Drivers Fatigue Detection Using EfficientDet In Advanced Driver Assistance Systems

1st Author: Riadh Ayachi
Laboratory of Electronics and Microelectronics,
Faculty of Sciences of Monastir,
University of Monastir,
Monastir 5000, Tunisia
riadh.ayachi@fsm.rnu.tn

2nd Author: Mouna Afif
Laboratory of Electronics and Microelectronics,
Faculty of Sciences of Monastir,
University of Monastir,
Monastir 5000, Tunisia
Mouna.afif@outlook.fr

3rd Author: Yahia Said
Laboratory of Electronics and Microelectronics, Faculty of Sciences of Monastir, University of Monastir, Monastir 5000, Tunisia
Electrical Engineering Department, College of Engineering, Northern Border University, Arar, Saudi Arabia
yahia.said@fsm.rnu.tn

4th Author: Abdessalem Ben Abdelali
Laboratory of Electronics and Microelectronics,
Faculty of Sciences of Monastir,
University of Monastir,
Monastir 5000, Tunisia
Abdessalem.BenAbdelali@enim.rnu.tn

2021 18th International Multi-Conference on Systems, Signals & Devices (SSD) | 978-1-6654-1493-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/SSD52085.2021.9429294

Abstract*—Modern cars have focused on road safety by guaranteeing driver, pedestrians, and other traffic object safety. Advanced driver assistance systems are a set of intelligent systems that support the driver by providing more information about the surrounding environment. The fatigue detection system is an intelligent system that detects the driver's face and gauges the driver's tiredness state. Such a system can prevent accidents by stropping the car if the driver is drowsy. In this paper, we propose a driver fatigue detection based on object detection model fatigue indicators. The efficientDet model was used to detect the state of the eye and mouth states then the eyes' closure duration/Percentage of eye closure (PERCLOS) and yawning frequency/frequency of mouth (FOM) were used to judge fatigue state. The efficientDet is a lightweight object detection model with high performance. The proposed approach was evaluated on the National Tsing Hua University Driver Drowsiness Detection (NTHU-DDD) dataset. The proposed approach has achieved an accuracy of 96.05% and real-time processing. The reported results show the efficiency of the proposed model for driver fatigue detection.

Keywords*—Driver fatigue detection, deep learning, efficientDet, advanced driver assistance system.

I. INTRODUCTION

Recently, the number of accidents caused by drowsy drivers has increased dramatically. According to The National Highway Traffic Safety Administration in the USA, 72000 crush accidents were caused by drowsy drivers. In 2013, 44000 injuries and 800 death were reported because of drowsy driver accidents. However, the number of accidents is increasing each year and additional 6000 accidents are reported. The main cause behind sleeping when driving is not exactly estimated and can happen for many reasons such as sleep disordering, night shifting works, medications, and many others. Sleeping while driving is a dangerous situation for the driver and other traffic components. So, it is prohibited to drive when feeling drowsy for safety purposes.

Advanced driver assistance systems (ADAS) were widely integrated into today's cars to assist the driver and to handle easy and repetitive actions such as high way driving and parking. ADAS is equipped with a huge number of sensors such as lidar, radar, cameras, and ultrasonic sensors. Cameras were the most important sensor more many systems such as traffic sign detection and recognition [1], traffic light detection [2], pedestrian detection [3], and fatigue detection [4]. The data provided by the cameras can be used for many tasks without the need for additional sensors in addition to the low cost of the camera sensor compared to other sensors such as lidar and radar.

For driver fatigue detection, the visual data provided by the camera inside the vehicle is processed by analyzing face landmarks. The eyes and moth positions can be used to estimate the fatigue state of the driver. Yawing and low blinking frequency are the main fatigue indicators. In effect, focusing on the mouth to detect yawing in addition to estimating the blinking frequency can help to detect fatigue. the eyes' closure duration/Percentage of eye closure (PERCLOS) is a fatigue detection metric based on calculating the duration of eyes closure per minute. If the percentage is equal to or higher than 80% then the driver is falling asleep. yawning frequency/frequency of mouth (FOM) can be used as a driver fatigue metric by calculating the percentage of mouth opening in a given period. Yawing has a well-defined moth position and timing. The mentioned metrics were combined to detect fatigue.

To detect and analyse facial landmarks, an object detection model based on a convolutional neural network (CNN) was deployed. CNN was widely used to solve computer vision applications such as object detection [5], scene recognition [6], indoor objects detection [7]. Unlike handcrafted methods, CNN models can learn directly from input data without being explicitly programmed. CNN was inspired by the visual cortex for visual data processing and its decision-making mimic the biological rain through a deep artificial neural network. The main limitation of CNN models that the need for large-scale datasets for training and the need for huge computation resources. In recent years, the world has witnessed a huge explosion in data. Large scale datasets were collected from the internet and through public surveillance cameras. Datasets such as ImageNet [8] and Microsoft Common objects in context (MSCOCO) [9] has provided huge data amount that exceeds millions of images. Besides, recent advances in graphical processing units (GPU) have enabled the deployment of CNN models easily and have boosted the performance to a higher level. In addition to desktop GPUs, embedded GPUs were developed to respond to real-world application needs. Furthermore, special GPUs were developed for ADAS and autonomous vehicles.

To detect facial landmarks, the efficientDet model [10] was proposed. It was designed to achieve high detection performance with a minimum of computation complexity. In comparison to other object detection models, the efficientDet has 4 to 9 times lower model size and 13 to 42 times fewer floating-point operations (FLOPS). The main idea behind the high detection performance is the proposed bidirectional features pyramid network (BiFPN) that allows detecting objects at different scales. In addition, a scaled backbone [11] was used to achieve lower computation complexity. The efficientDet D7 has achieved average precision (AP) of 55.1% on the MSCOCO dataset and a latency of 232 milliseconds (ms) per image on the Nvidia Titan V GPU.

In this work, we propose to use the efficientDet D0 for facial landmark detection in order to achieve good performance while guaranteeing real-time processing on an embedded device. The efficientDET D0 has achieved an AP of 34,6 on the MSCOCO dataset and a latency of 12 ms per image.

For driver fatigue detection the model was fine-tuned using the National Tsing Hua University Driver Drowsiness Detection (NTHU-DDD) dataset [12]. The model was initialized with the MSCOCO dataset weights then finetuned on the proposed dataset. The achieved results convincing with an accuracy of 96.05% and a processing speed of 43 frames per second (FPS) on the Nvidia GTX960 GPU.

The main contributions of this paper are the followings:

- Proposing a fatigue detection system for ADAS
- Proposing the use of efficientDet D0 for facial landmark detection
- Using the PERCLOS and FOM metrics to judge driver fatigue y analysing face landmarks.
- Fine-tuning the proposed model on the NTHU-DDD dataset.

The rest of the paper is organized as a fellow, section 2 is reserved for related works. In section 3, the proposed approach was detailed. The experiment and results are discussed in section 4. Conclusions and future works are presented in section 5.

II. RELATED WORKS

Road safety is a global concern and an important research field. Many intelligent systems were invented to elevate the security level in vehicles. Driver fatigue detection has been studied for a long period and many works have been proposed but this system needs more improvement to be reliable and trusted. For a complete overview of driver fatigue detection, readers can refer to the survey proposed in [13].

Park et al. [14] proposed a drowsiness detection system based on a deep neural network. The proposed network is composed of three convolutional neural networks. The networks were used to build a robust facial movement and head gesture detector. The output of all networks was concatenated and fed to a classification layer based on the softmax function. The proposed network was evaluated on the NTHU-DDD dataset and an accuracy of 73.06% was achieved.

A CNN model was proposed in [15] for drowsy detection. The network was used to detect facial landmarks positions. The CNN model is composed of an input layer, three convolution layers, and a binary classifier with a softmax layer function was used. The dropout optimization was deployed after each layer to avoid overfitting. The model can predict fatigue and not fatigue classes. An accuracy of 81% was achieved when evaluating the model on the NTHU-DDD dataset.

A driver sleepiness detection system was proposed in [16] based on CNN models ensemble. Three CNN models were used for different feature extraction to perform different tasks. The AlexNet model [17] was used to detect the background and environmental variations. VGG-faceNet [18] was used for face features detection such as gender and ethnicities. FlowimageNet [19] was used to detect the head gesture. The ResNet [20] was used for hand gesture detection. The output of all CNN models was concatenated and a softmax layer was used for drowsiness prediction. The proposed CNN ensemble was evaluated on the NTHU-DDD dataset and an accuracy of 85%.

Reddy el al. [21] proposed a drowsiness detection system based on the combination of two CNN models. First, the MTCNN model [22] was used to detect face landmarks (eyes and mouth). Then Two separated CNN models were used to predict the drowsiness state. The first CNN was used to analyse the mouth features and the second CNN was used for analysing the eyes feature. Outputs from both models were concatenated to generate the final predictions. The proposed approach achieved an accuracy of 89.5%

III. PROPOSED APPROACH

In this work, we proposed a driver fatigue detection system based on an object detection model and fatigue indicators, which are PERCLOS and FOM. The object detection model was used to detect facial landmarks in the image and predict its state (open or closed). The main landmarks used for fatigue detection are the eyes and the mouth. By analysing the continuous of those landmarks state in a given period, fatigue state can be predicted.

For facial landmarks detection, we propose the use of the efficientDet model [10] thank to its high detection performance and its low computation complexity for embedded implementation suitability. The efficientDet D0 was used since it has the lowest model size and the fastest processing speed. The efficientDet D0 was based on the

efficientNet B0 CNN model [11] as a backbone. The efficienNet B0 [11] model is composed of a convolution layer followed by 16 mobile inverted bottlenecks (MBConv) [23] with squeeze and excitation connections [24] then a 1x1 convolution was used followed by a pooling layer.
The efficientDet proposed a new features pyramid network named bi-directional features pyramid network (BiFPN). The BiFPN was based on the NAS-FPN where more connections were added and the network was scaled based on the compound scaling technique proposed in [11]. 3 BiFPN layers were used in the efficientDet model. The architecture of the efficientDet model is presented in figure 1.

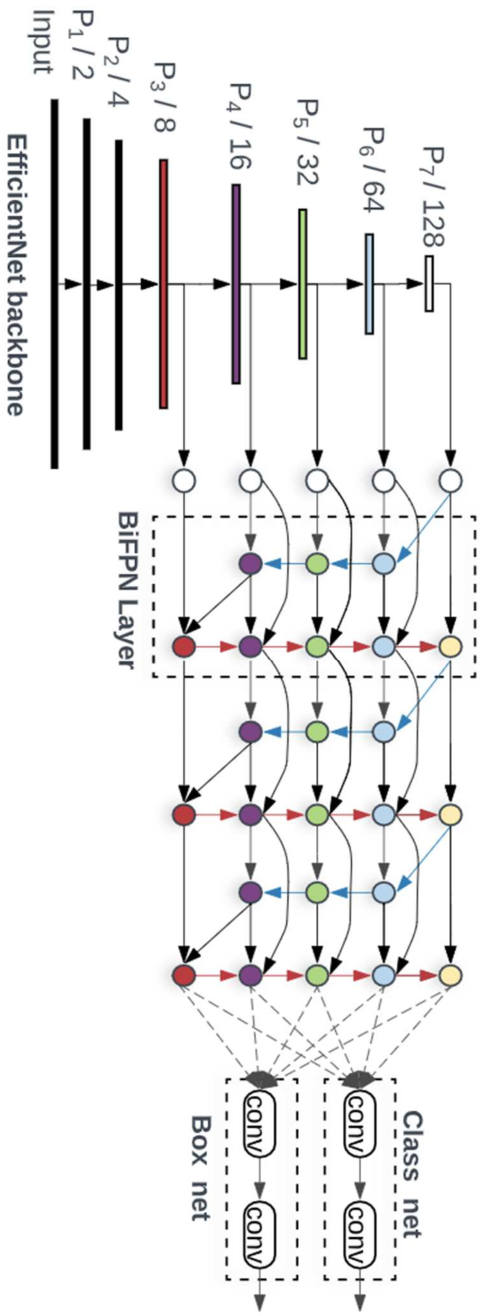

Figure 1: efficientDet model

The model was used to detect facial landmarks and recognize the state of those landmarks. The efficient used two separated sub-networks where a sub-network is used to detect and localize objects and the second sub-network was used to classifier the detected objects. In our case, the eyes and mouth were detected and the state of those landmarks (open/close) were classified. The state of the facial landmarks was used for fatigue detection. The fatigue state cannot be judged with a single image, the state of the facial landmarks must be analysed for a given period to avoid false alarm.
To generate trusted predictions, the drowsiness metrics were employed. The PERCLOS is the eye's closure duration/percentage of eye closure. It is to say the duration of the closure of the eye in a given duration. The PERCLOS was used as a drowsiness metric if the eye was closed in 80% of the evaluation period. In addition, a normal driver blinks approximately 10 times per minute. So, if the blinking frequency is less than 5, then the driver has drowsiness indications.

The FOM is the yawning frequency/frequency of mouth which defines the state of the mouth in a given period. If the driver is talking then the frequency of mouth is high but if the mouth was opened for a while then closed then the driver was yawning. Those metrics were used to judge fatigue state based on the analysed facial landmarks.

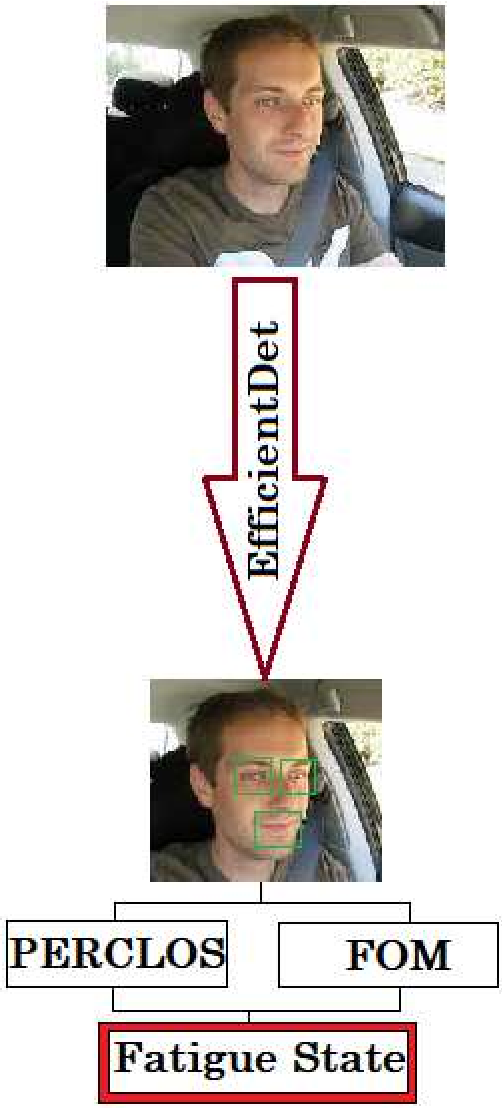

Figure 2: proposed approach for driver fatigue detection.

The proposed driver fatigue detection is a light model, which makes it suitable for implementation on a mobile device for an ADAS. The model has fewer FLOPS compared to other object detection models with the same performance. Thus, the model can achieve real-time processing on the mobile device.

IV. EXPERIMENT AND RESULTS

The proposed driver fatigue detection was evaluated on a desktop with an Intel i7 CPU with 32 GB of RAM and an Nvidia GTX960 GPU. The efficientDet model was developed based on the Keras deep learning framework with GPU acceleration through the CUDA library. The open cv library was used to load and display images and videos.

To train the model, the gradient descent algorithm was used. Adam optimizer was used with an initial learning rate of 0.001 and a batch size of 16. In addition to optimizing the weights, the adam optimizer optimizes the learning rate to achieve better results faster. As loss function for the regression output, the smooth l1 was used and for the classification, the focal loss was used.

The NTHU-DDD dataset was used to train and test the proposed approach. The dataset is composed of 36 videos. The videos present different situation were both male and female with different ethnicities has participle in the recording. Besides, each video has drowsiness behaviour such as low blinking frequency, yawning, and nodding. The videos were recorded where the drivers were wearing sunglasses and without sunglasses to present more challenges of real-world conditions. A total of 450 videos were captured in different situations and conditions. The dataset was divided into three sets, training set, evaluation set, and testing set. 360 videos resulting in 722223 frames were used for the training, 20 videos (173257 frames) were used for evaluation and 70 videos generating 736132 frames were used for testing. The videos have a resolution of 640x480 with a 30 FPS AVI format without audio. Figure 3 presents samples of the HTHU-DDD dataset.

Figure 3: samples of the NTHU-DDD dataset

To improve the performance and to increase the generalization power of the proposed model, we propose to apply a data augmentation technique on the training data. First, intensity manipulation with different ratios (+ 15%, 30% and – 15%, 30%) was used as a data augmentation technique. Second, a vertical merroir was used for data augmentation. Finally, a horizontal translation was applied for data augmentation.

The model weights were initialized using the MSCOCO weights and the output layers were fine-tuned suing the proposed data. The model was trained for 20 epochs and each epoch has 45123 iterations. The training process has lasted for 3 days on the proposed desktop. The model's loss function was optimized to 0.00284.

Since fatigue state cannot be judged with a single frame, we propose to fix the number of consecutive frames to 60 frames for both validation and testing. The model has achieved a validation accuracy of 96.45% and a testing accuracy of 96.05%. Table 1 present the achieved accuracies for different situations and conditions on the testing set.

Table 1: obtained accuracies in different situations

situation	Accuracy (%)
Barefaced	97.86
Glasses	95.98
Sunglasses	94.82
Night-barefaced	97.43
Night-glasses	94.28
Average	96.05

As shown in table 1 the lowest accuracies were achieved when wearing sunglasses because it reduces the model polity to analyse the eyes. The availability of the facial landmarks improves the model accuracy but the proposed model has achieved acceptable accuracy with occluded parts. The proposed model has proved its efficiency based on the achieved results. For better evaluation, the proposed approach was compared against state-of-the-art works on the NTHU-DDD dataset. Tale 2 present a comparison against existing works.

Table 2: comparison against state-of-the-art works.

Approach	Accuracy (%)
CNN model [15]	81
CNN ensemble [16]	85
Multi-CNN [21]	89.5
EfficientDet (ours)	96.05

The proposed approach outperformed state-of-the-art works on the NTHU-DDD dataset. In addition, the proposed approach has a low model size and few FLOPS compared to the existing works and runs in real-time with a processing speed of 43 FPS on the Nvidia GTX960 GPU. Unlike existing works, the proposed approach is suitable for real-world application through implementation on mobile devices

Based on the achieved results, the proposed approach has proved its efficiency. The use of the efficientDet model for facial landmark detection in addition to the use of the drowsiness indicators (PERCLOS and FOM) was very helpful for driver fatigue detection. Besides, the proposed approach has achieved real-time processing and it is suitable for embedded implementation.

V. CONCLUSION

Driver fatigue detection system plays a big role in accident prevention through warning drowsy drivers or stop the vehicle automatically if the driver falls in sleep. Building a reliable driver fatigue detection system is a hard challenge and many situations must be taken into account. In this paper, we proposed a driver fatigue detection system for ADA based on combining an object detection model and drowsiness metrics. The efficientDet was used to detect and classify the state of different facial landmarks. The PERCLOS and FOM were used as drowsiness metrics to generate final predictions on the driver's fatigue state. The proposed approach achieved state-of-the-art performance in addition to real-time processing and its suitability for real-world applications. The achieved results have proved the robustness of the proposed approach. As future works, the proposed approach will be evaluated on real conditions with an embedded implementation.

References

[1] Ayachi, Riadh, Mouna Afif, Yahia Said, and Mohamed Atri. "Traffic signs detection for real-world application of an advanced driving assisting system using deep learning." Neural Processing Letters 51, no. 1 (2020): 837-851.

[2] Kulkarni, Ruturaj, Shruti Dhavalikar, and Sonal Bangar. "Traffic Light Detection and Recognition for Self Driving Cars Using Deep Learning." In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), pp. 1-4. IEEE, 2018.

[3] Ayachi, Riadh, Yahia Said, and Abdessalem Ben Abdelaali. "Pedestrian Detection Based on Light-Weighted Separable Convolution for Advanced Driver Assistance Systems." Neural Processing Letters 52, no. 3 (2020): 2655-2668.

[4] Chang, Wan-Jung, Liang-Bi Chen, and Yu-Zung Chiou. "Design and implementation of a drowsiness-fatigue-detection system based on wearable smart glasses to increase road safety." IEEE Transactions on Consumer Electronics 64, no. 4 (2018): 461-469.

[5] Ayachi, Riadh, Yahia Said, and Mohamed Atri. "A Convolutional Neural Network to Perform Object Detection and Identification in Visual Large-Scale Data." Big Data (2020).

[6] Afif, Mouna, Riadh Ayachi, Yahia Said, and Mohamed Atri. "Deep Learning Based Application for Indoor Scene Recognition." Neural Processing Letters (2020): 1-11.

[7] Afif, Mouna, Riadh Ayachi, Edwige Pissaloux, Yahia Said, and Mohamed Atri. "Indoor objects detection and recognition for an ICT mobility assistance of visually impaired people." Multimedia Tools and Applications 79, no. 41 (2020): 31645-31662.

[8] Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. "Imagenet: A large-scale hierarchical image database." In 2009 IEEE conference on computer vision and pattern recognition, pp. 248-255. Ieee, 2009.

[9] Lin, Tsung-Yi, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. "Microsoft coco: Common objects in context." In European conference on computer vision, pp. 740-755. Springer, Cham, 2014.

[10] Tan, Mingxing, Ruoming Pang, and Quoc V. Le. "Efficientdet: Scalable and efficient object detection." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10781-10790. 2020.

[11] Tan, M. & Le, Q. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. Proceedings of the 36th International Conference on Machine Learning, in PMLR 97:6105-6114

[12] Weng, Ching-Hua, Ying-Hsiu Lai, and Shang-Hong Lai. "Driver drowsiness detection via a hierarchical temporal deep belief network." In Asian Conference on Computer Vision, pp. 117-133. Springer, Cham, 2016.

[13] Ramzan, Muhammad, Hikmat Ullah Khan, Shahid Mahmood Awan, Amina Ismail, Mahwish Ilyas, and Ahsan Mahmood. "A survey on state-of-the-art drowsiness detection techniques." IEEE Access 7 (2019): 61904-61919.

[14] Park, Sanghyuk, Fei Pan, Sunghun Kang, and Chang D. Yoo. "Driver drowsiness detection system based on feature representation learning using various deep networks." In Asian Conference on Computer Vision, pp. 154-164. Springer, Cham, 2016.

[15] Jabbar, Rateb, Khalifa Al-Khalifa, Mohamed Kharbeche, Wael Alhajyaseen, Mohsen Jafari, and Shan Jiang. "Real-time driver drowsiness detection for android application using deep neural networks techniques." Procedia computer science 130 (2018): 400-407.

[16] Dua, Mohit, Ritu Singla, Saumya Raj, and Arti Jangra. "Deep CNN models-based ensemble approach to driver drowsiness detection." Neural Computing and Applications (2020): 1-14.

[17] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Communications of the ACM 60, no. 6 (2017): 84-90.

[18] Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 815-823. 2015.

[19] Donahue, Jeffrey, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, and Trevor Darrell. "Long-term recurrent convolutional networks for visual recognition and description." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2625-2634. 2015.

[20] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[21] Reddy, Bhargava, Ye-Hoon Kim, Sojung Yun, Chanwon Seo, and Junik Jang. "Real-time driver drowsiness detection for embedded system using model compression of deep neural networks." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 121-128. 2017.

[22] Zhang, Kaipeng, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. "Joint face detection and alignment using multitask cascaded convolutional networks." IEEE Signal Processing Letters 23, no. 10 (2016): 1499-1503.

[23] Sandler, Mark, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. "Mobilenetv2: Inverted residuals and linear bottlenecks." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4510-4520. 2018.

[24] Hu, Jie, Li Shen, and Gang Sun. "Squeeze-and-excitation networks." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7132-7141. 2018.