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

Minakova, S.

# System-level Design For Efficient Execution of CNNs at the Edge

Svetlana Minakova

This work has received funding from the European Unions Horizon 2020 Research and Innovation project under grant agreement No. 780788.

System-level Design For Efficient Execution of CNNs at the Edge. Svetlana Minakova. - Dissertation Universiteit Leiden.


This dissertation was typeset using LaTeX.
Cover design: from images generated using DALL·E mini Deep Learning algorithm [20]. The images are combined and post-processed by Anna Minakova.

# System-level Design For Efficient Execution of CNNs at the Edge

**Proefschrift**

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op donderdag 24 november 2022
klokke 11.15 uur

door

Svetlana Minakova
geboren te Ryazan, Rusland
in 1993

**Promotores:**
Dr. T.P. Stefanov
Prof.dr. H.A.G. Wijshoff

**Promotiecommissie:**
Prof.dr. S. Ha (Seoul National University)
Prof.dr. J. Castrillon (Technical University of Dresden)
Prof.dr. H.E. Bal (Vrije Universiteit Amsterdam)
Prof.dr. A. Plaat
Prof.dr. N. Mentens
Prof.dr. M.S.K. Lew

*To my family and friends*

# Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| BFS | Breadth-First Search |
| CNN | Convolutional Neural Network |
| CPU | Central Processing Unit |
| CSDF | Cyclo-Static Data Flow |
| DL | Deep Learning |
| DSE | Design Space Exploration |
| FC | Fully Connected |
| FLOP | floating-point operation |
| FPGA | Field-programmable Gate Array |
| GA | Genetic Algorithm |
| GPU | Graphics Processing Unit |
| HAR | Human Activity Recognition |
| IoT | Internet-of-Things |
| KD | Knowledge Distillation |
| MB | MegaBytes |
| MoC | Model of Computation |
| MPSoC | Multi-Processor System-on-Chip |
| NAS | Neural Architecture Search |

| | |
|---|---|
| ONNX | Open Neural Network Exchange Format |
| SBRS | Scenario-Based Run-time Switching |
| SDF | Synchronous Data Flow |
| SOTA | State Of The Art |
| SSR | Scenario Switch Request |
| TPU | Tensor Processing Unit |
| UAV | Unmanned Aerial Vehicle |