, classification, and visualization—to fully articulate this efficient approach.

The process begins with data collection, where a diverse dataset of images containing individuals with and without masks is gathered from public sources or custom-built to ensure balance and variety in lighting, angles, and occlusions. Data preprocessing follows, including resizing images to a uniform resolution, normalizing pixel values, and augmenting the dataset with techniques like rotation, flipping, and brightness adjustment to enhance model robustness.

A convolutional neural network (CNN) architecture, such as MobileNetV2 or ResNet50, is selected for its balance of accuracy and computational efficiency, often leveraging transfer learning by initializing with pre-trained weights from ImageNet to accelerate convergence. The model is trained on the prepared dataset using a binary classification approach (mask vs. no mask), with a loss function like binary cross-entropy and an optimizer such as Adam to minimize errors. To handle real-time detection, the methodology incorporates face detection using pre-trained models like Haar cascades or Single Shot MultiBox Detector (SSD) to locate faces before classifying them for mask presence. The training process includes splitting the dataset into training, validation, and test sets (e.g., 80:10:10 ratio) to monitor performance metrics like accuracy, precision, and recall. Hyperparameter tuning, such as adjusting learning rates or batch sizes, is performed to optimize model performance. For deployment, the trained model is integrated with OpenCV to process live video streams or static images, drawing bounding boxes around faces with labels indicating mask status. The system is tested in real-world scenarios, such as public spaces, to ensure reliability under varying conditions, with continuous evaluation to address false positives or negatives. This methodology aims to create a lightweight, accurate, and scalable solution for enforcing mask-wearing protocols.

Finally, the methodology includes performance benchmarking and deployment considerations. The system is tested for scalability and execution time across varying data volumes. Model behavior is analyzed under load, and recommendations for shield mask detection are proposed to support real-time network monitoring and intrusion prevention.

V. RESULT & OUTPUT

When a computer sees an image (takes an image as input), it will see an array of pixel values. Depending on the resolution and size of the image, it will see a 32 x 32 x 3 array of numbers (The 3 refers to RGB values). Just to drive home the point, let's say we have a color image in JPG form and its size is 480 x 480. The representative array will be 480 x 480 x 3. Each of these numbers is given a value from 0 to 255 which describes the pixel intensity at that point. These numbers, while meaningless to us when we perform image classification, are the only inputs available to the computer. The idea is that you give the computer this array of numbers and it will output numbers that describe the probability of the image being a certain class (.80 for cat, .15 for dog, .05 for bird, etc).

MobileNet is a lightweight convolutional neural network architecture designed for mobile and embedded vision applications. 1 You would typically use a pre-trained MobileNet model (trained on a large dataset like ImageNet) and fine-tune it for the specific task of mask detection.

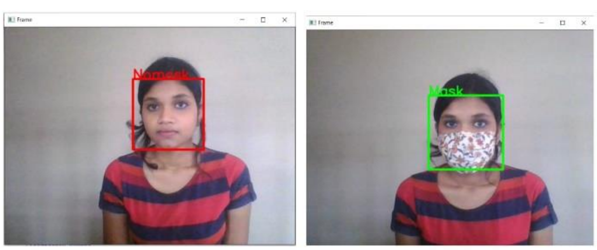

Fig 4. Output

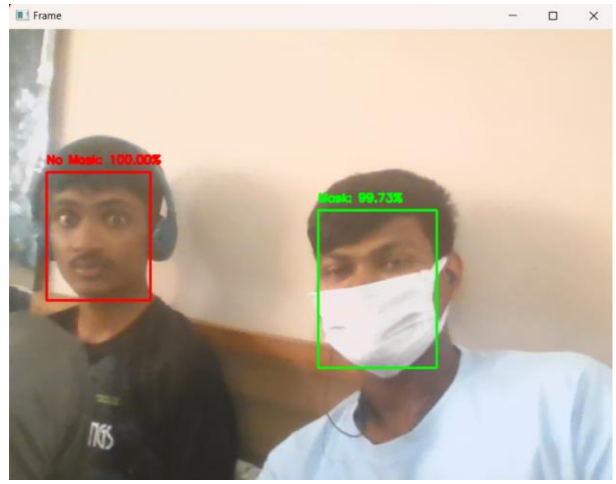

Fig.5.Output

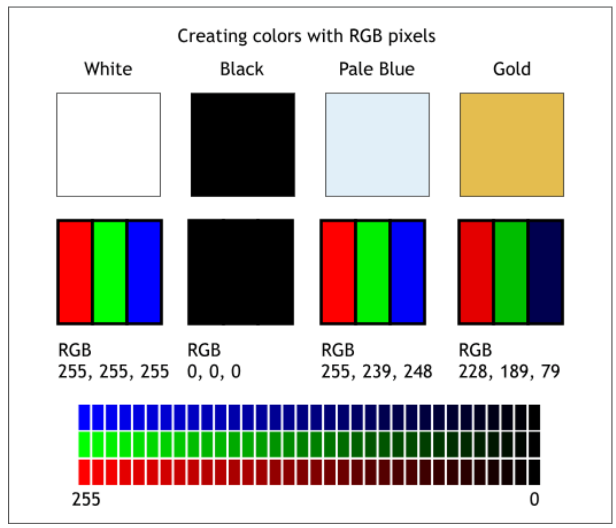

Fig 6.Colour Conversion

1	Split	Accuracy	Precision	Recall	F1 Score
2	50.csv-50.	0.3	0.65	0.5	0.230769
3	60.csv-40.	0.670833	0.835417	0.5	0.401496
4	70.csv-30.	0.485714	0.478927	0.475433	0.462658
5	80.csv-20.	0.7	0.541667	0.504895	0.440192

Fig 7.lossless compression results

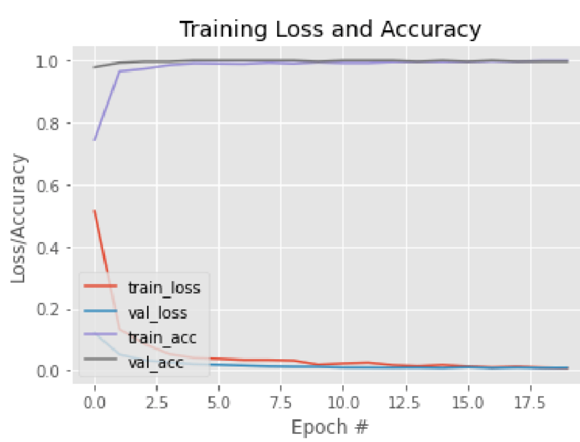

Fig 8.Training and Testing Accuracy

VI.CONCLUSION

In this work, a deep learning-based approach for detecting masks over faces in public places to curtail the community spread of Corona virus is presented. The proposed technique efficiently handles occlu sions in dense situations by making use of an ensemble of single and two- stage detectors at the pre-processing level. The ensemble approach not only helps in achieving high accuracy but also improves detection speed considerably Furthermore, the application of transfer learning on pre- trained models with extensive experimentation over an unbiased data set resulted in a highly robust and low-cost system. The identity detec tion of faces, violating the mask norms further, increases the utility of the system for public benefits.

Finally, the work opens interesting future directions for researchers. Firstly, the proposed technique can be integrated into any high- resolution video surveillance devices and not limited to mask detec tion only. Secondly, the model can be extended to detect facial land marks with a facemask for biometric purposes.

VII. FUTURE SCOPE

The proposed system for shield mask detection utilizes a deep learning-based approach, leveraging a convolutional neural network (CNN) architecture to accurately identify individuals wearing shield masks. Future improvements for shield mask detection using deep learning can include the development of more robust and accurate models that can detect shield masks in various environments and lighting conditions. Integrating multimodal inputs, such as RGB-D images or thermal imaging, can enhance detection accuracy. Additionally, exploring explainability techniques, like attention maps or feature importance, can provide insights into model decisions. Real-time processing and edge deployment can be achieved through model optimization and hardware acceleration. Furthermore, expanding the scope to detect other personal protective equipment (PPE) and integrating with existing surveillance systems can enable comprehensive public health monitoring. Continuous data collection and model updating can ensure adaptability to new scenarios and improved performance. By addressing these areas, shield mask detection using deep learning can become more effective, efficient, and widely applicable in various settings These enhancements make shield mask detection using deep learning more accurate, robust, efficient, adaptable, and widely applicable, ultimately contributing to enhanced public health monitoring, increased safety, and better prevention of infectious disease spread, thereby creating a safer environment for communities worldwide. Integration with CCTV Surveillance Systems: Deploy the system in public areas to automate monitoring without human intervention. Mobile Application Development: Extend the solution into a smartphone app for on-the-go detection and crowd compliance analytics.

VIII. REFERENCES

World Health Organization et al. Coronavirus disease 2019 (covid-19): situation report, 96. 2020. - Google Search. (n.d.). https://www.who.int/docs/default-sour ce/coronaviruse/situation-reports/20200816-covid-19-sitrep-209.pdf?sfvrsn=5 dde1ca2_2.

[2] Social distancing, surveillance, and stronger health systems as keys to controlling COVID-19 Pandemic, PAHO Director says - PAHO/WHO | Pan American Health Organization. (n.d.). https://www.paho.org/en/news/2-6-2020-social-distancing- surveillance-and-stronger-health-systems-keys-controlling-covid-19.

L.R. Garcia Godoy, et al., Facial protection for healthcare workers during pandemics: a scoping review, BMJ, Glob. Heal. 5 (5) (2020), e002553, https://doi. org/10.1136/bmjgh-2020-002553.

[4] S.E. Eikenberry, et al., To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic, Infect. Dis. Model. 5 (2020) 293–308, https://doi.org/10.1016/j.idm.2020.04.001.

[5] Wearing surgical masks in public could help slow COVID-19 pandemic's advance: Masks may limit the spread diseases including influenza, rhinoviruses and coronaviruses – ScienceDaily. (n.d.). https://www.sciencedaily.com/releases/ 2020/04/200403132345.htm.

[6] L. Nanni, S. Ghidoni, S. Brahnam, Handcrafted vs. non-handcrafted features for computer vision classification, Pattern Recogn. 71 (2017) 158–172, https://doi. org/10.1016/j.patcog.2017.05.025

. [7] Y. Jia et al., Caffe: Convolutional architecture for fast feature embedding, in: MM 2014 - Proceedings of the 2014 ACM Conference on Multimedia, 2014, doi: 10.1145/2647868.2654889.

[8] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. Lecun, OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks, 2014.

[9] D. Erhan, C. Szegedy, A. Toshev, D. Anguelov, Scalable Object Detection using Deep Neural Networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 2147–2154, https://doi.org/10.1109/ CVPR.2014.276.

[10] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real- time object detection, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016, vol. 2016-Decem, pp. 779–788, doi: 10.1109/CVPR.2016.91.

[11] M. Jiang, X. Fan, and H. Yan, RetinaMask: A Face Mask detector, 2020, http:// arxiv.org/abs/2005.03950.

[12] M. Inamdar, N. Mehendale, Real-Time Face Mask Identification Using Facemasknet Deep Learning Network, SSRN Electron. J. (2020), https://doi.org/10.2139/ ssrn.3663305

. [13] S. Qiao, C. Liu, W. Shen, A. Yuille, Few-Shot Image Recognition by Predicting Parameters from Activations, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2018, https://doi.org/ 10.1109/CVPR.2018.00755.

[14] A. Kumar, Z.J. Zhang, H. Lyu, Object detection in real time based on improved single shot multi-box detector algorithm, J. Wireless Com. Netw. 2020 (2020) 204, https://doi.org/10.1186/s13638-020-01826-x.

[15] ´A. Morera, ´A. S´anchez, A.B. Moreno, ´A.D. Sappa, J.F. V´elez, SSD vs. YOLO for detection of outdoor urban advertising panels under multiple variabilities, Sensors (Switzerland) (2020), https://doi.org/10.3390/s20164587.

[16] R. Girshick, J. Donahue, T. Darrell, J. Malik, Region-based Convolutional Networks for Accurate Object Detection

and Segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 38 (1) (2015) 142–158, https://doi.org/10.1109/TPAMI.2015.2437384.
[17] K. He, X. Zhang, S. Ren, J. Sun, Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition, IEEE Trans. Pattern Anal. Mach. Intell. (2015), https://doi.org/10.1109/TPAMI.2015.2389824.
[18] R. Girshick, Fast R-CNN, in: Proc. IEEE Int. Conf. Comput. Vis., vol. 2015 Inter, 2015, pp. 1440–1448, doi: 10.1109/ICCV.2015.169.
[19] N.D. Nguyen, T. Do, T.D. Ngo, D.D. Le, An Evaluation of Deep Learning Methods for Small Object Detection, J. Electr. Comput. Eng. 2020 (2020), https://doi.org/ 10.1155/2020/3189691.
[20] Z. Cai, Q. Fan, R.S. Feris, N. Vasconcelos, A unified multi-scale deep convolutional neural network for fast object detection, Lect. Notes Comput. Sci. (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (2016), https://doi.org/10.1007/978-3-319-46493-0_22.
[21] C.-Y. Fu, W. Liu, A. Ranga, A. Tyagi, A.C. Berg, DSSD : Deconvolutional Single Shot Detector, 2017, arXiv preprint arXiv:1701.06659 (2017).
[22] A. Shrivastava, R. Sukthankar, J. Malik, A. Gupta, Beyond Skip Connections: Top- Down Modulation for Object Detection, 2016, arXiv preprint arXiv:1612.06851 (2016).
[23] N. Dvornik, K. Shmelkov, J. Mairal, C. Schmid, BlitzNet: A Real-Time Deep Network for Scene Understanding, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, doi: 10.1109/ICCV.2017.447.
[24] Z. Liang, J. Shao, D. Zhang, L. Gao, Small object detection using deep feature pyramid networks, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2018, vol. 11166 LNCS, pp. 554–564, doi: 10.1007/978-3-030-00764-5_51
. [25] K. He, G. Gkioxari, P. Dollar, R. Girshick, Mask R-CNN, in: Proc. 5_51. [25] K. He, G. Gkioxari, P. Dollar, R. Girshick, Mask R-CNN, in: Proc. IEEE Int. Conf. Comput. Vis., vol. 2017-Octob, 2017, pp. 2980–2988, doi: 10.1109/ ICCV.2017.322.