ANALYSIS OF SMARTPHONE IMAGES FOR VISION SCREENING OF REFRACTIVE ERRORS

YANG ZHONGQI

MPhil

The Hong Kong Polytechnic University

2022

The Hong Kong Polytechnic University

Department of Computing

Analysis of Smartphone Images for Vision Screening of Refractive Errors

Yang Zhongqi

A thesis submitted in partial fulfillment of the requirements

for the degree of Master of Philosophy

October 2021

Certificate of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____(Signed)

Yang Zhongqi (Name of student)

Abstract

Refractive error is the most common of visual impairments and impacts millions of people globally. Regular vision screening is the recommended best strategy to ensure timely diagnosis and treatment, however, many people do not have access to optometric care and a comprehensive vision examination is inaccessible to many people. There is therefore a need for fast, low-cost and easily-operate vision screening approaches. In this thesis, we aim to investigate the possibility of conducting photorefraction, a common vision screening procedure, on the mobile platform, to address the challenge.

Our approach exploits machine learning algorithms and computer vision techniques. Starting from principles from optometry and prior studies, we create several hand-crafted features corresponding detection methods. The experiment results indicate that our detection methods outperform contemporary approaches, leading to a better performance of refractive error measurement and amblyopia risk factor detection. We then move on to pre-trained features extracted by convolutional neural networks (CNN). We employ the convolutional layers from multiple pre-trained CNN models to encode features and train machine learning models to predict the refractive error. The experiments show promising results, even though the CNN models were not trained on photorefraction datasets.

Given these encouraging results, we further investigate the possibility of data augmentation. One of our challenges is that it is not possible to collect a large amount of data which is enough to train a well-performing CNN model from scratch. Therefore, we investigate the use of synthetic data for augmentation. We develop a model of the eye based on the principle of photorefraction, and use it to generate synthetic pupil images with pre-determined refractive errors. Evaluation results show that models trained on these synthetic pupil images can achieve similar per-

formance as real images on multiple experiments, which provides solid evidence for the correctness of our photorefraction model.

We finally apply transfer learning to solve the insufficient data issue. CNN models pre-trained on large-scale public image datasets are finetuned with photorefraction images and the experiments results show large improvement. The CNN models are then trained on more than 10,000 images of synthetic eyes generated via our eye model, and finetuned using real images, achieving performances that outperform all of the previous models. These results support the feasibility of the proposed photorefraction model, and provides a novel direction to obtain training data, which may be extensible to other similar domains.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

Refractive error refers to the error occurred when eyeball refracts the lights incorrectly that makes the patients see blurred image. The most common types of refractive error can be categorized to myopia (nearsightedness), hyperopia (farsightedness) and astigmatism (a mixture of various error in different meridians). A recent report from WHO indicates that refractive error is a primary cause of visual impairment, impacting 285 million people globally, of whom 39 million are blind [1]. It affects not only adults and teenagers, but also children and preverbal infants. In 2018, the refractive error prevalence among children ranged from 16.2% in South-East Asia to 59.7% in America [2]. In Hong Kong, the prevalence of myopia is 18.3% and 61.5% for 6 and 12-year-old children respectively [3]. Furthermore, some research indicates that, due to the changes in lifestyles, the prevalence of refractive error is increasing faster in this decade [4] [5]. For pre-school children, the uncorrected refractive error can lead to worse vision disabilities such as amblyopia, also know as lazy-eye, which is the leading cause of vision loss in kids, but it is almost always treatable if detected early. Frequent eye exams

are important, especially during the school years, as the degree of error changes frequently during this time [6], and uncorrected refractive error can significantly impact academic [7], physical and even mental [8] development. Early discovery and treatment is therefore recommended to improve treatment outcomes [9].

Although a comprehensive eye examination is the recommended best strategy to ensure proper diagnosis and treatment, the requirement of professional expertise is a significant barrier to regular diagnosis in daily life. Generally speaking, the process of an eye exam often consists of several complex steps from completing necessary health history forms to multiple refraction tests. Special eye drops are also utilized to ensure the pupil's dilation to allow a better view of the structure inside the eyeball. The testing process also requires devices including visual field machines, optometry tonometer, phoropter, autorefractor, etc., which are usually not accessible to normal users. Currently, the most accurate vision screening techniques rely on subjective refraction, which requires real-time feedback from patients. This may be difficult, especially for young children and infants. Photorefraction-based refractive error diagnosis, which requires only a single photo, is a better choice in these situations. However, most of these approaches require expensive devices, and trained personnel to operate them.

One critical challenge is the lack of professionals, especially in underdeveloped areas. A recent report [10] presents that there are only 17 established institutions offering optometry degree programs in Africa, 14 of which are fully accredited. The optometry manpower is extremely insufficient in such a continent with a population of more than one billion and 55 recognized countries. As a consequence, it is hard for people in these underdeveloped areas to get access to regular vision screening due to the severe situation of vision healthcare.

Up to this point, contemporary vision screening methods are not easily acces-

sible enough, especially for impoverished populations worldwide. The solution to these problems clearly has to address two issues: the financial (equipment) cost, and also the human cost, in terms of trained personnel.

The development of e-health offers an alternative approach. During the past few decades, e-health technology has been significantly developed and the applications bring strategic benefits including the improvement of the healthcare delivery quality, better access to up-to-date health information, the reduction of healthcare costs, etc. As one way to provide an efficient healthcare service, e-health tools are not only widely invested in developed countries, the benefits make it significantly relevant for developing countries [11–13].

As an almost ubiquitous device, using smartphones for clinical practice is quite attractive for healthcare [14–18]. Equipped with different kinds of sensors, modern smartphones are increasingly programmed with algorithms that pave the way for diagnosing diseases. Specifically, for vision screening, there are several hand-held systems including SVone, GoCheckkids, EyeNetra, etc. These systems simulate the process of traditional photorefraction that uses a external light source to spot patient's pupil, and diagnose the refractive error by some properties of the light reflection. Smartphones with embedded camera and flashlight naturally fit in the photorefraction settings. The flashlight is capable to be the light source, with the camera to capture the image for further diagnosis by optometrists or computer vision models.

The drawback of contemporary systems is that they all require an external lens to capture clear eye images, which may bring much inconvenience and make it almost impossible to utilize in underdeveloped areas. Few recent studies are proposed to address this challenge. Kwok et al. [19] introduce the data-driven approach based on a smartphone as the sole device to take photos. To calibrate the

sizes, users can employ a simply modified glasses frame while taken pictures. Nevertheless, even the glasses frame largely reduces the cost and significantly improves convenience, it still relies on external attachment for assistance. Another work introduced by Chun et al. [20] utilizes several neural networks trained with eye images captured by smartphone cameras to conduct pre-diagnosis of myopia and hyperopia. Since collecting suitable eye images and the corresponding optometry result is extremely time-consuming and expensive, the challenge of this work is the lack of enough data to train well-performing CNN models. Their solution is to apply pre-trained models as a start point of training and then finetune the models with a limited amount of photorefraction images. As a result, the performance of the proposed system still has room for improvement.

Our motivation is to build a democratized e-health vision screening system. The system should be easily operated, meanwhile maintaining satisfactory accuracy. We propose that the smartphone can be used to make vision screening accessible to untrained individuals. To achieve this, we require a method that does not depend on professional expertise and expensive machines. Therefore, our intention is eventually to develop a low-cost, machine-learning-based automatic refractive error detection approach that uses a smartphone as the primary interaction device.

1.2 Study Overview

We establish our study by investigating the feasibility and efficiency of analysing smartphone images via machine learning algorithms for vision screening. Inspired by previous work [19], we follow the optometry principle to design hand-crafted features based on the size of eye structures. When there is a single flashlight and a proper distance between camera and eyes, the pictures will show a bright reflection region inside the pupil of myopic or hyperopic eyes. The optometry principle tells

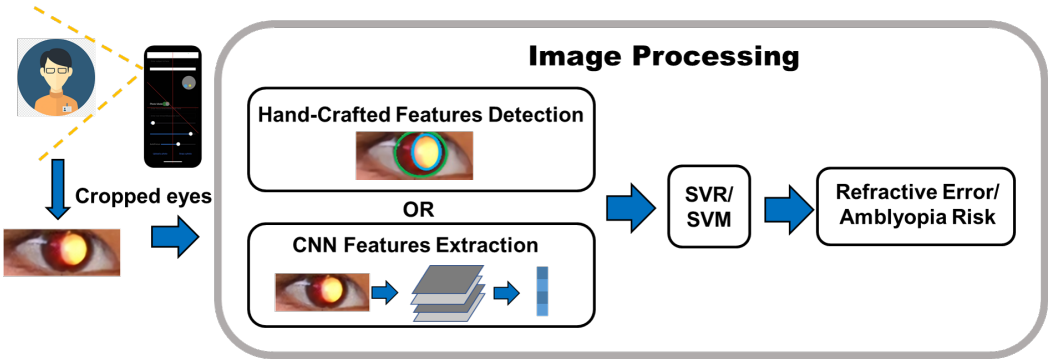


Figure 1.1: The overview of the proposed photorefraction system via smartphone images. This thesis focuses on image processing for eye images to conduct vision screening.

us that the refractive error can be measured by the sizes of the bright area on the retina (which is called crescent) and other related structures.

When a smartphone is used to capture the image, the non-optimal image quality brings challenges to the precision of estimating feature sizes. We propose multiple detection techniques which are low-cost and outperform the contemporary methods. In addition, we exploit the implementation of traditional machine learning algorithms to predict refractive error and detect amblyopia risk through the detected features. In order to further improve the convenience, we explore the possibility of reducing the need for add-on devices. The previous method requires the use of a calibrated pair of lens-less glasses (e.g. just the frame) that is used to measure the sizes of the eye structure in images. We therefore propose a novel method to perform this calibration through the iris size, which means that the smartphone is the only device that is required.

In addition to the features guided by the optometry principle, we also explore the use of convolutional neural networks (CNNs) to extract more abstract features. CNN has been shown to be a powerful tool to encode image information. The features extracted by CNN are often hard to interpret but are able to achieve better

performance than hand-crafted features. Several pre-trained CNN models and the utilization of the extracted features are investigated.

Training a well-performing CNN model often requires large-scale data. However, it is almost impossible to collect a large enough dataset for our problem because of its cost. To address this challenge, we develop a photorefraction model based on the optical principle. We start from simplifying the eyes, camera and flashlight into one optical geometry framework to calculate the route of light reflection and refraction to generate the crescent. The proposed photorefraction model makes it possible to generate synthetic eye images with determined refractive error and pupil size.

Finally, we aim to CNN training itself. The previous results have shown the potential of the CNN for analysing smartphone images for photorefraction. Since we are faced with insufficiency of data, we explore transfer learning, which is a widely used paradigm to deal with insufficient data issues especially in the medical image processing area. We implement transfer learning from other large-scale datasets to the photorefraction domain, and the experiment shows promising results.

Although we cannot collect enough real eye images, the proposed photorefraction model is able to generate a large-scale dataset of synthetic photorefraction images. We managed a synthetic dataset containing more than 10,000 images with refractive error ranging from 0D to -10.0D. The corresponding optometry results for the synthetic eyes include sphere error, cylinder error and astigmatism axis. The synthetic dataset fills the refractive error gap of real dataset with its variety.

We then implement the synthetic dataset on transfer learning to transfer the knowledge from the synthetic photorefraction dataset to real photorefraction dataset. Similar to the previous approach, the CNN models are firstly pre-trained on synthetic datasets to gain the ability to encode image information in synthetic eyes.

The results reveal that the model pre-trained on synthetic images outperforms currently published state of the art performance. From the view of transfer learning, this result suggests that our synthetic photorefraction domain is close to the real photorefraction domain, in the sense that the CNN models pre-trained on the synthetic dataset are a better start point model compared with the previous models. In addition, this result suggests our proposed photorefraction model provides a direction to generalize our smartphone vision screening system to other devices without collecting a large amount of data.

The contributions of this thesis are the following:

- Propose low-cost feature detection methods that achieve state-of-the-art performance on iris detection and crescent detection on images taken by a smartphone for photorefraction.
- Propose a novel calibration technique based on eye structure, making the smartphone photorefraction system totally free of external devices.
- Develop a photorefraction model through optometry principle to generate synthetic pupil with the crescent on refractive error.
- Investigate the use of transfer learning on this task, including transfer learning from commonly used large scale image dataset ImageNet and the synthetic photorefraction dataset generated by the proposed model to real photorefraction dataset.

The remaining chapters of this thesis will cover the following:

Chapter 2 provides the literature reviews from the e-health studies, photorefraction to smartphone photorefraction and transfer learning. It covers the implementation of e-health on chronic diseases, mental health and the efforts during the

COVID-19 pandemic. Then we extend the content to photorefraction and the combination of e-health and vision screening. We then introduce transfer learning and its applications in medical image processing.

Chapter 3 introduces the construction of the smartphone photorefraction dataset. We collect more than 4,000 photorefraction eye images using two common smartphones: Microsoft Lumia 950 XL and iPhone X. In addition, we conduct comprehensive vision screening for the subjects and obtain the optometry results. The collection was conducted under various scenarios including universities, primary schools and hospitals. To the best of our knowledge and belief, it is the largest smartphone photorefraction dataset so far.

Chapter 4 presents the implementation of machine learning algorithms. We proposed two types of features which are hand-crafted and CNN extracted respectively with their detection methods. The experiment results indicate that the proposed hand-crafted features are useful to measure refractive error and outperform contemporary studies. This chapter also demonstrates a technique making the smartphone photorefraction system free of external devices. We propose a novel calibration method through eye structure. The experiment results reveal that the system can totally relieve external devices with no loss on performance.

Chapter 5 demonstrates the implementation of traditional data augmentation approaches on the photorefraction dataset. We utilize several geometric transformations and noise injection to generate more training samples without disturbing the critical information related to label refractive error. The experimental results illustrate that after data augmentation, the CNN models are capable to encode efficient features and achieve higher performance with the model pipelining.

Chapter 6 intends to develop a photorefraction model to generate synthetic eye images as preparation for the following transfer learning. We summarize the proce-

dure of photorefraction into a geometrical optical framework. We conduct experiments with the synthetic pupils to evaluate the correctness of the proposed model. The results suggest our synthetic pupils perform well in multiple tasks. Through the photorefraction model, we are able to generate a large-scale dataset for the following transfer learning approaches. We then exploit the data augmentation through the synthetic photorefraction dataset. We investigate the effect of data amount and data distribution. The results reveal that pre-training CNN models on synthetic images significantly improves performance compared with the traditional data augmentation. In addition, the synthetic dataset with uniform diopter distribution outperforms the other original diopter distribution and achieves the state-of-the-art accuracy of refractive error estimation, which provides a deeper understanding on this task.

Chapter 7 summarizes our contributions and limitations of this thesis.

Chapter 2

Literature Review

This thesis aims to investigate photorefraction vision screening by processing smartphone images. To this end, we utilize the embedded camera and flashlight to simulate the manual photorefraction process. This chapter presents a review of the development of traditional photorefraction vision screening and the recent studies on vision screening by the images captured by smartphones. As insufficient data is a common issue when developing data-driven models in previous studies on this task, this chapter also reviews the transfer learning technique and how its implementations address this problem in the computer-aided healthcare area. Based on that, this chapter outlines the rationales for the proposed studies.

2.1 Photorefraction Vision Screening

As a fast and accurate objective technique for measuring the refractive error of the eye, photorefraction was first introduced by Kaakinen and Molteno et al. to determine amblyopia risk factors in 1979. They measure refractive status and accommodation of the eye by interpreting the fundus reflex (the crescent) with flash light [21, 22]. The early photorefraction with white light was used to look at ac-

commodation in owls but not for measuring refractive error since the constricted pupil is not feasible for an optometry. With the advent of infrared light-emitting diodes that will not make pupil constrict, it becomes possible to measure refractive error for human [23]. At this stage, photorefraction relied on a single flashlight. The theory behind photorefraction is shown in Figure 2.1. A light source is located eccentrically to the camera lens. The camera is focused on the pupils of the subject so that a crescent reflex is seen in the eye. When there are multiple light sources, an overlapped crescent with gradually changing intensity is shown. Based on the principle, an analytical description of reflex can be determined by the following parameters: refractive error of the eye, the distance of the eye to the camera, pupil size, and eccentricity. With all the parameters above, the refractive error of an eye should be revealed.

The accuracy of photorefraction is then validated by Atkinson et al. used isotropic photorefraction to screen 1096 infants aged 6–9 months in the City of Cambridge, and 5% of them were found to have large hypermetropic errors. This result was soon confirmed by retinoscopy screening, which is the most accurate measurement of refractive error. In 1985, Bobier et al. and Howland provided formulations of the refractive state of eyes, and named this technique *Eccentric Photorefraction* [24, 25].

Since the working range of a single external light source is limited by the fixed eccentricity (the distance from camera to flashlight), Schaeffel et al. modified the system with multiple eccentric light sources, which enabled the method to measure larger diopter ranges [26]. After that, using the fact that the slope of the intensity distribution across the pupil varied linearly, Scheaffel improved the method by measuring the intensity gradient of the reflex, which is proved to be a more accurate indicator [27] for refractive error. However, in this study we tend to conduct

photorefraction with smartphone, which usually has no such designed multiple eccentric flashlights. Therefore the theory of single flashlight photorefraction is implemented.

In addition to the manual photorefraction vision screening, in 1990, Uozato developed a Topcon PR-1000 photorefractor with multiple infrared light sources for vision screening of infants [28] by early stage computer vision techniques. And mainly because it is possible to do rapid screening and does not require the patient to understand language or provide feedback, computer-aided automatic photorefraction became widely used in vision screening for children and infants [29]. As a consequence, some commercial devices with the similar machenism like *Power Refractor* have been developed. Nevertheless, in practice, some optical characteristics of the eye will affect the calibration of instruments, including reflectance of the retina, the distance between the retina and cornea, and higher-order monochromatic aberrations [29, 30]. Therefore, each instrument model requires a specific calibration adjustment, which further increases the cost and complexity. [31]. Also, these devices often require external light sources and cameras, leading to high prices and inconvenience during use.

2.2 Vision Screening on Smartphone Images

Smartphones have become essential devices in daily life. With the performance improving rapidly and more kinds of embedded sensors being incorporated, there has been more health care research focusing on leveraging medical services to mobile devices [32–35]. Compared with conventional health care tools, smartphones can perform more timely and regular measurements of physical signals. Newer smartphones also possess enough computing power to deal with these signals or medical images locally, which allows smartphones to perform remote diagnostics

in real-time.

If we narrow the scope to vision screening, there are several commercial products to leverage vision screening with mobile devices as well. *SVOne* developed by Smart Vision lab, uses a handheld structure clipped onto the iPhone is used to capture the eye images. Nevertheless, SVOne is expensive ($3,950) because of the extra lens and requires professionals to operate. Another portable, smartphone-based autorefractor named *EyeNetra* consists of a smartphone, handheld phoropter, and a digital Lensometer to assist the optometry. The price of this set of refraction kits is $2,995. The high price of the commercial devices makes them not suitable for daily vision screening for normal users, especially for people in underdeveloped areas.

Previous studies investigate the feasibility of leveraging smartphones images to vision screening. Kwok et al. [19] firstly present a smartphone-based photorefraction vision screening approach. Inspired by the principle of photorefraction, they utilize the flashlight and embedded camera to conduct the photorefraction procedure. Towards building a data-driven approach to measure refractive error, several efficient hand-crafted features and the corresponding detection methods are demonstrated. Machine learning models are trained with the proposed features to process the captured eye images. The experiments show promising results on refractive measurement, which reveal the feasibility of a smartphone to conduct vision screening without professional devices. However, the demonstrated models highly rely on the efficiency of the selected hand-crafted features. Also, the proposed feature detection methods are easily affected by the noise of the image, which is hard to avoid in real applications.

Instead of hand-crafted features, another related work introduced by Chun et al. [20] investigated the implementation of deep learning models on refractive error

range classification. They collected 300 eye images and divided them into 5 classes according to their refractive error. However, the size of the dataset is not enough to train a well-performing CNN model. To address this issue, they applied several pre-trained CNN models and fine-tuned them with real eye images. The experiment results of fine-tuned CNN models show encouraging performance on the specific classification task and indicate the feasibility of CNN models on this task even with insufficient data amount. However, there still lacks a comprehensive investigation of how to build well-performing CNN models for this task in this study. Also, classifying the refractive error range is different from accurate measurement, where the latter is our actual ultimate goal.

As there is no large-scale dataset on smartphone image vision screening task, our previous work [38] aimed to address the insufficient data problem by exploiting active learning techniques to reduce the manual effort on image annotating. In this study, all eye images are assigned an *information score* computed by a model committee. Instead of the whole dataset, only the images with high information scores are selected to train models. The experimental results show that after applying active learning, we can reduce the amount of data by 18% without a performance drop. These previous studies have demonstrated the feasibility of several image processing methods and the implementation of transfer learning, active learning techniques on the smartphone image vision screening task.

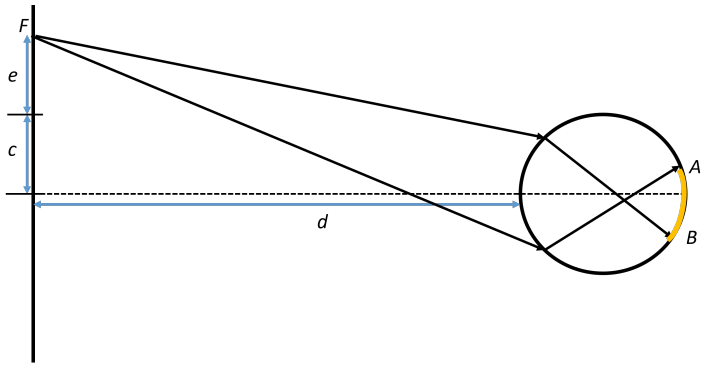
2.3 Transfer Learning

Existing research has shown the performance of deep learning networks is highly correlated with the amount of data. [39]. To train well-performing deep learning models, there are several developed large-scale databases such as ImageNet, PASCAL, VOC, MS COCO, etc. These databases contain millions of manually

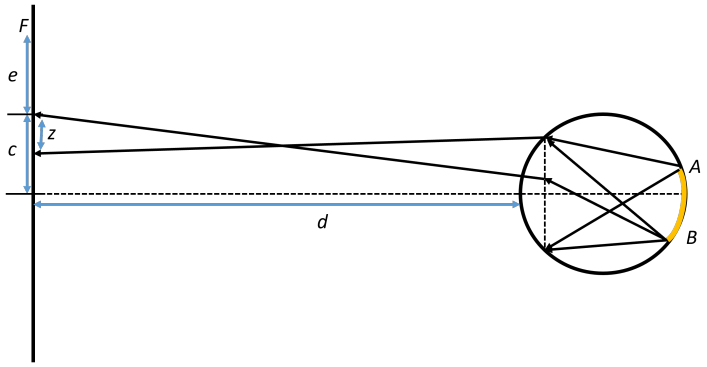
annotated images for classification, object detection, semantic segmentation tasks. However, not all the research could collect enough image instances for their specific task. Transfer learning is then proposed to address this challenge and has been widely utilized in training machine learning models.

Transfer learning can be roughly categorized into three types: 1) instance-based transfer learning: measure and weight the similarity of instances in the source domain and target domain. 2) feature-based transfer learning: measure and project features of the source domain and target domain to a higher space where the distance of them are closer. 3) parameter-based transfer learning: transfer the knowledge learned on a source domain to the target domain on a parameter level by sharing the parameters of the source domain model with the target domain model.

Since instance-based and feature-based transfer learning is hard to realize on deep learning models because of the scale of the dataset and the abstract features, parameter-based methods have become the most common way to utilize transfer learning on convolutional neural networks (CNN). As a typical parameter sharing method, fine-tuning is widely used in CNN transfer learning, especially in the computer vision area. Many works on image processing implement pre-trained CNN models which are able to extract general image features, and then fine-tune them on other specific projects. Fine-tuning is also the most common way to conduct transfer learning on vision screening via smartphone images. As mentioned above, pre-trained models [20, 38] are usually employed to fine-tune with smartphone eye image performance. The experimental results from previous studies also indicate the feasibility of fine-tuning on smartphone images vision screening tasks.



(a) Forming blurred image on retina



(b) Manifesting a bright area (z) on the picture

Figure 2.1: (a) A flashlight is located at a distance *e* above the camera. *d* is the distance from the camera to the eye being tested. Light rays enter the myopic eye and are refracted by lens. The refracted light rays focus in front of the retina and forms an blurred image AB. (b) If the eye is myopic, the light returning from this image will enter the camera and manifests as a bright area (*z*) on the captured picture. The photograph of the eye shows a bright crescent on the same side as the flashlight.

(a) (b)

Figure 2.2: (a) Device for *SVOne* (Source: [36]) (b) Refraction Kits for *EyeNetra* (Source: [37])

Chapter 3

Dataset Construction

In order to build up and evaluate the data-driven smartphone images-based vision screening system, we develop a dataset containing eye images that are taken by smartphones and possibly to be analyzed for vision screening. On our study, our data collection is carried out with two common smartphones: 1) Microsoft Lumia 950 XL and 2) iPhone X.

According to the principle of photorefraction, an eye's refractive error can be computed on the size of its pupil and crescent area. Since the resolution of a smartphone camera is limited, better effects can be observed if patients' pupils are dilated enough such that an observable crescent can be generated. To this end, we collected all the images under a dim illumination environment to make sure subjects' pupils were dilated as much as possible. Our experiments can be replicated in any dark environment such as a room with the light off and/or curtains drawn.

The distance between the eyes and the smartphone embedded camera also needs to be carefully deliberated. According to the principle of photorefraction, the lowest measurable refractive error is $\frac{1}{distance}$ diopter, where diopter is the unit of refractive error. As a result, a larger distance from the eye to the camera can bring a larger measurable refractive error range. On the other hand, unlike the single reflex

camera used in traditional photorefraction vision screening, it is hard for the smartphone embedded camera to capture high-quality images at a distance, as a larger distance means a smaller eye area, leading to lower image quality, which would affect the performance of the following image processing methods. Therefore, there is a trade-off between image quality and measurable refractive error range. We conducted several testing and found that 1 meter is a good compromise for our experiments, which can provide both acceptable image quality and a relatively large measurable refractive range.

During data collection, subjects are asked to stand or sit 1 meter away from the smartphone camera, and asked to look directly at it without blinking (as much as they can) while the photos are taken.

A perfect eyeball is completely round and symmetrical. However, for people with astigmatism, the eyeball is shaped more like an football that is tilted at some angle to the horizontal axis, leading to the multiple focus points of the refracted lights as shown in Figure 3.1. Standard optometry practice models the eyeball as a composition of a *Sphere* and a *Cylinder*, which is tilted at a particular *Axis*. The refractive error is thus recorded using three measures: the sphere error, cylinder error, and the astigmatism axis.

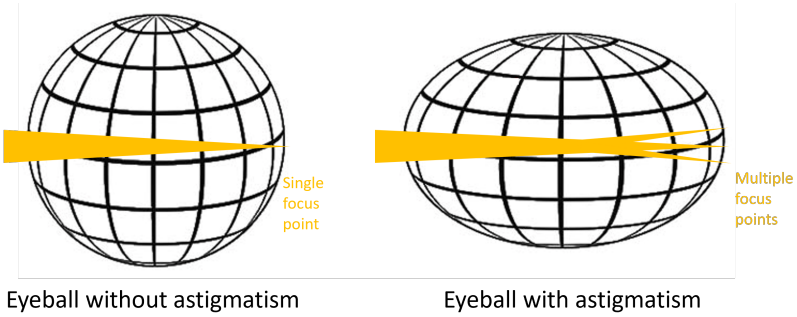


Figure 3.1: The difference of perfect eyeball and the eyeball with astigmatism

The gold standard for each subject is the optometry report with the sphere error,

cylinder error, and astigmatism axis. Nevertheless, to calculate all of the three measurements, we need the pictures captured at three meridians. With one picture in a specific meridian, only the refractive error at that meridian can be measured. We thus reduce these three measurements into one number and calculate the ground truth refractive error of the eye image as follows:

$$X = S + C \times \sin(A - m)^2 \tag{3.1}$$

where S refers to the sphere error, C refers to the cylinder error, A is the astigmatism axis and m is which meridian in which the crescent is located. The meridian here is the angle between the axis connecting the light source and the camera, and the horizontal axis, as shown in Figure 3.2.

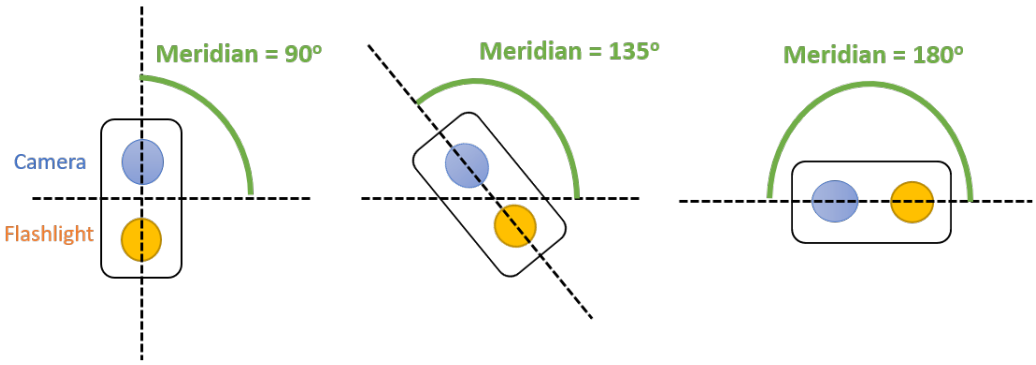


Figure 3.2: The meridian as defined in photorefraction.

There are two datasets taken by smartphones, which are noted as Dataset-2015 and Dataset-2020 according to the year during which the datasets are constructed:

3.1 Dataset-2015

The Dataset-2015 was collected and investigated by the previous study [19] in 2015. The device used to collect images for the Dataset-2015 is a Microsoft Lumia 950 XL, of which the eccentricity (distance from flashlight to camera) is 10mm.

The embedded camera and flashlight are utilized to simulate the photorefraction vision screening process. Images are taken by the default camera program of the operating system.

The experiments are conducted in a classroom with curtains down during the daytime. The smartphone is located at the same horizontal line as the subject's eyes. They took the eye images with the portrait orientation, which means the flashlight is on the right side of camera (from photographer's view). As a result, for myopic eyes the crescent will be located at the right side of pupil. Meanwhile, for the sake of calibration for later processing, subjects were asked to wear a glasses frame with two green labels as shown in Figure 3.3. The distance of the two labels is fixed at 10cm, which will be used as a scale to calculate actual sizes of eye structure from their pixels amount in images.

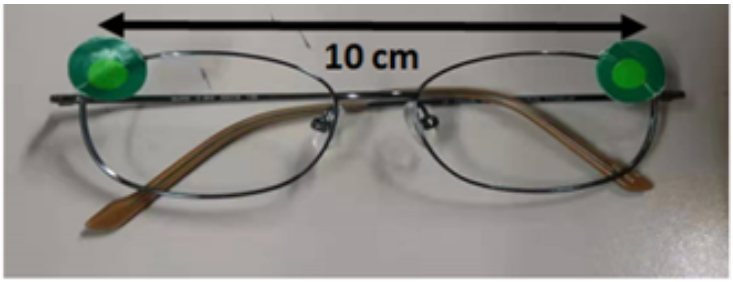


Figure 3.3: The modified glasses frame with two green labels. The distance between two labels is 10cm, which is used as scale to compute actual size.

The gold standard refractive error measurement for the subjects is measured by an open-field autorefractor: Grand Seiko WAM-5500, whose effectiveness has been demonstrated in previous clinical studies [40]. The eye structures including iris, pupil and crescent are annotated manually.

97 primary school students with myopia were recruited as the subjects for Dataset-2015 (age:11.34$\pm$0.96 years). Informed consent was obtained before data collection. For each subject, they take one picture and crop two eye images. The

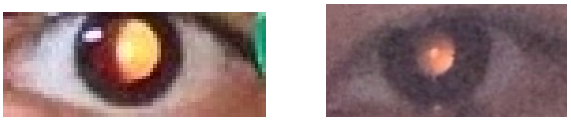
final Dataset-2015 contains 194 eye images in total.

3.2 Dataset-2020

The Dataset-2020 was collected by the collaboration with Dr. Chi-wai Do and Lily Chan from the School of Optometry. The staff of Tianjin Medical University Eye Hospital recruited the subjects and provided the optometry equipment. We use an updated iPhone X to capture the images for Dataset-2020. The embedded camera and the flashlight are applied to simulate the photorefraction vision screening. The same image-taking procedure as Dataset-2015 was conducted with 203 subjects in total (116 female), involving patients from different age groups (7 to 30 years old). Informed consent is obtained before the data collection. The data collection for 82 of the subjects takes place in a dark office with curtains down and the remaining 121 subjects in a totally dark room where no illumination exists, except the flashlight on the smartphone. The subjects sit 1 meter away from the smartphone and wear a glasses frame for calibration. We take multiple images for each subject and crop the eye area, giving us 5483 eye images for Dataset-2020. An experienced optometrist measured subjects' refractive errors using a state-of-the-art autorefractor. The eye structures on images including iris, pupil and crescent are annotated manually.

Unlike the Lumia 950 XL which uses a single LED light for the flash, the iPhone X flashlight is composed of 4 tiny LEDs arranged in a square configuration. As a consequence, the resulting crescent will be composed of 4 overlapping crescents, which creates a larger crescent with blurred boundaries. The eccentricity of the iPhone X is 6mm, which is much smaller than the eccentricity of the Lumia 950 XL. According to the optometry principle, given the same pupil size and refractive error, the size of crescents in Dataset-2020 will be larger than those in Dataset-2015.

Image Samples from Dataset-2015 and Dataset-2020



(a) (b)

Figure 3.4: (a) The eye image captured by Lumia 950XL; (b) the eye image captured by iPhone X. The boundary of crescent is more clear in Dataset-2015 compared with the crescent taken by iPhone X.

3.3 Data Cleaning

After collecting the eye images, we manually filter Dataset-2015 and Dataset-2020 to ensure their quality is acceptable. Since we focus on estimating the refractive error for myopia, the eye images with positive refractive error (hyperopic) are not used. Besides, images are removed from the datasets under these four conditions:

1. image is blurred due to misuse of the smartphone (e.g. hand tremor, out of focus, etc)

2. subject did not look at the camera,

3. subject blinked or otherwise closed their eyes during the photo, resulting in only a small area of the eye being exposed,

4. pupil is extremely contracted (usually caused by operation error, such as turning on a light accidentally or taking the photo too soon after the subject walks into the darkened room)

After the filtering, we have 179 images for Dataset-2015, and the refractive error ranges from 0D (normal eye) to -8.5D as shown in Figure 3.5. The Dataset-2020 contains 4872 images, whose refractive error range from 0D to -10.0D as shown in Figure 3.6. Table 3.1 depicts the statistics of the two datasets.

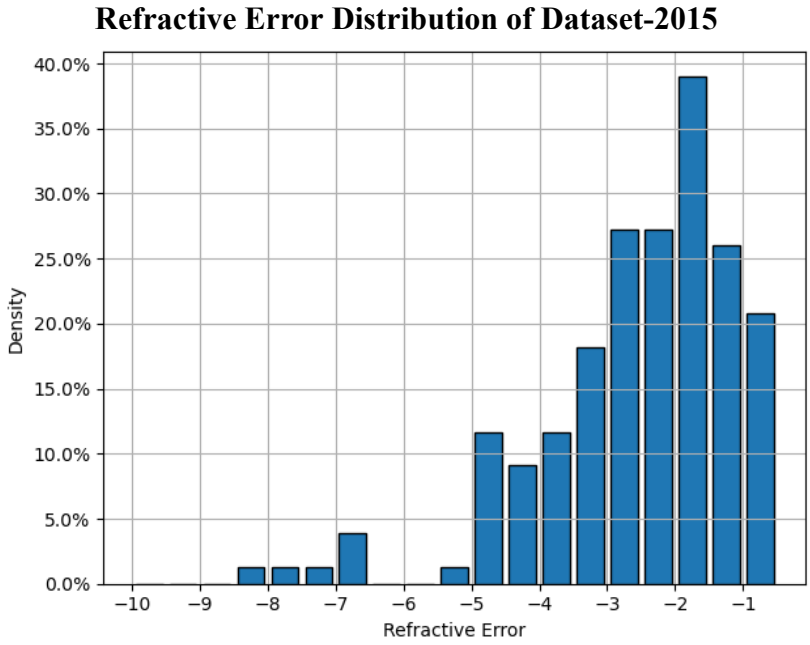


Figure 3.5: The distribution of refractive error in diopter for the Dataset-2015. Lower value refers to more severe myopia.

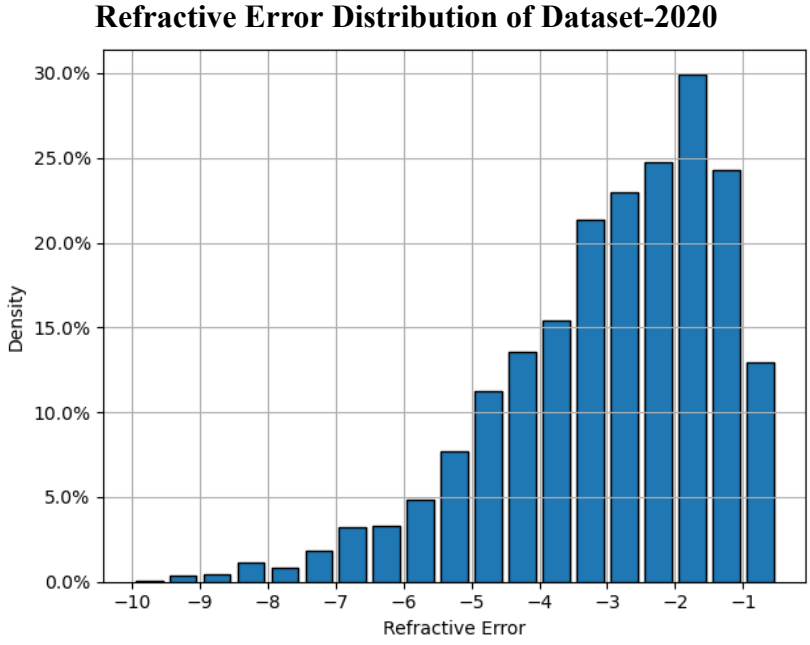


Figure 3.6: The distribution of refractive error in diopter for Dataset-2020. Lower value refers to more severe myopia.

Statistics	Dataset	
	Dataset-2015	Dataset-2020
	Refractive error	Refractive error
Mean	-2.20D	-2.67D
Standard deviation	1.58	1.87
Range	[-8.5D , 0D]	[-10.0D , 0D]

Table 3.1: Statistics of the Dataset-2015 and Dataset-2020

Chapter 4

Machine Learning on Smartphone Images for Photorefraction

4.1 Hand-Crafted Features

Our study starts by investigating the implementations of a machine learning algorithm on photorefraction vision screening via smartphone images. Inspired by the principle of photorefraction, the refractive error can be computed by the size of crescent and the other eye structures. So prior to the final refractive error measurement, we first attempt to extract features from eye structures, which are referred to as ”Hand-crafted features” including the sizes of iris, pupil and crescent shown in Figure 4.1.

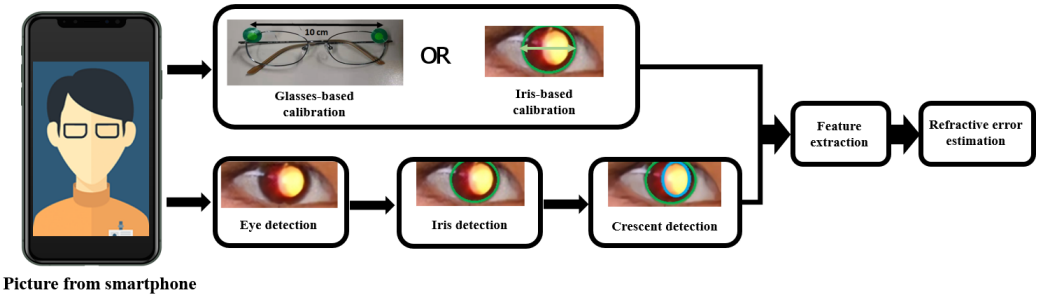


Figure 4.2: An overview of the approach with hand-crafted features

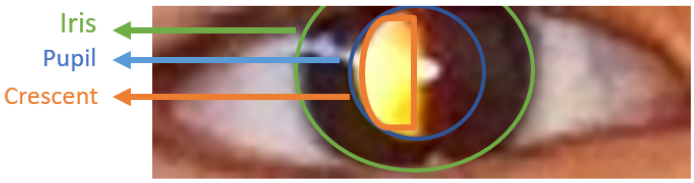


Figure 4.1: the hand-crafted features. The refractive error can be computed by the pupil size and crescent size, where the pupil size is measured in by the detected iris and crescent.

The overall feature detection tasks consist of iris detection and crescent detection, and the pupil size will be computed by the results of these two detections. Figure 4.2 illustrates our approach. From a myopic eye image, we extract the iris and crescent areas. Then we will compute the actual size of pupil radius, crescent width, etc. These features will be provided to a Support Vector Regression (SVR) [41] and Support Vector Machine (SVM) [42] engine for training together with the ground truth refractive error.

4.1.1 Iris Detection

Iris segmentation is widely performed in biometrics studies [43]. The most common and effective approach is to look for the area that most closely resembles a circle or ellipse in eye images and fit the corresponding shape to the area. Depend-

ing on how the fitting is performed these methods can be summarized as:

- Daugman's operator-based [44–46] that searches for the circle whose boundary has maximum gradient in intensity, in the 3-d space of coordinate of the center (x, y) and the radius r. The operator is applied iteratively to obtain accurate localization.

- Hough transform-based [47–50], which also conducted the searching in 3-d spcae of the three parameters of a circle, and finds the circle intersecting the most edge pixels.

- RANSAC-based [51, 52] approaches, which traverses all the combination of edge pixels to fit a circle intersecting with the most pixels.

Nevertheless, contemporary iris detection methods easily fail in our task due to two major challenges. The first is the reflex crescent. For instance, Daugman's operator-based methods find the most iris-like circle following the integral-differential of intensity. This can be significantly affected by the crescent due to the dramatic intensity change along its edge (Fig 4.3). The other challenge is brought by image quality. To induce the crescent and to stay within our constraint of using only low-cost devices, the eye images have to be taken with a smartphone camera at a distance, as well in low-light conditions. These settings result in relatively low quality and low resolution compared with eye images in existing common-used database for biometrics, which are often taken by single-lens reflex cameras. For example, the image resolution in our dataset is about 90×50, which is much lower than those in the SDUMLA-HMT [53] and CASIA v4-lamp [54] datasets, which are 768×576 and 640×480 respectively. In addition, the lower resolution leads to more noise, which affects the performance of the Hough transform-based methods (Fig 4.3).

(a) (b) (c)

Figure 4.3: Iris detection by Daugman's operator-based (a), Hough transform-based (b) and (c) proposed RANSAC-based method. Daugman's method can be misled by the crescent edge and Hough methods affected by outlier points

We therefore propose an iris detection method based on RAndom SAmple Consensus (RANSAC), which is more robust to the outlier points that frequently appear in our relatively low-quality eye images. The proposed method can be roughly divided into three steps:

The first step is to detect a coarse edge of iris according to the brightness difference between the sclera and iris limbus. For a given cropped eye image, we set a Cartesian coordinate system whose origin is located at the bottom left corner. The unit of coordinate is the pixel, i.e., a pixel at the third column(counted from left) and the fifth row(counted from the bottom) has a coordinate of (3,5). Then this pixel can be represented by $p_{3,5}$. We search for Right and Left Iris Edge Pixels as follows:

Definition 4.1.1 *A pixel* $p_{xr,y}$ *is an* Right Iris Edge Pixel *if* $\forall\ x$ *s.t.* $I(p_{x,y}) - I(p_{x-1,y}) > threshold$, $xr \geq x$.

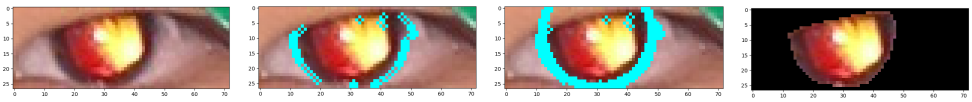
Definition 4.1.2 *A pixel* $p_{xl,y}$ *is an* Left Iris Edge Pixel *if* $\forall\, x$ *s.t.* $I(p_{x,y}) - I(p_{x+1,y}) > threshold$, $xl \leq x$.

where $I(p_{x,y})$ is the intensity (brightness) of the pixel located at (x, y) in the given image. This gives us one Right Iris Edge Pixel and one Left Iris Edge Pixel in each row (if such pixels exist). For denoising, we delete candidate pairs where $xr \leq xl$,

which indicates that these two detected pixels are not reliable. In the experiments we set the intensity threshold as 25 to get the optimal performance.

Then in the second step, we use RANSAC method to fit the most iris-like circle based on the filtered edge pixels.

Our final step is to improve our iris segmentation based on the obtained iris-like circle. The fitted circle may contain part of the eyelid and/or sclera, etc. We, therefore, apply Otsu's threshold [55] to delete these noisy areas, keeping only the iris area. Figure 4.4 illustrates the our whole iris detection process.



(a) (b) (c) (d)

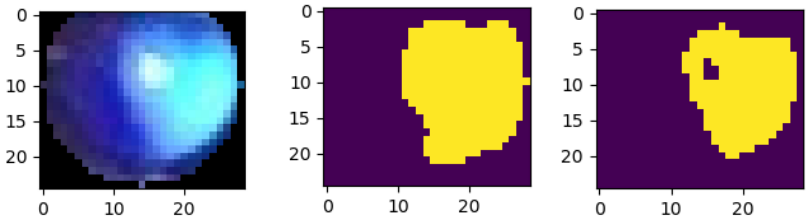
Figure 4.4: The iris detection process. (a) The original eye images. (b) Candidate Iris Edge Pixels are identified in every row. (c) fit a iris-like circle through RANSAC with detected Iris Edge Pixels. (d) Otsu's threshold is used to remove noise, leaving only the iris area.

4.1.2 Crescent Detection

The state-of-the-art approach [19] detects the crescent using an intensity (brightness) threshold. However, the best threshold may vary among the eye images due to the different reflecting ability among individuals, which lead to low accuracy of crescent detection.

To address this challenge, we propose a data-driven approach to detect crescent by classifying whether the pixels are in crescent. Firstly, all the pixels in iris region are annotated as Crescent or Non-Crescent. We then train a Support Vector Machine (SVM) as a pixel-level crescent detection model with the following features: (1)

distance from iris center; (2) intensity value in R, G, B channel; and (3) intensity gradient. This model thus labels all pixels in a target image as Crescent or Non-Crescent. We then select the largest contiguous Crescent cluster, which is further fitted to an ellipse by RANSAC. The width of the crescent is thus the length of the minor axis of the fitted ellipse. Then the radius of the pupil is computed as the largest distance between the iris center and pixels within the detected crescent area. An example of crescent detection in cropped iris is shown in Figure 4.5.



(a) (b) (c)

Figure 4.5: (a) The cropped iris image. (b) The manually annotated crescent. (c) Crescent detected by proposed method.

4.1.3 Refractive Error Estimation

After detecting iris and crescent from eye images, we are able to extract size-based features to build the refractive error detection model. The conventional photorefraction theory uses the crescent width, pupil radius, and iris radius as parameters for refractive error estimation. However, since the distance from eyes to the camera may not always be exactly 1 meter in practice, the number of pixels may not represent the actual size of the detected crescent, pupil or iris.

To address this issue, we propose to obtain the actual size of features using external calibration tool. Subjects are asked to wear a pair of lens-less glasses with two green labels separated by 10 cm (Figure 4.2). These two labels function as a calibration reference. Given the number of pixels in the line segment between

the centers of the labels and in the iris, pupil and crescent, it is then possible to determine the actual sizes of these eye features.

In addition, we observe that the brightness of the crescent and the proportion taken up by the crescent within the iris are also indicative of the refractive error. Therefore, besides the size features, the *sum of the intensity* values within the detected crescent area and the *ratio* of crescent width to pupil radius are also computed as features. Our final set of features for each eye are, therefore:

1. iris radius (Ir)
2. pupil radius (r)
3. crescent width (z)
4. Sum of Intensity (SoI)
5. ratio of crescent width to pupil radius (Ra)

Some of the features are illustrated in Figure 4.6. To make the model robust to outliers, we select Support Vector Regression (SVR) and train it on these features to estimate refractive error. The performance of this model and the contribution of different combinations of these features will be presented in the next section.

4.1.4 Results

Performance of Iris Detection

We adopt two performance metrics to evaluate the performance of our iris detection method. The first one is intersection-over-union (IoU). We regard a detected iris as correct, if the IoU of the labeled iris area and detected iris area is greater than a certain threshold. The second metric is mean-error-rate, which is computed as

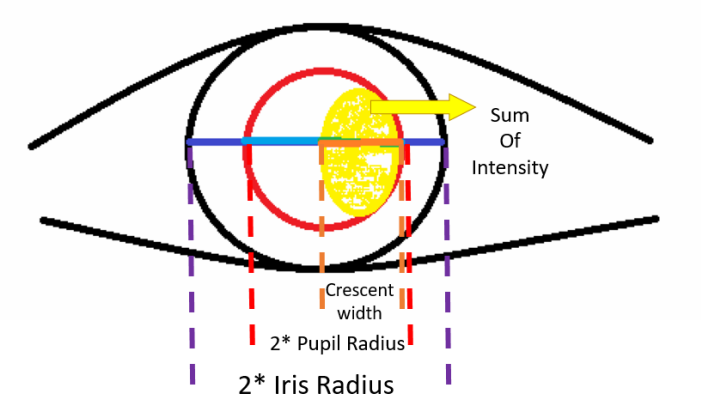


Figure 4.6: The captured features for SVR: (1) crescent width. (2) pupil radius. (3) iris radius. (4) SoI

follows:

$$e = \frac{1}{N \times w \times h} \sum_{x \in w} \sum_{y \in h} T(x, y) \oplus M(x, y) \tag{4.1}$$

where N is the total number of images, w and h are the width and height of one image, T and M are the labeled and detected iris respectively. The symbol $\oplus$ represents an exclusive OR operation to identify the segmentation error.

Figure 4.7 shows the performance of the proposed method where the Area Under the Curve (AUC) is 0.85, against both Daugman's (AUC=0.77) and Hough method (AUC=0.62). It can be seen that our method outperforms both competitors for all thresholds. Table 4.1 corroborates this finding with the mean error rate of iris detection, which is consistent with IoU. This indicates that our proposed iris detection method is more effective than the current state-of-the-art.

It is observed that both Daugman's Operator and Hough Transforming are heavily affected by the presence of crescent. As a result, the performance of these two techniques are not satisfactory for most of photorefraction eye images with myopia. We conduct the same iris detection experiments on the eye images without crescent, whose results are shown in Table 4.2. It can be seen that on normal eye images, the Daugman's Operator and Hough Transforming are able to achieve the

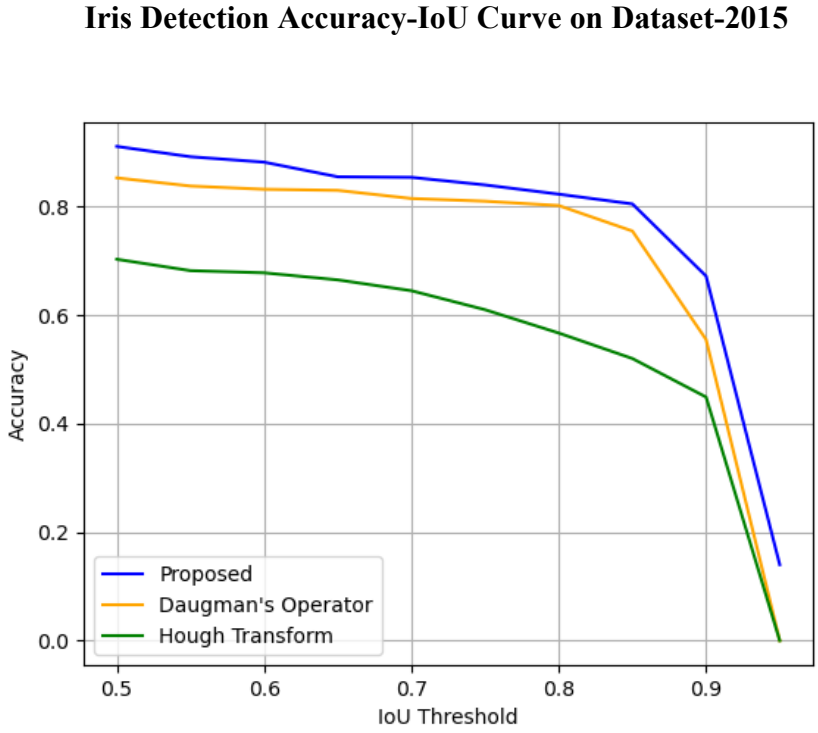


Figure 4.7: Accuracy along the IoU threshold of three methods. The proposed method achieve higher accuracy than the others under all the IoU threshold

close performance to the proposed method. This result again reveals that for photorefraction images, the proposed approach is more robust when crescent exists. It is also can be explained by the mechanics of the proposed method, which detect iris edge from the outside to avoid the various brightness distribution inside the iris.

Performance of Crescent Detection

We evaluate crescent detection on the images which exhibit a crescent on Dataset-2015 and Dataset-2020. For comparison, we also evaluate the performance of the state-of-the-art crescent detection approach [19] and the proposed method by the same metrics as iris detection. The proposed crescent detection method gets AUC=0.70, while state-of-the-art [19] gets AUC=0.32, which is also consistent with the mean error rate metrics as shown in Table 4.3. Both the results show that the proposed approach outperforms the state-of-the-art.

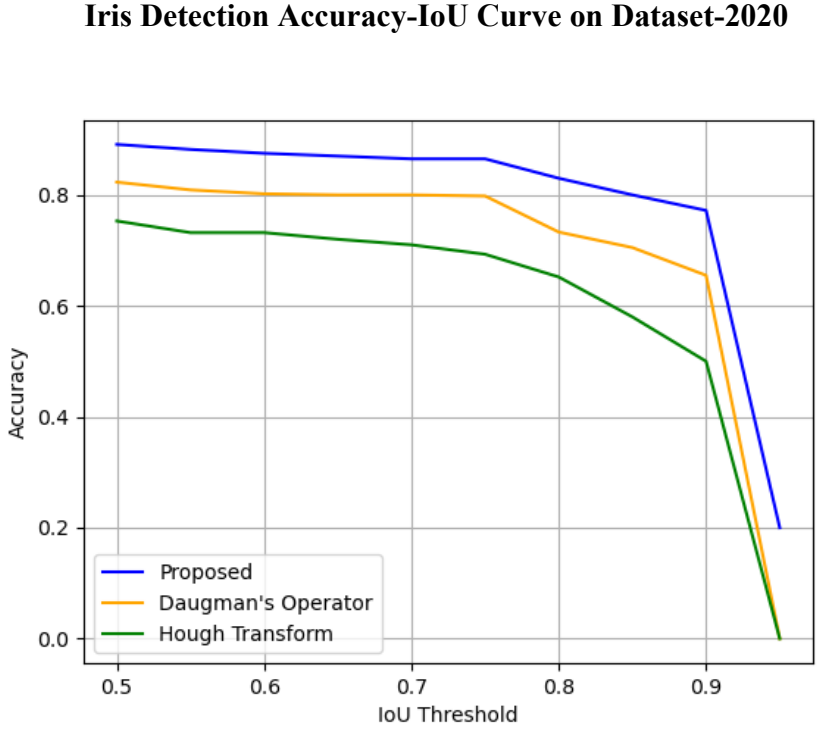


Figure 4.8: Accuracy along the IoU threshold of three methods.

From the example shown in Figure 4.5, we can see that beside detecting the crescent region accurately, the proposed method also can recognize the cornea reflection (bright spot near pupil center), which should not be part of the crescent. This result further shows that compared with the previous approach, the data-driven method is more robust than manually determined threshold.

In practice, the illumination conditions during taking the pictures are not exactly the same. As a result, the manually determined brightness threshold may fail on the images whose overall brightness is too dark or too bright. Also, the performance of the previous method is affected by the various brightness of crescent itself. Even under the same illumination, the ability to reflect light still vary among individuals. This phenomenon leads to more dark crescents without clear boundary.

Methods	Mean error rate	
	Dataset-2015	Dataset-2020
Daugman's Operator-Based	8.17%	6.75%
Hough-Based	9.46%	10.51%
Proposed	**4.02%**	**3.74%**

Table 4.1: Performance of Iris detection – mean error rate

Methods	Mean error rate	
	Dataset-2015	Dataset-2020
Daugman's Operator-Based	4.21%	3.52%
Hough-Based	4.77%	3.79%
Proposed	**4.10%**	**3.00%**

Table 4.2: Performance of Iris detection – mean error rate on eye images without crescent

Performance of Refractive Error Detection

In this study, we measure the overall performance of our refractive error detection approach by computing the mean absolute error (MAE), which is defined as the difference between actual and estimated refractive error. For comparison, we test the state-of-the-art model demonstrated by Kwok [19], which trained SVR with the detected iris, pupil and crescent sizes. We also employ their iris, pupil and crescent detection methods to our proposed feature combination to compare the performance. The results on Dataset-2015 and Dataset-2020 are presented in Table 4.4.

In general, models based on our proposed iris and crescent detection methods outperform state-of-the-art. The best performance of our refractive error detection

Methods	Mean error rate	
	Dataset-2015	Dataset-2020
Kwok [19]	15.61%	23.90%
Proposed	**5.21%**	**8.23%**

Table 4.3: Crescent mean error rate

Features Combination	Detected by Kwok [19]		Detected by Proposed	
	2015	2020	2015	2020
z, r	0.962 D	1.850 D	0.841 D	1.634 D
z, r, Ir (Kwok [19])	**0.890 D**	1.803 D	0.828 D	1.590 D
SoI, Ra	1.172 D	1.855 D	0.841 D	1.728 D
z, r, Ir, SoI, Ra	0.931 D	**1.800 D**	**0.785 D**	**1.575 D**
Theory-driven	1.145 D	2.163 D	0.883 D	1.612 D

Table 4.4: Mean absolute error

achieves MAE of 0.785 D, using the features of Crescent (z), Pupil (r), Iris (Ir), SoI and Ratio (Ra), detected by the proposed methods.

Our methods achieve more accurate iris and crescent detection and therefore outperforms the state-of-the-art on all combinations of features. This result also demonstrates that the new features SoI and $Ratio$ can contribute to the model, though on their own they do not perform as well as the size-based features.

It is interesting to see that these two features can improve the accuracy if detected by proposed methods, but are harmful if detected by Kwok [19]. Since these two features are directly computed based on crescent, the efficiency of them heavily rely on how accurate the crescent is segmented. Therefore, the more accurate approach proposed in this study makes the two features be beneficial. These results

also agree with the previous comparison on crescent detection.

We also observe that the theory-driven method performs worse than the SVR. This is perhaps because the theory-driven method is based on an ideal scenario and requires very precise measurements of features. The machine-learning-based method is, in contrast, more robust.

It is also should be noted that the accuracy on Dataset-2020 is much worse than Dataset-2015. The poor performance could be caused by the relatively low accuracy on crescent detection of iPhone images according to the previous results shown in Table 4.3. However, as a more updated smartphone model, iPhone X is supposed to provide higher quality images than Lumia smartphone, which is also supported by the iris detection results (shown in Table 4.2). Based on our observation, we hypothesize that the hand-crafted crescent detection methods suffer from the ambiguous crescent of iPhone eye images. The flashlight in the iPhone X is composed of four LEDs, which work together when taking pictures. Each LED is a single light source, which, according to the optical principle, can generate a crescent area on the eye images. There thus will be four overlapping crescent areas in the eye images. This creates a blurred boundary on the ”overall” big crescent. As a result, the crescent edge can only be recognized by tiny differences in the brightness gradient, greatly increasing the difficulties of extracting the exact size of crescent. The images in Dataset-2015, which is taken by the Lumia smartphone, has a single LED as the flashlight, which makes the crescent edge much more easy to identify (Fig. 4.9).

Refractive Amblyopia Risk Factor Detection

Amblyopia, or poor vision, is an important measure of eye health. Prior studies have demonstrated that children older than 48 months with a refractive error smaller

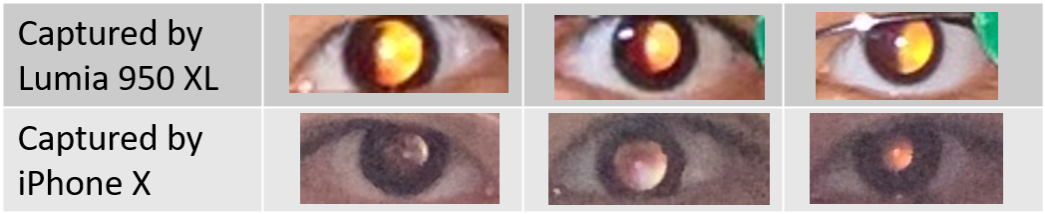


Figure 4.9: Comparison of eye images captured by Lumia 950 XL (the upper row) and iPhone X (the lower row).

than -1.5D have a high risk of developing amblyopia [56,57]. Since photorefraction can measure the refractive error of individuals, it can also be used to estimate the risk of amblyopia.

We further evaluate our method on estimating the risk of amblyopia. In this experiment, the myopic eye images are annotated as Risk or Non-Risk according to their refractive error. We train an SVM for refractive amblyopia risk factor detection with the best features combination presented in the previous experiment. Similarly, we explore the performance of different combinations of features. The performance is shown in Table 4.5, which indicates that our model can attain a much higher sensitivity with reasonable specificity.

Methods	Sensitivity	Specificity	Accuracy
Kwok [19]	64.57%	83.16%	76.22%
Proposed	**77.04%**	**83.05%**	**81.00%**

Table 4.5: Amblyopia factor detection on Dataset-2015

4.1.5 Iris Size-based Calibration

Our proposed methods described thus far require the use of an external calibration tool to obtain the actual sizes of eye structures. Although the calibrating glasses

Methods	Sensitivity	Specificity	Accuracy
Kwok [19]	62.12%	63.79%	62.55%
Proposed	**66.06%**	**76.65%**	**70.21%**

Table 4.6: Amblyopia factor detection on Dataset-2020

are cheap and easy to make, the best scenario would be a system that is completely independent from external devices. In this section, we aim to demonstrate a novel calibration technique which only depends on eye structure, without the external tools.

Method

The eyeball of a human is almost fully grown when born and there is no significant difference between gender and age groups [58]. In addition, the size of iris also does not vary too much from person to person. According to a survey [59], the range of the iris diameter varies from 10.2 to 13.0 mm and the mean size is 12mm. This constancy of iris size gave us the insight to use the iris itself for calibrating sizes. To confirm this, we manually measure the iris size for the eye images in our dataset, and we find that the mean and standard deviation of the iris size are 11.4mm and 0.82 respectively.

Given this information, we run two experiments to test our hypothesis. The first experiment simply uses the number of pixels in the iris as the parameter, rather than the size. The second experiment assumes that the size of the iris is the mean over the dataset (i.e. 11.4mm), and uses that to calibrate the sizes of the other features (e.g. crescent and pupil.).

Calibration Tools	Mean absolute error	Error rate
Glasses Label Distance	7.53 pixels	2.58%
Detected iris radius	0.74 pixels	4.03%

Table 4.7: Mean absolute error in pixels and error rate of different calibration tools

Evaluation

The accuracy of iris radius detection is essential if it is used for calibration. The mean absolute error of proposed iris radius detection method is 0.74 pixels. For the distance between labels on glasses frame, the mean absolute error is 7.53 pixels. In addition, we define the error rate ER as follows:

$$ER = \frac{1}{N}\sum_{n=1}^{N}\frac{|r_n - \hat{r_n}|}{r_n} \tag{4.2}$$

Where N is the total number of eyes, r_n is the manually determined radius of the nth eye (or the distance between labels for glasses calibration), $\hat{r_n}$ is the detected iris radius. The error rate ER of iris radius and the distance between labels on glasses frame are 4.03% and 2.58% respectively. A summary of the result is shown in Table 4.7. Considering that the size of the iris is much smaller than the glasses frame, iris-based calibration is less tolerant to error than glass-based calibration. Though the absolute error of iris radius detection is lower than the error incurred with automatic measurement of the distance between the calibration labels, the error rate is still a issue if we use iris for calibration, which may affect the performance of refractive error prediction.

For comparison, we train an SVR with selected features from 1) Automatic Glasses-based calibration, 2) Manual Glasses-based calibration, 3) Automatic Iris-based calibration, 4) Manual Iris-based calibration and 5) No calibration. *Automatic* means that the measurement are estimated automatically and *no calibration*

simply uses the size in pixels, without attempting to convert to a physical measurement in millimeters. The comparison results are shown in Table 4.8.

Calibration Method	Measurement	Features Used		Theory-driven
		z, r	z, r, SoI, Ra	
Glasses-based	Automatic	0.841 D	0.796 D	0.883 D
Glasses-based	Manual	0.842 D	0.805 D	0.887 D
Iris-based	Automatic	0.852 D	0.814 D	1.075D
Iris-based	Manual	0.812 D	0.806 D	1.112 D
None	-	0.873 D	0.849 D	*

Table 4.8: Comparison of performance in refractive error detection with different calibration methods (MAE, lower is better). Features used: Crescent (z), Pupil (r), Iris (Ir), Sum of Intensity inside Crescent (SoI) and Ratio (Ra)

We can see the performance of iris-based calibration is close to Glasses-based calibration. In particular, for the case where manually-determined iris sizes are used, and the model features are restricted to crescent width and pupil radius, the performance of the model with iris-based calibration is better than the corresponding model with glasses-based calibration. Note that the theory requires actual sizes to compute refractive error. Without calibration, one can only get the size in pixels. So we cannot calculate the theory-driven result for the case where calibration is not used. Considering that the error rate (ER) of the the current iris radius detection is higher than detecting the calibration labels on the glasses, we believe the differences of performance may be eliminated by a more accurate iris detection process. Since the iris-based calibration is is totally independent of external devices, this will enable us to create a more flexible and convenient vision screening system.

4.2 Applying Convolutional Neural Networks

In the previous section, we proposed several hand-crafted features based on the optometry principle and the corresponding detection methods to predict the refractive error. These hand-crafted features include the size of the crescent, pupil radius, and iris radius, and are used as features for SVR and SVM models. Nevertheless, due to the limited image quality, the proposed hand-crafted features are hard to precisely measure even manually by professionals. To further improve the accuracy of the photorefraction model via eye images, there is a need for more robust feature extraction approaches.

Convolutional neural networks (CNN) have shown their strong ability to encode the information of the image and are widely used in computer vision tasks. Some recent studies investigated the feasibility of CNN models on photorefraction tasks [20, 38]. However, there still lacks a comprehensive investigation of implementations of CNN models on this task.

It is also known that light-weight CNN structures are more effective in learning from less training data and taking less running time to generate well-performing models. We therefore propose to apply two state-of-the-art light-weight CNNs, namely, DenseNet [60] and MobileNet [61] and analyze their performance. We will compare the performance against models trained by ResNet-18 and another commonly used CNN structure, VGG-16 [62].

One foreseeable challenge of applying CNN models is that our datasets do not contain enough images for training from scratch. To address the data insufficiency issues, we propose to fine-tune the CNN models that were pre-trained on ImageNet database [63]. We hope the fine-tuned CNN model can encode high-level features that are indicative to refractive error, on the basis of the prior learned knowledge. To reduce the number of trainable parameters and avoid over-fitting, we freeze the

Model	Layers	Parameters (Million)
ResNet18	5 / 18	8.45 / 11.69
VGG16	6 / 16	127.03 / 138.36
DenseNet	36 / 121	5.82 / 7.98
MobileNet	4 / 15	2.49 / 3.50

Table 4.9: Number of layers and parameters (Remain/Total) used for fine-tuning.

first few layers of the CNN models, and fine-tune the remaining ones in the training process. Table 4.9 shows the number of remaining trainable layers and parameters for fine-tuning.

4.2.1 Pipelining CNN features

Well-trained CNN models are able to learn indicative features for the target problem. Prior studies also demonstrated that pipelining learned CNN features to traditional machine learning algorithms often attains better performance than using an end-to-end CNN model [64]. Inspired by these works, we also propose a similar approach. Specifically, we take the outputs (after global average pooling) of the last convolutional layer of the CNN models, and provide them as features for another machine learning model. Fig. 4.10 illustrates the pipeline. For each eye image, the CNN extracted features are obtained as following:

$$F_k = \frac{1}{h \times w} \sum_{i=1}^{h,w} M(k, i, j) \tag{4.3}$$

where the F_k refers to the kth element of final feature vector; the three-dimensional matrix $M(k, h, w)$ is the feature map from kth convolution kernel; h and w are height and width of the feature map respectively. We then train machine learning models on these CNN extracted features for refractive error *Estimation* and ambly-

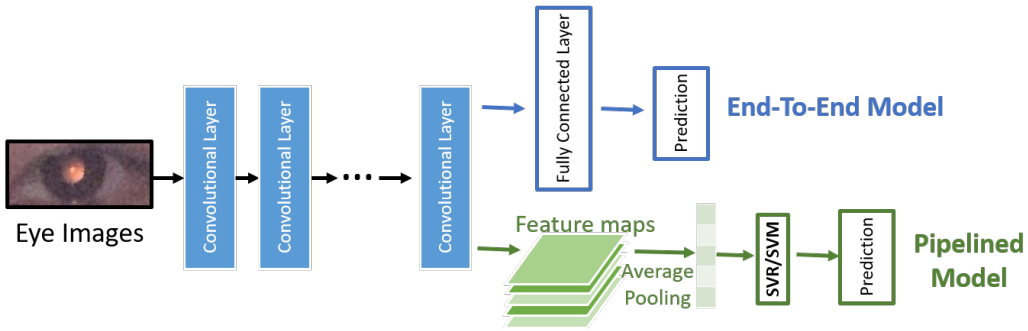


Figure 4.10: Extracting and applying learned features from CNN model.

opia risk factor *Classification*.

As presented in Table 4.9, we work on 4 CNNs. Each one could be trained from scratch or fine-tuned from pre-trained model. We thus could have 8 potential CNN models. We then can extract the CNN features learned by these models to train with other machine learning algorithms. In our method, we adopt support vector regression [41] to train *Estimation* models, and support vector machine (SVM) [65] to train *Classification* models.

4.3 Summary

This chapter firstly investigates the traditional machine learning approaches on smartphone image vision screening. We propose the hand-crafted features and improve the existing feature detection methods. The experiments results reveal that our data-driven models achieve promising accuracy and outperform the theory-driven method that computes the refractive error by optometry principle. Nevertheless, we also observe the hand-crafted features suffer from the image noise, and the ambiguous crescent especially on Dataset-2020. In real-life contexts, it is impractical to expect a smartphone to capture perfect quality eye images. This constrains the efficiency of hand-crafted features, which relies on extracting precise size information of crescent and other areas. To address this challenge, this

chapter then presents an investigation into CNN-based refractive error detection models. In addition, as CNN has the ability to learn and encode indicative features in images, this chapter thus demonstrates that piping the CNN extracted features to SVM/SVR to implement the features more efficiently.

Chapter 5

Data Augmentation

Although convolutional neural networks (CNN) is widely-used and achieves remarkable performance on computer vision tasks, it still faces the drawbacks of overfitting due to the huge amount of trainable parameters when there is no large scale dataset. Data augmentation focuses on the source of this problem by artificially enlarging training dataset. For image data, many works have demonstrated the effectiveness of the traditional augmentations based on the geometric manipulations such as rotating, horizontal and vertical flipping, random cropping, etc. The geometric manipulations make the models to encode the invariance of the images and are good methods to address the positional biases in data. Its application is efficient on many image recognition tasks, but unsuitable on the tasks where the bias of image are complex and needed to be supervised by domain knowledge. For example, on medical image processing, random cropping may alter the label of an image if some critical details are removed. Besides, the traditional augmentations also include non-geometric methods such as noise injection, color transformation, etc. These augmentations lead the models to learn more robust features by inducing the potential noises that present in real world. Their drawback is that some features based on intensity or color may be altered.

Other than traditional data augmentation, synthetic data generated by some models such as GANs, Neural Style Transfer etc. is another way to enlarge the training dataset. The synthetic data is able to fill the gap of unbalanced data with the unseen synthetic samples. In medical area, synthetic data augmentation are successfully implemented on accelerating Magnetic Resonance Imaging (MRI) scanning, quality positron emission tomography estimation, super resolution of retinal vasculature segmentation, etc. [66–69]. These implementations address the problem of unbalance, insufficient and unlabeled data. The validation of structured synthetic medical data also shows the predictive correlation of results derived from synthetic data and the results from real data. [70]

On the photorefraction task, it is hard to collect a lot of data from real patients. To address this challenge, we propose to implement both the traditional and synthetic approaches to artificially generate more photorefraction images for training. In this chapter, we focus on the implementation of traditional data augmentation methods. The augmentation based on synthetic data is introduced in the following chapter. One thing to be noted is that some of the manipulations may be invalid on specific data. For example for the hand-written digit MNIST dataset, the 6 can be transformed to 9, which is in totally different class. On our photorefraction dataset, flipping image is not used since the it will change the meridian at which the crescent shows. As a result, the corresponding astigmatism axis will be changed. Also rotating the image with a large angle is not valid due to the same reason. Random cropping can not be implemented since the refractive error is determined by the details of a specific small region of the eye (the area inside iris). The image would be useless if any of the detail of that region is cropped. Since the intensity of the crescent area are indicative of refractive error, our data augmentation methods need to avoid changing these values. This means that some common data augmentation

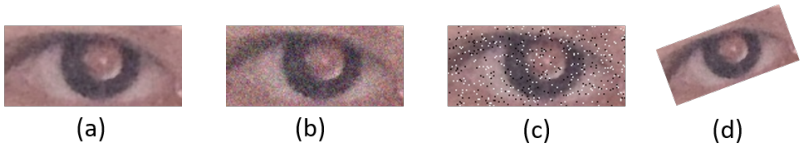


Figure 5.1: Examples of augmented eye images: (a) Original image, (b) Gaussian noise ($\sigma = 15$), (c) impulse noise ($nr = 0.06$), and (d) rotation ($20°$)

techniques such as histogram equalization, white balance and sharpen, etc. are not included in our augmentation method.

Our method adopts rotation, and adding noise (Gaussian and impulse noise) to augment the image data. For each eye image, we first rotate it by $\{-40°, -30°, ..., 40°\}$ respectively, to generate 7 images (including the original one, i.e. rotate by $0°$). We then further add noise to these images. For each of them, we add noise by the following 7 methods respectively: *None*; Gaussian ($\mu = 0$) with $\sigma \in \{15, 30, 45\}$; and impulse with $nr \in \{0.03, 0.06, 0.09\}$ (nr denotes noise rate). Finally, we generate $7 \times 7 = 49$ images for one eye. Fig. 5.1 illustrates some examples of our data augmentation.

5.1 Evaluation

We implemented our models using PyTorch. The models are trained with SGD optimizer with an initial learning rate of 1e-3, which decreases by 0.95 for every 30 epochs. The max training epochs is 600 with a batch size of 16. A momentum of 0.9 is used in the training process. In our method, all the input images are pre-processed for better training, including re-scaling the image resolution to 224×224, and normalizing the pixel intensity to the range of [0, 1].

We adopted 10-fold cross validation for evaluation. The dataset is divided into 10 partitions, and the models trained on 9 of the partitions and evaluated on the

remaining one. This process is repeated 10 times, and the average performance of all the partitions taken as the model performance. Since one subject may contribute to more than one image in our dataset, we ensure that images from the same subject are allocated to the same partition, i.e. we will not have images from the same subject in the training and evaluation set at the same time.

5.1.1 Results of CNN models

Table 5.1 presents the performance of different models, including the models trained with the four CNN structures, with and without fine-tuning (FT) from pre-trained models. The performance is measured in terms of mean absolute error (MAE) in diopter (D) for the *Estimation* task (the lower the better), and classification accuracy for amblyopia risk factor identification (the higher the better).

We first look at the performance of CNN models which are trained from scratch. In general, the CNN models, especially DenseNet and MobileNet, achieve better performance than models with hand-crafted features. The accuracy of hand-crafted methods rely on precise measurements of the sizes of the crescent, pupil, and iris, etc. Most of the time, the boundary of the crescent is not clear especially on the images in Dataset-2020 which are taken by iPhone X with the multiple-LED flashlight.

On the contrary, CNN models are trained to encode indicative information automatically. They are not constrained to size features. It is possible for a CNN to learn more powerful features such as the brightness gradient, shape of crescent area, etc. This makes CNN-based models more robust, especially in real applications where noise is unavoidable. This is supported by the observation that all of the data-driven approaches achieve better performance than the theory-driven approach across the board.

Method	Training (CNN)	Performance on Task/Dataset			
		Estimation (D)		Classification (%)	
		2015	2020	2015	2020
VGG16	from scratch	1.037	1.254	71.18	66.84
ResNet18	from scratch	0.831	1.121	75.30	68.14
DenseNet	from scratch	0.756	1.100	76.12	72.32
MobileNet	from scratch	0.739	1.081	77.20	69.19
VGG16	fine-tuned	0.853	0.825	78.85	78.38
ResNet18 [20]	fine-tuned	0.755	0.811	79.88	79.11
DenseNet	fine-tuned	**0.700**	**0.758**	**82.33**	**81.72**
MobileNet	fine-tuned	0.704	0.773	82.33	80.41
Hand-crafted	-	0.785	1.375	81.00	68.92
Theory-driven	-	0.883	1.612	67.35	62.14

Table 5.1: Evaluating refractive error detection models on different tasks.

Though CNNs can encode powerful features, they require large amounts of data and the small size of our dataset may still affect their performance. This is also revealed in our results: ResNet and VGG do not perform as well as the other two networks. According to Table 4.9, they have more trainable parameters, and thus are more susceptible to over-fitting when the training dataset is small. The experimental results suggest that light-weight CNNs are more appropriate for training well-performing refractive error detection models in our context.

To combat the challenge of data insufficiency, some of our CNN models are pre-trained on ImageNet dataset and our dataset used for fine-tuning. In the fine-tuning process, the knowledge of recognizing low-level features will be kept and adapted for encoding more indicative features for refractive error detection, which

hopefully makes the small dataset more efficient for training. In a way, fine-tuning reduces the required data amount for training well-performing models.

We observe that models with fine-tuning consistently achieve better performance than those trained from scratch, for all the CNN structures and on all tasks. Among the models, DenseNet achieves the best performance across the board, achieving an MAE of 0.755 for refractive error estimation, and accuracy of 81.72% for classification. The results indicate that fine-tuning from pre-trained models can boost the performance where additional data is not available.

We employ the best end-to-end model here, which is the fine-tuned DenseNet to evaluate the Estimation performance improvement brought by the data augmentation methods. The results are illustrated in the Table 5.2, where the digits stand for the performance gain between the model with data augmentation and without data augmentation. The Gaussian noise with sigma as 15 provides the largest improvement for its ability to simulate the noises induced by camera in real scenario. Surprisingly the rotations can also provide relatively large improvement. Theoretically the rotation will change the *axis* that indicates the meridian with maximum diopter. However in this task we only consider the value of refractive error, so the changes on *axis* is not critical. Furthermore the rotation give a simulation of the unstable smartphone while taking the pictures, leading to the eye images not exactly horizontal.

5.1.2 Pipelining from CNN to SVM/SVR

To further improve the model performance, we experiment with pipelining the CNN-learned features to SVR/SVM for the final training and evaluation. We experiment with features learned from the 8 potential models (4 CNN structures, with and without fine-tuning).

Augmentations	Dataset-2015	Dataset-2020
Gaussian, sigma=15	**0.092**	**0.225**
Gaussian, sigma=30	0.068	0.103
Gaussian, sigma=45	0.045	0.111
Impulse, nr=0.03	0.018	0.039
Impulse, nr=0.06	0.003	0.050
Impulse, nr=0.09	0.015	0.031
Rotation, 10	0.037	0.128
Rotation, 20	0.024	0.133
Rotation, 30	0.022	0.096
Rotation, 40	0.010	0.050

Table 5.2: Performance gain of different data augmentation methods.

Table 5.3 shows the results. We again see the benefit of fine-tuning from pre-trained model, which consistently contributes to more indicative features and better performance than the models trained from scratch.

In addition, we can also see that features learned by the fine-tuned DenseNet model outperforms other counterparts across the board, including the best end-to-end trained model (also based on DenseNet). We are able to precisely estimate refractive error with MAE around 0.72, and classify amblyopia risk with 85% accuracy.

It is observed that for all the tested CNN structures, providing the CNN-learned features to an SVR/SVM achieves better performance than an end-to-end-trained CNN model. When we train an end-to-end CNN model, fully connected layers are used to map learned features to the final predictions. Compared to that, SVR/SVM kernels, such as the Gaussian kernels, are able to project features to higher dimen-

Method	Training (CNN)	Performance on Task/Dataset			
		Estimation (D)		Classification (%)	
		2015	2020	2015	2020
VGG16	from scratch	0.833 (0.204)	1.151 (0.103)	72.30 (1.12)	68.48 (1.64)
ResNet18	from scratch	0.796 (0.034)	0.870 (0.251)	76.81 (1.51)	73.36 (5.22)
DenseNet	from scratch	0.720 (0.036)	0.810 (0.290)	82.73 (6.61)	76.24 (3.92)
MobileNet	from scratch	0.724 (0.015)	0.825 (0.256)	81.95 (4.75)	75.78 (6.59)
VGG16	fine-tuned	0.818 (0.035)	0.803 (0.022)	77.62 (1.23)	79.97 (1.59)
ResNet18 [20]	fine-tuned	0.736 (0.019)	0.792 (0.019)	79.88 (0)	81.46 (2.35)
DenseNet	fine-tuned	**0.662** (0.038)	**0.722** (0.036)	**87.95** (5.62)	**84.85** (3.13)
MobileNet	fine-tuned	0.675 (0.029)	0.746 (0.027)	86.52 (4.19)	83.85 (3.44)
Best end-to-end	fine-tuned	0.700	0.758	82.33	81.72
Hand-crafted	-	0.785	1.375	80.95	68.92
Theory-driven	-	0.883	1.612	67.35	62.14

Table 5.3: Performance attained by pipelining CNN-learned features to SVM/SVR. Figures in brackets denote performance gain over end-to-end CNN.

sional spaces. This enables a more non-linear way to model the features, leading to better performance.

It is encouraging to see that the CNN-learned features are effective. However, even for the same CNN structure, the learned features would vary with the training task. Our study includes two training tasks, namely, *Estimation* and *Classification*. Therefore, for one CNN model, there could be two possible sets of learned features. In order to have a deeper understanding on the problem, we also investigate the performance of applying these sets of features on different tasks. We take the fine-tuned DenseNet model, which attains the best performance on all tasks, and feed the features learned for the *Estimation* and *Classification* tasks into the SVM and SVR for training.

Features learned for	Performance on Estimation Task (MAE)		Performance on Classification task (%)	
	2015	2020	2015	2020
Estimation	0.662	**0.722**	89.30	**85.90**
Classification	0.681	0.730	87.95	84.85

Table 5.4: Comparing efficacy of features learned by the fine-tuned DenseNet model for different tasks

Table 5.4 shows the results. It is interesting to note that features trained on the *Estimation* task yield better performance than the *Classification* task. One possible reason for this may be the complexity of the two tasks. We note that the models trained for *Classification* task focus on the data located at the boundary between *Risk* and *Non-Risk* classes (i.e. refractive error close to -1.50D) so that they can distinguish the two classes as much as possible. This is also consistent with findings

from previous studies [38]. On the other hand, models trained for *Estimation* have to focus on all the data points. As a result, they have to encode features that better distinguish data with different refractive error, which are more powerful and lead to more robust performance. In summary, training CNN models for a more fine-grained task appears to be able to learn more powerful features.

From the experiment results, we also observe that the overall performance of the proposed model is heavily affected by its low accuracy on the images with high refractive error. The absolute error is negatively correlated with the amount of images in each refractive error range ($r^2 = 0.72$). Figure 5.2 illustrates the relation between them.

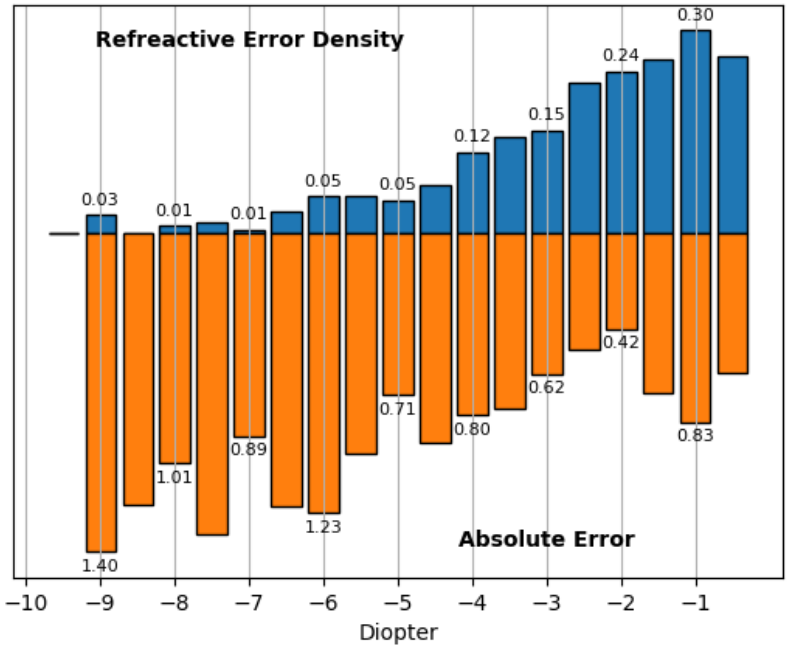


Figure 5.2: The distribution of ground truth refractive error (in diopters) and the absolute error of model for the corresponding images (fine-tuned DenseNet with pipelining).

To further study the effect of data distribution, we down-sample Dataset-2020 such that it exhibits a uniform distribution, and for comparison a dataset with the

same distribution as the original. These two down-sampled datasets contain the same number of images (250), but with different distributions. We then conduct the same training and testing process using the fine-tuned DenseNet with model pipelining. The experiment is repeated 10 times, each time with a different data sampling following the same distribution.

The results are shown in Table 5.5 and Figure 5.3. We can see that under uniform distribution, the model can achieve a similar overall performance as the original one without significant difference (p=0.19), but the performance on high refractive error images is improved by 0.228D (p=0.00).

Distribution	Overall MAE	High Diopter (<-5.0) MAE
Uniform	0.807 D	**0.883 D**
Original	**0.799 D**	1.21 D
Original (No Down-sample)	0.722 D	0.805 D

Table 5.5: The mean absolute error on down-sampled datasets.

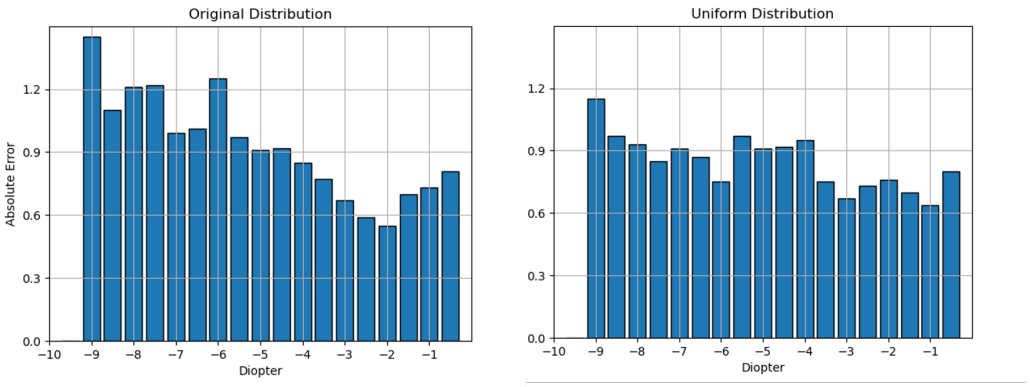


Figure 5.3: The comparison of absolute error distribution of down-sampled datasets.

These results show that that model performance is related to the data amount. Firstly, the total data amount of the two down-sampled dataset is the same, which

leads to a similar overall accuracy. However, the uniform distribution sample contains more images with high refractive error, which makes the model achieve better accuracy on that refractive error range. Similarly, fewer images leads to worse performance on the low refraction error images.

In addition, the comparison of down-sampling and without down-sampling also reveals the effect of data amount. With the same distribution, the dataset without downsampling contains much more data points, leading to much higher accuracy. This result further indicates that more data samples is correlated with better performance.

These encouraging results of the data amount effect show that the CNN models can be improved with more data, especially the images with high refractive error that the current dataset lacks. However, as people with severe myopia only represent only a small part of the population, collecting images with high refractive error is difficult in practice. Therefore, it is hard for CNN models to learn enough information from the current dataset.

5.2 Summary

The difficulty of collecting large-scale datasets is a limitation of training well-performing CNN models. In this chapter, we propose to augment our dataset by rotating and adding noise to the collected images. Our experimental results show that, with data augmentation, CNN models are able to encode more powerful features and achieve better performance for refractive error detection. In addition, our study demonstrates that piping the learned features to SVR/SVM is a more efficient way to make use of the features.

Nevertheless, it is still challenging to train heavy-weight models (i.e. models with too many trainable parameters). Our results suggest that light-weight CNNs

(e.g. DenseNet, MobileNet, etc.) are more appropriate for refractive error detection. Fine-tuning from pre-trained models can further address the data insufficiency challenge and gain much better performance. In our study, we find that fine-tuned DenseNet model pre-trained on the ImageNet dataset yields the best performance among the end-to-end models. Another interesting observation is that CNN models that are trained for a more fine-grained task can contribute to more indicative features. Finally, by applying the features learned by fine-tuned DenseNet model that is trained for the *Estimation* task, we are able to precisely detect refractive error with an estimation error about 0.72, and accuracy of 86% for classification respectively.

Nevertheless, the performance of CNN models is still affected by the imbalanced data. Because few people have severe myopia, the accuracy of CNN models is not satisfactory with insufficient data with high diopter. We conduct experiments to reveal that the accuracy of CNN model is heavily affected by data amount and data distribution. Therefore, more eye images, especially those with severe refractive error are therefore needed to further improve the performance.

Chapter 6

Data Augmentation through Synthetic Photorefraction Images

In the previous chapter, we propose the implementation of several machine learning methods on the photorefraction vision screening via eye images. However, the performance of CNN models is limited by insufficient data. To address this challenge, we develop a photorefraction model based on optical theory. We then use this model to generate synthetic data for data augmentation. This allows us to generate virtually unlimited data with specified refraction error ranges.

The extremely small number of hyperopic eyes in our dataset means that we could not test for correctness even if we could generate synthetic training data. Therefore, our current model handles only the myopic case. We will extend to hyperopic eyes in future work.

6.1 Constructing a Synthetic Eye Image

6.1.1 Eccentric Photorefraction Process

We start by simplifying the eyes, camera, and light source into one optical geometry framework. Under this framework, the crescent formation mechanism can be modeled as a 3-phase process:

- The Retina Phase traces the path of travel of the light from the light source to the retina, where a Retinal Image is formed.
- The Camera Phase traces the reverse path where the light is reflected back from the retina to the image capture device (sensor or film) of the camera.
- The Image Capture Phase models how the light rays that enter the camera manifest in the final image.

We start from geometry and optics to simulate a pupil image with a given refractive error, defined as a combination of spherical error, cylindrical error, and astigmatism axis. The spherical error denotes the magnitude of the myopia or hyperopia and is non-zero when either myopia or hyperopia is present. The cylindrical error denotes the magnitude of the astigmatism and is non-zero when astigmatism is present. The axis refers to where on the cornea the astigmatism is located, and is only non-zero when the cylindrical error is non-zero.

Our derivation is based on an early analysis of eccentric photorefraction with geometrical optics [71]. In addition, we make the following simplifying assumptions:

- the cornea-lens-vitreous structure of human eyes can be simplified to a single convex lens
- the pupil is round

- the curvature of the retina can be neglected (i.e. the retina is modelled as a vertical plane)
- light energy distribution on the surface of lens is uniform.
- the light source (the flash bulb of the camera) can be modelled as a point light source
- the image capture mechanism (light sensor or film), the flash bulb of the camera, and the lens of the camera are all on the same vertical plane

Retina Phase

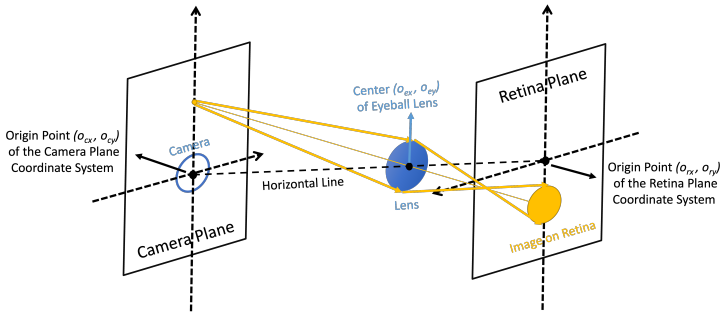


Figure 6.1: The Retina Phase: Formation of the image on retina

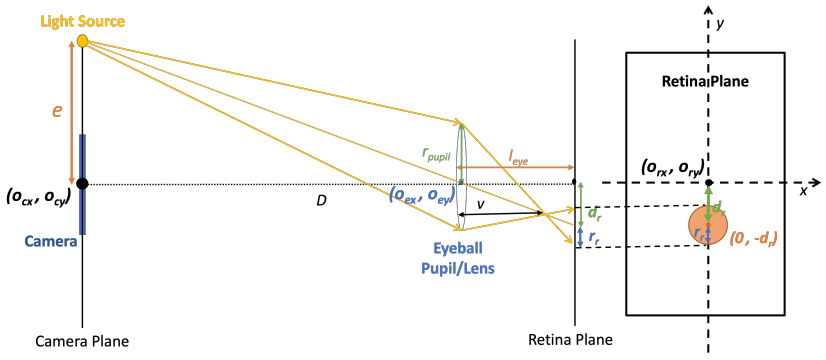


Figure 6.2: The cross section of the Retina Phase process.

Figure 6.1 presents the Retina Phase and Figure 6.2 illustrates the cross section which traces the light route from the light source to the retina. We first define two

coordinate systems, one on the plane of the Camera Lens, and one on the plane of the retina. The origin of the camera plane coordinate system (o_{cx}, o_{cy}) is set at the center of the Camera Lens. The origin of the retinal plane coordinate system (o_{rx}, o_{ry}) is set at the horizontal axis formed between (o_{cx}, o_{cy}), and the center of the eyeball lens (o_{ex}, o_{ey}). Physically, light rays from the light source is projected by the lens onto the retina, forming an image on the retina (which we shall call the *Retinal Image*), which is then sent to the brain via the optic nerve for processing.

Given a point light source, in eyes with perfect vision, the light rays converge to one point on the retina to create a sharp image. In eyes with myopia, the light rays converge to one point in front of the retina, and then diverge again to hit the retina over an area. This results in the brain "seeing" a blurred image. The Retinal Image is a circle in myopic-only cases, and an ellipse if astigmatism is present together with myopia.

Camera Phase

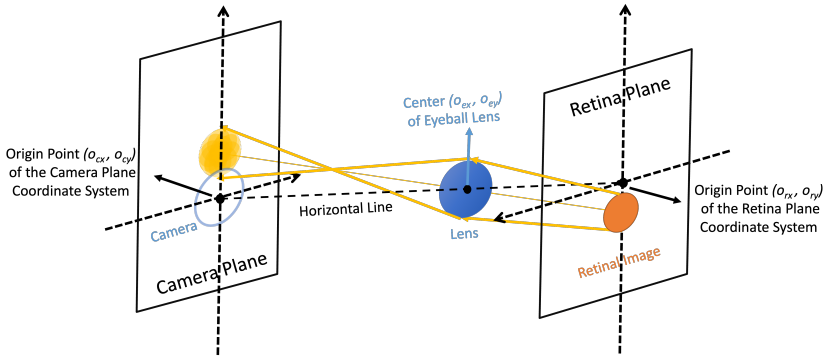


Figure 6.3: The Camera Phase: Formation of the Camera Image.

In the Camera Phase (Figure 6.3), we model the process in which the light rays from the Retinal Image are reflected back to the camera plane, which manifests as a crescent in the pupil of the eye in the captured image. Essentially, in the Camera Phase, the Retinal Image acts as an extended (i.e. non-point) light source. This light

source emits light, which is refracted by the eyeball lens, back in the direction of the camera in a similar way as when the light first entered the eyeball, and eventually generates an image on the Camera Plane.

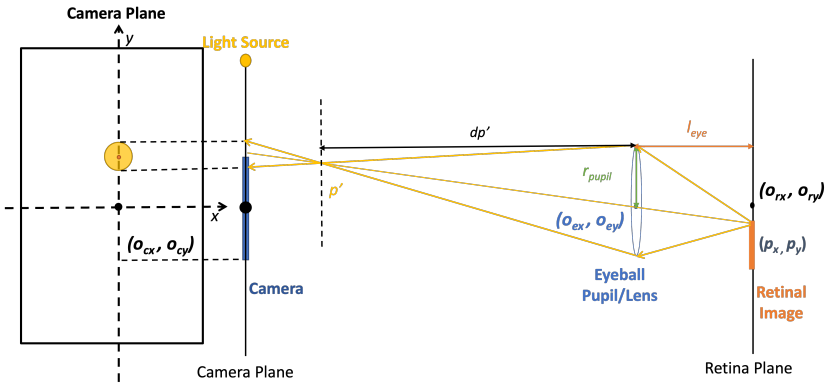


Figure 6.4: Cross section of Camera Phase considering a single point on Retinal Image as light source.

We consider the light rays emitting from one single point on the edge of the Retinal Image, located at (p_x,p_y). The light rays from this point will generate a circle on the camera plane in a similar process as in the Retina Phase, as shown in Figure 6.4. Since the Retinal Image is a circular area, this process is repeated many times, over all the points present within the area of the Retinal Image. Therefore, the whole image on the Camera Plane will be composed of multiple overlapping circles, as shown in Figure 6.5 and Figure 6.6.

Image Capture Phase

In the Image Capture Phase, we model the location of the crescent on the pupil in the final image captured by the camera. Since the image of the crescent passes through the eye lens on the way back from the retina to the camera plane, it is focused onto a point p', called the convergence point, in front of the camera plane. Hence, the projected image on the camera plane is an inverted image, as illustrated in Figure 6.7. However, since the image has already been inverted when it passed

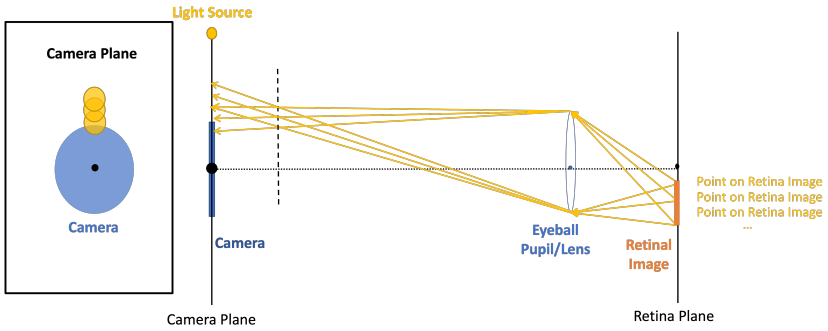


Figure 6.5: Cross section of the overlapping Camera Images generated by multiple points on Retina Plane

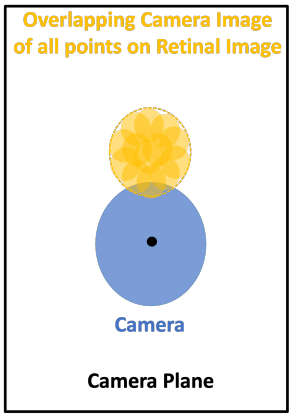


Figure 6.6: Illustration of the overlapping Camera Images on Camera Plane considering all the points on Retinal Image

through the lens during the camera phase, this means that the final image on the camera plane is an inversion of the inverted image (Figure 6.8).

The Camera Image, composed of multiple overlapping circles generated by the projection of light from the Retinal Image, is generated over an area that is larger than the Camera Lens. Hence, only the part of the image that is located in the intersection between Camera Image and the Camera Lens will be captured and manifest as crescent in the final eye picture. (Figure 6.9)

Considering the whole Retinal Image as an extended light source, we can obtain the final crescent by summing up the intensity over all points in all crescents generated by all points in the intersection between the Camera Image and the Camera

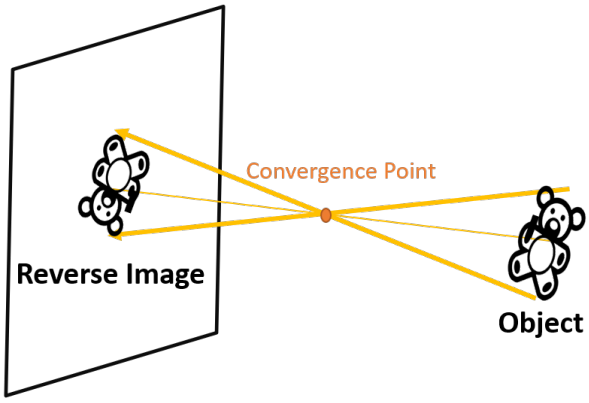


Figure 6.7: The illustration of generating reverse image of the object if the light rays converge in front of the plane.

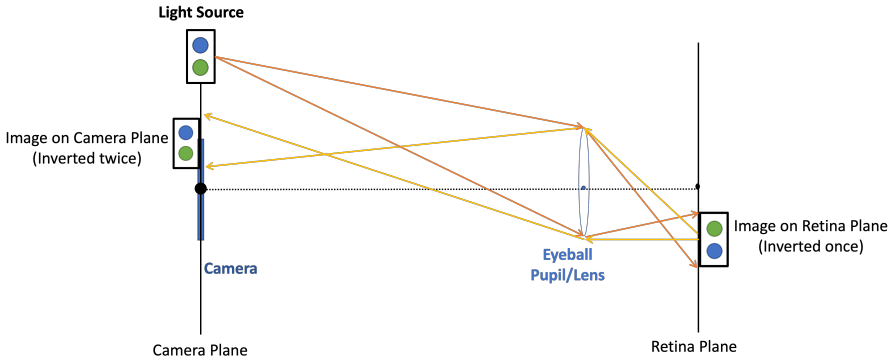


Figure 6.8: The illustration of the formation of the reverse images on Retina Plane and Camera Plane

Lens.

6.1.2 Constructing the Mathematical Model

Retina Phase

We model the light source, camera lens and the eye ball under an geometry optical framework. For the convenience of computing, we present the symbols and notations adopted in the Retina Phase in the Table 6.1.

As the Retinal Image generated by a point light source manifests as a circle, we can compute the position of its center and the radius in the cross-section of the

Variable	Description
(o_{cx}, o_{cy})	Origin of camera plane coordinate system. (Center of the Camera Lens)
(o_{ex}, o_{ey})	Center of the eyeball lens
(o_{rx}, o_{ry})	Origin of retinal plane coordinate system. Located on the retina on the horizontal axis formed between (o_{cx}, o_{cy}) and (o_{ex}, o_{ey})
$(0, -d_r)$	The coordinates of the center of the image on the retina plane
d_r	Distance from center of the Retinal Image $(0, -d_r)$ to the origin of retina plane coordinate system (o_{rx}, o_{ry})
r_r	Radius of the Retinal Image
e	Eccentricity (distance between light source and (o_{cx}, o_{cy}))
r_{pupil}	Pupil radius
D	Distance between center of the eye pupil/lens (o_{ex}, o_{ey}) and Camera Lens (o_{cx}, o_{cy})
l_{eye}	Distance from eyeball lens to the image of the flash inside the eyeball
f_e	Focal length of myopic eyeball lens
f_n	Focal length of perfect (no refractive error) eyeball lens
f_c	Focal length of concave (diverging) corrective lens for myopia

Table 6.1: Notations used for the calculation in Retina Phase.

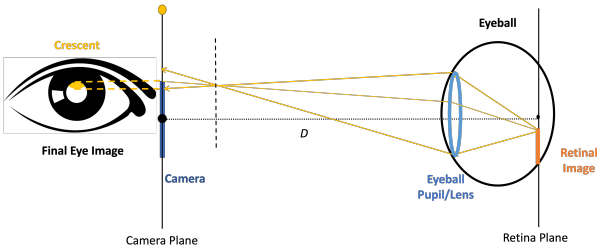


Figure 6.9: The crescent, as manifested in the pupil in the final image.

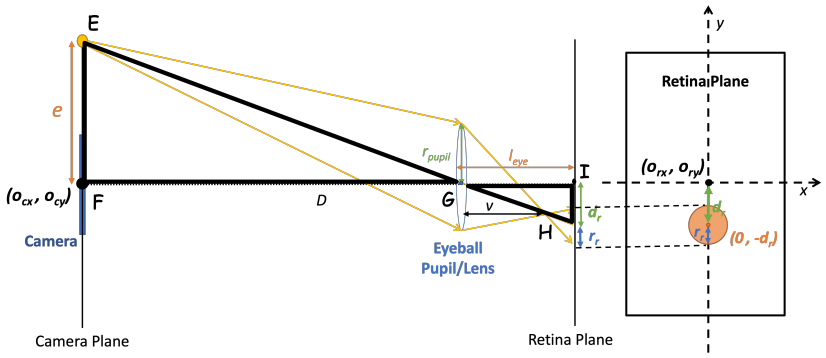


Figure 6.10: Calculation of distance from the origin from the center of the Retinal Image

geometry optical framework. Figure 6.10 illustrates the cross-sectional view for the Retina Phase. Our calculation are based on the *lens equation* from optics:

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f} \tag{6.1}$$

where u is the object distance, v is the image distance, and the f is the focal length.

We first compute the focal length f_n of a normal (no refractive error) eye lens that can correctly focus the light rays onto the retina. By definition, we have:

$$\frac{1}{D} + \frac{1}{l_{eye}} = \frac{1}{f_n} \tag{6.2}$$

$$\Rightarrow f_n = \frac{1}{\frac{1}{D} + \frac{1}{l_{eye}}} \tag{6.3}$$

In a completely relaxed state, the distance from the eyeball lens to the retina is $\frac{1}{60}$ meter [72]. This gives us $l_{eye} = \frac{1}{60}$m.

We take the case of a myopic-only eye with a measured spherical refractive error of R (R is negative in myopic cases). This refractive error measurement

means that the individual would require concave (diverging) corrective lenses with optical power of diopter R to correct the refractive error in the eyeball lens i.e. that the eyeball lens and the glasses together will be able to focus the light rays correctly to a point on the retina. The relationship between the optical power R (in m^{-1}, diopters) and the focal length of a corrective lens is

$$R = \frac{1}{f_c} \tag{6.4}$$

We make the "thin lens" simplifying assumption, which allows us to apply the rule of compound lens:

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} \tag{6.5}$$

Since the corrective lens work together with the eyeball lens to correct the individual's vision to that of a "normal" lens, this gives us:

$$\frac{1}{f_n} = \frac{1}{f_e} + \frac{1}{f_c} \tag{6.6}$$

$$\frac{1}{f_n} = \frac{1}{f_e} + R \tag{6.7}$$

This equation can be solved to give us the focal length of the (uncorrected) eyeball lens, where R is the spherical refractive error (myopia only):

$$\Rightarrow f_e = \frac{1}{\frac{1}{f_n} - R} \tag{6.8}$$

We then determine the position of the Retinal Image by computing the distance d_r between the center of the Retinal Image and (o_{rx}, o_{ry}). In Figure 6.10, consider the similar triangles ΔEFG and ΔHIG. we get:

$$\frac{d_r}{l_{eye}} = \frac{e}{D} \tag{6.9}$$

$$\Rightarrow d_r = \frac{e \times l_{eye}}{D} \tag{6.10}$$

Similarly, in Figure 6.11, consider similar triangles ΔKGJ and ΔLHJ, where their sides and the corresponding heights are in proportion. We can get the radius r_r of the Retinal Image as:

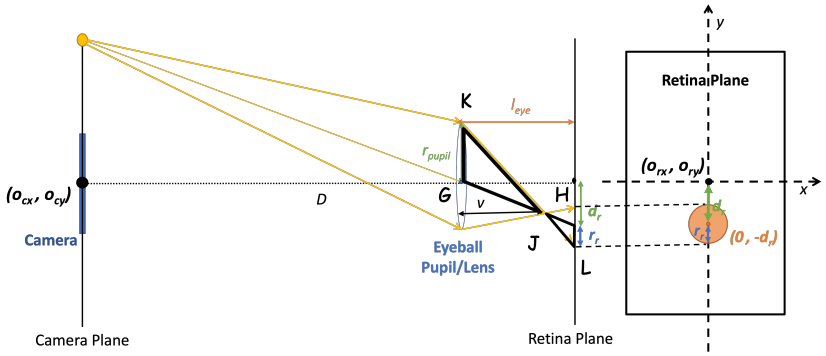


Figure 6.11: Calculation of radius of the image on retina.

$$\frac{r_r}{l_{eye} - v} = \frac{r_{pupil}}{v} \tag{6.11}$$

$$\Rightarrow r_r = \frac{r_{pupil} \times (l_{eye} - v)}{v} \tag{6.12}$$

By this, we can determine a circle on the retina, centered at $(0, -d_r)$ on the retina plane coordinate system, with radius r_r. This is our Retinal Image.

Camera Phase

To compute the position and area of the Camera Image induced by one single point on the retina, there are some additional symbols and notations presented as Table 6.2, which are used in the calculation of Camera Phase.

We first consider one arbitrary point $p \in Im_{retina}$ on the Retinal Image, located at (p_x, p_y). This point will generate a circle on the camera plane in a similar process as in the Retina Phase, as shown in Figure 6.12.

In Figure 6.12, considering the similar triangles ΔNFG and ΔMIG:

$$\frac{p_x}{l_{eye}} = \frac{p''_x}{D} \tag{6.13}$$

$$\Rightarrow p''_x = \frac{p_x \times D}{l_{eye}} \tag{6.14}$$

$$\frac{p_y}{l_{eye}} = \frac{p''_y}{D} \tag{6.15}$$

Notions	Description
Im_{Retina}	The Retinal Image
(p''_x, p''_y)	The coordinates of the center of the Camera Image
$r_{p''}$	Radius of the captured image
p	One single point on the Retinal Image
(p_x, p_y)	The coordinate of the single point p
D	Distance between center of the eye pupil/lens (o_{ex}, o_{ey}) and Camera Lens (o_{cx}, o_{cy})
p'	Convergence point
$d_{p'}$	Distance from eye pupil/lens (o_{ex}, o_{ey}) to the p' of lights in Camera Phase

Table 6.2: Notations used for the calculation in Camera Phase.

$$\Rightarrow p''_y = \frac{p_y \times D}{l_{eye}} \tag{6.16}$$

We proceed to calculate the radius of the image on the camera frame. In Figure 6.13, consider the similar triangles ΔNQP and ΔGKP, where their sides and the corresponding heights are in proportion. we can get:

$$\frac{r_{pupil}}{d_{p'}} = \frac{r_{p''}}{D - d_{p'}} \tag{6.17}$$

$$\Rightarrow r_{p''} = \frac{r_{pupil} \times (D - d_{p'})}{d_{p'}} \tag{6.18}$$

This allows us to determine, on the camera plane, a circle P'' centered at (p''_x, p''_y) with radius $r_{p''}$. This circle will form part of the *camera plane image*. We will name this function, which maps a point from the Retinal Image to a circle on the camera plane, $CP(p)$.

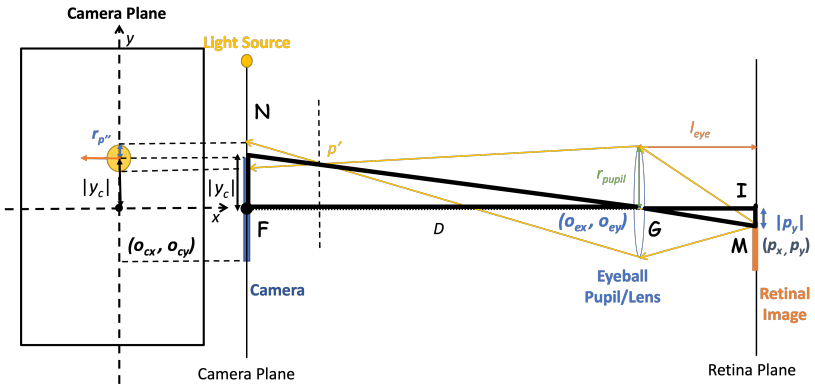


Figure 6.12: Calculation of the distance from the origin of camera plane to the center of image on camera plane

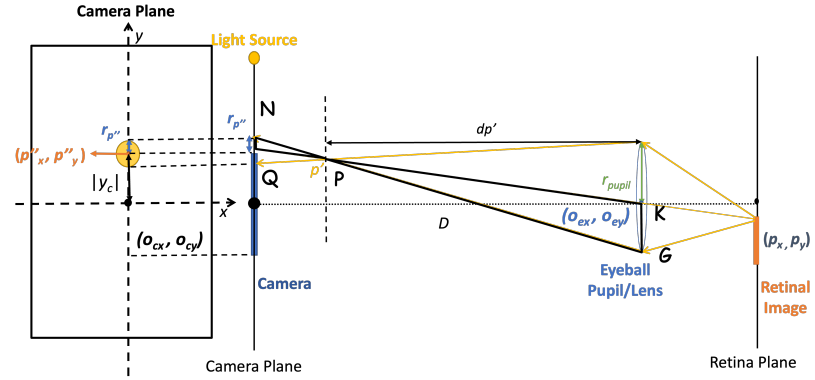


Figure 6.13: Calculation of the radius of the reflected image in the camera plane generated by one point in the Retinal Image

Image Capture Phase

To compute the final image captured by camera, some additional notations are presented in Table 6.3 .

As not the whole image will fall into the Camera Lens, only part of it will be captured by the camera to form the crescent. Reverse tracing the rays from the part of the camera plane image that is captured by the lens, back to its originating location on the Retinal Image, will determine which parts of the eye image will manifest a crescent, and the intensity of the crescent in that location.

In other words, from the Retinal Image, we trace all the light rays that finally

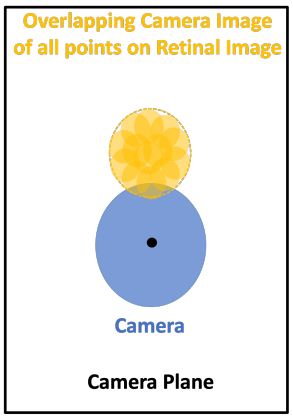


Figure 6.14: Illustration of the final camera plane image

Notions	Description
r_c	The radius of Camera Lens
P	The Camera Image generated by the single point p on Retinal Image
p_f	One point on the final image captured by camera

Table 6.3: Notations used for the Image Capture Phase.

reach the Camera Lens. The locations of these rays as they pass through the lens will give us the locations of the pupil which manifest a bright spot. Summing up the intensity of all eligible rays that pass through that location on the pupil give us the intensity of that point within the crescent. The final entire crescent is constituted of all points which were passed through by light rays that finally enter the Camera Lens.

Given the summation of all of the points generated by the projections of all points in the Retinal Image back to the camera plane. We can get a brightness function of all points overlapping camera plane image:

$$In(p) = \iint_{\Omega} CP(p)dp \tag{6.19}$$

where Ω is the Retinal Image, and $CP()$ is the transfer function that maps a

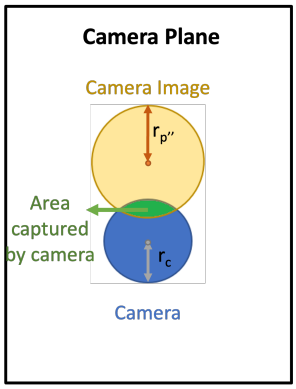


Figure 6.15: Illustration of the final camera plane image.

point p in the Retinal Image to a set of points in a circle on the camera plane as per the process in the camera phase.

As shown in Figure 6.7, since the image of the crescent passes through the eye lens on the way back from the retina to the camera plane, and is focused onto a point in front of the camera plane, it is thus an inverted image. In other words, the final captured image is the Camera Image, flipped horizontally and vertically about the x- and y-axis respectively – in other words, mapping each point (x, y) on the Camera Image to $(-x, -y)$. The image is then scaled such that it is the same size of the image of the pupil.

Mathematically, given the coordinate of a point (x, y), we can determine its position (x', y') in the inverted image as:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = M \times \begin{bmatrix} x \\ y \end{bmatrix} \tag{6.20}$$

where the function of M is to firstly flip (x, y) to $(-x, -y)$, and then scale the image to the same size of actual pupil:

$$M = \begin{bmatrix} \frac{D-d_{p'}}{d_{p'}} & 0 \\ 0 & \frac{D-d_{p'}}{d_{p'}} \end{bmatrix} \times \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \tag{6.21}$$

If we only consider the crescent brightness in the pupil captured by camera without scaling (the scaling does not affect the brightness), the inverted Camera

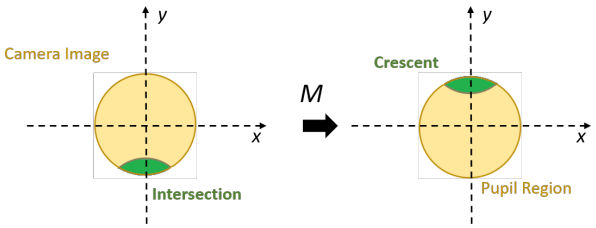


Figure 6.16: Illustration of retrieving the crescent.

Image can be directly regarded as the pupil region (yellow) and the crescent (green) generated by one single point p in the Retinal Image, as shown in Figure 6.16. The final image captured by camera is the summation of the captured area of inverted Camera Images generated by all the points on Retinal Image.

We then consider one point p_f in the pupil region of the final image captured by camera. Without scaling, we consider the final image captured by camera as the same size as the Camera Image. Then the brightness of the point p_f in the pupil region of final captured image can be computed as following:

$$In'(p_f) = \iint_{\Omega} I(p, p_f)\Delta\sigma \tag{6.22}$$

Where p is a point in the pupil, $In'(p_f)$ refers to the intensity at p_f, p is one single point inside $\Delta\sigma$, Ω refers to the whole image on retina, and the $I(p, p_f)$ is indicator function defined as follows:

$$I(p, p_f) = \begin{cases} 1, & \text{if } p_f \text{ is in the captured part of the } P \text{ induced by } p. \\ 0, & \text{if } p_f \text{ is not in the captured part of the } P \text{ induced by } p. \end{cases} \tag{6.23}$$

We set a coordinate system centered as the center of the pupil region of the final image captured by camera, and the coordinate of p_f is (p_{f_x}, p_{f_y}). Since the Camera Image is the inverted final image captured by camera, the coordinate of p_f's corresponding point on any Camera Image when we take the center of Camera Image as the origin is $(-p_{f_x}, -p_{f_y})$. Then if we consider one Camera Image on camera plane, the coordinate of Camera Image center is (x_c, y_c). Thus the coordinate of

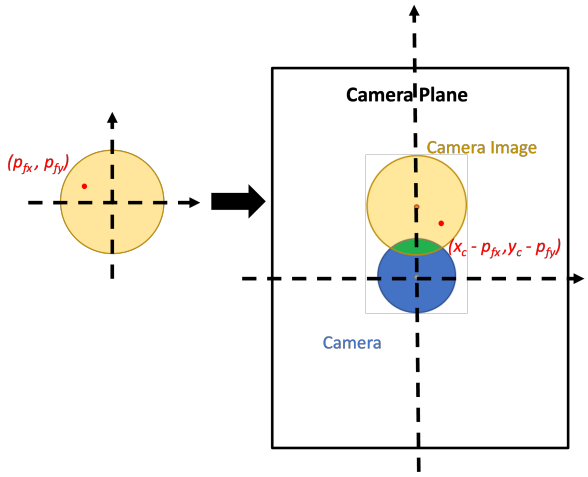


Figure 6.17: Illustration of the coordinate p_f's corresponding point on Camera Image

p_f's corresponding point on camera plane is $(x_c - p_{f_x}, y_c - p_{f_y})$. Figure 6.17 shows the calculation.

Then we can define whether one point is in the captured part of one Camera Image as:

Definition 6.1.1 *A point K of the final captured image is in the captured part of a Camera Image P if:* $||(x_c - K_x, y_c - K_y)||_2 \leq r_c$

where the x_c, x_y is the coordinate of the Camera Image center on camera plane; the K_x and K_y is the coordinate the K's corresponding point on Camera Image; r_c is the radius of Camera Lens. Thus the brightness of each point on the final image captured by camera can be determined by the integral above.

6.1.3 Generalizing for Astigmatism

As mentioned earlier in Section 6.1, standard optometry practice models the eyeball as a composition of a *Sphere* and a *Cylinder*, which is tilted at a particular *Axis*. In myopic-only cases, the *Cylinder* and *Axis* are both zero, thus implying a perfect sphere for the eyeball. When astigmatism is presented, the *Cylinder* value

gives the severity of the astigmatism, and the *Axis* value gives the meridian along which the astigmatism is presented. For the convenience of computing, we present some additional symbols and notations used for the calculation with astigmatism, as shown in Table 6.4. Our synthetic eye construction simplified the eyeball to a perfect sphere, which is an adequate approximation as the eyeball in our synthetic eye is used only for the purposes of simulating the Retinal Image. For the astigmatic case, we model the *lens* such that it has a different refractive error between the meridians, leading to different focal length. In this case, the refractive error on the astigmatism axis remains the *Sphere* value, while the refractive error on its perpendicular meridian is *Sphere* + *Cylinder* value. Consequently, the eyeball lens produces an elliptical Retinal Image and elliptical Camera Image, whose major axis lays on the astigmatism axis and the minor axis is perpendicular to it.

Table 6.4 presents the symbols and notations adopted in the generalization for the astigmatic case.

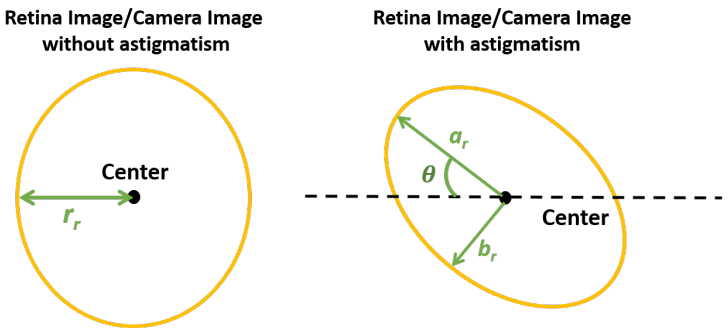


Figure 6.18: The image on retina without astigmatism is a circle (left); an ellipse with astigmatism (right).

Given the *Sphere* and *Cylinder*, we can get the refractive error R_a on the astigmatism *Axis* as and R_p on the axis perpendicular to it.

$$R_a = Sphere \tag{6.24}$$

$$R_p = Sphere + Cylinder \tag{6.25}$$

Then the f_a and f_p can be computed as:

$$f_a = \frac{1}{\frac{1}{f_n} - R_a} \tag{6.26}$$

$$f_p = \frac{1}{\frac{1}{f_n} - R_p} \tag{6.27}$$

On the axis with higher refractive error, the focal length is smaller so that the light rays converge farther away from retina, leading to a larger radius. As a result, on the axis perpendicular to the astigmatism *Axis*, since its refractive error is the largest among all the meridians, the major axis of the ellipse Retinal Image is formed. Ditto for the minor axis formation. We can get the $v_{r,mi}$ and $v_{r,ma}$ by:

$$v_{r,mi} = \frac{1}{\frac{1}{f_a} - \frac{1}{D}} \tag{6.28}$$

$$v_{r,ma} = \frac{1}{\frac{1}{f_p} - \frac{1}{D}} \tag{6.29}$$

Similar to the calculation of the center and radius of circle, we can determine the image on the retina as follows:

$$d_r = \frac{e \times d}{D} \tag{6.30}$$

$$b_r = \frac{r_{pupil} \times (d - v_{r,mi})}{v_{r,mi}} \tag{6.31}$$

$$a_r = \frac{r_{pupil} \times (d - v_{r,ma})}{v_{r,ma}} \tag{6.32}$$

In the Camera Phase, for every point (p_x, p_y) on the Retinal Image, we can determine a corresponding image on camera plane.

$$p''_x = \frac{p_x \times D}{l_{eye}} \tag{6.33}$$

$$p''_y = \frac{p_y \times D}{l_{eye}} \tag{6.34}$$

$$b_c = \frac{r_{pupil} \times (D - v_{c,mi})}{v_{c,mi}} \tag{6.35}$$

$$a_c = \frac{r_{pupil} \times (D - v_{c,ma})}{v_{c,ma}} \tag{6.36}$$

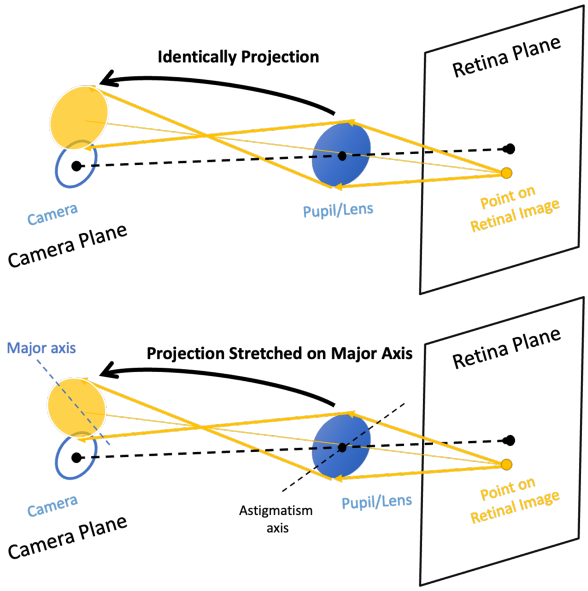


Figure 6.19: Comparison of Camera Image without (upper) and with astigmatism (lower).

When astigmatism is not present, the inverted Camera Image can be directly regarded as the final image since the shape is unchanged, and only needs to be flipped and scaled. However, when astigmatism is presented, the projection is stretched in the major axis direction since the points on this meridian are projected farther away from the center due to the larger refractive error along that axis, as shown in Figure 6.19. Therefore a simple inversion of the Camera Image will also give us a stretched image. To retrieve the originating points (i.e. the locations of the light source Retinal Image), we can compress the elliptical Camera Image in the same scale to a circle in the major axis direction, which is the direction that is stretch during refraction. Then the position of intersection part (green) in the pupil region (yellow) is the originating area of the lights captured by camera, as shown in Figure 6.20.

Similarly, to generate the crescent in the final image, we set a coordinate system centered at the elliptical Camera Image. For the convenience of computation, we can separate the reverse operation into 5 steps:

1. rotate the Camera Image by θ to make the major axis lay on the x-axis of the coordinate system;

2. compress the rotated ellipse along the x-axis to create a circle;

3. rotate the compressed image by $-\theta$ to the original direction;

4. obtain the inverse image of the circle provided by the 3rd step;

5. scale the image to the size of actual pupil.

Thus for every point (x, y) on Camera Image, we can get its corresponding position (x', y') in the reverse image through Equation 6.20 with a modified operator M to realize the four steps:

$$M_a = \begin{bmatrix} \frac{D-v_{c,mi}}{v_{c,mi}} & 0 \\ 0 & \frac{D-v_{c,mi}}{v_{c,mi}} \end{bmatrix} \times \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \times \begin{bmatrix} -cos(\theta) & -sin(\theta) \\ sin(\theta) & -cos(\theta) \end{bmatrix}^{-1} \times \begin{bmatrix} \frac{b_r}{a_r} & 0 \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} -cos(\theta) & -sin(\theta) \\ sin(\theta) & -cos(\theta) \end{bmatrix} \quad (6.37)$$

where the first matrix is identical to the operator used in Equation 6.20, which maps (x, y) to $(-x, -y)$ as in Step 4; the second step is to rotate the point by $-\theta$, which corresponds to Step 3; the third matrix compresses the ellipse along the major axis into a circle by reducing (x, y) to $(\frac{b_r}{a_r}x, y)$, which is Step 2; the last matrix rotates the point by θ. Similarly, the final image can be computed as the summation of all the reverse Camera Images as Equation 6.22.

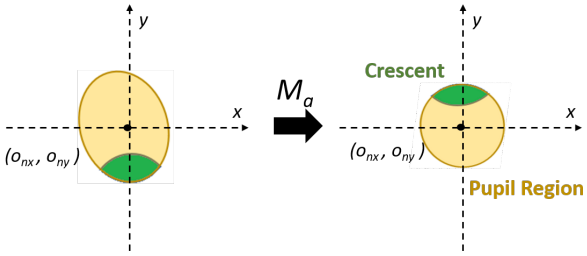


Figure 6.20: Illustration of retrieving the crescent with astigmatism

6.2 Applying the Synthetic Eye Generation

6.2.1 Generation of a Synthetic Eye Dataset

In the previous section, we model the crescent formation mechanism that is used in photorefraction vision screening. Through the analysis of reflection and refraction of lights emitted from light source, as well as how the lights are captured by the camera, we can synthesize pupils with crescents for any desired refractive error.

However, our final objective is to generate photorefraction eye images for data augmentation, which requires that we generate an image of the whole eye, not just the pupils. To this end, the other structure of eye image needs to be constructed. To keep the properties of smartphone-captured eye images in our dataset such as the resolution, noises, etc. we employ the eye images in the real dataset as the templates. Since some of the pupils in real images are too small to identify precisely, we manually select 65 images that appear suitable to be used as templates from both Dataset-2015 (Chapter 3.1) and Dataset-2020 (Chapter 3.2). The requirements are: the images must contain both upper and lower eyelid, and an iris that is not blocked too much. For each image, we convert it into an *eye template* as follows: we first identify and annotate the pupil region manually. Then the identified pupil is cropped out, leaving behind a blank region in the image.

The synthetic pupils are then generated with desired refractive error and scaled

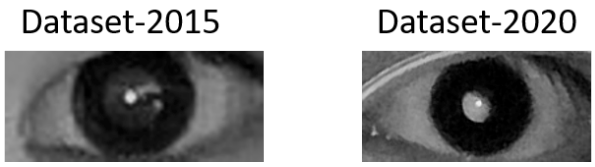


Figure 6.21: Samples of selected eye templates from Dataset-2015 and Dataset-2020

to the same size as the cropped pupil. In addition, the corneal reflex – the phenomenon that is manifested when light reflects off the *surface* of the cornea instead of entering the pupil – has to be considered. This corneal reflex usually manifests as a bright point on the center of the pupil in real images (Figure 6.22). To simulate

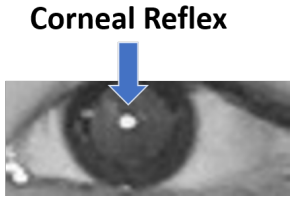


Figure 6.22: The corneal reflex (bright spot in the center of the cornea).

this phenomenon, we add a bright spot on the center of the synthetic pupil. As the cornea reflection is close to a direct reflection of the light source, which is usually the brightest region of the image, we set the brightness of the spot to the highest brightness value of the pixels in the original eye image. In addition, based on our observation, we set the radius of the bright spot as 2 pixels, which appears to be closest to the real corneal reflex by a visual inspection.

In real scenarios, it is usually hard to distinguish the pupil and iris, which are both black (because all of our eyes are Asian eyes). This similarity indicates that the pupil and iris have the same brightness and noise, making it possible to create more realism to our synthetic pupil by adding in noise based on the noise on the iris area, as follows:

We assume that the noise in the iris (and by extension the pupil) follows a

Gaussian distribution. We model the brightness of a given iris pixel as a constant B, plus some noise ϵ, where we note the noise distribution K follows a Gaussian distribution ($K \sim N(0, \sigma^2)$.

Since B is a constant, this means that $var(K) = var(B+K)$. We can calculate $var(B + K) = \frac{1}{n-1}\Sigma_{i=1}^{n}(b_i - \bar{b})^2$, where n is the total number of pixels, b_i is the brightness of one given pixel and $\bar{b}$ is the average brightness over all the pixels. This allows us to calculate $\sigma^2 = var(B + K)$.

For each synthetic pupil generated, we generate some noise ϵ following the distribution $\sim N(0, \sigma^2)$. This noise is then added to the brightness of that pupil.

The statistics of the variance of the distribution of noise of the eye templates is shown in Table 6.5. The brightness of image ranges from 0 to 255.

We then fill the blank space in the eye template with the synthetic pupil. As a result, an eye image is synthesized with the same eye structures of the original image except for the pupil, which is replaced by a synthetic pupil with the desired refractive error. Figure 6.23 illustrates the steps of generating multiple synthetic eye images under different refractive errors with one real eye image used as a template.

Figure 6.24: Synthetic eye image with partially blocked pupil. The synthetic pupil is also blocked on the same region.

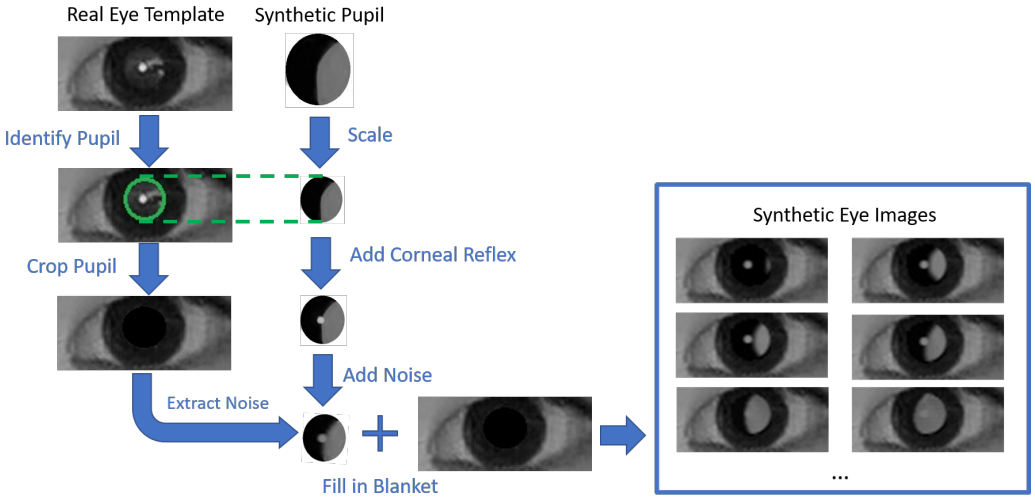


Figure 6.23: Illustration of the synthetic eye image generation process from real eye templates.

6.2.2 Real Evaluation and Experimental Study

The previous subsection described the process of generating eye pupil images with a given refractive error. In this subsection, we describe how to utilize the synthetic images to help to predict refractive error and evaluate their performance.

Automatic Annotation for Eye Features in Real Images

In real eye images, it is usually difficult for people to identify the edge of the pupil or crescent. This difficulty can be due to a number of reasons including: 1) the color of the pupil is too similar to the color of iris (especially for the eyes of Asians); 2) the complex illuminance condition with multiple light sources, which creates an image

with blurred crescents; 3) weak reflection ability of the fundus, which results in a dim crescent. As a consequence, manually annotating these eye features is usually challenging. However, since wrongly annotated images will affect the performance of machine learning models, the quality of the annotated images is also important.

Our synthetic pupils offer a solution to this challenge. We demonstrate a method to automatically annotate the crescent in training set images when the ground truth refractive error is available. Given an eye image, we utilize the proposed photorefraction model to synthesize a pupil with the same refractive error. This pupil is then used to replace the real pupil in the eye image. We then identify the pixels in the image of the real pupil which are in the same corresponding position as the crescent in the synthetic pupil. These pixels are then all annotated as being part of the crescent. An example of automatic annotation and the synthetic eye is shown in Figure 6.25.

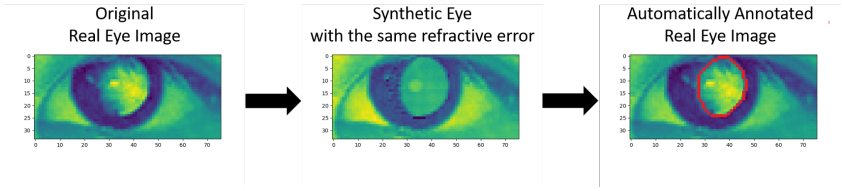


Figure 6.25: The illustration of automatically annotating process. Left: Real eye image with a crescent with ambiguous edges. Center: A synthetic eye, generated with the same refractive error. Right: The real eye image, with the pixels in the corresponding locations as the synthetic crescent annotated (red edge).

We then train the same crescent and iris detection models described in Chapter 4.1 to extract the hand-crafted features and predict refractive error, but with the automatic annotations. The experiment result is shown in Table 6.6.

We first note that on Dataset-2015, the performance with automatic annotation is close to that of the manual annotation with no significant difference (p=0.61).

Furthermore, we also observe that on Dataset-2020, the model trained with automatic annotation outperforms the manual annotation, which suggests that the automatic annotation is more precise than manual annotation. This result is an indirect measure of correctness for our synthetic pupils and the photorefraction model.

Table 6.7 presents the average interaction over union (IoU) between the manual and automatic annotations, which is a measure of the agreement between the manual and automatic annotations, on Dataset-2020 as compared to Dataset-2015. It can be seen that the average IoU is higher on Dataset-2015 than Dataset-2020, which suggests that Dataset-2015 is easier to annotate for humans. The fact that the average IoU is lower on Dataset-2020, but the *performance* of the model is better (lower MAE) on the automatically annotated eyes in Dataset-2020, suggests that human annotators are more likely to make mistakes annotating the blurred crescent boundaries exhibited by the eyes in Dataset-2020.

6.3 Contribution of the Synthetic Eyes

Due to the high-cost of data collection, it is difficult to acquire an adequate amount of photorefraction images to train well-performing CNN models. This problem is especially serious for rare cases (e.g. severe myopia), as these cases by definition present only in a small part of the population. As a consequence, the CNN models' performance is not satisfactory when evaluating on images with severe refractive error.

One possible solution is data augmentation. Nevertheless, there are drawbacks of traditional image augmentation approaches on our problem. Firstly, the operations that apply distortion on the image structure are not appropriate for our problem, since eye structures such as iris, pupil, crescent, etc. have physical meaning. For example, the flipping operation makes the crescent appear on the opposite side

of the pupil, which is a critical discriminator between myopia and hyperopia. Similarly, rotation with a large angle will change the astigmatism degree of eyes; scaling too much will change the shape and size of crescent. All these may be indicative of a different refractive error.

Furthermore, traditional image augmentation approaches generate data based on existing eye images – i.e. with the same refractive error. It is impossible to create samples with refractive errors that have not been seen in the original dataset. As a consequence, the insufficient data issue, especially where images with high refractive error are concerned, still can not be addressed.

In Chapter 4.2, we described a strategy to pre-train the CNN models on other large-scale datasets to let them learn basic information, and then fine-tune them on photorefraction images. We have demonstrated the efficacy of this method by applying CNN models pre-trained on a large-scale general image database (ImageNet), and achieving encouraging results. However, ImageNet samples are not similar to our eye dataset. We therefore postulate that if we could directly train the CNN on photorefraction images, it would be able to reach a better performance.

Given the photorefraction model proposed in Chapter 6.1, it is possible to generate large-scale synthetic photorefraction images for any given refractive error. In this chapter, we aim to exploit the implementation of synthetic photorefraction images on training CNN models.

6.3.1 Method

The synthetic eyes generated by our proposed photorefraction model provides a potential solution to address the insufficient data challenge. For the sake of a comprehensive investigation of the synthetic images, we generate two datasets with (1) uniform refractive error distribution and (2) the same refractive error distribution

of the real dataset. Each synthetic dataset contains 10,000 eye images generated by the proposed photorefraction model. Based on the results in Chapter 4.2, we select the best CNN model DenseNet with model pipelining, where the features extracted by CNN are then fed to SVR/SVM for refractive error *Measurement* and risk factor *Classification*.

Following the pipelining methodology presented in Chapter 4.2, the CNN model is first pre-trained on a synthetic dataset and then fine-tuned with the real eye images. The same experiment settings as chapter 4.2 are used for training, where the model is updated with SGD optimizer with a learning rate of 1e-3, decreasing by 0.95 for every 30 epochs along with a momentum of 0.9.

10-fold cross-validation is adopted for evaluation, where the model is trained on 9 of the partitions and evaluated on the remaining one. To avoid any cheating here, we ensure that the eyes in each testing partition will not be used as an eye template for generating synthetic eyes, fine-tuning models or training SVM/SVR. This process is repeated 10 times, and the average performance of all the partitions is taken as the model's overall performance.

6.3.2 Evaluation

The experiment results on Dataset-2015 are presented in Table 6.8. The synthetic dataset with distribution identical to the real dataset is denoted as Synthetic-I, and the one with uniform distribution as Synthetic-U.

We can see that the model pre-trained on synthetic images outperforms the one on ImageNet on both datasets. The first possible reason is that synthetic images are the eye images captured by smartphones, which is exactly our target domain. It is a lot easier to transfer knowledge between two domains that are close to each other. The pre-trained CNN learns information about eyes, pupils, crescent and

other information relevant to refractive error estimation on the synthetic images, while the models pre-trained on ImageNet database learn basic features for general image classification.

We also note that the synthetic image dataset with the uniform distribution achieves a better overall accuracy than that with distribution identical to the original dataset (i.e. with an imbalanced distribution.) This observation agrees with the findings of Chapter 4.2.

Data Amount Effect

We further explore the effect of the number of synthetic images. Figure 6.26 and Figure 6.27 show the trend of MAE while increasing the size of the synthetic dataset. We can see on both Dataset-2015 and Dataset-2020 that a larger amount of training images leads to better performance before the improvement flattens out. However, we can see that to achieve the best performance, the model on Dataset-2020 requires more data (about 6000 samples) compared with the model on Dataset-2015 (about 4500 samples). This observation also agrees with the finding from Chapter 4.1, where it was found that the model with hand-crafted features finds it harder to reach acceptable performance on Dataset-2020 than Dataset-2015. It is interesting to find that, even though the CNN models were not explicitly designed to extract the hand-crafted features, they still have more difficulty handling the iPhone images with blurred crescent boundaries. On the other hand, the similar behavior between the models with hand-crafted features and CNN model may indicate that the CNN model actually learned to encode information similar to the hand-crafted features.

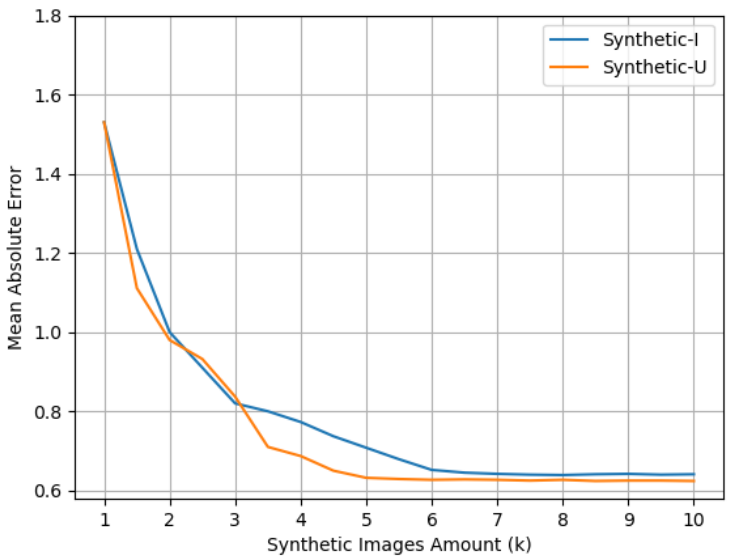


Figure 6.26: The effect of data amount on Dataset-2015. The MAE drops (performance increase) as more data is used for training.

Visualization of CNN Extracted Features

Although the CNN features are able to achieve better performance, they are harder to understand. Unlike the hand-crafted features which are intuitive and directly related to optometry principles, it is hard to explain what happened inside the CNN models and why the features benefit their final performance.

We visualize the CNN extracted features by calculating the Grad-Cam heat map, which is a weighted average of activation map. Given the feature maps from the last convolutional layer M_k of kernel k, we take the partial derivative $\frac{\partial R}{\partial P}$ of the average refractive error (R) of all pixels $P(k, x, y)$ on M_k as its weight W_k. Then we sum all feature maps along with the weights. The final visualization map V can be computed as $V = \sum_{k=1} W_k \times M_k$.

Some samples of the heat map of DenseNet are illustrated in Figure 6.28, where the red regions indicate higher values. We can see that the model pre-trained from ImageNet database focuses more on general image information such as corner,

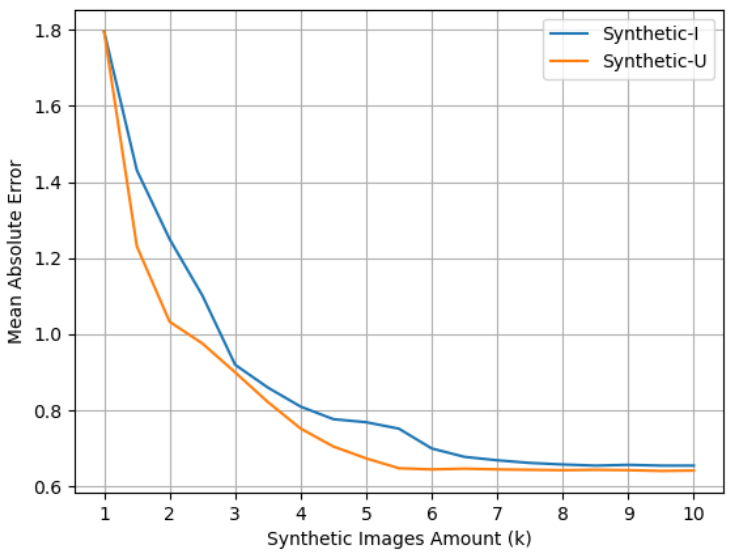


Figure 6.27: Effect of data amount on Dataset-2020. The MAE drops (performance increases) as more data is used for training.

edge, etc. As a comparison, the DenseNet pre-trained on synthetic dataset seems to pay more attention to the regions that are critical for computing crescent, such as eyes boundary, crescent, and pupil. This observation supports our hypothesis that training on the synthetic eye data leads the model to encode information directly related to refractive error.

This visualization provides an interpretation of the better performance of synthetic augmentation. As the models pre-trained on ImageNet tend to encode the basic structures of the image rather than those having optometry meaning, it is possible for the model to 'remember' an image and its refractive error, but not to learn the relation between crescent and refractive error. In contrast, the synthetic dataset contains a large number of images generated from the same eye image, but with different crescents and refractive error. This property could address the over-fitting issue by forcing the models to learn the crescent information which is actually useful to computing the refractive error.

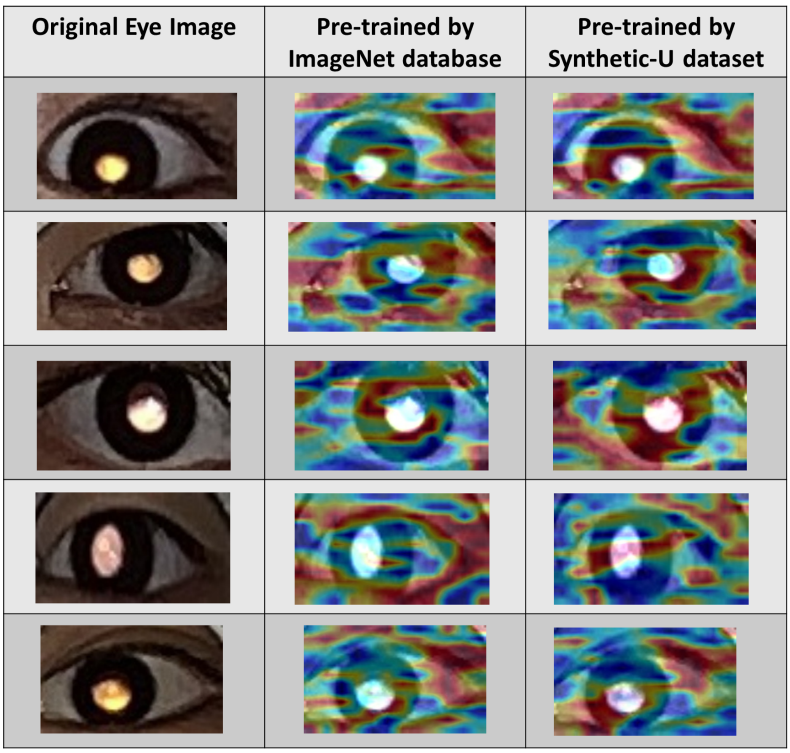


Figure 6.28: The comparison between heat map of DenseNet pre-trained on ImageNet (middle) and pre-trained on Synthetic-U (right).

Figure 6.29 presents a visualization of the effect of dataset size. We see that that as the number of synthetic photorefraction images increases, the red region of the visualization map converges onto the critical eye structures. In addition, the heat map does not change much after 6000 images are added. This agrees with the performance curves from Figures 6.26 and 6.27.

6.4 Summary

In this chapter, we develop a photorefraction model to generate synthetic pupil images and eye images. The contribution of this model is to provide a data augmentation approach of both photorefraction images and corresponding vision screening results. We also generalize the model to the cases with astigmatism. Following the findings in chapter 5, the photorefraction model provides a direction to address the

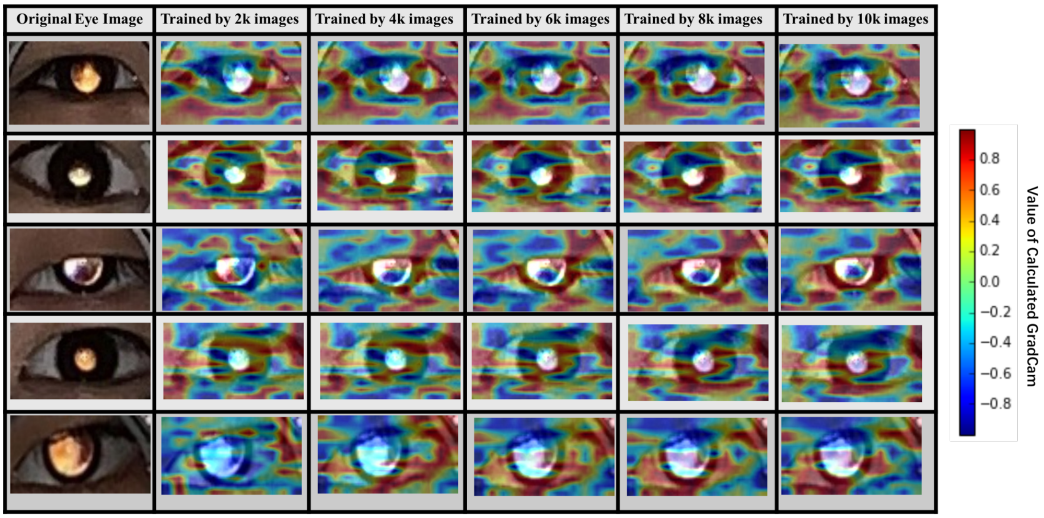


Figure 6.29: Visualizing the effect of the synthetic dataset size using Grad-cam heat map (DenseNet pre-trained on Synthetic-U).

insufficient data problem. It is possible to generate a large-scale dataset containing images and the corresponding refractive error for CNN training. To evaluate the proposed photorefraction model, we introduce the automatic annotation method for eye features to test the synthetic eye images. The experiment shows promising results, especially on the images captured by iPhone X, in which the crescents are hard for humans to identify precisely. It supports the correctness of our proposed photorefraction model and the generated synthetic images.

We then propose to utilize synthetic photorefraction images to pre-train CNN models. The results suggest that our synthetic dataset is a better source dataset than ImageNet for pre-training, and the proposed models achieve state-of-the-art performance. As a critical factor for CNN models, the effect of data amount is also investigated. We conduct the experiments on different synthetic image set sizes. The experiment results reveal that synthetic dataset with uniform refractive error distribution can contribute more to the CNN model because it provides more images on with severe refractive error, which are lacking in the real population and thus rarely seen in real datasets. We also find that Dataset-2020 requires more

images to reach the performance ceiling than Dataset-2015. This coincidence with results from models using hand-crafted features suggests that the CNN models are learning similar structural information from the synthetic data. This is supported by the visualization of the CNN extracted features, which suggests that the CNN models are trained to pay more attention to the crescent and iris areas.

Notions	Description
a_r	Major axis of the Retinal Image
b_r	Minor axis of the Retinal Image
$v_{r,mi}$	Distance from pupil to the convergence point of lights at minor axis in Retina Phase
$v_{r,ma}$	Distance from pupil to the convergence point of lights at major axis in Retina Phase
a_c	Major axis of the image on camera plane
b_c	Minor axis of the image on camera plane
$v_{c,mi}$	Distance from pupil to the convergence point of light rays at minor axis in Camera Phase
$v_{c,ma}$	Distance from pupil to the convergence point of light rays at major axis in Camera Phase
θ	The angle from the major axis of Camera Image or Retinal Image to the horizontal line
f_a	The focal length of eye lens with astigmatism on the astigmatism *Axis*
f_p	The focal length of eye lens with astigmatism on the axis perpendicular to the astigmatism *Axis*

Table 6.4: Additional variables in the case with astigmatism

	Dataset-2015	Dataset-2020
Mean	224.1	210.2
Standard Deviation	77.2	83.5
Range	[51.9 , 993.4]	[39.1 , 721.2]

Table 6.5: Statistics of the variance of the distribution of noise.

Dataset	Automatic Annotation	Manual Annotation
Dataset-2015	0.805D	0.785D
Dataset-2020	1.26D	1.575D

Table 6.6: Mean Absolute Error (MAE) of the hand-crafted features with automatic annotation

Dataset	Average IoU
Dataset-2015	0.921
Dataset-2020	0.873

Table 6.7: Average Intersection over Union (IoU) between the manual and automatic pupil annotations

Pre-trained on	Estimation (D)		Classification (%)	
	Dataset-2015	Dataset-2020	Dataset-2015	Dataset-2020
ImageNet	0.662	0.722	87.95	84.85
Synthetic-I	0.640	0.655	89.37	87.40
Synthetic-U	**0.625**	**0.642**	**90.51**	**89.33**
Train from scratch	0.720	0.810	82.73	76.24
Hand-crafted [73]	0.785	1.375	81.00	68.92
Theory-driven	0.883	1.612	67.35	62.14

Table 6.8: Evaluation results of the DenseNet pre-trained by synthetic dataset (with model pipelining).

Chapter 7

Conclusions

7.1 Contributions

Regular vision screening is critical to early diagnosis and treatment of refractive error but is hard to access by everyone. Recent studies have taken attempts to address this problem through e-health tools with the combination of healthcare and artificial intelligence. Nevertheless, there are still some issues that need to be addressed, including the unsatisfactory performance, inconvenience brought by external attachments and insufficient data, etc. This thesis investigates techniques for a deeper understanding of photorefraction on smartphone images, and develops state-of-the-art models for this problem.

We investigate the feasibility of using machine learning algorithms to analyze smartphone images for photorefraction. To address the non-optimal image quality, we proposed several hand-crafted features and methods to automatically extract them accurately. We train machine learning models to predict refractive error of eye images that outperform the previous studies. We then exploit knowledge about the eye structure to improve the convenience of the photorefraction system through a novel calibration method is that gets rid of the external calibration tools. The

performance without external calibration devices is close to our previous result, which is useful as it means that photorefraction can be conducted with a smartphone alone without suffering a performance drop.

In addition to hand-crafted features which are inspired by theory from optometry, we also investigate the feature extracted by pre-trained CNN models. Experiments are conducted to evaluate the CNN features and their results show promising performance. We also employ CNN models that are pre-trained on a large-scale image dataset and have already gained the ability to encode basic information of images, and pipeline them into other machine learning algorithms that are better able to learn from limited data. Our experiment results show that our approaches acquire much improvement of accuracy.

Inspire by the results of the previous section, we believe that CNN models trained on a large dataset could reach higher performance. The challenge here is on insufficient data. To address this issue, we propose a photorefraction model to calculate the light reflection process from the light source to the eye retina, and back to the camera plane – i.e. what happens during the process of photorefraction. The proposed model provides an approach to generate synthetic photorefraction images for a given refractive error. We conduct experiments to evaluate the generated pupil images and obtain satisfactory results. The experiment's outcome provides solid evidence for the correctness of the proposed photorefraction model.

Next, we are aiming to further improve the performance by applying data augmentation with synthetic images on CNN models. We implement the photorefraction model proposed above to generate a large-scale photorefraction dataset with selected refractive error distribution. The CNN models are pre-trained on the synthetic dataset and go through the same fine-tuning and model pipelining procedure. The experiments result reveal that the CNN models trained on our synthetic dataset,

especially with uniform distribution, achieve the best performance on both refractive error measurement and amblyopia detection.

Overall, this thesis demonstrates a photorefraction method via the eye images captured by smartphone without external devices. We investigate the implementation of various computer vision techniques to address the challenges brought by the image quality. A comprehensive analysis of photorefraction is provided and lead to optometry model to synthesis photorefraction images for data augmentation. We finally manage to achieve the state-of-the-art performance on this task through transfer leaning based on the synthetic photorefraction images.

7.2 Limitations

This thesis aims to develop an eye images-based photorefraction system and the result of experiment are promising and achieve the state of the art. However, there are still limitations on several aspects.

The first limitation of this study is the insufficient data amount. Although we propose a photorefraction model to generate synthetic data, it still cannot totally replace the real dataset. In this thesis, we construct the smartphone photorefraction images dataset with more than 4,000 images. However, the subjects recruited are less than 300, which means there are many eye images are from the same person. The limited amount of subjects leads to limited refractive error samples. In addition, all of the subjects are from a similar ethnic group whose iris are dark. As a consequence, our hand-crafted feature detection methods have not been tested on the eye images with various color of iris. Also, the CNN models require a large number of photorefraction images from other ethnicities to improve their robustness. Furthermore, our study is constrained by diagnosing myopic eyes, whose refractive error is lower than zero. However, hyperopia and astigmatism are also

common visual impairment. There is a need to extend our study to more types of refractive error especially astigmatism.

In addition, the hand-crafted feature detection approaches depend on traditional machine learning algorithms. The performance of these approaches is highly correlated with the corresponding features used to train machine learning models. In other words, the further improvement of the models requires a deeper understanding of eye structure and abundant optometry experience. Another direction is to implement CNN models to detect hand-crafted features including iris pupil sizes, width and shape of crescent. The challenge here is that there is not enough data to train the deep models and there is existing works on the low-quality eye images.

Our system also faces some drawbacks on user experience. The current image-taking operation requires an embedded camera on the back of the smartphone. In addition, the distance from eye to image has to be fixed to be 1 meter. Both of the constraints make it is impossible to operate the system by a single person. One solution is to find a way to make use of the front camera and utilize the screen as a light source. Unfortunately, the brightness of the screen in current commonly used smartphones is not bright enough to generate a recognizable crescent in our experiments, and the screen can manifest a much larger cornea reflection that invalidates the diagnosis. This is something that we intend to investigate in future work as more advanced models are released.

Finally, our system estimates a single value for the refractive error. For real usages, the system should follow real-world conventions as much as possible. Hence, in future work, we will investigate the possibility of converting the single refractive error to the component Sphere, Cylinder and Axis values.

References

[1] Donatella Pascolini and Silvio Paolo Mariotti, "Global estimates of visual impairment: 2010," *British Journal of Ophthalmology*, vol. 96, no. 5, pp. 614–618, 2012.

[2] Hassan Hashemi, Akbar Fotouhi, Abbasali Yekta, Reza Pakzad, Hadi Ostadimoghaddam, and Mehdi Khabazkhoob, "Global and regional estimates of prevalence of refractive errors: systematic review and meta-analysis," *Journal of current ophthalmology*, vol. 30, no. 1, pp. 3–22, 2018.

[3] C. Lam, C. H. Lam, S. Chi-Kwan Cheng, and L. Chan, "Prevalence of myopia among hong kong chinese schoolchildren: Changes over two decades," *Ophthalmic physiological optics : the journal of the British College of Ophthalmic Opticians (Optometrists)*, vol. 32, pp. 17–24, 01 2012.

[4] PJ Foster, , and Y Jiang, "Epidemiology of myopia," *Eye*, vol. 28, no. 2, pp. 202, 2014.

[5] Elie Dolgin, "The myopia boom," *Nature News*, vol. 519, no. 7543, pp. 276, 2015.

[6] Sean P Donahue and CN Nixon, "Visual system assessment in infants, children, and young adults by pediatricians," *Pediatrics*, vol. 137, no. 1, pp. 28–30, 2016.

[7] WC Maples et al., "Visual factors that significantly impact academic performance," *OPTOMETRY-ST LOUIS-*, vol. 74, no. 1, pp. 35–49, 2003.

[8] K. Zadnik, L. T Sinnott, S. A Cotter, L. A Jones-Jordan, R. N Kleinstein, R. E Manny, J D. Twelker, and D. O Mutti, "Prediction of juvenile-onset myopia," *JAMA ophthalmology*, vol. 133, no. 6, pp. 683–689, 2015.

[9] Michael X. Repka, Raymond T. Kraker, Jonathan M. Holmes, Allison I. Summers, Stephen R. Glaser, Carmen N. Barnhardt, David R. Tien, and for the Pediatric Eye Disease Investigator Group, "Atropine vs Patching for Treatment of Moderate Amblyopia: Follow-up at 15 Years of Age of a Randomized Clinical TrialTreatment of Moderate AmblyopiaTreatment of Moderate Amblyopia," *JAMA Ophthalmology*, vol. 132, no. 7, pp. 799–805, 07 2014.

[10] Olalekan A Oduntan, Khathutshelo P Mashige, Franklin E Kio, and Samuel B Boadi-Kusi, "Optometric education in africa: Historical perspectives and challenges," *Optometry and Vision Science*, vol. 91, no. 3, pp. 359–365, 2014.

[11] Daniel Luna, Alfredo Almerares, John Charles Mayan III, Fernán González Bernaldo de Quirós, and Carlos Otero, "Health informatics in developing countries: going beyond pilot practices to sustainable implementations: a review of the current challenges," *Healthcare informatics research*, vol. 20, no. 1, pp. 3, 2014.

[12] Frank Verbeke, Gustave Karara, and Marc Nyssen, "Evaluating the impact of ict-tools on health care delivery in sub-saharan hospitals," in *MEDINFO 2013*, pp. 520–523. IOS Press, 2013.

[13] Sajda Qureshi, "Creating a better world with information and communication technologies: health equity," 2016.

[14] Ariane Kerst, Jürgen Zielasek, and Wolfgang Gaebel, "Smartphone applications for depression: a systematic literature review and a survey of health care professionals' attitudes towards their use in clinical practice," *European archives of psychiatry and clinical neuroscience*, vol. 270, no. 2, pp. 139–152, 2020.

[15] Rachel Perry, Roshan M Burns, Rebecca Simon, and Julie Youm, "Mobile application use among obstetrics and gynecology residents," *Journal of Graduate Medical Education*, vol. 9, no. 5, pp. 611–615, 2017.

[16] Joaquin A Blaya, Hamish SF Fraser, and Brian Holt, "E-health technologies show promise in developing countries," *Health Affairs*, vol. 29, no. 2, pp. 244–251, 2010.

[17] Borja Martínez-Pérez, Isabel de la Torre-Díez, Miguel López-Coronado, Beatriz Sainz-De-Abajo, Montserrat Robles, and Juan Miguel García-Gómez, "Mobile clinical decision support systems and applications: a literature and commercial review," *Journal of medical systems*, vol. 38, no. 1, pp. 4, 2014.

[18] Santosh Krishna, Suzanne Austin Boren, and E Andrew Balas, "Healthcare via cell phones: a systematic review," *Telemedicine and e-Health*, vol. 15, no. 3, pp. 231–240, 2009.

[19] T. C. K. Kwok, N. C. M. Shum, G. Ngai, H. V. Leong, G. A. Tseng, H. Choi, K. Mak, and C. Do, "Democratizing optometric care: A vision-based, data-driven approach to automatic refractive error measurement for vision screening," in *2015 IEEE International Symposium on Multimedia (ISM)*, Dec 2015, pp. 7–12.

[20] Jaehyeong Chun, Youngjun Kim, Kyoung Yoon Shin, Sun Hyup Han, Sei Yeul Oh, Tae-Young Chung, Kyung-Ah Park, and Dong Hui Lim, "Deep learning–based prediction of refractive error using photorefraction images captured by a smartphone: Model development and validation study," *JMIR medical informatics*, vol. 8, no. 5, pp. e16225, 2020.

[21] Kari Kaakinen, "A simple method for screening of children with strabismus, anisometropia or ametropia by simultaneous photography of the corneal and the fundus reflexes," *Acta ophthalmologica*, vol. 57, no. 2, pp. 161–171, 1979.

[22] AC Molteno, J Hoare-Nairne, JC Parr, Anne Simpson, IJ Hodgkinson, NE O'Brien, and SD Watts, "The otago photoscreener, a method for the mass screening of infants to detect squint and refractive errors.," *Transactions of the Ophthalmological Society of New Zealand*, vol. 35, pp. 43–49, 1983.

[23] Howard C. Howland, Velma Dobson, and Nancy Sayles, "Accommodation in infants as measured by photorefraction," *Vision Research*, vol. 27, no. 12, pp. 2141 – 2152, 1987.

[24] WR Bobier and OJ Braddick, "Eccentric photorefraction: optical analysis and empirical measures.," *American journal of optometry and physiological optics*, vol. 62, no. 9, pp. 614–620, 1985.

[25] Howard C Howland, "Optics of photoretinoscopy: results from ray tracing.," *American journal of optometry and physiological optics*, vol. 62, no. 9, pp. 621–625, 1985.

[26] Frank Schaeffel, Howard C Howland, and Leslie Farkas, "Natural accommodation in the growing chicken," *Vision Research*, vol. 26, no. 12, pp. 1977–1993, 1986.

[27] F Schaeffel, H Wilhelm, and E Zrenner, "Inter-individual variability in the dynamics of natural accommodation in humans: relation to age and refractive errors.," *The Journal of Physiology*, vol. 461, no. 1, pp. 301–320, 1993.

[28] Teruko KUBO, Chiaki TAMURA, Hiroaki HIRAI, Hiroshi UOZATO, and Mototsugu SAISHIN, "Clinical application of photorefractor pr-1000 for screening of refractive anomalies in infants," *JAPANESE ORTHOPTIC JOURNAL*, vol. 19, pp. 69–73, 1991.

[29] Florian Gekeler, Frank Schaeffel, Howard C Howland, and John Wattam-Bell, "Measurement of astigmatism by automated infrared photoretinoscopy.," *Optometry and vision science: official publication of the American Academy of Optometry*, vol. 74, no. 7, pp. 472–482, 1997.

[30] Austin Roorda, Melanie CW Campbell, and WR Bobier, "Geometrical theory to predict eccentric photorefraction intensity profiles in the human eye," *JOSA A*, vol. 12, no. 8, pp. 1647–1656, 1995.

[31] David A Atchison and George Smith, "Chromatic dispersions of the ocular media of human eyes," *JOSA A*, vol. 22, no. 1, pp. 29–37, 2005.

[32] Andrew S Paterson, Balakrishnan Raja, Vinay Mandadi, Blane Townsend, Miles Lee, Alex Buell, Binh Vu, Jakoah Brgoch, and Richard C Willson, "A low-cost smartphone-based platform for highly sensitive point-of-care testing with persistent luminescent phosphors," *Lab on a Chip*, vol. 17, no. 6, pp. 1051–1059, 2017.

[33] Neha Srivathsa and Dhananjaya Dendukuri, "Automated abo rh-d blood type detection using smartphone imaging for point-of-care medical diagnostics," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2016, pp. 4345–4348.

[34] Alex Mariakakis, Jacob Baudin, Eric Whitmire, Vardhman Mehta, Megan A Banks, Anthony Law, Lynn Mcgrath, and Shwetak N Patel, "Pupilscreen: Using smartphones to assess traumatic brain injury," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, pp. 81, 2017.

[35] Alex Mariakakis, Megan A Banks, Lauren Phillipi, Lei Yu, James Taylor, and Shwetak N Patel, "Biliscreen: smartphone-based scleral jaundice monitoring for liver and pancreatic disorders," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 2, pp. 20, 2017.

[36] SVOne Official Website, "https://www.smartvisionlabs.com/autorefractors/svone-specifications/," .

[37] Eyenetra Official Website, "https://eyenetra.com/," .

[38] Eugene Yujun Fu, Zhongqi Yang, Hong Va Leong, Grace Ngai, Chi-wai Do, and Lily Chan, "Exploiting active learning in novel refractive error detection with smartphones," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 2775–2783.

[39] Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai, et al., "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354–377, 2018.

[40] Dorothy M Win-Hall, Jaime Houser, and Adrian Glasser, "Static and dynamic measurement of accommodation using the grand seiko wam-5500 autorefractor," *Optometry and vision science: official publication of the American Academy of Optometry*, vol. 87, no. 11, 2010.

[41] Alex J Smola and Bernhard Schölkopf, "A tutorial on support vector regression," *Statistics and computing*, vol. 14, no. 3, pp. 199–222, 2004.

[42] Marti A. Hearst, Susan T Dumais, Edgar Osuna, John Platt, and Bernhard Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their applications*, vol. 13, no. 4, pp. 18–28, 1998.

[43] Z. He, T. Tan, Z. Sun, and X. Qiu, "Toward accurate and fast iris segmentation for iris biometrics," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 9, pp. 1670–1684, 2008.

[44] J. Daugman, "How iris recognition works," in *The essential guide to image processing*, pp. 715–739. Elsevier, 2009.

[45] D. S. Jeong, J. W. Hwang, B. J. Kang, K. R. Park, C. S. Won, D. K. Park, and J. Kim, "A new iris segmentation method for non-ideal iris images," *Image and vision computing*, vol. 28, no. 2, pp. 254–260, 2010.

[46] N. F Soliman, E. Mohamed, F. Magdi, F. E A. El-Samie, and M AbdElnaby, "Efficient iris localization and recognition," *Optik*, vol. 140, pp. 469–475, 2017.

[47] R. P Wildes, "Iris recognition: an emerging biometric technology," *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1348–1363, 1997.

[48] L. Ma, Y. Wang, and T. Tan, "Iris recognition using circular symmetric filters," in *Object recognition supported by user interaction for service robots*. IEEE, 2002, vol. 2, pp. 414–417.

[49] D. M Monro, S. Rakshit, and D. Zhang, "Dct-based iris recognition," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, , no. 4, pp. 586–595, 2007.

[50] BH Shekar and S. S Bhat, "Multi-patches iris based person authentication system using particle swarm optimization and fuzzy c-means clustering," *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, pp. 243, 2017.

[51] M. A Fischler and R. C Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[52] T. Thomas, A. George, and KP I. Devi, "Effective iris recognition system," *Procedia Technology*, vol. 25, pp. 464–472, 2016.

[53] Y. Yin, L. Liu, and X. Sun, "Sdumla-hmt: a multimodal biometric database," in *Chinese Conference on Biometric Recognition*. Springer, 2011, pp. 260–268.

[54] Biometrics Ideal Test, "Casia iris image database," http://biometrics.idealtest.org/, 2010.

[55] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[56] Robert W Arnold, James W O'Neil, Kim L Cooper, David I Silbert, and Sean P Donahue, "Evaluation of a smartphone photoscreening app to de-

tect refractive amblyopia risk factors in children aged 1–6 years," *Clinical Ophthalmology (Auckland, NZ)*, vol. 12, pp. 1533, 2018.

[57] M M. W Peterseim, R. S Rhodes, R. N Patel, M E. Wilson, L. E Edmondson, S. A Logan, E. W Cheeseman, E. Shortridge, and R. H Trivedi, "Effectiveness of the gocheck kids vision screener in detecting amblyopia risk factors," *American journal of ophthalmology*, vol. 187, pp. 87–91, 2018.

[58] DA Leighton and A Tomlinson, "Changes in axial length and other dimensions of the eyeball with increasing age," *Acta ophthalmologica*, vol. 50, no. 6, pp. 815–826, 1972.

[59] Patrick Caroline, "The effect of corneal diameter on soft lens fitting," *Global Insight*, 2016.

[60] Forrest Iandola, Matt Moskewicz, Sergey Karayev, Ross Girshick, Trevor Darrell, and Kurt Keutzer, "Densenet: Implementing efficient convnet descriptor pyramids," *arXiv preprint arXiv:1404.1869*, 2014.

[61] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[62] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[63] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[64] Xiao-Xiao Niu and Ching Y Suen, "A novel hybrid cnn–svm classifier for recognizing handwritten digits," *Pattern Recognition*, vol. 45, no. 4, pp. 1318–1325, 2012.

[65] Ingo Steinwart and Andreas Christmann, *Support vector machines*, Springer Science & Business Media, 2008.

[66] Jelmer M Wolterink, Tim Leiner, Max A Viergever, and Ivana Išgum, "Generative adversarial networks for noise reduction in low-dose ct," *IEEE transactions on medical imaging*, vol. 36, no. 12, pp. 2536–2545, 2017.

[67] Ohad Shitrit and Tammy Riklin Raviv, "Accelerated magnetic resonance imaging by adversarial neural network," in *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pp. 30–38. Springer, 2017.

[68] Yan Wang, Biting Yu, Lei Wang, Chen Zu, David S Lalush, Weili Lin, Xi Wu, Jiliu Zhou, Dinggang Shen, and Luping Zhou, "3d conditional generative adversarial networks for high-quality pet image estimation at low dose," *Neuroimage*, vol. 174, pp. 550–562, 2018.

[69] Dwarikanath Mahapatra and Behzad Bozorgtabar, "Retinal vasculature segmentation using local saliency maps and generative adversarial networks for image super resolution," *arXiv preprint arXiv:1710.04783*, 2017.

[70] Anat Reiner Benaim, Ronit Almog, Yuri Gorelik, Irit Hochberg, Laila Nassar, Tanya Mashiach, Mogher Khamaisi, Yael Lurie, Zaher S Azzam, Johad Khoury, et al., "Analyzing medical research results based on synthetic data and their relation to real data results: systematic comparison from five ob-

servational studies," *JMIR medical informatics*, vol. 8, no. 2, pp. e16492, 2020.

[71] Wolfgang Wesemann, Anthony M Norcia, and Dale Allen, "Theory of eccentric photorefraction (photoretinoscopy): astigmatic eyes," *JOSA A*, vol. 8, no. 12, pp. 2038–2047, 1991.

[72] Daniel Palanker, "Optical properties of the eye," *AAO One Network*, vol. 48, 2013.

[73] Zhongqi Yang, Eugene Yujun Fu, Grace Ngai, Hong Va Leong, Chi-wai Do, and Lily Chan, "Screening for refractive error with low-quality smartphone images," in *Proceedings of the 18th International Conference on Advances in Mobile Computing & Multimedia*, 2020, pp. 119–128.