2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops

Real-time Driver Drowsiness Detection for Embedded System Using Model Compression of Deep Neural Networks

Bhargava Reddy, Ye-Hoon Kim, Sojung Yun, Chanwon Seo, Junik Jang
{tb.reddy, yehoon.kim, sojung15.yun, cw0323.seo, ji.jang}@samsung.com

Abstract

Driver's status is crucial because one of the main reasons for motor vehicular accidents is related to driver's inattention or drowsiness. Drowsiness detector on a car can reduce numerous accidents. Accidents occur because of a single moment of negligence, thus driver monitoring system which works in real-time is necessary. This detector should be deployable to an embedded device and perform at high accuracy. In this paper, a novel approach towards real-time drowsiness detection based on deep learning which can be implemented on a low cost embedded board and performs with a high accuracy is proposed. Main contribution of our paper is compression of heavy baseline model to a light weight model deployable to an embedded board. Moreover, minimized network structure was designed based on facial landmark input to recognize whether driver is drowsy or not. The proposed model achieved an accuracy of 89.5% on 3-class classification and speed of 14.9 frames per second (FPS) on Jetson TK1.

Keywords: Driver Monitoring System, Drowsiness Detection, Deep Learning, Knowledge Distillation, Real-time Deep Neural Network, Model Compression.

1. Introduction

Driver Drowsiness is one of the leading causes of motor vehicular accidents. In 2014, 846 fatalities related to drowsy drivers were recorded in NHTSA's reports [1]. These fatalities have remained largely consistent across the past decade. There was an estimated average of 83,000 crashes each year related to drowsy driving between 2005 and 2009. For these reasons, risk alert system for drivers using a detector which can determine drowsiness is highly recommended. The alert system can awaken the drowsy driver or hand over the control to autonomous vehicle.

Various techniques have been implemented to measure driver drowsiness. The techniques can be broadly classified into 3 categories

i. Driving pattern of the vehicle
ii. Psychophysiological characteristics of drivers
iii. Computer Vision techniques for driver monitoring

In the first group of techniques, various state of the art techniques are implemented based on monitoring steering wheel movement [2][3]. Some of the techniques in this group focus on acceleration or breaking time series, lane departure to determine the level of drowsiness in [4]-[6]. The techniques in the second category focus on electrical bio-signals such as EEG (Electroencephalography), ECG (Electrocardiography) and EOG (Electrooculogram) [7]. However, the techniques in the two previously mentioned classes have severe limitations. The former class of techniques can only be used in certain driving condition and are not robust in nature, whereas the latter is difficult for practical purposes, since it is uncomfortable for the driver to wear various signal measuring tools on the body. Thus, driver monitoring based on Computer Vision is becoming popular [8][9]. Computer Vision techniques mainly concentrate on detecting eye closure, yawning patterns and the overall expression of the face and movement of head.

This paper presents a Computer Vision based deep learning approach for driver drowsiness. This method takes driver's face as input and classifies the drowsiness behavior into 3 classes (normal, yawning and drowsy). The biggest advantage of the proposed model is the model is compressed small enough that it can be deployed on an embedded board while preserving reasonable accuracy. To deploy a driver drowsiness detection system in a daily use vehicle, a compressed model is significant. Since a person can fall asleep at any moment, it is highly necessary to have a real-time classifier for drowsiness detection, which consumes low power and can be deployed easily on a vehicle similarly with ECU (Electronic Control Unit).

2. Related Work

In this chapter, we summarize previous approaches on drowsiness detection. In order to improve accuracy and speed of drowsiness detection, various methods have been proposed. Conventional approaches on drowsiness detection are listed, followed by the latest approaches using deep learning. Furthermore, deep learning model compression methods to overcome run-time issues are described.

2160-7516/17 $31.00 © 2017 IEEE
DOI 10.1109/CVPRW.2017.59

2.1. Conventional Approaches for Drowsiness Detection

Driving pattern can be calculated by measuring steering wheel movement or deviation from lane or lateral position. Micro adjustments to the steering wheel are necessary when driving to keep the car in a given lane. Krajweski et al. [3] achieved an accuracy of 86% in drowsiness detection based on correlations between micro adjustments and drowsiness. In the other case of driving pattern recognition, deviation in lane position is used. This monitors the car's position with respective to the lane and analyze the deviation [6]. However the driving pattern based techniques are highly dependent on the driving skills, road conditions and vehicle characteristics.

The second class of techniques uses data taken from physiological sensors, like EEG, ECG and EOG data. EEG signals contain information about brain's activity. Three main signals in EEG for measuring driver's drowsiness are alpha, delta and theta signals. When a driver is drowsy, delta and theta signals spikes up, alpha signal increasing slightly. In [7], this technique gives the best accuracy among all the three methods (more than 90%). However the major drawback of this method is the intrusiveness which disturbs drivers by attaching many sensors on the body. Non-intrusive methods for bio-signals exist, but are less accurate.

The last one is based on facial feature extraction using Computer Vision, where behaviors such as eye closure, head movement, yawning duration, gaze or facial expression have been used. Danisman et al. [8] used distance between eyelids to measure drowsiness of 3 levels. The distinguishing was done based on the number of blinks per minute, under the assumption the count increases as the person becomes drowsier. In [9], behaviors of mouth and yawning are adopted as drowsiness measurement where the modified Viola-Jones object detection algorithm is used for face and mouth detection.

2.2. Drowsiness Detection using Deep Learning

Recently, deep learning is widely used to resolve difficult problems which cannot be handled properly using conventional methods. Deep learning based on Convolutional Neural Networks (CNNs) makes a breakthrough especially for Computer Vision tasks such as image classification, object detection, emotion recognition, scene segmentation [10]-[13] etc.

Dwivedi et al. [14] adopted shallow CNNs for drowsy driver detection with accuracy of 78%. As the latest research, S. Park et al. [15] proposed a new architecture using three networks. In the first network, image feature is learnt by using AlexNet which consists of 5 CNNs and 3 FC layers [16]. 16-layered VGG-FaceNet [17] is utilized to extract facial feature in the second network. The last network works to extract behavior features by using FlowImageNet [18]. As a result, 73% detection accuracy is achieved. Both [14] and [15] focused on improving drowsiness detection accuracy employing binary classification. In real-time applications, performance in terms of speed is also a crucial point.

2.3. Compression Algorithms of Deep Learning Model

Although deep learning is powerful on various classification tasks, it is a burden to deploy deep learning algorithms to practical applications on embedded systems since model size of deep learning is generally large and high computational complexity is required. Therefore, in the recent years algorithms to reduce model size and improve speed have been proposed by using various ways [19]-[29].

Methods to reduce model size have been proposed in [19]-[23]. Generally, trained networks include redundant information, so some of weights can be discarded by applying pruning without accuracy drop. To reduce model size further, quantization techniques have been introduced such as bit-quantization. In bit-quantization, the least number of bits are utilized for representing information of model while minimizing accuracy loss. In some of researches, they adopted binary networks. Even though these works have advantages in terms of model size and speed, accuracy cannot be maintained because of the simplicity of binary operations [21][22].

Moreover, Low-rank decomposition has been proposed in [24] to decompose a tensor and reduce number of matrix operations.

R. Caruana et al. [26] introduced the concept of applying ensemble selection from libraries of models. According to this work, researchers in [25][27][28][29][30] developed learning algorithms by adopting knowledge distillation approach between two networks. A role of one is teacher and that of the other network is student.

Teacher networks are large and have high computation requirements, which can learn patterns from a large dataset. On the contrary student networks are small requiring less computation and can learn only from teacher network. Due to its smaller size, student network is suitable to be implemented on embedded devices and has capability to run at real time on portable devices. Hinton et al. [25] proposed how to transfer weights from teacher network to student network using knowledge distillation. As a result the student network can successfully learn from the teacher.

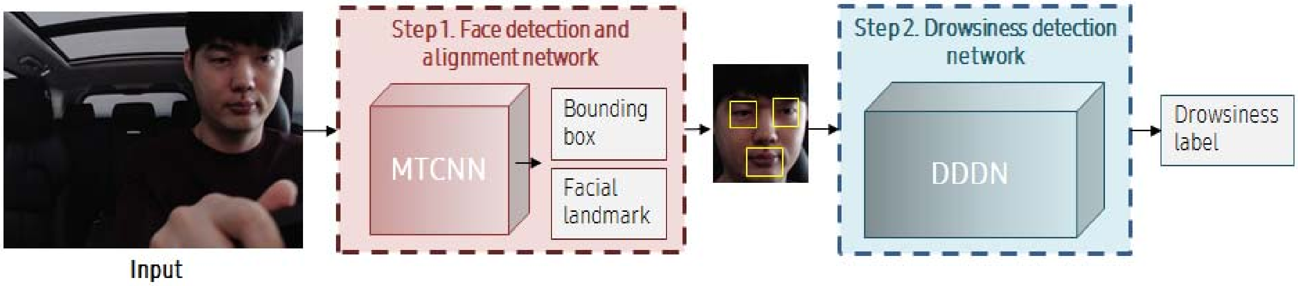

Figure 1: Overall framework of drowsiness detection: Step 1 consists of face and landmark detector and step 2 consists of drowsiness detection network from the detected face and landmark

3. Methodology

This section presents the proposed network architecture. The baseline architecture is described first. Afterwards, two compressed models are introduced. Overall, we propose three types of models, which include the baseline 4-stream drowsiness detection model, 2-stream drowsiness detection model and its compressed version using teacher-student technique[25]-[30] with minimum accuracy drop.

3.1. Architecture

The overall architecture of the proposed drowsiness detection consists of two steps as illustrated in Figure 1. It is a two-step process which the first step is the joint face detection and alignment and the second is the drowsiness detection model. For the face detection and alignment task, Multi-Task Cascaded Convolutional Networks (MTCNN) [32] is used since it is known as one of the fastest and accurate face detector. Exploiting cascaded structure, it can achieve high speed in joint face detection and alignment. As a result of face detection and alignment, face boundary coordinates and five landmark points containing locations of left-eye, right-eye, nose, left-lip-end and right-lip-end are obtained.

Driver Drowsiness Detection Network (DDDN) in second step indicates the proposed models for detecting driver's drowsiness. DDDN takes in the output of the first step (face detection and alignment) as its input. The following subsections describe various experiments on the proposed models for drowsy driver detection in detail. Experimental results of drowsiness detection based on the three proposed models are described in section 4.

3.1.1 Baseline-4 Model

First, a 4-stream deep neural network is proposed as the baseline model as shown in Figure 2. This network is named as baseline-4 model. In this network the inputs are left-eye, right-eye, mouth and face obtained from the detection network. The input images are resized into size of 224 x 224. The baseline-4 model is a neural network consisting 5 convolutional layers for each 4-stream input. Each stream of the network structure is similar to the AlexNet [16] architecture with filter sizes of 11 x 11, 5 x 5, 3 x 3, 3 x 3 and 3 x 3. The number of kernels for each layer is stated in Figure 1. Models similar to AlexNet architectures are beneficial for deployment on embedded board since execution speed of the network is faster than that of the other modern networks such as GoogleNet [36] and ResNet [10].

Motivated by the architecture for gaze tracking [33], the convolution layers of eyes share the same weight. This was proposed because the features from the eyes will approximately be the same. Each stream of convolutional layers ends with fully connected (FC) layers. Each size of the FC layers is stated in the Figure 2. All the FC layers are connected to last two FC layers.

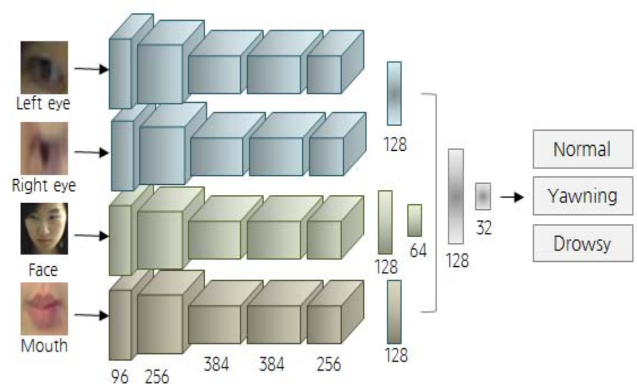

Figure 2: Structure of Baseline-4 Model. This consists of 4 streams with inputs of left-eye, right-eye, face and mouth.

We finally predict 3 classes of outputs as described earlier, which are normal, yawning and drowsy. *Normal* means that the driver is conscious and not in any state of fatigue. *Yawning* indicates the driver might become in danger of drowsy driving in a short time. *Drowsy* means that driver is having severe drowsiness or fatigue condition and immediately needs to take rest.

3.1.2 Baseline-2 Model

However, baseline-4 model has its limitation in terms of speed. Thus an improvement in speed is required for real-time applications.

As an initial step to improve the speed, we focus on reducing the number of streams of the network. Baseline-4 model takes four inputs to separate network streams. The wide shape of this network makes its size huge and significantly affects its efficiency of speed. For this reason, our first approach is to reduce the number of inputs and streams of the network.

Previous works have shown that partial face image can possess comparable features with whole face image on face recognition task [37]. Using sub-images of face as input for a face recognition model is useful in situations with occlusion or limited field of view and is less dependent on accurate face alignment. From this observation, we reduced the number of inputs and streams by half to make the best use of partial features on the face to make the network smaller for faster recognition speed. Instead of using four inputs as the basline-4 model, the newly proposed baseline-2 model is a 2-stream network structure which takes two inputs, the cropped images of left eye and the mouth, excluding the whole face from the set of inputs. The structure of the baseline-2 Model is illustrated in Figure 3.

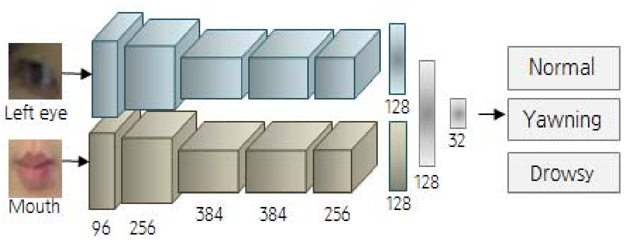

Figure 3: Structure of Baseline-2 Model. This consists of 2 streams with inputs of left-eye and

The model takes only one eye as the input since the movements of the two eyes (if involuntary) are generally identical to each other. Therefore, it can be assumed that the accuracy would not be degraded even though one-eye image is utilized for driver's drowsiness detection. Also, the mouth plays a major role when the driver is yawning, thus the mouth can be a main triggering point. To validate our claim, we passed only two inputs (left eye and mouth) to the 4-stream model, by making the other two stream inputs as null. We obtained just 1% reduction in validation accuracy.

Similar to the baseline-4 model, baseline-2 model consists of 5 convolutional layers for each 2-stream input with filter sizes of 11 x 11, 5 x 5, 3 x 3, 3 x 3 and 3 x 3 . The number of kernels for each layer is stated in Figure 3.

3.1.3 Compressed-2 Model

The initial approach to reduce the number of streams from four to two almost halved the execution time compared to baseline-4 model. However, the speed of the model could be further improved. Therefore, we finally propose the compressed-2 model.

This model adopts a compression method described in [25] using distillation of neural network. "Distillation" of neural network refers to an approach transferring the knowledge from a superfluously huge model to a small model. The technique introduces two concepts, namely teacher network and student network. Student network are smaller in size and does faster computation in contrary to the teacher network. The teacher network is the original large network which is directly trained from the dataset and the student network is a small network which learns features from the soft targets produced by the teacher network.

This technique was introduced because it is difficult to train a small network directly from hard-classified labels, since small networks don't converge easily. Large networks can easily be trained on huge datasets with discrete, hard-classified labels. Instead of training discrete value outputs to the student network, the above paper trained soft value outputs from the teacher network to student network. Since these soft-valued outputs will have more information about the input than discrete values, the smaller network can converge on the same dataset with fewer iterations and maintaining accuracy. In this paper, baseline-2 and compressed-2 models are teacher and student networks, respectively. The filter sizes of baseline-2 and compressed-2 models are the same, whereas the numbers of kernels for convolutional layers are reduced to 72, 128, 192, 192 and 128.

4. Experimental Results

In this section, experimental conditions and results are presented. First, experimental conditions including the dataset and specifications for hardware and software are described. Second, driver drowsiness detection accuracy and execution speed are discussed in detail. To demonstrate the effectiveness of the proposed algorithms, we give a comparison with experimental results on Faster R-CNN [11], which is one of the well-known algorithms for simultaneous object detection and recognition, on our dataset.

4.1. Dataset

There are publically accessible datasets for drowsiness detection. One is DROZY database [34], which contains multiple types of drowsiness-related data including signals such as EEG, EOG, ECG, EMG and near-infrared (NIR) images.

The dataset contains data of only 11 subjects, which is very few to train on CNNs. Also, sensor patches attached on the subjects' faces for collecting electronical bio-signals are shown in the image data, which can interrupt accurate recognition on Computer Vision-based approaches. These limitations make it difficult to apply this dataset on the model proposed in this paper. Another dataset is NTHU Driver Drowsiness Detection dataset [35] which contains both RGB and IR videos in various driving scenarios. However, camera angle and class label are different with our experimental environment. Thus, our experiment was performed on the custom dataset that we collected on our own. Custom dataset was collected using Logitech C920 HD Pro Webcam. All the images were recorded in 640x480 resolutions.

Participants were asked to act 3 states of behavior including normal, yawning and drowsy, exhibiting each behavior for 20 ~30 seconds. The complete video was stored in individual frames forming a total of more than 70,000 images. The total number of subjects is 33 including 11 people with glasses. For further robustness of the system, the dataset includes subjects of diverse ethnic groups and gender and the participants were asked to change the head pose. The 33 subjects were split into 25 for training, 4 for validation and 4 for testing. Each of the validation and test dataset has 2 subjects with glasses and two of them without it. No two of the train, validation and test sets have a common subject; this will provide better interpretation of results. Figure 4 shows example images of the custom data.

Figure 4: Examples of custom dataset

4.2. Hardware and Software Environments

In our experiments, GTX 1080 GPU was used for training and testing, whereas embedded board NVIDIA Jetson TK1 was used for deployment. GTX 1080 has 2560 CUDA cores, each core with a base clock speed of 1607MHz and boost clock of up to 1733MHz; it can do computations up to 9 TFLOPS (Tera floating point operations per second). The PC has i7-2600 CPU which can clock till 3.4GHz and 16GB of RAM. Jetson TK1 has 192 CUDA cores, each core with a base clock speed of 870MHz and it can do computations up to 326 GFLOPS (Giga FLOPS), which is 26 times lesser when compared to GTX 1080. The device has Cortex-A15 CPU. TK1 consumes 2.2 Watt of power to run the deep learning on it. The device is much cheaper than the other GPU boards for embedded purpose and consumes less power. Thus, Jetson TK1 is appropriate to deploy the proposed models on motor vehicles as a small unit.

Ubuntu 14.04 for OS and Caffe for deep learning framework were used. The complete dataset is stored in hd5 format for easy and fast training. No pre-training is done to the proposed models and the training is done completely from scratch. Starting learning rate of 0.01 is taken since the initial weights are all randomly initialized. 40 epochs of training is done, by decreasing the learning rate by a factor of 0.1 for 10 epochs.

4.3. Results on the proposed and benchmark models

In section 3.1, the proposed models were described. To validate performance of the models, accuracy of driver's drowsiness detection and execution speed on GPU boards were tested under various conditions. The following subsections describe results of the proposed and benchmark models.

4.3.1 Baseline-4 Model

As described in section 3.1.1, the baseline-4 model is a 4-stream network with each network receiving the two eyes, mouth and face crops as the input individually. The model size was 56MB on disk, with the deployed model taking nearly 600MB of GPU memory.

Execution times to run baseline-4 model are 3.4ms (milliseconds) and 88.5ms on GTX 1080 and Jetson TK1, respectively. Including face detection and alignment end to end speeds of approximately 72.0fps (frames per second) and 6.1fps are achieved on GTX 1080 and Jetson TK1, respectively. During the training phase we achieved a validation accuracy of 91.6%. The model when run on the test subjects reported an accuracy of 91.3%.

		Faster RCNN (VGG-16)	Faster RCNN (AlexNet)	Baseline-4 (ours)	Baseline-2 (ours)	Compressed-2 (ours)
Accuracy (%)	Validation	76.6	70.9	91.6	**94.8**	91.2
	Test	90.5	82.8	91.3	**93.8**	89.5
Compression (MB)	Model size	547	236	56	28	10
	GPU Memory	3183	845	600	443	353
Drowsiness Detection time (ms)	GTX 1080	-	-	3.7	2.3	**1.4**
	Jetson TK1	-	-	88.5	28.4	**18.9**
Overall speed (fps)	GTX 1080	9.1	22.7	72.0	82.0	**90.1**
	Jetson TK1	-	1.1	6.1	12.5	**14.9**

Table 1: Summary of overall experimental results ('-' means 'not applicable')

4.3.2 Baseline-2 Model

The baseline-2 model is proposed in section 3.1.2 where it adopts 2-stream network instead of 4-stream and left-eye and mouth are used as inputs. Since the number of streams has been reduced, disk and GPU usage has been reduced. The model takes 28MB in disk space and 443MB in GPU memory while testing.

2.3ms and 28.4ms are taken to run the baseline-2 model on GTX 1080 and Jetson TK1 respectively. Speeds for the end to end process (including face detection and alignment) are 82.0fps and 12.5fps on GTX 1080 and Jetson TK1, respectively. The baseline-2 model achieved a validation accuracy of 94.8% and a test accuracy of 93.84%.

4.3.3 Compressed-2 Model

The final model is the compressed-2 model stated in section 3.1.3 where the model has similar structure to baseline-2 model. Size of this Model is 10MB. This is 3 times smaller than its teacher network. It takes 353MB of GPU memory for testing.

Run times for running the compressed-2 model on GTX 1080 and Jetson TK1 are 1.4ms and 18.9ms, respectively. Speeds for the end to end process are 90.1fps and 14.9fps on GTX 1080 and Jetson TK1, respectively. The validation accuracy during training is 91.2% and the test accuracy is 89.5%.

4.3.4 Benchmark model

To evaluate the proposed models, Faster RCNN proposed in [11] is used to compare performance. Faster RCNN models are trained and tested using the custom dataset used in experiments for the proposed model. To compare with the above proposed networks, VGG-16 [38] and AlexNet [16] architectures for faster-RCNN are used. VGG16 based faster-RCNN takes 547MB of disk space, requiring approximately 3GB of GPU memory during run time. AlexNet based faster-RCNN takes 236MB of disk space with 845MB of GPU memory during run time.

Faster-RCNN takes end to end run time of 9.1fps and 22.7fps on GTX 1080 GPU with VGG16 and AlexNet based architectures. Since VGG16 architecture based faster-RCNN model takes more than 2GB of memory it is not possible to deploy on TK1 board, even though it has a reported 90% test accuracy on the dataset. Whereas AlexNet architecture can be deployed on TK1, but it runs really slow at 1.1fps. This makes it impractical for faster-RCNN models to run on embedded board.

4.4. Discussion

The summary of the overall experimental results are shown in Table 1. Experiments were performed using three types of models which are baseline-4, baseline-2 and compressed-2 models .Faster RCNN with VGG-16 and AlexNet architectures are used as benchmarks for comparison study.

In terms of accuracy, baseline-2 model achieves the best among all the proposed models at 93.8%. The best accuracy of faster-RCNN model is for VGG based architecture at 90.5%.

Compressed-2 model has the smallest model size and GPU memory usage. The disk size of compressed-2 model is about 52 times smaller than VGG16 based faster-RCNN and consumes 10 times lesser GPU memory. By maintaining the accuracy of nearly 90%, the compressed-2 model has a 3 times lesser disk space than baseline-2 model.

As we can see in Table 1, compressed-2 model is the fastest among the models. If we consider only drowsiness detection, compressed-2 model takes 2.6 and 4.5 times lesser time than baseline-4 on GTX 1080 and TK1 respectively. When comparing compressed-2 with baseline-2, the time reduced by 1.6 and 1.5 times on GTX 1080 and TK1.

In terms of end to end speed, compressed-2 is still the fastest version. The overall speed of compressed-2 model when compared with baseline-4 is 25% faster on GTX 1080 and 144% faster on Jetson TK1. The percentages are 10% and 19% when we compare between compressed-2 and baseline-2. The numbers are not as significant as the drowsiness only speed because face detection is the main bottleneck in this problem. The proposed model successfully outruns faster-RCNN (AlexNet based architecture) by 4 times and 13.5 times on GTX 1080 and TK1.

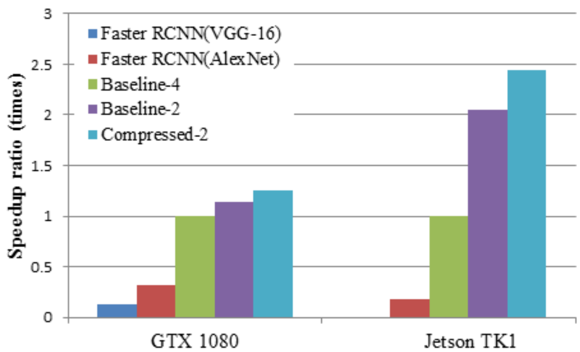

Figure 5: Speedup ratio of models on GTX 1080 and Jetson TK1

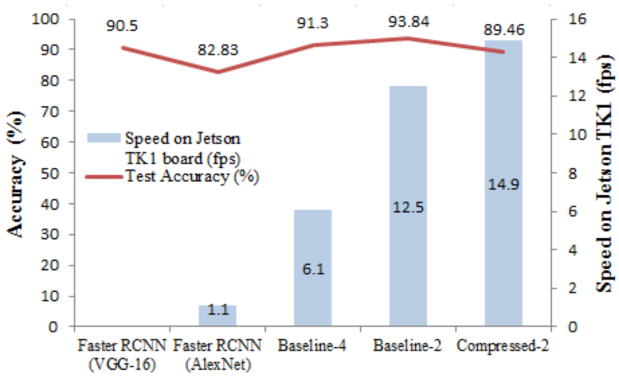

Figure 6: Comparison of speed (fps, blue-bar graph, right axis) and the test accuracy (%, red-line graph, left axis)

Figure 5 shows speed comparison graphs on both GPU boards. Here speedup is calculated based on the speed of baseline-4 model. Improvements of baseline-2 and compressed-2 models are larger on Jetson TK1 than on GTX 1080. It means that the proposed models are more effective to improve speed on embedded boards than high performance devices. Generally, accuracy and speed have a trade-off. Nevertheless, speed improvement is achieved by using the proposed models while maintaining reasonable accuracy as shown in Figure 6.

5. Conclusion and Future Works

In this paper, highly optimized deep neural network model for driver's drowsiness detection is designed and compressed for embedded system. The minimum facial landmarks are utilized as inputs to detect driver's drowsiness and a compression technique of knowledge distillation is applied to be implemented on real-time embedded system.

The experimental results under various circumstances supported possibility of implementation for real-time driver's drowsiness detector. Results showed that eyes and mouth play the major roles in drowsiness classification. Use of an eye and mouth gives additional accuracy of 3% comparing to that of eyes, mouth and face. This can happen when the model tries to learn unnecessary data from face. The results thus conclude that our optimized deep neural networks model can be used for driver's drowsiness detection on embedded devices with a high accuracy for safety with Advanced Driver Assistance System (ADAS) and Driver Monitoring System (DMS).

As a future work, Infra-red camera can be used to capture driver's behavior at night situation. Moreover, casual hear-rate sensor and image can be analyzed as multimodal deep learning approach and more recent model compression and knowledge distillation techniques can be adapted to reducing runtime more.

References

[1] Drowsy Driving NHTSA reports. (2017, June 02). Retrieved from https://www.nhtsa.gov/risky-driving/drowsy-driving.

[2] K. Fagerberg, "Vehicle-based detection of inattentive driving for integration in an adaptive lane departure warning system Drowsiness detection," M.S. thesis, KTH Signals Sensors and Systems, Stockholm, Sweden, 2004

[3] J. Krajewski, D. Sommer, U. Trutschel, D. Edwards and M. Golz, "Steering Wheel Behavior Based Estimation of Fatigue", in Proceedings of the Fifth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, pp. 118-124

[4] K. Mattsson, "In-vehicle prediction of truck driver sleepiness. Lane positions variables," M.S. thesis, Division of Media Technology, Dept. of Computer Science and Electrical Engineering, Lulea Univ. of Technology, Sodertalje, Sweden, 2007.

[5] H. Malik, F. Naeem, Z. Zuberi, and R. ul Haq, "Vision based driving simulation," in Proc. 2004 Int. Conf. Cyberworlds, 18–20 Nov. 2004, pp. 255–259

[6] Driver Alert Control (DAC). (2016, Feb 10). Retrieved from http://support.volvocars.com/uk/cars/Pages/owners-manual.aspx?mc=Y555&my=2015&sw=14w20&article=2e82f6fc0d1139c2c0a801e800329d4e

[7] Z. Mardi, S. N. Ashtiani, and M. Mikaili, "EEG-based drowsiness detection for safe driving using chaotic features and statistical tests," J. Med. Signals Sens., vol. 1, pp. 130–137, 2011.

[8] T. Danisman, I.M. Bilasco, C. Djeraba and N. Ihaddadene, "Drowsy driver detection system using eye blink patterns," Universite Lille 1 & Telecom Lille 1, Marconi, France, 2010.

[9] B. Hariri, S. Abtahi, S. Shirmohammadi, and L. Martel, "A yawning measurement method to detect driver drowsiness," Distrib. Collab. Virtual Environ. Res. Lab., Univ. Ottawa, Ottawa, ON, Canada, 2011

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. IEEE Conference on Computer Vision and Pattern Recognition, 2015.
[11] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In NIPS, 2015.
[12] B.-K. Kim, J. Roh, S.-Y. Dong, and S.-Y. Lee, Hierarchical committee of deep convolutional neural networks for robust facial expression recognition, Journal on Multimodal User Interfaces, 10 (2016), pp. 173–189.
[13] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 3431-3440.
[14] K. Dwivedi, K. Biswaranjan, A. Sethi, "Drowsy Driver Detection using Representation Learning," Advance Computing Conference (IACC), 2014 IEEE International, 21-22 Feb. 2014.
[15] S. Park, F. Pan, S. Kang and C. D. Yoo, "Driver drowsiness detection system based on feature representation learning using various deep networks," The ACCV Workshop on Driver Drowsiness Detection from Video 2016, Taipei, Taiwan, ROC, 2016.
[16] Krizhevsky, A., Sutskever, I., Hinton, G. E.: Imagenet classification with deep convolutional neural networks. NIPS, (2012) 1097–1105
[17] Parkhi, O. M., Vedaldi, A., Zisserman, A.: Deep face recognition. BMVC, 1 (2015) 6
[18] Donahue, J., Anne Hendricks, L., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K., Darrell, T.: Long-term recurrent convolutional networks for visual recognition and description. CVPR, (2015) 2625–2634
[19] S. Han, H. Mao, and W. J. Dally, "Deep Compression: Compressing Deep Neural Network with Pruning, Trained Quantization and Huffman Coding," Computing Research Repository (CoRR), 2015.
[20] P. Gysel, M. Motamedi, and S. Ghiasi, "Hardware-Oriented Approximation of Convolutional Neural Networks," ICLR Workshop 2016.
[21] M. Courbariaux, Y. Bengio, Binarynet: Training deep neural networks with weights and activations constrained to +1 or -1, CoRR abs/1602.02830, 2016
[22] Deep Networks on classification tasks using Torch. (2016, March 21). Retrieved from https://github.com/itayhubara/BinaryNet
[23] M. Rastegari, V. Ordonez, J. Redmon, and A. Farhadi. Xnornet: Imagenet classification using binary convolutional neural networks. In ECCV, pages 525–542. Springer, 2016.
[24] M. Jaderberg, A. Vedaldi, and A. Zisserman. Speeding up convolutional neural networks with low rank expansions. In BMVC, 2014.
[25] G. Hinton, O. Vinyals, and J. Dean, "Distilling the Knowledge in a Neural Network, " NIPS 2014, Dec. 2014.
[26] R. Caruana, A. N. Mizil, G. Crew, and A. Ksikes, "Ensemble Selection from Libraries of Models," ICML '04, 2004.
[27] J. Ba, and R. Caruana, "Do Deep Nets Really Need to be Deep," NIPS 2014, Dec. 2014.
[28] B. B. Sau, and V. N. Balasubramanian, "Deep Model Compression: Distilling Knowledge from Noisy Teachers," Computing Research Repository (CoRR) 2016.
[29] A. Romero, N. Ballas, S. E. Kahou, A. Chassang, C. Gatta, and Y. Bengio, "FitNets: Hints for Thin Deep Nets," Computing Research Repository (CoRR) 2014
[30] A. Romero, N. Ballas, S. Ebrahimi Kahou, A. Chassang, C. Gatta and Y. Bengio, "FitNets: Hints for Thin Deep Nets", ICLR 2015.
[31] B. B. Sau and V. N. Balasubramanian, "Deep Model Compression: Distilling Knowledge from Noisy Teachers," arxiv.org, 2016.
[32] K. Zhang, Z. Zhang, Z. Li, "Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks", IEEE Signal Processing Letters, 2016
[33] K. Krafka, A. Khosla, P. Kellnhofer, H. Kannan, "Eye Tracking for Everyone", CVPR 2016
[34] Q. Massoz, T. Langohr, C. Francois, J. G. Verly, "The ULg Multimodality Drowsiness Database (called DROZY) and Examples of Use, WACV 2016
[35] Ching-Hua Weng, Ying-Hsiu Lai, Shang-Hong Lai, "Driver Drowsiness Detection via a Hierarchical Temporal Deep Belief Network", In Asian Conference on Computer Vision Workshop on Driver Drowsiness Detection from Video, Taipei, Taiwan, Nov. 2016
[36] Szegedy, Christian, et al. "Going deeper with convolutions." CVPR 2015.
[37] Liao, Shengcai, Anil K. Jain, and Stan Z. Li. "Partial face recognition: Alignment-free approach." IEEE Transactions on pattern analysis and machine intelligence 35.5 (2013)
[38] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," ICLR 2015.