ADVANCES IN EYE BLINK DETECTION:

EVALUATING CONVOLUTIONAL NEURAL NETWORK BASED BLINK DETECTION ON REAL WORLD DATA

By

Andres Barragan

A Thesis Submitted to The Honors College

In Partial Fulfillment of the Bachelors degree
With Honors in

Computer Science

THE UNIVERSITY OF ARIZONA

M A Y 2 0 2 0

Approved by:

Andres Barragan

Honors Thesis

Dr. Kobus

ABSTRACT:

Subjects with varying races, wearing glasses, hair in the way of their eyes and periodically moving faces makes it difficult for convolutional neural networks (CNN) to detect eye blinks in scenarios such as detecting driver fatigue. However, if the accuracy of the CNN is too low, then using it for eye blink detection in life like scenarios would not be advisable because the CNN would be incorrectly classifying to many times. To increase the accuracy of the CNN researchers have tried varying race of subjects and situational profiles. The research goal in this paper is to show a similar method to increase the accuracy of a premade convolutional neural network that detects eye blink by varying the training data with different races and situational profiles. The races tested included Caucasian, Hispanic and Asian. Situational profiles included eyeglasses, multiple people in frame, and hair being in the way of the eye. The training data also incorporated frames of video data; this meant the subject periodically moved their face in different directions. Being able to detect eye blink is an important task by itself; therefore, with an accurate CNN we could further the field in eye blink detection. Additionally, I have shown that training with a variety of races, wearing glasses, hair in the way of their eyes and periodically moving faces increased the accuracy (Figures 5,8). However, even though I found an increase of accuracy, many mysteries still remain.

INTRODUCTION:

Our work for eye blink detection initially started when considering eye blink detection for financial risk taking as mentioned by Sherman (Sherman et al, 2016). In the paper published by Sherman, they showed a strong correlation between blink rate and risk taking. Therefore, by knowing someone's blink rate we would have some evidence that they are making a risky decision. Detecting an eye blink can also be used for driver assistance and would help prevent driver fatigue, one of the major causes of accidents all over the world, driver fatigue (Du et al., 2008). In a paper from Punitha they created a real-time fatigue monitoring system which is able to use eye blinking to estimate driver fatigue (Punitha et al., 2014). Since they are able to monitor the eyes with a camera, they are able to monitor the symptoms needed for early detections of driver fatigue which could prevent an accident. Another example, from Mohammed showed a motivation for eye blink detection that would help people with disabilities to interact with computers and phones (Mohammed et al., 2014). The software they are creating is using eye tracking but uses it in conjunction with eye blink detection. The main goal of this study was to help disabled individuals without hands to control phone calls with eye blinking.

RELATED WORK:

A paper by Anas showed a novel eye status detection method with a CNN (Anas et al., 2013). In this paper, they were able to outperform current eye blink detection techniques. They were able to show that through different illuminations, distances, head poses, and blur they were able to detect an eye blink where other techniques failed. The reason this is possible is because the neural network is able to train on difficult data sets and learn the characteristics needed to classify eye blink in harder detection environments. In a paper from Du they presented a method for classifying eye blink using the eye area, average height of the pupil, and the width to height ratio. With this method they were able to accomplish a detection accuracy around 91% (Du et al., 2008). In another paper, Pan used a support vector machine that uses features retrieved from the image, they were able to achieve an accuracy of 92% (Pan et al., 2007). Another method for eye blink detection was introduced by Fazil and Esfenhani. Here they used the measured amount of white colour pixels, representing the sclera, and black pixels, representing the iris and eyelash, to detect the eye blink with an average accuracy between 95-100% (Fazli and Esfehani, 2012).

The approach that I used in my work for detecting eye blink is from a GitHub repository(Lee, 2018). In this GitHub repository they used a data set called Closed Eyes in The Wild (CEW) (Song et al, 2014) and they followed a CNN detecting eye blink blog (Paraskevopoulos, 2018) when programming their own neural network. In the blog post by Paraskevopoulous it does not seem like he did any research to optimize the CNN for eye blinking, like they did in the Anas paper (Anas et al., 2013). In the Anas paper, they showed that it is possible to detect eye blinks in real-time at variable head poses, appearances and illumination conditions. The Anas paper also provided graphs on what type of data made the neural network perform poorly. However, in the Paraskevopoulous blog, they simply retrieved the CEW data set and created a binary CNN to detect eye blink. Simply grabbing a data set from an online source and creating a CNN will not perform well in real-time scenarios and that is checked in this paper. Work is also done to understand the failure of the neural network from the blog to perform accurately.

DATA:

In the Blink Comparison Patient data (BCP) set, a data set collected from our collaborator Dr. Wilson to analyze their theory on the effect of blink rate on risk raking (Sherman et al, 2016), 12 people in their lab were recorded watching a 3-5-minute video. To process the BCP data set, other students and I watched the videos frame by frame and classified the frames as either open, closing, closed or open. After the frames were classified, I removed the frames labeled as either opening or closing. I then used the package dlib (King et. al., 2009) to grab 26 by 34 images of the right eye and the left eye for each frame. The reason this size was chosen was because the older data set, mentioned later, used this size for their training. After cropping the images, I converted them to greyscale and used data augmentation to increase the amount of data. The data augmentation included rotation, width shift, height shift, and shearing. These classifications and cropped eyes were then used to train a CNN network with parameters that are identical to the CNN from Lee, a CNN I found in a GitHub repository (Lee et al., 2018). Lee did not publish a paper on their neural network; however, it was clear through their documentation that the neural network was intended for eye blink detection. Examples of the cropped eyes in the BCP data set is included below (Figure 1).

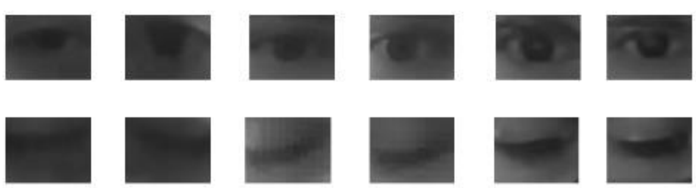

Figure 1. Examples of closed and open eyes from 2 different subjects from the BCP data set

The Closed Eyes in The Wild (CEW) data set (Song et al, 2014) contains 2423 different subjects where 1192 subjects are closed eyes from the internet and the other 1231 are open eyes from another data based called the Labeled Face in the Wild (LFW) (Huang et al, 2014). Eye patches were extracted from the subjects to 24 by 36 images centered along the eye position.

Classifications were created based on the subject either being in the closed eyes set from the internet or the open eyes from LFW. These classifications and cropped eyes were also used to train a CNN network with parameters that are identical to the CNN from Lee. Data augmentation also occurred with this data set. The data augmentation included rotation, width shift, height shift, and shearing. Examples of the cropped eyes in the CEW data set is included below (Figure 2).

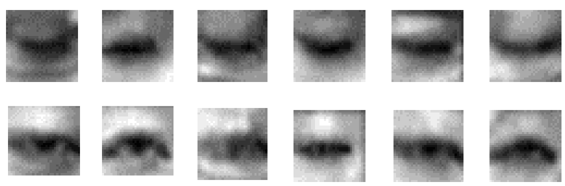

Figure 2. Examples of closed and open eyes from 12 different subjects from the CEW data set. Top row is closed eyes and bottom row is open eyes.

To check if the CEW data set would perform better on still Google images I collected still images from Google varying image sizes of individuals with open and closed eyes. Each Google image collected had similar characterizes as a headshot and did not include other people in the frame and had a clear sight of the face. Some examples of the headshot photos are included in below (Figure 3).

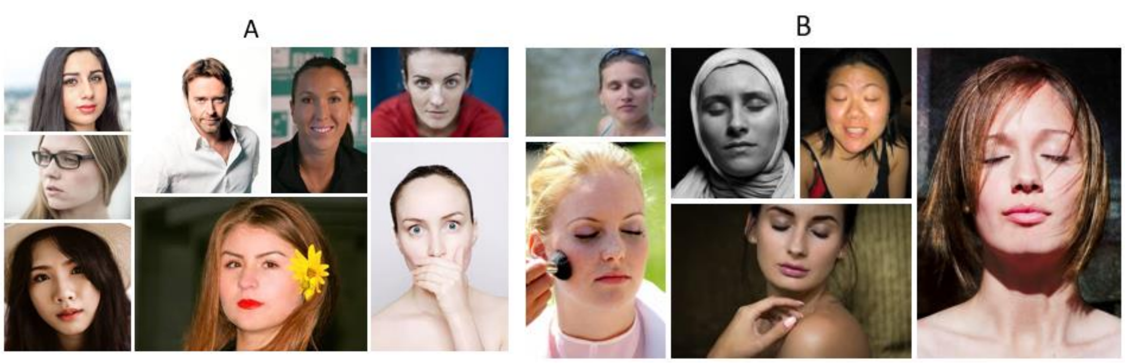

Figure 3. Open and closed eye still images. In part A some of the open eyes are included with varying sizes. In part B some of the closed eyes are including with varying sizes.

After the images were collected, I ran the same process for collecting the patches of the eyes as the BCP data set. Examples of the cropped patches from the Google still images are included below (Figure 4).

Figure 4. Examples of closed and open eyes from 12 different subjects from the SIG data.

METHODS:

In this research I used a CNN from CEW (Lee, 2018) to classify cropped eyes extracted from the frames of 12 different subjects from the BCP data set. The experiments conducted were training the neural network and collecting results in nine different scenarios. The first experiment was training the CNN on BCP data and testing with leave one out cross validation (LOOCV). The BCP data set has a ratio of 1 to 5 of images of closed eyes to open eyes. Also, frames that were in between the eye opening and closing were removed from the training data. Examples of the cropped eyes in the BCP data set are included below. For the cross validation, one patient was left out for testing and this was repeated for all patients. The second experiment involved training with all subjects from the BCP data set and testing with the CEW data set. The third experiment involved training the neural network on the CEW data and doing a 10-fold cross validation to test the overall accuracy of the neural network. In the experiment, the number of training, validation, and testing data was 2327/288/258 respectively. The fourth experiment was to test the CNN trained on the CEW data on the BCP data set. The fifth experiment involved training on the CEW data and testing on the BCP data with difficult testing data such as glasses and hair in the way of the eye removed. The sixth experiment involved training on the CEW data and testing on the BCP data with difficult testing data such as glasses, hair in the way of the eye and the difficult races with less visible pupils removed. The seventh experiments involved training the neural network with the CEW data and testing on the still images from Google (SIG). The eight experiment incorporated the CEW data and the BCP concatenated together for training with the SIG data for testing. The last experiment was training on the BCP data set and the SIG data.

The way that accuracy was calculated in the confusion matrices is included below.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \qquad \text{(Eq 1)}$$

In the formula, TP (True negative) refers to the correctly predicted open eyes, TN (True negative) is the correctly predicted closed eye, FP (False Positive) is the incorrectly predicted closed eyes, and FN (false negative) are the incorrectly predicted open eyes. Included in all the figures is a calculation of accuracy using Equation 1.

RESULTS:

Trained On BCA Data and Tested On BCA Data

n = 13250	Predicted: No	Predicted: Yes
Actual: No	Trug Neg(TN) 2725	False Pos(FP) 367
Actual: Yes	False Neg(FN) 190	True Pos(TP) 9968

Figure 5. Confusion Matrix for the neural network obtained by performing leave one out cross validation. The cross validation trained on 11 of the patients and tested on 1. This was done with all combinations of 11 to 1 subjects from the BCP data set. Each of the partition's accuracy was summed to get the above cross validation. The accuracy of this test was 95.8%.

Trained On Video Data and Tested On Kairess Data

n = 2874	Predicted: No	Predicted: Yes
Actual: No	Trug Neg 1155	False Pos 219
Actual: Yes	False Neg 22	True Pos 1478

Figure 6. Confusion Matrix with the neural network trained on all subjects in the BCP data set and tested on CEW data. The accuracy of this was gave 91.6%.

Trained On Kairess Data and Tested On Kairess Data

n = 2584	Predicted: No	Predicted: Yes
Actual: No	Trug Neg 1225	False Pos 6
Actual: Yes	False Neg 10	True Pos 1343

Figure 7. Confusion Matrix with the neural network trained on CEW data and tested on CEW data with a 10-fold cross validation. Each of the partition's accuracy was summed to get the above cross validation. The accuracy of this test was 99.4%

Trained On Kairess Data and Tested On Video Data

n = 12714	Predicted: No	Predicted: Yes
Actual: No	Trug Neg 1942	False Pos 158
Actual: Yes	False Neg 2049	True Pos 8565

Figure 8. Confusion Matrix with the neural network trained on all CEW data sets and tested on all subjects in the BCP data set. The accuracy of this test was 82.6%.

Trained On Kairess Data and Tested On No Difficult Video Data

n = 7913	Predicted: No	Predicted: Yes
Actual: No	Trug Neg 1311	False Pos 25
Actual: Yes	False Neg 569	True Pos 6008

Figure 9. Confusion Matrix with the neural network trained on all CEW data sets and tested on 9 subjects from the BCP data set without the subjects that had obstacles in the way of the eyes; such as glasses and hair. The accuracy of this test was 92.5%.

Trained On Kairess Data and Tested On Highly Visible Pupil Video Data

n = 6595	Predicted: No	Predicted: Yes
Actual: No	Trug Neg 1050	False Pos 51
Actual: Yes	False Neg 133	True Pos 5361

Figure 10, Confusion Matrix with the neural network trained on all CEW data sets and tested on 5 subjects from the BCP data set with only races Caucasian and Indian. These subjects did not have any obstacles or much movement. The data sets included also had the most visible pupil and iris. The accuracy of this test was 97.2%.

Trained On Kairess Data and Tested On the SIG Data

n = 80	Predicted: No	Predicted: Yes
Actual: No	Trug Neg 40	False Pos 0
Actual: Yes	False Neg 14	True Pos 26

Figure 11, Confusion Matrix with the neural network trained on all CEW Data and tested on the 40 images that were collected from Google. The accuracy of this test was 82.5%.

Trained On BCA Data with Kairess Data and the SIG Data

n = 80	Predicted: No	Predicted: Yes
Actual: No	Trug Neg 40	False Pos 0
Actual: Yes	False Neg 10	True Pos 30

Figure 12, Confusion Matrix with the neural network trained on all BCP Data with the CEW data concentered together and tested on the 40 images that were collected from Google. The accuracy of this test was 87.5%.

Trained On BCA Data and Tested On the SIG data

n = 80	Predicted: No	Predicted: Yes
Actual: No	Trug Neg 40	False Pos 0
Actual: Yes	False Neg 3	True Pos 37

Figure 13, Confusion Matrix with the neural network trained on all BCP Data and tested on the 40 images that were collected from Google. The accuracy of this test was 96.25%.

When testing the neural networking, trained on the CEW data set, on the BCP data set I saw an accuracy of 82.6% (Figure 8). By then training the neural network on the BCP data set that included complex data sets such as, hair in the way of the eyes, glasses, different races and periodically moving faces I saw an increase of accuracy to 95.8% when performing a LOOCV for the accuracy (Figure 5). This BCP data set detection increase of 13.2% may be occurring from a possible difference in the processing of the data sets so more experiments needed to be conducted.

To check if the accuracy when trained on the CEW data set was affected by the difficult subjects in the testing data such as those from races with less visible pupils, or glasses or hair being in the way of the eye from the BCP data set, I created two different confusion matrices without difficult subjects in the testing data (Figure 9). The first simplified test set had the subjects with glasses as well as hair getting in the way of eyes removed. By removing those difficult images, I saw an accuracy of 92.5% (Figure 9) up from 82.6% (Figure 8) when training the CNN on the CEW data. Another check to see if the difficult datasets in the testing data was causing the drop in the original accuracy, I removed the data sets that had less visible pupils and irises from the BCP testing set. These data sets included the Asian patients and a Native American patient. The previous sets with glasses and hair in the way of their eye were also removed. When the pupils that were less recognizable were removed from the testing data set, I saw an accuracy increase to 97% (Figure 10) from the 92.5% accuracy (Figure 9) when the testing BCP set only had the data sets with glasses and hair in the way of their eye removed.

Some form of overfitting may be occurring when training on all the subjects from the BCP dataset. Subsequently, when the CNN trains on all the subjects from the BCP dataset and is tested solely on the CEW data set, I saw an accuracy of 91.6% (Figure 6) in comparison to the accuracy of 99.4% (Figure 7) gathered from when the neural network was trained on CEW data set and tested on the CEW data set with a 10-fold cross validation. A reason why the accuracy may be so different is because partial closed eyes were not included in the BCP data set; therefore, partially closed eyes are being classified as closed in the CEW data set. This would be occurring because the CEW data set has partially closed eyes classified as closed in the data set.

Another experiment I wanted to check was to see if random face shots from Google would also test poorly on the CNN trained on the CEW data set. Since the CEW data set is composed of Google images and videos I expected the CNN trained on the BCP data set to perform worse because the BCP data set is only composed of videos. However, I saw the CNN trained on the CEW data and tested on the SIG data perform worse with an accuracy of 82.5% (Figured 11) versus when the CNN was trained on the BCP data, giving an accuracy of approximately 96.3% (Figure 13). Another unclear result was when I merged the CEW data and the BCP data set for training. When the SIG data were tested on SIG data, I saw an accuracy increase of 87.5% (Figure 12) from training solely on the CEW data set. However, the accuracy of training on the merged data did not beat training solely on the BCP data set and testing on the SIG data.

CONCLUSION/FUTUREWORK

With more data it would be possible to train the neural network on a proportional amount of different data sets with subjects that are not directly looking at the camera, different races, glasses or hair in the way of their eyes Also, it would make the accuracy better if the number of blinks to open eyes in the data were 50/50 instead of 20/100 (Du et al., 2008). To fix this problem, our neural network requires more data and more blink frames. Another possibility would be to create a neural network that keeps the frames that are in between blinking and to classify those frames as well. This form of neural network would need to include a time correlation to train along the subject from the BCP data set instead of single frames being open or closed. Another future approach that would increase accuracy is to include if an eye is partially closed or open within the data set as to not classify it as being closed. An unsolved mystery in our data set was training on the BCP data set and getting a higher performed than training on the CEW when the testing data was the still images from Google. More work needs to be done to understand what is going on in the testing data and to find the exact reason why CEW data set is performing poorly.

Work Cited:

A. Punitha, M. K. Geetha and A. Sivaprakash, "Driver fatigue monitoring system based on eye state analysis," 2014 International Conference on Circuits, Power and Computing Technologies *[ICCPCT-2014]*, Nagercoil, 2014, pp. 1405-1408.

Anas, E.R., Henríquez, P., Matuszewski, B.J., 2017. Online eye status detection in the wild with convolutional neural networks. In: VISIGRAPP (6: VISAPP), pp. 88–95 (2017)

Davis E. King. Dlib-ml: A Machine Learning Toolkit. Journal of Machine Learning Research 10, pp. 1755-1758, 2009

Du, Y., Ma, P., Su, X., Zhang, Y., 2008. Driver Fatigue Detection based on Eye State Analysis. In: Proceedings of the 11th Joint Conference on Information Sciences

G. B. Huang, M. Ramesh, T. Berg and E. Learned-Miller, Labeled faces in the wild: A database for studying face recognition in unconstrained environments, Tech. Rep. 07-49, University of Massachusetts, Amherst (October 2007).

Fazli, S., Esfehani, P., 2012. Automatic Fatigue Detection Based On Eye States. In: Proc. of Int. Conf. on Advances in Computer Engineering

F.Song, X.Tan, X.Liu and S.Chen, Eyes Closeness Detection from Still Images with Multi-scale Histograms of Principal Oriented Gradients, Pattern Recognition, 2014.

Lee, Eye Blink Detector, (2018), GitHub repository, https://github.com/kairess/eye_blink_detector

Le, V., Brandt, J., Lin, Z., Bourdev, L., Huang, T., 2012. Interactive facial feature localization. ECCV’12, pp.679-692.

Malik, K., Smolka, B., 2014. Eye blink detection using local binary patterns. In: 2014 International Conference on Multimedia Computing and Systems (ICMCS), pp. 385–390

Mohammed, A., Anwer, S., 2014. Efficient Eye Blink Detection Method for disabled-helping domain. Inter. Jour. of Adv. In: Computer Science and Applications

Pan, G., Sun, L., Wu, Z., Lao, S., 2007. Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In: 11th IEEE International Conference on Computer Vision. ICCV '07, pp. 1–8.

Paraskevopoulos, Iasonas. "A Simple Blink Detector Using a Small Convolutional Neural Network with Python." *Medium,* Medium, 26 Feb. 2018, medium.com/@iaswnparaskev/a-simple-blink-detector-using-a-small-convolutional-neural-network-with-python-ba8171ae0e0b.

Sherman, Emily, and Robert C. Wilson "Spontaneous Blink Rate Correlates with Financial Risk Taking." *bioRxiy* (2016): 046821