Communication

Pupil Size Prediction Techniques Based on Convolution Neural Network

Allen Jong-Woei Whang [1], Yi-Yung Chen [2,*], Wei-Chieh Tseng [1], Chih-Hsien Tsai [3], Yi-Ping Chao [4,5,6], Chieh-Hung Yen [4,7,8], Chun-Hsiu Liu [7,8] and Xin Zhang [1]

1. Department of Electronic and Computer Engineering, National Taiwan University of Science and Technology, Taipei City 106335, Taiwan; whang@mail.ntust.edu.tw (A.J.-W.W.); D10502301@mail.ntust.edu.tw (W.-C.T.); M10802312@mail.ntust.edu.tw (X.Z.)
2. Graduate Institute of Color & Illumination Technology, National Taiwan University of Science and Technology, Taipei City 106335, Taiwan
3. Graduate Institute of Electro-Optical Engineering, National Taiwan University of Science and Technology, Taipei City 106335, Taiwan; peter8411124146@gmail.com
4. Graduate Institute of Biomedical Engineering, Chang Gung University, Taoyuan City 333323, Taiwan; yiping@mail.cgu.edu.tw (Y.-P.C.); chiehhungyen@gmail.com (C.-H.Y.)
5. Department of Computer Science and Information Engineering, Chang Gung University, Taoyuan City 333323, Taiwan
6. Department of Neurology, Chang Gung Memorial Hospital at Linkou, Taoyuan City 333423, Taiwan
7. Department of Ophthalmology, Chang Gung Memorial Hospital at Linkou, Taoyuan City 333423, Taiwan; cdeegg@gmail.com
8. College of Medicine, Chang Gung University, Taoyuan City 333323, Taiwan
* Correspondence: yiyung-chen@mail.ntust.edu.tw

Abstract: The size of one's pupil can indicate one's physical condition and mental state. When we search related papers about AI and the pupil, most studies focused on eye-tracking. This paper proposes an algorithm that can calculate pupil size based on a convolution neural network (CNN). Usually, the shape of the pupil is not round, and 50% of pupils can be calculated using ellipses as the best fitting shapes. This paper uses the major and minor axes of an ellipse to represent the size of pupils and uses the two parameters as the output of the network. Regarding the input of the network, the dataset is in video format (continuous frames). Taking each frame from the videos and using these to train the CNN model may cause overfitting since the images are too similar. This study used data augmentation and calculated the structural similarity to ensure that the images had a certain degree of difference to avoid this problem. For optimizing the network structure, this study compared the mean error with changes in the depth of the network and the field of view (FOV) of the convolution filter. The result shows that both deepening the network and widening the FOV of the convolution filter can reduce the mean error. According to the results, the mean error of the pupil length is 5.437% and the pupil area is 10.57%. It can operate in low-cost mobile embedded systems at 35 frames per second, demonstrating that low-cost designs can be used for pupil size prediction.

Keywords: biomedical imaging; computational intelligence; engineering in medicine and biology; machine learning

1. Introduction

The irises of lower vertebrates are intrinsically photosensitive, so a pupillary light reflex (PLR) does not need to be controlled by the brainstem. However, the PLR of higher vertebrates is governed by the brainstem [1]. The pupil size of the human eye is between 1.5 mm and 9 mm [2] and is controlled by the autonomic nerve. Thus, the optic nerve function of the central and peripheral nervous systems can be evaluated [3]. In clinical practice, the pupillary response to light stimuli evaluates the retina, optic nerve function, and brainstem [4]. PLR is essential in the diagnosis of eye diseases and nervous system

research. PLR mainly measures the size (diameter or area) of the pupil [5], and the size of the pupil is controlled by the circular (sphincter) and radial muscles of the iris. The parasympathetic nervous system (PNS) innervates the circular muscle, and the sympathetic nervous system (SNS) controls the radial muscle [6]. Both PNS and SNS can be used as parameters to predict a patient's physical condition. When a patient has inconsistent responses on both sides of the pupil, or the contraction response is different from ordinary people, it may be a sign of certain diseases [6–13].

Many studies have shown that, under the stimulation of red light and blue light, the pupil contraction in primary open-angle glaucoma (POAG) is either smaller than that in ordinary people or non-existent [7,8,13]. In addition to glaucoma and a series of retinopathy caused by diabetes, including patients with diabetic retinopathy (DR), non-proliferative diabetic retinopathy (NPDR), and proliferative diabetic retinopathy (PDR), the pupil response decreases as the disease worsens under the stimulation of a light source [6,9,13]. Retinopathy is usually accompanied by the reduced function of intrinsically photosensitive retinal ganglion cells (ipRGC). In addition to abnormal post illumination pupil response (PIPR), it can also cause the dysregulation of circadian rhythms [9,13]. In addition to the lesions mentioned above, conditions caused by abnormalities of the central nervous system can also affect the pupil. For example, the pupils of patients under general anesthesia and suffering from Horner syndrome will respond differently to light source stimulation than those of ordinary people [10–12]. Pupil size is also very helpful in psychology; the size of the pupil changes with mood. The pupil dilates when a subject is in a pleasant mood; otherwise, it constricts [14,15]. In clinical practice, the size of the pupil is predicted by the experience of the medical staff. The behavior will not be unified since it depends on the subjective consciousness or the degree of fatigue among medical staff [16]. Medical equipment for pupil measurement is expensive and must rely on the cooperation of patients, which is inconvenient for medical staff.

For these reasons, this paper suggests that there must be a way to quantify pupil size, which is convenient and has a low cost. In the past, most studies have focused on pupil centering or eye-tracking [17–23]. There are few studies on pupil size. Concerning the study of pupil size, for example, Garcia et al., used OpenCV to preprocess an image and then used the Otsu threshold and contour detection to complete pupil size detection [16]. De Souza et al., proposed using two complementary independent algorithms. They then used the results of these two separate algorithms to infer the center of the pupil and looked outward to the edge to complete the pupil size prediction [24]. De Santis et al., used the level set theory to realize fully automatic segmentation and then used this result to predict pupil size [25]. Thasina Tabashum et al., used the Kalman Filter to develop a real-time prototype that can simultaneously extract pupil size over time and enable adjustment frame by frame [26]. S Navaneethan et al., proposed a human eye pupil detection system to quickly recognize and diagnose the human eye pupil area. Double threshold, logical OR, morphological closing, and average black pixel density modules are involved in the proposed solution [27]. These studies about pupil size most used rule-based algorithms. Additionally, Taehyung Kim et al., used a convolutional neural network-based semantic segmentation method for accurate pupil detection [28] that mixes the rule-based algorithm with an AI-based program. Deep learning is a type of artificial intelligence (AI). In terms of AI acceptance, 84.2% of healthcare workers agree that AI can assist the imaging and pathology department, and 76.3% of non-healthcare workers agree that AI is helpful [29]. Therefore, this paper proposes an algorithm based on deep learning that allows real-time calculations in a low-cost mobile embedded system. The research will help solve clinical problems in pupil measurement, including the fact that pupil size is unable to be quantified in real-time, involves inconvenient operation, and is often expensive. The main contributions of this paper include:

- We have proposed pupil size detection based on a convolution neural network that allows real-time calculation in a low-cost mobile embedded system.

- We have evaluated the performance of the proposed approach with multiple realistic datasets for optimizing the structure.

2. Materials and Methods

2.1. Dataset

This study chose three different datasets, labeled pupils, from the wild (LPW), CASIA-Iris, and Świrski datasets. The purpose was to prevent the training of the only dataset, which would create overfitting. In addition, we selected images from datasets, except for eye images with complete pupils for training, and struck off any photographs that showed the pupil partially covered. Additionally, the training data added complex pupil images, such as those which included glasses and makeup. Those taken outdoor also increased the challenge of the model.

2.1.1. Labeled Pupils in the Wild Dataset

The dataset used in this study is labeled as pupils in the wild (LPW) and was provided by Tonsen et al. [30]. The first dataset had nine different conditions for high-quality eye images. It provided a label for the center of the pupil, which strengthened the challenge and accuracy of our model. The images of this dataset are shown in Figures 1 and 2 with ellipse fitting.

Figure 1. The sample images of the LPW dataset.

LPW was provided in the form of videos, so this study needed to extract the images from the video in an appropriate sampling frequency. The videos were continuous frames, so the similarity of adjacent photos was high. The high similarity could easily cause over-training. Therefore, this study performed a structural similarity analysis to ensure that the sampling frequency was not too high. The calculation method of the structural similarity index is as follows:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{1}$$

where the μ_x and μ_y are mean values, σ_x and σ_y are standard deviations, σ_{xy} is the covariance, and C_1 and C_2 are constants used to control the overall stability. If the structural similarity index is 1, the two images do not have a difference.
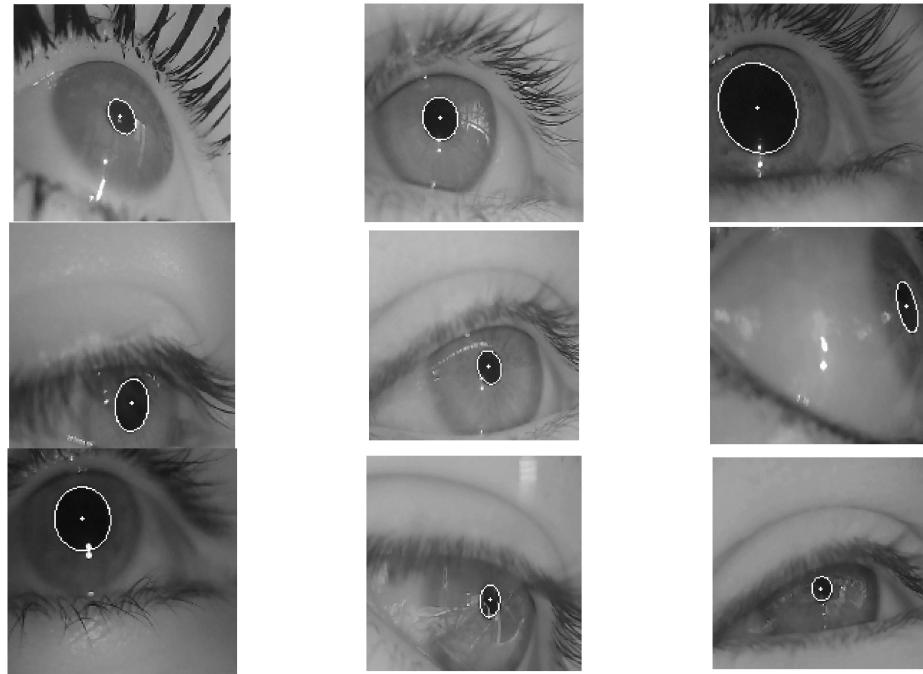

Figure 2. The result of the LPW dataset ellipse fitting.

This study compared the structural similarity index with different intervals of 10, 15, and 20 frames, and the average value of SSIM is shown in Table 1. Finally, this study chose the interval of 15 frames, in which the average value of the structural similarity index was closest to 0.7992 [31].

Table 1. The average value of SSIM at the different intervals of sampling.

Interval of Sampling	Average Value
10	0.8406
15	0.8090
20	0.7781

2.1.2. CASIA-IrisV4-Thousand Dataset

The LPW dataset was taken with the eye tracker located on the edge of glasses, so the oblique image was obtained. In addition, since the pictures of LPW are from videos, photos from the same person may be split into training and test datasets, and we need other datasets to reduce the impact. Therefore, for the second dataset, we added CASIA-Iris-Thousand. These data come from the Chinese Academy of Sciences' Institute of Automation (CASIA), a subset of CASIA-IrisV4 [32]. Additionally, CASIA-Iris-Thousand used an IKEMB-100 camera produced by IrisKing, which captured 1000 subjects to obtain 20,000 high-quality positive iris images. The images of this dataset are shown in Figures 3 and 4 with ellipse fitting.
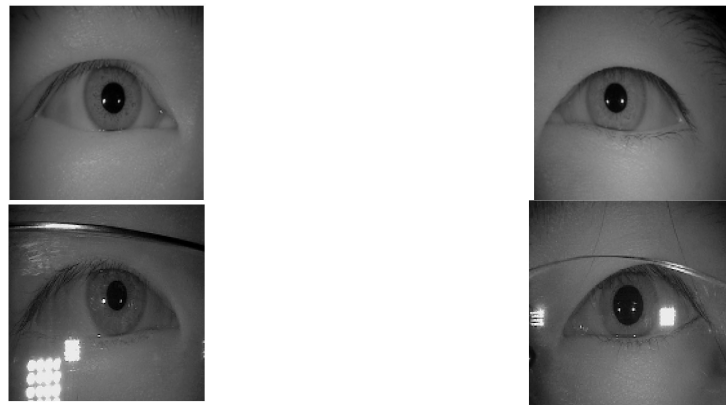
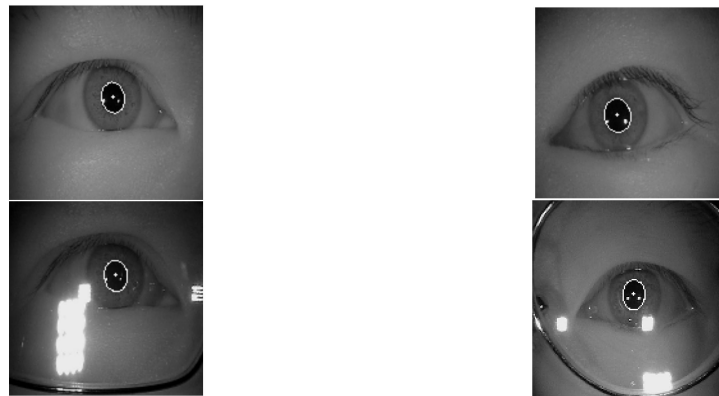**Figure 3.** The sample images of the CASIA-IrisV4-Thousand dataset.

Figure 4. The result of the CASIA-IrisV4-Thousand dataset ellipse fitting.

2.1.3. ŚWirski Dataset

Since the LPW and CASIA-IrisV4-Thousand datasets were pretty different, we chose to add a third dataset. This dataset was provided by Lech Świrski et al. [33]. These data contained four datasets as single images. As with the LPW dataset, both were shot obliquely with the eye tracker on the glasses. Similar to the CASIA-IrisV4-Thousand dataset, the eye position was not exaggerated. It was appropriate to join and neutralize the two datasets. The images of this dataset are shown in Figures 5 and 6 with ellipse fitting.
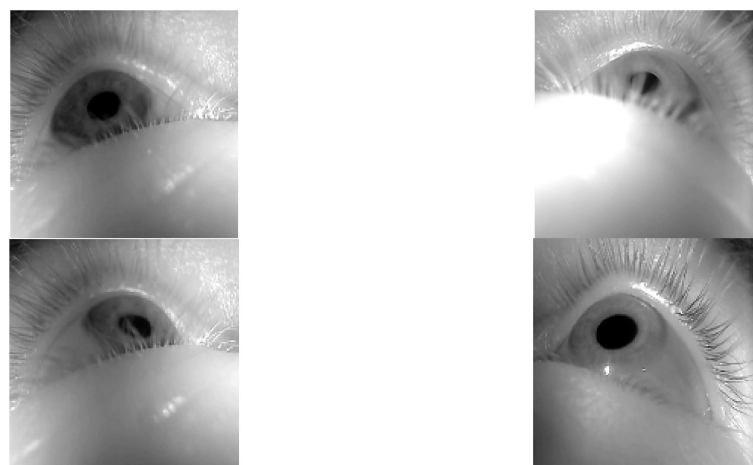

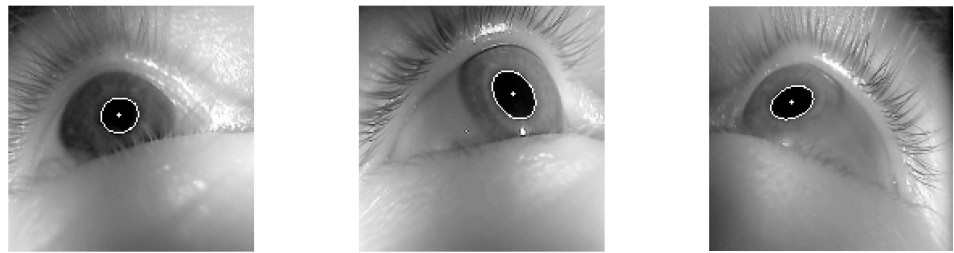**Figure 5.** The sample images of the Świrski dataset.

Figure 6. The result of the Świrski dataset ellipse fitting.

2.1.4. Preprocess Details

In addition to the similar structure index, data augmentation was also applied to ensure that the network could learn more styles during training. For data balance, the dataset will use data augmentation to solve different amounts of data. For data augmentation, this study used scales between 0.7 and 1.3 to produce different pupil sizes. The data were also shifted 10% in each axis to allow the pupil edge to be at the image edge or even beyond the image. Moreover, we used a rotation of $\pm 30°$ and flipped vertically and horizontally to cope with various situations. The items we used in data augmentation will not change the image's label to simplify the data procession. After that, we have 60,000 images, and the size is 152 by 152 pixels. All of them could enhance the richness of the dataset. Finally, the dataset was split into the training set and the testing set in a ratio of 7:3, which would be shuffled before data split.

Of the three datasets, some only provided the label of the pupil center, so we needed a new label for length. In both dark and light environments, human pupils have an average non-circularity of 0.0166. The ellipse's fitting abilities in both the dark and light were 59.6% and 47.7%, respectively, so the fitting shape was set to ellipse [34]. The ground truth of the network was calculated by the pixels on the major and minor axes of the ellipse fitted by a direct least-squares method [35].

2.2. Method and Network Structure

Before training, we divided the dataset into two parts. One was the training set, and the other was used for testing the models after training. As this study was limited by hardware that could not put all of the training data into the network simultaneously, the training set was split into three parts and loaded into the network in stages. That ensured that the training data in each stage were unknown. In addition to different training data, the learning rate of each step was also different. The learning rate was dropped from 1×10^{-3} to 1×10^{-5} at each step, and the batch size was 128. Every stage-trained for 100 epochs and used Adam as the optimizer. The network structure is shown in Figure 7, and the * symbol means multiplication to avoid confusion with the letter x.
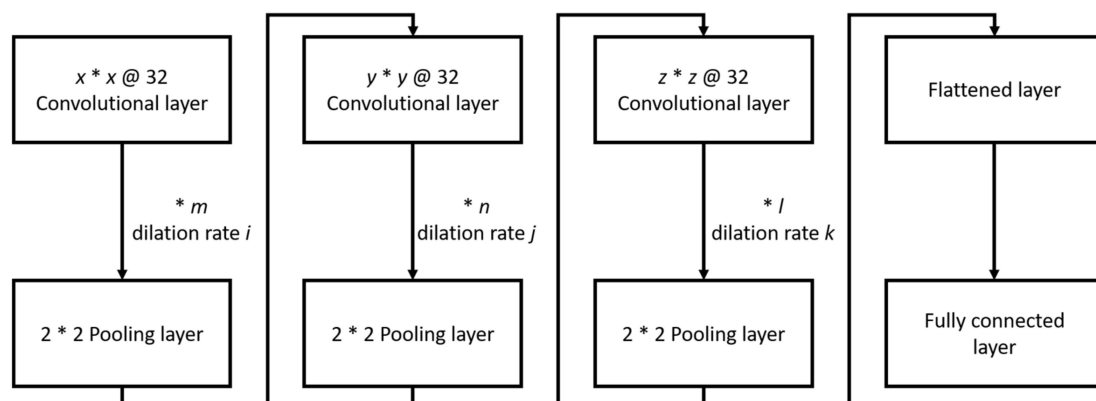

Figure 7. Network structure.

The network parameters are shown in Table 2. Concerning the symbols used, x, y, and z of $K_{(x,y,z)}$ are the size of the three-dimensional filters, and the amount of filters in each layer is fixed at 32. Besides, m, n, and l of $C_{(m,n,l)}$ are the number of convolution layers before the pooling layer, and i, j, and k of D(i, j, k) are the dilation rates of the convolution layer. After the convolution layers, one 2D max-pooling layer reduces noise, and the size is two by two.

Table 2. Network parameters.

		Comparison Type		
		Type I	**Type II**	**Type III**
Network depth	Shallow	$K_{(3,3,3)}C_{(1,1,1)}D_{(1,1,1)}$	$K_{(3,3,3)}C_{(1,1,1)}D_{(3,2,1)}$	$K_{(7,5,3)}C_{(1,1,1)}D_{(1,1,1)}$
	Middle	$K_{(3,3,3)}C_{(4,2,1)}D_{(1,1,1)}$	$K_{(3,3,3)}C_{(4,2,1)}D_{(3,2,1)}$	$K_{(7,5,3)}C_{(4,2,1)}D_{(1,1,1)}$
	Deep	$K_{(3,3,3)}C_{(8,4,2)}D_{(1,1,1)}$	$K_{(3,3,3)}C_{(8,4,2)}D_{(3,2,1)}$	$K_{(7,5,3)}C_{(8,4,2)}D_{(1,1,1)}$

Type I uses regular convolution; Type II uses dilation convolution; Type III uses regular convolution. The filter size has the same FOV as Type II.

The purpose of this study was to achieve a real-time prediction in a low-cost mobile embedded system, so the number of parameters could not be too large. A large model, or too many parameters, would not achieve the required speed using a low-cost mobile embedded system, so this study used dilation convolution to accomplish this purpose. The dilation convolution is shown in Figure 8.
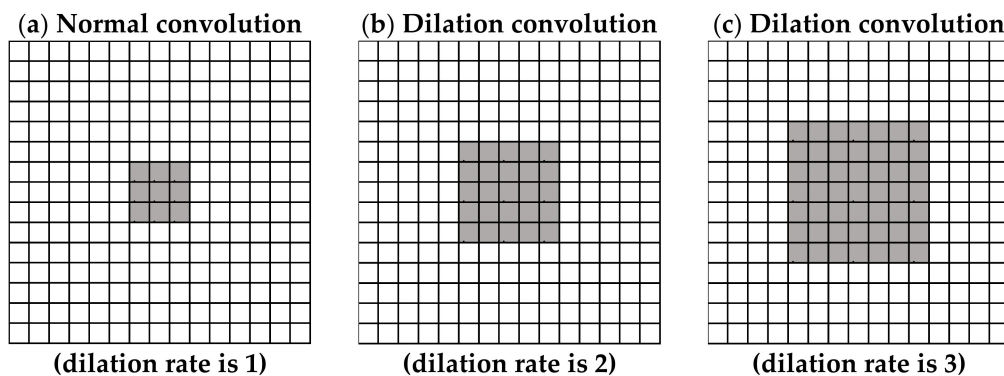

(a) Normal convolution   **(b) Dilation convolution**   **(c) Dilation convolution**

(dilation rate is 1)   **(dilation rate is 2)**   **(dilation rate is 3)**

Figure 8. Diagram convolution with a 3 × 3 filter. The black dot is the position where the convolution occurs, and the gray area is the FOV of the filter. The dilation rates are (**a**) = 1, (**b**) = 2, and (**c**) = 3.

The dilation convolution can enlarge the FOV of the filter without increasing the number of parameters. Widening the FOV can enable a broader range of information to be accepted [36]. Additionally, it can effectively prevent the misjudgment of the network caused by some similar features. This study also compared the network models with the same FOV between the regular filter and the dilated convolution to understand the effect of the decrease in the number of parameters by dilation convolution. Therefore, in addition to three different depth networks in Table 2, there were also three types with varying convolution structures.

3. Results

This study used the test datasets for testing the networks and used the mean error as metrics, which are defined as follows:

$$\text{mean error} = \frac{1}{n}\sum \frac{|\text{prediction(pix)} - \text{ground truth(pix)}|}{\text{ground truth(pix)}} \times 100\% \qquad (2)$$

As the mean error simultaneously calculates the error of the major and minor axes, n is equal to twice the total number of images. The results of the test are shown in Table 3:

Table 3. Mean error of each network.

		Comparison Type		
		Type I	**Type II**	**Type III**
Network depth	Shallow	5.437%	4.549%	4.321%
	Middle	3.442%	2.838%	2.677%
	Deep	2.662%	2.660%	18.165%

The mean error of length of all networks is mostly within 6%, and the best model even reaches 2.660%.

For testing the speed on a low-cost mobile embedded system, this study converted the original Keras model to the TensorFlow Lite model, which can be used for Raspberry Pi to (a) Intel E3-1230 v5 and (b) ARM v8 Cortex-A72 (Raspberry Pi 4 Model B). This study also tested the speed on the original Keras model of (c) Intel E3-1230 v5 and (d) NVIDIA RTX 2080 Ti. The results are shown in Figure 9.
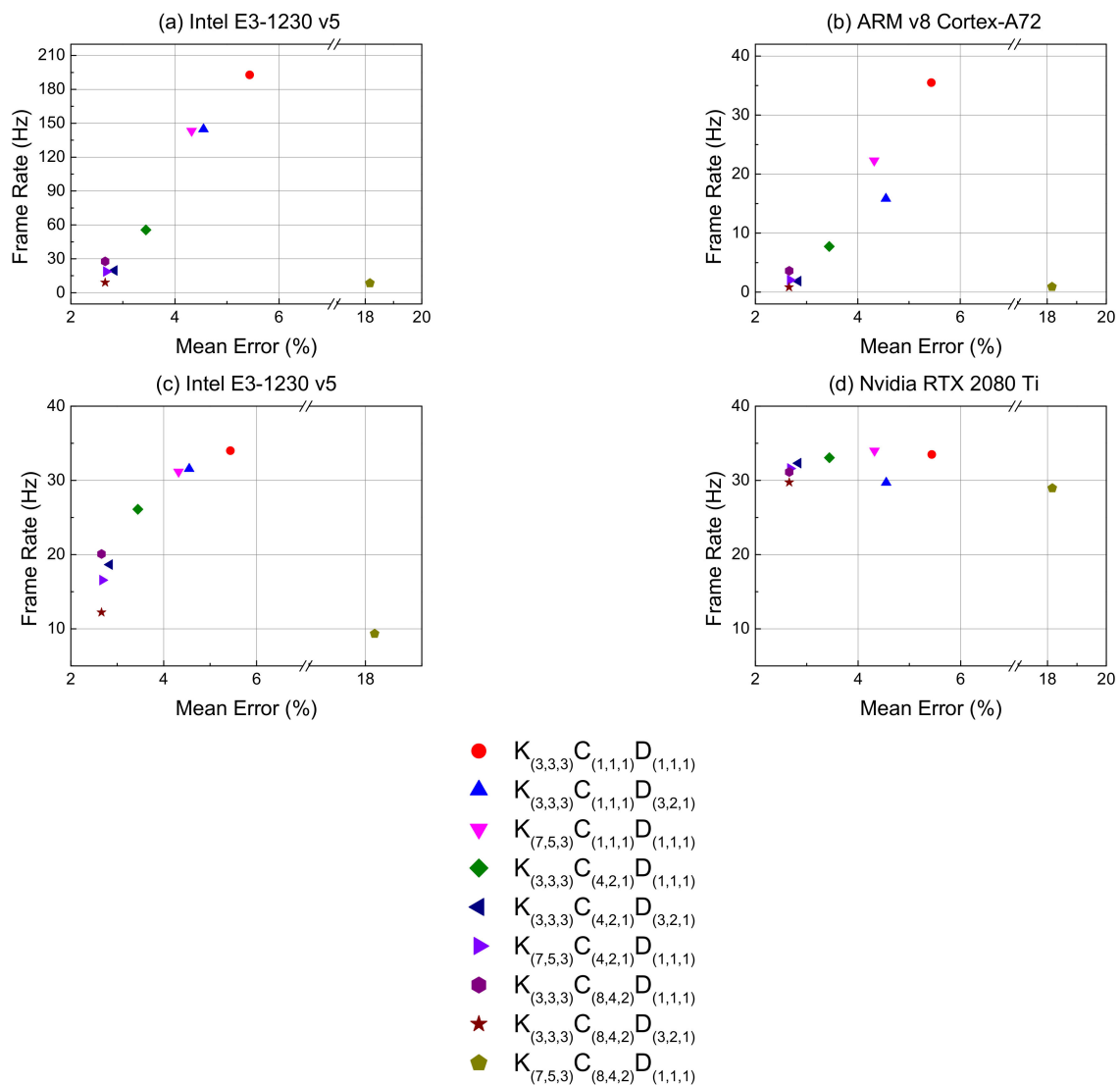

Figure 9. The relationship between frame rate and mean error under different hardware conditions for each network. (**a**,**b**) use the TensorFlow Lite model; (**c**,**d**) use the original Keras model.

On the (a) Intel E3-1230 v5 and (b) ARM v8 Cortex-A72, the highest speeds were around 55 FPS and 35 FPS. On the (c) Intel E3-1230 v5 and (d) NVIDIA RTX 2080 Ti, the highest speed was around 31 FPS and 31 FPS.

This study considered that the real-time calculation was faster than 30 FPS [37] in the low-cost mobile embedded system, so the model with the lowest mean error was $K_{(3,3,3)}C_{(1,1,1)}D_{(1,1,1)}$. Therefore, this study suggests using $K_{(3,3,3)}C_{(1,1,1)}D_{(1,1,1)}$ as the final model for this study. Under the same conditions, we recommend choosing $K_{(3,3,3)}C_{(4,2,1)}D_{(1,1,1)}$ and $K_{(3,3,3)}C_{(8,4,2)}D_{(3,2,1)}$ for (c) Intel E3-1230 v5 and (d) NVIDIA RTX 2080 Ti as the suggested model for different hardware.

4. Discussion

4.1. Model Evaluation

Although the model can reach a maximum of 35 FPS in a low-cost mobile embedded system, its mean error is too high. In addition, the model with the lowest mean error of 2.660% is too slow. When the pupil size changes, it may not detect all changes in pupil size in real-time because the calculation speed is insufficient. As such, this study used 30 FPS as the condition.

In addition to the length of major and minor axes, the area can also express the pupil size. The equation is as follows:

$$\text{Ellipse Area} = a \times b \times \pi \tag{3}$$

where a is the semi-major axis, and b is the semi-minor axis. In this study, the mean error of the length is not counted separately for the major and minor axes, so the ellipse area equation is used with the mean error of length. In the recommended model, the mean error of the length and area are 5.437% and 10.57%, respectively. Besides, the comparison with previous research [16] is shown in Table 4.

Table 4. The comparison with previous research.

	The Recommended Model	The Previous Research
Mean Error	5.437%	6.587%

4.2. Feature Map Visualization

This paper uses the GRAD-CAM [38] method to calculate its weight (w) by backpropagation to confirm the relevance of the feature map and the parameters. It can be observed from the image that the area of interest of the network architecture is the pupil area, as shown in Figure 10.

Figure 10. The feature map visualization.

4.3. Model Speed Trend

Figure 11 shows the relationship between the frame rate of each model and its number of parameters. According to the results, the frame rate decreases as the number of parameters increases, and its fitting curve shows an exponential trend. The coefficient of determination (R^2) of the fitting curve in Figure 11 is 0.8168, and it can be concluded that the frame rate and the number of parameters are highly correlated. Therefore, this study

tried to reduce the number of parameters by dilation convolution, which is conducive to improving the network's speed.
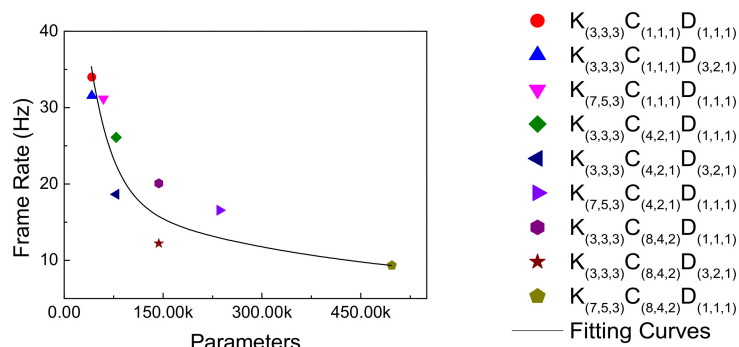

Figure 11. Trend curves of frame rate and parameters.

4.4. Model Comparison

Figure 12 shows the effect of changing the convolution filter at the same depth. Both Type II and Type III can effectively reduce the mean error in most cases. However, the speed decreases. The declining rate on Type II models is more considerable at the shallower structure than Type III, but the difference between the two becomes smaller in the middle, as shown in Figure 12a,b. In Figure 12c, the Type III model may have too many useless features leading to errors in judgment, although the convolution filter has a larger FOV. Therefore, the overall mean error is higher than Type II.
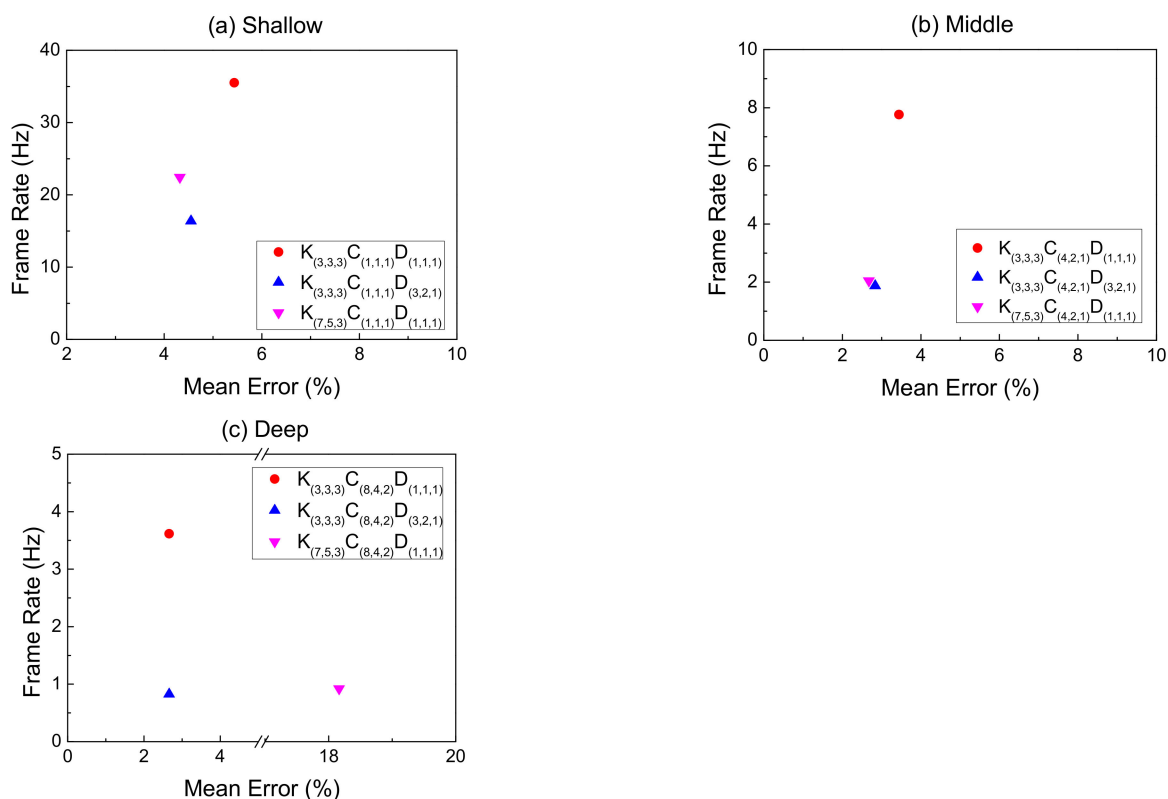

Figure 12. The effect of the filter's FOV changes on speed and mean error with (**a**) Shallow structure, (**b**) Middle structure, (**c**) Deep structure.

4.5. Model Revision

From the experimental results, this study found that in the Type II and Type I models with the same parameters, while the mean error of Type II is significantly reduced the calculation speed is slightly reduced. That is probably due to the fact that the Type II model has a larger FOV than Type I. Although the part with more FOV seems to skip calculation when performing convolution operations, the convolution kernel size is not reduced. The neglected part can be regarded as a fixed-zero weight calculation, so the part will not be trained, and the number of parameters will not increase. However, complementing the part where the weight of the filter is regarded as zero will create too many parameters. The larger FOV of the convolution kernel will decrease operating efficiency, resulting in a slower speed. In addition to wasting performance, it is also easy to affect the prediction because of some useless features.

This study conducted another experiment to account for the suggestion that the number of parameters causes the mean error to rise. This study tried to reduce the number of parameters in the deepest network of Type III. As the depth of the network was fixed, we adjusted the number of filters in each layer. We then changed the 32 filters to use between 4 and 10 filters, and the mean error was reduced to within 10%. However, if more than 11 convolution filters were used, the mean error would not decrease significantly. According to the results, the mean error will reduce by lowering the number of parameters, as expected. Therefore, due to too many network parameters, the deepest network of Type III will cause the mean error to fail to achieve the desired result, resulting in serious misjudgment.

5. Conclusions

This paper proposes an algorithm that can calculate pupil size that is based on the CNN model. This study compared the mean error with different network depths and FOV of the convolution filter for optimizing the structure. According to the results, both deepening the network and widening the FOV can effectively reduce the mean error. However, both will reduce speed. For the ARM v8 Cortex-A72 (RPi 4B), we recommend the model be used at 35 FPS. The mean error of the pupil length is 5.437%, and the pupil area is 10.57%. The model can calculate, in real-time, pupil size in a low-cost mobile system. It will help doctors and nurses to obtain quantified results. This study also provides models with lower mean error under different hardware conditions and respective improvements in calculation speeds.

We will make an in-depth study to optimize the neural network structure of the system to reduce the computational complexity and improve pupil detection speed in future works. For the algorithm to identify pupils in different fields, the richness of the dataset will be increased to enhance the generalization of real situations. We also need to consider that we split the training and test sets after the data augmentation may allow the model to peek at the test data during the training process and cause deviation. Besides, arranging clinical experiments to compare with the results of machines and doctors is also necessary.

Author Contributions: Conceptualization, A.J.-W.W. and Y.-Y.C.; methodology, W.-C.T. and Y.-P.C.; software, W.-C.T. and C.-H.T.; validation, C.-H.Y. and C.-H.L.; formal analysis, C.-H.T. and X.Z.; investigation, Y.-P.C.; resources, C.-H.Y.; data curation, C.-H.Y. and C.-H.L.; writing—original draft preparation, C.-H.T.; writing—review and editing, Y.-Y.C.; visualization, W.-C.T.; supervision, A.J.-W.W.; project administration, Y.-Y.C.; funding acquisition, Y.-Y.C. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data of LPW was obtained from Tonsen, M., etc., and are available from http://doi.org//10.1145/2857491.2857520 with openly available. The data of CASIA-IrisV4 was obtained from the Chinese Academy of Sciences' Institute of Automation (CASIA) and available from http://www.cbsr.ia.ac.cn/china/Iris%20Databases%20CH.asp with permission of the Center for Biometrics and Security Research (CBSR). The data of Świrski was obtained from Świrski, L., etc., and are available from http://www.cl.cam.ac.uk/research/rainbow/projects/pupiltracking/ with openly available.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Xue, T.; Do, M.T.H.; Riccio, A.; Jiang, Z.; Hsieh, J.; Wang, H.C.; Merbs, S.L.; Welsbie, D.S.; Yoshioka, T.; Weissgerber, P.; et al. Melanopsin signalling in mammalian iris and retina. *Nat. Cell Biol.* **2011**, *479*, 67–73. [CrossRef]
2. Kret, M.E.; Sjak-Shie, E.E. Preprocessing pupil size data: Guidelines and code. *Behav. Res. Methods* **2019**, *51*, 1336–1342. [CrossRef] [PubMed]
3. Bitsios, P.; Prettyman, R.; Szabadi, E. Changes in Autonomic Function with Age: A Study of Pupillary Kinetics in Healthy Young and Old People. *Age Ageing* **1996**, *25*, 432–438. [CrossRef] [PubMed]
4. Canver, M.C.; Canver, A.C.; Revere, K.E.; Amado, D.; Bennett, J.; Chung, D.C. Novel mathematical algorithm for pupillometric data analysis. *Comput. Methods Programs Biomed.* **2014**, *113*, 221–225. [CrossRef] [PubMed]
5. Lu, W.; Tan, J.; Zhang, K.; Lei, B. Computerized mouse pupil size measurement for pupillary light reflex analysis. *Comput. Methods Programs Biomed.* **2008**, *90*, 202–209. [CrossRef] [PubMed]
6. Jain, M.; Devan, S.; Jaisankar, D.; Swaminathan, G.; Pardhan, S.; Raman, R. Pupillary Abnormalities with Varying Severity of Diabetic Retinopathy. *Sci. Rep.* **2018**, *8*, 5363. [CrossRef] [PubMed]
7. Rukmini, A.V.; Milea, D.; Baskaran, M.; How, A.C.; Perera, S.A.; Aung, T.; Gooley, J.J. Pupillary Responses to High-Irradiance Blue Light Correlate with Glaucoma Severity. *Ophthalmology* **2015**, *122*, 1777–1785. [CrossRef]
8. Chang, D.S.; Boland, M.V.; Arora, K.S.; Supakontanasan, W.; Chen, B.B.; Friedman, D.S. Symmetry of the Pupillary Light Reflex and Its Relationship to Retinal Nerve Fiber Layer Thickness and Visual Field Defect. *Investig. Opthalmol. Vis. Sci.* **2013**, *54*, 5596–5601. [CrossRef]
9. Reutrakul, S.; Crowley, S.J.; Park, J.C.; Chau, F.Y.; Priyadarshini, M.; Hanlon, E.C.; Danielson, K.K.; Gerber, B.S.; Baynard, T.; Yeh, J.J.; et al. Relationship between Intrinsically Photosensitive Ganglion Cell Function and Circadian Regulation in Diabetic Retinopathy. *Sci. Rep.* **2020**, *10*, 1560. [CrossRef]
10. Larson, M.D.; Behrends, M. Behrends, and Analgesia, Portable infrared pupillometry: A review. *Anesth. Analg.* **2015**, *120*, 1242–1253. [CrossRef]
11. O'Neill, W.; Trick, K. The narcoleptic cognitive pupillary response. *IEEE Trans. Biomed. Eng.* **2001**, *48*, 963–968. [CrossRef] [PubMed]
12. Yoo, Y.J.; Yang, H.K.; Hwang, J.-M. Efficacy of digital pupillometry for diagnosis of Horner syndrome. *PLoS ONE* **2017**, *12*, e0178361. [CrossRef] [PubMed]
13. Adhikari, P.; Zele, A.J.; Feigl, B. The Post-Illumination Pupil Response (PIPR). *Investig. Opthalmol. Vis. Sci.* **2015**, *56*, 3838–3849. [CrossRef] [PubMed]
14. Mitz, A.R.; Chacko, R.V.; Putnam, P.T.; Rudebeck, P.H.; Murray, E.A. Using pupil size and heart rate to infer affective states during behavioral neurophysiology and neuropsychology experiments. *J. Neurosci. Methods* **2017**, *279*, 1–12. [CrossRef]
15. Wang, C.-A.; Baird, T.; Huang, J.; Coutinho, J.D.; Brien, D.C.; Munoz, D.P. Arousal Effects on Pupil Size, Heart Rate, and Skin Conductance in an Emotional Face Task. *Front. Neurol.* **2018**, *9*, 1029. [CrossRef] [PubMed]
16. Garcia, R.G.; Avendano, G.O.; Agdeppa, D.B.F.; Castillo, K.J.; Go, N.R.S.; Mesina, M.A. Automated Pupillometer Using Edge Detection in OpenCV for Pupil Size and Reactivity Assessment. In Proceedings of the 2019 3rd International Conference on Imaging, Signal Processing and Communication (ICISPC), Singapore, 27–29 July 2019; pp. 143–149.
17. Fuhl, W.; Rosenstiel, W.; Kasneci, E. 500,000 Images Closer to Eyelid and Pupil Segmentation. In *Transactions on Petri Nets and Other Models of Concurrency XV*; Springer Science and Business Media LLC.: Berlin, Germany, 2019; pp. 336–347.
18. Miron, C.; Pasarica, A.; Bozomitu, R.G.; Manta, V.; Timofte, R.; Ciucu, R. Efficient Pupil Detection with a Convolutional Neural Network. In Proceedings of the 2019 E-Health and Bioengineering Conference (EHB), Iasi, Romania, 21–23 November 2019; pp. 1–4.
19. Vera-Olmos, F.; Pardo, E.; Melero, H.; Malpica, N. DeepEye: Deep convolutional network for pupil detection in real environments. *Integr. Comput. Eng.* **2018**, *26*, 85–95. [CrossRef]
20. Yiu, Y.-H.; Aboulatta, M.; Raiser, T.; Ophey, L.; Flanagin, V.L.; zu Eulenburg, P.; Ahmadi, S.-A. DeepVOG: Open-source pupil segmentation and gaze estimation in neuroscience using deep learning. *J. Neurosci. Methods* **2019**, *324*, 108307. [CrossRef] [PubMed]
21. Fuhl, W.; Santini, T.; Kasneci, G.; Rosenstiel, W.; Kasneci, E. Pupilnet v2.0: Convolutional neural networks for cpu based real time robust pupil detection. *arXiv* **2017**, arXiv:1711.00112.
22. Vera-Olmos, F.J.; Malpica, N. Deconvolutional Neural Network for Pupil Detection in Real-World Environments. In *Transactions on Petri Nets and Other Models of Concurrency XV*; Springer Science and Business Media LLC.: Berlin, Germany, 2017; pp. 223–231.

23. Fuhl, W.; Santini, T.C.; Kübler, T.; Kasneci, E. Else: Ellipse selection for robust pupil detection in real-world environments. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*; Association for Computing Machinery: New York, NY, USA, 2016; pp. 123–130.

24. De Souza, J.K.S.; Pinto, M.A.D.S.; Vieira, P.G.; Baron, J.; Tierra-Criollo, C.J. An open-source, FireWire camera-based, Labview-controlled image acquisition system for automated, dynamic pupillometry and blink detection. *Comput. Methods Programs Biomed.* **2013**, *112*, 607–623. [CrossRef]

25. De Santis, A.; Iacoviello, D. Optimal segmentation of pupillometric images for estimating pupil shape parameters. *Comput. Methods Programs Biomed.* **2006**, *84*, 174–187. [CrossRef] [PubMed]

26. Tabashum, T.; Zaffer, A.; Yousefzai, R.; Colletta, K.; Jost, M.B.; Park, Y.; Chawla, J.; Gaynes, B.; Albert, M.V.; Xiao, T. Detection of Parkinson's Disease Through Automated Pupil Tracking of the Post-illumination Pupillary Response. *Front. Med.* **2021**, *8*, 645293. [CrossRef] [PubMed]

27. Navaneethan, S.; Nandhagopal, N. RE-PUPIL: Resource efficient pupil detection system using the technique of average black pixel density. *Sādhanā* **2021**, *46*, 114. [CrossRef]

28. Kim, T.; Lee, E.C. Experimental Verification of Objective Visual Fatigue Measurement Based on Accurate Pupil Detection of Infrared Eye Image and Multi-Feature Analysis. *Sensors* **2020**, *20*, 4814. [CrossRef] [PubMed]

29. Xiang, Y.; Zhao, L.; Liu, Z.; Wu, X.; Chen, J.; Long, E.; Lin, D.; Zhu, Y.; Chen, C.; Lin, Z.; et al. Implementation of artificial intelligence in medicine: Status analysis and development suggestions. *Artif. Intell. Med.* **2020**, *102*. [CrossRef]

30. Tonsen, M.; Zhang, X.; Sugano, Y.; Bulling, A. Labelled pupils in the wild: A dataset for studying pupil detection in unconstrained environments. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*; Association for Computing Machinery: New York, NY, USA, 2016; pp. 139–142.

31. Zhang, J.; Liang, X.; Wang, M.; Yang, L.; Zhuo, L. Coarse-to-fine object detection in unmanned aerial vehicle imagery using lightweight convolutional neural network and deep motion saliency. *Neurocomputing* **2020**, *398*, 555–565. [CrossRef]

32. Portions of the Research in This Paper Use the CASIA-IrisV3 Collected by the Chinese Academy of Sciences' Institute of Automation (CASIA) and a Reference to CASIA Iris Image Database. Available online: http://biometrics.idealtest.org/ (accessed on 25 November 2020).

33. Świrski, L.; Bulling, A.; Dodgson, N. Robust real-time pupil tracking in highly off-axis images. In *Proceedings of the Symposium on Applied Computing*; Association for Computing Machinery: New York, NY, USA, 2012; p. 173.

34. Wyatt, H.J. The form of the human pupil. *Vis. Res.* **1995**, *35*, 2021–2036. [CrossRef]

35. FitzGibbon, A.; Pilu, M.; Fisher, R. Direct least square fitting of ellipses. *IEEE Trans. Pattern Anal. Mach. Intell.* **1999**, *21*, 476–480. [CrossRef]

36. Yu, F.; Koltun, V. Multi-scale context aggregation by dilated convolutions. *arXiv* **2015**, arXiv:1511.07122.

37. Rickmann, A.; Waizel, M.; Kazerounian, S.; Szurman, P.; Wilhelm, H.; Boden, K.T. Digital Pupillometry in Normal Subjects. *Neuro Ophthalmol.* **2016**, *41*, 12–18. [CrossRef]

38. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 618–626.