#### ABSTRACT

The prediction of individual eye gaze is a research topic that has gained the interest of researchers with its wide range of applications because neural networks majorly increase the rate of accuracy of individual gaze. In our research work, we have been able to predict individual gaze using MPIIGaze dataset. We categorize the gaze prediction into four (4) directions as to whether an individual is looking downwards, left and right directions and also centre.

We use CNN to train and validate a hundred (100) images. Firstly, we train and validate our dataset as ordinary as they are. Secondly, we apply image enhancement processing technique. For the ordinary image, our model did not improve from 69%. On the other hand, validation accuracy with image enhancement resulted to be 72%. The difference in the accuracy of result between the original and enhanced dataset is simply 3%. With the image brightness enhancement technique, we achieved a higher rate of gaze prediction accuracy. Hence, we have seen that image enhancement has proved its purpose by providing image interpretation with better quality.

Keywords: Gaze detection; gaze direction; individual eyes; image enhancement; deep learning

# INDIVIDUAL EYE GAZE PREDICTION WITH THE EFFECT OF IMAGE ENHANCEMENT USING DEEP NEURAL NETWORKS

# A THESIS SUBMITTED TO THE GRADUATE SCHOOL OF APPLIED SCIENCES OF NEAR EAST UNIVERSITY

# By OLUWASEUN PRISCILLA OLAWALE

In Partial Fulfillment of the Requirements for the Degree of Master of Science in Software Engineering

NICOSIA, 2020

# INDIVIDUAL EYE GAZE PREDICTION WITH THE EFFECT OF IMAGE ENHANCEMENT USING DEEP NEURAL NETWORKS

# A THESIS SUBMITTED TO THE GRADUATE SCHOOL OF APPLIED SCIENCES OF NEAR EAST UNIVERSITY

# By OLUWASEUN PRISCILLA OLAWALE

# In Partial Fulfillment of the Requirements for the Degree of Master of Science in Software Engineering

NICOSIA, 2020

#### Oluwaseun Priscilla OLAWALE: INDIVIDUAL EYE GAZE PREDICTION WITH THE EFFECT OF IMAGE ENHANCEMENT USING DEEP NEURAL NETWORKS

**Approval of Director of Graduate School of Applied Sciences** 

Prof. Dr. Nadire ÇAVUŞ

## We certify this thesis is satisfactory for the award of the degree of Master of Science in Software Engineering

**Examining Committee in Charge:** 

Assoc. Prof. Dr. Yoney KIRSAL EVER Committee Chairman, Software Engineering Department, NEU

Assist. Prof. Dr. Boran ŞEKEROĞLU Commitee Member, Information Systems Engineering Department, NEU

Assoc. Prof. Dr. Kamil DİMİLİLER Supervisor, Electrical & Electronics Engineering Department, NEU

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Oluwaseun Priscilla Olawale

Signature:

Date:

To my parents...

#### **AKNOWLEDGEMENTS**

I acknowledge Alewi Zurishaddai, Olutoju Ileri for his complete help and support.

With respect and a heart of gratitude, I thank my Supervisor Assoc. Prof. Dr. Kamil Dimililer, the chairman of the Software Engineering Department Assoc. Prof. Dr. Yoney Kirsal-Ever and my Advisor, Assisc. Prof. Dr. Boran Sekeroglu. I acknowledge you all for helping me to secure the height I am able to attain now.

I would not have come this far without my awesome parents. Thank you for everything, your labour of love will not be forgotten. I love you guys for ever.

Also, to my wonderful siblings, the love of my life at this time, and friends, I appreciate every single one of you for your love, care and support.

#### ABSTRACT

The prediction of individual eye gaze is a research topic that has gained the interest of researchers with its wide range of applications because neural networks majorly increase the rate of accuracy of individual gaze. In our research work, we have been able to predict individual gaze using MPIIGaze dataset. We categorize the gaze prediction into four (4) directions as to whether an individual is looking downwards, left and right directions and also centre.

We use CNN to train and validate a hundred (100) images. Firstly, we train and validate our dataset as ordinary as they are. Secondly, we apply image enhancement processing technique. For the ordinary image, our model did not improve from 69%. On the other hand, validation accuracy with image enhancement resulted to be 72%. The difference in the accuracy of result between the original and enhanced dataset is simply 3%. With the image brightness enhancement technique, we achieved a higher rate of gaze prediction accuracy. Hence, we have seen that image enhancement has proved its purpose by providing image interpretation with better quality.

Keywords: Gaze detection; gaze direction; individual eyes; image enhancement; deep learning

## ÖZET

Bireysel göz bakışının tahmini, geniş uygulama yelpazesi ile araştırmacıların ilgisini çeken bir araştırma konusudur, çünkü sinir ağları bireysel bakışların doğruluk oranını büyük ölçüde artırmaktadır. Araştırma çalışmamızda MPIIGaze veri kümesini kullanarak bireysel bakışları tahmin

etme işlemini gerçekleştirdik. Bakış tahminini, bir bireyin aşağı, sol ve sağ yönlere bakıp bakmadığına ve merkeze bakıp bakmadığına dair dört (4) yöne ayırdık ve CNN'i yüz (100) görüntüyü eğitmek ve doğrulamak için kullandık. İlk olarak, veri setimizde herhangi bir işlem gerçekleştirmeden eğittik ve doğruladık. İkinci olarak, görüntü işleme tekniğini uygulayarak öğretme işlemini tamamladık. Görüntü işleme tekniğini kullanmadığımız deneyimizde modelimiz %69 tanıma oranında kalırken, görüntü işleme tekniği kullanılarak yapılan deneyimizde %72'lik bir başarı oranına ulaşıldı. Orijinal ve geliştirilmiş veri kümesi arasındaki sonuç doğruluğundaki fark sadece %3'tür. Görüntü parlaklığı artırma tekniği ile daha yüksek bir bakış oranı hassasiyeti elde ettik. Bu nedenle, görüntü işleme tekniklerinin daha verimli veri hazırlığı ve görüntü yorumlama sağlayarak amacını kanıtladığını gördük.

Anahtar Kelimeler: Bakış tespiti; bakış yönü; bireysel gözler; görüntü geliştirme; derin öğrenme

# TABLE OF CONTENTS

AKNOWLEDGEMENTS	ii
ABSTRACT	iii
ÖZET	iv
TABLE OF CONTENTS	v
LIST OF FIGURES	viii
LIST OF ABBREVIATIONS	X

# **CHAPTER 1: INTRODUCTION**

1.1 Background of Study	1
1.2 Problem Statement	2
1.3 Aims and Objectives	2
1.4 Scope of Study	2
1.5 Methodology	2
1.6 Expected Result	3

# **CHAPTER 2: LITERATURE REVIEW**

2.1 The Human Visual System	4
2.2 Eye Tracking	5
2.2.1 Fixations	5
2.2.2 Saccades	5
2.2.3 Scanpath	6
2.2.4 Gaze duration	6
2.3 How Does Gaze Tracking work?	6
2.4 Deep Learning	7
2.4.1 Supervised learning	8
2.4.2 Unsupervised learning	9

2.4.3 Semi-supervised learning	9
2.4.4 Reinforced learning	10
2.5 Image Enhancement	11
2.6 Related Works	16

# **CHAPTER 3: DEEP NEURAL NETWORKS**

3.1 Comparison of Deep Learning Over Machine Learning	17
3.2 Components of a Neural Network	17
3.3 Deep Learning Architecture	18

# **CHAPTER 4: METHODOLOGY**

4.1 Image Enhancement Flowchart	20
4.2 Conceptual Model	21
4.3 Tools Used	22
4.3.1 Dataset	22
4.3.2 Python programming language	24
4.3.3 Image preprocessing with image enhancement using brightness	25
4.4 System Specification Requirements	26
4.4.2 Functional requirements	26

# **CHAPTER 5: RESULTS AND DISCUSSIONS**

5.1 Results	28
5.1.1 Random plots of images used	28
5.1.2 Model summary	29
5.2 Discussions	30
5.2.1 Original results without the effect of image enhancement	30
5.2.2 Original results with the effect of image enhancement	32

5.3 Challenges	34
<b>CHAPTER 6: CONCLUSIONS AND FUTURE WORK</b>	
6.1 Conclusions	35
6.2 Future Work	35

REFERENCES
------------

PPENDICES	
APPENDIX 1	43
IMAGE ENHANCEMENT CODE IN PYTHON PROGRAMMING LANGU	JAGE43
APPENDIX 2	44
CNN MODEL CODE IN PYTHON PROGRAMMING LANGUAGE	44

# LIST OF FIGURES

Figure 2.1: A cross section of the human eye	4
Figure 2.2: How gaze works with eye trackers	. 7
Figure 2.3: Machine learning techniques	8
Figure 2.4: Supervised learning model	8
Figure 2.5: Unsupervised learning model	9
Figure 2.6: Semi-supervised learning model	10
Figure 2.7: Reinforcement learning model	10
Figure 2.8: Image restoration	12
Figure 2.9: Image Enhancement	12
Figure 2.10: Image recognition	13
Figure 2.11: Image segmentation	13
Figure 2.12: Image resizing	14
Figure 2.13: Image compression	14
Figure 2.14: Image processing activity with Machine Learning/ Deep Learning/	
Neural Networks	15
Figure 3.1: A basic neural network	18
Figure 3.2: Deep learning architecture	18
Figure 4.1: Basic flowchart for image enhancement	20
Figure 4.2: CNN Model	21
Figure 4.2: Sample center view image	22
Figure 4.4: Sample down view image	23
Figure 4.5: Sample left view image	23
Figure 4.6: Sample right view image	24

Figure 4.7: Sample original image and enhanced image without visual aid	25
Figure 4.8: Sample original image and enhanced image with visual aid	25
Figure 5.1: Image Dataset	28
Figure 5.2: CNN model summary	29
Figure 5.3: Right view gaze prediction	30
Figure 5.4: Evolution of loss and accuracy of original dataset	31
Figure 5.5: Validation prediction accuracy	31
Figure 5.6: Evolution of loss and accuracy of enhanced dataset	32
Figure 5.7: Validation prediction accuracy with image enhancement technique	33

#### LIST OF ABBREVIATIONS

- **API:** Application Programming Interface
- **BioID:** Biometric Identity
- CIA: Central Intelligence Agency
- **CT:** Computed Tomography
- **CNN:** Convolutional Neural Networks
- **DCT:** Discrete Cosine Transform
- **DNN:** Deep Neural Networks
- **FBI:** Federal Bureau of Investigation
- FERET: Face Recognition Technology
- **IMM:** Informatics and Mathematical Modeling
- **IDE:** Integrated Development Environment
- MRI: Magnetic Resonance Imaging
- **PIL:** Python Image Library

# CHAPTER 1 INTRODUCTION

#### 1.1 Background of Study

The prediction of individual eye gaze is a research topic that has gained the interest of researchers with its wide range of applications because neural networks majorly increase the rate of accuracy of individual gaze. Our eyes play a very vital role when it comes to focusing on something. The eyes easily respond to anything viewable (Parikh & Kalva, 2018) and an individual's attention generally depends on the direction in which the individual is looking at. Eyesight which is obviously an old means of communication (Mohamed, Silva, & Courboulay, 2007) is important for individuals to communicate with machines or computers (Barbuceanu & Antonya, 2009) as it collects input data for the human brain to process activities.

According to (Recasens, Khosla, Vondrick, & Torralba 2015), individuals are capable of paying attention to each other's gaze so as to realize the focus of that particular individual. This act can be termed 'gaze following'. Gaze following permits us as humans to interpret the thoughts of other individuals, their current and future activity. So, we can say that gaze is useful for interpreting emotions, focus and interactions of individuals (Xiong, Kim, & Singh, 2019).

Gaze are important in media applications (Feng, Cheung, Tan, Callet, & Ji, 2013), business (Wąsikowska, 2014), robotics and computer human interaction (Recasens, Khosla, Vondrick, & Torralba, 2015), 3D gaming (Koulieris, Drettakis, Cunningham, & Mania, 2016), medicine (Harezlak & Kasprowski, 2017), virtual reality (Wang, Woods, Costela, & Luo, 2017) marketing and psychology (Stember, et al., 2019), smart home and mobile device authentication. Aside the applications mentioned above, we can apply gaze prediction in:

- *i. Security:* To monitor individual abnormal behaviour in situations where an individual intends to cause chaos in public areas such as airports, hospitals, schools, highways, malls, etc.
- *ii. Interrogation:* To determine the accuracy of statements made by suspected individuals in custody of the police, FBI or CIA.
- *iii. Human behaviour:* To further predict individual emotion, thought and / or future activity and even health conditions of individuals in hospitals.

For gaze research, image pre-processing activities are usually carried out together with machine learning or neural networks (Dimililer, et al., 2018) which are categorized as deep learning. Image pre-processing activities such as image enhancement aims at producing better results when implemented in any form of computer related research. Gaze estimation methods usually regress gaze directions directly from a single face or eye images.

#### **1.2 Problem Statement**

So many activities go on in our everyday life as humans and most of them are determined by our gaze. But is it actually possible to predict our gaze? i.e. the gaze of individuals. It might seem difficult to predict the gaze direction of individuals because as humans, we have the choice to view just anything at our own discretion. So, we propose an enhanced individual eye gaze prediction using neural networks.

## **1.3 Aims and Objectives**

- To predict direction in which an individual is focusing on.
- To determine the accuracy of individual eye gaze, with original image.
- To determine the accuracy of individual eye gaze, with enhanced version of the original image.
- To compare the accuracy of the original gaze image with enhanced gaze image.

## 1.4 Scope of Study

In this study, we will review research work that relates to gaze and its applications, obtain dataset for the purpose of our experiment after which the results of the experiment will be analysed. Only the prediction of individuals' gaze will be done by apply image enhancement technique.

## 1.5 Methodology

The dataset that will be obtained will be trained and tested using python programming language. After image acquisition, we shall carry out image pre-processing with image

enhancement technique and finally use deep neural network to predict individual eye gaze direction. The purpose of image enhancement is to improve the quality of our image dataset.

# **1.6 Expected Result**

It is expected that at the end of this project, we would have come up with a basic model, using convolutional neural network to:

- predict the gaze direction of individuals.
- determine the accuracy of individual eye gaze, with original image.
- determine the accuracy of individual eye gaze, with enhanced version of the original image.
- compare the accuracy of the original gaze image with enhanced gaze image.

# CHAPTER 2 LITERATURE REVIEW

## 2.1 The Human Visual System

The human eye is a vital organ that is very important for sight. It interprets visual data from our physical environment into a particular and precise image. The image that the human eye produces is not a direct graphic image. All of what we see is interpreted by our brain. So, we can say that an individual's eye feeds the brain with input which in turn processes by sending it to image processing points in the brain (Manrow, 2019) that is further translated to the image that we see (Idrees, 2015).



Figure 2.1: A cross section of the human eye (Braille Institute, 2018)

Usually, with the aid of light, the retina creates the images that we see from our physical environment (LaValle, 2019). The light which enters the eye is controlled by the iris (Singh & Singh, 2012) which acts as a muscle (Manrow, 2019) and converted into electrical impulses by the retina which is situated at the back of the eye (Idrees, 2015), as shown in Figure 2.1 above. The retina holds photoreceptors that helps transform light from the physical world into neural pulses. The fovea which sits on the centre of the retina (Manrow, 2019) is loaded with photoreceptors which makes it really sensitive to colour (Majaranta & Bulling, 2014).

#### 2.2 Eye Tracking

Eye tracking, also known as point of gaze (Farnsworth, 2019) or oculography (Borys & Plechawska-Wójcik, 2017) refers to the process of tracing individual eye movement so as to determine the gaze direction of the same individual (Singh & Singh, 2012). Eye tracking or gaze usually may include individual gaze direction prediction as seen in the research work of Dimililer, et al., (2018) or prediction of an individual's point of gaze. Eye tracking is important because certain unique information can be obtained from individual facial characteristics (Yua, et al., 2018). Systems developed for the purpose of eye tracking are composed of software programs that detect pupil, carry out image processing and data filtering operations, and finally ensure recording of individual eye transitions (Lupu & Ungureanu, 2013).

According to (Kar & Corcoran 2017), eye movement being studied in the applications and research of eye gaze can be categorised into the following:

#### 2.2.1 Fixations

Fixation may refer to the point of attention of indivual gaze. We can say that it is a measure of an individual's optical attention. Krauzlis, Goffart, and Hafed, (2017) states that individuals are usually in full control of the position of their eye movemment during fixation. The gaze fixation of individuals usually continues for about 100 milliseconds to 1000 milliseconds, leaving most gaze fixation from about 200 milliseconds to 500 milliseconds (Singh & Singh, 2012).

#### 2.2.2 Saccades

Saccades are brief eye movements that are quick. They can last up to tens of milliseconds, measuring peak speeds of hundreds of degrees per second (Krauzlis, Goffart, &

Hafed, 2017). Saccades are necessary in order for individuals to correctly recognize any visual content (Majaranta & Bulling, 2014).

#### 2.2.3 Scanpath

According to (Drusch, Bastien, & Paris2014), scanpath comprises of a finit number of both fixation points that are linked by saccades. Scanpaths are important to help understand the optical characteristics of individuals.

#### 2.2.4 Gaze duration

This is the total number of gaze fixation all summed up. Studies reveals that gaze duration usually takes long for long words when individuals read or study (Hohenstein, 2013).

Gaze detection and prediction has been researched on by several researchers and has been applied in computer vision, the area of human computer interaction (Dimililer, et al., 2018) and many other fields such as medicine, military, education, security, government, etc. Mohamed, Silva, and Courboulay (2007) states that its application has led to the subsistence of many ways to detect and track the direction of individual eye gaze. Eye tracking serves as the root for observing targets (Majaranta & Bulling, 2014). According to Singh and Singh (2012), there has been a rapid increase in the utilization of eye tracking systems due to the fall in price of eye tracking systems. There is a large market for researchers using eye tracking systems for research purpose, especially in human computer interaction.

Eye gaze research comes with difficulties such as deciding the gaze fixation and determining the point of focus of individual gaze (Dimililer, et al., 2018). The main purpose of gaze tracking is to obtain useful information from the gaze of individuals (Farnsworth, 2019) in order to determine the action of the individuals.

#### 2.3 How Does Gaze Tracking work?

Usually, with eye trackers, infrared light is reflected in an individual's eyes which are picked up by the camera of the eye tracker. The eye tracker now determines the direction of the individual's gaze through calculations and filtering (How Eye Tracking Works, n.d.). Figure 2.2 below describes the process of gaze tracking.



Figure 2.2: How gaze works with eye trackers (Tobii Pro, n.d)

#### 2.4 Deep Learning

Although they are related, machine learning together with deep learning and neural networks are all sub unit areas of artificial intelligence.

Machine learning allows computers or machines to function in such a way, performing some specific kinds of assignment with the aid of intelligent software (Mohammed, Khan, & Bashier, 2016) thereby mimicking the intelligence of higher animals (humans). Machines with this software are able to sort out compound patterns and carefully determine actions based on data. Machine learning is employed in applications such as audio recognition (Ogidan, 2017), medical imaging with neural networks (Dimililer, 2013), etc. The different machine learning techniques are described in figure 2.3 below:



Figure 2.3: Machine learning techniques (Mohammed, Khan, & Bashier, 2016).

#### 2.4.1 Supervised learning

Supervised learning is basically concerned with all forms of labeled data. Hence, its algorithms can easily be applied to problems related classification of patterns and regression of data. It is used to discover mapping rules for the prediction of outputs with unfamiliar inputs (Wang & Sng, 2015). Labeled data requires an agent/supervisor to be present during learning process.



Figure 2.4: Supervised learning model (i2tutorials, 2019)

#### 2.4.2 Unsupervised learning

Mishra and Saroha (2016) mentioned that unlike supervised learning, unsupervised learning is applicable when unlabeled data are used. In other words, it is useful where labeled data are not required. This learning algorithm deals with a form of network learning that gives the right output without the help without any interaction from an external behavior. In unsupervised learning, the unlabeled data requires no agent/supervisor to be present during learning process.



Figure 2.5: Unsupervised learning model (Ramesh, 2018)

#### 2.4.3 Semi-supervised learning

Semi-supervised learning is a comprises both unsupervised and supervised learning, i.e. it requires both unlabeled and labeled data. According to (Reddy, Pulabaigari, & Eswara 2018), semi-supervised learning is usually used when labeled data are hard to obtain. The Figure below gives the description of unsupervised learning.



Figure 2.6: Semi-supervised learning model (Moltzau, 2019)

## 2.4.4 Reinforced learning

An agent is usually present and its learning process is an interactive one, usually between the agent present and the environment (Campos, 2018). Reinforcement learning is based on trial and error learning process (Sathya & Abraham, 2013). Figure 2.7 below describes the reinforcement learning model.



Figure 2.7: Reinforcement learning model (Wagner, 2018)

Deep learning which is also referred to as hierarchical learning is a section of machine learning algorithms (Hordri, Samar, Yuhaniz, & Shamsuddin, 2017). When compared to machine learning, deep learning has the ability to exhibit high rate of intelligence. Although, they do not fully comprehend narratives or long information like expert systems, they are excellent in the way (Bhatia & Rana, 2015).

According (Bhatia & Rana, 2015) perceptron algorithm is regarded as the first machine to depict human intelligence but it was limited in its ability to learn. This brought neural networks which is in vogue today. Neural networks have been applied in nearly all fields to obtain so many results in all facets of life. They are useful for grouping unlabeled data by similarities in the set of given inputs. Neural networks are designed to function just like the human brain. They make use of some machine perception to interpret sensory data.

#### 2.5 Image Enhancement

Like machine learning and neural networks, image processing techniques are also applied in the area of computer vision, optical character recognition, biometric verification, remote sensing, face detection and digital video processing (Dewangan, 2016) and medical imaging such as MRI (Dimililer & Ihan, ICAFS, 2016), X-ray images (Dimililer, Backpropagation Neural Network Implementation for Medical Image Compression, 2013), CT and medical palmistry (Dewangan, 2016).

The first step for any image processing activity is usually image acquisition, image preprocessing, application of image processing technique and finally result. The image processing technique to be applied may be:

#### i. Image restoration



Original image with noise

Restored image

Figure 2.8: Image restoration (Color Experts, 2013)

This technique is applied when an image contains noise or is blur. It is used to restore a corrupt image back to its original form by removing the blur or noise contained in the image (Rani, Jindal, & Kaur, 2016), as described in figure 2.8 above.



## *ii.* Image enhancement

Original image

Restored image

Figure 2.9: Image Enhancement (Robert, 2018)

Although it is usually regarded as image pre-processing technique, it is applied in improving the details in and quality of an image, as shown in figure 2.9 above. Image enhancement is briefly explained, just before section 2.6.

## iii. Image recognition

This procedure is useful for defining objects (including human, actions and locations) in an image (Gupta, 2018), as shown in figure 2.10 below.



Figure 2.10: Image recognition (Azati Software, 2019)

iv. Image segmentation



Original image

nageSegmented imageFigure 2.11: Image segmentation (Kumar, 2019)

Image segmentation is the act of dividing specific areas in an image with like features, that are useful for information analysis.

# v. Image resizing

This technique is applied when there is need to format the dimension of an image either by increasing or decreasing the number of pixels in the image.



Figure 2.12: Image resizing (Sajjad, et al., 2017)

#### vi. Image compression



Original Lena Image (12KB size)



Lena Image, Compressed (85% less information, 1.8KB)



Lena Image, Highly Compressed (96% less information, 0.56KB)

Figure 2.13: Image compression (DBA Blog, 2015)

We usually apply this technique when there is need to reduce the amount of storage data, in bits for an image.

Each of the image processing techniques mentioned above also have special methods by which they can be used to achieve the necessary results.

Image processing techniques in most cases are used with machine learning/deep learning/neural networks. The machine learning/deep learning/neural network used may be for the purpose of either prediction or classification, as the case may be. Image processing generally provide aids to:

- i. Make visible objects that are invincible.
- ii. Obtain an improved quality of the original image.
- iii. Study certain features in an image.
- iv. Measure the pattern of different objects in an image.
- v. Distinguish between objects and features in an image.

The Figure below illustrates how machine learning/deep learning/neural networks can be combined with image processing techniques.



# Figure 2.14: Image processing activity with Machine Learning/ Deep Learning/ Neural Networks

The term "image enhancement" may be described as a method that involves the manipulation of pixels from an image to achieve a clearer interpretation from that image or carry out other image processing techniques (Hussain & Lone, 2018) to gather useful information. Although image enhancement task poses to be a challenging task (Shukla, Potnis, & Dwivedy, 2017), its main purpose in image processing is to obtain an image that is far better than its original, in terms of quality (Bhardwaj, Kaur, & Singh, 2018).

Although image enhancement techniques are categorised into two; spatial domain and frequency domain, an image may be enhanced using one or more of the following techniques:

- i. Brightness
- ii. Contrast
- iii. Colour
- iv. Sharpness

For an image to be visually interpreted, at least one of these techniques may be applied. The above mentioned image enhancement techniques either manipulate the pixels in an image directly or Fourier transform of the image (Kaur & Taqdir, 2016). Some individuals

interchange brightness and contrast, but they are not the same. The brightness in an image refers to the ratio mix of lightness and darkness it contains while contrast refers to the difference value of the image from its colour and brightness. For most imaging tasks, image enhancement techniques are the most popular techniques employed for pre-processing images (Singh, Seth, Sandhu, & Samdani, 2019).

#### 2.6 Related Works

Yua, et al. (2018) presented an article in which they combined both CNN and support vector machines as a model to detect individual eyes. The dataset used included face database from IMM, FERET, ORL and BioID. With this model, they obtained a better detection accuracy, even with different eye defects.

Park, Spurr, and Hilliges (2018), organised a one of a kind research in which deep neural networks was used to estimate individual gaze based on one eye. This, they referred to as gazemaps with 3D point of gaze estimation. With dataset obtained from MPIIGaze, it was discovered that the estimation of the point of gaze is inflenced by the use of both individual eyes and the resolution of the image. The results obtained were interpreted to be extremely accurate.

DCT was applied by (Dimililer, et al., 2018) to detect and predict the direction of individual eye gaze. Image compression is an image pre-processing technique that they employed using back propagation neural networks. These techniques were used to categorize the gaze dataset into right, left and centre as gaze directions.

Jaques, Conati, Harley, and Azevedo (2014), carried out a research to using feature selection and machine learning to forecast the emotions of students with eye tracking as regards learning curiosity and enthuaism. The data that was used was gathered through an Intelligigent Tutoring System (ITS) called MetaTutor.

Jerry, Lam, and Eizenman, (2008) used the CNN algorithm for detection of eyes in gaze estimation systems. Their work did not include any form of image pre-processing. Three participants were used and it was discovered that for head movements, the CNN was able to detect individual eyes.

# CHAPTER 3 DEEP NEURAL NETWORKS

We briefly described deep learning in chapter 2 but in this chapter, we will focus more on it, stating and explaining some its algorithms.

In section 2.4, we stated that both machine learning and deep learning are both subsets of artificial intelligence. Deep neural networks may also be categorized as a type (Shaikh, 2017) or higher version of machine learning, because when used in an application they tend to achieve better results than the traditional machine learning. DNNs simply comprise of algorithms with neural network, forming deep layers of learning patterns. This is how we derive the term "deep learning".

## 3.1 Comparison of Deep Learning Over Machine Learning

- i. Deep learning depends on high level machines, unlike machine learning which can function on low level machines.
- ii. Machine learning cannot perform well to obtain high accuracy of results if the data is extremely large. Deep learning on the other hand will obtain high accuracy result with extremely large data but the reverse is the case with small data.
- iii. Deep learning can go as far as learning optimum pattern features of data, this is unlike machine learning.
- iv. Deep learning algorithms take really long time to execute training. Training might even go on for a few weeks. While machine learning takes less time to execute.

Deep neural networks may also be referred to an Artificial neural networks, with several different layers (Albawi, Mohammed, & Alzawi, 2017).

#### **3.2** Components of a Neural Network

A neural network is simply composed of an input layer, multiple hidden layers and an output layer, as shown in figure 3.1 below.



Figure 3.1: A basic neural network (WikiStat, n.d)

Like machine learning, neural networks are designed to function like the human brain. In a basic neural network, each layer includes a set of nodes, where computation is carried out. Each layer serves as input for the next layer until the final (output) layer is reached. In this process where the neural network is learning patterns, weights are generated, computed and neurons are activated with some activation function. The more the layers involved in neural networks, the deeper the rate of accuracy. Through this, they can accept, process and interpret sensory data. They are able to perform the following tasks:

- i. Classification
- ii. Prediction
- iii. Clustering

## **3.3 Deep Learning Architecture**

Like machine learning, some of the algorithms of deep neural networks can be categorised



Figure 3.2: Deep learning architecture (Simplilearn, 2019)

into deep supervised, classification for example and deep unsupervised learning, clustering for example (Nicholson, n.d), as shown in figure 3.2 above. Deep learning may either use labelled or unlabelled data for both training and validation. Although they depend on neural networks that are modeeled after the human brain, they are capable of self learing at some point. During the training phase, the algorithms extract and learn patterns of features from the data used, regroup them and learn some more patterns, before obtaining required results based on the kind of model used i.e. classification, prediction or clustering.

According to (Simplilearn, 2019), we can employ deep learning in fields such as fraud detection, audio and speech recognition, medical imaging, business management, computer vision, security surveillance, bioinformatics, etc. Some of the deep neural networks algorithm that exist include:

#### *i.* The Multilayer Perceptron

A perceptron is simply a one neural model and "the first" in the series of neural network algorithms (Brownlee, 2016). This algorithm is used to train and classify non linearly separable problems, thereby solving difficult computational operations and may be applied in machine translation, image verification, data classification and e-commerce.

#### ii. Recurrent Neural Networks

This algorithm is usually employed in Natural Language Processing (Britz, 2015). It is called RNN because for each neuron (node) in each layer that serves as input for the next layer, it repeats the same computational operation sequentially.

#### iii. Convolutional Neural Networks

CNN is a type of supervised learning algorithm with multilayer perceptrons performing feed-forward operation. When working with images, CNN model is the best algorithm to select. We later discus its components in section 4.2, since it is the algorithm that we adopt for the purpose of our research.

#### iv. Recursive Neural Networks

Recurrent neural networks are unlike recurrent neural networks and are usually applied on structured data input.

# CHAPTER 4 METHODOLOGY

Firstly, after we acquire the selected images from our dataset, we performed image preprocessing with image enhancement technique by applying brightness and finally use CNN which is a type of deep learning algorithm to predict individual eye gaze direction.

We divided the program execution into two parts. The first part of the program execution trains and validates our image dataset without the application of image enhancement. While image enhancement was applied in the later. Figure 4.2 describes the conceptual model of both parts of the program execution

#### **4.1 Image Enhancement Flowchart**

First of all, we input each image into our program, after which we apply a brightness of 1.5<sub>units</sub> and further save the image if the process is successful.



Figure 4.1: Basic flowchart for image enhancement

Figure 4.1 above describes the workflow of how image enhancement is applied. This is to enable us further our research with this image processing technique. Python programming language enables us to enhance images with the four (4) different techniques as mentioned in

section 2.5 above. We choose brightness for enhancing our dataset for better pictorial interpretation. Brightness is usually applied by decreasing or increasing the matrix of the image either by subtraction or addition.

#### 4.2 Conceptual Model

For each of our experimental sections, i.e. ordinary image dataset and enhanced version of our image data set, we make use of a basic CNN model with three (3) different layers of convolutions. For each layer, we carry out max pooling and connect the last layer to a fully connected layer before we finally get our output.





CNNs are made up of layers such as:

## i. Convolutional layers

This is the initial set of layers in a CNN model. Here, the convolutional layers apply filtering to summarize the pixels in an image. The main function of the convolutional layer is to reveal distinct visual features such as colour drops, lines, edges, etc. By doing this, the CNN learns specific characteristics, hierarchies of several patterns in an image.
## *ii.* Pooling layers

There is a pooling layer after each convolutional layer. Two techniques that may be used for pooling are the average pooling or maximum pooling.

## *iii.* Fully connected layers

These are the important layers. It does the actual leaning in the deep neural network (CNN in this case). It comprises of several perceptron layers and identifies the object in a class.

For our CNN model, we use 3 convolutional layers and for each layer, we carry out pooling operations and connect to one full layer before the actual prediction is done.

## 4.3 Tools Used

In this section, we give brief descriptions of some of the tools we use in accomplishing the results of our research.

## 4.3.1 Dataset

Our dataset consists of 100 images randomly obtained from MPIIGaze dataset. These images are divided into four (4) categories for our work, with 25 images in each category depicting:

*i. Center view:* We categorize this view as a fixed point for an individual's horizontal gaze direction in any scene as shown in the figure below.



Figure 4.3: Sample center view image

*ii. Down view:* We categorize images with individuals looking downwards in any scene, as described in the figure below.



Figure 4.4: Sample down view image

*iii. Left view:* Images in this category contain scenes of individuals facing the left side view, as described in figure 4.5 below.



Figure 4.5: Sample left view image

#### iv. Right view



Figure 4.6: Sample right view image

Images in this category contain scenes of individuals facing the right side view, as described in figure 4.6 above.

#### **4.3.2** Python programming language

We choose python programming language because we discovered that it is most useful when implementing deep learning and machine learning algorithms. It comprises of several libraries/modules and framework, some of which are briefly described below.

## i. Keras

Keras is a deep learning library enabled with high level neural networks API for generating easy and quick models. It supports CNN and recurrent networks.

### *ii.* Tensorflow framework

Tensorflow is a framework that that represents numerical computation with dataflow graphs. It is used for research experimentation on data that is to be trained and validated or tested with machine learning and deep learning algorithms. We can say that it serves as a platform for training neural networks. It requires a specific line of code to be installed with command prompt or python terminal. There are also IDEs for ease of its installation and the installation of other libraried required for projects executed in python programming language. One of such IDEs, which we made use of is the JetBrains Pycharm Community Editor.

## *iii.* TensorBoard module

The 'TensorBoard' module constitutes a set of applications that enables users to view tensorflow graphs.

## iv. PIL module

The PIL enables users of python programming language to perform image processing techniques. We employed it by using it to enhance our image dataset.

## 4.3.3 Image preprocessing with image enhancement using brightness

For preprocessing our dataset, we make use of the brightness image enhancement technique with just a few lines of code. As we mentioned earlier in section 2.5 above, the main purpose of employing the image enhancement technique with brightness is to improve on the quality of our dataset to achieve a higher accuracy of gaze prediction. The following Figures below gives a pictorial view of how the original image and enhanced image.



Original image



Enhanced image

Figure 4.4: Sample original image and enhanced image without visual aid





Original image

Enhanced image

Figure 4.5: Sample original image and enhanced image with visual aids

Usually, we encounter individual faces with or without spectacles or glasses as we know them to be. From the above shown figures, i.e. figure 4.4 and 4.5 we have applied image enhancement on these two individual faces. The use of spectacles does not affect the application of image processing techniques or deep learning algorithms, so long as the eyes are detectable.

## 4.4 System Specification Requirements

In this section, we give a brief description of the expected behaviour and features of our software.

## 4.4.1 Non-functional requirements

## i. Hardware requirements

- A computer system
- Minimum of Pentium IV / AMD A8-7410 APU or higher CPU
- Minimum of 512MB of RAM
- On-camera Monitors

## ii. Software Requirements

- Windows 7, Windows 8, Windows 10 or any other Operating System that supports python.
- 32-bit / 64-bit Operating System
- Python 3.6
- Anaconda/miniconda
- PyCharm IDE

## **4.4.2 Functional requirements**

Since we are adopting a basic technique for our CNN model, our CNN model shall be able to perform the following functions:

- Our model shall be able to predict individual gaze, based on the directions that we defined, i.e. centre view, left view, right view and down view.
- Our model shall obtain accuracy of individual gaze prediction with ordinary image.

- Our model shall obtain accuracy of individual gaze prediction with the enhanced version of the original image.
- Our program shall be able to apply brightness technique of image enhancement.

# CHAPTER 5 RESULTS AND DISCUSSIONS

## 5.1 Results

With the basic CNN model employed and the brightness image technique used, we have been able to obtain some amount of reasonable results, ranging from the random selection of images used in our research to the task of obtaining all the accuracy required. In this section, we generally display the results that we obtained



#### 5.1.1 Random plots of images used

Figure 5.1: Image dataset

Figure 5.1 above depicts how the selected images look like, with individual gaze facing the four directions that we defined earlier (center view, down view, left view and right view). The images have a 48 x 48 dimension in size. We choose to display a random selection of 20 images in all gaze direction categories.

#### 4.1.2 Model summary

The figure below describes the whole parameters in each layer from our CNN model. The whole parameter sums up to 5,654,276 parameters, the total number of trainable parameters are 5,651,332 parameters and the total number of untrainable parameters are 2,944 parameters.

Layer (type) Output Shape Param #	max_pooling2d_3 (MaxPooling2 (None, 6, 6, 512) 0
conv2d_1 (Conv2D) (None, 48, 48, 64) 640	dropout_3 (Dropout) (None, 6, 6, 512) 0
batch_normalization_1 (Batch (None, 48, 48, 64) 256	flatten_1 (Flatten) (None, 18432) 0
activation_1 (Activation) (None, 48, 48, 64) 0	dense_1 (Dense) (None, 256) 4718848
max_pooling2d_1 (MaxPooling2 (None, 24, 24, 64) 0	batch_normalization_4 (Batch (None, 256) 1024
dropout_1 (Dropout) (None, 24, 24, 64) 0	activation_4 (Activation) (None, 256) 0
conv2d_2 (Conv2D) (None, 24, 24, 128) 204928	dropout_4 (Dropout) (None, 256) 0
batch_normalization_2 (Batch (None, 24, 24, 128) 512	dense_2 (Dense) (None, 512) 131584
activation_2 (Activation) (None, 24, 24, 128) 0	batch_normalization_5 (Batch (None, 512) 2048
max_pooling2d_2 (MaxPooling2 (None, 12, 12, 128) 0	activation_5 (Activation) (None, 512) 0
dropout_2 (Dropout) (None, 12, 12, 128) 0	dropout_5 (Dropout) (None, 512) 0
conv2d_3 (Conv2D) (None, 12, 12, 512) 590336	dense_3 (Dense) (None, 4) 2052
batch_normalization_3 (Batch (None, 12, 12, 512) 2048	Total params: 5,654,276
activation_3 (Activation) (None, 12, 12, 512) 0	Non-trainable params: 2,944

Figure 5.2: CNN model summary

For the purpose of our research, we have used three convolutional layers, each with max pooling of 2 factor and one fully connected layer. Table 5.2 describes the summary of our model, while Figure 5.3 describes the conceptual view of how our CNN model was implemented.

#### **5.2 Discussions**

In this section, we discuss the results we obtained both for the original image dataset and the enhanced version of the dataset.

For our first aim which is to predict gaze, we were able to achieve results as shown in figure 5.3 below.



Figure 5.3: Right view gaze prediction

With just a picture on the wall, our CNN model was able to predict the individual's eye gaze as a right view direction. Hence, we can say that our CNN model is now capable of predicting individual eye gaze. The following sections in this chapter reviews the accuracy results that we obtained.

#### 5.2.1 Original results without the effect of image enhancement

The model starts to stabilize at about 61% without the application of image enhancement as shown in Figure 5.4 below. Figure 5.5 describes the best model that we obtained with the four (4) different classes, giving the validation accuracy as 69%, since our model did not improve from 69%. This model is tested and validated without the aid of the image enhancement technique.

#### Optimizer : Adam



Figure 5.4: Evolution of loss and accuracy of original dataset

Here, we obtain results for both loss and accuracy of the training and validation (testing) phases. The difference of errors that we have obtained for our dataset is 0.68% while we obtain the difference in value of accuracy is 7%.

```
Epoch 00095: val_accuracy did not improve from 0.67000
Epoch 96/100
Epoch 00096: val_accuracy did not improve from 0.67000
Epoch 97/100
1/1 [------] - 39s 39s/step - loss: 0.8876 - accuracy: 0.5800 - val loss: 0.9049 - val accuracy: 0.6700
Epoch 00097: val_accuracy did not improve from 0.67000
Epoch 98/100
Epoch 00098: val_accuracy improved from 0.67000 to 0.69000, saving model to model_weights.h5
Epoch 99/100
Epoch 00099: val_accuracy did not improve from 0.69000
Epoch 100/100
Epoch 00100: val_accuracy did not improve from 0.69000
```

### Figure 5.5: Validation prediction accuracy

#### 5.2.2 Original results with the effect of image enhancement

The model starts to stabilize at 64% with the application of image enhancement as shown in Figure 5.6 below. Figure 5.7 describes the best model that we obtained with the four (4) different classes, giving the validation accuracy as 72%. This model is tested and validated with the aid of the image enhancement technique.



Figure 5.6: Evolution of loss and accuracy of enhanced dataset

Here, we obtain results for both loss and accuracy of the training and validation (testing) phases. The difference of errors that we have obtained for our dataset is 3.45% while we obtain the difference in value of accuracy is 7%.

Epoch 00096: val\_accuracy did not improve from 0.72000 Epoch 97/100

1/1 [==========] - 36s 36s/step - loss: 0.9575 - accuracy: 0.6200 - val\_loss: 0.9194 - val\_accuracy: 0.7000
Epoch 00098: val\_accuracy did not improve from 0.72000

Epoch 00098: Val\_accuracy did not improve from 0.7200 Epoch 99/100

1/1 [======] - 36s 36s/step - loss: 0.8757 - accuracy: 0.6400 - val\_loss: 0.9102 - val\_accuracy: 0.7100
Epoch 00100: val\_accuracy did not improve from 0.72000

Figure 5.7: Validation prediction accuracy with image enhancement technique

For both phases, i.e., original results with and without the effect of image enhancement, we used a total number of two hundred (200) epochs, with 100 epochs belonging to each phase. Each epoch represents a full training cycle for the image dataset. There is no limit to choosing the total number of epochs for training and validating, but the purpose of choosing and using one hundred (100) epochs is to ensure optimal learning of the image dataset since we randomly selected a total of one hundred (100) images. We included a batch size of 10 images in each batch for both training and validation.

We use pooling to reduce the volume of each feature but main the relevant information each convolutional layer. The graph results obtained in Figures 5.2 and 5.4 were plotted with training data saved by the model in '. json' format. Since the loss decreased and accuracy increased in both phases, it is proper to say that our model was successful in learning and not cramming.

The difference in the accuracy of result between the original and enhanced dataset is simply 3%. The more we trained, validated and saved our model, the higher the accuracy of gaze prediction we obtained. With the result that we obtained, it is revealed that with image enhancement, there is a higher rate of gaze prediction accuracy.

## 5.3 Challenges

We faced challenges in obtaining dataset. Original datasets are not easily accessible and these datasets vary, even for the required research domain.

A HP windows 8.1 was used to carry out the code execution of our research. Because of its specifications, the runtime was intensively slow. As a result, for the purpose of our research, we had to randomly select one hundred (100) images, with twenty-five images in each of the four categories that we previously defined.

## CHAPTER 6 CONCLUSIONS AND FUTURE WORK

#### **6.1 Conclusions**

We use a basic CNN model to train and validate a hundred (100) images so as to predict the gaze direction of individuals. This is majorly because we were restricted with the processing speed of our computer. As a result, we could not go further to compare our results with the work of other researchers as regards this area.

For the ordinary image, our model did not improve from 69%. On the other hand, validation accuracy with image enhancement resulted to be 72%. The difference in the accuracy of result between the original and enhanced dataset is simply 3%. The more we trained, validated and saved our model, the higher the accuracy of gaze prediction we obtained. With the result that we obtained, it is revealed that with image enhancement, there is a higher rate of gaze prediction accuracy.

Hence, we have been able to achieve our aims and objects; and finally, we can say that image enhancement has proved its purpose by providing image interpretation with better quality.

#### 6.2 Future Work

At this time, we cannot say that everything from neural networks to deep learning to machine learning and finally, artificial intelligence is all new. This is because they have existed for over a decade. They have been applied in virtually all areas, including gaze detection.

The evolution of all gaze technologies over time has yielded useful results for researchers in general, some of which have helped to predict the intentions and actions of individuals. However, there is more work to be done in terms of security with the aid of machine learning or deep learning and image processing techniques.

In our future work, we shall compare the use of other image processing techniques for gaze prediction with other deep learning algorithms.

#### REFERENCES

- Albawi, S., Mohammed, T. A., & Alzawi, S. (2017). Understanding of a Convolutional Neural Network. In The International Conference on Engineering and Technology. Antalya, Turkey: IEEE.
- Azati Software. (2019). Image Detection, Recognition, and Classification With Machine Learning. Retrieved from https://azati.ai/image-detection-recognition-andclassification-with-machine-learning/
- Barbuceanu, F., & Antonya, C. (2009). Eye Tracking Applications. *Bulletin of the Transilvania* University of Braşov, 2(51), 17-24.
- Bhardwaj, N., Kaur, G., & Singh, P. K. (2018). A Systematic Review on Image Enhancement Techniques. Sensors and Image Processing. Advances in Intelligent Systems and Computing, 651, 227-235. doi:https://doi.org/10.1007/978-981-10-6614-6\_23
- Bhatia, N., & Rana, C. (2015). Deep Learning Techniques and its Various Algorithms and Techniques. *International Journal of Engineering Innovation & Research*, 4(5).
- Borys, M., & Plechawska-Wójcik, M. (2017). Eye-Tracking Metrics in Perception and Visual Attention Research. *European Journal of Medical Technologies*, *3*(36), 11-23.
- Britz, D. (2015). Recurrent Neural Networks Tutorial, Part 1 Introduction to RNNs. Retrieved from http://www.wildml.com/2015/09/recurrent-neural-networks-tutorialpart-1-introduction-to-rnns/
- Braille Institute, (2018). *The Aging Eye*. Retrieved from https://www.brailleinstitute.org/event/the-aging-eye
- Brownlee, J. (2016). *Crash Course On Multi-Layer Perceptron Neural Networks*. Retrieved from https://machinelearningmastery.com/neural-networks-crash-course/
- Campos, L. (2018). Deep Reinforcement Trading. Retrieved from https://quantdare.com.
- Color Experts. (2013). *Purpose of Image Restoration*. Retrieved from https://www.colorexpertsbd.com/blog/image-restoration
- DBA Blog. (2015). *File Compression*. Retrieved from https://bayleegunnell.weebly.com/blog/file-compression

- Dewangan, S. K. (2016). Importance & Applications of Digital Image Processing. International Journal of Computer Science & Engineering Technology, 7(07), 316-320.
- Dimililer, K. (2013). Backpropagation Neural Network Implementation for Medical Image Compression. *Journal of Applied Mathematics*, doi:10.1155/2013/453098
- Dimililer, K., & Ihan, A. (2016). Effect of Image Enhancement on MRI Brain Images with Neural Networks. In 12th International Conference on Application of Fuzzy Systems and Soft Computing, (vol. 102, pp. 39-44). Vienna, Austria.
- Dimililer, K., Ever, Y. K., Somturk, C., Ergun, F., Urun, G., & Kara, M. (2018). Effect of DCT Image Compression on Eye Gaze Direction Detection. In International Conference on Applied Mathematics, Computational Science and Systems Engineering. doi:https://doi.org/10.1051/itmconf/20181601003
- Drusch, G., Bastien, J. C., & Paris, S. (2014). Analysing Eye-Tracking Data: From Scanpaths and Heatmaps to the Dynamic Visualisation of Areas of Interest. In W. K. T. Ahram (Ed.), *In Proceedings of the 5th International Conference on Applied Human Factors* and Ergonomics, (pp. 19-23). Kraków, Poland.
- Farnsworth, B. (2019). *What is Eye Tracking and How Does it Work?* Retrieved from iMotions.com: https://imotions.com/blog/eye-tracking-work/
- Feng, Y., Cheung, G., Tan, W.-t., Callet, P. L., & Ji, Y. (2013). Low-Cost Eye Gaze Prediction System for Interactive Networked Video Streaming., 15(8), 1865-1879. doi:10.1109/TMM.2013.2272918
- Gupta, S. (2018). *nderstanding Image Recognition and Its Uses*. Retrieved from https://www.einfochips.com
- Harezlak, K., & Kasprowski, P. (2017). Application of Eye Tracking in Medicine: A survey, Research Issues and Challenges. *Computerized Medical Imaging and Graphics*, 65, 176-190. doi:https://doi.org/10.1016/j.compmedimag.2017.04.006
- Hohenstein, S. (2013). Eye Movements and Processing of Semantic Information in the Parafovea During Reading (Thesis).
- Hordri, N. F., Samar, A., Yuhaniz, S., & Shamsuddin, S. M. (2017). A Systematic Literature Review on Features of Deep Learning in Big Data Analytics. *International Journal of Advances in Soft Computing and its Applications*, 9(1), 32-49.

- Hussain, S., & Lone, M. M. (2018). Image Enhancement Techniques: A Review. *International Research Journal of Engineering and Technology (IRJET)*, 5(9).
- i2tutorials. (2019). What are the differences between Supervised Machine Learning and Unsupervised Machine Learning? Retrieved from https://www.i2tutorials.com/top-machine-learning-interview-questions-and-answers/what-are-the-differences-between-supervised-machine-learning-and-unsupervised-machine-learning/
- Idrees, M. (2015). Fundamental Optics of The Human Eye and Aging Effects on Visual Acuity: An Overview. *International Journal of Preclinical & Pharmaceutical Research*, 6(1).
- Jaques, N., Conati, C., Harley, J. M., & Azevedo, R. (2014). Predicting Affect from Gaze Data during Interaction with an Intelligent Tutoring System. In International Conference on Intelligent Tutoring Systems (pp. 29-28). Springer, Cham. doi:https://doi.org/10.1007/978-3-319-07221-0\_4
- Jerry, Lam, C. L., & Eizenman, M. (March 2008). Convolutional Neural Networks for Eye Detection in Remote Gaze Estimation Systems. In Proceedings of the International MultiConference of Engineers and Computer Scientists, (vol. 1, pp. 19-21). Hong Kong.
- Kar, A., & Corcoran, P. (2017). A Review and Analysis of Eye-Gaze Estimation Systems, Algorithms and Performance Evaluation Methods in Consumer Platforms. *In IEEE Access (vol. 5)*, doi:10.1109/access.2017.2735633
- Kaur, R., & Taqdir. (2016). Image Enhancement Techniques- A Review. *International Research Journal of Engineering and Technology (IRJET)*, 3(3), 1308-1315.
- Koulieris, G. A., Drettakis, G., Cunningham, D., & Mania, K. (2016). Gaze Prediction Using Machine Learning for Dynamic Stereo Manipulation in Games. 2016 IEEE Virtual Reality (VR). Greenville, SC, USA: IEEE. doi:10.1109/VR.2016.7504694
- Krauzlis, R. J., Goffart, L., & Hafed, Z. M. (2017). Neuronal Control of Fixation and Fixational Eye Movements. *Philosophical Transactions of the Royal Society of London. Series B*, *Biological Sciences*, 372(1718). doi:10.1098/rstb.2016.0205
- Kumar, V. V. (2019). *Panoptic Segmentation with UPSNet*. Retrieved from https://towardsdatascience.com/panoptic-segmentation-with-upsnet-12ecd871b2a3

- LaValle, S. M. (2019). Virtual Reality. In S. M. LaValle, *Virtual Reality* (pp. 125-150). Oulu, Finland: Cambridge University Press.
- Lupu, R. G., & Ungureanu, F. (2013). Asurvey of eye tracking Methods and Applications. Buletinul Institutului Politehnic din iaşi , 59(63), 71-86. Gheorghe Asachi" din Iaşi : Universitatea Tehnică.
- Manrow, S. (2019). What is the difference between human vision and computer vision? Retrieved from https://www.quora.com.
- Mishra, P. K., & Saroha, G. (2016). A Study on Classification for Static and Moving Object in Video Surveillance System. *I.J. Image, Graphics and Signal Processing, 5*, 76-82.
- Mohamed, A. O., Silva, M. P., & Courboulay, V. (2007). *A History of Eye Gaze Tracking*. Retrieved from https://hal.archives-ouvertes.fr/
- Mohammed, M., Khan, M. B., & Bashier, E. B. (2016). *Machine Learning: Algorithms and Applications*. Boca Raton, CRC Press. doi: https://doi.org/10.1201/978131537165
- Moltzau, A. (2019). Advancements in Semi-Supervised Learning with Unsupervised Data Augmentation. Retrieved from https://towardsdatascience.com/advancements-in-semisupervised-learning-with-unsupervised-data-augmentation-fc1fc0be3182
- Ogidan, T. E. (2017). Software System For Audio Sample Recognition (Masters Thesis) Near East University, Nicosia.
- Parikh, S., & Kalva, H. (2018). Eye Gaze Feature Classification for Predicting Levels of Learning. *In AIED 2018 workshop proceedings*.
- Park, S., Spurr, A., & Hilliges, O. (2018). Deep Pictorial Gaze Estimation. Computer Vision and Pattern Recognition. doi: 10.1007/978-3-030-01261-8\_44
- Tobii Pro, (n.d). *How do Tobii Eye Trackers work?* Retrieved September 16, 2019, from https://www.tobiipro.com/learn-and-support/learn/eye-tracking-essentials/how-do-tobii-eye-trackers-work/
- Ramesh. (2018). *Introduction to Machine Learning*. Retrieved from https://www.fcodelabs.com/2018/12/13/Machine-Learning-Intro/

- Rani, S., Jindal, S., & Kaur, B. (2016). A Brief Review on Image Restoration Techniques. *International Journal of Computer Applications*, 150(12), 30-33. doi:10.5120/ijca2016911623
- Recasens, A., Khosla, A., Vondrick, C., & Torralba, A. (2015). Where are they looking? Advances in Neural Information Processing Systems. In Proceedings of the 28th International Conference on Neural Information Processing Systems (vol. 1, pp. 199-207).
- Reddy, Y. C., Pulabaigari, V., & Eswara, R. B. (2018). Semi-Supervised Learning: A Brief Review. International Journal of Engineering & Technology, 81- 85. doi:10.14419/ijet.v7i1.8.9977
- Robert, A. (2018). *Improving the Quality of Photo through Image Enhancement*. Retrieved from https://www.techgyd.com/improve-image-quality-by-imageenhancement/37428/
- Sajjad, M., Shah, A., Jan, Z., Shah, S. I., Baik, S. W., & Mehmood, I. (2017). Facial appearance and texture feature-based robust facial expression recognition framework for sentiment knowledge discovery. *Cluster Computing*, 21(5). doi:10.1007/s10586-017-0935-z
- Sathya, R., & Abraham, A. (2013). Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification. *International Journal of Advanced Research in Artificial Intelligence*, 2(2), 34-38.
- Shukla, K. N., Potnis, A., & Dwivedy, P. (2017). A Review on Image Enhancement Techniques. International Journal of Engineering and Applied Computer Science (IJEACS), 2(7), 232-235. doi:10.24032/ijeacs/0207/05
- Simplilearn. (2019). *Deep Learning Algorithms You Should Know About*. Retrieved from https://www.simplilearn.com/deep-learning-algorithms-article
- Singh, H., & Singh, J. (2012). Human Eye Tracking and Related Issues: A Review. International Journal of Scientific and Research Publications, 2(9).
- Singh, K., Seth, A., Sandhu, H. S., & Samdani, K. (2019). A Comprehensive Review of Convolutional Neural Network based Image Enhancement Techniques. In 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN). Pondicherry, India, India: IEEE. doi:10.1109/ICSCAN.2019.8878706

- Stember, J. N., Celik, H., Krupinski, E., Chang, P. D., Mutasa, S., Wood, B. J., Bagci, U. (2019). Eye Tracking for Deep Learning Segmentation Using Convolutional Neural Networks. *Journal of Digital Imaging*, 32(4), 597-604. doi:https://doi.org/10.1007/s10278-019-00220-4
- Wagner, S. (2018). Reinforcement Learning and Supervised Learning: A brief comparison. Retrieved from https://hackernoon.com/reinforcement-learning-and-supervisedlearning-a-brief-comparison-1b6d68c45ffa
- Wang, L., & Sng, D. (2015). Deep Learning Algorithms with Applications to Video Analytics for A Smart City: A Survey. *Computer Vision and Pattern Recognition*, 8
- Wang, S., Woods, R. L., Costela, F. M., & Luo, G. (2017). Dynamic gaze-position prediction of saccadic eye movements using a Taylor series. *Journal of Vision*, 17(14). doi:10.1167/17.14.3
- Wąsikowska, B. (2014). The Application of Eye Tracking in Business. Retrieved from https://depot.ceon.pl/handle/123456789/11248
- WikiStat. (n.d). Neural Networks and Introduction to Deep Learning. Retrieved from https://www.math.univ-toulouse.fr/~besse/Wikistat/pdf/st-m-hdstat-rnn-deep-learning.pdf
- Xiong, Y., Kim, H. J., & Singh, V. (2019). Mixed Effects Neural Networks (MeNets) with Applications to Gaze Estimation. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (pp. 7743-7752).
- Yua, M., Tang, X., Lin, Y., Schmidt, D., Wang, X., Guo, Y., & Liang, B. (2018). An Eye Detection Method Based on Convolutional Neural Networks and Support Vector Machines. *Intelligent Data Analysis*, 22, 345–362. doi:10.3233/IDA-173361
- Zhao, Q., Chang, S., Harper, F. M., & Konstan, J. A. (2016). Gaze Prediction for Recommender Systems. In Proceedings of the 10th ACM Conference on Recommender Systems, (pp. 131-138). doi:10.1145/2959100.2959150

APPENDICES

#### **APPENDIX 1**

## IMAGE ENHANCEMENT CODE IN PYTHON PROGRAMMING LANGUAGE

```
from keras.preprocessing.image import load_img, img_to_array
import matplotlib.pyplot as plt
import os
import time
from PIL import ImageEnhance
# size of the image: 48*48 pixels
pic size = 48
# input path for the images
base_path = "C:/PythonFiles/new_gaze/MPIIGaze_Dataset/train/CentreView/"
newpath="C:/PythonFiles/new_gaze/MPIIGaze_Dataset/enhancedTrain/CentreView/"
NAME = "Views_vs_Views".format(int(time.time()))
plt.figure(0, figsize=(12,20))
cpt = 0
for fig in os.listdir(base_path):
    fulpath= base path + fig
    img = load img(fulpath)
    # enhance image brightness
    enh_bri = ImageEnhance.Brightness(img)
    brightness = 1.5
    image brightened = enh bri.enhance(brightness)
    # image brightened.show()
    image_brightened.save(newpath + os.path.basename(fulpath))
    print(fulpath)
    print(os.path.basename(fulpath))
    plt.imshow(img, cmap="gray")
```

#### **APPENDIX 2**

## CNN MODEL CODE IN PYTHON PROGRAMMING LANGUAGE

import numpy as np import seaborn as sns from keras.preprocessing.image import load\_img, img\_to\_array import matplotlib.pyplot as plt import os from keras.preprocessing.image import ImageDataGenerator from keras.callbacks import ModelCheckpoint from keras.layers import Dense, Input, Dropout, GlobalAveragePooling2D, Flatten, Conv2D, BatchNormalization, Activation, MaxPooling2D from keras.models import Model, Sequential from keras.optimizers import Adam import time from tensorflow.keras.callbacks import TensorBoard

# size of the image: 48\*48 pixels
pic\_size = 48

```
# input path for the images
base_path = "C:/PythonFiles/new gaze/MPIIGaze_Dataset/"
NAME = "Views_vs_Views".format(int(time.time()))
plt.Figure(0, Figuresize=(12,20))
cpt = 0
```

```
for expression in os.listdir(base_path + "train1/"):
    for i in range(1,6):
        cpt = cpt + 1
        plt.subplot(7,5,cpt)
        img = load_img(base_path + "train1/" + expression + "/" + os.listdir(base_path +
"train1/" + expression)[i], target_size=(pic_size, pic_size))
        plt.imshow(img, cmap="gray")
```

```
plt.tight_layout()
# plt.show()
```

for expression in os.listdir(base\_path + "train1"):
 print(str(len(os.listdir(base\_path + "train1/" + expression))) + " " + expression + "
images")

```
# number of images to feed into the NN for every batch batch_size = 10
```

datagen\_train = ImageDataGenerator()
datagen\_validation = ImageDataGenerator()

# cnn nb\_classes = 4

```
# Initialising the CNN
model = Sequential()
```

```
model.add(Conv2D(64, (3, 3)))
model.add(Activation('relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))
```

```
model.add(Conv2D(64, (3, 3)))model.add(Activation('relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))
```

model.add(Flatten()) # this converts our 3D feature maps to 1D feature vectors

```
model.add(Dense(64))
```

```
model.add(Dense(1))
model.add(Activation('sigmoid'))
```

# plot the evolution of Loss and Acuracy on the train and validation sets

```
plt.Figure(Figuresize=(20,10))
plt.subplot(1, 2, 1)
plt.suptitle('Optimizer : Adam', fontsize=10)
plt.ylabel('Loss', fontsize=16)
plt.plot(history.history['loss'], label='Training Loss')
plt.plot(history.history['val_loss'], label='Validation Loss')
plt.legend(loc='upper right')
```

```
plt.subplot(1, 2, 2)
plt.ylabel('Accuracy', fontsize=16)
plt.plot(history.history['acc'], label='Training Accuracy')
plt.plot(history.history['val_acc'], label='Validation Accuracy')
plt.legend(loc='lower right')
plt.show()
```

## CHAPTER 1 INTRODUCTION

#### 1.1 Background of Study

The prediction of individual eye gaze is a research topic that has gained the interest of researchers with its wide range of applications because neural networks majorly increase the rate of accuracy of individual gaze. Our eyes play a very vital role when it comes to focusing on something. The eyes easily respond to anything viewable (Parikh & Kalva, 2018) and an individual's attention generally depends on the direction in which the individual is looking at. Eyesight which is obviously an old means of communication (Mohamed, Silva, & Courboulay, 2007) is important for individuals to communicate with machines or computers (Barbuceanu & Antonya, 2009) as it collects input data for the human brain to process activities.

According to (Recasens, Khosla, Vondrick, & Torralba 2015), individuals are capable of paying attention to each other's gaze so as to realize the focus of that particular individual. This act can be termed 'gaze following'. Gaze following permits us as humans to interpret the thoughts of other individuals, their current and future activity. So, we can say that gaze is useful for interpreting emotions, focus and interactions of individuals (Xiong, Kim, & Singh, 2019).

Gaze are important in media applications (Feng, Cheung, Tan, Callet, & Ji, 2013), business (Wąsikowska, 2014), robotics and computer human interaction (Recasens, Khosla, Vondrick, & Torralba, 2015), 3D gaming (Koulieris, Drettakis, Cunningham, & Mania, 2016), medicine (Harezlak & Kasprowski, 2017), virtual reality (Wang, Woods, Costela, & Luo, 2017) marketing and psychology (Stember, et al., 2019), smart home and mobile device authentication. Aside the applications mentioned above, we can apply gaze prediction in:

- *i. Security:* To monitor individual abnormal behaviour in situations where an individual intends to cause chaos in public areas such as airports, hospitals, schools, highways, malls, etc.
- *ii. Interrogation:* To determine the accuracy of statements made by suspected individuals in custody of the police, FBI or CIA.

*iii. Human behaviour:* To further predict individual emotion, thought and / or future activity and even health conditions of individuals in hospitals.

For gaze research, image pre-processing activities are usually carried out together with machine learning or neural networks (Dimililer, et al., 2018) which are categorized as deep learning. Image pre-processing activities such as image enhancement aims at producing better results when implemented in any form of computer related research. Gaze estimation methods usually regress gaze directions directly from a single face or eye images.

#### **1.2 Problem Statement**

So many activities go on in our everyday life as humans and most of them are determined by our gaze. But is it actually possible to predict our gaze? i.e. the gaze of individuals. It might seem difficult to predict the gaze direction of individuals because as humans, we have the choice to view just anything at our own discretion. So, we propose an enhanced individual eye gaze prediction using neural networks.

## 1.3 Aims and Objectives

- To predict direction in which an individual is focusing on.
- To determine the accuracy of individual eye gaze, with original image.
- To determine the accuracy of individual eye gaze, with enhanced version of the original image.
- To compare the accuracy of the original gaze image with enhanced gaze image.

## 1.4 Scope of Study

In this study, we will review research work that relates to gaze and its applications, obtain dataset for the purpose of our experiment after which the results of the experiment will be analysed. Only the prediction of individuals' gaze will be done by apply image enhancement technique.

## 1.5 Methodology

The dataset that will be obtained will be trained and tested using python programming language. After image acquisition, we shall carry out image pre-processing with image enhancement technique and finally use deep neural network to predict individual eye gaze direction. The purpose of image enhancement is to improve the quality of our image dataset.

## **1.6 Expected Result**

It is expected that at the end of this project, we would have come up with a basic model, using convolutional neural network to:

- predict the gaze direction of individuals.
- determine the accuracy of individual eye gaze, with original image.
- determine the accuracy of individual eye gaze, with enhanced version of the original image.
- compare the accuracy of the original gaze image with enhanced gaze image.

# CHAPTER 2 LITERATURE REVIEW

### 2.1 The Human Visual System

The human eye is a vital organ that is very important for sight. It interprets visual data from our physical environment into a particular and precise image. The image that the human eye produces is not a direct graphic image. All of what we see is interpreted by our brain. So, we can say that an individual's eye feeds the brain with input which in turn processes by sending it to image processing points in the brain (Manrow, 2019) that is further translated to the image that we see (Idrees, 2015).



Figure 2.1: A cross section of the human eye (Braille Institute, 2018)

Usually, with the aid of light, the retina creates the images that we see from our physical environment (LaValle, 2019). The light which enters the eye is controlled by the iris (Singh & Singh, 2012) which acts as a muscle (Manrow, 2019) and converted into electrical impulses by the retina which is situated at the back of the eye (Idrees, 2015), as shown in Figure 2.1 above. The retina holds photoreceptors that helps transform light from the physical world into neural pulses. The fovea which sits on the centre of the retina (Manrow, 2019) is loaded with photoreceptors which makes it really sensitive to colour (Majaranta & Bulling, 2014).

#### 2.2 Eye Tracking

Eye tracking, also known as point of gaze (Farnsworth, 2019) or oculography (Borys & Plechawska-Wójcik, 2017) refers to the process of tracing individual eye movement so as to determine the gaze direction of the same individual (Singh & Singh, 2012). Eye tracking or gaze usually may include individual gaze direction prediction as seen in the research work of Dimililer, et al., (2018) or prediction of an individual's point of gaze. Eye tracking is important because certain unique information can be obtained from individual facial characteristics (Yua, et al., 2018). Systems developed for the purpose of eye tracking are composed of software programs that detect pupil, carry out image processing and data filtering operations, and finally ensure recording of individual eye transitions (Lupu & Ungureanu, 2013).

According to (Kar & Corcoran 2017), eye movement being studied in the applications and research of eye gaze can be categorised into the following:

#### 2.2.1 Fixations

Fixation may refer to the point of attention of indivual gaze. We can say that it is a measure of an individual's optical attention. Krauzlis, Goffart, and Hafed, (2017) states that individuals are usually in full control of the position of their eye movemment during fixation. The gaze fixation of individuals usually continues for about 100 milliseconds to 1000 milliseconds, leaving most gaze fixation from about 200 milliseconds to 500 milliseconds (Singh & Singh, 2012).

#### 2.2.2 Saccades

Saccades are brief eye movements that are quick. They can last up to tens of milliseconds, measuring peak speeds of hundreds of degrees per second (Krauzlis, Goffart, & Hafed, 2017). Saccades are necessary in order for individuals to correctly recognize any visual content (Majaranta & Bulling, 2014).

#### 2.2.3 Scanpath

According to (Drusch, Bastien, & Paris2014), scanpath comprises of a finit number of both fixation points that are linked by saccades. Scanpaths are important to help understand the optical characteristics of individuals.

#### 2.2.4 Gaze duration

This is the total number of gaze fixation all summed up. Studies reveals that gaze duration usually takes long for long words when individuals read or study (Hohenstein, 2013).

Gaze detection and prediction has been researched on by several researchers and has been applied in computer vision, the area of human computer interaction (Dimililer, et al., 2018) and many other fields such as medicine, military, education, security, government, etc. Mohamed, Silva, and Courboulay (2007) states that its application has led to the subsistence of many ways to detect and track the direction of individual eye gaze. Eye tracking serves as the root for observing targets (Majaranta & Bulling, 2014). According to Singh and Singh (2012), there has been a rapid increase in the utilization of eye tracking systems due to the fall in price of eye tracking systems. There is a large market for researchers using eye tracking systems for research purpose, especially in human computer interaction.

Eye gaze research comes with difficulties such as deciding the gaze fixation and determining the point of focus of individual gaze (Dimililer, et al., 2018). The main purpose of gaze tracking is to obtain useful information from the gaze of individuals (Farnsworth, 2019) in order to determine the action of the individuals.

#### 2.3 How Does Gaze Tracking work?

Usually, with eye trackers, infrared light is reflected in an individual's eyes which are picked up by the camera of the eye tracker. The eye tracker now determines the direction of the individual's gaze through calculations and filtering (How Eye Tracking Works, n.d.). Figure 2.2 below describes the process of gaze tracking.



Figure 2.2: How gaze works with eye trackers (Tobii Pro, n.d)

## 2.4 Deep Learning

Although they are related, machine learning together with deep learning and neural networks are all sub unit areas of artificial intelligence.

Machine learning allows computers or machines to function in such a way, performing some specific kinds of assignment with the aid of intelligent software (Mohammed, Khan, & Bashier, 2016) thereby mimicking the intelligence of higher animals (humans). Machines with this software are able to sort out compound patterns and carefully determine actions based on data. Machine learning is employed in applications such as audio recognition (Ogidan, 2017), medical imaging with neural networks (Dimililer, 2013), etc. The different machine learning techniques are described in figure 2.3 below:



Figure 2.3: Machine learning techniques (Mohammed, Khan, & Bashier, 2016).

## 2.4.1 Supervised learning

Supervised learning is basically concerned with all forms of labeled data. Hence, its algorithms can easily be applied to problems related classification of patterns and regression of data. It is used to discover mapping rules for the prediction of outputs with unfamiliar inputs (Wang & Sng, 2015). Labeled data requires an agent/supervisor to be present during learning process.



#### 2.4.2 Unsupervised learning

Mishra and Saroha (2016) mentioned that unlike supervised learning, unsupervised learning is applicable when unlabeled data are used. In other words, it is useful where labeled data are not required. This learning algorithm deals with a form of network learning that gives the right output without the help without any interaction from an external behavior. In unsupervised learning, the unlabeled data requires no agent/supervisor to be present during learning process.



Figure 2.5: Unsupervised learning model (Ramesh, 2018)

#### 2.4.3 Semi-supervised learning

Semi-supervised learning is a comprises both unsupervised and supervised learning, i.e. it requires both unlabeled and labeled data. According to (Reddy, Pulabaigari, & Eswara 2018), semi-supervised learning is usually used when labeled data are hard to obtain. The Figure below gives the description of unsupervised learning.



Figure 2.6: Semi-supervised learning model (Moltzau, 2019)

### 2.4.4 Reinforced learning

An agent is usually present and its learning process is an interactive one, usually between the agent present and the environment (Campos, 2018). Reinforcement learning is based on trial and error learning process (Sathya & Abraham, 2013). Figure 2.7 below describes the reinforcement learning model.



Figure 2.7: Reinforcement learning model (Wagner, 2018)

Deep learning which is also referred to as hierarchical learning is a section of machine learning algorithms (Hordri, Samar, Yuhaniz, & Shamsuddin, 2017). When compared to machine learning, deep learning has the ability to exhibit high rate of intelligence. Although, they do not fully comprehend narratives or long information like expert systems, they are excellent in the way (Bhatia & Rana, 2015).

According (Bhatia & Rana, 2015) perceptron algorithm is regarded as the first machine to depict human intelligence but it was limited in its ability to learn. This brought neural networks which is in vogue today. Neural networks have been applied in nearly all fields to obtain so many results in all facets of life. They are useful for grouping unlabeled data by similarities in the set of given inputs. Neural networks are designed to function just like the human brain. They make use of some machine perception to interpret sensory data.

#### 2.5 Image Enhancement

Like machine learning and neural networks, image processing techniques are also applied in the area of computer vision, optical character recognition, biometric verification, remote sensing, face detection and digital video processing (Dewangan, 2016) and medical imaging such as MRI (Dimililer & Ihan, ICAFS, 2016), X-ray images (Dimililer, Backpropagation Neural Network Implementation for Medical Image Compression, 2013), CT and medical palmistry (Dewangan, 2016).

The first step for any image processing activity is usually image acquisition, image preprocessing, application of image processing technique and finally result. The image processing technique to be applied may be:
# i. Image restoration



Original image with noise

Restored image

Figure 2.8: Image restoration (Color Experts, 2013)

This technique is applied when an image contains noise or is blur. It is used to restore a corrupt image back to its original form by removing the blur or noise contained in the image (Rani, Jindal, & Kaur, 2016), as described in figure 2.8 above.

*ii.* Image enhancement



Original image

Restored image

# Figure 2.9: Image Enhancement (Robert, 2018)

Although it is usually regarded as image pre-processing technique, it is applied in improving the details in and quality of an image, as shown in figure 2.9 above. Image enhancement is briefly explained, just before section 2.6.

# iii. Image recognition

This procedure is useful for defining objects (including human, actions and locations) in an image (Gupta, 2018), as shown in figure 2.10 below.



Figure 2.10: Image recognition (Azati Software, 2019)

# iv. Image segmentation



Original image Segmented image Figure 2.11: Image segmentation (Kumar, 2019)

Image segmentation is the act of dividing specific areas in an image with like features, that are useful for information analysis.

# v. Image resizing

This technique is applied when there is need to format the dimension of an image either by increasing or decreasing the number of pixels in the image.



Figure 2.12: Image resizing (Sajjad, et al., 2017)

### vi. Image compression



Original Lena Image (12KB size) Lena Image, Compressed (85% less information, 1.8KB)

Lena Image, Highly Compressed (96% less information, 0.56KB)

Figure 2.13: Image compression (DBA Blog, 2015)

We usually apply this technique when there is need to reduce the amount of storage data, in bits for an image.

Each of the image processing techniques mentioned above also have special methods by which they can be used to achieve the necessary results.

Image processing techniques in most cases are used with machine learning/deep learning/neural networks. The machine learning/deep learning/neural network used may be for the purpose of either prediction or classification, as the case may be. Image processing generally provide aids to:

- i. Make visible objects that are invincible.
- ii. Obtain an improved quality of the original image.
- iii. Study certain features in an image.
- iv. Measure the pattern of different objects in an image.
- v. Distinguish between objects and features in an image.

The Figure below illustrates how machine learning/deep learning/neural networks can be combined with image processing techniques.



# Figure 2.14: Image processing activity with Machine Learning/ Deep Learning/ Neural Networks

The term "image enhancement" may be described as a method that involves the manipulation of pixels from an image to achieve a clearer interpretation from that image or carry out other image processing techniques (Hussain & Lone, 2018) to gather useful information. Although image enhancement task poses to be a challenging task (Shukla, Potnis, & Dwivedy, 2017), its main purpose in image processing is to obtain an image that is far better than its original, in terms of quality (Bhardwaj, Kaur, & Singh, 2018).

Although image enhancement techniques are categorised into two; spatial domain and frequency domain, an image may be enhanced using one or more of the following techniques:

- i. Brightness
- ii. Contrast
- iii. Colour
- iv. Sharpness

For an image to be visually interpreted, at least one of these techniques may be applied. The above mentioned image enhancement techniques either manipulate the pixels in an image directly or Fourier transform of the image (Kaur & Taqdir, 2016). Some individuals interchange brightness

and contrast, but they are not the same. The brightness in an image refers to the ratio mix of lightness and darkness it contains while contrast refers to the difference value of the image from its colour and brightness. For most imaging tasks, image enhancement techniques are the most popular techniques employed for pre-processing images (Singh, Seth, Sandhu, & Samdani, 2019).

#### 2.6 Related Works

Yua, et al. (2018) presented an article in which they combined both CNN and support vector machines as a model to detect individual eyes. The dataset used included face database from IMM, FERET, ORL and BioID. With this model, they obtained a better detection accuracy, even with different eye defects.

Park, Spurr, and Hilliges (2018), organised a one of a kind research in which deep neural networks was used to estimate individual gaze based on one eye. This, they referred to as gazemaps with 3D point of gaze estimation. With dataset obtained from MPIIGaze, it was discovered that the estimation of the point of gaze is inflenced by the use of both individual eyes and the resolution of the image. The results obtained were interpreted to be extremely accurate.

DCT was applied by (Dimililer, et al., 2018) to detect and predict the direction of individual eye gaze. Image compression is an image pre-processing technique that they employed using back propagation neural networks. These techniques were used to categorize the gaze dataset into right, left and centre as gaze directions.

Jaques, Conati, Harley, and Azevedo (2014), carried out a research to using feature selection and machine learning to forecast the emotions of students with eye tracking as regards learning curiosity and enthuaism. The data that was used was gathered through an Intelligigent Tutoring System (ITS) called MetaTutor.

Jerry, Lam, and Eizenman, (2008) used the CNN algorithm for detection of eyes in gaze estimation systems. Their work did not include any form of image pre-processing. Three participants were used and it was discovered that for head movements, the CNN was able to detect individual eyes.

# CHAPTER 3 DEEP NEURAL NETWORKS

We briefly described deep learning in chapter 2 but in this chapter, we will focus more on it, stating and explaining some its algorithms.

In section 2.4, we stated that both machine learning and deep learning are both subsets of artificial intelligence. Deep neural networks may also be categorized as a type (Shaikh, 2017) or higher version of machine learning, because when used in an application they tend to achieve better results than the traditional machine learning. DNNs simply comprise of algorithms with neural network, forming deep layers of learning patterns. This is how we derive the term "deep learning".

# 3.1 Comparison of Deep Learning Over Machine Learning

- i. Deep learning depends on high level machines, unlike machine learning which can function on low level machines.
- ii. Machine learning cannot perform well to obtain high accuracy of results if the data is extremely large. Deep learning on the other hand will obtain high accuracy result with extremely large data but the reverse is the case with small data.
- iii. Deep learning can go as far as learning optimum pattern features of data, this is unlike machine learning.
- iv. Deep learning algorithms take really long time to execute training. Training might even go on for a few weeks. While machine learning takes less time to execute.

Deep neural networks may also be referred to an Artificial neural networks, with several different layers (Albawi, Mohammed, & Alzawi, 2017).

### **3.2** Components of a Neural Network

A neural network is simply composed of an input layer, multiple hidden layers and an output layer, as shown in figure 3.1 below.



Figure 3.1: A basic neural network (WikiStat, n.d)

Like machine learning, neural networks are designed to function like the human brain. In a basic neural network, each layer includes a set of nodes, where computation is carried out. Each layer serves as input for the next layer until the final (output) layer is reached. In this process where the neural network is learning patterns, weights are generated, computed and neurons are activated with some activation function. The more the layers involved in neural networks, the deeper the rate of accuracy. Through this, they can accept, process and interpret sensory data. They are able to perform the following tasks:

- i. Classification
- ii. Prediction
- iii. Clustering

# **3.3 Deep Learning Architecture**



Like machine learning, some of the algorithms of deep neural networks can be categorised

Figure 3.2: Deep learning architecture (Simplilearn, 2019)

into deep supervised, classification for example and deep unsupervised learning, clustering for example (Nicholson, n.d), as shown in figure 3.2 above. Deep learning may either use labelled or unlabelled data for both training and validation. Although they depend on neural networks that are modeeled after the human brain, they are capable of self learning at some point. During the training phase, the algorithms extract and learn patterns of features from the data used, re-group them and learn some more patterns, before obtaining required results based on the kind of model used i.e. classification, prediction or clustering.

According to (Simplilearn, 2019), we can employ deep learning in fields such as fraud detection, audio and speech recognition, medical imaging, business management, computer vision, security surveillance, bioinformatics, etc. Some of the deep neural networks algorithm that exist include:

# i. The Multilayer Perceptron

A perceptron is simply a one neural model and "the first" in the series of neural network algorithms (Brownlee, 2016). This algorithm is used to train and classify non linearly separable problems, thereby solving difficult computational operations and may be applied in machine translation, image verification, data classification and e-commerce.

# ii. Recurrent Neural Networks

This algorithm is usually employed in Natural Language Processing (Britz, 2015). It is called RNN because for each neuron (node) in each layer that serves as input for the next layer, it repeats the same computational operation sequentially.

# iii. Convolutional Neural Networks

CNN is a type of supervised learning algorithm with multilayer perceptrons performing feed-forward operation. When working with images, CNN model is the best algorithm to select. We later discus its components in section 4.2, since it is the algorithm that we adopt for the purpose of our research.

## iv. Recursive Neural Networks

Recurrent neural networks are unlike recurrent neural networks and are usually applied on structured data input.

# CHAPTER 4 METHODOLOGY

Firstly, after we acquire the selected images from our dataset, we performed image pre-processing with image enhancement technique by applying brightness and finally use CNN which is a type of deep learning algorithm to predict individual eye gaze direction.

We divided the program execution into two parts. The first part of the program execution trains and validates our image dataset without the application of image enhancement. While image enhancement was applied in the later. Figure 4.2 describes the conceptual model of both parts of the program execution

### 4.1 Image Enhancement Flowchart

First of all, we input each image into our program, after which we apply a brightness of  $1.5_{units}$  and further save the image if the process is successful.



Figure 4.1: Basic flowchart for image enhancement

Figure 4.1 above describes the workflow of how image enhancement is applied. This is to enable us further our research with this image processing technique. Python programming language enables us to enhance images with the four (4) different techniques as mentioned in section 2.5 above. We choose brightness for enhancing our dataset for better pictorial interpretation.

Brightness is usually applied by decreasing or increasing the matrix of the image either by subtraction or addition.

#### **4.2 Conceptual Model**

For each of our experimental sections, i.e. ordinary image dataset and enhanced version of our image data set, we make use of a basic CNN model with three (3) different layers of convolutions. For each layer, we carry out max pooling and connect the last layer to a fully connected layer before we finally get our output.





CNNs are made up of layers such as:

#### i. Convolutional layers

This is the initial set of layers in a CNN model. Here, the convolutional layers apply filtering to summarize the pixels in an image. The main function of the convolutional

layer is to reveal distinct visual features such as colour drops, lines, edges, etc. By doing this, the CNN learns specific characteristics, hierarchies of several patterns in an image.

# ii. Pooling layers

There is a pooling layer after each convolutional layer. Two techniques that may be used for pooling are the average pooling or maximum pooling.

# iii. Fully connected layers

These are the important layers. It does the actual leaning in the deep neural network (CNN in this case). It comprises of several perceptron layers and identifies the object in a class.

For our CNN model, we use 3 convolutional layers and for each layer, we carry out pooling operations and connect to one full layer before the actual prediction is done.

# 4.3 Tools Used

In this section, we give brief descriptions of some of the tools we use in accomplishing the results of our research.

# 4.3.1 Dataset

Our dataset consists of 100 images randomly obtained from MPIIGaze dataset. These images are divided into four (4) categories for our work, with 25 images in each category depicting:

*i. Center view:* We categorize this view as a fixed point for an individual's horizontal gaze direction in any scene as shown in the figure below.



Figure 4.3: Sample center view image

*ii. Down view:* We categorize images with individuals looking downwards in any scene, as described in the figure below.



Figure 4.4: Sample down view image

*iii. Left view:* Images in this category contain scenes of individuals facing the left side view, as described in figure 4.5 below.



Figure 4.5: Sample left view image

iv. Right view



#### Figure 4.6: Sample right view image

Images in this category contain scenes of individuals facing the right side view, as described in figure 4.6 above.

#### **4.3.2** Python programming language

We choose python programming language because we discovered that it is most useful when implementing deep learning and machine learning algorithms. It comprises of several libraries/modules and framework, some of which are briefly described below.

# i. Keras

Keras is a deep learning library enabled with high level neural networks API for generating easy and quick models. It supports CNN and recurrent networks.

#### *ii.* Tensorflow framework

Tensorflow is a framework that that represents numerical computation with dataflow graphs. It is used for research experimentation on data that is to be trained and validated or tested with machine learning and deep learning algorithms. We can say that it serves as a platform for training neural networks. It requires a specific line of code to be installed with command prompt or python terminal. There are also IDEs for ease of its installation and the installation of other libraried required for projects executed in python programming language. One of such IDEs, which we made use of is the JetBrains Pycharm Community Editor.

### *iii.* TensorBoard module

The 'TensorBoard' module constitutes a set of applications that enables users to view tensorflow graphs.

#### iv. PIL module

The PIL enables users of python programming language to perform image processing techniques. We employed it by using it to enhance our image dataset.

#### 4.3.3 Image preprocessing with image enhancement using brightness

For preprocessing our dataset, we make use of the brightness image enhancement technique with just a few lines of code. As we mentioned earlier in section 2.5 above, the main purpose of employing the image enhancement technique with brightness is to improve on the quality of our dataset to achieve a higher accuracy of gaze prediction. The following Figures below gives a pictorial view of how the original image and enhanced image.



Original image



Enhanced image

#### Figure 4.4: Sample original image and enhanced image without visual aid



Original image



Enhanced image

Figure 4.5: Sample original image and enhanced image with visual aids

Usually, we encounter individual faces with or without spectacles or glasses as we know them to be. From the above shown figures, i.e. figure 4.4 and 4.5 we have applied image enhancement on these two individual faces. The use of spectacles does not affect the application of image processing techniques or deep learning algorithms, so long as the eyes are detectable.

# 4.4 System Specification Requirements

In this section, we give a brief description of the expected behaviour and features of our software.

# 4.4.1 Non-functional requirements

# i. Hardware requirements

- A computer system
- Minimum of Pentium IV / AMD A8-7410 APU or higher CPU
- Minimum of 512MB of RAM
- On-camera Monitors

# ii. Software Requirements

- Windows 7, Windows 8, Windows 10 or any other Operating System that supports python.
- 32-bit / 64-bit Operating System
- Python 3.6
- Anaconda/miniconda
- PyCharm IDE

# **4.4.2 Functional requirements**

Since we are adopting a basic technique for our CNN model, our CNN model shall be able to perform the following functions:

- Our model shall be able to predict individual gaze, based on the directions that we defined, i.e. centre view, left view, right view and down view.
- Our model shall obtain accuracy of individual gaze prediction with ordinary image.
- Our model shall obtain accuracy of individual gaze prediction with the enhanced version of the original image.
- Our program shall be able to apply brightness technique of image enhancement.

# CHAPTER 5 RESULTS AND DISCUSSIONS

## 5.1 Results

With the basic CNN model employed and the brightness image technique used, we have been able to obtain some amount of reasonable results, ranging from the random selection of images used in our research to the task of obtaining all the accuracy required. In this section, we generally display the results that we obtained



# 5.1.1 Random plots of images used

Figure 5.1: Image dataset

Figure 5.1 above depicts how the selected images look like, with individual gaze facing the four directions that we defined earlier (center view, down view, left view and right view). The images have a 48 x 48 dimension in size. We choose to display a random selection of 20 images in all gaze direction categories.

# 4.1.2 Model summary

The figure below describes the whole parameters in each layer from our CNN model. The whole parameter sums up to 5,654,276 parameters, the total number of trainable parameters are 5,651,332 parameters and the total number of untrainable parameters are 2,944 parameters.

Layer (type)	Output Shape	Param	#	-	max_pooling2d_3 (MaxPooling2 (None, 6, 6, 512) 0			
conv2d_1 (Conv.	2D) (None, 48, 4	8, 64)	640	-	dropout_3 (Dropout)	(None, 6, 6, 512)	0	
batch_normalizat	tion_1 (Batch (None, 48	, 48, 64)	256	-	flatten_l (Flatten)	(None, 18432)	0	
activation_1 (Act	tivation) (None, 48, 48	, 64)	0	-	 dense_l (Dense)	(None, 256)	4718	848
max_pooling2d_	l (MaxPooling2 (None,	24, 24, 64	•) 0	-	batch_normalization_/	4 (Batch (None, 256)		1024
dropout_1 (Drop	out) (None, 24, 24	, 64) (	)	-	activation_4 (Activati	on) (None, 256)	0	
conv2d_2 (Conv.	2D) (None, 24, 24	4, 128)	204928	-	dropout_4 (Dropout)	(None, 256)	0	
batch_normalizat	tion_2 (Batch (None, 24	, 24, 128)	512	-	dense_2 (Dense)	(None, 512)	1315	84
activation_2 (Act	tivation) (None, 24, 24	, 128)	0	-	 batch_normalization_;	5 (Batch (None, 512)		2048
max_pooling2d_	2 (MaxPooling2 (None,	12, 12, 12	8) 0	-	activation_5 (Activati	on) (None, 512)	0	
dropout_2 (Drop	out) (None, 12, 12	, 128)	0	-	dropout_5 (Dropout)	(None, 512)	0	
conv2d_3 (Conv.	2D) (None, 12, 12	2, 512)	590336	-	dense_3 (Dense)	(None, 4)	2052	
batch_normalizat	tion_3 (Batch (None, 12	, 12, 512)	2048	-	Total params: 5,654,2	76		
activation_3 (Act	tivation) (None, 12, 12	, 512)	0	-	Non-trainable params: 5,6	: 2,944		

Figure 5.2: CNN model summary

For the purpose of our research, we have used three convolutional layers, each with max pooling of 2 factor and one fully connected layer. Table 5.2 describes the summary of our model, while Figure 5.3 describes the conceptual view of how our CNN model was implemented.

#### **5.2 Discussions**

In this section, we discuss the results we obtained both for the original image dataset and the enhanced version of the dataset.

For our first aim which is to predict gaze, we were able to achieve results as shown in figure 5.3 below.



Figure 5.3: Right view gaze prediction

With just a picture on the wall, our CNN model was able to predict the individual's eye gaze as a right view direction. Hence, we can say that our CNN model is now capable of predicting individual eye gaze. The following sections in this chapter reviews the accuracy results that we obtained.

# 5.2.1 Original results without the effect of image enhancement

The model starts to stabilize at about 61% without the application of image enhancement as shown in Figure 5.4 below. Figure 5.5 describes the best model that we obtained with the four (4) different classes, giving the validation accuracy as 69%, since our model did not improve from 69%. This model is tested and validated without the aid of the image enhancement technique.

#### Optimizer : Adam



Figure 5.4: Evolution of loss and accuracy of original dataset

Here, we obtain results for both loss and accuracy of the training and validation (testing) phases. The difference of errors that we have obtained for our dataset is 0.68% while we obtain the difference in value of accuracy is 7%.

```
Epoch 00095: val_accuracy did not improve from 0.67000
Epoch 96/100
Epoch 00096: val_accuracy did not improve from 0.67000
Epoch 97/100
Epoch 00097: val_accuracy did not improve from 0.67000
Epoch 98/100
Epoch 00098: val_accuracy improved from 0.67000 to 0.69000, saving model to model_weights.h5
Epoch 99/100
Epoch 00099: val_accuracy did not improve from 0.69000
Epoch 100/100
Epoch 00100: val_accuracy did not improve from 0.69000
```

#### Figure 5.5: Validation prediction accuracy

#### 5.2.2 Original results with the effect of image enhancement

The model starts to stabilize at 64% with the application of image enhancement as shown in Figure 5.6 below. Figure 5.7 describes the best model that we obtained with the four (4) different classes, giving the validation accuracy as 72%. This model is tested and validated with the aid of the image enhancement technique.



Figure 5.6: Evolution of loss and accuracy of enhanced dataset

Here, we obtain results for both loss and accuracy of the training and validation (testing) phases. The difference of errors that we have obtained for our dataset is 3.45% while we obtain the difference in value of accuracy is 7%.

Epoch 00096: val\_accuracy did not improve from 0.72000 Epoch 97/100 1/1 [========] - 39s 39s/step - loss: 0.9386 - accuracy: 0.6100 - val\_loss: 0.9237 - val\_accuracy: 0.7000 Epoch 00097: val\_accuracy did not improve from 0.72000 Epoch 98/100 1/1 [=========] - 36s 36s/step - loss: 0.9575 - accuracy: 0.6200 - val\_loss: 0.9194 - val\_accuracy: 0.7000 Epoch 00098: val\_accuracy did not improve from 0.72000 Epoch 00099: val\_accuracy did not improve from 0.72000 Epoch 00099: val\_accuracy did not improve from 0.72000 Epoch 00099: val\_accuracy did not improve from 0.72000 Epoch 00099: val\_accuracy did not improve from 0.72000 Epoch 00099: val\_accuracy did not improve from 0.72000 Epoch 00099: val\_accuracy did not improve from 0.72000 Epoch 100/100 1/1 [===========] - 36s 36s/step - loss: 0.8757 - accuracy: 0.6400 - val\_loss: 0.9102 - val\_accuracy: 0.7100 Epoch 00100: val\_accuracy did not improve from 0.72000

Figure 5.7: Validation prediction accuracy with image enhancement technique

For both phases, i.e., original results with and without the effect of image enhancement, we used a total number of two hundred (200) epochs, with 100 epochs belonging to each phase. Each epoch represents a full training cycle for the image dataset. There is no limit to choosing the total number of epochs for training and validating, but the purpose of choosing and using one hundred (100) epochs is to ensure optimal learning of the image dataset since we randomly selected a total of one hundred (100) images. We included a batch size of 10 images in each batch for both training and validation.

We use pooling to reduce the volume of each feature but main the relevant information each convolutional layer. The graph results obtained in Figures 5.2 and 5.4 were plotted with training data saved by the model in '. json' format. Since the loss decreased and accuracy increased in both phases, it is proper to say that our model was successful in learning and not cramming.

The difference in the accuracy of result between the original and enhanced dataset is simply 3%. The more we trained, validated and saved our model, the higher the accuracy of gaze prediction we obtained. With the result that we obtained, it is revealed that with image enhancement, there is a higher rate of gaze prediction accuracy.

# **5.3 Challenges**

We faced challenges in obtaining dataset. Original datasets are not easily accessible and these datasets vary, even for the required research domain.

A HP windows 8.1 was used to carry out the code execution of our research. Because of its specifications, the runtime was intensively slow. As a result, for the purpose of our research, we had to randomly select one hundred (100) images, with twenty-five images in each of the four categories that we previously defined.

# CHAPTER 6 CONCLUSIONS AND FUTURE WORK

#### **6.1 Conclusions**

We use a basic CNN model to train and validate a hundred (100) images so as to predict the gaze direction of individuals. This is majorly because we were restricted with the processing speed of our computer. As a result, we could not go further to compare our results with the work of other researchers as regards this area.

For the ordinary image, our model did not improve from 69%. On the other hand, validation accuracy with image enhancement resulted to be 72%. The difference in the accuracy of result between the original and enhanced dataset is simply 3%. The more we trained, validated and saved our model, the higher the accuracy of gaze prediction we obtained. With the result that we obtained, it is revealed that with image enhancement, there is a higher rate of gaze prediction accuracy.

Hence, we have been able to achieve our aims and objects; and finally, we can say that image enhancement has proved its purpose by providing image interpretation with better quality.

#### 6.2 Future Work

At this time, we cannot say that everything from neural networks to deep learning to machine learning and finally, artificial intelligence is all new. This is because they have existed for over a decade. They have been applied in virtually all areas, including gaze detection.

The evolution of all gaze technologies over time has yielded useful results for researchers in general, some of which have helped to predict the intentions and actions of individuals. However, there is more work to be done in terms of security with the aid of machine learning or deep learning and image processing techniques. In our future work, we shall compare the use of other image processing techniques for gaze prediction with other deep learning algorithms.