



Universiteit
Leiden

The Netherlands

System-level design for efficient execution of CNNs at the edge

Minakova, S.

Citation

Minakova, S. (2022, November 24). *System-level design for efficient execution of CNNs at the edge*. Retrieved from <https://hdl.handle.net/1887/3487044>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3487044>

Note: To cite this publication please use the final published version (if applicable).

STELLINGEN

Propositions belonging to the Ph.D. dissertation:

System-level Design For Efficient Execution of Convolutional Neural Networks at the Edge

by Svetlana Minakova

1. To demonstrate high inference throughput when executed on a multi-core platform, a Convolutional Neural Network should exploit both task-level parallelism and data-level parallelism available in it. (Chapter 3)
2. The memory footprint of a Convolutional Neural Network (CNN) can be reduced by splitting the data processed by layers of the CNN in parts, efficiently scheduling the processing of the data parts, and efficiently reusing the platform memory among the data parts. (Chapter 4)
3. If an application based on a Convolutional Neural Network (CNN) needs to adapt its resource utilization at run-time, it should: 1) interchangeably use multiple alternative CNNs, optimized individually and jointly in terms of both accuracy and platform resources utilization; 2) have means for efficient run-time switching among the alternative CNNs. (Chapter 5)
4. To significantly reduce the memory footprint of an application which uses multiple Convolutional Neural Networks (CNNs) and to avoid decreasing accuracy of the application, efficient reuse of platform memory between and within layers of the CNNs is required. (Chapter 6)
5. To ensure efficient execution of a Convolutional Neural Network (CNN) on an edge platform, the characteristics of the edge platform must be considered during the CNN design.
6. Artificial intelligence (AI) at the Edge is at the intersection between the field of AI and the field of embedded systems. Therefore, research in this area should (re-)use methodologies and tools not only from the field of AI, but also from the field of embedded systems.
7. To efficiently distribute computations in a Convolutional Neural Network (CNN) over computational resources of an (edge) platform, one should represent the CNN as a parallel model of computation (e.g., a Cyclo-Static Dataflow model) rather than as the widely used graph CNN model.
8. In computer science research, a good (software) implementation is often just as important as the idea or model it implements.
9. A successful research requires both competition and collaboration with other researchers as well as a good balance between the competition and the collaboration.
10. The ability to come up with bright ideas does not entail the ability to turn an idea into a good research paper, and vice versa.