International Journal of Technology and Engineering Studies
volume 6 issue 2 pp. 52-68 doi: https://dx.doi.org/10.20469/ijtes.6.40002-2

Research on Eye Pupil Location and Eye Tracking by Designing a Fully Conventional Neural Network (FCNN)

Muhammad Raza*
School of Computer Science,
Xian Technological University, Weiyang District,
XiAn, Shaanxi, China

Changyuan Wang
School of Computer Science,
Xian Technological University, Weiyang District,
XiAn, Shaanxi, China

Pengxiang XUE
School of Computer Science,
Xian Technological University, Weiyang District,
XiAn, Shaanxi, China

Shahzadi Bano
School of Information Engineering,
Zhengzhou University, Zhengzhou City,
Henan Province, China

Abstract: Eye-tracking has become a significant tool in various areas, incorporating interaction of humans with a computer, computer vision, psychology, and medical diagnostics. Various protocols have been utilized to track the gaze. Although, some may not be accurate in the real world, while others may need explicit custom calibration, which can cause problems. Few of these techniques are associated with low image grades and varying lighting situations. The latest success and popularity of deep studying has dramatically increased the effectiveness of eye-tracking. The convenience of huge datasets additionally improves the function of deep learning-depend techniques. This technical paper introduces the latest deep studying-depend gaze assessment technology with a concentration on Fully Convolution Neural Networks (FCNN). This technical paper also gives an overview of other machine-based eye assessment methods. This research objective to enable the research population to generate significant and useful horizons that can improve the structure and growth of better, more effective deep learning depends on eye-tracking methods. This paper also gives information on different pertained models, network structures, and open origin datasets to help train deep learning models.

Keywords: *Neural Networks, medical diagnostics, FCNN, datasets*

Received: 18 April 2020; ***Accepted:*** 03 June 2020; ***Published:*** 29 August 2020

I. INTRODUCTION

There are currently several kinds of commercial and non-commercial eye capturing solutions, which include appearance and model, depend on techniques. Although, few of these solutions are costly or imprecise in true life, while others need explicit utilization calibration, which can cause problems [1]. Therefore, the latest eye-tracking research has focused on making deep studying-depend eye-tracking methods that do not need explicit utilization calibration [1]. With the advent of deep studying, FCNN-depend eye scoring models have become very famous. It was released and gained popularity. [2] Well used for classifying handwritten numbers. The FCNN model is also suitable for processing large amounts of data and is used in various fields like computer imagination [3], speech identification [4], and language modeling [4]. The FCNN model allows image functions (such as pupil and blink position) to be mapped directly to the point of view without using manually developed functions [5]. This article focuses on FCNN and introduces methods for evaluating gaze based on appearance. The co-editors of this article revised the manuscript review and approved its

*Correspondence concerning this article should be addressed to Muhammad Raza, School of Computer Science, Xian Technological University, Weiyang District, XiAn, Shaanxi, China. E-mail: kendoumei017@gmail.com

publication. Kim Jong Hoon also introduced another machine learning-based eye assessment method. This article provides an overview of the latest FCNN gaze tracking technology and provides researchers with valuable and beneficial horizons that can improve the structure of better and more effective deep learning gaze tracking methods.

II. BASIC CONCEPTS

This section describes some of the basic concepts related to the research, especially FCNN's self-esteem. An Eye Assessment is a technique aimed at understanding the intentions and interests of the user [6]. Gaze assessment technology focuses on the link between data image and gaze direction [7, 8]. The direction of the line of sight is assessed based on the specific characteristics of the eye (for example pupil and corneal reflexes) drawn from the eye area, image data gathered by one or more cameras [9]. In general, eye assessment techniques can be classified into two types: model-depend techniques and appearance-depend methods [7]. See the next section for more information on these two technologies. Few practical uses of gaze assessment include relation between eyes [10, 11], driver's eye observation [3, 12], on-screen keyboard [13, 14] and Virtual Reality (VR) [15].

A. *Assessment of the Eyes by their Appearance*

Appearance-depend methods are based on the photometric look of the eye to assess gaze [9]. They usually require a camera to take a snapshot of the eye and use it to create a gaze assessment model that can match the appearance of a captured image to a specific gaze direction. In the real world, appearance-based gaze assessment models are noticed to give good outcomes [11]. They are used to directly assume screen correlates. That is, it can only be utilized in one tool and one orientation [16]. This is because the position of the line of sight is determined directly in the correlated system of the selected screen [17]. Several researchers have developed methods to predict the position of looking at the camera [10, 18].

B. *Model-depend Eye Assessment*

Model-depend methods allow gaze assessment by combining geometric eye models with eye properties like corneal reflex and pupil center [19]. Fig 1 shows you the geometric structure of the eye. Angle κ is explained as the angle between the visual line and the optical line. The angle variation among the optical line and the visual line is represented as $\theta = (\alpha, \beta)$. Here θ is a non-changeable vector of every person [1]. Individual calibration of model-depend methods is commonly utilized to calculate the value of θ of every object.

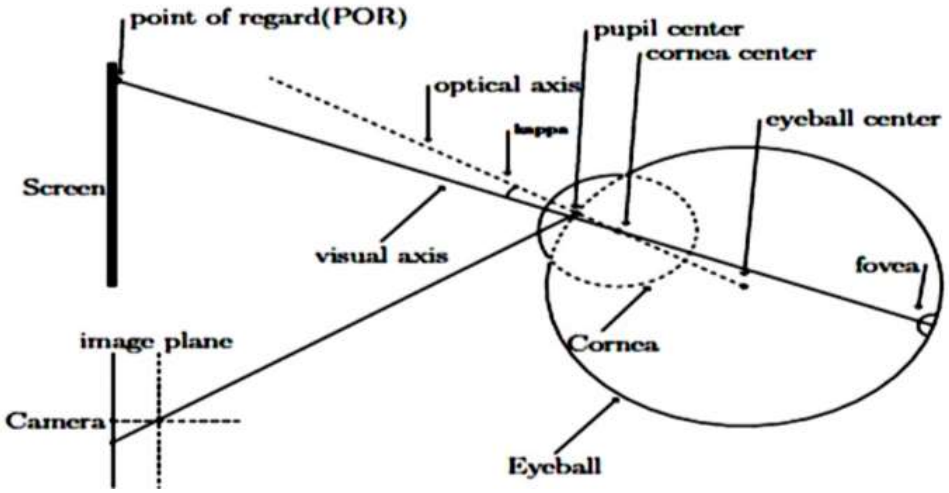

Fig. 1. Schematic Diagram of Eye

III. PERFORMANCE METRICS FOR GAZE CALCULATION

Gaze searching precision criteria are angle accuracy in degrees [1, 15, 20], distance precision in cm or mm [10, 21], and gaze rating [3, 13, 22]. This paper uses the words "precision" and "assumption error" to explain the function of different line-of-sight scoring systems. A truly categorized frame is one in which the predicted fixed area is equivalent to the real fixed area. Prediction errors are mainly described for regression issues. It is utilized to estimate the degree to which the model predicts line-of-sight correlates [23].

A. Gaze Point Coordinates in Pixels

The estimation of the coordinates of the viewpoint in pixels can be acquired from Equation 1 and Equation 2 [1].

$$\text{GazeX} = \text{mean}\,(\text{X}_{\text{left}} + \text{Xright}\,) \tag{1}$$

$$\text{GazeY} = \text{mean}(\text{Yleft} + \text{Yright}) \tag{2}$$

Here (Xleft; Xright; Yleft; Yright) shows the real coordinates of the line of sight of the left and right eyes received by the eye tracking device.

B. Gaze Position

You can use Equation 3 to 4 to calculate the gaze position (mm) distance on the screen [10].

$$X\text{Position(mm)} = \mu^{*}\text{Gaze}X \tag{3}$$

$$Y\text{Position(mm)} = \mu^{*}\text{GazeY} \tag{4}$$

Of these, μ is the size of the pixel of the display utilized for eye searching. The calculation depends on monitor screen size and pixel clarity. The calculation of μ can be obtained from Equation 5 [10].

$$\mu = \text{dim}_{\overline{4}}/\text{dim}_{x} \tag{5}$$

Here, as shown in Equation 6, DIMM is the screen diagonal in millimeters (in inches), and dim is the screen diagonal in pixels [1].

$$\text{dim}_{p} = \sqrt{width_{p}^{2} + \text{height}_{p}^{2}} \tag{6}$$

Where height and width$_p$ are the height and width of the screen (in pixels).

IV. ON-SCREEN DISTANCE

As shown in Fig. 7 on the next page, when the origin of the gazel system (X_{pixel}; Y_{pixel}), the distance to the user's viewpoint on the screen is the origin and the specific viewpoint. The offset is explained as the length between the eye searching sensor and the bottom edge of the screen. If the searcher is directly connected to the bottom of the display, the offset will be 0 and the source is at the center of the display.

V. GAZE ANGLE RELATIVE TO THE EYE

The viewing angle of a site on the display related to the user's eyes can be determined in Equation 7 and Equation 8 [15].

$$\text{gaze angle}(\theta) = \tan^{-1}\text{OSDist IZ} \tag{7}$$

Here, Z is the length among the eyes and the display.

$$\text{EstGP}(mm) = \sqrt{\left((\text{Gaze}X)^{2} + (Gaze\_Y)^{2} + (Z)^{2}\right)} \tag{8}$$

A. Pixel Distance

At the time of calibration, few points will be showed on the display screen and the user will be prompted to view the showed points. The coordinates (x; y) showed on the display screen constitute the basic truth of the recorded image. 9 [21] can be used to calculate the offset between the actual ground coordinates (GTx; GTy) and the assumed viewpoint (Gaze_X; Gaze_Y).

$$\text{pix}_{shift}(\text{ pixels }) = \sqrt{\left((GT_{x} - \text{Gaze\_}X)^{2} + (GT_{y} - \text{ Gaze }\_Y)^{2}\right)} \tag{9}$$

B. Angular Accuracy

The line-of-sight estimate (or assumption error) of an eye-searcher (in degrees) can be showed as the angular variation among the actual line-of-sight position and the calculated line-of-sight position. Using Equations 7 8, 9, you can use Equation 10 [20] to calculate the forecast error.

$$\text{ang}_{acc} = \left(\mu \times \text{pix}_{shift} \times \cos\left(\text{mean}(\theta)^{2}\right)/E\,\text{stGP}\right. \tag{10}$$

C. Euclidean Distance in cm or mm

Several researches [10] and [18] use Euclidean distances to assess the accuracy of technologies. They give the Average Euclidean Distance (AED) from the actual gaze position. The mean Euclidean distance among the calculated gaze coordinates (x, y) and ground gaze coordinates can be acquired from 9.

$$AED = \frac{1}{n}\sum_{i=1}^{n}\sqrt{(gt\_xi - e\_xi)^{2} + (gt\_yi - e\_yi)^{2}} \tag{11}$$

D. *Loosely Correct Estimation Rate*

Line-of-sight precision can be calculated depending on the attenuated true rate estimate [3]. Loosely Correct Estimation Rate (LCER) is acquired by splitting the nearly correct number of frames by the all number of frames. A near-true frame is one in which the predicted line-of-sight is close to the actual.

The performance on the metrics for eye assessment has explained by L.Kr et. Al [23]. Much of the performance indicators used in this paper are dissimilar and cannot be contrasted with each other. Therefore, the designer must select an appropriate performance counter.

VI. FULLY CONVOLUTIONAL NEURAL NETWORKS

In the deep study, FCNNs is a type of in-depth neural network that is mainly used to access visual pictures [24]. They are motivated by the organization of the visual cortex [21]. The visual cortex is the major area of the mind that receives signals and processes the visual details transferred by the eye [25]. As described in Fig. 2, FCNNs are too likely to traditional neural programs, consisting of various neurons with trainable weights and distortions. Every neuron in the network takes input, performs an internal product operation, and by choice follows the dot product through a nonlinear operation. The FCNN architecture explicitly assumes that whole inputs are pictures, allowing utilizes to encode few attributes in the structure [26]. The first 3 levels are accountable for quality extraction, and fully interconnected layers are responsible for categorization [13].

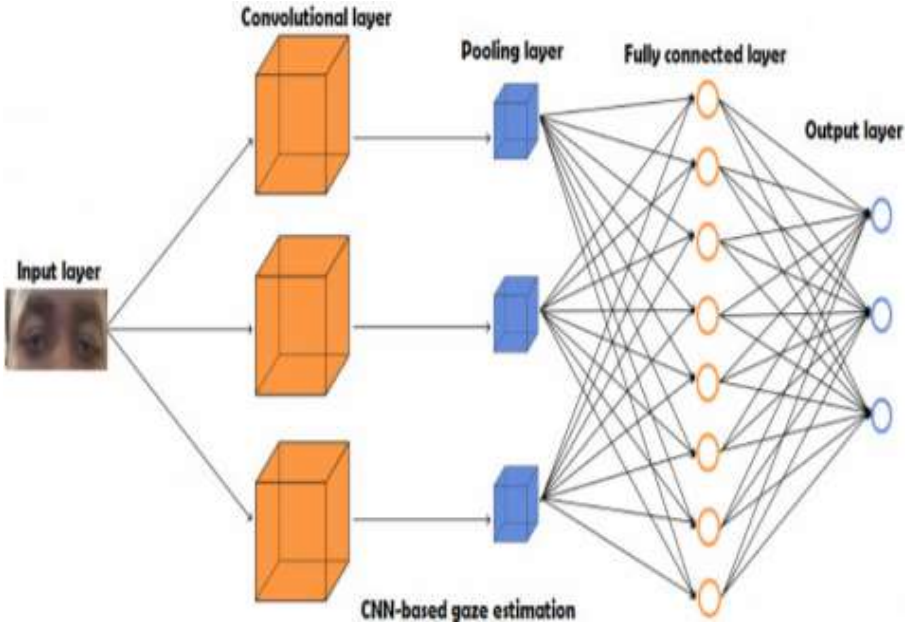

Fig. 2. A regular 3-layer and FCNN

A. *Conolutional Layer*

A convolution layer is made up of a group of trainable parameters, and these parameters are made up of a set of trainable filters. The width and height of each filter (also known as the core) are small but extend to the whole depth of the inlet volume [26]. When the input data is passed to the network, every filter depends on the height and width of input capacity and uses Equation **??** [26] to compute the point operation between the filter input and the input element. As the filter moves the input volume also moves with the height and width, as a result, a two-dimensional function map is generated.

$$H[i,j] = (m*k)[i,j] = \sum_{a}\sum_{b} k[a,b]m[i-a,j-b] \tag{12}$$

Here, m indicating the input picture, k represents the kernel, and a and b indicate the rows and columns of the result matrix.

B. Pooling Layer

The merged layer down the height and width of input data that results in a smaller volume. There are many types of pooling methods, which include maximum pooling and medium pooling. Takes the maximum value for each area represented by the 2 x 2 filter. Create a value in this area (for maximum union) or mean (for mean union) and make a new matrix with every entry in a new matrix. Both matrices are 4 matrices with a maximum or medium range of 4x2x2. As shown in Fig. 3, moving the input filter without overlapping the input area results in a 2x2 matrix.

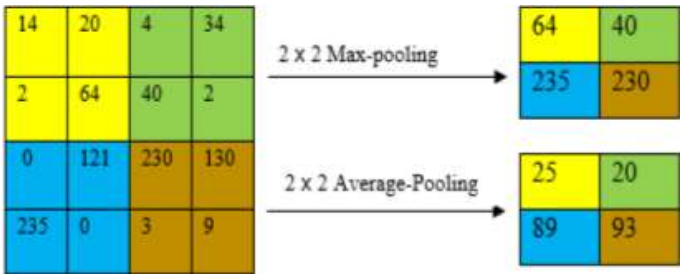

Fig. 3. Example of maximum and average pooling

C. Non-Linearity Layer

The non-linear layer is used to apply the element-wise activation activity to the input data. The activation activity determines either a neuron in the network needs to have functioned. Nonlinear activation performs nonlinear transformations on the input data so that the network can learn the complex associations in the dataset. There are several types of non-linear activation functions such as softmax, sigmoid, tanh, and linear rectifier unit Rectified Linear Unit (ReLu). The ReLu activation activity is most commonly used because it gives good results. The ReLu activation activity is defined by formula Equation 13 [13], as defined in Fig. 4. As you can see, the ReLu activity is straight for all positive numbers, but not for negative values. Displays all negative values as zero.

$$\text{relu}(x) = \max(0, x) \tag{13}$$

Where x is the input of the function

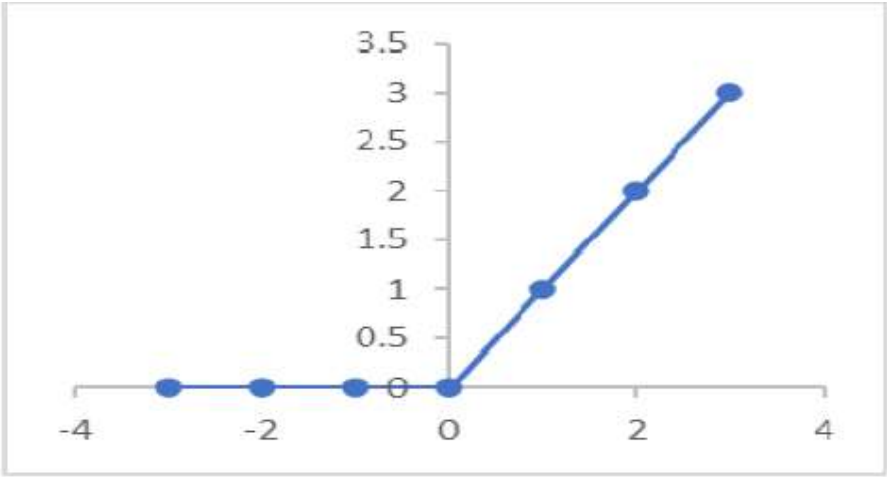

Fig. 4. ReLu activation function [27]

D. Fully Connected Layer

Fully connected layers perform the final categorization. It is consists of various neurons such as Neurons in artificial neural networks. Equation 14 Use [27] to compute the output of the neural layer Fully connected layers perform linear and nonlinear transformations of one-dimensional arrays. Use Equation 14 for linear transformations [22] and Equation 15 [13] for nonlinear transformations.

$$Z = W^T, X = b \tag{14}$$

Where X represents the input, b represents the variance, and W represents the studied parameter (or weight) in the network. W is a matrix of randomly initialized numbers.

$$\text{output } = \sigma(Z) \tag{15}$$

Where σ represents an activation function and Z represents the result of a linear transformation operation.

VII. MACHINE LEARNING BASED CLASSIFICATION METHOD

A. *Line-of-Sight Assessment Method Based on Two-Dimensional Classification*

1) Classification method for assessing the driver's gaze area: Various studies have developed classification models for assessing gaze. [12] proposed a five-layer FCNN-based technology to classify the driver's field of view and assess the head position. They created datasets using different images of male and female drivers (including glasses-wearing drivers). This study uses version [28] of the AlexNet architecture, and experimental results show that this method can provide 95% correct classification accuracy for eye detection. Significantly improved design of the personalized classification system for the driver's field of view. However, the design of singular systems lags [29]. A typical system should be able to classify the display area by different items and angles. [29] took it one step further by developing a universal eye recognition technology based on FCNN. The first strategy extracts the entire face for training, and the second strategy extracts the upper half of the face for training.

2) Line-of-sight classification method for input data: Several studies have proposed methods for assessing vision for an on-display screen keyboard. These methods use blinking as input to the eye searcher. They are very useful for text communication using Zhangetal eye movements. [13] Proposed an appearance-based calibration-free technology that permits users to enter the text simply by viewing at an on display of screen keyboard and blinking. These photos were taken with webcams, mobile phones, and electronic cameras. They applied various data enhancement techniques to enhance image clarity, lighting, little head rotation, and color stability of the subject's skin. They divided eye diseases into 10 groups, including 9 directions (see above) and 1 closed eye disease. They created an FCNN line-of-sight classification model based on a dataset that can learn 10 eye conditions from 2 eye pictures. The model was checked and its precision was 95%. [30] proposed a non-contact method for entering a line of sight using a head gesture. To enter a string, the user can display the desired character on the virtual QWERTY keyboard and perform a series of gestures. Gesture sequences are recorded by a face recognition system based on a deep neural network given by OpenCV [31]. The refined video is utilized as an input source to the pre-trained HopeNet model [32]. The HopeNet Model is an innovative FCNN-based head positioning model that calculates clear angles from EGB images. Then use the output of the pre-trained model to train the RNN model. This allows you to predict the order of clusters that users will see. The suggested cluster of the sequence is utilized to suppose the correct vocabulary that can be built from the sequence. This model was checked using data from 2234 videos from 22 objects with the precision of 21.81%.

3) Method for assessing line of sight based on three-dimensional classification: Eye-tracking technology provides high precision in assessing eye tracking in the x and y directions but does not depend on depth [33] Experimental results show that in the case of the model, the line-of-sight accuracy obtained by this method is 90.1%, and the line-of-sight accuracy obtained using the generalized model is 89.7%. [33] have suggested a neural network-depend method for estimating the depth of field of a forehead eye searcher. They utilized a two ocular eye searcher with 2 cameras to catch gaze details from an object observed at a fixed point of 1-5 meters. The recorded visual vector is utilized to train the NN model. The model is assessed and the outcomes showed that the mean classification fault generated by the model is < 10%.

4) Summary: Knowing the direction of the gaze can give important insight into the problems of various users. Various other methods for classifying gaze assessment have been proposed, including user access queue (EAC) detection methods. EAC refers to a specific eye movement pattern that can provide information about cognitive processes in the human mind [34, 35]. George and Rutray [22] developed an FCNN model based on real-time regression and low computational cost to estimate the likelihood of eye intrusion into the desktop surrounding. In this study, eye pictures were drawn from a data set involving facial images of various individuals.

VIII. LINE-OF-SIGHT CALCULATION DEPEND ON MACHINE LEARNING REGRESSION METHOD

A. *Line-of-Sight Calculation Method Depend on Two-Dimensional Regression*

1) Regression method by handwriting or in combination with external functions: The regression gaze scoring method attempts to evaluate the ability to map an input variable (x) to a numeric output parameter (y). Various regression solving methods have been developed to estimate the line of sight. [18] proposed a regression-depend gaze tracking method for devices called iTracker. They generated an extensive dataset containing 2 million images from over 1,450 mobile phones and tablets. Collect recordings in different lighting conditions, in different backgrounds, in different directions, and different head positions. The face mesh is a 25x25 binary input mask that can be utilized to define the position of the eyes and head for every picture. Depend on the data material; the iTracker system has been continuously trained to assume camera distance for mobile phones and tablets in the X and Y directions. Experimental outcomes show that the model assumption errors for mobile phones and tablets are 1.77 cm and 2.83 cm, respectively. Set up the trained model for each device and orientation. For mobile phones and tablets, the forecast error is reduced by 1.71 cm and 2.53 cm, respectively.

TABLE 1
OVERVIEW OF CLASSIFICATION METHODS BY APPEARANCE FOR PREDICTIVE CALCULATIONS

Rd.	Accuracy	Architecture	Data.wt	Image Resolution	Added Feature	Network Input	Application Area
[12]	95%	Own architecture	Own dataset	42 X 50	-	Full face	Automobile
[15]	6.7	Le Net	CAVE. Own dataset	28 X 28	-	Two eyes	VR/AR systems
[22]	89.81%	Own architecture	Eye Chimera	42 X 50	-	Two eyes	Desktop environments
[29]	93.36%	Alex Net. VGGI6	Own damsel	224 X 224. 227 X 227	•	Full face	Automobile
[13]	95.01%	Own architecture	Own douse	32 X 128		Two eyes	Gaze-based typing
[3]	92.8 - 99.6%	VGGFace16	CAVE, Own dataset	224 X 224	PCCR Vector	Full face and two eyes	Automobile
[36]	95.18%	Alex Net. VGG16. Res Net. Squeeze Net	Own dataset	224 X 224, 227 X 227	-	Full face	Automobile
[37]	89.7% -90.1%	Own Architecture	Own dataset	-	-	Two eyes	VR/AR systems
[38]	EER: 1%- 21.45%	SegNet	Own dataset	240 X 320		Two eyes	Forensics (post-mortem iris recognition)
[39]	EER: 0.85%	-	Own dataset	-	-	A sequence of gaze points from the two eyes	Biometrics
[40]	-	Own architecture	Own dataset	-	-	Two eyes	Biometrics

B. *3d Regression-Built Techniques for Gaze Calculation*

1) Regression techniques for free head 3D eye tracking: Few existing line-of-sight calculation methods can only estimate the line-of-sight position on the screen [7, 41] and do not give detail about the line-of-sight vector and eye position. These methods did not catch the linkage among head position, eye movements, and line-of-sight vectors. [42] They only need to study this linkage from the dataset, which leads to over fitting of the top-down relationship [42]. Several technologies offer advanced performance. However, using these methods, it is possible to predict only the gaze area on the screen, but not the three-dimensional surface [42]. Few applications require detail about a specific subject or area that the user is viewing [42]. These uses require a line-of-sight calculation method that can compute line-of-sight intersections between 3D line-of-sight vectors and different objects in a 3D view [42]. Examples of these applications consist of monitoring of attention of drivers, advertising investigation, and research [42]. Free heads line-of-sight technology evaluates eye position and line-of-sight vector in three-dimensional space [43].

TABLE 2
A SHORT SUMMARY OF APPEARANCE-DEPENDENT REGRESSION METHODS FOR GAZE CALCULATION

Ref	Accuracy	Architecture	Dataset	Image Resolution	Added Features	Network Input	Application area
[18]	$\tilde{2}$ cm, about 3	iTracker	Gaze Capture	224 X 224	Face grid	Full face and two eyes	Mobile phones and tablets
[10]	4.85 cm	iTracker	Gaze Capture	144 X 144	HOG	Full face and two eyes	Mobile phones
[21]	4.63	Own architectures	MPH Gaze, UT Multi view	-	-	Two eyes	Consumer electronic systems
[20]	1.53	DeepID	Own dataset	40 x 70	-	One eye	In the wild environments
[17]	4.8 , 6.0	Alex Net	MPH Gaze, EYED1AP	448 X448	Spatial weights	Full face	217 and 3D gaze estimation
[44]	29.53 mm	iTracker	Own Dataset	400 X 120	-	Face	-
[1]	1.0 , 1.3	Own Architecture	MIT	100 X 100	-	Fixation patch	Indoor environments
[11]	13.9	Own Architecture	MPII Gaze	36 X60	-	Full face	In the wild environments
[42]	4.3	Alex Net	MPII Gaze. UT Multi view	224x224	Gaze transform layer	Full face and one eye	3D gaze tracking
[43]	6.5	Own architecture	Own dataset	15 X 9	-	Synthesized eye images	3D gaze tracking
[45]	5.1 and 6.2	VGG-I6	EYEDIAP	128X48	Facial landmarks	Full face and two eyes	3D gaze tracking

IX. LINE-OF-SIGHT ESTIMATION METHOD BASED ON THREE-DIMENSIONAL REGRESSION

A. *Regression Method For 3D Line-of-Sight Tracking*

Some existing line-of-sight estimation methods can only estimate the line-of-sight position on the screen [7, 41] and cannot provide information on the line-of-sight vector and eye position. These methods do not capture the relationship between head position, eye movements, and gaze vector. [42] Merely examining this relationship from the dataset results in over fitting of the top-down relationship [42]. Examples of such applications include driver attention monitoring, advertising analysis, and research [42]. Free head line-of-sight technology can estimate eye position and line-of-sight vector in three-dimensional space [43]. [42] have proposed an efficient and inexpensive three-dimensional eye assessment method for tracking eye movements. They proposed a strategy for creating a gaze scoring model that can effectively capture the free movement of the head and eyes. In this strategy, gaze scoring is divided into two separate modeling tasks, and two different FCNN models are trained to simulate eye and head movements. Table 2 lists several selected regression-based gaze estimation methods.

X. DISCUSSION

This section describes an FCNN-based gaze assessment method. This section is divided into four subsections. The first part describes the existing method for assessing vision, and the second part describes the comparison between the calibration-based method and the FCNN method. The third subsection contains the various datasets used to evaluate the performance of the FCNN model, and the last subsection contains information about the network parameters used to train the FCNN gaze estimation model.

A. *Machine-Based Line-of-Sight Tracking*

Eye-tracking is a very interesting area that attracts the attention of many researchers. It is an excellent tool for studying human behavior in various fields such as psychology, medicine, marketing, and automobiles [3, 46]. Since vision can be used for navigation and control, it can also be used to improve human-computer interaction. Many methods have been proposed and used in the literature for torque estimation [19]. However, we recently discovered that it differs from traditional eye-tracking methods, which are mostly model-based. Compared to model-based methods, gaze-based assessment has not yet yielded many benefits. Several existing appearance-based methods solve the problem of 2D and 3D line-of-sight estimation.

B. *Calibration Based Eye Assessment and Deep Learning Eye Assessment*

This section compares calibration-based eye estimation methods with deep learning-based eye estimation methods. Calibration-based methods can give better results than non-calibrated methods. As shown in Table 3, calibration-based models are well suited for object- and

posture-independent applications. This is because some heads have a fixed position [17] and the head cannot move freely. Most of them also require training in specific areas, and some of them are assessed in specific areas, which limits the effectiveness of generalization [47, 48]. In contrast, deep learning gaze assessment models are posture and object independent and can perform gaze assessments that do not require complex adjustments or calibration procedures. They range from 50°C [49, 42, 43] [79] to over 1000 points [18] and are therefore more versatile (compared to calibration-based methods). Shows calibration properties. The head can move freely, so it can be used for unlimited line-of-sight tracking. It can also be extended to multiple devices and directions.

TABLE 3
COMPARISON OF TECHNOLOGIES BASED ON CALIBRATION AND FCNN

Rd	Accuracy	Configuration	1 Lead Movement
		CNN-based Techniques	
[18]	$\tilde{2}$ cm	Eye and Face Images, Face grid	Yes
[21]	4.63	Eye images	Yes
[20]	1.53	Eye images	Yes
[17]	4.8, 6.0	Image dataset	Yes
[11]	1.0, 1.3	Image dataset	Yes
[12]	95%	Full face images	Yes
[42]	4.3	Face and eye images	Yes
[29]	93.36%	Face images	Yes
[12]	95.01%	Eye images	Yes
[3]	92.8 - 99.6%	Face and eye images	Yes
[36]	95.18%	Face images Calibration - based techniques	Yes
[50]	1.25	2 cameras, 2 light sources	Yes
[47]	3.78	1 Kinect sensor integrated LEDs	Yes
[51]	0.6	4 Cameras, 2 LEDs. Stereo systems	Yes
[52]	1.0	1 camera. Stereo system, and LED	Yes
[53]	0.87	1 camera, 1 infrared light source	No
[54]	1.5	2 video cameras	Yes
[55]	1.11	2 IR light Sources	No
[48]	1.3	4 LEDs and Webcam	Yes
[55]	1.4	1 Video Camera	No

C. *Eye Assessment Protocol*

Some early research used small datasets to create appearance-based models. However, it is impossible to generalize these models to real estimates. Visual field assessment is characterized by changes in light, eye images, and head position. Thus, the new appearance-based model trains the model on large datasets. For example, the Gaze Capture dataset contains 1.4 million available images, collected by over 1,400 users in a variety of head positions, looks, and lighting conditions. The MPIIGaze dataset contains 213,000 images of 15 people with 15 different eye positions. The dataset was collected over 3 months of daily laptop use. The position and coverage of some of the heads in these datasets are still limited. Proposed a new learning framework that combines multiple images of the eye area. In this study, a combination of a 3D eye model and a real-time rendering structure was used to create one million composite images. By comparing the performance of a dataset to other datasets using real images, you can gauge the quality and ease of use of the dataset and get competitive results. This is because the ImageNet based model has captured various features such as curves and edges (original planes). This is very useful for most classification tasks.

TABLE 4
SEVERAL DATASETS SELECTED FOR EYE ASSESSMENT

Rd.	Dataset	Year	Size	Head Poses	Gaze Direction	Identities	Purpose
[56]	PUT	2008	9971	5	-	100	Face recognition
[41]	HPEG	2009	20 videos	2	-	10	Evaluation of eye gaze and head pose estimation.
[57]	RS-DMV	2010	10 videos	-	-	-	Eye gaze and had pose estimation
[58]	WD Ref	2012	99773	NA	NA	2,995	Face recognition
[49]	CAVE-DB	2013	5880	5	105	56	Face recognition
[43]	Celeb face	2014	202599	Continuous	NA	10.177	Image classification
[59]	Eye dip	2014	videos	Continuous	Continuous	16	Face recognition
[43]	UT Multi view	2014	64000	8+synthesized	160	50	Eye gaze estimation
[60]	VGG-Face	2015	2.6 million	Continuous	NA	2622	Appearance-based gaze estimation.
[61]	Tablet Gaze	2015	816 videos	4	35	51	Face recognition
[62]	Google	2015	200 million	Continuous		8 million	Unconstrained gaze estimation
[11]	MPII Gaze	2015	213659	Continuous	Continuous	15	Evaluation of gaze tracking techniques
[18]	Gaze Capture	2016	2445504	Continuous	13+ Continuous	1474	Gaze estimation
[3]	DDGC- DBI	2018	39108	Continuous	17	20	Driver classification Gaze Near-eye gaze estimation
[63]	Nv gaze	2019	2 million	Continuous	-	35	
[63]	Nv gaze	2019	2.5 million	-	-	-	Near-eye gaze estimation

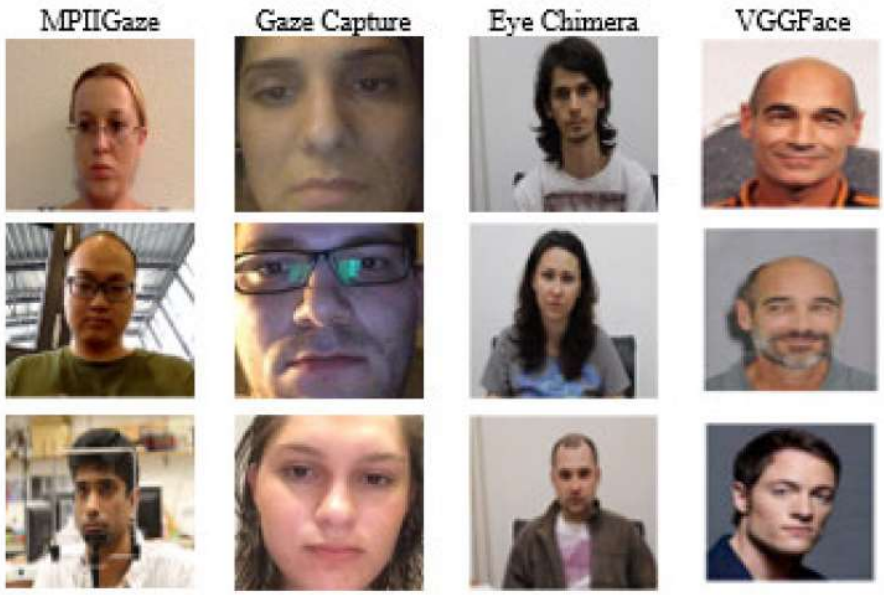

Fig. 5. Sample images from four datasets [64]

D. *Network Parameters Used to Train Appearance-Based Gaze Assessment Models*

As shown in Table 5, FCNN training requires parameters for heart rate, number of iterations, mini batch size, learning rate, and weight loss. Select a value for each setting based on the problem you want to solve. Most research has shown that the Stochastic Gradient Descent (SGD) optimizer performs well at 0.9 momentums. Momentum helps speed up the gradient vector in the right direction, speeding up convergence. Especially when fine-tuning the model, it is recommended to use a lower learning rate to maintain the weight of the previous workout. In Fig. 6-9 also show some views of the deep learning architecture.

TABLE 5
SOME PARAMETERS SELECTED FOR TRAINING THE DEEP LEARNING GAZE ASSESSMENT MODEL

Ref	Momentum	Initial Learning Rate	Weight Decay	Mini-Batch Size	Epoch	Optimizer
[18]	0.9	0.001, reduced to 0.0001 after 75,000 iterations.	0.0005	256	150,000	-
[3]	0.9	0.000001	0.0005	20	16	SGD
[10]	0.9	0.00001	0.9epoch	20	10	SGD
[29]	-	0.0001	-	32 (Alex Net) & 64 (V0016)	5	Adam
[36]	-	4 x 10-4 (for Squeeze Net) and 0.0001 (for Alex Net, VGG16 and ResNet50)	-	64 (Alex Net& Squeeze Net 32 (VGG16) & 64 ResNet50)	50	Gradient Descent
[29]	0.9	0.001, decreased 3 times	0.0005	256	74	SGD
[65]	0.9	0.1. divided by 10 when the error plateaus.	0.0001	256	60 x 104	SGD
[46]	0.9	0.01, manually decreased by order of magnitude as soon as the validation error stops decreasing. Decreased finally to 0.0001.	-	128	15	SGD

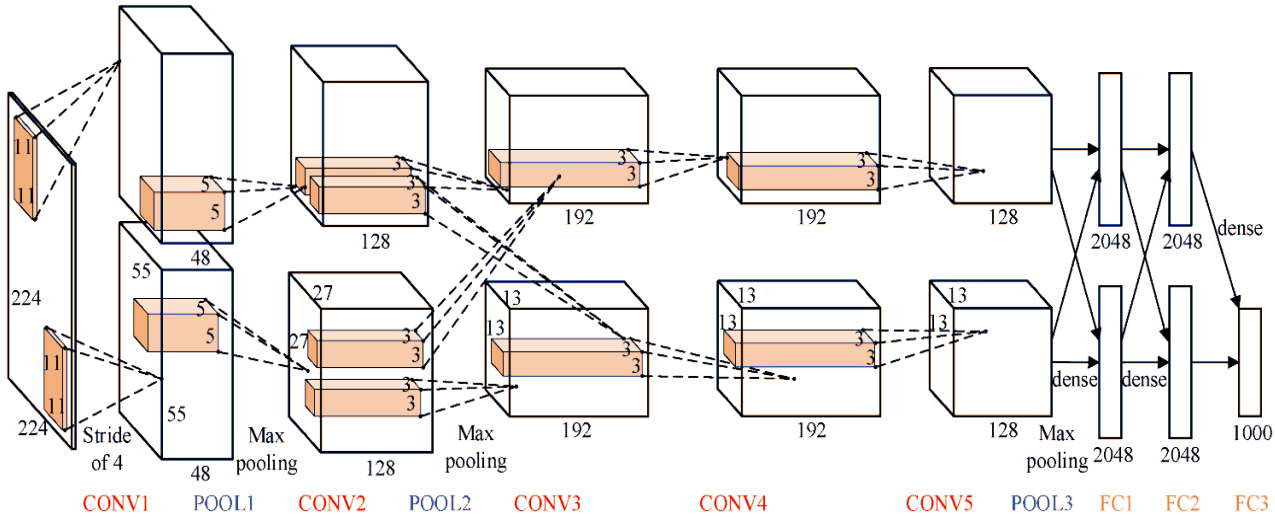

Fig. 6. Alex Net network architecture [28]

XI. LIMITATIONS, FUTURE WORK INSTRUCTIONS AND SUMMARIES

A. *Limitations and Future Work Instructions*

The experimental results given in [1] show that some of the proposed methods for assessing visual acuity give the most recent results. However, all research usually has limitations that form the basis of future research. This section contains some restrictions on FCNN-based gaze assessment methods.

The calibration-based gaze estimation method has a better prediction error than the FCNN-based method. The prediction error of most FCNN-based in-line estimation methods is greater than 0.5 mm. One of the main reasons is the lack of large balanced variable data. Most of the datasets available do not reflect design and instructions very well. Images in these datasets are captured by multiple devices, but the variability and amount of training data (displayed in the dataset) are imbalanced for each device.

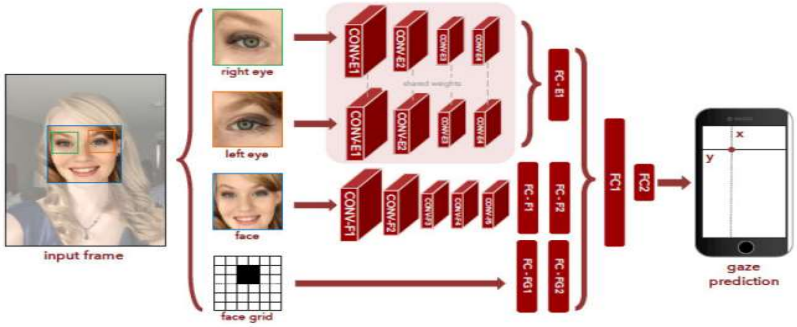

Fig. 7. Tracker architecture [18]

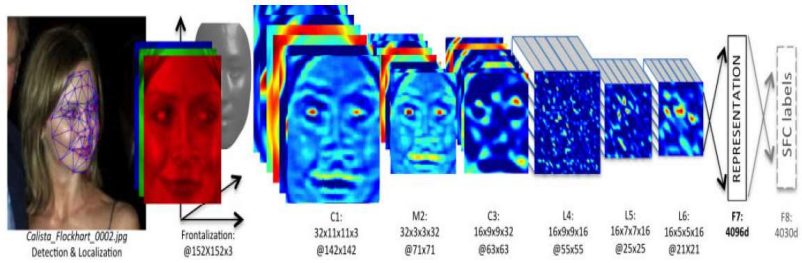

Fig. 8. Deep face architecture [46]

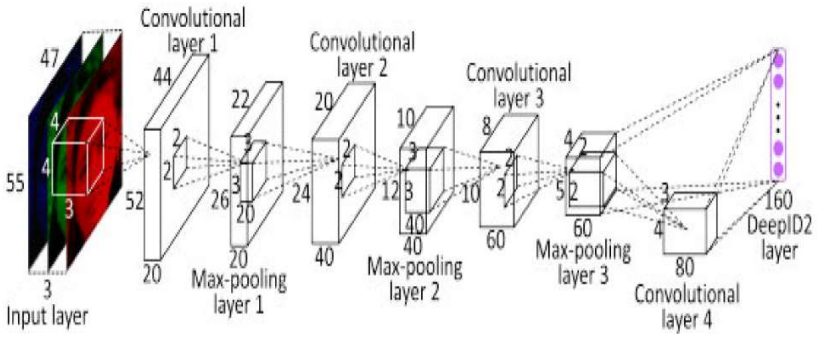

Fig. 9. DeepID2 architecture

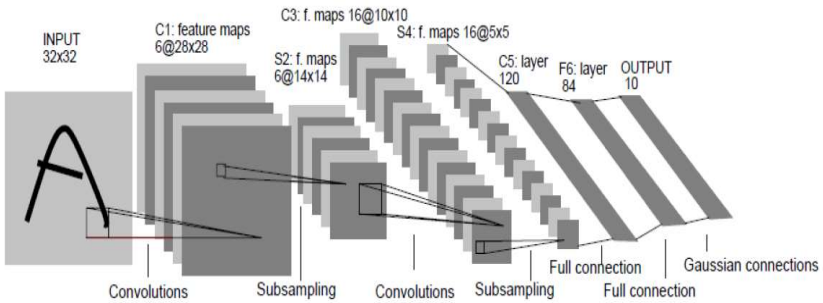

Fig. 10. LeNet-5 architecture

Some observations in the literature suggest that building a robust and efficient FCNN model can be time-consuming, time-consuming, and expensive to compute. FCNN is very time-consuming because the developers perform many resource-intensive operations that require access to very fast computers, for example, B. A GPU-accelerated cluster computer or a high-speed multi-core processor.

Several studies [17] have offered a scatter plot gaze rating model besides a general two-dimensional gaze rating model. Experimental conclusions presented in the material show that the performance of the heat map model is lower than that of the 2D star point model.

TABLE 6
VGG-16 NETWORK ARCHITECTURE [66]

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
Input (224 x 224 RGB image)					
conv3-64 LRN	conv3-64	conv3-64 conv3-64	conv3-64	conv3-64	conv3-64
			conv3-64	conv3-64	conv3-64
Maxpool					
conv3-128	conv3-128	conv3-128	conv3-128	corn/3-128	conv3-128
		conv3-128	conv3-128	conv3-128	I conv3-128
Maxpool					
conv3-256	conv3-250	conv3-250	conv3-250	conv3-256	conv3-250
conv3-256	conv3-256	conv3-256	conv3-256	conv3-256	conv3-256
			conv1-256	conv3-256	conv3-256
					conv3-256
Maxpool					
conv3-512	conv3-512	conv3-512	conv3-512	conv3-512	conv3-512
coav3-512	conv3-512	conv3-512	conv3-512	conv3-512	conv3-512
			conv1-512	conv3-512	conv3-512
					conv3-512
Maxpool					
conv3-512	conv3-512	conv3-512	conv3-512	conv3-512	
conv3-512	conv3-512	conv3-512	conv3-512	conv3-512	
	conv3-512		conv3-512	conv3-512	
	conv3-512			conv3-512	
	convl-512				
Maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Various the FCNN-based technologies presented in this paper are a representation for an unrestricted environment where different head postures, appearance, lighting, face closure, etc. may affect the captured images. Their conclusions indicate that there is yet a lot of area for improvement. Studies in the future may focus on developing reliable, improved, unlimited technologies that can handle many practical changes.

However, the same settings were made to the ROI image. Differences between the complete model and the ROI imaging model can accelerate changes that negatively impact the viewing correct of the final model. For better results, models are drawing for ROI pictures should further use pertained with techniques ROI images rather than full images.

Training FCNN with low-quality images can improve the learning rate, but the results presented in [15] show that poor quality pictures significantly impact the gaze correct of instantaneous FCNN-based estimates. Study in the future needs to take into account the development of technologies that complement speed accuracy.

Several gaze assessment models put forward in the literature have directed on corresponding small datasets .Consequently, they did not yield competitive conclusions.

Building efficient and powerful deep learning models typically require large, well-balanced datasets. This should be the focus of future research.

XII. CONCLUSION

In this article, we'll show you how to gauge your visibility by your appearance and focus on FCNN. Line-of-sight estimation techniques can be classified into model-depend methods and appearance-depend methods. Model-depend methods can be divided into temporal reflex methods and shape-depend methods. They use eye reflexes (also known as blinking) to assess the eyes to determine the position of the eye in three-dimensional space. The corneal reflex method relies on using an external light source to detect features, while the shape-based method relies on the observed shape of the eye to assess the line of sight, such as the edges of the iris or the center of the pupil. Unfortunately, these techniques do not accurately handle poor quality images or pictures with varying lighting order. The appearance-based classification depends on the appearance of the assessing look. Displays the functionality of the image directly to the gaze point without using manually created functions. Unlike a model-based method, you can process low-resolution images reliably. You can also summarize new profiles nicely without use custom proof. Early appearance-depend classifications introduce survey assessment methods that allow inference about stuck head positions and user training proof. Subsequent papers presented improved techniques that could handle various head postures, light, and surfaces. However, all of these methods need subject learning of the model. Recent studies [11] and [17] have focused on establishing techniques that can handle deferential head postures or the assessment of object-independent gaze. However, some of these methods require large priming datasets. To meet that demand, various studies [11] have provided large gaze assessments, consisting of pictures of various head positions and lighting situations. Because of its popularity and effectiveness, this paper focuses on deep learning conclusions (and other machine learning conclusions). The success of deep learning algorithms is driven by the accessibility of large datasets and modular computing assets such as thousands of CPU cores and GPUs. The main objective of this learning is to provide the experimentation community with an exclusive benchmark that is appropriate for improving the development of modern methods for assessing gaze based on appearance. We hope this learning will help the experimentation community and facilitate the development of improved gape assessment methods.

REFERENCES

[1] K. Wang, S. Wang, and Q. Ji, "Deep eye fixation map learning for calibration-free eye gaze tracking," in *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications,* New York, NY, 2016. doi: https://doi.org/10.1145/2857491.2857515

[2] R. A. Naqvi, M. Arsalan, G. Batchuluun, H. S. Yoon, and K. R. Park, "Deep learning-based gaze detection system for automobile drivers using a NIR camera sensor," *Sensors*, vol. 18, no. 2, p. 456, 2018. doi: https://doi.org/10.3390/s18020456

[3] S. Robertson, G. Penn, and Y. Wang. (2019) Exploring spectro-temporal features in end-to-end convolutional neural networks. [Online]. Available: https://arxiv.org/pdf/1901.00072.pdf

[4] Y. Kim, Y. Jernite, D. Sontag, and A. M. Rush, "Character-aware neural language models," in *Thirtieth AAAI Conference on Artificial Intelligence,* Menlo Park, CA, 2016.

[5] W. Wang and J. Shen, "Deep visual attention prediction," *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2368–2378, 2017. doi: https://doi.org/10.1109/TIP.2017.2787612

[6] A. Tsukada, M. Shino, M. Devyver, and T. Kanade, "Illumination-free gaze estimation method for first-person vision wearable device," in *International Conference on Computer Vision Workshops (ICCV Workshops),* Barcelona, Spain, 2011. doi: https://doi.org/10.1109/ICCVW.2011.6130505

[7] D. W. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, pp. 478–500, 2009. doi: https://doi.org/10.1109/TPAMI.2009.30

[8] C. N. M. Locsin and R. J. Ferolin, "Neural networks application for water distribution demand-driven decision support system," *Journal of Advances in Technology and Engineering Research*, vol. 4, no. 4, pp. 160–175, 2018. doi: https://doi.org/10.20474/jater-4.4.3

[9] H. Chennamma and X. Yuan. (2013) A survey on eye-gaze tracking techniques. [Online]. Available: https://arxiv.org/abs/1312.6410

[10] M. Quazi, "An overview of laser welding of high strength steels for automotive application," *International Journal of Technology and Engineering Studies*, vol. 6, no. 1, pp. 23–40, 2020. doi: https://doi.org/10.20469/ijtes.6.10004-1

[11] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, "Appearance-based gaze estimation in the wild," in

Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, 2015.

[12] S. F. Haider, M. Quazi, J. Bhatti, M. N. Bashir, and I. Ali, "Effect of Shielded Metal Arc Welding (SMAW) parame-ters on mechanical properties of low-carbon, mild and stainless-steel welded joints: A review," *Journal of Advances in Technology and Engineering Research*, vol. 5, no. 5, pp. 191–198, 2019. doi: https://doi.org/10.20474/jater-5.5.1

[13] C. Zhang, R. Yao, and J. Cai, "Efficient eye typing with 9-direction gaze estimation," *Multimedia Tools and Applications*, vol. 77, no. 15, pp. 19 679–19 696, 2018. doi: https://doi.org/10.1007/s11042-017-5426-y

[14] N. Z. T. Abdulnabi and O. Altun, "Batch size for training convolutional neural networks for sentence classification," *Journal of Advances in Technology and Engineering Studies*, vol. 2, no. 5, pp. 156–163, 2016. doi: https://doi.org/10.20474/jater-2.5.3

[15] I. Ali, N. Lin, M. Quazi, M. N. Bashir, H. Sadiq, and F. Sharaf, "Investigating the wear characteristics of metal matrix composite coating deposited on AA5083 Al-alloy by laser surface engineering technique," *North American Academic Research*, vol. 3, no. 1, pp. 138–146, 2019. doi: https://doi.org/10.5281/zenodo.3626448

[16] Y. Sugano, Y. Matsushita, Y. Sato, and H. Koike, "Appearance-based gaze estimation with online calibration from mouse operations," *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 6, pp. 750–760, 2015. doi: https://doi.org/10.1109/THMS.2015.2400434

[17] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, "It's written all over your face: Full-face appearance-based gaze estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops,* Honolulu, HI, 2017.

[18] M. Gul, N. Zulkifli, M. Kalam, H. Masjuki, M. Mujtaba, S. Yousuf, M. N. Bashir, W. Ahmed, M. Yusoff, S. Noor *et al.*, "RSM and artificial neural networking based production optimization of sustainable cotton bio-lubricant and evaluation of its lubricity & tribological properties," *Energy Reports*, vol. 7, pp. 830–839, 2021. doi: https://doi.org/10.1016/j.egyr.2021.01.033

[19] M. M. Quazi, M. Ishak, A. Arslan, M. Nasir Bashir, and I. Ali, "Scratch adhesion and wear failure characteristics of pvd multilayer crti/crtin thin film ceramic coating deposited on AA7075-T6 aerospace alloy," *Journal of Adhesion Science and Technology*, vol. 32, no. 6, pp. 625–641, 2018. doi: https://doi.org/10.1080/01694243.2017.1373988

[20] Y. Wang, T. Shen, G. Yuan, J. Bian, and X. Fu, "Appearance-based gaze estimation using deep features and random forest regression," *Knowledge-Based Systems*, vol. 110, pp. 293–301, 2016. doi: https://doi.org/10.1016/j.knosys.2016.07.038

[21] M. M. Quazi, M. Ishak, A. Arslan, M. Fazal, F. Yusof, B. Sazzad, M. N. Bashir, and M. Jamshaid, "Mechanical and tribological performance of a hybrid MMC coating deposited on Al-17Si piston alloy by laser composite surfacing technique," *RSC Advances*, vol. 8, no. 13, pp. 6858–6869, 2018. doi: https://doi.org/10.1039/C7RA08191J

[22] A. George and A. Routray, "Real-time eye gaze direction classification using convolutional neural network," in *International Conference on Signal Processing and Communications (SPCOM),* Bangalore, India, 2016.

[23] A. Kar and P. Corcoran, "Performance evaluation strategies for eye gaze estimation systems with quantitative metrics and visualizations," *Sensors*, vol. 18, no. 9, pp. 1–35, 2018. doi: https://doi.org/10.3390/s18093151

[24] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: An overview and application in radiology," *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, 2018. doi: https://doi.org/10.1007/s13244-018-0639-9

[25] T. Huff, N. Mahabadi, and P. Tadi, *Neuroanatomy, Visual Cortex*. Treasure Island, FL, StatPearls Publishing, 2020.

[26] S. Wakeel, S. Ahmad, S. Bingol, M. N. Bashir, T. C. Paçal, and Z. A. Khan, "Supplier selection for high temperature die attach by hybrid entropy-range of value MCDM technique: A semiconductor industry," in *21st International Conference on Electronic Packaging Technology (ICEPT),* Guangzhou, China, 2020. doi: https://doi.org/10.1109/ICEPT50128.2020.9202994

[27] B. Ramsundar and R. B. Zadeh, *TensorFlow for deep learning: From linear regression to reinforcement learning*. Sebastopol, CA: O'Reilly Media, Inc, 2018.

[28] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.

[29] S. Vora, A. Rangesh, and M. M. Trivedi, "On generalizing driver gaze zone estimation using convolutional neural networks," in *Intelligent Vehi-*

cles Symposium (IV), Los Angeles, CA, 2017. doi: https://doi.org/10.1109/IVS.2017.7995822

[30] S. Rustagi, A. Garg, P. R. Anand, R. Kumar, Y. Kumar, and R. R. Shah, "Touchless typing using head movement-based gestures," in *IEEE Sixth International Conference on Multimedia Big Data (BigMM),* New Delhi, India, 2020. doi: https://doi.org/10.1109/BigMM50055.2020.00025

[31] G. Bradski, "The openCV library," *Dr. Dobb's Journal: Software Tools for the Professional Programmer*, vol. 25, no. 11, pp. 120–123, 2000.

[32] S. Wakeel, S. Bingol, S. Ahmad, M. N. Bashir, M. S. M. M. Emamat, Z. Ding, and F. Hussain, "A new hybrid LGPMBWM-PIV method for automotive material selection," *Informatica*, vol. 45, no. 1, pp. 105–115, 2021.

[33] Y. Lee, C. Shin, A. Plopski, Y. Itoh, T. Piumsomboon, A. Dey, G. Lee, S. Kim, and M. Billinghurst, "Estimating gaze depth using multi-layer perceptron," in *International Symposium on Ubiquitous Virtual Reality (ISUVR),* Nara, Japan, 2017. doi: https://doi.org/10.1109/ISUVR.2017.13

[34] L. L. Di Stasi, R. Renner, A. Catena, J. J. Cañas, B. M. Velichkovsky, and S. Pannasch, "Towards a driver fatigue test based on the saccadic main sequence: A partial validation by subjective report data," *Transportation Research Part C: Emerging Technologies*, vol. 21, no. 1, pp. 122–133, 2012.

[35] S. D. Goldinger and M. H. Papesh, "Pupil dilation reflects the creation and retrieval of memories," *Current Directions in Psychological Science*, vol. 21, no. 2, pp. 90–95, 2012. doi: https://doi.org/10.1177/0963721412436811

[36] S. Vora, A. Rangesh, and M. M. Trivedi, "Driver gaze zone estimation using convolutional neural networks: A general framework and ablative analysis," *IEEE Transactions on Intelligent Vehicles*, vol. 3, no. 3, pp. 254–265, 2018. doi: https://doi.org/10.1109/TIV.2018.2843120

[37] C. Shin, G. Lee, Y. Kim, J. Hong, S.-H. Hong, H. Kang, and Y. Lee, "Evaluation of gaze depth estimation using a wearable binocular eye tracker and machine learning," *Journal of The Korea Computer Graphics Society*, vol. 24, no. 1, pp. 19–26, 2018.

[38] M. Trokielewicz, A. Czajka, and P. Maciejewicz, "Post-mortem iris recognition with deep-learning-based image segmentation," *Image and Vision Computing*, vol. 94, pp. 1–34, 2020. doi: https://doi.org/10.1016/j.imavis.2019.103866

[39] S. Jia, A. Seccia, P. Antonenko, R. Lamb, A. Keil, M. Schneps, M. Pomplun *et al.*, "Biometric recognition through eye movements using a recurrent neural network," in *International Conference on Big Knowledge (ICBK),* Singapore, 2018. doi: https://doi.org/10.1109/ICBK.2018.00016

[40] Z. Liang, F. Tan, and Z. Chi, "Video-based biometric identification using eye tracking technique," in *International Conference on Signal Processing, Communication and Computing (ICSPCC 2012),* Hong Kong, China, 2012. doi: https://doi.org/10.1109/ICSPCC.2012.6335584

[41] O. Ferhat and F. Vilariño, "Low cost eye tracking: The current panorama," *Computational Intelligence and Neuroscience*, vol. 2016, pp. 1–14, 2016. doi: https://doi.org/10.1155/2016/8680541

[42] W. Zhu and H. Deng, "Monocular free-head 3d gaze tracking with deep learning and geometry constraints," in *Proceedings of the IEEE International Conference on Computer Vision,* Venice, Italy, 2017.

[43] Y. Sugano, Y. Matsushita, and Y. Sato, "Learning-by-synthesis for appearance-based 3d gaze estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* Columbus, OH, 2014, pp. 1821–1828.

[44] D. Masko, "Calibration in eye tracking using transfer learning," KTH Royal Institute of Technology, Stockholm, Sweden, M.S. thesis, 2017.

[45] C. Palmero, J. Selva, M. A. Bagheri, and S. Escalera. (2018) Recurrent cnn for 3d gaze estimation using appearance and shape cues. [Online]. Available: https://arxiv.org/abs/1805.03064

[46] J. Niemann and C. Fussenecker, "Analysis of eye tracking usage in different domains and possible applications in the engineering environment," in *International Conference on Quality and Innovation in Engineering and Management,* Napoca, Romania, 2014.

[47] X. Zhou, H. Cai, Z. Shao, H. Yu, and H. Liu, "3d eye model-based gaze estimation from a depth sensor," in *International Conference on Robotics and Biomimetics (ROBIO),* Qingdao, China, 2016. doi: https://doi.org/10.1109/ROBIO.2016.7866350

[48] M. Gul, M. Kalam, M. Mujtaba, S. Alam, M. N. Bashir, I. Javed, U. Aziz, M. R. Farid, M. T. Hassan, and S. Iqbal, "Multi-objective-optimization of process parameters of industrial-gas-turbine fueled with natural gas by using Grey-Taguchi and ANN methods for better performance," *Energy Reports*, vol. 6, pp. 2394–2402, 2020. doi: https://doi.org/10.1016/j.egyr.2020.08.002

[49] B. A. Smith, Q. Yin, S. K. Feiner, and S. K. Nayar, "Gaze locking: passive eye contact detection for human-object interaction," in *Proceedings of the 26th annual ACM symposium on User Interface Software and Technology,* New York, NY, 2013. doi: https://doi.org/10.1145/2501988.2501994

[50] I. Ali, M. Quazi, E. Zalnezhad, A. A. Sarhan, N. L. Sukiman, and M. Ishak, "Hard anodizing of aerospace AA7075-T6 aluminum alloy for improving surface properties," *Transactions of the Indian Institute of Metals*, vol. 72, no. 10, pp. 2773–2781, 2019. doi: https://doi.org/10.1007/s12666-019-01754-5

[51] Vision Aware. (2020) Driving with low vision. [Online]. Available: https://bit.ly/3myBojj

[52] Y. LeCun. (2013) LeNet-5, convolutional neural networks. [Online]. Available: https://bit.ly/3BwOphx

[53] C. Ma, K.-A. Choi, B.-D. Choi, and S.-J. Ko, "Robust remote gaze estimation method based on multiple geometric transforms," *Optical Engineering*, vol. 54, no. 8, p. 083103, 2015. doi: https://doi.org/10.1117/1.OE.54.8.083103

[54] C.-C. Lai, S.-W. Shih, and Y.-P. Hung, "Hybrid method for 3-D gaze tracking using glint and contour features," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 1, pp. 24–37, 2014.

[55] R. Newman, Y. Matsumoto, S. Rougeaux, and A. Zelinsky, "Real-time stereo tracking for head pose and gaze estimation," in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580),* Grenoble, France, 2000. doi: https://doi.org/10.1109/AFGR.2000.840622

[56] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* Boston, MA, 2015.

[57] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter, "A 3d face model for pose and illumination invariant face recognition," in *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance,* Genova, Italy. Ieee, 2009, pp. 296–301.

[58] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proceedings of British Machine Vision Association,* Swansea, UK, 2015. doi: https://dx.doi.org/10.5244/C.29.41

[59] A. Galante and P. Menezes, "A gaze-based interaction system for people with cerebral palsy," *Procedia Technology*, vol. 5, pp. 895–902, 2012. doi: https://doi.org/10.1016/j.protcy.2012.09.099

[60] P. Blignaut, "Mapping the pupil-glint vector to gaze coordinates in a simple video-based eye tracker," *Journal of Eye Movement Research*, vol. 7, no. 1, pp. 1–11, 2014. doi: https://doi.org/10.16910/jemr.7.1.4

[61] J. Zhu and J. Yang, "Subpixel eye gaze tracking," in *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition,* Washington, DC, WA, 2002. doi: https://doi.org/10.1109/AFGR.2002.1004144

[62] Z. Zhu, Q. Ji, and K. P. Bennett, "Nonlinear eye gaze mapping function estimation via support vector regression," in *18th International Conference on Pattern Recognition (ICPR'06),* Hong Kong, China, vol. 1, 2006. doi: https://doi.org/10.1109/ICPR.2006.864

[63] J. Kim, M. Stengel, A. Majercik, S. De Mello, D. Dunn, S. Laine, M. McGuire, and D. Luebke, "Nvgaze: An anatomically-informed dataset for low-latency, near-eye gaze estimation," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems,* New York, NY, 2019. doi: https://doi.org/10.1145/3290605.3300780

[64] Y.-G. Shin, K.-A. Choi, S.-T. Kim, C.-H. Yoo, and S.-J. Ko, "A novel 2-D mapping-based remote eye gaze tracking method using two IR light sources," in *IEEE International Conference on Consumer Electronics (ICCE),* Las Vegas, NV, 2015. doi: https://doi.org/10.1109/ICCE.2015.7066375

[65] J. M. Franchak, K. S. Kretch, K. C. Soska, and K. E. Adolph, "Head-mounted eye tracking: A new method to describe infant looking," *Child Development*, vol. 82, no. 6, pp. 1738–1750, 2011. doi: https://doi.org/10.1111/j.1467-8624.2011.01670.x

[66] K. Simonyan and A. Zisserman. (2014) Very deep convolutional networks for large-scale image recognition. [Online]. Available: https://arxiv.org/abs/1409.1556