Marek Kowal, Ph.D., D.Sc.
Institute of Control and Computation Engineering
Faculty of Engineering and Technical Sciences
University of Zielona Góra
email: M.Kowal@issi.uz.zgora.pl

Zielona Góra, 17.01.2026r.

Review of the doctoral dissertation

Title: **Using a Camera to Determine Human Gaze Point**

Author: **Mohd Faizan Ansari, M.Sc.**

Supervisor: **Paweł Kasprowski, Ph.D., D.Sc.**

Field: **Technical Sciences**

Discipline: **Information and Communication Technology**

1. Evaluation of the structure of the doctoral dissertation, including information about its individual components

The dissertation consists of 7 chapters, a bibliography, a list of figures, and a list of tables. Chapter one is an introduction to the topic of the doctoral dissertation, where the goals, hypotheses, and scope of the work are presented.

Chapter two presents the theoretical background of the work. It briefly describes the physiology of the eye structure and the process of its movement. Next, basic concepts related to artificial neural networks are defined, such as an artificial neuron, artificial neural network, forward propagation, and backpropagation. The architecture of convolutional neural networks and the concept of transfer learning are also generally discussed. However, this chapter treats the topic too superficially and presents facts about artificial neural networks that are commonly known and usually included in introductions to many textbooks on this subject. It would be better to include here a more in-depth analysis of modern neural networks used for eye movement tracking and gaze direction estimation.

Chapter three presents a review of modern approaches to gaze direction estimation. The review is complete, many literature sources are cited, and different approaches to tracking gaze direction based on eye images are presented. However, a drawback of this part is the rather superficial description of individual methods, especially the limited discussion of approaches using a webcam, which, due to their direct relation to the topic of the dissertation, could be described in more detail. The chapter also includes a review of datasets used in gaze estimation research. Additionally, a review of practical applications of gaze estimation is presented.

From chapter four, the experimental part of the dissertation begins, whose goal is to verify the hypotheses stated in chapter one. Chapter four presents research and results confirming the first hypothesis, according to which a convolutional neural network can successfully classify regions on the screen that the user is looking at. The chapter describes the assumptions of the experiment, the process of data collection and preparation, the architectures of neural networks used for gaze estimation, the training process, and the experimental scenarios. Then the results, their discussion, and conclusions are presented.

Chapter five presents an experiment whose results confirm the second hypothesis, according to which adapting the model to a specific user allows obtaining more accurate gaze estimation results. In this case, the exact gaze direction was estimated and represented by screen coordinates. The chapter is prepared in a style similar to chapter four. It includes an introduction to the experiment, a description of data collection and preparation, a discussion of neural network architectures, and the process of hyperparameter tuning. Then extensive results for different experimental scenarios are presented. At the end, a discussion of the results, conclusions, and the contribution to the development of the discipline resulting from the conducted research are included.

Chapter six is prepared in a style similar to chapters four and five. Its goal is to present the results of an experiment confirming the third hypothesis, related to the use of transfer learning. In particular, it is shown that the use of transfer learning allows reducing the amount of data needed for training, speeding up the learning process, and improving its stability. The data used for pre-training may come from many users, and then the model is adapted to a specific person during the fine-tuning process. After introducing the assumptions of the experiment, a description of the datasets and their initial processing is presented. Next, the model architecture and its variants depending on the type of input data are described. In the next part, detailed experimental results are presented, and the chapter ends with their discussion, conclusions, and indication of the contribution to the development of the discipline.

Chapter seven contains a summary of the work and conclusions resulting from the conducted research. It also presents possible directions for future research. In this chapter, the contribution of the work to the development of the discipline is summarized in several key points. Moreover, the possibilities of practical use of the obtained results and their further development are discussed.

The overall structure of the doctoral dissertation is appropriate. The work begins with an introduction to the topic and a proper formulation of the goal and research hypotheses. A notable drawback is that the theoretical section presents overly elementary issues related to artificial neural networks, rather than offering a more in-depth discussion of solutions used in gaze estimation. The comprehensive review of eye-tracking methods should also be evaluated positively. The following chapters, describing the three main experiments, are prepared clearly. They present important aspects of the research, including results, their discussion, and conclusions. The contribution of the work to the development of the discipline is clearly defined and is consistent with the presented experimental part.

2. Evaluation of the bibliography

The bibliography is extensive and contains 183 references. It is selected correctly, and many of the cited publications refer to current achievements in the topic of the dissertation. The literature review was carried out in a reliable way and includes both classical and the most recent publications in the field of gaze estimation. The doctoral candidate correctly identifies research gaps related to the popularization of gaze estimation methods using commonly available webcams, which forms the basis for the presented original contribution.

In the literature review, different achievements in the area of gaze tracking are considered, dividing them into model-based methods, feature-based methods, and appearance-based methods, and examples of scientific works representing these approaches are provided. The doctoral candidate especially focuses on methods based on deep learning, which currently dominate in this research area. The doctoral candidate also demonstrates a good understanding of the current state of knowledge in this field. A valuable element of the review is the summary of datasets used in gaze estimation research. Additionally, a short review of practical applications of gaze tracking methods is also presented. From the analysis of this part of the work, it can be concluded that the doctoral candidate does not only cite the literature but also critically analyzes it, indicating both important achievements and limitations of the presented solutions.

In the bibliography, sources from the last 10 years dominate, coming from well-known journals and recognized scientific conferences in the field of computer science.

3. Indication and evaluation of the goal of the dissertation

The goal of the dissertation was to investigate the possibility of estimating the user's gaze direction based on images of the eyes and face recorded using a standard webcam with low image quality, without using specialized eye-tracking equipment. This goal was specified by formulating three research hypotheses. The first hypothesis assumes that it is possible to develop a model based on convolutional neural networks that classifies gaze direction into specific regions of the screen based on low-resolution and low-quality images. The second hypothesis concerns the use of models adapted to a specific user and assumes that model personalization allows achieving higher accuracy of gaze estimation, potentially comparable to solutions using specialized equipment. The third hypothesis relates to the use of transfer learning and assumes that using models pre-trained on data from many users makes it possible to reduce the amount of data required to adapt the model to a specific person and to speed up the training process.

My evaluation of the goal of the dissertation is positive. The selected topic is current and important from the point of view of the development of human–computer interaction methods and their practical applications in many areas, such as medical systems, user behavior analysis, or virtual and augmented reality applications. The dissertation correctly identifies the important problem of limited availability of expensive eye-tracking systems and the need to develop solutions based on commonly available devices, such as webcams in laptops or smartphones.

The goal of the dissertation is formulated in a clear way, and its realization is properly connected with the planned research experiments. An important advantage is also the consideration of practical aspects, such as computational limitations of edge devices,

which is reflected in the selection of relatively lightweight neural network architectures. Thanks to this, the dissertation also has some application potential.

4. Evaluation of whether the doctoral dissertation constitutes an original solution to a scientific problem

In the dissertation, three research hypotheses were defined. The hypotheses were formulated correctly, and their verification makes it possible to answer important questions in the field of Information and Communication Technology, especially in the context of using artificial intelligence methods for gaze direction estimation under conditions of limited input data quality. To verify the research hypotheses, an appropriate research plan was prepared, which allows their empirical confirmation. The doctoral candidate demonstrated that he is able to formulate research questions that are important for a given scientific field, translate them into specific hypotheses, and then plan and conduct research to verify them.

The doctoral candidate conducted comprehensive experimental studies, in which the quality of gaze direction estimation was verified for three variants that differ in the way the models were trained. In the first variant, the possibility of classifying gaze direction into defined screen regions was analyzed, while in the following variants the problem was considered as a regression task, estimating the exact coordinates of the gaze point. The studies also included the influence of model personalization (models adapted to a specific user) and the use of transfer learning, which allows using knowledge learned from data of many users to efficiently adapt the model to a new person with a limited number of samples.

In each of the analyzed variants, different models were tested, which differ both in architecture and in the type of input data (eye images and face images), and the influence of selected preprocessing techniques was also considered. Such designed experiments allowed for a multi-aspect analysis of the considered problem. Finally, it was shown that the adopted hypotheses are true and that it is possible to use relatively small convolutional neural network models to estimate gaze direction based on images from a webcam, with accuracy sufficient to consider practical applications.

The originality of the dissertation lies in a comprehensive and systematic study of the possibility of using commonly available webcams to perform gaze estimation, taking into account important factors such as data quality, model personalization, and the use of transfer learning. In my opinion, the doctoral dissertation constitutes an original solution to a scientific problem in an application-research sense, providing valuable conclusions about the effectiveness of the analyzed approaches under conditions close to real-world scenarios.

5. Indication and evaluation of the applied research methods

In the dissertation, experimental research methods were used, which are based on building, training, and comparative analysis of artificial intelligence models designed for gaze estimation. This approach is appropriate for this type of problem and is currently the dominant research methodology in the field of machine learning, where the effectiveness of proposed solutions is verified through numerical experiments and analysis of the obtained results.

The research was carried out using custom datasets, which were collected using a standard webcam. The data included images of users' faces and eyes, and at the same time the point on the screen that the participant was looking at was recorded, implemented by pointing to the appropriate location with a cursor. Such an approach to data acquisition should be considered correct and appropriate for the considered problem, although the description of the structure of the obtained datasets is in some places too general and could be presented in a more systematic way.

In individual experiments, different variants of splitting the data into training and test sets were used (including proportions 70%:30% and 80%:20%), while in the case of experiments using transfer learning, additional subsets for pre-training and fine-tuning of the models were separated. However, it should be noted that the method of data splitting, especially in the context of fine-tuning, was not clearly described, which makes it difficult to assess the potential influence of this process on the obtained results.

To evaluate the quality of the models, appropriate metrics were used, adjusted to the nature of the problem being solved. In the case of the classification task, classification accuracy was used, which is a justified choice, while in the case of the regression problem, mean absolute error (MAE) and angular error were applied, which are standard measures in gaze estimation tasks. The selection of metrics should be assessed as correct, although the dissertation could be improved by including a wider set of quality measures, which would allow for a more complete analysis of the obtained results.

Within the research, different variants of convolutional neural network architectures and different configurations of input data were analyzed, which made it possible to compare their influence on estimation quality. In one of the experiments, a comparison with selected modern methods was also performed, however, the scope of these comparisons is limited and could be extended, especially in relation to methods considered as state of the art.

In summary, the applied research methods are appropriate for the stated problem and were selected correctly. Despite the indicated remarks concerning the description of data, validation procedures, and the scope of comparisons, the overall research approach should be assessed positively, as it allowed for a reliable analysis of the effectiveness of the proposed solutions.

6. Evaluation of the part of the dissertation concerning the discussion of research results

The doctoral candidate devoted a significant part of the dissertation to the presentation and discussion of the research results. A detailed analysis of the results is presented in Chapters 4, 5, and 6, where partial results, their interpretation, discussion, and final conclusions are consistently shown, together with an indication of the contribution of individual experiments to the development of the discipline.

The research results are presented according to accepted standards for doctoral dissertations, in tabular and graphical form, supplemented with textual descriptions and analysis of selected cases. This way of presentation allows the reader to follow the influence of individual factors, such as the choice of model architecture, type of input data, or the use of transfer learning, on the quality of gaze estimation.

It is worth emphasizing that the doctoral candidate does not limit himself to presenting numerical results, but also attempts to interpret them, indicating possible reasons for the observed relationships. This applies in particular to the comparison of classification and regression approaches, the analysis of the influence of model personalization on estimation accuracy, and the evaluation of the effectiveness of transfer learning under conditions of a limited number of training data. The conclusions formulated in individual chapters are consistent with the obtained results and form a logical summary of the conducted experiments.

However, it should be noted that the dissertation could benefit from a more detailed analysis of the model training process, especially by presenting plots of changes in loss and quality metrics over successive training epochs. This would allow for a more complete evaluation of the stability and convergence of the training process.

Despite this remark, I assess that the part of the dissertation devoted to the analysis of research results has been prepared in a reliable way. The presented results and their interpretation provide a sufficient basis to conclude that the stated research hypotheses have been reliably verified.

7. Practical applications of the obtained research results

The doctoral candidate does not present a direct implementation of the developed methods in a specific application system, however, in the dissertation he indicates potential areas of their practical application. In particular, in subsection 3.5 a wide range of possible applications of gaze direction estimation is presented, including areas such as human-computer interaction (HCI), virtual reality (VR) and augmented reality (AR) applications, analysis of user behavior in marketing, applications in healthcare, education and learning, as well as in computer games and digital entertainment, monitoring and security systems, sports, and art.

The presented examples are convincing and well justified in the context of current trends in technology development. It is especially important to point out the possibility of using solutions based on standard webcams, which significantly reduces the entry barrier and implementation costs compared to solutions requiring specialized eye-tracking equipment.

It is worth emphasizing that the obtained research results may form a basis for building systems that operate in real conditions, especially in the context of devices with limited computational resources. In this way, the dissertation fits into the research direction focused on practical use of artificial intelligence methods and may contribute to the development of solutions that enable automation of processes and reduction of costs of technologies based on gaze direction analysis.

8. Evaluation of whether the dissertation presents general theoretical knowledge of the candidate in the discipline and the ability to conduct independent scientific work

The doctoral candidate demonstrated that he is able to independently identify an important research problem in the discipline of Information and Communication Technology, and then place it in the current state of scientific knowledge. In the dissertation, knowledge of the basics related both to gaze direction estimation and to

machine learning methods is visible, especially convolutional neural networks and transfer learning.

Based on the conducted literature review and his own analysis, the doctoral candidate formulated correct research hypotheses and then planned and carried out a consistent program of experimental research. This includes data preparation, selection and implementation of models, as well as analysis of the obtained results. An important element of the work is also the ability to interpret the results and to formulate conclusions that are logically connected with the conducted experiments.

The entire conducted research shows that the doctoral candidate has the ability to carry out independent scientific work, from the stage of identifying the problem, through its formalization, to experimental verification and critical analysis of the results. Despite some remarks concerning selected methodological aspects, the overall level of the work confirms the scientific maturity of the candidate.

It is also worth emphasizing the publication achievements of the doctoral candidate, which include articles published in peer-reviewed scientific journals (5 publications), conference proceedings (1 publication), and a chapter in a monograph. This confirms scientific activity and the ability to present research results in the scientific community.

9. Issues identified in the doctoral dissertation

Despite the overall positive evaluation of the dissertation and recognition of its contribution to the development of the discipline of Information and Communication Technology, several critical remarks can be indicated, mainly related to the adopted research methodology and the way the results are presented:

1. The doctoral candidate demonstrates the possibility of building effective gaze estimation models based on relatively small convolutional neural networks. These results were obtained on custom datasets prepared as part of the work (in total three datasets, with the largest including data from 19 participants). The creation of such datasets should be considered a valuable element of the dissertation. However, the work would gain value if experiments were also conducted on publicly available datasets that serve as benchmarks in this field. Especially since in subsection 3.4 their review is presented. The use of such datasets would allow for a more reliable comparison of the proposed solutions with methods considered as state of the art. In the current form, such comparisons are very limited, and only one comparison (Table 5.9) is included in the dissertation, which is additionally based on reference results obtained on different datasets, which significantly makes comparative interpretation more difficult.

2. The description of the custom datasets is not sufficiently detailed. In many places, basic information is missing, such as the number of samples, dataset structure, distribution of data between participants, or data acquisition conditions (lighting, background, camera setup). There are also no examples illustrating the diversity of the data. For example, in Chapter 4 the data collection procedure is described in detail, however, a consistent characterization of the created dataset (especially dataset D2) is not provided. Only in Chapter 6 (Table 6.1) more detailed information about the data structure is presented, which indicates a lack of consistency in their description.

3. The procedures of splitting data into training and test sets are not described in sufficient detail. Based on the provided information, it is not possible to clearly determine whether the test sets contained samples from the same participants as the training sets. Such a situation could lead to data leakage and overestimation of experimental results. A more appropriate approach would be to use a subject-level split, where data from one person appears only in one of the sets.

4. In the experiment described in Chapter 4, classification accuracy was used as the main quality metric. This choice is justified, however, other standard metrics used in classification tasks, such as precision, recall, F1-score, confusion matrix, or ROC/AUC curves, were not used in the dissertation. Their use would allow for a more complete evaluation of model performance.

5. In Tables 4.3-4.5, classification accuracy values are presented, but it is not clearly indicated whether they refer to training, validation, or test data. Additionally, a non-standard way of reporting results was used by averaging accuracy over groups of training epochs. A more standard approach would be to present the training process using plots (accuracy and loss over epochs) and to report final results for the best model on the test set.

6. The choice of the number of training epochs (for example 70 epochs in the experiment in Chapter 4) is not justified. From the presented results, it can be seen that model accuracy increases with the number of epochs, which suggests that further training could improve the results. There is no analysis of training convergence and no plots illustrating the training process.

7. In the experiment described in Chapter 5, the training process of the models is not presented, and the lack of plots showing changes of loss or error over epochs makes it difficult to assess the stability and quality of the training.

8. In the experiments described in Chapters 4 and 6, detailed information about the training hyperparameters and the method of their selection is not provided. This limits the possibility of reproducing the experiments and evaluating the influence of these parameters on the results.

9. In the experiments in Chapters 5 and 6, the MAE error is presented in absolute units (screen pixels). A more universal approach would be to normalize this metric with respect to screen resolution, which would allow better comparison of results between different configurations.

10. For eye region detection, a classical approach based on Haar cascades was used. The dissertation does not justify this choice, while modern lightweight models based on neural networks (for example YOLO) could be an alternative offering higher effectiveness and robustness to changing conditions.

11. The dissertation also lacks a broader justification of the method used to select the architectures of the applied neural networks. It is not clear whether this choice resulted from a systematic analysis or had an experimental character.

10. Remarks on editorial and formatting errors

In the dissertation, there are also minor editorial and formatting issues, which, however, do not affect its overall substantive evaluation:

p. 5 – typo: "weighs"
p. 18 – incorrect or incomplete reference: "equation ??"
p. 40 – incorrect reference to literature: "citezhang2021eye"
p. 58 – typo: "aystem"
p. 61 – repeated paragraph of text
p. 63 – incorrect combination of captions: "Figure 5.3 Figure 5.4"
p. 64 – incorrect table numbering: "Table 4.2"
p. 69 – paragraph repeated on p. 71
p. 70 – repeated fragment of text
p. 84 – typo: "particiapnt"

11. Final conclusions

The reviewed doctoral dissertation addresses a current and important scientific problem from the point of view of the discipline of Information and Communication Technology, related to gaze estimation using low-quality images obtained from standard webcams. The research results presented in the dissertation constitute a valuable contribution of an application-research nature, especially in the area of analyzing the possibility of using lightweight convolutional neural network models, model personalization, and the use of transfer learning under conditions of limited data availability.

The doctoral candidate demonstrated the ability to independently formulate research problems, plan and conduct experiments, and critically analyze the obtained results. The conducted research is systematic and allows for reliable verification of the stated hypotheses. Despite the critical remarks indicated in this review, mainly concerning selected methodological aspects and the way of presenting results, the overall evaluation of the dissertation remains positive.

The dissertation confirms that it is possible to develop solutions that enable gaze direction estimation using commonly available devices, which is important for the further development of human–computer interaction systems and their practical applications.

In conclusion, I state that the reviewed doctoral dissertation meets the requirements for doctoral theses specified in the applicable regulations. Therefore, I recommend admitting the dissertation to the public defense.