

Weakly-Supervised Degree of Eye-Closeness Estimation

Eyasu Mequanint, Shuai Zhang, Bijan Forutanpour, Yingyong Qi, Ning Bi
 Qualcomm AI Research *
 San Diego, CA, USA.

{emequani, shuazhan, bijanf, yingyong, nbi}@qti.qualcomm.com

Abstract

Following recent technological advances there is a growing interest in building non-intrusive methods that help us communicate with computing devices. In this regard, accurate information from eye is a promising input medium between a user and computing devices. In this paper we propose a method that captures the degree of eye closeness. Although many methods exist for detection of eyelid openness, they are inherently unable to satisfactorily perform in real world applications. Detailed eye state estimation is more important, in extracting meaningful information, than estimating whether eyes are open or closed. However, learning reliable eye state estimator requires accurate annotations which is cost prohibitive. In this work, we leverage synthetic face images which can be generated via computer graphics rendering techniques and automatically annotated with different levels of eye openness. These synthesized training data images, however, have a domain shift from real-world data. To alleviate this issue, we propose a weakly-supervised method which utilizes the accurate annotation from the synthetic data set, to learn accurate degree of eye openness, and the weakly labeled (open or closed) real world eye data set to control the domain shift. We introduce a data set of 1.3M synthetic face images with detail eye openness and eye gaze information, and 21k real-world images with open/closed annotation. The dataset will be released online upon acceptance. Extensive experiments validate the effectiveness of the proposed approach.

1. Introduction

Several advanced input technologies have been proposed to simplify user's interactions with computing devices. Information from an eye is one of the input techniques which improves the experience of working with a computer. While measuring a users visual line of gaze (where s/he is looking

in space) has been improving, the degree of eye closeness - which is rich in information for applications such as user-computer dialogue [18] - has not been well studied. Detection of human eyelid openness or blink state is a key step for effective eye-based vision systems. There are plenty of applications that require accurate eye states estimator: user-computer interaction [18], face authentication system where eye states are used for user's attention assessment and anti-spoofing [24, 32, 23], photography, deceit detection [22, 25], emotion analysis, eye tracking, avatar animation, gaming, virtual reality, and driver's drowsiness detection which help avoid impairment that leads to, according to The American National Highway Traffic Safety Administration (NHTSA), \$109 billion in damages annually.

Typically, a computer vision approach for eye states detection first extracts features around eyes part, and then classify the eye states. Existing approaches are designed only for binary eye states detection (open or closed). In an eye blink detection system, the process needs to collect multiple image frames as input. However, due to the speed of an eye blink, a fully eye-closed image may not be captured/sampled, and thus a binary state system could easily lead to incorrect decision due to missed blinks. A binary state is often insufficient for accurate user-computer dialogue and other similar applications that require higher speed and accuracy, such as photography, anti-spoofing, and others. For example, one of the goals of advanced driving assistance systems is early detection of driver's drowsiness. An alert is raised when the eyes are at least 80% closed over a certain time period [6]. As in prior examples, a two-state eye openness system is insufficient for high accuracy. Estimation of eyelid openness with more granularity allows for the extraction of more meaningful information for addressing these real-world applications. An example of detailed and binary eye states is shown in Figure 1.

In this work, we develop a deep neural network (DNN) based framework that can detect the degree of eye-openness with high granularity. It provides more accurate and detailed information than current binary states (open/closed) systems. Using deep learning for eye openness requires

*Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc.

highly granular and accurately annotated training data. Such training data is often scarce and cost prohibitive. To address this problem, we introduce a large data set of synthetic face images rendered using advanced graphics techniques with accurately controlled degree of eye openness (Figure 1), and a limited set of real face images with binary eye states labels. One issue that arises is the domain shift between these synthetically generated data vs authentic real world face images. To overcome this, we propose a weakly-supervised training method which utilizes the accurate annotation from the synthetic data, and weak annotation (open or closed) on recorded data for eye openness estimation. The contribution of our work is listed as follows:

- A computer vision based system to detect eye openness with high granularity for several applications such as human computer interaction. Our approach achieves high granularity results from low-granularity, binary labeled (opened or closed eye) real-world images.
- Augmenting, using weakly-supervised learning, the real-world training images with binary annotation (opened or closed eye) with synthetically generated images with detail information on the degree of eye openness. We introduce 1.3M synthetic face images (Figure 1) and 21K real-world images.
- We conduct experiments which show that the proposed approach effectively estimates the degree of eye openness for real-world image with high accuracy and granularity.

2. Related Works

Several eye-based systems have been proposed in the literature which use the percent of closeness (PERCLOSE) and average eye closure speed (AECS) measures for different decisions, such as drowsiness detection where PERCLOSE increases [8, 10, 28, 30, 21, 9, 26, 7, 5] and AECS decreases [12, 3, 4], for a drowsy driver. Existing eye-based approaches mostly use eye and face detectors, such as Viola Jones algorithm [33], and detect the eye state using classical computer vision techniques. [8] trained a Support Vector Machine (SVM) for eye state classification. Tomas *et al.* [11] divided the eye region into 3×3 cells where local motion vectors are estimated whose variance of the vertical components is used to determine the eye state. [10] detected the eye states equalizing the eyes using a Hat transformation followed by eye tracking strategy in a sequence of frames. [24] introduced an appearance based image feature to detect the eye openness using the AdaBoost algorithm. [16] detect the eye states by analyzing the response of a horizontal Laplacian filter around the eyes; numerous vertical

line segments should be visible, due to the pupils and eye corners, when the eyes are open, and only horizontal lines are observed when eyes are closed. [20] detects the closed and open eye states based on the number of black pixels the eye has; in a binarized eye region, closed eye image has higher number of black pixels compared to an open eye image. [31] first detects 98 facial landmarks and the average height-width eye ratio is used to determine the eye's state in a given frame. A real-time eye state detector, designed for very low near-infrared image, is proposed in [19]. More recently, [13] introduced deep learning into the field of fatigue detection. The method detects the face and feature point locations using multi-task cascaded convolutional neural network (MTCNN). The eye region is obtained according to the geometric relationship between eye feature points, and then eye state is classified by convolutional neural network (CNN).

All the above referenced works and other several eye state based systems detect only two levels of eye opening, which is not enough to model a practical system. From a perspective of practical real world system applicability, the eye state detection system should satisfy several constraints which could not be solved just by using only the two eye states. Drowsiness is a very good practical example which is a state a driver might be in with a partially closed eye. Very few works introduced percentage of eye openness which is more accurate than methods that detects only two levels of openness [2, 15, 1]. [2] just added a third level (partially opened) to the two eye states (opened and closed). [1] and [15] are drowsiness detection methods that uses the notion of percentage of eye openness (various states of eye openness). Both of them, to detect detailed eye states, use classical computer vision techniques. [1] is a geometry shape-based approach which uses Circular Hough Transform method to localize iris and eyelids. Since it's a geometric-based approach, a very small variation in eyelids localization leads to a wrong decision, and it also easily gets affected by illumination variation. Our approach, since we use deep learning, does not need iris and eyelids detection and is more robust for illumination variations. [15] is a video-based solution for eyelids movement detection. Classical approaches are used to detect the face and eye and only the left eye part is then taken as input to the next stage which vectorizes the input, does dimension reduction and input the result to a single linear model which detects the eye openness score. The eye openness score is then passed to a clustering module which helps the system detect some pattern based on which eyelids movement is detected. The method is not a general model, not fully automated and is subject dependent, and the structure of features extraction scheme needs to be defined by the user. It needs different levels of feature clustering whose number of clusters should be known for different level of feature extractions. The sys-

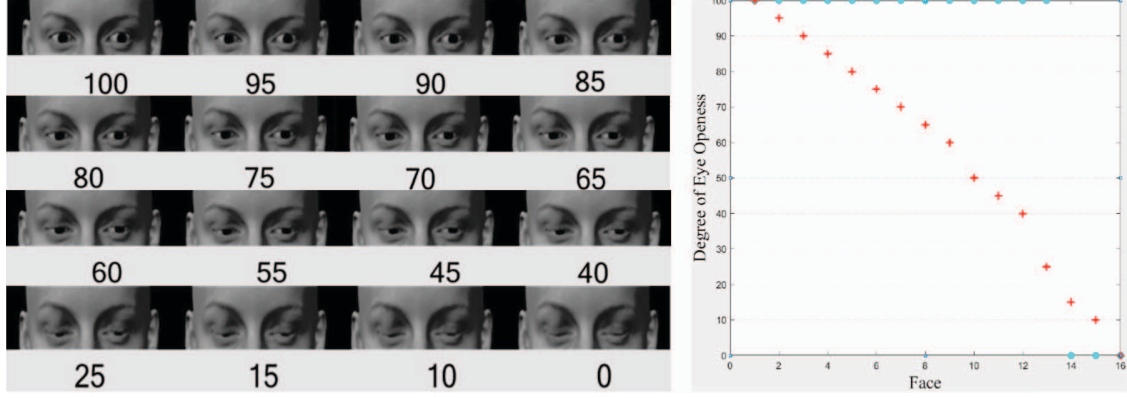


Figure 1. **Left:** eye portion of synthetic faces with labeled degree of openness (100 and 0 refers fully open and fully closed respectively). **Right:** cyan simulates results of available eye openness detection and red simulates results from our proposed approach.

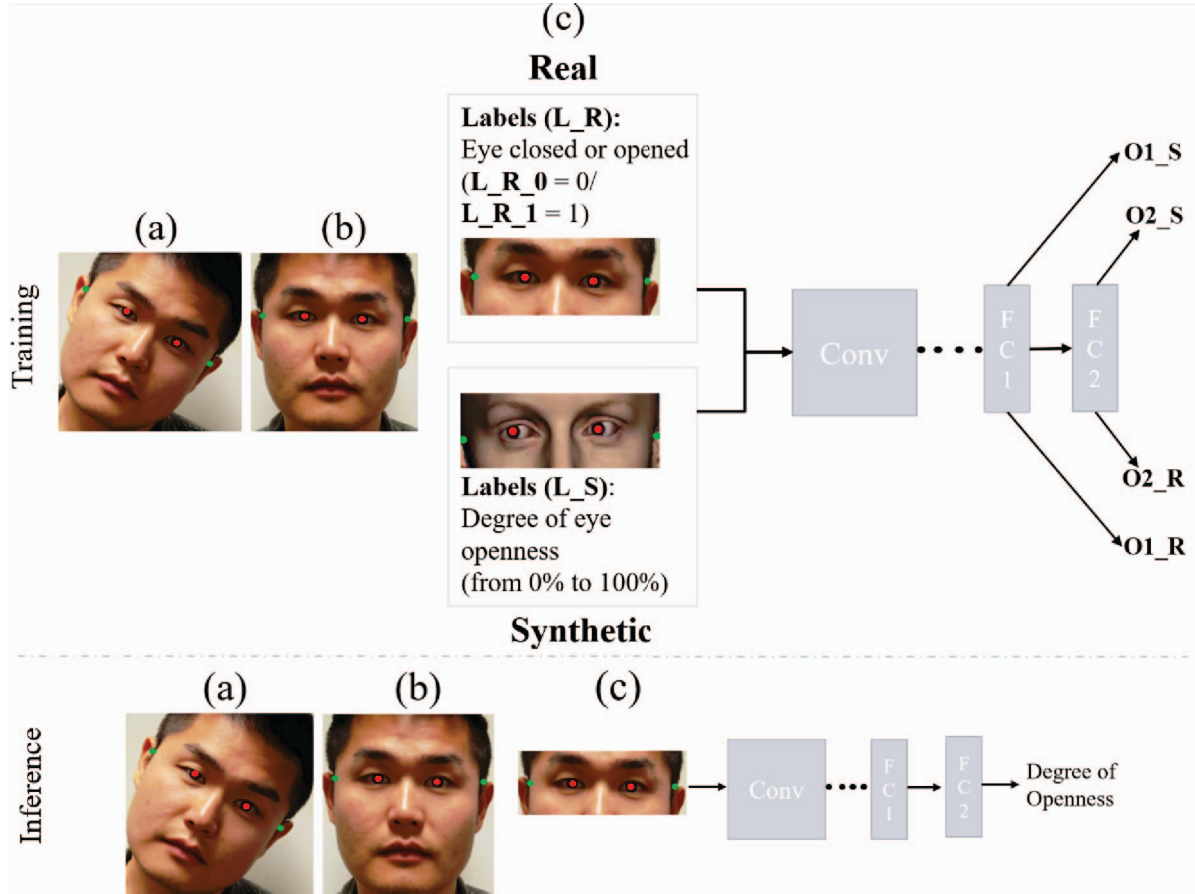


Figure 2. The proposed architecture for estimating degree of eye openness (training upper and inference bottom). (a) raw face image with landmarks (could be real or synthetic), (b) normalized version of the face, and (c) last stage of the preprocessing, contains cropped eye portion of the face. The preprocessed real and synthetic data are separated into two different groups with their corresponding labels (L_R and L_S). 'Conv', 'FC1' and 'FC2' represent a shared convolution block, and two fully connected blocks respectively. The output of FC2 estimates the degree of eye openness. $O1_S$ = output of the synthetic data at FC1, $O1_R$ = output of the real data at FC1, $O2_S$ = output of the synthetic data at FC2, $O2_R$ = output of the real data at FC2. $O2$ is a scalar which represents openness amount and $O1$ is feature vector of size 256.

tem, in general, has 6 parameters to tune which make it harder to be used in fully automated systems.

3. Proposed approach

We propose a deep learning solution to estimate the degree of eye openness. One of the things which make solving this problem using deep learning difficult is data. The high performance of deep learning results from abundant labelled training data. Collecting a data with accurate degree of eye openness is a difficult task. To best serve our purpose we used a synthetic data with different levels of eye openness. Using synthetically generated data, though solves the data scarcity, has a problem in training a general-purpose deep learning model for our task. Models trained on the synthetic data, due to the domain gap the data has with the real data, fail when tested on a real dataset. Domain adaptation studies the domain shift problem for the better use of available training data for new testing domains. In this work we propose to train a model, using a synthetic data that have known levels of eye openness and real data that only have open/closed annotation, which help us get the different eye states on real test dataset addressing the domain shift our data has with the synthetic data.

During training, the input batch to the network contains images from both the synthetic and real ones. The network, as shown in figure 2, consists of an input which comes from both synthetic and real data blocks, a convolution block (Conv) and two fully connected blocks (FC1 and FC2). Given an image, in the first step we detect the face and landmarks utilizing dlib toolkit [17]. In step (b) we align and normalize the face based on the landmarks shown by red and green dots, and the eye portion of the face is cropped to be passed through the network. The output of the FC2 block regresses the degree of eye openness (0 means closed, and positive numbers represent different levels of eye openness). The real data has no detail eye openness annotation. It just has information if the eye is closed or open, a kind of dataset much cheaper than a dataset with labeled degree of eye openness. We train the network leveraging all the available information both from the synthetic and real data. To this end, we propose a loss combined from three different losses.

Architecture: Very light network architecture based on Max-Feature-Map (MFM) operation, neural inhibition operation proposed in [35], is used. MFM operation is a special case of maxout [14] to learn a light convolutional neural network (CNN) with a small number of parameters. It is an alternative of ReLU which adopts a competitive relationship to suppress low-activation neurons in each layer. It not only is able to separate noisy and informative signals but also does feature selection between two feature maps. We refer the reader to [35] for better understanding of the MFM operation. 'Conv' (Figure 2) is constructed by 5 convolu-

tion layers with Max-Feature-Map operations and 3 max-pooling layers, shown in Figure 3. The 256-D deep features are extracted from the output of fully connected layer after MFM operation (layer MFM 6).

Loss Functions: The main challenge for training our model is the lack of ground truth for the real data. To address the problem we use a combination of three losses which leverages recent ideas from the problem of domain adaptation [29].

First, mean square error (MSE) loss is used, for inputs from the synthetic data, to reduce the gap between the regressed degree of eye openness and the ground truth labels, see *Loss1* in the follow paragraph. This loss is the key part to facilitate the proposed weakly supervised training, it relieve the painful detailed granular eye openness annotations for real-world dataset.

The second loss uses information we have for the real data, eye is closed or open, please see *Loss2*. If the real input data is labelled as closed, the output prediction from our network should be zero, otherwise the network should output a number greater than a predefined threshold, openness threshold (OT). OT = Openness threshold is a hyper parameter, used with the formula:

$$\begin{cases} 1 \text{ (open eyes), if } \mathbf{O2_R} > \mathbf{OT}; \\ 0 \text{ (closed eyes), Otherwise;} \end{cases}$$

Recent approaches which train a network to bring the source and target distribution together show excellent performance [29]. This inspired us to introduce a distribution loss, a loss which help us train our network using the gradients from the change in the distribution from the synthetic and real input data, see *Loss3*. This helps us bring the source and target distributions closer in the feature space learned by the network.

The proposed loss functions for cross domain (synthetic and real) network training consists of the following three terms:

- **Loss1:** MSE loss for accurately predicting level of openness on synthetic data

$$Loss1 = \text{MSE}(\mathbf{O2_S}, \mathbf{L_S}),$$

where $\mathbf{O2_S}$ is the estimated eye degree of synthetic data and $\mathbf{L_S}$ is the accurately labeled eye degree, see Fig 2;

- **Loss2:** Binary loss for accurately predicting binary (opened/closed eye) labelled real data with OT

$$Loss2 = \frac{1}{N} \sum_i \left\{ \|\mathbf{O2_R}_i\|^2 * (1 - \mathbf{L_R}_i) + \max((\mathbf{OT} - \mathbf{O2_R}_i), 0) * \mathbf{L_R}_i \right\},$$

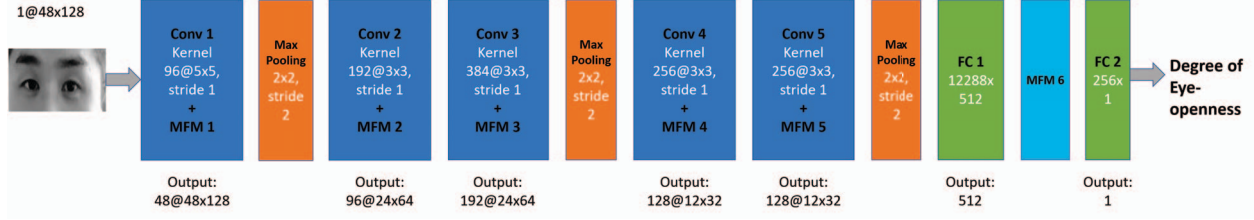


Figure 3. Model architecture.

where $O2_R$ is the estimated eye degree of real image data and L_R are the corresponding binary open/close labels. OT is the eye openness threshold;

- **Loss3:** Distribution loss for controlling domain mismatch between synthetic and real data; the distribution from synthetic and real data should be similar.

$$Loss3 = abs(mean(O1_S) - mean(O1_R)) + abs(var(O1_S) - var(O1_R)),$$

where $O1_S$ and $O1_R$ are the feature vectors of synthetic/real data from the 'FC1' layer of the model, see Fig 2.

The final loss is computed as

$$Loss = \lambda_1 Loss1 + \lambda_2 Loss2 + \lambda_3 Loss3.$$

4. Q_ECE: Eye Openness Estimation Dataset

Although problems related to eye-based systems have received a lot of attention, their performance still has not reached an acceptable level for practical use cases. The lack of high quality training data is one of the issues which limits the development of such systems. Existing databases, such as ZJU [24], Eyelink8 [11] and Silesian5 [27] lack the information necessary to address essential challenges. The datasets mentioned above do not take into account important characteristics such as human pose or image illumination. Furthermore, the samples are captured from a limited number of subjects. Another requirement of high quality datasets is high quality annotation of the data. In order to work well in real-world conditions, most eye-based systems require knowledge of all possible eye states. One of the main reasons existing eye-based approaches do not perform well for real-world applications is the low number of 'eye states' they use. This is typically only two: closed or open. Many existing real-world data sets require additional annotation, which can be cost prohibitive.

To alleviate the data annotation burden we create a dataset of 1.3M synthetic data by rendering face images using computer graphics techniques. The dataset was created using high quality 3D scans of 13 human head models 6 female and 7 male from different age groups and ethnicities

¹. Eyeball models were separate, and placed in the 3D eye sockets. Additionally, the subjects' eyelids were animated in conjunction with the up and down rotation of the eyeball. The dataset included 198 eye directions, via look-at target points (11 vertical x 18 horizontal, in a grid pattern), +/- 25 degrees vertical, +/- 35 degrees horizontal, in 5 degree increments. In addition, there were 11 different states of eye openings (100 % open to close in 10 % increments). Finally 49 camera positions were generated (-30 to 30 degree in 10 degree increments, horizontally and vertically). Some exemplar images from this synthetic dataset are shown in Figure 4. Our work is the first to create and use such a large dataset of high quality rendered face images with controlled eyelid movement. Since the dataset contains eyeball rotation and 2D and 3D look at point information as well, the data is not only useful for eye openness estimation, but also for gaze estimation and attention detection.

In addition to the synthetic data we also collected a real data from 16 subjects, different age groups and ethnicity, using NIR and RGB sensors. In the real data collection we tried to consider several situations as pose, illumination, (sun) glasses and others. For the case of pose variations we asked the subject to move the head 360 degree and in four different directions (Left, right, up and down). We consider the illumination variation collecting the data from different environments as 'indoor full light', 'indoor low light', 'in a dark' (NIR sensor only), 'outdoor shade' and 'outdoor sunlight'. '(Sun) glass' and 'No (sun) glass' situations are also included in most of the subject's data. In total we collected around 21k real images (17k NIR and 4k RGB images) with more than 12k closed and around 9k opened eyes. We also collected, from four of the subjects' a data which covers detail eye states, we asked the subjects to close and open the eye very slowly. A good model which estimates detail eye openness should give us 'U' kind of shape on plotting the frames versus degree of eye openness, please see the result on figure 5

5. Experiments and results

Since, to the best of our knowledge, there is no available dataset and recently proposed methods related to detailed

¹The 3D scans could be found here: www.3dscanstore.com



Figure 4. Sample synthetic face models.

eye states estimation, we conduct experiments on Q_ECE dataset introduced in this work.

5.1. Implementation Details

We implemented our method with Pytorch. For all training we take 80 epochs with an initial learning rate of 0.0001 and a batch size of 256. In case of joint training the percentage of input data (from a batch) is fixed as 25% real and 75% synthetic. Based on the generated synthetic data, as shown in figure 1, and using some experimental validation, we observed that an openness threshold (OT) of 15 gives the best result. λ_1 is set to 0.01 and both λ_2 and λ_3 are set to 1. For all experiments with synthetic data, we learn eye openness estimation using the synthetic data branch only. For the real-world datasets we use a joint training with the synthetic dataset, any input batch to the network contains images from both synthetic and real images. For an effective transfer learn from synthetic data, we follow similar ideas from existing works [34, 36]; it is important that both the synthetic and real dataset have the same distribution of eye openness. In our model, gray-scale face images are used instead of RGB images. The face images are aligned to 144×144 by the landmarks and the eye portion of 48×128 is cropped and used as inputs to the 'Conv' layer. Besides, each pixel value is normalized to be between $[0, 255]$.

5.2. Results and Metrics

The performance of degree of eye openness is measured using the mean squared error (MSE) computed between the regressed degree of eye openness and the ground truth eye states. We believe that MSE is the best metric for our purpose (detail eye openness estimation). When we label the

dataset, to avoid the resolution and the camera distance issues, we consider the normalized face image based on detected landmarks; a person whose image is taken from different camera distances and different sensors with different resolution should have similar openness amount for similar eye states. The openness amount from our data ranges from 0 (fully closed) to 100 (fully open). 100 (fully open eye state) is given to the biggest eye from all eyes in our dataset. That means if we test using a bigger eye than the biggest eye from our dataset, the eye openness estimation result will be beyond 100, and openness amount of a fully open small eye will be much smaller than 100. The performance of eye openness and closeness is measured using accuracy metric.

Training data	Test data	Degree of eye openness (MSE)	Open/Close (Accuracy)
Synthetic	Synthetic	9	100
Synthetic	Real	—	47.5
Synthetic + Real	Synthetic	9	100
Synthetic + Real	Real	—	99.62
Real	Synthetic	8094.8	52.3
Real	Real	—	96.30

Table 1. Results on Q_ECE dataset, with varying training source

Table 1 shows quantitative results for the various input configurations of our eye openness estimation network. The degree of eye openness, evaluated using MSE metric, is same with the two input configurations. The result tells us that the degree of eye openness for the synthetic test data deviates, on average, from the ground truth only by 3%. Considering 8% variations, in degree of eye openness, from the

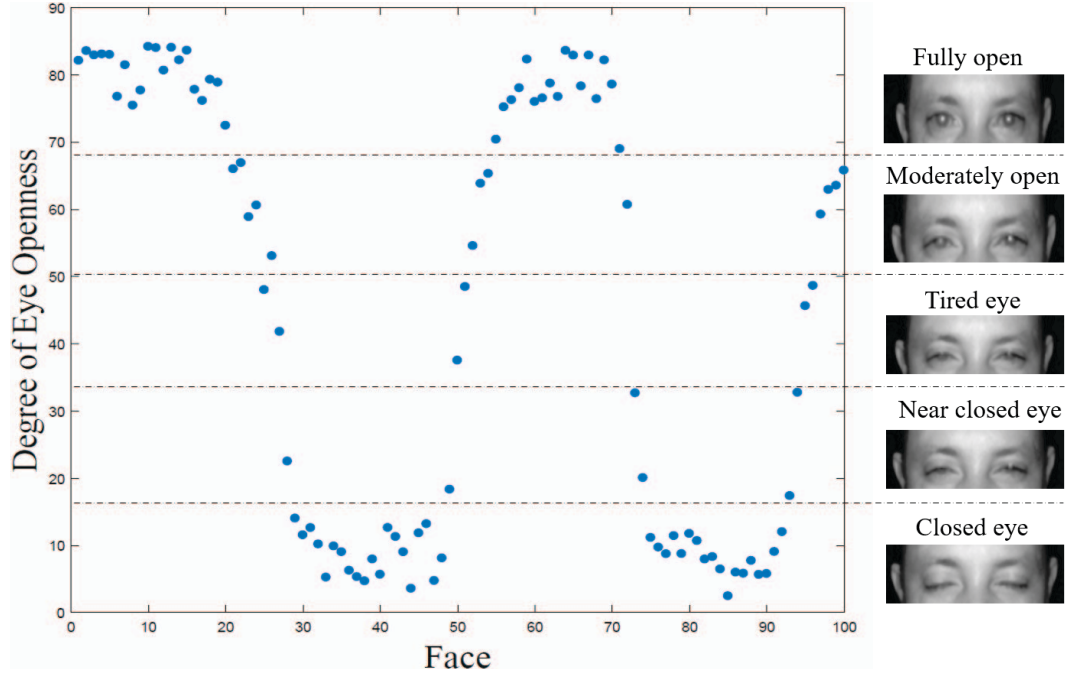


Figure 5. Performance on (100) video frames captured with "Close-Open-Close-Open" sequence moving the eyelids very slowly.

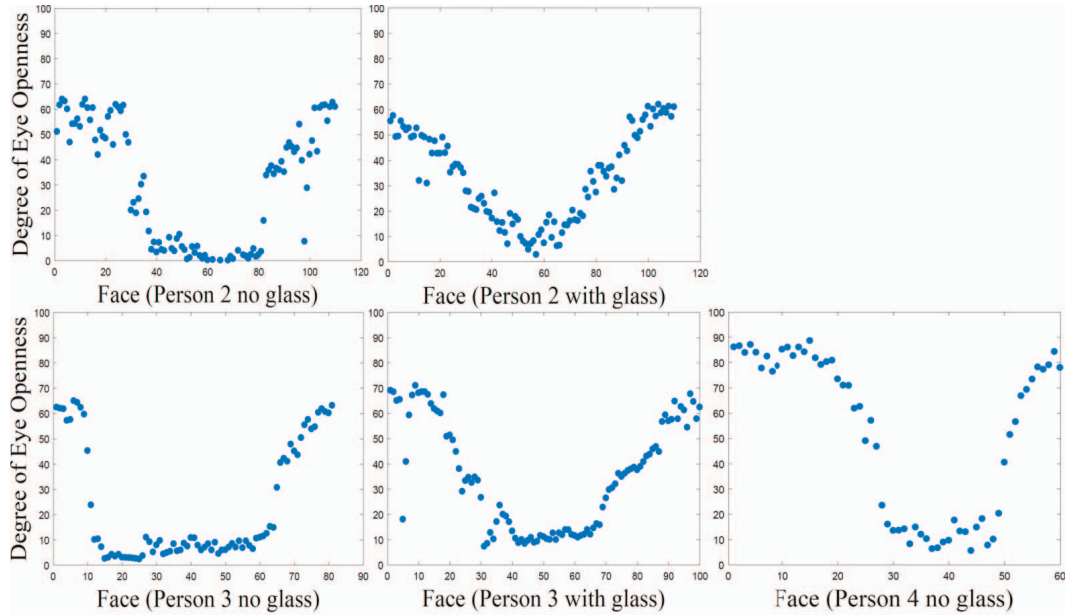


Figure 6. Performance on a video with "Close-Open" sequence moving the eyelids very slowly.

ground truth annotation as correct eye openness estimation we end up with 100 % accuracy in estimating the openness or closeness of the eyes. We observe that providing the real dataset together with the synthetic as input to our network (for joint training) results improved accuracy of eye openness and closeness detection on the real data. Training only using one (real or synthetic) and testing on the other end up

with a result close to random (open/close) decision. It is also noticed that the joint training, since it helps to augment training samples and regularize model from overfitting, improves the accuracy on the real-real experiment scenario.

We also evaluated our model by assessing its performance on video sequences collected from four of the subjects which we have asked to close and open their eye with

very slow motion which help us have various eye states. Figure 5 shows an example which is plotted from the predicted degree of eye openness from the video frames. A blue point represents the degree of eye openness (y-axis) of a face (x-axis). The entire process of closing and opening the eye could be represented by a curve with ‘U’ kind of shape. This help us extract meaningful information which could be leveraged for different applications. As can be seen from the figure, our model was able to capture the detail eye states, and we could divide the states (based on the degree of openness on the y-axis) into different meaningful information as ‘fully open’, ‘moderately open’, ‘tired eye’, ‘near closed eye’ and ‘closed eye’.

Experimental results on the other three subjects whose data is collected with “Close-Open” sequence moving the eyelids very slowly is shown on figure 6. As shown from the figure the proposed framework is able to capture detail eye openness states for all the persons with and without glass.

The last experiment which we conducted is using the small subset of our real dataset that are annotated with detail eye openness, let us call it Real’. Real’ consists of 2000 images. We first compute the red and green points as shown on figure 2 and used them to align and normalize the face and then we annotate the upper and lower eyelid points which is used to compute degree of eye openness that is used as our ground truth. 75% of the data is used to fine-tune the model trained using ‘Synthetic + Real’ and the rest of the data is used for testing. During training, since in this case we have detail eye openness annotations, we compute the loss for this dataset as losses that consider synthetic data annotations. The performance is then measured using ‘MSE’ metric. As shown on Table 2 the joint training help us boost the test result on real-world images. Moreover, adding few samples with detail annotations has a minor improvement over the binary labelled real data.

Training data	Test data	Degree of eye openness (MSE)
Synthetic	Real’ (test)	2045.80
Synthetic + Real	Real’ (test)	45.35
Synthetic + Real + Real’ (train)	Real’ (test)	34.20

Table 2. Result on Real’ face database

6. Conclusion

In this work, we shed the light to the research field of degree of eye openness estimation which help us estimate detail eye states, a problem which has not been well studied. We have addressed essential issues of the problem in-terms of practical and theoretical contributions. First, we created fully annotated synthetic data for estimation of the degree of eye openness which release the burden of detail eye state annotation of real images. The dataset will be released online upon acceptance. Secondly, we introduce a weakly-

supervised problem of leveraging low-cost binary labelled (opened or closed eye) real images together with the synthetic data for accurate estimation of the degree of eye openness on the real images. To this end we also collected real data which considers different practical situations. The experiments verify that the proposed approach effectively estimates the degree of eye openness for real world image. The proposed method, leveraging the cheap synthetic images, adapt easily to a weakly labelled real-world images.

References

- [1] B. Akrouit and W. Mahdi. Spatio-temporal features for the automatic control of driver drowsiness state and lack of concentration. *Machine Vision and Applications*, 26(1):1–13, 2015.
- [2] E. R. Anas, P. Henríquez, and B. J. Matuszewski. Online eye status detection in the wild with convolutional neural networks. In *VISIGRAPP (6: VISAPP)*, pages 88–95, 2017.
- [3] L. Barr, H. Howarth, S. Popkin, and R. J. Carroll. A review and evaluation of emerging driver fatigue detection measures and technologies. *National Transportation Systems Center, Cambridge. US Department of Transportation, Washington. Disponível em; http://www.ecse.rpi.edu/~qji/Fatigue/fatigue_report_dot.pdf*, 2005.
- [4] L. M. Bergasa, J. Nuevo, M. A. Sotelo, R. Barea, and M. E. Lopez. Real-time system for monitoring driver vigilance. *IEEE Transactions on Intelligent Transportation Systems*, 7(1):63–77, 2006.
- [5] D. Borza, R. Itu, and R. Danescu. In the eye of the deceiver: Analyzing eye movements as a cue to deception. *J. Imaging*, 4(10):120, 2018.
- [6] R. G. D Dinges. PERCLOS: a valid psychophysiological measure of alertness as assessed by psychomotor vigilance. *TechBrief NHTSA*. Publication No. FHWA-MCRT-98-006, 1998.
- [7] T. Danisman, I. M. Bilasco, C. Djeraba, and N. Ihaddadene. Drowsy driver detection system using eye blink patterns. In *2010 International Conference on Machine and Web Intelligence*, pages 230–233. IEEE, 2010.
- [8] S. Darshana, D. Fernando, S. Jayawardena, S. Wickramanayake, and C. DeSilva. Efficient perclos and gaze measurement methodologies to estimate driver attention in real time. In *2014 5th International Conference on Intelligent Systems, Modelling and Simulation*, pages 289–294. IEEE, 2014.
- [9] A. Dasgupta, A. George, S. Happy, and A. Routray. A vision-based system for monitoring the loss of attention in automotive drivers. *IEEE Transactions on Intelligent Transportation Systems*, 14(4):1825–1838, 2013.
- [10] I. G. Daza, N. Hernández, L. M. Bergasa, I. Parra, J. J. Yebes, M. Gavilán, R. Quintero, D. F. Llorca, and M. Sotelo. Drowsiness monitoring based on driver and driving data fusion. In *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 1199–1204. IEEE, 2011.

- [11] T. Drutarovsky and A. Fogelton. Eye blink detection using variance of motion vectors. In *European Conference on Computer Vision*, pages 436–448. Springer, 2014.
- [12] C. Fors, C. Ahlström, P. Sörner, J. Kovaceva, E. Hasselberg, M. Krantz, J.-F. Grönvall, K. Kircher, and A. Anund. *Camera-based sleepiness detection: final report of the project SleepEYE*. Statens väg-och transportforskningsinstitut, 2011.
- [13] L. Geng, Z. Hu, and Z. Xiao. Real-time fatigue driving recognition system based on deep learning and embedded platform. *American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)*, 53(1):164–175, 2019.
- [14] I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, and Y. Bengio. Maxout networks. *arXiv preprint arXiv:1302.4389*, 2013.
- [15] W. Han, Y. Yang, G.-B. Huang, O. Sourina, F. Klanner, and C. Denk. Driver drowsiness detection based on novel eye openness recognition method and unsupervised feature learning. In *2015 IEEE International Conference on Systems, Man, and Cybernetics*, pages 1470–1475. IEEE, 2015.
- [16] J. Jiménez-Pinto and M. Torres-Torriti. Optical flow and drivers kinematics analysis for state of alert sensing. *Sensors*, 13(4):4225–4257, 2013.
- [17] D. E. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10:1755–1758, 2009.
- [18] A. Królak and P. Strumiłło. Eye-blink detection system for human–computer interaction. *Universal Access in the Information Society*, 11(4):409–419, 2012.
- [19] M. Lalonde, D. Byrns, L. Gagnon, N. Teasdale, and D. Laurendeau. Real-time eye blink detection with gpu-based SIFT tracking. In *Fourth Canadian Conference on Computer and Robot Vision (CRV 2007)*, 28-30 May 2007, Montreal, Quebec, Canada, pages 481–487, 2007.
- [20] W. O. Lee, E. C. Lee, and K. R. Park. Blink detection robust to various facial poses. *Journal of neuroscience methods*, 193(2):356–372, 2010.
- [21] B. Manu. Facial features monitoring for real time drowsiness detection. In *2016 12th International Conference on Innovations in Information Technology (IIT)*, pages 1–4. IEEE, 2016.
- [22] F. M. Marchak. Detecting false intent using eye blink measures. *Frontiers in psychology*, 4:736, 2013.
- [23] J. Oh, S.-Y. Jeong, and J. Jeong. The timing and temporal patterns of eye blinking are dynamically modulated by attention. *Human movement science*, 31(6):1353–1365, 2012.
- [24] G. Pan, L. Sun, Z. Wu, and S. Lao. Eyeblink-based anti-spoofing in face recognition from a generic webcam. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8. IEEE, 2007.
- [25] J. Peth, J. S. Kim, and M. Gamer. Fixations and eye-blinks allow for detecting concealed crime related memories. *International Journal of Psychophysiology*, 88(1):96–103, 2013.
- [26] B. G. Pratama, I. Ardiyanto, and T. B. Adji. A review on driver drowsiness based on image, bio-signal, and driver behavior. In *2017 3rd International Conference on Science and Technology-Computer (ICST)*, pages 70–75. IEEE, 2017.
- [27] K. Radlak, M. Bozek, and B. Smolka. Silesian deception database: Presentation and analysis. In *Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection*, pages 29–35. ACM, 2015.
- [28] K. Rezaee, S. R. Alavi, M. Madanian, M. R. Ghezelbash, H. Khavari, and J. Haddadnia. Real-time intelligent alarm system of driver fatigue based on video sequences. In *2013 First RSI/ISM International Conference on Robotics and Mechatronics (ICRoM)*, pages 378–383. IEEE, 2013.
- [29] S. Sankaranarayanan, Y. Balaji, C. D. Castillo, and R. Chellappa. Generate to adapt: Aligning domains using generative adversarial networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [30] W. Sheng, Y. Ou, D. Tran, E. Tadesse, M. Liu, and G. Yan. An integrated manual and autonomous driving framework based on driver drowsiness detection. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4376–4381. IEEE, 2013.
- [31] F. M. Sukno, S.-K. Pavani, C. Butakoff, and A. F. Frangi. Automatic assessment of eye blinking patterns through statistical shape models. In *International Conference on Computer Vision Systems*, pages 33–42. Springer, 2009.
- [32] M. Szwoch and P. Pieniażek. Eye blink based detection of liveness in biometric authentication systems using conditional random fields. In *International Conference on Computer Vision and Graphics*, pages 669–676. Springer, 2012.
- [33] P. Viola and M. J. Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004.
- [34] E. Wood, T. Baltrusaitis, X. Zhang, Y. Sugano, P. Robinson, and A. Bulling. Rendering of eyes for eye-shape registration and gaze estimation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3756–3764, 2015.
- [35] X. Wu, R. He, Z. Sun, and T. Tan. A light cnn for deep face representation with noisy labels. *IEEE Transactions on Information Forensics and Security*, 13(11):2884–2896, 2018.
- [36] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling. Mpiigaze: Real-world dataset and deep appearance-based gaze estimation. *IEEE transactions on pattern analysis and machine intelligence*, 41(1):162–175, 2019.