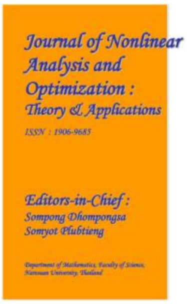
Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 1, No.1 : 2024
ISSN : **1906-9685**



Journal of Nonlinear Analysis and Optimization : Theory & Applications
ISSN : 1906-9685
Editors-in-Chief :
Sompong Dhompongsa
Somyot Plubtieng
Department of Mathematics, Faculty of Science, Naresuan University, Thailand

DEVELOPMENT OF FACE OCCLUSION DETECTION MODEL THROUGH DEEP LEARNING TECHNIQUE

Deepak Jangid, Ph.D. Scholar, MBM University, Jodhpur
Dr. Alok Singh Gehlot, Assistant Professor, MBM University, Jodhpur

Abstract:
Face Occlusion or the presence of extraneous things (such as a beard, cap, scarf, glasses, or other items) on the face that prevent facial recognition, are one of the biggest problems in face recognition systems. Face occlusion detection is a crucial task in computer vision with applications ranging from facial recognition systems to surveillance and human-computer interaction. This research paper proposes the development of a face occlusion detection model using the MobileNet and You Only Look Once (YOLO) object detection framework. To create a dataset for training and evaluation multiple images were collected from different source. The dataset is having four types of occlusion i.e. mask, glasses, hand, and sun glasses. These are the common occlusion that people have in their daily life. Also the types of occlusion selected for this work have both the negative and positive aspects. The proposed models are developed and trained on the high-speed computing hardware platform such as Quadro P5000 Graphics Processing Unit. In this work three pretrained model, Mobilenet v2, yolo v8 nano and yolo v8 small, were used to perform training on face occlusion detection dataset. These models were trained by using the xml and txt annotations file, generated by the labeling tool LabelImg. Experimental results show that yolo v8 nano model is having highest precision among these three models.

Keywords: Face occlusion, Mobilenet, Yolo, Deep learning, Transfer learning

I. INTRODUCTION

Occlusion or the presence of extraneous things on the face that prevent facial recognition, is one of the biggest problems facing face recognition systems. For many years, computer vision researchers have explored face recognition in great detail [1]. The face has a far higher potential for non-intrusive identity recognition than other widely used biometrics like the fingerprint, iris, palm, and vein. As a result, face recognition is extensively employed in numerous application areas, including border control, forensics, and surveillance. Face recognition performance has significantly improved with the advent of deep learning techniques [2–5] and publicly available large-scale face datasets [6,7].

Face occlusion refers to the situation where part of a person's face is obstructed or covered, making it difficult to see or recognize the facial features. This can happen in various contexts, such as in photographs, videos, or real-life interactions. Face occlusion can occur due to a variety of reasons, including objects in the foreground, shadows, accessories like hats or scarves, or intentional obfuscation for privacy reasons.

In the field of computer vision and facial recognition, dealing with face occlusion is a significant challenge. Advanced facial recognition systems aim to overcome these challenges by developing algorithms that can accurately identify and verify individuals even when parts of their faces are obscured. This involves employing sophisticated techniques such as 3D facial modeling, facial landmark detection, and machine learning to enhance recognition accuracy in the presence of occlusions.

Understanding and addressing face occlusion is crucial in various applications, including security systems, user authentication, and human-computer interaction. Researchers and engineers continually work on improving algorithms and technologies to handle face occlusion effectively.

II. CHALLENGES IN FACE OCCLUSION DETECTION

Face occlusion detection refers to the task of identifying and locating obstructed or hidden parts of a person's face in an image or video. This is a challenging problem with several factors that make it complex. Face occlusion detection poses several key challenges in computer vision and facial recognition systems. One of the primary challenges is the diverse nature of occlusions that can occur in real-world scenarios. Occlusions can range from partial blockages caused by accessories like sunglasses or hats to more complex scenarios where faces are obscured by objects or other individuals. Additionally, variations in lighting conditions and environmental factors further complicate the task of accurately detecting face occlusions.

Another significant challenge is the need for robust algorithms that can adapt to different face orientations, poses, and facial expressions. Faces can undergo various transformations, and detecting occlusions in non-frontal or expressive faces requires sophisticated techniques to ensure accuracy and reliability.

Furthermore, the presence of noise and artifacts in images, such as shadows or reflections, can interfere with occlusion detection. Developing algorithms that can distinguish between true occlusions and irrelevant visual disturbances is a critical aspect of improving the overall performance of face occlusion detection systems.

Privacy concerns also contribute to the complexity of face occlusion detection. Striking a balance between effective occlusion detection and respecting privacy by avoiding unnecessary intrusion raises ethical considerations and necessitates the development of privacy-preserving techniques.

In conclusion, addressing the challenges in face occlusion detection requires advancements in algorithmic robustness, adaptability to diverse scenarios, handling variations in facial features, and incorporating privacy-conscious methodologies. Overcoming these obstacles is essential for the continued improvement and deployment of accurate and reliable face recognition systems in real-world applications.

Addressing these challenges often involves a combination of advanced computer vision techniques, deep learning models, and careful consideration of the specific application requirements.

III. TECHNIQUES USED FOR FACE OCCLUSION DETECTION

Various techniques are employed to address the challenge of detecting occluded faces in images or video streams. One common approach involves leveraging deep learning architectures, such as convolutional neural networks (CNNs), which are trained on diverse datasets containing both occluded and non-occluded facial images. These networks learn hierarchical representations that enable them to discern subtle patterns associated with occlusion. Another technique involves the use of facial landmark detection, where key facial points are identified and tracked over time. Sudden changes or inconsistencies in the expected positions of these landmarks can indicate face occlusion. Additionally, depth sensing technologies, like stereo vision, are employed to capture the three-dimensional structure of the face, helping to identify occluded regions by analyzing disparities in depth information. Ensemble methods, combining the strengths of multiple algorithms, are also popular for robust face occlusion detection. These techniques collectively contribute to enhancing the reliability and accuracy of face recognition systems in real-world scenarios where occlusion is a common challenge.

IV. LITERATURE REVIEW

A thorough survey on occlusion has been done in detail in [8]. This article offers a comprehensive overview of face recognition approaches in the presence of occlusion. The classification of existing methods into three categories: Occlusion robust feature extraction, Two methods of face recognition are occlusion aware and occlusion recovery-based. Innovative and recently released works have been discussed, with a focus on deep learning techniques for occluded face recognition. Additionally, the comparative analyses of the performances for both face recognition and obstructed face detection has been shown.

The author's [9] developed a framework which is robust for its superiority in detecting faces with severe occlusion. It makes use of the gradient and form cues in a deep learning model. The three primary contributions of this work as follows: First, a novel energy function-based face detection approach has been discussed then a CNN models is proposed to generate deep features for faces that are obscured. Lastly, a unique sparse classification model with deep learning technique is built to determine whether the discovered face is obscured. It is demonstrated that the suggested head identification technique is reliable in identifying faces in any position. Furthermore, the facial occlusion verification technique that has been suggested is capable of accurately determining if the face area is obscured or not. According to experimental results, the proposed occlusion verification system can achieve 97.25% accuracy rate at a speed of 10 frames per second, and the created head detection algorithm can achieve 98.89% accuracy rate even in the presence of various forms of severe facial occlusions.

The work in [10] combines the region proposal approach with convolutional neural networks (CNN) to provide a robust and efficient end-to-end facial occlusion detection system. Two CNNs make up the coarse-to-noise technique used by the framework. While the second CNN determines which region of the face is obscured from the head image, the first CNN finds the head element inside an upper body image. Since the model works directly with image pixels and the architecture is built on the end-to-end principle, using CNN is more optimal from a system perspective than earlier techniques. A face occlusion database containing more than 50,000 photos with identified facial features was used for evaluation. According to experimental findings, the suggested framework works incredibly well. This study suggested a method for facial occlusion detection to improve ATM surveillance security. A face occlusion classifier and a head detector make up the coarse-to-fine technique. CNN, MLP, and region proposal are the Edge-Boxes that are used to implement the head detector. Using the face occlusion database, the LFW dataset, and the AR face database, the approach achieved detection accuracies of 97.58%, 85.61%, and 100%.

Target detection has become a ubiquitous technology in many facets of life in recent years, owing to its significance in deep learning. This work [11] uses a deep learning target detection technique called SSD (Single Shot MultiBox Detector) to address the issue of occlusions in face recognition by classifying and locating face occlusions. Using a self-built data set of seven typical face occlusion types, the average precision of all categories (mAP) was 95.46%. Tests indicate that this approach may successfully identify facial occlusion, offering a fresh concept for intelligent face recognition that is autonomous and has a wide range of potential applications.

An occlusion face identification technique based on picture segmentation was proposed in [12]. The main goal is to acquire the typical face image from the training set, followed by the distinction between the image to be measured and the average face image acquire the error face image; the occlusion region is then identified by segmenting the error face image; and lastly, the unocclusion area is coordinated that shows the classification.

Occlusion is identified in [13] according to skin color based on SSD Algorithm. The suggested approach is broken down into three stages: face detection is done in the first phase using a Hough transform, and occlusion detection in the second step involves using skin color to detect occlusion in an image. In order to classify skin tones and other colors in an image, an SVM classifier is utilized to train a system.

In this paper [14], the authors present the Real World Occluded Faces (ROF) dataset, which includes faces with both upper face occlusion and lower face occlusion. It proposes two evaluation protocols for this dataset. Using benchmark experiments on the dataset, to find that the performance of deep face representation models significantly degrades when tested on real-world occluded faces, but not when tested on artificially generated occluded faces. It has been demonstrated that artificially produced occlusions are not representative of occlusions found in the actual world. The testing deep face models on real-world occluded faces using masks or sunglasses causes noticeable performance decreases. The results' visualization shows that the inner face region is the primary focus of the deep face models. As a result, when this region is obscured, the models perform worse.

This paper [16] introduced a novel face recognition model named AAN-Face, which might be the first effort to enhance the training photos for different face identification tasks. The suggested attention erasing (AE) method is used in attention maps to replicate different degrees of occlusion, strengthening the models' resistance to changes in occlusion or position. The same attention map can be controlled by the suggested attention center loss (ACL) to focus on the same facial features, capturing important local patches and ignoring the uninformative ones. Differentiated local representations are encouraged to be learned by models when the AE scheme is merged with the ACL. Experimental results show that innovative technique outperforms the state-of-the-art models on various difficult tasks, particularly on masked face datasets.

V. MOBILENET AND YOLO

The authors Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, Liang-Chieh Chen presented the MobileNet[17] model. Since AlexNet won the ImageNet Challenge: ILSVRC 2012 , that popularized deep convolutional neural networks, which are now widely used in computer vision. To attain improved accuracy, the overall trend has been to create deeper and more complex networks. However, networks are not necessarily becoming more efficient in terms of size and speed as a result of these advancements in accuracy. Many real-world applications, such as augmented reality, robotics, and self-driving cars, require the recognition tasks to be completed quickly on a platform with limited processing resources. This study describes a low latency, highly efficient network architecture and two hyper-parameters that may be readily tailored to meet the design needs of embedded and mobile vision applications.

MobileNet architectures are known for their low latency, small model size, and good performance on tasks like image classification and object detection. The key features of MobileNet include depthwise separable convolutions, which help reduce the number of parameters and computations, making the models well-suited for resource-constrained environments. MobileNet has different versions, such as MobileNetV1, MobileNetV2, and MobileNetV3, each introducing improvements in terms of efficiency and performance. MobileNetV2, for example, introduced inverted residuals and linear bottlenecks to further enhance the efficiency of the architecture. MobileNetV3 focused on improving the accuracy of the models while maintaining efficiency. The choice of which version to use depends on the specific requirements of the task and the available computational resources.

YOLO Model: YOLO, or "You Only Look Once," was first published in [18]. YOLO is faster and more precise than earlier cutting-edge object identification algorithms, according to the authors. A novel method for object detection is called YOLO. Classifiers are repurposed to perform detection in previous object detection operations. Rather, they formulate object detection as a regression issue to bounding boxes that are geographically separated and the corresponding class probabilities. Bounding boxes and class probabilities are directly predicted by a single neural network from complete images in a single assessment. The detection pipeline may be immediately improved from start to finish based on detection performance because it is essentially one network. Their unified architecture operates very quickly. At 45 frames per second, the baseline YOLO model processes images in real time. Fast YOLO, a condensed form of the network, achieves double the mAP of other real-time detectors while processing an amazing 155 frames per second. While YOLO is less likely to forecast false positives on background, it produces more localization errors than the most advanced detection algorithms. Lastly, YOLO picks up extremely basic object representations. When extrapolating from natural images to other domains such as artwork, it performs better than other detection techniques like DPM and R-CNN.

All layers employ a linear activation function, with the exception of the last layer, which uses ReLU as the activation function. Several other techniques, like batch normalization and dropout, regularize the model and keep it from overfitting.

VI. DATASET COLLECTION

Collecting a diverse and well-annotated dataset is crucial for training an effective face occlusion detection model. To create a robust dataset for face occlusion detection, it is essential to gather a diverse set of images that represent various real-world scenarios.

To create a dataset for training and evaluation multiple images were collected from different source[15]. The dataset contains 7514 images having face with occlusion

Table 1: Dataset Collection

Occlusion Type	Number of Images
Glasses	2011
Mask	1412
Sunglasses	2185
Hand	1906

The dataset is having four types of occlusion i.e. mask, glasses, hand and sun glasses. These are the common occlusion that people have in their daily life. Also the types of occlusion selected for this work have both the negative and positive aspects. A person with mask is considered to have positive occlusion when he enter into a hospital but the same occlusion i.e. mask is considered to have negative aspect when goes in ATM machine

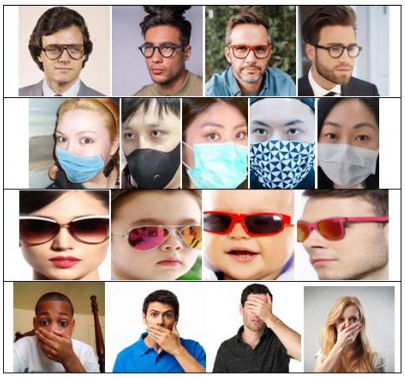


Figure 1: Sample Images [15]

VII. EXPERIMENTAL SETUP AND RESULTS

In this work three pretrained model, Mobilenet , yolov8 nano and yolov8 Small, were used to perform training on face occlusion detection dataset. These models were trained by using the xml and txt annotations file, generated by the labeling tool labelImg.

A. Implementation of Face Occlusion Detection Model

The proposed face occlusion detection model is implemented in the tensorflow and python environment. The model is trained on HP Z6 Workstation having Windows 10 pro Workstation operating system, Intel Xeon Silver 4110 two CPUs of 2.1 GHz with 32GB RAM. The Z6 workstation is equipped with Quadro P5000 GPU, having 16GB GDDR5X GPU Memory, 2560 NVIDIA CUDA Cores, and 8.9 Teraflops Computing Power. The face occlusion image dataset of a face occlusion contains a total of 7514 images. The face occlusion images are labeled using LabelImg image processing toolbox. The image labeler app LabelImg is very useful for generating the xml labels, in the form of rectangular regions of interest. After labeling the face occlusion images, labels are saved in the .xml amd .txt file format. The labeled xml image data is loaded in the tensorflow and python programming environment. The dataset with labels is randomly split into training and test set.

The training dataset consists of 6772 labeled images and the test dataset consists of 742 image datasets. The training dataset is used for feature extraction through the pre-trained deep networks. The feature maps obtained through the training process of pre-trained deep networks are used as the input for Region Proposal Network (RPN). The region proposal is generated by the RPN. These region proposals are either an object or background. The trained object detector is used to detect the object of interest within the unseen face occlusion images.

B. Face occlusion detection through mobilenet_v2 Pretrained Network

The mobilenet_v2 model was trained on the images and annotations generated by LabelImg tool. Figure 2 shows different losses during training process. Table 2 shows the performance in terms of precision.

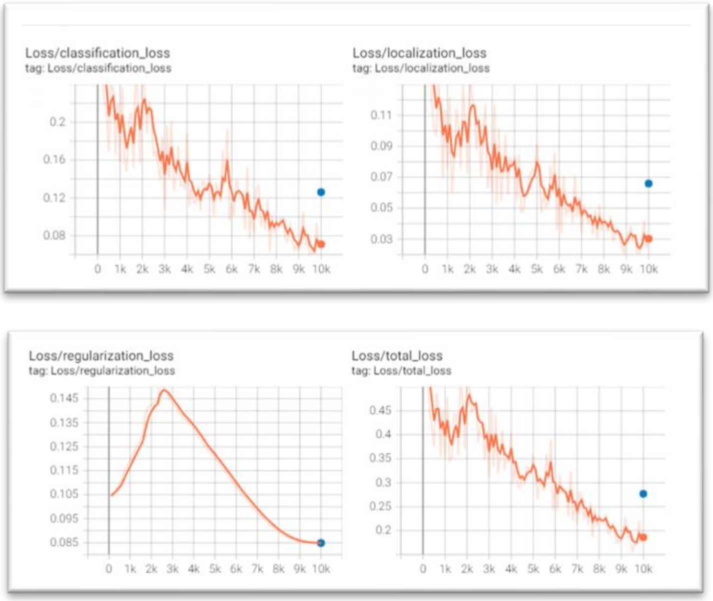


Figure 2: Different losses for mobilenet_v2

As shown in table 2 the precision is good for sun glasses, glass and hand category.

Table 2: Precession for mobilenet v2

Class	Average Precision
Glasses	0.980134
Sun Glasses	0.998844
Hand	0.912509

C. Face occlusion detection through yolo v8 small Pretrained Network

The yolo v8 small model was trained on the images and annotations generated by LabelImg tool. Figure3 shows different losses during training process. Table 3 shows the performance in terms of precision.

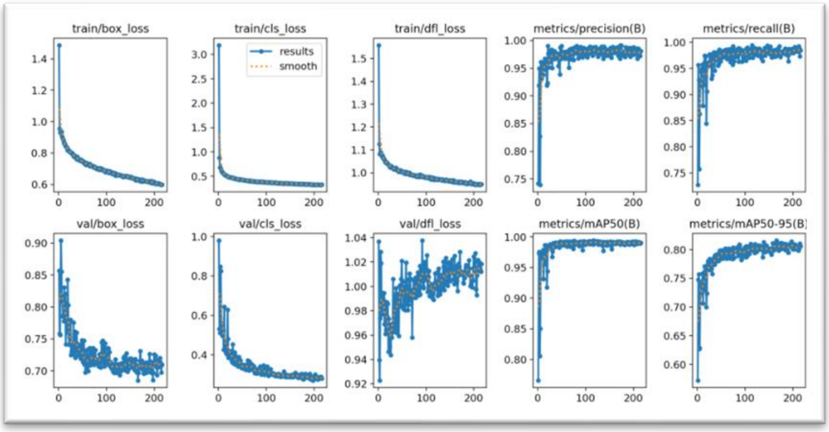


Figure 3: Different losses for yolo v8 small

As shown in table 3 the precision by using the yolo v8 small is much better for all the classes viz. sun glasses, glass, hand and mask.

Table 3: Precession for yolo v8 small

Class	Average Precision
Glasses	0.992
Sun Glasses	0.992
Hand	0.995
Mask	0.984

D. Face occlusion detection through yolo v8 nano Pretrained Network

The yolo v8 nano model was trained on the images and annotations generated by LabelImg tool. Figure 4 shows different losses during training process. Table 4 shows the performance in terms of precision.

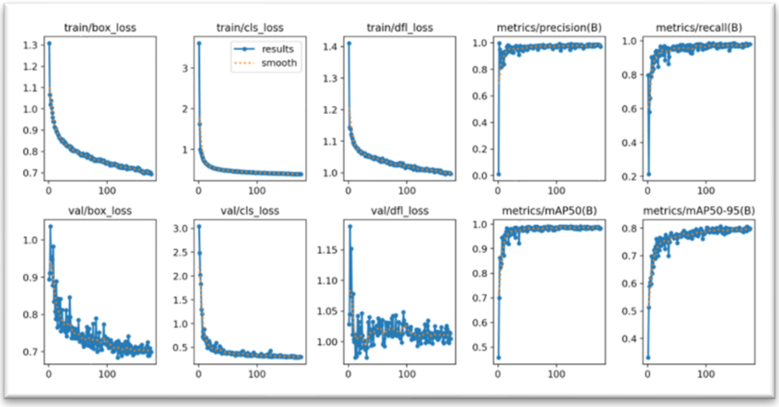


Figure 4: Different losses for yolo v8 nano

As shown in table 4 the precision by using the yolo v8 nano is very close to the yolo v8 small model, but as the aim of this model development is to deploy the trained model in embedded system, we use the yolo v8 nano model.

Table 4: Precession for yolo v8 nano

Class	Average Precision
Glasses	0.991
Sun Glasses	0.995
Hand	0.995
Mask	0.983

VII. CONCLUSION

Face occlusion detection is a crucial task in computer vision with applications ranging from facial recognition systems to surveillance and human-computer interaction. This research paper proposed the development of a face occlusion detection model using the MobileNet architecture and You Only Look Once (YOLO). The dataset is used for this work is having four types of occlusion i.e. mask, glasses, hand and sun glasses. The proposed models are developed and trained on the high-speed computing hardware platform such as Quadro P5000 Graphics Processing Unit. The face occlusion detection dataset trained on three pretrained model, Mobilenet v2, yolo v8 nano and yolo v8 Small. Experimental results show that yolo v8 nano model is having highest precision among these three models.

REFERENCES

1. Best-Rowden, L., Anil, K.J.: Longitudinal study of automatic face recognition. IEEE Trans. Pattern Anal. Mach. Intell. 40(1), 148–162 (2018)
2. He, K., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770–778 (2016)

3. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 1097–1105 (2012)
4. Liu, W., et al.: Sphereface: deep hypersphere embedding for face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 212–220 (2017)
5. Wang, H., et al.: Cosface: large margin cosine loss for deep face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 5265–5274 (2018)
6. Guo, Y., et al.: Ms-celeb-1m: a dataset and benchmark for large-scale face recognition. In: European Conference on Computer Vision, 87–102. Springer (2016)
7. Guo, Y., et al.: Ms-celeb-1m: a dataset and benchmark for large-scale face recognition. In: European Conference on Computer Vision, 87–102. Springer (2016)
8. Zeng, D., Veldhuis, R., Spreeuwers, "A survey of face recognition techniques under occlusion. IET Biom. 10(6), 581–606 (2021)".
9. L. Mao, F. Sheng and T. Zhang, "Face Occlusion Recognition With Deep Learning in Security Framework for the IoT," in IEEE Access, vol. 7, pp. 174531-174540, 2019
10. S. M. Sghaier and A. O. Elfaki, "Efficient Techniques For Human Face Occlusions Detection and Extraction," 2021 International Conference of Women in Data Science at Taif University (WiDSTaif), Taif, Saudi Arabia, 2021, pp. 1-5
11. X. Ziwei et al., "Face Occlusion Detection Based on SSD Algorithm," 2020 IEEE 10th International Conference on Electronics Information and Emergency Communication (ICEIEC), Beijing, China, 2020, pp. 362-365
12. Z. Gao, D. Li, C. Xiong, J. Hou and H. Bo, “Face recognition with contiguous occlusion based on image segmentation,” 2014 International Conference on Audio, Language and Image Processing, Shanghai, 2014, pp. 156-159, doi: 10.1109/ICALIP.2014.7009777.
13. G.Suvarna Kumar, P.V.G.D. Prasad Reddy , M.Srinadh Swamy and Sumit Gupta “Skin based occlusion detection and face recognition using machine learning technique,” International Journal of Computer Applications, March 2012.
14. Mustafa Ekrem Erakın, Ugur Demir, Hazım Kemal Ekenel, "On Recognizing Occluded Faces in the Wild", 2021 International Conference of the Biometrics Special Interest Group (BIOSIG).
15. https://www.kaggle.com/ datasets: face-obstructions, Glasses or No Glasses, MAFA_data, Face Mask Detection, Covid Mask, Sunglasses / No Sunglasses, Glasses and Coverings, With/Without Mask, Face Mask Detection, Masked Face Detection In the Wild dataset
16. Wang and G. Guo, "AAN-Face: Attention Augmented Networks for Face Recognition," in IEEE Transactions on Image Processing, vol. 30, pp. 7636-7648, 2021
17. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. -C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 4510-4520
18. J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 779-788