iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 Issue: V Month of publication: May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52287

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Eye Blink Oriented Chat

Muthyala Sathvika[1], Gopathi Pravallika[2], Kotamarthi Manasa[3], Ch. Vijayabhaskar[4]
[1, 2, 3, 4]Sreenidhi Institute of Science and Technology Yamnampet, Ghatkesar, Hyderabad, Telangana, Hyderabad-501301

Abstract: Digitalization has become an essential aspect of modern-day long-distance communication and is used extensively in daily life. As a result, it is crucial that this technology is accessible to everyone, including differently-abled individuals. The objective of this study is to create an eye-controlled virtual keyboard that enables differently-abled individuals to use text features effectively. To achieve this, the study utilizes Dlib, OpenCV, and Convolutional Neural Networks (CNNs) to monitor eye movement and blinks, allowing the user to select desired keys on the virtual keyboard. By accurately predicting the eye's state using neural networks, the system operates the virtual keyboard with precision. Through this study, a highly efficient method for differently-abled individuals to communicate through text has been established, paving the way for future advancements and research in this field.
Keywords: Eye-controlled, Convolutional Neural Network, OpenCV, and Dlib.

I. INTRODUCTION

Human beings have an inherent need for communication, which has been facilitated by computer-mediated communication (CMC). CMC allows people to communicate across great distances, to an unlimited number of people at a low cost, and with ease in creating and sharing documents and other material. Although computers have evolved significantly in terms of power and potentiality, people still use keyboards and mouse to communicate and work with them. However, for individuals with severe physical disabilities such as paralysis and amputation, it is difficult to use computers for communication. According to the Christopher and Dana Reeve Foundation's study in 2013, , almost 1 out of 50 people live with paralysis, and their physical activities are often limited to eye blinking. This study aims to design an application that allows people with physical disabilities to use a virtual keyboard using their eye movement. While there have been advancements in the field, such as the Tobii Eye Tracker, these devices are not economically feasible for the general public. The application designed in this study provides a cheaper and more convenient alternative for differently-abled people, which can be operated using a simple laptop and a webcam.

II. RELATED WORKS

Eye tracking technology has come a long way in the past few decades, and it has become an important tool for researchers studying human behavior and interaction with technology. This technology involves tracking the position and movements of the eye to determine the direction of gaze. There are several different technologies used to track eye movements, including infrared-oculography, scleral search coil method, electrooculography, and video-oculography.

Currently, most eye tracking research for Human-Computer Interaction (HCI) is based on video-oculography because it is less invasive to the user. However, there are still some challenges to be addressed in this field, such as attention diverting problems and accuracy issues. Researchers have been working on developing new systems to address these problems.

One such system is the EASE (Eye Assisted Selection and Entry) system, designed by Wang et al. This system uses eye-tracking to assist with text entry, making it easier and more efficient for users. Another system developed by MacKenzie and Ashtiani uses eye-blinking to control a scanning ambiguous keyboard. Chau and Betke have also developed a system that detects eye blinks and analyzes their pattern and duration.

In addition to these systems, there have been studies exploring different eye-gazing techniques, algorithms, and models. Grauman and Magee have surveyed and described different types of eye blinks, while Kro´lak and Strumiłło have developed a system that allows for operation dependent on the eye without requiring muscle movements. The eye-blink controlled systems are able to distinguish between voluntary and involuntary blinks and interpret single voluntary blinks or their sequences.

Seki et al. have categorized vision-based eye blink detection techniques into two types: active and passive eye-blink detection. Active eye-blink detection relies on special illumination and uses the retro-reflective property of the eye to provide accurate results quickly and robustly. Passive eye blink detection techniques, on the other hand, do not require additional light sources and detect the blinking ratio from sequences of images within the visible spectrum.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Chakrabor-thy et al. have provided a system that uses similar groundworks to the present study to address issues with eye tracking. However, the current study aims to improve efficiency and accuracy by using a different methodology. As eye-tracking technology continues to evolve, it is likely that more systems and techniques will be developed to improve its accuracy and usability.

III. DATA ACQUISITION

To train the models for blink and gaze detection, a dataset was acquired through a webcam. The dataset contained approximately 5000 images captured from 10 individuals, which were then preprocessed and filtered manually. For gaze detection, the dataset was categorized into three classes based on the position of the eyeball: left gaze, right gaze, and center gaze. The data collection process involved instructing the individuals to follow a dot that moved randomly to left, right, and center positions on the screen, with the corresponding eye images collected as corresponding classes.

On the other hand, for blink detection, the "open percentage" of each image from the previously collected dataset was manually assigned. The "open percentage" is a measure of how open the eyes are, with 100% indicating wide open and 0% indicating completely closed. Any image with an open percentage below 10% was considered as a blink. It is important to note that the data collection process was performed using a script to ensure consistency across the dataset.

IV. METHODOLOGY

The methods used for accomplishment of eye- controlled virtual keyboard is shown in Fig. 1.

A. *Video Capture*

First of all, video is captured from the webcam,which has the actual resolution. From the capturedvideo, we grab frames. The captured frame is passed for the grayscale conversion to reduce the computationpower for further processing.

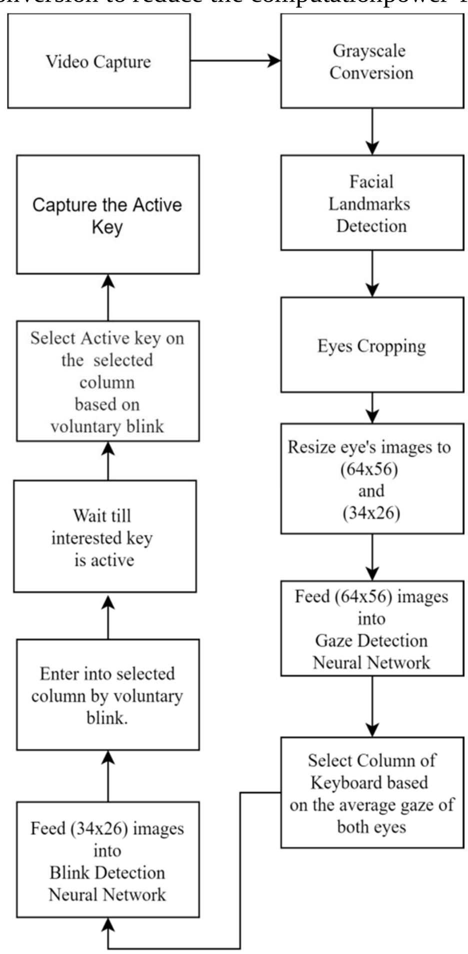

Fig. 1. System Overview

B. Grayscale Conversion

The grayscale conversion has many ways of con- verting colorful images into gray form. The luminositymethod is used to convert the color image into gray.It can be carried out by following steps:

1) Read the color image.
2) For any pixel read the intensity values of Red, Blue and Green channels as R, G and B respec- tively.
3) Calculate the gray valueGr=0.299*R+0.587*G+0.114*B
4) Set,Gr as intensity.
5) Repeat steps from 2 to 4 until all pixels arescanned.

C. Facial Landmarks Detection

Facial landmarks are used to localize and represent salient regions of the face. There are two steps while detecting the facial landmarks:

1) Localize the face in the image
2) Detect the key facial structures on the face ROI There are various landmark detection algorithms avail-able, but for this application, we use Dlib. Dlib is a cross-platform library written in C++, which uses

histogram-oriented gradient (HOG) to detect the faceand linear SVM for detecting 68 landmarks points of the face. The 68 landmarks point detected by Dlib isshown in figure 2.

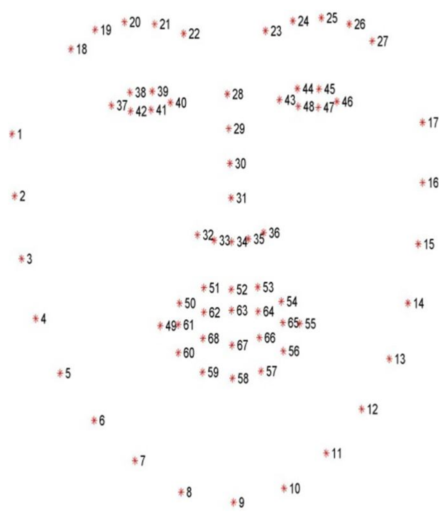

Fig. 2. Visualization of the 68 facial landmark points from [11].

D. Eye Cropping And Resizing

The left eye is cropped from the face using thelandmarks points: 37,38,39,40, 41,and 42 and the righteye is cropped from the face using the landmarks points: 43,44,45,46,47,and 48 as shown in Fig. 3. Nowobtained cropped left and right eyes are resized into size 64x56 and 34x26 to feed gaze detection network and blink detection network, respectively.

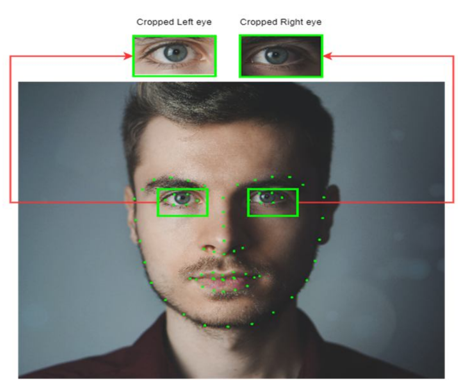

Fig. 3. Eye cropping using 37,38,39,40,41,42 and 43,44,45,46,47,48facial landmark points.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

E. Neural Network Architectures

1) *Blink Detection Architecture :* In the beginning, the blink detection architecture consists of a convolu- tion layer that applies 32 filters, each of size 3×3 on the input image. Max pooling is then used to reduce the spatial dimensions of the feature map. Convolutionand max-pooling operations are repeated twice with64 and 128 filters, each of size 3×3. A single fully connected layer with 512 nodes is appended to the convolution layer. The flatten layer transforms a two- dimensional matrix of features into a vector fed into a fully connected neural network classifier between the convolutional and fully connected layers. Here, Dropout is applied to the fully connected layer to randomly drop out nodes during training, reducing overfitting and improving generalization error. Next, we have a single Dense layer to generate the classifi- cation. Each of the Convolution layer and dense layer uses ReLU as the activation function but the single output node uses Sigmoid as activation function.
2) *Gaze Detection Architecture :* The gaze de- tection architecture is almost identical to the blinkdetection architecture, with only a difference in the output layer, which consists of three output nodes. The output nodes uses SoftMax as the activation function.

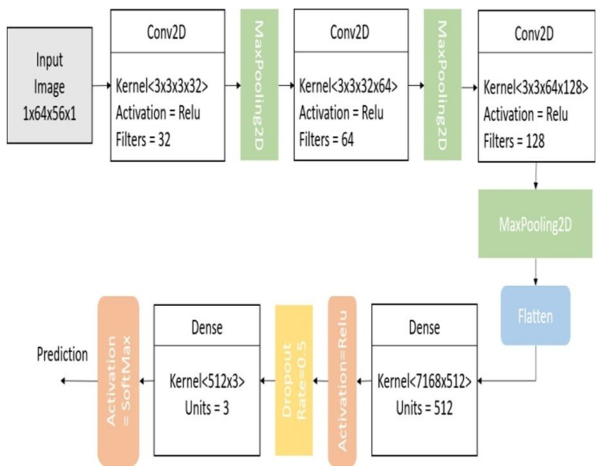

Fig. 4. Visualization of Blink Detection Architecture

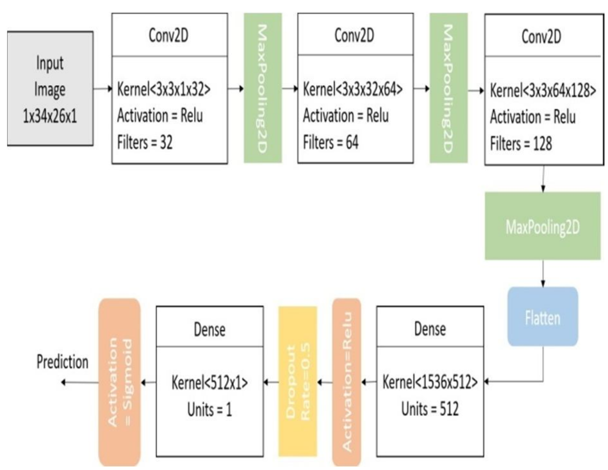

Fig. 5. Visualization of Gaze Detection Architecture

3) *Data Augmentation and Training*

Despite their apparent similarity, both architectures in this study were trained on different datasets with different dimensions. However, the dataset size for both models was inadequate, so Image Data Augmentation was performed to expand the dataset size. This technique involves applying various transformations to the original images, resulting in multiple copies of each image, each with unique features depending on the augmentation technique used, such as shifting, rotating, or flippingWe utilized random rotation up to 10%, random height and width shift up to 20%, and random zoom up to 10% to augment the original images.

This augmentation technique increased the size of our dataset and introduced more variations in the data, enhancing the model's ability to make accurate predictions on unseen data. Additionally, training on slightly altered images makes the model more robust.We reserved 20% of the dataset for validation purposes, and the remaining dataset was used for training purposes. To ensure efficient training, we set the batch size to 16 and trained the models for a total of 32 iterations or epochs. We also configured the EarlyStopping callback to halt training once the model performance stops improving on a hold-out validation dataset.total of 32 iterations, or epochs.

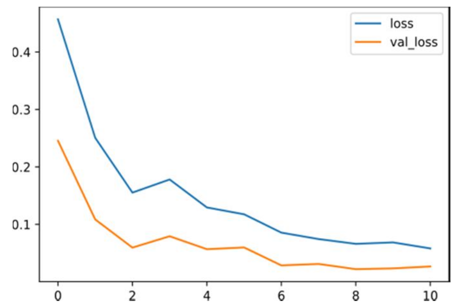

Fig. 6. Plot of Model Loss on Training and Validation Datasets for Blink Detection Model

The last layer of gaze detection architecture doesn’t activate with Sigmoid but with SoftMax instead, which allows us to generate an accurate multiclass probability distribution as shown in equation 4. And Categorical Cross-Entropy loss is used as a loss function as shown in equation 5.

e^{si}

$$f(s) = \quad c \quad e \quad j \qquad (4)$$

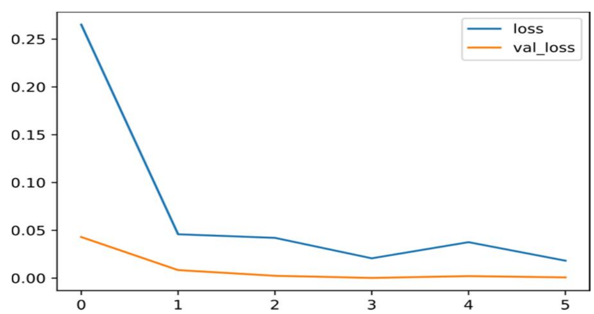

Fig. 7. Plot of Model Loss on Training and Validation Datasets for Gaze Detection Model

a) Activation function, Loss function, and the opti-mizer

For both Architecture, each of the convolution layerand fully connected layer uses the ReLU activation function shown in equation 1

$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x => 0 \end{cases}$$

For the blink detection architecture, the last layer uses Sigmoid as activation function and Binary Cross-Entropy Loss as loss function as shown in equation 2 and 3.

$$f(s)_i = \frac{1}{1 + e^{-s_i}} \qquad (2)$$

$$BCE = -t_1 \log (f(s_1)) - (1 - t_1) \log (1 - f(s_1)) \qquad (3)$$

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

where s_1 and t_1 are the score and the ground-truth
Where s_i refers to one of the class and s_j are thescores inferred by the network for each class in

$$CE = -\sum_{i}^{c} t_i \log (f(s)_i) \quad (5)$$

Where t_i and s_i are the ground truth and the CNNscore for each class.
Here we use Adam optimizer [9] to modify both network's parameters to minimize the cost function onthe training set as shown in equation 6.

$$\theta_{t+1} = \theta_t - \frac{\eta \cdot \hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}$$

where

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t} \qquad (6)$$

$$\hat{v}_t = \frac{v_t}{1-\beta_2^t}$$

and where

$$m_t = (1-\beta_1) g_t + \beta_1 m_{t-1} \quad v_t = (1-\beta_2) g^2 + \beta_2 v_{t-1}$$

- θ is the parameter to be updated
- is the learning rate, which is set to η=0.001as default
- g refers to gradient $g = J(\theta_{t,i})$, where is the gradient which is taken of J and J refers to our cost function.
- (Epsilon), which is a small term preventing division by zero $\in$
- β_1 and β_2 are two decay terms, also called the exponential decay rates
- m is first momentum vector for g.
  m̂ t is bias-corrected first momentum.
- v is second momentum vector for g.
  $\hat{}_t$ is bias-corrected second momentum..
  Where:

– First momentum β_1 =0.9 $f(x) =$
– Second momentum β_2 =0.999

If we take the value of t and put it in theexponent,
i.e., if t=5 then, $\beta^{t=5} = 0.9^5 = 0.59049$ 1

b) Keyboard key selection

The keyboard designed for this application contains all the alphanumeric keys along with the repeatedly used symbols and backspace. The keys of the keyboard are grouped into 10 columns, as shown in figure 8. Each column has 6 keys. For example, column 1 has '1', 'q', 'a', 'z', '+' and '%' keys. Initially, column 1 is selected, which is indicated by highlighting the column with green color. Another column is selected according to both eye's gaze, i.e., if the eyes are looking left, select the left column. If both eyes are looking right, select the right column to the previouslyselected column. For example, column 7 is selected previously, and if the person looks left, then column 6 will be selected, and if the person looks right, then column 8 will be selected. To select a specific key, the key containing the column must be selected via the eye's gaze . In order to enter into that column a person has to blink both eyes, as a result each keyon the column is highlighted for a certain period of time one after another.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

If the person blink, the keyon the column which is highlighted at the time ofeyes blink is selected as the pressed key. However, only the voluntary eye blink is used for entering inside the interested column, and to select the key, andinvoluntary eye blinks are rejected. This is done via tracking the frames. For example, let's say 1 secondis divided into 16 frames; if a person blink only for5 frames, it is regarded as an involuntary blink. If a person blink for more than 12 frames, it is regardedas a voluntary blink. Now the selected key is captured and copied to the clipboard.

V. RESULT

People who have physical disabilities face limitations in their ability to move around, perform manual tasks, and participate in certain activities. These impairments can create significant challenges when using information technology, making it necessary to create adaptations that allow people with disabilities to communicate effectively with computers. Information education is a crucial component of modern society, and it is vital to address communication problems between physically handicapped individuals and computers.

This study aims to address the deficiencies and disparities faced by individuals with physical disabilities by proposing an "Eye-controlled virtual keyboard" solution. This technology is designed to help physically challenged individuals communicate more effectively with computers, enabling them to participate fully in information education and other aspects of modern society.

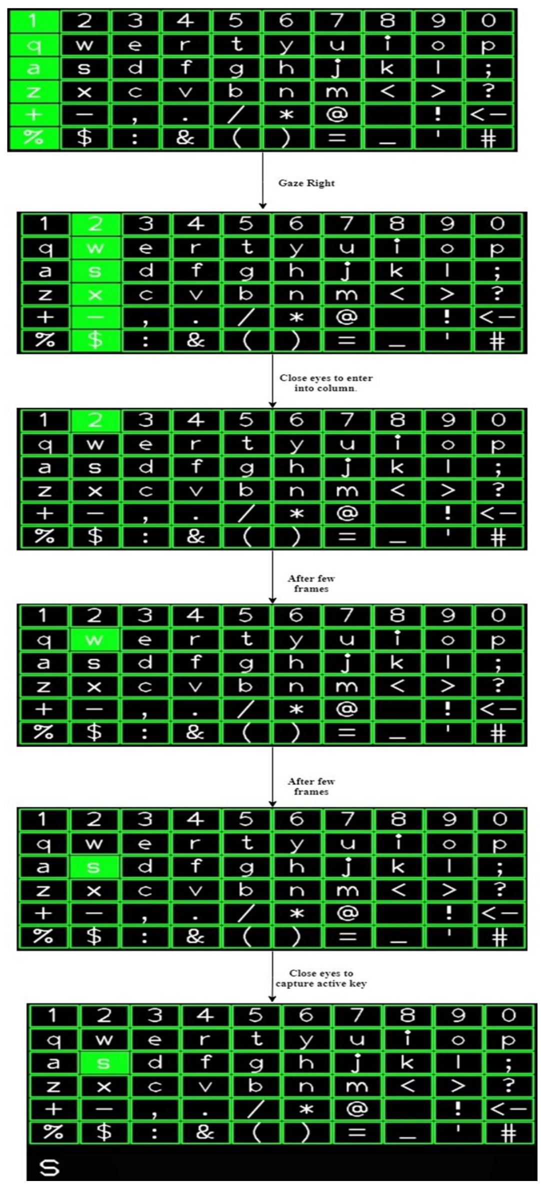

Fig. 8. Overall steps to select 's' key.

Due to the factors like lighting condition and camera quality. And also blink detection was more accurate than the gaze detection model. We were able to type 6 characters per minute on average under proper lightingconditions.

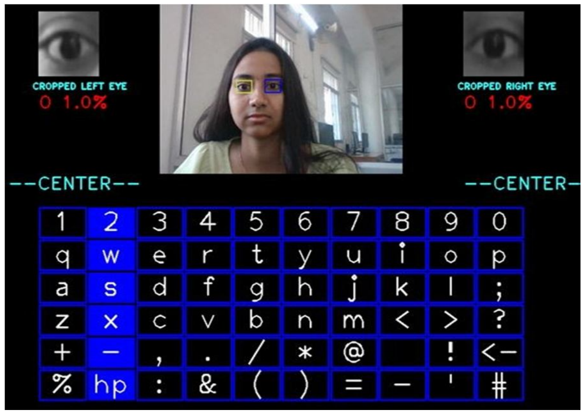

Fig. 9. Final window.

VI. CONCLUSION

Individuals with physical disabilities often experience difficulties in carrying out physical tasks and participating in certain life activities, which can limit their ability to use information technology effectively. Therefore, it is crucial to develop adaptations that allow individuals with physical impairments to communicate fully with computers. Given that information technology is an essential aspect of modern society, particular attention should be paid to addressing communication barriers between physically handicapped individuals and computers.

To reduce disparities and address deficiencies, this study proposes the development of an "Eye-controlled virtual keyboard" for individuals with physical disabilities. This solution aims to enable physically challenged individuals to communicate with computers more effectively, allowing them to participate fully in information education and other areas of modern society.

REFERENCES

[1] P. Chakraborty, D. Roy, M. Z. Rahman, and S. Rahman, “Eye Gaze Controlled Virtual Keyboard,” International Journal of Recent Technology and Engineering (IJRTE), 2019.
[2] M. L. Mele and S. Federici, “Gaze and eye-tracking solutions for psychological research”, Cognitive Processing, 2012
[3] M. Chau and M. Betke, “Real time eye tracking and blink detection with USB cameras”, 2005
[4] A. Sharma and P. Abrol, “Eye gaze techniques for human computer interaction: A research survey”, International Journalof Computer Application, 2013
[5] K. Grauman, M. Betke, J. Lombardi, J. Gips, and G. Bradski, “Communication via eye blinks and eyebrow raises: Video based human-computer interfaces”, Universal Access in the Information Society, np. 2(4), p. 359-473, 2003
[6] A. Królak and P. Strumiłło, “Eye-blink detection system for human-computer interaction”, Universal Access in the Infor- mation Society, 2011
[7] M. Seki, M. Shimotani, and M. Nishida, “A study of blink detection using bright pupils” 1998-2001
[8] Y. LeCun and Y. Bengio, “Convolutional Networks for Images, Speech”, and Time-Series. 1997.
[9] D. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” International Conference on Learning Repre- sentations, 2014.
[10] B. Triggs and N. Dalal, “Histograms of Oriented Gradients forHuman Detection,” in 2013 IEEE Conference on Computer Vision and Pattern Recognition, San Diego, California, 2005 pp. 886-893. doi: 10.1109/CVPR.2005.177
[11] A. Rosebrock, “Facial landmarks with dlib, OpenCV, and Python,” Pyimagesearch.com, 03-Apr-2017. [Online]. Available: https://www.pyimagesearch.com/2017/04/03/facial- landmarks-dlib-opencv-python/.
[12] J. Wang, S. Zhai, and H. Su, “Chinese Input with Keyboard and Eye-Tracking - An Anatomical Study”, 2001.
[13] B. S. Armour, E. A. Courtney-Long, M. H. Fox, H. Fredine, and A. Cahill, “Prevalence and Causes of Paralysis-United States,” 2013.