

2026.07.13.

Review of the PhD dissertation

„Novel deep learning methods for optical recognition”

submitted by Richárd Bence Rádli

Andrea Manno-Kovács

The dissertation addresses three computer vision problems connected to deep learning models. The first two are mainly theoretical, however both of them have significant practical applications, the third task is more practical and introduces an improved solution for pill recognition. The proposed models are mathematically sound and validated on multiple databases with convincing experimental results. The comparisons with the state-of-the-art also show the introduced models' superiority.

I. Presentation

The dissertation consists of five chapters. The introduction (Chapter 1) gives a general overview on artificial neural networks, learning methods and selected machine learning tasks. It also introduces the working environment and the three main tasks concerned in the dissertation. Chapter 2 – 4 present the main scientific contributions:

- A novel framework to improve visual defect detection efficiency of autoencoders by integrating spatial transformer networks to normalize image orientation.
- A novel framework for rapid training using scalable construction and iterative refinement of deep randomized neural networks.
- A multi-stream, two-phase neural network model for pill recognition employing multihead attention modules and high-level (free-text and visual appearance) cues used in the training.

Finally, Chapter 5 concludes the scientific contributions and lists the author's publications.

The manuscript is well-written and clearly organized, it is easy to read and follow. The whole work is very thorough and the figures and tables are well constructed. However, it should be noted that some of my previous comments were only addressed in the response letter, but the manuscript has not been extended or updated, which would have been also useful.

II. Scientific content

The scientific contributions of the manuscript are summarized in three thesis groups. These contributions include novel findings and also have practical applications in optical recognition and anomaly detection using deep learning techniques.

Thesis 1

The motivation is to enhance autoencoder architectures to improve visual defect detection abilities. The work is inspired by a real-world problem, existing for example during manufacturing. As a novelty, the thesis introduces a framework for visual defect detection applying autoencoders and as a further improvement, spatial transformer networks (STNs) are integrated into the framework for

orientation normalization. The performance of the proposed framework is thoroughly tested and verified on multiple datasets.

Questions, comments on Thesis 1:

How robust is the proposed STN in case of bigger rotations or distortions?

Thesis 2

A novel deep randomized neural network based framework is introduced with faster training and more accurate classification abilities. The method is further improved by a scalable pruning algorithm for eliminating the least contributing neurons and replacing them with the most contributing neurons of an auxiliary network. The proposed networks are thoroughly tested on multiple datasets and compared to the two most widely used gradient descent-based optimizers, Adam and SGD.

Questions, comments on Thesis 2:

Is the proposed method scalable for other models, like Vision Transformers? Can deep randomized neural networks replace backpropagation techniques in other deep learning models?

Thesis 3

A multi-stream two-phase neural network is introduced for metric-based pill recognition. As a further improvement, high-level free-text and visual appearance cues are integrated into the training of such models to enhance accuracy. The integration of these cues is a novel and smart idea, the performance improvement is verified on two databases. It should be mentioned that one of the two databases is completely collected and created by the author (and his colleagues) which is an enormous effort.

Questions, comments on Thesis 3:

The author responded earlier that the created OGYElv2 database is public and available to the general public. It would have been useful to add the link of the database to the dissertation as well.

Why was the EfficientNetV2-S network used as backbone? Were there any experiments or other considerations behind this choice?

How sensitive is the proposed model to incomplete or inaccurate textual description information?

III. Conclusion

The author has already two published journal papers and a third one is under revision on Thesis group 2. He also has 6 international conference papers. It should be mentioned that only 2 conference papers correspond to Thesis 1, while the other two theses are (planned to be) supported by journal papers. It would have been beneficial to publish the novelties of Thesis 1 also in a journal paper. In summary, the introduced scientific contributions, the presented manuscript and the supporting publications are up to standards required by a PhD degree. **Therefore, I recommend proceeding with the oral defense of the thesis and I support the award of the PhD degree to the candidate.**

Budapest, 2026. 07. 13.



Dr. Manno-Kovács Andrea