Automating, Analyzing and Improving Pupillometry with Machine Learning Algorithms

György Kalmár[a], Alexandra Büki[b], Gabriella Kékesi[b], Gyöngyi Horváth,[b] and László G. Nyúl[a]

Abstract

The investigation of the pupillary light reflex (PLR) is a well-known method to provide information about the functionality of the autonomic nervous system. Pupillometry, a non-invasive technique, was applied to study the PLR alterations in a new, schizophrenia-like rat substrain, named WISKET. The pupil responses to light impulses were recorded with an infrared camera; the videos were automatically processed and features were extracted from the pupillograms. Besides the classical statistical analysis (ANOVA), feature selection and classification were applied to reveal the significant differences in the PLR parameters between the control and WISKET animals. Based on these results, the disadvantages of this method were analyzed and the measurement setup was redesigned and improved. The pupil segmentation method has also been adapted to the new videos. 2564 images were annotated manually and used to train a fully-convolutional neural network to produce pupil mask images. The method was evaluated on 329 test images and achieved 4% median relative error. With the new setup, the pupil detection became reliable and the new data acquisition offers robustness to the experiments.

Keywords: pupillometry, classification, curve properties, U-Net

1 Introduction

Patients with schizophrenia, besides the well-known behavioral symptoms, also show autonomic dysregulation, including impaired pupillary function, which is a sensitive and reliable source of information about the function of the nervous system [1].

Pupillometry is a simple, non-invasive technique for the assessment of the autonomic nervous system function by testing the pupillary light reflex (PLR), meaning

[a]Department of Image Processing and Computer Graphics, Faculty of Science and Informatics, University of Szeged, E-mail: {kalmargy,nyul}@inf.u-szeged.hu

[b]Department of Physiology, Faculty of Medicine, University of Szeged, E-mail: {buki.alexandra,kekesi.gabriella,horvath.gyongyi}@med.u-szeged.hu

the contraction of the pupil in response to light. During the test, the changes of the pupil diameter are recorded and the size of the pupil is measured offline in each video frame, producing the *pupillogram.*

The measurement of the diameter in each frame manually is a labor-intensive and slow process. Automated, software-based methods can speed-up the detection and improve the analyzing process.

Our recent research investigated the PLR to reveal the schizophrenia-related alterations in the autonomic nervous system of WISKET rats and the related *medical* results were published in [3]. Contributions in the current work: the explanation of novel pupillogram features; presentation of a decision tree based classifier and the discussion of the results; introduction of a redesigned measurement setup and process; description and evaluation of the new pupil segmentation method.

Related works are summarized in Section 2. The experiments, the novelties of the feature extraction method and data analysis are described in Section 3. In Section 4 the results of the analysis and classification are presented and discussed. Section 5 describes the redesigned measurement setup and a manually annotated pupil segmentation dataset. A deep learning based pupil detection algorithm and the detection results on a test dataset are shown in Section 6. Section 7 concludes the paper.

2   Related works

Developing reliable and predictive animal models for any complex psychiatric diseases, such as schizophrenia, is essential to understand the neurobiological basis of the disorder. Recently, a new, selectively bred rat model of schizophrenia has been developed, named WISKET [7, 8, 11, 19].

Clinical studies in schizophrenic patients using pupillometry revealed impaired autonomic regulation [1, 2, 6, 13, 22]. PLR was also investigated in rodents [16]; however, only our recent study provided data from animal models of schizophrenia [3].

Speeding-up pupillometry with automated softwares is essential because manual methods are slow and labor-intensive. In cases, when the image quality is acceptable, circle or ellipse detection algorithms can be used [15, 21]. These methods rely on simple techniques, e.g. the Hough-transform with an ellipse model. When low contrast and reflections occur, complex and more sophisticated algorithms are needed [14, 23]. These solutions combine several simple methods with specialized extensions, e.g. in [14], the simple Hough-transform based ellipse detection was extended with randomization and was implemented in an iterative scheme to filter out the noise and to handle outliers. The recent rapid development of artificial intelligence and machine learning led to more accurate solutions to the pupil detection problem [4, 5]. Among these methods, the convolutional neural networks are particularly interesting as they can treat their inputs as images. When the expected outputs are binary masks, which is a common goal in biomedical applications, fully-convolutional neural networks are used. They utilize 'de-convolutional' layers to

up-sample and combine feature maps with low resolutions to binary masks having the same sizes as the input images. The most popular such a network structure, the U-Net, was introduced in [20].

3   Automated pupillometry

3.1   Data acquisition

As it was described recently [3], two series of experiments were performed in sedated (n=54) and anesthetized (n=20) control Wistar, and WISKET rats. After a 10-minute long dark adaptation period, the recordings lasted for $15s$ in sedated, and for $60s$ in anesthetized animals. The animals were positioned close to a camera, and an intensive visible light stimulus (approx. $300cd/m^2$ for $600ms$) was flashed into their left eyes. The IR-camera recorded pupillary responses at a speed of 24 frames-per-second under infrared illumination.

During the PLR measurements, the rats showed several minor movements, which affected the quality of the recorded videos. Furthermore, albino rats lack pigments in their body including their eyes, which reduces the contrast between the iris and the pupil. A specifically designed pupil detection and measurement algorithm was used to handle these quality drawbacks [10]. The input of the algorithm was the video recording; and the output was a curve of the determined relative pupil diameters in each frame - the pupillogram. The relative pupil diameter is the ratio of the pupil and iris diameters, expressed in percentage.

3.2   Feature extraction

To compare the responses of the animals, descriptive features from the PLR curves were extracted, which are relevant from the pathophysiological point of view and suitable to emphasize the differences between the groups regarding the autonomic nervous system activities.

An automated feature extraction method was designed, which produced 40 features from the pupillogram. Many of them were basic, traditional parameters like the initial diameter, which is the size of the pupil before the light impulse; minimum diameter; reaction time; maximum of the redilated diameter; etc. Several new features were implemented to obtain information about the dynamics of the response, thus 11 velocity related descriptors were introduced including, the average and the maximal contraction velocities and the times required to reach the latter. The velocities of the redilation phase at different time points were also calculated.

Novel, smoothness related descriptors were introduced, as well. A polynomial curve with a given order was fitted to the redilation part of the response. The area between the original and fitted curves served as a measure of non-smoothness. Fifth order polynomials were chosen, which were flexible enough to follow the slow perturbations of the original curve and indicated only the short, abnormal swings. In Figure 1, a representative response curve and a marked subset of the extracted features are presented.
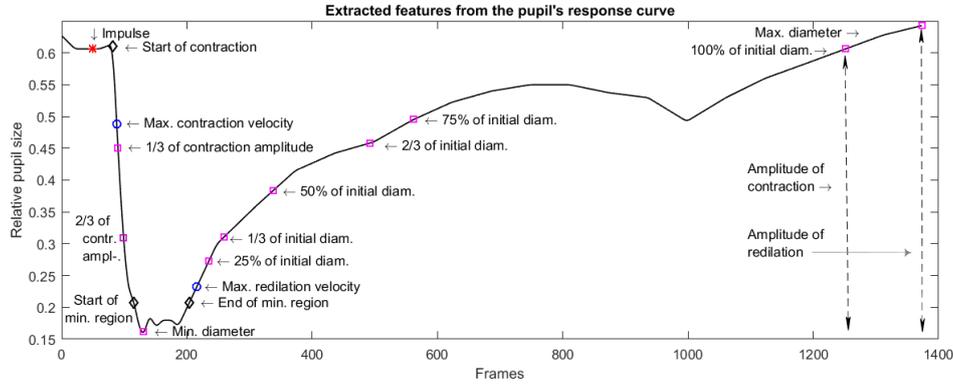
Figure 1: Pupillary light response curve and a marked subset of the 40 features.

3.3 Data analysis

The sedated and the anesthetized animals were analyzed separately.

One-way ANOVA was used for the analysis of differences between the control and test groups. The relationships between pupil parameters were assessed by linear regression analysis and calculation of the Pearson correlation coefficient. Only probabilities lower than 0.05 were considered significant.

Besides the investigation of the differences in PLR parameters, the other goal was to use pupillometry as a quick examination to facilitate the selection of the animals during the breeding process. Therefore, a binary decision tree was trained to investigate the possibility of classification and the model was evaluated by using cross-validation.

4 Evaluation

4.1 Classification results

The detailed discussion of the results of the statistical analysis is found in our recent study [3]. In accordance with these results, the trained decision tree selected almost the same features as predictor variables as the statistical analysis suggested.

In the sedated group, which has the greater cardinality, the fitted decision tree achieved 71% accuracy measured by cross-validation. The algorithm selected the following predictors: minimum diameter (relative pupil size); initial diameter (relative pupil size), and average redilation speed (change of relative pupil size/frame time). Figure 2 shows the fitted decision tree.

With the combination of the two analysis methods (statistical analysis, decision tree), significant differences were noticed between the control and test groups. The initial and minimum pupil diameters were larger and the degree of the constriction was lower in the WISKET rats. The flatness of the curve (length of the minimum
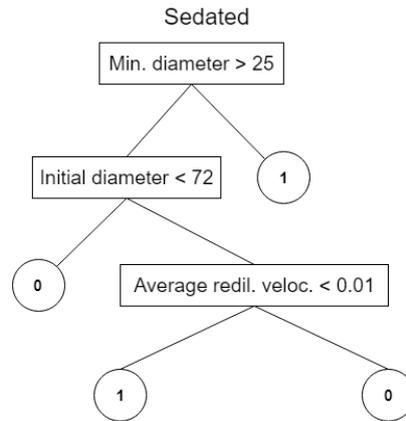
Sedated

Figure 2: Decision tree fitted to the sedated group data. Leaf nodes containing 0 correspond to control group and nodes with values 1 correspond to the test group.

region) and the contraction time were shorter in the control group. These results are in accordance with previous studies mentioned in Section 2.

The analysis of the anesthetized animals showed that they cannot be divided into two classes reliably. It is assumed that while anesthesia can prevent stress and allows a convenient investigation of pupillary reactions for a longer period, it also diminishes the differences between the two groups in this autonomic response. The classifier achieved only 60% accuracy. The selected predictors were the amplitude of contraction, the average redilation speed and the time required to reach the maximal redilation speed.

4.2   Discussion

The results showed that anesthetized animals cannot be used in these experiments, so the attention was focused on the sedated animals. However, the sedated animals were still able to move their heads during the experiments, which affected negatively the recording of the pupil with the closely placed camera.

Based on the decision tree analysis, the initial diameter and the minimum diameter seemed to be the most reliable features for the classification, although the dynamic, time-related features seemed to be less relevant (or more samples (rats) are required to detect the potential differences).

From these observations, it could be concluded that a more robust measurement process was required, which had a relatively short period and provided multiple light stimuli to induce reoccurring reflex responses and minimum diameter occasions. thus enabling a complex and more detailed analysis.

5 Improved pupillometry

5.1 Improved data acquisition

To improve the robustness of the recordings, the measurement setup was redesigned. Around the camera lens, an IR LED (infrared light-emitting diode) ring was attached, setting the camera and the illuminating IR LEDs (nearly) on the same optical axis. With this setting, the camera is able to detect the light reflected from the retina, causing the so-called 'Bright pupil effect' [17], and the camera can be placed farther from the animal, thus tiny animal movements do not affect the recording quality. In Figure 3 the differences between the old and new setups are presented.

A new embedded system was also developed to schedule the experiments, which is controlled from a graphical user interface on the PC to initiate different visible light impulse sequences. This scheduler implements accurate time synchronization and manages the timing of light stimuli. The system uses an RGB LED, which allows us to produce different light intensities and colors.

With these improvements, the signal-to-noise ratio was enhanced (the pupil is better detectable) in these images and the recordings are more robust and precise. The improved image quality can be observed in Figure 3.b.
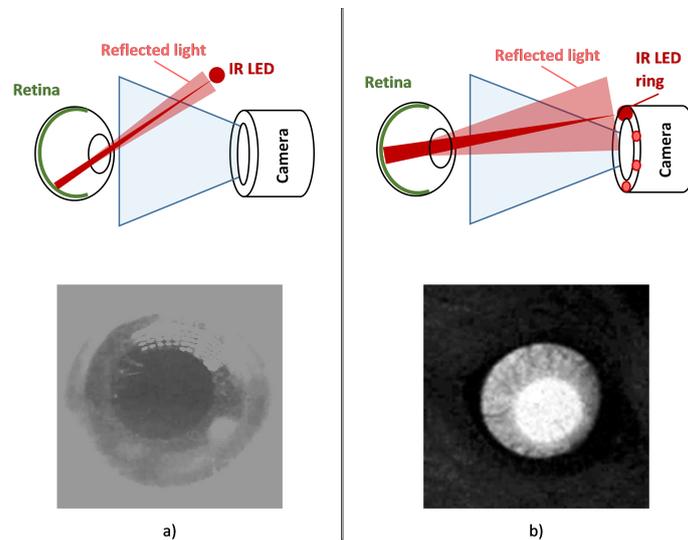
Figure 3: Difference between the old (a) and new (b) measurement setups. a) The camera is close to the eye and the illuminating infrared LED placed far from the camera's optical axis. The camera records black pupil region (reflected light never reaches the camera). b) The camera is placed farther from the eye and the illuminating LEDs are close to the optical axis. The camera records bright pupil region (reflected light reaches the camera).

5.2   Pupil segmentation dataset

The images recorded with the new measurement setup have a completely different nature, as it can be seen in Figure 3.b. Therefore, our previously developed method [10] cannot be used (different intensity levels, different resolution, etc.). To support the development of a new pupil segmentation algorithm and to validate the new setup, 56 experimental videos were recorded each containing more than 5000 frames. From these videos, 2564 randomly chosen frames were manually annotated (ellipses manually fitted to the pupil regions). An additional set of 329 images were selected and annotated to form a challenging test dataset.

Traditional image preprocessing methods were implemented to detect the eye region in the image, crop it and enhance the contrast. The preprocessed dataset contains 128×128 pixels-sized images and the corresponding binary pupil masks. Figure 4 shows some samples and the corresponding pupil masks from the pupil segmentation dataset.

The dataset is publicly available [9].
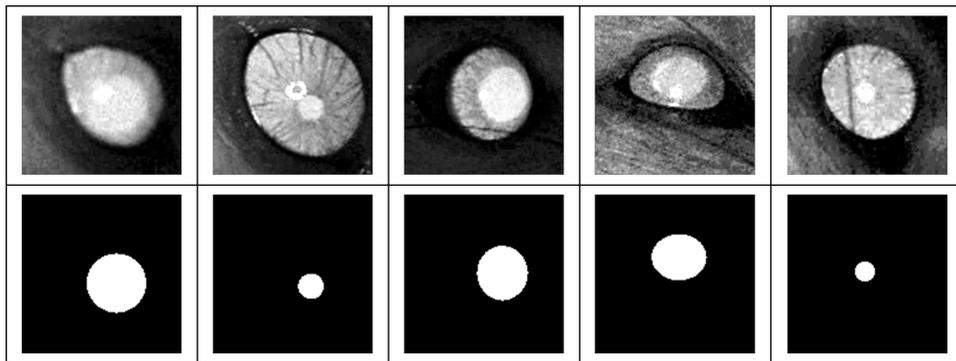
Figure 4: Example samples from the pupil segmentation dataset. Top row: original samples. Bottom row: corresponding pupil mask images.

6   Pupil segmentation

6.1   Pupil segmentation using neural network

A modern way to solve a segmentation problem is to use neural networks. Fully-convolutional networks are frequently used in segmentation problems as was explained in Section 2. In this paper, a U-shaped structure designed for biomedical applications, the U-Net structure [20] was used to perform the pupil segmentation task – the calculation of binary pupil masks. The training of the network required manually annotated data.

Our dataset contains input images with 128×128 resolutions. The originally published version on the U-Net worked with different input sizes, thus adaptations

were required. Three layers (each performs convolution and max-pooling) were applied and the number of channels was doubled after each pooling, as the original paper suggested. At the bottom of the U-shaped structure, 512 pieces of 16×16 sized feature maps were calculated. The abstract visualization of the used structure can be observed in Figure 5. Instead of random initialization, the "de-convolution" layers' weights were initialized to perform bilinear up-sampling. To compare the predicted and ground-truth pupil masks, binary cross-entropy was used as the loss function. Adam method [12] was used as the optimizer and the batch size was 64.

Data augmentation was employed to prevent model overfitting (random flips, random crop and resize). No other regularization techniques were used. The training process was optimized by the analysis of the training loss and validation loss curves. 10% of the training dataset served as the validation set. Early-stopping was applied to terminate the training process when the validation loss stopped decreasing. This setup was tested with different learning rates. The best performing model ran for 200 epochs, with the learning rate of 0.001. The algorithm was implemented in PyTorch [18].

The output of the neural network was an almost binary image. The pixel values were close to 1.0 when a pixel was considered as part of the pupil. The output was binarized with an empirically set threshold value. In the resulting binary images, contours were found and ellipses were fitted on these points. The final output was the mask of the fitted ellipse. If multiple ellipses were identified, the false detections were filtered out by simple rules based on the sizes, shapes, and locations of the ellipses. A set of example images and the predicted ellipses (without false detection filtering) are presented in Figure 6.
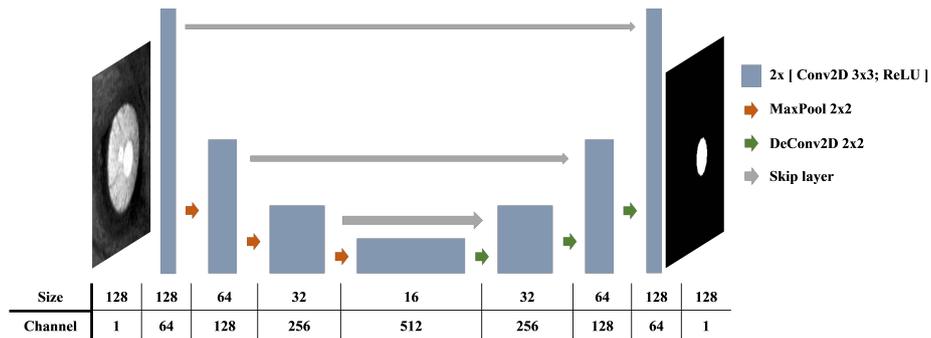
Figure 5: The neural network structure similar to [20].
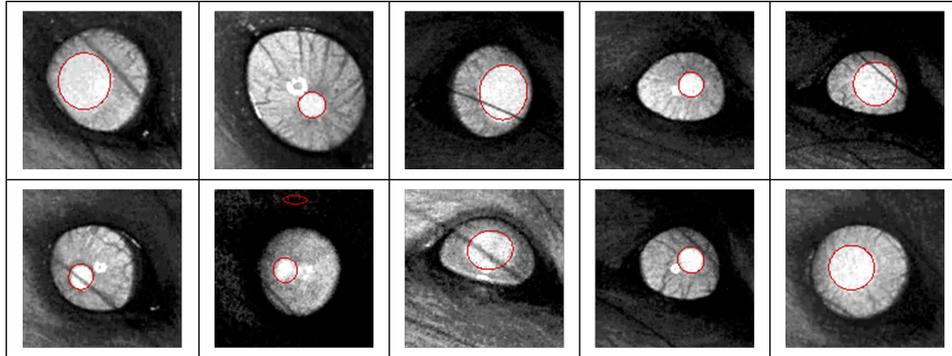
Figure 6: Example result images. Red ellipses present the predicted ellipses' borders.

6.2 Results

The trained model was evaluated on the test dataset containing 329 images. The Jaccard-index (Intersection of Union) was determined to compare the ground-truth and predicted pupil masks. The average Jaccard-index value on the test dataset was 0.815. This test set is challenging and there were samples, which were completely misclassified by the algorithm. Therefore the median value, which is more robust to these outliers was calculated too. The median relative error was 0.883.

Besides the pupil masks, the major axes lengths (diameters) of the annotated and predicted ellipses were also compared, which was important, because the original output of the system was the pupillogram, the curve of pupil diameters. The average relative diameter error was 12%, the median relative error was 4%. In Figure 7.a the sorted original and the corresponding predicted diameters are presented. It can be seen that most of the errors occurred when the pupils' diameters were smaller than 20 pixels. This information loss might be caused by the down-sampling section of the U-shaped structure. In Figure 7.b the relationship between the Jaccard-index and the relative diameter error can be observed. Small Jaccard-index value not necessarily implies considerable relative error. This can be explained because if only a small part of the mask is missing, this will reduce the Jaccard-index value (error of the masks). However, it is possible that the missing part only has an influence on the minor axis length of the ellipse, and the major axis length – the extracted diameter value – remains the same.

The diameter measurement accuracy can be improved with post-processing because additional filtering is available while the algorithm is being run on consecutive video frames. This time domain correlation could be utilized in more complex neural network structures too, which could accept a part of a video as their input. This possibility will be investigated in our future work.

Besides the objectively measured improvements, the pupillometry with the new measurement setup and segmentation algorithm will have a positive effect on the
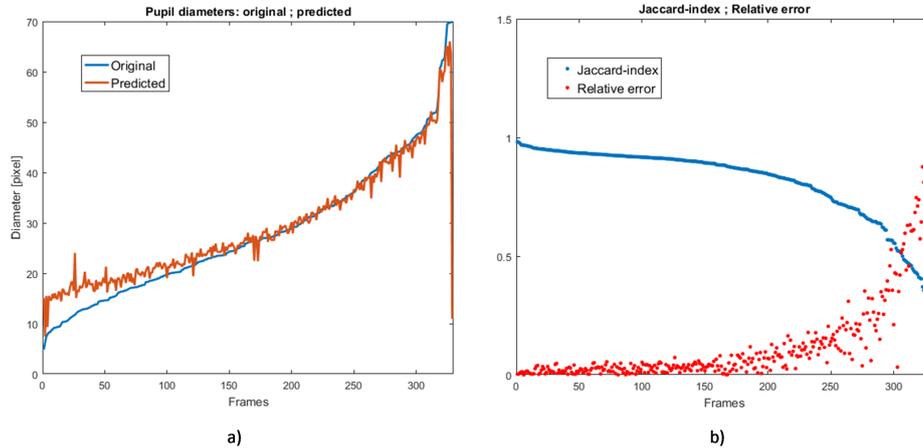
Figure 7: a) Sorted original diameters and the corresponding predicted diameters. b) Sorted Jaccard-indecies and the corresponding relative diameter errors.

classification of the animals too. As the former setup (one impulse/experiment) is inherited in the current measurements (impulse sequences), at least the same features can be extracted. However, the relations between the consecutive impulse responses will unfold more complex details about the alterations of the autonomic nervous system.

7   Concluding remarks

Pupillometry was applied to study the potential schizophrenia-like alterations in the PLR in WISKET rats. A 40-dimensional feature vector was assigned to each recorded video and the dataset was analyzed with decision tree-based classification. The results suggested that pupillary control showed significant alterations in WISKET rats, and the classification based on pupillometry might be applied as an additional examination during the breeding process. Based on the observations, the measurement process was redesigned to induce the "bright pupil effect", which offers more robustness to the experiments. To develop a new pupil segmentation algorithm, a publicly available pupil segmentation dataset was created. A fully-convolutional neural network was trained to produce binary pupil masks. On the test dataset, the relative pupil diameter predictor achieved low, 4% median error.

References

[1] Bär, Karl-Jürgen, Boettger, Michael Karl, Schulz, Steffen, Harzendorf, Christina, Agelink, Marcus Willy, Yeragani, Vikram K, Chokka, Prtap, and

Voss, Andreas. The interaction between pupil function and cardiovascular regulation in patients with acute schizophrenia. *Clinical Neurophysiology*, 119(10):2209–2213, 2008. DOI: `10.1016/j.clinph.2008.06.012`.

[2] Bär, Karl-Jürgen, Koschke, Mandy, Boettger, Michael Karl, Berger, Sandy, Kabisch, Alexander, Sauer, Heinrich, Voss, Andreas, and Yeragani, Vikram K. Acute psychosis leads to increased qt variability in patients suffering from schizophrenia. *Schizophrenia research*, 95(1):115–123, 2007. DOI: `10.1016/j.schres.2007.05.034`.

[3] Büki, Alexandra, Kalmár, György, Kekesi, Gabriella, Benedek, Gyorgy, Nyúl, László G., and Horvath, Gyongyi. Impaired pupillary control in "schizophrenia-like" WISKET rats. *Autonomic Neuroscience*, 213:34–42, 2018. DOI: `10.1016/j.autneu.2018.05.007`.

[4] Fuhl, Wolfgang, Santini, Thiago, Kasneci, Gjergji, and Kasneci, Enkelejda. Pupilnet: convolutional neural networks for robust pupil detection. *arXiv preprint arXiv:1601.04902*, 2016.

[5] Fuhl, Wolfgang, Santini, Thiago, Kasneci, Gjergji, Rosenstiel, Wolfgang, and Kasneci, Enkelejda. Pupilnet v2. 0: Convolutional neural networks for cpu based real time robust pupil detection. *arXiv preprint arXiv:1711.00112*, 2017.

[6] Hermesh, Haggai, Shiloh, Roni, Epstein, Yoram, Manaim, Hillel, Weizman, Abraham, and Munitz, Hanan. Heat intolerance in patients with chronic schizophrenia maintained with antipsychotic drugs. *American Journal of Psychiatry*, 157(8):1327–1329, 2000. DOI: `10.1176/appi.ajp.157.8.1327`.

[7] Horváth, Gyöngyi, Liszli, Péter, Kékesi, Gabriella, Büki, Alexandra, and Benedek, György. Characterization of exploratory activity and learning ability of healthy and 'schizophrenia-like' rats in a square corridor system (ambitus). *Physiology & behavior*, 169:155–164, 2017. DOI: `10.1016/j.physbeh.2016.11.039`.

[8] Horvath, Gyongyi, Petrovszki, Zita, Kekesi, Gabriella, Tuboly, Gabor, Bodosi, Balazs, Horvath, Janos, Gombkötő, Peter, Benedek, Gyorgy, and Nagy, Attila. Electrophysiological alterations in a complex rat model of schizophrenia. *Behavioural brain research*, 307:65–72, 2016. DOI: `10.1016/j.bbr.2016.03.051`.

[9] Kalmár, György, Büki, Alexandra, Kékesi, Gabriella, Horváth, Gyöngyi, and Nyúl, László G. Pupil segmentation dataset. Available at: `https://github.com/Gyoorey/PupilDataset`.

[10] Kalmár, György, Büki, Alexandra, Kékesi, Gabriella, Horváth, Gyöngyi, and Nyúl, László G. Image processing-based automatic pupillometry on infrared videos. *Acta Cybernetica*, 23(2):599–613, 2017. DOI: `10.14232/actacyb.23.2.2017.10`.

[11] Kékesi, Gabriella, Petrovszki, Z, Benedek, György, and Horváth, Gyöngyi. Sex-specific alterations in behavioral and cognitive functions in a "three hit" animal model of schizophrenia. *Behavioural brain research*, 284:85–93, 2015. DOI: `10.1016/j.bbr.2015.02.015`.

[12] Kingma, Diederik P. and Ba, Jimmy. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.

[13] Lidsky, Arnold, Hakerem, Gad, and Sutton, Samuel. Pupillary reactions to single light pulses in psychiatric patients and normals. *Journal of Nervous and Mental Disease*, 1971. DOI: `10.1097/00005053-197110000-00007`.

[14] Lu, Wei, Tan, Jinglu, Zhang, Keqing, and Lei, Bo. Computerized mouse pupil size measurement for pupillary light reflex analysis. *Computer Methods and Programs in Biomedicine*, 90(3):202–209, 2008. DOI: `10.1016/j.cmpb.2008.01.002`.

[15] McLaughlin, Robert A. Randomized hough transform: better ellipse detection. In *TENCON'96. Proceedings., 1996 IEEE TENCON. Digital Signal Processing Applications*, volume 1, pages 409–414. IEEE, 1996. DOI: `10.1109/tencon.1996.608850`.

[16] Mohan, Kabhilan, Harper, Matthew M, Kecova, Helga, Ye, Eun-Ah, Lazic, Tatjana, Sakaguchi, Donald S, Kardon, Randy H, and Grozdanic, Sinisa D. Characterization of structure and function of the mouse retina using pattern electroretinography, pupil light reflex, and optical coherence tomography. *Veterinary ophthalmology*, 15(s2):94–104, 2012. DOI: `10.1111/j.1463-5224.2012.01034.x`.

[17] Morimoto, C.H, Koons, D, Amir, A, and Flickner, M. Pupil detection and tracking using multiple light sources. *Image and Vision Computing*, 18(4):331–335, 2000. DOI: `10.1016/s0262-8856(99)00053-0`.

[18] Paszke, Adam, Gross, Sam, Chintala, Soumith, Chanan, Gregory, Yang, Edward, DeVito, Zachary, Lin, Zeming, Desmaison, Alban, Antiga, Luca, and Lerer, Adam. Automatic differentiation in pytorch. In *NIPS-W*, 2017.

[19] Petrovszki, Zita, Adam, Gabor, Tuboly, Gabor, Kekesi, Gabriella, Benedek, Gyorgy, Keri, Szabolcs, and Horvath, Gyongyi. Characterization of gene–environment interactions by behavioral profiling of selectively bred rats: The effect of NMDA receptor inhibition and social isolation. *Behavioural brain research*, 240:134–145, 2013. DOI: `10.1016/j.bbr.2012.11.022`.

[20] Ronneberger, O., P.Fischer, and Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *LNCS*, pages 234–241. Springer, 2015. DOI: `10.1007/978-3-319-24574-4_28`.

[21] Yuen, HK, Illingworth, John, and Kittler, Josef. Detecting partially occluded ellipses using the Hough transform. *Image and Vision Computing*, 7(1):31–37, 1989. DOI: `10.1016/0262-8856(89)90017-6`.

[22] Zahn, Theodore P and Pickar, David. Autonomic activity in relation to symptom ratings and reaction time in unmedicated patients with schizophrenia. *Schizophrenia research*, 79(2):257–270, 2005. DOI: `10.1016/j.schres.2005.05.025`.

[23] Zhu, Danjie, Moore, Steven T, and Raphan, Theodore. Robust pupil center detection using a curvature algorithm. *Computer methods and programs in biomedicine*, 59(3):145–157, 1999. DOI: `10.1016/s0169-2607(98)00105-9`.